Durham E-Theses

Thermodymiamics and inflation

Langbein, Rollo Foster

How to cite:

Use policy

Thermodynamics
and
Inflation

A thesis presented for the degree of

Doctor of Philosophy

by

Rollo Foster Langbein

University of Durham

Department of Physics

September 1992

2 DEC 1992

Abstract

The standard model of particle physics is introduced, and extensions of it, which may be of cosmological relevance, are considered. The inflationary paradigm is reviewed as an extension of the standard cosmological model. In particular, the natural inflation mechanism resulting from a thermal phase change in a field theory with a spontaneous symmetry breaking potential, is examined.

The question of when thermal equilibrium is likely to be a valid assumption in the early universe is considered in some detail. For inflation models, this question is answered by a self-consistency argument involving the total number of interactions per inflaton particle. In order to describe thermal-phase-change inflation models further, the temperature-dependent effective potential resulting from finite-temperature field theory is reviewed.

The self-consistency test is developed into a numerical procedure which may be used to discuss the likelihood of thermal state generation in specific inflation models in a quantitative way. Alternatively, the method can be used to provide bounds on the parameters in the inflation potential from the requirement that a thermal state should occur. This procedure is applied to several example potentials and in particular it is easily verified that the "new inflation" model (relying on a phase change) is not viable. The method is quite general and can be applied to any inflation model for which a finite temperature effective potential can be defined.

The procedure is generalised to the recently proposed extended inflation. Bounds on the extra free parameters which must be introduced in extended inflation are discussed. It is concluded that despite these extra free parameters the difficulties of generating a thermal state are just as great as they are in conventional inflation.

Acknowledgements

First of all I would like to thank my supervisor Peter Collins who, despite a diary packed full of meetings, was still able to regularly find time to advise and encourage. Even though discussions had to be booked a week in advance they were always worthwhile and most often fruitful.

I must also thank the other students who have made my stay in Durham so memorable. In particular, David Barclay and Peter Sutton, officemates and companions for the last three years, and Neil Shaban, squash partner and closest of friends. Also of course, Duncan Curtis, David Summers, Mark Oakden and all the others on the roof of the Physics Department.

My warmest gratitude goes to my parents, Ian Langbein and Rosalind Gillespie, for both financial and emotional support, despite being so very far away. And most of all, special love and deepest thanks to my dearest friend Julie.

Declaration

I declare that no material in this thesis has previously been submitted for a degree at this or any other University.

All the research work in this thesis was carried out in collaboration with Dr P.D.B. Collins. Material in Chapters 4 and 6 has been published in

> *Thermodynamics of Inflation*
>
> P.D.B. Collins and R.F. Langbein, *Phys. Rev. D.* **45**, 3429, 1992

and that in Chapter 7 in

> *Thermodynamic Aspects of Extended Inflation*
>
> P.D.B. Collins and R.F. Langbein, Durham preprint DTP-92/46

Contents

1

Introduction

Nature possesses four fundamental types of interaction — gravitation, electro-magnetism, and the weak and strong "nuclear" forces. The last three of these are described by quantum field theories, and it is expected that gravity too has a fundamental quantum nature. However, the only scale relevant to gravitational interactions is provided by Newton's constant G $(= 6.67259(85) \times 10^{-11}\,\mathrm{m^3 kg^{-1} s^{-1}})$, which means that quantum gravitational effects need only be considered if the interaction has a characteristic energy, mass, length or time scale of the order

$$
\begin{aligned}
E_{Pl} &= \left(\frac{\hbar c^5}{G}\right)^{1/2} \approx 1.2 \times 10^{19}\,\mathrm{GeV}, \\
M_{Pl} &= \left(\frac{\hbar c}{G}\right)^{1/2} \approx 2.1 \times 10^{-8}\,\mathrm{kg}, \\
l_{Pl} &= \left(\frac{\hbar G}{c^3}\right)^{1/2} \approx 1.6 \times 10^{-35}\,\mathrm{m}, \\
t_{Pl} &= \left(\frac{\hbar G}{c^5}\right)^{1/2} \approx 5.4 \times 10^{-44}\,\mathrm{s},
\end{aligned}
\tag{1.1}
$$

the so called "Planck units". These are very far removed from the characteristic scales of the other interactions and, when considering everyday particle physics processes, gravity may be safely ignored. Of course, on macroscopic scales it is gravity that dominates, since it is the only force which is both of infinite range and not cancelled

out in bulk matter by equal and opposite positive and negative charges, masses only ever being positive. Macroscopic scales are far from (1.1) however, and at such distances gravity will be perfectly well described by classical theory, as defined by Einstein's equation. This is fortunate since no successful quantum gravity theory has yet been constructed.

It is expected therefore, that the Universe at large can be described by classical gravity. Indeed, Einstein's equation, with the simplifying assumptions of large scale homogeneity and isotropy, readily reveals the fundamental cosmological expansion. For the early universe, two further elements must be added; the particle physics' standard model and the assumption that the high density conditions that would have been present imply a state of thermodynamic equilibrium. The resulting synthesis, the "hot big bang model", may plausibly be used as a basis for discussing cosmological history from just after the Planck time t_{pl} to the present. This model has had so many successes that it is now widely accepted as *the* standard cosmological model. One obvious measure of its success is that its consequences for particle physics are nowadays taken very seriously. It provides bounds on various parameters, for example mass/lifetime limits for neutrinos, and any new particles that are proposed must be "cosmologically compatible".

Since the fundamental equation in cosmology is Einstein's equation, then the natural scales are provided by G and the Planck quantities above. But if the Planck units are natural this would seem to indicate that the Universe, with an age of 10–20$\times 10^9$ years $\sim 10^{17}$ s and in which cosmological distances are measured in Mpc $= 3 \times 10^{21}$ m, is an exceedingly unnatural state of affairs. Two other pressing problems are the so-called "horizon problem", due to the fact that the universal expansion is less rapid than the speed of light, and the related question of how the Universe avoids dominance by superheavy "topological defects" which are predicted to occur if a grand unified theory in particle physics undergoes spontaneous symmetry breakdown. The standard model of cosmology can provide no explanation for any of these problems. However, a simple extension can be made which neatly resolves all three at once. The solution is to postulate that there was short period of "inflation" during which the universe expanded very rapidly.

Inflation arises very naturally from Einstein's equation. If ever a cosmological constant term Λ becomes dominant then it is found that the characteristic size of the Universe increases exponentially as $e^{(\Lambda/3)^{1/2}t}$. The mechanism for generating this dominance is another place where a connection with fundamental particle physics may be made — Λ is almost always regarded as originating from the "vacuum energy" generated by some scalar field. It was originally envisaged to be closely related to the symmetry breaking in the Higgs field of the particle physics standard model. The point is that because Einstein's equation couples gravity to energy, the absolute value of a potential becomes relevant. In particle physics the Higgs effect relies on the fact that the field is localized in an absolute minimum of its potential which is asymmetric, i.e., it is a spontaneous symmetry breaking potential. This defines the vacuum of the theory, and the scalar particles are quantum fluctuations about this minimum. If the absolute minimum of this potential has zero energy as one would expect, to agree with the observation that today the cosmological constant is zero or at least insignificant on the Planck scale, then away from this minimum, at the origin say, there will be a non-zero energy. Hence vacuum energy, which acts like a cosmological constant, will occur if the field becomes localized away from the absolute minimum of the potential.

Unfortunately, it has turned out to be rather more difficult to pin the scalar particle that would be required to cause inflation to any specific particle physics model. These days its exact identity is usually left indeterminate and it has become known as the "inflaton". Of course, the method that has just been outlined for generating vacuum energy could apply to practically any potential, even if its absolute minimum is at the origin. There have been many suggestions for ways of achieving the localization of the field away from the absolute minimum. No mechanism yet proposed, however, can be considered particularly compelling. Inflation is still very much a "paradigm in search of a model" with the hope, yet to be attained, that the scalar field will arise naturally from a fundamental particle physics theory.

Probably the most attractive way of generating the vacuum energy is via a thermal phase transition. A phase transition develops automatically when a scalar field theory with a spontaneous symmetry breaking potential is in thermal equilibrium. At sufficiently high temperatures symmetry restoration will occur. The transition

from the symmetric to the broken phase, as the Universe cools, then defines a transition from a state with a large vacuum energy to one where it is zero, and it is thus that a short era of inflation is generated. This effect is conveniently described by a temperature dependent effective potential.

Yet inflation is manifestly a non-equilibrium process — such rapid expansion will very quickly make interactions between the scalar particles exceedingly rare. It is not therefore possible for a thermal state to form whilst inflation is occurring and so, if this thermal phase change mechanism is to be believed, it must have been present from the start. Just how likely is it that an initial thermal state would have arisen? This thesis sets out to try and answer this question in detail.

Since it is believed that the eventual identity of the inflaton will be revealed as some kind of scalar quantum field, a review of basic field theory, gauge theory and the Higgs mechanism, is given in chapter 2. It is also shown in this chapter how these elements are employed in the standard model of particle physics. Chapter 3 gives an overview of the cosmological standard model, and the question of why and how inflation is to be added to this model is considered in more detail. Thermodynamic equilibrium is further explored in chapter 4, and of particular interest is the question of how to determine when equilibrium is likely to arise. A self consistency argument involving a calculation of the total number of interactions per inflaton is employed to try and answer this question. Chapter 5 shows how thermal equilibrium can be incorporated into field theory as "finite temperature field theory", and introduces the temperature dependent effective potential, the fundamental object for describing phase changes in the early universe. The results of chapters 4 and 5 are combined in chapter 6 to create a numerical procedure for testing thermal state generation in specific inflation models. The formalism is then generalized to extended inflation in chapter 7. Some conclusions are provided in chapter 8.

In this thesis "natural units" are used where $\hbar = c = k = 1$. Occasionally, however, factors of $\hbar$ will be included where it makes the exposition clearer.

2

The Standard Model of Particle Physics

Physics on the very smallest scales known is described by Quantum Field Theory. This chapter is a highly selective overview of the subject, mainly intended to give definitions of quantities which will be useful in later chapters, and to highlight elements which are relevant in a cosmological context. Section 2.1 describes quantum field theory, and shows how one obtains experimentally measurable cross sections from it, via the standard perturbative expansion. An outline of gauge theories, which form the basis of a description of all the fundamental interactions in the standard model of particle physics, is given in section 2.2. The other major feature of this standard model, spontaneous symmetry breakdown, is illustrated in section 2.3, and the standard model itself is then briefly surveyed in the following section. This particle-physics standard model is extremely successful, but it raises several further and more fundamental questions. Some of the theoretical extensions designed to try to answer such questions, which may also prove pertinent to a description of the early universe, are included in the final section.

2.1 Quantum Field Theory

The development of Quantum Field Theory is presented in this section through the simple example of a self-interacting scalar field ϕ, both because it simplifies the analysis and because it will be of particular importance in later chapters. More complicated cases will be given as modifications to this general framework. Even

though quantum theory often runs contrary to classical expectations it is still firmly rooted in a classical background. The classical dynamics of a field can be derived from the fundamental action principle (see for example [1]), expressed in terms of the action S which is defined by

$$S = \int d^4x \, \mathcal{L}(\partial_\mu \phi, \phi) \tag{2.1}$$

where $\mathcal{L}$ is the Lagrangian density of the theory. The action principle states that the fields described by $\mathcal{L}$ will evolve in such a way that S is a minimum. This is expressed symbolically as

$$\delta S = 0. \tag{2.2}$$

where δS describes the variation of (2.1) under small variations of both the coordinates $x^\mu \to x^\mu + \delta x^\mu$ and of the field $\phi(x) \to \phi(x) + \delta\phi(x)$. Minimisation of (2.1) implies the fields evolve according to the Euler-Lagrange equations of motion;

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi\right)} - \frac{\partial \mathcal{L}}{\partial \phi} = 0. \tag{2.3}$$

The requirement that the Lagrangian be translationally invariant implies, through Noether's theorem, that[1]

$$T^{\mu\nu} = \frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \phi\right)} \partial^\nu \phi - \eta^{\mu\nu} \mathcal{L} \tag{2.4}$$

is a conserved current and $T^{\mu\nu}$ is identified as the energy-momentum tensor. The Minkowski space metric $\eta^{\mu\nu}$ is defined as

$$\eta^{\mu\nu} \equiv \mathrm{diag}(1, -1, -1, -1). \tag{2.5}$$

Conservation of $T^{\mu\nu}$ implies the energy-momentum conservation law,

$$\partial_\mu T^{\mu\nu} = 0, \tag{2.6}$$

and that there is a conserved "charge"

$$P^\nu = \int T^{0\nu} \, d^3x \tag{2.7}$$

which is the total momentum of the field. The conjugate momentum to $\phi(x)$

$$\pi(x) = \frac{\partial \mathcal{L}}{\partial(\partial_0 \phi)} \tag{2.8}$$

and the total energy or Hamiltonian

$$H = \int T^{00} \, d^3x = \int d^3x \, (\pi(x)\partial_0\phi(x) - \mathcal{L}(x)) \tag{2.9}$$

can also be introduced in the usual way.

Quantisation of theory is achieved by replacing $\pi(x)$ and $\phi(x)$, the canonical variables of the theory, by quantum operators which act on a Hilbert space of square integrable functions, and which satisfy the same equations of motion as the classical variables. They are also postulated to satisfy the equal time canonical commutation relations (see e.g., [2])

$$[\pi(\mathbf{x}, t), \phi(\mathbf{y}, t)] = -i\delta^3(\mathbf{x} - \mathbf{y})$$
$$[\pi(\mathbf{x}, t), \pi(\mathbf{y}, t)] = [\phi(\mathbf{x}, t), \phi(\mathbf{y}, t)] = 0. \tag{2.10}$$

The time evolution of these canonical quantities is described by the Hamiltonian

$$\partial_0\phi(\mathbf{x}, t) = i[H, \phi(\mathbf{x}, t)]$$
$$\partial_0\pi(\mathbf{x}, t) = i[H, \pi(\mathbf{x}, t)]. \tag{2.11}$$

For a self-interacting scalar field the Lagrangian is

$$\mathcal{L} = \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - V(\phi) \tag{2.12}$$

where $V(\phi)$ is its interaction potential. The Euler-Lagrange equations for this case give

$$\partial^\mu\partial_\mu\phi + \frac{dV(\phi)}{d\phi} = 0 \tag{2.13}$$

and the mass of the field is given by the second derivative of the potential, $d^2V/d\phi^2$. Exact solutions of (2.13) are known only for the case where the potential describes a

mass exclusively, i.e., where there are no interaction terms, and $V = \mu^2 \phi^2 / 2$ where μ is the mass. In that case (2.13) reduces to the Klein-Gordon equation, $(\Box + \mu^2)\phi = 0$ and the solutions for ϕ are superpositions of plane waves $e^{\pm k \cdot x}$, i.e.,

$$\phi = \int \frac{d^3 k}{((2\pi)^3 2k_0)^{1/2}} \left(a(\mathrm{k}) e^{k \cdot x} + a^\dagger(\mathrm{k}) e^{-k \cdot x} \right) \qquad (2.14)$$

where $k_0 = (\mathrm{k}^2 + \mu^2)^{1/2}$ and the Fourier components $a^\dagger(\mathrm{k})$ and $a(\mathrm{k})$ are operators interpreted as creation and annihilation operators of k-states. A ground-state vacuum $|0\rangle$ is defined so that $a(\mathrm{k})|0\rangle = 0$. Thus the single particle states are $|k_1\rangle = a^\dagger(\mathrm{k}_1)|0\rangle$ and superpositions $|k_1 k_2\rangle = a^\dagger(\mathrm{k}_1)a^\dagger(\mathrm{k}_2)|0\rangle$, and so on. In this non-interacting case the Hamiltonian may be written

$$H = \int d^3 k \, k_0 a^\dagger(\mathrm{k}) a(\mathrm{k}) \qquad (2.15)$$

plus an irrelevant constant. The number operator

$$N = \int d^3 k \, a^\dagger(\mathrm{k}) a(\mathrm{k}) \qquad (2.16)$$

counts the number particles in a given state, and it clearly commutes with the Hamiltonian (2.15).

Of course, the theories of physical interest are not simply free fields but also have interactions with themselves and/or with other fields. In order to calculate observable quantities in such theories some kind of approximation scheme is required. The major technique for investigating high energy physics is particle scattering. Because the energy is high, the interaction time is extremely short and the detailed progression of the scattering process cannot be followed. It is therefore assumed that long before the collision the particles can be described asymptotically by the free field k-states above. The collision then takes place and a long time afterward the resultant particles are again described by a superposition of free-field states. The scattering probability is expressed in terms of the S-matrix which describes how to get from the state before the collision to the state after the collision, i.e., $|t = \infty\rangle = S|t = -\infty\rangle$[1]. This matrix

is further broken down into $S = I + iT$ where I is an identity corresponding to no scattering, and T contains the information about the scattering process. A relevant example is two particles, of momenta p_1 and p_2, scattering to give two particles of momenta p_3 and p_4. The quantity of interest is then

$$\langle p_3 p_4 | T | p_1 p_2 \rangle \equiv (2\pi)^4 \delta^4 (p_3 + p_4 - p_1 - p_2) \mathcal{M}. \qquad (2.17)$$

The matrix element $\mathcal{M}$ contains all the physics of the interaction and is calculated approximately from the field theory using a "perturbative expansion".

The quantity actually measured at accelerators when two particles collide and produce two subsequent particles is the cross section σ. It can be obtained from the matrix element $\mathcal{M}$ by[3]

$$d\sigma = \frac{|\overline{\mathcal{M}}|^2}{\mathcal{F}} dQ \qquad (2.18)$$

where the overbar indicates any appropriate averaging over the spins of the particles, $\mathcal{F}$ is the flux of the incident particles, i.e.,

$$\mathcal{F} \equiv |\mathbf{v}_1 - \mathbf{v}_2| \cdot 2E_1 \cdot 2E_2 \qquad (2.19)$$

and dQ is the Lorentz invariant phase space factor, i.e.,

$$dQ \equiv (2\pi)^4 \delta^{(4)}(p_3 + p_4 - p_1 - p_2) \frac{d^3 p_3}{(2\pi)^3 2E_3} \frac{d^3 p_4}{(2\pi)^3 2E_4} \qquad (2.20)$$

where the particle four-momenta $p_i = (E_i, E_i \mathbf{v}_i)$. The total cross section, for example, would be obtained by integrating (2.18) over $d^3 p_3$ and $d^3 p_4$.

The normal procedure used to calculate $\mathcal{M}$ from the field theory starts by splitting the Lagrangian (2.12) into "free" and "interacting" parts (see [4] for a complete treatment)

$$\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\text{int}}. \qquad (2.21)$$

For example, a scalar field, mass μ, with a $\lambda \phi^4$ interaction has

$$\mathcal{L}_0 = \partial^\mu \phi \partial_\mu \phi + \frac{1}{2}\mu^2 \phi^2, \quad \text{and} \quad \mathcal{L}_{\text{int}} = \frac{\lambda}{4!}\phi^4. \qquad (2.22)$$

From the interaction Lagrangian an interaction Hamiltonian can be defined, and if the

interaction part is small a perturbative expansion can be made. So for the example, if λ in (2.22) is small, i.e., if $\lambda \ll 1$, then a series expansion in powers of λ can be made. The interaction Hamiltonian is responsible for transitions between the free-field states, described by $\mathcal{M}$ in (2.17), and the method of calculating the coefficients in this perturbative expansion is succinctly summarized by the Feynman rules (see e.g., [3]). Some examples are given in table 2.1 and a sample perturbative series for elastic scattering in $\lambda\phi^4$ theory is shown in fig. 2.1(a).

In the particle physics standard model all of the matter fields are fermionic, and it is of course possible to develop a similar formalism as the above for them. In this case the fields are four-component spinors governed by a Clifford algebra of anti-commuting "γ-matrices" and the postulated commutation relations (2.10) are replaced by anti-commutation relations[1].

It is also possible to develop field theory in an equivalent but conceptually rather different way. This is the so called path integral formalism, a generalisation of Feynman's sum over histories formulation of quantum mechanics[5]. This scheme is neatly encompassed by the generating functional Z defined as (see e.g., [2])

$$Z[J] = N \int [d\phi] \exp \left\{ iS + \int d^4x \, J(x)\phi(x) \right\} \tag{2.23}$$

in which it is understood that all the quantities are in Euclidean space, i.e., $t \to -it$, so S is the Euclidean action, rather than (2.1). $\phi(x)$ is once again an ordinary function, N is a normalisation constant and the integration is over the (infinite dimensional) space of functions. The term in $J(x)$ describes an arbitrary source. $Z[J]$ is called the generating functional because it can be expanded in a power series in J, i.e.,

$$Z[J] = \sum_{n=0}^{\infty} \frac{i^n}{n!} \int d^4x_1 \cdots d^4x_n \, J(x_1) \cdots J(x_n) G^{(n)}(x_1, \ldots, x_n) \tag{2.24}$$

where these $G^{(n)}$ are Green's functions of the theory. The n-point Green's function is given, from the previously defined ϕ-operators, by

$$G^{(n)}(x_1, \ldots, x_n) = \langle 0|T(\phi(x_1) \cdots \phi(x_n))|0 \rangle \tag{2.25}$$

where T is the time ordering operator which arranges the operators in its argu-

$\lambda\phi^4$ Theory:
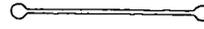
Scalar Propagator $\qquad\qquad\qquad$ o———o $\qquad\qquad$ $\dfrac{i}{q^2 - \mu^2}$

Vertex $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $-i\lambda$

Scalar QED:

Vector Propagator
(Feynman Gauge) $\qquad\qquad\qquad$ $\dfrac{i\eta_{\mu\nu}}{q^2}$

Vertices
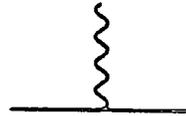
$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $-ie(p + p')^\mu$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $2ie^2\eta^{\mu\nu}$

in addition to the $\lambda\phi^4$ rules.

Loops: Integrate over the loop momentum $\qquad\qquad$ $\displaystyle\int \dfrac{d^4l}{(2\pi)^4}$

Table 2.1: Feynman Rules for $\lambda\phi^4$ theory and scalar QED (described in section 2.2). The matrix element $\mathcal{M}$ is calculated by drawing all topologically distinct diagrams for the process to the required order (see for example fig. 2.1(a)) and applying the above rules.

(a)

(b)   (c)
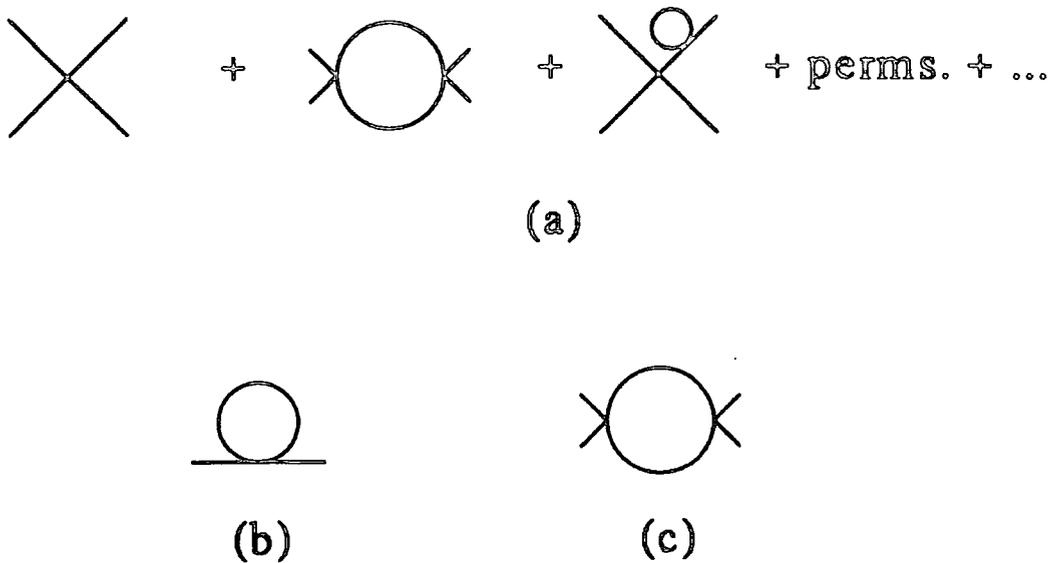
Figure 2.1: (a) Feynman diagrams for calculating the perturbative series for the elastic scattering of ϕ particles in $\lambda\phi^4$ theory. (b) the two-point function, and (c) the four point function, the two one-particle-irreducible diagrams for $\lambda\phi^4$ theory to one-loop order. General connected diagrams in the cross section in (a) may be obtained from appropriate combinations of the 1PI diagrams, (b) and (c).

ment from right to left with increasing times. These Green's functions are closely related to the S-matrix elements by the Lehmann-Symanzik-Zimmermann reduction formalism[1]. Thus, a knowledge of the generating functional fully defines the theory.

It turns out that these particular Green's functions are not quite what is required for calculating physical processes. This is because they contain disconnected graphs, whereas only connected graphs are required in a calculation like fig. 2.1. However, Green's functions for just the connected graphs are easily obtained by formally defining[2]

$$W[J] = \ln Z[J] \tag{2.26}$$

which has a similar expansion to (2.24) but now in terms of $G_c^{(n)}$, the required connected Greens functions. By definition, individual Green's functions may be derived

from the generating functional by functional differentiation, hence,

$$G_c^{(n)}(x_1 \cdots x_n) = \frac{1}{i^n} \left. \frac{\delta^n W[J]}{\delta J(x_1) \cdots \delta J(x_n)} \right|_{J=0} . \tag{2.27}$$

The anti-commuting nature of fermions is incorporated into this scheme by considering generating functionals in terms of anti-commuting Grassman variables (see e.g., [2]).

For actual calculations, it is in fact more convenient to work with momentum space Greens functions defined as

$$(2\pi)^4 \delta^4(k_1 + \cdots + k_n) G^{(n)}(k_1, \ldots, k_n) \equiv \int \prod_{i=1}^n d^4 x_i \, e^{-i p_i \cdot x_i} G^{(n)}(x_1, \ldots, x_n) \tag{2.28}$$

and truncated Greens functions for which the propagators have been removed from the external lines of the graph. The calculation is simplified further still by considering only Green's functions for one-particle-irreducible (1PI) graphs, i.e., Feynman graphs which cannot be split in two by cutting a single internal line. These are usually labelled $\Gamma^{(n)}$, and it is clear that any general connected graph can be constructed out of the simpler 1PI graphs. For the $\lambda \phi^4$ theory there are just two important Greens functions; the two-point Green's function $\Gamma^{(2)}$ which is the inverse of the Feynman propagator function, and the four-point function $\Gamma^{(4)}$ corresponding to the interaction. The one-loop diagrams for these two functions are shown in figs. 2.1(b) and (c).

Unfortunately, perturbative calculations in quantum field theory, for example the one represented pictorially in fig. 2.1(a), usually encounter serious difficulties. When the calculation is pursued above tree level the integrals that arise from any loops in the diagrams (see table 2.1) are found to be ultra-violet divergent, i.e., they diverge for large values of the loop momentum. Such behaviour requires a "renormalization scheme" of which there are many[6]. The idea is to define a "renormalized" Lagrangian in terms of a renormalized wave function, mass, and coupling all of which are finite but related to their "bare" (the name for quantities in the original or bare Lagrangian) counterparts by an infinite multiplicative constant. The point is that it

is only observables expressed in terms of the renormalized quantities that are actually measurable.

It may be shown that this procedure is equivalent to adding "counterterms" to the original Lagrangian which will cancel out the infinite parts of the 1PI loop integrals; this is the Bogoliubov-Parasiuk-Hepp (BPH) prescription. Only 1PI graphs need to be considered because if they are finite, then the more general diagrams built from them will be also. For the $\lambda\phi^4$ example above, the required counterterms have the same structure as the terms in the bare Lagrangian, i.e.,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}\mu^2\phi^2 - \frac{\lambda}{4!}\phi^4 + \frac{1}{2}A(\partial_\mu\phi)^2 - \frac{1}{2}B\phi^2 - \frac{1}{4!}C\phi^4 \qquad (2.29)$$

where A, B and C are the wave-function, mass and coupling constant renormalization counterterms. The renormalization prescription, given (2.29), proceeds as follows. First of all, the integrals resulting from the first-order 1PI loop diagrams are regulated in some way, for example, by cutting them off at some large momentum. The 1PI diagrams needed for (2.29) are fig. 2.1(b), the first order diagram for the two-point function, and fig. 2.1(c) that of the four-point function. Next, some renormalization conditions must be specified which can also be expressed in terms of Green's functions. The conditions

$$\Gamma^{(2)}(p^2)\Big|_{p^2=0} = -\mu^2, \qquad (2.30)$$

$$\frac{\partial\Gamma^{(2)}(p^2)}{\partial p^2}\Bigg|_{p^2=0} = 0 \qquad (2.31)$$

and

$$\Gamma^{(4)}(0,0,0,0) = -i\lambda \qquad (2.32)$$

serve to renormalize $\lambda\phi^4$ theory. Using these, it is now possible to determine the constants A, B and C in (2.29) and hence physical quantities can be calculated to first order. An example of this scheme in action will be seen in Chapter 5. The new Lagrangian (2.29) can then be used to generate the second order 1PI diagrams

(with up to two loops) for which the counterterms cancelling these divergences are calculated, and so on. If the number of counterterm types that have to be added to the Lagrangian remains finite, as higher and higher orders of perturbation theory are calculated, then the theory is said to be renormalizable. For example, only terms in A, B and C in (2.29) are required to renormalize $\lambda\phi^4$ theory to all orders. Gauge theories (see section 2.2) are another important example of renormalizable field theories.

It must be noted that the "subtraction point" of $p_i^2 = 0$ in (2.30)–(2.32) is arbitrary, and it is possible to choose any $p^2 = \kappa^2$ to renormalize at. Different choices of κ will give different definitions of the renormalized quantities. Of course, quantities such as the coupling can only really be defined from experiment by specifying it from the cross section at some value of the kinematical parameters, so there is no problem with this. This independence of physical observables with the renormalization scale κ turns out to be a non-trivial feature of renormalizable field theory. The behaviour of the theory under changes in this scale is embodied in the "renormalization group equation" (see e.g., [2]) which leads to the concept of "running couplings" — the variation of the coupling with the energy scale of the collision process.

2.2   Gauge Field Theory

A fundamental assumption of Quantum theory is that the absolute phase of a wave function is unobservable and therefore arbitrary. Consider the Lagrangian (2.22) but where ϕ is replaced by a complex scalar $\phi = \phi_1 + i\phi_2$. It is unchanged by the transformation

$$\phi(x) \to e^{i\alpha}\phi(x), \quad \phi^*(x) \to e^{-i\alpha}\phi^*(x) \tag{2.33}$$

Just as for the energy-momentum case above, such a symmetry implies, via Noether's theorem, that there must be a conserved current. In this case it is $J^\mu \equiv i(\phi^*\partial^\mu\phi - \phi\partial^\mu\phi^*)$ which in turn implies that the charge $Q \equiv \int d^3x\, J^0$ is conserved.

If it is demanded on aesthetic grounds that this symmetry is true for *local* phase transformations, or "gauge transformations", i.e.,

$$\phi(x) \to e^{i\alpha(x)}\phi(x), \quad \phi^*(x) \to e^{-i\alpha(x)}\phi^*(x) \tag{2.34}$$

then it is found that the Lagrangian must be modified to include a "gauge field" A^μ (see e.g., [3]), i.e.,

$$\mathcal{L} = \frac{1}{2}(\partial_\mu + ieA_\mu)^2\phi^2 - \frac{1}{2}\mu^2\phi^2 - \frac{\lambda}{4!}\phi^4 - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \qquad (2.35)$$

where the additional term in the field strength tensor $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$ is included so that the gauge field has an energy and is thus physical. In perturbative calculations (2.35) implies that extra Feynman rules need to be added to account for the coupling of the ϕ-field to the gauge field (see table 2.1). Such transformations (2.34) are obviously Abelian and the complete set of them forms the $U(1)$ group, the simplest possible Lie group. If a fermion field theory is provided with this symmetry then Quantum Electrodynamics (QED), the theory of electromagnetic interactions with matter, is obtained.

It might seem at first sight, that the gauge field A^μ has four degrees of freedom but this is not really so. Two "gauge fixing" conditions have to be specified corresponding to the gauge freedoms of A_μ which do not affect the physics. The first of these is usually the Lorentz condition, $\partial_\mu A^\mu = 0$, which implies that a non-interacting gauge particle obeys the wave equation $\Box A^\mu = 0$. It also allows its propagator to be specified as[3]

$$\frac{i}{q^2}\left(-\eta_{\mu\nu} + (1-\xi)\frac{q_\mu q_\nu}{q^2}\right) \qquad (2.36)$$

where ξ is arbitrary. (A common choice is the "Feynman gauge", $\xi = 1$.) The solutions of the free particle wave equation are the plane waves $A^\mu = \epsilon^\mu(q)e^{-iq\cdot x}$ where ϵ^μ is the polarization vector, and the second gauge condition that is often set is $\epsilon^0 = 0$. This (noncovariant) choice limits the particle to two transverse polarization degrees of freedom. This is a well known property of the photon, the gauge particle of the electromagnetic interaction.

It is of course possible to generalise this gauge structure to an arbitrary Lie group so that (2.33) becomes[7]

$$\phi \rightarrow \exp\left(i\frac{1}{2}\sum_i \lambda_i\alpha_i\right)\phi \qquad (2.37)$$

where the sum is over the generators λ_i of the Lie group multiplied by the parameters α_i. These more general groups are non-Abelian, an interesting consequence of which is that the gauge particles acquire a self interaction. The group $SU(2)$, for which the generators may be represented by the three Pauli matrices, in combination with a $U(1)$ group (after spontaneous symmetry breaking, see next section), may be used to describe the electroweak interactions[7]. The group $SU(3)$, with eight generators, is used in Quantum Chromodynamics (QCD) which describes the interactions between quarks with three different "colour charges"[8].

2.3 Spontaneous Symmetry Breaking

A study of the physical processes which occur via the weak force seems to indicate that they are mediated by a massive gauge boson[7]. Unfortunately, it is not possible to construct a renormalizable field theory which includes gauge bosons which have an explicit mass. The propagator for such particles would have the form

$$i\frac{-\eta_{\mu\nu} + q_\mu q_\nu / M^2}{q^2 - M^2} \tag{2.38}$$

where M is the gauge-particle mass, which approaches a constant as $q^2 \to \infty$. This means that as larger numbers of loops are calculated in perturbation theory, ever more serious divergences appear so that ever more counter terms must be added to the Lagrangian, and the theory is unrenormalizable. The way out is to generate the masses by spontaneous symmetry breaking (SSB).

Consider a Lagrangian like (2.21)–(2.22) but where the sign of μ^2 has been changed, i.e.,

$$\mathcal{L} = \left(\partial_\mu \phi\right)^2 + \frac{1}{2}\mu^2 \phi^2 - \frac{\lambda}{4!}\phi^4. \tag{2.39}$$

The potential term $V(\phi) = -\mu^2 \phi^2 / 2 + \lambda \phi^4 / 4!$ is shown in fig. 2.2. It would appear from (2.39) that the particle has an imaginary mass. This is an illusion, however; it is a consequence of the fact that the perturbative expansion has not been made around the true vacuum of the theory. If instead, (2.39) is expanded about the minimum of
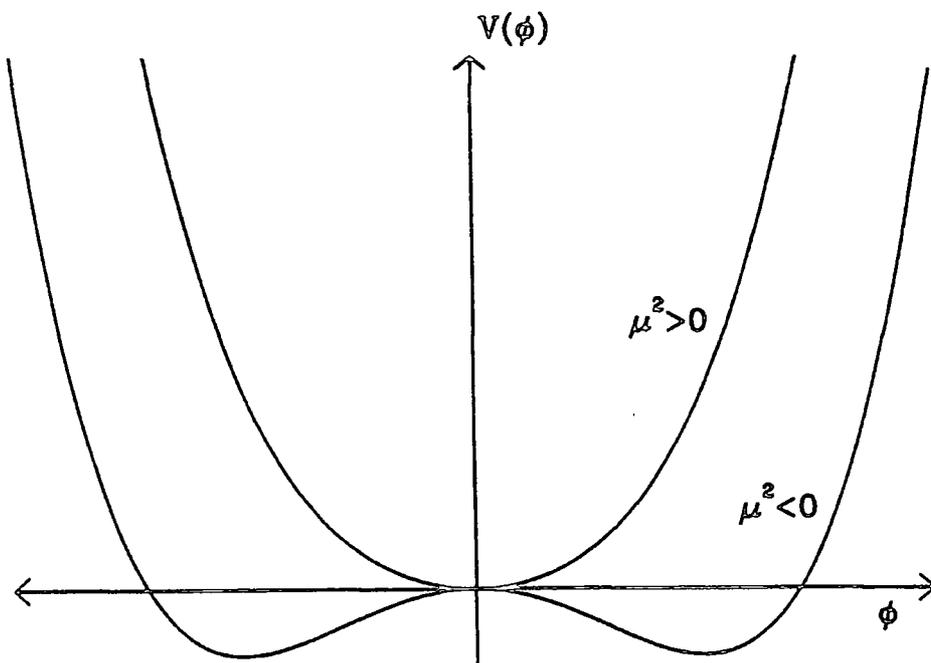
Figure 2.2: The potential $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4!}\phi^4$ ($\lambda > 0$). Comparison of the conventional form, $\mu^2 > 0$, with the spontaneous symmetry breaking form, $\mu^2 < 0$.

the potential at $v = \sqrt{-\mu^2/\lambda}$ by putting

$$\phi(x) = v + \eta(x) \tag{2.40}$$

it is found that

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu\eta\right)^2 - \lambda v^2\eta^2 - \lambda v\eta^3 - \frac{1}{4}\lambda\eta^4 \tag{2.41}$$

plus a constant so that a mass term with the correct sign is recovered corresponding to a particle of mass $\sqrt{2\lambda}v$.

To extend this argument to a gauge symmetry, a Lagrangian with a complex scalar singlet (2.35) but with a symmetry breaking negative μ^2 term in the potential is considered. $V(\phi)$ now has the "mexican hat" form shown in fig. 2.3. Making the perturbative expansion around a point in the "brim" of the hat where $|\phi| = v$, with
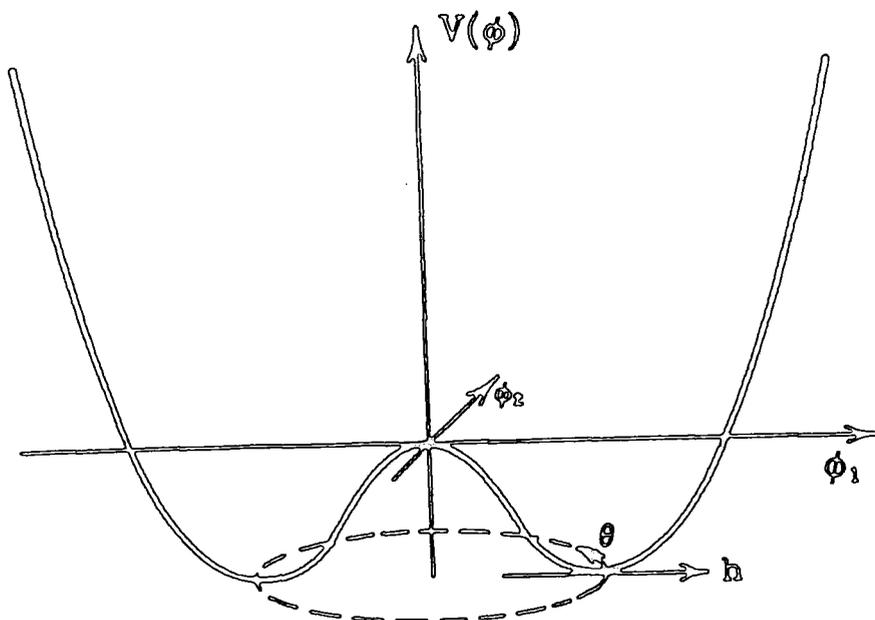
Figure 2.3: A symmetry breaking potential for a two component Higgs field —
the "mexican hat" potential. The Lagrangian should be expanded in the directions
θ and h at some point on the circle of minima to observe the Higgs mechanism.

the particular choice of gauge

$$\phi = \frac{1}{\sqrt{2}}(v + h(x))e^{i\theta(x)/v} \tag{2.42}$$

and

$$A_\mu \rightarrow A_\mu + \frac{1}{ev}\partial_\mu\theta, \tag{2.43}$$

it is found that

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu h\right)^2 - \lambda v^2 h^2 + \frac{1}{2}e^2 v^2 A_\mu^2 - \lambda v h^3 - \frac{1}{4}\lambda h^4$$

$$+\frac{1}{2}e^2 A_\mu^2 h^2 + v e^2 A_\mu^2 h - \frac{1}{4}F^{\mu\nu}F_{\mu\nu}. \tag{2.44}$$

This is independent of the field $\theta(x)$ but it can be seen that there is now a mass term
for the gauge particle of magnitude ev. A massive gauge particle has three degrees of

freedom, two transverse like the photon, and an extra longitudinal polarization. This is where the degree of freedom associated with the field $\theta(x)$ has disappeared to — it is said to be "eaten" when the gauge field becomes massive. This is known as the Higgs mechanism.

Gauge theories that generate masses in this way can be shown to retain their renormalizability and can therefore be used to provide a consistent, renormalizable theory of electroweak interactions. In addition to the photon there are the massive weak gauge particles $W^{\pm}$ and Z^0[7].

2.4 The Standard Model

All the ingredients necessary for a complete specification of the standard model have now been illustrated. The construction of this model will now be sketched (see e.g., [7] for a more complete description).

There are three fundamental gauge symmetries; $SU(3)_C$ of QCD with coupling g_s, and $SU(2)_L \times U(1)_Y$ of weak isospin and hypercharge with couplings g' and g which describe the electroweak interactions. There are two types of fundamental fermions; quarks and leptons. Only the quarks experience the couplings of QCD and transform as $SU(3)$ triplets with three different "colours". This colour force is mediated by gauge bosons called "gluons" of which there are eight. The QCD colour force is very strong at low energies and cannot be described by perturbation theory — the quarks are confined inside "colourless" hadrons, either triplets of quarks, i.e., baryons or quark-antiquark pairs, i.e., mesons. However, due to the running of the coupling mentioned in section 2.1, at energies above about 200 MeV, g_s becomes small enough for a perturbative description to be valid, thus allowing tests of QCD at high energy colliders. Both quarks and leptons feel the force associated with the $SU(2)_L \times U(1)_Y$ gauge group, but the left and right handed projections transform differently and have different weak isospin and hypercharge assignments. The complete set of standard model particles are given in table 2.2. The electroweak interactions would be mediated by four gauge particles, i.e., W^i ($i = 1, 2, 3$) for the $SU(2)$ of weak isospin and B for the $U(1)$ of hypercharge. But the symmetry is spontaneously broken by a complex

Particles

Family	$SU(3)$ singlets leptons	$SU(3)$ triplets quarks
1	$\begin{pmatrix} e \\ \nu_e \end{pmatrix}_L \quad e_R$	$\begin{pmatrix} u \\ d \end{pmatrix}_L \quad u_R, d_R$
2	$\begin{pmatrix} \mu \\ \nu_\mu \end{pmatrix}_L \quad \mu_R$	$\begin{pmatrix} c \\ s \end{pmatrix}_L \quad c_R, s_R$
3	$\begin{pmatrix} \tau \\ \nu_\tau \end{pmatrix}_L \quad \tau_R$	$\begin{pmatrix} t \\ b \end{pmatrix}_L \quad t_R, b_R$

Scalar Higgs boson, H

Forces

Force	Interaction mediated by	Fundamental Group
Electromagnetism	photon, γ	$U(1)_Q$
Weak	vector bosons, $W^\pm$ & Z^0	$SU(2)_L$
Strong (QCD)	8 gluons, g	$SU(3)_C$

Table 2.2: The Standard model particles. The fundamental fermions in the upper table are: the electron e, muon μ and tau τ with their respective neutrinos ν_e, ν_μ and ν_τ and the up u, down d, charm c, strange s, top t and bottom b, quarks. The left handed components of these fermions transform as $SU(2)$ doublets whilst the right handed components are singlets.

doublet Higgs field $\Phi = (\phi_1 + i\phi_2, \phi_3 + i\phi_4)$ and three of the Higgs-field's degrees of freedom are eaten giving three massive gauge particles ($W^\pm$ and Z^0) and leaving a single Higgs scalar. The fourth gauge particle remains massless and is identified with the photon — it corresponds to the only remaining symmetry of the original $SU(2)_L \times U(1)_Y$ group, the $U(1)_Q$ of electromagnetism, i.e., QED. The scale of this symmetry breaking is known as the electroweak scale and is $v = M_W/g \approx 246$ GeV.

The vacuum expectation value of the Higgs field is responsible for generating the masses of all the fermions as well. Due to the fact that left and right handed components of the fermions transform differently under the electroweak group, it is not possible to include explicit fermion mass terms like $m(\bar\psi_L \psi_R + \bar\psi_R \psi_L)$ in the

Lagrangian (which must be a singlet). The Higgs doublet rescues the situation by permitting the mass terms $h\bar{\psi}_L\Phi\psi_R$ where h is an arbitrary "Yukawa coupling". So in the standard model this one Higgs doublet very efficiently generates masses for the gauge bosons and the fundamental fermions through its vacuum expectation value.

Another interesting property of the weak interactions is that the quark eigenstates that undergo weak interactions are not the same as the quark mass eigenstates. This means that it is possible for transitions to occur between the different families of quarks, e.g., $c \rightarrow d$, $b \rightarrow u$, etc, and these transitions are embodied in the Kobayashi-Maskawa (KM)-matrix. CP violation is also allowed by this mixing between the three families, three being the minimum number for which this can happen. In the minimal standard model the neutrinos are massless and so mixing between the lepton families cannot occur. In fact, the right handed neutrinos are redundant since they are massless, chargeless, $SU(2)$-singlets which can never be produced by standard model processes.

