Classical and quantum cosmology of the Starobinsky inflationary model

Alexander Vilenkin*
Lyman Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138
(Received 19 April 1985)

Starobinsky has suggested an inflationary cosmological scenario in which the inflation is driven by quantum corrections to the vacuum Einstein's equations. Here a detailed review of the Starobinsky scenario is given and an observational constraint on the parameters of the model is derived. The quantum mechanics of the model is studied first using the instanton method, and then by solving the corresponding Wheeler-DeWitt equation. A cosmological wave function is obtained describing a universe tunneling from "nothing" to the Starobinsky inflationary phase. The curvature fluctuations in the tunneling universe are calculated. This quantum analysis determines the initial conditions for the classical evolution of the model.

I. INTRODUCTION

The idea that a closed universe can be created spontaneously as a quantum fluctuation was first suggested by Tryon[1] in 1973 (see also Ref. 2). He noted that a closed universe has zero charge, energy, and momentum, and can have all other conserved charges equal to zero. Hence, no conservation law prevents such a universe from being spontaneously created. Tryon suggested no mathematical description for the nucleation process, and so the initial size and the content of the universe were left undetermined.

Intuitively, one expects a universe created as a quantum fluctuation to be very small. This was a serious difficulty in 1973, since it was not clear how to make a large universe we live in out of a tiny closed universe. The problem has disappeared with the advent of inflationary scenarios[3–5] in which the universe passes through a de Sitter phase of exponential expansion. As a result of inflation, all scales in the universe are increased by a huge factor $Z=\exp(H\tau)$, where H is the constant expansion rate and τ is the duration of the inflationary phase.

In most inflationary scenarios the inflation is driven by the false-vacuum energy density ρ_v. In this case the expansion rate is

$$H=(8\pi G\rho_v/3)^{1/2} . \tag{1.1}$$

Starobinsky[5] has suggested an alternative model in which the de Sitter phase is obtained as a self-consistent solution of the vacuum Einstein's equations modified by the one-loop corrections due to quantized matter fields.[6] In this scenario the expansion rate depends on the number of elementary fields in the model, but is typically $\sim m_p$, where m_p is the Planck mass.[7]

In both versions of inflation, the de Sitter phase is unstable and has a finite lifetime. Therefore, one is confronted with the question of what happened before inflation and how the universe got into the de Sitter phase. A false-vacuum-dominated phase can be reached in the hot-big-bang model as a result of a first-order phase transition with strong supercooling.[3] No such mechanism was suggested for the Starobinsky model, and it appears that the model is consistent only if the universe is spontaneously created.[8]

The first attempt to describe the quantum creation of universes in the framework of quantum gravity was made in Ref. 9 and was further developed in Refs. 10—13. The picture emerging from this work[14] is that the universe tunnels quantum mechanically from "nothing" to a de Sitter space with a scale factor

$$a(t)=H^{-1}\cosh(Ht) . \tag{1.2}$$

Here, "nothing" means "no classical space-time." At the moment of nucleation $(t=0)$ the universe has size $a(0)=H^{-1}$ and zero "velocity," $\dot{a}(0)=0$. This is the beginning of time, and from that point on the universe evolves along the lines of the inflationary scenario. The initial hot phase is absent in this model. The semiclassical tunneling probability to a false vacuum of energy density ρ_v is given by[11,10]

$$P\propto\exp(-3/8G^2\rho_v) . \tag{1.3}$$

Quantum tunneling to the Starobinsky phase has been briefly discussed in Ref. 13.

It should be understood, of course, that the interpretation of probabilities like (1.3) is far from being clear. It appears that the best we can do is to compare the probabilities of tunneling to different states and assume that we live in a "typical" universe which starts in a state with the highest tunneling probability. Equation (1.3) suggests that of all the false vacua the tunneling is most probable to the one with the highest energy density, ρ_v. As ρ_v is increased, the tunneling action $(S=3/8G^2\rho_v)$ decreases and becomes ~ 1 near the Planck energy density, $\rho_v\sim G^{-2}$. This seems to suggest that for $\rho_v \gtrsim G^{-2}$ the barrier for the nucleation of the universe disappears and the semiclassical approximation breaks down. We note, however, that for energy densities and curvatures near the Planck scale, quantum corrections to Einstein's action become important. These corrections include quadratic and higher-order terms in curvature and can stop and reverse the decrease of the tunneling action at high curvatures. Thus, one is led to consider quantum cosmological models in which the universe nucleates with curvature and matter energy density near the Planck scale, and quantum gravitational corrections play an important role. The simplest model of this sort is the Starobinsky model which involves no classical matter fields.

The main purpose of this paper is to study the quantum cosmology of the Starobinsky model. The effective action for this model is nonlinear in curvature and the corresponding field equations involve fourth-order derivatives. Quantization of higher-derivative theories of gravity has been discussed in Refs. 16 and 17. In this paper we shall employ a simple minisuperspace model in which the infinite number of gravitational degrees of freedom is restricted to two pairs of conjugate variables (related to the scale factor and to the scalar curvature). Having in mind future applications, we shall develop the formalism for a general class of models with a higher-derivative action of the form

$$S=\int L(R)(-g)^{1/2}d^4x \;, \tag{1.4}$$

where $L(R)$ is an arbitrary function of the scalar curvature.

In the next section, I begin by reviewing the Starobinsky inflationary scenario. This analysis is more detailed and complete than that in Ref. 5. In particular, I give a detailed description of the decay of the Starobinsky phase and derive the spectrum of gravitational waves generated during inflation. Requiring that the background temperature fluctuations induced by the gravitational waves are not too large, I obtain an observational constraint on the parameters of the model. After this work was completed, I learned that Starobinsky obtained similar results in Ref. 18. Still, I think this review will be helpful, since Starobinsky's analysis in Refs. 5 and 18 is rather terse.

Sections III—V are devoted to the description of the quantum tunneling of the universe from "nothing" to the Starobinsky phase. In Sec. III the semiclassical tunneling probability is calculated using the instanton method and an order-of-magnitude estimate is obtained for the curvature fluctuations in the newly born universe. The knowledge of curvature fluctuations is important, since the Starobinsky phase is unstable, and its decay time depends on the initial magnitude of the curvature fluctuations.

A more detailed description of the tunneling universe is obtained by solving the Wheeler-DeWitt equation[19] which plays the role of the Schrödinger equation in quantum gravity. The Wheeler-DeWitt equation for higher-derivative models of the form (1.4) is derived in Sec. IV. Its solution for the Starobinsky model is found in Sec. V. The wave function describing a tunneling universe is appreciably different from zero only in a narrow range of curvatures around the instanton curvature $R=R_0$. The mean-square fluctuation of the curvature calculated from this wave function is in agreement with the order-of-magnitude estimate of Sec. III.

After the nucleation, the universe is described by the classical field equations with a very high accuracy. The role of quantum cosmology is, therefore, to determine the probability distribution of the initial conditions for the following classical evolution.

II. STAROBINSKY SCENARIO

A. de Sitter solution

The Starobinsky scenario is based on a self-consistent solution of the semiclassical Einstein's equations,

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=-8\pi G\langle T_{\mu\nu}\rangle \;, \tag{2.1}$$

where $\langle T_{\mu\nu}\rangle$ is the expectation value of the energy-momentum tensor. The metric is assumed to be of the Robertson-Walker form

$$ds^2=dt^2-a^2(t)d\sigma_\kappa{}^2 \;, \tag{2.2}$$

where $d\sigma_\kappa{}^2$ is the metric on a three-sphere, three-plane, and three-hyperboloid for closed ($\kappa=+1$), flat ($\kappa=0$), and open ($\kappa=-1$) Robertson-Walker metrices, respectively. In a curved space-time, even in the absence of classical matter or radiation, quantum fluctuations of matter fields give nontrivial contributions to $\langle T_{\mu\nu}\rangle$. (This effect is similar to vacuum polarization effects in quantum electrodynamics.) These quantum corrections take a particularly simple form in the case of free, massless, conformally invariant fields:[20]

$$\langle T_{\mu\nu}\rangle=k_1{}^{(1)}H_{\mu\nu}+k_3{}^{(3)}H_{\mu\nu} \;. \tag{2.3}$$

