Quantum Theory of Gravity. I. The Canonical Theory*

BRYCE S. DEWITT
Institute for Advanced Study, Princeton, New Jersey
and
Department of Physics, University of North Carolina, Chapel Hill, North Carolina†
(Received 25 July 1966; revised manuscript received 9 January 1967)

Following an historical introduction, the conventional canonical formulation of general relativity theory is presented. The canonical Lagrangian is expressed in terms of the extrinsic and intrinsic curvatures of the hypersurface $x^0=$constant, and its relation to the asymptotic field energy in an infinite world is noted. The distinction between finite and infinite worlds is emphasized. In the quantum theory the primary and secondary constraints become conditions on the state vector, and in the case of finite worlds these conditions alone govern the dynamics. A resolution of the factor-ordering problem is proposed, and the consistency of the constraints is demonstrated. A 6-dimensional hyperbolic Riemannian manifold is introduced which takes for its metric the coefficient of the momenta in the Hamiltonian constraint. The geodesic incompletability of this manifold, owing to the existence of a frontier of infinite curvature, is demonstrated. The possibility is explored of relating this manifold to an infinite-dimensional manifold of 3-geometries, and of relating the structure of the latter manifold in turn to the dynamical behavior of space-time. The problem is approached through the WKB approximation and Hamilton-Jacobi theory. Einstein's equations are revealed as geodesic equations in the manifold of 3-geometries, modified by the presence of a "force term." The classical phenomenon of gravitational collapse shows that the force term is not powerful enough to prevent the trajectory of space-time from running into the frontier. The as-yet unresolved problem of determining when the collapse phenomenon represents a real barrier to the quantum-state functional is briefly discussed, and a boundary condition at the barrier is proposed. The state functional of a finite world can depend only on the 3-geometry of the hypersurface $x^0=$constant. The label x^0 itself is irrelevant, and "time" must be determined intrinsically. A natural definition for the inner product of two such state functionals is introduced which, however, encounters difficulties with negative probabilities owing to the barrier boundary condition. In order to resolve these difficulties, a simplified model, the quantized Friedmann universe, is studied in detail. In order to obtain nonstatic wave functions which resemble a universe evolving, it is necessary to introduce a clock. In order that the combined wave functions of universe-*cum*-clock be normalizable, it turns out that the periods of universe and clock must be commensurable. Wave packets exhibiting quasiclassical behavior are constructed, and attention is called to the phenomenological character of "time." The inner-product definition is rescued from its negative-probability difficulties by making use of the fact that probability flows in a closed finite circuit in configuration space. The article ends with some speculations on the uniqueness of the state functional of the actual universe. It is suggested that a viewpoint due to Everett should be adopted in its interpretation.

1. INTRODUCTION

ALMOST as soon as quantum field theory was invented by Heisenberg, Pauli, Fock, Dirac, and Jordan, attempts were made to apply it to fields other than the electromagnetic field which had given it—and indeed quantum mechanics itself—birth. In 1930 Rosenfeld[1] applied it to the gravitational field which, at the time, was still regarded as the *other* great entity of Nature. Rosenfeld was the first to note some of the special technical difficulties involved in quantizing gravity and made some early attempts to develop general methods for handling them. As an application of his methods he computed the gravitational self-energy of a photon in lowest order of perturbation theory. He obtained a quadratically divergent result, confirming that the divergence malady of field theory, which had already been discovered in connection with the electron's electromagnetic self-energy, was widespread and deep seated. It is tempting, and perhaps no longer premature, to read into Rosenfeld's result a forecast that quantum gravidynamics was destined, from the very beginning, to be inextricably linked with the difficult issues lying at the theoretical foundations of particle physics.

During physics's great boom of the thirties the difficult issues of field theory were inevitably often bypassed. Moreover, it was recognized early that as far as the gravitational field is concerned its quanta (assuming they exist) can produce no *observable* effects until energies of the order of 10^{28} eV are reached, this fantastic energy corresponding to the so-called "Planck length" $(\hbar G/c^3)^{1/2}\approx 10^{-33}$ cm, where G is the gravitation constant. Hence, after Rosenfeld's initial studies years passed before anything essentially new was done in quantum gravidynamics, and even today interest in this area of research is confined to a very small group of workers.

In 1950 the author[2] reperformed Rosenfeld's self-energy calculation in a manifestly Lorentz-covariant and gauge-invariant manner. This work was stimulated by the then new "renormalization" methods, which

* This research was supported in part by the Air Force Office of Scientific Research under Grant No. AFOSR-153-64.
† Permanent address.
[1] L. Rosenfeld, Ann. Physik **5**, 113 (1930); Z. Physik **65**, 589 (1930).
[2] B. S. DeWitt, Ph.D. thesis, Harvard University, 1950 (unpublished).

had been developed by Tomonaga, Schwinger, and Feynman, and had as its aim a demonstration that Rosenfeld's result implies merely a renormalization of charge rather than a nonvanishing photon mass. An unanticipated source of potential difficulty arose in this calculation from the fact that not one but *two* gauge groups are simultaneously present (the group associated with gravity in addition to the familiar electromagnetic group) and that these groups are not combined in the form of a direct product but rather in the form of a *semidirect product* based on the automorphisms of the electromagnetic gauge group under general coordinate transformations. This means that if a fixed choice of gauge is to be maintained, every coordinate transformation must be accompanied by an electromagnetic gauge transformation. The calculation was pushed, however, again only to the lowest order of perturbation theory; in this order, which involves only single closed Feynman loops, the ensuing complications are easily dealt with.

At about the same time investigations of a more ambitious kind were undertaken by Bergmann.[3] Although the renormalization philosophy had proved a resounding success in quantum electrodynamics it was still under critical attack because the methods then (and frequently even now) in use involved the explicit manipulation of divergent quantities. Similar (although more elementary) difficulties also persisted in classical particle theories with one important exception, namely, the theory of the interaction of point masses with gravity. In 1938, Einstein, Infeld, and Hoffmann[4] had shown that the laws of motion of such particles follow from the gravitational field equations alone, without divergent quantities ever appearing or such concepts as self-mass intervening at any time. Moreover, this result had been subsequently extended to include electrically charged particles, and gave promise of being applicable to spinning particles as well. The gravitational field thus appeared as a kind of classical regulator, and Bergmann reasoned that the same might be true in the quantum theory. Since the fields are basic, in the Einstein-Infeld-Hoffmann view, and the particles are merely singularities in the fields, Bergmann's first task was to quantize the gravitational field. It was to be hoped that commutation relations for particle position and momentum would then follow as corollaries.

The obstacles which Bergmann faced were enormous. First of all, since the laws of particle motion depend crucially on the nonlinear properties of the Einstein field equations, it was necessary to quantize the full nonlinear gravitational field. Secondly, it was necessary to find some way of defining particle position and momentum in terms of field variables alone. Thirdly, it would eventually be necessary to include spin, so that quantized particles obeying a Dirac-like equation could be described. Fourthly, it would be necessary to extract Fermi statistics (for the particles) out of the Bose statistics obeyed by the gravitational field. Finally, it would be necessary ultimately to remove the asymmetry between particle and field inherent in the Einstein-Infeld-Hoffmann approach, so as to be able, as in quantum electrodynamics, to account for pair production and vacuum polarization. It is not surprising that Bergmann's goal today remains as elusive as ever.

To achieve this goal Bergmann set out upon the classical canonical road in search of a Hamiltonian. Despite the fact that canonical methods, by singling out the time for special treatment, run counter to the spirit of any relativistic theory—above all, such a completely covariant theory as general relativity—such a procedure seemed a good one for several reasons. Firstly, no other method was then known. Secondly, canonical methods afford quick insights into certain aspects of any theory. Thirdly, it seemed that standard perturbation methods would become available for certain types of calculations.

However, Bergmann immediately ran into major difficulties (some of which had already been foreseen by Rosenfeld) in the first stages of his program. These are referred to as "the problem of *constraints*," and are manifested in the following ways: Some of the field variables possess no conjugate momenta; the momenta conjugate to the remaining field variables are not all dynamically independent; the field equations themselves are not linearly independent, and some of them involve no second time derivatives, thus complicating the Cauchy problem. These difficulties are all related and arise from the existence of the general coordinate-transformation group as an invariance group for the theory.

Similar difficulties had already been encountered with the electromagnetic field and methods for handling them were well known. The same methods, however, proved to be much more difficult to apply in the case the gravitational field. An obstacle is created, for example, by the fact that not all of the relations between the momenta (i.e., the constraints) are linear. Moreover, because the invariance group of gravity is non-Abelian (in contrast to the gauge group of electrodynamics) tedious calculations must be performed to check that the commutators of the various constraints lead to no inconsistencies.

Bergmann and his co-workers performed much valuable ground work in formulating the difficulties in a precise way and in partially resolving them. In the meantime additional help came from an unexpected quarter. In 1950 Dirac[5] published the outline of a general Hamiltonian theory which is in principle applicable to any system describable by an action functional. Dirac's methods were quickly seized upon by Pirani and Schild[6]

[3] A bibliography of Bergmann's early work will be found in P. G. Bergmann, Helv. Phys. Acta Suppl. 4, 79 (1956).

[4] A. Einstein, L. Infeld, and B. Hoffmann, Ann. Math. **39**, 65 (1938).

[5] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).

[6] F. A. E. Pirani and A. Schild, Phys. Rev. **79**, 986 (1950).

for application to the gravitational field. Unfortunately, these authors chose to develop the theory within the framework of a "parameter formalism," in the hope, which eventually proved to be misplaced, of retaining a manifest covariance which Dirac's methods would otherwise destroy. The complexity of the resulting algebra prevented them from computing all of the constraints.

The theory remained in this incomplete state for several years. It was not until impetus was provided by the first international relativity conference in Bern in 1955 (Jubilee of Relativity Theory) and the second one in Chapel Hill in 1957 that things began to move again. A small step forward was made by the author,[7] who showed, using the Pirani-Schild formalism, that the four so-called "primary" constraints could, by a phase transformation, be changed into pure momenta. This meant that the state functional for gravity must be independent of the $g_{0\mu}$ components of the metric tensor ($\mu=0, 1, 2, 3$). Shortly afterward Higgs[8] showed that three of the so-called "secondary" or "dynamical" constraints are the generators of infinitesimal transformations of the three "spatial" coordinates x^1, x^2, x^3. The implication of this was that the state functional must be independent of the coordinates chosen in the spacelike cross sections $x^0=$ constant and hence cannot be taken to be an *arbitrary* functional of the metric components g_{ij} ($i, j=1, 2, 3$). Developments thereafter came rapidly. Dirac himself had by this time begun to apply his methods to the gravitational field.[9] As a result of simplifications and clarifications which he introduced, it became easy to show that the fourth dynamical constraint is consistent with the others, and the formal theory achieved for the first time a state of technical completion. It was then possible to begin asking "What does it all mean?"

On the classical side, the problems of physical interpretation were soon resolved by the work of Arnowitt, Deser, and Misner,[10] who showed how to use the canonical theory to provide a rigorous characterization of gravitational radiation and "energy." In the quantum domain, however, the interpretation of the formalism remained puzzling and obscure for several years, because one did not know the right questions to ask. It is only recently that the relevant issues have begun to come into focus, largely as a result of the patient researches of Wheeler,[11] whose ideas have proved a great source of stimulation to many workers, including the author.

[7] Reported at a meeting at Stevens Institute of Technology in January, 1958 (unpublished).

[8] P. W. Higgs, Phys. Rev. Letters **1**, 373 (1958); **3**, 66 (1959).

[9] P. A. M. Dirac, Proc. Roy. Soc. (London) **A246**, 326 (1958); **A246**, 333 (1958); Phys. Rev. **114**, 924 (1959).

[10] R. Arnowitt, S. Deser, and C. W. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962).

[11] The work of Wheeler and his associates is well described in J. A. Wheeler, *Relativity Groups and Topology, 1963 Les Houches Lectures* (Gordon and Breach Science Publishers, Inc., New York, 1964). This reference contains a large bibliography of additional papers on quantization, collapse, and many other related topics.

The present paper is the direct outcome of conversations with Wheeler,[12] during which one fundamental question in particular kept recurring: *What is the structure of the domain manifold for the quantum-mechanical state functional?* The attempt to answer this question has required a more far-reaching analysis of the technical structure of the canonical theory than can be found in the previous literature. The results of this analysis are here presented and used to develop an interpretative framework which, although tentative, is perhaps capable of serving in a variety of contexts.

Attention is mainly confined to the case of closed finite worlds, firstly because the issues which finite worlds raise are more critical and bizarre, and secondly because the case of infinite worlds is better handled within the framework of the so-called manifetlsy covariant theory which will be treated in two subsequent papers of this series. The latter theory, which has also achieved a state of technical completion following the pioneering work of Feynman,[13] differs utterly in its structure from the canonical theory, and so far no one has established a rigorous mathematical link between the two. At the present time the two theories play complementary roles, the canonical theory describing the quantum behavior of 3-space regarded as a time-varying geometrical object, and the covariant theory describing the behavior of real and virtual gravitons propagating in this object.

Section 2 of the present paper begins with the derivation of the canonical Lagrangian. Its structure in terms of the extrinsic and intrinsic curvatures of the hypersurface $x^0=$ constant is displayed, and attention is called to its relation to the total field energy in an asymptotically flat world. Section 3 is devoted to the primary and secondary constraints of the theory and to the independent question of coordinate conditions. Quantization is introduced in Sec. 4. Here the puzzling question of the role of a vanishing Hamiltonian is resolved by emphasizing the distinction between finite and infinite worlds. Asymptotic energy is an indispensable concept in an infinite world, and the Hamiltonian must be chosen accordingly. In a finite world there is no asymptotic energy, and an intrinsic description of the dynamics must be found, based on the constraints alone. The consistency of the constraints is demonstrated by straightforward computation of their commutators. The factor-ordering problem is disposed of by formal arguments which in effect assert that field variables taken at the same space-time point should be regarded as freely commutable. The "χ constraints" are shown to be the generators of 3-dimensional coordinate transformations.

In Sec. 5 the metric representation is introduced. The

[12] Any errors or wrong conjectures it contains are the author's own.

[13] R. P. Feynman, Mimeographed letter to V. F. Weisskopf dated January 4 to February 11, 1961 (unpublished); Acta Phys. Polon. **24**, 697 (1963).

distinction between finite and infinite worlds is again noted, and it is emphasized that the state functional in the former case depends only on the 3-geometry of the hypersurface $x^0=$ constant and not on the label x^0 itself. The concept of a manifold $\mathfrak{M}$ of 3-geometries is introduced, and the role played by the Hamiltonian constraint in determining its geometrical structure is suggested. The coefficient of the momenta in the Hamiltonian constraint may be regarded as a metric of a 6-dimensional hyperbolic Riemannian manifold M. The structure of this manifold is studied in detail, and its geodesic incompletability, owing to the existence of a frontier of infinite curvature, is noted. The possibility of relating M to the question of "intrinsic time" for the state functional is discussed, and a natural definition for the inner product of two state functionals is proposed.

In Sec. 6 a natural metric based on M is assigned to the infinite-dimensional manifold $\mathfrak{M}$, and some of the properties of geodesics in $\mathfrak{M}$ are examined. An attempt is then made to indicate the extent to which the dynamical properties of the quantized gravitational field are determined by the structure of $\mathfrak{M}$. The attempt is heuristic and far from complete, and much work remains to be done. The problem is approached through the WKB approximation and Hamilton-Jacobi theory. Einstein's equations are revealed as geodesic equations in $\mathfrak{M}$, modified by the presence of a "force term." The classical phenomenon of gravitational collapse shows that the force term is not powerful enough to prevent the trajectory of 3-space from striking the frontier of $\mathfrak{M}$. The problem of determining when the collapse phenomenon represents a real barrier to the quantum state functional is briefly discussed, and a boundary condition (vanishing state functional) at the barrier is proposed.

The barrier boundary condition raises difficulties with the definition of probability. In order to study these difficulties it is useful to test the theory on a simplified model. In Sec. 7 the quantized Friedmann universe is studied in detail, and its static wave functions in the WKB approximation are obtained. In order to obtain nonstatic wave functions which resemble a dynamical universe evolving it is necessary to introduce a clock. The combined wave functions of universe-*cum*-clock are studied, and it is pointed out that normalizability of the wave functions requires precise commensurability between the periods of universe and clock.

Wave packets exhibiting quasiclassical behavior are constructed in Sec. 8, in three different representations. Two of these make use of proper times defined by the clock and the universe respectively; the third treats universe and clock symmetrically through their mutual correlations. Attention is called to the deficiencies of the first two representations arising from the fact that, in a covariant theory, time is only a phenomenological concept. In the third representation probability flows in a closed finite circuit in configuration space, and wave packets do *not* ultimately spread in time. Use is made of this fact in Sec. 9 to show how the inner-product definition can be rescued from the negative probability difficulties arising from the barrier boundary condition $\Psi=0$ at $R=0$ ($R=$ radius of universe). It is also shown that the conventional Cauchy data for the wave function suffice to determine the quantum state completely.

Section 10 is devoted to speculations on the general theory. An interpretation of quantum mechanics due to Everett (see Ref. 52) is described and proposed for dealing with the concept of "a wave function for the universe." Such an interpretation is essential if the wave function is unique. Evidence is presented that the Hamiltonian constraint may indeed have only one solution. The problem of time-reversal invariance and entropy is briefly discussed. Two technical appendices follow at the end of the article.

Attention is called to the following points of notation: Latin indices range over the values 1, 2, 3 and Greek indices over the values 0, 1, 2, 3. Differentiation is denoted by a comma. The coordinates x^0 and x^i are assumed to be timelike and spacelike, respectively,and the geometry of space-time is assumed to be such that the hypersurfaces $x^0=$ constant are capable of carrying a complete set of Cauchy data. So-called "absolute units" in which $\hbar=c=16\pi G=1$ (G being the gravitation constant) are used throughout, as is also the signature $-+++$ for the space-time metric $g_{\mu\nu}$. The Riemann and Ricci tensors, and the curvature scalar, are taken in the respective forms

$$R_{\mu\nu\sigma}{}^{\tau}\equiv\Gamma_{\nu\sigma}{}^{\tau}{}_{,\mu}-\Gamma_{\mu\sigma}{}^{\tau}{}_{,\nu}+\Gamma_{\nu\sigma}{}^{\rho}\Gamma_{\mu\rho}{}^{\tau}-\Gamma_{\mu\sigma}{}^{\rho}\Gamma_{\nu\rho}{}^{\tau}, \quad (1.1)$$

$$R_{\mu\nu}\equiv R_{\sigma\mu\nu}{}^{\sigma}, \quad (1.2)$$

$${}^{(4)}R\equiv R_{\mu}{}^{\mu}\equiv g^{\mu\nu}R_{\mu\nu}, \quad (1.3)$$

where

$$\Gamma_{\mu\nu}{}^{\sigma}\equiv\tfrac{1}{2}g^{\sigma\tau}(g_{\mu\tau,\nu}+g_{\nu\tau,\mu}-g_{\mu\nu,\tau}), \quad g_{\mu\sigma}g^{\sigma\nu}=\delta_{\mu}{}^{\nu}. \quad (1.4)$$

The corresponding tensors in the spacelike cross sections $x^0=$ constant are distinguished by means of a prefixed superscript (3). These conventions have the property that ${}^{(4)}R$ is non-negative in a space-time containing normal matter and satisfying Einstein's equations, and that ${}^{(3)}R$ is positive in a 3-space of positive curvature.

2. EXTRINSIC AND INTRINSIC CURVATURE. CLASSIC FORM OF THE LAGRANGIAN

The canonical theory begins with the following decomposition of the metric tensor:

$$(g_{\mu\nu})=\begin{pmatrix}-\alpha^2+\beta_k\beta^k & \beta_j\\ \beta_i & \gamma_{ij}\end{pmatrix},$$
$$(g^{\mu\nu})=\begin{pmatrix}-\alpha^{-2} & \alpha^{-2}\beta^j\\ \alpha^{-2}\beta^i & \gamma^{ij}-\alpha^{-2}\beta^i\beta^j\end{pmatrix}, \quad (2.1)$$

$$\gamma_{ik}\gamma^{kj}=\delta_i{}^j, \quad \beta^i=\gamma^{ij}\beta_j. \quad (2.2)$$

When the conventional Einstein Lagrangian density is reexpressed in terms of the new variables, it is found, after some calculation, to take the form

$$\mathfrak{L}\equiv g^{1/2}\,{}^{(4)}R=\alpha\gamma^{1/2}(K_{ij}K^{ij}-K^2+{}^{(3)}R) -2(\gamma^{1/2}K)_{,0}+2(\gamma^{1/2}K\beta^i-\gamma^{1/2}\gamma^{ij}\alpha_{,j})_{,i} \quad (2.3)$$

where

$$g\equiv-\det(g_{\mu\nu})=\alpha^2\gamma, \qquad \gamma\equiv\det(\gamma_{ij}), \quad (2.4)$$

$$K_{ij}\equiv\tfrac{1}{2}\alpha^{-1}(\beta_{i.j}+\beta_{j.i}-\gamma_{ij,0}), \quad K^{ij}\equiv\gamma^{ik}\gamma^{jl}K_{kl}, \quad K\equiv\gamma^{ij}K_{ij}, \quad (2.5)$$

the dots denoting covariant differentiation based on the 3-metric γ_{ij}.

The quantity K_{ij}, which transforms as a symmetric tensor under spatial coordinate transformations, is known as the *second fundamental form*. It describes the curvature of the hypersurface $x^0=$ constant as viewed from the 4-dimensional space-time in which it is embedded. It is therefore also frequently called the *extrinsic curvature tensor* of the hypersurface, as opposed to the *intrinsic curvature tensor* ${}^{(3)}R_{ij}$, which depends only on γ_{ij} *in* the hypersurface. In a flat space-time ${}^{(3)}R_{ij}$ is completely determined by K_{ij}, but in a manifold of arbitrary curvature there need be no relationship between the two. The contracted forms ${}^{(3)}R$ and $K_{ij}K^{ij}-K^2$ will be referred to as the intrinsic and extrinsic curvatures, respectively.

The last two terms of Eq. (2.3), being total derivatives, are dynamically irrelevant and may be dropped. The Lagrangian then becomes

$$L\equiv\int\alpha\gamma^{1/2}(K_{ij}K^{ij}-K^2+{}^{(3)}R)d^3x, \quad (2.6)$$

which has the classic form "kinetic energy minus potential energy," with the extrinsic curvature playing the role of kinetic energy and the negative of the intrinsic curvature that of potential energy.

The form (2.6) is manifestly invariant under 3-dimensional general coordinate transformations. Precisely for this reason it differs from the Lagrangian of ordinary field theories, for the ${}^{(3)}R$ term of its integrand contains linearly occurring second spatial derivatives of the field variables. With an ordinary field theory in an infinite universe this would be of no significance. The usual assumption that the field vanishes outside some arbitrarily large but finite spatial domain permits linearly occurring second derivatives to be eliminated by partial integration without affecting either the dynamical equations or the canonical definition of energy. In the case of gravity, however, the field never vanishes outside a finite domain unless space-time is flat, and although such a partial integration leaves the dynamical equations unaffected it does change the definition of energy. It is easy to verify, in fact, that it subtracts from the Lagrangian (2.6) a surface integral E_∞ given by

$$E_\infty=\int_\infty\alpha\gamma^{1/2}\gamma^{ij}(\gamma_{ik,j}-\gamma_{ij,k})dS^k, \quad (2.7)$$

and hence adds a corresponding quantity to the canonical energy. In an asymptotically flat world it is always possible to find an asymptotically Minkowskian reference frame in which α, β_i, and γ_{ij} take the static Schwarzschild forms

$$\alpha\underset{r\to\infty}{\to}1-\frac{M}{16\pi r}, \quad \beta_i\underset{r\to\infty}{\to}0, \quad \gamma_{ij}\underset{r\to\infty}{\to}\delta_{ij}+\frac{M}{8\pi}\frac{x^ix^j}{r^3}, \quad (2.8)$$

where $r^2\equiv x^ix^i$ and M is the effective gravitational mass of the field distribution. Substitution of (2.8) into (2.7) yields

$$E_\infty=M. \quad (2.9)$$

It is to be noted that the removal of E_∞ from the Lagrangian does not correspond to a mere redefinition of the energy zero point. E_∞ is not a fixed constant but depends on the state of the field. In fact it is *the* energy, for as we shall see presently the canonical "energy" based on (2.6) always vanishes. (Indeed, E_∞ is the energy even when other fields are present.) Since neither (2.6) nor (2.7) have any explicit dependence on x^0, the quantity E_∞ is conserved. General relativity is unique among field theories in that its energy may always be expressed as a surface integral. This was the source of Bergmann's hope to use gravity as a regulator, but it is also a source of difficulties. We note in particular that the surface integral vanishes for a closed finite world. It is only for infinite asymptotically flat worlds that the energy concept has meaning.

3. THE CONSTRAINTS

The momenta conjugate to α,β_i, and γ_{ij} will be denoted by π, π^i, and π^{ij}, respectively. They have the explicit forms

$$\pi=\frac{\delta L}{\delta\alpha_{,0}}=0, \quad (3.1)$$

$$\pi^i=\frac{\delta L}{\delta\beta_{i,0}}=0, \quad (3.2)$$

$$\pi^{ij}=\frac{\delta L}{\delta\gamma_{ij,0}}=-\gamma^{1/2}(K^{ij}-\gamma^{ij}K), \quad (3.3)$$

Eqs. (3.1) and (3.2) being known as the *primary constraints*. The primary constraints are purely formal statements, which express the fact that the Lagrangian (2.6) is independent of the "velocities" $\alpha_{,0}$ and $\beta_{i,0}$.[14]

[14] Failure to bring the Lagrangian into the form (2.6) was responsible for the difficulties originally encountered with the primary constraints.

These "velocities" are arbitrary and cannot be re-expressed in terms of momenta. They therefore cannot be removed from the Hamiltonian, which, with the aid of (3.3), takes the form

$$H=\int(\pi\alpha_{,0}+\pi^i\beta_{i,0}+\pi^{ij}\gamma_{ij,0})d^3x-L$$

$$=\int(\pi\alpha_{,0}+\pi^i\beta_{i,0}+\alpha\mathcal{H}+\beta_i\mathcal{X}^i)d^3x, \tag{3.4}$$

where

$$\mathcal{H}\equiv\tfrac{1}{2}\gamma^{-1/2}(\gamma_{ik}\gamma_{jl}+\gamma_{il}\gamma_{jk}-\gamma_{ij}\gamma_{kl})\pi^{ij}\pi^{kl}-\gamma^{1/2}\,{}^{(3)}R \tag{3.5a}$$

$$\equiv\gamma^{1/2}(K_{ij}K^{ij}-K^2-{}^{(3)}R), \tag{3.5b}$$

$$\mathcal{X}^i\equiv-2\pi^{ij}{}_{,j}\equiv-2\pi^{ij}{}_{,j}-\gamma^{il}(2\gamma_{jl,k}-\gamma_{jk,l})\pi^{jk}. \tag{3.6}$$

The momenta, as well as $\mathcal{H}$ and $\mathcal{X}^i$, are all 3-densities of unit weight.

It is not hard to show that Einstein's empty-space field equations may be obtained by taking the Poisson bracket of the various dynamical variables with the Hamiltonian (3.4) and then imposing Eqs. (3.1) and (3.2) as external constraints. Since the undermined "velocities" $\alpha_{,0}$ and $\beta_{i,0}$ are multiplied in (3.4) by π and π^i, their Poisson brackets with anything may be ignored. If desired, one can always assign definite values to α and β_i which may be purely numerical or may depend on the γ_{ij} and π^{ij}. Each choice corresponds to the imposition of certain conditions on the space-time coordinates. For example, one may choose

$$\alpha=1,\quad \beta_i=0, \tag{3.7a}$$

which reduces the Hamiltonian to

$$H=\int\mathcal{H}d^3x. \tag{3.7b}$$

Another favorite choice is

$$K\equiv\tfrac{1}{2}\gamma^{-1/2}\gamma_{ij}\pi^{ij}=0,\quad(\gamma^{1/2}\gamma^{ij})_{,j}=0, \tag{3.8a}$$

which corresponds to the requirement that the volume of every hypersurface $x^0=$ constant be stationary under small timelike deformations,[15] and that the spatial coordinates in each hypersurface be harmonic. To obtain the explicit forms of the conditions which Eqs. (3.8a) impose upon α and β_i, one notes that these equations imply the vanishing not only of their left-hand sides but of all their space-time derivatives as well. Taking the Poisson brackets of K and $(\gamma^{1/2}\gamma^{ij})_{,j}$ with the Hamiltonian (3.4) one finds the conditions

$$\begin{aligned}&\alpha_{,}{}^{i}-{}^{(3)}R\alpha=0,\\&[\gamma^{1/2}(\beta^i{}_{,}{}^{j}+\beta^j{}_{,}{}^{i}-\gamma^{ij}\beta^k{}_{,k})+2\alpha\pi^{ij}]_{,j}=0.\end{aligned} \tag{3.8b}$$

In an infinite asymptotically flat world these equations, which are of the elliptic type, may be solved subject to the boundary conditions $\alpha\to1$, $\beta_i\to0$ at infinity in asymptotically Minkowskian coordinates. In a finite world of nonvanishing curvature, however, they usually possess either no physically admissible solutions, i.e., solutions for which α remains everywhere positive, or no solutions at all. Since the Laplace-Beltrami operator has a negative spectrum, the first equation, for example, cannot be solved in a 3-sphere.

