Action Functional of General Relativity as a Path Length. I. Closed Empty Universes*

ROBERT H. GOWDY†
Yale University, New Haven, Connecticut 06520
(Received 23 July 1970)

For the case of closed empty universes it is established (up to a uniqueness conjecture) that the action functional of general relativity is a Riemannian path length in superspace, an infinite-dimensional manifold whose points represent three-dimensional geometries. Each procedure for representing spacetimes by trajectories in superspace yields a corresponding superspace metric. These metrics are just the "geodesic sheaf" metrics found by DeWitt. A general expression for all of these metrics is obtained in terms of arbitrary coordinates on superspace. The expression requires an explicit and unique solution to the spacelike constraints ($G^{0i}=0$, $i=1, 2, 3$) of general relativity. Supertrajectories are separated into "timelike" ones which are permitted because they have real actions and "spacelike" ones which are forbidden because their actions are imaginary. By varying the path-length action, one finds that solutions of Einstein's source-free field equations correspond to a class of timelike geodesics in superspace. This class of geodesics is subject to conserved constraints which are homogeneous and quadratic in time derivatives. The truncated superspace which contains the empty "mixmaster" universes studied by Misner is discussed as an application of the supergeometric approach. The approach is particularly useful for analyzing the maximum expansion stage of the universe. Universes at this stage of their evolution encounter a real singularity of superspace. Time-symmetric universes end their supertrajectories on this singularity, while all other universes penetrate it in a well-defined way. The singularity, together with the causal structure of superspace, limits the anisotropy of empty mixmaster universes which are a finite number of e-foldings from the stage of maximum expansion.

I. INTRODUCTION

SUPERSPACE is the name which Wheeler has given to a manifold whose points represent three-dimensional geometries.[1,2] Arnowitt, Deser, and Misner,[3] as well as Baierlein, Sharp, and Wheeler,[4] have shown that spacetimes can usefully be regarded as trajectories in such a superspace. Each point along a superspace trajectory represents the intrinsic three-geometry of a spacelike hypersurface in the corresponding spacetime. This paper will consider only single-parameter trajectories—curves in superspace.[5] Along such a trajectory, a point at parameter value t represents the intrinsic geometry of the hypersurface $\Sigma(t)$ defined in the corresponding spacetime by the equation $x^0=t$.

DeWitt has noted that trajectories corresponding to solutions of Einstein's field equations may be regarded as geodesics in superspace.[6] Taken by itself, this fact is not surprising, because a large class of dynamical systems may be geometrized by a trivial procedure (see Appendix A). The present paper will show that there is one important respect in which the superspace geometrization of general relativity differs from the trivial cases: The *physical* action functional of general relativity is a path length in a *hyperbolic* metric. Because the physical action functional can only be real valued, trajectories with imaginary path lengths are forbidden. This causal structure gives physical meaning to the concept of spacelike and timelike trajectories in superspace and is a strong argument for the superspace geometrization of general relativity.

Each distinct procedure for defining the equal-time hypersurfaces $\Sigma(t)$ in a given spacetime provides a distinct mapping of spacetimes into superspace trajectories and a superspace metric in which the action functional of general relativity is a path length. These superspace metrics are just the ones defined by DeWitt —the metrics in which Einstein spacetimes appear as sheafs of geodesics.[6] The results of the present paper can be used to find any one of these metrics provided that one has an expression for the three-dimensional metric tensor as a function of space coordinates and supercoordinates (coordinates which identify points in superspace) and provided that one can solve the spacelike constraints of general relativity. In the general case, these provisions are difficult to meet. However, there are interesting subspaces of superspace which lend themselves to straightforward analysis.

The truncated superspace which contains Misner's "mixmaster" universes provides a simple application of the supergeometry approach to general relativity. Here one finds the basic singularity structure of superspace, the determination of superspace topology from the path-length form of the action, causality in super-

* Research (Yale University Report No. 3935-1) supported by the U. S. Atomic Energy Commission under Contract No. AT(30-1)-3935.

† Present address: Department of Physics and Astronomy, University of Maryland, College Park, Md. 20742.

[1] J. A. Wheeler, in *Relativity Groups and Topology*, edited by B. S. DeWitt and C. M. DeWitt (Gordon and Breach, New York, 1964).

[2] J. A. Wheeler, in *Battelle Rencontres*, edited by C. M. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[3] R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: an Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962).

[4] R. F. Baierlein, D. H. Sharp, and J. A. Wheeler, Phys. Rev. **126**, 1864 (1962).

[5] For some purposes it is desirable to represent spacetimes by many-parameter trajectories. See, for example, U. H. Gerlach, Phys. Rev. **177**, 1929 (1969).

[6] B. S. DeWitt, in *Relativity: Proceedings of the Relativity Conference in the Midwest*, edited by L. Witten (Plenum, New York, 1970).

space, and finally, a demonstration of the power of supergeometry to analyze complex solutions of Einstein's equations.

This paper shows that a closed empty universe may be thought of as a "particle" moving freely in a curved superspace geometry. This curved superspace possesses some of the familiar properties of ordinary spacetime—light cones and local causality. In the absence of gravitational effects, it is the Minkowski spacetime of special relativity which provides the fixed geometrical framework of physics. For the purely gravitational physics of closed empty universes, it is curved superspace which provides the fixed geometrical framework.

II. NOTATION

The notation of this paper is essentially that of Refs. 1 and 3. The time coordinate in a spacetime is denoted by x^0 while space coordinates are x^1, x^2, and x^3. Greek indices run from 0 to 3 while lower-case Latin indices run from 1 to 3. Spacetime-covariant objects which do not contain 0 or Greek-letter indices are prefixed by a superscript 4 to distinguish them from the corresponding space-covariant objects. The spacetime metric tensor components $g_{\mu\nu}$ have the space-covariant representation

$$\begin{aligned} {}^4g_{ij} &\equiv g_{ij} \equiv \text{intrinsic metric on } \Sigma(t)\,, \\ g_{0i} &\equiv N_i \equiv \text{shift vector}\,, \\ g_{00} &\equiv N_iN^i - N^2\,, \end{aligned}$$

where

$$N \equiv (-g^{00})^{-1/2}$$

is called the lapse function and g^{ij}, the matrix inverse of g_{ij}, is used to raise all lower-case Latin indices. Space-covariant derivatives are defined in terms of the metric g_{ij} and are denoted by vertical bars, as in $N_{i|j}$. The determinant of g_{ij} is written as g and the invariant three-volume element $d^3x\, g^{1/2}$ is denoted by σ. In addition to the extrinsic curvature tensor K_{ij} and the momenta

