Short distance and initial state effects in inflation: Stress tensor and decoherence

Paul R. Anderson*

Department of Physics, Wake Forest University, Winston-Salem, North Carolina, 27109, USA

Carmen Molina-París†

*Department of Applied Mathematics, University of Leeds, Leeds LS2 9JT, United Kingdom,
and Departamento de Matemáticas, Física Aplicada y Físico-química, Facultad de Farmacia,
Universidad San Pablo CEU, E-28660 Madrid, Spain*

Emil Mottola‡

Theoretical Division, T-8, Los Alamos National Laboratory, Los Alamos, New Mexico, 87545, USA

We present a consistent low energy effective field theory framework for parametrizing the effects of novel short distance physics in inflation, and their possible observational signatures in the cosmic microwave background. We consider the class of general homogeneous, isotropic initial states for quantum scalar fields in Robertson-Walker (RW) spacetimes, subject to the requirement that their ultraviolet behavior be consistent with renormalizability of the covariantly conserved stress tensor which couples to gravity. In the functional Schrödinger picture such states are coherent, squeezed, mixed states characterized by a Gaussian density matrix. This Gaussian has parameters which approach those of the adiabatic vacuum at large wave number, and evolve in time according to an effective classical Hamiltonian. The one complex parameter family of α squeezed states in de Sitter spacetime does not fall into this UV allowed class, except for the special value of the parameter corresponding to the Bunch-Davies state. We determine the finite contributions to the inflationary power spectrum and stress tensor expectation value of general UV allowed adiabatic states, and obtain quantitative limits on the observability and backreaction effects of some recently proposed models of short distance modifications of the initial state of inflation. For all UV allowed states, the second order adiabatic basis provides a good description of particles created in the expanding RW universe. Because of the absence of particle creation for the massless, minimally coupled scalar field in de Sitter space, there is no phase decoherence in the simplest free field inflationary models. We apply adiabatic regularization to the renormalization of the decoherence functional in cosmology to corroborate this result.

I. INTRODUCTION AND OVERVIEW

Inflationary models were introduced principally to account for the observed large scale homogeneity, isotropy and flatness of the Universe in a causal way, independently of detailed initial conditions [1]. Because of the exponential expansion of an initially small causal patch, the inflationary de Sitter epoch dominated by the vacuum equation of state $p = -\varepsilon$ suppresses any classical inhomogeneities in the initial conditions by many orders of magnitude, and leads to a primordial power spectrum that is both scale invariant and featureless. Most inflationary models assume that the quantum fluctuations which lead to this scale invariant spectrum originate in the maximally $O(4, 1)$ symmetric Bunch-Davies state of scalar fields in de Sitter spacetime [2], although the possibility that other states may play a role was considered by some authors [3–5]. In the last few years there has been a renewed interest in the possible effects of different initial states in inflation [6–20], fueled largely by the speculation that more precise observations of the cosmic microwave background (CMB) might make such effects observable, thus opening up the possibility of using CMB observations to probe novel short distance physics in the very early Universe.

The primary purpose of this paper is to present a consistent low energy effective field theory (EFT) framework for parametrizing such short distance and initial state effects in cosmological spacetimes. Although the elements of quantum field theory in curved space upon which this EFT framework relies have been known for some time, a comprehensive treatment of general homogeneous, isotropic initial states in Robertson-Walker spacetimes has not been given previously to our knowledge. Such a treatment of general initial states requires both covariant and canonical methods, as well as a dictionary to translate between them. Establishing the EFT framework, the relationship between the covariant and canonical approaches, and the form of the state-dependent terms in the covariant stress tensor occupies Secs. II, III, and IV of the paper. The paper is designed so that after becoming acquainted with the definitions and conventions in Sec. II, the reader may skip the detailed development of Secs. III and IV if desired,

*Electronic address: anderson@wfu.edu
†Electronic address: carmen@maths.leeds.ac.uk
‡Electronic address: emil@lanl.gov

and go directly to the applications in later sections, referring back to the previous sections for the derivation of the formulas as necessary. Readers interested only, in particular, modifications of the initial state of inflation and their effects on the CMB may wish to skip directly to Sec. V.

The secondary purpose of this paper is to apply the EFT methods developed in Secs. II, III, and IV to the processes of particle creation and decoherence in semiclassical cosmology. Certain problems with the definitions of particle number and the decoherence functional are resolved by the same adiabatic methods used to define the general class of UV allowed initial states in the EFT approach. Readers interested primarily in these applications may wish to go directly to Secs. VI or VII respectively, likewise referring to the earlier sections as needed. The remainder of this introduction contains a general overview of the issues addressed in the paper and our approach to them. As a further guide to the content of the paper, the concluding Sec. VIII contains a point-by-point summary of our main results.

As we consider initial state modifications of inflation it is perhaps worth emphasizing from the outset that any sensitivity of present day observations to initial state effects runs counter to some of the original motivations for and attractiveness of inflation. A scale invariant spectrum is one of inflation's most generic predictions, precisely because of the presumed late time *insensitivity* to perturbations of the initial state. If there are features in the power spectrum of the CMB today which are not erased by the exponential redshift of the inflationary epoch and which bear the imprint of new physics at short distance scales, then one might ask what prevents short distance physics from affecting other large scale properties of the Universe, such as its homogeneity, isotropy or flatness. Further, since it is not clear which inflationary model (if any) is correct, fine-tuning a specific model to make particular modifications observable in the CMB power spectrum results in a diminishing of the overall predictive power of inflation. As long as it is possible to accommodate any observable features in the power spectrum by appropriately fine-tuning the inflationary model, the physical origin of these features as true signatures of new high energy physics must remain uncertain [20]. Finally, the remarkable detection of a nonzero cosmological dark energy in the Universe today [21], at a level very different from estimates based on considerations of "naturalness" from short distance physics, should caution that present cosmological models are as yet far from complete, and the connection between microphysics and macroscopic structure in the Universe is still to be elucidated.

Despite these fine-tuning and naturalness problems, it is nevertheless true that in almost any given inflationary model there can be surviving initial state effects in the primordial power spectrum at *some* level, and the advent of more precise CMB data makes quantifying the sensitivity

of inflationary models to such initial state effects potentially worthwhile.

A quantitative treatment of short distance and initial state effects in inflation requires a *predictive, low energy* framework in which such effects can be parametrized and studied with a minimum of assumptions about the unknown physics at ultrashort distances. Effective field theory provides exactly that framework in other contexts, and we assume in this paper that EFT methods may be applied in gravity and cosmology as well. The EFT approach to perturbative gravitational scattering amplitudes was discussed in Ref. [22]. Here we extend EFT methods to the nonperturbative regime of semiclassical cosmology.

Since a fully predictive quantum theory of gravity is still lacking, we are virtually compelled to adopt an EFT approach to cosmology. Since all scales are presumed to be redshifted to lower energy scales where an EFT description eventually becomes applicable, the EFT appropriate for cosmology which respects general coordinate invariance and the equivalence principle is the Einstein theory together with its quantum corrections. In the EFT framework of semiclassical gravity we can study general (i.e., scale noninvariant) initial states in Robertson-Walker (RW) spacetimes in a well-defined way, without detailed knowledge of the short distance physics which may have generated them. Although more general initial states and more general matter EFT interactions could be considered, we focus in this paper on the specific gravitational effects of quantum matter fields, and restrict ourselves for simplicity to free scalar fields in spatially homogeneous and isotropic initial states consistent with the symmetry of the RW geometry.

The redshifting of short distance scales to larger ones as the Universe expands distinguishes semiclassical gravity from other effective field theories, which possess a *fixed* physical cutoff. In the case of a fixed cutoff, the shorter distance modes of the effective theory can be excluded from consideration, and their effects subsumed into a finite number of parameters of the low energy description. In practice, absorption of the cutoff dependence of quantum corrections into a finite set of parameters of the effective action is no different in an EFT from that in a renormalizable theory, except for the allowance of higher dimensional interactions and the corresponding parameters which are suppressed by the cutoff scale in the EFT description.

Implicit in the EFT framework is the assumption that the effects of short distance degrees of freedom *decouple* from the long distance ones. However, in an expanding Universe decoupling is a delicate matter. New short distance modes are continually coming within the purview of the low energy description, and some additional information is required to handle these short distance degrees of freedom as they newly appear. If these ultraviolet (UV) modes carry energy-momentum, as they should when their wavelength becomes larger than the short distance cutoff and EFT

methods apply, then simply excluding their contribution to the energy-momentum tensor at earlier times will lead to violations of energy conservation, a well-known point that has been emphasized anew in Ref. [13]. Energy nonconservation occurs with a physical momentum cutoff because energy is being supplied (by an unspecified external mechanism) to the new degrees of freedom as they redshift below the cutoff, although they carried no energy-momentum formerly. This is an essential point: arbitrary short distance modifications that violate energy conservation are unacceptable in a low energy EFT respecting general coordinate invariance, since the resulting energy-momentum tensor $\langle T_{ab} \rangle$ cannot be a consistent source for the semiclassical Einstein equations at *large* distances. Upsetting the macroscopic energy conservation law by a short distance physical cutoff affects the cosmological evolution at all scales, hence violating decoupling as well.

These considerations show that the necessary existence of a conserved source for the semiclassical Einstein equations provides an important constraint on the class of possible short distance modifications of the initial state of inflation, quite independent of the matter field content and its EFT. Let us emphasize that there is no problem modifying the initial state of a quantum field at a *fixed* time t_0 for momenta below some physical scale M at that time. However the quantum state is not completely specified and a conserved energy-momentum tensor cannot be computed unambiguously, until information is given also for physical momenta initially much greater than M, which will redshift below M at later times $t > t_0$. Without some prescription consistent with general coordinate invariance for dealing with these arbitrarily high energy "trans-Planckian" modes, which will become physical low energy modes eventually, the low energy effective theory of semiclassical gravity is not complete or predictive.

General covariance of the low energy effective theory of gravity coupled to quantum matter is the key technical assumption which we make in this paper. General coordinate invariance determines the form of the effective action, and therefore the counterterms which are available to absorb the ultraviolet divergences of the energy-momentum tensor of the quantum matter fluctuations. The renormalization of $\langle T_{ab} \rangle$ with the standard local covariant counterterms up to dimension four is possible if and only if the short distance properties of the vacuum fluctuations are the standard ones, as expressed, for example, in the Hadamard conditions on the two-point function $\langle \Phi(x) \Phi(x') \rangle$ as $x \to x'$ [23,24]. These UV conditions on the structure of the vacuum should be viewed as an extension of the equivalence principle to semiclassical gravity, since they amount to assuming that the local behavior of quantum matter fluctuations are determined in a curved spacetime by those of flat spacetime and small (calculable) deviations therefrom. With this physical assumption about the local properties of the vacuum, the ultrashort distance

modes are necessarily adiabatic vacuum modes and are fully specified as they redshift below the UV cutoff scale. Then covariant energy conservation is ensured, there are no state-dependent divergences in $\langle T_{ab} \rangle$, and the effective field theory of semiclassical gravity applied to cosmology becomes well-defined and predictive within its domain of validity.

Several authors have considered specific short distance modifications, such as modified dispersion relations for the modes [11] or spacetime noncommutativity [12]. We do not consider in this paper these or any other possible specific short distance modifications that would take us out of the framework of the covariant low energy EFT of gravity, without providing a completely consistent quantum alternative. Once the low energy EFT of semiclassical gravity applies, any imprint of UV physics can be encoded only in the parameters of the initial state up to some large but finite energy scale M. While many papers have discussed possible imprints of new short distance physics on the CMB power spectrum, a few authors have considered also the constraints that may arise from the energy-momentum tensor of the fields in a state other than the Bunch-Davies (BD) state, making use of various order of magnitude estimates [9,14,16–18,20]. If the initial state modifications are parametrized by adding irrelevant higher dimensional operators of the scalar EFT at the boundary, there is apparent disagreement between several authors about the order of the corrections these modifications induce in the energy-momentum tensor [16,17]. Quantitative control of the finite state-dependent terms in the energy-momentum tensor is potentially important for determining whether the short distance modifications can be observable in the power spectrum without upsetting other features of inflation. If the scalar field energy density is too large, it could prevent an inflationary phase from occurring at all [9,13,20]. In this paper we present the framework necessary for the unambiguous evaluation of initial state effects in both the power spectrum and the renormalized $\langle T_{ab} \rangle$ for any homogeneous and isotropic state. We then use the general framework to find specific constraints on initial states, such as those proposed in the boundary action formalism of Ref. [25]. We find by explicit computation that the terms in the matter EFT boundary action linear in the first higher dimensional operators that one can add, do enter the energy-momentum tensor and may place more restrictive bounds on the parameters than reported in [17].

A systematic study of more general initial states in RW spacetimes and their stress tensor expectation values was initiated in two recent papers [26,27]. Although it had generally been assumed that initial state inhomogeneities in $\langle T_{ab} \rangle$, different from the BD expectation value, would redshift on an expansion time scale H^{-1}, the proper treatment of the initially ultrahigh frequency (trans-Planckian) modes is critical to proving this result, and also for dem-

onstrating how it may break down in certain special cases. The conditions on homogeneous and isotropic initial states of a free scalar field in de Sitter space, necessary for its two-point function and energy-momentum tensor to be both finite in the IR and renormalizable in the UV were defined in Ref. [27]. When these conditions are satisfied it was shown that $\langle T_{ab} \rangle$ for a free scalar field of mass m and curvature coupling ξ does approach the BD value with corrections that decay as $a_{\mathrm{dS}}^{-3+2\,\mathrm{Re}\,\nu}$ for $\mathrm{Re}\,\nu < 3/2$, where a_{dS} is the RW scale factor in de Sitter space, given by Eq. (5.1) and ν is defined by Eq. (5.2) below. For sufficiently massive fields ν is purely imaginary and these fields have energy-momentum tensor expectation values which decay to the BD value as a_{dS}^{-3}, just as would be expected for classical nonrelativistic dust with negligible pressure. However, the light or massless cases in which $\nu^2 > 0$, show a quite different late time behavior. The massless conformally coupled scalar field, for which $\nu = \pm 1/2$, has in addition to the expected subdominant a_{dS}^{-4} behavior of classical massless radiation with $p = \varepsilon/3$, a much more slowly falling a_{dS}^{-2} component, with $p = -\varepsilon/3$, arising from quantum squeezed state effects. In the massless, minimally coupled case, relevant for the slowly rolling inflaton field, as well as the graviton itself, $\nu = 3/2$, and there is an additional a_{dS}^0 constant coherent component in the late time behavior, with $p = -\varepsilon$, signaling the breakdown of the $O(4, 1)$ invariant BD state in the IR, and the change in the stress tensor from the BD to the Allen-Folacci (AF) value [28]. These features could not be so readily anticipated by purely classical considerations, but are quite straightforward to obtain with the general properly renormalized $\langle T_{ab} \rangle$.

We present here a comprehensive treatment of general initial states in arbitrary RW spacetimes begun in Refs. [26,27], from both a canonical and covariant viewpoint. We emphasize throughout the paper that in a general RW spacetime all such states are on an equal footing *a priori*. There is no need to resolve the vacuum "ambiguity" often said to exist in curved space field theory. Different physical initial data will simply select different physical states. Only the local behavior of these states at very short distances needs to be restricted by the equivalence principle. Because the structure of the vacuum is most explicit in its wave functional representation, we review the Schrödinger description of arbitrary RW states first. The Schrödinger description in RW spacetimes has previously been investigated for pure states in [29,30] and for mixed states in [31]. The general state of a free scalar field is that of a mixed, squeezed state Gaussian density matrix $\hat{\rho}$, given by Eq. (3.10), evolving according to the quantum Liouville equation (3.29) in the Schrödinger representation. The parameters that specify this general Gaussian density matrix are in one-to-one correspondence with the amplitudes that define the two-point Wightman function and power spectrum of the field. We show that the

quantum Liouville equation or the scalar wave equation in the covariant description imply that these Gaussian parameters evolve with time according to an effective classical Hamiltonian, given by Eqs. (A10)–(A12), in which $\hbar$ appears as a parameter. This demonstrates that the evolution of an arbitrary initial state is completely unitary and time reversible in any RW spacetime. Any apparent discrepancy between the Hamiltonian and covariant approaches is resolved by including the RW scale factor on an equal footing with the matter field(s) in the Hamiltonian description. The correct source for Einstein's equations and backreaction considerations is not the canonical or effective Hamiltonian of the Schrödinger representation but the expectation value of the covariant energy-momentum tensor, $\langle T_{ab} \rangle$.

The fourth order adiabaticity condition [32] on the short distance components of the wave functional defines the class of UV allowed RW states consistent with the low energy effective field theory satisfying general covariance. We exhibit the finite state-dependent contributions to $\langle T_{ab} \rangle$ for a general homogeneous, isotropic RW state in (4.16). The $O(4, 1)$ invariant BD state is a UV allowed fourth order adiabatic vacuum state in de Sitter space, but the one complex parameter family of squeezed state generalizations of the BD state [4,5] (sometimes called α vacua) are *not* UV allowed RW states [33]. All such states except the BD state are therefore excluded as possible modified initial states in the low energy description, unless they are cut off at some physical momentum scale M [8,13], and are then no longer de Sitter invariant. Various possible modifications of the inflaton initial state up to some physical scale M at the initial time t_0 are considered in Sec. V, and their power spectrum and backreaction effects are computed in a consistent way. We compare our treatment of short distance and initial state effects with previous work involving α vacua [8], adiabatic states [7], and a boundary action approach [17,25]. The precise connection between the boundary action approach and the initial state specification is established. Readers interested in only these initial state effects in inflation may wish to skip directly to Sec. V, where these short distance effects are considered, and bounds on the short distance modifications are obtained from backreaction considerations.

In addition to the power spectrum and energy-momentum source for the gravitational field, the adiabatic method provides a consistent framework to discuss particle creation and decoherence in semiclassical cosmology. Although the definition of particle number in an expanding Universe is inherently nonunique, we show that the adiabatic number basis matched to the second order adiabatic terms in $\langle T_{ab} \rangle$ is the minimum one that allows for a finite total particle number with conserved energy-momentum. Matching the particle basis to fourth or higher adiabatic order is possible but of decreasing physical and practical importance. The EFT description also suggests that one

should limit the particle number definition to the minimal one that requires the fewest number of derivatives of the scale factor, i.e., *two*, which are sufficient to eliminate all power law cutoff dependences in the stress tensor. Hence the second order adiabatic basis is selected by the short distance covariance properties of the vacuum, together with the local derivative expansion characteristic of a low energy EFT description of Einstein's equations, which are themselves second order in derivatives of the metric. As is well known, in the case of the massless, conformally coupled scalar field in an *arbitrary* RW spacetime the zeroth order adiabatic vacuum modes become *exact* solutions of the wave equation, and hence no mixing between positive and negative frequency modes occurs. We show that no particle creation occurs also for the massless, minimally coupled scalar field (which sometimes serves as the inflaton field) in the special case of de Sitter spacetime.

Particle creation may be described as a squeezing of the density matrix parameters and corresponds to a basis in which the off-diagonal elements of $\hat{\rho}$ are rapidly oscillating in phase, and may be replaced by zero with a high degree of accuracy. To the extent that this approximation is valid and the information contained in these rapidly oscillating phases cannot be recovered, the evolution is *effectively* dissipative at a macroscopic level, despite being microscopically time reversible. The macroscopic irreversibility is measured by the von Neumann entropy (6.10) of the phase averaged density matrix in the adiabatic particle basis. In the two special massless cases of cosmological interest mentioned above, namely, the conformally coupled, scalar field in an arbitrary RW spacetime and the minimally coupled scalar field in de Sitter spacetime, this phase averaging effect is absent, since no particle creation occurs.

The decoherence functional is defined in the Schrödinger representation as the wave function overlap between two states with similar initial conditions but different macroscopic RW scale factors, i.e., it measures the quantum (de)coherence between different semiclassical realizations of the Universe [34]. Physical expectations of a very nearly classical Universe suggest that this quantum overlap between different macroscopic states in cosmology should be finite in principle but extremely small. However, a naive computation of the decoherence functional in RW cosmology is plagued by divergences, qualitatively similar to those encountered in $\langle T_{ab} \rangle$. Moreover, previous authors have found that the exact form and degree of these divergences depend upon the parametrization used for the matter field variables [34–36]. These divergences and ambiguities have prevented up until now the straightforward application and physical interpretation of the decoherence functional in semiclassical cosmology. By analyzing the general form of the divergences in the decoherence functional and relating them to the divergences in

the effective closed time path (CTP) action of semiclassical gravity [37], we show that these divergences can be regulated and removed by a slightly modified form of the adiabatic subtraction procedure [38] used to define both the renormalized $\langle T_{ab} \rangle$ and the finite particle number basis. This gives an unambiguous definition of a physical, UV *finite* decoherence functional for RW spacetimes which is displayed in Eq. (7.13) and which is free of field parametrization dependence and other ambiguities previously noted in the literature. The finite decoherence functional does fall rapidly to zero with time in the general case in which particle creation takes place, in accordance with physical intuition. In the special massless cases in which no particle creation occurs, the renormalized decoherence functional *vanishes*, showing that no decoherence of quantum fluctuations between different semiclassical RW universes occurs in these cases. This corroborates the close connection between the particle creation and decoherence effects which has been found in other contexts [39], and shows that the emergence of a classical Universe from initial conditions on a massless field must be due to other effects, such as interactions, which are neglected in the free field treatment presented in this paper.

The outline of the paper is as follows. In the next section we establish notation and define the general class of homogeneous, isotropic RW states in a RW spacetime. In Sec. III, we review the Hamiltonian description of the evolution of these states and give the form of the mixed state Gaussian density matrix of the Schrödinger representation, as well as the Wigner function and effective classical Hamiltonian which describes the evolution. In Sec. IV we evaluate the expectation value of the energy-momentum tensor, and the low energy effective action for gravity of which it is part. We specify the conditions on the short distance components of a general RW state in order for $\langle T_{ab} \rangle$ to be UV renormalizable with geometric counterterms of the same form as the effective action, and obtain an expression for the finite contributions of arbitrary UV allowed RW states. In Sec. V we consider three types of modified initial states in inflation, evaluating the power spectrum and energy-momentum tensor for each in turn. In Sec. VI we define the adiabatic particle number basis and show how particle creation leads to an effective dissipation in the density matrix description. In Sec. VII we define a finite renormalized decoherence functional for semiclassical cosmology, and corroborate the nondecoherence of massless inflaton fluctuations in de Sitter space. We conclude with a detailed summary and discussion of our results. Technical details of the Gaussian parametrization of the density matrix and its properties, the evaluation of integrals needed in Sec. V and the comparison of adiabatic bases used in squeezing and decoherence calculations by previous authors are relegated to Appendices A, B, and C respectively. Throughout we set $c = 1$ and use the metric and curvature conventions of Misner, Thorne, and Wheeler [40].

II. GENERAL RW INITIAL STATES

Homogeneous and isotropic RW spacetimes can be described by the line element,

$$ds^2 = -dt^2 + a^2(t)d\Sigma^2 = -dt^2 + a^2(t)\gamma_{ij}dx^i dx^j, \quad (2.1)$$

with t the comoving (or cosmic) time, and γ_{ij} the metric of the three-dimensional spacelike sections Σ of constant spatial curvature, which may be open, flat, or closed. It is also useful to introduce the conformal time coordinate,

$$\eta = \int^t \frac{dt}{a(t)}, \quad (2.2)$$

so that the line element (2.1) may be expressed in the alternative form,

$$ds^2 = a^2(\eta)(-d\eta^2 + d\Sigma^2), \quad (2.3)$$

where a is now viewed as a function of conformal time η. We take a to have dimensions of length with η dimensionless. The scalar curvature is

$$R = 6\left(\dot{H} + 2H^2 + \frac{\epsilon}{a^2}\right), \qquad H \equiv \frac{\dot{a}}{a}. \quad (2.4)$$

The overdot denotes differentiation with respect to t and $\epsilon = -1, 0, +1$ depending on whether the spatial sections are open, flat, or closed, respectively.

A free scalar field with mass m obeys the scalar wave equation,

$$(-\Box + m^2 + \xi R)\Phi = 0, \quad (2.5)$$

where $\Box = g^{ab}\nabla_a\nabla_b$ and ξ is the arbitrary dimensionless coupling to the scalar curvature. Since the RW three-geometry Σ is spatially homogeneous and isotropic, the wave equation (2.5) may be solved by decomposing $\Phi(t, \mathbf{x})$ into Fourier modes in the general form,

$$\Phi(t, \mathbf{x}) = \int[d\mathbf{k}](a_{\mathbf{k}}\phi_k(t)Y_{\mathbf{k}}(\mathbf{x}) + a_{\mathbf{k}}^{\dagger}\phi_k^*(t)Y_{\mathbf{k}}^*(\mathbf{x})). \quad (2.6)$$

The $Y_{\mathbf{k}}$ are the eigenfunctions of the three-dimensional Laplace-Beltrami operator Δ_3 on Σ, satisfying

$$-\Delta_3 Y_{\mathbf{k}}(\mathbf{x}) \equiv -\frac{1}{\sqrt{\gamma}}\frac{\partial}{\partial x^i}\gamma^{ij}\sqrt{\gamma}\frac{\partial}{\partial x^j}Y_{\mathbf{k}}(\mathbf{x})$$

$$= (k^2 - \epsilon)Y_{\mathbf{k}}(\mathbf{x}), \quad (2.7)$$

and the $\phi_k(t)$ are functions only of time and the magnitude of the wave vector, $k \equiv |\mathbf{k}|$.

For flat spatial sections, $\epsilon = 0$, $\gamma_{ij} = \delta_{ij}$, $\gamma \equiv \det\gamma_{ij} = 1$, and the $Y_{\mathbf{k}}(\mathbf{x})$ are simply plane waves $e^{i\mathbf{k}\cdot\mathbf{x}}$. The integration measure in Eq. (2.6) for this case is $\int[d\mathbf{k}] = \int d^3\mathbf{k}/(2\pi)^3$.

In the case of compact spatial sections, $\epsilon = +1$, the wave number k takes on discrete values which we label by the positive integers $k \geq 1$, and the harmonic functions

in Eq. (2.7) are the spherical harmonics of the sphere S^3. These S^3 harmonics denoted by Y_{klm} depend on three integers $\mathbf{k} \leftrightarrow (k, \ell, m)$, the first of which may be identified with $|\mathbf{k}|$, while (ℓ, m) refer to the usual spherical harmonics on S^2 with $\ell \leq k - 1$. Since $\sum_{\ell=0}^{k-1}(2\ell + 1) = k^2$, a given eigenvalue of the Laplacian (2.7) labeled by k is k^2-fold degenerate. The scalar spherical harmonics Y_{klm} may be chosen to satisfy $Y_{\mathbf{k}}^*(\mathbf{x}) = Y_{-\mathbf{k}}(\mathbf{x}) \equiv Y_{k\ell-m}(\mathbf{x})$, and normalized on S^3 so that [41]

$$\int_{S^3} d^3\Sigma Y_{k'\ell'm'}^*(\mathbf{x})Y_{k\ell m}(\mathbf{x}) = \delta_{\mathbf{k}\mathbf{k}'} = \delta_{kk'}\delta_{\ell\ell'}\delta_{mm'}, \quad (2.8)$$

and

$$\sum_{\ell=0}^{k-1}\sum_{m=-\ell}^{\ell}|Y_{k\ell m}(\mathbf{x})|^2 = \frac{k^2}{2\pi^2}, \quad (2.9)$$

which is independent of $\mathbf{x}$.

