Particle creation in de Sitter space

E. Mottola
Institute for Theoretical Physics, University of California, Santa Barbara, California 93106
(Received 15 August 1984)

In this, the first of a series of papers on quantum field theory in de Sitter spacetime, the invariant vacuum state appropriate for inflationary models of the early universe is identified and shown to decay due to the Hawking effect. The created pairs have an energy-momentum which leads to a first-order decrease of the effective cosmological constant, independently of any matter phase transition. A mechanism for dynamically relaxing $\Lambda_{\rm eff}\to 0$ is thereby suggested.

I. INTRODUCTION

In classical general relativity the possibility of adding a cosmological vacuum energy density, Λ, to Einstein's equations naturally raises the question of why this term is not present in the universe we observe ($\Lambda_{\rm obs}<10^{-122}\,M_{\rm PL}{}^2$). This question becomes an acute one when the notion of spontaneous symmetry breaking of the vacuum state is introduced into high-energy physics to describe unified field theories of apparently diverse interactions. In such a theory it is impossible to demand zero vacuum energy in both the symmetric and asymmetric phases and difficult to see why it should be zero in the present broken-symmetry phase.

A universe with $\Lambda\neq 0$ rapidly approaches the de Sitter solution.[1] In this series of papers a systematic study of quantum field theory in de Sitter spacetime is carried out. The importance of the causal structure of the space and the Hawking effect to an understanding of the cosmological-constant problem and inflation will be the main point of emphasis. Details of the matter dynamics are completely ignored—at least at first. Thus a *non*interacting scalar quantum matter field is considered. Since the matter Lagrangian is quadratic in Φ, the field theory is completely characterized by its two-point function. There is no phase transition and the physics is seemingly "trivial." However, this is not the case at all.

The vacuum state of the matter field must first be carefully specified by consideration of the realistic time-dependent problem posed by inflation. This "vacuum" is unstable; it decays via the emission of scalar particle pairs due to the Hawking effect in de Sitter space. The treatment will be that of canonical quantization of the Φ field in a fixed classical gravitational background. This requires defining appropriate "in" and "out" states and finding the Bogoliubov mixing transformation between them, in analogy to Hawking's original treatment of black-hole radiance.[2] We begin in Sec. II with a review of the relevant methods of quantum field theory in curved backgrounds. The correct state for inflation is seen to be completely analogous to the Unruh vacuum in the Schwarzschild case, when the field is massive.

In Sec. III the Bogoliubov transformation associated with the particle creation is exhibited and the decay rate calculated. Section IV contains a discussion of the various de Sitter-invariant states and a proof that the $|\,\text{in}\rangle$, $|\,\text{out}\rangle$, and Euclidean states are three different members of a one-parameter class of invariant "vacua." In Sec. V the energy momentum of the created particles is evaluated by computing the difference in the vacuum energy between the $|\,\text{in}\rangle$ and $|\,\text{out}\rangle$ states. This change in vacuum energy can also be regarded as a change in the effective cosmological constant. The sign is such as to imply a spontaneous *decrease in* $\Lambda_{\rm eff}$ as the particle creation proceeds. Again the situation is similar to black-hole decay or electric-field decay by charged particle creation "shorting out" the background field. The existence of de Sitter-invariant states with arbitrarily low energy density is demonstrated in Sec. V also. Thus the classic conditions for instability are realized.

In order to make the paper as self-contained and readable as possible, technical results extraneous to the main line of argument are relegated to the Appendix. The conventions are those of Misner, Thorne, and Wheeler (Ref. 21). The units are Planck units: $G=1/M_{\rm Pl}{}^2=\hbar=k=c=1$.

II. CANONICAL QUANTIZATION IN CURVED SPACE: GENERAL METHOD

Consider a noninteracting scalar field Φ in a classical gravitational background. The Lagrangian is

$$L_\Phi=-\tfrac{1}{2}g^{ab}(\partial_a\Phi)(\partial_b\Phi)-\frac{m^2}{2}\Phi^2-\frac{\xi R}{2}\Phi^2\;. \tag{1}$$

In order to quantize the field Φ we should first consider the solutions of the scalar wave equation

$$(-D_aD^a+m^2+\xi R)\Phi \equiv\left[-\frac{1}{\sqrt{-g}}\partial_a(g^{ab}\sqrt{-g}\;\partial_b)+m^2+\xi R\right]\Phi=0\;. \tag{2}$$

If it were possible to uniquely define positive- and negative-frequency solutions to this equation, then the field could be expanded in a Fock-space representation

$$\Phi=\sum_\lambda(a_\lambda\Phi_{\lambda(+)}+a_\lambda^\dagger\Phi_{\lambda(-)})\;. \tag{3}$$

The vacuum state would then be uniquely specified by $a_\lambda\,|\,\text{vac}\rangle=0,\in\lambda$ exactly as in flat space.

As is well known, the specification of positive- and negative-frequency modes is dependent on the choice of time slicing of the four-geometry. If the gravitational background field were "switched" on and then off, adiabatically, a preferred time slicing would be specified, asymptotic noninteracting "in" and "out" states would then exist, and the canonical quantization procedure again becomes well defined.

Actually, the adiabatic-switching method is rather awkward, and often it cannot be easily implemented, even in flat-space problems. Schwinger, in effect, proposed an elegant way around these difficulties by replacing the canonical adiabatic prescription with his proper-time formalism, together with suitable analyticity conditions on the proper-time kernel. By this method, he was able to calculate the decay rate of a constant, uniform electric field into charged particle-antiparticle pairs—without any adiabatic-switching arguments.[3]

In canonical language, Schwinger's method amounts to a definition of positive and negative frequency [by which the concepts of "particle" and "vacuum" are also given meaning through Eq. (3)], for $t \to +\infty$ and $t \to -\infty$ separately. If the decompositions of the solution space of the scalar wave equation into positive and negative subspaces at $t = \pm\infty$ are inequivalent, i.e., if the linear transformation relating the two sets of mode functions has off-diagonal elements, then particle creation occurs. This canonical reformulation of Schwinger's method has been successfully applied and extended to electromagnetic background-field problems by several authors.[4]

The same method works equally well for background gravitational field problems, as has been emphasized by Rumpf.[5] The importance of this observation is that the Schwinger method involves only analytic continuation in m^2. Therefore, it does not presume any analyticity of the metric, nor even the existence of asymptotically flat regions in the background spacetime. Once the in and out states are specified by careful consideration of which time-dependent problem we actually wish to solve, the Bogoliubov mixing transformation is uniquely determined in an observer-independent manner.

The cosmological-constant problem is essentially a question of determining the energy of the vacuum, which is not very meaningful in curved space until we specify what vacuum we should be discussing. Schwinger's method, in either its proper-time or canonical formulation, provides a precise definition of vacuum in curved backgrounds which is formally consistent, physically reasonable (for example, in possessing a semiclassical limit and predicting particle creation where it is expected), and which is equivalent to adiabatic methods, whenever the latter can be implemented.

The basic idea of the canonical method is to define positive- and negative-frequency solutions to Eq. (2), $\Phi_{\lambda\pm}$ according to whether the inner product

$$(f,g)_\Sigma = \int_\Sigma d\Sigma^a i f^* (\overrightarrow{\partial}_a - \overleftarrow{\partial}^a) g \tag{4}$$

is positive or negative. If Σ is a spacelike surface in the past of every point in the spacetime (i.e., a complete Cauchy surface for the classical time evolution), then the modes $\Phi_{\lambda(\pm)}$, assumed complete and orthonormal, and obeying

$$(\Phi_{\lambda(\pm)}, \Phi_{\lambda'(\pm)})_\Sigma = \pm\delta(\lambda,\lambda') \tag{5}$$

will be called incoming particle (antiparticle) modes if they can be analytically continued in m^2 to the upper (lower) half m^2 plane and are regular at past infinity. The corresponding canonical operators a_λ, $a_\lambda^\dagger$ in (3) satisfy

$$\begin{gathered}[a_\lambda, a_{\lambda'}^\dagger] = \delta(\lambda,\lambda') \ , \\ [a_\lambda, a_{\lambda'}] = [a_\lambda^\dagger, a_{\lambda'}^\dagger] = 0 \ ,\end{gathered} \tag{6}$$

and define the in state

$$a_\lambda \,|\, \text{in}\rangle = 0 \ . \tag{7}$$

This specification of the $|\,\text{in}\rangle$ state corresponds to a precise set of initial conditions for the time evolution of the field. In the same way, if Σ is a complete Cauchy surface for the reversed time evolution, i.e., a surface in the future of every point in the spacetime, then the outgoing modes $\Phi_\lambda^{(\pm)}$ and the corresponding operators $b_\lambda, b_\lambda^\dagger$ define the out state

$$b_\lambda \,|\, \text{out}\rangle = 0 \ . \tag{8}$$

An outgoing particle (antiparticle) mode can be analytically continued into the lower (upper) half m^2 plane and is regular at future infinity. Since the wave equation (2) is second order, there must exist a linear relation between the two sets of solutions,

$$\begin{bmatrix} \Phi_{\lambda(+)} \\ \Phi_{\lambda(-)} \end{bmatrix} = \begin{bmatrix} \alpha_\lambda & \beta_\lambda \\ \beta_\lambda^* & \alpha_\lambda^* \end{bmatrix} \begin{bmatrix} \Phi_\lambda^{(+)} \\ \Phi_\lambda^{(-)} \end{bmatrix} \tag{9}$$

with $|\alpha_\lambda|^2 - |\beta_\lambda|^2 = 1$. In (9) the assumption has been made that the matrix is diagonal in λ, as this is the case of interest. Also, we have used the fact that

$$\Phi_\lambda^{(+)} = [\Phi_\lambda^{(-)}]^*, \quad \Phi_{\lambda(+)} = [\Phi_{\lambda(-)}]^* \tag{10}$$

in a gravitational background.

