Origin of time asymmetry

S. W. Hawking*
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge, CB3 9EW, United Kingdom

R. Laflamme†
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge, CB3 9EW, United Kingdom and Theoretical Astrophysics, T-6, MSB288, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

G. W. Lyons‡
Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge, CB3 9EW, United Kingdom
(Received 18 December 1992)

It is argued that the observed thermodynamic arrow of time must arise from the boundary conditions of the Universe. We analyze the consequences of the no-boundary proposal, the only reasonably complete set of boundary conditions that has been put forward. We study perturbations of a Friedmann model containing a massive scalar field, but our results should be independent of the details of the matter content. We find that gravitational wave perturbations have an amplitude that remains in the linear regime at all times and is roughly time symmetric about the time of maximum expansion. Thus gravitational wave perturbations do not give rise to an arrow of time. However, density perturbations behave very differently. They are small at one end of the Universe's history, but grow larger and become nonlinear as the Universe gets larger. Contrary to an earlier claim, the density perturbations do not get small again at the other end of the Universe's history. They therefore give rise to a thermodynamic arrow of time that points in a constant direction while the Universe expands and contracts again. The arrow of time does not reverse at the point of maximum expansion. One has to appeal to the weak anthropic principle to explain why we observe the thermodynamic arrow to agree with the cosmological arrow, the direction of time in which the Universe is expanding.

PACS number(s): 98.80.Hw, 98.80.Bp

I. INTRODUCTION

The laws of physics do not distinguish the future from the past direction of time. More precisely, the famous *CPT* theorem [1] says that the laws are invariant under the combination of charge conjugation, space inversion, and time reversal. In fact, effects that are not invariant under the combination *CP* are very weak, so to a good approximation, the laws are invariant under the time reversal operation *T* alone. Despite this, there is a very obvious difference between the future and past directions of time in the Universe we live in. One only has to see a film run backward to be aware of this.

There are several expressions of this difference. One is the so-called psychological arrow, our subjective sense of time, the fact that we remember events in one direction of time but not the other. Another is the electromagnetic arrow, the fact that the Universe is described by retarded solutions of Maxwell's equations, and not advanced ones. Both of these arrows can be shown to be consequences of the thermodynamic arrow, which says that entropy is increasing in one direction of time. It is a nontrivial feature of our Universe that it should have a well-defined thermodynamic arrow which seems to point in the same direction everywhere we can observe. Whether the direction of the thermodynamic arrow is also constant in time is something we shall discuss shortly.

There have been a number of attempts to explain why the Universe should have a thermodynamic arrow of time at all. Why shouldn't the Universe be in a state of maximum entropy at all times? And why should the direction of the thermodynamic arrow agree with that of the cosmological arrow, the direction in which the Universe is expanding? Would the thermodynamic arrow reverse if the Universe reached a maximum radius and began to contract?

Some authors have tried to account for the arrow of time on the basis of dynamic laws. The discovery that *CP* invariance is violated in the decay of the K^0 meson [2] inspired a number of such attempts, but it is now generally recognized that *CP* violation can explain why the Universe contains baryons rather than antibaryons, but it cannot explain the arrow of time. Other authors [3] have questioned whether quantum gravity might not violate *CPT*, but no mechanism has been suggested. One would

*Electronic address: swh1@phx.cam.ac.uk
†Electronic address: laf@gov.lanl.tdo-serv
‡Electronic address: gwl10@phx.cam.ac.uk

not be satisfied with an *ad hoc CPT* violation that was put in by hand.

The lack of a dynamical explanation for the arrow of time suggests that it arises from boundary conditions. The view has been expressed that the boundary conditions for the Universe are not a question for science, but for metaphysics or religion. However, that objection does not apply if there is a sense in which the Universe has no boundary. We shall therefore investigate the origin of the arrow of time in the context of the no-boundary proposal of Hartle and Hawking [4]. This was formulated in terms of Einsteinian gravity which may be only a low-energy effective theory arising from some more fundamental theory such as superstrings. Presumably, it should be possible to express a no-boundary condition in purely string theory terms, but we do not yet know how to do this. However the recent Cosmic Background Explorer (COBE) observations [5] indicate that the perturbations that lead to the arrow of time arise at a time during inflation when the energy density is about 10^{-12} of the Planck density. In this regime, Einstein gravity should be a good approximation.

In most currently accepted models of the early Universe there is some scalar field ϕ whose potential energy causes the Universe to expand in an exponential manner for a time. At the end of this inflationary period, the scalar field starts to oscillate and its energy is supposed to heat the Universe and to be transformed into thermal quanta of other fields. However, this thermalization process involves an implicit assumption of the thermodynamic arrow of time. In order to avoid this, we shall consider a Universe in which the only matter field is a massive scalar field. This will not be a completely realistic model of the Universe we live in because it will be effectively pressure free after the inflationary period, rather than radiation dominated. However, it has the great advantage of being a well-defined model without hidden assumptions about the arrow of time. One would expect that the existence and direction of the arrow of time should not depend on the precise matter content of the Universe. We shall therefore consider a model in which the action is given by the Einstein-Hilbert action

$$I_g=\frac{1}{16\pi G}\int_{\mathcal{M}}d^4x(-g)^{1/2}R+\frac{1}{4\pi G}\int_{\partial\mathcal{M}}d^3x(h)^{1/2}K \tag{1.1}$$

plus the massive scalar field action

$$I_\Phi=-\frac{1}{2}\int_{\mathcal{M}}d^4x(-g)^{1/2}(g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi+m^2\phi^2)\ . \tag{1.2}$$

In accordance with the no-boundary proposal, we shall take the quantum state of the Universe to be defined by a path integral over all compact metrics with this action. This means that the wave function $\Psi[h_{ij},\phi_0]$ for finding a three-metric h_{ij} and scalar field ϕ_0 on a spacelike surface S is given by

$$\Psi(h_{ij},\phi_0)=\int_{\mathcal{C}}d[g_{\mu\nu}]d[\phi]e^{-I_e[g_{\mu\nu},\phi]}\ , \tag{1.3}$$

where the path integral is taken over all metrics and scalar fields on compact manifolds M with boundary S that induce the given values on the boundary. In general, the metrics in the path integral will be complex, rather than purely Lorentzian or purely Euclidean.

There are a number of problems in defining a path integral over all metrics, two of which are (1) the path integral is not perturbatively renormalizable and, (2) the Einstein-Hilbert action is not bounded below.

These difficulties may indicate that Einstein gravity is only an effective theory. Nevertheless, for the reasons given above, we feel the saddle-point approximation to the path integral should give reasonable results. We shall therefore endeavor to evaluate the path integral at stationary points of the action, that is, at solutions of the Einstein equations. These solutions will be complex in general.

The behavior of perturbations of a Friedmann model according to the no-boundary proposal was first investigated by Halliwell and Hawkins [6], and we shall adopt their notation. The perturbations are expanded in hyperspherical harmonics. There are three kinds of harmonics.

(1) Two degrees of freedom in tensor harmonics. These are gauge invariant and correspond to gravitational waves.

(2) Two degrees of freedom in vector harmonics. In the model in question they are pure gauge.

(3) Three degrees of freedom in scalar harmonics. Two of them correspond to gauge degrees of freedom and one to a physical density perturbation.

One can estimate the wave functions for the perturbation modes by considering complex metrics and scalar fields that are solutions of the Einstein equations whose only boundary is the surface S. When S is a small three-sphere, the complex metric can be close to that of part of a Euclidean four-sphere. In this case, the wave functions for the tensor and scalar modes correspond to them being in their ground state. As the three-sphere S becomes larger, these complex metrics change continuously to become almost Lorentzian. They represent universes with an initial period of inflation driven by the potential energy of the scalar field. During the inflationary phase, the perturbation modes remain in their ground states until their wavelengths become longer than the horizon size. The wave function of the perturbations then remains frozen until the horizon size increases to be more than the wavelength again during the matter-dominated era of expansion that follows the inflation. After the wavelengths of the perturbations come back within the horizon, they can be treated classically.

This behavior of the perturbations can explain the existence and direction of the thermodynamic arrow of time. The density perturbations, when they come within the horizon, are not in a general state, but in a very special state with a small amplitude that is determined by the parameters of the inflationary model; in this case, the mass of the scalar field. The recent observations by COBE indicate this amplitude is about 10^{-5}. After the density perturbations come within the horizon, they will grow until they cause some regions to collapse as protogalaxies and clusters. The dynamics will become highly nonlinear and chaotic, and the coarse-grained entropy will increase. There will be a well-defined thermodynamic arrow of time that points in the same direction everywhere in the Universe and agrees with the direction of time in which the Universe is expanding, at least during this phase.

The question then arises: If and when the Universe reaches maximum size, will the thermodynamic arrow reverse? Will entropy decrease and the Universe become smoother and more homogeneous during the contracting phase? In Ref. [7] it was claimed that the no-boundary proposal implied that the thermodynamic arrow would reverse during the contraction. This is now recognized to be incorrect, but it is instructive to consider the arguments that led to the mistake and see why they do not apply. The anatomy of error is not ruled by logic, but there were three arguments which together seemed to point to reversal.

(1) The no-boundary proposal implied that the wave function of the Universe was invariant under *CPT*.