Here, k_1 and k_3 are numerical coefficients and I have used the standard notations

$${}^{(1)}H_{\mu\nu}=2R_{;\mu;\nu}-2g_{\mu\nu}R_{;\sigma}{}^{;\sigma}+2RR_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R^2 \;, \tag{2.4a}$$

$${}^{(3)}H_{\mu\nu}=R_\mu{}^\sigma R_{\nu\sigma}-\tfrac{2}{3}RR_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R^{\sigma\tau}R_{\sigma\tau}+\tfrac{1}{4}g_{\mu\nu}R^2 \;. \tag{2.4b}$$

The tensor ${}^{(1)}H_{\mu\nu}$ is identically conserved, ${}^{(1)}H^{\nu}_{\mu;\nu}=0$. It can be obtained by varying a local action,

$${}^{(1)}H_{\mu\nu}=\frac{2}{\sqrt{-g}}\frac{\delta}{\delta g^{\mu\nu}}\int d^4x\sqrt{-g}\,R^2 \;. \tag{2.5}$$

To cancel infinities in $\langle T_{\mu\nu}\rangle$, one has to introduce infinite counterterms in the gravitational Lagrangian. One of these counterterms has the form $CR^2\sqrt{-g}$, where C is a logarithmically divergent constant. Since an arbitrary finite part can be added to C, the coefficient k_1 in Eq. (2.3) can take any value and has, in principle, to be determined by experiment. The tensor ${}^{(3)}H_{\mu\nu}$, on the other hand, is conserved only in conformally flat space-times (in particular, in Robertson-Walker space-times) and cannot be obtained by varying a local action. Its coefficient k_3 is uniquely determined:

$$k_3=\frac{1}{1440\pi^2}(N_0+\tfrac{11}{2}N_{1/2}+31N_1) \;, \tag{2.6}$$

where N_0, $N_{1/2}$, and N_1 are the numbers of quantum fields with spins 0, $\frac{1}{2}$, and 1, respectively. It will be convenient to introduce the notations

$$H_0=(8\pi k_3 G)^{-1/2}\ ,\quad M=(48\pi k_1 G)^{-1/2}\ . \tag{2.7}$$

Then Eq. (2.3) takes the form

$$8\pi G\langle T_{\mu\nu}\rangle = H_0^{-2\,(3)}H_{\mu\nu}+\tfrac{1}{6}M^{-2\,(1)}H_{\mu\nu}\ . \tag{2.8}$$

We shall assume that $H_0>0$, $M>0$.

Although the trace of the energy-momentum tensor vanishes for classical conformally invariant fields, the expectation value (2.8) has a nonzero anomalous trace:

$$8\pi G\langle T_\nu^\nu\rangle = H_0^{-2}(\tfrac{1}{3}R^2-R_{\nu\sigma}R^{\nu\sigma})-M^{-2}R_{;\nu}{}^{;\nu}\ . \tag{2.9}$$

This is the so-called trace anomaly;[20,21] the conformal invariance is broken by the regularization of infinities in $\langle T_{\mu\nu}\rangle$.

A model with free, massless, conformally invariant fields may appear rather artificial. We note, however, that the masses of the fields can be neglected in the high-curvature limit, $R \gg m^2$. In asymptotically free gauge theories, interactions become negligible in the same limit.[22] Besides, in a typical grand unified model, the main contribution to k_3 comes from the vector fields (because of the large factor multiplying N_1). Massless vector fields (as well as spinor fields) are described by conformally invariant equations, and their contribution to $\langle T_{\mu\nu}\rangle$ is of the form (2.8). Thus Eq. (2.8) may give a reasonable approximation for $\langle T_{\mu\nu}\rangle$ in a grand unified model for $R>\mu^2$, where μ is the unification energy scale. Corrections to Einstein's equations due to gravitons can be neglected compared to those due to matter fields if the number of matter fields is sufficiently large.

It is easily verified that Eq. (2.1) with $\langle T_{\mu\nu}\rangle$ from Eq. (2.8) has a de Sitter solution.[23] In a de Sitter space

$$R_{\mu\nu}=\tfrac{1}{4}g_{\mu\nu}R\ ,\quad R=\text{const}\ . \tag{2.10}$$

Substituting (2.10) in (2.1) and (2.8) and disregarding the trivial solution $R=0$, we obtain $R=12H_0^2$. The corresponding de Sitter solutions are

$$a(t)=H_0^{-1}\cosh(H_0t)\ ,\quad \kappa=+1 \tag{2.11a}$$

$$a(t)=a_0\exp(H_0t)\ ,\quad \kappa=0 \tag{2.11b}$$

$$a(t)=H_0^{-1}\sinh(H_0t)\ ,\quad \kappa=-1 \tag{2.11c}$$

for closed, flat, and open models, respectively. These solutions describe an inflationary phase driven entirely by the quantum corrections to Einstein's equations.

The magnitude of H_0 depends on the numbers of fields in Eq. (2.6), but is not, typically, much different from the Planck mass, m_P. For example, in the minimal SU(5) model, $N_0=34$, $N_{1/2}=45$, $N_1=24$, $8\pi k_3=1.8$, and $H_0=0.7m_P$.

B. Instability of the de Sitter phase

The evolution equation for the scale factor obtained from Eqs. (2.1) and (2.8) is[5,24]

$$\frac{\dot a^2+\kappa}{a^2}=\frac{1}{H_0^2}\left[\frac{\dot a^2+\kappa}{a^2}\right]^2-\frac{1}{M^2}\left[2\frac{\dot a\dddot a}{a^2}-\frac{\ddot a^2}{a^2}+2\frac{\ddot a\dot a^2}{a^3}-3\left[\frac{\dot a}{a}\right]^4-2\kappa\frac{\dot a^2}{a^4}+\frac{\kappa^2}{a^4}\right]\ . \tag{2.12}$$

In the de Sitter solutions (2.11) the scale factor grows exponentially, and for $H_0t>1$ the κ-dependent terms in (2.12) become negligible. It is, therefore, sufficient to study the flat-space model with $\kappa=0$, which is much simpler to analyze.

Introducing a new variable $H(t)=\dot a/a$, we can rewrite Eq. (2.12) with $\kappa=0$ as

$$H^2(H^2-H_0^2)=\frac{H_0^2}{M^2}(2H\ddot H+6H^2\dot H-\dot H^2)\ . \tag{2.13}$$

The de Sitter solution (2.11b) corresponds to $H=H_0$. To show that this solution is unstable, consider a small deviation from $H=H_0$:

$$H=H_0(1+\delta)\ . \tag{2.14}$$

Substituting this in Eq. (2.13) and linearizing in δ we obtain

$$\ddot\delta+3H_0\dot\delta-M^2\delta=0\ . \tag{2.15}$$

The two linearly independent solutions of (2.15) are given by $\delta=\exp(\alpha t)$ with

$$\alpha=-\frac{3H_0}{2}\pm\left[\frac{9H_0^2}{4}+M^2\right]^{1/2}\ .$$

The existence of a growing mode with $\alpha>0$ indicates the instability of the de Sitter solution (2.11). Note that flat space-time, $H=0$, is a stable solution of (2.13). Linearizing in H, we get $2\ddot H=-M^2H$, which does not have growing solutions for $M^2>0$.

Of course, the linear approximation (2.15) breaks down when δ becomes ~ 1. To study the nonlinear evolution of the model, we shall find approximate solutions of Eq. (2.13) in various regimes. Suppose that initially H is near H_0 and $\dot H$ is small, $\dot H \ll H_0^2$. If $H>H_0$, then H grows without bound. Such solutions are obviously unrelated to our universe, and we shall concentrate on the case $H<H_0$.

We shall first solve Eq. (2.13) assuming $H(t)$ to be slowly varying:

$$\dot H \ll H^2\ ,\quad \ddot H \ll H\dot H\ . \tag{2.16}$$

Then Eq. (2.13) takes the form

$$H^2-H_0^2=6(H_0^2/M^2)\dot H\ . \tag{2.17}$$

The solution of this equation is

$$\frac{H_0-H}{H_0+H}=\tfrac{1}{2}\delta_0\exp(M^2t/3H_0)$$

or

$$H = H_0 \tanh\left[\gamma - \frac{M^2 t}{6H_0}\right], \quad (2.18)$$

where $\gamma = \frac{1}{2}\ln(2/\delta_0)$ and δ_0 is the magnitude of $|H - H_0|/H_0$ at the initial moment, $t=0$. The range of probable values of δ_0 will be determined by quantum analysis in the following sections. [See Eq. (5.20).] From Eq. (2.18) we see that $H(t)$ changes on a time scale $\sim 6H_0/M^2$. Long inflation is obtained if $M^2 \ll 6{H_0}^2$. Below we shall assume that this is the case. In fact, we will see in Sec. II D that a much stronger constraint has to be imposed to avoid a conflict with observations.