This standard model has met with great success in explaining results found at particle accelerators around the world, and many of its parameters have been measured with unprecedented accuracy[9]. Indeed, no confirmed deviations from the standard model have yet been observed. The top quark has not yet been produced because of its high mass, but it would be very surprising if it were not found soon. It has proved possible to put upper and lower bounds on its mass by studying radiative corrections to other processes[10]. The only other part of the model which has not been tested satisfactorily is the Higgs sector. There are no particularly stringent bounds on its mass, and actually detecting Higgs particles in accelerators will be quite challenging[11]. The Higgs field is the least satisfactory part of the model, being a fairly arbitrary, some might say inelegant, addition to the Lagrangian. Nevertheless, it would seem that something like spontaneous symmetry breakdown via a Higgs mechanism does actually occur, and so something like the Higgs particle must exist because it is needed to ensure renormalizability of the theory[7]. So while the details may not be quite correct it seems difficult to refute the general principle. Perhaps the next most important task at accelerator laboratories is therefore to map out this Higgs sector.

2.5 Extensions

If the standard model is so successful why should extensions to it be considered? One reason is the unsatisfactory nature of the Higgs sector just mentioned, but there are many others. For example why are there three families of quarks and leptons? The fact that this is the minimum number which allows the CP violation that is observed, seems insufficient. It would seem that the gauge group $SU(3) \times SU(2)_L \times U(1)_Y$ is not the full story either. Its three factors require three independent couplings, not a particularly unified description — the electroweak "unification" is really a redescription in light of spontaneous symmetry breaking — it still requires an experimentally determined value of $\sin^2 \theta_W$ to establish the relationship between g and g'. There is also the question of why the weak force should distinguish between left and right handed fermions. And why *are* there two types of fundamental fermion, the coloured quarks and the colourless leptons, which then, apparently coincidentally, turn out to give the electron and proton exactly equal and opposite charges? Yet another unsatisfactory feature is that gravity is not included in the model. Unfortunately, attempts to quantise gravity by itself encounter very many problems[12], which motivates attempts to include it in a unified "theory of everything".

There are a multitude of extensions to the standard model which attempt to explain some or all of the problems outlined above. Some of those which will be relevant in later chapters are briefly sketched below.

Perhaps the least revolutionary extension is the inclusion of neutrino masses[13]. These can easily be included in the standard model Lagrangian in the same way the other fermion masses are, as terms involving the vacuum expectation of the Higgs. The general standard model picture in the previous section remains the same, but there can now be right handed neutrinos, and mixing is possible between the lepton families in a similar way to the quarks. There is as yet no definite evidence for mixing or indeed for neutrino masses[14].

Attempts to unify the gauge group $SU(3) \times SU(2) \times U(1)$ into a single group G with a single unified coupling g are know as grand unified theories (GUTs)[15]. The idea is that, in analogy with the electroweak breaking of $SU(2) \times U(1)_Y$ to $U(1)_Q$ at

a scale M_W, spontaneous symmetry breakdown would occur in G at some larger scale M_X (or perhaps this might be a multi-stage process) to give the gauge groups of the standard model. The scale for M_X may be estimated by running the three standard model couplings, using the renormalization group equation, to an energy where they all have roughly the same magnitude (see fig. 2.4). It is found that this occurs at $M_X \sim 10^{14}$ GeV. There are a number of candidates for the grand unified group, the most natural choices starting with $SU(5)$, the minimal model, then $SO(10)$ and E_6 though there are many more complicated choices that could be made. Besides unifying the strong and electroweak forces, GUTs have some other nice features; e.g. the equal and opposite charges of electrons and protons is explained by the fact that quarks and leptons occur in the same representation of G. They can (apart from the minimal SU(5) model) also explain, via the "see-saw mechanism"[15], why the neutrino masses are so small. The major problem with GUT theories is the difficulty they have in keeping separate the two very different mass scales, the GUT scale and the electroweak scale. Quadratic divergences, which occur in the renormalization of scalar loop diagrams (see fig. 2.1(b)), will drive all masses up to the GUT scale (unless the parameters are finely-tuned) whereas in reality all the known particles have masses less than the electroweak scale. This serious difficulty is known as "the hierarchy problem".

Another widely considered extension of the standard model is supersymmetry (SUSY)[16]. This postulates a symmetry between fermions and bosons so that every fundamental particle has a supersymmetric partner obeying the opposite statistics. Despite the fact that this requires a doubling the number of particles, the extra symmetry confers a number of desirable advantages. One is that the quadratic divergence usually associated with a scalar particle is automatically cancelled by a similar term from its fermionic partner. It is therefore possible to construct SUSY GUTs which avoid the hierarchy problem. Of course, SUSY must itself be broken because no supersymmetric pairs of particle with degenerate masses have been seen. The search continues for the lightest supersymmetric partner (LSP) to the known particles which is predicted to be stable. The scale of SUSY breaking M_S has variously been estimated to be in the range 10^3 GeV$< M_S < 10^{11}$ GeV. Hints at this scale have been

60

50

40

$1/\alpha_i$

30

20

10

0

$1/\alpha_3$

$1/\alpha_{GUT}$

$1/\alpha_2$

$1/\alpha_1$

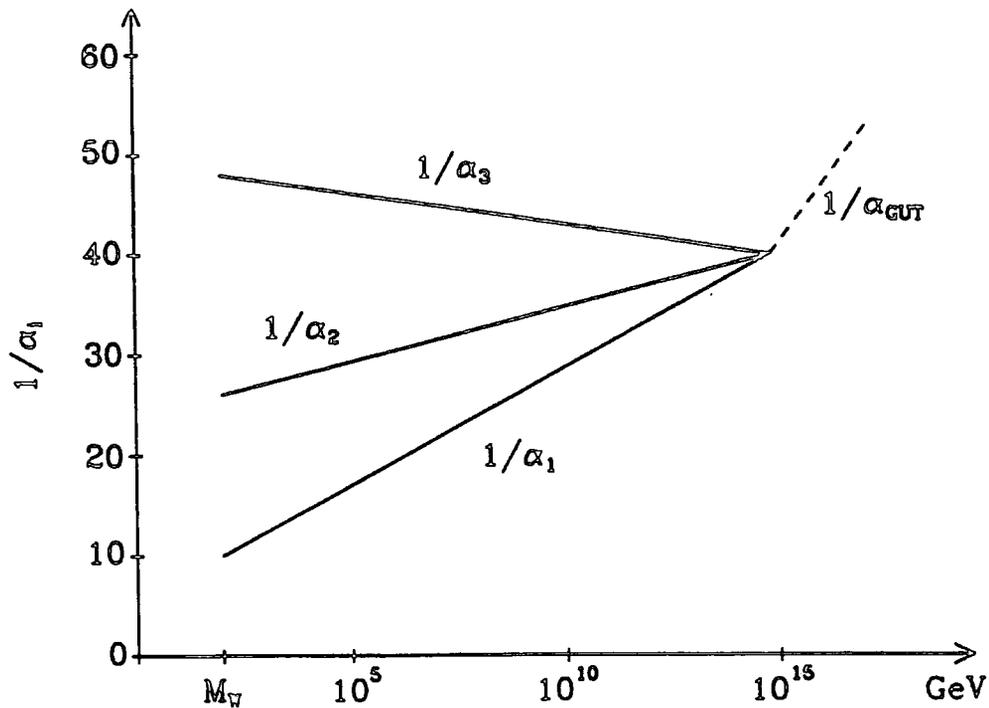M_W 10^5 10^{10} 10^{15} GeV

Figure 2.4: The running of the three standard model couplings with the energy scale. Where $\alpha_i = g_i^2/4\pi$, $g_3 \equiv g_s$, $g_2 \equiv g'$ and $g_1 \equiv g$.

obtained from plots like 2.4 which incorporate the additional effect of a SUSY breaking scale. Such plots show that ordinary grand unified SU(5) is ruled out but SUSY SU(5) can be supported and low values of $M_S \sim 1$ TeV are favoured[17].

If an attempt is made to make a theory invariant under local SUSY transformations, it is found that spin-2 particles must be introduced (as well as spin-3/2 particles) and these are identified with the graviton. The resulting theory is known as Supergravity (SUGRA)[18] and offers the possibility of unifying gravity with the other fundamental forces.

Two other attempts to unify all the forces of nature should also be mentioned. The first, Kaluza-Klein theories[19], attempt unification in a purely geometrical way by proposing that the gauge symmetries of the standard model arise from space-time symmetries (as in general relativity, see next chapter) in extra compactified dimensions. A realisation of the standard model would require seven extra compact dimensions. The second is superstrings[20], a quantum theory of one-dimensional

objects. This theory is finite, i.e., does not require renormalization, in versions of the theory with either 10 or 26 dimensions, which would mean 6 or 22 dimensions are compactified in a similar manner to the Kaluza-Klein theories. It is also possible to incorporate both the SUSY and GUT ideas into superstring models. There are no hints as yet of whether such theories have anything to do with reality.

There are of course many other additions and extensions that are not mentioned here. Very important to the rest of this thesis, however, is the concept that the fundamental theories of particle physics, when considered at different energy scales, may suffer transitions in the symmetries that they naturally exhibit. Whether such transitions are due to a spontaneous symmetry breaking mechanism or something more exotic is probably immaterial. But, they could have had quite a dramatic effect during the evolution of the very early universe as will be seen in the next chapter.

References

1. C. Itzykson and J-B. Zuber, *Quantum Field Theory*, (McGraw Hill, Singapore, 1985)

2. T-P. Cheng and L-F. Li, *Gauge Theory of Elementary Particle Physics*, (Oxford University Press, Oxford, 1984)

3. F. Halzen and A.D. Martin, *Quarks and Leptons*, (Wiley, Singapore, 1984)

4. J.D. Bjorken and S.D. Drell, *Relativistic Quantum Fields*, (McGraw-Hill, New York, 1965)

5. R.P. Feynman and A.R. Hibbs, *Quantum Mechanics and Path Integrals*, (McGraw Hill, 1965)

6. J. Collins, *Renormalization*, (Cambridge University Press, Cambridge, 1984)

7. P.D.B. Collins, A.D. Martin and E.J. Squires, *Particle Physics and Cosmology*, (Wiley, New York, 1989)

8. M.R. Pennington, *Rep. Prog. Phys.* **46**, 393 (1983)

9. D. Treille, CERN-PPE-92-107, C92-03-22, (1992)

10. J. Ellis, G.L. Fogli and E. Lisi, BARI-TH-111-92, CERN-TH-6568-92, (1992)

11. J.F. Gunion et al., *The Higgs Hunters Guide*, (Addison-Wesley, Redwood City, CA, 1990)

12. C.J. Isham, R. Penrose and D.W. Sciama (Eds.), *Quantum Gravity 2* (Oxford University Press, New York, 1981)

13. J.D. Vergados, *Phys. Rep.* **133**, 1 (1986)

14. J. Schechter, SU-4240-513, C92-05-25.1, (1992)

15. G.G. Ross, *Grand Unified Theories*, (Benjamin/Cummings, Reading, MA, 1984)

16. S. Ferrara (Ed.), *Supersymmetry*, (North-Holland/World Scientific, Singapore, 1986)

17. U. Amaldi et al., *Phys. Lett. B* **260**, 447 (1991)

18. J. Wess and J. Bagger, *Supersymmetry and Supergravity*, (Princeton University Press, Princeton, 1983)

19. H.C. Lee (Ed.), *An Introduction to Kaluza-Klein Theories*, (World Scientific, Singapore, 1984)

20. M.B. Green, J.H. Schwarz and E. Witten *Superstring Theory*, Vols. 1 and 2, (Cambridge University Press, Cambridge, 1987)

3

The Standard Model of Cosmology

The physics of the Universe on the very largest scales seems to be well described by General Relativity (GR). Despite the fact that GR is a purely classical theory, because its coupling to matter is so weak compared with the other fundamental forces, it is expected to provide a good description of the gravitational interaction up to energy scales of E_{Pl}, and thus of the evolution of the Universe since approximately t_{Pl}. The aim of this chapter is to introduce concepts and define symbols in common cosmological use which will be employed later. General Relativity together with the assumption of large scale homogeneity and isotropy, described in section 3.1, and embodied in the Friedmann equations, section 3.2, is the foundation upon which the standard model of cosmology is based. This standard model — "the hot big bang" — is remarkably successful and is briefly reviewed in sections 3.3 and 3.4 along with some of the supporting evidence.

There are still a number of fundamental problems which big bang cosmology does not address satisfactorily, though it is the model's very success that enables such questions to be asked at all. Some of these outstanding questions are described in section 3.5. One of the better motivated additions to the standard model, known as inflation, which can neatly resolve several of these problems at once, is described in the penultimate section. At present inflation is very much a paradigm, and though it seems quite probable that it is deeply rooted in the fundamental theories of particle physics, a compelling model has not yet been obtained. An account of some of the

better known inflation mechanisms concludes the chapter.

3.1 Einstein's Equation

The fundamental equation of General Relativity, Einstein's Equation, like the equations of motion of all classical theories, can be derived from an action principle (2.2). The action to consider is[1]

$$S = \int d^4 x \sqrt{g} \left(-\frac{1}{G} \left(\mathcal{R} + 2\Lambda \right) + 16\pi \mathcal{L}_{\text{matter}} \right) \tag{3.1}$$

where $\sqrt{g} \equiv \det \left(g^{\mu\nu} \right)$, $g^{\mu\nu}$ being the metric tensor, and Λ is a cosmological constant. The Ricci scalar $\mathcal{R}$ is defined as a contraction of the Ricci tensor $\mathcal{R}^{\mu\nu}$ which is in turn defined in terms of the Riemann curvature tensor $\mathcal{R}^{\mu}{}_{\nu\alpha\beta}$, i.e.,

$$\mathcal{R} = \mathcal{R}^{\mu}{}_{\mu}, \quad \mathcal{R}_{\mu\nu} = \mathcal{R}^{\sigma}{}_{\mu\nu\sigma} \tag{3.2}$$

the curvature tensor being

$$\mathcal{R}^{\mu}{}_{\nu\alpha\beta} = \partial_\alpha \Gamma^{\mu}{}_{\nu\beta} - \partial_\beta \Gamma^{\mu}{}_{\nu\alpha} + \Gamma^{\mu}{}_{\sigma\alpha} \Gamma^{\sigma}{}_{\nu\beta} - \Gamma^{\mu}{}_{\sigma\beta} \Gamma^{\sigma}{}_{\nu\alpha} \tag{3.3}$$

where

$$\Gamma^{\mu}{}_{\alpha\beta} = \frac{1}{2} g^{\mu\sigma} \left(\partial_\beta g_{\sigma\alpha} + \partial_\alpha g_{\sigma\beta} - \partial_\sigma g_{\alpha\beta} \right) \tag{3.4}$$

is the connection. $\mathcal{L}_{\text{matter}}$ contains all the matter and radiation fields in the universe, so for example it will contain the standard particle physics model Lagrangian described in section 2.4, but presumably only as a subset. It is interesting to note that the invariance of the purely geometrical part of (3.1) (i.e., all but the term in $\mathcal{L}_{\text{matter}}$) under local Lorentz transformations and general coordinate transformations can be shown to possess a gauge-like structure remarkably similar to the gauge construction resulting from the symmetry under local phase transformations that is used for the other fundamental forces reviewed in section 2.2[2].

Variation of (3.1) with respect to the metric leads to

$$\delta S = \int d^4x \sqrt{g} \left(\frac{1}{G} \left(\mathcal{R}^{\mu\nu} - \frac{1}{2}\mathcal{R}g^{\mu\nu} - \Lambda g^{\mu\nu} \right) + 8\pi T^{\mu\nu} \right) \delta g_{\mu\nu} \qquad (3.5)$$

where $T^{\mu\nu}$ is the stress-energy tensor for the fields in $\mathcal{L}_{\text{matter}}$, i.e., (2.4). (Actually, (2.4) is not necessarily a symmetric tensor whereas it must be in (3.5) since the metric and Ricci tensors are. The energy momentum tensor defined by (2.4) is not unique, however. It can be made symmetric by adding an appropriately chosen term $\partial_\lambda f^{\lambda\mu\nu}$ where $f^{\lambda\mu\nu}$ satisfies $f^{\lambda\mu\nu} = -f^{\mu\lambda\nu}$ and hence $\partial_\mu\partial_\lambda f^{\lambda\mu\nu} = 0$ which implies the total momentum (2.7) and energy-momentum conservation law (2.6) are not affected by such an addition[3].) The principle (2.2) thus yields

$$G^{\mu\nu} \equiv \mathcal{R}^{\mu\nu} - \frac{1}{2}\mathcal{R}g^{\mu\nu} = 8\pi G T^{\mu\nu} + \Lambda g^{\mu\nu}; \qquad (3.6)$$

which is Einstein's equation, $G^{\mu\nu}$ being the Einstein tensor. These equations are highly non-trivial, and to solve them assumptions must be made about the form of the metric and energy-momentum tensors. These assumptions depend on the case under study. For example, in the weak field limit where $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, $h_{\mu\nu}$ being a small correction to the flat metric (2.5), Newton's law of Gravitation is recovered from (3.6). Note also that gravity couples to the *energy* in (3.6), so that absolute values of potential energy affect the equations of motion, in contrast to all the other fundamental interactions for which only gradients of the potential (i.e., forces) have any consequence.

In a cosmological context the assumption usually made concerning the metric is that on the very largest scales (i.e., much larger than the typical supercluster size of 10^{24} m $\approx$ 30 Mpc) the Universe is, to a good approximation, homogeneous and isotropic. The major observation supporting this assumption is the homogeneity and isotropy of the cosmic microwave background radiation (CMBR). The temperature difference between microwave antennas separated by angles between 10 arc seconds and $180°$ being uniform to better than one part in 10^4 (though minute fluctuations at a level of about one part in 10^5 have recently been reported from the Cosmic Background Explorer (COBE) satellite[4]). Additional evidence is provided by the X-ray
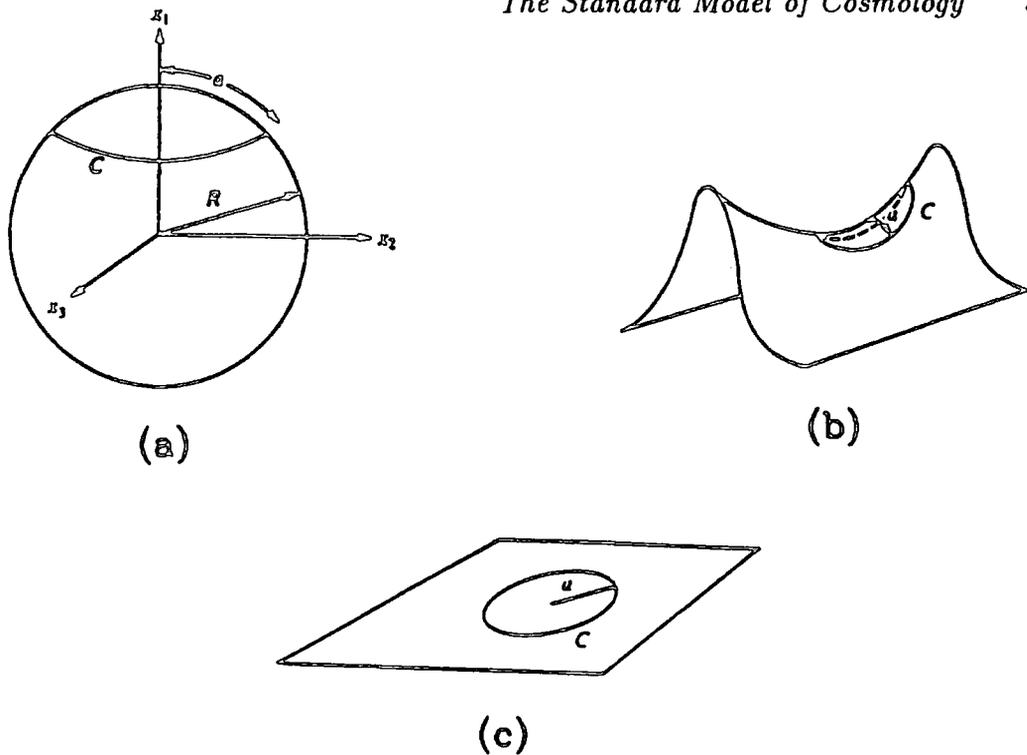
Figure 3.1: Two dimensional surfaces illustrating the difference between: (a) positive curvature, a sphere $x_i^2 = R^2$. On the sphere a circle of radius a has a circumference $C < 2\pi a$. (b) negative curvature, a saddle. Here a circle of radius a has a circumference $C > 2\pi a$. (c) zero curvature, a plane. The circle has a radius $C = 2\pi a$.

background radiation[5], the distribution of faint radio sources[6] and, with rather less certainty, the distribution of the galaxies themselves[7]. Homogeneity and isotropy imply that the spatial coordinates may be parameterised by a single scale describing the characteristic sign and "size" of a three dimensional hypersurface of constant curvature. Positive, negative or zero curvature, can be visualised for two dimensional surfaces as a sphere, a saddle or a plane as shown in fig. 3.1 (though actually the analogy is somewhat incomplete for negative curvature because the hyperbolic plane cannot be completely embedded in three dimensional space[8]). Furthermore, if all observers on this hypersurface are considered equal — the "cosmological principle" — the time coordinate may be separated from the rest and the scale size is then just a function of time. Under these assumptions the line element has the Robertson-Walker form[1]

$$ds^2 = dt^2 - R(t)^2 \left\{ \frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right\} \tag{3.7}$$

where $R(t)$ is the time dependent scale size of the universe and the parameter $k = +1$, -1 or 0 describes universes with positive, negative or zero curvature. The "comoving coordinates" (r, θ, ϕ) vary over the hypersurface for a given value of the scale size, the coordinate r being dimensionless and chosen to vary between 0 and 1. The metric is thus diagonal with components

$$g_{tt} = 1, \quad g_{rr} = \frac{-R(t)^2}{1 - kr^2}, \quad g_{\theta\theta} = -R^2 r^2, \quad g_{\phi\phi} = -R^2 r^2 \sin^2 \theta \tag{3.8}$$

and the non-zero components of the Ricci tensor are (from (3.2)–(3.4))

$$\mathcal{R}_{tt} = \frac{3\ddot{R}}{R}, \quad \mathcal{R}_{ii} = \frac{R\ddot{R} + 2\dot{R}^2 + 2k}{R^2} \tag{3.9}$$

while the Ricci scalar is (3.2)

$$\mathcal{R} = \frac{6}{R^2} \left(R\ddot{R} + \dot{R}^2 + k \right). \tag{3.10}$$

The other major ingredient needed to build a cosmological model is the nature of the energy-momentum tensor. For a Friedmann-Robertson-Walker (FRW) Universe, all matter and radiation is treated as if it were a uniform perfect fluid of energy density ρ and pressure p. Moreover, to be consistent with the symmetries of the metric it must be at rest with respect to the comoving coordinates, so that

$$T^\mu_{\ \nu} = \begin{pmatrix} \rho & & & \\ & -p & & \\ & & -p & \\ & & & -p \end{pmatrix}. \tag{3.11}$$

This assumption becomes an even more accurate description during the early universe when all the particles are considered to be in thermal equilibrium, to be discussed in section 3.4. Note that (3.11) is not the only energy-momentum tensor which satisfies

the symmetries of the metric (3.7); a fluid with bulk viscosity would also serve. The form (3.11) allows the cosmological constant to be included in the energy-momentum tensor $T^{\mu\nu}$ as a contribution $g^{\mu\nu}\Lambda/8\pi G$. It is written in terms of a "vacuum energy density" ρ_v, i.e.,

$$\frac{\Lambda}{3} \equiv \frac{8\pi G}{3}\rho_v \tag{3.12}$$

which implies from (3.11) that "vacuum pressure" p_v is negative, i.e., $p_v = -\rho_v$. In the standard picture of the universe Λ or ρ_v is assumed to be zero, or at least very small on the Planck scale, to agree with observational evidence today, but there will be more to say on this point when inflation is considered.

The 0-component of the energy-momentum conservation law $D_\nu T^{\mu\nu} = 0$ (where D_ν indicates a covariant derivative with respect to x_ν, $D_\nu T^{\mu\nu} = \partial_\nu T^{\mu\nu} + \Gamma^\mu_{\ \sigma\nu}T^{\sigma\nu} + \Gamma^\nu_{\ \sigma\nu}T^{\mu\sigma}$ i.e., it is the equivalent of (2.6) in curved coordinates), together with (3.11), gives the first law of thermodynamics — usually expressed in the form[1]

$$\frac{d}{dt}(\rho R^3) = -p\frac{d}{dt}(R^3). \tag{3.13}$$

In other words, the change in the energy equals (the negative of) the pressure times the change in volume, or $dE = -p\,dV$.

3.2 The Friedmann Equations

If the Robertson-Walker metric (3.7) and the perfect fluid stress-energy (3.11) are substituted into Einstein's equation (3.6), then the 0-0 component gives

$$H^2 \equiv \left(\frac{\dot{R}}{R}\right)^2 = \frac{8\pi G\rho}{3} - \frac{k}{R^2} + \frac{\Lambda}{3} \tag{3.14}$$

whilst the i-i component minus (3.14) gives

$$\frac{\ddot{R}}{R} = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}. \tag{3.15}$$

The equations (3.13), (3.14) and (3.15) can in fact be related by the Bianchi identities and so only two are independent. With the assumption $\Lambda = 0$ (3.14) and (3.15) are
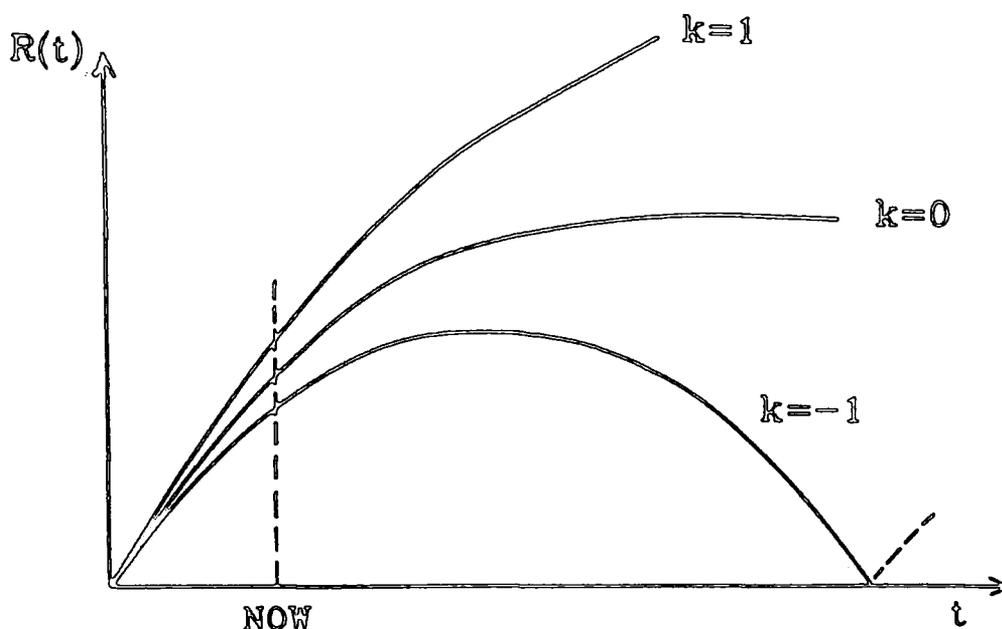
Figure 3.2: The evolution of the scale size of the Universe R as a function of time t for open $k = -1$, closed $k = +1$ and flat $k = 0$ FRW universes.

related by differentiation. For any k these equations predict expansion of the scale size R from a singularity at $R = 0$, the rate of this expansion being characterised by the Hubble parameter H. This universal expansion was discovered by Hubble in 1929 and today, the Hubble constant H_0, has a value

$$H_0 = h_0 \times 100 \ \mathrm{km\,s^{-1}Mpc^{-1}} \qquad (3.16)$$

where h_0 represents the uncertainty in the many observational determinations of H_0. The range $0.4 < h_0 < 1$ encompasses all recent observations[8].

When $\Lambda = 0$ in (3.14) it is known as the Friedmann equation and is often rewritten in terms of a critical density ρ_C defined as the density of the $k = 0$ universe, i.e., from (3.14)

$$\rho_C \equiv \frac{3H^2}{8\pi G} \qquad (3.17)$$

so that the Friedmann equation becomes

$$\frac{k}{H^2 R^2} = \Omega - 1 \tag{3.18}$$

where $\Omega \equiv \rho/\rho_C$ is the ratio of the actual density to the critical density. There is thus a connection between k and Ω. For $k = -1$, i.e., $\Omega < 1$ ($\rho < \rho_C$), known as an "open universe", (3.14) becomes dominated by the curvature at large R (i.e., late times), so that eventually $R \sim t$ and the expansion continues forever. For $k = +1$, i.e., $\Omega > 1$ ($\rho > \rho_C$), a "closed universe", (3.14) implies there will be a value of R when $\dot{R} = 0$ and so the expansion will stop and the universe will thereafter contract, eventually ending in a "big crunch". The special case $k = 0$ or $\Omega = 1$ ($\rho = \rho_C$) is thus the dividing line between eternal expansion and eventual recollapse (see fig. 3.2). In this case (3.14) gives

$$H = \sqrt{\frac{8\pi G\rho}{3}} \tag{3.19}$$

which if ρ is just matter density (i.e., $\rho \sim R^{-3}$, see next section) implies $R \sim t^{2/3}$ for large t, so the universe continues to expand but ever more slowly. The parameter Ω thus determines the eventual fate of the universe. Unfortunately, the limits on Ω from observations are something like[2]

$$0.1 \lesssim \Omega \lesssim 4 \tag{3.20}$$

which does not distinguish between these three cases. In fact, as will be discussed in section 3.5, the surprise is that Ω should be within an order of magnitude of unity.

3.3 The Hot Big Bang

It is clear from the Friedmann equations (3.14), and (3.13), that to understand in detail the evolution of the scale factor R an equation of state is needed relating ρ and p. The very simple equation of state $p = w\rho$, where w is constant, suffices to describe the behaviour of a broad range of contributions that might appear in the energy-momentum tensor, so long as just one such contribution is dominant. Two

examples of particular importance in standard cosmology are radiation (i.e., a gas of either massless particles or massive particles with relativistic internal velocities), with $p = \rho/3$ (see next chapter), i.e., $w = 1/3$, and bulk matter at rest with $p = 0$, i.e., $w = 0$. By using (3.13), the behaviour of ρ with R can be obtained[1]

$$\rho \propto R^{-4} \quad \text{for radiation dominance } (i.e., p = \rho/3),$$
$$\rho \propto R^{-3} \quad \text{for matter dominance (i.e., } p = 0). \tag{3.21}$$

The curvature term in (3.14) has an R^{-2} dependence so for early enough times, as $R \rightarrow 0$, it can be ignored compared to (3.21). Using (3.21) in (3.14) thus gives

$$R \propto t^{1/2}, \quad H = \frac{1}{2t} \quad \text{for radiation dominance,}$$
$$R \propto t^{2/3}, \quad H = \frac{2}{3t} \quad \text{for matter dominance.} \tag{3.22}$$

So, it is clear that the radiation energy density is likely to be dominant at the earliest times. It is expected that for much of the early evolution of the universe, thermal equilibrium will be a good approximation. A gas of relativistic particles which is in thermal equilibrium can be described by a temperature T defined by $\rho_r = g a_{SB} T^4 / 2$ where a_{SB} is the Stefan-Boltzmann constant, $a_{SB} = \pi^2/15$, as will be seen in the following chapter. There are a number of reasons why thermal equilibrium is considered to be a good description for the Universe's earliest history. One is that at such high densities, the rates of interaction between particles would have been large enough to maintain equilibrium. This point will be analysed in greater detail in the next chapter. The two major pieces of observational evidence pointing to a thermal bath at earlier epochs are the blackbody spectrum of the CMBR[9] and the cosmological abundances of the light elements[10], both of which are discussed further in the next section.

Even if particles which were originally in equilibrium subsequently go out of thermal equilibrium due to the expansion, (whereupon they are said to "decouple"), then because they no longer undergo any interactions their initial thermal distribution of

velocities will persist, modified only by the universal expansion. It is thus usually possible to assign an effective temperature to this remnant distribution. It is found that $T \propto R^{-1}$ or $T \propto R^{-2}$ depending on whether the particles decouple when relativistic, $T \gg m$, or when non-relativistic, $T \ll m$ (see next chapter).

So, it is possible to characterise the history of early universe in terms of a temperature which is common to all the particle species in thermal equilibrium with each other, though the effective temperature of a species may become different once it decouples from the thermal plasma. This is the reason fundamental particle physics is important in the study of the early universe — a high temperature implies that high energy collisions occur frequently. This thermal history and its connection with particle physics is considered next.

3.4 Thermal History of the Universe

The standard cosmological "big bang" model is summarised by the thermal history time line shown in fig. 3.3. The early events in this history are obviously speculative, whilst many of the later ones are quite well understood. Some of the more important of these events are now reviewed in a little more detail.

As the universe cools it should undergo a number of phase transitions. If the GUT and/or SUSY ideas of high energy physics (section 2.5) are correct, then phase transitions due to the spontaneous symmetry breaking of these symmetries should occur at about $T \sim 10^{14}$ GeV for GUTs and somewhere between 10^{11} and 10^{3} GeV for SUSY. It also seems likely that a transition corresponding to electroweak symmetry breaking will have occurred at about 10^2 GeV. Many interesting phenomena may be associated with these phase transitions, for example, baryosynthesis[11] and inflation (see section 3.6). A further transition resulting from the confinement of the quarks in colour-singlet hadronic states, when α_s becomes strong (see section 2.4) should also have occurred at around 100 to 300 MeV[12].

After this comes the very important epoch of nucleosynthesis at temperatures of around 10 to 0.1 MeV. Below these energies, processes like

$$n + \nu \longleftrightarrow p + e^-, \quad e^+ + n \longleftrightarrow p + \bar{\nu}, \quad n \longleftrightarrow p + e^- + \bar{\nu}, \qquad (3.23)$$
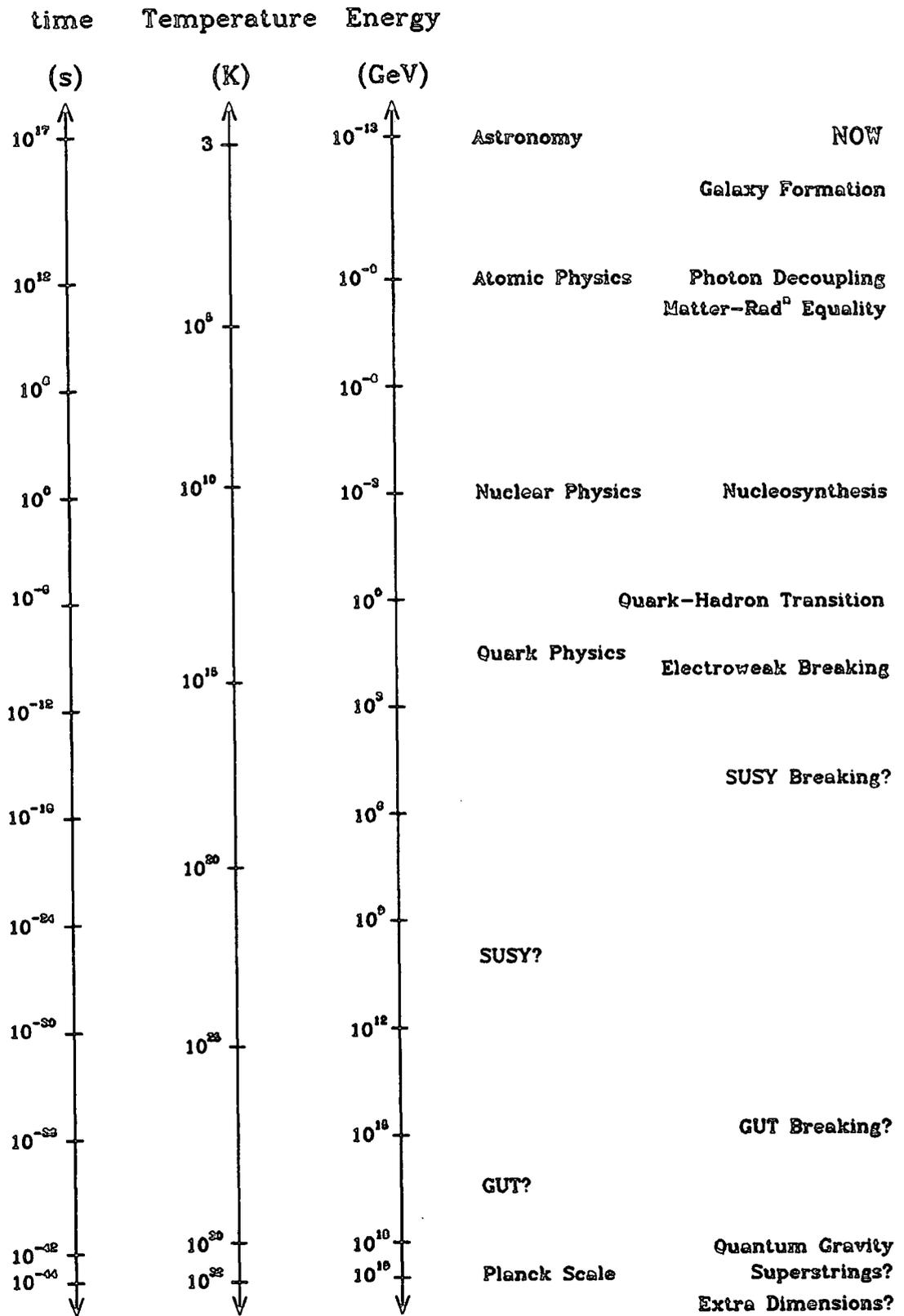
Figure 3.3: A thermal history timeline for the standard big bang model of cosmology.

become forbidden and light elements will be formed out of the plasma of neutrons and protons. Nucleosynthesis has been studied in very great detail[10] and it provides stringent tests of the big bang model. It successfully predicts the primordial abundances of the light elements D, ^{3}He, ^{4}He and ^{7}Li as well as the absence of others. Successful predictions require that the number of light neutrino species N_ν is 3 ± 1 and that the ratio of the baryon to photon number density $\eta = n_B/n_\gamma$ is $4 \times 10^{-10} < \eta < 7 \times 10^{-10}$ in agreement with observation. This is a very strong indication that thermodynamic equilibrium is a good approximation to the conditions present at this time ($t \sim 10^{-2}$ to 10^2 s). It also places bounds on the contribution of ordinary (baryonic) matter to the cosmic density[8],

$$0.011 \leq \Omega_B \leq 0.21. \tag{3.24}$$

Also at about this time ($T \sim 1$ MeV) the neutrinos can no longer be kept in equilibrium via the leptonic reactions

$$\nu e \longleftrightarrow \nu e, \quad \nu \bar{\nu} \longleftrightarrow e^+ e^-, \quad \text{etc.} \tag{3.25}$$

They therefore decouple, suffering no further interactions and persisting until today with their own characteristic temperature.

At a time $\sim 10^{11} s$ or $T \sim 10$ eV the total energy density in non-relativistic particles will start to dominate over the total energy density in radiation. This is the time t_{eq} of matter-radiation equality and the start of the epoch of matter dominance in (3.14). It is the time at which structure formation can begin to grow via the gravitational collapse of locally overdense regions[6].

Shortly afterwards, at the decoupling temperature $t_d \approx 10^{13}$ s, photons will no longer be able to ionise Hydrogen atoms and will thus decouple from matter. The resulting thermal distribution of photons will, like the neutrinos, continue undistorted, but redshifted by the cosmic expansion, until today. It makes up the CMBR first discovered in 1965 by Penzias and Wilson. It has recently been observed by the COBE satellite[9] to have a perfect (within experimental error) blackbody spectrum

with $T_0 = 2.735 \pm 0.060$ so confirming thermodynamic equilibrium at the photon decoupling time to very great accuracy.

3.5 Problems with the Big Bang

There can be no doubt that the cosmological standard model is very successful, even if it has not yet been quantitatively tested to quite the extent of the particle physics model. It is also true that, in common with its particle physics counterpart, there are no confirmed observations which contradict it. Nevertheless, there are again a number of unsatisfactory, or at least unnatural features, which indicate that the big bang is probably not the whole story. Some of these are[13, 2, 8]:

The Horizon Problem: The maximum distance a photon can travel since $t = 0$ is known as the horizon distance d_H and can be calculated;

$$d_H(t) = R(t) \int\limits_0^t \frac{dt'}{R(t')}. \tag{3.26}$$

This is the maximum distance over which events can influence each other within time t, i.e., it defines a volume which is in "causal contact". Since, using (3.22) in (3.26), $d_H \approx t$ or $d_H \approx H^{-1}$ for both radiation and matter dominance, the fraction of the universe that can be observed at time t is

$$\frac{d_H(t)}{R(t)} \sim \begin{cases} t^{1/3} & \text{for matter dominated evolution,} \\ t^{1/2} & \text{for radiation dominated evolution} \end{cases} \tag{3.27}$$

which increases with time because the expansion rate is slower than that of light. The number of causally disconnected domains at the era of photon decoupling that can be observed now is $t_0/t_d \sim 10^{17}s/10^{13}s = 10^4$. It seems rather odd then, that the microwave background photons which have not interacted at all since this time, have the same temperature in each of these domains to an accuracy of almost one part in 10^5. The problem gets worse if earlier epochs are considered; the number of nucleosynthesis domains that can be observed now is $\sim 10^{25}$, so it is strange that in all of these domains the precise details of light element synthesis seems to have been the same.
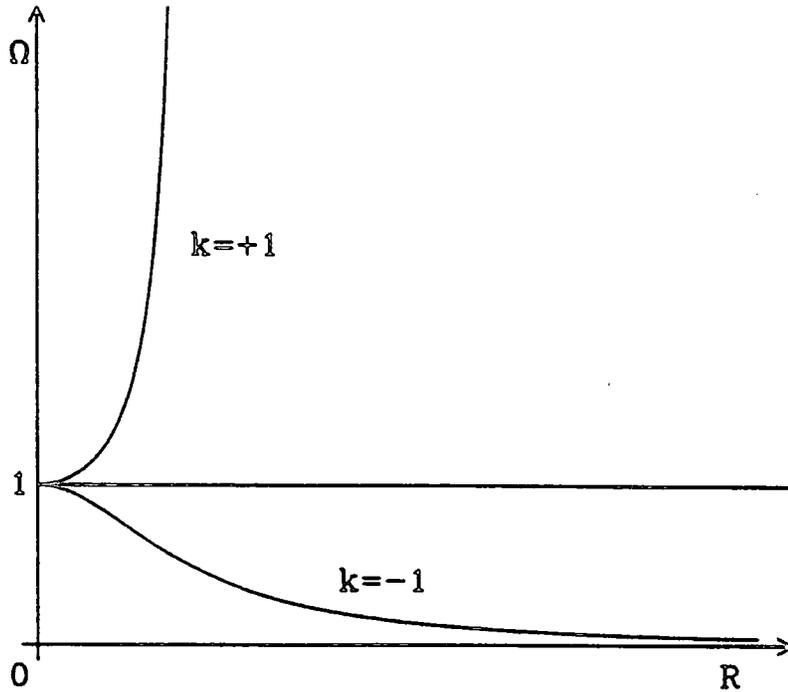
Figure 3.4: The variation of Ω versus R for either radiation or matter dominance illustrating the unstable nature of $\Omega = 1$. For $\Omega > 1$, Ω quickly diverges to ∞ while if $\Omega < 1$ it rapidly vanishes.

Another way of expressing this problem is that all the evidence points to a homogeneous, isotropic Universe. This high degree of smoothness means that the Universe can be described by a Robertson-Walker metric as indicated in sections 3.1 and 3.2. Yet this is obviously not the most general solution to Einstein's equation, there are many other examples including anisotropic and/or inhomogeneous space. Why then is this very special metric satisfactory?

Flatness or Naturalness: The Friedmann equation (3.14) may be rewritten in terms of Ω

$$\Omega^{-1} = 1 - \frac{3k}{8\pi G \rho R^2} \tag{3.28}$$

and using (3.21) for ρ shows that $\Omega = 1$ is an unstable fixed point. For Ω slightly larger than 1 it quickly diverges to ∞, whilst if it is slightly less it rapidly approaches zero (see fig. 3.4). In order that Ω be $O(1)$ today (which it is by observation, see (3.20)), it must have been $O(1 \pm 10^{-60})$ at the Planck time, in other words the density was

extremely close to critical density or, equivalently, the curvature was extremely close to zero.

Expressed as a naturalness condition; if all the parameters in the Friedmann equation (3.14) were of order one, then if $k = 1$ it would recollapse in a few Planck times whereas if $k = -1$ then it would reach a temperature of 3 K in about $10^{-11}s$. So the question could also be asked; why is the universe so old?

This problem could, of course, be solved by setting $k = 0$ or $\rho = \rho_c$ exactly, but the standard model offers no explanation of why this should be so.

Monopoles: A possible consequence of a symmetry that becomes spontaneously broken as the universe cools is the creation of "topological defects". In each causal domain at the epoch at which the breaking occurs, the details of symmetry breaking will be different. For example, if the symmetry breaking occurred via the mexican hat potential in fig. 2.3, a different point on the circle of minima would be chosen in each domain. Defects, where locally the field does not lie in the minimum of the potential, would form at the boundaries of these domains as the horizon expands. Depending on the topology of the space of minima and the group which is being broken, such defects may take the form of domain walls, cosmic strings, monopoles[14, 15], or textures[16]. GUTs in particular, notoriously produce monopoles which will have a mass of the order of the grand unified scale of $\sim 10^{14}$ GeV and will survive in such numbers that their density today will be something like $10^9 \rho_c$, i.e., orders of magnitude larger than the observed bound (3.20). The same problem can in fact be shown to occur for any stable superheavy relic particle. If a particle physics theory were proved to produce such objects there is no way that the standard big bang model could get rid of them.

Structure Formation: Despite the indications that the universe is smooth on the very largest scales there is obviously an abundance of structure on smaller scales; from stars to superclusters. Given a spectrum of initial density fluctuations in the thermal bath, it is possible to track the way in which these fluctuations would develop. This gravitational evolution is generally included in the standard model. However, the end result is completely dependent upon the initial fluctuation spectrum, and the standard cosmology gives no clue as to what this spectrum should be. It would

appear that some extension is required to determine the pattern of fluctuations. It seems likely that the recently observed fluctuations in the CMBR by COBE will help to narrow down the possibilities.

