Space-time Structure and the Origin of Physical Law

by

Martin Andrew Green

A Thesis submitted in conformity with the requirements
for the Degree of Doctor of Philosophy
in the University of Toronto.

ABSTRACT

The laws of physics are viewed as mathematical statements which should follow from some set of fundamental principles. Included amongst these principles are basic notions of space, time and, since the development of relativity theory, space-time. In the first part of the thesis a traditional world-view is adopted, with space-time a topologically simple geometrical manifold, matter being represented by smooth classical fields, and space a Riemannian submanifold of space-time. Using a completely coordinate-free notation, it is shown how to characterize the space-time geometry in terms of fields defined on 3-dimensional space. Accepting only a finite number of the fields induced on space as independent initial data, a procedure is then given for constructing dynamical and constraint equations which will consistently and unambiguously propagate these fields forward in time. When the geometrical initial data is restricted to include only the hypersurface metric, ${}^3g$, and the extrinsic curvature, K , the resulting dynamical and constraint equations combine to form the Einstein gravitational field equations (with the cosmological term).

This is a new and very direct approach to general relativity, which shows quite clearly that the raison d'être of the Einstein field equations is to propagate the spatial metric forward in time in a consistent fashion. Higher order gravitational equations cannot be ruled out, however, nor does this investigation of the

space-time geometry provide the basis for a theory of matter. In an attempt to remove some of this arbitrariness, it is conjectured that matter fields are not observed directly, but only indirectly through their influence on the space-time geometry. This would imply the existence of a "super" already unified theory, modelled after the Misner - Wheeler already unified theory of gravity and electromagnetism [9], and it would provide an intuitive physical argument for the correctness of the Einstein equations.

The problem of synthesizing gravitational and quantum physics is approached by adopting a new and radically different world-view. It is proposed that the objective world underlying all our perceptions is a 4-dimensional topological manifold, $\mathcal{W}$, with no physically significant field structure, but instead an unconstrained and extremely complex global topology. Conventional space-time, with its geometry and quantum fields, is then a topologically simple replacement manifold for $\mathcal{W}$, with the fields on space-time replacing the topological complexities of $\mathcal{W}$. A preliminary outline of the correspondence is presented, using as its basis a remarkable similarity between a natural graphical representation of $\mathcal{W}$ and the Feynman graphs of quantum field theory. The technical problems are formidable, but if they can be overcome then this theory may be able to explain the origin of quantum phenomena and the detailed phenomenology of the elementary particles.

... It is part of the martyrdom which I endure for the cause of the Truth that there are seasons of mental weakness, when Cubes and Spheres flit away into the background of scarce-possible existences; when the Land of Three Dimensions seems almost as visionary as the Land of One or None; nay, when even this hard wall that bars me from my freedom, these very tablets on which I am writing, and all the substantial realities of Flatland itself, appear no better than the offspring of a diseased imagination, or the baseless fabric of a dream.

Edwin A. Abbott

From Flatland

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

This thesis is a theoretical investigation of the fundamental nature and form of the physical world. Holding fast to the belief that the phenomena we perceive are manifestations of some exceptionally coherent and conceptually simple mathematical structure, I have placed the emphasis throughout on philosophical ideas rather than on phenomenological details. Elementary notions of time and evolution are shown to place significant constraints on the geometric structure of space-time, yielding, in the simplest case, a new derivation of general relativity. Geometrical field theory need not be fundamental, however; and preliminary arguments are given to support the conjecture that all of field theory, including quantum phenomena and particle physics phenomenology, is subordinate to and derivable from the unconstrained topological structure of space-time.

1. First Principles

Tightly bound into our modern world-view are several intuitive notions regarding time, space, space-time, and evolution. These constitute a (still evolving) set of guiding principles that we feel should be embodied in any reasonable physical theory. Recognizing that they are subject to change and personal differences, I shall place particular emphasis, in my statement of these principles, on our <u>perception</u> of the world rather than on the world itself.

1.1 Uniqueness - All of the physical phenomena that we perceive are manifestations of a unique and definite structure, which I call the <u>world</u>.

1.2 Time - There is a natural partial ordering for the phenomena that we are capable of perceiving (henceforth called events). If two events are ordered, then one lies either to the future or to the past of the other, and they are said to be time-ordered. A future directed time-like path from an event a to a future event b is a time-ordered set of events, $T(a,b) = \{a,\ldots,b\}$, which is maximal in the sense that the addition of any event, x, that lies neither to the past of a nor the future of b would necessarily destroy the time-ordering of $T(a,b)$. When given the natural topology induced by the time-ordering, each such path is homeomorphic with the closed interval $[0,1]$ in $\mathbb{R}$.†

1.3 The Universe - Two events, x_1 and x_2, are said to be <u>space-like</u> if and only if there exist events x_3 to the past and x_4 to the future of x_2 such that neither x_3 nor x_4 lies to the past or future of x_1. If $\mathcal{U}$ is a set of

† A summary of much of the mathematics and the notation used in this thesis is provided in Appendix I.

mutually space-like events that is maximal with respect to the space-like condition, then all events in U may be simultaneously perceived, and U is called the universe at some time.

1.4 History - If U_1 and U_2 are two non-intersecting universes such that $x_1 \in U_1$ lies to the future (past) of some $x_2 \in U_2$, then U_1 is said to lie to the future (past) of U_2. A history is a time-ordered one-parameter family, H, of non-intersecting universes, $U(t)$, $t \in [a,b]$, which is maximal in the sense that every event x that lies to the future of some event in $U(a)$ and to the past of some event in $U(b)$ is contained in $U(t)$ for some $t \in (a,b)$. The sequence of universes of events of which we progressively become aware as time passes is a history.

1.5 Space and Matter - The universe appears to us to be a synthesis of two very different kinds of structures: space and matter. Serving as a stable repository for matter, space is a three-dimensional differential manifold, S, of indeterminate global topology. Matter, the stuff that "resides in space" and gives events their distinguishing features, is characterized by a countable (and perhaps finite) number of tensor fields, ${}^3F_i$, $i \in \omega$, of class C^∞ on S.

1.6 Geometry - Defined on S is a Riemannian (ie. positive definite) metric tensor field, ${}^3g$, of class C^∞. This is manifested in the sizes and shapes that we perceive to be characteristic features of all matter distributions.

1.7 Continuity - As time passes, the geometry and the matter distributions that are defined on space undergo continual change, but these changes are never discontinuous. Denote by ${}^3F_i(t)$ and ${}^3g(t)$ the matter and metric tensor fields on S that correspond to the universe $U(t) \in H$. Then for each $x \in S$ the tensors ${}^3F_i(t;x)$ and ${}^3g(t;x)$ are continuous functions of t. Moreover, there always exists a parameterization, $t = t(t')$, such that ${}^3F_i(t';x)$ and ${}^3g(t';x)$ are differentiable functions of t' of class C^∞.

1.8 Space-time - There is a unique, smooth, 4-dimensional manifold, M, which has defined on it a set of tensor fields, F_j, $j \in \omega$, such that any universe $U(t)$ may be realized as a submanifold, $S(t) \subset M$, with the fields ${}^3F_i(t)$, ${}^3g(t)$ induced by the embedding $e_t:S \to M$; $e_t(S) = S(t)$. Each history is a foliation of some region $M' \subset M$, M' having the topology of $\mathbb{R}\times S$. The manifold M, together with whatever geometric structure it may have, is called space-time.

1.9 Evolution Rules - Let $\{S(t);t \in (a,b)\}$ be a foliation of $M' \subset M$ by space-like hypersurfaces such that for each $t_0 \in (a,b)$ and all $x \in S(t_0)$, every past directed time-like path through x intersects every one of the hypersurfaces $S(t)$, for $t \in (a,t_0)$. Then, for each $t_0 \in (a,b)$, a knowledge of the fields ${}^3F_i(t_0)$ and ${}^3g(t_0)$, and a finite number of their derivatives with respect to t, on $S(t_0)$ is sufficient to completely determine these fields on any future hypersurface $S(t)$,

$t \in (t_0,b)$, and also to determine the fields F_j at all points of M' to the future of $S(t_0)$. The rules that govern this "evolution" do not depend in any way on the field configurations.

In brief form, the first of these principles, (1.1), asserts my belief in the existence of an objective reality; items (1.2) to (1.8) establish the conventional space-time-matter picture (without clocks) as the appropriate model for the space of our perceptions; and (1.9) stipulates that the universe must evolve in a predictable fashion.

2. Classical Physics

It should be clear that all of the principles (1.1) to (1.9) originate in classical (as opposed to quantum) physics, and so I shall explore their implications only in the classical context. The main results that I shall obtain in Chapter 2 are:

2.1 The space-time appropriate for modelling perceived (classical) phenomena has defined on it a locally Minkowskian metric tensor field, g , whose projection onto any space-like hypersurface provides the metric, ${}^3g$, for that universe; and

2.2 In the absence of additional fields, the metric g satisfies a covariant set of local partial differential equations in which the highest derivative of g enters linearly. When restricted to second order, these are Einstein's vacuum gravitational field equations (with the cosmological term).

These are certainly not new results, but I believe that the novel approach that I have used in their derivation provides some new and useful insight into why general relativity (GR), or a related higher order theory, should be considered as fundamentally correct.

Many different approaches have been used to derive GR in the past. Einstein made use, primarily, of the classical correspondence principle, requiring that Newtonian gravity emerge as the weak field, small velocity, limit of a generally covariant theory [1]. A Lagrangian formulation was found by Hilbert [2], and recently Hojman, Kuchař, and Teitelboim (HKT) have derived general relativity by investigating the integrability conditions of the Hamiltonian equations of motion [3]. Several authors have started from a quantum picture and used (Lorentz) gauge invariance arguments to obtain the Einstein equations [4], while Boulware and Deser have used aspects of quantum <u>particle</u> physics as their starting point [5]. A good summary of the earlier derivations is provided by Misner, Thorne, and Wheeler [6].

The present derivation is closest, in spirit, to that of HKT. Rather than assuming, at the outset, that the space-time metric, g , satisfies some set of covariant field equations on M , they focussed attention on the space-like hypersurfaces of the space-time manifold. They sought a set of evolution rules that would propagate data, defined on some initial hypersurface, forward in time onto a future space-like slice, and they stipulated that the result should be independent of the choice of intermediate space-like hypersurfaces used in the time integration. This much of the HKT formalism I have kept, although I have abandoned their notation in favour of a more intuitive, coordinate-free notation.

I deviate from HKT in the assumptions that actually fix the dynamics. They assume that (in the vacuum case) the spatial metric, ${}^3g$, and a conjugate super-momentum provide all the necessary initial data, and they postulate that the equations governing the evolution of these fields take the Hamiltonian form [7], with the Poisson bracket algebra of the super-Hamiltonian and super-momentum closing exactly as the commutator algebra of the generators of hypersurface deformations. The philosophical motivation for these assumptions is not at all clear, and in the first part of my derivation I eliminate them completely. Without any new assumptions to replace them, I obtain (2.1) and the first half of (2.2). In order to pick the Einstein equations from the myriad of possibilities, a new assumption is necessary, however, and I have chosen the more conventional route of restricting the initial data to be ${}^3g$ and its "velocity", which is denoted by K . The obvious advantage in choosing this data, as opposed to that used by HKT, is that this choice obviates the need to assume anything about a Hamiltonian structure.

Space-times with matter fields, that is, fields F_j in addition to g , are considered in Chapter 3. It is shown, using the formalism developed in Chapter 2, that matter fields are quite able to influence the evolution of the spatial geometry, and that the space-time geometry, in turn, influences the evolution of the matter fields. When the geometrical initial data is once again restricted to include only ^3g and K , the standard Einstein coupling of matter and geometry is recovered - a result that is no surprise.

The most troublesome aspect of matter, at this stage, is that we have no philosophically appealing (non-empirical) criteria for deciding what kinds of matter fields should be considered, nor what equations they should satisfy (aside from the geometric constraints mentioned above). Many attempts have been made to construct unified field theories using alternative (non-Riemannian) geometries, and so obtain a geometric picture of electromagnetism or other kinds of matter fields. But these theories can always be reformulated in terms of Riemannian geometry plus tensor fields, making the physical significance of the alternative geometries uncertain at best.

Rainich [8], and later Misner and Wheeler [9], showed that the electromagnetic field need not be considered as something in addition to the metric, however, since it could actually be extracted from a suitably constrained space-time metric. A short discussion of this "already unified" theory of gravitation and electromagnetism is presented in Chapter 3. A somewhat different interpretation of the Misner - Wheeler results is then proposed and it is argued that, if all primary physical measurements are measurements of the metric geometry of space-time, then all matter fields, not just the electromagnetic field, must leave distinctive imprints in the space-time geometry, from which all of the observable characteristics of the various fields may be recovered. Fields which do not leave such distinct imprints could be eliminated from consideration, reducing the arbitrariness of the theory eventually adopted.

3. Quantum and Particle Physics

While classical field theory, and general relativity in particular, provides a very elegant and simple formalism for describing some of the features that we perceive to be characteristic of the world, it is unable to give us any insight into the nature or origin of quantum phenomena. It cannot tell us what kinds of (quantum) particles the world is made of or how they interact with each other. Normally, it is just assumed that quantum field theory (QFT) is the correct formalism to use when investigating quantum effects; and the particular particles that are known to exist are described with the use of phenomenological models, which are constantly being updated.

I find this situation highly unsatisfactory on two counts: (1) there is no compelling logical rationale (aside from its empirical successes) to indicate that QFT (or even ordinary QM) is a reasonable formalism to adopt; and

(2) the particle physics phenomenology we adopt has no philosophical basis to guide us in our search for a fundamental theory. At the same time, though, I cannot help but be impressed by the remarkable achievements of quantum theory - achievements that must indicate to even the strongest opponents of quantum mechanics that there is something correct about this inscrutable formalism.

Why does quantum theory work? What are the guiding principles from which it follows? How can we replace particle physics phenomenology by fundamental theory? In Chapter 4, a radically new and strikingly simple view of the world is presented, which I believe may eventually provide satisfying answers to all of these questions. Carrying to its logical conclusion Einstein's lead in removing restrictive assumptions about the structure of space-time, it is proposed that <u>the objective world underlying all of our perceptions is a 4-dimensional topological manifold</u>, $\mathcal{W}$, <u>which has no physically significant geometry or field structure</u>, <u>but instead an unconstrained (and extremely complex) global topology</u>. Direct perception of the detailed structure of $\mathcal{W}$ would be analogous to observing all of the virtual particles in the quantum mechanical vacuum, which is clearly impossible. But many characteristic features of the topology of $\mathcal{W}$ are perceptible. These we interpret as fields (metric and quantum fields) on a topologically simple 4-dimensional manifold, that we call space-time. Our conventional space-time thus emerges, in this picture, as a <u>replacement manifold</u> for the objective world, $\mathcal{W}$, with the fields on space-time capturing as much information about the topology of $\mathcal{W}$ as possible (but not nearly all of that information).

The only restriction placed on the form of $\mathcal{W}$ is that it must be a 4-dimensional topological manifold. No additional physical laws are necessary, or even allowed, and therein lies the great beauty of this new world-view. Although this work is still very preliminary, it is argued that (an improved version of) quantum field theory and all of the phenomenology of the elementary particles should follow directly from a detailed study of the topology of 4-manifolds. The space-time geometry emerges in a natural way, yielding an intuitive understanding of the semi-classical coupling of gravity and matter first proposed by Møller [10], and indicating most emphatically that gravity should not be considered as another quantum field.

Unfortunately, due to the extreme difficulty of the mathematics involved and the attendant paucity of knowledge about three- and four-dimensional manifolds, it is not possible, at the present, to make testable predictions based on this new world-view. The potential for a great wealth of predictions is there, however, and I can only hope that the prospect of applications in fundamental physics will stimulate mathematicians to develop the relevant tools more rapidly.

CHAPTER 2

THE GEOMETRIC INITIAL VALUE PROBLEM

In this chapter and the next, I shall attempt to establish very general criteria that any acceptable theory of gravity must satisfy if it is to be compatible with the first principles (1.1.1) to (1.1.9). The present chapter is devoted to the investigation of vacuum space-times (on which the metric is the only fundamental field), with matter fields being added in Chapter 3. General relativity will emerge as the simplest possible acceptable theory in both the vacuum and general cases, and conjectural arguments will be given in Chapter 3 to support the claim that GR is the "correct" theory to use.

1. Space-time, Space, and Its Time Evolution

I shall begin, in this section, by developing a suitable mathematical formalism for describing the evolution of space through space-time. The conceptual picture I shall draw is not new, having been used by several authors [3],[6],[11] , but in the past a coordinate representation has always been used. My notation, here and throughout the thesis, is completely coordinate free. Topological and geometrical concepts are thus kept to the fore, and complete covariance is assured. For simplicity it is assumed that space and space-time are diffeomorphic with $\mathbb{R}^3$ and

$\mathbb{R}^4$, respectively. Global topological considerations are discussed, briefly, at the end of Chapter 3.

Let S be an open cell in 3-dimensional space; let M be the space-time manifold; and let $e:S \to M$ be a smooth embedding. Then it is always possible to construct a smooth map

$$E_D : S \to C^\infty(I_\delta, M) \quad , \tag{1.1}$$

from S into the set of C^∞ maps from the interval $I_\delta = (-\delta, \delta)$ into M , such that each of the maps

$$D_t : S \to M \quad ; \quad D_t(x) = (E_D(x))(t) \quad , \; t \in I_\delta \; , \; x \in S \; , \tag{1.2}$$

is a smooth embedding, and

$$D_0 = e \quad . \tag{1.3}$$

The hypersurfaces so defined are neighbouring in the sense that as t approaches t_0 the hypersurface $D_t(S)$ approaches $D_{t_0}(S)$ arbitrarily closely.

A differential description of the motion of $D_t(S)$ through space-time is provided by the vector valued maps

$$\vec{D}_t : S \to T(M) \quad ; \quad \vec{D}_t(x) \in T_{D_t(x)}(M)$$

which are defined by

$$\vec{D}_t f(x) = \frac{d}{dt}(f(D_t(x))) \tag{1.4}$$

for all differentiable test functions, $f:M \to R$, and all $x \in S$.

As t is increased, the hypersurface $D_t(S)$ is deformed continuously through M, with each point $D_t(x)$ following the trajectory of the vector $\vec{D}_t(x)$.

In general, the hypersurfaces $D_t(S)$ need not be disjoint surfaces in M. The image space, $D_t^{\vdash}(S)$, could even be the same for all of them, each map D_t being obtained from e by composing it with a diffeomorphism of S. I shall consider, however, only those maps E_D for which $D_t^{\vdash}(S) \cap D_s^{\vdash}(S) = \emptyset$ for all $t \in I_\delta$ and $s \neq t$. This means that the hypersurfaces $D_t(S)$ provide a foliation of some open region $M' \subset M$ which has the topology $S \times I_\delta \simeq \mathbb{R}^4$. The vector fields $\vec{D}_t$ may now all be combined to form a smooth <u>deformation vector field</u>, $\vec{D}:M \to T(M)$, defined on M':

$$\vec{D}(X) = \vec{D}_t(D_t^{-1}(X)) \quad , \quad X \in D_t(S) \quad . \tag{1.5}$$

Specification of $\vec{D}$ and e uniquely determines E_D, so in that which follows no direct reference will be made to E_D. For convenience, I shall assume that a particular choice of e and $\vec{D}$ has been made.

Consider now, a vector field $\vec{u}$ on S. Each of the maps D_t may be used to push $\vec{u}$ forward onto M, yielding parallel vector fields $D_{t*}\vec{u}$ defined on the surfaces $D_t(S)$. Because the hypersurfaces provide a foliation of M', each point in M' will have associated with it exactly one vector, thus yielding a smooth vector field on M', also denoted by $\vec{u}$, which is everywhere parallel to the hypersurfaces $D_t(S)$ and satisfies

$$\mathcal{L}_{\vec{D}}\vec{u} = 0 \quad . \tag{1.6}$$

(Wherever it is practical in that which follows, I shall use the symbols $\vec{u},\vec{v},\vec{w},\ldots$ to denote vector fields on S or parallel vector fields on M , and the symbols $\vec{U},\vec{V},\vec{W},\ldots$ to denote general (not necessarily parallel) vector fields on M . Other symbols will be treated individually.)

Now, if M is the space-time manifold, then in accordance with (1.1.8) it must have defined on it a set of smooth tensor fields, F_j , $j \in \omega$, such that each universe U may be realized as a submanifold of M , with the fields of U determined by the fields on M . In particular, since I want to consider $e(S)$ as space at some time, there must be a symmetric field g of type (0,2) defined on a region of M containing $e(S)$, such that the pullback e^*g is a Riemannian metric on S . For the purposes of this chapter, I shall assume that g is the only fundamental physical field defined on M' .

Knowing e^*g , it is always possible to construct a triad field of orthonormal basis vectors, $\vec{h}_i$, on S :

$$e^*g(\vec{h}_i,\vec{h}_j) = \delta_{ij} \quad , \quad i,j = 1,2,3 \quad . \tag{1.7}$$

These can be pushed forward onto $e(S)$, and used there to construct an operator, Π , which projects space-time vector fields $\vec{U},\vec{V},\ldots,\vec{W}$ onto the hypersurface:

$$\Pi(\vec{U}) = \sum_{i=1}^{3} g(\vec{U}, e_*\vec{h}_i) e_*\vec{h}_i \quad . \tag{1.8}$$

Space-time vectors defined at points of $e(S)$ may thus be decomposed into <u>parallel</u> and <u>perpendicular</u> parts, even if g is singular:

$$\vec{U} = \vec{U}^{\parallel} + \vec{U}^{\perp} \quad , \qquad \vec{U}^{\parallel} = \Pi(\vec{U}) \quad . \tag{1.9}$$

The vector field $\vec{u}$ defined above is, of course, already parallel: $\vec{u} = \Pi(\vec{u})$. We may also use Π to construct, on $e(S)$, a hypersurface metric, ${}^3g$, which will prove to be much more convenient than e^*g . It is defined by

$${}^3g(\vec{U},\vec{V}) = g(\vec{U}^{\parallel},\vec{V}^{\parallel}) = g(\Pi(\vec{U}),\Pi(\vec{V})) \quad . \tag{1.10}$$

The definition of space-like given in (1.1.3) guarantees that if $e(S)$ is space-like (which I assume to be the case), then any hypersurface obtained from $e(S)$ by a smooth infinitesimal deformation is also space-like; but this property does not necessarily hold for finite deformations. I shall assume, however, that $\vec{D}$ has been carefully chosen so that all of the hypersurfaces $D_t(S)$ are space-like, deferring until later a discussion of the constraints imposed on $\vec{D}$ in order to assure this. With this assumption, it is clear that the pullback, ${}^3g_t = D_{t*}g$, of g onto S , corresponding to each of the hypersurfaces $D_t(S)$, is positive definite; and both Π and ${}^3g$ may be extended smoothly to all of M' . Of course, this extension is not unique, depending as it does on the particular choice of $\vec{D}$.

2. Initial Data

The tensor field ^3g_t provides a complete description of the Riemannian geometry of space at the time labelled by t . As t increases, the field changes continuously, its rate of change being given by

$$\frac{d}{dt}(^3g_t(\vec{u},\vec{v}))(x) = \pounds_{\vec{D}}g(\vec{u},\vec{v})(D_t(x))$$

$$= \pounds_{\vec{D}^{\parallel}}{}^3g(\vec{u},\vec{v})(\dot{D}_t(x)) + \pounds_{\vec{D}^{\perp}}g(\vec{u},\vec{v})(D_t(x)) \quad , \tag{2.1}$$

where $\vec{D}^{\parallel} = \Pi(\vec{D})$, $\vec{D}^{\perp} = \vec{D} - \vec{D}^{\parallel}$, and $\vec{u},\vec{v}$ on the left are arbitrary vector fields on S , and on the right they are the corresponding induced fields on M , which satisfy $\pounds_{\vec{D}}\vec{u} = \pounds_{\vec{D}}\vec{v} = 0$. Considering just the "initial surface", $e(S)$, we see that once $\vec{D}^{\parallel}$ has been chosen, the first term on the right is uniquely determined by the hypersurface metric, ^3g , which is equivalent to e^*g . The second term, however, is not determined by known data, nor is it invariantly defined, depending as it does on the component, $\vec{D}^{\perp}$, of the deformation vector field that cannot (yet) be specified from within the initial surface. Higher order derivatives, constructed by iterating (2.1), depend on $\vec{D}$ in still more complicated ways.

All by itself, the field ^3g , restricted to $e(S)$, provides insufficient data to determine the geometry of space at any other time; but the only other tensor fields available on $e(S)$ are the time derivatives discussed above. If we are to be able to construct the dynamical theory that is demanded by (1.1.9), then we must find some way of extracting the deformation dependence from these time derivatives to leave behind invariantly defined physical data. It is not difficult to see that this separation can only be achieved if $\vec{D}^{\perp}$ can be expressed in the form

$$\vec{D}^{\perp} = D^{\perp}\vec{n} \quad , \tag{2.2}$$

where $D^{\perp}$ is a scalar field (which can be specified from within $e(S)$) and $\vec{n}$ is a perpendicular vector field that is completely determined on $e(S)$ by g and the embedding, e. (In Newtonian gravity $\vec{n}$ would be an absolute time-like vector field, but that is excluded here by the assumption that all fields are dynamic.) Assuming such an $\vec{n}$, the offending last term in (2.1) can be written in the form

$$\begin{aligned} \mathcal{L}_{\vec{D}^{\perp}} g(\vec{u},\vec{v}) &= \vec{D}^{\perp}(g(\vec{u},\vec{v})) - g(\mathcal{L}_{\vec{D}^{\perp}}\vec{u},\vec{v}) - g(\vec{u},\mathcal{L}_{\vec{D}^{\perp}}\vec{v}) \\ &= D^{\perp}\vec{n}(g(\vec{u},\vec{v})) - g(D^{\perp}\mathcal{L}_{\vec{n}}\vec{u} - \vec{u}D^{\perp}\vec{n},\vec{v}) - g(\vec{u},D^{\perp}\mathcal{L}_{\vec{n}}\vec{v} - \vec{v}D^{\perp}\vec{n}) \\ &= D^{\perp}\mathcal{L}_{\vec{n}} g(\vec{u},\vec{v}) \quad . \end{aligned} \tag{2.3}$$

The tensor ${}^3g'$, defined by

$${}^3g'(\vec{U},\vec{V}) = \mathcal{L}_{\vec{n}} g(\Pi(\vec{U}),\Pi(\vec{V})) \tag{2.4}$$

is thus the invariantly defined piece of initial data that allows us to complete the specification, from within $e(S)$, of the rate of change of ^3g :

$$\frac{d}{dt}({}^3g_t(\vec{u},\vec{v}))\Big|_{t=0} = \pounds_{\vec{D}^{\parallel}}\,{}^3g(\vec{u},\vec{v}) + D^{\perp}\,{}^3g'(\vec{u},\vec{v}) \quad . \tag{2.5}$$

The fields $\vec{n}$ and $^3g'$, just like Π and ^3g , may be extended to all of M' by constructing them on each of the hypersurfaces $D_t(S)$. Equation (2.4) remains the defining equation for $^3g'$, while $\vec{n}$ always satisfies $\Pi(\vec{n}) = 0$.

As stated above, the field $\vec{n}$ must be completely determined by the space-time field g and the set of embedding maps, D_t . Since space is of co-dimension one in space-time, it is possible, using only the properties of g assumed in the previous section, to uniquely determine the direction of $\vec{n}$ at each point of M' . However, in order to fix the magnitudes of these vectors we must make the additional assumption that g is everywhere non-singular. With this condition, $\vec{n}$ may be chosen to be a field of unit vectors that are everywhere normal to the hypersurfaces $D_t(S)$:

$$\left.\begin{aligned} &g(\vec{n},\vec{n}) = \varepsilon \quad , \quad \varepsilon = \pm 1 \quad ; \\ &g(\vec{n},\Pi(\vec{U})) = 0 \quad \text{for all } \vec{U} \in \mathcal{T}(M) \quad . \end{aligned}\right\} \tag{2.6}$$

The sign, ε , of $g(\vec{n},\vec{n})$ need not be specified at this point, but must eventually be set to -1 in order to accommodate Dirac spinors and to prevent space-like events from being causally related.

Being non-singular, g is a (pseudo-)Riemannian metric on M'. It uniquely determines a covariant derivative, ∇, while the hypersurface metric, ${}^3g$, determines a covariant derivative, ${}^3\nabla$. These satisfy

$$\left.\begin{aligned} &\vec{U}(g(\vec{V},\vec{W})) = g(\nabla_{\vec{U}}\vec{V},\vec{W}) + g(\vec{V},\nabla_{\vec{U}}\vec{W}) \quad , \\ &\vec{u}({}^3g(\vec{v},\vec{w})) = {}^3g({}^3\nabla_{\vec{u}}\vec{v},\vec{w}) + {}^3g(\vec{v},{}^3\nabla_{\vec{u}}\vec{w}) \quad , \\ &{}^3\nabla_{\vec{u}}\vec{v} = \Pi({}^3\nabla_{\vec{u}}\vec{v}) \quad , \end{aligned}\right\} \tag{2.7}$$

and are uniquely determined if their torsions are set to zero:

$$[\vec{U},\vec{V}] = \nabla_{\vec{U}}\vec{V} - \nabla_{\vec{V}}\vec{U} \quad , \qquad [\vec{u},\vec{v}] = {}^3\nabla_{\vec{u}}\vec{v} - {}^3\nabla_{\vec{v}}\vec{u} \quad . \tag{2.8}$$

They are related by

$$\nabla_{\vec{u}}\vec{v} = {}^3\nabla_{\vec{u}}\vec{v} + K(\vec{u},\vec{v})\vec{n} \quad , \tag{2.9}$$

where $K(\vec{U},\vec{V}) = K(\vec{V},\vec{U}) = K(\Pi(\vec{U}),\Pi(\vec{V}))$ is the extrinsic curvature tensor. K is actually not new, since

$$\begin{aligned} {}^3g'(\vec{U},\vec{V}) &= \pounds_{\vec{n}}g(\Pi(\vec{U}),\Pi(\vec{V})) \\ &= \vec{n}(g(\Pi(\vec{U}),\Pi(\vec{V})) - g(\pounds_{\vec{n}}(\Pi(\vec{U})),\Pi(\vec{V})) - g(\Pi(\vec{U}),\pounds_{\vec{n}}(\Pi(\vec{V}))) \\ &= g(\nabla_{\Pi(\vec{U})}\vec{n},\Pi(\vec{V})) + g(\Pi(\vec{U}),\nabla_{\Pi(\vec{V})}\vec{n}) \\ &= -g(\vec{n},\nabla_{\Pi(\vec{U})}\Pi(\vec{V})) - g(\nabla_{\Pi(\vec{V})}\Pi(\vec{U}),\vec{n}) \\ &= -2\varepsilon K(\vec{U},\vec{V}) \quad . \end{aligned} \tag{2.10}$$

The second line, here, is obtained with the use of Leibniz' rule for $\pounds$; the third line uses Leibniz' rule for ∇ and the first half of (2.8); and the fourth line again uses Leibniz' rule for ∇ and the second half of (2.6). This repeated use of Leibniz' rule and the free interchange of Lie and covariant derivatives, made possible by equations (2.8), will characterize many of the calculations that follow.