In the open case $\epsilon = -1$, the sums over (ℓ, m) remain, but k becomes a continuous variable with range $[0, \infty)$. After integration over the direction of $\mathbf{k}$ in the $\epsilon = 0$ case, one is also left with the integration over the magnitude k with the scalar measure $\int dk k^2/(2\pi^2)$. Because of Eq. (2.9), the compact $\epsilon = +1$ case is simply related to the noncompact cases of $\epsilon = 0, -1$ by the replacement of the integral $\int dk k^2/(2\pi^2)$ by the discrete sum, $\sum_{k=1}k^2/(2\pi^2)$. Beginning the sum from $k = 1$ (so that the spatially homogeneous mode on S^3 has eigenvalue $k = 1$ instead of $k = 0$) makes this correspondence between the discrete and continuous cases most immediate. We define the scalar measure,

$$\int[dk] \equiv \begin{cases} \int_0^\infty dk & \text{if } \epsilon = 0, -1 \\ \sum_{k=1}^\infty & \text{if } \epsilon = 1 \end{cases} \quad (2.10)$$

in order to cover all three cases with a single notation.

The time-dependent mode functions $\phi_k(t)$ satisfy the ordinary differential equation,

$$\frac{d^2\phi_k}{dt^2} + 3H\frac{d\phi_k}{dt} + \frac{(k^2 - \epsilon)}{a^2}\phi_k + (m^2 + \xi R)\phi_k = 0. \quad (2.11)$$

If one defines $f_k(t) \equiv a^{3/2}\phi_k$, then this equation is equivalent to

$$\ddot{f}_k + \left[\omega_k^2 + \left(\xi - \frac{1}{6}\right)R - \frac{1}{2}\left(\dot{H} + \frac{H^2}{2}\right)\right]f_k = 0, \quad (2.12)$$

where

$$\omega_k^2(t) \equiv \frac{k^2}{a^2} + m^2. \quad (2.13)$$

Equation (2.12) is the equation for a harmonic oscillator with time-dependent frequency. Note that the comoving momentum index k of the mode is constant, while the

physical momentum $p = k/a$ redshifts as the Universe expands.

An analogous time-dependent harmonic oscillator equation may be derived also in conformal time under the substitution, $\chi_k(\eta) \equiv a\phi_k$, viz.,

$$\chi_k'' + \left[k^2 + m^2a^2 + (6\xi - 1)\left(\frac{a''}{a} + \epsilon\right)\right]\chi_k = 0, \quad (2.14)$$

where the primes denote differentiation with respect to conformal time η.

In view of the completeness and orthonormality of the spatial harmonic functions $Y_{\mathbf{k}}(\mathbf{x})$, it is easily verified that the equal time commutation relation

$$\left[\Phi(t, \mathbf{x}), \frac{\partial\Phi}{\partial t}(t, \mathbf{x}')\right] = \frac{i\hbar}{a^3}\delta_\Sigma(\mathbf{x}, \mathbf{x}') \equiv \frac{i\hbar}{a^3}\sqrt{\gamma}\delta^3(\mathbf{x} - \mathbf{x}'),$$

$$(2.15)$$

is satisfied, provided the creation and annihilation operators obey

$$[a_{\mathbf{k}}, a_{\mathbf{k}'}^\dagger] = \delta_{\mathbf{k}\mathbf{k}'}, \quad (2.16)$$

in the discrete notation, and the complex mode functions satisfy the Wronskian condition,

$$a^3(\dot{\phi}_k\phi_k^* - \phi_k\dot{\phi}_k^*) = \dot{f}_k f_k^* - f_k\dot{f}_k^* = \chi_k'\chi_k^* - \chi_k\chi_k^{*\prime}$$
$$= -i\hbar. \quad (2.17)$$

From the equation of motion (2.11), (2.12) or (2.14) this Wronskian condition is preserved under time evolution. Hence any initial condition for the second order equation of motion satisfying (2.17) is *a priori* allowed by the commutation relations. Given any two solutions of (2.11), we define their Klein-Gordon inner product as

$$(\psi_k, \phi_k) \equiv \frac{ia^3}{\hbar}(\psi_k^*\dot{\phi}_k - \dot{\psi}_k^*\phi_k), \quad (2.18)$$

which is independent of time.

Let $v_k(t)$ be some particular set of time-dependent mode functions satisfying Eq. (2.11) and the Wronskian condition (2.17). These can be used to define a vacuum state. Any other set of solutions ϕ_k satisfying the same Wronskian condition can be expressed as a linear superposition of v_k and its complex conjugate,

$$\phi_k = A_k v_k + B_k v_k^*, \quad (2.19)$$

which is the form of a Bogoliubov transformation. Because of Eq. (2.17) the time-independent complex Bogoliubov coefficients must satisfy

$$|A_k|^2 - |B_k|^2 = 1, \quad (2.20)$$

for each k. This is one real condition on the two complex numbers A_k and B_k. Since multiplication of both A_k and B_k by an overall constant phase has no physical consequences, there are only two real parameters needed to specify the mode function for each k.

The inner product (2.18) is preserved under the Bogoliubov transformation (2.19), i.e.,

$$(\phi_k, \phi_k) = (v_k, v_k) = 1, \quad (2.21a)$$
$$(\phi_k^*, \phi_k) = (v_k^*, v_k) = 0. \quad (2.21b)$$

Thus, we can invert (2.19) and solve for the Bogoliubov coefficients at an arbitrary initial time, $t = t_0$ or $\eta = \eta_0$, with the result,

$$A_k = (v_k, \phi_k) = \frac{ia_0^3}{\hbar}(v_k^*\dot{\phi}_k - \dot{v}_k^*\phi_k)_0, \quad (2.22a)$$

$$B_k = (v_k^*, \phi_k) = -\frac{ia_0^3}{\hbar}(v_k\dot{\phi}_k - \dot{v}_k\phi_k)_0. \quad (2.22b)$$

Interactions may be incorporated in this treatment as well [42,43], within the semiclassical large N approximation, but in order to keep the discussion as simple as possible we shall not consider scalar self-interactions in this paper.

We restrict our attention to initial states of a free scalar field, which like the RW geometry (2.1) itself, are also spatially homogeneous and isotropic, and call such states *RW states*. Spatial homogeneity of the RW states, i.e., invariance under spatial translations in Σ implies that the bilinear expectation values, $\langle a_{\mathbf{k}}^\dagger a_{\mathbf{k}'}\rangle$ and $\langle a_{\mathbf{k}}a_{\mathbf{k}'}^\dagger\rangle$, can be nonvanishing if and only if $\mathbf{k} = \mathbf{k}'$, while the expectation values, $\langle a_{\mathbf{k}}^\dagger a_{\mathbf{k}'}^\dagger\rangle$ and $\langle a_{\mathbf{k}}a_{\mathbf{k}'}\rangle$, can be nonvanishing if and only if $\mathbf{k} = -\mathbf{k}'$. In addition, isotropy of the RW states under spatial rotations implies that the expectation value of the number operator for $\mathbf{k} = \mathbf{k}'$,

$$\langle a_{\mathbf{k}}^\dagger a_{\mathbf{k}}\rangle = n_k \equiv \langle a_{\mathbf{k}}a_{\mathbf{k}}^\dagger\rangle - 1, \quad (2.23)$$

can be a function only of the magnitude k. This constant number n_k in each Fourier mode is the consequence of the unmeasurable $U(1)$ phase of the mode functions, and becomes the third real parameter needed for each k to specify the initial quantum state of the scalar field. We show in the next section that if the free field density matrix for a RW state is described by the general Gaussian ansatz in the Hamiltonian description, then the state is necessarily a *mixed* state if $n_k \neq 0$.

Because of the two parameter freedom to redefine ϕ_k according to the Bogoliubov transformation, (2.19) and (2.20), it is always possible to fix the parameters so that the remaining bilinears are equal to zero [43], i.e.,

$$\langle a_{\mathbf{k}}a_{-\mathbf{k}}\rangle = \langle a_{\mathbf{k}}^\dagger a_{-\mathbf{k}}^\dagger\rangle = 0. \quad (2.24)$$

If Φ is expanded in terms of the vacuum modes v_k instead of ϕ_k, then the corresponding annihilation and creation operators are

$$c_{\mathbf{k}} = A_k a_{\mathbf{k}} + B_k^* a_{-\mathbf{k}}^\dagger, \quad (2.25a)$$
$$c_{\mathbf{k}}^\dagger = A_k^* a_{\mathbf{k}}^\dagger + B_k a_{-\mathbf{k}}. \quad (2.25b)$$

This is the characteristic form of a Bogoliubov transformation to a squeezed state. If the arbitrary overall phase is

fixed by requiring A_k to be real, then the general squeezed state parameters r_k and θ_k are defined by

$$A_k = \cosh r_k, \qquad (2.26a)$$

$$B_k = e^{i\theta_k} \sinh r_k. \qquad (2.26b)$$

The bilinear expectation values,

$$\langle c_{\mathbf{k}} c_{-\mathbf{k}} \rangle = (2n_k + 1)A_k B_k^* = \sigma_k A_k B_k^*, \qquad (2.27a)$$

$$\langle c_{\mathbf{k}}^\dagger c_{-\mathbf{k}}^\dagger \rangle = (2n_k + 1)A_k^* B_k = \sigma_k A_k^* B_k, \qquad (2.27b)$$

are nonzero, and

$$N_k \equiv \langle c_{\mathbf{k}}^\dagger c_{\mathbf{k}} \rangle = |B_k|^2 + (2|B_k|^2 + 1)n_k = n_k + \sigma_k |B_k|^2, \qquad (2.28)$$

is the average occupation number of the general mixed, squeezed state with respect to the vacuum modes v_k. We have introduced the shorthand notation $\sigma_k \equiv 2n_k + 1$ for the Bose-Einstein factor in Eqs. (2.27) and (2.28).

The two-point (Wightman) function of the scalar field may be expressed in terms of the mode functions ϕ_k and n_k in the form,

$$\langle \Phi(t, \mathbf{x})\Phi(t', \mathbf{x}') \rangle = \int [d\mathbf{k}](n_k \phi_k^*(t)\phi_k(t') + (n_k + 1)$$
$$\times \phi_k(t)\phi_k^*(t'))Y_{\mathbf{k}}(\mathbf{x})Y_{\mathbf{k}}^*(\mathbf{x}'), \qquad (2.29)$$

in the case that the expectation value $\langle \Phi(t, \mathbf{x}) \rangle = 0$. When $\mathbf{x} = \mathbf{x}'$, the sums or integrals over the angular part of $\mathbf{k}$ can be evaluated with the result that

$$\langle \Phi^2(t, \mathbf{x}) \rangle = \int [dk]\frac{P_\phi(k; t)}{k}, \qquad (2.30)$$

is independent of $\mathbf{x}$. Here

$$P_\phi(k; t) \equiv \frac{k^3}{2\pi^2}\sigma_k |\phi_k(t)|^2 \qquad (2.31)$$

is the power spectrum of scalar fluctuations in the general RW mixed, squeezed initial state. It may also be expressed in terms of v_k in the form,

$$P_\phi(k; t) = P_v(k; t) + \frac{k^3}{\pi^2}(N_k |v_k(t)|^2$$
$$+ \sigma_k \operatorname{Re}[A_k B_k^* v_k^2(t)]) \qquad (2.32a)$$

$$\text{where} \quad P_v(k; t) \equiv \frac{k^3}{2\pi^2}|v_k(t)|^2 \qquad (2.32b)$$

is the fluctuation power spectrum in the selected vacuum state, and Eqs. (2.19) and (2.28) have been used.

In inflationary models this scalar power spectrum becomes the source for scalar metric fluctuations, and it is the power spectrum of the metric fluctuations that is actually measured in the CMB [44]. Because of the linearity of the metric fluctuations and the assumed spatial homogeneity of the classical inflaton field, the resulting power spectrum in the CMB is the same (up to an overall normalization) as the

quantum scalar field spectrum (2.32) that generates it. For example, in slow roll inflationary models, the relation between the linearized curvature perturbation $\mathcal{R}_{\mathbf{k}}$ and the quantum scalar field perturbation $\delta\phi_{\mathbf{k}}$ is [1]

$$\mathcal{R}_{\mathbf{k}} = -\left[\frac{H}{\dot{\bar{\phi}}}\delta\phi_{\mathbf{k}}\right]_{t=t_*}, \qquad (2.33)$$

where $\bar{\phi}(t) = \langle \Phi(t, \mathbf{x}) \rangle$ is the classical inflaton field (assumed independent of position) and t_* is a time a few e-folds after the perturbation has exited the horizon. From (2.33) we find that the power spectrum of Gaussian curvature fluctuations $P_{\mathcal{R}}(k; t_*)$ that is actually observed in the temperature fluctuations of the CMB is related to the power spectrum of scalar field fluctuations $P_\phi(k; t_*)$ by

$$P_{\mathcal{R}}(k; t_*) = \left(\frac{H}{\dot{\bar{\phi}}}\right)^2 P_\phi(k; t_*) = \frac{1}{8\pi^2\epsilon}\left(\frac{H}{M_{\text{Pl}}}\right)^2 \frac{P_\phi(k; t_*)}{P_\phi^{\text{BD}}(k; t_*)}. \qquad (2.34)$$

In the latter relation we have introduced the standard definition of the slow roll parameter ϵ [1], not to be confused with the $\epsilon = 0, \pm 1$ defined in (2.4) denoting flat, closed or open spatial sections. We have also normalized the spectrum to the scale invariant Bunch-Davies vacuum power spectrum $P_v(k; t_*) = P_\phi^{\text{BD}}(k; t_*)$, given explicitly by Eq. (5.6) of Sec. V below. Since our intention in this paper is to address the short distance and initial state effects of the scalar field in a general, model independent way, we focus on the scalar field power spectrum $P_\phi(k; t)$ of Eqs. (2.32) exclusively in the succeeding sections, leaving the model dependent connection to the scalar metric power spectrum $P_{\mathcal{R}}(k; t)$ unspecified. We do not discuss tensor perturbations of the metric at all in this paper.

III. DENSITY MATRIX AND HAMILTONIAN

The development of the previous section in terms of expectation values of the Heisenberg field operator $\Phi(t, \mathbf{x})$, with initial data specified by the time-dependent complex mode functions ϕ_k is well suited to a treatment of the covariant energy-momentum tensor and its renormalization. We consider this in detail for general RW states in the next section. However, the evolution of a quantum system from given initial conditions may be expressed just as well in the Schrödinger representation, and it is the latter approach which makes most explicit the specification of the initial state and its unitary evolution in configuration space. The Schrödinger representation is also the one best suited to discussions of decoherence of cosmological perturbations and the quantum to classical transition in inflationary models, which we discuss in Sec. VII. We present the Hamiltonian form of the evolution and corresponding density matrix for a general RW initial state in this section, demonstrating its full equivalence with the covariant formulation.

For a free field theory (or an interacting one treated in the semiclassical large N or Hartree approximations) it is clear that the two-point function (2.29) contains all the nontrivial information about the dynamical evolution of the general RW initial state. Even if nonzero, higher order connected correlators do not evolve in time either in a free field theory or in the leading order large N approximation to an interacting theory. Hence at least as far as the time evolution is concerned, a time-dependent Gaussian ansatz for the Schrödinger wave functional (or density matrix) of the scalar field about its mean value is sufficient in both cases. A proof of the equivalence between the large N semiclassical equations and the evolution of a Gaussian density matrix has been given for flat Minkowski spacetime in Refs. [43,45]. The general RW case differs from the flat spacetime case mainly by the appearance of the time-dependent scale factor, so that with the appropriate modifications a mixed state Gaussian density matrix also exists for the scalar field evolution in cosmology.

In order to obtain this Gaussian density matrix, let us first derive the Hamiltonian form of the evolution equations (2.5) and (2.11). In its Hamiltonian form, the variational principle both for the matter and metric degrees of freedom, should begin with a classical action that contains only *first* time derivatives of both Φ and a. However, the standard classical action for the scalar field plus gravity system, viz.,

$$S_\Phi + S_{EH} = -\frac{1}{2} \int d^4x \sqrt{-g}[(\nabla_a\Phi)g^{ab}(\nabla_b\Phi) + m^2\Phi^2 + \xi R\Phi^2] + \frac{1}{16\pi G_N} \int d^4x \sqrt{-g}(R - 2\Lambda)$$

(3.1)

contains the Ricci scalar R in both the matter action S_Φ (when $\xi \neq 0$) and the Einstein-Hilbert action S_{EH}. For RW spacetimes R contains second order time derivatives of the scale factor in the $\dot{H}$ term of Eq. (2.4). In order to remove these second order time derivatives from the action one should add a surface term to the action functional above, thereby replacing $S_\Phi + S_{EH}$ in Eq. (3.1) by

$$S_{cl}[\Phi, \dot{\Phi}; a, \dot{a}] = S_\Phi + S_{EH} + 3 \int dt d^3\Sigma \frac{d}{dt} \times \left[a^3 H \left(\xi\Phi^2 - \frac{1}{8\pi G_N} \right) \right],$$

(3.2)

which modifies both the matter and gravitational parts of the classical action at the end points of the time integration, but otherwise leaves the Lagrangian evolution equations away from the end points unchanged. In fact, it is this classical action S_{cl} modified by the surface term in Eq. (3.2), and *not* $S_\Phi + S_{EH}$ whose Euler-Lagrange variation (which by definition has vanishing $\delta\Phi$ and δa at the end points) leads to the scalar field equation of motion (2.5), as well as the Friedmann equation for the scale factor.

The surface term addition to the gravitational action for a general spacetime has been given in Ref. [46].

With this corrected classical action S_{cl}, the conjugate momentum for the scalar field is

$$\Pi_\Phi \equiv \frac{\delta S_{cl}}{\delta \dot{\Phi}} = a^3(\dot{\Phi} + 6\xi H\Phi).$$

(3.3)

If we ignore the Friedmann equation for the scale factor for the moment, treating $a(t)$ as an externally specified function of time, then the classical Hamiltonian density of the scalar field alone is

$$\mathcal{H}_\Phi \equiv \dot{\Phi}\Pi_\Phi - \mathcal{L}_{cl}$$

$$= \frac{\Pi_\Phi^2}{2a^3} - 3\xi H(\Pi_\Phi\Phi + \Phi\Pi_\Phi) + \frac{a}{2}\gamma^{ij}(\partial_i\Phi)(\partial_j\Phi)$$

$$+ \frac{a^3}{2}m^2\Phi^2 + 3\xi a^3\left[(6\xi - 1)H^2 + \frac{\epsilon}{a^2}\right]\Phi^2, \quad (3.4)$$

where we have symmetrized the second term in this expression involving $\Pi_\Phi\Phi$, in anticipation of the replacement of Φ and Π_Φ by noncommuting quantum operators.

In the Hamiltonian framework the three independent symmetric quadratic variances, $\langle\Phi^2\rangle$, $\langle\Phi\Pi_\Phi + \Pi_\Phi\Phi\rangle$ and $\langle\Pi_\Phi^2\rangle$ at coincident times determine the Gaussian density matrix $\hat{\rho}$. The one antisymmetric variance, $\langle\Phi\Pi_\Phi - \Pi_\Phi\Phi\rangle$ is fixed by the canonical commutation relation,

$$[\Phi(t, \mathbf{x}), \Pi_\Phi(t, \mathbf{x}')] = i\hbar\delta_\Sigma(\mathbf{x}, \mathbf{x}'),$$

(3.5)

which using Eq. (3.3) and $[\Phi, \Phi] = [\Pi_\Phi, \Pi_\Phi] = 0$ is equivalent to Eq. (2.15). Let us introduce the definitions,

$$\sigma_k \equiv 2n_k + 1,$$

(3.6a)

$$\zeta_k(t) \equiv \sqrt{\sigma_k}|\phi_k|,$$

(3.6b)

$$\pi_k(t) \equiv a^3(\dot{\zeta}_k + 6\xi H\zeta_k),$$

(3.6c)

for the time-independent Bose-Einstein factor, σ_k, and the two real functions of time, $\zeta_k(t)$ and $\pi_k(t)$. We show in Appendix A that these definitions allow us to express the three bilinear Fourier field mode amplitudes in the form,

$$\sigma_k|\phi_k|^2 = \zeta_k^2,$$

(3.7a)

$$\sigma_k \text{Re}(\phi_k^*\dot{\phi}_k) = \zeta_k\dot{\zeta}_k,$$

(3.7b)

$$\sigma_k|\dot{\phi}_k|^2 = \dot{\zeta}_k^2 + \frac{\hbar^2\sigma_k^2}{4a^6\zeta_k^2},$$

(3.7c)

and this allows in turn for the three independent Gaussian variances at coincident spacetime points to be written as

$$\langle\Phi^2\rangle = \frac{1}{2\pi^2} \int [dk]k^2\zeta_k^2,$$

(3.8a)

$$\langle\Phi\Pi_\Phi + \Pi_\Phi\Phi\rangle = \frac{1}{\pi^2} \int [dk]k^2\zeta_k\pi_k,$$

(3.8b)

$$\langle\Pi_\Phi^2\rangle = \frac{1}{2\pi^2} \int [dk]k^2\left(\pi_k^2 + \frac{\hbar^2\sigma_k^2}{4\zeta_k^2}\right).$$

(3.8c)

Thus, the three independent bilinears depend on a set of three real functions of k, $(\zeta_k, \pi_k; \sigma_k)$, as expected from our discussion of the initial data in the Heisenberg representation of the previous section. The usefulness of this particular set is that (ζ_k, π_k) will turn out to be canonically conjugate variables of the effective Hamiltonian that describes the semiclassical time evolution of the general Gaussian density matrix, while σ_k is strictly a constant of the motion. Notice also that the power spectrum defined in (2.31) can be written in terms of the Gaussian width parameter $\zeta_k(t)$ directly as

$$P_\phi(k; t) = \frac{k^3}{2\pi^2} \zeta_k^2(t), \qquad (3.9)$$

for both the pure and more general mixed state cases. It is

independent of $\mathbf{x}$ and the direction of $\mathbf{k}$ by the spatial homogeneity and isotropy of the RW state.

The Hamiltonian and corresponding pure state Schrödinger wave functional for scalar field evolution in cosmology has been previously given in Refs. [29,30]. However, a pure state Gaussian ansatz for the wave functional imposes a constraint on the three parameters $(\zeta_k, \pi_k; \sigma_k)$, in fact implying $\sigma_k = 1$ for all k [43,45]. To remove this restriction one must allow for the Gaussian ansatz also to contain mixed terms, so that the density matrix $\hat{\rho} \neq |\Psi\rangle\langle\Psi|$ in general. By simply keeping track of the powers of $a(t)$ and its derivatives, it is straightforward to generalize the Minkowski spacetime density matrix to the RW case with the result,

$$\langle q|\hat{\rho}(t)|q'\rangle = \langle q_0|\hat{\rho}_0(\bar{\phi}, \bar{p}; \zeta_0, \pi_0; \sigma_0)|q'_0\rangle \prod_{\mathbf{k} \neq \mathbf{0}} \langle q_{\mathbf{k}}|\hat{\rho}(\zeta_k, \pi_k; \sigma_k)|q'_{\mathbf{k}}\rangle$$

$$= \rho_0 \prod_{\mathbf{k} \neq \mathbf{0}} (2\pi\zeta_k^2)^{-1/2} \exp\left\{-\frac{\sigma_k^2 + 1}{8\zeta_k^2}(|q_{\mathbf{k}}|^2 + |q'_{\mathbf{k}}|^2) + \frac{i\pi_k}{2\hbar\zeta_k}(|q_{\mathbf{k}}|^2 - |q'_{\mathbf{k}}|^2) + \frac{\sigma_k^2 - 1}{8\zeta_k^2}(q_{\mathbf{k}} q'^*_{\mathbf{k}} + q'_{\mathbf{k}} q^*_{\mathbf{k}})\right\}, \qquad (3.10)$$

where the $\{q_{\mathbf{k}}\}$ are the set of complex valued Fourier coordinates of the scalar field which are time-independent in the Schrödinger representation, i.e., the matrix elements of the Heisenberg field operator $\Phi(t_0, \mathbf{x})$ at an arbitrary initial time t_0 are defined by

$$\langle q|\Phi(t_0, \mathbf{x})|q'\rangle = \left(\int [d\mathbf{k}] Y_{\mathbf{k}}(\mathbf{x}) q_{\mathbf{k}}\right)\langle q|q'\rangle. \qquad (3.11)$$

The latter matrix element is nonvanishing only for $q_{\mathbf{k}} = q'_{\mathbf{k}}$ and is defined precisely by Eq. (3.19) below. In this Schrödinger coordinate representation the action of the conjugate momentum operator Π_Φ is given by

$$\langle q|\Pi_\Phi(t_0, \mathbf{x})|q'\rangle = -i\hbar\left(\int [d\mathbf{k}] Y_{\mathbf{k}}(\mathbf{x}) \frac{\partial}{\partial q^*_{\mathbf{k}}}\right)\langle q|q'\rangle. \qquad (3.12)$$

Since Φ is a real field, the complex coordinates are related by $q^*_{\mathbf{k}} = q_{-\mathbf{k}}$ and occur in conjugate pairs. Hence we have the rule,

$$\frac{\partial q_{\mathbf{k}}}{\partial q^*_{\mathbf{k}'}} = \delta_{\mathbf{k}, -\mathbf{k}'} \qquad (3.13)$$

and the $\pm\mathbf{k}$ terms in the density matrix (3.10) are identical, and may be combined. The $q_{\mathbf{k}=0}$ field coordinate is real, and we have separated off the $\mathbf{k} = \mathbf{0}$ component of the density matrix $\hat{\rho}$ in Eq. (3.10), denoting it by ρ_0. In this spatially homogeneous and isotropic Fourier component we may also allow for the possibility of a nonvanishing real mean value of the scalar field,

$$\bar{\phi}(t) = \langle \Phi(t, \mathbf{x})\rangle \equiv \text{Tr}(\Phi(t, \mathbf{x})\hat{\rho})$$

$$= \int_{-\infty}^{\infty} dq_0 q_0 \langle q_0|\hat{\rho}_0(\bar{\phi}, \bar{p}; \zeta_0, \pi_0; \sigma_0)|q_0\rangle, \qquad (3.14)$$

which because of the RW symmetry is a function of time only. This spatially homogeneous expectation value is the classical inflaton field in inflationary models. Separating this mode explicitly from the rest is possible strictly only in a discrete basis, such as that corresponding to closed spatial sections, $\epsilon = +1$. The density matrix in the spatially homogeneous sector is

$$\rho_0(t) \equiv \langle q_0|\hat{\rho}_0(\bar{\phi}, \bar{p}; \zeta_0, \pi_0; \sigma_0)|q'_0\rangle = (2\pi\zeta_0^2)^{-1/2} \exp\left\{i\frac{\bar{p}}{\hbar}(q_0 - q'_0)\right.$$

$$\left. - \frac{\sigma_0^2 + 1}{8\zeta_0^2}[(q_0 - \bar{\phi})^2 + (q'_0 - \bar{\phi})^2] + \frac{i\pi_0}{2\hbar\zeta_0}[(q_0 - \bar{\phi})^2 - (q'_0 - \bar{\phi})^2] + \frac{\sigma_0^2 - 1}{4\zeta_0^2}(q_0 - \bar{\phi})(q'_0 - \bar{\phi})\right\}, \qquad (3.15)$$

where

$$\bar{p}(t) \equiv a^3(\dot{\bar{\phi}}(t) + 6\xi H\bar{\phi}(t)) \qquad (3.16)$$

is the momentum conjugate to the spatially homogeneous mean field $\bar{\phi}(t)$.