The Cosmological Constant: In the Friedmann equation (3.14) the cosmological constant Λ, equivalent to a vacuum energy, is conventionally set to zero or at least very small to agree with observation. However, there is no theoretical reason that explains this, only empirical evidence. Indeed, on naturalness grounds, a value of $\Lambda \sim m_{Pl}^4$ might be expected, a difference of something like 120 orders of magnitude. If the inflation mechanism, described in the next section, is accepted as valid, it becomes even more pressing to find a solution to this problem. Unfortunately, this seems to be a fundamental difficulty, probably only soluble in a complete quantum theory which includes gravity properly.

3.6   Inflation

It is possible to solve the first three and quite possibly the fourth of the above problems by a simple extension of the big bang model known as inflation. Inflation postulates that there was a brief era when vacuum energy ρ_v (or an effective cosmological constant Λ) was the dominant contribution in the right hand side of (3.14). ("Vacuum dominance" of (3.12) can be included as another case in (3.21) with $w = -1$, since vacuum pressure $p_v = -\rho_v$.) If this happens then the Friedmann equations, (3.14) and (3.15), become

$$\frac{\ddot{R}}{R} = \left(\frac{\dot{R}}{R}\right)^2 = \frac{\Lambda}{3} = \frac{8\pi G}{3}\rho_v = H_\Lambda^2 = const. \tag{3.29}$$

and hence

$$R(t) = R(t_s)e^{H_\Lambda(t-t_s)} \tag{3.30}$$

where $R(t_s)$ is the scale size at t_s. The universe would thus inflate exponentially during this period as in the cosmological model proposed by de Sitter. If inflation lasts until t_e then over the epoch of inflation $\Delta t \equiv t_e - t_s$ the universe will expand

by a factor $e^{H_\Lambda \Delta t}$. In (3.28), with $\rho = \rho_v = const.$ and using (3.30), it is found that $\Omega \to 1$ very rapidly, solving the flatness problem provided $e^{H_\Lambda \Delta t} > 10^{30}$ (or 50 "e-folds"). This would also solve the horizon problem because a single causal domain before inflation starts could be expanded exponentially far out of contact. So the fact that, for example, the same temperature is observed coming from domains out of causal contact at photon decoupling, is simply a reflection of the fact that they were in contact before the inflationary epoch. This also solves the monopole difficulty because any undesirable heavy relic will be diluted by $\left(e^{H \Delta t}\right)^3$ if it is produced before inflation starts. For 50 e-folds this implies there will be at most $O(1)$ primordial monopoles in the observable Universe.

It should be noted that the inflation scenario is very difficult to test. A period of inflation would all but erase the initial conditions from which it arose. Really, its only firm prediction is that $\Omega = 1$ to very high accuracy. Note however, that the possibility exists that this Ω may contain a contribution from a residual vacuum energy, i.e., using (3.14), the inflationary prediction would become $\Omega_{matter} + \Omega_\Lambda = 1$ where $\Omega_\Lambda \equiv \rho_v / \rho_C$. So for example, it may be the case that $\Omega_{matter} = 0.1$ and $\Omega_\Lambda = 0.9$, but such a tiny vacuum energy would be extremely difficult to detect. If it is true that $\Omega_{matter} = 1$ only (i.e., $\Lambda \equiv 0$), then because the upper bound on the density of ordinary baryonic matter (3.24) is significantly less than this, extra "dark matter" must be postulated to make up the difference. The interesting possibility is thus that most of the matter in the universe is actually non-baryonic.

If the mechanism for inflation were known precisely then it would be possible to predict the spectrum of primordial density fluctuations that it would generate. This might perhaps offer a solution to the structure formation problem above and enable the idea to be tested by comparing numerical simulations of structure evolution with real observations. This procedure has been tried for many inflation model candidates, but many of them generate similar spectra. A more direct way of testing such ideas should be the fluctuation spectrum at the epoch of decoupling, such as that which is now becoming available as a result of the temperature fluctuations in the CMBR observed by the COBE satellite[4]. (Though actually fluctuations measurements on smaller angular scales than those COBE is sensitive to, would be required for a proper

test here. Such small scale fluctuations should hopefully soon be detected by ground based experiments.)

3.7 Inflation Mechanisms

Providing the mechanism for inflation is one of the most important ways in which the link between cosmology and fundamental particle physics can be made. The standard scenario involves a spontaneous-symmetry-breaking potential V of some Higgs-like scalar field ϕ (see section 2.3). The relevant symmetry breaking is usually assumed to occur at a much larger energy scale than the electro-weak scale, e.g., the GUT or SUSY scale. Finite temperature field theory shows that in general such a potential will have its symmetry restored at high temperature, resulting in a temperature dependent potential like that of fig. 3.5, as will be discussed in Chapter 5. At high temperatures the absolute minimum is thus at $\phi = 0$ but as the universe cools through the critical temperature T_c for the transition, this changes to its broken symmetry value and it thus becomes energetically favourable for the field to evolve to this new minimum. As has already been mentioned, the size of the cosmological constant today is usually set to zero so, since gravity couples to absolute values in the potential, it has to be assumed that the minimum of the the zero temperature potential is at $V = 0$. Hence, before the field evolves to the new absolute minimum, it will have a vacuum energy $V(\phi = 0) - V(\phi = \phi_{\min}) \equiv \rho_v$. Of course, ignorance of a proper "theory-of-everything" has been hidden here, in that gravity has been coupled classically to the quantum field theory.

When the field falls into its new minimum, it will oscillate about this value and radiate other particles, so the Universe will reheat to a temperature which depends on the strength of the couplings to these particles but is no larger than at the start of inflation[17]. The vacuum energy is thus converted into matter and radiation and the Universe will evolve thereafter as in the standard big bang model.

The shape of the zero-temperature potential determines the way the evolution from $\phi = 0$ to $\phi = \phi_{\min}$ occurs. If there is a local maximum between the two minima it can be treated as a tunnelling process[13], a first-order phase transition. If this
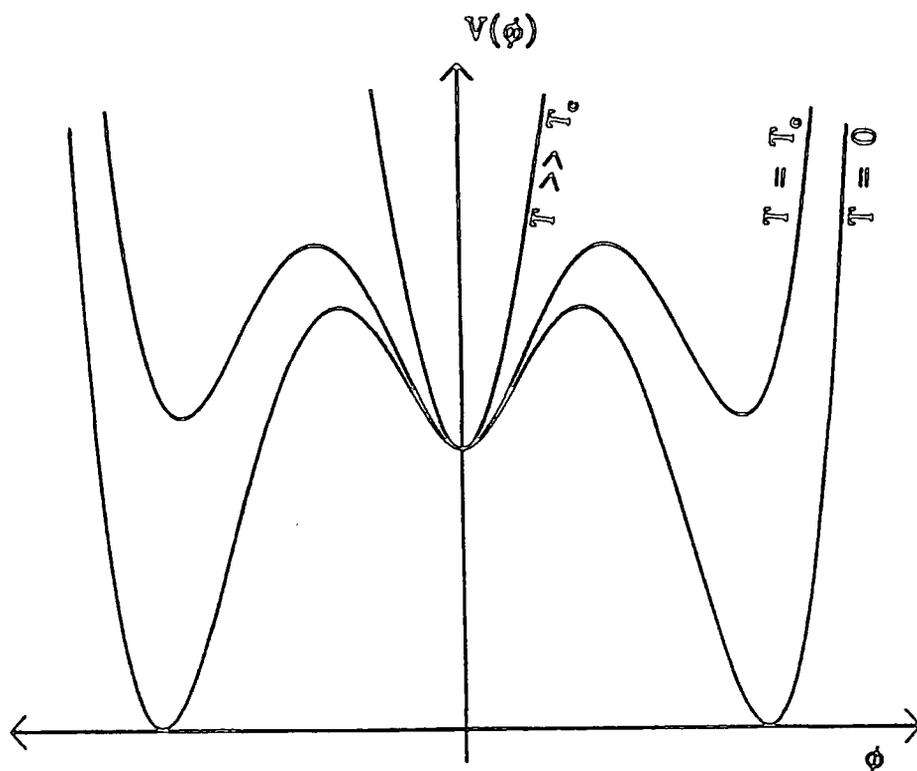
Figure 3.5: A typical temperature dependent effective scalar potential which can induce inflation. The symmetry broken by the $T = 0$ potential gets restored at high temperature. The field may get trapped in the $\phi = 0$ minimum as the universe cools below the critical temperature T_c, the point at which the asymmetric minimum and the minimum at $\phi = 0$ are degenerate. In the $\phi = 0$ minimum, a vacuum energy is generated.

local maximum is absent or at least vanishingly small, it can be modelled by "slow rolling"[18] or by "wave function evolution"[19], resulting in a second order transition.

The original inflation model proposed by Guth[13] was based on a first order phase transition. As long as the universe remains trapped in the false vacuum it will undergo inflation. The hoped for picture was that bubbles of normal phase would nucleate out of the inflating phase, grow, and eventually coalesce to form an expanse of normal-phase Universe. Collisions between bubble walls would serve to reheat the universe and also produce the spectrum of initial density fluctuations. There is a fatal flaw in this idea however; because it can be shown that the growth rate of bubbles of normal phase cannot compete with the exponential expansion of the inflating phase. Individual bubbles will thus only very rarely coalesce, and so percolation of the new phase will not occur and an inhomogeneous universe would be the end result. This
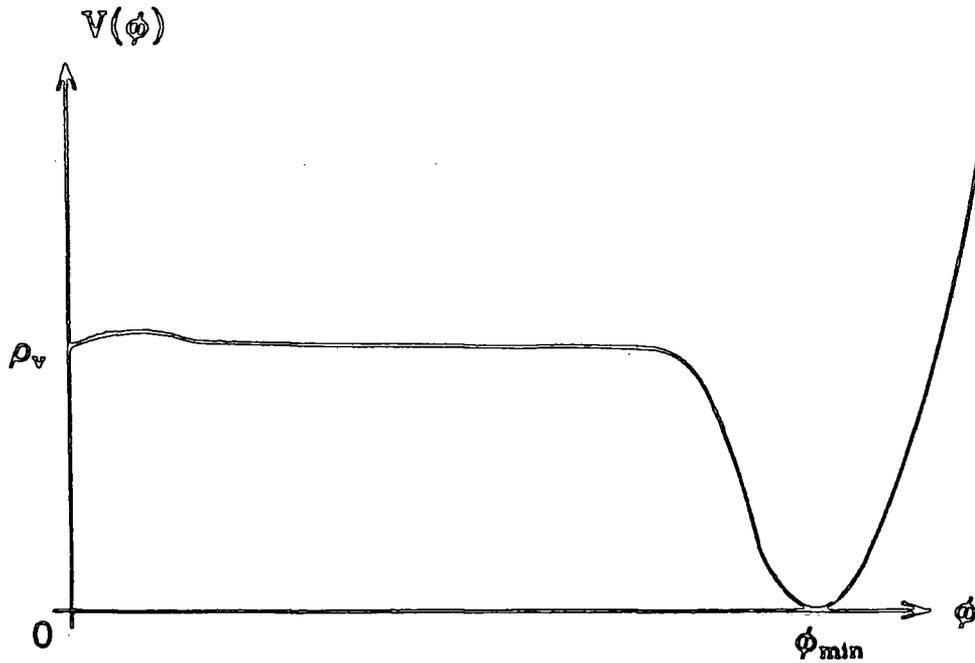
Figure 3.6: A zero temperature potential suitable for new inflation. A long flat plateau is required near the origin to obtain a long enough "slow-rolling" epoch.

is known as the "graceful exit" problem[20].

A solution to this problem, suggested by Albrecht, Steinhardt and Linde[18], is "new inflation" based on a second order phase transition. In this model the potential is assumed to be very flat near the origin with only perhaps a small barrier separating the minima (see fig. 3.6). The evolution of the vacuum along this flat part of the potential is described classically as "slow rolling". It is the time it takes the field to slow roll to its new minimum that determines the length of the period of inflation. The graceful exit problem is solved in theories of this type by postulating that the whole of the observable universe arises from within a single bubble of the new phase. In slow-rolling inflation, primordial density fluctuations arise because since ϕ is a quantum field it will undergo quantum fluctuations about its classical value which inflation then expands to observable sizes. The pattern of such fluctuations, due to their quantum origin, is Gaussian and it may be shown that the model predicts an approximately scale invariant spectrum of density perturbations (known as the
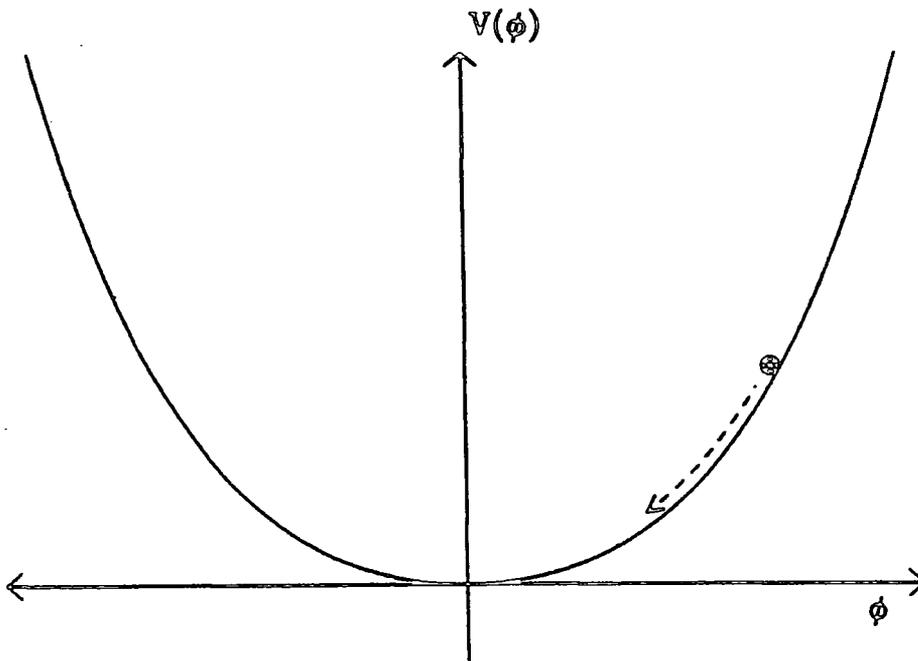
Figure 3.7: A simple potential suitable for chaotic inflation, with the field localised away from the minimum of the potential by "chaotic initial conditions".

Harrison-Zel'dovich spectrum[21]).

Unfortunately this new inflation suffers from a fine tuning problem. In order that the abovementioned fluctuations in the ϕ-field should not give rise to unacceptable anisotropies in the microwave background, the characteristic coupling in the potential for ϕ needs to be $O(10^{-14})$[2]. It is difficult to see how such small couplings could arise in a natural way from simple unified particle physics theories. (An exception is the superpotential in certain SUGRA models). Another serious objection is that with such a small coupling it would appear very unlikely that thermal equilibrium would be established at any stage, and there would then seem to be no reason why the ϕ-field should be initially localised near the origin[22, 23]. This matter shall be explored more fully in chapter 6.

Such concerns have lead to inflationary mechanisms which do not rely on a thermal phase change at all. The foremost example of this type is Linde's chaotic inflation[22]. In this model only a very simple form for the potential, say $V = \lambda\phi^4$ or even just

$m^2\phi^2/2$, need be assumed. With a potential like this there is no reason why the field should be localised at a minimum of the absolute minimum at the Planck time because it has not been in existence for a sufficient time to settle there. The scenario now runs as follows: consider a "patch" (this concept is, at the Planck scale, necessarily rather vague) of universe where the field is approximately localised, but is *not* at the minimum of the potential (see fig 3.7). If the patch is large enough and the starting point for ϕ is sufficiently far up the potential then it seems probable that the required slow-rolling inflation will occur as the field evolves towards the minimum at $\phi = 0$. Chaotic inflation certainly shows that inflation is a rather general consequence of coupling field theories to gravity. But there remains a concern about how valid it can be to couple the potential to gravity classically so near to the Planck scale when only a proper quantum theory of gravity can hope to provide a satisfactory description. Attempts have been made in this direction by linking chaotic inflation with the "wavefunction of the universe" description[17, 8].

To many people however, the most natural inflation mechanism is still the one based on a phase change, particularly the original first-order idea. There have thus been several attempts to overcome the graceful exit problem. The two possible solutions are to modify the expansion rate H during inflation or to modify the bubble nucleation rate. The very popular example of the first type is the recently proposed extended inflation[24] which involves modifying the gravitational action of GR (3.1) and will be considered in greater detail in Chapter 7.

References

1. S. Weinberg, *Gravitation and Cosmology*, (Wiley, New York, 1972)

2. P.D.B. Collins, A.D. Martin and E.J. Squires, *Particle Physics and Cosmology*, (Wiley, New York, 1989)

3. L.H. Ryder, *Quantum Field Theory*, (Cambridge University Press, Cambridge,

1985)

4. G.F. Smoot et al., COBE preprint "Structure in the *COBE* DMR First Year Maps" (1992)

5. E. Boldt, *Phys. Rep.* **146**, 215 (1987)

6. P.J.E. Peebles, *The Large Scale Structure of the Universe*, (Princeton University Press, New Jersey, 1980)

7. S.J. Maddox, et al., *Mon. Not. Roy. Astron. Soc.*, **242**, 43p, (1990)

8. E.W. Kolb and M.S. Turner, *The Early Universe*, (Addison-Wesley, Reading, MA, 1990)

9. G.F. Smoot et al., *Ap. J.*, **371**, L1, (1991)

10. A. Boesgaard and G. Steigman *Ann. Rev. Astron. Astrophys.*, **23**, 319, (1985)

11. M.E. Shaposhnikov, V.A. Rubakov and V.A. Kuzmin, *Phys. Lett. B* **155**, 36 (1985)

12. K.A. Olive, UMN-TH-1023-92, C92-01-13.2

13. A.H. Guth *Phys. Rev. D* **23**, 347 (1981)

14. A. Vilenkin, *Phys. Rep.* **121**, 263 (1985)

15. T.W.B. Kibble, *J. Phys.*, **A9**, 1387, (1976)

16. N. Turok, *Phys. Rev. Lett.* **63**, 2625 (1989)

17. A.D. Linde, *Particle Physics and Inflationary Cosmolgy*, (Harwood Academic, Chur, Switzerland, 1990)

18. A.D. Linde *Phys. Lett. B* **108**, 389 (1982); *Phys. Lett. B* **114**, 431 (1982); A. Albrecht and P.J. Steinhardt *Phys. Rev. Lett.* **48**, 1220 (1982)

19. A. Guth and S.Y. Pi, *Phys. Rev. D* **32**, 1899 (1985)

20. A. Guth and E. Weinberg, *Nucl. Phys.* **B212**, 321 (1983)

21. E.R. Harrison, *Phys. Rev. D* **1**, 2726 (1970); Ya. B, Zel'dovich, *Mon. Not. Roy. Astron. Soc.*, **160**, 1p, (1972)

22. A.D. Linde *Phys. Lett. B* **129**, 177 (1983)

23. G. Mazenko, W. Unruh, R. Wald, *Phys. Rev. D* **31**, 273 (1985)

24. D. La and P.J. Steinhardt, *Phys. Rev. Lett.* **62**, 376 (1989)

4

Thermodynamics

There can be no doubt that whether or not thermodynamic equilibrium became established in the various epochs of the early Universe is of great importance for understanding how it has evolved, as was seen in sections 3.3 and 3.4. The purpose of this chapter is to examine when the assumption of equilibrium is likely to be a valid. This issue is of particular relevance to inflation because it is a manifestly non-equilibrium process — the exponential expansion would quickly make particle interactions very rare. Section 4.1 reviews briefly the fundamentals of statistical mechanics and thermodynamics and introduces some of the thermodynamic variables which will be used later. The validity of equilibrium in the expanding universe is then explored in section 4.2 and expressed in terms of the particle interaction rate and the total number of interactions per particle. The last section explains how to calculate the interaction rate for a gas of inflaton particles.

4.1 Thermodynamic Equilibrium

Thermodynamic properties are naturally described as arising from the statistical behaviour of a system containing a large number of possible microstates. Here, it is required to know the behaviour of a system of particles described asymptotically by the free-field k-states considered in section 2.1. Interactions between the particles, necessary for establishing and maintaining equilibrium, are assumed to occur very

briefly and infrequently compared to the average time between such interactions. This is the ideal gas approximation[1].

Consider a complete set of orthonormal states $|nj\rangle = [a^\dagger(\mathbf{k}_1)]^{\bar{n}_1} \ldots [a^\dagger(\mathbf{k}_m)]^{n_m}|0\rangle$ i.e., states containing n_1 particles with momentum $\mathbf{k}_1$, and so on, in which j represents all the possible momentum eigenstates for states which have $n = n_1 + \cdots + n_m$ particles. This set of states is normalized in a large box of volume V so that the momentum eigenstates may be enumerated, and in the limit $V \to \infty$ the k-states and the annihilation and creation operators a and $a^\dagger$ become those introduced in section 2.1. Now consider a statistical ensemble of $\mathcal{N}$ wavefunctions, the lth member of which is $|r\rangle_l$, $l = 1, \ldots, \mathcal{N}$, where each $|r\rangle$ may be an arbitrary linear combination of the states $|nj\rangle$. The average properties of an ensemble are conveniently expressed in terms of the density operator which is defined by the matrix elements[1]

$$\langle nj|\hat{\rho}|n'j'\rangle = \frac{1}{\mathcal{N}} \sum_{l=1}^{\mathcal{N}} \langle nj|r\rangle_l \,_l\langle r|n'j'\rangle \tag{4.1}$$

i.e., it is the ensemble average of $\langle nj|r\rangle\langle r|n'j'\rangle$. So there are two probabilistic interpretations here. One is due to the quantum nature of the wavefunction — the diagonal elements inside the sum, i.e., $|\langle nj|r\rangle_l|^2$, being the probability of observing the lth system in the eigenstate $|nj\rangle$ — while the other is due to the statistical aspect of the ensemble. The expectation value of a physical observable is therefore determined by a double averaging

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{N}} \sum_{l=1}^{\mathcal{N}} \,_l\langle r|\mathcal{O}|r\rangle_l = \sum_{n,j} \langle nj|\hat{\rho}\mathcal{O}|nj\rangle \tag{4.2}$$

performed by introducing complete sets of states and using (4.1). This can be formally written

$$\langle \mathcal{O} \rangle = \frac{\text{Tr}[\hat{\rho}\mathcal{O}]}{\text{Tr}\hat{\rho}} \tag{4.3}$$

where the denominator is a normalization.

It can be shown[1] that for a system to be in equilibrium $\hat{\rho}$ must commute with the Hamiltonian H (and N), i.e., (2.15) (and (2.16)). This means that $\hat{\rho}$ must be diagonal in the orthonormal basis of states $|nj\rangle$ since they are eigenfunctions of the Hamiltonian. For a system in *thermal* equilibrium the chance that a system picked from the ensemble will have energy E_j and contain n particles is given by the Boltzmann factor $e^{-\beta(E_j - \mu n)}$ where $\beta \equiv 1/T$ and μ is the chemical potential. The diagonal elements of $\hat{\rho}$ are thus

$$\langle nj|\hat{\rho}|nj\rangle = \frac{1}{\mathcal{Z}} e^{-\beta(E_j - \mu n)} \tag{4.4}$$

so that (assuming H and N commute)

$$\hat{\rho} = \frac{e^{-\beta(H - \mu N)}}{\text{Tr}[e^{-\beta(H - \mu N)}]} \tag{4.5}$$

where in the large volume limit H and N are given by (2.15) and (2.16) respectively. The normalization factor $\mathcal{Z} \equiv \text{Tr}[e^{-\beta(H - \mu N)}]$, being the sum over E_j and n of the Boltzmann factor, is the grand partition function for the system. The thermal average of an operator, from (4.3), is thus

$$\langle \mathcal{O} \rangle = \frac{\text{Tr}[e^{-\beta(H - \mu N)} \mathcal{O}]}{\text{Tr}[e^{-\beta(H - \mu N)}]}. \tag{4.6}$$

Writing out the grand partition function in terms of the eigenvalues gives[2]

$$\begin{aligned}
\mathcal{Z} &= \sum_{n_1, n_2, \cdots} \exp \beta[\mu(n_1 + \cdots) - (n_1 E_1 + \cdots)] \\
&= \prod_i \sum_{n_i} e^{\beta(\mu - E_i) n_i}.
\end{aligned} \tag{4.7}$$

For bosonic particles n_i is unrestricted ($n_i = 0, 1, \ldots$) whereas for fermions $n_i = 0, 1$ only due to the exclusion principle. Summing the series in either case gives

$$\mathcal{Z} = \prod_i [1 \pm e^{\beta(\mu - E_i)}]^{\pm 1} \tag{4.8}$$

where the upper sign is for fermions and the lower sign for bosons. The mean number

of particles is thus

$$\langle N \rangle = -T \frac{\partial \ln \mathcal{Z}}{\partial \mu} = \sum_i \frac{1}{e^{\beta(E_i - \mu)} \pm 1} \tag{4.9}$$

In the infinite volume limit this sum may be replaced by an integral over momentum modes $V d^3 k / (2\pi)^3$ (the density of states) in the standard way. The equilibrium number density n is thus

$$n = \frac{\langle N \rangle}{V} = \frac{g}{(2\pi)^3} \int d^3 k \frac{1}{e^{\beta(\mu - E)} \pm 1} \equiv \frac{g}{(2\pi)^3} \int d^3 k \, f(E). \tag{4.10}$$

where g is the number of spin degrees of freedom for the particles. $f(E)$ is known as the equilibrium phase space distribution function and is the fundamental object in kinetic theory. The energy-momentum tensor for the system may defined in terms of f as[3]

$$T^{\mu\nu} = \int f \frac{k^\mu k^\nu}{k_0} \frac{d^3 k}{(2\pi)^3}. \tag{4.11}$$

In the comoving frame, comparison of (4.11) with (3.11) gives the energy density ρ and pressure p. Changing variables in the integrals ($E^2 = k^2 + m^2$) yields[4]

$$n = \frac{g}{2\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{1/2} E \, dE}{e^{(E-\mu)/T} \pm 1} \tag{4.12}$$

$$\rho = \frac{g}{2\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{1/2} E^2 \, dE}{e^{(E-\mu)/T} \pm 1} \tag{4.13}$$

$$p = \frac{g}{6\pi^2} \int_m^\infty \frac{(E^2 - m^2)^{3/2} \, dE}{e^{(E-\mu)/T} \pm 1} \tag{4.14}$$

whilst the entropy density is[4]

$$s = \frac{g}{2\pi^2} \int_m^\infty \left(\frac{(E - \mu)/T}{e^{(E-\mu)/T} \pm 1} \pm \ln\left(1 \pm e^{(E-\mu)/T}\right) \right) (E^2 - m^2)^{1/2} E \, dE, \tag{4.15}$$

consistent with the thermodynamic relation $s = (\rho + p - \mu n)/T$. The average energy

of the particles a distribution at temperature T is just

$$E_{\mathrm{av}} = \frac{\rho}{n}. \tag{4.16}$$

In the relativistic limit $(T \gg m)$ the integrals in these thermodynamic quantities (4.12)–(4.15) reduce to (with $T \gg \mu$)[5]

$$n = \frac{g}{2\pi^2} \int_0^\infty \frac{E^2\, dE}{e^{E/T} \pm 1}, \tag{4.17}$$

$$\rho = 3p = \frac{g}{2\pi^2} \int_0^\infty \frac{E^3\, dE}{e^{E/T} \pm 1} \tag{4.18}$$

and

$$s = \frac{g}{2\pi^2} \int_0^\infty \left(\frac{E/T}{e^{E/T} \pm 1} \pm \ln\left(1 \pm e^{E/T}\right) \right) E^2\, dE \tag{4.19}$$

which may be evaluated from a table of standard integrals[6]. So, for bosons (b) and fermions (f) it is found that

$$n_b = \frac{\zeta(3)}{\pi^2} T^3, \quad n_f = \frac{3}{4}\frac{\zeta(3)}{\pi^2} T^3 \tag{4.20}$$

where Riemann $\zeta(3) \approx 1.202$,

$$\rho_b = \frac{g\pi^2}{30} T^4 = \frac{1}{2} g a_{SB} T^4, \quad \rho_f = \frac{7}{16} g a_{SB} T^4 \tag{4.21}$$

where $a_{SB} = \pi^2/15$ is the Stefan-Boltzmann constant, and for both bosons and fermions

$$p = \rho/3 \tag{4.22}$$

and

$$s = \frac{2}{3} g a_{SB} T^3. \tag{4.23}$$

For a mixture of relativistic bosons and fermions in equilibrium with each other,

the total effective number of spin degrees of freedom is[7]

$$g_* \equiv g_b + \frac{7}{8}g_f \tag{4.24}$$

and the total radiation energy density is therefore

$$\rho_r = \frac{1}{2}g_* a_{SB} T^4. \tag{4.25}$$

The value of g_* will change as the Universe cools and particular species go out of equilibrium. The standard model values of g_* at different temperatures are given in table 4.1. Note that the integrals in (4.17)–(4.19) can account for the effect of a finite horizon scale by introducing a minimum energy cutoff corresponding to horizon-sized wavelengths. This would result in a smaller blackbody constant, an effect which becomes appreciable as the Planck scale is approached. It is possible to allow for this circumstance by reducing the effective value of g_* that is used. In the numerical work in chapters 6 and 7, however, the general expressions (4.12)–(4.15) are used and as the lower limit in these integrals is similar to this cutoff, the difference with the above procedure will be negligible.

Comparing (4.25) with (3.21) for radiation dominance implies that

$$T \sim \frac{1}{R} \quad \text{and hence} \quad \frac{\dot{T}}{T} = \frac{\dot{R}}{R}. \tag{4.26}$$

Substituting this and (4.25) into the Friedmann equation (3.14) and integrating gives (for early enough times)

$$T = \left(\frac{3}{16\pi G g_* a_{SB}}\right)^{1/4} t^{-1/2} \tag{4.27}$$

where the initial value used is $T \to \infty$ as $t \to 0$.

$T \gtrsim$	Particles in Equilibrium	g_*
1 MeV	γ, e^+, e^-, ν_e, ν_μ, ν_τ	10.75
100 MeV	$+\mu^+$, μ^-	14.25
140 MeV	$+\pi^\pm$, π^0	17.25
Λ_C	γ, 3ν s, $e^\pm$, $\mu^\pm$, u, $\bar{u}$, d, $\bar{d}$, $8g$ s	51.25
150 MeV	$+s$, $\bar{s}$	61.75
1.2 GeV	$+c$, $\bar{c}$	72.25
1.7 GeV	$+\tau^\pm$	75.75
4.7 GeV	$+b$, $\bar{b}$	86.25
90 GeV	$+W^\pm$, Z^0	96.25
130 GeV	$+t$, $\bar{t}$	105.75

Table 4.1: The effective number of spin degrees of freedom g_* (given by (4.24)), for the standard particle physics model (see table 2.2), for different temperatures. The Higgs particle (an additional 1 degree of freedom) is not included since its mass is unknown. At temperatures above 100 to 300 MeV, the QCD confinement scale Λ_C (i.e., before the assumed quark-hadron phase transition), the hadronic spin degrees of freedom are replaced by those of free quarks and gluons.

In the non-relativistic limit $(m \gg T)$ again the integrals (4.12)–(4.15) are easily evaluated giving[5]

$$n = g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu)/T}, \tag{4.28}$$

$$\rho = mn, \tag{4.29}$$

$$p = nT, \tag{4.30}$$

$$s = g \left(\frac{m^5 T}{(2\pi)^3}\right)^{1/2} e^{-m/T}, \tag{4.31}$$

these expressions being common to both bosons and fermions. There are a number of cases when the chemical potential μ is non-negligible, for example degenerate fermions

T	N_h $g_*^{-1/2} \times$	$\sqrt{N_h}$ $g_*^{-1/4} \times$	$\sqrt{N_h}/N_h$ $g_*^{1/4} \times$
10^{17} GeV	6×10^4	2.5×10^2	4×10^{-3}
10^{16} GeV	6×10^7	8×10^3	1.25×10^{-4}
10^{15} GeV	6×10^{10}	2.5×10^5	4×10^{-6}
10^{14} GeV	6×10^{13}	8×10^6	1.25×10^{-7}

Table 4.2: The total number of particles of a scalar species inside the horizon N_h, i.e., (4.32), at various energy scales. Statistical fluctuations about this value are expected to be of the order $\sqrt{N_h}$ and thus of relative size $\sqrt{N_h}/N_h$.

or bose condensates[5]. However, in the rest of this work only systems with vanishing chemical potential will be considered.

4.2 Thermal Equilibrium?

Due to the fact that the universe is a system which evolves with time, thermal equilibrium can never hold exactly. Indeed, it can be proved that it is not possible to define an equilibrium phase space distribution function for the Robertson-Walker metric at all[3]. As was seen in the last chapter, however, there are indications that for lengthy periods in the early universe, thermal equilibrium was a very good approximation.

First of all the accuracy of the statistical mechanical approach used in the previous section is briefly considered. The total number of particles inside the horizon N_h, of some scalar species at a given temperature close to the Planck scale, may be estimated by using (4.20), (4.27) and $d_H \approx t$, i.e.,

$$N_h \sim n d_H^3 = \sqrt{\frac{45}{16\pi^3} \frac{\zeta(3)}{\pi^2}} \left(\frac{M_{Pl}}{T}\right)^3 \approx 3.7 \times 10^{-2} g_*^{-1/2} \left(\frac{M_{Pl}}{T}\right)^3. \qquad (4.32)$$

The magnitude of the statistical fluctuations about this value are expected to be of order $\sqrt{N_h}$[1]. Some specific values are given in table 4.2 and it can be seen that for $g_* \sim 100$ statistical effects are likely to become appreciable above about 10^{17} GeV. However, at the GUT scale (10^{15} GeV) and below, i.e., the scales that will be considered in chapters 6 and 7, such statistical variations are expected to be small.

How can it be decided whether a given particle species will be in thermal equilibrium? The criterion usually adopted[7,8,5,3] is $\Gamma \gtrsim H$ where Γ, to be defined more fully in the following section, is the reaction rate for interactions involving the species in question. The simplest way of understanding this relation is to note that it requires that the reaction rate must be large compared the approximate age of the universe, H^{-1}. In other words it says that thermal equilibrium can be maintained only if particle interactions occur rapidly enough for the thermal distribution to be able to adjust itself to the changing conditions resulting from the expansion.

For $\Gamma \gg H$ then, a gas of particles should evolve in accordance with the thermodynamic quantities in section (4.1). For $\Gamma \ll H$ (i.e., when they have decoupled) it is also usually possible to predict their evolution so long as they start from equilibrium at the time they decouple t_d. Thus for massless particles, the starting point may be an equilibrium distribution at decoupling of the form $f(E, t_d) = (\exp(E/T_d) \pm 1)^{-1}$. Subsequently the number density will decrease with the expansion $n \propto R^{-3}$, and the wavelength and hence energy of each particle will be redshifted, $E = 1/\lambda = E'R_d/R$. So, since $dn = f(E)\,d^3k$, (see (4.10)), at some later time t, $f(E, t) = f(ER/R_d, t_d) = (\exp(ER/R_dT_d) \pm 1)^{-1}$. In a similar way, for massive particles which decouple when non-relativistic, the momentum $|\mathbf{k}|$ redshifts as R^{-1} which implies the kinetic energy redshifts as R^{-2} and hence $f(E, t) \propto \exp(-E_kR^2/R_d^2T_d)$. These distributions thus imply that[3]

$$T \propto R^{-1}, \quad \text{massless}$$
$$T \propto R^{-2}, \quad \text{massive.}$$
$$\tag{4.33}$$

Note that these arguments do not apply if particles decouple when semi-relativistic.

Another way of looking at the problem is to consider what happens if $\Gamma < H$[5]. There are two problems to consider. (i) Will thermal equilibrium be established? As

a typical example consider a GUT, mediated by massless gauge bosons (i.e., massless in the sense that $T \gg m$). It is found $\Gamma \propto T \propto t^{-1/2}$ and hence the total number of interactions per particle by time t is just

$$N_{pre} = \int_0^t \Gamma(t')\, dt' \approx \left. \frac{\Gamma}{H} \right|_t \qquad (4.34)$$

by using (3.22). So, a typical particle will interact less the once during the period before the time Γ becomes equal to H, i.e., such interactions will not occur quickly enough to generate or maintain thermal equilibrium. Of course, T and hence H and Γ cannot really be defined like this without assuming a thermal state to begin with, so this is really a *reductio ad absurdum* argument. (ii) Will a species already in equilibrium go out of thermal equilibrium? For example, massless particles which are in thermal equilibrium but enter a period when the universe inflates, with $H = H_\Lambda$ a constant, and $T \propto e^{-H_\Lambda t}$. Then the total number of interaction per particle after a time t is

$$N_{post} = \int_t^\infty \Gamma(t')\, dt' \approx \left. \frac{\Gamma}{H} \right|_t . \qquad (4.35)$$

So again a particle will interact less than once subsequent to the time when $\Gamma \approx H$. Though of course, the particles will retain their thermal distribution of velocities, evolving as (4.33).

The condition $\Gamma \gtrless H$ is perfectly adequate for either of the problems considered above separately. But, if both are to be taken into account simultaneously then it may be that Γ is greater than H for too short a period for the particles to achieve a thermal state. In this case the criterion for establishing a thermal state would seem to be a sufficiently large total number of interactions per particle. This can be estimated by integrating Γ over the appropriate time period, i.e.,

$$N \equiv \int_{\Gamma > H} \Gamma\, dt. \qquad (4.36)$$

Note again that this is a self-consistency condition. Thermal equilibrium is assumed to calculate N, in order to see if N is large enough to support a thermal state. Clearly

N must be significantly greater than 1 for a thermal state to form — in some sense it measures how likely thermal state generation is; the larger is N the more likely it is that a thermal distribution will be achieved.

Of course, it must be noted that the correct way of evolving the distribution function f is through the Boltzmann equation. It is in fact possible to justify the $\Gamma > H$ condition self-consistently by carefully considering a "perturbation" of the equilibrium distribution resulting from the fact that such a distribution cannot hold exactly in a Friedmann universe[3]. The Boltzmann equation is particularly useful for phenomena which arise during the decoupling era itself, for example to calculate the abundances of stable relic particles. But obviously, in order to be able to use the Boltzmann equation an initial thermal distribution is required. This is why it cannot be used here — the distribution before equilibrium has been established is not known, and so more approximate means must be resorted to.

Equation (4.36) is the basis of the method used in chapters 6 and 7 for examining questions to do with thermal equilibrium in specific inflationary models.

4.3   Reaction Rate

In order to be able to evaluate the integral (4.36) an expression for the reaction rate Γ is required. It is defined by[9, 5]

$$\Gamma = n\langle\sigma|v|\rangle, \tag{4.37}$$

where $\langle\sigma|v|\rangle$ is the thermally averaged interaction cross-section multiplied by the velocity. The question of interest is whether the scalar particles responsible for inflation, i.e., the ϕ-particles of section 3.7, will form themselves into a thermal state. The reaction rate for the elastic scattering of ϕ-particles is therefore required. For two identical scalar bosons (of four-momenta p_1 and p_2) interacting to produce a final state containing two identical scalars (of four-momenta p_3 and p_4), $\langle\sigma|v|\rangle$ is given by

$$\langle \sigma |v| \rangle = \frac{1}{4n^2} \int \frac{d^3 p_1}{(2\pi)^3 2E_1} \frac{d^3 p_2}{(2\pi)^3 2E_2} \frac{1}{(e^{E_1/T} - 1)} \frac{1}{(e^{E_2/T} - 1)}$$

$$\times \left[\frac{d^3 p_3}{(2\pi)^3 2E_3} \frac{d^3 p_1}{(2\pi)^3 2E_1} (2\pi)^4 \delta^4 (p_3 + p_4 - p_1 - p_2) |\overline{\mathcal{M}}|^2 \right] \qquad (4.38)$$

where $|\overline{\mathcal{M}}|^2$ is the squared averaged matrix element for the process which is calculated from the Feynman rules as in section 2.1. The factor $\frac{1}{4}$ is the symmetry factor associated with the phase space reduction from the identical particles in the initial and final states.

The density factors $(e^{E/T} - 1)^{-1}$ apply in the comoving frame and hence to perform the integration over the initial momenta we need an expression for the quantity in square brackets in this frame. Fortunately, this quantity is Lorentz-invariant and depends only on the centre-of-momentum energy because it is just the total cross section (from (2.18)) without the flux-factor, i.e., it is $\mathcal{F}(E)\sigma(E)$, where $\sigma(E)$ is the total cross-section as a function of centre-of-momentum energy per particle, E, and $\mathcal{F}(E)$ is the flux of ingoing particles. Since $\mathcal{F} = 2E_1 2E_2 |\mathbf{v}_1 - \mathbf{v}_2|$ (as each four-momentum p_i has components $(E_i, \mathbf{p}_i)$ in the co-moving frame and $\mathbf{p}_i = E_i \mathbf{v}_i$), then

$$\mathcal{F}(E) = 8E\sqrt{E^2 - m^2} \qquad (4.39)$$

in the centre-of-momentum frame. The quantity $\mathcal{F}\sigma$ is therefore calculated in the centre-of-momentum frame and transformed back to the co-moving frame by rewriting E in terms of the initial particles' momenta in the co-moving frame; i.e.,

$$E^2 = \frac{1}{2}(m^2 + E_1 E_2 - |\mathbf{p}_1||\mathbf{p}_2| \cos \psi) \qquad (4.40)$$

where ψ is the angle between $\mathbf{p}_1$ and $\mathbf{p}_2$.

Once the scattering amplitude for a specific model is known $\mathcal{F}\sigma$ may be calculated and the reaction rate is then given by

$$\Gamma = \frac{1}{8.(2\pi)^4 n} \int\limits_{m}^{\infty} dE_1 \int\limits_{m}^{\infty} dE_2 \int\limits_{0}^{\pi} d\psi \frac{|\mathbf{p}_1||\mathbf{p}_2| \sin \psi}{(e^{E_1/T} - 1)(e^{E_2/T} - 1)} \mathcal{F}(E)\sigma(E) \qquad (4.41)$$

where E is defined by (4.40).

In order to go further with the programme of calculating (4.36), specific details of the interactions, in order to calculate Γ, and of the vacuum energy, in order to calculate H are required. In models of inflation based on a thermal phase transition, both of these require the temperature-dependent effective field theory which is the subject of the next chapter.

References

1. R.K. Pathria *Statistical Mechanics*, (Pergamon Press, Exeter, 1972)

2. A.L. Fetter and J.D. Walecka *Quantum Theory of Many Particle Systems*, (McGraw-Hill, New York, 1971)

3. J. Bernstein, *Kinetic Theory in the Expanding Universe*, (Cambridge University Press, Cambridge, 1988)

4. J. Ehlers, in *General Relativity and Cosmology*, International School of Physics "Enrico Fermi," edited by R.K. Sachs, Course 47 (Academic Press, New York, 1971)

5. E.W. Kolb and M.S. Turner, *The Early Universe*, (Addison-Wesley, Reading, MA, 1990)

6. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, (Academic, New York, 1980)

7. P.D.B. Collins, A.D. Martin and E.J. Squires, *Particle Physics and Cosmology*, (Wiley, New York, 1989)

8. R.V. Wagoner, in *Physical Cosmology*, eds. J. Audouze, R. Balian, and D.N. Schramm, (North-Holland, Amsterdam, 1980)

9. J. Bernstein, L.S. Brown, G. Feinberg, *Phys. Rev. D* **32**, 3261 (1985)

5

Finite Temperature Field Theory

It was seen in section 3.7 that inflation could be the result of a Higgs-like symmetry breaking potential appearing in some Quantum Field Theory. Usually Quantum Field Theory is used to describe the behaviour of particles in a vacuum, for example scattering events at colliders. However, this is likely to be a bad approximation to the high density conditions which are expected at early times in the big bang model. It is far more likely that thermal equilibrium held for much of the time, as was seen in section 3.3. To describe this situation it is therefore necessary to reformulate Quantum Field Theory, replacing the zero-particle vacuum by a thermal bath, whereupon it becomes known as "finite temperature field theory".

In the context of the early universe, one of the more transparent ways of studying finite temperature field theory is via the effective potential obtained from the effective action in the path integral formulation of field theory. This is briefly reviewed in the following section which ends with the usual perturbative loop expansion of the effective potential. The next section shows how the temperature can be incorporated into this formalism and in particular, how to calculate the finite temperature corrections to the effective potential. The resulting finite-temperature effective potential is of fundamental importance to the study of the phase changes which might occur in the early universe. Explicit calculations of this potential for two specific examples that will be used in chapters 6 and 7 are given in section 5.3. The chapter concludes with a survey of general properties of the temperature-dependent effective potential.

5.1 The Effective Action

Since the inflationary mechanism is almost exclusively supposed to arise from the dynamics of a scalar field, the effective potential formalism will be developed for the simple case of a single real scalar field ϕ as in section 2.1. Where appropriate, this will be specialised further to the case where the potential takes the standard Higgs form (see fig. 2.3), i.e.,

$$V(\phi) = -\frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4!}\phi^4, \tag{5.1}$$

to demonstrate explicitly how calculations are to be performed. More involved examples are discussed in section 5.3.