Also used extensively are the following relations. Let $\vec{u}$ be a parallel vector field: $\Pi(\vec{u}) = \vec{u}$. Then †

$$[\vec{n},\vec{u}] = \frac{1}{D^{\perp}}(\pounds_{\vec{D}^{\perp}}\vec{u} + \vec{u}D^{\perp}\vec{n}) \quad . \tag{2.11}$$

It is obvious that both $\pounds_{\vec{D}}\vec{u}$ and $\pounds_{\vec{D}^{\parallel}}\vec{u}$ are parallel; but this implies that $\pounds_{\vec{D}^{\perp}}\vec{u} = \pounds_{\vec{D}}\vec{u} - \pounds_{\vec{D}^{\parallel}}\vec{u}$ is also a parallel vector field. Thus

$$\Pi([\vec{n},\vec{u}]) = \frac{1}{D^{\perp}}\pounds_{\vec{D}^{\perp}}\vec{u} \quad . \tag{2.12}$$

Since $\vec{n}$ is a unit vector it must satisfy

$$g(\vec{n},\nabla_{\vec{n}}\vec{n}) = g(\vec{n},\nabla_{\vec{u}}\vec{n}) = 0 \quad ; \tag{2.13}$$

so the perpendicular part of (2.11), taken with (2.8), gives

$$g(\vec{n},\nabla_{\vec{n}}\vec{u}) = \varepsilon\frac{1}{D^{\perp}}\vec{u}D^{\perp} \quad . \tag{2.14}$$

Now let $\vec{h}_i$, $i = 1,2,3$, be a triad of orthonormal parallel

† I will use the notations $\pounds_{\vec{U}}\vec{V}$ and $[\vec{U},\vec{V}]$ interchangeably.

vector fields on M' :

$$\Pi(\vec{h}_i) = \vec{h}_i \quad , \qquad {}^3g(\vec{h}_i,\vec{h}_j) = g(\vec{h}_i,\vec{h}_j) = \delta_{ij} \quad . \tag{2.15}$$

The quadruple $(\vec{n},\vec{h}_i)$ is then a vierbein for g , in terms of which any vector field may be expanded. In particular

$$\begin{aligned} \nabla_{\vec{u}}\vec{n} &= \sum_i g(\nabla_{\vec{u}}\vec{n},\vec{h}_i)\vec{h}_i + \varepsilon g(\nabla_{\vec{u}}\vec{n},\vec{n})\vec{n} \\ &= -\sum_i g(\vec{n},\nabla_{\vec{u}}\vec{h}_i)\vec{h}_i \\ &= -\sum_i K(\vec{u},\vec{h}_i)\vec{h}_i \quad , \end{aligned} \tag{2.16}$$

$$\begin{aligned} \nabla_{\vec{n}}\vec{n} &= \sum_i g(\nabla_{\vec{n}}\vec{n},\vec{h}_i)\vec{h}_i + \varepsilon g(\nabla_{\vec{n}}\vec{n},\vec{n})\vec{n} \\ &= -\sum_i g(\vec{n},\nabla_{\vec{n}}\vec{h}_i)\vec{h}_i \\ &= -\varepsilon\frac{1}{D^{\perp}} \sum_i \vec{h}_i D^{\perp}\vec{h}_i \quad . \end{aligned} \tag{2.17}$$

A feature of these relations that will arise repeatedly is that any derivative along $\vec{n}$, if it is not determined by some identity (as in (2.13)), must depend explicitly on the deformation vector field.

Associated with ∇ is the Riemann curvature tensor, R , of space-time:

$$R(\vec{U},\vec{V})\vec{W} = \nabla_{\vec{U}}\nabla_{\vec{V}}\vec{W} - \nabla_{\vec{V}}\nabla_{\vec{U}}\vec{W} - \nabla_{[\vec{U},\vec{V}]}\vec{W} \quad . \tag{2.18}$$

Similarly, the hypersurface curvature tensor, ${}^3R$, is constructed with the use of ${}^3\nabla$. Its non-zero components are given by

$${}^3R(\vec{u},\vec{v})\vec{w} = {}^3\nabla_{\vec{u}}{}^3\nabla_{\vec{v}}\vec{w} - {}^3\nabla_{\vec{v}}{}^3\nabla_{\vec{u}}\vec{w} - {}^3\nabla_{[\vec{u},\vec{v}]}\vec{w} \quad . \tag{2.19}$$

If we maintain the convention that $\vec{u},\vec{v},\vec{w},\ldots$ are hypersurface (parallel) vector fields, then we find that there are four distinct kinds of terms into which R may be decomposed: $R(\vec{u},\vec{v})\vec{w}$, $R(\vec{u},\vec{v})\vec{n}$, $R(\vec{n},\vec{v})\vec{w}$, and $R(\vec{n},\vec{v})\vec{n}$. These may be expanded, in terms of hypersurface fields, as follows:

$$R(\vec{u},\vec{v})\vec{w} = {}^3R(\vec{u},\vec{v})\vec{w} + \{{}^3\nabla_{\vec{u}}K(\vec{v},\vec{w}) - {}^3\nabla_{\vec{v}}K(\vec{u},\vec{w})\}\vec{n}$$
$$+ \varepsilon\sum_i\{K(\vec{u},\vec{w})K(\vec{v},\vec{h}_i) - K(\vec{v},\vec{w})K(\vec{u},\vec{h}_i)\}\vec{h}_i \quad , \tag{2.20}$$

$$R(\vec{u},\vec{v})\vec{n} = -\varepsilon\sum_i\{{}^3\nabla_{\vec{u}}K(\vec{v},\vec{h}_i) - {}^3\nabla_{\vec{v}}K(\vec{u},\vec{h}_i)\}\vec{h}_i \quad , \tag{2.21}$$

$$R(\vec{n},\vec{v})\vec{n} = \nabla_{\vec{n}}\nabla_{\vec{v}}\vec{n} - \nabla_{\vec{v}}\nabla_{\vec{n}}\vec{n} - \nabla_{[\vec{n},\vec{v}]}\vec{n}$$
$$= \nabla_{\vec{n}}(-\varepsilon\sum_i K(\vec{v},\vec{h}_i)\vec{h}_i) - \nabla_{\vec{v}}(-\varepsilon\frac{1}{D^\perp}\sum_i \vec{h}_i D^\perp \vec{h}_i)$$
$$+ \varepsilon\frac{1}{D^\perp}\sum_i K(\pounds_{\vec{D}^\perp}\vec{v},\vec{h}_i)\vec{h}_i + \varepsilon\left(\frac{1}{D^\perp}\right)^2 \vec{v}D^\perp\sum_i \vec{h}_i D^\perp \vec{h}_i$$
$$= -\varepsilon\sum_i\{\nabla_{\vec{n}}K(\vec{v},\vec{h}_i) - \frac{1}{D^\perp}(\vec{h}_i\vec{v}D^\perp - {}^3\nabla_{\vec{h}_i}\vec{v}D^\perp)$$
$$- \varepsilon\sum_j K(\vec{v},\vec{h}_j)K(\vec{h}_i,\vec{h}_j)\}\vec{h}_i \quad , \tag{2.22}$$

$$R(\vec{n},\vec{v})\vec{w} = \sum_i g(\vec{h}_i,R(\vec{n},\vec{v})\vec{w})\vec{h}_i + \varepsilon g(\vec{n},R(\vec{n},\vec{v})\vec{w})\vec{n}$$
$$= \sum_i g(\vec{n},R(\vec{h}_i,\vec{w})\vec{v})\vec{h}_i - \varepsilon g(\vec{w},R(\vec{n},\vec{v})\vec{n})\vec{n}$$
$$= \varepsilon\sum_i\{{}^3\nabla_{\vec{h}_i}K(\vec{w},\vec{v}) - {}^3\nabla_{\vec{w}}K(\vec{h}_i,\vec{v})\}\vec{h}_i + \{\nabla_{\vec{n}}K(\vec{w},\vec{v})$$
$$- \frac{1}{D^\perp}(\vec{v}\vec{w}D^\perp - {}^3\nabla_{\vec{v}}\vec{w}D^\perp) - \varepsilon\sum_i K(\vec{v},\vec{h}_i)K(\vec{w},\vec{h}_i)\}\vec{n} \quad . \tag{2.23}$$

Equations (2.20) and (2.21) are the classical equations of Gauss and Codazzi, and are derived in Appendix I. Equation (2.22),on the other hand, is a new result made possible by the introduction of the deformation vector field. In it, the implicit deformation dependence of $\nabla_{\vec{n}}K$ is balanced by the terms that depend explicitly on $\vec{D}$.

The Ricci tensor, S , of space-time is obtained by contracting the curvature, R :

$$S(\vec{U},\vec{V}) = \sum_i g(\vec{h}_i, R(\vec{h}_i,\vec{U})\vec{V}) + \varepsilon g(\vec{n}, R(\vec{n},\vec{U})\vec{V}) \quad ; \tag{2.24}$$

and its components may be written in the expanded forms

$$S(\vec{u},\vec{v}) = {}^3S(\vec{u},\vec{v}) - \varepsilon K(\vec{u},\vec{v})\sum_i K(\vec{h}_i,\vec{h}_i) + \nabla_{\vec{n}}K(\vec{u},\vec{v}) - \frac{1}{D^\perp}(\vec{u}\vec{v}D^\perp - {}^3\nabla_{\vec{u}}\vec{v}D^\perp) \quad , \tag{2.25}$$

$$S(\vec{u},\vec{n}) = \varepsilon\sum_i\{ {}^3\nabla_{\vec{u}}K(\vec{h}_i,\vec{h}_i) - {}^3\nabla_{\vec{h}_i} K(\vec{u},\vec{h}_i)\} \quad , \tag{2.26}$$

$$S(\vec{n},\vec{n}) = -\sum_{i,j} K(\vec{h}_i,\vec{h}_j)K(\vec{h}_i,\vec{h}_j) + \varepsilon\sum_i\{\nabla_{\vec{n}}K(\vec{h}_i,\vec{h}_i) - \frac{1}{D^\perp}(\vec{h}_i\vec{h}_i D^\perp - {}^3\nabla_{\vec{h}_i}\vec{h}_i D^\perp)\} \quad , \tag{2.27}$$

where ${}^3S(\vec{u},\vec{v})$ is the Ricci tensor associated with ${}^3R$. Contracting again yields the Ricci scalar:

$$\begin{aligned} S &= \sum_i S(\vec{h}_i,\vec{h}_i) + \varepsilon S(\vec{n},\vec{n}) \\ &= {}^3S - \varepsilon\sum_{i,j}\{K(\vec{h}_i,\vec{h}_j)K(\vec{h}_i,\vec{h}_j) + K(\vec{h}_i,\vec{h}_i)K(\vec{h}_j,\vec{h}_j)\} \\ &\quad + 2\sum_i\{\nabla_{\vec{n}}K(\vec{h}_i,\vec{h}_i) - \frac{1}{D^\perp}(\vec{h}_i\vec{h}_i D^\perp - {}^3\nabla_{\vec{h}_i}\vec{h}_i D^\perp)\} \quad , \end{aligned} \tag{2.28}$$

in which ${}^3S = \sum_i {}^3S(\vec{h}_i,\vec{h}_i)$.

Substituting (2.25) into (2.22) and (2.23) gives, respectively

$$R(\vec{n},\vec{v})\vec{n} = -\varepsilon\sum_i\{S(\vec{v},\vec{h}_i) - {}^3S(\vec{v},\vec{h}_i)$$

$$+ \varepsilon\sum_j(K(\vec{v},\vec{h}_i)K(\vec{h}_j,\vec{h}_j) - K(\vec{v},\vec{h}_j)K(\vec{h}_i,\vec{h}_j))\}\vec{h}_i \quad ; \qquad (2.29)$$

$$R(\vec{n},\vec{v})\vec{w} = \varepsilon\sum_i\{{}^3\nabla_{\vec{h}_i}K(\vec{w},\vec{v}) - {}^3\nabla_{\vec{w}}K(\vec{h}_i,\vec{v})\}\vec{h}_i + \{S(\vec{w},\vec{v}) - {}^3S(\vec{w},\vec{v})$$

$$+ \varepsilon\sum_i(K(\vec{w},\vec{v})K(\vec{h}_i,\vec{h}_i) - K(\vec{w},\vec{h}_i)K(\vec{v},\vec{h}_i))\}\vec{n} \quad . \qquad (2.30)$$

Thus, once the hypersurface metric, ${}^3g$, and the extrinsic curvature, K , have been determined, the only new (independent) data needed to complete the specification of R from within a given hypersurface are the hypersurface components, $S(\vec{u},\vec{v})$, of the space-time Ricci tensor.

If all the components of the space-time curvature are known, then they can be used to change the order of multiple covariant derivatives, allowing the calculation of quantities that would be otherwise inaccessible. I shall give a few examples which are of later importance. The simplest example is $\nabla_{\vec{n}}{}^3\nabla_{\vec{u}}\vec{v}$:

$$\nabla_{\vec{n}}{}^3\nabla_{\vec{u}}\vec{v} = \nabla_{\vec{n}}(\nabla_{\vec{u}}\vec{v} - K(\vec{u},\vec{v})\vec{n})$$

$$= R(\vec{n},\vec{u})\vec{v} + \nabla_{\vec{u}}\nabla_{\vec{n}}\vec{v} + \nabla_{[\vec{n},\vec{u}]}\vec{v} - \nabla_{\vec{n}}K(\vec{u},\vec{v})\vec{n} - K(\nabla_{\vec{n}}\vec{u},\vec{v})\vec{n}$$

$$- K(\vec{u},\nabla_{\vec{n}}\vec{v})\vec{n} - K(\vec{u},\vec{v})\nabla_{\vec{n}}\vec{n}$$

$$= \varepsilon\sum_i\{{}^3\nabla_{\vec{h}_i}K(\vec{v},\vec{u}) - {}^3\nabla_{\vec{v}}K(\vec{h}_i,\vec{u})\}\vec{h}_i - \varepsilon\sum_i K(\vec{u},\vec{h}_i)K(\vec{v},\vec{h}_i)\vec{n}$$

$$+ \nabla_{\vec{u}}(\frac{1}{D^\perp}\pounds_{\vec{D}^\perp}\vec{v} + \frac{1}{D^\perp}\vec{v}D^\perp\vec{n} + \nabla_{\vec{v}}\vec{n}) + \frac{1}{D^\perp}\nabla_{\pounds_{\vec{D}^\perp}\vec{u}}\vec{v} + \frac{1}{D^\perp}\vec{u}D^\perp\nabla_{\vec{n}}\vec{v}$$

$$-\{\frac{1}{D^\perp}(\vec{u}\vec{v}D^\perp - {}^3\nabla_{\vec{u}}\vec{v}D^\perp) + \frac{1}{D^\perp}K(\pounds_{\vec{D}^\perp}\vec{u},\vec{v}) + K(\nabla_{\vec{u}}\vec{n},\vec{v}) + \frac{1}{D^\perp}K(\vec{u},\pounds_{\vec{D}^\perp}\vec{v})$$

$$+ K(\vec{u},\nabla_{\vec{v}}\vec{n})\}\vec{n} + \varepsilon\frac{1}{D^\perp}K(\vec{u},\vec{v})\sum_i \vec{h}_i D^\perp \vec{h}_i$$

$$= \varepsilon\sum_i\{{}^3\nabla_{\vec{h}_i}K(\vec{v},\vec{u}) - {}^3\nabla_{\vec{v}}K(\vec{h}_i,\vec{u})\}\vec{h}_i + \frac{1}{D^\perp}{}^3\nabla_{\vec{u}}\pounds_{\vec{D}^\perp}\vec{v} - \varepsilon\frac{1}{D^\perp}\vec{v}D^\perp\sum_i K(\vec{u},\vec{h}_i)\vec{h}_i$$

$$- \varepsilon\nabla_{\vec{u}}(\sum_i K(\vec{v},\vec{h}_i)\vec{h}_i) + \frac{1}{D^\perp}{}^3\nabla_{\pounds_{\vec{D}^\perp}\vec{u}}\vec{v} - \varepsilon\frac{1}{D^\perp}\vec{u}D^\perp\sum_i K(\vec{v},\vec{h}_i)\vec{h}_i + \frac{1}{D^\perp}{}^3\nabla_{\vec{u}}\vec{v}D^\perp\vec{n}$$

$$+ \varepsilon\sum_i K(\vec{u},\vec{h}_i)K(\vec{v},\vec{h}_i)\vec{n} + \varepsilon\frac{1}{D^\perp}K(\vec{u},\vec{v})\sum_i \vec{h}_i D^\perp \vec{h}_i$$

$$= \varepsilon\sum_i\{{}^3\nabla_{\vec{h}_i}K(\vec{v},\vec{u}) - {}^3\nabla_{\vec{v}}K(\vec{h}_i,\vec{u}) - {}^3\nabla_{\vec{u}}K(\vec{v},\vec{h}_i) - K({}^3\nabla_{\vec{u}}\vec{v},\vec{h}_i)\}\vec{h}_i$$

$$+ \frac{1}{D^\perp}\{{}^3\nabla_{\pounds_{\vec{D}^\perp}\vec{u}}\vec{v} + {}^3\nabla_{\vec{u}}\pounds_{\vec{D}^\perp}\vec{v} + {}^3\nabla_{\vec{u}}\vec{v}D^\perp\vec{n}\}$$

$$+ \varepsilon\frac{1}{D^\perp}\sum_i\{\vec{h}_i D^\perp K(\vec{u},\vec{v}) - \vec{u}D^\perp K(\vec{h}_i,\vec{v}) - \vec{v}D^\perp K(\vec{u},\vec{h}_i)\}\vec{h}_i \quad . \qquad (2.31)$$

Before proceeding to more complicated examples, I must digress for a moment to establish new notation. Let T be an arbitrary tensor field on M', and let T_{co} be the associated covariant tensor. I shall denote by ΠT the tensor field on M' that has the same type as T and whose covariant components are defined by

$$(\Pi T)_{co}(\vec{U},\vec{V},\ldots) = T_{co}(\Pi(\vec{U}),\Pi(\vec{V}),\ldots) \quad . \qquad (2.32)$$

We already have $\Pi\,{}^3g = {}^3g$ and $\Pi K = K$. By assuming that ${}^3\nabla$ has the generalized action

$${}^3\nabla_{\vec{U}}\vec{V} = {}^3\nabla_{\Pi(\vec{U})}\Pi(\vec{V}) \quad , \qquad (2.33)$$

we obtain, in addition, $\Pi\,{}^3R = {}^3R$ and $\Pi\,{}^3S = {}^3S$. This assumption is necessary in order to avoid ambiguities.

I shall also adopt a simplifying notation for the hypersurface derivatives of hypersurface tensors. Let $T = \Pi T$ be an arbitrary covariant hypersurface tensor field. Then the tensor field T_1 will be defined by

$$T_1(\vec{U};\vec{V},\vec{W},\ldots) = {}^3\nabla_{\vec{U}}T(\vec{V},\vec{W},\ldots) \quad , \tag{2.34}$$

the tensor field T_2 by

$$T_2(\vec{X},\vec{U};\vec{V},\vec{W},\ldots) = {}^3\nabla_{\vec{X}}{}^3\nabla_{\vec{U}}T(\vec{V},\vec{W},\ldots) - {}^3\nabla_{{}^3\nabla_{\vec{X}}\vec{U}}T(\vec{V},\vec{W},\ldots) \quad , \tag{2.35}$$

and so on for higher derivatives. By virtue of (2.33), all of these fields satisfy $\Pi T_i = T_i$, $i \in \omega$.

Repeated application of the techniques used in the derivation of (2.31) now yields the following additional examples:

$$\begin{aligned}
\nabla_{\vec{n}}K_1(\vec{w};\vec{u},\vec{v}) &= (\Pi S)_1(\vec{w};\vec{u},\vec{v}) - {}^3S_1(\vec{w};\vec{u},\vec{v}) \\
&+ \varepsilon\sum_i\{K(\vec{h}_i,\vec{h}_i)K_1(\vec{w};\vec{u},\vec{v}) + K(\vec{u},\vec{v})K_1(\vec{w};\vec{h}_i,\vec{h}_i) + K(\vec{w},\vec{h}_i)K_1(\vec{h}_i;\vec{u},\vec{v}) \\
&+ K(\vec{u},\vec{h}_i)[K_1(\vec{v};\vec{w},\vec{h}_i) - K_1(\vec{h}_i;\vec{w},\vec{v})] + K(\vec{v},\vec{h}_i)[K_1(\vec{u};\vec{w},\vec{h}_i) \\
&- K_1(\vec{h}_i;\vec{w},\vec{u})]\} + \frac{1}{D^\perp}\vec{w}D^\perp\{\Pi S(\vec{u},\vec{v}) - {}^3S(\vec{u},\vec{v}) + \varepsilon K(\vec{u},\vec{v})\sum_i K(\vec{h}_i,\vec{h}_i)\} \\
&+ \varepsilon\frac{1}{D^\perp}\sum_i\{\vec{u}D^\perp K(\vec{w},\vec{h}_i)K(\vec{v},\vec{h}_i) + \vec{v}D^\perp K(\vec{w},\vec{h}_i)K(\vec{u},\vec{h}_i) \\
&- \vec{h}_iD^\perp[K(\vec{w},\vec{u})K(\vec{v},\vec{h}_i) + K(\vec{w},\vec{v})K(\vec{u},\vec{h}_i)]\} + \frac{1}{D^\perp}\{\vec{w}\vec{u}\vec{v}D^\perp - \vec{w}\,{}^3\nabla_{\vec{u}}\vec{v}D^\perp \\
&- {}^3\nabla_{\vec{w}}\vec{u}\vec{v}D^\perp + {}^3\nabla_{{}^3\nabla_{\vec{w}}\vec{u}}\vec{v}D^\perp - {}^3\nabla_{\vec{w}}\vec{v}\vec{u}D^\perp + {}^3\nabla_{{}^3\nabla_{\vec{w}}\vec{v}}\vec{u}D^\perp\} \quad ,
\end{aligned} \tag{2.36}$$

$$\begin{aligned}
\nabla_{\vec n}{}^3R(\vec u,\vec v)\vec w = {} & \varepsilon\sum_i\{K(\vec u,\vec h_i)\,{}^3R(\vec h_i,\vec v)\vec w + K(\vec v,\vec h_i)\,{}^3R(\vec u,\vec h_i)\vec w \\
& + K(\vec w,\vec h_i)\,{}^3R(\vec u,\vec v)\vec h_i\} + \varepsilon\sum_i\{-K_2(\vec u,\vec v;\vec w,\vec h_i) - K_2(\vec u,\vec w;\vec v,\vec h_i) \\
& + K_2(\vec u,\vec h_i;\vec v,\vec w) + K_2(\vec v,\vec u;\vec w,\vec h_i) + K_2(\vec v,\vec w;\vec u,\vec h_i) - K_2(\vec v,\vec h_i;\vec u,\vec w) \\
& + \frac{1}{D^\perp}\vec u D^\perp[K_1(\vec h_i;\vec v,\vec w) - K_1(\vec w;\vec v,\vec h_i)] - \frac{1}{D^\perp}\vec v D^\perp[K_1(\vec h_i;\vec u,\vec w) \\
& - K_1(\vec w;\vec u,\vec h_i)] + \frac{1}{D^\perp}\vec w D^\perp[K_1(\vec v;\vec u,\vec h_i) - K_1(\vec u;\vec v,\vec h_i)] \\
& + \frac{1}{D^\perp}\vec h_i D^\perp[K_1(\vec u;\vec v,\vec w) - K_1(\vec v;\vec u,\vec w)] + \frac{1}{D^\perp}(\vec v\vec w D^\perp - {}^3\nabla_{\vec v}\vec w D^\perp)K(\vec u,\vec h_i) \\
& - \frac{1}{D^\perp}(\vec u\vec w D^\perp - {}^3\nabla_{\vec u}\vec w D^\perp)K(\vec v,\vec h_i) + \frac{1}{D^\perp}(\vec u\vec h_i D^\perp - {}^3\nabla_{\vec u}\vec h_i D^\perp)K(\vec v,\vec w) \\
& - \frac{1}{D^\perp}(\vec v\vec h_i D^\perp - {}^3\nabla_{\vec v}\vec h_i D^\perp)K(\vec u,\vec w) - K({}^3R(\vec u,\vec v)\vec w,\vec h_i)\}\vec h_i \\
& + \frac{1}{D^\perp}{}^3R(\vec u,\vec v)\vec w D^\perp\vec n \quad, \qquad (2.37)
\end{aligned}$$

$$\begin{aligned}
\nabla_{\vec n}{}^3S(\vec u,\vec v) = {} & \varepsilon\sum_i\{K(\vec u,\vec h_i)\,{}^3S(\vec h_i,\vec v) + K(\vec v,\vec h_i)\,{}^3S(\vec u,\vec h_i) + K_2(\vec h_i,\vec h_i;\vec u,\vec v) \\
& - K_2(\vec h_i,\vec u;\vec h_i,\vec v) - K_2(\vec h_i,\vec v;\vec u,\vec h_i) + K_2(\vec u,\vec v;\vec h_i,\vec h_i)\} \\
& + \varepsilon\frac{1}{D^\perp}\sum_i\{\vec u D^\perp[K_1(\vec v;\vec h_i,\vec h_i) - K_1(\vec h_i;\vec h_i,\vec v)] + \vec v D^\perp[K_1(\vec u;\vec h_i,\vec h_i) \\
& - K_1(\vec h_i;\vec u,\vec h_i)] + \vec h_i D^\perp[2K_1(\vec h_i;\vec u,\vec v) - K_1(\vec u;\vec h_i,\vec v) - K_1(\vec v;\vec u,\vec h_i)] \\
& + (\vec u\vec v D^\perp - {}^3\nabla_{\vec u}\vec v D^\perp)K(\vec h_i,\vec h_i) - (\vec u\vec h_i D^\perp - {}^3\nabla_{\vec u}\vec h_i D^\perp)K(\vec h_i,\vec v) \\
& - (\vec v\vec h_i D^\perp - {}^3\nabla_{\vec v}\vec h_i D^\perp)K(\vec u,\vec h_i) + (\vec h_i\vec h_i D^\perp - {}^3\nabla_{\vec h_i}\vec h_i D^\perp)K(\vec u,\vec v)\} \;. \qquad (2.38)
\end{aligned}$$

We saw above (cf. Equations (2.20),(2.21),(2.29), and (2.30)) that R may be expressed completely in terms of ${}^3g$, K , and ΠS , and hypersurface derivatives (i.e. derivatives along hypersurface vector fields $\vec{u},\vec{v},\ldots$) of these quantities. The components $\nabla_{\vec{u}}R(\vec{V},\vec{W})\vec{X}$ and (with the use of Bianchi's second set of identities) $\nabla_{\vec{n}}R(\vec{v},\vec{w})\vec{X}$ of ∇R may also be expressed in terms of this same data. However, whenever there are two or more derivatives in the normal direction, as in $\nabla_{\vec{n}}R(\vec{n},\vec{v})\vec{X}$, more data is required. Direct calculations, in which Equations (2.36), (2.37), and (2.38) are used, give the following results:

$$\begin{aligned}
\nabla_{\vec{n}}R(\vec{n},\vec{v})\vec{x} = {} & \varepsilon\sum_i\{(\Pi S)_1(\vec{h}_i;\vec{x},\vec{v}) - {}^3S_1(\vec{h}_i;\vec{x},\vec{v}) - (\Pi S)_1(\vec{x};\vec{h}_i,\vec{v}) \\
& + {}^3S_1(\vec{x};\vec{h}_i,\vec{v})\}\vec{h}_i + \sum_{i,j}\{K(\vec{h}_j,\vec{h}_j)[K_1(\vec{h}_i;\vec{x},\vec{v}) - K_1(\vec{x};\vec{h}_i,\vec{v})] \\
& + K(\vec{x},\vec{v})K_1(\vec{h}_i;\vec{h}_j,\vec{h}_j) - K(\vec{h}_i,\vec{v})K_1(\vec{x};\vec{h}_j,\vec{h}_j) + K(\vec{x},\vec{h}_j)[K_1(\vec{v};\vec{h}_i,\vec{h}_j) \\
& - 2K_1(\vec{h}_j;\vec{h}_i,\vec{v})] + K(\vec{v},\vec{h}_j)[K_1(\vec{x};\vec{h}_i,\vec{h}_j) - K_1(\vec{h}_i;\vec{x},\vec{h}_j)] \\
& + K(\vec{h}_i,\vec{h}_j)[2K_1(\vec{h}_j;\vec{x},\vec{v}) - K_1(\vec{v};\vec{x},\vec{h}_j)]\}\vec{h}_i \\
& + \varepsilon\sum_i\{K(\vec{x},\vec{v})[S(\vec{h}_i,\vec{h}_i) - {}^3S(\vec{h}_i,\vec{h}_i) + \varepsilon K(\vec{h}_i,\vec{h}_i)\sum_j K(\vec{h}_j,\vec{h}_j)] \\
& \quad + K(\vec{h}_i,\vec{h}_i)[S(\vec{x},\vec{v}) - {}^3S(\vec{x},\vec{v}) + \varepsilon K(\vec{x},\vec{v})\sum_j K(\vec{h}_j,\vec{h}_j)] \\
& \quad - K(\vec{x},\vec{h}_i)[S(\vec{v},\vec{h}_i) - {}^3S(\vec{v},\vec{h}_i) + \varepsilon K(\vec{v},\vec{h}_i)\sum_j K(\vec{h}_j,\vec{h}_j)] \\
& \quad - K(\vec{v},\vec{h}_i)[S(\vec{x},\vec{h}_i) - {}^3S(\vec{x},\vec{h}_i) + \varepsilon K(\vec{x},\vec{h}_i)\sum_j K(\vec{h}_j,\vec{h}_j)] \\
& \quad - K(\vec{h}_i,{}^3R(\vec{h}_i,\vec{x})\vec{v}) - K(\vec{h}_i,{}^3R(\vec{h}_i,\vec{v})\vec{x}) + K_2(\vec{x},\vec{h}_i;\vec{h}_i,\vec{v})
\end{aligned}$$

$$+ K_2(\vec{v},\vec{h}_i;\vec{x},\vec{h}_i) - K_2(\vec{x},\vec{v};\vec{h}_i,\vec{h}_i) - K_2(\vec{h}_i,\vec{h}_i;\vec{x},\vec{v})\}\vec{n}$$

$$+ \{\nabla_{\vec{n}}(\Pi S)(\vec{x},\vec{v}) - \frac{1}{D^\perp}\vec{x}D^\perp S(\vec{n},\vec{v}) - \frac{1}{D^\perp}\vec{v}D^\perp S(\vec{x},\vec{n})\}\vec{n} \quad , \tag{2.39}$$

$$\nabla_{\vec{n}}R(\vec{n},\vec{v})\vec{n} = \sum_{i,j}\{K(\vec{h}_j,\vec{v})[S(\vec{h}_i,\vec{h}_j) - {}^3S(\vec{h}_i,\vec{h}_j)$$

$$+ 2\varepsilon K(\vec{h}_i,\vec{h}_j)\sum_k K(\vec{h}_k,\vec{h}_k)] - K(\vec{h}_i,\vec{v})[S(\vec{h}_j,\vec{h}_j) - {}^3S(\vec{h}_j,\vec{h}_j)$$

$$+ 2\varepsilon K(\vec{h}_j,\vec{h}_j)\sum_k K(\vec{h}_k,\vec{h}_k)] + K(\vec{h}_i,\vec{h}_j)[S(\vec{h}_j,\vec{v}) - {}^3S(\vec{h}_j,\vec{v})]$$

$$- K(\vec{h}_j,\vec{h}_j)[S(\vec{h}_i,\vec{v}) - {}^3S(\vec{h}_i,\vec{v})] + K(\vec{h}_j,{}^3R(\vec{h}_j,\vec{h}_i)\vec{v})$$

$$+ K(\vec{h}_j,{}^3R(\vec{h}_j,\vec{v})\vec{h}_i) + K_2(\vec{h}_j,\vec{h}_j;\vec{h}_i,\vec{v}) - K_2(\vec{h}_i,\vec{h}_j;\vec{h}_j,\vec{v})$$

$$- K_2(\vec{v},\vec{h}_j;\vec{h}_i,\vec{h}_j) + K_2(\vec{h}_i,\vec{v};\vec{h}_j,\vec{h}_j)\}\vec{h}_i$$

$$- \varepsilon\sum_i\{\nabla_{\vec{n}}(\Pi S)(\vec{h}_i,\vec{v}) - \frac{1}{D^\perp}\vec{h}_i D^\perp S(\vec{n},\vec{v}) - \frac{1}{D^\perp}\vec{v}D^\perp S(\vec{h}_i,\vec{n})\}\vec{h}_i \quad . \tag{2.40}$$

with $S(\vec{n},\vec{v})$ defined by (2.26). In both of these equations, the new data involved is the (deformation dependent) field $\nabla_{\vec{n}}(\Pi S)$. By noting that

$$\nabla_{\vec{n}}S(\vec{u},\vec{v}) = \vec{n}(S(\vec{u},\vec{v})) - S(\nabla_{\vec{n}}\vec{u},\vec{v}) - S(\vec{u},\nabla_{\vec{n}}\vec{v})$$

$$= \nabla_{\vec{n}}(\Pi S)(\vec{u},\vec{v}) + \Pi S(\nabla_{\vec{n}}\vec{u},\vec{v}) - S(\nabla_{\vec{n}}\vec{u},\vec{v}) + \Pi S(\vec{u},\nabla_{\vec{n}}\vec{v}) - S(\vec{u},\nabla_{\vec{n}}\vec{v})$$

$$= \nabla_{\vec{n}}(\Pi S)(\vec{u},\vec{v}) - \frac{1}{D^\perp}\vec{u}D^\perp S(\vec{n},\vec{v}) - \frac{1}{D^\perp}\vec{v}D^\perp S(\vec{u},\vec{n}) \quad , \tag{2.41}$$

all of the deformation dependent terms may be combined into the one physical (deformation independent) field $\Pi(\nabla_{\vec{n}}S)$, yielding results completely analogous to (2.29) and (2.30).

