Fermions in quantum cosmology

P. D. D'Eath and J. J. Halliwell
Department of Applied Mathematics and Theoretical Physics, Silver Street, Cambridge CB3 9EW, United Kingdom
(Received 23 June 1986)

This paper considers the extension of the ideas of quantum cosmology and, in particular, the proposal of Hartle and Hawking for the boundary conditions of the Universe, to models which incorporate fermions in a realistic manner. We consider inhomogeneous fermionic perturbations about a homogeneous, isotropic minisuperspace background model, by expanding the fermion fields in spinor harmonics on the spatial sections, taken to be three-spheres. The Dirac action is thus found to take the form of an infinite sum of terms, each describing a time-dependent Fermi oscillator. On quantization, we find that the Wheeler-DeWitt equation for the wave function of the Universe may be decomposed into a set of time-dependent Schrödinger equations, one for each fermion mode, and a background minisuperspace Wheeler-DeWitt equation, which includes a term in its potential describing the back reaction of the fermionic perturbations on the homogeneous modes. Our quantization procedure employs the holomorphic representation for the fermion modes, which permits them to be treated in a manner very similar to the case of bosonic perturbations considered by Halliwell and Hawking. We set initial conditions for the Schrödinger equations by applying the proposal of Hartle and Hawking that the quantum state of the Universe is defined by a path integral over compact four-metrics and regular matter fields. We find this to imply that the fermion modes start out in their ground state. Particles are created in the subsequent (inflationary) evolution, and their number, defined with respect to instantaneous Hamiltonian diagonalization, is calculated and is found to be finite. We calculate the back-reaction term and find, after regularization, that its effect is negligible. We construct a model of a fermionic particle detector and, in the case of an exact de Sitter background, examine its response to the state picked out by the Hartle-Hawking proposal. We show that it experiences a thermal spectrum at the de Sitter temperature, with a distribution of the Fermi-Dirac form, although the distribution does not have the correct density-of-states factor to be precisely Planckian.

I. INTRODUCTION

There has been considerable interest in the recent proposal of Hartle and Hawking that one may provide boundary conditions for the wave function in quantum cosmology by defining it by a path integral over compact metrics and regular matter fields.[1-4] This proposal has been applied to a number of minisuperspace models, in which the problem of quantizing the infinite number of modes of the gravitational and matter fields is drastically simplified by restricting attention to a finite number, typically the homogeneous modes, on the assumption that these will in some sense dominate. The wave functions thus obtained have been interpreted, in the classical limit, as corresponding to a superposition of solutions to the classical field equations, which involve less arbitrary parameters than the general solution. In most of the minisuperspace models considered so far, the Hartle-Hawking proposal has been seen to pick out the solutions with the most desirable cosmological properties.[3-5]

In most of these models, the matter source is taken to be a scalar field, which is reasonable since it is this type of matter field that one would expect to give the most significant dynamical effects at early times. However, most of the theories we know also involve other types of fields, such as gauge fields, antisymmetric tensor fields, and fermion fields, and thus it is of interest to extend the current repertoire of quantum cosmological models to incorporate these other types of fields in a realistic manner. In this paper, we consider the problem of including fermion fields in quantum cosmological models.

The first attempt to include fermions in quantum cosmology was by Isham and Nelson[6] and, more recently, by Christodoulakis and Zanelli.[7] These authors considered minisuperspace models involving a homogeneous fermion field in a homogeneous, isotropic universe. In such a universe, every point in space is equivalent to every other point, so that one is effectively considering a fermion field at a single point in space. The number of permissible particle states is then severely restricted by the exclusion principle—for a Dirac field, for example, there are just 16 possible states, including the vacuum state. With such a small number of fermion states, one could not claim that such models provide a physically realistic description of the Universe in which we live.

It is clear from the above that in order to include fermions in a realistic manner in quantum cosmology, it is necessary to go beyond minisuperspace to include the infinite number of inhomogeneous modes of the fields. The main subject of this paper is precisely that—to construct a quantum cosmological model with fermions which includes all the modes of the fields. Our model may be regarded as an extension of the model of Halliwell and Hawking,[8] in which the full infinite number of inhomogeneous modes of the gravitational field coupled to a scalar field were considered, to lowest nontrivial order in in-

 ©1987 The American Physical Society

homogeneous perturbations. At this order in the perturbations, the inhomogeneous modes of the bosonic fields entirely decouple from the fermion modes, so that we may consider the fermion modes separately.

We begin in Sec. II with a description of the Hamiltonian formulation of Einstein-Dirac theory. The Hamiltonian formalism for general relativity coupled to a scalar field was outlined in Ref. 8. For the Einstein-Dirac theory, however, the situation is considerably more complicated for at least three reasons. Firstly, there is the necessity to work with the tetrad e^a_μ, rather than the metric. This introduces six new degrees of freedom, namely, the freedom to perform local Lorentz rotations. Associated with this extra freedom are six first-class constraints $J_{ab} \approx 0$, where J_{ab} are the Lorentz generators. The second complication is due to the fact that the Dirac action involves derivative couplings. The Dirac field couples to the tetrad through the spin connection ω^{ab}_μ, which occurs in the covariant derivative on spinors, $D_\mu \psi$. This has the effect of altering the relationship between the velocities and the momenta in the Hamiltonian formalism. The third complication is due to the fact that the Dirac action is first order in time derivatives of the fermion field, and hence involves second-class constraints which must be dealt with by the Dirac procedure. Our treatment of the Hamiltonian formalism follows principally that of Nelson and Teitelboim.[9]

We wish our model to describe fermionic perturbations about a homogeneous isotropic minisuperspace model. A particular model is described in Sec. III: namely, Hawking's massive scalar field model.[3,4,10] It consists of a $k=+1$ Robertson-Walker metric described by a single scale factor e^α and driven by a homogeneous massive scalar field ϕ. The wave function for this model corresponds, in the classical limit, to a set of classical solutions which have an initial inflation and then go over to a matter-dominated phase. They research a maximum size and then undergo recollapse. It is to be emphasized, however, that the main results of the perturbed model do not depend in a crucial way on our particular choice of background.

In Sec. IV, we describe the inclusion of fermions in the model. The Dirac action is expanded in spinor harmonics on the spatial sections, which are three spheres, and the Hamiltonian is derived in terms of the coefficients of the harmonics. The Wheeler-DeWitt equation for the wave function of the model is given in Sec. V and is found to involve an infinite sum of Hamiltonian operators, each of which describes a time-dependent Fermi oscillator. The eigenstates of these Hamiltonians are derived and their interpretation as particle states is discussed, using instantaneous Hamiltonian diagonalization. Our quantization procedure employs the holomorphic representation, in which the wave functions for the fermion field modes are analytic functions of odd elements of a Grassmann algebra.[11,12]

In Sec. VI, it is shown that the Wheeler-DeWitt equation may be approximated by a set of time-dependent Schrödinger equations, one for each fermion mode, and a minisuperspace Wheeler-DeWitt equation, which differs from that of Sec. III by an extra term in its potential which represents the back reaction of the perturbations. This result was also given in Ref. 8, but is derived more carefully here, since we are interested in a more detailed treatment of the back reaction and the handling of divergences.

Boundary conditions for the Wheeler-DeWitt equation are considered in Sec. VII. We adopt the Hartle-Hawking proposal, which demands that one sum over all compact four-metrices and regular matter fields which match prescribed values on a given three-surface. In the semiclassical approximation, this involves solving the Euclidean-Dirac equation on the interior of a three-sphere, subject to the field matching prescribed values on the three-sphere boundary. This boundary-value problem is more involved than in the bosonic case, and is discussed in some detail. It is shown that the Hartle-Hawking proposal implies that the fermion modes start out in their ground state, as was found also to be the case for the bosonic modes in Ref. 8, and this constitutes our main result.

This result is used as the initial condition for the Schrödinger equations in Sec. VIII, which are evolved to obtain the wave functions of the perturbation modes at the end of the inflationary phase. The particle creation, defined with respect to instantaneous Hamiltonian diagonalization, is calculated and is found to be finite.

The back reaction of the perturbation modes on the behavior of the minisuperspace background is considered in Sec. IX, both in the massive and massless cases. The expression for this back reaction is formally divergent and is therefore replaced by its regularized value. The back reaction is found to be negligible, in both massless and massive cases. It is shown that in the massive case, the result is what one would expect, given the number of particles created.

In Sec. X, we consider a very different notion of a particle. Restricting attention to the approximation of quantized fermion fields on an exact classical de Sitter background, we consider a model of a particle detector designed to detect fermions and examine its response to the state picked out by the Hartle-Hawking proposal. The detector response depends on the properties of the Green's functions in this state. It is argued that the state and hence the Green's functions are de Sitter invariant, thus allowing us to obtain expressions for these Green's functions in closed form as a function only of the geodesic distance between the two points. The Green's functions are shown to be antiperiodic in imaginary time, and have a pole only when the two points are connected by a null geodesic, not when they are antipodal. We thus show that the spectrum measured by the detector is thermal, with the de Sitter temperature and with the correct Fermi-Dirac distribution, but does not have the correct density-of-states factor to be exactly of the Planck form. Our conclusions are presented in Sec. XI.

II. EINSTEIN-DIRAC THEORY: ACTION AND HAMILTONIAN FORM

We begin with the action of Einstein-Dirac theory, constructed from even (commuting) gravitational variables and odd (anticommuting Grassmann) fermion fields.

Since the spinor harmonics on S^3, into which the spinor fields will be decomposed, have definite chirality, we use two component spinors. The spinor conventions are those of D'Eath,[13] outlined in Appendix A. In addition to the gravitational and fermion fields, the full action of our model also involves a scalar field. For simplicity of exposition we do not include the scalar field in this section. Only its homogeneous mode is needed for the background model, described by the simple Hamiltonian formalism of Sec. III. Its inhomogeneous perturbations, while coupled to gravity, do not couple at lowest order to the fermionic perturbations studied subsequently.

The Lorentzian action of Einstein-Dirac theory has the form

$$I = I_V + I_B \ , \tag{2.1}$$

the sum of a volume contribution I_V and a boundary contribution I_B. I_V is taken to be

$$I_V = \frac{1}{2\kappa^2}\int d^4x\, eR - \frac{i}{2}\int d^4x\, e(\bar{\phi}^{A'} e^{\mu}_{AA'} D_\mu \phi^A + \bar{\chi}^{A'} e^{\mu}_{AA'} D_\mu \chi^A) + \text{H.c.} - \frac{m}{\sqrt{2}}\int d^4x\, e(\chi_A \phi^A + \bar{\phi}^{A'}\bar{\chi}_{A'}) \ . \tag{2.2}$$

We use units in which $\hbar = c = 1$ and $\kappa^2 = 8\pi G$. The gravitational field is described by the tetrad e^a_μ, where $a,b,\ldots = 0,1,2,3$ are tetrad indices and $\mu,\nu,\ldots = 0,1,2,3$ are world indices, or equivalently by the Hermitian spinor-valued forms $e^{AA'}_\mu$. Unprimed spinor indices $A,B,\ldots$ take values 0,1 and primed spinor indices A',B' take values $0',1'$. Here $e = \det(e^a_\mu)$ and R is the Ricci scalar of the metric $g_{\mu\nu} = \eta_{ab} e^a_\mu e^b_\nu$, where η_{ab} is the Minkowski metric. The Dirac field is described by the pair of spinor fields ϕ^A, χ^A with Hermitian conjugates $\bar{\phi}^{A'}$, $\bar{\chi}^{A'}$. Derivatives such as $D_\mu \phi^A$ are defined by

$$D_\mu \phi^A = \partial_\mu \phi^A + \omega^A_{\mu B} \phi^B \ , \tag{2.3}$$

where the connection forms ω^{ab}_μ are expanded spinorially as

$$\omega^{AA'BB'}_\mu = \omega^{AB}_\mu \epsilon^{A'B'} + \bar{\omega}^{A'B'}_\mu \epsilon^{AB} \tag{2.4}$$

with $\omega^{AB}_\mu = \omega^{(AB)}_\mu$ and its conjugate $\bar{\omega}^{A'B'}_\mu = \bar{\omega}^{(A'B')}_\mu$.

Boundary contributions I_B to the action must be included whenever boundary surfaces are present. Classically, they are needed in order that the variational condition $\delta I = 0$ should lead to the correct classical solution when data are suitably posed on the boundaries. The form of I_B depends on the choice of boundary data. Suppose, for example, that the boundary consists of an initial surface S_I and final surface S_F. As can be seen from the Hamiltonian formulation below, one natural choice of boundary conditions is to specify the spatial tetrad variables $e_i^{AA'}$ ($i,j,\ldots = 1,2,3$) together with half of the fermion variables ϕ^A, χ^A, $\tilde{\phi}^{A'}$, $\tilde{\chi}^{A'}$ on each surface. In this context, we free $\tilde{\phi}^{A'}$ and $\tilde{\chi}^{A'}$ from the restriction of being the Hermitian conjugates of ϕ^A, χ^A, and use a modified notation with a tilde instead of a bar. If, say, we specify $e_i^{AA'}$, $\tilde{\phi}^{A'}$, $\tilde{\chi}^{A'}$ on S_I and $e_i^{AA'}$, χ^A, ϕ^A on S_F then the boundary contribution is

$$I_B = \frac{1}{\kappa^2}\left[\int_{S_F} - \int_{S_I}\right] d^3x\, h^{1/2}\text{tr}K + \frac{i}{2}\left[\int_{S_F} + \int_{S_I}\right] d^3x\, h^{1/2}(\tilde{\phi}^{A'} n_{AA'}\phi^A + \tilde{\chi}^{A'} n_{AA'}\chi^A) \ . \tag{2.5}$$

Here, $h = \det(h_{ij})$, $h_{ij} = -e_{AA'i} e_j^{AA'}$ is the spatial metric, and $\text{tr}K = h_{ij}K^{ij}$, where K_{ij} is the second fundamental form (extrinsic curvature) of the surface. Also, $n^{AA'}$ is the spinor version of the unit timelike future-directed normal n^μ to the surface. It is determined by the $e_i^{AA'}$ through the relations

$$n_{AA'} e_i^{AA'} = 0, \quad n_{AA'} n^{AA'} = 1 \ . \tag{2.6}$$

The variational condition $\delta I = 0$, subject to the specified data, leads to the Dirac equations

$$e^{\mu}_{AA'} D_\mu \phi^A = i\frac{m}{\sqrt{2}}\tilde{\chi}_{A'} \ , \tag{2.7}$$

$$e^{\mu}_{AA'} D_\mu \chi^A = i\frac{m}{\sqrt{2}}\tilde{\phi}_{A'} \ , \tag{2.8}$$

$$e^{\mu}_{AA'} D_\mu \tilde{\phi}^{A'} = -i\frac{m}{\sqrt{2}}\chi_A \ , \tag{2.9}$$

$$e^{\mu}_{AA'} D_\mu \tilde{\chi}^{A'} = -i\frac{m}{\sqrt{2}}\phi_A \ , \tag{2.10}$$

together with the Einstein equations with the energy-momentum tensor formed from the Dirac field. Since ϕ^A,χ^A and $\tilde{\phi}^{A'},\tilde{\chi}^{A'}$ are no longer Hermitian conjugates, the classical solution $e^{AA'}_\mu$ will in general no longer be Hermitian. This feature is expected even in the absence of fermion fields, since the classical solution for the metric $g_{\mu\nu} = e^a_\mu e^b_\nu \eta_{ab}$ may not be Lorentzian, but rather Euclidean or complex corresponding to complex e^a_μ.

If instead we had specified the variables $e_i^{AA'}$, ϕ^A, χ^A on S_I and $e_i^{AA'}$, $\tilde{\phi}^{A'}$, $\tilde{\chi}^{A'}$ on S_F, then the fermionic boundary condition in (2.5) would appear with the opposite sign. Generally, a change in the chirality of a variable being specified results in a sign change in the corresponding boundary condition in the action. The example of boundary data given here results from a local splitting of the complete fermionic data ϕ^A, χ^A, $\tilde{\phi}^{A'}$, $\tilde{\chi}^{A'}$. However, it will be most natural when considering the wave function of the Universe and specifying fermionic data on a single surface, to use a nonlocal decomposition based on the eigenfunction expansion of spinor fields on S^3.

The boundary contributions I_B are also needed to obtain the correct path-integral expressions for quantum amplitudes. For example, following the path-integral treatment of fermions by Faddeev and Slavnov,[12] the amplitude to go from initial data ($e_i^{AA'}, \tilde{\phi}^{A'}, \tilde{\chi}^{A'}$) specified on S_I to final data ($e_i^{AA'}, \phi^A, \chi^A$) specified on S_F is given formally by

$$K(e_F,\tilde{\phi}_F,\tilde{\chi}_F;e_I,\phi_I,\chi_I) = \int \exp(iI)d[e]d[\phi]d[\tilde{\phi}]d[\chi]d[\tilde{\chi}] \; . \tag{2.11}$$

The action I is given by Eqs. (2.1), (2.2), (2.5), and the infilling fields $e_\mu^{AA'}$, ϕ^A, $\tilde{\phi}^{A'}$, χ^A, $\tilde{\chi}^{A'}$ in the path integral should agree with the given data on the boundaries. The path integral can be regarded as a contour integral over $e_\mu^{AA'}(x)$, and the contour deformed such that the $e_\mu^{AA'}(x)$ describe positive-definite metrics $g_{\mu\nu}$, in the case that the boundary data $e_i^{AA'}$ describe positive-definite three-metrics h_{ij} on S_I and S_F. Strictly, the path integral should also include gauge-fixing and ghost terms. Provided that the boundary terms I_B are included, the theory will have the property that the amplitude to go from specified data on S_I to specified data on S_F can be recovered by inserting an intermediate surface S_J and composing the amplitude to go from data on S_I to data on S_J with the amplitude to go from data on S_J to data on S_F, summing over a complete set of states on S_J.

In describing the Hamiltonian formulation of Einstein-Dirac theory, we follow principally the treatment of Nelson and Teitelboim.[9] The gravitational variables are split into $e_0^{AA'}$ and $e_i^{AA'}$, when we decompose the action (2.1) with respect to a family of surfaces $t=x^0=\text{const}$. Spatial indices are lowered and raised using the spatial metric h_{ij} and its inverse h^{ij}. In particular, we use the notation $e^{AA'i}=h^{ij}e_j^{AA'}$. A basis for the space of spinors with one unprimed and one primed index is given by the $e_i^{AA'}$ together with the normal spinor $n^{AA'}$. The variables $e_0^{AA'}$ can be expanded as

$$e_0^{AA'}=Nn^{AA'}+N^ie_i^{AA'} \; , \tag{2.12}$$

where N is the lapse function and N^i is the shift vector.

