The Dissertation Committee for James Brian Pitts certifies that this is the approved version of the following dissertation:

Gravitation with a Flat Background Metric

Committee:

William C. Schieve, Supervisor

Lawrence P. Horwitz

Cory F. Juhl

Richard A. Matzner

Lawrence C. Shepley

Matthew W. Choptuik

Gravitation with a Flat Background Metric

by

James Brian Pitts, B.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2001

This dissertation is dedicated to my parents William and Jan Pitts for all their support.

Acknowledgments

J. B. P. thanks Prof. W. C. Schieve for dissertation supervision in its many facets.

He also thanks Profs. L. P. Horwitz, M. Choptuik, C. Juhl, R. Matzner, and L. Shepley for service on the dissertation committee and discussions; Prof. W. L. Craig for introduction to the philosophy of time and for correspondence; R. Hayes for making the null cone figure and for computer assistance; Dr. S. H. Hawley for computer assistance and for reading manuscripts; Prof. I Goldman for discussions and for providing results from his unpublished and untranslated dissertation; Prof. M. Visser for correspondence and manuscript reading; Profs. C. Brans, S. Deser, T. Durt, R. Kane, F. Kronz, A. A. Logunov, C. Rovelli, D. C. Salisbury, G. Scharf, L. E. Szabó, and Drs. J. Barbour, L. Lehner and A. Pruss for helpful discussions or correspondence; Prof. S. Weinberg for discussion; R. Takenaga, R. Luter, Dr. T. Yudichak, and Prof. V. A. Petrov for assistance; the Center for Relativity for the use of computer facilities; the University libraries; and the Ilya Prigogine Center for Studies in Statistical Mechanics and Complex Systems and the Welch Foundation for financial support.

James Brian Pitts

The University of Texas at Austin

August 2001

Gravitation with a Flat Background Metric

Publication No. ____________________

James Brian Pitts, Ph.D.
The University of Texas at Austin, 2001

Supervisor: William C. Schieve

Although relativistic physics tends to omit nondynamical "absolute objects" such as a flat metric tensor or a preferred time foliation, there exist interesting questions related to such entities, such as worries about the "flow" of time in special relativity, and the apparent disappearance of time altogether in canonical general relativity. This latter problem is related to the lack of a fixed causal structure with respect to which one might posit "equal-time" commutation relations, for example.

In view of these issues, we consider whether including a flat background metric, and perhaps a preferred foliation, is physically worthwhile. We show how a derivation of Einstein's equations from flat spacetime can be generalized to include a preferred foliation, the possible significance of which we discuss, though ultimately we suggest why such a foliation might be present in metaphysics and yet absent from physics. We also derive a new "slightly bimetric" class of theories using the flat spacetime approach.

However, such derivations are only *formally* special relativistic, because they give no heed to the flat metric's causal structure, which the curved effective metric might well violate. After reviewing the history of this problem, we introduce new variables to give a kinematic description of the relation between the two null cones. Then we propose a method to enforce special relativistic causality by using the gauge freedom to restrict the configuration space suitably. Consequences for exact solutions, such as the Schwarzschild solution and its 'singularity', are discussed.

Advantages and difficulties regarding adding a mass term to the theory are discussed briefly.

Contents

List of Figures

Chapter 1

Introduction

While relativistic physics tends to omit nondynamical "absolute objects" such as a flat metric tensor or a preferred time foliation[1], there exist interesting questions related to such entities. Concerning general relativity, the lack of an *a priori* fixed causal structure is merely technically demanding at the classical level, but it constitutes a real puzzle at the quantum level, for one no longer knows how to write equal-time commutation relations, for example, because one needs to know the metric in order to determine equal times, but the metric is itself quantized: a chicken-and-the-egg problem. This is an aspect of the problem of time in quantum gravity. Furthermore, one wonders whether gravity need be formulated in a way that differs so drastically from the other forces; can it not be treated as a force in flat spacetime? In chapter 2 we therefore review the flat spacetime approach to Einstein's equations, presenting a derivation based on universal coupling and gauge invariance, which then serves as the model for various generalizations. This work will appear in the near future in *General Relativity and Gravitation* in a paper cowritten with W. C. Schieve, and is used with the permission of Kluwer Academic Publishing.

Concerning special relativity, the relativity of distant simultaneity fits poorly

[1]Absolute objects have been discussed in ([1–3]).

with intuitive notions of the "flow" of time. With that issue in mind, in the third chapter we generalize the previous derivation to consider how a preferred foliation might manifest itself in the flat spacetime approach to gravitation in either a $4+1$-dimensional "parametrized" theory with an invariant global time τ, or in a $3+1$-dimensional preferred-frame theory. The functional forms that the actions for such theories might take are suggested. The relevance of these theories to the actual world, however, is not so obvious, given the remarkable empirical adequacy of both special relativity and Einstein's theory. In the $4+1$-dimensional case, experiments suggest that the τ-dependence of the world must be rather slow. Otherwise, light dispersion should be observed due to the modified dispersion relation. In the $3+1$-dimensional case, in view of the constructive role of Lorentz invariance in theoretical physics, and the limited (though perhaps nonzero) physical evidence for difficulties with Lorentz invariance, one is reluctant to give up that symmetry needlessly. However, there are philosophical arguments that might motivate one to posit a preferred foliation. Can one then suggest that physics is Lorentz-invariant but metaphysics is not? Such a suggestion has been thought to run afoul of Einstein's maxim "Subtle is the Lord, but not malicious." On the contrary, we attempt to suggest why such a suggestion need not imply a perverse world, based on the very utility of Lorentz invariance. This work will appear (jointly written with W. C. Schieve) in the near future in an issue of *Foundations of Physics* devoted to a conference proceedings edited by L. P. Horwitz, and is used with the permission of Kluwer Academic Publishing.

Having perhaps reconciled the flow of time with special relativity, we turn back to gravitation in special relativity in chapter 4. Flat spacetime derivations of general relativity have been known for almost half a century, so merely presenting another would seem a bit sterile. It is natural to ask whether any interesting generalizations of this procedure can be found. The answer is "yes." We find that restricting the gauge invariance and requiring only the traceless part of the stress tensor to be coupled to gravity yields a larger family of "slightly bimetric" theories, in which the determinant of the flat background metric appears. Such theories have an arbitrary cosmological constant and can have a third degree of freedom per space point. This work will appear in the near future in the above-mentioned *General Relativity and Gravitation* paper cowritten with W. C. Schieve.

Such derivations of general relativity and slightly bimetric theories, however, are only *formally* special relativistic, because the curved null cone might not respect the flat one. This difficulty afflicts not merely our derivation, but in fact all derivations in this tradition, and implies that the alleged resemblance of Einstein's theory to other field theories in this approach is merely formal. In chapter 5 we survey in some detail the treatment of this fundamental question over the last six decades. As it happens, this issue has in general been ignored, explained away, postponed with the hope that it would go away, or mishandled in one of several ways, although there have been positive signs in recent years. We critique claims that the problem is insoluble and claims that it has already been solved, and conclude that the issue remains quite open.

In chapter 6 we undertake to solve the problem. The kinematic issue of the relationship between the two null cones is handled using the work of G. S. Hall and collaborators on the Segré classification of symmetric rank 2 tensors with respect to a Lorentzian metric. For our purposes, we classify the curved metric with respect to the flat one, and find necessary and sufficient conditions for a suitable relationship. Requiring that flat spacetime causality not be violated, and not be arbitrarily close to being violated, a condition that we call "stable η-causality", implies that all suitable curved metrics have a complete set of generalized eigenvectors with respect to the flat metric, and that the causality conditions take the form of *strict* inequalities. Given strict inequalities, one is in a position to solve such conditions, which are somewhat analogous to the "positivity conditions" of canonical gravity, which have been discussed by J. Klauder, F. Klotz, and J. Goldberg. In these new variables, stable η-causality holds *identically*, because the configuration space has been reduced. This reduction implies the need for reconsidering the gauge freedom of the theory. It turns out that gauge transformations no longer form a group, because multiplication is not defined between some elements. A portion of the work in chapters 5 and 6 will appear in the *Proceedings of the 20th Texas Symposium on Relativistic Astrophysics*[117] in a paper cowritten with W. C. Schieve. This portion is used by permission of the American Institute of Physics.

Given the need to respect the flat metric's null cone and reduction in gauge freedom, the usual treatment of the Schwarzschild 'singularity' is called into question, but the resolution is not yet clear, we find in chapter 6. While a causally acceptable solution containing a region of small r from which not even light can escape is known, the relevance of this solution to objects formed from gravitational collapse needs to be explored. Moreover, making the curved metric respect the flat null cone ensures that the resulting spacetime is *globally hyperbolic*. Global hyperbolicity implies that any regions of no escape lack some of the typical properties of black holes in the geometrical formulation. Given that global hyperbolicity apparently pulls the fangs from the Hawking black hole information loss paradox, it appears that this paradox does not afflict the special relativistic approach to gravitation.

Finally, we discuss some noteworthy exact solutions and contemplate the possibility that a rest mass term should be added to the theory. "Caustic plane wave" exact solutions of Einstein's equations appear to be an embarrassment for our program of taking the flat metric seriously, because these solutions seem to imply severe η-null cone violation. We suggest, however, that one can make a gauge transformation to alter parts of these solutions to an acceptable form, while the rest is excluded from spacetime as unphysical. As in the case of the Schwarzschild singularity, long-standing (and resolved) coordinate singularities in the geometrical theory might imply true singularities in our approach. The field of an infinite plate is also discussed. It turns out that such a plate must be repulsive, a result which is not so surprising in view of an analogy to the Schwarzschild solution. If one gives gravitation a finite range using the Maheshwari-Logunov mass term, then an attractive plate is possible. In view of the many desirable features of massive general relativity from the bimetric point of view (such as that the flat metric becomes observable, that energy and momentum are localized, and that the theory in a sense lacks constraint equations), it is disappointing the massive theory's lack of gauge freedom prevents the use of our approach to ensuring null cone consistency. Suggestions for restoring enough gauge freedom to use our procedure are made, but they face difficulties of a physical or at least technical nature.

Chapter 2

Flat Spacetime and Einstein's Equations

2.1 Previous Use of a Flat Background Metric in Gravitation

A number of authors have discussed the utility of a flat background metric $\eta_{\mu\nu}$ in general relativity or the possibility of deriving that theory, approximately or exactly, from a flat spacetime theory [4–73, 75–114]. Some have permitted the background metric to be curved [69, 82, 94, 95, 118–122], but our interest is in flat backgrounds only, because they are uniquely plausible as nondynamical entities. The use of a flat background metric enables one to formulate a gravitational stress-energy tensor [93], not merely a pseudotensor, so gravitational energy and momentum are localized in a coordinate-independent (but gauge-variant) way. It also enables one to derive general relativity and other generally covariant theories from plausible special-relativistic postulates, rather than postulating them. (We call a theory "generally covariant" if no nondynamical fields appear in the Euler-Lagrange equations, even if some do appear in the action.) As W. Thirring observed, it is not clear *a priori* why Riemannian geometry is to be preferred over all the other sorts of geometry that exist, so a derivation of effective Riemannian geometry is attractive [29]. It also seems appealing to try to cast gravitation in the same form as the other forces [29]. Furthermore, a non-geometrical form of gravitation can facilitate introduction of supersymmetry [83]. We will also find that this approach avoids the difficulty of the lack of an *a priori* causal structure for defining dynamics in quantum gravity.

2.2 Generally Covariant Theories from Universal Coupling and Infinitesimal Free Field Gauge Invariance

To such a derivation of generally covariant theories we now turn. Our derivation combines elements familiar from the work of Kraichnan [20, 21] and Deser [65], but it has improvements as well. It is based upon universal coupling and an assumed initial infinitesimal invariance (up to a boundary term) of the free gravitational action. This derivation will also serve as the model for the new derivation of slightly bimetric theories and theories with a preferred foliation. The assumption of gauge invariance requires that the field be massless. However, it can easily be modified to produce massive theories, too.[1]

2.3 Free Field Action

Let S_f be the action for a free symmetric tensor field $\gamma_{\mu\nu}$ (of density weight 0) in Minkowski spacetime with metric tensor $\eta_{\mu\nu}$ in arbitrary coordinates. The torsion-free η-compatible covariant derivative is denoted by ∂_μ. The field $\gamma_{\mu\nu}$ will turn out to be the gravitational potential. We require that S_f change only by a boundary term under the infinitesimal gauge transformation $\gamma_{\mu\nu} \rightarrow \gamma_{\mu\nu} + \delta\gamma_{\mu\nu}$, where

$$\delta\gamma_{\mu\nu} = \partial_\mu \xi_\nu + \partial_\nu \xi_\mu, \tag{2.1}$$

[1] J. B. Pitts, in preparation.

ξ_ν being an arbitrary covector field. In the special case that the Lagrangian density is a linear combination of terms quadratic in first derivatives of the $\gamma_{\mu\nu}$, and free of algebraic and higher-derivative dependence on $\gamma_{\mu\nu}$, the requirement of gauge invariance uniquely fixes coefficients of the terms in the free field action up to a boundary term, giving linearized vacuum general relativity [304].[2]

For any S_f invariant in this sense under (2.1), the free field equation is identically divergenceless, as we now show. With arbitrary divergences $e^\mu{}_{,\mu}$ and $f^\mu{}_{,\mu}$ permitted, the action changes by

$$\delta S_f = \int d^4x \left[\frac{\delta S_f}{\delta \gamma_{\mu\nu}} (\partial_\nu \xi_\mu + \partial_\mu \xi_\nu) + e^\mu{}_{,\mu} \right] = \int d^4x f^\mu{}_{,\mu} \, . \tag{2.2}$$

The explicit forms of the boundary terms are not needed for our purposes. Integrating by parts, letting ξ^μ have compact support to annihilate the boundary terms (as we shall do throughout this dissertation), and making use of the arbitrariness of ξ^μ, we obtain the identity

$$\partial_\mu \frac{\delta S_f}{\delta \gamma_{\mu\nu}} = 0. \tag{2.3}$$

2.4 Metric Stress-Energy Tensor

If the energy-momentum tensor is to be the source for the field $\gamma_{\mu\nu}$, consistency requires that the *total* stress tensor be used, including gravitational energy and momentum, not merely the nongravitational ("matter") sort, for only the total stress tensor is divergenceless in the sense of ∂_ν [65], or, equivalently, in the sense of a Cartesian coordinate divergence. To obtain a global conservation law, one needs a vanishing *coordinate* divergence for the 4-current. In general relativity in its geometrical form, one must choose between tensorial expressions and global conservation laws. If one employs only tensors (or tensor densities), one can write $\nabla_\mu T^{\mu\nu}_{\mathrm{mat}} = 0$ for the matter stress tensor (where ∇_μ is the usual torsion-free g-compatible covariant

[2] For related work, one might see Wald [128] and Heiderich and Unruh [129].

derivative). But this equation typically does not yield a global conservation law [152], because in general it cannot be written as a coordinate divergence. (From the flat spacetime viewpoint, this equation is best regarded as a force law, not a conservation equation.) If coordinate-dependent expressions are admitted, then one can write $\tau^{\mu\nu}{}_{,\mu} = 0$, where $\tau^{\mu\nu}$ is some nontensorial complex that includes gravitational as well as matter stress [208, 209]. But these objects behave oddly under coordinate transformations [1, 210–213, 215, 216]. A flat background metric, in contrast, yields *tensorial global* conservation laws, as Rosen has emphasized [216, 217]. However, the gravitational energy and momentum localization is now gauge-dependent.

An expression for the total stress tensor can be derived from S using the metric recipe [1, 20, 21, 93] in the following way. The action depends on the flat metric $\eta_{\mu\nu}$, the gravitational potential $\gamma_{\mu\nu}$, and "bosonic" matter fields u. Here u represents an arbitrary collection of dynamical tensor (density) fields of arbitrary rank, index position, and weight. Under an arbitrary infinitesimal coordinate transformation described by a vector field ξ^{μ}, the action changes by the amount

$$\delta S = \int d^4x \left(\frac{\delta S}{\delta \gamma_{\mu\nu}} \pounds_{\xi} \gamma_{\mu\nu} + \frac{\delta S}{\delta u} \pounds_{\xi} u + \frac{\delta S}{\delta \eta_{\mu\nu}} \pounds_{\xi} \eta_{\mu\nu} + g^{\mu}{}_{,\mu} \right). \tag{2.4}$$

But S is a scalar, so $\delta S = 0$. Letting the matter and gravitational field equations hold gives

$$\delta S = \int d^4x \left(\frac{\delta S}{\delta \eta_{\mu\nu}} |\gamma \right) \pounds_{\xi} \eta_{\mu\nu} = 0, \tag{2.5}$$

or

$$\partial_{\mu} \frac{\delta S}{\delta \eta_{\mu\nu}} |\gamma = 0. \tag{2.6}$$

(The bar notation in $\frac{\delta S}{\delta \eta_{\mu\nu}}|\gamma$ emphasizes that the other independent variable is $\gamma_{\mu\nu}$, a fact that will become important shortly.)

This metric stress tensor density $\mathfrak{T}^{\mu\nu} = 2\frac{\delta S}{\delta \eta_{\mu\nu}}$ agrees with the symmetrized canonical tensor in the case of electromagnetism, up to a trivial factor (assuming the electromagnetic potential to be a covector of vanishing density weight, *i.e.*, a 1-form; otherwise, terms that vanish when the equations of motion hold also arise). In more general cases, the relation between the metric and symmetrized canonical results is more complicated, so some ambiguity in the term "stress tensor" exists; one could try to resolve this ambiguity by introducing further criteria [1, 93, 218].

2.5 Choice of Dynamical Variables

Deser treated the gravitational potential and $\{^{\mu}_{\alpha\beta}\}$ as independent variables, giving a first-order Lagrangian formalism [65]. This approach, which lacks Lagrange multipliers to enforce the Levi-Civita character of the connection, can be made to work if one is clever, but we prefer using only $\gamma_{\mu\nu}$ as the independent variable, as in Kraichnan's second-order Lagrangian approach [20]. There are several reasons for our preference. First, the second order approach seems more natural [219] and physical because it avoids unnecessary variables (40 extra ones). In Deser's derivation, the connection is just Levi-Civita's on-shell, so its dynamics is not interesting. Second, as Deser's approach simply *verifies* that an assumed from is correct, it requires either a lucky guess or knowledge of the answer in advance, whereas the second-order recipe does not. Furthermore, the second-order approach is cleaner and more elegant, for no messy calculations are required. Finally, this second-order approach is more general in two respects. First, all generally covariant theories, including those with higher derivatives, manifestly fall within its scope, rather than remaining latent possibilities in the form of other lucky guesses. Second, the first-order approach either fails if the matter action depends on the connection [220], as it does for a perfect fluid [221], or requires the introduction of still more variables (perhaps another 40) to serve as Lagrange multipliers. In contrast, the second order approach always works using only 10 variables. For these reasons, we find a second-order principle preferable.

2.6 Full Universally-Coupled Action

We seek an action S obeying the physical requirement that the Euler-Lagrange equations be just the free field equations for S_f augmented by the total stress tensor:

$$\frac{\delta S}{\delta \gamma_{\mu\nu}} = \frac{\delta S_f}{\delta \gamma_{\mu\nu}} - \lambda \frac{\delta S}{\delta \eta_{\mu\nu}}, \tag{2.7}$$

where $\lambda = -\sqrt{32\pi G}$. In this respect our derivation follows Deser's more than Kraichnan's, for Kraichnan made no use of a free field action, but only of postulated free field equations.

The basic variables in this approach are the gravitational potential $\gamma_{\mu\nu}$ and the flat metric $\eta_{\mu\nu}$. But one is free to make a change of variables in S from $\gamma_{\mu\nu}$ and $\eta_{\mu\nu}$ to $g_{\mu\nu}$ and $\eta_{\mu\nu}$, where

$$g_{\mu\nu} = \eta_{\mu\nu} - \lambda \gamma_{\mu\nu}. \tag{2.8}$$

Equating coefficients of the variations gives

$$\frac{\delta S}{\delta \eta_{\mu\nu}}|\gamma = \frac{\delta S}{\delta \eta_{\mu\nu}}|g + \frac{\delta S}{\delta g_{\mu\nu}} \tag{2.9}$$

and

$$\frac{\delta S}{\delta \gamma_{\mu\nu}} = -\lambda \frac{\delta S}{\delta g_{\mu\nu}}. \tag{2.10}$$

Putting these two results together gives

$$\lambda \frac{\delta S}{\delta \eta_{\mu\nu}}|\gamma = \lambda \frac{\delta S}{\delta \eta_{\mu\nu}}|g - \frac{\delta S}{\delta \gamma_{\mu\nu}}. \tag{2.11}$$

Equation (2.11) splits the stress tensor into one piece that vanishes when gravity is on-shell and one piece that does not. Using this result in (2.7) gives

$$\lambda \frac{\delta S}{\delta \eta_{\mu\nu}}|g = \frac{\delta S_f}{\delta \gamma_{\mu\nu}}, \tag{2.12}$$

which says that the free field Euler-Lagrange derivative must equal (up to a constant factor) that part of the total stress tensor that does not vanish when the gravitational field equations hold. Recalling (2.3), one derives

$$\partial_\mu \frac{\delta S}{\delta \eta_{\mu\nu}} | g = 0, \tag{2.13}$$

which says that the part of the stress tensor not proportional to the gravitational field equations has identically vanishing divergence (on either index), *i.e.*, is a (symmetric) "curl" [1]. This result concerning the splitting of the stress tensor will be used in considering the gauge transformations of the full theory. It also ensures that the gravitational field equations *alone* entail conservation of energy and momentum, without any separate postulation of the matter equations. Previously the derivation of a conserved stress tensor required that gravity *and matter* obey their field equations, as in (2.5). This is possible only if the gravitational potential encodes considerable information about the matter fields through constraints. The Hamiltonian and momentum constraints imply this very fact [152], so one sees the origin of constraints from another angle.

It is worth recalling a conclusion of E. R. Huggins [42], who was a student of Feynman. Huggins found that the requirement that energy be a spin-two field coupled to the stress-energy tensor does not lead to a unique theory, because of terms of this curl form. Rather, "an additional restriction is necessary. For Feynman this restriction was that the equations of motion be obtained from an action principle; Einstein required that the gravitational field have a geometrical interpretation. Feynman showed these two restrictions to be equivalent." [42] (p. 3) Because we have built in the requirement of an action principle already, it is no surprise that we will find Riemannian geometrical theories to be the unique result.

We observe that the quantity $\frac{\delta S}{\delta \eta_{\mu\nu}} | g$, being symmetrical and having identically vanishing divergence on either index, is of the form

$$\frac{\delta S}{\delta \eta_{\mu\nu}}|g = \frac{1}{2}\partial_\rho \partial_\sigma (\mathcal{M}^{[\mu\rho][\sigma\nu]} + \mathcal{M}^{[\nu\rho][\sigma\mu]}) + b\sqrt{-\eta}\eta^{\mu\nu} \tag{2.14}$$

[152] (pp. 89, 429), where $\mathcal{M}^{\mu\rho\sigma\nu}$ is a tensor density of weight 1 and b is a constant. This result follows from the converse of Poincaré's lemma in Minkowski spacetime. (It is not strictly necessary to separate the b term out, but doing so is convenient, because getting this term from $\mathcal{M}^{\mu\rho\sigma\nu}$ would require that $\mathcal{M}^{\mu\rho\sigma\nu}$ depend on the position 4-vector.) We gather all dependence on $\eta_{\mu\nu}$ (with $g_{\mu\nu}$ independent) into one term, writing

$$S = S_1[g_{\mu\nu}, u] + S_2[g_{\mu\nu}, \eta_{\mu\nu}, u]. \tag{2.15}$$

One easily verifies that if [20]

$$S_2 = \frac{1}{2}\int d^4x R_{\mu\nu\rho\sigma}(\eta)\mathcal{M}^{\mu\nu\rho\sigma}(\eta_{\mu\nu}, g_{\mu\nu}, u) + \int d^4x \alpha^\mu,_\mu + 2b\int d^4x\sqrt{-\eta}, \tag{2.16}$$

then $\frac{\delta S_2}{\delta \eta_{\mu\nu}}|g$ has just the desired form, while S_2 does not affect the Euler-Lagrange equations. While Kraichnan's derivation has the advantage of not needing the physical answer beforehand, it does require clever mathematical use of the flat spacetime Riemann tensor to obtain superpotential-like terms. This quantity tends to be overlooked because it vanishes, but it is useful because its variation does not. The boundary and 4-volume terms are novel and useful, though not essential. The boundary term is necessary for showing that Rosen's action (with no second derivatives of the dynamical variables) [6] can be derived *via* universal coupling in flat spacetime, not merely postulated.[3] The 4-volume term can cancel the 0th order term in the action, so that the action vanishes when there is no gravitational field.

Thus,

$$S = S_1[g_{\mu\nu}, u] + \frac{1}{2}\int d^4x R_{\mu\nu\rho\sigma}(\eta)\mathcal{M}^{\mu\nu\rho\sigma} + 2b\int d^4x\sqrt{-\eta} + \int d^4x \partial_\mu \alpha^\mu. \tag{2.17}$$

[3] Although such a derivation was never presented by Rosen, to our knowledge, he did indicate that such a derivation would be desirable and intended to complete the project himself [6] (p. 153). As he notes, deriving the theory from flat spacetime seems more appealing than merely grafting the flat metric onto general relativity after the fact.

The boundary term is at our disposal. α^{μ} is a weight 1 vector density, because we require that S be a scalar. For S_1, we choose the Hilbert action for general relativity plus minimally coupled matter and a cosmological constant:

$$S_1 = \frac{1}{16\pi G}\int d^4x\sqrt{-g}R(g) - \frac{\Lambda}{8\pi G}\int d^4x\sqrt{-g} + S_{\mathrm{mat}}[g_{\mu\nu}, u]. \tag{2.18}$$

As is well-known, the Hilbert action is the simplest (scalar) action that can be constructed using only the metric tensor. If the gravitational field vanishes everywhere, then the gravitational action ought to vanish also, so we set $b = \Lambda/16\pi G$.

Rosen [6] noted that

$$R_{\mu\nu}(g) = R_{\mu\nu}(\eta) + E_{\mu\nu}(g, \partial), \tag{2.19}$$

where $E_{\mu\nu}(g, \partial)$ is identical in form to the Ricci tensor for $g_{\mu\nu}$, but with η-covariant derivatives ∂_{μ} replacing partial derivatives. Thus one finds that

$$E_{\mu\nu}(g, \partial) = \partial_{\sigma}\Delta^{\sigma}_{\rho\mu} - \partial_{\mu}\Delta^{\sigma}_{\rho\sigma} + \Delta^{\alpha}_{\rho\mu}\Delta^{\sigma}_{\alpha\sigma} - \Delta^{\alpha}_{\rho\sigma}\Delta^{\sigma}_{\alpha\mu}, \tag{2.20}$$

where the field strength tensor $\Delta^{\beta}_{\mu\alpha}$ is defined by

$$\Delta^{\beta}_{\mu\alpha} = \left\{{}^{\beta}_{\mu\alpha}\right\} - \Gamma^{\beta}_{\mu\alpha}. \tag{2.21}$$

Here $\left\{{}^{\beta}_{\mu\alpha}\right\}$ and $\Gamma^{\beta}_{\mu\alpha}$ are the Christoffel symbols for $g_{\mu\nu}$ and $\eta_{\mu\nu}$, respectively. Using (2.19) in the Hilbert term and using the product rule on the second derivatives in $E_{\mu\nu}(g, \partial)$ leaves first derivatives of the gravitational field and a boundary term. The boundary term is canceled if one chooses

$$16\pi G\alpha^{\mu} = -\Delta^{\mu}_{\rho\sigma}\mathfrak{g}^{\sigma\rho} + \Delta^{\sigma}_{\rho\sigma}\mathfrak{g}^{\mu\rho}, \tag{2.22}$$

where $\mathfrak{g}^{\mu\rho}$ is the contravariant metric density of weight 1. Using another of Rosen's results concerning the bimetric formalism [6], one readily expresses the g-covariant derivative of a tensor density in terms of the η-covariant derivative and terms involving $\Delta^{\mu}_{\rho\sigma}$ in place of the partial derivative and terms involving $\left\{{}^{\mu}_{\rho\sigma}\right\}$. A (1,1) tensor density of weight w is illustrative. For such a field, the η-covariant derivative [222] is

$$\partial_\mu \phi^\alpha_\beta = \phi^\alpha_\beta,_\mu + \phi^\sigma_\beta \Gamma^\alpha_{\sigma\mu} - \phi^\alpha_\sigma \Gamma^\sigma_{\beta\mu} - w\phi^\alpha_\beta \Gamma^\sigma_{\sigma\mu}, \tag{2.23}$$

and the g-covariant derivative $\nabla_\mu \phi^\alpha_\beta$ is analogous, with connection $\{^{\alpha}_{\sigma\mu}\}$. Recalling equation (2.21), one writes Rosen's result as

$$\nabla_\mu \phi^\alpha_\beta = \partial_\mu \phi^\alpha_\beta + \phi^\sigma_\beta \Delta^\alpha_{\sigma\mu} - \phi^\alpha_\sigma \Delta^\sigma_{\beta\mu} - w\phi^\alpha_\beta \Delta^\sigma_{\sigma\mu}. \tag{2.24}$$

The action to date takes the form

$$\begin{aligned} S = \frac{1}{16\pi G} \int d^4x \mathfrak{g}^{\mu\rho} R_{\mu\rho}(\eta) + \frac{1}{2} \int d^4x R_{\mu\nu\rho\sigma}(\eta) \mathcal{M}^{\mu\nu\rho\sigma}(\eta_{\mu\nu}, g_{\mu\nu}, u) \\ + \frac{1}{16\pi G} \int d^4x \mathfrak{g}^{\mu\rho} (\Delta^\sigma_{\mu\alpha} \Delta^\alpha_{\rho\sigma} - \Delta^\sigma_{\rho\mu} \Delta^\alpha_{\alpha\sigma}) + S_{\mathrm{mat}}[g_{\mu\nu}, u]. \end{aligned} \tag{2.25}$$

One can make $R_{\mu\nu\rho\sigma}(\eta)$ disappear from S by setting

$$\mathcal{M}^{\mu\nu\rho\sigma} = -\eta^{\nu\sigma} \mathfrak{g}^{\mu\rho}/8\pi G. \tag{2.26}$$

The contravariant weight 1 metric density $\mathfrak{g}^{\mu\rho}$ distinguishes itself here. This quantity has often appeared to be the preferred variable, not only in flat spacetime forms of general relativity (*e.g.*, [11, 17]), but also in other contexts. The DeDonder gauge condition, also known as the harmonic coordinate condition, prefers this variable [223, 224]; the desirability of this gauge was strongly urged by Fock. More recently, A. Anderson and J. York have found the "slicing density" [225], a weight -1 densitized version of the ADM lapse, to be quite useful. The slicing density is simply related to the 0-0 component of $\mathfrak{g}^{\mu\nu}$, as York has noted. One reason that we do not use $\mathfrak{g}^{\mu\nu}$ (or rather, $\frac{\mathfrak{g}^{\mu\nu} - \sqrt{-\eta}\eta^{\mu\nu}}{\lambda}$) as the gravitational potential is to make clear that no preference for this variable is built in by hand.

The total action is therefore Rosen's tensorial one with no second derivatives:

$$S = \frac{1}{16\pi G} \int d^4x \mathfrak{g}^{\mu\rho} (\Delta^\sigma_{\mu\alpha} \Delta^\alpha_{\rho\sigma} - \Delta^\sigma_{\rho\mu} \Delta^\alpha_{\alpha\sigma}) + S_{\mathrm{mat}}[g_{\mu\nu}, u]. \tag{2.27}$$

This action should be compared to those available in geometrical general relativity, where one chooses either to include second derivatives of the dynamical variables, or to give up the scalar character of the action, at least if the metric is the dynamical variable .

Babak and Grishchuk [93] have proposed a different principle for specifying $\mathcal{M}^{\mu\nu\rho\sigma}$, with different results. Their proposal gives an attractive form to the metric stress tensor, *viz.*, a tensorial relative of the Landau-Lifshitz pseudotensor [226], which is the only symmetric pseudotensor with no second derivatives. This tensor had been previously obtained in a conservation law for bimetric general relativity by Rosen [6], but that derivation did not involve Noether's theorem [227]. (Another desirable stress tensor has been discussed by Petrov and Katz recently [94, 95].)

There are two key ingredients in the derivation of generally covariant theories in this way, apart from the use of an action principle. One is universal coupling, which says that the source for the field equations must be the total stress-energy tensor. The other key ingredient can be either free field gauge invariance of the assumed form or gravitation-induced conservation of energy and momentum. Gauge invariance might be motivated, if in no other way, by a desire for Lorentz invariance and positive energy. However, as unimodular general relativity and the slightly bimetric theories with dynamical $\frac{\sqrt{-g}}{\sqrt{-\eta}}$ below show, this specific form of gauge invariance is more restrictive than necessary for positive energy and Lorentz invariance. This fact follows from the fact that slightly bimetric theories behave like scalar-tensor theories (as will be shown below), and at least some of the latter have positive energy [228]. This condition is therefore weaker than that required by Fierz [4] and van Nieuwenhuizen [75], who were interested in good behavior of *free* fields. This fact has also been pointed out by G. Cavalleri and G. Spinelli [80]. However, below it will appear that the full gauge invariance might be needed if the special relativistic causal structure is to be upheld.

2.7 Gauge Invariance and Gauge Fixing

It is instructive to determine what has become of the original free field gauge invariance. The scalar character of the action entails

$$\delta S_{coord} = \int d^4x [\frac{\delta S}{\delta g_{\mu\nu}} \pounds_\xi g_{\mu\nu} + \frac{\delta S}{\delta u} \pounds_\xi u + (\frac{\delta S}{\delta \eta_{\mu\nu}} | g) \pounds_\xi \eta_{\mu\nu} + h^\mu{}_{,\mu}] = 0 \qquad (2.28)$$

under a coordinate transformation, where the form of $h^\mu{}_{,\mu}$ is not important. (The same will hold for the other boundary terms below.) But in a flat spacetime theory, invariance under coordinate transformations is trivial. A *gauge* transformation, on the other hand, would be a transformation that changes the action only by a boundary term, but is not a coordinate transformation. Using the coordinate transformation formula and noting that the terms involving the absolute objects do not contribute more than a divergence, one easily verifies that a (pure) gauge transformation is given by $\delta g^{\mu\nu} = \pounds_\xi g^{\mu\nu}$, $\delta u = \pounds_\xi u$, $\delta \eta^{\mu\nu} = 0$, with ξ^μ arbitrary. (See also ([89]), but we do not impose any gauge condition *a priori* as Logunov *et al.* do. If one does impose a gauge condition, it seems best to implement it in the action principle.) Thus,

$$\begin{aligned} \delta S_{gauge} &= \delta S_{coord} - \int d^4x [(\frac{\delta S}{\delta \eta_{\mu\nu}} | g) \pounds_\xi \eta_{\mu\nu} + i^\mu{}_{,\mu}] \\ &= 0 - \int d^4x (-2\xi^\alpha \eta_{\alpha\mu} \partial_\nu \frac{\delta S}{\delta \eta_{\mu\nu}} | g + j^\mu{}_{,\mu}). \end{aligned} \qquad (2.29)$$

Recalling from (2.13) above that

$$\partial_\mu \frac{\delta S}{\delta \eta_{\mu\nu}} | g = 0 \qquad (2.30)$$

identically, one sees that δS_{gauge} is indeed merely a boundary term, so our guessed form of the gauge invariance is verified. In this case, gauge transformations change (bosonic) dynamical fields in the same way that coordinate transformations do, but leave the nondynamical object $\eta_{\mu\nu}$ unchanged. If one performs simultaneously a gauge transformation and a coordinate transformation in the 'opposite direction,' then the dynamical variables are unchanged, but the absolute object $\eta_{\mu\nu}$ is altered.

Given that coordinate-independent localization of gravitational energy and momentum is one of the attractive features of the bimetric approach to general relativity, it must be emphasized that the tensorial character has not removed the original arbitrariness, but merely transformed it into the gauge-variance of the gravitational stress-energy tensor [83, 85]. However, several authors, including Rosen and Papapetrou, have suggested that the bimetric approach ought to be gauge-fixed [6, 11, 108]. Still, such gauge fixing might not lead to complete localization [83, 85, 130, 131]. Full discussion of the gauge freedom, however, will require consideration of the meaning of the formalism and the relation between the null cone structures of the two metrics, an important but largely neglected issue that will be taken up below.

Chapter 3

A Preferred Foliation in Flat Spacetime Gravitation?

3.1 Introduction

As we have seen, general relativity can be formally derived as a flat spacetime theory. Here we introduce on an n-dimensional space a preferred temporal foliation $\partial_\mu \theta$ along side the n-dimensional flat metric $\eta_{\mu\nu}(x)$ (in arbitrary coordinates) as nondynamical objects. While the dimension n is not specified, we expect that $n = 4$ with metric signature $-+++$ and $n = 5$ with signature $-+++-$ [132–149] will be of the greatest interest. (If $n = 4$, then Greek indices run from 0 to 3 and $\theta = x^0 = t$ for some preferred inertial frame in the natural coordinates. If $n = 5$, then Greek indices run over 0,1,2,3,5 with $\theta = x^5 = \tau$ in natural coordinates, τ being an invariant supertime.) The foliation $\partial_\mu \theta$ obeys $(\partial_\mu \theta)\eta^{\mu\nu}\partial_\nu \theta = -1$ and $\partial_\nu \partial_\mu \theta = 0$. We define the unit normal covector by $n_\mu = -\partial_\mu \theta$. Its η-raised counterpart $n^\nu = n_\mu \eta^{\mu\nu}$ is future-pointing [151]. The metric and foliation induce a projection tensor $h_{\mu\nu} = \eta_{\mu\nu} + n_\mu n_\nu$. If we agree to refer to θ as "time" (perhaps meaning τ) and the remaining dimensions as "space" (perhaps having signature $-+++$) for convenience, then the projection tensor serves as the "spatial" metric. The symmetric gravitational potential is $\gamma^{\mu\nu}$ (of density weight 0), and bosonic matter fields are denoted by u (with all indices suppressed). One could take γ to be either a contravariant density of any weight (except $\frac{1}{2}$) or a covariant density of any weight (except $-\frac{1}{2}$). (These weights are special because the resulting inverse metric tensor density or metric tensor density, respectively, has a determinant of -1, so there is one fewer independent component. Thus, invertibility issues arise.) In the massless case, this choice makes no difference. If it were desired to add a gauge-symmetry-breaking mass term, the choice of index position and weight would affect the results.

3.2 Spatial Gauge Invariance and Universal Momentum Coupling

This derivation is an improved version of previous work of ours on this subject [145], based on the derivation of general relativity given above. Unlike our previous parametrized derivation [145], this one does not assume that $n_\mu n_\nu \gamma^{\mu\nu} = 0$, so one can retain the lapse function N of an ADM split [151] as a nontrivial quantity, as opposed to requiring $N = 1$ *a priori* and thus not varying it. The assumption of gauge invariance requires that the field be massless, except for the time-time part $n_\mu n_\nu \gamma^{\mu\nu}$, which can be massive. One can show this fact directly by writing the most general algebraic quantity that is quadratic in the gravitational potentials. (As in the case without a preferred foliation [260], there is no mathematical objection to adding a mass term that breaks the gauge invariance. A possible physical difficulty with negative energy will be noted below. Including both mass terms would imply that the time-time part of the field had a different rest mass from the rest of the field.)

Let S_f be the action for a free gravitational field (in the sense that the gravitational coupling constant vanishes). We require that S_f change only by a boundary term under the infinitesimal gauge transformation

$$\gamma^{\mu\nu} \to \gamma^{\mu\nu} + \partial^\mu \xi^\nu + \partial^\nu \xi^\mu, \tag{3.1}$$

ξ^ν being a vector field obeying $\xi^\nu n_\mu = 0$.

For any S_f invariant in this sense under (3.1), the free field equations' divergence is purely temporal, for its spatial projection vanishes, as we now show. The action changes by

$$\delta S_f = \int d^n x [\frac{\delta S_f}{\delta \gamma^{\mu\nu}} (\partial^\nu \xi^\mu + \partial^\mu \xi^\nu) + e^\mu{}_{,\mu}] = \int d^n x f^\mu{}_{,\mu} . \tag{3.2}$$

The forms of the boundary terms $e^\mu{}_{,\mu}$ and $f^\mu{}_{,\mu}$ are not needed. Integrating by parts, letting ξ^μ have compact support to annihilate the boundary terms (as we shall do throughout the paper), and recalling the purely spatial character of ξ^μ, we obtain the identity

$$h^\nu_\alpha \partial^\mu \frac{\delta S_f}{\delta \gamma^{\mu\nu}} = 0. \tag{3.3}$$

An expression for the total energy-momentum tensor can be derived from S using the metric recipe [1, 20, 93] in the following way, making allowances for the presence of another absolute object. The action depends on the flat metric $\eta_{\mu\nu}$, the foliation n_μ, the gravitational potential $\gamma^{\mu\nu}$, and bosonic matter fields u, the last representing any number of tensor densities of arbitrary weights and index positions. Under an arbitrary infinitesimal coordinate transformation described by a vector field ψ^μ, the action changes by the amount

$$\delta S = \int d^n x (\frac{\delta S}{\delta \gamma^{\mu\nu}} \pounds_\psi \gamma^{\mu\nu} + \frac{\delta S}{\delta u} \pounds_\psi u + \frac{\delta S}{\delta \eta^{\mu\nu}} \pounds_\psi \eta^{\mu\nu} + \frac{\delta S}{\delta n_\mu} \pounds_\psi n_\mu + g^\mu{}_{,\mu}), \tag{3.4}$$

where $\pounds_\psi$ is the Lie derivative with respect to ψ^μ and $g^\mu{}_{,\mu}$ is another boundary term of no present interest. But S is a scalar, so $\delta S = 0$. Letting the matter and gravitational field equations hold gives

$$\delta S = \int d^n x (\frac{\delta S}{\delta \eta^{\mu\nu}} \pounds_\psi \eta^{\mu\nu} + \frac{\delta S}{\delta n_\mu} \pounds_\psi n_\mu) = 0, \tag{3.5}$$

or, upon rewriting the Lie derivatives in terms of the covariant derivative compatible with the flat metric [152],

$$\partial^\mu \frac{\delta S}{\delta \eta^{\mu\nu}} - \frac{1}{2} n_\nu \partial_\mu \frac{\delta S}{\delta n_\mu} = 0. \tag{3.6}$$

One can identify $\frac{\delta S}{\delta \eta^{\mu\nu}} - \frac{1}{2} n_\nu \eta_{\mu\alpha} \frac{\delta S}{\delta n_\alpha}$ as the energy-momentum tensor. Its asymmetry reflects the preferred character of n_μ. If we take the spatial projection of this quantity, we obtain the tensor of momentum density and its flux $h^\nu_\alpha \frac{\delta S}{\delta \eta^{\mu\nu}}$.

For the full theory, we postulate that the spatial projection of the Euler-Lagrange equations should be just the spatial projection of the free field equations for S_f augmented by the momentum tensor (including gravitational momentum):

$$h^\nu_\alpha \frac{\delta S}{\delta \gamma^{\mu\nu}} = h^\nu_\alpha \frac{\delta S_f}{\delta \gamma^{\mu\nu}} - \lambda h^\nu_\alpha \frac{\delta S}{\delta \eta^{\mu\nu}}, \tag{3.7}$$

where λ is a coupling constant. If $n = 4$ and the theory in question is general relativity, then $\lambda = -\sqrt{32\pi G}$; for other theories, the Newtonian limit would need consideration. If $n = 5$, then λ has different dimensions from Newton's constant G, with an additional length entering [144]. Horwitz *et al.* have previously noted the entrance of an additional length in their parametrized electromagnetism [141].

