QC
174.45
.B54
1982

QUANTUM FIELDS IN CURVED SPACE

N. D. BIRRELL

Logica Pty Ltd, Australia

AND

P. C. W. DAVIES

Professor of Theoretical Physics, University of Newcastle upon Tyne

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

© Cambridge University Press 1982

First published 1982
First paperback edition (with corrections) 1984
Reprinted 1989, 1992, 1994

Library of Congress catalogue card number: 81-3851

British Library cataloguing in publication data

Birrell, N.D.
Quantum fields in curved space

1. Quantum field theory
I. Title II. Davies, P.C.W.
530,1'43 QC174.45

ISBN 0 521 27858 9 paperback

Transferred to digital printing 1999

Contents

Preface

The subject of quantum field theory in curved spacetime, as an approximation to an as yet inaccessible theory of quantum gravity, has grown tremendously in importance during the last decade. In this book we have attempted to collect and unify the vast number of papers that have contributed to the rapid development of this area. The book also contains some original material, especially in connection with particle detector models and adiabatic states.

The treatment is intended to be both pedagogical and archival. We assume no previous acquaintance with the subject, but the reader should preferably be familiar with basic quantum field theory at the level of Bjorken & Drell (1965) and with general relativity at the level of Weinberg (1972) or Misner, Thorne & Wheeler (1973). The theory is developed from basics, and many technical expressions are listed for the first time in one place. The reader's attention is drawn to the list of conventions and abbreviations on page ix, and the extensive references and bibliography.

In preparing this book we have drawn upon the material of a very large number of authors. In adapting certain published material (including that of the authors) we have gratuitously made what we consider to be corrections, occasionally without explicitly warning the reader that our use of that material differs from the original publications.

The bulk of the text was written while we worked together at the Department of Mathematics, Kings College, London. We are greatly indebted to many colleagues there and elsewhere for assistance. Special thanks are extended to T.S. Bunch, S.M. Christensen, N.A. Doughty, J.S. Dowker, M.J. Duff, L.H. Ford, S.A. Fulling, C.J. Isham, G. Kennedy, L. Parker and R.M. Wald for critical reading of sections of manuscript.

Finally we should like to thank Mrs J. Bunn for typing the manuscript and the Science Research Council for financial support.

Preface to the paperback edition

Since the book first went to press, there have been several important advances in this subject area. The topic of interacting fields in curved space has been greatly developed, especially in connection with the phenomenon of symmetry breaking and restoration in the very early universe, where both high temperatures and spacetime curvature are significant. A direct consequence of this work has been the formulation of the so-called inflationary universe scenario, in which the universe undergoes a de Sitter phase in the very early stages. This work has focussed attention once more on quantum field theory in de Sitter space, and on the calculation of $\langle \phi^2 \rangle$. A comprehensive review of the inflationary scenario is given in *The Very Early Universe*, edited by G.W. Gibbons, S.W. Hawking and S.T.C. Siklos (Cambridge University Press, 1983).

Further results of a technical nature have recently been obtained concerning a number of the topics considered in this book. Mention should be made of the work of M.S. Fawcett, who has finally calculated the quantum stress tensor for a Schwarzschild black hole (*Commun. Math. Phys.*, **81** (1983), 103), and of W.G. Unruh & R.M. Wald, who have clarified the thermodynamic properties of black holes by appealing to the effects of accelerated mirrors close to the event horizon (*Phys. Rev. D*, **25** (1982), 942; **27** (1983), 2271). Interest has also arisen over field theories in higher-dimensional spacetimes, in which Casimir and other vacuum effects become important. For a review, see E. Witten, *Nucl. Phys. B*, **186** (1981), 412. Finally, much further work has been done on the properties of particle detectors (see, for example, the paper by K.J. Hinton in *J. Phys. A: Gen. Phys.*, **16** (1983), 1937).

We are grateful to K.J. Hinton, J. Pfautsch, S.D. Unwin and W.R. Walker for assistance in revising the text.

Note added at 1986 reprinting

We would like to thank Professor H. Minn for providing corrections to the original printing.

Conventions and abbreviations

Our notation for quantum field theory mainly follows that of Bjorken & Drell (1965). The sign conventions for the metric and curvature tensors are $(- - -)$ in the terminology of Misner, Thorne & Wheeler (1973). That is, the metric signature is $(+ - - -)$; $R^{\alpha}{}_{\beta\gamma\delta} = \partial_{\delta}\Gamma^{\alpha}{}_{\beta\gamma} - \ldots$; $R_{\mu\nu} = R^{\alpha}{}_{\mu\alpha\nu}$. Formulae can be changed from our notation to the often used Misner, Thorne & Wheeler $(+ + +)$ conventions by changing the signs of $g_{\mu\nu}$, $\Box \equiv g^{\mu\nu}\nabla_{\mu}\nabla_{\nu}$, $R^{\alpha}{}_{\beta\gamma\delta}$, $R_{\mu\nu}$, $T_{\mu}{}^{\nu}$ but leaving $R_{\alpha\beta\gamma\delta}$, $R_{\mu}{}^{\nu}$, R and $T_{\mu\nu}$ unchanged. For the majority of the book we use units in which $\hbar = c = G = 1$.

The following special symbols and abbreviations are used throughout:

$*$	complex conjugate
$\dagger$ or h.c.	Hermitian conjugate
$\bar{}$	Dirac adjoint
$\dfrac{\partial}{\partial x^{\mu}}$ or ∂_{μ} or $,\mu$	partial derivative
∇_{μ} or $;\mu$	covariant derivative
Re (Im)	real (imaginary) part
tr	trace
ln	natural logarithm
k_{B}	Boltzmann's constant
γ	Euler's constant
$[A,B]$	$AB - BA$
$\{A,B\}$	$AB + BA$
$a_{(\mu,\nu)}$	$\frac{1}{2}(a_{\mu,\nu} + a_{\nu,\mu})$
$\simeq$	approximately equal to
$\sim$	order of magnitude estimate
$\approx$	asymptotically approximate to
$\equiv$	defined to be equal to
$: :$	normal ordering

1
Introduction

The last decade has witnessed remarkable progress in the construction of a unified theory of the forces of nature. The electromagnetic and weak interactions have received a unified description with the Weinberg–Salam theory (Weinberg 1967, Salam 1968), while attempts to incorporate the strong interaction as described by quantum chromodynamics into a wider gauge theory seem to be achieving success with the so-called grand unified theories (Georgi & Glashow 1974, for a review see Cline & Mills 1978).

The odd one out in this successive unification is gravity. Not only does gravity stand apart from the other three forces of nature, it stubbornly resists attempts to provide it with a quantum framework. The quantization of the gravitational field has been pursued with great ingenuity and vigour over the past forty years (for reviews see Isham 1975, 1979*a*, 1981) but a completely satisfactory quantum theory of gravity remains elusive. Perhaps the most hopeful current approaches are the supergravity theories, in which the graviton is regarded as only one member of a multiplet of gauge particles including both fermions and bosons (Freedman, van Nieuwenhuizen & Ferrara 1976, Deser & Zumino 1976; for a review see van Nieuwenhuizen & Freedman 1979).

In the absence of a viable theory of quantum gravity, can one say anything at all about the influence of the gravitational field on quantum phenomena? In the early days of quantum theory, many calculations were undertaken in which the electromagnetic field was considered as a classical background field, interacting with quantized matter. Such a semiclassical approximation readily yields some results that are in complete accordance with the full theory of quantum electrodynamics (see, for example, Schiff 1949, chapter 11). One may therefore hope that a similar regime exists for quantum aspects of gravity, in which the gravitational field is retained as a classical background, while the matter fields are quantized in the usual way. Adopting Einstein's general theory of relativity as a description of gravity, one is led to the subject of quantum field theory in a curved background spacetime, which is the subject of this book.

It was originally pointed out by Planck (1899) that the universal constants G, $\hbar$ and c could be combined to give a new fundamental unit of length, the Planck length $(G\hbar/c^3)^{\frac{1}{2}} = 1.616 \times 10^{-33}$ cm, and time, the Planck time $(G\hbar/c^5)^{\frac{1}{2}} = 5.39 \times 10^{-44}$ s. If the gravitational field is treated as a small perturbation, and attempts are made to quantize it along the lines of quantum electrodynamics (Q.E.D.), then the square of the Planck length appears in the role of coupling constant. Unlike Q.E.D., however, in which the coupling constant $e^2/\hbar c$ is dimensionless (and small) the Planck length has dimensions. Effects can become large when the length and time scales of the quantum processes of interest fall below the Planck value. When this happens, the higher orders of perturbation theory become comparable with the lowest order, and the whole concept of a small perturbation expansion breaks down.

The Planck values therefore mark the frontier at which a full theory of quantum gravity, preferably non-perturbative in character, must be invoked. Nevertheless, one might hope that when the distances and times involved are much larger than the Planck values, the quantum effects of the gravitational field will be negligible. As the Planck length is so small (twenty powers of ten below the size of an atomic nucleus) this appears to leave much scope for a semiclassical theory.

There is, however, a problem with this naive reasoning. According to the equivalence principle, which lies at the very foundation of metric theories of gravity, all forms of matter and energy couple equally strongly to gravity. This includes, of course, gravitational energy; crudely speaking, gravity gravitates. In quantum language, we may say that the graviton is just as much subject to an external gravitational field as, say, a photon. Consequently, whenever a classical background gravitational field produces important effects involving (real or virtual) photons, one must allow for equally important effects involving gravitons. It follows that quantum gravity will enter in a non-trivial way *at all* scales of distance and time, whenever interesting quantum field effects occur. Thus, the basic non-linearity of gravity frustrates all attempts to ignore quantum gravity (Duff 1981).

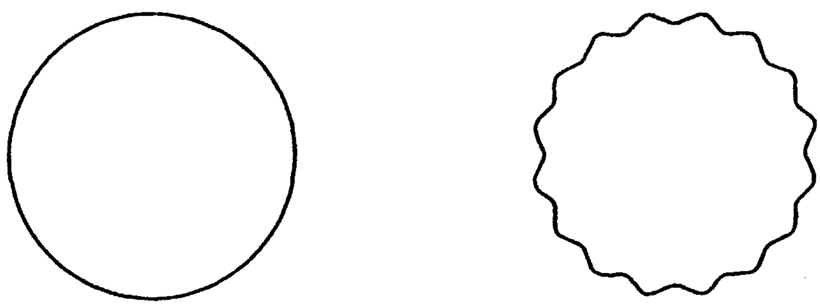
In spite of this complication, it may still be possible to proceed with a semiclassical description. In ordinary classical relativity one frequently wishes to discuss the propagation of gravitational waves in a curved background spacetime. Although one is dealing with the vacuum Einstein equation, the small disturbance that represents the gravitational wave can be separated off from the background spacetime:

$$g_{\mu\nu} = g^{\mathrm{c}}_{\mu\nu} + \bar{g}_{\mu\nu} \tag{1.1}$$

where $\bar{g}_{\mu\nu}$ represents the wave and $g^c_{\mu\nu}$ the background spacetime. The waves can then be treated as a null fluid much like any other, and their contribution to the left-hand side of Einstein's equations can be cast in a form that enables them to be taken over to the right and treated as a part of the source, i.e., as a part of $T_{\mu\nu}$.

In this spirit, it seems reasonable to suppose that the 'graviton' field, representing linearized perturbations on the background spacetime, can be included along with all the other quantum fields as part of the matter rather than the geometry. So long as one remains well clear of Planck dimensions, such a linearized approximation should work. However, until a full theory of quantum gravity is available, this approximation will be open to question. (Whether or not one should call a theory that makes provision for gravitons 'semiclassical' is, of course, a matter for debate.) It is analogous to treating photon emission by an atom immersed in a background electric or magnetic field.

DeWitt (1967*a*,*b*) has used (1.1) as the starting point for an approach to the complete quantization of gravity – the so called 'background field' method. In this approach $g^c_{\mu\nu}$ is the classical metric of some background spacetime, and $\bar{g}_{\mu\nu}$ is taken to be a quantum field propagating in this background. The Einstein action and the matter field actions can be expanded in powers of $\bar{g}$ about g^c, and Feynman rules derived. The lowest order, one-loop, quantum processes are shown in the Feynman diagrams in fig. 1. The wavy line represents the propagator for the field $\bar{g}_{\mu\nu}$ propagating in the background $g^c_{\mu\nu}$, while the uniform line represents the corresponding propagator for a matter field. Since neither fig. 1*a* nor 1*b* contains any graviton vertices, they both represent expressions which are independent of the gravitational coupling constant G, and so are always of comparable magnitude. Thus, at the one-loop level, the quantization of the



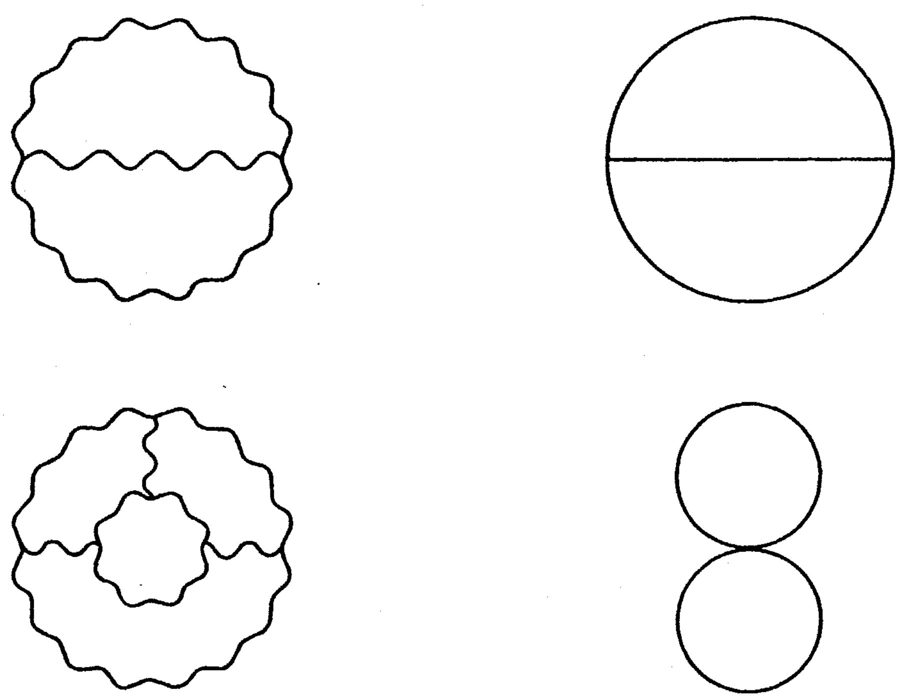
(a) *(b)*

Fig. 1. Lowest order (one-loop) contribution to the vacuum energy from matter fields (*a*) and linearized gravitons (*b*) propagating in a prescribed background gravitational field.

gravitational field in the background $g^c_{\mu\nu}$ is equally as important as the quantization of the matter fields.

Physically, the single closed loops represent an infinite vacuum or zero-point energy, that in the case of a flat background spacetime (i.e., conventional quantum field theory) is artificially removed by subtraction, or by so-called normal ordering (see §2.4). When the background is curved, however, a more elaborate procedure is necessary involving the dynamics of the gravitational field. This is the device known as *renormalization*, and is familiar from Q.E.D., which is also plagued by divergences.

In the latter case, the divergences are removed by renormalization of particle masses, charges and wavefunctions. Only a finite number of quantities needs to be renormalized, a feature that qualifies Q.E.D. for the status of a 'renormalizable theory'. This 'renormalizability' depends crucially on the fact that the coupling constant $e^2/\hbar c$ is dimensionless. In contrast, G has units of (length)^2 (in natural units $\hbar = c = 1$), which gives rise, via a simple power counting argument, to an unending sequence of new divergences at each order. Higher order terms in the expansion of the



(a) *(b)*

Fig. 2. (*a*) Higher order, multiple-loop graviton diagrams containing vertices. These represent contributions to the vacuum energy that diverge more strongly than that from fig. 1(*b*), and render quantum gravity non-renormalizable. (*b*) When non-gravitational interactions are introduced, multiple loops can also occur in the matter field diagrams.

gravitational action in powers of $\bar{g}$ produce graviton Feynman diagrams with multiple loops (see fig. 2*a*). There is a simple relationship between the degree of divergence and the number of such loops in a diagram (see, for example, Duff 1975), such that with increasing numbers of loops one encounters more and more virulent divergences. This fact renders quantum gravity 'unrenormalizable' – with each new order more new physical quantities have to be invented to absorb the infinities. It is for this reason that the quantization of the gravitational field has not been satisfactorily accomplished.

If one truncates the expansion of the combined theory of gravity and matter at some particular number of loops, then the finite number of divergent quantities can be removed by renormalization of a finite number of physical quantities. For example, as we shall see in chapter 6, at the one-loop level, renormalization of G, the cosmological constant Λ, and the coupling constants of two new geometrical tensors suffices to render the theory finite. Thus, in a sense, this truncated theory could be considered renormalizable.

It should be noted that if one attempts to carry out quantization by expanding around flat spacetime, taking $g^{c}_{\mu\nu} = \eta_{\mu\nu}$ (the Minkowski space metric) in (1.1), then the renormalization cannot be carried out in a generally covariant way (see, for example, Duff 1975). This, combined with other difficulties (Christensen & Duff 1980), obliges one to consider fields propagating in a background with arbitrary metric $g^{c}_{\mu\nu}$.

