PHYSICAL REVIEW D VOLUME 28, NUMBER 12 15 DECEMBER 1983

Wave function of the Universe

J. B. Hartle
Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637
and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

S. W. Hawking
Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge, England
and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106
(Received 29 July 1983)

The quantum state of a spatially closed universe can be described by a wave function which is a functional on the geometries of compact three-manifolds and on the values of the matter fields on these manifolds. The wave function obeys the Wheeler-DeWitt second-order functional differential equation. We put forward a proposal for the wave function of the "ground state" or state of minimum excitation: the ground-state amplitude for a three-geometry is given by a path integral over all compact positive-definite four-geometries which have the three-geometry as a boundary. The requirement that the Hamiltonian be Hermitian then defines the boundary conditions for the Wheeler-DeWitt equation and the spectrum of possible excited states. To illustrate the above, we calculate the ground and excited states in a simple minisuperspace model in which the scale factor is the only gravitational degree of freedom, a conformally invariant scalar field is the only matter degree of freedom and $\Lambda>0$. The ground state corresponds to de Sitter space in the classical limit. There are excited states which represent universes which expand from zero volume, reach a maximum size, and then recollapse but which have a finite (though very small) probability of tunneling through a potential barrier to a de Sitter-type state of continual expansion. The path-integral approach allows us to handle situations in which the topology of the three-manifold changes. We estimate the probability that the ground state in our minisuperspace model contains more than one connected component of the spacelike surface.

I. INTRODUCTION

In any attempt to apply quantum mechanics to the Universe as a whole the specification of the possible quantum-mechanical states which the Universe can occupy is of central importance. This specification determines the possible dynamical behavior of the Universe. Moreover, if the uniqueness of the present Universe is to find any explanation in quantum gravity it can only come from a restriction on the possible states available.

In quantum mechanics the state of a system is specified by giving its wave function on an appropriate configuration space. The possible wave functions can be constructed from the fundamental quantum-mechanical amplitude for a complete history of the system which may be regarded as the starting point for quantum theory.[1] For example, in the case of a single particle a history is a path $x(t)$ and the amplitude for a particular path is proportional to

$$\exp(iS[x(t)]) , \tag{1.1}$$

where $S[x(t)]$ is the classical action. From this basic amplitude, the amplitude for more restricted observations can be constructed by superposition. In particular, the amplitude that the particle, having been prepared in a certain way, is located at position x and nowhere else at time t is

$$\psi(x,t)=N\int_C \delta x(t)\exp(iS[x(t)]) . \tag{1.2}$$

Here, N is a normalizing factor and the sum is over a class of paths which intersect x at time t and which are weighted in a way that reflects the preparation of the system. $\psi(x,t)$ is the wave function for the state determined by this preparation. As an example, if the particle were previously localized at x' at time t' one would sum over all paths which start at x' at t' and end at x at t thereby obtaining the propagator $\langle x,t\,|\,x',t'\rangle$. The oscillatory integral in Eq. (1.2) is not well defined but can be made so by rotating the time to imaginary values.

An alternative way of calculating quantum dynamics is to use the Schrödinger equation,

$$i\partial\psi/\partial t=H\psi . \tag{1.3}$$

This follows from Eq. (1.2) by varying the end conditions on the path integral. For a particular state specified by a weighting of paths C, the path integral (1.2) may be looked upon as providing the boundary conditions for the solution of Eq. (1.3).

A state of particular interest in any quantum-mechanical theory is the ground state, or state of minimum excitation. This is naturally defined by the path integral, made definite by a rotation to Euclidean time, over the class of paths which have vanishing action in the far past. Thus, for the ground state at $t=0$ one would write

$$\psi_0(x,0)=N\int \delta x(\tau)\exp(-I[x(\tau)]) , \tag{1.4}$$

where $I[x(\tau)]$ is the Euclidean action obtained from S by sending $t \to -i\tau$ and adjusting the sign so that it is positive.

In cases where there is a well-defined time and a corresponding time-independent Hamiltonian, this definition of ground state coincides with the lowest eigenfunction of the Hamiltonian. To see this specialize the path-integral expression for the propagator $\langle x,t \mid x',t' \rangle$ to $t=0$ and $x'=0$ and insert a complete set of energy eigenstates between the initial and final state. One has

$$\langle x,0 \mid 0,t' \rangle = \sum_n \psi_n(x)\bar{\psi}_n(0)\exp(iE_n t') = \int \delta x(t)\exp(iS[x(t)]) \ , \tag{1.5}$$

where $\psi_n(x)$ are the time-independent energy eigenfunctions. Rotate $t' \to -i\tau'$ in (1.5) and take the limit as $\tau' \to -\infty$. In the sum only the lowest eigenfunction (normalized to zero energy) survives. The path integral becomes the path integral on the right of (1.4) so that the equality is demonstrated.

The case of quantum fields is a straightforward generalization of quantum particle mechanics. The wave function is a functional of the field configuration on a spacelike surface of constant time, $\Psi = \Psi[\phi(\vec{x}),t]$. The functional Ψ gives the amplitude that a particular field distribution $\phi(\vec{x})$ occurs on this spacelike surface. The rest of the formalism is similarly generalized. For example, for the ground-state wave functional one has

$$\Psi_0[\phi(\vec{x}),0] = N \int \delta\phi(x)\exp(-I[\phi(x)]) \ , \tag{1.6}$$

where the integral is over all Euclidean field configurations for $\tau < 0$ which match $\phi(\vec{x})$ on the surface $\tau = 0$ and leave the action finite at Euclidean infinity.

In the case of quantum gravity new features enter. For definiteness and simplicity we shall restrict our attention throughout this paper to spatially closed universes. For these there is no well-defined intrinsic measure of the location of a spacelike surface in the spacetime beyond that contained in the intrinsic or extrinsic geometry of the surface itself. One therefore labels the wave function by the three-metric h_{ij} writing $\Psi = \Psi[h_{ij}]$. Quantum dyanmics is supplied by the functional integral

$$\Psi[h_{ij}] = N \int_C \delta g(x)\exp(iS_E[g]) \ . \tag{1.7}$$

S_E is the classical action for gravity including a cosmological constant Λ and the functional integral is over all four-geometries with a spacelike boundary on which the induced metric is h_{ij} and which to the past of that surface satisfy some appropriate condition to define the state. In particular for the amplitude to go from a three-geometry h'_{ij} on an initial spacelike surface to a three-geometry h''_{ij} on a final spacelike surface is

$$\langle h''_{ij} \mid h'_{ij} \rangle = \int \delta g \exp(iS_E[g]) \ , \tag{1.8}$$

where the sum is over all four-geometries which match h'_{ij} on the initial surface and h''_{ij} on the final surface. Here one clearly sees that one cannot specify time in these states. The proper time between the surfaces depends on the four-geometries in the sum.

As in the mechanics of a particle the functional integral (1.7) implies a differential equation on the wave function. This is the Wheeler-DeWitt equation[2] which we shall derive from this point of view in Sec. II. With a simple choice of factor ordering it is

$$\left[-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}} - {}^3R(h)h^{1/2} + 2\Lambda h^{1/2} \right]\Psi[h_{ij}] = 0 \ , \tag{1.9}$$

where G_{ijkl} is the metric on superspace,

$$G_{ijkl} = \tfrac{1}{2}h^{-1/2}(h_{ik}h_{jl} + h_{il}h_{jk} - h_{ij}h_{kl}) \tag{1.10}$$

and ${}^3R$ is the scalar curvature of the intrinsic geometry of the three-surface. The problem of specifying cosmological states is the same as specifying boundary conditions for the solution of the Wheeler-DeWitt equation. A natural first question to ask is what boundary conditions specify the ground state?

In the quantum mechanics of closed universes we do not expect to find a notion of ground state as a state of lowest energy. There is no natural definition of energy for a closed universe just as there is no independent standard of time. Indeed in a certain sense the total energy for a closed universe is always zero—the gravitational energy canceling the matter energy. It is still reasonable, however, to expect to be able to define a state of minimum excitation corresponding to the classical notion of a geometry of high symmetry. This paper contains a proposal for the definition of such a ground-state wave function for closed universes. The proposal is to extend to gravity the Euclidean-functional-integral construction of nonrelativistic quantum mechanics and field theory [Eqs. (1.4) and (1.6)]. Thus, we write for the ground-state wave function

$$\Psi_0[h_{ij}] = N \int \delta g \exp(-I_E[g]) \ , \tag{1.11}$$

where I_E is the Euclidean action for gravity including a cosmological constant Λ. The Euclidean four-geometries summed over must have a boundary on which the induced metric is h_{ij}. The remaining specification of the class of geometries which are summed over determines the ground state. Our proposal is that the sum should be over compact geometries. This means that the Universe does not have any boundaries in space or time (at least in the Euclidean regime) (cf. Ref. 3). There is thus no problem of boundary conditions. One can interpret the functional integral over all compact four-geometries bounded by a given three-geometry as giving the amplitude for that three-geometry to arise from a zero three-geometry, i.e., a single point. In other words, the ground state is the amplitude for the Universe to appear from nothing.[4] In the following we shall elaborate on this construction and show in simple models that it indeed supplies reasonable wave functions for a state of minimum excitation.

The specification of the ground-state wave function is a constraint on the other states allowed in the theory. They must be such, for example, as to make the Wheeler-DeWitt equation Hermitian in an appropriate norm. In analogy with ordinary quantum mechanics one would expect to be able to use these constraints to extrapolate the boundary conditions which determine the excited states of the theory from those fixed for the ground state by Eq. (1.7). Thus, one can in principle determine all the allowed cosmological states.

The wave functions which result from this specification will not vanish on the singular, zero-volume three-geometries which correspond to the big-bang singularity. This is analogous to the behavior of the wave function of the electron in the hydrogen atom. In a classical treatment, the situation in which the electron is at the proton is singular. However, in a quantum-mechanical treatment the wave function in a state of zero angular momentum is finite and nonzero at the proton. This does not cause any problems in the case of the hydrogen atom. In the case of the Universe we would interpret the fact that the wave function can be finite and nonzero at the zero three-geometry as allowing the possibility of topological fluctuations of the three-geometry. This will be discussed further in Sec. VIII.

After a general discussion of this proposal for the ground-state wave function we shall implement it in a minisuperspace model. The geometrical degrees of freedom in the model are restricted to spatially homogeneous, isotropic, closed universes with S^3 topology, the matter degrees of freedom to a single, homogeneous, conformally invariant scalar field and the cosmological constant is assumed to be positive. A semiclassical evaluation of the functional integral for the ground-state wave function shows that it indeed does possess characteristics appropriate to a "state of minimum excitation."