The final outcome of these rather lengthy calculations is that if we wish to characterize the geometry of space-time from within a given space-like hypersurface, say $e(\mathcal{S})$, then the only a priori independent fields that we must specify on $e(\mathcal{S})$ are the hypersurface metric, ${}^3g$, the extrinsic curvature, K, the projection, ΠS, of the space-time Ricci tensor onto $e(\mathcal{S})$, and the hypersurface components of the covariant derivatives along $\vec{n}$, to all orders, of the space-time Ricci tensor: $\Pi(\nabla_{\vec{n}}S)$, $\Pi(\nabla_{\vec{n}}\nabla_{\vec{n}}S - \nabla_{\nabla_{\vec{n}}\vec{n}}S)$, etc.. Equation (2.10) shows that K is just the derivative of ${}^3g$ along $\vec{n}$, and (2.25) indicates that ΠS is (roughly speaking) the derivative of K along $\vec{n}$; so the independent data is effectively ${}^3g$ and all of its (normalized) time derivatives.

Although all of these fields are well defined and a priori independent of each other, physical space-time is such that only a finite number of them need be specified in order to determine the complete set (cf. (1.1.9)). The <u>initial data</u> on $e(\mathcal{S})$ then consists of ${}^3g$ and those derivatives of ${}^3g$ along $\vec{n}$ up to some finite order, say m, that cannot be obtained as functionals of the others. By implication, the $(m+1)$th and higher derivatives of ${}^3g$ can be obtained as explicit functionals of the initial data fields and their hypersurface derivatives.

3. Gravitational Field Equations

The physical assumptions that I have made so far are insufficient to determine at what differential order the initial data cuts off and dynamical equations begin. I shall assume, therefore, that, as above, the initial data includes derivatives of ^3g along $\vec{n}$ up to and including the m-th order. If this data is known on $e(S)$, then on an infinitesimally close hypersurface, $D_{\delta t}(S)$, the metric is given by (cf. (2.1))

$$^3g_{\delta t}(\vec{u},\vec{v}) = {}^3g_0(\vec{u},\vec{v}) + \frac{d}{dt}({}^3g_t(\vec{u},\vec{v}))\Big|_{t=0}\cdot\delta t \quad , \qquad (3.1)$$

and the time derivatives up to (m-1)th order are given by similar expressions. By iterating this process, the hypersurface metric can be carried forward m infinitesimal steps in time, but $m\cdot\delta t$ is still infinitesimal. In order to be able to integrate ahead a finite distance in time, we must carry all m derivatives forward onto each successive hypersurface, thus making it equivalent to its predecessor. This can be done only if the (m+1)th time derivative of ^3g is assumed to be an explicit functional of ^3g and the lower derivatives, on each of the hypersurfaces, with the functional form being the same on all hypersurfaces. Once the (m+1)th derivative has been determined on $e(S)$, with the use of these dynamical equations, the m-th derivative can be constructed on $D_{\delta t}(S)$, and the process can be repeated ad infinitum.

At first glance, it might seem as though any functional of the (m+1) hypersurface fields that comprise the initial data, and their hypersurface derivatives to all orders, should yield a consistent set of dynamical field equations. However, things aren't quite that simple. Let us suppose, for the moment, that $m = 2$. The initial data on $e(S)$ is then ${}^3g$, K , and ΠS , and the dynamical field equations give $\Pi(\nabla_{\vec{n}}S)$ as a functional of the initial data:

$$\Pi(\nabla_{\vec{n}}S) = \Pi(\nabla_{\vec{n}}S)[{}^3g, K, \Pi S, {}^3R, {}^3\nabla K, {}^3\nabla(\Pi S), {}^3\nabla\,{}^3R, \ldots] \quad . \tag{3.2}$$

If this functional is known, then, with the use of (2.41) and (2.26), we can determine $\nabla_{\vec{n}}(\Pi S)$ as a functional of the initial data and the deformation vector field, $\vec{D}$; and by repeatedly applying the techniques demonstrated in (2.31) we can compute the functional form of the covariant derivative along $\vec{n}$ of each of the fields upon which $\Pi(\nabla_{\vec{n}}S)$ depends. The dynamical field equations, (3.2), thus determine their own derivative:

$$\nabla_{\vec{n}}(\Pi(\nabla_{\vec{n}}S)) = \nabla_{\vec{n}}(\Pi(\nabla_{\vec{n}}S))[{}^3g, K, \Pi S, \ldots; \vec{D}] \quad . \tag{3.3}$$

But we also have the general result:

$$\begin{aligned}\nabla_{\vec{n}}(\Pi(\nabla_{\vec{n}}S))(\vec{u},\vec{v}) &= \vec{n}(\Pi(\nabla_{\vec{n}}S)(\vec{u},\vec{v})) - \Pi(\nabla_{\vec{n}}S)(\nabla_{\vec{n}}\vec{u},\vec{v}) \\ &\quad - \Pi(\nabla_{\vec{n}}S)(\vec{u},\nabla_{\vec{n}}\vec{v}) \\ &= \vec{n}(\nabla_{\vec{n}}S(\vec{u},\vec{v})) - \nabla_{\vec{n}}S(\nabla_{\vec{n}}\vec{u},\vec{v}) - \nabla_{\vec{n}}S(\vec{u},\nabla_{\vec{n}}\vec{v}) + \frac{1}{D^{\perp}}\vec{u}D^{\perp}\,\nabla_{\vec{n}}S(\vec{n},\vec{v})\end{aligned}$$

$$+ \frac{1}{D^{\perp}}\vec{v}D^{\perp}\nabla_{\vec{n}}S(\vec{u},\vec{n})$$

$$= \{\nabla_{\vec{n}}\nabla_{\vec{n}}S(\vec{u},\vec{v}) - \nabla_{\nabla_{\vec{n}}\vec{n}}S(\vec{u},\vec{v})\} - \varepsilon\frac{1}{D^{\perp}}\sum_{i}\vec{h}_{i}D^{\perp}\nabla_{\vec{h}_{i}}S(\vec{u},\vec{v})$$
$$+ \frac{1}{D^{\perp}}\vec{u}D^{\perp}\nabla_{\vec{n}}S(\vec{n},\vec{v}) + \frac{1}{D^{\perp}}\vec{v}D^{\perp}\nabla_{\vec{n}}S(\vec{u},\vec{n}) \quad . \tag{3.4}$$

Taken together, the two terms in parentheses on the right hand side of (3.4) constitute the fourth (normalized) time derivative of ^{3}g , a quantity which <u>must not depend in any way on the deformation</u>, $\vec{D}$. However, when (3.4) is subtracted from (3.3) the resulting equation may be solved to give an expression for $\{\nabla_{\vec{n}}\nabla_{\vec{n}}S - \nabla_{\nabla_{\vec{n}}\vec{n}}S\}$ which does depend explicitly on $\vec{D}$. This apparent contradiction is resolved by <u>constraining the initial data</u> to satisfy functional relations which make the deformation dependence of $\{\nabla_{\vec{n}}\nabla_{\vec{n}}S - \nabla_{\nabla_{\vec{n}}\vec{n}}S\}$ vanish identically. If the functional form of $\Pi(\nabla_{\vec{n}}S)$ (Equation (3.2)) has been chosen appropriately, then the associated constraint equations will be sufficiently weak that ΠS may still be considered as part of the initial data (i.e. ΠS may not be obtained as a functional of ^{3}g , K , and their hypersurface derivatives).

Once the <u>primary constraint equations</u> (introduced in the previous paragraph) have been found, they may be used to aid in the construction of the next covariant derivative of S along $\vec{n}$: $\{\nabla_{\vec{n}}\nabla_{\vec{n}}\nabla_{\vec{n}}S - 3\nabla_{\vec{n}}\nabla_{\nabla_{\vec{n}}\vec{n}}S + 2\nabla_{\nabla_{\vec{n}}\nabla_{\vec{n}}\vec{n}}S\}$. As with $\{\nabla_{\vec{n}}\nabla_{\vec{n}}S - \nabla_{\nabla_{\vec{n}}\vec{n}}S\}$, this will also depend explicitly on $\vec{D}$, and new <u>secondary constraints</u> on the initial data must be chosen to make the deformation dependent terms vanish. The calculations are exactly analogous to those for the primary constraints, but longer due to the extra derivative.

Finally, the fourth derivative of S along $\vec{n}$ may be computed with the use of the dynamical equations and the primary and secondary constraints, and once again new constraints must be imposed on the initial data to eliminate the explicit deformation dependence. Still higher derivatives will automatically be independent of $\vec{D}$.

Looking back at (3.4), we see that the primary constraints place restrictions on $\nabla_{\vec{w}}S(\vec{u},\vec{v})$ and $\nabla_{\vec{n}}S(\vec{u},\vec{n})$. Equations analogous to (3.4) for the higher derivatives would show that the secondary constraints restrict $\nabla_{\vec{w}}S(\vec{u},\vec{n})$ and $\nabla_{\vec{n}}S(\vec{n},\vec{n})$, and that the tertiary constraints (which are often also called secondary) restrict $\nabla_{\vec{w}}S(\vec{n},\vec{n})$. Thus the complete set of dynamic plus constraint equations determines the form of all the components of ∇S , as functionals of the initial data. Moreover, since ∇S is a space-time tensor field that is completely independent of the choice of hypersurface, $e(\mathcal{S})$, or deformation, $\vec{D}$, the functionals that make up its components must fit together to form a space-time tensor field that is also independent of e or $\vec{D}$, but which is nonetheless constructed from the initial data on $e(\mathcal{S})$.

When we look at the initial data, though, we see that it is itself derived from g and its space-time derivatives, so any space-time tensor field that is constructed from the initial data (and is independent of e and $\vec{D}$) must ultimately be a functional of g, R, ∇R, etc.. Of these latter fields, only g and R can be constructed directly from the initial data on $e(S)$, in the particular case that we are considering $(m = 2)$. The complete set of geometrical field equations (dynamical equations and constraints) must therefore take the form:

$$\nabla S = \nabla S[g,R] \quad . \tag{3.5}$$

Aside from the requirement that it yield a tensor field ∇S of the correct form (i.e. third rank, covariant, symmetric in the last two indices, and satisfying the contracted Bianchi identities), no further restrictions are placed on this functional by the physical assumptions made so far.

The calculations for other values of m $(m \geqslant 1)$ are very much the same as for $m = 2$. If the initial data is assumed to include ${}^3g$ and its invariant derivatives along $\vec{n}$ up to and including m-th order, then the dynamical equations give the $(m+1)$th derivative as a functional of the initial data. Higher derivatives of ${}^3g$ are then obtained by differentiating the field equations, and the deformation dependence of the invariant terms (space-time tensors) is eliminated by imposing constraints on the initial data. Although the sequence may terminate earlier, there are, in general, $m + 1$ orders of constraint equations. When the dynamical equations and the constraints are all satisfied, derivatives of g to all orders may be computed, and the system of equations is integrable.

The purpose of the constraints is to guarantee that the predicted geometry of any future hypersurface, $D_t(S)$, depends only on the initial data defined on $e(S)$, and not on the sequence of intermediate hypersurfaces used in the time integration. Their net effect, however, is to supplement the dynamical equations, building them up into a set of covariant equations in the space-time fields, g , R , ∇R , etc., in which the highest derivative of g is of $(m+1)$th order and enters linearly (cf. (3.5)).

4. The Einstein Vacuum Equations

Throughout modern physics it is assumed that dynamical systems are characterized completely by their instantaneous "coordinates" and "velocities", with their "accelerations" being determined by dynamical equations. For the geometrical field theory being discussed in this chapter, the coordinates are the components of ^3g on $e(S)$, and the velocities are the components of K ; so in this section I shall investigate the class of theories for which $m = 1$.

As outlined above, the dynamical equations must take the form

$$\Pi S = \Pi S[{}^3g, K, {}^3R, {}^3\nabla K, \ldots] \quad . \tag{4.1}$$

Once this functional has been chosen, it may be used with (2.25) to find $\nabla_{\vec{n}}K$ in terms of the initial data and the deformation vector field $\vec{D}$, and with (2.36),(2.37),(2.38), and other similar equations to compute $\nabla_{\vec{n}}$ of each of the other fields upon which ΠS depends. Knowing all these derivatives, we can use the chain rule to compute $\nabla_{\vec{n}}(\Pi S)$.

On the other hand, though, equation (2.41) gives

$$\nabla_{\vec{n}}(\Pi S)(\vec{u},\vec{v}) = \nabla_{\vec{n}}S(\vec{u},\vec{v}) + \frac{1}{D^{\perp}}\vec{u}D^{\perp}S(\vec{n},\vec{v}) + \frac{1}{D^{\perp}}\vec{v}D^{\perp}S(\vec{u},\vec{n}) \quad . \tag{4.2}$$

Subtracting from this the expression obtained from (4.1) for $\nabla_{\vec{n}}(\Pi S)(\vec{u},\vec{v})$ yields an equation that may be solved for $\nabla_{\vec{n}}S(\vec{u},\vec{v})$:

$$\nabla_{\vec{n}}S(\vec{u},\vec{v}) = \nabla_{\vec{n}}S[{}^3g,K,{}^3R,{}^3\nabla K,\ldots;\vec{D}](\vec{u},\vec{v}) \quad . \tag{4.3}$$

Because $\nabla_{\vec{n}}S$ may be constructed directly from g , it is clear that the right hand side of (4.3) must actually be independent of $\vec{D}$; but an examination of the terms in (2.25),(2.36),(2.37), and (2.38) that depend explicitly on $\vec{D}$ shows that no matter how the functional (4.1) is chosen, its derivative, $\nabla_{\vec{n}}(\Pi S)$, will not have (explicitly) deformation dependent terms of the form $\frac{1}{D^{\perp}}\vec{u}D^{\perp}T(\vec{v})$ and $\frac{1}{D^{\perp}}\vec{v}D^{\perp}T(\vec{u})$ (with T independent of $\vec{D}$) capable of cancelling the last two terms in (4.2). The only way in which the deformation dependence in (4.3) can be eliminated is thus to constrain the initial data to satisfy

$$S(\vec{n},\vec{v}) \equiv \varepsilon\sum_i\{{}^3\nabla_{\vec{v}}K(\vec{h}_i,\vec{h}_i) - {}^3\nabla_{\vec{h}_i}K(\vec{v},\vec{h}_i)\} = 0 \quad , \tag{4.4}$$

and to choose the functional (4.1) so that its derivative, $\nabla_{\vec{n}}(\Pi S)$, is completely independent of $\vec{D}$.

The constraints (4.4) serve to limit the configurations of the initial data fields on $e(S)$. However, there is nothing special about this particular space-like hypersurface, so equations (4.4) should also be satisfied on each subsequent hypersurface, $D_t(S)$. To this end I require that

$$\vec{n}(S(\vec{n},\vec{v})) = 0 \quad , \tag{4.5}$$

which leads, through a straightforward calculation, to

$$\nabla_{\vec{n}} S(\vec{n},\vec{v}) + \frac{1}{D^{\perp}}\sum_i \vec{h}_i D^{\perp}\{g(\vec{v},\vec{h}_i)S(\vec{n},\vec{n}) - \varepsilon S(\vec{v},\vec{h}_i)\} = 0 \quad . \tag{4.6}$$

Since this must be satisfied for all choices of $\vec{D}$, I find that

$$\nabla_{\vec{n}} S(\vec{n},\vec{v}) = 0 \tag{4.7}$$

and

$$\Pi S(\vec{u},\vec{v}) = \varepsilon\, {}^3g(\vec{u},\vec{v})S(\vec{n},\vec{n}) \quad . \tag{4.8}$$

With the use of (2.25) and (2.27), equation (4.8) can be solved to give

$$\Pi S(\vec{u},\vec{v}) = \tfrac{1}{2}\, {}^3g(\vec{u},\vec{v})\{\sum_i {}^3S(\vec{h}_i,\vec{h}_i) - \varepsilon \sum_{i,j}[K(\vec{h}_i,\vec{h}_i)K(\vec{h}_j,\vec{h}_j) - K(\vec{h}_i,\vec{h}_j)K(\vec{h}_i,\vec{h}_j)]\} \quad . \tag{4.9}$$

Using (2.25) and (2.38), it can then be shown that

$$\nabla_{\vec{n}}(\Pi S)(\vec{u},\vec{v}) = 0 \quad (= \nabla_{\vec{n}} S(\vec{u},\vec{v})) \quad . \tag{4.10}$$

Derivatives of S in directions parallel to the hypersurface are given by

$$
\begin{aligned}
&{}^3\nabla_{\vec{w}}(\Pi S)(\vec{u},\vec{v}) \quad (= \nabla_{\vec{w}} S(\vec{u},\vec{v})) \\
&= \tfrac{1}{2}\ {}^3g(\vec{u},\vec{v})\{\vec{w}\,{}^3S - 2\varepsilon \sum_{i,j} [{}^3\nabla_{\vec{w}}K(\vec{h}_i,\vec{h}_i)K(\vec{h}_j,\vec{h}_j) \\
&- {}^3\nabla_{\vec{w}}K(\vec{h}_i,\vec{h}_j)K(\vec{h}_i,\vec{h}_j)]\} \quad ; \qquad (4.11)
\end{aligned}
$$

however, the contracted Bianchi identities for S tell us that these must vanish:

$$
\begin{aligned}
0 &= \sum_i \nabla_{\vec{h}_i} S(\vec{h}_i,\vec{v}) + \varepsilon\nabla_{\vec{n}} S(\vec{n},\vec{v}) - \tfrac{1}{2}\vec{v}S \\
&= \sum_i \nabla_{\vec{h}_i} S(\vec{h}_i,\vec{v}) - \tfrac{1}{2}\vec{v}\{(4/3)\sum_i S(\vec{h}_i,\vec{h}_i)\} \\
&= \sum_i \{{}^3\nabla_{\vec{h}_i}(\Pi S)(\vec{h}_i,\vec{v}) - (2/3)\,{}^3\nabla_{\vec{v}}(\Pi S)(\vec{h}_i,\vec{h}_i)\} \\
&= \sum_i \{-\tfrac{1}{2}\vec{v}\,{}^3S + \varepsilon\sum_j [{}^3\nabla_{\vec{v}}K(\vec{h}_i,\vec{h}_i)K(\vec{h}_j,\vec{h}_j) - {}^3\nabla_{\vec{v}}K(\vec{h}_i,\vec{h}_j)K(\vec{h}_i,\vec{h}_j)]\} \\
&= -\sum_i {}^3\nabla_{\vec{v}}(\Pi S)(\vec{h}_i,\vec{h}_i) \quad . \qquad (4.12)
\end{aligned}
$$

Now it is well known that any symmetric, second rank tensor with vanishing covariant derivative must be proportional to the metric tensor, so we finally obtain the dynamical equations

$$(\Pi S)(\vec{u},\vec{v}) = -\lambda\,{}^3g(\vec{u},\vec{v}) \quad , \qquad (4.13)$$

with λ a constant on $e(S)$. Equation (4.10) indicates that λ must also be a constant in time; and (4.8) now reduces to the secondary constraint:

$$S(\vec{n},\vec{n}) = -\varepsilon\lambda \quad . \tag{4.14}$$

When collected together, equations (4.4),(4.13), and (4.14) are quickly recognized as Einstein's vacuum equations for the gravitational field (with the cosmological term):

$$S(\vec{U},\vec{V}) = -\lambda g(\vec{U},\vec{V}) \quad . \tag{4.15}$$

For the restricted set of initial data fields, ${}^3g$ and K , they are the only field equations capable of unambiguously propagating the metric of space forward in time.

A much shorter, but less instructive, derivation of these same equations follows from the conclusions of Section 3 of this chapter. They indicate that the Ricci tensor S must be an explicit functional of tensors formed from g and its first derivatives (since $m = 1$). But it is impossible to form any tensor field from the first derivatives of g , so we must have

$$S = S[g] \quad , \tag{4.16}$$

which leads immediately to (4.15).

It is also interesting to note that even if we were willing to add ΠS to the initial data (cf. Section 3) we would be frustrated in all attempts to do so. Because g and R are both of even rank, any non-trivial functional of these fields must also be an even rank tensor. But ∇S is a third rank tensor, so the only solution to (3.5) is

$$\nabla S = 0 \quad , \tag{4.17}$$

which leads us once again to the field equations (4.15).

5. Metric Signature, and Causality

Throughout the foregoing discussions the sign ε of $g(\vec{n},\vec{n})$ has been left undetermined, but definite. Whether ε is $+1$ or -1 makes little difference to the form of the field equations. However, if g is to give rise to the (partial) time ordering of physical events that is demanded by (1.1.2), then we must make the standard assignment:

$$\varepsilon = -1 \quad . \tag{5.1}$$

More pragmatic reasons for making this choice are provided by the empirical successes of special relativity and Maxwell's theory of electromagnetism.

Our original motivation for introducing the field g was to induce a positive definite metric on each space-like hypersurface of space-time (the term "space-like" being defined in (1.1.3)), and if we had decided that $\varepsilon = +1$, then every hypersurface of space-time would have had such a metric. Starting with the initial space-like hypersurface, $e(S)$, <u>any</u> deformation vector field, $\vec{D}$, would have then led to a sequence of hypersurfaces $D_t(S)$ with induced Riemannian metrics. But with the locally Minkowskian metric of physical space-time, the field

$\vec{D}$ is severely restricted by the requirement that both $\vec{D}^{\parallel}$ and $\vec{D}^{\perp}$ be smooth. This is illustrated by the following example.

Let M be a space-time endowed with a locally Minkowskian metric g that satisfies some known set of (predictive) field equations; let M' be an open cell in M ; and let σ' be a space-like hypersurface of M' which extends to a space-like hypersurface σ of M . Then, just as in Section 1, it is always possible to generate a parameterized family of space-like hypersurfaces σ_t by deforming σ along a smooth vector field $\vec{D}$ on M . For the purposes of this example I shall choose $\vec{D}$ such that on σ' it satisfies $D^{\perp} > 0$, and on $\sigma\backslash\overline{\sigma}'$ its perpendicular part vanishes. I shall denote the portion of σ_t that does not coincide with σ by σ'_t ($\sigma'_t = (\sigma_t \cup \sigma)\backslash\sigma$). This is illustrated in Figure 5.1. It follows immediately from the above assumptions that for each point x of σ'_t , and for all t , every past (future) directed time-like path through x intersects σ' and each of the intermediate surfaces σ'_s , $0 < s < t$ (cf. (1.1.9)).

In conjunction with the field equations, the initial data induced on σ can be used to predict what the geometry of each of the subsequent hypersurfaces σ_t is. The fields predicted to exist on σ'_t , however, depend only on the initial data defined on σ' and are independent of the field configurations on $\sigma\backslash\sigma'$. Without this result, one could not make confident predictions about the future (or past) without first gathering

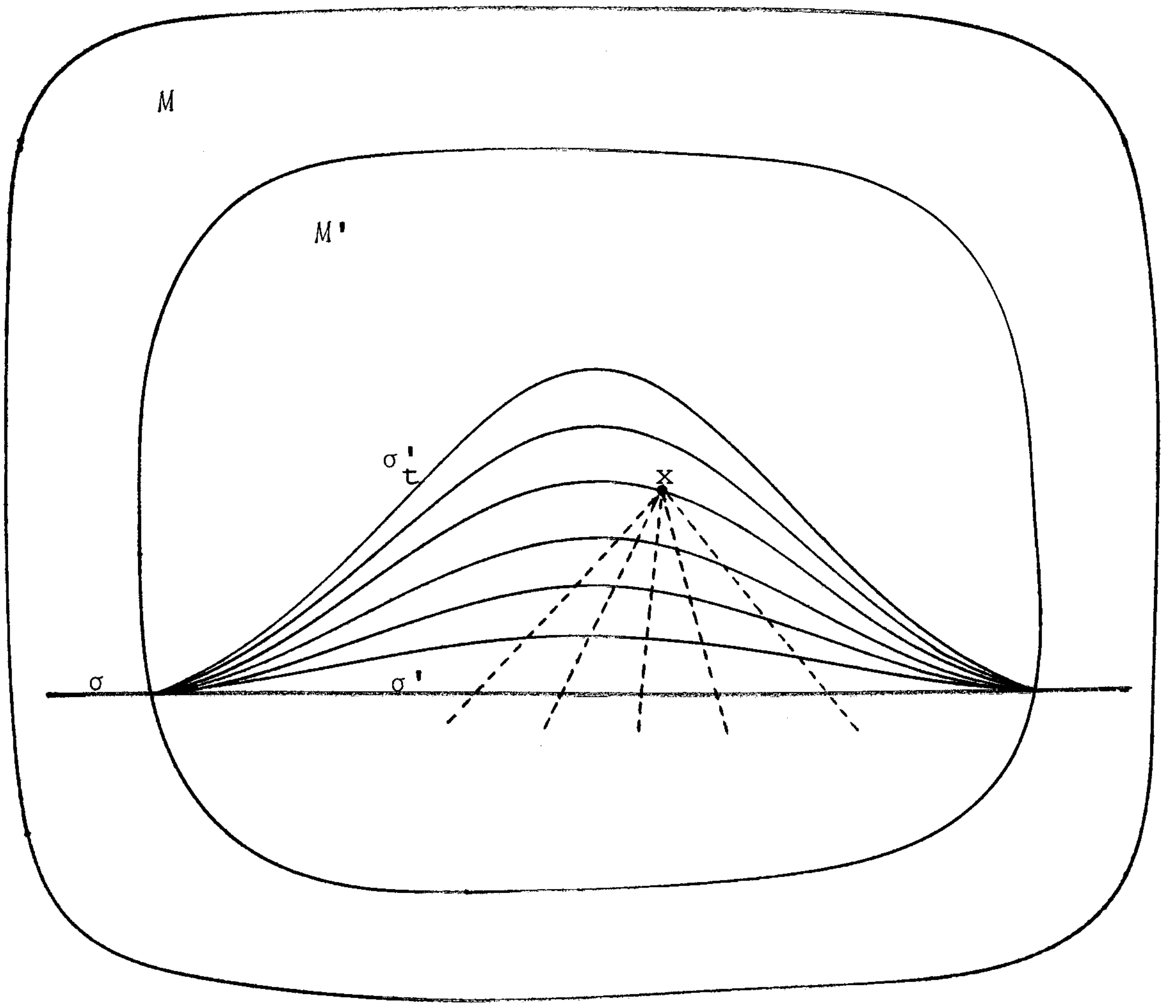


Figure 5.1 Every past directed time-like path through x (broken lines) intersects σ' and each of the intermediate space-like hypersurfaces σ_s' .

information about the entire present universe, rather than just some local neighbourhood.

Conversely, if the fields on σ'_t are to be independent of the initial data on $\sigma\backslash\sigma'$, then the deformation vector field $\vec{D}$ must (1) satisfy $D^{\perp} = 0$ on $\sigma\backslash\overline{\sigma}'$, and (2) leave each of the hypersurfaces σ_t space-like. This latter condition is assured by requiring that both $\vec{D}^{\parallel}$ and $D^{\perp}$ be smooth fields.

CHAPTER 3

MATTER FIELDS AND THE GEOMETRY OF SPACE-TIME

While much of the universe appears to be vacuous (or very nearly so), we are unable to make direct observations of the vacuum. Instead, we observe the matter that is contained in the universe, and deduce from the distribution of matter the geometry of both the vacuous and non-vacuous regions. It is thus essential to include matter in any complete discussion of the geometrical structure of space and space-time.

1. Initial Value Problem

The distribution of matter in space and its evolution in time is characterized by a set of smooth tensor fields F_j, $j \in \omega$, on space-time (cf. (1.1.8)). Since the space-time metric g may always be used to lower indices, I shall assume (without loss of generality) that the fields F_j are all covariant tensors.