Proceeding in a standard way, one can calculate the momenta $p^i_{AA'}$ conjugate to the variables $e_i^{AA'}$, while finding that the momenta $p^0_{AA'}$ vanish. Fermionic momenta conjugate to $\phi^A,\chi^A,\tilde{\phi}^{A'},\tilde{\chi}^{A'}$ are defined by $\pi_{\phi A}=\delta I/\delta\dot{\phi}^A$, etc., with the convention that odd variables must be brought to the left using anticommutation before the functional differentiation is carried out. Since the Lagrangian is of first order in derivatives of the fermion fields, these momenta are related to the original variables:

$$\pi_{\phi A}=-\frac{i}{2}h^{1/2}n_{AA'}\tilde{\phi}^{A'} \; , \tag{2.13}$$

$$\pi_{\chi A}=-\frac{i}{2}h^{1/2}n_{AA'}\tilde{\chi}^{A'} \; , \tag{2.14}$$

while $\pi_{\tilde{\phi}A'}$ and $\pi_{\tilde{\chi}A'}$ are minus the Hermitian conjugates of $\pi_{\phi A}$ and $\pi_{\chi A}$. The basic dynamical variables in the theory can then be taken to be $e_i^{AA'}$, $p^i_{AA'}$, ϕ^A, χ^A, $\tilde{\phi}^{A'}$, and $\tilde{\chi}^{A'}$. These are related by the primary constraints

$$J_{AB}\approx 0, \quad \bar{J}_{A'B'}\approx 0 \tag{2.15}$$

as a consequence of the invariance of the Lagrangian under local Lorentz transformations. Here

$$J_{AB}=e_{(A}{}^{A'i}p_{B)A'i}+\phi_{(A}\pi_{\phi B)}+\chi_{(A}\pi_{\chi B)} \tag{2.16}$$

and $\bar{J}_{A'B'}$ is its Hermitian conjugate. Once the fermionic momentum variables are eliminated through the second-class constraints (2.14) and corresponding equations for the Hermitian conjugates, the original Poisson brackets among the dynamical variables must be replaced by suitable Dirac brackets.

The Hamiltonian has the form

$$H=\int d^3x(N\mathcal{H}_*+N^i\mathcal{H}_i+M_{AB}J^{AB}+\bar{M}_{A'B'}\bar{J}^{A'B'}) \; . \tag{2.17}$$

Here the quantities N,N^i together with $M_{AB}=M_{(AB)}$, $\bar{M}_{A'B'}=\bar{M}_{(A'B')}$ introduced for the primary constraints (2.15), occur as Lagrange multipliers, freely specified during dynamical evolution. Together they specify the amount of displacement applied normally and tangentially and the amount of local Lorentz rotation applied to dynamical data per unit time.

The generators $\mathcal{H}_*$ and $\mathcal{H}_i$ are

$$\begin{aligned}\mathcal{H}_* &= 2\kappa^2h^{-1/2}[\pi_{ij}\pi^{ij}-\tfrac{1}{2}(\text{tr}\pi)^2]-\frac{1}{2\kappa^2}h^{1/2}\,{}^3R \\ &\quad +\frac{i}{2}h^{1/2}e^i_{AA'}(\tilde{\phi}^{A'(3)}D_i\phi^A+\tilde{\chi}^{A'(3)}D_i\chi^A)+\text{H.c.} \\ &\quad +\frac{m}{\sqrt{2}}h^{1/2}(\chi_A\phi^A+\tilde{\phi}^{A'}\tilde{\chi}_{A'}) \; ,\end{aligned} \tag{2.18}$$

$$\begin{aligned}\mathcal{H}_i &= -2h_{ij}\,{}^{(3)}\nabla_k\pi^{jk} \\ &\quad +\frac{i}{2}h^{1/2}n_{AA'}(\tilde{\phi}^{A'(3)}D_i\phi^A+\tilde{\chi}^{A'(3)}D_i\chi^A)+\text{H.c.} \\ &\quad +\tfrac{1}{4}h_{ik}\partial_j[(\tilde{\phi}^{A'}\phi^A+\tilde{\chi}^{A'}\chi^A)\epsilon^{kjl}e_{AA'l}] \; .\end{aligned} \tag{2.19}$$

Here

$$\pi^{ij}=\tfrac{1}{2}e^{AA'(i}p_{AA'}{}^{j)}=-\frac{h^{1/2}}{2\kappa^2}(K^{ij}-h^{ij}\text{tr}K) \tag{2.20}$$

and $\text{tr}\pi=h_{ij}\pi^{ij}$. The Ricci scalar of the three-dimensional metric h_{ij} is denoted by ${}^3R$. The spatial spinor derivative is given, e.g., by

$${}^{(3)}D_i\phi^A=\partial_i\phi^A+{}^{(3)}\omega^A_{iB}\phi^B \; , \tag{2.21}$$

where the spatial connection forms ${}^3\omega_i^{AB}$ (without torsion) are given in Refs. 9 and 13. The spatial covariant derivative on tensor densities is denoted by ${}^{(3)}\nabla_i$. Classically the generators $\mathcal{H}_*$ and $\mathcal{H}_i$ vanish:

$$\mathcal{H}_*\approx 0, \quad \mathcal{H}_i\approx 0 \; . \tag{2.22}$$

The form of the theory is considerably simplified by working in the time gauge. In which the condition

$$e_i^0=0 \tag{2.23}$$

is imposed on the tetrad components e_i^a, leaving only the triad variables e_i^α (with greek letters $\alpha,\beta,\ldots$ from the beginning of the alphabet denoting triad indices, taking values 1,2,3). Equivalently, $n^a=\delta_0^a$ in this gauge. The conjugate momenta p_0^i are also eliminated by being set to zero:

$$p_0^i=0 \; . \tag{2.24}$$

This eliminates the generators $J_{*i}=n^ae_i^bJ_{ab}$ of Lorentz boosts, where $J_{ab}\leftrightarrow J_{AB}\epsilon_{A'B'}+\bar{J}_{A'B'}\epsilon_{AB}$ are the generators of local Lorentz transformations, leaving only the freedom of local triad rotations with generators $J_{\alpha\beta}$ or J_{ij}. Classically, these generators still vanish

$$J_{\alpha\beta}\approx 0 \ . \tag{2.25}$$

In terms of the variables e_i^α, their conjugate momenta p^i_α and the rescaled fermionic variables

$$\psi^A = h^{1/4}\phi^A, \ \ \lambda^A = h^{1/4}\chi^A \ , \tag{2.26}$$

together with $\bar\psi^{A'}$ and $\bar\lambda^{A'}$, the only nonzero brackets are

$$[e_i^\alpha(x), p^j_\beta(x')]^* = \delta^\alpha_\beta \delta^j_i \delta(x,x') \ , \tag{2.27}$$

$$[\psi^A(x), \bar\psi^{A'}(x')]^* = -2in^{AA'}\delta(x,x') \ , \tag{2.28}$$

$$[\lambda^A(x), \bar\lambda^{A'}(x')]^* = -2in^{AA'}\delta(x,x') \ . \tag{2.29}$$

The Hamiltonian is still given by Eq. (2.17), with e_i^0 and p^i_0 set to zero.

The theory may be quantized in the time gauge by starting from a maximal (anti)commuting set of variables, for example, $e_i^\alpha(x)$, $\psi^A(x)$, and $\lambda^A(x)$, and representing states by Grassmann-valued wave functionals $\Psi(e_i^\alpha(x), \psi^A(x), \lambda^A(x))$. Dirac brackets become (anti)commutators, following the rules given by Casalbuoni[14] for systems including fermions, and the remaining variables p^i_α, $\bar\psi^{A'}$, and $\bar\lambda^{A'}$ can then be represented by the momentumlike operators

$$p^i_\alpha(x) \to -i\frac{\delta}{\delta e_i^\alpha(x)} \ , \tag{2.30}$$

$$\bar\psi^{A'}(x) \to 2n^{AA'}\frac{\delta}{\delta\psi^A(x)} \ , \tag{2.31}$$

$$\bar\lambda^{A'} \to 2n^{AA'}\frac{\delta}{\delta\lambda^A(x)} \ . \tag{2.32}$$

A formal inner product between wave functions can be found such that $p^i_\alpha(x)$ is Hermitian and such that $\bar\psi^{A'}(x)$. $\bar\lambda^{A'}(x)$ are the Hermitian adjoints of $\psi^A(x)$, $\lambda^A(x)$.

In the quantum theory, the classical constraints (2.22) and (2.25) becomes constraints on physically allowed wave functionals

$$\mathscr{H}_*\Psi = 0 \ , \tag{2.33}$$

$$\mathscr{H}_i\Psi = 0 \ , \tag{2.34}$$

$$J_{\alpha\beta}\Psi = 0 \ . \tag{2.35}$$

For the constraints (2.34) and (2.35) which are only linear in "momentum" operators, it is natural to choose a factor ordering such that "momentum" operators appear on the right in $\mathscr{H}_i$ and $J_{\alpha\beta}$. Then Eqs. (2.34) and (2.35) can be shown to imply that Ψ is invariant under spatial coordinate transformations and local rotations applied to its arguments e_i^α, ψ^A, and λ^A. Corresponding properties will hold in other representations. It is unlikely that any factor ordering can be found for the operator $\mathscr{H}_*$ which involves momenta quadratically, such that the complete set of constraint generators $\mathscr{H}_*$, $\mathscr{H}_i$, and $J_{\alpha\beta}$ form a closed algebra. This difficulty may possibly only be overcome in supergravity models. The factor ordering in $\mathscr{H}_*$ will, however, not make a substantial difference to the results of this paper, to the order of approximation to which we work and will be chosen later for convenience.

In the application to quantum cosmology considered here, the action and Hamiltonian can be expanded out to quadratic order in perturbations of the gravitational field about a Friedmann metric with S^3 spatial sections, and to quadratic order in fermionic variables. At this order the bosonic perturbations and fermion fields are noninteraction and can be treated separately. The wave function for the bosonic perturbations at this order will be that found in Ref. 8. If we fix the gauge freedom of the remaining SO(3) local rotations. This can be done by taking a normalized left-invariant basis E_i^α of one-forms on S^3, such that $E_i^\alpha E_{\alpha j}$ gives the metric of a unit sphere, and writing

$$e_i^\alpha = a(E_i^\alpha + \epsilon_i^\alpha) \ , \tag{2.36}$$

where a is the radius of the background three-sphere, and ϵ_i^α are triad perturbations. If we impose the gauge condition

$$E_i^\alpha \epsilon_{\alpha j} - E_j^\alpha \epsilon_{\alpha i} = 0 \ , \tag{2.37}$$

then the constraints $J_{\alpha\beta}$ become second class. After elimination of these constraints the gravitational variables can be taken to be h_{ij} and π^{ij}, while the same fermionic variables can be used. At lowest order the only nonzero Dirac brackets are those expected between h_{ij} and π^{ij}, together with the preceding fermionic brackets (2.28) and (2.29). Thus, it is sufficient to study only the fermionic perturbations, as will be done in the rest of this paper.

III. THE BACKGROUND MINISUPERSPACE MODEL

In this section, we describe the background minisuperspace model about which we wish to do fermionic perturbations. This is Hawking's massive scalar field model.[3,4,10]

The metric is taken to be

$$ds^2 = \sigma^2(-N^2(t)dt^2 + e^{2\alpha(t)}d\Omega_3^2) \ , \tag{3.1}$$

where $d\Omega_3^2$ is the metric on the unit three-sphere, and $\sigma^2 = 2/(3\pi m_p^2)$. The model involves a massive scalar field $(2\pi^2\sigma^2)^{-1/2}\phi$, of mass $\sigma^{-1}M$ which is taken to be homogeneous, $\phi=\phi(t)$. The action is

$$I = -\frac{1}{2}\int dt\, Ne^{3\alpha}\left[\frac{\dot\alpha^2}{N^2} - e^{-2\alpha} - \frac{\dot\phi^2}{N^2} + M^2\phi^2\right] \tag{3.2}$$

from which one may derive the field equations

$$N\frac{d}{dt}\left[\frac{\dot\alpha}{N}\right] + 3\dot\phi^2 - N^2e^{-2\alpha} = 0 \ , \tag{3.3}$$

$$N\frac{d}{dt}\left[\frac{\dot\phi}{N}\right] + 3\dot\alpha\dot\phi + N^2M^2\phi = 0 \ , \tag{3.4}$$

and the constraint

$$\frac{\dot\alpha^2}{N^2} + e^{-2\alpha} - \frac{\dot\phi^2}{N^2} - M^2\phi^2 = 0 \ . \tag{3.5}$$

The Hamiltonian is

$$H_0 = \tfrac{1}{2}Ne^{-3\alpha}(-\pi_\alpha^2 + \pi_\phi^2 + e^{6\alpha}M^2\phi^2 - e^{4\alpha}) \ . \tag{3.6}$$

The classical Hamiltonian constraint is $H_0=0$.

The Hamiltonian constraint is quantized to yield the Wheeler-DeWitt equation

$$H_0\Psi=\tfrac{1}{2}Ne^{-3\alpha}\left[\frac{\partial^2}{\partial\alpha^2}-\frac{\partial^2}{\partial\phi^2}+V(\alpha,\phi)\right]\Psi(\alpha,\phi)=0\ , \quad (3.7)$$

where

$$V=e^{6\alpha}M^2\phi^2-e^{4\alpha}\ . \quad (3.8)$$

This is a hyperbolic equation on the two-dimensional minisuperspace with metric

$$ds^2=f_{ab}dq^a dq^b=e^{3\alpha}(-d\alpha^2+d\phi^2)\ . \quad (3.9)$$

Here, we have introduced the notation q^a and f_{ab} for the minisuperspace coordinates and metric to emphasize that the results of the perturbed model, described in the following sections, will also be true of a number of such models with different minisuperspace backgrounds, and are not sensitive to the details of this particular choice of background.

The solution to (3.8) picked out by the Hartle-Hawking proposal is described in Refs. 3, 4, 10, and 15. The wave function is exponential in behavior for $V<0$, and is then interpreted as corresponding to a Euclidean four-geometry in the classical limit. This region will thus be referred to as the Euclidean region, and may be thought of as a classically forbidden region. The wave function is oscillatory throughout most of the region $V>0$ and may be interpreted as corresponding to a Lorentzian four geometry in the classical limit. This will be referred to as the Lorentzian region. In this region, one may use the WKB approximation, in which one writes

$$\Psi=\mathrm{Re}(Ce^{iS})\ , \quad (3.10)$$

where S is a rapidly varying phase and C is a slowly varying amplitude. S obeys the Hamilton-Jacobi equation corresponding to (3.7), which has the approximate solution

$$S\approx-\frac{1}{3M^2\phi^2}(e^{2\alpha}M^2\phi^2-1)^{3/2} \quad (3.11)$$

for $|\phi|\gg 1$. An equation for C may be derived, but C will not be required.

In the classical limit, the wave function (3.10) may be interpreted as corresponding to a set of trajectories $q^a(t)$, which are the integral curves of the vector field

$$\frac{1}{N}\frac{d}{dt}=\nabla S\cdot\nabla \quad (3.12)$$

the dot product being with respect to the metric f_{ab}. These trajectories are solutions to the field equations (3.3)–(3.5) parametrized by the coordinate time t. They begin at a minimum size $e^\alpha=1/M\phi$ and then undergo a long inflation with $e^\alpha=(M\phi)^{-1}\cosh(M\phi t)$ while ϕ remains approximately constant. The solution then goes over to a matter-dominated phase with e^α proportional to $t^{2/3}$, and ϕ oscillates about zero with frequency m. The solution subsequently reaches a maximum size and then collapses in a similar manner.

IV. EXPANSION IN HARMONICS

In this section we describe the fermionic perturbations about the background minisuperspace model of the preceding section. One may construct a complete set of spinor harmonics, $\rho_A^{np}(\mathbf{x})$, $\bar{\sigma}_A^{np}(\mathbf{x})$, and $\bar{\rho}_{A'}^{np}(\mathbf{x})$, $\sigma_{A'}^{np}(\mathbf{x})$ for the expansion of any spinor field and its Hermitian conjugate on the three-sphere. The construction and properties of these harmonics are described in Appendix B. Since the boundary conditions for the spinor fields must be given in terms of the weighted fields, it is most convenient to expand these in harmonics, rather than their unweighted counterparts. One may thus write

$$\phi_A=\frac{e^{-3\alpha/2}}{2\pi}\sum_{np}\sum_q \alpha_n^{pq}[m_{np}(t)\rho_A^{nq}(\mathbf{x})+\bar{r}_{np}(t)\bar{\sigma}_A^{nq}(\mathbf{x})]\ , \quad (4.1)$$

$$\bar{\phi}_{A'}=\frac{e^{-3\alpha/2}}{2\pi}\sum_{np}\sum_q \alpha_n^{pq}[\bar{m}_{np}(t)\bar{\rho}_{A'}^{nq}(\mathbf{x})+r_{np}(t)\sigma_{A'}^{nq}(\mathbf{x})]\ , \quad (4.2)$$

$$\chi_A=\frac{e^{-3\alpha/2}}{2\pi}\sum_{np}\sum_q \beta_n^{pq}[s_{np}(t)\rho_A^{nq}(\mathbf{x})+\bar{t}_{np}(t)\bar{\sigma}_A^{nq}(\mathbf{x})]\ , \quad (4.3)$$

$$\bar{\chi}_{A'}=\frac{e^{-3\alpha/2}}{2\pi}\sum_{np}\sum_q \beta_n^{pq}[\bar{s}_{np}(t)\bar{\rho}_{A'}^{nq}(\mathbf{x})+t_{np}(t)\sigma_{A'}^{nq}(\mathbf{x})]\ , \quad (4.4)$$

where we have included the weight factor $e^{-3\alpha/2}$ and

$$\sum_{np}=\sum_{n=0}^{\infty}\sum_{p=1}^{(n+1)(n+2)}\ . \quad (4.5)$$

The label q runs over the same range as p. The time-dependent coefficients m_{np}, r_{np}, t_{np}, s_{np}, and their complex conjugates $\bar{m}_{np}$, $\bar{r}_{np}$, $\bar{t}_{np}$, $\bar{s}_{np}$, are taken to be odd elements of a Grassmann algebra. The constant coefficients α_n^{pq}, β_n^{pq} have been included for convenience, to avoid couplings between different values of p in the expansion of the action. They are given in Appendix B.

