PHYSICAL REVIEW D VOLUME 32, NUMBER 12 15 DECEMBER 1985

Vacuum states in de Sitter space

Bruce Allen*
Department of Physics, University of California, Santa Barbara, California 93106
(Received 27 February 1985; revised manuscript received 26 August 1985)

We examine possible vacuum states for scalar fields in de Sitter space, concentrating on those states (1) invariant under the de Sitter group O(1,4) or (2) invariant under one of its maximal subgroups E(3). For massive fields there is a one-complex-parameter family of de Sitter-invariant states, which includes the "Euclidean" vacuum state as a special case. We show these states are generated from the Euclidean vacuum by a frequency-independent Bogoliubov transformation, and obtain formulas for the symmetric, antisymmetric, and Feynman functions. In the massless minimally coupled case we prove that there exists no de Sitter-invariant Fock vacuum state. However one can find Fock states which are E(3) invariant. These states include the Bunch-Davies and Ottewill-Najmi vacua as special cases.

INTRODUCTION

de Sitter space is interesting for two reasons. First, it provides a promising "inflationary" model of the very early universe.[1] Second, it is a highly symmetric curved space, in which one can quantize fields and obtain simple exact solutions.[2] This paper is about the various vacuum states which a quantized field can have in de Sitter space. The choice of such a state is a necessary boundary condition (or initial condition) in the construction of a realistic inflationary model, and it sheds light on more general questions. In particular, the problem of how to select a physically meaningful vacuum state in a general curved space is an outstanding unsolved problem.

There are potentially two kinds of vacuum states in de Sitter space: those which respect de Sitter invariance, and those which break it. The de Sitter-invariant states are those which "look the same" to any freely falling observer, anywhere in de Sitter space. The other possible states are those which break de Sitter invariance by singling out some family of observers. These states have special properties as seen by those observers, and are appropriate for certain inflationary models.

The paper has four sections. In the first section, we describe de Sitter space, its symmetry group, and some of its properties. In a de Sitter-invariant state, the two-point function $\langle \Phi(x)\Phi(y) \rangle$ only depends upon the geodesic distance between the two points x and y. For this reason, we introduce the geodesic distance as a spacetime coordinate. We also discuss the properties of antipodal points, which are used in later sections. The antipodal point of x is the unique point "directly across de Sitter space" from x.

In the second section, we show why a unique vacuum state is *not* selected *only* by requiring that it be de Sitter invariant. This is because (for a massive scalar field) the set of de Sitter-invariant states forms a one-complex-parameter family. The familiar vacuum state known as the "Euclidean" vacuum is a particular member of this family. We show that the other members of this family can be generated from the Euclidean vacuum by a simple frequency-independent mode-mixing Bogoliubov transformation. If one chooses a basis of Euclidean modes with the correct transformation property under point $\rightarrow$ antipodal point, it is trivial to obtain formulas for the two-point functions (symmetric, antisymmetric, Feynman) in the general vacuum state.

However, the situation is rather different for the massless (minimally coupled) case, which the rest of the paper is about. The third section contains our main result. We prove that there is *no* Fock vacuum state in which the massless two-point function is de Sitter invariant. The proof exploits the antipodal point transformations of Sec. I and the de Sitter-invariant two-point functions of Sec. II.

Section IV examines the implications of this result. Although no de Sitter-invariant vacuum state exists, one may still look for states with as much symmetry as possible. These states break de Sitter invariance but are invariant under some *maximal subgroup* of the full de Sitter group. The largest (i.e., maximal) subgroups of the de Sitter group are O(4), O(1,3), and E(3). We treat the last case, and find a family of E(3)-invariant states. The O(4)- and O(1,3)-invariant vacua are left for a later date.

Note that throughout this paper we use Planck units $\hbar=c=G=1$.

I. DE SITTER SPACE

de Sitter space can be constructed as follows. Take a five-dimensional flat space R^5 with a metric $\eta_{ab}=\mathrm{diag}(-1,1,1,1,1)$ and consider the four-dimensional surface defined by all five-vectors X^a which satisfy

$$X^a X^b \eta_{ab} = H^{-2} . \tag{1.1}$$

This surface is a hyperboloid, as shown in Fig. 1. If we now induce the natural metric on this surface, by considering it as a subspace of (R^5,η_{ab}) then the resulting spacetime is called de Sitter space.[3] The Hubble constant H fixes the rate of expansion of the spatial sections.

By construction, de Sitter space is a maximally symmetric homogeneous space which has constant positive

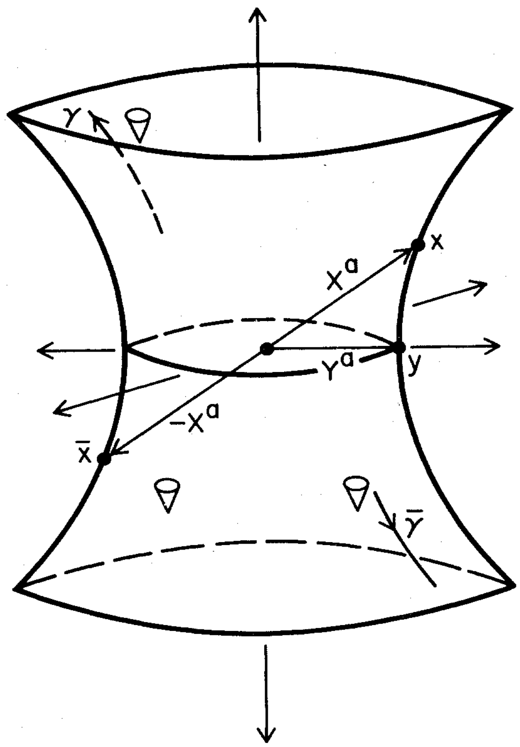

FIG. 1. de Sitter space is a hyperboloid defined by (1.1). Just as for a sphere, every point x has an antipodal point $\bar{x}$. The isometry $x \to \bar{x}$ is not continuously connected to the identity, because it reverses the direction of time. A future directed curve γ is sent to a past-directed curve $\bar{\gamma}$ by this operation.

scalar curvature $R = 12H^2$. Its symmetry group is the de Sitter group O(1,4), which is composed of four disconnected components. Let G denote the component which contains the identity element. G is called the connected de Sitter group, and is analagous to the proper orthochronous Poincaré group in flat space.

The other three components of O(1,4) can be easily obtained from G. Consider two elements of O(1,4),

$$T \equiv \text{diag}(-1,1,1,1,1)\ ,$$
$$S \equiv \text{diag}(1,-1,1,1,1)\ , \tag{1.2}$$

which correspond to time reversal and space reflection, respectively. If we multiply all the elements of G by T one gets a set which one can denote G_T,

$$G_T \equiv \{ g \cdot T \mid g \in G \}\ , \tag{1.3}$$

and in a similar fashion one can form G_S and G_{TS}. These sets are the other three components of O(1,4). We define a de Sitter-invariant state as one which is invariant under the action of all four components of O(1,4). We will also encounter some states which are invariant under the connected part of the de Sitter group G, but which are *not* invariant under the action of the other three components of O(1,4). We will be using the following element of G_T, which is called "antipodal transformation":

$$A = \text{diag}(-1,-1,-1,-1,-1)\ . \tag{1.4}$$

This operation sends a point x, located by the five-vector $X^a(x)$ to its antipodal point (which we denote by $\bar{x}$) located by the five-vector $-X^a$, so $X^a(\bar{x}) = -X^a(x)$. This is shown in Fig. 1.

Later on it will also be useful to have a simple expression for the geodesic distance between two points x and y. If we define the real function of x and y

$$Z(x,y) \equiv H^2 \eta_{ab} X^a(x) Y^b(y) \tag{1.5}$$

then the geodesic distance between the points x and y is

$$d(x,y) = H^{-1} \cos^{-1} Z(x,y)\ . \tag{1.6}$$

This is similar to the expression for a sphere (where $\cos^{-1} Z$ becomes $\cos^{-1}\theta$) so one can think of $\cos^{-1} Z$ as a hyperbolic "angle" between the points x and y. $Z(x,y)$ has the important property that it changes sign if one point is sent to its antipodal point,

$$Z(x,y) = -Z(\bar{x},y)\ . \tag{1.7}$$

This follows from the definition of Z, since the operation of sending x to $\bar{x}$ simply changes the overall sign of the five-vector X^a.

The range of Z is shown on a conformal diagram[3] in Fig. 2. Since all points in de Sitter space are equivalent, we fix the point y as shown in Fig. 1, and allow x to move around. If x is on the light cone of y then $Z = 1$, if x is to the future or past of y then $Z > 1$, and if x and y are spacelike separated then $Z < 1$. Hence, from (1.6) the timelike geodesics have imaginary length and the spacelike geodesics have real length.

The geodesic distance

$$d(x,y) = \int_x^y (\eta_{ab} \dot{X}^a \dot{X}^b)^{1/2} d\lambda$$

is a function which is invariant under both G and G_T. However, it is also possible to define a "signed" distance

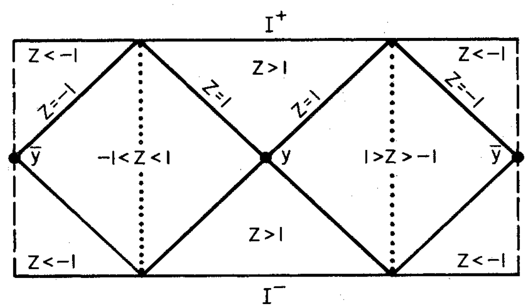

FIG. 2. de Sitter space is conformal to a cylinder, which is shown "unwrapped." The left and right edges of this diagram must be identified along the dashed line. The range of $Z-1$ is positive for points timelike related to y and negative for points spacelike related to y. Without loss of generality, the point y can be taken to be *any* point of de Sitter space. The surface $Z = 0$ is shown as a vertical dotted line. It is "halfway" between the point y and its antipodal point $\bar{y}$.

function $\tilde{d}(x,y)$ which is only invariant under G. This signed function carries a bit of extra information, because if x and y are timelike separated, it tells you which one lies to the future of the other.