Extrapolating the boundary conditions which allow the ground state to be extracted from the Wheeler-DeWitt equation, we are able to go further and identify the wave functions in the minisuperspace models corresponding to excited states of the matter field. These wave functions display some interesting features. One has a complete spectrum of excited states which show that a closed universe similar to our own and possessed of a cosmological constant can escape the big crunch and tunnel through to an eternal de Sitter expansion. We are able to calculate the probability for this transition.

In addition to the excited states we make a proposal for the amplitudes that the ground-state three-geometry consists of disconnected three-spheres thus giving a meaning to a gravitational state possessing different topologies.

Our conclusion will be that the Euclidean-functional-integral prescription (1.7) does single out a reasonable candidate for the ground-state wave function for cosmology which when coupled with the Wheeler-DeWitt equation yields a basis for constructing quantum cosmologies.

II. QUANTUM GRAVITY

In this section we shall review the basic principles and machinery of quantum gravity with which we shall explore the wave functions for closed universes. For simplicity we shall represent the matter degrees of freedom by a single scalar field ϕ, more realistic cases being straightforward generalizations. We shall approach this review from the functional-integral point of view although we shall arrive at many canonical results.[5] None of these are new and for different approaches to the same ends the reader is referred to the standard literature.[6]

A. Wave functions

Our starting point is the quantum-mechanical amplitude for the occurrence of a given spacetime and a given field history. This is

$$\exp(iS[g,\phi]) \,, \tag{2.1}$$

where $S[g,\phi]$ is the total classical action for gravity coupled to a scalar field. We are envisaging here a fixed manifold although there is no real reason that amplitudes for different manifolds may not be considered provided a rule is given for their relative phases. Just as the interesting observations of a particle are not typically its entire history but rather observations of position at different times, so also the interesting quantum-mechanical questions for gravity correspond to observations of spacetime and field on different spacelike surfaces. Following the general rules of quantum mechanics the amplitudes for these more restricted sets of observations are obtained from (2.1) by summing over the unobserved quantities.

It is easy to understand what is meant by fixing the field on a given spacelike surface. What is meant by fixing the four-geometry is less obvious. Consider all four-geometries in which a given spacelike surface occurs but whose form is free to vary off the surface. By an appropriate choice of gauge near the surface (e.g., Gaussian normal coordinates) all these four-geometries can be expressed so that the only freedom in the four-metric is the specification of the three-metric h_{ij} in the surface. Specifying the three-metric is therefore what we mean by fixing the four-geometry on a spacelike surface. The situation is not unlike gauge theories. There a history is specified by a vector potential $A_\mu(x)$ but by an appropriate gauge transformation $A_0(x)$ can be made to vanish so that the field on a surface can be completely specified by the $A_i(x)$.

As an example of the quantum-mechanical superposition principle the amplitude for the three-geometry and field to be fixed on two spacelike surfaces is

$$\langle h''_{ij},\phi'' \,|\, h'_{ij},\phi' \rangle = \int \delta g\, \delta\phi \exp(iS[g,\phi]) \,, \tag{2.2}$$

where the integral is over all four-geometries and field configurations which match the given values on the two spacelike surfaces. This is the natural analog of the propagator $\langle x'',t' \,|\, x',t' \rangle$ in the quantum mechanics of a single particle. We note again that the proper time between the two surfaces is not specified. Rather it is summed over in the sense that the separation between the surfaces depends on the four-geometry being summed over. It is not that one could not ask for the amplitude to have the three-geometry and field fixed on two surfaces *and* the proper time between them. One could. Such an amplitude, however, would not correspond to fixing observations on just two surfaces but rather would involve a set of intermediate observations to determine the time. It would therefore not be the natural analog of the propagator.

Wave functions Ψ are defined by

$$\Psi[h_{ij},\phi] = \int_C \delta g\, \delta\phi \exp(iS[g,\phi]) \,. \tag{2.3}$$

The sum is over a class C of spacetimes with a compact boundary on which the induced metric is h_{ij} and field configurations which match ϕ on the boundary. The remaining specification of the class C is the specification of the state.

If the Universe is in a quantum state specified by a wave function Ψ then that wave function describes the correlations between observables to be expected in that state. For example, in the semiclassical wave function describing a universe like our own, one would expect Ψ to be large when ϕ is big and the spatial volume is small, large when ϕ is small and the spatial volume is big, and small when these quantities are oppositely correlated. This is the only interpretative structure we shall propose or need.

B. Wheeler-DeWitt equation

A differential equation for Ψ can be derived by varying the end conditions on the path integral (2.3) which defines it. To carry out this derivation first recall that the gravitational action appropriate to keeping the three-geometry fixed on a boundary is

$$l^2 S_E = 2\int_{\partial M} d^3x\, h^{1/2} K + \int_M d^4x(-g)^{1/2}(R-2\Lambda)\ . \tag{2.4}$$

The second term is integrated over spacetime and the first over its boundary. K is the trace of the extrinsic curvature K_{ij} of the boundary three-surface. If its unit normal is n^i, $K_{ij}=-\nabla_i n_j$ in the usual Lorentzian convention. l is the Planck length $(16\pi G)^{1/2}$ in the units with $\hbar=c=1$ we use throughout. Introduce coordinates so that the boundary is a constant t surface and write the metric in the standard $3+1$ decomposition:

$$ds^2=-(N^2-N_iN^i)dt^2+2N_i dx^i dt+h_{ij}dx^i dx^j\ . \tag{2.5}$$

The action (2.4) becomes

$$l^2 S_E=\int d^4x\, h^{1/2} N[K_{ij}K^{ij}-K^2+{}^3R(h)-2\Lambda]\ , \tag{2.6}$$

where explicitly

$$K_{ij}=\frac{1}{N}\left[-\frac{1}{2}\frac{\partial h_{ij}}{\partial t}+N_{(i\,|\,j)}\right] \tag{2.7}$$

and a stroke and ${}^3R$ denote the covariant derivative and scalar curvature constructed from the three-metric h_{ij}. The matter action S_M can similarly be expressed as a function of N, N_i, h_{ij}, and the matter field.

The functional integral defining the wave function contains an integral over N. By varying N at the surface we push it forward or backward in time. Since the wave function does not depend on time we must have

$$0=\int \delta g\,\delta\phi\left[\frac{\delta S}{\delta N}\right]\exp(iS[g,\phi])\ . \tag{2.8}$$

More precisely, the value of the integral (2.3) should be left unchanged by an infinitesimal translation of the integration variable N. If the measure is invariant under translation this leads to (2.8). If it is not, there will be in addition a divergent contribution to the relation which must be suitably regulated to zero or cancel divergences arising from the calculation of the right-hand side of (2.8).

Classically the field equation $H\equiv\delta S/\delta N=0$ is the Hamiltonian constraint for general relativity. It is

$$H=h^{1/2}(K^2-K_{ij}K^{ij}+{}^3R-2\Lambda-l^2T_{nn})=0\ , \tag{2.9}$$

where T_{nn} is the stress-energy tensor of the matter field projected in the direction normal to the surface. Equation (2.8) shows how $H=0$ is enforced as an operator identity for the wave function. More explicitly one can note that the K_{ij} involve only first-time derivatives of the h_{ij} and therefore may be completely expressed in terms of the momenta π_{ij} conjugate to the h_{ij} which follow from the Lagrangian in (2.6):

$$\pi_{ij}=-h^{1/2}(K_{ij}-h_{ij}K)\ . \tag{2.10}$$

In a similar manner the energy of the matter field can be expressed in terms of the momentum conjugate to the field π_ϕ and the field itself. Equation (2.8) thus implies the operator identity $H(\pi_{ij},h_{ij},\pi_\phi,\phi)\Psi=0$ with the replacements

$$\pi^{ij}=-i\frac{\delta}{\delta h_{ij}}\ ,\quad \pi_\phi=-i\frac{\delta}{\delta\phi}\ . \tag{2.11}$$

These replacements may be viewed as arising directly from the functional integral, e.g., from the observation that when the time derivatives in the exponent are written in differenced form

$$-i\frac{\delta}{\delta h_{ij}}\int\delta g\,\delta\phi e^{iS}=\int\delta g\,\delta\phi\,\pi^{ij}e^{iS}\ . \tag{2.12}$$

Alternatively, they are the standard representation of the canonical commuation relations of h_{ij} and π^{ij}.

In translating a classical equation like $\delta S/\delta N=0$ into an operator identity there is always the question of factor ordering. This will not be important for us so making a convenient choice we obtain

$$\left\{-G_{ijkl}\frac{\delta^2}{\delta h_{ij}\delta h_{kl}}+h^{1/2}\left[-{}^3R(h)+2\Lambda+l^2T_{nn}\left[-i\frac{\delta}{\delta\phi},\phi\right]\right]\right\}\times\Psi[h_{ij},\phi]=0\ . \tag{2.13}$$

This is the Wheeler-DeWitt equation which wave functions for closed universes must satisfy. There are also the other constraints of the classical theory, but the operator versions of these express the gauge invariance of the wave function rather than any dynamical information.[6]

We should emphasize that the ground-state wave function constructed by a Euclidean functional-integral prescription [(Eq. (1.11)] will satisfy the Wheeler-DeWitt equation in the form (2.13). Indeed, this can be demonstrated explicitly by repeating the steps in the above demonstration starting with the Euclidean functional integral.

C. Boundary conditions

The quantity G_{ijkl} can be viewed as a metric on superspace—the space of all three-geometries (no connection with supersymmetry). It has signature

$(-,+,+,+,+,+)$ and the Wheeler-DeWitt equation is therefore a "hyperbolic" equation on superspace. It would be natural, therefore, to expect to impose boundary conditions on two "spacelike surfaces" in superspace. A convenient choice for the timelike direction is $h^{1/2}$ and we therefore expect to impose boundary conditions at the upper and lower limits of the range of $h^{1/2}$. The upper limit is infinity. The lower limit is zero because if h_{ij} is positive definite or degenerate, $h^{1/2} \geq 0$. Positive-definite metrics are everywhere spacelike surfaces; degenerate metrics may signal topology change. Summarizing the remaining functions of h_{ij} by the conformal metric $\tilde{h}_{ij} = h_{ij}/h^{1/3}$ we may write an important boundary condition on Ψ as

$$\Psi[\tilde{h}_{ij}, h^{1/2}, \phi] = 0, \quad h^{1/2} < 0 . \tag{2.14}$$

Because $h^{1/2}$ has a semidefinite range it is for many purposes convenient to introduce a representation in which $h^{1/2}$ is replaced by its canonically conjugate variable $-\frac{4}{3}Kl^{-2}$ which has an infinite range. The advantages of this representation have been extensively discussed.[7] In the case of pure gravity since $-\frac{4}{3}Kl^{-2}$ and $h^{1/2}$ are conjugate, we can write for the transformation to the representation where $\tilde{h}_{ij}$ and K are definite

$$\Phi[\tilde{h}_{ij}, K] = \int_0^\infty \delta h^{1/2} \exp\left[-i\tfrac{4}{3}l^{-2} \int d^3x\, h^{1/2}K\right] \Psi[h_{ij}] \tag{2.15}$$

and inversely,

$$\Psi[h_{ij}] = \int_{-\infty}^{+\infty} \delta K \exp\left[+i\tfrac{4}{3}l^{-2} \int d^3x\, h^{1/2}K\right] \Phi[\tilde{h}_{ij}, K] . \tag{2.16}$$

In each case the functional integrals are over the values of $h^{1/2}$ or K at each point of the spacelike hypersurface and we have indicated limits of integration.