In this book we consider, for the most part, the quantum theory of gravity plus matter truncated at the one-loop level. For free matter fields, there are no higher loop processes anyway and fig. 1*a* gives the exact contribution of these fields (to the effective action defined in chapter 6). Fig. 1*b* gives only the contribution from the gravitons which is of zeroth order in G. As the loop expansion is an expansion in $\hbar$ (Nambu 1966) the theory truncated at the one-loop level contains all terms of the complete theory to order $\hbar$, and is in that sense the first order quantum correction to general relativity.

If self- or mutually-interacting matter fields are included in the theory then the one-loop contribution of the matter fields shown in fig. 1*a* is no longer exact, as there now exist multiple-loop diagrams involving vertices (fig. 2*b*). As it is the effects of interactions between matter fields ('particles') that are most often observed in the laboratory, such multiple-loop matter field diagrams should be included in our discussions. However, if we wish to work to a consistent order in $\hbar$, we should then have to include *graviton* diagrams also with arbitrary numbers of loops and so be confronted with the non-renormalizability of gravity. However, each extra graviton loop in a connected Feynman diagram in general introduces a factor of G, while each matter field loop introduces a factor of the relevant coupling constant (such as e^2). If l is a typical length or time scale for the system under consideration, then provided $l^{-2}G \ll e^2$, the effect of additional graviton loops will be insignificant compared with that of additional matter loops. Thus, even in the case of interacting matter fields, there is a large regime in which quantum gravity can be limited to the one-loop level with some justification.

Given that we possess at least a reasonable approximation to a theory of gravitational effects on quantum fields, how important are the processes described thereby? Crudely speaking, non-trivial gravitational effects occur in quantum field modes for which the wavelength λ is comparable with some characteristic length scale of the background spacetime. Thus, near a black hole of radius r, the quantum field modes with wavelength $\gtrsim r$ are seriously disturbed by the presence of the hole. Similarly, if the gravitational field changes on a timescale t, then quantum field modes with frequency $\lesssim t^{-1}$ are seriously disturbed. Thus, if one regards 10^{-13} cm and 10^{-23} s as the length and time scales characteristic of important quantum processes, one finds that only in the vicinity of microscopic black holes or in the earliest epochs of the big bang can important effects be expected.

The weakness of gravity therefore effectively precludes phenomena that can be studied in the laboratory and, unless microscopic black holes are much more numerous than present estimates suggest, it precludes any possibility at all of direct observational verification. Quantum field theory in curved spacetime must, it seems, rest entirely upon theoretical considerations.

The paucity of experimental checks renders all the more significant the results of Hawking (1975). His study of quantum black holes and the discovery of their thermal emission is a cornerstone of the theory developed in this book. Hawking's result is compelling for two reasons. First, it appears to be very fundamental, and has been derived in several different ways. Second, it establishes a strong connection between black holes and thermodynamics that was suspected before the application of quantum theory to black holes. It is therefore tempting to suppose that the Hawking effect has exposed a small corner of a broad new area of fundamental physics in which gravity, quantum field theory and thermodynamics are closely interwoven. If this is the case, then their synthesis would almost certainly lead to important new advances in physics, including some with observational consequences.

The situation can be compared with the early days of kinetic theory. The atomic hypothesis was not really open to direct experimental verification in the mid-nineteenth century due to the smallness of atomic effects. Nevertheless, the fully developed theory was capable of reaching beyond the atomic domain and predicting new phenomena in gas dynamics that could be checked. Similarly, one hopes that quantum gravity would, if it became properly understood, intrude into other, more accessible, areas of physics.

Investigation of the effects of gravity on quantum fields dates at least since the work of Schrödinger (1932). After the Second World War there was a surge of interest in quantizing the gravitational field, but direct investigation of particle creation effects in a background gravitational field really began in earnest with the work of Parker in the late 1960s followed by the investigations of Zel'dovich and coworkers. These early investigations dwelt on the cosmological consequences of particle creation. They were hampered by the lack of systematic techniques for studying the stress–energy–momentum tensor, $T_{\mu\nu}$, of the gravitationally disturbed quantum fields. The stress-tensor is important for two reasons. It can be used to assess the importance of quantum effects on the dynamics of the gravitational field itself (i.e., the back-reaction problem). Also, it is frequently a more useful probe of the physical situation than a particle count. In regions of strong gravity, vacuum polarization effects, akin to those in Q.E.D., can lead to important phenomena even in the absence of actual particle creation.

In the mid-seventies, a great deal of effort was expended in developing rigorous techniques for computing $\langle T_{\mu\nu} \rangle$. Being formally infinite, these techniques involve renormalization, so some of them were borrowed from ordinary (Minkowski space) quantum field theory and from the full quantum gravity theory, while others were developed specially.

An essential feature of these techniques is that they all yield a covariantly conserved (i.e. divergenceless) $\langle T_{\mu\nu} \rangle$, which is therefore a suitable candidate for the right-hand side of a semiclassical Einstein equation. It might seem that if particle pairs are created from the vacuum then $\langle T_{\mu\nu} \rangle$ should not be conserved. Indeed, Hawking (1970) showed that a conserved $\langle T_{\mu\nu} \rangle$, subject to the dominant energy condition (basically that energy and pressure should always remain positive – see Hawking & Ellis 1973) was incompatible with particle creation. However, one of the unusual features of curved space quantum field theory is that spacetime curvature can induce negative stress–energy–momentum in the vacuum, thereby violating the dominant energy condition, and circumventing Hawking's result (Zel'dovich & Pitaevsky 1971). Thus particle creation is compatible with a conserved $\langle T_{\mu\nu} \rangle$. The possibility of violating the energy conditions also opens the way to the avoidance of spacetime singularities (see §7.4).

With Hawking's discovery of thermal black hole emission, combined with these improving techniques for computing $\langle T_{\mu\nu} \rangle$, the subject of quantum field theory in curved spacetime, or one-loop quantum gravity, enjoyed a period of very rapid expansion, with several hundred papers appearing in the literature. While some of the interpretative issues have been contentious (most notably the physical significance to be attached to particles versus $\langle T_{\mu\nu} \rangle$, the effect of spacetime singularities, the existence of so-called conformal anomalies, and the criteria to be used in the specification of appropriate quantum states) there is now broad agreement on most of the technical results. In the chapters that follow the treatment will inevitably reflect some of the interpretative opinions of the authors, but we have tried as far as possible to adhere to what we understand to be the 'majority view'. We have tried to avoid making vague statements of a physical nature about 'particles' and 'energy', where possible stating our conclusions in operational terms – what a hypothetical observer moving in such-and-such a way would actually measure with a particular piece of apparatus.

The organization of material is conventional. In chapter 2 we review the basic concepts of ordinary Minkowski space quantum field theory and establish our notation. Chapter 3 generalizes these ideas to curved spacetime and introduces the notion of particle creation by gravitational fields. We also give what is to date the most complete and careful description of the concepts of adiabatic states, and of particle detectors. Applications to flat and curved spacetime follow in chapters 4 and 5, with several explicit examples worked in detail.

A central feature of the book, chapter 6, gives an exhaustive account of regularization and renormalization techniques. Several methods have been published, and we have endeavoured not only to discuss them all (with examples) but to unify them as much as possible. Armed with these results, the reader can tackle chapter 7, that applies the methods of chapter 6 to a number of important examples.

The full range of techniques developed in the earlier chapters are then deployed for a detailed discussion of black holes, which occupies the whole of chapter 8. Considerations of space compelled us to limit many of the important and fascinating physical implications to a catalogue of references. We have been unable, for example, to dwell at length on the thermodynamic aspects of black holes and their possible extension to more general gravitational fields; nor have we been able to give a detailed account of time asymmetry, Poincaré cycles, black holes versus white holes and the implications of quantum spacetime singularities. In addition, nearly all the exciting work on the astrophysical and cosmological consequences of quantum black holes has been omitted.

The final chapter attempts to go beyond what may be considered as the 'first round' of results in the subject of quantum field theory in curved spacetime. Here we briefly outline the generalization of the theory that will be necessary to accommodate non-gravitational interactions. Special interest centres on issues such as the effect of self- and mutual-field interactions on particle creation and vacuum stress, and whether a theory that is renormalizable in Minkowski space, such as Q.E.D. or $\lambda\phi^4$ theory, remains renormalizable in the presence of a non-trivial topology and/or spacetime curvature.

It is inevitable in a book of this sort that there will be some ragged edges. The subject is still rapidly evolving, and many gaps remain to be filled. Nevertheless, the last year or so has witnessed a period of consolidation and reflection, so that we have been able to present a reasonably coherent and self-contained account.

2

Quantum field theory in Minkowski space

In this chapter we shall summarize the essential features of ordinary Minkowski space quantum field theory, with which we assume the reader has a working knowledge. A great deal of the formalism can be extended to curved spacetime and non-trivial topologies with little or no modification. In the later chapters we shall follow the treatment given here.

Most of the detailed analysis will refer to a scalar field, but the main results will be listed for higher spins also. This restriction will enable the important features of curved space quantum field theory to emerge with the minimum of mathematical complexity.

Much of the chapter will be familiar from textbooks such as Bjorken & Drell (1965), but the reader should take special note of the results on the expectation value of the stress–energy–momentum tensor and vacuum divergence (§2.4), as these will play a central role in what follows. Special importance also attaches to Green functions, treated in detail in §2.7. The reader may be unfamiliar with thermal Green functions and metric Euclideanization. As these will be essential for an understanding of the quantum black hole system, an outline of this topic is given here.

Finally, although we shall not develop a lot of our formalism using the Feynman path-integral technique, we do make use of the basic structure of the path integral in the work on renormalization in chapter 6, and again on interacting fields in curved space in chapter 9. While it is not necessary for the reader to master the path-integral formulation, the basic outline given in §2.8 may be found helpful.

2.1 Scalar field

Consider a scalar field $\phi(t, \mathbf{x})$ defined at all points $(t, \mathbf{x})$ of an n-dimensional Minkowski spacetime, satisfying the field equation

$$(\Box + m^2)\phi = 0 \tag{2.1}$$

where $\Box \equiv \eta^{\mu\nu}\partial_\mu \partial_\nu$ and $\eta^{\mu\nu}$ is the Minkowskian metric tensor. The quantity m is to be interpreted as the mass of the field quanta when the theory is quantized. In what follows we shall frequently abbreviate the spacetime point $(t,\mathbf{x}) = (x^0,\mathbf{x})$ as x.

Equation (2.1) may be obtained from the Lagrangian density

$$\mathscr{L}(x) = \tfrac{1}{2}(\eta^{\alpha\beta}\phi_{,\alpha}\phi_{,\beta} - m^2\phi^2) \tag{2.2}$$

by constructing the action

$$S = \int \mathscr{L}(x)\mathrm{d}^n x \tag{2.3}$$

and demanding that for variations with respect to ϕ

$$\delta S = 0. \tag{2.4}$$

One set of solutions of (2.1) is

$$u_{\mathbf{k}}(t,\mathbf{x}) \propto \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x} - \mathrm{i}\omega t} \tag{2.5}$$

where

$$\omega \equiv (k^2 + m^2)^{\frac{1}{2}} \tag{2.6}$$

$$k \equiv |\mathbf{k}| = \left(\sum_{i=1}^{n-1} k_i^2\right)^{\frac{1}{2}} \tag{2.7}$$

and the Cartesian components of $\mathbf{k}$ can take the values

$$-\infty < k_i < \infty, \quad i = 1, \ldots, n-1.$$

The modes (2.5) are said to be positive frequency with respect to t, being eigenfunctions of the operator $\partial/\partial t$:

$$\frac{\partial}{\partial t}u_{\mathbf{k}}(t,\mathbf{x}) = -\mathrm{i}\omega u_{\mathbf{k}}(t,\mathbf{x}), \quad \text{with } \omega > 0. \tag{2.8}$$

Define the scalar product

$$\begin{aligned}(\phi_1,\phi_2) &= -\mathrm{i}\int\{\phi_1(x)\partial_t\phi_2^*(x) - [\partial_t\phi_1(x)]\phi_2^*(x)\}\mathrm{d}^{n-1}x \\ &= -\mathrm{i}\int_t \phi_1(x)\overleftrightarrow{\partial_t}\phi_2^*(x)\mathrm{d}^{n-1}x,\end{aligned} \tag{2.9}$$

where t denotes a spacelike hyperplane of simultaneity at instant t. Then the $u_{\mathbf{k}}$ modes (2.5) are orthogonal

$$(u_{\mathbf{k}}, u_{\mathbf{k}'}) = 0, \quad \mathbf{k} \neq \mathbf{k}'. \tag{2.10}$$

If we choose

$$u_{\mathbf{k}} = [2\omega(2\pi)^{n-1}]^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t} \tag{2.11}$$

then the $u_{\mathbf{k}}$ functions are normalized in the scalar product (2.9):

$$(u_{\mathbf{k}}, u_{\mathbf{k}'}) = \delta^{n-1}(\mathbf{k} - \mathbf{k}'). \tag{2.12}$$

For many purposes it is more convenient to restrict the solutions $u_{\mathbf{k}}$ to the interior of a spacelike $(n-1)$-torus of side L (i.e., choose periodic boundary conditions). Then

$$u_{\mathbf{k}} = (2L^{n-1}\omega)^{-\frac{1}{2}} e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t} \tag{2.13}$$

where

$$k_i = 2\pi j_i/L, \quad j_i = 0, \pm 1, \pm 2, \ldots, \quad i = 1, \ldots, n-1.$$

Thus

$$(u_{\mathbf{k}}, u_{\mathbf{k}'}) = \delta_{\mathbf{k}\mathbf{k}'}. \tag{2.14}$$

To convert from continuum to discrete (box) normalization one should replace each $\int d^{n-1}k$ by

$$(2\pi/L)^{n-1} \prod_{i=1}^{n-1} \sum_{j_i} \equiv (2\pi/L)^{n-1} \sum_{\mathbf{k}}.$$

2.2 Quantization

The system is quantized in the canonical quantization scheme by treating the field ϕ as an operator, and imposing the following equal time commutation relations

$$\left.\begin{aligned} [\phi(t,\mathbf{x}), \phi(t,\mathbf{x}')] &= 0 \\ [\pi(t,\mathbf{x}), \pi(t,\mathbf{x}')] &= 0 \\ [\phi(t,\mathbf{x}), \pi(t,\mathbf{x}')] &= i\delta^{n-1}(\mathbf{x} - \mathbf{x}') \end{aligned}\right\} \tag{2.15}$$

where π is the canonically conjugate variable to ϕ defined by

$$\pi = \frac{\partial \mathscr{L}}{\partial(\partial_t \phi)} = \partial_t \phi. \tag{2.16}$$

The field modes (2.11) or (2.13) and their respective complex conjugates form a complete orthonormal basis with scalar product (2.9), so ϕ may be expanded as

$$\phi(t,\mathbf{x}) = \sum_{\mathbf{k}} [a_{\mathbf{k}} u_{\mathbf{k}}(t,\mathbf{x}) + a_{\mathbf{k}}^{\dagger} u_{\mathbf{k}}^{*}(t,\mathbf{x})]. \tag{2.17}$$

The equal time commutation relations for ϕ and π are then equivalent to

$$\left.\begin{aligned} [a_{\mathbf{k}}, a_{\mathbf{k}'}] &= 0 \\ [a_{\mathbf{k}}^{\dagger}, a_{\mathbf{k}'}^{\dagger}] &= 0 \\ [a_{\mathbf{k}}, a_{\mathbf{k}'}^{\dagger}] &= \delta_{\mathbf{k}\mathbf{k}'}. \end{aligned}\right\} \tag{2.18}$$

In the Heisenberg picture, the quantum states span a Hilbert space. A convenient basis in this Hilbert space is the so-called Fock representation. The normalized basis ket vectors, denoted $|\rangle$, can be constructed from the vector $|0\rangle$, called the vacuum, or no-particle state, the physical significance of which will be discussed shortly. The state $|0\rangle$ has the property that it is annihilated by all the $a_{\mathbf{k}}$ operators:

$$a_{\mathbf{k}}|0\rangle = 0, \quad \forall \mathbf{k}. \tag{2.19}$$

The state obtained by operating on $|0\rangle$ with $a_{\mathbf{k}}^{\dagger}$ is called a one-particle state, and is denoted by $|1_{\mathbf{k}}\rangle$

$$|1_{\mathbf{k}}\rangle = a_{\mathbf{k}}^{\dagger}|0\rangle. \tag{2.20}$$

Similarly one may construct many-particle states

$$|1_{\mathbf{k}_1}, 1_{\mathbf{k}_2}, \ldots, 1_{\mathbf{k}_j}\rangle = a_{\mathbf{k}_1}^{\dagger} a_{\mathbf{k}_2}^{\dagger} \ldots a_{\mathbf{k}_j}^{\dagger}|0\rangle, \tag{2.21}$$

if all $\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_j$ are distinct. If any $a_{\mathbf{k}}^{\dagger}$ are repeated, then

$$|{}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, {}^jn_{\mathbf{k}_j}\rangle = ({}^1n!\,{}^2n! \ldots {}^jn!)^{-\frac{1}{2}} (a_{\mathbf{k}_1}^{\dagger})^{{}^1n} (a_{\mathbf{k}_2}^{\dagger})^{{}^2n} \ldots (a_{\mathbf{k}_j}^{\dagger})^{{}^jn}|0\rangle, \tag{2.22}$$

the $n!$ terms being necessary to accommodate the Bose statistics of identical scalar particles. Also

$$a_{\mathbf{k}}^{\dagger}|n_{\mathbf{k}}\rangle = (n+1)^{\frac{1}{2}}|(n+1)_{\mathbf{k}}\rangle \tag{2.23}$$

$$a_{\mathbf{k}}|n_{\mathbf{k}}\rangle = n^{\frac{1}{2}}|(n-1)_{\mathbf{k}}\rangle. \tag{2.24}$$

The basis vectors are normalized according to

$$\begin{aligned} &\langle {}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, {}^rn_{\mathbf{k}_r} | {}^1m_{\mathbf{k}'_1}, {}^2m_{\mathbf{k}'_2}, \ldots, {}^sm_{\mathbf{k}'_s}\rangle \\ &\quad = \delta_{rs} \sum_{\alpha} \delta_{{}^1n^{\alpha(1)}m} \ldots \delta_{{}^rn^{\alpha(s)}m} \delta_{\mathbf{k}_1\mathbf{k}'_{\alpha(1)}} \ldots \delta_{\mathbf{k}_r\mathbf{k}'_{\alpha(s)}} \end{aligned} \tag{2.25}$$

where the sum is over all permutations α of the integers $1 \ldots s$.