To estimate the limits of validity of our approximate solution (2.18), we substitute it in Eq. (2.13) and find when the neglected terms become comparable to those we kept. This happens at

$$t \sim t_* = 6\gamma H_0/M^2 ,$$

when $H \sim M$. Hence, in the course of inflation, the expansion rate gradually changes from $\sim H_0$ to $\sim M \ll H_0$.

The scale factor $a(t)$ is found from Eq. (2.18) by a simple integration:

$$a(t) = {H_0}^{-1}[\cosh\gamma/\cosh(\gamma - M^2 t/6H_0)]^{6{H_0}^2/M^2} .$$

For $t_* - t \gtrsim 6H_0/M^2$ this gives $a(t) = H_0 \exp(H_0 t)$, and for $t_* - t \ll 6H_0/M^2$

$$a(t) = {H_0}^{-1}(\cosh\gamma)^{6{H_0}^2/M^2} \exp[-\tfrac{1}{12}M^2(t_* - t)^2] . \quad (2.19)$$

The expansion rate during this period is

$$H(t) = \tfrac{1}{6}M^2(t_* - t) . \quad (2.20)$$

To investigate the further evolution of the model, we note that for $H \ll H_0$ the term proportional to H^4 in Eq. (2.13) can be neglected:

$$2H\ddot{H} + 6H^2\dot{H} - \dot{H}^2 + M^2H^2 = 0 . \quad (2.21)$$

The "friction" term, $6H^2\dot{H}$, is also small for $H \ll M$. Neglecting this term and introducing a new variable $z = \dot{H}/H$, we bring Eq. (2.21) to the form

$$2\dot{z} + z^2 + M^2 = 0 . \quad (2.22)$$

The solution is $z = -M\tan(Mt/2)$ and

$$H = B\cos^2(Mt/2) , \quad (2.23)$$

where $B = \text{const}$.

The effect of the friction term in Eq. (2.21) is to gradually damp the oscillations of H described by Eq. (2.23). An approximate solution of Eq. (2.21) is

$$H = \frac{4}{3t}\cos^2\left[\frac{Mt}{2}\right]\left[1 - \frac{\sin Mt}{Mt}\right] + O(t^{-3}) . \quad (2.24)$$

Although $(Mt)^{-1}\sin Mt \ll 1$, this term has to be kept, since its contribution to the derivatives of H is not negligible. Two arbitrary integration constants are not shown explicitly in Eq. (2.24). They correspond to the choice of the origin of t and to the change of Mt to $(Mt+\alpha)$ in the trigonometric functions. The scale factor is given by[5]

$$a(t) = \text{const} \times t^{2/3}[1 + (2/3Mt)\sin Mt + O(t^{-2})] . \quad (2.25)$$

The expansion rate averaged over the oscillation period is

$$\bar{H} = 2/3t . \quad (2.26)$$

It corresponds to the expansion law of a matter-dominated universe, $\bar{a}(t) \propto t^{2/3}$. This behavior can be understood in the following way. The oscillations of the expansion rate in Eq. (2.23) can be thought of as coherent oscillations of a massive field describing scalar particles of mass M (Starobinsky calls them scalarons). The gravitational effect of such "particles" with zero momentum ($\nabla H = 0$) is similar to that of pressureless gas and leads to the expansion law $a \propto t^{2/3}$. It is interesting that the particles and the expansion of the universe are both described by the same variable, $a(t)$.

C. Thermalization

Rapid oscillations of the expansion rate result in particle production. This process can be thought of as a decay of scalarons into other particles. It is well known[20] that the particle-production rate vanishes for massless, conformally coupled fields in a Robertson-Walker universe. Therefore, we have to take into account deviations from conformal invariance. For simplicity, we shall consider a real scalar field ϕ of mass m and conformal parameter ξ,

$$\phi_{;\mu}{}^{;\mu} + (m^2 + \xi R)\phi = 0 , \quad (2.27)$$

assuming that $m \ll M$ and $|\xi - \frac{1}{6}| \ll 1$. (Note that for $m > M/2$ the particle production is exponentially suppressed.)

To estimate the rate of particle production, at some time t_0, we note that for $Mt_0 \gg 1$ the period of oscillations ($2\pi M^{-1}$) is much shorter than the average Hubble time ($3t_0/2$). For a time interval Δt such that $t_0 \gg \Delta t \gg M^{-1}$, we can neglect the effect of power-law expansion and consider only the oscillatory contribution to $a(t)$. With proper normalization,

$$a(t) \approx 1 + (2/3Mt_0)\sin Mt . \quad (2.28)$$

The oscillatory term in (2.28) is small and can be treated as a perturbation. A perturbative technique for calculating the particle production has been developed in Ref. 25.

Let us first consider the case of a massless, nonconformally coupled field: $m=0$, $\xi \neq \frac{1}{6}$. Zel'dovich and Starobinsky[25] have shown that in this case the rate of particle production per unit volume per unit time is

$$\dot{n} = \frac{1}{16\pi}(\xi - \tfrac{1}{6})^2 R^2 , \quad (2.29)$$

where

$$R \approx 6\ddot{a} = -(4M/t_0)\sin Mt \quad (2.30)$$

is the (linearized) scalar curvature. Averaging over the period of oscillation and using the fact that the particles are produced in pairs with energy $M/2$ per particle, we can write

$$\bar{\dot{\rho}} = \tfrac{1}{2}M\bar{\dot{n}} = \frac{M^3}{4\pi {t_0}^2}(\xi - \tfrac{1}{6})^2 , \quad (2.31)$$

where ρ is the energy density of the particles. (I have verified that a direct calculation of $\rho = \langle T_0{}^0 \rangle$ gives the same answer.)

The rate at which the energy of scalarons is dissipated is

$$\Gamma = \bar{\dot{\rho}}/\bar{\rho} = \tfrac{3}{2} G M^3 (\xi - \tfrac{1}{6})^2 , \tag{2.32}$$

where $\bar{\rho} = (6\pi G t_0{}^2)^{-1}$ is the mass density of scalarons. At $t \sim \Gamma^{-1}$ the oscillations of the scale factor (scalarons) are damped, the created particles thermalize, and the universe becomes radiation dominated at the temperature

$$T_{\text{th}} \sim (\Gamma m_P)^{1/2} \sim |\xi - \tfrac{1}{6}| M^{3/2} m_P{}^{-1/2} , \tag{2.33}$$

where $m_P = G^{-1/2}$ is the Planck mass. (For $T_{\text{th}} \lesssim 10^{16}$ GeV the thermalization time is much shorter than the Hubble time, t.)

It is shown in Appendix A that the particle-production rate for a field of mass $m \ll M$ can be obtained from that for a massless field by replacing

$$(\xi - \tfrac{1}{6}) M^2 \to (\xi - \tfrac{1}{6}) M^2 - \tfrac{1}{3} m^2 . \tag{2.34}$$

Hence, in the general case,

$$\Gamma = G \tilde{m}^4 / 6M , \tag{2.35}$$

$$T_{\text{th}} \sim |\tilde{m}^2| (M m_P)^{-1/2} , \tag{2.36}$$

where

$$\tilde{m}^2 = m^2 - 3(\xi - \tfrac{1}{6}) M^2 . \tag{2.37}$$

Although Eqs. (2.35) and (2.36) were obtained assuming that $m \ll M$, they should give a correct order of magnitude for $m \lesssim M$. Of course, different particles will have different values of $\tilde{m}^2$. To estimate Γ and T_{th}, one should use the largest value of $\tilde{m}^2$ satisfying $m < M/2$.

