

THE VACUUM OF DE SITTER SPACE

A.D. DOLGOV*, M.B. EINHORN, AND V.I. ZAKHAROV

The Randall Laboratory of Physics
University of Michigan
Ann Arbor, MI 48109, USA

(Received October 28, 1994)

To resolve infrared problems with the de Sitter invariant vacuum, de Sitter symmetry must be broken. We discuss how the history of the de Sitter phase is related to the infrared cutoff. We illustrate how either (1) the diagonalization of the Hamiltonian for long-wavelength modes or (2) an explicit modification of the metric in the distant past leads to natural infrared cutoffs. The former case resembles a bosonic superconductor in which graviton-pairing occurs between non-adiabatic modes. While the dynamical equations respect de Sitter symmetry, the vacuum is not de Sitter invariant because of the introduction of an initial condition at a finite time. In the latter case, we indicate just how infrared cutoffs are associated with modifications of the metric. The implications for the one-loop stress tensor and the production of particles are also discussed.

PACS numbers: 98.80. Hw

1. Introduction

One of the most perplexing problems of cosmology concerns the magnitude of the cosmological constant¹, for which a stringent observational upper limit exists. Although supersymmetry requires the cosmological constant to vanish, supersymmetry is either explicitly or spontaneously broken in nature². As a result, the low-energy effective field theory is not supersymmetric, and the natural scale of the cosmological term in the gravitational

* Permanent address: ITEP, 113259, Moscow, Russia.

¹ For a review, see *e.g.* Refs [1, 2].

² Spontaneously broken global SUSY demands a nonzero vacuum energy in the broken phase. In supergravity a nonvanishing ρ_{vac} is not obligatory and one can adjust the parameters so that it vanishes when the symmetry is broken. However, the required fine-tuning is unnatural, typically required a precision of about 10^{-100} .

action should be no smaller than the magnitude of the vacuum energy associated with the quanta of matter and gravitons themselves. As the universe cools, such a cosmological term would eventually predominate over radiation and other forms of matter. Indeed, the most popular cosmology today, the inflationary scenario (for a review see [3]), assumes that our universe passed through an era in which the cosmological term dominated, and it is a total mystery why, when this phase terminated, we should be left in a universe with a tiny or vanishing vacuum energy. We do not advance a solution to the cosmological constant problem here, but address the simpler but nevertheless vexing question of what is the correct vacuum state during the inflationary era. The answer to this question is a prelude to addressing consistently the issue of whether such a phase is intrinsically unstable due to gravitational fluctuations, as some have suggested [4–8]. It also is crucial for understanding the quantum field theory of a free, massless, minimally coupled scalar field in a de Sitter background [4]. Before developing our approach, we shall review the problems associated with the vacuum state and previous attempts to deal with them.

The general form for the action may be written as the sum of gravitational and matter terms:

$$S = S_g + S_m. \quad (1)$$

The gravitational action S_g with a cosmological constant Λ is

$$S_g = \frac{1}{\kappa^2} \int d^4x \sqrt{g} (R - 2\Lambda), \quad (2)$$

with $\kappa^2 \equiv 16\pi/m_P^2$. The corresponding field equations are

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}(R - 2\Lambda) = -\frac{\kappa^2}{2}T_{\mu\nu}, \quad (3)$$

where $T_{\mu\nu}$ is the stress tensor of matter. It is presumed that the cosmological term, which may actually arise in part from vacuum fluctuations of particles, dominates the stress tensor of these particles, so we may neglect the matter action and $T_{\mu\nu}$ for the moment. de Sitter space is the maximally symmetric solution of the resulting equations. In conformal coordinates appropriate to the spatially flat sections, the metric takes the form

$$g_{\mu\nu} = a^2(\tau)\eta_{\mu\nu}, \quad (4)$$

where $\eta_{\mu\nu}$ is the usual Minkowski space metric, $a(\tau) \equiv -1/H\tau$, with $H = \sqrt{\Lambda/3}$.

The physical fluctuations in the metric, that is to say, the gravitons, in the neglect of self-interactions, obey the same field equation as the free,

massless, minimally coupled scalar field [9, 10]. Although our ultimate interest is in the purely gravitational case, for simplicity, we will consider the quantum field theory of a massless, minimally coupled scalar field in the de Sitter background. This field obeys the same equation of motion as certain modes of the graviton field, including the physical modes, but avoids the intricacies of tensor algebra, gauge fixing, unphysical modes, and ghosts that complicate the treatment of the graviton field itself. In this case, it is well-known that a de Sitter invariant vacuum state does not exist [11, 12, 4]. The claim that the same is true for quantum gravity, at least, linearized quantum gravity, is more controversial, but the two-point function (or propagator) certainly increases logarithmically for large spacelike separations [13, 5]. There are discussions [14, 15, 30, 7] suggesting that quantum gravity shares the problems of the massless, minimally coupled scalar, and we too believe that gravitons share these properties. To establish notation, we will review the situation briefly. The action for the scalar field is simply

$$S_m = \int d^4x \sqrt{g} \phi_{,\mu} \phi^{,\mu}. \quad (5)$$

In the conformal coordinates introduced above, the field equation takes the form

$$\phi_{,\mu}^{,\mu} + \frac{2\partial_\tau a}{a} \partial_\tau \phi = \phi_{,\mu}^{,\mu} - \frac{2}{\tau} \partial_\tau \phi = 0. \quad (6)$$

The solution of this equation with definite momentum \vec{k} is

$$\phi_k(\tau, \vec{x}) = \phi_k(\tau) \exp(i\vec{k} \cdot \vec{x}); \quad \text{with} \quad \phi_k(\tau) \equiv H(\tau - \frac{i}{k}) \exp(-ik\tau). \quad (7)$$

The corresponding field operator may be written as

$$\phi(x) = \int \frac{d^3k}{(2\pi)^3 \sqrt{2k}} \left[a_{\vec{k}} \phi_k(\tau) e^{i\vec{k} \cdot \vec{x}} + a_{\vec{k}}^\dagger \phi(\tau)^* e^{-i\vec{k} \cdot \vec{x}} \right], \quad (8)$$

with $a_{\vec{k}}$ and $a_{\vec{k}}^\dagger$ satisfying the commutation relations

$$[a_{\vec{k}}, a_{\vec{q}}^\dagger] = (2\pi)^3 \delta^{(3)}(\vec{k} - \vec{q}). \quad (9)$$

All other commutators are zero. The natural, de Sitter invariant no-particle state $|0, \text{in}\rangle$ that suggests itself is that state annihilated by the $a_{\vec{k}}$:

$$a_{\vec{k}} |0, \text{in}\rangle \equiv 0. \quad (10)$$

Unfortunately, although this would appear to be a perfectly acceptable, adiabatic in-state, this choice of vacuum leads to infrared divergences, illustrating the difficulties first pointed out in Ref. [11]. For example, the

Feynman propagator in this state is ill-defined because of an infrared divergence. The source of the problem is the i/k term in the wave function, which is a feature not present in Minkowski space (or for the conformally coupled massless scalar in curved space as well). The propagator takes the form [7]

$$G_F(x, x') = \frac{H^2}{4\pi^2} \left[\frac{\tau\tau'}{(x-x')^2} - \frac{1}{2} \ln(x-x')^2 \right]. \quad (11)$$

The scale of the argument of the log is not specified, reflecting the absence of an infrared cutoff. This propagator grows for large spacelike separations; simply inserting a scale into the logarithm would not change that. In an acceptable vacuum state, such correlations should fall to zero at large distances (so-called cluster decomposition.) It is noteworthy that the commutator of the fields does not display such infrared problems, a feature that we have exploited elsewhere [16]. This emphasizes that this is a uniquely quantum effect, without a classical counterpart. One may object that the Feynman propagator is irrelevant, because of complications defining the out-state in these coordinates. But perturbation theory requires the computation of other interesting quantities such as the time-ordered product between in-vacua or, similarly, the anticommutator G_1 defined by

$$G_1(\tau, \tau', \vec{x} - \vec{x}') \equiv \langle 0, \text{in} | \{ \phi(\tau, \vec{x}), \phi(\tau', \vec{x}') \} | 0, \text{in} \rangle. \quad (12)$$

These, like the Feynman propagator, suffer infrared divergences. This infrared problem is associated with the behavior at large distances and is not peculiar to large times³. The existence of a physically reasonable propagator (*i.e.*, two-point correlation functions) is a prerequisite to formulating a sensible perturbation expansion. Only then can one begin to ask, for example, whether to resolve infrared problems with the de Sitter invariant vacuum, de Sitter symmetry must be broken. We discuss how the history of the de Sitter phase is related to the infrared cutoff. We illustrate how either (1) the diagonalization of the Hamiltonian for long-wavelength modes or (2) an explicit modification of the metric in the distant past leads to natural infrared cutoffs. The former case resembles a bosonic superconductor in which graviton-pairing occurs between non-adiabatic modes. While the dynamical equations respect de Sitter symmetry, the vacuum is not de Sitter invariant because of the introduction of an initial condition at a finite time. In the latter case, we indicate just how infrared cutoffs are associated with modifications of the metric. The implications for the one-loop stress tensor and the production of particles are also discussed.