To define $\tilde{d}(x,y)$ take

$$\tilde{d}(x,y)=H^{-1}\cos^{-1}\tilde{Z}(x,y)$$

where

$$\tilde{Z}(x,y)=\begin{cases} H^2\eta_{ab}X^a(x)X^b(y)+i\epsilon & \text{if } x \text{ to future of } y\ , \\ H^2\eta_{ab}X^a(x)X^b(y)-i\epsilon & \text{if } x \text{ to past of } y\ , \end{cases} \tag{1.8}$$

where ϵ is a positive real infinitesimal. Then for $g\in G$, $\tilde{d}(gx,gy)=\tilde{d}(x,y)$ so $\tilde{d}$ is still invariant under G. However, for $A\in G_T$, one can show that

$$\tilde{d}(x,y)\neq\tilde{d}(\bar{x},\bar{y})=\tilde{d}(Ax,Ay)$$

so that $\tilde{d}$ is not invariant under G_T. Suppose we take two timelike-separated points x and y, with x to the future of y. We can smoothly send (x,y) to $(\bar{x},\bar{y})$ in the following way. First bring x close to y, keeping it within y's forward light cone. Carry them together to $\bar{y}$, keeping x inside y's forward light cone during the journey. With y now at $\bar{y}$, carry x *through* y, into $\bar{y}$'s past light cone. Then, bring x to $\bar{x}$ along a path that stays within $\bar{y}$'s past light cone. During this journey, $\tilde{Z}(x,y)$ starts above the real axis, passes around $\tilde{Z}=1$, and ends up below the real axis, as shown in Fig. 3. Now $\cos^{-1}\tilde{Z}$ has a branch cut from $\tilde{Z}=1$ to ∞ along the real axis, and changes sign across it. Consequently, for timelike-separated points x and y

$$\tilde{d}(x,y)=-\tilde{d}(\bar{x},\bar{y}) \tag{1.9}$$

so $\tilde{d}(x,y)$ is not invariant under G_T.

Eventually, we will need a coordinate system that covers de Sitter space. There are several well-known systems.[3] For our purpose it is convenient to use two spatially flat coordinate patches. These are shown on the conformal diagram in Fig. 4. The metric is

$$ds^2\equiv H^{-2}t^{-2}(-dt^2+d\mathbf{x}^2)\ , \tag{1.10}$$

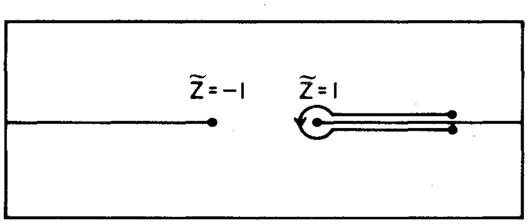

FIG. 3. One sheet of the Riemann surface for $\cos^{-1}\tilde{Z}$, which has branch points at $\tilde{Z}=\pm1$, and cuts to $\pm\infty$ as shown. Just above the right-hand cut the function is $i\,|\cos^{-1}\tilde{Z}\,|$ and just below the cut it is $-i\,|\cos^{-1}\tilde{Z}\,|$. When a spacetime point x passes from the future to the past of y, $\tilde{Z}(x,y)$ circles around $\tilde{Z}=1$ as shown. Consequently, the distance $\tilde{d}=H^{-1}\cos^{-1}\tilde{Z}$ changes sign, and $\tilde{d}(x,y)=-\tilde{d}(\bar{x},\bar{y})$.

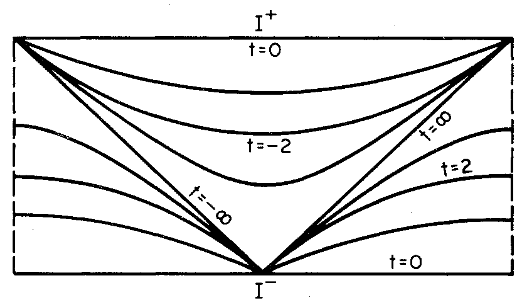

FIG. 4. This conformal diagram of de Sitter space shows how two patches of spatially flat coordinates cover de Sitter space. The left and right edges of this diagram must be identified along the dotted lines. If a point has coordinates $(t,\mathbf{x})$ then its antipodal point has coordinates $(-t,\mathbf{x})$.

where $d\mathbf{x}^2=dx_1{}^2+dx_2{}^2+dx_3{}^2$. The patch $t\in(-\infty,0)$ covers the "upper-half" of de Sitter space up to I^+ and the patch $t\in(0,\infty)$ covers the "lower-half" down to I^-. (Formally one must cover a strip around $t=\pm\infty$ with another patch, but in practice this is unnecessary.) A notable feature of our coordinate system is that if a point has coordinates $(t,\mathbf{x})$ then its antipodal point has coordinates $(-t,\mathbf{x})$. If a point x has coordinates $(t,\mathbf{x})$ and another point y has coordinates $(t',\mathbf{x}')$ then[4,5]

$$Z(x,y)=\frac{t^2+t'^2-(\mathbf{x}-\mathbf{x}')^2}{2tt'}\ , \tag{1.11}$$

where $\mathbf{x}^2$ means the three-dimensional Cartesian dot product $\mathbf{x}\cdot\mathbf{x}$. Note that changing the sign of either t or t' changes the sign of Z, so that $Z(x,y)=-Z(\bar{x},y)$.

II. DE SITTER-INVARIANT VACUUM STATES FOR A SCALAR FIELD

In this section, we examine the de Sitter-invariant states for a real scalar field. These states were first described by Chernikov[6] and Tagirov[7] and have also been studied by Schomblond[4,5] and most recently by Mottola.[8] In Sec. II A we find all symmetric two-point functions which are de Sitter invariant. In Sec. II B we show how the Euclidean vacuum state can be used to obtain all de Sitter-invariant states, and in Sec. II C we examine the effect of time reversal T on these states.

A. Time-symmetric solutions of the wave equation for $G^{(1)}(x,y)$

Consider the symmetric two-point function

$$G^{(1)}_\lambda(x,y)=\langle\lambda\,|\,\Phi(x)\Phi(y)+\Phi(y)\Phi(x)\,|\,\lambda\rangle \tag{2.1}$$

in a de Sitter-invariant state $|\,\lambda\rangle$. (The label λ is to remind the reader that there is more than one such state.) We assume $|\,\lambda\rangle$ is invariant under the *full* disconnected group O(1,4), which implies that $G^{(1)}_\lambda(x,y)$ can only depend upon the two spacetime points x and y via the geodesic distance $d(x,y)$. [In flat space, Lorentz invariance means that it can only depend on $\sigma=(\mathbf{x}-\mathbf{y})^2-(x^0-y^0)^2$.] Because the geodesic distance $d(x,y)$ is a function of $Z(x,y)$, one may write

$$G^{(1)}(x,y)=G^{(1)}(d(x,y))=F(Z(x,y))=F(Z) . \quad (2.2)$$

Note that if $|\lambda\rangle$ were only invariant under the connected part $G\subset SO(1,4)$ then $G^{(1)}$ could be of the form

$$F_1(Z)\theta(x,y)+F_2(Z)\theta(y,x) ,$$

where $\theta(x,y)=(0,\frac{1}{2},1)$ if x is (past, spacelike, future) separated from y. The two-point function obeys the massive scalar field equation

$$(-\Box_x+m^2)G^{(1)}(x,y)=0 \quad (2.3)$$

which can be expressed in terms of the variable Z as[4]

$$\left[(Z^2-1)\frac{d^2}{dZ^2}+4Z\frac{d}{dZ}+m^2H^{-2}\right]F(Z)=0 . \quad (2.4)$$

This second-order equation admits two fundamental real solutions. Suppose that the first solution is $f(Z)$. Then since the differential equation (2.4) is invariant under $Z\to -Z$, another solution is $f(-Z)$.

The hypergeometric function provides a fundamental real solution to (2.4). Let c be either root of the quadratic equation

$$c(c-3)+m^2H^{-2}=0 . \quad (2.5)$$

Then

$$f(Z)={}_2F_1(c,3-c,2,\tfrac{1}{2}(1+Z)) \quad (2.6)$$

is a solution of the wave equation (2.4). For $m^2>0$, this solution has as its *only* singularity a simple pole at $Z=1$, and unless $m^2H^{-2}=2$ (the conformal massless case) it also has a branch point at $Z=1$. The branch cut runs along the real axis from $Z=1$ to ∞. In order that $G^{(1)}(x,y)$ be real, the average value across the cut (or equivalently, the real part) should be taken when $Z>1$. Note that (2.6) does not depend upon which root of (2.5) one takes for c, because changing roots of (2.5) flips the first two arguments of (2.6) and

$${}_2F_1(a,b,c,Z)={}_2F_1(b,a,c,Z) .$$

For $m^2>0$, $f(Z)$ and $f(-Z)$ are linearly independent. Hence the general solution to (2.4) is of the form

$$F(Z)=af(Z)+bf(-Z) . \quad (2.7)$$

Here the choice of real constants a and b determines the particular solution. The general solution (2.7) has simple poles at $Z=1$ and -1, and the residues at these points are determined by a and b, respectively. If x is on the light cone of y then $Z=1$ and the first term blows up. Similarly, if x is on the light cone of $\bar{y}$ (the antipodal point of y) then $Z=-1$ and the second term blows up. On the Euclidean section of de Sitter space, which is a four-sphere, the situation is exactly the same. If x is at the north pole, then $Z=1$ when y is also at the north pole, and $Z=-1$ when y is at the south pole.

The "Euclidean" vacuum, which we will denote by $\lambda=0$, is the one in which the two-point function is only singular if x is on the light cone of y (or $x=y$ on S^4). In this state $b=0$ and the two-point function has only one singular point. This is also the "unique" vacuum of Schomblond.[5] The value of the constant a in the "Euclidean" vacuum state is determined by the canonical commutation relations of Φ and $\dot{\Phi}$, and is[5]

$$a=(8\pi)^{-1}H^2(m^2H^{-2}-2)\sec\pi(\tfrac{9}{4}-m^2H^{-2})^{1/2} .$$

For this value of a, the σ^{-1} and $\ln\sigma$ short-distance singularities in $G_0^{(1)}$ have the Hadamard form.[2] For example, the coefficient of the σ^{-1} term is the same as it would be in flat space.

B. The general de Sitter-invariant state obtained from the Euclidean vacuum

There is a particular set of modes $\phi_n(x)$ which are orthonormal and which, via canonical quantization, serve to define the "Euclidean" vacuum.[4–8] The fundamental idea of this section is that from these "Euclidean" modes ϕ_n one can define a *new* set of modes $\tilde{\phi}_n$, and that the new "vacuum" defined by these modes is also a de Sitter-invariant state.