D. Generation of gravitational waves

A stringent constraint on the Starobinsky model can be derived from the observational limit on the amplitude of very-long-wavelength gravitational waves.[26] Starobinsky has shown[27] that quantum fluctuations of the gravitational field during the inflationary era can give rise to long-wavelength gravitational waves of appreciable amplitude at present. He found that the initial amplitude of the waves is $h \sim H_h / m_P$, where H_h is the expansion rate at the time when the comoving wavelength of the wave crossed the horizon during the inflationary period. The amplitude remains constant until much later, the wave reenters the horizon. The gravitational waves disturb the isotropy of the microwave background through the Sachs-Wolfe effect.[26] They produce a $\delta T/T$ of order their amplitude h. Since observations indicate a large-scale anisotropy of less than 10^{-4}, one can derive an upper bound on the expansion rate at the time when the relevant scales crossed the horizon:[26]

$$H_h \lesssim 5 \times 10^{-4} m_P . \tag{2.38}$$

The results of Refs. 26 and 27 do not directly apply to the Starobinsky model, since the equation describing the propagation of gravitational waves gets modified by the quadratic terms (2.3) in the Einstein equations. In the presence of weak gravitational waves, the metric can be written as

$$ds^2 = a^2(\eta)[d\eta^2 - (\delta_{ij} + h_{ij}) dx^i dx^j] , \tag{2.39}$$

where η is the conformal time which is related to t by

$$\eta = \int dt / a(t) . \tag{2.40}$$

The gravitational wave perturbation h_{ij} can be decomposed into modes of the form $h_{ij} = e_{ij} h_k(\eta) \exp(i\mathbf{k}\cdot\mathbf{x})$ where the tensor e_{ij} corresponds to one of the two independent polarizations. The functions $h_k(\eta)$ satisfy the equation[28]

$$\begin{aligned} h_k'' &\left[1 + \frac{R}{3M^2} + \frac{2R_0^0 - R}{3H_0{}^2} \right] \\ &+ h_k' \left[2\frac{a'}{a} \left[1 + \frac{R}{3M^2} - \frac{2R_0^0}{3H_0{}^2} \right] + \frac{R'}{3M^2} \right] \\ &+ k^2 h_k \left[1 + \frac{R}{3M^2} + \frac{R - 6R_0^0}{3H_0{}^2} \right] = 0 , \end{aligned} \tag{2.41}$$

where primes stand for derivatives with respect to η. For $M^2 \ll H_0{}^2$, we can neglect terms proportional to $H_0{}^{-2}$, and Eq. (2.41) takes the form

$$h_k'' + \left[2\frac{a'}{a} + \frac{b'}{b} \right] h_k' + k^2 h_k = 0 , \tag{2.42}$$

where

$$b = 1 + R/3M^2 . \tag{2.43}$$

The term b'/b is due to the one-loop corrections (2.8) and was not taken into account in Ref. 27.

The analysis of the spectrum of gravitational waves in our model can be reduced to that in Refs. 26 and 27 if we note that Eq. (2.42) can be thought of as a regular wave equation for gravitons, but in a modified metric with a scale factor $\tilde{a} = a\sqrt{b}$. For wavelengths much shorter than the horizon, $-k\eta \gg 1$, the graviton mode functions have the form

$$h_k(\eta) = (2G/k)^{1/2} (\pi a b^{1/2})^{-1} \exp(-ik\eta) . \tag{2.44}$$

The choice of pure positive-frequency modes corresponds to the absence of gravitons in the initial state. When the wavelength becomes much greater than the horizon, $-k\eta \ll 1$, $h_k(\eta)$ approaches a constant. The magnitude of this constant can be determined by solving Eq. (2.42) in the transition region, $-k\eta \sim 1$.

Since H and b are slowly varying functions of time, we can replace them in the transition region by their values at the time of horizon crossing, H_h and $b_h = 1 + 4H_h{}^2/M^2$. In other words, we approximate our nearly de Sitter space by an exact de Sitter space. The corresponding scale factor is $a(\eta) = -(H_h \eta)^{-1}$, and the solution of Eq. (2.42) which matches to Eq. (2.44) is

$$h_k(\eta) = -\left[\frac{2G}{k b_h} \right]^{1/2} \frac{H_h \eta}{\pi} \left[1 - \frac{i}{k\eta} \right] e^{-ik\eta} . \tag{2.45}$$

We see that at late times, $-k\eta \ll 1$, $h_k(\eta) = i(H_h/\pi)(2G/k^3 b_h)^{1/2}$. This is different from the results of Ref. 27 by a factor $b_h^{-1/2} = M/2H_h$. [In the regime of interest to us, $H_h \gg M$, and we can neglect 1 in Eq. (2.43) for b.]

Note that the resulting spectrum of gravitational waves is exactly flat,

$$|h_k| = (GM^2/2\pi^2)^{1/2} k^{-3/2} , \tag{2.46}$$

and is independent of H_0. A constraint on the parameter M of the model can be obtained by replacing $H_h \to M/2$ in Eq. (2.38):

$$M \lesssim 10^{-3} m_P \sim 10^{16} \text{ GeV} . \tag{2.47}$$

I have learned that Starobinsky has derived an even stronger constraint, $M \lesssim 10^{14}$ GeV, by requiring that the density fluctuations resulting from inflation are sufficiently small.[18]

Thus, to give a successful inflationary scenario, the Starobinsky model has to include two vastly different scales, $H_0 \sim 10^{19}$ GeV and $M \lesssim 10^{14}$ GeV. In terms of the coefficients k_1 and k_3 in Eq. (2.3) we must have $k_1 \gtrsim 10^9 k_3$. As we learned in Sec. II A, k_3 is fixed by the model, while k_1 is arbitrary, and so it is consistent to assume that $k_1 \gg k_3$. This assumption does not strike one as very natural, but, to be fair, one has to add that all known inflationary models require an unnatural adjustment of parameters of one sort or another.

With $M \lesssim 10^{14}$ GeV, the thermalization temperature is $T_{\text{th}} \lesssim 10^{12}$ GeV, which is well below the grand unification scale. The grand unification symmetry breaking occurs during inflation, and the superheavy magnetic monopoles are inflated away.

III. TUNNELING PROBABILITY

We now turn to the quantum cosmology of the Starobinsky model. A universe created by quantum tunneling should be closed, since otherwise the tunneling action is infinite, and from now on we specialize to the closed Robertson-Walker model with $\kappa = +1$.

The semiclassical tunneling probability can be calculated using the instanton method.[9,13] Omitting the pre-exponential factor, the probability is given by[29]

$$P \propto \exp(-|S_0|) , \tag{3.1}$$

where S_0 is the instanton action. The instanton is a solution of Euclidean field equations which can be obtained from the de Sitter solution (2.11a) by changing $t \to it$:

$$a(t) = H_0^{-1} \cos(H_0 t) . \tag{3.2}$$

This is the well-known de Sitter instanton[30] describing a four-sphere of radius H_0^{-1}.

I shall now demonstrate how one can calculate the tunneling action for a model described by the action of the general form (1.4),

$$S = \int L(R) d\Omega . \tag{3.3}$$

Here, $d\Omega = (-g)^{1/2} d^4x$ is the four-volume element and $L(R)$ is an arbitrary function of the scalar curvature. For a de Sitter instanton, the scalar curvature is constant,

$$R = R_0 = 12 H_0^2 , \tag{3.4}$$

and we can write

$$S(R) = L(R)\Omega_R = 384\pi^2 R^{-2} L(R) ,$$

where

$$\Omega_R = 384\pi^2 R^{-2} \tag{3.5}$$

is the volume of a four-dimensional sphere of scalar curvature R.

The curvature of the de Sitter instanton R_0 can be found by solving the field equations as we did in Sec. II. Alternatively, it can be found by minimizing the action (3.5):

$$R_0 L'(R_0) - 2L(R_0) = 0 . \tag{3.6}$$

Once R_0 is found, we can calculate $S(R_0)$ and use it to find the tunneling probability from Eq. (3.1).

