Quantum field theory methods and inflationary universe models

Robert H. Brandenberger

Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

This paper reviews the theory of inflationary universe models, giving particular emphasis to the question of origin and growth of energy-density fluctuations in these new cosmologies. The first four sections constitute a pedagogical introduction to some of the important quantum field theory methods used in inflationary universe scenarios: calculation of the effective potential, finite-temperature quantum field theory, analysis of the decay of a metastable quantum state, and free field theory in curved space-time.

CONTENTS

I. INTRODUCTION

The discovery by Guth (1981) that inflationary universe models elegantly solve some important cosmological problems of the standard "big bang" scenario has sparked a lot of interest in these new models among both physicists and cosmologists.

The crucial observation was that, if matter is described in terms of a quantum field with nonvanishing potential, there is a natural way to generate an effective cosmological constant in the early universe. This cosmological constant leads to a phase of exponential growth of the universe (hence the word "inflationary"), which in turn explains the homogeneity and flatness of the universe.

In the new models, gravity is described classically by a Friedmann-Robertson-Walker (FRW) metric, while matter is treated quantum mechanically, using the qualitative behavior of quantum field theories in flat space-time. We consider grand unified theories (see, for example, Langacker, 1981, for a detailed review) with gauge symmetry spontaneously broken at zero temperature by a nonvanishing expectation value of a scalar Higgs field. At high temperatures the symmetry is restored, but at the cost of generating a large cosmological constant. As the universe cools down it will therefore go through one or more phase transitions.

The explanation of the flatness and horizon problems (Guth, 1981, and Sec. VI) is not the first striking success of an approach combining particle physics and cosmology. Several years ago it was realized that grand unified theories can explain the observed nonvanishing baryon-to-entropy ratio of the universe. The key point is to consider baryon number-violating processes that occur out of equilibrium. References on this topic include Sakharov (1967), Yoshimura (1978), Ignatiev *et al.* (1978), Dimopoulos and Susskind (1978), Toussaint *et al.* (1979), Weinberg (1979), and the review article by Dolgov and Zeldovich (1981).

The objective of this paper is twofold. We want to give an introduction to and survey of inflationary universe models. We place particular emphasis on the generation and growth of energy-density fluctuations in inflationary cosmologies. To our knowledge, this is the first review of fluctuations. Inflationary universe models have been reviewed by Linde (1984) and in several conference proceedings (e.g., Guth, 1982; Linde, 1982e; Albrecht, 1984;

Guth, 1984) and popular articles (Barrow and Turner, 1982; Guth and Steinhardt, 1984). The second objective is to present a pedagogical introduction to quantum field theory methods used in the recent literature on particle physics and cosmology, and thus to bridge the gap between a graduate-level quantum field theory course and current research. We begin with a nontechnical survey of the main results in the first four sections and will mention how these results are used in inflationary universe models.

Why should matter be described in terms of quantum fields? In the early universe, matter is highly compressed and very hot. For times earlier than 10^{-6} sec after the big bang, the thermal energy of particles exceeds 1 GeV. From accelerator experiments we know that at these energies elementary particle interactions are correctly described by quantum field theory. Therefore, it seems mandatory to describe matter in terms of quantum fields in the early universe. It is most certainly incorrect to use an ideal-gas approximation for matter at these high temperatures.

In quantum field theory there is a natural way to obtain a nonvanishing cosmological constant. Consider a theory with a large positive potential $V(\varphi)$ for a given range of φ values. Consider a quantum state homogeneous in space and stationary, concentrated at φ values in the above range. Then, as discussed in Sec. VI.A, the contribution of the potential energy to the energy-momentum tensor dominates and acts like a cosmological constant in the Einstein equations.

It is therefore important to be able to determine the state of the quantum field theory at any given time in the evolution of the universe. The theory of the effective potential, discussed in detail in Sec. II, is an important tool in determining the quantum state. Loosely speaking, the effective potential includes the quantum corrections to the classical potential that appears in the Lagrangian of the field theory. To be more specific, consider a Lagrangian containing a scalar field φ with potential $V(\varphi)$ coupled to other fields A_μ. $V(\varphi)$ is the energy of a homogeneous stationary state with $\varphi(x)=\varphi$ and $A_\mu=0$. The main reason for our interest in the effective potential $V_{\text{eff}}(\bar{\varphi})$ in the context of inflationary universe models is the following: $V_{\text{eff}}(\bar{\varphi})$ is the minimum expectation value of the energy density in the class of all homogeneous stationary states with expectation value $\bar{\varphi}$. We derive this result in Sec. II.B. Provided we know that the state of the universe is homogeneous, we can use the effective potential to determine it. Since a stationary state will minimize the energy density, the state of the universe will have its expectation value of the operator Φ minimize $V_{\text{eff}}(\bar{\varphi})$. Summarizing the above in less precise language: the value of $\bar{\varphi}$ in the quantum state at a fixed time will minimize $V_{\text{eff}}(\bar{\varphi})$ at that time.

The reason that V_{eff} and V differ is obvious: in a quantum theory self-interactions of φ and interactions with the other fields A_μ will influence the energy of the state and cause it to differ from the classical value $V(\varphi)$.

In Sec. VI the effective potential will be used in a second way, namely to determine the equation of motion of the expectation value $\bar{\varphi}$ of Φ:

$$\frac{\partial^2\bar{\varphi}}{\partial t^2}-\nabla^2\bar{\varphi}=-\frac{\partial}{\partial\bar{\varphi}}V_{\text{eff}}(\bar{\varphi}) \ . \tag{1.1}$$

As explained in Sec. II.B, Eq. (1.1) is the leading contribution in an expansion about a homogeneous and stationary state. We can obtain Eq. (1.1) from the equation of motion of the classical field $\varphi(x)$ in the absence of interactions with other fields,

$$\frac{\partial^2\varphi}{\partial t^2}-\nabla^2\varphi=\frac{\partial}{\partial\varphi}V(\varphi) \ , \tag{1.2}$$

by replacing the classical potential $V(\varphi)$ with the effective potential $V_{\text{eff}}(\bar{\varphi})$. The difference is due to the fact that the evolution of $\bar{\varphi}$ is influenced by self-interactions of the φ field and by interactions with other fields.

In Sec. II.A we give the mathematical definition of the effective potential and of the effective action, the quantity of interest when considering nonhomogeneous or nonstationary processes. In Sec. II.B we derive the general properties of the effective potential we discussed above. Sections II.C and II.D are devoted to computing an approximate answer for $V_{\text{eff}}(\varphi)$, the one-loop approximation. In Sec. II.C we discuss the computation for the simplest "toy model" field theories; in Sec. II.D we extend the analysis to realistic particle physics models, unified gauge field theories.

The main reason for describing matter in terms of unified gauge field theories comes from particle physics. Since we expect that all the particles observed at high energies in particle accelerators should have been copiously produced in the early universe, we must consider the model that best describes the interactions of observed elementary particles. A second observation renders grand unified models very attractive for cosmology: today the cosmological constant Λ vanishes ($\Lambda/m_{\text{pl}}^2<10^{-120}$; see, for example, Hawking, 1984). In order for the cosmological constant to have been large in the early universe, a phase transition from $\langle V(\varphi)\rangle\neq 0$ to $\langle V(\varphi)\rangle=0$ must have taken place ($\langle\ \rangle$ denotes the quantum expectation value). Grand unified theories already have phase transitions built in.

Before describing these phase transitions, we must step back and explain the time dependence of the effective potential. From the mathematical definition in Sec. II.A it follows that the effective action is the free energy of the quantum system, and the effective potential therefore the free-energy density, provided again we restrict our attention to homogeneous states. In statistical mechanics it is well known that the free energy is temperature dependent. Thus the effective potential will depend on temperature as well. In Sec. III we compute the finite-temperature corrections to the effective potential. We find a temperature-dependent mass term proportional to φ^2T^2 which leads to symmetry restoration at high temperatures. At high temperature $\varphi=0$ will be the minimum of the effective potential, even if it was not a minimum at zero temperature. The high- and low-temperature ground states will then be different.

Most grand unified field theories are of the same structure as the models discussed in Sec. II.D: one or more scalar fields (Higgs fields) are coupled to gauge fields and Fermi fields. The Lagrangian has a large symmetry group. The group is much larger than the symmetry group observed at low energies and, for example, explicitly prohibits mass terms for fermions. Since the fermions we observe are massive, they must obtain their masses through symmetry breaking; even though the Lagrangian has a given symmetry, the theory will not be symmetric if the ground state breaks the symmetry. In the Higgs mechanism (Higgs, 1964a,1964b; Englert and Brout, 1964; Guralnik, Hagen, and Kibble, 1964) the symmetry is broken by a nonvanishing ground-state expectation value of φ. The condition for this to happen is that the effective potential take on its minimum at a nontrivial value of φ. In such grand unified models there is a natural phase transition: due to the finite-temperature corrections to the effective potential (discussed above), the high-temperature ground state has a vanishing expectation value of φ. Gauge symmetry is unbroken. Below a critical temperature T_c, which is, in general, of the order of the scale of symmetry breaking, a new absolute minimum of $V_{\text{eff}}(\varphi)$ appears. As the universe cools down in a standard big bang scenario, a phase transition must take place below T_c (see Sec. III.C). Thereafter the quantum ground state is localized at the absolute minimum $\varphi=\sigma$ of $V_{\text{eff}}(\varphi)$. Since we must fine-tune the cosmological constant to be zero today, i.e., $V_{\text{eff}}(\sigma)=0$, it must have been greater than zero in the high-temperature ground state, i.e., $V_{\text{eff}}(0)>0$. Thus grand unified theories with symmetry breaking via the Higgs mechanism automatically lead to phase transitions which can lead to an inflationary universe.

Any phase transition is described in terms of an order parameter, a parameter that distinguishes the various phases. In our case the order parameter is $\bar{\varphi}$, the expectation value of φ. We consider two types of phase transitions. In transitions of the first type, the order parameter jumps discontinuously in time from its value in the first phase to that in its second. In transitions of the second type the order parameter changes continuously. The former were used in "old" inflationary universe models (see Sec. VI.B). They arise when $\varphi=0$ is a local minimum of the zero-temperature effective potential (see Fig. 19 below). In these models the phase transition proceeds by quantum tunneling: the expectation value $\bar{\varphi}$ tunnels through the potential barrier. A bubble of the new phase nucleates in a surrounding sea of the old phase.

Phase transitions via bubble nucleation in quantum field theory are analyzed in Sec. IV. We first show that the decay rate of an unstable state is given by the imaginary part of its energy. Most of the section is devoted to the computation of the decay rate. For notational simplicity we first consider a one-dimensional quantum mechanics problem. We express the energy of the unstable state in terms of a functional integral and then use semiclassical techniques to evaluate the integral and pick out the imaginary part. In Sec. IV.E we show explicitly that, in the semiclassical approximation, the state after the decay event corresponds to a bubble of the stable phase with vanishing cosmological constant expanding at the speed of light in a surrounding sea of the unstable phase. This process is often called bubble nucleation.

Transitions of the second type occur when $\varphi=0$ is a local maximum of the zero-temperature effective potential (see Fig. 20 below). In this case the order parameter will simply evolve according to Eq. (1.1). This scenario arises in new inflationary universe models (see Sec. VI.B).

All the tools discussed so far are Minkowski space-time methods. In curved space-time there are new effects. We discuss some of them in Sec. V. The main point is that there is no unique vacuum state, even for a free scalar field. The origin of this ambiguity is the fact that in general relativity there is no distinguished coordinate frame. The most famous consequence of this ambiguity is Hawking radiation: an observer with a particle detector will at late times in certain space-times observe a nonvanishing thermal flux of particles in a state which at an initial time was set up to be empty of particles (Hawking, 1974,1975). In Secs. V.A and V.B we discuss Hawking radiation in the de Sitter phase of a cosmological model. A further consequence, which may be of even greater importance for inflationary universe scenarios, is the nonvanishing expectation value of the energy-momentum tensor $T_{\mu\nu}$, the quantity that is coupled to gravity when describing matter in terms of quantum fields. This is discussed in Sec. V.D. Finally, the results for Green's functions of a free scalar field theory in an expanding universe are used to determine the initial values of adiabatic energy-density fluctuations in cosmological models (see Sec. VII.C).

Section VI is a discussion of inflationary universe models. We explain the cosmological problems that motivated the original proposals, describe the two standard examples, the old and the new inflationary universes, and summarize many alternate models.

Besides solving the horizon and flatness problems, the main success of inflationary universe scenarios is that they provide a mechanism which for the first time explains from first principles the origin of primordial energy-density fluctuations. In Sec. VII we describe the mechanism, summarize the formalism to analyze the growth of fluctuations before they enter our horizon (gauge-invariant linear gravitational perturbation theory), and apply it to inflationary universe models.

The crucial point is that scales of present cosmological interest originate inside the causal horizon in the period of exponential expansion of the universe. This is in marked contrast to the standard big bang model, in which perturbations on all scales originate outside the causal horizon. The difference stems from the fact that, in a phase of exponential expansion, the causal horizon is constant while the physical wavelength of a perturbation increases exponentially. By contrast, in the usual cosmological models the causal horizon expands faster than the physical distance between two comoving points. We explain the basic mechanism in Sec. VII.A. Inside the causal horizon, energy-density fluctuations are caused by quantum fluctuations in the scalar field φ (as discussed in Sec. VII.C). The magnitude of these fluctuations at formation time is very small, in our sample grand unified models of the order 10^{-20}. In Sec. VII.C we discuss how the change in the equation of state from $p \simeq -\rho$ in the phase of exponential expansion to $p=\frac{1}{3}\rho$ at later stages leads to a large amplification factor for $\delta\rho/\rho$.

On the basis of this introduction, Secs. VI and VII are self-contained. Readers who are interested only in a review of inflationary universe models and fluctuations in such models can start with these sections.

A few words concerning notation: greek indices run from 0 to 3, latin ones only over the spatial indices. We use the Einstein summation convention. $m_{\rm pl}$ stands for the Planck mass. Unless otherwise indicated $\hbar=k=c=1$ in our units. We write the Robertson-Walker (RW) metric in the form

$$g_{\mu\nu}=\mathrm{diag}(-1,a^2(t),a^2(t),a^2(t)) \ . \tag{1.3}$$

$a(t)$ is the scale factor, R the Ricci scalar.

II. THE EFFECTIVE POTENTIAL

A. Generating functionals

The effective potential has long been recognized as an important concept in modern quantum field theory, in particular when questions of symmetry breaking arise. The concept was originally introduced by Heisenberg and Euler (1936) and by Schwinger (1951). It was applied to problems of symmetry breaking by Goldstone, Salam, and Weinberg (1962) and by Jona-Lasinio (1964). The physical meaning of the effective potential was explored in detail by Symanzik (1970). Good reviews of the early work are those by Zumino (1970) and (in a particularly lucid way) by Coleman (1973). Coleman and Weinberg (1973) also discovered elegant techniques for computing the one-loop effective potential, while Jackiw (1974) and Iliopoulos, Itzykson, and Martin (1975) developed a functional integral approach to computing the effective potential also to higher order.

There are many developments in the theory of the effective potential that we shall only mention briefly or not have space to discuss at all. They include generalization to composite operators (Cornwall, Jackiw, and Tomboulis, 1974), useful approximation schemes that go beyond one-loop calculations (e.g., Jackiw, 1975), and analysis of the gauge dependence of the effective action (e.g., Boulware, 1981; DeWitt, 1981; Abbott, 1981a).

The effective potential includes all quantum corrections to the classical field theory potential. Loosely speaking, minimizing the effective potential gives us the field configuration with minimal energy, the vacuum of the theory. Thus by studying the effective potential we may obtain information about the symmetries of the full theory, not just those of the Lagrangian.

We begin our discussion of the effective potential by reviewing the formalism of generating functionals. The generating functional $Z(J)$ for the full Green's functions $G_n(x_1,\ldots,x_n)$ is defined by

$$Z(J)=\left\langle 0 \left| T\exp\left[i\int J(x)\varphi(x)d^4x \right] \right| 0 \right\rangle \ , \tag{2.1}$$

where $|0\rangle$ is the physical vacuum and $\varphi(x)$ the Heisenberg field, in both cases for the theory without the source term $J(x)\varphi(x)$. Expanding $Z(J)$ as a power series in J gives the representation in terms of Green's functions:

$$\begin{aligned} Z(J) &= \sum_{n=0}^{\infty} \frac{i^n}{n!}\int d^4x_1\cdots d^4x_n J(x_1)\cdots J(x_n)\langle 0\,|\,T[\varphi(x_1)\cdots\varphi(x_n)]\,|\,0\rangle \\ &= \sum_{n=0}^{\infty} \frac{i^n}{n!}\int d^4x_1\cdots d^4x_n J(x_1)\cdots J(x_n) G_n(x_1\cdots x_n) \ . \end{aligned} \tag{2.2}$$

In functional integral representation (see, for example, Coleman, 1973),

$$\begin{aligned} Z(J) &= N\int [d\varphi]\exp iS(\varphi,J) \ , \\ S(\varphi,J) &= \int d^4x\{\mathscr{L}[\varphi(x)]+J(x)\varphi(x)\} \ . \end{aligned} \tag{2.3}$$

$\mathscr{L}(\varphi)$ is the Lagrange density of the theory, S its action, and N a normalization constant. $[d\varphi]$ is the (formal) measure on function space. For notational simplicity we express all formulas in terms of a single scalar field.

One of the main advantages of generating functionals is the fact that the transition between full, connected, and one-particle irreducible (1PI) Green's functions is given by simple algebraic operations on the level of generating functionals. As discussed below, the logarithm of $Z(J)$ is the generating functional $iW(J)$ for the connected Green's functions. The Legendre transform of $W(J)$ is the generating functional for 1PI graphs.

The generating functional for connected Green's functions $G_n^c(x_1,\ldots,x_n)$ is defined as

$$iW(J)=\sum_{n=0}^{\infty}\frac{i^n}{n!}\int d^4x_1\cdots d^4x_n J(x_1)\cdots J(x_n) \times G_n^c(x_1,\ldots,x_n) \ . \tag{2.4}$$

The connection between $Z(J)$ and $W(J)$ is given by

$$iW(J)=\ln Z(J) \ . \tag{2.5}$$

This result is well known. Its proof, since it is unrelated to the main focus of this paper, will be presented in Appendix A.

Before defining the 1PI generating functional, we introduce momentum-space 1PI Green's functions:

$$\tilde{\Gamma}^{(2)}(k,-k)=\frac{i}{D(k)}\ . \tag{2.6}$$

$D(k)$ is the full propagator and is related to the full two-point function $G^{(2)}$ by

$$G^{(2)}(k_1,k_2)=(2\pi)^4\delta^{(4)}(k_1+k_2)D(k_1)\ . \tag{2.7}$$

For $n>2$

$$i\tilde{\Gamma}^{(n)}(k_1,\ldots,k_n)=\Sigma \tag{2.8}$$

where Σ are all 1PI graphs with n external lines with momenta $k_1,\ldots,k_n$. In evaluating the graphs, the following conventions must be observed:

(a) no propagators for external lines;
(b) no overall energy-momentum-conserving δ function;
(c) all momenta directed inwards.

The first two conventions are crucial in establishing the functional relation between $W(J)$ and $\Gamma(\bar{\varphi})$. In position space

$$\Gamma^{(n)}(x_1,\ldots,x_n)=\int\prod_{i=1}^{n}\left[\frac{d^4k_i}{(2\pi)^4}e^{ik_ix_i}\right]\tilde{\Gamma}^{(n)}(k_1,\ldots,k_n)(2\pi)^4\delta^{(4)}\left[\sum_{i=1}^{n}k_i\right]. \tag{2.9}$$

Then the generating functional $\Gamma(\bar{\varphi})$ for 1PI Green's functions is defined by

$$\Gamma(\bar{\varphi})=\sum_{n=2}^{\infty}\frac{1}{n!}\int d^4x_1\cdots d^4x_n\bar{\varphi}(x_1)\cdots\bar{\varphi}(x_n)\Gamma^{(n)}(x_1,\ldots,x_n)\ . \tag{2.10}$$

At this stage $\bar{\varphi}$ stands for some source function. Why we choose this notation will become obvious later. The main result is the following.

Theorem 1. To lowest order in a,

$$\exp\left[\frac{i}{a}W(J)\right]=N\int[d\bar{\varphi}]\exp\left[\frac{i}{a}\left[\Gamma(\bar{\varphi})+\int d^4x\,J(x)\bar{\varphi}(x)\right]\right]. \tag{2.11}$$

The right-hand side of this equation is the generating functional for a theory with action $\Gamma(\bar{\varphi})$ (Γ theory). Since the Γ theory is given in terms of an unconstrained functional integral, the Feynman rules can be read off in the usual way (see, for example, Coleman, 1973): The Feynman propagator $D_F(k)$ for the Γ theory is obtained by inverting the operator coupling the two fields in the quadratic term of $\Gamma(\bar{\varphi})$:

$$\frac{1}{2}\int\bar{\varphi}(x)\Gamma^{(2)}(x,y)\bar{\varphi}(y)d^4x\,d^4y=\frac{1}{2}\int\frac{d^4k}{(2\pi)^4}\bar{\varphi}^*(k)\tilde{\Gamma}^{(2)}(k,-k)\bar{\varphi}(k)\ ,$$
$$D_F(k)\tilde{\Gamma}^{(2)}(k,-k)=i\ . \tag{2.12}$$

Thus $D_F(k)=D(k)$. This result justifies defining $\tilde{\Gamma}^{(2)}(k,-k)$ in an independent and at first sight somewhat unnatural way. The n point vertex factors are given by

$$i\tilde{\Gamma}^{(n)}(k_1,\ldots,k_n)(2\pi)^4\delta^4\left[\sum_{i=1}^{n}k_i\right], \tag{2.13}$$

which explains why we inserted the factor i in Eq. (2.8).

The theorem asserts that the connected Green's functions of the original theory may be obtained by calculating the corresponding connected Green's functions of the Γ theory in tree approximation. (The expansion in powers of a is equivalent to the loop expansion. We demonstrate this in Appendix B.)

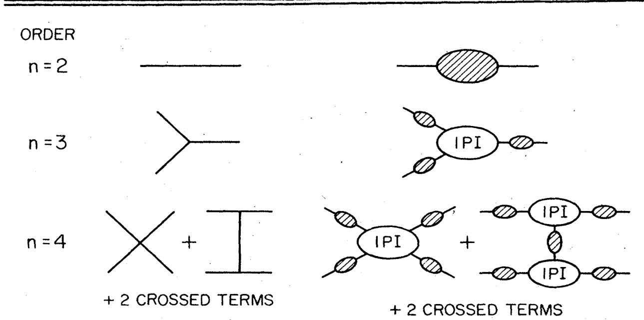
The idea underlying the proof of Theorem 1 is to write down all tree graphs of the Γ theory, reexpress Γ propagators and Γ vertex factors in terms of Green's functions of the original theory, and observe that we thus obtain the sum of all connected diagrams of the original theory. In Table I this is demonstrated for the lowest-order Green's functions, and it is straightforward to generalize the procedure to all orders. From the theorem it is easy to derive the analytic form of the relation between $W(J)$ and $\Gamma(\bar{\varphi})$. For small values of a, the right-hand side of Eq. (2.11) may be evaluated using the stationary phase approximation. Comparing the exponents yields

$$W(J)=[\Gamma(\bar{\varphi})+J\bar{\varphi}]\,|_{\delta\Gamma/\delta\bar{\varphi}=-J}\ . \tag{2.14}$$

[We shall use the shorthand notation $J\bar{\varphi}=\int d^4x \times J(x)\bar{\varphi}(x)$ frequently in the rest of the paper.] Since

$$\frac{\delta W}{\delta J}=\frac{\delta\Gamma}{\delta\bar{\varphi}}\frac{\delta\bar{\varphi}}{\delta J}+\bar{\varphi}+J\frac{\delta\bar{\varphi}}{\delta J}=\bar{\varphi}\ , \tag{2.15}$$

TABLE I. Comparison of Green's functions. The left-hand column gives the Γ theory (tree level). The right-hand column gives the result reexpressed in terms of quantities of the original theory.



the inverse transformation of Eq. (2.14) is given by

$$\Gamma(\bar{\varphi})=[W(J)-J\bar{\varphi}]\,|_{\delta W/\delta J=\bar{\varphi}}\ . \tag{2.16}$$

$\Gamma(\bar{\varphi})$ is thus the Legendre transform of $W(J)$. It is called the effective action. We can now see the reason for denoting the source of Γ by $\bar{\varphi}$: From Eq. (2.15) it follows immediately that $\bar{\varphi}$ is the average field,

$$\bar{\varphi}(x)=\frac{\delta W}{\delta J}(x)=\frac{\int [d\varphi]\varphi(x)\exp[iS(\varphi,J)]}{\int [d\varphi]\exp[iS(\varphi,J)]}\ . \tag{2.17}$$

In a translationally invariant theory $\bar{\varphi}(x)$ will be constant. In this case it is convenient to extract the infinite space-time volume arising in each term of $\Gamma(\bar{\varphi})$ and to define the effective potential $V_{\rm eff}(\bar{\varphi})$ by

$$\Gamma(\bar{\varphi})=\int d^4x[-V_{\rm eff}(\bar{\varphi})]\ . \tag{2.18}$$

We conclude the discussion of generating functionals with a different proof of the relation between $W(J)$ and $\Gamma(\bar{\varphi})$ (see Zumino, 1970). Our starting point is the definition of $\Gamma(\bar{\varphi})$ as the Legendre transform of $W(J)$. It immediately follows that

$$k(x,x')=\frac{\delta^2\Gamma(\bar{\varphi})}{\delta\bar{\varphi}(x)\delta\bar{\varphi}(x')}=-\frac{\delta J(x)}{\delta\bar{\varphi}(x')} \tag{2.19}$$

is the inverse of the connected two-point function

$$-G(x,x')=\frac{\delta^2 W}{\delta J(x)\delta J(x')}=\frac{\delta\bar{\varphi}(x)}{\delta J(x')}\ . \tag{2.20}$$

Hence we obtain the useful relations

$$\frac{\delta G(x,x')}{\delta J(x'')}=G(x,y)\frac{\delta k(y,y')}{\delta J(x'')}G(y',x') \tag{2.21}$$

and

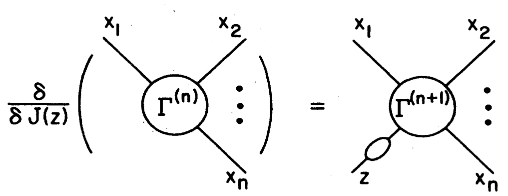
$$\frac{\delta}{\delta J}=G\frac{\delta}{\delta\bar{\varphi}}\ . \tag{2.22}$$

These equations form the basis for an inductive proof that $\Gamma(\bar{\varphi})$ is the generating functional of 1PI graphs. Consider first the three-point function:

$$\begin{aligned}\frac{\delta^3 W}{\delta J(x_1)\delta J(x_2)\delta J(x_3)}&=\frac{\delta}{\delta J(x_1)}G(x_2,x_3)=G(x_2,y_2)\frac{\delta k(y_2,y_3)}{\delta J(x_1)}G(y_3,x_3)\\&=G(x_1,y_1)G(x_2,y_2)G(x_3,y_3)\frac{\delta^3\Gamma}{\delta\bar{\varphi}(y_1)\delta\bar{\varphi}(y_2)\delta\bar{\varphi}(y_3)}\ ,\\&=G(x_1,y_1)G(x_2,y_2)G(x_3,y_3)\Gamma^{(3)}(y_1,y_2,y_3)\ .\end{aligned} \tag{2.23}$$

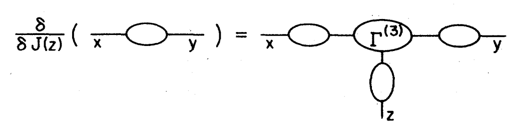
In terms of Feynman graphs this result is presented in Fig. 1. Obviously, $\Gamma^{(3)}(y_1,y_2,y_3)$ is the 1PI three-point function.

The two essential ingredients of the induction proof are



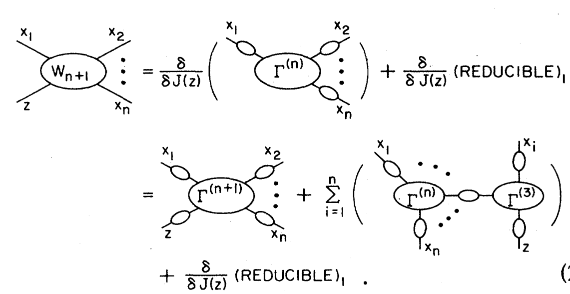
(2.24)

[Eq. (2.22)], and



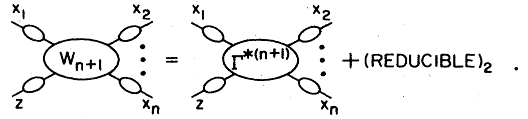
(2.25)

The $\Gamma^{(n)}(x_1,\ldots,x_n)$ are defined as functional derivatives of $\Gamma(\bar{\varphi})$. We must prove that $\Gamma^{(n)}$ is the sum of all 1PI graphs. Assume the claim has been shown to order n. Now



(2.26)

$\Gamma^{(n+1)}$ is defined as a functional derivative of $\Gamma^{(n)}$. On the other hand, W_{n+1} can also be expressed pictorially as a 1PI piece $\Gamma^{*(n+1)}$ defined in terms of the graphical expansion plus a sum of reducible graphs:



(2.27)

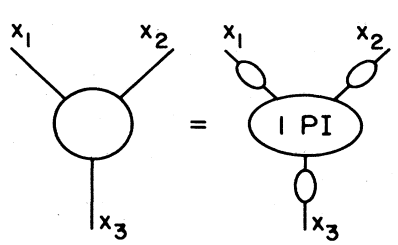
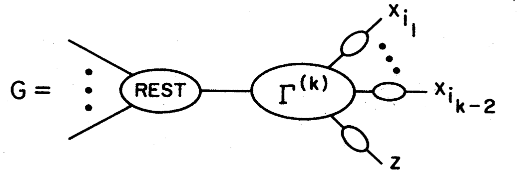


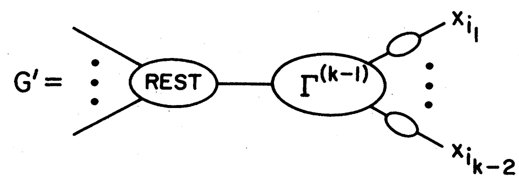
FIG. 1. Decomposition of the connected three-point function.

We must show that $\Gamma^{*(n+1)}$ and $\Gamma^{(n+1)}$ are identical. Equivalently we prove that there is a one-to-one correspondence between graphs in the two last terms on the right-hand side of Eq. (2.26) and those in $(\text{reducible})_2$. Obviously, all of the former graphs are reducible. Differentiation cannot produce more connectedness. To construct the inverse map, consider a reducible $n+1$ point graph G. Pick the z vertex and isolate the irreducible component it couples to, $\Gamma^{(k)}$ with $3 \le k \le n$:



(2.28)

If $k=3$ and the "rest" is 1PI, then G stems from taking the functional derivative of a $\Gamma^{(n)}$ term in $W^{(n)}$. Otherwise, G is one term in the derivative of the element



(2.29)

of the set $(\text{reducible})_1$; and G can be obtained in no other way as an element of the set of graphs of Eq. (2.26). Hence we have constructed the one-to-one map and proved that $\Gamma^{(n+1)}$ and $\Gamma^{*(n+1)}$ are identical.

B. Physical meaning of the effective potential and effective action

The effective potential is a very important concept in field theory. It is the quantum field potential energy in a sense that will be stated precisely below. In addition, it has the nice feature of being easily calculable. We shall discuss the latter point first.

For constant $\bar{\varphi}$ we can write

$$\begin{aligned}\Gamma(\bar{\varphi}) &= \sum_{n=2}^{\infty} \frac{1}{n!}\bar{\varphi}^n \int \prod_{i=1}^{n} d^4x_i \Gamma^{(n)}(x_1,\ldots,x_n) \\ &= \sum_{n=2}^{\infty} \frac{1}{n!}\bar{\varphi}^n \int \prod_{i=1}^{n} \left[d^4x_i \frac{d^4k_i}{(2\pi)^4} e^{ik_ix_i} \right] \tilde{\Gamma}^{(n)}(k_1,\ldots,k_n)(2\pi)^4 \delta^{(4)}\left[\sum_{i=1}^{n} k_i\right] \\ &= \sum_{n=2}^{\infty} \frac{1}{n!}\bar{\varphi}^n \int d^4x_1 \tilde{\Gamma}^{(n)}(0,\ldots,0) \; . \end{aligned} \tag{2.30}$$

Therefore,

$$V_{\text{eff}}(\bar{\varphi}) = -\sum_{n=2}^{\infty} \frac{1}{n!}\bar{\varphi}^n \tilde{\Gamma}^{(n)}(0,\ldots,0) \; . \tag{2.31}$$

$V_{\text{eff}}(\bar{\varphi})$ can thus be viewed as the generating functional for 1PI graphs with vanishing external momenta. We can calculate it by summing all 1PI graphs and inserting a factor $\bar{\varphi}$ for each external line. The effective potential for a theory with Lagrange density

$$\mathcal{L}(\varphi,\partial_\mu\varphi) = \tfrac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - U(\varphi) \tag{2.32}$$

to lowest order in $\hbar$ (see Appendix B) reduces to the classical potential $U(\bar{\varphi})$, i.e.,

$$V_{\text{eff}}(\bar{\varphi}) = U(\bar{\varphi}) + \mathcal{O}(\hbar) \; . \tag{2.33}$$

The proof is easy: To lowest order in $\hbar$ only tree graphs contribute. All tree graphs with more than one vertex are not 1PI. Thus each term $(g_i/i!)\varphi^i$ in $U(\varphi)$ gives rise to exactly one tree-level 1PI graph whose value is $-g_i\bar{\varphi}^i$ ($-g_i$ is the vertex factor, and the $\bar{\varphi}^i$ comes from the i external lines). In $V_{\text{eff}}(\bar{\varphi})$ the minus signs cancel and the factor $1/i!$ is reinserted, yielding $U(\bar{\varphi})$.

The physical interpretation of the effective potential is summarized in the following theorem (see, for example, Symanzik, 1970, and Coleman, 1973).

Theorem 2. $V_{\text{eff}}(\bar{\varphi})$ is the minimum of the energy-density expectation value in the class of all normalized states $|a\rangle$ satisfying $\langle a|\varphi|a\rangle = \bar{\varphi}$, i.e.,

$$V_{\text{eff}}(\bar{\varphi}) = \langle a|\mathcal{H}|a\rangle \tag{2.34}$$

for a state $|a\rangle$ which obeys

$$\delta\langle a|\mathcal{H}|a\rangle = 0$$

subject to the constraints

$$\langle a|a\rangle = 1, \quad \langle a|\varphi|a\rangle = \bar{\varphi} \; . \tag{2.35}$$

The proof of this assertion proceeds in two steps. In the first we solve the corresponding quantum mechanics problem. Then we translate the results into field theory.

In quantum mechanics the problem is to find a state $|a\rangle$ such that $\langle a|H|a\rangle$ is stationary subject to the constraints $\langle a|a\rangle = 1$ and $\langle a|A|a\rangle = A_c$ for some self-adjoint operator A. We can incorporate the constraints into the variational equation by introducing Lagrange multipliers E and J. The new variational problem becomes

$$\begin{aligned} &\delta\langle a|(H-E-JA)|a\rangle = 0 \quad (\text{unconstrained}) \; , \\ &\langle a|a\rangle = 1 \; , \\ &\langle a|A|a\rangle = A_c \; . \end{aligned} \tag{2.36}$$

Since H is assumed to be self-adjoint, the variational equation is

$$(H-E-JA)\,|\,a\,\rangle=0\ , \tag{2.37}$$

which has some (in general unnormalized) solution $|\,a\,\rangle=|\,a(E,J)\rangle$. Using the normalization condition $\langle a\,|\,a\,\rangle=1$, we can solve for E as a function of J. The operator constraint $\langle a\,|\,A\,|\,a\,\rangle=A_c$ will give J as a function of A_c. Reinserting these relations into Eq. (2.37) we obtain

$$[H-E(J)-JA]\,|\,a(J)\rangle=0\ . \tag{2.38}$$

$|\,a(J)\rangle$ is a normalized eigenstate of $H-JA$. In field theory we shall be able to make a stronger statement: Since the only normalizable eigenstate of a Hamiltonian is the vacuum state, we shall be able to identify $|\,a(J)\rangle$ with the vacuum state of $H-JA$. If we take the scalar product of (2.38) with $\langle a(J)\,|$ and functionally differentiate the resulting equation with respect to J we get

$$\begin{aligned}0&=\frac{\delta}{\delta a}\langle a(J)\,|\,(H-JA)\,|\,a(J)\rangle\frac{\delta a}{\delta J}\\&\quad-\langle a(J)\,|\,A\,|\,a(J)\rangle-\frac{\delta E}{\delta J}\\&=2\langle\delta a\,|\,(H-JA)\,|\,a\,\rangle\frac{\delta a}{\delta J}-\frac{\delta E}{\delta J}-A_c\ .\end{aligned} \tag{2.39}$$

Since $\langle a\,|\,a\,\rangle=1$ implies

$$\langle\delta a\,|\,(H-JA)\,|\,a\,\rangle=E\langle\delta a\,|\,a\,\rangle=0\ , \tag{2.40}$$

the first term of Eq. (2.39) vanishes, and we conclude

$$\frac{\delta E}{\delta J}=-A_c\ ; \tag{2.41}$$

from Eq. (2.38),

$$\langle a(J)\,|\,H\,|\,a(J)\rangle=E(J)-J\frac{\delta E}{\delta J}\ . \tag{2.42}$$

$E(J)-J\delta E/\delta J$ is therefore the energy of the ground state of H in the space of states satisfying the constraint

$$\langle a(J)\,|\,A\,|\,a(J)\rangle=A_c(J)\ . \tag{2.43a}$$

The translation of this result to field theory is based on Eq. (2.16),

$$\Gamma(\overline{\varphi})=W(J)-J\frac{\delta W}{\delta J},\quad \frac{\delta W}{\delta J}=\overline{\varphi}\ . \tag{2.43b}$$

Formally these equations are identical to Eqs. (2.41) and (2.42). We obtain the physical interpretation of $\Gamma(\overline{\varphi})$ by recalling the physical meaning of $W(J)$ and invoking the formal equivalence of the equations. Consider adding a source term $J(x)\varphi(x)$ to the Hamiltonian density. If the source $J(x)$ is adiabatically turned on at time zero and off at time T, and if it is localized in a volume V, then

$$\exp[iW(J)]=\langle 0^+\,|\,0^-\rangle=\exp[-iVT\varepsilon(J)]\ . \tag{2.44}$$

The vacuum $t\rightarrow-\infty$ Schrödinger state $|\,0^-\rangle$ of the sourceless theory shifts adiabatically into the ground state of the theory with source, then propagates with a phase velocity determined by its energy density $\varepsilon(J)$ and returns adiabatically to the unperturbed ground state $|\,0^+\rangle$ at $t>T$, picking up a total phase $VT\varepsilon(J)$. Thus $-W(J)$ is the ground-state action of the theory with source $J(x)\varphi(x)$, and in analogy with the quantum mechanics problem considered before we conclude that $-\Gamma(\overline{\varphi})$ is the ground-state action of the sourceless theory in the space of states satisfying

$$\langle a\,|\,\varphi\,|\,a\,\rangle=\overline{\varphi}\ . \tag{2.45}$$

Dividing by $-VT$ we have thus verified that

$$V_{\rm eff}(\overline{\varphi})=\langle a\,|\,\mathscr{H}\,|\,a\,\rangle \tag{2.46}$$

for the lowest-energy state (i.e., $\delta\langle a\,|\,\mathscr{H}\,|\,a\,\rangle=0$) in the space of states satisfying

$$\langle a\,|\,a\,\rangle=1,\ \langle a\,|\,\varphi\,|\,a\,\rangle=\overline{\varphi}\ . \tag{2.47}$$

We can draw important conclusions from Theorem 2. First, minimizing $V_{\rm eff}(\overline{\varphi})$ with respect to $\overline{\varphi}$ gives the ground-state energy of the theory. Furthermore, the value of $\overline{\varphi}$ at which the minimum of the effective potential is taken on yields information about spontaneous symmetry breaking. If the minimum occurs at $\overline{\varphi}\neq0$, then all symmetries of $\mathscr{L}(\varphi,\partial_\mu\varphi)$ which do not leave $\overline{\varphi}$ invariant are spontaneously broken.

Above, we have shown that the effective potential can be obtained by a variational principle on static field configurations, supplemented by constraints on the states. Jackiw and Kerman (1979) generalized the idea and obtained a variational definition of the full effective action. We state their result as a theorem.

Theorem 3. $\Gamma(\varphi)$ is the stationary value of the time-integrated matrix element of $i\partial_t-H$ taken between time-dependent states $|\,\psi_\pm,t\,\rangle$ subject to the constraints

$$\begin{aligned}&\langle\psi_-,t\,|\,\Phi(\mathbf{x})\,|\,\psi_+,t\,\rangle=\varphi(t,\mathbf{x})\ ,\\&\langle\psi_-,t\,|\,\psi_+,t\,\rangle=1\ ,\\&\lim_{t\rightarrow\pm\infty}|\,\psi_\pm,t\,\rangle=|\,0\rangle\ ,\end{aligned} \tag{2.48}$$

i.e., $\Gamma(\varphi)$ is the stationary value of

$$\Gamma(\varphi)=\int_{-\infty}^{\infty}dt\langle\psi_-,t\,|\,i\partial_t-H\,|\,\psi_+,t\,\rangle \tag{2.49}$$

subject to the above constraints.

