Integration contours for the no-boundary wave function of the universe

Jonathan J. Halliwell*
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

James B. Hartle
Department of Physics, University of California, Santa Barbara, California 93106
(Received 10 October 1989)

In the no-boundary proposal for the initial conditions of a closed cosmology, the wave function of the universe is the integral of exp(-action) over a contour of four-geometries and matter-field configurations on compact manifolds having only that boundary necessary to specify the arguments of the wave function. There is no satisfactory covariant Hamiltonian quantum mechanics of closed cosmologies from which the contour may be derived, as there would be for defining the ground states of asymptotically flat spacetimes. No compelling prescription, such as the conformal rotation for asymptotically flat spacetimes, has been advanced. In this paper it is argued that the contour of integration can be constrained by simple physical considerations: (1) the integral defining the wave function should converge; (2) the wave function should satisfy the constraints implementing diffeomorphism invariance; (3) classical spacetime when the universe is large should be a prediction; (4) the correct field theory in curved spacetime should be reproduced in this spacetime; (5) to the extent that wormholes make the cosmological constant dependent on initial conditions the wave function should predict its vanishing. We argue that the convergence criterion is readily satisfied by choosing a suitable *complex* contour. The constraints will be satisfied if the end points of the contour are suitably restricted. For classical spacetime to be a prediction, the contour must be dominated by one or more saddle points at which the four-metric is complex. We discuss the conditions under which such complex solutions to the Einstein equations arise and their interpretation. Because the action is double valued in the space of complex metrics, every solution of the Einstein equations corresponds to *two* saddle points: one with $\mathrm{Re}(\sqrt{g}\,)>0$, the other with $\mathrm{Re}(\sqrt{g}\,)<0$. They differ only in the sign of their action. We find that criteria (4) and (5) imply that the contour should not be dominated by a saddle point with $\mathrm{Re}(\sqrt{g}\,)<0$. This restriction may be difficult to satisfy in the path-integral forms of the "tunneling" boundary condition proposals of Linde, Vilenkin, and others. Although all of these physical considerations constrain the contour and largely determine the semiclassical predictions of the wave function, there is still remaining freedom. Until fixed by more fundamental considerations, the remaining freedom in the contour means that there are many corresponding no-boundary proposals.

I. INTRODUCTION

To apply quantum mechanics to the universe as a whole we need a quantum mechanics of cosmology and within that quantum cosmology a theory of initial conditions.[1] There are no practical predictions of any kind which do not involve a theory of initial conditions even if only very weakly, and predictions of very-large-scale observations may be testably sensitive to its details. We do not yet have a satisfactory and manageable quantum theory of gravity readily applicable to cosmology. We may imagine, however, that any theory of initial conditions within such a quantum gravity will specify, approximately, for scales larger than the Planck length, the wave function of a closed universe. This is the amplitude

$$\Psi_0[h_{ij}(\mathbf{x}),\chi(\mathbf{x}),\partial M] \tag{1.1}$$

that the universe contains a connected spacelike surface which is a three-manifold ∂M with a metric $h_{ij}(\mathbf{x})$ and matter field $\chi(\mathbf{x})$ upon it. From this amplitude one can extract most predictions of the initial conditions on accessible scales including, in particular, a prediction of classical spacetime.

In the absence of a complete theory of quantum gravity it is reasonable to try and identify principles which specify a wave function of the universe which is adequate for predictions on scales larger than the Planck scale using a low-energy effective gravitational theory. The hope would be that such principles could be generalized to a complete theory. The unique low-energy effective gravitational theory is Einstein's general relativity with cosmological constant coupled to some number of matter fields.[2] We are thus led in quantum cosmology to propose theories of initial conditions which fix Ψ_0 using general relativity coupled to matter for dynamics.

Observations show that at earlier times the universe was more nearly homogeneous, more nearly isotropic, and more nearly in thermal equilibrium. Natural candidates for the initial conditions of our universe are, therefore, those which specify qualitatively a state of minimum excitation. Were there a Hamiltonian quantum mechanics of closed cosmology, the obvious candidate for a state of minimum excitation would be the ground state.[3] However, there are serious difficulties with applying Hamiltonian quantum mechanics to closed cosmologies. For example, such a formulation would rule out the possibility of topology change which may be important for the values of the physical constants. More importantly, to form a Hamiltonian quantum mechanics of cosmology one must distinguish a time variable.[3] No natural time variable is suggested by the classical theory of general relativity because it does not prefer one set of spacelike surfaces to another. The choice of a time variable for Hamiltonian quantum mechanics therefore involves singling out a special set of surfaces not preferred by the classical theory. York[4] has argued that a natural preferred time for quantum mechanics is the trace of the extrinsic curvature, K. Ashtekar, and others, have proposed another attractive choice.[5] However, whatever the choice, it seems likely that the resulting Hamiltonian will be time dependent because the constraints of general relativity are quadratic in the momenta and have a nontrivial potential term. There will thus not be one unique ground state picked out by the theory but rather one for each time.

Feynman's sum-over-histories framework for quantum mechanics provides an alternative starting point for constructing a quantum mechanics for cosmology which can avoid these difficulties. A sum-over-histories quantum mechanics can have an equivalent Hamiltonian formulation when, among other conditions, from among the physical variables specifying a history a time can be identified which uniquely parametrizes the histories as a curve in the configuration space of the remainder of the variables. Such histories are said to "move forward in time." In cosmology the histories are four-geometries. There are no geometric variables which uniquely label a hypersurface in a general four-geometry. Certainly K is not such a variable. Therefore, from a sum-over-histories quantum cosmology which treats all four-geometries equally we do not expect to recover an equivalent Hamiltonian formulation.[6] There is no natural preferred foliation by hypersurfaces of a general compact spacetime; therefore, there is no natural notion of time; therefore, there is no natural generally covariant Hamiltonian formulation of the quantum mechanics of these spacetimes.

A natural candidate for a theory of initial conditions in sum-over-histories quantum cosmology is the "no-boundary" proposal.[7,8] The wave function of the universe is specified by a Euclidean sum over histories of the form

$$\Psi_0[h,\chi,\partial M]=\sum_M \int_{\mathcal{C}} \mathcal{D}g\, \mathcal{D}\Phi \exp(-I[g,\Phi,M]) \ . \qquad (1.2)$$

I is the Euclidean action for metric g and matter-field configuration Φ on a four-manifold M. The sum over manifolds is over a class which have the one boundary ∂M on which the arguments of the wave function are specified and *no other boundary*. The functional integral is over metrics g and matter fields Φ on M which induce h and χ on ∂M.

To make a construction such as (1.2) definite, the class of manifolds, the measure for the functional integrals, and the contour $\mathcal{C}$ over which these integrations are carried out must be given. There thus may be many "no-boundary" proposals depending on how these aspects of the construction are specified. The essence of the no-boundary proposals is topological: the specification of the topology of the manifold on which the tensor fields of integration g and Φ lie. It is this essentially topological character which gives some hope that the idea may be generalizable to many theories of quantum gravity.

Various possibilities have been discussed for the class of manifolds[9] and for the measure.[10] The object of this paper is to discuss the possibilities for the contour $\mathcal{C}$. It may be that this contour of integration is determined by a yet more fundamental quantum theory of gravity or by some synthesis between a theory of dynamics and a theory of initial conditions. However, as far as the low-energy theory is concerned, in the absence of a Hamiltonian quantum mechanics to specify the physical degrees of freedom or an alternative construction of a state of minimum excitation, the contour in the construction (1.2) is up for grabs. It is therefore reasonable to consider the class of contours *generally* and to ask which contours lead to no-boundary proposals which satisfy sensible criteria for consistency and physical predictions.

This situation should be contrasted with that in the quantum theory of asymptotically flat spacetimes. In that case, there is the conformal rotation prescription of Gibbons, Hawking, and Perry[11] for the construction of a ground-state wave function by an integration over asymptotically flat spacetimes analogous to (1.2). Briefly stated, their prescription is to decompose the metrics $g_{\mu\nu}$ that are integrated over into conformal equivalence classes represented by a metric $\tilde{g}_{\mu\nu}$ satisfying $R(\tilde{g}_{\mu\nu})=0$ and a conformal factor Ω defined by $g_{\mu\nu}=\Omega^2\tilde{g}_{\mu\nu}$. The conformal factor is written $\Omega=1+Y$ with Y vanishing asymptotically and Y is rotated to imaginary values. The resulting action is real. The integral over imaginary values of Y is explicitly convergent and the integral over conformal equivalence classes is likewise by the positive action theorem of Schoen and Yau.[12] As shown by Schleich and others,[13] this complex contour gives a ground state which coincides with that of the Hamiltonian theory for all orders of perturbation theory of Einstein gravity linearized about flat space. It also coincides with the results of Mazur and Mottola for gravity linearized about curved backgrounds.[14] These supporting calculations together with the strength of the Schoen and Yau result make the Gibbons, Hawking, and Perry conformal rotation a compelling prescription for asymptotically flat spacetimes.

Despite these reassuring results, the prescription of Gibbons, Hawking, and Perry is not without its limitations. The chief difficulty lies in the generality of the equivalence class condition $R(\tilde{g}_{\mu\nu})=0$. Not every asymptotically flat four-metric can be conformally transformed to one for which this condition holds.[15] In using this prescription, therefore, certain metrics are being missed out in the sum over geometries. However, the prescription does at least encompass a large class of interesting metrics, in that all metrics sufficiently close to flat space, or to solutions of the vacuum Einstein equations, may be conformally transformed to $R=0$. There is a further potential difficulty when gravity is coupled to matter: The conformal rotation ensuring convergence of the gravitational part of the path integral could destroy the positivity of a conformally noninvariant matter action. The matter integrals would therefore fail to converge unless contour rotations were applied to those integrations also.

The difficulties with the conformal rotation are particularly acute if the prescription is applied to the case of closed cosmologies under investigation here. There is no singly obvious candidate for the conformal equivalence class condition $R(\bar{g})=0$. For the case of compact four-manifolds without boundary, the condition $R=4\Lambda$ has been suggested,[16] but again this is problematic because not all compact four-metrics can be conformally transformed to this form.[17] More serious is the fact that there is no known analogue of the positive action theorem. For particular choices of equivalence class condition, the rotated action is complex and there is no guarantee that the sum over conformal equivalence classes will converge (see, however, Ref. 18). The contour problem for the wave function of closed cosmologies is therefore a very different problem from that in asymptotically flat spacetimes.

For these reasons, rather than seek a prescription for the contour, we are led instead to search generally for contours which satisfy sensible criteria for consistency and physical predictions. Five criteria naturally suggest themselves as reasonable restrictions on the contour defining a wave function of the universe Ψ_0. Two of them should be required of any contour for mathematical consistency.

(1) The integral defining Ψ_0 should converge.

(2) The resulting Ψ_0 should satisfy the constraints implementing diffeomorphism invariance. In a theory where Einstein's action governs the dynamics of spacetime these are the Wheeler-DeWitt equation and the momentum constraints associated with invariance under spatial diffemorphisms. These are, in some suitable operator ordering, functional differential constraints on Ψ_0. Equation (1.2) may be viewed as an integral representation for Ψ_0. Whether a given integral representation satisfies a certain differential constraint depends on the contour and in particular on its end points.

Any contour satisfying (1) and (2) defines a possible wave function of the universe. But there are further criteria which single out, from the many possible wave functions, those generally sensible for physical prediction. Three requirements are so general that they may be reasonably imposed.

(3) The wave function should imply classical spacetime on familiar scales when the universe is large. Classical spacetime is a manifest fact of the late universe but a property of only very special quantum states of the universe. Classical spacetime is predicted by a theory of initial conditions when two requirements are satisfied.

(i) There is negligible interference between alternative histories for spacetime geometry determined on scales far above the Planck length. That is, the alternative histories *decohere*.[19,20]

(ii) The histories are highly correlated according to classical laws.[21]

There are, of course, many wave functions which lead to classical correlations (e.g., Ref. 22). However, classical correlations are most commonly signaled in quantum cosmology when the wave function is well approximated by a certain type of semiclassical approximation. A typical form corresponding to classical spacetime and quantum matter is

$$\Psi_0[h_{ij},\chi,\partial M]\approx \sum_p \Delta_p[h_{ij},\partial M]\exp(iS_p[h_{ij},\partial M]) \times\psi_p[\chi,\hat{g}_p,\partial M] \; . \qquad (1.3)$$

Here, S_p is a classical action obeying the Lorentzian Hamiltonian-Jacobi equation for gravity coupled to the expectation value of matter fields. For this approximation to be valid, S_p must vary much more rapidly with h_{ij} than the prefactor Δ_p or ψ_p. The wave function (1.3) is then strongly peaked about the ensemble of classical spacetimes $\hat{g}_p$ defined by the integral curves of S_p (Ref. 21). The $\psi_p[\chi,\hat{g}_p,\partial M]$ are the associated states of the matter field in these spacetime backgrounds. They should be normalizable in the variables $\chi(\mathbf{x})$ on a spacelike surface of $\hat{g}_p$ (Ref. 23). Mathematically, a semiclassical approximation to Ψ_0 of the form (1.3) arises when, in a steepest-descent approximation to the functional integral (1.2), the dominating saddle points are complex. The integral will frequently have many saddle points, and for that reason we have included the sum over the discrete label p. The number of saddle points supplying the *dominant* contribution to the integral, however, is typically very small so p runs over only a very small number of discrete values (one or two in simple examples) (Ref. 24). For each p, the prefactor Δ_p is of the form $e^{-I_R}\Delta_{\mathrm{WKB}}$ where Δ_{WKB} is the usual WKB prefactor. Because the dominanting saddle points are generally complex, the prefactor also includes the factor e^{-I_R}, where I_R is the real part of the complex action.