The corresponding transformation for the operators a_λ and b_λ is

$$\begin{bmatrix} a_\lambda \\ a_\lambda^\dagger \end{bmatrix} = \begin{bmatrix} \alpha_\lambda^* & -\beta_\lambda^* \\ -\beta_\lambda & \alpha_\lambda \end{bmatrix} \begin{bmatrix} b_\lambda \\ b_\lambda^\dagger \end{bmatrix} . \tag{11}$$

The two basis sets are inequivalent, i.e., particle creation occurs, if and only if $\beta_\lambda \neq 0$. The canonical transformation between the two bases may be realized formally by a unitary transformation in Hilbert space

$$b_\lambda = \mathcal{U} a_\lambda \mathcal{U}^{-1} \ . \tag{12}$$

Actually, this transformation will turn out to be purely formal because the Fock spaces spanned by a_λ and b_λ are generally unitarily inequivalent, i.e., $\langle \text{out}\,|\,\text{in}\rangle \to 0$ in the infinite-volume limit. This causes no serious difficulty, provided we are careful to ask only physically sensible questions of the background-field formalism. Thus, the total number of created quanta in infinite volume and infinite time is infinite, but the creation probability per unit four volume is finite. This may be calculated from an expression of the form

$$|\langle \text{out} | \text{in} \rangle|^2 \rightarrow \exp(-\Gamma V_4) ,$$

where V_4 is the four-volume in which the particle creation takes place. Thus

$$\Gamma = 2\text{Im}\mathscr{L}_{\text{eff}} = -\lim_{V_4 \rightarrow \infty} \frac{1}{V_4} \ln |\langle \text{out} | \text{in} \rangle|^2 . \quad (13)$$

In addition to the decay rate we will also be interested in the two-point functions in various states. The antisymmetric two-point (commutator) function is unchanged by the canonical transformation (12). Information about the particle creation resides instead in the symmetric function

$$G^{(1)}_{\text{vac}}(x,x') = \langle \text{vac} | \Phi(x)\Phi(x') + \Phi(x')\Phi(x) | \text{vac} \rangle ,$$

where $|\text{vac}\rangle$ is any of the vacuum states to be considered. For simplicity, we will hereafter drop the superscript on the symmetric two-point function.

Let us apply these formal considerations to the familiar example of the Schwarzschild black hole,

$$ds^2 = -dt^2 \left[1 - \frac{2M}{r} \right] + \frac{dr^2}{1 - 2M/r} + r^2 d\Omega^2 . \quad (14)$$

The Penrose diagram for this geometry is shown in Fig. 1. In the static coordinates (14) the background appears time-independent and indeed it possesses a timelike Killing field $\partial/\partial t$. However, the horizon at $r = 2M$ gives rise to nontrivial mixing between positive- and negative-frequency modes. If a Fock decomposition is specified with respect to incoming spherical waves on $\mathscr{I}^-$ and $\mathscr{H}^-$, then the quantum state of the matter field is fixed. It is the $|\text{in}\rangle$ vacuum state considered by Unruh,[6] and corresponds to the time-dependent problem of spherically symmetric stellar collapse (cf. Fig. 2). When this *same* state is reexpressed in terms of the outgoing modes of the field on $\mathscr{H}^+$ and $\mathscr{I}^+$, we find a thermal distribution of created particles with temperature $T = 1/8\pi M$. The outward flux of radiation at late times implies a corresponding decrease in the mass M: the hole decays.

A very different state is obtained if one considers the time-independent Schwarzschild spacetime on the Euclidean section. Since the quantum states of the Φ field are in one to one correspondence to the Feynman's Green's function solutions of

$$(-D^a D_a + m^2 + \xi R) G_F(x,x') = \delta(x,x') \quad (15)$$

and since the Euclidean section $t \rightarrow it$ turns the hyperbolic differential operator in (15) into an elliptic operator, the continuation uniquely specifies a vacuum state—the Hawking-Hartle vacuum.[7] This state corresponds to an equilibrium situation where the rate of emission from the hole is exactly compensated by the rate of absorption of matter into the hole. In particular, it is completely time symmetric.

In the language of Green's functions the Unruh in state corresponds to adding a particular (symmetric) solution of the *homogeneous* version of Eq. (15) to the Euclidean Green's function, in order to account for the different initial conditions on $\mathscr{I}^-$ and $\mathscr{H}^-$. The resulting Unruh Green's function possesses all of the same symmetries as the Euclidean function (such as spherical symmetry) but it selects a preferred direction of time corresponding to the collapse problem and decay of the black hole. Completely analogous statements can be made for the de Sitter case to which we turn.

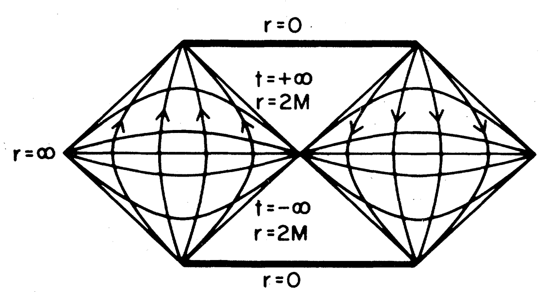

FIG. 1. The Penrose diagram for the globally extended Schwarzschild spacetime, with the angular coordinates θ,ϕ suppressed.

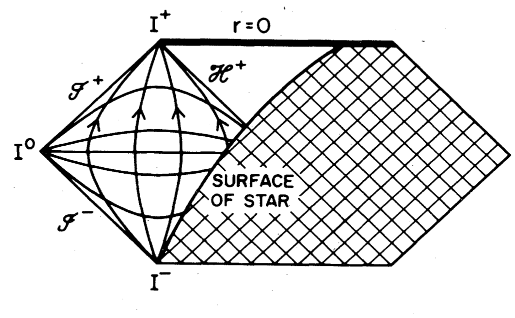

FIG. 2. The Penrose diagram for the exterior geometry of spherically symmetric gravitational collapse. $I^{\pm}$ denotes the timelike infinities, I^0 spacelike infinity, $\mathscr{I}^{\pm}$ null infinities, and $\mathscr{H}^+$ the future horizon at $r = 2M$. The shaded region must be replaced by the interior geometry appropriate to the detailed collapse problem. The Bogoliubov mode mixing at late time is completely determined by the exterior, unshaded section of the spacetime, however.

III. CANONICAL QUANTIZATION IN DE SITTER SPACETIME

The Penrose diagram for global de Sitter spacetime is shown in Fig. 3. A convenient coordinate system that covers the whole spacetime is the one in which the spatial sections are closed three-spheres ($K = 1$):

$$ds^2 = -dt^2 + H^{-2}\cosh^2 Ht (d\chi^2 + \sin^2\chi \, d\Omega^2) . \quad (16)$$

The transformations to other frequently considered coordinates are given in the Appendix for reference. The particle creation will turn to be independent of the value of

K, i.e., whether the universe is spatially closed, open, or flat.

In coordinates (16) the wave equation (2) becomes

$$\left[\frac{1}{\cosh^3 Ht}\frac{\partial}{\partial t}\left[\cosh^3 Ht\frac{\partial}{\partial t}\right]-\frac{H^2\Delta_3}{\cosh^2 Ht}+M^2\right]\Phi=0 . \tag{17}$$

Here the notation

$$M^2\equiv m^2+\xi R \tag{18}$$

and

$$\Delta_3\equiv\frac{1}{\sin^2\chi}\left[\frac{\partial}{\partial\chi}\left[\sin^2\chi\frac{\partial}{\partial\chi}\right]+\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left[\sin\theta\frac{\partial}{\partial\theta}\right]+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right] , \tag{19}$$

the Laplacian on the unit three-sphere has been introduced. Since $R=4\Lambda=12H^2$ defines the constant H (in Planck units), M^2 is simply a constant. It is clear that apart from overall scales, physics depends only on the dimensionless ratio M/H, which is the ratio of the horizon scale of the background to the Compton wavelength of the matter field. For future use, we define the quantity

$$\gamma\equiv\left[\frac{M^2}{H^2}-\frac{9}{4}\right]^{1/2} . \tag{20}$$

The solutions to (17) are easily written down in terms of standard functions of analysis

$$\Phi(t,\chi,\theta,\phi)\sim y_k(t)Y_{klm}(\chi,\theta,\phi) . \tag{21}$$

The Y_{klm} are S_3 spherical harmonics obeying

$$-\Delta_3 Y_{klm}=k(k+2)Y_{klm} \tag{22}$$

which are $(k+1)^2$-fold degenerate $(m=-l,\ldots,+l;l=0,1,\ldots,k)$. The orthogonality and completeness relations as well as useful addition theorem for the Y_{klm} may be found in the Appendix. The $y_k(t)$ then obey

$$\frac{1}{\cosh^3 Ht}\frac{d}{dt}\left[\cosh^3 Ht\frac{dy_k}{dt}\right]+\frac{H^2k(k+2)}{\cosh^2 Ht}y_k+M^2y_k=0 \tag{23}$$

which is a form of Legendre's equation. For purposes of analyzing the asymptotics of the solutions as $t\to\pm\infty$, however, it is more convenient to write the solutions in the form[8]

$$y_k^{(\pm)}(t)\sim\cosh^k(Ht)\exp[(-k-\tfrac{3}{2}\mp i\gamma)Ht]F(k+\tfrac{3}{2},k+\tfrac{3}{2}\pm i\gamma;\ 1\pm i\gamma;-e^{-2Ht}) \tag{24}$$

or

$$y_{k(\pm)}(t)\sim\cosh^k(Ht)\exp[(k+\tfrac{3}{2}\mp i\gamma)Ht]F(k+\tfrac{3}{2},k+\tfrac{3}{2}\mp i\gamma;\ 1\mp i\gamma;-e^{+2Ht}) , \tag{25}$$

where F is the hypergeometric function ${}_2F_1$. The first set of solutions (24) have simple asymptotic properties as $t\to+\infty$:

$$y_k^{(\pm)}(t)\to\exp(-\tfrac{3}{2}Ht\mp i\gamma Ht),\quad t\to+\infty . \tag{26}$$