The idea of the proof is similar to that employed for Theorem 2, so we shall just sketch the main points. To incorporate the first two constraints we add Lagrange multiplier terms to $\Gamma(\varphi)$. Variation with respect to $|\,\psi_+,t\,\rangle$ and $|\,\psi_-,t\,\rangle$ then yields

$$\begin{aligned}&\left[i\partial_t-H+\int dx\,J\varphi\right]|\,\psi_+,t\,\rangle=w(t)\,|\,\psi_+,t\,\rangle\ ,\\&\left[i\partial_t-H+\int dx\,J\varphi\right]|\,\psi_-t\,\rangle=w^*(t)\,|\,\psi_-,t\,\rangle\ .\end{aligned} \tag{2.50}$$

$w(t)$ is a Lagrange multiplier. We define new states $|\,\pm,t\,\rangle$ by

$$| +,t\rangle = \exp\left[i\int_{\infty}^{t} dt' w(t')\right] | \psi_+ t\rangle ,$$
$$| -,t\rangle = \exp\left[-i\int_{t}^{\infty} dt' w^*(t')\right] | \psi_- t\rangle . \quad (2.51)$$

These new states are the asymptotic ground states. Hence

$$\langle -,t | +,t\rangle = e^{iW(J)} = \exp\left[i\int_{-\infty}^{\infty} dt\, w(t)\right] , \quad (2.52)$$

and

$$W(J) = \int_{-\infty}^{\infty} dt \left\langle \psi_-,t \left| i\partial_t - H + \int d^3x\, J\varphi \right| \psi_+,t\right\rangle . \quad (2.53)$$

It is not hard to combine the equations of motion (2.50) and the constraints (2.48) to conclude that

$$\frac{\delta W(J)}{\delta J} = \varphi . \quad (2.54)$$

Hence

$$W(J) - \int d^4x\, J\varphi = \Gamma(\varphi)$$
$$= \int_{-\infty}^{\infty} dt \langle \psi_-,t | i\partial_t - H | \psi_+,t\rangle . \quad (2.55)$$

To conclude this section we should like to stress the physical interpretation of

$$\frac{\delta\Gamma}{\delta\bar{\varphi}} = J , \quad (2.56)$$

which follows from the definition of Γ as a Legendre transform. For vanishing source terms, the variation of Γ must be zero. This is the quantum-mechanical equation of motion for $\varphi(\mathbf{x},t)$.

Instead of expanding in powers of φ, we can expand in powers of momentum about the point where all external momenta vanish (see, for example, Coleman, 1973):

$$\Gamma(\varphi) = \int d^4x [-V_{\text{eff}}(\varphi) + \tfrac{1}{2}(\partial_\mu\varphi\partial^\mu\varphi)Z(\varphi) + \cdots] . \quad (2.57)$$

Working to this order in momenta and renormalizing the field such that $Z(\varphi)$ becomes unity, we obtain the following equation of motion:

$$\Box\varphi = -V'_{\text{eff}}(\varphi) \quad (2.58)$$

(vanishing external sources). Thus, close to the static limit, the quantum corrections to the classical scalar field equation of motion mean replacing the potential by the effective potential. This is another justification for using the notion of effective potential.

C. Computation of the effective potential and effective action

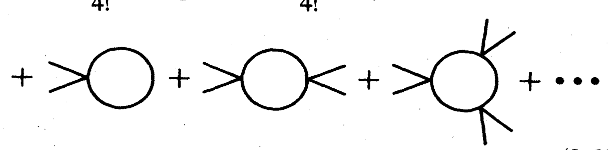
We first present the elegant technique of Coleman and Weinberg (1973) for summing all graphs contributing to the one-loop effective potential, following the exposition in Coleman (1973). We develop the combinatorial method in the case of the theory of a single scalar field and then generalize to theories involving Fermi and gauge fields. The combinatorial method cannot easily be generalized to higher orders in the loop expansion. Therefore, we sketch a method that can, the functional integral method developed by Jackiw (1974) and Iliopoulos, Itzykson, and Martin (1975). We conclude the section by briefly mentioning other approximation schemes.

Consider first a scalar field theory given by the Lagrange density

$$\mathscr{L}(\varphi) = \tfrac{1}{2}\partial_\mu\varphi\partial^\mu\varphi - \frac{\lambda}{4!}\varphi^4 + \tfrac{1}{2}A\,\partial_\mu\varphi\partial^\mu\varphi - \tfrac{1}{2}B\varphi^2 - \frac{1}{4!}C\varphi^4 . \quad (2.59)$$

The final three terms are renormalization counterterms and hence of order $\hbar$. Since $\bar{\varphi}$ is constant, no A-type counterterms are required in computing the effective potential. Let $V_{\text{eff}}^{(i)}(\bar{\varphi})$ denote the effective potential to order $\hbar^i$ and $B^{(i)}$ ($C^{(i)}$) the B (C) counterterms to order $\hbar^i$. Using Eqs. (2.31) and (2.33), we see that $V_{\text{eff}}^{(1)}(\bar{\varphi})$ is obtained by adding to the classical potential $U(\bar{\varphi})$ to order $\hbar$ all one-loop graphs with vanishing external momenta, i.e.,

$$V_{\text{eff}}^{(1)}(\bar{\varphi}) = \frac{\lambda}{4!}\bar{\varphi}^4 + \tfrac{1}{2}B^{(1)}\bar{\varphi}^2 + \frac{1}{4!}C^{(1)}\bar{\varphi}^4$$



(2.60)

Since all external momenta are zero, the loop integration is the same for each graph. This is the main reason we are able to sum the 1PI one-loop graphs explicitly. Since there is a factor $\bar{\varphi}$ for each external line and a factor $1/2n$ from the indistinguishability of external lines (n is the number of vertices), the sum of the graphs is

$$i\sum_{n=1}^{\infty}\int\frac{d^4k}{(2\pi)^4}\frac{1}{2n}\left[\frac{\frac{1}{2}\lambda\bar{\varphi}^2}{k^2+i\varepsilon}\right]^n = -\frac{i}{2}\int\frac{d^4k}{(2\pi)^4}\ln\left[1-\frac{\frac{1}{2}\lambda\bar{\varphi}^2}{k^2+i\varepsilon}\right]$$
$$= \frac{1}{2}\int\frac{d^4k_E}{(2\pi)^4}\ln\left[1+\frac{\lambda\bar{\varphi}^2}{2k_E^2}\right] . \quad (2.61)$$

The initial i stems from the fact that $i\Gamma$ and not Γ is the sum of all 1PI graphs [cf. Eq. (2.8)]. Wick rotation to Euclidean space produces the compensating i. While it is already remarkable that the infinite sum of graphs can be performed explicitly, it is even more remarkable that in the process the infrared behavior is improved. Each individual graph has a polynomial infrared singularity which gets worse as we proceed to higher orders in the perturbative expansion. The sum of graphs, on the other hand, has only a logarithmic infrared singularity.

Renormalization of the theory proceeds along the usual lines. First, we regularize by imposing a momentum-space cutoff at $k_E^2=\Lambda^2$. Then the integral (2.61) is elementary and yields

$$\frac{\lambda\Lambda^2}{64\pi^2}\bar{\varphi}^2+\frac{\lambda^2\bar{\varphi}^4}{256\pi^2}\left[\ln\frac{\lambda\bar{\varphi}^2}{2\Lambda^2}-\frac{1}{2}\right] . \tag{2.62}$$

Mass and coupling-constant renormalization fix $B^{(1)}$ and $C^{(1)}$. Our mass renormalization condition demands that the renormalized mass vanish:

$$\left.\frac{d^2V_{\rm eff}^{(1)}}{d\bar{\varphi}^2}\right|_{\bar{\varphi}=0}=0 . \tag{2.63}$$

This implies

$$B^{(1)}=-\frac{\lambda\Lambda^2}{32\pi^2} . \tag{2.64}$$

Due to the infrared divergence, we must perform coupling-constant renormalization at a nonsymmetric point $\bar{\varphi}=M\neq0$:

$$\left.\frac{d^4V_{\rm eff}^{(1)}}{d\bar{\varphi}^4}\right|_{\bar{\varphi}=M}=\lambda . \tag{2.65}$$

An elementary computation yields

$$C^{(1)}=-\frac{11}{32}\frac{\lambda^2}{\pi^2}-\frac{3\lambda^2}{32\pi^2}\ln\frac{\lambda M^2}{2\Lambda^2} . \tag{2.66}$$

Combining these results, we find

$$V_{\rm eff}^{(1)}(\bar{\varphi})=\frac{\lambda}{4!}\bar{\varphi}^4+\frac{\lambda^2\bar{\varphi}^4}{256\pi^2}\left[\ln\frac{\bar{\varphi}^2}{M^2}-\frac{25}{6}\right] . \tag{2.67}$$

For small values of $\bar{\varphi}$ this is negative. Hence this calculation gives us the important result that radiative corrections may destabilize the tree-level ground state $\bar{\varphi}=0$ (see Fig. 2; Coleman and Weinberg, 1973).

A necessary condition for spontaneous symmetry breaking via radiative corrections is that the renormalized mass be zero (such theories are commonly called Coleman-Weinberg-type models). Another condition is that the effective loop-expansion parameter $|\lambda\ln(\bar{\varphi}^2/M^2)|$ be small (only then is the loop expansion reliable). In the scalar quantum field theory given by Eq. (2.59) the second condition is violated for all values of λ, since at the minimum of $V_{\rm eff}^{(1)}(\bar{\varphi})$

$$\lambda\ln\frac{\bar{\varphi}^2}{M^2}=-\tfrac{32}{3}\pi^2+\mathscr{O}(\lambda) . \tag{2.68}$$

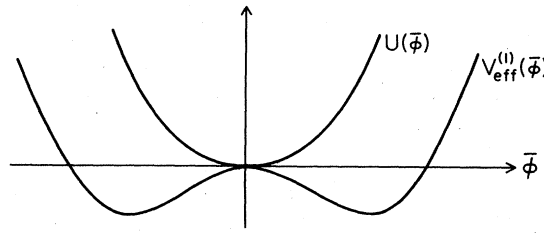


FIG. 2. Effective potential for a sample scalar field theory. $U(\bar{\varphi})$ is the (tree-level effective) potential, $V_{\rm eff}^{(1)}(\bar{\varphi})$ the one-loop effective potential.

The mechanism of spontaneous symmetry breaking via radiative corrections does work for massless scalar electrodynamics (Coleman and Weinberg, 1973) and for many non-Abelian gauge theories.

The computation of the one-loop effective potential is conceptually the same for non-Abelian gauge theories, although the technical details are more tedious. Consider a theory given by

$$\mathscr{L}(A,\psi,\varphi)=-\tfrac{1}{4}{\rm Tr}(F_{\mu\nu}F^{\mu\nu})+i\,{\rm Tr}\bar{\psi}\not{D}\psi+{\rm Tr}\bar{\psi}M\psi$$
$$-{\rm Tr}\bar{\psi}\Gamma\varphi\psi+\tfrac{1}{2}{\rm Tr}(D_\mu\varphi)^+D^\mu\varphi-U(\varphi) . \tag{2.69}$$

We are using the usual notation (see, for example, Itzykson and Zuber, 1980): Given a non-Abelian gauge group G with generators τ_α, a set of Dirac bispinor fields ψ^a, and a set of scalar (Higgs) fields φ_i, each of which forms a representation space for G, then

$$A_\mu=A_\mu^\alpha\tau_\alpha ,$$
$$F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu+ig[A_\mu,A_\nu] , \tag{2.70}$$
$$D_\mu=\partial_\mu+igA_\mu .$$

In D_μ the τ_α are the matrices that represent the generators of G in the appropriate representation space. The trace also refers to the appropriate representation space. If the Higgs and Fermi fields are represented as column vectors, the trace reduces to the scalar product, e.g.,

$${\rm Tr}(D_\mu\varphi)^+D^\mu\varphi=(D_\mu\varphi)^+D^\mu\varphi . \tag{2.71}$$

Finally, Γ denotes the matrix of Yukawa coupling constants.

The Feynman rules in general involve extra fields, the ghost fields ω and $\bar{\omega}$, scalar fields obeying Fermi statistics (Faddeev and Popov, 1967). They arise as a consequence of the field-dependent functional determinant in the Faddeev-Popov ansatz for the generating functional. In covariant gauges, the ghost term in the effective Lagrangian is

$${\rm Tr}\bar{\omega}\partial_\mu A^\mu\omega \tag{2.72}$$

and contains no Higgs couplings to ghost fields. Hence to one-loop order no ghosts arise in our calculation of the effective potential. Therefore,

$$V_{\rm eff}^{(1)}(\bar\varphi)=U(\bar\varphi)+V_{\rm c.t.}^{(1)}+V_{1\,\rm loop}\ . \tag{2.73}$$

$V_{\rm c.t.}^{(1)}$ contains all the counterterms to order $\hbar$, and $V_{1\,\rm loop}$ is the sum of all one-loop graphs with external φ lines at zero momenta. The particle propagating around the loop can be a Higgs boson, a fermion, or a gauge boson. (By G charge conservation the particle type cannot change within a single loop.) The corresponding contributions to $V_{1\,\rm loop}$ are denoted by $V_S^{(1)}$, $V_f^{(1)}$, and $V_g^{(1)}$. We can calculate each of these contributions by following the procedure outlined in the scalar field theory example.

For $V_S^{(1)}$ the only change is the presence of more than one scalar particle. The relevant diagrams are those of Eq. (2.60). Consider a vertex with one internal line associated with φ_i, the other with φ_j. Then the effective coupling constant, which includes the physical coupling constant, the $\bar\varphi$ factors for external legs, and the Bose combinatorial factors for internal links, equals

$$W^{ij}=\left.\frac{\partial^2 U(\varphi)}{\partial\varphi_i\partial\varphi_j}\right|_{\varphi=\bar\varphi}\ . \tag{2.74}$$

In our single scalar field theory given by Eq. (2.59),

$$W=\tfrac{1}{2}\lambda\bar\varphi^2\ . \tag{2.75}$$

If $U(\varphi)$ is an even quartic polynomial (the standard Coleman-Weinberg potential for a renormalizable theory with $\varphi\to-\varphi$ symmetry), then $W(\bar\varphi)$ is a quadratic form and hence can be diagonalized. In the diagonal basis the different Higgs fields decouple, and $V_S^{(1)}$ can be obtained by adding the contributions of the new basis fields. Each individual term is given by Eq. (2.62). Neglecting the λ^2 corrections to terms already contained in $V_{\rm eff}^{(0)}(\bar\varphi)$, the cutoff-independent part of Eq. (2.62) becomes

$$\frac{1}{64\pi^2}\left[\frac{\lambda\bar\varphi^2}{2}\right]^2\ln\frac{\lambda\bar\varphi^2}{2}\ . \tag{2.76}$$

Therefore, in the diagonal basis

$$V_S^{(1)}(\bar\varphi)=\frac{1}{64\pi^2}\sum_i (W^{ii})^2\ln W^{ii}$$
$$=\frac{1}{64\pi^2}{\rm Tr}(W^2\ln W)\ . \tag{2.77}$$

The latter expression is basis independent.

$V_f^{(1)}(\bar\varphi)$ is given by the sum of the diagrams in Fig. 3 (we are considering the case $M=0$). Since the trace of an odd number of γ matrices vanishes, the diagrams with an odd number of vertices are zero. If we insert all indices explicitly, the Yukawa coupling term in Eq. (2.69) is

$$\sum_{a,b,i}\bar\psi^a\Gamma^i_{ab}\varphi_i\psi^b\ . \tag{2.78}$$

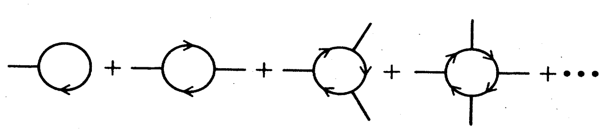


FIG. 3. One-loop graphs contributing to $V_f^{(1)}(\bar\varphi)$.

The second diagram thus gives a contribution of

$$-\tfrac{1}{2}\sum_{i,j,a,b}{\rm Tr}_S\bar\varphi_i\Gamma^i_{ab}\frac{k^2}{k^4}\Gamma^j_{ba}\bar\varphi_j\ . \tag{2.79}$$

We write ${\rm Tr}_S$ for a trace over spinor indices. If we take this trace and simplify the notation by introducing a trace ${\rm Tr}_f$ in the fermion representation space, we get

$$-\frac{1}{2}\frac{4}{k^2}{\rm Tr}_f(\bar\varphi\Gamma)^2 \tag{2.80}$$

with $(\bar\varphi\Gamma)_{ab}=\bar\varphi_i\Gamma^i_{ab}$. Analogously the graph with $2n$ vertices yields

$$-\frac{1}{2n}\frac{4}{(k^2)^n}{\rm Tr}_f(\bar\varphi\Gamma)^{2n}\ . \tag{2.81}$$

Thus (dropping the subscript f for the trace)

$$V_f^{(1)}(\bar\varphi)=-4i\,{\rm Tr}\sum_{n=1}^{\infty}\int\frac{d^4k}{(2\pi)^4}\frac{1}{2n}\left[\frac{(\bar\varphi\Gamma)^2}{k^2}\right]^n\ . \tag{2.82}$$

This result is analogous to Eq. (2.61), and therefore further evaluation is identical to the procedure in the scalar example. We obtain for the cutoff-independent contribution [cf. Eq. (2.76)]

$$V_f^{(1)}(\bar\varphi)=-\frac{1}{64\pi^2}4\,{\rm Tr}[(\bar\varphi\Gamma)^4\ln(\bar\varphi\Gamma)^2]\ . \tag{2.83}$$

Finally we come to the calculation of $V_g^{(1)}(\bar\varphi)$: The contributing diagrams are given in Fig. 4. The only vertex that contributes to one-loop order is the $A^2\varphi^2$ vertex. It stems from the following term in the Lagrange density:

$$\tfrac{1}{2}{\rm Tr}[(igA_\mu\varphi)^+igA^\mu\varphi]=\tfrac{1}{2}g_\alpha g_\beta{\rm Tr}[(\tau_\alpha A^\alpha_\mu\varphi)^+\tau_\beta A^{\mu\beta}\varphi]$$
$$=\tfrac{1}{2}A^\alpha_\mu A^{\mu\beta}M^2_{\alpha\beta} \tag{2.84}$$

with

$$M^2_{\alpha\beta}=g_\alpha g_\beta{\rm Tr}[(\tau_\alpha\varphi)^+\tau_\beta\varphi]\ . \tag{2.85}$$

First a few general comments on Eq. (2.84). (1) If G is not a simple group, then in general there will be a different coupling constant associated with each field A^α_μ. (2) The trace is a trace in the Higgs representation space. (3) $M^2_{\alpha\beta}|_{\varphi=\bar\varphi}$ is the mass matrix for the gauge bosons in the broken phase with vacuum expectation value $\bar\varphi$. At the same time $M^2_{\alpha\beta}$ is the gauge boson effective-coupling-constant matrix in the sense of Eq. (2.74). The last comment also shows that the calculation of $V_g^{(1)}$ is analogous to that of $V_S^{(1)}$. Since in Landau gauge the gauge fields have three degrees of freedom, we obtain (with a Lie algebra trace)

$$V_g^{(1)}(\bar\varphi)=\frac{3}{64\pi^2}{\rm Tr}(M^4\ln M^2)|_{\varphi=\bar\varphi}\ . \tag{2.86}$$

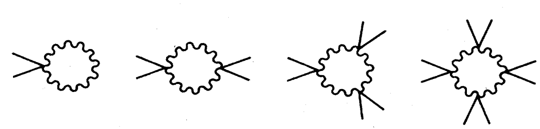


FIG. 4. One-loop graphs contributing to $V_g^{(1)}(\bar\varphi)$.

The combinatorial technique outlined above is an elegant and conceptually simple way to calculate the one-loop effective potential. Unfortunately, we know of no easy way to generalize the method to higher orders in $\hbar$. The functional integral approach sketched below does admit an easy generalization. At any given order in $\hbar$ the computation of the effective potential will be reduced to the calculation of a finite number of Feynman graphs with respect to a new interaction.

The crucial point in the functional integral approach (Jackiw, 1974; Iliopoulos, Itzykson, and Martin, 1975) is the fact that the loop expansion is equivalent to an expansion about a stationary point of the action. We shall use the background-field method, originally due to DeWitt (1964) and extended by 't Hooft (1975), DeWitt (1981), Boulware (1981), and Abbott (1981a). See Abbott (1982) for an introduction.

Consider the shifted generating functional

$$\widetilde{Z}(J,\varphi)=\int[d\varphi']e^{iS(\varphi'+\varphi)+iJ\varphi'}N\ . \tag{2.87}$$

Here φ is the classical background field, N is the normalization constant,

$$N^{-1}=\int[d\varphi']e^{iS(\varphi')}\ . \tag{2.88}$$

Similarly we can define $\widetilde{W}(J,\varphi)$ by

$$\widetilde{Z}(J,\varphi)=e^{i\widetilde{W}(J,\varphi)}\ . \tag{2.89}$$

The generating functional for 1PI graphs in the given background field φ is defined in analogy to Eq. (2.16) as the Legendre transform of $\widetilde{W}(J,\varphi)$:

$$\widetilde{\Gamma}(\widetilde{\varphi},\varphi)=\widetilde{W}(J,\varphi)-J\widetilde{\varphi}\ , \tag{2.90}$$

where on the right-hand side J is expressed in terms of $\widetilde{\varphi}$ by inverting

$$\widetilde{\varphi}=\frac{\delta\widetilde{W}(J,\varphi)}{\delta J}\ . \tag{2.91}$$

By shifting the integration variable in Eq. (2.87) it is easy to check that

$$\widetilde{\Gamma}(\widetilde{\varphi},\varphi)=\Gamma(\widetilde{\varphi}+\varphi)\ . \tag{2.92}$$

In particular,

$$\widetilde{\Gamma}(0,\varphi)=\Gamma(\varphi)\ . \tag{2.93}$$

Equation (2.93) yields an alternate procedure for computing the effective action $\Gamma(\varphi)$: $\widetilde{\Gamma}(0,\varphi)$ is the sum of all 1PI vacuum graphs of the theory in the background field φ.

We now prove that, to any given order in $\hbar$, $\widetilde{\Gamma}(0,\varphi)$ is a finite sum of graphs in the background theory. Clearly we must keep track of powers of $\hbar$. Recall that both terms in the exponent of Eq. (2.87) contain a factor $\hbar^{-1}$. We rescale the quantum field φ' by defining $\psi'=\hbar^{-1/2}\varphi'$. For our standard example, a $(\lambda/4!)\varphi^4$ theory, we obtain

$$\hbar^{-1}S(\varphi'+\varphi)=\hbar^{-1}S(\varphi)-\hbar^{-1/2}\int d^4x\,\psi'\left[\partial_\mu\partial^\mu\varphi-\mu^2\varphi-\frac{\lambda}{6}\varphi^3\right]$$
$$+\int d^4x\left[\tfrac{1}{2}\partial_\mu\psi'\partial^\mu\psi'+\frac{1}{2}\left[\mu^2+\frac{\lambda}{2}\varphi^2\right]\psi'^2+\frac{\lambda\hbar^{1/2}}{6}\varphi\psi'^3+\frac{\lambda}{4!}\psi'^4\right]\ . \tag{2.94}$$

The second term vanishes if we expand about a classical solution φ. In this case

$$\hbar^{-1}S(\varphi'+\varphi)=\hbar^{-1}S(\varphi)+\int d^4x\left[\tfrac{1}{2}\partial_\mu\psi'\partial^\mu\psi'+\frac{1}{2}\left[\mu^2+\frac{\lambda}{2}\varphi^2\right]\psi'^2+\frac{\lambda\hbar^{1/2}}{6}\varphi\psi'^3+\frac{\lambda}{4!}\psi'^4\right]\ . \tag{2.95}$$

We thus obtain a new theory expanded about a classical background solution with mass

$$m^2=u^2+\tfrac{1}{2}\lambda\varphi^2 \tag{2.96}$$

and interactions

$$\frac{\lambda}{6}\hbar^{1/2}\varphi\psi'^3+\frac{\lambda}{4!}\hbar\psi'^4\ . \tag{2.97}$$

In the new theory the propagator is independent of $\hbar$, while both vertices carry positive powers of $\hbar$. Since to a given order of perturbation theory there are only a finite number of Feynman graphs, and since the expansion to fixed order p in $\hbar$ will involve only graphs with fewer than $2p$ vertices, we have demonstrated our initial claim.

The background-field method also yields an easy way to determine the one-loop effective potential. From the definition as a Legendre transform we have

$$\Gamma_1(0,\varphi)=W_1(0,\varphi)\ , \tag{2.98}$$

where subscripts denote the order in the loop expansion. Thus $\exp[i\Gamma_1(0,\varphi)]$ is just the Gaussian approximation of Eq. (2.87) about the background field φ. By Eq. (2.95), therefore,

$$\Gamma_1(\varphi)=-\tfrac{1}{2}\ln\det[(-\partial_\mu\partial^\mu+\mu^2+\tfrac{1}{2}\lambda\varphi^2)\mathbb{1}]\times\det{}^{-1}[(-\partial_\mu\partial^\mu+\mu^2)\mathbb{1}]\ . \tag{2.99}$$

The functional integral approach yields a closed-form answer for the one-loop effective action. If φ is a constant field we can evaluate $\Gamma_1(\varphi)$ in momentum space and obtain

$$\Gamma_1(\varphi)=-\tfrac{1}{2}\int d^4x\int\frac{d^4k}{(2\pi)^4}\ln\left[\frac{k^2+\mu^2+\frac{1}{2}\lambda\varphi^2}{k^2+\mu^2}\right]\ . \tag{2.100}$$

For $\mu=0$ we immediately obtain our original result Eq.

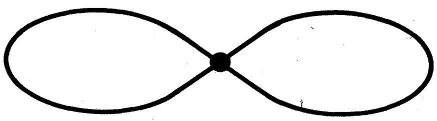


FIG. 5. The two-loop graph included in the Hartree-Fock approximation.

(2.61) for $V_1(\varphi)$.

The one-loop approximation to the effective potential is just one of many conceivable approximation schemes. It is the semiclassical approximation. Other approximation schemes are discussed in Jackiw (1975) and in textbooks on quantum-field-theoretical methods in statistical physics (e.g., Fetter and Walecka, 1971). Two popular methods are the Hartree-Fock approximation (see, for example, Jackiw, 1975), which amounts to improving beyond one-loop order by [in the language of the modified action of Eq. (2.94)] including the graph of Fig. 5, and the Rayleigh-Ritz method. The Hartree-Fock approximation is not systematic in λ, since it does not include the graph of Fig. 6, but it gives the leading large-N contributions if λ scales as N^{-1} and φ as $N^{1/2}$ (Dolan and Jackiw, 1974; Schnitzer, 1974; Coleman, Jackiw, and Politzer, 1974).

The Rayleigh-Ritz method is an approximation scheme based not on a graphical expansion, but rather on the variational definition of the effective potential. The idea is to give a multiparameter family of state wave functionals obeying the constraints of Eq. (2.47) and to minimize the energy expectation value in this subclass of states. The minimum will be an approximation for the effective potential.

D. Two examples

To illustrate the formalism established in the previous sections and to make contact with the recent literature on new scenarios in cosmology inspired by modern unified gauge field theories, we shall explicitly calculate the one-loop effective potential for two important models, the Glashow (1961) -Weinberg (1967) -Salam (1969) model of weak and electromagnetic interactions and the minimal SU(5) Georgi-Glashow (1974) model. We shall consider both models in the Coleman-Weinberg mode, i.e., without an explicit Higgs mass term in $U(\varphi)$.

Before analyzing the models, it is useful to obtain an estimate for the order of magnitude of the three one-loop correction terms for $V_{\text{eff}}(\bar{\varphi})$. For a Higgs potential $V(\varphi)=\lambda P(\varphi)$ with $P(\varphi)$ a fourth-order polynomial, we can use Eqs. (2.77), (2.83), and (2.86) to conclude that

$$\begin{aligned} V_S^{(1)}(\bar{\varphi}) &\sim (\lambda\bar{\varphi}^2)^2 \ln \lambda\bar{\varphi}^2 , \\ V_f^{(1)}(\bar{\varphi}) &\sim (\Gamma\bar{\varphi})^4 \ln(\Gamma\bar{\varphi})^2 , \\ V_g^{(1)}(\bar{\varphi}) &\sim (g\bar{\varphi})^4 \ln(g\bar{\varphi})^2 . \end{aligned} \tag{2.101}$$

If the minimum of the effective potential occurs at $\bar{\varphi}=\sigma$, then $\Gamma\sigma$ will be the order of magnitude of the fermion masses [see Eq. (2.69)]. From Eq. (2.85) it follows that the heavy gauge particle masses will be of order $g\sigma$. If our theories contain no heavy unobserved fermions, then $\Gamma \ll g$ and $V_f^{(1)}(\bar{\varphi})$ can be neglected. In most gauge theories we can take λ to be small. In that case the position σ at which the minimum is taken on is determined by balancing $\lambda\bar{\varphi}^4$ against $V_g^{(1)}(\bar{\varphi})$. Below we shall show that λ must be of the order g^4. Therefore, $V_S^{(1)}(\bar{\varphi})$ is negligible. Hence for a wide class of models, which includes our two examples, the one-loop effective potential is given by

$$V_{\text{eff}}^{(1)}(\bar{\varphi}) = U(\bar{\varphi}) + V_g^{(1)}(\bar{\varphi}) . \tag{2.102}$$

In these models the one-loop result predicts spontaneous symmetry breaking. The effective potential is again given by Fig. 2. In contrast to a scalar field theory, symmetry breaking still persists at higher orders of the loop expansion. The reason for the difference is that for gauge theories the effective loop-expansion parameter is very small for field values φ near the minimizing point σ. To prove this assertion, we first solve for λ in terms of σ [this step is called dimensional transmutation (Coleman and Weinberg, 1973)]. For notational simplicity we consider theories with a single Higgs field. After regularization and renormalization following Eqs. (2.62)—(2.67), Eqs. (2.86) and (2.102) become

$$V_{\text{eff}}^{(1)}(\bar{\varphi}) = \frac{\lambda}{4!}\bar{\varphi}^4 + \frac{3c}{64\pi^2}(g\bar{\varphi})^4 \left[\ln\frac{\bar{\varphi}^2}{M^2} - \frac{25}{6} \right] , \tag{2.103}$$

where c is a constant of order unity arising from performing the traces in Eq. (2.86). The natural renormalization point is $M=\sigma$. Then the condition for σ to be an extremum of $V_{\text{eff}}^{(1)}(\bar{\varphi})$ is

$$V_{\text{eff}}'^{(1)}(\sigma) = \left[\frac{\lambda}{6} - \frac{11g^4}{16\pi^2}c \right] \sigma^3 = 0 . \tag{2.104}$$

Hence

$$\lambda = \frac{33}{8\pi^2} g^4 c . \tag{2.105}$$

Each loop will introduce another factor $g^2\ln(\varphi/\sigma)$. (This is the reason we call it the effective loop-expansion parameter.) Since it is very small for field values φ close to σ, the loop expansion is indeed reliable for this range of field values.

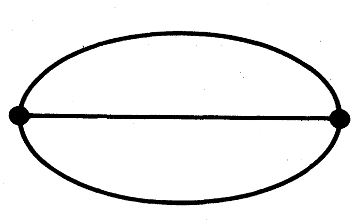


FIG. 6. A two-loop graph that does not contribute in the Hartree-Fock approximation.

The gauge group G of the Glashow-Weinberg-Salam model is $SU(2)\times U(1)$. Let g and g' denote the coupling constants of the two factors and W^α ($\alpha=1,2,3$) and V the associated gauge fields. The scalar fields form a complex doublet,

$$\varphi=\begin{bmatrix}\varphi^+\\ \varphi_0\end{bmatrix}, \tag{2.106}$$

on which SU(2) acts via its defining representation, while the U(1) is an overall phase rotation. The Higgs vacuum expectation value is

$$\langle\varphi\rangle=\begin{bmatrix}0\\ a\end{bmatrix}. \tag{2.107}$$

Since

$$D_\mu\langle\varphi\rangle=\frac{ia}{2}\begin{bmatrix}gW_\mu^1-igW_\mu^2\\ -gW_\mu^3+g'V_\mu\end{bmatrix} \tag{2.108}$$

the U(1) subgroup generated by the generator associated with

$$\frac{g'W_\mu^3+gV_\mu}{(g'^2+g^2)^{1/2}}\equiv\sin\theta_W W_\mu^3+\cos\theta_W V_\mu\equiv A_\mu \tag{2.109}$$

leaves $\langle\varphi\rangle$ invariant and thus remains unbroken. A_μ is the electromagnetic field, θ_W the Weinberg angle. In terms of A_μ, $W_\mu^{1,2}$, and $Z_\mu\equiv\cos\theta_W W_\mu^3-\sin\theta_W V_\mu$, the gauge boson mass matrix in the broken phase is already diagonal:

$$\overline{D_\mu\langle\varphi\rangle}D^\mu\langle\varphi\rangle=\frac{a^2g^2}{4}(W_\mu^{1^2}+W_\mu^{2^2})+\frac{a^2(g^2+g'^2)}{4}Z_\mu^2\ , \tag{2.110}$$

which implies

$$m^2(W_{1,2})=\tfrac{1}{4}g^2a^2\ ,\qquad m^2(Z)=\tfrac{1}{4}(g^2+g'^2)a^2\ . \tag{2.111}$$

Thus for $\bar\varphi=\bar\varphi\langle\varphi\rangle$ the contribution of the gauge field loops to the one-loop effective potential is

$$V_g^{(1)}(\bar\varphi)=\frac{3}{1064\pi^2}[2g^4+(g^2+g'^2)^2]\bar\varphi^4\ln\bar\varphi^2+\mathcal{O}(\bar\varphi^4)\ . \tag{2.112}$$

In terms of the measurable quantities e and θ_W defined by

$$g=\frac{e}{\sin\theta_W}\quad\text{and}\quad g'=\frac{e}{\cos\theta_W}\ , \tag{2.113}$$

Eq. (2.112) becomes

$$V_g^{(1)}(\bar\varphi)=\frac{3}{64}\left[\frac{e^2}{4\pi}\right]^2\left[\frac{2+\cos^{-4}\theta_W}{\sin^4\theta_W}\right]\bar\varphi^4\ln\bar\varphi^2+\mathcal{O}(\bar\varphi^4)\ . \tag{2.114}$$

Regularization and renormalization proceed exactly as in the scalar field theory case [see Eqs. (2.62)–(2.67)]. Thus our final result is

$$V_g^{(1)}(\bar\varphi)=\frac{3}{64}\left[\frac{e^2}{4\pi}\right]^2\left[\frac{2+\cos^{-4}\theta_W}{\sin^4\theta_W}\right]\times\bar\varphi^4\left[\ln\frac{\bar\varphi^2}{M^2}-\frac{25}{6}\right]. \tag{2.115}$$

In the Georgi-Glashow minimal SU(5) model the Higgs field is in the adjoint representation. For most values of the free parameters, the energetically favored symmetry-breaking channel is (see Billoire and Tamvakis, 1982, and Guth and Weinberg, 1980)

$$SU(5)\rightarrow SU(3)\times SU(2)\times U(1)\ . \tag{2.116}$$

A Higgs vacuum expectation value of the form

$$\varphi=\varphi\,\mathrm{diag}(1,1,1,-\tfrac{3}{2},-\tfrac{3}{2}) \tag{2.117}$$

will break SU(5) in the required way. For field configurations of this form, the effective one-loop potential can easily be calculated using the adjoint representation formula

$$M_{\alpha\beta}^2=g^2\mathrm{Tr}([\tau_\alpha,\varphi],[\tau_\beta,\varphi])\ . \tag{2.118}$$

The 12 gauge bosons associated with the residual symmetry remain massless, the others pick up identical masses,

$$M_U^2=\tfrac{25}{8}g^2\bar\varphi^2\ . \tag{2.119}$$

Inserting this result into Eq. (2.86) we find

$$V_g^{(1)}(\bar\varphi)=\frac{5625}{1024\pi^2}g^4\bar\varphi^4\ln\bar\varphi^2+\mathcal{O}(\bar\varphi^4)\ , \tag{2.120}$$

which after regularization and renormalization yields

$$V_g^{(1)}(\bar\varphi)=\frac{5625}{1024\pi^2}g^4\bar\varphi^4\left[\ln\frac{\bar\varphi^2}{M^2}-\frac{25}{6}\right]. \tag{2.121}$$

This completes our discussion of the zero-temperature effective potential in sample field theory models.

E. Comments

In this section we comment briefly on two questions of importance in inflationary universe models—first, the extension of the formalism to composite operators, and second, the question of convexity.

In many systems symmetry forces the expectation value of the scalar field to vanish, i.e., $\bar\varphi=0$. In this case the ordinary effective action $\Gamma(\bar\varphi)$ is uninteresting. Instead we should like to compute the minimal energy of states with a fixed two-point function $G(\mathbf{x},\mathbf{y})$. The generalization of the effective action formalism to cover this case has been worked out by Cornwall, Jackiw, and Tomboulis (1974).

We start by considering the vacuum persistance amplitude in the presence of two source terms, the original $J(x)\varphi(x)$ and a new term $\frac{1}{2}\varphi(x)\varphi(y)K(x,y)$,

$$Z(J,K)=N\int[d\varphi]\exp\left[i\int d^4x\{\mathscr{L}[\varphi(x)]+J(x)\varphi(x)+\tfrac{1}{2}K(x,y)\varphi(x)\varphi(y)\}\right]. \quad (2.122)$$

$W(J,K)$ is defined as usual by

$$Z(J,K)=\exp[iW(J,K)]. \quad (2.123)$$

The generalized effective action $\Gamma(\overline{\varphi},G)$ is the double Legendre transform of $W(J,K)$:

$$\Gamma(\varphi,G)=W(J,K)-\int d^4x\,J(x)\overline{\varphi}(x)-\tfrac{1}{2}\int d^4x\,d^4y[\overline{\varphi}(x)\overline{\varphi}(y)+G(x,y)]K(x,y) \quad (2.124)$$

with $J(x)$ and $K(x,y)$ determined by

$$\frac{\delta W(J,K)}{\delta J(x)}=\overline{\varphi}(x),\qquad \frac{\delta W(J,K)}{\delta K(x,y)}=\tfrac{1}{2}[\overline{\varphi}(x)\overline{\varphi}(y)+G(x,y)]. \quad (2.125)$$

In the absence of sources the equations of motion are

$$\frac{\delta\Gamma(\overline{\varphi},G)}{\delta\overline{\varphi}(x)}=0,\quad \frac{\delta\Gamma(\overline{\varphi},G)}{\delta G(x,y)}=0. \quad (2.126)$$

If we restrict our attention to static configurations, we obtain

$$\Gamma(\overline{\varphi},G)=-E(\overline{\varphi},G)\int dt, \quad (2.127)$$

where $E(\overline{\varphi},G)$ is the minimum of the energy when varying over all normalized states with the constraints

$$\langle\varphi(\mathbf{x})\rangle=\overline{\varphi}(\mathbf{x})$$

and

$$\langle\varphi(\mathbf{x})\varphi(\mathbf{y})\rangle=\overline{\varphi}(\mathbf{x})\overline{\varphi}(\mathbf{y})+G(\mathbf{x},\mathbf{y}). \quad (2.128)$$

This statement is clearly a generalization of Theorem 2. For homogeneous states $[\overline{\varphi}(x)=\text{const},\ G(x,y)=G(x-y)]$ we can write

$$E(\overline{\varphi},G)=V(\overline{\varphi},G)\int d^3x. \quad (2.129)$$

$V(\varphi,G)$ is the generalization of the effective potential.

Hawking and Moss (1983) and Vilenkin (1983) have used this formalism to study the evolution of the quantum field in inflationary universe models.

The convexity properties of the effective potential have caused some confusion in the literature. It is well known (see, for example, Arnold, 1978) that the Legendre transform of any function, whenever defined, is convex. Thus, as emphasized already by Symanzik (1970), the effective potential computed nonperturbatively will be convex. This is true in particular for the double-well potential of Fig. 7. The Maxwell construction gives an easy intuitive understanding for the flat segment of the effective action for $-\sigma<\varphi<\sigma$. Given the value $\varphi=\alpha\sigma+(1-\alpha)(-\sigma)$ for $0\le\alpha\le1$, we consider a state $|\Omega\rangle=\alpha^{1/2}|\Omega_1\rangle+(1-\alpha)^{1/2}|\Omega_2\rangle$ where $|\Omega_1\rangle$ ($|\Omega_2\rangle$) is the state with $\varphi=\sigma$ ($\varphi=-\sigma$). Then the expectation value of the quantum field operator φ in the state $|\Omega\rangle$ is the given classical value φ, and the energy of the state vanishes. Note that the effective action given in curve (b) of Fig. 7 is nonanalytic.

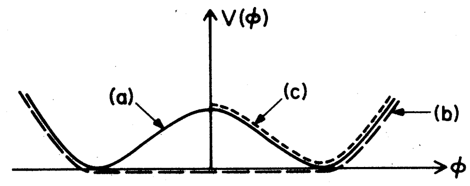


FIG. 7. Double-well potential and its effective potential: (a) tree-level potential; (b) effective potential by the Maxwell construction; (c) perturbative result.

The effective potential we computed to one-loop order in the previous sections is obviously nonconvex. Since a perturbative calculation always yields an analytic function, the best we can hope to obtain is the analytic continuation of the nontrivial branch of the true effective potential. It was pointed out by Langer (1967,1969) that the analytic continuation is important in the analysis of metastability of statistical mechanical systems.

Also in quantum field theory the analytic continuation of the nontrivial branch of the exact effective potential, i.e., the function we approximated by the one-loop effective potential, plays an important role for questions concerning the existence and decay of metastable and unstable states. From the background-field approach to the effective potential discussed earlier it is obvious that the one-loop effective potential $V^{(1)}(\overline{\varphi})$ gives the free-energy density of the state consisting of Gaussian fluctuations with prescribed mass about $\varphi(x)=\overline{\varphi}$. This has been stressed by Hill (1983) and more recently by Mazenko *et al.* (1984). If $V^{(1)}(\overline{\varphi})$ has a local minimum at the origin, it is possible to prepare a metastable state. If $V^{(1)}(\overline{\varphi})$ has a local maximum at the origin, then the equation of motion for a scalar field with potential $V^{(1)}(\overline{\varphi})$ will approximate the evolution of the expectation value of a state initially set up to be localized at some $\overline{\varphi}$ close to the origin. These are the circumstances under which $V^{(1)}(\overline{\varphi})$ is relevant for cosmology.

Whether the initial state in a cosmological model is localized in the above sense or not is a completely separate question. In all models of the type of the new inflationary universe it is assumed that the initial state will be localized. Mazenko *et al.* (1984) have recently given arguments to the contrary. In the opinion of the author this crucial question is still unresolved. Work towards resolving the issue is in progress.