The new modes $\tilde{\phi}_n(x)$ are defined by a Bogoliubov transformation

$$\tilde{\phi}_n(x)=A\phi_n(x)+B\phi_n^*(x) \quad (2.8)$$

with two properties. First, it is "frequency" or "mode" independent, so A and B are *constants* which are independent of n. Second, the transformation preserves orthonormality and thus defines a new "vacuum" state. Since

$$(\tilde{\phi}_m,\tilde{\phi}_n)=(|A|^2-|B|^2)(\phi_m,\phi_n)$$
$$=(|A|^2-|B|^2)\delta_{nm} ,$$

the second condition implies that $|A|^2-|B|^2=1$. We seek the *most general* transformation with this property. If A and B were real, then we could let $A=\cosh\alpha$ and $B=\sinh\alpha$ for some constant real parameter α. In fact, A and B are complex so one needs $A=(\cosh\alpha)e^{i\gamma}$ and $B=(\sinh\alpha)e^{i(\gamma+\beta)}$ for real constants α,β,γ. Because the *overall* phase $e^{i\gamma}$ of $\tilde{\phi}_n$ ultimately disappears from the expectation values, it is sufficient to consider the two-parameter (α,β) family of A and B given by $\gamma\equiv 0$:

$$\tilde{\phi}_n(x)=(\cosh\alpha)\phi_n(x)+(e^{i\beta}\sinh\alpha)\phi_n^*(x) . \quad (2.9)$$

The complete set of A,B satisfying the two conditions given above, without the overall phase $e^{i\gamma}$, corresponds to the parameter range

$$\alpha\in[0,\infty),\ \ \beta\in(-\pi,\pi) . \quad (2.10)$$

The "Euclidean" vacuum $A=1$ and $B=0$ corresponds to $\alpha=0$. Now we have to prove that the new "vacuum" state defined by (2.9) is in fact de Sitter invariant when $\alpha\neq 0$.

To prove this would not be easy, except for the following trick. The mode functions $\phi_n(x)$ which define the "Euclidean" vacuum can be *chosen* to satisfy

$$\phi_n(\bar{x})=\phi_n^*(x) , \quad (2.11)$$

where $\bar{x}$ is the antipodal point to x. [Note that Appendix A shows that a basis of modes satisfying (2.11) can be obtained from the conventional "Euclidean" vacuum modes of Tagirov[7] and Mottola[8] by a trivial Bogoliubov transformation. The transformation does not mix positive- and negative-norm states, and therefore defines a physically equivalent vacuum state.] Now we examine the two-point functions in the (α,β) state. Because the free field theory may be defined in terms of its two-point functions, we need to show that these two-point functions are de Sitter invariant.

The symmetric two-point function (2.1) in the (α,β) state is given by the mode sum

$$G^{(1)}_{\alpha,\beta}(x,y)=\langle \alpha,\beta \mid \Phi(x)\Phi(y)+\Phi(y)\Phi(x) \mid \alpha,\beta\rangle = \sum_n [\tilde{\phi}_n(x)\tilde{\phi}^*_n(y)+\tilde{\phi}^*_n(x)\tilde{\phi}_n(y)] \quad (2.12)$$

and the commutator function by the mode sum

$$iD_{\alpha,\beta}(x,y)=\langle \alpha,\beta \mid \Phi(x)\Phi(y)-\Phi(y)\Phi(x) \mid \alpha,\beta\rangle = \sum_n [\tilde{\phi}_n(x)\tilde{\phi}^*_n(y)-\tilde{\phi}^*_n(x)\tilde{\phi}_n(y)] \,. \quad (2.13)$$

We denote the two-point functions in the "Euclidean" vacuum $\alpha=0$ by $G^{(1)}_0$ and iD_0. Now substituting (2.9) into (2.12) one obtains

$$G^{(1)}_{\alpha,\beta}(x,y)=(\cosh 2\alpha)\sum_n [\phi_n(x)\phi^*_n(y)+\phi^*_n(x)\phi_n(y)]+(\sinh 2\alpha)\cos\beta\sum_n [\phi_n(x)\phi_n(y)+\phi^*_n(x)\phi^*_n(y)] + i(\sinh 2\alpha)\sin\beta\sum_n [\phi^*_n(x)\phi^*_n(y)-\phi_n(x)\phi_n(y)] \,. \quad (2.14)$$

The first term in (2.14) is proportional to $G^{(1)}_0(x,y)$. The second and third terms would be intractable, *but* the "Euclidean" modes have been chosen to obey $\phi^*_n(x)=\phi_n(\bar{x})$ (2.11). Using this relation, these terms become proportional to $G^{(1)}_0(\bar{x},y)$ and $D_0(\bar{x},y)$. Hence, one obtains from (2.14)

$$\begin{aligned}G^{(1)}_{\alpha,\beta}(x,y)&=\cosh(2\alpha)G^{(1)}_0(x,y)+\sinh(2\alpha)[\cos\beta G^{(1)}_0(\bar{x},y)-\sin\beta D_0(\bar{x},y)]\\ &=\cosh(2\alpha)G^{(1)}_0(Z)+\sinh 2\alpha[\cos\beta G^{(1)}_0(-Z)-\sin\beta D_0(\bar{x},y)] \,,\end{aligned} \quad (2.15)$$

where $G^{(1)}_0(-Z)$ has been obtained in (2.15) by using $Z(\bar{x},y)=-Z(x,y)$ and the result from Sec. II A that $G^{(1)}_0(x,y)$ is only a function of Z. The right-hand side (RHS) of (2.15) is clearly de Sitter invariant when $\beta=0$ because it is only a function of Z. The case $\beta\neq 0$ will be discussed shortly.

Turning now to the other two-point functions, one can easily verify that the commutator function is the same for all the different de Sitter-invariant states

$$iD_{\alpha,\beta}(x,y)=\sum_n [\tilde{\phi}_n(x)\tilde{\phi}^*_n(y)-\tilde{\phi}^*_n(x)\tilde{\phi}_n(y)]=iD_0(x,y) \,. \quad (2.16)$$

The Feynman function is the time-ordered expectation value

$$iG^F(x,y)=\theta(x,y)\langle \Phi(x)\Phi(y)\rangle+\theta(y,x)\langle \Phi(y)\Phi(x)\rangle \,, \quad (2.17)$$

where $\theta(x,y)=1$ if x is in the forward light cone of y, $\theta(x,y)=0$ if x is in the past light cone of y, and $\theta(x,y)=\frac{1}{2}$ if x and y are null or spacelike separated. This time-ordered product can be expressed in terms of $G^{(1)}$ and D as

$$iG^F_{\alpha,\beta}(x,y)=\tfrac{1}{2}G^{(1)}_{\alpha,\beta}(x,y)+\tfrac{1}{2}i\epsilon(x,y)D_{\alpha,\beta}(x,y) \,, \quad (2.18)$$

where $\epsilon(x,y)=\theta(x,y)-\theta(y,x)$. One can then use the expressions for $G^{(1)}_{\alpha,\beta}$ and $iD_{\alpha,\beta}$ [Eqs. (2.15) and (2.16)] to find $G^F_{\alpha,\beta}$, in terms of G^F_0, the Feynman function in the "Euclidean" vacuum:

$$iG^F_{\alpha,\beta}(x,y)=iG^F_0(x,y)+\tfrac{1}{2}[G^{(1)}_{\alpha,\beta}(x,y)-G^{(1)}_0(x,y)] \,. \quad (2.19)$$

One can see that when $\alpha\neq 0$ a homogeneous piece is added to the Feynman function. [Note that the $i\epsilon$ prescription for G^F_0 is given in Ref. 5, Eq. (A16).]

C. Properties of the de Sitter-invariant states

We have obtained the form (2.15) of the symmetric two-point function $G^{(1)}_{\alpha,\beta}$ in the (α,β) state. However, it is invariant under the full disconnected de Sitter group O(1,4) only if $\beta\equiv 0$. To see this, note that $G^{(1)}_{\alpha,\beta}$ is the sum of three terms, proportional to $G^{(1)}_0(Z)$, $G^{(1)}_0(-Z)$, and $D_0(\bar{x},y)$. As discussed in Sec. II A, the first two terms are O(1,4) invariant. However, since $D_0(x,y)=-D_0(y,x)$, the commutator $D_0(x,y)$ cannot be a function of Z, since $Z(x,y)=Z(y,x)$. It then follows that $D(\bar{x},y)$ is also not a function only of Z, since $Z(\bar{x},y)=-Z(x,y)$. Thus, since $D_0(\bar{x},y)$ is not a function only of Z, it is not O(1,4) invariant. It is only when the coefficient $\sin\beta$ of this term vanishes (i.e., for $\beta=0$) that the state (α,β) is invariant under the disconnected de Sitter group O(1,4).

To understand this better, one must consider the properties of $G^{(1)}$, iD, and G^F under time reversal T. Since T and the antipodal transformation A (1.2) are both contained in the same disconnected component of O(1,4), we can equally well send both arguments to their antipodal points. Hence, in the "Euclidean" vacuum, $G^{(1)}_0$ is time-reversal invariant:

$$\begin{aligned}G^{(1)}_0(Tx,Ty)&=G^{(1)}_0(\bar{x},\bar{y})\\ &=\sum_n [\phi_n(\bar{x})\phi^*_n(\bar{y})+\phi^*_n(\bar{x})\phi_n(\bar{y})]\\ &=\sum_n [\phi^*_n(x)\phi_n(y)+\phi_n(x)\phi^*_n(y)]\\ &=G^{(1)}_0(x,y) \,,\end{aligned} \quad (2.20)$$

where we have used (2.11) and (2.12). Similarly, one can see from (2.11) and (2.13) that the commutator function changes sign

$$iD_0(Tx,Ty)=iD_0(\bar{x},\bar{y})=-iD_0(x,y)\ . \tag{2.21}$$

This, of course, means that $iD_0(x,y)$ is not a function of $Z(x,y)$ since $Z(Tx,Ty)=Z(x,y)$. In fact, the commutator function is of the form $\epsilon(x,y)\omega(Z)$. It can also be represented as a function of the *signed* distance function $\tilde{Z}$ defined in Sec. I. This function $iD_0(\tilde{Z})$ has a branch cut along the real $\tilde{Z}$ axis from $\tilde{Z}=1$ to ∞ and changes sign across the cut. Using (2.15), (2.20), and (2.21) one can see that

$$G^{(1)}_{\alpha,\beta}(Tx,Ty)=G^{(1)}_{\alpha,-\beta}(x,y) \tag{2.22}$$

and, in fact, the time reversal of the state (α,β) is the state $(\alpha,-\beta)$. Only the $\beta=0$ states are time-reversal invariant.

Starting from the "Euclidean" vacuum state, invariant under the full disconnected de Sitter group O(1,4), we have constructed a two-real-parameter (α,β) family of states invariant under the connected de Sitter group G. There is a one-real-parameter $(\alpha,0)$ family of time-symmetric states invariant under the disconnected group O(1,4). This is in agreement with the work of Chernikov and Tagirov.[6,7] There, the complex parameter λ labeling the vacuum states is related to our (α,β) by $\alpha(\lambda)=\cosh^{-1}(1-|\lambda|^2)^{-1/2}$ and $\beta(\lambda)=i\ln(-\lambda/|\lambda|)$, and the time reversal of the state $|\lambda\rangle$ is the state $|\lambda^*\rangle$.