Real field coordinates $(q^R_{\mathbf{k}}, q^I_{\mathbf{k}})$ for the $\mathbf{k} \neq \mathbf{0}$ modes may be introduced by

$$q_{\mathbf{k}} = \frac{1}{\sqrt{2}}(q^R_{\mathbf{k}} - iq^I_{\mathbf{k}}), \qquad \mathbf{k} \neq \mathbf{0} \qquad (3.17)$$

and the functional integration measure over the field coordinate space defined by

$$[\mathcal{D}q] \equiv dq_0 \prod_{\mathbf{k}>0} dq_{\mathbf{k}}^{\mathrm{R}} dq_{\mathbf{k}}^{\mathrm{I}}. \qquad (3.18)$$

The inner product appearing in (3.11) is defined by

$$\langle q | q' \rangle \equiv \delta(q_0 - q_0') \prod_{\mathbf{k}>0} \delta(q_{\mathbf{k}}^{\mathrm{R}} - q_{\mathbf{k}}'^{\mathrm{R}}) \delta(q_{\mathbf{k}}^{\mathrm{I}} - q_{\mathbf{k}}'^{\mathrm{I}}) \quad (3.19)$$

and the general mixed state Gaussian density matrix (3.10) is properly normalized,

$$\mathrm{Tr}\,\hat{\rho} = \int [\mathcal{D}q] \langle q | \hat{\rho} | q \rangle = 1, \qquad (3.20)$$

with respect to this measure.

It is clear from (3.10) that if $\sigma_k = 1$ for all k, then the mixed terms vanish and the density matrix reduces to the product,

$$\langle q | \hat{\rho} | q' \rangle |_{\sigma_k=1} = \langle q | \Psi \rangle \langle \Psi | q' \rangle \equiv \prod_{\mathbf{k}} \Psi(q_{\mathbf{k}}) \Psi^*(q_{\mathbf{k}}'),$$
$$(3.21)$$

characteristic of a pure state, with

$$\Psi(q_{\mathbf{k}}) = \Psi(q_{-\mathbf{k}}) = (2\pi\zeta_k^2)^{-1/4} \exp\left\{-\frac{|q_{\mathbf{k}}|^2}{4\zeta_k^2} + i\pi_k \frac{|q_{\mathbf{k}}|^2}{2\hbar\zeta_k}\right\},$$
$$\mathbf{k} \neq \mathbf{0}, \qquad (3.22a)$$

$$\Psi(q_0) = (2\pi\zeta_0^2)^{-1/4} \exp\left\{i\frac{\bar{p}}{\hbar}(q_0 - \bar{\phi}) - \frac{(q_0 - \bar{\phi})^2}{4\zeta_0^2}\right.$$
$$\left. + i\pi_0 \frac{(q_0 - \bar{\phi})^2}{2\hbar\zeta_0}\right\}, \qquad \mathbf{k} = \mathbf{0}, \qquad (3.22b)$$

which is the Gaussian pure state Schrödinger wave functional in the Fourier representation of the complex field coordinate basis (3.11). The pure state case corresponds to $n_k = 0$, $\sigma_k = 1$, and requires only the two real functions of k and t, (ζ_k, π_k) for its full specification.

The Wigner function(al) representation of the Gaussian density matrix is obtained by shifting $q_{\mathbf{k}} \rightarrow q_{\mathbf{k}} + x_{\mathbf{k}}/2$ and $q_{\mathbf{k}}' \rightarrow q_{\mathbf{k}} - x_{\mathbf{k}}/2$ in (3.10), and Fourier transforming $\hat{\rho}$ with respect to the difference variables $x_{\mathbf{k}}$. This yields

$$F_W[q, p] \equiv \int [\mathcal{D}x] \prod_{\mathbf{k}} (2\pi\hbar)^{-1} \exp\left(-\frac{i}{\hbar} p_{\mathbf{k}}^* x_{\mathbf{k}}\right)$$
$$\times \left\langle q_{\mathbf{k}} + \frac{x_{\mathbf{k}}}{2} \middle| \hat{\rho} \middle| q_{\mathbf{k}} - \frac{x_{\mathbf{k}}}{2} \right\rangle$$
$$= F_0(q_0, p_0) \prod_{\mathbf{k} \neq \mathbf{0}} (\pi\hbar\sigma_k)^{-1}$$
$$\times \exp\left\{-\frac{|q_{\mathbf{k}}|^2}{2\zeta_k^2} - \frac{2}{\hbar^2\sigma_k^2}|\zeta_k p_{\mathbf{k}} - \pi_k q_{\mathbf{k}}|^2\right\}, \quad (3.23)$$

where $p_{\mathbf{k}}^* = p_{-\mathbf{k}}$, and

$$F_0(q_0, p_0) = (\pi\hbar\sigma_0)^{-1} \exp\left\{-\frac{(q_0 - \bar{\phi})^2}{2\zeta_0^2}\right.$$
$$\left. - \frac{2}{\hbar^2\sigma_0^2}[\zeta_0(p_0 - \bar{p}) - \pi_0(q_0 - \bar{\phi})]^2\right\} \quad (3.24)$$

is the Wigner function in the spatially homogeneous $\mathbf{k} = \mathbf{0}$ sector. Note that the normalization of the Gaussian Wigner functional is constant in time, as required for a Hamiltonian evolution in phase space. For a given $q_{\mathbf{k}}$ this Gaussian function is peaked on the phase space trajectory,

$$p_{\mathbf{k}} \approx \frac{\pi_k}{\zeta_k} q_{\mathbf{k}}, \qquad \mathbf{k} \neq \mathbf{0}, \qquad (3.25a)$$

$$p_0 \approx \bar{p} + \frac{\pi_0}{\zeta_0}(q_0 - \bar{\phi}), \qquad \mathbf{k} = \mathbf{0}, \quad (3.25b)$$

becoming very sharply peaked on this trajectory in the formal classical limit $\hbar \rightarrow 0$, σ_k fixed, although the width of the peak becomes larger for mixed states with larger σ_k (with $\hbar$ fixed) [47]. The Wigner functional (3.23) is positive definite for Gaussian states and may be interpreted as a normalized probability distribution for any $\hbar\sigma_k$ [20].

The functional integration measure (3.18) implies an inner product,

$$\langle \Psi_2 | \Psi_1 \rangle = \int [\mathcal{D}q] \langle \Psi_2 | q \rangle \langle q | \Psi_1 \rangle \equiv \exp(i\Gamma_{12}) \quad (3.26)$$

between pure states, and a coherence probability functional,

$$\mathrm{Tr}\,(\hat{\rho}_1 \hat{\rho}_2) = \int [\mathcal{D}q] \int [\mathcal{D}q'] \langle q | \hat{\rho}_1 | q' \rangle \langle q' | \hat{\rho}_2 | q \rangle$$
$$\equiv \exp(-2\tilde{\Gamma}_{12}) \qquad (3.27)$$

for general mixed states in the Schrödinger picture. In the case of pure states the real functional $\tilde{\Gamma}_{12}$ becomes $\mathrm{Im}\Gamma_{12}$ of (3.26). In the case $\hat{\rho}_1 = \hat{\rho}_2$, performing the Gaussian integrations in the coordinate representation gives

$$\mathrm{Tr}(\hat{\rho}^2) = \int [\mathcal{D}q] \int [\mathcal{D}q'] \langle q | \hat{\rho} | q' \rangle \langle q' | \hat{\rho} | q \rangle = \left(\prod_{\mathbf{k}} \sigma_{\mathbf{k}}\right)^{-1}$$
$$= \exp\left(-\frac{1}{\pi^2} \int [dk] k^2 \ln\sigma_k\right) \leq 1. \qquad (3.28)$$

The inequality is saturated if and only if $\sigma_k = 1$ for all k, in which case $\hat{\rho} = |\Psi\rangle\langle\Psi|$, and the equality is simply a consequence of the normalization condition, $\langle\Psi|\Psi\rangle = 1$ on the pure state wave functional. If for any k, $\sigma_k > 1$, $\mathrm{Tr}\hat{\rho}^2 < 1$, which is characteristic of a mixed state density matrix.

For either a pure or mixed state the Gaussian density matrix satisfies the quantum Liouville equation,

$$i\hbar \frac{\partial \hat{\rho}}{\partial t} = [\mathcal{H}_\Phi, \hat{\rho}], \qquad (3.29)$$

provided the time-dependent parameters $(\bar{\phi}, \bar{p}; \zeta_k, \pi_k)$ ap-

pearing in $\hat{\rho}$ satisfy the first order equations,

$$\dot{\bar{\phi}} = \frac{\bar{p}}{a^3} - 6\xi H \bar{\phi}, \tag{3.30a}$$

$$\dot{\bar{p}} = 6\xi H \bar{p} - a^3 \left[m^2 + 6\xi \frac{\epsilon}{a^2} + 6\xi(6\xi - 1)H^2 \right] \bar{\phi}, \tag{3.30b}$$

$$\dot{\zeta}_k = \frac{\pi_k}{a^3} - 6\xi H \zeta_k, \tag{3.30c}$$

$$\dot{\pi}_k = 6\xi H \pi_k - a^3 \left[\frac{k^2}{a^2} + m^2 + (6\xi - 1)\left(\frac{\epsilon}{a^2} + 6\xi H^2 \right) \right] \zeta_k$$
$$+ \frac{\hbar^2 \sigma_k^2}{4a^3 \zeta_k^3}. \tag{3.30d}$$

The first two of these equations are equivalent to the second order equation for the spatially homogeneous mean field,

$$\ddot{\bar{\phi}} + 3H\dot{\bar{\phi}} + (m^2 + \xi R)\bar{\phi} = 0, \tag{3.31}$$

while the latter two may be combined to yield the second order equation for the Gaussian width parameter,

$$\ddot{\zeta}_k + 3H\dot{\zeta}_k + \left(\frac{k^2 - \epsilon}{a^2} + m^2 + \xi R \right)\zeta_k = \frac{\hbar^2 \sigma_k^2}{4a^6 \zeta_k^3}. \tag{3.32}$$

The last equation is derived in Appendix A from Eq. (2.11) and the defining relation (3.6b).

This establishes the equivalence between the Hamiltonian evolution of the Gaussian density matrix (3.10), according to (3.29), and the Lagrangian evolution of general RW initial states described in the previous section. Since the Hamiltonian (3.4) is Hermitian with respect to the field coordinate measure (3.18), the time evolution of $\hat{\rho}$ is unitary and the normalization (3.20) is preserved. Hence there is no dissipation in the system and the evolution is fully time reversible in principle. The time evolution of the density matrix parameters $(\bar{\phi}, \bar{p}; \zeta_k, \pi_k)$ may also be derived from an effective *classical* Hamiltonian, $H_{\text{eff}} = \text{Tr}(\mathcal{H}_\Phi \hat{\rho})$, given explicitly by Eq. (A10) of Appendix A. This effective Hamiltonian is just the expectation value of the quantum Hamiltonian $\mathcal{H}_\Phi$ in the general Gaussian state, in which $\hbar$ appears as a parameter. Notice that the role of the $\hbar$ term in Eq. (3.32) is to act as a "centrifugal barrier" for the coordinate ζ_k, preventing the Gaussian width parameter from ever shrinking to zero. This width depends on the product $\hbar\sigma_k$, so that the classical high temperature limit $\hbar\sigma_k \rightarrow k_B T/\omega_k$ is treated on the same footing as the quantum zero temperature limit $\hbar\sigma_k \rightarrow \hbar$. Both classical thermal and quantum uncertainty principle effects contribute to the width of the Gaussian in general.

The Hamiltonian evolution and the density matrix description of RW states is not manifestly covariant under general coordinate transformations, depending as it does on a particular slicing of the four-dimensional geometry into three-dimensional surfaces Σ. Since initial data must be specified on such a spacelike Cauchy surface, this is the

natural $3 + 1$ splitting for initial value problems in RW cosmology. The equation of motion (2.5) is certainly invariant under general coordinate transformations, whereas the initial data must be specified on some three surface Σ for any particular physical initial state.

We note that the canonical effective Hamiltonian generating the time evolution of the wave functional in the Schrödinger representation is not simply related to the expectation value $\langle T_{tt} \rangle$ of the covariant energy-momentum tensor. The key point in reconciling the canonical and covariant energy densities is that the *full* system of matter plus metric fields must be taken into account. After the addition of the surface terms to the standard classical action in Eq. (3.1) to remove the second derivatives of the metric, the momentum conjugate to the scale factor a is

$$\Pi_a \equiv \frac{\delta S_{\text{cl}}}{\delta \dot{a}} = -\frac{3}{4\pi G_N} \dot{a} a + 6\xi \dot{a} a \Phi^2 + 6\xi a^2 \dot{\Phi}\Phi. \tag{3.33}$$

Then the total Hamiltonian density, constructed in the canonical prescription is

$$\mathcal{H}_{\text{tot}} = \Pi_\Phi \dot{\Phi} + \Pi_a \dot{a} - \mathcal{L}_{\text{cl}} = -\frac{1}{8\pi G_N} a^3 G_{tt} + a^3 T_{tt}, \tag{3.34}$$

where

$$G_{tt} = 3\left(H^2 + \frac{\epsilon}{a^2} \right), \tag{3.35}$$

is the *tt* component of the Einstein tensor, $G_{ab} = R_{ab} - g_{ab}R/2$ and

$$T_{tt} = \frac{1}{2}\dot{\Phi}^2 + 6\xi H \dot{\Phi}\Phi + \frac{1}{2a^2}\gamma^{ij}(\partial_i\Phi)(\partial_j\Phi)$$
$$+ \frac{m^2}{2}\Phi^2 + 3\xi\left(H^2 + \frac{\epsilon}{a^2} \right)\Phi^2, \tag{3.36}$$

is the *tt* component of the covariant energy-momentum tensor

$$T_{ab} \equiv -\frac{2}{\sqrt{-g}}\frac{\delta}{\delta g^{ab}}S_\Phi$$
$$= (\nabla_a\Phi)(\nabla_b\Phi) - \frac{g_{ab}}{2}g^{cd}(\nabla_c\Phi)(\nabla_d\Phi)$$
$$- 2\xi\nabla_a(\Phi\nabla_b\Phi) + 2\xi g_{ab}g^{cd}\nabla_d(\Phi\nabla_c\Phi)$$
$$+ \xi G_{ab}\Phi^2 - \frac{m^2}{2}g_{ab}\Phi^2. \tag{3.37}$$

Invariance under transformations of the time coordinate leads to the classical equation of constraint,

$$\mathcal{H}_{\text{tot}} = 0. \tag{3.38}$$

Because of Eq. (3.34) this coincides with the *tt* component of the classical Einstein equations, viz., the Friedmann equation for RW cosmologies. This constraint equation is

equivalent to the requirement that the classical theory be invariant under arbitrary time reparametrizations, $t \rightarrow t'(t)$, a condition which Hamiltonian evolution in a fixed background $a(t)$ does not require. The three momentum constraints of spatial coordinate transformations, equivalent to the ti components of Einstein's equations are automatically satisfied in any homogeneous, isotropic RW state.

Hence the Hamiltonian and covariant approaches agree, only for the *full* system of gravity plus matter, i.e., provided that the RW scale factor is treated as a dynamical degree of freedom, on the same footing as the matter field Φ. In contrast, the Hamiltonian $\mathcal{H}_\Phi$ of Eq. (3.4) generates the correct evolution of the Φ field in a fixed RW background, whether or not the scale factor $a(t)$ satisfies Einstein's equations, and with no requirement of invariance under reparametrizations of time. It is the covariant energy-momentum tensor T_{ab} that is conserved and should be used whenever the full cosmological theory of matter and gravitational degrees of freedom are under consideration. With this important proviso the canonical and covariant formulations of the initial value problem are completely equivalent.

IV. ENERGY-MOMENTUM TENSOR OF UV ALLOWED RW STATES

In the previous sections we have defined and described general homogeneous and isotropic RW initial states, with no restriction on the set of three density matrix parameters $(\zeta_k, \pi_k; \sigma_k)$ which describes the state and its evolution. However, because of the Wronskian condition (2.17), that enforces the canonical commutation relations of the quantum field, the state parameters do not approach zero rapidly enough at large k for the integrals in (3.8) or the expectation value of the covariant energy-momentum tensor $\langle T_{ab} \rangle$, to converge. Hence these expressions are purely formal, and a definite renormalization prescription is necessary to extract the finite state-dependent terms. This is a necessary prerequisite for any discussion of short distance, initial state, or backreaction effects in inflation, at least within a conventional effective field theory framework.

Because the energy-momentum tensor is an operator of mass dimension four, it contains divergences up to fourth order in the comoving momentum cutoff k_M. Requiring that the forms of the integrands at large k match those expected for the vacuum up to fourth order in derivatives of the metric, allows for all the divergences in $\langle T_{ab} \rangle$ to be absorbed into counterterms of the relevant and marginally irrelevant terms of the local gravitational effective action [32]. This adiabatic order four condition on the initial state imposes restrictions on the set of parameters $(\zeta_k, \pi_k; \sigma_k)$ at large k, and guarantees that the renormalized expectation value $\langle T_{ab} \rangle_R$ will remain finite and well-defined at all subsequent times [38]. Conversely, failure to impose these short distance restrictions on the initial state leads to cutoff dependence which cannot be identified with covariant local

counterterms up to dimension four in the gravitational action, and which violate the assumptions of a low energy EFT for gravity consistent with the symmetries of general covariance implied by the equivalence principle.

The available counterterms up to dimension four in the coordinate invariant effective action for gravity are the four local geometric terms Λ, R, R^2 and $C_{abcd}C^{abcd}$ (the square of the Weyl conformal tensor), which can be added to the one-loop action of the scalar field, $S^{(1)}[g]$. Hence the low energy gravitational effective action is formally

$$S_{\text{eff}}[g] = S^{(1)}[g] + \frac{1}{16\pi G_N} \int d^4x \sqrt{-g}(R - 2\Lambda)$$
$$- \frac{1}{2} \int d^4x \sqrt{-g}(\alpha C_{abcd}C^{abcd} + \beta R^2), \quad (4.1)$$

where

$$S^{(1)}[g] = \frac{i\hbar}{2} \text{Tr} \ln(-\Box + \xi R + m^2), \quad (4.2)$$

and Λ, G_N, α, and β are bare parameters which are chosen to cancel the corresponding divergences in $S^{(1)}[g]$. A fully covariant renormalization procedure is one that removes all divergences in $S^{(1)}[g]$ by adjustment of the scalar parameters Λ, G_N, α, and β of (4.1), and *only* those parameters, in such a way that the total effective action $S_{\text{eff}}[g]$ and the renormalized energy-momentum tensor derived from it,

$$\langle T_{ab} \rangle_R = -\frac{2}{\sqrt{-g}} \frac{\delta}{\delta g^{ab}} S_R^{(1)}[g], \quad (4.3)$$

is finite (i.e., independent of the cutoff k_M) and covariantly conserved:

$$\nabla^b \langle T_{ab} \rangle_R = 0. \quad (4.4)$$

Thus the renormalized expectation value $\langle T_{ab} \rangle_R$ is strictly well-defined only by reference to the full low energy effective action $S_{\text{eff}}[g]$ and the equations of motion of the gravitational field following from it,

$$\frac{1}{8\pi G_N}(G_{ab} + \Lambda g_{ab}) = \langle T_{ab} \rangle_R + \alpha_R{}^{(C)}H_{ab} + \beta_R{}^{(1)}H_{ab},$$
$$(4.5)$$

of which it is a part.

The local conserved tensors,

$${}^{(1)}H_{ab} \equiv \frac{1}{\sqrt{-g}} \frac{\delta}{\delta g^{ab}} \int d^4x \sqrt{-g}R^2$$
$$= 2g_{ab}\Box R - 2\nabla_a\nabla_b R + 2RR_{ab} - \frac{g_{ab}}{2}R^2, \quad (4.6a)$$

$${}^{(C)}H_{ab} \equiv \frac{1}{\sqrt{-g}} \frac{\delta}{\delta g^{ab}} \int d^4x \sqrt{-g}C_{abcd}C^{abcd}$$
$$= 4\nabla^c\nabla^d C_{acbd} + 2R^{cd}C_{acbd}, \quad (4.6b)$$

derived from the fourth order terms in the effective action are similar to those which appear in any EFT, whose equations of motion involve a local expansion in the num-

ber of derivatives. Provided that we restrict our attention to the low momentum region of validity of the EFT, these terms in Eq. (4.3) should be negligible compared to those involving fewer derivatives of the metric. Their only role is to provide the covariant UV counterterms necessary to remove the subleading logarithmic divergences in $\langle T_{ab} \rangle$. Conversely, if when multiplied by finite renormalized parameters α_R and β_R of order unity, they are not negligibly small, then the applicability of the EFT framework for low energy gravity is in question. In cosmology certainly a necessary condition for this framework to be applicable is that the Riemann curvature tensor components and their contractions are negligibly small in Planck length units, i.e.,

$$\ell_{\mathrm{Pl}}^2 |R_{ab}{}^{cd}| = \hbar G_N |R_{ab}{}^{cd}| \ll 1, \qquad (4.7)$$

and we restrict ourselves to this regime.

The independent (unrenormalized) components of the energy-momentum tensor with nonvanishing expectation values in a general RW initial state are the energy density,

$$\varepsilon_u \equiv \langle T_{tt} \rangle_u$$

$$= \frac{1}{4\pi^2} \int [dk] k^2 \sigma_k \left[|\dot{\phi}_k|^2 + \left(\frac{k^2 - \epsilon}{a^2} + m^2 \right) |\phi_k|^2 \right]$$

$$+ \frac{3\xi}{2\pi^2} \int [dk] k^2 \sigma_k \left[2H \mathrm{Re}(\phi_k^* \dot{\phi}_k) + \left(\frac{\epsilon}{a^2} + H^2 \right) |\phi_k|^2 \right], \qquad (4.8a)$$

and the trace,

$$T_u = \frac{(6\xi - 1)}{2\pi^2} \int [dk] k^2 \sigma_k \left[-|\dot{\phi}_k|^2 + \left(\frac{k^2 - \epsilon}{a^2} \right. \right.$$

$$\left. \left. + m^2 + \xi R \right) |\phi_k|^2 \right] - \frac{m^2}{2\pi^2} \int [dk] k^2 \sigma_k |\phi_k|^2. \quad (4.8b)$$

The other nonvanishing components of $\langle T_{ab} \rangle$ in a general RW state are the diagonal spatial components, $\langle T_{ij} \rangle = p g_{ij}$. The isotropic pressure p may be obtained from the energy density ε and trace T, by $p = (\varepsilon + T)/3$. The conservation equation (4.4) in the case of RW symmetry has only a time component which is nontrivial, namely,

$$\dot{\varepsilon} + 3H(\varepsilon + p) = \dot{\varepsilon} + H(4\varepsilon + T) = 0. \qquad (4.9)$$

The unrenormalized expressions (4.8) satisfy this relation by use of the equation of motion (2.11), provided that a comoving momentum cutoff k_M, introduced to render the integrals finite, is itself independent of time. An important criterion for any renormalization procedure is that it preserve this property so that (4.9) remains valid for the fully renormalized quantities as well. Notice that a fixed cutoff in the physical momentum $p_M = k_M/a$ will *not* preserve the form of the covariant conservation equation (4.9), because of the nonvanishing time derivative operating on the upper limit of the integrals at $k_M = p_M a$, if $p_M = M$ is assumed to be independent of time.

In the case of spatially homogeneous and isotropic RW spacetimes the adiabatic method has been shown to be equivalent to a fully covariant treatment of the divergences of the energy-momentum tensor which preserves its conservation [48]. The starting point of this method is the WKB-like form of the exact mode functions,

$$\phi_k(t) \equiv \sqrt{\frac{\hbar}{2a^3 \Omega_k}} \exp\left(-i \int^t dt' \Omega_k(t') \right), \qquad (4.10)$$

which when substituted into (2.11), yields the second order equation for Ω_k,

$$\Omega_k^2 = \omega_k^2 + \left(\xi - \frac{1}{6} \right) R - \frac{1}{2} \left(\dot{H} + \frac{H^2}{2} \right) + \frac{3}{4} \frac{\dot{\Omega}_k^2}{\Omega_k^2} - \frac{\ddot{\Omega}_k}{2\Omega_k}. \qquad (4.11)$$

From this expression a systematic asymptotic expansion of the frequency Ω_k in time derivatives of the metric scale factor $a(t)$ can be developed. At leading order, neglecting all time derivatives, $\Omega_k \simeq \omega_k$. Substituting this into the right-hand side of (4.11), one finds to second order,

$$\Omega_k \simeq \omega_k + \frac{(\xi - \frac{1}{6})}{2\omega_k} R - \frac{m^2}{4\omega_k^3}(\dot{H} + 3H^2) + \frac{5}{8} \frac{m^4}{\omega_k^5} H^2 + \dots, \qquad (4.12)$$

where the ellipsis consists of terms third and higher order in derivatives of the metric. It is clear that this asymptotic expansion is valid at large k, i.e., at distance scales much shorter than the characteristic scale of the variation of the geometry H^{-1}. Hence requiring the exact solutions of the mode equation (2.11) to match this asymptotic expansion to some order implies that the quantum state density matrix of the scalar field (3.10) should match that of the local vacuum to that order. It is a statement of the equivalence principle in the low energy EFT that the *local*, short distance properties of the quantum vacuum at a point x should approximate that of the nearly flat space vacuum constructed in a local neighborhood of x. Hence the wave functional (3.22) must have large k components characterized by $\{\zeta_k, \pi_k\}$ which are *universal*, corresponding to local geometric invariants at x in the effective action, and the same for all physically realizable states, independent of the geometry of the spacetime at larger scales.

When (4.12) is substituted into (4.10), and the resulting mode function is substituted into (4.8) with σ_k set equal to 1, one obtains integrands which match the quartic and quadratic divergent behavior of the unrenormalized stress tensor components [38]. Up to adiabatic order two these are explicitly given by

$$\varepsilon^{(2)} = \frac{\hbar}{4\pi^2 a^3} \int [dk] k^2 \varepsilon_k^{(2)}, \qquad (4.13a)$$

$$T^{(2)} = \frac{\hbar}{4\pi^2 a^3} \int [dk] k^2 T_k^{(2)}, \qquad (4.13b)$$

with

$$\varepsilon_k^{(2)} = \omega_k + \frac{m^4}{8\omega_k^5}H^2 + \frac{(6\xi-1)}{2\omega_k}\left[\frac{\epsilon}{a^2} - H^2 - \frac{m^2}{\omega_k^2}H^2\right], \quad (4.14a)$$

$$T_k^{(2)} = -\frac{m^2}{\omega_k} - \frac{m^4}{4\omega_k^5}(\dot{H}+3H^2) + \frac{5m^6}{8\omega_k^7}H^2 + \frac{(6\xi-1)}{\omega_k}$$

$$\times \left[\dot{H}+H^2+\frac{m^2}{2\omega_k^2}\left(2\dot{H}+3H^2+\frac{\epsilon}{a^2}\right) - \frac{3m^4}{2\omega_k^4}H^2\right]. \quad (4.14b)$$

Notice that these expressions are state-independent and universal, depending only upon the RW geometry and the parameters m, ξ of the matter Lagrangian. Although not manifestly covariant in form, the results of Ref. [48] indicate that subtracting these second order asymptotic terms from the unrenormalized energy density (4.8a) and trace (4.8b) corresponds to adjustment of the generally covariant counterterms up to two derivatives in the low energy effective action (4.1). Consistent with this, it may be checked that the second order energy density $\varepsilon^{(2)}$ of Eq. (4.14a) and the second order pressure $p^{(2)}$, satisfy the covariant conservation equation (4.9), provided any cutoff of the k integrals is again independent of time. Hence the (partially) renormalized energy density $\varepsilon_u - \varepsilon^{(2)}$, and trace $T_u - T^{(2)}$, which are free of quartic and quadratic divergences, also obey the conservation equation (4.9).