(2) The analogy between spacetime and the surface of the Earth suggested that if the North Pole were regarded as the beginning of the Universe, the South Pole should be its end. One would expect conditions to be similar near the North and South Poles. Thus, if the amplitude of perturbations was small at early times in the expansion, it should also be small at late times in the contraction. The Universe would have to get smoother and more homogeneous as it contracted.

(3) In studies of the Wheeler-DeWitt equation on minisuperspace models [8], it was thought that the no-boundary condition implied that $\Psi(a)\rightarrow 1$ as the radius $a\rightarrow 0$. In the case of a Friedmann model with a massive scalar field, this seemed to imply that the classical solutions that corresponded to the wave function through the WKB approximation would bounce and be quasiperiodic. This could be true only if the solutions were restricted to those in which the perturbations became small again as the Universe contracted.

Page [9] pointed out that the first argument about the *CPT* invariance of the wave function did not imply that the individual histories had to be *CPT* symmetric, just that if the quantum state contained a particular history, then it must also contain the *CPT* image of that history with the same probability. Thus, this argument did not necessarily imply that the thermodynamic arrow reversed in the contracting phase. It would be equally consistent with *CPT* invariance for there to be histories in which the thermodynamic arrow was pointed forward during both the expansion and contraction, and for there to be other histories with equal probability in which the arrow was backward. With a relabeling of time and space directions and of particles and antiparticles, these two classes of histories would be physically identical. Both would correspond to a steady increase in entropy from one end of time, which can be labeled the big bang, to the other end, which can be labeled the big crunch.

The second argument, about the North and South Poles being similar, is really a confusion between real and imaginary time. It is true that there is no distinction between the positive and negative directions of time. In the Euclidean regime, the imaginary time direction is on the same footing as spatial dimensions. So one can reverse the direction of imaginary time by a rotation. Indeed, this is the basis of the proof that the no-boundary quantum state is *CPT* invariant. But as noted above, this does not imply that the individual histories are symmetric in real time or that the big crunch need be similar to the big bang.

The third argument, that the boundary condition for the Wheeler-DeWitt equation should be $\Psi\rightarrow 1$ for small three-spheres S in a homogeneous isotropic minisuperspace model, was the one that really led to the error of suggesting that the arrow of time reversed. The motivation behind the adoption of this boundary condition was the idea that the dominant saddle point in the path integral for a very small three-sphere would be a small part of a Euclidean four-sphere. The action for this would be small. Thus, the wave function would be about one, irrespective of the value of the scalar field. With this boundary condition, the minisuperspace Wheeler-DeWitt equation gave a wave function that was constant or exponential for small radii, and which oscillated rapidly for larger radii. From the WKB approximation, one could interpret the oscillations as corresponding to Lorentzian geometries. That fact that the oscillating region did not extend to very small radii was taken to indicate that these Lorentzian geometries would not collapse to zero radii but would bounce. Thus, they would correspond to quasiperiodic oscillating universes. In such universes, the perturbations would have to obey a quasiperiodic boundary condition, and be small whenever the radius of the universe was small. Otherwise, the universe would not bounce. This would mean that the thermodynamic arrow would have to reverse during the contraction phase, so that the perturbations were small again at the next bounce.

This boundary condition on the wave function became suspect when Laflamme [10,11] found other minisuperspace models in which a bounce was not possible. Then Page [9] pointed out that for small three-surfaces S, there was another saddle point that could make a significant contribution to the wave function. This was a complex metric that started almost like half of a Euclidean four-sphere and was followed by an almost Lorentzian metric that expanded to a maximum radius, and then collapsed to the small three-surface S. The long Lorentzian period would give the action of these metrics a large imaginary part. This would lead to a contribution to the wave function that oscillated very rapidly as a function of the radius of the three-surface S and the value of the scalar field on it. Thus, the boundary condition of the Wheeler-DeWitt equation would not be exactly $\Psi\rightarrow 1$ as the radius tends to zero. There would also be a rapidly oscillating component of the wave function.

As before, the wave functions for perturbations about the Euclidean saddle-point metric would be in their ground states, but there is no reason for this to be true for perturbations about the saddle-point metric with a long Lorentzian period that expanded to a large radius and then contracted again.

To find out what the wave functions for perturbations in the contracting phase are, one has to solve the relevant Schrödinger equation during the expansion and contraction. This we do in Secs. III A and III B. We find that the tensor modes have wave functions that correspond to gravitational waves that oscillate with an adiabatically varying amplitude. This amplitude will depend on the radius of the Universe. It will be the same at the same radius in the expanding and contracting phases, and it will be small compared to one whenever the wavelength is less than the horizon size. Thus, these gravitational wave modes will not become nonlinear and will not give rise to a thermodynamic arrow of time.

In contrast, scalar modes between the Compton wavelength of the scalar field and the horizon size will not oscillate, but will have power-law behavior. There are two independent solutions of the perturbation equations, one which grows and one which decreases with time. The boundary condition provided by the no-boundary proposal picks out the solution that is a small perturbation about the Euclidean saddle point for small three-spheres. It does not require that the perturbation about the saddle point with a long Lorentzian period remain small. So the no-boundary proposal picks out the solution of the density perturbation equation that starts small, but grows during the expansion and continues to grow during the contraction. At some point during the expansion, the amplitude will grow so large that the linearized treatment will break down. This, however, does not prevent one from using linear perturbation theory to draw conclusions about the thermodynamic arrow of time. The arrow of time is determined by when the evolution becomes nonlinear. The linear treatment and the no-boundary proposal enable one to say that this will happen during the expansion. After that, the evolution will become chaotic and the coarse-grained entropy will increase. It will continue to increase in the contracting phase because there is no requirement that the perturbations become small again as the Universe shrinks. Thus, the thermodynamic arrow will not reverse. It will point the same way while the Universe expands and contracts.

The thermodynamic arrow will agree with the cosmological arrow for half the history of the Universe, but not for the other half. So why is it that we observe them to agree? Why is it that entropy increases in the direction that the Universe is expanding? This is really a situation in which one can legitimately invoke the weak anthropic principle, because it is a question of where in the history of the Universe conditions are suitable for intelligent life. The inflation in the early Universe implies that the Universe will expand for a very long time before it contracts again. In fact, it is so long that the stars will have all burnt out and the baryons will have all decayed. All that will be left in the contracting phase will be a mixture of electrons, positrons, neutrinos, and gravitons. This is not a suitable basis for intelligent life.

The conclusion of this paper is that the no-boundary proposal can explain the existence of a well-defined thermodynamic arrow of time. This arrow always points in the same direction. The reason we observe it to point in the same direction as the cosmological arrow is that conditions are suitable for intelligent life only at the low-entropy end of the Universe's history.

II. THE HOMOGENEOUS MODEL

In this section we review the homogeneous model with the metric

$$ds^2=\sigma^2[-N(t)^2dt^2+a(t)^2d\Omega_3^2] , \qquad (2.1)$$

where $\sigma^2=2/(3\pi m_P^2)$, N is the lapse function, a is the scale factor, and $d\Omega_3^2$ is the standard three-sphere metric. Expressing the scalar field as $\sqrt{2}\pi\sigma\phi$ with the quadratic potential $2\pi^2\sigma^2m^2\phi^2$, the Lorentzian action is

$$I=-\frac{1}{2}\int dt\, Na^3\left[\frac{\dot a^2}{N^2a^2}-\frac{1}{a^2}-\frac{\dot\phi^2}{N^2}+m^2\phi^2\right] , \qquad (2.2)$$

where the dot denotes derivative with respect to Lorentzian Friedmann-Robertson-Walker (FRW) time (if not explicitly stated, throughout the paper time derivatives are Lorentzian). There are no time derivatives of the lapse function N in this action; it is a Lagrange multiplier. Varying the action with respect to N leads to the constraint

$$H=\frac{N}{2a^3}[-a^2\pi_a^2+\pi_\phi^2-a^4(1-a^2m^2\phi^2)]=0 , \qquad (2.3)$$

where the momenta π_a and π_ϕ are defined as

$$\pi_a=-\frac{a}{N}\dot a \quad\text{and}\quad \pi_\phi=\frac{a^3}{N}\dot\phi , \qquad (2.4)$$

and H is the Hamiltonian. This constraint is a consequence of the invariance under time reparametrization. Varying the action with respect to the field ϕ we obtain the reduced Klein-Gordon equation

$$N\frac{d}{dt}\left[\frac{\dot\phi}{N}\right]+3\frac{\dot a}{1}\dot\phi+N^2m^2\phi^2=0 . \qquad (2.5)$$

This latter equation together with the Hamiltonian constraint $H=0$ is sufficient to describe the classical dynamics. The second-order equation for a can be derived from these equations. In the inhomogeneous model there are also momentum constraints, but these are trivially satisfied in the homogeneous background.

The quantum theory is obtained by replacing the different variables by operators. We will follow the Dirac method and impose the classical constraints as quantum operators. The Hamiltonian constraint thus becomes

$$\left[a^2\frac{\partial^2}{\partial a^2}-\frac{\partial^2}{\partial\phi^2}-a^4(1-a^2m^2\phi^2)\right]\Psi_0(a,\phi)=0 , \qquad (2.6)$$

and is called the Wheeler-DeWitt equation. The solution of this equation, $\Psi_0(a,\phi)$ is the wave function of the Universe. There is a factor ordering ambiguity, but it is not important for the conclusions of our paper, which rely on the classical limit.