The second set (25) have simple asymptotics at $t\to-\infty$:

$$y_{k(\pm)}(t)\to\exp(+\tfrac{3}{2}Ht\mp i\gamma Ht),\quad t\to-\infty . \tag{27}$$

Any of the four solutions may be obtained from any of the others via the relations

$$y_k^{(-)}(t)=[y_k^{(+)}(t)]^* ,\qquad y_{k(-)}(t)=[y_{k(+)}(t)]^* , \tag{28}$$

and

$$y_{k(\pm)}(t)=y_k^{(\mp)}(-t) . \tag{29}$$

Because of (26) and (27), $y_k^{(\pm)}$ and $y_{k(\pm)}$ may be used as basis functions for the canonical quantization of the Φ field in the sense of Sec. II. In other words, $y_{k(\pm)}$ define positive-and negative-frequency solutions with which a Fock basis may be constructed at $t=-\infty$:

$$\Phi=\sum_{klm}(a_{klm}\Phi_{(+)klm}+a_{klm}^\dagger\Phi_{(-)klm}) . \tag{30}$$

The $y_k^{(\pm)}$ functions define a similar decomposition at $t=+\infty$:

$$\Phi=\sum_{klm}(b_{klm}\Phi_{klm}^{(+)}+b_{klm}^\dagger\Phi_{klm}^{(-)}) . \tag{31}$$

The $|\text{in}\rangle$ and $|\text{out}\rangle$ states are defined by Eqs. (7) and (8). The Bogoliubov mixing coefficients of Eq. (9) are found from Eqs. (24) and (25) and the inversion transformation for the hypergeometric function.[9] The result is

$$\alpha_k=\frac{\Gamma(1-i\gamma)\Gamma(-i\gamma)}{\Gamma(k+\frac{3}{2}-i\gamma)\Gamma(-k-\frac{1}{2}-i\gamma)} ,\qquad \beta_k=\frac{\Gamma(1-i\gamma)\Gamma(i\gamma)}{\Gamma(k+\frac{3}{2})\Gamma(-\frac{1}{2}-k)}=\frac{i(-)^k}{\sinh\pi\gamma} . \tag{32}$$

These coefficients satisfy $|\alpha_k|^2-|\beta_k|^2=1$ for all k, showing that the transformation is canonical. The magnitudes $|\alpha_k|$ and $|\beta_k|$ are independent of k as well. Thus we may set

$$\alpha_k=e^{-2i\delta_k}\cosh2\bar{\theta},\quad \beta_k=i(-)^k\sinh2\bar{\theta} ,$$

where $\bar{\theta}$ is a fixed constant given by

$$\sinh2\bar{\theta}=\text{csch}\,\pi\gamma . \tag{33}$$

The relative amplitude for producing a pair of particles in the final states (klm) and $(kl\,-m)$ if none were present in the initial state is

$$\frac{\langle \text{out} \,|\, b_{klm} b_{kl-m} \,|\, \text{in} \rangle}{\langle \text{out} \,|\, \text{in} \rangle} .$$

By making use of (12) and (33), the square of this amplitude becomes

$$w_{klm} = |\beta_k/\alpha_k|^2 = \tanh^2 2\bar{\theta} = \operatorname{sech}^2 \pi\gamma \tag{34}$$

which gives the relative probability of creating a pair in the given mode. Absolute probabilities are obtained by requiring the total probability of creating 0,1,2,... pairs to be unity:

$$N_{klm}(1 + w_{klm} + w_{klm}{}^2 + \cdots) = 1$$

or

$$N_{klm} = 1 - w_{klm} . \tag{35}$$

This is also the absolute probability of creating no particles in the given mode. The absolute probability that no particles are created in any mode, i.e., that the vacuum remains the vacuum is

$$|\langle \text{out} \,|\, \text{in} \rangle|^2 = \prod_{klm} N_{klm} = \exp\left[\sum_{klm} \ln(\tanh^2 \pi\gamma)\right] \tag{36}$$

by (34) and (35).

Since w_{klm} is independent of k,l,m, the sum in (36) is quite divergent. This is the divergence referred to in Eq. (13). We cut off the sum at $k = N$ and then let the cutoff change. Then

$$\Delta \sum_{k=0}^{N} \sum_{l=0}^{k} \sum_{m=-l}^{l} 1 = (N+1)^2 \Delta N \to e^{3\ln N} \frac{\Delta N}{N} \quad \text{as } N \to \infty . \tag{37}$$

This can be related to the change in the four volume by the following consideration. The physical momentum of a state with quantum number N is

$$k_{\text{phys}} \to \frac{N}{\cosh Ht} \quad \text{as } N \to \infty .$$

This is clear from the wave equation (23), for example. For any realistic probe of when decay of the vacuum has occurred, k_{phys} will be bounded from above. For fixed k_{phys}, ΔN is then related to the total elapsed time since the beginning of the decay process when the initial state was prepared (i.e., the beginning of the inflationary epoch) and we have

$$\frac{\Delta N}{N} = \frac{\Delta(\cosh Ht)}{\cosh Ht} \to H \Delta t$$

or

$$\ln N \to Ht \tag{38}$$

as N and t both become large. This is independent of the value of k_{phys}, which may be taken to infinity without affecting the result.

The slice of four-volume in the time interval Δt is

$$\Delta V_4 = \frac{2\pi^2}{H^3} \cosh^3 Ht \, \Delta t \to \frac{\pi^2}{4H^4} e^{3Ht} H \Delta t \to \frac{\pi^2}{4H^4} e^{3\ln N} \frac{\Delta N}{N} \tag{39}$$

in the same limit. Thus, Eqs. (37)—(39) yield

$$\Delta \sum_{klm} \ln(\tanh^2 \pi\gamma) \to -\Gamma \Delta V_4$$

with

$$\Gamma = \frac{8H^4}{\pi^2} \ln(\coth \pi\gamma) . \tag{40}$$

This result, though implicit in earlier work [the transformation, Eqs. (32) was written down by Gutzwiller[8]] has not been explicitly understood in terms of the problem of vacuum decay. By relating the de Sitter problem to a more standard canonical quantization problem, along lines precisely paralleling background-field methods in electromagnetism, the interpretation of the Bogoliubov mixing coefficients is made apparent. Some remarks as to the limits of the validity of Eq. (40) and its significance for inflation are in order.

First, let us state the obvious: the calculation leading to (40) is based on field quantization in a classical background. Thus, quantum gravitational effects must be negligible or the calculation is meaningless. Quantum gravitational effects will be of order $GH^2 = H^2$, in Planck units. Thus, a necessary condition for the validity of (40) is $H \ll 1$.

Second, the background metric has been taken to be fixed, i.e., and effects of the particle creation on the metric have been ignored. This amounts to neglecting the energy-momentum of the particles as a source term in Einstein's equations. In the final section, we will calculate T_{ab} and discuss the effects of the particle production on the geometry, to first order in $\hbar$.

Finally, the limitations of the canonical method must be addressed. The $|\text{in}\rangle$ vacuum represents a well-defined specification of the quantum state of the Φ field over a complete Cauchy surface. For a noninteracting field theory with Lagrangian (1) in the metric (16), no approximation whatever is involved in this specification. However, in any realistic model of inflation the universe is not globally de Sitter. *If* the particle creation rate is slow compared to the expansion of the universe, then the matter density will be attenuated exponentially rapidly and the $|\text{in}\rangle$ vacuum is indeed the state of the field that is left behind. In the Appendix it is shown that the initial data on the $t = \text{constant} \to -\infty$ hypersurface, which specifies the $|\text{in}\rangle$ vacuum in coordinates (16), is equivalent to zero particle content in the open and flat $(k = -1, 0)$ coordinate frames as well. Thus, the $|\text{in}\rangle$ vacuum really is analogous to the Unruh vacuum state in the Schwarzschild case. It is precisely the state specified by initial data on a time slice which is the beginning of the inflationary epoch, independently of what went on before.

If we confine our interest in the future evolution to a three-volume which began the inflation with a scale of order H^{-3}, then the difference between the globally defined $|\text{in}\rangle$ state and the actual state of the system must be exponentially small over this region—provided that the particle creation rate is small.