In order to remove the remaining logarithmic divergences in the energy density and trace in a general RW spacetime, the terms containing up to four derivatives of the metric must be subtracted as well in four spacetime dimensions. The expressions for the adiabatic order four terms in the mode expansion (4.12), or $\varepsilon^{(4)}$ and $T^{(4)}$ can be found in [26,41,48]. We shall not need their explicit form here, and simply assume that one can identify a particular solution v_k to (2.11), whose frequency function Ω_k possesses an asymptotic expansion for large k which agrees with (4.12), up to fourth adiabatic order, and stress tensor components (4.8), which agree with $\varepsilon^{(4)}$ and $T^{(4)}$ up to fourth adiabatic order.

In the general case, this is the necessary and sufficient condition for the renormalized stress tensor to be finite and conserved in the RW state corresponding to this particular mode function v_k. For example, in de Sitter spacetime,

these modes, v_k, could be taken to be the BD modes [2], since these are adiabatic order four modes and the BD state is a candidate vacuum state. A general set of modes ϕ_k can be written then as a Bogoliubov transformation (2.19) of these vacuum modes. The difference of the renormalized stress tensor in this general state with that given by the particular choice of $\phi_k = v_k$ and $\sigma_k = 1$ then define the finite state-dependent terms in the stress tensor in the general RW initial state. In order for the initial state defined by this general set of modes to remain a UV allowed RW initial state, the state-dependent terms in the renormalized stress tensor should not spoil the fourth order approach to the local vacuum which we required of the vacuum modes v_k. Hence we impose the condition that the integrals with state-dependent integrands must be convergent as well. Pure or mixed states satisfying this condition will be called *UV allowed RW states.*

These UV allowed states are described by mode functions, ϕ_k and corresponding density matrix parameters, $(\zeta_k, \pi_k; \sigma_k)$ for which the integrands in the stress tensor components (4.8) agree with the fourth order adiabatic integrands $\varepsilon_k^{(4)}$ and $T_k^{(4)}$ at large k. We may choose any particular fourth order adiabatic v_k with respect to which to define the renormalized vacuum energy-momentum tensor components,

$$\varepsilon_v \equiv \varepsilon_u \Big|_{\substack{\phi_k=v_k \\ \sigma_k=1}} - \varepsilon^{(4)}, \quad (4.15a)$$

$$T_v \equiv T_u \Big|_{\substack{\phi_k=v_k \\ \sigma_k=1}} - T^{(4)}. \quad (4.15b)$$

The definition of the class of UV allowed states then guarantees that the difference of stress tensors for any UV allowed RW state with respect to this choice of vacuum are well-defined and finite. To identify these terms we have only to introduce the form of the Bogoliubov transformation (2.19) for the general mode function ϕ_k into (4.8), and using (2.20), separate off the vacuum terms evaluated at $A_k = 1$, $B_k = 0$, and $\sigma_k = 1$, which are renormalized by (4.15). The remaining terms are the finite terms for arbitrary UV allowed RW states with respect to the given vacuum choice. Collecting these remaining state-dependent terms gives the fully renormalized result,

$$\varepsilon \equiv \langle T_{tt}\rangle_R = \varepsilon_v + \frac{1}{2\pi^2}\int [dk]k^2(N_k|\dot{v}_k|^2 + \sigma_k\text{Re}[A_kB_k^*\dot{v}_k^2]) + \frac{1}{2\pi^2}\int [dk]k^2\left(\frac{k^2-\epsilon}{a^2}+m^2\right)(N_k|v_k|^2 + \sigma_k\text{Re}[A_kB_k^*v_k^2])$$

$$+ \frac{6\xi H}{\pi^2}\int [dk]k^2(N_k\text{Re}[v_k^*\dot{v}_k] + \sigma_k\text{Re}[A_kB_k^*v_k\dot{v}_k]) + \frac{3\xi}{\pi^2}\int [dk]k^2\left(\frac{\epsilon}{a^2}+H^2\right)(N_k|v_k|^2 + \sigma_k\text{Re}[A_kB_k^*v_k^2]), \quad (4.16a)$$

and

$$T \equiv \langle T\rangle_R = T_v + \frac{(1-6\xi)}{\pi^2}\int [dk]k^2(N_k|\dot{v}_k|^2 + \sigma_k\text{Re}[A_kB_k^*\dot{v}_k^2])$$

$$+ \frac{(6\xi-1)}{\pi^2}\int [dk]k^2\left(\frac{k^2-\epsilon}{a^2}+m^2+\xi R\right)(N_k|v_k|^2 + \sigma_k\text{Re}[A_kB_k^*v_k^2]) - \frac{m^2}{\pi^2}\int [dk]k^2(N_k|v_k|^2 + \sigma_k\text{Re}[A_kB_k^*v_k^2]), \quad (4.16b)$$

where N_k is defined by Eq. (2.28). The vacuum terms denoted by the subscript v defined by Eqs. (4.15) and the additional state-dependent terms in Eqs. (4.16) are separately conserved. Because the state is assumed to be UV allowed, N_k also approaches zero faster than k^{-4} as $k \to \infty$, and all terms in Eq. (4.16) are finite, i.e., there is no cutoff dependence and the integrals may be extended to infinity. Note also that a pure state with $n_k = 0$, $\sigma_k = 1$ remains a pure (coherent) state under the Bogoliubov transformation (2.19), notwithstanding the nonzero value of $N_k = |B_k|^2$ for this state in the v_k basis. The quantum coherence effects of the Bogoliubov transformation appear also in the rapidly oscillating interference terms involving $A_k B_k^*$ in Eq. (4.16), which must be retained in the general UV allowed RW coherent state in order to retain the strict time reversibility of the evolution, as we shall see in Sec. VI.

V. SHORT DISTANCE EFFECTS IN INFLATION

The development of the previous sections applies to general RW initial states of the scalar field of any mass and ξ in an arbitrary RW spacetime. In this section we apply this general framework to the special case relevant for slow roll inflationary models, namely, de Sitter space with a massless minimally coupled inflaton field. If spatially flat sections are used, then the scale factor for de Sitter space takes the form,

$$a_{\mathrm{dS}} = \frac{1}{H} e^{Ht} = -\frac{1}{H\eta}, \qquad -\infty < \eta < 0, \qquad (5.1)$$

with H a spacetime constant related to the scalar curvature by $R = 12H^2$. The entire de Sitter manifold may be represented as a hyperboloid of revolution embedded in a five-dimensional flat Minkowski spacetime [32]. The hyperboloid possesses an $O(4, 1)$ invariance group of isometries, which can be made manifest if spatially closed coordinates ($\epsilon = 1$) are used. The flat coordinates ($\epsilon = 0$) with the scale factor given by (5.1) cover only one-half of the full de Sitter hyperboloid. None of the results presented in this section will depend on the choice of flat, open, or geodesically complete closed spatial sections, so we treat only the flat sections ($\epsilon = 0$) in detail.

In the flat sections under the transformation of variables $y = -k\eta = k \exp(-Ht)$, the mode equation (2.12) becomes Bessel's equation with index,

$$\nu^2 = \frac{9}{4} - \frac{m^2}{H^2} - 12\xi. \qquad (5.2)$$

The BD state [2,49,50] is the unique RW allowed state which is completely invariant under the full $O(4, 1)$ isometry group of de Sitter space. In the coordinates where the scale factor is given by (5.1) the BD state is specified by the particular solution of (2.11) given by

$$\phi_k^{\mathrm{BD}} \equiv \sqrt{\frac{\pi\hbar}{4Ha_{\mathrm{dS}}^3}} e^{i\pi\nu/2} e^{i\pi/4} H_\nu^{(1)}(y)$$

$$\to \frac{1}{a_{\mathrm{dS}}} \sqrt{\frac{\hbar}{2k}} e^{-ik\eta} \quad \text{as } y \to \infty. \qquad (5.3)$$

For $\nu^2 < 0$, ν is purely imaginary and Eq. (5.3) is independent of the choice of sign of $\mathrm{Im}(\nu)$. Note that the asymptotic form for $y = |k\eta| \to \infty$, holds independently of the value of ν. From the subleading terms in this asymptotic expansion of the Hankel function $H_\nu^{(1)}$ for large values of its argument, it is straightforward to show that the BD state (with $n_k = 0$) is an adiabatic order four UV allowed RW state for any ν. Hence taking $v_k = \phi_k^{\mathrm{BD}}$ is UV allowed and the adiabatic order four subtractions of Eqs. (4.15) yield a UV finite vacuum energy-momentum tensor expectation value in the BD state, which satisfies $T_v = -4\varepsilon_v$ or $p_v = -\varepsilon_v$, as a consequence of the de Sitter invariance of this state. The calculation of the renormalized ε_v as a function of the parameters m, H, ξ is given in Refs. [2,33,50].

The massless minimally coupled field is of particular interest both because slow roll inflationary models rely on such a field, and because it obeys the same mode equation in a RW spacetime as gravitons in a certain gauge [51]. For this field $m = \xi = 0$, $\nu = 3/2$ and Eq. (5.3) becomes

$$\phi_k^{\mathrm{BD}} \Big|_{\substack{m=0 \\ \xi=0}} = H \sqrt{\frac{\hbar}{2k^3}} e^{-ik\eta} (i - k\eta). \qquad (5.4)$$

Although this state is perfectly UV finite, an *infrared* divergence occurs in the two-point function, Eq. (2.29). The BD state must therefore be modified at very small values of k [52] which means that, strictly speaking, it is not possible to take $v_k = \phi_k^{\mathrm{BD}}$ for all k. An IR finite vacuum state is the AF state [28], which is actually a family of IR finite states. Because these states are not de Sitter invariant, the energy-momentum tensor for the massless minimally coupled scalar field in any of these states is also not de Sitter invariant [28,53]. Nevertheless it has been proven [27] that the energy-momentum tensor of the $m = 0$, $\xi = 0$ scalar field for any of the AF states and indeed for any UV finite, homogeneous, isotropic state, asymptotically approaches the de Sitter invariant energy-momentum tensor found by Allen and Folacci [28], namely, $p_v = -\varepsilon_v = 119\hbar H^4/960\pi^2$.

Since an AF state is just the BD state modified at very small values of k, the short distance or UV properties of the AF and BD states are identical. In this paper we are only concerned with these short distance effects, so we take as our preferred vacuum state $v_k = \phi_k^{\mathrm{BD}}$, even in the $m = \xi = 0$ case, ignoring the infrared divergences in the two-point function which this generates. This is not a problem for the power spectrum provided that the $k \approx 0$ modes are not the ones that dominate today. There are no infrared

divergences in the energy-momentum tensor of the BD state and the finite difference in $\langle T_{ab} \rangle_R$ between the BD state and any realistic AF state are of order H^4 and small if $H \ll M_{\text{Pl}}$. Hence this distinction will play no role in our analysis of short distance modifications of the initial state.

With $v_k = \phi_k^{\text{BD}}$ the power spectrum for a general UV allowed pure state given by (2.32) becomes

$$P_\phi(k;t)|_{n_k=0} = P_\phi^{\text{BD}} + \frac{k^3}{\pi^2}(|B_k|^2|\phi_k^{\text{BD}}|^2$$
$$+ \text{Re}[A_k B_k^*(\phi_k^{\text{BD}})^2]) \qquad (5.5)$$

where

$$P_\phi^{\text{BD}} = \frac{k^3}{2\pi^2}|\phi_k^{\text{BD}}|^2 = \hbar\left(\frac{H}{2\pi}\right)^2(1 + k^2\eta^2), \qquad (5.6)$$

is the spectrum of the Bunch-Davies state for the massless, minimally coupled field. In the late time limit, $\eta \to 0^-$,

$$P_\phi^{\text{BD}} \to \frac{\hbar H^2}{4\pi^2}, \qquad (5.7)$$

the BD power spectrum becomes completely independent of k, i.e., scale invariant. If one evaluates the power spectrum at the time of horizon crossing instead, $k|\eta| = k/aH \sim 1$, one also obtains a scale invariant spectrum with a normalization differing slightly from (5.7) by a constant factor of order unity [20,44].

The terms in (5.5) dependent on $|B_k|^2$ and on $A_k B_k^*$ are the contributions to the value of the power spectrum for states different from the BD vacuum state. The first important point to notice is that if the state is UV allowed then $|B_k|$ must approach zero and P_ϕ must approach P_ϕ^{BD} at large k. From this fact we can draw an immediate conclusion, namely, if P_ϕ is evaluated at horizon crossing, $k = Ha = e^{Ht}$, then *the power present in any UV allowed initial state always reverts to its scale invariant BD value for fluctuations with large enough k which cross the horizon at sufficiently late times.* To make this statement more quantitative suppose that the initial state is nonadiabatic up to some physical scale M at the initial time t_0 with $a(t_0) \equiv a_0$, while above that scale it is the same as the BD state. The comoving wave number corresponding to this scale is $k_M = Ma_0$. The horizon crossing time for a mode with this wave number is

$$t_M = H^{-1}\ln k_M = t_0 + H^{-1}\ln\left(\frac{M}{H}\right). \qquad (5.8)$$

Fluctuations which leave the horizon at times $t > t_M$ will have $k > k_M$ and the standard BD power spectrum, P_ϕ^{BD}. Thus, if inflation goes on for longer than $\ln(M/H)$ e-foldings, the initial state effects at the physical scale M inflate to scales far outside the horizon. If the scales we observe in the CMB now correspond to $k > k_M$, i.e., to modes which left the horizon of the de Sitter epoch at times $t > t_M$, then there will be *no* imprint of the short distance

initial state effects at scale M in the present day CMB observations. Conversely, if $k \leq k_M$ for the currently observable modes then the initial state modifications of the spectrum at scale M may be observable. Taking the present horizon crossing scale to be of the order of the present Hubble parameter H_{now}, this would imply that the condition

$$Ma_0/a_{\text{now}} \simeq H_{\text{now}} \qquad (5.9)$$

is satisfied. This condition on Ma_0/a_{now} is a general constraint on the present observability of any initial state effects in inflation at the scale M in the low energy EFT framework, regardless of their short distance origin. Additional constraints and additional parameters may arise in any given inflationary model. For example in the slow roll scenario the measured CMB power spectrum depends on the slow roll parameter ϵ in Eq. (2.34), so that observational constraints on the CMB power spectrum generally depend on more parameters than simply those of the initial state of the scalar field.

A constraint which does not depend on other parameters of the inflationary model is that arising from the energy-momentum tensor of initial states different from the BD state. If these contributions to the stress tensor are too large the model will deviate significantly from de Sitter space and may not inflate at all. Specializing our general results (4.16) to the case of de Sitter space with flat spatial sections and a scalar field that is massless and minimally coupled, the relevant energy-momentum tensor components are

$$\varepsilon = \varepsilon_{\text{BD}} + I_1 + I_2, \qquad (5.10a)$$
$$T = T_{\text{BD}} + 2I_1 - 2I_2, \qquad (5.10b)$$
$$p = \frac{\varepsilon + T}{3} = p_{\text{BD}} + I_1 - \frac{I_2}{3}, \qquad (5.10c)$$

where the finite state-dependent integrals I_1 and I_2 for the case $n_k = 0$, $\sigma_k = 1$ are

$$I_1 \equiv \frac{1}{2\pi^2}\int_0^\infty dk k^2(|B_k|^2|\dot\phi_k^{\text{BD}}|^2$$
$$+ \text{Re}[A_k B_k^*(\dot\phi_k^{\text{BD}})^2]), \qquad (5.11a)$$
$$I_2 \equiv \frac{1}{2\pi^2 a^2}\int_0^\infty dk k^4(|B_k|^2|\phi_k^{\text{BD}}|^2$$
$$+ \text{Re}[A_k B_k^*(\phi_k^{\text{BD}})^2]). \qquad (5.11b)$$

In the following subsections we consider various examples of modifications of the initial data for inflation and make use of these general expressions to compute the modified power spectrum and energy-momentum tensor components they generate.

A. de Sitter invariant α states

The BD state for the massive scalar field described by the mode functions (5.3) is a special RW allowed state, since it is invariant not only under spatial translations and

rotations, but also under the full $O(4, 1)$ isometry group of globally extended de Sitter spacetime [2]. For that reason it has seemed the most natural analog of the Poincaré invariant vacuum state of quantum field theory in flat Minkowski spacetime, and is usually assumed, explicitly or implicitly, to be the relevant "vacuum" state of the scalar field in inflationary models. However, as maximally extended de Sitter spacetime is very different from flat spacetime globally, global $O(4, 1)$ invariance is a much stronger condition than local flat space behavior.

Because it is a RW allowed state the BD state indeed has a two-point correlation function $\langle \Phi(x)\Phi(x') \rangle$ with short distance properties as $x \to x'$ that depend only on the local geometry at x, and more specifically are of the Hadamard form [33]. Since $\langle \Phi(x)\Phi(x') \rangle$ is of mass dimension two, this is equivalent to the statement that the BD mode functions (5.3) possess an asymptotic expansion for large k which agrees with Eqs. (4.10) and (4.12) up to second adiabatic order. However, *any UV allowed state* satisfies this property. What is special about the BD state is that its asymptotic expansion for large k agrees with the adiabatic expansion (4.12) to *all* orders. This is a much stronger statement than that it goes over to the local Poincaré invariant vacuum state in the flat space limit, since an infinite order adiabatic state carries information about the geometry of the background spacetime at *all* scales, including correlations on causally disconnected scales much larger than that of the horizon H^{-1}, a situation which has no analog in flat space.

The fourth order adiabatic condition on the state guarantees that the stress tensor in that state possesses no new divergences, and can be renormalized accordingly by the standard local counterterms of the low energy EFT of gravity. None of these RW allowed states are de Sitter invariant except the BD state. Nevertheless, we showed in Ref. [27] that as a consequence of the redshifting of short distance modes to large distances in de Sitter space, all RW allowed initial states for $\mathrm{Re}\nu < 3/2$ have a renormalized $\langle T_{ab} \rangle$ which approaches the BD value at late times. All such fourth order RW states are equivalent locally and are *a priori* equally possible initial states for an inflationary model.

The only possible way to generalize the BD state while maintaining de Sitter invariance, for $\mathrm{Re}\nu < 3/2$, would be to require $|A_k| = |A|$ and $|B_k| = |B|$ be independent of k and satisfy (2.19) [4,5,49]. Because of the unmeasurable overall phase of the Bogoliubov coefficients, we can choose A to be purely real and parametrize these squeezed states by a single complex number, $z \equiv re^{i\theta}$ as

$$A_k = \cosh r = \frac{1}{\sqrt{1 - |\lambda|^2}} = \frac{1}{\sqrt{1 - e^{\alpha+\alpha^*}}}, \quad (5.12a)$$

$$B_k = e^{i\theta}\sinh r = \frac{\lambda}{\sqrt{1 - |\lambda|^2}} = \frac{e^\alpha}{\sqrt{1 - e^{\alpha+\alpha^*}}}. \quad (5.12b)$$

The two alternate parametrizations shown in terms of λ and e^α are sometimes employed [8,13,49,54].

The Wightman function (2.29) for this general one complex parameter family of squeezed states is de Sitter invariant, in the sense that it is a function only of the $O(4, 1)$ de Sitter invariant distance between the points. However, as pointed out in Ref. [55] none of these squeezed α states are truly invariant under the de Sitter isometry group except for the BD state, $r = 0$, since the $r \neq 0$ states transform under $O(4, 1)$ transformations by a nonzero (in fact, infinite) phase. Since the Bogoliubov coefficient B_k in (5.12) does not approach zero at large k, this class of r or α states are *not* UV allowed states unless $r \equiv 0$ identically. If Eqs. (5.12) were taken literally for all k, the integrals (5.10) and (5.11) would diverge *quartically*. Although these states were used in various contexts, such as studying the sign of backreaction effects of particle creation in de Sitter space [4], and have been reconsidered lately by several authors [8,13,54], this severe UV divergence is unacceptable for a physical initial state within the low energy effective theory of gravity described by Eq. (4.1). This is clear even at the level of field theory with no self-interactions, provided only that it is covariantly coupled to gravity, since the stress tensor in the general α state has divergences which depend on α, and thus requires state-dependent counterterms for its renormalization [33]. When self-interactions are considered still other unphysical features become manifest [54].

Computing the power spectrum for a general (r, θ) state by using Eq. (2.31) with (5.12), we obtain

$$P_\phi(k; r, \theta) = P_\phi^{\mathrm{BD}}\bigg\{1 + 2\sinh^2 r - \frac{\sinh 2r}{1 + k^2\eta^2}$$
$$\times \mathrm{Re}((1 + ik\eta)^2 e^{-2ik\eta - i\theta})\bigg\}. \quad (5.13a)$$

At late times, $\eta \to 0^-$ (with k fixed)

$$P_\phi(k; r, \theta) \to \frac{\hbar H^2}{4\pi^2}(1 + 2\sinh^2 r - \sinh 2r\cos\theta). \quad (5.13b)$$

Because of the nonadiabatic UV modification of the BD state by r at arbitrarily large k, the effects of this modification do not redshift away, and essentially the same result is obtained if Eq. (5.13) is evaluated at horizon crossing time $\eta = -1/k \to 0^-$ with $k\eta = -1$ fixed. The scale invariant modification (5.13) is equivalent to that found in Ref. [8] with the choice,

$$\sinh r = \frac{H}{2M}, \quad (5.14a)$$

$$\theta = \frac{\pi}{2} - \frac{2M}{H} - \tan^{-1}\bigg(\frac{H}{2M}\bigg), \quad (5.14b)$$

which gives

$$P_\phi = P_\phi^{\mathrm{BD}}\left\{1 - \frac{H}{M}\sin\left(\frac{2M}{H}\right) + \frac{H^2}{M^2}\sin^2\left(\frac{M}{H}\right)\right\}$$

$$\simeq P_\phi^{\mathrm{BD}}\left\{1 - \frac{H}{M}\sin\left(\frac{2M}{H}\right)\right\}, \tag{5.15}$$

where the last approximate equality holds if $H \ll M$, with M the physical UV scale of new physics, denoted by Λ in [8]. In the case of exact de Sitter invariance the general de Sitter invariant squeezed state gives a simple multiplicative correction to P_ϕ^{BD} sinusoidally varying with H/M, which is itself unobservable in the CMB, since it has no k dependence. Hence it could be interpreted as a redefinition of the inflation scale H, with no observable consequences.

More importantly, the k independence of the strictly de Sitter invariant r and θ state is untenable in the EFT framework, since it leads to state-dependent divergences in the energy-momentum tensor. These are avoided if and only if the Bogoliubov coefficient B_k approaches zero fast enough at large k, for the state to be a UV allowed state. Thus one could consider a cutoff version of the (r, θ) states in which $r = 0$ above some large but finite comoving cutoff,

$$k_M = Ma_0, \tag{5.16}$$

with M the physical cutoff (in units of inverse length) at some arbitrary initial time t_0. One can then assume that modes with $k > k_M$ are in the adiabatic BD state while modes with smaller values of k are in an (r, θ) state. It is clear that such a state is no longer de Sitter invariant and has a power spectrum identical to (5.13) or (5.15) for $k < k_M$, but reverting back to its BD value for $k > k_M$. Thus in such a state,

$$P_\phi = P_\phi^{\mathrm{BD}}\{1 + \theta(k_M - k)[2\sinh^2 r - \sinh 2r\cos\theta]\}, \tag{5.17}$$

instead of (5.13). There is now a sharp break in the power spectrum, which could be observable in principle, if we are fortunate enough to have access to the right values of $k \sim k_M$ in the present CMB. *If* we assume that this condition is satisfied by the wave number of the present CMB, then observations would put a constraint on the magnitude of the deviations from scale invariance of the spectrum of the form (5.17). However, since this constraint is model dependent in any given inflationary model, we do not consider it further, and turn instead to the constraints arising from the contributions of such a cutoff r state to the energy-momentum tensor during the de Sitter phase.

It is clear from the divergence of the energy-momentum tensor at infinite k_M that the dominant contribution to the integrals in (5.11) comes from those modes close to the UV cutoff. Substituting (5.12) into (5.11) and cutting the integrals off at $k_M = Ma_0$, gives

$$I_1 = \frac{\hbar M^4}{16\pi^2}\left(\frac{a_0}{a}\right)^4 \sinh^2 r$$
$$- \frac{\hbar M^4}{64\pi^2}\left(\frac{a_0}{a}\right)^4 \sinh 2r F_\theta^{(3)}(x)|_{x=Ma_0/Ha} \tag{5.18a}$$

$$I_2 = \frac{\hbar M^4}{16\pi^2}\left[\left(\frac{a_0}{a}\right)^4 + \frac{2H^2}{M^2}\left(\frac{a_0}{a}\right)^2\right]\sinh^2 r$$
$$- \frac{\hbar M^4}{64\pi^2}\left(\frac{a_0}{a}\right)^4 \sinh 2r$$
$$\times \left[F_\theta^{(3)}(x) - \frac{4}{x}F_\theta^{(2)}(x) + \frac{4}{x^2}F_\theta^{(1)}(x)\right]_{x=Ma_0/Ha} \tag{5.18b}$$

with

$$F_\theta^{(p)}(x) \equiv \frac{\partial^p}{\partial x^p}\left(\frac{\sin x \sin(x-\theta)}{x}\right). \tag{5.19}$$

The properties of the functions $F_\theta^{(p)}(x)$ are discussed in Appendix B. All contributions to Eqs. (5.18) are finite (for finite M) at all times, including the initial time. All terms redshift with the expansion at least as rapidly as a^{-2}, in accordance with our general theorem in Ref. [27]. The terms involving $F_\theta^{(p)}$ are rapidly oscillating for early times, $Ma_0 \gg Ha$, but redshift to zero as fast or faster than the nonoscillatory terms for late times, $Ma_0 \ll Ha$. The transition from the oscillatory to damping behavior occurs at a time when $Ma_0 \sim Ha$ which is of the same order parametrically in M/H as t_M, defined in Eq. (5.8). By that time all the oscillatory terms give contributions to I_1 and I_2 which are already of order H^4 and negligible.

Since the maximum of $F_\theta^{(p)}(x)$ is of order unity, at an x of order unity, while $|F_\theta^{(p)}(x)|$ is bounded by $1/x$ as $x \to \infty$, the oscillating terms are never larger parametrically than HM^3, while the nonoscillating terms make a maximum contribution to the energy density or the pressure of order,

$$\frac{\hbar M^4}{16\pi^2}\sinh^2 r \tag{5.20}$$

at the initial time, $a = a_0$. Comparing this with the energy density of the inflaton field at the onset of inflation, $3\hbar H^2 M_{\mathrm{Pl}}^2/8\pi$ and requiring that the backreaction from the additional terms (5.18) be smaller gives the bound,

$$\sinh r < \sqrt{6\pi}\frac{H}{M}\frac{M_{\mathrm{Pl}}}{M}. \tag{5.21}$$

It is possible for the right side of this inequality to be larger than unity, even for $H \ll M$. Hence $\sinh r$ could be quite large and the break in the power spectrum (5.17) large enough to be observable, without creating too large a backreaction. This kind of a nonadiabatic initial state modification of the BD state at short distances produces the largest effects in the power spectrum (5.17) of the CMB

without giving an unacceptably large backreaction during inflation.