The condition (2.14) implies through (2.15) that $\Phi[\tilde{h}_{ij}, K]$ is analytic in the lower-half K plane. The contour in (2.16) can thus be distorted into the lower-half K plane. Conversely, if we are given $\Phi[\tilde{h}_{ij}, K]$ we can reconstruct the wave function Ψ which satisfies the boundary condition (2.14) by carrying out the integration in (2.16) over a contour which lies below any singularities of $\Phi[\tilde{h}_{ij}, K]$ in K.

In the presence of matter K and $\tilde{h}_{ij}$ remain convenient labels for the wave functional provided the labels for the matter-field amplitudes $\tilde{\phi}$ are chosen so that a multiple of K is canonically conjugate to $h^{1/2}$. In cases where the matter-field action itself involves the scalar curvature this means that the label $\tilde{\phi}$ will be the field amplitude rescaled by some power of $h^{1/2}$. For example, in the case of a conformally invariant scalar field the appropriate label is $\tilde{\phi} = \phi h^{1/6}$. With this understanding we can write for the functionals

$$\Psi = \Psi[h_{ij}, \tilde{\phi}], \quad \Phi = \Phi[\tilde{h}_{ij}, K, \tilde{\phi}] \tag{2.17}$$

and the transformation formulas (2.15) and (2.16) remain unchanged.

D. Hermiticity

The introduction of wave functions as functional integrals [Eq. (2.3)] allows the definition of a scalar product with a simple geometric interpretation in terms of sums over spacetime histories. Consider a wave function Ψ defined by the integral

$$\Psi[h_{ij}, \phi] = N \int_C \delta g\, \delta\phi \exp(iS[g, \phi]) , \tag{2.18}$$

over a class of four-geometries and fields C, and a second wave function Ψ' defined by a similar sum over a class C'. The scalar product

$$(\Psi', \Psi) = \int \delta h\, \delta\phi\, \overline{\Psi}'[h_{ij}, \phi]\Psi[h_{ij}, \phi] \tag{2.19}$$

has the geometric interpretation of a sum over all histories

$$(\Psi', \Psi) = \overline{N}'N \int \delta g\, \delta\phi \exp(iS[g, \phi]) , \tag{2.20}$$

where the sum is over histories which lie in class C to the past of the surface and in the time reversed of class C' to its future.

The scalar product (2.19) is not the product that would be required by canonical theory to define the Hilbert space of physical states. That would presumably involve integration over a hypersurface in the space of all three-geometries rather than over the whole space as in (2.19). Rather, Eq. (2.19) is a mathematical construction made natural by the functional-integral formulation of quantum gravity.

In gravity we expect the field equations to be satisfied as identities. An extension of the argument leading to Eq. (2.8) will give

$$\int \delta g\, \delta\phi\, H(x) \exp(iS[g, \phi]) = 0 \tag{2.21}$$

for any class of geometries summed over and for any intermediate spacelike surface on which $H(x)$ is evaluated. Equation (2.21) can be evaluated for the particular sum which enters Eq. (2.20). $H(x)$ can be interpreted in the scalar product as an operator acting on either Ψ' or Ψ. Thus,

$$(H\Psi', \Psi) = (\Psi', H\Psi) = 0 . \tag{2.22}$$

The Wheeler-DeWitt operator must therefore be Hermitian in the scalar product (2.19).

Since the Wheeler-DeWitt operator is a second-order functional-differential operator, the requirement of Hermiticity will essentially be a requirement that certain surface terms on the boundary of the space of three-metrics vanish and, in particular, at $h^{1/2} = 0$ and $h^{1/2} = \infty$. As in ordinary quantum mechanics these conditions will prove useful in providing boundary conditions for the solution of the equation.

III. GROUND-STATE WAVE FUNCTION

In this section, we shall put forward in detail our proposal for the ground-state wave function for closed cosmologies. The wave function depends on the topology and the three-metric of the spacelike surface and on the values of the matter field on the surface. For simplicity we shall begin by considering only S^3 topology. Other possibilities will be considered in Sec. VIII.

As discussed in the Introduction, the ground-state wave function is to be constructed as a functional integral of the form

$$\Psi_0[h_{ij},\phi]=N\int \delta g\,\delta\phi\,\exp(-I[g,\phi])\,, \qquad (3.1)$$

where I is the total Euclidean action and the integral is over an appropriate class of Euclidean four-geometries with compact boundary on which the induced metric is h_{ij} and an appropriate class of Euclidean field configurations which match the value given on the boundary. To complete the definition of the ground-state wave function we need to give the class of geometries and fields to be summed over. Our proposal is that the geometries should be compact and that the fields should be regular on these geometries. In the case of a positive cosmological constant Λ any regular Euclidean solution of the field equations is necessarily compact.[8] In particular, the solution of greatest symmetry is the four-sphere of radius $3/\Lambda$, whose metric we write as

$$ds^2=(\sigma/H)^2(d\theta^2+\sin^2\theta\, d\Omega_3{}^2)\,, \qquad (3.2)$$

where $d\Omega_3{}^2$ is the metric on the three-sphere. $H^2=\sigma^2\Lambda/3$ and we have introduced the normalization factor $\sigma^2=l^2/24\pi^2$ for later convenience. Thus, it is clear that compact four-geometries are the only reasonable candidates for the class to be summed over when $\Lambda>0$.

If Λ is zero or negative there are noncompact solutions of the field equations. The solutions of greatest symmetry are Euclidean space ($\Lambda=0$) with

$$ds^2=\sigma^2(d\theta^2+\theta^2\, d\Omega_3{}^2) \qquad (3.3)$$

and Euclidean anti–de Sitter space ($\Lambda<0$) with

$$ds^2=(\sigma/H)^2(d\theta^2+\sinh^2\theta\, d\Omega_3{}^2)\,. \qquad (3.4)$$

One might therefore feel that the ground state for $\Lambda\leq 0$ should be defined by a functional integral over geometries which are asymptotically Euclidean or asymptotically anti–de Sitter. This is indeed appropriate to defining the ground state for scattering problems where one is interested in particles which propagate in from infinity and then out to infinity again.[9] However, in the case of cosmology, one is interested in measurements that are carried out in the interior of the spacetime, whether or not the interior points are connected to some infinite regions does not matter. If one were to use asymptotically Euclidean or anti–de Sitter four-geometries in the functional integral that defines the ground state one could not exclude a contribution from four-geometries that consisted of two disconnected pieces, one of which was compact with the three-geometry as boundary and the other of which was asymptotically Euclidean or anti–de Sitter with no interior boundary. Such disconnected geometries would in fact give the dominant contribution to the ground-state wave function. Thus, one would effectively be back with the prescription given above.

The ground-state wave function obtained by summing over compact four-geometries diverges for large three-geometries in the cases $\Lambda\leq 0$ and the wave function cannot be normalized. This is because the Λ in the action damps large four-geometries when $\Lambda>0$, but it enhances them when $\Lambda<0$. We shall therefore consider only the case $\Lambda>0$ in this paper and shall regard $\Lambda=0$ as a limiting case of $\Lambda>0$.

An equivalent way of describing the ground state is to specify its wave function in the $\tilde{\phi},\tilde{h}_{ij},K$ representation. Here too it can be constructed as a functional integral:

$$\Phi_0[\tilde{h}_{ij},K,\tilde{\phi}]=N\int \delta g\,\delta\phi\,\exp(-I^K[g,\phi])\,. \qquad (3.5)$$

The sum is over the same class of fields and geometries as before except that now $\tilde{\phi}$, $\tilde{h}_{ij}$, and K are fixed on the boundary rather than ϕ and h_{ij}. The action I^K is therefore the Euclidean action appropriate to holding $\tilde{\phi}$, $\tilde{h}_{ij}$, and K fixed on a boundary. It is a sum of the appropriate pure gravitational action which up to an additive constant is

$$l^2 I_E^K[g]=-\tfrac{2}{3}\int_{\partial M} d^3x\, h^{1/2}K-\int_M d^4x\, g^{1/2}(R-2\Lambda) \qquad (3.6)$$

and a contribution from the matter. The latter is well illustrated by the action of a single conformally invariant scalar field, an example which we shall use exclusively in the rest of this paper. We have

$$I_M^K[g,\phi]=\tfrac{1}{2}\int_M d^4x\, g^{1/2}[(\nabla\phi)^2+\tfrac{1}{6}R\phi^2]\,. \qquad (3.7)$$

These actions differ from the more familiar ones in which ϕ and h_{ij} are fixed only in having different surface terms. Indeed, these surface terms are just those required to ensure the equivalence of (3.1) and (3.5) as a consequence of the transformation formulas (2.15) and (2.16). In the case of the matter action of a conformally invariant scalar field with $\tilde{\phi},h_{ij},K$ fixed the additional surface term conveniently cancels that required in the action when ϕ and h_{ij} are fixed.

It is important to recognize that the functional integral (3.5) does not yield the wave function at the Lorentzian value of K but rather at a Euclidean value of K. For the moment denote the Lorentzian value by K_L. If the hypersurfaces of interest were labeled by a time coordinate t in a coordinate system with zero shift [$N_i=0$ in Eq. (2.5)] then the rotation $t\rightarrow i\tau$ and the use of the traditional conventions $K_L=-\nabla\cdot n$ and $K=\nabla\cdot n$ will send $K_L\rightarrow -iK$. In terms of the Euclidean K the transformation formulas (2.15) and (2.16) can be rewritten to read

$$\Phi[\tilde{h}_{ij},K,\tilde{\phi}]=\int_0^\infty \delta h^{1/2}\exp\left[-\tfrac{4}{3}l^{-2}\int d^3x\, h^{1/2}K\right]\times\Psi[h_{ij},\tilde{\phi}]\,, \qquad (3.8)$$

$$\Psi[h_{ij},\phi]=-\frac{1}{2\pi i}\int_C \delta K\,\exp\left[\tfrac{4}{3}l^{-2}\int d^3x\, h^{1/2}K\right]\times\Phi[\tilde{h}_{ij},K,\tilde{\phi}]\,, \qquad (3.9)$$

where the contour C runs from $-i\infty$ to $+i\infty$. At the risk of some confusion we shall continue to use K in the remainder of this paper to denote the Euclidean K despite having used the same symbol in Secs. I and II for the Lorentzian quantity.