One is free to make a change of variables in S from the flat metric $\eta^{\mu\nu}$ and gravitational potential $\gamma^{\mu\nu}$ to $\eta^{\mu\nu}$ and $g^{\mu\nu}$, where

$$g^{\mu\nu} = \eta^{\mu\nu} + \lambda \gamma^{\mu\nu}. \tag{3.8}$$

Equating coefficients of the variations gives

$$\frac{\delta S}{\delta \eta^{\mu\nu}}|\gamma = \frac{\delta S}{\delta \eta^{\mu\nu}}|g + \frac{\delta S}{\delta g^{\mu\nu}} \tag{3.9}$$

and

$$\frac{\delta S}{\delta \gamma^{\mu\nu}} = \lambda \frac{\delta S}{\delta g^{\mu\nu}}. \tag{3.10}$$

Putting these two results together gives

$$\lambda \frac{\delta S}{\delta \eta^{\mu\nu}}|\gamma = \lambda \frac{\delta S}{\delta \eta^{\mu\nu}}|g + \frac{\delta S}{\delta \gamma^{\mu\nu}}. \tag{3.11}$$

Equation (3.11) splits the energy-momentum tensor into one piece that vanishes when the gravitational Euler-Lagrange equations hold and one piece that does not. Using this result in (3.7) gives

$$\lambda h^{\nu}_{\alpha} \frac{\delta S}{\delta \eta^{\mu\nu}}|g = -h^{\nu}_{\alpha} \frac{\delta S_f}{\delta \gamma^{\mu\nu}}, \tag{3.12}$$

which says that the spatial projection of the free field Euler-Lagrange derivative must equal (up to a constant factor) that part of the momentum tensor that does not vanish when the gravitational field equations hold. Recalling (3.3), one derives

$$h^{\nu}_{\alpha} \partial^{\mu} \frac{\delta S}{\delta \eta^{\mu\nu}}|g = 0, \tag{3.13}$$

which says that the part of the momentum tensor not proportional to the gravitational field equations has identically vanishing divergence. This result concerning the splitting of the momentum tensor will be used in considering the gauge transformations of the full theory. It also ensures that the gravitational field equations *alone* entail conservation of momentum, without any separate postulation of the matter equations, though the conservation of energy still depends upon the matter field equations.

Expanding the projection tensor gives

$$\partial^{\mu} \frac{\delta S}{\delta \eta^{\mu\alpha}}|g + n^{\nu} n_{\alpha} \partial^{\mu} \frac{\delta S}{\delta \eta^{\mu\nu}}|g = 0. \tag{3.14}$$

Let us divide the action into 3 pieces: a piece S_1 independent of $\eta^{\mu\nu}$ (the other variable being $g^{\mu\nu}$, not $\gamma^{\mu\nu}$), a piece S_2 that contributes a symmetric curl (*i.e.*, a quantity with identically vanishing divergence on either index) to the stress tensor, and another piece S_3. We therefore write

$$S = S_1[g, u, n] + S_2[g, u, \eta, n] + S_3[g, u, \eta, n], \tag{3.15}$$

suppressing all indices in the arguments. Given the assumed property of S_2, one can write

$$-\eta^{\mu\alpha}\eta^{\nu\beta}\frac{\delta S_2}{\delta\eta^{\alpha\beta}}|g = \frac{1}{2}\partial_\rho\partial_\sigma(\mathcal{M}^{[\mu\rho][\sigma\nu]} + \mathcal{M}^{[\nu\rho][\sigma\mu]}) + b\sqrt{-\eta}\eta^{\mu\nu} \tag{3.16}$$

[152] (pp. 89, 429), where $\mathcal{M}^{\mu\rho\sigma\nu}$ is a tensor density of weight 1 and b is a constant. This result follows from the converse of Poincaré's lemma in flat spacetime. One easily verifies that if [20]

$$S_2 = \frac{1}{2}\int d^n x R_{\mu\nu\rho\sigma}(\eta)\mathcal{M}^{\mu\nu\rho\sigma}(g,u,\eta,n) + \int d^n x \mathcal{X}^\mu{}_{,\mu} + 2b\int d^n x\sqrt{-\eta}, \tag{3.17}$$

then $\frac{\delta S_2}{\delta\eta^{\mu\nu}}|g$ has just the desired form, while S_2 does not affect the Euler-Lagrange equations because its value is 0. $\mathcal{X}^\mu$ is a weight 1 vector density, because we require that S be a scalar. For convenience we deposit all boundary terms into S_2.

One can also show that if

$$S_3 = -2\int d^n x \partial_\beta\psi^{\alpha\beta}(g,u,\eta,n)n_\alpha\partial_\mu n^\mu, \tag{3.18}$$

then (3.14) is satisfied. Now both terms of that equation are used, not just one as for S_2. Here $\psi^{\alpha\beta}(g,u,\eta,n)$ is any (2,0) weight 1 tensor density constructed from all the fields, dynamical and nondynamical. It turns out that

$$\frac{\delta S_3}{\delta\eta^{\mu\nu}} = -n^\rho n_\alpha\eta_{\mu\nu}\partial_\beta\partial_\rho\psi^{\alpha\beta} + 2n_\alpha n_{(\nu}\partial_{\mu)}\partial_\beta\psi^{\alpha\beta}. \tag{3.19}$$

Combining these 3 terms gives the total action

$$\begin{aligned} S = S_1[g^{\mu\nu},u,n_\mu] + \frac{1}{2}\int d^n x R_{\mu\nu\rho\sigma}(\eta)\mathcal{M}^{\mu\nu\rho\sigma}(g,u,\eta,n) + 2b\int d^n x\sqrt{|\eta|} + \\ -2\int d^n x\partial_\beta\psi^{\alpha\beta}(g,u,\eta,n)n_\alpha\partial_\mu n^\mu + \int d^n x\mathcal{X}^\mu(g,u,\eta,n)_{,\mu}\,. \end{aligned} \tag{3.20}$$

One sees that S_1 can contain mass and self-interaction terms for the time-time component of the gravitational potential, using $\sqrt{-g}$ to the first power and any suitable function of $g^{\mu\nu}n_\mu n_\nu + 1$. With this action, one could write down a number of theories, including (of course) general relativity. One would still need to verify that any given theory made sense theoretically (such as by having adequate positive energy properties) and empirically, for the universal coupling principle does not address such questions fully.

It is instructive to determine what has become of the original free field gauge invariance. The scalar character of the action $S[g, u, \eta, n]$ implies that the variation of the action under a coordinate transformations vanishes:

$$\delta S_{ct} = \int d^n x (\frac{\delta S}{\delta g^{\mu\nu}} \pounds_{\psi} g^{\mu\nu} + \frac{\delta S}{\delta u} \pounds_{\psi} u + \frac{\delta S}{\delta \eta^{\mu\nu}} | g \pounds_{\psi} \eta^{\mu\nu} + h^{\mu}{}_{,\mu}) = 0. \tag{3.21}$$

But as we saw above in a flat spacetime theory, invariance under coordinate transformations is trivial, so a gauge transformation must differ somehow. Using the coordinate transformation formula and noting that the terms involving the absolute objects do not contribute more than a divergence, one easily verifies that a (pure) gauge transformation is given by $\delta g^{\mu\nu} = \pounds_{\xi} g^{\mu\nu}$, $\delta u = \pounds_{\xi} u$, $\delta \eta^{\mu\nu} = 0$, $\delta n_{\mu} = 0$, where $\xi^{\mu} n_{\mu} = 0$, but ξ^{μ} is otherwise arbitrary. In showing that the term for the flat metric does not contribute nontrivially, one must recall from (3.13) above that

$$h^{\nu}_{\alpha} \partial^{\mu} \frac{\delta S}{\delta \eta^{\mu\nu}} | g = 0 \tag{3.22}$$

identically. The term for the foliation does not contribute at all because $\pounds_{\xi} n_{\mu} = 0$ on account of the constancy of n_{μ} and the assumption that $\xi^{\mu} n_{\mu} = 0$. Thus, gauge transformations change (bosonic) dynamical fields in the same way that 'spatial' coordinate transformations do, but leave the nondynamical objects unchanged.

Taking the independent variables to be those in $S[g^{\mu\nu}, u, \eta^{\mu\nu}, n_{\mu}]$, one can easily derive the Bianchi identities. Letting the Euler-Lagrange equations hold, one finds an additional equation that holds as a consequence. This phenomenon is not unprecedented: in unimodular general relativity (which also has a nondynamical object present, namely, a volume element), this additional equation restores the trace of the Einstein equations (up to an arbitrary cosmological constant) that failed to appear in the Euler-Lagrange equations [239]. So a constraint not admitted through the front door might still reenter through the back door, if this extra equation is nontrivial.

From a Hamiltonian point of view, one says that dynamical preservation of the momentum constraints (roughly, the time-space Einstein equations) implies the Hamiltonian constraint (roughly, the time-time component of the Einstein equations), up to an arbitrary constant, in the unimodular theory. On the other hand, we find that in the theory obtained from the Hamiltonian for general relativity but with $N = 1$ *a priori* (so the Hamiltonian constraint does not follow from the variation of the Hamiltonian), no further constraint, including any portion of the Hamiltonian constraint, is required to preserve the momentum constraints. At least that is the case in the vacuum theory; we expect that minimally coupled matter would behave similarly. (Also see ([153]).) But it would appear that this theory would have trouble with the negative energy degree of freedom constituted by the determinant of the curved spatial metric, so it does not seem physically viable.

We return now to the Lagrangian formulation for theories with a flat background metric and a preferred foliation. Making a coordinate transformation, letting the matter and gravitational field equations hold, discarding the boundary terms, and using the arbitrariness of the coordinate transformation yields the relation

$$\partial^\mu \frac{\delta S}{\delta \eta^{\mu\nu}}|g - \frac{1}{2} n_\nu \partial_\mu \frac{\delta S}{\delta n_\mu} = 0. \tag{3.23}$$

On account of (3.13), the spatial projection of this equation already holds identically. The remainder, the temporal component, takes the form

$$2n^\nu \partial^\mu \frac{\delta S}{\delta \eta^{\mu\nu}}|g + \partial_\mu \frac{\delta S}{\delta n_\mu} = 0. \tag{3.24}$$

When the action is expanded as using $S = S_1 + S_2 + S_3$, as in (3.15), several terms vanish, namely, $\frac{\delta S_1}{\delta \eta^{\mu\nu}}|g$, $\partial^\mu \frac{\delta S_2}{\delta \eta^{\mu\nu}}|g$, and $\frac{\delta S_2}{\delta n_\mu}$. After showing that $\frac{\delta S_3}{\delta n_\mu} = 2\partial^\mu n_\alpha \partial_\beta \psi^{\alpha\beta}$, one finds that the contributions from S_3 cancel each other. The final result takes the rather simple form

$$\frac{\partial}{\partial x^{\mu}} \frac{\delta S_1}{\delta n_{\mu}} = 0. \tag{3.25}$$

The physical meaning of this equation will depend on the precise form of S_1.

3.3 Interpretation of the $-+++-$ Formalism

In 5-dimensional form this work suggests itself as a route to a theory of "parametrized" gravitation, the extra dimension being the invariant supertime τ. If the extra time dimension is to be interesting, it is necessary that physical fields be permitted to depend on τ [141]. But then one faces the question of relating a 5-dimensional description to the observed 4 dimensions. One common approach in an electromagnetic context has been "concatenation," in which the τ-dependent vector potential is integrated over all τ (from eternity past to eternity future) to give standard Maxwell potentials, the latter supposedly being tied to experiments [133, 134, 136, 138, 139, 141, 142]. We have previously argued that concatenation is unsatisfactory [144], *pace* ([141]). One reason is that it makes essential use of the linearity of the field equations, but this linearity does not hold generically; in particular, it is violated by any reasonable theory of gravity [144]. It is doubtful that any variant of concatenation of a nonlinear parametrized theory would give a plausible nonlinear nonparametrized theory. Even Yang-Mills theories, though formally similar to electromagnetism, cannot be concatenated. Another reason is that observations are said to be influenced by all values of τ, including future ones, yet τ is thought to be related to the process/flowing aspect of time [141]. The latter fact implies that experiments performed by real people in ordinary life ought to occur at some definite moment (or finite interval) of τ. So concatenation introduces the paradox of backwards causation in τ [144]. It also introduces a curious distinction between measurement and evolution, like certain versions of quantum mechanics, rather than regarding measurement as a specific kind of evolution.

Motivated by these criticisms, we previously suggested that omitting concatenation and interpreting all experiences as involving a convective derivative with respect to τ along a worldline might yield an adequate interpretation of the parametrized formalism [144]. However, the non-concatenated view has a drawback of its own, for such a theory generically agrees with standard well-confirmed 4-dimensional electromagnetism (*e.g.*, light speed measurements) only if the dependence on τ is quite weak [1] This limit is similar to the zero mode limit considered by Frastai and Horwitz [143], who have been aware of some of the difficulties with concatenation.[2] They observed that the zero mode limit is a sufficient condition for agreement of the parametrized theory with experiment. But our suggestion is that approaching the zero mode limit is a *necessary* condition. For example, the lack of observed dispersion in light propagation indicates that the τ derivatives are much smaller than the t derivatives in such contexts. If τ were to be associated with temporal becoming only, with no physical content (*cf.* ([147]), then perhaps a more attractive and economical solution exists in the $n = 4$ context, as we will explain below.

A recent development in membrane theory might possibly be of interest here. Recently the somewhat analogous question "Can there be 'large' extra dimensions?" has received a surprising positive answer. The Randall-Sundrum scenario [150] appears to permit higher-dimensional theories to have large extra dimensions while giving empirically reasonable results. Perhaps one can imagine a parametrized analog of this move.

In conclusion, it is clear that the parametrized formalism still faces fundamental interpretive questions.

[1] We thank Profs. R. Matzner and M. Choptuik for making this point and Prof. D. Salisbury for related thoughts.

[2] We thank Prof. Horwitz for a discussion of this matter.

3.4 Interpretation of the $-+++$ Formalism

In a 4-dimensional context this formalism corresponds to the existence of a preferred reference frame. It is generally assumed that no such thing exists, though the subject has received some attention [154–156], especially from W.-T. Ni [244] and H. B. Nielsen and collaborators [157–159]. It is also striking that Dirac, on seeing how the Hamiltonian formalism for general relativity could be simplified (in the simplification of the primary constraints to the vanishing of four momenta, which permits one to eliminate the conjugate parts of the metric from the formalism) by the addition of a non-Lorentz-invariant divergence to the Lagrangian density, could write, "*This is a substantial simplification, but it can be obtained only at the expense of giving up four-dimensional symmetry.* I am inclined to believe from this that four-dimensional symmetry is not a fundamental property of the physical world." [160] (emphasis in the original) This remark is especially striking given that the resulting Lagrangian density was still within a divergence of Lorentz invariance, so the violation does not even appear in the Euler-Lagrange equations.

While it might complicate canonical formalisms, the presumed nonexistence of a preferred frame is in many important respects quite helpful, because of the resulting tight restrictions on the number of theories that can be conceived. With a very few possible exceptions, all known physical processes are consistent with the orthodox relativistic view that there is no preferred foliation (and that backwards-in-time causation does not occur). In view of the apparently limited gains and substantial losses realized by giving up Lorentz invariance, one might wonder what is the purpose of considering a preferred foliation in physics. It would seem that a rather good argument is needed to justify such work. We now consider whether such an argument is available. We warn the reader that the remainder of this chapter will be rather philosophical.

One apparent difficulty for standard Lorentz-invariant physics is the remarkable behavior seen in certain quantum mechanical experiments, such as by Aspect *et al.*, in which 2 particles in an entangled superposed state seem to be able to 'communicate' superluminally. Much of the physics community seems to believe that locality is doomed, and has given up on it, at least when its mind is on quantum mechanics. However, this would be a tremendous loss, and so it ought not to be accepted unnecessarily [161].

What can be said in defense of locality? We do not claim to give a comprehensive review, but only provide two suggestions. Evidently the detector efficiency loophole is still open [162, 163]. Szabó and Fine's model even works for experiments testing the GHZ scenario [162], a more recent and perhaps more potent threat to local hidden variables than Bell's theorem. Detector efficiencies are still too low to close this loophole [163]. Other sources of deviation from the ideal experiment might also be considered [164]. One also knows that the experiments violating the Bell inequalities are compatible with the orthodox relativity if one is prepared to embrace "superdeterminism" [165–168], which violates the inequalities by introducing correlations between the hidden variables and the detector settings. By positing a common cause for these correlations, one can preserve orthodox relativity, the Aspect experiments notwithstanding. Because the GHZ theorem involves similar locality assumptions to those involved in Bell's theorem [169], we suspect that it can be subverted analogously. However, this view's demanding philosophical underpinnings, such as its denial of (libertarian) free will[3] and evident need for an all-determining Agent to correlate the initial conditions of the world, might limit its appeal (see, *e.g.*, Bell's attitude [170] (pp. 100-103, 110, 154)).[4] The detector

[3]Free will faces a potent long-standing conceptual objection that an action that isn't fully caused is to that extent merely random and thus un-free [171–173], so the denial of free will might be inevitable on other grounds. If so, then the entrance fee for superdeterminism will decrease.

[4]On the other hand, the 3 major near-Eastern monotheistic traditions all have (or had) strands that affirm theological determinism: Pharisaic Judaism [174], Reformed/Calvinist Christianity, and Islam. That there might be a natural affinity here is suggested by the language (*e.g.*, ([166]) about events being "already 'written in a book'." The resemblance to Psalm 139:16 (NASB) cannot be accidental:

efficiency loophole has also seemed unappealing to some, such as Bell [170] (p. 109). However, it is at least worthwhile to show that these strategies exist, because they show that even in this peculiar aspect of quantum mechanics, nothing is presently known with certainty that requires a preferred frame.

Another trouble spot for the usual relativistic view of time is quantum gravity's "problem of time" [175], which consists in the *prima facie* disappearance of time from quantum versions of general relativity. However, it seems that the problem lies not in the lack of a particular preferred frame, a feature shared with special relativity–the success of standard field theory in other contexts suggests that this feature is not at fault–but in the lack of a preferred class of inertial frames peculiar to the form of gauge invariance of general relativity. Because plausibly the 'fault' lies in how general relativity differs from special relativity (general covariance, *i.e.*, lack of nondynamical objects), so one needn't add structures unknown to special relativity to address the issue. One expects that the problem of time would disappear if one suitably introduced a nondynamical background metric into the equations of motion. Adding a small rest mass to the theory would be an obvious way to implement this procedure (and thereby obtain a nonvanishing Hamiltonian), if the traditional negative-energy objection to massive gravity [234] (appendix on "ghost" theories) can be overcome. M. Visser has recently suggested that this worry has been somewhat overstated [235]. One might also prefer that the curved metric respect the flat background's null cone structure, a nontrivial condition that, to our knowledge, has not been successfully imposed in an attractive way.

Recently another suggested habitat for the violation of Lorentz invariance has appeared. T. Jacobson and D. Mattingly have suggested that there is "reason

Thine eyes have seen my unformed substance;
And in Thy book they were all written,
The days that were ordained for me,
When as yet there was not one of them.

to doubt exact Lorentz invariance: it leads to divergences in quantum field theory associated with states of arbitrarily high energy and momentum. This problem can be cured with a short distance cutoff which, however, breaks Lorentz invariance" [176]. They then introduce an "aether" consisting of a dynamical unit timelike vector (or covector) field. Their aether, being dynamical and failing in general to define a preferred foliation (because the covector typically is not a gradient), differs essentially from what we consider. Their divergence argument might give a good reason to consider a preferred foliation, such as we have considered here. But it seems premature to put too much reliance on this proposal.

If these areas of physics do not provide sufficiently strong evidence for the existence of an observable preferred foliation in physics, then one might ask if there are extra-physical reasons for considering a preferred temporal foliation in physics. It so happens that in the 20th century's central debate in the philosophy of time [177], one of the two views, if established, would show that a preferred foliation exists at the most fundamental level. (Below we will find authors arguing that if a preferred foliation exists, then presumably it manifests itself in physics.) This is the debate about the objectivity or otherwise of temporal becoming, that is, the 'flow' of time [178, 179]. Some physicists and philosophers, based to a large degree on the influence of relativity [179], incline toward the "block universe" view that regards all moments of time as ultimately equal in status; the notions of past, present, and future are regarded as illusory or mind-dependent. If the block view (also known as "stasis," "B-theory," or "tenseless") is correct, then all facts can in principle be displayed on a (single) spacetime diagram.[5] On the other hand, if a spacetime diagram (perhaps augmented by timeless mathematical and logical truths) cannot display all facts, because some facts have temporal properties inconsistent with

[5]If one is reluctant to make statements such as "$2 + 2 = 4$ in 1980," one can admit a class of timeless truths also, but that seems unnecessary [180]. Facts such as "I am John Perry" [181], if they are nontrivial, might not fit on a spacetime diagram, but since their temporal properties are not the problem, we can set them aside.

such a representation, then the "process" ("A-theory," "tensed") view that affirms objective becoming will be established.

The process view of time receives some unexpected assistance from stasis advocate D. H. Mellor, who wrote [182] (pp. 4-5):

> The tenseless camp often offers only weak inducements to join it: the relative simplicity of tenseless logic, for example, or its consonance with relativity's unification of space and time. But tenseless time needs a stronger sales pitch than that. Tense is so striking an aspect of reality that only the most compelling argument justifies denying it: namely, that the tensed view of time is self-contradictory and so cannot be true."

Mellor claims to find this needed contradiction in McTaggart's paradox, but this claim is not generally accepted and indeed appears to be false [183]. If McTaggart's paradox fails to demonstrate a contradiction, but Mellor's judgement is otherwise to be trusted, then the process view of time wins already. But some will require a more compelling case, which we believe can be made.

There is an argument [184–186] that appears to disprove the block view by showing that it cannot accommodate certain facts [189, 190]. In order to make one's appointments on time, one frequently needs to know what time it is *now*. For example, if one wants to pay taxes to the American government in a timely way, one might want to know that "It is now April 2." Otherwise, one might file many weeks or even years late, because one just would not know when to file [181]. This sort of fact, along with more general facts about what is occurring *now*, cannot be represented tenselessly [191, 192], such as on a spacetime diagram, or known by a timeless being (even a divine one, as Kretzmann had in mind) [193], *i.e.*, one lacking temporal location and duration. Let us see why this is the case. On such a diagram, one might make a mark at "April 2" on the time axis, but this mark will soon be outdated, so it will no longer represent "now." Trying to keep the "now" mark current would require continually erasing and drawing on the spacetime diagram, which is of course illegal, for one then has a *succession* of diagrams (a movie), not a single one. Neither will the fact "It is April 2 on April 2" be of any use, both because it is a tautology [177] and thus cannot inspire any action at all, and because it is always true [194] and thus cannot motivate action at any special moment. If one finds the use of a date label such as April 2 troubling (as if a substantivalist view of time might be to blame), one could substitute some ordinary occurrence: "A rooster is crowing now" would serve, provided that the rooster only crows shortly before tax day. So "now" points to one or more facts that the block universe cannot accommodate. We must therefore reject the following claim by H. Reichenbach [195] (pp. 16,17):

> There is no other way to solve the problem of time than the way through physics ... If time is objective the physicist must have discovered that fact, if [sic] there is Becoming the physicist must know it ... It is a hopeless enterprise to search for the nature of time without studying physics. If there is a solution to the philosophical problem of time, it is written down in the equations of mathematical physics.

We must further differ from Reichenbach, who asserted that determinism would exclude becoming [195], for irreducibly tensed facts provide a ground for an objective flow of time, even if determinism is true. (Interestingly, Reichenbach concluded that physics in fact does ground time flow.) Thus, we conclude that a preferred foliation generated by the "moving now" exists.

It seems natural to assume that if a preferred foliation exists, it ought to manifest itself in physics fairly readily.[6] Such an intuition indicates that Reichen-

[6] In the positivist era of the 20th century, the question in the title of this subsection might have received the answer "of course," because the verificationist criterion for meaning would have said that it was meaningless to talk about entities that are unobservable in principle. But such replies need not detain us today.

bach's claim, though too strong, was not wholly misguided. In contemplating the notion of "beables" for quantum field theory, J. S. Bell wrote:

> As with relativity before Einstein, there is then a preferred frame in the formulation of the theory ... but it is experimentally indistinguishable. It seems an eccentric way to make a world.

[170] (p. 180, ellipses in the original; see also p. 155). (This seems to have been Bell's *a priori* judgement of the idea. While he thought it somewhat odd, he nevertheless thought it worthwhile to consider as "the cheapest resolution" of what he saw *a posteriori* to be a real difficulty posed by the Aspect experiments [168].) Philosopher T. Maudlin, in attempting to make sense of quantum mechanics and reconcile it to relativity, suggests that backwards causation or a preferred reference frame might be the least unacceptable ways of doing so [197]. Concerning the possibility of a preferred frame in making sense of quantum mechanics, Maudlin, perhaps having in view Einstein's line that God is subtle but not malicious, writes:

> One way or another, God has played us a nasty trick. The voice of Nature has always been faint, but in this case it speaks in riddles and mumbles as well. Quantum theory and Relativity seem not to directly contradict each other, but neither can they easily be reconciled. Something has to give: either Relativity or some foundational element of our world-picture must be modified ... the real challenge falls to the theologians of physics, who must justify the ways of a deity who is, if not evil, at least extremely mischievous.[197] (p. 242)

W. L. Craig has also suggested that "the deep-seated conviction that comes to expression in Einstein's aphorism" is "[p]robably at the root of many physicists' rejection of a neo-Lorentzian approach to relativity theory" [187] (p. 184).

So if one is persuaded that the flow of time is objective and that a preferred foliation ought to show itself readily in physics, then one might consider the formalism above for gravitation with $n = 4$, or perhaps some other way of including the foliation in physics. D. Bohm's nonlocal deterministic version of quantum mechanics is perhaps presently the most vibrant work that assumes a preferred foliation of 4-dimensional spacetime; the theory is presently being applied to quantum gravity and quantum cosmology [199, 200]. (But its nonlocality is not easy to embrace, even if one can tolerate a preferred frame.) However, as Butterfield and Isham note, "[m]ost general relativists feel [that] this response is too radical to countenance: they regard foliation independence as an undeniable insight of relativity" [201].[7]

We suggest that the following explanatory strategy might relieve this tension between the philosophical support for objective becoming and the dearth of physical support for a preferred foliation, at least if theism is plausible. That is, let us take Einstein's remark about God as a theological truth claim and evaluate it.[8]

Maudlin has framed the issue as a problem of evil. The "theologians of physics" might therefore naturally review the usual types of answers to problems of evil, and hit upon a greater-good defense. We suggest the following as one such attempt: Rather than regarding the inclusion of a physically invisible (or nearly so) preferred foliation as an "eccentric" [170] or "if not evil, at least extremely mischievous" [197] way to make a world, one might suggest that the Maker rendered the preferred foliation physically invisible as an act of *benevolence* to physicists. To put it more plainly, the world was made Lorentz-invariant to make physics easier. As we noted above, the requirement of Lorentz invariance so restricts the possible theories that the principle of Lorentz invariance answers a vast number of questions that would otherwise require laborious experimentation to settle. Writing down, *e.g.*, all possible terms in linearized gravity, first given Lorentz invariance, and then given a

[7]We thank Prof. C. Rovelli for stimulating correspondence on this issue.

[8]Einstein's "God" was that of Spinoza [198], so we are perhaps taking Einstein more literally than he intended. However, theism is a common idea whose explanatory powers are widely thought to be worth discussing, so we proceed.

preferred foliation, would give one a clear sense of the economy afforded by Lorentz invariance. (Perhaps other symmetries are amenable to a similar interpretation.) We will content ourselves with the simpler scalar and vector field cases. Viewed from a spacetime perspective, rendering the foliation invisible amounts to keeping the foliation from appearing nontrivially in the action, with only the flat metric present. Viewed from the perspective of space-at-a-time, it means that the foliation only appears in concert with the spatial metric, in such a way that, along with a fundamental constant with dimensions of velocity, neither the foliation nor the spatial metric appears alone, but only the two combined into an effective spacetime metric.

We will now write down all possible terms that could appear in the Lagrangian density for a real scalar field ϕ, restricting the equations of motion to be linear and to have (at most) second derivatives. The existence of fundamental constants with dimensions of velocity (c) and angular momentum ($\hbar$) will be assumed, although we choose units in which these constants have unit value. Because the equations of motion, which are perhaps more important than the Lagrangian density, are unchanged by the addition of a divergence, we will regard terms that are equal up to a divergence as equivalent to avoid overcounting, and will drop terms that are themselves divergences. For brevity, we use a vertical bar to denote the flat metric's covariant derivative. We also use the overdot notation to indicate the time derivative $n^{\mu}\partial_{\mu}$. The possible terms lacking n_{μ} are $\phi_{|\mu}\phi^{|\mu}$ and ϕ^2. The first term is so basic that we will assume that it must always be present, even if the foliation appears in the theory. The values of all other coefficients are relative to this one, which depends on the normalization only. (Our results thus are not inflated, but rather perhaps a bit pessimistic.) If n_{μ} is also present, then $\dot{\phi}^2$ is also available. Thus there are 2 unspecified constants in the case with a preferred foliation observable, but there is only 1 with it absent, in the case of a scalar field. This is a savings already, though a modest one. The vector case will be more compelling.

In the case of a vector field, the possible terms in the Lorentz-invariant case, taking into account the restrictions above are these: $A^{\mu}_{|\nu}A^{|\nu}_{\mu}$, $A^{\mu}_{|\mu}A^{\nu}_{|\nu}$, $A^{\mu}A_{\mu}$, and $\epsilon^{\mu\nu\alpha\beta}A_{\mu|\nu}A_{\alpha|\beta}$, leaving 3 unspecified coefficients. If Lorentz invariance is not required, then there appear in addition $\dot{A}^{\mu}\dot{A}^{\nu}n_{\mu}n_{\nu}$, $A^{\mu}_{|\sigma}A^{\nu|\sigma}n_{\mu}n_{\nu}$, $\dot{A}^{\mu}\dot{A}_{\mu}$, $\dot{A}^{\mu}A^{\nu}_{|\nu}n_{\mu}$, $A^{\mu}A^{\nu}n_{\mu}n_{\nu}$, $A^{\mu}_{|\mu}A^{\nu}n_{\nu}$, and $\epsilon^{\mu\nu\alpha\beta}A_{\mu|\nu}A_{\alpha}n_{\beta}$, giving 10 unspecified coefficients without Lorentz invariance. The economy afforded by Lorentz invariance is thus considerable in the case of a vector field. One expects that it would be substantial for higher-rank fields as well. We have not considered complex fields or fermionic matter, but we imagine that Lorentz invariance provides a respectable simplification in those cases as well. Complex fields have an interesting property not shared by a single real scalar field, *viz.*, they admit first-order-in-time equations of motion such as the Schrödinger equation. (For a single real field, the most similar term in the Lagrangian density, $\phi\dot{\phi}$, is merely a divergence.) So the investigation of these other types of fields, and of sets of real fields with internal symmetry groups, might be of interest.

In view of these simplifications, it is not too implausible to think that temporal becoming is objective, and yet physics is exactly Lorentz invariant, if the existence of a benevolent God who supports the enterprise of physics is plausible. Such a divine motivation does fit naturally within traditional monotheistic religion. According to the traditional story (Genesis 1:28), God tells the human race to reproduce and to fill the earth and subdue it, and to rule over other creatures. This command, which has been dubbed the "cultural mandate," has been seen as encouraging the scientific study of natural phenomena. If this conclusion is correct, then physicists and philosophers can pursue their respective visions of time without mutual interference. The approach outlined above, if successful, permits one to take an antirealist attitude toward relativity in metaphysics, while remaining quite comfortable with the 4-dimensionalist picture urged by Minkowski and so useful in physics. One might find this resolution of the tension physically preferable to Craig's approach [187], which will send one looking for "physical explanations" of relativistic effects, a search which, as we suggested above, has not obviously been fruitful, and which leaves one suspicious of Lorentz invariance and all its benefits. With a plausibility argument in hand for antirealism about any metaphysical claims for relativity, but contentment with relativity as a piece of physics–one might call this a gentle and sympathetic antirealism–one can perhaps say that the great gulf that Stapp [202] and Horwitz *et al.* [141] have found between "Einstein time" and "process time" is not so fixed. The need for a preferred foliation in physics would need to be shown on grounds other than the flow of time, if at all. The preferred foliation would be an "irrelevant variable" in the terminology of ([2]).

This suggestion just made also answers an objection against the idea that God is temporal, *i.e.*, has a location and a duration in time [203–206]. A temporal God's knowledge of which events are objectively simultaneous defines a preferred foliation. But we have already addressed the relativistic objection to a preferred foliation. But if a metaphysically preferred foliation plausibly needn't have physically observable effects, then one can regard God as temporal without embarrassment about relativity.

We close by concluding that even if one is persuaded of the existence of a metaphysically preferred foliation of spacetime, there presently seems to be no compelling reason for rejecting standard Lorentz-invariant physics. However, there are some reasons strong enough to make consideration of the violation of Lorentz invariance an interesting pursuit. With an eye to our discussion of the relation between the flat and curved metrics' null cones below, it seems proper to point out how a preferred foliation might be included, assuming that backward causation (with respect to the preferred time) is to be avoided. If physics is Lorentz invariant but a metaphysically preferred foliation exists, the dynamical null cone of the curved metric is still threatens backward causation. The reason is that if the curved metric's null cone violates the flat metric's null cone in some reference frame, this violation could just as easily occur in the preferred frame as any other. Thus, the null cone discussion below will apply unchanged even if a physically unobservable preferred foliation is present.

Chapter 4

Slightly Bimetric Theories of Gravity

4.1 Slightly Bimetric Theories Derived using Traceless Universal Coupling and Restricted Free Field Invariance

The possibility of deriving general relativity in flat spacetime is fairly well-known, though we believe the above derivation to be especially clear. One naturally asks, can anything new, something besides general relativity and other generally covariant theories (with higher derivatives), be obtained from a procedure along these lines? In fact, other theories can be derived. We will now show a larger family of theories that can be obtained by making two modifications. One relaxes universal coupling to apply only to the traceless part of the stress tensor, while the other restricts the free field gauge invariance to divergenceless vector fields.

Under conformal transformations, a metric tensor factors into two pieces. One is the conformally invariant part, the densitized metric $\hat{\eta}_{\mu\nu}$ of weight $-\frac{1}{2}$, which has determinant $\hat{\eta} = -1$. This quantity determines the flat metric's null cone structure. Its inverse, the weight $\frac{1}{2}$ density $\hat{\eta}^{\mu\nu}$, also has determinant -1. Using the matrix relation $\delta \det(A) = (\det A) Tr(A^{-1} \delta A)$, one sees that $\delta \hat{\eta}_{\mu\nu}$ and consequently $\frac{\delta S}{\delta \hat{\eta}_{\mu\nu}}$ are traceless. The other, conformally variant factor is $\sqrt{-\eta}^{\frac{1}{2}}$, where η is the determinant of $\eta_{\mu\nu}$. (We shall work with $\sqrt{-\eta}$ rather than its square root, but nothing important depends on this choice.) Recalling the derivation of the metric stress tensor above, one sees that (apart from trivial factors) the traceless part of the stress tensor comes from $\hat{\eta}_{\mu\nu}$ and the trace comes from $\sqrt{-\eta}$. As was just shown, universal coupling to the total stress tensor yields an effectively Riemannian theory. It is known that in massless scalar gravity, universal coupling to the trace of the stress tensor yields a conformally flat Riemannian theory: the determinant of the flat metric is completely "clothed" by the gravitational field [20, 236, 237]. Thus, one suspects that treating the traceless and trace parts of the stress tensor differently might yield interesting results. Anticipating some of our results, we observe the pattern that whatever part of the stress tensor (the whole, the trace, or the traceless part) is universally coupled to gravity, the corresponding part of the flat metric (the whole, the determinant, or the conformally invariant part, respectively) is entirely "clothed" by the gravitational field and rendered unobservable (if the field is massless).

We therefore write a general action for a gravitational field and bosonic matter as $S[\hat{\eta}_{\mu\nu}, \sqrt{-\eta}, \gamma_{\mu\nu}, u]$, with the gravitational field $\gamma_{\mu\nu}$ taken as a density of weight $-\frac{1}{2}$ to match $\hat{\eta}_{\mu\nu}$. The Lie derivative of tensor densities requires care. For a (1, 1) density of weight w, the form is [222]

$$\pounds_{\xi} \phi^{\alpha}_{\beta} = \xi^{\mu} \phi^{\alpha}_{\beta},_{\mu} - \phi^{\mu}_{\beta} \xi^{\alpha},_{\mu} + \phi^{\alpha}_{\mu} \xi^{\mu},_{\beta} + w \phi^{\alpha}_{\beta} \xi^{\mu},_{\mu} . \tag{4.1}$$

The form for any tensor density is readily generalized from this expression.

The metric stress tensor can be split up into traceless and trace parts by reworking the earlier derivation. One has

$$\delta S = \int d^4x (\frac{\delta S}{\delta \gamma_{\mu\nu}} \pounds_\xi \gamma_{\mu\nu} + \frac{\delta S}{\delta u} \pounds_\xi u + \frac{\delta S}{\delta \hat{\eta}_{\mu\nu}} \pounds_\xi \hat{\eta}_{\mu\nu} + \frac{\delta S}{\delta \sqrt{-\eta}} \pounds_\xi \sqrt{-\eta}) = 0. \tag{4.2}$$

Letting the matter and gravitational field equations hold gives

$$\delta S = \int d^4x (\frac{\delta S}{\delta \hat{\eta}_{\mu\nu}} \pounds_\xi \hat{\eta}_{\mu\nu} + \frac{\delta S}{\delta \sqrt{-\eta}} \pounds_\xi \sqrt{-\eta}) = 0. \tag{4.3}$$

Local energy-momentum conservation takes the form

$$\partial_\mu (2 \frac{\delta S}{\delta \hat{\eta}_{\mu\nu}} + \frac{\delta S}{\delta \sqrt{-\eta}} \sqrt{-\eta} \hat{\eta}^{\mu\nu}) = 0. \tag{4.4}$$

It is convenient to introduce the following change of variables:

$$S[\hat{\eta}_{\mu\nu}, \sqrt{-\eta}, \gamma_{\mu\nu}, u] = S[\hat{\eta}_{\mu\nu}, \sqrt{-\eta}, \tilde{g}_{\mu\nu}, u], \tag{4.5}$$

where

$$\tilde{g}_{\mu\nu} = \hat{\eta}_{\mu\nu} - \lambda \gamma_{\mu\nu}. \tag{4.6}$$

The reason for taking the gravitational field to be (0,2) weight $-\frac{1}{2}$ is now clear: doing so makes it easy to add the gravitational potential to the conformally invariant part of the flat metric. (Plainly a (2,0) weight $\frac{1}{2}$ field would work equally well, *mutatis mutandis.*) Taking care with the trace, one finds that

$$\frac{\delta S}{\delta \hat{\eta}_{\mu\nu}} |\gamma = \frac{\delta S}{\delta \hat{\eta}_{\mu\nu}} |g + \frac{\delta S}{\delta \tilde{g}_{\alpha\beta}} P^{\mu\nu}_{\alpha\beta} \tag{4.7}$$

and

$$\frac{\delta S}{\delta \gamma_{\mu\nu}} = -\lambda \frac{\delta S}{\tilde{g}_{\mu\nu}}, \tag{4.8}$$

where

$$P^{\mu\nu}_{\alpha\beta} = \delta^{(\mu}_{\alpha} \delta^{\nu)}_{\beta} - \frac{1}{4} \eta^{\mu\nu} \eta_{\alpha\beta} \tag{4.9}$$

is the traceless symmetric projection tensor with respect to $\eta_{\mu\nu}$. Combining these two results gives

$$\lambda \frac{\delta S}{\delta \hat{\eta}_{\mu\nu}}|\gamma = \lambda \frac{\delta S}{\delta \hat{\eta}_{\mu\nu}}|g - \frac{\delta S}{\delta \gamma_{\alpha\beta}} P^{\mu\nu}_{\alpha\beta}, \tag{4.10}$$

which splits the traceless part of the stress tensor into a part that vanishes on-shell and another that depends on how much of the conformally invariant part of the flat metric remains after the change of variables.

We now introduce the physical postulate of traceless universal coupling:

$$\frac{\delta S}{\delta \gamma_{\alpha\beta}} P^{\mu\nu}_{\alpha\beta} = \frac{\delta S_f}{\delta \gamma_{\alpha\beta}} P^{\mu\nu}_{\alpha\beta} - \lambda \frac{\delta S}{\delta \hat{\eta}_{\mu\nu}}|\gamma; \tag{4.11}$$

in words, the traceless part of the full field equations equals the traceless part of the free field equations coupled to the traceless part of the stress tensor. This postulate will let us explore what theories, besides Riemannian and conformally flat Riemannian theories, can be obtained from a slightly relaxed version of universal coupling. Combining equations (4.10) and (4.11) gives

$$\lambda \frac{\delta S}{\delta \hat{\eta}_{\mu\nu}}|g = \frac{\delta S_f}{\delta \gamma_{\alpha\beta}} P^{\mu\nu}_{\alpha\beta}. \tag{4.12}$$

The traceless part of the free field equations must equal a term derived from how the flat metric remains in the action after the change to the bimetric variables.

This result suggests that it would be useful to have a result concerning $\partial_\mu \frac{\delta S_f}{\delta \gamma_{\alpha\beta}} P^{\mu\nu}_{\alpha\beta}$ derived from an infinitesimal invariance. In order that only the traceless part of the free field equations be involved, the variation of the gravitational field ought itself to be traceless. We require that S_f change at most by a boundary term under the infinitesimal transformation $\gamma_{\mu\nu} \rightarrow \gamma_{\mu\nu} + \delta\gamma_{\mu\nu}$, where $\delta\gamma_{\mu\nu} = \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$, but with ξ_μ restricted so that

$$\partial_\mu \xi^\mu = 0. \tag{4.13}$$

Now ξ_ν is a density of weight $-\frac{1}{2}$. Others using a similarly restricted invariance have restricted $\gamma_{\mu\nu}\eta^{\mu\nu}$ to vanish[238–242], but we leave it arbitrary, anticipating that another degree of freedom might appear. This gauge invariance is consistent with a non-zero mass and self-interaction potential for the trace part of the gravitational field. Given the various reasons for which scalar fields are presently postulated, such as inflation and dark matter, it would be welcome to find an extra scalar field without postulating it *ad hoc*. If these uses require that the new degree of freedom be massive, then it is easier to see why the three classical tests of general relativity have not detected it, whereas a massless scalar ought to have yielded incorrect light bending properties. That is because the scalar would attract (or perhaps repel, if the 'wrong' sign could be tolerated, though that seems doubtful) slow-moving objects, but would not bend light [20] If such a scalar exists, perhaps tests of gravity at short range would find more (or less) attraction than expected.

One can write

$$\xi^{\mu} = \partial_{\nu}\mathcal{F}^{\mu\nu}, \tag{4.14}$$

with $\mathcal{F}^{\mu\nu}$ an arbitrary antisymmetric field of suitable weight. Repeated integration by parts and the arbitrariness of $\mathcal{F}^{\mu\nu}$ entail that

$$\partial_{\mu}\partial^{[\rho}P^{\nu]\mu}_{\alpha\beta}\frac{\delta S_f}{\delta\gamma_{\alpha\beta}} = 0, \tag{4.15}$$

which means that the divergence of the traceless part of the free field equations equals the gradient of some function. Recalling equation (4.12), one shows that $\partial_{\mu}\frac{\delta S}{\delta\hat{\eta}_{\mu\nu}}|g$ is a gradient. If one splits the full action S into S_1 and S_2, then S_2 can take the same form as above for general relativity. S_1 can have the form $S_1[\tilde{g}_{\mu\nu}, \sqrt{-\eta}, u]$, with the $\hat{\eta}_{\mu\nu}$ absent. We have not found any other solutions to equation (4.12).