The harmonics are eigenfunctions of the Dirac operator on the three-sphere, $n_{AB'}e^{BB'j\,(3)}D_j$. We have used a notation in which the unbarred harmonics have positive eigenvalues and the barred harmonics have negative eigenvalues. Associated with each unbarred harmonic in (4.1)–(4.4) is an unbarred coefficient and associated with each barred harmonic is a barred coefficient; thus, the unbarred coefficients correspond to the positive half of the spectrum of the operator $n_{AB'}e^{BB'j\,(3)}D_j$ and the barred coefficients correspond to the negative half of the spectrum. As will be discussed in Sec. VII, it is convenient to express the boundary conditions for the fermion fields in terms of the positive and negative halves of the spectrum of the Dirac operator on the bounding surfaces—hence, our choice of definition of the barred and unbarred variables in (4.1)–(4.4). In what follows, we shall assume that the barred variables $\bar{m}_{np}$, $\bar{s}_{np}$, $\bar{t}_{np}$, $\bar{r}_{np}$ are fixed on the initial surface S_I, and the unbarred variables m_{np}, s_{np}, t_{np}, r_{np} are fixed on the final surface S_F.

Inserting (4.1)–(4.4) into the total action (2.1) and using the properties of the spinor harmonics, one finds that the action is

$$I = I_0 + I_f + I_{fB} \ , \tag{4.6}$$

where I_0 is the action (3.2) of the background model. The fermion volume part I_f may be written

$$I_f = I_f^{(1)}[m,s] + I_f^{(2)}[t,r] \ , \tag{4.7}$$

where

$$I_f^{(1)} = \sum_{np} \int dt\, N \left[\frac{i}{2N}(\bar{m}_{np}\dot{m}_{np} + m_{np}\dot{\bar{m}}_{np} + \bar{s}_{np}\dot{s}_{np} + s_{np}\dot{\bar{s}}_{np}) + e^{-\alpha}(n+\tfrac{3}{2})(\bar{m}_{np}m_{np} + \bar{s}_{np}s_{np}) - m(s_{np}m_{np} + \bar{m}_{np}\bar{s}_{np}) \right] \tag{4.8}$$

and

$$I_f^{(2)} = \sum_{np} \int dt\, N \left[\frac{i}{2N}(\bar{t}_{np}\dot{t}_{np} + t_{np}\dot{\bar{t}}_{np} + \bar{r}_{np}\dot{r}_{np} + r_{np}\dot{\bar{r}}_{np}) + e^{-\alpha}(n+\tfrac{3}{2})(\bar{t}_{np}t_{np} + \bar{r}_{np}r_{np}) - m(r_{np}t_{np} + \bar{t}_{np}\bar{r}_{np}) \right] . \tag{4.9}$$

The boundary term appropriate to the boundary conditions described above is

$$I_{fB} = \frac{i}{2}\sum_{np}(\bar{m}_{np}m_{np} + \bar{s}_{np}s_{np} + \bar{t}_{np}t_{np} + \bar{r}_{np}r_{np})_{S_F} + \frac{i}{2}\sum_{np}(\bar{m}_{np}m_{np} + \bar{s}_{np}s_{np} + \bar{t}_{np}t_{np} + \bar{r}_{np}r_{np})_{S_I} \ . \tag{4.10}$$

The fermion action, including boundary terms, is thus a sum over n,p of actions, each of the form

$$I_n[x,\bar{x},y,\bar{y}] = \int dt\, N \left[\frac{i}{2N}(\bar{x}\dot{x} + x\dot{\bar{x}} + \bar{y}\dot{y} + y\dot{\bar{y}}) + e^{-\alpha}(n+\tfrac{3}{2})(\bar{x}x + \bar{y}y) - m(yx + \bar{x}\bar{y}) \right] + \frac{i}{2}(\bar{x}x + \bar{y}y)_{S_F} + \frac{i}{2}(\bar{x}x + \bar{y}y)_{S_I} \ . \tag{4.11}$$

In this expression, and in what follows, the Grassmann variables x and y will be used generically, to denote m_{np} and s_{np}, respectively, or to denote t_{np} and r_{np}, respectively. The total fermion action may thus be expressed in terms of I_n:

$$I_f + I_{fB} = \sum_{np} [I_n(m_{np},\bar{m}_{np},s_{np},\bar{s}_{np}) + I_n(t_{np},\bar{t}_{np},r_{np},\bar{r}_{np})] \ . \tag{4.12}$$

The action (4.11) describes two time-dependent Fermi oscillators $x,\bar{x}$ and $y,\bar{y}$ which couple together through the mass term. The problem of solving our quantum cosmological model is thus essentially reduced to that of solving this simple system.

From (4.11), the following field equations may be derived:

$$i\frac{\dot{x}}{N} + \nu x - m\bar{y} = 0 \ , \tag{4.13}$$

$$i\frac{\dot{\bar{x}}}{N} - \nu\bar{x} + my = 0 \ , \tag{4.14}$$

$$i\frac{\dot{y}}{N} + \nu y + m\bar{x} = 0 \ , \tag{4.15}$$

$$i\frac{\dot{\bar{y}}}{N} - \nu\bar{y} - mx = 0 \ , \tag{4.16}$$

where $\nu = e^{-\alpha}(n+\frac{3}{2})$. These are of course, just the components of the Dirac equations (2.7)—(2.10), when expanded in harmonics. It is also convenient to have a second-order form for these equations. One finds that x and y obey the same second-order equation, which is

$$\frac{1}{N}\frac{d}{dt}\left[\frac{\dot{x}}{N}\right] + \left[\frac{\dot{\nu}}{iN} + \nu^2 + m^2\right]x = 0 \tag{4.17}$$

and $\bar{x}$ and $\bar{y}$ obey the conjugate equation

$$\frac{1}{N}\frac{d}{dt}\left[\frac{\dot{\bar{x}}}{N}\right] + \left[-\frac{\dot{\nu}}{iN} + \nu^2 + m^2\right]\bar{x} = 0 \ . \tag{4.18}$$

Following the Dirac procedure, one may obtain the Hamiltonian

$$H_n(x,\bar{x},y,\bar{y}) = N[\nu(x\bar{x} + y\bar{y}) + m(yx + \bar{x}\bar{y})] \ , \tag{4.19}$$

where $x,y,\bar{x},\bar{y}$ obey the Dirac-brackets relations

$$[x,\bar{x}]^* = -i, \quad [y,\bar{y}]^* = -i \ . \tag{4.20}$$

All other brackets relations yield zero. The total fermion Hamiltonian may thus be expressed in terms of (4.19):

$$H_f = \sum_{np} H_{np} = \sum_{np} [H_n(m_{np},\bar{m}_{np},s_{np},\bar{s}_{np}) + H_n(t_{np},\bar{t}_{np},r_{np},\bar{r}_{np})] \tag{4.21}$$

which is of course just the expansion of (2.17) in harmonics.

For the remaining bosonic variables, variation of the action (4.6) yields the field equations (3.3)—(3.5), but in the case of (3.3) and (3.5), modified on the right-hand side by terms quadratic in the fermion fields. Since the fermions are regarded as perturbations, these modifications may be ignored, to the order at which we are working. The total Hamiltonian is now the sum of the background Hamiltonian (3.6) and the fermion Hamiltonian (4.21), and vanishes,

$$H_0+H_f=0\ , \tag{4.22}$$

this being the Hamiltonian constraint.

We have replaced the problem of working with the fields $\phi_A, \chi_A, \bar{\phi}_{A'}, \bar{\chi}_{A'}$ with that of working with the coefficients m_{np}, s_{np}, etc., defined by Eqs. (4.1)–(4.4). To lowest nontrivial order in perturbations, these coefficients are invariant under local Lorentz transformations and diffeomorphisms in the three-surface. It follows that the generators of these transformations, namely, the Lorentz generators J_{ab} and the fermion part of the $\mathscr{H}_i$ play no role in this model. It is therefore sufficient to consider only the constraint (4.22), to the order at which we are working.

V. QUANTIZATION

The quantum state of our cosmological model may be be described by a wave function Ψ, which is a function of the gravitational and matter-field configurations on a given three-surface. The wave function obeys the Wheeler-DeWitt equation

$$[H_0+H_f]\Psi=0\ , \tag{5.1}$$

where H_0 is the operator of the background Wheeler-DeWitt equation (3.7). H_f is obtained from the classical fermion Hamiltonian (4.21) by replacing the dynamical variables with operators in a manner consistent with the anticommutation relations

$$\{x,\bar{x}\}=1,\quad \{y,\bar{y}\}=1\ . \tag{5.2}$$

These relations follow from the Dirac-bracket's relations (4.20) according to the usual quantization rules for anticommuting variables.[14] They are satisfied by the representation

$$\bar{x}\rightarrow\frac{\partial}{\partial x},\quad \bar{y}\rightarrow\frac{\partial}{\partial y} \tag{5.3}$$

known as the holomorphic representation.[11,12] With this choice, the wave function is a function of the background variables q^a and the unbarred variables, i.e., $\Psi=\Psi(q^a,m,s,t,r)$ where m denotes all the m_{np}, etc. This choice of representation is consistent with the formalism developed so far, in which we have assumed that the barred variables are fixed on S_I and the unbarred variables are fixed on S_F, bearing in mind that we eventually wish to calculate the wave function by a path integral over a set of paths ending at the point specified by its argument, (q^a,m,s,t,r).

There is the usual operator-ordering ambiguity in going from the classical constraint (4.22) to the Wheeler-DeWitt equation (5.1). For the background terms, this will not affect the results presented here, so it will be chosen for ease of calculation. For the fermionic terms, there is an ambiguity for terms of the form $x\bar{x}$. For such terms, we shall adopt the so-called Weyl ordering,[16,17] which in this case involves the substitution

$$x\bar{x}\rightarrow\frac{1}{2}\left[x\frac{\partial}{\partial x}-\frac{\partial}{\partial x}x\right]\ . \tag{5.4}$$

With the above choice of operator ordering, the Hamiltonian (4.19) becomes the operator

$$H_n=N\left[-\nu+\nu\left[x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}\right]+m\left[yx+\frac{\partial^2}{\partial x\,\partial y}\right]\right]\ . \tag{5.5}$$

It is useful to determine the eigenstates of the operator $N^{-1}H_n$. The eigenvalue equation $N^{-1}H_n\psi=E\psi$ is easily solved by expanding ψ in the Grassmann numbers x,y. One thus obtains the following eigenstates and eigenvalues:

$$\psi^{(0)}=N_0\left[1+\frac{m}{(\nu+\omega)}xy\right],\quad E_0=-\omega\ , \tag{5.6}$$

$$\psi^{(1)}=N_1x,\quad E_1=0\ , \tag{5.7}$$

$$\psi^{(2)}=N_2y,\quad E_2=0\ , \tag{5.8}$$

$$\psi^{(3)}=N_3\left[1+\frac{m}{(\nu-\omega)}xy\right],\quad E_3=+\omega\ , \tag{5.9}$$

where $\omega=(\nu^2+m^2)^{1/2}$ and $N_0,\ldots,N_3$ are normalization factors. These eigenstates are orthogonal with respect to the inner product appropriate to the holomorphic representation, defined for any pair of functions f,g of the Grassmann variables x,y (Refs. 11 and 12):

$$(f,g)=\int\overline{f(x,y)}g(x,y)e^{-x\bar{x}-y\bar{y}}dx\,d\bar{x}\,dy\,d\bar{y}\ . \tag{5.10}$$

Integration over $x,y,\bar{x},\bar{y}$ is performed according to the usual rules of Berezin integration:[12]

$$\int dx=0,\quad \int x\,dx=1,\quad \int d\bar{x}=0,\quad \int\bar{x}\,d\bar{x}=1\ , \tag{5.11}$$

and likewise for y and $\bar{y}$.

The eigenstates (5.6)–(5.9) may be normalized using (5.10), thus fixing the normalization factors. One finds that $N_1=N_2=1$ and

$$N_0=\left[\frac{\omega+\nu}{2\omega}\right]^{1/2},\quad N_3=\left[\frac{\omega-\nu}{2\omega}\right]^{1/2}\ . \tag{5.12}$$

One may also see that x and $\partial/\partial x$ are adjoints of each other with respect to (5.10), and likewise y and $\partial/\partial y$. It follows that the Hamiltonian operator (5.5) is self-adjoint in the inner product (5.10).

To interpret the states (5.6)–(5.9), it is convenient to introduce the operators

$$a=\frac{1}{(2\omega)^{1/2}}\left[(\nu+\omega)^{1/2}\frac{\partial}{\partial y}+\frac{m}{(\nu+\omega)^{1/2}}x\right]\ , \tag{5.13}$$

$$b=\frac{1}{(2\omega)^{1/2}}\left[(\nu+\omega)^{1/2}\frac{\partial}{\partial x}-\frac{m}{(\nu+\omega)^{1/2}}y\right]\ , \tag{5.14}$$

and their adjoints

$$a^{\dagger}=\frac{1}{(2\omega)^{1/2}}\left[(\nu+\omega)^{1/2}y+\frac{m}{(\nu+\omega)^{1/2}}\frac{\partial}{\partial x}\right]\ , \tag{5.15}$$

$$b^{\dagger}=\frac{1}{(2\omega)^{1/2}}\left[(\nu+\omega)^{1/2}x-\frac{m}{(\nu+\omega)^{1/2}}\frac{\partial}{\partial y}\right] . \quad (5.16)$$

These operators satisfy the anticommutation relations

$$\{a,a^{\dagger}\}=1,\ \ \{b,b^{\dagger}\}=1 , \quad (5.17)$$

with all others equal to zero. It is readily verified that these operators are creation and annihilation operators between the states (5.6)—(5.9). Thus one has, for example,

$$\begin{aligned} &a\psi^{(0)}=0,\ \ a^{\dagger}\psi^{(0)}=\psi^{(2)} ,\\ &a\psi^{(2)}=\psi^{(0)},\ \ a^{\dagger}\psi^{(2)}=0 , \end{aligned} \quad (5.18)$$

plus 12 more similar relations. The operators a and $a^{\dagger}$ step between $\psi^{(0)}$ and $\psi^{(2)}$ and between $\psi^{(1)}$ and $\psi^{(3)}$, while b and $b^{\dagger}$ step between $\psi^{(0)}$ and $\psi^{(1)}$ and between $\psi^{(2)}$ and $\psi^{(3)}$.

It is natural to introduce the number operators N_a and N_b, defined by $N_a=a^{\dagger}a$, $N_b=b^{\dagger}b$, and the Hamiltonian operator (5.5) may then be expressed in terms of these operators by

$$N^{-1}H_n=\omega(N_a+N_b)-\omega . \quad (5.19)$$

Since N_a and N_b each commute with H_n, they possess a set of simultaneous eigenstates. In fact, the states (5.6)—(5.9) are eigenstates of N_a and N_b:

$$\begin{aligned} &N_a\psi^{(0)}=0,\ \ N_a\psi^{(1)}=0 ,\\ &N_a\psi^{(2)}=\psi^{(2)},\ \ N_a\psi^{(3)}=\psi^{(3)} . \end{aligned} \quad (5.20)$$

Identical relations hold for N_b but with $\psi^{(1)}$ and $\psi^{(2)}$ interchanged.

The interpretation of the states $\psi^{(0)}$, $\psi^{(1)}$, $\psi^{(2)}$, and $\psi^{(3)}$ is now clear. Regard $a^{\dagger}$ and $b^{\dagger}$ as creation operators of particle and antiparticle states, respectively, in the mode labeled by some value of n,p. Then $\psi^{(0)}$ represents the vacuum state, $\psi^{(1)}$ represents a one-antiparticle state, $\psi^{(2)}$ represents a one-particle state, and $\psi^{(3)}$ represents a two-particle state containing a particle and an antiparticle. Note that here we are referring to particle states as defined by instantaneous Hamiltonian diagonalization, which leads to different definitions of what one means by a particle state for different values of the scale factor e^{α}. Moreover, this definition of a particle state will not in general agree with the measurements made by a detector moving along a geodesic, as we shall see in Sec. X. Note also that the total Hamiltonian (4.21) contains two terms of the type (5.5), for each mode n,p, in order to account for both helicity states of the particle.

VI. THE SCHRÖDINGER EQUATIONS

The Wheeler-DeWitt equation (5.1) will be very difficult to solve, because it involves an infinite number of variables. However, as was shown in Ref. 8. Wheeler-DeWitt equations of this type may be approximated by time-dependent Schrödinger equations for each mode along the trajectories of the minisuperspace background model. Here we repeat the derivation, partly because we are interested in a more careful treatment of divergences, but also because the derivation of Ref. 8 involved some inaccurate statements, although the final result is correct.

The Wheeler-DeWitt equation (5.1) may be written

$$\left[-\frac{1}{2}\nabla^2+\frac{1}{2}e^{-3\alpha}V+\sum_{np}H_{np}\right]\Psi=0 , \quad (6.1)$$

where ∇^2 is the Laplacian in the two-dimensional metric (3.9). Since the perturbation modes do not couple to each other, a natural ansatz for the wave function is

$$\Psi=\Psi_0(q^a)\prod_{np}\Psi_{np}(q^a,m_{np},s_{np},t_{np},r_{np}) , \quad (6.2)$$

where each wave function Ψ_{np} depends only on the individual perturbation modes m_{np}, s_{np}, t_{np}, r_{np}. Inserting (6.2) into (6.1), one obtains

$$\frac{1}{2}\left[-\frac{\nabla^2\Psi_0}{\Psi_0}+e^{-3\alpha}V+\sum_{np}\Omega_{np}-\left[\sum_{np}\frac{\nabla\Psi_{np}}{\Psi_{np}}\right]^2\right]+\sum_{np}\left[\frac{H_{np}\Psi_{np}}{\Psi_{np}}-\frac{1}{2}\Omega_{np}-\frac{(\nabla\Psi_0)}{\Psi_0}\cdot\frac{(\nabla\Psi_{np})}{\Psi_{np}}-\frac{1}{2}\nabla^2(\ln\Psi_{np})\right]=0 , \quad (6.3)$$

where we have introduced the quantity Ω_{np} in the square brackets and subtracted it in the second set of large parentheses. Its value will be chosen later, essentially to subtract off the vacuum energies of the perturbation modes. It depends on the background variables q^a and the label n but will turn out not to depend on the degeneracy label p.

The wave functions Ψ_{np} will turn out to be slowly varying functions of the background variables q^a, and thus the term $\nabla^2(\ln\Psi_{np})$ will be small. It represents a higher-order correction in the Wentzel-Kramers-Brillouin (WKB) solution and so may be dropped, as one can verify after explicit expressions for the Ψ_{np} have been obtained.

Consider next the last term in the square brackets of (6.3), involving the square of a sum over modes. We are anticipating the Ψ_{np} to be of the form $[1+f(q^a)xy]$ for some function f. One should of course also include an overall background-dependent factor in this expression, but such a factor may always be absorbed into the definition of Ψ_0 in (6.2) and it is convenient to assume that this has been done. It follows that the term in question is a sum over terms which are quartic in the perturbation variables and moreover, involves a coupling between different modes. However, we are working only to quadratic order in the perturbation variables, at which the different modes do not interact, and it would therefore be inconsistent to include this term in our considerations without including the quartic order contributions to the gravitational and matter fields, which is beyond the scope of this calculation. This term will therefore be dropped. One would expect that it is consistent to neglect such quartic contributions provides that the inhomogeneities are sufficiently small.