³ This divergence is, however, the origin of the instability discussed in Ref. [5]. Other infrared problems at large times are the basis for the instabilities discussed in Refs. [6, 7, 8].

A useful discussion of this problem has been given by Allen and Folacci [17], who concentrate on a different foliation of the de Sitter manifold, *viz.*, spatially closed coordinates. The advantage of these as opposed to flat (or open) coordinates is that the modes are discrete. Thus, the zero mode may be separated and modified, and they introduced modified vacua that were suggested by a construction of DeWitt [18] in which the zero mode is treated differently. However, such a construction does not resolve infrared problems for the flat or open coordinates, in which the momentum is continuous. Modification of the treatment of the zero mode alone will not change the cluster properties of the correlation functions. Consequently, people often resort to the introduction of an *ad hoc* infrared cutoff (such as $|\vec{k}| > k_0$) whose physical origin is somewhat obscure. Although the stress tensor is not infrared divergent, there is little hope that all observables will be cutoff-independent. Even for the closed coordinates, it is not so clear that the so-called Hadamard vacua discussed in Ref. [17] are physically selected.

It is well-known [11, 19, 20] that matter can provide modifications of the metric that resolve these infrared difficulties. In Refs [11, 19] it is shown that, at least for spatially-flat, Robertson–Walker spacetime, infrared divergences cannot develop in the course of time evolution. Therefore, such divergences are associated with the supposition that the de Sitter phase is eternal. If, in the distant past, the cosmological term (or vacuum energy) does not dominate, this should provide an infrared cutoff. In Ref. [20], an example is given in which the universe evolves from being radiation dominated to vacuum dominated, but the emphasis there is on the behavior of $\langle \phi^2 \rangle$, rather than on the propagator. Also, it seems that the momentum cutoff was associated with the Hubble constant H rather than with the time of transition to exponential expansion.

In this paper, we revisit this issue, emphasizing the way in which the infrared cutoff depends on the history of the de Sitter phase and the way this affects the two-point correlation function and particle production. We discuss this in several ways. In the first, we show that, even without reference to the physics at early times, one may be led to define a good vacuum state by consideration of the low-frequency, nonadiabatic modes and their coupling *via* the Hamiltonian. The relation between the indistinguishability among nonadiabatic states and the infrared problem is similar to infrared problems that occur elsewhere in field theory, such as in quantum electrodynamics. Subsequently, we consider explicit modifications of the metric at early times, in a manner similar to Ref. [20], and discuss how the infrared cutoff comes about in both a WKB treatment and in a model that can be solved exactly analytically.

The outline of the remainder of this paper is as follows: in the next section, we discuss the treatment of the nonadiabatic modes. In Section 3,

we evaluate the change in the stress tensor for our new vacuum state. In Section 4, we consider a different method of treating the history, by modifying the metric in the distant past, and using a WKB expansion for an evaluation of the effects. In Section 5, we consider yet another modification of the metric, one which lends itself to analytic solution. In Section 6, we summarize our results and comment on their significance.

2. Nonadiabatic modes

Even though the stress tensor is known to be free of infrared problems, an examination of the Hamiltonian does in fact provide some insight into the nature of the problem. Because the metric is explicitly time-dependent, momentum but not energy is conserved. Nevertheless, even though the Hamiltonian is not time-independent, it still governs the time-evolution of the field. It is useful to construct the Hamiltonian corresponding to the matter action Eq. (5). We find

$$H(\tau) = \int \frac{d^3k}{(2\pi)^3} k \left[\frac{r_k}{2} \left(a_{\vec{k}} a_{\vec{k}}^\dagger + a_{\vec{k}}^\dagger a_{\vec{k}} \right) + \frac{s_k}{2} a_{\vec{k}} a_{-\vec{k}} + \frac{s_k^*}{2} a_{\vec{k}}^\dagger a_{-\vec{k}}^\dagger \right], \quad (13)$$

where

$$r_k \equiv \left(1 + \frac{1}{2(k\tau)^2} \right), \quad s_k \equiv -\frac{\exp(-2ik\tau)}{k\tau} \left(i + \frac{1}{2k\tau} \right). \quad (14)$$

The coefficients r_k, s_k depend only on the magnitude $k = |\vec{k}|$. Notice that they are in fact functions of the product $k\tau$ only. From this expression, we see that in general, the Hamiltonian causes transitions increasing or decreasing the number of quanta by two. In particular, the Hamiltonian makes transitions between the de Sitter invariant vacuum $|0, in\rangle$ defined in Eq. (10) and the two-particle states

$$|\vec{k}, -\vec{k}\rangle \equiv a_{\vec{k}}^\dagger a_{-\vec{k}}^\dagger |0, in\rangle. \quad (15)$$

Note that this form of the Hamiltonian is precisely that of a BCS superconductor in quadratic approximation⁴, except that the quanta here are bosons rather than electrons. Nevertheless, one is tempted to diagonalize the Hamiltonian at some particular time by a Bogoliubov transformation [21, 22] and to try to define no-particle states with respect to the transformed basis. This naive expectation is not correct, and the vacuum state

⁴ See any modern text on superconductivity [21].

would be extremely sensitive to the time chosen. Such a procedure has been a highly disreputable thing to suggest for many years [23, 24], especially because it leads to an infinite density of particles at any later time. In any case, the off-diagonal s_k vanishes in the distant past ($\tau \rightarrow -\infty$), so one would be inclined to choose the in-state at $\tau = -\infty$ to be precisely $|0, in\rangle$, especially since the rate of change of the metric in conformal coordinates is so slow in the distant past, and the Hamiltonian approaches a finite, time-independent limit.

This line of reasoning is perfectly sensible and would be the end of the story were it not for the fact that the theory remains infrared divergent in that state. Moreover, in our present thinking about cosmology, in which the inflationary phase arises either after cooling from a hot initial phase or from some sort of phase transition at very early times, in fact it is an idealization to presume that the universe is in the de Sitter phase at all times. Let us consider how our description might change as a result of the supposition that the idealized Lagrangian that we have been discussing is only valid for times exceeding some early time τ_0 . Our concern is with what the state of the universe must be. At first, one would think that the no-particle state will correspond to the usual adiabatic vacuum⁵. In this construction, at any given time, the modes of the field may be divided into high- and low-momentum modes, the transition being set by the time rate of change of the scale factor. In conformal time, this is

$$\frac{d}{d\tau} \log(a(\tau)) = \frac{1}{\tau}. \quad (16)$$

Since we may not entertain times earlier than τ_0 , we may think of those momenta with $k|\tau_0| < 1$ as non-adiabatic and those with $k|\tau_0| > 1$ as adiabatic. For the adiabatic modes, we will assume that the standard construction prevails, so that, to a good approximation, the correct vacuum state will be like $|0, in\rangle$ in that it will be annihilated by the $a_{\vec{k}}$, but only for $k > 1/|\tau_0|$.

However, the nonadiabatic modes require further discussion. It is reasonable to ask whether an observer might design an experiment to probe these low frequency or very long wavelength modes. In fact, because of the horizon in de Sitter space, they are impossible to detect. We allude to the well-known fact that, no matter how long one waits, an observer at the origin can only receive signals that originate later than τ_0 at comoving coordinates $|\vec{x}| < |\tau_0|$. Correspondingly, by the uncertainty principle, a detector could not be built to probe (conformal) momenta $|\vec{k}| < 1/|\tau_0|$, since no signal can be communicated over coordinates greater than $|\tau_0|$. Consider

⁵ The adiabatic vacuum was developed by Parker [25] and others and is reviewed in Ref. [24].

in particular the so-called two-particle states defined in Eq. (15). Normally, we think of them as consisting of two distinct quanta moving in opposite directions. However, for frequencies $k < 1/|\tau_0|$, it would be impossible to make such a determination. In fact, such nonadiabatic modes are essentially indistinguishable from the no-particle state $|0, in\rangle$ itself. Thus, we have a situation in which the Hamiltonian causes transitions between physically indistinguishable states, the classic situation discussed by Lee and Nauenberg [26]. They argued quite generally that one must use techniques of degenerate perturbation theory if one is to avoid encountering infrared divergences. In particular, they showed that one must either sum over degenerate states in forming observables or *diagonalize the Hamiltonian* in the space of degenerate states. In the present context, this suggests that, while the definition of the vacuum $|0, in\rangle$ given in Eq. (10) may be correct for the adiabatic modes, one should modify the vacuum *for the nonadiabatic modes* by reference to a basis in which the Hamiltonian is diagonal.