There are other forms of the semiclassical approximation which can be appropriate. For example, there may be classical matter fields, in which case S_p and Δ_p would depend on them as well. Or some modes of the gravitational field could form a classical background while others form quantum fluctuations about it. The important point is that there be a division between variables behaving classically and included in the rapidly varying factor $\exp(iS_p)$ and the remaining ones which enter into the more slowly varying rest.

Not every saddle point of the action is a potential contributor to a semiclassical approximation of the no-boundary wave function. These saddle points must correspond to solutions on a *compact* manifold M with a single boundary ∂M on which real boundary data h_{ij} and χ are prescribed. If these saddle points are known, then the question of whether a no-boundary proposal predicts classical spacetime is the question of whether the contour passes through the appropriate complex saddle points to make an approximation such as (1.3) valid when the universe is large.

(4) A closely related requirement is the reproduction of familiar quantum field theory for matter when spacetime is approximately classical. This is connected with the matter fluctuations in the steepest-descent approximation about the complex saddle points. More precisely, it is the question of whether the wave functions $\psi_p[\chi,\hat{g}_p,\partial M]$ in (1.3) correctly describe quantum field theory in the spacetime $\hat{g}_p$.

(5) Recent work by Hawking,[25] Coleman,[26] Giddings and Strominger,[27] and others has shown that if nontrivial "wormhole" topologies are included in the Euclidean sum over histories (1.2) then the coupling constants of the effective low-energy theory accessible to us may depend as much on the initial conditions of the universe as they do on the form of the fundamental Lagrangian. In particular wormholes may provide a mechanism for making the cosmological constant vanish. A reasonable restriction on a no-boundary proposal is that it predict this. We measure the cosmological constant only through the evolution of the universe itself, that is, through the dynamics of the classical spacetime of the late universe. A wave function of the universe predicts a distribution of cosmological-constants when it predicts a family of classical spacetimes, each obeying Einstein's equation, but with *different* possible cosmological-constant terms Λ, that is, a semiclassical approximation of the form

$$\Psi_0[h,\chi,\partial M]=\int d\Lambda \sum_p \Delta_p[\Lambda,h_{ij}]\exp(iS_p[\Lambda,h_{ij},\partial M]) \times\psi_p(\chi,\hat{g}_{\Lambda p},\partial M) \ . \qquad (1.4)$$

Here, S_p is the classical action for gravity with cosmological constant Λ coupled to expectation values of matter fields in the state ψ_p. Roughly speaking, in the effective theory, the integral over metrics in Eq. (1.2) contributes the few terms in (1.4) labeled by different values of p, while the sum over manifolds becomes the sum over Λ. The distribution of the cosmological constant implied by $\Delta_p[\Lambda,h_{ij}]$ is thus closely connected with the semiclassical approximation and therefore with the contour of integration of a no-boundary proposal.

In the following sections, we shall discuss each of these criteria in turn and we shall argue that there are contours of integration which meet them.

II. CONSISTENCY: CONVERGENCE AND CONSTRAINTS

The Euclidean action for gravity is unbounded below on the space of real metrics on compact manifolds. The integration contour in (1.2) defining the wave function of the universe cannot, therefore, be over real metrics. The resulting integral would diverge. Indeed, were the integral over real metrics to converge, it would define a never oscillatory wave function and thus be inconsistent with one of the necessary predictions of quantum cosmology—classical spacetime when the universe is large [cf. Eq. (1.3)]. To converge, the contour must be over complex geometries. A complex contour defining a state of minimum excitation in closed cosmologies should not be surprising since, as discussed in the Introduction, such complex contours are needed to define the analogous states in the case of asymptotically flat spacetimes, in agreement with Hamiltonian quantum mechanics.

The requirement of convergence is not a strong constraint on the contours of integration. In minisuperspace models,[28–30] it can be satisfied in a variety of ways. It is also easily satisfied in lattice implementations of the sum over histories following the methods of the Regge calculus.[31,32] In a simplicial approximation to (1.2), the action becomes a function of the squared edge lengths of the simplicial net, some fixed by the boundary geometry, the n_i interior ones integrated over. A complex metric is specified by complex values of the squared edge lengths. A complex contour of integration is then specified as an n_i-dimensional contour in the $2n_i$-dimensional space of complex interior squared edge lengths. In most directions in the space of complex edge lengths, the asymptotic behavior of the action is dominated by the total volume term, this being the sum of the square roots of polynomials in the squared edge lengths. This volume term will have a real part which will become positively infinite in some directions, negatively infinite in others while in certain special directions, the real part vanishes. Such special directions occur for real edge lengths when the edges of all four-simplices become large but the simplices themselves are nearly degenerate. It is exactly such real directions along which the gravitational action on real metrics is not bounded from below. However, there are many more complex directions along which the real part of the action increases for all large edge lengths. While we offer no definite prescription, it should not be difficult to choose a contour following these directions along which the integrals converge. Indeed, in simple minisuperspace models it seems possible to use one-dimensional steepest-descent contours along which the integral converges as fast as possible.[28–32]

A related but more stringent requirement for the contour than convergence is the requirement of diffeomorphism invariance. In classical physics, invariance of the action under a group parametrized by functions of time implies constraints between canonical coordinates and canonical momenta. The invariance of the action of general relativity under four-dimensional diffeomorphisms implies the four constraints

$$\mathcal{H}_\mu(\pi^{ij}(\mathbf{x}),h_{ij}(\mathbf{x}),\pi_\chi(\mathbf{x}),\chi(\mathbf{x}))=0 \qquad (2.1)$$

between the components of the three-metric on a spacelike surface, $h_{ij}(\mathbf{x})$, and their canonically conjugate momenta $\pi^{ij}(\mathbf{x})$, together with the matter-field configurations $\chi(\mathbf{x})$ on the surface and their conjugate momenta $\pi_\chi(\mathbf{x})$.

In quantum mechanics, wave functions Ψ constructed as *invariant* sums over histories satisfy the constraints as operator identities, e.g.,

$$\mathcal{H}_\mu\left[-i\frac{\delta}{\delta h_{ij}(\mathbf{x})},h_{ij}(\mathbf{x}),-i\frac{\delta}{\delta\chi(\mathbf{x})},\chi(\mathbf{x})\right] \times\Psi[h_{ij}(\mathbf{x}),\chi(\mathbf{x}),\partial M]=0 \qquad (2.2)$$

on each disconnected part of the boundary ∂M. This has been discussed in many places, perhaps in the most detail by the authors in Ref. 33, where further references are given. We briefly summarize these results as they affect the choice of contour.

An invariantly constructed sum over histories is an integral of the general form

$$\Psi[h_{ij}(\mathbf{x}),\chi(\mathbf{x}),\partial M]=\int_{\mathcal{C}}\mathcal{D}z\,\exp\{-I[z(x)]\}\ , \tag{2.3}$$

where the integral is over some possibly extended space of variables $z(x)$, defined on the manifold M which match the arguments of the wave function on its boundary ∂M. We include in $\mathcal{D}z$ any gauge-fixing apparatus. $\mathcal{C}$ is a contour of integration. An invariantly constructed integral is one for which, under transformations for which the action changes by at most a boundary term on ∂M, the measure $\mathcal{D}z$ and class of paths $\mathcal{C}$ are invariant. Under these assumptions, Eq. (2.2) may be formally derived from (2.3).

The important point for the contour in this result is that it be invariantly defined. In particular, this means that the contour of integration over *gauge-dependent* variables cannot have gauge-variant end points. Were the end points changeable by a gauge transformation the integral would not be invariant and (2.2) would not follow. In typical implementations, such as those considered in Refs. 28–30, this means that the contour of integration for each field component will have infinite range, be closed, or have a range coinciding with a periodicity of the functions entering the sum. We shall illustrate this below.

That the requirement of invariance should constrain the end points of the contour of integration is not a surprise. Mathematically, (2.3) is an integral representation for the function Ψ. It is characteristic of integral representations that whether they satisfy differential relations such as (2.2) depends on the end points of the contour of integration.

The connection between wave functions satisfying the constraints implementing diffeomorphism invariance and invariant ranges of their defining functional integrals has been very clearly discussed by Teitelboim,[34] and since elaborated on in related contexts by the present authors.[6,35] We now recall the model of Ref. 35, to illustrate in a simple way how these issues usually arise. Reference 35 was concerned with homogeneous minisuperspace models defined on a configuration space of n coordinates $q^\alpha(t)$. In Hamiltonian form, their defining action is

$$S=\int_0^1 dt\,[p_\alpha\dot{q}^{\,\alpha}-NH(p_\alpha,q^\alpha)]\ . \tag{2.4}$$

Here, $N(t)$ is a "lapse" multiplier for the super-Hamiltonian

$$H=\tfrac{1}{2}f^{\alpha\beta}(q)p_\alpha p_\beta+U(q)\ , \tag{2.5}$$

where $f^{\alpha\beta}(q)$ is a metric on minisuperspace and $U(q)$ is a potential. The action (2.4) is invariant under reparametrization of the proper time

$$\delta q^\alpha=\epsilon(t)\{q^\alpha,H\},\quad \delta p_\alpha=\epsilon(t)\{p_\alpha,H\},\quad \delta N=\dot{\epsilon}(t) \tag{2.6}$$

provided $\epsilon(0)=0=\epsilon(1)$. Classically, there is therefore a constraint [found immediately by varying (2.4) with respect to N] which is

$$H=0\ . \tag{2.7}$$

Consider the wave function constructed from a path integral of the form

$$\Psi(q^\alpha)=\int_{\mathcal{C}}\mathcal{D}p_\alpha\mathcal{D}q^\alpha\mathcal{D}N\,\Delta[p,q,N]\delta[\dot{N}-\chi(p,q,N)]\times\exp(iS[p,q,N])\ , \tag{2.8}$$

where $\chi(p,q,N)$ is an arbitrary function entering the parametrization-fixing condition in the argument of the delta function. Δ is the associated Faddeev-Popov determinant. The action is invariant under the reparametrization (2.6). The measure combined with the gauge-fixing delta function is likewise invariant. Equation (2.8) will therefore define an *invariant* path-integral construction provided the range of integration defines an invariant class of paths to integrate over. To see the consequences of different choices of range, specialize to the gauge $\chi=0$, i.e., $\dot{N}=0$. Δ may then be shown to equal a constant. The integral over the p_α,q^α alone defines a wave function $\Psi(q^\alpha,N)$, which satisfies the familiar Schrödinger equation

$$i\frac{\partial\Psi}{\partial N}=H\Psi\ . \tag{2.9}$$

The remaining integral in (2.8) is over the constant value of N:

$$\Psi(q^\alpha)=\int_{\mathcal{C}}dN\,\Psi(q^\alpha,N)\ . \tag{2.10}$$

This is effectively an integral over the time. Evidently from (2.9), $\Psi(q^\alpha)$ defined by (2.10) will satisfy the Wheeler-DeWitt equation

$$H\Psi(q^\alpha)=0 \tag{2.11}$$

if the range of N is chosen to be from $-\infty$ to $+\infty$, or if the N contour is closed. For ranges with finite end points it will not. However, from (2.6) it is clear that reparametrization transformations amount to translation of N. No range with finite end points is left invariant by (2.6). The invariant range of integration is the real line.

III. COMPLEX SADDLE POINTS AND THEIR INTERPRETATION

Our third criterion is that the wave function should predict classical spacetime when the universe is large. As discussed in Sec. I, classical spacetime is predicted when, among other conditions, the wave function is oscillatory, of the form (1.3), where S_p is a (possibly approximate) solution to the Lorentzian Hamilton-Jacobi equation. Mathematically, an expression of this form can emerge as a steepest-descent approximation to the path integral (1.2) only if the dominating saddle-points of the integral over metrics are Lorentzian, or more generally, complex. This section is therefore concerned with a discussion of complex solutions to the Einstein equations—the conditions under which they exist, how to find them, and their interpretation.

In Sec. III A we discuss complex metrics, the complex Einstein equations, and complex diffeomorphisms. In Sec. III B we discuss the interpretation of complex solutions and the role they play in the prediction of classical spacetime. In Sec. III C we give some simple examples of complex solutions, obtained by joining together real Euclidean and real Lorentzian solutions. In Sec. III D we go to the restricted context of minisuperspace. There, it is possible to discuss complex solutions in considerable detail, although at the expense of generality. In Sec. III E we speculate on the extension of the considerations of Sec. III D to the general case.

A. The complex Einstein equations and their solutions

First some definitions. We consider a fixed (real) four-manifold M with a single connected boundary ∂M. A complex metric on M is an invertible second-rank complex-valued tensor field. Complex matter field configurations are similarly defined complex-valued tensor fields. We seek the extrema of the action for gravity coupled to matter fields. This has the form

$$l^2 I[g,\phi,M] = -\int_M d^4x\sqrt{g}\,(R-2\Lambda) - 2\int_{\partial M} d^3x\sqrt{h}\,K + l^2\int_M d^4x\sqrt{g}\,\mathcal{L}_{\mathrm{m}} \,. \tag{3.1}$$

Here, $l=(16\pi G)^{1/2}$ is the Planck length, K is the trace of the extrinsic curvature of ∂M, and $\mathcal{L}_{\mathrm{m}}$ is the matter Lagrangian. The saddle points which may potentially contribute to a steepest-descent evaluation of (1.2) are the extrema of this action with the *real* values of $h_{ij}(\mathbf{x})$ and $\chi(\mathbf{x})$ fixed on the boundary which are the arguments of Ψ_0. The extrema will thus satisfy the Einstein equation on M,

$$R_{\alpha\beta} - \tfrac{1}{2}Rg_{\alpha\beta} + \Lambda g_{\alpha\beta} = \tfrac{1}{2}l^2 T_{\alpha\beta} \,, \tag{3.2}$$

together with the appropriate matter-field equations. In general these solutions will be complex. The solutions relevant for a steepest-descent evaluation of a no-boundary integral are those which have finite action. As we shall see below, these need not necessarily even be continuous.