2.3 Energy–momentum

To explore the significance of these Fock states, it is instructive to examine the Hamiltonian and momentum operators for the field. These quantities are obtained from the stress–energy–momentum tensor, $T_{\mu\nu}$, henceforth abbreviated as stress-tensor. $T_{\mu\nu}$ may be constructed in a standard manner (see (3.189)) to be

$$T_{\alpha\beta} = \phi_{,\alpha}\phi_{,\beta} - \tfrac{1}{2}\eta_{\alpha\beta}\eta^{\lambda\delta}\phi_{,\lambda}\phi_{,\delta} + \tfrac{1}{2}m^2\phi^2\eta_{\alpha\beta} \qquad (2.26)$$

from which one obtains for the Hamiltonian density

$$T_{tt} = \tfrac{1}{2}\left[(\partial_t\phi)^2 + \sum_{i=1}^{n-1}(\partial_i\phi)^2 + m^2\phi^2\right] \qquad (2.27)$$

and for the momentum density

$$T_{ti} = \partial_t\phi\partial_i\phi, \quad i = 1, \ldots, n-1, \qquad (2.28)$$

in terms of Minkowski coordinates.

Substituting ϕ from (2.17) into (2.27) and (2.28), and integrating over all space, yields

$$H \equiv \int_t T_{tt}\mathrm{d}^{n-1}x = \tfrac{1}{2}\sum_{\mathbf{k}}(a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + a_{\mathbf{k}}a_{\mathbf{k}}^\dagger)\omega \qquad (2.29)$$

$$P_i \equiv \int_t T_{ti}\mathrm{d}^{n-1}x = \sum_{\mathbf{k}} a_{\mathbf{k}}^\dagger a_{\mathbf{k}} k_i \qquad (2.30)$$

for the Hamiltonian and momentum component operators, respectively. Using the commutation relations (2.18), equation (2.29) may be recast in a more suggestive form

$$H = \sum_{\mathbf{k}}(a_{\mathbf{k}}^\dagger a_{\mathbf{k}} + \tfrac{1}{2})\omega. \qquad (2.31)$$

Clearly, both H and P_i commute with the operators

$$\left.\begin{aligned} N_{\mathbf{k}} &\equiv a_{\mathbf{k}}^\dagger a_{\mathbf{k}} \\ &\text{and} \\ N &\equiv \sum_{\mathbf{k}} N_{\mathbf{k}} \end{aligned}\right\} \qquad (2.32)$$

$$[N, H] = [N, P_i] = 0. \qquad (2.33)$$

The significance of N is revealed by taking its expectation values for the

Fock states. From (2.19) and (2.24) one obtains

$$\langle 0|N_{\mathbf{k}}|0\rangle = 0, \quad \forall \mathbf{k} \tag{2.34}$$

$$\langle {}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, {}^jn_{\mathbf{k}_j}|N_{\mathbf{k}_i}|{}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, {}^jn_{\mathbf{k}_j}\rangle = {}^in. \tag{2.35}$$

Thus, the expectation value of the operator $N_{\mathbf{k}_i}$ is the integer ${}^in$, that is, the entry in the ket vector under the label $\mathbf{k}_i$. Similarly, if $N_{\mathbf{k}_i}$ in (2.35) is summed over all i

$$\langle|N|\rangle = \sum_i {}^in. \tag{2.36}$$

This simple relationship between $N_{\mathbf{k}}$ and n suggests the name 'number operator for the mode **k**' for $N_{\mathbf{k}}$ and 'total number operator' for N. Because of relations (2.33), eigenstates of N are also eigenstates of H and **P**. For each increment of one in the number ${}^in$, $\langle|H|\rangle$ and $\langle|\mathbf{P}|\rangle$ increase by ω_i and $\mathbf{k}_i$ respectively. We can therefore interpret the ${}^in$ as labelling the *number of quanta*, each of energy ω_i and momentum $\mathbf{k}_i$, in the mode labelled by $\mathbf{k}_i$. Thus, the state $|{}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, {}^jn_{\mathbf{k}_j}\rangle$ is a state containing ${}^1n$ quanta in the mode with momentum $\mathbf{k}_1$, ${}^2n$ quanta in the mode with momentum $\mathbf{k}_2$ and so on.

Returning to (2.23) and (2.24), a useful physical interpretation is now available for the operators $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^\dagger$. The former reduces the number of quanta in mode **k** by one, while the latter increases this number by one. Thus $a_{\mathbf{k}}$ is referred to as an *annihilation operator* and $a_{\mathbf{k}}^\dagger$ as a *creation operator*, for quanta in the mode **k**.

2.4 Vacuum energy divergence

Special interest attaches to the state $|0\rangle$. This is the no-particle, or *vacuum state*. It carries zero momentum

$$\langle 0|\mathbf{P}|0\rangle = 0, \tag{2.37}$$

a result which follows immediately from (2.30). We should also expect it to carry zero energy, as no field quanta are present. However inspection of (2.31) reveals a term $\sum_{\mathbf{k}} \frac{1}{2}\omega$, so

$$\langle 0|H|0\rangle = \langle 0|0\rangle \sum_{\mathbf{k}} \tfrac{1}{2}\omega = \sum_{\mathbf{k}} \tfrac{1}{2}\omega \tag{2.38}$$

where we have used the normalization condition $\langle 0|0\rangle = 1$.

Not only is the right-hand side of (2.38) nonzero; it is actually infinite

$$\sum_{\mathbf{k}} \tfrac{1}{2}\omega = \tfrac{1}{2}(L/2\pi)^{n-1}\int \omega \mathrm{d}^{n-1}k$$
$$= (L^2/4\pi)^{(n-1)/2}\frac{1}{\Gamma((n-1)/2)}\int_0^\infty (k^2+m^2)^{\frac{1}{2}}k^{n-2}\mathrm{d}k \quad (2.39)$$

which diverges like k^n for large k. This divergence can be usefully analysed by performing the integral in (2.39) with n continued away from integral values to obtain

$$-L^{n-1}2^{-n-1}\pi^{-n/2}m^n\Gamma(-n/2).$$

The Γ-function contains poles at all even integral values of $n \geq 0$. This method of temporarily making divergent quantities finite by continuing the dimension of the spacetime away from integer values forms the basis of dimensional regularization (see chapter 6).

The fact that (2.39) is divergent apparently indicates that the vacuum contains an infinite density of energy. The trouble comes from the $\frac{1}{2}\omega$ zero-point energy associated with each simple harmonic oscillator mode of the scalar field. As ω has no upper bound the zero-point energy can be arbitrarily large. This is a problem which will plague the subject of quantum fields in curved spacetime throughout. However, in flat spacetime, it is easily circumvented. Energy as such is not measurable in non-gravitational physics, so we can rescale – or *renormalize* – the zero point of energy, even by an infinite amount, without affecting observable quantities. This may be accomplished by simply throwing away the $\frac{1}{2}\sum_{\mathbf{k}}\omega$ term in (2.31) or, more elegantly, by defining a *normal ordering* operation, denoted by : :, in which one demands that wherever a product of creation and annihilation operators appears, it is understood that all annihilation operators stand to the right of the creation operators. Thus, returning to the form (2.29), the normal ordering operation demands

$$:a_{\mathbf{k}}a_{\mathbf{k}}^{\dagger}: = a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}} \quad (2.40)$$

whence

$$:H: = \sum_{\mathbf{k}} a_{\mathbf{k}}^{\dagger}a_{\mathbf{k}}\omega \quad (2.41)$$

and the troublesome $\frac{1}{2}\omega$ term has disappeared.

Finally, let us return to $T_{\mu\nu}$ and expression (2.26). Using the expansion

(2.17) for ϕ in terms of the modes $u_{\mathbf{k}}$, we have

$$\phi_{,\alpha}\phi_{,\beta} = \sum_{\mathbf{k}} \sum_{\mathbf{k}'} (a_{\mathbf{k}}\partial_\alpha u_{\mathbf{k}} + a_{\mathbf{k}}^\dagger \partial_\alpha u_{\mathbf{k}}^*)(a_{\mathbf{k}'}\partial_\beta u_{\mathbf{k}'} + a_{\mathbf{k}'}^\dagger \partial_\beta u_{\mathbf{k}'}^*).$$

From (2.19) and the associated result

$$\langle 0 | a_{\mathbf{k}}^\dagger = 0 \tag{2.42}$$

together with

$$\langle 0 | a_{\mathbf{k}} a_{\mathbf{k}'}^\dagger | 0 \rangle = \delta_{\mathbf{k}\mathbf{k}'}$$

one obtains

$$\langle 0 | \phi_{,\alpha}\phi_{,\beta} | 0 \rangle = \sum_{\mathbf{k}} u_{\mathbf{k},\alpha} u_{\mathbf{k},\beta}^*.$$

In general

$$\langle 0 | T_{\alpha\beta} | 0 \rangle = \sum_{\mathbf{k}} T_{\alpha\beta}[u_{\mathbf{k}}, u_{\mathbf{k}}^*], \tag{2.43}$$

where $T_{\alpha\beta}[\phi, \phi]$ denotes the bilinear expression (2.26) for $T_{\alpha\beta}$. Similarly

$$\langle {}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, | T_{\alpha\beta} | {}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots \rangle = \sum_{\mathbf{k}} T_{\alpha\beta}[u_{\mathbf{k}}, u_{\mathbf{k}}^*] + 2\sum_i {}^i n T_{\alpha\beta}[u_{\mathbf{k}_i}, u_{\mathbf{k}_i}^*]. \tag{2.44}$$

2.5 Dirac spinor field

So far attention has been restricted to the scalar (spin-zero) field. The quantization of higher-spin fields proceeds in close analogy. In particular the spin $\frac{1}{2}$ fermion field ψ possesses the Lagrangian density

$$\mathscr{L} = \tfrac{1}{2}\mathrm{i}(\bar{\psi}\gamma^\alpha\psi_{,\alpha} - \bar{\psi}_{,\alpha}\gamma^\alpha\psi) - m\bar{\psi}\psi \tag{2.45}$$

where $\bar{\psi}$ is the Dirac adjoint of ψ (i.e., $\psi^\dagger\gamma^0$), and γ^μ are Dirac matrices that satisfy the anticommutation relations

$$\{\gamma^\alpha, \gamma^\beta\} = 2\eta^{\alpha\beta}. \tag{2.46}$$

(See, for example, Delbourgo & Prasad 1974 for the properties of γ-matrices in n-dimensions.)

Variation of $\bar{\psi}$ in the action $S = \int \mathscr{L}\,\mathrm{d}^n x$ leads to the Dirac equation

$$\mathrm{i}\gamma^\alpha\psi_{,\alpha} - m\psi = 0 \tag{2.47}$$

for a particle of mass m. The field ψ carries spinor labels (ψ^a) which have been suppressed in the above expressions. Similarly we suppress labels on the Dirac matrices (γ_{ab}).

A complete set of mode solutions of the Dirac equation is given (in discrete normalization) by

$$\left.\begin{aligned} u_{\mathbf{k},s}(t,\mathbf{x}) &= N u(\mathbf{k},s)\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}-\mathrm{i}\omega t} \\ v_{\mathbf{k},s}(t,\mathbf{x}) &= N v(\mathbf{k},s)\mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}+\mathrm{i}\omega t}, \end{aligned}\right\} \tag{2.48}$$

where

$$N = \begin{cases} (m/\omega L^{n-1})^{\frac{1}{2}}, & m \neq 0 \\ (2\omega L^{n-1})^{-\frac{1}{2}}, & m = 0. \end{cases} \tag{2.49}$$

The familiar constant positive and negative energy spinors $u(\mathbf{k},s), v(\mathbf{k},s)$ (see, for example, Bjorken & Drell 1965, chapters 3 and 10) exist in two independent spin states and are normalized according to

$$u^\dagger(\mathbf{k},s)u(\mathbf{k},s') = v^\dagger(\mathbf{k},s)v(\mathbf{k},s') = \begin{cases} (\omega/m)\delta_{ss'}, & m \neq 0 \\ 2\omega\delta_{ss'}, & m = 0. \end{cases} \tag{2.50}$$

We can thus expand the field ψ as

$$\psi(t,\mathbf{x}) = \sum_{\pm s}\sum_{\mathbf{k}}[b_{\mathbf{k}}(s)u_{\mathbf{k},s}(t,\mathbf{x}) + d_{\mathbf{k}}^\dagger(s)v_{\mathbf{k},s}(t,\mathbf{x})] \tag{2.51}$$

which is normalized with respect to the inner product

$$(\psi,\phi) = \int_t \mathrm{d}^{n-1}x\,\bar{\psi}(t,\mathbf{x})\gamma_0\phi(t,\mathbf{x}). \tag{2.52}$$

The operators $b_{\mathbf{k}}(s)$, $d_{\mathbf{k}}(s)$, $b_{\mathbf{k}}^\dagger(s)$, $d_{\mathbf{k}}^\dagger(s)$ all anticommute except in the following cases:

$$\{b_{\mathbf{k}}(s), b_{\mathbf{k}'}^\dagger(s')\} = \{d_{\mathbf{k}}(s), d_{\mathbf{k}'}^\dagger(s')\} = \delta_{ss'}\delta_{\mathbf{k}\mathbf{k}'}. \tag{2.53}$$

By constructing Hamiltonian, momentum and angular momentum operators and considering their expectation values in the Fock basis, one finds that $b_{\mathbf{k}}^\dagger(s)$ is the creation operator for quanta in a mode of momentum $\mathbf{k}$, energy ω and spin s, while $d_{\mathbf{k}}^\dagger(s)$ annihilates quanta in a mode of momentum $(-\mathbf{k})$, energy $(-\omega)$, spin s. Physically $b_{\mathbf{k}}^\dagger(s)$ and $d_{\mathbf{k}}^\dagger(s)$ represent creation operators for electrons and positrons respectively, while $b_{\mathbf{k}}(s)$, $d_{\mathbf{k}}(s)$ are the corresponding annihilation operators.

More detailed examination of the stress-tensor for the Dirac field is postponed until its treatment in curved spacetime in §3.8.

2.6 Electromagnetic field

The electromagnetic (massless, spin 1) field is described by the Lagrangian density

$$\mathcal{L} = -\tfrac{1}{4}F_{\alpha\beta}F^{\alpha\beta} \tag{2.54}$$

where

$$F_{\alpha\beta} = A_{\alpha,\beta} - A_{\beta,\alpha} \tag{2.55}$$

is the Maxwell field strength tensor. Variation of the action $S = \int \mathcal{L}\,\mathrm{d}^n x$ then yields

$$F^{\alpha\beta}{}_{,\beta} = 0, \tag{2.56}$$

which along with the identity

$$F_{\alpha\beta,\gamma} + F_{\beta\gamma,\alpha} + F_{\gamma\alpha,\beta} = 0 \tag{2.57}$$

constitutes Maxwell's equations.

The field strength tensor (2.55) and hence the Lagrangian (2.54) are invariant under local gauge transformations

$$A_\alpha \to A_\alpha^\Lambda = A_\alpha + \partial_\alpha \Lambda(x) \tag{2.58}$$

where $\Lambda(x)$ is an arbitrary differentiable scalar function. This gauge invariance prevents the straightforward quantization of the theory and must be broken, usually by adding to the Lagrangian density a gauge-fixing term

$$\mathcal{L}_G = -\tfrac{1}{2}\zeta^{-1}(A^\alpha{}_{,\alpha})^2, \tag{2.59}$$

where ζ is a parameter determining the choice of gauge; $\zeta = 1$ being the Feynman gauge and $\zeta \to 0$ the Landau gauge.