$$\pi^{ij} \equiv g^{1/2}(g^{ij}K_r{}^r - K^{ij})$$

defined by Arnowitt, Deser, and Misner[3], this paper adopts a "velocity tensor" which is defined by

$$v_{ij} \equiv \partial g_{ij}/\partial t - N_{i|j} - N_{j|i}\,.$$

The extrinsic curvature K_{ij} is related to this velocity tensor by

$$v_{ij} = -2NK_{ij}\,.$$

In this paper, some further conventions are adopted. Capital Latin indices run from 1 to ∞ in expressions describing the full superspace and from 1 to the appropriate number in truncations of superspace. Terms containing repeated capital Latin indices are to be summed over the appropriate range. Matrix expressions are interpreted according to the rules

$$\begin{aligned} (\mathbf{A})_{ij} &\equiv A_{ij} && \text{for } \mathbf{A} \text{ covariant}\,, \\ &\equiv A_i{}^j && \text{for } \mathbf{A} \text{ mixed}\,, \\ &\equiv A^{ij} && \text{for } \mathbf{A} \text{ contravariant}\,, \\ A &\equiv g^{ij}A_{ij}\,, && |A|^2 \equiv A_{ij}A^{ij}\,, \\ (\mathbf{AB})_{ij} &\equiv (\mathbf{A})_{ir}(\mathbf{B})_{rj}\,. \end{aligned}$$

The total derivative dx/dt of a function x with respect to a parameter t along a curve in superspace will be denoted by $\dot{x}$. The curve parameter is denoted by t only when it is arbitrary. When the parameter is not arbitrary but is fixed by some constraint, it will be denoted by a Greek letter such as ω and $\dot{x}$ will denote $dx/d\omega$.

III. PATH-LENGTH FORM OF ACTION

Because of its general covariance, the action functional of general relativity is already a path length in the most general sense: It is an invariant integral along a curve in superspace. However, the invariance of this integral under the time parameter changes $t \to T(t)$ is not manifest because it depends upon the transformation properties of the g_{00} and g_{0i}. To find out if the path length is Riemannian and to evaluate the corresponding supermetric tensor components, one must make the parameter independence explicit. One could solve the full set of constraint equations for the N and N_i, thus ensuring the correct transformation properties for g_{00} and g_{0i}. However, that approach would require fixing the spacelike hypersurfaces $\Sigma(t)$ by a particular coordinate condition. This paper solves the spacelike constraints [see Eq. (3) below] for the N_i, thus fixing the shift

$$x^i(t+\Delta t) = x^i(t) + N^i\Delta t$$

of spatial coordinates along any hypersurface-orthogonal trajectory, but leaving the hypersurfaces themselves unrestricted. With the spacelike constraints solved, the generator of *position-independent* time-parameter changes can be found easily. This generator is essentially the *total* Hamiltonian of general relativity. Solving the single constraint which sets this generator equal to zero then ensures a manifestly parameter-independent form of the action functional. The Hamiltonian constraints which remain are to be solved *after* varying the action.

The action functional of general relativity is

$${}^4I = \int d^4x(-{}^4g)^{1/2}({}^4R)\,.$$

It may be written in the space-covariant form[7]

$$\begin{aligned} {}^4I &= \int_1^2 dt \int \sigma[-2\dot{K} + 2N^bK_{|b} + N(R + K^2 + |K|^2)] \\ &= \int_1^2 dt \int \sigma[N(R - K^2 + |K|^2)] - 2\int_2 \sigma K + 2\int_1 \sigma K\,. \end{aligned}$$

[7] D. R. Brill and R. H. Gowdy, Rept. Progr. Phys. (to be published).

The end-point terms can be dropped without affecting the dynamics. Indeed, these end-point terms must be dropped in order to obtain Einstein's field equations from variations which hold only the initial and final hypersurface three-geometries fixed. Thus, the action functional of general relativity can be taken to be

$$I=\int dt\, L\,, \tag{1}$$

where

$$L=\int \sigma\left[\frac{1}{4N}(|v|^2-v^2)+NR\right]. \tag{2}$$

Varying this action with respect to N and N_i yields the constraint equations and varying it with respect to g_{ij} yields the dynamical equations of general relativity.[8] This action is the *physical* action of the gravitational field in the sense that the action of a system with both gravitational and nongravitational degrees of freedom is just the *sum* of this gravitational action and the action which describes the nongravitational variables. Here we assume that the nongravitational action is expressed in units such that $c^3/16\pi k=1$. If one chose a different gravitational action such as $I'=\sqrt{I}$, this simple addition of actions would not work for nongravitational actions obtained by the usual minimal-coupling arguments.[9]

Throughout this paper, it is assumed that the spacelike constraints

$$\begin{aligned}\delta I/\delta N_i &\equiv \chi^i\\ &= g^{1/2}[N^{-1}(g^{ij}v-v^{ij})]_{|j}\\ &=0\end{aligned} \tag{3}$$

have been solved and used to eliminate superfluous degrees of freedom. All variations are assumed to be subject to these three constraints per space point. However, the Hamiltonian constraints

$$\begin{aligned}\delta I/\delta N &\equiv \mathcal{H}\\ &= g^{1/2}[(1/4N^2)(v^2-|v|^2)+R]\\ &=0\end{aligned}$$

are not assumed to be solved.

To obtain the generator for the position-independent time-parameter changes $t\rightarrow T(t)$, note that $\dot{g}_{ij}$ transforms according to the rule

$$\partial g_{ij}/\partial t\rightarrow \dot{T}^{-1}\partial g_{ij}/\partial t$$

and use the spacelike constraints (3) and the position independence of T to find

$$v_{ij}\rightarrow \dot{T}^{-1}v_{ij}\,.$$

Finally, use the rule $dt\rightarrow dT=\dot{T}dt$ to obtain the transformed Lagrangian

$$L=\int \sigma\left[\frac{1}{4N}\dot{T}^{-1}(v^2-|v|^2)-\dot{T}NR\right].$$

Because the unsolved constraints $\mathcal{H}(P)=0$ all commute with one another when the spacelike constraints are satisfied, one can generate canonical transformations with ordinary Poisson brackets.[10] The momentum conjugate to the variable T can then be found from the textbook formula

$$H=\partial L/\partial\dot{T}\,,$$

which yields

$$\begin{aligned}H&=\int d^3x\, N\mathcal{H}\\ &=\int\sigma\left[\frac{1}{4N}(v^2-|v|^2)+NR\right]\end{aligned}$$

when it is applied to the transformed Lagrangian and T is set equal to t. By solving the single constraint $H=0$, one can ensure the explicit time-parameter invariance of the action.