Let F be a typical representative of the matter fields defined on M'. (The notation here is the same as in Chapter 2 - i.e. M' is an open cell in the space-time manifold, M.) Then, on the initial surface, $e(S)$, the instantaneous configuration of F is described by the hypersurface tensor fields

$$\Pi F = F(\Pi(\_),\ldots,\Pi(\_))$$
$$F(\vec{n},\Pi(\_),\ldots,\Pi(\_)) \ , \quad F(\Pi(\_),\vec{n},\Pi(\_),\ldots,\Pi(\_)) \ , \ \ldots$$
$$F(\vec{n},\vec{n},\Pi(\_),\ldots,\Pi(\_)) \ , \quad F(\vec{n},\Pi(\_),\vec{n},\Pi(\_),\ldots,\Pi(\_)) \ , \ \ldots$$
etc.

The instantaneous rate of change of F is described by analogous hypersurface fields constructed from $\nabla_{\vec{n}}F$; the accelerations are constructed from $\{\nabla_{\vec{n}}\nabla_{\vec{n}}F - \nabla_{\nabla_{\vec{n}}\vec{n}}F\}$; and so on.

Although all of the fields that are induced on $e(S)$ in this way are *a priori* independent of each other, only a small subset of them need be specified (along with the geometrical initial data) in order to determine the complete set. For a given matter field, F_i , only a finite number of the induced hypersurface fields may be considered as initial data; and, as with the metric, all of the time derivatives of F_i beyond some given order, say m_i , can be obtained as explicit functionals of the lower derivatives of F_i , the initial data for the other matter fields, F_j , and the geometrical initial data (discussed in Chapter 2).

Now suppose, once again, that the geometrical initial data is given by ${}^3g$ and K , corresponding to $m = 1$. In the vacuum case the dynamical equations for g took the form

$$\Pi S = \Pi S[{}^3g,K,{}^3R,{}^3\nabla K,\ldots] \quad ; \qquad (2.4.1)$$

but when matter fields are present there is much more initial data upon which ΠS can depend. To avoid confusion between the field, ΠS , and the functional, I shall write

$$\Pi S = {}^3E[\ \text{I.D.}\ ] \tag{1.1}$$

with I.D. representing the complete set of initial data fields and their hypersurface derivatives.

Once the functional ${}^3E$ has been chosen, equation (1.1) may be substituted into (2.2.25) to find $\nabla_{\vec{n}}K$ in terms of the initial data and the deformation vector field, $\vec{D}$; and this, together with the (as yet undetermined) dynamical equations for the matter fields, allows us to compute $\nabla_{\vec{n}}{}^3E$ as a functional of the initial data and $\vec{D}$. We know from (2.4.2), however, that this new functional must take the form

$$\nabla_{\vec{n}}{}^3E(\vec{u},\vec{v}) = {}^3E'(\vec{u},\vec{v}) + \frac{1}{D^\perp}\vec{u}D^\perp\,{}^3P(\vec{v}) + \frac{1}{D^\perp}\vec{v}D^\perp\,{}^3P(\vec{u}) \quad , \tag{1.2}$$

where ${}^3E'$ and ${}^3P$ are again explicit functionals of the initial data. Further comparison with (2.4.2) gives

$$\nabla_{\vec{n}}S(\vec{u},\vec{v}) = {}^3E'(\vec{u},\vec{v})[\ \text{I.D.}\ ] \quad , \tag{1.3}$$

and the primary constraint equations

$$S(\vec{u},\vec{n}) = {}^3P(\vec{u})[\ \text{I.D.}\ ] \quad . \tag{1.4}$$

These constraints on the initial data must always hold on $e(S)$. But since $e(S)$ is arbitrarily chosen they must also be satisfied on any other space-like hypersurface. It immediately follows that

$$\vec{n}(S(\vec{u},\vec{n}) - {}^3P(\vec{u})) = 0 \quad . \tag{1.5}$$

Expanding this and using (1.1) and (1.4) we find that (with ${}^3P(\vec{n})=0$)

$$\nabla_{\vec{n}}{}^3P(\vec{u}) = \nabla_{\vec{n}}S(\vec{u},\vec{n}) + \frac{1}{D^\perp}\vec{u}D^\perp S(\vec{n},\vec{n}) - \varepsilon\frac{1}{D^\perp}\sum_i \vec{h}_i D^\perp\, {}^3E(\vec{u},\vec{h}_i) \quad . \tag{1.6}$$

But $\nabla_{\vec{n}}S$ must be independent of $\vec{D}$, so when $\nabla_{\vec{n}}{}^3P(\vec{u})$ is computed directly it must take the form

$$\nabla_{\vec{n}}{}^3P(\vec{u}) = {}^3P'(\vec{u}) + \frac{1}{D^\perp}\vec{u}D^\perp F - \varepsilon\frac{1}{D^\perp}\sum_i \vec{h}_i D^\perp\, {}^3E(\vec{u},\vec{h}_i) \quad , \tag{1.7}$$

where F and ${}^3P'$ are functionals of the initial data. Comparison with (1.6) then gives the secondary constraint

$$S(\vec{n},\vec{n}) = F \quad . \tag{1.8}$$

The equations (1.1),(1.4), and (1.8) are easily recognized as the Einstein field equations:

$$S(\vec{U},\vec{V}) = E(\vec{U},\vec{V}) \quad , \tag{1.9}$$

where E is the symmetric space-time tensor defined by

$$\begin{aligned} E(\vec{u},\vec{v}) &= {}^3E(\vec{u},\vec{v}) \quad , \\ E(\vec{u},\vec{n}) &= {}^3P(\vec{u}) \quad , \\ E(\vec{n},\vec{n}) &= F \quad . \end{aligned} \tag{1.10}$$

By defining the Einstein tensor

$$G(\vec{U},\vec{V}) = S(\vec{U},\vec{V}) - \tfrac{1}{2}g(\vec{U},\vec{V})S \quad , \tag{1.11}$$

and the stress-energy tensor (in natural units)

$$T(\vec{U},\vec{V}) = (1/8\pi)(E(\vec{U},\vec{V}) - \tfrac{1}{2}g(\vec{U},\vec{V})\{\sum_i E(\vec{h}_i,\vec{h}_i) + \varepsilon E(\vec{n},\vec{n})\} \quad , \quad (1.12)$$

which, like E , is an explicit functional of the initial data, the field equations may be recast into their standard form:

$$G(\vec{U},\vec{V}) = 8\pi T(\vec{U},\vec{V}) \quad . \qquad (1.13)$$

The cosmological term appearing in the vacuum equations, (2.4.15), has here been absorbed into T .

The role of the stress-energy as the source of the gravitational (metric) field is now manifest. But we are not finished yet. If the tensor S is to be a genuine Ricci tensor, then it must satisfy the contracted Bianchi identities:

$$\sum_i \nabla_{\vec{h}_i} S(\vec{h}_i,\vec{V}) + \varepsilon \nabla_{\vec{n}} S(\vec{n},\vec{V}) = \tfrac{1}{2}\vec{V}S \quad . \qquad (1.14)$$

Through (1.13), these give us conservation laws that the stress-energy must satisfy:

$$\sum_i \nabla_{\vec{h}_i} T(\vec{h}_i,\vec{V}) + \varepsilon \nabla_{\vec{n}} T(\vec{n},\vec{V}) = 0 \quad ; \qquad (1.15)$$

or, in terms of hypersurface fields,

$$\sum_i \{{}^3\nabla_{\vec{h}_i} {}^3E(\vec{h}_i,\vec{v}) - \tfrac{1}{2}\, {}^3\nabla_{\vec{v}} {}^3E(\vec{h}_i,\vec{h}_i) - K(\vec{h}_i,\vec{h}_i)\, {}^3P(\vec{v})$$

$$- K(\vec{h}_i,\vec{v})\, {}^3P(\vec{h}_i)\} = \varepsilon(\tfrac{1}{2}\vec{v}F - {}^3P'(\vec{v})) \quad , \qquad (1.16)$$

and

$$\vec{n}F = -2\varepsilon\frac{1}{D^\perp}\sum_i \vec{h}_i D^\perp\, {}^3P(\vec{h}_i) + \varepsilon\sum_i\{{}^3E'(\vec{h}_i,\vec{h}_i) + 2K(\vec{h}_i,\vec{h}_i)F$$

$$- 2\,{}^3\nabla_{\vec{h}_i}\,{}^3P(\vec{h}_i) - 2\varepsilon\sum_j K(\vec{h}_i,\vec{h}_j)\,{}^3E(\vec{h}_i,\vec{h}_j)\} \quad . \tag{1.17}$$

Equations (1.16) and (1.17) are constraints on the form of the equations that govern the evolution of the matter fields, so through the Bianchi identities gravity exerts a back-reaction on matter.

Example. To illustrate this coupling of gravity and matter, I shall assume that the only matter field defined on space-time is a real scalar field, ϕ . On the initial hypersurface, $e(S)$, ϕ is characterized by the fields

$$\phi \ , \ \vec{n}\phi \ , \ \{\vec{n}\vec{n}\phi - \nabla_{\vec{n}}\vec{n}\phi\} \ , \text{ etc.}$$

For the functional ${}^3E = \Pi\, {}^3E$, I choose the form defined by

$${}^3E(\vec{u},\vec{v}) = \vec{u}\phi\vec{v}\phi + \tfrac{1}{2}\,{}^3g(\vec{u},\vec{v})\mu^2\,\phi^2 \quad , \tag{1.18}$$

where μ is a constant. Differentiating this along $\vec{n}$ gives

$$\nabla_{\vec{n}}{}^3E(\vec{u},\vec{v}) = \vec{n}({}^3E(\vec{u},\vec{v})) - {}^3E(\nabla_{\vec{n}}\vec{u},\vec{v}) - {}^3E(\vec{u},\nabla_{\vec{n}}\vec{v})$$

$$= \vec{n}\vec{u}\phi\vec{v}\phi + \vec{u}\phi\vec{n}\vec{v}\phi + \tfrac{1}{2}\vec{n}({}^3g(\vec{u},\vec{v}))\mu^2\,\phi^2 + {}^3g(\vec{u},\vec{v})\mu^2\,\phi\vec{n}\phi$$

$$- {}^3E(\tfrac{1}{D^\perp}\pounds_{\vec{D}^\perp}\vec{u} - \varepsilon\sum_i K(\vec{u},\vec{h}_i)\vec{h}_i,\vec{v}) - {}^3E(\vec{u},\tfrac{1}{D^\perp}\pounds_{\vec{D}^\perp}\vec{v} - \varepsilon\sum_i K(\vec{v},\vec{h}_i)\vec{h}_i)$$

$$= \frac{1}{D^\perp}\vec{u}D^\perp\vec{n}\phi\vec{v}\phi + \vec{u}\vec{n}\phi\vec{v}\phi + \frac{1}{D^\perp}\vec{v}D^\perp\vec{n}\phi\vec{u}\phi + \vec{u}\phi\vec{v}\vec{n}\phi + \tfrac{1}{2}\,{}^3g(\nabla_{\vec{n}}\vec{u},\vec{v})\mu^2\,\phi^2$$

$$+ \tfrac{1}{2}\,{}^3g(\vec{u},\nabla_{\vec{n}}\vec{v})\mu^2\,\phi^2 + {}^3g(\vec{u},\vec{v})\mu^2\,\phi\vec{n}\phi - \tfrac{1}{2}\,{}^3g(\tfrac{1}{D^\perp}\pounds_{\vec{D}^\perp}\vec{u},\vec{v})\mu^2\,\phi^2$$

$$+ \ \varepsilon\sum_i K(\vec{u},\vec{h}_i)\{\vec{h}_i\phi\vec{v}\phi + \tfrac{1}{2}\ {}^3g(\vec{h}_i,\vec{v})\mu^2\ \phi^2\} - \tfrac{1}{2}\ {}^3g(\vec{u},\tfrac{1}{D^{\perp}}L_{\vec{D}^{\perp}}\vec{v})\mu^2\ \phi^2$$

$$+ \ \varepsilon\sum_i K(\vec{v},\vec{h}_i)\{\vec{h}_i\phi\vec{u}\phi + \tfrac{1}{2}\ {}^3g(\vec{u},\vec{h}_i)\mu^2\ \phi^2\}$$

$$= \{\vec{u}(\vec{n}\phi)\vec{v}\phi + \vec{u}\phi\vec{v}(\vec{n}\phi) + {}^3g(\vec{u},\vec{v})\mu^2\ \phi\vec{n}\phi + \varepsilon\sum_i(K(\vec{u},\vec{h}_i)\vec{v}\phi$$

$$+ \ K(\vec{v},\vec{h}_i)\vec{u}\phi)\vec{h}_i\phi\} + \frac{1}{D^{\perp}}\vec{v}D^{\perp}\vec{n}\phi\vec{u}\phi + \frac{1}{D^{\perp}}\vec{u}D^{\perp}\vec{n}\phi\vec{v}\phi \qquad , \qquad (1.19)$$

where extensive use has been made of (2.2.11) and (2.2.16). As required, this final expression takes the form stipulated in (1.2); and by identifying the appropriate terms we find

$$^3E'(\vec{u},\vec{v}) = \vec{u}(\vec{n}\phi)\vec{v}\phi + \vec{u}\phi\vec{v}(\vec{n}\phi) + {}^3g(\vec{u},\vec{v})\mu^2\ \phi\vec{n}\phi$$

$$+ \ \varepsilon\sum_i\{K(\vec{u},\vec{h}_i)\vec{v}\phi + K(\vec{v},\vec{h}_i)\vec{u}\phi\}\vec{h}_i\phi \qquad , \qquad (1.20)$$

$$^3P(\vec{u}) = \vec{n}\phi\vec{u}\phi \qquad . \qquad (1.21)$$

If we set $^3P(\vec{n}) = 0$, then (1.21) can be differentiated to give

$$\nabla_{\vec{n}}{}^3P(\vec{u}) = \vec{n}({}^3P(\vec{u})) - {}^3P(\nabla_{\vec{n}}\vec{u})$$

$$= \vec{u}\phi(\vec{n}\vec{n}\phi - \nabla_{\vec{n}}\vec{n}\phi) + \vec{n}\phi(\vec{u}\vec{n}\phi + \varepsilon\sum_i K(\vec{u},\vec{h}_i)\vec{h}_i\phi)$$

$$+ \ \frac{1}{D^{\perp}}\vec{u}D^{\perp}(\vec{n}\phi\vec{n}\phi + \tfrac{1}{2}\varepsilon\mu^2\ \phi^2) - \varepsilon\frac{1}{D^{\perp}}\sum_i\vec{h}_iD^{\perp}(\vec{u}\phi\vec{h}_i\phi + \tfrac{1}{2}\ {}^3g(\vec{u},\vec{h}_i)\mu^2\ \phi^2) \ . \qquad (1.22)$$

This takes the required form, (1.7), and again identifying terms we obtain

$$F = \vec{n}\phi\vec{n}\phi + \tfrac{1}{2}\varepsilon\mu^2\ \phi^2 \qquad , \qquad (1.23)$$

$$^3P'(\vec{u}) = \vec{u}\phi(\vec{n}\vec{n}\phi - \nabla_{\vec{n}}\vec{n}\phi) + \vec{n}\phi(\vec{u}\vec{n}\phi + \varepsilon\sum_i K(\vec{u},\vec{h}_i)\vec{h}_i\phi) \quad . \qquad (1.24)$$

The functional forms of ^3E, $^3E'$, ^3P, $^3P'$, and F can now be substituted into the conservation laws (1.16) and (1.17) to give, respectively

$$\{\sum_i(\vec{h}_i\vec{h}_i\phi - {}^3\nabla_{\vec{h}_i}\vec{h}_i\phi - K(\vec{h}_i,\vec{h}_i)\vec{n}\phi) + \varepsilon(\vec{n}\vec{n}\phi - \nabla_{\vec{n}}\vec{n}\phi) - \mu^2\phi\}\vec{u}\phi = 0 \quad , \qquad (1.25)$$

$$\{\sum_i(\vec{h}_i\vec{h}_i\phi - {}^3\nabla_{\vec{h}_i}\vec{h}_i\phi - K(\vec{h}_i,\vec{h}_i)\vec{n}\phi) + \varepsilon(\vec{n}\vec{n}\phi - \nabla_{\vec{n}}\vec{n}\phi) - \mu^2\phi\}\vec{n}\phi = 0 \quad . \qquad (1.26)$$

These equations must be satisfied everywhere on $e(S)$, but since at generic points $\vec{n}\phi$ and/or $\vec{u}\phi$ are non-zero we must set

$$\sum_i(\vec{h}_i\vec{h}_i\phi - \nabla_{\vec{h}_i}\vec{h}_i\phi) + \varepsilon(\vec{n}\vec{n}\phi - \nabla_{\vec{n}}\vec{n}\phi) - \mu^2\phi = 0 \quad . \qquad (1.27)$$

This last equation is easily recognized as the Klein Gordon equation, and from it we deduce that the initial data for ϕ consists of just the two fields, ϕ and $\vec{n}\phi$, on $e(S)$. The space-time metric must satisfy the field equations (1.9), with the explicit functionals obtained above being used in (1.10) to define E.

It is clear from the equations (1.13) (or (1.9)) that, as in the vacuum case, the dynamical equations governing the evolution of the metric are always supplemented by a set of primary and secondary constraint equations in such a way that the complete set is covariant in the space-time fields from which the initial data is constructed. This is equally true when it is supposed that time derivatives of g beyond the first are included in the initial data (i.e. $m>1$). Thus, for $m = 2$, the vacuum equations (2.3.5) generalize immediately to

$$\nabla S = \nabla S[g,R,F_j,\nabla F_j,\ldots,\nabla^{m_j}F_j] \qquad (1.28)$$

when matter fields are present.

The back-reaction of the space-time geometry on the matter fields is also present, but less obvious, when $m > 1$. Once the geometrical field equations (i.e. equations (1.28) for $m = 2$) have been chosen, one must always check to see that they are compatible with the Bianchi identities (1.14); imposing restrictions on the matter field equations to assure this. These restrictions, when they are necessary, represent geometry's reaction on matter.

Before proceeding to the next section, a few brief remarks regarding gauge fields are in order. For convenience in the foregoing discussions, I have implicitly assumed that the distribution of matter in space-time is characterized by a unique configuration of the fields F_j. However, it is well known that many different configurations of the same set of fields (here more appropriately called gauge potentials) may actually provide equivalent, complete characterizations of the same matter distribution [12]. Moreover, it may be necessary to define the potentials, F_j, piecewise on overlapping neighbourhoods, in order to cover the entire space-time manifold [13]. (If $F_j^{(1)}$ and $F_j^{(2)}$ are the field configurations on the overlapping neighbourhoods $\mathcal{U}_1$ and $\mathcal{U}_2$ of M, respectively, then on the overlap region, $\mathcal{U}_1 \cap \mathcal{U}_2$, both $F_j^{(1)}$ and $F_j^{(2)}$ characterize the same matter distribution.)

The degrees of freedom in the fields F_j that are not needed to uniquely specify the matter distribution are called the gauge degrees of freedom, and the associated (gauge fixing) fields have no physical significance. When written in their four dimensional form in terms of space-time fields, the physical field equations make no reference to these non-physical fields. Nonetheless, in order to cast the field equations into an initial value form, specific gauge fixing conditions, which will have no ultimate effect on the physical predictions, must be chosen. In the first part of this section, no mention was made of these arbitrary gauge conditions, but since the gauge conditions have no influence on the physics, no generality was lost.

2. Alternative Geometries and Unified Field Theories

A great number of researchers have tried, during the past sixty-five years, to develop a new theory that maintains the philosophical and empirical successes of GR while either extending its domain of validity or else evading some of the philosophical problems that plague GR. The main premise of almost all such efforts is that the pseudo-Riemannian geometry of GR is too restrictive to provide a complete description of the world and, in particular, that the "physical" covariant derivative has non-vanishing torsion.

Einstein himself was never completely happy with GR, primarily because of its singular solutions. Considering GR to be just a macroscopic theory, he hoped to be able to find a more complex geometric theory that would yield a singularity-free model for an elementary particle. As early as 1928 Einstein suggested a theory of gravity with non-vanishing torsion, but zero curvature [14]. His later efforts to construct a unified theory of gravity and electromagnetism [15] presumed a still more complicated geometry, with a non-symmetric fundamental tensor, the symmetric part of which was a locally Minkowskian metric, and again a non-trivial torsion tensor. Although Einstein never developed a completely acceptable model, it has been shown recently by Moffat and co-workers [16] that all of the phenomenology of gravitation and (classical) electromagnetism may be understood within the context of the (pseudo-)hermitian geometry of the Einstein-Schrodinger theory ([1],[17]). Moffat [18] has also shown that a variation on the Einstein-Strauss theory [15] can lead to particle-like solutions which are non-singular in the sense that they are world-line complete, even though there are singularities in some of the field invariants.

The desire to obtain a renormalizable quantum theory of gravity seems to be the main reason for renewed interest in

Einstein-Cartan type theories [19]. Several different models have been proposed [20], with spin being coupled to gravity, through the torsion, in a non-trivial way. However, all such efforts seem to lead to a torsion field which is algebraically related to spin, and which, therefore, does not propagate as an independent field.

On the surface, it may seem as though the formalism I have developed excludes from consideration any kind of geometric structure for space-time other than the pseudo-Riemannian geometry of GR, and thus all of the "generalized" or "unified" theories based on alternative kinds of geometry. This is not the case, however. All that I have done was to separate the metric defined on space-time from any other tensor fields that are pertinent to physics, and then determine what sorts of equations are capable of propagating the metric forward in time. Since, in each of the theories discussed above, the alternative geometries always include a metric tensor, and since the metric must always propagate, the results of Section 1 remain applicable even for theories with non-Riemannian geometry, provided additional geometric fields such as the torsion or the skew part of a non-symmetric fundamental tensor are treated as "matter" fields.

This general applicability of the (pseudo-)Riemannian results should not be surprising, and has actually been known for a long time [21]. It follows from the well known fact that the difference of any two affine connections is a tensor field.

What it implies is that any theory that is based on a non-Riemannian geometry (which includes a metric) may always be reformulated in terms of (pseudo-)Riemannian geometry plus tensor fields; and, in particular, any physical theory whose field equations include derivatives of the metric up to and including second order, but no higher, is mathematically equivalent to Einstein's general theory of relativity (with sources). Thus, while they cannot be dismissed altogether, the advantages of introducing alternative geometries seem limited to the motivation of field equations different from those that would normally be investigated, and of new interpretations for physical fields.

3. Already Unified Theory

Rather than probing new kinds of geometries, Misner and Wheeler [9] followed the early work of Rainich [8] and showed that the conventional (pseudo-)Riemannian space-time already provides a sufficiently rich structure to accommodate both gravity and electromagnetism.

For compactness in the exposition of their findings, I shall now adopt a component notation, with indices $i,j,\ldots$ and $\alpha,\beta,\ldots$ ranging from 0 to 3, and repeated indices being summed. The vectors $\vec{h}_i$, $i=0,1,2,3$, will now represent a vierbein field:

$$g(\vec{h}_i,\vec{h}_j) = \eta_{ij} \quad , \tag{3.1}$$

while $\vec{\partial}_\alpha$, $\alpha=0,1,2,3$, are the coordinate basis vectors for some implicit coordinate chart. Corresponding 1-forms $\underline{h}^i$ and $\underline{dx}^\alpha$ are defined by

$$\underline{h}^i(\vec{h}_j) = \delta^i{}_j \quad \text{and} \quad \underline{dx}^\alpha(\vec{\partial}_\beta) = \delta^\alpha{}_\beta \quad . \tag{3.2}$$

The "already unified" theory of Misner and Wheeler is not a new theory, but just standard Einstein-Maxwell theory (with a source-free electromagnetic field) written in a purely geometrical form. Let $\underline{F} = \frac{1}{2}F_{\alpha\beta}\underline{dx}^\alpha \wedge \underline{dx}^\beta = \frac{1}{2}F_{ij}\underline{h}^i \wedge \underline{h}^j$ be the 2-form representing the electromagnetic field. Then its Poincaré dual is the 2-form $*F = \frac{1}{2}{*F}_{\alpha\beta}\underline{dx}^\alpha \wedge \underline{dx}^\beta$ whose components are defined by

$$*F_{\alpha\beta} = F^{\mu\nu}\underline{\mathrm{Det}}(\vec{\partial}_\alpha, \vec{\partial}_\beta, \vec{\partial}_\mu, \vec{\partial}_\nu) \quad , \tag{3.3}$$

where $F^{\mu\nu} = g^{\mu\lambda}\, g^{\nu\sigma}\, F_{\lambda\sigma}$ and $\underline{\mathrm{Det}}$ is the volume 4-form

$$\underline{\mathrm{Det}} = \underline{h}^0 \wedge \underline{h}^1 \wedge \underline{h}^2 \wedge \underline{h}^3 \quad . \tag{3.4}$$

In terms of $\underline{F}$ and $*\underline{F}$, the source-free Maxwell equations take the simple form

$$\underline{dF} = 0 \quad \text{and} \quad \underline{d}*\underline{F} = 0 \quad , \tag{3.5}$$

and the Maxwell stress-energy tensor has components

$$T_{\alpha\beta} = F_{\alpha\mu}\, F_\beta{}^\mu + *F_{\alpha\mu}\, *F_\beta{}^\mu \quad . \tag{3.6}$$

The complete Einstein-Maxwell system thus consists of equations (3.5) and (1.13), with the stress-energy tensor in (1.13) being given by (3.6).

Rainich [8] showed that, quite independent of the Maxwell equations, (3.5), any Ricci tensor arising from (1.13) with the stress-energy tensor (3.6) must satisfy

$$S \equiv S_\alpha{}^\alpha = 0 \quad , \tag{3.7}$$

$$S_\alpha{}^\beta \, S_\beta{}^\gamma = \delta_\alpha{}^\gamma (S_{\sigma\tau} \, S^{\sigma\tau}/4) \quad , \tag{3.8}$$

$$S_{00} \geqslant 0 \quad . \tag{3.9}$$

Misner and Wheeler proved the converse, showing that any geometry whose Ricci tensor satisfies the Rainich conditions, (3.7), (3.8), and (3.9), can be represented as the "Maxwell square", (3.6), of some skew field $\underline{F}$. They showed, moreover, that the field $\underline{F}$ is uniquely determined by S (using equations (3.6) and (1.13)) up to a global <u>duality rotation</u>:

$$\underline{F} \to e^{*\alpha} \, \underline{F} = \underline{F} \, \mathrm{Cos}\, \alpha + {}^*\underline{F} \, \mathrm{Sin}\, \alpha \quad . \tag{3.10}$$

Defining the 1-form $\underline{\alpha} = \alpha_\mu \underline{dx}^\mu$ by the equation

$$\alpha_\tau = (\mathrm{Det})_{\tau\lambda\mu\nu} \, S^{\lambda\beta;\mu} \, S_\beta{}^\nu / (S_{\gamma\delta} S^{\gamma\delta}) \quad , \tag{3.11}$$

they then showed that if (3.7), (3.8), and (3.9) are satisfied and

$$\underline{d\alpha} = 0 \quad , \tag{3.12}$$

then the field $\underline{F}$ whose Maxwell square is S will automatically satisfy the Maxwell equations, (3.5); and they gave an explicit procedure (the details of which are not important here) for finding $\underline{F}$, given S , in the restricted case that S is not null ($S_{\sigma\tau}S^{\sigma\tau} \neq 0$).

The equations (3.12) are fourth order differential equations in the components of the space-time metric, g , and taken together with the Rainich conditions, (3.7), (3.8), and (3.9), and the Misner-Wheeler procedure for finding $\underline{F}$, given S , they constitute a purely geometrical, "already unified" way of representing the Einstein-Maxwell equations. The electromagnetic field, in this picture, is a derivative quantity and never enters on a fundamental level. The only fundamental field is g .

While its development was a great achievement, the already unified theory of gravity and electromagnetism is not without problems. The first of these is that it is unable to cope with electromagnetic fields that are null on any set of measure greater than zero. This is not too severe a restriction, though, since in reasonable physical situations one would expect the electromagnetic field to have a coulomb component that is non-vanishing at generic points. More serious problems are the lack of a Lagrangian formulation for the theory, and the certainty that any linearized version or initial value formalism of the already unified theory would be indistinguishable from corresponding treatments of the Einstein-Maxwell theory.

Aside from its failings, and the obvious fact that no experiment can distinguish it from Einstein-Maxwell theory, I find the already unified theory very interesting for the following reason. Suppose that Einstein-Maxwell theory actually provides a "correct" description of the mutually interacting gravitational and electromagnetic fields, and that in the region M' of space-time there are no matter fields other than the electromagnetic. Then the work of Rainich, Misner, and Wheeler indicates that if we make thorough measurements of the metric on M', then we can *deduce* from those measurements the configuration of the electromagnetic field (up to an overall duality rotation). *There is no need to make measurements of the electromagnetic field independently of the measurements of the space-time geometry.*

Even if one wanted to make direct measurements of the electromagnetic field, how would one do it? The simplest procedure would be to take a known charged particle, say an electron; set it adrift with some initial velocity, $\vec{v}$; and make careful observations of its trajectory. But in charting its trajectory through space-time we would be measuring distances - that is, measuring the space-time geometry - so we really wouldn't be making direct measurements of the electromagnetic field.

I have not studied the problem sufficiently to make a definitive statement, but I suspect that all attempts to make direct measurements of the electromagnetic field would be similarly doomed. Jumping far beyond the domain of electromagnetism I shall adopt the following hypothesis:

Hypothesis: The only physical field that may be measured directly is the space-time metric. The configurations of all matter fields defined on space-time must be deduced from the space-time geometry.

If this is the case, and if we still want to think of the matter fields as being somehow fundamental (as they are in quantum mechanics), then should we not think of the metric as being just the messenger of the matter fields, shouting out their existence and their characteristics as clearly as possible without colouring or obscuring the message unnecessarily with its own idiosyncracies? But the metric would fulfil this task most readily if it were to couple to the matter fields in the simplest possible way, through the Einstein equations (1.13), leaving the determination of the space-time geometry completely up to the matter fields. In any higher order dynamical equations for the metric ($m > 1$), all of the components of the curvature would be included in the initial data (albeit constrained) making it potentially impossible to decide what portion of the curvature is "gravitational" in origin and what portion should be ascribed to the matter fields.

It also follows from my hypothesis, that if the various matter fields are to be perceived, and distinguished from each other, then they must each leave a distinctive "imprint", analogous to the imprint created by the electromagnetic field, on the space-time geometry; and accordingly, that there must exist a <u>super already unified theory</u> capable of providing a purely geometrical description of all forms of matter. Any reasonable theory of interacting metric and matter fields must therefore be able to be recast into an already unified form, and if it cannot be, then it must be rejected.

The last two paragraphs were, clearly, quite conjectural, and I do not intend that they be taken as more than that. Nonetheless, I believe that these conjectures deserve further investigation, firstly because of the remarkable Rainich, Misner, Wheeler results, and secondly because only by pursuing such ideas can we ever hope to gain an understanding, based on physical ideas rather than mathematical conveniences, of why the particular equations we use to describe the world should be more appropriate than any other set.