The field equation resulting from the inclusion of the gauge-breaking term is

$$[\eta_{\alpha\beta}\Box - (1 - \zeta^{-1})\partial_\alpha\partial_\beta]A^\beta = 0. \tag{2.60}$$

In the Feynman gauge these equations reduce to

$$\Box A_\alpha = 0 \tag{2.61}$$

with solution

$$A^\alpha(t, \mathbf{x}) = \sum_{\mathbf{k}\lambda} [a_{\mathbf{k}\lambda} u^\alpha_{\mathbf{k}\lambda}(t, \mathbf{x}) + a^\dagger_{\mathbf{k}\lambda} u^{\alpha}_{\mathbf{k}\lambda}{}^*(t, \mathbf{x})], \tag{2.62}$$

where the plane wave modes $u^{\alpha}_{\mathbf{k}\lambda}$ are given by

$$u^{\alpha}_{\mathbf{k}\lambda}(t, \mathbf{x}) = (2L^{n-1}\omega)^{-\frac{1}{2}} e^{\alpha}_{\mathbf{k}\lambda} e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t}, \tag{2.63}$$

with $e^{\alpha}_{\mathbf{k}\lambda}$, $\lambda = 1, 2, 3, 4$ labelling independent polarization vectors associated with mode $\mathbf{k}$. These vectors can be chosen to form an orthonormal system with

$$\eta_{\alpha\beta} e^{\alpha}_{\mathbf{k}\lambda} e^{\beta}_{\mathbf{k}\lambda'} = \eta_{\lambda\lambda'}. \tag{2.64}$$

The quantization of the fields now proceeds in much the same way as for the scalar field; however, because the physical photon has only two independent (transverse) degrees of freedom, while the Lorentz covariant choice of gauge used here has kept all four potentials A^{α} on an equal footing, the results so obtained do not immediately make physical sense. To obtain a physical interpretation of the quantized theory, the so-called Gupta–Bleuler formalism can be used (Gupta 1950, Bleuler 1950; see also the textbook treatments of Jauch & Rohrlich 1955, Bogolubov & Shirkov 1959 or Schweber 1961). We shall not adopt this formalism here, but shall instead treat the quantization of the electromagnetic field by the path-integral approach, which introduces a new feature, i.e., the so-called ghost fields, that assume considerable importance in the curved spacetime treatment. Before turning to this, however, the subject of Green functions must be considered.

2.7 Green functions

Vacuum expectation values of various products of free field operators can be identified with various Green functions of the wave equation. Treating first the scalar field, of particular importance are the expectation values of the commutator and anticommutator of the fields, denoted respectively by

$$iG(x, x') = \langle 0|[\phi(x), \phi(x')]|0\rangle \tag{2.65}$$

$$G^{(1)}(x, x') = \langle 0|\{\phi(x), \phi(x')\}|0\rangle. \tag{2.66}$$

G is known as the Pauli–Jordan or Schwinger function while $G^{(1)}$ is sometimes called Hadamard's elementary function. These Green functions can be split into their positive and negative frequency parts as

$$\left.\begin{aligned} iG(x, x') &= G^{+}(x, x') - G^{-}(x, x') \\ G^{(1)}(x, x') &= G^{+}(x, x') + G^{-}(x, x') \end{aligned}\right\} \tag{2.67}$$

where $G^{\pm}$, known as the Wightman functions, are given by

$$\left.\begin{aligned} G^{+}(x,x') &= \langle 0|\phi(x)\phi(x')|0\rangle \\ G^{-}(x,x') &= \langle 0|\phi(x')\phi(x)|0\rangle. \end{aligned}\right\} \tag{2.68}$$

The Feynman propagator G_{F} is defined as the time-ordered product of fields

$$\begin{aligned} \mathrm{i}G_{\mathrm{F}}(x,x') &= \langle 0|T(\phi(x)\phi(x'))|0\rangle \\ &= \theta(t-t')G^{+}(x,x') + \theta(t'-t)G^{-}(x,x') \end{aligned} \tag{2.69}$$

where

$$\theta(t) = \begin{cases} 1, & t>0 \\ 0, & t<0. \end{cases}$$

Finally, retarded and advanced Green functions are defined respectively by

$$\left.\begin{aligned} G_{\mathrm{R}}(x,x') &= -\theta(t-t')G(x,x') \\ G_{\mathrm{A}}(x,x') &= \theta(t'-t)G(x,x') \end{aligned}\right\} \tag{2.70}$$

and their average is denoted as

$$\bar{G}(x,x') = \tfrac{1}{2}[G_{\mathrm{R}}(x,x') + G_{\mathrm{A}}(x,x')], \tag{2.71}$$

which is related to G_{F} by

$$G_{\mathrm{F}}(x,x') = -\bar{G}(x,x') - \tfrac{1}{2}\mathrm{i}G^{(1)}(x,x'). \tag{2.72}$$

Using the field equation (2.1) it is clear that G, $G^{(1)}$, $G^{\pm}$ all satisfy the homogeneous equation

$$(\Box_x + m^2)\mathscr{G}(x,x') = 0. \tag{2.73}$$

Also, using $\partial_t\theta(t-t') = \delta(t-t')$ and the equal time commutators (2.15) one obtains the following equations for G_{F}, G_{R} and G_{A}:

$$(\Box_x + m^2)G_{\mathrm{F}}(x,x') = -\delta^n(x-x') \tag{2.74}$$

$$(\Box_x + m^2)G_{\mathrm{R,A}}(x,x') = \delta^n(x-x'). \tag{2.75}$$

The Green functions $G_{\mathrm{F,R,A}}$ describe the propagation of field disturbances subject to certain boundary conditions.

Integral representations for the Green functions can be obtained by substituting the mode decomposition (2.17) for ϕ into the definitions of the Green functions as vacuum expectation values. One finds that all the Green

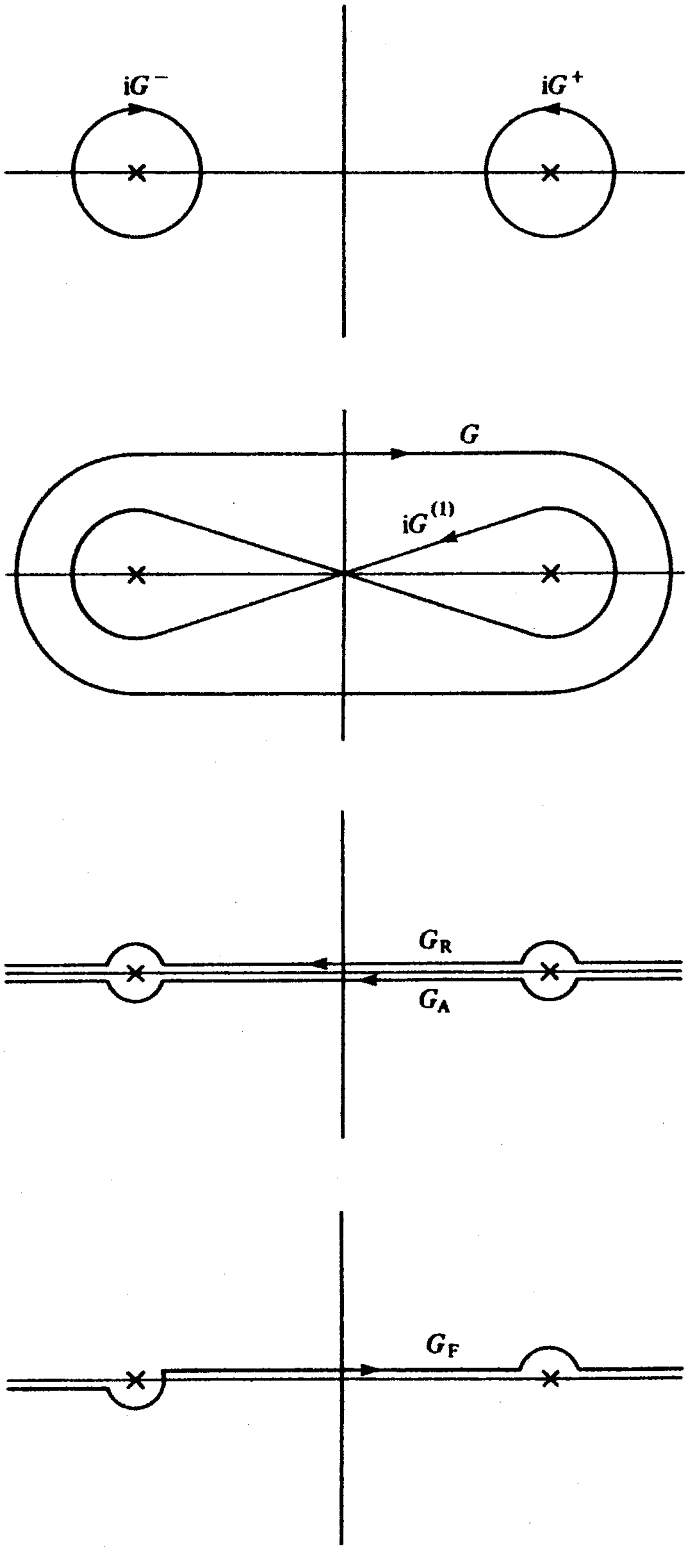


Fig. 3. The various Green functions are each associated with the above contours in the complex k^0 plane. The poles on the real axis at $k^0 = \pm(|\mathbf{k}|^2 + m^2)^{\frac{1}{2}}$ are marked with crosses. The open contours should be envisaged as closed by infinite semicircles in the upper- or lower-half planes.

functions can be represented as

$$\mathscr{G}(x, x') = (2\pi)^{-n} \int \frac{\exp[\mathrm{i}\mathbf{k}\cdot(\mathbf{x} - \mathbf{x}') - \mathrm{i}k^0(t - t')]}{(k^0)^2 - |\mathbf{k}|^2 - m^2} \mathrm{d}^n k. \tag{2.76}$$

The integral has poles at $k^0 = \pm(|\mathbf{k}|^2 + m^2)^{\frac{1}{2}}$. Considered as a contour integral, the k^0 integration may be performed by deforming the contour around the poles. The way in which this deformation is performed (see fig. 3) depends on the boundary conditions on the field and determines which of the various Green functions is obtained from (2.76).

For example, the integral corresponding to G_F yields

$$G_\mathrm{F}(x, x') = \frac{-\mathrm{i}\pi}{(4\pi\mathrm{i})^{n/2}} \left(\frac{2m^2}{-\sigma + \mathrm{i}\varepsilon} \right)^{(n-2)/4} H^{(2)}_{\frac{1}{2}n-1}\{[2m^2(\sigma - \mathrm{i}\varepsilon)]^{\frac{1}{2}}\} \tag{2.77}$$

where $\sigma = \frac{1}{2}(x - x')^2 = \frac{1}{2}\eta_{\alpha\beta}(x^\alpha - x'^\alpha)(x^\beta - x'^\beta)$ and $H^{(2)}$ is a Hankel function of the second kind. The $-\mathrm{i}\varepsilon$ is added to σ to indicate that G_F is really the boundary value of a function which is analytic in the lower-half σ plane.

In the massless limit the Green functions are customarily denoted by D rather than G. In this limit the Feynman propagator in four dimensions reduces to

$$D_\mathrm{F}(x, x') = (\mathrm{i}/8\pi^2\sigma) - (1/8\pi)\delta(\sigma) \tag{2.78}$$

while Hadamard's elementary function becomes

$$D^{(1)}(x, x') = -1/4\pi^2\sigma. \tag{2.79}$$

For the fields of spin $\frac{1}{2}$ and 1 we note here only a few of the Green functions. For Dirac spinors one defines

$$\mathrm{i}S_\mathrm{F}(x, x') = \langle 0|T(\psi(x)\bar{\psi}(x'))|0\rangle \tag{2.80}$$

$$S^{(1)}(x, x') = \langle 0|[\psi(x), \bar{\psi}(x')]|0\rangle \tag{2.81}$$

which satisfy

$$(\mathrm{i}\gamma^\alpha \partial_\alpha - m)S_\mathrm{F}(x, x') = \delta^n(x - x'), \tag{2.82}$$

$$(\mathrm{i}\gamma^\alpha \partial_\alpha - m)S^{(1)}(x, x') = 0 \tag{2.83}$$

and can be written in terms of G_F and $G^{(1)}$ as

$$S_\mathrm{F}(x, x') = (\mathrm{i}\gamma^\alpha \partial_\alpha + m)G_\mathrm{F}(x, x') \tag{2.84}$$

$$S^{(1)}(x, x') = -(\mathrm{i}\gamma^\alpha \partial_\alpha + m)G^{(1)}(x, x'). \tag{2.85}$$

Note that, for example, S_F is a matrix in the spinor indices of the fields in

(2.80), which when written in full reads

$$\mathrm{i}S_{\mathrm{F}}(x,x')_{ab} = \langle 0|\psi_a(x)\psi_c^{\dagger}(x')|0\rangle(\gamma^0)^c{}_b\theta(t-t') - \langle 0|\psi_c^{\dagger}(x')\psi_a(x)|0\rangle(\gamma^0)^c{}_b\theta(t'-t).$$

The Feynman propagator for the electromagnetic field is defined by

$$\mathrm{i}D_{\mathrm{F}\alpha\beta}(x,x') = \langle 0|T(A_\alpha(x)A_\beta(x'))|0\rangle \tag{2.86}$$

which is obviously gauge dependent. Using (2.60) one finds that

$$[\eta_{\alpha\lambda}\Box_x - (1-\zeta^{-1})\partial_\alpha^x\partial_\lambda^x]D_{\mathrm{F}}^{\lambda\beta}(x,x') = \delta_\alpha^\beta\delta^n(x-x') \tag{2.87}$$

which gives an integral representation for the propagator:

$$D_{\mathrm{F}\alpha\beta}(x,x') = (2\pi)^{-n}\int\frac{[-\eta_{\alpha\beta}+(1-\zeta)k_\alpha k_\beta/k^2]}{(k^0)^2-|\mathbf{k}|^2}\exp[\mathrm{i}\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')-\mathrm{i}k^0(t-t')]\mathrm{d}^n k. \tag{2.88}$$

The contour for the k^0 integral in (2.88) is the same as for G_{F} (see fig. 3). In particular, in the Feynman gauge ($\zeta = 1$)

$$D_{\mathrm{F}\alpha\beta}(x,x') = -\eta_{\alpha\beta}D_{\mathrm{F}}(x,x'). \tag{2.89}$$

Note also that if the gauge-breaking term is removed by letting $\zeta\to\infty$ then (2.88) becomes infinite. That is, unless the gauge-breaking term is present the differential operator on the left-hand side of (2.87) is not invertible.

Examination of the integral representations of the Feynman Green functions, such as (2.76) and (2.88), reveals a useful mathematical property that is frequently exploited in practical calculations. Inspection of the contour for G_{F} in the complex k^0 plane shows that the topological relation between the contour (assumed closed by an infinite semicircle) and the poles remains unchanged if it is rotated anticlockwise through 90^0 to lie along the imaginary k^0 axis from $-\mathrm{i}\infty$ to $\mathrm{i}\infty$. If the integration variable k^0 is now changed to $\kappa = -\mathrm{i}k^0$ and the variables $\tau = -\mathrm{i}t$ and $\tau' = -\mathrm{i}t'$ replace t and t', then the contour of integration will once again run along the real axis but will no longer intersect the poles.

For example, in the scalar case one obtains

$$G_{\mathrm{F}}(t,\mathbf{x};t',\mathbf{x}') = -\mathrm{i}G_{\mathrm{E}}(\mathrm{i}\tau,\mathbf{x};\mathrm{i}\tau',\mathbf{x}') \tag{2.90}$$

where

$$G_{\mathrm{E}}(\tau,\mathbf{x};\tau',\mathbf{x}') = (2\pi)^{-n}\int_{-\infty}^{\infty}\frac{\exp[\mathrm{i}\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')+\mathrm{i}\kappa(\tau-\tau')]}{\kappa^2+|\mathbf{k}|^2+m^2}\mathrm{d}\kappa\,\mathrm{d}^{n-1}k. \tag{2.91}$$

G_E is the 'Euclidean' Green function which satisfies

$$(\Box_x - m^2)G_E(x, x') = -\delta^n(x - x'). \tag{2.92}$$

In (2.92), $\Box$ is the elliptic operator

$$\frac{\partial^2}{\partial \tau^2} + \frac{\partial^2}{\partial (x^1)^2} + \ldots + \frac{\partial^2}{\partial (x^{n-1})^2},$$

which is the d'Alembertian on an n-dimensional *Euclidean* space, rather than on Minkowski space. This corresponds to considering the properties of the field ϕ in Euclidean space.

The advantage of Euclidean field theory is that the elliptic operator has a unique, well-defined inverse, because the poles in the integral representation (2.91) lie on the imaginary rather than the real axis. Hence, it is often mathematically convenient to work in Euclidean space, and 'rotate' back to pseudo-Euclidean spacetime, using (2.90), at the end of the calculation. The boundary conditions for the Feynman propagator are automatically imposed by this procedure. (Note that none of the other contours in fig. 3 can be so rotated without intersecting the poles.) For a more detailed discussion (in the context of curved spacetimes) see Candelas & Raine (1977*a*) and Wald (1979*b*).