If on the other hand r is assumed small, as, for example, in (5.14), then no term in (5.18) is larger in magnitude than

$$\frac{\hbar H^2 M^2}{16\pi^2},\qquad(5.22)$$

and we obtain only the weaker condition,

$$M < \sqrt{6\pi}M_{\text{Pl}}.\qquad(5.23)$$

In this case the effects on the power spectrum would be small, though observable in principle, and the physical momentum scale of the cutoff $k_M/a_0 = M$ at the onset of inflation need only be somewhat smaller than the Planck scale, beyond which there would be no justification for using the low energy effective action for gravity (4.1) in any case.

Summarizing, the r states do not match the adiabatic expansion of the mode functions or the energy-momentum tensor at any k for which $r \neq 0$. They are therefore completely nonadiabatic states. For that very reason the size of their effects on the CMB power spectrum for k below the cutoff scale k_M can be arbitrarily large in principle. Observations of the CMB may provide the strongest constraints on this kind of initial state modification, but the quantitative bound depends on the inflationary model. The only model independent constraint for these nonadiabatic modifications of the initial state comes from the magnitude of the backreaction produced, which is a relatively weak constraint, giving Eq. (5.21) or (5.23). The largest power of the physical cutoff M allowed by dimensional analysis appears in the stress tensor, i.e., M^4 in (5.20) for these states. To illustrate how these results change if adiabatic conditions are imposed on the initial state, we consider next zeroth order adiabatic states.

B. Adiabatic order zero states

A state of given adiabatic order can be obtained by first substituting the expansion (4.12) into (4.10) and expanding to that adiabatic order. The result, evaluated at some arbitrary time t_0 serves as the initial condition for the exact modes ϕ_k. These modes will remain adiabatic to this order for all time [32].

A zeroth order adiabatic vacuum state for the massless minimally coupled scalar field can be obtained by setting $\Omega_k = \omega_k$ in Eq. (4.10) and omitting terms proportional to $\dot{a}$ in the resulting expression for $\dot{\phi}_k$ since they are of first adiabatic order. At the time t_0 one has then

$$\phi_k(0) = -H\eta_0\sqrt{\frac{\hbar}{2k}}e^{-ik\eta_0},\qquad(5.24a)$$

$$\dot{\phi}_k(0) = \dot{v}_k(0) = -iH^2\eta_0^2\sqrt{\frac{\hbar k}{2}}e^{-ik\eta_0}.\qquad(5.24b)$$

Substituting into (2.22) with $v_k = \phi_k^{\text{BD}}$ gives:

$$A_k = \left(1 + \frac{i}{2k\eta_0}\right),\qquad(5.25a)$$

$$B_k = \frac{i}{2k\eta_0}e^{-2ik\eta_0}.\qquad(5.25b)$$

If the relations (5.25) are substituted into Eq. (5.5), one finds the late time power spectrum,

$$P_\phi(k; \eta \to 0^-) = P_\phi^{\text{BD}}\left[1 - \frac{\sin(2k\eta_0)}{k\eta_0} + \frac{\sin^2(k\eta_0)}{k^2\eta_0^2}\right].\qquad(5.26)$$

In this case the initial state effects produce sinusoidal modulations of the power spectrum with wave number which vanish as $k \to \infty$. Also note that if the adiabatic order zero initial condition is taken in the infinite past, $\eta_0 \to -\infty$, the modifications vanish as well. This is because in that limit the adiabatic order zero initial state becomes the BD state with $A_k = 1$, $B_k = 0$, and the spectrum reverts to the standard BD value.

Danielsson [7] has considered initial data which are of the same form as Eqs. (5.24) and (5.25), but despite this apparently adiabatic construction, rather than viewing η_0 as a fixed time Cauchy surface, where initial conditions are imposed on the state for all k, he takes η_0 to depend on k in such a way that $k\eta_0 = -M/H$, with M a fixed physical scale. The motivation seems to have been to avoid making any statement about modes whose physical wavelength is shorter than the cutoff M^{-1}, and indeed $\eta_0(k) = -M/(Hk)$ is the conformal time at which the mode with comoving wave number k first falls below the physical cutoff M. However, inspection of (5.25) with this substitution, shows that $|B_k|$ now behaves as a constant, $H/2M$, as $k \to \infty$. Hence this prescription yields a state which is *not* adiabatic at all, but amounts to populating the highest frequency modes considered with a constant particle occupation number, and choosing a cutoff (r, θ) state with parameters given by (5.14). Thus the results of the previous subsection apply. These initial conditions taken literally for all k lead to an energy-momentum tensor which is quartically dependent on the cutoff, just as in the previous subsection. As discussed there, this is not a physically allowed UV state if extended to arbitrarily large k, i.e., arbitrarily late times $\eta_0(k) \to 0^-$.

This example illustrates the shortcomings of considering modifications of the initial state of inflation and their effects on the CMB power spectrum alone, without also considering the associated effects on the energy-momentum tensor and backreaction. When one considers only the power spectrum for some finite range of k, it may seem perfectly reasonable to restrict attention to only those modes with a physical wavelength larger than the short distance cutoff scale M^{-1}, since no sum or integral over k

is required for the power spectrum. However, the stress tensor does require such a sum over all k, and some prescription for the ultrashort distance modes has to be given as these modes will redshift to wavelengths larger than the short distance cutoff at later times. The essential question is not at what time η_0 these UV modes have physical wavelengths larger than M^{-1}, but rather what contribution do these short distance modes make to the energy-momentum tensor, which involves a sum/integration over *all* k, at *any* time. This question requires that a choice be made about whether the stress tensor is to be a consistent source for Einstein's equations and only if it is, can the magnitude of the backreaction be reliably estimated. General covariance of semiclassical gravity requires the state to be adiabatic at the very highest trans-Planckian energies, and this adiabaticity condition in turn constrains the possible effects of short distance initial state modifications on the power spectrum at late times, which might otherwise be overlooked.

Instead of taking η_0 to be a function of k let us assume that all the modes are determined at the same arbitrary but *fixed* time η_0, independent of k, by Eq. (5.24). This defines a true adiabatic order zero state. Although B_k given by (5.25) now does decrease with increasing k, its magnitude still does not fall off fast enough to make the state fourth order adiabatic and UV allowed. Since $k^4|B_k| \sim k^3$ as $k \rightarrow \infty$, the energy-momentum tensor can depend as much as cubically on the comoving momentum cutoff of the mode sum. The cubic divergence in the state-dependent mode sum means that there is no local (state-independent) counterterm available to absorb this divergence. The necessity of imposing a physical cutoff on the behavior of (5.25) implies that the power spectrum (5.26) cannot be valid for arbitrarily large k either, but instead must approach the BD spectrum more rapidly than (5.26) as $k \rightarrow \infty$. If we insert a cutoff k_M, as in the previous subsection such that the modes are the BD modes for $k > k_M$, then the condition (5.9) is necessary for these initial state modifications to be observable in the CMB today.

If Eqs. (5.25) are substituted into Eqs. (5.11) with a cutoff k_M one finds that

$$I_1 = \frac{\hbar H M^3}{32\pi^2}\left(\frac{a_0}{a}\right)^4$$
$$\times \left[F^{(2)}(x) + \frac{H}{M} - \frac{H}{M}F^{(1)}(x) \right]_{x=k_M(\eta-\eta_0)}, \quad (5.27a)$$

$$I_2 = -\frac{\hbar H M^3}{32\pi^2}\left(\frac{a_0}{a}\right)^4\left[F^{(2)}(x) - \frac{H}{M} - \frac{H}{M}F^{(1)}(x) \right.$$
$$+ \frac{4H}{M}\left(\frac{a}{a_0}\right)F^{(1)}(x) + \frac{4H^2}{M^2}\left(\frac{a}{a_0}\right)^2 F(x) - \frac{4H^2}{M^2}\left(\frac{a}{a_0}\right)F(x)$$
$$\left. - \frac{4H^3}{M^3}\left(\frac{a}{a_0}\right)^2 F^{(-1)}(x) \right]_{x=k_M(\eta-\eta_0)}. \quad (5.27b)$$

Here $F^{(p)} \equiv F_0^{(p)}$ with the latter defined in Eq. (5.19), $F \equiv$ $F^{(0)}$, and

$$F^{(-1)}(x) \equiv \int_0^x dy F(y) = \int_0^x dy \frac{\sin^2 y}{y}. \quad (5.28)$$

As expected the effects of the state-dependent terms redshift away like a^{-2} and a^{-4}, and are largest at or near the initial time $\eta = \eta_0$ when they are of order,

$$\frac{\hbar H M^3}{32\pi^2}. \quad (5.29)$$

Requiring this to be less than the energy density of the inflaton field gives the bound,

$$M < (12\pi H M_{\text{Pl}}^2)^{1/3}, \quad (5.30)$$

for the adiabatic order zero state, in place of (5.21) for the nonadiabatic state.

It is clear that Eq. (5.29), softer by one power of H/M compared to the previous case (5.20) is the result of the fact that the adiabatic order zero state has a Bogoliubov coefficient, $|B_k|$, which approaches zero at large k with one power of $1/k$ in (5.25). If we had chosen a state which matches the adiabatic vacuum mode v_k to first, second, or third order, i.e., with Bogoliubov coefficient $|B_k|$ approaching zero at large k like k^{-2}, k^{-3} or k^{-4}, respectively, then we should expect to obtain leading contributions to the stress tensor components that behave like $H^2 M^2$, $H^3 M$ or $H^4 \ln(M/H)$, respectively, for large M/H. When the state is a UV allowed state, the stress tensor components are independent of the upper limit $k_M = M a_0$ of the mode integrals for large M, so that the integral may be extended to infinity. In that case the stress tensor components are of order H^4, independent of the cutoff M, and negligible compared to the energy density driving the inflation for all $H \ll M_{\text{Pl}}$.

C. Boundary action approach

The authors of Ref. [25] have discussed setting conditions of the form,

$$(\partial_t \phi_k + \kappa \phi_k)|_{t=t_0} = 0, \quad (5.31)$$

on the initial state mode functions, motivated by the addition of boundary terms to the low energy EFT action functional. Here κ is in general a complex function of k. For the BD state,

$$\kappa_{\text{BD}} = -\frac{\dot{\phi}_k^{\text{BD}}(t_0)}{\phi_k^{\text{BD}}(t_0)} = \frac{H k^2 \eta_0^2}{1 + i k \eta_0} \quad (5.32)$$

becomes purely imaginary in the limit $k|\eta_0| \rightarrow \infty$. If we make use of (2.22) we find that the Bogoliubov coefficients are given by

$$A_k = \frac{-i(\kappa - \kappa_{BD}^*)}{2\sqrt{(\mathrm{Im}\kappa)(\mathrm{Im}\kappa_{BD})}} \frac{\phi_k^{BD}(0)^*}{|\phi_k^{BD}(0)|}, \qquad (5.33a)$$

$$B_k = \frac{i(\kappa - \kappa_{BD})}{2\sqrt{(\mathrm{Im}\kappa)(\mathrm{Im}\kappa_{BD})}} \frac{\phi_k^{BD}(0)}{|\phi_k^{BD}(0)|}, \qquad (5.33b)$$

up to an overall phase.

If attention is restricted to modifications of the BD state corresponding to the lowest dimension local operator in the scalar EFT on the initial time surface at $t = t_0$, namely $\beta(\nabla_i \Phi)^2/M$ where M is again the physical cutoff scale, then the authors of Ref. [25] argue that this would lead to a modified initial condition of the form (5.31) with

$$\kappa = \kappa_{BD} + \frac{\beta k^2}{a_0^2 M}. \qquad (5.34)$$

If the effective action on the boundary is real for real time it would seem that β must be real. The authors of Ref. [25] treat β as a real parameter, obtaining corrections to the (real) power spectrum which are linear in β. Treating β as an arbitrary complex parameter, we obtain the Bogoliubov coefficients for the case $k|\eta_0| \gg 1$,

$$A_k B_k^* \simeq -i\beta^* \frac{k}{2Ma_0} e^{2ik\eta_0}\left(1 - \frac{i\beta k}{2Ma_0}\right)$$
$$\times \left(1 + \frac{k}{Ma_0}\mathrm{Im}\beta\right)^{-1}, \qquad (5.35a)$$

$$|B_k|^2 \simeq \frac{|\beta|^2 k^2}{4M^2 a_0^2}\left(1 + \frac{k}{Ma_0}\mathrm{Im}\beta\right)^{-1}. \qquad (5.35b)$$

As in the α state case discussed in Sec. VA, these Bogoliubov coefficients are nonadiabatic and would lead to a divergent stress tensor if continued to arbitrarily large

k. Substituting the cutoff $k_M = Ma_0$ into the power spectrum (5.5) as before, we obtain at late times,

$$P_\phi(k; \eta \to 0^-) = P_\phi^{BD}\left\{1 + \frac{\theta(k_M - k)k}{(k_M + k\mathrm{Im}\beta)}\right.$$
$$\left. \times \left(\mathrm{Re}(i\beta^* e^{2ik\eta_0}) + |\beta|^2\frac{k}{k_M}\cos^2(k\eta_0)\right)\right\}. \qquad (5.36)$$

If $|\beta|$ is of order one then both terms in Eq. (5.36) are of the same order at $k \simeq k_M$, and the (large) deviations from a scale invariant spectrum may be observable in present CMB data. The same remarks about fine-tuning to $k \sim k_M$ and dependence on the specific features of the inflationary model apply to this initial state modifications as to the cutoff r states of Sec. VA.

If $|\beta|$ is assumed to be much less than unity, we can write $\beta = |\beta|e^{i\gamma}$ and obtain the modified power spectrum to linear order in $|\beta|$,

$$P_\phi(k) \simeq P_\phi^{BD}\left\{1 - |\beta|\theta(k_M - k)\frac{k}{k_M}\sin(2k\eta_0 - \gamma)\right.$$
$$\left. + \mathcal{O}(|\beta|^2)\right\}. \qquad (5.37)$$

The finite state-dependent contributions to the stress tensor given by the integrals (5.11) are easily written down for the case of general complex β, but because of the denominators in (5.35) they are rather complicated. In the case that β is purely real the integrals simplify and may be evaluated in terms of the functions $F^{(p)} \equiv F_0^{(p)}$ with the latter defined in Eq. (5.19). The result is

$$I_1 = \frac{\hbar\beta M^4}{128\pi^2}\left(\frac{a_0}{a}\right)^4 F^{(4)}(x)|_{x=k_M(\eta-\eta_0)} + \frac{\hbar\beta^2 M^4}{96\pi^2}\left(\frac{a_0}{a}\right)^4\left[1 + \frac{3}{16}F^{(5)}(x)\right]_{x=k_M(\eta-\eta_0)} \qquad (5.38a)$$

$$I_2 = -\frac{\hbar\beta M^4}{128\pi^2}\left(\frac{a_0}{a}\right)^4\left[F^{(4)}(x) + \frac{4H}{M}\frac{a}{a_0}F^{(3)}(x) + \frac{4H^2}{M^2}\left(\frac{a}{a_0}\right)^2 F^{(2)}(x)\right]_{x=k_M(\eta-\eta_0)}$$
$$+ \frac{\hbar\beta^2 M^4}{96\pi^2}\left(\frac{a_0}{a}\right)^4\left[1 + \frac{3H^2}{2M^2}\left(\frac{a}{a_0}\right)^2 - \frac{3}{16}F^{(5)}(x) - \frac{3H}{4M}\frac{a}{a_0}F^{(4)}(x) - \frac{3H^2}{4M^2}\left(\frac{a}{a_0}\right)^2 F^{(3)}(x)\right]_{x=k_M(\eta-\eta_0)}. \qquad (5.38b)$$

All terms are again finite at all times and redshift at late times at least as rapidly as a^{-2}, in accordance with our general theorem in Ref. [27]. The oscillatory integrals $F^{(p)}(x)$ are discussed in Appendix B, and the illustrative particular case of $F^{(4)}(x)$ is plotted in Fig. 1.

Since the maximum of $F^{(p)}(x)$ is of order 3 to 4 at $x = k_M(\eta - \eta_0) \sim 1$ (for p even) or $x = 0$ (for p odd), the maximum value of either the terms linear or quadratic in β is of order,

$$\max(\beta, \beta^2)\frac{\hbar M^4}{32\pi^2}, \qquad (5.39)$$

in contrast to (5.18), where the oscillatory terms were

smaller than the nonoscillatory ones. If the maximum of the contributions (5.39) are required to be smaller than the energy density driving the expansion, and $\beta \ll 1$ then the terms linear in β give the largest contribution and the strongest bound, viz.,

$$\beta < 12\pi\left(\frac{H}{M}\right)^2\left(\frac{M_{Pl}}{M}\right)^2. \qquad (5.40)$$

If $\beta < 1$ this bound on the term linear in β is of the same order of magnitude as that obtained in Ref. [16], but disagrees with the bound in Ref. [17], whose authors argue that the term linear in β gives no bound on the bulk stress tensor away from the initial boundary surface at $t = t_0$.
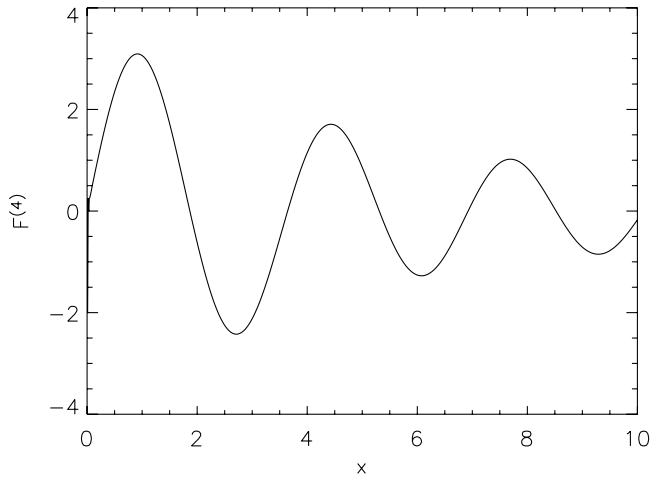
FIG. 1.   The oscillatory function $F^{(p)}$ for $p = 4$ defined by Eq. (B5), as a function of $x = k_M(\eta - \eta_0)$.

The authors of Ref. [17] evaluate integrals such as $F^{(4)}$ with a Gaussian cutoff in k, rather than with a hard cutoff at $k = k_M$ used here. However this by itself does not account for the disagreement. A Gaussian cutoff amounts to a slightly different choice of initial state which is also perfectly UV allowed, since it approaches the BD state faster than any power of k. This different state gives a different finite contribution to the renormalized stress tensor, which is of the same order in M as the contribution (5.39), differing only in its numerical coefficient at the initial time, and with a smoother, less strongly oscillatory behavior in time than that shown in Fig. 1 for the hard cutoff k_M. In neither case does the contribution to the energy-momentum tensor fall off exponentially in the conformal time difference $\eta - \eta_0$ as claimed in Ref. [17]. At late times $\eta \rightarrow 0^-$ either the Gaussian or hard cutoff of the initial state momentum integral yields a finite stress tensor with components that fall off as a^{-4}, a^{-3} and a^{-2}, as expected by the redshift of the RW expansion and in accordance with Ref. [27]. Hence the finite terms in the renormalized stress tensor which are first order in β are no more subject to renormalization ambiguities or localized on the boundary than those which are second order in β. Which terms give the stronger bound on β depends entirely on the values assumed for the parameters $k_M\eta_0$, H and M.

For general values of β, if the ratios H/M and M/M_{Pl} are the same order of magnitude, then requiring that the maximum contributions (5.39) be smaller than the energy density driving the expansion yields the bound,

$$\max(\beta, \beta^2) < 10 \leftrightarrow 100. \qquad (5.41)$$

If $\beta < 1$ this bound is easily satisfied, while if $\beta > 1$ the maximum value comes from the terms quadratic in β, and the results of Ref. [17] are recovered. We conclude that although the precise evaluation of the backreaction effects of initial state modifications differs from the estimates given in either [16] or [17], the qualitative final conclusion

that the backreaction constraints are not a very severe restriction on the parameter(s) of the boundary value action is similar to the conclusions reached by these authors.

VI. ADIABATIC PARTICLE CREATION AND DEPHASING

One of the principal physical effects that can be described by the semiclassical Gaussian density matrix is particle creation by a time varying RW scale factor [56]. Although the definition of a "particle" is intrinsically nonunique in a time varying background, it is possible to use the adiabatic nature of the UV allowed RW states to constrain this nonuniqueness considerably. The adiabatic particle concept was studied in some detail in Ref. [26], where a proposal was made for the adiabatic basis. The deficiency with that earlier proposal is that the two parts of the stress-energy tensor corresponding to vacuum and particle contributions are not separately conserved. Here we remedy that defect and in the process remove almost all of the nonuniqueness in the definition of adiabatic particle number.

Let us remark first that it is always possible to express the density matrix $\hat{\rho}$ in the *time-independent* number basis in which the number operator $a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}$ of Eq. (2.23) is diagonal. The transformation to this particle number basis may be derived by the methods of Refs. [43,57] with the result,

$$\langle n|\hat{\rho}|n'\rangle = \prod_{\mathbf{k}} \frac{2\delta_{n_{\mathbf{k}}n_{\mathbf{k}}'}}{\sigma_k + 1}\left(\frac{\sigma_k - 1}{\sigma_k + 1}\right)^{n_{\mathbf{k}}}, \qquad (6.1)$$

where n labels the set of integers $\{n_{\mathbf{k}}\}$, one for each distinct $\mathbf{k}$. In this occupation number representation, the density matrix is time-independent and diagonal, since the σ_k are constants of motion. The positive diagonal matrix elements of $\hat{\rho}$ in this discrete number representation may be viewed as the probabilities of finding exactly $n_{\mathbf{k}}$ particles in the mode labeled by wave number $\mathbf{k}$, with the particle number basis defined by the time-independent operator, $\hat{n}_{\mathbf{k}} = a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}$, i.e.,

$$\langle a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\rangle = \mathrm{Tr}(\hat{n}_{\mathbf{k}}\hat{\rho}) = \frac{2}{\sigma_k + 1} \sum_{n_{\mathbf{k}}=0}^{\infty} n_{\mathbf{k}}\left(\frac{\sigma_k - 1}{\sigma_k + 1}\right)^{n_{\mathbf{k}}}$$

$$= \frac{\sigma_k - 1}{2}, \qquad (6.2)$$

which is equivalent to Eq. (2.23), together with (3.6a).

Since the unitary transformation to the $n_{\mathbf{k}}$ basis exactly undoes the action of the time evolution operator, the preceding definition of particle number is always time-independent, no matter how rapidly the geometry changes with time. Hence it carries no information about particle creation in the time evolving RW geometry, or indeed about any features of the time evolution of the system whatsoever. Moreover, if one makes a Bogoliubov trans-

formation from one exact set of eigenfunctions $\phi_k(t)$ to another exact set $v_k(t)$, then the particle number N_k with respect to the new basis remains exactly time-independent, cf. (2.28). Useful as these are in the description of the density matrix and energy-momentum tensor of the field, any such time-independent parametrization of particle number is not the appropriate one to describe particle creation in time varying RW cosmologies. For that purpose one needs to define an appropriate time-dependent number operator with respect to an *approximate* adiabatic vacuum basis, which agrees with the exact number basis only at very large k.

The definition of the approximate adiabatic basis is constrained by several physical and practical requirements. First the basis should be defined by a set of WKB-like approximate mode functions,

$$
\begin{aligned}
\tilde{\phi}_k &= a^{-3/2} \tilde{f}_k \\
&= \sqrt{\frac{\hbar}{2a^3 W_k}} \exp\left(-i \int^t dt' W_k(t') - i\delta_k(t)\right), \quad (6.3)
\end{aligned}
$$

similar in form to Eq. (4.10), for some adiabatic frequency function W_k and additional time-dependent phase $\delta_k(t)$, to be specified. If $W_k = \Omega_k$, with Ω_k a solution to Eq. (4.11), and δ_k is constant in time, then $\tilde{f}_k$ would be an exact solution to the mode equation (2.12), and we would obtain the time-independent number basis (6.1). However, there is no constraint on the $\tilde{\phi}_k$ to satisfy the equation of motion (2.11), except asymptotically for large k, where they should approach the exact mode functions. Hence we have considerable freedom to choose the two functions, $W_k(t)$ and $\delta_k(t)$ which define the adiabatic basis. The most physically meaningful choice of basis is that which requires the vacuum energy-momentum tensor defined by $\tilde{\phi}_k$ to agree with the adiabatic expansion of the vacuum terms in the covariant energy-momentum tensor to a fixed adiabatic order. It is this tensor which couples to gravity and is covariantly conserved (as distinguished from the canonical Hamiltonian), so that matching the adiabatic basis (6.3) to this tensor will guarantee separate conservation of the vacuum polarization and particle contributions to the stress-energy.

It is clear that the leading (adiabatic order zero) terms in Eqs. (4.14) are vacuum polarization terms, and not particle contributions, since they appear even in static or flat space-times, where there can be no particle creation whatsoever. Thus, the frequency W_k of the adiabatic modes (6.3) should certainly match the zeroth order term ω_k in order for the energy-momentum tensor of these modes to match the adiabatic order zero terms in Eq. (4.14). However, adiabatic order zero matching is not sufficient. If we do not require matching $\tilde{\phi}_k$ to at least *second* order in the adiabatic expansion of the vacuum v_k, then we will find particle contributions to the energy-momentum tensor which *diverge quadratically* in the UV cutoff k_M. These divergent

adiabatic order two terms are clearly part of the *state-independent* vacuum polarization contribution to $\langle T_{ab}\rangle$, and not the particle content of the state. Choosing the adiabatic frequency and phase in (6.3) to match the adiabatic order two expansion terms (4.14) *exactly* guarantees that they will be completely removed by the subtractions in (4.15). It is this physical requirement that the power law divergences in the conserved stress tensor components should be associated with the state-independent geometrical contributions and not the state-dependent particle content that determines the adiabatic basis functions $\tilde{\phi}_k$ and renders the definition of adiabatic particle number (almost) unique. The only remaining nonuniqueness of the particle definition arises from the possibility of matching the vacuum contributions to the stress tensor to higher than second order in the adiabatic expansion.