In an interacting field theory, diagonalizing the Hamiltonian is a formidable task, but since we are working here with a quadratic Hamiltonian, it can be carried out explicitly by means of a Bogoliubov transformation

$$b_{\vec{k}} = \alpha_k a_{\vec{k}} + \beta_k a_{-\vec{k}}^\dagger, \quad (17)$$

with coefficients subject to the constraints

$$|\alpha_k|^2 - |\beta_k|^2 = 1. \quad (18)$$

α_k and β_k depend only on the frequency $k = |\vec{k}|$ and not the direction⁶.

This discussion suggests the following improved approximation to the vacuum state of de Sitter space. For an arbitrary Bogoliubov transformation, Eq. (17), one may define a vacuum state $|\tau_0, in\rangle$ by

$$b_{\vec{k}} |\tau_0, in\rangle = 0. \quad (19)$$

The state is defined by the prescription one gives for the coefficients α_k and β_k . As suggested earlier, we choose the usual prescription $\alpha_k = 1, \beta_k = 0$ for adiabatic modes $k > 1/|\tau_0|$. Thus, for these modes, this vacuum is the same as the de Sitter vacuum. However, for the non-adiabatic modes, $k < 1/|\tau_0|$, we choose these coefficients such that they diagonalize the Hamiltonian. (We shall come back to discuss the transition between the two regimes later.) While the physical motivation based on the Lee–Nauenberg theorem seems quite compelling, it remains to be shown that this prescription is precisely of the sort described by Ford and Parker [11], *i.e.*, it resolves the

⁶ We will display the coefficients α_k, β_k below.

infrared problems associated, for example, with the two-point correlation functions. In particular, since the Hamiltonian varies with time, it may not be immediately obvious that diagonalizing the Hamiltonian at some initial time τ_0 will remove the infrared divergences for all later times.

As noted earlier, the Hamiltonian is explicitly a function of time, but it is the Hamiltonian $H(\tau_0)$ at the initial time τ_0 that we wish to diagonalize. It is a simple exercise to obtain the form of the Bogoliubov transformation. We define an angle θ_k by the phase of s_k , i.e., $s_k \equiv |s_k| \exp(-i\theta_k)$. Then we find

$$\frac{\beta_k}{\alpha_k} = R_k \exp(-i\theta_k), \text{ with } R_k \equiv \frac{r_k}{|s_k|} - \sqrt{\left(\frac{r_k}{|s_k|}\right)^2 - 1}. \quad (20)$$

A useful form for the phase angle is

$$\theta_k = 2k\tau_0 - \gamma_k + \pi, \text{ where } \tan \gamma_k \equiv 2k\tau_0. \quad (21)$$

Without loss of generality, we may choose the coefficient α_k to be real, so that the phase $-\theta_k$ is associated with β_k . Then, for the magnitudes, we find

$$\alpha_k = \frac{1}{\sqrt{1 - R_k^2}}, \quad |\beta_k| = \frac{R_k}{\sqrt{1 - R_k^2}}, \quad k < 1/|\tau_0|. \quad (22)$$

The Hamiltonian for the nonadiabatic modes H_{na} then takes the diagonal form

$$H_{na}(\tau_0) = \int_{k < 1/|\tau_0|} \frac{d^3k}{(2\pi)^3} k \left\{ b_{\vec{k}}, b_{\vec{k}}^\dagger \right\}, \quad (23)$$

As remarked earlier the state so constructed resembles superconductivity, as the vacuum state involves pairing of quanta of equal and opposite momenta. It is important to recognize that the pairing of nonadiabatic modes does not correspond to condensation of the scalar field. The symmetry breaking arises from off-diagonal long-range order. Thus, in terms of the original quanta, there are extremely long-range correlations in the vacuum which, as we shall see, resolve the infrared divergences. Note, however, that unlike ordinary superconductors, diagonalization does not result in a mass for the elementary excitations. Of course, the expectation value of the Hamiltonian $H(\tau_0)$ is lower in the state $|\tau_0, in\rangle$ than in the original de Sitter invariant state $|0, in\rangle$. So we may think of this as spontaneous breaking of de Sitter symmetry, but it also exhibits aspects of explicit breaking, since it involves a parameter (τ_0) that is not part of the original Lagrangian. On the other hand, if we regard the gravitational action as being defined only on the

interval of time $\tau_0 < \tau$, then one may say it was present in the theory from the beginning.

In the transformed basis, the expansion of the field becomes

$$\phi(x) = \int \frac{d^3k}{(2\pi)^3 \sqrt{2k}} \left[b_{\vec{k}} u_k(\tau) e^{i\vec{k} \cdot \vec{x}} + b_{\vec{k}}^\dagger u(\tau)^* e^{-i\vec{k} \cdot \vec{x}} \right], \quad (24)$$

where the modified wave functions are

$$u_k(\tau) = \alpha_k \phi_k(\tau) - \beta_k^* \phi_k^*(\tau). \quad (25)$$

It is useful to have approximate expressions for these new wave functions for small momenta. In the nonadiabatic regime, $k|\tau_0| \ll 1$, we then find $R_k \approx 1 - 2(k\tau_0)^2$, so that

$$\begin{aligned} \alpha_k &= \frac{1}{2k|\tau_0|} \left(1 + \frac{1}{2}(k\tau_0)^2 + \dots \right) \\ |\beta_k| &= \frac{1}{2k|\tau_0|} \left(1 - \frac{3}{2}(k\tau_0)^2 + \dots \right) \\ \theta_k &= \pi + \frac{4}{3}(k\tau_0)^2 + \dots \end{aligned} \quad (26)$$

In particular, $\alpha_k + \beta_k \rightarrow O(k|\tau_0|)$ as $k \rightarrow 0$, as required for removal of the infrared singularity [11]. To see this explicitly, we may use these results to obtain the behavior of the wavefunction, Eq. (25), in the nonadiabatic regime. Since the conformal time is in the range $\tau_0 < \tau < 0$, we may expand for all time to obtain

$$u_k(\tau) \approx iH\tau_0 \left[1 + \frac{1}{2}((k\tau)^2 + (k\tau_0)^2) \right]. \quad (27)$$

Thus the $1/k$ infrared behavior as $k \rightarrow 0$ characteristic of the de Sitter wave function, Eq. (7), has been replaced in the new basis by a constant $a(\tau_0)^{-1}$, no worse than the flat space or conformal wave function. Working in this modified vacuum, the infrared singularities will be removed, and quantities such as the propagator in our new vacuum will be well-defined.

For example, the anticommutator G_1 defined with respect to the state $|\tau_0\rangle$ becomes

$$G_1(\tau, \tau', \vec{x} - \vec{x}') = \int \frac{d^3k}{(2\pi)^3 2k} \left[u_k(\tau) u_k^*(\tau') e^{i\vec{k} \cdot (\vec{x} - \vec{x}')} + \text{c.c.} \right]. \quad (28)$$

For $k|\tau_0| > 1$, the integrand is unchanged from the de Sitter-invariant vacuum, but for $k|\tau_0| < 1$, the integrand is modified in such a way that it tends to a constant. If one considers large spacelike separations, e.g.,

$\tau = \tau'$, $|\vec{x} - \vec{x}'| \rightarrow \infty$, the behavior will be dominated by wave numbers $k < 1/|\vec{x} - \vec{x}'|$. For this quantum amplitude, one is not restricted to the causal region, so one may consider the regime in which $|\vec{x} - \vec{x}'| \gg |\tau_0|$. In this regime, the dominant contribution will come from the nonadiabatic modes where the wave functions may be approximated by Eq. (27). Thus, for large separations, G_1 will fall to zero,

$$G_1(\tau, \tau, \vec{x} - \vec{x}') \sim \frac{H^2 \tau_0^2}{(\vec{x} - \vec{x}')^2}. \quad (29)$$

Therefore, this vacuum state satisfies cluster decomposition of the fields. On the other hand, the correlator G_1 will increase logarithmically out to separations on the order of $|\tau_0|$ before turning over, so that pairing only significantly changes the behavior outside the causally connected domain. Nevertheless, this is important for understanding the behavior of radiative corrections. This damping of long wavelength modes could also have implications for the sort of vacuum fluctuations that emerge at the completion of an inflationary epoch.