There is one advantage to constructing the ground-state wave function from the functional integral (3.5) rather than (3.1) and it is the following: the integral in Eq. (3.9) will always yield a wave function $\Psi_0[h_{ij},\phi]$ which vanishes for $h^{1/2}<0$ if the contour C is chosen to the right of any singularities of $\Phi_0[\tilde{h}_{ij},K,\tilde{\phi}]$ in K provided Φ does not diverge too strongly in K. The boundary condition (2.14) is thus automatically enforced. This is a considerable advantage when the wave function is only evaluated approximately.

The Euclidean gravitational action [Eq. (3.6)] is not positive definite. The functional integrals in Eqs. (3.1) and (3.5) therefore require careful definition. One way of doing this is to break the integration up into an integral over conformal factors and over geometries in a given conformal equivalence class. By appropriate choice of the contour of integration of the conformal factor the integral can probably be made convergent. If this is the case a properly convergent functional integral can be constructed.

This then is our prescription for the ground state. In the following sections we shall derive some of its properties and demonstrate its reasonableness in a simple minisuperspace model.

IV. SEMICLASSICAL EXPECTATIONS

An important advantage of a functional-integral prescription for the ground-state wave function is that it yields the semiclassical approximation for that wave function directly. In this section, we shall examine the semiclassical approximation to the ground-state wave function defined in Sec. III. For simplicity we shall consider the case of pure gravity. The extension to include matter is straightforward.

The semiclassical approximation is obtained by evaluating the functional integral by the method of steepest descents. If there is only one stationary-phase point the semiclassical approximation is

$$\Psi_0[h_{ij}]=N\Delta^{-1/2}[h_{ij}]\exp(-I_{\rm cl}[h_{ij}]) \ . \tag{4.1}$$

Here, $I_{\rm cl}$ is the Euclidean gravitational action evaluated at the stationary-phase point, that is, at that solution $g_{\mu\nu}^{\rm cl}$ of the Euclidean field equations

$$R_{\mu\nu}=\Lambda g_{\mu\nu} \ , \tag{4.2}$$

which induces the metric h_{ij} on the closed three-surface boundary and satisfies the asymptotic conditions discussed in Sec. III. $\Delta^{-1/2}$ is a combination of determinants of the wave operators defining the fluctuations about $g_{\mu\nu}^{\rm cl}$ including those contributed by the ghosts. We shall focus mainly on the exponent. For further information on Δ in the case without boundary see Ref. 10.

If there is more than one stationary-phase point, it is necessary to consider the contour of integration in the path integal more carefully in order to decide which gives the dominant contribution. In general this will be the stationary-phase point with the lowest value of ReI although it may not be if there are two stationary-phase points which correspond to four-metrics that are conformal to one another. We shall see an example of this in Sec. VI. The ground-state wave function is real. This means that if the stationary-phase points have complex values of the action, there will be equal contributions from stationary-phase points with complex-conjugate values of the action. If there is no four-geometry which is a stationary-phase point, the wave function will be zero in the semiclassical approximation.

The semiclassical approximation for Ψ_0 can also be obtained by first evaluating the semiclassical approximation to Φ_0 from the functional integral (3.5) and then evaluating the transformation integral (3.9) by steepest descents. This will be more convenient to do when the boundary conditions of fixing $\tilde{h}_{ij}$ and K yield a unique dominant stationary-phase solution to (4.2) but fixing h_{ij} does not.

One can fix the normalization constant N in (4.1) by the requirement

$$\int \delta h\, \overline{\Psi}_0[h_{ij}]\Psi_0[h_{ij}]=1 \ . \tag{4.3}$$

As explained in Sec. II, one can interpret (4.3) geometrically as a path integral over all four-geometries which are compact on *both* sides of the three-surface with the metric h_{ij}. The semiclassical approximation to this path integral will thus be given by the action of the compact four-geometry without boundary which is the solution of the Einstein field equation. In the case of $\Lambda>0$ the solution with the most negative action is the four-sphere. Thus,

$$N^2=\exp\left[-\frac{2}{3H^2}\right] . \tag{4.4}$$

The semiclassical approximation for the wave function gives one considerable insight into the boundary conditions for the Wheeler-DeWitt equation, which are implied by the functional-integral prescription for the wave function. As discussed in Sec. II, these are naturally imposed on three-geometries of very large volumes and vanishing volumes.

Consider the limit of small three-volumes first. If the limiting three-geometry is such that it can be embedded in flat space then the classical solution to (4.2) when $\Lambda>0$ is the four-sphere and remains so as the three-geometry shrinks to zero. The action approaches zero. The value of the wave function is therefore controlled by the behavior of the determinants governing the fluctuations away from the classical solution. These fluctuations are to be computed about a vanishingly small region of a space of constant positive curvature. In this limit one can neglect the curvature and treat the fluctuations as about a region of flat space. The determinant can therefore be evaluated by considering its behavior under a constant conformal rescaling of the four-metric and the boundary three-metric. The change in the determinant under a change of scale is given by the value of the associated ζ function at zero argument.[11]

Regular four-geometries contain many hypersurfaces on which the three-volume vanishes. For example, consider the four-sphere of radius R embedded in a five-dimensional flat space. The three-surfaces which are the intersection of the four-sphere with surfaces of x^5 equals constant have a regular three-metric for $|x^5|<R$. The volume vanishes when $|x^5|=R$ at the north and south poles even though these are perfectly regular points of the four-geometry. One therefore would not expect the wave function to vanish at vanishing three-volume. Indeed, the three-volume will have to vanish somewhere if the topolo-

gy of the four-geometry is not that of a product of a three-surface with the real line or the circle. When the volume does vanish, the topology of the three-geometry will change. One cannot calculate the amplitude for such topology change from the Wheeler-DeWitt equation but one can do so using the Euclidean functional integral. We shall estimate the amplitude in some simple cases in Sec. VIII.

A qualitative discussion of the expected behavior of the wave function at large three-volumes can be given on the basis of the semiclassical approximation when $\Lambda>0$ as follows. The four-sphere has the largest volume of any real solution to (4.2). As the volume of the three-geometry becomes large one will reach three-geometries which no longer fit anywhere in the four-sphere. We then expect that the stationary-phase geometries become complex. The ground-state wave function will be a real combination of two expressions like (4.1) evaluated at the complex-conjugate stationary-phase four-geometries. We thus expect the wave function to oscillate as the volume of the three-geometry becomes large. If it oscillates without being strongly damped this corresponds to a universe which expands without limit.

The above considerations are only qualitative but do suggest how the behavior of the ground-state wave function determines the boundary conditions for the Wheeler-DeWitt equation. In the following we shall make these considerations concrete in a minisuperspace model.

V. MINISUPERSPACE MODEL

It is particularly straightforward to construct minisuperspace models using the functional-integral approach to quantum gravity. One simply restricts the functional integral to the restricted degrees of freedom to be quantized. In this and the following sections, we shall illustrate the general discussion of those preceding with a particularly simple minisuperspace model. In it we restrict the cosmological constant to be positive and the four-geometries to be spatially homogeneous, isotropic, and closed so that they are characterized by a single scale factor. An explicit metric in a useful coordinate system is

$$ds^2=\sigma^2[-N^2(t)dt^2+a^2(t)d\Omega_3{}^2]\ , \tag{5.1}$$

where $N(t)$ is the lapse function and $\sigma^2=l^2/24\pi^2$. For the matter degrees of freedom, we take a single conformally invariant scalar field which, consistent with the geometry, is always spatially homogeneous, $\phi=\phi(t)$. The wave function is then a function of only two variables:

$$\Psi=\Psi(a,\phi),\quad \Phi=\Phi(K,\tilde{\phi})\ . \tag{5.2}$$

Models of this general structure have been considered previously by DeWitt,[12] Isham and Nelson,[13] and Blyth and Isham.[14]

To simplify the subsequent discussion we introduce the following definitions and rescalings of variables:

$$\phi=\frac{\tilde{\phi}}{a}=\frac{\chi}{(2\pi^2\sigma^2)^{1/2}a}\ , \tag{5.3}$$

$$\Lambda=3\lambda/\sigma^2,\quad H^2=|\lambda|\ . \tag{5.4}$$

The Lorentzian action keeping χ and a fixed on the boundaries is

$$S^a=\tfrac{1}{2}\int dt\left[\frac{N}{a}\right]\left[\left[\frac{a}{N}\frac{da}{dt}\right]^2+a^2-\lambda a^4 + \left[\frac{a}{N}\frac{d\chi}{dt}\right]^2-\chi^2\right]. \tag{5.5}$$

From this action the momenta π_a and π_χ conjugate to a and χ can be constructed in the usual way. The Hamiltonian constraint then follows by varying the action with respect to the lapse function and expressing the result in terms of a, χ, and their conjugate momenta. One finds

$$\tfrac{1}{2}(-\pi_a{}^2-a^2+\lambda a^4+\pi_\chi{}^2+\chi^2)=0\ . \tag{5.6}$$

The Wheeler-DeWitt equation is the operator expression of this classical constraint. There is the usual operator-ordering problem in passing from classical to quantum relations but its particular resolution will not be central to our subsequent semiclassical considerations. A class wide enough to remind oneself that the issue exists can be encompassed by writing

$$\pi_a{}^2=-\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p\frac{\partial}{\partial a}\right], \tag{5.7}$$

although this is certainly not the most general form possible. In passing from the classical constraint to its quantum operator form there is also the possibility of a matter-energy renormalization. This will lead to an additive arbitrary constant in the equation. We thus write for the quantum version of Eq. (5.6)

$$\frac{1}{2}\left[\frac{1}{a^p}\frac{\partial}{\partial a}\left[a^p\frac{\partial}{\partial a}\right]-a^2+\lambda a^4-\frac{\partial^2}{\partial\chi^2}+\chi^2-2\epsilon_0\right]\times\Psi(a,\chi)=0\ . \tag{5.8}$$

A useful property stemming from the conformal invariance of the scalar field is that this equation separates. If we assume reasonable behavior for the function Ψ in the amplitude of the scalar field we can expand in harmonic-oscillator eigenstates

$$\Psi(a,\chi)=\sum_n c_n(a)u_n(\chi)\ , \tag{5.9}$$

where

$$\frac{1}{2}\left[-\frac{d^2}{d\chi^2}+\chi^2\right]u_n(\chi)=(n+\tfrac{1}{2})u_n(\chi)\ . \tag{5.10}$$