In order to define the adiabatic basis precisely let the exact mode functions be expressed in terms of the adiabatic modes (6.3) as

$$
f_k = \alpha_k(t)\tilde{f}_k + \beta_k(t)\tilde{f}_k^*, \quad (6.4a)
$$

in terms of time varying Bogoliubov coefficients, $\alpha_k(t)$ and $\beta_k(t)$. The time-dependent Bogoliubov coefficients are required to satisfy

$$
|\alpha_k(t)|^2 - |\beta_k(t)|^2 = 1, \quad (6.4b)
$$

for each k, in order to guarantee that the transformation of bases is a canonical one. As in the case of the time-independent change of bases discussed in Sec. II, this allows a two real parameter freedom in the choice of the $\tilde{f}_k$ for each k (up to an overall irrelevant phase). Rather than using the Bogoliubov coefficients or W_k and δ_k as the two independent parameters, it is more convenient to use W_k and a different independent function of time $V_k(t)$, defined by the relation on the first time derivative of Eq. (6.4a). We define V_k implicitly by the relation,

$$
\dot{f}_k = \left(-iW_k + \frac{V_k}{2}\right)\alpha_k(t)\tilde{f}_k + \left(iW_k + \frac{V_k}{2}\right)\beta_k(t)\tilde{f}_k^*. \quad (6.4c)
$$

The canonical transformation from the exact mode functions f_k to the approximate adiabatic functions $\tilde{f}_k$ is now completely specified by W_k and V_k, while δ_k of Eq. (6.3) is fixed (implicitly) in terms of these two. Before determining W_k and V_k explicitly from the stress tensor components we first compute the particle number $\mathcal{N}_k$ and density matrix in the general adiabatic basis determined by Eqs. (6.4).

If the original field operator $\Phi(t, \mathbf{x})$ is expanded in terms of the approximate adiabatic mode functions, $\tilde{f}_k$ and $\tilde{f}_k^*$, the corresponding annihilation and creation operators,

$$
\tilde{a}_{\mathbf{k}}(t) = a_{\mathbf{k}}\alpha_k(t) + a_{-\mathbf{k}}^\dagger \beta_k^*(t), \quad (6.5a)
$$

$$
\tilde{a}_{\mathbf{k}}^\dagger(t) = a_{\mathbf{k}}^\dagger \alpha_k^*(t) + a_{-\mathbf{k}}\beta_k(t), \quad (6.5b)
$$

are generally time-dependent. Because of Eq. (6.4b) and the freedom to choose the phase of α_k we may set

$$\alpha_k(t) = \cosh r_k(t), \tag{6.6a}$$

$$\beta_k(t) = e^{i\theta_k(t)} \sinh r_k(t), \tag{6.6b}$$

in analogy with Eq. (2.26). With respect to this time-dependent adiabatic basis, the number density of particles in mode k is

$$\mathcal{N}_k(t) \equiv \langle \tilde{a}_{\mathbf{k}}^\dagger \tilde{a}_{\mathbf{k}} \rangle = n_k + \sigma_k |\beta_k(t)|^2$$
$$= \sinh^2 r_k(t) + n_k \cosh 2r_k(t), \tag{6.7}$$

which also depends on time in general. However, if W_k and V_k are chosen properly at large k, $\mathcal{N}_k$ will be an adiabatic invariant with respect to the effective classical Hamiltonian (A10), and therefore slowly varying in time for a slowly varying RW scale factor. The number density of particles spontaneously created from the vacuum with $n_k = 0$ is $|\beta_k(t)|^2$. An expression for this quantity in terms of W_k and V_k is easily obtained from Eq. (6.4) in the same way as (2.22), with the result,

$$|\beta_k(t)|^2 = \sinh^2 r_k(t) = \frac{1}{2\hbar W_k} \left| \dot{f}_k + \left(iW_k - \frac{V_k}{2} \right) f_k \right|^2$$
$$= \frac{1}{4W_k \Omega_k} \left[(W_k - \Omega_k)^2 + \frac{1}{4} \left(V_k + \frac{\dot{\Omega}_k}{\Omega_k} \right)^2 \right], \tag{6.8}$$

where the last expression follows from inserting the WKB-form (4.10) for the exact mode function f_k in terms of the *exact* frequency Ω_k. From this we observe that if (and only if) the adiabatic frequency matches the exact frequency, $W_k = \Omega_k$ and $V_k = -\dot{\Omega}_k/\Omega_k$, then the Bogoliubov coefficient β_k vanishes identically. In that case there is no particle creation and $\mathcal{N}_k = n_k$ is strictly a constant of the motion.

Using the value of $|\beta_k|^2$ the density matrix may be expressed in the adiabatic particle basis, by the methods of Refs. [43,57]. Unlike Eq. (6.1) the off-diagonal matrix elements of $\hat{\rho}$ do not vanish in this basis, and are quite complicated in the general case [43,57]. Although the diagonal elements are time-dependent, they depend on time only through the function $r_k(t)$. This means that they are relatively much more slowly varying than the corresponding phase variables $\theta_k(t)$, upon which the off-diagonal elements of $\hat{\rho}$ also depend. Thus, in this adiabatic number basis it becomes possible to argue that particle creation is related to phase decoherence or *dephasing* of the state: since macroscopic observables are generally relatively insensitive to the process of averaging over the rapidly varying phases, one can replace the exact $\hat{\rho}$ in this basis by its more slowly varying diagonal elements only [58]. In the pure state case, with adiabatic vacuum conditions in the infinite past, these diagonal elements simplify and are given explicitly by [43]

$$\rho_{n_k=2l_k}(k; t) = \frac{(2l_k - 1)!!}{2^{l_k} l_k!} \operatorname{sech} r_k \tanh^{2l_k} r_k. \tag{6.9}$$

The diagonal matrix elements are nonvanishing only for even integers, corresponding to the fact that particles are created from the vacuum in pairs. The positive numbers (6.9) have the interpretation of the probabilities of finding exactly l_k pairs of adiabatic particles at time t in the mode labeled by $k = |\mathbf{k}|$, with vacuum initial conditions as $t \to -\infty$. If this replacement of the exact pure state density matrix (3.10) by its phase averaged diagonal elements is justified, then the von Neumann entropy suffers the replacement,

$$S \equiv -\operatorname{Tr}(\hat{\rho} \ln \hat{\rho})$$
$$\to -\frac{1}{2\pi^2} \int dk k^2 \sum_{l_k=0}^{\infty} \rho_{2l_k}(k; t) \ln \rho_{2l_k}(k; t), \tag{6.10}$$

which becomes time-dependent. Although, in general the effective von Neumann entropy does not grow strictly monotonically in time, starting in an initial pure state with all of the $r_k = 0$ leads to a larger effective entropy at late times when some of the modes have $r_k \neq 0$ [39,58,59]. This shows that particle creation is directly related to increased squeezing of the initial state, and the growth of entropy this entails corresponds to the effective loss of information resulting from averaging over the rapidly varying phases $e^{\pm i\theta_k(t)}$ in macroscopic physical observables.

The validity of a truncation of the density matrix to its diagonal terms only in the adiabatic number basis and the associated loss of phase information will depend on the initial state and the details of the evolution [58]. When particle creation takes place from initial vacuumlike states this would seem to be quite a good approximation in the cases that it has been tested quantitatively [39,43]. Conversely, if one starts from a different type of initial state the squeezing parameters need not increase monotonically with time, and phase averaging is not justified. If it should happen in some special case(s) that the squeezing coefficients r_k are constant in time for all k, so that the adiabatic particle number basis becomes an exact vacuum basis, then by making the appropriate time-independent Bogoliubov transformation (2.19) to that basis one can set all the $r_k = 0$. Then it is clear that the θ_k become undefined and no phase decoherence of the initial state can occur by particle creation or dephasing effects in θ_k.

With these general remarks on the adiabatic particle number basis and dephasing let us fix the still undetermined functions W_k and V_k. Following Ref. [26] let us replace the exact mode functions ϕ_k by $\tilde{\phi}_k$, and their time derivatives $\dot{\phi}_k$ by $d\tilde{\phi}_k/dt = (-iW_k + V_k/2 - 3H/2)\tilde{\phi}_k$ with $\alpha_k = 1$, $\beta_k = 0$, and $n_k = 0$ in Eqs. (4.8) for the stress tensor components. In this way we obtain the (cutoff dependent) adiabatic vacuum contributions in this basis to the energy density and trace, namely,

$$\tilde{\varepsilon} = \frac{\hbar}{4\pi^2 a^3} \int [dk] k^2 \tilde{\varepsilon}_k, \qquad (6.11a)$$

$$\tilde{T} = \frac{\hbar}{4\pi^2 a^3} \int [dk] k^2 \tilde{T}_k, \qquad (6.11b)$$

with $\tilde{\varepsilon}_k$ and $\tilde{T}_k$ the following functions of W_k and V_k:

$$\tilde{\varepsilon}_k = \frac{1}{2W_k} \left[\omega_k^2 + W_k^2 + \frac{(V_k - H)^2}{4} + (6\xi - 1) \right.$$
$$\left. \times \left(HV_k - 2H^2 + \frac{\epsilon}{a^2} \right) \right], \qquad (6.12a)$$

$$\tilde{T}_k = \frac{1}{W_k} \left[-m^2 + (6\xi - 1)\left(\omega_k^2 - W_k^2 - \frac{V_k^2}{4} + \frac{3HV_k}{2} \right. \right.$$
$$\left. \left. - \frac{H^2}{4} + \dot{H} \right) + (6\xi - 1)^2 \frac{R}{6} \right]. \qquad (6.12b)$$

The adiabatic vacuum basis and particle number may be fixed now by setting these two expressions in terms of W_k and V_k *precisely equal* to the corresponding state-independent vacuum contributions to energy-momentum tensor in a general RW spacetime to second adiabatic order, given previously by Eq. (4.14). That is, we require

$$\tilde{\varepsilon}_k = \varepsilon_k^{(2)}, \qquad (6.13a)$$

$$\tilde{T}_k = T_k^{(2)}, \qquad (6.13b)$$

thus giving two relations for the two functions, W_k and V_k which determine the adiabatic particle basis. Solving for W_k we find

$$W_k = \frac{-m^2 + (6\xi - 1)(2\omega_k^2 + HV_k + \dot{H}) + (6\xi - 1)^2(HV_k + \dot{H} + \frac{2\epsilon}{a^2})}{2(6\xi - 1)\varepsilon_k^{(2)} + T_k^{(2)}}. \qquad (6.14)$$

If this is substituted back into either of Eqs. (6.13), we obtain a quadratic equation for V_k, so that the two functions W_k and V_k which determine the adiabatic vacuum modes can be determined algebraically for any ξ and m for the general RW geometry.

In this way the energy-momentum tensor components may be written in a form analogous to (4.16), where because of the exact matching (6.13) the vacuum terms up to second adiabatic order are now *identically zero*. Therefore the remaining nonvacuum terms defined with respect to the time-dependent adiabatic vacuum necessarily satisfy the covariant conservation equation (4.9). The nonvacuum terms in the adiabatic particle basis can be obtained from (4.16) with the replacements of $N_k \rightarrow \mathcal{N}_k$, $v_k \rightarrow \tilde{\phi}_k$, $A_k \rightarrow \alpha_k$ and $B_k \rightarrow \beta_k$. The nonvacuum energy-momentum tensor components for general W_k and V_k are given explicitly in [26], and we do not repeat them here.

Since Eq. (6.13) matches the vacuum energy-momentum tensor to second adiabatic order, the adiabatic frequency W_k obtained from solving Eqs. (6.13) agrees with the adiabatic expansion (4.12) up to and including second adiabatic order, differing from Ω_k only at adiabatic order four. On the other hand since V_k^2 and HV_k appear in Eqs. (6.13), and V_k contains terms of odd adiabatic orders, V_k needs to agree with $-\dot{\Omega}_k/\Omega_k \simeq -\dot{\omega}_k/\omega_k$ only up to first adiabatic order, i.e.,

$$V_k = -\frac{\dot{\omega}_k}{\omega_k} + \ldots = H\left(1 - \frac{m^2}{\omega_k^2}\right) + \ldots \qquad (6.15)$$

where the ellipsis includes terms of third adiabatic order and higher. Hence $V_k + \dot{\Omega}_k/\Omega_k$ is in general nonvanishing at adiabatic order three, and the lowest order term in the adiabatic expansion which appears in the expression for

$|\beta_k|^2$ in Eq. (6.8) is *sixth* order, i.e., the adiabatic particle number $\mathcal{N}_k$ defined by Eqs. (6.7), (6.8), and (6.13) is a fourth order adiabatic invariant. Time derivatives of $\mathcal{N}_k$ are correspondingly highly suppressed, and particle creation is small in slowly varying backgrounds, particularly at the highest wave numbers. In flat spacetime, H, ϵ and V_k vanish, and (6.14) together with (4.14) give $W_k = \omega_k$ which is time-independent. Hence the adiabatic modes become exact modes, $\beta_k = 0$, and no particles at all are created.

In the general case, Eqs. (6.13) match the vacuum energy-momentum tensor contributions to second adiabatic order and therefore satisfy a weaker condition than the fourth order UV allowed state condition. However the second order condition (6.13) is sufficient to render the total number of adiabatic particles created finite, i.e.,

$$\mathcal{N}(t) = \frac{1}{2\pi^2} \int [dk] k^2 \mathcal{N}_k(t) < \infty. \qquad (6.16)$$

Note that this would *not* be the case had we matched the energy-momentum tensor components only to the lowest (zeroth) order adiabatic expansion, or followed the Hamiltonian diagonalization procedure of Ref. [56]. The requirement of matching the stress tensor contributions (6.12) *exactly* to a fixed adiabatic order as in (6.13) (rather than reexpanding the algebraic expressions for W_k and V_k) removes essentially all of the ambiguity in the definition of the particle number, and guarantees that the nonvacuum particle contribution is separately conserved.

Because of the weighting of the momentum integrals for the energy-momentum tensor components (4.16) by an extra power of k with respect to Eq. (6.16), the identification (6.13) still leaves a logarithmic cutoff dependence in the vacuum energy-momentum tensor components. Unlike

the power law divergences which are noncovariant in form, the remaining logarithmic cutoff dependence is proportional to the geometric tensor $^{(1)}H_{ab}$ of Eq. (4.6a). Hence it can be viewed as a geometric vacuum polarization contribution in the low energy EFT equipped with a short distance cutoff. Even if the short distance cutoff is of the order of the Planck length, the higher derivatives in $^{(1)}H_{ab}$ mean that this contribution is negligible for all geometries varying more slowly than the Planck scale. Of course one is free to define the adiabatic particle basis by replacing the second order terms in (6.13) with the corresponding adiabatic order four (or higher) components $\varepsilon^{(4)}$ and $T^{(4)}$, which would remove also the covariant logarithmic cutoff dependence in $\langle T_{ab}\rangle$. Matching (6.3) to fourth order would make the resulting particle number (6.7) an adiabatic invariant of even higher (in fact, eighth) order. Matching to higher orders requires accurate knowledge of higher and higher order time derivatives of the metric scale factor, and hence of the entire spacetime evolution. Because only the power law divergences are noncovariant in form in the adiabatic procedure, matching to higher than second order is both unnecessary from the point of view of the covariant UV divergence structure of the stress tensor, and contrary to an EFT approach in terms of the minimal number of spacetime derivatives whenever EFT and the limits (4.7) apply. The tensor $^{(1)}H_{ab}$ actually vanishes in the important special case of de Sitter space, which renders the distinction between matching to second or fourth adiabatic order to remove the remaining logarithmic divergences in the stress tensor superfluous in this case. Finally, the second order definition of particle number through Eqs. (6.7), (6.8), and (6.13) is also the *minimal* one necessary to yield a finite total number of created particles (6.16) in a general RW spacetime. We emphasize that such a definition of particle number is intrinsic, based on the physical requirement of identifying and removing the vacuum contributions to the conserved stress tensor, and does not depend on the existence of flat *in* and *out* regions of the spacetime, or any extraneous notion of particle detectors [2].

We now apply this general second order definition of adiabatic particle number to two important special cases with zero mass. When $m = 0$ the relations (6.13) simplify

considerably. For the massless conformally coupled field ($m = 0$, $\xi = 1/6$), the condition on the trace (6.13b) becomes empty but the first condition (6.13a) gives directly,

$$m = 0,\ \xi = \frac{1}{6}: \qquad (W_k - \omega_k)^2 + \frac{(V_k - H)^2}{4} = 0, \tag{6.17}$$

which is solved uniquely for real W_k and V_k by

$$W_k = \omega_k = \frac{k}{a}, \tag{6.18a}$$

$$V_k = H = -\frac{\dot{W}_k}{W_k}. \tag{6.18b}$$

Because the fourth order adiabatic terms in $\varepsilon^{(4)}$ and $T^{(4)}$ vanish for the massless conformally coupled field in an arbitrary RW spacetime, this result for W_k and V_k remains unchanged if the fourth order adiabatic basis is used. Indeed as a result of the relations (6.18), the adiabatic basis functions (6.3) become *exact* solutions of Eq. (2.11) for the massless, conformally coupled field in a *general* RW spacetime. This means that the Bogoliubov coefficients α_k and β_k become time-independent. As in Eq. (2.19), $\mathcal{N}_k$ may be identified with the time-independent N_k and there is no production of massless, conformally coupled scalar particles.

In a second important special case, the massless minimally coupled field ($m = 0$, $\xi = 0$), V_k drops out of Eq. (6.14), and either of Eqs. (6.13) gives a quadratic equation for V_k, which is easily solved. Thus for $m = 0$ and $\xi = 0$ we obtain

$$W_k = \frac{2(k^2 - \epsilon)}{2k^2 - \epsilon + a^2(\dot{H} + 2H^2)}\frac{k}{a}, \tag{6.19a}$$

$$V_k = 3H - \frac{2(k^2 - \epsilon)^{1/2}}{2k^2 - \epsilon + a^2(\dot{H} + 2H^2)}$$
$$\times \left[4H^2k^2 - (a^2\dot{H} + \epsilon)\left(\dot{H} + 2H^2 - \frac{\epsilon}{a^2}\right)\right]^{1/2}, \tag{6.19b}$$

in a general RW spacetime. Although this definition of the adiabatic basis was determined by matching only to second adiabatic order, by Eqs. (6.13), in the special case of de Sitter space it becomes *exact*. Explicitly in flat spatial sections, $\epsilon = 0$, for which $\dot{H} = 0$, we have

$$m = 0,\ \xi = 0,\ a = a_{\rm dS}: \qquad W_k = \frac{k^2}{k^2 + H^2 a_{\rm dS}^2}\frac{k}{a_{\rm dS}}, \tag{6.20a}$$

$$V_k = H\frac{k^2 + 3H^2 a_{\rm dS}^2}{k^2 + H^2 a_{\rm dS}^2} = -\frac{\dot{W}_k}{W_k}, \tag{6.20b}$$

with $a_{\rm dS}$ given by Eq. (5.1). As a result of these relations, the adiabatic mode function $\tilde{\phi}_k$ is an *exact* solution of the mode equation (2.11) in de Sitter space. Indeed from Eqs. (5.1) and (6.20),

$$\int^t dt'\, W_k(t') = k^3 \int^\eta \frac{d\eta'\, \eta'^2}{1 + k^2\eta'^2} = k\eta - \tan^{-1}(k\eta). \tag{6.21}$$

Hence,

$$\sqrt{\frac{\hbar}{2a^3 W_k}} \exp\left(-i \int^t dt' W_k(t')\right)$$

$$= H \sqrt{\frac{\hbar(1 + k^2 \eta^2)}{2k^3}} \exp(-ik\eta + i\tan^{-1}(k\eta))$$

$$= -i\phi_k^{\text{BD}}|_{\substack{m=0 \\ \xi=0}} \qquad (6.22)$$

by Eq. (5.4). The fact that the second order adiabatic mode functions $\tilde{\phi}_k$ of Eq. (6.3) coincide with the exact BD mode functions up to a constant phase for the massless, minimally coupled field in de Sitter space implies that the Bogoliubov coefficients α_k and β_k are time-independent. This may be verified explicitly by computing Eq. (6.8) with W_k and V_k given by Eqs. (6.20). Thus $\mathcal{N}_k$ may be identified with N_k, and there is no production of massless, minimally coupled scalar particles in the special case of exact de Sitter spacetime. This implies that there is also no phase decoherence of the free massless inflaton field in the de Sitter epoch, with respect to the adiabatic particle basis defined by Eqs. (6.13). In the next section we will corroborate the absence of decoherence for the massless inflaton by computing the decoherence functional directly. A comparison of this result with the results of earlier work, such as that of Ref. [47] is given in Appendix C.

To conclude this section we remark that the adiabatic particle number basis should provide an efficient approximation of the low energy semiclassical limit of the energy-momentum tensor. If we neglect the phase correlated bilinears,

$$\langle \tilde{a}_{\mathbf{k}} \tilde{a}_{-\mathbf{k}} \rangle = \sigma_k \sinh r_k \cosh r_k e^{-i\theta_k}, \qquad (6.23)$$

which should make a relatively small contribution to $\langle T_{ab} \rangle$ in the mode sum over k, compared to the terms involving $\mathcal{N}_k$, then the energy density becomes simply

$$\varepsilon \simeq \tilde{\varepsilon} \equiv \varepsilon_v + \frac{1}{2\pi^2 a^3} \int [dk] k^2 \varepsilon_k^{(2)} \mathcal{N}_k. \qquad (6.24)$$

This should provide a useful analytic approximation to the energy density in a general UV allowed RW state whenever the RW scale factor varies slowly enough for the higher order adiabatic corrections to $\varepsilon_k^{(2)}$ and $\mathcal{N}_k$, to be negligible.

The quantity $\varepsilon_k^{(2)}$ in Eq. (6.24), given by (4.14a) has the interpretation of the single particle energy in the time varying RW background. Because this second order single particle energy satisfies

$$\dot{\varepsilon}_k^{(2)} = -H\varepsilon_k^{(2)} - HT_k^{(2)}, \qquad (6.25)$$

where $T_k^{(2)}$ is given by Eq. (4.14b), and the finite vacuum contributions of Eqs. (4.15) are separately conserved, it follows that the conservation equation (4.9) is exactly satisfied, provided that the trace in the quasiparticle approximation corresponding to Eq. (6.24) is

$$T \simeq \tilde{T} \equiv T_v + \frac{1}{2\pi^2 a^3} \int [dk] k^2 T_k^{(2)} \mathcal{N}_k$$

$$- \frac{1}{2\pi^2 H a^3} \int [dk] k^2 \varepsilon_k^{(2)} \dot{\mathcal{N}}_k. \qquad (6.26)$$

Defining the ideal fluid pressure in this approximation to be that in the absence of any particle creation, i.e.,

$$\tilde{p} \equiv \frac{\varepsilon_v + T_v}{3} + \frac{1}{6\pi^2 a^3} \int [dk] k^2 (\varepsilon_k^{(2)} + T_k^{(2)}) \mathcal{N}_k, \qquad (6.27)$$

the conservation equation (4.9) becomes then

$$\dot{\tilde{\varepsilon}} + 3H(\tilde{\varepsilon} + \tilde{p}) = \frac{1}{2\pi^2 a^3} \int [dk] k^2 \varepsilon_k^{(2)} \dot{\mathcal{N}}_k, \qquad (6.28)$$

after transposing the last term of Eq. (6.26) to the right-hand side of Eq. (6.28). Thus in the quasiparticle limit where it is valid to make the replacements (6.24) and (6.26), the term involving $\dot{\mathcal{N}}_k$ carries the interpretation of the rate of heat dissipation per unit volume due to the nonconservation of adiabatic particle number $\mathcal{N}_k$. If the particles are in quasistationary local thermodynamic equilibrium at the slowly varying effective temperature $T_{\text{eff}}(t)$, then this rate of heat dissipation may be equated to T_{eff} times the rate of effective entropy density generation s_{eff},

$$T_{\text{eff}} \frac{ds_{\text{eff}}}{dt} = \frac{1}{2\pi^2 a^3} \int [dk] k^2 \varepsilon_k^{(2)} \dot{\mathcal{N}}_k, \qquad (6.29)$$

by the first law of thermodynamics. The effective entropy generation gives rise to an effective bulk viscosity in the energy-momentum tensor due to particle creation, even in the absence of self-interactions of the quasiparticles.

Let us emphasize that the entropy generation and bulk viscosity are only *effective*, i.e., an approximation valid only to the extent that the phase information contained in the off-diagonal elements of the exact density matrix (3.10) in the adiabatic particle basis cannot be recovered. Likewise the correlations (6.23), which also depend on the rapidly varying phases θ_k conjugate to $\mathcal{N}_k$ should make a negligible contribution to macroscopic physical quantities. If the exact phase information is retained, then the evolution remains unitary, as required by the equivalence of the evolution to that of the effective classical Hamiltonian (A10). However, the extension of the usual description of the cosmological fluid by nonideal terms is suggested by the adiabatic particle creation rate and phase averaging in the low energy EFT. This may provide a useful phenomenological description in some circumstances, and also shows the approximations which are necessary in principle to pass from the fully reversible field theory description of matter in cosmological spacetimes to an effective, irreversible kinetic theory with a definite arrow of time.

VII. DECOHERENCE

The density matrix description of the evolution of arbitrary RW states in Sec. III allows us to describe the quantum to classical transition, i.e., decoherence, in a cosmological setting. The fundamental quantity of interest is the decoherence functional between two different histories, Γ_{12} or $\tilde{\Gamma}_{12}$ [35], given by Eqs. (3.26) or (3.27), of the pure or mixed state cases, respectively. Evaluating the Gaussian integrals needed to compute $\text{Tr}(\hat{\rho}_1\hat{\rho}_2)$ with the measure (3.18) gives

$$\tilde{\Gamma}_{12} = \frac{1}{4\pi^2}\int[dk]k^2\ln\left\{\frac{1}{4}\left(\sigma_{1k}\frac{\zeta_{2k}}{\zeta_{1k}} + \sigma_{2k}\frac{\zeta_{1k}}{\zeta_{2k}}\right)^2 + \frac{(\zeta_{1k}\pi_{2k} - \zeta_{2k}\pi_{1k})^2}{\hbar^2} + \frac{(\sigma_{1k} - \sigma_{2k})^2}{4}\right\}, \quad (7.1)$$

where $\{\zeta_{1k}, \pi_{1k}; \sigma_{1k}\}$ and $\{\zeta_{2k}, \pi_{2k}; \sigma_{2k}\}$ are the Gaussian state parameters of the two different histories. This simplifies somewhat in the case of two different pure state histories,

$$\text{Im}\,\Gamma_{12} = \tilde{\Gamma}_{12}\big|_{\sigma_{k1}=\sigma_{k2}=1}$$
$$= \frac{1}{4\pi^2}\int[dk]k^2\ln\left\{\frac{1}{4}\left(\frac{\zeta_{2k}}{\zeta_{1k}} + \frac{\zeta_{1k}}{\zeta_{2k}}\right)^2 + \frac{(\zeta_{1k}\pi_{2k} - \zeta_{2k}\pi_{1k})^2}{\hbar^2}\right\}. \quad (7.2)$$

In this pure state case the decoherence functional may be expressed in terms of one complex frequency function,

$$\Upsilon_k \equiv \frac{\hbar}{2\zeta_k^2} - \frac{i\pi_k}{\zeta_k} \quad (7.3a)$$

$$= -ia^3\frac{\dot{\phi}_k^*}{\phi_k^*} - 6i\xi Ha^3, \quad (7.3b)$$

where the last relation is derived in Appendix A. It is straightforward to rewrite Eq. (7.2) in the two alternate forms,

$$\text{Im}\Gamma_{12} = \frac{1}{4\pi^2}\int[dk]k^2\ln\left\{\frac{|\Upsilon_{1k} + \Upsilon_{2k}^*|^2}{4(\text{Re}\,\Upsilon_{1k})(\text{Re}\,\Upsilon_{2k})}\right\} \quad (7.4a)$$

$$= \frac{1}{4\pi^2}\int[dk]k^2\ln\{|a_1^{3-6\xi}(a_1^{6\xi}\phi_{1k})\dot{\phi}_{2k}^* - a_2^{3-6\xi}(a_2^{6\xi}\phi_{2k}^*)\dot{\phi}_{1k}|^2/\hbar^2\}. \quad (7.4b)$$

Because of the infinite product of integrations in the functional measure (3.18), it is clear that some condition(s) will have to be imposed on the two states or density matrices in these expressions, in order to insure a convergent result for large k. The otherwise ill-defined divergent nature of (3.26) has been noted by several authors [34–36]. The divergences in the decoherence functional are similar to those encountered in the unrenormalized expressions for the energy-momentum tensor components, Eqs. (4.8). In that case, the superficial degree of divergence was quartic,

whereas that of Eq. (7.1) or (7.2) is reduced by one power of k, and can be no more than cubically divergent at large k. A method to handle this cubic and subleading linear cutoff dependence of the decoherence functional (7.1) or (7.4) is needed before meaningful results in the low energy EFT description can be obtained.