Of the remaining terms in (6.3), the terms independent of the perturbation modes must vanish separately. Moreover, since the individual modes do not couple, each term in the sum over all modes must vanish independently. Equation (6.3) thus implies

$$\left[-\nabla^2+e^{-3\alpha}V+\sum_{np}\Omega_{np}\right]\Psi_0=0 , \tag{6.4}$$

$$(H_{np}-\tfrac{1}{2}\Omega_{np})\Psi_{np}=\frac{(\nabla\Psi_0)}{\Psi_0}\cdot\nabla\Psi_{np} . \tag{6.5}$$

Equation (6.4) is of the form of the background Wheeler-DeWitt equation (3.7), but modified by the term $\sum\Omega_{np}$. This term represents the back reaction of the perturbation modes on the homogeneous modes. We shall show in Sec. IX that it is small after regularization, so that Ψ_0 will be approximately the same as the wave function for the minisuperspace background, described in Sec. III. In regions where the wave function oscillates rapidly, one may use the WKB approximation for Ψ_0 and write $\Psi_0=C\exp(iS)$. The operator on the right-hand side of (6.5) may then be written

$$\frac{\nabla\Psi_0}{\Psi_0}\cdot\nabla\approx i\nabla S\cdot\nabla=i\frac{\partial}{\partial t} , \tag{6.6}$$

where we have introduced the vector $\partial/\partial t$, for some parameter t. Equation (6.5) now reads

$$(H_{np}-\tfrac{1}{2}\Omega_{np})\Psi_{np}=i\frac{\partial\Psi_{np}}{\partial t} . \tag{6.7}$$

The integral curves of $\partial/\partial t$ are the classical trajectories corresponding to the classical limit of the minisuperspace background model and are parametrized by the coordinate time t. Equation (6.7) is thus a time-dependent Schrödinger equation for each perturbation mode along the trajectories of the minisuperspace background.

If the back reaction is large, the above derivation will still hold and (6.7) will still be a Schrödinger equation along the integral curves of $\partial/\partial t$. These integral curves, however, will be the trajectories corresponding to the classical limit of the solution Ψ_0 of (6.4), which could be very different from the classical trajectories described in Sec. III.

The Hamiltonian H_{np} is the sum of two terms of the form (5.6) with x,y denoting m_{np},s_{np} in the first and denoting t_{np},r_{np} in the second it follows that the wave function Ψ_{np} may be further decomposed

$$\Psi_{np}(\alpha,\phi,m_{np},s_{np},t_{np},r_{np})=\psi_n(\alpha,\phi,m_{np},s_{np})\times\psi_n(\alpha,\phi,t_{np},r_{np}) \tag{6.8}$$

and each wave function ψ_n will thus obey the Schrödinger equation

$$(H_n-\tfrac{1}{4}\Omega_{np})\psi_n=i\frac{\partial\psi_n}{\partial t} , \tag{6.9}$$

where H_n is given by (5.6). Note that $\psi_n(\alpha,\phi,\ldots)$ is the same function for both terms in the product in (6.8), since the Schrödinger equation for each term has the same form and, as we shall see in the next section, each term satisfies the same initial conditions. It will also turn out that the function ψ_n is independent of p; it is thus labeled only by n.

VII. BOUNDARY CONDITIONS

We come now to the most central feature of the model, namely, the application of the Hartle-Hawking proposal to set boundary conditions on the wave function. This will provide initial conditions for the Schrödinger equations derived in the preceding section.

We seek the solution to the Wheeler-DeWitt equation defined by the path integral over the class C of compact four-metrics and regular matter fields:

$$\Psi=\int_C d[e^a_\mu]d[\phi_A]d[\tilde{\phi}_{A'}]d[\chi_A]d[\tilde{\chi}_{A'}]e^{-\tilde{I}} . \tag{7.1}$$

The Euclidean action $\tilde{I}$ is obtained by choosing the lapse function N to be negative imaginary in the Lorentzian action I, and then $\tilde{I}=-iI$. For our model, the class C of paths will be a set of paths in the infinite-dimensional superspace with coordinates $(\alpha,\phi,m,\tilde{m},s,\tilde{s},t,\tilde{t},r,\tilde{r})$ which may be parametrized by the Euclidean time coordinate $\tau=\int iN\,dt$, and their initial point is conveniently taken to be $\tau=0$. To ensure that the paths correspond to compact metrics and regular matter fields, conditions must be imposed on the superspace coordinates at $\tau=0$. For α and ϕ, the appropriate conditions are

$$e^\alpha=0,\quad \frac{de^\alpha}{d\tau}=1,\quad \phi=\phi_0,\quad \frac{d\phi}{d\tau}=0 , \tag{7.2}$$

at $\tau=0$.

Consider next the boundary conditions for the fermion field. We have developed a formalism for which $\tilde{m},\tilde{s},\tilde{t},\tilde{r}$ are fixed on S_I and m,s,t,r are fixed on S_F. However, the Hartle-Hawking proposal demands that the initial surface S_I be shrunk to zero in a regular manner, thus leaving S_F as the only bounding surface. One therefore requires that the fermion field is regular on the interior of S_F, which is a three-sphere in our model, and matches the prescribed values of m,s,t,r on S_F. Now consider what this implies for the coefficients $m,\tilde{m}$, etc. These are defined through the expansions (4.1)–(4.4), which involve the weight factor $e^{-3\alpha/2}$ and this is singular at $\tau=0$. Regularity of ϕ_A, χ_A, $\tilde{\phi}_{A'}$ and $\tilde{\chi}_{A'}$ on the interior of the three-sphere therefore implies that

$$\begin{aligned} m_{np}=s_{np}=t_{np}=r_{np}=0 , \\ \tilde{m}_{np}=\tilde{s}_{np}=\tilde{t}_{np}=\tilde{r}_{np}=0 , \end{aligned} \tag{7.3}$$

at $\tau=0$. The Hartle-Hawking proposal thus demands that one sums over all paths satisfying (7.2) and (7.3) at the initial point $\tau=0$ which match the prescribed values of α, ϕ, m_{np}, s_{np}, t_{np}, r_{np} at the final point $\tau=\tau'$, say, at which one wishes to know the value of the wave function.

It is enlightening to discuss the boundary conditions for the fermion fields purely at the classical level since, at first sight, it might appear that the classical solution will not possess sufficient freedom to satisfy all the conditions. In fact, since the fermion action is quadratic in the fermion fields, the evaluation of the path integral over the fermion modes is essentially semiclassical. Thus, for the model considered here, it is sufficient to understand the boundary conditions at the classical level.

Let us write each of the fields ϕ_A, χ_A, $\tilde{\phi}_{A'}$, and $\tilde{\chi}_{A'}$ in the form $\phi_A=\phi_A^{(+)}+\phi_A^{(-)}$ where, in the decomposition (4.1)—(4.4), the (+) part corresponds to the modes with unbarred coefficients and the (−) part corresponds to the modes with barred coefficients. The boundary value problem is thus to find the solution to the Euclidean-Dirac equation such that $\phi_A^{(+)}$, $\chi_A^{(+)}$, $\tilde{\phi}_{A'}^{(+)}$, and $\tilde{\chi}_{A'}^{(+)}$ match prescribed values on S_F and ϕ_A, χ_A, $\tilde{\phi}_{A'}$, and $\tilde{\chi}_{A'}$ are regular on the interior of S_F. From the Euclidean version of the Dirac equations (2.7)—(2.10), one may derive second-order equations for the fields $\phi_A^{(+)}$, $\chi_A^{(+)}$, $\tilde{\phi}_{A'}^{(+)}$, and $\tilde{\chi}_{A'}^{(+)}$. These may be solved on the interior of S_F subject to the above boundary conditions and the solution is then uniquely determined. Now the Dirac equations (2.7)—(2.10) relate the derivatives of the (+) variables to the (−) variables and vice versa, as one may see from the expanded form (4.13)—(4.16). Equations (2.7)—(2.10) may thus be used to determine the fields $\phi_A^{(-)}$, $\chi_A^{(-)}$, $\tilde{\phi}_{A'}^{(-)}$, and $\tilde{\chi}_{A'}^{(-)}$ in terms of the derivatives of $\phi_A^{(+)}$, $\chi_A^{(+)}$, $\tilde{\phi}_{A'}^{(+)}$, and $\tilde{\chi}_{A'}^{(+)}$. One should then ask whether or not the (−) parts of the fields satisfy the condition of regularity on the interior of S_F. Since the (+) parts of the fields are regular solutions to elliptic equations on the interior of S_F they will be analytic, as will their derivatives, so it follows that the (−) parts of the fields will also be analytic on the interior of S_F. The variables $\phi_A^{(-)}$, $\chi_A^{(-)}$, $\tilde{\phi}_{A'}^{(-)}$, and $\tilde{\chi}_{A'}^{(-)}$ will therefore satisfy the requirement of regularity on the interior of S_F. A unique solution to the classical field equations is thus obtained, satisfying all the conditions, showing that the boundary value problem is well posed.

It is appropriate at this point to explain why we chose to fix the unbarred variables on S_F, rather than the barred variables, or some combination of barred and unbarred variables. The reason for our choice is to ensure that a regular solution exists in the massless limit. If $m=0$, all the variables decouple and each one obeys a first-order equation. The general solution for the unbarred variables is regular on the interior of S_F, and one then has the freedom to require the solution to match a prescribed value on S_F. The general solution for the barred variables, however, is singular at $\tau=0$ and one is forced to take the barred variables to be zero identically, this being the only regular solution on the interior of S_F. There is no freedom to match a prescribed value on S_F. In the massless case therefore, one is compelled to fix the unbarred variables on S_F with the barred variables free. In the massive case, the variables each obey a second-order equation, one of whose solution is always regular on the interior of S_F, and thus one has more freedom in the choice of boundary conditions. At a deeper level, what we are encountering here is the general theory of Atiyah *et al.* on the spectral theory of elliptic operators.[18]

We have made the ansatz (6.2) for the total wave function, and have shown that Ψ_0 is already known. We now show how approximate expressions for the perturbation wave functions Ψ_n at small geometries may be obtained from the path integral, and hence used as initial conditions for the Schrödinger equations.

The path integral over the fermion modes will be of the form

$$\int d[x]\,d[\tilde{x}]\,d[y]\,d[\tilde{y}]e^{-\tilde{I}_n} , \tag{7.4}$$

where $\tilde{I}_n$ is the Euclidean version of the action (4.11). According to the boundary conditions described above, the integral is taken over a set of paths $(x(\tau),\tilde{x}(\tau),y(\tau),\tilde{y}(\tau))$ satisfying $\tilde{x}=\tilde{y}=0$ at $\tau=0$ and $x=x'$, $y=y'$ at $\tau=\tau'$, where x',y' is the point at which we wish to evaluate the wave function. Equation (7.4) may be evaluated to yield an expression of the form

$$A\exp(-\tilde{I}_n^{\,\mathrm{cl}}) , \tag{7.5}$$

where $\tilde{I}_n^{\,\mathrm{cl}}$ is the action of the solution to the classical Euclidean field equations which satisfies the above boundary conditions and A is a prefactor, evaluated by integrating over the fluctuations about the extremizing path. The x',y' dependence is contained entirely in the exponent, so the prefactor will not be needed.

To obtain the wave function $\psi_n(\alpha',\phi',x',y')$, it is necessary to perform a functional integration of (7.5) over paths $(\alpha(\tau),\phi(\tau))$ satisfying the initial conditions (7.2) and matching (α',ϕ') at $\tau=\tau'$. One would expect the dominant contribution to come from paths close to solutions of the classical Euclidean field equations. For such paths, one may employ the adiabatic approximation, also used in Ref. 8, in which one assumes that α is a slowly varying function of τ. In particular, one assumes that

$$\left|\frac{d\alpha}{d\tau}\right| \ll (n+\tfrac{3}{2})e^{-\alpha} . \tag{7.6}$$

Equation (7.6) is satisfied by all paths in the neighborhood of $\tau=0$ by virtue of the initial conditions (7.2). Moreover, if $\alpha(\tau)$ is the solution to the field equations satisfying the initial conditions (7.2), then (7.6) is satisfied throughout the whole of Euclidean region. In the Lorentzian region, Eq. (7.6) is the condition that the mode labeled by n is inside the horizon of the de Sitter phase. This approximation may be used to solve the Dirac equation and hence obtain the exponent in (7.5).

On evaluating the Euclidean action $\tilde{I}_n$ on the classical path, one discovers that the volume term vanishes leaving just the boundary term; hence,

$$\tilde{I}_n^{\,\mathrm{cl}}=\tfrac{1}{2}[\tilde{x}(\tau')x'+\tilde{y}(\tau')y'] . \tag{7.7}$$

It is now necessary to solve the Euclidean-Dirac equation, subject to the above boundary conditions, to find $\tilde{x}(\tau')$ and $\tilde{y}(\tau')$ in terms of x' and y'. The Euclidean version of (4.17) is

$$\frac{d^2x}{d\tau^2}-\left[\frac{d\nu}{d\tau}+\nu^2+m^2\right]x=0 . \tag{7.8}$$

In terms of ν, the adiabatic approximation (7.6) is $|d\nu/d\tau| \ll \nu^2$ whence the approximate solution to (7.8) satisfying $x(0)=0$, $x(\tau')=x'$, is

$$x(\tau)=\frac{\sinh(\omega\tau)}{\sinh(\omega\tau')}x' . \tag{7.9}$$

The Euclidean version of (4.13) is

$$-\frac{dx}{d\tau}+\nu x-m\tilde{y}=0\ , \tag{7.10}$$

from which one may determine $\tilde{y}(\tau')$ in terms of x':

$$\tilde{y}(\tau')=\frac{1}{m}[\nu-\omega\coth(\omega\tau')]x'\ . \tag{7.11}$$

A similar calculation for y and $\tilde{x}$ yields

$$\tilde{x}(\tau')=\frac{1}{m}[-\nu+\omega\coth(\omega\tau')]y'\ . \tag{7.12}$$

The Euclidean action (7.8) of the classical solution is thus given by

$$\tilde{I}_n^{\rm cl}=\frac{1}{m}[\nu-\omega\coth(\omega\tau')]x'y'\ . \tag{7.13}$$

For large values of n, $\coth(\omega\tau')\approx 1$, whence

$$\tilde{I}_n^{\rm cl}\approx\frac{(\nu-\omega)}{m}x'y'=\frac{-m}{(\omega+\nu)}x'y' \tag{7.14}$$

and one obtains the following approximate expressions for the perturbation wave functions:

$$\psi_n(\alpha',\phi',x',y')\approx\exp\left[\frac{mx'y'}{(\omega+\nu)}\right]$$

$$=1+\frac{mx'y'}{(\omega+\nu)}\ , \tag{7.15}$$

apart from a possible prefactor, independent of x' and y'. Comparing this with the eigenstates of the Hamiltonian (5.6)–(5.9), we see that the Hartle-Hawking proposal picks out the lowest-energy eigenstate for the fermion modes, that is, the ground state.

In this section we have solved the Dirac equation to obtain the semiclassical expressions for the perturbation wave functions, using the adiabatic approximation (7.6). This method indicates the generality of the model, that is, the fact that it does not depend in a crucial way on the detailed behavior of the minisuperspace background. However, in the limit of an exact de Sitter background, which is a good approximation to the initial behavior of the model, the Dirac equation may be solved exactly in terms of hypergeometric functions. This is carried out in Appendix C, partly to check the approximation, but also because such a solution turns out to be needed in order to solve the Schrödinger equation outside the horizon, as we shall see in the next section.

VIII. EVOLUTION AND PARTICLE CREATION

We now consider the evolution of the perturbation wave functions according to the Schrödinger equation (6.9), subject to the initial condition derived from the Hartle-Hawking proposal in the preceding section. We make the following ansatz for the wave function:

$$\psi_n=1+\frac{m}{(u+\nu)}xy \tag{8.1}$$

for some function $u(t)$. Inserting this into (6.9), one obtains

$$-\tfrac{1}{4}\Omega_{np}-\nu-\frac{m^2}{(u+\nu)}+\left[\frac{(\nu+\frac{1}{4}\Omega_{np})}{(u+\nu)}-m\right]xy$$

$$=-im\frac{(\dot{u}+\dot{\nu})}{(u+\nu)^2}xy\ . \tag{8.2}$$

In order that the terms independent of x and y cancel, it is necessary to make the following choice for Ω_{np}:

$$\Omega_{np}=-4\left[\nu+\frac{m^2}{(u+\nu)}\right] \tag{8.3}$$

and (8.2) will then be satisfied if u obeys the equation

$$i\dot{u}+u^2=\nu^2-i\dot{\nu}+m^2 \tag{8.4}$$

which is of the Riccati type.

Inside the horizon, $|\dot{\nu}|\ll\nu^2$, and (8.4) then has the approximate solution $u\approx\pm(\nu^2+m^2)^{1/2}=\pm\omega$. The initial condition is that ψ_n is in the ground state initially, so $u\approx+\omega$ and ψ_n remains in the ground state for as long as the mode remains inside the horizon.

To consider the subsequent evolution outside the horizon, more detailed calculation is necessary. The inflationary phase is accurately described by taking the scale factor e^α to be of the de Sitter form, and in this case it is possible to solve (8.4) exactly in terms of hypergeometric functions. This is done in Appendix C. Outside the horizon, i.e., for $\nu\ll H$, one finds that the approximate solution is

$$u=-\nu-m\tanh(Ht)-\frac{2im}{H}\left[1-\frac{2im}{H}\right]^{-1}\nu\ . \tag{8.5}$$

Given the solution (8.5), one may calculate particle creation during the de Sitter phase. The average number of particles in the state ψ_n at fixed α and ϕ is given by

$$\langle N_a\rangle=\frac{(\psi_n,N_a\psi_n)}{(\psi_n,\psi_n)}=\frac{m^2}{2\omega(\omega+\nu)}\frac{|\omega-u|^2}{(|\nu+u|^2+m^2)}\ . \tag{8.6}$$

This is equal to the average number of antiparticles $\langle N_b\rangle$, as one would expect, since the creation takes place as particle-antiparticle pairs. This also corresponds to the fact that the single particle and antiparticle states $\psi^{(2)}$ and $\psi^{(1)}$ do not play any role. The total number of particles created, Γ, is then obtained by summing over all modes and over both helicity states:

$$\Gamma=2\sum_{np}(\langle N_a\rangle+\langle N_b\rangle)=4\sum_{n=0}^{\infty}(n+1)(n+2)\langle N_a\rangle\ . \tag{8.7}$$

It is convenient to split the sum over n into three regions:

(i) $n\ll n_1=me^\alpha$ (i.e., $\nu\ll m$) ,

(ii) $n_1\ll n\ll n_2=He^\alpha$ (i.e., $m\ll\nu\ll H$) ,

(iii) $n_2\ll n$ (i.e., $\nu\gg H$) .