To solve the $H=0$ constraint, express the lapse function N as the product of a function n and a position-independent *lapse scale factor* s: $N=sn$. For the lapse scale factor s to be well defined, the function n must be normalized in some way. Thus, we call n the *normalized lapse function*. One way to normalize n is to require that the universal time parameter t agree with the proper time of a particular observer at rest (i.e., on a hypersurface-orthogonal trajectory) at a space point P_0. One would then require $n(P_0)=1/s$. However, the choice of normalization for n is purely a matter of convenience. All of the equations in this section leave this normalization arbitrary. Now reexpress L and H in a way which displays their dependence upon s by defining a kinetic term

$$\mathcal{T}\equiv\int\sigma\left[\frac{1}{4n}(|v|^2-v^2)\right] \tag{4}$$

and a potential term

$$\mathcal{V}\equiv-\int\sigma nR\,. \tag{5}$$

The resulting expressions for L and H are

$$L=s^{-1}\mathcal{T}-s\mathcal{V} \tag{6}$$

and

$$H=-(s^{-1}\mathcal{T}+s\mathcal{V})=0\,. \tag{7}$$

Use Eq. (7) to obtain L in the form

$$L=2s^{-1}\mathcal{T}\,. \tag{8}$$

[8] See Ref. 3 for the appropriate 3+1 form of these equations.

[9] Here, minimal coupling means that nongravitational actions should be of minimal differential order in $g_{\mu\nu}$ and should reduce to their special relativity values in a Minkowski spacetime.

[10] A discussion of the canonical formalism of general relativity and further references may be found in B. S. DeWitt, Phys. Rev. **160**, 1113 (1967); and also T. Kimura, Progr. Theoret. Phys. (Kyoto) **27**, 747 (1962).

To eliminate the $H=0$ constraint, solve it for the lapse scale factor

$$s=\pm(-\mathcal{T}/\mathcal{V})^{1/2}. \tag{9}$$

Now combine Eqs. (9) and (8) to obtain

$$L=\begin{cases}\pm 2(-\mathcal{V}\mathcal{T})^{1/2} & \text{for } \mathcal{T}>0\\ \mp 2(-\mathcal{V}\mathcal{T})^{1/2} & \text{for } \mathcal{T}<0,\end{cases} \tag{10}$$

where the sign of L is fixed by the sign in Eq. (9).

Equation (10) shows L to be the square root of a quadratic in the velocity tensor v_{ij}. If Ldt is to be a Riemannian line element on superspace, then v_{ij} must be linear and homogeneous in the total derivatives of supercoordinates with respect to the time parameter t. To show that v_{ij} actually has this structure, one must first express $\partial g_{ij}/\partial t$ in terms of coordinates on superspace and then solve the spacelike constraints for the shift components N_i.

Let the functionals S^A be coordinates on part of superspace. Describe the geometry $\mathcal{G}(S)$ which is located at the supercoordinate values $(S^1,S^2,\ldots)$ by choosing space-coordinate functions x^1, x^2, and x^3 on $\mathcal{G}(s)$ and displaying the corresponding metric components $g_{ij}(x,S)$. When a spacetime is represented as a curve in superspace, all of the time dependence of g_{ij} is carried by the supercoordinates and one has

$$\partial g_{ij}/\partial t=g_{ij,A}\dot{S}^A. \tag{11}$$

Here, the subscript $(,A)$ denotes partial differentiation with respect to S^A, with x and S^B $(B\neq A)$ held fixed.

From Eq. (11), the spacelike constraints (3) can be presented in the form

$$2G^{ijrs}(n^{-1}N_{r|s})_{|j}=\dot{S}^A G^{ijrs}(n^{-1}g_{rs,A})_{|j}, \tag{12}$$

where DeWitt's notation

$$G^{ijrs}\equiv\tfrac{1}{2}(g^{ir}g^{js}+g^{is}g^{jr}-2g^{ij}g^{rs})$$

has been adopted.[6] These constraints may be regarded as a system of inhomogeneous differential equations for the shift components N_i. Because the source terms are all linear in the $\dot{S}^A$, there is a particular solution of the form

$$N_i=\zeta_{iA}(x,S)\dot{S}^A.$$

To obtain the general solution of Eq. (12), let n_i solve the homogeneous system

$$G^{ijrs}(n^{-1}n_{r|s})_{|j}=0 \tag{13}$$

and take

$$N_i=\zeta_{iA}\dot{S}^A+n_i. \tag{14}$$

Equations (11) and (14) together with the definition of v_{ij} yield

$$v_{ij}=(g_{ij,A}-\zeta_{iA|j}-\zeta_{jA|i})\dot{S}^A-n_{i|j}-n_{j|i}. \tag{15}$$

At this point, one must ask if Eq. (14) can supply a *unique* solution to Eq. (12) for a given set of boundary conditions on the shift vector N_i. The conjecture that the solution is unique up to a Killing vector is part of what is often called the "thin-sandwich conjecture." The full thin-sandwich conjecture requires that both N and N_i be determined by the Hamiltonian and spacelike constraints.[1,2,4] The full conjecture leads to a nonlinear and singular Cauchy problem.[11] Here we are not concerned with determining the lapse function and have only a regular, linear Cauchy problem. In the closed universes being considered by this paper, the "unique shift conjecture" means that Eq. (13) has only trivial (zero or Killing-vector) solutions so that $n_{r|s}+n_{s|r}$ must always vanish. The validity of this conjecture as well as other aspects of the spacelike constraints will be treated in a later paper. The present paper adopts the unique shift conjecture without proof.

From Eqs. (4), (10), (15), and the unique shift conjecture, one obtains the path-length form of the action for closed empty universes:

$$I=\int d\Sigma, \tag{16}$$

where

$$d\Sigma^2=G_{AB}dS^AdS^B \tag{17}$$

and

$$G_{AB}=-\mathcal{V}\int\sigma[n^{-1}G^{ijrs}(g_{ij,A}-2\zeta_{iA|j})\times(g_{rs,B}-2\zeta_{rB|s})]. \tag{18}$$

This form of the action functional is to be varied with respect to the supertrajectory which is specified by the functions $S^A(t)$ to obtain the dynamical equations of general relativity as geodesic equations in superspace. The Hamiltonian constraints which remain after $H=0$ has been eliminated are obtained by varying I with respect to the normalized lapse function n. These constraints

$$\dot{S}^A[\delta G_{AB}/\delta n(P)]\dot{S}^B=0 \quad \text{for all positions } P \tag{19}$$

are similar in structure to the null condition which is imposed upon the trajectories of massless particles in spacetime. Like the null condition, these constraints are preserved by the geodesic equations.