It is useful to make a further change of variables from a densitized curved metric to an ordinary one by

$$g_{\mu\nu} = \tilde{g}_{\mu\nu}\sqrt{-\eta}^{\frac{1}{2}}. \tag{4.16}$$

The Euler-Lagrange equations change trivially: $\frac{\delta S}{\delta\tilde{g}_{\mu\nu}} = \frac{\delta S}{\delta g_{\mu\nu}}\sqrt{-\eta}^{\frac{1}{2}}$. We conclude that the general action is

$$S = S_1[g_{\mu\nu}, \sqrt{-\eta}, u] + \frac{1}{2}\int d^4x R_{\mu\nu\rho\sigma}(\eta_{\mu\nu})\mathcal{M}^{\mu\nu\rho\sigma}(\eta_{\mu\nu}, g_{\mu\nu}, u)$$
$$+ \int d^4x(\partial_\mu \alpha^\mu + 2b\sqrt{-\eta}). \qquad (4.17)$$

We call this form "slightly bimetric": "slightly" because only the determinant of $\eta_{\mu\nu}$ enters the Euler-Lagrange equations essentially, not the whole flat metric, and "bimetric" because the whole of $\eta_{\mu\nu}$ is present somewhere in the theory, *viz.*, in the action, in the definition of the stress tensor, and in the definition of ideal lengths and times for objects unaffected by gravity (of which there are none). The restriction of the initial invariance has the consequence that the gravitational field equations alone no longer suffice to yield conservation of energy-momentum; the matter fields u must also obey their equations of motion, at least in part. This last result bears a resemblance to the result of Lee *et. al.* [243] that the "matter response equations" $\nabla_\mu T^{\mu\nu}_{\text{mat}} = 0$ follow from the gravitational field equations if and only if no absolute objects are present in the field equations. These equations still follow, of course, from the matter field equations, assuming that matter couples only to a curved metric [152].

4.2 Gauge Invariance

We now turn to consider the gauge invariance of slightly bimetric theories. Going through the same procedure as for generally covariant theories, we guess that a gauge transformation is given by $\delta g_{\mu\nu} = \pounds_\xi g_{\mu\nu}$, $\delta u = \pounds_\xi u$, $\delta\eta_{\mu\nu} = 0$, but with ξ^μ obeying some restriction. Here ξ^μ has vanishing weight. Thus,

$$\delta S_{gauge} = \delta S_{coord} - \int d^4x(\frac{\delta S}{\delta \eta_{\mu\nu}}|g \pounds_\xi \eta_{\mu\nu} + i^\mu{}_{,\mu})$$
$$= 0 - \int d^4x(-2\xi^\alpha \eta_{\alpha\mu}\partial_\nu \frac{\delta S}{\delta\eta_{\mu\nu}}|g + j^\mu{}_{,\mu}). \qquad (4.18)$$

Recalling that

$$\partial_\mu \frac{\delta S}{\delta \eta_{\mu\nu}}|g = \partial^\nu \psi \tag{4.19}$$

for some scalar density ψ, one sees that δS_{gauge} is indeed a boundary term if and only if $\partial_\mu \xi^\mu = 0$ (unless ψ vanishes, in which case the theory is really generally covariant). Thus, our assumed form of the invariance is verified, and the restriction on ξ^μ is known. The same restriction holds for the full nonlinear theory as held for the linear theory. In this slightly bimetric case, gauge transformations change (bosonic) dynamical fields in the same way that $\eta_{\mu\nu}$-volume-preserving coordinate transformations do, but leave the absolute object $\eta_{\mu\nu}$ unchanged.

4.3 Slightly Bimetric Theories Are Equivalent to Generally Covariant Theories plus a Scalar Field

Having proposed the addition of a flat background metric to general relativity and noted the possibility of constructing alternative theories with this extra ingredient, Rosen himself subsequently devoted considerable energy to a particular bimetric theory of gravity (e.g., [207]), hoping to avoid singularities, which afflict general relativity, and to give simpler partial differential equations than Einstein's. Although Rosen's theory passes a considerable number of empirical tests, it has difficulty with the binary pulsar [244]. More generally, theories into which the flat metric enters the action nontrivially will display various effects which can be tested against experiment. Concerning the matter action, experiment strongly restricts how the flat metric can enter [244], so it makes sense to let matter see only a curved metric, with the unclothed conformally invariant part of the flat metric absent, apart from a term containing the flat metric's Riemann tensor; such a term merely alters the stress tensor by a curl, and does not affect the field equations. (But see [246–248] for recent interest in nonminimal coupling to scalar fields. The assumption of minimal coupling will not be used.) Requiring that the matter stress tensor appear on the right side of the gravitational Euler-Lagrange equations substantially imposes the same condition [65]. The gravitational action has more room for a flat metric to enter, but one expects that theories with more exposed background geometry will have more trouble agreeing with experiment. If only the determinant of the flat metric $\sqrt{-\eta}$ appears in the action nontrivially, then the effects should be testable, but not as constrained as if the whole metric appears. Slightly bimetric theories therefore are perhaps the best chance for empirically viable continuation of Rosen's bimetric program. However, they do not satisfy Rosen's desire for simpler partial differential equations. Whether slightly bimetric theories help to avoid singularities is tied to the success of scalar-tensor theories in doing the same. On this point, reports are mixed [246, 249]. It is known that a suitably nonminimally coupled scalar field can violate all the standard energy conditions, up to and including the averaged null energy condition, which is tied to the possibility of traversable wormholes [235, 341]. Given that slightly bimetric theories permit any desired coupling for the 'extra scalar' (although the relevance of wormholes to a bimetric approach to gravity is questionable, as we will remark below), it follows that slightly bimetric theories can be chosen so as to violate such energy conditions, the satisfaction of which plays a role in the singularity theorems [152]. On the other hand, such scalar fields can permit large negative energy fluxes, which raise the question of violating the generalized second law of thermodynamics [342] and suggest the lack of a stable ground state [261].

It is convenient to split the action into effectual and ineffectual pieces, so we write

$$S = S_e[g_{\mu\nu}, \sqrt{-\eta}, u] + S_i[g_{\mu\nu}, \eta_{\mu\nu} u], \tag{4.20}$$

both terms being scalars. The effectual terms are those that affect the Euler-Lagrange equations. All terms that do not affect the (gravitational or matter)

Euler-Lagrange equations and that contribute at most a curl to $\frac{\delta S}{\eta_{\mu\nu}}$, *viz.*, divergences, flat space 4-volume terms, and terms involving $R_{\mu\nu\rho\sigma}(\eta_{\mu\nu})$, are gathered into the ineffectual term S_i.

Making use of the properties of the action under coordinate transformations, one can derive generalized Bianchi identities [243]. Under an arbitrary infinitesimal coordinate transformation described by a vector field ξ^μ, the action changes by the amount

$$\delta S = \int d^4x (\frac{\delta S_e}{\delta g_{\mu\nu}} \pounds_\xi g_{\mu\nu} + \frac{\delta S_i}{\delta g_{\mu\nu}} \pounds_\xi g_{\mu\nu} + \frac{\delta S_e}{\delta\sqrt{-\eta}} \pounds_\xi \sqrt{-\eta} \\ + \frac{\delta S_i}{\delta \eta_{\mu\nu}} \pounds_\xi \eta_{\mu\nu} + \frac{\delta S_e}{\delta u} \pounds_\xi u + \frac{\delta S_i}{\delta u} \pounds_\xi u) = 0. \tag{4.21}$$

By construction $\frac{\delta S_i}{\delta g_{\mu\nu}}$ and $\frac{\delta S_i}{\delta u}$ vanish identically, so the second and sixth terms do not contribute. One observes that $\frac{\delta S_i}{\delta \eta_{\mu\nu}}$ is a curl, so the fourth term contributes only a boundary term. Letting the matter field equations $\frac{\delta S}{\delta u} = 0$ and the gravitational field equations $\frac{\delta S}{\delta g_{\mu\nu}} = 0$ hold annihilates the first and third terms, so only the second remains:

$$\delta S = \int d^4x \frac{\delta S_e}{\delta\sqrt{-\eta}} \pounds_\xi \sqrt{-\eta} = 0. \tag{4.22}$$

Thus, upon integration, one obtains

$$\frac{\delta S_e}{\delta\sqrt{-\eta}} = J, \tag{4.23}$$

where J is a constant of integration.

This last equation is sufficiently similar in appearance to an Euler-Lagrange equation that one can consider another theory with a dynamical metric, matter fields, and a *dynamical* weight 1 density ψ, with ψ replacing $\sqrt{-\eta}$, plus an additional term:

$$S'[g_{\mu\nu}, u, \psi] = S_e[g_{\mu\nu}, u, \psi] - \int d^4x J\psi. \tag{4.24}$$

The Euler-Lagrange equations for this action are

$$\frac{\delta S'}{\delta g_{\mu\nu}} = 0, \tag{4.25}$$

$$\frac{\delta S'}{\delta u} = 0, \tag{4.26}$$

and

$$\frac{\delta S'}{\delta \psi} = 0. \tag{4.27}$$

The metric and matter equations are identical to those for the original action S_e. The equation for ψ is equivalent to the integrated on-shell identity $\frac{\delta S_e}{\delta\sqrt{-\eta}} = J$ above. The theories differ in substance, for one has an absolute (*i.e.*, nondynamical) object, and J is an integration constant, while the other has no absolute objects, and J is a parameter in the action. But they do not differ in the forms and solutions of the equations: they are empirically indistinguishable. Thus, scalar density-tensor theories are equivalent to slightly bimetric theories in this sense. We emphasize that the coupling of the scalar field to the curved metric is of arbitrary form, not necessarily minimal.

Ordinarily one considers theories with a scalar field, not a scalar density field. In fact the scalar density-tensor theories above can be recast as theories with a scalar field. This recasting involves a change of variables $\phi = \psi/\sqrt{-g}$. Reexpressing the action S' in the new variables gives

$$S''[g_{\mu\nu}, u, \phi] = S'[g_{\mu\nu}, u, \psi]. \tag{4.28}$$

The matter field equations are untouched by this transformation. Under a variation of $g_{\mu\nu}$ and ψ, one obtains

$$\frac{\delta S'}{\delta g_{\mu\nu}}\delta g_{\mu\nu} + \frac{\delta S'}{\delta \psi}\delta\psi = \frac{\delta S''}{\delta g_{\mu\nu}}\delta g_{\mu\nu} + \frac{\delta S''}{\delta \phi}\delta\phi. \tag{4.29}$$

Using $\psi = \phi\sqrt{-g}$ and equating coefficients of δg and of $\delta\phi$ gives

$$\frac{\delta S'}{\delta g_{\mu\nu}}|\psi + \frac{\delta S'}{\delta \psi} g^{\mu\nu}\phi\sqrt{-g}/2 = \frac{\delta S''}{\delta g_{\mu\nu}}|\phi \tag{4.30}$$

and

$$\frac{\delta S'}{\delta \psi}\sqrt{-g} = \frac{\delta S''}{\delta \phi}. \tag{4.31}$$

One sees that the scalar-tensor equations are just linear combinations of the scalar density-tensor equations. Thus every slightly bimetric theory has a scalar-tensor "twin" and *vice versa.*

4.4 General Form for a Slightly Bimetric Theory

If one prohibits derivatives higher than second order (and permits those only linearly) in the Lagrangian density, then the most general slightly bimetric action is of the form

$$S = \frac{1}{16\pi G}\int d^4x[a(\kappa)\sqrt{-g}R(g) + f(\kappa)\sqrt{-g}g^{\mu\nu}\Delta^{\alpha}_{\mu\alpha}\Delta^{\beta}_{\nu\beta} + e(\kappa)\sqrt{-g}]$$
$$+\frac{1}{2}\int d^4x R_{\mu\nu\rho\sigma}(\eta_{\mu\nu})\mathcal{M}^{\mu\nu\rho\sigma}(\eta_{\mu\nu}, g_{\mu\nu}, u) + \int d^4x\partial_\mu\alpha^\mu + S_{\mathrm{mat}}[g_{\mu\nu}, \sqrt{-\eta}, u]. \tag{4.32}$$

The term $2b\sqrt{-\eta}$ has been absorbed into $e(\kappa)\sqrt{-g}$, while the possible term

$$c(\kappa)\sqrt{-g}g^{\mu\nu}\nabla_\mu\nabla_\nu\kappa \tag{4.33}$$

has been absorbed by redefinition of $f(\kappa)$ and α^μ. Employing Rosen's results as above, one can rewrite this action in a Rosen-style form with no second derivatives of either dynamical or absolute variables:

$$S = \frac{1}{16\pi G}\int d^4x\mathfrak{g}^{\mu\rho}[a(\kappa)\Delta^{\sigma}_{\mu\alpha}\Delta^{\alpha}_{\rho\sigma} - \left(a + \kappa\frac{da}{d\kappa}\right)\Delta^{\sigma}_{\rho\mu}\Delta^{\alpha}_{\alpha\sigma} + \left(f(\kappa) + \kappa\frac{da}{d\kappa}\right)\Delta^{\sigma}_{\rho\sigma}\Delta^{\alpha}_{\mu\alpha}]$$
$$+\frac{1}{16\pi G}\int d^4x e(\kappa)\sqrt{-g}] + S_{\mathrm{mat}}[g_{\mu\nu}, \sqrt{-\eta}, u].$$

In writing this form, we have set

$$16\pi G\alpha^{\mu} = -a(\kappa)\Delta^{\mu}_{\rho\sigma}\mathfrak{g}^{\sigma\rho} + a(\kappa)\Delta^{\sigma}_{\rho\sigma}\mathfrak{g}^{\mu\rho} \tag{4.34}$$

and

$$\mathcal{M}^{\mu\nu\rho\sigma} = -a(\kappa)\eta^{\nu\sigma}\mathfrak{g}^{\mu\rho}/8\pi G. \tag{4.35}$$

Using (4.32) one finds the Euler-Lagrange equations of motion to be

$$\begin{aligned}\frac{16\pi G}{\sqrt{-g}}\frac{\delta S}{\delta g_{\mu\nu}} &= -aG^{\mu\nu} + \frac{\kappa}{2}a' g^{\mu\nu}R + \nabla^{\mu}\nabla^{\nu}a - g^{\mu\nu}\nabla^2 a \\ &\quad -\nabla_{\sigma}(f g^{\mu\nu}\Delta^{\alpha}_{\rho\alpha}g^{\sigma\rho}) - f\Delta^{\sigma}_{\alpha\sigma}\Delta^{\rho}_{\beta\rho}g^{\alpha\mu}g^{\beta\nu} \\ +\left(\frac{f+f'\kappa}{2}\right)\Delta^{\sigma}_{\alpha\sigma}\Delta^{\rho}_{\beta\rho}g^{\alpha\beta}g^{\mu\nu} &+ \frac{1}{2}g^{\mu\nu}(e+e'k) + \frac{16\pi G}{\sqrt{-g}}\frac{\delta S_{\mathrm{mat}}}{\delta g_{\mu\nu}} = 0.\end{aligned} \tag{4.36}$$

One can split S into S_e and S_i as before. Employing the machinery used above in finding the generalized Bianchi identities and using the matter and gravitational equations of motion, one obtains

$$\frac{\delta S_e}{\delta\sqrt{-\eta}},_{\mu} = 0, \tag{4.37}$$

or, upon integration,

$$\frac{\delta S_e}{\delta\sqrt{-\eta}} = J, \tag{4.38}$$

where J is a constant of integration. The explicit form of $\frac{\delta S_e}{\delta\sqrt{-\eta}}$ is

$$\frac{\delta S_e}{\delta\sqrt{-\eta}} = \frac{\delta S_{\mathrm{mat}}}{\delta\sqrt{-\eta}} + \frac{1}{16\pi G}(-a'\kappa^2 R + \partial_{\nu}(2f\nabla^{\nu}\kappa) - f' g^{\mu\nu}\kappa,_{\mu}\kappa,_{\nu} - e'\kappa^2). \tag{4.39}$$

By making a conformal transformation to the Einstein frame, one can typically set $a = 1$. One reason not to do so at this stage is because the above action contains degenerate cases related to *unimodular* general relativity [238–242, 250], which involve $a = \sqrt{\kappa}$. In these cases, the Ricci scalar term pertains to a curved metric whose determinant is just that of the flat metric and thus nondynamical; in searching for new theories, one wants not to lose sight of any special cases. Also, nongravitational experiments are governed by the metric which is conformally coupled to matter (as will be discussed below), if one exists; typically that is not the Einstein frame's metric. Otherwise, setting $a = 1$ is convenient.

4.5 Some Special Cases and Empirical Consequences

Slightly bimetric theories, or their scalar-tensor twins, split into a number of cases, among which are (theories equivalent to) generalized Brans-Dicke [244, 245] theories, general relativity without or with a scalar field, unimodular general relativity [238–242, 250], and some others, including an interesting scalar-tensor theory with no explicit kinetic term for the scalar field [252]. General relativity itself is of course a trivial example of a slightly bimetric theory. An attractive example of general relativity with a scalar field was briefly considered by Avakian and Grigorian [251]; however, their refutation of the theory, which corresponds to an unspecified constant a_3 in their notation, cannot be accepted because the theory manifestly *includes* general relativity, and thus every solution of the Einstein field equations, as a special case. This theory is very similar to the "restricted gravity" of Dragon and Kreuzer, who find a massive "dilaton" in the metric [233]. Unimodular general relativity sets $\sqrt{-g} = \sqrt{-\eta}$ *a priori*, so the traceless part of the Einstein equations are the Euler-Lagrange equations. The Bianchi identities restore the trace of the Einstein equations, up to an integration constant. It is interesting to note that in considering the "most general linear theory of gravitation", Nachtmann, Schmidle, and Sexl omitted the case in which matter is coupled only to the traceless part of the gravitational field [35, 36]. Such a case corresponds to coupling to a covariantly unimodular matter metric in the nonlinear theory.

One readily sees that some slightly bimetric theories contain general relativity (perhaps with the covariantly unimodular condition $\kappa = 1$) as a special case. Full consideration of the empirical properties of the theories requires dividing the family of theories into natural cases; the theories do not even all have the same number of degrees of freedom. Various equivalence principles are satisfied, or violated, as the case may be, for particular slightly bimetric theories, so different versions might provide tests of various equivalence principles. Theories in which matter is not universally coupled will tend to violate the weak equivalence principle [2]. Because some slightly bimetric theories reduce to general relativity in a suitable limit, these versions ought to remain viable as long as general relativity's outstanding track record persists, at least if the general relativistic limit is stable. Full consideration of these matters awaits another time.

The bimetric scalar tensor theory of the Armenian school (for example, [253]) is not slightly bimetric. This fact is clear because the field equations contain the tensor $\Delta^{\alpha}_{\mu\nu}$, and not merely its contraction $\Delta^{\mu}_{\mu\nu}$.

4.6 A Built-in 'Scalar Field' and Cosmological Problems

The scalar degree of freedom present in some slightly bimetric theories could perhaps be detected once gravitational wave astronomy is well under way [254]. In any case, there exists monopole radiation in general in such theories. The issue of gravitational radiation for scalar tensor theories has been investigated. by R. V. Wagoner [255]. In addition, it might facilitate inflationary cosmological models, because it can be nonminimally coupled [256]. Or it might serve as a form of dark matter. There have been a number of studies of scalar field dark matter recently [257]. For minimally coupled matter, the scalar field acts as "noninteracting dark matter," which interacts only with itself and gravity. This form of dark matter has recently been considered by Peebles and Vilenkin [258].

Using the scalar-tensor twin of a slightly bimetric theory should permit carrying over many results from scalar-tensor theories to slightly bimetric theories, such as issues of positive scalar field energy [228].

Concerning the cosmological constant, theorists have been interested in explaining the difference between its quantum-mechanically predicted large value and its observed small value—this is the "cosmological constant problem" [259]. (At least, this is the "old cosmological constant problem"; recently new cosmological constant problems have arisen [250].) One approach that has attracted attention is unimodular general relativity [238–241, 250], because the cosmological constant is in that case not a coupling constant in the action, but a datum in the initial conditions. Other slightly bimetric theories behave in the same fashion, the integration constant J being related to an effective cosmological constant, so they retain this advantage in addressing this problem. Based on recent supernova measurements, it might be necessary to include an effective cosmological constant [229, 230]. Receiving it as a constant of integration is much more appealing than the traditional way by putting a term linear in the gravitational field into the action, for such an action defines a theory in which the field about a point source grows with distance, behavior which is difficult to accept [260]. Finally, we observe that A. N. Petrov has considered in a rather general fashion the possibility of obtaining the cosmological constant as a constant of integration [267].

Recently, some variable speed-of-light (VSL) theories have been proposed [231]. One type fixes the gravitational 'light' speed, while allowing the electromagnetic light speed to vary, there being two different curved metrics in the theory [231]. One could define an electromagnetic metric using a scalar field χ by the equation

$$[g_{em}]_{\alpha\beta} = g_{\alpha\beta} - K(\nabla_\alpha \chi)\nabla_\beta \chi, \tag{4.40}$$

K being constant. As we have seen above, slightly bimetric theories can mimic the presence of a scalar field, so it is not surprising that VSL theories can also be slightly bimetric. If one replaces χ by $\ln\sqrt{\frac{g}{\eta}}$, then one gets an electromagnetic metric of the form $[g_{em}]_{\alpha\beta} = g_{\alpha\beta} - K\Delta_\alpha\Delta_\beta$, where $\Delta^\mu_{\alpha\mu} = \Delta_\alpha$.

4.7 Classical Local Empirical Aspects of Bimetric Theories

It might be useful to explain why the bimetric/field approach to general relativity is empirically equivalent to the geometrical form, at least locally and at the classical level. (This issue has also been addressed by Thirring [29], Freund *et al.* [260], and Zel'dovich and Grishchuk [83]. But we will argue that the theories are in fact distinct, so the qualifications "at least locally and at the classical level" will be important.) Questions might arise due to the fact that measurements of times and lengths in the geometrical theory are *assumed* to be governed by $g_{\mu\nu}$, there being no other metric tensor to choose; but if $\eta_{\mu\nu}$ is also present, then other choices might seem possible. This discussion will also help to give the empirical interpretation of slightly bimetric theories.

If one considers what an 'ideal' rod or clock might be, the geometrical view says that it is one governed by $g_{\mu\nu}$ [151], whereas the bimetric approach says that it is one that is unaffected by gravity and thus governed by $\eta_{\mu\nu}$. But it is real rods and clocks, not ideal ones, that are used in experiments. J. L. Anderson has recently argued that a metric in general relativity is unnecessary, because the behavior of rods and clocks can be determined *via* the Einstein-Infeld-Hoffmann procedure [271]. Even if such a procedure were impossible in practice, it would remain true that the behavior of real rods and clocks would be completely determined (classically) by the partial differential equations obeyed by all the fields, for, in light of modern field theory, real rods and clocks are just congealed field excitations. Conceptually, there is no room for a separate postulate of the behavior of length and time measurements.

Because the bimetric and geometrical approaches to general relativity yield identical partial differential equations for $g_{\mu\nu}$ and matter fields u, it follows that the two approaches are empirically equivalent. Thus, once the obsolete dualism between matter and field is removed, it becomes clear that these two approaches to general relativity are equivalent empirically, at least locally and classically.

In the case of slightly bimetric theories, it is no longer the case that the flat background metric is entirely clothed. So how does one interpret measurements? Here the existence of a scalar-tensor "twin" for each slightly bimetric theory is useful. Assuming that the usual postulated relation between measurements in general relativity and the partial differential equations of general relativity is consistent, the same results can be carried over to slightly bimetric theories *via* their scalar-tensor twins. Scalar-tensor theories are specific examples of general relativity coupled to a scalar field. In some theories, there exists a "Jordan frame" in which matter is minimally coupled, as in general relativity. General relativity assumes nongravitational experiments to be described by the metric minimally coupled to matter. The scalar field should not make any difference, for one could regard it as merely another matter field. So the relevant metric for typical experiments is the one minimally coupled to matter, if such a thing exists.

This claim, it should be noted, does not entail any claim about the controversy regarding which of the conformally related metrics is "the physical metric." This question has been reviewed recently [261] by V. Faraoni *et al.*, who cite worries about negative energy in the Jordan frame as one reason to believe that, as they conclude, the Jordan conformal 'frame' is unacceptable. If this argument is sound, then the number of viable scalar-tensor theories might be rather limited. These authors state that the issue of the empirical viability of scalar-tensor theories needs to be reconsidered in view of the result that the Einstein frame is the physical one.

However, we find the concept of energy in a nonminimally coupled theory without a flat background metric to be unclear. For example, the vanishing of the *covariant* divergence of the matter stress tensor is repeatedly described as a conservation law by Faraoni *et al.*, without any hesitation, but we saw above that this equation typically fails to describe the global conservation of anything. Further, the heritage of the geometrical form of gravitation is that non-gravitational energy density is well-defined, but gravitational energy density is not. Scalar-tensor theories bring about a situation in which the scalar field is neither clearly gravity, nor clearly matter, and conformal transformations tend to shift its status one way or the other. Thus, such arguments about scalar field energy densities as Faraoni *et al.* cite might be unreliable. We suggest that a field formulation of scalar-tensor theories on a flat background might illuminate this issue. As the variable transformation $g_{\mu\nu} = \eta_{\mu\nu} - \lambda\gamma_{\mu\nu}$ in our derivation of general relativity showed, a field redefinition involving the flat background metric should change the stress tensor only by terms proportional to the equations of motion, if the flat metric itself is untouched; clearly redefinitions not involving the flat metric will not change the stress tensor at all. Given that gauge transformations with a flat background metric resemble diffeomorphisms in some respects, as we have seen, but conformal transformations generally do not [152], one need not worry that the conformally transformation will lead to ambiguity regarding the *flat* metric. In this way can define a stress-energy tensor including gravity and the scalar field, and whatever other fields might be present. Such a tensor will of course be gauge-variant, as in general relativity, but the integrals of motion have proven adequately gauge-invariant for Einstein's theory. The results will be independent of all notions of which curved metric is "the physical one", so the possibility of determining which metric is physical based on positive energy issues will not arise, *pace* Varaoni *et al.* Some results on energy-momentum complexes or scalar field stress tensors have been found [228, 262–264], which results presumably will have field formulation analogs.

Chapter 5

Bimetric General Relativity and Null Cone Consistency: A History Since 1939

As we have seen, the use of a flat metric tensor $\eta_{\mu\nu}$ in gravitation has received a fair amount of attention over the last six decades or so. However, the interpretation of the resulting bimetric or field formulation of general relativity has not been adequately clarified, due to an ambiguous notion of causality: the effective curved metric which determines matter propagation is not obviously consistent with the flat background causal structure. Having a consistent relationship is clearly a *necessary* condition for a true special-relativistic theory. Examples of consistent and inconsistent relationships are given in figure 5.1. Whether it is *sufficient* is unclear, because the propagation of gravity itself, or of matter fields coupled in a nonstandard way, can lead to more subtle behavior [231, 265]. However, this question of null cone consistency clearly is important. This issue holds for slightly bimetric theories, also, although our proposed solution is less plausible for slightly bimetric theories than for generally covariant ones.

We will attempt to sketch the history of the issue from roughly the late 1930s till the present. We do not consider the period between 1905 and the late 1930s, though doubtless that would be an interesting project, which would consider the time after special relativity had solidified and include the invention of Einstein's theory of gravitation. (L. P. Grishchuk mentions a bit of the history in works of Poincaré and Einstein [85]. Fang and Fronsdal sketch the history of the flat spacetime approach up to 1979 [79], but neglect to consider the null cone issue.) Rather, we start with the *rebirth* of the flat spacetime approach to gravity, with works by Fierz, Rosen *et al.* While the importance of the problem perhaps seems evident in retrospect, the neglect of it in the literature suggests that it in practice was not so obvious, or that philosophical considerations disposed some to disregard it. One can crudely divide the issue's history into three periods (though at times we will disregard the historical boundaries to be able to discuss an author's whole work in a unified way). For the first 20 years (1939-1959), the problem seems not to have been recognized or mentioned, at least not in print (to our knowledge). For the next 20 years (1959-1979), it was sometimes mentioned, but either resolved incorrectly, dismissed as unimportant, or postponed with the hope that it would disappear. More recently (1979-2001), it has been recognized more often, and occasionally regarded as worthy of sustained attention. A few authors have attempted either to solve it or to prove it insoluble. However, we disagree that either of these goals has been achieved.

One should perhaps distinguish between two null cone problems. The first is: given that one regards Einstein's equations as describing the evolution of an effective curved metric in Minkowski spacetime, what does one make of the potential violation of Minkowski causality by matter responding to the curved metric? The second is: given that one chooses to quantize the geometrical (single-metric) theory, what does one make of causality without a metric to define equal-time commutation relations? However, these problems are related, and we believe that the SRA as presented here solves them both, so we will treat them together.

5.1 The Years 1939 to 1959: the Null Cone Consistency Problem Ignored

Around 1940, in his seminal papers on the bimetric description of general relativity, N. Rosen suggested that there ought to be some (gauge-fixing) relation between the flat and curved metrics, because one expects that the two coincide if the gravitational field vanishes [6] (p. 149). While this paper did not consider the meaning of the bimetric formalism in detail, its companion paper (p. 150) considered interpretive issues. Rosen wrote (apart from a change in notation to match ours),

> [f]rom the standpoint of the general theory of relativity, one must look upon $\eta_{\mu\nu}$ as a fiction introduced for mathematical convenience. However, the question arises whether it may not be possible to adopt a different point of view, one in which the metric tensor $\eta_{\mu\nu}$ is given a real physical significance as describing the geometrical properties of space, which is therefore taken to be flat, whereas the tensor $g_{\mu\nu}$ is to be regarded as describing the gravitational field. [6] (p. 150).

Rosen recognized that the flat spacetime view implies that the speed of light (measured with ideal rods and clocks, which are not distorted by gravity) will tend to differ from unity [6] (p. 153), but he seems not to have addressed the possibility that it might *exceed* 1. Finally, while his approach merely postulated bimetric general relativity, he did suggest that it would be desirable to derive it independently [6] (p. 153). His intention to carry out this procedure himself [6] (p. 153) seems not to have been realized, but many others have done it since that time, as did we in an earlier chapter.

During the 1940s, with some war-time inconvenience in Greece, A. Papapetrou was able to express general relativity in an attractive form resembling electromagnetism, with the theories being expressed in the tensorial DeDonder and Lorentz gauges, respectively [11]. He emphasized the improved nature of the conservation laws, especially for angular momentum, and found that certain attractive relations that have no invariant meaning in the geometrical view become perspicuous given the flat spacetime interpretation. Papapetrou held that for the flat spacetime approach, gauge-fixing to tie together the two metrics was "indispensable" (p. 20), because the energy-momentum and angular momentum localization would suggest physically distinct systems given different relations between the two metrics. He was aware of Rosen's result that the flat spacetime interpretation implies a varying speed of light (using unrenormalized instruments), but seems also to have failed to entertain the possibility that the gravitational field might make light travel *faster* than in special relativity.

The neglect of the null cone issue continued well into the 1950s in the important works of S. N. Gupta [16–19] and R. H. Kraichnan [20, 21]. At this stage the derivation of the exact nonlinearities of general relativity, which Rosen had desired, was achieved. Concerning the special-relativistic nature of the theory, both authors seem to have regarded the Lorentz covariance of the theory as sufficient for special relativity. If the theory's gauge invariance and the unobservability of the flat metric are mentioned, the idea that the observable effective curved metric might well *conflict* with the flat metric is not. This is an important distinction that will also be overlooked repeatedly by later authors. One could imagine that the flat metric might fail to appear in the equations of motion, but still have its null cone serve as a bound on the curved metric's null cone, so this distinction is not trivial.

F. J. Belinfante, interested in the work of Papapetrou and Gupta and in particular in the solidifying covariant perturbation approach to quantizing gravity, contemplated the use of a flat metric in "Einstein's curved universe", which evidently meant the geometrical theory of gravity [22]. Working in the context of the static Schwarzschild solution (in which it is difficult to get the relation between the two null cones wrong, at least outside the Schwarzschild radius, unless one tries to do so), Belinfante only had occasion to consider the null cone relationship incidentally. But the fact that r becomes g-temporal and t becomes g-spatial for very small radii, juxtaposed with the *a priori* fixed character of these quantities with respect to the flat metric, does give him occasion for thought. Belinfante gives indications (including in the paper's title) that he does not believe deeply in the flat spacetime approach, so perhaps the null cone issue would not have interested him. While he is prepared to suggest that the "Swiss-cheese"-like behavior of the Schwarzschild solution in the bimetric context might help eliminate field theory's divergences, it is clear from review papers on quantum gravity [25, 26] that the flat metric is just a tool–perhaps a useful one, but more likely not–for Belinfante. It is thus not too surprising that the null cone issue is ignored. The "[r]eal problem" is not to be found in "[t]heories, usually in flat space, which seek to be approximations to Einstein's theory, or a perturbation-theoretical treatment of Einstein's theory", but in "[q]uantization of Einstein's theory itself." (pp. 198, 192) [25]. For Belinfante, spacetime might have a Swiss cheese structure, contain worm holes, or have a closed spatial topology [26]. Some of Belinfante's work with Swihart on linear gravity also neglects to discuss the null cone issue [23, 24].

5.2 The Years 1959-1979: the Problem Dismissed or Postponed

The null cone consistency issue is perhaps first discussed in print by W. Thirring in 1959 [28–30], but then dismissed with a resolution that does not permit a true special relativistic interpretation. Thirring clearly recognizes the apparent conflict between the two null cones [29], writing, "Another feature of the equations of motion ... we want to point out is that the velocity $|d\mathbf{x}/dt|$ is not required to be < 1 ... Thus [assuming the curved metric to be diagonal] there is a limiting velocity c but it is space dependent and may exceed unity." (pp. 100, 101) A bit later, he writes "Since c is also gauge dependent and will exceed unity in some gauge systems [the matter equation of motion] even admits an apparently acausal behavior." (p. 101) However, Thirring thinks that this acausal behavior is *only* apparent, for he is satisfied with the fact that the "renormalized" velocity (measured physically using real clocks, which are distorted by the gravitational field) is not greater than unity: "However, we shall see shortly that c also corresponds to the velocity of light and that it becomes unity when measured with real measuring rods and clocks since they all are affected by the [gravitational] field." (p. 101) Evidently the unobservable nature of the intervals governed by $\eta_{\mu\nu}$ satisfies Thirring that the apparently acausal behavior is not a problem: "The real metric [interval corresponding to $g_{\mu\nu}$] is gauge invariant whereas [the interval corresponding to $\eta_{\mu\nu}$] is not and therefore has no physical significance. Space-time measured with real objects will show a Riemannian structure whereas there are no measuring rods which could measure the original pseudoeuclidean space." (p. 103) Thirring's argument is doubtful because the same distinction that was neglected by Gupta and Kraichnan is also neglected here: the non-measurability of the flat metric does not entail that it lacks physical significance. Generally one considers causality to be an important physical concept. At the risk of stating the obvious, we recall that in special relativity, the relevant speed for causality is not the speed at which electromagnetic radiation actually propagates, but the value of the universal velocity constant (ordinary called "the speed of light" and written as c, but to do so here would invite confusion) which appears in Lorentz transformations, that is essential. As is well-known, to permit propagation faster than that speed in one frame is to admit backward causation in another frame, and generally one rejects backward causation. Given the violation of the flat spacetime null cone, it is not clear what is supposed to be the meaning of this field theoretic approach. Yet, according to Thirring, the field theoretic approach gives "a theory following the pattern of well understood field theories, in particular electrodynamics." (p. 116) Thus, Thirring's list of advantages and disadvantages of the field and geometric approaches to gravity (pp. 116, 117) is notably incomplete, because the obvious notion of causality for the field approach has been discarded. Thirring comes very close to noticing the problem of null cone inconsistency, but then stops short, apparently due to a prejudice against unobservable entities.

One might hope that Thirring's almost-recognition of the problem would have inspired his successors to recognize the seriousness of the problem. That, however, did not occur. In particular, although L. Halpern made a rather minute study of Thirring's paper [40], the light cone issue receives only a single sentence (p. 388), one sufficiently noncommittal that no discomfort with Thirring's purported resolution of the causality issue is obvious. Halpern was not an advocate of the flat spacetime approach to gravitation [39], so it is the more remarkable that he overlooked a potentially serious difficulty. R. Sexl also was aware of the Thirring's work and even presented it at a conference [33], yet he also accepted Thirring's ostensible resolution of the null cone conflict [33, 34].

The covariant perturbation program for quantizing general relativity yielded a large number of works based on expanding the curved metric into a background part and a dynamical part. Thus, one might expect the question of the relation of the two metrics to be considered in some way. Commonly the background metric was flat, leading to equations at least formally special relativistic.

A notable exception is the work of B. S. DeWitt, who made great use of non-flat background metrics and found various benefits in doing so [119]. While DeWitt could make use of a background metric, to him it was always at most a tool, not a deep part of nature. In an article entitled "The Quantization of Geometry," he wrote:

> The problem of [quantizing the gravitational field] may be approached from either of two viewpoints, loosely described as the "flat space-time approach" and "the geometrical approach." In the flat spacetime approach, which has been investigated by several authors ... the gravitational field is regarded as just one of several known physical fields, describable within the Lorentz-invariant framework of a flat space-time. Its couplings with other fields ... lead to a contraction or elongation of "rigid" rods and a retardation or advancement of "standard" clocks .... Both the geometrical and flat space-time points of view have the same *real* physical content. However, it has been argued that the flat space-time approach provides more immediate access to the concepts of conventional quantum field theory and allows the techniques of the latter to be directly applied to gravitation. While there is merit in this argument, too strong an insistence upon it would constitute a failure to have learned the lessons which special relativity itself has already taught. Just as it is now universally recognized as inconvenient (although *possible*) to regard the Lorentz-Fitzgerald contraction from relativistic modifications in the force law between atoms, so it will almost certainly prove inconvenient at some stage to approach space-time geometry, even in the quantum domain, in terms of fluctuations of standard intervals which are the same for all physical devices and hence unobservable. [118] (pp. 267,268).

Concerning the question of a well-defined causal structure, which his approach appeared to lack, he suggested, "Critics of the program to quantize gravity frequency [*sic*] ask, 'What can this mean?' A good answer to this question does not yet exist. However, there are some indications where the answer may lie." [119]. DeWitt's vision for the program, which he was prepared to call "covariant quantum geometrodynamics" in a volume honoring J. A. Wheeler (the title itself suggesting sympathy for a geometrical view of gravitation, much as "The Quantization of Geometry" did), included that it "should be able to handle any topology which may be imposed on 3-space"[120] (p. 437). Of the covariant perturbation formalism, he wrote that the "most serious present defect of the covariant formalism is its foundation in scattering theory, with spacetime being assumed asymptotically flat. The method of the background field, which we have introduced, indicates a way in which this defect may be removed" [120] (p. 437). It seems very likely that the null cone issue, as we have formulated it, would not be important to DeWitt, given that the background metric was merely a tool for investigating a truly geometrical theory.

Other authors, especially in the particle physics tradition, seem at least somewhat more content with a flat background metric. In his lectures on gravitation, R. Feynman shows himself ambivalent about the interpretation of gravitation. After deriving Einstein's equations from a flat spacetime field theory, he concluded from the unobservability of the flat metric that the latter was not essential. Using an analogy with curiously intelligent insects walking on a tiled floor, he says that "[t]here is no need to think of processes as occurring in a space which is truly Euclidean, since there is nothing physical which can ever be measured in this fictional space. The tiles represent simply a labelling of coordinates, and any other labelling would have done just as well" [43] (p. 101). Concerning the "assumption that space is truly flat," he concludes that "[i]t may be convenient in order to write a theory in the beginning to assume that measurements are made in a space that is in principle Galilean, but after we get through predicting real effects, we see that the Galilean space has no significance" (p. 112), but serves only as a "bookkeeping device" (p.

113). Concerning the "relations between different approaches to gravity theory,"

> [i]t is one of the peculiar aspects of the theory of gravitation, that it has both a field interpretation and a geometrical interpretation ... these are truly two aspects of the same theory ... the fact is that a spin-two field has this geometrical interpretation; this is not something readily explainable–it is just marvelous. The geometric interpretation is not really necessary or essential to physics. It might be that the whole coincidence might be understood as representing some kind of gauge invariance. It might be that the relationship between these two points of view about gravity might be transparent after we discuss a third point of view .... (p. 113)

Feynman seems to feel free to switch between the two views as he sees fit. Questions about nontrivial topologies or the desire to have a transparent notion of causality seem not to have occupied him. Had they, he might have hesitated in proclaiming them to be "the same theory", given the competition between causality and gauge invariance. Later, in developing the covariant perturbation theory, Feynman did not address these issues, but wrote as if no conceptual difficulties existed. He wrote: "The questions about making a 'quantum theory of geometry' or other conceptual questions are all evaded by considering the gravitational field as just a spin-2 field nonlinearly coupled to matter and itself (one way, for example, is expanding $g_{\mu\nu} = \delta_{\mu\nu} + h_{\mu\nu}$ and considering $h_{\mu\nu}$ as the field variable) and attempting to quantize this by following the prescription of quantum field theory, as one expects to do with any other field. The central difficulty springs from the fact that the Lagrangian is invariant under a gauge group," but this issue, he finds, can be resolved by adding a gauge-fixing term, the result being "completely satisfactory" at the level of tree diagrams (which correspond to the classical theory) [45]; see also ([44]). If the main difficulty is gauge invariance–which in fact competes with special relativistic causality, as Thirring nearly realized–and if the gauge-fixing terms lead to a "completely satisfactory" result without regard to the light cone relationship, then, unless we are to charge Feynman with oversight, clearly the flat metric is merely a useful tool for him. However, the claim to have avoided all conceptual questions cannot be sustained, because the light cone issue is just such a question, and the meaning of parts of Lorentz-covariant field theory remains obscure if the problem is ignored. Huggins, a student of Feynman, also neglects to consider the null cones issue [42].

S. Mandelstam presented a critique of the flat-space covariant perturbation program as Gupta had developed it [37]. Gupta had imposed the DeDonder coordinate (gauge) condition. Let us see how close Mandelstam comes to identifying the null cones issue. He writes: "Quantization in flat space can only be regarded as a provisional solution of the problem for several reasons," such as its approximate (at least at that stage of development) character, the use of an indefinite metric, and the presence of unphysical states. "But the main objection to this method of quantization lies surely in the physical sacrifices it makes by going to flat space. The variable specifying the coordinates are numbers without physical significance which can be chosen in an infinite variety of ways. On the other hand, distances in space-time, which are physically significant entities, are related to the coordinates in a manner which has not been elucidated when the metric is quantized." However, perhaps these objections can be met: "It may be possible to add to the theory a prescription for interpreting its results physically. If it could then be shown that the predictions of the theory were independent of the coordinate conditions used, and that they tended to the predictions of the unquantized theory in the classical limit, we would have a satisfactory theory. Some progress has actually been made in this direction by Thirring", which "indicates the connection of the Gupta variables to the metric," though "the basic difficulties of the 'flat space' approach remain." Clearly one of Mandelstam's worries is the question of gauge invariance in a procedure that makes use of coordinate conditions. It is difficult to tease out a clear statement of worry about rival null cones from these remarks, though the issue might have been intended among the "the basic difficulties of the 'flat space' approach" that remain.

A moment of considerable clarity occurred in 1962 with the appearance of a paper by J. R. Klauder [41], whose abstract opens with the statement, "[i]n any quantum theory, in which the metric tensor of Einstein's gravitational theory is also quantized, it becomes meaningless to ask for an initial space-like surface on which to specify the conventional field commutators." Klauder elaborates:

> In so far as [certain] formalisms [for quantizing gravity] are transcriptions of techniques successful in a flat Lorentz space-time, they ignore a unique problem peculiar to general relativity. Conventional field theories deal, in particular, with commutation rules, which, when employed for the fields separated by a space-like interval, have an especially simple form. Whether two nearby points are or are not space-like is a *metric* question that can be asked (and in principle answered) not only in flat space but also in any space with a preassigned curved metric as well. However as soon as the space-time metric $g_{\mu\nu}(x)$ becomes a dynamical variable–as in Einstein's theory–then an initial space-like surface on which to specify commutators of any two fields becomes a meaningless concept.

Klauder's approach to handling this problem was to propose an alternative formalism in which fields can fail to commute at most only at the same *event.* Unfortunately, Klauder's acute awareness of the null cone issue did not spread too widely immediately.

S. Weinberg did considerable work on gravitation considered as a Lorentz-invariant theory [48–56]. Concerning the geometric interpretation of general relativity, Weinberg could write that "the geometric interpretation of the theory of gravitation has dwindled to a mere analogy, which lingers in our language ... but is not otherwise very useful. The important thing is to be able to make predictions about images on photographic plates, frequencies of spectral lines, and so on, and it simply doesn't matter whether we ascribe these predictions to the physical effect of gravitational fields or to a curvature of space and time." [55] (p. 147) This ambivalence about the meaning of the theory perhaps helps to explain why the null cone consistency issue appears to be ignored in Weinberg's writings. However, the meaning of concepts used in Lorentz-invariant field theory in which Weinberg's work is rooted, or at least its relation to an underlying classical theory, does seem somewhat obscure if this issue is neglected. Somewhat more recently, R. Penrose reported that Weinberg was "no longer convinced that the anti-geometrical viewpoint is necessarily the most fruitful" [57], on account of some impossibility theorems [56]. In recent personal communication with us, Weinberg stated that he is no longer a strong advocate of any view on the subject, though it is quite interesting that the flat spacetime approach reproduces general relativity.