Thus (8.7) may be written

$$\Gamma=\left[\sum_{n=0}^{n_1}+\sum_{n=n_1}^{n_2}+\sum_{n=n_2}^{\infty}\right]4(n+1)(n+2)\langle N_a\rangle \ . \qquad (8.8)$$

To determine whether or not Γ is finite, it is necessary to consider the sum over modes in region (iii), in which $\nu \gg H$ so these modes are inside the horizon (the de Sitter phase is finite in length in this model, and there will be an infinite number of very-short-wavelength modes which never leave the horizon). For such modes, we have shown that the wave function is approximately in the ground state $u\approx\omega$, so it appears that $\langle N_a\rangle\approx 0$. The question, however, is whether or not it is sufficiently close to the ground state for the sum $\sum n^2\langle N_a\rangle$ to be finite. To this end, we need the more accurate solution given in Appendix C, which is $u\approx\omega-iH$. Equation (8.6) thus yields

$$\langle N_a\rangle\approx\frac{m^2}{2\omega(\omega+\nu)}\frac{H^2}{[(\omega+\nu)^2+H^2+m^2]} \ . \qquad (8.9)$$

For large n, it is easily seen that $\langle N_a\rangle=O(n^{-4})$ and hence that $\sum n^2\langle N_a\rangle$ is finite. The contribution to Γ in region (iii) is of order m^2e^{α}/H which will turn out to be considerably smaller than the contribution from regions (i) and (ii).

Consider next regions (i) and (ii). Using (8.5), one finds that $\langle N_a\rangle\approx 1$ in region (i) and this yields a contribution to Γ to order $m^3e^{3\alpha}$. In region (ii), $\langle N_a\rangle\approx\frac{1}{2}$, and this yields a contribution to Γ of order $H^3e^{3\alpha}$, which is considerably greater than the contribution from region (i). To summarize, the particle production resulting from the de Sitter phase is finite, and is dominated by the modes for which $m\ll\nu\ll H$, which yield a value of Γ of order $H^3e^{3\alpha}$.

Finally, we consider the massless case. If $m=0$, the variables $x,\bar{x}$ decouple from $y,\bar{y}$, and the Hamiltonian operator for each mode, H_{np}, is a sum of four terms, each of the form $-\nu/2+\nu x\partial/\partial x$. The eigenstates of this operator are 1 and x with eigenvalues $-\nu/2$ and $\nu/2$, respectively. It is straightforward to show that the Hartle-Hawking proposal picks out the lowest-energy eigenstate, so that the initial condition for the Schrödinger equation (6.7) is $\Psi_{np}=1$. In fact, $\Psi_{np}=1$ for all time is the exact solution to the Schrödinger equation, as is easily verified, and it is necessary to choose $\Omega_{np}=-4\nu$. For the massless case, therefore, the fermion modes remain in their ground state throughout the subsequent evolution, so there will clearly be no particle production. This is not really surprising, bearing in mind the fact that the massless Dirac action is conformally invariant. One may therefore conformally transform the problem from the Friedman-Robertson-Walker- (FRW) type universe considered here to the Einstein static universe, in which there is no particle production because it is static.

IX. THE BACK REACTION

In the preceding section, we solved the Schrödinger equations and obtained an expression for the quantity Ω_{np}. This feeds back into the Wheeler-DeWitt equation (6.4) to give the fermionic contribution to the back reaction, $\sum\Omega_{np}$. As one would expect, however, this expression is formally divergent.

In an earlier version of this paper, we regularized the quantity $\sum\Omega_{np}$ by inserting a cutoff in the summation. We then subtracted the divergent parts following a prescription previously used by Ford in the Einstein static universe.[19] However, there are reasons for believing that this method is not really appropriate to Robertson-Walker space-times. The main objection is that it is not covariant, and is therefore at variance with the widely held belief that the regularization ought to be covariant in order to obtain the right answer. This objection is valid, as may be seen by considering particular calculations. For example, this noncovariant method preserves the trace of the energy-momentum tensor at every stage, so no trace anomaly is obtained, contrary to the results of covariant calculations. (There is no trace anomaly in the Einstein static universe, so we are not questioning Ford's calculations.) Similarly, the method yields an expression for the vacuum energy of massless fermions in de Sitter space different from that obtained using covariant methods.[20] For these and other reasons, we now believe that our original calculation was based on a method that does not apply.

It appears, however, that the problem of treating divergences in quantum cosmology is perhaps a little more difficult than it is in quantum field theory on a fixed background. Firstly, there is the issue of regularizing the divergences. In quantum field theory on a fixed background, one works at the level of the action, or the field equations, at which general covariance is manifest, so covariant regularization is reasonably straightforward. In quantum cosmology, on the other hand, one works at the level of the Hamiltonian, at which it is more difficult to maintain general covariance. Indeed, as we have discussed, one appears to be led naturally into regularizing the divergences in the Wheeler-DeWitt equation in a noncovariant manner.

A possible way to avoid this difficulty is to use dimensional regularization. One could quite simply do the entire calculation in d space-time dimensions, so that a covariant regularization scheme is in use from the very beginning. In connection with this, the results of Hill are encouraging.[21] Hill considered scalar field quantization in de Sitter space, using the functional Schrödinger quantization method and dimensional regularization. The results of his calculation of $\langle T_{\mu\nu}\rangle$ are in agreement with other covariant methods.[20]

There is a second difficulty concerning the subtraction of the divergent terms, after a regulator has been introduced. In quantum field theory on a fixed background, one usually postulates the existence of higher derivative terms in the field equations with coefficients denoted by γ, say. By renormalization of the cosmological term Λ, the Planck mass m_P, and γ, one can absorb all the divergences of $\langle T_{\mu\nu}\rangle$. One then assumes that γ_R, the renormalized value of γ, is zero, or at least negligible.

It is not obvious, however, that such a scheme will go through in quantum cosmology. One may include higher derivative terms in quantum cosmological models and indeed, models of this type has been considered.[22] The problem is that the inclusion of such terms introduces extra dynamical variables into the theory, completely changing the Hamiltonian formalism, and hence the Wheeler-DeWitt equation. It may well be possible to correctly regularize the divergences in the Wheeler-DeWitt equation, and then absorb them by normalization. It is not clear, however, that one may then smoothly take the limit $\gamma_R \rightarrow 0$ and recover the Wheeler-DeWitt equation of conventional Einstein gravity coupled to a correctly renormalized energy-momentum tensor. The limit may be singular.

The handling of divergences in quantum cosmology thus appears to be more problematic than in quantum field theory on a fixed background. Nevertheless, the problems are not insoluble and we hope to address them in more detail in a future publication. Here, as a temporary measure to avoid the difficulties outlined above, we will relate the formalism we have been using to the semiclassical Einstein equations, in which the back-reaction problem is better understood. In Appendix D, we show that when quantum corrections due to the matter are taken into account, the Hamilton-Jacobi equation for the background Wheeler-DeWitt equation is

$$(\nabla \operatorname{Re} S)^2 + e^{-3\alpha} V + \int d^3x \, h^{1/2} \langle T_{00} \rangle = 0 \ . \tag{9.1}$$

This is equivalent to the time-time component of the semiclassical Einstein equations. This is a useful result since all the divergences are contained in $\langle T_{00} \rangle$, to which covariant regularization techniques may be applied.

Consider therefore $\langle T_{\mu\nu} \rangle_{\text{ren}}$, the renormalized expectation value of the energy-momentum tensor in the quantum state defined by the Hartle-Hawking proposal. For simplicity we will restrict attention to an exact de Sitter background, which describes the early part of the evolution quite well. As will be argued in Sec. X, the state defined by the Hartle-Hawking proposal is de Sitter invariant. It follows that $\langle T_{\mu\nu} \rangle_{\text{ren}}$ is of the form

$$\langle T_{\mu\nu} \rangle_{\text{ren}} = C g_{\mu\nu} \tag{9.2}$$

for some constant C, which has the dimensions of an energy density. In the massless case, the only scale available is H, the Hubble parameter of the de Sitter phase, thus C is proportional to H^4. In fact, explicit calculation yields the result $C = 11H^4/1920\pi^2$ (Ref. 20). The potential in (9.1) is thus given by

$$e^{-3\alpha} V + 2\langle H_f \rangle = e^{3\alpha} \left[M^2\phi^2 + \frac{11H^4}{1920\pi^2} \right] - e^{4\alpha} \ . \tag{9.3}$$

Since $H \approx M\phi$, and a typical value of H in a realistic inflationary scenario is about 10^{-5} (Planck units), the back reaction is totally negligible.

For the massive case, one again has the result (9.2), but now another scale is involved, namely m, the mass of the fermion field. However, the classical energy-momentum tensor is linear in m, so assuming that this dependence on m is not affected by the renormalization, one would expect C to be of order mH^3. This is supported by the results of Sec. VIII in which it was shown that the number of particles produced in the de Sitter phase is of order $H^3 e^{3\alpha}$. If the energy density is dominated by the rest mass, then the energy density of particles produced is of order mH^3. A typical value of m would be, say, the electron mass, about 10^{-22} (Planck units) and thus the back region is negligible in the massive case also.

One would expect that the particle production during the post-inflationary evolution is small, since when the modes reenter the horizon, the adiabatic approximation again becomes valid, so the quantum state of the fermion modes is approximately a stationary state. It follows that the back reaction will also be small during this era.

In this section we have shown that the back reaction of the fermionic modes has negligible effect on the behavior of the background modes. This means that our calculation of the particle production is self-consistent, since the particles produced do not significantly change the gravitational background that caused them to be produced in the first place.

Finally, it is instructive to compare these results with those obtained by attempting to regularize the quantity $\sum \Omega_{np}$ directly. In the massless case, one has the relation

$$\int d^3x \, h^{1/2} \langle T_{00} \rangle = 2\langle H_f \rangle = \sum_{np} \Omega_{np} = -4 \sum_{np} \nu \ . \tag{9.4}$$

This is only a formal relation since these expressions are formally divergent, so no meaning can be attached to them unless one specifies a regularization scheme. If one extracts a finite expression from the term on the far right-hand side using the noncovariant method outlined at the beginning of this section, then one obtains a result proportional to $e^{-\alpha}$. Covariant methods applied to the expression on the far left-hand side, on the other hand, yield a result proportional to $e^{3\alpha}$. This clearly illustrates the care that needs to be taken in handling divergences.

X. PARTICLE DETECTOR RESPONSE

In Sec. VIII we considered the evolution of the state defined by the Hartle-Hawking proposal and showed that there is particle production in the massive case but no particle production in the massless case. Particles were defined by instantaneous Hamiltonian diagonalization, which is the most convenient definition in the formalism we have been using. It is well known, however, that in curved space-time there are many different ways to define particles and these definitions will not in general agree. In this section, we consider a quite different definition of the particle concept by considering a particle detector moving along a geodesic, and examining its response to the quantum state defined by the Hartle-Hawking proposal. We will show that the spectrum measured by the detector is thermal, with the correct Fermi-Dirac distribution, but not exactly Planckian.

We showed in Sec. VI that in the Lorentzian region, the full Wheeler-DeWitt equation may be approximated by a set of time-dependent Schrödinger equations along the classical trajectories of the homogeneous modes. We will work in this approximation, which is equivalent to doing quantum field theory for the fermion fields on a fixed classical background. This background is taken to be de Sitter space, which is formally achieved by holding the homogeneous scalar field at a constant value. For simplicity, we will restrict attention to the case of massless fermions, so that we need only consider a single Weyl spinor ϕ_A and its conjugate $\bar{\phi}_{A'}$.

Particle detectors designed to detect scalar particles have been considered by Unruh[23] and DeWitt.[24] The idea is that one considers a simple pointlike system with internal energy levels labeled by E moving along a geodesic $x^\mu(\tau)$, where τ is the detector's proper time. The detector is assumed to interact weakly with the scalar field Φ through an interaction Lagrangian of the form $cm(\tau)\Phi[x(\tau)]$, where c is a small coupling constant and $m(\tau)$ is the detector's monopole moment. One may extend this idea to detectors which detect fermions by considering an interaction Lagrangian of the form

$$L_I = c\{\eta^A(\tau)\phi_A[x(\tau)] + \bar{\phi}_{A'}[x(\tau)]\bar{\eta}^{A'}(\tau)\} , \tag{10.1}$$

where η^A and $\bar{\eta}^{A'}$ describe a fermionic source which couples weakly to the fermion field ϕ_A, $\bar{\phi}_{A'}$. One may now proceed to calculate the detector response in a manner very similar to the scalar case.[20] A more realistic model of a detector would involve coupling to a fermionic current, but as discussed further in Sec. XI this will not alter the basic thermal character of the results.

The fermion field starts in the state defined by the Hartle-Hawking proposal, which we will denote by $|0_{HH}\rangle$, and the detector is taken to start in the state $|E_0\rangle$. (Note that we are using the Heisenberg picture here, in which the states are independent of time, whereas in the preceding sections we used the Schrödinger picture.) For a general trajectory, the detector will undergo a transition to an excited state $|E\rangle$, while the fermion field is excited to a state $|\psi\rangle$. From first-order perturbation theory, the amplitude for this transition is

$$A = i\left\langle \psi,E \left| \int_{-\infty}^{\infty} d\tau L_I \right| 0_{HH},E_0 \right\rangle . \tag{10.2}$$

One may solve the Heisenberg equations of motion for η^A and $\bar{\eta}^{A'}$ and the amplitude factorizes to yield

$$\begin{aligned} A = & ic\langle E|\eta^A(0)|E_0\rangle \int_{-\infty}^{\infty} d\tau\, e^{i(E-E_0)\tau}\langle\psi|\phi_A[x(\tau)]|0_{HH}\rangle \\ & + ic\langle E|\bar{\eta}^{A'}(0)|E_0\rangle \int_{-\infty}^{\infty} d\tau\, e^{i(E-E_0)\tau}\langle\psi|\bar{\phi}_{A'}[x(\tau)]|0_{HH}\rangle . \end{aligned} \tag{10.3}$$

The quantity of interest is the transition probability to all possible E and ψ, obtained by summing $|A|^2$ over all E and over a complete set $|\psi\rangle$. The resulting probability is divergent as a consequence of the double integration over τ. One thus usually considers the transition probability per unit proper time, which is given by

$$c^2\sum_E \langle E|\eta^A(0)|E_0\rangle\langle E_0|\bar{\eta}^{B'}(0)|E\rangle F^{(+)}_{B'A}(E-E_0) + c^2\sum_E \langle E|\bar{\eta}^{A'}(0)|E_0\rangle\langle E_0|\eta^B(0)|E\rangle F^{(-)}_{A'B}(E-E_0) , \tag{10.4}$$

where, introducing $\Delta E = E - E_0$, $F^{(+)}_{B'A}$ and $F^{(-)}_{A'B}$ are given by

$$F^{(+)}_{B'A}(\Delta E) = \int_{-\infty}^{\infty} d\tau\, e^{-i\Delta E\tau} G^{(+)}_{B'A}(x(\tau),x(0)) \tag{10.5}$$

and

$$F^{(-)}_{A'B}(\Delta E) = \int_{-\infty}^{\infty} d\tau\, e^{-i\Delta E\tau} G^{(-)}_{A'B}(x(0),x(\tau)) . \tag{10.6}$$

Equations (10.5) and (10.6) are the spinor versions of the detector response function and are independent of the structure of the detector. $G^{(+)}_{B'A}$ and $G^{(-)}_{B'A}$ are fermion Green's functions and are defined by

$$G^{(+)}_{B'A}(x_2,x_1) = \langle 0_{HH}|\bar{\phi}_{B'}(x_2)\phi_A(x_1)|0_{HH}\rangle , \tag{10.7}$$

$$G^{(-)}_{B'A}(x_2,x_1) = \langle 0_{HH}|\phi_A(x_1)\bar{\phi}_{B'}(x_2)|0_{HH}\rangle . \tag{10.8}$$

Strictly, we should use a notation which indicates that the spinor index B', lives in the tangent space at x_2 while A lives in the tangent space at x_1. The reader should bear this in mind, but no confusion should arise since the only properties of the Green's functions we will use do not depend on such details.

Equations (10.7) and (10.8) are calculated by inserting the expressions (4.1) and (4.2) for ϕ_A and $\bar{\phi}_{A'}$, with the result

$$\begin{aligned} G^{(+)}_{B'A}(x_2,x_1) = & \frac{1}{4\pi^2} e^{-3\alpha(t_2)/2} e^{-3\alpha(t_1)/2} \\ & \times \sum_{n_1p_1q_1}\sum_{n_2p_2q_1} \alpha_{n_1}^{p_1q_1}\alpha_{n_2}^{p_2q_2}[\langle 0_{HH}|\bar{m}_{n_2p_2}(t_2)m_{n_1p_1}(t_1)|0_{HH}\rangle \bar{\rho}_{B'}^{n_2q_2}(\mathbf{x}_2)\rho_A^{n_1q_1}(\mathbf{x}_1) + \text{three similar terms}] . \end{aligned} \tag{10.9}$$

The expectation value in (10.9) is calculated using the holomorphic inner product (5.10), bearing in mind that $\bar{m}_{np}(t)$ and

$m_{np}(t)$ are time-dependent operators in the Heisenberg picture and must first be transformed to the Schrödinger picture using the unitary evolution operator. One thus finds

$$\langle 0_{HH} \,|\, \bar{m}_{n_2 p_2}(t_2) m_{n_1 p_1}(t_1) \,|\, 0_{HH} \rangle = \delta_{n_1 n_2} \delta_{p_1 p_2} \exp\left[-i(n+\tfrac{3}{2}) \int_{t_1}^{t_2} e^{-\alpha} dt \right] , \tag{10.10}$$

where

$$\int_{t_1}^{t_2} e^{-\alpha} dt = 2[\arctan(e^{Ht_2}) - \arctan(e^{Ht_1})] . \tag{10.11}$$

Similarly, one may show that the three other terms in (10.9) are zero. The Green's function $G^{(+)}_{B'A}$ is thus given by

$$G^{(+)}_{B'A}(x_2,x_1) = \frac{1}{2\pi^2} e^{-3\alpha(t_2)/2} e^{-3\alpha(t_1)/2} \sum_{np} \exp\left[-i(n+\tfrac{3}{2}) \int_{t_1}^{t_2} e^{-\alpha} dt \right] \bar{\rho}^{np}_{B'}(\mathbf{x}_2) \rho^{np}_{A}(\mathbf{x}_1) . \tag{10.12}$$

Similarly, for $G^{(-)}_{B'A}$ one finds

$$G^{(-)}_{B'A}(x_2,x_1) = \frac{1}{2\pi^2} e^{-3\alpha(t_2)/2} e^{-3\alpha(t_1)/2} \sum_{np} \exp\left[+i(n+\tfrac{3}{2}) \int_{t_1}^{t_2} e^{-\alpha} dt \right] \sigma^{np}_{B'}(\mathbf{x}_2) \bar{\sigma}^{np}_{A}(\mathbf{x}_1) . \tag{10.13}$$

The infinite summations in (10.12) and (10.13) will not converge as they stand. To make them converge, one may add a factor $-i\epsilon$ to $\int e^{-\alpha} dt$ in (10.12) and $+i\epsilon$ in (10.13). These factors indicate how the poles of the Green's functions are to be displaced, after the series (10.12) and (10.13) have been summed.