Two special cases of these two-point functions have appeared in the literature. In the time symmetric case $\beta=0$, equations identical to (2.15) and (2.16) have been obtained by Schomblond and Spindel, in coordinates that cover half of de Sitter space [take $d=\cosh\alpha$ and $c=i\sinh\alpha$ in Eqs. (39) and (43) of Ref. 5]. The parameter α can be thought of as determining the strength of the singularity in $G^{(1)}(x,y)$ when $x=\bar{y}$. The ratio of the residues at $Z=-1$ and $Z=+1$ is $\tanh 2\alpha$, and for $\alpha=0$ one obtains the "Euclidean" two-point function which has no pole at $Z=-1$. Schomblond and Spindel rejected the vacua for which $\beta\neq 0$, because they claimed that $G^{(1)}$ was then a distribution and not a function of Z.

The second special case of these formulas which has appeared in the literature is in the work of Mottola[8] who has considered the cases $\beta=\pm\pi/2$. In that work the vacuum angle θ is related to the parameters α and λ by

$$\alpha=\theta,\quad \lambda=-i\tanh\theta\ . \tag{2.23}$$

For $m^2H^{-2}>\frac{9}{4}$, Mottola picks out the vacua

$$\begin{aligned}\alpha&=|\sinh^{-1}[\operatorname{csch}\pi(m^2H^{-2}-\tfrac{9}{4})^{1/2}]|\ ,\\ \beta&=\pm\pi/2\ ,\end{aligned} \tag{2.24}$$

and identifies them as in/out vacuum states. Because $\beta=\pm\pi/2$ these in/out states are time reverses of each other. However, formula (57) given in Ref. 8, for the Feynman function in this state, is incorrect because it is only a function of the variable Z and is therefore time symmetric.

An interesting property of the $\alpha\neq 0$ states is that they contain an infinite number of particles relative to one another, or relative to the "Euclidean" $\alpha=0$ vacuum. This is because in the Bogoliubov transformation (2.9) $\sum_n|\beta_n|^2=\infty$. Gibbons and Hawking[11] showed that this is a necessary property of any de Sitter-invariant state. If a state contains any quanta of a *given* momentum, then by de Sitter invariance (which includes boosts) it must contain an equal number of quanta of *every* momentum.

In the $\alpha\neq 0$ states, the two-point functions do not have the same short-distance behavior as the conventional flat-space two-point functions do. For example, the two-point function $G^{(1)}_{\alpha,\beta}$ given in (2.15) diverges $\cosh 2\alpha$ times as rapidly as it would in flat space. Mottola[8] argues that this is because the fields do not obey boundary conditions at ∞ which are analogous to the ordinary flat-space ones, because in flat space one imposes boundary conditions at spatial ∞, but in de Sitter space there is *no* spatial ∞. Whether or not $\alpha\neq 0$ defines physically acceptable states is unclear.

III. THE MASSLESS CASE HAS NO DE SITTER-INVARIANT FOCK VACUUM

The remainder of this paper is almost entirely devoted to the massless case. If $m^2=0$, the treatment in the previous section breaks down. In particular, if $m^2=0$ then the hypergeometric function (2.6) is a constant. This is because $m^2=0\to c=0$ and

$$f(Z)={}_2F_1(0,3;2;\tfrac{1}{2}(1+Z))=1\ . \tag{3.1}$$

Obviously a constant is a trivial solution to $\Box G(Z)=0$, but now $f(Z)$ and $f(-Z)$ are equal, and not linearly independent. Hence (2.7) is not the general solution to $\Box G=0$.

If $m^2=0$, the second fundamental real solution is

$$P(Z)=(1+Z)^{-1}-(1-Z)^{-1}+\ln\left|\frac{Z-1}{Z+1}\right|\ . \tag{3.2}$$

This solution has the important property that $P(-Z)=-P(Z)$. The absolute value sign ensures that $P(Z)$ is always real. It arises in the following way. The complex-analytic function $\ln[(Z-1)/(Z+1)]$ has branch cuts along the real axis from $Z=\pm 1$ to $\pm\infty$. Since $G^{(1)}(Z(x,y))$ is by definition symmetric in x and y, the $\pm i\epsilon$ convention of (1.8) corresponds to taking the average value of $\ln[(Z-1)/(Z+1)]$ across the cuts. This yields the absolute value sign in $P(Z)$. Hence when $m^2=0$ the general solution to $\Box G(Z)=0$ is

$$G^{(1)}(Z)=\alpha P(Z)+\beta\ , \tag{3.3}$$

where the choice of real constants α and β determines the particular solution. If $\alpha=H^2/4\pi^2$ then this has the same short-distance behavior as a massless two-point function has in flat space.

One can already see hints that the massless case is going to be fundamentally different than the massive one. Since the function $P(Z)$ has two singular points at $Z=\pm 1$, our general de Sitter-invariant solution has either (1) no singular points or (2) two singular points. There is nothing corresponding to a de Sitter-invariant "Euclidean" vacuum state with only one singular point. Because $P(-Z) = -P(Z)$, any de Sitter-invariant massless two-point function (3.3) has the property that

$$G^{(1)}(Z)+G^{(1)}(-Z)=2\beta=\text{real constant} . \quad (3.4)$$

We will use this property to prove that in the massless case there is no de Sitter-invariant Fock vacuum state.

At this point we assume familiarity with the standard method of canonical quantization.[12] In Appendix B we prove the following theorem.

Theorem

(1) In de Sitter space, define the inner product of two scalar functions $\phi_1(x)$ and $\phi_2(x)$ as

$$(\phi_1,\phi_2)=i\int_\Sigma (\phi_1^*\nabla_\mu\phi_2-\phi_2\nabla_\mu\phi_1^*)d\Sigma^\mu \quad (3.5)$$

where Σ is any Cauchy surface (hence having topology S^3).

(2) Let $\phi_n(x)$ be a set of complex scalar functions satisfying (a) $(\Box-m^2)\phi_n(x)=0$ for a given real m^2, (b) $(\phi_n,\phi_m^*)=0$ and $(\phi_n,\phi_m)=\delta_{nm}$, and (c) $\phi_n(x)$ and $\phi_n^*(x)$ are a basis for the space of smooth functions $f(x)$ satisfying $(\Box-m^2)f(x)=0$.

(3) Define

$$G^{(1)}(x,y)=\sum_n [\phi_n(x)\phi_n^*(y)+\phi_n^*(x)\phi_n(y)] . \quad (3.6)$$

Then the following is true.

(A) $G^{(1)}(x,y)+G^{(1)}(x,\bar{y})\neq 0$ everywhere (note $\bar{y}$ antipodal point of y).

(B) $G^{(1)}(x,y)-G^{(1)}(x,\bar{y})\neq 0$ everywhere.

(C) If $m^2=0$ and C is a real constant, then $G^{(1)}(x,y)+G^{(1)}(x,\bar{y})\neq C$ everywhere.

The significance of this theorem is not hard to explain. A Fock vacuum state $|0\rangle$ is defined by a set of mode functions $\phi_n(x)$ which satisfy condition (2). If the field operator is

$$\Phi(x)=\sum_n [a_n\phi_n(x)+a_n^\dagger\phi_n^*(x)] \quad (3.7)$$

and the annihilation and creation operators a_n and $a_m^\dagger$ satisfy the usual algebra

$$[a_n,a_m]=0,\ \ [a_n,a_m^\dagger]=\delta_{nm} , \quad (3.8)$$

then the Fock vacuum is defined by $a_n|0\rangle\equiv 0$. In this vacuum state, the symmetric two-point function is simply a sum of the mode functions

$$\begin{aligned} G^{(1)}(x,y)&=\langle 0|\Phi(x)\Phi(y)+\Phi(y)\Phi(x)|0\rangle \\ &=\sum_n [\phi_n(x)\phi_n^*(y)+\phi_n^*(x)\phi_n(y)] . \end{aligned} \quad (3.9)$$

Since $Z(x,y)=-Z(x,\bar{y})$, theorems (A) and (B) show that there exists no de Sitter-invariant Fock vacuum (for any m^2) for which $G^{(1)}(Z)\pm G^{(1)}(-Z)\equiv 0$.

In the massless case, we showed in (3.4) that de Sitter invariance required the two-point function to obey $G^{(1)}(Z)+G^{(1)}(-Z)\equiv C$, where C is a real constant. Theorem (C) shows that there exists no Fock vacuum state in which this equation can hold. Hence, this establishes that there is no de Sitter-invariant Fock vacuum state for $m^2=0$.

It is often believed that "what goes wrong" when $m^2=0$ has something to do with the fact that the wave equation has a constant solution, which is often called a "zero mode." This is simply not true. Before concluding this section, we would like to show that this so-called "zero mode" is simply the real part of an otherwise undistinguished normalizable mode.

Let us work in global coordinates defined by the $(H=1)$ metric

$$ds^2=-dt^2+\cosh^2 t\, d\Omega^2 , \quad (3.10)$$

where $d\Omega^2$ is the metric on a unit three-sphere. Consider mode functions $\phi(t)$ which are only functions of time, and are constant on each three-sphere. For these modes, the equation $\Box\phi=0$ then becomes

$$\cosh^{-3}t\frac{d}{dt}\cosh^3 t\frac{d}{dt}\phi=0 \quad (3.11)$$

or simply $\cosh^3 t\dot{\phi}=\text{const}$. The two linearly independent solutions are

$$\begin{aligned} &e(t)=1 , \\ &f(t)=\int_0^t \frac{dt}{\cosh^3 t}=\frac{1}{2}\left[\frac{\sinh t}{\cosh^2 t}+\arctan(\sinh t)\right] . \end{aligned} \quad (3.12)$$

Note that $e(-t)=e(t)$ and $f(-t)=-f(t)$. Because the modes are constant on each S^3 their inner products are easily evaluated:

$$(e,e)=(f,f)=0,\ \ (e,f)=(f,e)^*=iV_3 , \quad (3.13)$$

where $V_3=2\pi^2$ is the volume of a unit three-sphere. From these two solutions one can form the true "zero mode" which is actually a perfectly ordinary *unit norm* complex mode:

$$\phi_0(t)=(2V_3)^{-1/2}[e(t)-if(t)] . \quad (3.14)$$

Note that $(\phi_0,\phi_0)=1$, $(\phi_0,\phi_0^*)=0$, and $(\phi_0^*,\phi_0^*)=-1$ as desired. Furthermore, this mode is orthogonal to all the higher modes, which are proportional to the spherical harmonic functions $Y_{klm}(\Omega)$ for $(k,l,m)\neq(0,0,0)$.

Why does no de Sitter-invariant Fock vacuum exist for a massless scalar field? Hawking and Moss[13] showed that if one perturbs the Euclidean vacuum for $m^2>0$ then the perturbations decay exponentially in time and the properties of the state quickly approach those of the Euclidean vacuum. This does not appear to happen if $m^2=0$. In the massless case a small spatially homogeneous perturbation grows linearly at first and then approaches a constant nonzero value. This is because the behavior of such a perturbation is governed by the mode function $f(t)$, and $f(\infty)=\pi/4$. We believe that in the quantum theory, vacuum fluctuations drive this type of undamped perturbation, and cause spontaneous breaking of de Sitter invariance. In the next section we will explore one of the ways in which such symmetry breaking can take place.