Based on the "spin limitation principle," which requires that only definite angular momenta be exchanged, V. I. Ogievetsky and I. V. Polubarinov have derived Einstein's equations and a family of massive relatives thereof in flat spacetime [60–62]. While this principle is quite attractive, it fails to pay any heed to whether the resulting theories yield propagation consistent with the causal structure of the flat metric. Given that some of their theories are massive and thus make the $\eta_{\mu\nu}$ *observable*, this shortcoming seems fairly serious, given the intent of deriving these theories from flat spacetime. While we can find no mention of the null cone issue in the work of Ogievetksy and Polubarinov, it would be interesting to see if the spin limitation principle could be generalized in such a way as to yield consistency of the null cones.

A large amount of work related to the field approach has been done by S. Deser, sometimes with collaborators such as D. G. Boulware, R. Nepomechie, A.

Waldron or others. In the course of papers which derived general relativity via self-interaction in flat spacetime [65] or curved [69], or general relativity from quantum gravity [66, 67], or supergravity from self-interaction [68], or which study bimetric theories for a festschrift for N. Rosen [72], we can find no mention of the issue of the null cone consistency issue. In particular, Deser finds the main issues for bimetric theories to be essentially the same problems that he and Boulware found in massive variants of general relativity [234], *viz.*, empirically falsified light bending properties, negative energy disasters, or both [72]. Unlike some authors who have a strong preference, Deser (after a rather pro-geometrical paper early on [266]) seems to admire both the geometric and field formulations: "The beautiful geometrical significance of general relativity is complemented by its alternate formulation as the unique consistent self-coupled theory arising from flat-space free gravitons, without appeal to general covariance." [69] However, depending on how one reads the flat spacetime approach, one might obtain some different features, as we will observe below, so one might prefer to see the meaning of the field formalism addressed.

Deser and R. Nepomechie have studied a somewhat related issue related to the anomalous propagation of gauge fields in some conformally flat spacetimes, compared to a flat background [70, 71] with the same null cone structure. In particular, backscattering off the geometry causes the propagation to lie not merely on the null cone, but inside it. However, "while our results are surprising, they do not imply any consistency problems" [70], as they would if the propagation were *outside* the null cone. Another related issue is work on massive spin $\frac{3}{2}$ electrodynamics [73]. We thank Prof. Deser for drawing our attention to these papers. The spin $\frac{3}{2}$ propagation issue was identified some time ago by G. Velo and D. Zwanziger as a problem. They found that the "main lesson to be drawn from our analysis is that special relativity is not automatically satisfied by writing equations that transform covariantly. In addition, the solutions must not propagate faster than light" [74], a lesson that needed, as it happens, application in altered form to gravitation (spin 2) as well.

During the 1970s, the relation between the two null cones continued to be neglected in practice in the context of the covariant perturbation approach to quantizing general relativity. However, quantum gravity review talks drew attention to this problem from time to time.

This service was performed with special clarity in a 1973 review by A. Ashtekar and R. Geroch [125]. They find that much of the difficulty in quantizing the theory arises from the fact that "the distinction between the arena and the phenomenon, characteristic of other physical theories, is simply not available in general relativity: the metric plays both roles." [125] (p. 1214) In discussing field theoretic approaches, they write that

> [i]t is normally the case in quantum field theory ... that two distinct fields come into play–a kinematical background field (the metric of Minkowski space) and a dynamical field .... One can certainly regard general relativity as a field theory, but in this case there is only a single field, the metric g_{ab} of spacetime, which must play both these roles. But the application of the techniques of quantum field theory apparently requires a non-dynamical background field. In quantum electrodynamics, for example, the causality of the Feynman propagators and the asymptotic states, in terms of which the S-matrix is defined, refer directly to the metric of Minkowski space. Thus, one does not expect to be able to carry over directly to general relativity, regarded as a classical field theory, the procedure which led for example from classical Maxwell theory to quantum electrodynamics. In order to apply the techniques of quantum field theory one must, apparently, either modify these techniques or reformulate the interpretation of general relativity as a field theory. (p. 1229)

This latter suggestion of reinterpreting general relativity does not strike us as being unimaginable or even excessively difficult, especially given that Kraichnan had presented a simple and clean derivation already in 1955 [20]. However, Ashtekar and Geroch do present some objections to this general line of attack. "It turns out, however, that this perturbation approach to obtaining a quantum theory of the gravitational field suffers from a number of difficulties. There exist, [126] for example, four-dimensional manifolds M on which there are metrics g_{ab} of Lorentz signature, but on which there are no flat metrics." (p. 1232) However, it is not clear why such examples must be regarded as physically admissible. If it could be shown that some exact solutions of obvious physical utility admit no flat metric, then the argument would be persuasive, but that argument seems not to have been made. Thus, it seems that this argument against the perturbation approach will be highly persuasive only if one is *already* committed to a geometrical view of Einstein's equations at the classical level. But one would exaggerate only slightly to say that that is the point at issue. The idea of *requiring* that the curved null cone be consistent with the flat one seems not to have been entertained, but Ashtekar and Geroch have shown powerfully, if reluctantly, why such an approach merits consideration.

The null cone consistency issue was also mentioned several times by C. J. Isham at the first Oxford quantum gravity symposium [285]. We quote from pp. 20, 21:

> One natural approach perhaps is to separate out the Minkowski metric $\eta_{\mu\nu}$ and write $g_{\mu\nu}(x) = \eta_{\mu\nu} + h_{\mu\nu}(x)$ where $h_{\mu\nu}(x)$ describes the deviation of the geometry from flatness ... [which approach has some advantages.] However, there are a number of objections to this point of view. For example: (i) The actual background manifold may not be remotely Minkowskian in either its topological or metrical properties, in which case the separation [above] ... is completely inappropriate. (ii) Even if

> [the equation above] is justified (from the point of view of i)) [*sic*] the procedure is still dubious because the lightcone structure of the physical spacetime is different from that of Minkowski space. For example, if the field $\hat{\phi}$ has some sort of microcausality property with respect to the metric $g_{\mu\nu}$ then this is <u>not</u> equivalent to microcausality with respect to the fictitious Minkowski background.

Once again a prior commitment to a geometrical view of general relativity is manifest. Isham seems not to entertain the idea of *requiring* that the spacetime be compatible with the flat background, but at least the consistency issue is clearly stated. At the second Oxford symposium, Isham observed that "one of the ambitions of the Riemannian programme is to free quantum gravity from perturbation theory based on the expansion $g_{\mu\nu} = g^{c}_{\mu\nu} + \sqrt{G}h_{\mu\nu}$. Expansions of this type are known to be bad in classical general relativity and they clearly misrepresent the global topological and lightcone structures of the pair $(M, g_{\mu\nu})$" [286] (p. 14). Isham has continued to mention this issue in more recent talks in the context of the problem of causality and time [201, 292]. The problem of time shows up in the light cone issue for the covariant perturbation approach to quantum gravity, but related difficulties show up elsewhere [292]. Still more recently, Isham has expressed the issue as follows:

> *The problem of time* The background metric η provides a fixed causal structure with the associated family of Lorentzian inertial frames. Thus, at this level, there is no problem of time. The causal structure also allows a notion of microcausality, thereby permitting a conventional type of relativistic quantum field theory to be applied to the field $h_{\alpha\beta}$.
>
> However, many people object strongly to an expansion [of the curved metric into a flat one plus a dynamical part] since it is unclear how this background causal structure is to be related to the physical one; or, indeed, what the latter really means . . . it is not clear what happens to the

> microcausal commutativity conditions in such circumstances; or, indeed, what is meant in general by 'causality' and 'time' in a system whose light cones are themselves the subject of quantum fluctuations.[201] (p. 58)

Another significant mention of the null cone consistency issue in a quantum gravity review talk comes from P. van Nieuwenhuizen at the first Marcel Grossmann meeting. After showing keen awareness of the problem, van Nieuwenhuizen shelves it. He writes:

> According to the particle physics approach, gravitons are treated on exactly the same basis as other particles such as photons and electrons. In particular, particles (including gravitons) are always in flat Minkowski space and move <u>as if</u> they followed their geodesics in curved spacetime because of the dynamics of multiple graviton exchange. This particle physics approach is entirely equivalent to the usual geometric approach. Pure relativists often become somewhat uneasy at this point because of the following two aspects entirely peculiar to gravitation:
>
> - In canonical quantization one must decide before quantization which points are spacelike separated and which are timelike separated, in order to define the basic commutation relations. However, it is only after quantization that the fully quantized metric field can tell us this spacetime structure. It follows that the concept of space-like or time-like separation has to be preserved under quantization, and it is not clear whether this is the case. (One might wonder whether the causal structure of spacetime need be the same in covariant quantization as in canonical quantization.)
> - Suppose one wanted to quantize the fluctuations (for example of a scalar field, or even of the gravitational field itself) about a

given curved classical background instead of about flat Minkowski spacetime. In order to write the field operators corresponding to these fluctuations in second-quantized form, one needs positive and negative frequency (annihilation and creation) solutions. In non-stationary spacetimes it is not clear whether one can define such solutions. (It may help to think of non-stationary space-time as giving rise to a time-dependent Hamiltonian.)

> The strategy of particle physicists has been to ignore these two problems for the time being, in the hope that they will ultimately be resolved in the final theory. Consequently we will not discuss them any further.[76, 77]

While quantization is not our immediate concern, a similar worry to the first of these two exists at the classical level if one wishes to take the flat metric seriously: there is no reason to expect that the dynamics will yield automatically a physical causal structure consistent with the *a priori* special-relativistic one.

5.3 The Years 1979-2001: the Problem Increasingly Attended and the Development of Three Views

More recently, the question of null cone consistency has come to be recognized as interesting somewhat more often. While a fair number continue to neglect the issue, those who have addressed it can be found to have one of three attitudes toward the flat metric: that it is a useful fiction, that it is a useless fiction, or that it is the truth. These views will be considered in turn. First we note some recent signs of the growing awareness of the problem.

In the 1984, the subject made its way into a standard text [152]. R. Wald writes: "The breakup of the metric into a background metric which is treated classically and a dynamical field γ_{ab}, which is quantized, is unnatural from the viewpoint of classical general relativity. Furthermore, the perturbation theory one obtains from this approach will, in each order, satisfy causality conditions with respect to the background metric η_{ab} rather than the true metric g_{ab}. Although the summed series (if it were to converge) still could satisfy appropriate causality conditions, the covariant perturbation approach would provide a very awkward way of displaying the role of the spacetime metric in causal structure." [152] (p. 384). Once again, a prior commitment to a geometrical understanding of classical gravity is evident. Some of Wald's negative attitude toward the "breakup" of g_{ab} results from assuming that the curved metric is fundamental, not derived. But given how easy it is to *derive* Einstein's equations from a flat spacetime theory [20], as we even showed above, why should one not regard the *curved* metric as derived? Be that as it may, one is pleased that the light cone issue is emerging from the neglect that it once suffered. It is intriguing that Wald suggests that the whole series might be g_{ab}-causal even though each term is η_{ab}-causal. An easy way for such to occur would be for the curved metric's null cone in fact to be confined on or within the flat one's. If that is the case, then it seems that Wald is almost suggesting (albeit reluctantly) what we will do below.

The recent contemplation of "naive quantum gravity" by S. Weinstein also has called attention to the lack of a fixed causal structure in quantum gravity [287]. If one is interested in full quantum gravity, as opposed to semiclassical work, then "we would expect that the metric itself is subject to quantum fluctuations ... But if the metric is [subject to quantum fluctuations], then it is by no means clear that it will be meaningful to talk about whether x and y are spacelike separated, unless the metric fluctuations somehow leave the causal (i.e. conformal) structure alone." (pp. 96-7) There appear to be two things that this last suggestion might mean. First, it might mean that the metric is conformally flat, so that the causal structure is just that of flat spacetime, while gravity is described by a scalar field. However, it is well-known that scalar gravity is empirically falsified by the classical tests of general relativity [20]. Second, it might mean that, although the full metric is allowed to vary, its variations are *bounded* so that the null cone of the nondynamical (and presumably flat) metric is respected. That is what we propose here. Weinstein does indeed consider "whether it is at all *possible* to construe gravitation as a universal interaction that nonetheless propagates in flat, Minkowski spacetime." (p. 91) He concludes that

> the short answer is, 'No,' for three reasons. First, the 'invisibility' of the flat spacetime means that there is no privileged way to decompose a given curved spacetime into a flat background and a curved perturbation about that background. Though this non-uniqueness is not particularly problematical for the classical theory, it is quite problematical for the quantum theory, because different ways of decomposing the geometry (and thus retrieving a flat background geometry) yield different quantum theories. Second, not all topologies admit a flat metric, and therefore spacetimes formulated on such topologies do not admit a decomposition into flat metric and curved perturbation. Third, it is not clear *a priori* that, in seeking to make a decomposition into background and perturbations about the background, the background should be *flat*. For example, why not use a background of constant curvature? (p. 92)

However, these arguments seem less than compelling. Concerning the first argument, Weinstein provides neither argument nor citation. It appears to be a claim that a suitably gauge-invariant theory cannot be constructed. Supposing that this claim is true, which is not obvious, it might be cause for "gauge-fixing" the theory at a fundamental level, a proposal which has in fact already been endorsed at the classical level by N. Rosen [6], A. Papapetrou [11], A. A. Logunov and collaborators such as A. A. Vlasov (for example, [105]), and H. Nikolić [108], or perhaps for adding mass term to the theory, if the negative energy and causality worries discussed elsewhere can be handled. Concerning the second objection, which resembles that of Ashtekar and Geroch, the advocate of flat spacetime will ask "why are nontrivial topologies necessary?" There are no *facts* or even good arguments that require them at present. In the absence of such, the insistence that nontrivial topologies are theoretically necessary is close to question-begging. Concerning the last objection, it seems clear that a flat background is the default choice because it is simpler than any other choice. While any other choice requires some argument for making that choice instead of the others and strongly suggests the question "why does spacetime have *this* geometry?", flat spacetime does not. Weinstein's specific alternative suggestion of a constant curvature spacetime, for example, suggests the question "why does the curvature take *this* value, as opposed to some other value?" We can agree with Weinstein that in "allowing metric fluctuations to affect causal structure, one is clearly at some remove from ordinary field-theoretic quantization schemes." (p. 97) But it seems unclear, *pace* Weinstein, that there is any need to renounce the use of a flat background causal structure.

We now discuss three major attitudes toward a flat metric that one finds.

5.4 Field Formulation: the Flat Metric as a Useful Fiction

Some authors have explicitly stated that the flat metric is merely an auxiliary object, formally useful but not tied to the causal structure of the theory [83–85, 102, 103]. The reasons given include the gauge-variance of the relationship between the null cones and the unobservability of the flat metric. The fact that the flat metric's null cone is sometimes violated using otherwise-convenient gauges appears to be another reason: it does not appear possible to fix the gauge to be, say, tensorial

DeDonder and have the null cone relationship be automatically satisfactory. L. P. Grishchuk has written that “the mutual disposition of the light cones of the $g_{\mu\nu}$ and $\eta_{\mu\nu}$ can be of interest only in the case when the attempt is made to interpret the metric relations of the world as observable,” which efforts are of course bound to fail [85]. Unfortunately, Grishchuk has overlooked the same distinction that Gupta, Kraichnan, Thirring, Feynman, and probably many others missed, and has failed to recognize that, if the null cones can be made consistent, then a conceptual difficulty posed by quantization would be eliminated. A. N. Petrov describes the same view (though we have taken the liberty of spelling out with words the abbreviations used):

> However, the background in the field formulation of general relativity is not observed. The movement of test particles and light rays is not connected with the geometry of the background spacetime. The light velocity in the background spacetime can approach an infinite value. In contrast, in the geometrical formulation of general relativity the test particles and the light rays define the geodesics in real physical spacetime. Thus, the background spacetime in the field formulation of general relativity is an auxiliary and nonphysical (fictitious) concept which is necessary for the description of true physical fields [103] (pp. 452, 453).

We admit inability to understand how a fictitious entity could be “necessary for the description of true physical fields”. It would appear that if the object in question is necessary for the description of true physical fields, then it is real; but if it is not real, then it is not necessary for the description of true physical fields. But let us continue with Petrov:

> We stress that the field formulation of general relativity and the geometrical formulation of general relativity are two different formalisms for a description of the same physical reality and they lead to the same physical conclusions . . . there are no obstacles in treating any solution to

> general relativity (spacetime) in the framework of the field formulation of general relativity. However, it is clear that a manifold which supports a physical metric will not coincide in general with a "manifold" which supports an auxiliary metric. As a result, in the field configuration on the auxiliary nonphysical background, "singularities," "membranes," "absolute voids," and others can appear. This leads to cumbersome and confused interpretations and explanations. Thus, the whole spirit of general relativity itself requires the investigation of many problems with the help of the geometrical formulation technique. However, there exists problems [sic] for an investigation in which the field formulation technique is more convenient [103] (p. 453).

Thus, the field formulation is seen as a tool that sometimes is helpful, but sometimes not so convenient, and in any case not to be trusted in addressing deep issues. If the flat metric tensor is to be praised so faintly, one might wonder if it is worth using.

A similar attitude has been taken by D. E. Burlankov [86], who did some early work using a flat background metric as a convenient fiction [38]. Burlankov objects to the fundamental status of Minkowski spacetime because of the gauge-variance of the null cone relation, and also because the curved null cone *differs* from the flat null cone [86]. The former argument will be addressed in due time. The latter argument, in Burlankov's hands, is said to imply that only curved metrics conformally related to the flat background would be acceptable. But this objection is just unpersuasive. It is not worrisome if the gravitational field slows light down below the universal velocity constant, as long as gravity does not speed light up. Burlankov's position [86] is fairly similar to that of Zel'dovich and Grishchuk, but a few points deserve special notice. Burlankov is sympathetic to idea (asserted by Logunov *et al.*) that general relativity has difficulties, noting "the collapse problem, the singularity problem, strong gauge invariance, and the absence of a 'natural' energy-momentum complex" (p. 176). However, Burlankov finds that the "solution of the amazing problems in gravity does not lie" in the bimetric formalism (p. 177): Minkowski space cannot be taken as fundamental because of the null cone difficulties.

5.5 Geometrical Formulation: the Flat Metric as a Useless Fiction

Other authors have taken the view that the flat metric is a blemish on the pure geometric beauty of general relativity, and thus is to be avoided in general. Such a description would seem to fit R. Penrose [81], J. Bičák [279], and L. Shepley[1]. This negative attitude toward the flat metric seems to have motivated Penrose to note that the null cone issue really must be handled if the Lorentz-covariant approach (which he associates with Weinberg) is to be considered satisfactory. Penrose, recognizing the connection between scattering theory and the Lorentz-covariant perturbation approach to gravity, poses a dilemma for the latter. Using global techniques, he shows that either the curved null cone locally violates the flat one, or the scattering properties become inconvenient because the geodesics for the two metrics continue to diverge even far away from a localized source. He concludes that a "satisfactory" relationship between the two null cones cannot be found. Concerning the horns of Penrose's dilemma, we simply accept the second one. It is known that long-range fields have inconvenient scattering properties [284]. We find that the root of the divergence between the geodesics is merely the long-range $\frac{1}{r}$ character of the potential in the conformally invariant part of the curved metric. If the fall-off were a power law of the form $\frac{1}{r^{1+\epsilon}}$, $\epsilon > 0$, then no difficulty would arise. So this objection is basically a reflection of the fact that a long-range symmetric tensor potential exists. But why is that a fundamental problem?

[1]We thank Prof. Shepley for discussing this issue.

5.6 Special Relativistic Approach: the Flat Metric as the Truth

Besides the field-in-fictitious flat spacetime and geometrical approaches, there is another attitude that one might take toward the flat metric approach, *viz.*, that special relativity is correct in its usual strict sense (global Lorentz invariance, trivial spacetime topology, and no violation of η-causality), and thus that the gravitational field must be made to respect the flat metric's causal structure. This view is more conservative than the other views [43] (p. 101), and is sufficiently obvious and attractive an idea that one might expect it to have been explored thoroughly, probably decades ago, and either sorted out or refuted. But as a matter of fact, we do not find that to be the case. Demonstrating this surprising fact was one purpose of the substantial review of the history of the subject above. Some authors have claimed to have sorted it out, and some to have refuted it, but we disagree on both points, as will appear below.

For the sake of convenience, this approach needs a name. We will use the term "special relativistic approach" (SRA). We have resisted calling this approach a "formulation" to match Petrov's "field formulation" and "geometric formulation," because it will appear below that the SRA is in fact physically distinct, though in a rather subtle and recondite way, from the geometrical approach. Some have objected to regarding a theory based on the Einstein equations as something other than general relativity [83–85]. Others, such as L. Shepley, have insisted that the SRA is distinct from general relativity[2]. Perhaps the common usage of the term "general relativity" is simply too vague to provide a resolution to this difficulty. If nothing else were at stake, one would avoid pretentious claims of a new theory. But as will appear below, there might in fact be a physical difference, which it would be logically possible to test, between the two approaches. In particular, the phenomena

[2]We thank Prof. Shepley for discussing this issue.

of collapse to form black holes seem to be altered somewhat in the SRA, not least because the SRA implies global hyperbolicity.

Some terminology will be helpful. Let us agree to refer to lengths and times measured using ordinary rods and clocks, which respond to $g_{\mu\nu}$, as "physical" lengths and times, but lengths and times measured using ideal rods and clocks, which are not affected by gravity and thus respond to $\eta_{\mu\nu}$, "metaphysical" lengths and times. Let us also write ds^2 for the g-interval and $d\sigma^2$ for the η-interval. The terms "physical" and "metaphysical" are parallel to the more traditional terms "renormalized" and "unrenormalized," but we find our terms more descriptive of our viewpoint. Suppose that one wants to carry a wristwatch around in a gravitational field in order to know what time it is and interact with others. It is not obvious simply from the terms used that one wants a "renormalized" watch, but it is obvious that one wants a "physical" watch in order to live successfully in the physical world. On the other hand, suppose that one wants to know whether to take seriously the infinite character of $t = \infty$ in naive coordinates for the Schwarzschild solution in the SRA. It is not obvious that one would trust a "unrenormalized" clock, and indeed many authors have heaped insults on the readings of such ethereal clocks. But it is indeed obvious that a "metaphysical" clock is to be trusted about ∞ in preference to a merely "physical" clock. Our terminology is thus well-suited to a program of taking the taking the flat metric, including its causal structure, seriously, a goal perhaps not shared by some authors who have trafficked in "renormalized" and "unrenormalized" measurements. It should be clear that while both the field formulation and the special relativistic approach use a flat background metric and share many mathematical results, the SRA takes a realist attitude toward the flat metric, whereas the field formulation takes an antirealist attitude.

The SRA has the advantage of simple and fixed notion of causality at the classical level, because the flat null cone serves as a bound on the curved one. In this view, one becomes less dependent on the study of topology, global techniques, careful definitions of causality of various sorts, and the like, such as modern texts contain [127, 152]. The SRA is therefore simpler in an obvious way than the alternatives. There seems to be no difficulty in extending this nondynamical causal structure into the quantum regime, so there should be no problem in writing down equal-time commutation relations, *etc.* in the usual way, which question worried Isham above [292]. Thus, the worries expressed by Ashtekar and Geroch, Isham, van Nieuwenhuizen, Wald and Weinstein above are resolved.

While the special relativistic approach to Einstein's equations is *locally* and *classically* equivalent to the usual theory, as we saw earlier, there might be different *global* or *quantum* properties. For example, though flat spacetime with trivial topology is stable in general relativity [289], closed flat space is unstable [290], so it appears that the usual topology is more than just a simple and convenient choice for the SRA. Also, it will turn out below that some regions of spacetime in complete exact solutions of the geometrical theory simply do not exist in the SRA. Moreover, the SRA of general relativity has less gauge freedom than the geometrical and field formulations, because any gauge choice that leads to an improper null cone relationship must be prohibited. (The use of a new set of variables, in which only the null-cone respecting field configurations are possible, would be a way to prohibit them. We propose such a set below.) For that reason, some old and settled issues in the geometrical framework would have to be reconsidered in view of the different postulates. In particular, given that the gauge freedom is traditionally used to dispose of the 'singularity' at the Schwarzschild radius, and also that there is no possibility in the SRA of adding more spacetime 'past infinity', it is clear that the Schwarzschild radius will need careful consideration. Below we begin to consider that issue.

The attitude of regarding the flat spacetime as fundamental has been most visibly promoted by A. A. Logunov and colleagues [89–91] (to name a few).[3] To distinguish their view clearly from any geometrical notions, they have given the name "relativistic theory of gravitation" (RTG) to the work. The nature of the RTG has evolved slightly over the years. For some time it consisted in Rosen's tensorial $\Gamma\Gamma$ action for general relativity and his tensorial DeDonder condition [6] postulated as necessary, presumably with specification of trivial topology for spacetime. There is also attached a "causality principle" that requires that the curved null cone not violate the flat one [89, 281, 282]. This causality principle, which seems to have appeared following criticisms by Zel'dovich and Grishchuk [84, 85], is the feature most relevant to our purposes.[4]

Logunov *et al.*, being committed to the flat spacetime view, regard the question of compatible null cones as worthy of solution. Furthermore, they believe it to be solved already by their causality principle, which we shall call the Logunov Causality Principle (LCP). The LCP states that field configurations that make the curved metric's null cone open wider than the flat metric's are physically meaningless [89, 281, 282]. As they observe, satisfaction is not guaranteed (even with their gauge conditions, notes Grishchuk [85]), which means that the set of partial differential equations is not enough to define the theory. The LCP is therefore enforced "by hand." Some causality principle is indeed needed, but the LCP strikes us as somewhat arbitrary and *ad hoc.* One would desire three improvements. First, one would prefer that the causality principle be tied somehow to the Lagrangian density, not separately appended [85]. Second, one wants a guarantee that there exist

[3]This school has also produced an energetic critique of geometrical general relativity as lacking physical meaning, in the sense of lacking conservation laws and failing to make definite predictions. One need not accept this critique to be partial to flat spacetime.

[4]More recently, the RTG has sometimes featured acquired a mass term, which makes the action that of Freund, Maheshwari, and Schonberg's massive general relativity [260], from which the tensorial DeDonder condition follows automatically from the gravitational (and matter) field equations, much as the Lorentz condition follows in Proca's massive electromagnetism. As Logunov *et al.* note, the promise of gravitational energy localization and like benefits of the flat spacetime formalism are best realized only if the gauge invariance is broken. However, we are interested especially in the massless version, and it will appear below that massive versions faces major difficulties.

enough solutions obeying the principle to cover all physically relevant situations. Third, one would prefer a more convenient set of variables to describe the physics. We begin to address these matters below.

Concerning the first shortcoming, it might be suggested [91] that the Logunov causality principle is analogous to the energy conditions [152] that one typically imposes. However, this analogy strikes us as weak. The dissimilarity is in how the two conditions accept or reject solutions. The energy conditions are used to exclude or include whole classes of matter fields, so any configuration with one sort of matter field–perhaps a minimally coupled massless scalar field with the correct sign in the Lagrangian density– is permitted, whereas any configuration with another sort of matter field–perhaps the scalar field with the wrong sign– is prohibited as unphysical. This criterion expresses the idea that some sorts of matter are physically reasonable, but others are not. Furthermore, there is no worry that a permissible sort of matter could evolve into a forbidden sort in accord with the field equations. On the other hand, the LCP cannot give (or at least has not given) a similarly general explanation for why it rejects some solutions of the field equations. A more serious problem is that it cannot give any assurance that it permits a sufficiently large number of solutions to cover all physical situations that arise.

Let us focus our attention on the second difficulty, the possible shortage of solutions, which is potentially very serious. *A priori* there is no reason to believe that one can (partially) fix the gauge, and then still reject some solutions in the appropriate gauge as unphysical. *A posteriori* there seems to be good evidence that this worry is serious. We will employ a somewhat homely example because of its obvious physical relevance. Suppose that a young man named Nicholas has a drum set and a pair of sticks. If Nicholas is a skilled drummer, then the motion of his sticks will be quite under his control, but nevertheless the position of his sticks as a function of time will be rather wild and violent from a kinematical point of view. In particular, we can have great confidence that the motion of his sticks (and arms) will be such that their quadrupole moment will have a nonvanishing second time derivative, in general. We can also be confident that the traceless part of this second time derivative, contracted with itself, tends not to vanish. But that means that Nicholas emits gravitational radiation, for this is just the formula for the average power radiated in general relativity, under suitable assumptions such as a slowly varying source [152]. There will be anisotropic but roughly spherical waves of gravitational radiation diverging from Nicholas. Far away from him, these waves will look approximately like plane waves obeying linearized gravity with the tensorial Hilbert gauge condition (or approach such waves for high frequencies, in the massive case). The behavior of plane monochromatic single-polarization waves in linearized general relativity is well-known [304]. In this gauge, the two energy-carrying polarizations both consist of alternately shrinking one transverse direction and stretching the other, while the time (lapse) and spatial propagation directions are unaffected [304]. (Below we will analyze this linear solution using a generalized eigenvalue-eigenvector formalism.) But the shrinking of one spatial eigenvalue while leaving the time lapse unaltered implies a violation of the flat metric's null cone. In short, it appears that, if the exact behavior of the plane waves is anything like the linearized behavior, then monochromatic gravitational radiation satisfying the tensorial DeDonder condition generically violates the Logunov causality principle. While monochromatic radiation is a rather idealized case, and thus perhaps need not obey η-causality by itself for a satisfactory theory, it is not at all obvious that the superposition of monochromatic waves which all individually violate η-causality yields a sufficiently generous set of realistic waves that satisfy that condition. Thus, there is reason to worry that the Logunov causality principle cannot be implemented in worlds in which Nicholas plays drums.

Evidently, arranging for wave solutions to obey the causality principle is rather more difficult than addressing most of the solutions that Logunov and collaborators have addressed to date.[5] Regarding standard homogeneous and isotropic cosmological solutions, perhaps one could learn to accept a requirement that the scale factor just could not take certain values [280]. Concerning the Kerr-Newman solution [281], the exterior causes no problem, and the physical significance of the vacuum interior is open to question–especially given a mass term or the reduction of the gauge freedom, as the RTG requires. As far as Kasner solutions are concerned [282], the world looks little like a Kasner solution; maybe somehow it just *couldn't have* looked like a Kasner solution–it's hard to say. Perhaps one could accept these requirements. But no one will accept the idea that Nicholas is unable to play his drums, because he does it every day. It is therefore rather likely that the Logunov causality principle does not admit enough solutions to account for manifestly physically relevant situations, such as Nicholas's drumming. Thus, some way of enforcing null cone consistency without excluding necessary solutions of the field equations must be sought.

The lesson that we draw from the apparent shortcomings of the RTG in its present and past forms is not, *pace* some authors [83–85], that the flat metric must be considered merely a useful fiction. Rather, if the SRA is to be maintained, then a more fundamental approach to securing consistency between the null cones must be sought. One will want to use the gauge freedom of general relativity to secure null cone consistency. In that way, one can be confident that a sufficiently large number of solutions exist, because one member from each equivalence class of solutions will be included.

We pause to note that some related matters pertaining to the light cone in string theory have been considered by E. Martinec [123].

[5] Recently, I. P. Denisova has investigated the effective metric of a plane polarized electromagnetic wave, but this solution applies only to the massive version of the RTG. We discuss it below.

5.7 The Role of Positivism in the Neglect of the Null Cone Consistency Issue

It is interesting to compare the slowly increasing attention given this problem from 1939 to 2001 to the slowly decreasing influence of positivist and other radical empiricist philosophies, which laid great emphasis on observability and denigrated unobservable entities. Such philosophies were very influential on physicists in the first half of the 20th century, in no small part due to Mach, Einstein (at least in the era of special and general relativity), Bohr, and Heisenberg, and persist to some extent even today in physics, especially in the "modern" physics that they helped to found, relativity and quantum mechanics [187, 188]. This remains the case though strict positivism has fallen on hard times among philosophers [187]. One conspicuous example has been noted by W. L. Craig, a marginal note in Misner, Thorne, and Wheeler's standard text on gravitation: "Newton's absolute space is unobservable, nonexistent." [151] (p. 19) [187] (p. 120). Our aim is not to defend Newton's absolute space, but merely to note the facility with which this highly influential text feels able to deduce the nonexistence of an entity from its unobservability, a sign of positivist influence.

These issues carry over rather plainly into remarks made below in the history of the null cone consistency issue. As was shown above, it has often been concluded that if the flat metric is not observable, then it does not matter, or even does not really exist, being at most a convenient fiction. The role of positivist philosophy in diverting attention from the null cone consistency issue is in fact explicitly supported in the work of N. Straumann [88, 124]. Straumann, appealing to Einstein in part, cites a positivistic motivation for discounting the flat metric sometimes used to derive general relativity [124]. He writes about theories of gravity in flat space-time:

> In spite of these arguments [such as how the bending of light shows that

> the effective spacetime is not conformally flat] one may ask, as many have done, how far one gets with a theory of gravity in Minkowski space, along the lines of electrodynamics, admitting the nonobservability of a flat metric. Such attempts have shown that a consequent development finally allows the elimination of the flat metric leading to a description in terms of a "curved" metric which has a direct physical interpretation. The originally postulated Poincaré invariance turns out to be physically meaningless and plays no useful role.[6] We may summarize as follows:
>
> In the presence of gravitational fields, Minkowski space can no longer be physically realized. If one requires from the theory that the defining concepts have an empirically verifiable meaning, then it is more sensible to relate its assertions to the orbits of point masses or light rays, rather than to an unobservable Minkowski space. (p. 87)

Alas, Straumann has made the same error as others above in identifying physical significance with appearance in the field equations, because the flat metric could perhaps still be physically significant by providing an *a priori* fixed causal structure which *bounds* the dynamical causal structure. But our present concern is not so much whether his statement is correct, as whether it reflects positivist philosophy. In case the clause about an "empirically verifiable meaning" is not enough to show positivist influence, Straumann, though affirming some other thought currents as having some role, concludes, "[i]n the previous sections we have, for the most part, taken a positivistic attitude" [124] (p. 87).

As the status of positivism has plummeted in the last few decades among philosophers [187], it seems appropriate to reconsider its legacy in other fields. Thus, it is quite fitting that the null cone issue receive increased attention nowadays, though much work remains to be done.

[6] Straumann's reference here appears to be misnumbered as 21, for the obvious relevant reference is S. Deser's 1970 paper, which he lists as 23.

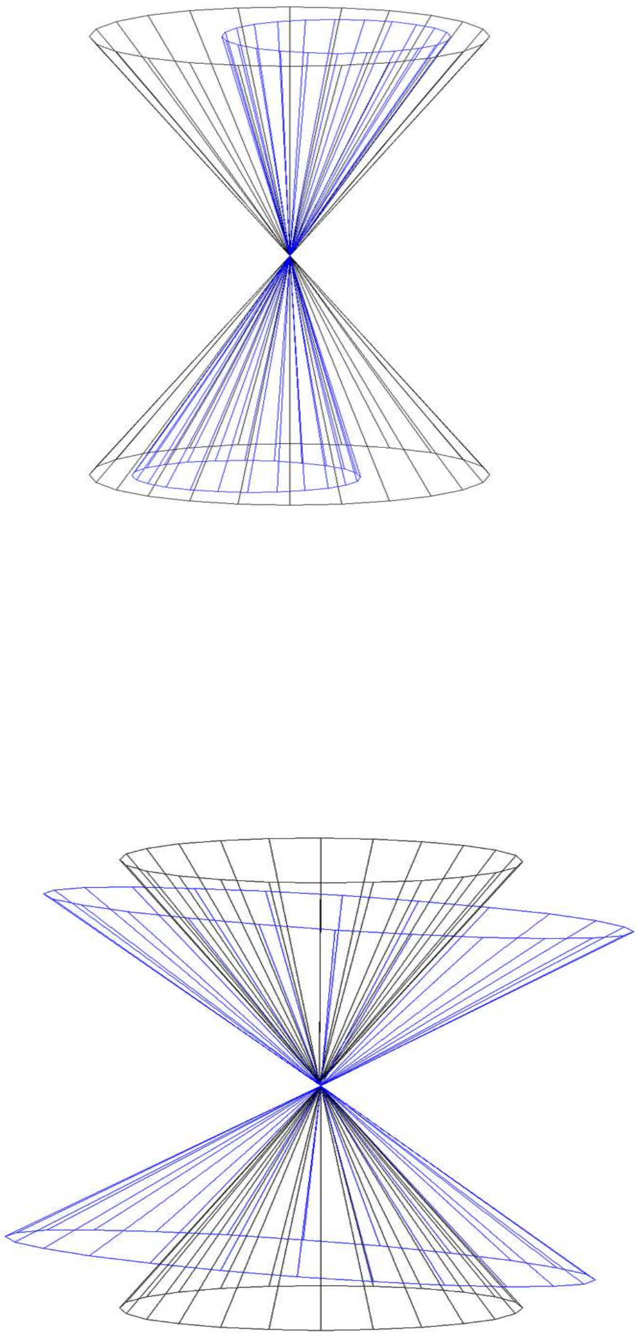

Figure 5.1: Examples of Consistent (above) and Inconsistent (below) Null Cone Relationships in 2 Dimensions. By Ryan Hayes using *Blender*

Chapter 6

Describing and Enforcing the Proper Null Cone Relationship

6.1 Consistent Null Cones by Suitable Gauge Restrictions?

As several authors above pointed out, the local relation between the two null cones is indeed gauge-dependent in general relativity [81, 85]. One might therefore hope to design a set of gauge-fixing conditions that yield the desired behavior, or at least to impose restrictions on the variables that exclude unsuitable gauge choices while permitting suitable ones. Rather than putting the conditions in arbitrarily by hand, one prefers to implement them in the action principle somehow.

It might be hoped that the ADM split of the metric [151, 152], which is quite useful in applications and in identifying the true degrees of freedom, would be a good language for discussing the null cone consistency issue. Let us see if that is the case. For convenience we choose Cartesian coordinates, so that $\eta_{\mu\nu} = diag(-1, 1, 1, 1)$. We therefore make an ADM split of Logunov's 4-dimensional analysis of the causality principle. In considering whether all the vectors V^{μ} lying on η's null cone are g-

timelike, g-null, or g-spacelike, it suffices to consider future-pointing vectors with unit time component; thus $V^{\mu} = (1, V^{i})$, where $V^{i}V^{i} = 1$ (the sum running from 1 to 3). The causality principle can be written $h_{ij}(\beta^{i} + V^{i})(\beta^{j} + V^{j}) - N^{2} \geq 0$ for all spatial unit vectors V^{i}. Here the spatial metric is h_{ij}, the lapse is N, and the shift is β^{i}. One could visualize this equation as the requirement that an ellipsoid (not centered at the origin if the shift β^{i} is nonzero) not protrude from the closed unit ball. Unfortunately, the "all" in "for all spatial unit vectors" is not too easy to handle, so we will in fact look for a better language than an ADM split for discussing null cone consistency.

If there to be any is hope for restricting the gauge freedom so as to ensure that the curved null cone stays consistent with the flat one, then there must be "enough" gauge freedom to transform any physically significant solution into a form that satisfies η-causality. Here we argue that gauges satisfying the causality principle likely *do* exist, so there is enough gauge freedom. Given a flat background metric and a Cartesian coordinate system for it, one can readily draw the flat and curved metrics' light cones on the tangent space at some event (apart from obvious difficulties with higher-dimensional pictures). One wants the curved cone to be located on or within the flat one. The flat cone has the usual ideal conical shape, whereas the curved one is distorted and tilted, in general. In a bimetric context, it is basically the case that the curved spatial metric controls the width of the light cone, while the shift vector determines its tilt from the vertical (future) direction and the lapse function determines its length. (Given that only the conformal part of the metric affects the null cone, the ADM description is a bit redundant.) For generally covariant theories such as general relativity, the spatial metric contains the physical degrees of freedom; the lapse and shift represent the gauge freedom, so they can be chosen arbitrarily, at least over some region. (For slightly bimetric theories, one has one fewer arbitrary function to choose, so the argument is considerably less plausible.) By analogy with conditions typically imposed in geometrical general relativity to avoid causality difficulties [152], one would prefer, if possible, that the curved light cone be strictly inside the flat light cone (*i.e.*, be η-timelike), not tangent to it, because tangency indicates that the field is on the verge of η-causality violation. Under quantization, one might expect fluctuations to push the borderline case into the unacceptable realm, so it seems best to provide a cushion to avoid the problem, if possible. This requirement we call "stable η-causality," by analogy to the usual condition of stable causality [152]. One might worry that this requirement would exclude all curved metrics conformally related to the flat one, and even the presumed "vacuum" $g_{\mu\nu} = \eta_{\mu\nu}$ itself. This worry is justified, but if one takes the message of gauge invariance seriously, then there is no fundamental basis for preferring $g_{\mu\nu} = \eta_{\mu\nu}$ over having the curved metric agree *up to a gauge transformation* with the flat metric. With this relaxed criterion, one can avoid the troubles with the folded surface discussed below.

Let the desired relation between the null cones hold at some initial moment. Also let the curved spatial metric and shift be such at some event in that moment that they tend to make the curved cone violate the flat one a bit later. By suitably reducing the lapse, one can lengthen the curved cone until it once again is safely inside the flat cone. By so choosing the lapse at all times and places, one should be able to satisfy the causality principle at every event, if no global difficulties arise. In a rough sense, one might use up $\frac{1}{4}$ of the gauge freedom of general relativity, while leaving the remainder.

Generally it is assumed that the reason that the gravitational Hilbert action is gauge-invariant is because such gauge invariance reflects a deep feature of the world. However, as we saw above, one can give a somewhat more humble explanation: it is known from the flat spacetime approach that eliminating the time-space components of the field is essential for positive energy properties in Lorentz-invariant theories

[4, 75], though in fact the time-time component need not be [80]. We might suggest that the gauging away of the time-time component is necessary rather to respect η-causality.

6.2 The Causality Principle and Loose Inequalities

As should be clear from the worries about conformally flat curved metrics, the desired relationship between the two null cones takes the form of some loose inequalities $a \leq b$. Such relations have been called "unilateral" [305–309] or "one-sided" [310–312], typical examples being nonpenetration conditions. Such constraints are rather more difficult to handle than the standard "bilateral" or "two-sided" constraint *equations* that most treatments of constraints in physics discuss. Loose inequalities are also more difficult to handle than strict inequalities $a < b$, such as the positivity conditions in canonical general relativity [313, 314, 332–334], which require that the "spatial" metric be spatial. One might eliminate the positivity conditions by a change of variables [314] that satisfies the inequalities identically, such as an exponential function $h = e^y$, as Klotz contemplates. While that is possible, it does not come for free: the "ground state" in which the curved metric equals the flat one is not permitted for any finite value of the argument. However, given the gauge freedom, this objection does not seem compelling, even if it complicates linearization. (Using an even function such as cosh to solve the inequality would seem to introduce even greater difficulties, so we prefer the exponential form.)

If one does leave causality constraints in the theory, rather than solving them as was suggested in the previous paragraph, then one must worry about impulsive constraint forces. That would be the case if one lets the curved metric evolve freely until it "hits" the flat null cone, at which point it might "bounce off," or perhaps become deformed like a face against a window, neither of which seems like an appealing prospect. To avoid such blemishes, which seem difficult to accept in an ostensibly fundamental classical field theory, one prefers to make the constraint "ineffective", so the constraint force vanishes on account of the constraint itself [319]. Better yet, let us change the configuration space of the theory so that the problem is avoided altogether.