Two solutions to (3.2) are physically equivalent if they are connected by a real diffeomorphism. Points on the manifold can be labeled by overlapping charts of real-valued coordinates. Real coordinate transformations therefore connect physically equivalent metrics. In a sum over geometries these equivalence classes should be counted only once, and a standard gauge-fixing machinery is available to effect this.

If the metric components are analytic functions of the coordinates,[36] for some family of coordinate charts, themselves connected by analytic transition functions, then new solutions can be generated from old by complex coordinate transformations. Explicitly, if $g_{\alpha\beta}(z^\gamma)$ is an analytic function of z^γ and a solution of (3.2) for real z^γ then, by analytic continuation, it is a solution for complex z^γ as well. Thus, given a complex-valued function $z^\gamma(x^\beta)$, the metric

$$g'_{\alpha\beta}(x) = \left[g_{\gamma\sigma}(z)\frac{\partial z^\gamma}{\partial x^\alpha}\frac{\partial z^\sigma}{\partial x^\beta} \right]_{z=z(x)} \tag{3.3}$$

will solve (3.2) if $g_{\alpha\beta}$ does.

With suitable restrictions on the transformation $z^\gamma(x^\beta)$ the action of the new metric will be the same as that of the old. In the space of four complex coordinates the action is defined by the integral over some purely real region. If the transformation $z^\gamma(x^\beta)$ leaves the boundary of this region fixed, then it may be thought of as a distortion of the integration contour defining the action integral to complex values of the coordinates. If this distortion encounters no singularities of the integrand of the action integral, then the value of the action will be unchanged. In particular all semiclassical predictions, which follow from the value of the action of the solutions, will be unaffected. Thus, as far as their semiclassical predictions are concerned, two extrema connected by a complex diffeomorphism of this type are physically equivalent. The contour of integration over metrics should therefore be chosen to pass through at most one of the extrema connected by complex diffeomorphisms which do not affect the value of the action. This means that one can use complex diffeomorphisms to simplify the form of the metric in searching for suitable solutions.

The Lagrangian density for the action (3.1) is an analytic function of the metric except where the metric becomes singular. In particular, there is a branch point at $g=0$ arising from the factor of $\sqrt{g}$. The action is therefore double valued on the space of complex metrics. Carrying $g_{\alpha\beta}$ once around the branch point at $g=0$ changes the sign of I. Each solution to (3.2) therefore corresponds to two different extrema on different sheets of I whose actions differ only in sign. This doubling of extrema was first noticed in a Regge calculus model[31] and subsequently in the minisuperspace model of Ref. 28. We shall have more to say about it in Secs. IV and V.

B. Interpretation

To predict classical spacetime in the late universe the contour of integration over geometries must be dominated by extrema of the action which represent classical spacetime. Which are they? For general matter we can only expect solutions to the Einstein equations which are complex. In particular the action of the solutions is complex:

$$I[h_{ij},\phi] = I_R[h_{ij},\chi] - iS[h_{ij},\chi] \,. \tag{3.4}$$

The whole action will satisfy the Euclidean Hamilton-Jacobi equation, which takes the form

$$-(\nabla I)^2 + U[h_{ij},\chi] = 0 \,, \tag{3.5}$$

where U is the Wheeler-DeWitt potential. The real and imaginary parts of the action will generally not satisfy it by themselves. In general, therefore, S does not define an ensemble of classical geometries and matter fields. Consider, however, the real and imaginary parts of Eq. (3.5):

$$-(\nabla I_R)^2 + (\nabla S)^2 + U[h_{ij},\chi] = 0 \,, \tag{3.6}$$

$$(\nabla S)\cdot(\nabla I_R)=0 \ . \tag{3.7}$$

From Eq. (3.6), one may see that if the gradient of S becomes much larger than the gradient of I_R, then S will be an *approximate* solution to the Lorentzian Hamilton-Jacobi equation. In this case it *does* define an ensemble of classical geometries and matter fields.

If S defines an ensemble of classical trajectories, then the real part of the action is also important. In a steepest-descent approximation leading to an expression such as (1.3) the real part of the action contributes the term e^{-I_R} to the prefactor Δ. Since it is exponential it can well be the most significant contribution to the prefactor. If this is the case, the measure provided by the prefactor on the ensemble of classical trajectories is, to leading order, of the form e^{-2I_R}. It is preserved along the classical trajectories by virtue of (3.7). In situations where the real part of the action becomes independent of the overall scale of the universe, the measure will still determine the relative weight in the semiclassical approximation of classical histories with that scale. For example, for anisotropic models it will determine the relative weight of different anisotropies.[37] In the Hawking scalar field model,[8] it would affect the relative weight of universes with different initial values of ϕ, and thus provides a measure of the likelihood of inflation.[38] In models in which the cosmological constant becomes a variable, it will determine the relative weight of universes with different values of the cosmological constant, as will be discussed in Sec. V.

Beyond the simple considerations we have discussed, very little is known about the complex solutions of the Einstein equations on a general manifold. In the next few subsections, we therefore turn to a few simple examples and special cases.

C. Simple solutions

To begin, we shall consider the simplest manifold contributing to the no-boundary wave function. This is the four-ball B^4 which may also be described as a part of the four-sphere with three-sphere boundary. We shall consider also the simplest model of matter, which is vacuum with cosmological constant.

We first consider real solutions. There are Euclidean signature $(+,+,+,+)$ real solutions on B^4. Consider, for example, the vacuum Einstein equation with cosmological constant Λ. For a round three-sphere metric of radius a on the boundary the solution is a round four-sphere metric on the interior of B^4 with radius $(3/\Lambda)^{1/2}$. In fact, there are two solutions corresponding to filling in the three-sphere with more than a hemisphere of the four-sphere or less. This example shows that there is not necessarily a unique solution for the three-metric fixed on the boundary.

There is a largest three-sphere, that of radius $(3/\Lambda)^{1/2}$, for which solutions of this type exist; thus, we do not expect Euclidean extrema for large nearly symmetric three-geometries. However, since arbitrarily large, irregular three-geometries can divide the round four-sphere, there is no upper limit to the size of these geometries which bound a purely Euclidean solution. There are no solutions of purely Lorentzian signature $(-,+,+,+)$ on B^4 which induce a spacelike metric on its boundary, because B^4 cannot carry a nonsingular timelike vector field pointing outward at the boundary.[39]

The absence of global Lorentzian solutions means that the extrema which predict classical spacetime must be complex, their action having both real and imaginary parts. The simplest class of such solutions can be constructed by joining together purely Lorentzian solutions with purely Euclidean ones giving a purely real metric with a discontinuous change in signature. As long as such solutions have finite action they are acceptable candidates for extrema of the sum over geometries. The action will be finite provided the standard junction conditions are satisfied across the three-surface which separates the Euclidean from the Lorentzian part of the solutions.[40] The junction conditions are that the induced three-metric h_{ij} and the extrinsic curvature K_{ij} should match across the surface. Matching the metrics shows that the surface must be spacelike in the Lorentzian geometry. The metric in the neighborhood of the surface can then be written in a standard 3+1 decomposition with, say, $\tau=0$ labeling the dividing surface as

$$ds^2=N^2d\tau^2+h_{ij}(dx^i+N^id\tau)(dx^j+N^jd\tau) \ . \tag{3.8}$$

The signature is controlled by N. Real metrics with Euclidean signature correspond to real N. Real metrics with Lorentzian signature correspond to purely imaginary N. For both cases N^i and h_{ij} are real. Conventionally, the extrinsic curvature is defined to be real in both Euclidean and Lorentzian spacetimes. Then, however, the Einstein equations take a different form for Euclidean and Lorentzian metrics when expressed in terms of h_{ij}, K_{ij} and their derivatives. As a consequence, the junction conditions between Euclidean and Lorentzian metrics would not be expressed as continuity of K_{ij}. Here, we use the definition

$$K_{ij}=\frac{1}{2N}\left[\frac{\partial h_{ij}}{\partial\tau}+D_{(i}N_{j)}\right] , \tag{3.9}$$

where D_i is the derivative in the three-surface, for *both* Euclidean and Lorentzian spacetimes. Then, six of the Einstein equations have the form

$$\frac{\partial}{\partial\tau}(K^i_j-K\delta^i_j)+F^i_j(h_{ij},{}^3R_{ij},K_{ij})=0 , \tag{3.10}$$

where F^i_j is a tensor function of $h_{ij},{}^3R_{ij},K_{ij}$, and is the same for both Euclidean and Lorentzian regions. From this one can deduce that a correct junction condition is the continuity of K_{ij} as defined by (3.9).

With the definition (3.9), K_{ij} is purely real for spacelike surfaces in Euclidean spacetimes and purely imaginary in Lorentzian spacetimes. It is, therefore, clear that a real Euclidean and real Lorentzian metric can be matched only across a spacelike surface where

$$K_{ij}=0 \ . \tag{3.11}$$

Across such a surface, the three-geometry also must be continuous. In the presence of matter, there will be other junction conditions for the matter fields. Together, these conditions are very restrictive, as we shall see.

If there is a solution with $K_{ij}=0$ on ∂B^4 then two such solutions could be joined together to give a nonsingular Einstein metric on the whole of S^4. The simplest example is the round metric on S^4 and this does have a $K_{ij}=0$ surface at its "equator." It is not known whether this example is unique. However, if it is, then the only metric on B^4 with a $K_{ij}=0$ surface is the round metric in which ∂M is the three-sphere "equator" itself with a round metric.

The Lorentzian solution which matches the Euclidean one is the evolution of the initial data $K_{ij}=0$ and the round metric on the three-sphere. This, of course, is de Sitter space with the matching surface being the de Sitter "throat" at the moment of time symmetry. The action of this complex solution is easily evaluated and is

$$I=-\frac{1}{3\lambda}[1\pm i(\lambda a^2-1)^{3/2}] , \tag{3.12}$$

where $\lambda=2\Lambda/9\pi G$ and $a^2 d\Omega_3^2$ is the round metric on ∂M (Ref. 41).

Thus, if the round metric on S^4 is unique, only for very special values of the boundary metric, namely, the most symmetric possibility, will there be any real extremum on B^4 with a real Lorentzian part and a real Euclidean part. For *general* boundary data we expect the extrema to be complex metrics which cannot be transformed into metrics consisting of just a pure real part and a pure Euclidean part. As will be discussed towards the end of this section, a general characterization of the conditions under which complex solutions arise is not available. To make further progress, therefore, we need to restrict attention to a manageable class of reasonably simple metrics. A sufficiently simple class with a history of utility in quantum cosmology is the class of models known as mini- superspace models, and it is these that we study next.

D. Minisuperspace models

In minisuperspace models, one severely restricts the four-metric in (3.8) so that the shift N^i, is zero, the lapse is homogeneous, $N=N(\tau)$ and the three-metric h_{ij} is restricted in such a way that it is described by a finite number of functions of $\tau, q^\alpha(\tau)$, say, where $\alpha=1,2,\ldots,n$. The no-boundary wave function $\Psi(q^\alpha)$ is given by a minisuperspace path integral of the form (2.8). This path integral is closely related to the more general minisuperspace propagator between fixed three-geometries[35] which, in the gauge $\dot N=0$, is of the form

$$G(q''^\alpha|q'^\alpha)=\int dN\int \mathcal{D}q\,\exp(-I[q^\alpha(\tau),N]) . \tag{3.13}$$

Here, I is the reduced version of the Einstein-Hilbert action and is of the form

$$I=\int_{\tau'}^{\tau''} d\tau\, N\left[\frac{1}{2N^2}f_{\alpha\beta}\dot q^\alpha \dot q^\beta+U(q)\right] . \tag{3.14}$$

$f_{\alpha\beta}$ is the DeWitt metric on minisuperspace and has signature $(-+++\cdots)$. $U(q)$ is potential which may take positive or negative values. The path integral (3.13) is over paths $q^\alpha(\tau)$ satisfying the boundary conditions

$$q^\alpha(\tau'')=q''^\alpha,\quad q^\alpha(\tau')=q'^\alpha , \tag{3.15}$$

where q''^α and q'^α are real. We are allowing the four-metric to be complex, so N and $q^\alpha(\tau)$ may be complex, subject to the restriction (3.15), but τ is strictly real.

The no-boundary wave function is obtained by summing over paths corresponding to four-geometries which close off the bounding three-surface. Closure of the four-geometry is achieved by imposing certain conditions at the initial point of the paths $q^\alpha(\tau)$. These conditions involve not just the q'^α's as in (3.15), but are generally conditions on some combination of the q'^α's and their conjugate momenta, p'_α (Ref. 42). The no-boundary wave function is thus obtained from (3.13) by setting some of the q'^α's to certain values, and then performing some kind of Fourier transform of (3.13) in the remaining q'^α's, with the corresponding momenta set to values determined by the closure condition. For convenience we will in what follows concentrate on the propagator between fixed three-geometries (3.13), but our conclusions will also hold for different choices of boundary conditions, such as those implied by the no-boundary proposal.