The Green functions introduced so far have all been calculated as expectation values of products of field operators in a pure state, namely the vacuum state. These Green functions are suitable for describing systems at zero temperature. However a system at nonzero temperature is not described by a pure state but one that is statistically distributed over all such states. The Green functions for systems at nonzero temperature are thus given by the average over all pure states of the expectation value of the products of field operators in those pure states (Kadanoff & Baym 1962, Abrikosov, Gorkov & Dzyaloskinskii 1963, Mattuck 1967, Fetter & Walecka 1971).

Suppose that $|\psi_i\rangle$ is a pure state, being an eigenstate of the Hamiltonian (2.41) with energy eigenvalue E_i. Then it will also be an eigenstate of the total number operator N of (2.32) with number eigenvalue n_i say. Since both the number of particles and the energy are variable, an equilibrium system at temperature T is described by a grand canonical ensemble of states. The probability that the system will be in the state $|\psi_i\rangle$ is given by

$$\rho_i = e^{-\beta(E_i - \mu n_i)}/Z \tag{2.93}$$

where

$$\beta = 1/k_B T \tag{2.94}$$

k_B being Boltzmann's constant, μ the chemical potential,

$$Z = \sum_j e^{-\beta(E_j - \mu n_j)} = e^{-\beta\Omega} \tag{2.95}$$

the grand partition function, and Ω the thermodynamic potential. The ensemble average at a temperature $T = (k_B\beta)^{-1}$ of any operator A is thus

$$\langle A \rangle_\beta = \sum_i \rho_i \langle \psi_i | A | \psi_i \rangle. \tag{2.96}$$

Introducing the quantum *density operator* defined by

$$\rho = \exp[\beta(\Omega + \mu N - H)] \tag{2.97}$$

then

$$\rho_i = \langle \psi_i | \rho | \psi_i \rangle. \tag{2.98}$$

The requirement of unit total probability becomes

$$\operatorname{tr} \rho \equiv \sum_i \langle \psi_i | \rho | \psi_i \rangle = 1, \tag{2.99}$$

and (2.96) reduces to

$$\langle A \rangle_\beta = \operatorname{tr} \rho A. \tag{2.100}$$

Now we can define nonzero temperature Green functions (also called thermal or temperature Green functions) simply by replacing the vacuum expectation values in the definitions of zero-temperature Green functions by the ensemble average $\langle\ \rangle_\beta$. For example, from (2.68) we define in the case of scalar fields

$$\left.\begin{aligned} G_\beta^+(x, x') &= \langle \phi(x)\phi(x') \rangle_\beta \\ G_\beta^-(x, x') &= \langle \phi(x')\phi(x) \rangle_\beta. \end{aligned}\right\} \tag{2.101}$$

Assuming for now that the chemical potential vanishes we have the following important property of these thermal Green functions:

$$G_\beta^\pm(t, \mathbf{x}; t', \mathbf{x}') = G_\beta^\mp(t + i\beta, \mathbf{x}; t', \mathbf{x}'). \tag{2.102}$$

This relation is obtained from the Heisenberg equations of motion

$$\phi(t, \mathbf{x}) = e^{iH(t - t_0)}\phi(t_0, \mathbf{x})e^{-iH(t - t_0)} \tag{2.103}$$

as follows:

$$\begin{aligned} G_\beta^+(t, \mathbf{x}; t', \mathbf{x}') &= \operatorname{tr}[e^{-\beta H}\phi(t, \mathbf{x})\phi(t', \mathbf{x}')]/\operatorname{tr}(e^{-\beta H}) \\ &= \operatorname{tr}[e^{-\beta H}\phi(t, \mathbf{x})e^{\beta H}e^{-\beta H}\phi(t', \mathbf{x}')]/\operatorname{tr}(e^{-\beta H}) \end{aligned}$$

$$= \operatorname{tr}[\phi(t+\mathrm{i}\beta,\mathbf{x})\mathrm{e}^{-\beta H}\phi(t',\mathbf{x}')]/\operatorname{tr}(\mathrm{e}^{-\beta H})$$

$$= \operatorname{tr}[\mathrm{e}^{-\beta H}\phi(t',\mathbf{x}')\phi(t+\mathrm{i}\beta,\mathbf{x})]/\operatorname{tr}(\mathrm{e}^{-\beta H})$$

$$= G_\beta^-(t+\mathrm{i}\beta,\mathbf{x};t',\mathbf{x}') \tag{2.104a}$$

and similarly for G_β^-. In arriving at the above result we have used the property $\operatorname{tr} AB = \operatorname{tr} BA$. From (2.102) similar properties for the other Green functions can be obtained; for example, from (2.67) it follows that

$$G_\beta^{(1)}(t,\mathbf{x};t',\mathbf{x}') = G_\beta^{(1)}(t+\mathrm{i}\beta,\mathbf{x};t',\mathbf{x}'). \tag{2.104b}$$

Had the chemical potential been retained it would merely have introduced the factor $\mathrm{e}^{\beta\mu}$ on the right-hand side. Note, however, that

$$\mathrm{i}G_\beta(x,x') = \mathrm{i}G(x,x') = [\phi(x),\phi(x')], \tag{2.105}$$

because the commutator of a free scalar field is a c-number (this follows from (2.18)) and thus its statistical and vacuum expectation values are equal. This will not in general be the case in interacting theories where the (unequal time) commutator can be an operator.

Using (2.105) and relations such as (2.104) for the other Green functions we can write integral representations for the thermal Green functions. Starting with G_β we can use (2.105) to write its Fourier transform as

$$\mathrm{i}G(x,x') = \mathrm{i}G_\beta(x,x') = (1/2\pi)\int_{-\infty}^{\infty} \mathrm{d}\omega\, c(\omega;x,x')\mathrm{e}^{-\mathrm{i}\omega(t-t')} \tag{2.106}$$

in which $c(\omega;x,x')$ can easily be calculated from (2.76) with the appropriate contour:

$$c(\omega;x,x') = (2\pi)^{1-n}\int \mathrm{d}^{n-1}k\,\delta(\omega^2-|\mathbf{k}|^2-m^2)[\theta(\omega)-\theta(-\omega)]\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot(\mathbf{x}-\mathbf{x}')}. \tag{2.107}$$

If we also write the Fourier transform of $G_\beta^\pm$ as

$$G_\beta^\pm(x,x') = (1/2\pi)\int_{-\infty}^{\infty} \mathrm{d}\omega\, g^\pm(\omega)\mathrm{e}^{-\mathrm{i}\omega(t-t')} \tag{2.108}$$

then, from (2.67),

$$c(\omega) = g^+(\omega) - g^-(\omega). \tag{2.109}$$

The relation (2.104a) implies that

$$g^+(\omega) = \mathrm{e}^{\beta\omega}g^-(\omega)$$

which, when used in conjunction with (2.109) yields

$$g^{\pm}(\omega) = \pm c(\omega)(1 - e^{\mp\beta\omega})^{-1}$$

giving the integral representations

$$G_{\beta}^{\pm}(x, x') = \pm \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} \frac{c(\omega)}{1 - e^{\mp\beta\omega}} e^{-i\omega(t-t')}. \tag{2.110}$$

From this equation the representations of other Green functions can be obtained. In particular, by explicitly evaluating the integrals, making use of expansions of the factors $(1 - e^{\mp\beta\omega})^{-1}$ in powers of $e^{\mp\beta\omega}$, one finds that

$$G_{\beta}^{(1)}(t, \mathbf{x}; t', \mathbf{x}') = \sum_{k=-\infty}^{\infty} G^{(1)}(t + ik\beta, \mathbf{x}; t', \mathbf{x}'). \tag{2.111}$$

That is, the thermal Green function can be written as an infinite imaginary-time image sum of the corresponding zero-temperature Green function.

In the spin $\frac{1}{2}$ case, because the fields anticommute rather than commute one finds that $S_{\beta}^{(1)}$, the nonzero temperature version of (2.81), satisfies an antiperiodicity condition

$$S_{\beta}^{(1)}(t, \mathbf{x}; t', \mathbf{x}') = -S_{\beta}^{(1)}(t + i\beta, \mathbf{x}, t', \mathbf{x}') \tag{2.112}$$

rather than the periodicity condition of (2.104). One then obtains an image sum similar to (2.111), but reflecting the antiperiodicity by including a factor $(-1)^k$:

$$S_{\beta}^{(1)}(t, \mathbf{x}; t', \mathbf{x}') = \sum_{k=-\infty}^{\infty} (-1)^k S^{(1)}(t + ik\beta, \mathbf{x}; t', \mathbf{x}'). \tag{2.113}$$

The treatment of commuting, spin 1 fields is similar to the scalar case (see, for example, Brown & Maclay 1969).

2.8 Path-integral quantization

The canonical quantization scheme outlined in §2.1 is only one of several approaches to quantum field theory. One could, for example, have started with the covariant commutation relations

$$[\phi(x), \phi(x')] = iG(x, x') \tag{2.114}$$

(recall (2.65), noting that G is a c-number), in place of the canonical commutation relations (2.15). The scheme (2.114) has the advantage of being closer to the spirit of general relativity because it does not single out a particular time t. It leads immediately to the same commutation relations (2.18) for the creation and annihilation operators. In a globally hyperbolic spacetime (see §3.1) the covariant and canonical approaches are equivalent.

Another quantization technique, especially well suited to the rigorous treatment of the functional analysis that is encountered in quantum field theory, is the C^* algebra approach of Segal (1967) (for a review of the relevance of this approach to curved spacetime, see Isham 1978*a*).

Finally, the path-integral quantization of Feynman (Feynman & Hibbs 1965) is a powerful approach to quantum gravity and to the quantization of interacting fields, with their attendant problems of renormalization. As we shall have occasion to use this technique in later chapters, we shall give a short review in this section. More detailed treatments can be found in Rzewuski (1969), Abers & Lee (1973), Taylor (1976), Frampton (1977), Nash (1978) and Itzykson & Zuber (1980).

The basic object of the theory is the functional integral for a field ϕ with action S

$$Z[J] = \langle \text{out}, 0|0, \text{in}\rangle = \int \mathscr{D}[\phi] \exp\{iS[\phi] + i\int d^n x\, J(x)\phi(x)\} \quad (2.115)$$

taken over the space of functions ϕ with an appropriate measure. The quantity Z is known as the generating functional for the theory and gives the transition amplitude from the initial vacuum $|0, \text{in}\rangle$ to the final vacuum $|0, \text{out}\rangle$ in the presence of a source $J(x)$ which is producing particles. When the source is switched off the two vacua reduce to the usual source-free Minkowski space vacuum $|0\rangle$ and one has

$$Z[0] = \langle 0|0\rangle \quad (2.116)$$

(usually normalized to unity).

It follows by functional differentiation of Z with respect to J that

$$i^j\langle 0|T(\phi(x_1)\ldots\phi(x_j))|0\rangle_c = \left(\frac{\delta^j \ln Z}{\delta J(x_1)\ldots\delta J(x_j)}\right)_{J=0} \quad (2.117)$$

giving the connected, time-ordered Green functions of the theory (c denoting that only connected Feynman diagrams are included in perturbation theory).

As an example, we consider the case of the free scalar field with

$$S[\phi] = \int d^n x[\mathscr{L}_0(x) + \tfrac{1}{2}i\varepsilon\phi^2(x)]$$

where $\mathscr{L}_0$ is the free field Lagrangian (2.2) and the infinitesimal factor (which is related to the boundary conditions on ϕ) can be used to make the functional integral convergent. Substituting (2.2) for $\mathscr{L}_0$ and integrating by parts, the action becomes

$$S[\phi] = \int \mathrm{d}^n x[-\tfrac{1}{2}\phi(\Box + m^2 - \mathrm{i}\varepsilon)\phi] \tag{2.118}$$

where we have discarded a boundary term. Using (2.118) the exponent in (2.115) can be written suggestively in the form

$$-\tfrac{1}{2}\int \mathrm{d}^n x\, \mathrm{d}^n y\, \phi(x)K_{xy}\phi(y) + \int J(x)\phi(x)\mathrm{d}^n x \tag{2.119}$$

where the symmetric operator

$$K_{xy} = (\Box_x + m^2 - \mathrm{i}\varepsilon)\delta^n(x-y) \tag{2.120}$$

can formally be treated as a symmetric matrix K with continuous indices x, y, having the properties

$$\int \mathrm{d}^n y\, K^{\frac{1}{2}}_{xy}K^{\frac{1}{2}}_{yz} = K_{xz} \tag{2.121}$$

$$\int \mathrm{d}^n y\, K^{\frac{1}{2}}_{xy}K^{-\frac{1}{2}}_{yz} = \delta^n(x-z) \tag{2.122}$$

$$K^{-1}_{xy} = -G_{\mathrm{F}}(x,y), \tag{2.123}$$

the last result following by inverting the definition of the Feynman propagator (2.74). These properties become well defined within the context of the functional integral $\mathscr{D}[\phi]$.

Changing the integration variable from ϕ to

$$\phi'(x) = \int \mathrm{d}^n y\, K^{\frac{1}{2}}_{xy}\phi(y), \tag{2.124}$$

the quadratic form (2.119) may be recast, using (2.121)–(2.123), as

$$-\tfrac{1}{2}\int \mathrm{d}^n x\left[\phi'(x) - \int \mathrm{d}^n y\, J(y)K^{-\frac{1}{2}}_{yx}\right]^2 - \tfrac{1}{2}\int \mathrm{d}^n x\, \mathrm{d}^n y\, J(x)G_{\mathrm{F}}(x,y)J(y). \tag{2.125}$$

When (2.125) is substituted into the exponent in the functional integral, the second term is independent of ϕ and can be removed from the integral, while the first term yields an integral of the Gaussian type, and can be performed to give a numerical factor.

Thus

$$Z(J) \propto (\det K^{\frac{1}{2}})^{-1} \exp\left[-\tfrac{1}{2}\mathrm{i}\int \mathrm{d}^n x\, \mathrm{d}^n y\, J(x) G_{\mathrm{F}}(x, y) J(y)\right] \quad (2.126)$$

where

$$(\det K^{\frac{1}{2}})^{-1} = [\det(-G_{\mathrm{F}})]^{\frac{1}{2}} = \exp[\tfrac{1}{2}\operatorname{tr}\ln(-G_{\mathrm{F}})] \quad (2.127)$$

is the Jacobian arising from the change of variable (2.124).

It follows by inspection of (2.126) that, for example

$$\left(\frac{\delta^2 \ln Z}{\delta J(x)\delta J(y)}\right)_{J=0} = -\langle 0|T(\phi(x)\phi(y))|0\rangle = -\mathrm{i}G_{\mathrm{F}}(x, y)$$

in agreement with the definition (2.69).

For the spin $\frac{1}{2}$ field ψ, the generating functional Z is taken to be

$$Z(\eta, \bar{\eta}) = \int \mathscr{D}[\psi]\mathscr{D}[\bar{\psi}] \exp\left\{\mathrm{i}\int \mathrm{d}^n x [\mathscr{L}_0(x) + \bar{\eta}(x)\psi(x) + \eta(x)\bar{\psi}(x)]\right\} \quad (2.128)$$

where $\eta, \bar{\eta}$ are anticommuting external currents and $\mathscr{L}_0$ is given by (2.45). In place of (2.126) one obtains

$$Z(\eta, \bar{\eta}) \propto (\det S_{\mathrm{F}})^{-1} \exp\left[-\mathrm{i}\int \mathrm{d}^n x\, \bar{\eta}(x) S_{\mathrm{F}}(x, y)\eta(y)\right]. \quad (2.129)$$

The case of the electromagnetic field runs into complications associated with the gauge symmetry. To see this, one examines the analogue of (2.115)

$$Z(J) = \int \mathscr{D}[A_\alpha] \exp\left\{\mathrm{i}\int \mathrm{d}^n x [\mathscr{L}_0(x) + J^\beta(x) A_\beta(x)]\right\} \quad (2.130)$$

with $\mathscr{L}_0(x)$ given by (2.54). The action in the exponent of (2.130) can be written

$$\int \mathrm{d}^n x\, \mathscr{L}_0(x) = -\tfrac{1}{4}\int F_{\alpha\beta}F^{\alpha\beta}\mathrm{d}^n x = -\tfrac{1}{2}\int \mathrm{d}^n x\, \mathrm{d}^n y\, A_\alpha(x) K^{\alpha\beta}_{xy} A_\beta(y), \quad (2.131)$$

$$K^{\alpha\beta}_{xy} = (\eta^{\alpha\beta}\Box_x - \partial^\alpha_x \partial^\beta_x)\delta^n(x - y). \quad (2.132)$$

Inspection of (2.87) and (2.88) in the limit $\zeta \to \infty$ shows that $K^{-1} = D_{\mathrm{F}}(\zeta \to \infty)$ is singular. Thus, as we shall see below, quantization of the electromagnetic field based on the above naive description cannot proceed. This difficulty was mentioned in connection with gauge invariance on page 19, where a gauge-breaking term (2.59) was introduced into the Lagrangian (this term is removed by taking the $\zeta \to \infty$ limit) which yields an invertible wave operator, and Green function (2.88), for finite ζ.