6.3 New Variables and the Segré Classification of the Curved Metric with Respect to the Flat

Previous formulations of the causality principle, which have used the metric [89, 90, 281, 282] or ADM variables as above, have been sufficiently inconvenient to render progress difficult. This was our third complaint about Logunov's formulation of η-causality. One could achieve a slight savings by using the conformally invariant weight $-\frac{1}{2}$ densitized part of the metric $g_{\mu\nu}(-g)^{-\frac{1}{4}}$. Then nine numbers at each event are required (the determinant being -1), which is a bit better than the 10 of the full metric, but still too many.

One would like to diagonalize $g_{\mu\nu}$ and $\eta_{\mu\nu}$ simultaneously by solving the generalized eigenvalue problem

$$g_{\mu\nu}V^{\mu} = \Lambda\eta_{\mu\nu}V^{\mu}, \tag{6.1}$$

or perhaps the related problem using $g_{\mu\nu}(-g)^{-\frac{1}{4}}$. However, in general that is impossible, because there is not a complete set of eigenvectors, due to the Lorentzian nature of the metric [320–324] (and references in ([324]). There are four Segré types for a real symmetric rank 2 tensor with respect to a Lorentzian metric, the several types having different numbers and sorts of eigenvectors [320–323].

To our knowledge, the only previous work to consider a generalized eigenvector decomposition of a curved Lorentzian metric with respect to a flat one[1] was done

[1]There is also a literature on choosing coordinates to diagonalize a curved metric [326]. According to K. P. Tod, "there is very little change for Lorentzian metrics" compared to Riemannian ones

by I. Goldman [325], in the context of Rosen's bimetric theory of gravity, which does not use Einstein's field equations and is a quite distinct theory. Goldman's work was tied essentially to Rosen's theory, so it does not address our concerns much. The lack of gauge freedom in Rosen's theory also ensured that the curved null cone was not subject to adjustment, unlike the situation in general relativity with a flat metric. As it happens, in Rosen's theory, for static spherically symmetric geometries, the causality principle is always *violated*[2], so that theory does not qualify for a SRA. However, Einstein's theory does evidently have enough gauge freedom to make a special relativistic approach possible.

An eigenvalue-eigenvector decomposition for the *spatial* metric was briefly contemplated by Klotz and Goldberg [313, 314]. For space, as opposed to spacetime, one has a positive definite background (identity) matrix, so the usual theorems apply. But Klotz and Goldberg, who did not assume a flat metric to exist, found little use for the eigenvector decomposition because of the nontensorial nature of the 3×3 identity matrix. Such a decomposition, even given a flat metric tensor, is still somewhat complicated if the ADM shift is nonvanishing ($g_{0i} \neq 0$), as it usually is. Diagonalization has been quite useful in the study of spatially homogeneous cosmologies [318], but our interest is not in specific solutions only, but the general case. However, one expects that studying such models would be instructive.

Let us now proceed with the diagonalization project. Given that a complete set of generalized eigenvectors might fail to exist, it is necessary to consider how many eigenvectors do exist and under which conditions. This problem has been substantially addressed in a different context by G. S. Hall and collaborators [320–323], who were interested in classifying the stress-energy or Ricci tensors with respect to the (curved) metric in (geometrical) general relativity. Such problems have in fact

[326]. But there is a large change for the eigenvalue problem of interest to us in changing from a Riemannian to a Lorentzian background metric, so the connection between these problems must be somewhat loose.

[2]Prof. Goldman has kindly provided this information from his dissertation in Hebrew.

been studied over quite a long period of time [324] (and references therein), but we find the work of Hall *et al.* to be especially convenient for our purposes. There exist four cases, corresponding to the four possible Segré types (apart from degeneracies) for the classified tensor. The case $\{1,111\}$ has a complete set of eigenvectors (1 timelike, 3 spacelike with respect to η), and is thus the most convenient case. The case $\{211\}$ has two spacelike eigenvectors and one null eigenvector (with respect to η), whereas the $\{31\}$ case has one spacelike eigenvector and one null one. The last case, $\{z\,\bar{z}11\}$ has 2 (real) spacelike eigenvectors and 2 complex eigenvectors.

We now consider the conditions under which metrics of each of these Segré classes obey η-causality. To give a preview of our results, we state that the $\{1,111\}$ and $\{211\}$ cases sometimes do obey it, although the $\{211\}$ metrics appear to be dispensable. But no metric of type $\{31\}$ or $\{z\,\bar{z}11\}$ obeys the causality principle, so these types can be excluded from consideration for the SRA.

Hall *et al.* introduce a real null tetrad of vectors $L^\mu, N^\mu, X^\mu, Y^\mu$ with vanishing inner products, apart from the relations $\eta_{\mu\nu}L^\mu N^\nu = \eta_{\mu\nu}X^\mu X^\nu = \eta_{\mu\nu}Y^\mu Y^\nu = 1$, so L^μ and N^μ are null, while X^μ and Y^μ are spacelike. (The signature is $-+++$.) Expanding an arbitrary vector V^μ as $V^\mu = V^L L^\mu + V^N N^\mu + V^X X^\mu + V^Y Y^\mu$ and taking the η-inner product with each vector of the null tetrad reveals that $V^L = \eta_{\mu\nu}V^\mu N^\nu$, $V^N = \eta_{\mu\nu}V^\mu L^\nu$, $V^X = \eta_{\mu\nu}V^\mu X^\nu$, and $V^Y = \eta_{\mu\nu}V^\mu Y^\nu$. Thus, the Kronecker delta tensor can be written as $\delta^\mu_\nu = L^\mu N_\nu + L_\nu N^\mu + X^\mu X_\nu + Y^\mu Y_\nu$, indices being lowered here using $\eta_{\mu\nu}$. For some purposes it is also convenient to define the timelike vector $T^\mu = \frac{L^\mu - N^\mu}{\sqrt{2}}$ and the spacelike vector $Z^\mu = \frac{L^\mu + N^\mu}{\sqrt{2}}$.

We employ the results of Hall *et al.* [320–323], who find that the four possible Segré types (ignoring degeneracies) for a (real) symmetric rank 2 tensor in a four-dimensional spacetime with a Lorentzian metric can be written in the following ways, using a well-chosen null tetrad. The type $\{1,111\}$ can be written as

$$g_{\mu\nu} = 2\rho_0 L_{(\mu}N_{\nu)} + \rho_1(L_\mu L_\nu + N_\mu N_\nu) + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu, \tag{6.2}$$

or equivalently

$$g_{\mu\nu} = -(\rho_0 - \rho_1)T_\mu T_\nu + (\rho_0 + \rho_1)Z_\mu Z_\nu + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu. \tag{6.3}$$

As usual, the parentheses around indices mean that the symmetric part should be taken [152]. The type $\{211\}$ can be written as

$$g_{\mu\nu} = 2\rho_1 L_{(\mu}N_{\nu)} + \lambda L_\mu L_\nu + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu, \tag{6.4}$$

with $\lambda \neq 0$, the null eigenvector being L^μ. The type $\{31\}$ can be written as

$$g_{\mu\nu} = 2\rho_1 L_{(\mu}N_{\nu)} + 2L_{(\mu}X_{\nu)} + \rho_1 X_\mu X_\nu + \rho_2 Y_\mu Y_\nu, \tag{6.5}$$

the null eigenvector again being L^μ. The final type, $\{z\,\bar{z}11\}$, can be written as

$$g_{\mu\nu} = 2\rho_0 L_{(\mu}N_{\nu)} + \rho_1(L_\mu L_\nu - N_\mu N_\nu) + \rho_2 X_\mu X_\nu + \rho_3 Y_\mu Y_\nu, \tag{6.6}$$

with $\rho_1 \neq 0$. The requirements to be imposed upon the curved metric for the moment are the following: all η-null vectors must be g-null or g-spacelike, all η-spacelike eigenvectors must be g-spacelike, $g_{\mu\nu}$ must be Lorentzian (which amounts to having a negative determinant), and $g_{\mu\nu}$ must be connected to $\eta_{\mu\nu}$ by a succession of small changes which respect η-causality and the Lorentzian signature. It convenient to employ a slightly redundant form that admits all four types in order to treat them simultaneously. Thus, we write

$$g_{\mu\nu} = 2AL_{(\mu}N_{\nu)} + BL_\mu L_\nu + CN_\mu N_\nu + DX_\mu X_\nu + EY_\mu Y_\nu + 2FL_{(\mu}X_{\nu)}. \tag{6.7}$$

Using this form for $g_{\mu\nu}$, one readily finds the squared length of a vector V^μ to be

$$g_{\mu\nu}V^\mu V^\nu = 2AV^L V^N + B(V^N)^2 + C(V^L)^2 + D(V^X)^2 + E(V^Y)^2 + 2FV^X V^N \tag{6.8}$$

It is not clear *a priori* how to express sufficient conditions for the causality principle in a convenient way. Obviously it is sufficient that every g-timelike vector be η-timelike and every g-null vector be η-null or η-timelike. One could alternatively say that every η-spacelike vector must be g-spacelike and every η-null vector be g-null or g-spacelike. However, timelike and spacelike vectors are not very convenient to use because of the inequalities inherent in the word "every". But it will turn out that in four dimensions, the necessary conditions that we can readily impose are also sufficient.

6.4 Necessary Conditions for Respecting the Flat Metric's Null Cone

The causality principle requires that the η-null vectors L^{μ} and N^{μ} be g-null or g-spacelike, so $B \geq 0, C \geq 0$. These conditions already exclude the type $\{z\,\bar{z}11\}$, because the form above requires that B and C differ in sign. It must also be the case that the η-spacelike vectors X^{μ} and Y^{μ} are g-spacelike, so $D > 0$ and $E > 0$.

Not merely L^{μ} and N^{μ}, but all η-null vectors must be g-null or g-spacelike. This requirement quickly implies that $E \geq A$, and also requires that

$$B(V^N)^2 + 2FV^XV^N + (D - A)(V^X)^2 \geq 0. \tag{6.9}$$

Here there are two cases to consider: $F = 1$ for type $\{31\}$, and $F = 0$ for types $\{1,111\}$ and $\{211\}$. Let us consider $F = 1$. The $\{31\}$ has $B = 0$, so the equation reduces to $2FV^XV^N + (D - A)(V^X)^2 \geq 0$, which implies that either $V^X = 0$ or, failing that, $2FV^N + (D - A)V^X \geq 0$. Clearly one could also consider a null vector with the opposite value of V^X, yielding the inequality $2FV^N - (D - A)V^X \geq 0$. Adding these two inequalities gives $4V^N \geq 0$, which simply cannot be made to hold for all values of V^N. Thus, the $F = 1$ case yields no η-causality-obeying curved metrics, and the $\{31\}$ type is eliminated. It remains to consider $F = 0$ for the $\{1,111\}$ and $\{211\}$ types. The resulting inequality is $B(V^N)^2 + (D - A)(V^X)^2 \geq 0$. Because $B \geq 0$ has already been imposed, it follows only that $D \geq A$.

Let us summarize the results so far. The inequalities $B \geq 0$ and $C \geq 0$ have excluded the type $\{z\,\bar{z}11\}$. We also have $D > 0$, $D \geq A$, $E > 0$, $E \geq A$. Finally, $F = 0$ excludes the type $\{31\}$, so only $\{1,111\}$ and $\{211\}$ remain.

We now impose the requirement of Lorentzian signature. At a given event, there can be no objection to finding a coordinate x such that $(\frac{\partial}{\partial x})^\mu = x^\mu$ and (flipping the sign of y^μ if needed for the orientation) a coordinate y such that $(\frac{\partial}{\partial y})^\mu = y^\mu$; these two coordinates can be regarded as Cartesian. Then the null vectors L^μ and N^μ lie in the $t-z$ plane of this sort of Cartesian system. The curved metric has a block diagonal part in the $x-y$ plane with positive determinant, so imposing a Lorentzian signature means ensuring a negative determinant for the 2×2 $t-z$ part. The vectors L^μ and N^μ in one of these coordinate systems take the form $L^\mu = (L^0, 0, 0, L^3)$ and $N^\mu = (N^0, 0, 0, N^3)$. Given the Cartesian form $\eta_{\mu\nu} = diag(-1, 1, 1, 1)$ and the nullity of these two vectors, it follows that $|L^0| = |L^3|$ and $|N^0| = |N^3|$. Therefore the relevant parts of the curved metric can be written in such a coordinate basis as

$$\begin{aligned} g_{00} &= 2AL^0N^0 + B(L^0)^2 + C(N^0)^2, \\ g_{03} = g_{30} &= -A(N^0L^3 + L^0N^3) - BL^0L^3 - CN^0N^3, \\ g_{33} &= 2AL^3N^3 + B(L^3)^2 + C(N^3)^2. \end{aligned} \tag{6.10}$$

Taking the determinant using *Mathematica*©[3] and recalling that $|L^0| = |L^3|$ and $|N^0| = |N^3|$, one finds that the condition for a negative determinant is $2(A^2 - BC)|L^3|^2|N^3|^2(sign(L^0L^3N^0N^3) - 1) < 0$. The linear independence of L^μ and N^μ implies that $sign(L^0L^3N^0N^3) = -1$, so the determinant condition is $A^2 - BC > 0$. Because B and C are both nonnegative, $A^2 - BC > 0$ implies that $A \neq 0$. But the requirement that the curved metric be smoothly deformable through a sequence of signature-preserving steps means that the curved metric's value of A cannot "jump" from one sign of A to another, but must agree with the flat metric's positive sign. It follows that $A > 0$.

[3] *Mathematica* is produced by Wolfram Research, www.wolfram.com.

We now summarize the necessary conditions imposed:

$$A > 0, \quad A^2 > BC, \quad B \geq 0,$$
$$C \geq 0, \quad D \geq A, \quad E \geq A,$$
$$F = 0. \tag{6.11}$$

6.5 Sufficient Conditions for Respecting the Flat Metric's Null Cone

Thus far, we have shown nothing of the sufficiency of these necessary conditions. We now prove that these conditions are in fact sufficient. It is helpful to consider the two types, $\{1, 111\}$ and $\{211\}$, separately.

For the type $\{1, 111\}$, the conditions on the coefficients A, B, *etc.* reduce to

$$A > 0, \quad A > B, \quad B \geq 0,$$
$$C = B, \quad D \geq A, \quad E \geq A. \tag{6.12}$$

For this form the following relations between variables hold:

$$A = \rho_0, \quad B = \rho_1, \quad D = \rho_2,$$
$$E = \rho_3. \tag{6.13}$$

It follows that this type can be expressed as

$$g_{\mu\nu} = -(A - B)T_\mu T_\nu + (A + B)Z_\mu Z_\nu + DX_\mu X_\nu + EY_\mu Y_\nu. \tag{6.14}$$

Writing the eigenvalues for T^μ, X^μ, Y^μ, and Z^μ as D_0^0, D_1^1, D_2^2, and D_3^3, respectively, one has

$$D_0^0 = A - B, \quad D_1^1 = A + B, \quad D_2^2 = D,$$
$$D_3^3 = E. \tag{6.15}$$

One sees that the inequalities imply that the eigenvalue for the timelike eigenvector T^μ (briefly, the "timelike eigenvalue") is no larger than any of the spacelike eigenvalues:

$$D^0_0 \leq D^1_1, \quad D^0_0 \leq D^2_2, \quad D^0_0 \leq D^3_3, \tag{6.16}$$

and that all the (generalized) eigenvalues are positive. Let us now see that these conditions are sufficient. Writing an arbitrary vector V^μ as $V^\mu = V^T T^\mu + V^X X^\mu + V^Y Y^\mu + V^Z Z^\mu$, one sees that its η-length (squared) is $\eta_{\mu\nu} V^\mu V^\nu = -(V^T)^2 + (V^X)^2 + (V^Y)^2 + (V^Z)^2$. Clearly this length is never more positive than $\frac{1}{D^0_0} g_{\mu\nu} V^\mu V^\nu = -(V^T)^2 + \frac{D^1_1}{D^0_0}(V^X)^2 + \frac{D^2_2}{D^0_0}(V^Y)^2 + \frac{D^3_3}{D^0_0}(V^Z)^2$, so the necessary conditions are indeed sufficient for type $\{1,111\}$.

For the type $\{211\}$, the conditions on the coefficients A, B, *etc.* reduce to

$$\begin{aligned} &A > 0, \quad B > 0, \quad C = 0, \\ &D \geq A, \quad E \geq A, \quad F = 0. \end{aligned} \tag{6.17}$$

One can write the curved metric in terms of T^μ, Z^μ, X^μ, and Y^μ, though T^μ and Z^μ are not eigenvectors. One then has

$$g_{\mu\nu} = -(A - \frac{1}{2}B)T_\mu T_\nu + (A + \frac{1}{2}B)Z_\mu Z_\nu + BZ_{(\mu}T_{\nu)} + DX_\mu X_\nu + EY_\mu Y_\nu. \tag{6.18}$$

Writing an arbitrary η-spacelike vector field V^μ as $V^\mu = GT^\mu + HZ^\mu + IX^\mu + JY^\mu$, with $H^2 + I^2 + J^2 > G^2$, one readily finds the form of $g_{\mu\nu}V^\mu V^\nu$. Employing the relevant inequalities and shuffling coefficients, one obtains the manifestly positive result $g_{\mu\nu}V^\mu V^\nu = A(H^2 + I^2 + J^2 - G^2) + \frac{1}{2}B(G - H)^2 + (D - A)I^2 + (E - A)J^2$. This positivity result says that all η-spacelike vectors are g-spacelike. Earlier the requirement that all η-null vectors be g-null or g-spacelike was imposed. These two conditions together comprise the causality principle, so we have obtained sufficient conditions for the $\{211\}$ type also.

The $\{211\}$ type, which has with one null and two spacelike eigenvectors, is a borderline case in which the curved metric's null cone is tangent to the flat metric's cone along a single direction [327]. Clearly such borderline cases of $\{211\}$ metrics obeying the causality principle form in some sense a measure 0 set of all causality principle-satisfying metrics. Given that they are so scarce, one might consider neglecting them. Furthermore, they are arbitrarily close to violating the causality principle. We recall the criterion of stable causality in geometrical general relativity [152] (where the issue is closed timelike curves, without regard to any flat metric's null cone), which frowns upon metrics which satisfy causality, but would fail to do so if perturbed by an arbitrarily small amount. One could imagine that quantum fluctuations might push such a marginal metric over the edge, and thus prefers to exclude such metrics as unphysical. By analogy, one might impose stable η-causality, which excludes curved metrics that are arbitrarily close to violating the flat null cone's notion of causality, though we saw that such a condition would exclude conformally flat metrics, also. Perhaps a better reason for neglecting type $\{211\}$ metrics is that they are both technically inconvenient and physically unnecessary. Because η-causality-respecting $\{211\}$ metrics are arbitrarily close to $\{1,111\}$ metrics, one could merely make a gauge transformation[4] to shrink the lapse a bit more and obtain a $\{1,111\}$ metric instead. Thus, every curved metric that respects η-causality either is of type $\{1,111\}$, or is arbitrarily close to being of type $\{1,111\}$ and deformable thereto by a small gauge transformation reducing the lapse.

It follows that there is no loss of generality in restricting the configuration space to type $\{1,111\}$ curved metrics, for which the two metrics are simultaneously diagonalizable. As a result, there exists a close relationship between the traditional orthonormal tetrad formalism and this eigenvector decomposition. In particular,

[4]Here we refer to gauge transformations in the field formulation, where the flat metric's null cone is ignored. It will become evident that the notion of a gauge transformation in the special relativistic approach, which respects η-causality, is more restrictive.

one can build a g-orthonormal tetrad field e^{μ}_{A} simply by choosing the normalization of the eigenvectors. This choice removes the local Lorentz freedom of the tetrad (except when eigenvalues are degenerate). One might enforce this choice using an ineffective constraint by including a term in the action of the form

$$\int d^4 \eta_{\mu\nu} e^{\mu}_{A} e^{\nu}_{B} \eta_{\alpha\rho} e^{\alpha}_{C} e^{\rho}_{E} \Lambda^{ABCE}. \tag{6.19}$$

Here Λ^{ABCE} is a Lagrange multiplier which vanishes when $A = B$ or when $C = E$ and which is symmetric on its first pair of indices, on its second pair of indices, and under interchange of the first pair with the second pair.

Rewriting the generalized eigenvector equation for the case in which a complete set exists, one can write $g_{\mu\nu} e^{\mu}_{A} = \eta_{\mu\nu} e^{\mu}_{B} D^{B}_{A}$, with the four eigenvalues being the elements of the diagonal matrix D^{A}_{B}. It is sometimes convenient to raise or lower the indices of this matrix using the matrix $\eta_{AB} = diag(-1, 1, 1, 1)$. The tetrad field has $\{e^{\mu}_{A}\}$ has inverse $\{f^{A}_{\mu}\}$. We recall the standard relations $g_{\mu\nu} = f^{A}_{\mu} \eta_{AB} f^{B}_{\nu}$ and $g_{\mu\nu} e^{\mu}_{A} e^{\nu}_{B} = \eta_{AB}$. It is not difficult to show the how the tetrad lengths are related to the eigenvalues: $\eta_{\mu\nu} e^{\mu}_{A} e^{\nu}_{B} = D^{-1}_{AB}$, and equivalently, $\eta^{\mu\nu} f^{A}_{\mu} f^{B}_{\nu} = D^{AB}$. It follows that $f^{A}_{\nu} = \eta_{\nu\alpha} e^{\alpha}_{B} D^{AB}$, which says that a given leg of the cotetrad f^{A}_{ν} can be expressed solely in terms of the corresponding leg of the tetrad e^{μ}_{A}, through a stretching, an index lowering, and possibly a sign change, without reference to the other legs. Simultaneous diagonalization implies that the tetrad vectors are orthogonal to each other with respect to *both* metrics.

It might be interesting to use these eigenvectors as the tetrad field in C. Møller's tetrad formalism. Concerning localization of gravitational energy, Møller concluded that a satisfactory solution within Riemannian geometry does not exist, but that one does exist in a tetrad form of general relativity, apart from the question of finding six extra equations to fix the freedom under local Lorentz transformations, because the localization does not depend on the world coordinate choice or the global Lorentz frame [211–215]. Recent work with teleparallel gravity, a modernized form of Møller's work, reaches similar conclusions about these invariances or their absence [335]. In the case of the present eigenvector formalism, the additional six equations require the several eigenvectors to be η-orthogonal to each other. Given a flat background metric tensor, one expects that world coordinate invariance might go over to invariance under gravitational gauge transformations, yielding a bimetric tetrad energy localization that depends only on the local Lorentz transformations, while imposing η-orthogonality of the tetrad would remove even the local Lorentz freedom (except in the case of degeneracies). Perhaps we have found a way to complete Møller's program. However, one might note that for Møller, the use of an unobservable flat background metric, at least as it appeared in the work of F. H. J. Cornish [47], "does not seem to me quite satisfactory" [215] (p. 10). Thus, we cannot be confident that Møller himself would have approved of our suggestion. On the other hand, given that our null cone formalism proves that the flat metric can be physically significant even without being directly observed in the field equations, perhaps Møller's doubts would have been allayed. In any case, the use of an η-orthogonal tetrad in gravitational energy localization deserves study.

As we have previously noted [111], one can also use the flat metric to find a symmetric tensorial 'square root' of the (curved) metric, which can be used in place of a tetrad to couple gravity to fermions.[5] This procedure was implemented by Huggins [42], and similar work was done a bit later by Ogievetskiĭ and Polubarinov [274]. The quantities used take the form of an infinite series in the gravitational potential, which is defined as something like the difference between the curved and flat metrics. (There are various possible definitions, either contravariant or covariant, and with varying density weights.) Being symmetric, such entities have only 10 components, which number seems preferable to the larger number (16) of a tetrad field. But

[5]The existence of this formalism suggests attempting to include fermions in the above universal coupling derivation, though the fact that the desired result is not expressed using the curved metric alone suggests some complication.

this advantage diminishes when the square root is compared to the "eigenvierbein" formalism described above, which tends to have 10 components also. Moreover, the infinite series is not too convenient for making gauge transformations. Finally, the square root of the metric is not known to have any close connection to the null cone issue in general, whereas the eigenvierbein is. But that relationship is a fundamental issue for the SRA, so the eigenvierbein is to be preferred over the square root of the curved metric in the SRA.

6.6 Linearized Plane Waves a Difficulty for the Logunov Causality Principle

As was stated in connection with Nicholas's drumming, monochromatic plane gravitational waves satisfying the linearization of the Einstein equations and the Hilbert (linearized DeDonder) gauge appear to violate η-causality in general. We will now show that in more detail, using Ohanian and Ruffini [304] as our guide, while making use of the eigenvalue technology introduced above. These results will also hold approximately for the Maheshwari-Logunov massive theory for large frequencies and weak fields.

Defining a trace-reversed potential $\phi^{\mu\nu} = \gamma^{\mu\nu} - \frac{1}{2}\eta^{\mu\nu}\gamma$ (with $\gamma^{\mu}_{\mu} = \gamma$) and imposing the Hilbert gauge $\partial_{\mu}\phi^{\mu\nu} = 0$, one puts the linearized Einstein equations in the form $\partial^2\phi^{\mu\nu} = 0$. Because we desire plane wave solutions, let $\lambda\phi^{\mu\nu} = H\epsilon^{\mu\nu}\cos(k_{\alpha}x^{\alpha} + \psi)$, where $\epsilon^{\mu\nu}$ is a constant polarization tensor, k^{α} a constant polarization vector, and H is a small number fixing the amplitude. We let the waves travel in the z-direction, so $k^{\mu} = \omega(1,0,0,1)$. We also define the vectors $\epsilon^{\mu}_{1} = (0,1,0,0)$ and $\epsilon^{\mu}_{2} = (0,0,1,0)$.

The gauge condition implies that the polarization tensor is orthogonal to the propagation vector: $\epsilon^{\mu\nu}k_{\mu} = 0$, leaving six independent solutions. One can take the six independent polarization tensors (with both indices lowered using the flat metric) to be:

$$\epsilon_{1\mu\nu} = \epsilon_{1\mu}\epsilon_{1\nu} - \epsilon_{2\mu}\epsilon_{2\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.20}$$

$$\epsilon_{2\mu\nu} = \epsilon_{1\mu}\epsilon_{2\nu} + \epsilon_{2\mu}\epsilon_{1\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{6.21}$$

$$\epsilon_{3\mu\nu} = \epsilon_{1\mu}\frac{1}{\omega}k_\nu + \epsilon_{1\nu}\frac{1}{\omega}k_\mu = \begin{bmatrix} 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \tag{6.22}$$

$$\epsilon_{4\mu\nu} = \epsilon_{2\mu}\frac{1}{\omega}k_\nu + \epsilon_{2\nu}\frac{1}{\omega}k_\mu = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \tag{6.23}$$

$$\epsilon_{5\mu\nu} = k_\mu k_\nu = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}, \tag{6.24}$$

$$\epsilon_{6\mu\nu} = \epsilon_{1\mu}\epsilon_{1\nu} + \epsilon_{2\mu}\epsilon_{2\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{6.25}$$

The first two, which are transverse-transverse and traceless, are the physical (energy-carrying) polarizations in the massless theory. One can show that they induce effective curved metrics of type $\{1111\}$, but, partly due to the oscillations of the cosine function, they violate η-causality, as the behavior of the eigenvalues shows. Let us be rather explicit and find an η-null vector that is on occasion g-timelike for the first polarization. The polarization tensor being traceless, one has $g_{\mu\nu} = \eta_{\mu\nu} - \lambda\gamma_{\mu\nu} = diag(-1,1,1,1) - H\cos(k_\alpha x^\alpha + \psi) diag(0,1,-1,0)$. The vector $U^\mu = (1,1,0,0)$ is η-null, but is g-timelike during every other half-period of the cosine function. Thus, this physical polarization violates η-causality.

A reply to a previous but less explicit version of this claim on our part has recently been made by Logunov and M. A. Mestvirishvili [369]. In addition to providing the above counterexample, we wish to explain why we do not accept their reply. In particular, their argument for equation (42) is invalid. The argument in question is that the Hilbert (linearized tensorial DeDonder) condition implies that the 4-velocity of a test particle is orthogonal to the gravitational potential. While the wave vector $k^\mu = (1,0,0,1)$ is indeed orthogonal to the gravitational potential, no such relation holds for all η-null vectors, such as $(1,1,0,0)$. The worry is not that the gravitational wave *itself* will propagate outside the flat metric's null cone, but that another massless test body–let it be a photon–in the vicinity, moving perpendicular to the gravitational wave propagation direction (in the x direction in this case), will violate the flat metric's null cone.[6] We recall from Logunov's formulation of the causality principle that *all* η-null vectors must be g-null or g-spacelike [89], but that condition fails here. The error in their equation (42) propagates into equation (43), from which they erroneously concluded that η-causality is satisfied for this solution. Thus, this gravitational wave polarization indeed violates η-causality, in the sense that any photon moving in the x-direction through this gravitational wave

[6] We thank Prof. Logunov for corresponding about this matter.

will propagate faster than c (using metaphysical rods and clocks), in contradiction of special relativistic causality.

Perhaps at this stage one could hope that the gauge waves will be more friendly toward η-causality, and that by including enough of the gauge waves, one could tame the bad properties of the physical waves. The polarizations $\epsilon_{3\mu\nu}$ and $\epsilon_{4\mu\nu}$, which are transverse-longitudinal and traceless, however, will probably not help, because the resulting metrics are of type $\{31\}$, and thus violate η-causality. The fifth polarization $k_\mu k_\nu$, which is longitudinal-longitudinal and traceless[7], is of type $\{211\}$, and thus could satisfy η-causality, but the fluctuating cosine implies that if causality is respected during one half-period, then it is violated during the next. So it does not help much either. The last polarization $\epsilon_{1\mu}\epsilon_{1\nu} + \epsilon_{2\mu}\epsilon_{2\nu}$ is transverse-transverse but not traceless. This wave is of type $\{1111\}$, but the oscillating eigenvalues once again violate η-causality.

Thus, all six polarizations of plane wave admitted by the linearized Einstein equations in the Hilbert (linearized DeDonder) gauge individually violate η-causality. It would be surprising if it is possible to make sufficiently general (for Nicholas's drumming, for example) superpositions of these waves that respect η-causality, given that each one violates it individually. Let us consider this question more explicitly. One can consider a general superposition

$$\lambda\gamma_{\mu\nu} = \sum_{A=1}^{6} H_A \epsilon_{A\mu\nu} \cos(k_\alpha x^\alpha + \psi_A) - \eta_{\mu\nu} H_6 \cos(k_\alpha x^\alpha + \psi_6), \tag{6.26}$$

where ψ_A are real phases. The characteristic polynomial $|g_{\mu\nu} - \Lambda\eta_{\mu\nu}| = 0$ can be simplified by defining $B_A = H_A \cos(k_\alpha x^\alpha + \psi_A)$ (with no summation over A). The

[7]The solution presented by I. P. Denisova, to be discussed below, resembles this form, except that her function W is strictly negative, so η-causality is preserved, whereas our cosine function fluctuates in sign.

result is

$$\begin{vmatrix} 1+B_5+B_6-\Lambda & -B_3 & -B_4 & -B_5 \\ -B_3 & -1+B_1+\Lambda & B_2 & B_3 \\ -B_4 & B_2 & -1-B_1+\Lambda & B_4 \\ -B_5 & B_3 & B_4 & -1+B_5-B_6+\Lambda \end{vmatrix} = 0. \tag{6.27}$$

It looks difficult to get practical results from this equation in the general case, it being a messy quartic polynomial. However, the third and fourth polarizations, affecting the only shift and an off-diagonal spatial component each, look rather unlikely to help satisfy η-causality, but probably would harm it. It seems fair to set $B_3 = B_4 = 0$. The characteristic polynomial then becomes block diagonal, with roots $\Lambda = 1 \pm \sqrt{B_1^2 + B_2^2}$ and (repeated) $\Lambda = 1 + B_6$. If $B_5 \neq 0$, then $\Lambda = 1 + B_6$ corresponds to the null eigenvector $(1, 0, 0, 1)$, yielding a $\{211\}$ metric (which at best only just satisfies causality), but setting $B_5 = 0$ gives another eigenvector and hence a $\{1111\}$ metric, which can satisfy stable causality, so we choose the latter. However, it is clear that one of the other two eigenvalues will at almost every moment be less than unity if a physical wave is present, while $1 + B_6$ will be greater than unity half the time. Thus, plausibly, a general superposition of gauge and physical polarizations will not rescue η-causality.

Clearly this argument is not a mathematical proof. It does not even use exact solutions of Einstein's equations. However, this argument does make it seem likely that plane wave-like solutions, which are necessary near the future light cone of Nicholas's drum set, violate η-causality. If that conclusion is correct, then the RTG's approach to causality indeed must be modified, as we urged above.

6.7 Dynamics of the Causality Principle

It is one thing to know the kinematic inequalities to describe the causality principle, but another to enforce them. A difficulty comes from trying to mate the causality principle with the condition that, as the gravitational field gets very weak,

$$g_{\mu\nu} \rightarrow \eta_{\mu\nu}, \tag{6.28}$$

which seems like a natural requirement on a flat spacetime theory of gravity. (However, it is not gauge-invariant.) Given that the derivation of general relativity involves a defining relation of the form $g_{\mu\nu} = \eta_{\mu\nu} - \lambda\gamma_{\mu\nu}$ [111] or the like for the curved metric in terms of the flat and the gravitational potential $\gamma_{\mu\nu}$, clearly $g_{\mu\nu} = \eta_{\mu\nu}$ is expected to be the state of no gravity (apart from gauge transformations).

The causality principle inequalities can be written as

$$D^0_0 \leq D^1_1, \quad D^0_0 \leq D^2_2, \quad D^0_0 \leq D^3_3. \tag{6.29}$$

One can abbreviate this set of inequalities by

$$D^0_0 \leq \min(D^1_1, D^2_2, D^3_3). \tag{6.30}$$

In three spacetime dimensions one can readily plot the function $D^0_0 \leq \min(D^1_1, D^2_2)$ using *Mathematica*. The result is seen in two spatial dimensions in figure 6.1. The greatest allowed values of the timelike eigenvalue lie on the folded surface shown. In three spatial dimensions, a similar (but more complicated) situation arises.

Perhaps motivated by the suggestion that bimetric theories ought to be gauge-fixed, one might attempt to enforce the causality principle by (partial) gauge-fixing, perhaps setting the timelike eigenvalue equal to some function of the spacelike eigenvalues only. This would be an algebraic or "ultralocal" gauge fixing. One might want to choose a specific function $D^0_0 = f(D^1_1, D^2_2, D^3_3)$ to fix the gauge. If $f(D^1_1, D^2_2, D^3_3) = \min(D^1_1, D^2_2, D^3_3)$, then the curved null cone is always on the verge of violating the flat cone, but never quite does so. By analogy with the requirement of stable causality in the standard causal analysis of geometrical general relativity, one would prefer that the metric not be arbitrarily close to unacceptable behavior, because a quantization might lead to fluctuations that push it over the edge [152]. A more attractive choice is therefore

$$f(D^1_1, D^2_2, D^3_3) = \frac{\min^2(D^1_1, D^2_2, D^3_3)}{\max(D^1_1, D^2_2, D^3_3)}, \tag{6.31}$$

which provides a cushion except when the curved metric equals the flat one. This choice also corresponds to the usual form of the Schwarzschild solution, while being consistent with the explicitly conformally flat form of the flat Robertson-Walker cosmological model [152].

If one were to write these functions explicitly, they would contain Heaviside step functions. Thus they are merely continuous, which is a rather weak level of differentiability. This difficulty is in fact inevitable. If the selected gauge at all permits the D^0_0 to take its maximum value along the fold, then the gauge can be continuous, but clearly is not twice continuously differentiable, as one would wish. Because the lapse is closely related to D^0_0, a merely continuous timelike eigenvalue implies a merely continuous lapse. The lapse being gauge freedom, one might hope that continuity is enough for the lapse. Unfortunately, the Einstein equations (in Hamiltonian form) contain second spatial derivatives of the lapse [152]. Two derivatives of a merely continuous function will yield a Dirac delta function in the $\dot{\pi}^{ab}$ equation. So the momentum will evolve discontinuously if D^0_0 ever touches the fold. Such behavior seems physically unreasonable, so the fold must be avoided if a sensible result is to emerge. Avoiding the fold under all circumstances means that a flat spacetime theory of gravity does not have $g_{\mu\nu} = \eta_{\mu\nu}$ as a solution, which seems surprising at first, but the gauge-variance of $g_{\mu\nu} = \eta_{\mu\nu}$ implies that this loss is not too worrisome.

In any case, if one wishes to fix the gauge, there is no need to restrict attention to such *ultralocal* forms as $D^0_0 = f(D^1_1, D^2_2, D^3_3)$, with no derivatives present. Why not a merely local gauge fixing, with a finite number of derivatives? Analogously, in geometrical general relativity, one considers not only Gaussian normal coordinates $g^{00} = -1, g^{0i} = 0$, but also harmonic coordinates, and in electromagnetism, one considers not only the temporal gauge $A_0 = 0$ [328], but also the Lorentz gauge. However, even given this enlarged set of gauge-fixing choices, it is not clear that there exists any way to include the solution ($g_{\mu\nu} = \eta_{\mu\nu}$) without encountering discontinuous evolution, which seems unacceptable.

It seems best to avoid these difficulties altogether at the theoretical level by abstaining from *fixing* the gauge (and instead merely *restrict* the gauge to respect η-causality), and to take the gauge freedom seriously enough to give up the solution $g_{\mu\nu} = \eta_{\mu\nu}$ as inessential. One might even aim to satisfy the causality-related restrictions on the gauge *identically* using a new set of variables adapted to this purpose.

6.8 Satisfying η-causality Identically Using New Variables

If one imposes stable η-causality, which excludes all metrics that satisfy η-causality but which are arbitrarily close to violating it, along with all that violate it, then the loose inequalities of the causality principle are changed to strict inequalities. These inequalities are then susceptible to being *solved and eliminated*, as Klotz proposed to do with the positivity conditions of canonical gravity [313, 314]. Let us see how this goal can be achieved. For generality, let us work in d spacetime dimensions, where $d = 2$, $d = 3$, and $d = 4$ are perhaps of the most interest, unless one is considering Kaluza-Klein theory. (For $d \neq 4$, one might want to verify that nothing disturbing happens regarding the eigenvector formalism happens, although that seems unlikely.) Given stable η-causality, there always exists a complete set of generalized eigenvectors with real eigenvalues and eigenvectors, as we proved earlier. Some time ago J. A. Schouten cryptically wrote, "The theorems of principal axes and of principal blades do not hold if the fundamental tensor is indefinite because in this case a real symmetric tensor or bivector may possibly have a special position with respect to the real nullcone. But if the symmetric tensor or bivector does not have any such special position the theorems remain valid." [288] (p. 47) Evidently stable η-causality prevents $g_{\mu\nu}$ from having a "special position" with respect to the fundamental tensor $\eta_{\mu\nu}$.

It is clear that the determinant of the effective curved metric, as long as it remains negative, is irrelevant to the relation between the two null cones. One could then split the curved metric into the (conformally invariant) unimodular part $g_{\mu\nu}/\sqrt[d]{-g}$ and the determinant g. However, it is no worse in Cartesian coordinates, and perhaps more convenient in non-Cartesian coordinates, to use the *covariantly* unimodular metric $\hat{g}_{\mu\nu} = g_{\mu\nu}/\sqrt[d]{\frac{g}{\eta}}$ and the coordinate scalar $\kappa = \sqrt{\frac{g}{\eta}}$.

Let us consider the generalized eigenvector problem in terms of the covariantly unimodular metric:

$$\hat{g}_{\mu\nu} u^{\mu} = \eta_{\mu\nu} u^{\mu} \hat{\Lambda}. \tag{6.32}$$

Given that we have already required that the curved metric have a complete set of eigenvectors with respect to the flat, we can multiply the relation $\hat{g}_{\mu\nu} u^{\mu}_{A} = \eta_{\mu\nu} u^{\mu}_{B} \hat{D}^{B}_{A}$ (where the diagonal matrix $\hat{D}^{B}_{A}$ has unit determinant, rendering $\hat{D}^{0}_{0}$ dependent on the other components) by the inverse matrix U^{-1} of the matrix U of unnormalized eigenvectors u^{μ}_{A}. The result is

$$\hat{g}_{\rho\nu} = \eta_{\mu\nu} u^{\mu}_{B} \hat{D}^{B}_{A} u^{-1A}_{\rho}. \tag{6.33}$$

If this equation is twice contracted with the eigenvalue matrix U, giving

$$\hat{g}(u_A, u_C) = \eta(u_B, u_C) \hat{D}^{B}_{A}, \tag{6.34}$$

then the freedom to normalize the eigenvectors can be put to work. It is useful to make u^{μ}_{A} be η-unit vectors.[8]

Let us now specialize to Cartesian coordinates and write u^{C}_{A}, if necessary, to remind ourselves of that fact. The normalization relation $\eta_{\mu\nu} u^{\mu}_{A} u^{\nu}_{B} = \eta_{AB}$, if expressed in Cartesian coordinates, is nothing other than the condition [344] for U to be in the complete Lorentz group $O(1,3)$ if $d = 4$. It is clear on physical grounds that the gravitational field ought not to reverse time or space, so only the subgroup $SO(1,3) \uparrow$, the proper orthochronous Lorentz group with $|U| = 1$ and $u^{0}_{0} \geq 0$, is relevant. This subgroup being connected to the identity, any of its matrices u^{C}_{A} can be obtained by exponentiating a matrix in the Lie algebra of the Lorentz group, that is, the algebra of real matrices W^{A}_{B} such that $\eta_{CA} W^{A}_{B}$ is antisymmetric [344]. Given this η-unit normalization convention and the use of Cartesian coordinates, we can use the relations

$$\begin{aligned} U &= e^{W}, \\ U^{-1} &= e^{-W} \end{aligned} \tag{6.35}$$

to eliminate the eigenvectors in favor of a matrix W describing how the eigenvectors are boosted or rotated relative to the coordinate basis. The orthogonality of the eigenvectors is satisfied identically using this construction. If one wishes, it is possible to write W in terms of a standard representation of the infinitesimal generators of the Lorentz group with known commutation relations [345] (pp. 538-541). Setting $d \neq 4$ would of course require using the Lorentz group in the appropriate dimensions. Recently progress with this sort of expression has been made [336].

Let us consider the eigenvalues $\hat{D}^{A}_{B}$. We saw above that the timelike eigenvalue for the covariantly unimodular metric is dependent upon the spacelike eigen-

[8]Earlier we saw that g-unit normalization was convenient for tying the eigenvectors to the standard orthonormal tetrad formalism. These two different normalizations are adapted to different purposes.

values. One can define reduced eigenvalues

$$\bar{D}^A_B = \frac{\hat{D}^A_B}{\hat{D}^0_0} = \frac{D^A_B}{D^0_0}; \tag{6.36}$$

clearly $\bar{D}^0_0 = 1$. It is interesting that now only three numbers (compared to nine components of $\hat{g}_{\mu\nu}$) are needed to describe the null cone relationship sufficiently to show whether η-causality is respected. Let us now eliminate $\hat{D}^A_B$ in favor of the reduced eigenvalues using the relation $\hat{D}^0_0 = \frac{1}{\sqrt[d]{|\bar{D}|}}$. At this stage, one can write the (covariantly) unimodular metric in matrix form as

$$\hat{g} = \frac{e^{-W}\bar{D}e^{W}\eta}{\sqrt[d]{|\bar{D}|}}. \tag{6.37}$$

While one could represent W by a 2-form [336] if one wished, which would permit the use of an arbitrary coordinate system, this eigenvalue formalism requires the introduction of a set of Cartesian coordinates for the eigenvalue matrix.

Imposing stable η-causality gives the *strict* inequalities

$$\bar{D}^1_1 > 1, \bar{D}^2_2 > 1, \bar{D}^3_3 > 1. \tag{6.38}$$

But strict inequalities can be solved, as Klotz suggested [313, 314], and then the causality constraints would be eliminated. This goal is achieved by setting, for example,

$$\bar{D}^1_1 = e^{\alpha} + 1, \bar{D}^2_2 = e^{\beta} + 1, \bar{D}^3_3 = e^{\gamma} + 1, \tag{6.39}$$

for the case $d = 4$, with obvious changes for other dimensions. Formally defining a matrix $M^B_A = diag(-\infty, \alpha, \beta, \gamma)$ lets one write $\bar{D} = I + e^M$. The covariantly unimodular metric is now $\hat{g} = \frac{(I+e^{-W}e^{M}e^{W})\eta}{\sqrt[d]{|I+e^M|}}$. One can also set $\kappa^{\frac{2}{d}} = e^{\delta}$. The full metric can then be written as

$$g = \frac{e^{\delta}(I + e^{-W}e^{M}e^{W})\eta}{\sqrt[d]{|I + e^M|}}. \tag{6.40}$$

We let the lower index indicate the row and the upper, the column.