Conditions of the above type correspond merely to restrictions on the coordinates and have no physical content. There exist conditions of yet another type which actually restrict the dynamical freedom of the field and which hold regardless of whether a specific choice has been made for α and β_i or not. These are obtained by noting that since the primary constraints hold for all time, the x^0 derivatives of π and π^i must vanish. Stating this in the form of a Poisson bracket with H, one arrives immediately at the so-called *secondary or dynamical constraints*:

$$\mathcal{H}=0, \tag{3.9}$$

$$\mathcal{X}^i=0. \tag{3.10}$$

Equation (3.9) will be called the *Hamiltonian constraint*[16] in virtue of the structure of the function $\mathcal{H}$, which appears in (3.5b) as the difference of the extrinsic and intrinsic curvatures, in analogy with the classic form of the Hamiltonian as the sum of the kinetic and potential energies. This Hamiltonian, however, vanishes, as does indeed the total Hamiltonian (3.4). That is to say, in any "Ricci-flat" space-time (i.e., one satisfying Einstein's empty-space equations) the extrinsic and intrinsic curvatures of any hypersurface are equal. As has been emphasized by Wheeler,[11] the converse of this theorem is also true, namely, if $\mathcal{H}$ vanishes over every hypersurface then space-time is Ricci-flat. This suggests, as will be verified later, that it is the Hamiltonian constraint which provides the essential description of the "intrinsic" (i.e., coordinate-independent) dynamics of the gravitational field.

4. QUANTIZATION, CONSISTENCY OF THE CONSTRAINTS, FACTOR ORDERING

In the quantum theory, Poisson brackets become commutators. This means that the constraint Eqs. (3.1), (3.2), (3.9), and (3.10) cannot become operator equations, for otherwise the Hamiltonian (3.4) would yield no dynamics at all, extrinsic or intrinsic. Instead they become conditions on the state vector Ψ[5,9]:

$$\pi\Psi=0, \tag{4.1}$$

$$\pi^i\Psi=0, \tag{4.2}$$

$$\mathcal{H}\Psi=0, \tag{4.3}$$

$$\mathcal{X}^i\Psi=0. \tag{4.4}$$

[15] If 3-space is infinite this applies to the volume inside every finite domain.

[16] All four constraints (3.9), (3.10) are sometimes referred to as "Hamiltonian constraints." We prefer to reserve the terminology for this particularly important constraint.

These quantum constraints are often a source of puzzlement and confusion. Consider the equation

$$\gamma_{ij}(x^0,\mathbf{x})=e^{iHx^0}\gamma_{ij}(0,\mathbf{x})e^{-iH^0},\quad \mathbf{x}\equiv(x^1,x^2,x^3) \tag{4.5}$$

which is the quantum-mechanical relation expressing the field operator γ_{ij} on an arbitrary hypersurface in terms of the corresponding operator on the hypersurface $x^0=0$. Suppose we choose α and β_i as in Eq. (3.7a). Then Eq. (4.3) and its conjugate imply[17]

$$H\Psi=0,\quad \Psi^\dagger H=0, \tag{4.6}$$

and hence

$$\Psi^\dagger\gamma_{ij}(x^0,\mathbf{x})\Psi=\Psi^\dagger\gamma_{ij}(0,\mathbf{x})\Psi. \tag{4.7}$$

A similar result holds for any other field operator or product of field operators. Since the statistical results of any set of observations are ultimately expressible in terms of expectation values, one therefore comes to the conclusion that nothing ever happens in quantum gravidynamics, that the quantum theory can never yield anything but a static picture of the world.[18]

To see what is wrong with this conclusion one must examine the behavior of H, or more precisely $\mathfrak{H}$, at infinity. In an infinite asymptotically flat world the field disperses ultimately to a state of infinite weakness. In the asymptotic region $\mathfrak{H}$ therefore tends to its dominant linear term $\gamma_{ii,jj}-\gamma_{ij,ij}=0$,which is the well-known fourth constraint of linearized gravity theory.[10] This term is the asymptotic limit of the term which is removed from the integrand of H by the partial integration discussed in Sec. 2, and which gives rise to the surface integral (2.7). In the linearized theory, however, it becomes a constraint which has no relation to the total energy. Therefore if the full theory is to be applicable not only in the nonlinear region but also at infinity where the linear theory holds sway, it must make use of the Hamiltonian

$$H_\infty\equiv H+E_\infty, \tag{4.8}$$

which results from the partial integration. The integrand of this Hamiltonian reduces, in the asymptotic region, to an expression *quadratic* in the γ's and π's, namely, the usual integrand of the linearized theory.

It follows that in an infinite asymptotically flat world Eq. (4.5) should be replaced by

$$\gamma_{ij}(x^0,\mathbf{x})=e^{iH_\infty x^0}\gamma_{ij}(0,\mathbf{x})e^{-iH_\infty x^0}. \tag{4.9}$$

Even with this replacement, however, the appearance of the world is still static whenever Ψ is an eigenstate of energy-momentum. To obtain nonstatic behavior one must construct *wave packets*, by superposing many different momenta. But this is precisely what one wants to do in order to provide S-matrix theory, for example, with a rigorous foundation and insure that the field really does disperse ultimately to a state of infinite weakness.

[17] $\mathfrak{H}$ is assumed to be ordered in an Hermitian fashion.

[18] Cf. A. Komar, Phys. Rev. **153**, 1385 (1967).

Although the above discussion makes use of the coordinate system defined by Eqs. (3.7a), the same problems arise in any other asymptotically Minkowskian coordinate system, and the same conclusions apply. To the extent that we can ignore the possible lack of commutativity of α and β_i with $\mathfrak{H}$ and χ^i in the construction of an Hermitian Hamiltonian, the same apparent static behavior of the field will occur whenever we incorrectly use H instead of H_∞ in Eq. (4.9).

It should be noted that coordinate conditions such as (3.7a) and (3.8a) are *operator* equations and not constraints on the state vector.[19] (This follows from the complete arbitrariness of α and β_i in the classical theory.) On the other hand, equations such as (3.8a), which hold only when α and β_i are suitably restricted, are *not* operator equations. Indeed, they are not even constraints, but become instead *expectation-value equations*

$$\Psi^\dagger K\Psi=0,\quad \Psi^\dagger(\gamma^{1/2}\gamma^{ij})_{,j}\Psi=0, \tag{3.8c}$$

which hold for all values of x^0 provided they hold at some initial instant and Eqs. (3.8b) are satisfied. They do not hold in all permissible states merely in virtue of (3.8b).

Although we know that the physical content of the classical theory is unaffected by the choice of coordinates, it is not so easy to prove, using the canonical theory, that the results of a calculation of some physical *quantum* amplitude is independent of the choice of coordinates. It is not enough merely to know, for example, that two different coordinate systems both take the Minkowskian form $\alpha\to 1$, $\beta_i\to 0$, $\gamma_{ij}\to\delta_{ij}$ at infinity, in order to conclude that the physical S matrix remains unchanged under the transformation from one system to the other, for the operator ΔH, which represents the change in the Hamiltonian in passing from one system to the other, produces effects which propagate to infinity. In order to prove invariance of the S matrix under coordinate transformations (including q-number coordinate transformations), one would have to show that ΔH affects only the nonphysical field modes at infinity. The obstacle to such a demonstration is the lack of commutativity of the operators appearing in the dynamical equations, particularly when α and β_i depend nonlocally on γ_{ij} and π^{ij}. Although noncommutativity has no effect on the scattering amplitudes in lowest order, it plays havoc with the radiative corrections. For the study of radiative corrections a manifestly covariant theory is almost essential. In the following paper of this series the theory of gravitational radiative

[19] There is an alternative approach to the quantum theory of gravity which makes use of an action functional which is not coordinate-invariant and which generates no primary or secondary constraints. In this approach the constraints must be imposed from the outside. They take the form of coordinate conditions whose form is not arbitrary but is determined by the action functional itself. In this case the coordinate conditions *are* constraints on the state vector. This is the approach which has been followed, for example, by Gupta [S. N. Gupta, in *Recent Developments in General Relativity* (Pergamon Press, Inc., New York, 1962)].

corrections will be displayed in all its complexity, and the S matrix in the mainfestly covariant theory will be proved to be fully coordinate invariant. This result has not yet been proved in the canonical theory, and for this reason we shall include little further discussion of the case of infinite asymptotically flat worlds in this paper, but will concentrate henceforth on finite worlds.

In the finite case there is no distinction between H and H_∞, and hence we must face up anew to the difficulties posed by Eq. (4.7). The following procedure will be adopted: Instead of regarding this equation as implying that the universe is static we shall interpret it as informing us that the coordinate labels x^μ are really irrelevant. Physical significance can be ascribed only to the intrinsic dynamics of the world, and for the description of this we need some kind of intrinsic coordinatization based either on the geometry or the contents of the universe. In the case of infinite asymptotically flat worlds the Minkowski coordinates at infinity have independent physical relevance as preferred coordinates (up to a Lorentz transformation) based on an *a priori assumed* isometry group (the Poincaré group) for the asymptotic region. One may say that they are intrinsically determined by an implicit laboratory or observer at infinity, and that the constraints serve merely to eliminate the nonphysical modes from the field. In the case of finite worlds, however, the constraints are everything; they and they alone must yield the complete quantum-mechanical description of the world geometry. One of our tasks in the remainder of this paper will be to try to convince the reader that the equations of constraint really do *saturate* the theory, that nothing else is needed.

We must first establish the fact that the constraints are consistent with each other, and this raises some issues of factor ordering.[20] Unfortunately, general agreement has not yet been reached on how to resolve these issues, and hence the proposals which follow must be regarded as tentative. We emphasize, however, our view that the factor-ordering question is not very important to the theory as a whole, and should in no case be permitted to impede attempts to apply the theory to concrete problems. It arises in every local-field theory possessing nontrivial spectral functions, and bears mainly on problems of interpreting divergences. The latter are always resolved by symmetry arguments or by removing infinities from divergent integrals in an invariant way. How such procedures operate in the case of gravity will appear in the papers devoted to the manifestly covariant theory, where questions of factor ordering will again be discussed.

Consistency of the constraints is established if it can be shown that commutators of the constraints lead to no new constraints. The basic commutation relations of the canonical variables themselves are

$$[\alpha, \pi'] = i\delta(\mathbf{x},\mathbf{x}'), \quad [\beta_i, \pi^{j'}] = i\delta_i{}^{j'}, \quad [\gamma_{ij}, \pi^{k'l'}] = i\delta_{ij}{}^{k'l'}; \quad \text{all other commutators vanish;} \tag{4.10}$$

in which a notation is employed which emphasizes the bitensor transformation character of the quantities on the right, with primes being used, either on indices or on the variables themselves, to distinguish different points of 3-space. Here $\delta(\mathbf{x},\mathbf{x}')$ denotes the 3-dimensional δ function, and

$$\delta_i{}^{j'} \equiv \delta_i{}^j \delta(\mathbf{x},\mathbf{x}'), \quad \delta_{ij}{}^{k'l'} \equiv \delta_{ij}{}^{kl}\delta(\mathbf{x},\mathbf{x}'), \quad \delta_{ij}{}^{kl} \equiv \tfrac{1}{2}(\delta_i{}^k\delta_j{}^l + \delta_i{}^l\delta_j{}^k). \tag{4.11}$$

(The δ function will ordinarily be viewed as a bidensity of zero weight at its first argument and of unit weight at its second.)

The primary constraints evidently give no trouble, since they commute with each other and with the secondary constraints. We therefore turn to the latter and look first at the $\mathcal{H}$ constraints. These will be taken precisely as written in Eq. (3.6), with the momentum factor π^{jk} standing to the right. However, the index will be lowered by defining

$$\mathcal{H}_i \equiv \gamma_{ij}\mathcal{H}^j, \tag{4.12}$$

which, since γ_{ij} stands to the left, yields an alternative form for Eq. (4.4):

$$\mathcal{H}_i\Psi = 0. \tag{4.13}$$

$\mathcal{H}_i$ has the important property of being homogeneous bilinear in the γ_{ij} and the π^{ij}, with the γ's to the left and the π's to the right. Therefore its commutator with any other $\mathcal{H}_{j'}$ has the same property. To compute this commutator it is helpful first to compute the following:

$$\left[\gamma_{ij}, i\int \mathcal{H}_{k'}\delta\xi^{k'}d^3x'\right] = -\gamma_{ij,k}\delta\xi^k - \gamma_{kj}\delta\xi^k{}_{,i} - \gamma_{ik}\delta\xi^k{}_{,j}, \tag{4.14}$$

$$\left[\pi^{ij}, i\int \mathcal{H}_{k'}\delta\xi^{k'}d^3x'\right] = -(\pi^{ij}\delta\xi^k)_{,k} + \pi^{kj}\delta\xi^i{}_{,k} + \pi^{ik}\delta\xi^j{}_{,k}, \tag{4.15}$$

which reveal the $\mathcal{H}$'s as generators of 3-dimensional coordinate transformations. Under the infinitesimal coordinate transformation $\bar{x}^i = x^i + \delta\xi^i$, the change in any function of the γ_{ij}, π^{ij} and their derivatives is given by commutation with $i\int \mathcal{H}_i\delta\xi^i d^3x$, provided the function has no explicit dependence on x. From this it follows at once that

$$[\mathcal{H}_i, \mathcal{H}_{j'}] = -i\int \mathcal{H}_{k''}C^{k''}{}_{ij'}d^3x'', \tag{4.16}$$

[20] See, for example, J. L. Anderson, in *Proceedings of the 1962 Eastern Theoretical Conference*, edited by M. E. Rose (Gordon and Breach Science Publishers, Inc., New York, 1963), p. 387. See also J. Schwinger, Phys. Rev. **130**, 1253 (1963); **132**, 1317 (1963).

where the c's are the structure constants of the general coordinate transformation group:

$$c^{k''}{}_{ij'} \equiv \delta^{k''}{}_{i,l'}\delta^{l''}{}_{j'} - \delta^{k''}{}_{j',l'}\delta^{l''}{}_{i}. \tag{4.17}$$

The same observations, combined with the fact that $\mathcal{H}$ is a scalar density, yield the formula

$$[\chi_i, \mathcal{H}'] = i\mathcal{H}\delta_{,i}(\mathbf{x},\mathbf{x}'). \tag{4.18}$$

Again the ordering of factors remains the same on both sides of the equation. The only term of $\mathcal{H}$ which might lead to difficulty is the one quadratic in the momenta. But all of the factors which appear in this term have homogeneous linear transformation laws under the 3-dimensional coordinate transformation group and hence remain undisturbed in position when commuted with χ_i. Thus, the commutators (4.16) and (4.18) yield no new constraints, and the choice of factor ordering for $\mathcal{H}$ is so far arbitrary.

Now note that all of the above results could have been obtained equally well had the opposite ordering been chosen for χ_i, with the π's standing to the left and the γ's to the right. The difference between the two choices for χ_i therefore commutes with everything and is evidently a c number. It is a c number, moreover, with definite transformation properties; namely, it is a covariant 3-vector density. From this we may conclude that it can only be zero, for otherwise 3-space would contain a preferred direction quite independently of any geometry which may be imposed on it. The reasonableness of this conclusion also follows from a straightforward formal computation of the difference between the two χ's, which yields derivatives of δ functions with coincident arguments. Any ordering may therefore be chosen for χ_i, and if γ_{ij} and π^{ij} are Hermitian so is χ_i.

The same conclusions do not automatically hold for χ^i, since the difference between two orderings for it involves an undifferentiated δ function. Let us therefore see what we can say about the formal symbol $\delta(\mathbf{x},\mathbf{x})$. Consider the third commutator in (4.10). If we set $\mathbf{x}'=\mathbf{x}$ and contract all the indices, we obtain

$$(\gamma_{ij}\pi^{ij} - \pi^{ij}\gamma_{ij}) = 6i\delta(\mathbf{x},\mathbf{x}). \tag{4.19}$$

The quantity on the right is certainly a c number. Therefore we may write

$$[6i\delta(\mathbf{x},\mathbf{x}), i\int \chi_{k'}\delta\xi^{k'}d^3x'] = 0. \tag{4.20}$$

On the other hand, if we apply the same commutator to the left we obtain

$$[(\gamma_{ij}\pi^{ij} - \pi^{ij}\gamma_{ij}), i\int \chi_{k'}\delta\xi^{k'}d^3x']$$

$$= -[(\gamma_{ij}\pi^{ij} - \pi^{ij}\gamma_{ij})\delta\xi^{k}]_{,k} = -6i[\delta(\mathbf{x},\mathbf{x})\delta\xi^{k}]_{,k}. \tag{4.21}$$

Equating the two results we find

$$[\delta(\mathbf{x},\mathbf{x})\delta\xi^{i}]_{,i} = 0. \tag{4.22}$$

This equation must hold for arbitrary $\delta\xi^i$. Therefore, although most people would say that $\delta(\mathbf{x},\mathbf{x})$ is infinite, we see that it is actually zero.

In order to understand how this formal result can be consistent with the rest of the theory one must first note that Eqs. (4.3) and (4.13) are really abbreviations for the correct forms

$$\int \mathcal{H}\xi d^3x\ \Psi = 0 \quad \text{for all } \xi, \tag{4.23}$$

$$\int \chi_i\xi^i d^3x\ \Psi = 0 \quad \text{for all } \xi^i, \tag{4.24}$$

where ξ and ξ^i are arbitrary but *smooth* c-number weight functions. The problem of taking commutators of field quantities at the same space-time point therefore never arises with pairs of constraints but only in connection with the definition of the functions $\mathcal{H}$ and χ_i themselves. This means that the δ function may, without inconsistency, be thought of as the limit of a sequence of successively narrower *twin-peaked* functions, all of which are smooth, have unit integral, and vanish at the point $\mathbf{x}'=\mathbf{x}$ in the valley between the peaks. An example of such a function in one dimension would be $\delta(x) = \lim(2\pi)^{-1}[f_\epsilon(x-\sqrt{\epsilon}) + f_\epsilon(x+\sqrt{\epsilon}) - 2f_\epsilon(x)/(1+\epsilon)]$, where $f_\epsilon(x) \equiv \epsilon(x^2+\epsilon^2)^{-1}$. In an infinite world, passage to the limit $\epsilon \to 0$ would correspond to the usual cutoff going to infinity in momentum space, while maintenance of the valley at $\mathbf{x}'=\mathbf{x}$ would yield a particular regularization of the resulting divergences. The answer to the question whether or not this regularization is equivalent to the quite different procedures which will prove useful in the manifestly covariant theory must await a demonstration of how to derive one theory from the other. In the meantime we shall in *this* paper simply adopt it as a rule that any two field operators taken at the same space-time point commute. The consistency question for the constratints then reduces to that of the classical theory.

There remains to be considered only the commutator $[\mathcal{H},\mathcal{H}']$. At first sight it might be thought that the commutator of the two quadratic-in-the-momenta terms, one from $\mathcal{H}$ and the other from $\mathcal{H}'$, leads to difficulties. However, these terms contain no derivatives (of the γ's or π's) with respect to the 3-space coordinates and hence they commute. Since the terms $\gamma^{1/2}\,{}^{(3)}R$ and $\gamma'^{1/2}\,{}^{(3)}R'$ contain no momenta, they likewise commute. The only commutators which remain are the cross commutators, and these can be evaluated by judicious use of the variational formula

$$\delta(\gamma^{1/2}\,{}^{(3)}R) = \gamma^{1/2}\gamma^{ij}\gamma^{kl}(\delta\gamma_{ik.jl} - \delta\gamma_{ij.kl}) - \gamma^{1/2}({}^{(3)}R^{ij} - \tfrac{1}{2}\gamma^{ij}\,{}^{(3)}R)\delta\gamma_{ij}. \tag{4.25}$$

The final result is

$$[\mathcal{H},\mathcal{H}'] = 2i\chi^i\delta_{,i}(\mathbf{x},\mathbf{x}') + i\chi^i{}_{,i}\delta(\mathbf{x},\mathbf{x}'), \tag{4.26a}$$

or, more correctly,

$$\left[\int \mathcal{H}\xi_1 d^3x, \int \mathcal{H}\xi_2 d^3x\right] = i\int \chi^i(\xi_1\xi_{2,i} - \xi_{1,i}\xi_2)d^3x. \quad (4.26b)$$

If we were still concerned about the order of factors we would find that a symmetric (Hermitian) ordering for $\mathcal{H}$ would yield a symmetric ordering for χ^i in (4.26), namely $\chi^i = \frac{1}{2}\{\gamma^{ij}, \chi_j\}$, and the problem at issue would then be to evaluate the commutator $[\gamma^{ij}, \chi_j]$. From our present point of view this commutator vanishes, and consistency is maintained.[21]

5. THE FUNCTIONAL WAVE EQUATION AND THE STATE-FUNCTIONAL DOMAIN MANIFOLD

Further analysis of the canonical theory requires the introduction of a specific representation for the quantum states. Wheeler[11] has chosen for this purpose what may be called the *metric representation*, in which Ψ becomes a functional of the metric components $g_{\mu\nu}$, and the momenta become functional differential operators:

$$\pi = \frac{\delta}{i\delta\alpha}, \quad \pi^i = \frac{\delta}{i\delta\beta_i}, \quad \pi^{ij} = \frac{\delta}{i\delta\gamma_{ij}}. \quad (5.1)$$

The primary constraints tell us that Wheeler's Ψ depends only on the γ's. We shall indicate this, for the present, by writing Ψ in the form $\Psi[\gamma]$. (Since we are working in a closed finite world, it would be meaningless to include also a dependence on x^0.)

Consider now the χ constraints. In the metric representation these take the form

$$2i(\delta\Psi[\gamma]/\delta\gamma_{ij})_{,j} = 0, \quad (5.2)$$

which are the necessary and sufficient conditions that $\Psi[\gamma]$ be an invariant under coordinate transformations. In a finite world this means that Ψ depends only on the geometry of 3-space. One possible way to express this dependence would be to regard Ψ as a function of a discrete infinity of variables, namely all the independent invariants, beginning with $\int\gamma^{1/2}d^3x$, $\int\gamma^{1/2}\,{}^{(3)}Rd^3x$, $\int\gamma^{1/2}\,{}^{(3)}R^2d^3x$, etc., which can be constructed out of products of the Riemann tensor and its covariant derivatives, with the topology of 3-space itself being separately specified.

Higgs[8] has pointed out that in an infinite world such a characterization of Ψ would be inadequate, for in this case the asymptotic coordinates also play a role. Ψ could instead be represented as a functional of any three of the six coordinate-invariant functions[22]:

$$\varphi^{AB}(\eta) \equiv \int \gamma^{1/2}\gamma^{ij}\zeta^A{}_{,i}\zeta^B{}_{,j}\delta^3(\eta - \zeta(\mathbf{x}))d^3x, \qquad A, B = 1, 2, 3, \quad (5.3)$$

where the ζ's are scalars satisfying the elliptic differential equation

$$\zeta^A{}_{,i}{}^{,i} = 0, \quad (5.4)$$

with the boundary conditions $\zeta^A \to x^A$ at infinity. The ζ's define a harmonic coordinate system, and Eqs. (5.4) yield $\partial\varphi^{AB}/\partial\eta^B = 0$ as a corollary. If φ^{11}, φ^{12}, φ^{22} are arbitrarily chosen then φ^{13}, φ^{23}, φ^{33} are determined by integrating successively the equations $\partial\varphi^{13}/\partial\eta^3 = -\partial\varphi^{11}/\partial\eta^1 - \partial\varphi^{12}/\partial\eta^2$, $\partial\varphi^{23}/\partial\eta^3 = -\partial\varphi^{12}/\partial\eta^1 - \partial\varphi^{22}/\partial\eta^2$, $\partial\varphi^{33}/\partial\eta^3 = -\partial\varphi^{13}/\partial\eta^1 - \partial\varphi^{23}/\partial\eta^2$. If space-time is asymptotically flat and the coordinates x^i are Minkowskian at infinity, then these equations can be consistently integrated with the asymptotic boundary conditions $\varphi^{AB} \to \delta_{AB}$.

The above example is cited in order to re-emphasize the fundamental difference between finite and infinite worlds. In the finite case we may replace the symbol $\Psi[\gamma]$ by $\Psi[{}^{(3)}\mathcal{G}]$ to display the fact that Ψ depends only on the 3-geometry, denoted here by ${}^{(3)}\mathcal{G}$, and on nothing else, whereas in the infinite case we must write something like $\Psi[{}^{(3)}\mathcal{G}, \mathfrak{L}]$, with $\mathfrak{L}$ symbolizing the surrounding laboratory which determines the asymptotic coordinates (including, in the Schrödinger picture, the coordinate x^0).

We shall denote by $\mathfrak{M}$ the set of all possible 3-geometries which a finite world may possess. The following question will arise: Can a topology be imposed upon $\mathfrak{M}$ which is both meaningful and at the same time useful in the context of the quantum theory of finite worlds? One possibility which suggests itself is to view $\mathfrak{M}$ as an infinite-dimensional vector space whose "points" are discrete sets of invariants mentioned earlier. The topology could be that defined by the Cartesian metric on this space, and the symbol ${}^{(3)}\mathcal{G}$ could be replaced by a set of vector components. In fact, this possibility is not very useful, and although we shall actively pursue the question of assigning a metric, and indeed a pseudo-Riemannian structure, to $\mathfrak{M}$, no advantage will be gained by attempting to make our symbolism more explicit. It will be sufficient simply to keep in mind the idea that $\mathfrak{M}$ is not just a mere set but is actually a

[21] J. Schwinger (Ref. 20) proposes an alternative resolution of the factor ordering problem which, in the notation of the present paper, runs essentially as follows: Replace $\mathcal{H}$ in the Hamiltonian constraint by $(\gamma^{3/2}\mathcal{H})$, where ( ) indicates that the factors are to be placed in some (arbitrary) symmetrical order. Then compute

$$[(\gamma^{3/2}\mathcal{H}), (\gamma'^{3/2}\mathcal{H}')] = \tfrac{1}{2}i[\{\gamma^3\gamma^{ij}, (\chi_j)\} + \{\gamma'^3\gamma'^{i'j'}, (\chi_{j'})\}]\delta_{,i}(\mathbf{x},\mathbf{x}')$$
$$= \tfrac{1}{2}i[\{\gamma^3\gamma^{ij}, (\chi_{j'})\} + \{\gamma'^3\gamma'^{i'j'}, (\chi_j)\}]\delta_{,i}(\mathbf{x},\mathbf{x}').$$

Since the commutator

$$[\gamma^3\gamma^{ij}, (\chi_{j'})] = i(\gamma^3\gamma^{ij} + \gamma'^3\gamma'^{i'j'})\delta_{,j}(\mathbf{x},\mathbf{x}')$$

is antisymmetric in $\mathbf{x}$ and $\mathbf{x}'$, it follows that

$$[\gamma^3\gamma^{ij}, (\chi_{j'})] + [\gamma'^3\gamma'^{i'j'}, (\chi_j)] = 0,$$

whence

$[(\gamma^{3/2}\mathcal{H}), (\gamma'^{3/2}\mathcal{H}')] = i[\gamma^3\gamma^{ij}(\chi_{j'}) + \gamma'^3\gamma'^{i'j'}(\chi_j)]\delta_{,i}(\mathbf{x},\mathbf{x}')$ in which the χ's stand to the right. Demonstration of consistency of the other commutators is elementary.