It is useful to relate distance in superspace to the proper time of an observer at rest at a space point P. By evaluating the lapse scale factor

$$s=\pm(-\mathcal{T}/\mathcal{V})^{1/2}=\pm\frac{d\Sigma}{dt}\Big/\mathcal{V}$$

and using the relation

$$d\tau(P)=N(P)dt=sn(P)dt,$$

one obtains the desired relation

$$d\tau(P)=\mathcal{V}^{-1}n(P)d\Sigma. \tag{20}$$

[11] E. P. Belasco and H. C. Ohanian, J. Math. Phys. 10, 1503 (1969); A. Komar, *ibid.* 11, 820 (1970).

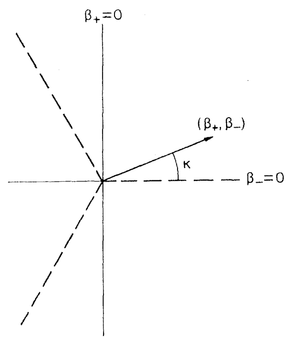


FIG. 1. Three-geometry belonging to the mixmaster truncation of superspace is represented by the point (β_+,β_-) in the β plane. The length of the position vector shown is a measure of the total anisotropy of the geometry—the departure of the metric from that of the three-sphere. The angle κ measures the departure from axial isotropy. Axially isotropic geometries which are space sections of Taub universes may be found along the dashed lines in this figure.

Equations (17) and (18) define a Riemannian metric on superspace for each choice of the normalized lapse function. These metrics are just the ones that DeWitt finds by requiring that the Hamilton-Jacobi equation of general relativity yields an eikonal equation in superspace.[6] All of these metrics contain the conformal factor $\mathcal{V}$ which vanishes everywhere on a hypersurface in superspace. This superhypersurface will be called the *nodal surface*. It includes the supersurface $R(P)=0$ which is the locus of time-symmetry points and the superhypersurface defined by $\int\sigma=0$, which is the locus of the final collapse of the universe as well as its initial "big bang".

The supermetric described by Eqs. (17) and (18) is hyperbolic and leads to three distinct classes of supercurves: *timelike* curves for which $\int d\Sigma$ is real, *null* curves for which $\int d\Sigma$ is zero, and *spacelike* curves for which $\int d\Sigma$ is imaginary. An important feature of the supergeometric approach to general relativity is that the spacelike curves are dynamically forbidden. By adopting the substitution $x^0 \to ix^0$ in spacetime coordinates, one can identify these forbidden supercurves with geometries of signature $(++++)$. Timelike supercurves represent geometries with the signature $(+++-)$ of spacetime. It is important to realize that this definition of "timelike" and "spacelike" depends *only* on the sign of $d\Sigma^2$. As the following example demonstrates, there can be several independent timelike directions in superspace and these directions are not necessarily those associated with the "minority" sign in the signature of the supermetric.

IV. TRUNCATED SUPERSPACE OF MIXMASTER UNIVERSES

Superspace is an infinite-dimensional manifold possessed of an infinite variety of metrics. To analyze this vast structure, one can truncate it by choosing a particular subspace of superspace and a particular set of normalized lapse functions. The most general type of truncation produces trajectories which correspond to approximate solutions of Einstein's equations. However, there is a simple way to construct exact truncations: If one chooses a truncated superspace which contains *all* three-geometries with a given group of motions and restricts the normalized lapse function to have the same group of motions, then the symmetry-preserving character of general relativity guarantees that solutions of the truncated geodesic equations and constraints will solve all of the supergeodesic equations and all of the constraints. An exact truncation which contains the closed mixmaster universes studied by Misner will now be considered.[12]

Description of Mixmaster Universes

The three-geometries which will be allowed in the mixmaster truncation of superspace are topologically three-spheres. They are homogeneous with respect to a three-dimensional group of motions—the $SU(2)$ group of left translations of unit quaternions—but are not isotropic. The homogeneity requirement forces n to be a constant which can be set equal to 1. In this truncation constraints (19) are already solved by the symmetry assumptions which leave no freedom to vary n. Because the homogeneity group $SU(2)$ is a covering group for the rotation group $SO(3)$, it is convenient to follow Misner in adopting the extended Euler angle coordinates $0\leqslant\psi<4\pi$, $0\leqslant\theta<\pi$, $0\leqslant\phi<2\pi$ on each three-geometry. In terms of the inexact differential forms

$$\sigma_1=\sin\psi d\theta-\cos\psi\,\sin\theta d\phi\,,$$
$$\sigma_2=\cos\psi d\theta+\sin\psi\,\sin\theta d\phi\,,$$
$$\sigma_3=-(d\psi+\cos\theta d\phi)\,,$$

the three-geometries are described by the line elements

$$ds^2=\tfrac{1}{4}a^2(e^{2\beta})_{ij}\sigma_i\sigma_j\,, \tag{21}$$

where a is the effective radius of the geometry $(\int\sigma=\frac{1}{8}a^3\int\sigma_1\wedge\sigma_2\wedge\sigma_3=2\pi^2a^3)$ and β is a position-independent traceless matrix which measures the anisotropy of the geometry ($\beta=0$ corresponds to complete isotropy).[12] The orientation of the space coordinates can be further specified by requiring σ_1, σ_2, and σ_3 to be the forms which are dual to the principal directions of β. The off-diagonal elements of β then vanish. The diagonal elements of β are conveniently expressed in the form

$$\beta_{11}=\beta_++\sqrt{3}\beta_-\,,\quad \beta_{22}=\beta_+-\sqrt{3}\beta_-\,,\quad \beta_{33}=-2\beta_+\,.$$

The three numbers a, β_+, and β_- are supercoordinates on this truncation of superspace.

It is helpful to understand in detail the way in which the supercoordinates β_+ and β_- characterize the anisotropy of a three-geometry. Consider the geometry represented in Fig. 1 as a point in the (β_+,β_-) plane. The length $|\beta|=(\beta_+{}^2+\beta_-{}^2)^{1/2}$ of the position vector shown is a measure of the total anisotropy of the geometry. At the origin of the β plane is found the metric of a three-sphere—a completely isotropic

[12] C. W. Misner, Gravity Award Essay, Gravity Research Foundation, New Boston, New Hampshire, 1967 (unpublished); C. W. Misner, Phys. Rev. **186**, 1319 (1969); and also I. M. Khalatnikov and E. M. Lifshitz, Zh. Eksperim. i Teor. Fiz. Pis'ma v Redaktsiyu **11**, 200 (1970) [JETP Letters **11**, 123 (1970)].

geometry. The angle which the position vector makes with the β_+ axis measures what may be called the axial anisotropy of the geometry—the departure from the case where there is, at each point in space, one axis about which the geometry is isotropic. Axial isotropy occurs when two of the eigenvalues of β are equal. At $\kappa=0°, \pm120°$, one finds "oblate" axially isotropic geometries for which the two equal eigenvalues are positive. At $\kappa=180°$, $\pm60°$, are the "prolate" axially isotropic geometries for which the two equal eigenvalues are negative.[13] Along these six directions in superspace lie those three-geometries which are space sections of Taub universes.