Before describing the effective action it is necessary to discuss Legendre transforms which are an important component of this formalism. The Legendre transform of a function $f(x)$ is defined as[1]

$$g(\eta) = f(x'(\eta)) - x'(\eta)\eta \tag{5.2}$$

where f is evaluated at the point x' defined by

$$\left.\frac{df}{dx}\right|_{x=x'} = \eta. \tag{5.3}$$

For brevity this is written as

$$g(\eta) = f(x) - x\eta \quad \text{and} \quad \frac{df}{dx} = \eta. \tag{5.4}$$

This definition can be understood in the following terms: to obtain the value of the Legendre transform at the point η, construct the tangent to the curve f which has slope η. The intercept of this tangent with the f axis then gives the value of g. In other words, the equation of the tangent, for fixed η, is $f = \eta x + g$. Rearranging gives (5.4).
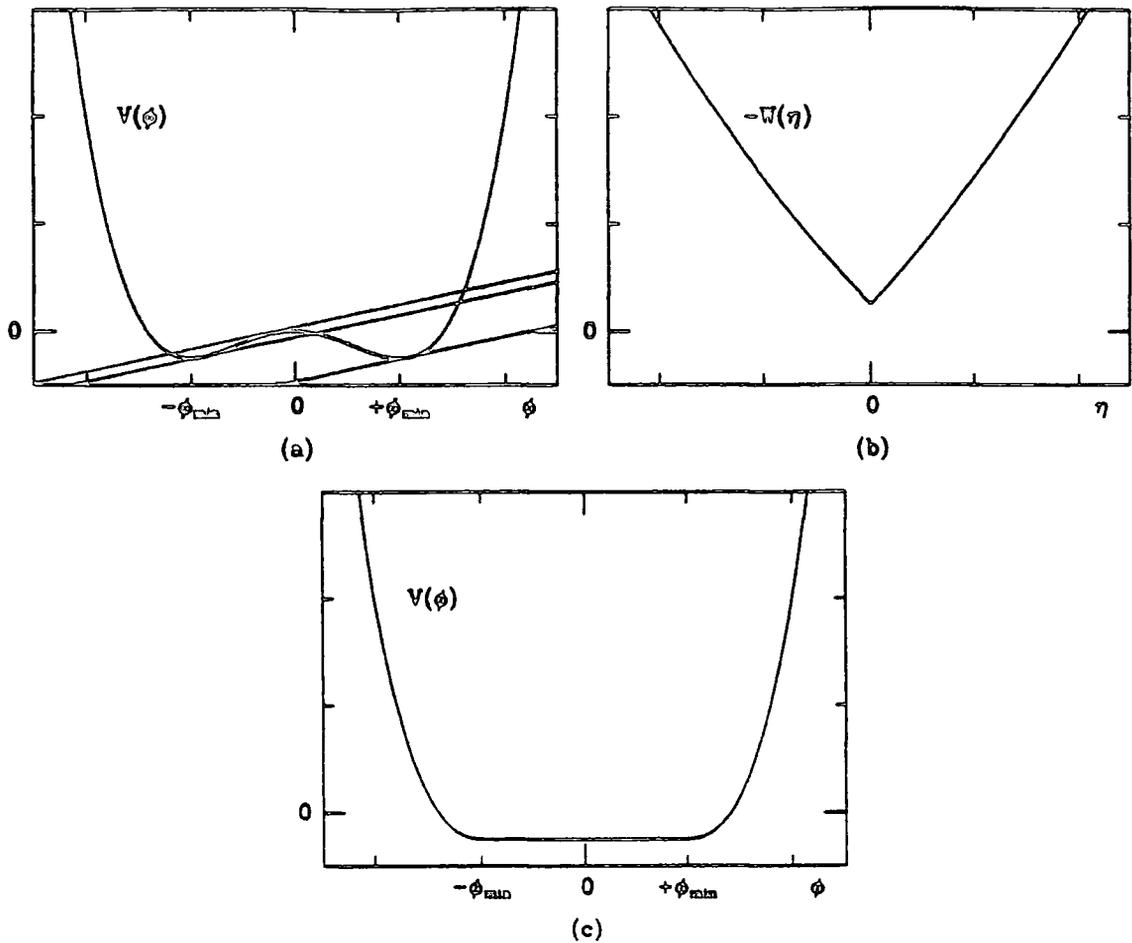
Figure 5.1: (a) The potential $V(\phi)$, (5.1) illustrating the difficulty of defining the Legendre transform of a non-convex function. (b) The Legendre transform W of V given by (5.4). (c) The inverse Legendre transform $\tilde{V}$ of W.

The inverse transformation of (5.4) is given by

$$\tilde{f}(x) = g(\eta) + \eta x \quad \text{and} \quad \frac{dg}{d\eta} = -x. \tag{5.5}$$

In ordinary circumstances, it would be hoped that f and $\tilde{f}$ are identical. This is indeed the case if f is everywhere convex. (For example, it can readily be checked for $f(x) = x^2$ that $g(\eta) = -\eta^2/4$ which implies $\tilde{f} = x^2$.) Problems arise however if the function is not entirely convex. As an example, consider taking the Legendre transform $W(\eta)$ of the potential $V(\phi)$ in (5.1). For small values of η, (5.4) does not define W uniquely[2] because there are three tangents to V with slope η (see

fig. 5.1(a)). This problem is rectified by selecting the minimum intercept of those defined by (5.4). With this definition it is found that the Legendre transform of (5.1) has the form shown in fig. 5.1(b). It can be seen that W is a convex function. This is in fact a general feature of the Legendre transform. Note the discontinuity in the first derivative at $\eta = 0$. Because of this, calculating $\tilde{V}$ from (5.5) gives, $\tilde{V}(\phi) = V(\phi)$ for $|\phi| > \phi_{\min} \equiv \sqrt{6\mu^2/\lambda}$ whilst $\tilde{V}(\phi) = V(\phi_{\min})$ for $|\phi| < \phi_{\min}$ (see fig. 5.1(c)). The figure shows that $\tilde{V}$ is in fact the convex envelope of V. This too is a general property of the Legendre transform.

After this preliminary, the effective action formalism can now be considered. The generating functional of the one particle irreducible (1PI) graphs is

$$\Gamma_{1PI}[\phi_c] = \sum_n \frac{1}{n!} \int d^4x_1 \cdots d^4x_n \, \phi_c(x_1) \cdots \phi_c(x_n) \Gamma^{(n)}(x_1, \ldots, x_n) \qquad (5.6)$$

where, for the moment, ϕ_c is just some arbitrary scalar source function. The quantities $\Gamma^{(n)}(x_1, \ldots, x_n)$ are the one particle irreducible n-point Green's functions in coordinate space, and are related to their counterparts in momentum space by

$$\Gamma^{(n)}(x_1, \ldots, x_n) = \int \frac{d^4k_1}{(2\pi)^4} \cdots \frac{d^4k_n}{(2\pi)^4} (2\pi)^4 \delta^4(k_1 + \cdots + k_n)$$

$$\times e^{i(k_1 \cdot x_1 + \cdots + k_n \cdot x_n)} \Gamma^{(n)}(k_1, \ldots, k_n) \qquad (5.7)$$

the $\Gamma^{(n)}(k_1, \ldots, k_n)$ being the n-point Green's functions of the one particle irreducible Feynman graphs mentioned in section 2.1.

Γ_{1PI} can be related to $W[J]$ defined in (2.26), the generating functional for general connected graphs, where $J(x)$ is an arbitrary source. The relation is given, to lowest order in $\hbar$, by the following theorem

$$\exp\left(\frac{i}{\hbar} W[J]\right) = N \int [d\phi] \exp\left[\frac{i}{\hbar} \left(\Gamma_{1PI}[\phi_c] + \int d^4x J(x)\phi_c(x)\right)\right]. \qquad (5.8)$$

which is often employed when deriving the Feynman rules from path integrals[3, 4]. Assuming that $\hbar$ is small, the right hand side of (5.8) may be evaluated using the

method of stationary phases to give[3]

$$W[J] = \Gamma_{1PI}[\phi_c] + \int d^4x \, J(x)\phi_c(x) \quad \text{at} \quad \frac{\delta\Gamma_{1PI}}{\delta\phi_c} = -J \qquad (5.9)$$

So, generalizing (5.5) to functionals, $W[J]$ is by definition the inverse Legendre transform of Γ_{1PI}. What is actually wanted however is a definition of Γ_{1PI} in terms of W. The generalization of the definition (5.4) can be used to invert (5.9), but as for the ordinary Legendre transform of functions above, the resultant functional may not be identical to Γ_{1PI}. This resultant functional is defined as the effective action $\Gamma_{eff}[\phi_c]$[2, 3],

$$\Gamma_{eff}[\phi_c] = W[J] - \int d^4x J(x)\phi_c(x) \quad \text{and} \quad \frac{\delta W}{\delta J} = \phi_c. \qquad (5.10)$$

The quantity $\phi_c(x)$ is known as the classical field and, from the definition of Green's functions, is the vacuum expectation value with the source term present i.e.,

$$\phi_c(x) \equiv \frac{\delta W}{\delta J} \equiv \left[\frac{\langle 0|\phi(x)|0\rangle}{\langle 0|0\rangle} \right]_J. \qquad (5.11)$$

Very often Γ_{eff} and Γ_{1PI} are not distinguished in the literature, though they are not identical. Their difference involves an interesting physical interpretation which will be discussed in a moment. First note that it follows from the definition (5.10) that

$$\frac{\delta\Gamma_{eff}[\phi_c]}{\delta\phi_c(x)} = -J(x) \qquad (5.12)$$

i.e., the same relation that holds for Γ_{1PI} at the stationary phase point in (5.9). As will be seen shortly, it is this relation that makes the effective action formalism so useful. Also, both the effective and 1PI actions may be expanded in powers of momentum instead of ϕ_c as in(5.6), i.e.,

$$\Gamma_{1PI \atop eff}[\phi_c] = \int d^4x \left(-V_{1PI \atop eff}(\phi_c) + \frac{1}{2}(\partial_\mu\phi_c)^2 Z_{1PI \atop eff}(\phi_c) + \cdots \right) \qquad (5.13)$$

where $V_{1PI \atop eff}$ and $Z_{1PI \atop eff}$ are now ordinary functions of ϕ_c. V_{eff} is called the effective potential and V_{1PI} the 1PI potential. By substituting (5.7) into (5.6) and comparing
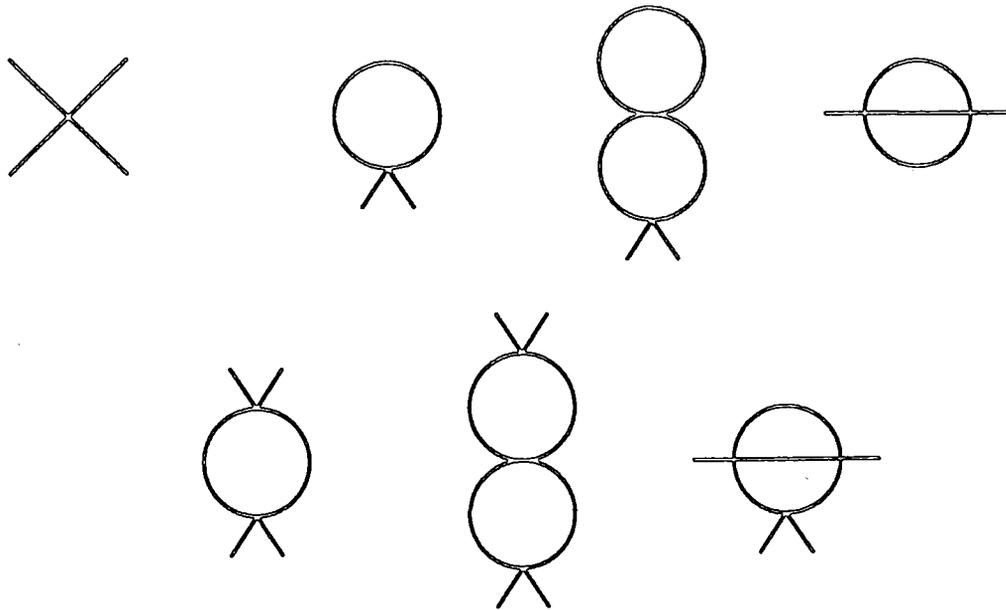
Figure 5.2: Graphs contributing to the 1PI potential (5.14) in $\lambda\phi^4$ theory.

it with (5.13) the 1PI potential may be written as the following expansion

$$V_{1\mathrm{PI}}(\phi_c) = -\sum_n \frac{1}{n!}\Gamma^{(n)}(0,\ldots,0)[\phi_c(x)]^n \qquad (5.14)$$

i.e., it can be calculated as the sum of all possible one particle irreducible diagrams with n external legs carrying zero momentum. Some of these diagrams are exhibited in fig. 5.2 for the potential (5.1).

What is the difference between $V_{1\mathrm{PI}}$ and V_{eff}, V_{eff} being the Legendre transform of the inverse transform of $V_{1\mathrm{PI}}$? Supposing $V_{1\mathrm{PI}}$ had a similar form to the $V(\phi)$ in (5.1) then, as explained above, the Legendre transform of it is convex and so V_{eff} is the convex envelope of $V_{1\mathrm{PI}}$ (see fig. 5.1). It would therefore appear that the effective potential cannot describe the type of symmetry breaking potential which would be needed to generate inflation via a phase change. Furthermore, on the grounds of physical interpretation, it seems likely that $V_{1\mathrm{PI}}$ will be a better approximation to the potential that is actually required to describe an inflationary phase change[2]. The

reason may be understood by again considering a V_{1PI} that is similar to (5.1), and its associated V_{eff} (see fig. 5.1). V_{eff} gives the minimum value of the energy density for a given spatial average of $\phi = \overline{\phi}$[5, 3, 2]. If $|\overline{\phi}| > \phi_{min}$ then the minimum energy is given by a homogeneous state with $\phi = \overline{\phi}$. However, for $|\overline{\phi}| < \phi_{min}$ the minimum energy state consists of many small domains inside each of which $\phi = \pm\phi_{min}$ but for which the spatial average value is arranged to be $\overline{\phi}$. In the large volume limit the surface energy of the domain boundaries may be neglected, and so any such state has minimum energy $V(\phi_{min})$ which is reflected in the effective potential as a straight line at $V_{(\phi_{min})}$ connecting the two original minima of V_{1PI}. If however, *homogeneous* states where $\phi \approx \overline{\phi}$ over the whole region (i.e., the field is localized on the potential) are demanded, then clearly the potential that appears in the Lagrangian of the theory plus its quantum corrections, i.e., V_{1PI}, gives the energy of this state and V_{eff} is no longer relevant[5]. Since in inflation models the ϕ field is practically always considered to be localized (see section 3.7), it is V_{1PI} that must be used.

It can however be shown that V_{1PI} is the analytic continuation of V_{eff} to the region $|\phi| < \phi_{min}$. This is really the sense in which V_{eff} is used to describe inflationary phase transitions[3]. Henceforth V_{1PI} will be renamed the effective potential V_{eff} to agree with the literature. Within this framework it may also be shown that the localization condition does not commute with the Hamiltonian, and hence such states are unstable. Their decay rates can be interpreted as the imaginary contributions to the energy which appear in the formalism[5]. Such a scheme has long been useful in the analysis of statistical mechanics systems[6].

The important property of the effective potential that makes it so useful for descriptions of phenomena related to symmetry breaking is evident from (5.12). When the source J vanishes then the minima of $\Gamma[\phi_c]$ give the vacuum expectation values of ϕ, because then

$$\frac{\delta\Gamma_{eff}[\phi_c]}{\delta\phi_c} = 0. \tag{5.15}$$

Moreover, since the vacuum is expected to be translationally invariant (i.e., momen-

tum conservation is not broken), from (5.13) the simple derivative

$$\frac{dV_{\text{eff}}(\phi_c)}{d\phi_c} = 0 \qquad (5.16)$$

gives all the information required. The crucial point is that V_{eff} may be calculated first, before deciding which point the Lagrangian must be expanded around e.g., in the event that, say, radiative corrections shift the absolute minimum of the potential. Hence, the issue of whether spontaneous symmetry breaking does or does not occur is easily decided from (5.16) depending on whether or not the minimum of V_{eff} is at $\phi_c = 0$[7].

Should it be discovered from the analysis of V_{eff} that symmetry breaking does indeed occur, then a new field ϕ' may conveniently be introduced, defined by

$$\phi' = \phi - \langle\phi\rangle \qquad (5.17)$$

where $\langle\phi\rangle$ is the expectation of ϕ in the asymmetric vacuum. The analysis then proceeds exactly as for the Higgs mechanism (see section 2.3). So, in the example (5.1) $\langle\phi\rangle = \phi_{\min}$ is required.

As might be anticipated, apart from a very few simple cases, V_{eff} cannot not be calculated exactly and so the question of how to expand the effective potential in a perturbative series must be addressed. Since the aim is to explore all possible vacuum states at once, an approximation scheme which reflects this is needed. Following Coleman and Weinberg[7], a parameter a is introduced which multiplies the total Lagrangian of the system,

$$\mathcal{L}(\phi, \partial_\mu\phi, a) \equiv a^{-1}\mathcal{L}(\phi, \partial_\mu\phi). \qquad (5.18)$$

Now consider an arbitrary 1PI graph that would appear in the effective potential (5.13), for example one of those in fig. 5.2, which has I internal lines and V vertices. It will be multiplied by a factor a^{I-V} because each vertex incurs a factor

a^{-1}, but each propagator provides a factor a because it is the inverse of the differential operator in the quadratic terms of $\mathcal{L}$. Furthermore, it is easy to see that the number of loops in a graph is $I - V + 1$, so a diagram with no loops will be multiplied by a^{-1}, with one loop by unity, with two loops by a and so on. Thus an expansion of the effective potential in which the first term is the sum of all zero-loop graphs (i.e., tree graphs giving the original potential), the second term is the sum of all one-loop graphs, and so on, is an expansion in powers of a. This is equivalent to an expansion in powers of $\hbar$[8], but a more important point is that since a multiplies the *total* Lagrangian the approximation is unaffected by field redefinitions or the details of how $\mathcal{L}$ is split into free and interacting parts. The loop expansion is therefore suitable for the purpose of surveying the vacua of $V(\phi)$ before any particular one is chosen.

The renormalization conditions which are required in perturbation theory may all be expressed in terms of the quantities in (5.13). For example, with the potential (5.1), the mass is renormalized to one loop level by setting it equal to the inverse of the propagator at zero momentum $\Gamma^{(2)}(0,0)$ i.e., (2.30), which, on using (5.14), implies that[7]

$$\mu^2 = \left.\frac{d^2 V_{\text{eff}}}{d\phi_c^2}\right|_0 .$$ (5.19)

Similarly, the coupling constant is usually defined as the four-point function at zero external momenta $\Gamma^{(4)}(0,0,0,0)$ i.e., (2.32), so again from (5.14)

$$\lambda = \left.\frac{d^4 V_{\text{eff}}}{d\phi_c^4}\right|_0 .$$ (5.20)

Likewise, though it is not required in calculating the effective potential, the field may be renormalized by setting

$$Z_{\text{eff}}(0) = 1.$$ (5.21)

The effective potential for (5.1) will now be calculated explicitly to one loop order. It proves to be notationally convenient to split up the Lagrangian into free and interacting parts in a slightly different way from normal (as in section 2.1).
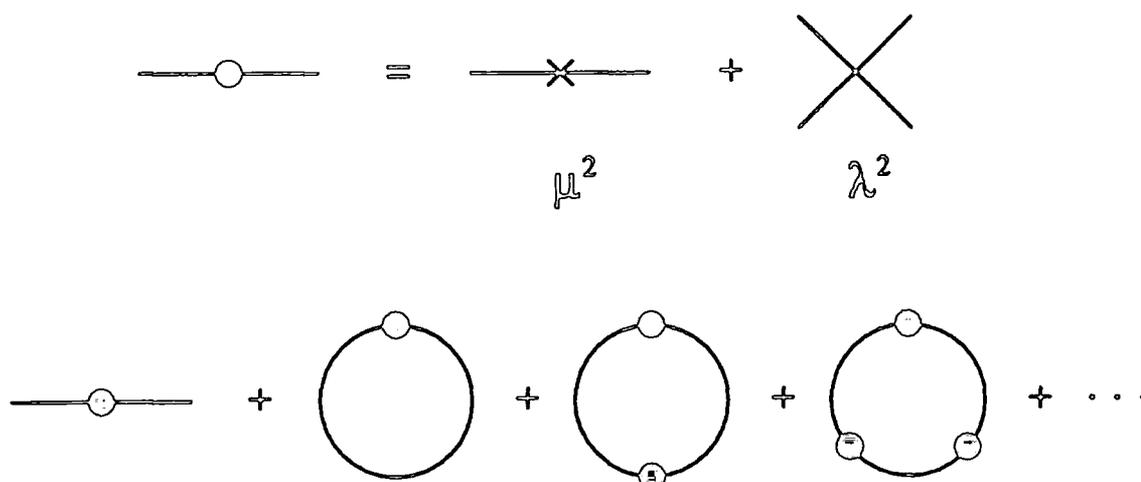
Figure 5.3: Diagrams for calculating the effective potential for (5.1) to one-loop order (5.22).

Instead, the mass term is considered to be part of the interaction Lagrangian so that there are two vertices with couplings μ and λ and the propagator is now massless, i.e., $1/k^2$. The one loop correction is thus the sum of diagrams shown in fig. 5.3. The nth term in the sum is the diagram which has n vertices, each of which provides a factor $\mu^2 + \frac{1}{2}\lambda\phi_c^2$ (the $\frac{1}{2}$ because exchanging the two external lines does not give a new diagram) and n internal lines, each of which provides a $1/k^2$ propagator. Reflections and rotations of the diagram do not give new diagrams either so this term also has a factor $1/2n$. Integrating over the loop momentum thus gives the one-loop effective potential as

$$V_{\text{eff}}(\phi_c) = V(\phi_c) + \frac{1}{2}B\phi_c^2 + \frac{1}{4!}C\phi_c^4 + i \int \frac{d^4k}{(2\pi)^4} \sum_{n=1}^{\infty} \frac{1}{2n} \left(\frac{\mu^2 + \frac{1}{2}\lambda\phi_c^2}{k^2 + i\epsilon} \right)^n, \qquad (5.22)$$

where the terms in B and C are the renormalization counterterms for the mass and coupling constant respectively (see (2.29), a term $\frac{1}{2}A(\partial_\mu\phi)^2$ would be included in the Lagrangian to renormalize the wavefunction). The series inside the integral may be summed yielding

$$V_{\text{eff}}(\phi_c) = V(\phi_c) + \frac{1}{2}B\phi_c^2 + \frac{1}{4!}C\phi_c^4 + \frac{1}{2} \int \frac{d^4k}{(2\pi)^4} \ln\left(1 + \frac{m^2(\phi_c)}{2k^2} \right), \qquad (5.23)$$

in which the integral has been rotated to Euclidean space and

$$m^2(\phi_c) = \left.\frac{\partial^2 V_{\text{eff}}}{\partial \phi^2}\right|_{\phi=\phi_c} = -\mu^2 + \frac{1}{2}\lambda\phi_c^2 \tag{5.24}$$

which is the reason the μ^2 term has been included as part of the interaction. It is straightforward to generalise this derivation to show that (5.23) applies to a scalar field with any polynomial self-interaction[2] (with the appropriate counterterms to match).

An interesting observation[9] can be made at this point. If just the k_0 integration is carried out in the correction term in (5.23) then it is found that

$$V_{\text{eff}}(\phi_c) = V(\phi_c) + \frac{1}{2} \int \frac{d^3k}{(2\pi)^3} \sqrt{\mathbf{k}^2 + m^2(\phi_c)}. \tag{5.25}$$

This adds weight to the claim that this quantity (really being $V_{1\text{PI}}$) gives the quantum corrections to the classical potential for a *localized* state. Fluctuations of the field ϕ about the value ϕ_c will have a frequency given by the curvature (5.24) at that value of ϕ_c, and the energy of such a mode will be $\sqrt{k^2 + m^2(\phi_c)}$. This is converted to an energy density by the integration, and so the correction term is the sum over all zero-point energies of fluctuations of ϕ about the classical value ϕ_c.

In order to fully evaluate the integral in (5.23) renormalization must be carried out (see section 2.1). It is first of all regulated by cutting it off at some large momentum $k^2 = \Lambda^2$, after which it is easily evaluated, giving

$$V_{\text{eff}}(\phi_c) = V(\phi) + \frac{1}{2}B\phi_c^2 + \frac{1}{4!}C\phi_c^4 + \frac{\Lambda^2 m^2(\phi_c)}{32\pi^2} + \frac{m^4(\phi_c)}{64\pi^2}\left(\ln\left(\frac{m^2(\phi_c)}{\Lambda^2}\right) - \frac{1}{2}\right) \tag{5.26}$$

after discarding terms that vanish for large Λ. Applying the conditions (5.19) and (5.20) it is found that

$$B = -\frac{\lambda\Lambda^2}{32\pi^2} - \frac{\lambda\mu^2}{32\pi^2}\ln\left(\frac{\mu^2}{\Lambda^2}\right) \tag{5.27}$$

and

$$C = -\frac{3\lambda^2}{32\pi^2}\left(\ln\left(\frac{\mu^2}{\Lambda^2}\right) + 1\right) \tag{5.28}$$

so that finally

$$V_{\text{eff}}(\phi_c) = \frac{1}{2}\mu^2\phi_c^2 + \frac{\lambda}{4!}\phi_c^4 + \frac{1}{64\pi^2}\left[m^4(\phi_c)\ln\left(\frac{m^2(\phi_c)}{\mu^2}\right) - \frac{3}{8}\lambda^2\phi_c^4 - \frac{\lambda\mu^2}{2}\phi_c^2\right] \quad (5.29)$$

after discarding irrelevant constants. On substituting (5.24) it is found that for small values of ϕ_c the argument of the logarithm term in (5.29) is negative, though it is positive near the minimum of (5.1) at $\phi_c^2 = 6\mu^2/\lambda$. This is the source of the imaginary contributions to the effective potential referred to above.

It should be noted that this method of summing all possible n-loop graphs, although conceptually clear, is extremely cumbersome beyond one-loop order. A more sophisticated method involving tadpole diagrams[10] is available for second and higher order calculations but such accuracy will not be needed here.

5.2 Finite Temperature

What happens if the usual zero particle vacuum is replaced by a thermal bath? In section 4.1 (for vanishing chemical potential μ), it was shown that the thermal average of an operator $\mathcal{O}$ may be written as (4.6)

$$\langle\mathcal{O}\rangle \equiv \frac{\text{Tr}[e^{-\beta H}\mathcal{O}]}{\text{Tr}\,e^{-\beta H}} \quad (5.30)$$

where β is the inverse temperature $1/T$. The trick is to treat the factor $e^{-\beta H}$ as an imaginary time development operator[3]. For example, consider the two point Green's function for particles in a heat bath with a temperature $1/\beta$,

$$\langle\Gamma^{(2)}(x-y)\rangle_\beta = \frac{\text{Tr}\left[e^{-\beta H}T[\phi(x)\phi(y)]\right]}{\text{Tr}\,e^{-\beta H}}. \quad (5.31)$$

If the usual definition of time ordering is extended to cover imaginary times, in a way which is compatible with the usual definition in Minkowski space, then

$$T[\phi(x)\phi(y)] = \begin{cases} \phi(x)\phi(y) & \text{if } ix_0 > iy_0 \\ \phi(y)\phi(x) & \text{if } ix_0 < iy_0. \end{cases} \quad (5.32)$$

Utilising this definition, and the cyclic property of traces, it is found that (5.31)

evaluated at, say, $x_0 = 0$ and $-i\beta \le y_0 \le 0$, is

$$\langle \Gamma^{(2)}(x-y)\rangle_\beta \Big|_{x^0=0} = \text{Tr}[e^{-\beta H}\phi(y_0,\mathbf{y})\phi(0,\mathbf{x})]/\text{Tr}e^{-\beta H}$$

$$= \text{Tr}[\phi(0,\mathbf{x})e^{-\beta H}\phi(y_0,\mathbf{y})]/\text{Tr}e^{-\beta H}$$

$$= \text{Tr}[e^{-\beta H}e^{\beta H}\phi(0,\mathbf{x})e^{-\beta H}\phi(y_0,\mathbf{y})]/\text{Tr}e^{-\beta H} \qquad (5.33)$$

$$= \text{Tr}[e^{-\beta H}\phi(-i\beta,\mathbf{x})\phi(y_0,\mathbf{y})]/\text{Tr}e^{-\beta H}$$

$$= \langle \Gamma^{(2)}(x-y)\rangle_\beta \Big|_{x^0=-i\beta}.$$

Hence it can be seen that operators at finite temperature become periodic in the imaginary time direction with period β. In a similar way it may be shown that fermions are anti-periodic in imaginary time[3], i.e.,

$$\langle \Gamma_f^{(2)}(x-y)\rangle_\beta \Big|_{x^0=0} = - \langle \Gamma_f^{(2)}(x-y)\rangle_\beta \Big|_{x^0=-i\beta}. \qquad (5.34)$$

It is a simple matter to generalise this to n-point Green's functions, and hence the conclusion is that the difference between the zero and finite temperature cases is a difference in the boundary conditions.

The path integral approach is particularly expedient for finite temperature field theory because it leaves the boundary conditions unspecified. The whole formalism of generating functionals and effective potentials derived in section 5.1 is thus directly applicable. The only change is the Euclidean time limits in the action integral[3], viz.,

$$\int d^4 x_E \rightarrow \int_0^\beta d\tau \int d^3 x \qquad (5.35)$$

and since periodic boundary conditions imply discrete values of momenta $k_0 \rightarrow 2\pi T$, so that in momentum space

$$\int d^4 k \rightarrow 2\pi T \sum_{n=-\infty}^{\infty} \int d^3 k. \qquad (5.36)$$

So, for example, the effective potential to one loop at finite temperature can be calculated from the same diagrams as the zero temperature case (fig. 5.3) but with the

Feynman rules modified by (5.36). So the effective potential (5.23) derived from (5.1) becomes[2]

$$V_{\text{eff}}(\phi_c) = V(\phi_c) + \frac{1}{2}T \sum_{n=-\infty}^{\infty} \int \frac{d^3k}{(2\pi)^3} \ln[\mathbf{k}^2 + (2\pi nT)^2 + m^2(\phi_c)] \qquad (5.37)$$

where again $m^2(\phi_c)$ is again given by (5.24) and the counterterms have been omitted temporarily. The sum over n diverges, but it can be re-expressed by using the following equality[11]

$$
\begin{aligned}
\frac{\partial}{\partial E} \left(\sum_n \ln \left((2\pi nT)^2 + E^2 \right) \right) &= \sum_n \frac{2E}{(2\pi nT)^2 + E^2} \\
&= \frac{1}{\pi T} \left(2 \sum_{n=1}^{\infty} \frac{E/2\pi T}{n^2 + (E/2\pi T)^2} + \frac{2\pi T}{E} \right) \qquad (5.38) \\
&= \frac{1}{T} \coth(E/2T) \\
&= \frac{1}{2} + \frac{1}{e^{E/T} - 1}
\end{aligned}
$$

where $E^2 = \mathbf{k}^2 + m^2(\phi_c)$ and the third line follows after summing the series[12]. Integrating (5.38) with respect to E and substituting in (5.37) gives

$$V_{\text{eff}}(\phi_c, T) = V(\phi_c) + \int \frac{d^3k}{(2\pi)^3} \left(\frac{E}{2} + T \ln(1 - e^{-E/T}) \right). \qquad (5.39)$$

The first term in the integrand is simply the zero-temperature one-loop correction obtained in (5.25) while the second term represents the contribution of the heat bath. Writing this second term in spherical polars and then substituting $x = |\mathbf{k}|/T$ gives the finite temperature effective potential as

$$V_{\text{eff}}(\phi_c, T) = V_{\text{eff}}(\phi_c) + \frac{T^4}{2\pi^2} \int_0^{\infty} dx\, x^2 \ln \left\{ 1 - \exp \left[-\sqrt{x^2 + m^2(\phi_c)/T^2} \right] \right\}, \qquad (5.40)$$

i.e., it is just the one-loop result at zero temperature plus a finite-temperature correction. Note that this temperature-dependent correction is *finite* which means that the counter terms (5.27) and (5.28) which make the zero-temperature potential finite also serve to renormalize the temperature-dependent potential.

A similar calculation shows that a gas of fermions which couple to the ϕ-particles and are in thermal equilibrium with them will give a contribution to the effective potential of the form

$$-4\frac{T^4}{2\pi^2}\int\limits_0^\infty dx\, x^2 \ln\left\{1 + \exp\left[-\sqrt{x^2 + m^2(\phi_c)/T^2}\right]\right\} \tag{5.41}$$

where now $m(\phi_c)$ is the fermion mass in the shifted vacuum. The factor 4 counts the fermion/antifermion spin states while the minus sign is due to the Fermi statistics.

The temperature dependent effective potential (5.40) can be considerably simplified at high temperature by approximating the integral. First of all the integral in (5.40) is written as

$$I(a^2) \equiv \int\limits_0^\infty dx\, x^2 \ln[1 - \exp(-(x^2 + a^2)^{1/2})] \tag{5.42}$$

where $a^2 \equiv m^2(\phi_c)/T^2$. High temperature corresponds to small a^2 so a Taylor expansion about $a^2 = 0$ is required. The first term is

$$\begin{aligned} I(a^2)\big|_{a^2=0} &= \int_0^\infty dx\, x^2 \ln(1 - e^{-x}) \\ &= -\frac{\pi^4}{45} \end{aligned} \tag{5.43}$$

and the second term is

$$\begin{aligned} a^2 \left.\frac{\partial I(a^2)}{\partial a^2}\right|_{a^2=0} &= a^2 \int_0^\infty dx\frac{x}{e^x - 1} \\ &= \frac{m^2(\phi_c)}{T^2}\cdot\frac{\pi^2}{12} \end{aligned} \tag{5.44}$$

from a table of standard integrals[12]. So the high temperature expansion of (5.40) begins

$$V_{\text{eff}}(\phi, T) = V_{\text{eff}}(\phi) - \frac{1}{90}\pi^2 T^4 + \frac{1}{24}m^2(\phi)T^2 + \cdots. \tag{5.45}$$

Further terms may be obtained with rather more effort[11].

Substituting (5.24) into (5.45) gives the high temperature expansion of the effective potential resulting from (5.1), i.e.,

$$V_{\text{eff}}(\phi, T) \approx -\frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4!}\phi^4 - \frac{\pi^2}{90}T^4 + \frac{1}{24}\mu^2 T^2 + \frac{\lambda}{48}\phi^2 T^2 \qquad (5.46)$$

where the zero-temperature one-loop terms (in (5.29)) have been ignored because they are small compared to the tree level part. This simplification has been introduced because it will make the further analysis in chapters 6 and 7 more tractable.

For completeness it should be added that it is possible to approach finite temperature field theory in a more physically intuitive way by using Minkowski space propagators in the so called "real time formalism" (the method just described is sometimes known as the "imaginary time formalism"). The boson propagator at finite temperature is then[11]

$$\langle D \rangle_\beta = \frac{i}{k^2 - m^2 + i\epsilon} + \frac{2\pi}{e^{-\beta H} - 1}\delta(k^2 - m^2) \qquad (5.47)$$

i.e., the propagator is the usual one in Minkowski space, but if the particle is emitted on mass shell ($k^2 = m^2$) then it must be indistinguishable from the particles in the heat bath and comply with their statistical distribution. The main advantage conferred by this approach is that the zero-temperature and finite-temperature parts are separate right from the start of the calculation which is often an aid when calculating physical processes[13].

5.3 Further Examples

The temperature-dependent effective potential to one-loop order will now be calculated for two particular models that will be used extensively in the next two chapters. The first is the Coleman-Weinberg potential which has some interesting features in its own right. It is the potential for a scalar field theory which has *no* explicit mass term and couples to gauge bosons. The simplest possible case is a massless complex scalar singlet field coupled to the gauge boson of a $U(1)$ gauge group — "massless
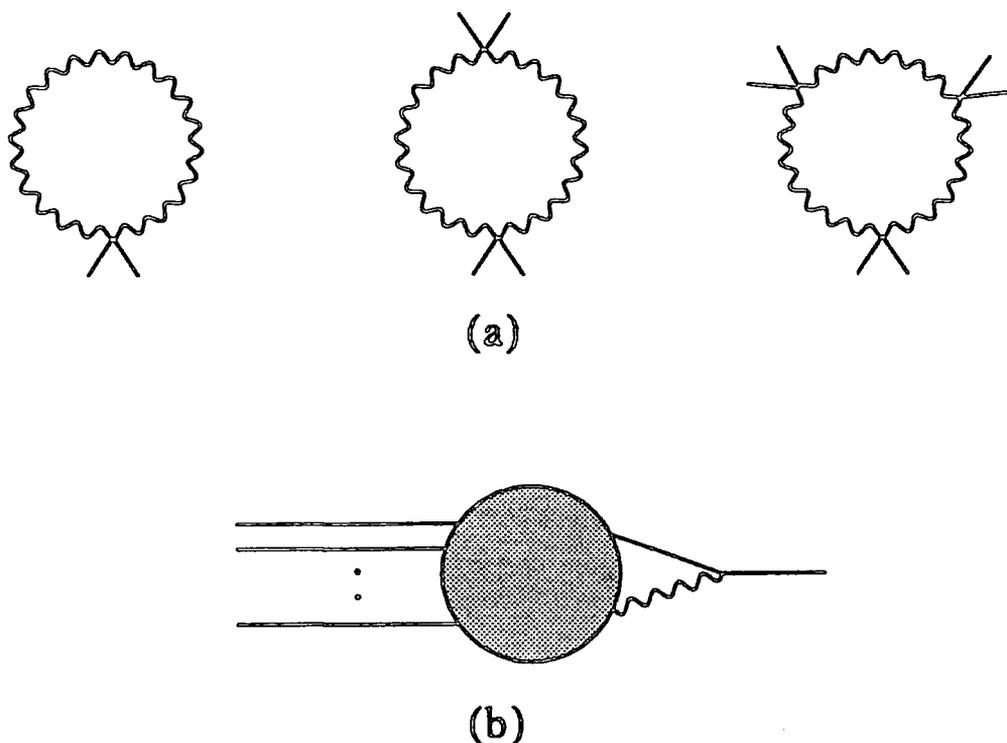
(a)

(b)

Figure 5.4: One-loop diagrams for the Coleman-Weinberg potential. In the Landau gauge (5.49) only the diagrams in (a) contribute to the effective potential, any diagram of the form in (b) vanishes.

scalar QED" — i.e., (2.35) without a term in μ^2. The Lagrangian is thus

$$\mathcal{L} = \frac{1}{2}(\partial_\mu + ieA_\mu)^2\phi^2 - \frac{\lambda}{4!}\phi^4 - \frac{1}{4}F_{\mu\nu}F^{\mu\nu} + \text{counterterms} \qquad (5.48)$$

where e is the coupling to the gauge field A_μ and $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ is the gauge field strength tensor. The scalar potential at tree level is $V(\phi) = \lambda\phi^4/4!$.

First the zero-temperature effective potential is calculated. It might seem that all the diagrams in fig. 5.4 need to be summed along with the pure scalar graphs of fig. 5.3, each scalar line being either a ϕ_1 or ϕ_2 particle. However the calculation can be simplified in two ways. Firstly, because the effective potential can only depend on $\phi_c = \sqrt{\phi_1^2 + \phi_2^2}$, only graphs with one of ϕ_1 or ϕ_2 on the external legs needs to be included. Secondly, working in the Landau gauge, where the gauge particle propagator is

$$D_{\mu\nu} = -i\frac{g_{\mu\nu} - k_\mu k_\nu/k^2}{k^2 + i\epsilon}, \qquad (5.49)$$

i.e., (2.36) with $\zeta = 0$, removes an entire class of diagrams[7]. The reason can be understood by considering the gauge-scalar-scalar vertex. If one scalar is internal and one external as in graphs like those of fig. 5.4(b), then, because the external scalar momentum is considered to vanish, the internal scalar and gauge particles have the same momenta. Contracting k^μ with (5.49) gives zero and so all diagrams with a vertex like fig. 5.4(b) vanish.

To summarise, the one loop correction may be calculated by summing the diagrams in fig. 5.3 with the internal loop containing either ϕ_1 or ϕ_2 plus all the diagrams in fig. 5.4(a) with the gauge particle in the internal loop. For each set of diagrams the sum is the same as (5.23) but with

$$m^2(\phi_c) = \frac{\lambda}{2}\phi_c^2 \tag{5.50}$$

and with a factor 3 to count the degrees of freedom of the gauge particle loops, so that to one loop V_{eff} is

$$V_{\text{eff}}^{(1)} = \frac{\lambda}{4!}\phi_c^4 + \frac{1}{2}B\phi_c^2 + \frac{1}{4!}C\phi_c^4\,5\int\frac{d^4k}{(2\pi)^4}\ln\left(1 + \frac{\lambda\phi_c^2}{2k^2}\right). \tag{5.51}$$

This can be regulated as before by cutting off the integral at $k^2 = \Lambda^2$ giving

$$V_{\text{eff}}^{(1)} = \frac{\lambda\Lambda^2}{64\pi^2}\phi_c^2 + \frac{\lambda^2\phi_c^4}{256\pi^2}\left(\ln\frac{\lambda\phi_c^2}{2\Lambda^2} - \frac{1}{2}\right). \tag{5.52}$$

Then using (5.19) with $\mu^2 = 0$ gives

$$B = -\frac{\lambda\Lambda^2}{32\pi^2}. \tag{5.53}$$

Unfortunately, (5.20) cannot be used to define the coupling because the log term in (5.52) implies that the fourth derivative of $V_{\text{eff}}^{(1)}$ is infinite at $\phi_c = 0$. Instead, an arbitrary mass scale is introduced by moving the renormalization point to $\phi_c = M$,

i.e.,

$$\frac{d^4 V_{\text{eff}}}{d\phi_c^4}\bigg|_{\phi_c = M} = \lambda \tag{5.54}$$

which gives

$$C = -\frac{11}{32}\frac{\lambda^2}{\pi^2} - \frac{3\lambda^2}{32\pi^2}\ln\frac{\lambda M^2}{2\Lambda^2}. \tag{5.55}$$

The effective potential is therefore

$$V_{\text{eff}}(\phi_c) = \frac{\lambda}{4!}\phi_c^4 + \left(\frac{5\lambda^2}{1152\pi^2} + \frac{3e^4}{64\pi^2}\right)\phi_c^4\left(\ln\frac{\phi_c^2}{M^2} - \frac{25}{6}\right). \tag{5.56}$$

If λ is assumed to be of order e^4 then the term in λ^2 may be ignored as negligible compared to the others. By differentiating (5.56) it is easy to see that there is a new absolute minimum of this potential which is not at the origin, and that there is a maximum at $\phi_c = 0$. It therefore appears that the radiative corrections to the effective potential have induced a symmetry breaking.

The renormalization parameter M is completely arbitrary — changing its value merely results in a reparameterisation of V_{eff}. It is equivalent to modifying the renormalization subtraction point (see section 2.1). So, M might as well be set to some convenient value, if only to remove it from further consideration. The natural choice is to set $M = \phi_{\text{min}}$ the new absolute minimum of V[7]. Since $(dV/d\phi)|_{\phi=\phi_{\text{min}}} = 0$ this means that

$$\lambda = \frac{33}{8\pi^2}e^4 \tag{5.57}$$

which implies that

$$V_{\text{eff}}(\phi_c) = \frac{3e^4}{64\pi^2}\phi_c^4\left(\ln\frac{\phi_c^2}{\phi_{\text{min}}^2} - \frac{1}{2}\right). \tag{5.58}$$

So it may be seen that the original specification of the potential in terms of two dimensionless parameters e and λ has been turned into a specification in terms of e and ϕ_{min}, a parameter which has the dimensions of mass. This phenomenon is called "dimensional transmutation"[7].

Usually one-loop corrections make no difference to the position of the vacuum in the effective potential, and so the perturbative expansion should remain trustworthy. But when such a radical change to V_{eff} as (5.58) is made the validity of the loop expansion must be checked. Higher order loop diagrams generate higher powers of $e^4 \ln(\phi_c/\phi_{\text{min}})$ (a factor such as this is known as the loop-expansion parameter) and near the new minimum ϕ_{min} these log terms will be small. (For precisely this reason it can be shown that a pure massless $\lambda\phi^4$ theory cannot acquire symmetry breaking in this way.) Hence, assuming that the coupling $e \ll 1$, the loop expansion should still be valid and so the radiative corrections do indeed induced symmetry breaking. The assumption that λ is of order e^4 can be relaxed by using the renormalization group equation[2, 7].

Next the finite temperature corrections must be included. As was noted previously, no further renormalization is needed at finite temperature — the one-loop T-dependent correction to (5.58) is therefore just five times the integral in (5.40), corresponding to the two scalars and one gauge particle running around the loop, but with $m^2(\phi_c)$ given by (5.50). Substituting (5.50) into (5.45) gives the high temperature expansion of the effective Coleman-Weinberg potential[3], i.e.,

$$V_{\text{eff}}(\phi_c, T) = \frac{3e^4}{64\pi^2}\phi_c^4 \left(\ln \frac{\phi_c^2}{\phi_{\text{min}}^2} - \frac{1}{2} \right) + \frac{11}{128\pi^2}e^4\phi_c^2 T^2 \tag{5.59}$$

where non-ϕ_c-dependent terms have been ignored.

The second example to be calculated in this section is an ordinary scalar potential (like (5.1) but with positive μ^2) which also has a cubic coupling $\delta\mu\phi^3$, i.e.,

$$V(\phi) = \frac{1}{2}\mu^2\phi^2 - \frac{\delta\mu}{3!}\phi^3 + \frac{\lambda}{4!}\phi^4. \tag{5.60}$$

This is the most general renormalizable scalar potential with only self interactions[14]. This case is rather less involved than the Coleman-Weinberg potential because the double-well vacuum structure (for suitable δ, μ and λ) is already inherent in (5.60). It must be noted however, that the cubic term means the usual reflection symmetry of $\mathcal{L}$ about $\phi = 0$ is lost at tree level.
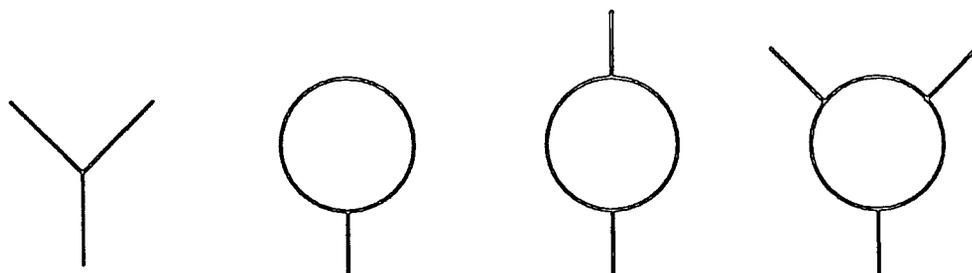
Figure 5.5: One loop diagrams involving the cubic coupling. Each diagram is added to the corresponding term in fig. 5.3.