In this paper, we investigate the predictions of the no-boundary proposal in a model where small inhomogeneities are taken into account. In order to impose this proposal, we return to a path-integral formulation of the wave function. It is very hard to calculate this path integral exactly. However, we can have a good idea of the resulting wave function by using a saddle-point approximation.

$$\Psi(h_{ij},\phi)\approx Ce^{-I_E^{SP}[g_{\mu\nu}\Phi]} , \quad (2.7)$$

where C is a prefactor and I_E^{SP} is the Euclidean saddle-point action. In this approximation it is clear how to impose the proposal of Hartle and Hawking. The regularity condition is imposed on the (complex) saddle points of the path integral. The semiclassical approximation to the path integral can then be used to estimate the wave function.

One of the problems in using the semiclassical approximation in this model is that we cannot simply deform the complex metric into purely real Euclidean and real Lorentzian sections, for real arguments of the wave function. This could only be achieved in this model if the time derivatives of both a and ϕ vanish simultaneously on the Euclidean axis [12], which is not possible, as ϕ increases monotonically if the no-boundary condition is imposed. Therefore, we must solve the background equations of motion for complex values of time and physical variables, obtaining complex solutions which satisfy the no-boundary proposal and have the given a and ϕ on the final hypersurface. The no-boundary proposal imposes the boundary conditions at one end of the four-geometry:

$$a=0, \ \frac{da}{d\tau}=1, \ \frac{d\phi}{d\tau}=0, \ \phi=\phi_0 ; \quad (2.8)$$

thus, we only have the freedom to choose the (complex) value of ϕ at the origin of complex time τ.

Lyons [13] found that there were many contours in the complex time plane which induced real end points a and ϕ. Some possibilities are obtained by choosing the initial value of ϕ to have an imaginary part much smaller than the real part such that $\phi_0^{Im}\approx-(1+2n)\pi/6\phi_0^{Re}$ (for integer n). In this paper we will only investigate the case $n=0$.

For small a, the complex metric can effectively be considered as a small real Euclidean section, with ϕ_0 approximately real, described by

$$\phi\approx\phi_0 \ \text{ and } a\approx\frac{1}{m\phi_0}\sin m\phi_0\tau , \quad (2.9)$$

where τ is the Euclidean time. When we consider gravitons below, it is a good approximation to assume the following behavior for the radius a when $\phi_0>1$. For small a ($<m\phi_0$), the background is part of a Euclidean four-sphere:

$$a\approx\frac{1}{m\phi_0\cosh\eta_E} , \quad -\infty<\eta_E<0 . \quad (2.10)$$

The Euclidean conformal time is given by $\eta_E=\int d\tau/a$. Although η_E has a semi-infinite range, notice that the proper distance is finite. The radius a starts at zero and increases to a maximum value of $a/m\phi_0$, the equator of the four-sphere. For larger a, the saddle point is well approximated by de Sitter space:

$$a\approx\frac{1}{m\phi_0\cos\eta}, \quad 0<\eta<\frac{\pi}{2}-\delta_e , \quad (2.11)$$

where η is the analytic continuation of $\eta_E=i\eta$. The Universe is then in an inflationary era. In terms of comoving time,

$$\phi\approx\phi_0-\frac{mt}{3} \ \text{ and } a\approx\frac{1}{m\phi_0}e^{m\phi_0 t-m^2t^2/6} , \quad (2.12)$$

where t is the analytic continuation of τ in the Lorentzian region.

The action is given by

$$I_e\approx-\frac{1}{3m^2\phi_0^2}[1-(1-m^2\phi_0^2a^2)^{3/2}] . \quad (2.13)$$

For large a ($\gg 1/m\phi_0$), the saddle point will have a large imaginary part. The wave function will therefore be of WKB type. After a suitable coarse graining [14], we can associate the phase of the wave function to the Hamilton-Jacobi function of general relativity. When this is possible, we will assume that the Universe behaves essentially classically. The wave function will be associated with the family of classical Lorentzian trajectories described by the Hamilton-Jacobi function.

Meanwhile, the scalar field is decreasing and inflation will end at $\eta=\pi/2-\delta_e$, when the scalar field reaches a value around unity, at which point the value of a will be

$$a_e\approx(1/m\phi_0)\exp(3\phi_0^2/2) .$$

δ_e is given by $\delta_e\approx\exp(-3(\phi_0)^2/2)$. For $\phi_0>1$, we have $\delta_e\ll1$. When $\eta>\pi/2-\delta_e$, the scalar field oscillates and behaves essentially as a pressureless fluid (i.e., dust):

$$\phi\approx\frac{1}{m}\left[\frac{a_{max}}{a^3}\right]^{1/2}\cos(mt) . \quad (2.14)$$

The scale factor of the Universe is then well described by

$$a\approx a_m\sin^2\left[\frac{\pi/2-3\delta_e-\eta}{2}\right], \quad \frac{\pi}{2}-\delta_e<\eta , \quad (2.15)$$

where the constants have been chosen to ensure a smooth transition between the inflationary and dust era. The Universe will therefore expand to a maximum radius

$$a_m\approx m^2a_e^3\approx\frac{\exp(9(\phi_0)^2/2)}{m(\phi_0)^3}$$

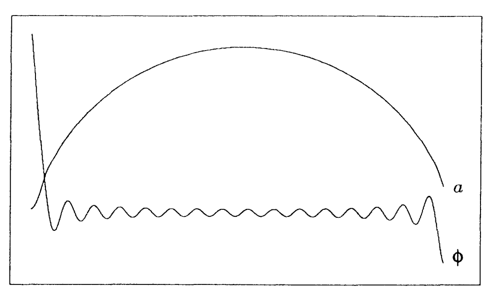

FIG. 1. Classical trajectories for the scale factor a and homogeneous scalar field ϕ corresponding to the no-boundary proposal.

and recollapse. It will be convenient later on to redefine the origin of conformal time during the dustlike era by setting $\eta_d=\eta-\pi/2+3\delta_e$. The scale factor will then evolve as

$$a\approx a_{\max}\sin^2\frac{\eta_d}{2},\quad 0<\eta_d<2\pi\ . \tag{2.16}$$

Figure 1 depicts a typical classical trajectory corresponding to the no-boundary proposal.

III. INHOMOGENEOUS PERTURBATIONS

Let us now consider the behavior of small perturbations around the homogeneous model described in the previous section. We write the metric as

$$g_{\mu\nu}(t,\mathbf{x})=g_{\mu\nu}(t)+\delta g_{\mu\nu}(t,\mathbf{x})\ . \tag{3.1}$$

The background part $g_{\mu\nu}(t)$ was described in the previous section by the line element (2.1).

One can decompose a general perturbation $\delta g_{\mu\nu}$ of a Robertson-Walker background metric into scalar (Q^n_{lm}), vector $((P_i)^n_{lm},(S^{o,e}_i)^n_{lm})$, and tensor $((P_{ij})^n_{lm},(S^{o,e}_{ij})^n_{lm},(G^{o,e}_{ij})^n_{lm})$ harmonics. This classification originates from the way they transform under rotations of the three-sphere. These harmonics are constructed from the scalar, vector, and tensor eigenfunctions of the Laplacian on the three-sphere, viz. Q^n_{lm}, $(S^{o,e}_i)^n_{lm}$, and $(G^{o,e}_{ij})^n_{lm}$. More details and properties of these harmonics are given in Refs. [15,16].

We can expand the inhomogeneous perturbations of the metric in terms of these harmonics (where the index n should be thought of as shorthand for nlm and o,e). The tensor perturbations are

$$\delta g^{(t)}_{\mu\nu}=\sum_n a^2\begin{bmatrix}0 & 0\\ 0 & 2d_nG^n_{ij}\end{bmatrix}\ . \tag{3.2}$$

G^n_{ij} are the transverse traceless tensor harmonics. The vector perturbations are

$$\delta g^{(v)}_{\mu\nu}=\sum_n\frac{a^2}{\sqrt{2}}\begin{bmatrix}0 & j_nS^n_i\\ j_nS^n_i & 2c_nS^n_{ij}\end{bmatrix}\ , \tag{3.3}$$

where the $S^n_{ij}=S^n_{i|j}+S^n_{j|i}$ are obtained from the transverse vector harmonics S^n_i. The scalar perturbations of the

$$\delta g^{(s)}_{\mu\nu}=\sum_n\frac{a^2}{\sqrt{6}}\begin{bmatrix}-2N_0^2g_nQ^n & k_nP^n_i\\ k_nP^n_i & 2a_n\Omega_{ij}Q^n+6b_nP^n_{ij}\end{bmatrix}\ , \tag{3.4}$$

where the $P^n_i=Q^n_{|i}/(n^2-1)$ and

$$P^n_{ij}=\frac{\Omega_{ij}Q^n}{3}+\frac{Q^n_{|ij}}{(n^2-1)}\ ,$$

are obtained from the scalar harmonics Q^n. We must also take into account the scalar perturbations of the scalar field:

$$\delta\phi=\sum_n\frac{1}{\sqrt{6}}f_nQ^n\ . \tag{3.5}$$

This expansion is in effect a Fourier transform adapted to the symmetry of the FRW background. The coefficients a_n, b_n, c_n, d_n, f_n, g_n, j_n, and k_n are functions of time, but not of the spatial coordinates of the three-sphere hypersurfaces. Spatial information is encoded in the harmonics.