4. Global Considerations

Throughout the foregoing discussions no assumptions have been made about the global topology of space or space-time. Instead, I have restricted my attention to some open cell, M', in space-time on which all physical fields are well defined and of class C^{∞}. For each generic point, x, of space-time there exists such a cell containing x, so all of the conclusions I have drawn respecting field equations hold at all generic points.

The global topology of space-time becomes important, however, when one starts investigating solutions to the field equations. The interrelationship between the pseudo-Riemannian geometry and the global topology of space-time, and the topology of space-like hypersurfaces of space-time, is discussed extensively by Hawking and Ellis [22]. Here, I would just like to point out that since any manifold may be constructed by piecing together open cells, any solution of a set of field equations may be (and in practice is) constructed by piecing together solutions defined on open cells.

CHAPTER 4

THE TOPOLOGICAL WORLD

1. Gravity in a Quantum World

Classical concepts are adequate for the description and understanding of all observed features of gravitation. However, it is firmly established that the behaviour of matter in the real world can be fully understood only within the context of quantum theory. Material phenomena that can be effectively described with the use of classical (as opposed to quantum) variables arise as a consequence of the quantum behaviour of large systems, and are only manifested in macroscopic systems. Since general relativity, as it was formulated by Einstein and as I have presented it above, couples the space-time geometry to classical descriptors of matter, it is reasonable to conclude that GR is only valid when macroscopic (classical) systems are being investigated.

This limitation of GR has long been recognized, and many attempts have been made to remove it by constructing new theories that are valid in the quantum domain and which reduce to GR in the classical limit. The most common line of attack is to quantize the space-time metric much as one would any other field [23], however, quantized GR has been shown to be non-renormalizable [24] and it seems unlikely that renormalizability can be restored within a model that has the correct classical behaviour [25].

An alternative "semi-classical" approach was proposed by Møller [10]. He suggested that, rather than quantizing the gravitational field, it might be more appropriate to continue to think of it as a c-number field, with the expectation value of the quantum stress-energy tensor as its source:

$$G(\vec{U},\vec{V}) = 8\pi \langle T_{op}(\vec{U},\vec{V}) \rangle \quad . \tag{1.1}$$

These equations are to be solved self-consistently, with the quantum fields that contribute to the stress-energy being defined on the curved space-time that they determine. The major difficulty with the coupling (1.1) is that the simple normal ordering procedure used in special relativistic QFT to eliminate the zero-point energy from $\langle T_{op} \rangle$ has no obvious unique analogue on curved space-time [26]. The situation is not hopeless, however, and recent results obtained by imposing physical renormalization conditions at each order in perturbation theory [36] may well lead to a resolution of the problem. One of the most interesting features of this gravity modified quantum theory is that the linear superposition principle ceases to be valid because equation (1.1) is non-linear. This certainly represents a dramatic break from conventional quantum theory, but, as has been demonstrated by Everett in his "many worlds" interpretation of quantum mechanics [27], it is not unreasonable to assume that the entire universe is described by a single, smoothly evolving wave function, thereby eliminating the need for a superposition principle.

In addition to the two traditional approaches to the reconciliation of curved space-time and quantum theory, there have been several suggestions advocating the adoption of radically different world-views. Most notable amongst these are the twistor theory [28] being developed by Penrose and co-workers, in which the spin group $SL(2,\mathbf{C})$ plays a central role, and Finkelstein's space-time code [29] in which quantum processes are considered as fundamental and space secondary. Unfortunately, these theories are extremely complicated and they seem to be quite arbitrary. Unless they can be made more intuitive it is unlikely that they will ever gain popular acceptance.

Much of the recent interest in quantum gravity seems to have been stimulated by developments in elementary particle theory. The non-abelian gauge theories have provided a single formalism capable of handling all strong, weak, and electromagnetic interactions; and, as a bonus, the fibre bundle picture of gauge fields makes them look (at least superficially) similar to the gravitational field [13]. If the graviton could be added to the elementary particle zoo, using gauge theory techniques, then particle theory would, in a sense, be complete. From a different viewpoint, gravity appeared as the only remaining physical phenomenon that might be able to eliminate the singularities that occur throughout quantum field theory. In either case, it was (and is) suspected that gravity and particle physics are linked in some fundamental way, and that neither one can be fully understood without the other.

In the remainder of this chapter I shall pursue this connection between gravitation and particle physics, looking for answers, not in quantum theory, but in the structure of space-time. I take my guidance from Einstein, who showed most elegantly that the removal of unnecessarily restrictive assumptions can reveal beautiful and exciting physics.

2. Pregeometry is No Geometry

The key to special relativity was the revelation that time need not be absolute. Einstein quickly realized, though, that even the Minkowski space-time was too restrictive - its absolute geometric structure could not be justified. Freeing up the geometry led naturally to GR and an understanding of gravity never before dreamed of.

In early investigations of curved space-time it was just assumed that space and space-time have the topologies of $\mathbb{R}^3$ and $\mathbb{R}^4$, respectively, but it was soon realized that this assumption was also too restrictive. Observations of the universe extend out only a finite distance, so on a very large scale the topology is indeterminate. Other topologies (than $\mathbb{R}^4$) were investigated and gave interesting results, and GR quickly assumed a central role in the field of cosmology [30].

In the other direction, at small distances rather than large, the situation is similar. Experimentally, we have only been able to probe to 10^{-16}cm , which is a long way from the Planck length, $L_P \approx 10^{-33}$cm , at which gravitational effects are expected to significantly influence quantum processes. There is thus no compelling physical reason to suppose that the space-time topology remains trivial at lengths less than 10^{-16}cm (or 10^{-33}cm if you wish to be more conservative). Indeed, it has often been suggested that the space-time topology becomes extremely complex at short distances, with the degree of complexity increasing as the length scale decreases. As an extension of their already unified theory, Misner and Wheeler [9] showed that charge could be recovered from source-free electrodynamics by assuming a multiply connected space-time. Looking more towards quantum gravity, Wheeler [31] conjectured that space is a "foam-like" structure whose topology is constantly changing due to quantum fluctuations at lengths of the order of L_P . He envisaged particles as being macroscopic collective modes of the fluctuating topology/geometry.

With the completely arbitrary topology of Wheeler's quantum geometrodynamics, it would seem as though Einstein's programme of removing restrictive assumptions about the structure of space and time has been brought to a conclusion. But Wheeler is not yet satisfied. He argues that if one can obtain electromagnetism without electromagnetism and charge without charge, then one should also be able to obtain geometry without geometry; and he has coined the word pregeometry to symbolize the structure from which geometry arises.

The nature of pregeometry is very vague. Wheeler surmises that it might be topological or even "pretopological", but he has no specific model for it. By drawing a picture in which each topological configuration of space is endowed with a geometric structure, he seems to imply, however, that pregeometry is something conceptually distinct from (and in addition to) the topology of space or space-time. I believe that this picture is unnecessarily complicated, and that the pregeometry which Wheeler seeks is nothing more than the topology of space-time.

Consider a 4-dimensional topological manifold, $\mathcal{W}$, whose global topology is extremely complex. Although it is always possible to assign to $\mathcal{W}$ some particular geometry or field structure, I shall assume that all such fields are irrelevant - $\mathcal{W}$ is completely characterized by its topology. Moreover, the global topology of $\mathcal{W}$ is not subject to any restrictions beyond those that are necessary to preserve the manifold structure. Now imagine trying to describe some of the gross features of the global structure of $\mathcal{W}$ without knowing all the fine features. The usual topological descriptors become useless because they depend on a complete knowledge of the details, but perhaps there is an alternative mode of description. If we consider a topologically simple 4-manifold, $\mathcal{M}$, then perhaps we can replace (or symbolically represent) some of the topological complexities of $\mathcal{W}$ with the use of appropriately chosen fields on $\mathcal{M}$.

This is the basic picture of physics that I shall begin to develop below. I think of the objective world underlying all of our perceptions as an unimaginably complex, 4-dimensional manifold, $\mathcal{W}$. All of physics for all of time is coded into the topology of $\mathcal{W}$, but even this vast amount of information represents only a tiny fraction of the information contained in $\mathcal{W}$. If $\mathcal{W}$ is the (real) objective world, then the space-time-matter world of our perceptions is but a faint shadow. Matter and all its properties, life, and even intellect are contained in that shadow. In this world of perceptions, most information about the topological complexities of $\mathcal{W}$ is lost, and the remainder is represented by matter fields defined on a topologically simple, geometrical space-time manifold. Conventional space-time thus emerges as a replacement manifold for $\mathcal{W}$ - a simplified version of $\mathcal{W}$ which has no objective existence.

Wheeler's vision of transcending geometry is realized, not by appealing to some new mathematical or logical structure, but by recognizing the stupendous amount of information that can be coded into the topology of a 4-dimensional manifold. If we think of $\mathcal{W}$ as space-time viewed on a deeper level, then pregeometry is the space-time topology. When geometry is born, topological complexities must die; so geometric space-time is topologically simple at small distances (in contrast with Wheeler's geometrical space-time foam).

3. Breaking the Topological Code

My conjecture is that all of physics is encoded in the mathematics of 4-dimensional topological manifolds, and that the particular world we are a part of corresponds to a particular manifold, $\mathcal{W}$. The problem now is to break the code and extract physical laws from an (almost) unconstrained mathematical system. I cannot claim to have done this, but I do have suggestions for a scheme that I consider to be worth pursuing.

Again, just as in Chapter 2 , I shall begin with an investigation of 3-dimensional manifolds. My principal reference here is "3-manifolds" by John Hempel [32], which is quite a complete survey of progress, up until 1976, on the problem of classifying all 3-manifolds. This is a very difficult problem in topology and most of the techniques being used to solve it are beyond the grasp of a novice like myself. Nonetheless, there are some general results that are easily apprehended, and which seem particularly useful for the physics problem I have set myself.

Let M_1 and M_2 be connected 3-manifolds (possibly with boundaries) and let B_1 , B_2 be closed 3-cells in the interiors of M_1 , M_2 , respectively. Removing the interiors of these cells leaves the remainders $R_i = M_i - \text{Int } B_i$, $i = 1,2$. A third 3-manifold M is said to be a <u>connected sum</u> of M_1 and M_2 if there exist embedding maps $e_i : R_i \to M$ such that $e_1(R_1) \cap e_2(R_2) = e_1(\partial B_1) = e_2(\partial B_2)$ and $M = e_1(R_1) \cup e_2(R_2)$. This is denoted by $M = M_1 \# M_2$. If either M_1 or M_2 is non-orientable then

$M_1 \# M_2$ is unique up to equivalence; but if both are oriented then two distinct manifolds may arise, corresponding to the cases when $e_1^{-1} \circ e_2$ is an orientation preserving and orientation reversing homeomorphism of the 2-sphere boundaries. In situations I shall consider, this ambiguity will never occur.

"Connected sum" is a well defined associative and commutative operation, so for any finite k the notation $M_1 \# M_2 \# \ldots \# M_k$ is unambiguous.

For any 3-manifold M it is obvious that $M \# S^3 = M$, so the 3-sphere behaves as an identity element for connected sums. M is said to be <u>prime</u> if $M = M_1 \# M_2$ implies that one of M_1, M_2 is a 3-sphere.

<u>3.1</u> <u>Theorem</u> Each compact 3-manifold can be expressed as a connected sum of a finite number of prime factors [32].

Prime decomposition is not unique. Hempel shows that if $M = M_1 \# (S^2 \times S^1)$ where M_1 is non-orientable, then $M = M_1 \# P$. Here P is the non-orientable S^2 bundle over S^1 (the 3-dimensional analogue of the Klein bottle) and both $S^2 \times S^1$ and P are prime. To get around this problem he defines a <u>normal</u> prime factorization of a 3-manifold M to be a prime factorization $M = M_1 \# \ldots \# M_k$ such that some M_i is $S^2 \times S^1$ only if M is orientable. This leads to the central result:

<u>3.2</u> <u>Theorem</u> Let $M = M_1 \# \ldots \# M_k = M^*_1 \# \ldots \# M^*_{k^*}$ be two

normal, prime factorizations of a compact 3-manifold M. Then $k = k^*$ and (after reordering) M_i is homeomorphic to M^*_i [32]. That is, there is a unique normal, prime factorization for each compact 3-manifold (with boundary).

With this last result, the problem of classifying all compact 3-manifolds is reduced to the problem of finding and classifying all prime 3-manifolds. However this is still a very difficult task which is far from complete. Even though an infinite number of prime 3-manifolds have already been identified there are many more yet to be found.

Now, what I want to do is to build up the topological space-time manifold, W, by stacking together 3-dimensional submanifolds. This will create a picture concordant with the perceived special status of space-like hypersurfaces in geometrical space-time. It will also introduce the concept of time on a fundamental level. Although it is not clear whether this assumption is necessary or not, I shall assume, for illustrative purposes, that W is endowed with a differential structure of class C^∞. The Whitney embedding theorem [33] then allows me to consider W as a smooth submanifold of $\mathbb{R}^8$; inducing on W a (non-physical) Riemannian metric, g_8.

Let B be a closed 8-cell in $\mathbb{R}^8$ such that $W' = W \cap B$ is a connected, compact (yet still extremely complex), 4-dimensional submanifold-with-boundary of W and $\partial W' = W \cap \partial B$ is a

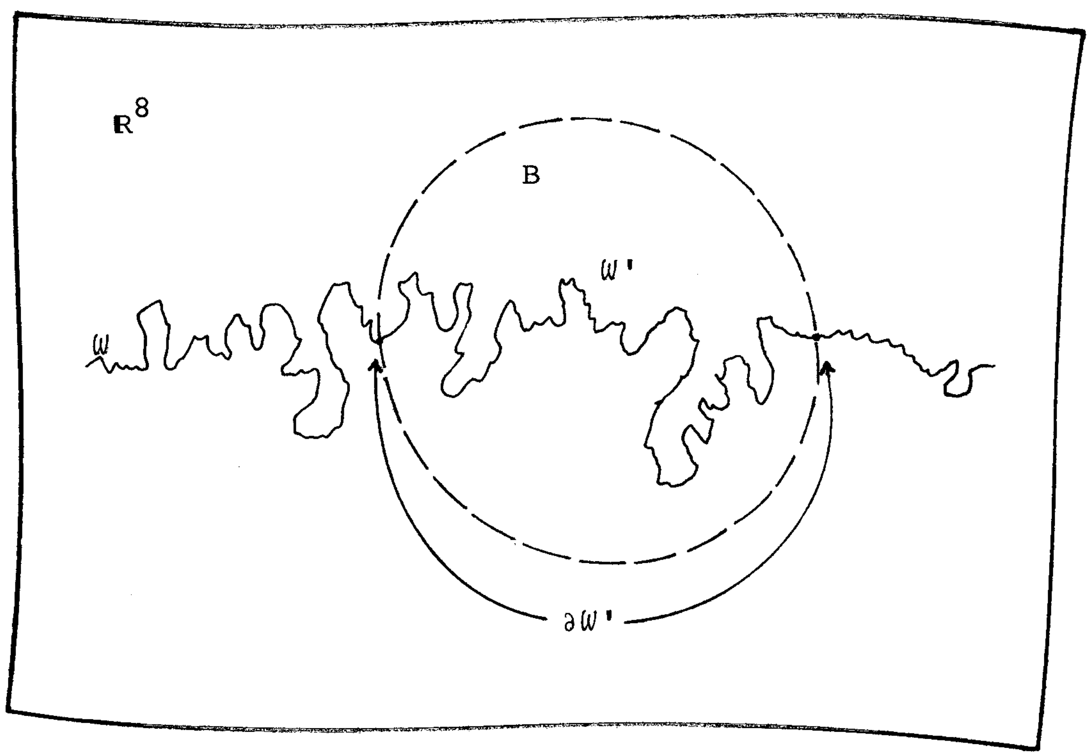


Figure 3.1 $\mathcal{W}$ may always be considered as a submanifold of $\mathbb{R}^8$. The closed 8-cell, B, is chosen such that $\mathcal{W}' = \mathcal{W} \cap B$ is connected and $\partial\mathcal{W}' = \mathcal{W} \cap \partial B$ is closed.

closed 3-manifold (Figure 3.1). W' will correspond to a closed 4-cell in <u>geometric</u> space-time, whereas the entire geometric space-time (corresponding to all of W) may have a non-trivial topology. Choose, for the "initial" 3-dimensional submanifold of W' , a submanifold S_0 such that $\partial S_0 = S_0 \cap \partial W'$ is a 2-sphere which divides $\partial W'$ into two pieces, one to the "past" and one to the "future" of ∂S_0 . With this condition on the boundary, the prime decomposition of S_0 must take the form

$$S_0 = M_1 \# \ldots \# M_k \# B^3 \quad , \tag{3.3}$$

where B^3 is the closed ball in 3-dimensions and the manifolds M_i have no boundaries [34].

Turn now to the geometry induced on S_0 by the embedding of W in $\mathbb{R}^8$. Since its geometry has no physical significance, we may deform W (and B) in $\mathbb{R}^8$ to make it assume whatever geometric configuration we wish, subject of course to the constraints imposed by topology. In particular, we can assume that the embedding has been chosen so that S_0 takes the form of a Euclidean space onto which a large number, k , of small, widely spaced, prime 3-manifolds, M_i , have been fastened (Figure 3.2). For comparison, the 2-dimensional analogue of S_0 is a Euclidean disc onto which have been fastened (by cutting and pasting) a large number of very small and widely separated handles (or crosscaps). The main difference is that 3 dimensions provides an infinite variety of distinct objects, rather than just

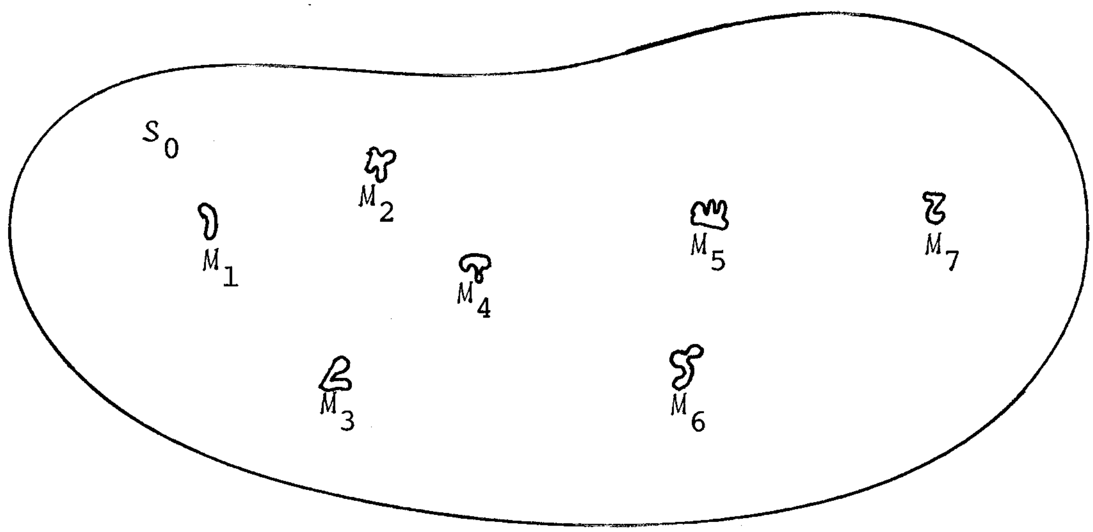


Figure 3.2 The geometry of S_0 is chosen to make it appear like a Euclidean space onto which small, widely separated, prime 3-manifolds, M_i , have been fastened.

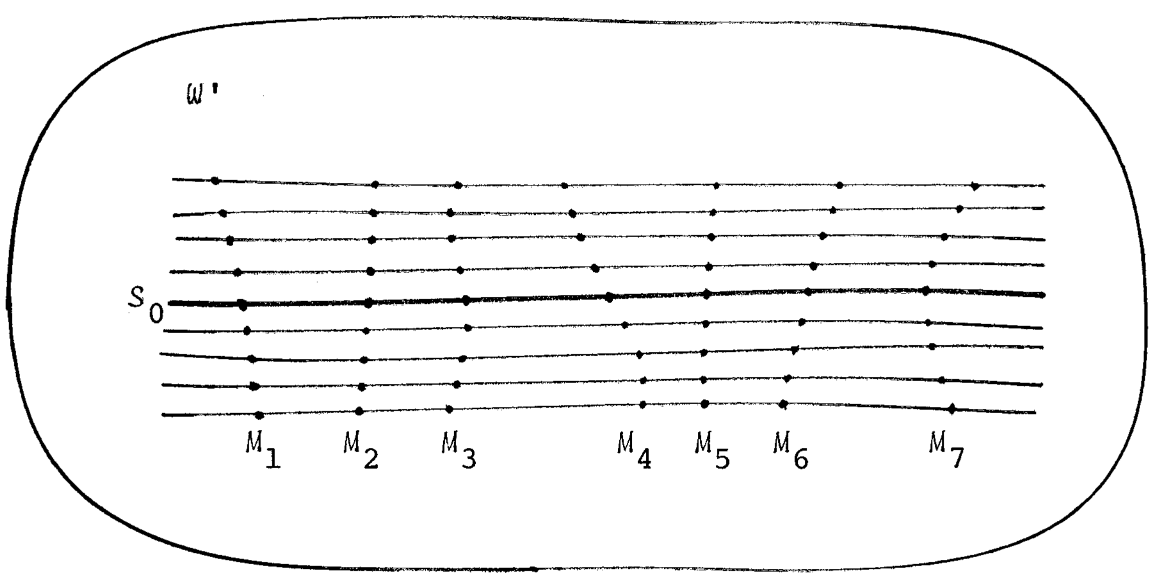


Figure 3.3 Hypersurfaces near S_0 are assigned similar geometries so that the topological anomalies appear to travel on smooth paths through W' .

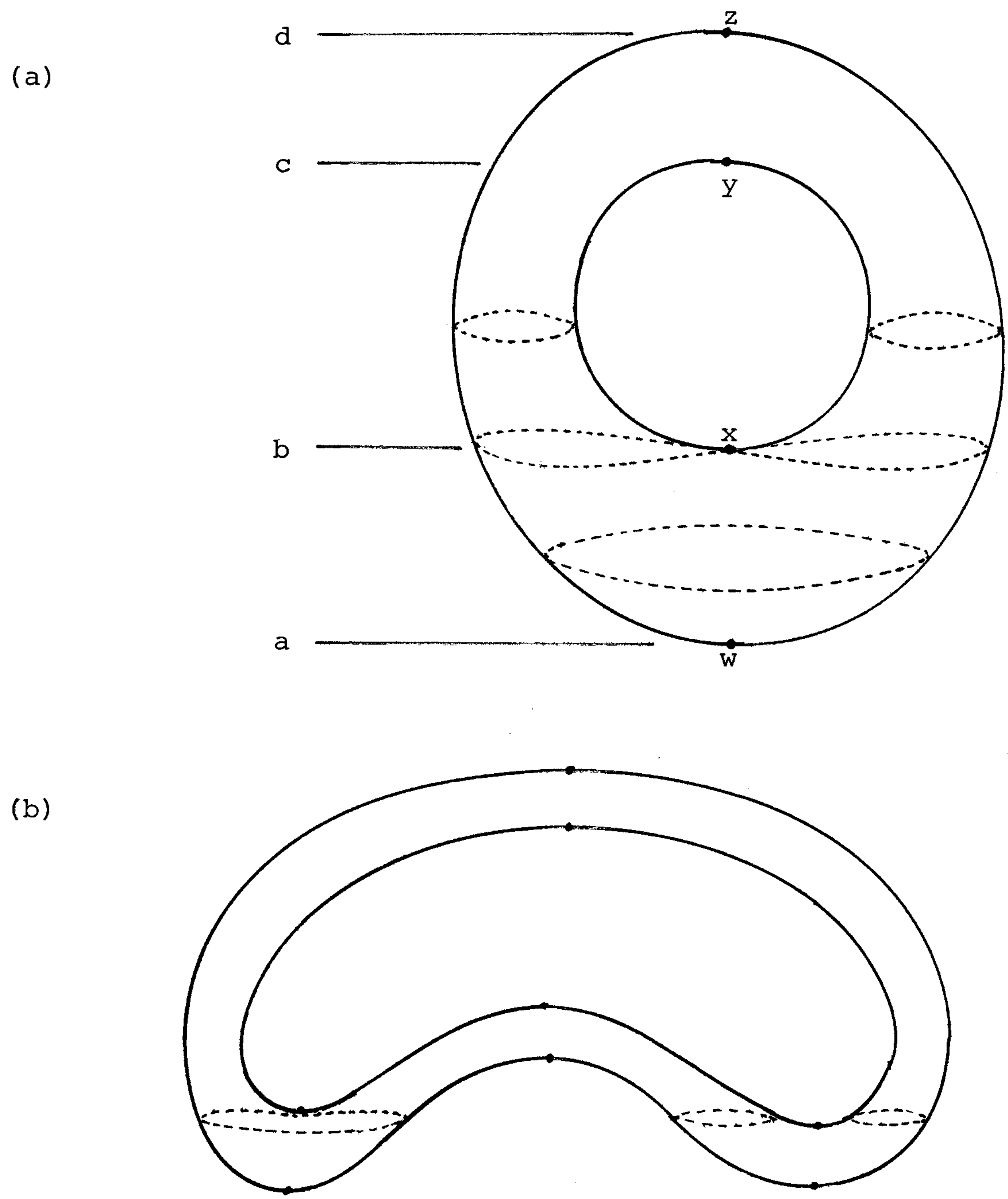


Figure 3.4 (a) The topology of 2-dimensional slices through the solid torus changes at the critical points w,x,y,z . (b) A different choice of geometry can produce additional critical points.

handles or crosscaps.

Sufficiently small deformations of S_0 in W' yield new hypersurfaces with the same topology as S_0. Let $\{S_t : t \in I_\delta\}$ be a continuous family of mutually disjoint hypersurfaces in W' such that, for $t > s$, S_s may be continuously deformed through W' into S_t, and $\partial S_t = S_t \cap \partial W'$ is a 2-sphere in the piece of $\partial W'$ lying to the future of ∂S_s. Each of the 3-manifolds S_t, $t \in I_\delta$, is homeomorphic to S_0 and has the same prime factorization: $M_1 \# \ldots \# M_k \# B^3$. By choosing the induced geometries to be similar to that already chosen for S_0, we arrive at a simple geometric picture of an open region of W' containing S_0 which portrays space-time as an (almost) Euclidean space that is being traversed by k very small and widely separated topological anomalies (Figure 3.3).

If we try to deform S_0 too far through W', however, we will run into topological obstructions because W' does not have the product topology $S_0 \times I$, with I a closed interval. This is best illustrated in three dimensions, rather than four, by looking at 2-dimensional slices through the solid torus (Figure 3.4(a)). Between a and b all of the slices have the topology of a disc, but at b the 2-dimensional section ceases to be a manifold (due to the singular point x), and between b and c each section is the disjoint sum of two discs. The point x, and also w,y,z are called critical points of the torus [35]. Although a different choice of geometry, such as in Figure 3.4(b), could have produced additional critical points (which need be neither isolated nor non-degenerate), only the four isolated and non-degenerate critical points of Figure 3.4(a) are demanded by the topology of the torus.

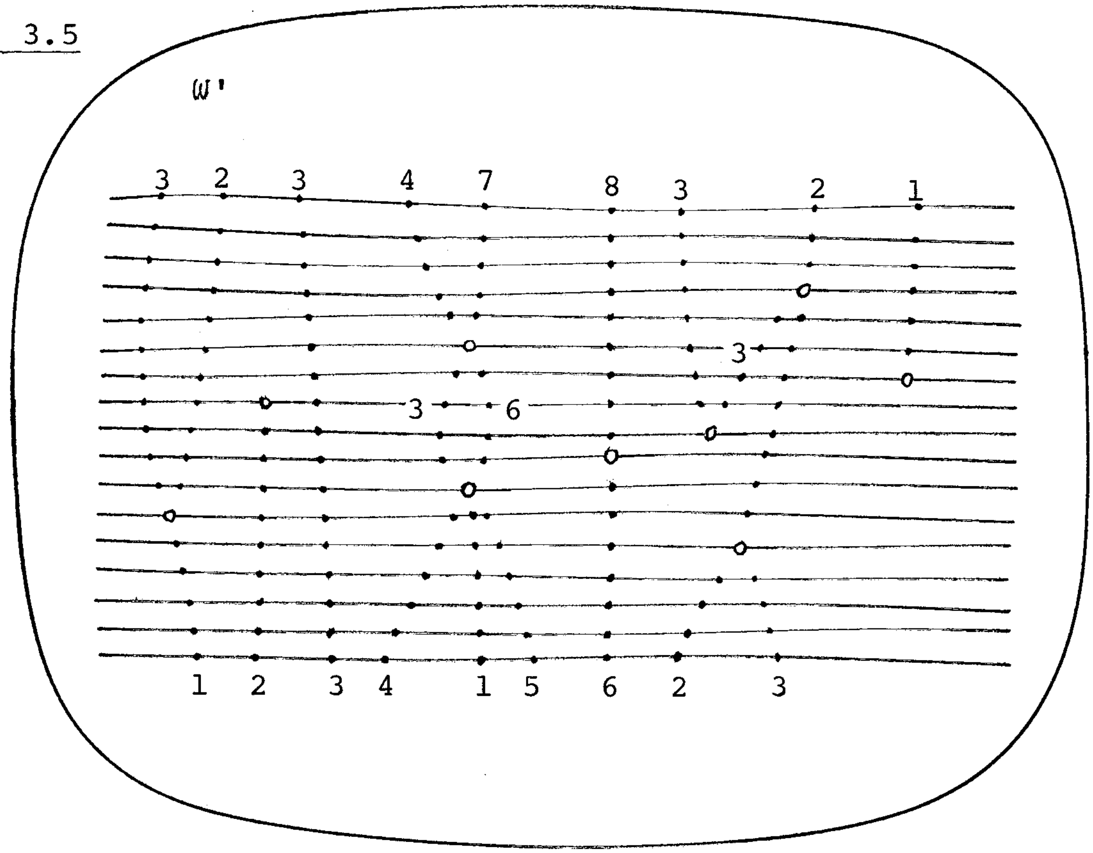
Returning to the topological space-time manifold, $\mathcal{W}'$, we see that we can continue constructing surfaces S_t, for ever increasing t, until at some $t_c > \delta$ a critical point of $\mathcal{W}'$ is reached. Beyond t_c the topology of the hypersurfaces differs from that of S_0. However, the change in topology that takes place at an isolated critical point is small compared to the tremendous complexity of S_0; indicating that most of the prime factors, M_i, that appear in the factorization (3.3) of S_0 do not participate in the topological changes and continue to appear in the prime factorization of S_t for $t > t_c$. Since the topological change takes place at an isolated critical point in $\mathcal{W}'$, those prime factors of $S_{t_c-\varepsilon}$ and $S_{t_c+\varepsilon}$ that do participate must "meet" at the critical point. Adapting to this situation the specialized geometry introduced above leads to the geometric representation of $\mathcal{W}'$ shown in Figure 3.5(a), in which distinct prime 3-manifolds are labelled by distinct integers.