At the other extreme, consider the behavior of G_1 at short distances. Of course, the correlator blows up as in flat space, but the first contribution to the short distance expansion that depends on the structure of spacetime comes from $\langle \phi(\tau, \vec{x})^2 \rangle$. By translation invariance of de Sitter space in conformal coordinates, we may choose $\vec{x} = 0$. Of course, an operator such as ϕ^2 requires renormalization in any case. However, it has been argued [27, 28, 30, 4] that, in the usual, de Sitter vacuum, the renormalized vacuum expectation value grows as $\ln|\tau|$ owing to the infrared divergence in this state. In our new vacuum $|\tau_0\rangle$, this is not IR divergent. Thus, any $\log|\tau|$ dependence must then come from ultraviolet (conformal) momenta, so that it becomes a question of renormalization. Since ϕ^2 is not an observable operator in the present theory, it is something of a moot point, but we will return to it in subsequent sections.

One might entertain other states that are equally good candidates as vacua. For example, in order to make the transition from the nonadiabatic to adiabatic modes smooth, one would be tempted simply to adopt the form β_k given in Eqs (20) and (22) for *all* k , not just the adiabatic modes, but then to multiply the adiabatic modes by a damping factor, such as $\exp(-bk|\tau_0|)$ for some constant b of order 1. Correspondingly, the magnitude of α_k would be modified so that the constraint Eq. (18) is satisfied. This seems to us to be a perfectly acceptable definition of an adiabatic vacuum, although probably more can be said about the transition between nonadiabatic and adiabatic modes. These issues would seem to depend on the details of the history of the de Sitter phase.

The usual objections to defining the no-particle state by reference to diagonalizing the Hamiltonian do not apply to our construction. Normally,

one would be concerned that there would be an infinite density of particles produced at all times later than the reference time τ_0 , but that is not the case here since it is only the very low frequency modes that are significantly modified. Moreover, we have argued that one may not observe the non-adiabatic modes as particles; at best, their effects could be felt indirectly, for example, through their contribution to the energy density. In fact, we will show that their contributions to the energy density are small, vanishing with τ_0 as τ_0^{-2} .

One consequence of the presumed logarithmic growth of the propagator for the massless, minimally coupled scalar was the impossibility of breaking continuous symmetries in de Sitter space [29]. We see that, in a good vacuum such as $|\tau_0\rangle$ that manifests cluster decomposition, this conclusion is invalidated. There is no reason why spontaneous symmetry breaking should not proceed as usual in de Sitter space.

The preceding considerations are naturally relevant to the vacuum state of quantum cosmological gravity (QCG), the theory of simply the gravitational field in the presence of a non-zero cosmological constant but without any matter. As mentioned earlier, the small fluctuations about the de Sitter background involve gravitons that include modes which, in the linear approximation, satisfy the same equation of motion as massless, minimally-coupled scalars⁷. Thus, we expect the vacuum state of QCG to resemble $|\tau_0\rangle$, in which the non-adiabatic gravitons of opposite helicity and momenta are paired. With some additional effort, such a description can be made gauge-invariant. The situation is analogous to a BCS superconductor; the gap parameter is not gauge-invariant, so pairing would appear to break gauge invariance. But it can be shown that the Ward identities are fulfilled, so it is only the global $U(1)$ -symmetry that is broken. Similarly here, covariant conservation of the stress tensor is intact, even though global de Sitter symmetry is broken. Unlike the superconductor, however, there are no Goldstone bosons associated with this symmetry breaking. In this respect, our situation resembles explicit breaking: even though the equations of motion remain de Sitter invariant, the action over the interval $\tau > \tau_0$ is not de Sitter invariant.

3. Stress tensor in the state $|\tau_0, in\rangle$

As another application of this vacuum, consider the renormalized, one-loop stress tensor $T_{\mu\nu}$. Even in the de Sitter-invariant state $|0, in\rangle$, it is not beset by infrared divergences, and its vacuum expectation value is finite, although its value depends on the renormalization prescription. To gain

⁷ This has been reviewed in Refs. [7, 30, 16].

insight into the effects of modifying the vacuum in this way, we evaluate its behavior at one-loop order in the state $|\tau_0, in\rangle$. A direct calculation of the *unrenormalized* contribution gives for the nonzero elements

$$\begin{aligned}\langle T_{00} \rangle &= \frac{1}{2} \int \frac{d^3 k}{(2\pi)^3 2k} (|\dot{u}_k(\tau)|^2 + k^2 |u_k(\tau)|^2) , \\ \langle T_{ii} \rangle &= \frac{1}{2} \int \frac{d^3 k}{(2\pi)^3 2k} \left(|\dot{u}_k(\tau)|^2 - \frac{k^2}{3} |u_k(\tau)|^2 \right) .\end{aligned}\quad (30)$$

We rewrite the various terms in the integrand as, for example,

$$|u_k(\tau)|^2 = (1 + 2|\beta_k|^2) |\phi_k(\tau)|^2 - 2 \operatorname{Re} \{ \alpha_k \beta_k \phi_k^2(\tau) \} , \quad (31)$$

where we used $|\alpha_k|^2 = 1 + |\beta_k|^2$ in the first term. The value when $\beta_k = 0$ is of course the de Sitter invariant contribution, whereas the terms involving β_k represent the change arising from our modification of the vacuum. The de Sitter invariant contribution is completely independent of τ_0 . It gives rise to quartic and quadratic ultraviolet divergences which, in a mass-independent renormalization scheme such as minimal subtraction, is completely removed by counterterms⁸. The process of renormalization also introduces a well-known anomalous contribution [31] to the stress-tensor which will be small for H^2/m_{Pl}^2 small. The remaining terms involving β_k depend on the initial time τ_0 and cannot be removed by renormalization. As our interest is in the difference between the stress tensor in the de Sitter vacuum and the stress tensor our new vacuum $|\tau_0, in\rangle$, we will ignore the anomalous contribution here. Thus, the change in the contributions to the one-loop renormalized stress tensor is

$$\begin{aligned}\langle T_{00} \rangle &= \frac{1}{2\pi^2} \int_0^{1/|\tau_0|} k dk \left\{ \left[|\beta_k \dot{\phi}_k|^2 \right. \right. \\ &\quad \left. \left. - \operatorname{Re}(\alpha_k \beta_k \dot{\phi}_k^2) \right] + k^2 [|\beta_k \phi_k|^2 - \operatorname{Re}(\alpha_k \beta_k \phi_k^2)] \right\} ,\end{aligned}\quad (32)$$

with a similar expression for $\langle T_{ii} \rangle$. Although the expressions for β_k are quite complicated, we may use simple scaling arguments to obtain the general form of $\langle T_{\mu\nu} \rangle$. First of all, note that $|k\phi_k|^2 = H^2(1 + (k\tau)^2)$ and $|\phi_k|^2 = H^2(k\tau)^2$. Since α_k and β_k are functions only of $k\tau_0$, we may rescale $k \rightarrow k/\tau_0$ to see that $\langle T_{00} \rangle$ is of the form $A/\tau_0^2 + B\tau^2/\tau_0^4$, where A and B are

⁸ People often adopt a renormalization scheme that retains a finite value for this piece, but it is arbitrary.

numerical constants that depend on the detailed form of the function β_k . Therefore, the energy density and pressure are

$$\langle T_0^0 \rangle \equiv \rho = H^4 \left[+A \frac{\tau^2}{\tau_0^2} + B \frac{\tau^4}{\tau_0^4} \right], \quad (33)$$

$$-\langle T_i^i \rangle \equiv p = H^4 \left[-\frac{A}{3} \frac{\tau^2}{\tau_0^2} + \frac{B}{3} \frac{\tau^4}{\tau_0^4} \right], \quad (\text{not summed on } i). \quad (34)$$

It is easy to check that these forms obey the covariant conservation constraint $T_{\mu\nu}^{;\nu} = 0$, providing an independent check on the relative signs appearing in the expression above. While our modified vacuum state contributes nonzero energy density (with $\langle T_{\mu\nu} \rangle$ not proportional to the metric tensor $g_{\mu\nu}$), these effects rapidly die away. (Recall that $-H\tau = \exp(-Ht)$, so they decay exponentially in time.) The signs of the pressure provide insight into the two terms. The B -term is traceless and has the form of a relativistic gas that is diluted as $a(\tau)^{-4}$. The A -term has negative pressure and falls only as $a(\tau)^{-2}$. These are the amplified waves originally discussed by Grishchuk [10]. Finally, as expected, as $\tau_0 \rightarrow -\infty$, these contributions vanish.