In [6] the actions (1.3) and (1.4) were expanded to second order around the homogeneous model. In Appendix A, we have reproduced it with the equations of motion for the various Fourier coefficients. After examining the perturbing Lagrangians (A2) and (A3), we find that the different types of harmonics decouple from each other. Their wave functions will therefore separate, so we can write

$$\Psi_n(a,\phi,a_n,b_n,c_n,d_n,f_n)=\psi^s_n(a,\phi,a_n,b_n,f_n)\psi^v_n(a,\phi,c_n)\psi^t_n(a,\phi,d_n)\ . \tag{3.6}$$

It is thus possible to investigate them separately. We will study the tensor and scalar modes in the next two subsections. For the vector modes, there are only two variables, c_n and j_n. The latter one, however, is a Lagrange multiplier, and thus induces a constraint for the only variable left. Thus, we find that the vector degrees of freedom are pure gauge and will only contribute to the phase of the total wave function.

A. Linear gravitons

Linear gravitons are the transverse and traceless parts of the three-metric, and are described by the variable d_n in the above notation. Using the background equation of motion we can derive the equation [17]

$$d_n''+2\mathcal{H}d_n'+(n^2-1)d_n=0 \tag{3.7}$$

for the modes d_n. Here the derivatives are evaluated with respect to Lorentzian conformal time and $\mathcal{H}=a'/a$. The gravitons are decoupled from the scalar- and vector-derived tensor harmonics and depend only on the behavior of the background.

We will calculate the wave function for the graviton modes using a saddle-point approximation, assuming the background wave function (2.7) and saddle-point action (2.13). The tensor part of the wave function [see (3.6)] can be written as

$$\psi^t_n(a,\phi_0,d_n)=\int[dd_n]e^{-(I_E)}\approx Ce^{-(I^{\rm ext}_E)}\ , \tag{3.8}$$

where

$$C=[\delta^2(I^{\rm ext}_E)/\delta d^i_n\delta d^f_n]^{1/2}$$

is the prefactor assuming the flat spacetime measure.

The Euclidean action for a mode d_n calculated along an extremizing path is given by the boundary term

$$I^{\rm ext}_E=\left.\left[\frac{a^2d_nd_n'}{2}+2aa'd_n^2\right]\right|^{\eta^f_E}_{\eta^i_E}\ , \tag{3.9}$$

where η_E is the Euclidean time, a function of the background variables a and φ_0 as described in [18]. It is possible to rewrite this action in terms of values of the field on the boundary d_n^i, d_n^f and solutions of the classical equation p_n,

$$\frac{d}{d\eta_E} a^2 \frac{d}{d\eta_E} p_n - (n^2-1)a^2 p_n = 0 \ , \tag{3.10}$$

evaluated on the boundary. The regularity condition for the no-boundary proposal implies that d_n must vanish when the three-geometry shrinks to zero, and this implies that the action will have the form

$$I_E^{\rm ext} = A d_n^2 = \frac{a^2}{2}\left[\frac{p_n'}{p_n} + 4\frac{a'}{a}\right] d_n^2 \ . \tag{3.11}$$

In regions of configuration space where the Universe is Lorentzian, the appropriate analytic continuation of (3.11) should be taken.

It is possible to find a good analytical approximation for p_n, and thus of the wave function, using (3.8) and (3.11) and assuming that the background is described by Eqs. (2.10), (2.11), and (2.15). The p_n are approximately

$$p_n \propto \begin{cases} \left[\cosh\eta_E - \dfrac{\sinh\eta_E}{\eta}\right] e^{n\eta_E}, & -\infty < \eta_E < 0 \ \text{in the Euclidean region} \ , \\ \left[\cos\eta + i\dfrac{\sin\eta}{n}\right] e^{-in\eta}, & 0 < \eta < \dfrac{\pi}{2} - \delta_e \ \text{in the inflationary era} \ , \\ \left[\dfrac{\cos[n(\eta - 3\pi/2 + 3\delta_e)]}{\cos^2[(\eta - 3\pi/2 + 3\delta_e)/2]} - \dfrac{\sin[(\eta - 3\pi/2 + 3\delta_e)/2]\sin[n(\eta - 3\pi/2 + 3\delta_e)]}{2e\cos^3[(\eta - 3\pi/2 + 3\delta_e)/2]}\right], & \\ \qquad \dfrac{\pi}{2} - \delta_e < \eta \ \text{in the dustlike phase} \ . \end{cases} \tag{3.12}$$

Modes with $n\delta_e \ll 1$ are those with wavelengths much larger than the Hubble radius at the end of inflation. At the onset of inflation, they are in their ground state, and thus oscillate adiabatically. These modes will no longer oscillate adiabatically when they leave the Hubble radius during inflation. However, all modes will reenter the Hubble radius during the dust era when

$$n \approx \tan\left[\frac{(\eta - \pi/2 + 3\delta_e)}{2}\right] ,$$

and start oscillating adiabatically again. Modes with $n\delta_e \gg 1$ oscillate adiabatically throughout the evolution. All the modes oscillate around the time of maximum expansion, and even if some do not have a phase which is exactly time symmetric, their amplitudes are.

The variance squared of the field and its momenta for modes with $n\delta_e \ll 1$ around the time of maximum expansion are given by

$$\langle d_n^2 \rangle = \frac{1}{2(A^* + A)} \approx \frac{1 + 2\gamma\cos(2n\eta) + \gamma^2}{2na^2(1-\gamma^2)} \ , \tag{3.13}$$

$$\langle \pi_{d_n}^2 \rangle = \frac{A^* A}{2(A^* + A)} \approx \frac{na^2}{2}\frac{(1-\gamma^2)^2 + 4\gamma^2\sin^2(2n\eta)}{[1 + 2\gamma\cos(2n\eta) + \gamma^2](1-\gamma^2)} \ , \tag{3.14}$$

and

$$\langle d_n \pi_{d_n} + \pi_{d_n} d_n \rangle = \frac{i(A - A^*)}{A + A^*} \approx \frac{4\gamma\sin(2n\eta)}{1-\gamma^2} \ , \tag{3.15}$$

where $\gamma = 1 - n^2\delta_e^2/2$. The expectation value of the Hamiltonian

$$H_n = \frac{1}{2a^3}\{\pi_{d_n}^2 + 4(d_n\pi_{d_n} + \pi_{d_n} d_n)a\pi_a + d_n^2[10a^2\pi_a^2 + 6\pi_\phi^2 - 6a^6 m^2\phi^2 + (n^2+1)a^4\} \tag{3.16}$$

is

$$\langle H_n \rangle \approx \begin{cases} \dfrac{n}{a} & \text{at the onset of inflation} \ , \\ \dfrac{n}{an^2\delta_e^2} & \text{near the maximum expansion} \ . \end{cases} \tag{3.17}$$

This shows that modes start in their ground state before the onset of inflation, and get excited during inflation and the dust phase.

A useful way to gain information about this state is to investigate the Wigner function

$$\mathcal{F}(\bar{d}_n, \bar{\pi}_n) = \frac{1}{2\pi}\int d\Delta\, e^{-2i\bar{\pi}\Delta}\psi^*(\bar{d}_n - \Delta)\psi(\bar{d}_n + \Delta) \ . \tag{3.18}$$

The Wigner function gives an idea of the phase-space probability distribution of possible classical perturbations (once decoherence has occurred). For the wave function (3.8) with action (3.11), it is given by

$$\mathcal{F}(\bar{d}_n,\bar{\pi}_n)=\frac{A+A^*}{2\pi}\exp\left[-\left(\frac{4AA^*}{A+A^*}\bar{d}_n^2+\frac{1}{A+A^*}\bar{\pi}_n^2-2i\frac{A-A^*}{A+A^*}\bar{d}_n\bar{\pi}_n\right)\right] . \tag{3.19}$$

At the onset of inflation, the Wigner function is a round Gaussian (factoring out the mode number and the radius of the Universe). A mode with $n < \tan(\pi/2-\delta_e)$ will go outside the Hubble radius and have frozen amplitude, and the Wigner function will then become an ellipse elongated in the momentum direction. When the mode comes back within the Hubble radius, it starts rotating with period $2\pi/n$ in phase space. This behavior lasts until $n \approx \tan\eta$ in the recontracting phase. The parameter characterizing the eccentricity of this ellipse is called the squeezing and has been studied by Grishchuk and Sidorov [19].

Typical classical perturbations $d_n^{\rm cl}$ resulting from the above Wigner function are small at the onset of inflation. Their amplitudes get frozen when they leave the Hubble radius. During this stage their energies increase. The perturbations will start oscillating again with amplitude proportional to a^{-1} when they come back within the Hubble radius in the dust phase. They behave like

$$d_n^{\rm cl}\approx\frac{\sin(n\eta+\epsilon)}{an^{3/2}\delta_e} , \tag{3.20}$$

where ϵ is an unimportant phase depending on the details of the matching of the p_n functions in (3.12). Around the time of maximum expansion, the amplitude of the graviton modes is symmetric, and thus their arrow of time agrees with the cosmological one. Figure 2 depicts a typical classical evolution of a linear graviton.

B. Linear scalar perturbations

1. Quantum mechanics of the physical degree of freedom

We have seen that gravitons are adiabatic near the time of maximum expansion, so that their amplitude is time symmetric with respect to that point. This is not special to gravitons, as the electromagnetic field and massless or conformally coupled scalar fields will also be adiabatic. In this section we will show, however, that perturbations of a massive scalar field will not behave adiabatically at the time of maximum expansion.