III. FINITE-TEMPERATURE QUANTUM FIELD THEORY

A. Formalism and modified Feynman rules

Conventional quantum field theory is set up to describe scattering events that take place in empty space, i.e., in a surrounding vacuum. By the Lehmann-Symanzik-Zimmermann (LSZ) formalism the scattering matrix elements are expressed in terms of (zero-temperature) Green's functions, i.e., vacuum expectation values of time-ordered strings of Heisenberg field operators (see, for example, Itzykson and Zuber, 1980). The assumption of an empty space in which a scattering event takes place is very well justified when studying particle interactions in accelerators, but in the early stages of the universe standard cosmology predicts a high matter and radiation density which renders the assumption totally inapplicable. In analogy with thermodynamics, the background state in which we study a scattering event should be a thermal bath at the temperature T of the universe. Replacing the vacuum by this thermal bath, we obtain the definition of finite-temperature Green's functions for a scalar field φ:

$$\begin{aligned} G_n^\beta(x_1,\ldots,x_n) &= \sum e^{-\beta E(\psi)} \\ &\quad \times \langle \psi \mid T[\varphi(x_1)\cdots\varphi(x_n)] \mid \psi \rangle N \\ &= \frac{\mathrm{Tr}\{e^{-\beta H} T[\varphi(x_1)\cdots\varphi(x_n)]\}}{\mathrm{Tr}e^{-\beta H}} . \end{aligned} \tag{3.1}$$

The first sum runs over a complete set of states $|\psi\rangle$ with energies $E(\psi)$ and thus can be written as a trace in the space of states.

Equation (3.1) can be interpreted in statistical mechanics language. It is the average value of the time-ordered product of the n field operators in the grand canonical ensemble with vanishing chemical potential. Thus there is a close formal relation between methods of finite-temperature quantum field theory and quantum statistical mechanics.

The use of Green's functions in quantum statistical mechanics goes back to the work of Matsubara (1955). The correct boundary conditions for Bose and Fermi systems were derived by Kubo (1957). Martin and Schwinger (1959) made important contributions, as did many Russian physicists (e.g., Abrikosov, Gorkov, and Dzyaloshinski, 1959; Fradkin, 1959; Akhiezer and Peletminskii, 1960). Quantum field theory methods in statistical mechanics are also discussed in at least three textbooks (Fetter and Walecka, 1971; Kadanoff and Baym, 1962; Abrikosov, Gorkov, and Dzyaloshinski, 1965).

Finite-temperature effects in quantum field theory and their implications for cosmology were first considered by Kirzhnits (1972) and Kirzhnits and Linde (1972,1974). The formalism of finite-temperature effective potentials on which we base our discussion was developed by Weinberg (1974), Bernard (1974), and Dolan and Jackiw (1974). Summaries of the early work on phase transitions in gauge theories and cosmology have been published by Kirzhnits and Linde (1975) and Linde (1979).

As emphasized by Bernard (1974), in gauge theories the finite-temperature formalism is not gauge invariant. In particular, the finite-temperature effective potential is not gauge invariant. The reason is obvious: only in physical gauges (gauges in which there are no spurious particles such as ghosts) does the naive trace correspond to a summation over all physical states of the system. The partition function must clearly be defined in a physical gauge. The functional integral for it can be written down in this gauge. The functional integral can then be extended to an arbitrary gauge by gauge invariance. Thus the finite-temperature Feynman rules in any gauge may be determined. We shall not need any technical details on this point in later sections and therefore shall not go into them.

In analogy with the zero-temperature case we shall derive the perturbation expansion for the finite-temperature Green's functions in terms of finite-temperature Feynman rules using the functional formalism. The finite-temperature generating functional is

$$Z^\beta(J) = \frac{\mathrm{Tr}\left[e^{-\beta H} T \exp\left[i \int J(x)\varphi(x) d^4x\right]\right]}{\mathrm{Tr}e^{-\beta H}} . \tag{3.2}$$

The important observation here is that in Minkowski space-time, as compared to the zero-temperature case, the only changes in the functional integral are the boundary conditions on the set of paths on which the measure has its support. Recall that whenever the coupling in the term in the Hamiltonian quadratic in momenta is field independent, the transition amplitude between two Heisenberg states ψ_1 and ψ_2 is given by (see, for example, Abers and Lee, 1973)

$${}_H\langle \psi_1 | \psi_2 \rangle_H = \int [d\varphi(s,\mathbf{x})] \psi_1^*[\varphi(t_f,\mathbf{x}_f)] \psi_2[\varphi(t_i,\mathbf{x}_i)] \exp\left[i \int_{t_i}^{t_f} ds \int d^3x \, \mathscr{L}(\varphi,\partial_\mu\varphi)\right] . \tag{3.3}$$

In the limit $t_i \to -\infty$, $t_f \to +\infty$ the state wave functionals $\psi_1^*(\varphi)$ and $\psi_2(\varphi)$ give the boundary conditions on the functional integral. It is now obvious that the difference between zero temperature and finite temperature is merely a difference in boundary conditions. The product of vacuum-state wave functionals

$$\Omega^*[\varphi(t_f,\mathbf{x}_f)]\Omega[\varphi(t_i,\mathbf{x}_i)] \tag{3.4}$$

must be replaced by

$$\sum e^{-\beta E(\psi)} \psi^*[\varphi(t_f,\mathbf{x}_f)]\psi[\varphi(t_i,\mathbf{x}_i)] . \tag{3.5}$$

The change in boundary conditions has a remarkable interpretation in Euclidean space: The finite-temperature Green's functions are periodic (for Bose fields) or antiperiodic (for Fermi fields) in Euclidean time with period β. Thus the paths contributing to the Euclidean functional integral must likewise be periodic or antiperiodic. We can derive these results by considering the two-point functions. Let φ stand for an arbitrary field and set $s=0$ (1) if φ is bosonic (fermionic).

The first step in deriving the periodicity results is to extend the definition of time ordering to Euclidean times $t\in[0,-i\beta]$. Since analytic continuation to Minkowski space must reproduce the usual definition, we define

$$T[\varphi(x)\varphi(y)]=\begin{cases}\varphi(x)\varphi(y) & \text{if } ix_0>iy_0\\ (-1)^s\varphi(y)\varphi(x) & \text{if } ix_0<iy_0\ .\end{cases} \tag{3.6}$$

The two-point function

$$G_2^\beta(x-y)=\frac{\mathrm{Tr}\{e^{-\beta H}T[\varphi(x)\varphi(y)]\}}{\mathrm{Tr}e^{-\beta H}} \tag{3.7}$$

can be transformed using cyclicity of the trace and field transformation properties under the Poincaré group:

$$\begin{aligned}(\mathrm{Tr}e^{-\beta H})G_2^\beta(x-y)\,|_{x^0=0}&=(-1)^s\mathrm{Tr}[e^{-\beta H}\varphi(y_0,\mathbf{y})\varphi(0,\mathbf{x})]\\ &=(-1)^s\mathrm{Tr}[e^{-\beta H}e^{\beta H}\varphi(0,\mathbf{x})e^{-\beta H}\varphi(y_0,\mathbf{y})]\\ &=(-1)^s\mathrm{Tr}[e^{-\beta H}\varphi(-i\beta,\mathbf{x})\varphi(y_0,\mathbf{y})]\\ &=(-1)^s(\mathrm{Tr}e^{-\beta H})G_2^\beta(x-y)\,|_{x^0=-i\beta}\ .\end{aligned} \tag{3.8}$$

Hence we have verified the claim

$$G_2^\beta(x-y)\,|_{x^0=0}=(-1)^sG_2^\beta(x-y)\,|_{x^0=-i\beta}\ . \tag{3.9}$$

If the generating functional of a theory is given by an unconstrained functional integral of the form

$$Z(J)=\int[d\varphi]e^{iS(\varphi,J)}N\ , \tag{3.10}$$

then we can see, by inspecting $S(\varphi,J)$, that the Feynman rules follow immediately as we discussed in the proof of Theorem 1. As in Eq. (2.12) the Feynman propagator is iA^{-1}, where A is the operator coupling the fields in the term of S quadratic in φ. The vertex factors are i times the coefficients of the interaction terms in the action and are thus obviously unaffected by modifications of the boundary conditions. Since A is in general a differential operator in position space, its inverse in momentum space will be a local operator and hence invariant under temperature changes. This means the momentum-space propagator will remain unchanged when introducing a finite temperature. The propagator in position space, on the other hand, will depend on the boundary conditions.

Periodicity in Euclidean time direction implies discretization of momentum space in the ik_0 direction. The possible k_0 values are given by $-i\beta\omega_n=\pi 2n$ for Bose fields, $-i\beta\omega_n=\pi(2n+1)$ for Fermi fields. This leads to changes in the Feynman rules summarized in Table II.

B. Finite-temperature effective potential

In Sec. II we have demonstrated that in a large class of gauge theories with a Higgs potential $U(\varphi)$ of the Coleman-Weinberg type, i.e., without an explicit mass term $\frac{1}{2}\mu^2\varphi^2$ and with minimum at $\varphi=0$, radiative one-loop corrections destabilize the minimum. A new nonsymmetric minimum at $\varphi=\sigma$ appears, and hence spontaneous symmetry breaking occurs. In this section we shall investigate the effects of finite temperature in these models. To this end we must compute the finite-temperature one-loop effective potential $V_{\mathrm{eff}}^{(1)\beta}(\bar{\varphi})$.

Consider first the scalar field theory [Eq. (2.59)]. Since the only change compared to the zero-temperature case is the discretization of the k_0 variable, the identical steps that led to Eq. (2.61) will give

TABLE II. Comparison of zero and finite temperature Feynman rules.

	Zero temperature	Finite temperature
loop integral	$\int\frac{d^4k}{(2\pi)^4}$	$\frac{1}{-i\beta}\sum_{n=-\infty}^{\infty}\int\frac{d^3k}{(2\pi)^3}$
vertex δ functions	$(2\pi)^4\delta^{(4)}\left[\sum k_i\right]$	$\frac{\beta}{i}(2\pi)^3\delta_{\sum\omega_i}\delta^{(3)}\left[\sum \mathbf{k}_i\right]$

$$V_{\rm eff}^{(1)\beta}(\bar\varphi)=U(\bar\varphi)-\frac{i}{2}\frac{1}{-i\beta}\sum_{n=-\infty}^{\infty}\int\frac{d^3k}{(2\pi)^3}\ln\left[1-\frac{\lambda\bar\varphi^2}{2(\omega_n^2-\mathbf{k}^2)}\right]$$
$$=U(\bar\varphi)+\frac{1}{2\beta}\sum_{n=-\infty}^{\infty}\int\frac{d^3k}{(2\pi)^3}\ln\left[1+\frac{\lambda\bar\varphi^2}{2\left[\frac{4\pi^2}{\beta^2}n^2+\mathbf{k}^2\right]}\right]$$
$$=U(\bar\varphi)+\frac{1}{2\beta}\int\frac{d^3k}{(2\pi)^3}\sum_{n=-\infty}^{\infty}\ln\left[\frac{4\pi^2}{\beta^2}n^2+E_{\mathbf{k}}^2\right]+\text{const} \tag{3.11}$$

with

$$E_{\mathbf{k}}^2=\mathbf{k}^2+\frac{\lambda\bar\varphi^2}{2} \tag{3.12}$$

(we have omitted all counterterms). The sum over n is divergent, but the infinite part is a constant. The finite part, which contains the only E dependence, can be calculated (see Dolan and Jackiw, 1974) by first differentiating with respect to E, summing the resulting series using the identity

$$\sum_{n=1}^{\infty}\frac{y}{y^2+n^2}=-\frac{1}{2y}+\frac{\pi}{2}\coth\pi y$$
$$=-\frac{1}{2y}+\frac{\pi}{2}+\frac{\pi e^{-2\pi y}}{1-e^{-2\pi y}}\,, \tag{3.13}$$

and integrating the resulting function. We get

$$V_{\rm eff}^{(1)\beta}(\bar\varphi)=U(\bar\varphi)+\int\frac{d^3k}{(2\pi)^3}\left[\frac{E_{\mathbf{k}}}{2}+\frac{1}{\beta}\ln(1-e^{-\beta E_{\mathbf{k}}})\right]. \tag{3.14}$$

Now

$$i\int_{-\infty}^{\infty}\frac{dx}{2\pi}\ln(-x^2+y^2-i\varepsilon)=y+\text{const}\,. \tag{3.15}$$

We can check Eq. (3.15) by the same trick of differentiating with respect to y inside the integral, performing the x integral, and at the end integrating with respect to y. Applying Eq. (3.15) to the first integrand in Eq. (3.14), we obtain

$$V_{\rm eff}^{(1)\beta}(\bar\varphi)=U(\bar\varphi)+\frac{i}{2}\int\frac{d^4k}{(2\pi)^4}\ln(-k_0^2+E_k^2-i\varepsilon)$$
$$+\frac{1}{\beta}\int\frac{d^3k}{(2\pi)^3}\ln(1-e^{-\beta E_{\mathbf{k}}})\,. \tag{3.16}$$

Rotating the second term to Euclidean space we find that, up to a constant,

$$V_{\rm eff}^{(1)}(\bar\varphi)=U(\bar\varphi)+\frac{1}{2}\int\frac{d^4k_E}{(2\pi)^4}\ln\left[1+\frac{\lambda\bar\varphi^2}{2k_E^2}\right]$$
$$+\frac{1}{2\pi^2\beta}\int_{\infty}^{\infty}dk\,k^2\ln(1-e^{-\beta E_{\mathbf{k}}})\,. \tag{3.17}$$

The first two terms are exactly the zero-temperature result. We conclude therefore that introducing a finite temperature gives an extra term

$$\Delta V_{\rm eff}^{(1)\beta}(\bar\varphi)=\frac{1}{2\pi^2\beta^4}\int_0^{\infty}x^2\ln\left\{1-\exp\left[-\left[x^2+\frac{\lambda\bar\varphi^2}{2}\beta^2\right]^{1/2}\right]\right\}dx \tag{3.18}$$

in the one-loop effective potential.

It is very important that this extra term is finite, since as a consequence the zero-temperature renormalization counterterms will also render the nonzero-temperature one-loop effective potential finite. Thus even on the level of renormalized potentials the relation

$$V_{\rm eff}^{(1)\beta}(\bar\varphi)=V_{\rm eff}^{(1)T=0}(\bar\varphi)+\frac{1}{2\pi^2\beta^4}I(y) \tag{3.19}$$

with

$$I(y)=\int_0^{\infty}x^2\ln(1-e^{-(x^2+y^2)^{1/2}})dx \tag{3.20}$$

and

$$y^2=\frac{\lambda}{2}\bar\varphi^2\beta^2 \tag{3.21}$$

remains valid.

The generalization to non-Abelian gauge theories is straightforward. Since the sum of all one-loop graphs is given by Eq. (3.12) with $\lambda\bar\varphi^2/2$ replaced by the effective coupling constant of the loop particle to the external scalar field, we obtain the finite-temperature correction to the one-loop effective potential by summing Eq. (3.18) over all loop particles (in the diagonal mass basis described in Sec. II.C).

In the Glashow-Weinberg-Salam model, therefore,

$$\Delta V_{\rm eff}^{(1)\beta}(\bar\varphi)=\frac{3}{2\pi^2\beta^4}\{2I(\tfrac{1}{2}g\bar\varphi\beta)+I[\tfrac{1}{2}(g^2+g'^2)^{1/2}\bar\varphi\beta]\}\,, \tag{3.22}$$

and in the minimal SU(5) Georgi-Glashow model

$$\Delta V_{\rm eff}^{(1)\beta}(\bar\varphi)=\frac{18}{\pi^2\beta^4}I(\sqrt{25/8}g\bar\varphi\beta) \tag{3.23}$$

[see Eqs. (2.111) and (2.119). Also recall that both Eqs. (3.22) and (3.23) only hold for a one-parameter family of field configurations given by Eqs. (2.111) and (2.117), respectively].

For high temperatures we can expand $I(y)$ in a power series in y^2 (more precisely, the condition for applicability of the expansion is $\bar\varphi\beta \ll 1$):

$$\begin{aligned} I(y) &= \int_0^\infty x^2 \ln(1-e^{-x})dx + \frac{y^2}{2}\int_0^\infty x\frac{e^{-x}}{1-e^{-x}}dx + \mathcal{O}(y^4) \\ &= -\sum_{n=1}^{\infty}\frac{1}{n}\int_0^\infty x^2 e^{-nx}dx + \frac{y^2}{2}\int_0^\infty \frac{xe^{-x}}{1-e^{-x}}dx + \mathcal{O}(y^4) \\ &= -\frac{\pi^4}{45}+y^2\frac{\pi^2}{12}+\mathcal{O}(y^4)\ . \end{aligned} \tag{3.24}$$

In this approximation the scalar field theory result is [see Eq. (3.19)]

$$V_{\rm eff}^{(1)\beta}(\bar\varphi)=V_{\rm eff}^{(1)T=0}(\bar\varphi)+\frac{\lambda}{48\beta^2}\bar\varphi^2-\frac{\pi^2}{90\beta^4}+\mathcal{O}(\bar\varphi^4)\ . \tag{3.25}$$

In the Glashow-Weinberg-Salam model we get

$$V_{\rm eff}^{(1)\beta}(\bar\varphi)= V_{\rm eff}^{(1)T=0}(\bar\varphi)+\frac{1}{32\beta^2}(3g^2+g'^2)\bar\varphi^2 -\frac{\pi^2}{10\beta^4}+\mathcal{O}(\bar\varphi^4) \tag{3.26}$$

[see Eq. (3.22)], and from Eq. (3.23) we deduce that for the minimal SU(5) model

$$V_{\rm eff}^{(1)\beta}(\bar\varphi)=V_{\rm eff}^{(1)T=0}(\bar\varphi)+\frac{75}{16\beta^2}g^2\bar\varphi^2-\frac{\pi^2}{15\beta^4}+\mathcal{O}(\bar\varphi^4)\ . \tag{3.27}$$

The finite-temperature corrections to the one-loop effective potential in the high-temperature limit give rise to a temperature-dependent mass term. This $T^2\bar\varphi^2$ term converts $\bar\varphi=0$ from a local maximum of the effective potential back to a local minimum. In fact, for high temperatures, the $T^2\bar\varphi^2$ and $\bar\varphi^4$ terms dominate for all values of $\bar\varphi$ and make $\bar\varphi=0$ the absolute minimum, whereas below some critical temperature T_c $\bar\varphi=0$ remains a relative minimum, a metastable false vacuum (see Fig. 8).

The critical temperature T_c is the temperature at which the minima are degenerate. We can roughly estimate T_c by assuming that the location $\bar\varphi=\sigma$ of the asymmetric minimum is temperature independent. In this approximation the criterion for T_c is

$$V_{\rm eff}^{(1)T_c}(\sigma)=0\ . \tag{3.28}$$

To be specific, we shall consider a theory with a purely quartic Higgs potential (e.g., the Glashow-Weinberg-Salam model in the Coleman-Weinberg mode). By Eqs. (2.103) and (3.25)

$$V_{\rm eff}^{(1)T}(\bar\varphi)=\frac{\lambda}{4!}\bar\varphi^4+B\bar\varphi^4\left[\ln\frac{\bar\varphi^2}{M^2}-\frac{25}{6}\right]+C\bar\varphi^2T^2\ , \tag{3.29}$$

with numerical constants B and C determined by the specific model. The natural choice for the renormalization point is $M=\sigma$. Replacing λ by σ via dimensional transmutation as discussed earlier [Eqs. (2.103) and (2.104)], we obtain

$$\lambda=88B \tag{3.30}$$

and hence

$$V_{\rm eff}^{(1)T}(\bar\varphi)=B\bar\varphi^4\left[\ln\frac{\bar\varphi^2}{\sigma^2}-\frac{1}{2}\right]+C\bar\varphi^2T^2\ . \tag{3.31}$$

Equation (3.28) now immediately gives

$$T_c^2=\frac{B\sigma^2}{2C}\ . \tag{3.32}$$

By dimensional analysis alone we immediately could have predicted the order of magnitude of T_c. Since the symmetry-breaking scale σ is the only mass scale in the theory, any critical temperature should be of the same order of magnitude.

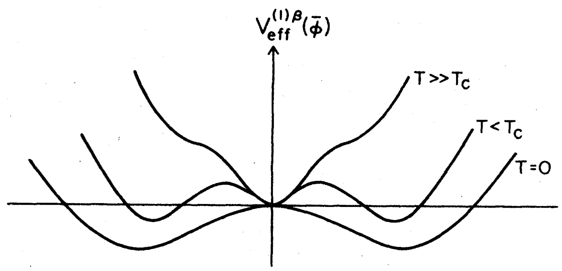


FIG. 8. Finite-temperature effective potential.

C. Cosmological applications

The importance of finite-temperature effects in field theories for cosmology was realized long before the idea of inflation was born. Kirzhnits (1972) and Kirzhnits and Linde (1972) pointed out that at high temperatures the SU(2)×U(1) symmetry of the Weinberg-Salam model is restored due to the effective mass term we discussed in the preceding section. Kirzhnits and Linde (1974) stressed the analogy with superconductivity. In superconductivity the order parameter vanishes at high temperatures. At low temperatures Bose condensates of Cooper pairs give a nonvanishing order parameter. The order parameter corresponds to the vacuum expectation value of the Higgs field in field theory.

Since in standard cosmological models the temperature starts out at infinity at the big bang, phase transitions in the quantum field theory as a function of temperature give rise to phase transitions in matter during the cosmological evolution. If we describe matter by a grand unified field theory there are several important phase transitions (see, for example, Guth, 1982): the grand unification transition at 10^{14} GeV, the Weinberg-Salam transition at 10^2 GeV, the confinement transition at 0.5 GeV, and recombination at 10^{-10} GeV. Each causes significant changes in the description of matter. Guth (1981), however, was the first to appreciate the potential of finite-temperature effects for solving cosmological problems.

A favorite application of finite-temperature field theory before the advent of inflationary universe models was the attempt to derive lower bounds on the mass m_H of the Weinberg-Salam Higgs boson from cosmological arguments. Weinberg (1976) and Linde (1976) derived a lower bound $m_H > 4.9$ GeV by requiring that the symmetry $SU(2)\times U(1)$ be spontaneously breakable, i.e., that the global minimum of the effective potential be at a value $\varphi=\sigma$ corresponding to broken symmetry. This bound holds assuming the universe started out in the symmetric phase. If, on the other hand, the universe started out in the asymmetric phase $\varphi=\sigma$ then, provided the tunneling probability were sufficiently small, it could have remained in this minimum. If this scenario were cosmologically acceptable (in the standard big bang scenario it is not), the upper bound would weaken to $m_H > 450$ MeV (Frampton, 1976; Linde, 1977; Steinhardt, 1980).

In standard cosmological models the lower bound on m_H can be improved by estimating the tunneling probability. The lower bound on the tunneling probability stemming from requiring the phase transition to have taken place by the present time leads to a new bound $m_H > 6.9$ GeV (Linde, 1977). Witten (1981) realized that the Weinberg-Salam phase transition leads to entropy production and hence to a dilution of the baryon-to-entropy ratio (see also Steinhardt, 1981). Guth and Weinberg (1980) translated this fact into a better lower bound $m_H \geq 9$ GeV. Since these topics lie outside of the main line of this paper, we shall not go into the details. We mention them only to give a flavor of applications other than the inflationary scenario in which quantum field theory and cosmology are combined.

IV. DECAY OF THE FALSE VACUUM

A. The problem

As a motivation for this section we shall describe how the problem of vacuum decay arises in cosmological scenarios. Big bang models predict that at very early times the universe was extremely hot and has since then cooled down to the presently observed temperature of 2.7 K. We shall assume matter can be described by a quantum field theory with a symmetry which at zero temperature is spontaneously broken by a scalar field acquiring a nontrivial vacuum expectation value.

Our analysis of the finite-temperature effective potential (see Fig. 8) shows that initially $\bar{\varphi}=0$ is the only ground state of the theory. As time increases, the temperature of the universe will decrease, and at some critical temperature $T=T_c$ the symmetric vacuum will cease to be stable and a new energetically favored ground state appears. Thermal, gravitationally induced, and quantum fluctuations will tend to force the theory into the new asymmetric ground state, the true vacuum. Provided the potential barrier is high compared to the thermal energy and the gravitational energy, quantum fluctuations will dominate. The relevant quantity is the decay rate per unit volume Γ/V of the false vacuum. Its value is crucial if we are interested in the cosmological implications of the models, since the universe will expand exponentially and supercool in the false vacuum for a period $(\Gamma/V)^{-1}$ until the phase transition occurs. These cosmological implications will be discussed in detail in Sec. VI.

Here we shall neglect gravitational and thermal fluctuations and focus on the quantum effects. We treat quantum fluctuations in a semiclassical approximation using functional integral methods. After quantum tunneling, the further evolution of the fields will be determined by solving the classical field equations.

B. Calculation of the decay rate in quantum mechanics

We present the functional integral approach to calculating the decay rate, an approach developed by Coleman (1977) and Callan and Coleman (1977), and simultaneously by Stone (1976,1977) and Frampton (1977) based on previous work by Voloshin, Kobzarev, and Okun (1974). The functional integral approach to problems of metastable states was pioneered by Langer (1967,1969) in condensed matter physics. We closely follow the excellent review article by Coleman (1979).

For notational simplicity let us first consider tunneling in quantum mechanics. Let $\psi(t)$ denote the amplitude for an unstable state with energy $E_0=\alpha+i\beta$. Then the decay probability per unit time Γ is given by

$$|\psi(t)|^2=e^{2\beta t/\hbar}|\psi(0)|^2=e^{-\Gamma t}|\psi(0)|^2 \tag{4.1}$$

so

$$\Gamma=-\frac{2}{\hbar}\,\mathrm{Im}E_0 \ . \tag{4.2}$$

We have thus reduced the problem of calculating the decay rate to one of determining the imaginary part of the energy of the false vacuum.

According to Eq. (4.2) we must calculate the energy E_0

of the false vacuum. In a first step we show that this problem is equivalent to calculating a certain transition matrix element. Consider a theory with Hamiltonian H and potential U. For two position eigenstates $|x_f\rangle$ and $|x_i\rangle$

$$\langle x_f|e^{-HT/\hbar}|x_i\rangle=\sum_n e^{-E_nT/\hbar}\langle x_f|n\rangle\langle n|x_i\rangle\ , \qquad (4.3)$$

where the sum runs over a complete set of energy eigenstates. Consider the subset of states not orthogonal to either $|x_f\rangle$ or $|x_i\rangle$, i.e., with $\langle x_f|n\rangle\neq 0$ and $\langle x_i|n\rangle\neq 0$. Let Ω denote the lowest-energy state in this subset and let E_0 be its energy. Then in the limit $T\to\infty$

$$\langle x_f|e^{-HT/\hbar}|x_i\rangle \underset{T\to\infty}{\longrightarrow} e^{-E_0T/\hbar}\langle x_f|\Omega\rangle\langle\Omega|x_i\rangle\ . \qquad (4.4)$$

Equation (4.4) provides an elegant way of reexpressing the energy E_0 of the false vacuum. We can always choose coordinates such that $x=0$ is the false vacuum. Now we choose $|x_i\rangle=|x_f\rangle=|0\rangle$. Then

$$E_0=-\hbar\lim_{T\to\infty}\frac{1}{T}\ln\langle 0|e^{-HT/\hbar}|0\rangle\ . \qquad (4.5)$$

The second step is to transform Eq. (4.3) into a functional integral. We use the Feynman-Kac formula (see, for example, Glimm and Jaffe, 1981; the Feynman-Kac formula is the Euclidean version of the standard functional integral representation for quantum-mechanical transition amplitudes described, for example, in Abers and Lee, 1973):

$$\langle x_f|e^{-HT/\hbar}|x_i\rangle=N\int[dx]e^{-S_E(x)/\hbar}$$

$$x(T/2)=x_f,\ x(-T/2)=x_i\ . \qquad (4.6)$$

S_E is the Euclidean action

$$S_E(x)=\int_{-T/2}^{T/2}dt\left[\frac{1}{2}\left[\frac{dx}{dt}\right]^2+U[x(t)]\right]\ . \qquad (4.7)$$

We evaluate the functional integral (4.6) in a semiclassical approximation based on the idea that the stationary points of S_E will give the dominating contribution. We therefore perform a Gaussian approximation about each stationary point. The extrema of S_E are given by

$$\ddot{x}=U'(x)\ , \qquad (4.8)$$

with the boundary conditions given in Eq. (4.6):

$$x(T/2)=x_f,\ \ x(-T/2)=x_i\ . \qquad (4.9)$$

Equation (4.8) is the classical equation of motion for a point particle in the inverted potential $-U(x)$. This interpretation will aid in the qualitative discussion of the solutions. Nontrivial solutions of Eq. (4.8) are called instantons.

The Gaussian approximation about an instanton $\bar{x}$ consists of expanding $S(x)$ about $\bar{x}$ and keeping only terms quadratic in the fluctuation $z=x-\bar{x}$. We can easily verify that

$$S_E(x)=S_E(\bar{x})+\tfrac{1}{2}\int_{-T/2}^{T/2}dt\,z(t)\left[-\frac{\partial^2}{\partial t^2}+U''[\bar{x}(t)]\right]z(t)+\mathcal{O}(z^3)\ . \qquad (4.10)$$

If we drop the correction terms of order $\mathcal{O}(z^3)$ the functional integral becomes a simple Gaussian integral and can be performed explicitly, yielding

$$N\int[dx]e^{-S_E(x)/\hbar}\simeq Ne^{-S_E(\bar{x})/\hbar}$$

$$\times\{\det[-\partial_t^2+U''(\bar{x})]\}^{-1/2}\ . \qquad (4.11)$$

In the general case (more than one instanton solution) we must sum (4.11) over all stationary points.

For potentials of physical interest there are indeed solutions of Eq. (4.8) satisfying the boundary conditions (4.9). Consider, for example, the potential of Fig. 9 and choose $x_f=x_i=0$. The trivial solution is $\bar{x}(t)\equiv 0$. For this extremum all eigenvalues of the second variational derivative of S_E are positive. Hence Eq. (4.11) is real and does not contribute to the decay rate [see Eqs. (4.2) and (4.5)]. This is what we expect, since the same solution $\bar{x}\equiv 0$ is the stable ground state for the harmonic-oscillator potential, and in that case $\mathrm{Im}E_0=0$. For $T\to\infty$ there is a nontrivial solution $\bar{x}$ of Eq. (4.8) satisfying boundary conditions (4.9). In fact, this solution is the standard example of an instanton. It corresponds to a particle rolling off the hill $x=0$ of $-U(x)$ at time $T=-\infty$ with zero initial kinetic energy, bouncing off the wall at $x=x^*$ at time t_c, and returning to the top of the hill at $T=\infty$. We can easily determine the Euclidean action of this instanton using Eq. (4.7) and conservation of (particle) energy:

$$E=\tfrac{1}{2}\dot{x}^2-U(x)=0\ . \qquad (4.12)$$

Hence

$$S_E(\bar{x})=\lim_{T\to\infty}\int_{-T/2}^{T/2}dt\left[\frac{d\bar{x}}{dt}\right]^2$$

$$=2\int_0^{x^*}dx\left.\frac{dt}{dx}\right|_{x=\bar{x}}\left[\frac{d\bar{x}}{dt}\right]^2$$

$$=2\int_0^{x^*}(2U)^{1/2}dx\equiv B\ . \qquad (4.13)$$

Obviously the solution $\bar{x}$ is not unique. First, the center t_c of the instanton is arbitrary. Furthermore, since combinations of instantons with centers located far apart are excellent approximate solutions, they should be included as well. This leads us to the dilute-gas approximation.

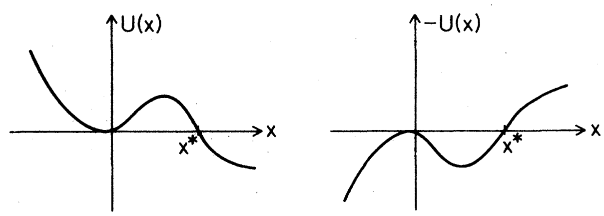


FIG. 9. A potential with an unstable ground state.

C. Dilute-gas approximation

In the dilute-gas approximation we take an approximate solution consisting of n widely separated instantons with centers $t_1 > t_2 > \cdots > t_n$, calculate the contribution of this configuration by taking the fluctuations about the n constituent instantons to be independent, and finally sum over all possible instanton centers as well as over n.

Let $\bar{x}$ denote the n-instanton configuration. $z = x - \bar{x}$ is the fluctuation about $\bar{x}$. As sketched in Fig. 10 we divide the interval $[-T/2, T/2]$ into nonoverlapping regions R_i on which the single instantons are concentrated. R^* is the time region for which $\bar{x} \approx 0$. The fluctuation field on R_i is denoted by z_i, on R^* by z^*. In our approximation we let the fluctuations z_i be totally unconstrained, i.e., we neglect the boundary condition

$$z^* \,|_P = (\bar{x} + z^i)\,|_P \tag{4.14}$$

for all points P at the border of two regions. We also perform a Gaussian approximation in each region. In greater detail,

$$N \int [dx] e^{-S_E(x)/\hbar} = N \int [dz] e^{-S_E(\bar{x}+z)/\hbar} \simeq \left[\prod_{i=1}^{n} \int [dz_i] e^{-S_E(\bar{x}+z_i)/\hbar} \,|_{R_i} \right] N \int [dz^*] e^{-S_E(\bar{x}+z^*)/\hbar} \,|_{R^*} \,. \tag{4.15}$$

In region R^* we are evaluating the fluctuations about $\bar{x}=0$. With

$$\omega^2 = \frac{d^2}{dx^2} U(x)\,|_{x=0} \tag{4.16}$$

Eq. (4.11) yields

$$N \int [dz^*] e^{-S_E(\bar{x}+z)/\hbar} \,|_{R^*} \simeq N \det{}^{-1/2}(-\partial_t^2 + \omega^2) \sim \left[\frac{\omega}{\pi\hbar} \right]^{1/2} e^{-\omega T/2} \tag{4.17}$$

for T large. (The determinant is evaluated in Appendix C.) In regions R_i the problem is the same as in the one-instanton case discussed above. Thus

$$\int [dz_i] e^{-S_E(\bar{x}+z_i)/\hbar} \,|_{R_i} \equiv e^{-B/\hbar} K \,. \tag{4.18}$$

If we incorrectly, as will shortly be revealed, assumed that the determinant in Eq. (4.11) were well defined, then we would have

$$K = \det[-\partial_t^2 + U''(\bar{x})]^{-1/2} \,. \tag{4.19}$$

Integration over all locations of the centers gives a factor

$$\int_{-T/2}^{T/2} dt_1 \int_{-T/2}^{t_1} dt_2 \cdots \int_{-T/2}^{t_{n-1}} dt_n = \frac{T^n}{n!} \,. \tag{4.20}$$

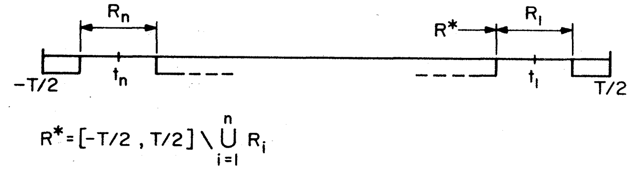


FIG. 10. Division of the time interval for an n-instanton configuration.

Note that a further inaccuracy in this approximation is the use of Eq. (4.15) even when the regions R_i (which have some fixed length) overlap.

We obtain the final result of the dilute-gas approximation by combining Eqs. (4.15)—(4.20):

$$N \int [dx] e^{-S_E(x)/\hbar} \simeq \left[\frac{\omega}{\pi\hbar} \right]^{1/2} e^{-\omega T/2} \sum_{n=0}^{\infty} \frac{T^n}{n!} e^{-nB/\hbar} K^n = \left[\frac{\omega}{\pi\hbar} \right]^{1/2} e^{-T(\omega/2 - K e^{-B/\hbar})} \,. \tag{4.21}$$

By Eq. (4.5) it follows that

$$E_0 = \frac{\hbar\omega}{2} - \hbar K e^{-B/\hbar} \tag{4.22}$$

and by Eq. (4.2) that

$$\Gamma = 2(\mathrm{Im} K) e^{-B/\hbar} \,. \tag{4.23}$$

We have thus reduced the computation of the decay rate to a calculation of the imaginary part of K. K is determined by demanding that Eqs. (4.15)—(4.18) give the correct answer for the one-instanton configuration:

$$K e^{-B/\hbar} \det{}^{-1/2}(-\partial_t^2 + \omega^2) = \int [dx] e^{-S(x)/\hbar} \,|_{\text{one instanton}} \,. \tag{4.24}$$

The basic reason that K has a nonzero imaginary part is that the operator

$$-\frac{\partial^2}{\partial t^2} + U''(\bar{x}) \tag{4.25}$$

which arises in Eq. (4.10) has a negative eigenvalue. Thus a naive Gaussian approximation is impossible. The negative eigenvalue appears for the following reason: Let $\bar{x}$ be the instanton with center $t=0$. Then

$$x(\tau,t)=\bar{x}(t+\tau)=\bar{x}(t)+\tau\frac{d\bar{x}(s)}{ds}\bigg|_{s=t}+\mathscr{O}(\tau^2) \quad (4.26)$$

is a fluctuation of $\bar{x}$ with parameter τ corresponding to shifting the center of the instanton by τ. Since the Euclidean action is independent of the instanton center, the above fluctuation, which we shall denote by $z_0(t)$, is an eigenmode of Eq. (4.25) with zero eigenvalue. But $z_0(t)$ obviously has one node corresponding to the turning point of $\bar{x}$. Therefore, there must be one eigenmode $z_1(t)$ of Eq. (4.25) with negative eigenvalue λ_1. Since the naive Gaussian approximation breaks down because of the existence of fluctuations with nonpositive curvature of $S_E(x)$ at $\bar{x}$, we must treat $z_0(t)$ and $z_1(t)$ separately and perform the naive Gaussian approximation only for the remaining fluctuation modes. By integrating over all possible instanton centers in Eq. (4.15) we have already integrated in the $z_0(t)$ direction (up to normalization). The integral in the direction of the $z_1(t)$ mode must be defined by analytic continuation. (This is no surprise, since we know that the energy of any unstable state can only be defined by analytic continuation.)

The first step is to extract the integration in the unstable direction $z_1(t)$ by splitting the path integral $[dx]$ into the product of a one-dimensional integral over a specific one-parameter family of paths and a path integral over the remaining set of paths. Let λ be our path parameter for the one-parameter family. Following Coleman (1979) we choose the family to contain $x(t)=0$ for $\lambda=0$, to contain the instanton $\bar{x}(t)$ for $\lambda=1$, and to proceed in the direction of the negative mode for $\lambda>1$, i.e.,

$$x_{1+\lambda'}(t)=\bar{x}(t)+\lambda' z_1(t)+\mathscr{O}(\lambda'^2) \; . \quad (4.27)$$

If $z_i(t)$ denotes the eigenmodes of Eq. (4.25) with eigen-

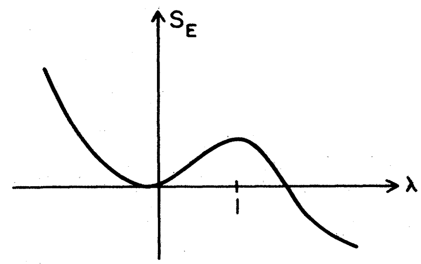


FIG. 11. Euclidean action for the one-parameter family of paths.

values $\lambda_i>0$ $(i=2,3,\ldots)$, then the functional integral appearing in Eqs. (4.6) and (4.11) can be rewritten as

$$\int [dx]=\int_{-\infty}^{\infty} d\lambda \prod_{i=2}^{\infty}\int [dz_i](2\pi\hbar)^{-1/2} \; . \quad (4.28)$$

In the Gaussian approximation the functional integral decouples into an infinite product of one-dimensional Gaussian integrals. Each gives a factor $\lambda_i^{-1/2}$. Therefore,

$$\int [dx]e^{-S_E(x)/\hbar}=\int_{-\infty}^{\infty} d\lambda\, e^{-S_E(x_\lambda)/\hbar} \times\prod_{i=2}^{\infty}\lambda_i^{-1/2}(2\pi\hbar)^{-1/2} \; . \quad (4.29)$$

The action S_E as a function of λ is sketched in Fig. 11. To keep the integral finite, we must distort the path into the complex plane for positive λ as indicated in Fig. 12. [This corresponds to analytically continuing the potential $U(x)$ so as to render $x=0$ the absolute minimum.] After the initial distortion the path integral can be evaluated using the method of steepest descent,

$$\int_{-\infty}^{\infty} d\lambda\, e^{-S_E(x_\lambda)/\hbar}=\int_{-\infty}^{1} d\lambda\, e^{-S_E(x_\lambda)/\hbar}+\int_{\gamma_2} d\lambda\, e^{-\lambda_1(1/2)(\lambda-1)^2/\hbar-B/\hbar}$$
$$=R+\tfrac{1}{2}\,|\lambda_1|^{-1/2} i e^{-B/\hbar}(2\pi\hbar)^{1/2} \; , \quad (4.30)$$

where R is real. The factor $\frac{1}{2}$ arises since we are only integrating over half the Gaussian peak. The imaginary part of Eq. (4.29) thus becomes

$$\tfrac{1}{2}e^{-B/\hbar}\prod_{i=1}^{\infty}|\lambda_i|^{-1/2}\equiv\tfrac{1}{2}e^{-B/\hbar}\det{}'[-\partial_t^2+U''(\bar{x})]^{-1/2} \; . \quad (4.31)$$

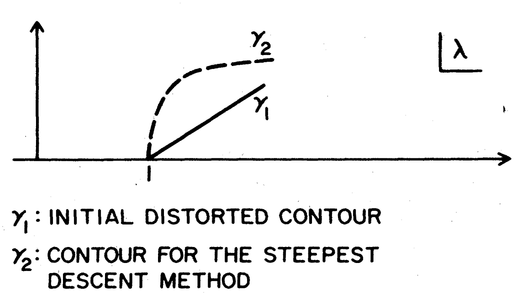


FIG. 12. Distortion of the contour into the complex λ plane. γ_1: initial distorted contour; γ_2: contour for the steepest-descent method.