[22] In the Schrödinger picture Ψ would also depend on x^0.

manifold. Thus we shall say: $\mathfrak{M}$ is the *domain manifold* for the state functional Ψ, and the ${}^{(3)}\mathcal{G}$ are its "points."

So far nothing has been said about dynamics. The only way in which dynamics can enter the picture is through the Hamiltonian constraint. This now takes the form

$$\left(G_{ijkl}\frac{\delta}{\delta\gamma_{ij}}\frac{\delta}{\delta\gamma_{kl}}+\gamma^{1/2}\,{}^{(3)}R\right)\Psi[{}^{(3)}\mathcal{G}]=0, \tag{5.5}$$

where

$$G_{ijkl}\equiv\tfrac{1}{2}\gamma^{-1/2}(\gamma_{ik}\gamma_{jl}+\gamma_{il}\gamma_{jk}-\gamma_{ij}\gamma_{kl}). \tag{5.6}$$

According to our rule of freely commuting field operators taken at the same space-time point, the functional differential operator $\delta/\delta\gamma_{ij}$ must always be understood to give zero when acting on a γ_{kl} at the same point.[23] If it were not for this rule, we might try to regard the first term in the parentheses of (5.5) as a kind of Laplace-Beltrami operator in a 6-dimensional Riemannian manifold having G_{ijkl} as its *contra*variant metric. Although such an interpretation is inappropriate for the operator itself, it *is* nevertheless useful to regard G_{ijkl} as a metric tensor and to study the properties of the manifold which it defines. These properties, which are derived in Appendix A, turn out to be quite interesting.

The manifold in question will be denoted by M. When γ_{ij} is positive definite (as it is for a spacelike hypersurface) M has the hyperbolic signature $-+++++$. A "pure dilation" of γ_{ij} (i.e., multiplication by a multiple of the unit matrix) constitutes a typical "timelike" displacement. It is convenient to introduce the timelike coordinate

$$\zeta\equiv(32/3)^{1/2}\gamma^{1/4} \tag{5.7}$$

and any five other coordinates ζ^A orthogonal to it. The covariant metric then takes the form

$$\begin{pmatrix}-1 & 0\\ 0 & (3/32)\zeta^2\bar{G}_{AB}\end{pmatrix}, \tag{5.8}$$

where

$$\bar{G}_{AB}\equiv\mathrm{tr}(\boldsymbol{\gamma}^{-1}\boldsymbol{\gamma}_{,A}\boldsymbol{\gamma}^{-1}\boldsymbol{\gamma}_{,B}), \tag{5.9}$$

$$\boldsymbol{\gamma}\equiv(\gamma_{ij}). \tag{5.10}$$

Expression (5.8) reveals M as a set of "nested" 5-dimensional submanifolds, all having the same intrinsic shape and differing only in the scale factor $(3/32)\zeta^2$. The shape is described by the positive-definite metric $\bar{G}_{AB}$ which, since expression (5.9) remains invariant under a dilation of the $\boldsymbol{\gamma}$'s, is independent of ζ.

The manifold having $\bar{G}_{AB}$ as a metric will be denoted by $\bar{M}$. It is shown in Appendix A that the geodesic equation in $\bar{M}$ takes the form

$$\frac{d^2\boldsymbol{\gamma}}{d\bar{s}^2}-\frac{d\boldsymbol{\gamma}}{d\bar{s}}\boldsymbol{\gamma}^{-1}\frac{d\boldsymbol{\gamma}}{d\bar{s}}=0,\quad \mathrm{tr}\left(\boldsymbol{\gamma}^{-1}\frac{d\boldsymbol{\gamma}}{d\bar{s}}\right)=0. \tag{5.11}$$

This has the general solution[24]

$$\boldsymbol{\gamma}(\bar{s})=\mathbf{M}^{\sim}e^{\mathbf{N}\bar{s}}\mathbf{M}, \tag{5.12}$$

where $\mathbf{M}$ is an arbitrary nonsingular 3×3 matrix and $\mathbf{N}$ is subject only to the restrictions

$$\mathbf{N}^{\sim}=\mathbf{N},\quad \mathrm{tr}\mathbf{N}=0,\quad \mathrm{tr}\mathbf{N}^2=1, \tag{5.13}$$

the last of which guarantees that $\bar{s}$ is the arc length. Since $e^{\mathbf{N}\bar{s}}$ is analytic for all values of $\bar{s}$, $\bar{M}$ is geodesically complete. It is not difficult to show that any two points of $\bar{M}$ may be joined by a unique geodesic and that if the two points are represented by symmetric matrices $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ having the same determinant then their distance of separation is $\{\mathrm{tr}[\ln(\boldsymbol{\gamma}_1^{-1}\boldsymbol{\gamma}_2)]^2\}^{1/2}$. The manifold $\bar{M}$ is evidently noncompact and diffeomorphic to Euclidean 5-space.

By straightforward computation one may verify that the Riemann and Ricci tensors of $\bar{M}$ have the respective forms

$$\bar{R}_{ABCD}=\mathrm{tr}[\boldsymbol{\gamma}^{-1}\boldsymbol{\gamma}_{,D}\boldsymbol{\gamma}^{-1}\boldsymbol{\gamma}_{,C}\boldsymbol{\gamma}^{-1}\times(\boldsymbol{\gamma}_{,A}\boldsymbol{\gamma}^{-1}\boldsymbol{\gamma}_{,B}-\boldsymbol{\gamma}_{,B}\boldsymbol{\gamma}^{-1}\boldsymbol{\gamma}_{,A})], \tag{5.14}$$

$$\bar{R}_{AB}=-\tfrac{3}{2}\bar{G}_{AB}. \tag{5.15}$$

From the latter it follows that $\bar{M}$ is an "Einstein space" of constant negative Gaussian curvature. It is furthermore not difficult to show that the Riemann tensor (5.14) has vanishing covariant derivative, which implies that $\bar{M}$ is, in fact, *a symmetric space*[25] with a certain group structure. The group structure may be deduced from the observation that the transformation

$$\boldsymbol{\gamma}'=\mathbf{L}^{\sim}\boldsymbol{\gamma}\mathbf{L}, \tag{5.16}$$

where $\mathbf{L}$ is an arbitrary constant nonsingular 3×3 matrix, leaves the metric (5.9) unchanged. The full linear group in three dimensions therefore acts isometrically on $\bar{M}$. Because of the dilation invariance of the points of $\bar{M}$, however, it is only the simple Lie Group $SL(3,R)$ which acts *effectively* on it. It is easily verified that $SL(3,R)$ acts transitively on $\bar{M}$ and, moreover, that the *isotropy subgroup*[25] at any point is isomorphic to $SO(3)$. $\bar{M}$ may therefore be identified as the coset space

$$\bar{M}=SL(3,R)/SO(3). \tag{5.17}$$

Although the manifold $\bar{M}$ is geodesically complete, the manifold M is not. It is shown in Appendix A that all geodesics in M ultimately hit a *frontier* of infinite

[23] There is nothing automatically pathological, however, about having two functional derivatives acting at the same point, as in (5.5). For example, if $I\equiv\frac{1}{2}\int dx\int dx'\varphi(x)K(x,x')\varphi(x')$, where φ is an arbitrary function and K is a fixed kernel, then $\delta^2I/\delta\varphi(x)\delta\varphi(x)=K(x,x)$. Pathology occurs only if $K(x,x')$ is singular at $x'=x$.

[24] The tilde "$\sim$" denotes the transpose. All matrices are assumed real.

[25] See, for example, S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic Press Inc., New York, 1962). The author is indebted to Professor Helgason for enlightment as to the group structure of $\bar{M}$.

curvature. "Timelike" and null geodesics hit it at one end; "spacelike" geodesics hit it at both ends. This frontier, which will be denoted by F, is located at $\zeta=0$, as may be inferred from the readily computed curvature scalar

$$^{(6)}R=-60/\zeta^2. \tag{5.18}$$

A question now arises as to what estent the Riemannian structure of M may be regarded as imposing a structure on $\mathfrak{M}$ by way of the Hamiltonian constraint. Without attempting to answer this question directly, we may point out certain very suggestive features of the theory. First of all, the existence of the timelike coordinate ζ in M suggests that a corresponding "intrinsic time" exists in $\mathfrak{M}$ and that the Hamiltonian constraint does indeed have dynamical content. This idea is given support by the following considerations: The specification of a given 3-geometry requires the assignment of essentially 3 independent quantities at each point of 3-space. If we regard the usual enumeration of the degrees of freedom possessed by the gravitational field, namely *two* for every point of 3-space, as being valid in a finite world, this leaves one quantity per 3-space point to play the role of intrinsic time. Baierlein, Sharp, and Wheeler[11,26] have shown in the classical theory that if the intrinsic geometry is given on any two hypersurfaces then, except in certain singular cases, the geometry of the entire space-time manifold, *and hence the absolute time lapse between the two hypersurfaces*, is determined. Moreover, it is determined solely by the constraints. Analogously, the quantum theory is completely determined by the transformation functional $\langle{}^{(3)}\mathcal{G}'|{}^{(3)}\mathcal{G}''\rangle$, where $|{}^{(3)}\mathcal{G}\rangle$ denotes that state of the gravitational field for which there exists at least one hypersurface having an infinitely precise geometry ${}^{(3)}\mathcal{G}$. Wheeler[11] has emphasized the imporatnce of the two-hypersurface formulation of gravidynamics (or "geometrodynamics" as he calls it) and has suggested the use of the Feynman sum-over-histories method to compute the transformation functional.[27]

Another suggestive feature of the theory is the following. Because of the hyperbolic character of M the Hamiltonian constraint (5.5) resembles a Klein-Gordon equation, with $-\gamma^{1/2}\,{}^{(3)}R$ playing the role of the mass-squared term. An important difference, however, is that ${}^{(3)}R$ can be either positive or negative, and hence the "wave" propagation of the state functional is not confined to timelike directions.

In spite of this difference the analogy with the Klein-Gordon theory suggests the following definition for the quantum-mechanical inner product of two states Ψ_a and Ψ_b:

$$(\Psi_b,\Psi_a)=Z\int_\Sigma \Psi_b{}^*[{}^{(3)}\mathcal{G}]$$
$$\times\prod_{\mathbf{x}}\left(d\Sigma^{ij}G_{ijkl}\frac{\vec{\delta}}{i\delta\gamma_{kl}}-\frac{\overleftarrow{\delta}}{i\delta\gamma_{kl}}G_{ijkl}d\Sigma^{ij}\right)\Psi_a[{}^{(3)}\mathcal{G}]. \tag{5.19}$$

The infinite product, which arises because (5.5) is really not just one equation but ∞^3 equations, is here taken over all the points of 3-space, and is to be understood in a formal sense as representing the result of a limiting process based on a sequence of lattices in 3-space, each lattice requiring the introduction of a corresponding normalizing constant Z. The symbol Σ denotes the topological product of a set of 5-dimensional M-hypersurfaces $\Sigma(\mathbf{x})$ (one chosen at each point of 3-space), the $d\Sigma^{ij}$ being their directed surface elements. It is an immediate consequence of the Hamiltonian constraint that this inner product is independent of the choice of $\Sigma(\mathbf{x})$'s provided some kind of appropriate boundary conditions are satisfied at the "edges" of Σ. It is also worth noting that since the G_{ijkl} do not involve any spatially differentiated γ's, the operators standing in the infinite product all commute, and hence no factor-ordering difficulties arise here.

In view of the coordinate invariance of the state functionals the inner product integral (5.19) contains a $3\times\infty^3$-fold redundancy arising from the geometrical indistinguishability of 3-metrics which differ only by coordinate transformations.[28] This produces a divergence which must be formally absorbed into the normalization constant Z, and reminds us that $\mathfrak{M}$ is not just the topological product of M with itself over all the points of 3-space, but is a subspace of the latter manifold.

Another difficulty with the definition (5.19) concerns the problem of "negative probability." This problem arises here, just as it does for the Klein-Gordon equation, from the fact that the Hamiltonian constraint involves a second derivative with respect to the "time" coordinate. If the $\Sigma(\mathbf{x})$'s are chosen "spacelike," then the only way to assure positive definiteness of (5.19), when $\Psi_b=\Psi_a$, is to restrict the content of Ψ_a to "positive frequency" components with respect to every "time" coordinate $\zeta(\mathbf{x})$. Restriction to such components, however, implies that Ψ_a vanishes nowhere in the range $-\infty<\zeta<\infty$, and this conflicts with the one-sided character of ζ, namely $\zeta>0$, which follows from the geometrical analysis revealing the existence of a frontier in M at $\zeta=0$. One might hope that an analytic continuation could be performed around $\zeta=0$, but

[26] R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962). There is nothing mysterious about the existence of a manifold of "time" variables. The same manifold exists in conventional field theory in those formulations which make the state functional depend on an arbitrary spacelike hypersurface.

[27] The sum-over-histories or "functional integral" method has not yet been applied to any "practical" problem of quantum gravidynamics. It will be encountered in heuristic and formal applications in the following papers of this series. Its consistency with the Dirac theory has been demonstrated by Leutwyler. [See H. Leutwyler, Phys. Rev. **134**, B1155 (1964).]

[28] A coordinate transformation generally produces a change in Σ, but this does not affect the integral.

whether this would have any physical meaning is unclear. The singularity in the Hamiltonian constraint at $\zeta=0$ is a strong one, as may be seen by rewriting (5.5) in the form

$$\left[-\frac{\delta^2}{\delta\zeta^2}+\frac{(32/3)}{\zeta^2}\bar{G}^{AB}\frac{\delta^2}{\delta\zeta^A\delta\zeta^B}+(3/32)\zeta^2\,{}^{(3)}R\right]\times\Psi[{}^{(3)}\mathcal{G}]=0, \quad (5.20)$$

which makes use of (5.7) and (5.8). The question at issue is whether the frontier in M generates a corresponding barrier in $\mathfrak{M}$ beyond which there is no possibility of extending the state functional. Unfortunately, in the present state of our knowledge no clear-cut answer can be given to this question. Some of the problems which have a bearing on it, however, can be identified. These will now be discussed.

6. THE METRIC OF $\mathfrak{M}$. THE HAMILTON-JACOBI EQUATION AND GRAVITATIONAL COLLAPSE

The most obvious way to approach $\mathfrak{M}$ is through the manifold M^{∞^3}, which is defined formally as the topological product of M with itself over the points of 3-space:

$$M^{\infty^3}\equiv\prod_{\mathbf{x}} M(\mathbf{x}). \quad (6.1)$$

The "points" of M^{∞^3} are the matrix functions $\gamma_{ij}(\mathbf{x})$. For brevity they will be denoted simply by γ. In practice the definition (6.1) must be supplemented by some sort of continuity requirements. For example, γ may be required to be continuous and piecewise differentiable. However, we do not wish to be precise about this here, since as yet no rigorous theory of the role of the manifold $\mathfrak{M}$ in the quantum theory exists. We wish merely to point out some of the issues involved, and to leave the formalism itself as unencumbered as possible. Thus we shall be willing to admit any sort of pathology for γ which we can get away with, i.e., for which some sort of physical interpretation exists, however idealized, which permits γ to be handled in a consistent fashion. For example, geometrical singularities at which the Riemann tensor behaves like a differentiated δ function, or for which integrals like $\int\gamma^{1/2}\,{}^{(3)}R dx$ still exist, will not be excluded *a priori*. In the same spirit, we shall not place any restrictions on coordinate transformations beyond perhaps requiring them to be differentiable (so that tensor transformation laws exist almost everywhere) and one-to-one (so that they form a group). Thus we shall not automatically exclude transformations for which the Jacobian either vanishes or diverges at certain points. The ultimate question will always be: What is the *barrier* beyond which we cannot go? In every case this will probably depend, to some extent at least, on the context, and we do not wish to prejudice the answer in advance.

There is, however, one trivial pathology which may may be avoided without loss of generality, namely, coordinate singularities which arise from the impossibility of covering compact manifolds with a single well-behaved coordinate system. We shall always assume that 3-space is covered with a finite set of overlapping coordinate patches, each of which can be put into one-to-one correspondence with a certain portion of the Cartesian mesh in Euclidean 3-space, and on the boundaries of which the coordinates are held fixed. In addition a set of supplementary connection formulas between patches must be assuemd to hold in the overlap regions. All of this paraphernalia is to be understood as included in the definition (6.1), which means that each function $\gamma_{ij}(\mathbf{x})$ is really a set of functions, one in each coordinate patch, and that the $\mathbf{x}$'s in Eq. (6.1) are to be understood as ranging over all values in all patches.

Now let γ to be a fixed point of M^{∞^3}. Consider the set of all points which may reached from γ by coordinate transformations. This set is known as the *orbit* of γ under the coordinate transformation group and will be denoted by "orb γ." There is a one-to-one correspondence between the orbits in M^{∞^3} and the points of $\mathfrak{M}$. In fact no generality is lost if they are identified:

$$\text{orb}\,\gamma\equiv{}^{(3)}\mathcal{G}. \quad (6.2)$$

Suppose M^{∞^3} is endowed with a metric. (That this is feasible will appear in a moment.) If this metric satisfies a certain condition then it will impose, in a natural way, a metric on $\mathfrak{M}$. The condition is that the coordinate transformation group in 3-space be an *isometry group* of M^{∞^3}, The associated metric in $\mathfrak{M}$ is then obtained by defining the distance between two neighboring orbits to be the shortest distance in M^{∞^3}.

It is shown in Appendix B that the above condition is satisfied if and only if the metric in M^{∞^3} transforms, under 3-dimensional coordinate transformations, contragrediently to the Kronecker product $\gamma_{ij}\gamma_{k'l'}$. This means that the metric in M^{∞^3}, which we shall denote by $\mathcal{G}^{ijk'l'}$, must be a contravariant bitensor density of weight at both x and x'.

There are infinitely many contravariant bitensor densities which can be constructed out of the γ's and which might serve as acceptable metrices for M^{∞^3}. Of these, however, there is only a single one-parameter family which is *local*, i.e., which involves only undifferentiated γ's and for which both $\mathcal{G}^{ijk'l'}$ and its inverse vanish when $x\neq x'$. This family is given by

$$\mathcal{G}^{ijk'l'}=\tfrac{1}{2}\gamma^{1/2}(\gamma^{ik}\gamma^{jl}+\gamma^{il}\gamma^{jk}+\lambda\gamma^{ij}\gamma^{kl})\delta(\mathbf{x},\mathbf{x}'), \quad (6.3)$$

where λ can assume any real value except $-\frac{2}{3}$. If we wished to impose a positive-definite metric on $\mathfrak{M}$, so that we could use, as the condition for the identity of two 3-geometries, the vanishing of the "distance" between them, then the metric of M^{∞^3} itself would have to be positive definite. In the present case this requires $\lambda>-\frac{2}{3}$, the simplest choice being $\lambda=0$. On the other

hand, the choice $\lambda=-2$ is the natural choice if we assume that Eq. (6.1) defines not merely a topological product but also a geometrical structure generated by the original metric on M. For then we have

$$\int \mathcal{G}_{ija''b''}\mathcal{G}^{a''b''k'l'}d^3x''=\delta_{ij}{}^{k'l'}, \tag{6.4}$$

where

$$\mathcal{G}_{ijk'l'}\equiv G_{ijkl}\delta(\mathbf{x},\mathbf{x}'), \tag{6.5}$$

with G_{ijkl} given by Eq. (5.6). In this case the "arc length" $d\$$ associated with a displacement $d\gamma_{ij}$ in $M^{\infty 3}$ is given by

$$d\$^2=\int d^3x\int d^3x'\,\mathcal{G}^{ijk'l'}d\gamma_{ij}d\gamma_{k'l'}$$

$$=\int \gamma^{1/2}(\gamma^{ik}\gamma^{jl}-\gamma^{ij}\gamma^{kl})d\gamma_{ij}d\gamma_{kl}d^3x. \tag{6.6}$$

Geodesics in $M^{\infty 3}$ are not, in general, geodesics in $\mathfrak{M}$. However, in Appendix B it is shown that if a geodesic in $M^{\infty 3}$ intersects one of the orbits in its path orthogonally then it *is* a geodesic in $\mathfrak{M}$, and, moreover, it intersects every other orbit in its path orthogonally. This means that it is in principle possible to use formula (A69) of the Appendix to determine the distance between two 3-geometries. In practice, of course, the amount of labor involved is formidable, assuming that the 3-geometries are given in the form of two matrix functions $\gamma_1(\mathbf{x})$ and $\gamma_2(\mathbf{x})$. One must integrate expression (A69) over 3-space and then find the minimum of the integral as one of the functions, say $\gamma_1(\mathbf{x})$, is held fixed while the other ranges over the various equivalent forms it can take under coordinate transformations. This means solving the complicated set of nonlinear partial differential equations which result from the corresponding variational principle and which, in effect, yield the coordinate transformation which "lines up" $\gamma_1(\mathbf{x})$ and $\gamma_2(\mathbf{x})$, so that a geodesic from one to the other intersects orbits orthogonally.

Such complications can be avoided if one merely wants to know the distance from a given 3-geometry to the *frontier*.[29] In this case, since the frontier is an extended object and is at different distances—spacelike, timelike, and null—in different directions, it is necessary to specify a direction $d\gamma_{ij}/d\$$ in addition to the 3-geometry itself. Here $\$$ is either the arc length (6.6) or, in the exceptional null case, an affine parameter, and $d\gamma_{ij}/d\$$ must satisfy the starting condition [cf. Eq. (B24)]

$$\gamma^{jk}\left[\left(\frac{d\gamma_{ij}}{d\$}\right)_{.k}-\left(\frac{d\gamma_{jk}}{d\$}\right)_{.i}\right]=0, \tag{6.7}$$

which guarantees that the starting direction will be orthogonal to the starting orbit. The square of the distance in the frontier in the assigned direction is then given by[30]

$$\$^2=\min 2\sigma(\gamma,d\gamma/d\$)\left[G^{ijkl}\frac{d\gamma_{ij}}{d\$}\frac{d\gamma_{kl}}{d\$}\right]^{-1}, \tag{6.8}$$

where G^{ijkl} and σ are defined in Appendix A, Eqs. (A1) and (A63), respectively, and "min" denotes the minimum value over 3-space. The metric tensor at the point (or points) in 3-space at which the minimum occurs will become singular when the "point" γ in $M^{\infty 3}$ has progressed a distance $\$$ along the geodesic. The geodesic can then go no further without changing the signature of a portion of 3-space. The frontier has been reached.

It does not automatically follow that 3-space acquires a *geometrical* singularity at the frontier. However, there are several facts worth noting.

(1) The occurrence of a singular metric cannot be avoided by changing the coordinates as one proceeds along the geodesic. Although it is true that a coordinate transformation can carry one from one point to another in $M^{\infty 3}$ and even, *seemingly*, away from the frontier, yet since expression (6.8) is a scalar, $\$$ remains unchanged. What happens is that the coordinate transformation also changes the direction $d\gamma_{ij}/d\$$. Moreover, the covariance of Eq. (6.7) ensures that the orthogonality of the geodesic to the orbits is left unaffected.

(2) As long as no coordinate transformations are performed while γ is moving along its orthogonal geodesic, the coordinate system in 3-space, if initially nonsingular, will remain nonsingular until the frontier is reached. No such statement can be made for nonorthogonal geodesics, which in some cases follow a circuitous route in $\mathfrak{M}$ from a given ${}^{(3)}\mathcal{G}$ back again to the same ${}^{(3)}\mathcal{G}$, but in a different coordinate system.[31]

(3) It is not necessary that the metric become singular simultaneously at all points of 3-space in order that

[29] By *frontier* we do not necessarily mean *barrier*. It must be repeatedly emphasized that very little is known about the general conditions under which extensions beyond the frontier can be carried out. Here we are defining the frontier to be simply the locus of points γ in $M^{\infty 3}$ for which the matrix $\gamma_{ij}(\mathbf{x})$ has one or more singularity points ($\gamma=0$) in 3-space, regardless of whether or not these singularity points represent real geometrical singularities. A formal definition would be

$$\mathbf{F}\equiv \bigcup_{\mathbf{x}} F(\mathbf{x}) \prod_{\mathbf{x}'\neq\mathbf{x}} M(\mathbf{x}'),$$

where $\prod$ and $\bigcup$ denote the topological product and union, respectively, and $F(\mathbf{x})$ is the frontier of $M(\mathbf{x})$.

[30] When $d\gamma_{ij}/d\$$ is a null vector in M, expression (6.8) becomes an indeterminate form 0/0. At such points in 3-space (6.8) may, in view of Eq. (A54) which implies $\$=\text{constant}\times\gamma^{1/2}$, be replaced simply by $\$^2=4(\gamma^{ij}d\gamma_{ij}/d\$)^{-2}$.

[31] In attempting to visualize $M^{\infty 3}$ and $\mathfrak{M}$ it is helpful to have a simpler model in mind. The following is suggested: Let the big manifold be Euclidean 3-space and let the group be rotations about an axis. The orbits are then circles concentric with the axis and at right angles to it, and the orbit manifold is the Euclidean half-plane. A straight line (i.e., geodesic) in the big manifold will be a geodesic in the orbit manifold if it intersects or is parallel to the axis, so that it intersects every circle in its path at right angles. A straight line which is skew to the axis, however, is a hyperbola in the orbit manifold, and a skew line at right angles to the axis returns again to each orbit which it intersects.

the frontier be reached. On the other hand, this *can* happen. It happens, for example, when $d\gamma_{ij}/d\mathcal{S}=$ constant$\times\gamma_{ij}$, which obviously satisfies (6.7). In this case the geodesic motion is one of pure dilation, and the square of the distance to the frontier is, apart from an unimportant factor, simply the volume of 3-space.

(4) In the case of a pure dilation it is obvious that a geometrical singularity (zero volume) *does* occur at the frontier. That geometrical singularities must also occur in many other cases as well follows from the readily verified relation

$$d\,{}^{(3)}R/d\mathcal{S}=-\,{}^{(3)}R^{ij}d\gamma_{ij}/d\mathcal{S}\,, \tag{6.9}$$

which holds as long as condition (6.7) is satisfied. In Appendix A it is shown that most geodesics (i.e., all but a set of measure zero) strike the frontier at points where some of the γ_{ij} (and hence some of the $d\gamma_{ij}/d\mathcal{S}$) become infinite, even though γ itself vanishes. Except in special cases, therefore, expression (6.9) will acquire singularities at the frontier.

With these mathematical preliminaries in mind let us now have a look at quantum dynamics. It is helpful to begin by analyzing Eq. (5.5) in the WKB approximation, so as to make the maximum possible use of classical ideas. We write

$$\Psi[{}^{(3)}\mathcal{G}]=\mathfrak{A}\exp(i\mathfrak{W})\,, \tag{6.10}$$

where $\mathfrak{A}$ and $\mathfrak{W}$ are assumed to be real functionals satisfying (roughly) the restriction

$$|\delta\mathfrak{A}/\delta\gamma_{ij}|\ll|\mathfrak{A}\delta\mathfrak{W}/\delta\gamma_{ij}|\,. \tag{6.11}$$

The phase then satisfies the Hamilton-Jacobi equation[32]

$$G_{ijkl}\frac{\delta\mathfrak{W}}{\delta\gamma_{ij}}\frac{\delta\mathfrak{W}}{\delta\gamma_{kl}}=\gamma^{1/2}\,{}^{(3)}R\,, \tag{6.12}$$

while the amplitude satisfies the conservation law

$$\delta(\mathfrak{A}^2G_{ijkl}\delta\mathfrak{W}/\delta\gamma_{kl})/\delta\gamma_{ij}=0\,. \tag{6.13}$$

In addition, the ∞ constraints impose the restrictions

$$\left(\frac{\delta\mathfrak{W}}{\delta\gamma_{ij}}\right)_{.j}=0,\quad\left(\frac{\delta\mathfrak{A}}{\delta\gamma_{ij}}\right)_{.j}=0\,. \tag{6.14}$$

Each solution of the Hamilton-Jacobi equation (6.12) determines a family of solutions of the classical field equations (i.e., a family of Ricci-flat 4-geometries) having the following property: For every 3-geometry there exists one and only one member of the family which has the 3-geometry as a spacelike hypersection, i.e., for which the 3-geometry is to be found among the infinity of spacelike hypersections which the member admits. Once $\mathfrak{W}$ is given, each 3-geometry determines a unique 4-geometry. The 4-geometry may be computed by making the identification $\pi^{ij}=\delta\mathfrak{W}/\delta\gamma_{ij}$ and integrating the equation

$$\partial\gamma_{ij}/\partial x^0=2\alpha G_{ijkl}\delta\mathfrak{W}/\delta\gamma_{kl}+\beta_{i.j}+\beta_{j.i}\,, \tag{6.15}$$

which follows from (2.5) and (3.3).[33] The quantities α and β_i are, as always, completely arbitrary, at least in sufficiently small finite regions. (Some global restrictions will generally exist.)