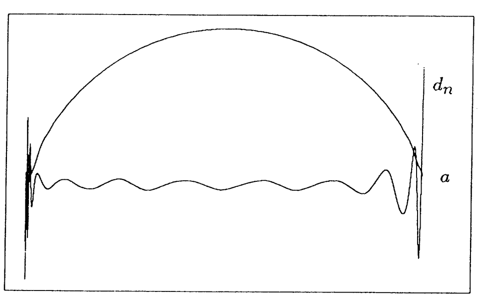

FIG. 2. Classical trajectories for the radius of the Universe a in the FRW model with scalar field. The amplitude of a particular mode of the gravitons d_n has time symmetric amplitude with respect to the maximum up to very small radii.

From the expansions (3.4) and (3.5), we see that there are five scalar degrees of freedom described by the time-dependent coefficients a_n, b_n, f_n, k_n, and g_n. However, the latter two appear as Lagrange multipliers in the Lagrangians (A2) and (A3) and induce two constraints, so overall there is only one true scalar degree of freedom. Without the presence of the scalar field, the scalar degrees of freedom would also be pure gauge. Care should be taken in the treatment of the scalar perturbations in order to avoid gauge dependent results. Let us first find the real degree of freedom.

Variations of the action with respect to the Lagrange multipliers N, g_n, and k_n result in the Hamiltonian, linear Hamiltonian, and momentum constraints. In Dirac quantization, which we follow here, these constraints are imposed as constraints on the quantum state. The wave function therefore depends only on a linear combination of the coefficients a_n, b_n, and f_n. The momentum constraints ensure that the wave function is invariant under diffeomorphisms of the spatial three-surfaces. The Hamiltonian and linear Hamiltonian constraints ensure time reparametrization invariance of the wave function.

Shirai and Wada [20] give an explicit form for the wave function, which automatically satisfies the momentum constraints. These are solved by making the judicious change of variables

$$\begin{aligned}\tilde{\alpha}&=\alpha+\frac{1}{2}\sum_n a_n^2-2\sum_n\frac{n^2-4}{n^2-1}b_n^2 ,\\ \tilde{\phi}&=\phi-3\sum_n b_nf_n ,\end{aligned} \tag{3.21}$$

where $\alpha=\ln a$. Once this transformation has been performed, the momentum constraints imply that the wave function is independent of the linear combination a_n-b_n. In terms of the two degrees of freedom left, the linear Hamiltonian constraint becomes

$$\pi_\phi\pi_{f_n}-\pi_\alpha\pi_{s_n}+e^{6\alpha}m^2\phi f_n+K_ns_n=0 , \tag{3.22}$$

where $s_n=a_n+b_n$ and

$$K_n=\tfrac{1}{3}[(n^2-4)\pi_\alpha^2-(n^2+5)\pi_\phi^2-(n^2-4)e^{6\alpha}m^2\phi^2] .$$

The remaining gauge degree of freedom can be eliminated by solving the linear Hamiltonian constraint using the canonical transformation

$$\begin{aligned}\begin{bmatrix}y_n\\ z_n\end{bmatrix}&=\begin{bmatrix}K_n & e^{6\alpha}m^2\phi\\ \pi_\phi & \pi_\alpha\end{bmatrix}\begin{bmatrix}s_n\\ f_n\end{bmatrix} ,\\ \begin{bmatrix}\pi_{s_n}\\ \pi_{f_n}\end{bmatrix}&=\begin{bmatrix}K_n & \pi_\phi\\ e^{6\alpha}m^2\phi & \pi_\alpha\end{bmatrix}\begin{bmatrix}\pi_{y_n}-\dfrac{y_n}{\Sigma}\\ \pi_{z_n}\end{bmatrix} ,\end{aligned} \tag{3.23}$$

where $\Sigma = -K_n \pi_\alpha + e^{6\alpha} m^2 \phi \pi_\phi$. The linear Hamiltonian constraint then implies that, imposed as a quantum constraint,

$$\pi_{y_n} \Psi(y_n, z_n) = 0 , \tag{3.24}$$

and so Ψ is independent of y_n. Therefore, the true degree of freedom has been isolated—the wave function is found to depend only on the single physical variable

$$z_n = \pi_\phi s_n + \pi_\alpha f_n = a^2(\phi' s_n - \mathcal{H} f_n) , \tag{3.25}$$

and on the background variables $\tilde{a}$ and $\tilde{\phi}$ (in the rest of the paper we will drop the tilde on a and ϕ). The expression for the Hamiltonian for the modes z_n is rather complicated, and is shown only in Appendix A.

We can find the wave function for the scalar perturbations in terms of the real degree of freedom by using the semiclassical approximation to the path-integral expression for the wave function as in the graviton case

$$\psi^s(a, \phi, z_n) \sim C(a, \phi) \exp(-I_E^{\rm cl}) . \tag{3.26}$$

The Euclidean action of the saddle-point contribution to the path integral is a boundary term (since the action is quadratic) given by

$$I_n^{\rm cl} = (M z_n z_n' - N z_n^2)\Big|_{\eta_E^i}^{\eta_E^f} \tag{3.27}$$

for

$$M = \frac{(n^2-4)}{2[(n^2-4)a'^2 + 3a^2\phi'^2]} ,$$

$$N = \frac{1}{4MUa^3}\left[K_n \left[2a^4 - 3a^6 m^2 \phi^2 + 3\frac{n^2-1}{n^2-4} a^4 \phi'^2 \right] + a^{12} m^4 \phi^2 + 3a^9 \phi\phi' a' \right] ,$$

$$U = K_n a a' + a^8 m^2 \phi\phi' ,$$

and the derivatives here are with respect to Euclidean conformal time.

It is difficult to find solutions of the equations for z_n. It is easier to return to the original variables and pick a particular gauge. In order to study the scalar perturbations, we shall choose the gauge $b_n = k_n = 0$, which is known as the longitudinal gauge. Once the result has been obtained in this gauge, it will be easy to recast it in terms of the true degree of freedom z_n, and therefore in a gauge-invariant way. Alternatively, we could use the gauge-invariant variables of Bardeen [21]. Their relationship with the formalism used here is described in Appendix B.

In the $b_n = k_n = 0$ gauge, we have the equations of motion (in Lorentzian time)

$$a_n'' + 3\mathcal{H} a_n' + (3m^2\phi^2 a^2 - 2) a_n = 3(m^2 \phi a^2 f_n - \phi' f_n') , \tag{3.28}$$

$$f_n'' + 2\mathcal{H} f_n' (n^2 - 1 + m^2 a^2) f_n = 2m^2 \phi a^2 a_n - 4\phi' a_n' , \tag{3.29}$$

and the constraints

$$a_n' + \mathcal{H} a_n = -3\phi' f_n , \tag{3.30}$$

$$a_n(n^2 - 4 - 3\phi'^2) = 3\phi' f_n' + 3m^2 \phi a^2 f_n + 9\mathcal{H}\phi' f_n . \tag{3.31}$$

This gauge corresponds to having a zero-shift function and a diagonal metric on the three-sphere. It fixes the metric up to a conformal transformation of the three-sphere corresponding to the $n=2$ mode. This eliminates the possibility of gauge artifacts for scalar modes with higher n. Equations (3.28)–(3.30) are just (A7), (A11), and (A8), noting that $g_n = -a_n$ in this gauge. The last equation follows from (A14), (3.30), and the background constraint. These equations are not independent; the first one can be obtained by taking a derivative of the first constraint and using the second equation and the background equation of motion. Equations (3.28), (3.30), and (3.31) can be combined to give the decoupled equation of motion for a_n:

$$a_n'' + 2\left[\mathcal{H} - \frac{\phi''}{\phi'} \right] a_n' + \left[2\mathcal{H}' - 2\mathcal{H}\frac{\phi''}{\phi'} + n^2 + 3 \right] a_n = 0 . \tag{3.32}$$

This equation is useful in the inflationary era, where $\phi' \neq 0$. It is also useful in the limit where the curvature of the three-space can be neglected, as we can solve it explicitly in either the adiabatic or nonadiabatic regime (see [22]). Once we have a solution for a_n, we can also find f_n using the constraint equation (3.30) or (3.31), and therefore the real degree of freedom z_n. In the region near the maximum expansion it is much harder to solve (3.32), and we return to (3.28)–(3.31).

2. No-boundary proposal mode function

Let us now construct the solutions of (3.28)–(3.32) selected by the no-boundary proposal. We focus only on modes which go outside the Hubble radius during inflation. These are the ones which get excited by the varying gravitational field. The very-high-frequency modes remain adiabatic throughout the history of the Universe, so the variation of their amplitudes will agree with the cosmological arrow of time. As in the graviton case, we divide the background saddle-point four-geometry into an approximately Euclidean section, followed by an inflationary one, which finally turns into dust. We have, however, to take into account the detailed behavior of the background scalar field ϕ as it couples directly to the perturbations. We first find the regular Euclidean solutions and match them up to the ones in the inflationary phase. This can be done by analytic continuation. In the inflationary era, the modes oscillate for a while until they leave the Hubble radius. At that point we match them to nonadiabatic solutions. Finally, the inflationary era comes to an end when ϕ becomes small and starts oscillating, behaving like a dust background. At this point we match on the solutions for the dustlike phase. It turns out that, for the Euclidean and inflationary solutions, the right-hand terms in (3.29) are negligible. We can solve for the scalar field modes f_n and calculate a_n from an integral version of the constraint (3.30), and check that this agrees with the approximate solutions of (3.32). If these terms were negligible during the whole of the dust era, the modes would oscillate adiabatically around the maximum expansion, as in the graviton case. However, we show that these terms do contribute to a monotonically increasing amplitude of the scalar field perturbations around maximum expansion.