It is important to notice that the only distinction between geometries which lie in different 60° intervals of κ is the orientation of their coordinate systems. By adopting β_+ and β_- as unbounded coordinates on superspace, one imposes a particular topology on superspace. In this topology, each distinct three-geometry is represented by several points (either three or six) in superspace. A superspace with this type of topology is called an *extended superspace*. It is a covering space for *simple superspace* which has only one point for each three-geometry.[6] From the point of view of this paper, the simplest justification for the extended topology is that it leads to the most regular superspace metric. However, there is a deeper reason—the existence of strata in simple superspace.[14] In a simple superspace, those points which represent geometries with larger symmetry groups have different neighborhood structures than points with smaller symmetry groups. In mixmaster superspace, simple superspace is the region defined by $0\leqslant\kappa\leqslant60°$. Within this region, the a axis is a stratum representing isotropic geometries while the $\kappa=0$ half-plane is a stratum made up of axially isotropic geometries. This region is not a manifold because each point of the $\kappa=0$ half-plane has only half a neighborhood and, within the $\kappa=0$ half-plane, the a axis has only half a neighborhood. DeWitt has proposed that a manifold be constructed by joining many copies of simple superspace along their strata.[6] Here, six copies join together to make up the mixmaster superspace manifold.

Metric on Mixmaster Superspace

Now compute the metric tensor on the mixmaster truncation of superspace. Because of homogeneity, the conformal factor $\mathcal{V}$ in Eq. (18) is just $\mathcal{V}=-R\int\sigma$. But the β matrix is traceless so that the volume element $\sigma=d^3x\ g^{1/2}$ is unaffected by the values of β_+ and β_-. The volume $\int\sigma$ is just that of a three-sphere—$2\pi^2a^3$.

[13] C. W. Misner, in *Relativity: Proceeding of the Relativity Conference in the Midwest*, edited by L. Witten (Plenum, New York, 1970); D. J. Okerson, B. S. thesis, Princeton University, 1969 (unpublished).

[14] For more about strata, see Ref. 7 as well as A. E. Fischer, Ph.D. thesis, Princeton University, 1969 (unpublished); and A. F. Fischer, in *Relativity: Proceedings of the Relativity Conference in the Midwest*, edited by L. Witten (Plenum, New York, 1970).

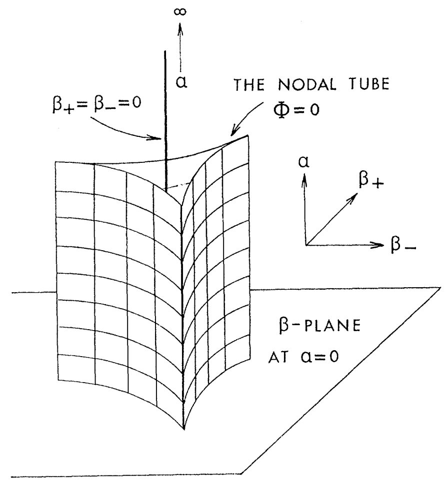


FIG. 2. Nodal surface (defined by $d\Sigma^2=0$) in mixmaster superspace is sketched here. The plane at $a=0$ is the site of spacetime singularities. The nodal tube is the place where all time-symmetric trajectories end and separates the "inside" region, where β_+ and β_- are the timelike coordinates, from the "outside" region, where a is the timelike coordinate. The maximum expansion stage of the universe always occurs within the nodal tube. This sketch shows the corners of the tube meeting at a finite distance from the a axis. Actually they meet at infinity.

The scalar curvature R may be computed easily by using the moving-frame techniques described by Flanders[15] and by Misner.[16] It may also be computed using results obtained by Khalatnikov and Lifshitz.[17] It is given by

$$R=(6/a^2)\Phi,$$

where

$$\Phi=\tfrac{1}{3}\operatorname{Tr}(2e^{-2\beta}-e^{4\beta}).$$

The positive definite "potential" V used by Misner[12] is related to Φ by $V=1-\Phi$. The functions ζ_{iA} which appear in Eq. (18) vanish because, in these homogeneous universes, the spacelike constraints (12) are solved by $N_r=0$. To perform the rest of the calculations called for by Eq. (18), use homogeneity to evaluate the space integral and obtain the derivatives $g_{ij,A}$ from Eq. (21). The line element on mixmaster superspace is then

$$d\Sigma^2=(6\pi^2a^2)^2\Phi d\mu^2, \tag{22}$$

where

$$d\mu^2=(d\beta_+)^2+(d\beta_-)^2-(a^{-1}da)^2. \tag{23}$$

The form of $d\mu^2$ suggests replacing a by its logarithm. To agree with Misner's choice of supercoordinates,[12,18] one can define a coordinate Ω which is related to a by

$$(6\pi)^{1/2}a=e^{-\Omega}. \tag{24}$$

[15] H. Flanders, *Differential Forms with Applications to the Physical Sciences* (Academic, New York, 1963), pp. 127–136.

[16] C. W. Misner, J. Math. Phys. 4, 924 (1963) (in an appendix).

[17] E. M. Lifshitz and I. M. Khalatnikov, Advan. Phys. **12**, 185 (1963).

[18] C. W. Misner, Phys. Rev. Letters 22, 1071 (1969).