Of course, because $\langle T_{\mu\nu} \rangle$ is non-zero, there will be feedback on the background metric, so the de Sitter form is no longer exact. However, for $H^2/m_{Pl}^2 \ll 1$, this is a small distortion which, according to Eqs (33) and (34), decays with time.

4. Modifications of the metric as an infrared cutoff

In this section we will pursue the same idea of introducing an infrared cutoff through the finite time of existence of de Sitter phase but in a different manner. Namely we will make an ad hoc assumption that metric has form (4) with the scale factor $a(\tau)$ equal to

$$a^2(\tau) = \frac{1}{H^2\tau^2} + \frac{\delta}{H^2} \quad (35)$$

so that at large times, or $\tau \rightarrow 0$ we recover de Sitter space. On the other hand assuming δ to be positive we see that as $\tau \rightarrow -\infty$, or in the distant past, the metric becomes that of Minkowski spacetime. To be consistent with the Einstein's equations we tacitly assume a nontrivial matter distribution which would drive to the solution (35) (see also below). We will demonstrate that finite δ does induce an infrared cut off both when one evaluates $\langle 0|\phi^2|0\rangle$ and $\langle 0|T_{\mu\nu}|0\rangle$. What might be most amusing, is that the infrared cutoff manifests itself in these two cases in different fashion.

Since the metric has been modified we need to find the corresponding modes of the scalar field which would replace now solutions (7). We can try either to solve the equations of motion for the scalar field or to apply WKB expansion. The former approach has a shortcoming of being somewhat cumbersome while the latter, at first sight, is not suited at all to study infrared problems we are interested in. In fact the both ways work well. In this section we apply the high-frequency, or WKB expansion, while discussion of the exact solution is postponed until the next section.

Let us first explain the general strategy of dealing with infrared problems within the WKB expansion and introduce to this end the necessary notation. We are looking for a solution ϕ_k with 3-momentum k which is represented as

$$\phi_k(x) = \frac{\exp(i\mathbf{k} \cdot \mathbf{x})}{(2\pi)^{3/2} a(\tau)} \chi_k(\tau), \quad (36)$$

where χ_k in massless case satisfies the equation (see, *e.g.*, [32, 33]):

$$\frac{d^2 \chi_k}{d\tau^2} + \{k^2 - (\xi - \frac{1}{6})a^2 R\} \chi_k = 0. \quad (37)$$

Here R is the Ricci scalar which can be computed as

$$R = a^{-2} \left(3\dot{D} + \frac{3}{2}D^2 \right), \quad D = \frac{2\ddot{a}}{a} \quad (38)$$

and the dot denotes the derivative with respect to the conformal time τ . The quasiclassical solution to (37) is

$$\chi_k = \frac{\exp(-i \int W_k(\eta') d\eta')}{(2W_k(\tau))^{1/2}}, \quad (39)$$

$$W_k^2 = k^2 + (\xi - \frac{1}{6})a^2 R - \frac{\ddot{W}_k}{2W_k} + \frac{3}{4} \left(\frac{\dot{W}_k}{W_k} \right)^2. \quad (40)$$

Eq. (40) allows to obtain W_k as an expansion in k^{-2} in a regular way.

In particular, the first two terms in the expansion are well known:

$$W_k^2 = k^2 \left(1 - \frac{(\xi - \frac{1}{6})a^2 R}{2k^2} + O(k^{-4}) \right), \quad (41)$$

or

$$\phi_k = (2\pi)^{-3/2} \exp(i\mathbf{k} \cdot \mathbf{x} - ik\tau) \left(\tau + i \frac{(\xi - \frac{1}{6})\tau a^2 R}{2k} \dots \right), \quad (42)$$

while further terms can be obtained iteratively.

What is special about de Sitter background and $\xi = 0, 1/6$ is that these first two terms (42) reproduce the exact result (see Eq. (7)). It is worth emphasizing that this termination of the expansion can be established within the quasiclassical expansion itself. The coefficient in front of k^{-3} term in the expansion (42) does turn out to be zero if $Ra^2 = 12/\tau^2$.

As a rule evaluation of a physical quantity involves integration over modes. Consider for example $\langle 0|\phi^2|0\rangle$. From Eq. (42) one immediately gets [32]

$$\langle 0|\phi^2|0\rangle = \frac{1}{(2\pi)^3} \int \frac{d^3k}{2k} \left(1 - \frac{(\xi - \frac{1}{6})a^2 R}{2k^2} + O(k^{-4}) \right). \quad (43)$$

We will return later to the question of the ultraviolet regularization. In fact, the integrand is obtained in the high momentum limit, and the exact one should be infrared finite. However, the signal of the infrared cutoff may be inferred from the higher-order terms like k^{-4} . In case that one starts with a quantity which is ultraviolet divergent (like $\langle 0|\phi^2|0\rangle$), the following two simple rules seem to apply. First, all power-like ultraviolet divergencies can be removed through subtractions so that we have to concentrate on the log term. Clearly, the coefficient in front of the log is uniquely fixed by the k^{-2} expansion. Second, the next term in the expansion after the logarithmic one, *i.e.* the first term which diverges as a power in the infrared provides with an estimate of the infrared cutoff in the log. Indeed the infrared cutoff can be estimated by finding at which k^2 the term producing the log and the next one are of the same order.

From this point of view, infrared problems in the de Sitter background arise not because subsequent terms in the k^{-2} expansion (see Eq. (42)) produce power-like infrared divergences but because for $\xi = 1/6$ the high-frequency expansion terminates, and there are in fact no further terms in (43). As a result there is no infrared cutoff for the log and $\langle 0|\phi^2|0\rangle$ is undetermined. This is just the conclusion which was drawn from analyzing the exact solution for ϕ_k . Now we expect that by introducing $\delta \neq 0$ (see Eq. (35)) we destroy the conspiracy of the coefficients which terminates the expansion (42) and there emerges an infrared cutoff. We proceed now to check these expectations explicitly. To first order in δ

$$Ra^2 = \frac{12}{\tau^2} (1 - \frac{1}{2}\delta\tau^2). \quad (44)$$

Let us note that as far as Eq. (40) is concerned such a modification of the metric is equivalent perturbatively to introducing positive $\xi \neq 0$,

$$\xi_{\text{eff}} \approx 3\delta\tau^2. \quad (45)$$

As is well known, [24] there is an important difference between positive and negative ξ . Namely, negative ξR is like a negative mass-squared inducing an

instability that leads to spontaneous breakdown of the de Sitter-invariant vacuum state. As a result, one cannot match the WKB expansion with the correct infrared behavior. If ξ is positive, there is no danger of mismatch between the infrared and ultraviolet expansions.

Keeping terms linear in δ and expanding ϕ_k in k^{-1} leads to the following results:

$$\begin{aligned} \frac{\phi_k}{H} = \frac{e^{(ik \cdot x - ik\tau)}}{(2\pi)^3} & \left(\tau - \frac{i}{k} + 0 \frac{1}{\tau k^2} + 0 \frac{i}{\tau^2 k^3} \right. \\ & \left. + 0 \frac{1}{\tau^3 k^4} - \frac{i\delta\tau}{2k} - \frac{\delta\tau}{k^2} + \frac{i3\delta}{4k^3} + 0 \frac{\delta}{\tau k^4} + O(k^{-5}) \right). \end{aligned} \quad (46)$$

Here we indicated the vanishing coefficients as well since they illustrate the statement above that the WKB expansion is perfectly consistent with the exact result for ϕ_k (at $\delta = 0$).