It is not hard to verify that (6.15) does indeed yield a solution of the classical field equations for each initial 3-geometry. One simply differentiates (6.15) with respect to x^0 and replaces $\delta\mathfrak{W}/\delta\gamma_{kl}$ by its expression in terms of α, $\beta_{i.j}$, and $\partial\gamma_{ij}/\partial x^0$. One finds

$$\begin{aligned}\gamma_{ij,00}=&(\ln\alpha)_{,0}(\gamma_{ij,0}-\beta_{i.j}-\beta_{j.i})+\frac{\partial G_{ijkl}}{\partial\gamma_{mn}}\\&\times G^{klrs}\gamma_{mn,0}(\gamma_{rs,0}-2\beta_{r.s})+4\alpha G_{ijkl}\int\frac{\delta^2}{\delta\gamma_{kl}\delta\gamma_{m'n'}}\\&\times\left(\alpha' G_{m'n'r's'}\frac{\delta\mathfrak{W}}{\delta\gamma_{r's'}}+\beta_{m'.n'}\right)d^3x'\,.\end{aligned} \tag{6.16}$$

The integration which appears in the last term of this equation may be performed with the aid of the identities

$$\begin{aligned}&G_{m'n'r's'}\frac{\delta^2\mathfrak{W}}{\delta\gamma_{kl}\delta\gamma_{m'n'}}\frac{\delta\mathfrak{W}}{\delta\gamma_{r's'}}\\&=-\frac{1}{2}\left[\frac{\partial G_{mnrs}}{\partial\gamma_{kl}}\frac{\delta\mathfrak{W}}{\delta\gamma_{mn}}\frac{\delta\mathfrak{W}}{\delta\gamma_{rs}}+\gamma^{1/2}({}^{(3)}R^{kl}-\gamma^{kl}\,{}^{(3)}R)\right]\\&\times\delta(\mathbf{x},\mathbf{x}')+\tfrac{1}{2}\gamma'^{1/2}\gamma^{m'n'}\gamma^{r's'}\\&\qquad\times(\delta_{m'r'}{}^{kl}{}_{.n's'}-\delta_{m'n'}{}^{kl}{}_{.r's'})\,,\end{aligned} \tag{6.17}$$

$$\left(\frac{\delta^2\mathfrak{W}}{\delta\gamma_{kl}\delta\gamma_{m'n'}}\right)_{.n'}=\frac{\delta^2\mathfrak{W}}{\delta\gamma_{n'r'}}(\tfrac{1}{2}\delta_{n'r'}{}^{kl}{}_{.}{}^{m'}-\delta^{m'}{}_{n'}{}^{kl}{}_{.r'})\,, \tag{6.18}$$

which are obtained by functionally differentiating Eqs. (6.12) and (6.14) and making use of (4.25). The result is a set of six local-field equations which, together with (6.12) and (6.14) re-expressed in terms of α, $\beta_{i.j}$, $\gamma_{ij,0}$, are equivalent to the ten Einstein empty-space equations.

Let us now make the simplifying assumptions $\alpha_{,i}=0$, $\beta_i=0$. We then have

$$G^{ijkl}\gamma_{ij,0}\gamma_{kl,0}=4\alpha^2\gamma^{1/2}\,{}^{(3)}R\,. \tag{6.19}$$

Let us also assume that the integral

$$I\equiv\int\gamma^{1/2}\,{}^{(3)}R\,d^3x \tag{6.20}$$

[32] The Hamilton-Jacobi equation for general relativity appears to have been first written down by A. Peres, Nuovo Cimento **26**, 53 (1962).

[33] The inverse problem of constructing the $\mathfrak{W}$ which corresponds to a given family of solutions of the classical field equations has been analyzed in detail by U. H. Gerlach (to be published). The author is indebted to Gerlach for the opportunity of studying this analysis in manuscript prior to publication.

(extended over the whole of 3-space) is nonvanishing. We may then choose

$$\alpha=\tfrac{1}{2}|I|^{-1/2}, \tag{6.21}$$

which permits x^0 to be identified with the arc length $\$$ in the manifold $M^{\infty 3}$ and permits the first of Eqs. (6.14) to be re-expressed in the form

$$(G^{ijkl}d\gamma_{kl}/d\$)_{;j}=0, \tag{6.22}$$

which is identical with (6.7), showing that x^0 is in fact also the arc length in $\mathfrak{M}$. Finally, Eq. (6.16) takes the form

$$\frac{d^2\gamma_{ij}}{d\$^2}-\frac{d\gamma_{ik}}{d\$}\gamma_{kl}\frac{d\gamma_{lj}}{d\$}+\frac{1}{2}\frac{d\gamma_{ij}}{d\$}\gamma_{kl}\frac{d\gamma_{kl}}{d\$}+\tfrac{1}{8}\gamma^{-1/2}\gamma_{ij}G^{mnrs}\frac{d\gamma_{mn}}{d\$}\times\frac{d\gamma_{rs}}{d\$}=\frac{d\ln\alpha}{d\$}\frac{d\gamma_{ij}}{d\$}-2\alpha^2({}^{(3)}R_{ij}-\tfrac{1}{4}\gamma_{ij}\,{}^{(3)}R). \tag{6.23}$$

If the right-hand side of Eq. (6.23) were zero, the sequence of 3-geometries (as $\$$ varies) would trace out a geodesic in $\mathfrak{M}$. The right-hand term may therefore be regarded as a "force" term which caused the actual "trajectory" of 3-space to deviate from a geodesic. The following important questions arise: Is this "force term" powerful enough to keep the trajectory from striking the frontier? If not, what does arrival at the frontier mean physically?

Before giving answers to these questions, let us first take a crude over-all look at some of the simple implications of Eq. (6.23). It is not difficult to verify that if this equation is multiplied by $\frac{1}{2}\gamma^{1/2}\gamma^{ij}$ and the result is integrated over 3-space, the following equation is obtained:

$$\frac{d^2V}{d\$^2}-\frac{d\ln\alpha}{d\$}\frac{dV}{d\$}=-\tfrac{1}{4}, \quad V\equiv\int\gamma^{1/2}d^3x, \tag{6.24}$$

or equivalently,

$$\frac{d^2V}{d\tau^2}=-I, \tag{6.25}$$

where τ is the "proper time":

$$d\tau/d\$=\alpha. \tag{6.26}$$

From this it follows that the curve of V as a function of τ is concave downward whenever I is positive. Under these circumstances an expanding world tends to "slow down" while a contracting world tends to accelerate towards collapse.

A case for which I is positive is that in which 3-space has the geometry of a 3-sphere. The geometry cannot, however, remain spherical more than instantaneously, since the right-hand side of Eq. (6.19) is then everywhere positive, which requires the vector $d\gamma_{ij}/d\$$ to be "spacelike" in M for all $\mathbf{x}$, thus ruling out the possibility of a pure dilation. The derivative $d\gamma_{ij}/d\$$ must contain shearing components corresponding to the presence of the gravitational radiation which is, in fact, needed in order to "close up" the universe. On the other hand, the 3-geometry may still be spherical in a coarse-grained sense. That is, although the sign of $\gamma^{1/2}\,{}^{(3)}R$ may fluctuate at a fine-grained level due to the presence of gravitational waves, its mean value may approximate that of a 3-sphere. In this case Eq. (6.25) takes the approximate form

$$\frac{d^2V}{d\tau^2}\approx-6(4\pi^4V)^{1/3}, \tag{6.27}$$

leading to a total lifetime of the universe given by[34]

$$T\approx\tfrac{2}{3}(2\pi^2)^{-1/3}\int_0^{V\max}\frac{dV}{(V_{\max}{}^{4/3}-V^{4/3})^{1/2}}=\sqrt{2}\,\mathrm{cn}^{-1}(0|\tfrac{1}{2})R_{\max}=2.62R_{\max}, \tag{6.28}$$

where $R_{\max}$ is the radius of maximum expansion. Here it is clear that the "force term" in Eq. (6.23) does not prevent the 3-geometry from striking the frontier.

In the general case there are two factors which govern the trajectory of 3-space. Firstly, the condition $\alpha_{,i}=0$, which has been adopted in the above discussion, is known to be a poor one for keeping the hypersurfaces, $x^0=$ constant, smooth. When these hypersurfaces are sandwiched together, as here, with spatially uniform intervals, they often quickly develop geometrical singularities which have nothing to do with the geometry of space-time.[35] Such singularities can usually be avoided simply by relaxing the condition $\alpha_{,i}=0$. However—and this is the second factor—it is now known from the work of Avez,[36] Penrose,[37] Hawking,[38] and Geroch[39] that a nontrivial *singularity in space-time* "almost always" occurs at some point in the history of any physically interesting universe. At such a point abandonment of the condition $\alpha_{,i}=0$ is of no use. 3-space will acquire a geometrical singularity anwyay. Thus, if the initial hypersurface is sufficiently close to the point of onset of a change in 3-space topology, or if a so-called "trapped 2-surface" [37] is on the point of being born within it, then it will develop a geometrical

[34] This is to be compared with $T=2R_{\max}$ for a Friedmann universe filled with radiation treated as an ideal gas. Note that it is not possible to use expression (6.28) as an upper bound on the lifetime of the universe. Although it is easy to show that it is the spherical geometry which, for fixed V, makes I stationary (i.e., independent of small variations in the metric), this stationary point is neither a maximum nor a minimum, and hence it is not possible to assert that $I\geq 6(4\pi^4V)^{1/3}$.

[35] The phenomenon occurs already in a flat space-time. It is not possible to construct a family of uniformly spaced *curved* spacelike hypersurfaces in Minkowski space without the members of the family developing a geometrical singularity either in the past or in the future. The singularity always develops in the *convex* direction, contrary to the situation in a Euclidean manifold.

[36] A. Avez, Ann. Inst. Fourier (Grenoble) **13**, 105 (1963).

[37] R. Penrose, Phys. Rev. Letters **14**, 57 (1965).

[38] S. W. Hawking, Phys. Rev. Letters **17**, 444 (1966).

[39] R. P. Geroch, Phys. Rev. Letters **17**, 445 (1966).

singularity which has nothing to do with the maintenance of the condition $\alpha_{,i}=0$. In this case the force term in (6.32) is again powerless to prevent the trajectory of 3-space from striking the frontier; indeed it may hasten the impact.

The occurrence of a singularity in space-time itself is known as *gravitational collapse*. Gravitational collapse may involve the whole of 3-space, as when the volume of the universe goes to zero, or it may involve only a small part of it (e.g., a collapsing superstar). It seems extremely likely that the almost universal inevitability of gravitational collapse is closely connected to the existence of the frontier in $\mathfrak{M}$. However, the establishment of this connection in rigorous terms is a major problem which remains unsolved. The existence proofs of Refs. 36–39 give no indication of the precise physical nature of the collapse singularity except for the statement that the normal causal properties of space-time break down there. This alone, of course, is enough to guarantee that the singularity represents a real barrier beyond which it is impossible to extend the solution of Einstein's equations. It means that in certain regions of the universe (or in the universe as a whole) time for the classical physicist, ultimately comes to an end beyond which he can make no further predictions.

The question now arises whether the classical collapse barrier, which we shall denote by $\mathfrak{B}$, is also a barrier for the solutions of the quantum equation (5.5). That the answer is not obvious may be seen as follows. Consider first a point ${}^{(3)}\mathcal{G}$ in $\mathfrak{M}$ which is not on $\mathfrak{B}$. If ${}^{(3)}\mathcal{G}$ has a singularity this must be due to the hypersurface $x^0=$ constant being chosen poorly. At the singular point in 3-space both the right and left sides of Eq. (6.12) will diverge. Correspondingly the two terms inside the parentheses of Eq. (5.5) will each contribute a divergence at this point. The two divergences will, however, cancel so that Eq. (5.5) is still satisfied. Consider now a point on $\mathfrak{B}$. Here something special happens which causes Eq. (6.12) to break down. However, it does not automatically follow that Eq. (5.5) likewise breaks down, for there exist possibilities for treating Eq. (5.5) which have no counterparts in the classical theory. For example, we note that if $\mathfrak{W}$ is a solution of Eq. (6.12) then so is $-\mathfrak{W}$. Moreover, the addition of an arbitrary constant to $\mathfrak{W}$ leaves Eq. (6.12) unaffected. Let this constant be adjusted so that $\mathfrak{W}$ vanishes at the point on $\mathfrak{B}$ in question, and choose for the WKB form of the solution of (5.5), the *superposition*

$$\Psi=\mathfrak{C}[\exp(i\mathfrak{W})-\exp(-i\mathfrak{W})]. \tag{6.29}$$

Then Ψ itself vanishes at the barrier, and this might conceivably alleviate the singularity in Eq. (5.5) which would otherwise occur, and permit an extension of Ψ beyond the barrier.

If one is looking for an example on which to practice hand-waving arguments he might consider a situation in which 3-space is about to undergo a change in topology. It can be shown that a change of topology requires (a) the development of a geometrical singularity in 3-space and (b) a breakdown in the causal structure (e.g., hyperbolic signature) of space-time at the onset of the singularity. Therefore topological transitions cannot be handled classically. However, since the singularity in ${}^{(3)}\mathcal{G}$ need occur at only a single point of 3-space it may develop in such a way that the corresponding singularities of Eq. (5.5) all cancel. We are careful, of course, not to say that the singularities *will* cancel. No one really knows whether topological transitions can be handled quantum mechanically.

Although the classical and quantum barriers may not be identical, and although each may depend to some extent on the particular solution of the Hamilton-Jacobi equation (6.12), or of the "wave equation" (5.5), under consideration at the moment, it seems very probable that there exists an irreducible *core* which is common to all barriers. We have suggested that it may be possible to continue Ψ past 3-geometries which contain isolated singularities. However, it is extremely difficult to imagine how such a continuation could be performed beyond a 3-geometry which has a dense set of singularities, or which is singular at *all* of its points, e.g., a 3-space of zero volume. It is therefore likely that the following set theoretical inequality holds:

$$\text{orb}\prod_{\mathbf{x}}F(\mathbf{x})\subseteq\mathfrak{B}_Q\subseteq\mathfrak{B}, \tag{6.30}$$

where $\mathfrak{B}_Q$ denotes the quantum barrier. In the remainder of the paper we shall assume that this inequality does hold.

The fact that $\mathfrak{B}$ is not the empty set is an embarrassment to the classical physicist, for it means that his theory breaks down. The fact that $\mathfrak{B}_Q$ is not the empty set, however, is not necessarily embarrassing to the quantum physicist, for he may be able to dispose of it by simply imposing, on the state functional, the following condition:

$$\Psi[{}^{(3)}\mathcal{G}]=0 \quad \text{for all } {}^{(3)}\mathcal{G} \text{ on } \mathfrak{B}_Q. \tag{6.31}$$

Provided it does not turn out to be ultimately inconsistent, this condition, which is already suggested by (6.29), yields two important results. Firstly, it makes the probability amplitude for catastrophic 3-geometries vanish, and hence gets the physicist out of his classical collapse predicament. Secondly, it may permit the Cauchy problem for the "wave equation" (5.5) to be handled in a manner very similar to that of the ordinary Schrödinger equation. Thus let Σ be a hypersurface like that which appears in Eq. (5.19). Since the dimensionality of $\prod_{\mathbf{x}}F(\mathbf{x})$ (the orbit of which forms the "core" of $\mathfrak{B}_Q$) is the same as that of Σ, namely $5\times\infty^3$, it would appear that the specification of Ψ on Σ, together with the boundary condition (6.31), is equivalent to its specification on two hypersurfaces and hence suffices to determine $\Psi[{}^{(3)}\mathcal{G}]$ completely for all ${}^{(3)}\mathcal{G}$.

If this heuristic argument [based on the analogy of Eq. (5.5) to the Klein-Gordon equation] is indeed valid, then it is not necessary to specify also the normal derivatives of Ψ on Σ, despite the fact that Eq. (5.5) is of the second differential order.[40]

The only obvious difficulty with condition (6.31) is that it makes the presence of "negative frequency" components in Ψ unavoidable (see the discussion at the end of Sec. 5) and hence leaves one very unclear as to how to use Eq. (5.19) to define inner products and at the same time maintain positive definiteness of probability. In the following sections we shall show how this difficulty can be resolved in a special case.

7. THE QUANTIZED FRIEDMANN UNIVERSE

The simplest classical model which exhibits the collapse phenomenon is the Friedmann universe. If the Friedmann universe is assumed to be closed it must be filled either with gravitational radiation or with some other form of energy. It is not difficult to show that when other forms of energy are present in addition to gravity, the Hamiltonian constraint condition (4.3) is replaced by

$$(\mathfrak{H}+\mathfrak{K})\Psi=0, \tag{7.1}$$

where $\mathfrak{K}$ is the Hamiltonian of the system (or systems) giving rise to the additional energy. In order to avoid having to deal with entities as complicated as gravitons, with their spin and orbital states and their mutual interactions, we shall make use of such additional energy in the form of noninteracting material particles "at rest". The Friedmann universe is obtained by distributing these particles uniformly throughout a 3-sphere and "freezing out" all the degrees of freedom of the gravitational field save one, namely, that which corresponds to the time-varying spherical radius R.

If $\gamma^0{}_{ij}$ denotes the metric (in some coordinate system) of a 3-sphere of unit radius, then the 4-metric of the Friedmann world may be written in the form

$$(g_{\mu\nu})=\begin{pmatrix}-\alpha^2 & 0\\ 0 & \gamma_{ij}\end{pmatrix}, \quad \gamma_{ij}=R^2\gamma^0{}_{ij}, \tag{7.2}$$

where α and R depend only on x^0. Substituting this into (2.6), integrating over the volume $2\pi^2R^3$ of the Friedmann universe, and remembering that ${}^{(3)}R$ for a 3-sphere is $6/R^2$, we obtain for the effective Lagrangian of the gravitational field

$$L=12\pi^2[-\alpha^{-1}R(R_{,0})^2+\alpha R]. \tag{7.3}$$

As for the material particles (dust) which fill the universe, we shall, for reasons which will become clear as the analysis proceeds, endow them with internal dynamical degrees of freedom which may be described by canonical coordinates q^i and Lagrangians of the form $l(q,\dot{q})$, the dot denoting differentiation with respect to *proper* time:

$$\dot{q}^i=\alpha^{-1}q^i{}_{,0}. \tag{7.4}$$

Just as we have done for the gravitational field, however, we shall "freeze out" all the internal degrees of freedom save a small number by requiring all the particles to be identical and to be in coherent identical states (i.e., "in step"). Under these conditions the effective particle Lagrangian becomes

$$\mathbf{L}=\alpha N l(q,\alpha^{-1}q_{,0}), \tag{7.5}$$

where N is the total number of the particles in the universe.

Adding (7.3) and (7.5) to obtain the total Lagrangian we see that once again we have the primary constraint

$$\pi=\partial(L+\mathbf{L})/\partial\alpha_{,0}=0. \tag{7.6}$$

The wave function of the quantized Friedmann universe therefore cannot depend on α.

The total Hamiltonian becomes

$$\begin{aligned} H+\mathbf{H}&=\pi\alpha_{,0}+\Pi R_{,0}+P_iq^i{}_{,0}-L-\mathbf{L}\\ &=\pi\alpha_{,0}+\alpha(\mathfrak{H}+\mathfrak{K}), \end{aligned} \tag{7.7}$$

where

$$\Pi=\partial L/\partial R_{,0}=-24\pi^2\alpha^{-1}RR_{,0}, \tag{7.8}$$

$$P_i=\partial\mathbf{L}/\partial q^i{}_{,0}=Np_i, \quad p_i=\partial l/\partial\dot{q}^i, \tag{7.9}$$

$$\mathfrak{H}\equiv-\Pi^2/48\pi^2R-12\pi^2R, \tag{7.10}$$

$$\mathfrak{K}\equiv Nm, \quad m\equiv p_i\dot{q}^i-l. \tag{7.11}$$

The symbol m is here used to denote the internal Hamiltonian of the particles because the Hamiltonian is, in fact, the rest mass, provided the arbitrary zero point of the Lagrangian l has been properly chosen. We note that the "kinetic energy" term in the gravitational Hamiltonian (7.10) has the opposite sign (i.e., negative) from that of conventional Hamiltonians. This is because the only motion permitted to a Friedmann universe is one of pure dilation, and hence the coordinate R is "timelike."

The condition $\pi_{,0}=0$ leads immediately to the dynamical constraint $\mathfrak{H}+\mathfrak{K}=0$ which, in the quantum theory, takes the form (7.1). In the R representation this becomes

$$\left(\frac{1}{48\pi^2}R^{-1/4}\frac{\partial}{\partial R}R^{-1/2}\frac{\partial}{\partial R}R^{-1/4}-12\pi^2R+Nm\right)\Psi=0, \tag{7.12}$$

where factors have been ordered in such a way that the first term inside the parentheses becomes a one-dimensional Laplace-Beltrami operator,[41] and m is now the particle mass operator. Equation (7.12), which is

[40] Alternative and more detailed heuristic arguments leading to the same conclusion have been given by H. Leutwyler, University of Bern report, 1965 (unpublished).

[41] Here a definite ordering must be chosen. Since the number of degrees of freedom is now finite the ordering question cannot be treated as a problem in interpreting formally divergent symbols.

the analog of (5.5), must account completely for all the physical properties of the Friedmann universe.

It is by no means obvious how the familiar properties of the Friedmann world are to be extracted, in the classical limit, from Eq. (7.12), nor is it obvious what significance is to be attached to Ψ in the purely quantum domain. Difficulties of this type are not new to physics. A similar problem faced Schrödinger when he first wrote down the equation of the hydrogen atom. In his case there was a period of intense discussion, largely guided by Bohr, which ultimately led most physicists, with only a few dissenters of whom Einstein was the champion, to accept what has come to be known as the "Copenhagen view." The Copenhagen view depends on the assumed *a priori* existence of a classical level to which all questions of observation may utlimately be referred. Here, however, the whole universe is the object of inspection; there is no classical vantage point, and hence the interpretation question must be re-argued from the beginning. While we do not wish to stress this point unduly, since, after all, the Friedmann model ignores the vast complexities of the real universe, it is nevertheless clear that the quantum theory of space-time must ultimately force a deviation from the traditional Copenhagen doctrine.

Leaving aside these questions for the moment, let us note some of the simple mathematical properties of Eq. (7.12). If we carry out the point transformation

$$X\equiv R^{3/2},\quad \Phi\equiv(\partial R/\partial X)^{1/2}\Psi=(\tfrac{2}{3})^{1/2}R^{-1/4}\Psi, \tag{7.13}$$

Eq. (7.12) is converted to

$$-(3/64\pi^2)\partial^2\Phi/\partial X^2+12\pi^2X^{2/3}\Phi=Nm\Phi. \tag{7.14}$$

If the particles are in eigenstates of mass, so that m may be treated as a c number, and if the boundary condition

$$\Phi=0 \text{ at } X=0 \text{ or, equivalently, } \Psi=0 \text{ at } R=0 \tag{7.15}$$

analogous to (6.31) is imposed, then Eq. (7.14) becomes simply the Schrödinger equation of a particle of mass $32\pi^2/3$ moving at energy Nm in the one-dimensional potential

$$\begin{aligned} V&=\infty\,, & X&<0,\\ V&=12\pi^2X^{2/3}, & X&>0. \end{aligned} \tag{7.16}$$

Now unless the mass eigenvalue m happens to be such that Nm is one of the allowed eigenvalues of Eq. (7.14), the function Φ will not be normalizable but will behave in an exponential manner for large values of X. This is not necessarily bad if we insist on viewing R, and hence X, as a "time" coordinate, for it is usually impossible to require that a state function be normalizable with respect to *time*. Moreover, we may hesitate to allow the universe as a whole to determine the spectrum of masses which we can put into it, for in the classical theory the universe exerts no such control. However, several convincing arguments can be adduced which suggest that Φ must nonetheless be normalizable. The most important of these is that a closed Friedmann universe has, in the classical theory, a maximum radius of expansion. Hence if a correspondence principle is to exist, based on a transition to a classical limit, R must be effectively bounded from above. The existence of the classical turning point, as is well known, corresponds to the restriction to normalizable state functions.

For present purposes it suffices to determine the normalizable solutions of Eq. (7.12) in the WKB approximation. From the phase integral condition[42]

$$\begin{aligned} n+\tfrac{3}{4}&=-(2\pi)^{-1}\oint \Pi dR\\ &=24\pi\int_0^{R_{\max}}[R(R_{\max}-R)]^{1/2}dR, \end{aligned} \tag{7.17}$$

$$R_{\max}\equiv Nm/12\pi^2, \tag{7.18}$$

we obtain the "energy" spectrum

$$Nm=[48\pi^2(n+\tfrac{3}{4})]^{1/2},\quad n=0,1,2,\cdots. \tag{7.19}$$

Computation of the normalized state function itself involves only elementary integrals. Inside the turning point it is found to have the form

$$\begin{aligned} \Psi=(2/\pi R_{\max})^{1/2}&[(R_{\max}/R)-1]^{-1/4}\\ &\times\sin\{6\pi^2[(2R-R_{\max})(R(R_{\max}-R))^{1/2}\\ &\quad+R_{\max}{}^2\sin^{-1}((R/R_{\max})^{1/2})]\}, \end{aligned} \tag{7.20}$$

while outside it falls off to negligible values at distances of the order of $R_{\max}{}^{-1/3}$ beyond $R_{\max}$.[43]

In realistic situations this function has an enormous number of nodes. For a Friedmann world approximating the actual universe one finds, very roughly,

$$n\sim 10^{120}, \tag{7.21}$$

and if all the degrees of freedom of the real world were taken into account the number would be vastly greater. However, despite the enormity of the quantum number, the function (7.20) does not provide a classical description of the universe, for it is a static function, composed of standing waves undergoing neither expansion nor contraction. The standing waves may, to be sure, be regarded as a superposition of waves "traveling" in opposite directions, those "traveling" in the direction of expansion (increasing R) corresponding, by virtue of the "timelike" character of R, to the "positive frequency" components mentioned at the end of Sec. 5, and those "traveling" in the direction of contraction corresponding to "negative frequency"

[42] Here $n+\frac{3}{4}$ is used in place of the usual $n+\frac{1}{2}$ because of the "hard wall" character of the potential (7.16) for $X<0$.

[43] Cf. Ref. 11, p. 462.

components.[44] However, in order to make this "travel" apparent we need some other coordinate besides R.

It is at this point that the internal particle dynamics enter the picture. The collective internal motion permits the particle ensemble to be used as a clock. Classically the temporal behavior of R may be determined by means of the correlation which exists between R and the q^i. This correlation is described by the solutions of the Hamilton-Jacobi equation

$$\mathfrak{H}(R,\partial\mathcal{W}/\partial R)+\mathfrak{K}(q,\partial\mathcal{W}/\partial q)=0. \tag{7.22}$$

We may assume the particle Lagrangian l to be that of a multiply periodic system. The constants of integration of Eq. (7.22) are then conveniently taken to be the action variables $\mathbf{J}_i$ of the collective Lagrangian $\mathbf{L}$, and since the equation is obviously separable we have solutions of the form

$$\mathcal{W}=-W(R,J)+\mathbf{W}(q,\mathbf{J})+\text{const.}, \tag{7.23}$$

where

$$J=-(2\pi)^{-1}\oint \Pi dR=J(\mathbf{J}), \quad \Pi=-\partial W/\partial R, \tag{7.24}$$

the integral being taken over a complete expansion-contraction cycle of the Friedmann universe. For these solutions Eq. (7.22) takes the separated form

$$\frac{1}{48\pi^2R}\left(\frac{\partial W}{\partial R}\right)^2+12\pi^2R=\mathfrak{K}(q,\partial\mathbf{W}/\partial q)=E(\mathbf{J}), \tag{7.25}$$

where $E(\mathbf{J})$ is a certain function of the $\mathbf{J}_i$.