IV. VACUUM STATES THAT BREAK DE SITTER INVARIANCE

In this section, we examine some vacuum states that are *not* invariant under the de Sitter group. These states are analogous to broken-symmetry vacuum states in Yang-Mills gauge theories.[14] Before going further, it is worthwhile to develop this analogy a bit further.

In Yang-Mills gauge theories, there exist vacuum states which are not invariant under the full symmetry group of the gauge theory. Under certain conditions (on the coupling constants and temperature) one can show that a particular broken symmetry state is the true physical vacuum, i.e., the lowest-energy state. In this situation the other vacuum states are known as "false" or "metastable" vacua. Usually the interesting vacua are those which are invariant under the action of some subgroup of the full symmetry group. Often, these subgroups are the *maximal* subgroups of the gauge group.[15]

The de Sitter group O(1,4) is ten dimensional, and its maximal subgroups are six dimensional. These maximal subgroups are O(4), O(1,3), and E(3) (the group of rigid motions of flat Euclidean R^3). The first is compact and the other two are noncompact. The three subgroups correspond to transformations of de Sitter space which leave invariant three different families of hypersurfaces. Those three families of hypersurfaces can be obtained by foliating de Sitter space with maximally symmetric spatial surfaces. These are the standard foliations with closed ($k=1$) or open ($k=-1$) or flat ($k=0$) spatial sections.[3]

For example, in the metric $ds^2=t^{-2}(-dt^2+d\mathbf{x}^2)$ the E(3) acts entirely on the spatial $\mathbf{x}$ sending it to $\mathbf{x}'=R\mathbf{x}+\mathbf{A}$ where R is an SO(3) rotation matrix and $\mathbf{A}$ is a constant three-vector. The matrix R and vector $\mathbf{A}$ are each specified by three parameters, and thus E(3) is a six-dimensional group. Similarly, in the metric $ds^2=-dt^2+\cosh^2 t\, d\Omega^2$, where $d\Omega^2$ is the metric on a unit three-sphere, the SO(4) subgroup acts entirely on Ω.

We do not know a systematic way to find Fock vacua which are invariant under these subgroups. However, for E(3) there is a simple way to obtain them. Schomblond[5] has constructed the de Sitter-invariant states for $m^2>0$ using spatially flat coordinates. We follow that treatment precisely, with one exception. Schlomblond evaluates the mode sums using a formula valid only if $m^2>0$. We do the $m^2=0$ case. The vacuum states constructed in this way are *not* de Sitter invariant (as they would be for $m^2>0$) but merely E(3) invariant. It seems possible that an analogous careful treatment of $m^2=0$ quantization with the other two choices of spatial coordinates would yield O(4)- and O(1,3)-invariant vacuum states.

A nice feature of the massless case is that the mode functions can always be expressed in terms of exponentials and polynomials, eliminating the need for special functions. If the mode functions $\phi_{\mathbf{k}}(t,\mathbf{x})$ are written in the coordinates of (1.10) as

$$\phi_{\mathbf{k}}(t,\mathbf{x})=V_{\mathbf{k}}(t)\frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{(2\pi)^{3/2}}\,, \tag{4.1}$$

then the wave equation $\Box\phi=0$ separates. The time functions $V_{\mathbf{k}}(t)$ must satisfy the equation

$$\left[\frac{d^2}{dt^2}-\frac{2}{t}\frac{d}{dt}+k^2\right]V_{\mathbf{k}}(t)=0\,, \tag{4.2}$$

where $k=(\mathbf{k}\cdot\mathbf{k})^{1/2}$ with a flat three-dimensional inner product. The index $\mathbf{k}$ is a coordinate wave number, and not a physical momentum. The general solution to (4.2) is

$$V_{\mathbf{k}}(t)=t^{3/2}[a(\mathbf{k})H^{(1)}(kt)+b(\mathbf{k})H^{(2)}(kt)]\,, \tag{4.3}$$

where $H^{(1)}$ and $H^{(2)}$ are the functions

$$H^{(1)}(z)=(z^{-1/2}+iz^{-3/2})e^{i(z-\pi/4)}\,,$$
$$H^{(2)}(z)=(z^{-1/2}-iz^{-3/2})e^{-i(z-\pi/4)}\,. \tag{4.4}$$

These functions are the Hankel functions given by Schomblond in Eq. (2.9) of Ref. 5 when $\nu=(\frac{9}{4}-m^2/H^2)^{1/2}=\frac{3}{2}$. For real arguments x, these functions form a complex-conjugate pair: $H^{(1)}(x)=[H^{(2)}(x)]^*$.

The modes defined by $H^{(1)}$ and $H^{(2)}$ are negative and positive frequency, respectively, with respect to the Klein-Gordon inner product

$$(\phi_1,\phi_2)=i\int_\Sigma(\phi_1^*\nabla_u\phi_2-\phi_2\nabla_u\phi_1^*)d\Sigma^u\,, \tag{4.5}$$

which is independent of the choice of spacelike Cauchy surface Σ. Here Σ is a $t=$const flat spatial surface with topology R^3. It is *not* a Cauchy surface for the complete de Sitter space, but only for the half-space covered by one of the two coordinate patches shown in Fig. 4. The norm of the mode function defined by (4.1) and (4.3) is

$$(\phi_{\mathbf{k}},\phi_{\mathbf{k}'})=2H^{-2}(|b|^2-|a|^2)\delta_{\mathbf{k}\mathbf{k}'}\,. \tag{4.6}$$

To carry out canonical quantization, we express the field operator Φ in terms of the mode functions $\phi_{\mathbf{k}}$ and creation and annihilation operators $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^\dagger$:

$$\Phi(x)=\sum_{\mathbf{k}}[\phi_{\mathbf{k}}(t,\mathbf{x})a_{\mathbf{k}}+\phi_{\mathbf{k}}^*(t,\mathbf{x})a_{\mathbf{k}}^\dagger]\,. \tag{4.7}$$

The field operator Φ is taken to be self-adjoint, so that it represents a real classical field. The creation and annihilation operators obey canonical commutation relations

$$[a_{\mathbf{k}},a_{\mathbf{k}'}]=0,\quad [a_{\mathbf{k}}^\dagger,a_{\mathbf{k}'}^\dagger]=0,\quad [a_{\mathbf{k}},a_{\mathbf{k}'}^\dagger]=\delta_{kk'} \tag{4.8}$$

if the modes $\phi_{\mathbf{k}}$ are chosen to have norm 1.

For each choice of functions a and b satisfying

$$|b(\mathbf{k})|^2-|a(\mathbf{k})|^2=\tfrac{1}{2}H^2\,,$$

there exists a "vacuum " state $|a,b\rangle$ which is defined by the property that it is annihilated by all of the operators $a_{\mathbf{k}}$:

$$a_{\mathbf{k}}|a,b\rangle=0 \quad \text{for all } \mathbf{k}\,. \tag{4.9}$$

When $m^2>0$, the de Sitter-invariant vacua are obtained by taking $a(\mathbf{k})$ and $b(\mathbf{k})$ to be constants, independent of $\mathbf{k}$.[5] This is because any dependence upon $\mathbf{k}$ would define a preferred rest frame, and hence a preferred state of motion. Here with $m^2=0$ we know it will not be possible to find a de Sitter-invariant vacuum state. However, taking $a(\mathbf{k})$ and $b(\mathbf{k})$ to be constant will turn out to define E(3)-invariant vacua.

Let us now specialize to the case where $a(\mathbf{k})=a$ and $b(\mathbf{k})=b$ are constant. In the vacuum state defined by this choice, the symmetric two-point function is

$$G^{(1)}(x,y)=\langle a,b\mid \Phi(x)\Phi(y)+\Phi(y)\Phi(x)\mid a,b\rangle \tag{4.10}$$

and can be evaluated from (4.7) and (4.8). It is

$$\begin{aligned}G^{(1)}(x,y)=&(\mid a\mid^2+\mid b\mid^2)P_1(x,y)+2\,\mathrm{Re}(ab^*)P_3(x,y)\\&+2i\,\mathrm{Im}(ab^*)P_4(x,y)\ ,\end{aligned} \tag{4.11}$$

where the P_i are defined as three-dimensional Fourier transforms ("mode sums")

$$P_i(t,\mathbf{x},t',\mathbf{x}')=(tt')^{3/2}\int \frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}Q_i(kt,kt')\ . \tag{4.12}$$

The $Q_i(\tau,\tau')$ are products of mode functions.

$$\begin{aligned}Q_1(\tau,\tau')&=H^{(1)}(\tau)H^{(2)}(\tau')+H^{(2)}(\tau)H^{(1)}(\tau')\ ,\\Q_2(\tau,\tau')&=H^{(1)}(\tau)H^{(2)}(\tau')-H^{(2)}(\tau)H^{(1)}(\tau')\ ,\\Q_3(\tau,\tau')&=H^{(1)}(\tau)H^{(1)}(\tau')+H^{(2)}(\tau)H^{(2)}(\tau')\ ,\\Q_4(\tau,\tau')&=H^{(1)}(\tau)H^{(1)}(\tau')-H^{(2)}(\tau)H^{(2)}(\tau')\ .\end{aligned} \tag{4.13}$$

In Appendix C, we show how to evaluate the integrals in (4.12). Because the integrands $Q_i(kt,kt')$ only depend upon the length k of the vector $\mathbf{k}$, their Fourier transforms P_i can only depend upon $\mathbf{r}=\mathbf{x}-\mathbf{x}'$ via the length $r=(\mathbf{r}\cdot\mathbf{r})^{1/2}$.

In terms of the variable Z given by (1.11) the P_i are

$$\begin{aligned}P_1&=(2\pi^2)^{-1}[(1-Z)^{-1}-\ln(1-Z)-\ln(2tt')+u(r)]\ ,\\P_4&=i(2\pi^2)^{-1}[(1+Z)^{-1}-\ln(-1-Z)-\ln(2tt')+u(r)]\ ,\\P_2&=i(2\pi)^{-1}[2rtt'(t-t')^{-1}\delta(r^2-(t-t')^2)\\&\qquad+\epsilon(t-t')\theta((t-t')^2-r^2)]\ ,\\P_3&=(2\pi)^{-1}[2rtt'(t+t')^{-1}\delta(r^2-(t+t')^2)\\&\qquad-\epsilon(t+t')\theta((t+t')^2-r^2)]\ ,\end{aligned} \tag{4.14}$$

where $u(r)$ is an unspecified (regularization) function which appears because of an infrared ($\mathbf{k}=0$) divergence in the integrals (4.12). Since the mode sums are constructed to be solutions of the wave equation, the functions must satisfy $\Box P_i=0$. In Appendix C, we show that this restricts $u(r)$ to be a function of the form

$$u(r)=\frac{c_1}{r}+c_2\ , \tag{4.15}$$

where c_1 and c_2 are arbitrary constants. We will see later that $u(r)$ cancels out of physical gauge-invariant quantities.