Because $\mathscr{L}_0$ is manifestly gauge-invariant, it is independent of the longitudinal and timelike components of A_α, i.e., K projects out the transverse field components. Hence, any variation of the longitudinal and timelike components of A_α will leave $\mathscr{L}_0$ unchanged. More generally, variation of A_α related to the original A_α by the gauge transformation (2.58) will leave $\mathscr{L}_0$ unchanged.

If one envisages the space of all functions A_α, then under the action of the gauge transformation (2.58), a point (i.e., function A_α) in the space will be mapped into points along a line, representing other functions related to the original A_α by the continuous gauge transformation. This line is called the *orbit* of the gauge group associated with (2.58). Problems with the path integral occur because variations of A_μ in $\mathscr{D}[A_\alpha]$ that lie *along* these orbits do not induce any variation in $\mathscr{L}_0(x)$. To see this, note that the functional integration in (2.130) is over an infinite volume of function space, so in order for the integral to converge it is necessary for $\mathscr{L}_0(x) \to \infty$ as $A_\alpha \to \infty$ to produce a declining exponential. However, when $A_\alpha \to \infty$ along the gauge orbits, $\mathscr{L}_0$ remains constant (and hence finite) and fails to produce the required convergence. As the volume of the 'orbit subspace' is also infinite, the functional integral is undefined, unless a way can be found to renormalize it by dividing out this infinite volume.

One way to cure this ill was pointed out by Fadeev & Popov (1967), following earlier work of Feynman (1963) and DeWitt (1965, 1967*b*). Variations in A_α are required to be restricted to functions belonging to distinct orbits. This can be achieved by choosing a 'hypersurface' in the space of A_α that intersects each orbit only once. Then rather than integrating over the whole space, one integrates only over the hypersurface.

The equation for such a hypersurface may be written

$$F[A_\alpha] = 0, \tag{2.133}$$

so it might be supposed that one need merely insert $\delta[F(A_\alpha)]$ into the integrand of (2.130). However, to ensure a gauge-invariant result, more care is needed. Two neighbouring hypersurfaces will be related by

$$F(A_\alpha^\Lambda(x)) = F(A_\alpha(x)) + \int \mathrm{d}^n y M_{xy} \Lambda(y) + \mathrm{O}(\Lambda^2) \tag{2.134}$$

where Λ parametrizes the gauge transformation, $F(A_\alpha) = [F(A_\alpha^\Lambda)]_{\Lambda=0}$, and M depends on the choice of gauge. For example, in the Landau gauge

$F(A_\alpha) = A^\alpha{}_{,\alpha}$, so

$$F(A^\Lambda_\alpha) = A^\alpha{}_{,\alpha} + \Box\Lambda = F(A_\alpha) + \Box\Lambda \tag{2.135}$$

from which it follows that

$$M_{xy} = \Box\delta^n(x-y). \tag{2.136}$$

It is convenient to define the quantity $\Delta_F(A_\alpha)$ by

$$\Delta_F^{-1}(A_\alpha) = \int \mathscr{D}[\Lambda]\delta[F(A^\Lambda_\alpha)] \tag{2.137}$$

which is easily shown to be gauge invariant. When restricted to the hypersurface defined by (2.133), the first term on the right of (2.134) vanishes so, symbolically, $F(A^\Lambda_\alpha) = M\Lambda$. The integral in (2.137) can then be performed immediately, by changing the variable from Λ to $M\Lambda$:

$$\Delta_F^{-1}(A_\alpha) = (\det M)^{-1} = \exp(-\operatorname{tr}\ln M) \tag{2.138}$$

$(\det M)^{-1}$ being the Jacobean arising from this change of variable.

Returning to the afflicted path integral (2.130) (with J set to zero) we may insert, using (2.137), the unit operator $\Delta_F(A_\alpha)\int\mathscr{D}[\Lambda]\delta[F(A^\Lambda_\alpha)]$ into the integrand without changing anything. Exploiting the gauge invariance of Δ_F we may change its argument to A^Λ_α, and then with a change of integration variable the path integral may be written

$$\int\mathscr{D}[A_\alpha]\int\mathscr{D}[\Lambda]\Delta_F(A_\alpha)\delta[F(A_\alpha)]\exp\left[\mathrm{i}\int\mathscr{L}_0(x)\mathrm{d}^nx\right] \tag{2.139}$$

from which it is obvious that the integrand of the $\mathscr{D}[\Lambda]$ integral is independent of Λ. Hence we may factor it out. Although infinite, this integral is independent of the fields A_α, so it may simply be divided out from (2.139) to obtain a new definition of Z that does not suffer from singularity problems:

$$Z = \int\mathscr{D}[A_\alpha]\delta[F(A_\alpha)]\exp\left\{\mathrm{i}\int\mathrm{d}^nx[\mathscr{L}_0(x) + J^\alpha(x)A_\alpha(x) - \mathrm{i}\operatorname{tr}\ln M]\right\}. \tag{2.140}$$

In arriving at (2.140) we have used the fact that the factor $\delta[F(A_\alpha)]$ in the integrand of (2.139) restricts Δ_F to the hypersurface (2.133), enabling us to use the result (2.138).

The restriction of the integration to the hypersurface $F(A_\alpha) = 0$ does not, therefore, merely introduce a factor $\delta[F(A_\alpha)]$ into the integrand, but also introduces an extra term into the field action. This additional contribution can be regarded as due to an additional fictitious field. In fact, it is not difficult to show that in the Landau gauge

$$\int \mathscr{D}[c]\mathscr{D}[c^*]\exp\left(\mathrm{i}\int \eta^{\alpha\beta}\partial_\alpha c\partial_\beta c^*\right) = \exp(-\operatorname{tr}\ln M) \qquad (2.141)$$

where c and c^* are massless scalar fields, but satisfying anticommutation relations, for which reason they are called 'Fadeev–Popov ghost fields'. As they do not couple to the vector fields A_α, the ghost fields are frequently ignored in flat spacetime quantum field theory. In curved spacetime, however, they play an important role.

The vector part of the path integral may be further reduced in the Landau gauge by also expressing $\delta[F(A_\alpha)] = \delta(A^\alpha{}_{,\alpha})$ as an exponential:

$$\delta(A^\alpha{}_{,\alpha}) = \lim_{\zeta\to 0}\exp\left[-\frac{\mathrm{i}}{2\zeta}\int (A^\alpha{}_{,\alpha})^2\mathrm{d}^n x\right]. \qquad (2.142)$$

Comparison with (2.59) reveals that the exponent here is simply $\mathrm{i}\mathscr{L}_G$ in the Landau ($\zeta\to 0$) gauge. Hence the vector generating functional is

$$\int \mathscr{D}[A_\alpha]\exp\left[\mathrm{i}\int \mathrm{d}^n x(\mathscr{L}_0 + \mathscr{L}_\mathrm{G})\right]. \qquad (2.143)$$

This functional integral may be evaluated along the same lines as the scalar case and is found to be proportional to

$$\exp\left[-\tfrac{1}{2}\mathrm{i}\int \mathrm{d}^n x\, \mathrm{d}^n y\, J_\alpha(x) D_\mathrm{F}^{\alpha\beta}(x-y) J_\beta(y)\right] \qquad (2.144)$$

where $D_\mathrm{F}^{\alpha\beta}$ is given from (2.88)

Although the total generating functional, (2.144) multiplied by (2.141), has been derived in the Landau gauge $\zeta\to 0$, it remains true for any ζ.

We end this section with some brief remarks about the convergence of the path integral (2.115). In general, because the action S is real, the exponent in the integrand is pure imaginary. Hence the integral is generally not properly defined when taken over the whole function space. In the foregoing this was ameliorated by the careful use of the iε factor. As remarked in §2.7, the employment of iε to define the Feynman Green functions is equivalent to passing to an imaginary time description, in which the field is defined in Euclidean space rather than Minkowski space. If the vacuum expectation values of the time-ordered field products, such as (2.117), are analytic in the complex t plane, then one can construct a field quantization in Euclidean space, in which the path integrals assume a well-defined, strongly convergent Gaussian form (at least for a wide class of Lagrangians), and recover the Minkowski space theory at the end by 'rotating' back from it to t. This technique is frequently used in practice. When passing to curved spacetime, however, it may happen that no 'Euclideanized' (i.e., positive definite metric) spacetime exists that corresponds to the original pseudo-Riemannian spacetime.

3

Quantum field theory in curved spacetime

The basic formalism of quantum field theory is generalized to curved spacetime in this chapter, in a straightforward way. The discussion is preceded by a very brief summary of pseudo-Riemannian geometry. The treatment is in no way intended to be complete, and we refer the reader to Weinberg (1972), Hawking & Ellis (1973), or Misner, Thorne & Wheeler (1973) for further details. Readers unfamiliar with conformal transformations and Penrose conformal diagrams are advised to read §3.1 carefully, however.

The basic generalization of the particle concept to curved spacetime is readily accomplished. What is not so easy is the physical interpretation of the formalism so developed. There has, in fact, been a certain amount of controversy over the meaning – and meaningfulness – of the particle concept when a background gravitational field is present. In some cases, such as for static spacetimes, the concept seems well defined, while in others (e.g. spacetimes that admit closed timelike world lines or do not everywhere possess Cauchy surfaces) the notion of particle can seem hopelessly obscure. We restrict consideration to 'well-behaved' spacetimes, and do not embark upon a philosophical discourse about the meaning of particles. Instead we relate the formalism directly to what an actual particle detector might be expected to register in the particular quantum state of interest. It is in this concrete operational sense that we define particles in curved spacetime. Although this approach has been studied before, we give the most developed treatment of particle detectors so far.

Building upon an explicit example of particle creation by a changing background gravitational field, we present a detailed and in-depth analysis of adiabaticity – the definition of particles in a quasi-static spacetime. It is here that one expects to make contact with standard quantum field theory, for we know that to be a good approximation in the (relatively) slowly expanding universe that we inhabit. The treatment reveals that the high frequency behaviour of the field is independent of the quantum state or the global structure of the spacetime, and depends purely on the local geometry. This turns out to be of crucial significance for the regularization and renormalization programme dealt with in later chapters.

We give what is intended to be a complete treatment of the concept of adiabatic states in §3.5. As a result, this section requires careful reading. What is essential is the application of the adiabatic limit to the Feynman propagator, dealt with in §3.6. This culminates in the so-called DeWitt–Schwinger representation of G_F, given by the expansion (3.141). Its rôle in renormalization theory is so crucial that the importance of this expansion cannot be overemphasized. We recommend that §3.5 be read briefly in the first instance, with attention concentrated on the physical remarks, and re-read in depth when the subsequent applications have been examined.

The section on the conformal vacuum, §3.7, is much easier to follow, and relatively important for later applications, especially to concrete calculations of cosmological particle creation and quantum vacuum stress.

A final section on higher-spin fields in curved spacetime is merely a summary of standard formalism, though the expressions for the stress–energy–momentum tensors (3.190)–(3.195) will be frequently used.

3.1 Spacetime structure

We assume spacetime to be a C^∞ n-dimensional, globally hyperbolic, pseudo-Riemannian manifold (for more detailed discussion see Hawking & Ellis 1973). These conditions may be more restrictive than necessary to construct a viable quantum field theory (for example, see Avis, Isham & Storey 1978 for quantum field theory in a non-globally hyperbolic spacetime). The differentiability conditions ensure the existence of differential equations and the global hyperbolicity ensures the existence of Cauchy hypersurfaces.

The pseudo-Riemannian metric $g_{\mu\nu}$ associated with the line element

$$ds^2 = g_{\mu\nu}(x)dx^\mu dx^\nu, \quad \mu, \nu = 0, 1, \ldots, (n-1)$$

has signature $n-2$. Several coordinate patches with associated $g_{\mu\nu}$ may be needed to cover the entire manifold. We define the determinant

$$g \equiv |\det g_{\mu\nu}|.$$

We shall frequently make use of Penrose conformal diagrams (Penrose 1964) for depicting the causal structure of spacetime. This is a device that enables the whole of an infinite spacetime to be represented as a finite diagram (compact manifold), by applying a *conformal transformation* to the metric structure. Conformal transformations, which shrink or stretch the manifold, must be distinguished from coordinate transformations $x^\mu \to x'^\mu$

which merely relabel the coordinates in some patch, leaving the geometry itself unchanged. A conformal transformation of the metric may be described by

$$g_{\mu\nu}(x) \to \bar{g}_{\mu\nu}(x) = \Omega^2(x) g_{\mu\nu}(x) \tag{3.1}$$

for some continuous, non-vanishing, finite, real function $\Omega(x)$.

From such a transformation of the metric, the conformal transformation properties of various other quantities can be derived. For example, the Christoffel symbol, Ricci tensor and Ricci scalar transform respectively as

$$\Gamma^{\rho}{}_{\mu\nu} \to \bar{\Gamma}^{\rho}{}_{\mu\nu} = \Gamma^{\rho}{}_{\mu\nu} + \Omega^{-1}(\delta^{\rho}_{\mu}\Omega_{;\nu} + \delta^{\rho}_{\nu}\Omega_{;\mu} - g_{\mu\nu}g^{\rho\alpha}\Omega_{;\alpha}) \tag{3.2}$$

$$\begin{aligned} R^{\nu}{}_{\mu} \to \bar{R}^{\nu}{}_{\mu} = {} & \Omega^{-2}R^{\nu}{}_{\mu} - (n-2)\Omega^{-1}(\Omega^{-1})_{;\mu\rho}g^{\rho\nu} \\ & + (n-2)^{-1}\Omega^{-n}(\Omega^{n-2})_{;\rho\sigma}g^{\rho\sigma}\delta^{\nu}_{\mu} \end{aligned} \tag{3.3}$$

$$\begin{aligned} R \to \bar{R} = {} & \Omega^{-2}R + 2(n-1)\Omega^{-3}\Omega_{;\mu\nu}g^{\mu\nu} \\ & + (n-1)(n-4)\Omega^{-4}\Omega_{;\mu}\Omega_{;\nu}g^{\mu\nu} \end{aligned} \tag{3.4}$$

from which one obtains the following useful transformation:

$$\begin{aligned} [\Box + \tfrac{1}{4}(n-2)R/(n-1)]\phi \to {} & [\bar{\Box} + \tfrac{1}{4}(n-2)\bar{R}/(n-1)]\bar{\phi} \\ & = \Omega^{-(n+2)/2}[\Box + \tfrac{1}{4}(n-2)R/(n-1)]\phi \end{aligned} \tag{3.5}$$

where

$$\Box\phi = g^{\mu\nu}\nabla_{\mu}\nabla_{\nu}\phi = (-g)^{-\frac{1}{2}}\partial_{\mu}[(-g)^{\frac{1}{2}}g^{\mu\nu}\partial_{\nu}\phi] \tag{3.6}$$

$$\bar{\phi}(x) \equiv \Omega^{(2-n)/2}(x)\phi(x). \tag{3.7}$$

As a simple illustration of a Penrose diagram, consider two-dimensional Minkowski space, which has the line element

$$\mathrm{d}s^2 = \mathrm{d}t^2 - \mathrm{d}x^2. \tag{3.8}$$

We frequently work in terms of null coordinates, u, v, defined by

$$\left.\begin{aligned} u &= t - x \\ v &= t + x \end{aligned}\right\} \tag{3.9}$$

in which the line element (3.8) becomes

$$\mathrm{d}s^2 = \mathrm{d}u\,\mathrm{d}v \tag{3.10}$$

so that

$$g_{\mu\nu} = \tfrac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{3.11}$$

Suppose we perform the *coordinate transformation*

$$\left.\begin{aligned} u' &= 2\tan^{-1}u \\ v' &= 2\tan^{-1}v \end{aligned}\right\} \tag{3.12}$$

where

$$-\pi \leq u',\ v' \leq \pi. \tag{3.13}$$

Then from (3.10)

$$ds^2 = \tfrac{1}{4}\sec^2\tfrac{1}{2}u' \sec^2\tfrac{1}{2}v' du' dv' \tag{3.14}$$

so that

$$g_{\mu\nu}(u', v') = \tfrac{1}{8}\sec^2\tfrac{1}{2}u' \sec^2\tfrac{1}{2}v' \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \tag{3.15}$$

If we now perform a *conformal transformation* with

$$\Omega^2(x) = (\tfrac{1}{4}\sec^2\tfrac{1}{2}u' \sec^2\tfrac{1}{2}v')^{-1}$$

then

$$g_{\mu\nu}(u', v') \to \bar{g}_{\mu\nu}(u', v') = \tfrac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{3.16}$$

and the conformally-related line element is given by

$$d\bar{s}^2 = du' dv'. \tag{3.17}$$

This has the same form as the original Minkowski space line element (3.10), but only covers the compact region (3.13) depicted in fig. 4. The effect of the conformal transformation (3.16) has been to shrink in infinity to the boundary lines on the diagram.

The boundary of the figure contains several features of interest. First note that all null rays remain at 45° in the Penrose diagram: conformal transformations leave the null cones invariant. Thus any causal analysis may proceed with the null rays drawn as is usual in Minkowski space. Clearly all null rays will terminate on the diagonal boundary lines labelled $\mathscr{I}^+$ and $\mathscr{I}^-$ and called future and past null infinity respectively. Asymptotically timelike lines converge on points marked i^+ (future timelike infinity) and i^- (past timelike infinity). Similarly, asymptotically spacelike lines converge on i^0 (spacelike infinity).