We now combine Eqs. (4.24), (4.30), and (4.31) to obtain

$$\mathrm{Im}K=\frac{1}{2}\left[\frac{\det(-\partial_t^2+\omega^2)}{\det{}'[-\partial_t^2+U''(\bar{x})]}\right]^{1/2}\left[\frac{B}{2\pi\hbar}\right]^{1/2} . \quad (4.32)$$

Thus by Eq. (4.23)

$$\Gamma=\left[\frac{\det(-\partial_t^2+\omega^2)}{\det{}'[-\partial_t^2+U''(\bar{x})]}\right]^{1/2} e^{-B/\hbar}\left[\frac{B}{2\pi\hbar}\right]^{1/2} . \quad (4.33)$$

D. Decay rate in quantum field theory

The decay rate calculation in quantum field theories is almost identical to the one just described for quantum mechanics. Following Eqs. (4.2)–(4.6) we can express the decay rate as a functional integral,

$$\Gamma = -2 \lim_{T\to\infty} \frac{1}{T} \ln \operatorname{Im}\langle \varphi_f | e^{-HT/\hbar} | \varphi_i \rangle$$
$$= -2 \lim_{T\to\infty} \frac{1}{T} \ln N \int_{\varphi(T/2)=\varphi_f, \varphi(-T/2)=\varphi_i} [d\varphi] e^{-S_E(\varphi)/\hbar} . \quad (4.34)$$

We shall evaluate the functional integral using a semiclassical approximation. The instanton with lowest action S_E will give the leading contribution. To determine the coefficient of the exponential, we shall perform a dilute-gas approximation based on the minimal-action instanton as the fundamental solution.

First consider the case of a scalar quantum field theory with two tree-level minima (Fig. 13).

The Euclidean action is given by

$$S_E(\varphi) = \int d^4x [\tfrac{1}{2}\partial_\mu\varphi\partial_\mu\varphi + U(\varphi)] . \quad (4.35)$$

The boundary conditions are:

$$\lim_{\tau\to\pm\infty} \varphi(\tau,\mathbf{x}) = \varphi_- ,$$
$$\lim_{|\mathbf{x}|\to\infty} \varphi(\tau,\mathbf{x}) = \varphi_- . \quad (4.36)$$

The first line is the field theory version of Eq. (4.9). It ensures that the lowest-energy state in the sector contributing to (4.34) is the false vacuum [at least to first order in $\hbar$; see Eq. (4.41) and the comments made at that point]. The second line is the condition of finite action: Only solutions with finite action give a nonvanishing contribution to Eq. (4.34) in a semiclassical approximation.

The equation of motion for the instantons, i.e., the condition that S_E be stationary, is

$$\partial_\mu\partial_\mu\varphi = U'(\varphi) . \quad (4.37)$$

Since we are only interested in the lowest-action instanton, we can use the following theorem to reduce the problem to one of one degree of freedom.

Theorem 3 (Coleman, Glaser, and Martin, 1979). If an $O(4)$ invariant solution of Eq. (4.37) exists, its action S_E will be lower than that of any $O(4)$ noninvariant solution.

Thus we can restrict our attention to $O(4)$ invariant solutions. In terms of $\rho = (x_\mu x_\mu)^{1/2} = (|\mathbf{x}|^2 + \tau^2)^{1/2}$, Eq. (4.37) becomes

$$\frac{d^2\varphi}{d\rho^2} + \frac{3}{\rho}\frac{d\varphi}{d\rho} = U'(\varphi), \quad \rho \geq 0 \quad (4.38)$$

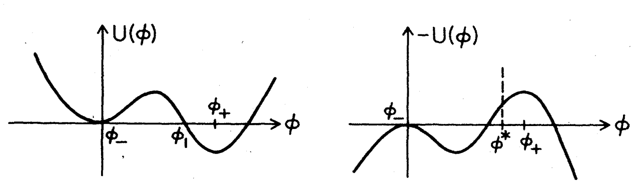


FIG. 13. Sample potential with an unstable vacuum. φ_- is the false vaccum, φ_+ the true vacuum.

with boundary condition

$$\lim_{\rho\to\infty} \varphi(\rho) = \varphi_- . \quad (4.39)$$

Equation (4.38) is the equation of motion for a classical point particle in the potential $-U$, subject to a time-dependent damping force $(3/\rho)d\varphi/d\rho$ [ρ plays the role of time in Eq. (4.38)].

We can easily argue by continuity that a solution of Eq. (4.38) satisfying boundary conditions (4.39) must exist. We are basing the following discussion on the notation in Fig. 13. Let the particle start at time $\rho=0$ with zero kinetic energy from some position φ^* with $\varphi_- < \varphi^* < \varphi_+$. If $\varphi^* < \varphi_1$, then by energy conservation the particle can never reach φ_-. If $\varphi^* \simeq \varphi_+$ then the particle will reach φ_- in finite time with nonzero kinetic energy and will overshoot. To see this in slightly greater detail, we observe that for $\varphi^* \simeq \varphi_+$ the particle will remain close to φ_+ for a long time. During this period, the damping will decrease and gradually become negligible. Hence again, by energy conservation, the particle will overshoot. By continuity there will be an intermediate φ^* for which the solution of Eq. (4.38) with initial velocity zero arrives at φ_- after infinite time with zero kinetic energy.

The dilute-gas approximation based on the minimal-action $O(4)$ invariant solution $\bar{\varphi}$ is almost identical to the approximation of Sec. IV.C for quantum mechanics. The main change is that we can translate the instanton center in all four space-time directions. Hence integration over all instanton centers for the n-instanton configuration gives a factor $(TV)^n/n!$. Therefore, the factor T in Eq. (4.21) must be replaced by TV. The extra factor V persists in Eq. (4.23). Denoting $B = S_E(\bar{\varphi})$, we obtain

$$\Gamma = 2(\operatorname{Im}K) e^{-B/\hbar} V . \quad (4.40)$$

As expected, it is only the decay rate per unit volume, the physical quantity of interest, that will be finite. By Eq. (4.33)

$$\frac{\Gamma}{V} = c e^{-B/\hbar} \left[\frac{\det[-d^2 + U''(\varphi_+)]}{\det'[-d^2 + U''(\bar{\varphi})]} \right]^{1/2} , \quad (4.41)$$

where $d^2 = \partial_\mu\partial_\mu$, c is some constant, and det′ is defined as in Eq. (4.31) by taking the product of the absolute values of the eigenvalues and omitting those corresponding to translation modes.

Up to this point the analysis has been to lowest order in $\hbar$ only. To higher orders we must replace the potential $U(\varphi)$ by the effective potential $V_{\text{eff}}(\varphi)$. The reason why the analysis up to this point was only an approximation is that for Eq. (4.34) to give the false vacuum energy it is necessary to introduce a source term which forces the false vacuum to be stable for some initial time interval. The correct procedure therefore is to consider Eq. (4.6) as the limit of the generating functional $Z(J)$:

$$\langle 0 | e^{-HT/\hbar} | 0 \rangle = \lim_{J\to 0} e^{-W(J)/\hbar}$$
$$\overset{\hbar\to 0}{\longrightarrow} \lim_{J\to 0} \int [d\bar{\varphi}] e^{-[\Gamma(\bar{\varphi})+J\bar{\varphi}]/\hbar}$$
$$= \int [d\bar{\varphi}] e^{-\Gamma(\bar{\varphi})/\hbar} . \qquad (4.42)$$

We can expand $\Gamma(\bar{\varphi})$ in powers of momentum. In position space this expansion is

$$\Gamma(\bar{\varphi}) = \int d^4x [-V_{\rm eff}(\bar{\varphi}) + \tfrac{1}{2}\partial_\mu\bar{\varphi}\partial^\mu\bar{\varphi} Z(\bar{\varphi}) + \mathcal{O}(\partial_\mu\bar{\varphi})^4] . \qquad (4.43)$$

Since we can renormalize the field to set $Z(\bar{\varphi})=1$, we have indeed verified that the potential $U(\varphi)$ must be replaced by the effective potential $V_{\rm eff}(\bar{\varphi})$.

Since for constant $\bar{\varphi}$ only the first term survives, the above definition of the effective potential agrees with the previous one (2.18). From a physical point of view, changing U to $V_{\rm eff}$ in the equations of motion means including quantum effects (e.g., the back reaction of gauge particles and fermions on the Higgs field) in determining the evolution of the Higgs field.

For theories in which the scalar field φ couples to other quantum fields there is another way to see how the effective potential naturally emerges as the correct potential to consider in tunneling problems. To take a specific example, consider an Abelian gauge theory. The transition amplitude between two specified Higgs field configurations $|\varphi_i\rangle$ and $|\varphi_f\rangle$ and asymptotically vanishing gauge field configurations is given by

$$I = \langle \varphi_f, A=0 | e^{-HT\hbar^{-1}} | \varphi_i, A=0 \rangle$$
$$= N \int_{\substack{\varphi(\infty)=\varphi_f \\ \varphi(-\infty)=\varphi_i \\ A(\pm\infty)=0}} [d\varphi][dA] \exp[-\hbar^{-1} S_E(\varphi, A)] \qquad (4.44)$$

with Euclidean action

$$S_E(\varphi, A) = \int d^4x [D_\mu\varphi D^\mu\varphi + V(\varphi) + F_{\mu\nu}F^{\mu\nu}] . \qquad (4.45)$$

We now integrate out the gauge fields to obtain an effective action for the Higgs field alone. In the case of an Abelian gauge theory the functional integral over gauge configurations is Gaussian and can be performed explicitly. We obtain

$$S_{\rm eff}(\varphi) = \int d^4x [\partial_\mu\varphi\partial^\mu\varphi + \hat{V}(\varphi)] . \qquad (4.46)$$

To lowest nontrivial order in $\hbar$ and neglecting nonrenormalizable terms, the same gauge loop graphs which determine the effective potential (see Sec. II) contribute to $\hat{V}(\varphi)$. Clearly there are other graphs that contribute to the effective potential but, as we argued in Sec. II, they will be subdominant.

The above analysis also shows that the potential relevant for tunneling calculations in gauge field theories is not the exact effective potential, but rather the perturbatively computed potential.

E. Bubble nucleation

So far we have calculated only the decay rate of the false vacuum. In our WKB approximation we shall describe the evolution of the field φ after tunneling by the classical equations of motion.

Consider first the decay of a false ground state in quantum mechanics (see Fig. 9). In the semiclassical description, the classical particle will make a quantum jump at some time $t=0$ from the local minimum of $U(x)$ to the escape point $x=\sigma$ characterized by equal potential and zero kinetic energy. For $t>0$ the particle propagates classically.

A similar analysis holds for quantum field theory. At some time $t=0$ the classical field will make a quantum jump to a state with zero kinetic energy and hence potential energy equal to that of the false vacuum. Such a state is the midpoint of the instanton $\bar{\varphi}$ (Fig. 14). Thus

$$\varphi(x_0=0,\mathbf{x}) = \bar{\varphi}(\mathbf{x},\tau=0) ,$$
$$\frac{\partial}{\partial t}\varphi(x_0=0,\mathbf{x}) = \frac{\partial}{\partial\tau}\bar{\varphi}(\mathbf{x},\tau=0) = 0 . \qquad (4.47)$$

For $t>0$ the field will evolve according to the classical equations with initial conditions given by Eq. (4.47). The Minkowski space equation is

$$\partial_\mu\partial^\mu\varphi = -U'(\varphi) . \qquad (4.48)$$

In Euclidean space, Eq. (4.48) becomes

$$\partial_\mu\partial_\mu\varphi = U'(\varphi) , \qquad (4.49)$$

which is identical to the instanton equation (4.37). A solution $\varphi(t,\bar{x})$ for $t>0$ describes the growth of a bubble of the true vacuum in a surrounding sea of false vacuum. In order to show this, we consider first $\rho^2=\tau^2+\mathbf{x}^2 = \mathbf{x}^2 - t^2$. In terms of ρ, Eq. (4.49) is our old $O(4)$ invariant instanton equation (4.38),

$$\frac{\partial^2\varphi}{\partial\rho^2} + \frac{3}{\rho}\frac{\partial\varphi}{\partial\rho} = U'(\varphi) . \qquad (4.50)$$

Since the initial conditions (4.39) and (4.47) agree as well, the classical field in Euclidean space is exactly the instanton

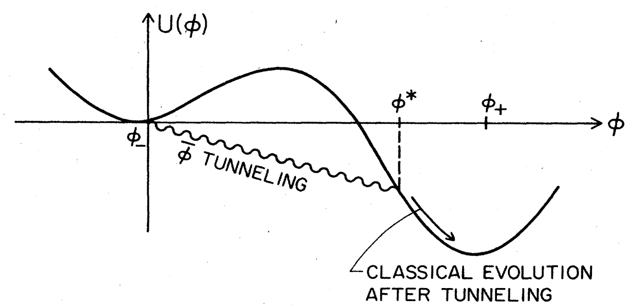


FIG. 14. Vacuum decay for quantum field theory.

$$\varphi(\mathbf{x},\tau)=\bar{\varphi}[(\mathbf{x}^2+\tau^2)^{1/2}], \quad \mathbf{x}^2+\tau^2>0 . \tag{4.51}$$

$\varphi(t,\mathbf{x})$ in this region is therefore the analytic continuation of the instanton solution back to Minkowski space, i.e.,

$$\varphi(t,\mathbf{x})=\bar{\varphi}[(\mathbf{x}^2-t^2)^{1/2}] \tag{4.52}$$

defines the field for $|\mathbf{x}|>t>0$. To obtain a qualitative picture of the solution, we note (see Fig. 15) that $\varphi(t,\mathbf{x})$ has the value φ^* at $|\mathbf{x}|=t$, and as $|\mathbf{x}|$ increases for fixed t, $\varphi(t,\mathbf{x})$ gradually approaches the false vacuum φ_-.

It is important to stress the dual role in which the instanton appears. On the one hand, it appears as the path that contributes most heavily to the functional integral calculation of the false vacuum decay rate; on the other hand, it reappears as the Euclidean space field configuration after tunneling.

To obtain the field equation for $|\mathbf{x}|<t$, we set $\lambda=i\rho$. In terms of λ, Eq. (4.50) becomes

$$\frac{d^2\varphi}{d\lambda^2}+\frac{3}{\lambda}\frac{d\varphi}{d\lambda}\equiv-U'(\varphi) , \tag{4.53}$$

which is the classical equation of motion for a particle in potential $U(\varphi)$ starting at $\varphi=\varphi^*$ with a time-dependent damping force. A discussion of the solution on a qualitative level is very easy (see Fig. 14): The particle will oscillate about the true vacuum φ_+ with damped amplitude.

To summarize, we have the following field configuration for fixed t:

$$\varphi(t,\mathbf{x})=\begin{cases}\varphi_-, & |\mathbf{x}|\to\infty \quad \text{false vacuum}\\ \varphi^*, & |\mathbf{x}|=t \quad \text{bubble wall}\\ \varphi_+, & |\mathbf{x}|=0 \quad \text{true vacuum (asymptotically for } t\to\infty \text{ only)} .\end{cases} \tag{4.54}$$

This is depicted in Fig. 15. We conclude this section with several remarks.

(1) The semiclassical analysis predicts that bubbles of true vacuum will form via quantum tunneling and expand with the speed of light.

(2) This analysis is independent of any assumptions concerning the ratio between potential barrier height on the one hand and potential difference between true and false vacua on the other. In particular, we have not made the thin-wall approximation (see, for example, Coleman, 1977,1979).

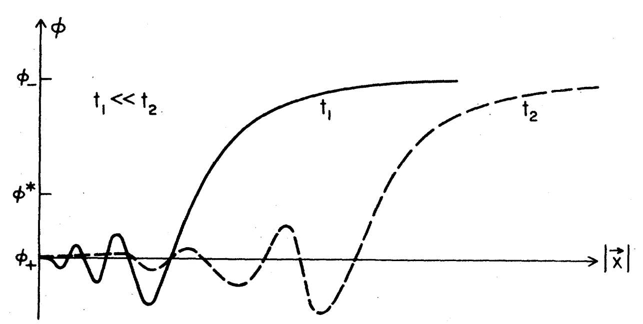


FIG. 15. Qualitative picture of bubble growth.

(3) Since we keep our discussion general, we get no information on the size of the bubbles formed. We note that φ^* is not what is conventionally called the bubble wall. Usually the bubble wall is defined to be the φ value for which $U(\varphi)$ attains its local maximum between φ_- and φ_+. On the other hand, semiclassical approximations are good only asymptotically (see the comments in Coleman, 1979, p. 87). Thus it is not useful to ask questions about measurements immediately after bubble nucleation in a semiclassical approximation.

(4) If we want to include quantum corrections to the classical equation of motion, we must replace the potential $U(\varphi)$ in Eq. (4.48) by the effective potential $V_{\text{eff}}(\varphi)$. Including these corrections leads to faster decay of the amplitude of oscillation. In addition, it provides the framework for calculating baryon generation and the back reaction of these baryons on the evolution of the Higgs field (see, for example, Abbott, Farhi, and Wise, 1982).

F. Finite-temperature and gravitational effects

Cosmological applications of vacuum decay have in the past years been considered by many authors, including Linde (1977,1981a), Guth and Weinberg (1980), Steinhardt (1980,1981a,1981b), Sher (1981), Cook and Mahanthappa (1981), Billoire and Tamvakis (1982), Tamvakis and Vayonakis (1982), Hawking, Moss, and Stewart (1982), and Izawa and Sato (1982).

In cosmology phase transitions occur at high temperature. As pointed out by Linde (1977,1981b,1983b), there are some minor changes in the formalism of vacuum decay. At nonzero temperatures vacuum decay is determined by the shape of the finite-temperature effective potential. In a semiclassical analysis based on the functional integral approach, we must determine the finite-temperature instantons. At zero temperature the lowest-energy instantons had $O(4)$ symmetry. Due to the periodicity in Euclidean time at nonzero temperature, the finite-temperature instantons can no longer have this symmetry. On a heuristic level it is easy to guess the form of the finite-temperature instantons: at low temperatures (T^{-1} is much greater than the radius of zero-temperature instanton) the field configuration will be a sequence of zero-temperature instantons with centers separated by T^{-1} in the Euclidean time direction. As the temperature increases the boundary energies will become more and more significant, and for sufficiently high temperatures the minimal energy field configuration will be constant in Euclidean time and have $O(3)$ symmetry (in space). Its action will therefore be

$$S_4(\varphi)=T^{-1}S_3(\varphi) , \tag{4.55}$$

where $S_3(\varphi)$ is the action of the instanton φ restricted to three space. The instanton equation is

$$\frac{d^2\varphi}{dr^2}+\frac{2}{r}\frac{d\varphi}{dr}=V(\varphi,T) . \tag{4.56}$$

It is important to stress that in contrast to the zero-temperature case there is no theorem on which to base the conjecture that the minimal energy configuration indeed has $O(3)$ symmetry.

Linde (1982d) has also investigated the effect of finite temperature on the coupling constants of gauge theories, a further way in which finite temperature can influence the vacuum decay.

In cosmology, curvature of space-time will be important at the time of grand unification phase transitions. The decay of a metastable state in curved space-time has been studied by Coleman and DeLuccia (1980) and Hawking and Moss (1982), using semiclassical techniques [see also the work by Shore (1980), Ford and Toms (1982), Denardo and Spallucci (1982), and Allen (1983)]. Coleman and DeLuccia (1980) consider the action

$$S=\int d^4x\sqrt{-g}\,[\tfrac{1}{2}g^{\mu\nu}\partial_\mu\varphi\partial_\nu\varphi-V(\varphi)-(16\pi G)^{-1}R] \tag{4.57}$$

and look for minimal-action stationary points of its analytic continuation to Euclidean space. We again make the reasonable but unproven assumption that the minimal-action solutions will have $O(4)$ symmetry. In this case the Euclidean equations of motion reduce to

$$\varphi''+\frac{3\rho'}{\rho}\varphi'=\frac{dV}{d\varphi} \tag{4.58}$$

and

$$\rho'^2=1+\frac{8\pi G}{3}\rho^2(\tfrac{1}{2}\varphi'^2-V)\ , \tag{4.59}$$

where the prime denotes $d/d\xi$ and ξ is the Euclidean radial coordinate. Coleman and DeLuccia construct an explicit solution in the thin-wall approximation (small energy-density difference between the two vacua). They show that in the case in which a state with positive cosmological constant decays into Minkowski space-time, the bubble action decreases and the radius of the bubble at the time of materialization diminishes compared to the flat space-time result.

Hawking and Moss (1982) pointed out that for sufficiently large curvature the radius of the bubble solution would exceed the de Sitter radius H^{-1} (H is the Hubble constant). In this case the only Euclidean solution of the instanton equations, apart from $\varphi=0$, is the homogeneous solution $\varphi=\varphi_1$, where φ_1 is the value of the local maximum of the potential. We interpret this solution as homogeneous tunneling of a horizon volume of space from $\varphi=0$ to $\varphi=\varphi_1$. φ_1 is unstable, and therefore after tunneling the field φ will classically move towards the global minimum (see Fig. 14). The Hawking-Moss scenario provides a realization for an inflationary universe (see Sec. VI). Related work has been done by Mottola and Lapedes (1983) and Abbott and Burges (1983).

Other work on implications of curvature for vacuum decay has been done by Linde (1982b), Hut and Klinkhamer (1981), Abbott (1981b), Parke (1983), and Vilenkin (1983b).

As pointed out by Lapedes and Mottola (1982; see also Patrasciou, 1981, and Shepard, 1983), one may be able to improve on the semiclassical analysis summarized above by taking complex instantons into account. The mathematical reason for this phenomenon was explained by Balian, Parisi, and Voros (1978): there exist functions whose asymptotic series converge to a numerical value which differs significantly from the exact value. The discrepancies are explained as contributions from singularities in the complex plane closer to the origin than the closest singularity along the real axis. For details we refer the reader to the papers cited above.

V. HAWKING RADIATION

A. Bogoliubov transformations

Standard quantum field theory, quantum field theory in Minkowski space-time, is based on the special theory of relativity. It crucially uses the existence of a distinguished inertial Lorentz frame. The entire second quantization prescription is based on this frame, which determines the space-time slicing and the definition of positive and negative frequency. Consider the theory of a free scalar field. The classical field $\varphi(\mathbf{x},t)$ satisfies the Klein-Gordon equation

$$(\Box+m^2)\varphi(\mathbf{x},t)=0\ . \tag{5.1}$$

The solution can be expanded into positive and negative frequency modes proportional to e^{+iwt} and e^{-iwt}, respectively,

$$\varphi(\mathbf{x},t)=e^{\pm iwt}\varphi_w(\mathbf{x})\ . \tag{5.2}$$

We next determine a complete set $\varphi_{wi}(\mathbf{x})$ of solutions of Eq. (5.1) for φ_w:

$$(\nabla^2+w^2-m^2)\varphi_w(\mathbf{x})=0\ . \tag{5.3}$$

Then the most general solution of the classical field equation (5.1) can be expanded as

$$\varphi^{\rm cl}(\mathbf{x},t)=\sum_{i,w}[a_{wi}f_{wi}(\mathbf{x},t)+a'_{wi}f^*_{wi}(\mathbf{x},t)] \tag{5.4}$$

with

$$f_{wi}(\mathbf{x},t)=e^{iwt}\varphi_{wi}(\mathbf{x})\ . \tag{5.5}$$

On the level of classical field theory, a_{wi} and a'_{wi} are numerical constants. The summation over w represents integration.

The quantum theory is obtained by the method of second quantization, by replacing the constants a_{wi} and a'_{wi} by operators a_{wi} and $a^\dagger_{wi}$. Here a_{wi} is the annihilation operator for a particle with frequency ω in mode i. The quantum field $\varphi^q(\mathbf{x},t)$ can be expanded as

$$\varphi^q(\mathbf{x},t)=\sum_{i,w}[a_{wi}f_w(\mathbf{x},t)+a^\dagger_{wi}f^*_w(\mathbf{x},t)]\ . \tag{5.6}$$

The vacuum state $|\Omega\rangle$ is defined by

$$a_{wi} \mid \Omega \rangle = 0, \quad \forall (w,i) \ . \tag{5.7}$$

Cosmology is based on the general theory of relativity. General coordinate invariance has dramatic effects if we attempt to construct a quantum field theory for matter in curved space-time. There is no distinguished Lorentz frame and hence no unique definition of the vacuum state. The notions of vacuum and particle number become observer dependent. An observer O picks a coordinate frame (more precisely, there is a physical prescription which defines an observer frame at every space-time point). In this frame, modes that are positive and negative frequency at the observer point may be defined in analogy to Eq. (5.2) (see below). In particular, O can define a vacuum state, the state that O sees as containing no particles, as the state which is annihilated by all annihilation operators of the mode expansion analogous to Eq. (5.6). Since the frames of two observers O_1 and O_2 at different points of the space-time manifold will in general be different, the positive frequency modes of O_1 will be nontrivial combinations of positive and negative frequency modes of O_2. Therefore, the vacuum of O_1 will contain particles from the point of view of O_2. This is the basic idea behind Parker particle production (Parker, 1966,1969).

The quantitative analysis proceeds as follows:

$$(\Box_g + m^2)\varphi(\mathbf{x},t) = 0 \tag{5.8}$$

is the equation of motion of a scalar field in a space-time manifold with metric g. In this equation, $\Box_g$ is the Laplace operator of g, expressed in terms of the coordinates $(\mathbf{x},t)$ of the observer O_1, situated at the origin of the coordinate system.

Only in static space-times or in asymptotically flat regions of a nonstatic universe is the definition of positive- and negative-frequency solutions straightforward. In these cases the positive- and negative-frequency modes are defined as having the time dependence of Eq (5.2). t is the parameter along the timelike Killing vector field. This definition may be generalized to Robertson-Walker metrics in the following way: we expand $\varphi(\mathbf{x},t)$ in terms of a basis of the spatial Laplace equation. For a flat universe this can be a Fourier decomposition in space or an expansion in terms of spherical harmonics and solutions of the remaining radial equation (an example for the second procedure is worked out in Sec. V.B). Inserting the expansion into Eq. (5.8) we obtain a "temporal equation," an equation for the time dependence of the expansion coefficients. A solution is called positive frequency at time t_0 if its phase is described by $\exp(iwt)$ for t near t_0. We denote the solutions defined to be positive and negative frequency at t_0 by $f_k(\mathbf{x},t)$ and $f_k^*(\mathbf{x},t)$ (k is a formal summation index). The expansion of the quantum field $\varphi(\mathbf{x},t)$ in terms of the basis of O_1 is

$$\varphi(\mathbf{x},t) = \sum_k [a_k f_k(\mathbf{x},t) + a_k^\dagger f_k^*(\mathbf{x},t)] \ . \tag{5.9}$$

The modes f_k are normalized with respect to the conserved scalar product

$$\langle f,g \rangle = \frac{1}{2i} \int_{t=\text{const}} d^3\mathbf{x}(\bar{f}\dot{g} - \dot{\bar{f}}g) \ . \tag{5.10}$$

A second observer O_2 at a different point in space-time will choose a different coordinate system and find a second set g_l and g_l^* of positive and negative frequency modes. $\varphi(\mathbf{x},t)$ can be expanded in this second basis:

$$\varphi(\mathbf{x},t) = \sum_l (b_l g_l + b_l^\dagger g_l^*) \ , \tag{5.11}$$

where b_l and $b_l^\dagger$ are annihilation and creation operators of particles in mode g_l, from the point of view of observer O_2.

The modes g_l, positive frequency from the point of view of O_2, will in general be nontrivial linear combinations of the modes f_k and f_k^* (using matrix notation):

$$g_l = \sum (\alpha_{lk} f_k + \beta_{lk} f_k^*) = (\alpha f + \beta f^*)_l \ . \tag{5.12}$$

The constants α_{lk} and β_{lk} are called Bogoliubov coefficients. They can be determined by taking inner products:

$$\alpha_{lk} = (f_k, g_l) \ , \qquad \beta_{lk} = -(f_k^*, g_l) \tag{5.13}$$

(negative frequency modes have negative norm). The Bogoliubov coefficients form a unitary matrix M

$$\begin{bmatrix} g \\ g^* \end{bmatrix} = M \begin{bmatrix} f \\ f^* \end{bmatrix}, \quad M = \begin{bmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{bmatrix} , \qquad \alpha\alpha^* - \beta\beta^* = 1 \ , \tag{5.14}$$

unitary since both bases are orthonormal. From Eq. (5.14), we immediately obtain the transformation of creation and annihilation operators,

$$(a \quad a^\dagger) = (b \quad b^\dagger) M \ . \tag{5.15}$$

If any of the β_{lk} coefficients are nonvanishing there is particle production in the following sense: observer O_2 will measure a nonzero particle number in the vacuum state prepared by O_1. The number operator for mode k of observer O_2 is

$$N_k = b_k^\dagger b_k \ . \tag{5.16}$$

The expectation value of N_k in the vacuum of O_1 becomes

$$\langle 0 \mid N_k \mid 0 \rangle = \sum_l \mid \beta_{kl} \mid^2 \ . \tag{5.17}$$

In a special case we obtain a thermal flux of particles. The condition is

$$\mid \alpha_{kl} \mid^2 = e^{\omega_k T_H^{-1}} \mid \beta_{kl} \mid^2, \quad \forall k,l \ , \tag{5.18}$$

where ω_k is the energy of the kth mode. In this case unitarity of M immediately gives

$$\langle 0 \mid N_k \mid 0 \rangle = (e^{T_H^{-1}\omega_k} - 1)^{-1} \ , \tag{5.19}$$

i.e., a blackbody spectrum at temperature T_H. In this case we speak of Hawking radiation (Hawking, 1975).

T_H is the Hawking temperature.

The first detailed investigation of particle creation in expanding universes is due to Parker (1966,1968, 1969,1977). He introduced the Bogoliubov mode-mixing technique and observed that particles are always produced in pairs. Other early work on the subject was done by Sexl and Urbantke (1967,1969), Zeldovich (1970), and Zeldovich and Starobinskii (1971). The last-mentioned paper investigated particle production in anisotropic universes as a means to isotropize an initially very anisotropic universe, a problem that was later analyzed in more detail using path-integral techniques by Fischetti, Hu, and Hartle (1979) and Hartle and Hu (1979,1980). Interest in quantum field theory in curved space-time increased dramatically after Hawking's discovery (1974,1975) that the spectrum of particles emitted from a collapsing black hole is thermal and could possibly explain the large entropy-to-baryon ratio. This discovery sparked interest in the analysis of the Hawking effect in other space-times. In particular, Davies (1975) and Unruh (1976), basing their work on earlier work by Fulling (1973), noted that a uniformly accelerated observer in flat space-time detects Hawking radiation in the usual vacuum state of an observer at rest. The mode-mixing analysis was also carried out for de Sitter space (Lapedes, 1978) and the de Sitter phase of an inflationary universe model (Brandenberger and Kahn, 1982).

Parallel to the analysis of particle production in specific space-times, alternate methods of deriving Hawking radiation were pursued. Hartle and Hawking (1976) gave a path-integral derivation of black hole radiance, which was generalized to de Sitter space (Gibbons and Hawking, 1977) and to the case of interacting scalar fields (Gibbons and Perry, 1976,1978). These functional integral methods are very elegant (see Sec. V.C), but they apply only to space-times that have special symmetries. Formal aspects of quantum field theory in curved space-time, such as the existence of the S matrix and the back reaction of particles on the geometry, have been studied, for example, by Wald (1975,1977,1979).

Complementary to the investigation of particle production in specific space-time manifolds is the further development of quantum field theory in curved space-times. The problem of renormalizing the energy-momentum tensor $T_{\mu\nu}$ is much more serious in curved space-time than in Minkowski space-time. In standard quantum field theory the vacuum expectation value of $T_{\mu\nu}$ is rendered finite (and zero) by normal ordering. Since normal ordering depends on having a fixed set of positive and negative frequency modes, the method is inapplicable in curved space-times. We shall briefly summarize the issues in Sec. V.D.

There are many good review articles on quantum field theory in curved space-time. Parker (1977) emphasizes the mode-mixing technique; DeWitt (1975) stresses in particular boundary effects (Casimir effect). Parker (1979) and Boulware (1979) discuss renormalization in detail. Two more recent review articles are those by Gibbons (1979) and by Sciama *et al.* (1981).

B. Hawking radiation in the de Sitter phase of an inflationary universe

As an example of particular relevance to inflationary universe models we shall analyze Hawking radiation in a de Sitter phase of a FRW cosmology, i.e., in a Robertson-Walker metric with scale factor

$$a(t)=e^{Ht} . \tag{5.20}$$

The analysis is based on the work of Lapedes (1978; see also Brandenberger and Kahn, 1982). This section is more technical than the rest of the paper and can be omitted by readers who are not particularly interested in Hawking radiation.

There are four main steps in the calculation. First, we must pick the state as the vacuum state of some initial frame (frame of O_1). Next, we must determine the observer frame (frame of O_2). The third step is to write down the coordinate transformation from one basis to the other. This is necessary in order to express all mode functions in the same coordinates. Finally, we can determine the Bogoliubov coefficients α_{kl} and β_{kl} as scalar products according to Eq. (5.13). If we want to prove that the radiation of particles is thermal, we must in addition verify condition (5.18).

As initial space-time slicing we take the FRW coordinates. By symmetry this is the only distinguished frame. There is also a good physical argument for taking the initial state as the vacuum in this frame. In inflationary universe models (see Sec. VI), the de Sitter phase is preceded by a hot radiation-dominated FRW era. The state after the big bang will be a thermal state in the coordinate frame. In the initial stages of the de Sitter phase, the thermal energy will red-shift away and the state will approach the vacuum state in the coordinate frame exponentially fast.

The frame of an observer O_2 at some space-time point in the de Sitter phase will be given by the static coordinates. In a more general metric the observer would pick locally static coordinates. This choice is based on an analysis of particle detectors (Unruh, 1976; Gibbons and Hawking, 1977). The observer carries a particular detector, modeled as a box which is Fermi-Walker transported with constant proper size along the world line of O_2. The box contains a Schrödinger particle ψ linearly coupled to the quantum field. An excitation of ψ is interpreted as the detection of a field quantum. The quantum mechanics calculation must be performed in static coordinates, otherwise the spatial eigenmodes of the Hamiltonian become time dependent and the measurement argument breaks down. This dictates the choice of (locally) static coordinates for O_2.

The static de Sitter coordinates $(\tilde{t},\tilde{r},\vartheta,\varphi)$ are given in terms of the FRW coordinates (t,r,ϑ,φ) by

$$\begin{aligned} \tilde{r}&=e^{tH}r , \\ \tilde{t}&=-\frac{1}{2H}\ln[e^{-2tH}-(rH)^2] . \end{aligned} \tag{5.21}$$

These relations are valid in the region $re^{tH} < H^{-1}$ and can easily be obtained by combining the well-known transformation between FRW and global de Sitter coordinates (Hawking and Ellis, 1973) with the transformation between global and static coordinates (Lapedes, 1978). For notational convenience the observer O_2 was placed at the origin of the FRW coordinate system. The general case may be recovered by translating the coordinates. In terms of the static coordinates, the metric (1.1) becomes

$$ds^2 = -(1-\tilde{r}^2H^2)d\tilde{t}^2 + (1-\tilde{r}^2H^2)^{-1}d\tilde{r}^2 + \tilde{r}^2 d\Omega^2 , \tag{5.22}$$

where $d\Omega^2$ is the usual line element on S^2

$$d\Omega^2 = d\vartheta^2 + \sin^2\vartheta\, d\varphi^2 . \tag{5.23}$$

To obtain the positive and negative frequency modes in the first basis, we use the method of separation of variables. We insert the ansatz

$$f_{wlm}(r,t,\vartheta,\varphi) = F_w(t)\tilde{f}_{wl}(r) y_{lm}(\vartheta,\varphi) \tag{5.24}$$

into the Klein-Gordon equation (5.1)

$$(-g^{-1})^{1/2}(\partial_a \sqrt{-g}\, g^{ab}\partial_b f) = 0 . \tag{5.25}$$

The resulting differential equation for $y_{lm}(\vartheta,\varphi)$ is the equation for spherical harmonics. Hence the solutions $y_{lm}(\vartheta,\varphi)$ are labeled by the usual angular momentum quantum numbers. The radial and temporal equations become

$$r^{-2}\partial_r(r^2\partial_r\tilde{f}) - l(l+1)r^{-2}\tilde{f} = -\omega^2\tilde{f} , \tag{5.26}$$

$$\ddot{F}_\omega + 3\dot{a}(t)a^{-1}(t)\dot{F}_\omega = -\omega^2 a^{-2}(t)F_\omega . \tag{5.27}$$

The radial equation (5.26) can be solved by using the standard methods to analyze singular linear differential equations (see, for example, Morse and Feshbach, 1953). The result is

$$\tilde{f}_{wl}(r) = r^l e^{iwr}\varphi(l+1, 2(l+1), -2iwr) , \tag{5.28}$$

where $\varphi(a,c,x)$ is the confluent hypergeometric function (Morse and Feshbach, 1953, p. 551). The temporal equation (5.27) admits both a positive- and a negative-frequency solution for given w. Their precise forms are unimportant for the following calculations.

To obtain the positive- and negative-frequency modes in the basis of O_2, we again apply separation of variables. We must solve Eq. (5.25) in the static de Sitter coordinates. We Fourier transform in t and separate the angular dependence, which as in the first basis can be decomposed into spherical harmonics. The remaining radial equation has three singular points, at $r=0$ and $r=\pm H^{-1}$. The solutions that are regular at these points can be determined by the Puiseuz method (Bieberbach, 1953). We obtain

$$g_{wlm}(\tilde{t},\tilde{r},\vartheta,\varphi) = \exp(-iw\tilde{u})\tilde{r}^{l}\hat{g}_{wl}(\tilde{r})y_{lm}(\vartheta,\varphi) \tag{5.29}$$

with

$$\tilde{u} = \tilde{t} - (2H)^{-1}\ln\left[\frac{1-\tilde{r}H}{1+\tilde{r}H}\right] . \tag{5.30}$$

$\hat{g}_{wl}(r)$ is the regular part of the solution of the radial equation. It satisfies a differential equation in which w enters linearly only in purely imaginary terms. Hence

$$\hat{g}^*_{wl}(r) = \hat{g}_{-wl}(r) . \tag{5.31}$$

The Bogoliubov coefficients (5.13) are determined by taking inner products on the $t=0$ hypersurface (the result is hypersurface independent). Since the inner product is obviously diagonal in the angular momentum quantum numbers, the only nonvanishing coefficients are

$$\alpha_{ww'lm} = (f_{w'lm}, g_{wlm}) ,$$
$$\beta_{ww'lm} = -(f^*_{w'lm}, g_{wlm}) = -\alpha^*_{-ww'lm} . \tag{5.32}$$

The final identity follows from Eq. (5.31). To prove that the radiation is thermal, it is thus sufficient to evaluate $\alpha_{ww'lm}$ and investigate the effect of changing w to $-w$.

After inserting the coordinate transformation (5.21) into the modes (5.29) and evaluating at $t=0$ we obtain

$$\dot{g}_{wlm}\,|_{t=0} = \left[lH - \frac{iw}{1-rH}\right] g_{wlm}\,|_{t=0} . \tag{5.33}$$

Therefore,

$$\alpha_{ww'lm} = -\tfrac{1}{2}\int_0^\infty dr\, r^{2(l+1)}e^{-iw'r}\varphi^*(l+1,2(l+1),-2iw'r)\hat{g}_{wl}(r)$$
$$\times(1-rH)^{iwH^{-1}}\left[F^*_{w'}(0)\left[ilH + \frac{w}{1-rH}\right] - i\dot{F}^*_{w'}(0)\right]$$
$$\equiv -\tfrac{1}{2}[F^*_{w'}(0)ilH - i\dot{F}^*_{w'}(0)]I_1 - \tfrac{1}{2}F^*_{w'}(0)I_2 . \tag{5.34}$$

The main idea of the actual computation is to write

$$I_1(w) = i^v I_1'(w) \quad \text{with } v = iwH^{-1} \tag{5.35}$$

and to show that to leading order in H^{-1}

$$I_1'(-w) = I_1'^*(w) . \tag{5.36}$$

After rotating the contour of integration to the imaginary r axis, I_1 becomes

$$I_1(w) = -i^{2l+3}\int_0^\infty dr\, r^{2(l+1)}e^{-w'r}(1+irH)^v\varphi^*(l+1,2(l+1),-2w'r)\hat{g}_{wl}(ir)$$
$$= -i^{2l+3}i^vH^ve^{-iw'H^{-1}}\int_{\gamma^0} ds(s+iH^{-1})^{2(l+1)}e^{-w's}s^v$$
$$\times\varphi^*(l+1,2(l+1),-2w'(s+iH^{-1}))\hat{g}_{wl}(is-H^{-1}) \ . \quad (5.37)$$

γ^0 is the positive real axis translated by $-iH^{-1}$. To leading order in H^{-1} this becomes

$$I_1(w) = -i^{2l+3}i^vH^v\int_0^\infty ds\, s^{2(l+1)}e^{-w's}s^v\varphi^*(l+1,2(l+1),-2w's)\hat{g}_{wl}(is)$$
$$\equiv -i^{2l+3}i^vI_1'(w) \ . \quad (5.38)$$

Finally, since φ^* is a real function and by Eq. (5.31)

$$I_1'(-w) = H^{-v}\int_0^\infty ds\, s^{2(l+1)}e^{-w's}s^{-v}\varphi^*(l+1,2(l+1),-2w's)g_{wl}^*(-is)$$
$$= I_1'^*(w) \ . \quad (5.39)$$

Similarly,

$$I_2(w) = -i^{2l+3}i^vI_2'(w) \quad (5.40)$$

with $I_2'(-w) = I_2'^*(w)$. Hence

$$|\alpha_{ww'lm}|^2 = i^{4v}|\alpha_{-ww'lm}|^2 = e^{2\pi H^{-1}w}|\alpha_{-ww'lm}|^2 \quad (5.41)$$

and thus by the thermality condition (5.18)

$$T_H = \frac{H}{2\pi} \ . \quad (5.42)$$

We have thus shown that an observer in the de Sitter phase of a FRW universe detects a thermal flux of particles at a temperature [Eq. (5.42)] given by the Hubble expansion constant. As we shall discuss in detail in Sec. VII, Hawking radiation in the de Sitter phase of an inflationary universe is the source of primordial matter fluctuations, classical energy-density perturbations that may explain the origin of galaxies and clusters of galaxies.