The q^i are obtained as functions of R and the $\mathbf{J}_i$ by integrating the simultaneous equations

$$dq^i/dR=\mathbf{V}^i(q,\mathbf{J})/V(R,J), \tag{7.26}$$

where

$$\mathbf{V}^i=(\partial\mathfrak{K}/\partial P_i)_{P=\partial\mathbf{W}/\partial q}, \tag{7.27}$$

$$V=(\partial\mathfrak{K}/\partial\Pi)_{\Pi=-\partial W/\partial R}=(24\pi^2R)^{-1}\partial W/\partial R. \tag{7.28}$$

The integrals of Eqs. (7.26) are not hard to obtain. If Eq. (7.25) is differentiated with respect to $\mathbf{J}_i$, one finds

$$V\frac{\partial^2W}{\partial R\partial J}\frac{\partial J}{\partial\mathbf{J}_i}=\mathbf{V}^j\frac{\partial^2\mathbf{W}}{\partial q^j\partial\mathbf{J}_i}=\frac{\partial E}{\partial\mathbf{J}_i}, \tag{7.29}$$

which, together with (7.26), yields

$$\frac{\partial^2\mathbf{W}}{\partial\mathbf{J}_i\partial q^j}dq^j=\frac{\partial J}{\partial\mathbf{J}_i}\frac{\partial^2W}{\partial J\partial R}dR, \tag{7.30}$$

whence

$$-\frac{\partial W}{\partial J}\frac{\partial J}{\partial\mathbf{J}_i}+\frac{\partial\mathbf{W}}{\partial\mathbf{J}_i}=\delta^i, \tag{7.31}$$

where the δ^i are "phase constants." Equations (7.31) may be solved algebraically to express the q's in terms of R and the constants of integration $\mathbf{J}_i$, δ^i. The δ_i determine the relative phases of the simultaneous oscillatory motions, and the $\mathbf{J}_i$ determine the amplitudes.

In the quantum theory an analogous correlation between R and the q^i can be established provided the state function has the form of a *superposition* of solutions of (7.12) corresponding to different eigenvalues of m. It is well known that a multiply periodic system cannot be used as a clock if it is in an eigenstate of energy. The uncertainty principle requires many different energy levels to be represented in its wave function. Here, however, we run into a very special difficulty which is peculiar to the quantum theory of space-time. The values which m can assume are already determined by the quantization condition (7.19) quite independently of the form of the particle Lagrangian l. Hence, unless the operators $\mathfrak{K}$ and $-\mathfrak{K}$ have at least one eigenvalue in common, *the Hamiltonian constraint* (7.1) *will have no solutions at all.* Equation (7.1) is unlike an ordinary time-independent Schrödinger equation in that it picks out only a single eigenvalue of the operator $\mathfrak{H}+\mathfrak{K}$. Moreover, the latter operator, being the sum of two operators having spectra bounded respectively from above and from below, has itself a spectrum which stretches from $-\infty$ to ∞.

For purposes of the present discussion we must *assume* not only that $\mathfrak{H}+\mathfrak{K}$ has a zero eigenvalue but that this eigenvalue is highly *degenerate.* We shall postpone until Sec. 10 a discussion of what the actual state of affairs may be in the real universe. For the present we concentrate on mathematical developments.

We shall confine ourselves to the WKB approximation and look for solutions of Eq. (7.1) of the form [cf. Eq. (6.11)]

$$\Psi=A\exp[i(-W+\mathbf{W})], \tag{7.32}$$

where A is a real amplitude satisfying (hopefully) the inequalities [cf. Eq. (6.12)]

$$\left|\frac{\partial A}{\partial R}\right|\ll\left|A\frac{\partial W}{\partial R}\right|, \quad \left|\frac{\partial A}{\partial q^i}\right|\ll\left|A\frac{\partial\mathbf{W}}{\partial q^i}\right|. \tag{7.33}$$

A differential equation for A may be obtained by substituting (7.32) into (7.1). One finds

$$[\mathfrak{H}(R,-i\partial/\partial R-\partial W/\partial R)+\mathfrak{K}(q,-i\partial/\partial q+\partial\mathbf{W}/\partial q)]A=0. \tag{7.34}$$

When the inequalities (7.33) are satisfied the "big" terms of (7.34) already add up to zero by virtue of the Hamiltonian-Jacobi equation (7.22). In order to obtain conditions on A we must include the smaller, "higher-order" terms, and for this purpose it is convenient to introduce a smooth real test function $\varphi(R,q)$ which vanishes outside a finite closed region in the $R-q$

[44] Owing to the negative character of the kinetic-energy term of the Hamiltonian (7.10), the directions of "travel" of the exponential components of a standing wave are opposite to the conventional ones.

manifold.[45] We then subtract the equation

$$\int_0^\infty dR \int dq\ \varphi A[\mathfrak{IC}(R, -i\partial/\partial R - \partial W/\partial R) + \mathfrak{K}(q, -i\partial/\partial q + \partial \mathbf{W}/\partial q)]A = 0 \quad (7.35)$$

from its complex conjugate and use the Hermiticity of $\mathfrak{IC}$ and $\mathfrak{K}$ to obtain

$$\int_0^\infty dR \int dq\ A[\varphi, \mathfrak{IC}(R, -i\partial/\partial R - \partial W/\partial R) + \mathfrak{K}(q, -i\partial/\partial q + \partial \mathbf{W}/\partial q)]A = 0. \quad (7.36)$$

When the inequalities (7.33) are satisfied, this becomes approximately

$$i\int_0^\infty dR \int dq[(\partial\varphi/\partial R)V + (\partial\varphi/\partial q^i)\mathbf{V}^i]A^2 = 0, \quad (7.37)$$

which, by virtue of the arbitrariness of φ, implies (after an integration by parts)

$$\partial(A^2V)/\partial R + \partial(A^2\mathbf{V}^i)/\partial q^i = 0. \quad (7.38)$$

This equation, which is the analog of (6.14), assures conservation of the "flux of probability" in the R-q manifold.

The general solution of Eq. (7.38) can be obtained by making use of the relations

$$V\partial\delta^i/\partial R + \mathbf{V}^j\partial\delta^i/\partial q^j = 0, \quad (7.39)$$

$$\partial(V\partial^2 W/\partial R\partial J)/\partial R = 0, \quad (7.40)$$

$$\partial(\mathbf{V}^i D)/\partial q^i = 0, \quad (7.41)$$

where

$$D \equiv \det(D_i{}^j), \quad D_i{}^j = \partial^2\mathbf{W}/\partial q^i\partial\mathbf{J}_j, \quad (7.42)$$

and where δ^i, in Eq. (7.39) is regarded not as a constant of integration but as a function of R and the q^i, defined by (7.31). It is easy to see that Eq. (7.39) follows from (7.29) and (7.31). Equation (7.40) is obtained by differentiating (7.29) with respect to R, while (7.41) is obtained by differentiating (7.29) with respect to q^k and multiplying by the matrix $D^{-1}{}_k{}^i$ inverse to (7.42). From these relations it follows immediately that

$$A^2 = (\partial^2 W/\partial R\partial J)DF(\delta), \quad (7.43)$$

where F is an arbitrary function of the δ^i.

Actually the form of F is not arbitrary, since there are other differential equations which the state function (7.32) must satisfy in addition to (7.12), namely, the eigenvalue equations

$$J(R, -i\partial/\partial R)\Psi = J\Psi, \quad (7.44)$$

$$\mathbf{J}_i(q, -i\partial/\partial q)\Psi = \mathbf{J}_i\Psi. \quad (7.45)$$

[45] It is aways to be understood that the R-q manifold is restricted to positive values of R (*excluding* zero).

The operators $J(R, -i\partial/\partial R)$ and $\mathbf{J}_i(q, -i\partial/\partial q)$ are obtained by solving the equations

$$\Pi = -\partial W/\partial R, \quad P_i = \partial\mathbf{W}/\partial q^i \quad (7.46)$$

for the J's in terms of Π, R, and the P's and q's, making the replacements $\Pi \to -i\partial/\partial R$, $P_i \to -i\partial/\partial q^i$, and carrying out appropriate Hermiticity symmetrizations. Now

$$J(R, -i\partial/\partial R)\Psi = \{\exp[i(-W+\mathbf{W})]\} \times J(R, -i\partial/\partial R - \partial W/\partial R)A, \quad (7.47)$$

$$\mathbf{J}_i(q, -i\partial/\partial q)\Psi = \{\exp[i(-W+\mathbf{W})]\} \times \mathbf{J}_i(q, -i\partial/\partial q + \partial\mathbf{W}/\partial q)A. \quad (7.48)$$

Because of the identities

$$J(R, -\partial W/\partial R) \equiv J, \quad \mathbf{J}_i(q, \partial W/\partial q) \equiv \mathbf{J}_i \quad (7.49)$$

the "big" terms of (7.47) and (7.48) already yield Eqs. (7.44) and (7.45). Hence the "higher-order" terms must vanish. With the aid of the inequalities (7.33) and a test function φ, as before, one easily finds that this implies

$$\partial(A^2\partial J/\partial\Pi)/\partial R = 0, \quad (7.50)$$

$$\partial(A^2\partial\mathbf{J}_i/\partial P_i)/\partial q^j = 0. \quad (7.51)$$

But $\partial J/\partial\Pi = -(\partial^2 W/\partial R\partial J)^{-1}$ and $\partial\mathbf{J}_i/\partial P_j = D^{-1}{}_i{}^j$. Hence, substituting (7.43) into (7.47) and (7.48), and making use of the identity

$$\partial(DD^{-1}{}_i{}^j)/\partial q^j = 0. \quad (7.52)$$

which can be shown to hold by virtue of the symmetry of $\partial D_k{}^l/\partial q^j$ in j and k, one finds that F must be a *constant*, independent of the δ_i.

In order to obtain normalizable solutions of (7.12) the J's must be quantized. In addition, the branch-point behavior of the functions W and $\mathbf{W}$ at the classical turning points must be taken into account, and superpositions of the form (7.32) corresponding to the different branches must be employed. These superpositions are the standard WKB solutions.

Suppose we freeze out all the collective particle degrees of freedom save one, and suppose this degree of freedom corresponds to a motion of libration in a smooth potential. Then we may distinguish two branches of the function $\mathbf{W}$, a branch $\mathbf{W}^+$ which increases with increasing q and a branch $\mathbf{W}^-$ which decreases with the increasing q. Similarly we may distinguish two branches, W^+ and W^-, of the function W. These branches are determined only up to arbitrary constants. The constants may be adjusted so that the WKB solutions take the form

$$\Psi_n = A_n[\exp(-iW_n{}^+) + \exp(-iW_n{}^-)] \times [\exp(i\mathbf{W}_n{}^+) + \exp(i\mathbf{W}_n{}^-)], \quad (7.53)$$

where

$$W_n{}^\pm \equiv W^\pm(R, n+\tfrac{3}{4}), \quad \mathbf{W}_n{}^\pm \equiv W^\pm(q, \mathbf{n}+\tfrac{1}{2}), \quad (7.54)$$

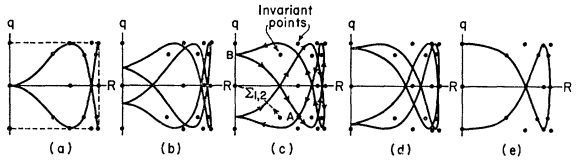

FIG. 1. Packet traces in the R-q plane. (Case $\Delta n=3$, $\Delta\mathbf{n}=2$.)

only those quantum numbers n, $\mathbf{n}$ being permitted for which

$$J=n+\tfrac{3}{4} \text{ when } \mathbf{J}=\mathbf{n}+\tfrac{1}{2}. \tag{7.55}$$

If the operator $\mathfrak{H}+\mathfrak{K}$ is to have a zero eigenvalue which is highly degenerate, it is clear from Eqs. (7.25) and (7.55) that the coherent internal dynamical behavior of the particles filling the universe must be precisely matched to that of the universe itself in such a way that the derivative $dJ/d\mathbf{J}$ is a *constant rational number* over a wide range of values of $\mathbf{J}$. We shall write

$$dJ/d\mathbf{J}=\Delta n/\Delta\mathbf{n}, \tag{7.56}$$

where Δn and $\Delta\mathbf{n}$ are relatively prime integers. Δn and $\Delta\mathbf{n}$ are the spacings between adjacent permitted values of the quantum numbers n and $\mathbf{n}$, respectively. An immediate consequence of Eq. (7.56) is that the angular frequencies (with respect to proper time) of the R and q motions are always (within the allowed range of $\mathbf{J}$ values) commensurable. These angular frequencies are given by

$$\omega\equiv dE/dJ=\boldsymbol{\omega}d\mathbf{J}/dJ, \tag{7.57}$$

$$\boldsymbol{\omega}\equiv dE/d\mathbf{J}, \tag{7.58}$$

and, in the allowed range, satisfy

$$\omega\Delta n=\boldsymbol{\omega}\Delta\mathbf{n}. \tag{7.59}$$

Use of the angular frequencies permits Eqs. (7.29) to be re-expressed in the form

$$V\frac{\partial^2 W}{\partial R\partial J}=\omega, \quad \mathbf{V}\frac{\partial^2 W}{\partial q\partial\mathbf{J}}=\boldsymbol{\omega}, \tag{7.60}$$

whence (7.43) becomes

$$A^2=F\omega\boldsymbol{\omega}/V\mathbf{V}. \tag{7.61}$$

For later convenience we shall choose

$$F=\Delta n/2\pi\boldsymbol{\omega}=\Delta\mathbf{n}/2\pi\omega. \tag{7.62}$$

The normalization constant A_n in expression (7.53) is then given by

$$A_n=(\omega_n\Delta n/2\pi|V_n\mathbf{V}_\mathbf{n}|)^{1/2} \approx[(E_{n+\Delta n}-E_n)/2\pi|V_n\mathbf{V}_\mathbf{n}|]^{1/2}, \tag{7.63}$$

the subscripts indicating that the quantized values of the quantities to which they are affixed are to be employed. No signs have been placed on the V's to correspond with the different branches of the W's, because motion in a potential is time-reversal invariant, and hence $|V^+|=|V^-|$, $|\mathbf{V}^+|=|\mathbf{V}^-|$.

8. A WAVE PACKET FOR THE UNIVERSE. THE CONCEPT OF TIME

We are now in a position to construct a state function exhibiting classical behavior. We do this by superposing many WKB solutions:

$$\Psi_a=\sum_n{}' a_n\Psi_n, \tag{8.1}$$

the prime indicating that the summation is to be carried out only over those quantum numbers which satisfy condition (7.55). If the a's are carefully chosen, Ψ_a will have the form of a "wave packet" which traces out a classical trajectory, namely, a generalized lissajous figure in the R-q plane. It is easy to see that Eq. (7.31) is just the condition for constructive interference, provided the symbols in the equation are understood to denote the "peak" values of the quantities to which they refer.

In view of the commensurability condition (7.59) the lissajous figure is necessarily closed and of finite length.[46] A typical set of packet traces is shown in Fig. 1 for the case $\Delta n=3$, $\Delta\mathbf{n}=2$. The action variable $\mathbf{J}$ is the same for each trace, but the phase constant δ varies from one to the other. The following facts may be inferred from the figures: Each trace lies within a rectangle having sides equal to the full amplitudes of the oscillations. Except in the degenerate cases depicted in Figs. 1(a) and 1(e), each trace divides the rectangle into $2\Delta n\Delta\mathbf{n}+\Delta n+\Delta\mathbf{n}+1$ disjoint regions. Although the size and shape of corresponding regions vary from figure to figure, each region contains an *invariant point* which is independent of δ. These points are shown in the figures.

The degenerate curves are those which have collapsed onto the invariant points. They are divided by the invariant points into a total of $4\Delta n\Delta\mathbf{n}$ segments, which will be called *invariant segments*. The invariant segments may be labeled in a systematic fashion, starting, say, from the "southwest" corner of the enclosing rectangle, by pairs of integers $(r,\mathbf{r})$ satisfying $1\leq r\leq 2\Delta n$, $1\leq\mathbf{r}\leq 2\Delta\mathbf{n}$. When an invariant segment is used as a contour of integration (see Sec. 9) it will be denoted by the symbol $\Sigma_{r,\mathbf{r}}$. [See Fig. 1(c).] Each invariant segment corresponds to a lapse of proper time of amount $T/4\Delta n=\mathbf{T}/4\Delta\mathbf{n}$ where T and $\mathbf{T}$ are the oscillation periods.

[46] Degenerate forms of closure [see, for example, Figs 1(a) and 1(e)], in which the packet "moves" back and forth along the same curve, are to be understood as included in this statement. Also, if successive groups of a's are chosen to vanish in such a way that the *effective* spacing between adjacent quantum numbers becomes a multiple of Δn (or $\Delta\mathbf{n}$), then the packet trace will consist of m separately closed Lissajous figures superimposed upon one another. Such a trace must be understood as representing a *single* packet which consists of m disconnected parts. Although the following discussion can be extended to include such situations, they will, for simplicity, be excluded from consideration.

A wave packet will be called *good* if its state function has negligible values throughout most of each region containing an invariant point. Except at intersection points or turning points only one of the branches of each of the functions W and $\mathbf{W}$ is involved in the constructive interference of the WKB functions at any one position along a good packet trace. Thus, for "motion" in a "northeasterly" direction the relevant branches are W^+, $\mathbf{W}^+$, provided we use the standard convention that time increases in the direction of increasing W and $\mathbf{W}$. Continuing around the compass we have W^+, $\mathbf{W}^-$ for SE; W^-, $\mathbf{W}^-$ for SW; and W^-, $\mathbf{W}^+$ for NW. The trace is thus divided into *branch segments* having definite quadrant orientations.

A given branch segment may be intersected by other branch segments which further subdivide it. The resulting pieces will be called *simple segments*. Each simple segment is intersected by precisely one invariant segment, and the two may therefore be labeled by the same integers. The branches involved in a given simple segment are W^+, $\mathbf{W}$, or W^-, $\mathbf{W}^-$ if $r+\mathbf{r}$ is odd and W^+, $\mathbf{W}^-$, or W^-, $\mathbf{W}^+$ if $r+\mathbf{r}$ is even, the choice depending on the direction of "motion." The direction in which the packet moves as time increases may be indicated by affixing arrows to the packet trace, as in Fig. 1(c). If the W's are adjusted so that $\delta=0$ corresponds to a degenerate trace, then the arrows are reversed by changing the sign of δ.

Proper time itself is defined by

$$\tau=\omega^{-1}\Theta, \quad \boldsymbol{\tau}=\boldsymbol{\omega}^{-1}\boldsymbol{\Theta}, \tag{8.2}$$

where the ω's are defined by Eqs. (7.57) and (7.58), and

$$\Theta=\partial W/\partial J, \quad \boldsymbol{\Theta}=\partial \mathbf{W}/\partial \mathbf{J}. \tag{8.3}$$

Classically the angle variables Θ and $\boldsymbol{\Theta}$ are canonically conjugate to $-J$ and $\mathbf{J}$, respectively, and hence τ and $\boldsymbol{\tau}$ are canonically conjugate to $\mathcal{H}$ and $\boldsymbol{\mathcal{H}}$, respectively. In the quantum theory this leads one to write the commutation relations

$$[\tau,\mathcal{H}]=i, \quad [\boldsymbol{\tau},\boldsymbol{\mathcal{H}}]=i. \tag{8.4}$$

It is important to remember, however, that the quantum τ's are *not Hermitian*. This follows not only from their periodic character, which arises from their dependence on the Θ's [Eqs. (8.2)], but also from the fact that their canonical conjugates, $\mathcal{H}$ and $\boldsymbol{\mathcal{H}}$, have discrete, "one-sided" spectra, bounded, respectively, from above and below. The usual eigenvector properties which hold for Hermitian operators therefore do not hold for the τ's, and we must distinguish between right and left eigenvectors.

Let us introduce the left eigenvectors $\langle\tau',\boldsymbol{\tau}'|$ which, in virtue of (8.4), may be chosen to satisfy

$$-i\frac{\partial}{\partial\tau'}\langle\tau',\boldsymbol{\tau}'|=\langle\tau',\boldsymbol{\tau}'|\mathcal{H}, \quad -i\frac{\partial}{\partial\boldsymbol{\tau}'}\langle\tau',\boldsymbol{\tau}'|=\langle\tau',\boldsymbol{\tau}'|\boldsymbol{\mathcal{H}}. \tag{8.5}$$

We may also introduce the corresponding conjugate vectors, denoted by $|\tau',\boldsymbol{\tau}'\rangle$, which are right eigenvectors of the conjugate operators $\tau^\dagger$ and $\boldsymbol{\tau}^\dagger$. Now let $\mathcal{P}$ denote the projection operator into the physical subspace of allowed state vectors. Using Eqs. (8.5) and the Hamiltonian constraint (7.1), which may be rewritten in the form

$$(\mathcal{H}+\boldsymbol{\mathcal{H}})\mathcal{P}=\mathcal{P}(\mathcal{H}+\boldsymbol{\mathcal{H}})=0, \quad \mathcal{P}^2=\mathcal{P}, \tag{8.6}$$

it is easy to see that $\langle\tau,\boldsymbol{\tau}|\mathcal{P}$ depends only on the difference $\tau-\boldsymbol{\tau}$. This simple dependence may be recognized as a quantum consequence of the classical correlation

$$\tau-\boldsymbol{\tau}=-\boldsymbol{\omega}^{-1}\delta, \tag{8.7}$$

which follows from (7.31) and (8.2).

The projection operator $\mathcal{P}$ is conveniently defined in terms of the eigenvectors $|n+\frac{3}{4}, \mathbf{n}+\frac{1}{2}\rangle$ of the J's. Writing $|n+\frac{3}{4}, \mathbf{n}+\frac{1}{2}\rangle\equiv|n\rangle$ whenever the quantum numbers are restricted as in (7.55), we have

$$\mathcal{P}=\sum_{n}{}'|n\rangle\langle n|, \tag{8.8}$$

provided the normalization $\langle n|n'\rangle=\delta_{nn'}$ is assumed. In virtue of Eqs. (8.5) we may also write

$$\langle\tau,\boldsymbol{\tau}|n\rangle=(\omega_n\Delta n/2\pi)^{1/2}\exp[-iE_n(\tau-\boldsymbol{\tau})]. \tag{8.9}$$

The normalization here is chosen so as to maximize the orthogonality properties of the vectors $\langle\tau,\boldsymbol{\tau}|$ relative to the physical subspace. Noting that $\omega_n\Delta n$ gives the approximate spacing between adjacent permitted "energy" eigenvalues E_n, we have

$$\begin{aligned}\langle\tau,\boldsymbol{\tau}|\mathcal{P}|\tau',\boldsymbol{\tau}'\rangle &\approx(2\pi)^{-1}\sum_{n}{}'(\exp\{-iE_n[(\tau-\boldsymbol{\tau})-(\tau'-\boldsymbol{\tau}')]\})\Delta E_n\\ &=\tilde{\delta}((\tau-\boldsymbol{\tau})-(\tau'-\boldsymbol{\tau}')).\end{aligned} \tag{8.10}$$

If the "energy" spectrum were continuous and ranged from $-\infty$ to ∞ the function $\tilde{\delta}$ would be the Dirac δ. In reality it is a function which although divergent at the origin does not completely vanish elsewhere. Thus the eigenvectors $\langle\tau,\boldsymbol{\tau}|$ are only approximately orthonormal, a fact which stems from the lack of strict Hermiticity of the operators τ and $\boldsymbol{\tau}$.

Instead of working with the vectors $\langle\tau,\boldsymbol{\tau}|$ it is more interesting to work with $\langle\tau,q|$, $\langle R,\boldsymbol{\tau}|$, and $\langle R,q|$, which are defined in an obvious fashion. The normalization of $\langle R,q|$ may be fixed by setting

$$\langle R,q|n\rangle=\Psi_n, \tag{8.11}$$

where Ψ_n is the function having the WKB approximation (7.53), with A_n given by (7.63). In a similar manner the normalization of $\langle\tau,q|$ and $\langle R,\boldsymbol{\tau}|$ may be fixed by giving the WKB approximations of their inner products with $|n\rangle$. We shall choose

$$\begin{aligned}\langle\tau,q|n\rangle\equiv\boldsymbol{\Phi}_n=(\boldsymbol{\omega}_n\Delta\mathbf{n}/2\pi|\mathbf{V}_n|)^{1/2}\\ \times e^{-iE_n\tau}[\exp(i\mathbf{W}_n^+)+\exp(i\mathbf{W}_n^-)],\end{aligned} \tag{8.12}$$

$$\begin{aligned}\langle R,\boldsymbol{\tau}|n\rangle=\Phi_n=(\omega_n\Delta n/2\pi|V_n|)^{1/2}[\exp(-iW_n^+)\\ +\exp(-iW_n^-)]\exp(iE_n\boldsymbol{\tau}).\end{aligned} \tag{8.13}$$

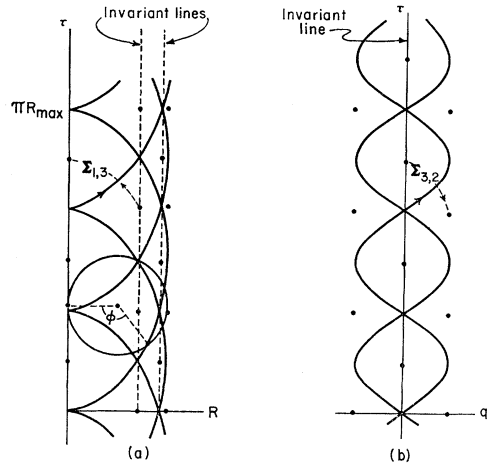

FIG. 2. Packet traces in the (a) R-τ and (b) τ-q planes. (Case $\Delta n=3$, $\Delta\mathbf{n}=2$.)

Denoting by $\mathbf{\Phi}$ and Φ arbitrary superpositions of the functions $\mathbf{\Phi}_n$ and Φ_n, respectively, we may write the Schrödinger equations

$$i\partial\mathbf{\Phi}/\partial\tau=\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\mathbf{\Phi}, \tag{8.14}$$

$$i\partial\Phi/\partial\tau=\mathfrak{H}\mathcal{C}(R,-i\partial/\partial R)\Phi. \tag{8.15}$$

When the function Ψ_a of (8.1) has the form of a wave packet so also have the corresponding functions Φ_a and $\mathbf{\Phi}_a$. The form of the packet trajectory in the case of the function Φ_a may be determined by noting that the condition for constructive interference, which establishes the correlation between R and τ, is

$$\partial W/\partial J=\omega\tau+\delta, \tag{8.16}$$

where δ is a phase constant and the other symbols denote the peak values of the quantities to which they refer.[47] Differentiation of Eq. (8.16) with respect to τ and use of Eqs. (7.28) and (7.60) yields

$$dR/d\tau=V=-(24\pi^2R)^{-1}\Pi$$
$$=R^{-1}[R(R_{\max}-R)]^{1/2}. \tag{8.17}$$

The integration of this equation is most easily carried out with the aid of an angle ϕ defined by

$$d\tau/d\phi=R. \tag{8.18}$$

This gives

$$d\phi=[R(R_{\max}-R)]^{-1/2}dR, \tag{8.19}$$

which yields the familiar cyclodial trajectory of the dust-filled Friedmann universe:

$$R=\tfrac{1}{2}R_{\max}(1-\cos\phi), \tag{8.20}$$

$$\tau=\tfrac{1}{2}R_{\max}(\phi-\sin\phi), \tag{8.21}$$

the constants of integration being chosen so that $R=0$, $\phi=0$ at $\tau=0$. We note that by virtue of the boundary condition (7.15) the packet rebounds repeatedly from the collapsed state until it ultimately loses its identity owing to spreading. Throughout the period of each rebound the width of the packet remains at all times finite, never suffering infinite compression. Transition through collapse thus becomes, in the quantum theory, a continuous process—something which cannot be achieved within the classical framework.