The zero temperature one-loop sum proceeds as before but with an extra set of diagrams involving the cubic coupling (see fig. 5.5) so that now

$$m^2(\phi_c) = \mu^2 - \delta\mu\phi_c + \frac{\lambda}{2}\phi_c^2 \tag{5.61}$$

instead of (5.24). The loop integral is again regulated in the standard way giving

$$V_{\text{eff}}(\phi_c) = V(\phi_c) + \frac{1}{2}B\phi_c^2 + \frac{1}{4!}C\phi_c^4 - \frac{1}{3!}D\phi_c^3 + E\phi + \frac{\Lambda^2 m^2(\phi_c)}{32\pi^2}$$

$$+ \frac{m^4(\phi_c)}{64\pi^2}\left(\ln\left(\frac{m^2(\phi_c)}{\Lambda^2}\right) - \frac{1}{2}\right) \tag{5.62}$$

where the counterterm D is needed to renormalize the cubic coupling, while E must be included because reflection symmetry can no longer be invoked to guarantee that it will remain zero in higher orders of perturbation theory. Corresponding to these two new counterterms are the two renormalization conditions

$$\left.\frac{d^3 V_{\text{eff}}}{d\phi_c^3}\right|_0 = -\delta\mu \quad \text{and} \quad \left.\frac{dV_{\text{eff}}}{d\phi_c}\right|_0 = 0 \tag{5.63}$$

along with (5.19) and (5.20) which are used to renormalize the mass and quartic

couplings. Applying these renormalization conditions to (5.62) gives

$$B = \frac{1}{32\pi^2}\left[-\lambda\Lambda^2 - \delta^2\mu^2 - (\lambda + \delta^2)\mu^2\ln\frac{\mu^2}{\Lambda^2}\right],$$

$$C = \frac{1}{32\pi^2}\left[-3\lambda^2 + 6\delta^2\lambda - \delta^4 + \lambda^2\ln\frac{\mu^2}{\Lambda^2}\right],$$

$$D = \frac{1}{32\pi^2}\left[-\delta^3\mu - 3\delta\lambda\mu - 3\delta\lambda\mu\ln\frac{\mu^2}{\Lambda^2}\right],$$

$$E = \frac{1}{32\pi^2}\left[\delta\mu\Lambda^2 + \delta\mu^3\ln\frac{\mu^2}{\Lambda^2}\right],$$

$$(5.64)$$

so that the one-loop effective potential is

$$V_{\text{eff}}(\phi_c) = V(\phi_c) + \frac{1}{64\pi^2}\left[\delta\mu^3\phi_c + \frac{1}{2}(\delta^2 - \lambda)\mu^2\phi_c^2 + \delta\mu\left(\frac{3}{2}\lambda + \frac{1}{3}\delta^2\right)\phi_c^3\right.$$

$$\left. + \left(\frac{1}{12}\delta^4 - \frac{1}{2}\delta^2\lambda - \frac{3}{8}\lambda^2\right)\phi_c^4 + m^4(\phi_c)\ln\left(\frac{m^2(\phi_c)}{\mu^2}\right)\right]. \tag{5.65}$$

The temperature dependent correction is just the integral in (5.40) but with $m^2(\phi)$ given by (5.61). The high temperature expansion for this case is

$$V_{\text{eff}}(\phi_c, T) = \frac{1}{2}\mu^2\phi_c^2 - \frac{\delta\mu}{3!}\phi_c^3 + \frac{\lambda}{4!}\phi_c^4 - \frac{\delta\mu}{24}\phi_c T^2 + \frac{\lambda}{48}\phi_c^2 T^2 \tag{5.66}$$

where again the zero-temperature one-loop corrections in (5.65) have been ignored because they are small compared to the tree level terms.

5.4   The Temperature Dependent Effective Potential

Very many features of $V_{\text{eff}}(T, \phi)$ can be understood by examining its high temperature expansion; for example (5.46) which stems from the potential (5.1). The term proportional to T^4 is the pressure due to the gas of particles (see (4.21) and (4.22)). The temperature dependent particle mass is defined as the coefficient of the quadratic
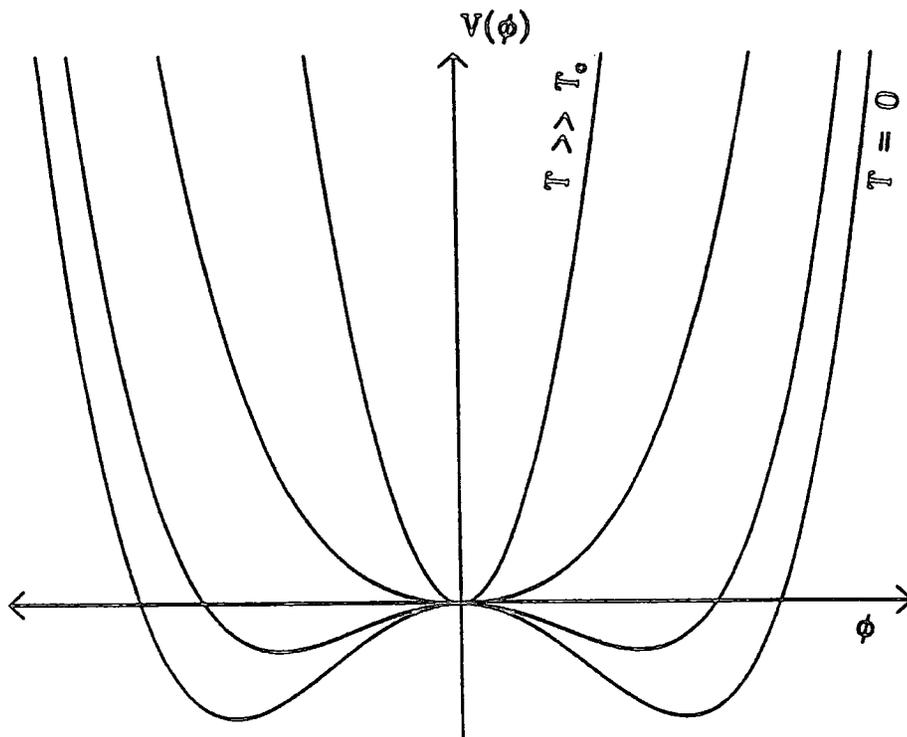
Figure 5.6: The finite-temperature effective potential for (5.1) given in (5.46).

terms in this potential at the minimum i.e., $m(\phi, T) = (\partial^2 V_{\text{eff}}/\partial\phi^2)_{\phi_{\text{min}}}$. When this is evaluated at $\phi = 0$ for (5.46) one obtains

$$m(\phi = 0, T) = -\mu^2 + \frac{\lambda}{24}T^2 \qquad (5.67)$$

which can be made positive by making the temperature sufficiently large. Also, this minimum at $\phi = 0$ will be unique if the temperature is high enough, and hence the symmetry of the Lagrangian will be restored.

This is in fact a generic feature of finite temperature field theory for any symmetry breaking potential[2, 3]. It can be compared with traditional statistical mechanics systems, like for example ferromagnets, which have a rotationally symmetric phase at high temperature which is spontaneously broken if the temperature falls below the Curie point, and spontaneous magnetisation results[15]. Indeed, a very similar mathematical formalism may be used to describe the two cases. It may be seen from fig. 5.6 that this symmetry restoration occurs with the first example of section 5.3,
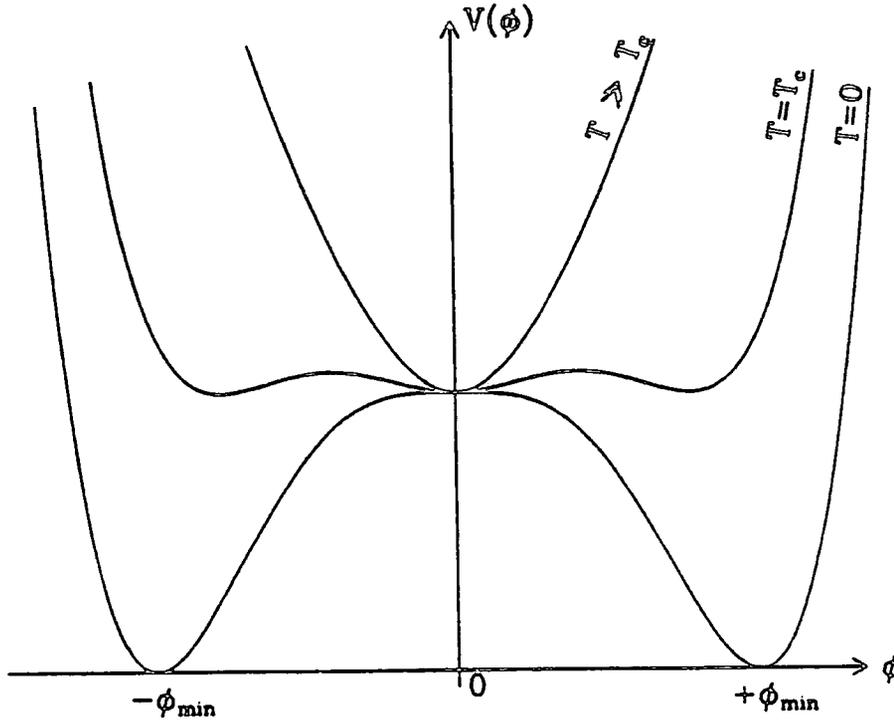
Figure 5.7: The finite-temperature effective potential (5.59) for the Coleman-Weinberg potential in section 5.3.

but not with the second due to the fact that the zero-temperature potential is not symmetric (i.e., is not a symmetry breaking potential).

Thus it seems that finite temperature field theory generates a potential whose evolution (see fig. 5.6) as the universe cools can describe the phase transitions which may have occurred in the early universe. Such a phase transition is characterised by a critical temperature T_c which is defined as the temperature at which the minimum at $\phi = 0$ and the asymmetric minimum become degenerate. On dimensional grounds T_c is expected to be similar in magnitude to the symmetry breaking scale of the potential. It can be estimated from the high temperature expansion by some simple algebra. For example, from (5.46) it found that

$$T_c \approx \sqrt{\frac{24\mu^2}{\lambda}}. \tag{5.68}$$

There are two reasons why this is likely to be only a rather crude estimate; first because the high temperature expansion (5.45)is not adequate at $T = T_c$ and secondly
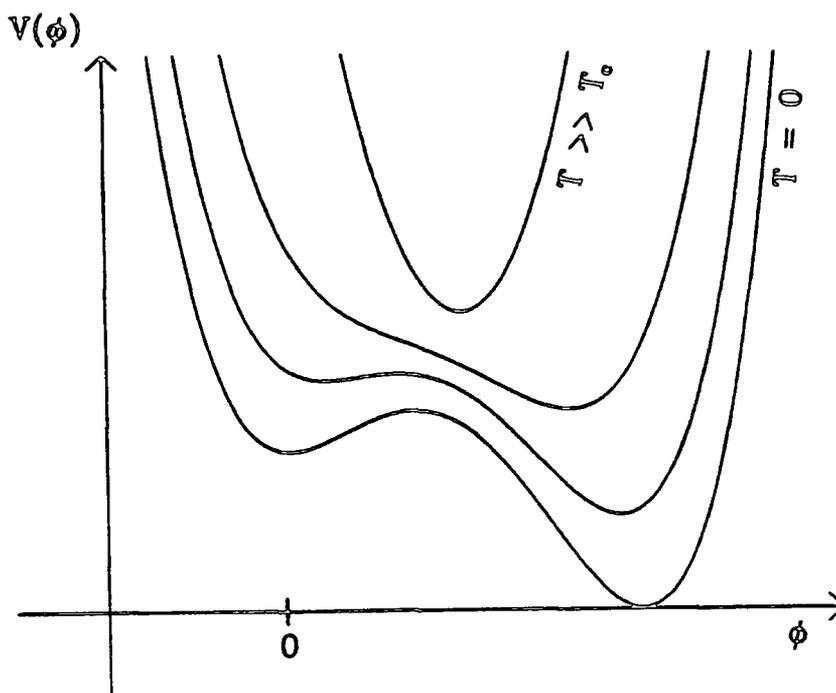
Figure 5.8: The finite temperature effective potential (5.66) for the general self-interacting scalar potential in section 5.3. $\phi = 0$ occurs at the left hand minimum of the $T = 0$ potential.

because near the critical temperature higher loop contributions will make significant contributions to the effective potential[9].

The two examples of the previous section may be used to illustrate the two generic types of phase transition. The first (5.59) can describe the second order phase transition of "new inflation" (see fig. 5.7). The flat plateau due to the log term, and the absence of an explicit mass term, make for a long period of "slow rolling" (assuming the field can be localized at the origin initially) which can generate the required amount of inflation.

The second example (5.66) could be used to describe a first order phase transition in which the field becomes trapped in the false vacuum near the origin and must tunnel out. Unfortunately, as is obvious from fig. 5.8, as it stands this potential cannot in fact generate the required false vacuum state because the high temperature minimum is skewed too heavily towards the low temperature absolute minimum by

the $\delta\mu\phi T^2$ term[14]. This can be rectified by the addition of fermions which have a Yukawa coupling to ϕ whose negative contributions, i.e., (5.41), will (for suitable couplings[14]) move the high temperature minimum back towards the origin without affecting the zero-temperature potential. Alternatively, (5.60) may be thought of as just an approximation to a more complex symmetry breaking potential (like, for example fig. 3.5) in which δ parameterises the size of the barrier between the two vacua.

These potentials together with the thermodynamic quantities of the previous chapter form the core upon which the investigations of chapters 6 and 7 are based.

References

1. L.H. Ryder, *Quantum Field Theory*, (Cambridge University Press, Cambridge, 1985)

2. M. Sher, *Phys. Rep.* **179**, 273 (1989)

3. R.H. Brandenberger, *Rev. Mod. Phys.* **57**, 1 (1985)

4. S. Coleman in: "Laws of Hadronic Matter" (Erice, 1973), ed. A. Zichichi, (Academic Press, New York, 1975)

5. E.J. Weinberg and A. Wu, *Phys. Rev.* D **36**, 2474 (1987)

6. J.S. Langer, *Ann. Phys. (NY)*, **41**, 108, (1967) and **54**, 258, (1969)

7. S. Coleman and E. Weinberg, *Phys. Rev.* D **7**, 1888 (1973)

8. T-P Cheng and L-F Li, *Gauge Theory of Elementary Particle Physics*, (Oxford University Press, Oxford, 1984)

9. A.D. Linde, Rep. Prog. Phys. **42**, 389, (1979)

10. K.T. Mahanthappa and M. Sher, *Phys. Rev.* D **22**, 1711 (1980)

11. L. Dolan and R. Jackiw, *Phys. Rev. D* **9**, 3320 (1974)

12. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, (Academic, New York, 1980)

13. J.-L Cambier, J.R. Primack and M. Sher, *Nucl. Phys.* **B209**, 372 (1982)

14. M. Gleiser, *Phys. Rev. D* **42**, 3350 (1990)

15. P.D.B. Collins, A.D. Martin and E.J. Squires, *Particle Physics and Cosmology*, (Wiley, 1989)

6

Inflation Models

All the elements that are required to determine whether a model of inflation can generate an initial thermal state have now been assembled. The general formalism designed to answer this question is demonstrated in section 6.1. Sections 6.2, 6.3 and 6.4 then apply this formalism to three specific inflation potentials, and illustrate the resulting constraints on the parameters of the potentials that the method implies. The final section provides some conclusions.

6.1 Methodology

As remarked in section 4.2, the method is basically a self-consistent estimate of the total number of interactions per inflaton N, (4.36). It is self-consistent in the sense that thermodynamic equilibrium is assumed so that a temperature T can be defined and hence the reaction rate Γ and the thermodynamic quantities defined in section 4.1 can be calculated. These quantities also determine the time evolution of T and H via the Friedmann equations (3.14) and (3.15). N may then be estimated by integrating Γ with respect to time between the points where Γ and H cross each other, this being roughly the time period over which equilibrium applies as discussed in section 4.2. Because this region of interest, through which the equations must be evolved, occurs precisely when the transition from radiation dominance to vacuum dominance is happening, analytic calculation of these quantities becomes quite intractable. Numerical evaluation by computer must therefore be resorted to. First of

all, however, the general principles will be illustrated with a greatly simplified "toy model".

It is convenient to combine the two Friedmann equations (3.14) and (3.15) by using the identity

$$\frac{d}{dt}\left(\frac{\dot{R}}{R}\right)^2 \equiv 2\left(\frac{\dot{R}}{R}\right)\left(\frac{\ddot{R}}{R} - \left(\frac{\dot{R}}{R}\right)^2\right). \tag{6.1}$$

The toy model assumes the dominance of a single massless scalar species so that, including a constant vacuum energy, $\rho = \rho_r + \rho_v$ where ρ_r is given by (4.21) with $g = 1$. Substituting this into the Friedmann equations (3.14) and (3.15) and these into (6.1) shows that $\dot{T}/T = -\dot{R}/R$, i.e., (4.26) still applies. So, from (3.14),

$$\frac{\dot{T}}{T} = \sqrt{\frac{3}{8\pi G}}\left(\rho_v + \frac{1}{2}a_{SB}T^4\right)^{1/2}. \tag{6.2}$$

Integrating with respect to t gives, instead of (4.27),

$$T = \left(\frac{2\rho_v}{a_{SB}}\right)^{1/4}\frac{1}{\sqrt{\sinh(2H_\Lambda t)}} \tag{6.3}$$

where $H_\Lambda^2 = 8\pi G\rho_v/3$. For small t, i.e., before ρ_v dominance, so that $\sinh(x) \approx x$, (4.27) is recovered, whereas for large t, after ρ_v comes to dominate

$$T \approx \left(\frac{8\rho_v}{a_{SB}}\right)^{1/4}e^{-H_\Lambda t}, \tag{6.4}$$

which is supercooling due to the inflation. From (6.3) it follows that

$$H = -\frac{\dot{T}}{T} = H_\Lambda\coth(2H_\Lambda t). \tag{6.5}$$

The other approximation to be made in the toy model is

$$\Gamma \approx n\sigma(E_{av})v(E_{av}) \tag{6.6}$$

where n is given in this case by (4.20) and E_{av}, (4.16), is

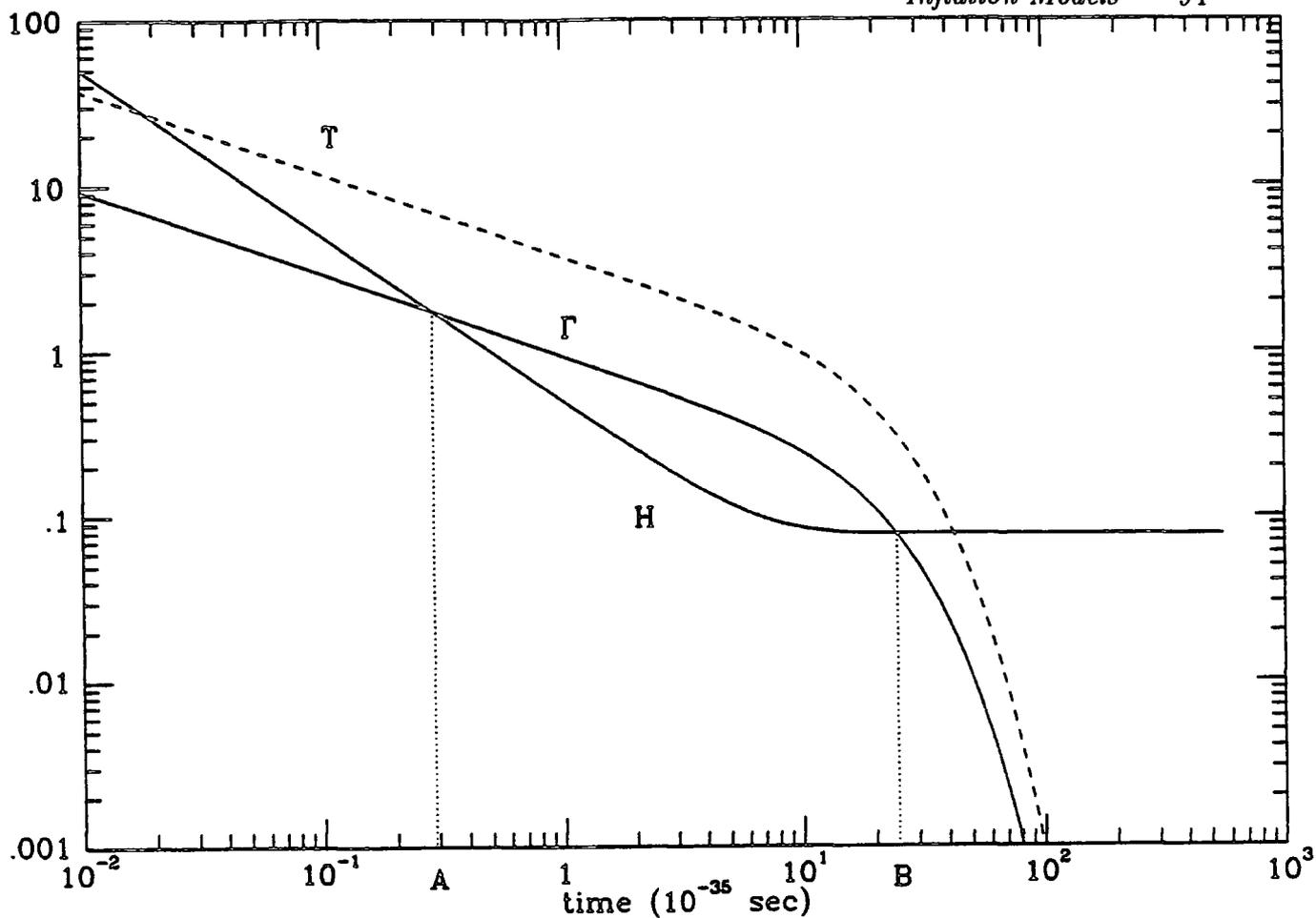$$E_{av} = \frac{\rho_r}{n} = \frac{\pi^4}{30\zeta(3)}T. \tag{6.7}$$

Figure 6.1: The behaviour of the Hubble parameter H, (6.5), reaction rate Γ, (6.9), and temperature T, (6.3), for the toy model. The choice of parameters is $\mu = 10^{14}$ GeV and $\lambda = 1$. The vertical axis units are 10^{35} s^{-1} for H and Γ and 10^{14} GeV for T. The total number of interactons N would be estimated by integrating Γ between the points A and B.

For massless particles $v = 1$, and the total cross section is taken as that given by a generic four point coupling λ (see table 2.1) so that $\sigma(E)$ to tree level, using (2.18), is

$$\sigma(E) = \frac{\lambda^2}{64\pi E^2}. \tag{6.8}$$

Putting these together yields

$$\Gamma = \frac{30^2 \zeta(3)^3}{64\pi^{11}} \lambda^2 T \approx 8.3 \times 10^{-5} \, \lambda^2 T. \tag{6.9}$$

A scale for the vacuum energy is now set by assuming the potential that generates inflation has the simple Higgs form (2.39), ignoring for the moment the fact that this implies a mass for the inflatons. Setting zero vacuum energy at the asymmetric minimum therefore implies

$$\rho_v = V(\phi = 0) - V(\phi = \phi_{\min}) = \frac{3\mu^4}{2\lambda}.$$ (6.10)

Fig. 6.1 shows the behaviour of H, T and Γ obtained if the parameters $\lambda = 1$ and $\mu = 10^{14}$ GeV are chosen, values that would be typical if this were a GUT symmetry breaking. This is the behaviour one would hope for to generate a thermal state before inflation starts. H starts out as $1/2t$ during radiation dominance, but becomes constant during inflation. During the radiation dominance epoch $\Gamma \propto t^{-1/2}$ and in this case crosses H. The quantity N would be calculated from this figure by integrating Γ between the points A and B and constraints could be obtained on the parameters μ and λ by making a contour plot of the resulting levels of N. Only parameters that produced a value of N significantly greater than 1 would be acceptable. Even for this simple model the integration would have to be done numerically as it will for the more complete models in the following sections, but the approximate shape of the contours may be predicted in advance by the following procedure. Dividing (6.9) by (6.5) using (6.10) (and replacing G by M_{Pl}^{-2} from (1.1)) gives

$$\frac{\Gamma}{H} = \frac{30^2 45^{1/4} \zeta(3)^3}{128\pi^{12}} \lambda^{9/4} \frac{M_{Pl}}{\mu} \frac{\tanh(2H_\Lambda t)}{\sqrt{\sinh(2H_\Lambda t)}}.$$ (6.11)

The maximum value of the function $\tanh(x)/\sqrt{\sinh(x)}$ is $1/\sqrt{2}$ and hence

$$\left.\frac{\Gamma}{H}\right|_{\max} = \frac{30^2 45^{1/4} \zeta(3)^3}{128\sqrt{2}\pi^{12}} \frac{M_{Pl}}{\mu} \lambda^{9/4} \approx 2.42 \times 10^{-5} \frac{M_{Pl}}{\mu} \lambda^{9/4}.$$ (6.12)

Now, constant values of $(\Gamma/H)|_{\max}$ imply roughly constant values of N and in particular, the condition $(\Gamma/H)|_{\max} = 1$ gives the value at which Γ and H just cross. Thus

the contours are expected to be fairly close to

$$\left.\frac{\Gamma}{H}\right|_{\max} \frac{\mu}{M_{\rm Pl}} = 2.42 \times 10^{-5} \, \lambda^{9/4} \tag{6.13}$$

and these are shown in fig. 6.2.

A note should be made here about the effective number of particle spin states g_*. From the point of view of inflation, which requires that ρ_v be generated by the effective scalar potential, it is only important for the scalar field to be in equilibrium with itself. So in the calculation of the reaction rate by (4.41) g has been set to one. However, when calculating the energy density to insert into Einstein's equation, contributions from all the other fields must also be included. These other fields will only contribute if they too are in thermal equilibrium, though it is possible to imagine cases where the ϕ-particles are in equilibrium and the other particles are not. It requires a further calculation (similar to the one that is made for the ϕ-particles) to check this. To keep things simple, only models where ϕ alone is the relevant field will be considered in the following sections, and so g_* can be set to one. Other reasonable choices of g_* would not affect the results very much.

The full numerical procedure that is to be followed in sections 6.2–6.4 will now be detailed. By assuming thermodynamic equilibrium ρ_T and p_T, the energy density and pressure of the inflatons, can be calculated. The general expressions are required because the inflatons are expected to have a mass of a similar scale to the vacuum energy. Since they are functions of T, following the same procedure that lead to (6.2) from (6.1), the general expression

$$dt = -\frac{1}{3}\sqrt{\frac{3}{8\pi G}} \frac{\frac{d}{dT}\rho_T}{(\rho_T + \rho_v)^{\frac{1}{2}}(\rho_T + p_T)} dT \tag{6.14}$$

is obtained, again assuming $\rho = \rho_T + \rho_v$ where ρ_v is constant. There are several inputs that are required from the temperature dependent effective potential of the particular model. First of all, the temperature dependent mass $m(T)$ enables ρ_T and p_T to be determined at any given temperature by numerically integrating (4.13) and (4.14)[1].
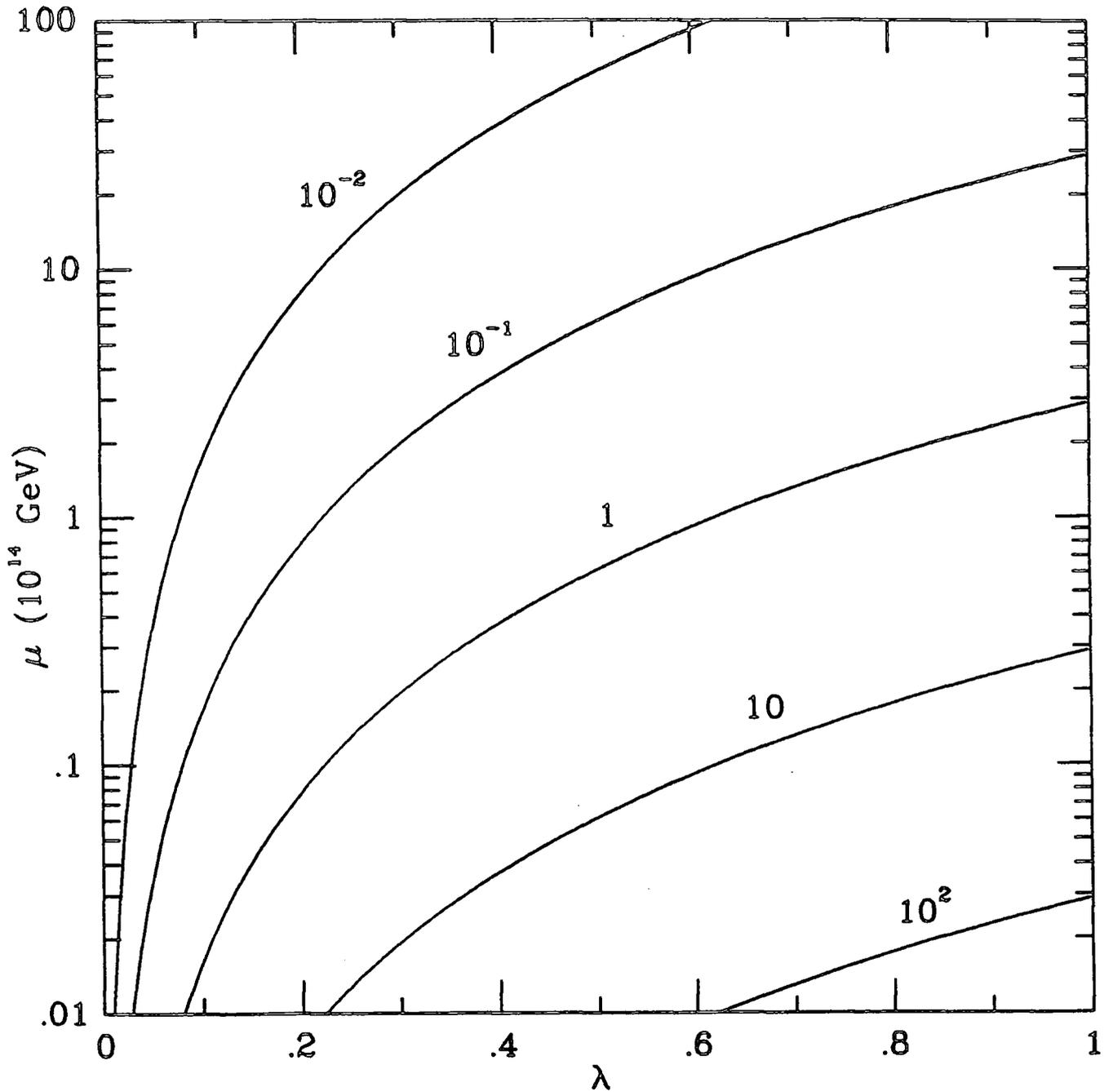
Figure 6.2: A contour plot of $(\Gamma/H)|_{\max}$ in μ-λ parameter space defined by (6.13). A contour plot of N as a function of the same parameters would be expected to look similar to this. The contour level $(\Gamma/H)|_{\max} = 1$ gives parameter values for which the curves for Γ and H just cross.

The second input is the value of ρ_v obtained by setting the vacuum energy in the asymmetric minimum of the $T = 0$ potential to zero. Dividing the temperature range of interest into a large number of steps and numerically integrating (6.14) over each step[1], now yields $T(t)$ given an initial $T(t_i)$, the starting point for the evolution. $H(t)$ can also be calculated from (3.14), since ρ_T is known at each temperature step. The initial condition $T(t_i)$ is determined by assuming radiation dominance when the evolution is started far above the ρ_v-scale. $T(t_i)$ is then determined from (4.27) which is equivalent to setting $T \to \infty$ as $t \to 0$. The third requirement of the potential is the coupling(s) of the inflatons which, along with their mass, allow the total cross section to be calculated from (2.18). $\Gamma(t)$ can then be computed by numerically evaluating the triple integral in (4.41)[2] at each step. As the evolution proceeds step by step, values of Γ which are found to be larger than H are accumulated. Finally, N is determined from these collected values of Γ by an area summation rule[3], assuming, of course, that the entire $\Gamma > H$ region has been captured in the evolution.

This complete procedure is then iterated, varying the values of the parameters which appear in the potential. A sufficient density of points in the space defined by these parameters allows contour levels of constant N to be interpolated.

The results of this programme are now described in detail for inflation models which rely on the specific potentials determined in chapter 5.

6.2   Pure Scalar Potential

Possibly the simplest model that could be considered is one where the inflaton sector is a pure scalar with the Higgs like potential (5.1), i.e., it is like the toy model but with the mass and reaction rate treated correctly. As was seen in section 5.2 the thermal corrections give[4]

$$V_{\text{eff}}(\phi, T) = -\frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4!}\phi^4 + \frac{1}{48}\lambda\phi^2 T^2 + \rho_v. \tag{6.15}$$

The fact that there is no false minimum at $\phi = 0$ in the zero temperature potential (see fig. 5.6) means that this potential can only be used to describe a second order, or "slow-rolling", phase transition as in new inflation[5].

The four-point coupling λ gives the tree-level cross section (6.8). Again, as in the toy model, the vacuum energy density ρ_v is determined by requiring that $V_{\text{eff}}(\phi = \phi_{\min}, T = 0) = 0$, to agree with observation today, giving (6.10). And, as in section 5.4, the temperature dependent mass when the field is localized at $\phi = 0$ and high temperature is (5.67), i.e.,

$$m^2(T) = -\mu^2 + \frac{\lambda}{24}T^2. \tag{6.16}$$

From this it may be seen that $m(T) \to 0$ as $T \to \sqrt{24\mu^2/\lambda}$, the critical temperature, so it might seem that there would be a problem defining a mass for the scalar particles below the critical temperature. This is simply due to the fact that there is no longer a minimum of V_{eff} at $\phi = 0$. The gradient $\partial V_{\text{eff}}/\partial\phi$ becomes negative near here, and so it is presumed that the vacuum expectation value will start evolving away from $\phi = 0$ as soon as the critical temperature is passed. Since it is difficult to know the precise details of how this evolution will occur, and since the one-loop potential is known to be untrustworthy near the critical point due to the contributions of higher orders in perturbation theory[6], the following approximation is used. As soon as the critical temperature is passed it is assumed that field becomes localized in the new minimum, in other words the evolution between vacua is assumed to be instantaneous. This approximation will be justified later. The mass in this new minimum given by $\partial^2 V_{\text{eff}}/\partial\phi^2$ is now $m^2(T) = 2\mu^2 - \lambda T^2/12$. Moreover, the value of the potential at the new minimum will not have reached zero so, for a time, there will be a temperature dependent vacuum energy density

$$\rho_v = \frac{1}{8}\left(\mu^2 - \frac{\lambda}{48}T^2\right). \tag{6.17}$$

The procedure outlined in section 6.1 can now be followed, though below the critical temperature (6.14) must be modified slightly to

$$dt = -\frac{1}{3}\sqrt{\frac{3}{8\pi G}} \frac{\frac{d}{dT}(\rho_T + \rho_v)}{(\rho_T + \rho_v)^{\frac{1}{2}}(\rho_T + p_T)} dT \tag{6.18}$$

to handle the temperature dependent ρ_v of (6.17). The consequent evolution of H, Γ, T, and s is plotted in fig. 6.3 for the choices $\lambda = 1$ and $\mu = 10^{14}$ GeV. The cusp-like
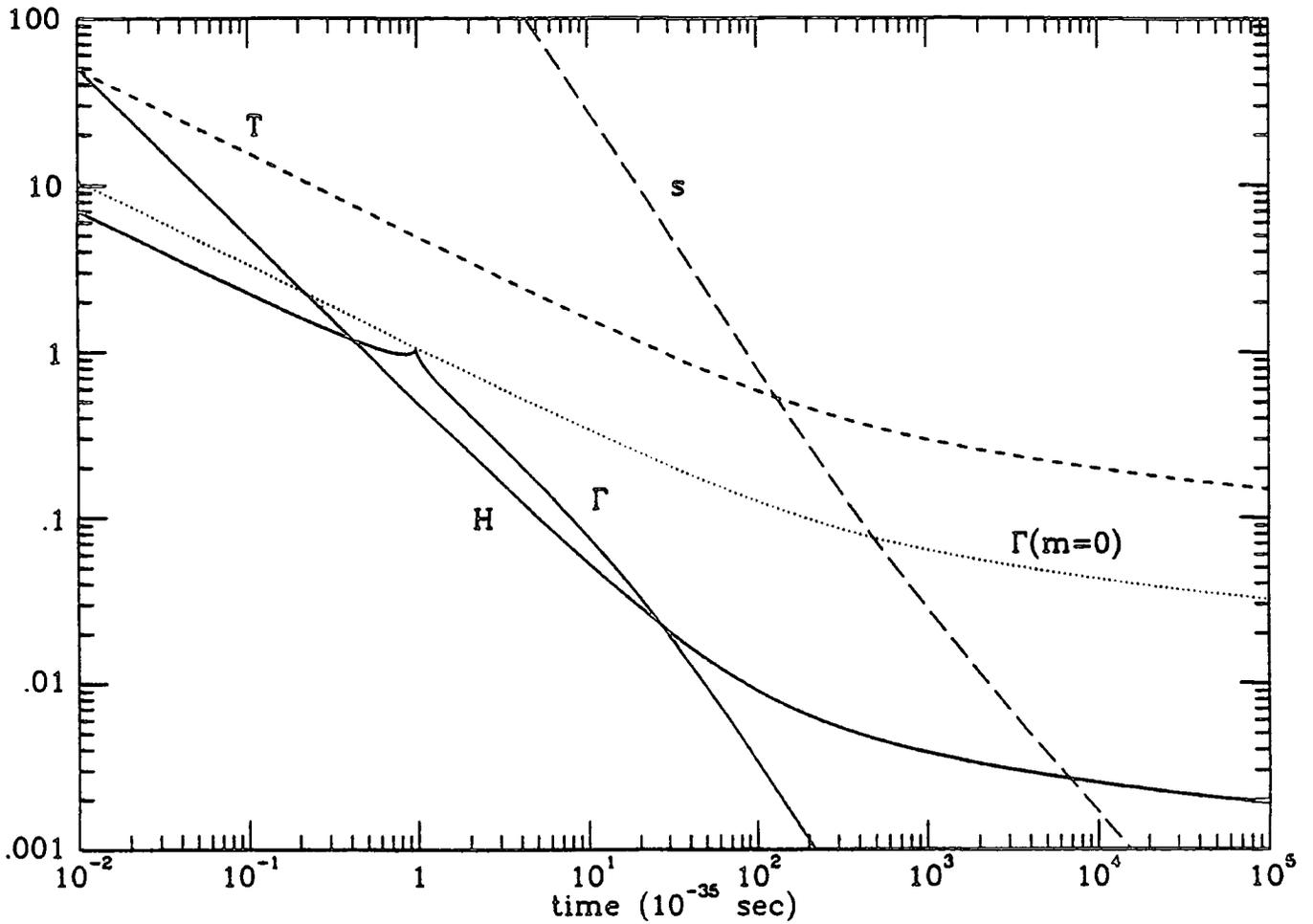
Figure 6.3: A plot of the expansion rate H, reaction rate Γ, temperature T and entropy density s as a function of time for a model based on the pure scalar potential of equation (6.15). $\Gamma(m = 0)$ is the massless limit of reaction rate (6.19). The choice of parameters in (6.15) are $\mu = 10^{14}$ GeV and $\lambda = 1$. The vertical axis units are 10^{35} s^{-1} for H and Γ, 10^{14} GeV for T and 10^{65} JK^{-1}m^{-3} for s.

point in the reaction rate at the critical point is due to the mass being swapped from $m^2 = -\mu^2 + \lambda T^2/24$ to $m^2 = 2\mu^2 - \lambda T^2/12$. In the limit $m \to 0$ the reaction rate (4.41) may be integrated exactly for the cross section (6.8) giving

$$\Gamma(m = 0) = \frac{\lambda^2 \pi}{(24)^2 .64\zeta(3)} T \qquad (6.19)$$

(where Riemann $\zeta(3) \approx 1.202$). This is also plotted in fig. 6.3 and it can be seen that Γ does take this value at the critical point where $m \to 0$. Despite the reaction-rate

cusp, the entropy density is continuous and smooth so that at least the model is consistent. In fact the real situation is not expected to be very different from this because the large value of λ ($= 1$) that has been used implies a very short period of evolution between vacua. The quantities displayed in fig. 6.3 can only possibly be valid during the time when $\Gamma > H$, and then only if sufficient interactions have occurred for a thermal state to form. If λ is reduced too much, Γ will not cross H at all. In fact $\Gamma < H$ for all t if the tip of the cusp is less than $H(T_c)$. By making the approximation that the critical temperature occurs during the radiation dominance period, and then calculating $H(T_c)$ from (3.14) by ignoring ρ_v and using (6.19) to find $\Gamma(T_c)$ (and replacing G with M_{Pl}^{-2}, (1.1)), it may be found that

$$\lambda > \left(\sqrt{\frac{96\pi}{45}}(24)^2 .64\zeta(3)\frac{\mu}{M_{Pl}} \right)^{2/5} \simeq 106 \left(\frac{\mu}{M_{Pl}} \right)^{2/5} \qquad (6.20)$$

for Γ to cross H. For $\mu = 10^{14}$ GeV this implies $\lambda > 0.975$. So it would seem that thermal state production requires a rather large value of λ, in contradiction with the tiny couplings ($\sim 10^{-14}$) required for new inflation to produce a homogeneous universe. This point was originally made in [7].

As would probably be expected with such a simple evolution between the vacuum states, inflation does not occur in this model — H does not become roughly constant because ρ_v does not. It is obviously the details of the vacuum evolution between minima that generates this ρ_v and for this evolution to be long enough a small value of λ is required, yet it is just such values of λ which preclude the formation of a thermal state. In contrast to the conclusions of the toy model, it appears a more complicated potential is needed to generate new inflation properly.

6.3 Coleman-Weinberg Potential

The next potential to consider is one where the scalar particle couples to a gauge field. If there is no mass term for the scalar, then symmetry breaking occurs via radiative corrections as was seen in section 5.3, which leads to the Coleman-Weinberg potential. For a complex scalar singlet coupled by a $U(1)$ gauge, the temperature
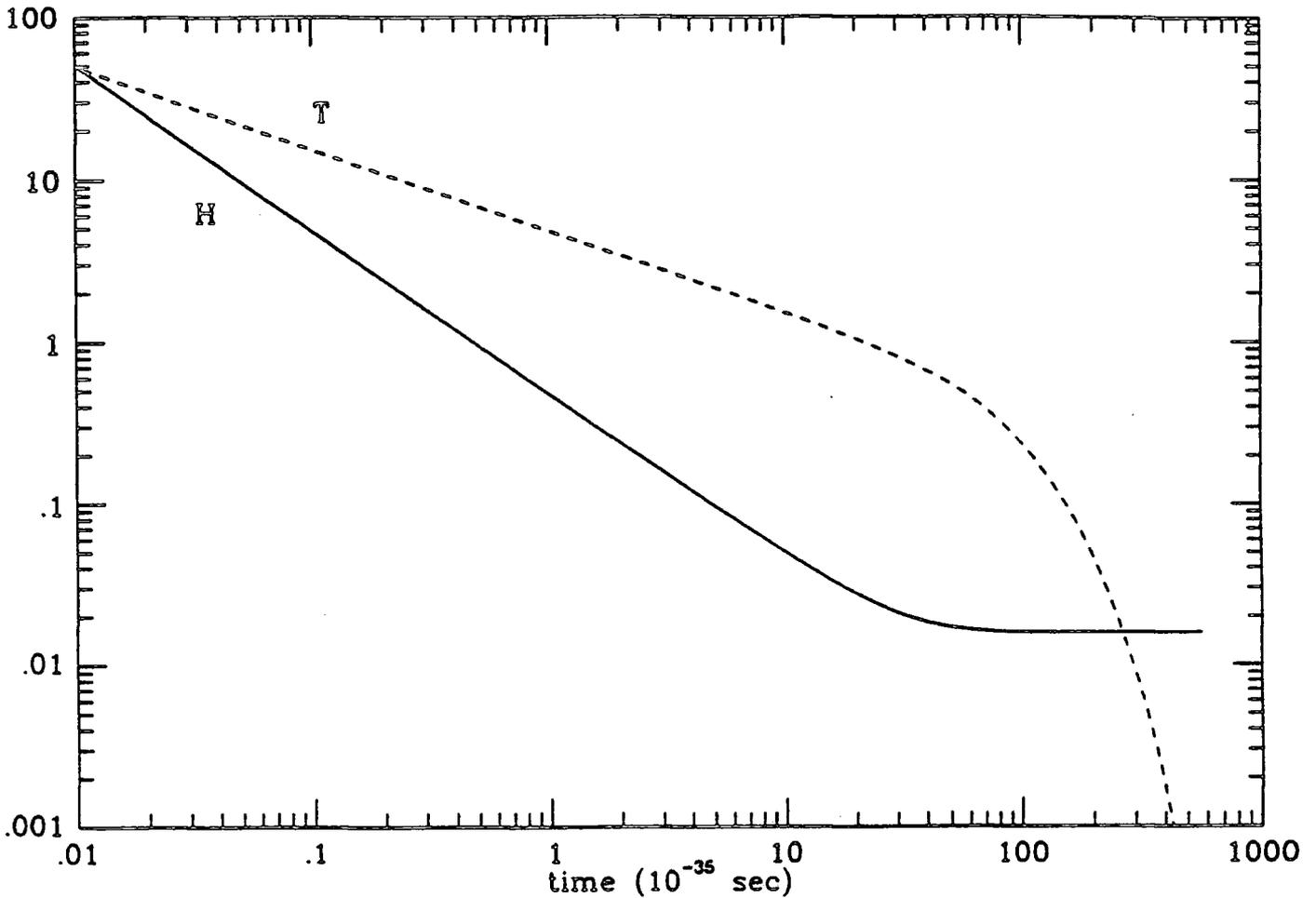
Figure 6.4: A plot of the expansion rate H and the temperature T for the Coleman-Weinberg model (equation (6.21)) with parameters $\phi_{min} = 10^{15}$ GeV and $e = 0.3$. The vertical scale is 10^{35} s^{-1} for H and 10^{14} GeV for T.

dependent effective potential is (5.59)[8, 4]

$$V_{\text{eff}}(\phi, T) = \frac{3}{64\pi^2}e^4\phi^4 \left(\ln\left(\frac{\phi^2}{\phi_{min}^2}\right) - \frac{1}{2}\right) + \frac{11}{128\pi^2}e^4\phi^2 T^2 + \rho_v. \qquad (6.21)$$

The lack of a mass term means that there is one less parameter in the potential which simplifies the analysis. It also implies that the potential is very flat near the origin so that the potential is suitable for the slow rolling transition of new inflation[5, 9].