C. Euclidean functional integral approach

The Bogoliubov mode-mixing method has many nice features. Foremost among these are its conceptual simplicity and physical clarity. It is based on an analysis of the process of measuring particles in curved space-time (Unruh, 1976; Gibbons and Hawking, 1977; DeWitt, 1979). On the negative side, however, the computation of the Bogoliubov coefficients is tedious and often requires approximations.

For space-times with special symmetries there exists a much more elegant method of deriving thermal radiation. It relies on establishing an analogy with finite-temperature field theory in flat space-time. To be more precise, it is shown that the Green's functions of the field theory are periodic in Euclidean time with some period τ. They are thus identical in form to finite-temperature Green's functions of temperature τ^{-1} (see Sec. III).

Intuitively it is very easy to understand our result of Sec. V.B concerning Hawking radiation in the de Sitter phase of an inflationary universe, at least for the case of an eternal de Sitter phase. The scale factor $a(t) = \exp(Ht)$ is periodic in imaginary time with period $2\pi H^{-1}$. Hence the Green's functions of the theory must be periodic as well; they will be finite-temperature Green's functions corresponding to the Hawking temperature

$$T_H = \tau^{-1} = \frac{H}{2\pi} \ . \quad (5.43)$$

Gibbons and Hawking (1977) use the path-integral approach to quantum field theory to prove that Green's functions in de Sitter space-time are periodic in imaginary time and yield the Hawking temperature (5.43). It is convenient to introduce Kruskal coordinates (Hawking and Ellis, 1973). We start from the de Sitter metric in static coordinates [see Eq. (5.22)]:

$$ds^2 = -(1-r^2H^2)dt^2 + (1-r^2H^2)^{-1}dr^2 + r^2d\Omega^2 \ . \quad (5.44)$$

First, we embed de Sitter space as the hyperboloid

$$-T^2+S^2+x^2+y^2+z^2 = H^{-2} \quad (5.45)$$

in five-dimensional Minkowski space-time, i.e., R^5 with the metric

$$ds^2 = -dT^2+dS^2+dx^2+dy^2+dz^2 \ . \quad (5.46)$$

The coordinate transformation is

$$T = (H^{-2}-r^2)^{1/2}\sinh(Ht) \ ,$$
$$S = (H^{-2}-r^2)^{1/2}\cosh(Ht) \ ,$$
$$x = r\cos\vartheta\cos\varphi \ , \quad (5.47)$$
$$y = r\sin\vartheta\sin\varphi \ ,$$
$$z = r\cos\vartheta \ .$$

The Kruskal coordinates U and V are defined by

$$T+S = 2H^{-1}U \ ,$$
$$T-S = 2H^{-1}V \ . \quad (5.48)$$

Hence

$$e^{2Ht}=VU^{-1}\ ,$$
$$r=(1+UV)(1-UV)^{-1}H^{-1}\ . \tag{5.49}$$

The main idea of the Gibbons and Hawking (1977) analysis is to prove that the functional integral for the Green's function can be given a well-defined meaning by analytically continuing back from a region in which the metric is positive definite, namely the five-dimensional sphere obtained by replacing T by $i\tau$ (τ real). Following Hartle and Hawking (1976), we construct the propagator for a scalar field of mass m by

$$G(x,x')=\lim_{\varepsilon\to 0}\int_0^\infty dW\,F(W,x,x')\times\exp[-(im^2W+\varepsilon W^{-1})]\ . \tag{5.50}$$

$F(W,x,x')$ is given by a functional integral over all paths $x(w)$ from x to x':

$$F(W,x,x')=\int [Dx(w)]\exp\left[\frac{i}{4}\int_0^W g(\dot{x},\dot{x})dw\right]\ . \tag{5.51}$$

We continue W to negative imaginary values. Then on the Euclidean 5 sphere F satisfies the diffusion equation

$$\frac{dF}{dW}=\nabla^2F\ , \tag{5.52}$$

where ∇^2 is the Laplacian on the 4 sphere. Using Kruskal coordinates it is easy to analyze the singularities of $G(x,x')$. Singularities occur when x and x' can be connected by a null geodesic. For other values of x and x' the propagator is given in terms of an analytic function in U and V. Hence by Eq. (5.49) the propagator will be periodic in t. The corresponding temperature is our by now familiar Hawking temperature (5.43).

The functional integral method we described above was pioneered for the black hole metric by Hartle and Hawking (1976). Gibbons and Perry (1976,1978) investigated the equivalence with thermal Green's functions. They also pointed out an important further result of the Euclidean approach: it shows that at least in perturbation theory Hawking radiation persists even for an interacting quantum field theory. Since all Green's functions can be expanded in terms of propagators and vertex factors, they will all be thermal provided that the free field propagator is thermal. Hence at least perturbatively the curved-space-time Green's functions will be thermal.

It is important, however, to realize that while the elegant functional integral method is only possible for metrics with special symmetries and event horizons, Hawking radiation is not restricted to such cases. Nontrivial mode mixing is obviously independent of special symmetries or event horizons. Brandenberger and Kahn (1982) work out the example of a de Sitter—like phase of a FRW universe with scale factor

$$a(t)=a(0)\exp\left[\frac{2H}{\Gamma}(1-e^{-\Gamma t/2})\right]\ . \tag{5.53}$$

The Hawking temperature in this example is time dependent,

$$T_H(t)=\frac{H}{2\pi}e^{-\Gamma t/2}\ , \tag{5.54}$$

and in particular is unrelated to the period of the scale factor in Euclidean time.

D. Stress tensor approach

The methods discussed in the previous sections all analyze the spectral distribution of a single quantity. In particular, the Bogoliubov mode-mixing method gives the number density of particles detected by an observer. In many examples the spectral distribution is thermal. The Bogoliubov method, however, does not allow any conclusions concerning the equation of state of matter.

Hawking radiation can be viewed as a consequence of vacuum fluctuations in a nontrivial background geometry (Unruh, 1974; Sciama *et al.*, 1981). An entirely different effect is the production of real particles in an external gravitational field, which is analogous to particle production in an external electromagnetic field (Schwinger, 1951). These vacuum polarization effects also induce nonzero "vacuum" expectation values, in particular for the energy-momentum tensor.

Evaluating the "vacuum" expectation value of $T_{\mu\nu}$ gives the equation of state of matter produced by polarization effects in the background metric. In general the equation of state will not be thermal in the sense that $p\neq\frac{1}{3}\rho$. This is true in particular in de Sitter space (Bunch and Davies, 1978) and in the approximate de Sitter phase of a FRW universe (Bunch and Davies, 1977a,1977b; Brandenberger, 1983). The stress tensor approach does not give the spectral distribution of physical quantities.

The main idea of the stress tensor approach is very simple; the implementation, however, is both tedious and conceptually delicate. The idea is to first pick the state we want to evaluate $T_{\mu\nu}$ in. The choice is determined by the dynamics of the previous evolution. Often the state is the vacuum state in a given coordinate frame defined as explained in Sec. V.A [see in particular Eqs. (5.6) and (5.7)]. Then the mode expansion (5.6) of the scalar field $\varphi(\mathbf{x},t)$ is inserted into the expression for $T_{\mu\nu}$ for a classical field. The final step is to evaluate the expectation values using Eq. (5.7) and the canonical commutation relations.

We immediately must confront a delicate conceptual problem. All expectation values diverge. They must be rendered finite by regularization and renormalization.

Even in flat Minkowski space-time infinities arise. They are usually interpreted as a consequence of the infinite vacuum energy of the ground state of a free scalar field (the sum of the ground-state energies of all the harmonic oscillators labeled by $\mathbf{k}$, which the free scalar field decomposes into in momentum space) and are removed by normal ordering.

Normal ordering is a noncovariant renormalization scheme. It is only defined once we have picked a frame and thus defined the meaning of creation and annihilation operators. In curved space-time a new renormalization prescription is needed.

The subject of renormalization in curved space-time is too complex an issue to admit a quick discussion as an appendix to this paper. A separate review article is required. Several reviews on the subject already exist (Boulware, 1979; Davies, 1980; DeWitt, 1975). We shall just say a few words in order to guide the reader to the literature.

There exist many renormalization schemes. They differ by the regularization method used. Once the theory is rendered finite by regularization, renormalization proceeds by dropping the terms which diverge as the regularization cutoff is removed. In addition, terms depending on quantities that enter via regularization must be dropped (one example being finite terms which depend on the direction of point splitting when using point-splitting regularization). This renormalization scheme seems *ad hoc* and not based on clear physical principles. *A priori* one would expect different answers for each scheme. If we restrict our attention to the renormalization of the stress-energy tensor the situation is slightly clearer. There are certain conditions that the renormalized expectation value of $T_{\mu\nu}$ must satisfy; it must be conserved, obey causality, give the usual result for off-diagonal terms, and vanish in Minkowski space-time. Wald (1977) has shown that, given a precise formulation of the above conditions, the answer is unique up to conserved geometrical tensors. Since most renormalization schemes have the above conditions built into them, it is not surprising they give the same result.

Covariant point splitting is a conceptually straightforward, albeit technically tedious, method. It is very generally applicable. It was pioneered by Fulling and Davies (1976), Davies, Fulling, and Unruh (1976), and Davies (1976), and was widely applied to two- and four-dimensional models by the same authors. A good general discussion is contained in Davies *et al.* (1977). The basic idea is to separate the points at which operators in $T_{\mu\nu}$ (or φ^2) are evaluated by moving them in opposite directions along a given geodesic. This eliminates the source of divergences, namely coincident point singularities.

Earlier methods included "n-wave regularization" (Zeldovich and Starobinskii, 1971), and the related "adiabatic regularization" scheme (Parker and Fulling, 1974; Fulling and Parker, 1974; Fulling, Parker, and Hu, 1974).

The dimensional regularization scheme was applied to curved space-time by Brown (1977) and Brown and Cassidy (1977a,1977b), the idea being to render Feynman integrals finite by evaluating them at noninteger dimensions.

A related approach is ζ function regularization, in which a formally divergent functional determinant is written as a generalized ζ function (Hawking, 1977), a function of the power u of the operator. While the naive sum of eigenvalues of the operator diverges, the ζ function converges for a certain range of u. The idea behind ζ function regularization is to extend the function by analytic continuation and then to define the finite, regularized value of the determinant in terms of this analytic continuation (see, for example, Parker, 1979). ζ function regularization is particularly useful in evaluating effective potentials. It is the method used by Allen (1983,1984) to compute the one-loop effective potential in de Sitter space.

Pauli-Villars regularization is another regularization scheme well known in regular quantum field theory. It involves rendering Feynman integrals finite by introducing heavy scalar particles with opposite statistics. The resulting Fermi minus sign for a closed loop introduces an ultraviolet cutoff. The method is mentioned in the review by Boulware (1979), but to our knowledge it has not often been used.

The background-field method, on the other hand, is a well-studied method, first applied by DeWitt (1975) to quantum field theory in curved space-time. It is based on the Schwinger (1951) -DeWitt (1964) proper-time formalism. Early work on this technique includes articles by Dowker and Critchley (1976), Christensen (1976), and Christensen and Fulling (1978).

The easiest example is the theory of a conformally coupled scalar field in a conformally flat FRW space-time. As vacuum state we choose the state that is conformally related to the standard flat space-time vacuum state. Davies *et al.* (1977) have analyzed this case in detail, using regularization by covariant point splitting. The expectation value of $T_{\mu\nu}$ must be a combination of two-index conserved geometric tensors with the correct dimensionality. This follows since the regularization scheme does not break locality. It is an easy linear algebra problem to find the two linearly independent conserved geometric tensors. Their coefficients, however, can only be determined by going through the entire regularization calculation. The result is

$$\langle 0|\, T_\mu^\nu \,|0\rangle = (2880\pi^2)^{-1}[a(2R_{,\mu}^{\;\nu} - 2g_\mu^\nu \Box R + 2RR_\mu^\nu - \tfrac{1}{2}g_\mu^\nu R^2) + b(R_{\mu\alpha}R^{\nu\alpha} - \tfrac{2}{3}RR_\mu^\nu - \tfrac{1}{2}g_\mu^\nu R_{\alpha\beta}R^{\alpha\beta} + \tfrac{1}{4}g_\mu^\nu R^2)] \,. \tag{5.55}$$

For a scalar field

$$a = -\tfrac{1}{6}, \quad b = 1 \,. \tag{5.56}$$

In particular, for the de Sitter phase of a FRW universe, the result becomes (Bunch and Davies, 1977b)

$$\langle 0|\, T_{\mu\nu} \,|0\rangle = \rho g_{\mu\nu} \tag{5.57}$$

with

$$\rho = \frac{H^4}{960\pi^2} \ . \tag{5.58}$$

The vacuum stress-energy tensor is not thermal; it is de Sitter invariant. Similar results hold for the de Sitter—like phase of an inflationary universe (Brandenberger, 1983).

VI. INFLATIONARY UNIVERSE MODELS

A. Problems with the standard model

In the past three years there has been a lot of interest in inflationary universe models. The original idea is due to Guth and Tye (1980) and Guth (1981), who proposed what is now called the old inflationary universe. Considerable progress towards a realistic scenario was made by Linde (1982a) and Albrecht and Steinhardt (1982) by introducing the new inflationary universe based on a Coleman-Weinberg (1973) potential for the scalar field that drives inflation. Other important contributions are due to Hawking and Moss (1982) and Press (1981).

Inflationary universe models arise from an attempt to modify standard big bang cosmology by treating matter, not as an ideal gas, but in terms of quantum fields. Formally, the models involve coupling quantum field theory to classical general relativity. The original motivation was the realization by Guth (1981) that the modified scenarios may solve several important cosmological problems in the standard model, foremost among these the horizon and flatness problems. We shall begin by reviewing the standard model and its problems.

The standard model of the early universe is based on a homogeneous and isotropic Robertson-Walker metric,

$$\begin{aligned} ds^2 &= -dt^2 + a(t)^2 \left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2 \right] , \\ d\Omega^2 &= d\vartheta^2 + \sin^2\vartheta \, d\varphi^2 , \end{aligned} \tag{6.1}$$

where $a(t)$ is the cosmic scale factor; $k=+1,-1,0$ corresponds to a closed, open, or spatially flat universe.

The matter energy-momentum tensor is

$$T_\mu^{\nu M} = \mathrm{diag}(-\rho, +p, +p, +p) \ . \tag{6.2}$$

In the standard model matter is treated as an ideal gas of particles. For high temperature (temperatures larger than the rest mass of the dominant particles in the universe) we have an ideal gas of massless particles. Then the energy density ρ, the pressure p, and the entropy density s are given by

$$\begin{aligned} \rho &= \frac{\pi^2}{30} N(T) T^4 , \\ p &= \frac{\rho}{3} , \\ s &= \frac{2\pi^2}{45} N(T) T^3 , \end{aligned} \tag{6.3}$$

where $N(T) = N_b(T) + \frac{7}{8} N_f(T)$ and $N_b(T)$ $[N_f(T)]$ is the number of bosonic (fermionic) spin degrees of freedom at temperature T. $N(T)$ will, in general, decrease as T decreases, since fewer particles remain in thermal equilibrium with radiation.

The Einstein equations

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu} \tag{6.4}$$

give the following first-order differential equation for the scale factor $a(t)$:

$$\left[\frac{\dot a}{a} \right]^2 + \frac{k}{a^2} = \frac{8\pi G}{3} \rho \tag{6.5}$$

[using Eqs. (6.1) and (6.2)]. $G = m_{\rm pl}^{-2}$ is Newton's constant, $m_{\rm pl}$ the Planck mass.

The third element in the standard model is the assumption of adiabatic expansion of the universe,

$$\frac{d}{dt}(sa^3) = 0 \ . \tag{6.6}$$

For constant $N(T)$ this becomes

$$aT = \text{const} \ . \tag{6.7}$$

The FRW equation (6.5) can then be rewritten as a differential equation for the temperature T:

$$\left[\frac{\dot T}{T} \right]^2 + \varepsilon T^2 = \frac{8\pi G}{3} \rho, \quad \varepsilon = \frac{k}{a^2 T^2} \ . \tag{6.8}$$

Entropy conservation means ε is constant.

It is unreasonable to expect that the classical description of matter as an ideal thermal gas will remain accurate at high temperatures. Then the quantum nature of matter will become important. The inflationary universe models are based on the description of matter by a quantum field theory, in particular a grand unified field theory. In this case Eq. (6.2) is not the only term in the classical energy-momentum tensor that couples to gravity through the Einstein equations (6.4). If $|\psi\rangle$ denotes the quantum state of the system, there is a new contribution,

$$T^Q_{\mu\nu} = \langle \psi | T^{\rm op}_{\mu\nu} | \psi \rangle \ . \tag{6.9}$$

$T^{\rm op}_{\mu\nu}$ stands for the energy-momentum tensor operator of the field theory. The total classical energy-momentum tensor is

$$T_{\mu\nu} = T^M_{\mu\nu} + T^Q_{\mu\nu} \ . \tag{6.10}$$

If the state $|\psi\rangle$ does not break the symmetries of the FRW background metric, then

$$T^Q_{\mu\nu} = \rho_0 g_{\mu\nu} \tag{6.11}$$

acts as a cosmological constant term in the Einstein equations.

In the standard big bang model $T^Q_{\mu\nu}$ is neglected. This is equivalent to assuming that the quantum state is one with $\rho_0 = 0$ (a true ground state). Since the present cosmological constant presumably is zero [the best upper bound is (Hawking, 1984) $|\Lambda| < 10^{-82}$ GeV^2 or $\Lambda / m_{\rm pl}^2$

$<10^{-120}$], ρ_0 must vanish today or be negligibly small. This need not hold in the early universe. As demonstrated in the previous sections, a large class of field theories undergo phase transitions as the temperature changes. ρ_0 changes during these transitions. Before elaborating on the cosmology of field theories with nonvanishing and nonconstant ground-state energy density, let us look at some of the cosmological problems of the standard model.

Both the horizon and the flatness problems are naturalness problems (Guth, 1981). It is unnatural that we observe homogeneous microwave background radiation from a region many orders of magnitude larger than that which could ever have been in causal contact. This is the horizon problem. The flatness problem is related to the observational fact that the present energy density of the universe ρ is less than an order of magnitude different from the critical energy density ρ_c, the borderline value which corresponds to a spatially flat universe,

$$0.1<\frac{\rho}{\rho_c}<4\ . \tag{6.12}$$

$\rho=\rho_c$ is an unstable fixed point under time evolution. Hence early in the universe the relative energy difference between ρ and ρ_c must have been much smaller. If we extrapolate the densities back to a temperature $T=10^{17}$ GeV, the relative energy-density difference becomes of the order of 10^{-55}, an unnaturally small number.

The quantitative analysis of the flatness problem is based on an equation for the relative energy-density difference,

$$\frac{|\rho-\rho_c|}{\rho}\ . \tag{6.13}$$

From Eq. (6.8) we have

$$H^2=\frac{8\pi G}{3}\rho_c \tag{6.14}$$

and

$$H^2+\varepsilon T^2=\frac{8\pi G}{3}\rho\ , \tag{6.15}$$

where, as usual, the Hubble parameter $H(t)$ is $\dot{a}(t)/a(t)$. Thus

$$\frac{|\rho-\rho_c|}{\rho}=\frac{3}{8\pi G}\frac{\varepsilon T^2}{\rho}=\frac{45\varepsilon}{4\pi^3 GNT^2}\sim T^{-2}\ , \tag{6.16}$$

which shows immediately that ρ_c is an unstable fixed point under expansion of the universe. By Eq. (6.3) ε is related to the total entropy S, which in turn can be experimentally determined (Guth, 1981),

$$|\varepsilon|=\frac{1}{(aT)^2}=O(1)N^{2/3}S^{-2/3}\le 10^{-58}N^{2/3}\ . \tag{6.17}$$

If we take $N\sim 10^2$ and evaluate (6.16) at $T=10^{17}$ GeV we obtain

$$\frac{|\rho-\rho_c|}{\rho}<10^{-55}\ . \tag{6.18}$$

The fine-tuning of initial conditions required to give this small value is related to the large present entropy density of the universe. Any solution of the flatness problem must reduce the entropy in the early universe.

The horizon problem is illustrated by Fig. 16. The concentric circles represent space at a fixed time (the radial axis being time). Constant angle corresponds to constant comoving coordinates. An observer at O determines the visible size of the universe by finding the most distant sources of radiation. Projecting their positions back to time t, at constant comoving coordinates, defines a quantity $L(t)$, the physical radius at time t of the presently observed universe. Is it possible to explain the observed homogeneity and isotropy of the universe within a radius $L(t)$ by a causal process? The horizon problem: it is impossible to do so. To see this we fix an initial time t_0, the time at which we want to set initial conditions [e.g., the Planck time $t_{\rm pl}$ defined by $T(t_{\rm pl})=m_{\rm pl}$]. Then the forward light cone of a point at t_0 has a physical radius $l(t)$ at time t. We understand microphysics for times larger than some intermediate time t_I (according to current particle physics models for energies lower than the grand unification scale). But at t_I the region of causal contact is much smaller than the observed horizon, i.e.,

$$l(t)\ll L(t)\ . \tag{6.19}$$

Thus homogeneity and isotropy of the universe must be postulated as initial conditions.

To give a quantitative analysis we consider the spatially flat case $\varepsilon=0$ and take $t_0=0$. Then Eqs. (6.3) and (6.8) yield

$$T^2(t)=\frac{1}{2\alpha t},\quad \alpha^2=\frac{8\pi^3 GN(T)}{90}\ , \tag{6.20}$$

or by entropy conservation $a(t)\sim t^{1/2}$. Hence the particle horizon (physical radius of causal contact) $l(t)$ is

$$l(t)=a(t)\int_0^t a^{-1}(s)ds=2t\ . \tag{6.21}$$

The physical radius $L(t)$ of the observed part of the universe, on the other hand, is given by the equation for constant comoving coordinates,

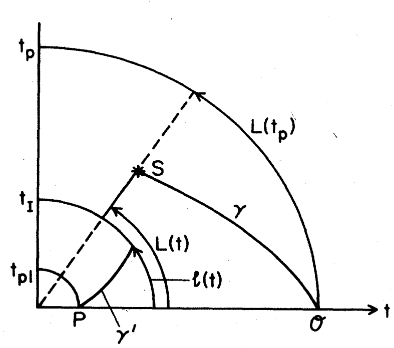


FIG. 16. The horizon problem: $L(t)\gg l(t)$. Concentric circles represent space at fixed time, the radial axis is time. γ indicates a light ray from a star S to the observer at O. γ' indicates the forward light cone of the point P at $t_{\rm pl}$.

$$L(t)=\frac{a(t)}{a(t_p)}L(t_p)=\frac{T(t_p)}{T(t)}L(t_p)\ , \qquad (6.22)$$

where t_p is the present time, $T(t_p)$ the present microwave background temperature. Hence by Eqs. (6.20)—(6.22)

$$\frac{l(t)}{L(t)}=[\alpha T(t)T(t_p)L(t_p)]^{-1}\ . \qquad (6.23)$$

With $N\sim 10^2$, $T(t_p)=2.7$ K, and $L(t_p)=10^{10}$ yr, and evaluating Eq. (6.23) at 10^{17} GeV, we obtain

$$\left[\frac{l}{L}\right]\sim 10^{-28}\ . \qquad (6.24)$$

The idea underlying the solution of the horizon problem is very simple. We assume a sufficiently long period of exponential growth of the scale factor $a(t)$ starting after $t_{\rm pl}$ and ending before $t(T_I)$. Thus $l(t)$ will expand exponentially during this period while $L(t)$ remains unchanged. The small number problem (6.24) is eliminated provided the period τ of exponential expansion satisfies

$$e^{H\tau}>10^{28}\sim e^{65}\ . \qquad (6.25)$$

The solution is sketched in Fig. 17. Between t_A and t_B the scale factor is increasing as $a(t)\sim e^{Ht}$.

It is not hard to realize the above idea if we describe matter in terms of a field theory. Restricting our attention to the spatially flat case $k=0$, it is obvious from Eq. (6.5) that a period of exponential expansion (i.e., a period of inflation) will occur if and only if $\rho(t)$ is constant. Since for a scalar field theory

$$\rho[\varphi(\mathbf{x},t)]=\tfrac{1}{2}\dot{\varphi}^2+\tfrac{1}{2}(\nabla\varphi)^2+V[\varphi(\mathbf{x},t)]\ , \qquad (6.26)$$

this will be true if the energy density is dominated by a constant potential-energy term. This in turn occurs in field theory models that undergo phase transitions. The high-temperature quantum state must have positive potential energy; the low-temperature vacuum state must have vanishing potential energy. During the phase transition, vacuum energy will be converted into thermal energy. This is equivalent to entropy generation. The entropy will increase by a factor Z^3 where

$$Z=e^{H\tau}\ . \qquad (6.27)$$

The temperature of the original thermal state decreases exponentially during the inflationary, or supercooling, phase. In a rapid transition (rapid compared to H^{-1}), all vacuum energy is converted into thermal energy. Thus the temperature after this "reheating" is of the same order of magnitude as the temperature before inflation (see Fig. 18).

Entropy generation also solves the flatness problem. ε is no longer constant. According to Eq. (6.8) its present value is smaller by a factor Z^2 than its original one. Hence for times prior to entropy production the right-hand side of Eq. (6.16) must be multiplied by Z^2. For the value of Z given by Eq. (6.25), the initial value of $|\rho-\rho_{\rm cr}|\rho^{-1}$ is no longer unnaturally small. The above can be explained much more intuitively: even if the overall curvature of the universe originally was large, due to the exponential expansion of the universe our presently observed part of the universe originates from such a small section as to render it essentially flat. It is remarkable that the same minimal value for Z is required to solve both the flatness and horizon problems.

Inflationary universe models can also solve a whole class of cosmological problems that arise when coupling grand unified field theories to cosmology. Many grand unified theories predict the existence of new heavy stable particles such as monopoles or photinos. Estimates of their number densities based on the standard model of cosmology exceed the experimental upper bounds. A period of inflation, however, will dilute the predicted density by a factor Z^3, provided the particles are created before inflation and cannot be produced after. With the additional factor, the predicted densities are far below experimental limits.

The most famous of these problems is the monopole problem (Zeldovich and Khlopov, 1978; Preskill, 1979). Unified field theories, theories in which a large gauge group G is spontaneously or dynamically broken down to $H=SU(3)\times U(1)$ and where the large group contains the U(1) factor nontrivially, predict the existence of extended topological structures, classical static finite-energy solutions of the equations of motion which correspond to the symmetry-breaking field (e.g., the Higgs field) defining a nontrivial topological map from the (spatial) sphere at infinity into G/H, i.e., having nontrivial winding number

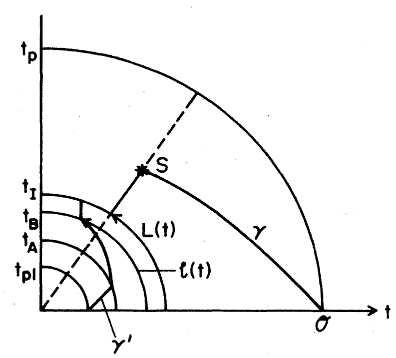


FIG. 17. Solution of the horizon problem: intermediate phase of exponential expansion of the universe for $t_A<t<t_B$. Now $l(t)>L(t)$.

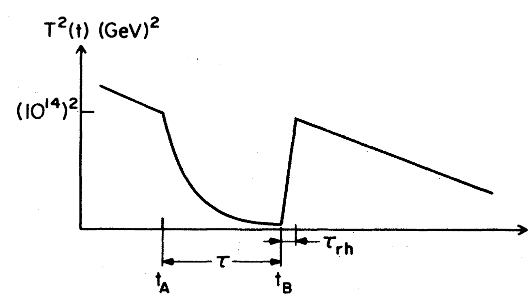


FIG. 18. Temperature of matter in an inflationary universe.

(see, for example, Coleman, 1975,1981). Cosmologically, monopoles form at the energy scale T_c when the U(1) factor first emerges in the symmetry-breaking process. Since microphysics cannot act coherently outside the horizon, there can be no correlations in the angle of symmetry breaking outside the horizon. Hence the monopole density n_M at $t(T_c)$ will be (Kibble, 1976)

$$n_M[t(T_c)] \sim l(t_c)^{-3} \tag{6.28}$$

(density per physical volume). Explicit estimates (Preskill, 1979) show that monopole-antimonopole annihilation is negligible. So are all other decay modes. Hence the present monopole density is given by redshifting:

$$n_M(t_p) \sim l(t_c)^{-3} \left[\frac{a(t_c)}{a(t_p)} \right]^3$$

$$= l(t_c)^{-3} \left[\frac{T_p}{T_c} \right]^3 . \tag{6.29}$$

For monopoles of mass $m_M \sim 10^{16}$ GeV which form at $T_c \sim 10^{16}$ GeV, the density is $n_M(t_p) \sim 10^{-29}$ GeV^4, which exceeds the critical density $\rho_T \sim 10^{-52}$ GeV^4 to close the universe by many orders of magnitude. ρ_T is the energy density of a radiation-dominated universe at $T \sim 10^{-13}$ GeV. The actual critical energy density is larger by a couple of orders of magnitude (see our discussion of the flatness problem).

Inflation solves the monopole problem in an elegant way (Guth, 1981): provided no monopoles can be produced after the period of exponential expansion, the monopole density gets suppressed by a factor Z^3 during inflation.

In gauge theories with a nontrivial vacuum structure there are other topological structures: one-dimensional strings and two-dimensional domain walls (Kibble, 1976). They arise in many grand unified theories (Kibble *et al.*, 1982; Olive and Turok, 1982) and form at the scale of symmetry breaking. Their energy density could easily exceed ρ_c. If strings and domain walls form before the onset of inflation, their energy density is reduced to a negligible value.

In supersymmetric theories, the lightest of the supersymmetic partners of observed particles is stable. In the standard cosmological model, the predicted densities of these particles often exceed experimental lower bounds (Ellis, Linde, and Nanopoulos, 1982). Inflation can cure this problem in exactly the same way as it cures the monopole problem, namely by suppressing n_M by a factor Z^3, provided the superpartners remain out of thermal equilibrium after inflation. Since the particles may be reproduced after inflation, the analysis is more complicated. See Khlopov and Linde (1984) and references therein.

A final problem that inflation easily cures is the "rotation problem" (Ellis and Olive, 1983), the absence of observed rotation in our universe. Due to inflation, it is possible that the universe originally was rotating, but that this effect is unobservable today since our part of the universe corresponds to a very small primordial region of the universe.

Inflation makes an important prediction concerning the flatness of the universe: As we saw above, any initial ratio $|\rho-\rho_c|/\rho$ of the order of one before inflation will lead to a present ratio that is infinitesimal. $\Omega \equiv \rho/\rho_c$ is predicted to be one. Luminous matter can only account for a small fraction of this energy density: $\Omega_{\rm lum} \sim 10^{-2}$. Thus inflationary universe models predict that the energy density in the universe is dominated by some form of dark matter (see, for example, Primack and Blumenthal, 1983a,1983b, and references therein). Galaxy formation in such models is then dominated by the dark matter (Primack and Blumenthal, 1983a,1983b; Bond and Szalay, 1983).

Cosmological implications of phase transitions in quantum field theories were studied before the seminal paper by Guth (1981). Linde (1974) and Bludman and Ruderman (1977) pointed out that there will be a large cosmological constant in the symmetric phase of a theory that undergoes spontaneous symmetry breaking. Einhorn, Stein, and Toussaint (1980), Einhorn and Sato (1981), and Guth and Tye (1980) realized that exponential expansion in the period of supercooling in the symmetric phase would dilute the monopole density and thus solve the monopole problem. Sato (1981a) argued that the expansion could also account for the fact that the sign of CP violation is the same everywhere in our observed universe. He also realized that in a phase transition by bubble nucleation there is no completion of the transition (Sato, 1981b). Kazanas (1980) finally suggested exponential expansion of the universe as the reason for the observed homogeneity of the universe.

B. Inflationary universe models

Guth (1981) originally suggested realizing an inflationary universe scenario by considering a field theory with a first-order phase transition by bubble nucleation, i.e., in which $\varphi=0$ is a local minimum of the one-loop effective potential. A "toy model" would be a pure scalar field theory with a quartic potential, with a local minimum at $\varphi=0$ and a global minimum at $\varphi=\sigma$. A more realistic model is the SU(5) Georgi-Glashow model of Sec. II, with a small explicit mass term for the Higgs field ($m \ll \sigma$). The finite-temperature one-loop effective potential is given by Eq. (3.31),

$$V_{\rm eff}^{(1)T}(\overline{\varphi}) = B\overline{\varphi}^4 \left[\ln \frac{\overline{\varphi}^2}{\sigma^2} - \frac{1}{2} \right] + \tfrac{1}{2} m^2 \overline{\varphi}^2$$

$$+ C\overline{\varphi}^2 T^2 + V(0) \tag{6.30}$$

and is sketched in Fig. 19 for both zero and high temperature. B and C are constants determined by coefficients of order unity and powers of the gauge coupling constant g (Linde, 1982a),

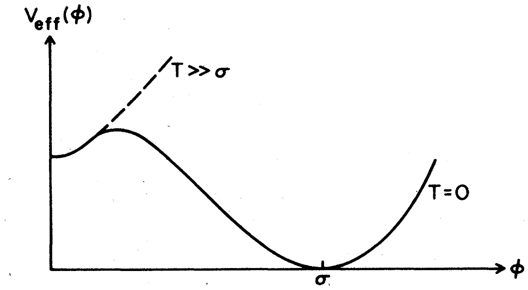


FIG. 19. Sketch of the effective potential in the old inflationary universe for zero and high temperature.

$$B=\frac{5625}{512\pi^2}g^4 ,$$
$$C=\frac{75}{16}g^2 . \tag{6.31}$$

The scenario, now commonly called the old inflationary universe, starts after the big bang with a thermal state at the Planck scale 10^{19} GeV. Above this temperature quantum gravity presumably becomes important, and all present theories break down. Since the scalar field energy is of the order σ^4 and hence much smaller than the radiation energy, the thermal $T^M_{\mu\nu}$ dominates the energy-momentum tensor. Hence there will be an initial radiation-dominated phase, during which $a(t)\sim t^{1/2}$. The finite-temperature corrections in Eq. (6.30) render the symmetric minimum $\varphi=0$ the absolute minimum of $V_{\rm eff}$. Gauge symmetry is unbroken. During this phase the thermal energy redshifts away. As soon as T falls below the critical temperature T_c the time-independent vacuum energy $V(0)$ begins to dominate $T_{\mu\nu}$. The universe starts a phase of exponential expansion. The temperature T_m of the original thermal ensemble supercools exponentially,

$$T_m(t)\sim T_c e^{-H(t-t_{\rm GUT})} . \tag{6.32}$$

After a few e foldings into this de Sitter phase, $\varphi=0$ ceases to be the stable minimum of the effective potential. $\varphi=0$ becomes a metastable "false" vacuum state. Its decay rate τ^{-1} can be determined by the methods of Sec. IV. Since $T_m(t)$ decreases exponentially, we can use the zero-temperature effective potential.

At this point it is appropriate to comment on the role of Hawking radiation. As discussed in Sec. V, there is Hawking radiation even in a de Sitter phase of finite length in a FRW universe. This radiation, however, does not correspond to a thermal bath of particles. In fact (Brandenberger, 1983), the equation of state is de Sitter. Hence it is not correct simply to apply finite-temperature field theory at the Hawking temperature T_H. Instead, one must compute the effects of curvature on the effective potential. Shore (1980) and Allen (1983) have computed these curvature effects for sample field theories. They conclude that there is a gravitational correction term in $V_{\rm eff}(\varphi)$. In the limit $g^2\varphi^2 \ll R$

$$\Delta V_{\rm eff}(\varphi)=O(1)g^2R\varphi^2\ln\frac{R}{\mu_2^2}+O(1)g^4\varphi^4\ln\frac{R}{\mu_3^2} . \tag{6.33}$$

Here g is the gauge coupling constant of the sample field theory, μ_i are renormalization constants, and the $\mathcal{O}(1)$ coefficients will in general depend on the coupling of the scalar field to gravity, i.e., on the value for ζ in the term $\frac{1}{12}\zeta R\varphi^2$ in the Lagrangian. Of particular interest is the mass term in Eq. (6.33). In the old inflationary universe ($H \ll m \ll \sigma$) curvature corrections are negligible.

The lifetime of the metastable vacuum $\varphi=0$ is τ. If $\tau > 65H^{-1}$, then the de Sitter phase is sufficiently long to solve the horizon and flatness problems. The metastable state will decay by tunneling through the potential barrier. Thereafter the evolution of the system is determined by the classical motion of the vacuum expectation value of the scalar field. In the old inflationary universe, the classical evolution starts at a large value of φ (the turning point of the instanton). It takes a short time for the field to evolve to the absolute minimum at σ (large and short are understood as compared to H and H^{-1}). During this brief period red-shift effects are negligible.

At values comparable to σ the Higgs field develops its large mass. Feynman graphs corresponding to the Higgs field decaying into a fermion-antifermion pair generate a complex phase in the effective action. As analyzed by Abbott, Farhi, and Wise (1982), this leads to an effective damping term $\Gamma\dot{\varphi}$ in the equation of motion for $\langle 0|\varphi|0\rangle$. Once $\langle 0|\varphi|0\rangle$ reaches σ, it will perform damped oscillations about σ and release its energy as thermal energy of the particles produced by the damping mechanism. Since on dimensional grounds Γ will be of the order σ, the damping period, i.e., reheating period, will be of the order σ^{-1} and thus negligible compared to H^{-1}. Reheating has also been analyzed by Albrecht, Steinhardt, Turner, and Wilczek (1982), Dolgov and Linde (1982), and Hosoya and Sakagami (1984).

In the old inflationary universe, the vacuum energy of the Higgs field before tunneling is transformed into both thermal and wall energy. If φ_e denotes the turning point of the instanton solution (see Fig. 19), then

$$\rho_{\rm wall}=V(0)-V(\varphi_e)=\mathcal{O}(1)\sigma^4 . \tag{6.34}$$

The rest of the energy density is thermalized,

$$\rho_{\rm thermal}=V(\varphi_e)-V(\sigma)=\mathcal{O}(1)\sigma^4 . \tag{6.35}$$

The universe thus reheats to a temperature $T_{\rm rh}$ of the order σ, only a factor of order 1 below the matter temperature at the beginning of the de Sitter phase (see Fig. 18). Reheating is the nonadiabatic process that generates entropy. The temperature of matter increases from $T_{\rm min}$ to $T_{\rm rh}$ at virtually constant scale factor [$T_{\rm min}$ is the matter temperature at the end of inflation and is given by $\mathcal{O}(1)\sigma e^{-H\tau}$]. Hence the entropy increase factor Z is

$$Z=\frac{T_{\rm rh}}{T_{\rm min}}=\mathcal{O}(1)e^{H\tau} . \tag{6.36}$$

After reheating, the scalar field is in its true vacuum state. The only nonvanishing contribution to the energy-

TABLE III. Cosmology of the old inflationary universe model.

Temperature	Time	Scale	Field theory	Cosmology
∞	0			
			quantum gravity	
$T_{\rm pl}=10^{19}$ GeV		Planck		
			$\varphi=0$: true vacuum	$a(t)\sim t^{1/2}$ standard model
$T_{\rm GUT}\simeq 10^{14}$ GeV	$t_{\rm GUT}$	GUT		
			$\varphi=0$: false vacuum	$a(t)\sim e^{Ht}$ de Sitter phase supercooling of matter
T_b	$t_b=t_{\rm GUT}+\tau$		quantum tunneling	reheating of universe
$T_{rh}\sim 10^{14}$ GeV	t_b		$\varphi=\sigma$: true vacuum	
				$a(t)\sim t^{1/2}$ standard model

momentum tensor is from thermal radiation. Hence further cosmological evolution is as in the standard model. In particular, baryosynthesis remains unaffected. The old inflationary universe is summarized in Table III.

Guth (1981) immediately realized that the old inflationary universe does not provide a realistic cosmology. One drawback is that it merely replaces the horizon problem by an inhomogeneity problem. As sketched in Fig. 20, the bubble walls, which carry a large fraction of the initial vacuum energy density, will remain inside our observed horizon. They would create energy-density perturbations of unacceptable magnitude. This problem stems from the fact that bubbles form after inflation, and therefore the bubble wall corresponds to the forward light cone of a point at $T_{\rm GUT}$ in the standard model, by which the observed horizon (6.19) is much too small to contain the observed universe.

A further problem is that the phase transition will never terminate. As time goes on, an arbitrarily large fraction of space will be in the new phase. However, since the physical size of the regions outside the bubble expands exponentially compared to the $t^{1/2}$ growth for regions of the new phase, clusters of the new phase will never join together to form large regions which could contain our observed universe. In technical terms: the bubbles do not

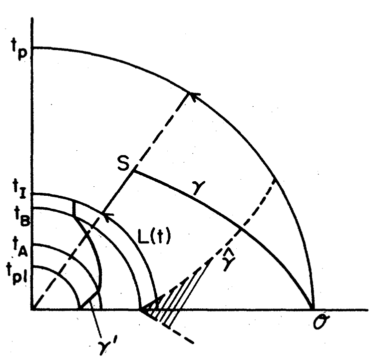


FIG. 20. The homogeneity problem in the old inflationary universe: the bubble of true vacuum phase (γ indicates the bubble walls) is much smaller than the observed horizon $L(t)$.

percolate. This problem has been analyzed in detail by Guth and Weinberg (1983).

To circumvent the above problems, Linde (1982a) and Albrecht and Steinhardt (1982) proposed a modified version of the scenario, the new inflationary universe. Since in this proposal bubbles form before inflation (if they form at all), the observed part of the universe lies within one large bubble, and the above problems do not arise. A similar proposal based on a homogeneous transition of the entire universe was made by Hawking and Moss (1982). The crucial feature of the new inflationary universe is the assumption that the one-loop effective potential of the scalar field driving inflation has a local maximum at $\varphi=0$, as is realized, for example, in a Coleman-Weinberg (1973) model. The transition will then not occur by bubble nucleation, but by spinodal decomposition (see, for example, Langer, 1974). For the SU(5) Georgi-Glashow (1974) model, the one-loop effective potential is given by Eq. (6.30) with $m^2=0$. As emphasized by Linde (1982b,1982c) this m^2 includes both the bare-mass term and the gravitational corrections given by Eq. (6.33):

$$V_{\rm eff}^{(1)T}(\varphi)=B\varphi^4\left[\ln\frac{\varphi^2}{\sigma^2}-\frac{1}{2}\right]+\tfrac{1}{2}B\sigma^4+CT^2\varphi^2 . \tag{6.37}$$

The critical temperature T_c is the energy at which the unified gauge group SU(5) breaks spontaneously. By Eq. (3.32),

$$T_c=T_{\rm GUT}=\left[\frac{B}{2C}\right]^{1/2}\sigma . \tag{6.38}$$

The unification scale σ is of the order of 10^{14}–10^{15} GeV. It can be calculated using data from low-energy physics (see, for example, Buras *et al.*, 1978).