The saddle points of (3.13) are the configurations $(q^\alpha(\tau),N)$ for which $\delta I/\delta q=0$ and $\partial I/\partial N=0$ subject to the boundary conditions (3.15), i.e., those for which

$$\frac{\ddot q^\alpha}{N^2}+\frac{1}{N^2}\Gamma^\alpha_{\beta\gamma}\dot q^\beta\dot q^\gamma-f^{\alpha\beta}\frac{\partial U}{\partial q^\beta}=0 \tag{3.16}$$

and

$$\int_{\tau'}^{\tau''} d\tau\left[\frac{1}{2N^2}f_{\alpha\beta}\dot q^\alpha\dot q^\beta-U(q)\right]=0 . \tag{3.17}$$

Because (3.13) is an ordinary, rather than functional, integral over N, (3.17) is not the usual Hamiltonian constraint. However, the integrand in (3.17) is constant, by virtue of (3.16), and from (3.17) the usual constraint then follows:

$$\frac{1}{2N^2}f_{\alpha\beta}\dot q^\alpha\dot q^\beta-U(q)=0 . \tag{3.18}$$

We are interested in the solutions to (3.16) and (3.17) subject to the boundary conditions (3.15). A useful strategy for obtaining and studying the complex solutions is as follows: First, one solves the second-order equations (3.16), subject to the boundary conditions (3.15). This yields a solution $\bar q^\alpha(\tau)$, say, depending explicitly on the as-yet undetermined constant N and on the end-point values of q^α and τ. It seems reasonable to assume that this solution always exists and is real when N is real, although it is not necessarily unique (see Ref. 24, for example). The second step involves inserting this solution into the constraint equation (3.18) [or (3.17)]. This gives a purely *algebraic* equation for the constant lapse N, which may be solved to give N as a function of the end-point values of q^α and τ. The solution for N may be purely real, implying that the corresponding four-metric is real Euclidean, but will generally be complex. Whether it is real or complex will depend on the end-point values of

q^α. If it is complex, it will in turn imply that the q^α's (i.e., the three-metric components) are complex, so the four-metric will be fully complex.

Before going any further it is perhaps useful to compare what we are doing here with tunneling calculations in nonrelativistic quantum mechanics. There, one also looks for solutions to a set of Euclidean equations of the form (3.16), (3.18) (there is a constraint equation because one considers fixed-energy solutions). A crucial difference, however, is that in ordinary quantum mechanics the metric $f_{\alpha\beta}$ is positive definite, whereas here, it is indefinite. With a positive-definite metric there are well-characterized conditions under which the solution connecting two values of q^α is real or complex: If the two values of q^α are in the region $U<0$, they are connected by a real Euclidean solution; if both are in $U>0$, they are connected by a purely imaginary Euclidean solution; if one is in $U<0$ and one in $U>0$, the solution is complex. With an indefinite metric, however, things are more complicated. Both real Euclidean and real Lorentzian trajectories can exist in both $U>0$ and $U<0$, and it may be the case that any two points in configuration space can be connected by a real Euclidean, a real Lorentzian, or a complex solution. These possibilities will be made manifest by the appearance of a number of solutions to the constraint equation for N, some real, some imaginary, some complex. Even in the restricted context of minisuperspace, it is not immediately clear whether there exist general conditions under which the solutions are real or complex, and it would be of interest to study this. Some simple models in which these conditions can be established are considered in Ref. 30.

The complex saddle points which dominate the path integral will generally not have any immediate interpretation in terms of real Euclidean or real Lorentzian metrics. For the purpose of calculating the action of these saddle points, however, there is a certain restricted class of solutions which may, in a sense to be explained, be regarded as combinations of real Euclidean and real Lorentzian solutions. Suppose we evaluate the action (3.14) of a complex solution $q^\alpha(\tau)$ with a complex value of N. The action (3.14) is the integral over a strictly real quantity τ of a complex-valued integrand. However, the solution $q^\alpha(\tau)$ and the action have the property that N and τ always occur in the combination $N\tau$. Introduce, therefore, a new *complex* integration variable $T=N(\tau-\tau')$. Equation (3.14) is then entirely equivalent to the complex integral

$$I=\int dT\left[\frac{1}{2}f_{\alpha\beta}\frac{dq^\alpha}{dT}\frac{dq^\beta}{dT}+U(q)\right], \qquad (3.19)$$

where the T contour is taken to be the straight line running from 0 to $\tilde{T}$, in the complex T plane, with $\tilde{T}=N(\tau''-\tau')$. The point now, is that this contour may be distorted into one running along the real axis from 0 to $\mathrm{Re}(\tilde{T})$ and then from there parallel to the imaginary axis up to $\tilde{T}$ (assuming there are no intervening poles). The integrand along the first section is purely real, corresponding to a real Euclidean four-metric. On the second section, under suitable conditions, one obtains a purely imaginary result corresponding to a real Lorentzian metric.[43] Let us determine what these conditions are.

Let $\tilde{T}=\tilde{T}_1+i\tilde{T}_2$, where $\tilde{T}_1$ and $\tilde{T}_2$ are real. On the second section of the contour, one may write $T=\tilde{T}_1+it\tilde{T}_2$, where $0\le t\le 1$. Then, since $dT=i\tilde{T}_2dt$, the action will be purely imaginary, and the corresponding four-metric real Lorentzian, if $q^\alpha(T)$ is real. Expand $q^\alpha(T)$ in t about the point $t=0$. One obtains

$$q^\alpha(\tilde{T}_1+it\tilde{T}_2)=q^\alpha(\tilde{T}_1)+it\tilde{T}_2\dot{q}^\alpha(\tilde{T}_1)+\tfrac{1}{2}(it\tilde{T}_2)^2\ddot{q}^\alpha(\tilde{T}_1)+\cdots . \qquad (3.20)$$

One may now see that the imaginary part of $q^\alpha(T)$ is contained entirely in the odd derivative terms, which will not in general be zero. Suppose, however, that $\tilde{T}_1$ is such that $\dot{q}^\alpha(\tilde{T}_1)=0$. Since q^α satisfies the second-order differential equation (3.16), which is unchanged by $\tau\to-\tau$, it follows that all the other odd derivatives of q^α also vanish at $\tilde{T}_1$, and $q^\alpha(T)$ is then real. The integrand in (3.19) along the second section of the contour will therefore be real, and the action will be purely imaginary. The condition under which the complex solution may be regarded as a combination of real Euclidean and Lorentzian metrics, therefore, is that $\dot{q}^\alpha(\tilde{T}_1)=0$. As mentioned at the beginning of this section, this distortion of the contour is essentially the same as a complex diffeomorphism transforming a complex metric into a discontinuous combination of real Euclidean and Lorentzian metrics. The discontinuity comes from the right-angle turn of the contour, at which the lapse effectively changes from purely real to purely imaginary values. Moreover, in terms of the three-metric, the condition $\dot{q}^\alpha(\tilde{T}_1)=0$ is just the condition that the extrinsic curvature K_{ij} vanishes at $T=\tilde{T}_1$. We have therefore arrived, by a rather different route, at the junction conditions (3.11).

A number of qualifying remarks should be made at this stage. First, we have assumed that the action of the solution may be separated into real and imaginary parts by distorting the contour into just two pieces, the first running up the real axis and the second running parallel to the imaginary axis. This is appropriate for models such as the de Sitter model, mentioned above and discussed in more detail below, in which the solution connecting a small three-sphere to a large three-sphere is a section of four-sphere matched onto a section of de Sitter space across a single transitional surface. More generally, one might expect to have to distort the contour into a sequence of more than just two straight segments. Each segment would run parallel to or along the real or imaginary axis, with $\dot{q}^\alpha=0$ at each right-angle bend in the contour. Such examples actually occur. An example slightly more complicated than the de Sitter model is the axionic wormhole solution with positive cosmological constant.[44] One can in a smooth way join a section of real Euclidean wormhole solution onto a small Lorentzian Tolman-like universe at one end, and to a large de Sitter-like universe at the other end. In evaluating the action of a solution connecting a very small three-sphere to a very large one, the T contour would run first along the imaginary axis, then parallel to the real axis and finally parallel to the imaginary axis.

The second point to be made concerns the condition $\dot{q}^{\alpha}(\tilde{T}_1)=0$. The condition that it be possible to distort the contour into a series of contours parallel to the real or imaginary axis with $\dot{q}^{\alpha}=0$ at the right-angle bends is actually rather stringent. Demanding that *all* of the $\dot{q}^{\alpha}$'s vanish together is a set of n conditions, and it is most unlikely that they can be satisfied in general, given that one has at best only the freedom to choose the point in the complex plane at which to take the right-angle bends. That is, it seems unlikely that one can in general find a contour along which the four-metric is either real Euclidean or real Lorentzian metrics. It may be possible to keep *some* of the q^{α}'s real along the contour, but not all of them. This strongly suggests that it is generally not possible to find a complex diffeomorphism which transforms a given complex metric into a combination of real Euclidean and Lorentzian. In the general case, that this is so may be seen from function counting. A complex four-metric has ten independent complex components. However, a complex diffeomorphism has available only eight real functions to eliminate the ten imaginary parts of the complex four-metric.

Note that this is in contrast with the situation in ordinary quantum mechanics mentioned above. There, because the metric is positive definite, *all* the $\dot{q}^{\alpha}$'s vanish together when $U=0$. This means that a complex solution can always be transformed into a combination of real Euclidean and Lorentzian solutions, and the complex methods introduced here reproduce the familiar quantum-mechanical results. With an indefinite metric, however, if $U=0$ it does not necessarily follow that all the $\dot{q}^{\alpha}$'s vanish. All that one can say is that either $\dot{q}^{\alpha}=0$ for all α, or the trajectories become momentarily null. As we have just argued, it will generally be the latter possibility that is realized.

Wormhole solutions again provide an illustrative example of the above point. In considering the propagation between fixed three-geometries and matter fields, wormhole solutions arise if the Lorentzian momentum conjugate to ϕ is fixed on the boundary.[45] In these solutions, complex diffeomorphisms may be used to reduce the scale factor to real Euclidean or real Lorentzian form, with the transition between the two at the ends of the wormholes, at which $\dot{a}=0$. This cannot be achieved for ϕ, however. The solution for ϕ is such that $\dot{\phi}$ never vanishes, and thus the condition that $\dot{\phi}/N$ is continuous cannot be satisfied for real ϕ. So although the solution for a is real Euclidean or real Lorentzian, the solution for ϕ is actually complex, and can in no way correspond to a combination of real Euclidean and real Lorentzian solutions.

To illustrate this discussion of complex solutions, let us consider a specific example, namely, the de Sitter minisuperspace model.[28] This model admits de Sitter space and the four-sphere as solutions, and these have already been discussed above, but it is of interest to discuss them in this model in which they arise in a somewhat different fashion. The four-metric is taken to be

$$ds^2=\frac{2G}{3\pi}\left[\frac{N^2}{q(\tau)}d\tau^2+q(\tau)d\Omega_3^2\right] . \qquad (3.21)$$

This unconventional parametrization of the four-metric simplifies the algebra. The Euclidean Einstein-Hilbert action with cosmological constant $\Lambda=9\pi G\lambda/2$ is

$$I=\frac{1}{2}\int_{\tau'}^{\tau''}d\tau N\left[-\frac{\dot{q}^2}{4N^2}+\lambda q-1\right] \qquad (3.22)$$

and the field equation and constraint are

$$\frac{\ddot{q}}{N^2}=-2\lambda , \qquad (3.23)$$

$$\frac{\dot{q}^2}{4N^2}+\lambda q-1=0 . \qquad (3.24)$$

Because the model is one dimensional, the minisuperspace "metric" is negative definite, and thus this model does not capture an important feature of the full theory: namely, the indefiniteness of the DeWitt metric. However, we are mainly interested in seeing how complex solutions arise and how they may be interpreted in terms of real Euclidean and Lorentzian solutions, and this we shall be able to see in this simple model.

Following the general stratergy described above, we first solve the field equation (3.23) subject to the boundary conditions (3.15). The solution is

$$\bar{q}(\tau)=-\lambda N^2(\tau-\tau')^2+\left[\frac{q''-q'}{N(\tau''-\tau')}+\lambda N(\tau''-\tau')\right]\times N(\tau-\tau')+q' . \qquad (3.25)$$

It depends on N explicitly and is real for real N. Inserting this solution into the constraint equation (3.24), one obtains the following algebraic equation for the lapse function:

$$\lambda^2\tilde{T}^4+2(\lambda q''+\lambda q'-2)\tilde{T}^2+(q''-q')^2=0 , \qquad (3.26)$$

where $\tilde{T}=N(\tau''-\tau')$ as above. We will apply the no-boundary proposal, which, for this model, involves choosing $q'=0$, so that the geometries represented by (3.21) close off at $\tau=\tau'$. Note that for every solution $\tilde{T}$ to (3.26) with $\mathrm{Re}(\tilde{T})>0$, $-\tilde{T}$ is a second solution. The metric is the same for these two solutions but they correspond to two different extrema of the action with opposite signs for $\sqrt{g}=Nq^2$ and hence opposite signs for the action at the extremum. This is the inevitable doubling of the saddle points discussed in Sec. III A. For the moment, we will restrict attention to the solutions with $\mathrm{Re}(\tilde{T})>0$. Whether the solutions to (3.26) are real or complex depends on the value of q''. In particular, when $\lambda q''<1$, the solutions are real:

$$\tilde{T}=\frac{1}{\lambda}[1\pm(1-\lambda q'')^{1/2}] . \qquad (3.27)$$

The corresponding four-metric is the real Euclidean metric on the four-sphere. The plus/minus sign corresponds to the three-sphere boundary being closed off by more than/less than half of a four-sphere. When $\lambda q''>1$, the solutions are complex:

$$\tilde{T}=\frac{1}{\lambda}[1\pm i(\lambda q''-1)^{1/2}] . \qquad (3.28)$$

With these values of N, the corresponding $q(\tau)$ may be written

$$q(\tau)=-\lambda N^2(\tau-\tau')^2+2N(\tau-\tau') \ . \tag{3.29}$$

We therefore have a complex four-metric (3.21) with N given by (3.28) and $q(\tau)$ given by (3.29). It is complex because N is complex.