It should be clear, from the above analysis, that all of the topological complexities of $\mathcal{W}'$ are now represented by the (labelled) graph-like structure shown in Figure 3.5(b). Each line corresponds to a prime, compact 3-manifold without boundary;

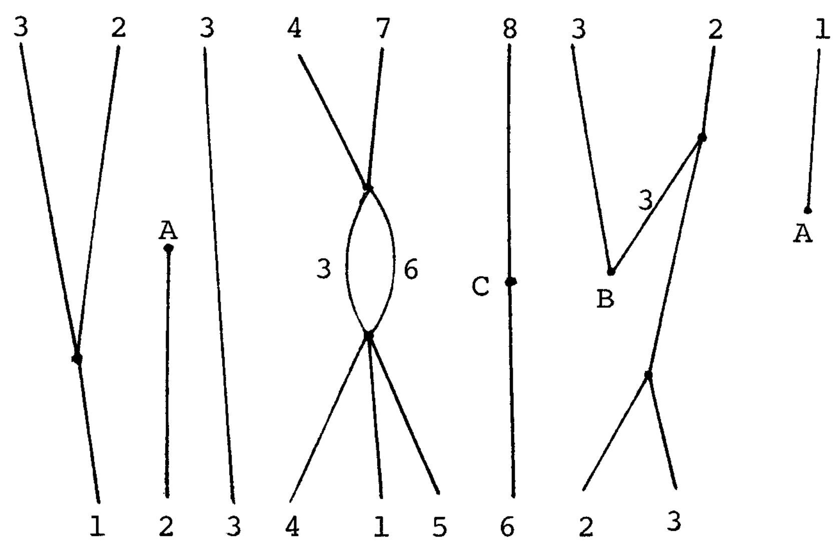
Figure 3.5 (a) Different hypersurfaces of $\mathcal{W}$ may have different topologies. Changes in the hypersurface topology take place at isolated critical points (small circles), with only a small number of prime 3-manifolds, M_i , meeting at each critical point. (b) Stripping the inessential details from the geometric picture, (a), leaves a simple graphical representation of $\mathcal{W}'$.

Figure 3.5

(a)



(b)



and the integer(s) associated with it identify (through some as yet unknown classification scheme) which element of the infinite collection of distinct prime 3-manifolds is being considered. The vertices, which correspond to critical points of W' , need no labels because they are completely characterized by the labelled lines that emanate from them. Critical points such as B are non-essential and can be eliminated by a more careful choice of the hypersurfaces, S_t . However, the remaining critical points (or vertices) are demanded by the topology of W' , just as the four critical points of Figure 3.4(a) are demanded by the topology of the torus. The minimal graph, obtained by eliminating all non-essential vertices (such as B), is thus unique up to the operation of flipping external lines from the past to future and vice versa (which corresponds to choosing a different 2-sphere in $\partial W'$ to bound S_0).

The vertices labelled A and C in Figure 3.5(b) are included only because I cannot prove that such vertices do not exist. The possibility of having vertices such as A seems remote, however, and I shall assume from now on that they never occur. If vertices such as C exist, which I also doubt, then they can be eliminated by assigning the same label to all lines that can be joined to each other by vertices like C . Thus, the lines labelled 6 and 8 in Figure 3.5(b) would be assigned the same label, say 6 . The new graph so obtained would be uniquely determined by W' , but it would no longer provide a faithful representation of this region of topological space-time.

By noting the remarkable similarity between the (not necessarily faithful) graphical representation of $\mathcal{W}'$ and the Feynman diagrams of quantum field theory, we can now initiate the transition from topological space-time to geometrical space-time. Think of $\mathcal{W}'$ as not just something that bears a resemblance to a Feynman diagram, but rather think of it as being a Feynman diagram. Think of each line segment in Figure 3.5(b) as a distinct virtual particle; and think of each vertex as an unrenormalized particle interaction vertex.

The integer labels on the lines in Figure 3.5(b) identify which prime 3-manifold is to be associated with each line, but they serve equally well to identify the elementary particles. There is thus a one-to-one correspondence between the distinct, prime, compact 3-manifolds without boundary (excluding $S^2 \times S^1$) and the elementary particles of quantum physics. It follows immediately that there is a countable infinity of distinct elementary particles, and not just a (small) finite number as is most often supposed.

Allowed particle interactions - that is, allowed vertices - are determined in conventional quantum theory by the phenomenological field theory which the Feynman diagrams represent. In the topological space-time, however, the rules which determine what lines may meet at a vertex (critical point) are purely topological in nature. They are imposed by the simple requirement that $\mathcal{W}$ be a 4-dimensional manifold, and they may, in principle, be derived. Unfortunately, this "selection rule" problem seems just as difficult as, and is clearly dependent upon, the classification of prime 3-manifolds.

Assume, nonetheless, that the classification and selection rule problems are solved, yielding a complete array of particles and interactions. The particles are naturally divided, by the orientability of their corresponding prime 3-manifolds, into two classes. One orientability class will yield bosons in geometrical space-time and the other class will yield fermions. Which is which must be decided with the aid of the selection rules. Ultimately, the selection rules must also be called on to identify the particular prime 3-manifold that corresponds to each known elementary particle (electron, photon, etc.).

Turn, at last, to field theory. Abandon $\mathcal{W}'$ and replace it by the topologically trivial manifold, $\mathcal{M}' \simeq B^4$ (with B^4 the closed ball in $\mathbb{R}^4$). Assume, for the time being, that $\mathcal{M}'$ has a globally Minkowskian metric, η ; and construct on $\mathcal{M}'$ a quantum field theory, $\mathcal{Q}$, with fields ψ_i , $i \in \omega$, such that a one-to-one correspondence between the ψ_i's and the prime 3-manifolds (without boundaries) may be found, which places the interaction vertices of the Feynman graphs of $\mathcal{Q}$ in one-to-one correspondence with the allowed vertices of $\mathcal{W}'$. The parameters (masses and coupling constants) of $\mathcal{Q}$ will, of course, be undetermined, but even when this freedom is ignored there may still be several theories, $\mathcal{Q}_1$, $\mathcal{Q}_2$, $\mathcal{Q}_3$,...., which satisfy the above requirements. From this collection of candidate theories choose the one theory, $\mathcal{Q}$, that is completely determined by its Feynman graphs.

Propagators in $\mathcal{Q}$ carry the virtual particles through M' with constant velocities; and interactions of the fields take the simplest possible form that is consistent with the required vertices. Whenever two or more different particles (fields) have exactly equivalent, yet distinct, allowed interactions the associated fields have identical masses and coupling constants in $\mathcal{Q}$ (even though these parameters are not yet known). "Internal" symmetries, such as colour SU(3) , thus arise out of the topology of W in a natural way.

In order to fix the masses and coupling constants, go back now and reconsider the geometry of M' . It was necessary to assume a c-number metric in the first instance because without it the whole quantum theory, $\mathcal{Q}$, would collapse. However, a physical metric, g , cannot be arbitrarily imposed, as η was. Instead, g must arise out of a logical analysis of the topology of W' and, in particular, the graphical representation of W' obtained above. Since all of this topological information has already been exploited in the construction of $\mathcal{Q}$, our only option is to have $\mathcal{Q}$ determine g in some self-consistent way. The correct coupling will give g the simplest possible form; and Møller's proposal,

$$G(\vec{U},\vec{V}) = 8\pi\langle T_{op}(\vec{U},\vec{V})\rangle \quad , \tag{1.1}$$

seems ideally suited for this purpose. All partial derivatives in $\mathcal{Q}$ are now converted to covariant derivatives (minimal coupling) and the state $\rangle$ is determined by the particular topology of $\mathcal{W}$. It is to be assumed, as well, that a unique procedure has been found for eliminating the zero-point energy from T_{op} (cf. Section 1); and that $\mathcal{Q}$ has been renormalized on the background g, leaving only "physical" masses and coupling constants.

The specific geometry obtained from (1.1) depends not only on the state, $\rangle$, but also on the values, m_i, c_α, that are chosen for the masses and coupling constants. To obtain a unique space-time geometry require that

$$\frac{\delta g}{\delta m_i} = 0 \quad \text{and} \quad \frac{\delta g}{\delta c_\alpha} = 0 \quad . \tag{3.4}$$

Solve these equations to find the unique set of physical masses and coupling constants and hence the unique geometrical space-time, $\{\mathcal{M}',g,\mathcal{Q}\}$, corresponding to $\mathcal{W}'$.

I have moved quickly through this formal construction of geometrical field theory, and in doing so I have passed over many very real problems, both technical and philosophical. Most of these are due to the non-linearity of the semi-classical field theory. As mentioned above, some progress has been made on the factor ordering (which must be solved to make $\langle T_{op} \rangle$ finite); but the problem of renormalizing interacting quantum fields on a self-consistently determined background geometry remains untouched. Also required by the non-linearity of (1.1) are significant changes, of both an interpretive and a mathematical nature, in the foundations of quantum theory.

A new problem, more directly related to my topological picture of the world, arises from equations (3.4). These equations may be thought of as a generalized bootstrap, with the mutual interactions of the elementary particles determining all the masses and coupling constants. However, because the metric is a tensor rather than a scalar field it may be impossible to satisfy all of the equations (3.4); and even if g had only one degree of freedom (such as the Newtonian gravitational potential) the solutions, m_i , c_α , would, in general, not be constants but rather functions of the space-time coordinates. In this latter case, one could reasonably expect counter-terms from the renormalization to suppress fluctuations of the parameters, with any remanent variations having a length scale much larger than the radius, $g/\nabla g$, characteristic of changes in the local geometry. No great problems will arise in the quantum theory as long as the geometric radius remains large when compared with the Compton wavelengths of the particles being considered; but if the masses and coupling constants change too rapidly, or if it becomes necessary to consider more than one component of g , then the entire field theory will collapse. It is worth noting that in our local region of the universe the metric is, in fact, adequately specified by the one component, g_{00}, and that fluctuations of g about η have characteristic lengths of about 10^{15}m (at the surface of the sun) or more. Thus, only in the neighbourhood of a gravitational shock wave or some other equally catastrophic gravitational event should the masses and coupling constants be expected to change noticeably and, perhaps, become ill defined.

4. Summary

Although the technical hurdles still to be cleared are immense, the rough outline presented above shows that the idea of extracting field physics from the topology of a 4-dimensional world is not so crazy as to be impossible. By constructing the graphical representation of $\mathcal{W}'$ (and, by extension, $\mathcal{W}$) we are actually led directly to quantum field theory. The intricate web of virtual particles and interactions is reduced systematically, through the renormalization procedure, to leave a "physical" graph that represents physical particles propagating and interacting in a geometry of their own creation. This geometry depends on the particular topology of $\mathcal{W}$ and is a c-number field - gravity is not quantized. Uniqueness of the fields is assured by choosing the masses and coupling constants such that infinitesimal variations of these parameters leave the geometry unchanged.

The splitting of space and time is essential in the construction of a semi-local (neither global nor local) representation of the topology of W . So also is the dimension, three, of space, because a non-trivial connected sum decomposition is possible only in three-dimensions [37]. (A similar decomposition is possible in 2-dimensions, but all factors are identical.)

In the end, though, the most remarkable and compelling feature of this topological world-view is its simplicity. Providing the theory is born out by further analysis, all perceived physical phenomena, including gravity, quantum effects, and the detailed behaviour of the elementary particles, will be understood as characteristics of an unconstrained 4-dimensional topological manifold.

CHAPTER 5

CONCLUSION

I have investigated the structure of space-time from two very different points of view. Adopting a traditional world-view in Chapters 2 and 3, and restricting attention to the classical domain, I have developed a general formalism for describing, in a coordinate-free way, the evolution of the metrical geometry of the universe. This evolution can be described in terms of tensor fields intrinsic to the space-like hypersurface only if the tensor field, g, from which the hypersurface metric, ${}^3g = \Pi g$, is constructed, is a (pseudo-)Riemannian metric on space-time. The invariantly defined, normalized time derivatives of ${}^3g$ are then the extrinsic curvature, K; the hypersurface projection of the space-time Ricci tensor, ΠS; and the hypersurface projections of the invariant derivatives of S, to all orders, along the unit normals to the hypersurface: $\Pi(\nabla_{\vec{n}}S)$, $\Pi(\nabla_{\vec{n}}\nabla_{\vec{n}}S - \nabla_{\nabla_{\vec{n}}\vec{n}}S)$, etc..

I have shown how to construct dynamical theories of the evolution of ${}^3g$ by supposing that ${}^3g$ and its time derivatives up to some finite order, m, are essential initial data fields, but that the $(m+1)$th time derivative of ${}^3g$ is some explicit functional of the lower derivatives and of additional initial data fields that characterize the distribution of matter in space-time. The integrability conditions are constraints on the initial data which allow one to consistently and unambiguously construct all time derivatives of ^3g , of order greater than $(m+1)$, in terms of the initial data. In the simplest case, $m = 1$, it is only possible to construct one consistent set of dynamical and constraint equations for the metric, and these are the Einstein gravitational field equations. Higher order gravitational theories (with $m>1$) cannot be ruled out, however, nor can any restrictions be placed on the functional form of the stress-energy tensor (when $m = 1$), aside from the obvious condition that it must be a symmetric tensor with vanishing covariant divergence.

With the hope of obtaining a unified theory of gravitation and quantum phenomena, I have proposed, in Chapter 4, that the objective world underlying all of our perceptions is a 4-dimensional topological manifold, $\mathcal{W}$, with no physically significant field structure, but instead an unconstrained and extremely complex global topology. Aided by the connected sum decomposition theorem for 3-manifolds, I have demonstrated that $\mathcal{W}$ may be uniquely represented by a labelled graph which has properties remarkably similar to those of the Feynman graphs of QFT. The lines of this graph correspond to prime 3-manifolds (without boundary) and the vertices correspond to isolated, non-degenerate critical points of $\mathcal{W}$. By exploiting this similarity with Feynman graphs, I have been able to show how the space-time of our perceptions, with its geometry and quantum fields, might arise as a replacement manifold for $\mathcal{W}$, and how the phenomenology and laws of field physics might emerge from the unconstrained topological structure of $\mathcal{W}$. Neither geometry nor quantum fields are fundamental - instead each arises to give meaning to the other, with geometry providing a substrate for the quantum fields and the stress-energy of the quantum fields (corresponding to the unique world, $\mathcal{W}$) determining the space-time geometry.

In both of these world-views - the geometrical and the topological - the splitting of space and time has played a central role. With each passing instant a new universe, a new face of the world, is revealed to us. To a very great extent the changes that take place in the appearance of the universe are predictable, and the rules used to make predictions are the laws of physics. Thus, in the geometrical world-view, the evolution of the geometry of space is predicted with the use of the Einstein field equations; while, in the topological world-view, changes in the topology of the 3-dimensional slices of $\mathcal{W}$, although random, are subject to the restrictions imposed by the topological selection rules. The geometrical replacement manifold for $\mathcal{W}$ bridges the gap between pure geometry and pure topology, capturing the topological selection rules in the quantum field theory, $\mathcal{Q}$, and resurrecting GR in the form of Møller's semi-classical theory of gravity.

ACKNOWLEDGEMENT

I would like to thank my supervisor, Professor J.W. Moffat, for the patient guidance he has provided, and for the many hours of stimulating and informative discussions that we have had, during the past six years. I am also grateful to Dr. Robert Coleman, who has helped me to develop an understanding of the techniques and concepts of modern differential geometry, and with whom I have also had many enjoyable discussions.

Financial support for my graduate education has been provided by the National Research Council, the Province of Ontario, the University of Toronto, and the Department of Physics. The most important support, however, has been the wonderful encouragement provided by my wife, Melanie, and my parents.

APPENDIX I

DIFFERENTIAL MANIFOLDS

Topology, Geometry, and Submanifolds

The purpose of this appendix is to provide an introduction to the mathematical structures used in the thesis, and also to establish notational conventions. While I recognize that most readers will be totally familiar with the items discussed, I have attempted to make the presentation fairly complete for the benefit of those who do not have a background in topology or differential geometry. The emphasis is on definitions and an understanding of the basic concepts. Most results are stated without proof.

Numerous comprehensible references exist for this material, but I shall mention here only those monographs that I have found particularly useful. 'General Topology' by Lipschutz [38] is a good primer for the basics of point set topology, leading up to but not including the definition of a manifold. Munkres [39] provides clear definitions of manifold, differential structure, and differential manifold, but does not indicate that it is the field structure, rather than the topology, of the real numbers which gives rise to the possibility of defining derivatives. The fundamental definition of derivative is contained in Porteous [40] along with the basics of algebra, topology and a great wealth of other information that should be of interest to many physicists, but which cannot be treated here. Guillemin and Pollack, in their well illustrated text 'Differential Topology'[33] , explore many aspects of that vaguely defined field. They develop both differential and integral calculus on manifolds and show how these relate to the global structure of differential manifolds. The properties of differentiable maps from one manifold into another are shown to depend significantly on the global topologies of the two manifolds.

Differential geometry, considered in its broadest sense as the mathematics of differentiable fields on manifolds, is perhaps the branch of modern mathematics most familiar to physicists. An excellent classical text is 'Ricci Calculus' by Schouten [21], but I much prefer the more modern treatment and notation of Kobayashi and Nomizu [41]. Throughout this thesis I will employ a coordinate free notation which is similar to that used in reference [41], however the elegant fibre bundle picture which they develop will not be used here because it would be overkill for the simple geometries to be considered.

1. Topology

Topology is the study of sets and their subsets. Let X be a non-empty set and let Sub X be the class consisting of all subsets of X . A subset $\mathcal{T}$ of Sub X is a topology on X iff $\mathcal{T}$ satisfies

(i) $\emptyset, X \in \mathcal{T}$;

(ii) for all $A, B \in \mathcal{T}$, $A \cap B \in \mathcal{T}$;

(iii) for all $A_i \in \mathcal{S} \subset \mathcal{T}$, $\bigcup_i A_i \in \mathcal{T}$,

where $\emptyset$ is the null set. The elements of $\mathcal{T}$ are called the open sets of the topology. A set X , together with a topology $\mathcal{T}$ on X , is called a topological space, $(X, \mathcal{T})$. Normally this will be denoted by X alone, with the topology $\mathcal{T}$ assumed to be known. It is important to recognize that the open subsets of X may generally be chosen in more than one way, each different choice giving rise to a different topology on X .

A base for the topology $\mathcal{T}$ on X is a subset $\mathcal{B} \subset \mathcal{T}$ such that every open set $A \in \mathcal{T}$ is the union of members of $\mathcal{B}$. An open cover or cover for a topological space $(X,\mathcal{T})$ is a subset $\mathcal{S}$ of $\mathcal{T}$ such that $\bigcup\mathcal{S} = X$. Every base is a cover, but not every cover is a base. If for each cover $\mathcal{S}$ for X a finite subset $\mathcal{S}'$ of $\mathcal{S}$ covers X then X is said to be compact.

A topological space X is a Hausdorff space iff for each pair of distinct points $a,b \in X$ there exist open sets $A,B \in \mathcal{T}$ such that $a \in A$, $b \in B$, and $A \cap B = \emptyset$.

We are often concerned with maps from one topological space to another. Let $(X,\mathcal{T})$ and $(X',\mathcal{T}')$ be topological spaces. A function $f:X \rightarrow X'$ is said to be continuous iff the inverse image of every open set of X' is an open set of X, that is, iff

$$f^{-1}[A] \in \mathcal{T} \quad \text{for all} \quad A \in \mathcal{T}' .$$

Two topological spaces X and X' are called homeomorphic or topologically equivalent if there exists a bijective map $f:X \rightarrow X'$ such that both f and f^{-1} are continuous. The map f is called a homeomorphism.

Let $g:W \rightarrow X$ be a continuous map with domain a subset of the topological space W and let $a \in W\backslash\text{dom } g$ (i.e. a is an element of the complement of dom g in W). Then g has a limit b at a if there exists a continuous map $f:W \rightarrow X$ such that $f(a) = b$ and $f(w) = g(w)$ for all $w \in \text{dom } g$. If dom g is a proper subset of W, a is an element of the closure of dom g, and X is a Hausdorff space, then b is unique. Porteous notes that this is one of the most important features of a

Hausdorff space [40].

If $Y \subset X$ and $T_Y = \{A_Y : A_Y = A \cap Y$ for some $A \in T\}$ then (Y, T_Y) is a topological <u>subspace</u> of (X, T) with the <u>induced</u> topology. Any subspace of a Hausdorff space is a Hausdorff space.

Let X and Y be topological spaces and let $W = X \times Y$ be the cartesian product of the sets X and Y . The <u>product topology</u> induced on W from X and Y consists of all those subsets of W that can be constructed as the union of sets of the form $A \times B$ where A is open in X and B is open in Y . Unless specified otherwise, the product $X \times Y$ of two topological spaces will be assumed to have the product topology.

Until now I have avoided reference to numbers. However, the number systems with which we are so familiar play an important role in topology. Porteous [40] starts with the null set and builds up the <u>natural</u> numbers through a constructive process. The non-existence of a largest natural number is the

Archimedian Order Axiom: The set $\omega = \{0,1,2,...\}$ of natural numbers is not bounded from above.

Addition, multiplication, and exponentiation are defined in a set theoretic fashion. Further constructions yield the <u>integers</u> Z and the <u>rational</u> numbers Q . The <u>real</u> numbers $\mathbb{R}$ are then defined to be the elements of an ordered field with the usual operations of addition and multiplication, containing Q as an ordered subfield, such that

(Least Upper Bound Axiom): If A is a subset of $\mathbb{R}$ bounded from above, then A has a least upper bound.

This is equivalent to the statement that the real numbers are complete, that is, that every Cauchy sequence of real numbers converges to a point in $\mathbb{R}$.

The topology of $\mathbb{R}$ is defined with the use of the open intervals $S = \{x : a < x < b \; ; \; x,a,b \in \mathbb{R}\}$. Let $A \subset \mathbb{R}$. A point $p \in A$ is an interior point of A iff p belongs to some open interval S_p which is contained in A . The set A is open iff each of its points is an interior point [38]. Note that the topology of $\mathbb{R}$ does not depend explicitly on the operations of addition or multiplication, but only on the well ordered property of the real numbers. As is usual, $\mathbb{R}^n$ will be used to denote the topological product $\mathbb{R}\times\mathbb{R}\times\ldots\times\mathbb{R}$ with n factors.

A point $x \in \mathbb{R}^n$ may be represented by the ordered n-tuple of its components $(x^1, \ldots ,x^n)$. If we exploit the field structure of the real numbers then we may use these components to define a norm on $\mathbb{R}^n$:

$$|x| = \sqrt{(x^1)^2 + \ldots + (x^n)^2} \quad . \tag{1.1}$$

For any $\delta > 0$ the sets $\{x \in \mathbb{R}^n : |x| < \delta\}$, $\{x \in \mathbb{R}^n : |x| \leq \delta\}$, and $\{x \in \mathbb{R}^n : |x| = \delta\}$ are called respectively the open ball, the closed ball, and the sphere in $\mathbb{R}^n$, centred on the origin, with radius δ . The origin or additive identity of $\mathbb{R}^n$ need not play such a fundamental role, however. Instead we can make use of the metric or distance function :

$$d(x,y) = |x - y| \quad , \tag{1.2}$$

which specifies the euclidean distance between any two points x and y in $\mathbb{R}^n$. An open ball with centre x and radius δ is then the set $B(x,\delta) = \{y \in \mathbb{R}^n: d(x,y)<\delta\}$. The class of all such open balls provides a base for a <u>metric topology</u> T_d on the set $\mathbb{R}^n$. For finite dimension n this metric topology coincides with the usual topology on $\mathbb{R}^n$. When no particular point of $\mathbb{R}^n$ is singled out as the origin, the space $(\mathbb{R}^n, T_d)$ is called an <u>affine space</u> and denoted by A^n .

One more word must be added to our mathematical vocabulary before we can define manifold. A set X is said to be <u>countable</u> iff there exists an injective map from X into ω , that is, iff the natural numbers may be used to uniquely label the elements of X .

A <u>topological manifold</u> M is a Hausdorff space with a countable basis, satisfying the following condition: There is a number $n \in \omega$ such that each point of M has a neighbourhood homeomorphic with an open set of $\mathbb{R}^n$. The number n is the <u>dimension</u> of M . If A is an open proper subset of M then $M \backslash A$ is a <u>manifold with boundary</u>. The set $\bar{A} \backslash A$, where $\bar{A}$ is the closure of A in M , is the common boundary of $\bar{A}$ and $M \backslash A$ and is an (n-1)-dimensional manifold. The notation ∂M is commonly used to denote the boundary of a manifold-with-boundary M . For any M , $\partial\partial M = \emptyset$.

The spaces A and $\partial\bar{A}$ are examples respectively of n and (n-1)-dimensional submanifolds of M . An m-dimensional <u>submanifold</u> M' of a manifold M , with $0 \leq m \leq n$, is a topological subspace of M which (with the induced topology) is an m-dimensional manifold. A submanifold of dimension zero has the discrete topology, and consists of isolated points in M .

The product space $M \times M'$ of two manifolds M and M' is a manifold with $\dim(M \times M') = \dim(M) + \dim(M')$.

It is possible for a manifold to be made up of several disjoint pieces. A topological space (manifold) is <u>connected</u> iff it is not the union of two non-empty disjoint open sets. Unless specified otherwise a manifold will be assumed to be connected.

Every connected manifold M is metrizable, that is, M admits a distance function $d: M \times M \to \mathbb{R}$ which is <u>compatible</u> with the topology of M and which satisfies, for all $x,y,z \in M$:

$$\left.\begin{array}{ll} \text{(i)} & d(x,y) \geq 0 \quad \text{and} \quad d(x,x) = 0 \quad ; \\ \text{(ii)} & d(x,y) = d(y,x) \quad ; \\ \text{(iii)} & d(x,z) \leq d(x,y) + d(y,z) \quad ; \\ \text{(iv)} & \text{If} \quad x \neq y \quad \text{then} \quad d(x,y) > 0 \quad . \end{array}\right\} \quad (1.3)$$

The open balls defined with the use of d provide a base for the topology of M . However the topology of M does not uniquely determine d , nor is every distance function necessarily compatible with the topology of M . This will be elaborated in the next section.

Let M be an n-dimensional manifold with $n \geq 1$. An <u>open n-cell</u> in M is an n-dimensional submanifold of M which is homeomorphic to an open ball in $\mathbb{R}^n$. A closed n-cell is a subspace $S \subset M$ which is homeomorphic to a closed ball in $\mathbb{R}^n$. The boundary ∂S of a closed n-cell S is an (n-1)-sphere. In general, an (n-1)-sphere S^{n-1} is a manifold homeomorphic with a sphere in $\mathbb{R}^n$.

All manifolds of a given dimension are locally the same: if M is an n-dimensional manifold, then every point $x \in M$ is contained in the interior of some closed n-cell in M . When considered in their entirety, however, two manifolds of the same dimension can be very different. It is thus the global structure of M which serves to differentiate it from other n-dimensional manifolds. This is best illustrated with specific examples in one and two dimensions:

(i) There are only two distinct (connected) 1-dimensional manifolds (without boundary). These are the line $A^1 = \mathbb{R}$ and the circle or 1-sphere S^1 (Figure 1.1). They are distinguished by the fact that S^1 is compact while A^1 is not. If we choose a point $x \in S^1$ and a closed 1-cell $C \subseteq S^1$ which contains x in its interior, then $S^1 \setminus C$ is an open 1-cell which is homeomorphic with A^1 . The line can thus be considered as a 1-sphere which has had a closed 1-cell removed or "cut out".

(ii) There is a countable infinity of distinct 2-manifolds. They fall into two natural classes, orientable and non-orientable, and may all be constructed from the 2-sphere S^2 by a process of "cutting and pasting" [42]. Although it is technically a purely

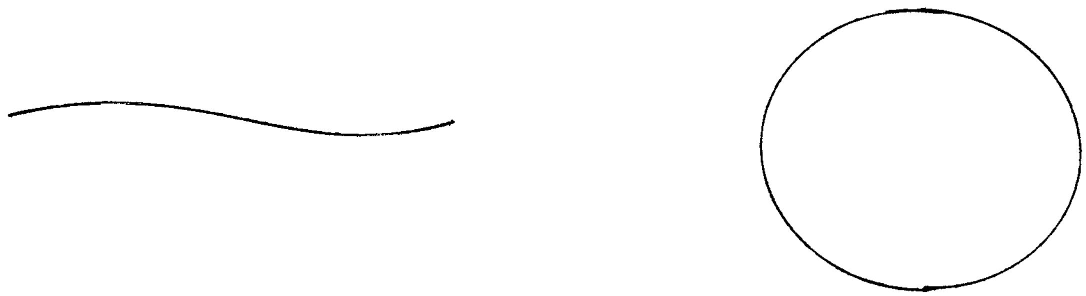


Figure 1.1 The line and circle are, up to homeomorphism, the only 1-dimensional manifolds.

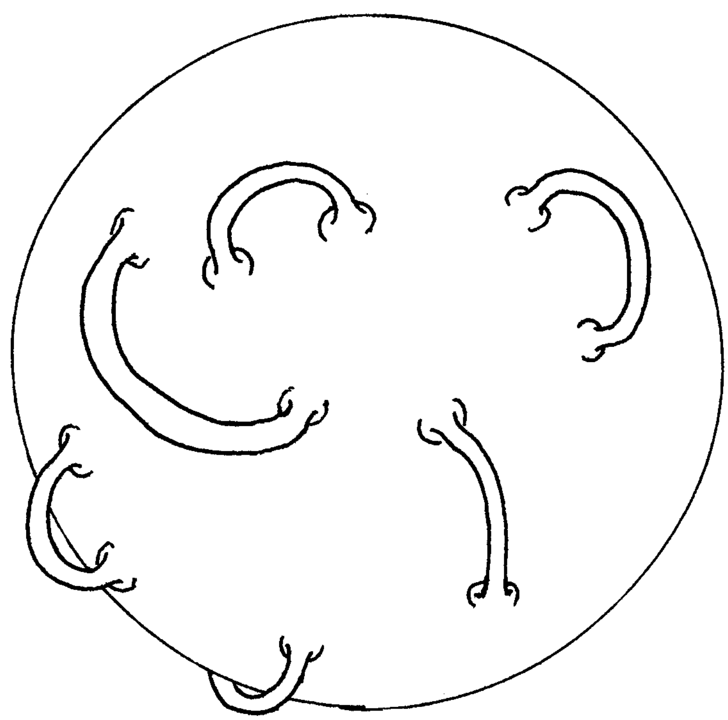
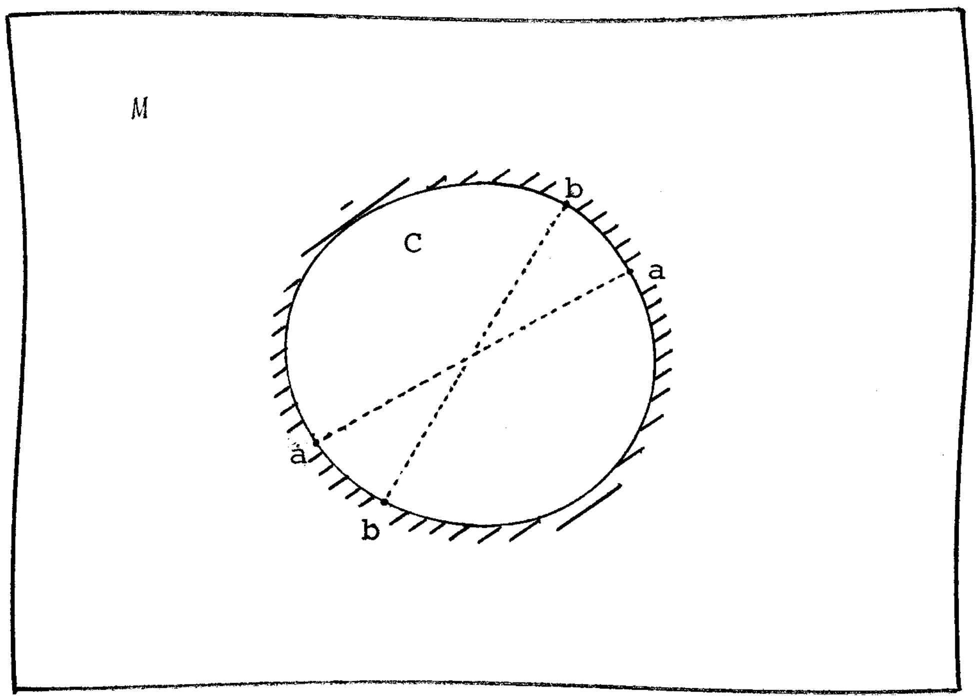


Figure 1.2 A sphere with (at least) six handles.