The semiclassical approximation is justified if the tunneling action is large, $|S_0| \gg 1$. The curvature fluctuations in the newly born universe can be estimated from

$$|\delta S| \sim 1 . \tag{3.7}$$

Expanding $S(R)$ around $R = R_0$ we have

$$S(R) = S_0 + \tfrac{1}{2} S''(R_0)(\delta R)^2 ,$$

and Eq. (3.7) gives

$$\left[\frac{\delta R}{R_0}\right]^2 \sim \frac{2}{R_0^2 |S''(R_0)|} . \tag{3.8}$$

As an example, consider a simple model described by the Einstein action with a cosmological term representing the false vacuum energy:

$$L(R) = R/16\pi G - \rho_v . \tag{3.9}$$

Extremizing the action (3.5) we find $R_0 = 32\pi G \rho_v$, $S(R_0) = 3/8G^2\rho_v$, and[11,10]

$$P \propto \exp(-3/8G^2\rho_v) .$$

We now turn to the Starobinsky model. The one-loop effective action corresponding to Eqs. (2.1) and (2.8) is not known in a closed form for a general Robertson-Walker metric. As I mentioned in Sec. II A, such an action should be nonlocal. However, an approximate local action can be found for the case of the Starobinsky model with $M \ll H_0$. This action is given by Eq. (3.3) with (see Appendix B)

$$L(R) = \frac{1}{16\pi G}\left[R + \frac{R^2}{6M^2} + \frac{R^2}{R_0} \ln\frac{R}{R_0}\right] , \tag{3.10}$$

where $R_0 = 12 H_0^2$. From Eq. (3.5),

$$S(R) = \frac{24\pi}{G}\left[\frac{1}{R} + \frac{1}{6M^2} + \frac{1}{R_0} \ln\frac{R}{R_0}\right] . \tag{3.11}$$

This action has an extremum at $R = R_0$, as it should. The tunneling probability is given by

$$P \propto \exp\left[-\frac{4\pi}{GM^2}\right], \tag{3.12}$$

where I have used the fact that $H_0^2 \gg M^2$.

The curvature fluctuations can be estimated from Eq. (3.8):

$$(\delta R/R_0)^2 \sim GH_0^2/\pi \,. \tag{3.13}$$

A better estimate for $\delta R/R_0$ will be found in Sec. V by solving the Wheeler-DeWitt equation for the wave function of the universe.

Let us now consider how the tunneling probability (3.12) is affected by adding a false vacuum energy term to the action:

$$S=\frac{24\pi}{G}\left[\frac{1}{R}+\frac{1}{6M^2}+\frac{1}{R_0}\ln\frac{R}{R_0}-\frac{16\pi G\rho_v}{R^2}\right]. \tag{3.14}$$

This action has two extrema (for sufficiently small ρ_v):

$$R_\pm = 6H_0^2\left[1\pm\left[1-\frac{32\pi G\rho_v}{3H_0^2}\right]^{1/2}\right]. \tag{3.15}$$

R_+ and R_- correspond to a local minimum and a local maximum of S, respectively, and

$$S(R_+)\le S(R_-)\,. \tag{3.16}$$

If $\rho_v \ll G^{-2}$, then the curvatures of the two instantons are $R_+ \approx 12H_0^2$ and $R_- \approx 32\pi G\rho_v$. The first instanton describes a tunneling to the Starobinsky phase (slightly modified by the presence of ρ_v), while the second corresponds to a tunneling to a false-vacuum-dominated phase in which the quantum-gravitational corrections are small. It follows from Eq. (3.16) that the tunneling to the Starobinsky phase has a higher probability.

At $\rho_v = 3H_0^2/32\pi G \equiv \rho_c$, the two instantons coincide, and for $\rho_v > \rho_c$ the action (3.14) has no stationary points. The critical density ρ_c is comparable to the Planck energy density $\rho_c \sim G^{-2}$.

If there are several false vacuum states to choose from, which of them gives the highest tunneling probability? From Eq. (3.14) we have

$$\frac{\partial}{\partial\rho_v}S(R_+) = -\frac{384\pi^2}{R_+^2} < 0\,, \tag{3.17}$$

which means that greater vacuum energies give a smaller tunneling action. Hence, the universe prefers to tunnel to the Starobinsky phase with a false vacuum of the highest energy density satisfying $\rho_v < \rho_c$.

This analysis suggests the following interesting possibility. Suppose we have a grand unified theory with a Higgs field Φ and effective potential $V(\Phi)$. Suppose that $V(\Phi)$ is unbounded from above (which is usually the case). Then $V(\Phi)=\rho_c$ at some value of Φ, $\Phi=\Phi_c$. If Φ_c were a stationary point of V, $V'(\Phi_c)=0$, then the instanton with $R=6H_0^2$, $\Phi=\Phi_c$ would give the greatest tunneling probability. If $V'(\Phi_c)\neq 0$, then $\Phi=\Phi_c$ is not a solution of the field equations. However, $\Phi=\Phi_c$ can still be an *approximate* instanton solution if the slope of the potential $V(\Phi)$ is sufficiently small near Φ_c. It is possible that such approximate instantons give the highest tunneling probability. To obtain a definitive answer, one has to study the solutions of the Wheeler-DeWitt equation for models including both quantum-gravitational corrections and scalar fields. This problem will not be addressed in the present paper. In the following sections we shall concentrate on the "pure" Starobinsky model without classical matter fields.

IV. CANONICAL QUANTIZATION

A more complete description for the nucleation of the universe can be obtained by solving the "Schrödinger equation" for the wave function of the universe. In quantum gravity, the wave function $\Psi(g_{ij})$ is defined on a space of all possible three-geometries (superspace). In higher-derivative theories of gravity the number of independent variables is doubled. The standard choice of additional variables is K_{ij}, the extrinsic curvature of the three-geometry.[16] The role of the Schrödinger equation is played by the Wheeler-DeWitt equation[19]

$$\mathscr{H}\Psi = 0\,, \tag{4.1}$$

where $\mathscr{H}$ is the Hamiltonian.

In this paper we will not attempt to analyze inhomogeneous fluctuations of the metric and will restrict the three-geometry to be homogeneous, isotropic, and closed. This corresponds to reducing the infinite number of gravitational degrees of freedom in the superspace to just two variables. Such a reduced superspace is called minisuperspace. Minisuperspace quantization of higher derivative theories with a term proportional to R^2 in the gravitational Lagrangian has been discussed in Ref. 17. Here we shall consider a general class of models with action of the form (3.3):

$$S=\int L(R)(-g)^{1/2}d^4x = 2\pi^2\int L(R)a^3dt\,. \tag{4.2}$$

The scalar curvature for a closed Robertson-Walker metric is

$$R = 6a^{-2}(1+\dot a^2 + a\ddot a)\,. \tag{4.3}$$

Substituting (4.3) in (4.2), the action can be written as

$$S=\int \mathscr{L}(a,\dot a,\ddot a)dt\,. \tag{4.4}$$

Before we discuss the general case, let us first consider a simple example of minisuperspace quantization—that of Einstein's gravity with a cosmological term, Eq. (3.8). Substituting Eq. (4.3) in the action and integrating by parts to remove the second derivative of a, we obtain

$$S=\frac{3\pi}{4G}\int (1-\dot a^2 - H_v^2 a^2)a\,dt\,, \tag{4.5}$$

where $H_v = (8\pi G\rho_v/3)^{1/2}$. The momentum conjugate to the scale factor a is

$$P_a = \frac{\partial\mathscr{L}}{\partial\dot a} = -\frac{3\pi}{2G}a\dot a \tag{4.6}$$

and the Hamiltonian is

$$\mathscr{H} = P_a\dot a - \mathscr{L} = -\frac{G}{3\pi a}P_a^2 - \frac{3\pi}{4G}a(1-H_v^2a^2)\,. \tag{4.7}$$

The Wheeler-DeWitt equation is obtained by replacing $P_a \to -i\partial/\partial a$:

$$\left[a^{-p}\frac{\partial}{\partial a}a^{p}\frac{\partial}{\partial a}-\left[\frac{3\pi}{2G}\right]^2 a^2(1-H_v{}^2a^2)\right]\Psi(a)=0 . \quad (4.8)$$

Here, the parameter p represents the ambiguity in the ordering of factors a and $\partial/\partial a$. Fortunately, a variation of p affects only the pre-exponential factor of the semiclassical wave function. The actual value of p will be unimportant for our calculations and can be adjusted for convenience.

Setting $p=0$, Eq. (4.8) takes the form of a one-dimensional Schrödinger equation for a "particle" described by a coordinate $a(t)$, having zero energy and moving in a potential

$$U(a)=\left[\frac{3\pi}{2G}\right]^2 a^2(1-H_v{}^2a^2) . \quad (4.9)$$

"Nothing" $(a=0)$ is separated from the classically allowed range $(a \geq H_v{}^{-1})$ by a potential barrier. The WKB tunneling probability is

$$P \propto \exp\left[-2\int_0^{H_v^{-1}}[U(a)]^{1/2}da\right] . \quad (4.10)$$

This, of course, gives the same answer as the instanton approach, Eq. (3.9).