Suppose one now attempts to evaluate the action of this complex configuration. In terms of the complex integration variable $T=N(\tau-\tau')$ introduced above, the four-metric (3.21) of this solution becomes

$$ds^2=\frac{2G}{3\pi}\left[\frac{dT^2}{-\lambda T^2+2T}+(-\lambda T^2+2T)d\Omega_3^2\right] \tag{3.30}$$

from which it is easy to see that $\dot{q}(T)=0$ at $T=1/\lambda$; i.e., at $T=\mathrm{Re}(\tilde{T})$. It follows from the above discussion that in evaluating the action of this complex solution, the T contour may be distorted into one running along the real axis from 0 to $1/\lambda$, and from there parallel to the imaginary axis to the complex value of T given by (3.28). The action is purely real along the first section of contour, and the metric is that of the four-sphere. One obtains $I=-1/3\lambda$, the correct result for the action of half a four-sphere. The action is purely imaginary along the second section, and the metric is that of real Lorentzian de Sitter space. One obtains $I=\pm(i/3\lambda)(\lambda q''-1)^{3/2}$, which is the correct result for the Euclidean action of a section of Lorentzian de Sitter space. The total action is complex and is given by (3.12).

The method described in this section, let us call it the "lapse method," while it has conceptual advantages, would be rather difficult to apply in practice. An alternative, perhaps more practical search for complex solutions might proceed as follows.

We are searching for the extrema of the action which imply classical spacetime on a region of a compact manifold M with a fixed large three-geometry on its single boundary ∂M. Such solutions must approach a Lorentzian solution in the neighborhood of the boundary. As there are no purely Lorentzian solutions on M, we investigate complex solutions which may be continuous or discontinuous as long as the action is finite. These complex solutions will consist of a number of different coordinate charts. Complex diffeomorphisms can be used to simplify the form of the metric in each chart to obtain single representatives of families of extrema of equal action. One can then attempt to solve the resulting differential equations and match the solutions between the individual charts to obtain a solution on the whole manifold.

To make this procedure more definite let us consider Hawking's scalar field model.[8] Matter is modeled by a scalar field with mass m minimally coupled to curvature. The cosmological constant is assumed to vanish. The manifold is B^4. The geometry is assumed to have three-sphere symmetry

$$ds^2=N^2(\tau)d\tau^2+a^2(\tau)d\Omega_3^2 \tag{3.31}$$

and the scalar field as well, $\phi=\phi(\tau)$. Regular geometries are locally flat near the center of the ball at $\tau=0$:

$$N\sim 1, \quad a(\tau)\sim\tau \tag{3.32}$$

with a suitable choice of scale for τ. To imply classical spacetime, the solution must be Lorentzian

$$N\sim\pm i, \quad a\sim a_0 \tag{3.33}$$

near the boundary with a suitable choice of scale for τ, where a_0 is the prescribed radius of the boundary three-geometry, say at $\tau=\tau_0$.

Complex transformations of the variable τ, which leave $\tau=0$ and $\tau=\tau_0$ fixed, connect functional forms for $N(\tau)$ satisfying (3.32) and (3.33). Various classes of N will give the same action and the same configurations as discussed in Sec. III A. The ones simplest to discuss are those represented by the choice

$$N=1, \quad 0<\tau<\tau_1 \ , \tag{3.34a}$$

$$N=\pm i, \quad \tau_1<\tau<\tau_0 \tag{3.34b}$$

for some τ_1. Let us look for solutions of this form. The differential equations for the extrema are

$$\ddot{\phi}+3\frac{\dot{a}}{a}\dot{\phi}-m^2N^2\phi=0 \ , \tag{3.35a}$$

$$\left[\frac{\dot{a}}{a}\right]^2=\dot{\phi}^2+m^2N^2\phi^2-\frac{N^2}{a^2} \ . \tag{3.35b}$$

One may conduct a search for a solution as follows: We integrate outwards from the origin with boundary conditions for regularity

$$a(0)=0, \quad \phi(0)=\phi_2, \quad \dot{\phi}(0)=0 \ , \tag{3.36}$$

where ϕ_2 is some complex constant. At some point τ_1 the value of N^2 changes according to (3.34). We match across this point with the matching conditions implied by the differential equations, namely,

$$[a]=0, \quad [\phi]=0, \quad \left[\frac{\dot{a}}{N}\right]=0, \quad \left[\frac{\dot{\phi}}{N}\right]=0 \tag{3.37}$$

for the discontinuities across τ_1. We carry on the integration until the value τ_0 is reached (if any) at which $\mathrm{Re}(a)=a_0$, its prescribed value. For given ϕ_2 and τ_1 it is unlikely that the remaining fields will assume their boundary values, namely, $\mathrm{Im}(a_0)=0$, $\mathrm{Re}(\phi)=\phi_0$, $\mathrm{Im}(\phi)=0$. These are three conditions which one can hope to satisfy by varying the three real constants in ϕ_2 and τ_1. In this way solutions can be constructed, which from the above naive counting argument we expect to be locally unique.

E. The general case

In the general case, free of the restrictions of minisuperspace considered above, one has a three-metric h_{ij}, a lapse function N, and a shift vector N^i, all of which have full spacetime dependence. The considerations of Sec. III D above suggest that a similar strategy may be applied in the general case, in order to study complex solutions. Let us take the gauge-fixing conditions $\dot{N}=0$, and

$\dot{N}^i=0$. These are the conditions that are most useful for the purposes of studying the path integral.[34] They will not fix the gauge completely, however, and some extra conditions will be necessary. Of the ten Einstein equations, the six space-space equations (3.10) involve second time derivatives of h_{ij}. The remaining four are constraint equations and are preserved by the space-space equations.

The natural generalization of the lapse method described above is to do the following. First, solve the six space-space equations for the three-metric $h_{ij}(\mathbf{x},\tau)$, subject to the boundary conditions $h_{ij}(\mathbf{x},\tau')=h'_{ij}(\mathbf{x})$, $h_{ij}(\mathbf{x},\tau'')=h''_{ij}(\mathbf{x})$. The solution will depend explicitly on the as-yet arbitrary functions $N(\mathbf{x})$ and $N^i(\mathbf{x})$. It seems reasonable to assume that a solution always exists, and is real for real N and N^i. The second step involves inserting the solution for h_{ij} into the four constraint equations. This will give four partial differential equations for the four functions $N(\mathbf{x}),N^i(\mathbf{x})$.

Clearly it would in general be very difficult to carry this through in practice, although it may allow one to say something about the nature of the solutions.[46] There is, however, a special case where this method makes contact with a perhaps more familiar problem. This is the case in which the initial and final surfaces are very close together. The method then essentially reduces to the "thin sandwich" problem.[47] For h''_{ij} and h'_{ij} very close, the six space-Einstein equations play no role, leaving only the constraints. By approximating the initial velocities by $\dot{h}_{ij}=(h''_{ij}-h'_{ij})/\delta\tau$, where $\delta\tau$ is the parameter time separation of initial and final surfaces, the problem reduces to an initial-value problem, namely, that of finding values of the lapse and shift, for given h_{ij} and $\dot{h}_{ij}$, such that the four constraints are solved. It is not difficult to show that the Hamiltonian constraint becomes a very simple purely algebraic equation, determining N^2 in terms of h_{ij}, $\dot{h}_{ij}$, and N^i. Substituting the solution into the three momentum constraints, one then obtains a rather complicated partial differential equation for N^i. It has no immediately recognizable mathematical character, but is "almost" elliptic. The conditions under which solutions exist do not appear to be known, however.

It is therefore at this point that our investigation runs up against the fact that very little is known about the Einstein equations as a boundary-value problem. We would like to know, for example, under what conditions the solution to the Einstein equations interpolating between two three-metrics (or between zero and a final three-metric) is real Euclidean, real Lorentzian, or fully complex. A reasonable conjecture is that this boundary-value problem always possesses at least one complex solution. The examples above support this. These solutions are locally unique (as we shall discuss in Sec. IV) although the issue of whether the solutions are related by complex diffeomorphisms is a more complicated problem. The behavior of the action is an important property of solutions for their semiclassical interpretation. When are there solutions and what are the properties of their action for arbitrarily large three-geometries? Does the real part of the action approach a constant as the boundary three-geometry is conformally scaled to larger and larger volumes? Under what conditions does the imaginary part of the action (regarded as a function of the boundary three-metric) vary much more rapidly than the real part? For fixed scale, which three-geometries have the least real action? For given boundary geometry which compact four-manifolds give rise to the least real action?

IV. THE RECOVERY OF QUANTUM FIELD THEORY IN CURVED SPACETIME

We have studied the complex saddle points of the path integral over four-metrics and their significance for the prediction of classical spacetime. However, the path integral involves not only a sum over four-geometries, but also a sum over matter fields on those geometries. At the saddle points of the integral over metrics therefore, there remains a functional integral over matter fields on a background geometry which is a (generally complex) solution to the Einstein equations. Such a functional integral is clearly closely related to that describing conventional quantum field theory in curved spacetime (QFTICS). Indeed, our fourth restriction on the contour is precisely that QFTICS be recovered in the limit that gravity becomes classical. The point of this section, therefore, is to discuss the extent to which this is the case. This takes us into a discussion of matter perturbations about complex saddle points.

To lowest order, one may regard the matter field Φ as a perturbation on the metric $g_{\mu\nu}$. It does not act as a source. In the leading-order saddle-point approximation, therefore, the path integral has the form

$$\Psi_0[h_{ij},\Phi'']=\int_{\mathcal{C}}\mathcal{D}g_{\mu\nu}\mathcal{D}\Phi\exp(-I_g[g_{\mu\nu}]-I_m[g_{\mu\nu},\Phi]) \approx\exp(-I_g[\hat{g}_{\mu\nu}])\int\mathcal{D}\Phi\exp(-I_m[\hat{g}_{\mu\nu},\Phi])\ , \tag{4.1}$$

where I_m is the matter action and $\hat{g}_{\mu\nu}$ is a (generally complex) solution to the Euclidean Einstein equations subject to the condition that it match the prescribed three-metric h_{ij} on the bounding three-surface. This solution may not be unique, in which case the right-hand side of (4.1) may be a sum of terms, one for each solution through which the contour passes. In general, there will only be a small number of solutions which contribute to the dominant semiclassical behavior. As mentioned in Sec. I, it may be argued that these separate terms decohere,[20] and so may be treated individually for all practical predictions.

Because the solution $\hat{g}_{\mu\nu}$ is generally complex, it has a complex action:

$$I_g[\hat{g}_{\mu\nu}]\equiv\hat{I}_g=I_R[h_{ij}]-iS[h_{ij}]\ . \tag{4.2}$$

The case we are interested in is that in which the background spacetime is approximately classical. As discussed in the previous section, this is typically realized when the imaginary part of the action S varies much more rapidly than the real part I_R. S is then an approximate solution to the Lorentzian Hamilton-Jacobi equation, and defines a set of classical trajectories with tangent vector

$$\frac{\partial}{\partial t} \equiv \int d^3\mathbf{x}\, G_{ijkl} \frac{\delta S}{\delta h_{ij}(\mathbf{x})} \frac{\delta}{\delta h_{kl}(\mathbf{x})} \equiv \nabla S \cdot \nabla \ , \tag{4.3}$$

where G_{ijkl} is the inverse DeWitt metric on superspace. Equation (4.1) may now be written

$$\Psi[h_{ij},\Phi''] \approx e^{-I_R} e^{iS} \psi_{\rm m}[h_{ij},\Phi''] \ , \tag{4.4}$$

where

$$\psi_{\rm m}[h_{ij},\Phi''] = \int \mathcal{D}\Phi \exp(-I_{\rm m}[\hat{g}_{\mu\nu},\Phi]) \ . \tag{4.5}$$

If, as we are assuming, I_R is a slowly varying function of h_{ij}, then using the fact that (4.4) is an approximate solution to the Wheeler-DeWitt equation, it may be shown that $\psi_{\rm m}$ is a solution to the functional Schrödinger equation

$$i\frac{\partial \psi_{\rm m}}{\partial t} = H_{\rm m}\left[-i\frac{\delta}{\delta\Phi}\Phi\right]\psi_{\rm m} \ , \tag{4.6}$$

where $H_{\rm m}$ is the matter Hamiltonian.[23] Alternatively, one would expect to be able to derive (4.6) from (4.5) directly, using the same approximations, but we will not do this here.

Equation (4.6) admits solutions which are normalizable in the matter modes only, in the inner product

$$(\psi_1,\psi_2) \equiv \int \mathcal{D}\Phi\, \psi_1^*(\Phi,t)\psi_2(\Phi,t) \ , \tag{4.7}$$

where $t = t[h_{ij}]$ and is defined by (4.3). This inner product is conserved under evolution by (4.6). Normalizability of the matter wave function $\psi_{\rm m}$ does not automatically follow from their path-integral representation (4.5), however. As we shall see, it depends on the nature of the metric at the saddle point, $\hat{g}_{\mu\nu}$.

The normalizable solutions to (4.6) correspond to the usual Fock-space states of QFTICS, in the Lorentzian background geometry defined by the integral curves of S. It is in this sense that QFTICS is recovered at the saddle points.[23] To illustrate this in more detail, let us consider a particular example. This example will highlight a particular way in which QFTICS may fail to be recovered.