One may easily verify that (10.12) and (10.13) are solutions to the homogeneous Weyl equation, as indeed they should be. One may also verify that if one takes the sum of (10.12) and (10.13) at $t_1 = t_2$, a spatial δ function is obtained, by virtue of the completeness relation for the spinor spherical harmonics. The equal-time anticommutation relations are thus respected.

Using (10.12) and (10.13), one may show that the Green's functions are antiperiodic in imaginary time, with period $2\pi i/H$:

$$G^{(\pm)}_{B'A}(t_2 + 2\pi i/H, \mathbf{x}_2; t_1, \mathbf{x}_1) = -G^{(\pm)}_{B'A}(t_2, \mathbf{x}_2; t_1, \mathbf{x}_1) . \tag{10.14}$$

This important property of the de Sitter space Green's functions is shared by thermal Green's functions in Minkowski space and is the main property one uses in showing that the detector experiences a thermal spectrum. From (10.12) one may also show that the Green's function $G^{(+)}_{B'A}(x_2,x_1)$ is analytic in the strip between the real axis and the line $\mathrm{Im}(t_2-t_1) = -\pi/H$, in the complex (t_2-t_1) plane, but is not analytic in the strip between the real axis and the line $\mathrm{Im}(t_2-t_1) = \pi/H$. This structure repeats itself antiperiodically throughout the whole complex plane (Fig. 1). Similarly for $G^{(-)}_{B'A}(x_2,x_1)$, one may show that it is analytic in precisely those regions where $G^{(+)}$ is not analytic, and it is not analytic in precisely those regions where $G^{(+)}$ is (Fig. 2).

To proceed further, we really need expressions for the Green's functions in closed form and to this end, it is useful to first show that the Green's functions are de Sitter invariant. Recall that we are working in the approximation in which the background is taken to be a fixed solution to the classical field equations, namely, de Sitter space; thus, one does not sum over background geometries when calculating the quantum state of the fermion field, only over the matter fields. Consider the Euclidean section of this space, which is a four-sphere S^4, of radius H^{-1}. Suppose one asks for the quantum state of the field on a three-sphere S^3 of radius $e^{\alpha} < H^{-1}$. The state defined by the Hartle-Hawking proposal is calculated by summing over all field configurations which are regular on the section of S^4 interior to the S^3 and match prescribed values on the S^3. The resulting state will depend on the geometry only through the radius of the S^3,

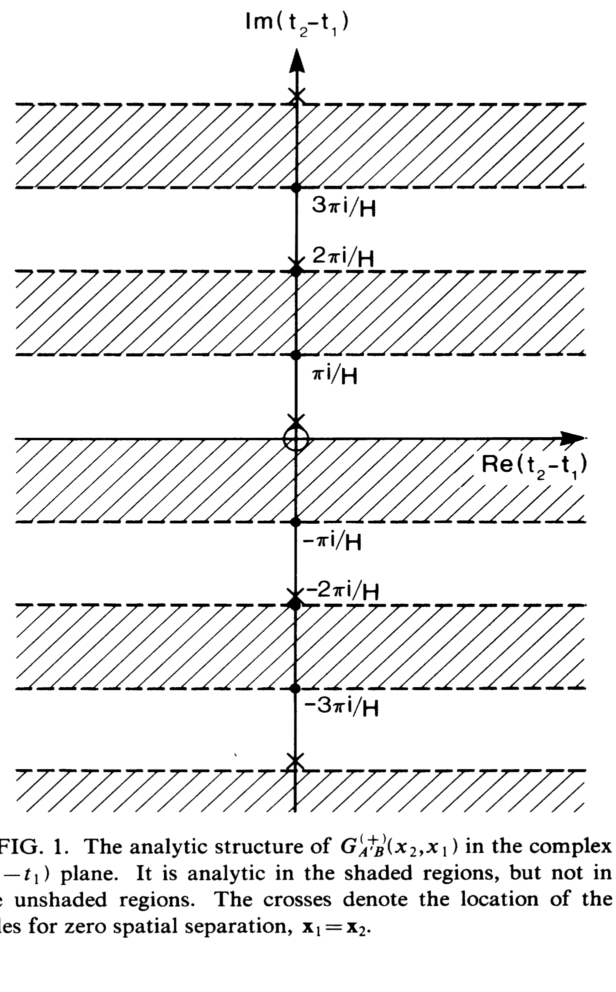

FIG. 1. The analytic structure of $G^{(+)}_{A'B}(x_2,x_1)$ in the complex (t_2-t_1) plane. It is analytic in the shaded regions, but not in the unshaded regions. The crosses denote the location of the poles for zero spatial separation, $\mathbf{x}_1 = \mathbf{x}_2$.

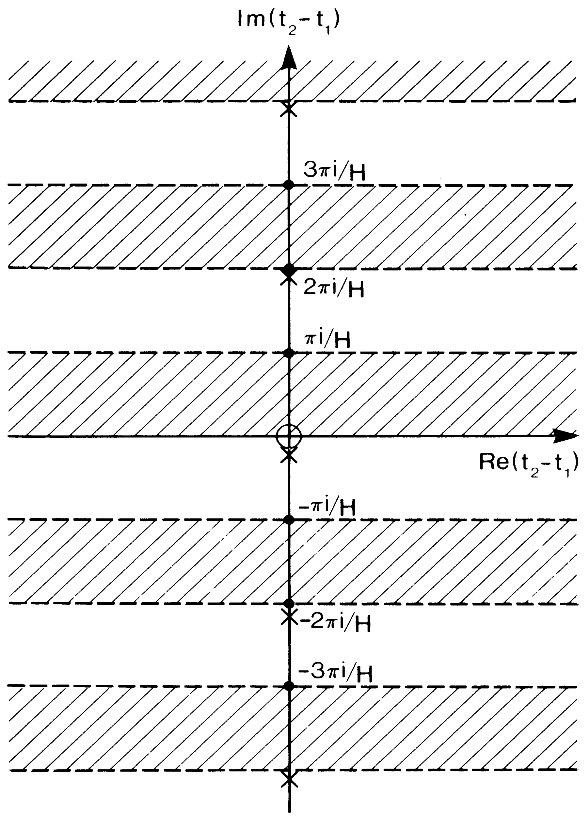

FIG. 2. The analytic structure of $G^{(-)}_{A'B}(x_2,x_1)$. It is analytic in precisely those regions where $G^{(+)}$ is not analytic.

not on its location or orientation on the S^4. One thus has the freedom to move the S^3 around to different locations on the S^4 without changing the quantum state. These different locations are related to each other by transformations under the group SO(5), the isometry group of S^4. It follows that the state is SO(5) invariant on the Euclidean section. On analytic continuation back to the Lorentzian section, one thus finds that the state is invariant under SO(4,1), the de Sitter group, and hence that the Green's functions which are constructed from this state are de Sitter invariant.

The main significance of de Sitter invariance is that we may now use the results of Allen and Lutken,[25] who obtained explicit expressions for the most general de Sitter-invariant spinor two-point functions. This they achieved by finding two-point functions which solved the homogeneous Dirac equation but depend only on μ, the geodesic distance between the points x_1 and x_2 (μ is imaginary for timelike separations). In the massless case, they found the Green's functions to take the form (symbolically):

$$G^{(\pm)}(x_2,x_1)=C\sin^{-3}(\mu H/2)\ , \qquad (10.15)$$

where C is a constant to be determined. The right-hand side of (10.15) should also include the spinor parallel propagator between x_1 and x_2 and the spinor corresponding to the tangent vector to the geodesic between the two points. These details may be found in Ref. 25. Equation (10.15) implies that the only poles in the Green's functions are at $\mu=\pm 2\pi n/H$, where n is an integer; thus there is a pole if and only if x_1 and x_2 are connected by a null geodesic in the complexified space-time, as one would expect. One might have expected that there is also a pole when x_1 and x_2 are antipodal, as can happen in the massive case[25] and also in the scalar case,[26] but this possibility is clearly ruled out here.

We have shown that (10.12) and (10.13) are de Sitter-invariant solutions to the homogeneous Weyl equation. It follows that in the regions where the series (10.12) and (10.13) converge, they will each converge to an expression of the form (10.15). Along the real axis in the t_2-t_1 plane, $G^{(+)}_{B'A}(x_2,x_1)$ is essentially the same as $-G^{(-)}_{B'A}(x_2,x_1)$, except that for $G^{(+)}$ the poles are displaced above the real axis, whereas for $G^{(-)}$ the poles are displaced below it. The Green's functions are consistent with the equal-time anticommutation relations which fixes their normalization, hence, the constant C may be determined. We thus have an expression for the Green's functions in closed form.

Given the above information, the response functions (10.5) and (10.6) may now be partially evaluated. Equation (10.5) involves the Green's function evaluated on the timelike geodesic $x^\mu(\tau)$. Since all timelike geodesics are related by de Sitter transformations, without loss of generality we may choose any, since the Green's functions are de Sitter invariant. A convenient geodesic to use is the one that goes straight up the de Sitter hyperboloid, with constant spatial coordinates, so that proper time τ is the same as coordinate time t. The response function (10.5) is thus given by

$$F^{(+)}_{B'A}(\Delta E)=\int_{-\infty}^{\infty} dt\, e^{-i\Delta E t}G^{(+)}_{B'A}(t,\mathbf{0};0,\mathbf{0})\ , \qquad (10.16)$$

where the integration in the t plane runs along the real axis. Suppose we now consider the quantity $\tilde{F}^{(+)}_{B'A}$ defined to be the same integral, but taken along the line $\mathrm{Im}(t)=2\pi/H$. From the antiperiodicity of $G^{(+)}$, it follows that this is given in terms of $F^{(+)}$ by

$$\tilde{F}^{(+)}_{B'A}(\Delta E)=-e^{2\pi\Delta E/H}F^{(+)}_{B'A}(\Delta E)\ . \qquad (10.17)$$

If one now joins these two paths together at each end to obtain a closed contour, the method of residues then implies

$$F^{(+)}_{B'A}(\Delta E)-\tilde{F}^{(+)}_{B'A}(\Delta E)=2\pi i R_{B'A}(\Delta E)\ , \qquad (10.18)$$

where $R_{B'A}$ is the sum of the residues at the poles of the integrand. Combining (10.17) and (10.18), and evaluating the residue using (10.15), one obtains

$$F^{(+)}_{B'A}(\Delta E)=-C\frac{\pi i}{2H}n_{B'A}\frac{(\Delta E)^2/H^2+\frac{1}{4}}{1+e^{2\pi\Delta E/H}}\ . \qquad (10.19)$$

An identical result is also found for $F^{(-)}_{B'A}(\Delta E)$ and it is convenient to write

$$F^{(+)}_{B'A}(\Delta E)=F^{(-)}_{B'A}(\Delta E)=n_{B'A}f(\Delta E)\ . \qquad (10.20)$$

Equation (10.19) implies that the spectrum is thermal, as we may see by considering the transition rate (10.4). The field η^A may be expanded in the normal way, in terms of particle creation operators and antiparticle annihilation operators, and similarly for $\bar{\eta}^{A'}$. Let $|E_0\rangle$ denote the vacuum state. Then it is not difficult to show that only the first term in (10.4) contributes when $|E\rangle$ is an antiparticle state and only the second term contributes when $|E\rangle$ is a particle state. The probability per unit time for the detector to absorb a particle of energy E is thus given by the second term:

$$\langle E|\bar{\eta}^{A'}(0)|E_0\rangle\langle E_0|\eta^B(0)|E\rangle n_{A'B} f(\Delta E) . \quad (10.21)$$

For the emission of a particle of energy E, one interchanges E_0 and E in (10.4), and then only the first term contributes. One finds that the probability per unit time for the detector to emit a particle of energy E is again (10.21) but with $f(\Delta E)$ replaced by $f(-\Delta E)$. From (10.19) it follows that

$$f(-\Delta E)=e^{2\pi\Delta E/H} f(\Delta E) . \quad (10.22)$$

An identical relation holds for the absorption and emission of antiparticles. The relation (10.22) between the rates of absorption and emission is precisely the condition that the detector experiences a thermal spectrum at the de Sitter temperature $H/2\pi$. This phenomenon was first discovered by Gibbons and Hawking in the context of scalar field theory in de Sitter space.[27]

The distribution has a denominator of the Fermi-Dirac form, but is not exactly Planckian because of the term $\frac{1}{4}$ which is added to $(\Delta E)^2/H^2$ in (10.19). This is to be compared with the results of Sciama, Candelas, and Deutsch.[28] In the case of massless spin s fields in the Rindler wedge, they observed a thermal spectrum but with a density-of-states factor of the form ν^2+s^2. They state that the reason one normally expects this factor to be just ν^2, is that it is computed on the assumption that the wavelengths involved are small compared to the size of the container. One would thus expect that in our case, the deviation from the Planck form is essentially due to the finite spatial size of de Sitter space.

The massive case is considerably more involved than the massless case considered here and we have not carried out explicit calculations. However, since the thermal nature of the spectrum depends only on very general properties of the Green's functions, one would expect the results obtained in the massless case also to hold in the massive case.

XI. DISCUSSION AND CONCLUSIONS

We have constructed a quantum cosmological model with a fermion matter source without neglecting any of the modes of the fermion field. An important ingredient in our approach was the use of the holomorphic representation for the fermion fields, in which the quantum state of the fermion modes is described by a set of wave functions which are analytic functions of odd elements of a Grassmann algebra. This allowed us to treat the problem in a manner very similar to the case of bosonic perturbations, considered by Halliwell and Hawking.[8] As in the bosonic case, our main result is that the fermion modes enter the Lorentzian region in the ground state.

In ordinary quantum mechanics, one may define the ground-state wave function by a Euclidean path integral over paths $x(\tau)$ which begin at $x=0$ and end at the point specified by the argument of the wave function, where τ is Euclidean time.[29] In view of this, it is reasonably obvious that in the bosonic case, the Hartle-Hawking proposal picks out the ground-state wave functions for the inhomogeneous modes, since regularity of the matter fields demands that the inhomogeneous modes go to zero at the initial point of the paths. For the fermion modes on the other hand, it is not so obvious that the ground state is picked out, partly because the boundary value problem for fermions is more complicated, but also because one generally has less intuitive feel for fermionic systems than for bosonic ones. However, given the correct way to set up the boundary value problem, and given a little experience with the path-integral description of simple fermionic systems, one could argue that our conclusions about the fermionic modes are as obvious as the bosonic case. Indeed, one of the motivations for developing the holomorphic representation was to highlight the likenesses between bosonic and fermionic systems.[12] One way or another, the important point is that we have demonstrated the unity of ideas contained in the Hartle-Hawking proposal, in that we now know that it picks out the ground state for both the bosonic modes and the fermionic modes.

In Sec. VIII given the above initial conditions, we evolved the wave functions using the time-dependent Schrödinger equations and found that in the massless case, the modes remained in the ground state. This is what one would expect, bearing in mind the conformal invariance of the massless theory and the fact that the background metric is conformally static. In the massive case, the fermion modes evolved away from their ground state, thus leading to particle production. Defining particle states using Hamiltonian diagonalization, we calculated the average number of particles produced and found it to be finite and of order H^3 per unit volume. What is notable is not the actual number, but the fact that it is finite, since the technique of Hamiltonian diagonalization is known to give divergent results in the case of scalar particle production.[30] The fact that our result is finite is essentially due to the initial conditions. The finiteness of the particle production depends on the behavior of the very-high-frequency modes. These modes never leave the horizon in our model, and thus remain in their initial state. Since this state is the ground state, they make no, or at least little, contribution to the particle production.

In Ref. 8, the effect of the scalar-field perturbations on the isotropy of the microwave background was investigated. This was done by choosing a gauge in which the perturbations in the scalar field are zero, but the scalar perturbations in the metric are nonzero. The resulting inhomogeneity in the metric on the surface of last scattering manifests itself as anisotropy in the microwave background. The above gauge choice—the "no density perturbations" gauge—is possible because there is a coupling between the inhomogeneous modes of the metric and the inhomogeneous modes of the scalar field. This coupling occurs because one writes the scalar field $\Phi(\mathbf{x},t)=\phi(t)+\delta\phi(\mathbf{x},t)$, where $\phi(t)$ is the "large" homogeneous mode and $\delta\phi$ is a small perturbation. In the expansion of the scalar field action to quadratic order in the perturbations, one thus encounters couplings of the form $\phi(t)\delta\phi(\mathbf{x},t)\delta g(\mathbf{x},t)$, where δg is the metric perturbation. In the fermion case, however, all the modes of the field are inhomogeneous perturbations—there is no large homogeneous mode analogous to $\phi(t)$. It follows that to this order, there is no coupling between the inhomogeneous modes of the metric and the fermion field modes; thus we conclude that the fermion modes have negligible effect on the isotropy of the microwave background. Similarly, one would expect that their effect on density perturbations is also negligible. The fundamental reason for both of these results is presumably the exclusion principle, which forbids too much fermionic matter from being assembled in one place.

In Sec. IX, we calculated the back reaction of the perturbation modes on the homogeneous minisuperspace background and found that it has a negligible effect in both the massless and massive cases. In the massless case, the back reaction was just the energy density of the vacuum, as one would expect, since the field remains in its ground state. In the massive case it was found to be the same order of magnitude as the energy density of the particles produced, again as one would expect. The fact that the back reaction is small implies that our calculation of the evolution of the wave function and particle production is self-consistent, since the behavior of the homogeneous background is not significantly modified.

It is worth noting that for a particular choice of operator ordering for the fermionic variables, there is no back-reaction term in the massless case. This is because one may shift the whole spectrum of the Hamiltonian by an arbitrary multiple of ν by taking an appropriate linear combination of $x\bar{x}$ and $\bar{x}x$ in place of the combination (5.4). In particular, one may shift the eigenvalue of the ground state to zero, whence the back reaction, which is essentially the vacuum energy density, is zero. One may therefore wonder why we did not do this. The point is, however, that there is only one particular choice of operator ordering which has zero vacuum energy—an arbitrarily made choice will lead to a back-reaction term. The choice (5.4) firstly, is quite a natural choice, and secondly, is representative of the general case. Strictly, one should go to the Hamiltonian from the path integral and derive the measure and, hence, the operator ordering, but this is beyond the scope of this paper and is clearly not necessary for the applications considered here.