In the general case, second derivatives of a cannot be removed from Eq. (4.2) using integration by parts. The standard approach to canonical quantization in this case is to introduce, in addition to a, another variable (we shall choose it to be the scalar curvature, R) and to express the action in the form

$$S=\int \mathscr{L}(a,\dot{a},R,\dot{R})dt .$$

This corresponds to rewriting a fourth-order differential equation as a system of two second-order equations.

With the action written as in Eq. (4.2) we can treat Eq. (4.3) as a constraint. Equivalently, we can rewrite the action using a Lagrange multiplier:

$$S=2\pi^2\int \{L(R)a^3-\beta[R-6a^{-2}(1+\dot{a}+a\ddot{a})]\}dt . \quad (4.11)$$

β is determined by varying S with respect to R:

$$\beta=a^3L'(R) . \quad (4.12)$$

Substituting (4.12) back into Eq. (4.11) and integrating by parts to eliminate $\ddot{a}$, we obtain

$$S=2\pi^2\int [L(R)-RL'(R)+6L'(R)a^{-2}(1-\dot{a}^2) -6L''(R)a^{-1}\dot{a}\dot{R}]a^3dt . \quad (4.13)$$

This form of the action is already suitable for canonical quantization. However, it will be convenient to "diagonalize" the derivative part of the Lagrangian by introducing new variables instead of a and R. We define

$$q=H_0a(L'/L_0')^{1/2} , \quad (4.14a)$$

$$x=\tfrac{1}{2}\ln(L'/L_0') , \quad (4.14b)$$

where $L'=L'(R)$, $L_0'=L'(R_0)$, H_0 is related to R_0 by $R_0=12H_0{}^2$ and R_0 is the curvature of the self-consistent de Sitter solution which can be found from Eq. (3.6). Note that, for $R=R_0$, $q=H_0a$ and $x=0$. The transformation (4.14) becomes singular at points where $L'=0$. However, we will be interested in a range of x near $x=0$ $(R=R_0)$ where the change of variables (4.14) is justified.

From (4.13) and (4.14) we find

$$S=\int \mathscr{L}(q,\dot{q},x,\dot{x})dt ,$$
$$\mathscr{L}=2\pi^2H_0{}^{-3}[L-RL'+\tfrac{1}{2}R_0(L'^2/L_0')q^{-2} +6L'(\dot{x}^2-\dot{q}^2/q^2)](L'/L_0')^{-3/2}q^3 , \quad (4.15)$$

where R has to be expressed in terms of x from Eq. (4.14b). The canonical momenta and the Hamiltonian are found from

$$P_q=\partial\mathscr{L}/\partial\dot{q}=-BL'^{-1/2}q\dot{q} , \quad (4.16a)$$

$$P_x=\partial\mathscr{L}/\partial\dot{x}=BL'^{-1/2}q^3\dot{x} , \quad (4.16b)$$

$$\mathscr{H}=\dot{q}P_q+\dot{x}P_x-\mathscr{L} , \quad (4.16c)$$

where $B=24\pi^2H_0{}^{-3}L_0'^{3/2}$. Replacing $P_q \to -i\partial/\partial_q$, $P_x \to -i\partial/\partial x$ in the Hamiltonian, we obtain the Wheeler-DeWitt equation:

$$\left[\frac{\partial^2}{\partial q^2}-\frac{1}{q^2}\frac{\partial^2}{\partial x^2}-V(q,x)\right]\Psi(q,x)=0 , \quad (4.17)$$

$$V(q,x)=\frac{1}{\lambda^2}q^2\left[1+\frac{2L_0'}{R_0L'^2}(L-RL')q^2\right] . \quad (4.18)$$

Here, $\lambda^{-1}=288\pi^2L_0'/R_0=576\pi^2L_0/R_0{}^2$, where I have used Eq. (3.6). Now, making use of Eq. (3.5) we can write

$$\lambda=2/3S_0 , \quad (4.19)$$

where S_0 is the instanton action. The WKB approximation which will be used in the next section is based on an expansion in powers of λ. It is justified if λ is small or, equivalently, if the instanton action is large, $S_0 \gg 1$.

For simplicity I have set the factor-ordering parameters equal to zero in Eq. (4.17). Nonzero values of these parameters introduce only unimportant modifications in the pre-exponential factor of the semiclassical wave function.

The formalism developed thus far applies to models with arbitrary $L(R)$. Now we specialize to the Starobinsky model with L given by Eq. (3.10). Using the condition $M^2 \ll H_0{}^2$, Eqs. (4.14b) and (4.18) take the form

$$x=\tfrac{1}{2}\ln(R/R_0) , \quad (4.20)$$

$$V(q,x)=\lambda^{-2}q^2[1-q^2+\mu^2(x)q^2] , \quad (4.21)$$

where

$$\mu^2(x)=\frac{M^2}{2H_0{}^2}(2x+e^{-2x}-1) , \quad (4.22)$$

$$\lambda=GM^2/6\pi . \quad (4.23)$$

In Eq. (4.22) we neglected terms $\sim(M/H_0)^4$.

For $x=0$, the potential for q is similar to that in Eq. (4.9). The classically forbidden range is from $q=0$ to

$q=1$. At nonzero values of x, the potential barrier ends ($V=0$) at $q\approx 1+\frac{1}{2}\mu^2(x)$. Since $x=0$ is the minimum of $\mu^2(x)$, it corresponds to the smallest width of the barrier, and one can expect the wave function to be peaked near $x=0$.

Equations (4.21) and (4.22) apply for values of x such that $\mu^2(x)\ll 1$:

$$-\ln(H_0/M)<x\ll H_0^2/M^2 . \tag{4.24}$$

Beyond this range, the approximations made in deriving these equations are no longer justified (one has to include higher-order terms in M^2/H_0^2).

V. THE WAVE FUNCTION OF THE UNIVERSE

In this section we shall solve Eqs. (4.17) and (4.21) and find the wave function describing a universe tunneling to the Starobinsky phase.

To solve the Wheeler-DeWitt equation (4.17), we have to specify the boundary condition at $q=0$. We shall require that $|\Psi(q=0)|<\infty$. It is clear from Eq. (4.17) that this condition can be satisfied only if $\partial^2\Psi/\partial x^2$ vanishes at $q=0$: $\partial\Psi/\partial x(0,x)=\text{const}$. Since $\Psi(0,x)$ can be bounded for all values of x only if this constant is zero, we require

$$\frac{\partial\Psi}{\partial x}(0,x)=0 . \tag{5.1}$$

This boundary condition does not fix $\Psi(q,x)$ uniquely. Below we shall impose an additional requirement which will single out the tunneling solution of Eq. (4.17).

For $q\ll 1$ we can neglect the terms proportional to q^4 in $V(q,x)$:

$$\left[\frac{\partial^2}{\partial q^2}-\frac{1}{q^2}\frac{\partial^2}{\partial x^2}-\frac{1}{\lambda^2}q^2\right]\Psi=0 . \tag{5.2}$$

Two linearly independent solutions of (5.2) satisfying the boundary condition (5.1) are $q^{1/2}I_{1/4}(q^2/2\lambda)$ and $q^{1/2}K_{1/4}(q^2/2\lambda)$, where $I_\nu(z)$ and $K_\nu(z)$ are modified Bessel functions. A tunneling solution should exponentially decrease in the classically forbidden range. Hence, we choose

$$\Psi=q^{1/2}K_{1/4}(q^2/2\lambda) , \quad q\ll 1 . \tag{5.3}$$

The asymptotic form of (5.3) for $q\gg\lambda^{1/2}$ is

$$\Psi=(\pi\lambda/q)^{1/2}\exp(-q^2/2\lambda) , \quad \lambda^{1/2}\ll q\ll 1 . \tag{5.4}$$

In most of the region $0<q<1$ we can use the WKB approximation. The condition of its applicability is $|V(q,x)|\gg 1$ or

$$q\gg\lambda , \quad 1-q\gg\lambda^2 . \tag{5.5}$$

The WKB approximation is based on expansion in powers of λ. Writing the Wheeler-DeWitt equation as

$$(\nabla^2-V)\Psi=0 , \tag{5.6}$$

we make the WKB ansatz

$$\Psi=A\,e^{-S} \tag{5.7}$$

and assume that S is of order λ^{-1}. Then, comparing terms with equal powers of λ, we obtain

$$\nabla S\cdot\nabla S=V , \tag{5.8}$$

$$2\nabla A\cdot\nabla S+A\nabla^2 S=0 . \tag{5.9}$$

Here, $\nabla^2=\partial_q{}^2-q^{-2}\partial_x{}^2$,

$$\nabla S\cdot\nabla S=(\partial S/\partial q)^2-q^{-2}(\partial S/\partial x)^2 ,$$

etc., and I have used the fact that $V\propto\lambda^{-2}$.