We will consider massless scalar field perturbations about the saddle points of the de Sitter minisuperspace model, discussed in Sec. III. The four-metric is taken to be (3.21). For convenience, we take $\tau'=0$, $\tau''=1$ and take $q'=0$ corresponding to closed four-geometries. We are interested solely in the complex saddle points arising when $\lambda q'' \gg 1$, so N ($=\tilde{T}$ with these conventions) may take either of two values (3.28), or minus these values. $q(\tau)$ and N are thus given by

$$q(\tau) = -\lambda N^2\tau^2 + 2N\tau \quad \text{if } N = +\frac{1}{\lambda}[1 \pm i(\lambda q''-1)^{1/2}] \ , \tag{4.8}$$

$$q(\tau) = -\lambda N^2\tau^2 - 2N\tau \quad \text{if } N = -\frac{1}{\lambda}[1 \pm i(\lambda q''-1)^{1/2}] \ . \tag{4.9}$$

We are going to study perturbations about the metric (3.21) with $q(\tau)$ given by either of these two expressions. Note that here and in what follows, τ is strictly real. Whether or not the four-metric is real or complex is controlled solely by the lapse, N. We will proceed to study perturbations about these complex saddle points *as if* we were studying perturbations on a section of four-sphere described by the metric (3.21) with N real. The fact that we will ultimately choose N to be complex will not worry us because all the expressions we will deal with are analytic in N.

The Euclidean action for a massless minimally coupled scalar field is

$$I_{\rm m}[\Phi,g_{\mu\nu}] = \tfrac{1}{2}\int d^4x\, g^{1/2} g^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi \ . \tag{4.10}$$

The spatial dependence of $\Phi(\mathbf{x},\tau)$ is most easily treated by expanding in harmonics on S^3 (Ref. 48). One thus writes

$$\Phi(\mathbf{x},\tau) = \left[\frac{3\pi}{2G}\right]^{1/2} \sum_{nlm} f_{nlm}(\tau) Q^n_{lm}(\mathbf{x}) \ , \tag{4.11}$$

where the $Q^n_{lm}(\mathbf{x})$ are eigenfunctions of the Laplacian operator on S^3:

$$\nabla^2 Q^n_{lm}(\mathbf{x}) = -(n^2-1)Q^n_{lm}(\mathbf{x}) \ . \tag{4.12}$$

For convenience we will exclude the constant ($n=1$) mode from our considerations. It is more appropriately thought of as part of the homogeneous background. In terms of the perturbation coefficients, f_{nlm}, the action (4.10) on the background (3.21) is

$$I_{\rm m}[\Phi,q] = \sum_{nlm} I_{nlm}[f_{nlm},q] = \tfrac{1}{2}\sum_{nlm}\int_0^1 d\tau\, N\left[q^2\frac{\dot{f}^2_{nlm}}{N^2} + (n^2-1)f^2_{nlm}\right] \ . \tag{4.13}$$

It is extremized by configurations satisfying the field equations

$$q^2\ddot{f}_{nlm} + 2q\dot{q}\dot{f}_{nlm} - N^2(n^2-1)f_{nlm} = 0 \ . \tag{4.14}$$

The perturbation wave functions on our minisuperspace background are given by a path integral of the form

$$\psi_{\rm m}[\Phi''(\mathbf{x}),q''] = \int \mathcal{D}\Phi \exp(-I_{\rm m}[\Phi,q]) \ . \tag{4.15}$$

The integral is taken over all paths $\Phi(\mathbf{x},\tau)$ that match the value $\Phi''(\mathbf{x})$ at $\tau=1$ and are regular on the background metric. Because the individual perturbation modes decouple, one may write

$$\psi_{\rm m}[\Phi''(\mathbf{x}),q''] = \prod_{nlm} \psi_{nlm}(f''_{nlm},q'') \ . \tag{4.16}$$

The wave function for each mode $\psi_{nlm}(f''_{nlm},q'')$ is therefore given by a path integral:

$$\psi_{nlm}(f''_{nlm},q'') = \int \mathcal{D}f_{nlm} \exp(-I_{nlm}[f_{nlm},q]) \ . \tag{4.17}$$

The path integral (4.17) is over paths $f_{nlm}(\tau)$ satisfying the boundary conditions $f_{nlm}(1) = f''_{nlm}$, $f_{nlm}(0)=0$, the latter condition following from the requirement of regularity at $q=0$.

Equation (4.17) may be evaluated exactly to yield the result

$$\psi_{nlm}(f''_{nlm},q'')=A(q'')\exp(-\hat{I}_{nlm}) , \quad (4.18)$$

where $\hat{I}_{nlm}$ is the action of the solution $\hat{f}_{nlm}(\tau)$ to the classical field equation (4.14) satisfying the above boundary conditions. The prefactor A may be evaluated by standard methods, but will be ignored in what follows. Using (4.14), one may obtain the following expression for $\hat{I}_{nlm}$:

$$\hat{I}_{nlm}=\frac{1}{2N}[q(\tau)\hat{f}_{nlm}(\tau)\dot{\hat{f}}_{nlm}(\tau)]_{\tau=0}^{\tau=1} . \quad (4.19)$$

Now let us find an explicit expression for the solution $\hat{f}_{nlm}$. Consider first the case of the two saddle points with $\mathrm{Re}(N)>0$. Then $q(\tau)$ is given by (4.8). There are two linearly independent solutions to the field equation (4.14). The first is

$$\hat{f}_{nlm}(\tau)=\mathrm{const}\times\tau^{(n-1)/2}(2-\lambda N\tau)^{-(n+1)/2} \times(\lambda N\tau-n-1) . \quad (4.20)$$

It is regular (indeed goes to zero) as $\tau\to 0$. The second solution is obtained from (4.20) by letting $n\to -n$. It blows up as $\tau\to 0$. Only the first of these two solutions is consistent with regularity; thus the solution $\hat{f}_{nlm}(\tau)$ is given by (4.20) with the constant chosen so that $\hat{f}_{nlm}(1)=f''_{nlm}$. Inserting this solution in (4.19), one obtains

$$\hat{I}_{nlm}=\frac{(n^2-1)q''^2 f''^2_{nlm}}{2N(\lambda N-2)(\lambda N-n-1)} . \quad (4.21)$$

This expression is valid for real or complex N, but only for the two saddle points (4.8) for which $\mathrm{Re}(N)>0$. One thus obtains the following expressions for the perturbation wave functions:

$$\psi_{nlm}(f''_{nlm},q'') =A(q'')\exp\left[-\frac{(n^2-1)q''^2[n\pm i(\lambda q''-1)^{1/2}]}{2(n^2-1+\lambda q''^2)}f''^2_{nlm}\right] . \quad (4.22)$$

Now consider the other saddle points with $\mathrm{Re}(N)<0$. $q(\tau)$ is then given by (4.9). A repeat of the above analysis then yields, in place of (4.21), the result

$$\hat{I}_{nlm}=\frac{-(n^2-1)q''^2 f''^2_{nlm}}{2N(-\lambda N-2)(-\lambda N-n-1)} \quad (4.23)$$

for the action of the classical solutions. For the wave functions about these saddle points with negative $\mathrm{Re}(N)$, one thus obtains

$$\psi_{nlm}(f''_{nlm},q'') =A(q'')\exp\left[+\frac{(n^2-1)q''^2[n\pm i(\lambda q''-1)^{1/2}]}{2(n^2-1+\lambda q''^2)}f''^2_{nlm}\right] . \quad (4.24)$$

The perturbation wave functions about the positive $\mathrm{Re}(N)$ saddle points, (4.22), differ from those about the negative $\mathrm{Re}(N)$ saddle points (4.24), in one and only one crucial respect—the sign of the exponent. The former *decay* exponentially for large fluctuations, whereas the latter *grow* exponentially. The latter, therefore, while they are solutions to the Schrödinger equation, are not normalizable in the inner product (4.7), so do not correspond to Fock space states. This means that, in this example, QFTICS *is not recovered at the saddle points for which* $Re(N)<0$.

From the discussion of these saddle points in Sec. III, it is easily seen that this failure to recover QFTICS is not an artifact of this particular model, but is quite general. Like the gravitational action, the matter action changes sign under $N\to -N$. Because the Euclidean matter action with N real and positive is positive definite, it will be negative definite when the sign of N is reversed. Clearly if one attempted to do quantum field theory with such an action, one would very quickly run into difficulties. We have therefore derived a restriction on the contour from the demand that QFTICS be recovered in the saddle-point approximation: it is that *the contour should not be dominated by a saddle point with* $\mathrm{Re}(\sqrt{g})<0$.

Although we have concentrated so far on the no-boundary proposal, virtually all of our conclusions apply to other proposals, provided they have a sum-over-histories formulation defined by a suitable choice of contour $\mathcal{C}$. In particular, the above restriction on the contour is of significance for the so-called "tunneling" boundary condition proposed by Linde[49] and by Vilenkin.[50,51] Vilenkin has presented his version of this boundary condition in various different ways, not all of which are obviously equivalent. For the purposes of this paper, which is to discuss integration contours, the most convenient expression of his proposal is that phrased in terms of a path integral. This is that the wave function be given by a path integral over Lorentzian geometries which close off in the past.[51] This form of the Vilenkin proposal was studied in Ref. 28, where it was shown that although it does define a unique contour of integration, and thus a unique wave function, the contour is dominated by a saddle point for which $\mathrm{Re}(\sqrt{g})<0$. As explained above, one would not, therefore, expect to recover QFTICS when considering matter perturbations about this saddle point. This appears to rule out *this particular form* of the Vilenkin proposal as a viable candidate for the boundary conditions on the wave function of the universe.

It is important to note that this conclusion applies only to one particular form of the Vilenkin proposal: namely, the form involving a sum of Lorentizan geometries. It is this and this alone that appears to fall foul of our fourth criterion. As mentioned above, there is another, perhaps better known form of the Vilenkin proposal: namely, that the wave function consist solely of outgoing modes at singular boundaries of superspace.[50] This is not by any means obviously equivalent to the other form criticized above. This version of the proposal appears to be perfectly consistent with the recovery of QFTICS.

The proposal of Linde appears, in simple examples, to give a wave function identical to that of Vilenkin.[49] Linde also seems to regard a purely Lorentzian path integral as his starting point. Because the usual Wick rotation to a Euclidean action for gravity leads to a minus sign in front of the kinetic term for the scale factor, he proposed that the Wick rotation should be performed in the "wrong" direction. In the de Sitter minisuperspace model of Ref. 28, this proposal appeared to pick out a contour identical to that picked out by the path-integral version of the Vilenkin proposal. It also, therefore, is dominated by a saddle point with $\mathrm{Re}(\sqrt{g}\,)<0$, and thus is inconsistent with the recovery of QFTICS. In fact, this failure to recover QFTICS is more obvious in this proposal—clearly if one does the Wick rotation in the wrong direction, although it gives a desirable sign to the conformal part of the gravitational action, any matter action will become negative definite, rather than positive definite. Linde responds to this difficulty by arguing that one should proceed phenomenologically and choose different contours, or different Wick rotations, for each part of the total action separately, depending on what it is that one is calculating.[52] This point of view, for which we see little motivation, is contrary to the point of view we are taking in this paper, which is that one and the same contour should be used for both the gravitational and matter parts of the path integral.

Finally, note that as mentioned in Sec. III, the solution (4.20) to the scalar field equation is complex for complex N. Furthermore, it does not have the property that $\hat{\bar{f}}_{nlm}=0$ at the value of τ for which $\dot{q}$ vanishes. This means that in the complex T plane (recall $T=N\tau$) there is no contour along which both q *and* f_{nlm} are real. The complex solution for f_{nlm}, therefore, is not transformed into a combination of real Euclidean and real Lorentzian solutions by the transformation which does this for q. It is therefore an example of an inescapably complex solution. We can say the same slightly differently: if we had carried out this perturbation calculation not on the complex background metric (3.21), but on the equivalent background metric consisting of a real Euclidean metric on half a four-sphere matched onto a real Lorentzian metric on a section of de Sitter space, then we would have found that the scalar field has to be complex to match the prescribed boundary values. The purely real tunneling solutions which dominate the no-boundary wave function for highly symmetrical points in its configuration space are, therefore, isolated points in an immediate neighborhood of generally complex solutions.

V. WORMHOLES AND THE INDEFINITENESS OF THE ACTION

As briefly discussed in the Introduction, recent work by Hawking,[25] Coleman,[26] Giddings and Strominger,[27] and others indicates that wormhole configurations in the path integral will modify the fundamental coupling constants. In particular, Coleman argued that they may provide an explanation as to why the cosmological constant is zero.[26] Of the many technicalities involved in these calculations, it seems likely that the choice of contour of integration will be important. The point of this section, therefore, is to try to establish the extent to which these calculations depend on the contour. It will then be possible to determine what restrictions, if any, are imposed on the contour by demanding that vanishing cosmological constant be a prediction.

Let us begin by reviewing the mechanism, identifying the points at which the contour is an issue. We follow the original version of the mechanism, as given by Coleman.[26] However, we anticipate that because what we have to say about it is rather general, our remarks will also apply to the various modified versions that have since been put forward. We are interested in calculating the wave function of the universe, $\Psi[h_{ij}]$, a functional of the three-metric h_{ij} on a three-surface ∂M. ∂M may consist of a number of disconnected pieces, as would be the case if the "initial" or "final" states included a number of disconnected baby universes. For convenience, we will consider only the case in which ∂M consists of a single connected piece. Also for convenience, we will not include matter fields. Using the no-boundary proposal, the wave function is defined by summing over four-metrics on compact four-manifolds whose only boundary is ∂M. One thus has

$$\Psi_0[h_{ij}]=\int_{\mathcal{C}}\mathcal{D}g_{\mu\nu}\exp(-I[g_{\mu\nu},\Lambda]) \ . \tag{5.1}$$

The manifolds summed over are connected ones. However, it is assumed that there is a separation of length scales so that we may talk about "large" manifolds and "small" interconnecting wormholes. It is then convenient to think of the sum over manifolds as the combination of (i) a sum over large connected manifolds which have ∂M as their only boundary, (ii) a sum over large disconnected manifolds which have no boundary at all, and (iii) a sum over small wormhole configurations connecting these two types of large manifolds to themselves and to each other. It is next argued that, for observables on scales much larger than the wormhole scale, one may replace the wormhole and baby universe configurations by an effective local field theory on the large manifolds. Using this effective field theory to sum over all the wormhole connections, one obtains a path integral in which the sum over wormholes is replaced by a sum over couplings. In particular, the fully renormalized, low-energy cosmological constant, observed in the dynamics of the universe on very large scales, is summed over

$$\Psi_0[h_{ij}]\approx\int d\Lambda\int_{\mathcal{C}}\mathcal{D}g_{\mu\nu}\exp(-I[g_{\mu\nu},\Lambda]) \ , \tag{5.2}$$

where we have ignored irrelevant factors. In (5.2) the sum over manifolds no longer involves a sum over wormholes but only the sums (i) and (ii) above over large manifolds. This sum over the cosmological constant is the main effect of interest that wormholes have.