Now we are in position to apply the rules of defining the infrared cutoff mentioned above. Indeed, we see that the series in k^{-2} which represents $\langle 0|\phi^2|0\rangle$ does have further terms. The infrared cutoff, k_{IR} can be estimated by equating two subsequent terms in the expansion:

$$k_{\text{IR}}^{-1} \sim \delta k_{\text{IR}}^{-3}, \quad k_{\text{IR}}^2 \sim \delta. \quad (47)$$

Note that the infrared cutoff (47) is time independent. Thus, for $\langle 0|\phi^2|0\rangle$ we find (omitting the subtraction term associated with the UV quadratic divergence)

$$\langle 0|\phi^2|0\rangle = \frac{H^2}{4\pi^2} \ln \Lambda_{\text{UV}}^2 / \delta, \quad \xi = \frac{1}{6}. \quad (48)$$

According to the standard arguments the ultraviolet cutoff, Λ_{UV} does depend on time in the conformal coordinates,

$$\Lambda_{\text{UV}} \sim M a(\tau), \quad (49)$$

where M is some mass parameter. Indeed, all the masses are scaled with $a(\tau)$ in de Sitter background. If so, we reproduce the well-known result [27, 28, 20] for $\langle 0|\phi^2|0\rangle$:

$$\langle 0|\phi^2|0\rangle \sim -\frac{H^2}{2\pi^2} \ln \tau. \quad (50)$$

We turn now to consideration of the vacuum expectation value $\langle 0|T_{\mu\nu}|0\rangle$. The change brought by $\delta \neq 0$ is a multiplicative factor $(1 - 3\delta/2k^2)$ as far as energy and momentum associated with each mode ϕ_k are concerned. If

we turn to the integral over modes determining the vacuum expectation value then this factor is crucial since the term proportional to δ gives a logarithmically divergent integral over k . Keeping this log term we find

$$\langle 0|T_{00}|0\rangle = -3\langle 0|T_{ii}|0\rangle = -\delta \frac{3H^2}{8\pi^2} \cdot \ln \frac{\Lambda_{UV}}{\Lambda_{IR}}, \quad (51)$$

where Λ_{UV} , Λ_{IR} are ultraviolet and infrared cutoffs and no summation over i is assumed. Moreover, as explained above, we do not keep power-like divergent terms.

For the ultraviolet cutoff we still have (49). It is most interesting that the estimate (47) for the infrared cutoff is now changed. The reason is that the terms kept in the expansion (46) are sufficient only to find the coefficient in front of the log term (51) while for the estimate of the infrared cutoff we need the next term in the k^{-2} expansion of the solutions ϕ_k . One can obtain the following estimate

$$\Lambda_{IR}^2 \tau^2 \sim 1. \quad (52)$$

It might be worth noting that the time dependence exhibited by (52) is so to say normal, since the WKB expansion is an expansion in inverse powers of $k\tau$. The fact that this time dependence did not manifest itself in (47) is quite unique and is due to the termination of the WKB expansion in pure de Sitter background.

Thus, we have for the contribution to the vacuum expectation value of the stress tensor:

$$\rho \equiv \langle 0|T_0^0|0\rangle \propto H^4 \tau^2 \delta, \quad p = -\frac{1}{3}\rho. \quad (53)$$

Two remarks concerning this result are now in order. First, the vacuum expectation value is to satisfy the covariant conservation condition

$$D_\mu \langle 0|T_\nu^\mu|0\rangle = 0, \quad \text{or} \quad \tau \frac{d}{d\tau} \rho = 3(\rho + p), \quad (54)$$

where the latter equation makes use of de Sitter metric, and the energy density ρ and pressure p are defined in Eqs (33), (34). The use of de Sitter metric is justified since $\langle 0|T_{\mu\nu}|0\rangle$ starts from terms proportional to δ and in the linear approximation one can neglect that $\delta \neq 0$ otherwise.

The condition (54) establishes a relation between time dependence and the coefficient of proportionality between ρ and p , in our case $p = -1/3\rho$. Moreover, it is worth emphasizing that constraint (54) should be satisfied both for each mode with fixed k and for the sum over the modes, or the vacuum expectation value. Since for a logarithmically divergent integral the relation between ρ and p for the sum over modes is the same as for each mode one immediately concludes that no extra time dependence can

be associated with the $\ln(\Lambda_{UV}/\Lambda_{IR})$. This is a nontrivial check of general covariance which is passed by the estimates above.

Second, Eq. (53) implies that $\langle 0|T_{\mu\nu}|0\rangle$ is not proportional to $g_{\mu\nu}$ which in de Sitter background would result in $\langle 0|T_{00}|0\rangle = -\langle 0|T_{ii}|0\rangle$. If we were considering pure de Sitter solution (*i.e.* $\delta = 0$) then $g_{\mu\nu}$ would be the only tensor structure available, and Eq. (53) would not be possible. However, in the approximation linear in δ , we do have a new tensor structure emerging in $R_{\mu\nu}$:

$$R_{00} = \frac{3}{\tau^2}(1 + \delta\tau^2), \quad R_{ii} = -\frac{3}{\tau^2}(1 - \delta\tau^2). \quad (55)$$

By virtue of equations for the gravitational field the term proportional to δ reflects the structure of $T_{\mu\nu}$ of matter needed to produce solution (35). This new tensor structure is also manifested in Eq. (53).

Thus, by assuming the scale factor to be given by (35) we have tacitly introduced nontrivial stress energy of matter which drives metric to such a solution. Then contribution (53) can be viewed as a radiative correction to the stress tensor of the matter. The plausible interpretation of this correction is that during the transition from Minkowski to de Sitter phases scalar particles are produced.

In conclusion of this section, let us compare this method with that of the preceding sections involving the introduction of the initial time τ_0 . Note that, although technically the prescriptions for the infrared cutoff look very similar since they modify the situation in the distant past, the physics is actually quite different in these two cases. Indeed, in the latter case we assume that there is some matter distribution which causes $\langle 0|T_{\mu\nu}|0\rangle \neq \text{const} \cdot g_{\mu\nu}$. In the former case, the deviations from pure de Sitter solution are entirely due to the initial state.

5. Exact solution for a modified metric

Here we will illustrate some of the results discussed above for the particular case of a modified de Sitter metric which was originally flat (at negative time infinity) and which admits an exact solution for the minimally coupled scalar field ϕ . We chose the metric to be of the usual Robertson–Walker form with the scale factor:

$$a(t) = 1 + \exp(Ht) = [1 - \exp(H\tau)]^{-1}, \quad (56)$$

where the conformal time τ is expressed through t as

$$\tau = -H^{-1} \ln[1 + \exp(-Ht)]. \quad (57)$$

As above the future infinity, $t \rightarrow +\infty$, corresponds to $\tau \rightarrow -0$. The scale factor (56) cannot be created by any realistic source. For large positive time

the source is close to the cosmological term and this can be achieved but in the past when both the energy and pressure density were close to zero the pressure should be negative and much larger by magnitude than energy. The latter cannot be realized by normal physical matter. Still this metric is convenient for the analysis of the quantum fluctuations in the de Sitter spacetime because the equation of motion (6) can be exactly solved. The solution which behaves as $\exp(ik\tau)$ when $\tau \rightarrow -\infty$, has the form:

$$\phi_k = y^{i\kappa} F(\alpha, \beta, \gamma; y), \quad (58)$$

where $\kappa = k/H$, $F(\alpha, \beta, \gamma; y)$ is the hypergeometric function as determined *e.g.* in Ref. [34], $\alpha = -1 + \sqrt{1 - \kappa^2} + i\kappa$, $\beta = -1 - \sqrt{1 - \kappa^2} + i\kappa$, $\gamma = 1 + 2i\kappa$, and $y = \exp(H\tau)$. When τ changes from $-\infty$ to 0 , y changes from 0 to 1 .

In the infrared limit when $k \rightarrow 0$ and y is fixed, the solution is well defined,

$$F \rightarrow 1 + i\kappa(-2y + y^2) \quad (59)$$

in contrast to expression (7) which diverges as $1/k$. One would expect however that this divergence is recovered after a long de Sitter stage when $y \rightarrow 1$ (or $\tau \rightarrow 0$). This particular hypergeometric function may be approximated near $y = 1$ as

$$F \approx \frac{\Gamma(\gamma)}{\Gamma(\gamma - \alpha)\Gamma(\gamma - \beta)} \left[2 + \gamma(y - 1) + \frac{\gamma(\gamma + 1)}{2}(y - 1)^2 \right]. \quad (60)$$

Once again the limit of small k is nonsingular and agrees with expression (59) for $y = 1$. We will observe the trace of the above mentioned infrared singularity in the ultraviolet expansion, $\kappa \rightarrow \infty$, of the solution. The result depends upon the way the limits $k \rightarrow \infty$ and $y \rightarrow 1$ are taken. For the case of $y \rightarrow 1$ first and then $k \rightarrow \infty$, expression (60) is convenient. We find from it (up to y -independent phase factor) that:

$$F \approx \left(1 - \frac{\pi}{4\kappa} + \frac{3 + 2\pi^2}{24\kappa^2} \right) \left[\frac{1}{\kappa} + \left(i + \frac{1}{2\kappa} \right) (y - 1) - \kappa \left(1 - \frac{i}{2\kappa} \right) \left(1 - \frac{i}{\kappa} \right) (y - 1)^2 \right]. \quad (61)$$

The previously found solution for the exact infinitely long de Sitter stage is reproduced by the first two terms in the expansion in $(y - 1)$ in the limit of large k . The infrared singularity at $k = 0$ in the exact de Sitter solution (7) is evidently absent in the distorted metric and appears here only in the ultraviolet expansion. Note that due to the factor in front of the r.h.s. of

Eq. (61) the expansion of $|F^2|$ contains not only even but also odd powers of $1/\kappa$. It would generate new types of divergences in perturbation theory.