Figure 2 shows the curves traced out in the R—τ and τ—q planes by the packet of Fig. 1, and reveals a slight complication which was overlooked in the above simple analysis. The curves in the two planes appear to depict Δn and $\Delta\mathbf{n}$ distinct packets, respectively, rather than only a single packet. The extra "ghost packets" arise because the complete spectra of $\mathfrak{H}\mathcal{C}$ and $\mathfrak{H}\mathcal{C}$ are not made use of in the superposition (8.1). Only every Δnth level of $\mathfrak{H}\mathcal{C}$ and every $\Delta\mathbf{n}$th level of $\mathfrak{H}\mathcal{C}$ occur. This means that τ and τ are determined modulo $T/\Delta n=\mathbf{T}/\Delta\mathbf{n}$ rather than modulo T and $\mathbf{T}$, respectively, and a given packet must consequently appear "simultaneously" in several places in order to allow for the resulting proper-time ambiguity.[48]

The multiple traces intersect themselves along $\Delta n-1$ and $\Delta n-1$ phase-invariant lines, respectively. (See the indicated lines in the figures.) These lines divide the *branch segments* (which are defined just as for the R-q lissajous figures) into simple segments. Each simple segment is straddled by a unique pair of points at which maximum destructive interference occurs. The pairs of points may be connected by arcs intersecting the associated simple segments. These arcs will be denoted by $\mathbf{\Sigma}_{r,\mathbf{t}}$ and $\mathbf{\Sigma}_{t,\mathbf{r}}$ in the R—τ and τ—q planes, respectively. The pairs of suffixes r, $\mathbf{t}$ and t, $\mathbf{r}$ have the ranges $r=1, 2\cdots\Delta n$; $\mathbf{r}=1, 2, \cdots\Delta\mathbf{n}$; t, $\mathbf{t}=\cdots-2, -1, 0, 1, 2, \cdots$, and may be used in an obvious manner to identify either the simple segments or their associated intersecting arcs. Examples of arcs and point pairs are shown in the figure.

With the introduction of the three wave functions Φ, $\mathbf{\Phi}$, and Ψ we now have at our disposal three distinct mathematical windows from which to view the Friedmann world. From one window the material content of the universe is seen as a clock for determining the dynamical behavior of the world geometry. From another it is the geometry which appears as a clock for determining the dynamical behavior of the material content. From the third the geometry and the material content appear on equal footing, each one correlated in a certain manner with the other.

It is the third window which is to be preferred as most accurately revealing the physics of the quantized Friedmann model. The variables τ and τ, because of their lack of Hermiticity, are not rigorously observable and hence cannot yield a measure of proper time which is valid under all circumstances. It is only with good

[47] Equation (8.16) also follows from (8.2), (8.3), and (8.7).

[48] This has also the consequence that $\langle R,\tau|\mathcal{P}|R',\tau\rangle$ and $\langle\tau,q|\mathcal{P}|\tau,q'\rangle$ do not have the form of simple δ functions, although they diverge at $R=R'$ and $q=q'$.

wave packets that these variables are useful. But even with a good packet the description in terms of τ and $\boldsymbol{\tau}$ is not perfect, as is revealed in a striking way by the fact that the wave packets Φ_a and $\boldsymbol{\Phi}_a$ inevitably spread in "time," whereas the packet Ψ_a does not. It is for this reason that we may say that "time" is only a phenomenological concept, useful under certain circumstances.

It is worth remarking that it is not necessary to drag in the whole universe to argue for the phenomenological character of time. If the principle of general covariance is truly valid then the quantum mechanics of every-day usage, with its dependence on Schrödinger equations of the form (8.14) or (8.15), is only a phenomenological theory. For the only "time" which a covariant theory theory can admit is an intrinsic time defined by the contents of the universe itself. Any intrinsically defined time is necessarily non-Hermitian, which is equivalent to saying that there exists no clock, whether geometrical or material, which can yield a measure of time which is operationally valid under *all* circumstances, and hence there exists no operational method for determining the Schrödinger state function with arbitrarily high precision. This statement also follows directly from the uncertainty principle. Because every clock has a "one-sided" energy spectrum, its ultimate accuracy must necessarily be inversely proportional to its rest mass. When the whole universe is cast in the role of a clock, the concept of time can of course be made fantastically accurate (at least in principle) because of the enormity of the masses and quantum numbers involved. But as long as the universe is finite, a theoretical limit to the accuracy nevertheless remains.

9. THE INNER PRODUCT

We shall now use the results of the two preceding sections to show how the definition (5.19) for inner products can be rescued from the negative-probability disaster, at least in the case of the quantized Friedmann model. First we must derive the form which (5.19) takes in this model. Consider the following integral:

$$\int \varphi\{[\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_b]^*\Psi_a-\Psi_b{}^*\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_a\}dq,$$

where Ψ_a and Ψ_b are arbitrary complex functions of R and the q^i, and φ is a real test function. Because of the Hermiticity of $\mathfrak{H}\mathcal{C}$ this integral may be rewritten in the form

$$\int \Psi_b{}^*[\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q),\varphi]\Psi_a dq = -i\int \Psi_b{}^*\mathbf{V}^i(q,-i\partial/\partial q)\bullet(\partial\varphi/\partial q^i)\Psi_a dq, \quad (9.1)$$

where $\mathbf{V}^i$ is defined by (7.27), but with the replacement $P_i\rightarrow -i\partial/\partial q^i$ instead of $P_i=\partial\mathbf{W}/\partial q^i$, and where the dot in the right-hand integrand indicates that the factor $\partial\varphi/\partial q^i$ is to be inserted between noncommuting factors in the terms of $\mathbf{V}^i$ in such a way as to yield the commutator on the left. If now the differential operators occurring in $\mathbf{V}^i$ are peeled to the left and right, via integrations by parts, in such a manner that they no longer act on $\partial\varphi/\partial q^i$, then the integral takes the form

$$-i\int(\partial\varphi/\partial q^i)(\Psi_b{}^*\overleftrightarrow{\mathbf{V}}{}^i\Psi_a)dq = i\int\varphi\partial(\Psi_b{}^*\overleftrightarrow{\mathbf{V}}{}^i\Psi_a)/\partial q^i dq, \quad (9.2)$$

where $\overleftrightarrow{\mathbf{V}}{}^i$ denotes the result of the peeling process. Because of the arbitrariness of φ it follows that

$$[\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_b]^*\Psi_a-\Psi_b{}^*\mathfrak{H}\mathcal{C}(q,-i\partial/\partial q)\Psi_a = \partial(\Psi_b{}^*\overleftrightarrow{\mathbf{V}}{}^i\Psi_a)/\partial q^i. \quad (9.3)$$

In a similar manner we find

$$[\mathfrak{H}\mathcal{C}(R,-i\partial/\partial R)\Psi_b]^*\Psi_a-\Psi_b{}^*\mathfrak{H}\mathcal{C}(R,-i\partial/\partial R)\Psi_a = \partial(\Psi_b{}^*\overleftrightarrow{\mathbf{V}}\Psi_a)/\partial R, \quad (9.4)$$

where in this case we can give an explicit form for $\overleftrightarrow{V}$:

$$\overleftrightarrow{V}=\frac{i}{48\pi^2}\left(R^{-3/4}\frac{\overrightarrow{\partial}}{\partial R}R^{-1/4}-R^{-1/4}\frac{\overleftarrow{\partial}}{\partial R}R^{-3/4}\right). \quad (9.5)$$

The analog of (5.19) is now obvious, namely,

$$(\Psi_b,\Psi_a)=\int_\Sigma(\Psi_b{}^*\overleftrightarrow{V}\Psi_a dq+\Psi_b{}^*\overleftrightarrow{\mathbf{V}}{}^i\Psi_a dR d\Sigma_i), \quad (9.6)$$

where Σ is an appropriate surface in the R-q manifold and $d\Sigma_i$ is the directed surface element of its projection into q space. From Eqs. (9.3) and (9.4) it follows that

$$\partial(\Psi_b{}^*\overleftrightarrow{\mathbf{V}}\Psi_a)/\partial R+\partial(\Psi_b{}^*\overleftrightarrow{\mathbf{V}}{}^i\Psi_a)/\partial q^i=0, \quad (9.7)$$

whenever Ψ_a and Ψ_b are physical state functions satisfying the Hamiltonian constraint (7.1). Therefore the integral (9.6) is independent of Σ provided the boundary of Σ remains in a region where Ψ_a and Ψ_b vanish.

When the coherent dust filling the Friedmann universe is restricted to only one degree of freedom the inner product (9.6) reduces to

$$(\Psi_b,\Psi_a)=\int_\Sigma(\Psi_b{}^*\overleftrightarrow{V}\Psi_a dq-\Psi_b{}^*\overleftrightarrow{\mathbf{V}}\Psi_a dR), \quad (9.8)$$

where Σ is an appropriate contour in the R-q plane. The key word here is "appropriate." In analogy with our previous treatment of the manifold $\mathfrak{M}$ of 3-geometrics in the general theory, we may view the R-q plane as endowed with a natural metric determined by the structure of the functions $\mathfrak{H}$ and $\mathfrak{H}\mathcal{C}$. With respect to this metric the coordinates R and q are "timelike" and "spacelike", respectively. If the Hamiltonian constraint

(7.1) were an ordinary wave equation we would naturally adopt for the contour Σ a "spacelike" line such as $R=$ constant. However, just as in the general theory, so also here, "wave" propagation is not restricted to timelike directions. Indeed, from the lissajous traces of Fig. 1, it is evident that the Friedmann universe not only executes "timelike" and "spacelike" motions with impartiality, but even turns around and "moves" backward with respect to the "time" coordinate. The distinction between "timelike" and "spacelike" clearly does not have the same pervasive significance here as it does in ordinary wave theories.

If we were actually to choose, for Σ, a line $R=$ constant, we would obtain the useless result $(\Psi_b,\Psi_a)=0$ for all Ψ_a and Ψ_b. This is because all physically admissible state functions have non-negligible values only in a finite domain of the R-q plane. Hence any line $R=$ constant can be deformed into a line along which Ψ_a and Ψ_b effectively vanish, without affecting the value of the ntegral (9.8) at all. The same is true if Σ is a "timelike" curve which starts at $R=0$ and goes out to infinity in the R-q plane. Since any normalizabile superposition of the functions (7.20) vanishes at $R=0$ at least as fast as $R^{7/4}$, such a curve can also be deformed into one along which Ψ_a and Ψ_b vanish, without affecting (9.8).

How then shall we choose Σ? The answer is to be found in the conservation laws (6.13), (7.38), and (9.7). From our analysis of the Lissajous traces of Fig. 1 it is evident that probability flows in a closed finite circuit in the R-q plane. Σ must therefore be a *finite* curve, chosen so as to intersect a unidirectional unit flux of probability of each of the two functions Ψ_a and Ψ_b. This means that Eq. (9.8) can be used to define inner products only when Ψ_a and Ψ_b both have the form of good packets. If they do not have this form or if they fall into the degenerate category depicted in Figs. 1(a) and 1(e), then some other representation must be employed. An analogous condition must hold in the general theory if Eq. (5.19) is to be valid. Whenever the condition is violated the usefulness of Wheeler's "metric representation" diminishes.

It is not difficult to show that Eq. (9.8) indeed yields an acceptable value for the inner product under the required conditions. The case in which the two packets do not overlap (except at intersections) may be disposed of at once; (Ψ_b, Ψ_a) then clearly vanishes. The only case which need concern us is that in which the two packets overlap, at least partially, throughout the entire length of their trajectories. The **J** values at their "peaks" then differ negligibly compared to the spread of values contained in their superpositions. As our initial contour we shall choose an invariant segment $\Sigma_{r,\mathrm{r}}$ corresponding to the average of the peak **J** values. Since the packets are "good" we know that Ψ_a and Ψ_b vanish at its endpoints. Suppose $\Sigma_{r,\mathrm{r}}$ intersects a NE-SW branch of the lissajous figure formed by the packet traces [e.g., the segment $\Sigma_{1,2}$ shown in Fig. 1(c)], and suppose the orientation of both packets is the same, say NE, at the point of intersection with $\Sigma_{r,\mathrm{r}}$. Then it is only the W^+, $\mathbf{W}^+$ branch of each packet which interferes constructively along $\Sigma_{r,\mathrm{r}}$. Approximating Ψ_a and Ψ_b by their WKB forms, keeping only those parts which refer to the branch in question, and taking note of the inequalities (7.33), we have

$$(\Psi_b,\Psi_a)=\int_{\Sigma_{r,\mathrm{r}}}(\Psi_b{}^*\overleftrightarrow{V}\Psi_a dq-\Psi_b{}^*\overleftrightarrow{\mathbf{V}}\Psi_a)dR$$

$$=\sum_{n,n'}{}' b_{n'}{}^*a_n\int_{\Sigma_{r,\mathrm{r}}}(\omega_{n'}\omega_n\Delta n'\Delta n/4\pi^2|V_{n'}V_n\mathbf{V}_{\mathrm{n}'}\mathbf{V}_{\mathrm{n}}|)^{1/2}\{[\exp(iW_{n'}{}^+)\overleftrightarrow{V}\exp(-iW_n{}^+)]$$

$$\times\exp[-i(\mathbf{W}_{\mathrm{n}'}{}^+-\mathbf{W}_{\mathrm{n}}{}^+)]dq-\exp[i(W_{n'}{}^+-W_n{}^+)][\exp(-i\mathbf{W}_{\mathrm{n}'}{}^+)\overleftrightarrow{\mathbf{V}}\exp(i\mathbf{W}_{\mathrm{n}}{}^+)]dR\}\,, \quad (9.9)$$

where the b's are the coefficients of the expansion of Ψ_b:

$$\Psi_b=\Sigma_n{}'b_n\Psi_n. \quad (9.10)$$

Having dropped the parts of the WKB functions which refer to irrelevant branches, we may now extend the contour $\Sigma_{r,\mathrm{r}}$ until its ends coincide with points at which maximum destructive interference (of the W^+, $\mathbf{W}^+$ parts) occurs [e.g., the points A and B in Fig. 1(c)]. The contour then corresponds to a proper time lapse of $T/2\Delta n$ instead of $T/4\Delta n$, and spans just the right number of nodes so that the integral in (9.9) vanishes except when $n=n'$. Expression (9.9) accordingly reduces to

$$(\Psi_b,\Psi_a)=\sum_n{}' b_n{}^*a_n\int_{\Sigma}[(\omega_{\mathrm{n}}\Delta\mathbf{n}/2\pi|\mathbf{V}_{\mathrm{n}}|)dq-(\omega_n\Delta n/2\pi|V_n|)dR]$$

$$=\sum_n{}' b_n{}^*a_n[(\omega_{\mathrm{n}}\mathbf{T}/4\pi)+(\omega_nT/4\pi)], \quad (9.11)$$

the positive sign of the final bracketed factor being obtained by appropriately orienting the original contour $\Sigma_{r,\mathrm{r}}$. The contour $\Sigma_{1,2}$ in Fig. 1(c) shows the correct orientation. If the direction of "motion" of the packets

is reversed then the orientation of the contour must be reversed.

For good packets the frequencies ω_n and $\boldsymbol{\omega}_n$ remain sensibly constant and equal to the peak frequencies $2\pi/T$ and $2\pi/\mathbf{T}$, respectively, over the range of effective n values in the sum (9.11). Therefore we have

$$(\Psi_b,\Psi_a)=\sum_n{}' b_n^* a_n\,, \tag{9.12}$$

which is just the accepted definition. By virtue of the Σ invariance of expression (9.8) the contour may now be displaced to any location, including turning points where the WKB approximation breaks down. All that is required is that the contour cut each wave packet only once and that Ψ_a and Ψ_b vanish at its endpoints. We therefore have quite generally

$$\int_\Sigma (\Psi_b^* \overleftrightarrow{\nabla} \Psi_a dq - \Psi_b^* \overleftrightarrow{\nabla} \Psi_a dR) \approx \sum_n{}' b_n^* a_n\,, \tag{9.13}$$

the relation "$\approx$" tending toward "$=$" the more precisely defined the packets Ψ_a and Ψ_b become.

If, in the above derivation, the two packets had been oppositely oriented then one of the pairs of functions W^+, $\mathbf{W}^+$ in Eq. (9.9) would have had to be changed to W^-, $\mathbf{W}^-$, and the integral, with the extended contour, would have vanished even when $n=n'$. This, however, does not conflict with (9.12) since, in the case of oppositely oriented packets, the relative phases of a_n and b_n vary so rapidly with n that the inner product vanishes anyway.

An entirely similar analysis can be carried out in the R-$\mathfrak{r}$ and τ-q planes. Here the inner product integrals are given by

$$(\Phi_b,\Phi_a)=\int_{\Sigma_{r,\,\mathfrak{t}}} (\Phi_b^* \overleftrightarrow{\nabla} \Phi_a d\tau - \Phi_b^* \Phi_a dR)\,, \tag{9.14}$$

$$(\boldsymbol{\Phi}_a,\boldsymbol{\Phi}_b)=\int_{\Sigma_{t,\,\mathfrak{r}}} (\boldsymbol{\Phi}_b^* \boldsymbol{\Phi}_a dq - \boldsymbol{\Phi}_b^* \overleftrightarrow{\nabla} \boldsymbol{\Phi}_a d\tau)\,, \tag{9.15}$$

which reduce to the familair $\int \Phi_b^* \Phi_a dR$ and $\int \boldsymbol{\Phi}_b^* \Phi_a dq$ when the contours are distorted to $\mathfrak{r}=$ constant and $\tau=$ constant, respectively. If the ranges of integration of the latter integrals are extended to include all permissible R and q values, the integrals must be divided by Δn and $\Delta\mathbf{n}$, respectively, because of the presence of the ghost packets.

The above analysis permits us to adopt a new viewpoint regarding the Cauchy problem for the Hamiltonian constraint. At the end of Sec. 6 it was conjectured that by virtue of the boundary condition (6.31) [(7.15) in the present context] the state function will be determined everywhere as soon as it is specified on a hypersurface. This is very easy to demonstrate in the present context, because of the separability of Eq. (7.12), which permits the eigenfunctions Ψ_n to be expressed in the product form

$$\Psi_n(R,q)=(2\pi\Delta n/\boldsymbol{\omega}_\mathbf{n})^{1/2} X_n(R)\mathbf{X}_\mathbf{n}(q)\,, \tag{9.16}$$

where X_n and $\mathbf{X}_\mathbf{n}$ have the WKB approximations

$$X_n=(\omega_n/2\pi|\,V_n\,|)^{1/2} \times[\exp(-iW_n{}^+)+\exp(-iW_n{}^-)]\,, \tag{9.17}$$

$$\mathbf{X}_\mathbf{n}=(\boldsymbol{\omega}_\mathbf{n}/2\pi|\,\mathbf{V}_\mathbf{n}\,|)^{1/2} \times[\exp(i\mathbf{W}_\mathbf{n}{}^+)+\exp(i\mathbf{W}_\mathbf{n}{}^-)]\,, \tag{9.18}$$

and satisfy the orthonormality conditions

$$\int_0^\infty X_n^* X_{n'} dR=\delta_{nn'}\,, \quad \int \mathbf{X}_\mathbf{n}^* \mathbf{X}_{\mathbf{n}'} dq=\delta_{\mathbf{n}\mathbf{n}'}\,. \tag{9.19}$$

Thus, we may write

$$\Psi(R,q)=\sum_n{}' X_n(R)\,\mathbf{X}_\mathbf{n}(q)\int \mathbf{X}_\mathbf{n}^*(q') \times\Psi(R',q')dq'/X_n(R') \tag{9.20a}$$

$$=\sum_n{}' X_n(R)\,\mathbf{X}_\mathbf{n}(q)\int_0^\infty X_n^*(R') \times\Psi(R',q')dR'/\mathbf{X}_\mathbf{n}(q')\,, \tag{9.20b}$$

which express Ψ everywhere in terms of its values on the infinite contour $R=R'$ or on the infinite contour $q=q'$.

However, when the function Ψ has the form of a wave packet Ψ_a, it should be equally possible to determine it completely by knowing its value over a *finite* contour Σ which intersects the packet only once. That this is indeed the case follows from the fact that for a good packet the integrals

$$\int_\Sigma (\Psi_n^* \overleftrightarrow{\nabla} \Psi_a dq - \Psi_n^* \overleftrightarrow{\nabla} \Psi_a dR)\,, \tag{9.21}$$

for all n, may to a high degree of accuracy be replaced simply by

$$\int_\Sigma (\Psi_n^*|\,V_n\,|\,\Psi_a dq - \Psi_n^*|\,\mathbf{V}_\mathbf{n}\,|\,\Psi_a dR)\,. \tag{9.22}$$

When the packet is good these integrals have non-negligible values only for a restricted range of n values centered on the peak of the packet, over which $|\,V_n\,|$ varies slowly. Let the peak n values be determined by evaluating (9.22) for all n. Then let the contour be extended until its ends reach points of maximum destructive interference, as determined by the peak n value and the slope of the packet branch where it intersects Σ.[49] Suppose the slope is NE-SW, corresponding to the classical functions W^+, $\mathbf{W}^+$ or W^-, $\mathbf{W}^-$. Denote by

[49] It may be objected that in choosing Σ to intersect the packet along a definite branch we are assuming some preliminary knowledge about the approximate "location" of the packet. This preliminary knowledge, however, differs in no fundamental respect from the knowledge which we always have in other more familiar instances, e.g., that a given particle is "somewhere in the laboratory."

Ψ_n^{++} and Ψ_n^{--} the parts of (the WKB approximation to) Ψ_n associated with these functions. Now compute the integrals

$$\int_{\Sigma'} (\Psi_n^{\pm\pm *} | V_n | \Psi_a dq - \Psi_n^{\pm\pm *} | \mathbf{V_n} | \Psi_a dR), \quad (9.23)$$

where Σ' denotes the extended contour. Of these integrals, only those corresponding to the previously determined significant n values need be included, and of these, in turn, only those corresponding to a definite choice of signs (either $++$ or $--$) will have non-negligible values (corresponding to a definite packet orientation, which becomes thus determined). The values in question are just the amplitudes a_n of Eq. (8.1) and from these the entire state function can be constructed. This means that for a good packet the Cauchy data are not only the same as for the ordinary Schrödinger equation but are also effectively taken from a compact domain of "configuration space."

10. DISCUSSION AND SPECULATION

Perhaps the most impressive fact which emerges from a study of the quantum theory of gravity is that it is an extraordinarily economical theory. It gives one just exactly what is needed in order to analyze a particular physical situation, but not a bit more. Thus it will say nothing about time unless a clock to measure time is provided, and it will say nothing about geometry unless a device (either a material object, gravitational waves, or some other form of radiation) is introduced to tell when and where the geometry is to be measured.[50] In view of the strongly operational foundations of both the quantum theory and general relativity this is to be expected. When the two theories are united the result is an operational theory *par excellence*.[51]

The economy of quantum gravidynamics is also revealed in the manner in which the formalism determines its own interpretation. We have seen how the Hamiltonian constraint, in the case of a finite universe, forces us to abandon all use of externally imposed coordinates (in particular x^0) and to look instead for an internal description of the dynamics. We have seen how the metric structure of the manifold $\mathfrak{M}$, with its frontier of infinite curvature, suggests a natural boundary condition for the state functional, which may simplify the Cauchy data needed to specify a state. And finally, if it be permitted to extend the results of our study of the Friedmann model to the general case, we have learned how (and when) to use the inner-product definition (5.19), by recognizing that probability flows in closed circuits in $\mathfrak{M}$.

This "principle of self-determination," which permeates even classical general relativity, has been elevated to the rank of a universal principle by Everett,[52] who applies it to ordinary nonrelativistic quantum mechanics. As conventionally formulated quantum mechanics comes in two packages: (1) formalism and (2) interpretation based on the existence of a classical level. According to Everett, package 2 should be thrown away. Quantum mechanics is a theory which attempts to describe in mathematical language a situation in which chance is not a measure of our ignorance but is *absolute*. Naturally it cannot avoid introducing things like wave functions which undergo repeated fission, corresponding to the many possible outcomes of a given physical process. According to Everett, the wave function nonetheless provides a faithful representation of reality; it is the universe itself which splits.

To those who would immediately object that they do not feel themselves split, Everett replies that this only confirms the theory; they are not supposed to feel it. Everett allows into the theory only those elements which are in the formalism itself, namely, a Hilbert space, a Hamiltonian, and a Schrödinger equation for vectors in the Hilbert space. From these meager beginnings one can show, by standard arguments, that the wave function for a Hamiltonian which, in conventional language, would be described as that of a system coupled to an apparatus, evolves into a superposition of vectors representing the possible values of some system variable together with corresponding apparatus "readings." Moreover, if the "measurement" is repeated on a large number N of identically prepared systems, the final superposition consists of vectors representing various possible sets of N values for the system variable together with corresponding apparatus "memory sequences" which record these values. No interpretation of the mathematics is admitted up to this point; in particular no *a priori* interpretation is given to the coefficients in the final superposition.

Now let the coefficients in the final superposition in the case of a single system be denoted by c_n. Then the coefficients in the case of N systems will be products of c's. It can be shown[53] that if one removes from the final N-system superposition all those vectors which correspond to memory sequences in which the recorded values of the system variable fail to meet the standard

[50] For details on the quantum theory of measurement in general relativity see B. S. DeWitt, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (John Wiley & Sons, Inc., New York, 1962).

[51] A notable failure to recognize this fact is to be found in P. W. Bridgman, in *Albert Einstein, Philosopher Scientist*, edited by P. A. Schilpp (Tudor Publishing Company, New York, 1949). Bridgman's confusion, which is shared by others, stems from the fact that in traditional formulations of general relativity one speaks about things, such as curvilinear coordinates, which have no operationally defined reality. This confusion would have been eliminated had modern coordinate-independent formulations of differential geometry been available in 1916. Modern methods make it plain that coordinate systems are precisely what general relativity is *not* talking about. General relativity is concerned with those attributes of physical reality which are coordinate-independent and is the rock on which present day emphasis on invariance principles will ultimately stand or fall.

[52] H. Everett, III, Rev. Mod. Phys. **29**, 454 (1957).

[53] N. R. Graham, Ph.D. thesis, University of North Carolina (unpublished).

requirements for a *random sequence with probabilities* $|c_n|^2$, to any arbitrary, but fixed, degree of accuracy, the resulting wave function is indistinguishable from the true final wave function in the limit $N \to \infty$. By "indistinguishable" we mean that the difference between it and the true wave function has vanishing norm.

The probability interpretation of quantum mechanics thus emerges from the formalism itself. Nonrandom memory sequences are "of measure zero" in the final superposition, in the limit $N \to \infty$. Each automaton (i.e., apparatus *cum* memory sequence) in the superposition sees the world obey the familiar quantum laws. However, there exists no outside agency which can designate which "branch" of the superposition is to be regarded as the *real* world. All are equally real, and yet each is unaware of the others. Thus if, within a given branch, an automaton, which has measured a given variable without changing it, subsequently checks his original observation, his memory sequence will not fail him. He will get his original value, and not that of some other branch. Moreover, if he communicates with another automaton who has simultaneously made the same measurement, their results will agree, which means that the two are in the same branch and that communication between different branches is impossible. The automaton therefore never feels himself split.

Everett's view of the world is a very natural one to adopt in the quantum theory of gravity, where one is accustomed to speak without embarassment of the "wave function of the universe." It is possible that Everett's view is not only natural but essential. For example, if the Hamiltonian constraint possesses only a single solution, so that the wave function for the universe is unique, then some conception like Everett's would appear to be needed in order to assess the physical significance of such uniqueness.[54]

In our discussion of the Friedmann model we assumed that the operator $\mathfrak{H}+\mathfrak{H}$ possesses a highly degenerate zero eigenvalue. How plausible is this assumption in the case of the actual world? In the case of the Friedmann model we were obliged to match the internal dynamics of the dust with that of the universe as a whole, with one hundred percent precision. Let us try to be a little more realistic. Suppose we replace the dust by a gas of noninteracting scalar bosons, but still maintain a rigid spherical geometry. Then we have an infinity of degrees of freedom. However, this infinity is *discrete*, because the universe is finite. Moreover, and this is important, there can never be more than a finite number of field quanta present in the state vector superposition, since the total energy (of the bosons) cannot be infinite. This is true even if the bosons are massless, since there is no infrared catastrophe in a finite world.

Now it is not at all difficult to verify that the Hamiltonian $\mathfrak{H}$ in this case does not match $\mathfrak{H}$ in any obviously commensurable way.[55] For each choice of boson quantum numbers $\mathfrak{H}$ becomes a well-defined function of R, and the combination $\mathfrak{H}+\mathfrak{H}$ has a well-defined spectrum. But only by the sheerest accident does this spectrum include zero. All the evidence points to the fact that the complete spectrum of $\mathfrak{H}+\mathfrak{H}$, although discrete, is everywhere dense on the real line and does not condense into a set of finitely separated, infinitely degenerate levels. A similar situation holds with vector bosons and with fermions, and it seems hardly likely that the switching on of interactions between the particles will change the picture.