Once again, setting the vacuum energy today $V_{\text{eff}}(\phi_{\text{min}}, 0) = 0$, yields

$$\rho_v = \frac{3e^4}{128\pi^2}\phi_{\text{min}}^4 \qquad (6.22)$$

in the $\phi = 0$ state. The temperature dependent mass $(\partial^2 V_{\text{eff}}/\partial \phi^2)|_{\phi=0}$ is

$$m^2(T) = \frac{22e^4}{128\pi^2}T^2 \tag{6.23}$$

which is greater than zero for all non-zero temperatures unlike the pure Higgs potential above. The evolution of H and T can now be followed, and a typical result is shown in fig. 6.4. The $1/2t$ radiation dominated dependence of H gives way to a constant when inflation starts.

In order to calculate the reaction rate from (4.41) the relevant cross section needs to be determined. In this case the picture is complicated somewhat by the fact that there are different sets of particles depending on whether it is the symmetric or broken symmetry phase which applies. If the symmetry is broken then there is a single scalar and the gauge particle gains a mass via the Higgs mechanism. In the symmetric phase there is a scalar with an anti-particle partner and these are coupled via a massless gauge particle, i.e., the situation is similar to scalar QED but the mass of the particles is given by (6.23).

So, in the symmetric phase the relevant cross section is the two-particle elastic cross section which is calculated from the four-point coupling and the two exchange diagrams shown in fig. 6.5 using the rules in table 2.1 and (2.18). The four-point coupling may be ignored as negligible compared to the gauge-boson diagrams because, as noted in section 5.3, λ is $O(e^4)$ so that the largest contribution to the cross section from this diagram, the interference term, is $O(e^6)$. The effect of the gauge particle production diagram, i.e., the t-channel version of 6.5(a), which involves both particle and antiparticle degrees of freedom, could in principle be included but in fact it turns out to have much smaller contribution to the reaction rate than the exchange diagrams, as will be shown below.

Zero mass gauge-boson exchange (fig. 6.5) presents difficulties because of the divergence at zero-momentum transfer. Such an infra-red divergence arises because a massless exchange allows the particles to scatter even when they are widely separated. However, the finite density of states allows a natural momentum cutoff to be defined.
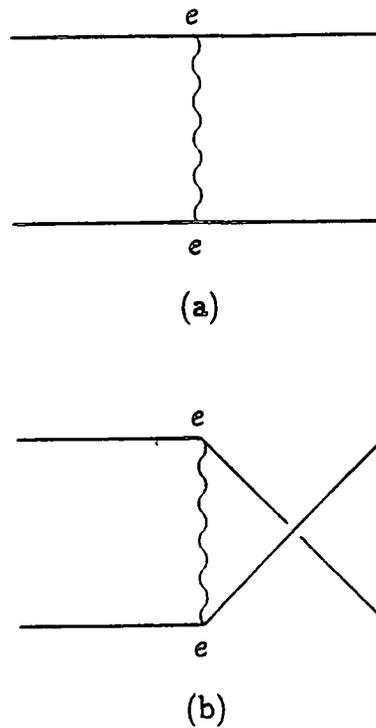
e

(a)

e

e

(b)

Figure 6.5: The gauge-boson exchange diagrams for the Coleman-Weinberg model used to calculate the cross-section in the reaction rate integral (4.41).

In particle physics language, the scattering of any two particles is unlikely to occur if their separation is greater than the average separation of the particles at that density, since there will then be other particles between them. Hence, an effective maximum impact parameter for the scattering of such states is

$$b_{\max} = n_{c.m.}^{1/3} \qquad (6.24)$$

corresponding to the average separation of the particles at the given temperature. The subscript c.m. refers to the centre-of-momentum frame which is where the cross section is to be evaluated, to insert into the reaction rate integral (4.41). $n_{c.m.}$ may be calculated from n (4.12), the value in the comoving frame, by a Lorentz contraction of the volume, i.e.,

$$n_{c.m.} = \frac{E_1 + E_2}{2E} n \qquad (6.25)$$

where E_1 and E_2 are the energies of the ingoing states and E is their centre-of-

momentum energy. Now $b_{\max}$ corresponds to a minimum 3-momentum cutoff

$$|\mathbf{k}_{\min}| = \frac{1}{b_{\max}} = n_{\text{c.m.}}^{1/3}. \tag{6.26}$$

This may be re-expressed in the centre-of-momentum frame as a minimum scattering angle, $\theta_{\min}$ between ingoing and outgoing states, i.e.,

$$\cos(\theta_{\min}) = -\frac{|\mathbf{k}_{\min}|^2}{2(E^2 - m^2)} + 1 \equiv y_m, \tag{6.27}$$

which can be used to cut off the phase space integral in (2.18) in the total cross section. The total cross section is thus

$$\sigma(E) = \frac{1}{32\pi E^2 p^4}\left\{ 2(2E^4 + 4E^2 p^2 + p^4)\frac{y_m}{1 - y_m^2} + 2p^4 y_m + (2E^4 - p^4)\ln\left(\frac{1 + y_m}{1 - y_m}\right) \right\}. \tag{6.28}$$

The cross section for the production diagram on the other hand is just

$$\sigma_{\text{prod}} = \frac{e^4}{64\pi E^2}. \tag{6.29}$$

The reaction rate Γ may now be calculated from (4.41) and compared to the Hubble parameter H. This is done in fig. 6.6 for the same choice of $\phi_{\min}$ and e as fig. 6.4. It may be seen that the contribution from the production diagram is indeed negligible. Also displayed in this figure is the approximation to the reaction rate (6.6) but where $E_{\text{av}} = \rho/n$ is the average value of the energy in the thermal distribution, i.e., (4.16). It is not obvious that such an approximation will always be valid, for example, when the energy density is changing from radiation to matter dominance and vacuum energy is present as well (this is why the general expression (4.16) is needed for E_{av}). The figure shows that for this model the approximation has the same form as the exact result but is too large by a factor of five or so.

Again it should be emphasised that these calculations of Γ are not valid when $\Gamma < H$ because then there are not enough interactions for thermal equilibrium to
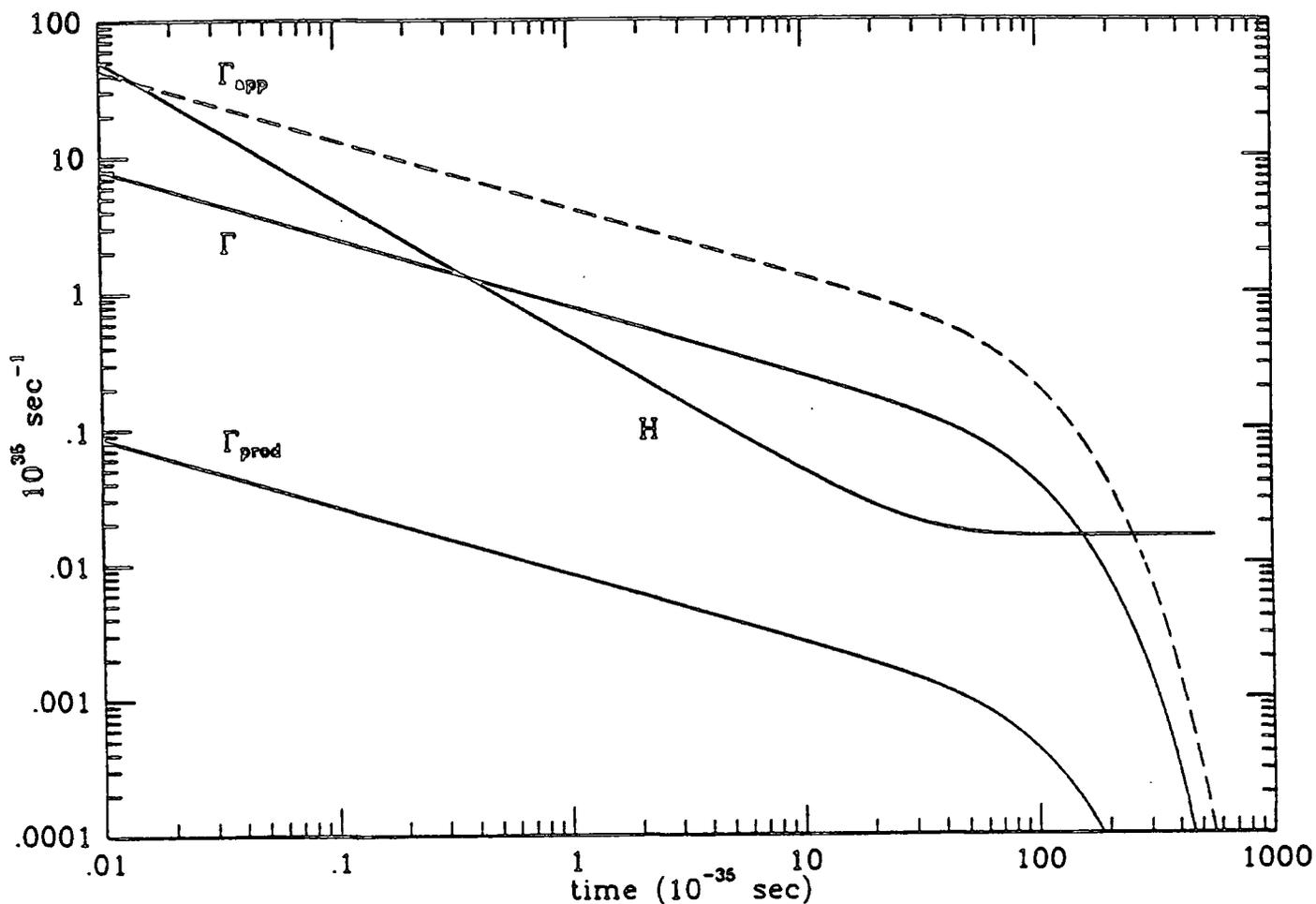
Figure 6.6: A plot of the expansion rate H, and reaction rates Γ, for the Coleman-Weinberg potential (6.21) with parameters $\phi_{min} = 10^{15}$ GeV and $e = 0.3$. Γ is the full calculation of the reaction rate (4.41) for the processes shown in fig. 6.5 and Γ_{app} is the approximation in equation (6.6). Γ_{prod} is the full calculation for the cross-section given by just the production diagram (t-channel diagram of fig. 6.5(a)). The average total number of interactions, N, is calculated by integrating Γ over the time interval between the points where it crosses H, and in this case is 14.2.

hold. Similarly, the calculation of H is only valid during radiation dominance if there is thermal equilibrium because the radiation energy density can only be calculated assuming a well defined temperature (if there were other particle species in equilibrium then of course they would define a temperature from which H could then be calculated). However, the constancy of H during inflation is still valid, even for $\Gamma < H$, because at zero temperature the effective potential still provides a non-zero vacuum

energy density, assuming the field is localized near the origin. The figure shows that the procedure for calculating N (section 6.1) can now be employed to determine whether a thermal state is likely, and that for the chosen parameters $N = 14.2$.

In order to determine constraints on $\phi_{\min}$ and e, this procedure has to be carried out for many different choices of these parameters. The results are conveniently summarised as a contour plot in $\phi_{\min}$-e space (see fig. 6.7). Certain regions of parameter space are excluded where the number of interactions is zero, i.e., where Γ does not cross over H. Also the larger the value of N the more likely it is that a thermal state can form, i.e., for large couplings and small symmetry breaking scales. For a symmetry breaking scale $\phi_{\min} \simeq 10^{15}$ GeV a coupling of $e \gtrsim 0.3$ is needed, and again the conflict with new inflation couplings is apparent. It should be noted for comparison however, that the coupling in typical GUT theories is $e_{\rm GUT} = (4\pi\alpha_{\rm GUT})^{1/2} \simeq 0.56$ for $\alpha_{\rm GUT} = 1/40$. Alternatively, if small couplings are demanded, $e < 0.01$ say, then the symmetry breaking scale is required to be small too, $\phi_{\min} \lesssim 5 \times 10^{12}$ GeV.

It would of course be possible to evade such constraints by adding extra features to the model. An example would be a very heavy gauge boson which couples strongly to the ϕ-particles and hence establishes a thermal state, but decays before the epoch of ρ_v-dominance (otherwise the false vacuum would decay too quickly). However, the only reason to invoke such extra couplings is to establish a thermal state after which their influence must vanish, which seems rather artificial.

Although it does not seem possible for such a model to generate inflation of the kind needed to solve the problems outlined in section 3.6, the possibility that such models are relevant to phase changes at much lower energies should not be ruled out.

6.4   Old Inflation Potential

Despite the well known graceful exit problem of old inflation[10], it is interesting to investigate the possibility of producing a thermal state in a model based on a potential which produces a first order phase transition. In fact, such "first order inflation" models, where the transition is completed by percolation of bubbles, have enjoyed a resurgence of interest lately. The foremost example of such a model will be
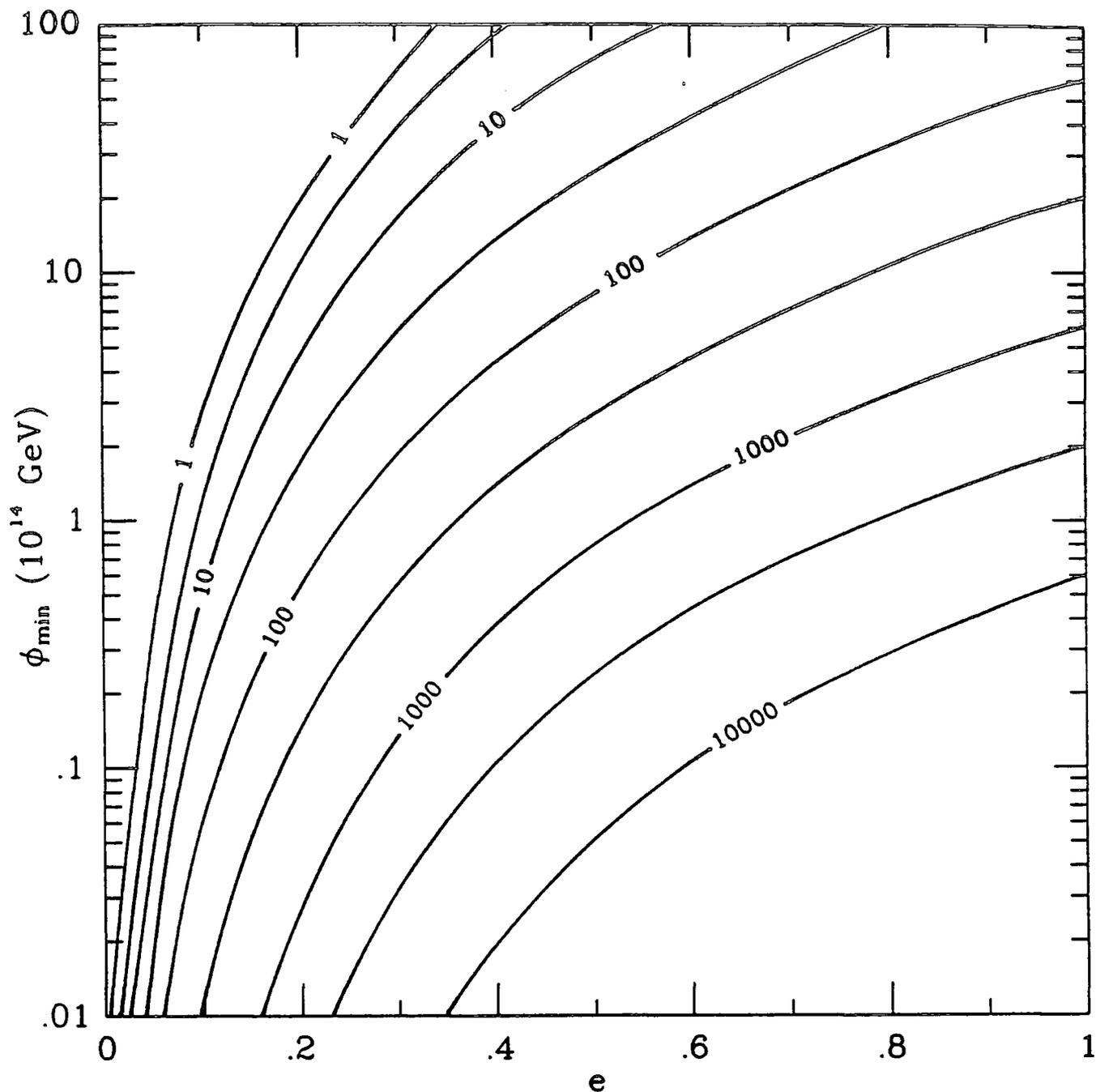
Figure 6.7: A contour plot of the average total number of interactions, N, calculated from equation (4.36) in the ϕ_{min}-e parameter space for the model based on the Coleman-Weinberg potential (6.21). The intermediate contours are at half orders of magnitude (i.e. $\sqrt{10}$ times the previous contour level). The parameter-space region to the left of the contour $N = 1$ is excluded because insufficient reactions occur to form a thermal state.

examined in the next chapter. The temperature dependent effective potential to be used is (5.66)[11], i.e.,

$$V_{\text{eff}}(\phi, T) = \frac{1}{2}\mu^2\phi^2 - \frac{\delta\mu}{3!}\phi^3 + \frac{\lambda}{4!}\phi^4 - \frac{\delta\mu}{24}\phi T^2 + \frac{\lambda}{48}\phi^2 T^2 + \rho_v. \qquad (6.30)$$

As discussed in section 5.4, in order for (6.30) to be a realistic inflation potential it must either have an extra fermion Yukawa coupling to modifying the final term, or it is only an approximation to some more complete symmetry breaking potential. (6.30) however, has all the properties that are required for the present purposes. A more complete treatment could be made if ever the inflation potential were known exactly, but the qualitative features are unlikely to differ from those described below.

The temperature dependent mass for (6.30) is

$$m^2(T) = \mu^2 + \frac{\lambda}{24}T^2 \qquad (6.31)$$

showing the finite mass at $\phi = 0$ and $T = 0$ consistent with a false minimum at the origin. The vacuum energy is determined in the usual way giving

$$\rho_v = \left(\frac{3}{2\lambda} - \frac{\delta\Delta}{3\lambda^2} - \frac{9\delta^2}{4\lambda^2} + \frac{9\delta^3\Delta}{8\lambda^3} + \frac{9\delta^4}{16\lambda^3}\right)\mu^4 \qquad (6.32)$$

where

$$\Delta = \sqrt{\frac{\delta^2}{4} - \frac{2\lambda}{3}}. \qquad (6.33)$$

The cross section to be inserted into the reaction rate (4.41) is calculated from the three diagrams in fig. 6.8 as well as the four-point coupling, which gives

$$\sigma(E) = \frac{\lambda^2}{64\pi E^2} + \frac{(\delta\mu)^2\lambda}{64\pi E^2}\left[\frac{2}{4E^2 - m^2} - \frac{1}{p^2}\ln\frac{4p^2 + m^2}{m^2}\right]$$

$$+\frac{(\delta\mu)^4}{128\pi E^2}\left[\frac{4}{m^2(4p^2 + m^2)} + \frac{2}{(4E^2 - m^2)^2} + \frac{m^2}{p^2(2p^2 + m^2)(4E^2 - m^2)}\ln\frac{4p^2 + m^2}{m^2}\right].$$
$$(6.34)$$

H and Γ are evolved using the procedure described in section 6.1 and for typical choices of λ, δ and μ the plot in fig. 6.9 is obtained. In this case the total number of
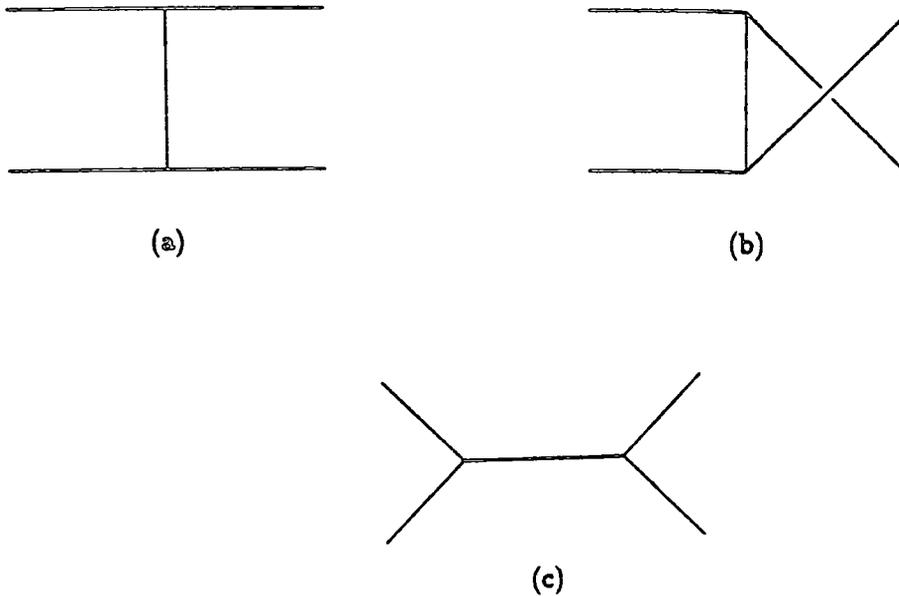
Figure 6.8: The Feynman diagrams due to the cubic coupling term in the "old-inflation" potential (6.30). The coupling is $\delta\mu$ at each vertex. These diagrams, as well as the quartic coupling (see table 2.1), are used to calculate the cross-section for this model and hence the reaction rate (4.41).

interactions per particle $N = 4.7$. Note the interesting behaviour of the temperature which arises directly from (6.14); (6.31) gives $m \rightarrow \mu$ at low temperature so the matter dominated form for ρ_T (4.29) and p_T (4.30) can be used in (6.14) which implies $T \sim 1/t$. Once $T < \mu$ however, Γ gets damped out by the Boltzmann factor $e^{-\mu/T}$. The approximate calculation of Γ has again been tested for this model in fig. 6.9 and it can be seen that this time it is even more different from the exact result, though its behaviour has similar features.

Constraints on λ, δ, μ, can now be deduced from the calculations of N. Unfortunately, since there are three parameters, a simple contour plot as in the previous section is not possible in this case. However, a single value of N can be selected to determine whether a thermal state has been produced, so one can plot the contour level with this value of N in the λ-δ plane for several different values of μ. Since the purpose here is to produce bounds on the parameters of the model, the $N = 1$
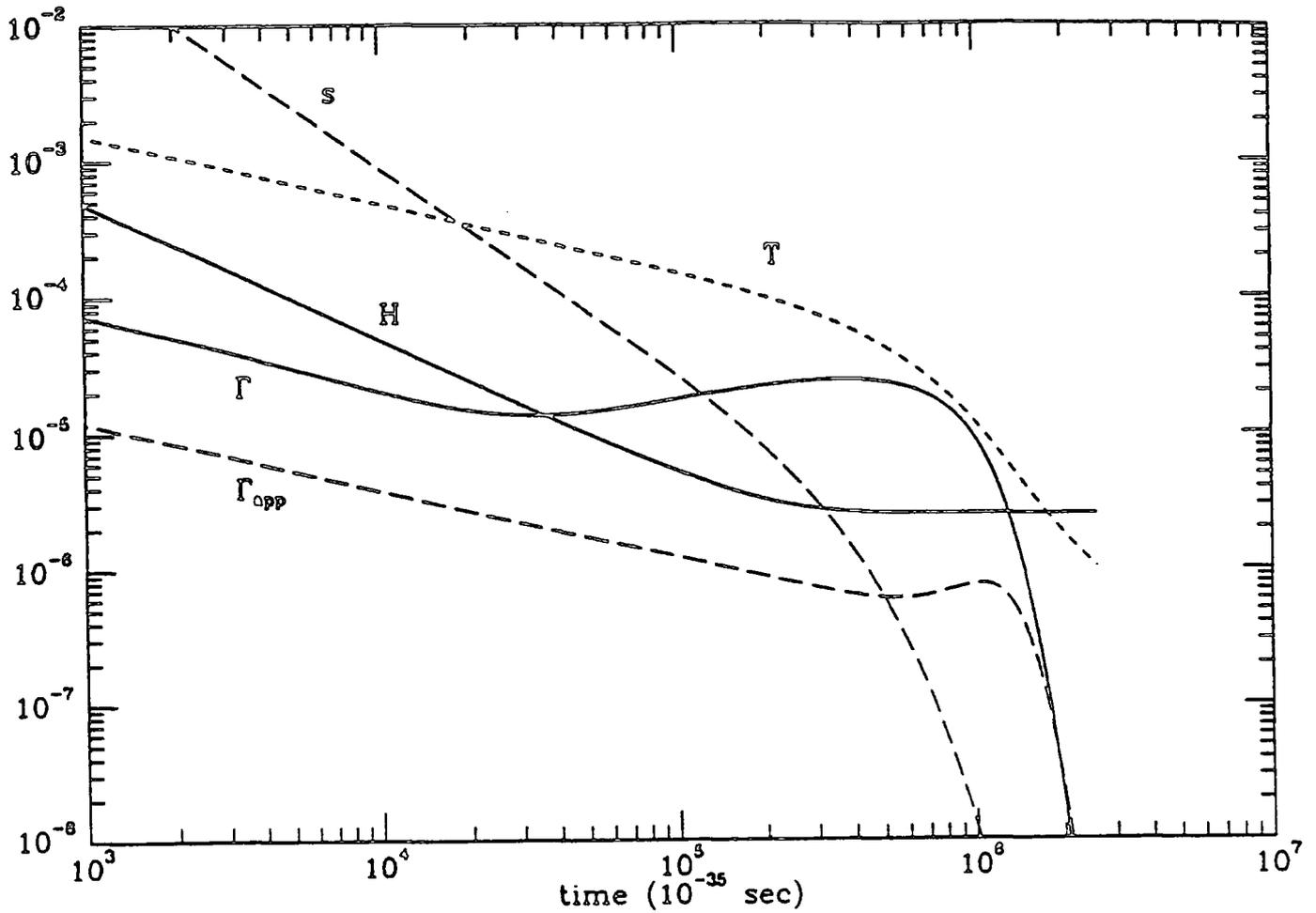
Figure 6.9: A plot of the expansion rate H, the temperature T, the full calculation of the reaction rate Γ from (4.41), and the approximate reaction rate Γ_{app} from (6.6), for the "old-inflation" potential (6.30). The choice of parameters is $\mu = 10^{11}$ GeV, $\lambda = 0.05$ and $\delta = 0.9$. The vertical axis units are 10^{35} s^{-1} for H and Γ, 10^{16} GeV for T and 10^{65} JK^{-1}m^{-3} for s. The average total number of interactions in this case is 4.7.

level contour has been selected (see fig. 6.10), the lowest value for the beginnings of a thermal state. Note that the parameters in the lower right-hand region are excluded because they give an unsuitable zero-temperature effective potential, either because the asymmetric minimum of V_{eff} is higher than that at $\phi = 0$ or because there is no asymmetric minimum at all.

It is obvious from fig. 6.10 that GUT scale inflation is impossible with this potential for any reasonable value of the couplings — the value of δ would have to be
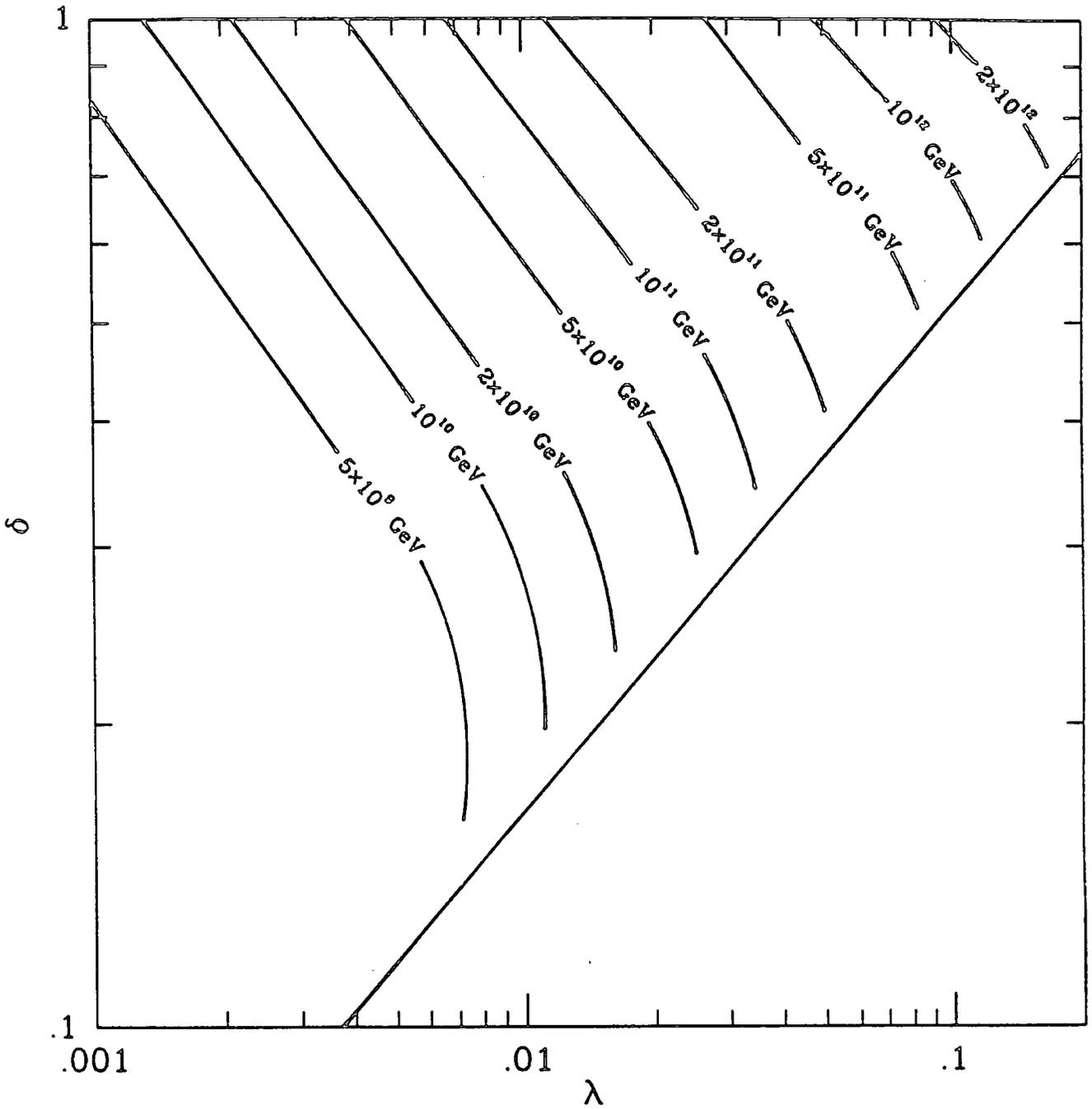
Figure 6.10: A plot of the $N = 1$ contours for various values of μ in δ-λ space for the "old-inflation" model (6.30). Regions to the left of this contour cannot form a thermal state for the given value of μ because insufficient interactions occur. The lower-right region is excluded because here the parameters do not give rise to a suitable zero-temperature symmetry-breaking potential.

so large that perturbation theory and hence the effective potential formalism would be invalid. To be consistent with perturbation theory, i.e., $\delta, \lambda < 1$, the symmetry breaking scale must be less than about 10^{12} GeV (assuming, naturally, that extra heavy gauge couplings are not invoked to generate thermal states initially) which is perhaps more suggestive of a broken SUSY model.

6.5   Summary

What has been shown in the preceding sections is that models of inflation which rely on an initial state of thermal equilibrium, can be tested in a self-consistent way to see if this state is possible. The method requires that the temperature dependent effective potential be known. This potential provides all the input data that is needed for the procedure described in section 6.1 to be carried out; viz., the size of the vacuum energy and the mass and couplings of the particles.

The specific models that have been tested in the above are all found to be lacking. The old inflation potential of section 6.4 of course suffers from the graceful exit problem. The fact that the tiny couplings required of new inflation potentials like those in sections 6.2 and 6.3 are inconsistent with the values required to establish an initial thermal state has also been pointed out before[12, 13]. This becomes very clear using the above formalism with which it can be seen just how far from establishing a thermal state such models are.

Even if the constraints relating to the duration of or exit from inflation are ignored, then the earliest time a thermal state could be generated in any of the above models is roughly the 10^{15} GeV scale for the largest couplings allowed by perturbation theory (in order that the effective potential approach is valid). To obtain a thermal state at higher energies the above models would require very strong couplings (unless of course additional features such as coupling via an unstable superheavy gauge boson is invoked). However, if such non-perturbative couplings are employed, it is not clear how the finite temperature effects should be estimated. Furthermore, GUT couplings in general become weaker at higher energies. It would therefore seem that if inflation is to occur at larger than GUT scales then some mechanism other than a

temperature-dependent phase change must be favoured, for example chaotic inflation mentioned in section 3.7[12].

Even though the models in sections 6.2–6.4 cannot describe inflation at the GUT scale it is still conceivable that they are relevant to phase changes nearer to the electroweak scale. If the traditional problems (i.e., the horizon and flatness/oldness problems) that inflation is supposed to solve are ignored at such scales, perhaps a phase change could generate false vacuum states for only short times. Such a localized (in space) inflationary effect might even be relevant to the formation of the bubbles and voids seen in recent observations of large scale structure[14].

However, the main conclusion here is that a formalism for studying thermal state generation has been successfully implemented. If a more complete model of inflation is constructed then it will be possible to test it in the manner described above. Such a model is the subject of the next chapter. The old inflation potential describes the most general renormalizable self-coupled scalar field, whilst the Coleman-Weinberg potential illustrates the effect of gauge couplings. Although fermions cannot generate a vacuum energy by themselves, they can modify the effective scalar potential at first order in perturbation theory via loop contributions (5.41), and this is in fact necessary to make the old inflation potential work properly (see section 5.4). Nevertheless, such modifications are usually not severe enough to change the qualitative features that have already been described. So the models used above seem to cover all the salient features relevant to the generation of phase changes and vacuum energies from the usual extensions of the standard model of particle physics. It may of course be that the picture of fundamental physics at the GUT scale is completely different from what has been described here, being more than ten orders of magnitude above the highest energies achieved in colliders. But if so, provided an effective potential formalism can be constructed it should still be possible to test thermodynamic equilibrium in a similar way.

References

1. Numerical Algorithms Group, Fortran Library Mk. 15 (1991), Routine D01BAF

2. Numerical Algorithms Group, Fortran Library Mk. 15 (1991), Routine D01FCF

3. Numerical Algorithms Group, Fortran Library Mk. 15 (1991), Routine D01GAF

4. R.H. Brandenberger, *Rev. Mod. Phys.* **57**, 1 (1985)

5. A.D. Linde *Phys. Lett. B* **108**, 389 (1982); *Phys. Lett. B* **114**, 431 (1982); A. Albrecht and P.J. Steinhardt *Phys. Rev. Lett.* **48**, 1220 (1982)

6. A.D. Linde, *Particle Physics and Inflationary Cosmolgy*, (Harwood Academic, Chur, Switzerland, 1990)

7. A.D. Linde, Rep. Prog. Phys. **42**, 389, (1979)

8. S. Coleman and E. Weinberg, *Phys. Rev. D* **7**, 1888 (1973)

9. P.D.B. Collins, A.D. Martin and E.J. Squires, *Particle Physics and Cosmology*, (Wiley, New York, 1989)

10. A. Guth and E. Weinberg, *Nucl. Phys.* **B212**, 321 (1983)

11. M. Gleiser, *Phys. Rev. D* **42**, 3350 (1990)

12. A.D. Linde *Phys. Lett. B* **129**, 177 (1983)

13. R.H. Brandenberger and H.A. Feldman, *Physica* **A 158**, 343 (1989)

14. M.J. Geller and J.P. Huchra, *Science* **246**, 897, (1989)

7

Extended Inflation

It has been appreciated already, in chapters 3 and 6, that inflation due to a phase change has some serious difficulties; the graceful exit problem for a first-order transition and the fine-tuning problem for the second-order version. Some of the alternatives to a transition originating from finite temperature field theory have also been briefly mentioned. However, it has also been found possible to achieve a graceful exit from inflation without such a radical modification of the mechanism. There are many who believe that this approach is the more natural.

One of the most popular recent ways of evading the graceful exit problem is to modify Einstein gravity to the more general theory of Brans and Dicke. In this theory the exponential expansion of conventional inflation during the period of vacuum energy dominance, is softened to a power law, whence it becomes known as "extended inflation"[1]. It can be shown that such power law inflation can allow percolation of the non-inflating phase to occur. The objective of this chapter is to understand how extended inflation modifies the constraints that arise from the requirement for an initial thermal state. To this end, Brans-Dicke theory is considered in more detail in section 7.1, along with its implications for cosmology. Section 7.2 explains how to evolve the resulting equations of motion so that a numerical procedure similar to that of the previous chapter is possible. From this analysis it is found that Brans-Dicke gravity enjoys several extra parameters and the constraints on these are given in section 7.3. In sections 7.4 and 7.5, detailed models based on the first order poten-

tial and Coleman-Weinberg potential are considered while the final section supplies a summary.

7.1 Brans-Dicke Action

If Newton's constant G in the gravitational action (3.1) is generalised to a field Φ, with dimensions of (energy)^2, then the action becomes

$$S_{EI} = \int d^4x\sqrt{g}\left(-\Phi\mathcal{R} + \omega\frac{\left(\partial_\mu\Phi\right)^2}{\Phi} + 16\pi\mathcal{L}_{\text{matter}}\right) \tag{7.1}$$

which is the scalar-tensor theory of Brans and Dicke[2]. As before, $\mathcal{L}_{\text{matter}}$ contains the scalar-particle symmetry-breaking potential which generates inflation. ω is a constant free parameter and GR is recovered in the limit $\omega \to \infty$. It is assumed that Φ does not contribute to the energy-momentum tensor $T^\mu{}_\nu$ so that the principle of equivalence is not violated and the usual conservation law $D_\nu T^{\mu\nu} = 0$, still holds. Using this assumption together with the field equation obtained by varying (7.1) with respect to Φ, yields[3]

$$D^\mu D_\mu\Phi = \frac{8\pi}{3+2\omega}T^\sigma{}_\sigma. \tag{7.2}$$

Variation of the action with respect to $g_{\mu\nu}$ gives the equations of motion for the gravitational field,

$$\mathcal{R}_{\mu\nu} = \frac{8\pi}{\Phi}\left(T_{\mu\nu} - \left(\frac{1+\omega}{3+2\omega}\right)g_{\mu\nu}T^\sigma{}_\sigma\right) - \frac{1}{\Phi}D_\mu D_\nu\Phi - \frac{\omega}{\Phi^2}D_\mu\Phi D_\nu\Phi. \tag{7.3}$$

When the energy-momentum tensor takes the perfect-fluid form (3.11) and the metric takes the Robertson-Walker form (3.7), so that it becomes possible to study the cosmology of the theory, the homogeneity and isotropy imply that Φ is a function of time only. The field equations then become[3]

$$\frac{d}{dt}\left(\rho R^3\right) = -p\frac{d}{dt}\left(R^3\right), \tag{7.4}$$

$$\ddot{\Phi} + 3H\dot{\Phi} = \frac{8\pi}{3 + 2\omega}(\rho - 3p), \tag{7.5}$$

$$H = -\frac{\dot{\Phi}}{2\Phi} + \left[\frac{2\omega + 3}{3}\left(\frac{\dot{\Phi}}{2\Phi}\right)^2 + \frac{8\pi}{3\Phi}\rho\right]^{1/2}, \tag{7.6}$$

where the first of these is just the usual energy-momentum conservation law (3.13), the second comes from the Φ-field equation (7.2) and the last arises from (7.3). This set of equations replaces the Friedmann equations of conventional Einstein gravity. So, as well as the effect of a changing gravitational constant $1/\Phi$, it can be seen from (7.6) that there is also a contribution to the energy density from the kinetic energy of the Φ field, i.e., terms in $\dot{\Phi}$. A general feature of these equations is that the rescalings

$$\Phi \rightarrow \chi^2 \Phi \tag{7.7}$$

and

$$t \rightarrow \chi t \tag{7.8}$$

leave the equations unchanged, a fact that will prove useful later on. There are also the two simpler cases; $\omega \rightarrow \infty$, which is the GR limit, and $p = \rho/3$, which is radiation dominance. In either case the right hand side of (7.5) is zero, i.e.,

$$\frac{d}{dt}\left(\dot{\Phi}R^3\right)^{\cdot} = 0, \tag{7.9}$$

so that at large times (assuming an expanding universe) $\dot{\Phi} \rightarrow 0$, and the usual Friedmann equations are recovered with the final value of Φ playing the rôle of $1/G$.

Because such a modification of Einstein's equation seems rather *ad hoc*, the hope is that an action like (7.1) might arise naturally as a low energy effective action of some more general theory such as super-strings or higher dimensional Kaluza-Klein theory. But this hope has not been realized in Kaluza-Klein theory despite its similarities with (7.1) on reduction to four dimensions[4]. There is still some debate about whether string theory could generate an action like (7.1) [5]. However, it has

proved possible to obtain (7.1) from a theory which is based on a broken global scale invariance[6].

The key feature of the action (7.1) is that during an era of false vacuum dominance, i.e., $\rho = -p = \rho_v$, the solutions to (7.4)–(7.6) are[7, 1]

$$\Phi(t) = \Phi(t_s)\big(1 + K(t - t_s)\big)^2 \tag{7.10}$$

$$R(t) = R(t_s)\big(1 + K(t - t_s)\big)^{\omega+1/2} = R(t_s)\left(\frac{\Phi(t)}{\Phi(t_s)}\right)^{(\omega+1/2)/2} \tag{7.11}$$

where

$$K = \sqrt{\frac{32\pi\rho_v}{(2\omega + 3)(6\omega + 5)\Phi(t_s)}} \tag{7.12}$$

and $R(t_s)$ and $\Phi(t_s)$ are the values at the beginning of inflation t_s. Note that this is not the general vacuum dominated solution of (7.4)–(7.6) but rather the large t limit. The general solution, given in [8], is dominated by the dynamics of the Φ-field at early times (as occurs in the radiation dominated solution shown in the next section). The Hubble parameter at the start of inflation is thus

$$H(t_s) = \left(\omega + \frac{1}{2}\right) K. \tag{7.13}$$

So, rather than exponential inflation the scale factor R expands as a power law. Since bubbles of true vacuum will still be filling space at an exponential rate they will be able to coalesce into a single region of true vacuum thus exiting gracefully from inflation. So, as originally envisaged for old inflation, it is the bubble wall collisions which reheat the universe and produce density inhomogeneities.

′ Brans-Dicke theory can be constrained by time-delay tests in the solar system. In order that it be sufficiently similar to GR (the limit $\omega \to \infty$), ω must be sufficiently large. The most recent constraint is $\omega \gtrsim 500$[9]. In the context of extended inflation however, it has been found that there are other constraints which can be put on ω that are due to the inflationary phase change itself[10, 11]. If the magnitude of adiabatic
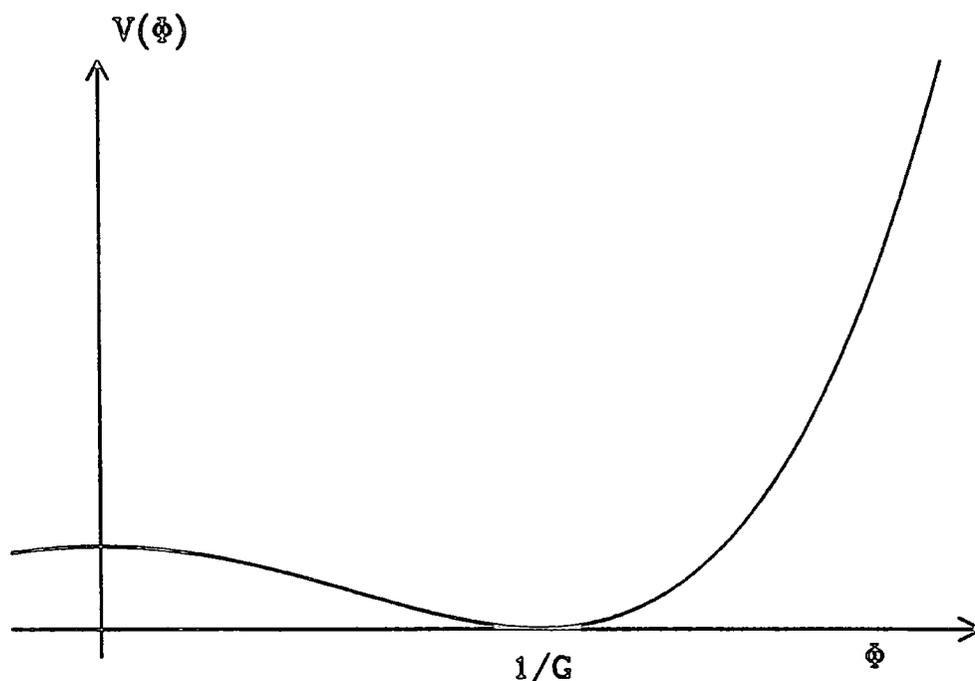
Figure 7.1: Symmetry breaking potential for $V(\Phi)$ for the induced gravity version of extended inflation. The asymmetric absolute minimum ensures $\Phi = 1/G$ at low energies whilst it can still take an arbitrary value at high energy.

density fluctuations and gravitational wave perturbations are not to conflict with the isotropy of the microwave background, then $\omega > 3/2$, while for the distribution of bubbles of normal phase generated by the subsequent phase transition to be consistent with this isotropy requires $\omega \lesssim 25$. So that it would appear that pure Brans-Dicke cosmology based on (7.1) is ruled out.