If $d = 2$, it is not difficult to write out this form of the curved metric explicitly. (In higher dimensions, the noncommutativity of the various rotations and boosts in W presents more difficulty, although some simplification has been achieved [336].) The matrix W can be put in the form

$$W = \begin{bmatrix} 0 & \epsilon \\ \epsilon & 0 \end{bmatrix}. \tag{6.41}$$

One quickly finds that

$$\begin{aligned} U &= \begin{bmatrix} \cosh\epsilon & \sinh\epsilon \\ \sinh\epsilon & \cosh\epsilon \end{bmatrix}, \\ U^{-1} &= \begin{bmatrix} \cosh\epsilon & -\sinh\epsilon \\ -\sinh\epsilon & \cosh\epsilon \end{bmatrix}. \end{aligned} \tag{6.42}$$

Finishing the elementary algebra gives

$$g = \frac{e^{\delta}}{\sqrt{1+e^{\alpha}}} \begin{bmatrix} -1 + e^{\alpha}\sinh^2\epsilon & -e^{\alpha}\cosh\epsilon\sinh\epsilon \\ -e^{\alpha}\cosh\epsilon\sinh\epsilon & 1 + e^{\alpha}\cosh^2\epsilon \end{bmatrix}. \tag{6.43}$$

With all three (or ten or $\frac{d(d+1)}{2}$) components of the curved metric expressed in a form along these lines, the following highly desirable properties of the curved metric are satisfied *automatically*:

1. the curved metric satisfies stable η-causality;
2. the curved metric has Lorentzian signature;
3. the gravitational field has not reversed time or space;
4. the curved metric respects global hyperbolicity.

It would seem that these are all the properties that one would desire from an effective curved metric in the SRA. The matrix W (apart from the near-antisymmetry) and α, β, γ, and δ can be any real numbers. Thus, as long as a curved metric can be written in terms of these variables, then it is satisfactory for the SRA, and *vice versa.* These variables are evidently well-suited for the SRA applied to a metric theory of gravity.

If it were not for the frightful amount of calculation involved, and probably the inconveniently long and complicated results of the calculations, it might be interesting to rewrite the entirety of gravitation in the SRA in these variables. One example of the difficulty is what becomes of the primary constraints in the canonical form of the theory, which show that not all 10 components of the metric have canonical momenta of the usual sort suited to Legendre transformations from the Lagrangian density [315]. Until Dirac [316] and J. L. Anderson [317] showed in 1958 how to reduce the primary constraints to the vanishing of some canonical momenta, the presence of primary constraints was a significant issue in canonical general relativity. However, the use of η-causality-adapted variables perhaps prevents reducing the primary constraints to the vanishing of some momenta, so this issue might be reopened.

The issue can be illustrated in a tractable form using a model theory suited to Kasner solutions [313, 314], in which the ansatz

$$g_{\mu\nu}(t, x^i) = diag(g_0(t), g_1(t), g_2(t), g_3(t)) \tag{6.44}$$

is made before varying the action. The resulting equations in fact reproduce the Einstein equations. In this model theory, the primary constraint is that π^0, the momentum conjugate to g_0, vanish. In the η-causality variables, this diagonal form of the metric is realized by setting $W = 0$ and letting the remaining fields depend on t only. One then obtains for (g_0, g_1, g_2, g_3),

$$\Bigg(\frac{-e^{\delta}}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}}, \\ \frac{e^{\delta}(1+e^{\alpha})}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}},$$

$$\frac{e^{\delta}(1+e^{\beta})}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}},$$
$$\left.\frac{e^{\delta}(1+e^{\gamma})}{\sqrt[4]{(1+e^{\alpha})(1+e^{\beta})(1+e^{\gamma})}}\right). \tag{6.45}$$

Given how the causality variables mix together the coordinates (g_0, g_1, g_2, g_3), one expects the momenta to be comparably mixed, which means that the primary constraint will also mix momenta together.

This model theory might be a good test bed for ascertaining what *explicitly* is the gauge freedom of the SRA. In other words, which vector fields ξ^{μ} generate gauge transformations that respect stable η-causality? With the Lagrangian density rewritten in terms of these η-causality-adapted variables, only this set of vector fields would be admitted as gauge transformations, while any others would be regarded as merely mathematical transformations. One fact that is already clear is that the set of admissible vector fields depends upon the field configuration prior to the transformation, because a vector field that preserves η-causality given one initial field configuration might violate it given another. This situation is rather different from the usual situation (in the geometrical or field formulation), in which just any (reasonable) vector field generates a gauge transformation.

With the requirement of η-causality imposed–perhaps stable η-causality using the causality variables–it follows that any "SRA spacetime" $(R^4, \eta_{\mu\nu}, g_{\mu\nu})$ is globally hyperbolic in the sense of Wald [152]. How so? It follows from η-causality that the future domain of dependence of an η-spacelike slice is in fact the whole of R^4. But global hyperbolicity just is the possession of a Cauchy surface [152], so any η-causal SRA spacetime $(R^4, \eta_{\mu\nu}, g_{\mu\nu})$ is globally hyperbolic.

One could further ask whether the SRA has a well posed initial value formulation. The use of harmonic coordinates has been a common technique for answering this question in the geometrical formulation [152], in which the choice of harmonic coordinates constitutes a gauge-fixing. Given that the SRA has restricted gauge freedom, and that the tensorial DeDonder gauge condition (which makes the coordinate g-harmonic when η-Cartesian) has causal difficulties for plane wave solutions, one might fear that the proofs of a well posed initial value formulation fail for the SRA. However, such a fear is groundless, as we see if we resist the temptation to use η-Cartesian coordinates, for which we have no need. The SRA has both coordinate freedom and gauge freedom. The choice of g-harmonic coordinates, if nothing is said about the functional form of the flat metric, is merely a coordinate choice, not a gauge choice. The eigenvalues, which express the relation between the two null cones, are coordinate scalars, and so are indifferent to the choice of g-harmonic coordinates (and possibly messy form for the flat metric). The gauge freedom has not been used at all, and thus is fully available for deforming the curved metric until it becomes consistent with the flat one. Therefore traditional harmonic coordinate approach to demonstrating a well posed initial value problem experiences no obstacles from the SRA. Recent work that permits coordinate freedom [346], though interesting, will not be essential for the SRA.

Some time ago D. Finkelstein and C. Misner noticed the existence of "homotopic" conservation laws in nonlinear field theories such as general relativity [347]. For general relativity, the conserved quantity is the "metric twist," which indicates how many times the null cone is ineliminably twisted around. However, given that "[i]t is readily seen that a metric twist necessarily contains regions in which the direction of time and causality are anomalous" [347] (p. 239), it is evident that the curved metric has vanishing twist for the SRA.

6.9 Infinitesimal Gauge Transformations and the Eigenvector Formalism

It will be interesting to know how the generalized eigenvalues and eigenvectors of $g_{\mu\nu}$ with respect to $\eta_{\mu\nu}$ change under an infinitesimal gauge transformation $\delta g_{\mu\nu} = \pounds_{\xi} g_{\mu\nu}$, $\delta\eta_{\mu\nu} = 0$. Then $g_{\mu\nu}V^{\mu} = \Lambda\eta_{\mu\nu}V^{\mu}$ becomes $(g_{\mu\nu} + \delta g_{\mu\nu})(V^{\mu} + \delta V^{\mu}) = (\Lambda + \delta\Lambda)\eta_{\mu\nu}(V^{\mu} + \delta V^{\mu})$. The $0th$ order terms vanish by assumption. (We now install the index A, over which we do not sum, to distinguish the several eigenvalues and their eigenvectors.) The result is

$$(g_{\mu\nu} - \Lambda_A\eta_{\mu\nu})\delta V^{\mu}_A + V^{\mu}_A(\delta g_{\mu\nu} - \eta_{\mu\nu}\delta\Lambda_A) = 0. \tag{6.46}$$

For fixed A, one can find the equation governing the variation in the eigenvalue by contracting the previous equation with V^{ν}_A, and thus obtain $V^{\mu}_A V^{\nu}_A(\delta g_{\mu\nu} - \eta_{\mu\nu}\delta\Lambda_A) = 0$. Using the result $g_{\mu\nu}V^{\mu}_A V^{\nu}_A = \Lambda_A\eta_{\mu\nu}V^{\mu}_A V^{\nu}_A$ which follows obviously from the original eigenvalue equation, one finds that

$$\frac{\delta\Lambda_A}{\Lambda_A} = \frac{V^{\mu}_A V^{\nu}_A \delta g_{\mu\nu}}{V^{\mu}_A V^{\nu}_A g_{\mu\nu}}, \tag{6.47}$$

where no choice of normalization has been employed, but the fact that all the eigenvalues are nonzero and all the eigenvectors nonnull has been used. Using the result for $\delta\Lambda_A$ in (6.46) and contracting with V^{ν}_B yields

$$V^{\mu}_B(g_{\mu\nu} - \Lambda_A\eta_{\mu\nu})\delta V^{\nu}_A = V^{\mu}_B V^{\nu}_A\left(\delta g_{\mu\nu} - g_{\mu\nu}\frac{V^{\alpha}_A V^{\rho}_A \delta g_{\alpha\rho}}{V^{\sigma}_A V^{\beta}_A g_{\sigma\beta}}\right). \tag{6.48}$$

If there is degeneracy among the eigenvalues, then use of this equation is more difficult than for the nondegenerate case. The causality principle will limit or forbid (if stable η-causality is required) the degeneracy between the timelike and the spacelike eigenvalues, but degeneracy among the spacelike eigenvalues will still be rather common, as in the Schwarzschild and Robertson-Walker cosmological solutions. For convenience, we henceforth confine our attention to the case with no degeneracy. It follows that

$$\eta_{\mu\nu} V^{\mu}_{B} (\Lambda_B - \Lambda_A) \delta V^{\nu}_{A} = V^{\mu}_{A} V^{\nu}_{B} \delta g_{\mu\nu}, \tag{6.49}$$

which gives three equations for δV^{μ}_{A} for a given value of A. The fourth equation for δV^{μ}_{A} results from one's choice of normalization of the eigenvectors.

The quantities most directly relevant to the causality principle are the "reduced" eigenvalues $\overline{\Lambda}_A = \frac{\Lambda_A}{\Lambda_0}$, of which obviously only the values $A = 1, 2, 3$ are interesting. We therefore replace the index A, which generally runs over $0, 1, 2, 3$, with the index j, which runs over $1, 2, 3$. One wishes for all the reduced eigenvalues to be no smaller than unit value. The reduced eigenvalues change under infinitesimal gauge transformations of the metric by

$$\delta \overline{\Lambda}_j = \overline{\Lambda}_j \left(\frac{V^{\mu}_{j} V^{\nu}_{j}}{V^{\alpha}_{j} V^{\rho}_{j} g_{\alpha\rho}} - \frac{V^{\mu}_{0} V^{\nu}_{0}}{V^{\alpha}_{0} V^{\rho}_{0} g_{\alpha\rho}} \right) \delta g_{\mu\nu}. \tag{6.50}$$

If one wishes to enforce the causality principle in a deep way, it seems natural to distinguish between ostensible gauge transformations that violate the causality principle and those that do not, as we saw above using the causality-adapted variables.

6.10 Finite Gauge Transformations and an Orthonormal Tetrad

The form of a *finite* gauge transformation for the densitized inverse metric tensor in the field formulation is known from the work of Grishchuk, Petrov, and A. D. Popova [82] to have the form

$$\mathfrak{g}^{\sigma\rho} \rightarrow e^{\pounds_{\xi}} \mathfrak{g}^{\sigma\rho}, u \rightarrow e^{\pounds_{\xi}} u, \eta_{\mu\nu} \rightarrow \eta_{\mu\nu} \tag{6.51}$$

in terms of the convenient variable $\mathfrak{g}^{\sigma\rho} = \sqrt{-g} g^{\sigma\rho}$, the flat metric tensor, and matter fields u described by some tensor densities (with indices suppressed).

We recall the bimetric form of the action above for a generally covariant theory, with the metric here expressed in terms of the weight 1 inverse metric:

$$S = S_1[\mathfrak{g}^{\mu\nu}, u] + \frac{1}{2}\int d^4x R_{\mu\nu\rho\sigma}(\eta)\mathcal{M}^{\mu\nu\rho\sigma} + 2b\int d^4x\sqrt{-\eta} + \int d^4x \partial_\mu \alpha^\mu. \quad (6.52)$$

Clearly the terms other than S_1 either do not change under the assumed field transformation, or do so at most by a boundary term, so our attention turns to $S_1 = \int d^4x \mathcal{L}_1$.

We now derive a useful formula. Writing out $e^{\pounds_\xi}A$ as a series $e^{\pounds_\xi}A = \sum_{i=0}^{\infty} \frac{1}{i!}\pounds_\xi^i A$ for some tensor density A will put us in a position to derive a useful 'product' rule for the exponential of Lie differentiation. One could write a similar series for another tensor density B. Multiplying the series and using the Cauchy product formula [343]

$$\sum_{i=0}^{\infty} a_i z^i \sum_{j=0}^{\infty} b_j z^j = \sum_{n=0}^{\infty}\sum_{k=0}^{n} a_k b_{n-k} z^k \quad (6.53)$$

and the n-fold iterated Leibniz rule [343]

$$[fg]^{(n)} = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} f^{(k)} g^{(n-k)}, \quad (6.54)$$

one recognizes the result as the series expansion of $e^{\pounds_\xi}(AB)$, so one has the result

$$(e^{\pounds_\xi}A)(e^{\pounds_\xi}B) = e^{\pounds_\xi}(AB) \quad (6.55)$$

In view of the matrix relationships among the various metric quantities, one has by definition that $(\mathfrak{g}^{\mu\nu} + \delta\mathfrak{g}^{\mu\nu})(\mathfrak{g}_{\rho\nu} + \delta\mathfrak{g}_{\rho\nu}) = \delta^\mu_\rho$ and various other relations. In that way, one can derive the form of $\delta\mathfrak{g}_{\rho\nu}$, δg, $\delta g_{\rho\nu}$, and the like. Let us show this fact explicitly for g, using $\mathfrak{g}^{\sigma\rho} + \delta\mathfrak{g}^{\sigma\rho} = e^{\pounds_\xi}\mathfrak{g}^{\sigma\rho}$. One knows from mathematics that a determinant is given by $|\mathfrak{g}^{\sigma\rho}| = \epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi\mathfrak{g}^{\alpha\beta}\mathfrak{g}^{\mu\chi}\mathfrak{g}^{\nu\psi}\mathfrak{g}^{\rho\phi}$, where $\epsilon(\alpha\mu\nu\rho)$ is the totally antisymmetric symbol with $\epsilon(0123) = 1$. Because this form for the determinant holds in any coordinate system, $\epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi$ is a *scalar*, so $e^{\pounds_\xi}(\epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi) = \epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi$. We therefore have

$$|e^{\pounds_\xi}\mathfrak{g}^{\sigma\rho}| = \epsilon(\alpha\mu\nu\rho)\delta^0_\beta\delta^1_\chi\delta^2_\psi\delta^3_\phi(e^{\pounds_\xi}\mathfrak{g}^{\alpha\beta})(e^{\pounds_\xi}\mathfrak{g}^{\mu\chi})(e^{\pounds_\xi}\mathfrak{g}^{\nu\psi})(e^{\pounds_\xi}\mathfrak{g}^{\rho\phi})$$
$$= e^{\pounds_\xi}|\mathfrak{g}^{\sigma\rho}|. \qquad (6.56)$$

Using $\mathfrak{g}^{\sigma\rho} = g^{\sigma\rho}\sqrt{-g}$, one finds that $|\mathfrak{g}^{\sigma\rho}| = |g_{\sigma\rho}|$, so

$$g + \delta g = e^{\pounds_\xi} g. \qquad (6.57)$$

The relation $-g - \delta g = (\sqrt{-g} + \delta\sqrt{-g})^2$ defines $\delta\sqrt{-g}$, so one quickly also finds that

$$\sqrt{-g} + \delta\sqrt{-g} = e^{\pounds_\xi}\sqrt{-g}, \qquad (6.58)$$

with which one readily finds the result for $g^{\sigma\rho}$ and so on. Again the transformed field is just the exponentiated Lie derivative of the original.

Grishchuk, Petrov, and Popova have exhibited a straightforward relationship between finite gauge transformations (with the exponentiated Lie differentiation) and the tensor transformation law [99, 100]. Evidently the former is fundamental, the latter derived. One can define a vector field ξ^α using the fact that under a gauge transformation, $\mathfrak{g}^{\mu\nu}$ changes in accord with the tensor transformation law, while the flat metric stays fixed. Let us follow them and define ξ^α in terms of the finite coordinate transformation

$$x'^\alpha = e^{\xi^\mu \frac{\partial}{\partial x^\mu}} x^\alpha. \qquad (6.59)$$

Then the tensor transformation law, which is fairly easy to use, gives the finite gauge transformation formula, which is difficult to use. One can therefore hope to avoid using the latter at all, and work only with the tensor transformation law, which will tend to be simpler and will relate to known results in the geometrical theory.

In case the form for a finite gauge transformation for a tetrad field has not appeared previously, we provide one here. One can 'derive' this formula using a

Baker-Hausdorf-Campbell-type trick [348] by neglecting some of the non-commuting quantities, and then verifying that the resulting form, taken as a "lucky guess," has the desired properties. It is not difficult to verify that the following form preserves both the completeness relation to the inverse metric $g^{\mu\nu} = e^{\mu}_{A}\eta^{AB}e^{\nu}_{B}$ and the orthonormality relation $g_{\mu\nu}e^{\mu}_{A}e^{\nu}_{B} = \eta_{AB}$:

$$e^{\mu}_{A} + \delta e^{\mu}_{A} = e^{\pounds_{\xi}}(e^{\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)})^{C}_{A}e^{\mu}_{C}. \tag{6.60}$$

The expression involving $\pounds_{\xi}^{-1}$ should be understood as a shorthand for a series, this factor simply reducing the number of Lie differentiations by one. The outer set of parentheses indicates that the Lie differentiation in the second factor does not act outside the parentheses, although the Lie differentiation in the first factor acts on everything to its right. F is a matrix field which, when an index is moved using η_{AB} or η^{AB}, is antisymmetric: $F^{C}_{A} = -\eta_{AE}F^{E}_{B}\eta^{BC}$. The expression $\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)$ should be compared to some similar expressions in Lie group theory [344] (p. 179), [349] (p. 20), [350] (p. 80). This relation, it should be noted, is not tied to the presence of a flat background metric tensor or the interpretation of e^{μ}_{A} as generalized eigenvectors.

Let us now verify the completeness relation $g^{\mu\nu} = e^{\mu}_{A}\eta^{AB}e^{\nu}_{B}$, by showing that this relation with the gauge-transformed tetrad yields the gauge-transformed curved metric. The equations $\pounds_{\xi}\eta^{AB} = 0$ and the $(e^{\pounds_{\xi}}A)(e^{\pounds_{\xi}}B) = e^{\pounds_{\xi}}(AB)$ will be used in simplification. One has by definition of a variation Δ induced by this tetrad transformation,

$$\begin{aligned} g^{\mu\nu} + \Delta g^{\mu\nu} &= (e^{\mu}_{A} + \delta e^{\mu}_{A})\eta^{AB}(e^{\nu}_{B} + \delta e^{\nu}_{B}) \\ &= [e^{\pounds_{\xi}}(e^{\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)})^{C}_{A}e^{\mu}_{C}]\eta^{AB}e^{\pounds_{\xi}}(e^{\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)})^{E}_{B}e^{\nu}_{E} \\ &= e^{\pounds_{\xi}}[(e^{\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)})^{C}_{A}e^{\mu}_{C}\eta^{AB}(e^{\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)})^{E}_{B}e^{\nu}_{E}]. \end{aligned} \tag{6.61}$$

Acting with $e^{-\pounds_{\xi}}$ gives

$$e^{-\pounds_{\xi}}(g^{\mu\nu} + \Delta g^{\mu\nu}) = (e^{\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)})^{C}_{A}e^{\mu}_{C}\eta^{AB}(e^{\pounds_{\xi}^{-1}(F - e^{-\pounds_{\xi}}F)})^{E}_{B}e^{\nu}_{E}. \tag{6.62}$$

Let us define the matrix field $H^C_E = \pounds_\xi^{-1}(F - e^{-\pounds_\xi}F)^C_E$. Clearly H inherits near-antisymmetry from F: $H^C_E = -\eta_{EJ}H^J_B\eta^{BC}$. One then has

$$e^{-\pounds_\xi}(g^{\mu\nu} + \Delta g^{\mu\nu}) = e^\mu_C(e^H)^C_A\eta^{AB}(e^H)^E_B e^\nu_E$$
$$= e^\mu_C(e^H)^C_A\eta^{AB}(I^E_B + H^E_B + H^E_J H^J_B + \ldots)e^\nu_E, \tag{6.63}$$

where the one factor is expanded as a series. Continuing by moving the Lorentz metric into strategic locations gives

$$e^\mu_C(e^H)^C_A(I^A_P + \eta^{AB}H^E_B\eta_{EP} + \eta^{AB}H^J_B\eta_{JK}\eta^{KL}H^E_L\eta_{EP} + \ldots)e^{P\nu}$$
$$= e^\mu_C(e^H)^C_A(I^A_J - H^A_J + H^A_K H^K_J - \ldots)e^{J\nu}, \tag{6.64}$$

where the near-antisymmetry of H has been employed. Reverting to the exponential form gives

$$e^\mu_C(e^H)^C_A(e^{-H})^A_J e^{J\nu}$$
$$= e^\mu_C I^C_E e^{E\nu}$$
$$= g^{\mu\nu}, \tag{6.65}$$

leading to the expected conclusion $g^{\mu\nu} + \Delta g^{\mu\nu} = g^{\mu\nu} + \delta g^{\mu\nu} = e^{\pounds_\xi}g^{\mu\nu}$. Thus, completeness holds, and the tetrad-induced variation Δ of the inverse curved metric agrees with the gauge transformation variation δ. By similar maneuvers, one establishes the orthonormality relation for this tetrad variation:

$$(e^{\pounds_\xi}g_{\mu\nu})(e^\mu_A + \delta e^\mu_A)(e^\nu_B + \delta e^\nu_B) = \eta^{AB}. \tag{6.66}$$

Finally, the inverse tetrad transforms as

$$f^A_\mu + \delta f^A_\mu = e^{\pounds_\xi}(e^{\pounds_\xi^{-1}(-F + e^{-\pounds_\xi}F)})^A_C f^C_\mu, \tag{6.67}$$

which looks much like the tetrad form, save for the sign of F.

It turns out that the proofs of orthogonality and completeness depend only on the near-antisymmetry of the exponentiated matrix containing F, not on its detailed form, which in the expression above is rather involved. If the form above could be derived properly in the Baker-Hausdorf-Campbell fashion, perhaps it would be shown to be equivalent to $e^{(I\pounds_\xi + F)^C_A} e^\mu_C$. Such a form would seem natural because of its symmetrical treatment of the Lie and local Lorentz terms. The nonuniqueness of the tetrad gauge transformation form seems to imply that a given vector field ξ^μ and matrix F with one tetrad gauge transformation formula correspond to the same ξ^μ but a different F using a different transformation formula.

In these relations, we have retained the full local Lorentz freedom. But in the eigenvierbein formalism in the SRA, the vectors, in addition to being g-orthonormal, will be η-orthogonal: $\eta_{\mu\nu} e^\mu_A e^\nu_B = 0, A \neq B$. This orthogonality can be ensured using the term $\int d^4x \eta_{\mu\nu} e^\mu_A e^\nu_B \eta_{\alpha\rho} e^\alpha_C e^\rho_E \Lambda^{ABCE}$ described above. This set of six orthogonality equations will typically render the tetrad unique, and thus determine the local Lorentz transformation matrix F for a gauge transformation ξ^μ (although solving for F does not look easy). There might be some complication when there is degeneracy among the eigenvalues, as one expects in the Schwarzschild and Robertson-Walker field configurations (barring a perverse gauge choice). Degeneracy between the temporal eigenvalue and a spatial one would mean that the curved metric is arbitrarily close to violating causality, which we prefer to avoid by requiring stable η-causality. Given stable η-causality, the timelike eigenvector never suffers degeneracy-induced ambiguity, unlike the spacelike eigenvectors in some cases, such as those named.

6.11 Gauge Transformations Not a Group

If one is not interested in taking η-causality seriously, then any suitably smooth vector field will generate a gauge transformation. However, in the SRA, respecting η-causality–indeed, preferably stable η-causality–is essential. This fact entails that only a subset of all vector fields generates gauge transformations in the SRA.

Let us be more precise in defining gauge transformations in the SRA, requiring stable η-causality. A gauge transformation in the SRA is a mathematical transformation generated by a vector field in the form described above, but which also has both the untransformed and transformed curved metrics respect stable η-causality. It is evident that a vector field that generates a gauge transformation given one curved metric and a flat metric, might not generate a gauge transformation given another curved metric (and the same flat metric), because in the second case, the transformation might move the curved metric out of the stable η-causality-respecting configuration space, which is only a subset of the naive configuration space. It follows that one cannot identify gauge transformations with generating vector fields alone; rather, one must also specify the field configuration (curved metric) assumed prior to the transformation. For thoroughness, one can also use the flat metric (which is not transformed) as a label, to ensure that the trivial coordinate freedom is not confused with the physically significant gauge freedom. Let us therefore provisionally write a gauge transformation as an ordered triple

$$(\xi^{\mu}(x), \eta_{\mu\nu}(x), g_{\mu\nu}(x)), \tag{6.68}$$

where both $g_{\mu\nu}$(x) and $e^{\pounds_{\xi}} g_{\mu\nu}(x)$ satisfy stable causality with respect to $\eta_{\mu\nu}(x)$. The former restriction limits the configuration space for the curved metric, whereas the latter restricts the vector field. (At this point we drop the indices and the spacetime position argument for brevity.)

One wants to compose two gauge transformations to get a third gauge transformation. At this point, the fact that a gauge transformation is not labelled merely by the vector field, but also by the curved metric (and the flat), has important consequences. Clearly the two gauge transformations to be composed must have the second one start with the curved metric with which the first one stops. We also want the flat metrics to be compatible. Thus, the 'group' multiplication operation is defined only in certain cases, meaning the gauge transformations in the SRA *do* *not form a group*, despite the inheritance of the mathematical form of exponentiating the Lie differentiation operator from the field formulation's gauge transformation group. Two gauge transformations (ψ, η_2, g_2) and (ξ, η_1, g_1) can be composed to give a new gauge transformation $(\psi, \eta_2, g_2) \circ (\xi, \eta_1, g_1)$ only if $g_2 = e^{\pounds_\xi} g_1$ and $\eta_2 = \eta_1$. Because the multiplication operation is not always defined between two gauge transformations, the *closure* property of groups fails to hold.

The failure of closure implies modifications of associativity. Associativity does not hold in the usual way, because composition is not always meaningful. However, whenever the composition of SRA gauge transformations is meaningful, associativity holds, due to inheritance from the field formulation gauge group.

The failure of closure also modifies the existence of the identity, at least given the definition of a gauge transformation as an ordered triple. Whereas a group has a single element that acts as the identity on all elements and from either side, gauge SRA transformations do not have any single element that can be multiplied with all other elements, so in particular there is no identity element. There are, rather, many "little identity elements", which all have vanishing vector field, but which differ in their curved (or flat) metric labels. However, this complication can be removed if one gerrymanders the definition of a gauge transformation in the following way: let a gauge transformation be an ordered triple of the sort described above if and only if its vector field is nonvanishing, but let there also be a trivial transformation that maps any gauge transformation to itself, never mind any curved or flat metric labels. We therefore identify all transformations $(0, \eta_A, g_B)$ and write them as (0).

The inverse property is essentially untouched by the failure of closure. The left inverse of (ξ, η_1, g_1) is $(-\xi, \eta_1, e^{\pounds_\xi} g_1)$, yielding $(-\xi, \eta_1, e^{\pounds_\xi} g_1) \circ (\xi, \eta_1, g_1) = (0, \eta_1, g_1)$, an identity transformation. The right inverse is also $(-\xi, \eta_1, e^{\pounds_\xi} g_1)$, yielding $(\xi, \eta_1, g_1) \circ (-\xi, \eta_1, e^{\pounds_\xi} g_1) = (0, \eta_1, e^{\pounds_\xi} g_1)$, which is also an identity transformation, or rather, *the* identity transformation (0), given our gerrymandered definition of the identity. Thus, every SRA gauge transformation has a two-sided inverse.

In summary, of the group properties of closure, associativity, the existence of an identity element for all elements, and the existence of a two-sided inverse for each element [328], SRA gauge transformations, with the gerrymandered identity transformation, have an identity element for all elements, a two-sided inverse for each element, and associativity in those cases where multiplication is defined, but not closure, because multiplication is not defined for every ordered pair of elements.

6.12 Canonical Quantization in the SRA

The relevance of the special relativistic approach to Einstein's equations to canonical quantum gravity deserves some consideration. The primary patrons of the flat metric in the context of Einstein's equations have been the particle physicists in the context of the old covariant perturbation program of quantization. However, this program famously proved to be nonrenormalizable, even (probably) with the addition of carefully chosen matter fields in the later supergravity era [351]. Therefore, the covariant perturbation program has largely been abandoned.[9]

With the strongest advocates of the flat metric having diverted their attention to strings, membranes, and the like, one might form the belief that the use of a flat background metric has nothing further to contribute to quantum gravity, and in particular, to canonical quantum gravity. Isham writes of the null cone issue in the covariant perturbation program: "This very non-trivial problem is one of the reasons why the canonical approach to quantum gravity has been so popular." [351] (p. 12) And again, "One of the main aspirations of the canonical approach to quantum gravity has always been to build a formalism with no background spatial, or

[9] An exception is some recent work by G. Scharf and collaborators such as I. Schorn, N. Grillo, and M. Wellmann (for example, [113, 116]). Their use of "causal" methods helps to achieve finite results. Another possibility, suggested some time ago by Weinberg, is that 4-dimensionally nonperturbatively renormalizable. Recently O. Lauscher and M. Reuter have argued that this situation likely is realized [337].

spacetime, metric." [351] (p. 18) The use of a flat background in canonical gravity indeed seems to be rather rare, apart from some work in the field formulation by Grishchuk and Petrov [100], which does not consider the flat metric's null cone. (We do not see any reason that the *formal* use of a flat background in the field formulation requires one to give up the spatial metric components as the canonical coordinates, as they do.)

However, it would be a mistake to conclude that the canonical formalism is immune to similar worries, worries which a flat background's null cone structure could address. Isham continues:

> However, a causal problem arises here [in the canonical approach] too. For example, in the Wheeler-DeWitt approach, the configuration variable of the system is the Riemannian metric $q_{ab}(x)$ on a three-manifold Σ, and the canonical commutation relations invariably include the set
>
> $$[\hat{q}_{ab}(x), \hat{q}_{cd}(x')] = 0 \tag{6.69}$$
>
> for all points x and x' in Σ. In normal canonical quantum field theory such a relation arises because Σ is a space-like subset of spacetime, and hence the fields at x and $x\prime$ should be simultaneously measurable. But how can such a relation be justified in a theory that has no fixed causal structure? The problem is rarely mentioned but it means that, in this respect, the canonical approach to quantum gravity is no better than the covariant one. It is another aspect of the 'problem of time' .... [351] (p. 12)

Evidently introducing a flat metric can help:

> The background metric η provides a fixed causal structure with the usual family of Lorentzian inertial frames. Thus, at this level, there is no problem of time. The causal structure also allows a notion of microcausality,

> thereby permitting a conventional type of relativistic quantum field theory . . . It is clear that many of the *prima facie* issues discussed earlier are resolved in an approach of this type by virtue of its heavy use of background structure. [351] (p. 17)

What then is the difficulty?

> However, many classical relativists object violently . . . , not least because the background causal structure cannot generally be identified with the physical one. Also, one is restricted to a specific background topology, and so a scheme of this type is not well adapted for addressing many of the most interesting questions in quantum gravity: black hole phenomena, quantum cosmology, phase changes *etc.* [351] (p. 17)

However, above we have presented a formalism which, if adopted, plausibly *does* ensure that the physical causal structure is consistent with the background one by construction. Thus, this first objection is largely answered. The second objection is strong only if one already knows that gravitation is geometrical at the classical level. But there is no necessity in taking such a view, not least because it is so easy to derive Einstein's equations for the geometry of an effectively curved spacetime within truly flat spacetime, as we saw above. We conclude that it would be interesting to investigate the canonical quantization of Einstein's equations within the special relativistic approach, because serious conceptual problems with standard approaches would evidently be resolved, whereas no serious problems would be generated, at least at the conceptual level.

On the other hand, the special relativistic approach might perhaps complicate certain technical issues in canonical quantum gravity. First, as we saw above, the use of the η-causality variables suggests that the primary constraints of the theory would be non-trivialized. Second, the fact that some vector fields do not generate gauge transformations in the SRA suggests that the custom of splitting spatial and temporal diffeomorphisms and treating them independently might be threatened. In the field formulation, with merely formal use of the flat metric, could at least make sense of spatial gauge transformations as those generated by solutions ξ^{α} of the equation $x'^{\alpha} = e^{\xi^{\mu} \frac{\partial}{\partial x^{\mu}}} x^{\alpha}$ given the restriction that $x'^{0} = x^{0}$. But it is not clear that such a separation is even possible if the flat metric's null cone structure is to be respected, as in the SRA. The reason is that whether a vector field generates a gauge transformation in the SRA depends on *both* the temporal and spatial parts of the vector field, because both influence the curved null cone.

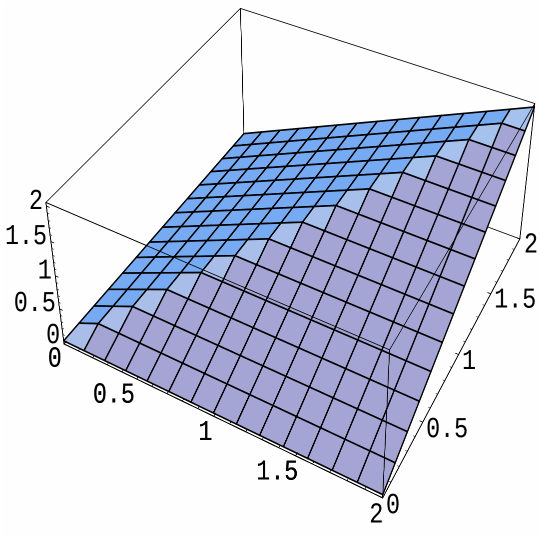

Figure 6.1: Bounding Surface for Temporal Eigenvalue as Function of Spatial Eigenvalues in 2 Dimensions

Chapter 7

The Schwarzschild Radius in the SRA

While everyone knows that the Schwarzschild radius harbors merely a pseudosingularity in geometrical general relativity, it is not so obvious what happens in the SRA. One reason is that the gauge freedom is restricted by the need to secure null cone consistency. Another is that there is no way of extending Minkowski spacetime by adding points 'past infinity', because the flat metric ensures that one knows in advance that one's naive coordinate system (such as Cartesian or spherical polar) covers the whole spacetime adequately. (For our non-geometrical purposes, it is permissible for a coordinate system not to cover regions of measure 0 well.)

7.1 Eddington-Finkelstein and Painlevé-Gullstrand Coordinates and Gauge Fixing

Here we will make two efforts to use the gauge freedom of the SRA to remove the apparent singularity in the curved metric, one based on Eddington-Finkelstein coordinates, and one based on Painlevé-Gullstrand coordinates, both of which introduce apparent time-reversal variance. Both of them will violate η-causality and thus fail to remove the Schwarzschild singularity in the SRA. In fact, they violate causality *everywhere* outside the Schwarzschild radius. While the failure of these two plausible efforts does not prove that all such efforts would fail, it does suggest that the usual geometrical approach to the Schwarzschild 'singularity' cannot be readily carried into the SRA.

Let us turn to the Eddington-Finkelstein form of the Schwarzschild metric. Our treatment will be based on Penrose's discussion of the subject in the context of the geometrical theory [352], apart from a change of signature to $-+++$. In coordinates $x^{\mu} = (t, r, \theta, \phi)$, the curved line element takes the form

$$ds^2 = -(1 - \frac{2m}{r})dt^2 + (1 - \frac{2m}{r})^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{7.1}$$

Let us choose the flat metric to have the standard line element

$$d\sigma^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{7.2}$$

Having specified both metrics, we have fixed the gauge.

Supposing that one makes a transformation to the Eddington-Finkelstein coordinates $x^{\mu'} = (v, r, \theta, \phi)$, where $v = t + r + 2m\ln(r - 2m)$, it is simple to use the tensor transformation law to find the metrics in the new coordinates. One finds the curved line element to take the form

$$ds^2 = -(1 - \frac{2m}{r})dv^2 + 2dvdr + 0dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{7.3}$$

which famously is nonsingular at $r = 2m$. In these coordinates, the flat line element takes the form

$$d\sigma^2 = -dv^2 + \frac{2r}{(r-2m)}dvdr + \frac{4mr - 4m^2}{-r^2 + 4mr - 4m^2}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{7.4}$$

Letting r approach $2m$ from without, one finds the flat line element to take the form

$$d\sigma^2 = -dv^2 + \infty dvdr - \infty dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{7.5}$$

which is of course singular. This is just as expected: any coordinate transformation that removes the Schwarzschild singularity from the curved metric must be singular [353], and thus that transformation in a bimetric context merely throws the singularity onto the flat background metric [115]. In our terminology, this coordinate transformation avoids detecting the singularity only by adopting a set of physical rods and clocks that are *infinitely* distorted relative to metaphysical rods and clocks, which is to say, broken.

To make a nontrivial effort to remove the singularity, we must make a *gauge* transformation, which looks like a coordinate transformation on the curved metric and the (bosonic) matter fields, but does not touch the flat metric. One can consider an "Eddington-Finkelstein gauge" in which the curved metric ends in the form

$$ds^2 = -(1 - \frac{2m}{r})dt^2 + 2dtdr + 0dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{7.6}$$

which is the same form as the Eddington-Finkelstein metric above, save that the letter v has been replaced by the letter t. The flat metric is still

$$d\sigma^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{7.7}$$

and the coordinates are (t, r, θ, ϕ). We have thus found a gauge that leaves both the curved and flat metrics nonsingular.

Have we therefore eliminated the singularity? We have not if the result violates η-causality. We must therefore solve the generalized eigenvalue problem for the eigenvalues. As we saw above, every field configuration obeying the causality principle has all real eigenvalues and at least 3 independent eigenvectors. The characteristic polynomial easily yields the solutions

$$\Lambda = 1, \Lambda = 1, \Lambda = \frac{1 - \frac{2m}{r} \pm \sqrt{(\frac{r-2m}{r})^2 - 4}}{2}. \tag{7.8}$$

The eigenvalues must all be real for there to be any hope of satisfying causality, but is plain that this curved metric in fact *violates* the causality principle everywhere outside the Schwarzschild radius, and even somewhat within it. Because causality is violated at least somewhere, this Eddington-Finkelstein gauge is forbidden in the SRA. Thus, the singularity has not been shown to be removable.

One could consider instead the Painlevé-Gullstrand coordinate system [354, 355], another coordinate system that renders the curved metric nonsingular at $r = 2m$. In coordinates (T, r, θ, ϕ), the Schwarzschild solution takes the form

$$ds^2 = -(1 - \frac{2m}{r})dT^2 + dr^2 + 2\sqrt{\frac{2m}{r}}dTdr + r^2(d\theta^2 + \sin^2\theta d\phi^2), \qquad (7.9)$$

which is also nonsingular at the Schwarzschild radius. This coordinate T can be expressed in terms of r and t *via*

$$dT = dt + \sqrt{2mr}\frac{dr}{r - 2m}, \qquad (7.10)$$

which is of course singular at the Schwarzschild radius. (We do not need the integrated form, though it is known [355].) The flat metric, assumed to have the form $d\sigma^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)$ in the (t, r, θ, ϕ) coordinates, naturally is singular at $r = 2m$ in the (T, r, θ, ϕ) system.

Making a gauge transformation such that the curved metric takes the Painlevé-Gullstrand form $ds^2 = -(1 - \frac{2m}{r})dt^2 + dr^2 + 2\sqrt{\frac{2m}{r}}dtdr + r^2(d\theta^2 + \sin^2\theta d\phi^2)$, but the flat metric still takes the form $d\sigma^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)$, one finds both metrics to be nonsingular. But η-causality must still be considered. Finding the generalized eigenvalues gives

$$\Lambda = 1, \Lambda = 1, \Lambda = 1 - \frac{M}{r} \mp \sqrt{-\frac{2m}{r} + \frac{M^2}{r^2}}. \qquad (7.11)$$

Reality of the eigenvalues, a necessary condition for good causal properties, is violated unless $M \geq 2r$, which is to say, everywhere outside the Schwarzschild radius and part of the region inside it. Thus, the Painlevé-Gullstrand gauge transformation also fails to remove the Schwarzschild singularity in the SRA.

While the failure of these two efforts to remove the singularity while respecting η-causality does not imply that it cannot be removed, such failure at least suggests that the *prima facie* staticity of the solution–its independence of t and t-reversal invariance–might possibly be genuine in the SRA.

7.2 Schwarzschild Problem in Tensorial DeDonder Gauge

In view of the status of the (tensorial) DeDonder gauge as the prince of gauge conditions (apart from our argument that plane waves show it SRA-inadmissible) and its role in the RTG, it is interesting to consider the Schwarzschild radius in the t-independent Schwarzschild problem. This was done some time ago by F. J. Belinfante [22] along with J. C. Garrison [27] in the field formulation, and more recently by Yu. P. Viblyi [270] in the context of the RTG.

Belinfante finds a constant of integration besides the mass in the solution, but argues that it must vanish to give physically acceptable behavior, and obtains the form also given by V. Fock [223]. In this form, expressed in the naive set of spherical coordinates on flat spacetime, the horizon occurs at $r = M$. Belinfante and Garrison, contemplating the uniqueness of Fock's harmonic coordinates, again find that this constant should vanish, and find even stronger grounds than Belinfante had earlier.

For the RTG, however, Viblyi reports that this constant of integration is physically significant and that the solutions with different values are physically distinct: "In the RTG and general relativity the equations are specified in general on different ranges of variation of the same variables, and therefore the number of harmonic solutions of general relativity that have physical meaning need not be equal to the number of corresponding solutions of the RTG. Thus, the unique harmonic solution of general relativity for a static centrally symmetric field (Fock's solution) is only a particular solution of the RTG equations, namely, a particular solution of" the tensorial DeDonder condition [270]. Viblyi finds that all values of this constant of integration imply that sources of finite size must be considered. The constant is determined by fitting to some interior solution. Thus, the external field depends on the structure of the source, unlike both Newtonian gravity and the geometrical theory [270]. One should note that the second minus sign in Viblyi's equation (7) should be a plus sign (*cf.* [22]).

At one time A. A. Vlasov and A. A. Logunov argued that the RTG excludes black holes because a particle reaches the horizon only at infinite metaphysical time (if the reader will pardon our ascribing our term "metaphysical" to them) [303]. Later it happened that a nonstatic Schwarzschild form that satisfies the tensorial DeDonder condition (almost everywhere [98]) was discovered by Yu. M. Loskutov [98], Vlasov [96] and A. N. Petrov [97]. This form makes the solution nonstatic and makes the curved metric nonsingular at the horizon. Vlasov took this solution to show that his earlier conclusions with Logunov, which had claimed to show that black holes can never exist in the RTG, were ill-founded, due to failure to consider all possible cases. He writes:

> Therefore, in the absence within the RTG ideology of any statement about the uniqueness of the spherically symmetric solution (including the case of a rotating charged source) the solution of the collapse problem in the RTG essentially depends on the form of the spherically symmetric solution used, i.e., on both the model of the matter considered and on the initial, boundary, asymptotic, etc., conditions specified in the original Minkowski-space coordinates. Here in the RTG it is possible to have solutions which do not have singularities on spherical surfaces and do not violate the requirements on the [signatures of the curved and flat metrics]. Therefore, in the RTG it is impossible to speak of the absence of catastrophic collapse and the appearance of singularities on the cor-

responding spherical surfaces, and also to draw conclusions solely on the basis of the simplest form (in the Minkowski-space coordinates) of the spherically symmetric solutions. [96]

The phrase "In the absence ..." leaves open the possibility that the original conclusion would eventually be verified (which Loskutov argued has occurred [98]). Petrov drew a similar conclusion to Vlasov's, but stronger, that "the Einstein equations with the supplementary harmonic conditions, and the structure of a flat background space-time, do not exclude black holes" [97] (p. 196).