Now we can do an easy and interesting check of our calculation. Suppose that we want the two-point function (4.11) to equal the de Sitter-invariant function (3.3). Using (4.14) and demanding equality leads to three conditions:

$$\begin{aligned}&\mathrm{Re}(ab^*)=0\ ,\\&\pi^{-2}\,\mathrm{Im}(ab^*)=\frac{H^2}{4\pi^2}\ ,\\&(2\pi^2)^{-1}(\mid a\mid^2+\mid b\mid^2)=\frac{H^2}{4\pi^2}\ .\end{aligned} \tag{4.16}$$

The only solution (up to a trivial overall phase) is

$$a=\tfrac{1}{2}H,\quad b=-\frac{i}{2}H\ . \tag{4.17}$$

Consequently a de Sitter-invariant two-point function is only obtained when a and b satisfy $\mid a\mid^2-\mid b\mid^2\equiv 0$. This means that the mode functions $\phi_{\mathbf{k}}$ cannot be chosen to be positive norm. They must have *zero* norm, since (4.6) vanishes. But this is *precisely* what the proof of theorem (A) (Sec. III) told us would go wrong. In that proof, we showed that de Sitter invariance implied that at least one of the mode functions had zero norm. Now we see that all of them do.

Because these mode functions have zero norm, the $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^\dagger$ in (4.7) cannot satisfy canonical commutation relations (4.8). For example, the commutator function

$$\begin{aligned}D(x,x')&=-i\langle \Phi(x)\Phi(x')-\Phi(x')\Phi(x)\rangle\\&=i(\mid b\mid^2-\mid a\mid^2)P_2(t,t',r)\end{aligned} \tag{4.18}$$

vanishes everywhere, since $\mid a\mid^2=\mid b\mid^2$. Note that when a and b take on these unphysical "de Sitter-invariant" values, the function $u(r)$ cancels out of expression (4.11) for $G^{(1)}$.

In the massive case, one obtains a one-parameter family of de Sitter-invariant states by taking[5]

$$\begin{aligned}a(k)&=-i(2)^{-1/2}H\sinh\alpha\ ,\\b(k)&=(2)^{-1/2}H\cosh\alpha\end{aligned} \tag{4.19}$$

where α is some real number. When $m^2=0$ this no longer defines a de Sitter-invariant state. However, it does define an E(3)-invariant state. To see this, consider $G^{(1)}$ in the vacuum defined by (4.19),

$$G^{(1)}_\alpha(t,\mathbf{x};t',\mathbf{x}')=\frac{H^2}{4\pi^2}\{\cosh2\alpha[(1-Z)^{-1}-\ln(1-Z)-\ln(2tt')]+\sinh2\alpha[(1+Z)^{-1}-\ln(-1-Z)-\ln(2tt')]+\mathrm{const}\}\ . \tag{4.20}$$

In this formula, $Z(t,\mathbf{x};t',\mathbf{x}')$ is defined by (1.11) and is completely de Sitter invariant. It only depends upon the geodesic distance between x and x'. However, $G^{(1)}$ is not just a function of Z; therefore, the state defined by (4.19) cannot be de Sitter invariant.

$G^{(1)}$ *is* invariant under those de Sitter transformations which leave t and t' unchanged. This is the subgroup of the de Sitter group which maps a spacelike surface defined by $t=\text{const}$ into itself. As discussed earlier, this is the subgroup E(3). Hence, the vacuum defined by (4.19) when $m^2=0$ is E(3) invariant.

When $m^2>0$, the Euclidean vacuum is defined by $\alpha=0$. In the literature this vacuum state is also called the Bunch-Davies[16,17] or Birrell-Davies[12,18] vacuum. When $m^2=0$ the Euclidean vacuum state no longer exists. This is because (1) this state is de Sitter invariant and (2) its two-point function has only one singular point on S^4. However, the Bunch-Davies vacuum, defined by $\alpha=0$ in (4.19) and (4.20) does exist for $m^2=0$. It is simply no longer de Sitter invariant.

Because the two-point functions behave badly at the surface $t=\pm\infty$, we believe that these vacua are only defined on the half-space covered by one spatially flat coordinate patch (see Fig. 4). In these states the expectation value of the square of the scalar field can be determined from the coincidence limit of $G^{(1)}(x,x')$. The de Sitter-invariant part of $G^{(1)}$ can only affect the constant part of $\langle\Phi^2(x)\rangle$, since the only de Sitter-invariant scalar function is a constant. The spacetime dependent part of $\langle\Phi^2(x)\rangle$ arises entirely from the $\ln(2tt')$ terms in (4.20). In co-moving coordinates $ds^2=-d\tau^2+e^{2H\tau}d\mathbf{x}^2$, where $\tau=-H\ln Ht$, the expectation value of Φ^2 is

$$\langle\Phi^2(x)\rangle=\frac{H^3}{4\pi^2}e^{2\alpha}(\tau-\tau_0)\ , \tag{4.21}$$

where all the constants have been absorbed into the "turn on" time τ_0. For $\alpha=0$ this result is in agreement with others[13,17,19,20] who have studied this effect.

The massless Lagrangian $(\nabla_\mu\Phi)(\nabla^\mu\Phi)$ is invariant under global gauge transformations $\Phi\to\Phi+\text{const}$. Consequently quantities like $\langle\Phi(x)\Phi(y)+\Phi(y)\Phi(x)\rangle$ and $\langle\Phi^2\rangle$ are devoid of physical significance since they are not gauge invariant. However, one can find many gauge-invariant objects, including the vacuum expectation values

$$\langle\Phi^2(x)\rangle-\langle\Phi^2(y)\rangle,\quad \langle\Phi(x)\Phi(y)-\Phi(y)\Phi(x)\rangle$$

$$\langle[\Phi(x_1)-\Phi(y_1)]\cdots[\Phi(x_i)-\Phi(y_i)]\rangle\ .$$

They are all free of infrared divergences as $|\mathbf{k}|\to 0$, and (remarkably) objects of the third type are easily shown to be *fully* de Sitter invariant under simultaneous transformation of all the x_i and y_i. However, the gauge-invariant quantity

$$\langle\Phi^2(x)\rangle-\langle\Phi^2(y)\rangle\propto(\tau_x-\tau_y)$$

still breaks de Sitter invariance, since de Sitter invariance would imply

$$\langle\Phi^2(x)\rangle-\langle\Phi^2(y)\rangle=0\ .$$

Ford[21] has explained this unusual behavior as due to the presence of an infrared divergence in the theory. One may also understand it in the following way. The massless field has no de Sitter-invariant vacuum state, and we are looking at a state with E(3) invariance. The property of this state that breaks de Sitter invariance is that $\langle\Phi^2\rangle$ is changing in time. The surfaces $\langle\Phi^2\rangle=\text{const}$ are flat spacelike surfaces which are E(3) invariant. We expect that the O(4)- and O(1,3)-invariant vacuum states will break de Sitter invariance in a similar way. In those vacua the surfaces $\langle\Phi^2\rangle=\text{const}$ are probably the $k=\pm 1$ spacelike surfaces of constant positive or negative curvature.

Unfortunately, the E(3)-invariant vacuum states we have defined are not unique. They only happen to be in some sense "as close as possible" to the de Sitter-invariant states. Another way to define E(3)-invariant vacua would be to imagine that de Sitter space were smoothly joined to a static flat spacetime along a slice $t=T$. For the modes and their derivatives to be continuous across the boundary they would have to behave like e^{ikt} at the boundary. Hence

$$\left.\frac{\dot V_{\mathbf{k}}(t)}{V_{\mathbf{k}}(t)}\right|_{t=T}=-ik \tag{4.22}$$

which determines a and b to be

$$\begin{aligned} a(\mathbf{k})&=-\tfrac{1}{4}(2)^{1/2}\frac{H}{kT}e^{-2ikT}\ ,\\ b(\mathbf{k})&=\tfrac{1}{2}(2)^{1/2}H\left[1+\frac{i}{2}k^{-1}T^{-1}\right]\ . \end{aligned} \tag{4.23}$$

This condition would ensure that a freely falling detector would not respond near $t=T$. The Bunch-Davies vacuum state ($\alpha=0$) is obtained when $T=-\infty$.

Another possibility has been studied by Ottewill and Najmi[22] following a prescription given by Ashtekhar and Magnon.[23] They define the vacuum as that state which minimizes the energy on a specified spacelike surface. If we choose a flat spacelike surface at $t=T$, their condition[22]

$$\left.\frac{\dot V_{\mathbf{k}}(t)}{V_{\mathbf{k}}(t)}\right|_{t=T}=-ik-\frac{1}{T} \tag{4.24}$$

defines an E(3)-invariant vacuum state. This vacuum is equivalent to

$$\begin{aligned} a(k)&=-\tfrac{1}{2}(2)^{1/2}Hk^{-1}T^{-1}\left[1-\frac{i}{2}k^{-1}T^{-1}\right]e^{-2ikT}\ ,\\ b(k)&=\tfrac{1}{2}(2)^{1/2}H(1+\tfrac{1}{2}k^{-2}T^{-2})\ . \end{aligned} \tag{4.25}$$

Note that when $T=-\infty$ this is the Bunch-Davies vacuum ($\alpha=0$). Thus, the (massless) Bunch-Davies vacuum is obtained from the lowest-energy initial state at $T=-\infty$. The expressions for the two-point functions in the "minimum energy" vacuum can be found in Ref. 22.

If the early universe was pure radiation $k=0$ Robertson-Walker for $t<T$ and de Sitter for $t>T$ then one can show that the minimum energy vacuum state at $t=T$ is obtained. During the radiation phase, the natural vacuum state for a massless field is the conformal vacuum, in which the mode functions are (r=radiation and f=flat)

$$\phi_k^r(x)=\Omega^{-1}(t)\phi_k^f(x)\ , \tag{4.26}$$

where $\Omega=t$ is the conformal factor. Since the flat-space mode functions have time dependence e^{-ikt}, for $t<T$,

$$\frac{\dot{\phi}_k^r}{\phi_k^r}=\frac{\dot{\phi}_k^f}{\phi_k^f}-\frac{\dot{\Omega}}{\Omega}=-ik-\frac{1}{t}\ . \qquad (4.27)$$

In order to be continuous and differentiable across the join at $t=T$, the mode functions in the de Sitter phase must satisfy (4.27) at $t=T$. This is identical to the minimum energy vacuum defined by (4.24).