Inspection of Eqs. (40) and (20) shows that the particle creation rate is *not* small when M becomes of the order of H. Physically, this means that although the formal specification of the $|\text{in}\rangle$ vacuum is completely unambiguous, the *actual* state of the system with light fields will be very far from the $|\text{in}\rangle$ state. In the conformal ($M^2=2H^2$) or massless ($M=0$) limits, the particle creation rates are high enough that in any realistic model of inflation with interactions, the created particles would have time to interact and thermalize before red-shifting. Then a much more reasonable choice of vacuum state is the thermal state, specified by continuation from the Euclidean section. Since we do not analyze this state in this paper, we should restrict ourselves to the massive limit, $M \gg H$. In that limit

$$\Gamma \to \frac{16H^2}{\pi^2} e^{-2\pi\gamma} \to \frac{16H^4}{\pi^2} e^{-M/T_H} , \tag{41}$$

where $T_H = H/2\pi$ is the Hawking-de Sitter temperature. Thus, the divergence of (40) at $\gamma=0$ is unphysical and means only that the $|\text{in}\rangle$ state cannot be the vacuum or ground state of the field, even approximately. In the literature this is sometimes referred to as the breakdown of the adiabatic method.[10] The additional condition $H \ll M$ is the one significant difference between the de Sitter and Schwarzschild backgrounds. The asymptotic flatness of the latter guarantees a finite, small Γ for any value of the test field mass.

IV. INVARIANT STATES AND GREEN'S FUNCTIONS

In flat space, Lorentz invariance plus regularity at spacelike infinity fixes a unique vacuum state and corresponding two-point function. In de Sitter space there is a one-parameter family de Sitter-invariant "vacuum" states (Ref. 16). The one that is usually discussed in the Euclidean vacuum, obtained by considering the de Sitter metric on the Euclidean section,

$$t \to \frac{-i}{H}\left[\frac{\pi}{2} - \psi\right] . \tag{42}$$

The geometry is then S_4 and the $G(x,x')$ solving Eq. (15) may be constructed from the harmonic functions regular on S_4 (Ref. 11). In order to show how this state is related to the $|\text{in}\rangle$ and $|\text{out}\rangle$ states considered in our previous discussion of particle creation and the other invariant vacua, let us introduce yet another representation of the scalar field operator:

$$\Phi(x) = \sum_{klm} y_k(t) Y_{klm}(\Omega_3) c_{klm} + \text{H.c.} \tag{43}$$

The $y_k(t)$ are now taken to be those solutions of Eq. (23), normalized by Eq. (5) which are regular on the Euclidean section. In terms of the variable $\zeta \equiv i \sinh Ht$ we may express $y_k(t)$ in the form of a Legendre function[12]

$$\begin{aligned} y_k(t) &= A_k(\zeta^2-1)^{-1/2} P_\nu^{k+1}(\zeta)[i^k\theta(t) + (-i)^k\theta(-t)] , \\ A_k &= \frac{H}{\sqrt{2}} |\Gamma(-k-\tfrac{1}{2}-i\gamma)| , \\ \nu &= -\tfrac{1}{2} + i\gamma . \end{aligned} \tag{44}$$

which is regular under (42), and as t passes through zero. If we denote by $\bar{x}$ the antipodal point to x, with coordinates $(-t, \pi-\chi, \pi-\theta, \phi+\pi)$, then y_k satisfies

$$y_k(\bar{t}) = y_k(-t) = (-)^k y_k^*(t) . \tag{45}$$

The state which is annihilated by c_{klm}, denoted by $|0\rangle$,

$$c_{klm}|0\rangle = 0 , \tag{46}$$

is the Euclidean vacuum. One way to show this is to evaluate the two-point function in this state explicitly:

$$G_0(x,x') = \langle 0 | \{\Phi(x), \Phi(x')\} | 0 \rangle \tag{47}$$

$$= \sum_{klm} y_k(t) Y_{klm}(\Omega_3) y_k^*(t') Y_{klm}^*(\Omega_3) + (x \leftrightarrow x') . \tag{48}$$

A more elegant method is to use the fact that (48) is de Sitter invariant. This follows immediately upon recognizing that the $\{y_k Y_{klm}\}$ are basis vectors for an infinite-dimensional representation of the de Sitter group SO(4,1), labeled by ν.[13] Then the sum in (48) is simply a group character in this representation. The de Sitter invariance implies that $G_0(x,x')$ can be a function only of the geodesic distance between x and x' or

$$z(x,x') = -\sinh Ht \sinh Ht' + \cosh Ht \cosh Ht' \cos\Omega , \tag{49}$$

where Ω is the angle between the spacelike components of x and x' on S_3. This is the only nontrivial invariant scalar function of two arguments. Since $G_0 = G_0(z)$, the partial differential operator in Eq. (15), viz., $D_a D^a$ may be replaced by an ordinary differential operator in z:[14]

$$-D^a D_a G_0(z(x,x')) = H^2 \left[(z^2-1)\frac{d^2}{dz^2} + 4z\frac{d}{dz} \right] G_0(z) . \tag{50}$$

Upon substituting $G_0 = (d/dz)w$, Eq. (15) then becomes Legendre's equation with $\nu(\nu+1) = 2 - M^2/H^2 = \frac{1}{4} - \gamma^2$. The general solution is of the form

$$A\frac{d}{dz}P_\nu(z) + B\frac{d}{dz}P_\nu(-z) . \tag{51}$$

Now, $P_\nu(-z)$ is singular at $z=1$ where Eq. (15) has a δ function source. On the Euclidean section $z=-1$ is possible only for $x=\bar{x}'$ where Eq. (15) is quite regular, but $P_\nu(z)$ is singular. Thus, the coefficient A in (51) must vanish in the Euclidean state. Then B may be fixed by comparison to the flat-space limit $H \to 0$, near $(x-x')^2=0$. The light-cone singularity structure must be $1/2\pi^2(x-x')^2$. Since

$$\frac{d}{dz}P_\nu(-z)\underset{z\to 1}{\sim}\frac{1}{\pi}\cosh\pi\gamma(1-z)^{-1}$$
$$\underset{H\to 0}{\sim}\frac{2}{H^2}\frac{\cosh\pi\gamma}{(x-x')^2}$$

we obtain $B=H^2/4\pi\cosh\pi\gamma$ and[15]

$$G_0(x,x')=\frac{H^2}{4\pi}\text{sech }\pi\gamma\frac{d}{dz}P_\nu(-z),\quad z<1\ . \tag{52}$$

When x and x' are timelike, $z>1$ and G_0 is obtained from (52) by taking the real part of the function analytically continued around the branch cut from $z=+1$ to ∞.

If $A\neq 0$, the two-point function and corresponding vacuum state is still de Sitter invariant. However, (51) shows that $A\neq 0$ corresponds to placing an "image source" at $x=\bar{x}'$ in addition to the $\delta(x,x')$ in (15). The direct source at x' results in causal propagation in the interior of the forward light cone of x'. Since $\bar{t}'=-t'$, the image source at $\bar{x}'$ results in propagation in the interior of the *backward* light cone of $\bar{x}'$. It is for this reason that the authors of Ref. 16 effectively set $A=0$ in order to eliminate the acausal behavior associated with the image source. However, in the *half* space covered by the coordinates of (A9), for example, there is no physically unacceptable acausality: cf. Fig. 5. It is true that only for $A=0$ are the real and imaginary parts of the Feynman propagator related by the ordinary $i\epsilon$ prescription. Changing A changes the symmetric part of G_F without changing the antisymmetric, commutator part. Again this causes no problems when restricted to the half space of Fig. 5.

Just as in electrostatics where the method of images provides the solution to Poisson's equation in half spaces with certain conditions on the boundary, the introduction of an image source in Eq. (15) provides the solution to a certain real-time problem specified by conditions on the past horizon. In other words, if $A\neq 0$, (51) is an acceptable two-point function in only half of the full de Sitter spacetime. The two-point function in the other half depends on additional information about the preinflationary epoch, just as the two-point function in the collapse problem (Fig. 2) depends on the interior solution.

The de Sitter-invariant states labeled by a real parameter θ may be characterized by means of a unitary operator acting on the Euclidean vacuum $|0\rangle$.[10,17] For *arbitrary* values of θ and δ_k define

$$\mathcal{U}(\theta;\delta_k)=\exp(iS\theta)\exp\left[i\sum_{klm}\delta_k c^\dagger_{klm}c_{klm}\right],$$
$$S=\tfrac{1}{2}\sum_{klm}(-)^k\,(c_{klm}c_{kl-m}+c^\dagger_{klm}c^\dagger_{kl-m})\ . \tag{53}$$

Under $\mathcal{U}(\theta;\delta_k)$, c_{klm} becomes

$$\mathcal{U}(\theta;\delta_k)c_{klm}\,\mathcal{U}^{-1}(\theta;\delta_k)$$
$$=e^{-i\delta_k}[\cosh\theta c_{klm}+i(-)^k\sinh\theta c^\dagger_{kl-m}]\ . \tag{54}$$

Therefore, using Eqs. (9), (11), and (45),

$$\mathcal{U}(\theta;\delta_k)\Phi(x)\,\mathcal{U}^{-1}(\theta;\delta_k)$$

has the same form as (43) with $y_k(t)$ replaced by

$$y_k(t)\to e^{i\delta_k}[\cosh\theta y_k(t)-i(-)^k\sinh\theta y_k^*(t)]\ . \tag{55}$$