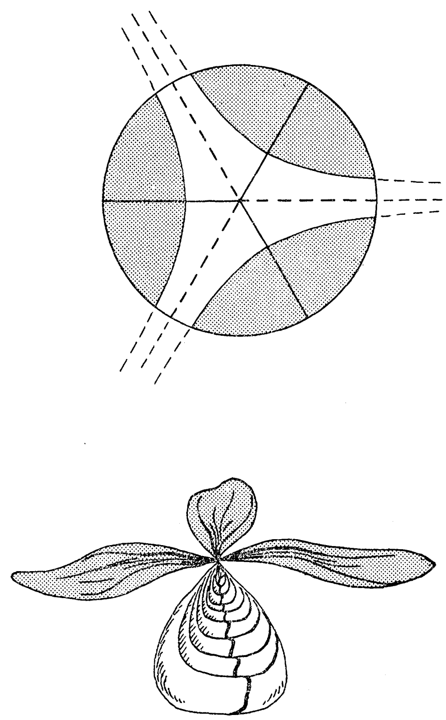


FIG. 3. Two ways to imbed an $a=\text{const}$ section of mixmaster superspace in three dimensions. Those parts of the section which lie outside of the nodal tube are shaded. If the metric $d\mu^2$ is to be preserved, then the section imbeds as a plane. A circular piece of this plane is shown. The straight dashed lines shown here correspond to the dashed lines in Fig. 1. If the metric $d\Sigma^2$ is to be preserved, then the conformal factor Φ shrinks distances along the nodal tube to zero. Thus, the section imbeds as a "bag" with the neck containing all of the points which are on the nodal tube. The heavy line which is wending its way among the folds of the bag corresponds to one of the dashed lines in the circular figure. It is the $a=\text{const}$ projection of the trajectory of a Taub universe.

The collapse singularity at $a=0$ is then removed from the supercoordinate patch and occurs at $\Omega=\infty$.

The key to understanding the geometry of this truncated superspace is the function $\Phi(\beta_+,\beta_-)$. This function has an absolute maximum value of 1 at the origin of the β plane. The level curves for $\Phi>0$ are continuous and enclose the origin. For $\Phi\leqslant 0$, the level curves lie at finite distance from the origin in all directions except for $\kappa=0$, $\pm\frac{2}{3}\pi$, where they go to infinity. In the latter three directions, Φ is positive and approaches zero as the distance from the origin of the β plane increases. A detailed description of the function Φ (actually, the related potential V) may be found elsewhere.[12,13,18,19] The curve defined by $\Phi=0$ is of particular interest. In the full (α,β) space, this curve traces out a nearly triangular tube which encloses the a axis. The resulting surface will be called the *nodal tube*.[20] It is displayed in Fig. 2 along with the branch of the nodal surface which lies at $a=0$. The figure does not show the edges of the tube in their proper location at infinity. Figure 3 shows an $a=\text{const}$ section of the superspace. The $\Phi=0$ hypersurface separates the superspace into two regions: The "*outside*" region is defined by $\Phi<0$ and is shaded in the figure. The "*inside*" region, defined by $\Phi>0$, is not shaded in the figure. Figure 3 also shows how an $a=\text{const}$ section of superspace may be imbedded in a flat three-space. Equations (22) and (23) show that, inside the nodal tube, a is a spacelike coordinate while β_+ and β_- are timelike. Outside of the nodal tube it is a which is the timelike coordinate while β_+ and β_- are spacelike.

[19] R. A. Matzner, L. C. Shepley, and J. B. Warren, Ann. Phys. (N. Y.) (to be published); M. P. Ryan, Ph.D. thesis, University of Maryland, 1970 (unpublished).

[20] The surface consists of three infinite sheets which form a tube if they are regarded as meeting at $|\beta|=\infty$, $\kappa=0°$, $\pm120°$.

Superspace Causality

Some of the properties of mixmaster universes can be deduced solely from the casual structure of superspace. It is this structure which outlaws empty isotropic universes—their trajectories all lie on the a axis and are spacelike. The point of maximum expansion of the universe can occur only within the nodal tube, where β_+ and β_- are timelike. Because the nodal tube is of limited extent in the β plane, the anisotropy of the universe at maximum expansion is correspondingly limited. Figure 4 shows how one can construct a null surface from the maximum expansion stage backwards towards $a=0$ or $\Omega=\infty$ to obtain similar limitations on the anisotropy of the universe at any given number of e-foldings from maximum expansion. Because of the shape of the nodal tube these limitations restrict only the type of anisotropy and not its total amount. Taub-like universes (at $\kappa=0°$, $\pm120°$) can have arbitrarily large amounts of anisotropy near maximum expansion.

Supergeodesic Equations

To obtain further information about mixmaster universes it is necessary to investigate the supergeodesic equations. A difficulty of principle appears as soon as one asks how to extend geodesics across the nodal tube. There is no way to read this information from the supermetric. Suppose, in Fig. 3, one asserts that geodesics are to cross the neck of the bag as smoothly in the imbedding space as possible. Now twist the bag about its neck by any angle and the prescription has been changed without altering the intrinsic metric of the mixmaster superspace. Fortunately, general relativity provides the additional information which is needed to cross the nodal tube uniquely: The three-geometry must evolve smoothly in universal proper time. By setting the curve parameter t equal to proper time τ, one can obtain a correct set of geodesic equations which remain regular across the nodal tube. A still more regular form of the correct geodesic equations can be obtained by setting t equal to a parameter ω, which is related to τ by $2d\tau=\pm a d\omega$.

So long as a is finite, smoothness in τ implies smoothness in ω.

From Eqs. (20), (22), and (23), the choice of ω as a curve parameter corresponds to the constraint

$$\dot{\beta}_+^2+\dot{\beta}_-^2-\dot{\Omega}^2=\Phi. \tag{25}$$

Now write the path-length action in the form

$$I=\int d\Sigma$$
$$=\int\left(6\pi^2a^2\Phi^{1/2}\frac{d\mu}{d\omega}\right)d\omega.$$

Varying this action with respect to Ω yields the equation

$$-2a^2\Phi^{1/2}\frac{d\mu}{d\omega}+\frac{d}{d\omega}\left(a^2\Phi^{1/2}\dot{\Omega}\frac{d\mu}{d\omega}\right)=0.$$

Now use constraint (25) in the form $d\mu/d\omega=\Phi^{1/2}$ to obtain one of the geodesic equations:

$$\ddot{\Omega}=2(\Phi+\dot{\Omega}^2). \tag{26}$$

Use constraint (25) again to find the alternate equation

$$\ddot{\Omega}=2(\dot{\beta}_+^2+\dot{\beta}_-^2). \tag{27}$$

Similarly, varying the action with respect to β_+ and β_- yields the remaining geodesic equations in the form

$$\begin{aligned}\ddot{\beta}_\pm-2\dot{\Omega}\dot{\beta}_\pm&=\tfrac{1}{2}\partial\Phi/\partial\beta_\pm\\&=-\tfrac{1}{2}\partial V/\partial\beta_\pm.\end{aligned} \tag{28}$$

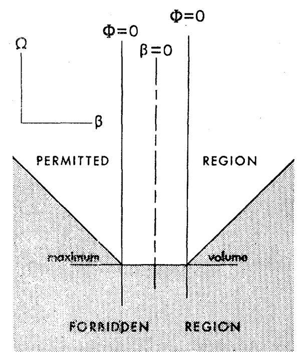
It is not difficult to show that the constraint (26) is conserved by the geodesic equations (26)–(28). Appendix B shows that these geodesic equations may be obtained from Misner's Hamiltonian formulation of mixmaster dynamics.