Let us consider this new coordinate system, which, when mated with a Minkowski metric of the usual form, yields a new gauge.

For reference, we recall that the usual static harmonic form is Fock's (and Belinfante's)

$$ds^2 = -\frac{r-M}{r+M}dt^2 + \frac{r+M}{r-M}dr^2 + (r+a)^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{7.12}$$

where $c = G = 1$, though it is in fact the Cartesian-like coordinates $x = r\sin\theta\cos\phi$, *etc.* that are harmonic [97]. This metric obeys the tensorial DeDonder condition in terms of the usual spherical coordinates in which a flat metric $\eta_{\mu\nu}$ is given by

$$d\sigma^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{7.13}$$

The new Loskutov-Vlasov-Petrov time-reversal-*variant* form of $g_{\mu\nu}$, nonsingular at the horizon, is

$$\begin{aligned} ds^2 = -\frac{r-M}{r+M}d\tau^2 + \frac{8M^2}{(r+M)^2}d\tau dr + \left(1 + \frac{2M}{r+M} + (\frac{2M}{r+M})^2 + (\frac{2M}{r+M})^3\right) dr^2 \\ +(r+a)^2(d\theta^2 + \sin^2\theta d\phi^2), \end{aligned}$$

along with the flat metric of the usual form

$$d\sigma^2 = -d\tau^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{7.14}$$

Given the assumption of the fundamental nature of flat spacetime, the coordinate r ranges over $[0, \infty)$. (The fact that the central singularity $r = -M$ is excluded will not matter if our doubts about the behavior of this gauge at the Schwarzschild radius are borne out.)

Given our experience with the Eddington-Finkelstein and Painlevé-Gullstrand gauges, our first order of business is to see if this gauge respects η-causality. In fact, one easily sees that the radial g-null geodesics [97] (p. 198)

$$\frac{d\tau}{dr} = \frac{r^2 + 2Mr + 5M^2}{r^2 - M^2}, \tag{7.15}$$

$$\frac{d\tau}{dr} = -\frac{r + 3M}{r + M} \tag{7.16}$$

satisfy η-causality. It is not difficult to show that the two non-angular eigenvalues are real and positive, and (with the help of *Mathematica*) that the larger one has a spacelike eigenvector, while the smaller has a timelike one. Thus, this non-static form of the Schwarzschild solution does satisfy η-causality, unlike the efforts above.[1]

One might worry, in sympathy with Rosen [359], that it seems unreasonable to destroy the staticity of the problem to remove the Schwarzschild singularity, because the physics gives one no reason at all to doubt staticity. This new solution is merely stationary, which is to say, constant in time, but not time-reversal invariant. Petrov observes that his system, which is defined on the coordinate ranges $r \in [0, \infty), \tau \in (-\infty, \infty)$, only covers half the Schwarzschild geometry (p. 198), so presumably the other half that he envisions would restore the time-reversal invariance. (However, from the perspective of the SRA (including the RTG), the region covered by $r \in [0, \infty), \tau \in (-\infty, \infty)$, with flat metric $d\sigma^2 = -d\tau^2 dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)$, just is the whole of Minkowski spacetime,

[1] One should note that the curved metric makes sense down to $r = -M$, so the limitation to $r \geq 0$ is made only to accomodate flat spacetime. If one does not require the tensorial DeDonder condition, then it is easy to choose a new flat metric $\eta'_{\mu\nu}$ with line element $d\sigma'^2 = -d\tau^2 + dr^2 + (r + M)^2(d\theta^2 + \sin^2\theta d\phi^2)$ which lets r have the range $r \geq -M$, and η-causality is respected over the entire range $r > -M$. The true singularity lying at $r = -M$ [97], this altered form evidently is a better candidate for satisfying the SRA analog of g-geodesic completeness than is Petrov's form.

and thus does not admit augmentation.) This Rosen-like objection, though it might count against the new form as the field of a point mass sitting sempiternally at the origin, will not prevent this form from arising in the context of a gravitational collapse process. Thus, for all that has been said so far, this new solution might in fact show that catastrophic collapse is possible in the SRA. At a minimum, it shows that spherically symmetric strong-gravity solutions with regions of no escape exist in the SRA. One can express Petrov's solution using the η-causality variables exhibited above.

However, it seems to us that some doubt attaches to the mathematics of Petrov's gauge transformation relating the static form to his nonstatic form. We recall the work above [99, 100] relating finite gauge transformations (with the exponentiated Lie differentiation) to the tensor transformation law. If the exponentiated form is fundamental, then results derived using the tensor transformation formula are valid only if there exists the relevant vector field ξ^{α} that generates the gauge transformation. The coordinate transformation that Petrov uses to get the new form of the curved metric is given on p. 204 as

$$\tau = t + 2M \ln \left| \frac{r-M}{r+M} \right|, r = r. \tag{7.17}$$

For large r, this coordinate transformation is small, but it grows as r shrinks. In fact, it diverges at $r = M$, which is the horizon in the static form. Without that divergence, this coordinate transformation would fail to bring the point $t = \infty$ to a finite value of τ, and his argument against Vlasov, Logunov and Mestvirishvili would collapse. But given the formula $x'^{\alpha} = e^{\xi^{\mu} \frac{\partial}{\partial x^{\mu}}} x^{\alpha}$ that defines the vector field generating the gauge transformation, it is not obvious that this coordinate transformation is allowed, because there must exist a reasonable solution ξ^{μ} defined on $r \in [0, \infty), t \in (-\infty, \infty)$ to the equations

$$t + 2M \ln \left| \frac{r-M}{r+M} \right| = e^{\xi^{\mu} \frac{\partial}{\partial x^{\mu}}} t, r = e^{\xi^{\mu} \frac{\partial}{\partial x^{\mu}}} r, \theta = e^{\xi^{\mu} \frac{\partial}{\partial x^{\mu}}} \theta, \phi = e^{\xi^{\mu} \frac{\partial}{\partial x^{\mu}}} \phi, \tag{7.18}$$

even with the divergence of the first equation at $r = M$. If this doubt proves warranted, then Petrov's gauge transformation will be unacceptable. Without this gauge transformation, the nonstatic solution will in fact be a physically distinct solution from the static one. Which solution is more relevant to gravitational collapse is unclear.

Loskutov argues that the problem can only be studied successfully when the interior is considered [98]. He asserts that the vacuum solution can only hold for $r > M$, lest the Logunov causality principle be violated. (However, as we have seen, this procedure of accepting or rejecting solutions "by hand" is questionable.) Loskutov then argues, using the gauge condition and the need to match the exterior solution to an interior solution at some $r > M$ (supposedly shown using the Logunov causality principle), that if the curved metric is stationary,[2] then the shift vector g_{0i} vanishes, which vanishing excludes this nonstatic form of the Schwarzschild metric. Loskutov therefore reaffirms the original conclusion of Vlasov and Logunov that catastrophic collapse is impossible because the Schwarzschild radius is unreachable.

It seems that our worries about the Logunov causality principle propagate into Loskutov's argument. Earlier we were not persuaded by Petrov's argument that the Schwarzschild is reachable, either. It would appear that a complete resolution of this issue might require a more adequate procedure for respecting η-causality while ensuring a "complete" set of solutions of the field equations. In any case, the fact that the SRA implies global hyperbolicity constrains the properties of the region of no escape in this nonstatic Schwarzschild solution.

[2] The translated paper says "static," but this cannot have been a proper use of "static" as a technical term, or else the problem is trivialized.

7.3 Existence of Schwarzschild 'Singularity' and Interior Dependent on Constant of Integration in Covariantly Unimodular Gauge

We turn now to study an interesting property of the static Schwarzschild solution in the covariantly unimodular gauge $\frac{\sqrt{-g}}{\sqrt{-\eta}}$. It might be the case that the Schwarzschild singularity is truly singular in the SRA. However, as we will find, it need not be a singular *sphere* at a finite radius, with an inscrutable interior. In fact, it plausibly need not *have* an interior at all.

We review and build upon the work of Ya. I. Pugachev and V. D. Gun'ko [101, 356, 357], who found that one can eliminate the interior of the Schwarzschild radius entirely by a suitable choice of a constant of integration, at least in the covariantly unimodular gauge $\frac{\sqrt{-g}}{\sqrt{-\eta}} = 1$. Their results are sufficiently important, at least given our SRA reinterpretation, and yet sufficiently unknown, that a recapitulation will be worthwhile. We will alter their signature to $- + ++$. Eventually our approach will diverge slightly from theirs.

Using the usual spherical coordinates for flat spacetime to give

$$d\sigma^2 = -dt^2 + dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{7.19}$$

for the metaphysical interval, Pugachev and Gun'ko consider the physical interval of the form

$$ds^2 = -e^{\nu(r)}dt^2 + e^{\lambda(r)}dr^2 + e^{\mu(r)}r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{7.20}$$

where the very general form of the angular metric is the most interesting part of this equation. Given this metric, the field equations take the following form (which we have verified using GRTensor II, but have corrected a typographical error in their paper):

$$\nu'' - \lambda'\nu' + \frac{\nu'}{2}(\nu' + \lambda' + 2\mu') + \frac{2\nu'}{r} = 0, \tag{7.21}$$

$$2\mu'' + \nu'' + \frac{4\mu'}{r} - \frac{\lambda'}{2}(\nu' + \lambda' + 2\mu') + \frac{1}{2}(\nu'^2 + \lambda'^2 + 2\mu'^2) = 0, \tag{7.22}$$

$$e^{\mu-\lambda}\left[1 + r(\mu' - \lambda') + \frac{r^2\mu''}{2} + \frac{r^2\mu'(\mu' - \lambda')}{2} + \frac{1}{2}(1 + \frac{r\mu'}{2})(\nu' + \lambda' + 2\mu') - \frac{r^2\mu'^2}{2}\right] = 1, \tag{7.23}$$

where a prime denotes differentiation with respect to r. They find these equations to be solved by the two relations

$$e^{\frac{\lambda+\nu}{2}} = (re^{\frac{\mu}{2}})', e^{\nu} = 1 + \frac{c_1}{r}e^{-\frac{\mu}{2}}. \tag{7.24}$$

Hoping to find a unique and physically meaningful result, they impose the gauge condition $\Delta^{\mu}_{\mu\alpha} = 0$, which implies that $\frac{\sqrt{-g}}{\sqrt{-\eta}}$ is a constant everywhere and always.[3] One naturally chooses the constant to be unity for give the expected properties at large r. Given the symmetries, only the r condition is nontrivial, leaving $\lambda' + \nu' + 2\mu' = 0$, which explains why such expressions were isolated in the field equations. They manage to solve the field equations in this gauge to get

$$e^{\nu} = -1 + \frac{c_1}{(r^3 + c_2)^{\frac{1}{3}}}, \tag{7.25}$$

$$e^{\lambda} = \frac{r^4}{(r^3 + c_2)[c_1 + (r^3 + c_2)^{\frac{1}{3}}]}, \tag{7.26}$$

$$e^{\mu} = (1 + \frac{c_2}{r^3})^{\frac{2}{3}}, \tag{7.27}$$

which GRTensor II confirms to have vanishing Ricci tensor.

The constants of integration c_1 and c_2 must now be determined. The Newtonian limit for large r quickly yields $c_1 = -2GM$. For c_2, we part with their work for the moment, though not greatly. It is interesting to see if one can choose c_2 in such a way that the solution is well-behaved all the way to $r = 0$. Given our interest in the flat metric's null cone, good behavior for us will imply satisfying the

[3]We cannot accept their claim that this is a Lorentz-like gauge, however.

causality principle, too. To keep all metric components well-behaved, clearly one needs $c_2 \geq 0$. One also needs $r^3 + c_2 \geq 8G^3M^3$; making this hold down to $r = 0$ gives $c_2 \geq 8G^3M^3$. What does the causality principle demand? All eigenvalues must be positive, and the timelike one must be at least as small as the others. Given the assumptions so far about c_2, the timelike eigenvalue is less than unity everywhere, whereas the spacelike ones are greater than unity everywhere. Thus, any value of c_2 such that $c_2 \geq 8G^3M^3$ eliminates the usual behavior of vanishing lapse, diverging radial metric, *etc.*, in favor of good behavior all the way down to (but not including) $r = 0$, as in electromagnetism and Newtonian gravity for a point source. (The typical form of the Schwarzschild metric is given by $c_2 = 0$.) If with Pugachev and Gun'ko we require that the source be characterized by only one parameter while retaining good behavior (a requirement that makes sense in the absence of sources), then the value $c_2 = 8G^3M^3$ is uniquely selected. With this choice, one finds some potentials diverging as $r \to 0$, which is just what Newtonian and Maxwellian intuitions lead one to expect. In that case, the Schwarzschild singularity corresponds to $r = 0$, merely a point, but the interior corresponds to nothing at all. Below we will recall some comparable results of others.

What these results show is that, given this attractive partial gauge fixing, the event horizon can be created or destroyed by the choice of the parameter c_2. It should be pointed out that the work is done by gauge-fixing in terms of the *radius*, not the time, as in the cases above. It should also be observed that adjusting the gauge fixing is vastly more natural in this context than is adjusting the temporal gauge. How so? The Schwarzschild problem is well-behaved for all t and static in t in the most obvious form, so there is nothing to suggest adjusting t; it is at some values of r, rather, that the solution appears to go bad, so adjusting the radial field seems like the natural approach, in sympathy with Rosen's intuitions.

In addition to the integration-constant dependence of the existence of the

Schwarzschild 'singularity', Pugachev and Gun'ko report that other gauge choices lead to physically distinct behaviors. It seems odd that physical behavior can be affected by the choice of gauge, so something must give.

The usual geometric reply, provided in this case by M. E. Gertsenshtein and M. Yu. Konstantinov [356], is that the above analysis fails to cover the whole spacetime. Why? It is largely because it fails to recognize that several coordinate systems are needed to cover the full solution, fails to reckon with its topology, and fails to extend the g geodesics until either they attain infinite g-affine length or they hit a singularity. If we were interested in the geometric approach to gravitation, these criticisms would be relevant, but they are irrelevant to the SRA.[4] Given that the spacetime is just Minkowski's, we know in advance that it has trivial topology, that it can be covered in single coordinate system, and that curves are complete if they have infinite *metaphysical* length. We can therefore discount this geometric reply, at least in letter.[5]

However, there might be some relevance to the spirit of their reply, because we must still reckon with the gauge-variance of physical results. One approach, taken by Pugachev and Gun'ko, is to argue that the bimetric formalism requires some sort of gauge fixing. The need to fix the gauge has been argued by Rosen [6], Papapetrou [11], Logunov and Vlasov (for example, [105]), and H. Nikolić [108]. It seems quite possible that different gauge fixings would give inequivalent theories.[6] Another

[4]Pugachev and Gun'ko did not make any clear statement of commitment to the flat spacetime approach, but perhaps believed themselves merely to be studying Einstein's theory and using the flat metric as a convenient tool to find physically reasonable solutions. Thus, it is understandable that such objections as these by Gertsenshtein and Konstantinov would arise to their work, though not to ours.

[5]The work of Z. Zakir [358] is somewhat similar to that of Pugachev and Gun'ko in finding the existence of the Schwarzschild singularity to depend on an additional constant. However, as Zakir works within a geometrical approach to gravitation, the objection that his solution is not geodesically complete seems unavoidable. We thank Dr. L. Lehner for comments on this subject.

[6]This outcome would perhaps suggests that the terms "gauge" and "gauge fixing" are inappropriate, but we will continue to use them until these issues have been sorted out. Fixing the gauge using ineffective constraints imposed by Lagrange multipliers seems like a natural approach. For the Logunov-Vlasov tensorial DeDonder gauge fixing, a term such as $\int d^4x \Lambda_{\mu\nu} \Delta^{\mu}_{\alpha\rho} g^{\alpha\rho} \Delta^{\nu}_{\beta\sigma} g^{\beta\sigma}$, with

option might be our proposal that admissible gauges must 'extend' the solution as far as possible, rather than gratuitously omit parts of it, but still respect η-causality. Thus, a good gauge fixing would have the vacuum Schwarzschild solution have a potential that diverges as $\frac{1}{r}$ at the center, rather than diverging at finite r or not diverging at all.

Given this charming property of the covariantly unimodular gauge, one could ask how this gauge condition relates to the null cone issue. Does this gauge supply all of our desiderata? We recall that this condition (or the unimodular condition $\sqrt{-g}$) has been coupled with $\frac{3}{4}$ of the tensorial (or ordinary) DeDonder condition from time to time as an attractive gauge fixing [6, 233], so it is already known to have some appealing properties. However, it looks as if the condition of covariant unimodularity typically fails to ensure null cone consistency, for it implies that if the determinant of the spatial metric (say, in Cartesian coordinates) falls below unity, then the lapse increases above unity. Such a situation might arise in a Robertson-Walker cosmological context, where it would imply η-causality violation. So the ultimate significance of these results about removing the Schwarzschild radius and its interior by choosing a constant of integration, in light of the unacceptability of this gauge condition, remains unclear. Clearly the issue is subtle. However, one might suggest that any good gauge *ought* to behave as the covariantly unimodular gauge behaves at the Schwarzschild radius, so such behavior might be a criterion for finding a good gauge.

7.4 The Schwarzschild Radius as a Point?

For a well-motivated value of the constant of integration in the covariantly unimodular gauge, we have seen that the Schwarzschild singularity can be merely the point $r = 0$. In fact, other authors, such as L. Bel [360] and A. I. Janis, E. T. Newman, and

Lagrange multipliers $\Lambda_{\mu\nu}$, might be suitable.

J. Winicour [361] and Rosen [359] discussed that result somewhat earlier apart from any bimetric formalism. (See also ([362]).) It would seem that, from a flat spacetime point of view, this is the most desirable result, because it satisfies η-causality, conforms to intuitions about point sources from Newtonian gravity and electromagnetism, manifests all the expected symmetries (including staticity), and 'extends' the Schwarzschild solution as far as possible without hitting a singularity–for surely η-causality violation constitutes a singularity in the SRA.

Rosen, commenting on the work of Janis, Newman, and Winicour approvingly, finds that "the interior of the Schwarzschild singularity has been excluded from the physical space." [359] (p. 233) He continues:

> There seem to be two ways of dealing with this non-physical space. One can simply exclude it from the physical space by requiring that the boundary given by $r = 2m$ be an impassible barrier for light or matter, or one can go over to a coordinate system which has been chosen so as to exclude the non-physical region without leaving any boundary. [359] (p. 233)

This latter approach will be employed below.

7.5 Finding a Relation Between Two Coordinate Systems to Fix the Gauge

Next we outline a procedure for implicit partial gauge-fixing *via* the use of two coordinate systems, one in which an interesting solution of Einstein's equations is given, and one which is known to cover the entire Minkowski spacetime and to give the Minkowski metric a simple known form. Imposing η-causality yields a highly nonlinear partial differential equation relating the two coordinate systems. By making use of the symmetries of the problem, one can hope to reduce the partial differential equation to a manageable form, so that some relationship between the two coordinate systems can be found such that the gravitational potential is well-behaved. Good behavior includes satisfying η-causality, of course. Probably there will be many satisfactory answers if there is one, because η-causality only imposes inequalities, not equations, on the eigenvalues. We perform this procedure for the Schwarzschild solution and find a simple solution, which illustrates in a different way how it is that the interior of the Schwarzschild radius is excluded by the relation between the two radial coordinates. It also becomes clear how one could make other choices, which seem physically less desirable, that exclude even part of the exterior of the Schwarzschild solution.

Given the matrix of coordinate basis components of a curved metric in some coordinate system, one can look for a coordinate transformation between the coordinates x^μ used and some simple coordinate system (perhaps Cartesian; here a ranges from 0 to 3) y^a that covers all of spacetime, gives the flat metric a simple form, and respects η-causality. Thus, we want to solve the generalized eigenvalue problem with characteristic polynomial

$$det(\Lambda \eta_{ab} - g_{\mu\nu} \frac{\partial x^\mu}{\partial y^a} \frac{\partial x^\nu}{\partial y^b}) = 0, \tag{7.28}$$

and then impose certain conditions on the eigenvalue roots.

Unfortunately, this equation looks rather difficult to solve in general. However, in the case of the Schwarzschild solution, we have spherical symmetry and evidently staticity as well. It seems reasonable to investigate the case with $y^0 = x^0 = t, y^2 = x^2 = \theta, y^3 = x^3 = \phi$, while the relationship between $y^1 = R$ and $x^1 = r$ is to be determined.

The characteristic polynomial yields the eigenvalues

$$\Lambda_0 = 1 - \frac{2m}{r}, \tag{7.29}$$

$$\Lambda_1 = (1 - \frac{2m}{r})^{-1}, \tag{7.30}$$

$$\Lambda_2 = \Lambda_3 = \frac{r^2}{R^2}. \tag{7.31}$$

We of course assume that $M \geq 0$. It is worth noting that the range of the coordinate r is not known *a priori.* While there are presumably many solutions $r(R)$ of this equation–the work above with the covariant unimodularity condition is another example–we choose for simplicity to consider solutions of the form $r = R + C$. It is clear that setting $C \geq 2M$ is necessary for η-respecting behavior of the eigenvalues, but choosing $C > 2M$, though also respecting η-causality, would gratuitously exclude part of the Schwarzschild solution. We therefore take $r = R+2M$. The system $y^a = (t, R, \theta, \phi)$ is just the usual spherical coordinate system on flat spacetime, so R ranges over $[0, \infty)$, implying that r ranges over $[2M, \infty)$. Thus, the value $r = 2M$ of the 'Schwarzschild singularity' is actually merely the point source at the origin $R = 0$, and the interior $r < 2M$ just does not exist, as was suggested above.

7.6 Gravitational Collapse and the Black Hole Information Loss Paradox

As we have seen, the SRA, which requires that the flat null cone be respected, implies that all SRA spacetimes are globally hyperbolic. While Petrov's nonstatic Schwarzschild solution shows that regions of no escape are permitted within the SRA, some of the more exotic properties of the black holes present in the geometrical formulation will not appear in the SRA.

It has been argued that global hyperbolicity renders the Hawking black hole information loss paradox innocuous [293]. That is because "the spacetime can be foliated by a family of Cauchy surfaces in such a way that there is no time at which the Lemma [proved therein] forces the state of the universe to be mixed" [293] (p. 204). Thus, there is no evolution from a pure state to a mixed state, and so no information is lost. If global hyperbolicity solves the problem, then so does the

SRA. It would be pleasant if so small an investment as accepting a special relativistic understanding of the Einstein equations were to yield such large dividends.

If the presence of black holes in the SRA is unclear, one might wonder what becomes of the work on black hole entropy. As J. Oppenheim has evidently shown recently [296], the proportionality of black hole entropy to area does not depend on the existence of an event horizon, but merely occurs in the limit as a gravitating system approaches its gravitational radius. Inclusion of the gravitational field in thermodynamics yields a correction term that violates entropy extensivity; in the limit as the radius approaches the Schwarzschild radius, the entropy is proportional to area rather than volume.

Chapter 8

Plane Wave and Infinite Plate Solutions and Finite Range Gravitation

Two further exact solutions in the SRA merit investigation. The plane wave solution appears inconsistent with the SRA in its usual form, though we suggest how it might be accommodated serviceably. An infinite plate in the SRA, it turns out, must be repulsive, which is perhaps not too surprising, given its connection to the Schwarzschild solution at close ranges. If the gravitational field is given a finite range through a graviton rest mass of the Maheshwari-Logunov form, an attractive infinite plate is possible. However, difficulties in producing an acceptable massive theory of gravity in the SRA are noted. Such difficulties might suggest that the massless theory is to be preferred even in the SRA.

8.1 Caustic Plane Waves?

As we saw above while Nicholas was playing drums, plane wave solutions appear to violate the η-causality in the tensorial DeDonder gauge. Given the RTG, one would have to reject such solutions as unphysical, but that seems like too high a price to pay, because we would expect such solutions to be physical, and thus we conclude that the RTG should be modified. We have argued above that a better approach would be to use the gauge freedom to try to recast the solution into a form that is consistent with the flat null cone, though that works only if gauge freedom is present.

The suggestion of reducing the lapse to keep the curved null cone in check faces an obstacle, if one intended to preserve this entire spacetime. It is not merely the case local violations of causality occur, though that is bad enough. Rather, a global violation occurs in the form of caustic plane waves [297, 302, 331]. Bondi and Pirani write,

> We call a plane gravitational wave caustic if it is capable of inducing accelerations of test particles so strong that two particles, initially at rest in flat spacetime and aligned suitably with respect to the wave, but arbitrarily far apart, will collide, within a finite time interval that is independent of their initial separation, after being struck by the wave. We shall show that all plane sandwich waves with fixed polarization are caustic, and describe some unusual optical properties of such waves .... (p. 395).

We recall that a sandwich wave is one having a limited duration, so the wave is 'sandwiched' between two regions of flat spacetime. As Bondi and Pirani observe, a special case of this caustic phenomenon is found in W. Rindler's text [331] (pp. 166-174), while Penrose called attention to this issue in 1965 [297]. Such a phenomenon is obviously inconsistent with the SRA, and thus constitutes a singularity of η-causality violation.

In a set of coordinates that one rather naturally chooses, there exists a coordinate singularity in this solution [331]. We recall the early (1937) worries of N. Rosen [300, 301] that plane waves in general relativity do not exist, because the metric becomes singular. Later it was shown by I. Robinson and by H. Bondi that there exists a coordinate transformation that removes Rosen's singularity, so in the geometrical theory, such waves are in fact nonsingular and acceptable [298, 299]. The evident violation of η-causality in this solution suggests that this apparent singularity is most likely genuine in the SRA. As a result, the only part of the plane wave solution that exists anywhere in Minkowski spacetime is the part before the apparent singularity. It would be interesting to see if there is a tendency for coordinate singularities of the geometrical theory to indicate physical singularities in the SRA, especially given that any solution in the SRA can be given in one coordinate chart.

Whatever the status of the ostensible singularity, the collision of particles initially arbitrarily far apart in a finite amount of physical time is not permissible in the SRA. To avoid this disaster, it seems that one must choose the lapse such that this collision 'event', though finitely removed in physical time, is infinitely distant in metaphysical time. This gravitational time dilation would be analogous to the behavior of a particle falling towards a very compact static spherically symmetric mass, and somewhat like attempts to avoid the Big Bang singularity by removing it to $t = -\infty$ [338–340]. With this choice of lapse, the test particles never collide. Rather, the world gradually slows to a halt, with the time dilation becoming arbitrarily large, so that a given interval of metaphysical time corresponds to ever-smaller intervals of physical time. As $t \to +\infty$, the particles become arbitrarily close to each other, but at no moment do they actually collide. Thus, part of the sandwich wave spacetime described by Bondi and Pirani and by Rindler just does not exist in the SRA.

We also note a midisuperspace (quantized) model of pure gravitational plane waves suggesting classically nontrivial behavior near the region where null cones are focused by the wave [329]. In particular, the quantum fluctuations in that region become significant, and coherent quantum states fail to approximate the classical spacetime well. Given that the classical limit is presumably an approximation to be derived from a fully quantum theory of gravity–which does not perhaps yet exist, but the symmetries of this model permit one to avoid some of the major problems involved–this result suggests that the classical geometric results are unreliable in this region. It is intriguing that a *canonical* quantization of Einstein's theory–which has no trace of a flat background metric–should suggest the breakdown of the classical limit in the same regime that the SRA finds a singularity.

8.2 The Gravitational Field of an Infinite Plate

It seems worthwhile to consider the gravitational field outside an infinite plate in the SRA. It turns out that the causality principle requires that gravity be repulsive, at least in the usual class of solutions. While repulsion might seem surprising, there might be some analogy to the gravitational force just outside the Schwarzschild radius in the spherically symmetric static case; one expects that as one approaches the Schwarzschild radius from the outside, the field would start to resemble that of an infinite plate. The repulsion is rooted in the choice of the non-affine parametrization of the geodesic equation, which implies that the 4-force is not uniquely defined in the direction of the 4-velocity. Our discussion of the plate solution is based on that of P. A. Amundsen and Ø. Grøn [363], apart from a change of signature and an interest in the SRA. One might also see F. Rohrlich's work [364].

The curved metric is assumed to have the form

$$ds^2 = -E(z)dt^2 + G(z)(dx^2 + dy^2) + F(z)dz^2, \tag{8.1}$$

while these coordinates are Cartesian with respect to the flat metric. Although there are two classes of solutions that one could consider [363], we restrict ourselves to the case with $G = 1$, which (apart from the conclusion of repulsion) behaves much as one would expect based on analogy to electrostatics. The field equations reduce to

$$F = \frac{E'^2}{4g^2 E}, \tag{8.2}$$

where g is some constant. It remains to choose a gauge and to satisfy the causality principle. One natural choice is $E = F$, though we postpone a description of its merits. The field equation then becomes trivial to solve, yielding $E(z) = E_0 e^{\pm 2gz}$, the constant g being positive without loss of generality.

Now let us recall the demands of flat spacetime causality: the timelike eigenvalue must be at least as small as the three spacelike eigenvalues. We can therefore discard the growing solution in favor of the decaying one. If we further choose (as do Amundsen and Grøn) to normalize the solution so that $g_{\mu\nu}$ takes the Minkowski form $(-1, 1, 1, 1)$ at the plate $z = 0$, then the curved metric is

$$ds^2 = -e^{-2g|z|}dt^2 + dx^2 + dy^2 + e^{-2g|z|}dz^2. \tag{8.3}$$

One readily shows that the physical length of a line of constant t, x, and y is just $\frac{2}{g}$, though its metaphysical length is of course infinite. Evidently the (repulsive) gravitational field has caused physical rods to grow arbitrarily large far from the plate, because it only takes a finite number of them to cover a metaphysically infinite distance. It is clear from this example that the geometric criterion of geodesic completeness for nonsingularity, that every g-geodesic extend to infinite g-length [152], is inappropriate for the SRA.

It is not difficult to find explicitly some of the geodesics of $g_{\mu\nu}$, the worldlines of test particles subject only to the gravitational force. We use the g-geodesic equation in non-affinely parametrized form [152]

$$\frac{d^2x^\alpha}{d\lambda^2} + \{^{\alpha}_{\mu\nu}\}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} + K(x)\frac{dx^\alpha}{d\lambda} = 0. \tag{8.4}$$

It seems convenient to choose the parameter λ be t, the metaphysical time in the rest frame. The time component of the geodesic equation gives $K = 2g\frac{dz}{dt}$. The x equation gives $\frac{d^2x}{dt^2} + 2g\frac{dz}{dt}\frac{dx}{dt} = 0$, and the y equation is analogous. The z equation, which is of the most interest, gives $\frac{d^2z}{dt^2} - g + g(\frac{dz}{dt})^2 = 0$. Consistency with flat spacetime causality has thus yielded an anti-Newtonian limit for slow velocities. However, particles with $u = 1$ are neither repelled nor attracted. It is not difficult to integrate the z equation.

Above it was suggested that repulsion could be made plausible by analogy to the Schwarzschild solution. We recall that a particle takes infinite metaphysical time to fall to the Schwarzschild singularity, at least in a static gauge, so the gravitational force, suitably defined, must be repulsive at short ranges, as indeed one can show. This repulsion follows from the choice of the non-affine parametrization of the g-geodesic equation, which permits adding a 'force' parallel to the 4-velocity with a fairly arbitrary coefficient. The repulsion is not confined to short ranges, but can occur at longer ranges given a large radial velocity. Using the standard forms for the Schwarzschild metric (and the flat metric), and parametrizing the motion of a radially-moving particle by t, the metaphysical time in the rest frame, the g-geodesic equation reduces to $\frac{d^2r}{dt^2} = \frac{M}{(r^3)(r-2M)}[-(r-2m)^2 + 3r^2(\frac{dr}{dt})^2]$. This form shows that repulsion can occur either for large radial velocities or for small distances from the Schwarzschild radius.

While the Schwarzschild analogy might reconcile one to the repulsive force of the plate, one might wonder if the plate result is gauge-variant, perhaps due merely to the choice $E = F$. In fact, given the natural assumptions that one wants to make (including $G = 1$), that is true for every gauge, as we now show. In order to achieve attraction for a the particle at rest, one needs $\frac{dE}{dz} > 0$. One would expect the plate to be attractive for all z, so this derivative must be positive for all $z > 0$. Given the assumption that $G = 1$, causality requires that $E \leq 1$. Thus, it must be the case that $0 < E \leq 1$ (although one could perhaps permit $E = 0$ at $z = 0$), and that E is a strictly increasing function of z for all positive z. Obviously it follows that $E' \to 0$ as $z \to \infty$. But then the field equation $F = \frac{E'^2}{4g^2 E}$ implies that $F \to 0$, in violation of causality for large z. Thus, at least given $G = 1$, there is no way of securing an attractive infinite plate in the SRA, at least not as the sole content of the universe. Given that the behavior of the curved metric at arbitrarily large values of z played a role in this argument, it might turn out that including other objects in the universe (apart from test bodies) would make an attractive infinite plate possible.

It is interesting that domain walls involve repulsive infinite-plate-like solutions [366].

8.3 Infinite Plate in Finite-Range Gravitation

We find it interesting that if one gives gravitation a mass term of the form employed by Freund, Maheshwari, and Schonberg [260] and by Logunov *et al.*, then, at least to linear order, the problem of the impossibility of an attractive infinite plate (alone in the universe) disappears. (We suspect that it holds in the nonlinear theory also, but have not solved the nonlinear ODE). This removal of the difficulty is not too surprising, because the rest mass gives gravitation a finite range, implying that most of the infinite plate has negligible effect on the gravitational potential at a given point. The plate thus acts something like a finite patch only.

Retaining $G = 1$ but using the massive field equations, we observe that *necessarily* $E = F$; the remaining equation is

$$E'' - \frac{E'^2}{E} + m^2(E - E^2) = 0, \tag{8.5}$$

where the speed of light and the reduced Planck constant have both been set to unit value. This equation is not easy to solve[1], so we linearize by setting $E = 1 + \epsilon$, where $|\epsilon| \ll 1$. The resulting field equation is just $\epsilon'' - m^2\epsilon = 0$. Given that $G = 1$, causality will exclude an exponentially growing solution (not to mention its violating the assumption of linearization), so we set

$$\epsilon = -ae^{-mz}. \tag{8.6}$$

The equation of motion for a test body at rest is $\frac{d^2z}{dt^2} + \left\{ {z \atop tt} \right\} = 0$. Using $\left\{ {z \atop tt} \right\} = \frac{E'}{2E}$ and the form of ϵ, one gets (at linear order) $\frac{d^2z}{dt^2} + \frac{ame^{-mz}}{2} = 0$. Clearly a massive theory will have the curved metric approximate the flat one as $z \to \infty$, so we cannot impose that requirement at the plate. Instead, we require the Newtonian limit to hold at the plate, and obtain $a = \frac{2g}{m}$. Given that $0 < a \leq 1$, it follows that for fixed Newtonian acceleration g, there is a minimum 'graviton' rest mass: $m \geq 2g$. Taking the limit as the mass vanishes therefore makes g vanish, too, so we can see why no infinite-range attractive plate solution exists. The linearization is a good approximation when a is small, which is to say, when $m \gg g$. It appears, then, that an attractive flat plate requires gravitational rest mass, indeed, a suitably large one.

8.4 Causality in Finite-Range Gravitation

While numerical solution of the differential equation for this massive theory above would be possible, such effort seems unwarranted given the problems that the theory faces. More generally, if massive relatives of Einstein's theories are to be accepted, then two difficulties must be overcome. The first is the negative-energy worry about

[1] Given that this theory faces both causality and ghost worries, solving the equations numerically does not seem worth the trouble.

massive "ghost" theories [234], which has been with us for the last three decades. The second is the null cone consistency issue, at least if one wishes to interpret the theory in accord with the SRA. Given the observability of flat spacetime in a massive theory, such an interpretation is very natural, so its absence would be somewhat embarrassing to a massive theory. While the negative energy question was raised some time ago, more recently M. Visser has doubted that this worry is quite so conclusive [235]. Also, Loskutov has made calculations indicating the positive-definiteness of the gravitational radiation intensity in the Maheshwari-Logunov theory [368] (and earlier work). However, the above-mentioned causality worries for linearized plane waves obeying the tensorial DeDonder condition apply to the Maheshwari-Logunov massive theory, because in the high-frequency limit, the rest mass will not affect the dispersion relation. As a result, it is not clear what to make of Loskutov's work in the SRA.

Recently I. P. Denisova has found an interesting electrovac plane wave solution for the massive RTG. (Solutions in the massive theory are more difficult to find than for Einstein's equations, in part because the field equations are all independent in the massive case.) The electromagnetic source gives rise to a gravitational "wave" that merely makes the curved metric represent a noninertial frame of reference as the wave passes by. As Denisova notes, this solution obeys η-causality. One difference between this solution and the linearized vacuum plane monochromatic wave that violates η-causality is that the former is longitudinal, whereas the vacuum waves are transverse: an electromagnetic wave moving along the z-axis affects only the $t-t$, $t-z$, and $z-z$ components of the effective curved metric. (However, it is quite unclear that the massive RTG has enough solutions obeying η-causality, so the presence of one wave that does will not touch that fundamental issue.) Denisova's solution because singular in the massless limit, so apparently it has no analog in the massless case. Using the eigenvector technology developed above, one can show that this solution is of Segré type $\{211\}$, the class of metrics that can just barely satisfy η-causality, and that our necessary and sufficient conditions for η-causality indeed hold. However, if our proposal for securing enough solutions by exploiting gauge freedom is accepted, then this solution would not be acceptable in its $\{211\}$ present form, but rather would require gauge-transforming to make it a $\{1111\}$ solution.

The reader will perhaps be puzzled by the call for a gauge transformation in the RTG, inasmuch as the theory has no nontrivial gauge freedom. However, some time ago Stueckelberg showed how to turn Proca's massive electromagnetism into a gauge theory by introducing a suitably-coupled scalar field [294, 295]. It would be interesting to see if a similar move can be made using a massive version of general relativity (such as the Maheshwari-Logunov theory) in order to secure enough gauge freedom to fix the null cone relationship. One might try a substituting a finitely-gauged transformed curved metric and matter fields, a procedure analogous to one that yields Stueckelberg's gauge-invariant version of Proca's massive electromagnetism. Probably a single arbitrary function would provide enough gauge freedom to yield consistency by adjusting the lapse, so it is not necessary to restore the full amount of gauge freedom to the massive theory–which would undo some of the benefit of adding the mass term. One might let the vector field(s) generating the transformation(s) be parallel to the relevant eigenvector(s) e^{μ}_{A}, or be the (contravariant) gradient of some scalar field(s). However, the resulting theories might tend to have higher derivatives, indeed, an infinite number of derivatives, appearing in the erstwhile graviton mass term, and thus be nonlocal in most gauges. One might also consider using the Batalin-Fradkin-Tyutin (BFT) procedure [370–372] (or some relative thereof) for converting second-class constraints (such as in massive electromagnetism and massive gravity in their simplest forms) into first-class constraints *via* the introduction of additional fields, though such an undertaking would not be technically trivial.

While a graviton mass would have profound theoretical consequences, its experimental effect, at least in many situations [235], would be limited by known bounds on the graviton mass [367].

Chapter 9

Conclusion

In this work we have contemplated the role that nondynamical or "absolute" objects–in particular, a flat metric and perhaps a preferred time foliation–ought to play in gravitation. We have found no excellent reason to include an invariant time foliation in physics, but good reasons to include one in metaphysics. While one might be disturbed that physics and metaphysics would differ so, we have suggested that there could be good reasons for this distinction to exist.

Concerning a flat metric, we have found another clean derivation of the Hilbert action of general relativity from a gauge-invariant free field theory which gets universally coupled to the stress tensor. We have also found that including the determinant of the flat metric in the field equations permits one to write slightly bimetric theories of gravity, which are phenomenologically similar to very general scalar-tensor theories. However, despite appearances, neither these theories nor general relativity derived from flat spacetime is manifestly consistent with special relativity, because they give no assurance that the flat metric's causal structure will be respected. We then reviewed the history of the discussion of this issue since 1939, identifying a number of stances and diagnosing them as unsatisfactory, or at least not in line with our interests.

We then undertook to solve the problem ourselves, and largely succeeded. The kinematic issues were resolved using a generalized eigenvector formalism, which we showed to exist in the relevant cases. The dynamical issue of enforcing stable η-causality can be solved by using specially-adapted variables in which it holds identically, a procedure inspired by Klotz's approach to the positivity conditions of canonical quantum gravity. This procedure reduces the configuration space for the curved metric so that only good configurations are possible. Given that a gauge transformation should connect one physically acceptable configuration with another, the SRA has rather restricted gauge freedom. Gauge transformations could not be identified merely with their generating vector fields, but also depend essentially on the curved metric configuration assumed prior to transformation. As a result, gauge transformations in the SRA do not form a group, because the multiplication operation is not always defined. Studying simple cases, such as homogeneous cosmological models, might give insight into the explicit form of the SRA gauge transformations.

Given the restricted gauge freedom and the need to respect stable η-causality, the issue of gauge *vs.* physical singularities had to be reconsidered. For the Schwarzschild solution, the jury is still out. But the SRA implies global hyperbolicity, which implies that any regions of no escape will not be as exotic as are the black holes of the geometrical theory. Furthermore, global hyperbolicity probably resolves the gravitational information loss paradox. Caustic plane waves, which seemed incompatible with the SRA and lack a Cauchy surface, in fact likely can be handled by reducing the lapse enough to postpone the unacceptable region forever and thus excluding it from the physical spacetime in the SRA.

For infinite plate masses, the SRA to Einstein's equations implied a repulsive plate, although adding a mass term permitted an attractive plate. While massive gravity seems very natural from a flat spacetime perspective, as Logunov has observed, we find a difficulty: the appealing properties are rooted in destroying the gauge freedom, whereas respecting η-causality (and ensuring a 'complete set' of solutions) appears to require some gauge freedom. Furthermore, it was not clear how to produce a satisfactory massive theory with gauge freedom. If this latter difficulty cannot be resolved, then it might follow that the massless theory is in fact preferable even in the SRA.

Bibliography

[1] J. L. Anderson, *Principles of Relativity Physics.* Academic, New York (1967).

[2] K. S. Thorne, D. L. Lee, A. P. Lightman, *Phys. Rev. D* **7** (1973), 3563.

[3] J. D. Norton, *Rep. Prog. Phys.* **56** (1993), 791.

[4] M. Fierz, *Helv. Phys. Acta* **12** (1939), 3.

[5] M. Fierz, W. Pauli, *Proc. Roy. Soc. A* **173** (1939), 211.

[6] N. Rosen, *Phys. Rev.* **57** (1940), 147, 150.

[7] N. Rosen, *Ann. Phys.* (N.Y.) **22** (1963), 1.

[8] M.-A. Tonnelat, *Ann. d. Phys.* **17** (1942), 159.

[9] M.-A. Tonnelat, in *Proceedings on Theory of Gravitation: Conference in Warszawa and Jabłonna*, 25-31 July, 1962, *ed.* L. Infeld. Gauthier-Villars, Paris (1964). This work refers to more of her papers.

[10] H. Weyl, *Am. J. Math.* **66** (1944), 591.

[11] A. Papapetrou, *Proc. Roy. Irish Acad. A* **52** (1948), 11.

[12] A. Papapetrou, W. Urich, *Annalen der Physik* **6** (1954), 220.

[13] M. Moshinsky, *Phys. Rev.* **80** (1950), 514.

[14] M. Kohler, *Zeit. Phys.* **131** (1952), 571.

[15] M. Kohler, *Zeit. Phys.* **134** (1953), 286, 306.