<u>Figure 1.3</u> A crosscap is formed by removing from M the interior of the 2-cell, C , and then identifying opposite points on the resulting boundary.

topological concept we shall delay the definition of orientability until the next section, relying for the time being on the intuitive notion that an orientable surface has two sides (e.g. the inside and outside of the unit sphere in $\mathbb{R}^3$) while a non-orientable surface has only one side. Examples of compact orientable 2-manifolds are the sphere S^2 , the _torus_ $S^1 \times S^1$ which may also be considered as a sphere with one handle, and more generally a surface of genus h or _sphere with h handles_ (Figure 1.2). The simplest compact non-orientable 2-manifold is a sphere with one crosscap. Being the non-orientable analogue of a handle, a _crosscap_ is constructed by removing from a 2-dimensional manifold M an open 2-cell C and then identifying (with the use of an orientation reversing homeomorphism $f:\partial\bar{C} \to \partial\bar{C}$) "opposite" points of the resulting circular boundary (Figure 1.3). The non-orientable analogue of a sphere with h handles is a _sphere with q crosscaps_. A sphere with 2 crosscaps is topologically equivalent to the well known Klein bottle. Non-compact 2-manifolds may be constructed from a compact 2-manifold M by removing from M any number r of closed 2-cells. The resulting manifold is then said to have r _contours_. The general result which emerges is that any 2-dimensional manifold (without boundary) is characterized topologically by its orientability class, its contour number r , and its number of handles or crosscaps, h or q . The Moebius band, for example, is a non-orientable surface with one contour and one crosscap.

For manifolds of dimension greater than two the problem of analyzing the global structure becomes much more complex. The vast field of algebraic topology [42],[43] provides general procedures for describing some aspects of the global structure of a given manifold M, but no scheme is known for uniquely identifying each distinct manifold. It is thought, in fact, that for $n \geq 4$ it is impossible to find a classification scheme which will uniquely label each and every n-dimensional manifold (up to homeomorphism) [32]. The 3-dimensional problem, on the other hand, may be solvable. Much progress has been made toward a partial solution and I believe that some of the results may have a direct and profound application in physics. This is discussed in detail in Chapter 4.

2. Differential Manifolds

We return now to the local properties of manifolds. Let U be an open set of an n-dimensional manifold M and let $\phi : U \to \mathbb{R}^n$ be a homeomorphism of U onto an open set in $\mathbb{R}^n$. The pair (U,ϕ) is called a chart on M and the n component functions of ϕ determine a local coordinate system in U. An atlas of M is a family of charts (U_i, ϕ_i) on M such that the open sets U_i cover M.

A mapping f of an open set of $\mathbb{R}^n$ into $\mathbb{R}^m$ is said to be of class C^r, $r \in \omega$, if its m component functions $f^1, \ldots, f^m$ are r times continuously differentiable. If f is real analytic then it is said to be of class C^{ω}. By C^0 we mean that f is continuous.

A <u>differential structure</u> $\mathcal{D}$ of class C^r on an n-manifold M is an atlas (U_i,ϕ_i) on M such that

(i) If (U_i,ϕ_i) and (U_j,ϕ_j) belong to $\mathcal{D}$, then

$$\phi_i\phi_j^{-1} : \phi_j(U_i \cap U_j) \to \mathbb{R}^n \tag{2.1}$$

is differentiable of class C^r; and

(ii) The atlas $\mathcal{D}$ is maximal with respect to property (i); i.e., if any chart not in $\mathcal{D}$ is adjoined to $\mathcal{D}$, then (i) fails.

The manifold M together with the differential structure $\mathcal{D}$ is called an n-dimensional <u>differential manifold</u> of class C^r. A differential manifold of class C^{∞} is called a <u>smooth</u> manifold. Although it does not appear explicitly, we have exploited for the first time the full field structure of the real numbers in the definition of a differential structure.

Let M be of class C^r, $r \geq 1$, and let $x \in U_i \cap U_j$. Denote by $a_{ij}(x)$ the $n \times n$-matrix of first partial derivatives of the functions (2.1) evaluated at $\phi_j(x)$, i.e. the Jacobian matrix of (2.1). The atlas $\mathcal{D}$ is called <u>oriented</u> if the determinant of $a_{ij}(x)$ is positive for all i,j and $x \in U_i \cap U_j$. If $\mathcal{D}$, $\mathcal{D}'$ are two distinct oriented atlases of M then the Jacobian matrices of $\phi_i'\phi_j^{-1}$ have determinants which, for all i,j and $x \in U_i' \cap U_j$, are either always positive or always negative. Accordingly, the orientations of $\mathcal{D}$ and $\mathcal{D}'$ are said to be either the same or opposite. The class of all atlases which have the same orientation as $\mathcal{D}$ is called an orientation of M. If M admits an oriented atlas then it is orientable, and has exactly two orientations. If no oriented atlas exists for M then M is non-orientable.

Given two differential manifolds M and M' of dimensions n and n' and of class C^r, a map $f:M \to M'$ is said to be of class C^k, $k \leq r$, if for all i,j the map

$$\phi_i' f \phi_j^{-1} : \phi_j(U_j \cap f^{-1}(U_i')) \to \mathbb{R}^{n'} \tag{2.2}$$

is of class C^k. The rank of f at $x \in U_j$ is defined to be the rank of the Jacobian matrix of the map (2.2) at $\phi_j(x)$. If rank $f = n$ at each point $x \in M$, f is said to be an immersion. The map f is proper if the preimage of every compact set in M' is compact in M. An immersion that is injective and proper is called an embedding. If f is a homeomorphism of M and M' and an immersion then it is called a diffeomorphism. In this case M and M' are said to be diffeomorphic. An embedding $f:M \to M'$ maps M diffeomorphically onto a submanifold of M'.

Not every manifold admits a differential structure [44], nor do two different differential structures on the same manifold always give rise to diffeomorphic differential manifolds. However, if $\dim M \leq 3$ or if M is homeomorphic to $\mathbb{R}^n$, $n \neq 4$, then M admits a differential structure [32] and all differential manifolds $(M,\mathcal{D})$ are diffeomorphic [39].

Two important theorems [39],[45] allow us to restrict our attention to differential manifolds of class C^∞ and to smooth (C^∞) maps between C^∞ manifolds. The first states that if M and M' are C^∞ manifolds and $f:M \to M'$ is a C^1 immersion, embedding, or diffeomorphism, then f may be approximated by a C^∞ immersion, embedding, or diffeomorphism, respectively. The second theorem states that every differential structure of class C^1 on a manifold M contains a C^∞ structure. From now on, unless indicated otherwise by the context, the term manifold will be taken to mean smooth differential manifold, and all maps between manifolds will be assumed to be smooth.

Differential manifolds are of special interest to physicists because they serve as the substrate for all of the geometrical structures with which we deal. The simplest such structure is a function or <u>scalar</u> <u>field</u> $f:M \to R$ on the manifold M. We shall denote the algebra of all such (smooth) functions by $\mathcal{F}(M)$. A <u>differentiable</u> <u>curve</u>, or simply <u>curve</u>, in M is a mapping of a closed interval $[a,b] \subset R$ into M. If $x(t)$, $t \in [a,b]$ is a curve in M, then the <u>vector</u> tangent to $x(t)$ at $p=x(t_0)$ is the derivative operator $\vec{X}_p$ defined by

$$\vec{X}_p f = (df(x(t))/dt)_{t_0} \qquad \text{for all} \quad f \in \mathcal{F}(M) \ . \tag{2.3}$$

The set of all vectors that can be constructed at p is, in a natural way, an n-dimensional vector space (where $n = \dim M$) called the <u>tangent</u> <u>space</u> to M at p, and is denoted by $T_p(M)$.

If $(U,\phi) \in \mathcal{D}$ is a chart on M such that $p \in U$, and the quantities $u^i = \phi^i(x)$, $i=1,\ldots,n$, are the local coordinates of a point x, then the curve $x(t)$ has the coordinate representation $x^i(t) = \phi^i(x(t))$. Within this coordinate picture we find that

$$(df(x(t))/dt)_{t_0} = \sum_i (\partial f/\partial u^i)_p \cdot (dx^i(t)/dt)_{t_0} \quad , \tag{2.4}$$

so that the vector $\vec{X}_p$ may be represented by the differential operator

$$\vec{X}_p = \sum_i (dx^i(t)/dt)_{t_0} (\partial \ /\partial u^i)_p \quad . \tag{2.5}$$

The partial derivative operators $(\partial \ /\partial u^i)_p$, or simply $(\partial_i)_p$, are linearly independent vectors at p and constitute a basis, called a <u>coordinate basis</u>, for $T_p(M)$. The numbers

$$X_p^i = (dx^i(t)/dt)_{t_0} \tag{2.6}$$

are called the <u>components</u> of the vector $\vec{X}_p$ in the coordinate basis. Although it is always possible to represent a vector in this fashion it is certainly not necessary, nor even desirable in many situations. In that which follows, a <u>coordinate free</u> formalism will be used almost exclusively so that the basic geometrical concepts being investigated will remain in the fore.

A <u>vector field</u> $\vec{X}$ on M is an assignment of a vector $\vec{X}_p$ to each point p of M. Acting pointwise on a function $f \in \mathcal{F}(M)$, $\vec{X}$ generates a new function $\vec{X}f$ defined by

$$(\vec{X}f)(p) = \vec{X}_p f \quad . \tag{2.7}$$

If $\vec{X}f$ is a smooth function on M for each $f \in \mathcal{F}(M)$ then $\vec{X}$ is a smooth vector field. Just as we are restricting our attention to smooth maps we shall consider only smooth vector fields, and we shall use the terms vector field and smooth vector field interchangeably. The (infinite dimensional) space of all vector fields on M will be denoted by $\mathcal{T}(M)$.

Let $T_p^*(M)$ denote the vector space dual to $T_p(M)$, that is the space of all linear maps $\omega_p : T_p(M) \to \mathbb{R}$. A 1-form $\underline{\omega}$ on M is an assignment to each point $p \in M$ of an element of $T_p^*(M)$. If $\vec{X} \in \mathcal{T}(M)$ and $\underline{\omega}$ is a 1-form on M , then $\underline{\omega}(\vec{X})$ is the function defined by

$$\underline{\omega}(\vec{X})(p) = \omega_p(\vec{X}_p) \quad , \quad p \in M \quad . \tag{2.8}$$

If $\underline{\omega}(\vec{X}) \in \mathcal{F}(M)$ for all $\vec{X} \in \mathcal{T}(M)$, then $\underline{\omega}$ is a C^∞ 1-form (differential form of degree 1). As usual, we shall consider only smooth 1-forms, denoting the space of all C^∞ 1-forms by $\mathcal{T}_1(M)$. The 1-form $\underline{du}^i$, $i \in \{1,\ldots,n\}$, dual to the coordinate basis vector field $\vec{\partial}_i$ in the neighbourhood $U \subset M$ is defined by

$$\underline{du}^i(\vec{\partial}_j) = \delta^i_j \quad , \text{ for all } j \in \{1,\ldots,n\} \tag{2.9}$$

where δ^i_j is the Kronecker δ function. Accordingly, $\underline{du}^1_p, \ldots ,$ $\underline{du}^n_p$ are linearly independent in $T_p^*(M)$ and are called the coordinate (basis) 1-forms. In the neighbourhood U , any $\underline{\omega} \in \mathcal{T}_1(M)$ may always be written in the form

$$\underline{\omega} = \sum_i \omega_i \underline{du}^i \tag{2.10}$$

with the components ω_i functions in $\mathcal{F}(M)$.

The second dual V** of a finite dimensional vector space V is naturally isomorphic to V. Thus, just as a 1-form $\underline{\omega}$ maps $\mathcal{T}(M)$ into $\mathcal{F}(M)$, we may think of a vector field $\vec{X}$ as a map from $\mathcal{T}_1(M)$ into $\mathcal{F}(M)$. As a natural generalization of this picture we define a <u>tensor</u> T_p <u>of type</u> (r,s) at $p \in M$ to be an (r+s)-linear map from $T_p^*(M) \times \ldots \times T_p^*(M) \times T_p(M) \times \ldots \times T_p(M)$ into the real numbers, where $T_p^*(M)$ appears r times and $T_p(M)$ appears s times. The space of all such maps is an n^{r+s}-dimensional vector space $(T^r_s)_p(M)$ called the tensor space of type (r,s) at p. A (smooth) <u>tensor field</u> T <u>of type</u> (r,s) on M is an assignment, to each point $p \in M$, of a tensor T_p of type (r,s) such that the function $T(\underline{\omega}_1, \ldots, \underline{\omega}_r, \vec{X}_1, \ldots, \vec{X}_s)$ given by

$$T(\underline{\omega}_1, \ldots, \underline{\omega}_r, \vec{X}_1, \ldots, \vec{X}_s)(p) = T_p(\underline{\omega}_{1p}, \ldots, \underline{\omega}_{rp}, \vec{X}_{1p}, \ldots, \vec{X}_{sp}) \tag{2.11}$$

is in $\mathcal{F}(M)$ for all $\underline{\omega}_1, \ldots, \underline{\omega}_r \in \mathcal{T}_1(M)$ and $\vec{X}_1, \ldots, \vec{X}_s \in \mathcal{T}(M)$. The space of all such tensor fields on M is denoted $\mathcal{T}^r_s(M)$. The components of T relative to the coordinate bases $\vec{\partial}_i$ and $\underline{du}^i$ defined on U are the functions

$$T^{i\ldots j}{}_{k\ldots l} = T(\underline{du}^i, \ldots, \underline{du}^j, \vec{\partial}_k, \ldots, \vec{\partial}_l) \ , \tag{2.12}$$

where $i,\ldots,j;k,\ldots,l = 1,\ldots,n$. Tensors of type (r,0) are said to be <u>contravariant</u> of degree r and tensors of type (0,s) are said to be <u>covariant</u> of degree s. A tensor of type (r,s) $(r,s \neq 0)$

is <u>mixed</u> of degree (r+s).

Although the definition of a tensor given above is perhaps the simplest, it is not the only way in which we can think of tensors. It is easy to see that a tensor T_p of type (r,s) may also be thought of as a (u+v)-linear map (with $u \leqslant r$, $v \leqslant s$) from $T^*_p(M) \times \ldots \times T^*_p(M) \times T_p(M) \times \ldots \times T_p(M)$ into $(T^{r-u}_{s-v})_p(M)$, where $T^*_p(M)$ now appears u times and $T_p(M)$ appears v times. With this interpretation, the tensor field T , now written T' , has the explicit action

$$T'(\underline{\omega}_1,\ldots,\underline{\omega}_u,\vec{X}_1,\ldots,\vec{X}_v) =$$

$$= \sum_{i\ldots j} \sum_{k\ldots l} T(\underline{\omega}_1,\ldots,\underline{\omega}_u,\underline{du}^i,\ldots,\underline{du}^j,\vec{X}_1,\ldots,\vec{X}_v,\vec{\partial}_k,\ldots,\vec{\partial}_l)$$
$$\cdot\ \vec{\partial}_i\otimes\ldots\otimes\vec{\partial}_j\otimes\underline{du}^k\otimes\ldots\otimes\underline{du}^l \qquad (2.13)$$

where the sums range from 1 to n and the quantities

$$(E_{i\ldots j}{}^{k\ldots l})_p = (\vec{\partial}_i\otimes\ldots\otimes\vec{\partial}_j\otimes\underline{du}^k\otimes\ldots\otimes\underline{du}^l)_p \quad , \qquad (2.14)$$

with (r-u) lower and (s-v) upper indices, are the basis vectors for $(T^{r-u}_{s-v})_p(M)$ defined by

$$E_{i\ldots j}{}^{k\ldots l}(\underline{du}^{i'},\ldots,\underline{du}^{j'},\vec{\partial}_{k'},\ldots,\vec{\partial}_{l'})$$
$$= \delta^{i'}_i\cdot\ \ldots\ \cdot\delta^{j'}_j\cdot\delta^k_{k'}\cdot\ \ldots\ \cdot\delta^l_{l'} \quad . \qquad (2.15)$$

The <u>tensor product</u> $\otimes$ used in (2.13) is an associative bilinear product used to create, from two given tensors of types (r,s) and (u,v), a new tensor of type (r+u,s+v). For example, if $\underline{\theta}$ and $\underline{\omega}$ are two 1-forms on M, then $\underline{\theta}\otimes\underline{\omega}$ is the tensor field of type (0,2) defined by

$$\underline{\theta}\otimes\underline{\omega}(\vec{U},\vec{V}) = \underline{\theta}(\vec{U})\cdot\underline{\omega}(\vec{V}) \quad . \tag{2.16}$$

The <u>exterior</u> or <u>wedge product</u> of $\underline{\theta}$ and $\underline{\omega}$ is the <u>2-form</u>

$$\underline{\theta}\wedge\underline{\omega} = (\underline{\theta}\otimes\underline{\omega} - \underline{\omega}\otimes\underline{\theta}) \quad , \tag{2.17}$$

which is an <u>antisymmetric</u> covariant tensor field of degree 2. If M is an n-manifold, then a <u>p-form</u>, $p \leq n$, on M is a completely antisymmetric covariant tensor field of degree p. In order to generalize the wedge product to products of p- and q-forms $(p+q \leq n)$ we require that $\wedge$ be associative and bilinear, and we define the p-form $\underline{\omega}_1\wedge \ldots \wedge\underline{\omega}_p$ constructed from p 1-forms to be the tensor that satisfies

$$\underline{\omega}_1\wedge \ldots \wedge\underline{\omega}_p(\vec{V}_1, \ldots ,\vec{V}_p) = \det \| \underline{\omega}_i(\vec{V}_j)\| \tag{2.18}$$

for all $\vec{V}_1, \ldots ,\vec{V}_p \in \mathcal{T}(M)$. If $\underline{\theta}$ is a p-form and $\underline{\phi}$ a q-form, then it follows that

$$\underline{\theta}\wedge\underline{\phi} = (-1)^{pq}\, \underline{\phi}\wedge\underline{\theta} \quad . \tag{2.19}$$

If $p+q > n$ then $\underline{\theta}\wedge\underline{\phi} = 0$. By convention, a function $f \in \mathcal{F}(M)$ is a 0-form and $f\wedge\underline{\theta} = f\cdot\underline{\theta}$.

The primary use of differential forms lies in the theory of integration on manifolds. Let M be an n-manifold with differential structure $\mathcal{D}$; let $\{(U_I,\phi_I)\}$ be a subset of $\mathcal{D}$ such that $\{U_I\}$ is a locally finite cover for M ; and let $\{\rho_\alpha\}$ be a partition of unity on M subordinate to the cover $\{U_I\}$ [33]. If $\underline{\omega}$ is an n-form on M then for any given α

$$\underline{\omega}_\alpha = \rho_\alpha \cdot \underline{\omega} \tag{2.20}$$

is an n-form on M which is non-zero only in some neighbourhood $U_J \in \{U_I\}$. Denoting by u^i , $i = 1,\ldots,n$, the local coordinates induced in this neighbourhood by the map ϕ_J , one can quickly verify that on U_J

$$\underline{\omega}_\alpha = f_\alpha \cdot \underline{du}^1 \wedge \ldots \wedge \underline{du}^n \tag{2.21}$$

for some unique $f_\alpha \in \mathcal{F}(U_J)$. The integral of $\underline{\omega}_\alpha$ on M is then defined to be

$$\int_M \underline{\omega}_\alpha = \int_{\phi_J(U_J)} f_\alpha(\phi_J^{-1}(x))\,d^n x \tag{2.22}$$

where the integral on the right hand side is the usual integral in Euclidean n-space. Noting that $\sum_\alpha \underline{\omega}_\alpha = \underline{\omega}$, we can now define the integral of $\underline{\omega}$ over M to be

$$\int_M \underline{\omega} = \sum_\alpha \int_M \underline{\omega}_\alpha \quad . \tag{2.23}$$

If $\underline{\omega}$ has compact support then this sum will converge and will be independent of the choice of cover $\{U_I\}$ or the partition of unity $\{\rho_\alpha\}$.

A very useful generalization of the classical theorems of Stokes and Gauss to an arbitrary manifold M may also be derived.

But before stating (without proof) this generalized Stokes' theorem we must define the exterior derivative operator $\underline{d}$. If M is an n-dimensional manifold and $f \in \mathcal{F}(M)$, then we define the 1-form $\underline{df} = \underline{d}f$ by

$$\underline{df}(\vec{V}) = \vec{V}f \qquad \text{for all} \quad \vec{V} \in \mathcal{T}(M) \quad . \tag{2.24}$$

More generally, if $\underline{\omega}$ is a p-form on M , then the exterior derivative $\underline{d\omega}$ of $\underline{\omega}$ is a (p+1)-form that is uniquely determined by the following properties of $\underline{d}$ [33]:

$$\text{(i)} \qquad \underline{d}(\underline{\theta}_1 + \underline{\theta}_2) = \underline{d\theta}_1 + \underline{d\theta}_2 \quad ; \tag{2.25}$$

$$\text{(ii)} \qquad \underline{d}(\underline{\theta} \wedge \underline{\phi}) = (\underline{d\theta}) \wedge \underline{\phi} + (-1)^p \underline{\theta} \wedge \underline{d\phi} \quad ; \tag{2.26}$$

$$\text{(iii)} \qquad \underline{d}(\underline{d\theta}) = 0 \quad ; \tag{2.27}$$

which must hold for all p-forms $\underline{\theta}, \underline{\theta}_1, \underline{\theta}_2$ and q-forms $\underline{\phi}$ on M . In particular, $\underline{d}(\underline{du}^i) = 0$, where $\underline{du}^i$ are the coordinate 1-forms in some neighbourhood U . Thus, when $\underline{\omega}$ is expanded in the form

$$\underline{\omega} = \sum_{i_1 < \ldots < i_p} \omega_{i_1 \ldots i_p} \cdot \underline{du}^{i_1} \wedge \ldots \wedge \underline{du}^{i_p} \quad , \tag{2.28}$$

its exterior derivative is just

$$\underline{d\omega} = \sum_{i_1 < \ldots < i_p} \underline{d}(\omega_{i_1 \ldots i_p}) \wedge \underline{du}^{i_1} \wedge \ldots \wedge \underline{du}^{i_p} \quad . \tag{2.29}$$

Now let $\underline{\omega}$ be an (n-1)-form on M and let M' be an n-dimensional submanifold of M with boundary $\partial M'$. A curve in $\partial M'$ is, in a natural way, also a curve in M ; so we may always think of a vector field $\vec{V} \in \mathcal{T}(\partial M')$ as the restriction to $\partial M'$ of some $\vec{V} \in \mathcal{T}(M)$. Similarly, the (n-1)-form $\underline{\omega}$ on M restricts to an (n-1)-form, also denoted by $\underline{\omega}$, on $\partial M'$. Since $\partial M'$ is an (n-1)-dimensional manifold, the integral of $\underline{\omega}$ over this boundary may be constructed as in (2.23). If M' is compact and oriented and $\partial M'$ is assigned the boundary orientation [33], then the <u>generalized Stokes' theorem</u> may be concisely expressed in the form

$$\int_{\partial M'} \underline{\omega} = \int_{M'} \underline{d\omega} \quad . \tag{2.30}$$

The exterior derivative operator $\underline{d}$ is intrinsic to the manifold M on which it is defined. It acts only on differential forms, raising the degree of a p-form to (p+1). In view of (2.24), it may be considered as a generalization of the gradient operator of vector calculus. Also intrinsic to each manifold is a second kind of differential operator $\mathcal{L}$ called the <u>Lie derivative</u>. If $\vec{X}$ and $\vec{Y}$ are vector fields on M , then the Lie derivative of $\vec{Y}$ along $\vec{X}$ is the vector field $\mathcal{L}_{\vec{X}}\vec{Y}$ defined by

$$(\mathcal{L}_{\vec{X}}\vec{Y})f = [\vec{X},\vec{Y}]f = \vec{X}(\vec{Y}f) - \vec{Y}(\vec{X}f) \tag{2.31}$$

for all $f \in \mathcal{F}(M)$. The Lie derivative of a function f along $\vec{X}$ is defined to be the function

$$\mathcal{L}_{\vec{X}}f = \vec{X}f \quad , \tag{2.32}$$

and if $\underline{\omega}$ is a 1-form, then $\mathcal{L}_{\vec{X}}\underline{\omega}$ is the 1-form which satisfies

$$(\mathcal{L}_{\vec{X}}\underline{\omega})(\vec{V}) = \vec{X}(\underline{\omega}(\vec{V})) - \underline{\omega}(\mathcal{L}_{\vec{X}}\vec{V}) \tag{2.33}$$

for all $\vec{V} \in \mathcal{T}(M)$. By requiring that $\mathcal{L}$ satisfy Leibniz' rule for the derivative of a product, we can define the Lie derivative of an arbitrary tensor field. For example, if T is a tensor field of type (1,1), then $\mathcal{L}_{\vec{X}}T$ is defined by

$$\mathcal{L}_{\vec{X}}T(\underline{\omega},\vec{V}) = \vec{X}(T(\underline{\omega},\vec{V})) - T(\mathcal{L}_{\vec{X}}\underline{\omega},\vec{V}) - T(\underline{\omega},\mathcal{L}_{\vec{X}}\vec{V}) \tag{2.34}$$

for all $\underline{\omega} \in \mathcal{T}_1(M)$ and $\vec{V} \in \mathcal{T}(M)$.

One more geometric structure deserves discussion here. A Riemannian metric on M is a (smooth) tensor field g of type (0,2) which is symmetric and positive definite. That is,

$$(g(\vec{U},\vec{V}))(x) = (g(\vec{V},\vec{U}))(x) \geqslant 0 \tag{2.35}$$

for all $\vec{U},\vec{V} \in \mathcal{T}(M)$ and $x \in M$, and $(g(\vec{U},\vec{V}))(x) = 0$ only if $\vec{U}_x = 0$ or $\vec{V}_x = 0$. Every differential manifold admits a Riemannian metric [45], and any Riemannian metric on M may be used to construct a distance function compatible with the topology of M . Let a and b be two points in M and let x(t) be a curve such that $x(t_a) = a$, $x(t_b) = b$, and $t_a < t_b$. Then, if g is a Riemannian metric on M , the length of the curve x(t) between a and b is

$$l(a,b)[x] = \int_{t_a}^{t_b} \sqrt{g(\vec{X}(t),\vec{X}(t))}\, dt \tag{2.36}$$

where $\vec{X}(t)$ is the vector tangent to the curve at x(t) (defined as in equation (2.5)). Each such curve joining a and b has a well defined length which is greater than zero if a and b are distinct. The metric <u>distance</u> between a and b is the least upper bound of the lengths of all smooth curves connecting a and b :

$$d(a,b) = \underset{[x]}{\text{lub}} \; (l(a,b)[x]) \qquad (2.37)$$

The distance function satisfies the conditions (1.3), and the open balls defined with the use of d provide a base for the topology of M . Each curve x(t) joining a and b such that l(a,b)[x] = d(a,b) is called a <u>geodesic</u> of the metric g on M .

If we relax the condition that g be a smooth tensor field, allowing it to be <u>divergent</u> at some point $x \in M$, then we may still be able to define a distance function as in (2.37), but d need no longer be compatible with the topology of M . If g fails to be positive definite, but is still symmetric and non-singular , then it is called a pseudo-Riemannian metric. In such a case the distance function d ceases to be well defined, and g may no longer be used to generate the open sets of M .