The calculation of the back reaction requires a regularization procedure to subtract off vacuum energy divergences. As discussed in Sec. IX, consistent application of this approach leads to higher-derivative gravitational theories, and hence possibly to a totally different quantum theory. One of the initial motivations for including fermions in quantum cosmology was the hope that the vacuum energies of the bosonic and fermionic modes would cancel, so that, in the one-loop approximation considered here, there would be no need for regularization and inclusion of counterterms in the action. However, even in locally supersymmetric theories with equal numbers of bosonic and fermionic degrees of freedom, the one-loop divergences do not in general cancel. Nevertheless, such theories, do have attractive properties, such as the possibility of the operator ordering being defined naturally by supersymmetry.[13] A model based on $N=1$ supergravity coupled to matter is currently under investigation.[31]

In Sec. X, we constructed a model of a particle detector designed to detect fermions and examined its response to the quantum state selected by the Hartle-Hawking proposal, in the approximation of a quantized fermion field on a classical de Sitter background. We showed that the detector experiences a thermal spectrum at the de Sitter temperature $H/2\pi$, with the correct Fermi-Dirac denominator. The spectrum deviated from the Planck form, however, as a consequence of the finite spatial size of de Sitter space.

After completion of this work, we became aware of similar works by Hinton[32] and Takagi,[33] who also considered models of particle detectors designed to detect fermions. The models considered by these authors are different from ours in that they used a trilinear interaction Lagrangian of the form $m\bar{\psi}\psi$, where ψ is a four-component spinor and m is the usual monopole moment. Their calculations are thus more complicated than ours, since they involve four-point functions rather than two-point functions. With the bilinear interaction Lagrangian (10.1), our derivation is really no more complicated than the scalar case. Hinton and Takagi found, as we did, that the detector response depends essentially on response functions of the form in (10.5) and (10.6); hence, their results are very similar to ours.

An important step in our derivation is the argument that the state, and hence the Green's function, is de Sitter invariant. It is worth remarking at this point that the argument is clearly not restricted to fermion matter sources and one would expect it to be true of any type of matter field. In other words, if for a given matter field, one or more de Sitter-invariant vacuum states exist, then the state picked out by the Hartle-Hawking proposal will correspond to one of these states.

It is of interest to enquire as to the relation between the definition of particles according to the Hamiltonian diagonalization technique of Sec. VIII, which yielded zero-particle production in the massless case, and the detector response of Sec. X, which was quite definitely nonzero. In instantaneous Hamiltonian diagonalization, one considers a spacelike section of de Sitter space and considers the quantum state of the field over the entire surface. We showed in fact, that all the modes of the field on such a surface are in the ground state, in the massless case. An observer moving along a geodesic, however, cannot measure all of these modes, since some of them have wavelengths greater than the size of his horizon. It is this lack of knowledge of the entire state of the field which causes him to perceive a thermal spectrum.

In conclusion, we have constructed a quantum cosmological model which incorporates all the modes of a fermion matter source. Our main result is that the boundary conditions proposal of Hartle and Hawking implies that the fermion modes start out in their ground state. We have also show how to calculate quantities of interest such as the back reaction and the particle production. From these calculations we may conclude that fermions do not play a significant dynamical role in quantum cosmology.

One may anticipate that their main role is technical in nature, such as, the cancellation of vacuum energy divergences, but this is a topic for future investigation.

ACKNOWLEDGMENTS

Throughout the course of this work we have benefited from conversations with Bruce Allen, Gary Gibbons, and Stephen Hawking. We are also grateful to Paul Davies and Ian Moss for bringing Refs. 32 and 33 to our attention.

APPENDIX A: SPINOR CONVENTIONS

We use Lorentzian conventions, in which the space-time metric $g_{\mu\nu}$ is given in terms of the tetrad e^a_μ of basis forms by

$$g_{\mu\nu}=\eta_{ab}e^a_\mu e^b_\nu \ , \tag{A1}$$

where $\eta_{ab}=\mathrm{diag}(-1,1,1,1)$ is the Minkowski metric. Space-time indices $\mu,\nu,\ldots$ and tetrad indices $a,b,\ldots$ run from 0 to 3. Space-time indices are raised and lowered using $g^{\mu\nu}$ and $g_{\mu\nu}$, while tetrad indices are raised and lowered using η^{ab} and η_{ab}. Because of the choice of signature $+2$ for the tetrad metric, our two-component spinor conventions differ from those of Penrose and Rindler.[34]

Tetrad vectors are related to spinors via the Infeld—van der Waerden translation symbols $\sigma_a^{AA'}$, here taken to be

$$\sigma_0=-\frac{1}{\sqrt{2}}I, \ \ \sigma_i=\frac{1}{\sqrt{2}}\Sigma_i \ \ (i=1,2,3) \ , \tag{A2}$$

where Σ_i are the Pauli matrices. For example, the basis forms e^a_μ correspond to the spinor-valued forms

$$e_\mu^{AA'}=e^a_\mu\sigma_a^{AA'} \ , \tag{A3}$$

where spinor indices $A,B,\ldots$ take values 0,1 and A',B' take values $0',1'$. Spinor indices are raised and lowered by the alternating spinors ϵ^{AB}, ϵ_{AB}, $\epsilon^{A'B'}$, $\epsilon_{A'B'}$, each of which is given by the matrix $\left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$. For example,

$$\lambda^A=\epsilon^{AB}\lambda_B, \ \ \lambda_A=\lambda^B\epsilon_{BA} \ , \tag{A4}$$

$$\zeta^{A'}=\epsilon^{A'B'}\zeta_{B'}, \ \ \zeta_{A'}=\zeta^{B'}\epsilon_{B'A'} \ . \tag{A5}$$

The inverse of Eq. (A3) is then

$$e^a_\mu=-\sigma^a_{AA'}e_\mu^{AA'} \ . \tag{A6}$$

Under an infinitesimal local Lorentz transformation acting on the tetrad indices of e^a_μ, the $e_\mu^{AA'}$ change by

$$\delta e_\mu^{AA'}=N^A_B e_\mu^{BA'}+\bar{N}\,^{A'}_{B'}e_\mu^{AB'} \ , \tag{A7}$$

where $N_{AB}=N_{(AB)}$ is symmetric and $\bar{N}_{(A'B')}$ is its conjugate. A similar transformation law holds for a spinor of any index type. Thus, any tensor $T^{\mu_1\cdots\mu_n}$ defines a spinor

$$T^{AA'\cdots DD'}=e_{\mu_1}^{AA'}\cdots e_{\mu_n}^{DD'}T^{\mu_1\cdots\mu_n} \tag{A8}$$

with inverse relation

$$T^{\mu_1\cdots\mu_n}=(-1)^n e^{\mu_1}_{AA'}\cdots e^{\mu_n}_{DD'}T^{AA''\cdots DD'} \ . \tag{A9}$$

In particular the metric $g_{\mu\nu}$ corresponds to the spinor—$\epsilon_{AB}\epsilon_{A'B'}$:

$$g^{\mu\nu}e_\mu^{AA'}e_\nu^{BB'}=-\epsilon^{AB}\epsilon^{A'B'} \ , \tag{A10}$$

while the inverse relation [spinor version of Eq. (A1)] is

$$g_{\mu\nu}=-\epsilon_{AB}\epsilon_{A'B'}e_\mu^{AA'}e_\nu^{BB'} \ . \tag{A11}$$

In a Hamiltonian decomposition, the spatial tetrad forms $e_i^{AA'}$ determine the spatial metric

$$h_{ij}=-e_{AA'i}e_j^{AA'} \ , \tag{A12}$$

where $i,j,\ldots=1,2,3$. They also determine the spinor version $n^{AA'}$ of the future-directed unit timelike normal n^μ to a spacelike hypersurface $x^0=\mathrm{const}$, through the relations

$$n_{AA'}e_i^{AA'}=0, \ \ n_{AA'}n^{AA'}=1 \ . \tag{A13}$$

The normal obeys the useful relations

$$n_{AA'}n^{AB'}=\tfrac{1}{2}\epsilon^{B'}_{A'}, \ \ n_{AA'}n^{BA'}=\tfrac{1}{2}\epsilon^B_A \ . \tag{A14}$$

Provided that the tetrad e^a_μ is real, corresponding to a Lorentzian metric $g_{\mu\nu}$, the $e_\mu^{AA'}$ will be Hermitian, as will $n^{AA'}$.

Passage to the Euclidean regime can be achieved by rotating the basis from $e^0_\mu\to -ie^0_\mu$, while still using Lorentzian conventions for the tetrad metric η_{ab} and for spinors: the space-time metric $g_{\mu\nu}$ then becomes positive definite. In the Hamiltonian treatment, one can instead rotate the lapse function $N\to -iN$. In that case, it is convenient to define the Euclidean normal spinor ${}_en^{AA'}=-in^{AA'}$, which corresponds to a unit spacelike normal vector ${}_en^\mu$. It will be anti-Hermitian, obeying ${}_e\bar{n}^{AA'}=-{}_en^{AA'}$.

APPENDIX B: SPINOR HARMONICS ON THE THREE-SPHERE

We seek a complete set of harmonics for the expansion of any spinor field on the three-sphere S^3. Scalar, vector and tensor harmonics on S^3 have been constructed by Lifshitz and Khalatnikov.[35] We will use their method to construct spinor harmonics.

One begins by considering flat Euclidean four-space E^4, which in Cartesian coordinate x^μ, $\mu=0,1,2,3$, has metric

$$ds^2=\delta_{\mu\nu}dx^\mu dx^\nu \ . \tag{B1}$$

The metric may also be written in spherical coordinates r,χ,θ,ϕ:

$$ds^2=dr^2+r^2d\Omega_3{}^2 \ , \tag{B2}$$

where $d\Omega_3{}^2$ is the metric on S^3:

$$d\Omega_3{}^2=d\chi^2+\sin^2\chi(d\theta^2+\sin^2\theta\, d\phi^2) \ . \tag{B3}$$

Harmonics on S^3 are then obtained by considering the following homogeneous polynomials in Cartesian coordinates $x^\mu=-\sigma^\mu_{AA'}x^{AA'}$:

$$r^n\rho_A(\chi,\theta,\phi)=T_{AA_1\cdots A_nA_1'\cdots A_n'}x^{A_1A_1'}\cdots x^{A_nA_n'} \ , \tag{B4}$$

where $T_{AA_1 \cdots A_n A'_1 \cdots A'_n}$ is a constant spinor of rank $(2n+1)$, totally symmetric in all its indices and hence traceless. It follows that (B4) is a solution to the Weyl equation in E^4:

$$e^{AA'\mu}D_\mu(r^n\rho_A)=0 \;, \tag{B5}$$

since D_μ is simply ∂_μ in Cartesian coordinates. This equation is now rewritten in terms of spherical coordinates (B2), yielding the result

$$\frac{1}{r}e^{AA'j\,(3)}D_j(r^n\rho_A)+{}_en^{AA'}\left[\frac{\partial}{\partial r}+\frac{3}{2r}\right](r^n\rho_A)=0 \;, \tag{B6}$$

where ${}^{(3)}D_i$ is the spinor covariant derivative on S^3. $i=1,2,3$ and ${}_en^{AA'}$ is the Euclidean normal. Differentiating out (B6), and then restricting to the three-sphere, one obtains the desired eigenvalue equation for spinor harmonics on the S^3:

$$e^{AA'j\,(3)}D_j\rho_A=-(n+\tfrac{3}{2})_en^{AA'}\rho_A \;. \tag{B7}$$

Spinor harmonics with primed indices may be obtained by considering the polynomials

$$r^n\sigma_{A'}(\chi,\theta,\phi)=S_{A_1\cdots A_nA'_1\cdots A'_nA'}x^{A_1A'_1}\cdots x^{A_nA'_n} \;, \tag{B8}$$

where $S_{A_1\cdots A_nA'_1\cdots A'_nA'}$ is a spinor of rank $(2n+1)$ which is totally symmetric on all its indices. As above, it is easily shown that $\sigma_{A'}$ satisfies

$$e^{AA'j\,(3)}D_j\sigma_{A'}=-(n+\tfrac{3}{2})_en^{AA'}\sigma_{A'} \;. \tag{B9}$$

Harmonics with eigenvalues of the opposite sign may be constructed from ρ_A and $\sigma_{A'}$ by complex conjugation. One thus obtains new harmonics $\bar\rho_{A'}$ and $\bar\sigma_A$ satisfying

$$e^{AA'j\,(3)}D_j\bar\rho_{A'}=(n+\tfrac{3}{2})_en^{AA'}\bar\rho_{A'} \;, \tag{B10}$$

$$e^{AA'j\,(3)}D_j\bar\sigma_A=(n+\tfrac{3}{2})_en^{AA'}\bar\sigma_A \;. \tag{B11}$$

These equations follow from (B7) and (B9) using the fact that the Euclidean normal is anti-Hermitian (see Appendix A). Harmonics with eigenvalues of the opposite sign may also be constructed from ρ_A and $\sigma_{A'}$ by multiplication by ${}_en^{AA'}$. Using the property that ${}^{(3)}D_j({}_en^{AA'})=0$, one thus obtains harmonics ${}_en_{A'}^{B}\rho_B$ and ${}_en_A^{B'}\sigma_{B'}$, satisfying

$$e^{AA'j\,(3)}D_j({}_en_{A'}^{B}\rho_B)=(n+\tfrac{3}{2})_en^{AA'}({}_en_{A'}^{B}\rho_B) \;, \tag{B12}$$

$$e^{AA'j\,(3)}D_j({}_en_A^{B'}\sigma_{B'})=(n+\tfrac{3}{2})_en^{AA'}({}_en_A^{B'}\sigma_{B'}) \;. \tag{B13}$$

There is a linear relation between the harmonics obtained by these two different methods, as we shall show, so it does not matter which ones we use. We choose to restrict attention to the ones obtained by complex conjugation, $\bar\sigma_A$ and $\bar\rho_{A'}$.

Each harmonic will be labeled by an integer n ($n=0,1,2,\ldots$) corresponding to its eigenvalue. Furthermore, it follows from (B4) and (B8) that the degeneracy for each n is $(n+1)(n+2)$. Each harmonic will thus be labeled by a second integer p, where $p=1,2,\ldots(n+1)(n+2)$.

The harmonics ρ_A^{np} and $\bar\sigma_A^{np}$ or $\sigma_{A'}^{np}$ and $\bar\rho_{A'}^{np}$ form a complete set for the expansion of any spinor field or its Hermitian conjugate on S^3.

Orthogonality and Normalization. We denote the integration measure on S^3 by

$$d\mu=\sin^2\chi\,\sin\theta\,d\chi\,d\theta\,d\phi \;. \tag{B14}$$

Using the eigenvalue equations (B7) and (B9)—(B11) one may derive the orthogonality relations

$$\int d\mu\,\rho_A^{np}n^{AA'}\sigma_{A'}^{mq}=0 \;, \tag{B15}$$

$$\int d\mu\,\rho_A^{np}\epsilon^{AB}\sigma_B^{mq}=0 \;, \tag{B16}$$

for all n,p,m,q. Hermitian conjugation yields similar equations. In these and the following orthogonality relations, it is more convenient to use the Lorentzian normal $n^{AA'}$, rather than its Euclidean counterpart ${}_en^{AA'}$, since it is the former that appears in the expressions we wish to expand.

One may also derive the relations

$$\int d\mu\,\rho_A^{np}n^{AA'}\bar\rho_{A'}^{mq}=\delta^{nm}H_n^{pq} \;, \tag{B17}$$

$$\int d\mu\,\rho_A^{np}\epsilon^{AB}\rho_B^{mq}=\delta^{nm}A_n^{pq} \;, \tag{B18}$$

$$\int d\mu\,\bar\sigma_A^{np}n^{AA'}\sigma_{A'}^{mq}=\delta^{nm}H_n^{pq} \;, \tag{B19}$$

$$\int d\mu\,\sigma_{A'}^{np}\epsilon^{A'B'}\sigma_{B'}^{mq}=\delta^{nm}A_n^{pq} \;. \tag{B20}$$

Hermitian conjugation of (B18) and (B20) yields two more sets of relations. The sets of numbers H_n^{pq} and A_n^{pq} depend on the normalization of the harmonics, which we discuss below. First, we show that they are related by a consistency condition.

From (B17), one may see that for each n, the H_n^{pq} constitute the elements of an Hermitian matrix, denoted by H_n, of dimension $(n+1)(n+2)$. Likewise, the A_n^{pq} constitute the elements of an antisymmetric matrix A_n, of the same dimension. Any spinor field $\Psi_A(\mathbf{x})$ on S^3 may be expanded in harmonics:

$$\Psi_A(\mathbf{x})=\sum_{np}[a_{np}\rho_A^{np}(\mathbf{x})+b_{np}\bar\sigma_A^{np}(\mathbf{x})] \tag{B21}$$

for some coefficients a_{np}, b_{np}. The coefficients a_{np} may then be determined using either (B17) or (B18). From (B17) one obtains

$$a_{np}=\int d\mu\,\Psi_A(\mathbf{x})n^{AA'}\sum_q\bar\rho_{A'}^{nq}(H_n{}^{-1})^{qp} \tag{B22}$$

and from (B18) one obtains

$$a_{np}=\int d\mu\,\Psi_A(\mathbf{x})\epsilon^{AB}\sum_q\rho_B^{nq}(A_n{}^{-1})^{qp} \;. \tag{B23}$$

Since Ψ_A is arbitrary, one may equate the integrands of (B22) and (B23), and after some rearrangement, one obtains

$$\bar\rho_{A'}^{np}=-2n_{A'}^{A}\sum_q\rho_A^{nq}(A_n{}^{-1}H_n)^{qp} \;. \tag{B24}$$

The existence of the inverses $A_n{}^{-1}$, $H_n{}^{-1}$ follows from the linear independence of the harmonics in the degeneracy label p. Taking the Hermitian conjugate of this equation and then substituting into the right-hand side of

(B24) one obtains the consistency condition

$$A_n{}^{-1}H_nA_n{}^{-1}H_n=-\tfrac{1}{2}l_n\ , \quad \text{(B25)}$$

where l_n is the $n\times n$ identity matrix. Note that (B24) also provides the relation between the harmonics $\bar{\rho}_{A'}$ and ${}_en_{A'}^{A}\rho_A$ mentioned earlier.