The analysis here applies also to four-dimensional Minkowski space if each point on the diagram is considered as a 2-sphere, except for points on

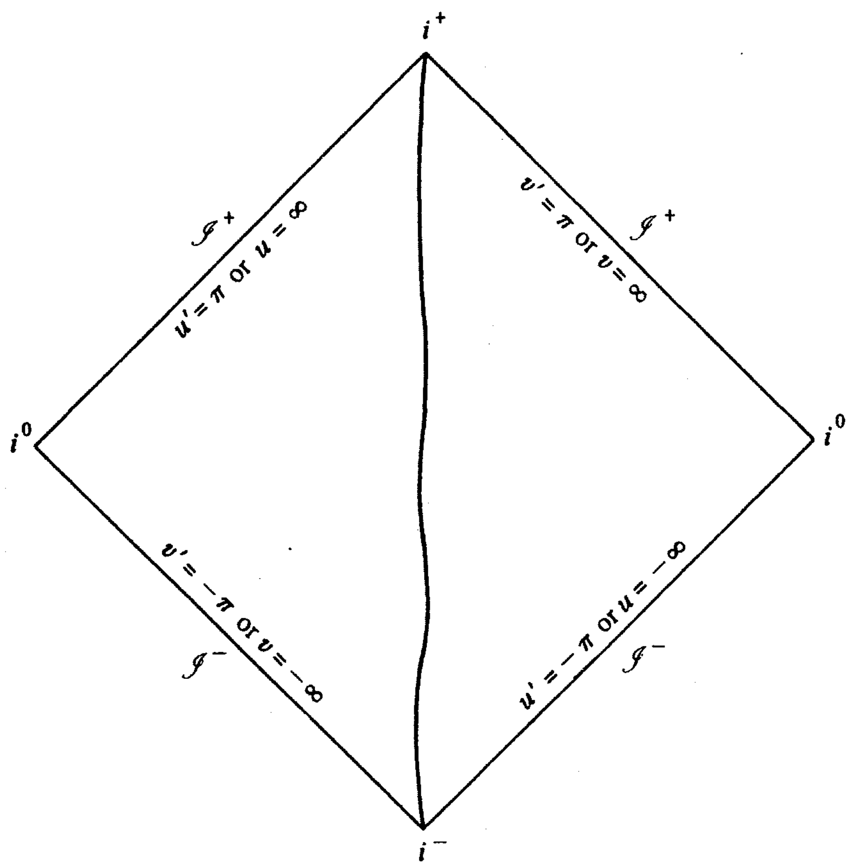


Fig. 4. Penrose conformal diagram of Minkowski space. The compact region $-\pi \leq u', v' \leq \pi$ is conformal to the whole of Minkowski space $(-\infty \leq u, v \leq \infty)$. Null rays $u, v =$ constant remain at 45°. The world line of an asymptotically timelike observer is shown.

the vertical symmetry axis and i^0, which represent spacetime points. Thus $\mathscr{I}^+$ and $\mathscr{I}^-$ are really null 3-surfaces.

A timelike line that is asymptotically null may occur, such as if a particle undergoes uniform acceleration for all time, approaching the speed of light as $t \to \infty$. The world line for such a particle is drawn in fig. 5. The null asymptote has the property that events that lie above it are forever causally inaccessible to the accelerated particle. That is, those events cannot communicate with the particle (though the converse may not be true). The null asymptote is therefore an *event horizon* for the accelerated particle, though not for an unaccelerated particle.

As another illustration of Penrose diagrams, we shall consider four-dimensional Schwarzschild spacetime, described by the line element

$$ds^2 = (1 - 2M/r)dt^2 - (1 - 2M/r)^{-1}dr^2 - r^2(d\theta^2 + \sin^2\theta \, d\phi^2). \quad (3.18)$$

This spacetime is the unique spherically symmetric vacuum solution of

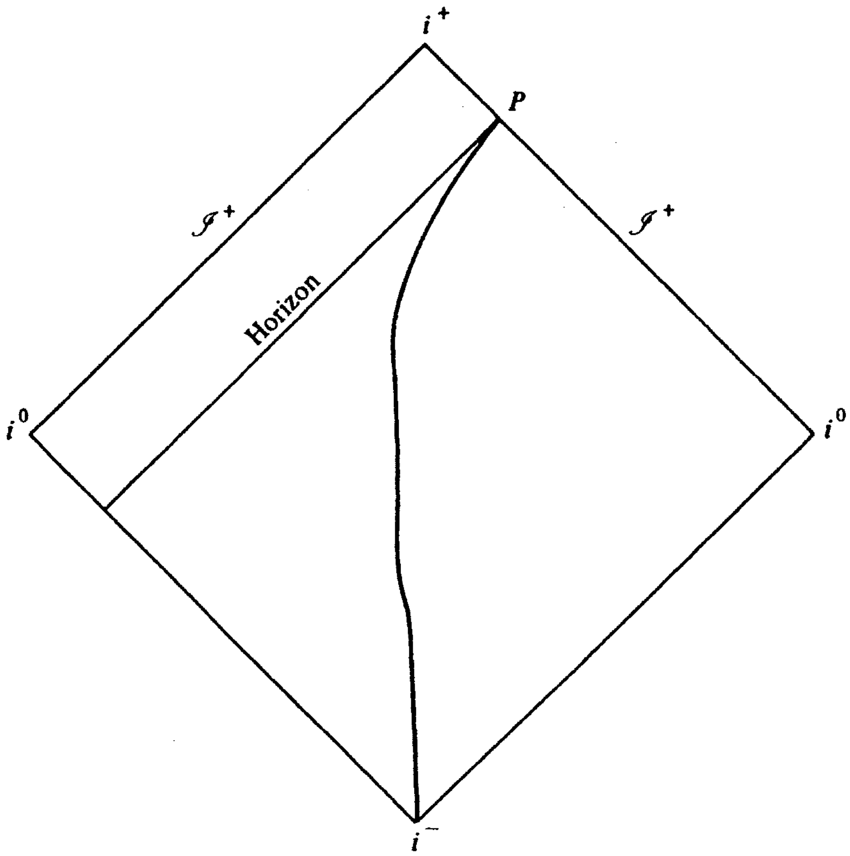


Fig. 5. The timelike world line represents an observer who accelerates continuously to the right, so that he asymptotically approaches the speed of light. This world line does not, therefore, reach i^+, but intersects $\mathscr{I}^+$ at P. The backward null ray through P thus acts as an event horizon, because events above it can never be witnessed by the observer: all null rays from such events intersect $\mathscr{I}^+$ between P and i^+.

Einstein's equation and is often used to represent the empty space region surrounding a spherical star or collapsing body of mass M.

Transforming to *Kruskal* coordinates defined (Kruskal 1960) by

$$\left.\begin{aligned} \bar{u} &= -4M\,\mathrm{e}^{-u/4M} \\ \bar{v} &= 4M\,\mathrm{e}^{v/4M} \end{aligned}\right\} \tag{3.19}$$

where $u = t - r^*, v = t + r^*$ and $r^* = r + 2M \ln|(r/2M) - 1|$, (3.18) becomes

$$ds^2 = (2M/r)\mathrm{e}^{-r/2M}\mathrm{d}\bar{u}\,\mathrm{d}\bar{v} - r^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2). \tag{3.20}$$

The $\mathrm{d}\bar{u}\mathrm{d}\bar{v}$ part of this metric is conformal to two-dimensional Minkowski space described by (3.10), which in turn may be compactified by the coordinate and conformal transformations (3.12) and (3.16). The resulting diagram is therefore identical to fig. 4. However, the left-hand edge of the diamond is not $\mathscr{I}$ in this case, for it follows from (3.19) that $\bar{u}, \bar{v}$ are defined only in the quadrant $-\infty < \bar{u} \leq 0, 0 \leq \bar{v} < \infty$. The left-hand edge therefore comprises segments $\bar{u} = 0$ and $\bar{v} = 0$, or $r = 2M$, $t = \pm\infty$. This is a singularity of the u, v(or t, r) coordinate system, but not the $\bar{u}, \bar{v}$ system, as may be seen by comparing the metrics (3.18) and (3.20). Therefore, the spacetime may be analytically extended beyond the left-hand edge, using the Kruskal coordinates $\bar{u}, \bar{v}$ defined over the whole plane $-\infty < \bar{u}, \bar{v} < \infty$. The resulting Penrose diagram is shown in fig. 6 and the spacetime is known as the maximally extended Kruskal manifold.

The horizontal zig-zag lines represent $r = 0$ in the past and future. This is a singularity in the Kruskal metric (3.20) and, in fact, in the spacetime, as may be seen from inspection of the Riemann tensor. The manifold cannot be analytically continued across these edges. There is another asymptotically Minkowskian spacetime region marked II, to the left of the diagram, containing the $\mathscr{I}$ surfaces $\bar{u} = +\infty, \bar{v} = -\infty$. This half of the manifold is geometrically identical to the right-hand diamond marked I, except that the time direction is formally reversed ($t \to -t$).

The null ray $\bar{u} = 0$ ($r = 2M, t = +\infty$) is the latest retarded null ray to reach $\mathscr{I}^+$: for $\bar{u} > 0$, all the null rays intersect the future singularity, $r = 0$. The ray $\bar{u} = 0$ (representing an outward directed 2-surface in the full four-dimensional picture) is therefore an event horizon for observers restricted to region I. Similarly $\bar{v} = 0$ ($r = 2M, t = -\infty$) is an event horizon for observers restricted to region II. The empty regions III and IV are thus black holes for these respective sets of observers (and white holes for their opposite numbers). The world lines of observers who avoid the black and white holes converge on i^+ and i^- in their respective regions. The existence of an event horizon prevents any communication between regions I and II.

The maximally extended Kruskal spacetime is everywhere (except at $r = 0$) a solution of the vacuum Einstein equation. In the real world, where a black hole is more likely to form from the implosion of a star, the vacuum equations only apply to the region outside the star and so only a fragment of fig. 6 will be relevant. Nevertheless, as we shall see in chapter 8, the difference as far as observers in region I are concerned is negligible at late times ($t \to \infty$), after the implosion has occurred.

Frequently our analysis will be restricted to spacetimes with special geometrical symmetries. These can be described using Killing vectors ξ^μ, which are solutions of Killing's equation

$$\mathscr{L}_\xi g_{\mu\nu}(x) = 0 \tag{3.21}$$

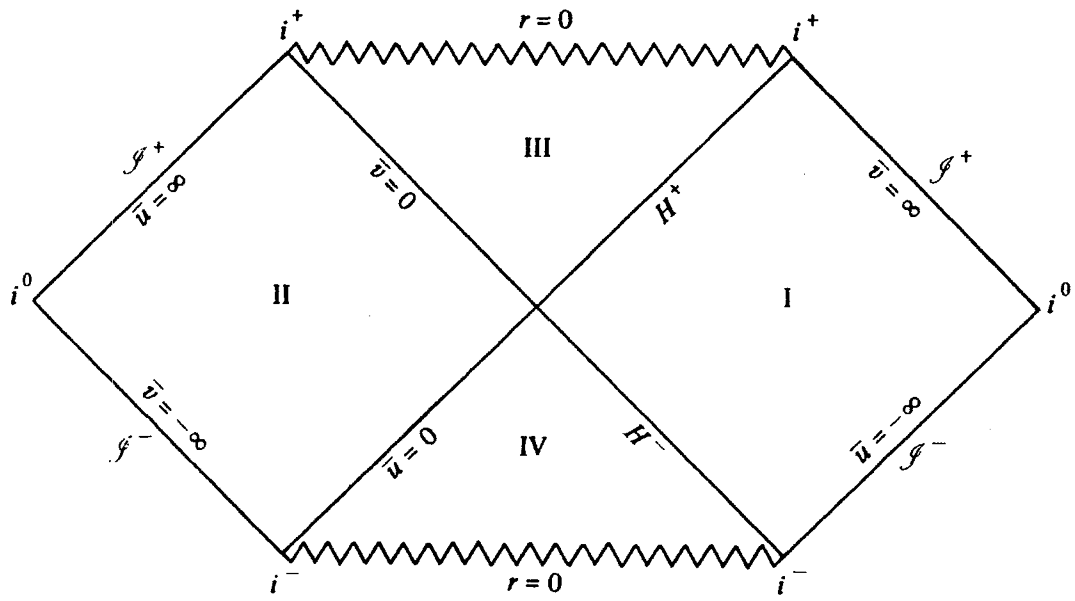


Fig. 6. Penrose diagram of the maximally extended Kruskal manifold. The physical singularity is marked by a wavy line. The future and past horizons, labelled $H^\pm$ respectively, are represented by the null rays $\bar{u}=0$, $\bar{v}=0$. Observers who do not fall into the black hole (or emerge from the white hole) are restricted to either of the diamond shaped regions marked I and II.

where $\mathscr{L}_\xi$ is the Lie derivative along the vector field ξ^μ. Equation (3.21) can be written as

$$\xi_{\mu;\nu} + \xi_{\nu;\mu} = 0. \tag{3.22}$$

We shall also be interested in the symmetries associated with conformal flatness, when the spacetime is conformal to Minkowski space. The geometry then admits a *conformal* Killing vector field, which satisfies the conformal generalization of (3.21):

$$\mathscr{L}_\xi g_{\mu\nu}(x) = \lambda(x) g_{\mu\nu}(x) \tag{3.23}$$

where $\lambda(x)$ is some (non-singular, non-vanishing) scalar function.

3.2 Scalar field quantization

Formally, field quantization in curved spacetime proceeds in close analogy to the Minkowski space case. We start with Lagrangian density (see §3.8)

$$\mathscr{L}(x) = \tfrac{1}{2}[-g(x)]^{\frac{1}{2}}\{g^{\mu\nu}(x)\phi(x)_{,\mu}\phi(x)_{,\nu} - [m^2 + \xi R(x)]\phi^2(x)\} \tag{3.24}$$

where $\phi(x)$ is the scalar field and m the mass of the field quanta. The coupling between the scalar field and the gravitational field represented by the term $\xi R\phi^2$, where ξ is a numerical factor and $R(x)$ is the Ricci scalar curvature, is included as the only possible local, scalar coupling of this sort with the correct dimensions. The resulting action is

$$S = \int \mathscr{L}(x)\mathrm{d}^n x \tag{3.25}$$

where n is the spacetime dimension. Setting the variation of the action with respect to ϕ equal to zero yields the scalar field equation

$$[\Box_x + m^2 + \xi R(x)]\phi(x) = 0 \tag{3.26}$$

where now $\Box$ is given by (3.6).

Two values of ξ are of particular interest: the so-called minimally coupled case, $\xi = 0$, and the conformally coupled case

$$\xi = \tfrac{1}{4}[(n-2)/(n-1)] \equiv \xi(n). \tag{3.27}$$

In this latter case, if $m = 0$ the action and hence the field equations are invariant under conformal transformations (3.1) if the field is assumed to transform as in (3.7). Indeed, from (3.5) it is clear that if

$$[\Box + \tfrac{1}{4}(n-2)R/(n-1)]\phi = 0,$$

then also

$$[\bar{\Box} + \tfrac{1}{4}(n-2)\bar{R}/(n-1)]\bar{\phi} = 0.$$

The scalar product (2.9) is generalized to

$$(\phi_1, \phi_2) = -\mathrm{i}\int_\Sigma \phi_1(x)\overleftrightarrow{\partial}_\mu \phi_2^*(x)[-g_\Sigma(x)]^{\frac{1}{2}}\mathrm{d}\Sigma^\mu \tag{3.28}$$

where $\mathrm{d}\Sigma^\mu = n^\mu\,\mathrm{d}\Sigma$, with n^μ a future-directed unit vector orthogonal to the spacelike hypersurface Σ and $\mathrm{d}\Sigma$ is the volume element in Σ. The hypersurface Σ is taken to be a Cauchy surface in the (globally hyperbolic) spacetime and one can show, using Gauss' theorem (see Hawking & Ellis 1973, §2.8), that the value of (ϕ_1, ϕ_2) is independent of Σ.

There exists a complete set of mode solutions $u_i(x)$ of (3.26) which are orthonormal in the product (3.28): i.e., satisfying

$$(u_i, u_j) = \delta_{ij}, \quad (u_i^*, u_j^*) = -\delta_{ij}, \quad (u_i, u_j^*) = 0. \tag{3.29}$$

The index i schematically represents the set of quantities necessary to label the modes. The field ϕ may be expanded as in (2.17):

$$\phi(x) = \sum_i [a_i u_i(x) + a_i^\dagger u_i^*(x)]. \tag{3.30}$$

The covariant quantization of the theory is implemented by adopting the commutation relations (2.18):

$$[a_i, a_j^\dagger] = \delta_{ij}, \text{ etc.} \tag{3.31}$$

The construction of a vacuum state, Fock space, etc. can then proceed exactly as described for the Minkowski space case in §2.1. This time, however, there is an inherent ambiguity in the formalism (Fulling 1973). In Minkowski space there is a natural set of modes, namely (2.11), that are closely associated with the natural rectangular coordinate system (t, x, y, z). In turn, these natural coordinates are associated with the Poincaré group, the action of which leaves the Minkowski line element unchanged. Specifically, the vector $\partial/\partial t$ is a Killing vector of Minkowski space, orthogonal to the spacelike hypersurfaces $t =$ constant, and the modes (2.11) are eigenfunctions of this Killing vector with eigenvalues $-i\omega$ for $\omega > 0$ (positive frequency). The vacuum is invariant under the action of the Poincaré group.