3. Affine and Riemannian Geometry

Let M be an n-manifold and let $\vec{U},\vec{V} \in \mathcal{T}(M)$ be any two vector fields. The Lie derivative $\pounds_{\vec{U}}\vec{V}$ is, in a natural way, also a vector field on M . However, this derivative cannot be considered as a derivative of $\vec{V}$ in the usual sense, since (when written in a coordinate representation) it depends on the derivatives of the components of $\vec{U}$ as well as on the components themselves. In fact

$$\pounds_{\vec{U}}\vec{V} = -\pounds_{\vec{V}}\vec{U} \quad , \tag{3.1}$$

so that $\vec{U}$ and $\vec{V}$ are really on an equal footing. In order to construct derivatives of $\vec{V}$ more akin to the directional derivatives of vectors in Euclidean space, we must define on M a new geometric structure ∇ called a <u>covariant derivative</u>. We define ∇ such that for any $\vec{U},\vec{V},\vec{W} \in \mathcal{T}(M)$ and $f \in \mathcal{F}(M)$, $\nabla_{\vec{U}}\vec{V} \in \mathcal{T}(M)$ and

$$\left.\begin{aligned} \nabla_{(\vec{U}+\vec{V})}\vec{W} &= \nabla_{\vec{U}}\vec{W} + \nabla_{\vec{V}}\vec{W} \quad , \\ \nabla_{\vec{U}}(\vec{V}+\vec{W}) &= \nabla_{\vec{U}}\vec{V} + \nabla_{\vec{U}}\vec{W} \quad , \\ \nabla_{f\vec{U}}\vec{V} &= f\nabla_{\vec{U}}\vec{V} \quad , \\ \nabla_{\vec{U}}(f\vec{V}) &= (\vec{U}f)\vec{V} + f\nabla_{\vec{U}}\vec{V} \quad . \end{aligned}\right\} \tag{3.2}$$

From ∇ we can construct two important tensors: a vector-valued 2-form $\underline{\Theta}$ called the <u>torsion</u>, which is defined by

$$\underline{\Theta}(\vec{U},\vec{V}) = \nabla_{\vec{U}}\vec{V} - \nabla_{\vec{V}}\vec{U} - [\vec{U},\vec{V}] \quad ; \tag{3.3}$$

and the <u>curvature</u> tensor R, which is of type (1,3) and has the action

$$R(\vec{U},\vec{V})\vec{W} = \nabla_{\vec{U}}\nabla_{\vec{V}}\vec{W} - \nabla_{\vec{V}}\nabla_{\vec{U}}\vec{W} - \nabla_{[\vec{U},\vec{V}]}\vec{W} \quad . \tag{3.4}$$

The operator $(R(\vec{U},\vec{V}))_x$, $x \in M$, is a linear transformation of $T_x(M)$. It is antisymmetric,

$$R(\vec{U},\vec{V}) = -R(\vec{V},\vec{U}) \quad , \tag{3.5}$$

and its trace is a symmetric tensor S called the Ricci tensor:

$$S(\vec{V},\vec{W})(x) = \mathrm{Trace}(\vec{U}_x \to (R(\vec{U},\vec{V})\vec{W})_x) \quad . \tag{3.6}$$

The curvature satisfies the cyclic Bianchi identity,

$$R(\vec{U},\vec{V})\vec{W} + R(\vec{W},\vec{U})\vec{V} + R(\vec{V},\vec{W})\vec{U} = \nabla_{\vec{U}}\underline{\Theta}(\vec{V},\vec{W}) + \nabla_{\vec{W}}\underline{\Theta}(\vec{U},\vec{V})$$

$$+ \nabla_{\vec{V}}\underline{\Theta}(\vec{W},\vec{U}) - \underline{\Theta}(\vec{U},\underline{\Theta}(\vec{V},\vec{W})) - \underline{\Theta}(\vec{W},\underline{\Theta}(\vec{U},\vec{V})) - \underline{\Theta}(\vec{V},\underline{\Theta}(\vec{W},\vec{U})) \quad , \tag{3.7}$$

which follows directly from the Jacobi identity

$$[[\vec{U},\vec{V}],\vec{W}] + [[\vec{W},\vec{U}],\vec{V}] + [[\vec{V},\vec{W}],\vec{U}] = 0 \quad . \tag{3.8}$$

By setting the covariant derivative of a function equal to its ordinary derivative:

$$\nabla_{\vec{U}} f = \vec{U} f \quad , \tag{3.9}$$

and requiring that ∇ satisfy Leibniz' rule for the derivative of a product, we can define the covariant derivative of an arbitrary tensor field. If $\underline{\omega}$ is a 1-form and T a tensor field of type (1,1), then the covariant analogues of (2.33) and (2.34) are

$$\nabla_{\vec{U}}\underline{\omega}(\vec{V}) = \vec{U}(\underline{\omega}(\vec{V})) - \underline{\omega}(\nabla_{\vec{U}}\vec{V}) \quad , \tag{3.10}$$

$$\nabla_{\vec{U}} T(\underline{\omega},\vec{V}) = \vec{U}(T(\underline{\omega},\vec{V})) - T(\nabla_{\vec{U}}\underline{\omega},\vec{V}) - T(\underline{\omega},\nabla_{\vec{U}}\vec{V}) \quad . \tag{3.11}$$

Bianchi's second identity,

$$\nabla_{\vec{U}} R(\vec{V},\vec{W}) + \nabla_{\vec{W}} R(\vec{U},\vec{V}) + \nabla_{\vec{V}} R(\vec{W},\vec{U}) = R(\vec{U},\underline{\Theta}(\vec{V},\vec{W})) + R(\vec{V},\underline{\Theta}(\vec{W},\vec{U})) + R(\vec{W},\underline{\Theta}(\vec{U},\vec{V})) \quad , \tag{3.12}$$

is obtained by noting that

$$\nabla_{\vec{U}} R(\vec{V},\vec{W})\vec{X} = \nabla_{\vec{U}}(R(\vec{V},\vec{W})\vec{X}) - R(\nabla_{\vec{U}}\vec{V},\vec{W})\vec{X} - R(\vec{V},\nabla_{\vec{U}}\vec{W})\vec{X}$$
$$- R(\vec{V},\vec{W})\nabla_{\vec{U}}\vec{X} \quad ,$$

expanding R in terms of the covariant derivative, and making use of (3.8) and the tensorial nature of the torsion.

In any coordinate neighbourhood $U \subset M$, with local coordinates u^i, the action of ∇ is completely determined by the functions Γ^i_{jk}, defined by

$$\Gamma^i_{jk} = \underline{du}^i(\nabla_{\vec{\partial}_j}\vec{\partial}_k) \quad . \tag{3.13}$$

These are called the components of the affine connection Γ associated with ∇, the name reflecting the fact that ∇ allows the comparison of vectors in the affine tangent spaces of distinct points along a curve. Let $x(t)$, $t \in \mathbb{R}$, be a curve in M and let $\vec{X}$ be a vector field on M such that $\vec{X}_{x(t_0)}$ is, for each t_0, the tangent vector to the curve at $x(t_0)$. A vector field $\vec{V} \in \mathcal{T}(M)$ is said to be <u>parallel along the curve</u> $x(t)$ if

$$(\nabla_{\vec{X}}\vec{V})_{x(t)} = f(t)\cdot\vec{V}_{x(t)} \tag{3.14}$$

for some function $f:\mathbb{R} \to \mathbb{R}$ and for each $t \in \mathbb{R}$. If the vector field $\vec{X}$ is itself parallel along the curve $x(t)$, then the curve is called a <u>path</u> of the affine connection Γ. A diffeomorphism $s:\mathbb{R} \to \mathbb{R}$ may be used to reparameterize the points of

$x(t)$, yielding the new curve

$$x'(t) = x(s(t)) \quad . \tag{3.15}$$

Moreover, if $x(t)$ is a path of Γ , then s may always be chosen so that

$$(\nabla_{\vec{X}'} \vec{X}')_{x'(t)} = 0 \qquad \text{for all} \quad t \in \mathbb{R} \quad , \tag{3.16}$$

where $\vec{X}'$ bears the same relation to $x'(t)$ as $\vec{X}$ bears to $x(t)$. The path $x'(t)$ is then called a <u>geodesic</u> of ∇ , and the parameter t is called an <u>affine parameter</u> for the geodesic.

Now let g be a Riemannian or pseudo-Riemannian metric on M . The covariant derivative ∇ is said to be <u>metrical</u> if, for all $\vec{U},\vec{V},\vec{W} \in \mathcal{T}(M)$,

$$\nabla_{\vec{U}} g(\vec{V},\vec{W}) = 0 \quad , \tag{3.17}$$

or, more simply, $\nabla g = 0$. If ∇ is metrical, then the norm $g(\vec{X}_{x(t)},\vec{X}_{x(t)})$ of the tangent vector to a geodesic $x(t)$ is independent of t . The length (as defined in (2.36)) of the segment of $x(t)$ between $x(t_a)$ and $x(t_b)$ is thus directly proportional to the affine length $(t_b - t_a)$. In the Riemannian case this will always be positive, but if g is not positive-definite then there will also exist <u>null</u> geodesics, which have length zero, and <u>time-like</u> geodesics, whose lengths are pure imaginary.

A metrical covariant derivative which has vanishing torsion:

$$\nabla_{\vec{U}}\vec{V} - \nabla_{\vec{V}}\vec{U} = [\vec{U},\vec{V}] \quad , \tag{3.18}$$

is called a Riemannian covariant derivative. Each metric g on M uniquely determines a Riemannian covariant derivative ∇ on M, and from now on it is this covariant derivative with which we shall deal. The manifold M, together with g and ∇, is called a (pseudo-)Riemannian manifold. The Riemann curvature tensor of M has, in addition to (3.5) and (3.7), the symmetries

$$g(\vec{X},R(\vec{U},\vec{V})\vec{W}) = -g(\vec{W},R(\vec{U},\vec{V})\vec{X}) \quad , \tag{3.19}$$

$$g(\vec{X},R(\vec{U},\vec{V})\vec{W}) = g(\vec{V},R(\vec{W},\vec{X})\vec{U}) \quad . \tag{3.20}$$

Since g is non-degenerate, it is possible to find, in a neighbourhood U_x of each point $x \in M$, a set of vector fields $\vec{h}_i$, $i = 1,\dots,n$, which are orthogonal,

$$g(\vec{h}_i,\vec{h}_j) = 0 \quad \text{for} \quad i \neq j \quad , \tag{3.21}$$

and which are normalized to plus or minus one,

$$g(\vec{h}_i,\vec{h}_i) = \varepsilon_i = \pm 1 \quad . \tag{3.22}$$

The sum $\sigma = \sum_i \varepsilon_i$ is the signature of the metric g and is an invariant quantity. In terms of the $\vec{h}_i$, the Ricci tensor of ∇ may be written

$$S(\vec{U},\vec{V}) = \sum_i \varepsilon_i g(\vec{h}_i,R(\vec{h}_i,\vec{U})\vec{V}) \quad . \tag{3.23}$$

It satisfies the contracted form,

$$\sum_i \varepsilon_i \nabla_{\vec{h}_i} S(\vec{h}_i,\vec{V}) = \tfrac{1}{2}\vec{\nabla} S \quad , \qquad (3.24)$$

of the Bianchi identity (3.12), where S on the right hand side is the Ricci scalar,

$$S = \sum_i \varepsilon_i S(\vec{h}_i,\vec{h}_i) \quad . \qquad (3.25)$$

For any metrical connection Γ the Ricci tensor is symmetric:

$$S(\vec{U},\vec{V}) = S(\vec{V},\vec{U}) \quad . \qquad (3.26)$$

4. Submanifold Geometry

Of especial importance in the physical discussions of chapters 2 and 3 is the relationship between the geometry of a (pseudo-) Riemannian manifold and that of its submanifolds. Let $\mathcal{S}$ and M be manifolds of dimensions n and $(n+p)$, respectively, and let $e:\mathcal{S} \to M$ be an embedding. By identifying $\mathcal{S}$ with its image $e(\mathcal{S})$ in M , one can immediately see that e induces, for each $x \in \mathcal{S}$, an injective map $de_x:T_x(\mathcal{S}) \to T_{e(x)}(M)$ whose co-domain is the subspace $T^{\parallel}_{e(x)}(M)$ of $T_{e(x)}(M)$ consisting of those vectors which are tangent to curves in $e(\mathcal{S})$. The image of a vector field $\vec{v} \in \mathcal{T}(\mathcal{S})$ is denoted $e_*\vec{v}$, and is an assignment to each $e(x) \in e(\mathcal{S})$ of a vector in $T^{\parallel}_{e(x)}(M)$. Such a vector field is said to be parallel to the submanifold.

Now let M have defined on it a metric g with Riemannian covariant derivative ∇ . The pullback $\tilde{g} = e^*g$ of g onto S is the symmetric tensor field defined by

$$\tilde{g}(\vec{u},\vec{v})(x) = g(e_*\vec{u},e_*\vec{v})(e(x)) \tag{4.1}$$

for all $x \in S$ and all $\vec{u},\vec{v} \in \mathcal{T}(S)$. If g is not positive definite, then $\tilde{g}$ need not be positive definite, nor even a metric on S . However, we shall consider here only those embeddings e for which $\tilde{g}$ is a Riemannian metric.

A general vector field $\vec{V}$ on $e(S)$ is a smooth assignment of a vector $\vec{V}_{e(x)} \in T_{e(x)}(M)$ to each point $e(x)$. If each $\vec{V}_{e(x)}$ is in $T^{\parallel}_{e(x)}(M)$, then $\vec{V}$ is a parallel vector field. (Note that this has nothing to do with "parallel along a curve".) On the other hand, a <u>perpendicular vector field</u> $\vec{V}$, is a vector field on $e(S)$ such that

$$g(\vec{V},\vec{u})(e(x)) = 0 \tag{4.2}$$

for all $x \in S$ and all parallel vector fields $\vec{u}$. The space of all vector fields on $e(S)$ will be denoted by $\mathcal{T}_e(S)$, the space of parallel vector fields by $\mathcal{T}^{\parallel}_e(S)$, and the space of perpendicular vector fields by $\mathcal{T}^{\perp}_e(S)$. For convenience, the distinction between $\vec{v} \in \mathcal{T}(S)$ and $e_*\vec{v} \in \mathcal{T}^{\parallel}_e(S)$ will be dropped.

It is always possible to choose, in some neighbourhood $U \subset e(S)$ of each point $e(x)$ of the submanifold, a set $\vec{h}_i$, $i = 1,\ldots,n$, of orthonormal vector fields in $\mathcal{T}^{\parallel}_e(S)$:

$$g(\vec{h}_i,\vec{h}_j) = \delta_{ij} \quad , \tag{4.3}$$

and a corresponding set $\vec{n}_\mu$, $\mu = 1,\ldots,p$, of orthogonal unit vectors in $\mathcal{T}_e^\perp(S)$:

$$g(\vec{n}_\mu,\vec{n}_\nu) = \varepsilon_\mu \delta_{\mu\nu} = \pm\delta_{\mu\nu} \quad . \tag{4.4}$$

The projection operator Π defined, on each such neighbourhood U , by

$$\Pi(\vec{V}) = \sum_i g(\vec{V},\vec{h}_i)\vec{h}_i \tag{4.5}$$

for all $\vec{V} \in \mathcal{T}_e(S)$,may then be used to project out the <u>parallel part</u> $\vec{V}^{\parallel} = \Pi(\vec{V})$ of the field $\vec{V}$. Similarly, the <u>perpendicular part</u> of $\vec{V}$ is $\vec{V}^{\perp} = \vec{V} - \vec{V}^{\parallel} = \sum_\mu \varepsilon_\mu g(\vec{V},\vec{n}_\mu)\vec{n}_\mu$. The metric $\tilde{g}$ can now be redefined by setting

$$\tilde{g}(\vec{U},\vec{V}) = g(\Pi(\vec{U}),\Pi(\vec{V})) \quad , \tag{4.6}$$

so that its arguments $\vec{U}$ and $\vec{V}$ need no longer be parallel vector fields.

For all $\vec{u},\vec{v} \in \mathcal{T}_e^{\parallel}(S)$, the vector field $\nabla_{\vec{u}}\vec{v}$ may always be written in the form

$$\nabla_{\vec{u}}\vec{v} = \tilde{\nabla}_{\vec{u}}\vec{v} + \alpha(\vec{u},\vec{v}) \tag{4.7}$$

where $\tilde{\nabla}_{\vec{u}}\vec{v} \in \mathcal{T}_e^{\parallel}(S)$ and $\alpha(\vec{u},\vec{v}) \in \mathcal{T}_e^{\perp}(S)$. It is easy to check that $\tilde{\nabla}$ satisfies the conditions (3.2) for a covariant derivative. Moreover, $\tilde{\nabla}$ has vanishing torsion,

$$\tilde{\nabla}_{\vec{u}}\vec{v} - \tilde{\nabla}_{\vec{v}}\vec{u} = [\vec{u},\vec{v}] \quad , \tag{4.8}$$

because ∇ has vanishing torsion: If $\vec{u},\vec{v} \in \mathcal{T}_e^{\parallel}(S)$, then

$$[\vec{u},\vec{v}] = \tilde{\nabla}_{\vec{u}}\vec{v} - \tilde{\nabla}_{\vec{v}}\vec{u} + \alpha(\vec{u},\vec{v}) - \alpha(\vec{v},\vec{u}) \tag{4.9}$$

will also be a parallel vector field; but this implies that

$$\alpha(\vec{u},\vec{v}) = \alpha(\vec{v},\vec{u}) \quad , \tag{4.10}$$

reducing (4.9) to (4.8) . Finally, $\tilde{\nabla}$ is metrical, and hence Riemannian, since

$$\begin{aligned} \vec{u}(\tilde{g}(\vec{v},\vec{w})) &= \vec{u}(g(\vec{v},\vec{w})) \\ &= g(\nabla_{\vec{u}}\vec{v},\vec{w}) + g(\vec{v},\nabla_{\vec{u}}\vec{w}) \\ &= \tilde{g}(\tilde{\nabla}_{\vec{u}}\vec{v},\vec{w}) + \tilde{g}(\vec{v},\tilde{\nabla}_{\vec{u}}\vec{w}) \end{aligned} \tag{4.11}$$

for all $\vec{u},\vec{v},\vec{w} \in \mathcal{T}_e^{\parallel}(S)$.

The operator α defined by (4.7) is called the <u>second fundamental form</u> of S for the embedding e . It is linear in its first argument since $\nabla_{f\vec{u}}\vec{v} = f\nabla_{\vec{u}}\vec{v}$, and the symmetry condition (4.10) indicates that α must also be linear in the second argument. In analogy with $\tilde{g}$, we set

$$\alpha(\vec{U},\vec{V}) = \alpha(\Pi(\vec{U}),\Pi(\vec{V})) \tag{4.12}$$

so that α is a tensorial map from $\mathcal{T}_e(S)\times\mathcal{T}_e(S)$ into $\mathcal{T}_e^{\perp}(S)$. Making use of the unit vector fields $\vec{n}_\mu$, $\alpha(\vec{U},\vec{V})$ may be expanded in the form

$$\alpha(\vec{U},\vec{V}) = \sum_\mu K^\mu(\vec{U},\vec{V})\vec{n}_\mu \quad . \tag{4.13}$$

The geometric objects K^{μ} introduced here are p real-valued symmetric tensor fields on $e(S)$, called the extrinsic curvatures of the submanifold in the directions $\vec{n}_{\mu}$.

If $\vec{\xi} = \sum_{\mu} \xi^{\mu} \vec{n}_{\mu}$ is a perpendicular vector field and $\vec{u}, \vec{v} \in \mathcal{T}_e^{\parallel}(S)$, then

$$\begin{aligned} g(\vec{v}, \nabla_{\vec{u}} \vec{\xi}) &= \vec{u}(g(\vec{v}, \vec{\xi})) - g(\nabla_{\vec{u}} \vec{v}, \vec{\xi}) \\ &= -g(\alpha(\vec{u}, \vec{v}), \vec{\xi}) \\ &= -\sum_{\mu} \varepsilon_{\mu} K^{\mu}(\vec{u}, \vec{v}) \xi^{\mu} \quad , \end{aligned} \tag{4.14a}$$

and

$$\begin{aligned} g(\vec{n}_{\mu}, \nabla_{\vec{u}} \vec{\xi}) &= \vec{u}(g(\vec{n}_{\mu}, \vec{\xi})) - g(\nabla_{\vec{u}} \vec{n}_{\mu}, \vec{\xi}) \\ &= \varepsilon_{\mu} \vec{u} \xi^{\mu} - \sum_{\nu} \xi^{\nu} g(\nabla_{\vec{u}} \vec{n}_{\mu}, \vec{n}_{\nu}) \quad . \end{aligned} \tag{4.14b}$$

Equations (4.7) and (4.14) are known respectively as the formulas of Gauss and Weingarten [41]. If $e(S)$ is a hypersurface of M , that is, if $p = 1$, then there is only one unit normal vector field $\vec{n}$, and one extrinsic curvature K . In this case, which is the only case that we shall consider from now on, the last term in (4.14b) vanishes, because $g(\nabla_{\vec{u}} \vec{n}, \vec{n}) = 0$.

The hypersurface curvature $\hat{R}$ is defined to have the action

$$\hat{R}(\vec{u}, \vec{v}) \vec{w} = \tilde{\nabla}_{\vec{u}} \tilde{\nabla}_{\vec{v}} \vec{w} - \tilde{\nabla}_{\vec{v}} \tilde{\nabla}_{\vec{u}} \vec{w} - \tilde{\nabla}_{[\vec{u}, \vec{v}]} \vec{w} \tag{4.15}$$

for all $\vec{u}, \vec{v}, \vec{w} \in \mathcal{T}_e^{\parallel}(S)$, and to satisfy

$$\hat{R}(\vec{U}, \vec{V}) \vec{W} = \Pi(\hat{R}(\Pi(\vec{U}), \Pi(\vec{V})) \Pi(\vec{W})) \quad . \tag{4.16}$$

It has the usual symmetries (3.5) , (3.7) , (3.19), and (3.20) and satisfies the hypersurface analogue,

$$\tilde{\nabla}_{\vec{u}}\tilde{R}(\vec{v},\vec{w}) + \tilde{\nabla}_{\vec{w}}\tilde{R}(\vec{u},\vec{v}) + \tilde{\nabla}_{\vec{v}}\tilde{R}(\vec{w},\vec{u}) = 0 \quad , \tag{4.17}$$

of the differential identity (3.12). The Ricci tensor of $\tilde{\nabla}$ is defined by

$$\tilde{S}(\vec{U},\vec{V}) = \sum_i \tilde{g}(\vec{h}_i,\tilde{R}(\vec{h}_i,\vec{U})\vec{V}) \quad , \tag{4.18}$$

and the Ricci scalar is

$$\tilde{S} = \sum_i \tilde{S}(\vec{h}_i,\vec{h}_i) \quad . \tag{4.19}$$

Any tensor which is left invariant by Π , as are $\tilde{R}$ and $\tilde{S}$, is called a <u>hypersurface</u> <u>tensor</u> <u>field</u>. Although in many cases, such as K , the tilde will be omitted, a superscript tilde indicates that the field under consideration is a hypersurface field.

The curvature R of ∇ is related to the hypersurface curvature and the extrinsic curvature through (4.7) . For all $\vec{u},\vec{v},\vec{w} \in \mathcal{T}^{\parallel}_e(S)$,

$$\begin{aligned} R(\vec{u},\vec{v})\vec{w} &= \nabla_{\vec{u}}\nabla_{\vec{v}}\vec{w} - \nabla_{\vec{v}}\nabla_{\vec{u}}\vec{w} - \nabla_{[\vec{u},\vec{v}]}\vec{w} \\ &= \nabla_{\vec{u}}(\tilde{\nabla}_{\vec{v}}\vec{w} + K(\vec{v},\vec{w})\vec{n}) - \nabla_{\vec{v}}(\tilde{\nabla}_{\vec{u}}\vec{w} + K(\vec{u},\vec{w})\vec{n}) \\ &\quad - \tilde{\nabla}_{[\vec{u},\vec{v}]}\vec{w} - K([\vec{u},\vec{v}],\vec{w})\vec{n} \\ &= \tilde{R}(\vec{u},\vec{v})\vec{w} + \varepsilon\sum_i\{K(\vec{u},\vec{w})K(\vec{v},\vec{h}_i) - K(\vec{v},\vec{w})K(\vec{u},\vec{h}_i)\}\vec{h}_i \\ &\quad + \{\tilde{\nabla}_{\vec{u}}K(\vec{v},\vec{w}) - \tilde{\nabla}_{\vec{v}}K(\vec{u},\vec{w})\}\vec{n} \quad , \end{aligned} \tag{4.20}$$

and, as a result of the symmetry (3.19),

$$R(\vec{u},\vec{v})\vec{n} = -\varepsilon\sum_{i}\{\tilde{\nabla}_{\vec{u}}K(\vec{v},\vec{h}_i) - \tilde{\nabla}_{\vec{v}}K(\vec{u},\vec{h}_i)\}\vec{h}_i \quad . \tag{4.21}$$

Equation (4.20) is equivalent to the classical equations of Gauss and Codazzi [41].

REFERENCES

1. A. Einstein, The Meaning of Relativity, Princeton University Press, Princeton (5th ed., 1956).
2. D. Hilbert, Konigl. Gesell. d. Wiss. Göttingen, Nachr., Math.-Phys. Kl., 395 (1915).
3. S. Hojman, K. Kuchař, and C. Teitelboim, Nature (London) 245 (1973),97; Ann. Phys. (N.Y.) 96 (1976),88; C. Teitelboim, Ann. Phys. (N.Y.) 79 (1973),542; "The Hamiltonian Structure of Spacetime", in: GRG Einstein centennial volume, (A. Held and P. Bergman, Eds.), Plenum, New York (to be published).
4. W.E. Thirring, Ann. Phys. (N.Y.) 16 (1961),96; R.P. Feynman, Lectures on Gravitation, unpublished, Calif. Inst. of Technology (1963); S. Weinberg, Phys. Rev. 138B (1965),988; S. Deser, Gen. Rel. and Grav. 1 (1970),9.
5. D.G. Boulware and S. Deser, Ann. Phys. (N.Y.) 89 (1975),193.
6. C.W. Misner, K.S. Thorne, and J.A. Wheeler, Gravitation, Freeman, San Francisco (1973).
7. General relativity has been cast into the Hamiltonian form by P.A.M. Dirac, Proc. Roy. Soc. (London) Ser.A 246 (1958), 326; 246 (1958),333; Phys. Rev. 114 (1959),924; and also by R. Arnowitt, S. Deser, and C.W. Misner, The Dynamics of General Relativity, in: Gravitation: An Introduction to Current Research, (L. Witten, Ed.), Wiley, New York (1962).

8. G.Y. Rainich, Trans. Am. Math. Soc. 27 (1925),106.

9. C.W. Misner and J.A. Wheeler, Ann. Phys. (N.Y.) 2 (1957),525.

10. C. Møller, The Energy Momentum Complex in General Relativity and Related Problems, in: Les Theories Relativistes de la Gravitation, CNRS, Paris (1962).

11. K. Kuchař, Canonical Quantization of Gravity, in: Relativity, Astrophysics, and Cosmology, (W. Israel, Ed.), Reidel, Dordrecht, Holland (1973).

12. C.N. Yang and R.L. Mills, Phys. Rev. 96 (1954),191; C.Llewellyn-Smith, New Scientist, 10 April 1975, 74.

13. Lu Qi-keng, Chin. J. Phys. 23 (1975),153.

14. A. Einstein, Preuss. Akad. d. Wiss., Phys.-Math. Kl., Sitzungberichte, (1928),217,224; K. Hayashi, Phys. Rev. D19 (1979),35.

15. A. Einstein, Ann. Math. 46 (1945),578; A. Einstein and E.G. Strauss, *ibid.* 47 (1946),731; op. cit. 1, Appendix II; A. Einstein, Rev. Mod. Phys. 20 (1948),35.

16. J.W. Moffat and D.H. Boal, Phys. Rev. D11 (1975),1375; D.H. Boal and J.W. Moffat, Phys. Rev. D11 (1975),2026; J.W. Moffat, Space-time Structure in a Generalization of Gravitation Theory, Univ. of Toronto Preprint (1976); G. Kunstatter and J.W. Moffat, Phys. Rev. D17 (1978),396.

17. E. Schrödinger, Proc. Roy. Irish Acad. L.I. A213 (1948); Space-Time Structure, Cambridge Univ. Press, Cambridge, England (1954).

18. J.W. Moffat, Phys. Rev. D19 (1979),3554.

19. E. Cartan, Ann. Ec. Norm. Sup. 40 (1923),325; 41 (1924),1; 42 (1925),17.

20. F.W. Heyl, P. von der Heyde, G.D. Kerlick, and J.M. Nester, Rev. Mod. Phys. 48 (1976),393.

21. J.A. Schouten, Ricci Calculus, Springer, Berlin (1954).

22. S.W. Hawking and G.F.R. Ellis, The large scale structure of space-time, Cambridge Univ. Press, Cambridge, England (1973).

23. B.S. DeWitt, Phys. Rev. 160 (1967),1113; 162 (1967),1195; 162 (1967),1239; C.J. Isham, An Introduction to Quantum Gravity, in: Quantum Gravity - An Oxford Symposium, (C.J. Isham, R. Penrose, and D.W. Sciama, Eds.), Oxford Univ. Press, Oxford (1975).

24. G. 't Hooft and M. Veltman, Ann. Inst. Henri Poincaré, Sect. A, XX (1974),69; S. Deser and P. van Nieuwenhuizen, Phys. Rev. D10 (1974),401; D10 (1974),411.

25. L. Smolin, What is the Problem of Quantum Gravity?, revised version of Part I of L. Smolin's Ph. D. thesis, Harvard Univ. (1979).

26. op. cit. 10; L. Parker and S.A. Fulling, Phys. Rev. D7 (1973),2357; Ann. Phys. (N.Y.) 87 (1974),176.

27. H. Everett III, Rev. Mod. Phys. 29 (1957),454; B.S. DeWitt, in: The Many-worlds Interpretation of Quantum Mechanics, (B.S. DeWitt and N. Graham, Eds.), Princeton Univ. Press, Princeton (1973).

28. R. Penrose, Twistor Theory, Its Aims and Achievements, in: Quantum Gravity - An Oxford Symposium, (C.J. Isham, R. Penrose, and D.W. Sciama, Eds.), Oxford Univ. Press, Oxford (1975).

29. D. Finkelstein, Phys. Rev. 184 (1969),126; D5 (1972),320; D5 (1972),2922; D9 (1974),2219; D. Finkelstein, G. Frye, and W. Susskind, Phys. Rev. D9 (1974),2231; D. Finkelstein, Process Philosophy and Quantum Dynamics, in: Physical Theory as Logico-Operational Structure, XXX - XXX, (Hooker, Ed.), Reidel, Dordrecht, Holland (1978).

30. Ya. B. Zeldovich, Cosmology and the Early Universe, in: General Relativity - An Einstein Centenary Survey, (S.W. Hawking and W. Israel, Eds.), Cambridge Univ. Press, New York (1979).

31. J.A. Wheeler, Ann. Phys. (N.Y.) 2 (1957),604; Superspace and the Nature of Quantum Geometrodynamics, in: Battelle Rencontres, (C.M. DeWitt and J.A. Wheeler, Eds.), Benjamin, New York (1968).

32. J. Hempel, 3-manifolds, Princeton Univ. Press, Princeton (1976).

33. V. Guillemin and A. Pollack, Differential Topology, Prentice-Hall, Englewood Cliffs, New Jersey (1974).

34. op. cit. 32, p. 27.

35. J. Milnor, Morse Theory, Princeton Univ. Press, Princeton (1969); op. cit. 33.

36. T.W.B. Kibble and S. Randjbar-Daemi, Jour. Phys. A 13 (1980), 141; S. Randjbar-Daemi, B.S. Kay, and T.W.B. Kibble, Phys. Lett. 91B (1980),417.

37. A.A. Markov, Insolubility of the problem of homeomorphy, in: Proc. Int. Congress Math. 1958, Cambridge Univ. Press, Cambridge, England.

38. S. Lipschutz, General Topology, McGraw-Hill, New York (1965).

39. J.R. Munkres, Elementary Differential Topology, Princeton Univ. Press, Princeton (revised edition, 1966).

40. I.R. Porteous, Topological Geometry, Van Nostrand Reinhold, London (1969).

41. S. Kobayashi and K. Nomizu, Foundations of Differential Geometry, Wiley, New York (Vol. I, 1963; Vol. II, 1969).

42. S.S. Cairns, Introductory Topology, Ronald Press, New York (1961).

43. E.H. Spanier, Algebraic Topology, McGraw-Hill, New York (1966).

44. M. Kervaire, Comment. Math. Helv. 34 (1960),257.

45. N. Steenrod, The Topology of Fibre Bundles, Princeton Univ. Press, Princeton (1951).