In the new inflationary universe there is no temperature-independent potential barrier to stabilize the symmetric state $\varphi=0$ once $T_m<T_c$. Even if we set only the bare mass equal to zero and keep the gravitationally induced mass term, the latter is of the order $H^2\varphi^2$, with coefficient D. If the evolution of the initial quantum state is determined by tunneling through this small poten-

tial barrier, the turning point of the instanton solution (which, as discussed in Sec. IV, equals the initial point for the classical evolution of the expectation value of φ) is very close to the origin, at a distance of the order H (see Fig. 21). A simple estimate of the distance may be obtained in the following way: we first approximate the effective potential (6.37) for $T=0$ by replacing the logarithm by a constant of order unity,

$$V_{\rm eff}^{(1)}(\varphi)=-\frac{\lambda}{4}\varphi^4+\tfrac{1}{2}B\sigma^4 \ . \tag{6.39}$$

Then the turning point is approximately given by the value of φ for which

$$V_{\rm eff}^{(1)}(\varphi)+DH^2\varphi^2=\tfrac{1}{2}B\sigma^4 \ . \tag{6.40}$$

The answer is

$$\varphi=\left[\frac{D}{\lambda/4}\right]^{1/2}H=\mathcal{O}(1)H \ . \tag{6.41}$$

λ is given by evaluating the logarithm at $\varphi=H$,

$$\lambda=4B\left[\frac{1}{2}-\ln\frac{H^2}{\sigma^2}\right]=120B \ , \tag{6.42}$$

where B is given by Eq. (6.31).

The cosmology of the new inflationary universe is summarized in Table IV. As the temperature decreases from the Planck scale $m_{\rm pl}$ to $T_0=\sigma[15B\pi^{-2}N(T)^{-1}]^{1/4}$, the thermal component of $T_{\mu\nu}$ dominates, and the universe is radiation dominated. As soon as the matter temperature T_m drops below T_0, the vacuum energy begins to dominate. Hence by the FRW equation (6.5) with $k=0$,

$$a(t)=e^{(t-t_0)H} \tag{6.43}$$

with

$$Hm_{\rm pl}=\left[\frac{4\pi}{3}B\right]^{1/2}\sigma^2=\mathcal{O}(1)\sigma^2 \ . \tag{6.44}$$

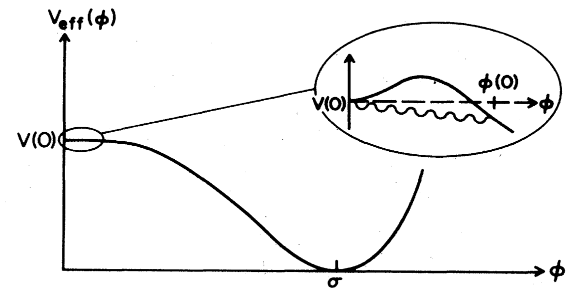


FIG. 21. Sketch of the effective potential in the new inflationary universe (mass term of order H included). The detailed sketch of the potential near the origin includes the instanton solution corresponding to tunneling.

In particular, for the SU(5) Georgi-Glashow (1974) model, $\sigma=\mathcal{O}(1)10^{14}$ GeV and hence $H=\mathcal{O}(1)10^{9}$ GeV. It is important to realize that, in grand unified theories with a Coleman-Weinberg (1973) potential for the scalar field, H and σ are not independent.

At this point the scenario begins to deviate from the old inflationary universe cosmology. Since the potential at the origin is flat (compared to the characteristic energy scale σ, even gravitational corrections are negligible), the evolution of the quantum state can be determined by considering free scalar field theory in the de Sitter phase of the FRW cosmology. The idea is first to treat the scalar field evolution quantum mechanically until quantum field nonlinearities become important, and then to continue semiclassically. Since in theories with symmetry under $\varphi\rightarrow-\varphi$ the vacuum expectation value of φ vanishes, the object to consider is the classical field

$$\varphi_{\rm cl}(t)=\langle 0\,|\,\varphi^2(\mathbf{x},t)\,|\,0\rangle^{1/2} \ , \tag{6.45}$$

renormalized by subtracting the flat-space result for a free scalar field before eliminating the cutoffs. For the appropriate vacuum state (we shall come back to this point

TABLE IV. Cosmology of the new inflationary universe model.

Temperature	Time	Scale	Field theory	Cosmology
∞	0			
			quantum gravity	
$T_{\rm pl}=10^{19}$ GeV		Planck		
			$\varphi=0$: true vacuum	$a(t)\sim t^{1/2}$ standard model
$T_{\rm GUT}\simeq10^{14}$ GeV $\simeq T_0$	$t_{\rm GUT}=t_0$ $\Delta t\sim t^*$	GUT		
			slow-rolling phase $\varphi\sim H$	$a(t)\sim e^{tH}$ de Sitter period
	$t_B\simeq t_0+t^*$ $\Delta t\sim H^{-1}$			
			acceleration phase $\varphi(t){:}H\rightarrow\sigma$	$a(t)\sim e^{tH}$ de Sitter period
	t_{rh} $t_{rh}+\tau_{rh}$			
$T_{rh}\sim10^{14}$ GeV			oscillation phase	reheating period
	$(\tau_{rh}\sim\sigma^{-1})$			
			$\varphi=\sigma$: true vacuum	$a(t)\sim t^{1/2}$ standard model

in Sec. VII) and in the case of minimal coupling, Vilenkin and Ford (1982), Linde (1982c), and Hawking and Moss (1983) find a linear growth of $\varphi_{cl}^2(t)$:

$$\varphi_{cl}^2(t)=\frac{H^3}{(2\pi)^2}t \tag{6.46}$$

(see also Vilenkin, 1983, and Brandenberger, 1984). This growth by far dominates any potential effects due to semiclassical tunneling through the finite-temperature barrier of the effective potential.

Hawking and Moss (1983) show that $\varphi_{cl}(t)$ evolves semiclassically according to the classical field equations of motion, with the potential $V(\varphi)$ replaced by the effective potential $V_{eff}(\varphi)$. Thus $\varphi_{cl}(t)$ can be described as a superposition of the homogeneous quantum growth (6.46) and the growth due to the nontrivial potential determined by the homogeneous solution of the Klein-Gordon equation,

$$\left[\frac{\partial^2}{\partial t^2}+3H\frac{\partial}{\partial t}-a^{-2}(t)\nabla^2\right]\varphi(\mathbf{x},t)=-V'_{eff}[\varphi(\mathbf{x},t)]\ . \tag{6.47}$$

Initially the quantum growth will dominate, later the semiclassical evolution (see also Moss, 1984).

The semiclassical equation for $\varphi_{cl}(t)$ is [by Eq. (6.39)]

$$\ddot{\varphi}+3H\dot{\varphi}=-\lambda\varphi^3\ . \tag{6.48}$$

Since the slope of V_{eff} is initially negligible we can in a first period (from t_0 to t_B) neglect the acceleration term (Guth and Pi, 1982; Hawking, 1982; Brandenberger and Kahn, 1984). The solution of the resulting equation is

$$\varphi(t)=\left[\frac{3H}{2\lambda}\right]^{1/2}\frac{1}{(t^*-t)^{1/2}},\quad t^*=\frac{3H}{2\lambda\varphi^2(0)} \tag{6.49}$$

(for simplicity we set $t_0=0$ in the following computations). The approximation under which Eq. (6.49) holds breaks down at the time t_B when $\ddot{\varphi}$ becomes equal to $3H\dot{\varphi}$. This occurs for

$$t^*-t_B=2H^{-1}\ , \tag{6.50}$$

and the value of φ at t_B is $\varphi(t_B)=(3\lambda^{-1})^{1/2}H$.

In the second period of the de Sitter phase, lasting from t_B to reheating at t_R, we can self-consistently neglect the Hubble redshift. As discussed in Brandenberger and Kahn (1984), the solution of Eq. (6.48) in this approximation which correctly matches solution (6.49) at t_B is

$$\varphi(t)=\left[\frac{2}{\lambda}\right]^{1/2}[\alpha-(t-t_B)]^{-1},\quad \alpha=(\tfrac{2}{3})^{1/2}H^{-1}\ . \tag{6.51}$$

α is approximately the length of this second period.

In both periods the kinetic energy of the scalar field is negligible compared to the vacuum energy. Hence the scale factor continues to increase exponentially according to Eqs. (6.43) and (6.44). Inflation takes place while the scalar field is slowly rolling down the "hill" of the potential (see Fig. 21). In the first period, the "slow-rolling" period, the value of φ remains of the order H; during the final e folding φ rapidly approaches the minimum of the potential.

Since the curvature of the potential at the minimum σ is large (of order σ^2), the damping term in the equation of motion for φ due to the imaginary part in the effective action discussed before will be large, at least in models such as the SU(5) model we are considering, in which the Higgs field couples directly to fermions. Hence the reheating period will be very short (of the order σ^{-1}), and the universe will reheat to a temperature equal to T_c up to a factor of order unity. Thenceforth cosmology evolves as in the standard model.

The new inflationary universe was a big step forward in attempting to improve our understanding of the evolution of the early universe, but it has its share of problems as well.

The most serious objection is the reliance of the scenario on the Coleman-Weinberg potential. A bare-mass parameter can be at most of the order H. This represents an unnatural fine-tuning of particle physics (and it is precisely fine-tuning problems, albeit of cosmological nature, that the model was designed to solve).

Next, it is difficult to obtain sufficient inflation without fine-tuning initial conditions for the evolution of the classical field, a problem emphasized by Starobinsky (1982). Considering first only the semiclassical evolution (6.49), we notice that t^*-2H^{-1} is the length of the slow-rolling period. Writing $\varphi(0)=\gamma H$ and setting $\lambda=0.5$ [see Eqs. (6.42) and (6.31) or Guth and Pi, 1982], we obtain

$$t^*-2H^{-1}=\left[\frac{3}{\gamma^2}-2\right]H^{-1}\ . \tag{6.52}$$

To obtain sufficient inflation, this period should be longer than $65H^{-1}$. Hence γ must be smaller than 0.2, while by dimensional analysis we should expect a value of order 1. In addition, the quantum growth of φ given by Eq. (6.46) will further shorten the period of inflation.

New inflationary universe models yield a mechanism that, from first principles, explains the origin of primordial energy-density fluctuations, a major breakthrough which will be discussed in detail in Sec. VII. In the simplest models, like our sample SU(5) model, however, the predicted amplitude of the fluctuations is 5 orders of magnitude too large (Guth and Pi, 1982; Hawking, 1982; Starobinsky, 1982; Bardeen, Steinhardt, and Turner, 1983). This is the by now famous fluctuation problem.

Next, in our entire analysis we considered only one ray in the configuration space of the SU(5) adjoint Higgs field, the ray in the direction corresponding to SU(5) $\to$SU(3)$\times$SU(2)$\times$U(1) symmetry breaking. Our justification for this was that the global minimum of the potential for the range of parameters we are interested in is in the SU(3)$\times$SU(2)$\times$U(1) phase (Guth and Weinberg, 1981; Albrecht, Jensen, and Steinhardt, 1984). Recently, Breit, Gupta, and Zaks (1983) have pointed out that the shape of the potential in the multidimensional configuration space is such that the inflationary transition will first lead to the phase with $SU(4)\times U(1)$ symmetry, which will later decay into the $SU(3)\times SU(2)\times U(1)$ symmetric minimum. Thus the new inflationary scenario cannot be realized at all in our sample SU(5) model. Gupta and Quinn (1984) have presented an SU(5) model with a full Higgs sector which avoids the above problem.

Finally, as emphasized by Mazenko *et al.* (1984), it is unclear whether realistic initial conditions will lead to an inflationary phase at all (see the final comments in Sec. II).

Since the new inflationary universe has so many striking successes to its credit, and none of the problems seems to invalidate the entire approach, there have been many recent efforts to construct realistic models. In particular, Steinhardt and Turner (1984) have investigated which conditions must be satisfied if the potential of a single scalar field φ is to produce a cosmologically acceptable inflationary scenario (see also Albrecht and Steinhardt, 1983).

First, the potential at the origin must be flat. This is crucial in order to have a new inflationary scenario at all. Second, the flat portion of the potential must be sufficiently long (length much greater than H), else the quantum growth of the classical field given in Eq. (6.46) will prevent enough inflation. Third, the slow-rolling phase must be long enough to give sufficient inflation. In our SU(5) toy model this last condition is not satisfied; the value of φ at which the "slow-rolling approximation" breaks down is too small $[(3\lambda^{-1})^{1/2}H]$, since the slope of the potential is increasing too rapidly. The problem is tied to the specific form of the Coleman-Weinberg potential with only one free parameter, namely σ. To ensure sufficient inflation, the potential must have very small curvature near the origin. Fourth, the vacuum energy in the de Sitter phase may not exceed $3\times 10^{-8} m_{\rm pl}^4$ (Starobinsky, 1979; Rubakov *et al.*, 1982; Fabbri and Pollock, 1983), else primordial gravitational waves will create anisotropies in the cosmic microwave background radiation which violate the observational upper bounds. Fifth, the curvature at the absolute minimum must be large enough to ensure reheating of the universe above temperatures of baryosynthesis. Finally, the potential may not be too flat near the origin, else fluctuations are too large. Our analysis in Sec. VII gives the following general formula for the amplitude of primordial energy-density fluctuations on a given length scale, when that scale enters the Hubble radius $H^{-1}(t)$ at t_f, in terms of the initial fluctuations and the initial equation of state when the scale leaves the Hubble radius in the de Sitter phase at t_i (Brandenberger and Kahn, 1984),

$$\frac{\delta\rho}{\rho}(t_f)=\mathcal{O}(1)\frac{1}{1+w(t_i)}\frac{\delta\rho}{\rho}(t_i)\ . \tag{6.53}$$

Here $w(t)$ is the ratio of pressure to energy density, and $\delta\rho/\rho$ is to be evaluated in comoving coordinates (all concepts will be explained in detail in Sec. VII). If at t_i the slow-rolling approximation is valid, we obtain

$$\frac{\delta\rho}{\rho}(t_f)=\mathcal{O}(1)\frac{H^2}{\dot\varphi(t_i)}\ . \tag{6.54}$$

Thus $\dot\varphi(t_i)$ should be of the order $10^{+4}H^2$ to give the required amplitude.

C. Alternative proposals

In the following we should like to mention briefly some of the models that have recently been suggested to avoid the problems of the new inflationary universe (see also the review by Albrecht, 1984). We shall start with the proposals that most closely follow the new inflationary universe.

Shafi and Vilenkin (1983) proposed a model with a separate "inflaton field." They start from the observation that the problems mentioned above for the new inflationary scenario disappear if λ can be tuned to a small value. [For our minimal SU(5) model this was not possible, since λ was determined by the gauge coupling constant g; see Eqs. (6.42) and (6.31).] By Eq. (6.49), λ must be smaller than 10^{-2} to give sufficient inflation. By Eqs. (6.54) and (6.49), and using $t^*-t_i=50H^{-1}$, we find that λ must be of the order 10^{-12} to give the correct amplitude for fluctuations. Shafi and Vilenkin (1983) consider a standard grand unified field theory and add an additional scalar field to generate inflation, the "inflaton field." A Coleman-Weinberg (1973) potential for the inflaton field is postulated. The self-coupling λ is freely specifiable, although in view of the above remarks it must be smaller than 10^{-12}. To be able to reheat the universe, the inflaton must couple to fermions. Also, the curvature must be large enough to generate sufficient reheating. This sets a lower bound on λ. A value for λ of the order 10^{-12} just barely meets the two constraints. In our opinion this is a simple but also somewhat artificial solution of the problems of the new inflationary universe, artificial since an extra field is introduced without any other physical motivation but to cure cosmological problems of the basis model. In a similar spirit, Pi (1984) has recently proposed a model in which the additional Higgs field also generates the Peccei-Quinn (1977) symmetry, a symmetry used in particle theory to solve the strong CP problem (see Wilczek, 1983, for a recent review).

We have already mentioned the model by Gupta and Quinn (1984) which gives a consistent cosmology and consistent low-energy physics in a full SU(5) model. An interesting feature of the model is that inflation occurs during the transition between the $SU(4)\times U(1)$ phase and the $SU(3)\times SU(2)\times U(1)$ phase. Drawbacks of the model are the severe fine-tunings necessary. Earlier work on the realization of inflation in SU(5) models includes that of Guth and Weinberg (1981) and Sher (1983).

Softly broken global supersymmetry (Dimopoulos and Georgi, 1981) in a natural way solves many of the problems of the new inflationary universe by introducing a second scale into the problem, the supersymmetry-breaking scale m_s (Ellis, Nanopoulos, Olive, and Tamvakis, 1982). Consider the supersymmetric SU(5) model with Coleman-Weinberg potential. Since the one-loop corrections to the effective potential have opposite signs for bosons and fermions, the contributions cancel up to supersymmetry-breaking effects. Thus in Eq. (6.39),

$$\tilde{\lambda}=\left[\frac{m_s}{\sigma}\right]^2\lambda . \tag{6.55}$$

(Quantities in the nonsupersymmetric model are written without a tilde, those in the supersymmetric version with a tilde.) This leads to the following rescalings:

$$\tilde{H}=\left[\frac{m_s}{\sigma}\right]H , \tag{6.56}$$

$$\tilde{t}^*-t_i=\left[\frac{m_s}{\sigma}\right]^{-1}(t^*-t_i) , \tag{6.57}$$

$$\tilde{t}^*=\left[\frac{\sigma}{m_s}\right]^3 t^* . \tag{6.58}$$

Since $\tilde{t}^*$ is basically the length of the de Sitter period, Eq. (6.58) makes it clear that there is no problem in obtaining enough inflation. Also, from Eqs. (6.57) and (6.54) we conclude that

$$\frac{\delta\tilde{\rho}}{\rho}(t_f)=\frac{m_s}{\sigma}\frac{\delta\rho}{\rho}(t_f) . \tag{6.59}$$

The fluctuation problem can be solved as well. These results can be seen intuitively as follows: the curvature of the potential near the origin is much smaller than in the nonsupersymmetric version of the model. Hence the slow-rolling phase will extend from $\varphi=H$ until $\varphi=\sigma^2/m_{\rm pl}$, leading to a much longer time period of inflation. Also, since the scalar field is evolving more slowly, scales of cosmological interest will leave the Hubble radius at a larger value of φ and with a "larger" velocity $\dot{\varphi}$ (larger compared to H^2). Models based on softly broken supersymmetry with $m_s\sim 10^3$ GeV (the value suggested by particle physics) have too small a vacuum energy to reheat sufficiently. Hence the cosmology in these models is wrong.

Steinhardt (1982) and Albrecht *et al.* (1982,1983) investigated the cosmology in reverse-hierarchy supersymmetric models (Witten, 1981), based on the Dimopoulos-Raby (1983) model of geometric hierarchy. This model generates the scales of both weak and strong symmetry breaking from an intermediate scale μ typically of the order 10^{12} GeV. For $\mu<\varphi<m_{\rm pl}$ the potential of the scalar field φ is given by

$$V(\varphi)=c_1\mu^4-c_2\mu^4\ln\frac{\varphi}{m_{\rm pl}} , \tag{6.60}$$

where c_1 and c_2 are constants of order unity (see Fig. 22). A natural choice of the constants in the superpotential gives

$$\frac{c_1}{c_2}>\mathcal{O}(1)10^2 . \tag{6.61}$$

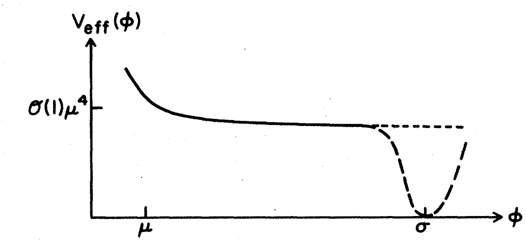


FIG. 22. Sketch of the effective potential for the reverse-hierarchy supersymmetric model with (dashed) and without (pointed) supergravity corrections.

In this case there is enough inflation while the Higgs field is in the flat part of the potential for large values of φ. Since the curvature is suppressed with respect to the slope by an additional factor φ^{-1}, the curvature can be very small while the slope simultaneously is larger, relatively speaking, than in the case of the new inflationary universe. Thus it is, in principle, possible simultaneously to obtain sufficient inflation and suppress fluctuations. To verify the final point we note that in the flat part of the potential beginning at $\varphi^2=(3c_2/8\pi c_1)m_{\rm pl}$ the solution of the space-independent Klein-Gordon equation (6.47) in the slow-rolling approximation is

$$\varphi^2(t)=\frac{3c_2}{8\pi c_1}(1+\tfrac{2}{3}Ht)m_{\rm pl}^2 \tag{6.62}$$

with

$$H^2=\frac{8\pi c_1\mu^4}{3m_{\rm pl}^2} . \tag{6.63}$$

Hence by Eq. (6.54)

$$\frac{\delta\rho}{\rho}(t_f)=\mathcal{O}(1)\frac{c_1}{c_2^{1/2}}\left[\frac{\mu}{m_{\rm pl}}\right]^2 , \tag{6.64}$$

which naturally lies below the upper bound of 10^{-4}.

From the point of view of cosmology, the problem with reverse-hierarchy supersymmetric models is the question of how to reheat the universe. The shape of the potential for $\varphi\sim m_{\rm pl}$ (i.e., near the absolute minimum) is determined by gravitational corrections. In a theory of global supersymmetry the curvature of the potential is too small to reheat the universe to above 10^9 GeV, the minimal value required to obtain the usual scenario of baryosynthesis. In addition, there are decoupling theorems (Dimopoulos and Raby, 1983; Albrecht *et al.*, 1982,1983) which prove that the Higgs field φ is decoupled from the low-energy sector of the theory, thus blocking reheating even if the curvature were large enough. Albrecht and Steinhardt (1983) have recently claimed that it is, in principle, possible to circumvent these theorems. They consider corrections due to local supersymmetry and give an example of a superpotential that leads to extra terms in the potential for φ; these terms generate a steep well near the absolute minimum (indicated in Fig. 22).

A group of authors originally at CERN (Ellis *et al.*, 1983a,1983b; Nanopoulos, Olive, and Srednicki, 1983; Nanopoulos *et al.*, 1983a,1983b; Olive, 1983) and Linde (1983c,1983d) proposed the idea of primordial supersymmetric inflaton, first in the context of global supersymmetry (Ellis *et al.*, 1983), then in simple $N=1$ supergravity (Linde, 1983c,1983d; Nanopoulos, Olive, and Srednicki, 1983; Nanopoulos *et al.*, 1983a,1983b; Olive, 1983). Their approach, similar to that of Shafi and Vilenkin (1983), is to introduce a separate inflaton sector to drive inflation. If the scale of symmetry breaking for this sector is pushed toward the Planck scale, a reasonable cosmological scenario can be obtained without unnatural fine-tunings of the coefficients of the inflaton part of the superpotential. We shall restrict our attention to the inflaton sector of the theory and consider the $N=1$ supergravity model. The tree-level effective potential of a general theory symmetric about $\varphi=0$ is (Linde, 1983d)

$$V(\varphi)=3\mu^6 m_{\rm pl}^{-2}\left[1-\alpha^2 m_{\rm pl}^{-2}\varphi^2+\frac{\alpha^4}{4}m_{\rm pl}^{-4}\varphi^4\right], \qquad (6.65)$$

with the addition of nonrenormalizable terms suppressed by higher powers of the Planck scale. The proponents of the model claim that values of $\hat{\mu}$ and α of the order of 10^{-1} will give a consistent cosmological scenario ($\mu=\hat{\mu}m_{\rm pl}$).

In the model given by Eq. (6.65) the Hubble parameter H, the inflaton mass m_φ, and the location φ_0 of the global minimum of $V(\varphi)$ are given by

$$H=\mathscr{O}(1)\hat{\mu}^3 m_{\rm pl}, \qquad (6.66)$$

$$m_\varphi=\mathscr{O}(1)\hat{\mu}^3\alpha m_{\rm pl}, \qquad (6.67)$$

$$\varphi_0=\sqrt{2}\alpha^{-1}m_{\rm pl}. \qquad (6.68)$$

The fact that φ_0 is much larger than $m_{\rm pl}$ is the basic reason why we obtain sufficient inflation. It should also be noted that in this case the theorem of Ovrut and Steinhardt (1983) concerning the existence of extra minima in the potential does not apply. The constraint on the vacuum energy from gravitational waves (Rubakov *et al.*, 1982; Fabbri and Pollock, 1983) is also (marginally) satisfied. Since scales of cosmological interest leave the Hubble radius when $\varphi(t)$ has a large value with a fairly large slope, fluctuations will be suppressed as well, despite the fact (Kahn and Brandenberger, 1984) that there is a prolonged period of reheating, and scales of interest leave the Hubble radius in a non–de Sitter phase. The result is

$$\frac{\delta\rho}{\rho}(t_f)=\mathscr{O}(1)\hat{\mu}^3\alpha. \qquad (6.69)$$

Nanopoulos *et al.* (1983a,1983b; Nanopoulos, Olive, and Srednicki, 1983) and Olive (1983) stress the fact that their model also reproduces the correct low-energy physics. The judgments of experts in the field on the merits of this model differ sharply, however. Goncharov and Linde (1984), Linde and Goncharov (1984), and Holman *et al.* (1984) have also discussed realistic supergravity models. One problem that reemerges in models of primordial inflation is the monopole problem. Since grand unified symmetry breaking occurs after inflation, there is no substantial dilution of the monopole density. A more careful analysis (see, for example, Linde, 1983c) shows that there is no problem.

For a review of supersymmetric grand unified theories we refer the reader to Ellis (1982). More recent developments in supersymmetric cosmology have been discussed by Olive (1983) and Nanopoulos (1983). Early work on general aspects of cosmology in supersymmetric models includes that of Srednicki (1982a,1982b), Ginsparg (1982), Pi (1982), Nanopoulos and Tamvakis (1982), and Nanopoulos, Olive, and Tamvakis (1982). Linde (1983f) and Kounnas *et al.* (1983) discuss supersymmetry breaking in the context of inflationary universe scenarios. Finite-temperature effects in supergravity models with inflation have been studied by Gelmini *et al.* (1983) and Olive and Srednicki (1984).

Finally, Linde (1983a,1983e) has recently proposed a model of chaotic inflation. For a sufficiently flat effective potential, all values of φ should be equally probable as initial values for the classical evolution of the scalar field φ, Linde argues. He suggests that the observed part of the universe originates from a fluctuation region with $\varphi(0)>3m_{\rm pl}$. Then a reasonable inflationary scenario could even be obtained for a $\lambda\varphi^4$ theory of a simple scalar field. The lower bound on $\varphi(0)$ ensures that the initial slow-rolling period [the period during which the $\ddot{\varphi}$ term in the equation of motion (6.47) is negligible and in which the scale factor is increasing rapidly due to the dominance of the potential energy] is long enough to generate sufficient inflation. It is possible to verify that despite the prolonged period of "fast rolling" Eq. (6.54) remains valid (Kahn and Brandenberger, 1984). Hence one can show that

$$\frac{\delta\rho}{\rho}=\mathscr{O}(1)\lambda^{1/2}. \qquad (6.70)$$

Provided $\lambda<10^{-8}$ there is no contradiction with the maximal amplitude consistent with observations. A more stringent bound $\lambda<10^{-12}$ stems from the constraint on the initial vacuum energy from gravitational waves (Rubakov *et al.*, 1982; Fabbri and Pollock, 1983).

To conclude, we mention the attempts by Hawking and Moss (1982) and Mottola and Lapedes (1983) to investigate the initial evolution of the quantum field in scalar field models with a general (not a Coleman-Weinberg type) potential in more detail. Hawking and Moss (1982) find an instanton solution, which they interpret semiclassically as a homogeneous fluctuation of the quantum field over a horizon volume to the maximum of the potential. A slow-rolling phase follows. Since at the value of φ for which scales of cosmological interest leave the Hubble radius the slope of the potential is larger than in the new inflationary universe, the fluctuations will be smaller (Mottola, 1983).

VII. FLUCTUATIONS IN INFLATIONARY UNIVERSE MODELS

A. The basic mechanism

The universe is not completely homogeneous. There are inhomogeneities which we today see on different length scales as clusters of galaxies, galaxies, stars, etc. On the other hand, we note the absence of observed anisotropies in the cosmic background radiation (for a review and further references, see Weiss, 1980; for the theory of the connection between energy-density fluctuations on large scales and cosmic background radiation anisotropies, see Sachs and Wolfe, 1967). In the standard model the assumption is that present anisotropies originate from small initial perturbations about a homogeneous and isotropic metric. A deep problem, closely related to the horizon problem discussed in Sec. VI, is that there is no ready way to determine the origin of these perturbations. As sketched in Fig. 23, perturbations on all scales originate outside the region of causal contact, the particle horizon.

The particle horizon $l_p(t)$ of an observer at time t is the maximal physical distance light could have traveled from the comoving position in space of the observer at the beginning of the universe ($t=0$) until time t,

$$l_p(t)=a(t)\int_0^t a^{-1}(s)ds \; . \tag{7.1}$$

A length scale that will be very important in what follows is the Hubble radius $H^{-1}(t)$. [As in previous sections, we are considering only spatially flat FRW universes with scale factor $a(t)$, in terms of which $H(t)=\dot{a}(t)a^{-1}(t)$.] The Hubble radius is the maximal distance over which microphysics can act coherently; it is sometimes also called the effective particle horizon (Bardeen, Steinhardt, and Turner, 1983). For larger distances, the time light takes to travel exceeds the characteristic cosmological expansion time. In scalar field theory language, spatial correlations are suppressed [in the case of a de Sitter–like metric $a(t)=\exp(Ht)$ exponentially] on length scales larger than $H^{-1}(t)$. This follows from the Klein-Gordon equation (6.47) in curved space-time applied to plane-wave solutions.

A major success of inflationary universe models first realized by Press (1980,1981), Sato (1981b), Lukash (1980), and Chibisov and Mukhanov (1981) and subsequently investigated quantitatively for the new inflationary universe by Guth and Pi (1982), Hawking (1982), Starobinsky (1982), and Bardeen, Steinhardt, and Turner (1983), is a potential solution to the problem of the origin of fluctuations. The same mechanism that solves the horizon problem, exponential expansion of the universe for a finite period, naturally explains that perturbations on all cosmologically interesting scales (scales inside the horizon at present, but large enough to be detectable cosmologically) originate inside the Hubble radius at some point in the de Sitter phase. This is illustrated in Fig. 24. Thus, provided there exists a mechanism that generates perturbations inside the Hubble radius in the de Sitter phase, we can explain the origin of the primordial fluctuations.

Astrophysicists are interested in the quantity $(\delta M/M)(k,t_f(k))$, the average relative rest mass excess on a comoving length scale $k^{-1}=r$ when this scale enters the horizon in the FRW phase at the time $t_f(k)$. In this expression $t_f(k)$ stands for final horizon crossing in the FRW phase; $t_i(k)$ indicates initial Hubble-radius crossing in the de Sitter phase. The rest mass excess $(\delta M/M)(k,t_f(k))$ can be simply related to the Fourier transform $\delta\rho(\mathbf{k})$ of the continuous energy-density distribution $\delta\rho(\mathbf{x})$ (see Peebles, 1980, Chap. 26). We compute the mass excess $\delta M_r(\mathbf{x}_0)$ in a sphere of radius r about some point $\mathbf{x}_0$, and average $[\delta M_r(\mathbf{x}_0)]^2$ over a cutoff volume V to obtain

$$\left[\frac{\delta M}{M}\right](\mathbf{k})^2=V^{-1}\int d^3k\frac{\delta\rho}{\rho}(k)\,|\,W_r(\mathbf{k})\,|^2 \; , \tag{7.2}$$

with

$$W_r(\mathbf{k})=\int_{V_r}d^3x\, e^{-i\mathbf{k}\cdot\mathbf{x}}\Big/\int_{V_r}d^3x \; , \tag{7.3}$$

where V_r is a sphere of radius r about the origin. Since $W_r(\mathbf{k})\approx 1$ (0) if $|\mathbf{k}|r<1$ ($|k|r>1$) we obtain (using the

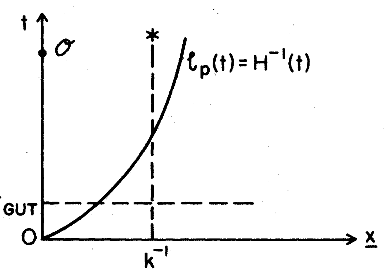


FIG. 23. Evolution in the standard model. Fluctuations on a given scale k^{-1} originate outside the particle horizon $l_p(t)$ of the observer O. The plot is in comoving coordinates.

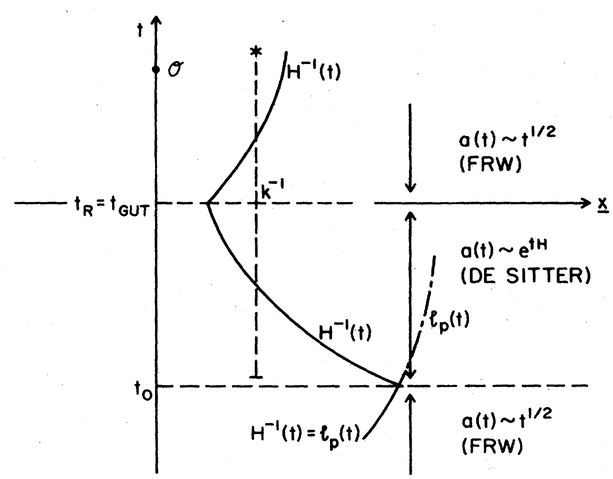


FIG. 24. Evolution of the Hubble radius $H^{-1}(t)$ and of the particle horizon $l_p(t)$ in the new inflationary universe. Fluctuations originate inside the Hubble radius in the de Sitter phase. The plot is in comoving coordinates.

random-phase approximation that $\delta\rho/\rho$ depends only on $|\mathbf{k}|$, and assuming a reasonable k dependence)

$$\left[\frac{\delta M}{M}\right]^2(\mathbf{k})\approx V^{-1}k^3\left[\frac{\delta\rho}{\rho}\right]^2(k)\;. \tag{7.4}$$

The normalization factor V^{-1} is cancelled by a corresponding factor V relating $\delta\rho(k)$ to the power spectrum [see Eq. (7.77)].

A second major success of the new inflationary universe is its prediction of a scale-invariant Zeldovich spectrum (Harrison, 1970; Zeldovich, 1972) for primordial energy-density fluctuations, namely

$$\frac{\delta M}{M}(k,t_f(k))=\text{const.} \tag{7.5}$$

It is very easy to understand qualitatively the reason for this (see Fig. 25). We assume some mechanism that generates perturbations inside the Hubble radius in the de Sitter phase. By time translation invariance of the de Sitter phase, the evolution of fluctuations on two different scales k_1 and k_2 up to the time when they leave the Hubble radius will be identical up to time translation. Hence

$$\frac{\delta M}{M}(k,t_i(k))=\text{const.} \tag{7.6}$$

Since microphysics cannot act coherently outside the Hubble radius, what physically characterizes the size of the perturbation should remain unchanged until horizon crossing $t_f(k)$ in the FRW phase. Thus we expect a scale-invariant Zeldovich spectrum [Eq. (7.5)].

A scale-invariant spectrum was originally postulated because it fit the experimental constraints fairly well and was the only power-law spectrum to do so. The observational constraints are twofold. First, the absence of observed anisotropies in the cosmic background radiation (see, for example, Weiss, 1980) imposes an upper bound on the amplitude of primordial perturbations on large scales (Sachs and Wolfe, 1967),

$$\frac{\delta M}{M}(k,t_f(k))<10^{-4}\quad\text{for } k\sim 10^{19}M_{\text{solar}}\;. \tag{7.7}$$

(We follow the astrophysical convention of labeling scales by the rest mass in a sphere of comoving radius k^{-1}.) On the other hand, clusters of galaxies and galaxies can only form via nonlinear processes. Linear perturbation theory breaks down when relative perturbations become of order 1. Thus knowing that perturbations on the scale of clusters of galaxies must have had time to grow to order 1 after horizon crossing imposes a lower bound on small scales,

$$\frac{\delta M}{M}(k,t_f(k))>10^{-6}\quad\text{for } k\sim 10^{15}M_{\text{solar}} \tag{7.8}$$

(see Silk, 1983,1984, for a review). This bound depends on the details of the cosmological model. In particular, the properties of the particles forming the dark matter of the universe will determine the length of the period during which perturbations on scales of interest grow, and thus will influence the lower bound. Equations (7.7) and (7.8) make a scale-invariant spectrum an obvious candidate, and the prediction of such a spectrum in the new inflationary universe is considered to be a very attractive feature.

The big disappointment is the discovery (Guth and Pi, 1982; Hawking, 1982; Starobinsky, 1982; Bardeen, Steinhardt, and Turner, 1983) that the predicted amplitude is several orders of magnitude too large, namely, larger than unity. This is the by now famous fluctuation problem.

The qualitative argument for a scale-invariant Zeldovich spectrum has its loopholes. First, it is not sufficient that the background metric be time translation invariant in the de Sitter phase. The initial conditions for perturbations must be invariant as well. If we consider imposing initial conditions at a fixed time (e.g., the beginning of the de Sitter phase), it is not at all obvious that the symmetry should be respected. Second, the quantity $\delta\rho/\rho$ does not remain constant outside the Hubble radius (a different quantity, ζ, to be defined later, does so). We must therefore prove that the amplification factor relating $\delta\rho/\rho(t_f)$ to $\delta\rho/\rho(t_i)$ is independent of k.

In the following sections we shall discuss the computation of energy-density fluctuations in inflationary universe models. We shall first summarize the formalism we consider to be most appropriate, namely, gauge-invariant linear perturbation theory (Bardeen, 1980; Brandenberger, Kahn, and Press, 1983), and subsequently outline its application to inflationary universe models.

We fix a comoving scale k and consider the evolution of perturbations on that scale. Our approach is to divide the evolution into two periods, a quantum period, while the perturbation is inside the Hubble radius $[t<t_i(k)]$, and a classical period $t_i(k)<t<t_f(k)$.

Perturbations are generated in the quantum period. The same vacuum fluctuations that give rise to Hawking radiation in the de Sitter phase of an inflationary universe will generate classical matter perturbations which couple to general relativity. Formally, the analysis will consider the quantum theory of a free scalar field on a fixed background metric in the spirit of Sec. V.

In the classical period $t_i(k)<t<t_f(k)$, perturbations are outside the Hubble radius. The classical matter fluc-

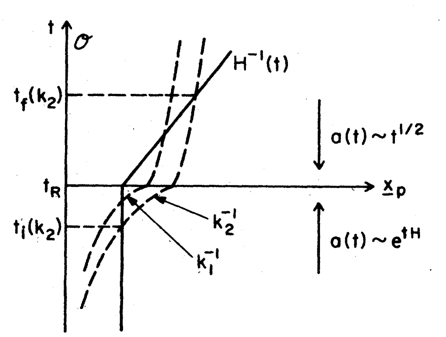


FIG. 25. Sketch of the evolution of fluctuations on two scales in physical coordinates $\mathbf{x}_p$. The evolution inside the Hubble radius in the de Sitter phase is time-translation invariant.

tuations couple to gravity. By the Einstein constraint equations, they will induce small metric fluctuations about the homogeneous and isotropic background metric. The perturbations will evolve according to the dynamical Einstein equations linearized about the background solution. We shall use Bardeen's (1980) gauge-invariant formalism (see also Brandenberger, Kahn, and Press, 1983).

B. The gauge-invariant formalism

We consider small perturbations about a homogeneous and isotropic solution of the Einstein equations

$$G_{\mu\nu}=8\pi G T_{\mu\nu}, \quad G_{\mu\nu}\equiv R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R \ . \tag{7.9}$$

The solution is given by a background metric,

$$g_{\mu\nu}^{(0)}=\mathrm{diag}(-1,a^2(t),a^2(t),a^2(t)) \ , \tag{7.10}$$

and a diagonal energy-momentum tensor,

$$T_{\mu\nu}^{(0)}=\mathrm{diag}(\rho^{(0)}(t),a^2(t)p^{(0)}(t),a^2(t)p^{(0)}(t) \ , \ a^2(t)p^{(0)}(t)) \ , \tag{7.11}$$

satisfying the FRW equations

$$\left[\frac{\dot a}{a}\right]^2=\frac{8\pi G}{3}\rho^{(0)} \ , \tag{7.12}$$

$$2\frac{\ddot a}{a}+\left[\frac{\dot a}{a}\right]^2=-8\pi G p^{(0)} \ . \tag{7.13}$$

The total perturbation is given by a perturbation $g_{\mu\nu}^{(1)}$ of the metric and a perturbation $T_{\mu\nu}^{(1)}$ of the energy-momentum tensor.

Perturbations can be classified according to how they transform under spatial transformations of the background coordinates [i.e., under the SO(3) subgroup of rotations of the Lorentz group]. There are scalar, vector, and tensor perturbations.

Tensor perturbations affect only the traceless, divergenceless parts of the spatial metric g_{ij} and of T_{ij}. If we write out the energy-momentum tensor for a perfect fluid,

$$T^{\mu\nu}=pg^{\mu\nu}+(p+\rho)u^\mu u^\nu \ , \tag{7.14}$$

where u^μ is the four-velocity of matter [its background form being $u^\mu=(1,\mathbf{0})$], then tensor perturbations correspond to

$$\rho^{(1)}=p^{(1)}=u_i^{(1)}=0 \ . \tag{7.15}$$

Hence tensor perturbations do not couple to energy-density perturbations. They are gravitational waves.