The potential (4.21) has two small parameters, λ and M^2/H_0^2. Keeping the term proportional to M^2/H_0^2 in (5.8) is justified only if $M^2/H_0^2\gg\lambda$ or

$$GH_0^2/6\pi\ll 1 , \tag{5.10}$$

where I have used Eq. (4.23). With the aid of Eqs. (2.6) and (2.7) we find

$$GH_0^2/6\pi=(48\pi^2k_3)^{-1}\lesssim N_1^{-1} ,$$

where N_1 is the number of gauge fields in the model. The smallest grand unification group, SU(5), has $N_1=24$, and thus the condition (5.10) is satisfied for all grand unified models.

From the facts that (i) $\Psi(q,x)$ is independent of x for small q and (ii) the x-dependent terms in $V(q,x)$ are multiplied by the small parameter M^2/H_0^2 it follows that $\partial\Psi/\partial x$ is also proportional to M^2/H_0^2. Neglecting terms $\sim(M/H_0)^4$ we can drop the term proportional to $(\partial S/\partial x)^2$ in Eq. (5.8):

$$\frac{\partial S}{\partial q}=\sqrt{V}=\lambda^{-1}q\,[1-q^2+\mu^2(x)q^2]^{1/2} . \tag{5.11}$$

(I have chosen the positive sign of the square root which corresponds to the tunneling solution.) Integrating Eq. (5.11) we obtain

$$S=\frac{1}{3\lambda}[1+\mu^2(x)]\{1-[1-q^2+\mu^2(x)q^2]^{3/2}\} , \tag{5.12}$$

where the arbitrary additive function of x has been chosen so that $S\to q^2/2\lambda$ for $q\to 0$ with accuracy up to terms $\sim(M/H_0)^4$, and so Eqs. (5.4) and (5.12) agree in the region where both approximations apply.

We shall find the amplitude of the wave function to the lowest order in M^2/H_0^2. Neglecting terms $\sim M^2/H_0^2$ in Eq. (5.9), we can rewrite it as

$$\frac{\partial}{\partial q}(A^2\sqrt{V_0})=0, \tag{5.13}$$

where

$$V_0(q)=\lambda^{-2}q^2(1-q^2) . \tag{5.14}$$

The solution of (5.13) is

$$A=\pi^{1/2}V_0{}^{-1/4} , \tag{5.15}$$

where the constant factor is determined from Eq. (5.4).

The WKB approximation breaks down near the turning point, $q=q_x\approx 1+\frac{1}{2}\mu^2(x)$. The wave function in the classically allowed range $q>q_x$, can be found by analytically continuing around the turning point in the complex q plane:[31]

$$\Psi=\pi^{1/2}|V_0(q)|^{-1/4}\exp\left[\frac{-1}{3\lambda}[1+\mu^2(x)]\{1+i[q^2-1-\mu^2(x)q^2]^{3/2}\}-\frac{i\pi}{4}\right]. \tag{5.16}$$

We can now discuss the curvature fluctuations in the newly born universe. From Eq. (5.16), the probability of nucleation with a certain value of x is proportional to

$$|\Psi|^2\propto\exp[-2\mu^2(x)/3\lambda]. \tag{5.17}$$

For $x\ll 1$, expanding $\mu^2(x)$ in powers of x, we obtain

$$|\Psi|^2\propto(-4\pi x^2/GH_0^2). \tag{5.18}$$

Now it follows from Eq. (4.20) that $x=\delta R/2R_0$, where $\delta R=R-R_0$ is the curvature fluctuation. Hence, we can write

$$\langle(\delta R/R_0)^2\rangle=4\langle x^2\rangle=GH_0^2/2\pi, \tag{5.19}$$

in a qualitative agreement with Eq. (3.7). A typical fluctuation in the expansion rate is

$$\delta_0\equiv\frac{\delta H}{H_0}=\frac{1}{2}\frac{\delta R}{R_0}\sim\left[\frac{GH_0^2}{8\pi}\right]^{1/2}. \tag{5.20}$$

As we saw in Sec. II, the magnitude of δ_0 determines the duration of the inflationary phase in the Starobinsky model,

$$t_*\sim(3H_0/M^2)\ln(2/\delta_0). \tag{5.21}$$

With H_0 corresponding to the minimal SU(5) model ($H_0\sim 0.7m_P$), Eqs. (5.20) and (5.21) give $\delta_0\sim 0.14$, $t_*\sim 8H_0/M^2\gtrsim 4\times 10^{10}H_0^{-1}$, where I have used Eq. (2.41). This value of t_* is more than sufficient to solve the horizon and flatness problems.

ACKNOWLEDGMENTS

I am grateful to Sidney Coleman and Gordon Smith for discussions. This work was supported in part by the National Science Foundation under Grants Nos. PHY83-51860, PHY82-06202, PHY82-15249, and by contributions from General Electric Company, Dennison Manfacturing Company, and Massachusetts Electric Company.

APPENDIX A: MASSIVE PARTICLE PRODUCTION

In this appendix it will be shown that Eqs. (2.31)—(2.33) can be generalized to the case of nonzero mass by the replacement (2.34). It will be convenient to use the conformal time η instead of the comoving time t,

$$\eta=\int dt/a(t)\approx t+(2/3M^2t_0)\cos Mt. \tag{A1}$$

Then the metric takes the form

$$ds^2=a^2(\eta)(d\eta^2-dx_1^2-dx_2^2-dx_3^2). \tag{A2}$$

In the linear approximation, the difference between η and t is negligible, and we shall use t in the rest of this appendix.

The field ϕ can be expanded in terms of creation and annihilation operators as

$$\phi(x)=\int d^3k[a_ku_k(x)+a_x^\dagger u_k^*(x)], \tag{A3}$$

where

$$u_k(x)=(2\pi)^{-3/2}a^{-1}(t)e^{i\mathbf{k}\cdot\mathbf{x}}\chi_k(t) \tag{A4}$$

and the functions $\chi_k(t)$ satisfy the equation[20]

$$\ddot{\chi}_k+k^2\chi_k+[m^2+(\xi-\tfrac{1}{6})R]a^2\chi_k=0. \tag{A5}$$

Linearizing this equation and using Eqs. (2.28) and (2.30) we obtain

$$\ddot{\chi}_k+\omega_k^2\chi_k-(4/3Mt_0)\tilde{m}^2\sin(Mt)\chi_k=0, \tag{A6}$$

where $\omega_k=(k^2+m^2)^{1/2}$ and $\tilde{m}^2$ is given by Eq. (2.37).

The main contribution to the particle production comes from the modes with $k\sim M/2\gg m$. We can, therefore, replace ω_k^2 by k^2 in Eq. (A6). Then the mass m and the factor $(\xi-\frac{1}{6})$ appear in Eq. (A6) only in the combination (2.37), and it is clear that all results for a massive field can be obtained from those for a massless field by the replacement (2.34).

We note also that, as pointed out by Starobinsky,[5] the production of gravitons by the oscillating metric (2.28) is suppressed. This is easily seen from Eq. (2.42) if we note that $2\dot{a}/a+\dot{R}/3M^2\approx 0$ for $a(t)$ given by Eq. (2.28).