The no-boundary proposal requires that the matter fields in the path integral be regular, so in the semiclassical approximation we look for solutions to the Euclidean perturbation equations which are regular as $\tau \to 0$. The regularity condition requires that f_n and a_n vanish as $\tau \to 0$. For $n \gg 1$, the dominant terms of Eq. (3.32) are the second derivative of a_n and $-n^2 \times a_n$, and one can construct a WKB solution. The approximate Euclidean solution selected by the no-boundary proposal is

$$a_n \approx A\frac{\phi'}{a}e^{n\eta_E}, \quad f_n \approx -\frac{An}{3}\frac{e^{n\eta_E}}{a}, \tag{3.33}$$

for some complex constant A. Here, the conformal time $\eta_E = 0$ corresponds to the juncture of Euclidean and Lorentzian spacetimes. Continuing the regular Euclidean solution into the Lorentzian section, taking $\eta_E \to i\eta$, gives

$$a_n \approx \tfrac{1}{3}im\,Ae^{in\eta}, \quad f_n \approx -\frac{An}{3}\frac{e^{in\eta}}{a}, \tag{3.34}$$

where we have used $\phi'/a = im/3$ during inflation (dash now denotes a Lorentzian time derivative). The analytical continuation holds into the inflationary era as long as the wavelength is smaller than the Hubble radius, i.e., $n \gg \mathcal{H}$. By this time inflation has begun, and we can match onto the inflationary solutions. When the modes move outside the Hubble radius, the modes a_n and f_n stop oscillating. They both have decaying and growing modes (the latter would be constant in the limit of exact de Sitter space). As the Universe inflates only the slowly growing mode remains [22], so that

$$a_n \approx \frac{D}{\phi^2}, \quad f_n \approx \frac{D}{\phi}, \tag{3.35}$$

where

$$D = \frac{1}{3}mi\,Ae^{in\eta_H}\left[\phi_H^2 + \frac{in\phi_H}{ma_H}\right]$$

is a constant depending on the detailed matching of the modes when they cross the Hubble radius at the time η_H. This solution is valid until the background scalar field decreases to $\phi \sim 1$. Figure 3 depicts the behavior of a_n during inflation and the beginning of the dust phase.

Eventually inflation ends and the background scalar field begins to oscillate. We expect that the background will behave effectively as a dust-filled Universe [see Eq. (2.16)] for perturbation modes with physical wavelengths much larger than the scalar field Compton wavelength ($n \ll ma$), since the pressure of the oscillating scalar field averages to zero over that wavelength scale. Therefore, the metric perturbations will behave like those of a pure dust universe (see, e.g., [22]). This is indeed what is found below.

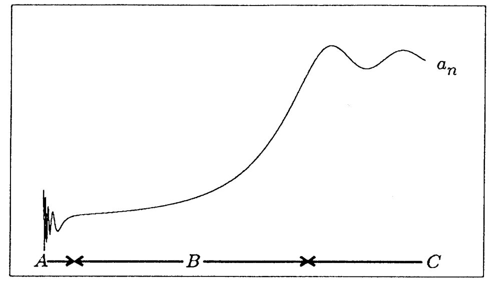

FIG. 3. The amplitude of a particular scalar perturbation a_n during the inflationary scenario and beginning of the dust period. (a) The mode oscillates adiabatically when the wavelength is smaller than the Hubble radius. (b) It freezes when it crosses the Hubble radius in the inflationary era. It slowly increases as $1/\phi^2$ until the end of inflation. (c) At the beginning of the dust era, the mode is essentially constant with a small oscillatory component due to the oscillations of the background scalar field.

During inflation the Hubble radius H^{-1} is roughly constant, but as the Universe evolves in the dust era the Hubble radius starts growing. When it becomes larger than the Compton wavelength $1/ma$, the dominant term in (3.29) is m^2a^2. The perturbation of the scalar field will start oscillating again. In this early state of the dust era when the curvature of the three-surface is negligible, it can be shown that the f_n oscillate exactly in phase with ϕ' as

$$f_n \approx -\frac{\phi'}{a}\int d\eta\, a a_n . \tag{3.36}$$

This will remain true in later stages of the dust era as long as $n < ma_e$. This condition ensures that the phase of f_n obtained by integrating (3.29) does not differ appreciably from that of ϕ'. Using (3.36), together with (3.30), we can establish that the metric perturbation a_n, time averaged over one oscillation period of π/m, is growing. The small oscillations around this average arise because

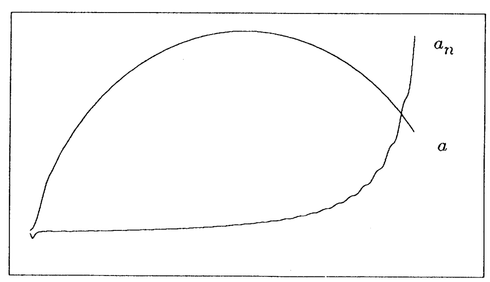

FIG. 4. Classical trajectories for the radius of the Universe a in the FRW model with scalar field. The amplitude of a particular scalar perturbation a_n does not have a time symmetric amplitude with respect to the maximum expansion.

the background energy-momentum tensor is not exactly that of dust, but that of an oscillating scalar field. The averaged gravitational perturbation a_n^A can be calculated by taking the derivative of the averaged version of (3.30) to obtain the differential equation

$$a_n^{A''}+3\mathcal{H}a_n^{A'}-2a_n^A=0\ . \tag{3.37}$$

The general solution is a linear combination of the solutions

$$a_n^{\text{anti}}\approx\frac{\sin\eta_d}{(1-\cos\eta_d)^3} \tag{3.38}$$

and

$$a_n^{\text{sym}}\approx\frac{2\sin^2\eta_d-6(\eta_d-\pi)\sin\eta_d-8\cos\eta_d+8}{(1-\cos\eta_d)^3}\ . \tag{3.39}$$

The conformal time is defined with the new origin at the beginning of the dust phase ($\eta_d\approx 0$). These solutions are antisymmetric and symmetric with respect to the maximum of expansion ($\eta_d=\pi$), and are the same solutions found for perturbations in a pressureless perfect fluid Universe, as expected. Both solutions diverge like η_d^{-5} in the beginning of the dust era as $\eta_d\to 0$. There is, however, a regular solution, given by

$$a_n^{\text{reg}}:=a_n^{\text{symm}}-6\pi a_n^{\text{anti}}\ ,$$

which approaches a constant in this limit. At the end of inflation, the a_n picked out by the no-boundary proposal are small as seen from (3.35). Therefore, the regular solution is the one selected by the no-boundary proposal, and this is asymmetric in the dust era: the perturbation amplitude steadily increases with time (See Fig. 4.). Matching the solutions for the dust era to (3.35) shows that, during the dust era,

$$a_n\approx Da_n^{\text{reg}}\ . \tag{3.40}$$

We can now use (3.36) to see that f_n is oscillating with monotonically increasing amplitude throughout the dust era:

$$f_n\approx-\frac{D\phi'}{(1-\cos\eta_d)^2}[4\eta_d-6\sin\eta_d+2\eta_d\cos\eta_d]\ . \tag{3.41}$$

With these solutions, we can construct the wave function (3.26). When the background saddle-point is approximately Lorentzian, the no-boundary wave function for the scalar perturbation is

$$\psi^s(z_n)\sim C(a,\phi)\exp\left\{-i\left[M\left[\frac{\mu_n'}{\mu_n}\right]+N(a,\phi)\right]z_n^2\right\}\ , \tag{3.42}$$

where M is given in (3.27) and μ_m is the mode function for z_n. It is a solution of the equation of motion for z_n picked out by the no-boundary proposal. It is explicitly given by the functions a_n and f_n, using (3.25) with z_n replaced by μ_n. From the solutions of a_n and f_n, we can see that it is clearly asymmetric about the time of maximum expansion. Considering points placed symmetrically about the maximum of expansion, the background will be the same at both points, so that the asymmetry in the mode function manifests itself as an asymmetry in the wave function. The variance of z_n is proportional to the modulus of μ_n, and is therefore asymmetric with respect to the time of maximum expansion. We therefore conclude that the wave function predicts the continuing growth of low-frequency scalar perturbations, even when the Universe begins to recollapse.

This result alone provides a time asymmetry so long as the modes stay in a regime where they can be treated in a linear approximation. However, most modes will also enter a nonlinear regime well before the maximum expansion occurs. When this occurs, the interaction terms in the Hamiltonian will become important, and hence the coarse-grained entropy will increase throughout the evolution.