Forming the state $\mathcal{U}(\theta;\delta_k)|0\rangle$ and evaluating the two-point function in that state gives

$$\sum_{klm}\{\cosh^2\theta y_k(t)y_k^*(t')+\sinh^2\theta y_k(t')y_k^*(t)+i(-)^k\sinh\theta\cosh\theta[y_k(t)y_k(t')-y_k^*(t)y_k^*(t')]\}\,Y_{klm}(\Omega_3)Y^*_{klm}(\Omega_3')+(x\leftrightarrow x')\ . \tag{56}$$

Notice that δ_k has dropped out of this expression so that it cannot affect the de Sitter invariance of the answer. The sums in (56) are evaluated in the Appendix. The result is

$$G_\theta(x,x')=\cosh2\theta G_0(x,x')-\sinh2\theta\cosh\pi\gamma G_0(x,\bar{x}')\ . \tag{57}$$

Since $z(x,\bar{x}')=-z(x,x')$, Eq. (57) is precisely of the form (51). We recognize θ as the one real parameter that distinguishes one de Sitter-invariant state from another. From the explicit formulas for the mode functions (24), (25), and (44), and the transformation law between them (55), it is possible to verify that the outgoing and incoming modes correspond to specific values of θ and δ_k (labeled by the subscripts out and in) obeying

$$2\bar{\delta}_k=(\delta_k)_{\rm out}-(\delta_k)_{\rm in}=2(\delta_k)_{\rm out}\ ,$$
$$2\bar{\theta}=\theta_{\rm out}-\theta_{\rm in}=2\theta_{\rm out}\ , \tag{58}$$

where $\bar{\theta}$ and $\bar{\delta}_k$ are the *constant* parameters defined by Eqs. (32) and (33). Then

$$|\text{in}\rangle=\mathcal{U}(-\bar{\theta};-\bar{\delta}_k)|0\rangle\equiv\mathcal{U}_{\rm in}|0\rangle\ ,$$
$$|\text{out}\rangle=\mathcal{U}(\bar{\theta};\bar{\delta}_k)|0\rangle\equiv\mathcal{U}_{\rm out}|0\rangle\ , \tag{59}$$

and the unitary operator which connects $|\text{in}\rangle$ to $|\text{out}\rangle$ states in Eq. (12) is

$$\mathcal{U}=\mathcal{U}_{\rm out}\mathcal{U}_{\rm in}{}^{-1}=\mathcal{U}_{\rm out}{}^2\ . \tag{60}$$

These relations prove *a posteriori* that the $|\text{in}\rangle$ and $|\text{out}\rangle$ states are de Sitter invariant, being two particular members of the one-parameter class of invariant states corresponding to equal and opposite values of the parameter θ. The Euclidean state is "halfway" between at $\theta=0$. The special property of the $|\text{in}\rangle$ and $|\text{out}\rangle$ states is that they correspond to adding an image source to Eq. (15) at the antipodal point of x' of the *same* magnitude as the direct source at x'. This is verified by substituting into the general form (57) the value θ given by (33) so that

$$G_{\rm in}^{\rm out}(x,x')=\frac{H^2}{4\pi\sinh\pi\gamma}\frac{d}{dz}[P_\nu(-z)\mp P_\nu(z)]\ . \tag{61}$$

V. ENERGY-MOMENTUM AND DECREASE IN $\Lambda_{\rm eff}$

The energy-momentum tensor of the scalar field involves bilinear operators in Φ. Thus, it is generally ill defined in the quantum theory, as is reflected in the quadratic divergences in $G(x,x')$ as $x\to x'$. However, in order to evaluate the effects of the particle-creation process described in Sec. III, we need only calculate the *difference* in the energy between the $|\,{\rm in}\rangle$ and $|\,{\rm out}\rangle$ vacuum states. Since the theory is "free," the $|\,{\rm in}\rangle$ state is the state of the field for all t (up to a phase). At late t the vacuum is specified by $|\,{\rm out}\rangle$, however; thus, the difference

$$\overline{T}_{ab}=\langle{\rm in}\,|\,T_{ab}\,|\,{\rm in}\rangle-\langle{\rm out}\,|\,T_{ab}\,|\,{\rm out}\rangle \tag{62}$$

must be the T_{ab} of the created particles.[18] The divergences in either term separately cancel when the difference is calculated. Indeed, we can express T_{ab} in terms of the symmetric two-point functions calculated in the previous section:

$$\overline{T}_{ab}=\tfrac{1}{2}\lim_{x\to x'}\left[(1-2\xi)\frac{\partial}{\partial x^a}\frac{\partial}{\partial x'^b}+\left[2\xi-\frac{1}{2}\right]g_{ab}g^{cd}\frac{\partial}{\partial x^c}\frac{\partial}{\partial x'^d}-2\xi\frac{\partial}{\partial x'^a}\frac{\partial}{\partial x'^b}+2\xi g_{ab}g^{cd}\frac{\partial}{\partial x'^c}\frac{\partial}{\partial x'^d}\right.$$
$$\left.+\xi(R_{ab}-\tfrac{1}{2}g_{ab}R)-\tfrac{1}{2}m^2g_{ab}\right][G_{\rm in}(x,x')-G_{\rm out}(x,x;)]\ . \tag{63}$$

Owing to the de Sitter invariance we may replace

$$\frac{\partial}{\partial x'^a}\frac{\partial}{\partial x'^b}\to -H^2g_{ab}\frac{d}{dz}\ ,$$
$$\frac{\partial}{\partial x^a}\frac{\partial}{\partial x'^b}\to +H^2g_{ab}\frac{d}{dz}\ , \tag{64}$$
$$R_{ab}-\tfrac{1}{2}g_{ab}R=-3H^2g_{ab}\ ,$$

to obtain

$$\overline{T}_{ab}=-\frac{H^2}{2}g_{ab}\lim_{z\to 1}\left[\frac{d}{dz}+3\xi+\frac{m^2}{2H^2}\right]\times[G_{\rm in}(z)-G_{\rm out}(z)]\ . \tag{65}$$

Substituting Eq. (61) gives the finite result

$$\overline{T}_{ab}=-\frac{H^4}{4\pi\sinh\pi\gamma}g_{ab}\left[\frac{d}{dz}+3\xi+\frac{m^2}{2H^2}\right]\frac{d}{dz}P_\nu(z)\Big|_{z=1}\ . \tag{66}$$

The derivatives are easily evaluated from

$$P_\nu(z)=F\left[-\nu,1+\nu;1;\frac{1-z}{2}\right] \tag{67}$$

and the expansion of the hypergeometric function near zero. The result is

$$\overline{T}_{ab}=\frac{m^2}{32\pi\sinh\pi\gamma}[m^2+12H^2(\xi-\tfrac{1}{6})]g_{ab}\ . \tag{68}$$

This vanishes in the conformal limit $M^2=2H^2$ as it must, because in that case $\langle T_{ab}\rangle$ is given entirely by the trace anomaly and is therefore independent of vacuum choice. Hence the difference (62) must be zero. Regarding $\overline{T}_{ab}$ as the energy-momentum of a perfect fluid with equation of state $\bar\rho+\bar p=0$ we find that

$$\bar\rho=-\frac{m^2}{32\pi\sinh\pi\gamma}[m^2+12H^2(\xi-\tfrac{1}{6})]<0\ ,$$
$$M^2>\tfrac{9}{4}H^2>2H^2\ . \tag{69}$$

Thus, the scalar particles actually have negative energy density with respect to the vacuum at late times. We can now treat this small (order $\hbar$) energy density as a source term in Einstein's equations to find out what the effect of the particle-creation process will be on the geometry, to first order in $\hbar$. Since $\overline{T}_{ab}=-\bar\rho g_{ab}$ with $\bar\rho<0$ we find

$$\Lambda_{\rm out}=\Lambda_{\rm in}+8\pi\bar\rho<\Lambda_{\rm in}\ . \tag{70}$$

The effective cosmological "constant," as measured by the horizon scale is *decreased* by the particle-creation effect, for $M^2>\frac{9}{4}H^2$.

Now, the actual calculation of this effect has involved a fixed background field and asymptotic states at $t=\pm\infty$. However, the bulk of the mode-mixing takes place on a time scale of the order of H^{-1}, as may be seen by constructing time-dependent mixing coefficients for positive frequency modes defined at an arbitrary time t. Thus, in reality, the geometry will begin to feel the effect of the particle creation after a few expansion times. This suggests that in a self-consistent treatment of the problem the horizon scale would begin to increase and continue to do so without bound, until the particle creation ceases at $H=0$. The coherent vacuum energy of $\Lambda\neq 0$ would then be dissipated into an arbitrarily diffuse cloud of scalar quanta, much as the coherent energy of an electric field is dissipated into e^+e^- pairs or that of a classical black hole is evaporated into Hawking radiation. The time scale for this process may be crudely estimated by taking

$$\frac{dH^2}{dt}\simeq\frac{8\pi}{3}\frac{\bar\rho}{H^{-1}}\ . \tag{71}$$

In the limit $H\ll m$ (the approximation only improves in time if it is initially satisfied) this gives

$$t \sim \frac{12H}{m^4} e^{\pi m/H} \tag{72}$$

for the characteristic time scale over which H changes appreciably. This would still allow for enormous inflation, if m/H is large. A consistent calculation of this evolution must go beyond the background-field formalism and explicitly violate de Sitter invariance by introducing a preferred time slicing.