The consequent equation for the $c_n(a)$ is

$$\frac{1}{2}\left[-\frac{1}{a^p}\frac{d}{da}\left[a^p\frac{dc_n}{da}\right]+(a^2-\lambda a^4)c_n\right]=(n+\tfrac{1}{2}-\epsilon_0)c_n\ . \tag{5.11}$$

For small a this equation has solutions of the form

$$c_n\approx\text{constant},\quad c_n\approx a^{1-p} \tag{5.12}$$

[if p is an integer there may be a $\log(a)$ factor]. For large a the possible behaviors are

$$c_n \sim a^{-(p/2+1)}\exp(\pm\tfrac{1}{3}iHa^3) \ . \tag{5.13}$$

To construct the solution of Eq. (5.11) which corresponds to the ground state of the minisuperspace model we turn to our Euclidean functional-integral prescription. As applied to this minisuperspace model, the prescription of Sec. III for $\Psi_0(a_0,\chi_0)$ would be to sum $\exp(-I[g,\phi])$ over those Euclidean geometries and field configurations which are represented in the minisuperspace and which satisfy the ground-state boundary conditions. The geometrical sum would be over compact geometries of the form

$$ds^2=\sigma^2[d\tau^2+a^2(\tau)d\Omega_3{}^2] \tag{5.14}$$

for which $a(\tau)$ matches the prescribed value of a_0 on the hypersurface of interest. The prescription for the matter field would be to sum over homogeneous fields $\chi(\tau)$ which match the prescribed value χ_0 on the surface and which are regular on the compact geometry. Explicitly we could write

$$\Psi_0(a_0,\chi_0)=\int \delta a\, \delta\chi \exp(-I[a,\chi]) \ , \tag{5.15}$$

where, defining $d\eta=d\tau/a$, the action is

$$I=\tfrac{1}{2}\int d\eta\left[-\left[\frac{da}{d\eta}\right]^2-a^2+\lambda a^4+\left[\frac{d\chi}{d\eta}\right]^2+\chi^2\right] . \tag{5.16}$$

A conformal rotation [in this case of $a(\eta)$] is necessary to make the functional integral in (5.15) converge.[15]

An alternative way of constructing the ground-state wave function for the minisuperspace model is to work in the K representation. Here, introducing

$$k=\sigma K/9 \tag{5.17}$$

as a simplifying measure of K, one would have

$$\Phi_0(k_0,\chi_0)=\int \delta a\, \delta\chi \exp(-I^k[a,\chi]) \ . \tag{5.18}$$

The sum is over the same class of geometries and fields as in (5.15) except they must now assume the given value of k on the bounding three-surface. That is, on the boundary they must satisfy

$$k_0=\frac{1}{3a}\frac{da}{d\tau} \ . \tag{5.19}$$

The action I^k appropriate for holding k fixed on the boundary is

$$I^k=k_0a_0{}^3+I \tag{5.20}$$

[cf. Eq. (3.6)]. Once $\Phi_0(k_0,\chi_0)$ has been computed, the ground-state wave function $\Psi_0(k_0,\chi_0)$ may be recovered by carrying out the contour integral

$$\Psi_0(a_0,\chi_0)=-\frac{1}{2\pi i}\int_C dk\, e^{ka_0{}^3}\Phi_0(k,\chi_0) \ , \tag{5.21}$$

where the contour runs from $-i\infty$ to $+i\infty$ to the right of any singularities of $\Phi_0(k_0,\chi_0)$.

From the general point of view there is no difference between computing $\Psi_0(a_0,\chi_0)$ directly from (5.15) or via the K representation from (5.21). In Sec. VI we shall calculate the semiclassical approximation to $\Psi_0(a_0,\chi_0)$ both ways with the aim of advancing arguments that the rules of Sec. III define a wave function which may reasonably be considered as the state of minimal excitation and of displaying the boundary conditions under which Eq. (5.11) is to be solved.

VI. GROUND-STATE COSMOLOGICAL WAVE FUNCTION

In this section, we shall evaluate the ground-state wave function for our minisuperspace model and show that it possesses properties appropriate to a state of minimum excitation. We shall first evaluate the wave function in the semiclassical approximation from the steepest-descents approximation to the defining functional integral as described in Sec. IV. We shall then solve the Wheeler-DeWitt equation with the boundary conditions implied by the semiclassical approximation to obtain the precise wave function.

It is the exponent of the semiclassical approximation which will be most important in its interpretation. We shall calculate only this exponent from the extrema of the action and leave the determination of the prefactor [cf. Eq. (4.1)] to the solution of the differential equation. Thus, for example, if there were a single real Euclidean extremum of least action we would write for the semiclassical approximation to the functional integral in Eq. (5.15)

$$\Psi_0(a_0,\chi_0)\approx N e^{-I(a_0,\chi_0)} \ . \tag{6.1}$$

Here, $I(a_0,\chi_0)$ is the action (5.16) evaluated at the extremum configurations $a(\tau)$ and $\chi(\tau)$ which satisfy the ground-state boundary conditions spelled out in Sec. III and which match the arguments of the wave function on a fixed-τ hypersurface.

A. The matter wave function

A considerable simplification in evaluating the ground-state wave function arises from the fact that the energy-momentum tensor of an extremizing conformally invariant field vanishes in the compact geometries summed over as a consequence of the ground-state boundary conditions. One can see this because the compact four-geometries of the class we are considering are conformal to the interior of three-spheres in flat Euclidean space. A constant scalar field is the only solution of the conformally invariant wave equation on flat space which is a constant on the boundary three-sphere. The energy-momentum tensor of this field is zero. This implies that it is zero in any geometry of the class (5.14) because the energy-momentum tensor of a conformally invariant field scales by a power of the conformal factor under a conformal transformation.

More explicitly in the minisuperspace model we can show that the matter and gravitational functional integrals in (5.15) may be evaluated separately. The ground-state boundary conditions imply that geometries in the sum are conformal to half of a Euclidean Einstein-static universe, i.e., that the range of η is $(-\infty,0)$. The boundary conditions at infinite η are that $\chi(\eta)$ and $a(\eta)$ vanish. The boundary conditions at $\eta=0$ are that $a(0)$ and $\chi(0)$ match the arguments of the wave function a_0 and χ_0. Thus, not only does the action (5.16) separate into a sum of a gravitational part and a matter part, but the boundary conditions on the $a(\eta)$ and $\chi(\eta)$ summed over do not depend on one another. The matter and gravitational integrals can thus be evaluated separately.

Let us consider the matter integral first. In Eq. (5.16) the matter action is

$$I_M=\tfrac{1}{2}\int d\eta\left[\left[\frac{d\chi}{d\eta}\right]^2+\chi^2\right] . \tag{6.2}$$

This is the Euclidean action for the harmonic oscillator. Evaluation of the matter field integral in (5.15) therefore gives

$$\Psi_0(a_0,\chi_0)=e^{-\chi_0^2/2}\psi_0(a_0) . \tag{6.3}$$

Here, $\psi_0(a)$ is the wave function for gravity alone given by

$$\psi_0(a_0)=\int \delta a \exp(-I_E[a]) , \tag{6.4}$$

I_E being the gravitational part of (5.16). Equivalently we can write in the K representation

$$\Phi_0(k_0,\chi_0)=e^{-\chi_0^2/2}\phi_0(k_0) , \tag{6.5}$$

where

$$\phi_0(k_0)=\int \delta a \exp(-I_E^k[a]) . \tag{6.6}$$

$I_E{}^k[a]$ is related to I_E as in (5.20) and the sum is over $a(\tau)$ which satisfy (5.19) on the boundary. Equation (6.3) shows that as far as the matter field is concerned, $\Psi_0[a_0,\chi_0]$ is reasonably interpreted as the ground-state wave function. The field oscillators are in their state of minimum excitation—the ground state of the harmonic oscillator. We now turn to a semiclassical calculation of the gravitational wave function $\psi_0(a_0)$.

B. The semiclassical ground-state gravitational wave function

The integral in (6.4) is over $a(\tau)$ which represent [through (5.14)] compact geometries with three-sphere boundaries of radius a. The integral in (6.6) is over the same class of geometries except that the three-sphere boundary must possess the given value of k. The compact geometry which extremizes the gravitational action in these cases is a part of the Euclidean four-sphere of radius $1/H$ with an appropriate three-sphere boundary. In the case where the three-sphere radius is fixed on the boundary there are two extremizing geometries. For one the part of the four-sphere bounded by the three-sphere is greater than a hemisphere and for the other it is less. A careful analysis must therefore be made of the functional integral to see which of these extrema contributes to the semiclassical approximation. We shall give such an analysis below but first we show that the correct answer is achieved more directly in the K representation from (6.6) because there is a single extremizing geometry with a prescribed value of k on a three-sphere boundary and thus no ambiguity in constructing the semiclassical approximation to (6.6).

For three-sphere hypersurfaces of the four-sphere with an outward pointing normal, k ranges from approaching $+\infty$ for a surface encompassing a small region about a pole to approaching $-\infty$ for the whole four-sphere (see Fig. 1). More exactly, in the notation of Eq. (3.7)

$$k=\frac{H}{3}\cot\theta . \tag{6.7}$$

The extremum action is constructed through (5.20) with the integral in (5.16) being taken over that part of the four-sphere bounded by the three-sphere of given k. It is

$$I_E^k(k)=-\frac{1}{3H^2}\left[1-\frac{\kappa}{(\kappa^2+1)^{1/2}}\right] , \tag{6.8}$$

where

$$k=\tfrac{1}{3}\kappa H . \tag{6.9}$$

The semiclassical approximation to (6.6) is now

$$\phi_0(k_0)\approx N\exp[-I_E^k(k_0)] . \tag{6.10}$$

The wave function $\psi_0(a_0)$ in the same approximation can be constructed by carrying out the contour integral

$$\psi_0(a_0)=-\frac{N}{2\pi i}\int_C dk \exp[ka_0{}^3-I_E{}^k(k)] \tag{6.11}$$

by the method of steepest descents. The exponent in the integrand of Eq. (6.11) is minus the Euclidean action for pure gravity with a kept fixed instead of k:

$$I_E(a)=-ka^2+I_E{}^k(k) . \tag{6.12}$$

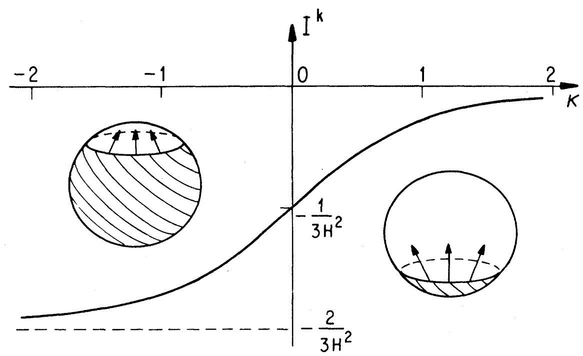

FIG. 1. The action I^k for the Euclidean four-sphere of radius $1/H$. The Euclidean gravitational action for the part of a four-sphere bounded by a three-sphere of definite K is plotted here as a function of κ (a dimensionless measure of K [Eq. (6.9)]). The action is that appropriate for holding K fixed on the boundary. The shaded regions of the inset figures show schematically the part of the four-sphere which fills in the three-sphere of given K used in computing the action. A three-sphere of given K fits in a four-sphere at only one place. Three-spheres with positive K (diverging normals) bound less than a hemisphere of four-sphere while those with negative K (converging normals) bound more than a hemisphere. The action tends to its flat-space value (zero) as K tends to positive infinity. It tends to the Euclidean action for all of de Sitter space as K tends to negative infinity.