While some authors have questioned the details of the derivation of (5.2) (e.g., Ref. 53), the general effect, that a sum over couplings is induced, appears to be a reasonably robust feature. It is not obviously affected by the choice of contour. The value of (5.2) will, however, be affected by this choice.

Let us defer for the moment discussion of the sum (ii) over disconnected manifolds without boundary. Because these manifolds have no boundary their contribution does not depend on h_{ij}, the three-metric on ∂M, and in fact leads only to a factor depending on Λ. This factor is very important but will be discussed below. Consider, then, the expression (5.2) where the sum is taken to be only over connected manifolds with boundary ∂M. We are interested in the wave function of the universe when the bounding three-geometry ∂M is large and spacetime is classical. As discussed in Sec. III, the dominant contribution to the functional integral over metrics is then expected to come from saddle points at which the four-metric is complex. They will have complex action, of the form $I=I_R(\Lambda,h_{ij})-iS(\Lambda,h_{ij})$, where I_R and S are real. In a region of superspace where the wave function predicts classical spacetime, the real part of the action will vary much more slowly than the imaginary part which satisfies the Lorentzian Hamilton-Jacobi equation. We neglect the dependence of I_R on h_{ij} in what follows. The saddle-point approximation (5.2) thus leads to an expression of the form

$$\Psi_0[h_{ij}]=\int d\Lambda\, e^{-I_R(\Lambda)}e^{iS(\Lambda,h_{ij})} . \tag{5.3}$$

More generally, it will lead to a sum of expressions of this form, depending on how many saddle points contribute, but this will not affect what follows.

Consider the integrand of (5.3). Each term $e^{iS(\Lambda,h_{ij})}$ is an approximate WKB solution to the Wheeler-DeWitt equation for a particular value of Λ. It is peaked about the set of Lorentzian solutions to the field equations with Hamilton-Jacobi function $S(\Lambda,h_{ij})$ and cosmological constant Λ. The total wave function of the Universe (5.3), therefore, represents an ensemble of classical universes in which all possible values of the cosmological constant are realized. The real part of the Euclidean action of the saddle point, $I_R(\Lambda)$, provides a weighting $e^{-I_R(\Lambda)}$ in the sum over Λ. More precisely, $e^{-2I_R(\Lambda)}$ may be thought of as the probability that a particular value of Λ will be realized along one of the classical trajectories about which the wave function is peaked.

Clearly to predict that $\Lambda=0$, the distribution $e^{-I_R(\Lambda)}$ must be very strongly peaked about $\Lambda=0$. To this end, it is argued that the leading-order contribution to the path integral comes from saddle points for which I_R is the action of half of a four-sphere. These saddle points have the crucial property that (the real part of) their action is negative. In particular, $I_R=-c/\Lambda$, where $c=24\pi^2/l^2$. It then follows that $e^{-I_R(\Lambda)}$ is indeed strongly peaked about $\Lambda=0$. This argument, involving just a single exponential, is often referred to as the Baum-Hawking argument.[25,54]

Now let us return to include the effect of the sum (ii) over disconnected manifolds without boundary. Coleman argues that this sum leads to the insertion into (5.3) of a factor of the form

$$Z(\Lambda)=\exp\left[\int \mathcal{D}g_{\mu\nu}e^{-I[g,\Lambda]}\right] , \tag{5.4}$$

where the functional integral in the exponent is now over *connected* manifolds without boundary. Equation (5.3) is now of the form

$$\Psi_0[h_{ij}]=\int d\Lambda\, Z(\Lambda)e^{-I_R(\Lambda)}e^{iS(\Lambda,h_{ij})} . \tag{5.5}$$

Again it is expected that the functional integral in (5.4) is dominated by four-sphere configurations. $Z(\Lambda)$ is thus of the form

$$Z(\Lambda)=\exp(e^{2c/\Lambda}) \tag{5.6}$$

(the factor of 2 is there because these are now whole spheres not just hemispheres). This is Coleman's celebrated double-exponential distribution. As a result of including it, the peak at $\Lambda=0$ observed in (5.3) becomes exponentially enhanced in (5.5).

The results described above depend primarily on the saddle-point approximation to the path integral. In particular, they rely crucially on the assumption that the path integral is dominated by saddle points whose action has negative real part. Clearly the *existence* of these saddle points cannot be affected by the choice of contour. There has, however, been some confusion in the literature surrounding this point. It is sometimes suggested that choosing a contour along which the real part of the action is everywhere positive definite—as one tries to do in the conformal rotation, for example—is not consistent with having a negative action saddle point, because that contour naively appears never to receive a contribution from configurations with negative action. There is no such inconsistency. Even if the prescribed contour is one along which the (real part of the) action is everywhere positive, it can always be distorted into one along which the action is at some stage negative. This is typically the case in one-dimensional examples when the prescribed contour is distorted into a steepest-descent contour, for example. There, the integral along a contour with everywhere positive action is approximated by an integral along the distortion of that contour in the immediate neighborhood of a negative action saddle point. A related point of confusion concerns the convergence of the path integral. It has been suggested that the fact that the action goes negative at some stage—at a saddle point, for example—is not consistent with convergence. This is no inconsistency here, either. Failure to converge would result only if action was allowed to become *arbitrarily* negative at some point along the contour.

Although the existence of negative action saddle points is not affected by the choice of contour, the question of whether or not a given negative action saddle point supplies the *dominant contribution* to the integral will depend very much on the contour. In particular, it is true only if the prescribed contour may be distorted into a contour for which the saddle point in question is a global maximum of the integrand. The sort of difficulty that can arise is that, as discussed in Sec. III, there exist saddle points with the "wrong" sign for the action. For every saddle point with $\mathrm{Re}(\sqrt{g})>0$ and negative action, as used in (5.3) and (5.5), there exists another saddle point with $\mathrm{Re}(\sqrt{g})<0$ and positive action. Clearly if the prescribed contour may be distorted into a contour dominated by one of the latter saddle points and not one of the former, then the weighting factors in the sum over Λ

would be of the form $e^{-c/\Lambda}$. $\Lambda=0$ would not then be predicted.

Fortunately, in the case of a Baum-Hawking argument, the possibility of such a contour can be ruled out by the considerations of the previous section. For if one were to choose a contour dominated by a saddle point with $\mathrm{Re}(\sqrt{g}\,)<0$, yielding a factor of the form $e^{-c/\Lambda}$, the real part of the matter action would be negative and, as already argued, QFTICS would not be recovered. The approximate wave function (5.3) is obtained by considering just one saddle point along the contour. One and the same saddle point determines the sign of both I_R and the matter action: negative I_R is tied to positive matter action, and vice versa.

This saving feature does not obviously apply to the Coleman argument, however. Although in the fundamental path-integral expression (5.1) the contour is supposedly fixed once and for all, when one passes to the effective theory the path integral splits into two parts. One considers the sum over large connected manifolds with boundary (i) and, separately, one considers the integral over disconnected manifolds (ii) leading to the factor (5.4). While it may, in principle, be possible to derive the appropriate contours for each of these integrals from the contour specified in the fundamental expression (5.1), it seems unlikely that this will be possible in practice. In the effective theory described by these two path integrals, therefore, the contour is once again up for grabs. In particular, it seems that one is allowed to choose it differently in each of these two integrals. Thus although the fourth criterion rules out the possibility of negative $\mathrm{Re}(\sqrt{g}\,)$ saddle points in the sum over connected manifolds with boundary, it does not obviously rule out this possibility in the sum over spheres leading to the double exponential factor (5.6). By an appropriate choice of contour, therefore, it is conceivable that a result of the form $\exp(e^{-2c/\Lambda})$ could be obtained in place of (5.6).

A related point concerns the sign of the action of the wormholes. If $\mathrm{Re}(\sqrt{g}\,)>0$, the wormholes have positive action. This is important because it means that large wormholes are suppressed, consistent with the assumption of a separation of length scales. In the same way that $\mathrm{Re}(\sqrt{g}\,)<0$ means positive action for spheres, $\mathrm{Re}(\sqrt{g}\,)<0$ also means *negative* action for the wormholes. Large wormholes would not then be suppressed.

These considerations all point to a possible restriction on the contour, if $\Lambda=0$ is to be predicted: it is that none of the dominating saddle points should have $\mathrm{Re}(\sqrt{g}\,)<0$. As already noted, this is related to the recovery of QFTICS discussed in Sec. IV.

VI. SUMMARY AND DISCUSSION

There is no satisfactory covariant Hamiltonian quantum mechanics for closed cosmologies from which the contour for the no-boundary wave function may be derived. No otherwise compelling prescription for its choice has been advanced, such as the Gibbons-Hawking-Perry prescription for asymptotically flat spacetimes. It therefore seems natural to search *generally* for suitable integration contours that define the no-boundary wave function, subject to five physically motivated criteria. This search was the subject of this paper.

The first two criteria—convergence, and that the wave function be annihilated by the constraints—were discussed in Sec. II. Because the Euclidean Einstein-Hilbert action is not bounded from below, the convergent contours are necessarily complex. Beyond this restriction, it is easily seen that convergent contours exist. They may be exhibited explicitly in simple models.

The demand that the wave function be annihilated by the constraints restricts the domain of integration of the contour. The domain must be diffeomorphism invariant. This means that for each field component, the contour must be infinite in length, closed, or have a range consistent with the periodicity of the functions entering the integrand. It may not have finite gauge-variant end points.

The requirement that classical spacetime be a prediction was discussed in Sec. III. For classical spacetime to be predicted, it is necessary that the semiclassical approximation to the wave function be oscillatory, and, hence, that the contour be dominated by one or more saddle points at which the four-metric is complex. Complex solutions to the Einstein equations are thus of interest. We discussed the conditions for their existence, how to find them, and how to interpret them. Through studying simple examples and special cases, we argued that a complex solution usually exists for generic boundary data, although real Euclidean or Lorentzian solutions exist only under special circumstances. This nonexistence of a real Euclidean solution for generic boundary data is perhaps the single most important fact for the prediction of classical spacetime. In the space of complex four-metrics the action is doubled valued. This means that to every solution of the Einstein equations, there correspond two saddle points of the path integral, one with $\mathrm{Re}(\sqrt{g}\,)>0$, the other with $\mathrm{Re}(\sqrt{g}\,)<0$. These two saddle points are identical in all respects except the sign of their action. We argued that because we allow the contour to be complex, both types of saddle point are candidate contributors to the path integral.

In Sec. IV we asked what restrictions might be imposed on the contour by insisting that in the limit that gravity becomes classical conventional quantum field theory be recovered in that spacetime background. At saddle points of the sum over metrics for which $\mathrm{Re}(\sqrt{g}\,)>0$, QFTICS is generally recovered. At the saddle points for which $\mathrm{Re}(\sqrt{g}\,)<0$, however, the (real part of the) matter action would generally be negative definite and QFTICS would not be recovered. This suggested that a sensible choice of contour should not be dominated by a saddle point with $\mathrm{Re}(\sqrt{g}\,)<0$. This restriction was found to rule out certain sum-over-histories forms of the tunneling boundary condition proposals of Linde and Vilenkin.

Finally, in Sec. V, we asked to what extent the wormhole calculations of Coleman, Giddings, and Strominger, Hawking and others are sensitive to the choice of contour. Or more precisely, what constraints on the contour could be imposed by demanding that $\Lambda=0$ be predicted? The main conclusion was a restriction on the contour identical to that of Sec. IV: that the path integral not be dominated by saddle points with $\mathrm{Re}(\sqrt{g}) < 0$. The reason here is that a distribution for Λ involving factors of the form $e^{1/\Lambda}$, peaked at $\Lambda = 0$, is a prediction only at the saddle points for which $\mathrm{Re}(\sqrt{g}) > 0$.

We may now draw together all that we have learned about the contour from the considerations of this paper. If one is to define the wave function of the universe using a path integral, then the contour should satisfy the following:

(a) The integral along it should converge;

(b) the contour for each gauge variant field component should be infinite, closed, or consistent with a periodicity of the integrand;

(c) when the bounding three-geometry is large, the integral along the contour should be dominated by one or more complex saddle points for which the imaginary part of the action varies much more rapidly than the real part;

(d) the dominating saddle points should in addition be such that $\mathrm{Re}(\sqrt{g}) > 0$.

These requirements are the essential content of the five criteria proposed in Sec. I, transcribed into a form directly applicable to specific examples.