Considering the stress tensor in the gauge-invariant formalism (see, e.g., [22]), we can show that the density contrast is

$$\frac{\delta\rho}{\rho}\approx\frac{2a}{3a_m}[(n^2-4)a_n-9\mathcal{H}\phi' f_n]\ . \tag{3.43}$$

Modes will cross the horizon ($\mathcal{H}\sim n$) when $\eta_d\sim 1/n$, and the recent COBE results [5] indicate that the density contrast at this time is of order 10^{-5}. Using Eqs. (3.40) and (3.41) in (3.43), we find that the constant D is of order 10^{-5}. At later times, only the first term in (3.43) is important, and we find that the density contrast behaves like

$$\frac{\delta\rho}{\rho}\approx 10^{-7}n^2\eta_d^2\ . \tag{3.44}$$

Consequently, when the density contrast is of order unity, we expect nonlinearity to be the dominant feature, and this occurs for $\eta_d^2\geq 10^7/n^2$. Modes with $n\geq 1000$ will therefore enter a nonlinear phase before they reach the maximum, and the coarse-grained entropy for these modes will grow.

IV. CONCLUSION

In this paper we have investigated the consequences of the no-boundary proposal for the arrow of time. In particular, we have investigated the behavior of small metric and matter perturbations around a homogeneous isotropic background. The no-boundary proposal predicts classical evolution with an inflationary area followed by a dustlike era. We found that perturbation modes are in their ground state at the beginning of the inflationary era. This can be interpreted as a statement that the Universe is born in a low entropy state. Modes which leave the Hubble radius during inflation become excited, then subsequently evolve in various ways in the dustlike era.

We find that gravitons oscillate adiabatically for most of the dustlike era, and consequently the amplitude of their oscillations is time symmetric with respect to the point of maximum scale factor. However, looking at the physical scalar degrees of freedom we find that those which have been excited by superadiabatic amplification during inflation have a time asymmetric evolution with respect to the maximum. In particular, the variance of the scalar modes predicted by the wave function is different at the same value of the scale factor before and after the maximum.

Thus we find that the wave function of the Universe distinguishes between symmetrically placed points on either side of maximum volume. The expanding phase has a smaller amplitude of the variance in the low-frequency scalar modes than does the corresponding point during the collapsing phase. In other words, the thermodynamic arrow coincides with the cosmological arrow before the maximum, but points in the opposite direction after the maximum. This is true for all the lowest-frequency modes, so that they induce a well-defined thermodynamic arrow of time. Amongst the modes which display this nonadiabatic behavior, higher-frequency modes will enter a nonlinear regime during the expansion, and consequently produce a growing coarse-grained entropy throughout expansion and recontraction, and hence also create a thermodynamic arrow of time.

ACKNOWLEDGMENTS

We would like to thank Nathalie Deruelle and Carsten Gundlach for useful conversations. R.L. acknowledges Peterhouse, Cambridge, for financial support. G.W.L. acknowledges the Science and Engineering Research Council for supporting this work.

APPENDIX A: ACTION AND FIELD EQUATIONS

In this Appendix, we reproduce the action and field equations of the perturbed FRW model driven by a massive minimally coupled scalar field from Ref. [6]. The homogeneous part of the Einstein-Hilbert Lagrangian is

$$L_0=-\frac{1}{2}N_0a^3\left[\frac{\dot a^2}{N^2a^2}-\frac{1}{a^2}-\frac{\dot\phi^2}{N^2}+m^2\phi^2\right] . \tag{A1}$$

The second-order perturbation of the Einstein-Hilbert Lagrangian is

$$\begin{aligned}L_g^n=\tfrac{1}{2}aN_0\Bigg\{&\frac{1}{3}\left[n^2-\frac{5}{2}\right]a_n^2+\frac{n^2-7}{3}\frac{n^2-4}{n^2-1}b_n^2-2(n^2-4)c_n^2-(n^2+1)d_n^2+\frac{2}{3}(n^2-4)a_nb_n\\&+\frac{2}{3}g_n[(n^2-4)b_n+(n^2+\tfrac{1}{2})a_n]+\frac{1}{N_0^2}\left[-\frac{1}{3(n^2-1)}k_n^2+(n^2-4)j_n^2\right]\Bigg\}\\&+\frac{1}{2}\frac{a^3}{N_0}\Bigg\{-\dot a_n^2+\frac{n^2-4}{n^2-1}\dot b_n^2+(n^2-4)\dot c_n^2+\dot d_n^2+g_n\left[2\frac{\dot a}{a}\dot a_n+\frac{\dot a^2}{a^2}(3a_n-g_n)\right]\\&+\frac{\dot a}{a}\left[-2a_n\dot a_n+8\frac{n^2-4}{n^2-1}b_n\dot b_n+8(n^2-4)c_n\dot c_n+8d_n\dot d_n\right]\\&+\frac{\dot a^2}{a^2}\left[-\frac{3}{2}a_n^2+6\frac{n^2-4}{n^2-1}b_n^2+6(n^2-4)c_n^2+6d_n^2\right]\\&+\frac{1}{a}\left[\frac{2}{3}k_n\left\{-\dot a_n-\frac{n^2-4}{n^2-1}\dot b_n+\frac{\dot a}{a}g_n\right\}-2(n^2-4)\dot c_nj_n\right]\Bigg\} .\end{aligned} \tag{A2}$$

The perturbation of the matter Lagrangian gives

$$\begin{aligned}L_m^n=\tfrac{1}{2}N_0a^3\Bigg\{&\frac{1}{N_0^2}(\dot f_n^2+6a_n\dot f_n\dot\phi)-m^2(f_n^2+6a_nf_n\phi)-\frac{1}{a^2}(n^2-1)f_n^2+\frac{\dot\phi^2}{N_0^2}g_n^2\\&+\frac{3}{2}\left[\frac{\dot\phi^2}{N_0^2}-m^2\phi^2\right]\left[a_n^2-4\frac{n^2-4}{n^2-1}b_n^2-4(n^2-4)c_n^2-4d_n^2\right]\\&-g_n\left[2m^2f_n\phi+3m^2a_n\phi^2+2\frac{\dot f_n\dot\phi}{N_0^2}+3\frac{a_n\dot\phi^2}{N_0^2}\right]-2\frac{1}{aN_0^2}k_mf_n\dot\phi\Bigg\} .\end{aligned} \tag{A3}$$

The field equations necessary to calculate the saddle-point approximation are given below. From (A1) we find the equations obeyed by the homogeneous background fields. The homogeneous scalar field φ_0 obeys

$$N_0\frac{d}{dt}\left[\frac{1}{N_0}\frac{d\varphi_0}{dt}\right]+3\frac{da}{a\,dt}\frac{d\varphi_0}{dt}+N_0m^2\varphi_0=\text{quadratic terms} , \tag{A4}$$

and the scale factor a obeys

$$N_0\frac{d}{dt}\left[\frac{1}{N_0 a}\frac{da}{dt}\right]+3\dot{\varphi}_0^2-\frac{N_0^2}{a^2}-\frac{3}{2}\left[-\frac{\dot{a}^2}{a^2}+\dot{\varphi}_0^2-\frac{N_0^2}{a^2}+N_0^2m^2\phi^2\right]=\text{quadratic terms}\ . \tag{A5}$$

The background variables a, φ_0, and their momenta are subject to the constraint

$$-\frac{\dot{a}^2}{a^2N_0^2}+\frac{\dot{\varphi}_0}{N_0^2}-\frac{1}{a^2}+m^2\varphi_0^2=\text{quadratic terms}\ . \tag{A6}$$

Let us now turn to the equation of motion of the small inhomogeneities. Variations with respect to a_n, b_n, c_n, d_n, and f_n give the following second-order field equations:

$$N_0\frac{d}{dt}\left[\frac{a^3}{N_0}\frac{da_n}{dt}\right]+\frac{1}{3}(n^2-4)N_0^2a(a_n+b_n)+3a^3(\dot{\varphi}_0\dot{f}_n-N_0^2m^2\varphi_0f_n)$$
$$=N_0^2[3a^3m^2\varphi_0^2-\tfrac{1}{3}(n^2+2)a]g_n+a^2\dot{a}\dot{g}_n-\frac{1}{3}N_0\frac{d}{dt}\left[\frac{a^2k_n}{N_0}\right]\ , \tag{A7}$$

$$N_0\frac{d}{dt}\left[\frac{a^3}{N_0}\frac{db_n}{dt}\right]-\frac{1}{3}(n^2-1)N_0^2a(a_n+b_n)=\frac{1}{3}N_0^2(n^2-1)ag_n+\frac{1}{3}N_0\frac{d}{dt}\left[\frac{a^2k_n}{N_0}\right]\ , \tag{A8}$$

$$N_0\frac{d}{dt}\left[\frac{a^3}{N_0}\frac{dc_n}{dt}\right]=\frac{d}{dt}\left[\frac{a^2j_n}{N_0}\right]\ , \tag{A9}$$

$$N_0\frac{d}{dt}\left[\frac{a^3}{N_0}\frac{dd_n}{dt}\right]+(n^2-1)N_0^2ad_n=0\ , \tag{A10}$$

and

$$N_0\frac{d}{dt}\left[\frac{a^3}{N_0}\frac{df_n}{dt}\right]+3a^3\dot{\varphi}_0\dot{a}_n+N_0^2[m^2a^3+(n^2-1)a]f_n=a^3\left[-2N_0^2m^2\varphi_0g_n+\dot{\varphi}_0\dot{g}_n-\frac{\varphi_0k_n}{a}\right]\ . \tag{A11}$$