In order to avoid this dilemma there are very many extensions which have been made to (7.1), some of which are claimed to be more natural. They emulate (7.1) (with $3/2 < \omega \lesssim 25$) when the temperature is of the inflationary scale, thus reproducing the power-law inflation required to counteract the graceful exit problem, but incorporate extra features to ensure compatibility with GR at low temperatures or late times. Some examples are:

i) Extended inflation with induced gravity[12] in which a potential $V(\Phi)$ is added to (7.1). It is assumed that V has a symmetry breaking form (see fig. 7.1) so that

at low energy Φ is tied to $1/G$ but at high energy it can be arbitrary.

ii) Variable ω models[13]. Possibly the most straightforward approach. It is assumed that ω is a function of the Φ-field and $\omega(\Phi)$ is then required to be < 25 during inflation, but then becomes > 500 today.

iii) Hyperextended inflation[14], which is closely related to variable ω models, in which the Φ-field couples to the Ricci-scalar $\mathcal{R}$ in a more general way than in (7.1). The most general form of coupling would be $f(\Phi')\mathcal{R}$ where $f(\Phi') = M^2 + \xi\Phi' + \xi'\Phi'^2/M^2 + \cdots$ for small Φ' where M is some mass scale. Redefining $\Phi = f(\Phi')$ gives an action like (7.1) but where ω depends on Φ. A mechanism which ensures $\omega > 500$ today must still be added to this scenario, however.

iv) Generalized extended inflation[15]. In this model it is assumed the scalar inflatons couple to the Φ-field in a different way than to ordinary matter. It is found that in addition to the variation of H during inflation the bubble nucleation rate is also time dependent. An ω-parameter can nevertheless be defined for the inflatons which obeys $\omega_I \lesssim 20$ whilst for ordinary "visible" matter $\omega_V \gtrsim 500$ still applies.

Since it is hoped that extended inflation will proceed via a first order phase transition, the most natural way of localizing the inflaton field in its false vacuum minimum seems to be via finite temperature field theory. Such a model is thus amenable to the initial thermal state analysis described in the previous chapter. In the rest of this chapter only pure Brans-Dicke extended inflation (with $3/2 < \omega \lesssim 25$) is considered since, in order to accomplish inflation in the prescribed way, all the common variations become very similar at the temperatures around the start of inflation i.e., the period when the thermal state is required.

7.2 Evolution

To be able to follow the procedure described in chapter 6, it is required that the equations of motion can be evolved when ρ and p obtain contributions from both thermal particles and the vacuum energy, i.e.,

$$\rho = \rho_T + \rho_v \tag{7.14}$$

and

$$p = p_T - \rho_v \tag{7.15}$$

where ρ_v is defined from the symmetry breaking potential in the usual way and contributions from the thermal particles ρ_T and p_T are defined by (4.13) and (4.14) respectively. Substituting these into (7.4)–(7.6) gives

$$\dot{T} = -3H \frac{\rho_T + p_T}{\frac{d}{dT}\rho_T} \tag{7.16}$$

$$\ddot{\Phi} + 3H\dot{\Phi} = \frac{8\pi}{3 + 2\omega}(4\rho_v + \rho_T - 3p_T) \tag{7.17}$$

$$H = -\frac{\dot{\Phi}}{2\Phi} + \left[\frac{2\omega + 3}{3} \left(\frac{\dot{\Phi}}{2\Phi} \right)^2 + \frac{8\pi}{3\Phi}(\rho_v + \rho_T) \right]^{1/2} . \tag{7.18}$$

which it is possible to solve numerically on a computer. Treating $\dot{\Phi}$ as an independent variable gives, from (7.16) and (7.17), three first order differential equations in T, Φ and $\dot{\Phi}$, with H defined by (7.18). The time period of interest is again divided up into a large number of steps and this set of three coupled differential equations is integrated over each step using a Runge-Kutta method[16]. Three initial conditions must therefore be specified for the three variables at some chosen initial time t_i. One condition determines when $T = \infty$, which sets the zero of time just as in GR, but the other two are arbitrary. The natural choice is to take $\Phi(t_i)$ and $\dot{\Phi}(t_i)$ as arbitrary and set $T(t_i)$ by putting $T = \infty$ at $t = 0$. Once all of these initial conditions have been set (7.16)–(7.18) can be evolved, and the method of calculating N, by comparing Γ and H at each step, follows exactly as in section 6.1.

Again, as in section 6.1, in order to set $T(t_i)$, the initial time t_i is set well before the inflation scale so that the only contribution to the energy-momentum tensor is radiation, i.e., (4.21) and (4.22). What is required, then, is the formula which

corresponds to (4.27). Unfortunately, this proves to rather more involved in Brans-Dicke gravity. With radiation only (7.16)–(7.18) become

$$H = \frac{\dot{R}}{R} = -\frac{\dot{T}}{T} \tag{7.19}$$

$$\frac{\ddot{\Phi}}{\dot{\Phi}} = -3H \tag{7.20}$$

and

$$H = -\frac{\dot{\Phi}}{2\Phi} + \left[\frac{2\omega + 3}{3} \left(\frac{\dot{\Phi}}{2\Phi} \right)^2 + \frac{8\pi a_{SB} T^4}{6\Phi} \right]^{1/2}. \tag{7.21}$$

It is still possible to solve these exactly and the solutions will now be discussed (see also [17, 11]) before it is shown how the initial condition $T(t_i)$ is to be set.

Equating (7.19) and (7.20) and integrating gives

$$\dot{\Phi} = BT^3 \tag{7.22}$$

where B is the integration constant, and by combining (7.19) and (7.21) and using (7.22) to eliminate $\dot{\Phi}$ and $\dot{T}$ gives

$$\frac{dT}{d\Phi} = \frac{T}{2\Phi} - \left[\frac{2\omega + 3}{3} \frac{T^2}{4\Phi^2} + \frac{8\pi a_{SB}}{6\Phi B^2} \right]^{1/2}. \tag{7.23}$$

To reduce this further, define

$$x \equiv \frac{T^2}{\Phi}, \tag{7.24}$$

which allows (7.23) to be rewritten as

$$\frac{dx}{d\Phi} = -\frac{2x^{\frac{1}{2}}}{\Phi} \left[\frac{2\omega + 3}{3} \frac{x}{4} + \frac{8\pi a_{SB}}{6B^2} \right]^{1/2}. \tag{7.25}$$

Separating the variables and integrating yields

$$T = \sqrt{\frac{3}{2\omega + 3}} \frac{1}{C^{1/2}B} \Phi^{\frac{1}{2}\left(1 + \sqrt{\frac{2\omega+3}{3}}\right)} \left(C\Phi^{-\sqrt{\frac{2\omega+3}{3}}} - \frac{8\pi a_{SB}}{6}\right) \tag{7.26}$$

where C is the integration constant times B^2. (7.22) can now be used to write

$$\dot{\Phi} = \frac{1}{\alpha^3 C^{3/2} B^2} \Phi^{3(1+\alpha)/2} \left(C\Phi^{-\alpha} - A\right)^3 \tag{7.27}$$

where $\alpha \equiv \sqrt{(2\omega + 3)/3}$ and $A \equiv 8\pi a_{SB}/6$ have been defined. Separating the variables again this becomes

$$dt = \alpha^3 C^{3/2} B^2 \Phi^{-3(1+\alpha)/2} \frac{d\Phi}{\left(C\Phi^{-\alpha} - A\right)^3}. \tag{7.28}$$

In order to integrate this expression the assumption that $T \to \infty$ as $t \to 0$ is used. From (7.22) and (7.26) it can be seen that this limit implies $\Phi \to 0(\infty)$ for $\dot{\Phi} > 0$ ($\dot{\Phi} < 0$), or from (7.22), $B > 0$ ($B < 0$). So there are two cases differing only in the limits of integration:

i) $\dot{\Phi} > 0$. From (7.28)

$$t = \alpha^3 C^{3/2} B^2 \int_0^\Phi \Phi'^{-3(1+\alpha)/2} \frac{d\Phi'}{\left(C\Phi'^{-\alpha} - A\right)^3}. \tag{7.29}$$

If the substitution

$$y = \frac{C}{A\Phi'^\alpha} \tag{7.30}$$

is made, then

$$t = -\alpha^2 B^2 C^{-1/2\alpha} A^{\frac{1}{2}\left(\frac{1}{\alpha} - 3\right)} \int_{y'}^\infty \frac{y^{\frac{1}{2}\left(\frac{1}{\alpha}+1\right)} dy}{(1 - y)^3} \tag{7.31}$$

where $y' = C/A\Phi^\alpha$.

ii) $\dot{\Phi} < 0$. Similarly

$$t = \alpha^3 C^{3/2} B^2 \int_\infty^\Phi \Phi'^{-3(1+\alpha)/2} \frac{d\Phi'}{(C\Phi'^{-\alpha} - A)^3} \qquad (7.32)$$

or

$$t = -\alpha^2 B^2 C^{-1/2\alpha} A^{\frac{1}{2}(\frac{1}{\alpha}-3)} \int_0^{y'} \frac{y^{\frac{1}{2}(\frac{1}{\alpha}+1)} dy}{(1-y)^3} \qquad (7.33)$$

using the substitution (7.30).

The expressions (7.31) and (7.33) may also be written as hypergeometric functions (also known as incomplete beta functions in this case), i.e.,

$$t = \frac{2\alpha^2}{3 \mp \frac{1}{\alpha}} \frac{B^2 C^{3/2}}{A^3} \Phi^{\frac{1}{2}(3\alpha\pm1)} \, {}_2F_1\left(3, \frac{1}{2}\left(3 - \frac{1}{\alpha}\right), \frac{1}{2}\left(3 - \frac{1}{\alpha}\right); \left(\frac{A\Phi^\alpha}{C}\right)^{\pm1}\right) \qquad (7.34)$$

where ${}_2F_1$ is the hypergeometric function in the notation of [18] with the upper sign for $\dot{\Phi} > 0$ and the lower for $\dot{\Phi} < 0$.

It is possible to use (7.29) or (7.32) to derive a relation between Φ, $\dot{\Phi}$ and t. From (7.29) it can be seen that $\Phi > 0$, trivially. In the limit $\Phi \to 0$ the $C\Phi^{-\alpha}$ term dominates in the denominator of (7.29) and the integration may be performed easily. Using (7.27) to remove B gives

$$\dot{\Phi}(t) < \frac{2}{3\alpha - 1} \frac{\Phi(t)}{t}. \qquad (7.35)$$

The opposite limit applies to (7.32) and so

$$\dot{\Phi}(t) > \frac{-2}{3\alpha + 1} \frac{\Phi(t)}{t}. \qquad (7.36)$$

These limits also give the behaviour of Φ in the $\dot{\Phi}$-dominated regime. From (7.29)

or (7.32) the expression

$$\Phi = \left(\frac{3\alpha - 1}{2\alpha^3 C^{-3/2} B^2} \right)^{2/(3\alpha-1)} t^{2/(3\alpha-1)}, \qquad \dot{\Phi} > 0 \tag{7.37}$$

or

$$\Phi = \left(\frac{(3\alpha + 1)A^3}{2\alpha^3 C^{3/2} B^2} \right)^{-2/(3\alpha+1)} t^{-2/(3\alpha+1)}, \qquad \dot{\Phi} < 0 \tag{7.38}$$

is obtained. Equations (7.26) and (7.19) now give T and H, i.e.,

$$T = \left(\frac{2}{B(3\alpha - 1)} \left(\frac{3\alpha - 1}{2\alpha^3 C^{-3/2} B^2} \right)^{2/(3\alpha-1)} \right)^{1/3} t^{\frac{1-\alpha}{3\alpha-1}} \qquad \dot{\Phi} > 0, \tag{7.39}$$

$$T = \left(\frac{-2}{B(3\alpha + 1)} \left(\frac{(3\alpha + 1)A^3}{2\alpha^3 C^{3/2} B^2} \right)^{-2/(3\alpha+1)} \right)^{1/3} t^{\frac{-1-\alpha}{3\alpha+1}} \qquad \dot{\Phi} < 0 \tag{7.40}$$

and

$$H = \frac{\alpha \mp 1}{3\alpha \mp 1} \frac{1}{t}. \tag{7.41}$$

If (7.26) is examined for small temperatures ($T \rightarrow 0$) one finds that $\Phi \rightarrow (A/C)^{-1/\alpha}$ a constant, and from (7.22) $\dot{\Phi} \rightarrow 0$, both of which may already be expected from (7.9) or (7.20) if radiation is the only energy density. In this limit (7.19)–(7.21) become the usual FRW equations, (3.13) and (3.14), and the conventional results for radiation dominated expansion, (4.27), (3.22), are recovered.

Returning to the question of setting $T(t_i)$; if t_i, $\Phi(t_i)$ and $\dot{\Phi}(t_i)$ are fixed, then from (7.22) and (7.29) (or (7.32)) the integration constants B and C may be determined numerically and hence from (7.26), $T(t_i)$ can be found.

To summarize, the full evolution of (7.16)–(7.18) can have three possible periods in which different forms of the energy density dominate. They are:

i) a possible early $\dot{\Phi}$-dominance period[19] where

$$H = \left(\frac{\alpha \mp 1}{3\alpha \mp 1}\right)\frac{1}{t}, \quad T \propto t^{(-\alpha \pm 1)/(3\alpha \mp 1)} \quad \text{and} \quad \Phi \propto t^{\pm 2/(3\alpha \mp 1)} \tag{7.42}$$

where the upper sign is for the case $\dot{\Phi} > 0$ and the lower sign for $\dot{\Phi} < 0$.

ii) an era which will be called thermal particle dominance, where ρ_T and p_T dominate the right hand side of (7.17)–(7.18) and the evolution is essentially the same as for the standard FRW model. For the radiation dominance limit (4.21) this gives

$$H = \frac{1}{2t}, \quad T \propto t^{-1/2} \quad \text{and} \quad \Phi \approx const, \tag{7.43}$$

though the expressions for H and T will be modified if the masses of the particles are not negligible at the relevant energy scale.

iii) a final period when the vacuum energy dominates and power-law inflation ensues with

$$H = \left(\omega + \frac{1}{2}\right)\frac{1}{t}, \quad R \sim t^{\omega + 1/2}, \quad \text{and} \quad \Phi \sim t^2. \tag{7.44}$$

The behaviour of the temperature T of the inflatons during this period depends in general on their temperature-dependent mass m. If $m \to 0$ then from (7.16) $T \sim t^{-(\omega + 1/2)}$, whereas if $m \to \mu$ a constant it is found that $T \sim 1/\ln(t/t_s)$ where t_s is the time inflation commences. Of course, once the expansion becomes so rapid that thermal equilibrium can no longer be maintained, the effective temperature of the particles will supercool in the usual way, i.e., (4.33).

The condition (7.35) (or (7.36)) governs the times when the first two periods of evolution apply and of course, the size of ρ_v determines when the last era starts. If $|\dot{\Phi}(t_i)|$ is set close to its maximum allowed value then the $\dot{\Phi}$-dominance epoch lasts right up until the time when the vacuum energy begins to dominate. On the other hand if $\dot{\Phi}(t_i)$ is set to a negligible value then the $\dot{\Phi}$-dominance period occurs long before the period of interest and one only need consider a thermal particle dominated period before the transition to vacuum dominance.

7.3 Limits

As compared with the conventional Friedmann equations (3.14) and (3.15), their Brans-Dicke theory counterparts (7.4)-(7.6) require three extra parameters to be specified; ω, $\Phi(t_i)$ and $\dot{\Phi}(t_i)$. Since the aim here is to determine bounds on the parameters of the inflaton potential such that a thermal state is just possible, Brans-Dicke parameters which maximise the value of the total number of interactions N are needed. Only a rough estimate of the reaction rate Γ is needed to understand how it will vary as the Brans-Dicke parameters are modified. The approximation (6.6) is thus used with the mass of the inflatons set at $\rho_v^{1/4}$. Fig. 7.2 is a typical graph of the variation of Γ, H, T and Φ during the three possible epochs of evolution described in the previous section. The three Brans-Dicke parameters will now be considered in turn, in an effort to maximise the value of N (given by (4.36)).

The difference between $\dot{\Phi} > 0$ and $\dot{\Phi} < 0$ is shown in fig. 7.3 for a large value of $|\dot{\Phi}(t_i)|$, but with Φ constrained to end up at the same value at late times. So, for a given final value of Φ (in pure Brans-Dicke theory this must be $1/G$), it is clear that this choice makes little difference, though the positive value produces a slightly greater N. Fig. 7.4 shows the effect of varying (positive) $\dot{\Phi}(t_i)$. A small initial value means that the evolution starts (at t_i) in a normal radiation dominance period which is succeeded by the inflation era. With $\dot{\Phi}(t_i)$ near to its maximum value (7.35), however, it is a $\dot{\Phi}$-dominance era which precedes inflation. An intermediate initial value of $\dot{\Phi}(t_i)$ gives all three epochs in sequence. As may be suspected from this graph, N cannot, in fact, be enhanced significantly, which is clearly seen from a plot of its variation as $\dot{\Phi}(t_i)$ is varied over its whole range, fig. 7.5. The reason is that as $\dot{\Phi}(t_i)$ increases towards its maximum value (7.35) (or (7.36)), $T(t_i)$ decreases and the pre-inflation $T(t)$ curve flattens (see fig. 7.4), i.e., T changes from radiation dominance (7.43) to $\dot{\Phi}$-dominance (7.42). The effect of this is to push the final crossing point of Γ and H to a later time which increases N slightly. Eventually though, as $\dot{\Phi}(t_i)$ is increased even closer towards its maximum (7.35), the temperature becomes so low that Γ does not cross H at all and N drops to zero.

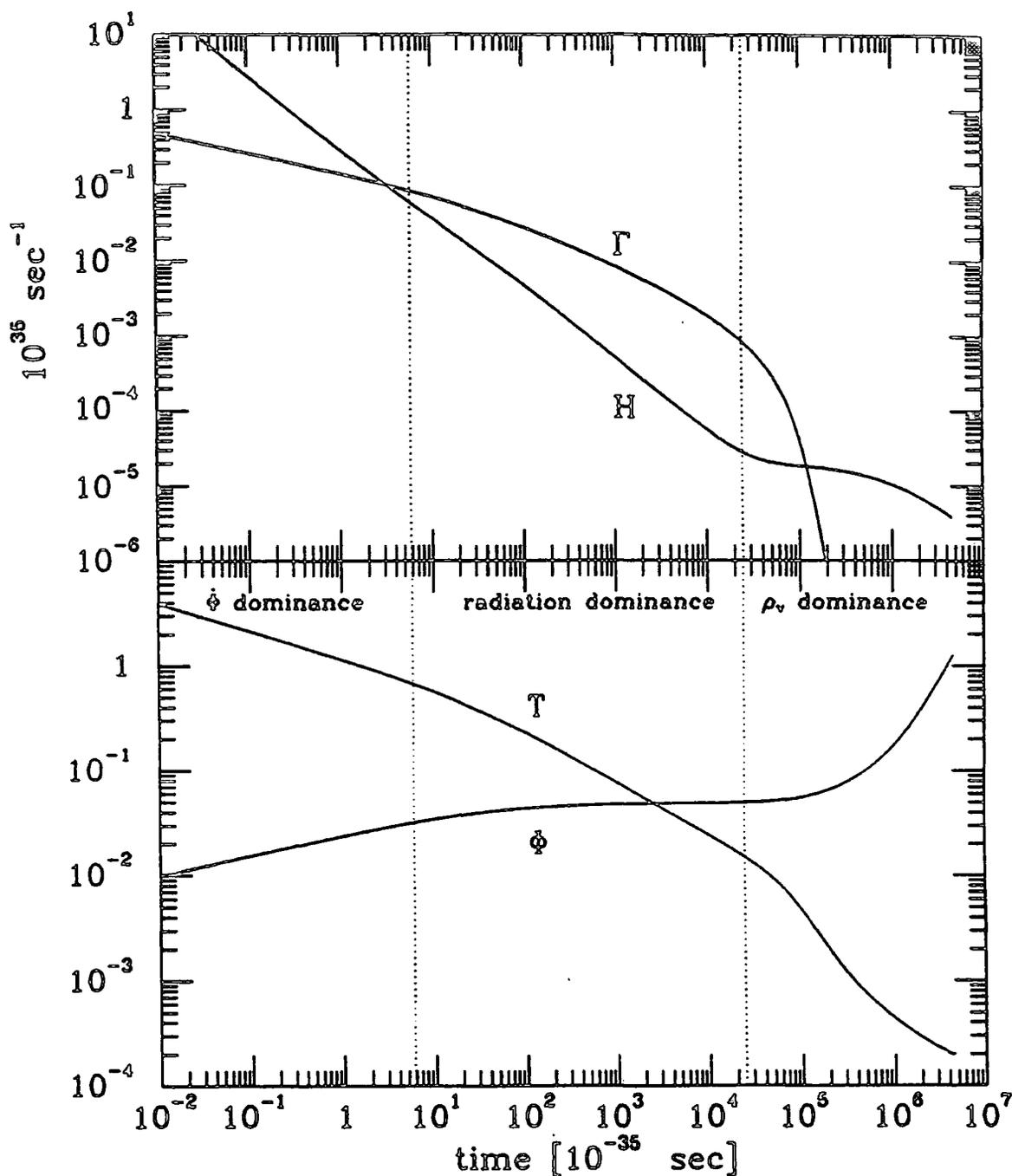As could perhaps have been predicted, N is far more dependent on the other

Figure 7.2: A plot of the expansion rate H, reaction rate Γ, temperature T and Brans-Dicke field Φ as functions of time with $\omega = 25$ and $\rho_v^{1/4} = 10^{12}$ GeV. Γ is calculated from the approximate expression (6.6) with a cross section derived from just the four point coupling λ (= 1). The vertical dotted lines divide successive regions of Φ-dominance, thermal particle dominance and vacuum dominance. The scales for the lower graph are 10^{14} GeV for T and $1/G$ for Φ.
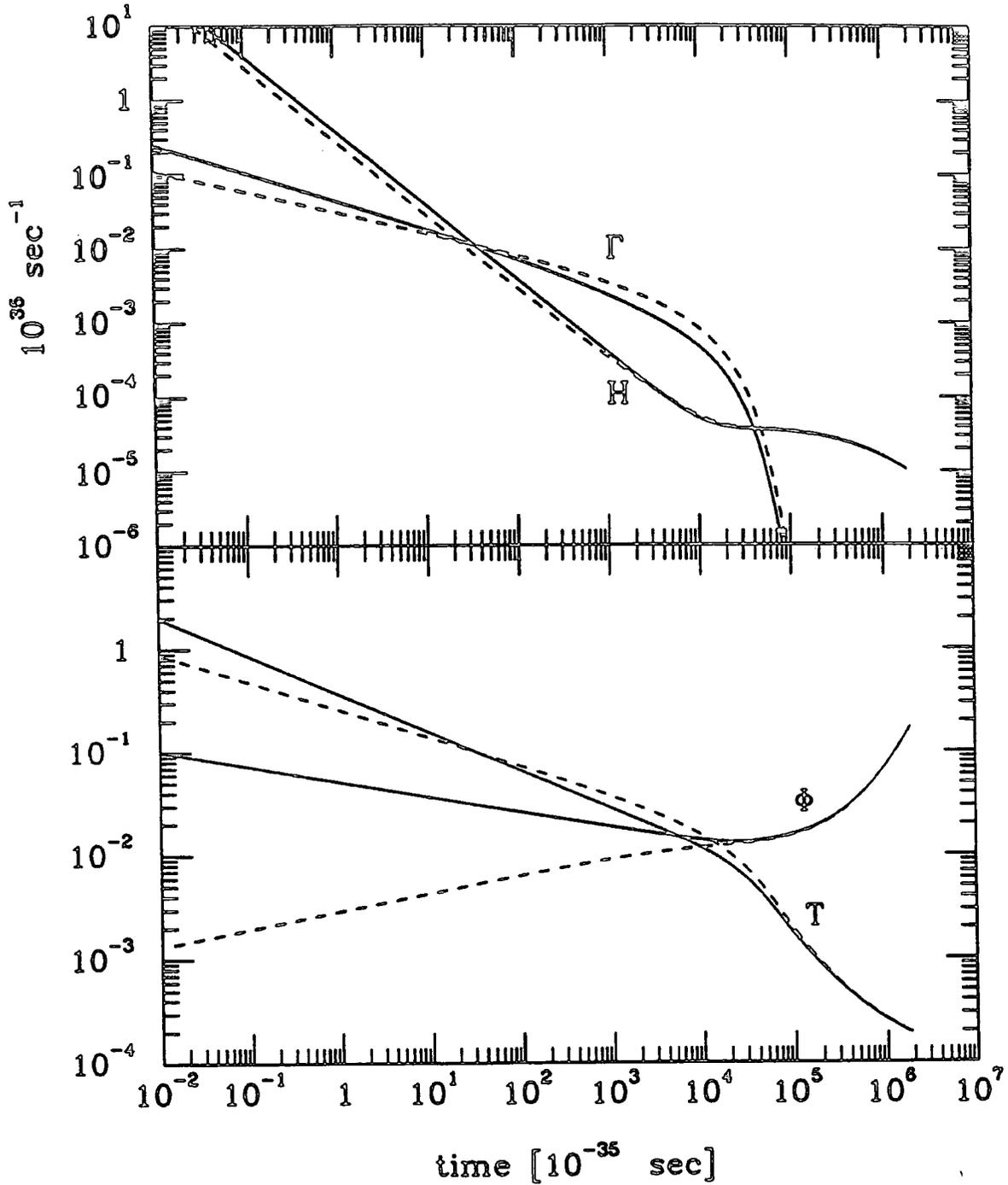
Figure 7.3: A plot of the expansion rate H, reaction rate Γ, temperature T and Brans-Dicke field Φ as functions of time (with $\omega = 25$ and $\rho_v^{1/4} = 10^{12}$ GeV) showing the difference between $\dot{\Phi}(t_i)$ negative (solid) and $\dot{\Phi}(t_i)$ positive (dashed) but with Φ constrained to end up at the same value. The scales for the lower graph are 10^{14} GeV for T and $1/G$ for Φ.
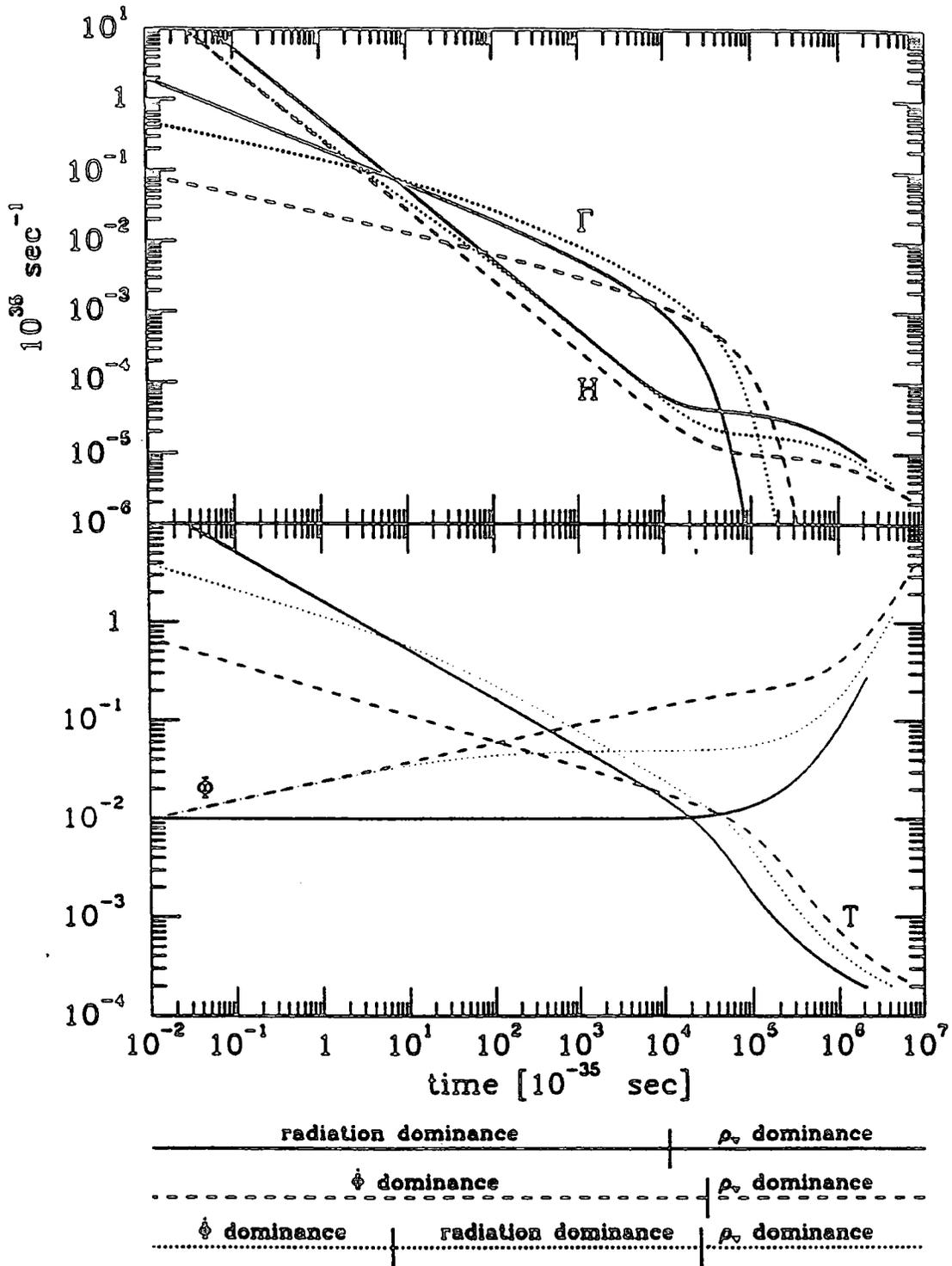
Figure 7.4: A plot of the expansion rate H, reaction rate Γ, temperature T and Brans-Dicke field Φ as functions of time (with $\omega = 25$ and $\rho_v^{1/4} = 10^{12}$ GeV) showing the effect of varying $\dot{\Phi}(t_i)$. The solid curves are for a negligible value of $\dot{\Phi}(t_i)$, the dashed curves for a value the near the maximum (7.35) and the dotted curve for an intermediate value. Below the diagram the periods where the different forms of energy density dominate for each set of curves is shown. The scales for the lower graph are 10^{14} GeV for T and $1/G$ for Φ.
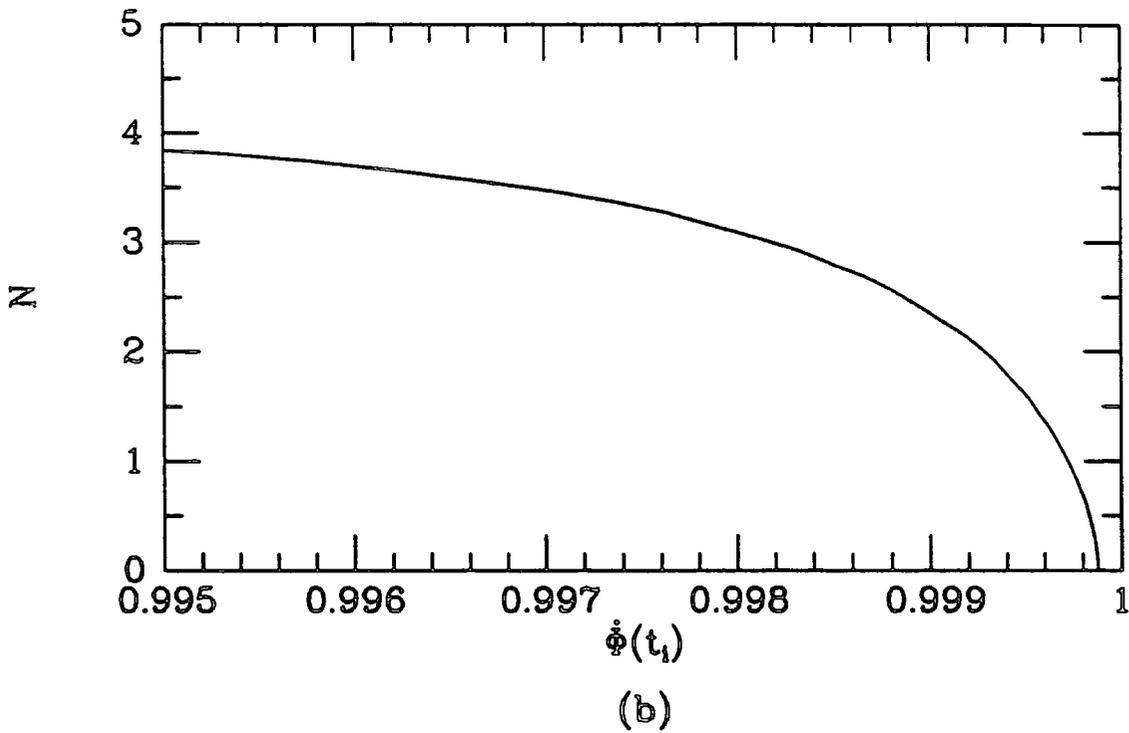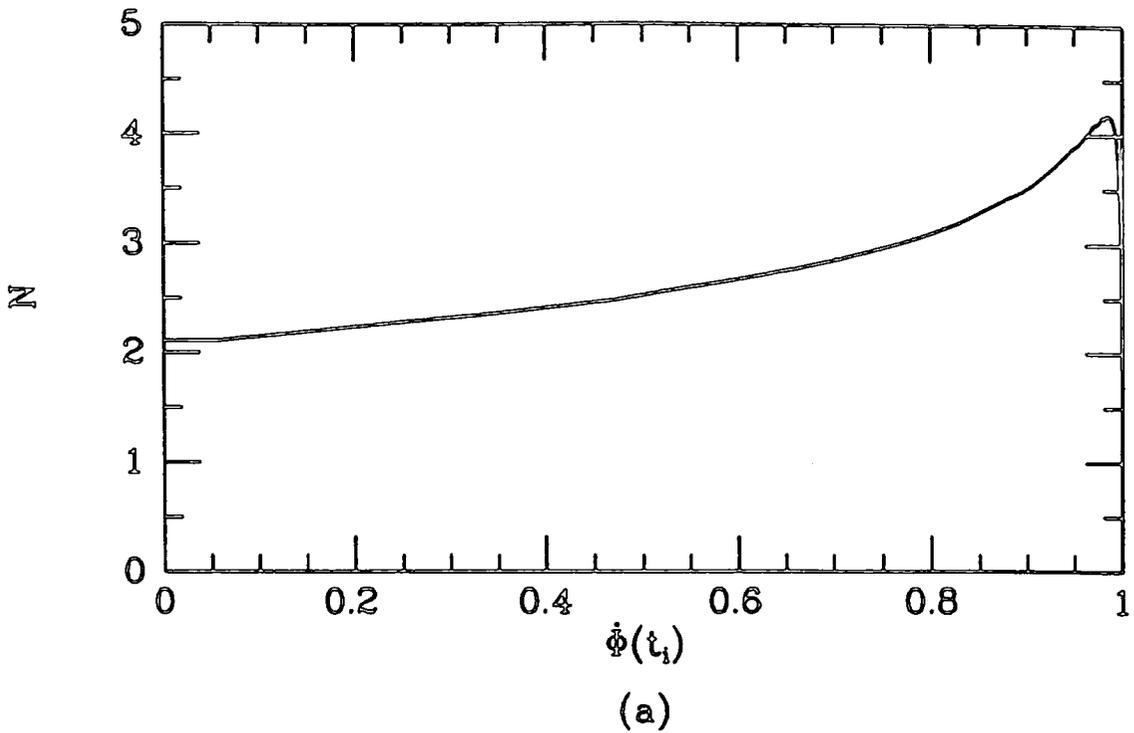
Figure 7.5: (a) The variation of the total number of interactions N with the starting value $\dot{\Phi}(t_i)$ for $\Phi(t_i) = 1/G$, $\omega = 20$, $\rho_v^{1/4} = 10^{12}$ GeV. The units for $\dot{\Phi}(t_i)$ are scaled such that the limiting value defined by (7.35) $= 1$. The lower graph (b) is an expanded view near this limit showing how N drops rapidly to zero. Note that the position of the maximum value of N will vary for different choices of $\Phi(t_i)$ and ω.

initial condition $\Phi(t_i)$. Unfortunately, this is also the most difficult parameter to put bounds on. In the pure Brans-Dicke model it is required that $\Phi \approx 1/G$ at the end of the inflationary era because Φ does not vary much during the subsequent radiation and matter dominated epochs which last until today. Inflation is required to last long enough for the whole of the presently observable universe to have been contained inside a single causal domain at the start of inflation, i.e., $R(t_s)/R(t_0) < H^{-1}(t_s)/H^{-1}(t_0)$, to solve the horizon problem, and during this time Φ increases as t^2, (7.10). If it is assumed that conventional expansion has occurred from the end of inflation t_e, until today, this implies $R(t_e)/R(t_0) = T_0/T_e$, where, assuming reheating is reasonably efficient, $T_e \sim M_F \equiv \rho_v^{1/4}$, the symmetry breaking scale. So from (7.11) and (7.13) [10]

$$\left(\frac{\Phi(t_e)}{\Phi(t_s)}\right)^{(\omega+1/2)/2} = \frac{R(t_e)}{R(t_s)} = \frac{R(t_e)}{R(t_0)}\frac{R(t_0)}{R(t_s)} > \frac{T_0}{M_F}\frac{H^{-1}(t_0)}{H^{-1}(t_s)} \sim \frac{M_F}{T_0}(G\Phi(t_s))^{-1/2} \tag{7.45}$$

where the results $H(t_s) \sim M_F^2/\Phi(t_s)$, from (7.10), and $H(t_0) \sim T_0^2/G^{1/2}$, for the CMBR photons, have been used in the last step. Rearranging this gives a maximum value for Φ at the start of inflation

$$\Phi(t_s) = \Phi_m \equiv \frac{1}{G}\left(\frac{T_0}{M_F}\right)^{2/(\omega-1/2)}. \tag{7.46}$$

The value of Φ_m defined in (7.46) also determines the maximum possible value of $\Phi(t_i)$ (i.e., $\Phi_{max}(t_i) = \Phi_m$) because if the evolution starts in the radiation dominance epoch then Φ stays almost constant. From (7.46), selecting for example $M_F = 10^{12}$ GeV, then $\Phi_{max}(t_i) = \Phi_m = 10^{-2} \times 1/G$. This is a conservative choice of $\Phi_{max}(t_i)$ for GUT theories (where $M_F \sim 10^{14}$ GeV) and is probably still an overestimate for $M_F = 10^{12}$ GeV since the value of $\dot{\Phi}(t_i)$ that gives the maximum N usually entails some $\dot{\Phi}$-dominated evolution. The only possibility for $\Phi(t_i)$ to be increased much above $10^{-2} \times 1/G$ is if M_F were reduced significantly below 10^{10} GeV. Thus in the following sections, $\Phi_{max}(t_i) = 10^{-2} \times 1/G$ is taken.

Unfortunately, the more general models described in section 7.1 do not necessarily obey constraints like (7.46), the obvious example being the induced gravity extension

which can have an arbitrary starting value for Φ. However, the only scale involved in the Φ-potential (see fig. 7.1) is provided by G, so it would seem unlikely that the value of Φ will exceed $1/G$ by more than a few orders of magnitude. The possibility that Φ may start a few orders of magnitude larger than $\Phi_{max}(t_i)$ will be considered in later sections. (Note that smaller starting values would only suppress N and make thermal equilibrium less likely.)

The relation (7.46) also enables the value of ω which gives the maximum N to be determined. As noted in section 7.1, to ensure that extended inflation does not conflict with observations of the microwave background, the value of ω must be constrained to lie in the region $3/2 < \omega < 25$. Fig. 7.6 demonstrates the effect of varying ω within this range with (7.46) providing the starting value $\Phi(t_i)$, which depends very strongly on ω. Also displayed is the observational limit from solar system experiments for pure Brans-Dicke gravity, $\omega = 500$. In figure 7.7 the variation of N with ω is shown and it is clear that $\omega = 25$ gives the maximum value.

The scaling behaviour mentioned in section 7.1, (7.7) and (7.8), is illustrated in fig. 7.8. Suppose the initial value of $\Phi(t_i)$ is scaled by a factor χ^2, the behaviour of (7.4)–(7.6) is essentially unchanged except that t is also scaled by a factor χ. In other words the evolution of the temperature T (and Φ and $\dot{\Phi}$ too) is similar but shifted in time by χ. Because H has units of t^{-1} in (7.4)–(7.6), its magnitude is scaled by χ^{-1} relative to the scaled temperature. This can be seen more clearly in fig. 7.9 where the t-axis has been scaled so that the temperatures coincide. Since Γ is approximately linearly dependent on T (see e.g., (6.19)) it will scale in approximately the same way as T and, as is obvious from fig. 7.9, an increase in $\Phi(t_i)$ increases N also. Numerical investigation reveals that, for large N, an increase of two orders of magnitude in $\Phi(t_i)$ results in an increase of at least an order of magnitude in N. The reason is that, with H and Γ curves which produce a large value of N, the effect of scaling $\Phi(t_i)$ by χ^2 is to shift the crossing point where Γ becomes greater than H by a factor $\chi^{p/(1-p)}$, where $p = -1/2$ for radiation dominance and $p = (-\alpha \pm 1)/(3\alpha \mp 1)$ for $\dot{\Phi}$-dominance. If the point where H again starts to dominate over Γ is assumed to occur well into the inflation period, then this second crossing point occurs more than χ times later. The region of integration is thus scaled by at least χ and, since the shape of Γ is almost
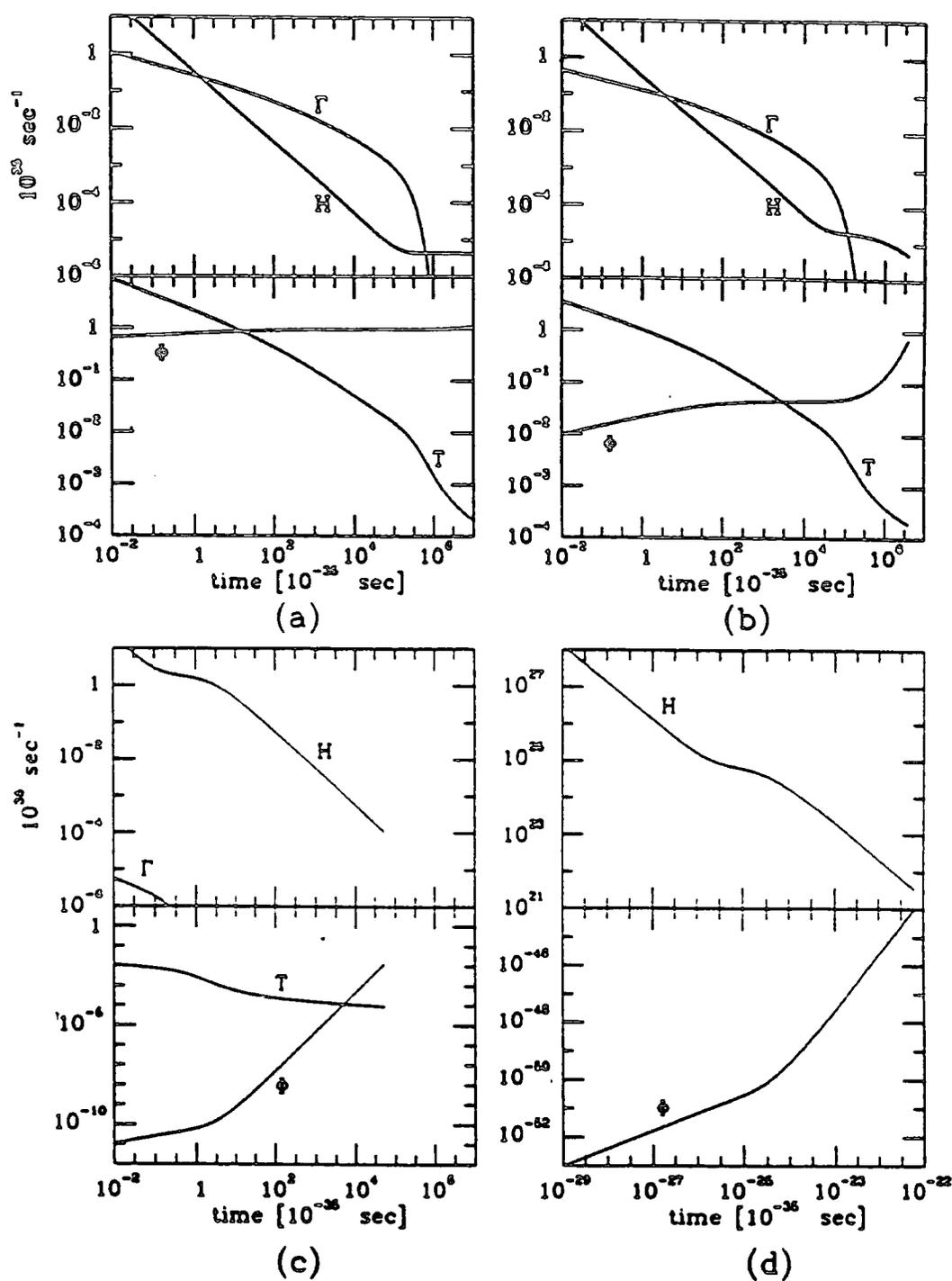
Figure 7.6: Plot of the expansion rate H, reaction rate Γ, temperature T and Brans-Dicke field Φ as functions of time (with $\rho_v^{1/4} = 10^{12}$ GeV) showing the effect of modifying the ω-parameter. (a) shows $\omega = 500$ the GR limit derived from solar system observations while in (b) $\omega = 25$, (c) $\omega = 5$ and (d) $\omega = 3/2$, values which span the range compatible with the isotropy of the microwave background. In each case the starting value for $\Phi(t_i)$ is determined from (7.46). The scales for the lower graph are 10^{14} GeV for T and $1/G$ for Φ. Note that in (d) Γ and T are unable to be displayed due to the choice of ρ_v — small values of ω ($= 3/2$) can only work with very small values of $\rho_v^{1/4}$, i.e., $\lesssim 10^2$ GeV.
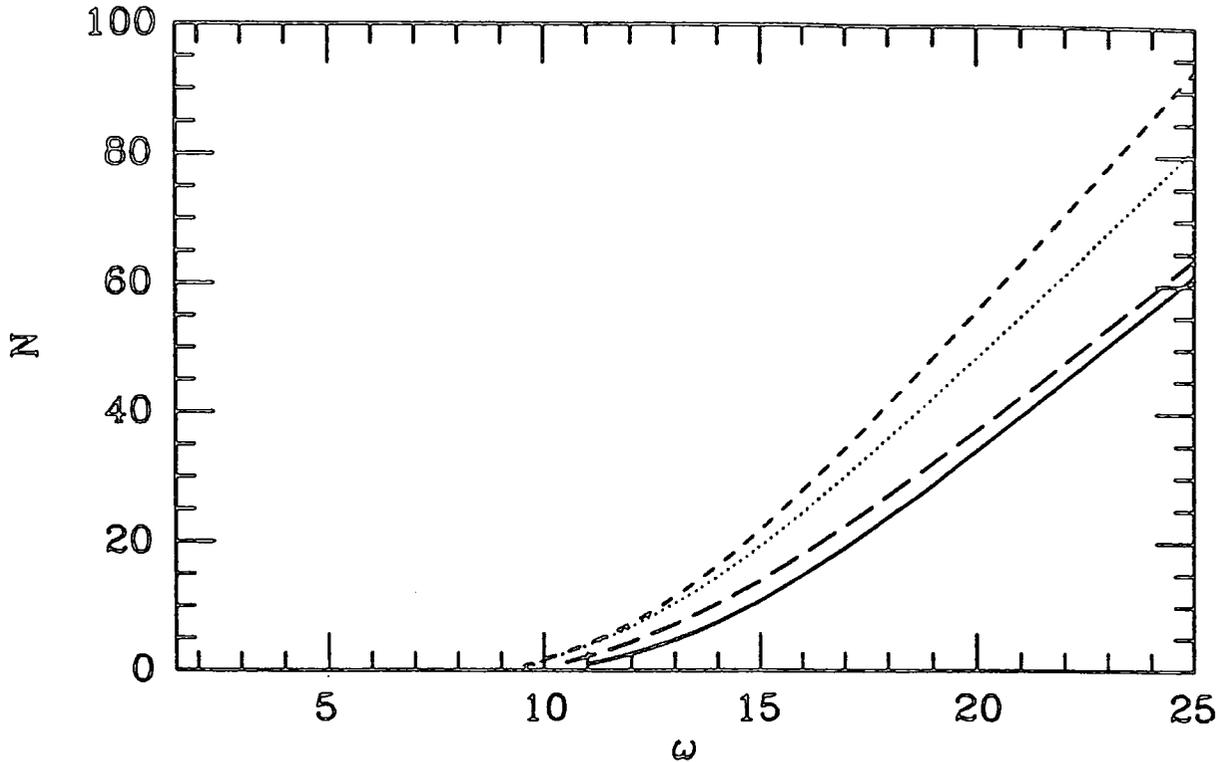
Figure 7.7: A plot of the total number of reactions N versus ω in the range $3/2 < \omega < 25$ for several different values of $\dot{\Phi}(t_i)$ spanning the peak near the limiting value illustrated in fig. 7.5(a). The plots are the limiting value (7.35) multiplied by 0.99 (solid), 0.999 (dotted), 0.9999 (short dashes, which gives the maximum value of N in this case) and 0.99999 (long dashes). $\Phi(t_i)$ is determined at each ω by (7.46).

unchanged, N is expected to be χ times greater too. This property can be used to extrapolate approximate limit values for the inflaton potential parameters which will form a thermal state for arbitrarily large values of $\Phi(t_i)$, as will be demonstrated in the following sections.