To evaluate (6.11) by steepest descents we must find the extrema of Eq. (6.12). There are two cases depending on whether Ha_0 is greater or less than unity.

For $Ha_0<1$ the extrema of $I_E(k)$ occur at real values of k which are equal in magnitude and opposite in sign. They are the values of k at which a three-sphere of radius a_0 would fit into the four-sphere of radius $1/H$. That is, they are those values of k for which Eq. (6.7) is satisfied with $a_0{}^2=(\sin\theta/H)^2$. This is not an accident; it is a consequence of the Hamilton-Jacobi theory. The value of I_E at these extrema is

$$I_\pm=-\frac{1}{3H^2}[1\pm(1-H^2a_0{}^2)^{3/2}]\ , \tag{6.13}$$

where the upper sign corresponds to $k<0$ and the lower to $k>0$, i.e., to filling in the three-sphere with greater than a hemisphere of the four-sphere or less than a hemisphere, respectively.

There are complex extrema of I_E but all have actions whose real part is greater than the real extrema described above. The steepest-descents approximation to the integral (6.11) is therefore obtained by distorting the contour into a steepest-descents path (or sequence of them) passing through one or the other of the real extrema. The two real extrema and the corresponding steepest-descents directions are shown in Fig. 2. One can distort the contour into a steepest-descents path passing through only one of them—the one with positive k as shown. The functional integral thus singles out a unique semiclassical approximation to $\psi_0(a_0)$ which is

$$\psi_0(a_0)\approx N\exp[-I_-(a_0)],\quad Ha_0<1\ , \tag{6.14}$$

corresponding to filling in the three-sphere with less than a hemisphere's worth of four-sphere.

From Eq. (4.4) we recover the normalization factor N:

$$N=\exp(-\tfrac{1}{3}H^{-2})\ . \tag{6.15}$$

Thus, for $Ha_0\ll 1$

$$\psi_0(a_0)=\exp(\tfrac{1}{2}a_0{}^2-\tfrac{1}{3}H^{-2})\ . \tag{6.16}$$

One might have thought that the extremum I_+, which corresponds to filling in the three-geometry with more than a hemisphere, would provide the dominant contribution to the ground-state wave function as $\exp(-I_+)$ is greater than $\exp(-I_-)$. However, the steepest-descents contour in the integral (6.7) does not pass through the extremum coresponding to I_+. This is related to the fact that the contour of integration of the conformal factor has to be rotated in the complex plane in order to make the path integral converge as we shall show below.

For $Ha_0>1$ there are no real extrema because we cannot fit a three-sphere of radius $a_0>1/H$ into a four-sphere of radius $1/H$. There are, however, complex extrema of smallest real action located at

$$k=\pm\frac{i}{3}H\left[1-\frac{1}{H^2a_0{}^2}\right]^{1/2}\ . \tag{6.17}$$

It is possible to distort the contour in Eq. (6.11) into a steepest-descents contour passing through both of them as shown in Fig. 3. The resulting wave function has the form

$$\psi_0(a_0)=2\cos\left[\frac{(H^2a_0{}^2-1)^{3/2}}{3H^2}-\frac{\pi}{4}\right]\ ,\quad Ha_0>1 \tag{6.18}$$

or for $Ha_0\gg 1$

$$\psi_0(a_0)\approx e^{+iHa_0{}^3/3}+e^{-iHa_0{}^3/3}\ . \tag{6.19}$$

The semiclassical approximation to the ground-state gravitational wave function $\psi_0(a)$ contained in Eqs. (6.16) and (6.19) may also be obtained directly from the functional integral (6.4) without passing through the k representation. We shall now sketch this derivation. We must consider explicitly the conformal rotation which makes the gravitational part of the action in (5.16) positive definite. The gravitational action is

$$I_E[a]=\tfrac{1}{2}\int d\eta\left[-\left[\frac{da}{d\eta}\right]^2-a^2+H^2a^4\right]\ . \tag{6.20}$$

If one performed the functional integration

$$\psi_0(a_0)=\int\delta a(\eta)\exp(-I_E[a]) \tag{6.21}$$

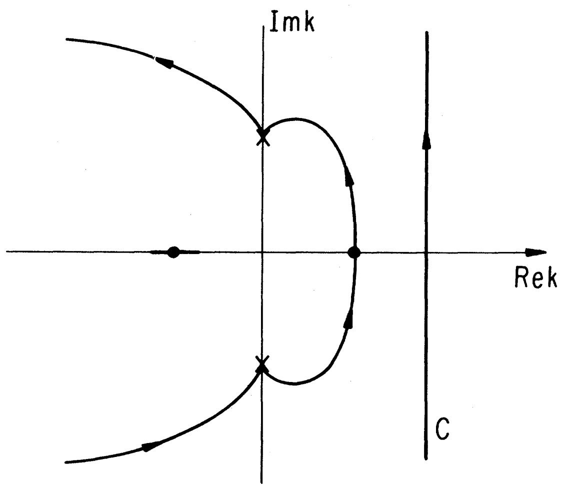

FIG. 2. The integration contour for constructing the semiclassical ground-state wave function of the minisuperspace model in the case $\Lambda>0$, $Ha_0<1$. The figure shows schematically the original integration contour C used in Eq. (6.11) and the steepest-descents contour into which it can be distorted. The branch points of the exponent of Eq. (6.11) at $\kappa=\pm i$ are located by crosses. There are two extrema of the exponent which correspond to filling in the three-sphere of given radius a with greater than a hemisphere of four-sphere or less than a hemisphere. For $Ha_0<1$ they lie at the equal and opposite real values of K indicated by dots. The contour C can be distorted into a steepest-descents contour through the extremum with positive K as shown. It cannot be distorted to pass through the extremum with negative K in the steepest-descents direction indicated. The contour integral thus picks out the extremum corresponding to less than a hemisphere of four-sphere (cf. Fig. 1) as the leading term in the semiclassical approximation.

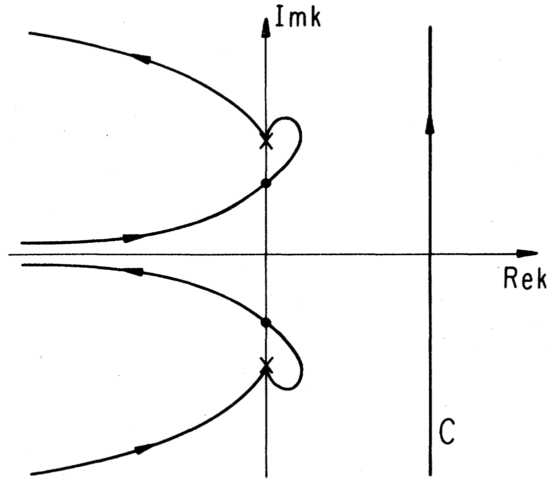

FIG. 3. The integration contour for constructing the semiclassical ground-state wave function of the minisuperspace model in the case $\Lambda > 0$, $Ha_0 > 1$. The figure shows schematically the original contour C used in Eq. (6.11) and the steepest-descents contour into which it can be distorted. The branch points of the exponent of Eq. (6.11) at $\kappa = \pm i$ are located by crosses. There are two complex-conjugate extrema of the exponent as indicated by dots and the contour C can be distorted to pass through both along the steepest-descents directions at 45° to the real axis as shown.

over real values of a, one would obtain a divergent result because the first term in (6.20) is negative definite. One could make the action infinitely negative by choosing a rapidly varying a. The solution to this problem seems to be to integrate the variable a in Eq. (6.21) along a contour that is parallel to the imaginary axis.[15] For each value of η, the contour of integration of a will cross the real axis at some value. Suppose there is some real function $\bar{a}(\eta)$ which maximizes the action. Then if one distorts the contour of integration of a at each value of η so that it crosses the real axis at $\bar{a}(\eta)$, the value of the action at the solution $\bar{a}(\eta)$ will give the saddle-point approximation to the functional integral (6.21), i.e.,

$$\psi_0(a_0) \approx N \exp(-I_E[\bar{a}(\eta)]) . \tag{6.22}$$

If there were another real function $\hat{a}(\eta)$ which extremized the action but which did not give its maximum value there would be a nearby real function $\hat{a}(\eta)+\delta a(\eta)$ which has a greater action. By choosing the contour of integration in (6.21) to cross the real a axis at $\hat{a}(\eta)+\delta a(\eta)$, one would get a smaller contribution to the ground-state wave function. Thus, the dominant contribution comes from the real function $\bar{a}(\eta)$ with the greatest value of the action.

It may be that there is no real $a(\eta)$ which maximizes the action. In this case the dominant contribution to the ground-state wave function will come from complex functions $a(\eta)$ which extremize the action. These will occur in complex-conjugate pairs because the wave function is real.

In the case of $Ha_0 < 1$, we have already seen that there are two real functions $a(\eta)$ which extremize the action and which correspond to less than or more than a hemisphere of the four-sphere. Their actions are I_- and I_+, respectively, given by (6.13). In fact, I_- is the maximum value of the action for real $a(\eta)$ and therefore gives the dominant contribution to the ground-state wave function. Thus, we again recover Eqs. (6.14) and (6.16). In the case of $Ha_0 > 1$, there is no maximum of the action for real $a(\eta)$. In this case the dominant contribution to the ground-state wave function comes from a pair of complex-conjugate $a(\eta)$ which extremize the action. Thus, we would expect an oscillatory wave function like that given by Eq. (6.19).