Vector perturbations are constructed from the divergenceless vector $Q_i^{(1)}$ by taking covariant derivatives with respect to the spatial background metric. The perturbation of g_{00} vanishes, that of g_{0i} is proportional to $Q_i^{(1)}$, and

$$g_{ij}^{(1)}\sim(Q_{i,j}^{(1)}+Q_{j,i}^{(1)}) \ . \tag{7.16}$$

Hence for vector perturbations, by Eq. (7.14)

$$\rho^{(1)}=p^{(1)}=\nabla\cdot\mathbf{U}^{(1)}=0 \ . \tag{7.17}$$

Vector perturbations are purely rotational modes.

Scalar modes are the only modes that couple to energy-density and pressure perturbations $\rho^{(1)}$ and $p^{(1)}$. Since we are interested in fluctuations that generate the primordial energy-density fluctuations necessary for galaxy formation, we can restrict our attention to scalar modes. Note, however, that primordial tensor fluctuations will influence the cosmic microwave background radiation. Since the amplitude of gravitational waves can grow by a process of superadiabatic amplification in a universe in which there is a period when the characteristic time for change in the background metric is less than the period of the wave and $p^{(0)}\neq\frac{1}{3}\rho^{(0)}$ (Grishchuk, 1974), even small initial tensor modes may become important. Rubakov *et al.* (1982) and Fabbri and Pollock (1983) have derived constraints on inflationary universe models due to primordial tensor perturbations.

The most general scalar metric perturbation can be constructed from scalar functions by multiplying with invariant tensors or taking covariant derivatives with respect to the spatial background metric:

$$g_{\mu\nu}^{(1)}=\alpha^2(t)\begin{bmatrix} E(\mathbf{x},t) & F(\mathbf{x},t)_{,i} \\ F(\mathbf{x},t)_{,i} & A(\mathbf{x},t)\delta_{ij}+B(\mathbf{x},t)_{,ij}\end{bmatrix} . \tag{7.18}$$

We can Fourier-expand the perturbation in space (for a nonflat FRW universe the generalization is the expansion in spherical harmonics; see Lifshitz and Khalatnikov, 1963, or Bardeen, 1980). In linear perturbation theory there is no mixing between the Fourier modes. We shall, therefore, consider perturbations on a fixed comoving scale and analyze their evolution outside the Hubble radius.

In principle the analysis is simple. It is based on linearizing the Einstein equations about the background solution [Eqs. (7.10) and (7.11)].

The problem is to identify the physical modes. There are many gauge modes, perturbations $g_{\mu\nu}^{(1)}$ which correspond to coordinate transformations of the background and do not give physical perturbations.

The most popular approach, based on the pioneering work of Lifshitz [1946; reviewed in Lifshitz and Khalatnikov (1963); the textbook treatments by Weinberg (1972) and Peebles (1980)], has been to partially fix the gauge by demanding

$$g_{00}^{(1)}=g_{0i}^{(1)}=0 \tag{7.19}$$

(synchronous gauge) and, by explicitly calculating the effect of residual gauge transformations, to separate the synchronous modes into physical and pure gauge modes.

In synchronous gauge, the interpretation of perturbations whose wavelength is larger than the Hubble radius is not always straightforward. While the difference in perturbation quantities such as the relative energy-density fluctuation is negligible inside the Hubble radius, it becomes dominant outside. Thus the dominant terms may be pure gauge artifacts. Press and Vishniac (1980) proposed a scheme for keeping track of gauge modes while continuing to calculate in standard synchronous gauge. The proposed scheme clarified a certain number of "tenacious myths" that had crept into the literature. For example, the Press-Vishniac scheme shows clearly that pressure perturbations on scales larger than the Hubble radius do not give rise, at lowest order, to growing density perturbations.

A conceptually straightforward and mathematically elegant approach to perturbation theory for scales outside the Hubble radius was proposed by Bardeen (1980), based on previous work of Hawking (1966) and Olson (1976). The idea is to eliminate the gauge degrees of freedom rather than just to specify and understand them. The gauge-invariant approach is also technically straightforward in the new derivation by Brandenberger, Kahn, and Press (1983), on which the following discussion will be based. We call this approach the Bardeen formalism.

A brief outline of the method: we first define and discuss the gauge-invariant gravitational potential Φ_H. Then we derive the complete set of equations of motion for both background and perturbative variables from a single variational principle. Subsequently we discuss the application of our framework. In particular, we recast the equation of motion for Φ_H as an approximate conservation law.

1. Gauge-invariant variables

A transformation

$$x^{\mu'}=x^{\mu}-\xi^{\mu} \tag{7.20}$$

of the background coordinates induces the following change in the metric:

$$\delta g_{\mu\nu}=\xi_{\mu;\nu}+\xi_{\nu;\mu} \tag{7.21}$$

in particular,

$$\begin{aligned}\delta g_{00}&=-2\dot{\xi}^0 ,\\ \delta g_{0i}&=-\xi^0_{,i}+a^2\dot{\xi}^i ,\\ \delta g_{ij}&=2\dot{a}a\xi^0\delta_{ij}+a^2(\xi^i_{,j}+\xi^j_{,i}) .\end{aligned} \tag{7.22}$$

Since we are only interested in scalar metric perturbations, it is sufficient to consider transformations that preserve the scalar character of the perturbation. Such gauge transformations are parametrized by two free functions $f(\mathbf{x},t)$ and $f^0(\mathbf{x},t)$ with

$$\begin{aligned}\xi^0&=f^0 ,\\ \xi_i&=f_{,i} .\end{aligned} \tag{7.23}$$

This gauge transformation induces the following scalar perturbation:

$$\begin{aligned}A&=2\frac{\dot{a}}{a}f^0 ,\\ B&=2a^{-2}f ,\\ E&=-2a^{-2}\dot{f}^0 ,\\ F&=-a^{-2}f^0+(a^{-2}f)^{\cdot} .\end{aligned} \tag{7.24}$$

There are four degrees of freedom for scalar metric perturbations, but two of them are pure gauge modes. Hence there are two physical degrees of freedom. We expect to obtain as dynamical equations either a system of two first-order coupled differential equations for two linearly independent gauge-invariant variables or a single second-order equation for one of the variables.

It is not hard to find a basis of gauge-invariant variables. They will be combinations of $A(\mathbf{x},t)$ through $F(\mathbf{x},t)$ which vanish for a pure gauge (7.24). The problem reduces to a simple exercise in linear algebra. A basis is

$$\Phi_H\equiv\tfrac{1}{2}(A+2\dot{a}aF-\dot{a}a\dot{B}) , \tag{7.25}$$

$$\Phi_k\equiv E+\left[\left[\frac{\dot{a}}{a}\right]^{-1}A\right]^{\cdot}a^{-2} . \tag{7.26}$$

Later on we shall prove that at Hubble-radius crossing

$$\Phi_H(t_f(\mathbf{k}))=\mathscr{O}(1)\left[\frac{\delta\rho}{\rho}\right]_c(t_f(\mathbf{k})) , \tag{7.27}$$

where the subscript c indicates that the right-hand side is to be evaluated in comoving coordinates [recall that $\delta\rho(\mathbf{x},t)$ is not gauge invariant]. $\mathbf{k}$ labels the Fourier mode under consideration. Since Φ_H is so simply related to the energy-density perturbation, the quantity we are interested in, we choose Φ_H as our gauge-invariant variable.

In principle, we could have chosen any combination of Φ_H and Φ_k as our gauge-invariant variable. Hinshaw (1984) has recently derived the equation of motion for Φ_k [to be more precise, for Φ_k multiplied by a function that depends only on the scale factor $a(t)$]. The resulting gauge-invariant equation of motion has a similar form to that of our result, Eq. (7.44). It is a second-order ordinary differential equation in time, with source terms that depend only on matter variables and the lapse function E. Again, the coefficients of $\dot{\Phi}_k$ and Φ_k depend only on the equation of state. An advantage of Hinshaw's equation is that the matter source terms depend only on T_{ij}, not on T_{0i}. In models in which the T_{ij} are freely specifiable this will be a significant advantage, since the source terms would then not depend on the preceding dynamical evolution of the geometry. When matter is described in terms of quantum fields, however, all components of $T_{\mu\nu}$ are influenced by the geometry. Hence we prefer to adopt the more frequently used variable Φ_H.

While Φ_H is mathematically gauge invariant, its physical interpretation is tied to a certain gauge. Bardeen (1980) demonstrates that in zero shear gauge (gauge in which the normal vectors to the constant time surfaces have zero shear)

$$\Phi_H = \frac{a^4}{k^2} R \ , \tag{7.28}$$

where R is the intrinsic curvature of the constant time surface.

2. Variational derivation of the equations of motion

Utilizing a technique familiar from the variational derivation of the FRW equations, we slightly generalize the background metric by adding a free function $\alpha(t)$:

$$g_{\mu\nu}^{(0)} = \mathrm{diag}(-\alpha^2(t), a^2(t), a^2(t), a^2(t)) \ . \tag{7.29}$$

We also introduce a formal expansion parameter ε:

$$g_{\mu\nu} = g_{\mu\nu}^{(0)} + \varepsilon g_{\mu\nu}^{(1)} \ . \tag{7.30}$$

All formulas will be evaluated at $\alpha = \varepsilon = 1$. All equations will be analyzed to lowest nonvanishing order in ε.

To be concrete (and for our later application) we shall assume that matter can be described by a single classical scalar field $\varphi(\mathbf{x},t)$, with matter Lagrangian

$$\mathscr{L}_M = -[\tfrac{1}{2}\varphi_{,\alpha}\varphi_{,\beta} g^{\alpha\beta} + V(\varphi)] \tag{7.31}$$

and energy-momentum tensor

$$T_\mu^\nu = \varphi_{,\mu}\varphi^{,\nu} - \delta_\mu^\nu [\tfrac{1}{2}\varphi_{,\alpha}\varphi_{,\beta} g^{\alpha\beta} + V(\varphi)] \ . \tag{7.32}$$

In the general case of a theory with matter Lagrangian $\mathscr{L}_M$, the energy-momentum tensor is given by

$$\frac{\delta}{\delta g_{\mu\nu}} (\sqrt{-g}\, \mathscr{L}_M) = -\tfrac{1}{2}\sqrt{-g}\, T^{\mu\nu} \ . \tag{7.33}$$

Here $\delta/\delta g_{\mu\nu}$ stands for variation with respect to $g_{\mu\nu}$, with all other components of the metric held fixed.

The starting point is the usual Einstein-Hilbert action,

$$I = \int d^4x \sqrt{-g} \left[\frac{1}{16\pi G} R + \mathscr{L}_M \right] . \tag{7.34}$$

We consider I as a functional of all the functions that define a scalar metric perturbation about a FRW background,

$$I = I[\alpha(t), a(t); A(\mathbf{x},t), B(\mathbf{x},t); E(\mathbf{x},t), F(\mathbf{x},t); \varphi(\mathbf{x},t)] \ . \tag{7.35}$$

The central idea is to obtain all equations of motion, both for the background and for the perturbation, as variational equations of the above functional. It is not hard to guess the structure of the answer. Since $\alpha(t)$ and $a(t)$ parametrize the background, variation of I with respect to them will yield the FRW equations (7.12) and (7.13). Since α and a are space independent, the variational equations will be space-averaged equations. Thus $\rho^{(0)}$ and $p^{(0)}$ in Eqs. (7.12) and (7.13) will be defined as space-averaged quantities. Since $A(\mathbf{x},t)$ and $B(\mathbf{x},t)$ determine the dynamical metric components g_{ij}, variations with respect to these functions will give the dynamical equations of motion for perturbations. Similarly, variation with respect to $E(\mathbf{x},t)$ and $F(\mathbf{x},t)$ will give the Einstein constraint equations (we are using the ADM language of the Hamiltonian approach to gravity; see Arnowitt, Deser, and Misner, 1962). Finally, the result of varying with respect to $\varphi(\mathbf{x},t)$ will be the Klein-Gordon equation in curved space-time, the equation of motion for matter.

It is straightforward to work out the exact form of the equations (see Brandenberger, Kahn, and Press, 1983). Angular brackets stand for space averaging. For notational convenience we write down tne resulting equations in synchronous gauge.

For $\delta/\delta\alpha(t)$:

$$\left[\frac{\dot a}{a}\right]^2 = \frac{8\pi G}{3}\rho, \quad \rho = -\langle T_0^0 \rangle \ . \tag{7.36}$$

For $\delta/\delta a(t)$:

$$2\frac{\ddot a}{a} + \left[\frac{\dot a}{a}\right]^2 = -8\pi G p, \quad p = \tfrac{1}{3}\langle T_i^i \rangle \ . \tag{7.37}$$

For $\delta/\delta A(\mathbf{x},t)$:

$$(3A + \nabla^2 B)^{\cdot\cdot} + 3\frac{\dot a}{a}(3A + \nabla^2 B)^{\cdot} - \frac{\nabla^2 A}{a^2} = -8\pi G T_i^{i(1)} \ . \tag{7.38}$$

For $\delta/\delta B(\mathbf{x},t)$:

$$\ddot A + 3\frac{\dot a}{a}\dot A = -8\pi G \nabla^{-2} a^2 T_{,ij}^{ij(1)} \ . \tag{7.39}$$

For $\delta/\delta E(\mathbf{x},t)$:

$$\frac{\nabla^2 A}{a^2} - 3\frac{\dot a}{a}\dot A - \frac{\dot a}{a}\nabla^2 \dot B = 8\pi G T_0^{0(1)} \ . \tag{7.40}$$

For $\delta/\delta F(\mathbf{x},t)$:

$$\nabla^2 \dot A = +8\pi G a^2 T_{,i}^{0i(1)} \ . \tag{7.41}$$

For $\delta/\delta\varphi(\mathbf{x},t)$:

$$(-g)^{-1/2}(\sqrt{-g}\, g^{\alpha\beta}\varphi_{,\beta})_{,\alpha} = V'(\varphi) \ . \tag{7.42}$$

It is not hard to combine the dynamical equations (7.38) and (7.39) to obtain the desired second-order dynamical equation of motion for Φ_H. We work with the synchronous-gauge form of the equations and derive an explicitly gauge-invariant equation, which will thus be independent of the synchronous-gauge derivation.

By writing the general ansatz

$$\ddot\Phi_H = C_1 \dot\Phi_H + C_2 \Phi_H + C_3 \ , \tag{7.43}$$

inserting the definition of Φ_H, computing $\dot\Phi_H$ and $\ddot\Phi_H$, and simplifying by using Eqs. (7.38) and (7.39), we obtain (see Brandenberger, Kahn, and Press, 1983, for more details)

$$\ddot\Phi_H + (4 + 3c_s^2) H \dot\Phi_H + 3(c_s^2 - w) H^2 \Phi_H = 4\pi G I(t) \ . \tag{7.44}$$

Here $w(t)$ and $c_s^2(t)$ determine the equation of state of the background,

$$w(t)=\frac{p(t)}{\rho(t)} , \tag{7.45}$$

$$c_s^2(t)=\frac{\dot{p}(t)}{\dot{\rho}(t)} . \tag{7.46}$$

Thus all coefficients on the left-hand side of Eq. (7.44) depend only on the background. $I(t)$ is a gauge-invariant combination of matter source terms,

$$I(t)= -P_1+3c_s^2Ha^2\nabla^{-2}P_3+a^2H\dot{P}_2 +3H^2a^2(\tfrac{2}{3}-w+c_s^2)P_2+3Ha^2c_s^2(p+\rho)F , \tag{7.47}$$

$$P_1=\nabla^{-2}a^2T^{ij(1)}_{,ij} ,$$
$$P_2=\nabla^{-2}a^2(\delta_{ij}T^{ij(1)}-3\nabla^{-2}T^{ij(1)}_{,ij}) , \tag{7.48}$$
$$P_3=T^{0i(1)}_{,i} ,$$

P_2 is an anisotropic stress term (the diagonal piece of $T^{ij(1)}$ cancels). The first two terms on the left-hand side of Eq. (7.47) form an entropy perturbation term. The final term is absent in synchronous gauge, but is necessary to ensure gauge invariance of the sum of terms under transformations which do not preserve the synchronous-gauge condition. An entropy perturbation is a perturbation $\rho^{(1)}$ and $p^{(1)}$ that obeys a different equation of state from the background equation. The precise form (7.47) and (7.48) of the matter source term will not be important in what follows.

We can also combine the constraint equations (7.40) and (7.41) to obtain a gauge-invariant constraint equation for Φ_H:

$$\Phi_H=4\pi Ga^2\nabla^{-2}[\,T^{0(1)}_0-3a\dot{a}\nabla^{-2}T^{k(1)}_{0,k} +3a\dot{a}(p+\rho)F] . \tag{7.49}$$

In particular, we can consider a fixed scale $\mathbf{k}$ and evaluate the right-hand side at Hubble-radius crossing, i.e., for

$$H^{-1}=k^{-1}a(t) . \tag{7.50}$$

In comoving coordinates the second term vanishes. At Hubble-radius crossing the third is proportional to the first, with proportionality constant of order 1. $a^2\nabla^{-2}$ can be replaced by $-H^{-2}$, which is equal to $4\pi G\rho$ up to a factor $\frac{3}{2}$. Hence at horizon crossing, Φ_H equals the energy-density perturbation in comoving coordinates up to a factor of order 1 [Eq. (7.27)].

The gauge-invariant formalism is summarized in Fig. 26. The dynamical system consists of three blocks: the background metric is given by the scale factor; all information about scalar metric perturbations is contained in the gauge-invariant function $\Phi_H(\mathbf{x},t)$; the third block is matter, described in our case by a scalar field $\varphi(\mathbf{x},t)$. The evolution of the background metric is determined by the space-averaged part of matter via the FRW equations. The dynamics of matter is given by the Klein-Gordon equation in curved space-time, and the evolution of Φ_H is described by Eq. (7.44). This evolution depends on the background via the equation of state (w,c_s^2) and H, and on matter perturbations through $I(t)$. The initial-value constraint equation (7.49) completes the dynamical system.

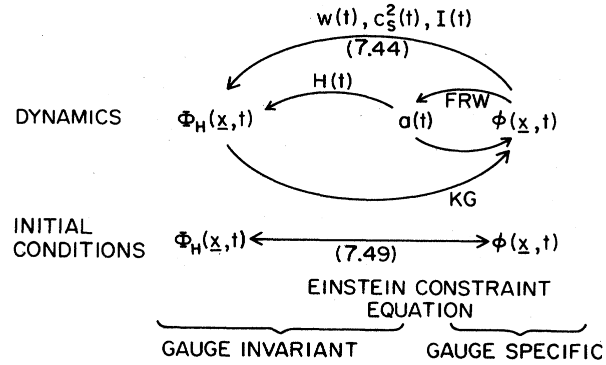


FIG. 26. Summary of the gauge-invariant formalism.

One technical point warrants further discussion. The perturbation $\delta\varphi(\mathbf{x},t)$ of a scalar function $\varphi(t)$ is not gauge invariant. It is impossible to define a gauge-invariant matter perturbation using matter variables alone. Sasaki (1983) advocates an approach that is completely gauge invariant. His matter perturbations contain gravitational terms to render the perturbation quantity gauge invariant. Brandenberger, Kahn, and Press (1983) chose the gauge-specific description of matter in which the physical division into geometrical and matter variables is kept. A complication in this approach is that in order to evolve $\varphi(\mathbf{x},t)$, we must first compute the metric perturbations in the specific gauge we choose. For example, in synchronous gauge, $A(\mathbf{x},t)$ and $B(\mathbf{x},t)$ are determined by explicit quadrature equations, which follow from Eqs. (7.41) and (7.25):

$$A(\mathbf{x},t)=-8\pi G\int^t a^2\nabla^{-2}P_3dt ,$$
$$B(\mathbf{x},t)=\int^t\frac{A-2\Phi_H}{a\dot{a}}dt . \tag{7.51}$$

The gauge-invariant formalism has recently been extended to the case of many uncoupled matter fields by Abbott and Wise (1984).

3. General comments and an approximate conservation law

We shall add a few comments on Eq. (7.44).

(a) Equation (7.44) is not the only form of the dynamical equation of motion. It is the natural one to emerge from our variational approach, and it is also mathematically simple, since it is an ordinary differential equation. On the other hand, to show clearly the two ways by which classical matter can generate metric perturbations, via entropy perturbations and via anisotropic stress perturbations, it is convenient to subtract the term

$$\frac{1}{a^2}\nabla^2\Phi_H \tag{7.52}$$

from both sides of Eq. (7.44) (using the gauge-invariant constraint equation). We obtain Bardeen's (1980) original equation of motion,

$$\ddot{\Phi}_H+(4+3c_s^2)H\dot{\Phi}_H+\left[\left[-\frac{\nabla^2}{a^2}+8\pi G\rho\right]c_s^2-8\pi Gp\right]\Phi_H=4\pi G[-P_1-c_s^2T_0^{0(1)}+a^2H\dot{P}_2+8\pi G(\tfrac{2}{3}\rho-p+\rho c_s^2)a^2P_2] . \tag{7.53}$$

The first two terms on the right-hand side of Eq. (7.53) are precisely the entropy perturbation. Equation (7.53) is also convenient for analyzing the evolution inside the Hubble radius. If there are no entropy and anisotropic stress perturbations, and if $w(t)=c_s^2(t)$, then Eq. (7.53) reduces to

$$\ddot{\Phi}_H+(4+3c_s^2)H\dot{\Phi}_H-\frac{\nabla^2}{a^2}c_s^2\Phi_H=0 . \tag{7.54}$$

Inside the Hubble radius, the Hubble damping term is negligible. We obtain the well-known result (see, for example, Peebles, 1980) that inside the Hubble radius Φ_H oscillates with the speed of sound c_s.

(b) Equation (7.44) clearly shows the two mechanisms by which Φ_H can increase outside the Hubble radius. First, there is homogeneous gravitational amplification of Φ_H due to a change in the equation of state of the universe (obtained by setting matter source terms to zero). In a single phase of the evolution of the universe, in general $w(t)=c_s^2(t)$, so the dominant mode of Eq. (7.44) will be time independent. During a phase transition, however, c_s^2-w will be nonzero and nonconstant, leading to a nontrivial amplification of Φ_H. The second mechanism is a nonvanishing matter source term. In particular, pressure perturbations outside the Hubble radius will generate energy-density perturbations at Hubble-radius crossing, in agreement with the Press-Vishniac (1980) analysis.

Note, however, that the division into the two mechanisms is not a gauge-invariant statement. Furthermore, the matter source terms are not freely specifiable at all times. If we specify them on some initial Cauchy surface, then time development will be determined by the complete dynamical system.

(c) Nonvanishing matter source terms can be taken into account by an easy Green's function method (Brandenberger and Kahn, 1984). We write

$$\Phi_H=\Phi_H^h+\Phi_H^p . \tag{7.55}$$

Here Φ_H^h is the solution of the homogeneous equation of motion with the given initial conditions, while Φ_H^p is the solution of the full equation (7.44) with vanishing initial conditions. If $f_1(t)$ and $f_2(t)$ are the two eigenmodes of the homogeneous equation, then

$$\Phi_H^h(t)=c_1f_2(t)+c_2f_2(t) , \tag{7.56}$$

$$\Phi_H^p(t)=-f_1(t)\int_{t_0}^t I(t')\varepsilon(t')f_2(t')dt' +f_2(t)\int_{t_0}^t I(t')\varepsilon(t')f_1(t')dt' , \tag{7.57}$$

where c_1 and c_2 are determined by the initial conditions. The factor $4\pi G$ on the right-hand side of Eq. (7.44) has been included in $I(t)$. Finally,

$$\varepsilon(t')=[f_2(t')\dot{f}_1(t')-\dot{f}_2(t')f_1(t')]^{-1} . \tag{7.58}$$

Bardeen, Steinhardt, and Turner (1983) argue that matter source terms are negligible outside the Hubble radius for inflationary universe models. It can, in fact, be shown (Brandenberger and Kahn, 1984) that keeping track of matter source terms using the above Green's function method does not change the result by more than a factor of order unity.

The analysis of the growth of perturbations outside the Hubble radius can be greatly simplified by deriving a conservation law for the growing mode solution of the homogeneous version of Eq. (7.44). We define a quantity ζ (Bardeen, Steinhardt, and Turner, 1983) by

$$\zeta=\frac{2}{3}\frac{\Phi_H+H^{-1}\dot{\Phi}_H}{1+w}+\Phi_H\left[1+\frac{2}{9}\left[\frac{k}{aH}\right]^2\frac{1}{1+w}\right] \tag{7.59}$$

Then, up to terms suppressed by an additional factor $(ka^{-1}H^{-1})^2$ outside the horizon,

$$\tfrac{2}{3}\dot{\zeta}H(1+w)=\ddot{\Phi}_H+(4+3c_s^2)H\dot{\Phi}_H+3(c_s^2-w)H^2\Phi_H . \tag{7.60}$$

Thus ζ is conserved outside the horizon [by Eq. (7.44), neglecting matter source terms]. In the following section we shall show that this conservation law leads to a very simple analysis of the growth of perturbations in inflationary universe models. This has also recently been noted by Lyth (1984).

C. Application to inflationary universe models

In this section we first apply the conservation law to give an analysis valid for a large class of inflationary universe models. Next, we present the arguments and problems concerning the generation of perturbations in the quantum period. We conclude by giving a more careful phase-by-phase analysis of the growth of Φ_H for the new inflationary universe model. Our analysis is based on the detailed analyses of Bardeen, Steinhardt, and Turner (1983) and Brandenberger and Kahn (1984). The first paper is reviewed by Turner (1982). Frieman and Turner

(1984) have independently confirmed the conclusions of Brandenberger and Kahn (1984). Earlier work on the growth of fluctuations in inflationary universe models is due to Kahn (1981) and Frieman and Will (1982). Kodama and Sasaki (1982) and Sasaki *et al.* (1982) have investigated the sources of perturbations in models with a first-order phase transition. Pagels (1983) considers vacuum fluctuations in quantum fields other than the Higgs field that drives inflation as the source of primordial energy-density contrasts.

1. Application of the conservation law

If we neglect matter source terms and hence can apply the conservation law $\zeta=$const it becomes trivial to obtain the value of Φ_H at final Hubble-radius crossing $t_f(k)$ in terms of initial data at Hubble-radius crossing $t_i(k)$ in the de Sitter phase (Brandenberger and Kahn, 1984). We usually drop the label k. Since in the radiation-dominated FRW phase $w=\frac{1}{3}$ and $\dot{\Phi}_H=0$, evaluating ζ at t_i and t_f yields

$$\Phi_H(t_f)=\frac{2}{5}\left[\frac{\frac{4}{3}\Phi_H(t_i)+H^{-1}\dot{\Phi}_H(t_i)}{1+w(t_i)}\right]+\tfrac{3}{5}\Phi_H(t_i)\ . \tag{7.61}$$

Since in inflationary universe models $1+w(t_i)\ll 1$, the second term is negligible.

The initial values can be expressed in terms of quantum field variables using the gauge-invariant constraint equation (7.49), the equation

$$1+w(t_i)=\frac{\dot{\varphi}^2(t_i)}{\rho}\ . \tag{7.62}$$

and the expressions for $T_0^{0(1)}$ and $T_{0,k}^k$ in terms of the scalar field (Brandenberger and Kahn, 1984). If we linearize in $\delta\varphi$ and use the equations of motion for the homogeneous background field φ_0 and for $\delta\varphi$,

$$\ddot{\varphi}+3H\dot{\varphi}_0=-V'(\varphi_0)\ , \tag{7.63}$$

$$\delta\ddot{\varphi}+3H\delta\dot{\varphi}=-V''(\varphi_0)\delta\varphi\ , \tag{7.64}$$

we obtain

$$\tfrac{4}{3}\Phi_H(t_i)+H^{-1}\dot{\Phi}_H(t_i)=\tfrac{1}{3}\Phi_H(t_i)=\frac{4\pi G}{3}a^2k^{-2}(\dot{\varphi}_0\delta\dot{\varphi}-\ddot{\varphi}_0\delta\varphi)\ . \tag{7.65}$$

When the slow-rolling approximation is valid, i.e., if $\ddot{\varphi}_0$ is negligible, then the first term in Eq. (7.65) dominates, and using the equation of motion for $\delta\varphi$, one obtains

$$\frac{\delta\rho}{\rho}(t_f)=\mathscr{O}(1)\frac{V''[\varphi(t_i)]\delta\varphi(t_i)}{\dot{\varphi}(t_i)H}\ . \tag{7.66}$$

When the slow-rolling approximation is not valid, then the second term in Eq. (7.65) dominates, and thus

$$\frac{\delta\rho}{\rho}(t_f)=\mathscr{O}(1)\frac{V'[\varphi(t_i)]}{\dot{\varphi}^2(t_i)}\delta\varphi(t_i)\ . \tag{7.67}$$

Equations (7.66) and (7.67) and are the main results of this section. They are applicable to a wide class of inflationary universe models, in particular models with a different cosmology than that of the new inflationary universe. The crucial requirement is that there exist an initial phase in which the scale factor increases more rapidly than the Hubble radius, so that scales of cosmological interest originate inside the Hubble radius.

If the scale that determines the curvature of the potential at the point $\varphi(t_i)$ when perturbations leave the Hubble radius is H, then Eq. (7.66) yields

$$\frac{\delta\rho}{\rho}(t_f)=\mathscr{O}(1)\frac{H}{\dot{\varphi}(t_i)}\delta\varphi(t_i)=\mathscr{O}(1)H\delta\tau(t_i)\ , \tag{7.68}$$

where $\delta\tau$ is the amplitude of the space-dependent time lag in the quantum field evolution given by

$$\varphi(\mathbf{x},t)=\varphi_0[t-\delta\tau(\mathbf{x})]\ . \tag{7.69}$$

The result (7.68) is originally due to Guth and Pi (1982).

Since in the new inflationary universe $\varphi(t_i)\sim H$ and $V''[\varphi(t_i)]\sim H^2$, Eq. (7.68) holds. There are academic examples of potentials for which Eq. (7.68) breaks down (Brandenberger and Kahn, 1984). An example is the reverse-hierarchy supersymmetric model [see Eq. (6.60)] with coefficients fine-tuned such that scales of cosmological interest leave the Hubble radius in the steep section of the potential, i.e., for $\varphi(t_i)\sim\mu$. In this case Eq. (7.67) gives

$$\frac{\delta\rho}{\rho}(t_f)=\mathscr{O}(1)\mu\delta\tau(t_i)\ . \tag{7.70}$$

2. Generation of perturbations

In our opinion the generation of classical matter perturbations in the de Sitter phase of an inflationary universe is not yet completely understood. The problem is to obtain classical fluctuations starting from a homogeneous and isotropic metric and a quantum state that does not break space-translational invariance.

Since the potential in the new inflationary universe is flat at the origin, most attempts so far have been to analyze the problem as one of a free, massless scalar field in the de Sitter phase of a FRW universe.

There is a simple but nonrigorous argument that gives the correct amplitude of primordial energy-density fluctuations (Bardeen, Steinhardt, and Turner, 1983): as discussed in Sec. V, there is Hawking radiation in the de Sitter phase of an inflationary universe. An observer detects a thermal flux of particles at a temperature

$$T_H=\frac{H}{2\pi}\ . \tag{7.71}$$

Hence, so the argument proceeds, there will be thermal energy-density fluctuation with amplitude

$$\frac{\delta\rho}{\rho}=\mathcal{O}(1)\frac{T^4}{\rho}=\mathcal{O}(1)\left[\frac{H}{\sigma}\right]^4 \tag{7.72}$$

(σ is the unification scale; see Sec. VI). The argument is nonrigorous, since there is no thermal bath of real particles, which would be necessary to complete the above analysis.

The source of Hawking radiation is vacuum fluctuations in the de Sitter phase. According to our present understanding, classical matter fluctuations are due to these vacuum fluctuations. Following Guth and Pi (1982), we introduce a classical field $\varphi_{\rm cl}(\mathbf{x},t)$ that consists of a homogeneous part $\varphi_0(t)$ and a perturbation $\delta\varphi(\mathbf{x},t)$,

$$\varphi_{\rm cl}(\mathbf{x},t)=\varphi_0(t)+\delta\varphi(\mathbf{x},t)\ . \tag{7.73}$$

$\varphi_0(t)$ will describe the spread of the vacuum-state wave functional; $\delta\varphi(\mathbf{x},t)$ will contain the information about spatial correlations in the vacuum-state wave functional. Since in a theory with symmetry under $\varphi\to-\varphi$ there is no drift in the expectation value of φ, i.e.,

$$\langle\psi_0\,|\,\varphi(\mathbf{x},t)\,|\,\psi_0\rangle=0\ , \tag{7.74}$$

$\varphi_0(t)$ will be given by the rms value (Hawking and Moss, 1983; see also Starobinsky, 1982; Vilenkin and Ford, 1982; Vilenkin, 1983; Linde, 1982c)

$$\varphi_0(t)=[\langle\psi_0\,|\,\varphi^2(\mathbf{x},t)\,|\,\psi_0\rangle]^{1/2}\ . \tag{7.75}$$

Since the Fourier transform of the spatial correlation function is the power spectrum $\langle\,|\varphi(\mathbf{k})|^2\rangle^{1/2}$, the Fourier transform of $\delta\varphi(\mathbf{x},t)$ must be

$$\delta\varphi(\mathbf{k},t)=\langle\,|\varphi(\mathbf{k},t)|^2\rangle^{1/2}\ . \tag{7.76}$$

The power spectrum, on the other hand, is equal to the coincident point—two-point function in momentum space, up to a normalization factor (for which we adopt the conventions of Peebles, 1980):

$$\langle\psi_0\,|\,\varphi^*(\mathbf{k})\varphi(\mathbf{l})\,|\,\psi_0\rangle=V^{-1}\delta^3(\mathbf{k}-\mathbf{l})\langle\,|\varphi(\mathbf{k},t)|^2\rangle\ . \tag{7.77}$$

This is essentially the definition given by Guth and Pi (1982).

Guth and Pi (1982) evaluate the formulas using the Green's functions of the Bunch-Davies (1978) vacuum of de Sitter space. We can also define a vacuum state appropriate for a FRW universe with a de Sitter phase of finite length (Brandenberger, 1984). Using the latter approach we obtain for a free, massless, minimally coupled scalar field

$$\delta\varphi(\mathbf{k},t_i)=V^{1/2}(2\pi)^{3/2}a^{-3/2}(t_i)H^{-1/2} \tag{7.78}$$

and (in agreement with Starobinsky, 1982; Vilenkin and Ford, 1982; Linde, 1982c; Vilenkin, 1983; and Hawking and Moss, 1983)

$$\varphi_0(t)=(2\pi)^{-1}H^{3/2}t^{1/2}\ . \tag{7.79}$$

The classical field $\varphi_{\rm cl}(\mathbf{x},t)$ can be used to construct a classical energy-momentum tensor for matter. In particular,

$$\frac{\delta\rho}{\rho}(\mathbf{k},t_i)=\dot\varphi_0(t_i)\delta\dot\varphi(\mathbf{k},t_i)\rho^{-1}\ . \tag{7.80}$$

By Eqs. (7.78) and (7.79) we obtain

$$\frac{\delta\rho}{\rho}(\underline{k},t_i)=\mathcal{O}(1)V^{1/2}k^{-3/2}\left[\frac{H}{\sigma}\right]^4\ . \tag{7.81}$$

By Eq. (7.4) this implies a scale-invariant Zeldovich spectrum:

$$\left[\frac{\delta M}{M}\right](\mathbf{k},t_i(\mathbf{k}))=\mathcal{O}(1)\left[\frac{H}{\sigma}\right]^4\ . \tag{7.82}$$

Definitions (7.75) and (7.76) are plausible, but not rigorously justified. This is the main problem with the approach discussed above. Such *ad hoc* definitions are, however, unavoidable in any semiclassical analysis of fluctuations. A further problem is to justify the time at which the transition to classical matter quantities is performed.

Hawking and Moss (1983) propose to extend the period during which matter is treated quantum mechanically. They must take quantum field nonlinearities into account. The details of this step have been, in our opinion correctly, questioned by Guth (1983) and Bardeen (1983). Both Bardeen (1983) and Moss (1984) have, in fact, shown that the semiclassical analysis outside the Hubble radius is valid. Guth and Pi (1984) and Bardeen and Hill (1984) are currently investigating inflationary scenarios in which the matter evolution is described quantum mechanically until reheating. Kahn and Press (1984) are taking a different approach and are trying to understand Hawking radiation and the generation of perturbations purely classically as an effect due to interactions and energy exchange between many classical fields (see also Press, 1983).

3. Analysis of the classical period for the new inflationary universe

The main goal of this section is to gain a better understanding of the physics of the amplification of perturbations.

Figure 26 shows the complete dynamical system. Its evolution is described by a complicated set of coupled differential equations, too complicated to admit an exact analytic solution. We propose the following approximation scheme: we pick a gauge in which metric perturbations are very small up to reheating, e.g., synchronous gauge. In this gauge, as a mathematical approximation, we solve the matter equation of motion without metric perturbations. This has already been done in Sec. VI [Eqs. (6.48)—(6.51)]. As a second step we determine the equation of state using

$$w(t)=-1+\frac{\dot\varphi^2(t)}{\rho}\ , \tag{7.83}$$

$$c_s^2(t)=-1-\frac{2\ddot\varphi(t)}{3H\dot\varphi(t)}\ . \tag{7.84}$$

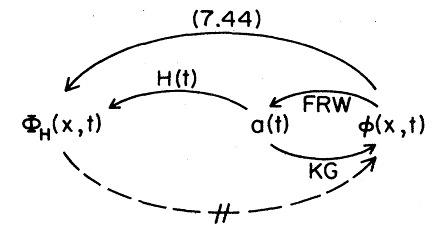


FIG. 27. Graphical sketch of the approximation scheme of Sec. VII.C for the determination of the growth of energy density fluctuations.

Finally, we insert the equation of state into the equation of motion (7.44) for Φ_H. Figure 27 paraphrases our approximation scheme graphically.

We now refer back to Eqs. (6.48)—(6.51) for an analysis of the evolution of $\varphi(t)$ in the two periods of the de Sitter phase in the new inflationary universe. In the first period $\varphi(t) \sim H$, $\dot\varphi(t) \sim H^2$ and thus $w(t) \sim -1$ and $c_s^2(t) \sim -1$. In the second period, for $t_B < t < t_i$, by Eq. (6.51)

$$\varphi(t) = \left[\frac{2}{\lambda}\right]^{1/2} f(t), \quad f(t) = [\alpha - (t - t_B)]^{-1} . \tag{7.85}$$

During this second period, $f(t)$ increases from $(\frac{3}{2})^{1/2} H$ to $\mathcal{O}(1)\sigma$. Hence $w(t)$ increases from -1 to 0, while $c_s^2(t)$ becomes very negative,

$$c_s^2(t) \sim -f(t) H^{-1} . \tag{7.86}$$

The change in the equation of state is sketched in Fig. 28. After reheating, $w(t)$ and $c_s^2(t)$ rapidly (period σ^{-1}) relax to their equilibrium values $\frac{1}{3}$ in a radiation-dominated FRW phase.

In the first de Sitter period $w(t) = c_s^2(t)$. From Eq. (7.44) it immediately follows that the dominant mode is the constant mode (the other is exponentially decaying). Hence Φ_H remains constant. In the second period $c_s^2(t)$ becomes very large and negative. To compensate in Eq. (7.44), $\ddot\Phi_H$ must become large and positive. The dominant mode will be a rapidly growing mode. Given the approximation $w = -1$ the eigenmodes of Eq. (7.44) in this period are

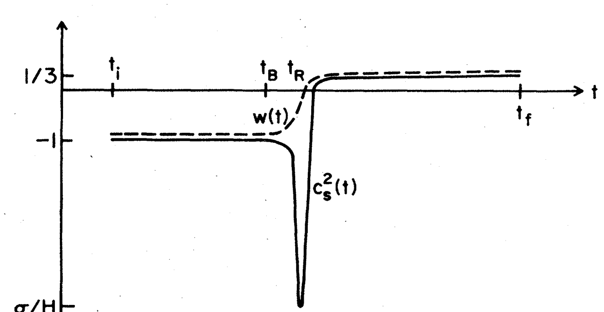


FIG. 28. Equation of state in the new inflationary universe (below -1 the scale is logarithmic).

$$f_1(t) = H \int_{t_B}^{t} \exp[H(t'-t)] \frac{E(t')}{E(t_B)} dt' , \tag{7.87}$$

$$f_2(t) = \exp[-H(t - t_B)] , \tag{7.88}$$

with $E(t) = \rho(t) + p(t) = \dot\varphi^2(t)$. Hence $\dot\Phi_H(t)$ increases by a factor $f^3(t_R)/f^3(t_B) = (\sigma/H)^3$. $\Phi_H(t)$ increases by $f^4(t_R)/f^4(t_B) = (\sigma/H)^4$.

In the FRW period, once again $w(t) = c_s^2(t)$, and therefore the dominant mode will be constant. Due to the mismatch in initial values of Φ_H and $\dot\Phi_H$ at t_R, Φ_H will increase by another factor σ/H immediately after reheating. The amplification of Φ_H is sketched in Fig. 29.

It is possible to check that the reheating period does not change the above analysis (Brandenberger and Kahn, 1984). Using the Green's-function method, we can show that matter source terms do not change the result by more than a factor of order 1.

Figures 27 and 28 clearly show that the rapid change in the equation of state is responsible for the growth of energy-density perturbations. On the other hand, the conservation equation (7.61) proves that the total amplification of perturbations is independent of the phase structure in the cosmological evolution between the times when perturbations of cosmological interest leave and reenter the Hubble radius. In particular, the amplification factor is independent of the details of reheating.

D. Conclusions

Inflationary universe models provide a mechanism which, for the first time, explains from first principles the origin of the primordial energy-density fluctuations required as initial conditions in theories of galaxy formation.

The crucial point is that in the de Sitter phase the causal horizon has constant physical distance, while the physical wavelength of a plane-wave perturbation expands exponentially. Thus, in contrast to the standard big bang model, fluctuations on all scales of present cosmological interest originate inside the causal horizon.

We have demonstrated that vacuum fluctuations, or more precisely spatial correlations in the wave functional of the quantum state of matter lead to classical matter inhomogeneities.