Variations with respect to k_n, j_n, and g_n lead to the constraints

$$\dot{a}_n+\frac{n^2-4}{n^2-1}\dot{b}_n+3f_n\dot{\varphi}_0=\frac{\dot{a}g_n}{a}-\frac{k_n}{a(n^2-1)}\ , \tag{A12}$$

$$\dot{c}_n=\frac{j_n}{a}\ , \tag{A13}$$

and

$$3a_n\left[\dot{\varphi}_0^2-\frac{\dot{a}^2}{a^2}\right]+2\left[\dot{\varphi}_0\dot{f}_n-\frac{\dot{a}\dot{a}_n}{a}\right]+N_0^2m^2(2f_n\varphi_0+3a_n\varphi_0^2)-\frac{2N_0^2}{3a^2}[(n^2-4)b_n+(n^2+\tfrac{1}{2})a_n]=\frac{2\dot{a}k_n}{3a^2}+2g_n\left[\dot{\varphi}_0^2-\frac{\dot{a}^2}{a^2}\right]\ . \tag{A14}$$

We also give the perturbation Hamiltonian in terms of the real degrees of freedom:

$$H_2^n(z_n,\pi_{z_n})=A\pi_{z_n}^2+Bz_n\pi_{z_n}+Cz_n^2\ ,$$

with

$$A=\frac{1}{2}\left[a\dot{a}^2+\frac{3a^3\dot{\phi}^2}{n^2-4}\right],\quad B=\frac{1}{2U}\left[K\left[2a-3a^3m^2\phi^2-3\frac{n^2-1}{n^2-4}a^3\dot{\phi}^2\right]+a^9m^4\phi^2-3a^8m^2\phi\dot{\phi}\dot{a}\right]\ ,$$

$$C=-\frac{1}{2U^2}\left[-\frac{3(n^2-1)K^3}{(n^2-4)a^3}+a^3m^2K^2-5a^9m^4\phi^2K+12a^{15}m^4\phi^2\dot{\phi}^2\right]\ , \tag{A15}$$

$$K=-3a^6\dot{\phi}^2-\frac{n^2-4}{3}a^4,\quad U=-Ka^2\dot{a}-a^9m^2\phi\dot{\phi}\ .$$

APPENDIX B: RELATION TO THE GAUGE-INVARIANT FORMALISM

There has recently been much interest in the gauge-invariant formalism [21,22] which cast the variables of the theory (the scalar perturbations of the gravitational and scalar fields) into ones which are invariant under infinitesimal gauge transformations. In this Appendix we relate the different harmonics in Eqs. (3.4) and (3.5) to the gauge-invariant variables (in particular we shall follow the approach of [22]).

Mukhanov *et al.* define the time- and space-dependent scalar metric perturbations as

$$ds^2=a^2(\eta)\{(1+2\phi)d\eta^2-2B_{|i}dx^i d\eta + [(1-2\psi)\gamma_{ij}+2E_{|ij}]dx^i dx_j\} , \tag{B1}$$

and the scalar field perturbations

$$\varphi(\mathbf{x},t)=\varphi_0(t)+\delta\varphi(\mathbf{x},t) . \tag{B2}$$

The above variables are related to the modes perturbations used in this paper in the following way:

$$\begin{aligned}\phi&=\sum_n \frac{g_n Q^n}{\sqrt{6}} ,\\ \psi&=\sum_n \frac{-(a_n+b_n)Q^n}{\sqrt{6}} ,\\ B&=\sum_n \frac{k_n Q^n}{(n^2-1)\sqrt{6}} ,\\ E&=\sum_n \frac{3b_n Q^n}{(n^2-1)\sqrt{6}} ,\end{aligned} \tag{B3}$$

where we have suppressed in the sum the indices lm corresponding to the angular momentum.

Under a general linear gauge transformation of the form

$$\eta\to\tilde{\eta}=\eta+\xi^0(\eta,\mathbf{x}) , \quad x^i\to\tilde{x}^i=x^i+\gamma^{ij}\xi_{|j}(\eta,\mathbf{x}) , \tag{B4}$$

the scalar perturbations transform as

$$\begin{aligned}\tilde{\phi}&=\phi-\frac{a'}{a}\xi^0-\xi^{0'} ,\\ \tilde{\psi}&=\psi+\frac{a'}{a}\xi^0 ,\\ \tilde{B}&=B+\xi^0-\xi' ,\\ \tilde{E}&=E-\xi ,\\ \tilde{\delta\phi}&=\delta\phi-\varphi_0'\xi^{0'} .\end{aligned} \tag{B5}$$

The idea of the gauge-invariant formalism is to make a linear combination of the different scalar perturbations such that the resulting variables are independent of the gauge. A possible choice is

$$\begin{aligned}\Phi&=\phi+\frac{1}{a}[(B-E')a]' ,\\ \Psi&=\psi-\frac{a'}{a}(B-E') ,\\ \delta\varphi^{(gi)}&=\delta\varphi+\varphi_0'(B-E') .\end{aligned} \tag{B6}$$

These gauge-invariant quantities obey the equations

$$\nabla^2\Phi-3\mathcal{H}\Phi'-(\mathcal{H}'+2\mathcal{H}^2-4K)\Phi=\frac{3l^2}{2}(\varphi'\delta\varphi^{(gi)'}+V_{,\varphi}a^2\delta^{(gi)}) , \tag{B7}$$

$$\Phi'+\mathcal{H}\Phi=3\frac{l^2}{2}\varphi'\delta\varphi^{(gi)} , \tag{B8}$$

$$\Phi''+3\mathcal{H}\Phi'+(\mathcal{H}'+2\mathcal{H}^2)\Phi=\frac{3l^2}{2}(\varphi'\delta\varphi^{(gi)'}-V_{,\varphi}a^2\delta\varphi^{(gi)}) , \tag{B9}$$

which are the gauge-invariant versions of the $\delta G_0^0=8\pi G\delta T_0^0$, $\delta G_i^0=8\pi G\delta T_i^0$, and $\delta G_j^i=8\pi G\delta T_j^i$ equations, and

$$\delta\varphi^{(gi)''}+2\mathcal{H}\delta\varphi^{(gi)'}-\nabla^2\delta\varphi^{(gi)}+V_{,\varphi\varphi}a^2\delta\varphi^{(gi)}-4\varphi_0'\Phi'+2V_{,\varphi}a^2\Phi=0 \tag{B10}$$

is the gauge-invariant version of the scalar field equation.

In the longitudinal gauge ($B=k_n=0$ and $E=b_n=0$) used in this paper, the gauge variables reduce to $\Phi=\phi$, $\Psi=\psi$, and $\delta\varphi^{(gi)}=\delta\varphi$, and, if we expand them in harmonics on the three-sphere, $\Phi_n=g_n/\sqrt{6}$, $\Psi_n=-a_n/\sqrt{6}$, and $\delta\varphi^{(gi)}=f_n/\sqrt{6}$. Indeed, it is easy to see that Eqs. (B9) and (B10) are equivalent to (3.31) and (3.32) respectively, and that the constraint (B8) is equivalent to (3.33).

[1] R. F. Streater and A. S. Wightman, *PCT, Spin Statistics and All That* (Benjamin, New York, 1964).
[2] J. H. Christenson, J. W. Cronin, V. L. Fitch, and R. Turley, Phys. Rev. Lett. **13**, 138 (1964).
[3] T. Banks, Nucl. Phys. **B249**, 332 (1985); R. Penrose, in *General Relativity—An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1979).
[4] J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).
[5] G. Smoot *et al.*, Astrophys. J. **396**, L1 (1992).
[6] J. J. Halliwell and S. W. Hawking, Phys. Rev. D **31**, 1777 (1985).
[7] S. W. Hawking, Phys. Rev. D **32**, 2489 (1985).
[8] S. W. Hawking and Z. C. Wu, Phys. Lett. **107B**, 15 (1985); D. N. Page, in *Quantum Concepts in Space and Time*, edited by R. Penrose and C. J. Isham (Clarendon Press, Oxford, 1986).
[9] D. Page, Phys. Rev. D **32**, 2496 (1985).
[10] R. Laflamme, Ph.D. thesis, Cambridge University, 1988.
[11] R. Laflamme and E. P. S. Shellard, Phys. Rev. D **35**, 2315 (1987).
[12] J. J. Halliwell and J. B. Hartle, Phys. Rev. D **41**, 1815 (1990).
[13] G. Lyons, Phys. Rev. D **46**, 1546 (1992).
[14] S. Habib and R. Laflamme, Phys. Rev. D **42**, 4056 (1990).
[15] E. M. Lifschitz and I. M. Khalatnikov, Adv. Phys. **12**, 185 (1963).
[16] U. H. Gerlach and U. K. Sengupta, Phys. Rev. D **18**, 1773 (1978).
[17] L. P. Grishchuk, Zh. Eksp. Teor. Fiz. **67**, 825 (1974) [Sov. Phys. JETP **40**, 409 (1975)].
[18] R. Laflamme and J. Louko, Phys. Rev. D **43**, 2730 (1991).
[19] L. P. Grishchuk and Yu. V. Sidorov, Class. Quantum Grav. **6**, L161 (1989).
[20] I. Shirai and S. Wada, Nucl. Phys. **B303**, 728 (1988).
[21] J. Bardeen, Phys. Rev. D **22**, 1882 (1980).
[22] V. F. Mukhanov, H. A. Feldman, and R. H. Brandenberger, Phys. Rep. **215**, 203 (1992).