C. Ground-state solution of the Wheeler-DeWitt equation

The ground-state wave function must be a solution of the Wheeler-DeWitt equation for the minisuperspace model [Eqs. (5.8) or (5.11)]. The $\exp(-\chi^2/2)$ dependence of the wave function on the matter field deduced in Sec. VI A shows that in fact $\psi_0(a)$ must solve Eq. (5.11) with $n=0$. There are certainly solutions of this equation which have the large-a combination of exponentials required of the semiclassical approximation by Eq. (6.19) as a glance at Eq. (5.13) shows. In fact the prefactor in these asymptotic behaviors shows that the ground-state wave function will be *normalizable* in the norm

$$(\psi_0', \psi_0) = \int_0^\infty da\, a^p \bar{\psi}_0(a) \psi_0(a) \tag{6.23}$$

in which the Wheeler-DeWitt operator is Hermitian.

The Wheeler-DeWitt equation enables us to determine the prefactor in the semiclassical approximation from the standard WKB-approximation formulas. With $p=0$, for example, this would give when $Ha_0 > 1$

$$\psi_0(a_0) = 2(H^2 a_0^4 - a_0^2 + \epsilon_0 + \tfrac{1}{2})^{-1/4} \times \cos\left[\frac{(H^2 a_0^2 - 1)^{3/2}}{3H^2} - \frac{\pi}{4}\right] . \tag{6.24}$$

We could also solve the equation numerically. Figure 4 gives an example when $p=0$ and $\epsilon_0 = -\frac{1}{2}$. There we have assumed that the wave function vanishes at $a=0$. The dotted lines represent graphs of the prefactor in Eq. (6.24) and show that the semiclassical approximation becomes rapidly more accurate as Ha increases beyond 1. We shall return to an interpretation of these facts below.

D. Correspondence with de Sitter space

Having obtained $\psi_0(a)$, we are now in a position to assess its suitability as the ground-state wave function. Classically the vacuum geometry with the highest symmetry, hence minimum excitation, is de Sitter space—the surface of a Lorentz hyperboloid in a five-dimensional Lorentz-signatured flat spacetime. The properties of the wave function contained in Eqs. (6.16) and (6.19) are those one would expect to be semiclassically associated with this geometry. Sliced into three-spheres de Sitter space contains spheres only with a radius greater than $1/H$. Equation (6.16) shows that the wave function is an exponential-

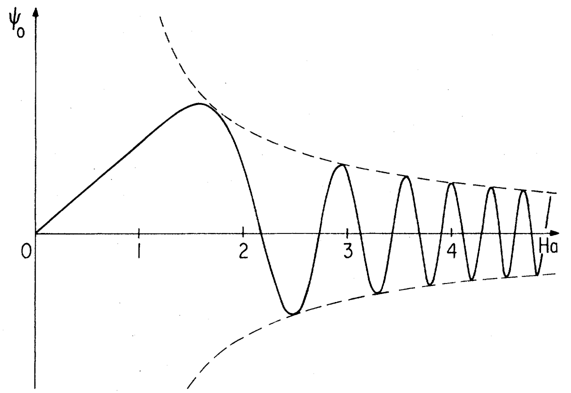

FIG. 4. A numerical solution of the Wheeler-DeWitt equation for the ground-state wave function $\psi_0(a)$. A solution of Eq. (5.11) is shown for $H=1$ in Planck units. We have assumed for definiteness $p=0$, $\epsilon_0=-\frac{1}{2}$, and a vanishing wave function at the origin. The wave function is damped for $Ha<1$ corresponding to the absence of spheres of radii smaller than H^{-1} in Lorentzian de Sitter space. It oscillates for $Ha>1$ decaying only slowly for large a. This reflects the fact that de Sitter space expands without limit. In fact, the envelope represented by the dotted lines is the distribution of three-spheres in Lorentzian de Sitter space: $[Ha(H^2a^2-1)^{1/2}]^{-1}$.

ly decreasing function with decreasing a for radii below that radius. Equation (6.24) shows the spheres of radius larger than $1/H$ are found with an amplitude which varies only slowly with the radius. This is a property expected of de Sitter space which expands both to the past and the future without limit. Indeed, tracing the origin of the two terms in (6.19) back to extrema with different signs of k one sees that one of these terms corresponds to the contracting phase of de Sitter space while the other corresponds to the expanding phase. The slow variation in the amplitude of the ground-state wave function reflects precisely the distribution of three-spheres in classical de Sitter space. Lorentzian de Sitter space is conformal to a finite region of the Einstein static universe

$$ds^2=\sigma^2a^2(\eta)(-d\eta^2+d\Omega_3^2)\ , \tag{6.25}$$

where $a(t)=(\cosh Ht)/H$ and $dt=a\,d\eta$. Three-spheres are evidently distributed uniformly in η in the Einstein static universe. The distribution of spheres in a in Lorentzian de Sitter space is therefore proportional to

$$[a(H^2a^2-1)^{1/2}]^{-1}\ . \tag{6.26}$$

This is the envelope of the probability distribution $a^p\,|\,\psi(a)\,|^2$ for spheres of radius a deduced from the semiclassical wave function and shown in Fig. 4. The wave function constructed from the Euclidean prescription of Sec. III appropriately reflects the properties of the classical vacuum solution of highest symmetry and is therefore reasonably called the ground-state wave function.

VII. EXCITED STATES

Our Universe does not correspond to the ground state of the simple minisuperspace model. It might be that the inclusion of more degrees of freedom in the model would produce a ground state which resembles our Universe more closely or it might be that we do not live in the ground state but in an excited state. Such excited states are not to be calculated by a simple path-integral prescription, but rather by solving the Wheeler-DeWitt equation with the boundary conditions that are required to maintain Hemiticity of the Hamiltonian operator between these states and the ground state. In this section, we shall construct the excited states for the minisuperspace model discussed in Sec. VI.

In the minisuperspace model where the spacelike sections are metric three-spheres all excitations in the gravitational degrees of freedom have been frozen out. We can study, however, excitations in the matter degrees of freedom. These are labeled by the harmonic-oscillator quantum number n as we have already seen [cf. Eq. (5.10)]. The issue then is what solution of Eq. (5.11) for $c_n(a)$ corresponds to this excited state. The equation can be written in the form of a one-dimensional Schrödinger equation

$$-\frac{1}{2a^p}\frac{d}{da}\left[a^p\frac{dc_n}{da}\right]+V(a)c_n=(n+\tfrac{1}{2}+\epsilon_0)c_n\ , \tag{7.1}$$

where

$$V(a)=\tfrac{1}{2}(a^2-\lambda a^4)\ . \tag{7.2}$$

At $a=0$ Eq. (7.1) will in general have two types of solutions one of which is more convergent than the other [cf. Eq. (5.12)]. The behavior for the ground state which corresponds to the functional-integral prescription could be deduced from an evaluation of the determinant in the semiclassical approximation as discussed in Sec. IV. Whatever the result of such an evalution, the solution must be purely of one type or the other in order to ensure the Hermiticity of the Hamiltonian constraint. The same requirement ensures a similar behavior for the excited-state solutions. In the following by "regular" solutions we shall mean those conforming to the boundary conditions arising from the functional-integral prescription. The exact type will be unimportant to us.

The potential $V(a)$ is a barrier of height $1/(4\lambda)$. At large a, the cosmological-constant part of the potential dominates and one has solutions which are linear combinations of the oscillating functions in (5.13). As we have already seen in the analysis of the ground state, the two possibilities correspond to a de Sitter contraction and a de Sitter expansion. With either of these asymptotic behaviors, a wave packet constructed by superimposing states of different n to produce a wave function with narrow support about some mean value of the scalar field would show this mean value increasing as one moved from large to small a.

Since each of the asymptotic behaviors in (5.13) is physically acceptable there will be solutions of (7.1) for all n. If, however, λ is small and n not too large, there are some values of n which are more important than others. These are the values which make the left-hand side of (7.1) at or close to those values of the energy associated with the metastable states (resonances) of the Schrödinger Hamiltonian on the right-hand side. To make this precise write

$$-\frac{1}{2}\frac{1}{a^p}\frac{d}{da}\left[a^p\frac{dc}{da}\right]+V(a)c=\epsilon c \ . \tag{7.3}$$

This is the zero angular momentum Schrödinger equation in $d=p+1$ dimensions for single-particle motion in the potential $V(a)$. Classically, for $\epsilon<1/(4\lambda)$ there are two classes of orbits: bound orbits with a maximum value of a and unbound orbits with a minimum value of a. Quantum mechanically there are no bound states. For discrete values of $\epsilon \ll 1/(4\lambda)$, however, there are metastable states. They lie near those values of ϵ which would be bound states if $\lambda=0$ and the barrier had infinite height. Since when $\lambda=0$ (7.3) is the zero angular momentum Schrödinger equation for a particle in a "radial" harmonic-oscillator potential in $d=p+1$ dimensions, these values are

$$\epsilon_N=2N+d/2, \quad N=0,1,2,\ldots \ . \tag{7.4}$$

For nonzero λ, if the particle has an energy near one of these values and much less than $1/(4\lambda)$ it can execute many oscillations inside the well but eventually it will tunnel out.

For the cosmological problem the classical Hamiltonian corresponding to (7.3) describes the evolution of homogeneous, isotropic, spatially closed cosmologies with radiation and a cosmological constant. The bound orbits correspond to those solutions for which the radiation density is sufficiently high that its attractive effect causes an expanding universe to recollapse before the repulsive effect of the cosmological constant becomes important. By contrast the unbound orbits correspond to de Sitter evolutions in which a collapsing universe never reaches a small enough volume for the increasing density of radiation to reverse the effect of the cosmological constant. There are thus two possible types of classical solutions. Quantum mechanically the Universe can tunnel between the two.

We can calculate the tunneling probability for small λ by using the usual barrier-penetration formulas from ordinary quantum mechanics. Let P be the probability for tunneling from inside the barrier to outside per transversal of the potential inside from minimum to maximum a. Then

$$P\approx e^{-B} \ , \tag{7.5}$$

where

$$B=2\int_{a_0}^{a_1} da[V(a)-\epsilon]^{1/2} \tag{7.6}$$

and a_0 and a_1 are the two turning points where $V(a)=\epsilon$. In the limit of $\epsilon \ll 1/(4\lambda)$ the barrier-penetration factor becomes

$$B=\frac{2}{3\lambda}=\frac{2}{3H^2} \ . \tag{7.7}$$

In magnitude this is just the total gravitational action for the Euclidean four-sphere of radius $1/H$ which is the analytic continuation of de Sitter space. This is familiar from general semiclassical results.[16]

Our own Universe corresponds to a highly excited state of the minisuperspace model. We know that the age of the Universe is about 10^{60} Planck times. The maximum expansion, assuming a radiation dominated model, is therefore at least of order ${a_{\max}}^2 \approx 10^{120}$. A wave packet describing our Universe would therefore have to be superpositions of states of definite n, with n at least $\approx {a_{\max}}^2 \approx 10^{120}$. As large as this number is, the dimensionless limit on the inverse cosmological constant is even larger. In order to have such a large radiation dominated Universe λ must be less than 10^{-120}. The probability for our Universe to tunnel quantum mechanically at the moment of its maximum expansion to a de Sitter-type phase rather than recollapse is $P\approx\exp(-10^{120})$. This is a very small number but of interest if only because it is nonzero.