All these phenomena may be related to a singular character of the limiting transition from the metric (56) to the exact de Sitter one. Indeed let us introduce a small parameter ε (instead of 1 in Eq. (56)) which characterizes a deviation from the de Sitter metric at a fixed finite time, so the metric has the form:

$$\bar{a}(t) = \varepsilon + \exp(Ht). \quad (62)$$

By shifting the zero of time $t = t' - t_1$ so that $\exp(-Ht_1) = \varepsilon$, we may rewrite $\bar{a}(t)$ as $\bar{a}(t') = \varepsilon[1 + \exp(Ht')]$. Now by rescaling the space coordinates $dx = \varepsilon dx'$ we return to the old metric (56). These transformations correspond to rescaling of the conformal momentum $k \rightarrow k/\varepsilon$. For finite ε this change of normalization is unobservable. But the de Sitter limit, when $\varepsilon \rightarrow 0$, is singular and gives rise to the peculiar behavior mentioned above.

It may be also interesting to get ultraviolet expansion of the solution (58) for a fixed y . To get this expansion it is convenient to present F in the form

$$F = 1 - y + \frac{i}{\kappa} f + \frac{1}{\kappa^2} g. \quad (63)$$

The functions f and g can be found as expansions in inverse powers of κ^2 : $f = f_1 + f_2/\kappa^2 + f_3/\kappa^4 \dots$ (and the similar expression for g) by the perturbative (in $1/\kappa^2$) solution of the equations:

$$(1 - y)f' + f = -\frac{1 + y}{2} + \frac{1 + y}{2\kappa^2} g' + \frac{y(1 - y)}{2\kappa^2} g'', \quad (64)$$

$$(1 - y)g' + g = -\frac{1 + y}{2} f' - \frac{y(1 - y)}{2} f''. \quad (65)$$

The terms which are logarithmically divergent in momentum are determined by the lowest order corrections f_1 and g_1 . The latter are singular, though not too strongly, near $y = 1$ but this divergence is cancelled out, as we see in what follows, in the expression for $\langle \phi^2 \rangle$. The calculation is straightforward and we get:

$$f_1 = -y - \frac{1 - y}{2} \ln(1 - y), \quad (66)$$

$$g_1 = \frac{y}{4} - \frac{y}{2} \ln(1 - y) - \frac{1}{8}(1 - y) \ln^2(1 - y). \quad (67)$$

The singularity in $(1 - y)$ is getting stronger with the rising order in $1/\kappa^2$. For example the most singular parts for f_2 and g_2 are respectively $f_2^{\text{sing}} = 3/8(1 - y)$ and $g_2^{\text{sing}} = -1/4(1 - y)^2$.

Using these expressions we may calculate the vacuum expectation value of ϕ^2 as a function of time. For a pure de Sitter universe, this quantity is known to rise as $\ln \tau = t$ [27, 28, 20]. This result is ascribed to the infrared instability of a minimally coupled massless scalar field. We will calculate the difference $\langle \delta \phi^2 \rangle \equiv \langle \phi^2(\tau) \rangle - \langle \phi^2(-\infty) \rangle$. One may hope that this automatically takes care of infinities in flat spacetime. In the limit of small y one may write keeping only terms of order y^2 :

$$\langle \phi^2(y) \rangle = \int \frac{d^3 k}{(2\pi)^3 2k} \left[(1-y)^2 + \frac{2y}{4\kappa^2 + 1} + \frac{(2\kappa^2 - 1)y^2}{(4\kappa^2 + 1)(\kappa^2 + 1)} \right], \quad (68)$$

where $y = \exp(H\tau)$. One has to subtract from this quantity the value at $y = 0$. The first term is quadratically divergent and to specify its value we have to define the regularization procedure. Usually it is just an arbitrary renormalization constant but in this case it becomes a function of time (or y). To avoid this time dependence one may change the variable of integration, $\kappa = a(\tau)p$ which means transition from the conformal momentum to the physical one. Written in this form the quadratically divergent integral in the expression for $\phi^2(y)$ is formally time independent and is cancelled out by the initial value in the flat spacetime.

There are also logarithmically divergent terms which are absent in the flat case. For small y they can be found from Eq. (68) which is exact with respect to the dependence on κ and correspondingly is not singular at $\kappa = 0$. Calculating the integral we find

$$\langle \delta \phi^2 \rangle = \frac{H^2}{16\pi^2} y(1+y) \ln \kappa_{\max}^2. \quad (69)$$

Here κ_{\max} is (unspecified) ultraviolet cutoff. For arbitrary y we may use Eqs (63), (66), (67) giving expansion in inverse powers of κ . Though the result is formally singular at $\kappa = 0$ we know that the effective infrared cutoff is $\kappa \approx 1$. The logarithmic dependence on $(1-y)$ which we see in f_1 and g_1 is cancelled out and we get exactly the same result (69). The logarithmically divergent terms are explicitly y -dependent and this dependence cannot be cancelled by a redefinition of the ultraviolet cutoff as it was done above for the quadratically divergent part. If we assume that there is a physical cutoff realized, say, by some massive particles, we would expect the cutoff in conformal momentum to be proportional to the scale factor $k = ma(\tau)$. This is because mass enters equations of motion in conformal coordinates always in the combination $ma(t)$. In this case we get the result $\langle \delta \phi^2 \rangle \sim y(1+y) \ln(1-y)$. At the de Sitter stage when $y \rightarrow 1$, we get the well-known result $\langle \delta \phi^2 \rangle \sim t$ plus a logarithmically divergent term. Note that the coefficient in front of the log is exactly $R = 6y(1+y)H^2$. Expressed in terms of R the combination $(1-y)$ looks rather unusual: $(1-y) = \sqrt{1 + 2R/3H^2} - 3$.

Let us turn now to the calculation of the energy-momentum tensor of the field ϕ in background (1). Though it is a more UV divergent quantity than ϕ^2 , it has the nice property of covariant conservation which helps greatly in a proper definition of the divergent parts. Let us start from a more simple example when the background metric is flat in the infinite past and future and changes as a function of time in between. Assume that the scale factor $a(-\infty) = 1$ and $a(+\infty) = a_f$. The mode decomposition of ϕ has the standard form (8) with $\phi_k(\tau) = \exp(-i\omega\tau)$ at negative time infinity and $\phi_k(\tau) = [\alpha_k \exp(-i\omega\tau) + \beta_k \exp(i\omega\tau)]a_f$ at positive time infinity. The Bogoliubov coefficients α_k and β_k satisfy the relation Eq. (18). We will use the Heisenberg representation and will calculate the variation with time of the vacuum energy density. One has in conformal coordinates

$$T_{\tau\tau} = \frac{1}{2}[(\partial_\tau\phi)^2 + (\partial_j\phi)^2]. \quad (70)$$

The physical energy density ρ is expressed through it as $\rho \equiv T_{tt} = T_{\tau\tau}/a^2$. We will consider the expectation value $\langle\rho(t)\rangle$ averaged over the initial vacuum state. It is an infinite quartically divergent quantity, but the difference $\delta\rho = \langle\rho(+\infty) - \rho(-\infty)\rangle$ is finite and is equal to

$$\delta\rho \equiv \langle\rho(+\infty) - \rho(-\infty)\rangle = \frac{1}{a_f^4} \int \frac{2|\beta_k|^2\omega^2}{2\omega(2\pi)^3} d^3k. \quad (71)$$

It is the energy density of the particles produced by the time varying gravitational field. In this description we do not need a precise meaning of "particle", the only essential quantity is the energy density. One can easily see that the pressure density of the final state is $\delta p = \delta\rho/3$. We use this simple example to illustrate that the finite part of the energy-momentum tensor averaged over the in-vacuum state is not proportional to the metric and, for the finite parts, nonlocal effects are essential. Note also that the finite time difference $\delta\rho(t) = \langle\rho(t) - \rho(-\infty)\rangle$ is quadratically divergent, and one needs curvature-dependent counterterms to eliminate these infinities.