[16] S. N. Gupta, *Proc. Phys. Soc. (Lon.) A* **65** (1952), 161 and 608.

[17] S. N. Gupta, *Phys. Rev.* **96** (1954), 1683.

[18] S. N. Gupta, *Rev. Mod. Phys.* **29** (1957), 334.

[19] S. N. Gupta, *Phys. Rev.* **172** (1968), 1303.

[20] R. Kraichnan, *Phys. Rev.* **98** (1955), 1118.

[21] R. Kraichnan, *Phys. Rev.* **101** (1956), 482.

[22] F. J. Belinfante, *Phys. Rev.* **98** (1955), 793.

[23] F. J. Belinfante, J. C. Swihart, *Ann. Phys.* (NY) **1** (1957), 168, 196; **2** (1957) 81.

[24] F. J. Belinfante, J. C. Swihart, *Ann. Phys.* (NY) **2** (1957) 81.

[25] F. J. Belinfante, *Rev. Mex. Fis.* 4 (1955), 192.

[26] F. J. Belinfante, D. I. Caplan, W. L. Kennedy, *Rev. Mod. Phys.* **29** (1957), 518.

[27] F. J. Belinfante, J. C. Garrison, *Phys. Rev.* **125** (1962), 1124.

[28] W. Thirring, *Fortsch. Phys.* **7** (1959), 79.

[29] W. Thirring, *Ann. Phys. (N.Y.)* **16** (1961), 96.

[30] W. Thirring, *Acta Phys. Aust.* **suppl. 5** (1968), 391.

[31] W. Thirring, in *Contemporary Physics: Trieste Symposium 1968*, International Atomic Energy Agency, Vienna (1968).

[32] I. I. Gutman, *Sov. Phys. J. E. T. P.* **37** (1960), 1162.

[33] R. Sexl, in *Evidence for Gravitational Theories: Proceedings of the International School of Physics "Enrico Fermi"*, June 19-July 1, 1961, Varenna on Lake Como, Villa Monastero, *ed.* C. Møller. Academic, New York (1962).

[34] R. U. Sexl, *Fortschr. Phys.* **15** (1967), 269.

[35] O. Nachtmann, H. Schmidle, R. U. Sexl, *Acta Phys. Austr.* **29** (1968), 289.

[36] O. Nachtmann, H. Schmidle, R. U. Sexl, *Commun. Math. Phys.* **13** (1969), 254.

[37] S. Mandelstam, *Ann. Phys. (N.Y.)* **19** (1962), 25.

[38] D. E. Burlankov, *Sov. Phys. J. E. T. P.* **17** (1963), 1306.

[39] L. Halpern, *Bull. Cl. Sci. Acad. R. Belg., 5e serie,* **49** (1963), 226.

[40] L. Halpern, *Ann. Phys. (N.Y.)* **25** (1963), 387.

[41] J. R. Klauder, *Nuov. Cim.* **25** (1962), 542.

[42] E. R. Huggins, *Quantum Mechanics of the Interaction of Gravity with Electrons: Theory of a Spin-Two Field Coupled to Energy.* Dissertation, California Institute of Technology (1962).

[43] R. Feynman, F. Morinigo, W. Wagner, B. Hatfield, *Feynman Lectures on Gravitation*, Addison-Wesley, Reading, Mass. (1995).

[44] R. Feynman, *Acta Phys. Polon.* **24** (1963), 697.

[45] R. P. Feynman, in *Magic Without Magic: John Archibald Wheeler, A Collection of Essays in Honor of His Sixtieth Birthday*, *ed.* J. R. Klauder, Freeman, San Franciso (1972).

[46] R. Arnowitt, in *Proceedings of the Eastern Theoretical Physics Conference, October 26-7, 1962*, *ed.* M. E. Rose. Gordon and Breach, New York (1963).

[47] F. H. J. Cornish, *Proc. Roy. Soc.* **A282** (1964) 358, 372.

[48] S. Weinberg, *Phys. Rev.* **133** (1964), B1318.

[49] S. Weinberg, *Phys. Rev.* **134** (1964), B882.

[50] S. Weinberg, *Phys. Lett.* **9** (1964), 357.

[51] S. Weinberg, *Phys. Rev.* **135** (1964), B1049.

[52] S. Weinberg, "The Quantum Theory of Massless Particles," in *Lectures on Particles and Field Theory, 1964, volume 2*, *ed.* S. Deser and K. W. Ford, Prentice Hall, Englewood Cliffs, N. J. (1965).

[53] S. Weinberg, *Phys. Rev.* **138** (1965), B988.

[54] S. Weinberg, in *Contemporary Physics: Trieste Symposium 1968*, International Atomic Energy Agency, Vienna (1968).

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

[56] S. Weinberg, E. Witten, *Phys. Lett. B* **96** (1980), 59.

[57] R. Penrose, in *Quantum Structure of Space and Time: Proceedings of the Nuffield Workshop, Imperial College London*, 3-21 August 1981, *ed.* M. J. Duff, C. J. Isham, Cambridge Univ., Cambridge (1981).

[58] W. Wyss, *Helv. Phys. Acta* **38** (1965), 469.

[59] W. Wyss, *Helv. Phys. Acta* **70** (1997), 336.

[60] V. I. Ogievetsky, I. V. Polubarinov, *Ann. Phys. (NY)* **35** (1965), 167.

[61] V. I. Ogievetskii, I. V. Polubarinov, *Sov. Phys. Dokl.* **11** (1966), 71.

[62] V. I. Ogievetskii, I. V. Polubarinov, *Sov. Phys. Dokl.* **11** (1966), 135.

[63] P. Mittelstaedt, J. B. Barbour, *Z. Phys.* (1967), 82.

[64] K. Westpfahl, *Fortschr. Phys.* **15** (1967), 309.

[65] S. Deser, *Gen. Rel. Grav.* **1** (1970), 9.

[66] D. G. Boulware, S. Deser, *Ann. Phys. (N.Y.)* **89** (1975), 193.

[67] S. Deser, in *Quantum Gravity: An Oxford Symposium*, *ed.* C. J. Isham, R. Penrose, and D. W. Sciama, Clarendon, Oxford (1975).

[68] D. G. Boulware, S. Deser, J. H. Kay, *Physica A* **96** (1979), 141.

[69] S. Deser, *Class. Quant. Grav.* **4** (1987), L99.

[70] S. Deser, R. I. Nepomechie, *Phys. Let. B* **132** (1983), 321.

[71] S. Deser, R. I. Nepomechie, *Ann. Phys. (NY)* **154** (1984), 396.

[72] S. Deser, in *Developments in General Relativity, Astrophysics, and Quantum Theory: A Jubilee Volume in Honour of Nathan Rosen*, *ed.* F. I. Cooperstock, L. P. Horwitz, J. Rosen. I.O. P., Bristol, and the Israel Physical Society, Jerusalem (1990).

[73] S. Deser, A. Waldron, V. Pascalutsa, *Phys. Rev. D* **62** (2000), 105031.

[74] G. Velo, Z. Zwanziger, *Phys. Rev.* **186** (1969), 1337.

[75] P. van Nieuwenhuizen, *Nuc. Phys. B* **60** (1973), 478.

[76] P. van Nieuwenhuizen, in *Proceedings of the First Marcel Grossmann Meeting on General Relativity*, 1975, Trieste, *ed.* R. Ruffini. North Holland, Amsterdam (1977).

[77] C. Rovelli, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0006061; presented at the 9th Marcel Grossmann Meeting in Rome, July 2000.

[78] H. J. Groenewold, *Int. J. Theor. Phys.* **9** (1974), 19.

[79] J. Fang, C. Fronsdal, *J. Math. Phys.* **20** (1979), 2264.

[80] G. Cavalleri, G. Spinelli, *Rivista del Nuovo Cimento* **3** (1980), 1.

[81] R. Penrose, in *Essays in General Relativity—A Festschrift for Abraham Taub, ed.* F. J. Tipler. New York, Academic (1980).

[82] L. P. Grishcuhk, A. N. Petrov, A. D. Popova, *Commun. Math. Phys.* **94** (1984), 379.

[83] Ya. B. Zel'dovich, L. P. Grishchuk, *Sov. Phys. Usp.* **29** (1986), 780.

[84] Ya. B. Zel'dovich, L. P. Grishchuk, *Sov. Phys. Usp.* **31** (1988), 666.

[85] L. P. Grishchuk, *Sov. Phys. Usp.* **33** (1990), 669.

[86] D. E. Burlankov, *Sov. J. Nucl. Phys.* **50** (1989), 174.

[87] R. M. Avakian, L. Sh. Grigorian, *Astro. Space Sci.* **146** (1988), 183.

[88] N. Straumann, *Los Alamos Preprints*, xxx.lanl.gov, astro-ph/0006423.

[89] A. A. Logunov, M. A. Mestvirishvili, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **86** (1991), 3.

[90] A. A. Logunov, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **92** (1992), 191.

[91] A. A. Logunov, *The Relativistic Theory of Gravitation.* Nova Science, Commack, New York (1998).

[92] I. Gottlieb, N. Ionescu-Pallas, *Rev. Roum. Phys.* **35** (1990), 25.

[93] S. Babak, L. Grishchuk, *Phys. Rev. D* **61** (2000), 024038.

[94] A. N. Petrov, J. Katz, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9905088; to appear in the Proceedings of the Conference on "Fundamental Interactions: From Symmetries to Black Holes" held in Brussels March 25 - 27, 1999.

[95] A. N. Petrov, J. Katz, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9911025.

[96] A. A. Vlasov, *Sov. J. Nucl. Phys.* **51** (1990), 1148; trans. *Yad. Fiz.*

[97] A. N. Petrov, *Astron. Astrophys. Trans.* **1** (1992), 195.

[98] Yu. M. Loskutov, *Sov. J. Nucl. Phys.* **54** (1991), 903.

[99] A. D. Popova, A. N. Petrov, *Int. J. Mod. Phys. A* **11** (1988), 2651,

[100] L. P. Grishchuk, A. N. Petrov, *Sov. Phys. J. E.T. P.* **65** (1987), 5; trans. *Zh. Eksp. Teor. Fiz.* **92** (1987), 9.

[101] Ya. I. Pugachev, V. D. Gun'ko, *Sov. Phys. Journ.*, trans. *Izvestiya Vysshikh Uchebnykh Zavednii, Fizika* **10** (1974), 46.

[102] A. N. Petrov, J. V. Narlikar, *Found. Phys.* **26**, 1201, 1996.

[103] A. N. Petrov, *Int. J. Mod. Phys. D* **4** (1995), 451.

[104] A. N. Petrov, *Int. J. Mod. Phys. D* **6** (1997), 239.

[105] A. A. Vlasov, *Moscow Unversity Physics Bulletin*, trans. *Vestnik Moskovskogo Universiteta. Fizika* **43** (1988), 87.

[106] E. Alvarez, *Rev. Mod. Phys.* **61** (1989), 561.

[107] Yu. V. Baryshev, *Gravitation* bf 2 (1996), 69; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9912003.

[108] H. Nikolić, *Gen. Rel. Grav.* **31** (1999), 1211.

[109] H. Nikolić, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9901045.

[110] A. I. Nikishov, *Los Alamos Preprrints*, xxx.lanl.gov, gr-qc/9912034.

[111] J. B. Pitts and W. C. Schieve, *Gen. Rel. Grav.* **33** (2001), forthcoming; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0101058.

[112] G. Magnano, L. M. Sokołowski, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0103072.

[113] G. Scharf, *Quantum Gauge Theories: A True Ghost Story.* Wiley, New York (2001).

[114] T. Borne, G. Lochak, H. Stumpf *Nonperturbative Quantum Field Theory and the Structure of Matter.* Kluwer, Dordrecht (2001).

[115] H. Fuchs, U. Kasper, D.-E. Liebscher, V. Müller, H.-J. Schmidt, *Fortschr. Phys.* **36** (1988), 427.

[116] G. Scharf, M. Wellmann, *Gen. Rel. Grav.* **33** (2001), 553.

[117] J. B. Pitts, W. C. Schieve, in *Proceedings of the 20th Texas Symposium on Relativistic Astrophysics, ed.* H. Martel, J. C. Wheeler, forthcoming; gr-qc/0101097.

[118] B. S. DeWitt, in *Gravitation: An Introduction to Current Research, ed.* L. Witten. Wiley, New York (1962).

[119] B. S. DeWitt, *Phys. Rev.* **162** (1967), 1195.

[120] B. S. DeWitt, in *Magic Without Magic: John Archibald Wheeler, A Collection of Essays in Honor of His Sixtieth Birthday, ed.* J. R. Klauder. Freeman, San Franciso (1972).

[121] T. A. Barnabey, *Phys. Rev. D* **10** (1974), 1741.

[122] L. Fatibene, M. Ferraris, M. Francaviglia, M. Raiteri, J. Math. Phys. **42** (2001), 1173; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0003019.

[123] E. Martinec, *Class. Quantum Grav.* **10** (1993), L187.

[124] N. Straumann, *General Relativity and Relativistic Astrophysics.* Springer, Berlin (1984).

[125] A. Ashtekar, R. Geroch, *Rep. Prog. Phys.* **37** (1974), 1211.

[126] R. Geroch, *J. Math. Phys.* **9** (1968), 1739.

[127] S. W. Hawking, G. F. R. Ellis, *The Large Scale Structure of Space-time.* Cambridge Univ., Cambridge (1973).

[128] R. Wald, *Phys. Rev. D* **33** (1986), 3613.

[129] K. R. Heiderich, W. G. Unruh, *Phys. Rev. D* **38** (1988), 490; also *Phys. Rev. D* **42** (1990), 2057.

[130] L. P. Grishchuk, A. D. Popova, *Sov. Phys. J. E. T. P.* **53** (1981), 1.

[131] L. P. Grishchuk, A. D. Popova, *J. Phys. A.: Math. Gen.* **15** (1982), 3525.

[132] J. R. Fanchi, *Parametrized Relativistic Quantum Theory.* Kluwer Academic, Dordrecht (1993).

[133] M. C. Land, L. P. Horwitz, *Found. Phys.* **21** (1991) 299.

[134] M. C. Land, L. P. Horwitz, *Found. Phys. Let.* **4** (1991) 61.

[135] M. C. Land, *Found.Phys.* **27** (1997) 19.

[136] N. Shnerb, L. P. Horwitz, *Phys. Rev. A* **48** (1993) 4068.

[137] M. C. Land, N. Shnerb, L. P. Horwitz, *J. Math. Phys.* **36** (1995) 3263.

[138] L. P. Horwitz, *Found. Phys.* **22** (1992) 421.

[139] D. Saad, L. P. Horwitz, R. I. Arshansky, *Found. Phys.* **19** (1989) 1125.

[140] L. P. Horwitz, C. Piron, *Helv. Phys. Acta* **46** (1973) 316.

[141] L. P. Horwitz, R. I. Arshansky, A. C. Elitzur, *Found. Phys.* **18** (1988) 1159.

[142] R. Arshansky, L. P. Horwitz, Y. Lavie, *Found. Phys.* **13** (1983) 1167.

[143] J. Frastai, L. P. Horwitz, *Found. Phys.* **25** (1995) 1495.

[144] J. B. Pitts and W. C. Schieve, *Found. Phys.* **28** (1998) 1417.

[145] J. B. Pitts and W. C. Schieve, *Found. Phys.* **29** (1999) 1977.

[146] J. B. Pitts and W. C. Schieve, *Found. Phys.* **31** (2001), forthcoming; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/gr-qc/0101099.

[147] P. Miller, in *Physics and the Ultimate Significance of Time*, *ed.* D. R. Griffin. SUNY, Albany (1986).

[148] C. Piron, F. Reuse, *Helv. Phys. Acta* **48** (1975), 631.

[149] M. A. Trump, W. C. Schieve, *Classical Relativistic Many-Body Dynamics.* Kluwer Academic, Dordrecht (1999).

[150] L. Randall, R. Sundrum, *Phys. Rev. Lett.* **83** (1999) 4690.

[151] C. Misner, K. Thorne, and J. Wheeler, *Gravitation.*. Freeman, New York (1973).

[152] R. Wald, *General Relativity.* Univ. Chicago, Chicago (1984).

[153] W. G. Unruh, R. M. Wald, *Phys. Rev. D* **40** (1989) 2598).

[154] F. Selleri, *Found. Phys.* **27** (1997), 1527.

[155] I. Schmelzer, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0001101.

[156] D. Z. Albert, in *Relativistic Quantum Measurement and Decoherence: Lectures of a Workshop Held at the Istituto Italiano per gli Studi Filosofici, Naples, April 9-11, 1999*, *ed.* H.-P. Breuer, F. Petruccione. Springer, Berlin (2000).

[157] H. B. Nielsen, M. Ninomiya, *Nuc. Phys. B* **141** (1978), 153.

[158] H. B. Nielsen, I. Picek, *Nuc. Phys. B* **211** (1983), 269.

[159] S. Chadha, H. B. Nielsen, *Nuc. Phys. B* **217** (1983), 125.

[160] P. A. M. Dirac, *Proc. Roy. Soy. (London) A* **246** (1958), 333.

[161] M. Ferrero, E. Santos, *Found. Phys.* **27** (1997), 765.

[162] L. E. Szabó, A. Fine, *Los Alamos Preprints*, xxx.lanl.gov, quant-ph/0007102.

[163] M. Genovese, G. Brida, C. Novero, E. Predazzi, *Los Alamos Preprints*, xxx.lanl.gov, quant-ph/0009067.

[164] W.A. Hofer, *Los Alamos Preprints*, xxx.lanl.gov, quant-ph/0103014.

[165] C. H. Brans, *Int. J. Theor. Phys.* **27**, 219 (1988).

[166] T. Durt, *Found. Phys.* **27** (1997), 415.

[167] T. Durt, *Int. J. Theor. Phys.* **35** (1996) , 2271.

[168] J. Bell, in *The Ghost in the Atom*, *ed.* P. C. W. Davies, J. R. Brown. Cambridge Univ., Cambridge (1986).

[169] R. K. Clifton, M. L. G. Redhead, J. N. Butterfield, *Found. Phys.* **21** (1991) 149.

[170] J. S. Bell, *Speakable and Unspeakable in Quantum Mechanics.* Cambridge Univ., Cambridge (1987).

[171] P. Van Inwagen, *An Essay on Free Will.* Oxford Univ., Oxford (1986). Though an advocate of free will, Van Inwagen finds no way to answer this “Mind” argument.

[172] R. Kane, *The Significance of Free Will.* Oxford Univ., Oxford, (1998). Kane, a defender of free will, wrestles with this argument through much of the book.

[173] J. Edwards, *The Freedom of the Will.* Soli Deo Gloria, Morgan, Pennsylvania (1996). Reprint of (Thomas Nelson, London, 1845).

[174] B. Metzger, *The New Testament: Its Background, Growth, and Content*, 2nd ed. Abingdon, Nashville (1983). Metzger cites Josephus.

[175] C. J. Isham, Lectures presented at the NATO Advanced Study Institute “Recent Problems in Mathematical Physics,” Salamanca, June 15-27, 1992. *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9210011.

[176] T. Jacobson, D. Mattingly, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0007031.

[177] Q. Smith, *Philosophical Studies* **52**, 371 (1987).

[178] M. Dorato, *Time and Reality: Spacetime Physics and the Objectivity of Temporal Becoming.* CLUEB, Bologna (1995).

[179] C. J. Isham, J. C. Polkinghorne, in *Quantum Cosmology and the Laws of Nature: Scientific Perspectives on Divine Action*, 2nd ed., *ed.* R. J. Russell, N. Murphey, and C. J. Isham. Vatican Observatory, Vatican City State, and the Center for Theology and the Natural Sciences, Berkeley (1999).

[180] R. M. Chisholm, D. W. Zimmerman, *Noûs* **31** (1997), 262.

[181] J. Perry, *Noûs* **13**, 3 (1979).

[182] D. H. Mellor, *Real Time.* Cambridge Univ., Cambridge (1981).

[183] Q. Smith, *Language and Time.* Oxford Univ., New York (1993).

[184] W. L. Craig, lectures given at Talbott Seminary, Biola Univ., Los Angeles (1995).

[185] W. L. Craig, *The Tensed Theory of Time.* Kluwer Academic, Dordrecht (2000).

[186] W. L. Craig, *The Tenseless Theory of Time.* Kluwer Academic, Dordrecht (2000).

[187] W. L. Craig, *Time and the Metaphysics of Relativity*, Kluwer Academic, Dordrecht (2001).

[188] E. Prugovečki, *Quantum Geometry: A Framework for Quantum General Relativity.* Kluwer Academic, Dordrecht (1992).

[189] A. N. Prior, *Philosophy* **34** (1959), 12.

[190] A. N. Prior, *Philosophy* **37** (1962) 114.

[191] D. Lewis, *The Philosophical Review* **88** (1979), 513.

[192] N. Wolterstorff, in *Contemporary Philosophy of Religion, ed.* S. M. Cahn, and D. Shatz. Oxford Univ., New York (1982).

[193] N. Kretzmann, *The Journal of Philosophy* **63** (1966), 409.

[194] D. H. Mellor, *Ratio* **23** (1981) 20. Mellor aims to refute this line of argument.

[195] H. Reichenbach, *The Direction of Time.* Univ. California, Berkeley (1956).

[196] D. R. Griffin, *Physics and the Ultimate Significance of Time.* SUNY, Albany (1986).

[197] T. Maudlin, *Quantum Non-Locality and Relativity: Metaphysical Intimations of Modern Physics.* Blackwell, Oxford (1994).

[198] P. E. Hodgson, *Cont. Phys.* **42** (2001), 117.

[199] F. Shojai, M. Golshani, *Int. J. Mod. Phys. A* **13** (1998), 2135.

[200] S. Goldstein, S. Teufel, in *Physics Meets Philosophy at the Planck Scale*, *ed.* C. Callender, N. Huggett. Cambridge Univ., Cambridge (2001); *Los Alamos Preprints*, xxx.lanl.gov, quant-ph/9902018.

[201] J. Butterfield, C. J. Isham, in *Physics Meets Philosophy at the Planck Scale*, *ed.* C. Callender, N. Huggett. Cambridge Univ., Cambridge (2001); *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9903072.

[202] H. P. Stapp, in *Physics and the Ultimate Significance of Time*, *ed.* D. R. Griffin. SUNY, Albany (1986).

[203] J. T. Wilcox, *The Journal of Religion* **41** (1961), 293.

[204] L. S. Ford, *The Journal of Religion* **48** (1968) 124.

[205] R. G. Gruenler, *The Inexhaustible God.* Baker, Grand Rapids (1983).

[206] W. L. Craig, *Religious Studies* **26** (1990) 335.

[207] N. Rosen, *Ann. Phys. (N.Y.)* **84** (1974), 455.

[208] R. Adler, M. Bazin, M. Schiffer, *Introduction to General Relativity.* McGraw-Hill, New York (1965).

[209] H. Stephani, *General Relativity*, 2nd ed. Cambridge Univ., Cambridge (1990).

[210] W. Pauli, *Theory of Relativity*, *trans.* G. Field. Pergamon, New York (1958).

[211] C. Møller, *Mat. Fys. Skr. Dan. Vid. Selsk.* **1** (1961), no. 1.

[212] C. Møller, in *Evidence for Gravitational Theories: Proceedings of the International School of Physics "Enrico Fermi," course XX*, Varenna on Lake Como, Villa Monastero, June 19-July 1, 1961, *ed.* C. Møller. Academic, New York (1963).

[213] C. Møller, in *Proceedings on Theory of Gravitation: Conference in Warszawa and Jabłonna*, 25-31 July, 1962, *ed.* L. Infeld. Gauthier-Villars, Paris (1964).

[214] C. Møller, *Mat. Fys. Medd. Dan. Vid. Selsk.* **34** (1964), no. 3.

[215] C. Møller, *Mat. Fys. Medd. Dan. Vid. Selsk.* **35** (1966), no. 3.

[216] N. Rosen, *Found. Phys.* **15** (1985), 997.

[217] N. Rosen, in *Proceedings on Theory of Gravitation: Conference in Warszawa and Jabłonna*, 25-31 July, 1962, *ed.* L. Infeld. Gauthier-Villars, Paris (1964).

[218] F. H. Hehl, *Rep. Math. Phys.* **9** (1976), 55.

[219] F. W. Hehl, G. D. Kerlick, *Gen. Rel. Grav.* **9** (1978), 691.

[220] J. B. Ray, *Nuov. Cim. B* **25** (1975), 706.

[221] U. Lindström, *Int. J. Mod. Phys. A* **3** (1988), 2401.

[222] W. Israel, *Differential Forms in General Relativity,* 2nd ed. Dublin Institute for Advanced Studies, Dublin (1979).

[223] V. A. Fock, *The Theory of Space, Time and Gravitation*, *trans.* N. Kemmer. Pergamon, New York (1959).

[224] K. V. Kuchař, C. G. Torre, *Phys. Rev. D* **44** (1991), 3116.

[225] A. Anderson, J. W. York, Jr., *Phys. Rev. Lett.* **81** (1998) 1154.

[226] L. D. Landau, E. M. Lifshitz. *The Classical Theory of Fields, 4th rev. English ed., trans.* M. Hamermesh. Pergamon, Oxford (1975).

[227] J. Katz, A. Ori, *Class. Quant. Grav.* **7** (1990), 787.

[228] D. I. Santiago, A. S. Silbergleit, *Gen. Rel. Grav.* **32** (2000), 565.

[229] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, J. Tonry, *Astron. J* **116** (1998), 1009.

[230] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, I. M. Hook, A. G. Kim, M. Y. Kim, J. C. Lee, N. J. Nunes, R. Pain, C. R. Pennypacker, R. Quimby, C. Lidman, R. S. Ellis, M. Irwin, R. G. McMahon, P. Ruiz-Lapuente, N. Walton, B. Schaefer, B. J. Boyle, A. V. Filippenko, T. Matheson, A. S. Fruchter, N. Panagia, H. J. M. Newberg, W. J. Couch, and The Supernova Cosmology Project, *Ap. J.* **517** (1999), 565.

[231] B. A. Bassett, S. Liberati, C. Molina-Paris, M. Visser, *Los Alamos Preprints*, xxx.lanl.gov, astro-ph/0001441.

[232] W. Buchmüller, N. Dragon, *Phys. Let. B* **223** (1989), 313.

[233] N. Dragon, M. Kreuzer, *Zeit. Phys. C* **41** (1988), 485.

[234] D. Boulware, S. Deser, *Phys. Rev. D* **6** (1972), 3368.

[235] M. Visser, *Gen. Rel. Grav.* **30** (1998), 1717; *Los Alamos Preprints*, xxx.lanl.gov, qr-qc/9705051.

[236] P. G. O. Freund, Y. Nambu, *Phys. Rev.* **174** (1968), 1741.

[237] S. Deser, L. Halpern, *Gen. Rel. Grav.* **1** (1970), 131.

[238] J. J. van der Bij, H. van Dam, *Physica A* **116** (1982), 307.

[239] W. G. Unruh, *Phys. Rev. D* **40** (1989), 1048.

[240] M. Henneaux, C. Teitelboim, *Phys. Let. B* **222** (1989), 195.

[241] W. Buchmüller, N. Dragon, *Phys. Let. B* **207** (1988), 292.

[242] A. Daughton, J. Louko, R. Sorkin, *Phys. Rev. D* **58** (1998), 084008.

[243] D. L. Lee, A. P. Lightman, W.-T. Ni, *Phys. Rev. D* **10** (1974), 1685.

[244] C. Will, *Theory and Experiment in Gravitational Physics,* rev. ed. Cambridge Univ., Cambridge (1993).

[245] T. Singh, L. N. Rai, *Gen. Rel. Grav.* **15** (1983), 875.

[246] D. Lohiya, M. Sethi, *Class. Quant. Grav.* **16** (1999), 1545.

[247] J. C. Fabris, N. Pinto-Neto, A. F. Velasco, *Class. Quant. Grav.* **16** (1999), 3807.

[248] N. Bartolo, M. Pietroni, *Phys.Rev. D* **61** (2000), 023518.

[249] N. Kaloper, K. A. Olive, *Phys. Rev. D* **57** (1998), 811.

[250] Y. J. Ng, H. van Dam, *Int. J. Mod. Phys. D* **10** (2001) 49; *Los Alamos Preprints*, xxx.lanl.gov, hep-th/9911102.

[251] R. M. Avakian, L. Sh. Grigorian, *Astrophys. Space Sci.* **146** (1988), 183.

[252] P. P. Fiziev, S. S. Yazadijev, T. L. Boyadijev, M. D. Todorov, *Phys. Rev. D* **61**, 124018; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0001103.

[253] L. Sh. Grigorian, A. A. Saharian, *Astrop. Space Sci.* **180** (1991), 39.

[254] F. Fucito, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0005089; presented at the 2nd SIGRAV school on gravity, Como, Italy.

[255] R. V. Wagoner, *Phys. Rev. D* **1** (1970), 3209.

[256] G. Venziano, lectures delivered in Les Houches, July 1999. *Los Alamos Preprints*, xxx.lanl.gov, hep-th/0002094.

[257] T. Matos, F. Guzmán, L. Ureña-López, *Class. Quant. Grav.* **17** (2000), 1707.

[258] P. J. E. Peebles, A. Vilenkin, *Phys. Rev. D* **60** (1999), 103506.

[259] S. Weinberg, *Rev. Mod. Phys.* **61** (1989), 1.

[260] P. G. O. Freund, A. Maheshwari, E. Schonberg, *Ap. J.* **157** (1969), 857.

[261] V. Faraoni, E. Gunzig, P. Nardone, *Fund. Cosmic Phys.* **20** (1999) 121; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9811047.

[262] Y. Nutku, *Ap. J.* **158** (1969), 991.

[263] H. B. Hart, *Phys. Rev. D* **5** (1972), 1256.

[264] D. L. lee, *Phys. Rev. D* **10** (1974), 2374.

[265] M. A. Clayton, *Class. Quantum Grav.* **13** (1996), 2851.

[266] S. Deser, *Rev. Mod. Phys.* **29** (1957), 417.

[267] A. N. Petrov, *Mod. Phys. Let. A* **6** (1991), 2107.

[268] L. P. Grishchuk, A. N. Petrov, *Sov. Astron. Lett.* **12** (1986), 179, trans. *Pis'ma Astron. Zh.* **12** (1986), 439.

[269] A. A. Vlasov, *Moscow Uni. Phys. Bull.* **43** (1988), 72, trans. *Vestnik Moskovskogo Universiteta. Fizika.*

[270] Yu. P. Viblyi, *Theoretical and Mathematical Physics*, trans. *Teoeticheskaya i Matematicheskaya Fizika* **88** (1991), 135.

[271] J. L. Anderson, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9912051.

[272] J. M. Nester, *Int. J. Mod. Phys. A* **4** (1989), 1755.

[273] J. M. Nester, R.-S. Tung, Y. Z. Zhang, *Class. Quantum Grav.* **11** (1994), 757.

[274] V. I. Ogievetskiĭ, I. V. Polubarinov, *Sov. Phys. J. E. T. P.* **21** (1965), 1093, trans. **48** (1965), 1625.

[275] W. R. Davis. *Classical Fields, Particles, and the Theory of Relativity.* Gordon and Breach, New York (1970).

[276] M. Visser, *Lorentzian Wormholes: from Einstein to Hawking.* American Institute of Physics, Woodbury, N.Y. (1995).

[277] S. Deser, C. J. Isham, *Phys. Rev. D* **14** (1976), 2505.

[278] J. M. Charap, *Class. Quant. Grav.* **4** (1987), 247.

[279] J. Bičák, in *Einstein Field Equations and Their Physical Implications–Selected essays in honour of Juergen Ehlers*, *ed.* B. G. Schmidt, Springer, Berlin (2000); *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0004016

[280] Yu. M. Loskutov, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **94** (1993), 515.

[281] Yu. V. Chugreev, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **88** (1991), 459.

[282] K. H. E. Pinson, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **89** (1991), 317.

[283] I. P. Denisova, *Doklady Physics* **43** (1998), 297; trans. *Doklady Akademii Nauk* **360** (1998), 335.

[284] H. Goldstein, *Classical Mechanics, 2nd ed.* Addison-Wesley, Reading, Mass. (1980).

[285] C. J. Isham, in *Quantum Gravity: An Oxford Symposium*, *ed.* C. J. Isham, R. Penrose, and D. W. Sciama, Clarendon, Oxford (1975).

[286] C. J. Isham, in *Quantum Gravity 2: A Second Oxford Symposium*, *ed.* C. J. Isham, R. Penrose, and D. W. Sciama, Clarendon, Oxford (1981).

[287] S. Weinstein, in *Physics Meets Philosophy at the Planck Scale*, *ed.* C. Callender, N. Huggett. Cambridge Univ., Cambridge (2001).

[288] J. A. Schouten, *Ricci-Calculus: An Introduction to Tensor Analysis and Its Geometrical Applications, 2nd ed.* Springer, Berlin (1954).

[289] Y. Choquet-Bruhat, S. Deser, *Ann. Phys. (NY)* **81** (1973), 165.

[290] D. R. Brill, S. Deser, *Comm. Math. Phys.* **32** (1973), 291.

[291] E. Verdaguer, *Phys. Reports* **229** (1993), 1.

[292] C. J. Isham, writeup of plenary lecture, GR14 conference, Florence, 1995; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9510063.

[293] G. Belot, J. Earman, L. Ruetsche, *Brit. J. Phil. Sci.* **50** (1999), 189.

[294] A. S. Vytheeswaran, *Int. J. Mod. Phys. A* **13** (1998), 765; *Los Alamos Preprints*, xxx.lanl.gov, hep-th/9701050.

[295] N. Banerjee, R. Banerjee, *Mod. Phys. Lett. A* **11** (1996), 1919.

[296] J. Oppenheim, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0105101.

[297] R. Penrose, *Rev. Mod. Phys.* **37** (1965), 215.

[298] H. Bondi, *Nature* **179** (1957), 1072.

[299] H. Bondi, F. A. E. Pirani, I. Robinson, *Proc. Roy. Soc. A* **251** (1959), 519.

[300] N. Rosen, *Phys. Zeit. Sow.* **12** (1937), 366.

[301] A. Einstein, N. Rosen, *J. Frank. Inst.* **223** (1937), 43.

[302] H. Bondi, F. A. E. Pirani, *Proc. Roy. Soc. A* **421** (1988), 395.

[303] A. A. Vlasov, A. A. Logunov, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **66** (1986), 163.

[304] H. Ohanian, R. Ruffini, *Gravitation and Spacetime*, 2nd ed. Norton, New York (1994).

[305] J. T. Oden, J. N. Reddy, *Variational Methods in Theoretical Mechanics.* Springer, Berlin (1976).

[306] G. Del Piero, F. Maceri, *ed.*, *Unilateral Problems in Structural Analysis: Proceedings of the Second [*sic*] Meeting on Unilateral Problems in Structural Analysis*, 1983, Ravello, *ed.* International Centre for Mechanical Science # 288, Springer, Vienna (1985).

[307] G. Del Piero, F. Maceri, *ed.*, *Unilateral Problems in Structural Analysis - 2: Proceedings of the Second [*sic*] Meeting on Unilateral Problems in Structural Analysis*, 1985, Prescudin, International Centre for Mechanical Science # 304, Springer, Vienna (1987).

[308] J. J. Moreau, P. D. Panagiotopoulos, G. Strang, *ed.*, *Topics in Nonsmooth Mechanics.* Birkhäuser, Basel (1988).

[309] M. D. P. Monteiro Marques, *Differential Inclusions in Nonsmooth Mechanical Problems: Shocks and Dry Friction*, Birkhäuser, Basel (1993).

[310] E. A. Lacomba, W. M. Tulczyjew, *J. Math. Phys.* **23** (1990), 2801.

[311] A. Ibort, M. de León, E. A. Lacomba, D. M. de Diego, P. Pitanga, *J. Phys. A.* **30** (1997), 5835.

[312] F. Cantrijn, M. de León, J. C. Marrero, D. M. de Diego, *J. Math. Phys.* **40** (1999), 795.

[313] F. S. Klotz, *A Study of General Relativity in Canonical Form.* Dissertation, Syracuse University (1970).

[314] J. N. Goldberg, F. S. Klotz, *Int. J. Theor. Phys.* **7** (1973), 31.

[315] K. Sundermeyer, *Constrained dynamics.* Springer, Berlin (1982).

[316] P. A. M. Dirac, *Proc. Roy. Soc. (London) A* **246** (1958), 333.

[317] J. L. Anderson, *Phys. Rev.* **111** (1958), 965.

[318] R. T. Jantzen, in *Proc. Int. Sch. Phys. "E. Fermi" Course LXXXVI (1982) on "Gamov Cosmology," ed.* R. Ruffini, F. Melchiorri. North Holland, Amsterdam (1987); *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0001069.

[319] J. M. Pons, D. C. Salisbury, L. C. Shepley, *J. Phys. A: Math. Gen.* **32**, 419 (1999).

[320] G. S. Hall, *Arab. J. Sci. Eng.* **9** (1984), 87.

[321] G. S. Hall, in *Differential Geometry*, Banach Center, PWN–Polish Scientific Publishers, Warsaw (1984).

[322] G. S. Hall, D. A. Negm, *Int. J. Theor. Phys.* **25** (1986), 405.

[323] G. S. Hall, M. J. Rebouças, J. Santos, A. F. F. Teixeira, *Gen. Rel. Grav.* **28** (1996), 1107.

[324] C. Bona, B. Coll, J. A. Morales, *J. Math. Phys.* **33** (1992), 670.

[325] I. Goldman, *Gen. Rel. Grav.* **9** (1978), 575.

[326] K. P. Tod, *Class. Quantum. Grav.* **9** (1992), 1693.

[327] R. V. Churchill, *Trans. Amer. Math. Soc.* **34** (1932), 784.

[328] M. Kaku, *Quantum Field Theory: A Modern Introduction.* Oxford, New York (1993).

[329] G. A. Mena Marugán , M. Montejo, *Phys. Rev. D* **61** (2000), 084019.

[330] G. W. Gibbons, *Comm. Math. Phys.* **45** (1975), 191.

[331] W. Rindler, *Essential Relativity: Special, General, and Cosmological*, 2nd ed. Springer, New York (1977).

[332] C. J. Isham, A. C. Kakas, *Class. Quantum Grav.* **1** (1984), 621, 633.

[333] K. V. Kuchař, in *General Relativity and Gravitation 1992: Proceedings of the Thirteenth International Conference on General Relativity and Gravitation held at Cordoba, Argentina, 28 June - 4 July 1992, ed.* R. J. Gleiser, C. N. Kozameh, O. M. Moreschi. Institute of Physics, Bristol (1993).

[334] J. R. Klauder, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0102041.

[335] G. Y. Chee, Y. Zhang, Y. Guo, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0106053.

[336] B. Coll, F. San José Martínez, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/01006052.

[337] O. Lauscher, M. Reuter, *Los Alamos Preprints*, xxx.lanl.gov, hep-th/0108040.

[338] C. W. Misner, *Phys. Rev. A* **186** (1969), 1328.

[339] A. G. Agnese, A. Wataghin, *Lett. Nuov. Cim.* 1 (1971), 857.

[340] J.-M. Levy-Leblond, *Amer. J. Phys.* **58** (1990), 156.

[341] C. Barceló, M. Visser, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0003025.

[342] L. H. Ford, T. A. Roman, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0009076 v. 2.

[343] R. V. Churchill, J. W. Brown, *Complex Variables and Applicaitons*, (5th ed.). McGraw-Hill, New York (1990).

[344] Y. Choquet-Bruhat, C. DeWitt-Morette, M. Dillard-Bleick, *Analysis, Manifolds, and Physics. Part I: Basics*, rev. ed. North-Holland, Amsterdam (1996).

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

[346] A. Anderson, Y. Choquet-Bruhat, J. W. York Jr., to be published in the Proceedings of the 2nd Samos Meeting; *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9907099.

[347] D. Finkelstein, C. Misner, *Ann. Phys. (NY)* **6** (1959), 230.

[348] R. L. Liboff, *Introductory Quantum Mechanics*, 2nd. ed. Addison-Wesley, Reading (1931).

[349] S. Helgason, in *Battelle Rencontres 1967 Lectures in Mathematics and Physics, ed.* C. DeWitt and J. A. Wheeler. Benjamin, New York (1968).

[350] B. S. DeWitt, *Dynamical Theory of Groups and Fields.* Gordon and Breach, New York (1965).

[351] C. J. Isham, in *Canonical Gravity: From Classical to Quantum. Proceedings, of the 117th WE Heraeus Seminar Held at Bad Honnef, Germany, 13-17 September 1993, ed.* J. Ehlers, H. Friedrich. Springer, Berlin (1994); *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9310031.

[352] R. Penrose, in *Battelle Rencontres 1967 Lectures in Mathematics and Physics, ed.* C. DeWitt and J. A. Wheeler. Benjamin, New York (1968).

[353] A. K. Raychaudhuri, S. Banerji, A. Banerjee, *General Relativity, Astrophysics, and Cosmology.* Springer, New York (1992).

[354] W. Israel, in *Three Hundred Years of Gravitation*, *ed.* S. W. Hawking, W. Israel. Cambridge Univ., Cambridge (1987).

[355] K. Martel, E. Poisson, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0001069.

[356] M. E. Gertsenshtein, M. Yu. Konstantinov, *Sov. Phys. Journ.*, trans. *Izvestiya Vysshikh Uchebnykh Zavednii, Fizika* **10** (1976), 131.

[357] Ya. I. Pugachev, V. D. Gun'ko, *Sov. Phys. Journ.*, trans. *Izvestiya Vysshikh Uchebnykh Zavednii, Fizika* **10** (1974), 133.

[358] Z. Zakir, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/9905102 v.2.

[359] N. Rosen, in *Relativity: Proceedings of the Relativity Conference in the Midwest, held at Cincinnati, Ohio, June 2-6, 1969*, *ed.* M. Carmeli, S. I. Fickler, L. Witten. Plenum, New York (1970).

[360] L. Bel, *J. Math. Phys.* **10** (1969), 1501.

[361] A. I. Janis, E. T. Newman, J. Winicour, *Phys. Rev. Lett.* **20** (1968), 878.

[362] F. I. Cooperstock, G. J. G. Junevicus, *Nuov. Cim.* **16B** (1973), 387.

(1978), no. 13.

[363] P. A. Amundsen, Ø. Grøn, *Phys. Rev. D.* **27** (1983), 1731.

[364] F. Rohrlich, *Ann. Phys.* (N.Y.) **22** (1963), 169.

[365] A. A. Logunov, *Part. Phys. Nucl.* **29** (1998), 1.

[366] P. J. E. Peebles, *Principles of Physical Cosmology.* Princeton Univ., Princeton (1993).

[367] C. M. Will, *Phys. Rev. D* **57** (1998), 2061.

[368] Yu. M. Loskutov, *Theor. Math. Phys.*, trans. *Teoreticheskaya i Matematicheskaya Fizika* **107** (1996), 686.

[369] A. A. Logunov, M. A. Mestvirishvili, *Los Alamos Preprints*, xxx.lanl.gov, gr-qc/0106055.

[370] M. Fleck, H. O. Girotti, *Los Alamos Preprints*, xxx.lanl.gov, hep-th/09906019.

[371] M.-I. Park, Y.-J. Park, *Los Alamos Preprints*, xxx.lanl.gov, hep-th/9702134 v. 3

[372] Y.-W. Kim, S.-K. Kim, Y.-J. Park, *Los Alamos Preprints*, xxx.lanl.gov, hep-th/9812001.

Vita

James Brian Pitts was born in Warner Robins, Georgia in 1973, the son of William E. and Jan D. Pitts. After graduating from Beavercreek High School, Beavercreek, Ohio, in 1991, he entered the Georgia Institute of Technology in Atlanta, Georgia. He received the degree of Bachelor of Science in Physics in 1995. In 1995 he entered the Graduate School in physics at The University of Texas at Austin.

Permanent Address: 4806 Avenue F, Austin, Texas 78751

This dissertation was typeset with LaTeX 2_ε[1] by the author.

[1] LaTeX 2_ε is an extension of LaTeX. LaTeX is a collection of macros for TeX. TeX is a trademark of the American Mathematical Society. Most of the macros used in formatting this dissertation were written by Dinesh Das, Department of Computer Sciences, The University of Texas at Austin.