In order to study the cutoff dependence of the inner product and decoherence functional, let us write

$$\Upsilon_{1k} = \Upsilon_k + \delta\Upsilon_k, \quad (7.5a)$$
$$\Upsilon_{2k} = \Upsilon_k - \delta\Upsilon_k, \quad (7.5b)$$

and expand the logarithm in Eq. (7.4a) to second order in $\delta\Upsilon_k$. We find that

$$\text{Im}\,\Gamma_{12} = \frac{1}{4\pi^2}\int[dk]k^2\frac{|\delta\Upsilon_k|^2}{(\text{Re}\,\Upsilon_k)^2} + \mathcal{O}(\delta\Upsilon_k)^4. \quad (7.6)$$

The leading behavior of $\text{Re}\,\Upsilon_k$ at large k may be read from the first order equation satisfied by this function, viz.,

$$\Upsilon_k^2 = ia^3\dot{\Upsilon}_k + a^6\left[\omega_k^2 + (6\xi - 1)\left(6\xi H^2 + \frac{\epsilon}{a^2}\right)\right] - 12i\xi Ha^3\Upsilon_k, \quad (7.7)$$

which is easily derived from the definition (7.3a) and the relations (3.30c) and (3.30d). From Eq. (7.7) we see that

$$\text{Re}\,\Upsilon_k \to ka^2\left[1 + \mathcal{O}\left(\frac{1}{k^2}\right)\right], \qquad k \to \infty. \quad (7.8)$$

Since $\delta\Upsilon_k$ is the same order as Υ_k generically at large k, the decoherence functional (7.6) will generally diverge as the UV cutoff is removed. Indeed the asymptotic behavior given by Eq. (7.8) implies at large k that

$$\delta\Upsilon_k \to 2ka(\delta a) + \dots \quad (7.9)$$

so that (7.6) will diverge cubically in the cutoff k_M, unless $\delta a \equiv 0$ identically. Hence no meaningful comparison between two different RW geometries in the low energy EFT can be made through the decoherence functional.

This divergent short distance behavior of the decoherence functional could have been anticipated from the relationship between Γ_{12} and the closed time path [37] action functional. Variation of the CTP action functional with respect to the metric, g_{ab}, produces connected correlation functions of the energy-momentum tensor with particular time orderings of their arguments. The first variation is the same as that of $S_{\text{eff}}[g]$ in Eq. (4.1), and produces the stress tensor expectation value (4.3). The second order variation (7.6) is formally proportional to

$$\text{Im}\int_0^t d^4x\sqrt{-g}\int_0^t d^4x'\sqrt{-g'}\,\delta g_{ab}(x)\Pi^{abcd}(x, x')\delta g_{cd}(x'). \quad (7.10)$$

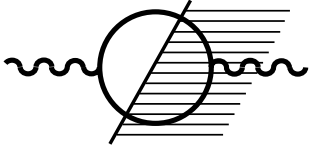The symmetrized expectation value of stress tensors,

FIG. 2. The imaginary part of the one-loop vacuum polarization given by Eqs. (7.10) and (7.11), which enters the decoherence functional at second order in the metric variation, δg_{ab} (represented by the wavy lines). The shaded part of the diagram represents the complex conjugation of the unshaded part, and the resulting squared amplitude is proportional to the probability for the creation of a scalar particle/antiparticle pair, represented by the diagonal cut through the diagram.

$$\mathrm{Im}\,\Pi^{abcd}(x.x') = \frac{1}{2}\langle T^{ab}(x)T^{cd}(x') + T^{cd}(x')T^{ab}(x)\rangle \tag{7.11}$$

is proportional to the noise kernel of fluctuations around the mean $\langle T^{ab}\rangle$ [60]. In flat spacetime, where one may transform conveniently to the momentum representation (7.10) is proportional to the cut in the one-loop polarization diagram of Fig. 2, corresponding to the squared matrix element for the creation of particle pairs by the perturbation δg_{ab}. We shall be particularly interested in the specific homogeneous variation, $\delta g_{ab} = 2(\delta a/a)g_{ab}$ which preserves the RW symmetries.

In coordinate space Eq. (7.11) is a singular distribution at coincident points $x = x'$, involving in general up to four derivatives of $\delta^{(4)}(x, x')$. Thus, the second variation (7.10) exists only for metric variations, δg_{ab}, which fall off rapidly enough in both space and time to permit integration by parts of the derivatives of $\delta^{(4)}(x, x')$. Since the time interval $[0, t]$ is finite in Eq. (7.10), the surface terms generated by these integrations by parts do not vanish at the end points in general, and can generate up to two derivatives of delta functions at equal spacetime arguments, i.e., formal cubic (and subleading linear) divergences in the decoherence functional, which is just what is obtained in from Eqs. (7.6), (7.8), and (7.9).

The unrenormalized decoherence functional is not only cutoff dependent in general but also ambiguous in that it depends upon the precise definition of the local measure (3.18) [36], which does not affect the physical content of the evolution described by the density matrix (3.10). In fact, had we used the $\tilde{q}_\mathbf{k} = aq_\mathbf{k}$ field to parametrize the density matrix rather than $q_\mathbf{k}$, then in order to normalize the state properly the measure would have to be replaced by a product of the $d\tilde{q}_\mathbf{k}$. This differs from Eq. (3.18) by an *infinite* number of local factors of $a(\eta)$. Hence, we should expect the inner product and decoherence functional defined by this measure on the field configuration space to differ from the previous one by cutoff dependent contact terms. Indeed, the frequency function Υ_k would be modified to

$$\tilde{\Upsilon}_k \equiv \frac{\Upsilon_k}{a^2}, \tag{7.12}$$

for which the leading term in the large k limit (7.9) cancels. The explicit form of $\delta\Upsilon_k$ is given below by Eq. (7.22), with the result that the decoherence functional (7.4a) generally diverges cubically as the comoving momentum cutoff $k_M \to \infty$. Because of the different large k behavior of $\tilde{\Upsilon}_k$, if the same steps leading to Eq. (7.22) are carried out for $\tilde{\Upsilon}_k$ instead, the resulting expression lacks the last term in Eq. (7.22), and for that reason its contribution to the decoherence functional is only linearly divergent [34,36]. This shows that the degree of divergence is dependent on the field parametrization and the definition of the inner product (3.26) on the field configuration space, which should not have any physical consequences for decoherence due to a slowly evolving geometry in the low energy EFT.

Because of the appearance of divergences of odd powers, associated with the boundaries of the region of integration, the divergences in the decoherence functional cannot be removed by renormalization of the bulk terms in the low energy effective action (4.1). Instead their renormalization requires introducing counterterms in the effective action, of dimension one and three, which are strictly localized on the boundaries. Adiabatic regularization may be used to define the necessary subtractions of the decoherence functional, corresponding to renormalization of these boundary terms, in a way quite analogous to the adiabatic subtractions used to define the renormalized energy-momentum tensor [38]. Since terms in the effective action up to dimension three are involved, we define the renormalized decoherence functional by subtracting from its unrenormalized value the adiabatic expansion of the functional up to and including its *third* adiabatic order asymptotic expansion for large k. Since the decoherence functional involves an absolute square, this requires subtracting only up to the *first* adiabatic order in the expression inside the absolute value signs in Eq. (7.4). That is, we define the renormalized decoherence functional by

$$\mathrm{Im}\,\Gamma_{12}^{(R)} \equiv \frac{1}{4\pi^2}\int[dk]k^2\left|\frac{\delta\Upsilon_k^*}{\mathrm{Re}\,\Upsilon_k} - \left(\frac{\delta\Upsilon_k^*}{\mathrm{Re}\,\Upsilon_k}\right)_1\right|^2, \tag{7.13}$$

where the subscript 1 denotes the expansion of the quantity in parentheses up to and including first order in its adiabatic expansion. This leaves the leading unsubtracted behavior at large k to be second adiabatic order under the absolute value signs, and fourth adiabatic order in its square in $\tilde{\Gamma}_{12}^{(R)}$. This corresponds to subtracting all surface divergences up to and including third adiabatic order in the CTP action functional, which is what is required.

To carry this out explicitly we begin by rewriting Eq. (7.3b) in conformal time in the form,

$$\Upsilon_k^* = ia^2\frac{\phi_k'}{\phi_k} + 6i\xi a'a = ia^2\frac{\chi_k'}{\chi_k} + i(6\xi - 1)a'a, \tag{7.14}$$

and computing its first variation,

$$\delta Y_k^* = 2ia\delta a \frac{\chi_k'}{\chi_k} + ia^2 \frac{\delta \chi_k'}{\chi_k} - ia^2 \frac{\chi_k' \delta \chi_k}{\chi_k^2}$$
$$+ i(6\xi - 1)\delta(a'a). \quad (7.15)$$

The variations of χ_k and its derivative are computed by varying Eq. (2.14), to obtain

$$\delta \chi_k'' + \left[k^2 + m^2 a^2 + (6\xi - 1)\left(\frac{a''}{a} + \epsilon\right) \right]\delta \chi_k$$
$$= -\left[m^2 \delta a^2 + (6\xi - 1)\delta\left(\frac{a''}{a}\right) \right]\chi_k, \quad (7.16)$$

which is solved in terms of the retarded Green's function of the differential operator on the left-hand side,

$$G_R(\eta, \eta') = \frac{i}{\hbar}(\chi_k(\eta)\chi_k^*(\eta') - \chi_k^*(\eta)\chi_k(\eta'))\Theta(\eta - \eta')$$
$$(7.17)$$

in the form,

$$\delta\chi_k(\eta) = -\int d\eta' G_R(\eta, \eta')\left[m^2 \delta a^2 \right.$$
$$\left. + (6\xi - 1)\delta\left(\frac{a''}{a}\right) \right]_{\eta'}\chi_k(\eta')$$
$$= \delta\alpha_k(\eta)\chi_k(\eta) + \delta\beta_k(\eta)\chi_k^*(\eta), \quad (7.18)$$

where

$$\delta\alpha_k(\eta) = -\frac{i}{\hbar}\int^\eta d\eta'\left[m^2 \delta a^2 + (6\xi - 1) \right.$$
$$\left. \times \delta\left(\frac{a''}{a}\right) \right]_{\eta'}|\chi_k(\eta')|^2, \quad (7.19a)$$

$$\delta\beta_k(\eta) = \frac{i}{\hbar}\int^\eta d\eta'\left[m^2 \delta a^2 + (6\xi - 1) \right.$$
$$\left. \times \delta\left(\frac{a''}{a}\right) \right]_{\eta'}[\chi_k(\eta')]^2. \quad (7.19b)$$

The lower limits of the integrals in Eq. (7.19) depend on the initial conditions of the wave functional and give a time-independent phase in the decoherence functional below which we shall not need to specify. Substituting Eq. (7.18) into (7.15) and using the Wronskian condition (2.17), we find that the $\delta\alpha_k$ term cancels and we are left with

$$\delta Y_k^* = 2ia\delta a \frac{\chi_k'}{\chi_k} - \frac{\hbar a^2}{\chi_k^2}\delta\beta_k + i(6\xi - 1)\delta(a'a). \quad (7.20)$$

Using Eq. (2.17) again, we have

$$\text{Re } Y_k = \frac{\hbar a^2}{2|\chi_k|^2}, \quad (7.21)$$

so that we secure finally,

$$\frac{\delta Y_k^*}{\text{Re}Y_k} = \frac{2i\chi_k^*}{\chi_k}\left(i\delta\beta_k + \frac{(6\xi - 1)}{\hbar}\delta\left(\frac{a'}{a}\right)\chi_k^2 \right.$$
$$\left. + \frac{a\delta a}{\hbar}(a^{12\xi}\phi_k^2)'a^{-12\xi} \right). \quad (7.22)$$

This form is valid for a scalar field of arbitrary mass and curvature coupling in a general RW spacetime. In order to renormalize it, we should subtract its asymptotic expansion up to adiabatic order one and substitute the square of the subtracted quantity in Eq. (7.13). We carry out this subtraction explicitly in two important special cases, namely, when the mass $m = 0$ and the curvature coupling ξ takes on either its conformal or minimally coupled value, $\xi = 1/6, 0$, respectively.

Using Eq. (7.22) together with (7.19) and the form of Eq. (2.14) in the massless, conformally coupled case, we obtain

$$\frac{\delta Y_k^*}{\text{Re}Y_k}\bigg|_{\substack{m=0 \\ \xi=1/6}} = \frac{2i\chi_k^*}{\hbar\chi_k}\frac{\delta a}{a}(\chi_k^2)' = \frac{4k|\chi_k|^2}{\hbar}\frac{\delta a}{a} = 2\frac{\delta a}{a},$$
$$(7.23)$$

which involves no time derivatives and hence is clearly of adiabatic order zero. If we were to substitute this directly into Eq. (7.4a), we would obtain a cubically divergent decoherence functional. Since this cubic divergence can be removed by simply redefining the canonical variables as in Eqs. (7.12) and (A16), it is clear that it can have no physical significance. However, if we first subtract off the adiabatic order zero part (which in this case is the entire expression), then the renormalized decoherence functional (7.13) in an arbitrary RW spacetime is identically *zero* for the massless, conformally coupled field. This lack of decoherence corresponds to the lack of particle creation for this field in any RW space which we have found in the previous section.

In the massless, minimally coupled case the corresponding expression is

$$\frac{\delta Y_k^*}{\text{Re}Y_k}\bigg|_{\substack{m=0 \\ \xi=0}} = \frac{2i\chi_k^*}{\chi_k}\int^\eta d\eta'\delta\left(\frac{a''}{a}\right)_{\eta'}[\chi_k(\eta')]^2$$
$$- \frac{2i}{\hbar}\delta\left(\frac{a'}{a}\right)|\chi_k|^2 + \frac{2i\chi_k^*}{\hbar\chi_k}a\delta a(\phi_k^2)', \quad (7.24)$$

which is generally nonzero. Note that the first term in this last expression depends upon the variation of the RW scale factor over the entire evolution from an arbitrary (unspecified) initial time at the lower limit of the η' integral to the final time η, while the last two terms of Eq. (7.24) depend upon the variation of the scale factor only at the final time. It is these two latter surface terms that generate divergences in the unsubtracted decoherence functional (7.6), when integrated over k. Such surface terms arise in the covariant expression (7.10) if the conservation of T^{ab} is used to express the tensorial noise correlator (7.11) in terms of

covariant derivatives of scalar quantities, and then an integration by parts is performed.

Specializing to de Sitter spacetime and using the form of the BD mode functions (5.3), the last term in Eq. (7.24) becomes

$$\frac{2i\chi_k^*}{\hbar\chi_k}a\delta a(\phi_k^2)'|_{dS} = 2\frac{\delta a}{a}\left(1 + \frac{i}{k\eta}\right). \qquad (7.25)$$

Here the first term is of adiabatic order zero as in the previous conformally coupled case, while the second term is of adiabatic order one. Hence both terms are fully subtracted in the renormalized decoherence functional. The other contact term becomes in de Sitter space,

$$-\frac{2i}{\hbar}\delta\left(\frac{a'}{a}\right)|\chi_k|^2|_{dS} = -\frac{i}{k}\delta\left(\frac{a'}{a}\right)\left(1 + \frac{1}{k^2\eta^2}\right), \qquad (7.26)$$

which consists of an adiabatic order one and order three term. Hence subtracting up to adiabatic order one removes the first term but leaves the $1/k^2\eta^2$ term unaffected. Since the first term in Eq. (7.24) involves two time derivatives of the scale factor or its variation, it is adiabatic order two, and likewise unaffected by the subtraction of up to adiabatic order one terms. Hence finally,

$$\begin{aligned}
\text{Im}\Gamma_{12}^{(R)}\big|_{{}^{m=0}_{\xi=0\,dS}} &= \frac{1}{4\pi^2}\int dk k^2 \left|\frac{\delta Y_k^*}{\text{Re}Y_k} - \left(\frac{\delta Y_k^*}{\text{Re}Y_k}\right)_1\right|^2\bigg|_{{}^{m=0}_{\xi=0\,dS}}\\
&= \frac{1}{4\pi^2}\int dk \left|\int^\eta d\eta' e^{2ik(\eta-\eta')}\left(1 - \frac{i}{k\eta'}\right)^2\right.\\
&\quad \times \delta\left(\frac{a''}{a}\right)_{\eta'} - \frac{1}{k^2\eta^2}\left(\frac{k\eta-i}{k\eta+i}\right)\delta\left(\frac{a'}{a}\right)\bigg|^2, \quad (7.27)
\end{aligned}$$

which is UV finite and nonzero for arbitrary variations of the scale factor.

If we consider the particular variation of the scale factor in which the de Sitter Hubble parameter H is varied in Eq. (7.27), while the conformal time η is held *fixed*, then $a'/a = -1/\eta$ and $a''/a = 2/\eta^2$ are fixed and Eq. (7.27) vanishes. Hence we find that under variations of the de Sitter curvature, which compare the wave functionals of the quantum field in macroscopically different de Sitter universes but at the same conformal time, the decoherence functional for the massless, minimally coupled field *vanishes* identically.

The ambiguous contact terms which are field parametrization dependent are removed by the adiabatic regularization and renormalization procedure in Eq. (7.13), as the two cases considered explicitly above show. Hence, Eq. (7.13) yields both a finite decoherence functional free of unphysical dependence on the short distance cutoff, and one that is independent of redefinitions of the scalar field variables and inner product. The renormalized decoherence functional proposed here vanishes in the two special massless cases of the conformally coupled field in a general RW background and the minimally coupled field in a

de Sitter background, the same two cases studied in the previous section where the adiabatic particle creation rate $\dot{\mathcal{N}}_k = 0$. Since the imaginary part of the polarization tensor (7.11) is just the cut one-loop diagram shown in Fig. 2, which is proportional to the probability for the metric fluctuation to create a particle/antiparticle pair from the vacuum, a close correspondence between the lack of particle creation and a vanishing decoherence functional for homogeneous metric perturbations in the Hubble parameter is not unexpected. The fact that the adiabatic subtraction procedure for the decoherence functional proposed here supports this correspondence suggests that it is the correct one to define a finite decoherence functional for semiclassical cosmology. In order to prove that this is indeed the *unique* procedure for defining a physical decoherence functional in the EFT approach, the adiabatic subtractions of first and third order should be related to definite boundary counterterms in the CTP effective action which reside exclusively on the surfaces at the initial and final times. These boundary terms may be related to those found recently by the authors of Ref. [19]. We leave the determination of these surface terms for a future investigation.

VIII. SUMMARY AND CONCLUSIONS

The principal objective of this paper has been to place semiclassical cosmology within a consistent EFT framework, in which possible short distance effects can be parametrized by well-defined initial conditions at the onset of inflation. Although the general principles and adiabatic methods underlying such an EFT framework have been available for some time, we have thought it worthwhile to make these assumptions completely explicit in this paper, and demonstrate how they can be applied and extended in a number of different ways, which may be useful for future cosmological models and observations. Because of the several different applications considered in the paper, we collect here and summarize the principal results, together with the relevant equations and sections where each point is discussed in detail.

(1) The general homogeneous, isotropic RW pure state is defined by field amplitudes ϕ_k obeying (2.11), which are linear combinations of vacuum modes (2.19) with Bogoliubov coefficients satisfying (2.20).

(2) These pure RW states are squeezed vacuum states annihilated by $a_{\mathbf{k}}$ in the mode expansion (2.6) and specified by two real time-independent squeezing parameters (2.26), up to an overall irrelevant phase.

(3) The wave functionals of these pure RW states are Gaussians (3.22) in the Schrödinger picture field coordinate basis.

(4) The general RW state with a nonzero occupation number (2.23) evolves as a mixed state described by the Gaussian density matrix (3.10) in the coordinate basis.

(5) The general RW mixed state requires three independent functions of k, $(\zeta_k, \pi_k; \sigma_k)$, which are related to the mode function ϕ_k by (3.6) or (3.7), and determine the three equal time symmetrized correlators of the field by (3.8) in the Hamiltonian description.

(6) The first two of these functions of k, (ζ_k, π_k) are time-dependent and form a canonically conjugate pair for the unitary evolution of the density matrix (3.29) and (3.30), described by the effective classical Hamiltonian (A10)–(A12), in which $\hbar$ appears as a parameter.

(7) The third function of k, $\sigma_k = 2n_k + 1$ is strictly a constant of the motion.

(8) The form of the Hamiltonian of the scalar field evolution in a fixed RW background depends on the field parametrization, and in general is not equal to the covariant energy density $\varepsilon = T_{tt}$.

(9) The covariant and Hamiltonian descriptions of the evolution are completely equivalent nonetheless, and the total Hamiltonian of the combined matter plus geometry system (3.34) vanishes by time reparametrization invariance for evolutions satisfying the classical Friedmann equation.

(10) The power spectrum of scalar field fluctuations in the general homogeneous, isotropic, mixed RW state is given by Eqs. (2.32).

(11) The spectrum of the actual scalar metric fluctuations observed in the CMB are dependent upon additional parameters which may be different for different inflationary models. An example is the dependence on the slow roll parameter ϵ in Eq. (2.34).

(12) The energy density and trace of the stress tensor in the general RW state is given by Eqs. (4.16), with ε_v and T_v the values in the fiducial vacuum state.

(13) In order to be a UV allowed RW state with short distance behavior consistent with general covariance of the low energy EFT and the equivalence principle, the fiducial vacuum state and all other physical states must be fourth order adiabatic states.

(14) Any modification of the fourth order adiabaticity condition at short distances has the potential to disturb the conservation of $\langle T_{ab} \rangle$, and/or violate the equivalence principle at arbitrarily large distances and late times, which would also violate the decoupling hypothesis of low energy EFT.

(15) The Bunch-Davies state is a UV allowed fourth order adiabatic state, which is also invariant under the full $O(4, 1)$ isometry group of global de Sitter spacetime.

(16) The general one complex parameter squeezed α states of the scalar field in de Sitter space are *not* UV allowed fourth order adiabatic states (even for non-self-interacting scalar fields), except for the single value of the parameter corresponding to the BD state.

(17) Because all UV allowed states are fourth order adiabatic, their power spectrum approaches that of the BD state for sufficiently large comoving wave numbers $k > k_M$, and sufficiently late times $t > t_M$ after the onset of inflation (5.8), when EFT methods should apply.

(18) As a consequence of this kinematic effect of the expansion, any modifications of the power spectrum due to initial state effects require a coincidence of fine-tuning (5.9) in order to be observable in the CMB today.

(19) Assuming such fine-tuning and cutting off the squeezed α state at a finite large comoving momentum scale k_M produces potentially observable scale dependent modifications of the CMB power spectrum (5.17), whose magnitude depends in general upon additional parameters of the inflationary model.

(20) Cutoff α states and nonadiabatic states generally produce the largest backreaction contributions during the inflationary epoch, given by (5.18), which are of order (5.20).

(21) States which are adiabatic order zero up to the cutoff scale k_M produce scale dependent modifications of the CMB power spectrum (5.26) which may be observable as modulations in the CMB power spectrum.

(22) Such states also produce backreaction effects during inflation which are somewhat smaller in amplitude than the cutoff α states, and which can be calculated exactly from Eqs. (5.10), (5.11), and (5.27).

(23) The modifications of the initial state given by the addition of a local higher dimension operator with coefficient β in the boundary action approach are nonadiabatic and yield in general modifications to the CMB power spectrum at linear order, Eqs. (5.36) and (5.37) in β.

(24) The backreaction contributions to the stress tensor during inflation are also linear in β in general and of order βM^4, which may be significant, depending on the cutoff and inflation scales M and H, but do not disturb inflation if (5.40) is satisfied.

(25) The adiabatic expansion of the stress tensor can be used also to define a time-dependent particle number basis for particles created by the RW expansion, (6.7), with parameters W_k and V_k defined by Eqs. (4.14), (6.12), and (6.13) matched to the stress tensor exactly at second adiabatic order.

(26) The total particle number defined in this way is the minimal one that is finite, (6.16), and gives separately conserved vacuum and particle contributions to the covariant stress tensor.

(27) In the general massive case the particle number is not conserved but is a sixth order adiabatic invariant, implying that the density matrix in the adiabatic particle representation has slowly varying diagonal components but much more rapidly varying off-diagonal components.

(28) Although the exact evolution is unitary and fully reversible, averaging over the rapidly varying off-diagonal elements of the density matrix in this basis gives rise to an approximation which is effectively dissipative, and in which the effective von Neumann entropy of the reduced density matrix (6.10) may increase with time.

(29) Neglect of these same phase correlations in the energy-momentum tensor via the approximations (6.24) and (6.27) gives an effective rate of heat dissipation due to particle creation (6.28) and (6.29), even in the absence of matter self-interactions.

(30) Notable exceptions to this dephasing occur in several special massless cases, due to the absence of particle creation for a conformally invariant scalar field in any RW spacetime, and for a massless, minimally coupled scalar field in de Sitter spacetime.

(31) The latter result implies that the quantum phase information in the density perturbations derived in slow roll inflationary models is *not* washed out by the expansion alone, so that the loss of phase decoherence in such models must be due to other effects not considered in this paper.

(32) The decoherence functional for arbitrary mixed Gaussian states given by Eq. (7.1) is related to the noise kernel, or imaginary part of the second variation CTP effective action (7.10).

(33) The adiabatic method may be used again to define the renormalized decoherence functional (7.13) in semiclassical cosmology, which is independent of the short distance cutoff and field reparametrizations.

(34) This renormalized decoherence functional vanishes in the special cases where there is no adiabatic particle creation, corroborating the close connection between particle creation, dephasing and decoherence.

(35) Comparison of the decoherence functional defined here with a previous result in the massless case is given in Appendix C.

(36) Verifying this adiabatic subtraction through a covariant subtraction of the surface terms in the CTP action functional would allow the study of decoherence effects and the quantum to classical transition quantitatively and reliably in general RW cosmologies for the first time.

APPENDIX A: EQUATIONS FOR ζ_k AND π_k AND THEIR EFFECTIVE HAMILTONIAN

In this appendix we compute the three independent and symmetric Gaussian variances $\langle \Phi^2 \rangle$, $\langle \Phi\Pi_\phi + \Pi_\phi\Phi \rangle$ and $\langle \Pi_\phi^2 \rangle$, and derive the equations of motion for the density matrix parameters ζ_k and π_k defined in Sec. III.