We can use our de Sitter invariant two-point functions to make one final point. Consider the result that would have been obtained for T_{ab} if we had compared the energies not of the $|\text{in}\rangle$ and $|\text{out}\rangle$ states in (62) but of general θ and $-\theta$. Using Eq. (57), the result (69) would be identical with the replacement of $\text{csch}\pi\gamma$ by its general value $\sinh 2\theta$. Thus, by choosing θ larger and larger we would find de Sitter invariant "vacua" with *arbitrarily negative* energy densities. Of course, we should not trust this conclusion when $|\rho|$ becomes comparable to the background Λ itself, without treating the back reaction self-consistently. But it does show that the de Sitter background is *unstable* to matter perturbations which change θ, for any value of $\Lambda > 0$. In particular, the Euclidean vacuum $\theta = 0$ is rendered unstable to matter fluctuations, although it is an equilibrium state with no decay rate in the sense of (13). The (in)stability analysis of the Euclidean vacuum will be the subject of the next paper in the series.

ACKNOWLEDGMENTS

I thank N. Mhyrvold for pointing out to me much of the literature cited in the references and B. Allen for useful discussions. This material is based upon research supported in part by the National Science Foundation under Grant No. PHY77-27084, supplemented by funds from the National Aeronautics and Space Administration.

APPENDIX

In this appendix the mathematical properties of de Sitter space and functions defined on it will be reviewed, and some new technical results presented.

Einstein's equations with a cosmological constant possess the maximally symmetric solution

$$R_{ab} = \Lambda g_{ab}, \quad R = 4\Lambda \equiv 12H^2 . \tag{A1}$$

This space is a four-manifold which may be embedded in a flat five-dimensional space

$$ds^2 = -dx_0{}^2 + dx_1{}^2 + dx_2{}^2 + dx_3{}^2 + dx_4{}^2 \tag{A2}$$

with the constraint

$$-x_0{}^2 + x_1{}^2 + x_2{}^2 + x_3{}^2 + x_4{}^2 = 1/H^2 . \tag{A3}$$

An explicit coordinatization satisfying (A3) and covering the entire space is given by

$$\begin{aligned} &x_0 = H^{-1}\sinh Ht_1, \quad x_1 = H^{-1}\cosh Ht_1 \cos\chi , \\ &x_2 = H^{-1}\cosh Ht_1 \sin\chi \cos\theta , \\ &x_3 = H^{-1}\cosh Ht_1 \sin\chi \sin\theta \cos\phi , \\ &x_4 = H^{-1}\cosh Ht_1 \sin\chi \sin\theta \sin\phi , \\ &t_1 \in [-\infty, \infty]; \quad \chi, \theta \in [0,\pi]; \quad \phi \in [0, 2\pi] . \end{aligned} \tag{A4}$$

When (A4) is substituted into (A2), Eq. (16) of the text follows with $t_1 = t$. The spatial sections are three-spheres of radius $H^{-1}\cosh Ht_1$ ($k=1$) and the space is an hyperboloid of revolution with isometry group O(4,1). Given two points on the manifold x and x', the quadratic form

$$z(x,x') = -x_0 x'_0 + x_1 x'_1 + x_2 x'_2 + x_3 x'_3 + x_4 x'_4 \tag{A5}$$

is invariant under this isometry group. Moreover, this is the only invariant function of two points, other than (A3) itself for x and x' separately. If (A4) is substituted into (A5), Eq. (49) is the result.

If instead of (A4) we set

$$\begin{aligned} &x_0 = H^{-1}\sinh Ht_{-1}\cosh\lambda, \quad x_1 = H^{-1}\cosh Ht_{-1} , \\ &x_2 = H^{-1}\sinh Ht_{-1}\sinh\lambda \cos\theta , \\ &x_3 = H^{-1}\sinh Ht_{-1}\sinh\lambda \sin\theta\cos\phi , \\ &x_4 = H^{-1}\sinh Ht_{-1}\sinh\lambda \sin\theta \sin\phi , \\ &t_{-1} \in [-\infty, \infty]; \quad \lambda \in [0, \infty] , \end{aligned} \tag{A6}$$

then the line element (A2) becomes

$$ds^2 = -dt_{-1}{}^2 + H^{-2}\sinh^2 Ht_{-1}(d\lambda^2 + \sinh^2\lambda\, d\Omega^2) . \tag{A7}$$

The spatial sections are now open ($k=-1$) hyperboloids. Comparing (A4) and (A6) these open coordinates can only cover the part of the space for which $\cos\chi \geq \text{sech} Ht_1$. This is illustrated in the Penrose diagram of Fig. 4.

Flat coordinates ($k=0$) may also be introduced by[19]

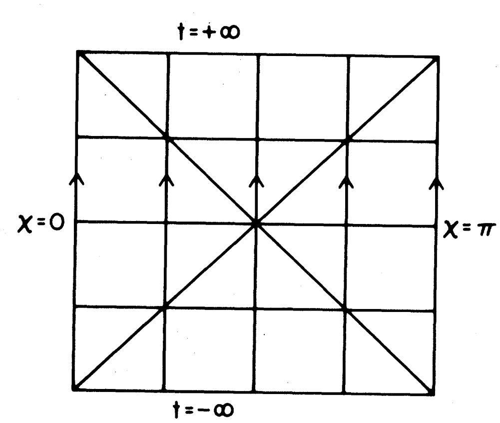

FIG. 3. The Penrose diagram for global de Sitter spacetime in coordinates (16). The vertical lines χ=const. are geodesic curves in the de Sitter metric.

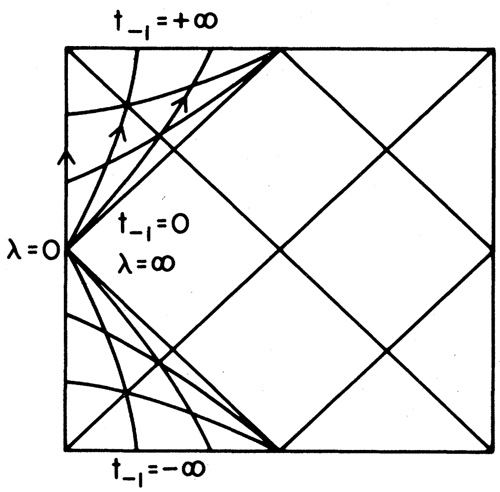

FIG. 4. The subspace of the full de Sitter spacetime covered by the open coordinates (A6) and (A7). The coordinates are singular as $t_{-1}\to 0$, which corresponds to the two different rays emanating from the central point at $\lambda=0$.

$$\cosh Ht_1 \sin\chi = \rho e^{H\hat{t}} ,$$
$$\sinh Ht_1 + \cosh Ht_1 \cos\chi = e^{H\hat{t}} , \quad \text{(A8)}$$
$$\hat{t}\in[-\infty,\infty],\rho\in[0,\infty] .$$

Then

$$ds^2 = -d\hat{t}^2 + H^{-2}e^{2H\hat{t}}(d\rho^2+\rho^2 d\Omega^2) \quad \text{(A9)}$$

which covers only the half space for which $\cos\chi \geq -\tanh Ht_1$. This is illustrated in Fig. 5.

Static coordinates are introduced by means of

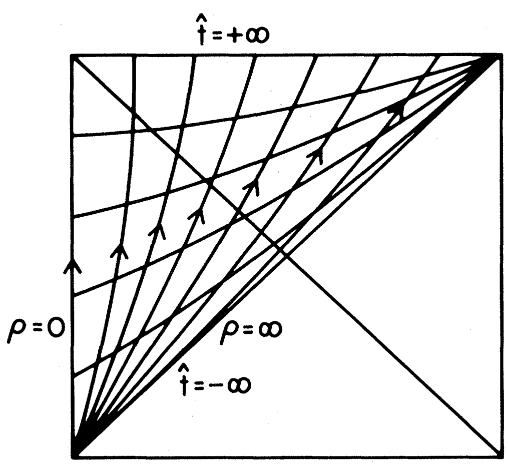

FIG. 5. The subset of the full de Sitter spacetime covered by the flat coordinates of (A8) and (A9). The curve $\hat{t}=-\infty$ divides the spacetime in half so that the antipodal point $\bar{x}$ lies in the lower right half if x lies in the upper left half. The lower half can be omitted, just as the shaded region of Fig. 2, to correspond to a preinflationary (non-de Sitter) phase of the early universe.

$$r = H^{-1}\cosh Ht_1 \sin\chi ,$$
$$\tanh H\tau = \tanh Ht_1 \sec\chi , \quad \text{(A10)}$$

so that (Fig. 6)

$$ds^2 = -d\tau^2(1-H^2r^2) + \frac{dr^2}{1-H^2r^2} + r^2 d\Omega^2 . \quad \text{(A11)}$$