One might now suggest that we look for a way out of this predicament by relaxing the spherical rigidity restriction. However, this would merely correspond to the introduction of a gas of interacting tensor bosons, i.e., gravitons. It therefore appears that the same situation holds even in the general theory and that the Hamiltonian constraint of the real world may indeed have only one solution.[56]

If the state functional of the universe is unique how can we interpret it? In the case of the Friedmann model a single eigenfunction Ψ_n certainly has no resemblance to the real world, nor to any other reasonable world for that matter. A plot of $|\Psi_n|^2$ in the R-q plane looks like a lot of bumps separated from one another by a rectangular array of nodal lines, certainly nothing like a lissajous figure. However, suppose an extra term were added to the Hamiltonian $\mathfrak{H}+\mathfrak{H}$ which had the effect of strongly correlating the phase of the coherent particle clock with the phase of the universe, without changing either the (zero) eigenvalue or the quantum numbers. Then the Hamiltonian would no longer be separable and the nodal lines of $|\Psi_n|^2$ would no longer form a rectangular array. The bumps would instead tend to cluster around the Lissajous figure having the favored correlation, the figure itself now being somewhat distorted due to the correlation interaction, but still definitely recognizable.

The Hamiltonian of the real world is highly nonseparable, and there is a high degree of correlation among its infinity of modes. This must express itself as a kind of "condensation" of the state functional into components having many of the attributes of the quasiclassical Friedmann packets.[57] At the same time, because of the size of the universe, we know that the "Everett process" must be occurring on a lavish scale: The quasiclassical components of the universal state

[54] See J. A. Wheeler, The Monist 47, 40 (1962).

[55] M. Miketinac (private communication).

[56] The spectrum of $\mathfrak{H}+\mathfrak{H}$, or of $\mathfrak{H}$ itself, can be shifted by the introduction of a "cosmological term" in the Einstein Lagrangian. If this spectrum is actually everywhere dense then we have the amusing result that a minute change in the cosmological constant can produce an enormous change in the zero-eigenvalue eigenvector and hence in the physical properties of the universe.

[57] If the state functional of the universe is unique then it is no longer possible or even meaningful to apply the inner-product definition (5.19) to the state functional as a whole. However, it might still be applied, in some reduced form, to its quasiclassical components.

functional must be constantly splitting into a stupendous number of branches, all moving in parallel without interfering with one another except insofar as quantum Poincaré cycles allow rare anomalies to occur. According to the Everett interpretation each branch corresponds to a possible world-as-we-actually-see-it.

We have seen that the Friedmann packets in the R-q plane do not ultimately spread in "time"; every expansion-contraction cycle is exactly like every other. Unless some form of leakage to other channels occurs (e.g., transitions to different 3-space topologies) the same must be true for the real universe (assuming it to be closed and finite). In the absence of such channels there could be only *one* expansion-contraction cycle, repeated over and over again, like a movie film, throughout eternity, the monotony of which would be alleviated only by the infinite variety to be found among the multitude of simultaneous parallel worlds all executing the cycle together. Such a conclusion holds, in fact, regardless of whether the total state functional is unique or not.

A question naturally arises in regard to entropy. Within a given branch of the universal state functional the entropy would be observed (by appropriate automata) to increase with time.[58] It might be supposed that this increase would continue only during the expansion phase of the universe and that it would reverse itself during the contraction phase. This is not so, for one has only to remember that the length of a Poincaré cycle for even a small part of the universe is vastly longer than a rebound cycle, and hence except for a vanishingly small fraction of branches the entropy must continue to increase (at least locally) until final collapse is reached, at which point the very concepts of entropy and probability, as well as time itself, cease to have meaning.

However, if the operator $\mathfrak{IC}+\mathfrak{K}$ is time-reversal invariant, and if its zero eigenvalue is nondegenerate, then the state functional of the universe is necessarily time-symmetric. This means that for every Everett branch in which entropy increases with time there must be another in which entropy decreases with time. To an observer in the second branch "time" in fact appears to be "flowing" in the opposite sense. Because of the extreme sensitivity of the state functional to slight changes in the operator $\mathfrak{IC}+\mathfrak{K}$ (see Ref. 56) it is difficult to say how these conclusions must be modified if, as recent experiments suggest, the real world is not invariant under time reversal. However, the world being as complicated (and hence ergodic) as it is, it is still quite possible that there is no *preferred* direction in time. The ensemble of Everett branches in which time has a given direction of flow may very well be balanced by another ensemble in which time flows oppositely, so that reality as a whole possesses no over-all time orientation despite the absence of time-reversal invariance.

[58] Each branch corresponds to a pure state in the traditional sense. This does not, however, prevent the assignment of an effective entropy to it. For a sufficiently complicated system even a pure state may be assigned an entropy based on the coarse-grained properties of the state rather than on an ensemble average. In the classical theory this is illustrated by computer calculations of n-body systems. Even though the position and velocity of every body is known, the system as a whole possesses effective thermodynamical properties, the determination of which is in fact the goal of the computation.

ACKNOWLEDGMENTS

I wish to express my warmest gratitude for the kindness of Professor Robert Oppenheimer and Professor Carl Kaysen in extending to me the hospitality of the Institute for Advanced Study.

APPENDIX A: THE MANIFOLD M

M is defined as the 6-dimensional space of "points" $\{\gamma_{ij}\}$ having as covariant and contravariant metric tensors, respectively, the expressions

$$G^{ijkl}=\tfrac{1}{2}\gamma^{1/2}(\gamma^{ik}\gamma^{jl}+\gamma^{il}\gamma^{jk}-2\gamma^{ij}\gamma^{kl}), \tag{A1}$$

$$G_{ijkl}=\tfrac{1}{2}\gamma^{-1/2}(\gamma_{ik}\gamma_{jl}+\gamma_{il}\gamma_{jk}-\gamma_{ij}\gamma_{kl}), \tag{A2}$$

satisfying

$$G_{ijab}G^{abkl}=\delta_{ij}{}^{kl}. \tag{A3}$$

Index pairs may, if desired, be mapped into single indices according to the rules

$$\begin{aligned}&\gamma_{11}=\gamma^1, \quad \gamma_{22}=\gamma^2, \quad \gamma_{33}=\gamma^3,\\ &\gamma_{23}=2^{-1/2}\gamma^4, \quad \gamma_{31}=2^{-1/2}\gamma^5, \quad \gamma_{12}=2^{-1/2}\gamma^6,\\ &\delta_{ij}{}^{kl}\rightarrow\delta^{\Gamma}{}_{\Delta}, \quad \Gamma,\,\Delta=1\cdots6, \text{ etc.}\end{aligned} \tag{A4}$$

although this is seldom convenient or necessary.

By straightforward computation one may verify the variational law

$$G_{ijkl}\delta G^{ijkl}=-\gamma^{ij}\delta\gamma_{ij}, \tag{A5}$$

from which it may be inferred that

$$G\equiv\det(G^{ijkl})=-a\gamma^{-1}, \tag{A6}$$

where a is some constant. In the special case $\gamma_{ij}=\delta_{ij}$ one easily finds that the roots of G^{ijkl} are $-\frac{1}{2}$, 1, 1, 1, 1, 1, from which it follows that $a=\frac{1}{2}$ and that the signature of M is $-+++++$. The components γ_{ij} evidently are "good" coordinates in M as long as $\gamma\neq0$.

If the ζ of Eq. (5.7) is chosen as a new coordinate then the surfaces of constant ζ have orthogonal trajectories whose tangent vectors are proportional to

$$\partial\zeta/\partial\gamma_{ij}=\tfrac{1}{4}\zeta\gamma^{ij}, \tag{A7}$$

or, in contravariant form,

$$G_{ijkl}\partial\zeta/\partial\gamma_{kl}=-\tfrac{4}{3}\zeta^{-1}\gamma_{ij}. \tag{A8}$$

If the orthogonal trajectories themselves are labeled by a set of five additional new coordinates ζ^A, $A=1\cdots5$, then

$$\gamma^{ij}\partial\gamma_{ij}/\partial\zeta^A=4\zeta^{-1}\partial\zeta/\partial\zeta^A=0, \tag{A9}$$

and, moreover, $\partial\gamma_{ij}/\partial\zeta$ must satisfy

$$G^{ijkl}\frac{\partial\gamma_{ij}}{\partial\zeta}\frac{\partial\gamma_{kl}}{\partial\zeta^A}=0. \tag{A10}$$

From these facts one may infer

$$G^{ijkl}\frac{\partial\gamma_{ij}}{\partial\zeta}\frac{\partial\gamma_{kl}}{\partial\zeta}=\left(G_{ijkl}\frac{\partial\zeta}{\partial\gamma_{ij}}\frac{\partial\zeta}{\partial\gamma_{kl}}\right)^{-1}=-1, \tag{A11}$$

$$\partial\gamma_{ij}/\partial\zeta=\tfrac{4}{3}\zeta^{-1}\gamma_{ij}, \tag{A12}$$

$$\gamma_{ij}\partial\zeta^A/\partial\gamma_{ij}=0, \tag{A13}$$

$$G^{ijkl}\frac{\partial\gamma_{ij}}{\partial\zeta^A}\frac{\partial\gamma_{kl}}{\partial\zeta^B}=\gamma^{1/2}\gamma^{ik}\gamma^{jl}\frac{\partial\gamma_{ij}}{\partial\zeta^A}\frac{\partial\gamma_{kl}}{\partial\zeta^B}, \tag{A14}$$

from which the metric (5.8) follows.

The above relations yield the following useful identities:

$$\mathrm{tr}(\gamma_{,A}\partial\zeta^B/\partial\gamma)=\delta_A{}^B, \tag{A15}$$

$$\mathrm{tr}(\gamma\partial\zeta^A/\partial\gamma)=0, \tag{A16}$$

$$\mathrm{tr}(\gamma^{-1}\gamma_{,A})=0, \tag{A17}$$

$$\mathrm{tr}[\gamma^{-1}(\gamma_{,AB}-\gamma_{,A}\gamma^{-1}\gamma_{,B})]=0, \tag{A18}$$

$$\frac{\partial\gamma_{ij}}{\partial\zeta^A}\frac{\partial\zeta^A}{\partial\gamma_{kl}}=\delta_{ij}{}^{kl}-\tfrac{1}{3}\gamma_{ij}\gamma^{kl}, \tag{A19}$$

$$\mathrm{tr}(\gamma_{,A}\mathbf{M})\mathrm{tr}(\mathbf{N}\partial\zeta^A/\partial\gamma)=\tfrac{1}{2}\,\mathrm{tr}(\mathbf{MN}+\mathbf{MN}^{\sim})-\tfrac{1}{3}\,\mathrm{tr}(\gamma\mathbf{M})\mathrm{tr}(\gamma^{-1}\mathbf{N}), \tag{A20}$$

$$\mathrm{tr}(\gamma_{,AB}\mathbf{M})\mathrm{tr}(\mathbf{N}\partial\zeta^B/\partial\gamma)+\mathrm{tr}(\gamma_{,B}\mathbf{M})\mathrm{tr}[\mathbf{N}(\partial\zeta^B/\partial\gamma)_{,A}]=-\tfrac{1}{3}\,\mathrm{tr}(\gamma_{,A}\mathbf{M})\mathrm{tr}(\gamma^{-1}\mathbf{N})+\tfrac{1}{3}\,\mathrm{tr}(\gamma\mathbf{M})\mathrm{tr}(\gamma^{-1}\gamma_{,A}\gamma^{-1}\mathbf{N}), \tag{A21}$$

$$\mathrm{tr}[\gamma_{,A}\mathbf{M}(\partial\zeta^A/\partial\gamma)\mathbf{N}]=\tfrac{1}{2}\,\mathrm{tr}(\mathbf{MN}^{\sim})+\tfrac{1}{2}\,\mathrm{tr}\mathbf{M}\,\mathrm{tr}\mathbf{N}-\tfrac{1}{3}\,\mathrm{tr}(\gamma\mathbf{M}\gamma^{-1}\mathbf{N}). \tag{A22}$$

Equations (A20) and (A22) follow from (A19), while Eqs. (A18) and (A21) are obtained from (A17) and (A20), respectively, by differentiation. **M** and **N** are arbitrary 3×3 matrices.

Using these identities and remembering the cyclic invariance of the trace, it is easy to compute the following:

$$\bar{G}^{AB}=\mathrm{tr}[\gamma(\partial\zeta^A/\partial\gamma)\gamma(\partial\zeta^A/\partial\gamma)], \qquad \bar{G}_{AC}\bar{G}^{CB}=\delta_A{}^B, \tag{A23}$$

$$\bar{\Gamma}_{ABC}\equiv\tfrac{1}{2}(\bar{G}_{AC,B}+\bar{G}_{BC,A}-\bar{G}_{AB,C})=\tfrac{1}{2}\mathrm{tr}[\gamma_{,C}\gamma^{-1}\times(-\gamma_{,A}\gamma^{-1}\gamma_{,B}-\gamma_{,B}\gamma^{-1}\gamma_{,A}+2\gamma_{,AB})\gamma^{-1}], \tag{A24}$$

$$\bar{\Gamma}_{AB}{}^C\equiv\bar{G}^{CD}\bar{\Gamma}_{ABD}=\mathrm{tr}[(-\gamma_{,A}\gamma^{-1}\gamma_{,B}+\gamma_{,AB})\partial\zeta^C/\partial\gamma], \tag{A25}$$

$$\bar{R}_{ABC}{}^D\equiv\bar{\Gamma}_{BC}{}^D{}_{,A}-\bar{\Gamma}_{AC}{}^D{}_{,B}+\bar{\Gamma}_{BC}{}^E\bar{\Gamma}_{AE}{}^A-\bar{\Gamma}_{AC}{}^E\bar{\Gamma}_{BE}{}^D=\mathrm{tr}[\gamma_{,C}\gamma^{-1}(\gamma_{,A}\gamma^{-1}\gamma_{,B}-\gamma_{,B}\gamma^{-1}\gamma_{,A})\times\partial\zeta^D/\partial\gamma], \tag{A26}$$

$$\bar{R}_{ABCD}\equiv\bar{R}_{ABC}{}^E\bar{G}_{ED}=\mathrm{tr}[\gamma^{-1}\gamma_{,D}\gamma^{-1}\gamma_{,C}\gamma^{-1}\times(\gamma_{,A}\gamma^{-1}\gamma_{,B}-\gamma_{,B}\gamma^{-1}\gamma_{,A})], \tag{A27}$$

$$\bar{R}_{AB}\equiv\bar{R}_{CAB}{}^C=-\tfrac{3}{2}\bar{G}_{AB}. \tag{A28}$$

The corresponding quantities in the full manifold M are

$$\Gamma_{AB}{}^C=\bar{\Gamma}_{AB}{}^C, \tag{A29}$$

$$\Gamma_{AB}{}^0=(3/32)\zeta\bar{G}_{AB}, \tag{A30}$$

$$\Gamma_{A0}{}^B=\zeta^{-1}\delta_A{}^B, \tag{A31}$$

$$\Gamma_{A0}{}^0=\Gamma_{00}{}^A=\Gamma_{00}{}^0=0, \tag{A32}$$

$$R_{ABC}{}^D=\bar{R}_{ABC}{}^D-(3/32)(\bar{G}_{AC}\delta_B{}^D-\bar{G}_{BC}\delta_A{}^D), \quad +(\text{all other components vanish}) \tag{A33}$$

$$R_{AB}=\bar{R}_{AB}+\tfrac{3}{8}\bar{G}_{AB}=-(9/8)G_{AB}, \tag{A34}$$

$$R_{A0}=R_{00}=0, \tag{A35}$$

$$^{(6)}R=-60\zeta^{-2}, \tag{A36}$$

where the index 0 is used for components in the direction of the "timelike" coordinate ζ.

The geodesic equation in $\bar{M}$ is obtained directly from (A25):

$$0=\frac{d^2\zeta^A}{d\bar{s}^2}+\bar{\Gamma}_{BC}{}^A\frac{d\zeta^B}{d\bar{s}}\frac{d\zeta^C}{d\bar{s}}=\mathrm{tr}\left[\frac{\partial\zeta^A}{\partial\gamma}\left(\frac{d^2\gamma}{d\bar{s}^2}-\frac{d\gamma}{d\bar{s}}\gamma^{-1}\frac{d\gamma}{d\bar{s}}\right)\right]. \tag{A37}$$

This reduces to (5.11) upon multiplication with $\gamma_{,A}$ and use of (A17), (A18), and (A19).

When one stays within the manifold $\bar{M}$ it is convenient to map the matrices γ into matrices **a** of unit determinant:

$$\mathbf{a}\equiv\gamma^{-1/3}\gamma. \tag{A38}$$

In the solution (5.12) of the geodesic equation, γ may be replaced by **a** provided **M** is restricted to have unit determinant. To find the geodesic connecting two matrices $\mathbf{a}_1$ and $\mathbf{a}_2$, let $\bar{s}=0$ at $\mathbf{a}_1$ and choose **M** in the form

$$\mathbf{M}=\mathbf{d}_1{}^{1/2}\mathbf{O}, \tag{A39}$$

where **O** is an orthogonal matrix which diagonalizes $\mathbf{a}_1$ and $\mathbf{d}_1{}^{1/2}$ is a diagonal square root of the resulting diagonal matrix. Then if $\bar{s}_{12}$ is the distance between $\mathbf{a}_1$ and $\mathbf{a}_2$ the matrix **N** satisfies

$$\mathbf{N}\bar{s}_{12}=\ln(\mathbf{d}_1{}^{-1/2}\mathbf{O}\mathbf{a}_2\mathbf{O}^{\sim}\mathbf{d}_1{}^{-1/2}). \tag{A40}$$

From the condition $tr\mathbf{N}^2=1$ one obtains

$$\bar{s}_{12}{}^2=\mathrm{tr}[\ln(\mathbf{d}_1{}^{-1/2}\mathbf{O}\mathbf{a}_2\mathbf{O}\mathbf{d}_1{}^{-1/2})]^2=\mathrm{tr}[\ln(\mathbf{a}_1{}^{-1}\mathbf{a}_2)]^2, \tag{A41}$$

which permits the matrix **N** itself to be determined. The logarithm of a matrix is an effectively unambiguous

concept, and the law of cyclic invariance of the trace applies to transcendental matrix functions as well as to rational functions. The uniqueness of the geodesic (5.12) is easily checked by noting that the matrix $\mathbf{O}$ is determined up to a transformation of the form $\mathbf{O}'=\mathbf{PO}$ where $\mathbf{P}$ is either a permutation matrix, if the roots of $\mathbf{a}_1$ are all distinct, or a more general orthogonal matrix, if some of the roots coincide. Such a transformation leaves (5.12) invariant. It is also easy to verify that it does not matter which of the eight possible square roots of $\mathbf{d}_1$ is chosen for $\mathbf{d}_1^{1/2}$.

The geodesic equations in M take the form

$$0=\frac{d^2\zeta}{ds^2}+(3/32)\zeta\left(\frac{d\bar{s}}{ds}\right)^2, \tag{A42}$$

$$0=\frac{d^2\zeta^A}{ds^2}+\bar{\Gamma}_{BC}{}^A\frac{d\zeta^B}{ds}\frac{d\zeta^C}{ds}+2\zeta^{-1}\frac{d\zeta^A}{ds}\frac{d\zeta}{ds}, \tag{A43}$$

where

$$\left(\frac{d\bar{s}}{ds}\right)^2=\bar{G}_{AB}\frac{d\zeta^A}{ds}\frac{d\zeta^B}{ds}. \tag{A44}$$

Differentiating the latter equation and making use of (A43) multiplied by $\bar{G}_{AB}d\zeta^B/ds$, one finds

$$\frac{d^2\bar{s}}{ds^2}=-\frac{2}{\zeta}\frac{d\bar{s}}{ds}\frac{d\zeta}{ds}, \tag{A45}$$

which may be integrated to yield

$$\frac{d\bar{s}}{ds}=\frac{\alpha}{\zeta^2}. \tag{A46}$$

α is an arbitrary integration constant which, without loss of generality, may be taken positive. When $\alpha\neq 0$ one may write

$$\frac{d\zeta^A}{d\bar{s}}=\frac{\zeta^2}{\alpha}\frac{d\zeta^A}{ds}, \tag{A47}$$

which, in combination with (A43), yields (A37), showing that geodesics in M project onto geodesics in $\bar{M}$.

Substitution of (A46) into (A42) yields

$$\frac{d^2\zeta}{ds^2}+\frac{\kappa^2\alpha^2}{\zeta^3}=0,\quad \kappa\equiv(3/32)^{1/2}, \tag{A48}$$

which integrates to

$$d\zeta/ds=\pm(\kappa^2\alpha^2\zeta^{-2}-\beta)^{1/2}, \tag{A49}$$

where β is another integration constant. Using the metric (5.8) it is easy to verify that standard normalization for the affine parameter s is obtained by choosing $\beta=-1$, 0, or 1 according as the geodesic is timelike, null, or spacelike.

Timelike geodesics. In this case ζ must always increase (or decrease) with s. Therefore choosing the positive root in (A49) and the boundary conditions $\zeta(0)=0$, $\bar{s}(\infty)=0$, one finds, upon setting $\beta=-1$ and integrating Eqs. (A46) and (A49),

$$\zeta(s)=[s(2\kappa\alpha+s)]^{1/2}=-\kappa\alpha\,\mathrm{csch}(\kappa\bar{s}), \tag{A50}$$

$$\bar{s}(s)=\frac{1}{2\kappa}\ln\frac{s}{2\kappa\alpha+s}. \tag{A51}$$

The ranges of the variables are

$$0\leqslant s<\infty,\quad -\infty\leqslant\bar{s}<0,\quad 0\leqslant\zeta<\infty, \tag{A52}$$

and one sees that the geodesic strikes the frontier at $s=0$.

In terms of the matrix $\boldsymbol{\gamma}$ the above results may be expressed in the form

$$\boldsymbol{\gamma}(s)=[-\kappa^2\alpha\,\mathrm{csch}(\kappa\bar{s})]^{4/3}\mathbf{M}^{\sim}e^{\mathbf{N}\bar{s}}\mathbf{M}, \tag{A53}$$

where $\mathbf{N}$ is restricted as in (5.13) and $\mathbf{M}$ is now required to have unit determinant.

Null geodesics. In this case it is the constant α which serves to fix the scale of s. Setting $2\kappa\alpha=1$, $\beta=0$, and choosing the positive root in (A49), one finds, with the boundary condition $\zeta(0)=0$,

$$\zeta(s)=s^{1/2}. \tag{A54}$$

There is no preferred zero point for the arc length $\bar{s}$ in $\mathbf{M}$. Hence the following integral of Eq. (A46) may be chosen:

$$\bar{s}=(2\kappa)^{-1}\ln s. \tag{A55}$$

This yields

$$\zeta(s)=e^{\kappa\bar{s}}, \tag{A56}$$

$$\boldsymbol{\gamma}(s)=(\kappa e^{\kappa\bar{s}})^{4/3}\mathbf{M}^{\sim}e^{\mathbf{N}\bar{s}}\mathbf{M}. \tag{A57}$$

The ranges of the variables are

$$0\leqslant s<\infty,\quad -\infty\leqslant\bar{s}<\infty,\quad 0\leqslant\zeta<\infty, \tag{A58}$$

and the geodesic is again seen to hit the frontier.

Spacelike geodesics. In this case there is a turning point at $\zeta=\kappa\alpha$ and both roots in (A49) can occur. Setting $\beta=1$, and choosing the boundary conditions $\zeta(0)=0$, $\bar{s}(\kappa\alpha)=0$, with $s\geqslant 0$, one finds

$$\zeta(s)=[s(2\kappa\alpha-s)]^{1/2}=\kappa\alpha\,\mathrm{sech}(\kappa\bar{s}), \tag{A59}$$

$$\bar{s}(s)=\frac{1}{2\kappa}\ln\frac{s}{2\kappa\alpha-s}, \tag{A60}$$

$$\boldsymbol{\gamma}(s)=[\kappa^2\alpha\,\mathrm{sech}(\kappa\bar{s})]^{4/3}\mathbf{M}^{\sim}e^{\mathbf{N}\bar{s}}\mathbf{M}. \tag{A61}$$

The ranges of the variables are

$$0\leqslant s\leqslant 2\kappa\alpha,\quad -\infty\leqslant\bar{s}\leqslant\infty,\quad 0\leqslant\zeta\leqslant\kappa\alpha, \tag{A62}$$

and the geodesic is seen to hit the frontier at both ends ($s=0$, 2κ).

Using the above results it is possible to obtain an expression for the "distance" to the frontier from any point in M. If $\boldsymbol{\gamma}$ is a fixed point and $\boldsymbol{\xi}$ is a contravariant vector at $\boldsymbol{\gamma}$, then

$$\sigma(\boldsymbol{\gamma},\boldsymbol{\xi})=\frac{16}{3}\frac{[\mathrm{tr}\,\mathbf{X}^2-\frac{1}{3}(\mathrm{tr}\,\mathbf{X})^2]^{1/2}+(\frac{2}{3})^{1/2}\,\mathrm{tr}\,\mathbf{X}}{[\mathrm{tr}\,\mathbf{X}^2-\frac{1}{3}(\mathrm{tr}\,\mathbf{X})^2]^{1/2}-(\frac{2}{3})^{1/2}\,\mathrm{tr}\,\mathbf{X}}\gamma^{1/2}, \tag{A63}$$

$$\mathbf{X}=\boldsymbol{\gamma}^{-1}\boldsymbol{\xi},$$

where σ is one-half the square of the distance from γ to the frontier along the geodesic which starts from γ in the direction of ξ. The formula holds as long as $\mathrm{tr}\mathbf{X} < (\mathrm{tr}\mathbf{X}^2)^{1/2}$, so that the denominator is positive, for otherwise the geodesic escapes to infinity in the direction of ξ. σ takes on its minimum value, $-(16/3)\gamma^{1/2}$, when the geodesic is a path of pure dilation which arrives at the frontier at the point $\gamma=0$. The condition for this is $\mathrm{tr}\mathbf{X}^2=\frac{1}{3}(\mathrm{tr}\mathbf{X})^2$.

It should be noted that when the geodesic is not a path of pure dilation, it usually strikes the frontier at a point where some of the γ_{ij} become infinite, in spite of the fact that γ itself vanishes there. To see this, first observe that expressions (A53), (A57), and (A61) all have the limiting behavior

$$\gamma(s) \to (2\kappa^2\alpha e^{\kappa\bar{s}})^{4/3}\mathbf{M}^{\sim}e^{\mathbf{N}\bar{s}}\mathbf{M} \text{ as } \bar{s}\to-\infty. \quad \text{(A64)}$$

Next note that in virtue of the conditions (5.13) $\mathbf{N}$ always has one root which is at least as negative as $-(\frac{1}{6})^{1/2}$. But $4\kappa/3=(\frac{1}{6})^{1/2}$. Therefore the only way in which a blow up of the exponential can be avoided in (A64) is for the roots of $\mathbf{N}$ to be precisely $-(\frac{1}{6})^{1/2}$, $-(\frac{1}{6})^{1/2}$, $(\frac{2}{3})^{1/2}$. Without loss of generality $\mathbf{N}$ may be chosen diagonal. The limiting form of $\gamma(s)$ in this special case is then

$$\gamma(0)=\mathbf{M}^{\sim}\begin{bmatrix}(2\kappa^2\alpha)^{4/3} & 0 & 0\\ 0 & (2\kappa^2\alpha)^{4/3} & 0\\ 0 & 0 & 0\end{bmatrix}\mathbf{M}. \quad \text{(A65)}$$

Since $\mathbf{M}$ and α are arbitrary (subject to $\det\mathbf{M}=1$) it follows that any singular symmetric matrix having an odd number of vanishing roots can be reached by a geodesic. Matrices having two vanishing roots can be reached (in a finite distance) from nonsingular points, but only along paths which suffer infinite absolute acceleration at the frontier.

It is not difficult to obtain an expression for the geodesic distance between two matrices γ_1 and γ_2 in M. The geometry of M may be summed up in the compact formula

$$ds^2=-d\zeta^2+\zeta^2 d\bar{s}^2, \quad \text{(A66)}$$

which follows from (5.8). With the introduction of the variables

$$t=\zeta\cosh\kappa\bar{s}, \quad x=\zeta\sinh\kappa\bar{s}, \quad \text{(A67)}$$

this is converted to

$$ds^2=-dt^2+dx^2, \quad \text{(A68)}$$

which is formally just the line element of 2-dimensional Minkowski space. Hence

$$\begin{aligned}\sigma(\gamma_1,\gamma_2)\equiv\tfrac{1}{2}(s_{12})^2&=-\tfrac{1}{2}(t_1-t_2)^2+\tfrac{1}{2}(x_1-x_2)^2\\ &=\tfrac{1}{2}(2\zeta_1\zeta_2\cosh\kappa\bar{s}_{12}-\zeta_1{}^2-\zeta_2{}^2)\\ &=(16/3)\{2(\gamma_1\gamma_2)^{1/4}\cosh[(3/32)^{1/2}\bar{s}_{12}]\\ &\qquad -\gamma_1{}^{1/2}-\gamma_2{}^{1/2}\}, \quad \text{(A69)}\end{aligned}$$

where $\bar{s}_{12}$ is given by (A41).