Do requirements (a)–(d) fix a class of contours leading to a unique wave function of the universe? It seems not. In the space of complex four-metrics, the integrand is a very complicated analytic function with many poles for the closed contours to encircle and many inequivalent directions in which the contour may go off to infinity. Requirements (c) and (d) on the nature of the dominating saddle points will reduce the available freedom to some extent, but it seems most unlikely that a unique wave function will remain. This is immediately apparent in simple models in which the contours may be explicitly exhibited. It means that the no-boundary proposal, as it is currently defined, does not imply a unique wave function of the universe. One could say that there are many no-boundary proposals depending on how the contour is chosen. A unique no-boundary wave function will be obtained only after extra information singling out a contour is put in. As in simple examples, however, it may be that inequivalent contours can give essentially equivalent semiclassical predictions. These are all we are likely to be able to deal with in quantum cosmology in the near future.

ACKNOWLEDGMENTS

Throughout the course of this work we have greatly benefited from conversations with Bob Geroch, Gary Horowitz, Karel Kuchař, Jorma Louko, Emil Mottola, Robert Myers, Don Page, Kristin Schleich, Andrew Strominger, Alex Vilenkin, Bob Wald, Bernard Whiting, and Jim York. J.J.H. was supported in part by NSF Grant No. PHY82-17853, supplemented by funds from the National Aeronautics and Space Administration, at the University of California at Santa Barbara. J.B.H. was supported in part by NSF Grant No. PHY85-06686 and by the John Simon Guggenheim Memorial Foundation. He is grateful for the hospitality of the Institute for Theoretical Physics, Santa Barbara, and of the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, where part of this work was carried out. We are both grateful to the Aspen Center for Physics for hospitality during the course of this work.

*Present address: Center for Theoretical Physics, Laboratory for Nuclear Science, Massachusetts Institute of Technology, Cambridge, MA 01239.

[1]For reviews of the current status of this enterprise, see J. J. Halliwell, ITP Report No. ITP-NSF-88-131, 1988 (unpublished); J. B. Hartle, in *Highlights in Gravitation and Cosmology*, edited by B. R. Iyer, A. Kembhavi, J. Narlikar, and C. V. Vishveshwara (Cambridge University Press, Cambridge, England, 1988); in *Proceedings of the 12th International Conference on General Relativity and Gravitation* (Cambridge University Press, Cambridge, England, 1990). For a bibliography of papers, see J. J. Halliwell, ITP Report No. NSF-ITP-88-132, 1988 (unpublished).

[2]D. Boulware and S. Deser, Ann. Phys. (N.Y.) **89**, 193 (1975); see, however, R. M. Wald, Phys. Rev. D **33**, 3613 (1986).

[3]By the phrase, "Hamiltonian quantum mechanics," we mean the quantum mechanics obtained by solving the constraints classically in a fixed gauge thereby exposing the true physical degrees of freedom. For parametrized theories such as general relativity, this process also involves identifying the true internal time of the system. It is this time that is also used as the preferred time in the Schrödinger equation in Hamiltonian quantum mechanics. Thus, in particular by "Hamiltonian quantum mechanics" we do not mean simply producing an operator form of the constraints which annihilates physical wave functions as in Dirac quantization. Such constraints still treat all dynamical variables equally (see Sec. II) and do not require a preferred time. However, operator constraints by themselves are not a complete quantum mechanics. For a detailed and lucid discussion of these points, see, for example, K. Kuchař in, *Quantum Gravity 2: A Second Oxford Symposium*, edited by C. J. Isham, R. Penrose, and D. Sciama (Clarendon, Oxford, 1981).

[4]J. W. York, Phys. Rev. Lett. **28**, 1082 (1972).

[5]A. Ashtekar, Phys. Rev. Lett. **57**, 2244 (1986); Phys. Rev. D **36**, 1587 (1987). See also A. Ashtekar, with invited contributions in *New Perspectives in Quantum Gravity* (Bibliopolis, Naples, 1988).

[6]J. B. Hartle, Phys. Rev. D **37**, 2818 (1988); **38**, 2985 (1988); UCSB report, 1989 (unpublished); in *Proceedings of the Osgood Hill Meeting on Conceptual Problems in Quantum Gravity*, edited by A. Ashtekar and J. Stachel (Birkhauser, Boston, 1989); or, in *Proceedings of the Fifth Marcel Grossmann Meeting on General Relativity*, Perth, Australia, 1988, edited by D. Blair and M. J. Buckingham (World Scientific, Singapore, 1989).

[7]S. W. Hawking, in *Astrophysical Cosmology*, edited by H. A. Brück, G. V. Coyne, and M. S. Longair (Pontifica Academia Scientarium, Vatican City, 1982); J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).

[8]S. W. Hawking, Nucl. Phys. **B239**, 257 (1984).

[9]See, for example, J. B. Hartle, Class. Quantum Grav. **2**, 707 (1985).

[10]See, for example, H. Leutwyler, Phys. Rev. **134**, 1155 (1964); B. S. DeWitt, in *Magic Without Magic: John Archibald Wheeler, a Collection of Essays in Honour of his 60th Birthday,* edited by J. Klauder (Freeman, San Francisco, 1972); E. Fradkin and G. Vilkovisky, Phys. Rev. D **8**, 4241 (1973); L. Faddeev and V. Popov, Usp. Fiz. Nauk. **111**, 427 (1973) [Sov. Phys. Usp. **16**, 777 (1974)]; M. Kaku, Phys. Rev. D **15**, 1019 (1977); C. Teitelboim, *ibid.* **28**, 310 (1983); H. Hamber, in *Critical Phenomena, Random Systems and Gauge Theories,* proceedings of the Les Houches Summer School, Les Houches, France, 1984, edited by K. Osterwalder and R. Stora (Les Houches Summer School Proceedings Vol. 43) (North-Holland, Amsterdam, 1986); M. Bander, Phys. Rev. Lett. **57**, 1825 (1986).

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

[12]R. Schoen and S. T. Yau, Phys. Rev. Lett. **61**, 263 (1988).

[13]K. Schleich, Phys. Rev. D **36**, 2342 (1987); J. B. Hartle, *ibid.* **29**, 2730 (1984); J. B. Hartle and K. Schleich, in *Quantum Field Theory and Quantum Statistics: Essays in Honour of the Sixtieth Birthday of E. S. Fradkin,* edited by I. A. Batalin, G. A. Vilkovisky, and C. J. Isham (Hilger, Bristol, 1987); H. Arisue, T. Fujuwara, M. Kato, and K. Ogawa, Phys. Rev. D **35**, 2309 (1987); D. Boulware (unpublished).

[14]P. O. Mazur and E. Mottola, Los Alamos Report No. LA-UR-89-340, 1989 (unpublished).

[15]To transform a metric of scalar curvature R to one for which $R=0$, the conformal factor Ω must satisfy $(-\nabla^2+R/6)\Omega=0$. If R is too negative, however, the operator $(-\nabla^2+R/6)$ will have a negative eigenvalue. Its zero mode will not therefore be the lowest mode and so will possess a node. The conformal transformation will therefore be singular. Further discussion of this point may be found in Ref. 16.

[16]S. W. Hawking, in *General Relativity: An Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).

[17]On a compact four-manifold without boundary, an arbitrary metric may be conformally transformed to one for which the scalar curvature is constant. This result was first conjectured and partially proved by H. Yamabe [Osaka Mat. J. **12**, 21 (1960)]. The proof has since been completed [for a useful review, see J. M. Lee and T. H. Parker, Bull. Am. Math. Soc. **17**, 37 (1987)]. The sign of the constant scalar curvature cannot, however, be prescribed—an arbitrary metric cannot be conformally transformed to one for which R is, say, positive.

[18]An interesting choice of equivalence class condition avoiding some of the aforementioned difficulties has been suggested by G. Horowitz, although this has not yet been fully developed into a prescription for the contour (G. Horowitz, talk presented at the Fermilab Workshop, 1989 (unpublished).

[19]C. Kiefer, Class. Quantum Grav. **4**, 1369 (1987); T. Fukuyama and M. Morikawa, Phys. Rev. D **39**, 462 (1989); M. Gell-Mann and J. B. Hartle (in preparation); F. Mellor, Newcastle report (unpublished); M. Morikawa, Phys. Rev. D **40**, 4023 (1989); T. Padmanabhan, Phys. Rev. D **39**, 2924 (1989); H. D. Zeh, Phys. Lett. A **116**, 9 (1986); **126**, 311 (1988).

[20]J. J. Halliwell, Phys. Rev. D **39**, 2912 (1989); in *Complexity, Entropy and the Physics of Information,* SFI Studies in the Sciences of Complexity, edited by W. H. Zurek (Addison-Wesley, Reading, Mass., 1990), Vol. IX.

[21]J. J. Halliwell, Phys. Rev. D **36**, 3626 (1987).

[22]C. Kiefer, Phys. Rev. D **38**, 1761 (1988).

[23]The recovery of quantum field theory in curved spacetime from quantum cosmology has been discussed by many people, including T. Banks, Nucl. Phys. **B249**, 332 (1985); B. DeWitt, Phys. Rev. **160**, 1113 (1967); J. J. Halliwell, DAMTP report, 1986 (unpublished); J. J. Halliwell and S. W. Hawking, Phys. Rev. D **31**, 1777 (1985); V. Lapchinsky and V. A. Rubakov, Acta Phys. Pol. **B10**, 1041 (1979). See also Ref. 48.

[24]J. J. Halliwell and R. Myers, Phys. Rev. D **40**, 4011 (1989).

[25]S. W. Hawking, Phys. Lett. **134B**, 403 (1984).

[26]S. Coleman, Nucl. Phys. **B310**, 643 (1988).

[27]S. Giddings and A. Strominger, Nucl. Phys. **B306**, 890 (1988); **B307**, 854 (1988); **B321**, 481 (1989).

[28]J. J. Halliwell and J. Louko, Phys. Rev. D **39**, 2206 (1989).

[29]J. J. Halliwell and J. Louko, Phys. Rev. D **40**, 1868 (1989).

[30]J. J. Halliwell and J. Louko, 1989 (in preparation).

[31]J. B. Hartle, J. Math. Phys. **30**, 452 (1989).

[32]J. Louko and P. Tuckey (in preparation).

[33]J. J. Halliwell and J. B. Hartle, ITP report, 1989 (unpublished).

[34]C. Teitelboim, Phys. Rev. D **25**, 3159 (1982); Phys. Rev. Lett. **50**, 705 (1983).

[35]J. J. Halliwell, Phys. Rev. D **38**, 2468 (1988).

[36]Euclidean solutions of the Einstein equations are generally analytic functions of harmonic or geodesic normal coordinates in a neighborhood of the real axis. See, e.g., A. L. Besse, *Einstein Manifolds* (Springer, Berlin, 1987), p. 145.

[37]P. Amsterdamski, Phys. Rev. D **31**, 3073 (1985); S. W. Hawking and J. C. Luttrell, Phys. Lett. **143B**, 83 (1984); I. G. Moss and W. A. Wright, *ibid.* **154B**, 115 (1985).

[38]S. W. Hawking and D. N. Page, Nucl. Phys. **B264**, 185 (1986); **B298**, 789 (1988); A. Vilenkin, Phys. Rev. D **37**, 888 (1988).

[39]R. Geroch, J. Math. Phys. **8**, 782 (1967).

[40]See, for example, C. Misner, K. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1970), p. 551 ff.

[41]There are real tunneling solutions such as this on manifolds other than B^4. These are important for arguments which seek to predict the large-scale topology of the universe from the no-boundary proposal. See G. W. Gibbons and J. B. Hartle (in preparation).

[42]These conditions are actually quite subtle. See, for example, J. Louko, Phys. Lett. B **202**, 201 (1988).

[43]D. Page has given a related construction (private communication).

[44]R. Myers, Phys. Rev. D **38**, 1327 (1988).

[45]C. P. Burgess and A. Kshirsagar, Nucl. Phys. **B324**, 157 (1989); J. D. Brown, C. P. Burgess, A. Kshirsagar, B. Whiting, and J. W. York, Nucl. Phys. **B328**, 213 (1989).

[46]A particular simplification could conceivably arise using the Ashtekar variables (Ref. 5). In these variables, the constraints, and thus the total Hamiltonian, are at most quadratic. This means that the analogue of the space-space equations in the Ashtekar variables could be particularly simple, and the main difficulties would then reside in the constraint equations for the Lagrange multipliers.

[47]J. Wheeler, in *Relativity, Groups and Topology,* Les Houches, France, 1963, edited by C. DeWitt and B. DeWitt (Gordon and Breach, New York, 1963).

[48]Inhomogeneous perturbations about minisuperspace have been studied by many people including T. Banks, W. Fischler, and L. Susskind, Nucl. Phys. **B262**, 159 (1985); P. D. D'Eath and J. J. Halliwell, Phys. Rev. D **35**, 1100 (1987); W. Fischler, B. Ratra, and L. Susskind, Nucl. Phys. **B259**, 730 (1985); J. J.

Halliwell and S. W. Hawking, Phys. Rev. D **31**, 1777 (1985); R. Laflamme, Phys. Lett. B **198**, 156 (1987); I. Shirai and S. Wada, Nucl. Phys. **B303**, 728 (1988); T. Vachaspati and A. Vilenkin, Phys. Rev. D **37**, 898 (1988); A. Vilenkin, *ibid.* **37**, 888 (1986); S. Wada, Nucl. Phys. **B272**, 729 (1986). See also Ref. 23.

[49]A. Linde, Zh. Eksp. Teor. Fiz. **87**, 369 (1984) [Sov. Phys. JETP **60**, 211 (1984)]; Lett. Nuovo Cimento **39**, 401 (1984); Rep. Prog. Phys. **47**, 925 (1984).

[50]A. Vilenkin, Phys. Rev. D **33**, 3560 (1986); **37**, 888 (1988).

[51]A. Vilenkin, Phys. Rev. D **30**, 509 (1984).

[52]A. Linde (private communication).

[53]W. G. Unruh, Phys. Rev. D **40**, 1053 (1989).

[54]E. Baum, Phys. Lett. **133B**, 183 (1983).