The initial energy-density fluctuations have a very

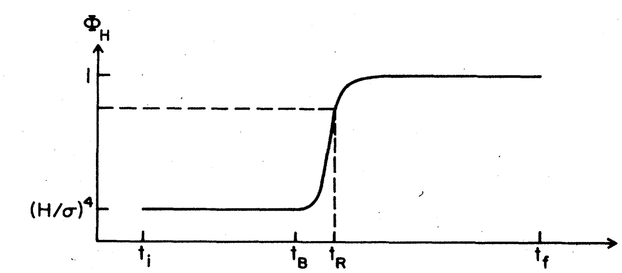


FIG. 29. Growth of Φ_H in the new inflationary universe (the vertical scale is logarithmic).

small magnitude. We measure them in terms of a gauge-invariant variable, which at Hubble-radius crossing is equal to the relative energy-density fluctuation in comoving coordinates up to a constant factor of order 1. In the original new inflationary universe model the value is of the order of 10^{-20}.

Due to the change in the equation of state of the background from almost de Sitter to radiation-dominated FRW, the perturbations increase by a large factor before they enter the Hubble radius. In many models, in particular in the case of the new inflationary universe, the amplification precisely offsets the initial suppression: the final amplitude of energy-density fluctuations is of the order 1, in conflict with requirements from the theories of galaxy formation. In order to obtain an acceptable magnitude, the particle physics models must in general be fine-tuned.

We show that the final amplitude of fluctuations depends only on the equation of state at the time t_i when scales of interest leave the Hubble radius in the de Sitter phase, and on the magnitude of perturbations at that time. In particular, the result does not depend on the details of reheating, nor on the presence of other possible phase transitions between time t_i and the time t_f when the scales reenter the horizon in the FRW phase.

The spectrum of fluctuations is predicted to be a scale-invariant Harrison (1970)-Zeldovich (1972) spectrum.

ACKNOWLEDGMENTS

It is my pleasure to thank David Groisser, Craig Iredale, Ronald Kahn, Rakesh Lowe, and Jennie Traschen for organizing and participating in the seminars at Harvard that generated my interest in the interplay between quantum field theory and cosmology. Working with Ronald Kahn and William Press on the theory of fluctuations in the early universe was lots of fun. Thanks again.

For comments on various drafts of this paper, for useful discussions, and for encouragement I am grateful to Andreas Albrecht, Luiz Alvarez-Gaume, Minos Axenides, James Bardeen, Douglas Eardley, Paul Ginsparg, Gary Hinshaw, Roman Jackiw, Arthur Jaffe, Palash Pal, Joseph Silk, Michael Turner, Simon White, and Mark Wise.

This research was supported in part by the National Science Foundation under Grant No. PHY77-27084, supplemented by funds from the National Aeronautics and Space Administration.

APPENDIX A: CONNECTION BETWEEN $W(J)$ AND $Z(J)$

In this appendix we shall review the well-known connection (2.5) between the full generating function $Z(J)$ and the generating function $iW(J)$ for connected Green's functions (see Itzykson and Zuber, 1980),

$$iW(J)=\ln Z(J)\ . \tag{A1}$$

This claim immediately reduces to a statement about the scattering matrix S,

$$S=T\exp\left[-i\int_{-\infty}^{\infty}dt\,H_I(t)\right]\ , \tag{A2}$$

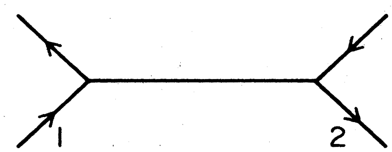
where $H_I(t)$ denotes the interaction picture Hamiltonian. After expanding S in a Taylor series we apply Wick's theorem to obtain a sum of normal ordered operators, each represented by a diagram (Wick diagram). In these diagrams a term in H_I is represented by a vertex. Each field operator in the term gives a leg attached to the vertex, and operator contractions are represented by linking the corresponding legs. In contrast to the case of Feynman diagrams, vertices in Wick diagrams are labeled. As an example, we consider

$$\mathscr{H}_I(x)=\lambda\bar\psi(x)\psi(x)\varphi(x)\ .$$

Then one of the terms of order λ^2, e.g.,

$$\lambda^2{:}\bar\psi(x_1)\psi(x_1)\overline{\varphi(x_1)\bar\psi(x_2)\psi(x_2)\varphi(x_2)}{:}$$

(where $\overline{\qquad}$ denotes contraction) is given by the Wick graph



In our case, H_I in Eq. (A2) includes the source term. The expansion of S yields

$$S=\sum\mathscr{W}\ , \tag{A3}$$

where $\mathscr{W}$ represents the Wick graphs. The Feynman diagrams arise by taking the expectation value of S between asymptotic in- and out-states. In particular

$$Z(J)=\langle 0\,|\,S\,|\,0\rangle=\sum\mathscr{W}_u\ , \tag{A4}$$

where $\mathscr{W}_u$ represents the Wick graphs with no uncontracted legs. Similarly,

$$iW(J)=\sum\mathscr{W}_{cu}\ , \tag{A5}$$

where $\mathscr{W}_{cu}$ represents the connected Wick graphs with no uncontracted legs.

Below we shall demonstrate that the sum of all Wick graphs is the exponential of the sum of all connected Wick graphs. If we consider the subset of graphs with n uncontracted legs, the corresponding statement naturally is false, except for $n=0$ (any combination of graphs with no uncontracted legs gives back a vacuum graph, and conversely all connected components of a vacuum graph have no external legs). Thus

$$\begin{aligned}Z(J)&=\langle 0\,|\,S\,|\,0\rangle=\sum\mathscr{W}_u\\&=\exp\sum\mathscr{W}_{cu}\\&=\exp iW(J)\ .\end{aligned} \tag{A6}$$

This completes the proof of (A1).

To prove that the sum of all Wick diagrams is the exponential of the sum of all connected Wick graphs, we shall for notational simplicity consider only the case in which H_I contains a single term and introduce the following notation: We define two diagrams as belonging to the same pattern P if they differ by a permutation of the vertices. The symmetry number $S(P)$ of P is the number of permutations that leave the diagram unchanged. Obviously

$$\sum_P \sum_{\text{graphs in pattern } P} \frac{1}{n(P)!}:O(P):=\sum_P \frac{:O(P):}{S(P)}(=S) , \quad \text{(A7)}$$

where $n(P)$ is the number of vertices [which equals the order in the Taylor expansion of Eq. (A2)], and $:O(P):$ is the operator the Wick diagram represents.

We now consider a pattern P consisting of n_i connected Wick diagrams of the (connected) pattern P_i ($i=1,2,\ldots,$ is an ordering of all connected patterns). Obviously

$$O(P)=\prod_{i=1}^{\infty} O(P_i)^{n_i} . \quad \text{(A8)}$$

Using Eqs. (A3), (A6), and (A7), we obtain

$$S=\sum_P \frac{:O(P):}{S(P)}=\sum_{\substack{n_i=0\\ i=1}}^{\infty\ \infty} \frac{:\prod_{i=1}^{\infty} O(P_i)^{n_i}:}{\prod_{i=1}^{\infty}[S(P_i)^{n_i} n_i!]}$$

$$=\prod_{i=1}^{\infty} \sum_{n_i=0}^{\infty} \frac{1}{n_i!}:\left[\frac{O(P_i)}{S(P_i)}\right]^{n_i}:=:\exp\sum_{i=1}^{\infty} \frac{O(P_i)}{S(P_i)}: . \quad \text{(A9)}$$

In the second step we used the facts that $S(P)$ factorizes as in (A8) and that permuting graphs of a given pattern will not give a different Wick graph. Now by (A7) and (A9),

$$S=:\exp\sum \mathscr{W}_c: , \quad \text{(A10)}$$

where $\mathscr{W}_c$ represents the connected Wick graphs.

APPENDIX B: LOOP EXPANSION AND $\hbar$ EXPANSION

In this section we shall prove the equivalence of the loop expansion and the $\hbar$ expansion. Consider a theory with generating functional

$$Z(J)=N\int[d\varphi]\exp\left[\frac{i}{\hbar}S(\varphi,J)\right] . \quad \text{(B1)}$$

The vertex factors are given by the coefficients of the interaction terms of $(i/\hbar)S(\varphi,J)$. Therefore, they are proportional to $\hbar^{-1}$. Propagators, on the other hand, are proportional to $\hbar$, since they are the inverse of operators in the exponent of Eq. (B1) [see Eq. (2.9)]. A graph with I internal lines and V vertices is thus proportional to

$$\hbar^{I-V} . \quad \text{(B2)}$$

For a connected graph, each vertex fixes one internal momentum except for one, which gives overall energy-momentum conservation. The number of loops equals the number of free momenta. Thus for connected graphs

$$L=I-V+1 . \quad \text{(B3)}$$

Combining Eqs. (B2) and (B3), we see that for connected graphs the $\hbar$ dependence is

$$\hbar^{L-1} , \quad \text{(B4)}$$

which demonstrates the equivalence of $\hbar$ and loop expansions.

APPENDIX C: DETERMINATION OF A DETERMINANT

Here we evaluate Eq. (4.13). Since only ratios of the determinants exist, we write

$$N\det{}^{-1/2}(B+\omega^2 I)=N'\det{}^{-1/2}(I+\omega^2 B^{-1}) , \quad \text{(C1)}$$

with $B=-\partial_t^2$. B acts on the space of functions with period T [see Eq. (4.2)]. Hence its eigenvalues are

$$\lambda_n=\left[\frac{\pi n}{T}\right]^2 , \quad n\in Z_+ .$$

Thus

$$\det(I+\omega^2 B^{-1})=\prod_{n=1}^{\infty}\left[1+\left[\frac{\omega T}{\pi n}\right]^2\right] . \quad \text{(C2)}$$

But the representation of $\sin\pi z$ as an infinite product (see, for example, Courant, 1937, p. 445) is

$$\sin\pi z=\pi z\prod_{n=1}^{\infty}\left[1-\frac{z^2}{n^2}\right] .$$

Therefore, setting $z=i\omega T/\pi$,

$$\det{}^{-1/2}(I+\omega^2 B^{-1})=\left[\frac{\sin i\omega T}{i\omega T}\right]^{-1/2}$$

$$=(\omega T)^{1/2}[\sinh(\omega T)]^{-1/2} . \quad \text{(C3)}$$

For ωT large,

$$\det{}^{-1/2}(I+\omega^2 B^{-1})\sim(2\omega T)^{1/2}\exp(-\omega T/2) . \quad \text{(C4)}$$

Now Eq. (4.13) follows by adjusting the normalization constant N.

REFERENCES

Abbott, L. F., 1981a, Nucl. Phys. B **185**, 189.
Abbott, L. F., 1981b, Nucl. Phys. B **185**, 233.
Abbott, L. F., 1982, Acta Phys. Pol. B **13**, 33.
Abbott, L. F., and C. Burges, 1983, Phys. Lett. **131B**, 49.
Abbott, L. F., E. Farhi, and M. B. Wise, 1982, Phys. Lett. **117B**, 29.
Abbott, L. F., and M. B. Wise, 1984, Nucl. Phys. B **237**, 226.

Abers, E., and B. Lee, 1973, Phys. Rep. **9**, 1.

Abrikosov, A. A., L. P. Gorkov, and I. E. Dzyaloshinski, 1959, Zh. Eksp. Teor. Fiz. **36**, 900 [Sov. Phys.—JETP **9**, 636 (1959)].

Abrikosov, A. A., L. P. Gorkov, and I. E. Dzyaloshinski, 1965, *Methods of Quantum Theory in Statistical Physics* (Prentice-Hall, Englewood Cliffs).

Akhiezer, I., and S. Peletminskii, 1960, Zh. Eksp. Teor. Fiz. **38**, 1829 [Sov. Phys.—JETP **11**, 1316 (1960)].

Albrecht, A., 1984, in *Fifth Workshop on Grand Unification*, edited by K. Kang, H. Fried, and P. Frampton (World Scientific, Singapore, in press).

Albrecht, A., S. Dimopoulos, W. Fischler, E. W. Kolb, S. Raby, and P. J. Steinhardt, 1982, in *Proceedings of the 3rd Marcel Grossman Meeting on the Recent Developments in General Relativity*, edited by Hu Ning (North-Holland, Amsterdam), p. 511.

Albrecht, A., S. Dimopoulos, W. Fischler, E. W. Kolb, S. Raby, and P. J. Steinhardt, 1983, Nucl. Phys. B **229**, 528.

Albrecht, A., L. Jensen, and P. J. Steinhardt, 1984, Nucl. Phys. B **239**, 290.

Albrecht, A., and P. J. Steinhardt, 1982, Phys. Rev. Lett. **48**, 1220.

Albrecht, A., and P. J. Steinhardt, 1983, Phys. Lett. **131B**, 45.

Albrecht, A., P. J. Steinhardt, M. S. Turner, and F. Wilczek, 1982, Phys. Rev. Lett. **48**, 1437.

Allen, B., 1983, Nucl. Phys. B **226**, 228.

Allen, B., 1984, Ann. Phys. (N.Y.) (in press).

Arnold, V. I., 1978, *Mathematical Methods of Classical Mechanics*, English ed. (Springer, New York).

Arnowitt, R., S. Deser, and C. W. Misner, 1962, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York), p. 227.

Balian, R., G. Parisi, and A. Voros, 1978, Phys. Rev. Lett. **41**, 1141.

Bardeen, James, 1980, Phys. Rev. D **22**, 1882.

Bardeen, James, 1983, "New Inflation Revisted—A Semiclassical Approach" (unpublished).

Bardeen, James, P. J. Steinhardt, and M. S. Turner, 1983, Phys. Rev. D **28**, 679.

Bardeen, W., and C. Hill, 1984, private communication.

Barrow, J. D., and M. S. Turner, 1982, Nature **298**, 801.

Bernard, C., 1974, Phys. Rev. D **9**, 3313.

Bieberbach, L., 1953, *Theorie der Gewoehnlichen Differentialgleichungen* (Springer, Berlin).

Billoire, A., and K. Tamvakis, 1982, Nucl. Phys. B **200**, 329.

Bludman, S., and M. Ruderman, 1977, Phys. Rev. Lett. **38**, 255.

Bond, J. R., and A. S. Szalay, 1983, Astrophys. J. **274**, 443.

Boulware, D. G., 1981, Phys. Rev. D **23**, 389.

Boulware, D. G., 1979, in *Recent Developments in Gravitation, Cargèse 1978*, edited by Maurice Lévy and S. Deser (Plenum, New York), p. 175.

Brandenberger, R., 1983, Phys. Lett. **129B**, 397.

Brandenberger, R., 1984, Nucl. Phys. B **245**, 328.

Brandenberger, R., and R. Kahn, 1982, Phys. Lett. **119B**, 75.

Brandenberger, R., and R. Kahn, 1984, Phys. Rev. D **29**, 2172.

Brandenberger, R., R. Kahn, and W. H. Press, 1983, Phys. Rev. D **28**, 1809.

Breit, J. D., S. Gupta, and A. Zaks, 1983, Phys. Rev. Lett. **51**, 1007.

Brown, L. S., 1977, Phys. Rev. D **15**, 1469.

Brown, L. S., and J. Cassidy, 1977a, Phys. Rev. D **15**, 2810.

Brown, L. S., and J. Cassidy, 1977b, Phys. Rev. D **16**, 1712.

Bunch, T. S., and P. C. W. Davies, 1977a, Proc. R. Soc. London Ser. A **356**, 569.

Bunch, T. S., and P. C. W. Davies, 1977b, Proc. R. Soc. London Ser. A **357**, 381.

Bunch, T. S., and P. C. W. Davies, 1978, Proc. R. Soc. London Ser. A **360**, 117.

Buras, A., J. Ellis, M. K. Gaillard, and D. V. Nanopoulos, 1978, Nucl. Phys. B **135**, 66.

Callan, C. G., and S. Coleman, 1977, Phys. Rev. D **16**, 1762.

Chibisov, G., and V. Mukhanov, 1981, Zh. Eksp. Teor. Fiz. Pis'ma Red. **33**, 549 [JETP Lett. **33**, 532 (1981)].

Christensen, S. M., 1976, Phys. Rev. D **14**, 2490.

Christensen, S. M., and S. A. Fulling, 1977, Phys. Rev. D **15**, 2088.

Coleman, S., 1973, in *Laws of Hadronic Matter (Erice, 1973)*, edited by A. Zichichi (Academic, New York, 1975), p. 139.

Coleman, S., 1975, in *New Phenomena in Subnuclear Physics, Part A (Erice, 1975)*, edited by A. Zichichi (Plenum, New York, 1977), p. 297.

Coleman, S., 1977, Phys. Rev. D **15**, 2929.

Coleman, S., 1977, in *The Whys of Subnuclear Physics (Erice, 1977)*, edited by A. Zichichi (Plenum, New York, 1979), p. 805.

Coleman, S., 1981, in *The Unity of the Fundamental Interactions (Erice, 1981)*, edited by A. Zichichi (Plenum, New York), p. 21.

Coleman, S., and F. DeLuccia, 1980, Phys. Rev. D **21**, 3305.

Coleman, S., V. Glaser, and A. Martin, 1978, Commun. Math. Phys. **58**, 211.

Coleman, S., R. Jackiw, and H. D. Politzer, 1974, Phys. Rev. D **10**, 2491.

Coleman, S., and E. J. Weinberg, 1973, Phys. Rev. D **7**, 1888.

Cook, G. P., and K. T. Mahanthappa, 1981, Phys. Rev. D **23**, 1321.

Cornwall, J. M., R. Jackiw, and E. Tomboulis, 1974, Phys. Rev. D **10**, 2428.

Courant, R., 1937, *Differential and Integral Calculus* (Interscience, New York), Vol. 1.

Davies, P. C. W., 1975, J. Phys. A **8**, 365.

Davies, P. C. W., 1976, Proc. R. Soc. London Ser. A **351**, 129.

Davies, P. C. W., 1980, in *General Relativity and Gravitation One Hundred Years after the Birth of Albert Einstein*, edited by A. Held (Plenum, New York), Vol. 1, p. 255.

Davies, P. C. W., S. A. Fulling, and W. G. Unruh, 1976, Phys. Rev. D **13**, 2720.

Davies, P. C. W., S. A. Fulling, S. M. Christensen, and T. S. Bunch, 1977, Ann. Phys. (N.Y.) **109**, 108.

Denardo, G., and E. Spallucci, 1982, Nuovo Cimento Lett. **33**, 336.

DeWitt, B. S., 1965, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York).

DeWitt, B. S., 1975, Phys. Rep. **19**, 295.

DeWitt, B. S., 1979, in *General Relativity, An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel (Cambridge University, Cambridge), p. 680.

DeWitt, B. S., 1981, in *Quantum Gravity 2, A Second Oxford Symposium*, edited by C. J. Isham, R. Penrose, and D. W. Sciama (Cambridge University, Cambridge), p. 449.

Dimopoulos, S., and H. Georgi, 1981, Nucl. Phys. B **193**, 150.

Dimopoulos, S., and S. Raby, 1983, Nucl. Phys. B **219**, 479.

Dimopoulos, S., and L. Susskind, 1978, Phys. Rev. D **18**, 4500.

Dolan, L., and R. Jackiw, 1974, Phys. Rev. D **9**, 3320.

Dolgov, A. D., and A. D. Linde, 1982, Phys. Lett. **116B**, 329.

Dolgov, A. D., and Ya. B. Zeldovich, 1981, Rev. Mod. Phys. **53**, 1.

Dowker, J., and R. Critchley, 1976, Phys. Rev. D **13**, 3224.
Einhorn, M. B., and K. Sato, 1981, Nucl. Phys. B **180**, 385.
Einhorn, M. B., D. L. Stein, and D. Toussaint, 1980, Phys. Rev. D **21**, 3295.
Ellis, J., 1982, in *The Very Early Universe, Proceedings of the Nuffield Workshop, Cambridge. . . 1982,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University, Cambridge, 1983), p. 91.
Ellis, J., A. D. Linde, and D. V. Nanopoulos, 1982, Phys. Lett. **118B**, 59.
Ellis, J., D. V. Nanopoulos, K. A. Olive, and K. Tamvakis, 1982, Phys. Lett. **118B**, 335.
Ellis, J., D. V. Nanopoulos, K. A. Olive, and K. Tamvakis, 1983a, Nucl. Phys. B **221**, 524.
Ellis, J., D. V. Nanopoulos, K. A. Olive, and K. Tamvakis, 1983b, Phys. Lett. **120B**, 331.
Ellis, J., and K. A. Olive, 1983, Nature **303**, 679.
Englert, F., and R. Brout, 1964, Phys. Rev. Lett. **13**, 321.
Fabbri, R., and M. D. Pollock, 1983, Phys. Lett. **125B**, 445.
Faddeev, L. D., and V. N. Popov, 1967, Phys. Lett. **25B**, 29.
Fetter, A. L., and J. D. Walecka, 1971, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York).
Fischetti, M., B. L. Hu, and J. B. Hartle, 1979, Phys. Rev. D **20**, 1757.
Ford, L. H., and D. J. Toms, 1982, Phys. Rev. D **25**, 1510.
Fradkin, E. S., 1959, Zh. Eksp. Teor. Fiz. **36**, 1286 [Sov. Phys.—JETP **9**, 912 (1959)].
Frampton, P. H., 1976, Phys. Rev. Lett. **37**, 1380.
Frieman, J. A., and M. S. Turner, 1984, Phys. Rev. D **30**, 265.
Frieman, J. A., and C. M. Will, 1982, Astrophys. J **259**, 437.
Fulling, S. A., 1973, Phys. Rev. D **7**, 2850.
Fulling, S. A., and P. C. W. Davies, 1976, Proc. R. Soc. London Ser. A **348**, 393.
Fulling, S. A., and L. Parker, 1974, Ann. Phys. (N.Y.) **87**, 176.
Fulling, S. A., L. Parker, and B. L. Hu, 1974, Phys. Rev. D **10**, 3905.
Gelmini, G. B., D. V. Nanopoulos, and K. A. Olive, 1983, Phys. Lett. **131B**, 53.
Georgi, H., and S. L. Glashow, 1974, Phys. Rev. Lett. **32**, 438.
Gibbons, G. W., 1979, in *General Relativity, An Einstein Centenary Survey,* edited by S. W. Hawking and W. Israel (Cambridge University, Cambridge), p. 639.
Gibbons, G. W., and S. W. Hawking, 1977, Phys. Rev. D **15**, 2738.
Gibbons, G. W., and M. Perry, 1978, Phys. Rev. Lett. **36**, 985.
Gibbons, G. W., and M. Perry, 1978, Proc. R. Soc. London Ser. A **358**, 467.
Ginsparg, P., 1982, Phys. Lett. **112B**, 45.
Glashow, S. L., 1961, Nucl. Phys. **22**, 579.
Glimm, J., and A. Jaffe, 1981, *Quantum Physics* (Springer, New York).
Goldstone, J., A. Salam, and S. Weinberg, 1962, Phys. Rev. **127**, 965.
Goncharov, A. S., and A. D. Linde, 1984, Phys. Lett. **139B**, 27.
Grishchuk, P., 1974, Zh. Eksp. Teor. Fiz. **67**, 825 [Sov. Phys.—JETP **40**, 409 (1974)].
Gupta, S., and H. R. Quinn, 1984, Phys. Rev. D **29**, 2791.
Guralnik, G. S., C. R. Hagen, and T. W. B. Kibble, 1964, Phys. Rev. Lett. **13**, 585.
Guth, A. H., 1981, Phys. Rev. D **23**, 347.
Guth, A. H., 1982, in *The Very Early Universe: Proceedings of the Nuffield Workshop, Cambridge. . . 1982,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University, Cambridge), p. 171.
Guth A. H., 1983, private communication.
Guth, A. H., 1984, "The New Inflationary Universe, 1984," paper delivered at the Inner Space/Outer Space Conference, Fermilab, May 2—5, 1984, to be published.
Guth, A. H., and S.-Y. Pi, 1982, Phys. Rev. Lett. **49**, 1110.
Guth, A. H., and S.-Y. Pi, 1984, "The Behavior of the Higgs Field in the New Inflationary Universe," paper delivered at the Inner Space/Outer Space Conference, Fermilab, May 2—5, 1984, to be published.
Guth, A. H., and P. J. Steinhardt, 1984, Scientific American **250**, No. 5, 116.
Guth, A. H., and S.-H. Tye, 1980, Phys. Rev. Lett. **44**, 631.
Guth, A. H., and E. J. Weinberg, 1980, Phys. Rev. Lett. **45**, 1131.
Guth, A. H., and E. J. Weinberg, 1981, Phys. Rev. D **23**, 876.
Guth, A. H., and E. J. Weinberg, 1983, Nucl. Phys. B **212**, 321.
Harrison, E., 1970, Phys. Rev. D **1**, 2726.
Hartle, J. B., and B. L. Hu, 1979, Phys. Rev. D **20**, 1772.
Hartle, J. B., and B. L. Hu, 1980, Phys. Rev. D **21**, 2756.
Hartle, J. B., and S. W. Hawking, 1976, Phys. Rev. D **13**, 2188.
Hawking, S. W., 1966, Astrophys. J. **145**, 544.
Hawking, S. W., 1974, Nature **248**, 30.
Hawking, S. W., 1975, Commun. Math. Phys. **43**, 199.
Hawking, S. W., 1977, Commun. Math. Phys. **55**, 133.
Hawking, S. W., 1982, Phys. Lett. **115B**, 295.
Hawking, S. W., 1984, Phys. Lett. **134B**, 403.
Hawking, S. W., and I. G. Moss, 1982, Phys. Lett. **110B**, 35.
Hawking, S. W., and I. G. Moss, 1983, Nucl. Phys. B **224**, 180.
Hawking, S. W., I. G. Moss, and J. M. Stewart, 1982, Phys. Rev. D **26**, 2681.
Hawking, S. W., and G. F. R. Ellis, 1973, *The Large Scale Structure of Space-Time* (Cambridge University, Cambridge).
Heisenberg, W., and H. Euler, 1936, Z. Phys. **98**, 714.
Higgs, P. W., 1964a, Phys. Lett. **12**, 132.
Higgs, P. W., 1964b, Phys. Rev. Lett. **13**, 508.
Hill, C., 1983, unpublished.
Hinshaw, G., 1984, private communication.
Holman, R., P. Ramond, and C. G. Ross, 1984, Phys. Lett. **137B**, 343.
Hosoya, A., and M. Sakagami, 1984, Phys. Rev. D **29**, 2228.
Hut, P., and F. R. Klinkhamer, 1981, Phys. Lett. **104B**, 439.
Ignatiev, A. Yu., N. V. Krasnikov, V. A. Kuzmin, and A. N. Tavkhelidze, 1978, Phys. Lett. **76B**, 436.
Iliopoulos, J., C. Itzykson, and A. Martin, 1975, Rev. Mod. Phys. **47**, 165.
Itzykson, C., and J. Zuber, 1980, *Quantum Field Theory* (McGraw-Hill, New York).
Izawa, M., and K. Sato, 1982, Prog. Theor. Phys. **68**, 1574.
Jackiw, R., 1974, Phys. Rev. D **9**, 1686.
Jackiw, R., 1975, in *Theories and Experiments in High-Energy Physics (Orbis Scientiae II),* edited by B. Kursunoglu, A. Perlmutter, and S. M. Widmayer (Plenum, New York), p. 371.
Jackiw, R., and A. Kerman, 1979, Phys. Lett. **71A**, 158.
Jona-Lasinio, G., 1964, Nuovo Cimento **34**, 1790.
Kadanoff, L. P., and G. Baym, 1962, *Quantum Statistical Mechanics* (Benjamin, New York).
Kahn, R., 1981, "The Decay of Correlations in De Sitter Space," unpublished.
Kahn, R., and R. Brandenberger, 1984, Phys. Lett. **141B**, 317.
Kahn, R., and W. H. Press, 1984, private communication.
Kazanas, D., 1980, Astrophys. J. **241**, L59.
Khlopov, M. Yu., and A. D. Linde, 1984, Phys. Lett. **138B**, 265.
Kibble, T. W. B., 1976, J. Phys. A **9**, 1387.

Kibble, T. W. B., G. Lazarides, and Q. Shafi, 1982, Phys. Lett. **113B**, 237.
Kirzhnits, D. A., 1972, Zh. Eksp. Teor. Fiz. Pis'ma Red. **15**, 745 [JETP Lett. **15**, 529 (1972)].
Kirzhnits, D. A., and A. D. Linde, 1972, Phys. Lett. **42B**, 471.
Kirzhnits, D. A., and A. D. Linde, 1974, Zh. Eksp. Teor. Fiz. **67**, 1263 [Sov. Phys.—JETP **40**, 628 (1974)].
Kirzhnits, D. A. and A. D. Linde, 1975, Ann. Phys. (N.Y.) **101**, 195.
Kodama, H., and M. Sasaki, 1982, Prog. Theor. Phys. **68**, 1398.
Kounnas, C., J. Leon, and M. Quiros, 1983, Phys. Lett. **129B**, 67.
Kubo, R., 1957, J. Phys. Soc. Jpn. **12**, 570.
Langacker, P., 1981, Phys. Rep. **72**, 185.
Langer, J. S., 1967, Ann. Phys. (N.Y.) **41**, 108.
Langer, J. S., 1969, Ann. Phys. (N.Y.) **54**, 258.
Langer, J. S., 1974, Physica **73**, 61.
Lapedes, A., 1978, J. Math. Phys. **19**, 2289.
Lapedes, A., and E. Mottola, 1982, Nucl. Phys. B **203**, 58.
Lifshitz, E. M., 1946, J. Phys. (Moscow) **10**, 116.
Lifshitz, E. M., and I. M. Khalatnikov, 1963, Adv. Phys. **12**, 185.
Linde, A. D., 1974, Zh. Eksp. Teor. Fiz. Pis'ma Red. **19**, 320 [JETP Lett. **19**, 183 (1974)].
Linde, A. D., 1976, Zh. Eksp. Teor. Fiz. Pis'ma Red. **23**, 73 [JETP Lett. **23**, 64 (1974)].
Linde, A. D., 1977, Phys. Lett. **70B**, 306.
Linde, A. D., 1979, Rep. Prog. Phys. **42**, 389.
Linde, A. D., 1981a, Phys. Lett. **99B**, 391.
Linde, A. D., 1981b, Phys. Lett. **100B**, 37.
Linde, A. D., 1982a, Phys. Lett. **108B**, 389.
Linde, A. D., 1982b, Phys. Lett. **114B**, 431.
Linde, A. D., 1982c, Phys. Lett. **116B**, 335.
Linde, A. D., 1982d, Phys. Lett. **116B**, 340.
Linde, A. D., 1982e, in *The Very Early Universe, Proceedings of the Nuffield Workshop, Cambridge. . . 1982,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University, Cambridge), p. 205.
Linde, A. D., 1983a, Phys. Lett. **129B**, 177.
Linde, A. D., 1983b, Nucl. Phys. B **216**, 421.
Linde, A. D., 1983c, Phys. Lett. **132B**, 137.
Linde, A. D., 1983d, Zh. Eksp. Teor. Fiz. Pis'ma Red. **37**, 606 [JETP Lett. **37**, 724 (1983)].
Linde, A. D., 1983e, Zh. Eksp. Teor. Fiz. Pis'ma Red. **38**, 149 [JETP Lett. **38**, 176 (1983)].
Linde, A. D., 1983f, Phys. Lett. **131B**, 330.
Linde, A. D., 1984, Rep. Prog. Phys. **47**, 925.
Linde, A. D., and A. Goncharov, 1984, Zh. Eksp. Teor. Fiz. **86**, 1594.
Lukash, V. N., 1980, Zh. Eksp. Teor. Fiz. **79**, 1601 [Sov. Phys.—JETP **52**, 807 (1980)].
Lyth, D. H., 1984, "Large Scale Energy Density Fluctuations and Inflation," Lancaster preprint.
Martin, P. C., and J. Schwinger, 1959, Phys. Rev. **115**, 1342.
Matsubara, T., 1955, Prog. Theor. Phys. **14**, 351.
Mazenko, G. F., W. G. Unruh, and R. M. Wald, 1984, "Does a Phase Transition in the Early Universe Produce the Conditions Needed for Inflation," Enrico Fermi preprint.
Morse, P., and H. Feshbach, 1953, *Methods of Theoretical Physics* (McGraw-Hill, New York), Vol. 1.
Moss, I. G., 1984, Nucl. Phys. B **238**, 436.
Mottola, E., 1983, Phys. Rev. D **27**, 2294.
Mottola, E., and A. Lapedes, 1983, Phys. Rev. D **27**, 2285.
Nanopoulos, D. V., 1983, "Unified Field Theories and the Early Universe," CERN preprint TH 3778.
Nanopoulos, D. V., K. A. Olive, and K. Tamvakis, 1982, Phys. Lett. **115B**, 15.
Nanopoulos, D. V., and K. Tamvakis, 1982, Phys. Lett. **110B**, 449.
Nanopoulos, D. V., K. A. Olive, and M. Srednicki, 1983, Phys. Lett. **127B**, 30.
Nanopoulos, D. V., K. A. Olive, M. Srednicki, and K. Tamvakis, 1983a, Phys. Lett. **123B**, 41.
Nanopoulos, D. V., K. A. Olive, M. Srednicki, and K. Tamvakis, 1983b, Phys. Lett. **124B**, 171.
Olive, K. A., and N. Turok, 1982, Phys. Lett. **117B**, 193.
Olive, K. A., 1983, in *Galaxies and the Early Universe: Proceedings of the 18th Rencontre de Moriond, Astrophysics Meeting,* edited by J. Audouze and J. Tran Thanh Van (Reidel, Dordrecht), p. 3; and lectures at the 1983 Aspen Summer School of Physics.
Olive, K. A., and M. Srednicki, 1984, "Finite Temperature Corrections in $N=1$ Supergravity," University of California at Santa Barbara preprint.
Olson, D. W., 1976, Phys. Rev. D **14**, 327.
Ovrut, B. A., and P. J. Steinhardt, 1983, Phys. Lett. **133B**, 161.
Pagels, H. R., 1983, Astrophys. Lett. **23**, 151.
Parke, S., 1983, Phys. Lett. **121B**, 313.
Parker, L., 1966, Ph.D. thesis, Harvard University.
Parker, L., 1968, Phys. Rev. Lett. **21**, 562.
Parker, L., 1969, Phys. Rev. **183**, 1057.
Parker, L., 1971, Phys. Rev. D **3**, 346.
Parker, L., 1977, in *Asymptotic Structure of Space-Time,* edited by F. Paul Esposito and Louis Witten (Plenum, New York), p. 107.
Parker, L., 1979, in *Recent Developments in Gravitation, Cargèse 1978,* edited by Maurice Lévy and S. Deser (Plenum, New York), p. 219.
Parker, L., and S. A. Fulling, 1974, Phys. Rev. D **9**, 341.
Patrasciou, A., 1981, Phys. Rev. D **24**, 496.
Peccei, R. D., and H. R. Quinn, 1977, Phys. Rev. D **16**, 1791.
Peebles, P. J. E., 1980, *The Large-Scale Structure of the Universe* (Princeton University, Princeton, N.J.).
Pi, S.-Y., 1982, Phys. Lett. **112B**, 441.
Pi, S.-Y., 1984, Phys. Rev. Lett. **152**, 1725.
Preskill, J. P., 1979, Phys. Rev. Lett. **43**, 1365.
Press, W. H., 1980, Phys. Scr. **21**, 702.
Press, W. H., 1981, in *Cosmology and Particles: Proceedings of the 16th Recontre de Moriond, Astrophysics Meeting,* edited by J. Audouze, P. Crane, T. Gaisser, D. Hegyi, and J. Tran Thanh Van (Éditions Frontières, Paris), p. 137.
Press, W. H., 1983, Harvard-Smithsonian Center for Astrophysics preprint 1811, to appear in "Jim Wilson 60th Birthday Volume," edited by J. Centrella.
Press, W. H., and E. T. Vishniac, 1980, Astrophys. J. **239**, 1.
Primack, J. R., and G. R. Blumenthal, 1983a, in *Fourth Workshop on Grand Unification, University of Pennsylvania. . . 1983,* edited by H. A. Weldon, P. Langacker, and P. J. Steinhardt (Birkhäuser, Boston), p. 256.
Primack, J. R., and G. R. Blumenthal, 1983b, in *Galaxies and the Early Universe: Proceedings of the 18th Rencontre de Moriond, Astrophysics Meeting,* edited by J. Audouze and J. Tran Thanh Van (Reidel, Dordrecht, 1983).
Rubakov, V. A., M. V. Sazhin, and A. V. Veryaskin, 1982, Phys. Lett. **115B**, 189.
Sachs, R. K., and A. M. Wolfe, 1967, Astrophys. J. **147**, 73.
Sakharov, A. D., 1967, Zh. Eksp. Teor. Fiz. Pis'ma Red. **5**, 32 [JETP Lett. **5**, 24 (1967)].

Salam, A., 1968, in *Elementary Particle Theory, Proceedings of the Eighth Nobel Symposium,* edited by N. Svartholm (Wiley, New York), p. 367.
Sasaki, M., 1983, Prog. Theor. Phys. **70**, 394.
Sasaki, M., H. Kodama, and K. Sato, 1982, Prog. Theor. Phys. **68**, 1561.
Sato, K., 1981a, Phys. Lett. **99B**, 66.
Sato, K., 1981b, Mon. Not. R. Astron. Soc. **195**, 467.
Schnitzer, H. J., 1974, Phys. Rev. D **10**, 1800.
Schwinger, J., 1951a, Proc. Natl. Acad. Sci. USA **37**, 452.
Schwinger, J., 1951b, Phys. Rev. **82**, 664.
Sciama, D., P. Candelas, and D. Deutsch, 1981, Adv. Phys. **30**, 327.
Sexl, R., and H. Urbantke, 1967, Acta Phys. Austriaca **26**, 339.
Sexl, R., and H. Urbantke, 1969, Phys. Rev. **179**, 1247.
Shafi, Q., and A. Vilenkin, 1984, Phys. Rev. Lett. **52**, 691.
Shepard, H. K., 1983, Phys. Rev. D **27**, 1288.
Sher, M., 1981, Phys. Rev. D **24**, 1699.
Sher, M., 1983, Phys. Lett. **124B**, 329.
Shore, G. M., 1980, Ann. Phys. (N.Y.) **128**, 376.
Silk, J., 1983, "Elementary Particles and the Large-Scale Structure of the Universe," University of California at Berkeley preprint.
Silk, J., 1985, "The Cosmic Microwave Radiation: Implications for Galaxy Formation," paper delivered at the Inner Space/Outer Space Conference, Fermilab, May 2—5, 1984, to be published.
Srednicki, M., 1982a, Nucl. Phys. B **202**, 327.
Srednicki, M., 1982b, Nucl. Phys. B **206**, 132.
Starobinsky, A. A., 1979, Zh. Eksp. Teor. Fiz. Pis'ma Red. **30**, 719 [JETP Lett. **30**, 682 (1979)].
Starobinsky, A. A., 1982, Phys. Lett. **117B**, 175.
Steinhardt, P. J., 1980, Phys. Lett. **97B**, 147.
Steinhardt, P. J., 1981a, Nucl. Phys. B **179**, 492.
Steinhardt, P. J., 1981b, Nucl. Phys. B **190**, 583.
Steinhardt, P. J., 1982, in *The Very Early Universe, Proceedings of the Nuffield Workshop, Cambridge. . . 1982,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University, Cambridge), p. 251.
Steinhardt, P. J., and M. S. Turner, 1984, Phys. Rev. D **29**, 2162.
Stone, M., 1976, Phys. Rev. D **14**, 3568.
Stone, M., 1977, Phys. Lett. **67B**, 186.
Symanzik, K., 1970, Commun. Math. Phys. **16**, 48.
Tamvakis, K., and C. Vayonakis, 1982, Phys. Lett. **109B**, 283.
't Hooft, G., 1975, Acta Univ. Wratislav. **38**, XII, *"Functional and Probabilistic Methods in Quantum Field Theory,"* Vol 1.
Toussaint, D., S. B. Treiman, F. Wilczek, and A. Zee, 1979, Phys. Rev. D **19**, 1036.
Turner, M. S., 1982, in *The Very Early Universe, Proceedings of the Nuffield Workshop, Cambridge. . . 1982,* edited by G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge University, Cambridge), p. 297.
Unruh, W., 1974, Phys. Rev. D **10**, 3194.
Unruh, W., 1976, Phys. Rev. D **14**, 870.
Vilenkin, A., 1983a, Nucl. Phys. B **226**, 527.
Vilenkin, A., 1983b, Nucl. Phys. B **226**, 504.
Vilenkin, A., and L. H. Ford, 1982, Phys. Rev. D **26**, 1231.
Voloshin, M. B., I. Yu. Kobzarev, and L. B. Okun, 1975, Yad. Fiz. **20**, 1229 [Sov. J. Nucl. Phys. **20**, 644 (1975)].
Wald, R. M., 1975, Commun. Math. Phys. **45**, 9.
Wald, R. M., 1977, Commun. Math. Phys. **54**, 1.
Wald, R. M., 1979, Commun. Math. Phys. **70**, 221.
Weinberg, S., 1967, Phys. Rev. Lett. **19**, 1264.
Weinberg, S., 1972, *Gravitation and Cosmology* (Wiley, New York).
Weinberg, S., 1974, Phys. Rev. D **9**, 3357.
Weinberg, S., 1976, Phys. Rev. Lett. **36**, 294.
Weinberg, S., 1979, Phys. Rev. Lett. **42**, 850.
Weiss, R., 1980, Annu. Rev. Astron. Astrophys. **18**, 489.
Wilczek, F., 1984, Erice lectures 1983, ITP preprint NSF-ITP-84-14.
Witten, E., 1981a, Phys. Lett. **105B**, 267.
Witten, E., 1981b, Nucl. Phys. B **177**, 477.
Yoshimura, M., 1978, Phys. Rev. Lett. **41**, 281; **42**, 746(E).
Zel'dovich, Ya. B., 1970, Zh. Eksp. Teor. Fiz. Pis'ma Red. **12**, 443 [JETP Lett. **12**, 307 (1970)].
Zel'dovich, Ya. B., 1972, Mon. Not. R. Astron. Soc. **160**, 1P.
Zel'dovich, Ya. B., and M. Yu. Khlopov, 1978, Phys. Lett. **79B**, 239.
Zel'dovich, Ya. B., and A. A. Starobinsky, 1971, Zh. Eksp. Teor. Fiz. **61**, 2161 [Sov. Phys.—JETP **34**, 1159 (1971)].
Zumino, B., 1970, in *Lectures in Elementary Particles and Quantum Field Theory:* 1970 Brandeis University Summer Institute in Theoretical Physics, edited by Stanley Deser, Marc Grisaru, and Hugh Pendleton (MIT, Cambridge, Mass.), Vol. 2, p. 437.