APPENDIX B: THE MANIFOLDS M^{∞^3} AND $\mathfrak{M}$

In discussing these manifolds it will be convenient to use an abbreviated notation which avoids the necessity of writing integral signs or excessive numbers of indices bearing various numbers of primes. This notation will be applicable to completely general manifolds and, in fact, will be used again in the following paper of this series in quite a different context, thus providing additional justification for its introduction here.

The functions $\gamma_{ij}(\mathbf{x})$ will be replaced by the symbol φ^i. More precisely, the symbol γ is replaced by φ, and the quintuple (i,j,x^1,x^2,x^3) by the single index i. In general applications the φ's may constitute either a finite discrete set of real numbers or, as here, a set of functions or "fields." When the index i has a continuous character, the summation convention for repeated indices will be understood to include integrations over the continuous labels for which it stands. In this Appendix no restriction will be placed on the range of values which the indices can assume.

The φ's are "coordinates" in a manifold (M^{∞^3} in the present case) on which a group acts. Group elements will be denoted by barred letters $\bar{x}$, $\bar{y}$, etc. and their components in some coordinate system in the group space will be denoted by $\bar{x}^\alpha$, $\bar{y}^\beta$, etc. For example, the 3-dimensional general coordinate transformation group may be coordinatized by the functions $\bar{x}^i(\mathbf{x})$ which define the coordinate transformation $x^i\to\bar{x}^i$. In the condensed notation the quadruplet (i,x^1,x^2,x^3) gets replaced by the single index α.

The multiplication table of the group defines a set of function(al)s $F^\alpha[\bar{y},\bar{x}]$ satisfying

$$F^\alpha[\bar{y},\bar{x}]=(\bar{y}\bar{x})^\alpha. \quad \text{(B1)}$$

For example, in the case of the coordinate transformation group this functional has the form

$$F^{i,\mathbf{x}}[\bar{y},\bar{x}]=\bar{y}^i(\bar{\mathbf{x}}(\mathbf{x})).$$

By virtue of the group postulates $F^\alpha[\bar{x},\bar{y}]$ also satisfies the following fundamental identities:

$$F^\alpha[\bar{x},e]=F^\alpha[e,\bar{x}]=\bar{x}^\alpha, \quad \text{(B2)}$$

$$F^\alpha[\bar{x},\bar{x}^{-1}]=F^\alpha[\bar{x}^{-1},\bar{x}]=e^\alpha, \quad \text{(B3)}$$

$$F^\alpha[\bar{z},\bar{y}\bar{x}]=F^\alpha[\bar{z}\bar{y},\bar{x}], \quad \text{(B4)}$$

where e is the identity element of the group and $\bar{x}^{-1}$ denotes the inverse of $\bar{x}$. In the case of the coordinate transformation group, we have $e^i(\mathbf{x})=x^i$.

Instead of dealing directly with $F^\alpha[\bar{y},\bar{x}]$, one more often makes use of a set of auxiliary funcation(al)s, together with the structure constants of the group. These are defined, respectively, by

$$L^\alpha{}_\beta[\bar{x}]\equiv(\delta F^\alpha[\bar{y},\bar{x}]/\delta\bar{y}^\beta)_{\bar{y}=e}, \quad \text{(B5)}$$

$$c^\alpha{}_{\beta\gamma}\equiv(\delta^2F^\alpha[\bar{y},\bar{x}]/\delta\bar{y}^\beta\delta\bar{x}^\gamma - \delta^2F^\alpha[\bar{y},\bar{x}]/\delta\bar{y}^\gamma\delta\bar{x}^\beta)_{\bar{x}=\bar{y}=e}. \quad \text{(B6)}$$

In the case of the coordinate transformation group one finds $L^{i}{}_{j'}[\bar{x}]=\delta^{i}{}_{j}\delta(\bar{\mathbf{x}}(\mathbf{x}),\mathbf{x}')$, while the structure constants are as given in Eq. (4.17). In the case of finite-dimensional Lie groups the functional derivatives in (B5) and (B6) become ordinary derivatives.

By repeatedly differentiating Eqs. (B2), (B3), (B4) and setting various elements equal to the identity, a number of important relations can be established. Among them we cite the following:

$$L^{\alpha}{}_{\beta}[e]=\delta^{\alpha}{}_{\beta}, \tag{B7}$$

$$L^{-1\alpha}{}_{\delta,\epsilon}-L^{-1\alpha}{}_{\epsilon,\delta}=-c^{\alpha}{}_{\beta\gamma}L^{-1\beta}{}_{\delta}L^{-1\gamma}{}_{\epsilon}, \tag{B8}$$

$$c^{\alpha}{}_{\beta\epsilon}c^{\epsilon}{}_{\gamma\delta}+c^{\alpha}{}_{\gamma\epsilon}c^{\epsilon}{}_{\delta\beta}+c^{\alpha}{}_{\delta\epsilon}c^{\epsilon}{}_{\beta\gamma}=0. \tag{B9}$$

In Eq. (B8) the arguments of the funcation(al)s have been suppressed, and differentiation is denoted by a comma. $L^{-1\alpha}{}_{\beta}$ denotes the matrix inverse to $L^{\alpha}{}_{\beta}$. In the case of the coordinate transformation group it is given by $L^{-1i}{}_{j'}[\bar{x}]=\delta^{i}{}_{j}\delta(\mathbf{x},\bar{\mathbf{x}}(\mathbf{x}'))\partial(\bar{\mathbf{x}}')/\partial(\mathbf{x}')$.

As a result of the action of the group the variables φ^{i} suffer a transformation which may be expressed in the form

$$\varphi'^{i}=\Phi^{i}[\bar{x},\varphi], \tag{B10}$$

where the function(al)s $\Phi^{i}[\bar{x},\varphi]$ satisfy the identities

$$\Phi^{i}[e,\varphi]=\varphi^{i}, \tag{B11}$$

$$\Phi^{i}[\bar{y}\bar{x},\varphi]=\Phi^{i}[\bar{y},\Phi[\bar{x},\varphi\ ]]. \tag{B12}$$

Differentiation of (B12) leads to

$$\Phi^{i}{}_{,\alpha}[\bar{x},\varphi]=R^{i}{}_{\beta}[\Phi[\bar{x},\varphi]]L^{-1\beta}{}_{\alpha}[\bar{x}], \tag{B13}$$

where

$$R^{i}{}_{\alpha}[\varphi]\equiv\Phi^{i}{}_{,\alpha}[e,\varphi]. \tag{B14}$$

The function(al)s $R^{i}{}_{\alpha}$ appear in the law of transformation of the φ's under infinitesimal group operations. Under the action of a group element having the coordinates $e^{\alpha}+\delta\xi^{\alpha}$, where the $\delta\xi$'s are infinitesimal, the φ's suffer the change

$$\delta\varphi^{i}=R^{i}{}_{\alpha}\delta\xi^{\alpha}. \tag{B15}$$

(Functional arguments are again suppressed.) For example, under the infinitesimal coordinate transformation $\bar{x}^{i}=x^{i}+\delta\xi^{i}$, the 3-metric γ_{ij} suffers the change

$$\delta\gamma_{ij}=\int R_{ijk'}\delta\xi^{k'}d^{3}x', \tag{B16}$$

where

$$R_{ijk'}\equiv-\gamma_{ij,k}\delta(\mathbf{x},\mathbf{x}')-\gamma_{kj}\delta_{,i}(\mathbf{x},\mathbf{x}')-\gamma_{ik}\delta_{,j}(\mathbf{x},\mathbf{x}') \tag{B17a}$$

$$=-\delta_{ik'.j}-\delta_{jk'.i}, \tag{B17b}$$

$$\delta_{ij'}\equiv\gamma_{ik}\delta^{k}{}_{j'}. \tag{B18}$$

If the transformation laws (B10) and (B14) are linear as in this case, then $R^{i}{}_{\alpha,jk}=0$. (The reader should avoid confusing differentiation with respect to the x's in the explicit notation and functional differentiation with respect to the φ's in the compact notation.)

With these preliminaries out of the way the question of imposing a metric on the manifold of φ's may now be considered. Let such a metric be denoted by $g_{ij}[\varphi]$. If the group is to generate isometric motions in the manifold then this metric must satisfy Killing's equation:

$$g_{ij,k}\delta\varphi^{k}+g_{kj}\delta\varphi^{k}{}_{,i}+g_{ik}\delta\varphi^{k}{}_{,j}=0, \tag{B19}$$

with $\delta\varphi^{i}$ given by Eq. (B15). It is not difficult to see that this equation may be regarded as a group transformation law for g_{ij}:

$$\delta g_{ij}\equiv g_{ij,k}\delta\varphi^{k}=-g_{kj}R^{k}{}_{\alpha,i}\delta\xi^{\alpha}-g_{ik}R^{k}{}_{\alpha,j}\delta\xi^{\alpha}. \tag{B20}$$

When the transformation law (B15) is linear and homogeneous, then

$$\delta\varphi^{i}=R^{i}{}_{\alpha,j}\varphi^{j}\delta\xi^{\alpha}, \tag{B21}$$

and Eq. (B20) says simply that g_{ij} must transform contragrediently to the Kronecker product $\varphi^{i}\varphi^{j}$. This is a necessary and sufficient condition for the isometry of group operations.

Now let $d\varphi^{i}$ be an arbitrary displacement, with the corresponding "arc length" ds given by

$$ds^{2}=g_{ij}d\varphi^{i}d\varphi^{j}. \tag{B22}$$

If $d\varphi^{i}$ happens to be orthogonal to the *orbit* of φ under the group then it satisfies the condition

$$R^{j}{}_{\alpha}g_{ij}d\varphi^{j}=0, \tag{B23}$$

and (B22) gives directly the distance between neighboring orbits. In the case of the manifold $M^{\infty^{3}}$, with the metric (6.5) and the transformation law (B16), (B17), the condition (B23) takes the form

$$\gamma^{jk}(d\gamma_{ij.k}-d\gamma_{jk.i})=0. \tag{B24}$$

More generally, the distance between orbφ and orb$(\varphi+d\varphi)$ is given by

$$d\bar{s}^{2}=g_{ij}\bar{d}\varphi^{i}\bar{d}\varphi^{j}, \tag{B25}$$

where $\bar{d}\varphi^{i}$ is the projection of $d\varphi^{i}$ normal to the orbit:

$$\bar{d}\varphi^{i}=(\delta^{i}{}_{j}-R^{i}{}_{\alpha}\gamma^{\alpha\beta}R^{k}{}_{\beta}g_{kj})d\varphi^{j}, \tag{B26}$$

$$\gamma_{\alpha\gamma}\gamma^{\gamma\beta}=\delta_{\alpha}{}^{\beta}, \tag{B27}$$

$$\gamma_{\alpha\beta}\equiv g_{ij}R^{i}{}_{\alpha}R^{j}{}_{\beta}. \tag{B28}$$

When the indices α, β include continuous labels the "matrix" $\gamma_{\alpha\beta}$ is tyipcally a differential operator (sum of differentiated δ functions) and its inverse $\gamma^{\alpha\beta}$ is a Green's function.

Equation (B25) may be written in the alternative forms

$$d\bar{s}^{2}=\bar{g}_{ij}d\varphi^{i}d\varphi^{j}=\bar{g}_{ij}\bar{d}\varphi^{i}\bar{d}\varphi^{j}, \tag{B29}$$

where

$$\bar{g}_{ij}\equiv g_{ij}-g_{ik}R^{k}{}_{\alpha}\gamma^{\alpha\beta}R^{l}{}_{\beta}g_{lj}. \tag{B30}$$

$\bar{g}_{ij}$ is the metric in the manifold of orbits, and the question arises how *it* transforms under the group. This question is answered by establishing the following

transformation laws:

$$\delta R^i{}_\alpha \equiv R^i{}_{\alpha,j}\delta\varphi^j = (R^i{}_{\beta,j}R^i{}_\alpha - c^\gamma{}_{\beta\alpha}R^i{}_\gamma)\delta\xi^\beta, \quad \text{(B31)}$$

$$\delta\gamma_{\alpha\beta} \equiv \gamma_{\alpha\beta,i}\delta\varphi^i = -(c^\gamma{}_{\delta\alpha}\gamma_{\gamma\beta} + c^\gamma{}_{\delta\beta}\gamma_{\alpha\gamma})\delta\xi^\delta, \quad \text{(B32)}$$

$$\delta\gamma^{\alpha\beta} \equiv \gamma^{\alpha\beta}{}_{,i}\delta\varphi^i = (c^\alpha{}_{\delta\gamma}\gamma^{\gamma\beta} + c^\beta{}_{\delta\gamma}\gamma^{\alpha\gamma})\delta\xi^\delta. \quad \text{(B33)}$$

Equations (B32) and (B33) are corollaries of (B19), (B27), and (B31), while (B31) itself is a consequence of the identity

$$R^i{}_{\alpha,j}R^j{}_\beta - R^i{}_{\beta,j}R^j{}_\alpha = R^i{}_\gamma c^\gamma{}_{\alpha\beta}, \quad \text{(B34)}$$

which is obtained by differentiating Eq. (B13) with respect to $\bar{x}^\beta$, setting $\bar{x}=e$, antisymmetrizing in α and β, and making use of (B8). With the aid of (B31) and (B33) it is straightforward to show that $\bar{g}_{ij}$ transforms just like g_{ij}. This means that group operations are isometries of $\bar{g}_{ij}$ just as they are of g_{ij}, and suggests that $\bar{g}_{ij}$ is effectively a function of orbφ alone. In order to make this fact explicit a coordinatization of the orbit manifold will be introduced.

This may be accomplished by first introducing a hypersurface in the φ manifold, defined by a set of simultaneous equations

$$f_\alpha[\varphi] = 0, \quad \text{(B35)}$$

where the index α ranges over the same continuum (or discrete set, as the case may be) as the group indices. The only requirement on the hypersurface is that it intersect the orbit of every point contained in (at least) some finite portion of the φ manifold. A coordinate system is then laid down in this hypersurface, with the coordinates denoted by z^A. If the hypersurface has been carefully chosen each orbit will intersect it in a single point, and the z's at that point may be used to label the orbit itself. For example, in the manifold $M^{\infty 3}$ one may choose for the equations (B35) the harmonic condition $(\gamma^{1/2}\gamma^{ij})_{,j}=0$; then any three of the functions $\varphi^{AB}(\eta)$ of Eq. (5.3) may be chosen as the z's.

A general point in the φ manifold will be reached by moving off the hypersurface along (i.e., within) an orbit thus:

$$\varphi^i[\bar{x},z] = \Phi^i[\bar{x},\varphi_0[z]], \quad \text{(B36)}$$

where $\varphi_0{}^i[z]$ is the starting point on the hypersurface. The group coordinates $\bar{x}^\alpha$ together with the z's provide a new labeling scheme for the points of the φ manifold, and the task before us is to compute the metrics g_{ij} and $\bar{g}_{ij}$ in this new coordinate system. For this purpose we shall need the relations

$$\varphi^i{}_{,\alpha} = R^i{}_\beta[\varphi]L^{-1\beta}{}_\alpha[\bar{x}], \quad \text{(B37)}$$

$$\varphi^i{}_{,\alpha A} = R^i{}_{\beta,j}[\varphi]\varphi^j{}_{,A}L^{-1\beta}{}_\alpha[\bar{x}], \quad \text{(B38)}$$

$$\varphi^i{}_{,\alpha\beta} = R^i{}_{\gamma,j}[\varphi]R^j{}_\delta[\varphi]L^{-1\gamma}{}_\alpha[\bar{x}]L^{-1\delta}{}_\beta[\bar{x}] + R^i{}_\gamma[\varphi]L^{-1\gamma}{}_{\alpha,\beta}[\bar{x}], \quad \text{(B39)}$$

which are obtained by applying Eq. (B13) to (B36). In the work which follows the arguments $\bar{x}$, φ, and z will be suppressed.

It is straightforward to compute

$$g_{\alpha\beta} \equiv g_{ij}\varphi^i{}_{,\alpha}\varphi^j{}_{,\beta} = \gamma_{\gamma\delta}L^{-1\gamma}{}_\alpha L^{-1\delta}{}_\beta, \quad \text{(B40)}$$

$$g_{\alpha A} = g_{A\alpha} \equiv g_{ij}\varphi^i{}_{,\alpha}\varphi^j{}_{,A} = g_{ij}R^i{}_\beta L^{-1\beta}{}_\alpha\varphi^j{}_{,A}, \quad \text{(B41)}$$

$$g_{AB} \equiv g_{ij}\varphi^i{}_{,A}\varphi^j{}_{,B} = \bar{g}_{AB} + g_{A\alpha}\bar{g}^{\alpha\beta}g_{\beta B}, \quad \text{(B42)}$$

where

$$\bar{g}_{AB} \equiv \bar{g}_{ij}\varphi^i{}_{,A}\varphi^j{}_{,B}, \quad \text{(B43)}$$

$$\bar{g}^{\alpha\beta} \equiv L^\alpha{}_\gamma L^\beta{}_\delta\gamma^{\gamma\delta}, \quad g_{\alpha\gamma}\bar{g}^{\gamma\beta} = \delta_\alpha{}^\beta. \quad \text{(B44)}$$

One then readily verifies that the contravariant metric, with components $g^{\alpha\beta}$, $g^{\alpha A}(=g^{A\alpha})$, g^{AB}, is given by

$$g^{\alpha\beta} = \bar{g}^{\alpha\beta} + \bar{g}^{\alpha\gamma}g_{\gamma A}g^{AB}g_{B\delta}\bar{g}^{\delta\beta}, \quad \text{(B45)}$$

$$g^{\alpha A} = -\bar{g}^{\alpha\beta}g_{\beta B}g^{BA}, \quad \text{(B46)}$$

$$\bar{g}_{AC}g^{CB} = \delta_A{}^B. \quad \text{(B47)}$$

It is now easy to show that g_{AB} and $\bar{g}_{AB}$ (and hence g^{AB}) are independent of the x's. Thus, using (B37) and (B38) one finds

$$g_{AB,\alpha} = g_{ij,k}\varphi^i{}_{,A}\varphi^j{}_{,B}\varphi^k{}_{,\alpha} + g_{ij}(\varphi^i{}_{,A\alpha}\varphi^j{}_{,B} + \varphi^i{}_{,A}\varphi^j{}_{,B\alpha}) = (g_{ij,k}R^k{}_\beta + g_{kj}R^k{}_{\beta,i} + g_{ik}R^k{}_{\beta,j}) \times \varphi^i{}_{,A}\varphi^j{}_{,B}L^{-1\beta}{}_\alpha. \quad \text{(B48)}$$

But this expression vanishes by virtue of the group transformation law for g_{ij} [Eq. (B20)]. Since $\bar{g}_{ij}$ obeys the same transformation law it follows that

$$\bar{g}_{AB,\alpha} = 0. \quad \text{(B49)}$$

That is, the metric $\bar{g}_{AB}$ of the orbit manifold depends only on the z's, as was expected.

For the study of geodesics in the φ manifold the following derivatives will also be needed:

$$g_{\alpha\beta,\gamma} = g_{ij,k}\varphi^i{}_{,\alpha}\varphi^j{}_{,\beta}\varphi^k{}_{,\gamma} + g_{ij}(\varphi^i{}_{,\alpha\gamma}\varphi^j{}_{,\beta} + \varphi^i{}_{,\alpha}\varphi^j{}_{,\beta\gamma}) = g_{\alpha\delta}L^\delta{}_\epsilon L^{-1\epsilon}{}_{\gamma,\beta} + g_{\beta\delta}L^\delta{}_\epsilon L^{-1\epsilon}{}_{\gamma,\alpha}, \quad \text{(B50)}$$

$$g_{\alpha A,\beta} = g_{ij,k}\varphi^i{}_{,\alpha}\varphi^j{}_{,A}\varphi^k{}_{,\beta} + g_{ij}(\varphi^i{}_{,\alpha\beta}\varphi^j{}_{,A} + \varphi^i{}_{,\alpha}\varphi^j{}_{,A\beta}) = g_{A\gamma}L^\gamma{}_\delta L^{-1\delta}{}_{\beta,\alpha}. \quad \text{(B51)}$$

These are obtained with the aid of (B19), (B37), (B38), and (B39).

The geodesic equations in the φ manifold may be written in the form

$$0 = g_{\alpha\beta}\frac{d^2\bar{x}^\beta}{ds^2} + g_{\alpha A}\frac{d^2z^A}{ds^2} + \Gamma_{\beta\gamma\alpha}\frac{d\bar{x}^\beta}{ds}\frac{d\bar{x}^\gamma}{ds} + 2\Gamma_{\beta A\alpha}\frac{d\bar{x}^\beta}{ds}\frac{dz^A}{ds} + \Gamma_{AB\alpha}\frac{dz^A}{ds}\frac{dz^B}{ds}, \quad \text{(B52)}$$

$$0 = g_{A\alpha}\frac{d^2\bar{x}^\alpha}{ds^2} + g_{AB}\frac{d^2z^B}{ds^2} + \Gamma_{\alpha\beta A}\frac{d\bar{x}^\alpha}{ds}\frac{d\bar{x}^\beta}{ds} + 2\Gamma_{\alpha BA}\frac{d\bar{x}^\alpha}{ds}\frac{dz^B}{ds} + \Gamma_{BCA}\frac{dz^B}{ds}\frac{dz^C}{ds}, \quad \text{(B53)}$$

where the Γ's are the Christoffel symbols. Multiplying

(B52) by $g_{A\delta}\bar{g}^{\delta\alpha}$, subtracting the result from (B53) and using the fact that $g_{AB,\alpha}=0$, one obtains

$$0=\bar{g}_{AB}\frac{d^2z^B}{ds^2}+\bar{\Gamma}_{BCA}\frac{dz^B}{ds}\frac{dz^C}{ds}+\tfrac{1}{2}[g_{\alpha A,\beta}+g_{\beta A,\alpha}-g_{\alpha\beta,A}-g_{A\delta}\bar{g}^{\delta\gamma}(g_{\alpha\gamma,\beta}+g_{\beta\gamma,\alpha}-g_{\alpha\beta,\gamma})]\frac{d\bar{x}^\alpha}{ds}\frac{d\bar{x}^\beta}{ds}$$

$$+[g_{\alpha A,B}-g_{\alpha B,A}-g_{A\gamma}\bar{g}^{\gamma\beta}(g_{\alpha\beta,B}+g_{\beta B,\alpha}-g_{\alpha B,\beta})]\frac{d\bar{x}^\alpha}{ds}\frac{dz^B}{ds}+\tfrac{1}{2}[(g_{B\alpha}\bar{g}^{\alpha\beta}g_{\beta A})_{,C}+(g_{C\alpha}\bar{g}^{\alpha\beta}g_{\beta A})_{,B}-(g_{B\alpha}\bar{g}^{\alpha\beta}g_{\beta C})_{,A}$$

$$-g_{A\beta}\bar{g}^{\beta\alpha}(g_{B\alpha,C}+g_{C\alpha,B})]\frac{dz^B}{ds}\frac{dz^C}{ds}, \quad \text{(B54)}$$

where the $\bar{\Gamma}$'s are the Christoffel symbols of the orbit manifold. The terms of this equation can be regrouped by judicious use of identities such as $g_{\alpha B,A}=(g_{\alpha\gamma}\bar{g}^{\gamma\delta}g_{\delta B})_{,A}$ and $\bar{g}^{\gamma\delta}{}_{,A}g_{\delta\beta}=-\bar{g}^{\gamma\delta}g_{\delta\beta,A}$ and by replacing derivatives of the form $g_{\alpha\beta,\gamma}$ and $g_{\alpha A,\beta}$ by their expressions (B50), (B51). The final useful result is

$$0=\bar{g}_{AB}\frac{d^2z^B}{ds^2}+\bar{\Gamma}_{BCA}\frac{dz^B}{ds}\frac{dz^C}{ds}+[(g_{A\gamma}\bar{g}^{\gamma\alpha})_{,C}-(g_{C\gamma}\bar{g}^{\gamma\alpha})_{,A}]\frac{dz^C}{ds}\left(g_{\alpha\beta}\frac{d\bar{x}^\beta}{ds}+g_{\alpha B}\frac{dz^B}{ds}\right)$$

$$+\tfrac{1}{2}\bar{g}^{\gamma\delta}{}_{,A}\left(g_{\gamma\alpha}\frac{d\bar{x}^\alpha}{ds}+g_{\gamma B}\frac{dz^B}{ds}\right)\left(g_{\delta\beta}\frac{d\bar{x}^\beta}{ds}+g_{\delta C}\frac{dz^C}{ds}\right)+g_{A\gamma}\bar{g}^{\gamma\beta}L^\delta{}_\epsilon(L^{-1\epsilon}{}_{\beta,\alpha}-L^{-1\epsilon}{}_{\alpha,\beta})\frac{d\bar{x}^\alpha}{ds}\left(g_{\delta\zeta}\frac{d\bar{x}^\zeta}{ds}+g_{\delta B}\frac{dz^B}{ds}\right). \quad \text{(B55)}$$

Now suppose the geodesic intersects one of the orbits orthogonally. The condition for this is [cf. Eq. (B23)]

$$R^i{}_\beta g_{ij}d\varphi^j/ds=0, \quad \text{(B56)}$$

which, when multiplied by $L^{-1\beta}{}_\alpha$, yields

$$g_{\alpha\beta}\frac{d\bar{x}^\beta}{ds}+g_{\alpha A}\frac{dz^A}{ds}=0. \quad \text{(B57)}$$

When this condition is satisfied we have

$$g_{AB}\frac{dz^A}{ds}\frac{dz^B}{ds}+2g_{\alpha A}\frac{d\bar{x}^\alpha}{ds}\frac{dz^A}{ds}+g_{\alpha\beta}\frac{d\bar{x}^\alpha}{ds}\frac{d\bar{x}^\beta}{ds}=\bar{g}_{AB}\frac{dz^A}{ds}\frac{dz^B}{ds}, \quad \text{(B58)}$$

and hence

$$ds^2=d\bar{s}^2, \quad \text{(B59)}$$

so that the arc length in the φ manifold becomes the same as in the orbit manifold. Moreover, by virtue of (B55) it follows that the z's in this case satisfy also the geodesic equation in the orbit manifold,

$$\bar{g}_{AB}\frac{d^2z^B}{d\bar{s}^2}+\bar{\Gamma}_{BCA}\frac{dz^B}{d\bar{s}}\frac{dz^C}{d\bar{s}}=0, \quad \text{(B60)}$$

provided the orthogonality condition (B57) is maintained along the entire length of the geodesic. But this is an immediate consequence of Eqs. (B50), (B51), and (B52), for by differentiating the left-hand side of (B57) with respect to s, one obtains

$$g_{\alpha\beta}\frac{d^2\bar{x}^\beta}{ds^2}+g_{\alpha A}\frac{d^2z^A}{ds^2}+g_{\alpha\beta,\gamma}\frac{d\bar{x}^\beta}{ds}\frac{d\bar{x}^\gamma}{ds}+(g_{\alpha\beta,A}+g_{\alpha A,\beta})\frac{d\bar{x}^\beta}{ds}\frac{dz^A}{ds}+g_{\alpha A,B}\frac{dz^A}{ds}\frac{dz^B}{ds}$$

$$=\tfrac{1}{2}g_{\beta\gamma,\alpha}\frac{d\bar{x}^\beta}{ds}\frac{d\bar{x}^\gamma}{ds}+g_{\beta A,\alpha}\frac{d\bar{x}^\beta}{ds}\frac{dz^A}{ds}=\left(g_{\delta\beta}\frac{d\bar{x}^\beta}{ds}+g_{\delta A}\frac{dz^A}{ds}\right)L^\delta{}_\epsilon L^{-1\epsilon}{}_{\alpha,\gamma}\frac{d\bar{x}^\gamma}{ds}, \quad \text{(B61)}$$

which vanishes by virtue of (B57) itself. Therefore, if the geodesic intersects one orbit orthogonally then it intersects every orbit in its path orthogonally, and, moreover, it traces out a geodesic curve in the orbit manifold.