VIII. TOPOLOGY

In the preceding sections we have considered the amplitudes for three-geometries with S^3 topology to occur in the ground state. The functional-integral construction of the ground-state wave function, however, permits a natural extension to calculate the amplitudes for other topologies. We shall illustrate this extension in this section with some simple examples in the semiclassical approximation.

There is no compelling reason for restricting the topologies of the Euclidean four-geometries which enter in the sum defining the ground-state wave function. Whatever one's view on this question, however, there must be a ground-state wave function for every topology of a three-geometry which can be embedded in a four-geometry which enters the sum. In the general case this will mean all possible three-topologies—disconnected as well as connected, multiply connected as well as simply connected. The general ground-state wave function will therefore have N arguments representing the possibility of N compact disconnected three-geometries. The functional-integral prescription for the ground-state wave function in the case of pure gravity would then read

$$\Psi_0[\partial M^{(1)},h_{ij}^{(1)},\ldots,\partial M^{(N)},h_{ij}^{(N)}]=\int \delta g \exp(-I_E[g]) \ , \tag{8.1}$$

where the sum is over all compact Euclidean four-geometries which have N disconnected compact boundaries $\partial M^{(i)}$ on which the induced three-metrics are $h_{ij}^{(i)}$. Since there is nothing in the sum which distinguishes one three-boundary from another the wave function must be symmetric in its arguments.

The wave function defined by (8.1) obeys a type of Wheeler-DeWitt equation in each argument but this is no longer sufficient to determine its form—in particular the correlations between the three-geometries. The functional integral is here the primary computational tool.

It is particularly simple to construct the semiclassical approximations to ground-state wave functions for those three-geometries with topologies which can be embedded in a compact Euclidean solution of the field equations. Consider for example the four-sphere. If the three-geometry has a single connected component and can be embedded in the four-sphere, then the extremal geometry at which the action is evaluated to give the semiclassical approximation is the smaller part of the four-sphere bounded by this three-geometry. The semiclassical ground-state wave function is

$$\Psi_0[h_{ij}] \approx N\Delta^{-1/2}[h_{ij}] \times \exp\left[2l^{-2}\left[\int_{\partial M} dx^3 h^{1/2} K + \int_M d^4x\, g^{1/2}\Lambda\right]\right] \tag{8.2}$$

where M is the smaller part of the four-sphere and K is the trace of the extrinsic curvature of the three-surface computed with outward-pointing normals. Since there is a large variety of topologies of three-surfaces which *can* be embedded in the four-sphere—spheres, toruses, etc.,—we can easily compute their associated wave functions. Of course, these are many interesting three-surfaces which cannot be so embedded and for which the extremal solution defining the semiclassical approximation is not part of the four-sphere. In general one would expect to find wave functions for arbitrary topologies since any three-geometry is cobordant to zero and therefore there is some compact four-manifold which has it as its boundary. The problem of finding solutions of the field equations on these four-manifolds which match the given three-geometry and are compact thus becomes an interesting one.

Similarly, the semiclassical approximation for wave functions representing N disconnected three-geometries are equally easily computed when the geometries can be embedded in the four-sphere. The extremal geometry defining the semiclassical approximation is then simply the four-sphere with the N three-geometries cut out of it. The symmetries of the solution guarantee that as far as the exponent of the semiclassical approximation is concerned, it does not matter where the three-geometries are cut out provided that they do not overlap. To give a specific example, we calculate the amplitude for two disconnected three-spheres of radius $a_{(1)}$ and $a_{(2)}$ assuming $a_{(1)} < a_{(2)} < H^{-1}$. One possible extremal geometry is two disconnected portions of a four-sphere attached to the two three-spheres. This gives a product wave function with no correlation. Another extremal geometry is the smaller half of the four-sphere bounded by the spheres of radius $a_{(2)}$ with the portion interior to a sphere of radius $a_{(1)}$ removed. This gives an additional contribution to the wave function which expresses the correlation between the spheres. The correlated part in the semiclassical approximation is

$$\Psi_0{}^c(a_{(1)}, a_{(2)}) = N\Delta^{-1/2}(a_{(1)}, a_{(2)}) \times \exp\left\{\frac{1}{3H^2}\left[-(1-H^2 a_{(2)}{}^2)^{3/2} + (1-H^2 a_{(1)}{}^2)^{3/2}\right]\right\}. \tag{8.3}$$

While the exponent is simple, the calculation of the determinant is now more complicated—it does not factor.

Equation (8.3) shows that the amplitude to have two correlated three-spheres of radius $a_{(1)} < a_{(2)} < H^{-1}$ is smaller than the amplitude to have a single three-sphere of radius $a_{(2)}$. In this crude sense topological complexity is suppressed. The amplitude for the Universe to bifurcate is of the order $\exp[-1/(3H^2)]$—a very large factor.

IX. CONCLUSIONS

The ground-state wave function for closed universes constructed by the Euclidean functional-integral prescription put forward in this paper can be said to represent a state of minimal excitation for these universes for two reasons. First, it is the natural generalization to gravity of the Euclidean functional integral for the ground-state wave function of flat-spacetime field theories. Second, when the prescription is applied to simple minisuperspace models, it yields a semiclassical wave function which corresponds to the classical solution of Einstein's equations of highest spacetime symmetry and lowest matter excitation.

The advantages of the Euclidean function-integral prescription are many but perhaps three may be singled out. First it is a complete prescription for the wave function. It implies not only the Wheeler-DeWitt equation but also the boundary conditions which determine the ground-state solution. The requirement of Hermiticity of the Wheeler-DeWitt operator extends these boundary conditions to the excited states as well.

A second advantage of this prescription for the ground-state wave function is common to all functional-integral formulations of quantum amplitudes. They permit the direct and explicit calculation of the semiclassical approximation. At the current stage of the development of quantum gravity where qualitative understanding is more important than precise numerical results, this is an important advantage. It is well illustrated by our minisuperspace model in which we were able to calculate semiclassically the probability of tunneling between a universe doomed to end in a big crunch and an eternal de Sitter expansion.

A final advantage of the Euclidean functional-integral prescription for the ground-state wave function is that it naturally generalizes to permit the calculation of amplitudes not usually considered in the canonical theory. In particular, we have been able to provide a functional-integral prescription for amplitudes for the occurrence of three-geometries with multiply connected and disconnected topologies in the ground state. In the semiclassical approximation we have been able to evaluate simple examples of such amplitudes.

The Euclidean functional-integral prescription sheds light on one of the fundamental problems of cosmology: the singularity. In the classical theory the singularity is a place where the field equations, and hence predictability, break down. The situation is improved in the quantum theory. An analogous improvement occurs in the problem of an electron orbiting a proton. In the classical theory there is a singularity and a breakdown of predictability when the electron is at the same position as the proton. However, in the quantum theory there is no singularity or breakdown. In an s-wave state, the amplitude for the electron to coincide with the proton is finite and nonzero, but the electron just carries on to the other side. Similarly, the amplitude for a zero-volume three-sphere in our minisuperspace model is finite and nonzero. One might interpret this as implying that the universe could continue through the singularity to another expansion period, although the classical concept of time would break down so that one could not say that the expansion happened after the contraction.

The ground-state wave function in the simple minisuperspace model that we have considered with a conformally invariant field does not correspond to the quantum state of the Universe that we live in because the matter wave function does not oscillate. However, it seems that this may be a consequence of using only zero rest mass fields and that the ground-state wave function for a universe with a massive scalar field would be much more complicated and might provide a model of quantum state of the observed Universe. If this were the case, one would have solved the problem of the initial boundary conditions of the Universe: the boundary conditions are that it has no boundary.[3]

ACKNOWLEDGMENTS

The authors are grateful for the hospitality of the Institute of Theoretical Physics, Santa Barbara, California, where part of this work was carried out. The research of one of us (J.H.) was supported in part by NSF Grants Nos. PHY81-07384 and PHY80-26043.

---

[1]See, e.g., R. P. Feynman, Rev. Mod. Phys. 20, 367 (1948); R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965) for discussions of quantum mechanics from this point of view.

[2]B. S. DeWitt, Phys. Rev. 160, 1113 (1967); J. A. Wheeler, in *Battelle Rencontres,* edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968).

[3]S. W. Hawking, in *Astrophysical Cosmology,* Pontificia Academiae Scientarium Scripta Varia, 48 (Pontificia Academiae Scientarium, Vatican City, 1982).

[4]For related ideas, see A. Vilenkin, Phys. Lett. 117B, 25 (1982); Phys. Rev. D 27, 2848 (1983).

[5]The connection between the canonical and functional-integral approaches to quantum gravity has been extensively discussed. See, in particular, H. Leutwyler, Phys. Rev. 134, B1155 (1964); L. Faddeev and V. Popov, Usp. Fiz. Nauk. 111, 427 (1973) [Sov. Phys. Usp. 16, 777 (1974)]; E. S. Fradkin and G. A. Vilkovisky, CERN Report No. TH-2332, 1977 (unpublished).

[6]For reviews of the canonical theory, see K. Kuchar, in *Quantum Gravity 2,* edited by C. Isham, R. Penrose, and D. W. Sciama (Clarendon, Oxford, 1981); A Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Academia Nazionale dei Lincei, Rome, 1976); K. Kuchar, in *Relativity, Astrophysics and Cosmology,* edited by W. Israel (Reidel, Dordrecht, 1973).

[7]J. M. York, Phys. Rev. Lett. 28, 1082 (1972).

[8]J. Milnor, *Morse Theory* (Princeton University Press, Princeton, New Jersey, 1962).

[9]For example, J. B. Hartle (unpublished).

[10]See, e.g., G. W. Gibbons and M. J. Perry, Nucl. Phys. B146, 90 (1978).

[11]S. W. Hawking, Commun. Math. Phys. 55, 133 (1977).

[12]B. DeWitt, Phys. Rev. 160, 1113 (1967).

[13]C. Isham and J. E. Nelson, Phys. Rev. D 10, 3226 (1974).

[14]W. E. Blyth and C. Isham, Phys. Rev. D 11, 768 (1975).

[15]G. W. Gibbons, S. W. Hawking, and M. J. Perry, Nucl. Phys. B138, 141 (1978).

[16]See, e.g., S. Coleman, Phys. Rev. D 15, 2929 (1977).