APPENDIX B: ACTION FOR STAROBINSKY MODEL

Although the one-loop gravitational action is not known for a general Robertson-Walker background, it has been calculated in some special cases. For a de Sitter space it is[32]

$$S=\frac{1}{16\pi}\int\left[\frac{R}{G}+\alpha R^2+\beta R^2\ln\frac{R}{\mu^2}\right](-g)^{1/2}d^4x. \tag{B1}$$

The coefficient α can be changed by varying the renormalization scale μ, while the coefficient β is related to the trace anomaly. Here it will be shown that (B1) is an approximate action in a more general case when the curvature is not assumed to be constant, but it is assumed that

$$\alpha\gg\beta. \tag{B2}$$

We will see that α and β are related to H_0 and M_0 of Eq. (2.8) with $\alpha/\beta\sim H_0^2/M^2$, and thus condition (B2) is satisfied for the Starobinsky model. If (B2) is satisfied, the value of μ in Eq. (B1) becomes unimportant, and we shall choose for definiteness $\mu^2=R_0=12H_0^2$.

The field equations obtained by varying the general action (3.3) are[33]

$$\frac{\partial L}{\partial R}R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}L+(\nabla_\mu\nabla_\nu-g_{\mu\nu}\nabla_\sigma\nabla^\sigma)\frac{\partial L}{\partial R}=0. \tag{B3}$$

Substituting (B1) in (B3) we obtain Eq. (2.1) with

$$8\pi\langle T_{\mu\nu}\rangle = 2\left[\alpha+\beta\ln\frac{R}{R_0}\right]R(R_{\mu\nu}-\tfrac{1}{4}g_{\mu\nu}R)$$
$$+\beta RR_{\mu\nu}+(\nabla_\mu\nabla_\nu-g_{\mu\nu}\nabla_\sigma\nabla^\sigma)$$
$$\times\left[2\alpha R+2\beta R\ln\frac{R}{R_0}+\beta R\right]. \tag{B4}$$

Equation (B2) allows us to neglect β compared to α. However, we have to keep the term $\beta RR_{\mu\nu}$, since this is the only term surviving in a de Sitter space. For a de Sitter space,

$$R_{\mu\nu}=\tfrac{1}{4}g_{\mu\nu}R\ ,\quad \nabla_\mu R=0\ , \tag{B5}$$

and all terms proportional to α vanish. Hence, we can rewrite Eq. (B4) as

$$8\pi\langle T_{\mu\nu}\rangle=\alpha\,{}^{(1)}H_{\mu\nu}+\beta RR_{\mu\nu}\ , \tag{B6}$$

where I have used Eq. (2.4a) for ${}^{(1)}H_{\mu\nu}$. From Eqs. (2.4b) and (B5) we find that in a de Sitter space

$${}^{(3)}H_{\mu\nu}=\tfrac{1}{48}g_{\mu\nu}R^2=\tfrac{1}{12}RR_{\mu\nu}\ , \tag{B7}$$

Since the term proportional to β in Eq. (B6) is important only when the metric is close to that of a de Sitter space, we can use Eq. (B7) and write

$$8\pi\langle T_{\mu\nu}\rangle=\alpha\,{}^{(1)}H_{\mu\nu}+12\beta\,{}^{(3)}H_{\mu\nu}\ . \tag{B8}$$

Comparing (B8) with Eq. (2.8), we find

$$\alpha=\frac{1}{6GM^2}\ ,\quad \beta=\frac{1}{12GH_0{}^2}=\frac{1}{GR_0}\ . \tag{B9}$$

Some properties of cosmological models derived from the action of the form (B1) have been discussed in Ref. 33.

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

[1]E. P. Tryon, Nature (London) **246**, 396 (1973).
[2]P. I. Fomin, Dok. Akad. Nauk. SSSR **9A**, 831 (1975).
[3]A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[4]A. D. Linde, Rep. Prog. Phys. **47**, 925 (1984).
[5]A. A. Starobinsky, Phys. Lett. **91B**, 99 (1980).
[6]In fact, Starobinsky's paper appeared before Guth (Ref. 3) suggested the standard version of inflation, although it was Guth who fully explained the advantages of inflationary scenarios.
[7]I use the system of units in which $\hbar=c=1$; the signature of the metric is $(+,-,-,-)$; $R^\mu{}_{\nu\sigma\tau}=\partial_\tau\Gamma^\mu_{\nu\sigma}-\cdots$. Note that the sign of the Riemann tensor is opposite to that chosen in the Starobinsky paper (Ref. 5).
[8]L. P. Grishchuk and Y. B. Zel'dovich, in *Quantum Structure of Space and Time,* edited by M. Duff and C. Isham (Cambridge University Press, Cambridge, 1982).
[9]A. Vilenkin, Phys. Lett. **117B**, 25 (1982); Phys. Rev. D **27**, 2848 (1983).
[10]A. Vilenkin, Phys. Rev. D **30**, 509 (1984).
[11]A. D. Linde, Lett. Nuovo Cimento **39**, 401 (1984).
[12]Y. B. Zel'dovich and A. A. Starobinsky, Pis'ma Astron. Zh. **10**, 323 (1984).
[13]A. Vilenkin, Nucl. Phys. **B252**, 141 (1985).
[14]For alternative approaches to quantum cosmology, see Ref. 15.
[15]J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983); S. W. Hawking, Nucl. Phys. **B239**, 257 (1984); D. N. Page, Pennsylvania State University report, 1984 (unpublished); T. Banks, W. Fischler, and L. Susskind, Report No. SLAC-PUB-3367, 1984 (unpublished); W. Fischler, B. Ratra, and L. Susskind, Stanford University Report No. ITP-775, 1984 (unpublished).
[16]D. Boulware, in *Quantum Theory of Gravity,* edited by S. M. Christensen (Adam Hilger, Bristol, 1983).
[17]S. W. Hawking and J. C. Luttrell, Nucl. Phys. **B247**, 250 (1984); G. Horowitz, Phys. Rev. D **31**, 1169 (1985).
[18]A. A. Starobinsky, Pis'ma Astron. Zh. **9**, 579 (1983).
[19]J. A. Wheeler, in *Battelle Rencontres,* edited by C. DeWitt and J. A. Wheeler (Benjamin, New York, 1968); B. S. DeWitt, Phys. Rev. **160**, 1113 (1967).
[20]N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge University Press, Cambridge, 1982).
[21]In the general case, $\langle T^\nu_\nu\rangle$ has an extra term proportional to the Weyl tensor squared. This term vanishes in conformally flat space times.
[22]L. Parker and D. J. Toms, Phys. Rev. D **29**, 1584 (1984).
[23]J. S. Dowker and R. Critchley, Phys. Rev. D **13**, 3224 (1976).
[24]Cosmological models of this sort including classical matter or radiation have been discussed in M. V. Fischetti, J. B. Hartle, and B. L. Hu, Phys. Rev. D **20**, 1757 (1979); B. L. Hu, in *Recent Developments in General Relativity,* edited by R. Ruffini (North-Holland, Amsterdam, 1980); P. Anderson, Phys. Rev. D **28**, 271 (1983).
[25]Y. B. Zel'dovich and A. A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. **26**, 373 (1977) [JETP Lett. **26**, 252 (1977)]; D. N. Birrell and P. C. W. Davies, J. Phys. A **13**, 2109 (1980).
[26]V. Rubakov, M. Sazhin, and A. Veryashin, Phys. Lett. **115B**, 189 (1982); L. F. Abbott, and M. B. Wise, Nucl. Phys. **B244**, 541 (1984).
[27]A. A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. **30**, 719 (1979) [JETP Lett. **30**, 682 (1979)].
[28]A. A. Starobinsky, Pis'ma Zh. Eksp. Teor. Fiz. **34**, 460 (1981) [JETP Lett. **34**, 438 (1981)].
[29]M. B. Voloshin, I. Yu. Kobzarev, and L. B. Okun, Yad. Fiz. **20**, 1229 (1974) [Sov. J. Nucl. Phys. **19**, 628 (1974)]; S. Coleman, Phys. Rev. D **15**, 2929 (1977); S. Coleman, in *The Whys of Subnuclear Physics,* edited by A. Zichichi (Plenum, New York, 1979).
[30]G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[31]C. M. Bender and T. T. Wu, Phys. Rev. D **7**, 1620 (1973); L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, Oxford, 1965).
[32]G. M. Shore, Ann. Phys. (N.Y.) **128**, 376 (1980).
[33]V. T. Gurovich and A. A. Starobinsky, Zh. Eksp. Teor. Fiz. **77**, 1683 (1979) [Sov. Phys. JETP **50**, 844 (1980)].