V. CONCLUSION

We have shown that when $m^2>0$, a one-complex-parameter family of de Sitter-invariant states can be obtained from the Euclidean vacuum by a frequency-independent Bogoliubov transformation. If $\phi_n(x)$ are the modes that define the Euclidean Fock vacuum $|0\rangle$ then the modes that define the $|\alpha,\beta\rangle$ state are

$$\tilde{\phi}_n(x)=\cosh\alpha\phi_n(x)+\sinh\alpha e^{i\beta}\phi_n^*(x)\ . \qquad (5.1)$$

In this state the two-point functions take the following simple forms (the subscript 0 denotes the Euclidean vacuum):

$$\begin{aligned}G^{(1)}_{\alpha,\beta}(x,y)&=\cosh2\alpha G^{(1)}_0(x,y)\\&\quad+\sinh2\alpha[\cos\beta G^{(1)}_0(\bar{x},y)-\sin\beta D_0(\bar{x},y)]\ ,\\ iD_{\alpha,\beta}(x,y)&=iD_0(x,y)\ ,\\ iG^F_{\alpha,\beta}(x,y)&=iG^F_0(x,y)+\tfrac{1}{2}[G^{(1)}_{\alpha,\beta}(x,y)-G^{(1)}_0(x,y)]\ ,\end{aligned} \qquad (5.2)$$

where $G^{(1)}$, iD, and iG^F are the symmetric, commutator and Feynman functions, and $\bar{x}$ is the antipodal point of x. The states are time-reversal invariant if and only if $\beta=0$.

The $|\lambda\rangle$ "vacua" contain an infinite number of quanta of the $|\lambda'\neq\lambda\rangle$ "vacua." Furthermore, for $\alpha\neq0$ the functions $G^{(1)}_{\alpha,\beta}$ and $G^F_{\alpha,\beta}$ have two loci of singular points, one if x is on the light cone of y, and the second if $\bar{x}$ is on the light cone of y. In the Euclidean vacuum ($\alpha=0$) the latter singularity is absent. This state is also the unique one in which the singularities as $x\to y$ are of Hadamard form. In flat space, carrying out a frequency-independent Bogoliubov transformation (5.1) on the vacuum state does not define physically acceptable vacuum states. However, Mottola[8] has suggested that for $4m^2>9H^2$ the vacuum states defined by

$$\alpha=|\sinh^{-1}[\mathrm{csch}\pi(m^2H^{-2}-\tfrac{9}{4})^{1/2}]|\ ,\quad \beta=\pm\frac{\pi}{2}$$

are analogous to the Unruh[24] in/out vacuum states of black holes. We do not know if this interpretation is correct.

In the massless (minimally coupled) case $m^2=0$ we proved rigorously that no de Sitter-invariant Fock vacuum state exists. This may be because if $m^2=0$ a small perturbation to ϕ does not damp away, but approaches a constant value at late times. Either (1) the standard Fock space construction must be abandoned or (2) the vacuum state must break de Sitter invariance. In considering the second possibility we looked for vacua invariant under a maximal subgroup of the de Sitter group. The maximal subgroups of SO(1,4) are E(3), SO(4), and SO(1,3).

We constructed several E(3)-invariant vacua, and showed that they broke de Sitter invariance because $\langle\Phi^2\rangle$ is constant on flat spacelike surfaces but grows in time. The Bunch-Davies vacuum, which is the de Sitter-invariant Euclidean vacuum for $m^2>0$, breaks de Sitter invariance when $m^2=0$ and defines an E(3)-invariant vacuum state. We believe that the O(4)- and O(1,3)-invariant vacuum states will be characterized by a time-dependent $\langle\Phi^2\rangle$ which is constant on spacelike surfaces of constant positive or negative curvature.

In de Sitter space, the quantum theory of linearized gravity (in conformal gauge) is exactly the same as that of two minimally coupled massless scalar fields.[21,26] However, our results do *not* imply that there is no de Sitter-invariant vacuum for gravity, because the choice of conformal gauge *explicitly* breaks de Sitter invariance. In fact, the Euclidean vacuum state for gravity is well defined and de Sitter invariant. This is because the wave operator on a four-sphere does not have any zero modes for massless fields with spin greater than zero.

ACKNOWLEDGMENTS

I would like to thank L. Ford, A. Guth, J. Hartle, W. Hiscock, G. Horowitz, T. Jacobson, and all of the others who have helped me with this paper. I would especially like to thank E. Mottola for pointing out that the previous version of this paper contained the unnecessary assumption that $\beta=0$, and hence excluded all possible time-reversal-noninvariant states. This work was supported in part by National Science Foundation Grant No. PHY 81-07384.

APPENDIX A

In this appendix, we show that the Euclidean vacuum state can be defined by choosing a complete set of orthonormal positive-norm modes $\phi_n(x)$ which obey $\phi_n(\bar{x})=\phi_n^*(x)$. Define global coordinates by the metric (with $H=1$)

$$ds^2=-dt^2+\cosh^2t\,d\Omega^2\ , \qquad (A1)$$

where $d\Omega^2$ is the metric on a unit three-sphere. The antipodal transformation $x\to\bar{x}$ sends $(t,\Omega)\to(-t,\bar{\Omega})$ where $\bar{\Omega}$ is the point on S^3 antipodal to Ω.

The conventional set of modes used to define the Euclidean vacuum[7,8] have the form

$$\psi_{klm}(x)=y_k(t)Y_{klm}(\Omega)\ . \qquad (A2)$$

The integers k,l,m label the modes on S^3, and the Y_{klm} are a complete set of scalar spherical harmonics on S^3. The range of m is $-|l|\le m\le|l|$. Under the transformation $x\to\bar{x}$

$$\begin{aligned}y_k(-t)&=y_k^*(t)\ ,\\ Y_{klm}(\bar{\Omega})&=(-1)^kY^*_{kl-m}(\Omega)\ ,\end{aligned} \qquad (A3)$$

and so the conventional Euclidean modes transform as

$$\psi_{klm}(\bar{x})=(-1)^k\psi^*_{kl-m}(x)\ . \qquad (A4)$$

Now define new modes by the trivial Bogoliubov transformation

$$\phi_{klm}(x)=(2)^{-1/2}e^{i(\pi/2)k}[e^{i\pi/4}\psi_{klm}(x)+e^{-i\pi/4}\psi_{kl-m}(x)]\ . \qquad (A5)$$

Then one may easily show that this is a complete set of orthonormal positive-norm modes. In particular,

(a) $\phi_{klm}(\bar{x})=\phi^*_{klm}(x)$,
(b) $(\phi_{klm},\phi_{k'l'm'})=\delta_{kk'}\delta_{ll'}\delta_{mm'}$ and $(\phi_{klm},\phi^*_{k'l'm'})=0$,
(c) The set of ϕ_{klm} is complete and spans the space of ψ_{klm},
(d) $\phi_{kl0}=e^{i(\pi/2)k}\psi_{kl0}$.

Property (a) follows from the definition (A5) and property (A4). Property (b) follows from (A5) and the fact that the ψ's are an orthonormal set:

$$(\psi_{klm},\psi_{k'l'm'})=\delta_{kk'}\delta_{ll'}\delta_{mm'}$$

and

$$(\psi_{klm},\psi^*_{k'l'm'})=0 .$$

Property (c) follows from

$$\psi_{klm}(x)=(2)^{-1/2}e^{-i(\pi/2)k}[e^{-i\pi/4}\phi_{klm}(x) + e^{i\pi/4}\phi_{kl-m}(x)] , \quad (A6)$$

since the ψ_{klm}'s are a complete set. The point is that there are "as many" ϕ's as ψ's because, for a given value of m, ϕ_{klm} and ϕ_{kl-m} are linearly independent. Property (d) follows from the definition (A5).

The general form of a Bogoliubov transformation is

$$\phi_n=\sum_m(\alpha_{nm}\psi_m+\beta_{nm}\psi^*_m) \quad (A7)$$

and it is called nonmixing or trivial if all the β_{nm} are zero. It is apparent from (A5) that our transformation is of this trivial type. There are two nonvanishing α's which are $\alpha_{klm,klm}$ and $\alpha_{klm,kl-m}$. The trivial Bogoliubov transformations define equivalent vacuum states;[12] hence, the Euclidean vacuum can be defined by a set of modes with $\phi_n(\bar{x})=\phi^*_n(x)$.

APPENDIX B

The following appendix contains a proof of theorems (A), (B), and (C), whose assumptions and conclusions are stated in Sec. III. We refer to the assumptions as (1), (2b), etc.

Proof by contradiction

(A) Assume

$$G^{(1)}(x,y)+G^{(1)}(x,\bar{y})=0$$

everywhere. Then from definition (3.6) this means

$$0=\sum_n\phi_n(x)[\phi^*_n(y)+\phi^*_n(\bar{y})]+\phi^*_n(x)[\phi_n(y)+\phi_n(\bar{y})] . \quad (B1)$$

Choose a Cauchy surface Σ and fix the point y. Regard (B1) as a scalar function of x, and form its inner product with any mode $\phi_m(x)$. Since the ϕ's form an orthonormal set (2b) this means

$$0=(\phi_m,G^{(1)}(y)+G^{(1)}(\bar{y}))=\phi^*_m(y)+\phi^*_m(\bar{y}) . \quad (B2)$$

Hence $\phi_m(x)+\phi_m(\bar{x})=0$ for each mode. Now choose a Cauchy surface Σ with the property that $x\in\Sigma\rightarrow\bar{x}\in\Sigma$. One can obtain such a Cauchy surface by intersecting de Sitter space with a spacelike four-plane that passes through the origin of the R^5 embedding space of Sec. I. Now let $(\partial/\partial t)^a$ be a future-pointing unit timelike vector orthogonal to Σ. Define

$$\dot{\phi}_m(x)=\frac{\partial}{\partial t}\phi_m(x) . \quad (B3)$$

The relation $\phi_m(x)=-\phi_m(\bar{x})$ then implies that

$$\dot{\phi}_m(x)=\dot{\phi}_m(\bar{x}) . \quad (B4)$$

The inner product of ϕ_m with itself is (1)

$$(\phi_m,\phi_m)=i\int_{S^3}(\phi^*_m\dot{\phi}_m-\phi_m\dot{\phi}^*_m)dV , \quad (B5)$$

where dV is the volume element on the three-sphere $\Sigma=S^3$. Consider the first integral in (B5). Since

$$\phi^*_m(x)\dot{\phi}_m(x)=-\phi^*_m(\bar{x})\dot{\phi}_m(\bar{x}) \quad (B6)$$

the integrand has opposite values on opposite sides of the sphere. Hence its integral over the sphere vanishes. But this implies $(\phi_m,\phi_m)=0$ which contradicts assumption (2b) that $(\phi_m,\phi_m)=1$. Q.E.D.