Properties of Geodesics in Mixmaster Superspace

Equation (27) shows that all supertrajectories experience an acceleration in the $+\Omega$ direction. The expansion rate of the universe must always be slowing until a maximum expansion point is reached. There must then be an accelerating collapse. By comparing Eqs. (26) and (27) one can also see that the following conditions are all equivalent. (1) Maximum expansion occurs on the nodal tube. (2) The maximum expansion occurs at a point of time symmetry. (3) The radius of the universe has an inflection point ($\ddot{\Omega}=0$) at the instant of maximum expansion.[21] Trajectories satisfying any one of these equivalent conditions are of particular interest. They are the time-symmetric solutions of Einstein's equations. At the point where it touches the nodal tube, such a trajectory has $\dot{\Omega}=\dot{\beta}_+=\dot{\beta}_-=0$ and

[21] These conditions are a special case of a general result which may be found in D. R. Brill, Nuovo Cimento Suppl. 2, 3 (1964).

FIG. 4. This diagram shows how superspace causality limits the anisotropy parameters (β_+ and β_-) of a universe which is a finite number of e-foldings (unit increments of Ω) from maximum volume.



its direction is given by the second derivatives $\ddot{\Omega}$ and $\ddot{\beta}_\pm$. From Eqs. (26) and (28), it can be seen that the trajectory is orthogonal to the nodal tube. This direction is permitted only when β_+ and β_- are timelike. Therefore, the time-symmetric trajectory must approach the nodal tube from within. The trajectory must stop on the nodal tube because there is no way to extend it further without violating superspace causality. The dynamical parameter ω reverses itself smoothly at the stopping point so that the collapse phase of the universe retraces the trajectory of the expansion phase.

The geodesic equations (28) reveal that the term $-2\dot{\Omega}$ acts as a damping coefficient or viscosity while

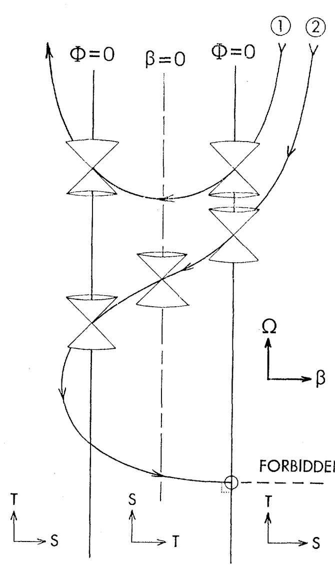


FIG. 5. Trajectories in mixmaster superspace. The symbols T and S at the bottom of the diagram indicate "timelike" and "spacelike," respectively. The nodal tube at $\Phi=0$ must be crossed in a lightlike direction. All trajectories are accelerated upward in this diagram. They lose energy during their downward travel and regain it as they fall upward. Trajectory 1 achieves its turnaround inside of the nodal tube and proceeds without interruption. Trajectory 2 achieves its minimum on the nodal tube. It cannot continue without violating superspace causality. Trajectory 1 corresponds to a universe undergoing expansion and collapse in a time-asymmetric way while trajectory 2 represents a time-symmetric universe—expanding as the trajectory is traced downward and collapsing as the trajectory is traced upward.

$\frac{1}{2}V$ is a potential function. Because V has the form of a well with steep walls (except for channels at $\kappa=0°$, $\pm 120°$), β_+ and β_- execute oscillations about the origin of the β plane. During the expansion phase of the universe, $-2\dot{\Omega}$ is positive and the oscillations are damped. During contraction the oscillations are amplified. As the $a=0$ singularity of superspace is approached, the oscillations become arbitrarily large and arbitrarily close together in proper time τ. This is the characteristic mixmaster singularity studied by Misner.[12,18]

The form of Eq. (28) suggests an anisotropy energy of the form

$$2E=\dot{\beta}_+^2+\dot{\beta}_-^2+V.$$

This is essentially the anisotropy energy defined by Misner.[12,18] During expansion, E is dissipated. During contraction, E increases again. Constraint (25) may be used to show that $2E$ never drops below 1 so that trajectories always have enough energy to climb out of the nodal tube. Figure 5 shows that trajectories which approach the nodal tube with $2E$ greater than 1 survive the reversal of spacelike and timelike coordinates by accelerating to the "speed of light." However, trajectories approaching the nodal tube from within with $2E$ equal to 1 are stopped by the singularity.

V. CONCLUSIONS

The physical action functional of general relativity, the action which is to be added to actions describing nongravitational degrees of freedom, may be regarded as a Riemannian path length in curved superspace. In this picture, the metric on curved superspace depends upon the lapse function N which measures the normal separation of equal-time hypersurfaces in spacetime. Einstein spacetimes are represented by timelike geodesics subject to conserved constraints in superspace. All of the superspace metrics are singular where $\int \sigma NR$ vanishes. The mixmaster truncation of superspace displays the two fixed branches of this singularity, the collapse branch at $a=0$ and the nodal tube which is the locus of time-symmetry points. In more complicated truncations of superspace one will also find singular hypersurfaces which depend upon the choice one makes for the lapse function N.

The basic virtue of the supergeometric approach to general relativity is that it brings the powerful tools of Riemannian geometry to bear upon the solution of Einstein's field equations. The familiar mathematical concepts that apply to particles moving in a curved spacetime may be used to analyze the dynamics of spacetime itself.

ACKNOWLEDGMENTS

I thank Professor Dieter R. Brill for his encouragement and helpful advice. I also thank Donald K. Walker for preparing the figures.

APPENDIX A: TRIVIAL GEOMETRIZATION

Consider a system with dynamical variables x^i ($i=1, 2, \ldots, n$) and Lagrangian

$$L=m_{ij}(x)\frac{dx^i}{dt}\frac{dx^j}{dt}-U(x)+R_k(x)\frac{dx^k}{dt}.$$

Now take the time t to be a variable on the same footing as the x^i by defining

$$x^0\equiv t,\quad m_{00}\equiv -U,\quad 2m_{0k}\equiv R_k.$$

The Lagrangian is then

$$L=m_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt},$$

where μ and ν are to be summed from 0 to n. Because the value of L is preserved whenever the equations of motion are satisfied, one can adopt the *new Lagrangian* $L'=L^{1/2}$ without altering the equations of motion. The resulting action functional $\int L'dt$ is a path length for the metric tensor $m_{\mu\nu}$ and solutions of the equations of motion are geodesics for this metric.