The square of the defining relation for ζ_k in Eq. (3.6) is

$$\sigma_k |\phi_k|^2 = \zeta_k^2. \tag{A1}$$

Hence using Eq. (2.29) and evaluating the sums or integrals over the angular part of **k**, we find directly for the first variance at coincident spacetime points,

$$\langle \Phi^2 \rangle = \frac{1}{2\pi^2} \int [dk] k^2 \sigma_k |\phi_k|^2 = \frac{1}{2\pi^2} \int [dk] k^2 \zeta_k^2, \tag{A2}$$

which is explicitly real and independent of **x**. In order to compute the second variance we differentiate Eq. (A1) to obtain

$$\frac{\sigma_k}{2}(\phi_k \dot{\phi}_k^* + \dot{\phi}_k \phi_k^*) = \sigma_k \operatorname{Re}(\phi_k \dot{\phi}_k^*) = \zeta_k \dot{\zeta}_k, \tag{A3}$$

which is the second of relations (3.7). Hence, using Eq. (3.3) the second symmetric variance at coincident points is

$$\begin{aligned}
\langle \Phi\Pi_\Phi + \Pi_\Phi\Phi \rangle &= a^3 \langle \Phi\dot{\Phi} + \dot{\Phi}\Phi + 12\xi H \Phi^2 \rangle \\
&= \frac{a^3}{\pi^2} \int [dk] k^2 \sigma_k [\operatorname{Re}(\phi_k \dot{\phi}_k^*) \\
&\quad + 6\xi H |\phi_k|^2] \\
&= \frac{1}{\pi^2} \int [dk] k^2 \zeta_k \pi_k.
\end{aligned} \tag{A4}$$

By squaring Eq. (A3) and using the Wronskian condition (2.17) in

$$\begin{aligned}
(\phi_k \dot{\phi}_k^* + \dot{\phi}_k \phi_k^*)^2 &= 4|\phi_k|^2 |\dot{\phi}_k|^2 + (\phi_k \dot{\phi}_k^* - \dot{\phi}_k \phi_k^*)^2 \\
&= 4|\phi_k|^2 |\dot{\phi}_k|^2 - \frac{\hbar^2}{a^6},
\end{aligned} \tag{A5}$$

we obtain

$$\sigma_k |\dot{\phi}_k|^2 = \dot{\zeta}_k^2 + \frac{\hbar^2 \sigma_k^2}{4a^6 \zeta_k^2}, \tag{A6}$$

which is the third member of Eq. (3.7). Hence the third Gaussian variance is

$$\langle \Pi_\Phi^2 \rangle = a^6 \langle \dot\Phi^2 + 6\xi H(\Phi\dot\Phi + \dot\Phi\Phi) + 36\xi^2 H^2 \Phi^2 \rangle$$

$$= \frac{a^6}{2\pi^2} \int dk k^2 \sigma_k [|\dot\phi_k|^2 + 12\xi H \, \mathrm{Re}(\phi_k \dot\phi_k^*) + 36\xi^2 H^2 |\phi_k|^2]$$

$$= \frac{a^6}{2\pi^2} \int dk k^2 \left[\dot\zeta_k^2 + \frac{\hbar^2 \sigma_k^2}{4a^6 \zeta_k^2} + 12\xi H \zeta_k \dot\zeta_k + 36\xi^2 H^2 \zeta_k^2 \right]$$

$$= \frac{1}{2\pi^2} \int dk k^2 \left(\pi_k^2 + \frac{\hbar^2 \sigma_k^2}{4\zeta_k^2} \right). \tag{A7}$$

This establishes Eqs. (3.8) of Sec. III.

The second order differential equation for ζ_k may be derived by differentiating (A3), and making use of Eqs. (2.11) and (A6) to obtain

$$\ddot\zeta_k + 3H\dot\zeta_k + \left(\frac{k^2 - \epsilon}{a^2} + m^2 + \xi R \right) \zeta_k = \frac{\hbar^2 \sigma_k^2}{4a^6 \zeta_k^3}, \tag{A8}$$

which is Eq. (3.32) of the text.

The equations for the parameters of the Gaussian density matrix may be compared with those arising from the purely classical Hamiltonian (3.4), viz.,

$$\dot\Phi = \frac{\Pi_\Phi}{a^3} - 6\xi H \Phi, \tag{A9a}$$

$$\dot\Pi_\Phi = 6\xi H \Pi_\Phi - a^3 \left[\frac{-\Delta_3 + \epsilon}{a^2} + m^2 + (6\xi - 1) \right.$$

$$\left. \times \left(\frac{\epsilon}{a^2} + 6\xi H^2 \right) \right] \Phi. \tag{A9b}$$

Note, in particular, that for $\hbar \neq 0$ the equation of motion for π_k (3.30d) differs from Eq. (A9b) of the purely classical evolution by the last centrifugal barrierlike term in Eq. (3.30d) which is a result of the uncertainty principle being enforced on the initial data through the Wronskian condition (2.17).

The first order evolution equations for the parameters $(\bar\phi, \bar p; \zeta_k, \pi_k; \sigma_k)$ may also be regarded as Hamilton's equations for the effective *classical* Hamiltonian,

$$H_{\mathrm{eff}}[\bar\phi, \bar p; \{\zeta_k, \pi_k; \sigma_k\}] = \mathrm{Tr}(\mathcal{H}_\Phi \hat\rho) = \overline{\mathcal{H}}_\Phi(\bar\phi, \bar p)$$

$$+ \frac{1}{2\pi^2} \int [dk] k^2 \mathcal{H}_k(\zeta_k, \pi_k; \sigma_k) \tag{A10}$$

where

$$\overline{\mathcal{H}}_\Phi(\bar\phi, \bar p) = \frac{\bar p^2}{2a^3} - 6\xi H \bar p \, \bar\phi + \frac{a^3}{2} \left[m^2 + 6\xi \frac{\epsilon}{a^2} \right.$$

$$\left. + 6\xi(6\xi - 1)H^2 \right] \bar\phi^2, \tag{A11}$$

is the classical Hamiltonian of the spatially independent mean values, $(\bar\phi, \bar p)$ and

$$\mathcal{H}_k(\zeta_k, \pi_k; \sigma_k) = \frac{\pi_k^2}{2a^3} - 6\xi H \pi_k \zeta_k + \frac{a^3}{2} \left[\omega_k^2 + (6\xi - 1) \right.$$

$$\left. \times \frac{\epsilon}{a^2} + 6\xi(6\xi - 1)H^2 \right] \zeta_k^2 + \frac{\hbar^2 \sigma_k^2}{8a^3 \zeta_k^2}, \tag{A12}$$

is the effective Hamiltonian describing the Gaussian fluctuations around the mean field for the Fourier mode k. It is straightforward then to verify that Hamilton's equations for this effective classical Hamiltonian (in which $\hbar$ appears as a parameter), viz.,

$$\dot{\bar\phi} = \frac{\partial \overline{\mathcal{H}}_\Phi}{\partial \bar p}, \tag{A13a}$$

$$\dot{\bar p} = -\frac{\partial \overline{\mathcal{H}}_\Phi}{\partial \bar\phi}, \tag{A13b}$$

$$\dot\zeta_k = \frac{\partial \mathcal{H}_k}{\partial \pi_k}, \tag{A13c}$$

$$\dot\pi_k = -\frac{\partial \mathcal{H}_k}{\partial \zeta_k}, \tag{A13d}$$

are identical with Eqs. (3.30) of the text. Hence ζ_k and π_k are conjugate variables with respect to the effective classical Hamiltonian (A10).

If we define the complex frequency Υ_k by (7.3a) of the text, then by differentiating that definition and using Eq. (3.30) we obtain its equation of motion (7.7). On the other hand Eq. (A3) with $\sigma_k = 1$, together with the Wronskian condition (2.17) imply

$$\hbar - 2ia^3 \zeta_k \dot\zeta_k = -2ia^3 \phi_k \dot\phi_k^*, \tag{A14}$$

so that dividing by $2\zeta_k^2 = 2\phi_k \phi_k^*$, and using the definition of π_k in Eq. (3.6) we obtain

$$\Upsilon_k = \frac{\hbar}{2\zeta_k^2} - \frac{i\pi_k}{\zeta_k} = -ia^3 \frac{\dot\phi_k^*}{\phi_k^*} - 6i\xi H a^3, \tag{A15}$$

which establishes Eq. (7.3b) of the text.

Finally we remark that $\mathcal{H}_\Phi$ depends on the choice of variables used to represent the scalar field. Indeed, if we choose the conformal field variable,

$$\chi = a\Phi, \tag{A16a}$$

instead of Φ and define the conjugate field momentum,

$$\Pi_\chi \equiv \frac{\partial S_{cl}}{\partial \chi'} = \chi' - \frac{a'}{a}\chi \qquad (m = 0, \xi = 0) \quad \text{(A16b)}$$

for the massless, minimally coupled field, the canonical Hamiltonian,

$$\mathcal{H}_\chi = \frac{1}{2}\Pi_\chi^2 + \frac{k^2}{2}\chi^2 + \frac{a'}{a}\chi\Pi_\chi = a\mathcal{H}_\Phi + \dot{a}\Phi\Pi_\Phi,$$
$$(m = 0, \xi = 0), \tag{A17}$$

differs from $\mathcal{H}_\Phi$ defined by Eq. (3.4), and neither is equal to the time component of the covariant stress tensor which couples to gravity for general m and ξ. This is to be expected since the canonical transformation from (Φ, Π_Φ) to (χ, Π_χ) is a time-dependent transformation, and neither Hamiltonian is a conserved quantity. This has the consequence that while every representation is physically equivalent, describing exactly the same physical time evolution, there is no spacetime or field coordinate independent meaning to the basis which diagonalizes the instantaneous canonical Hamiltonian in a particular set of coordinates [56], and no reason to prefer any such choice over any other as a physical particle basis.

APPENDIX B: EVALUATION OF INTEGRALS

In evaluating the integrals I_1 and I_2 in Eqs. (5.11) which contribute to the energy density and pressure of the massless, minimally coupled scalar field, we encounter integrals of the form,

$$\int_0^{k_M} dk\, k^{2n} \sin(2ku - \theta)$$
$$= (-)^n \frac{k_M^{2n+1}}{2^{2n}} F_\theta^{(2n)}(k_M u), \tag{B1a}$$

$$\int_0^{k_M} dk\, k^{2n+1} \cos(2ku - \theta)$$
$$= (-)^n \frac{k_M^{2n+2}}{2^{2n+1}} F_\theta^{(2n+1)}(k_M u). \tag{B1b}$$

Integrals of this kind are easily evaluated by successive differentiation with respect to u of the elementary integral,

$$\int_0^{k_M} dk \sin(2ku - \theta) = \frac{\cos\theta - \cos(2k_M u - \theta)}{2u}$$
$$= k_M F_\theta(k_M u), \tag{B2}$$

where

$$F_\theta(x) \equiv F_\theta^{(0)}(x) = \frac{\cos\theta - \cos(2x - \theta)}{2x}$$
$$= \frac{\sin x \sin(x - \theta)}{x}. \tag{B3}$$

Thus, in Eq. (B1) we have

$$F_\theta^{(p)}(x) = \frac{\partial^p}{\partial x^p}\left(\frac{\sin x \sin(x - \theta)}{x}\right), \tag{B4}$$

and, in the particular case $\theta = 0$,

$$F^{(p)}(x) \equiv F_0^{(p)}(x) = \frac{\partial^p}{\partial x^p}\left(\frac{\sin^2 x}{x}\right). \tag{B5}$$

We also define

$$F^{(-1)}(x) \equiv \int_0^x dy\, F(y) = \int_0^x dy\, \frac{\sin^2 y}{y}. \tag{B6}$$

For any θ and $p \geq 0$, $F_\theta^{(p)}(x)$ are damped oscillatory functions whose maxima occur at $x = 0$ for p odd and on the first oscillation for p even. The values of x and $F_\theta^{(p)}$ at the maximum are of order unity. The leading behavior as $x \to \infty$ is obtained by differentiating the oscillatory numerator only, i.e.,

$$F_\theta^{(p)}(x) \to \frac{1}{x}\frac{\partial^p}{\partial x^p}\sin x \sin(x - \theta) + \mathcal{O}\left(\frac{1}{x^2}\right). \tag{B7}$$

Thus the absolute value of $F_\theta^{(p)}(x)$ is bounded by $1/x$ for $x \gg 1$. Hence the maximum value of the integrals (B1) are of order k_M^{2n+1} and k_M^{2n+2} respectively for $u \sim k_M^{-1}$, while they behave like k_M^{2n} and k_M^{2n+1} respectively, multiplied by a rapidly oscillating function of $k_M u$, as $k_M u \to \infty$. The form of $F^{(4)}(x)$ as a function of x in the special case of $p = 4$ and $\theta = 0$ is given by Fig. 1 of the text.

APPENDIX C: COMPARISON OF SQUEEZING IN DIFFERENT BASES

In Sec. VI we found that in the second order adiabatic particle basis defined by Eqs. (6.13) there is no large squeezing and no particle creation for a massless, minimally coupled scalar field in exact de Sitter spacetime. In Sec. VII we corroborated the lack of true decoherence for this field. In this appendix we compare this result to earlier work, in particular, to Ref. [47], whose authors use a field amplitude "pointer basis."

Let us first consider a zeroth order adiabatic basis rather than the second order adiabatic basis defined by Eqs. (6.13). The zeroth order basis function is obtained by replacing W_k of Eq. (6.20) by $\omega_k = k/a_{dS}$, resulting in

$$\tilde{\phi}_k^{(0)} = \sqrt{\frac{\hbar}{2a^3\omega_k}}\exp\left(-i\int^t dt'\,\omega_k(t')\right)$$
$$= \frac{1}{a}\sqrt{\frac{\hbar}{2k}}\exp(-ik\eta) \tag{C1}$$

instead of Eq. (6.22). When the exact BD mode is expressed as a linear combination of these zeroth order adiabatic modes in the form of the Bogoliubov transformation,

$$\phi_k^{\mathrm{BD}} = \alpha_k^{(0)} \tilde{\phi}_k^{(0)} + \beta_k^{(0)} \tilde{\phi}_k^{(0)*}, \qquad (C2)$$

we now obtain the nontrivial time-dependent coefficients,

$$\alpha_k^{(0)} = \left(1 - \frac{i}{2k\eta}\right), \qquad (C3a)$$

$$\beta_k^{(0)} = -\frac{i}{2k\eta} \exp(-2ik\eta). \qquad (C3b)$$

Clearly in this adiabatic zero basis $\sinh r_k^{(0)} = |\beta_k^{(0)}| = 1/|2k\eta|$ which does not fall off fast enough at large k to be fourth order adiabatic, and which also approaches infinity as $\eta \to 0^-$.

The problem with the zeroth order adiabatic basis (C1) is that the particle number (6.16),

$$\frac{1}{2\pi^2} \int dk\, k^2 |\beta_k^{(0)}|^2 = \frac{1}{8\pi^2 \eta^2} \int dk \to \infty \qquad (C4)$$

is divergent at any finite time η. Corresponding to this linear divergence in the total number, the energy-momentum tensor of these "particles" is quadratically divergent at large k. Clearly this quadratic divergence is a residual divergence of the *vacuum* stress tensor and has nothing at all to do with physical particles, which are unambiguously well-defined in the UV limit $k \to \infty$. In that limit of wavelength much smaller than the horizon scale, the region of spacetime the modes sample may be approximated as flat, and the effects of the time variation of the geometry should be negligible. The η^{-2} factor in Eq. (C4) shows that this mismatch only grows more severe at late times, as the zeroth order basis (C1) becomes more and more different from the second order basis (6.22).

The quadratic vacuum divergences can be subtracted only by matching the energy-momentum tensor to second adiabatic order as in Eq. (6.13). When that is done the particle number is finite and no squeezing, phase decoherence or particle creation effect at all is obtained. Moreover, since the second order adiabatic modes (6.22) are already exact for the massless, minimally coupled scalar field in de Sitter space, any ambiguity in the particle concept at wavelengths of the order of the horizon scale is irrelevant here. Going to higher orders in the adiabatic expansion will not change the result obtained at second order.

The authors of Ref. [47] define a field pointer basis for the massless, minimally coupled field in de Sitter space. Comparing Eqs. (15), (19), and (47) of Ref. [47] with relations (C3) above, we find that the Bogoliubov coefficients, α_k and β_k of Eq. (15) in Ref. [47] are precisely *equal* to $\alpha_k^{(0)}$ and $\beta_k^{(0)}$ respectively, of the zeroth order adiabatic basis given by (C3). As we have seen, the squeezing parameter with respect to this basis, $\sinh r_k^{(0)} = 1/|2k\eta| \to \infty$ does become very large for superhorizon modes in the late time limit. The authors of Ref. [47] argue that the large squeezing in this basis leads to an effective decoherence of modes of the scalar field much larger than the de Sitter horizon.

It is clear that the squeezing is very much dependent on the basis in which it is computed. In the second order basis (6.22), determined by the structure of the short distance expansion of the covariantly conserved stress tensor through Eqs. (6.13), there is no mixing of positive and negative frequency modes, and no large squeezing of superhorizon modes. True decoherence of these modes should occur through other effects, such as those considered in Ref. [61] at the time these modes reenter the horizon.

[1] See e.g., A. R. Liddle and D. H. Lyth, *Cosmological Inflation and Large-Scale Structure* (Cambridge University Press, Cambridge, England, 2000), and references therein.
[2] T. S. Bunch and P. C. Davies, Proc. R. Soc. A **360**, 117 (1978).
[3] A. Vilenkin and L. H. Ford, Phys. Rev. D **26**, 1231 (1982); A. D. Linde, Phys. Lett. **116B**, 335 (1982); A. A. Starobinsky, Phys. Lett. **117B**, 175 (1982).
[4] E. Mottola, Phys. Rev. D **31**, 754 (1985).
[5] B. Allen, Phys. Rev. D **32**, 3136 (1985).
[6] R. H. Brandenberger and J. Martin, Mod. Phys. Lett. A **16**, 999 (2001); Int. J. Mod. Phys. A **17**, 3663 (2002); Phys. Rev. D **71**, 023504 (2005); J. Martin and R. H. Brandenberger, Phys. Rev. D **63**, 123501 (2001); **65**, 103514 (2002); **68**, 063513 (2003); R. H. Brandenberger, S. E. Joras, and J. Martin, Phys. Rev. D **66**, 083514 (2002);

R. Easther, B. R. Greene, W. H. Kinney, and G. Shiu, Phys. Rev. D **64**, 103502 (2001); **66**, 023518 (2002); **67**, 063508 (2003); G. Shiu and I. Wasserman, Phys. Lett. B **536**, 1 (2002); C. P. Burgess, J. M. Cline, F. Lemieux, and R. Holman, J. High Energy Phys. 02 (2003) 48; C. P. Burgess, J. M. Cline, and R. Holman, J. Cosmol. Astropart. Phys. 10 (2003) 4.
[7] U. H. Danielsson, Phys. Rev. D **66**, 023511 (2002); L. Bergstrom and U. H. Danielsson, J. High Energy Phys. 12 (2002) 38.
[8] U. H. Danielsson, J. High Energy Phys. 07 (2002) 40; 12 (2002) 25.
[9] T. Tanaka, astro-ph/0012431.
[10] N. Kaloper, M. Kleban, A. Lawrence, and S. Shenker, Phys. Rev. D **66**, 123510 (2002); L. Hui and W. H. Kinney, Phys. Rev. D **65**, 103507 (2002).
[11] L. Mersini, M. Bastero-Gil, and P. Kanti, Phys. Rev. D **64**,

043508 (2001); J. C. Niemeyer, Phys. Rev. D **63**, 123502 (2001); A. Kempf and J. C. Niemeyer, *ibid.* **64**, 103501 (2001); A. A. Starobinsky, JETP Lett. **73**, 415 (2001); M. Lemoine, M. Lubo, J. Martin, and J.-P. Uzan, Phys. Rev. D **65**, 023510 (2002); M. Lemoine, J. Martin, and J.-P. Uzan, *ibid.* **67**, 103520 (2003).

[12] F. Lizzi, G. Mangano, G. Miele, and M. Peloso, J. High Energy Phys. 06 (2002) 049.

[13] N. Kaloper, M. Kleban, A. Lawrence, S. Shenker, and L. Susskind, J. High Energy Phys. 11 (2002) 37.

[14] A. A. Starobinsky and I. I. Tkachev, JETP Lett. **76**, 235 (2002).

[15] M. Bastero-Gil and L. Mersini, Phys. Rev. D **67**, 103519 (2003); J. C. Niemeyer, R. Parentani, and D. Campo, Phys. Rev. D **66**, 083510 (2002); V. Bozza, M. Giovannini, and G. Veneziano, J. Cosmol. Astropart. Phys. 05 (2003) 1.

[16] M. Porrati, Phys. Lett. B **596**, 306 (2004).

[17] B. R. Greene, K. Schalm, G. Shiu, and J. P. van der Schaar, J. Cosmol. Astropart. Phys. 02 (2005) 1.

[18] U. H. Danielsson, Phys. Rev. D **71**, 023516 (2005).

[19] H. Collins and R. Holman, Phys. Rev. D **71**, 085009 (2005).

[20] L. Sriramkumar and T. Padmanabhan, Phys. Rev. D **71**, 103512 (2005).

[21] A. G. Riess *et al.*, Astron. J. **116**, 1009 (1998); **607**, 665 (2004); S. Perlmutter *et al.*, Astrophys. J. **517**, 565 (1999); J. L. Tonry *et al.*, Astrophys. J. **594**, 1 (2003).

[22] J. F. Donoghue, Phys. Rev. D **50**, 3874 (1994).

[23] S. L. Adler, J. Lieberman, and Y. J. Ng, Ann. Phys. (N.Y.) **106**, 279 (1977).

[24] R. M. Wald, Commun. Math. Phys. **45**, 9 (1975); Phys. Rev. D **17**, 1477 (1978).

[25] K. Schalm, G. Shiu, and J. P. van der Schaar, J. High Energy Phys. 04 (2004) 76; F. Nitti, M. Porrati, and J.-W. Rombouts, hep-th/0503247.

[26] S. Habib, C. Molina-París, and E. Mottola, Phys. Rev. D **61**, 024010 (2000).

[27] P. R. Anderson, W. Eaker, S. Habib, C. Molina-París, and E. Mottola, Phys. Rev. D **62**, 124019 (2000).

[28] B. Allen and A. Folacci, Phys. Rev. D **35**, 3771 (1987).

[29] J. Guven, B. Lieberman, and G. T. Hill, Phys. Rev. D **39**, 438 (1989).

[30] O. J. P. Eboli, S.-Y. Pi, and M. Samiullah, Ann. Phys. (N.Y.) **193**, 102 (1989).

[31] D. Boyanovsky, H. J. de Vega, and R. Holman, Phys. Rev. D **49**, 2769 (1994).

[32] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, England, 1982), and references therein.

[33] D. Bernard and A. Folacci, Phys. Rev. D **34**, 2286 (1986).

[34] R. Laflamme and J. Louko, Phys. Rev. D **43**, 3317 (1991); C. Kiefer, Phys. Rev. D **46**, 1658 (1992).

[35] J. P. Paz and S. Sinha, Phys. Rev. D **44**, 1038 (1991).

[36] A. O. Barvinsky, A. Yu. Kamenshchik, C. Kiefer, and I. V. Mishakov, Nucl. Phys. **B551**, 374 (1999).

[37] J. Schwinger, J. Math. Phys. (N.Y.) **2**, 407 (1961); L. V. Keldysh, Zh. Eksp. Teor. Fiz. **47**, 1515 (1964) [Sov. Phys. JETP **20**, 1018 (1965)]; K.-C. Chou, Z.-B. Su, B.-L. Hao, and L. Yu, Phys. Rep. **118**, 1 (1985); E. Calzetta and B.-L.

Hu, Phys. Rev. D **35**, 495 (1987); **37**, 2878 (1988); **49**, 6636 (1994).

[38] Ya. B. Zeldovich and A. A. Starobinsky, Zh. Eksp. Teor. Fiz. **61**, 2161 (1971) [Sov. Phys. JETP **34**, 1159 (1972)]; L. Parker and S. A. Fulling, Phys. Rev. D **9**, 341 (1974); S. A. Fulling and L. Parker, Ann. Phys. (N.Y.) **87**, 176 (1974); S. A. Fulling, L. Parker, and B.-L. Hu, Phys. Rev. D **10**, 3905 (1974).

[39] S. Habib, Y. Kluger, E. Mottola, and J. P. Paz, Phys. Rev. Lett. **76**, 4660 (1996).

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

[41] T. S. Bunch, J. Phys. A **13**, 1297 (1980).

[42] F. Cooper and E. Mottola, Mod. Phys. Lett. A **2**, 635 (1987).

[43] F. Cooper, S. Habib, Y. Kluger, and E. Mottola, Phys. Rev. D **55**, 6471 (1997).

[44] See e.g., T. Padmanabhan, *Structure Formation in the Universe* (Cambridge University Press, Cambridge, England, 1993).

[45] F. Cooper and E. Mottola, Phys. Rev. D **36**, 3114 (1987).

[46] G. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2752 (1977).

[47] D. Polarski and A. A. Starobinsky, Classical Quantum Gravity **13**, 377 (1996).

[48] N. D. Birrell, Proc. R. Soc. A **361**, 513 (1978); P. R. Anderson and L. Parker, Phys. Rev. D **36**, 2963 (1987).

[49] N. A. Chernikov and E. A. Tagirov, Ann. Inst. Henri Poincaré, A **9**, 109 (1968); E. A. Tagirov, Ann. Phys. (N.Y.) **76**, 561 (1973).

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

[51] L. P. Grishchuk, Zh. Eksp. Teor. Fiz. **67**, 825 (1974) [Sov. Phys. JETP **40**, 409 (1975)].

[52] L. H. Ford, Phys. Rev. D **31**, 710 (1985).

[53] K. Kirsten and J. Garriga, Phys. Rev. D **48**, 567 (1993).

[54] M. B. Einhorn and F. Larsen, Phys. Rev. D **67**, 024001 (2003); **68**, 064002 (2003); T. Banks and L. Mannelli, Phys. Rev. D **67**, 065009 (2003); K. Goldstein and D. A. Lowe, Nucl. Phys. **B669**, 325 (2003); H. Collins, R. Holman, and M. R. Martin, Phys. Rev. D **68**, 124012 (2003); H. Collins and R. Holman, Phys. Rev. D **70**, 084019 (2004).

[55] R. Floreanini, C. T. Hill, and R. Jackiw, Ann. Phys. (N.Y.) **175**, 345 (1987).

[56] A. A. Grib, S. G. Mamayev, and V. M. Mostepanenko, Fortschr. Phys. **28**, 173 (1980).

[57] L. S. Brown and S. J. Carson, Phys. Rev. A **20**, 2486 (1979); Y. Kluger, E. Mottola, and J. M. Eisenberg, Phys. Rev. D **58**, 125015 (1998).

[58] B. L. Hu and H. E. Kandrup, Phys. Rev. D **35**, 1776 (1987); H. E. Kandrup, Phys. Rev. D **37**, 3505 (1988).

[59] B. L. Hu and D. Pavon, Phys. Lett. B **180**, 329 (1986).

[60] E. Mottola, Phys. Rev. D **33**, 2136 (1986); B. L. Hu and S. Sinha, Phys. Rev. D **51**, 1587 (1995); N. G. Phillips and B. L. Hu, Phys. Rev. D **63**, 104001 (2001); B. L. Hu and E. Verdaguer, Living Rev. Relativity **7**, 3 (2004).

[61] C. Kiefer, D. Polarski, and A. A. Starobinsky, Phys. Rev. D **62**, 043518 (2000).