We now consider the normalization of the harmonics. Since H_n is Hermitian, it may be diagonalized by taking linear combinations of the ρ_A^{np} for each n, and then by suitable normalization, we may take $H_n^{pq}=\delta^{pq}$. This choice still leaves the freedom to bring the antisymmetric matrix A_n to block-diagonal form, with blocks $\left(\begin{smallmatrix}0 & a\\ -a & 0\end{smallmatrix}\right)$ for some constant a. By multiplying the harmonics by suitable phases, a may be taken to be real. The consistency condition (B25) then implies that $a=\pm\sqrt{2}$. We make the choice $a=+\sqrt{2}$, and write $A_n^{pq}=\sqrt{2}C_n^{pq}$ where C_n^{pq} is block diagonal with blocks $\left(\begin{smallmatrix}0 & 1\\ -1 & 0\end{smallmatrix}\right)$,

Finally, we give the coefficients $\alpha_n^{pq},\beta_n^{pq}$ introduced in Sec. IV to avoid couplings between different values of p in the expansion of the action. For each n, they may be regarded as block-diagonal matrices α_n, β_n, say, of dimension $(n+1)(n+2)$ with blocks $\left(\begin{smallmatrix}1 & 1\\ 1 & -1\end{smallmatrix}\right)$ for α_n and $\left(\begin{smallmatrix}1 & -1\\ -1 & -1\end{smallmatrix}\right)$ for β_n. They satisfy the relations

$$\alpha_n{}^2=\beta_n{}^2=2l_n,\ \ \alpha_n\beta_n=-\beta_n\alpha_n=2C_n\ , \quad \text{(B26)}$$

$$\alpha_nC_n\beta_n=-\beta_nC_n\alpha_n=2l_n\ , \quad \text{(B27)}$$

where C_n is the matrix of coefficients C_n^{pq}.

To summarize, the only nonzero orthogonality relations are (B17)—(B20), and their Hermitian conjugates, where $H_n^{pq}=\delta^{pq}$ and $A_n^{pq}=\sqrt{2}C_n^{pq}$. These relations provide sufficient information to perform the derivations presented in the main text.

APPENDIX C: SOLUTION OF THE DIRAC AND SCHRÖDINGER EQUATIONS

In this appendix we give exact solutions to the Dirac equation (7.9) and the Schrödinger equation (6.8), taking the homogeneous background to be an exact de Sitter space.

The x,y dependence of the perturbation wave function Ψ_n is contained entirely in the saddle-point approximation to the path integral

$$\Psi_n(\alpha,\phi,x,y)=\exp(-\tilde{I}_n^{\,\text{cl}})\ , \quad \text{(C1)}$$

where $\tilde{I}_n^{\,\text{cl}}$ is the Euclidean action of the solution $\alpha(\tau),\phi(\tau),x(\tau),y(\tau)$ to the Euclidean field equations satisfying the initial conditions described in Sec. VII and ending at the point (α,ϕ,x,y). To a good approximation, the Euclidean version to the field equations for the background (3.2)—(3.4) have the solutions

$$\phi(\tau)=\text{const},\ \ e^{\alpha(\tau)}=\frac{1}{H}\sin(H\tau) \quad \text{(C2)}$$

for large $|\phi|$ where $H=M\phi$, which is just the Euclidean section of de Sitter space. Moreover, this solution will continue to hold when continued into the Lorentzian region, for the duration of the inflationary phase. This continuation may be achieved by writing $\tau=\pi/2H+it$, yielding $e^{\alpha}=H^{-1}\cosh(Ht)$.

One may proceed, as in Sec. VII, to derive the following form for the wave function (C1):

$$\Psi_n(\alpha,\phi,x,y)=1+\frac{mxy}{f'(\tau)/f(\tau)+\nu}\ , \quad \text{(C3)}$$

where $f(\tau)$ is a solution to the Euclidean-Dirac equation (7.9) satisfying $f(0)=0$. In Sec. VII, we used the adiabatic approximation to solve (7.9), partly to avoid an excess of technical details, but also to indicate the generality of our approach, yielding the result $f'(\tau)/f(\tau)\approx\omega$. Here, however, given the form (C2) for e^{α}, one may solve exactly for $f(\tau)$ in terms of hypergeometric functions, thus allowing a check of the adiabatic approximation and also providing the solution in regions where it is not valid.

The Euclidean-Dirac equation (7.9) is

$$\frac{d^2f}{d\tau^2}-\left[-\frac{d\nu}{d\tau}+\nu^2+m^2\right]f=0 \quad \text{(C4)}$$

recalling that $\nu=(n+\frac{3}{2})e^{-\alpha}$. Introduce the variable $z=\frac{1}{2}(1-\cos H\tau)$ and let

$$y(z)=\exp\left[\int\nu\,d\tau\right]=(n+\tfrac{3}{2})\ln(\tan H\tau)f(\tau)\ . \quad \text{(C5)}$$

Then (C4) becomes

$$z(1-z)\frac{d^2y}{dz^2}-(n+1+z)\frac{dy}{dz}-\frac{m^2}{H^2}y=0\ . \quad \text{(C6)}$$

Of the two linearly independent solutions, only one is consistent with the initial condition $f(0)=0$. This is

$$y(z)=z^{n+2}{}_2F_1\left[n+2+\frac{im}{H},n+2-\frac{im}{H};n+3;z\right]\ , \quad \text{(C7)}$$

where ${}_2F_1$ is the hypergeometric function.[36]

From (C5), the denominator in (C3) is given by

$$\frac{f'(\tau)}{f(\tau)}+\nu=\frac{H}{2}\sin(H\tau)\frac{y'(z)}{y(z)} \quad \text{(C8)}$$

and (C7) then implies that

$$\frac{y'(z)}{y(z)}=\frac{(n+2)}{z}+\frac{(n+2)^2+m^2/H^2}{(n+3)(1-z)}\times\frac{{}_2F_1(1-im/H,1+im/H;n+4;z)}{{}_2F_1(1-im/H,1+im/H;n+3;z)}\ , \quad \text{(C9)}$$

where we have used the following property of the hypergeometric function:[36]

$${}_2F_1(a,b,c)=(1-z)^{c-a-b}{}_2F_1(c-a,c-b;c;z)\ . \quad \text{(C10)}$$

The solution (C8) may be continued from the Euclidean region ($e^{\alpha}<H^{-1}$) to the Lorentzian region ($e^{\alpha}>H^{-1}$). For modes inside the horizon, i.e., for $\nu\gg H$, one may use the fact that $|z|/n$ is small to expand the hypergeometric functions in (C9). One finds that, in an obvious notation,

$$\frac{{}_2F_1(n+4)}{{}_2F_1(n+3)}=1-\frac{(1+m^2/H^2)}{(n+4)(n+3)}z+O\left[\frac{z^2}{n^3}\right] \quad \text{(C11)}$$

and through (C8) and (C9) one obtains

$$\frac{f'(\tau)}{f(\tau)}+\nu=2\nu+\frac{1}{2}\frac{m^2e^\alpha}{n}-iH+\cdots \ . \tag{C12}$$

Assuming $\nu\gg m$, one may write

$$\omega+\nu=2\nu+\frac{1}{2}\frac{m^2e^\alpha}{n}+\cdots \ . \tag{C13}$$

It follows that, inside the horizon, the perturbation wave function (C3) is

$$\Psi_n(\alpha,\phi,x,y)=1+\frac{mxy}{\omega+\nu-iH} \ . \tag{C14}$$

Thus, apart from a small imaginary part, the perturbation wave functions are indeed of the ground-state form inside the horizon, as was indicated in Sec. VII. Note that it is reasonable to assume that $\nu\gg m$ inside the horizon, i.e., for $\nu\gg H$ since a typical value of H is the grand-unified-theory scale, 10^{-5} (Planck units),[8] whereas a typical value of m would be for example, the electron mass, about 10^{-22}. Outside the horizon, however, $\nu\ll H$ and there will be modes for which $\nu\ll m$.

The solution to the Dirac equation also yields the solution to the Schrödinger equation. It was shown in Sec. VIII that (8.1) is a solution to the Schrödinger equation providing $u(t)$ satisfies the Riccati equation (8.4). Comparing (8.1) with (C3). It is easily seen that $u(t)=f'(\tau)/f(\tau)$ is a solution to (8.2), where $\tau=\pi/2H+it$ and moreover, this solution satisfies the correct initial conditions, since we have shown that $f'(\tau)/f(\tau)\approx\omega$ inside the horizon. The expression (C3) thus provides the correct solution to the Schrödinger equation throughout the inflationary era. The solution outside the horizon may thus be obtained by expanding the hypergeometric functions in (C9) for $|z|\rangle n$. One obtains

$$\frac{{}_2F_1(n+4)}{{}_2F_1(n+3)}=\frac{n+3}{n+2+im/H}\left[1-\frac{1+im/H}{1-2im/H}\frac{1}{z}+O\left[\frac{n}{z^2}\right]\right] \tag{C15}$$

and one may then show through (C8) that the solution to the Riccati equation is

$$u+\nu=-m\tanh(Ht)-\frac{2im}{H}\left[1-\frac{2im}{H}\right]^{-1}\nu \tag{C16}$$

for $\nu\ll H$.

APPENDIX D: DERIVATION OF THE SEMICLASSICAL EINSTEIN EQUATIONS

In this appendix we indicate how the semiclassical Einstein equations emerge from the semiclassical limit of the Wheeler-DeWitt equation. The wave function (6.2) obeys the Wheeler-DeWitt equation (6.1). This was decomposed into the background Wheeler-DeWitt equation (6.4) and the Schrödinger equation (6.7). It is convenient to expand Ψ_0 in a WKB expansion

$$\Psi_0=\exp(iS_0+iS_1+\cdots) , \tag{D1}$$

where the ellipses indicate higher-order terms in the WKB expansion. Inserting this in (6.4), one obtains

$$-i\nabla^2S_0+(\nabla S_0)^2+2(\nabla S_0)\cdot(\nabla S_1)+e^{-3\alpha}V+\sum_{np}\Omega_{np}+\cdots=0 \ . \tag{D2}$$

S_0 is chosen to satisfy the Hamilton-Jacobi equation

$$(\nabla S_0)^2+e^{-3\alpha}V=0 \ . \tag{D3}$$

One may then split the remaining terms in (D2) into real and imaginary parts, yielding

$$-\nabla^2S_0+2\frac{\partial}{\partial t}(\mathrm{Im}S_1)+\mathrm{Im}\left[\sum_{np}\Omega_{np}\right]=0 , \tag{D4}$$

$$2\frac{\partial}{\partial t}(\mathrm{Re}S_1)+\mathrm{Re}\left[\sum_{np}\Omega_{np}\right]=0 , \tag{D5}$$

where $\partial/\partial t$ is defined by (6.6). $\mathrm{Im}S_1$ gives the usual WKB prefactor, with a correction due to the term $\mathrm{Im}(\sum\Omega_{np})$. We are concerned solely with corrections to the oscillatory part of the wave function, which are given by $\mathrm{Re}S_1$.

Consider now the Schrödinger equations (6.7). It is useful to define a new wave function $\tilde\Psi_{np}$, given by

$$\tilde\Psi_{np}=\exp\left[i\beta_{np}(t)+\tfrac{1}{2}\int\mathrm{Im}\Omega_{np}dt\right]\Psi_{np} , \tag{D6}$$

where $\beta_{np}(t)$ is a (real) phase, which is to be chosen. $\tilde\Psi_{np}$ satisfies the condition $(\tilde\Psi_{np},\tilde\Psi_{np})=1$, which Ψ_{np} did not satisfy, since we made the ansatz (8.1). $\tilde\Psi_{np}$ satisfies the Schrödinger equation

$$\left[H_{np}-\tfrac{1}{2}\mathrm{Re}\Omega_{np}-\frac{d\beta_{np}}{dt}\right]\tilde\Psi_{np}=i\frac{\partial\tilde\Psi_{np}}{\partial t} \ . \tag{D7}$$

It is convenient to choose β_{np} such that

$$\tfrac{1}{2}\mathrm{Re}\Omega_{np}+\frac{d\beta_{np}}{dt}=(\tilde\Psi_{np},H_{np}\tilde\Psi_{np}) \ . \tag{D8}$$

Equation (D8) may now be used to eliminate Ω_{np} from (D5), yielding

$$2\frac{\partial}{\partial t}(\mathrm{Re}S_1)-2\sum_{np}\frac{d\beta_{np}}{dt}+\sum_{np}(\tilde\Psi_{np},H_{np}\tilde\Psi_{np})=0 \ . \tag{D9}$$

β_{np} may then be absorbed into S_1 by defining

$$\tilde S_1=S_1-\sum_{np}\beta_{np} \ . \tag{D10}$$

$\tilde S_1$ thus satisfies

$$2\frac{\partial}{\partial t}(\mathrm{Re}\tilde S_1)+2\langle H_f\rangle=0 , \tag{D11}$$

where $\langle H_f\rangle$ is the expectation value of H_f in the state $\prod_{np}\tilde\Psi_{np}$.

The point of all these seemingly complicated redefinitions is quite simply to transfer part of the phase of Ψ_0 to the perturbation wave functions, which originally had no phase. For, as a result of these redefinitions, the total wave function may now be written

$$\Psi=\exp(iS)\prod_{np}\tilde\Psi_{np} , \tag{D12}$$

where $S = S_0 + \tilde{S}_1 + \cdots$ and $\tilde{\Psi}_{np}$ satisfies

$$(H_{np} - \langle H_{np} \rangle)\tilde{\Psi}_{np} = i\frac{\partial \tilde{\Psi}_{np}}{\partial t} . \quad \text{(D13)}$$

Most importantly, the real part of (D2) now has the form

$$(\nabla \operatorname{Re} S)^2 + e^{-3\alpha} V + 2\langle H_f \rangle + \cdots = 0 . \quad \text{(D14)}$$

Equation (D14) is the Hamilton-Jacobi equation for the background variables modified by a back-reaction term. One has the relation

$$\langle H_f \rangle = \frac{1}{2} \int d^3x h^{1/2} \langle T_{00} \rangle , \quad \text{(D15)}$$

thus (D14) corresponds to the time-time component of the semiclassical Einstein equations.

It is to be noted that only by splitting the phase of Ψ between Ψ_0 and the $\tilde{\Psi}_{np}$ in a particular way is one able to cast the real part of (D2) in the form of the semiclassical Hamilton-Jacobi equation. The Wheeler-DeWitt equation (6.1) determines only the total phase, so the phase may be divided between Ψ_0 and Ψ_{np} in an arbitrary way, depending on what one wishes to regard as the background wave function and what one wishes to regard as a perturbation wave function. The choice made here is a desirable one since it achieves what we set out to achieve, namely, to show how the semiclassical Einstein equations emerge from the Wheeler-DeWitt equation. Similar derivations have previously been given by Moss[37] and Hartle.[38]

[1]S. W. Hawking, Astrophs. Cosmology, Pontificiae Academiae Scientarium Varia **48**, 563 (1982).
[2]J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).
[3]S. W. Hawking, Nucl. Phys. **B239**, 257 (1984).
[4]S. W. Hawking, in *Relativity, Groups and Topology II,* Les Houches Session XL, edited by B. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
[5]P. Amsterdamski, Phys. Rev. D **31**, 3073 (1985); U. Carow and S. Watamura, *ibid.* **32**, 1290 (1985); P. F. Gonzalez-Diaz, Phys. Lett. **159B**, 19 (1985); J. J. Halliwell, Nucl. Phys. **B266**, 228 (1986); S. W. Hawking and J. C. Luttrell, Nucl. Phys. **B247**, 250 (1984); Phys. Lett. **143B**, 83 (1984); G. T. Horowitz, Phys. Rev. D **31**, 1169 (1985); X. M. Hu and Z. C. Wu, Phys. Lett. **149B**, 87 (1984); I. G. Moss and W. A. Wright, Phys. Rev. D **29**, 1067 (1984); Z. C. Wu, Phys. Lett. **146B**, 307 (1984).
[6]C. J. Isham and J. E. Nelson, Phys. Rev. D **10**, 3226 (1974).
[7]T. Christodoulakis and J. Zanelli, Phys. Rev. D **29**, 2738 (1984).
[8]J. J. Halliwell and S. W. Hawking, Phys. Rev. D **31**, 1777 (1985).
[9]J. E. Nelson and C. Teitelboim, Ann. Phys. (N.Y.) **116**, 86 (1978).
[10]D. N. Page, Pennsylvania State University report, 1985 (unpublished).
[11]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[12]L. D. Faddeev and A. A. Slavnov, *Gauge Fields: Introduction to Quantum Theory* (Benjamin, New York, 1980); L. D. Faddeev, in *Methods in Field Theory,* Les Houches, 1975, Session XXVIII, edited by R. Balian and J. Zinn-Justin (North-Holland, Amsterdam, 1976).
[13]P. D. D'Eath, Phys. Rev. D **29**, 2199 (1984).
[14]R. Casalbuoni, Nuovo Cimento **33A**, 115 (1976)
[15]S. W. Hawking and Z. C. Wu, Phys. Lett. **151B**, 15 (1985).
[16]F. A. Berezin and M. S. Marinov, Ann. Phys. (N.Y.) **104**, 336 (1977).
[17]M. Henneaux and C. Teitelboim, Ann. Phys. (N.Y.) **143**, 127 (1982).
[18]M. F. Atiyah, V. K. Patodi, and I. M. Singer, Math. Proc. Camb. Philos. Soc. **77**, 43 (1975).
[19]L. H. Ford, Phys. Rev. D **14**, 3304 (1976).
[20]N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England, 1982).
[21]C. T. Hill, Fermilab report, 1984 (unpublished).
[22]G. T. Horowitz, Phys. Rev. D **31**, 1169 (1985); S. W. Hawking and J. C. Luttrell, Nucl. Phys. **B247**, 250 (1984).
[23]W. G. Unruh, Phys. Rev. D **14**, 870 (1976).
[24]B. S. DeWitt, in *General Relativity: An Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).
[25]B. Allen and C. A. Lutken, Tufts University report, 1986 (unpublished).
[26]B. Allen, Phys. Rev. D **32**, 3136 (1985).
[27]G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[28]D. W. Sciama, P. Candelas, and D. Deutsch, Adv. Phys. **30**, 327 (1981).
[29]S. Coleman, in *The Whys of Subnuclear Physics,* edited by A. Zichichi (Plenum, New York, 1979).
[30]L. Parker, in *Asymptotic Structure of Spacetime,* edited by F. P. Esposito and L. Witten (Plenum, New York, 1977).
[31]P. D. D'Eath and D. I. Hughes (in preparation).
[32]K. Hinton, Ph.D. thesis, University of Newcastle, 1984.
[33]S. Takagi, Prog. Theor. Phys. **74**, 501 (1985).
[34]R. Penrose and W. Rindler, *Spinors and Space-Time* (Cambridge University Press, Cambridge, England, 1984).
[35]E. M. Lifshitz and I. M. Khalatnikov, Adv. Phys. **12**, 185 (1963).
[36]*Handbook of Mathematical Functions,* edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1965).
[37]I. G. Moss, in *Proceedings of the Fourth Marcel Grossman Conference, Rome, 1985,* edited by R. Ruffini (North-Holland, Amsterdam, 1986).
[38]J. B. Hartle, Lectures at the Cargese Summer School on Relativistic Astrophysics, 1986 (unpublished).