The situation is more complicated for the metric (56) when the space-time is not flat in the infinite future. Still we may use the similar approach to calculate the evolution of the energy density of the field ϕ starting with the initial vacuum value which we assume to be zero. The average value of the energy density over the initial vacuum state is

$$\rho = \frac{1}{a^2(t)} \int \frac{d^3k}{(2\pi)^3} \frac{\omega}{2} \left[|F^2| + \frac{Hy}{k} \text{Im}(F^* F') + \frac{H^2 y^2}{k^2} |F'|^2 \right]. \quad (72)$$

Here $\rho = T_{tt}$ is the energy density in the physical frame. Substituting expressions (63), (66), (67) into this equation we get

$$\rho = \frac{1}{a^2} \int \frac{d^3k}{(2\pi)^3} \frac{\omega}{2} \left[(1-y)^2 + \frac{H^2 y^2}{k^2} + \frac{H^4}{2k^4} K(y) \right]. \quad (73)$$

The first two terms, quartically and quadratically divergent respectively, may be eliminated by renormalization (see discussion in Sec. 3). The renormalized quantities are at worst constants so their contribution into ρ is not rising with time. The last term is logarithmically divergent, but nonsingular at $y = 1$. The function $K(y)$ might have in principle terms behaving like $1/(1 - y)$ or $\ln^2(1 - y)$, etc., as one can see from Eqs (66), (67) but all they have cancelled out. Hence the energy density of a massless minimally coupled scalar field in de Sitter background does not rise with time but goes down at least as $1/a^2$ and its back reaction on metric is not essential. Higher loop contributions probably will not change this conclusion, but this remains an open question.

6. Conclusions

In this paper, we have discussed how the history of the de Sitter phase affects the vacuum and removes the IR divergences of fluctuations in a de Sitter background. As previous workers have remarked, there is no solution without breaking de Sitter invariance. Through specific examples, we have related the IR cutoff to the breaking of de Sitter invariance in the past, a cosmologically well-motivated assumption.

We have offered several approaches that may be useful in different contexts. In the first, we suggested a treatment of the nonadiabatic modes that differs from the conventional one. The single new parameter is the initial time τ_0 before which the vacuum energy is presumed not to be dominant. In a manner analogous to the classic Lee-Nauenberg discussion of mass singularities, [26] we argued that one must diagonalize the Hamiltonian before embarking on a perturbation expansion. We found this to be an economical formalism, in which the correlation functions differ significantly only at distances large compared to the distance that a signal might have traveled since the initial time τ_0 . We showed that the correlation functions satisfy cluster decomposition, so that this is an acceptable framework for discussing quantum corrections.

In the second approach, we explicitly modified the metric at early times, in such a way that, in the distant past, the space is Minkowski spacetime where the vacuum is well-understood, and we explored the consequences at later times. While this cannot be solved analytically, a WKB solution provided insight into the nature of the infrared cutoff. In the third approach, we simply modified the scale factor in a simple manner, again so the past is Minkowskian, but in a way that the equations of motion can be solved analytically exactly.

Each approach has its strengths and weaknesses, but the first involving pairing has the advantage that it presumes little about the universe prior to

the de Sitter phase, and the dynamical equations remain unchanged, so that their symmetry properties remain exact. Whether the transition between nonadiabatic and adiabatic modes can be refined in some general way, for example, through a path integral approach⁹, is not clear. Whether, in the case of gravity, this construction requires modification to take into account interactions remains to be investigated. While a Bogoliubov transformation may not be sufficient, there is a related approach that may well be adequate. As in electrodynamics or Yang-Mills theory, [36] the true in- or out-states may be regarded as coherent states built up from creation operators applied the de Sitter invariant vacuum defined in Eq. [Ref. 10]. This provides the possibility for a general treatment of nonadiabatic modes in the presence of interactions based on the Hamiltonian.

The second approach in which the metric was explicitly modified may be useful when we know the universe is approximately Minkowski spacetime in the distant past. It also illustrates how the WKB method may provide insight into the nature of the IR cutoff in situations where the nonadiabatic modes cannot be so simply treated.

We emphasize that the considerations of this paper do not resolve the issue of whether there might be long-time singularities that destabilize the de Sitter metric [6, 7, 8]. It remains to be seen what consequences, if any, our results have on interacting field theories such as those considered by Ford [4] in which the behavior of ϕ^2 directly influences the stress tensor.

A. Dolgov is grateful to the Department of Physics of the University of Michigan for the hospitality during the period when this work was done. This work was supported in part by the U.S. Department of Energy.

REFERENCES

- [1] S. Weinberg, *Rev. Mod. Phys.* **61**, 1 (1989).
- [2] A. D. Dolgov, Proc. Rencontre de Moriond, Les Arcs, France, (1989), p. 227.
- [3] A. Linde, *Particle Physics and Inflationary Cosmology*, Harwood Academic Publishers, Philadelphia 1990.
- [4] L.H. Ford, *Phys. Rev.* **D31**, 710 (1985).
- [5] E.G. Floratos, J. Iliopoulos, T.N. Tomaras, *Phys. Lett.* **B197**, 373 (1987).
- [6] N.C. Tsamis, R.P. Woodard, *Phys. Lett.* **B301**, 351 (1993).
- [7] N.C. Tsamis, R.P. Woodard, UFIFT-HEP-93-17.
- [8] N.C. Tsamis, R.P. Woodard, UFIFT-HEP-92-24 (revised February, 1994).
- [9] E.M. Lifshitz, *Zh. Teor. Exper. Fiz.* **16**, 587 (1946).
- [10] L.P. Grishchuk, *Zh. Teor. Exper. Fiz.* **67**, 825 (1974).

⁹ See, e.g., Ref. [35].

- [11] L.H. Ford, L. Parker, *Phys. Rev.* **D16**, 245 (1977).
- [12] B. Allen, *Phys. Rev.* **D32**, 3136 (1985).
- [13] B. Allen, M. Tureyn, *Nucl. Phys.* **B292**, 813 (1987).
- [14] A. Higuchi, *Clas. Quantum Grav.* **8**, 1961 (1991).
- [15] G. Kleppe, *Phys. Lett.* **B317** 305 (1993).
- [16] A.D. Dolgov, M.B. Einhorn, V.I. Zakharov, UM-TH-94-11, March, 1994 (gr-qc/9403056).
- [17] B. Allen, A. Folacci, *Phys. Rev.* **D35**, 3771 (1987).
- [18] B.S. DeWitt, in *Les Houches 1983*, eds. C. DeWitt and B. S. DeWitt, Gordon and Breach, New York 1983.
- [19] S.A. Fulling, M. Sweeny, R.M. Wald, *Comm. Math. Phys.* **63**, 257 (1978).
- [20] A. Vilenkin, L. Ford, *Phys. Rev.* **D26**, 1231 (1982).
- [21] J.R. Schrieffer, *Theory of Superconductivity*, 1st revised ed. by J.R. Schrieffer, Benjamin, New York 1983; G. Rickayzen, *Green's Functions and Condensed Matter*, Academic Press Ltd., London 1980.
- [22] N.N. Bogoliubov, *J. Phys.* **9**, 23 (1947); *Vestn. MGU* **7**, 43 (1947); *Zh. Teor. Exper. Fiz.* **34**, 58 (1958); *Nuovo Cimento* **7**, 794 (1958).
- [23] S. Fulling, *Gen. Relativ. Gravit.* **10**, 807 (1979).
- [24] N.D. Birell, P.C.W. Davies, *Quantum Fields in Curved Space*, Cambridge University Press, Cambridge 1982.
- [25] L. Parker, *Phys. Rev.* **183**, 1057 (1969); see also his lectures in F.P. Esposito, L. Witten (eds.), *Symposium on Asymptotic Structure of Space-Time*, Plenum Press, New York 1976.
- [26] T.D. Lee, M. Nauenberg, *Phys. Rev.* **133**, 1549 (1964).
- [27] A.D. Linde, *Phys. Lett.* **116B**, 335 (1982).
- [28] A.A. Starobinsky, *Phys. Lett.* **117B**, 175 (1982).
- [29] B. Ratra, *Phys. Rev.* **D31**, 1931 (1985).
- [30] N.C. Tsamis, R.P. Woodard, UFIFT-HEP-92-14 (to appear in *Comm. Math. Phys.*).
- [31] M.J. Duff, Texas A&M preprint CTP-TAMU-06/93 (hep-th/9308075.)
- [32] T.S. Bunch, L. Parker, *Phys. Rev.* **D20**, 2499 (1979).
- [33] T.S. Bunch, P. Panagaden, L. Parker, *J. Phys.* **A 13**, 901 (1981).
- [34] I.S. Gradshteyn, I.M. Ryzhik, *Tables of Integrals, Series, and Products*, Academic Press, New York 1965.
- [35] D.M. Chitre, J.B. Hartle, *Phys. Rev.* **D16**, 251 (1977).
- [36] This is reviewed, for example, in H. F. Contopanagos and M. B. Einhorn, *Phys. Rev.* **D45**, 1291 (1992).