7.4   First Order Phase Change

It has already been noted that one of the prime motivations for considering extended inflation is to resurrect the old-inflation idea[1] — the power-law expansion during inflation allowing the region of new phase to percolate, thereby evading the graceful exit problem. It is worthwhile therefore, to test the scalar particle potential (6.30) in extended inflation. The vacuum energy ρ_v given by (6.32) and (6.33), the mass (6.31), and the cross section (6.34) are all defined as before. The analysis
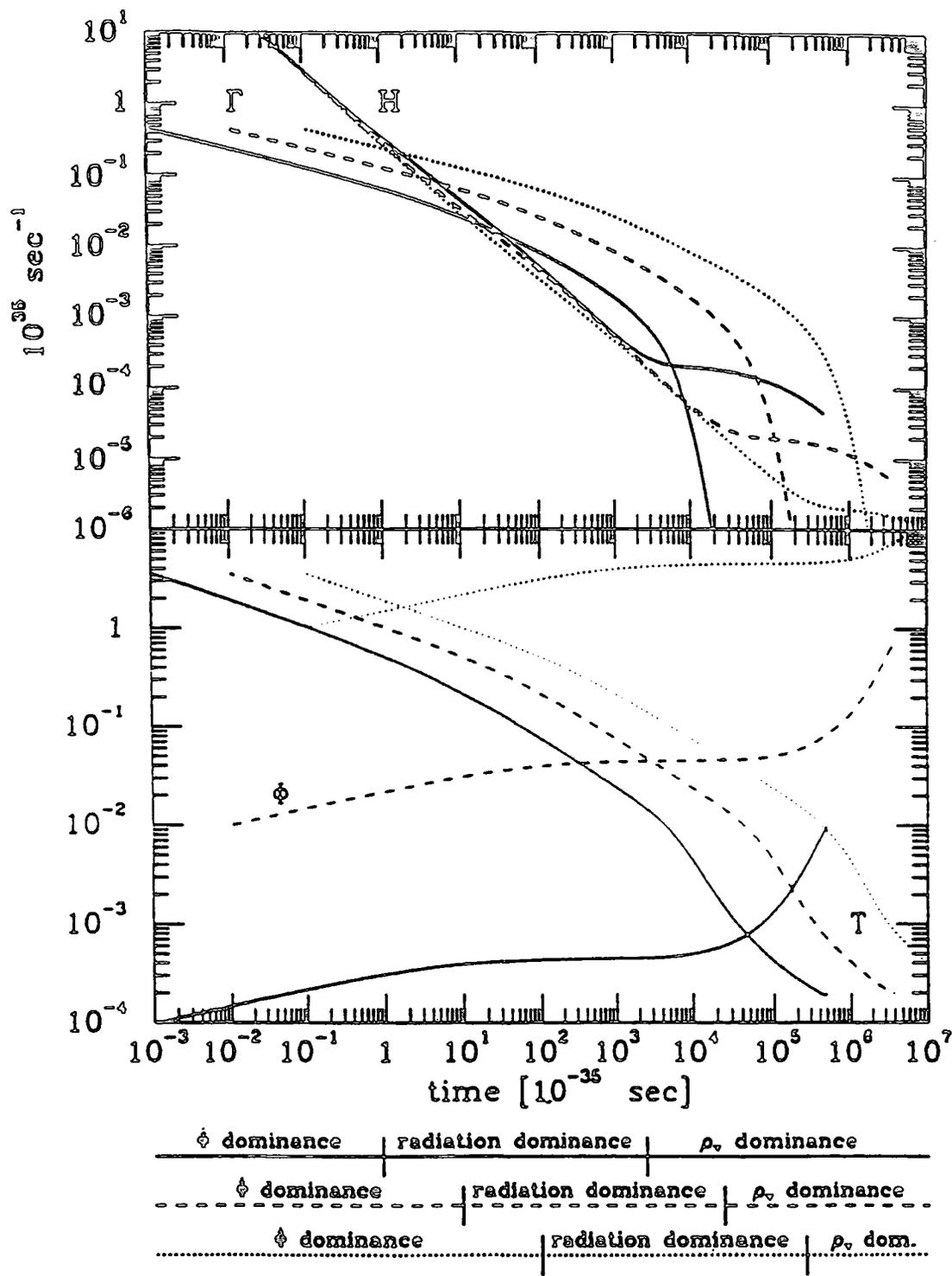
Figure 7.8: A plot of the expansion rate H, reaction rate Γ, temperature T and Brans-Dicke field Φ as functions of time (with $\omega = 25$ and $\rho_v^{1/4} = 10^{12}$ GeV) showing the scaling behaviour described at the end of section 7.3. The dotted curves (which give $N = 790$) have $\Phi(t_i)$ set at 100 times the value for the dashed curves ($N = 78$) which is in turn set at 100 times the value for the solid curves ($N = 6.6$). Below the diagram the periods where the different forms of energy density dominate for each set of curves are shown. The scales for the lower graph are 10^{14} GeV for T and $1/G$ for Φ.
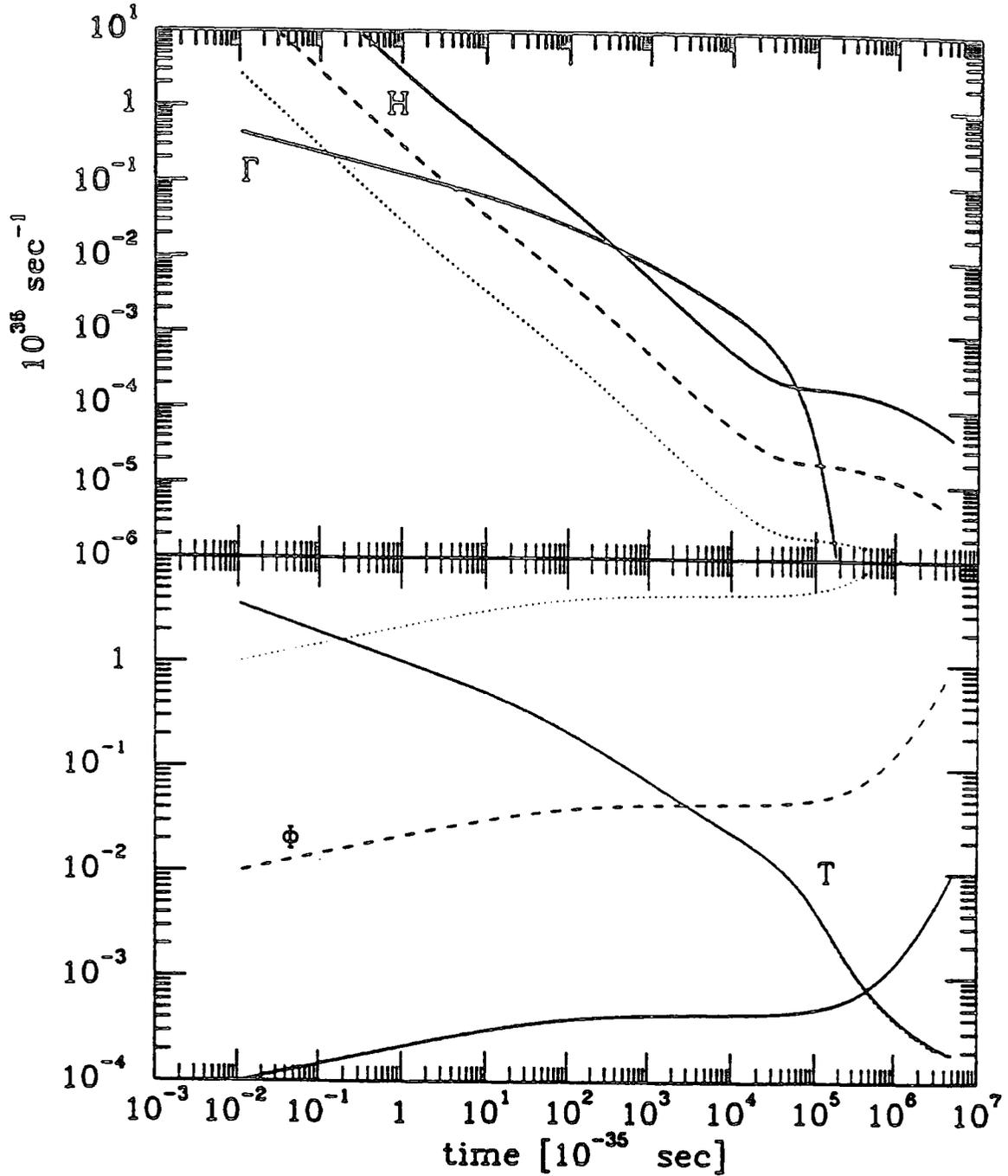
Figure 7.9: A plot of the expansion rate H, reaction rate Γ, temperature T and Brans-Dicke field Φ as functions of time (with $\omega = 25$ and $\rho_v^{1/4} = 10^{12}$ GeV) illustrating the scaling behaviour described at the end of section 7.3. It is the same plot as fig. 7.8 but the time axis for each case has been scaled so that the temperatures of all three coincide. The dotted curves (which give $N = 790$) have $\Phi(t_i)$ set at 100 times the value for the dashed curves ($N = 78$) which is in turn set at 100 times the value for the solid curves ($N = 6.6$). Below the diagram the periods where the different forms of energy density dominate for each set of curves are shown. The scales for the lower graph are 10^{14} GeV for T and $1/G$ for Φ.

proceeds in parallel with section 6.4 except that now the evolution of H and T are governed by (7.16)–(7.18) rather than the Friedmann equations.

Once again the value N calculated for given values of λ, δ and μ determines whether a thermal state is likely and, for the purposes of deriving bounds on these parameters, the condition $N > 1$ is appropriate. For pure Brans-Dicke theory, the $N = 1$ boundary lines in λ-δ space are displayed in fig. 7.10 for different values of μ using the most conducive values of the Brans-Dicke parameters which were determined in section 7.3, i.e., $\omega = 25$, $\Phi(t_i) = \Phi_{\max}(t_i) = 10^{-2} \times 1/G$ and $\dot{\Phi}(t_i)$ positive and selected to maximise N. Fig. 7.11 is similar to fig. 7.10 but with $\Phi(t_i) = 10^4 \times \Phi_{\max}(t_i)$, a value which seems well beyond the maximum likely starting point for the induced gravity version of extended inflation unless some special mechanism is introduced to generate unnaturally large starting values.

The consequences of even larger values of $\Phi(t_i)$ can be estimated by appealing to the scaling argument of the previous section. Thus, if the complete set of contours for a single value of μ is considered (see for example fig. 7.12), scaling says that an increase of $\Phi(t_i)$ by two orders of magnitude would increase N by one order of magnitude. The effect is therefore that the $N = 1000$ level contour for the larger $\Phi(t_i)$ will be at the position of the old $N = 100$ level contour, and so on. This enables one to deduce an approximate position for the new $N = 1$ level contour.

The contour plots in figs. 7.10 and 7.11 are very similar to that obtained in section 6.4. Evidently it is slightly more difficult to obtain a thermal state in pure Brans-Dicke extended inflation, though this problem is lessened if larger values of $\Phi(t_i)$ are used, as perhaps is allowed in, for example, the induced gravity generalisation of extended inflation. It must be concluded that inflation through a GUT-scale phase transition cannot be achieved by this model. To generate a thermal state a symmetry breaking scale $\lesssim 10^{12}$ GeV is necessary. Again this is suggestive of a broken SUSY model, or even electro-weak symmetry breaking at $M_F = 10^2$ GeV, assuming baryogenesis can occur at such a low scale.
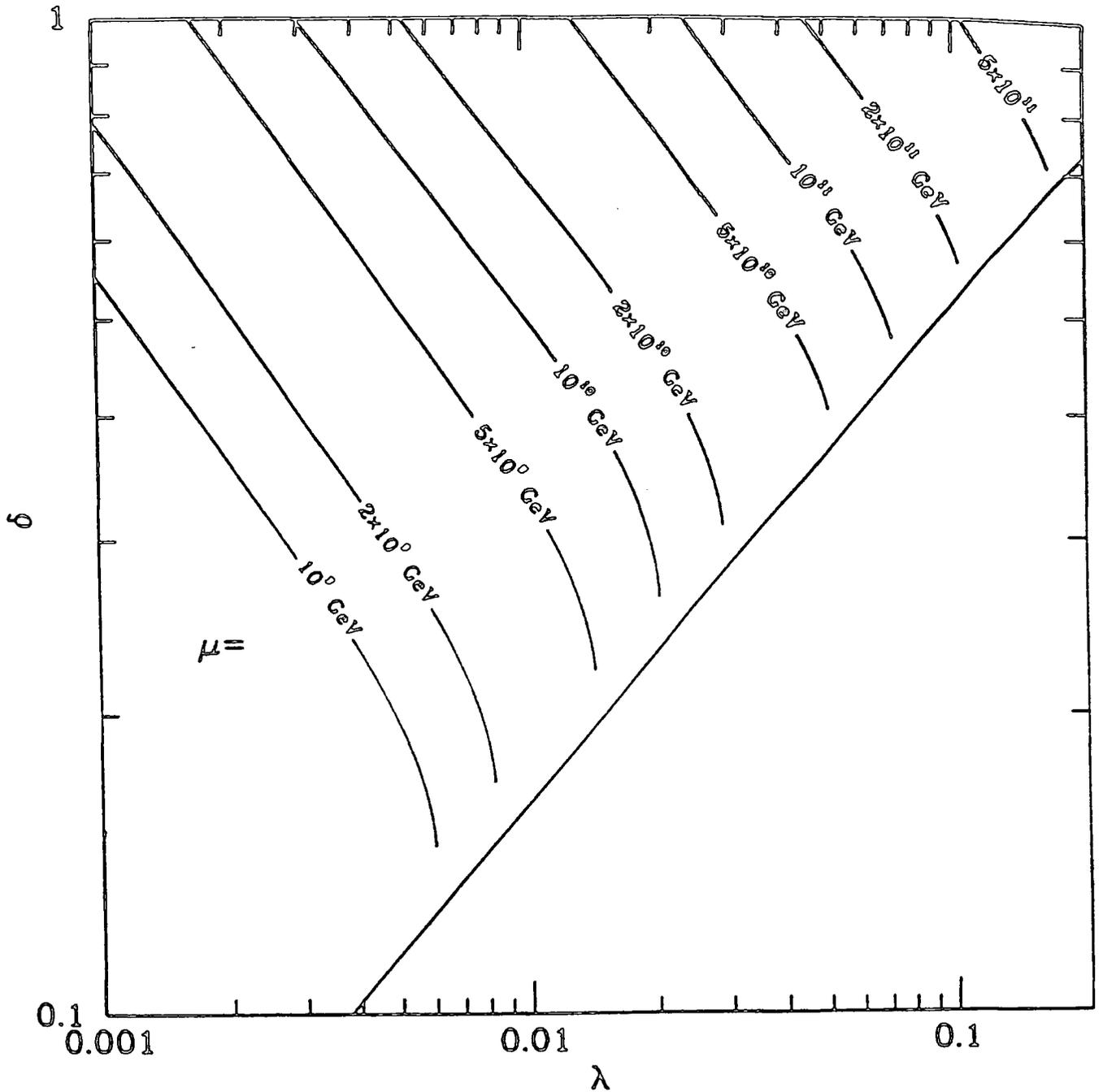
Figure 7.10: A plot of the $N = 1$ contours for the indicated values of μ in δ-λ space for the model based on first order potential (6.30) with $\Phi(t_i) = 10^{-2} \times 1/G$, i.e., the maximum possible for pure Brans-Dicke extended inflation (the other Brans-Dicke parameters, $\dot{\Phi}(t_i)$ and ω are set such that they maximise N for the given value of λ, δ and μ). Regions to the left of this contour cannot form a thermal state for the given value of μ because insufficient interactions occur. The lower-right region is excluded because here the parameters do not give rise to a suitable zero-temperature symmetry-breaking potential.
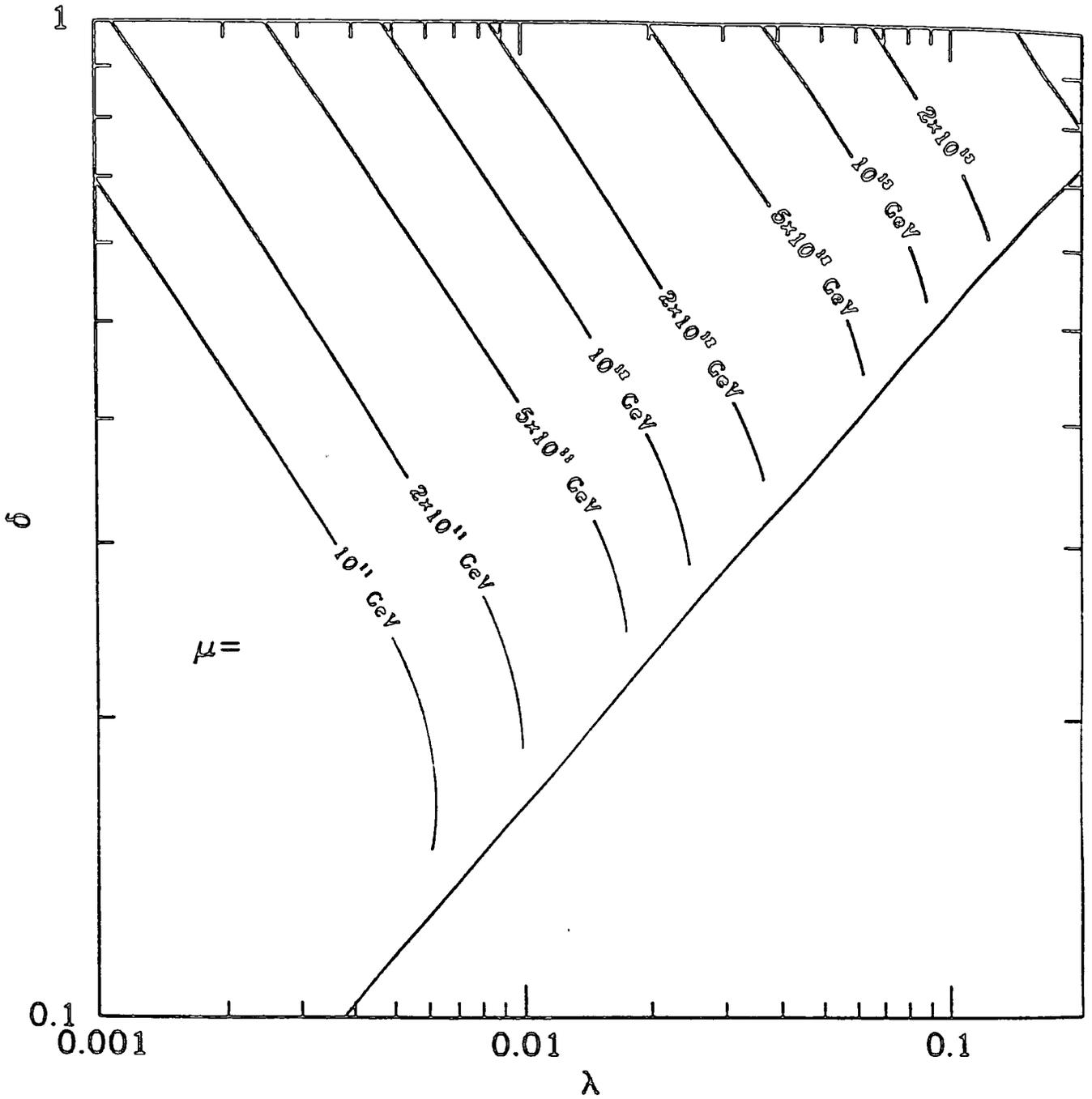
Figure 7.11: A plot of the $N = 1$ contours for various values of μ in δ-λ space for the model based on the first order potential (6.30) with $\Phi(t_i) = 10^2 \times 1/G$ (the other Brans-Dicke parameters, $\dot{\Phi}(t_i)$ and ω are set such that they maximise N for the given value of λ, δ and μ). Regions to the left of this contour cannot form a thermal state for the given value of μ because insufficient interactions occur. The lower-right region is excluded because here the parameters do not give rise to a suitable zero-temperature symmetry-breaking potential.
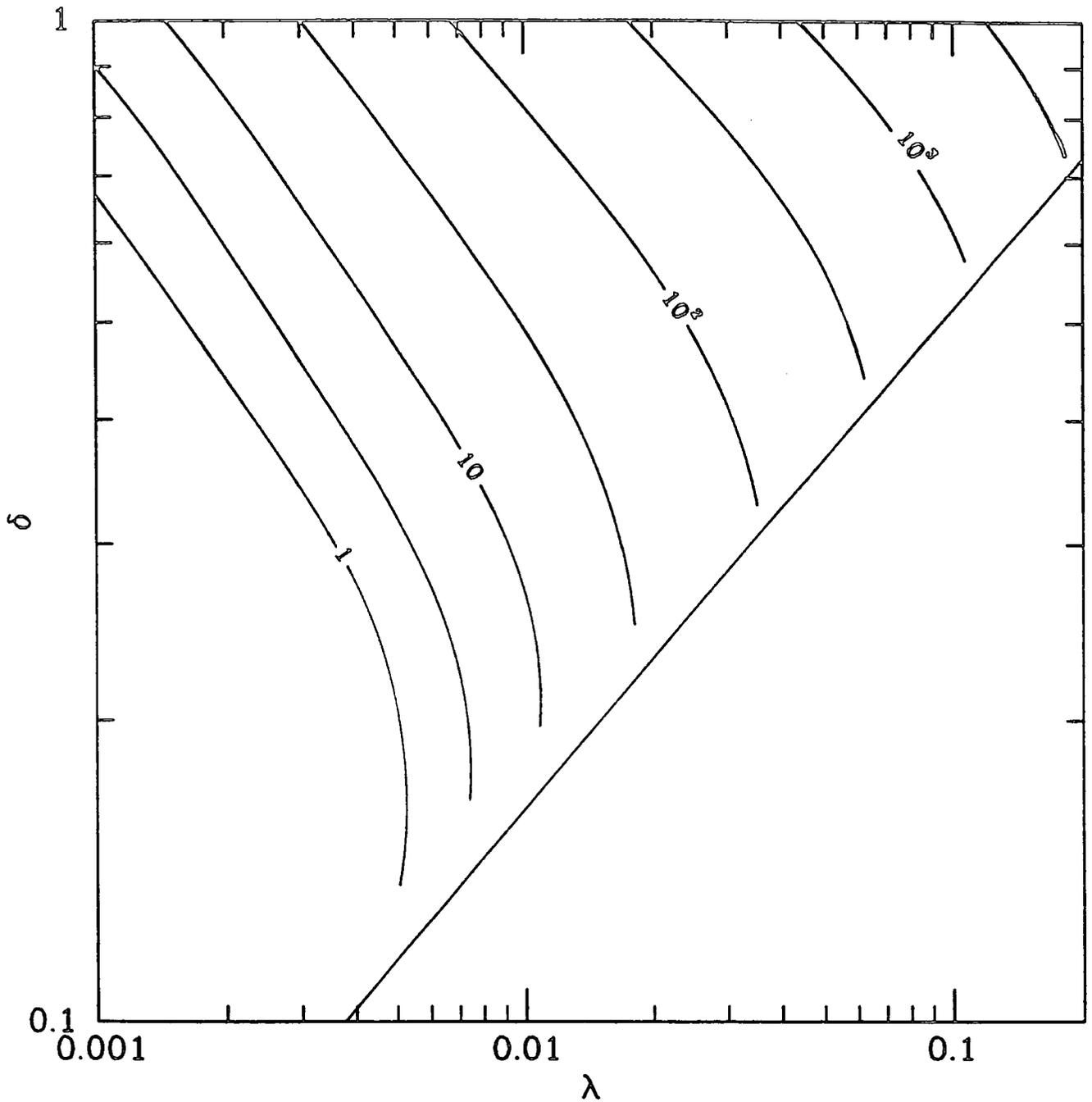
Figure 7.12: A plot of the complete set of contours of constant N with $\mu = 10^{11}$ in δ-λ space for the model based on the first order potential (6.30) with $\Phi(t_i) = 10^2 \times 1/G$ (the other Brans-Dicke parameters, $\dot{\Phi}(t_i)$ and ω are set such that they maximise N for the given value of λ, δ and μ). The intermediate contours are at half orders of magnitude (i.e., $\sqrt{10}$ times the previous contour level). The set of contours that would result if $\Phi(t_i)$ were set 100 times greater may be estimated by moving the $N = 100$ level contour to the position of the $N = 10$ level contour, and so on. The lower-right region is excluded because here the parameters do not give rise to a suitable zero-temperature symmetry-breaking potential.

7.5 Coleman-Weinberg Potential

The effect of coupling gauge particles to the inflatons can be considered in extended inflation by using the Coleman-Weinberg potential of section 6.3. Of course, now that the universe can exit gracefully from inflation by the percolation of bubbles of true vacuum, the new-inflation type limits on the couplings, which arise from the requirement that the entire observable universe must be inside a single bubble, no longer apply.

The finite-temperature effective potential used is, however, exactly the same as that in section 6.3, i.e., (6.21). The vacuum energy density ρ_v, mass and cross section are hence just given by (6.22), (6.23) and (6.28) respectively. The contour plot of N in e-ϕ_{min} space for a pure Brans-Dicke model with $\Phi(t_i) = \Phi_{max}(t_i)$ and the other two parameters set as in section 7.3, is given in fig. 7.13. A similar plot is provided in fig. 7.14 but with $\Phi(t_i) = 10^4 \times \Phi_{max}(t_i)$.

The scaling procedure may again be used here on figs. 7.13 and 7.14 (since they are complete sets of contours) to consider yet larger values of $\Phi(t_i)$. The contour plots are very similar to those obtained from conventional exponential inflation (see fig. 6.7). For a ϕ_{min} at the GUT scale a coupling of $e \gtrsim 0.3$ is needed to generate a thermal state. Such a coupling is at variance with the small couplings required for a slow-rolling transition[20, 21] and furthermore, it seems unlikely that ϕ would be able to stay localized at $\phi = 0$ for a sufficient length of time[21]. It is again concluded that for the pure Brans-Dicke case it is generally slightly harder to generate a thermal state than was found in section 6.3, but it becomes somewhat easier if $\Phi(t_i)$ is greatly increased.

7.6 Summary

There is no doubt that extended inflation has been successful in re-establishing the first-order phase transition as a viable mechanism for achieving inflation by eradicating the graceful exit problem. However for such a phase change to be possible a thermal state must be established and hence the sort of analysis described in chapter six may be employed.
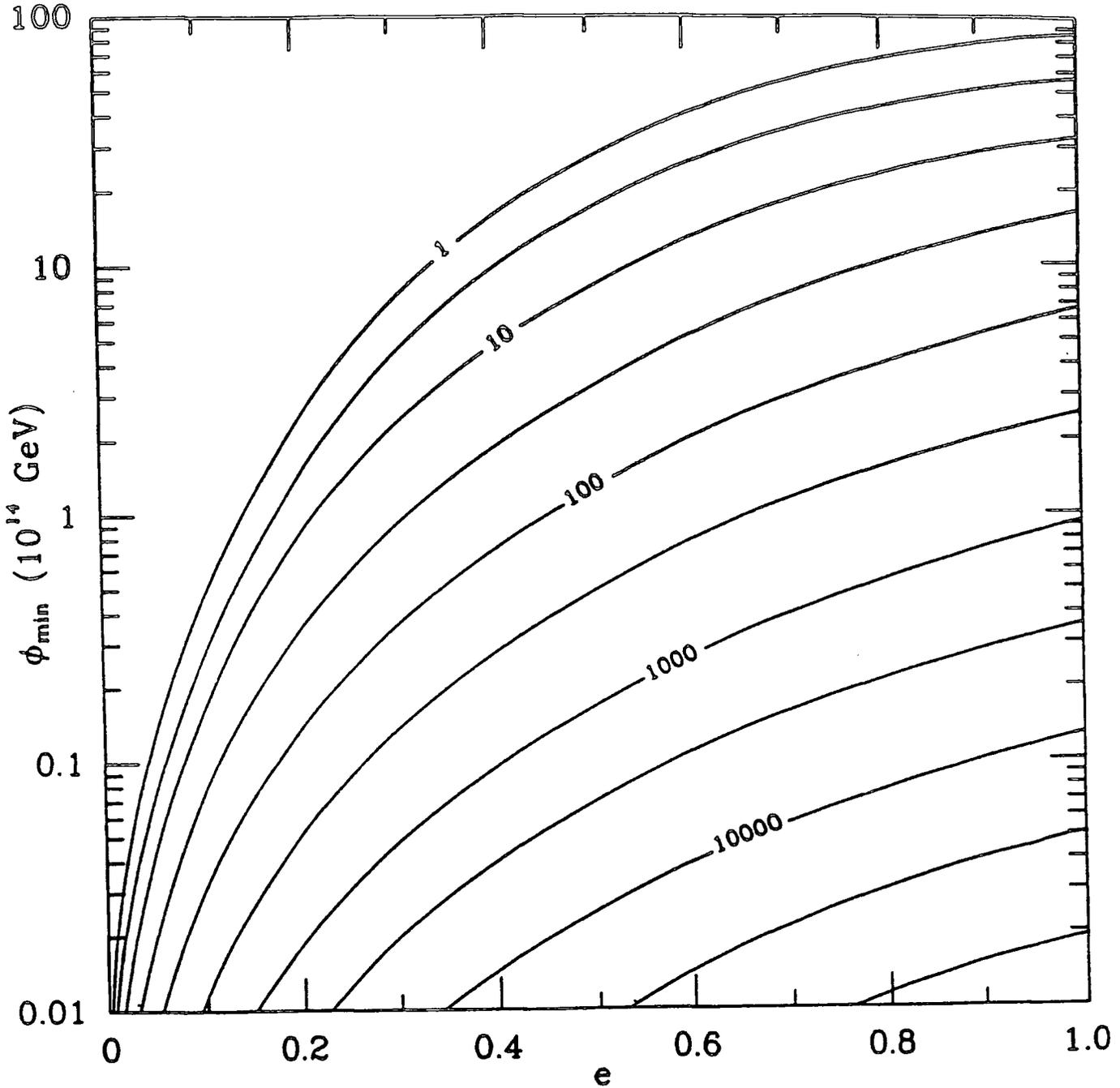
Figure 7.13: A contour plot of the average total number of interactions N in the ϕ_{min}-e parameter space for the model based on the Coleman-Weinberg potential (6.21) with $\Phi(t_i) = 10^{-2} \times 1/G$ (the other Brans-Dicke parameters, $\dot{\Phi}(t_i)$ and ω are set such that they maximise N for the given value of λ, δ and μ). The intermediate contours are at half orders of magnitude (i.e. $\sqrt{10}$ times the previous contour level). The parameter-space region to the left of the contour $N = 1$ is excluded because no interactions would occur to generate a thermal state.
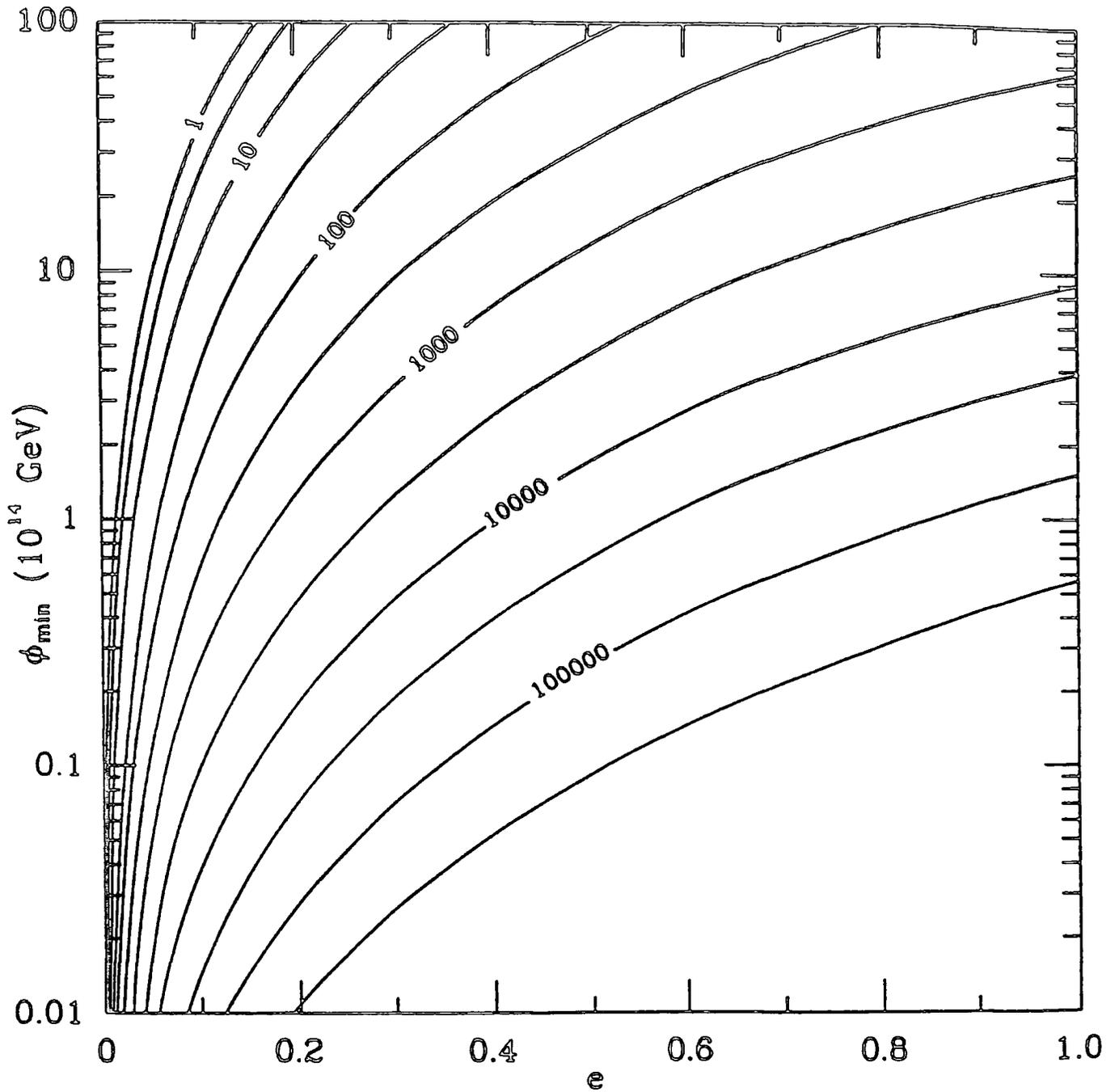
Figure 7.14: A contour plot of the average total number of interactions N in the ϕ_{min}-e parameter space for the model based on the Coleman-Weinberg potential (6.21) with $\Phi(t_i) = 10^2 \times 1/G$ (the other Brans-Dicke parameters, $\dot{\Phi}(t_i)$ and ω are set such that they maximise N for the given value of λ, δ and μ). The intermediate contours are at half orders of magnitude (i.e. $\sqrt{10}$ times the previous contour level). The parameter-space region to the left of the contour $N = 1$ is totally excluded because no interactions would occur to generate a thermal state.

The main conclusion of so doing is that, despite the extra free parameters available in the Brans-Dicke theory, the constraints on the inflaton potential parameters are remarkably similar to those obtained with conventional inflation. In other words, it seems just as difficult to produce a thermal state in extended inflation if not slightly more so. The reasons for this are fairly straightforward. Section 7.3 showed that the only Brans-Dicke parameter whose variation had any real impact on the total number of interactions N was the initial value of Φ, $\Phi(t_i)$. This parameter is constrained in the pure Brans-Dicke version of inflation by the fact that today $\Phi = 1/G$. Because Φ increases as t^2 during inflation it is found that $\Phi < 10^{-2} \times 1/G$ at the start of inflation (see (7.46)), and the smaller the value of Φ the more difficult it is to establish a thermal state (since it implies a larger effective value of G). It is thus marginally more difficult to establish a thermal state in the pure Brans-Dicke model than in standard inflation with GR. As already discussed, it is possible for the more general versions of extended inflation to get around this constraint, the foremost example being the induced gravity extension[12] (see fig. 7.1). However, since the only scale in such a potential is provided by G itself, it seems unlikely that the maximum starting value of Φ could be increased by more than a few orders of magnitude[10] without recourse to some special initial condition. The contour plots with such enhanced starting values of Φ (figs. 7.11 and 7.14) show that thermal state production is then somewhat less difficult than for pure Brans-Dicke extended inflation but the improvement is not very great.

A consequence of this similarity with conventional inflation is that the GUT scale is again an upper bound to the scale at which the generation of a thermal state could occur, and that symmetry breaking at scales $\lesssim 10^{12}$ GeV are preferable. So again, if inflation is to happen at higher energy scales than this then mechanisms which do not rely on a temperature dependent phase change seem to be favoured.

References

1. D. La and P.J. Steinhardt, *Phys. Rev. Lett.* **62**, 376 (1989)

2. C. Brans and C.H. Dicke, *Phys. Rev.* **24**, 925 (1961)

3. S. Weinberg, *Gravitation and Cosmology*, (Wiley, New York, 1972), Chapter 16

4. R. Holman *et al.*, *Phys. Rev. D* **43**, 995 (1991)

5. B.A. Campbell, A. Linde, K.A. Olive, *Nucl. Phys.* **B355**, 146 (1991); J. García-Bellido and M. Quirós, *Nucl. Phys.* **B368**, 463 (1992)

6. R. Holman *et al.*, *Phys. Rev. D* **43**, 3833 (1991)

7. C. Mathiazhagen and V. Johri, *Class. Q. Grav.*, **1**, L29, (1984)

8. J.D. Barrow and K. Maeda, *Nucl. Phys.* **B341**, 294 (1990)

9. C.M. Will, *Phys. Rep.* **113**, 345 (1984)

10. D. La, P.J. Steinhardt and E.W. Bertschinger, *Phys. Lett. B* **231**, 231 (1989)

11. E.J. Weinberg, *Phys. Rev. D* **40**, 3950 (1989)

12. F.S. Accetta and J.J. Trester, *Phys. Rev. D* **39**, 2854 (1989)

13. J. García-Bellido and M. Quirós, *Phys. Lett. B* **243**, 45 (1990)

14. P.J. Steinhardt and F.S. Accetta, *Phys. Rev. Lett.* **64**, 2740 (1990)

15. R. Holman, E.W. Kolb, Y. Wang, *Phys. Rev. Lett.* **65**, 17 (1990)

16. Numerical Algorithms Group, Fortran Library Mk. 15 (1991), Routine D02BAF

17. L.E. Gurevich, A.M. Finkelstein and V.A. Ruban, *Astrophys. and Space Science*, **22**, 231, (1973)

18. I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series, and Products*, (Academic, New York, 1980)

19. M. Kamionkowski and M.S. Turner, *Phys. Rev. D* **42**, 3310 (1990)

20. A.D. Linde *Phys. Lett. B* **129**, 177 (1983)

21. G. Mazenko, W. Unruh, R. Wald, *Phys. Rev. D* **31**, 273 (1985)

8

Conclusions

The standard hot big bang model does not provide a fully satisfactory account of the early development of the Universe. As discussed in section 3.5, some quite fundamental problems still remain. However, the inflationary scenario gives the promise of solving a number of these problems all at once, including the flatness/oldness/naturalness problem, the horizon problem and the monopole problem. It may also provide the initial spectrum of density inhomogeneities, from which the large scale structure of the Universe eventually formed.

As was seen in section 3.7, it is commonly supposed that the vacuum energy responsible for inflation is generated when the vacuum expectation value of a scalar field, the inflaton, is located away from the absolute minimum of its potential. One of the better motivated mechanisms for so doing relies on a symmetry breaking potential, qualitatively of similar form to the Higgs potential of the standard particle physics model described in chapter 2. In chapter 5 it was shown that when a model with such a potential is in thermal equilibrium it is transformed into a temperature-dependent effective potential. And this potential then generates a phase transition in a natural way, as the universe cools, from a state with a large vacuum energy to one where vacuum energy is negligible. Thermodynamics is a cornerstone of the hot big bang model (sections 3.3 and 3.4), and this is perhaps one reason why the thermal phase change seems so plausible. However, as explained in chapter 4, the assumption of a thermal state needs careful re-consideration when inflation is involved.

A self consistent calculation of the total number of interactions per particle (4.36) can be used to investigate the generation of an initial thermal state of inflatons in a quantitative way. A numerical implementation of this idea was given in chapter 6 where it was further applied to several specific example potentials. The generality of this procedure means that it can be applied to essentially any inflation model which relies on a thermal phase transition and extended inflation is a case in point. In chapter 7, it was shown how, after suitable modifications, the method can again be employed when the Brans-Dicke action of extended inflation replaces the Einstein action.

It was found that in both normal and extended inflation, with an inflaton potential based solely on self-interactions, inflation at the GUT scale is not favoured (see figs. 6.10, 7.10 and 7.11). In order to be able to form the required thermal initial state, a symmetry breaking scale $\lesssim 10^{11}$ GeV is required, which is perhaps more suggestive of a SUSY-breaking or an even lower energy transition. It is easier to generate a thermal state at the GUT scale with the Coleman-Weinberg potential (sections 6.3 and 7.5), but although the potential has the form required for "new-inflation", the size of the couplings needed to produce a thermal state (see figs. 6.7, 7.13 and 7.14) would not allow the ϕ-field to evolve by slow-rolling during the phase transition. It thus seems likely that the effective cosmological constant would not persist for long enough to solve the cosmological problems for which it has been invoked.

These conclusions, drawn from the analysis in chapters 6 and 7, can perhaps provide hints as to the nature of the inflationary transition if a thermal phase change mechanism is to be retained. Well below the GUT scale, i.e., $\lesssim 10^{12}$ GeV, there seems to be little or no restriction on the inflaton potential or the order of the phase change. However, at the GUT scale it seems that only a first order transition is allowed, since the couplings required for a second order transition are too large to allow slow-rolling. Furthermore, this first order transition cannot rely just on the self-coupling of the scalar. It may be possible to contrive a model with a first-order transition at the GUT scale if the scalar also couples to gauge particles, as in section 6.3. In this case it is the gauge coupling which ensures thermal equilibrium since the scalar self-coupling cannot be large enough to achieve this on its own. But note that a realistic

potential for this purpose would have to have five independent parameters; a mass and two scalar couplings, a fermion Yukawa coupling to ensure the potential has the correct form (see section 5.4), and the coupling to the gauge field.

For inflation above the GUT scale it seems that a temperature dependent phase change mechanism cannot readily be generated in any way. At such high scales all the models that have been tried in chapters 6 and 7 would require very strong, i.e., non-perturbative, couplings. In this case, the definition of temperature is by no means certain because it seems unlikely that the ideal gas approximation would apply. Certainly, the effective potential formalism of chapter 5 is no longer valid. It would seem therefore that mechanisms which do not rely on a thermal phase change, like chaotic inflation, have to be preferred at such early times.

Of course, the real problem here is a difficulty that has plagued inflation from the start — there is no clue as to the identity of the scalar inflaton field. If more realistic or more natural models existed then it would be far easier to draw more precise conclusions. Increasingly however, it is becoming apparant that inflation is a difficult scenario to realise in a straightforward manner, though this may just be a reflection of an incomplete knowledge of the physics that lies above the electroweak scale. The real question is whether there is an inflaton field in the true theory of elementary particle physics, be it a GUT, SUSY, superstring or something entirely different.