Now consider the wave equation (15) in the coordinates (A4). The equation separates as in (21). The Y_{klm} are eigenfunctions of the Laplacian (19) on S_3, with $k=0,1,\ldots,$ which are orthonormal,

$$\int d\Omega_3 Y^*_{k'l'm'} Y_{klm} = \delta_{kk'}\delta_{ll'}\delta_{mm'} \quad \text{(A12)}$$

and complete

$$\sum_{k=0}^{\infty}\sum_{l=0}^{k}\sum_{m=-l}^{l} Y^*_{klm}(\Omega_3')Y_{klm}(\Omega_3)$$
$$=\delta(\Omega_3,\Omega_3')$$
$$=\frac{1}{\sin^2\chi\sin\theta}\delta(\chi-\chi')\delta(\theta-\theta')\delta(\phi-\phi') \quad \text{(A13)}$$

with respect to the measure on S_3, $d\Omega_3=\sin^2\chi\sin\theta\, d\chi\, d\theta\, d\phi$. The Y_{klm} form a basis for the $(k+1)^2$-dimensional representation of SO(4,1) (Ref. 13). From this fact and (A12), (A13) follows the addition theorem

$$\sum_{l=0}^{k}\sum_{m=-l}^{l} Y^*_{klm}(\Omega_3')Y_{klm}(\Omega_3)$$
$$=\frac{k+1}{2\pi^2}\frac{\sin[(k+1)\Omega]}{\sin\Omega}$$
$$=-\frac{1}{2\pi^2}\frac{1}{\sin\Omega}\frac{\partial}{\partial\Omega}\cos[(k+1)\Omega] , \quad \text{(A14)}$$

where

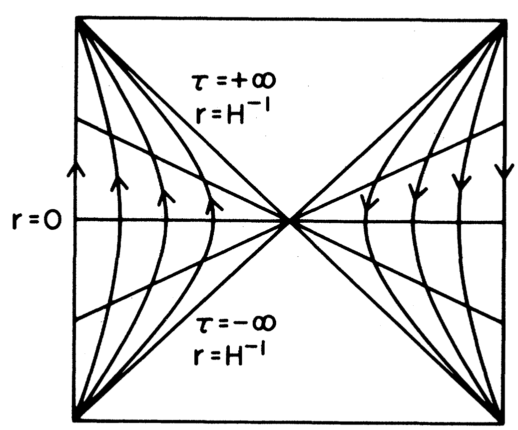

FIG. 6. The static coordinates (A10) and (A11) showing the curves of the timelike Killing field $\partial/\partial\tau$ and the particle horizon, similar to the event horizon of Figs. 1 and 2.

$$\cos\Omega \equiv \cos\chi \cos\chi' + \sin\chi \sin\chi'[\cos\theta \cos\theta' + \sin\theta \sin\theta' \cos(\phi-\phi')] \quad \text{(A15)}$$

is the cosine of the angle between Ω_3 and Ω_3' on S_3 in Eq. (49). The Y_{klm} also satisfy

$$Y^*_{klm}(\Omega_3) = Y_{kl-m}(\Omega_3) ,$$
$$Y_{klm}(\bar{\Omega}_3) = (-)^k Y_{klm}(\Omega_3) , \quad \text{(A16)}$$

where $\bar{\Omega}_3$ is the antipodal point to Ω_3 on S_3.

For the functions y_k, it is straightforward to verify that the substitution

$$y_k(t) \sim \cosh^k(Ht)\exp[\pm(k+\tfrac{3}{2})Ht \pm i\gamma Ht]F$$

leads to a hypergeometric equation for F, so that (24) and (25) are possible solutions. To demonstrate directly the de Sitter invariance of the $|\,\text{in}\rangle$ and $|\,\text{out}\rangle$ states defined by the asymptotic forms (26) and (27), consider the positive (negative) frequency condition

$$(y_{k(\pm)}, y_{k(\pm)})_\Sigma > (<) 0 \quad \text{(A17)}$$

applied to a $t_1 =$(finite) constant spacelike slice. This condition is unaffected by purely spatial rotations or translations. However, the de Sitter group O(4,1) also contains the analogs of boosts and time translations. A finite boost in the coordinates (16) sends

$$\sinh Ht_1 \to \cosh\alpha \sinh Ht_1 + \sinh\alpha \cosh Ht_1 \cos\chi . \quad \text{(A18)}$$

As $t_1 \to \pm\infty$, (A18) approaches $\pm\infty$ independently of α. Therefore, under any finite boost α, $t_1 = \pm\infty$ are invariant, although $t_1 =$finite are not. Consequently, the decomposition into positive (or negative) frequency modes according to (A17) is invariant under boosts as $t_1 \to \pm\infty$. Likewise, in the same limit, $t_1 \to t_1 + \Delta t_1$ cannot mix positive and negative frequency modes for finite Δt_1. Thus, the $|\,\text{in}\rangle$ and $|\,\text{out}\rangle$ states are invariant under the full de Sitter group SO(4,1).

To show that this decomposition is also independent of whether the spatial sections are closed, open, or flat in the limit $M \gg H$, it is sufficient to note[10] that the y_k behave like $e^{\mp iI}$ in that limit, I being the classical action for a freely falling particle:

$$I = \tfrac{1}{2}\int_{x'}^{x} g_{ab}\frac{dx^a}{ds}\frac{dx^b}{ds}ds$$
$$= \tfrac{1}{2}[\cosh^{-1}z(x,x')]^2 . \quad \text{(A19)}$$

Thus

$$\frac{\partial I}{\partial t_1} = \frac{\cosh^{-1}z}{(z^2-1)^{1/2}}\frac{\partial z}{\partial t_1} \quad \text{(A20)}$$

gives the instantaneous "energy" of the particle on a fixed t_1 slice. Using (A4), (A5), and (A8), we find a definite sign for (A20) as $t_1 \to -\infty$, t_1', χ' fixed:

$$\left[\frac{\partial I}{\partial t_1}\right]_{t_1 \to -\infty} < 0 . \quad \text{(A21)}$$

On the other hand, in the coordinates (A8) and (A9) which cover the half space in Fig. 5,

$$\left[\frac{\partial z}{\partial \hat{t}}\right]_{\hat{t}\to-\infty} = -\cosh Ht_1 e^{H\hat{t}'}\left[\frac{\partial t_1}{\partial \hat{t}}\right]_{\hat{t}\to-\infty} . \quad \text{(A22)}$$

Inspection of Fig. 5 or direct computation shows that along a line of fixed $\rho, \partial t_1/\partial\hat{t} > 0$ so that (A22) is negative and

$$\left[\frac{\partial I}{\partial \hat{t}}\right]_{\hat{t}\to-\infty} < 0 . \quad \text{(A23)}$$

Since the spectral decomposition according to (A17) is equivalent to the sign of the derivative of I normal to Σ (for $M \gg H$), the $|\,\text{in}\rangle$ state defined with respect to $t_1 \to -\infty$ (closed spatial section) is identical to the vacuum state defined with respect to $\hat{t} \to -\infty$ (flat spatial section). The open case follows immediately from Fig. 4 since $t_1 \to -\infty$ implies $t_{-1} \to -\infty$ as well. Likewise, $t_1 \to +\infty$ is equivalent to $\hat{t} \to +\infty, t_{-1} \to +\infty$, so the $|\,\text{out}\rangle$ states are the same in all three cases also.

Thus the Green's functions calculated in Sec. IV for the $|\,\text{in}\rangle$ and $|\,\text{out}\rangle$ states are de Sitter invariant and applicable in the half space of Fig. 5—independently of the excluded half spacetime. This is analogous to the excluded region of Fig. 2 and shows that the $|\,\text{in}\rangle$ state is the analog of the Unruh state of the Schwarzschild case, when $M \gg H$, i.e., when the creation rate (41) is small. This is precisely the initial condition we would want for a realistic model of inflation, where the excluded half space must be replaced by the spacetime which preceded the inflationary epoch, about which we presume nothing.

For evaluating the two-point functions in Sec. IV, we first note another useful representation of the $y_k(t)$, obtained by setting

$$\zeta = i\sinh Ht, \quad y_k(t) = (\zeta^2-1)^{-1/2}W_k(\zeta) . \quad \text{(A24)}$$

Then Eq. (23) may be put in the form

$$\frac{d}{d\zeta}\left[(1-\zeta^2)\frac{dW_k}{d\zeta}\right] + \left[\left[2-\frac{M^2}{H^2}\right] - \frac{(k+1)^2}{1-\zeta^2}\right]W_k = 0 \quad \text{(A25)}$$

which is Legendre's equation, with $\mu^2 = (k+1)^2, \nu(\nu+1) = 2 - M^2/H^2$. Set

$$\mu = k+1, \quad \nu = -\tfrac{1}{2} + i\gamma , \quad \text{(A26)}$$

γ given by Eq. (10).

Then, using Sec. 3.2 of Ref. 9, the solutions (24) and (25) may be expressed in terms of Legendre functions. For example,

$$y_k^{(+)}(t<0) \sim (\zeta^2-1)^{-1/2}Q^{\mu}_{-\nu-1}(\zeta) . \quad \text{(A27)}$$

By virtue of (28) and (29) similar relations hold for the other solutions.

In the complex ζ plane the functions $P_\nu^\mu(\zeta)$ and $Q_\nu^\mu(\zeta)$ are analytic everywhere except on the cut along the real axis from $-\infty$ to $+1$. When μ is an even integer P_ν^μ is analytic from -1 to $+1$ as well.

However, the function $Q_\nu^\mu(\zeta)$ has logarithmic singularities at both $\zeta=\pm1$. Thus, from (A27), the positive and negative frequency decomposition of the in and out basis defined by (24) and (25) is *not* regular on the Euclidean section (ζ real, between -1 and $+1$). For constructing the Euclidean vacuum we should use only the P_ν^μ functions, as in (44).