(B) Now suppose $G^{(1)}(x,y)-G^{(1)}(x,\bar{y})=0$ everywhere. This implies that

$$\phi_m(x)=\phi_m(\bar{x}) , \quad \dot{\phi}_m(x)=-\dot{\phi}_m(\bar{x}) \quad (B7)$$

so (B6) still holds. Hence one obtains the same contradiction as in (A).

(C) This case requires a bit more work. Again, suppose $G(x,y)+G(x,\bar{y})=C$ everywhere. Since C is a real constant, it is a solution to $\Box C=0$, and can be expressed in terms of the modes as

$$C=\sum_n[c_n\phi_n(x)+c^*_n\phi^*_n(x)] , \quad (B8)$$

where $c_n=(\phi_n,C)$. Hence $G(x,y)+G(x,\bar{y})-C=0$ means that

$$0=\sum_n\{\phi_n(x)[\phi^*_n(y)+\phi^*_n(\bar{y})-c_n] +\phi^*_n(x)[\phi_n(y)+\phi_n(\bar{y})-c^*_n]\} \quad (B9)$$

and hence

$$\phi_n(y)+\phi_n(\bar{y})=c^*_n \quad (B10)$$

for each mode. Note that if just *one* of the $c_n=0$, we have a contradiction, because the argument in (A) then implies that $(\phi_n,\phi_n)=0$. We will now show that at least one c_n must vanish. Define a new set of "tilde" functions

$$\tilde{\phi}_n(x)=\phi_n(x)-\frac{c^*_n}{2} . \quad (B11)$$

It follows from (B10) that

$$\tilde{\phi}_n(x)=-\tilde{\phi}_n(\bar{x}) \quad (B12)$$

and hence that

$$\dot{\tilde{\phi}}_n(x)=\dot{\tilde{\phi}}_n(\bar{x}) . \quad (B13)$$

Then since

$$-\widetilde{\phi}_n^*(x)\dot{\widetilde{\phi}}_m(x)=\widetilde{\phi}_n^*(\bar{x})\dot{\widetilde{\phi}}_m(\bar{x}) , \tag{B14}$$

by the argument given in (A), $(\widetilde{\phi}_n,\widetilde{\phi}_m)=0$. Choose $n\neq m$. Then from the definition (B11) of ϕ_n, this implies that

$$\begin{aligned}0&=(\phi_n,\phi_m)-\tfrac{1}{2}(c_n^*,\phi_m)-\tfrac{1}{2}(\phi_n,c_m^*)+\tfrac{1}{4}(c_n^*,c_m^*) ,\\ 0&=-\frac{1}{2}\frac{c_n}{C}(C,\phi_m)-\frac{1}{2}\frac{c_m^*}{C}(\phi_n,C) ,\\ 0&=-\frac{c_nc_m^*}{C} \quad \text{for all } n\neq m .\end{aligned} \tag{B15}$$

Hence at least one of the c_n must vanish. Q.E.D.

APPENDIX C

In this appendix we show how to evaluate integrals of the form encountered in (4.12). The integrands depend only upon the length of $\mathbf{k}$ and not upon its direction. Consequently the Fourier transforms only depend upon the length of $\mathbf{r}=\mathbf{x}-\mathbf{x}'$:

$$\begin{aligned}I(\mathbf{r})&=\int e^{i\mathbf{k}\cdot\mathbf{r}}Q(k)d^3\mathbf{k}\\ &=\frac{4\pi}{r}\int_0^\infty Q(k)k\sin(kr)dk .\end{aligned} \tag{C1}$$

The integrands $Q(k)$ can all be written as linear combinations of the six functions

$$\begin{bmatrix}k^{-1}\\k^{-2}\\k^{-3}\end{bmatrix}\sin k\tau,\quad \begin{bmatrix}k^{-1}\\k^{-2}\\k^{-3}\end{bmatrix}\cos k\tau , \tag{C2}$$

where $\tau=(t\pm t')$. For example,

$$\begin{aligned}Q_1(kt,kt')=2(tt')^{-3/2}[&tt'k^{-1}\cos k(t-t')\\ &+(t-t')k^{-2}\sin k(t-t')\\ &+k^{-3}\cos k(t-t')] .\end{aligned} \tag{C3}$$

Consequently, we must evaluate six integrals, which are, for $n=0,1,2$,

$$\begin{aligned}C_n(r,\tau)&=\int_0^\infty \frac{\sin kr\cos k\tau}{k^n}dk ,\\ S_n(r,\tau)&=\int_0^\infty \frac{\sin kr\sin k\tau}{k^n}dk .\end{aligned} \tag{C4}$$

The functions S_1, S_2, and C_1 are all well defined because the integrands are nonsingular at $k=0$ and fall off at $k=\infty$. However, S_0, C_0, and C_2 are not well defined because S_0 and C_0 oscillate at infinity and C_2 blows up at $k=0$. As we shall see, it is only the last case that presents any challenge.

A simple scheme for defining the ambiguous cases presents itself if we remember that these integrals are defined from solutions to a differential equation (the wave equation). Notice that from (C4)

$$\begin{aligned}\frac{\partial}{\partial\tau}S_n(r,\tau)&=C_{n-1}(r,\tau) ,\\ -\frac{\partial}{\partial\tau}C_n(r,\tau)&=S_{n-1}(r,\tau) .\end{aligned} \tag{C5}$$

Consequently, two of the ambiguous cases, C_0 and S_0, can be found from C_1 and S_1, which are well defined. The only remaining case is C_2.

To define C_2 we will use the relation (C5) in the opposite direction: we begin with $-S_1(r,\tau)$ and integrate it with respect to τ. The "constant" of integration introduced in this procedure is an undetermined function $\psi(r)$. We will then see that $\psi(r)$ is restricted to being of a specified form.

First let us evaluate S_2, C_1, and S_0:

$$\begin{aligned}S_2(r,\tau)&=\int_0^\infty \frac{\sin rx\sin\tau x}{x^2}dx\\ &=\tfrac{1}{8}\int_{-\infty}^\infty x^{-2}(e^{i(r-\tau)x}+e^{i(\tau-r)x}\\ &\qquad -e^{i(r+\tau)x}-e^{i(-r-\tau)x})dx .\end{aligned} \tag{C6}$$

Choose a contour from $-\infty$ to ∞ that passes below $x=0$ in the complex plane. For each term, the contour must be closed in the upper- or lower-half plane, depending upon the sign of the imaginary part of the exponential. Evaluating the residues at $x=0$ and using Cauchy's theorem

$$\begin{aligned}S_2(r,\tau)&=\frac{\pi}{4}[(r-\tau)\theta(\tau-r)+(\tau-r)\theta(r-\tau)\\ &\qquad -(r+\tau)\theta(-r-\tau)+(r+\tau)\theta(r+\tau)]\\ &=\frac{\pi}{2}[(r+\tau)\theta(r+\tau)-(\tau-r)\theta(\tau-r)-r] .\end{aligned} \tag{C7}$$

It then follows from (C5) that

$$\begin{aligned}C_1(r,\tau)&=\frac{\pi}{2}[\theta(r+\tau)-\theta(\tau-r)] ,\\ S_0(r,\tau)&=\frac{\pi}{2}[\delta(\tau-r)-\delta(\tau+r)] .\end{aligned} \tag{C8}$$

This completes the first half of our task.

Now let us evaluate C_2, S_1, and C_0. We begin with a standard result[25] for S_1

$$S_1(r,\tau)=\tfrac{1}{4}\ln\left[\frac{r+\tau}{r-\tau}\right]^2 \tag{C9}$$

from which it immediately follows from (C5) that

$$C_0(r,\tau)=r(r^2-\tau^2)^{-1} . \tag{C10}$$

Now, if we integrate $S_1(r,\tau)$ we obtain

$$\begin{aligned}C_2(r,\tau)&=-\int S_1(r,\tau)d\tau+\psi(r)\\ &=-\tfrac{1}{2}r\ln(r^2-\tau^2)-\tfrac{1}{4}\tau\ln\left[\frac{r+\tau}{r-\tau}\right]^2+\psi(r) ,\end{aligned} \tag{C11}$$

where $\psi(r)$ is some undetermined, arbitrary function of r. We can restrict the form of $\psi(r)$ since $P_1(t,t',r)$ must satisfy the wave equation. This turns out to be equivalent to demanding that

$$\frac{\partial^2}{\partial r^2}C_2(r,\tau)=-C_0(r,\tau) \tag{C12}$$

which implies that $\psi''(r)=0$. Consequently $\psi(r)$ must be of the form

$$\psi(r)=k_1+k_2r\ , \tag{C13}$$

where k_1 and k_2 are constants. If we also demand antisymmetry in r, then $k_1=0$. This means that $u(r)$ in (4.14) is restricted to be a constant, since $u(r)=\psi(r)/r$.

---

*Present address: Department of Physics, Tufts University, Medford, Massachusetts 02155.

[1] *The Very Early Universe,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University Press, London, 1983).
[2] P. Candelas and D. J. Raine, Phys. Rev. D **12**, 965 (1975).
[3] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, London, 1980).
[4] J. Géhéniau and C. Schomblond, Acad. R. Belg. Bull. Cl. Sci. **54**, 1147 (1968).
[5] C. Schomblond and P. Spindel, Ann. Inst. Henri Poincaré **A25**, 67 (1976).
[6] N. A. Chernikov and E. A. Tagirov, Ann. Inst. Henri Poincaré **A9**, 109 (1968).
[7] E. A. Tagirov, Ann. Phys. (N.Y.) **76**, 561 (1973).
[8] E. Mottola, Phys. Rev. D **31**, 754 (1984).
[9] 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).
[10] H. Rumpf, Phys. Rev. D **24**, 275 (1981).
[11] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[12] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, London, 1982).
[13] S. W. Hawking and I. G. Moss, Nucl. Phys. **B224**, 180 (1983).
[14] E. S. Abers and B. W. Lee, Phys. Rep. **9C**, 1 (1973).
[15] J. S. Kim, Nucl. Phys. **B196**, 285 (1982).
[16] T. S. Bunch and P. C. W. Davies, Proc. R. Soc. London **A360**, 117 (1978).
[17] T. S. Bunch and P. C. W. Davies, J. Phys. A **11**, 1315 (1978).
[18] N. D. Birrell, Proc. R. Soc. London **A361**, 513 (1978).
[19] L. H. Ford and A. Vilenkin, Phys. Rev. D **26**, 1231 (1982).
[20] A. D. Linde, Phys. Lett. **116B**, 335 (1982).
[21] L. H. Ford, Phys. Rev. D **31**, 710 (1985).
[22] A. H. Najmi and A. C. Ottewill, Phys. Rev. D **30**, 1733 (1984).
[23] A. Ashtekar and A. Magnon, Proc. R. Soc. London **A346**, 375 (1975).
[24] W. G. Unruh, Phys. Rev. D **14**, 870 (1976).
[25] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980).
[26] L. H. Ford and L. Parker, Phys. Rev. D **16**, 1601 (1977).