The idefinite metric $m_{\mu\nu}$ introduces a distinction between timelike and spacelike trajectories. However, in this trivial geometrization, L' is not the physical Lagrangian and the spacelike trajectories are allowed. To eliminate the spurious light-cone structure, one can add an infinite constant to the potential U. The apparent speed of light then becomes infinite.[22]

APPENDIX B: DERIVING GEODESIC EQUATIONS FROM MISNER'S HAMILTONIAN

In the canonical formalism of Arnowitt, Deser, and Misner, the Hamiltonian conjugate to Ω time is

$$H=(p^2-k^2a^4\Phi)^{1/2}, \tag{B1}$$

where $k=6\pi$ and

$$p^2\equiv p_+^2+p_-^2.$$

Hamilton's equations are then

$$\beta_\pm'=H^{-1}p_\pm, \tag{B2}$$

$$p_\pm'=-\partial H/\partial\beta_\pm, \tag{B3}$$

where primes denote derivatives with respect to Ω. From Eqs. (B1) and (B2) one obtains

$$p^2=k^2a^4\gamma^2\beta'^2, \tag{B4}$$

where

$$\gamma\equiv[\Phi/(\beta'^2-1)]^{1/2}. \tag{B5}$$

The numerical value of H is then given by

$$H=ka^2\gamma \tag{B6}$$

and $p_\pm$ has the expression

$$p_\pm=ka^2\gamma\beta_\pm' \tag{B7}$$

[22] P. Havas, Rev. Mod. Phys. **36**, 938 (1964).

when Eq. (B2) is satisfied. Equation (B3) can then be written in the form

$$a^{-2}\gamma(a^2\gamma\beta_\pm')' = -\tfrac{1}{2}\partial V/\partial\beta_\pm. \tag{B8}$$

To see if Eq. (B8) is equivalent to the geodesic equation (28), transform the curve parameter from Ω to ω. Constraint (25) yields the equation

$$d\Omega/d\omega = \pm\gamma\,, \tag{B9}$$

which connects the parameters Ω and ω. It is then a straightforward matter to compute $d^2\Omega/d\omega^2$ and show that the first geodesic equation (27) follows from the definition of ω alone. By using Eq. (B9) one can then show that

$$a^{-2}(a^2\dot{\beta}_\pm)^{\cdot} = a^{-2}\gamma(a^2\gamma\beta_\pm')'$$

so that Eq. (B8) reduces to the remaining geodesic equations (28).

PHYSICAL REVIEW D VOLUME 2, NUMBER 12 15 DECEMBER 1970

Insolubility of the Quantum Measurement Problem*

ARTHUR FINE
Sage School of Philosophy, Cornell University, Ithaca, New York 14850
(Received 27 January 1970)

The problem of whether a measurement interaction can leave the joint object-apparatus system in a mixture of states, in each state of which the apparatus's observable displays a definite value, is set within the most general quantum-theoretic framework for treating measurements. It is shown that the question posed by this problem admits only a negative answer. Some schemes for approximating the true object-apparatus state by means of such mixtures are examined. It is argued that such schemes constitute fundamental changes in the interpretation of quantum theory.

I. INTRODUCTION

THE quantum theory of measurement pursues the idealization where the measured object, the measuring apparatus, and the interaction between the two are each treated within the formalism of quantum theory. If both object and apparatus have, as measurement begins, a pure state, then, since the interaction between them is represented by a unitary motion on the joint object-apparatus space, the terminal state of the joint object-apparatus system will be a pure case in which, generally, neither the object nor the apparatus has a definite state. If one thinks of the apparatus as a macroscopic device—say, a pointer and scale—then the result that the apparatus has no state function is unacceptable. One may try to avoid this result by treating the initial state of the apparatus (more realistically, one may argue) as a mixed state and then hoping that the final state of the joint system will be a mixture of pure states in each of which the apparatus is itself in a pure state. The question of whether this can successfully be done is known as "the problem of measurement." For measurements satisfying von Neumann's account,[1] Wigner has shown that the problem of measurement cannot be solved affirmatively.[2] D'Espagnat[3] and Earman and Shimony[4] have generalized Wigner's argument for the broader class of measurements that fall under Landau's analysis.[5] I shall outline below the most general theory of measurement consistent with elementary quantum theory, an account which includes as special cases the theories of von Neumann and Landau, and by a somewhat different argument I shall show that no affirmative solution to the problem of measurement is possible. The remaining section will investigate the prospects for an approximate solution.

II. PROBLEM

We shall consider an object system with associated Hilbert space $\mathbf{H}_o$ and an apparatus system with space $\mathbf{H}_a$. An apparatus observable $\mathbf{A}$ with spectral resolution $\mathbf{A} = \sum \mu_n \mathbf{A}_n$ will be used to measure an object observable $\mathbf{O}$ with spectral resolution $\mathbf{O} = \sum \lambda_n \mathbf{O}_n$. The interaction will be treated in the tensor product space $\mathbf{H} = \mathbf{H}_o \otimes \mathbf{H}_a$. For generality, "states" will always be mixed, unless otherwise indicated, and the density operator-trace formalism will be used. Thus if initially the object has state $\mathbf{W}_o$ and the apparatus has state $\mathbf{W}_a$, the joint system will have state $\mathbf{W}_o \otimes \mathbf{W}_a$. The measurement is effected by means of a unitary motion $\mathbf{U}$ on $\mathbf{H}$ so that when the measurement terminates, the joint system has state $\mathbf{U}(\mathbf{W}_o \otimes \mathbf{W}_a)\mathbf{U}^{-1}$.

It is widely believed that there are no measurements that will leave both object and apparatus in definite

* Work supported by NFS under Grant No. GS-2034.

[1] J. von Neumann, *Mathematical Foundations of Quantum Mechanics*, translated by Robert T. Beyer (Princeton U.P., Princeton, 1955), Chaps. 5 and 6.

[2] E. P. Wigner, Am. J. Phys. **31**, 6 (1963). The same result is contained in A. Komar, Phys. Rev. **126**, 135 (1962).

[3] B. d'Espagnat, Nuovo Cimento Suppl. **4**, 828 (1966).

[4] J. Earman and A. Shimony, Nuovo Cimento **54B**, 332 (1968).

[5] L. Landau and R. Peierls, Z. Physik **69**, 56 (1931); L. Landau and E. Lifshitz, *Quantum Mechanics* (Pergamon, London, 1958), pp. 21–24.