In order to evaluate the Euclidean two-point function, Eq. (48), we first employ (A14) to perform the sums over l and m. Then the addition theorem of the Legendre functions[9,20]

$$P_\nu(\zeta\zeta'-(\zeta^2-1)^{1/2}(\zeta'^2-1)^{1/2}\cos\Omega)=P_\nu(\zeta)P_\nu(\zeta')+2\sum_{k=0}^{\infty}(-)^{k+1}P_\nu^{k+1}(\zeta)P_\nu^{-(k+1)}(\zeta')\cos[(k+1)\Omega]$$

valid for

$$\mathrm{Re}\,\zeta(\zeta')>0,\quad |\arg(\zeta(\zeta')-1)|<\pi \tag{A28}$$

may be used. Because of the restrictions on ζ and ζ', some care is necessary in applying the theorem. If we let $t=t_R+it_I$, the restrictions will be satisfied if (44) is understood as defining $y_k(t)$ in the strip of the complex t plane given by $-\pi/2H<t_I<0$, with $\theta(\pm t)$ replaced by $\theta(\pm t_R)$. In principle, the theorem must be applied to the four cases $t_R\gtrless 0, t_R'\gtrless 0$ separately. Actually, since $y_k(t)$ is regular as $t_R\to 0$ from above or below, the result in all four cases is the same and it suffices to consider only $t_R>0, t'_R>0$. In that case

$$\zeta=-\cosh Ht_R\sin Ht_1+i\sinh Ht_R\cos Ht_I \tag{A29}$$

so $\mathrm{Im}\zeta>0$ but $\mathrm{Im}\zeta'^*<0$.

Since $y_k(t)y_k^*(t')$ appears in (48) the addition theorem must be used with ζ' replaced by $\zeta'^*\to-\zeta'$ as $t_I'\to 0$. The result, as calculated in Ref. 12, is identical to Eq. (5.2) of the text.

When the general de Sitter invariant state is considered, the two-point function, (56) involves two kinds of terms. The first kind features $y_k(t)y_k^*(t')$ or $y_k(t')y_k^*(t)$ which is of the same form of $G_0(x,x')$ itself. Thus, the first two terms of (56) sum to give the first term of (57). The last two terms, involving $y_k(t)y_k(t')$ give

$$-2\sinh2\theta\,\mathrm{Im}\left[\sum_{klm}(-)^k y_k(t)y_k(t')Y_{klm}(\Omega_3)Y_{klm}^*(\Omega_3')\right]. \tag{A30}$$

Again we consider $t=t_R+it_I$ with $t_R>0$, $Ht_I\in(-\pi/2,0)$, and similarly for t'. From (44), the product y_ky_k contains a factor i^{2k} which compensates the $(-)^k$ in (A30).

As t_I,t_I' are varied from $-\pi/2H$ to zero, ζ and ζ' pick up the same phase factor, $e^{i\pi/2}$. This means that $Z=\zeta\zeta'-(\zeta^2-1)^{1/2}(\zeta'^2-1)^{1/2}\cos\Omega$, initially real and greater than one, picks up a factor of $e^{i\pi}$ after continuation to real t,t'. Since $P_\nu(Z)$ has a logarithmic branch cut from $-\infty$ to -1, the continuation carries the function onto its second Riemann sheet. The value of the sum (A30) can then be found by expressing the result of application of the addition theorem (A28) to (A30) in terms of the hypergeometric function, viz.,

$$-\frac{H^2}{8\pi}\mathrm{sech}\pi\gamma\frac{dP_\nu(Z)}{dZ}=-\frac{\nu(\nu+1)}{16\pi}H^2\mathrm{sech}\pi\gamma F\left[-\nu+1,\nu+2;2;\frac{1-Z}{2}\right]. \tag{A31}$$

Then, using the analytic continuation of the hypergeometric series[9] for $Z\to-1+|Z+1|e^{i\pi}$ we find

$$-\nu(\nu+1)\frac{H^2}{16\pi}\frac{\mathrm{sech}\pi\gamma(i\pi)}{\Gamma(-\nu)\Gamma(\nu+1)}F\left[\nu+1,\nu+2;2;\frac{1+Z}{2}\right] \tag{A32}$$

for the result on the second sheet. Since $\nu=-\frac{1}{2}+i\gamma$, $\Gamma(-\nu)\Gamma(\nu+1)=\pi\,\mathrm{sech}\pi\gamma$, and the imaginary part in (A30) is simply

$$\frac{H^2}{8\pi}\frac{d}{dz}P_\nu(-Z).$$

After continuation, $Z\to e^{i\pi}Z=-z(x,x')$ so that (A30) becomes finally

$$2\sinh2\theta\frac{H^2}{8\pi}\frac{d}{dz}P_\nu(+z)=-\frac{\sinh2\theta}{\mathrm{sech}\pi\gamma}G_0(x,\bar{x}). \tag{A33}$$

Thus, Eq. (57) is the result of the calculation. To demonstrate Eqs. (58) it is necessary to multiply Eqs. (24) and (25) by their proper normalization factor of $2^{k+1}H\gamma^{-1/2}$ and then use (44) to express the Euclidean y_k in terms of the correctly normalized $y_{k(\pm)}$ or $y_k(\pm)$, respectively. This requires the relation (3.2) (21) of Ref. 9. For example, defining $\alpha_k^{\mathrm{out}},\beta_k^{\mathrm{out}}$ by analogy with (9),

$$y_k=\alpha_k^{\mathrm{out}}y_k^{(+)}+\beta_k^{\mathrm{out}}y_k^{(-)} \tag{A34}$$

gives

$$-\frac{\beta_k^{\mathrm{out}*}}{\alpha_k^{\mathrm{out}}}=i(-)^k e^{-\pi\gamma}=-(-)^k\tanh\bar{\theta} \tag{A35}$$

and

$$\frac{\alpha_k^{\mathrm{out}}}{\alpha_k^{\mathrm{out}*}}=\frac{(-)^k e^{-3i\pi/2}\Gamma(-i\gamma)\Gamma(-\frac{1}{2}-k+i\gamma)}{\Gamma(i\gamma)\Gamma(-\frac{1}{2}-k-i\gamma)}=e^{-2i\delta_k} \tag{A36}$$

with $\bar{\theta}$ and $\bar{\delta}_k$ defined by Eqs. (32) and (33). Then defining $\alpha_k^{\rm in},\beta_k^{\rm in}$ by

$$y_{k(+)}=\alpha_k^{\rm in}y_k+\beta_k^{\rm in}y_k^* \quad \text{(A37)}$$

and using Eq. (29) we easily demonstrate that

$$\alpha_k^{\rm in}=\alpha_k^{\rm out}, \quad \beta_k^{\rm in}=-\beta_k^{\rm out*} , \quad \text{(A38)}$$

and, furthermore, that the α_k, β_k of Eqs. (9) and (11) are given by

$$\alpha_k=(\alpha_k^{\rm out})^2-(\beta_k^{\rm out*})^2, \quad \beta_k=2i\,{\rm Im}(\alpha_k^{\rm out}\beta_k^{\rm out}) . \quad \text{(A39)}$$

Together, relations (A35)—(A39) imply Eqs. (58) and (60) of the text.

[1]W. Boucher and G. W. Gibbons, in *The Very Early Universe,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University Press, Cambridge, 1983).

[2]S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).

[3]J. Schwinger, Phys. Rev. **82**, 664 (1951); **93**, 615 (1954).

[4]A. I. Nikishov, Zh. Eksp. Teor. Fiz. **57**, 1210 (1969) [Sov. Phys. JETP **30**, 660 (1970)]; N. B. Narozhnyi and A. I. Nikishov, Teor. Mat. Fiz. **26**, 16 (1976) [Theor. Math. Phys. **26**, 9 (1976)], and references cited therein. See also E. Brezin and C. Itzykson, Phys. Rev. D **2**, 1191 (1970).

[5]H. Rumpf, Phys. Lett. **61B**, 272 (1976); Nuovo Cimento **35B**, 321 (1976); H. Rumpf and H. K. Urbantke, Ann. Phys. (N.Y.) **114**, 332 (1978).

[6]W. G. Unruh, Phys. Rev. D **14**, 870 (1976).

[7]J. B. Hartle and S. W. Hawking, Phys. Rev. D **13**, 2188 (1976).

[8]M. Gutzwiller, Helv. Phys. Acta **29**, 313 (1956).

[9]*Higher Transcendental Functions* (Bateman Manuscript Project), Vol. 1, edited by A. Erdélyi (McGraw-Hill, New York, 1953).

[10]L. Parker, Phys. Rev. **183**, 1057 (1969).

[11]J. S. Dowker and R. Critchley, Phys. Rev. D **13**, 224 (1976); **13**, 3224 (1976).

[12]E. A. Tagirov, Ann. Phys. (N.Y.) **76**, 561 (1973). Equation (16) of this paper contains a typographical error.

[13]N. J. Vilenkin, *Special Functions and the Theory of Group Representations,* Translations of Mathematical Monographs, Vol. 22 (Amer. Math. Soc., Providence, 1968).

[14]P. Candelas and D. J. Raine, Phys. Rev. D **12**, 965 (1975).

[15]T. S. Bunch and P. C. W. Davies, Proc. R. Soc. London **A360**, 117 (1978).

[16]N. A. Chernikov and E. A. Tagirov, Ann. Inst. Henri Poincaré **9**, 109 (1968).

[17]C. J. C. Burges, Brandeis Report No. 158, 1984 (unpublished).

[18]B. S. DeWitt, Phys. Rep. **19C**, 297 (1975); N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, New York, 1982).

[19]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, New York, 1973).

[20]E. W. Hobson, *The Theory of Spherical and Ellipsoidal Harmonics* (Cambridge University Press, Cambridge, 1931).

[21]C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).