In curved spacetime the Poincaré group is no longer a symmetry group of the spacetime (see in this connection Urbantke 1969). Indeed, in general there will be no Killing vectors at all with which to define positive frequency modes. In some special classes of spacetime there may be symmetry under certain restricted transformations, for example rotations or translations, or the de Sitter group. In these cases, there may exist 'natural' coordinates associated with the Killing vectors – analogues of rectangular coordinates in Minkowski space. But even if such coordinates do exist, we shall see that they do not enjoy the same central physical status in quantum field theory as their Minkowski space counterparts. In general, though, no such privileged coordinates are available and no natural mode decomposition of ϕ based on the separation of the wave equation (3.26) in these privileged coordinates will present itself. Indeed, the whole spirit of general relativity, expressed through the principle of general covariance, is that coordinate systems are physically irrelevant.

Consider, therefore, a second complete orthonormal set of modes $\bar{u}_j(x)$. The field ϕ may be expanded in this set also

$$\phi(x) = \sum_j [\bar{a}_j \bar{u}_j(x) + \bar{a}_j^\dagger \bar{u}_j^*(x)]. \tag{3.32}$$

This decomposition of ϕ defines a new vacuum state $|\bar{0}\rangle$:

$$\bar{a}_j|\bar{0}\rangle = 0, \quad \forall j \tag{3.33}$$

and a new Fock space.

As both sets are complete, the new modes $\bar{u}_j$ can be expanded in terms of the old:

$$\bar{u}_j = \sum_i (\alpha_{ji} u_i + \beta_{ji} u_i^*). \tag{3.34}$$

Conversely

$$u_i = \sum_j (\alpha_{ji}^* \bar{u}_j - \beta_{ji} \bar{u}_j^*). \tag{3.35}$$

These relations are known as Bogolubov transformations (Bogolubov 1958). The matrices α_{ij}, β_{ij} are called Bogolubov coefficients, and by using (3.34) and (3.29) they can be evaluated as

$$\alpha_{ij} = (\bar{u}_i, u_j), \quad \beta_{ij} = -(\bar{u}_i, u_j^*). \tag{3.36}$$

Equating the expansions (3.30) and (3.32) and making use of (3.34), (3.35) and the orthonormality of the modes, (3.29), one obtains

$$a_i = \sum_j (\alpha_{ji} \bar{a}_j + \beta_{ji}^* \bar{a}_j^\dagger) \tag{3.37}$$

and

$$\bar{a}_j = \sum_i (\alpha_{ji}^* a_i - \beta_{ji}^* a_i^\dagger). \tag{3.38}$$

The Bogolubov coefficients possess the following properties

$$\sum_k (\alpha_{ik} \alpha_{jk}^* - \beta_{ik} \beta_{jk}^*) = \delta_{ij}, \tag{3.39}$$

$$\sum_k (\alpha_{ik} \beta_{jk} - \beta_{ik} \alpha_{jk}) = 0. \tag{3.40}$$

It follows immediately from (3.37) that the two Fock spaces based on the two choices of modes u_i and $\bar{u}_j$ are different so long as $\beta_{ji} \neq 0$. For example $|\bar{0}\rangle$ will not be annihilated by a_i:

$$a_i |\bar{0}\rangle = \sum_j \beta_{ji}^* |\bar{1}_j\rangle \neq 0 \tag{3.41}$$

in contrast to (3.33). In fact, the expectation value of the operator $N_i = a_i^\dagger a_i$ for the number of u_i-mode particles in the state $|\bar{0}\rangle$ is

$$\langle \bar{0} | N_i | \bar{0} \rangle = \sum_j |\beta_{ji}|^2, \tag{3.42}$$

which is to say that the vacuum of the $\bar{u}_j$ modes contains $\Sigma_j |\beta_{ji}|^2$ particles in the u_i mode.

Note that if u_j are positive frequency modes with respect to some timelike Killing vector field ξ, satisfying

$$\mathscr{L}_\xi u_j = -\mathrm{i}\omega u_j, \quad \omega > 0 \tag{3.43}$$

(cf. (2.8) which can be written $\mathscr{L}\partial_t u_k = -\mathrm{i}\omega u_k$), and $\bar{u}_k$ are a linear combination of u_j alone (not u_j^*), i.e., containing only positive frequencies with respect to ξ, then $\beta_{jk} = 0$. In that case $\bar{a}_k|0\rangle = 0$ as well as $a_j|0\rangle = 0$. Thus, the two sets of modes u_j and $\bar{u}_k$ share a common vacuum state. If any $\beta_{jk} \neq 0$, the $\bar{u}_k$ will contain a mixture of positive-(u_j) and negative-(u_j^*) frequency modes, and particles will be present.

More generally the Fock space based on $|0\rangle$ can be related to that based on $|\bar{0}\rangle$ using the completeness of the Fock space basis elements:

$$|{}^1n_{i_1}, {}^2n_{i_2}, \ldots\rangle$$

$$= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j_1 \ldots j_k} |\bar{1}_{j_1}, \bar{1}_{j_2}, \ldots, \bar{1}_{j_k}\rangle \langle \bar{1}_{j_1}, \bar{1}_{j_2}, \ldots, \bar{1}_{j_k}|{}^1n_{i_1}, {}^2n_{i_2}, \ldots\rangle. \tag{3.44}$$

In the notation used here we have, for example,

$$|{}^1n_{i_1}\rangle = |1_{i_1}, 1_{i_1}, \ldots, 1_{i_1}\rangle/({}^1n_i!)^{\frac{1}{2}}$$

where '1_{i_1}' is repeated ${}^1n_i$ times. The matrix element $\langle \bar{1}_{j_1}, \bar{1}_{j_2}, \ldots, \bar{1}_{j_k}|{}^1n_{i_1}, {}^2n_{i_2}, \ldots\rangle$ can be thought of as the transition amplitude or S-matrix element for a transition from the state $|{}^1n_{i_1}, {}^2n_{i_2}, \ldots\rangle$ to $|\bar{1}_{j_1}, \bar{1}_{j_2}, \ldots, \bar{1}_{j_k}\rangle$. These S-matrix elements can be written in terms of the Bogolubov coefficients. In particular, for the vacuum to many-particle amplitudes one finds (DeWitt 1975)

$$\langle \bar{0}|1_{j_1}, 1_{j_2}, \ldots, 1_{j_k}\rangle = \begin{cases} \mathrm{i}^{k/2}\langle \bar{0}|0\rangle \sum_{\rho} \Lambda_{\rho_1\rho_2} \cdots \Lambda_{\rho_{k-1}\rho_k} & k \text{ even} \\ 0 & k \text{ odd} \end{cases} \tag{3.45}$$

$$\langle \bar{1}_{j_1}, \bar{1}_{j_2}, \ldots, \bar{1}_{j_k}|0\rangle = \begin{cases} \mathrm{i}^{k/2}\langle \bar{0}|0\rangle \sum_{\rho} V_{\rho_1\rho_2} \cdots V_{\rho_{k-1}\rho_k} & k \text{ even} \\ 0 & k \text{ odd} \end{cases} \tag{3.46}$$

where ρ represents all distinct permutations of $\{j_1 \ldots j_k\}$ and

$$\left.\begin{aligned} \Lambda_{ij} &= -\mathrm{i}\sum_k \beta_{kj}\alpha_{ik}^{-1} \\ V_{ij} &= \mathrm{i}\sum_k \beta_{jk}^{*}\alpha_{ki}^{-1}. \end{aligned}\right\} \tag{3.47}$$

The many-particle to many-particle amplitudes may be found in Birrell & Taylor (1980). We defer consideration of the vacuum-to-vacuum amplitude until chapter 6.

In addition to the particle states and Bogolubov coefficients, we shall need various Green functions introduced in §2.7. The same notation and definitions will be used for the Green functions as in §2.7, except that now the field $\phi(x)$ satisfies the curved spacetime equation (3.26), and care must be taken in the specification of the vacuum state $|0\rangle$. Using (3.26), curved spacetime generalizations of the Green function equations (2.73)–(2.75) can be obtained. For example, for the Feynman propagator

$$iG_F(x, x') = \langle 0|T(\phi(x)\phi(x'))|0\rangle \tag{3.48}$$

one obtains

$$[\Box_x + m^2 + \xi R(x)]G_F(x, x') = -[-g(x)]^{-\frac{1}{2}}\delta^n(x - x'), \tag{3.49}$$

in place of (2.74). It is important to note that (3.49) by itself does not specify the state $|0\rangle$ appearing in (3.48), nor does it ensure that the solution has the properties of a time-ordered product. To fix the state and impose the time ordering, boundary conditions must be imposed on the solution of (3.49). In Minkowski space these boundary conditions take the form of the choice of contour used in (2.76). In curved spacetime the specification of boundary conditions will not be so simple, and will depend on the global features of the particular case under consideration. (For example, radiation which is retarded at its source will not remain so elsewhere because of 'backscattering' off the spacetime curvature.) Detailed discussions of wave propagation in curved spacetime, and the properties of Green functions, have been given for example, by DeWitt & Brehme (1960), and Friedlander (1975).

3.3 Meaning of the particle concept: particle detectors

The question naturally arises as to which set of modes furnishes the 'best' description of a physical vacuum, i.e., corresponds most closely to our actual experience of 'no particles'. It turns out that this question cannot be answered as stated, because it is necessary to specify also the details of the quantum measurement process that is used to detect the presence of quanta. In particular the state of motion of the measuring device can affect whether or not particles are observed to be present. For example, a free-falling detector will not always register the same particle density as a non-inertial, accelerating detector. In fact, this is even true in Minkowski space: an accelerated detector will register quanta even in the vacuum state defined by (2.19).

The special feature of Minkowski space is not that there is a unique vacuum (there is not), but that the conventional vacuum state as defined in terms of the modes (2.11) is the agreed vacuum for *all inertial* measuring devices, throughout the spacetime. This is because the vacuum defined by (2.19) is invariant under the Poincaré group and so are the set of inertial observers in Minkowski space.

One of the lessons learned from the development of this subject has been the realization that the particle concept does not generally have universal significance. Particles may register their presence on some detectors but not others, so there is an essential observer-dependent quality about them. One is still free to assert the presence of particles, but without specifying the state of motion of the detector, the concept is not very useful, even in Minkowski space.

Part of the reason for the nebulousness of the particle concept is its *global* nature. The modes are defined on the whole of spacetime (or at least a large patch) so that a particular observer's specification of the field mode decomposition, and hence the number operator describing the response of a particle detector carried by him, will depend, for example, on the observer's entire past history. To obtain a more objective probe of the state of a field one must construct locally-defined quantities, such as $\langle \psi | T_{\mu\nu}(x) | \psi \rangle$, which assumes a particular value at the point x of spacetime. The stress-tensor is objective in the sense that, for a fixed state $|\psi\rangle$, the results of different measuring devices can be related in the familiar fashion by the usual tensor transformation. For example, if $\langle \psi | T_{\mu\nu}(x) | \psi \rangle = 0$ for one observer, it will vanish for all observers. This is in contrast to the particle concept, where one observer may detect no particles while another may, as we shall see, disagree.

In many problems of interest the spacetime can be treated as asymptotically Minkowskian in the remote past and/or future. Under these circumstances the choice of the 'natural' Minkowskian vacuum defined by (2.19) has a well-understood physical meaning, i.e., the absence of particles according to *all inertial* observers in the asymptotic region – usually taken to be the commonly accepted idea of a vacuum. We refer to the remote past and future as the *in* and *out* regions respectively. This terminology is borrowed from Minkowskian quantum field theory where it is assumed that as $t \rightarrow \pm \infty$ all the field interactions approach zero. The analogous situation here would be to suppose that in the in and out regions spacetime admits natural particle states and a privileged quantum vacuum. This can either be Minkowski space, or some other spacetime of high symmetry such as the Einstein static universe. Whether a particular spacetime constitutes a suitable in or out region may also depend on the quantum field of interest. In the case of massless, conformally coupled fields, a conformally flat spacetime, even if not static, may still be a good candidate (see §3.7).

Since we work in the Heisenberg picture, if we choose the state of the quantum field in the in region to be the vacuum state, then it will remain in that state during its subsequent evolution. However, as will soon be demonstrated, at later times, outside the in region, freely falling particle detectors may still register particles in this 'vacuum' state. In particular, if there is also an out region, then the in vacuum may not coincide with the out vacuum. In that case a natural (e.g., inertial) class of observers in the out region will detect the presence of particles. We can therefore say that particles have been 'created' by the time-dependent external gravitational field. This is an especially useful description if the in and out regions are Minkowskian so that all inertial observers in the out region register the presence of quanta. Analogous processes of particle creation by external electromagnetic fields are well known (see for example, Gitman 1977, where references to earlier works are given). The possibility of similar particle production due to spacetime curvature was discussed over forty years ago by Schrödinger (1939), while other early work is due to DeWitt (1953), Takahashi & Umezawa (1957) and Imamura (1960). The first thorough treatment of particle production by an external gravitational field was given by Parker (1966, 1968, 1969), and Sexl & Urbantke (1967, 1969).

To illustrate these considerations we shall treat a model of a particle detector due to Unruh (1976) and DeWitt (1979). It consists of an idealized point particle with internal energy levels labelled by the energy E, coupled via a monopole interaction with a scalar field ϕ. We work in four-dimensional Minkowski space.

Suppose the particle detector moves along the world line described by the functions $x^\mu(\tau)$, where τ is the detector's proper time. The detector–field interaction is described by the interaction Lagrangian $cm(\tau)\phi[x(\tau)]$, where c is a small coupling constant and m is the detector's monopole moment operator. Suppose the field ϕ is in the vacuum state $|0_M\rangle$ defined by (2.19), where the subscript M stands for 'Minkowski vacuum'. For a general trajectory, the detector will not remain in its ground state E_0, but will undergo a transition to an excited state $E > E_0$, while the field will make a transition to an excited state $|\psi\rangle$. For sufficiently small c the amplitude for this transition may be given by first order perturbation theory (see §9.1) as

$$\mathrm{i}c\langle E, \psi| \int_{-\infty}^{\infty} m(\tau)\phi[x(\tau)]\,\mathrm{d}\tau\,|0_M, E_0\rangle$$

(The limits of integration may be confined to a smaller interval provided that the detector coupling is switched off adiabatically outside that interval.)

Using the equation for the time evolution of $m(\tau)$

$$m(\tau) = e^{iH_0\tau}m(0)e^{-iH_0\tau},$$

where $H_0|E\rangle = E|E\rangle$, the above transition amplitude factorizes to give

$$ic\langle E|m(0)|E_0\rangle \int_{-\infty}^{\infty} e^{i(E-E_0)\tau}\langle\psi|\phi(x)|0_M\rangle d\tau. \tag{3.50}$$

If ϕ is expanded in terms of standard Minkowski plane wave modes (2.17) it is clear that, to this order of perturbation theory, transitions can only occur to the state $|\psi\rangle = |1_{\mathbf{k}}\rangle$ containing one quantum of frequency $\omega = (|\mathbf{k}|^2 + m^2)^{\frac{1}{2}}$, for some $\mathbf{k}$. Then (in the continuum normalization (2.11))

$$\langle 1_{\mathbf{k}}|\phi(x)|0_M\rangle = \int d^3k'(16\pi^3\omega')^{-\frac{1}{2}}\langle 1_{\mathbf{k}}|a^{\dagger}_{\mathbf{k}'}|0_M\rangle e^{-i\mathbf{k}'\cdot\mathbf{x}+i\omega' t}$$

$$= (16\pi^3\omega)^{-\frac{1}{2}}e^{-i\mathbf{k}\cdot\mathbf{x}+i\omega t} \tag{3.51}$$

We must now take into account that $\mathbf{x}$ in (3.51) is not an independent variable but is determined by the detector's trajectory. Suppose it follows an inertial world line, i.e.,

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{v}t = \mathbf{x}_0 + \mathbf{v}\tau(1 - v^2)^{-\frac{1}{2}} \tag{3.52}$$

where $\mathbf{x}_0 =$ constant, $\mathbf{v} =$ constant, $|\mathbf{v}| < 1$, then the integral in (3.50) (with $\psi = 1_{\mathbf{k}}$) becomes

$$(16\pi^3\omega)^{-\frac{1}{2}}e^{-i\mathbf{k}\cdot\mathbf{x}_0}\int_{-\infty}^{\infty} e^{i(E-E_0)\tau}e^{i\tau(\omega-\mathbf{k}\cdot\mathbf{v})(1-v^2)^{-\frac{1}{2}}}d\tau$$

$$= (4\pi\omega)^{-\frac{1}{2}}e^{-i\mathbf{k}\cdot\mathbf{x}_0}\delta(E - E_0 + (\omega - \mathbf{k}\cdot\mathbf{v})(1 - v^2)^{-\frac{1}{2}}). \tag{3.53}$$

However as $\mathbf{k}\cdot\mathbf{v} \leq |\mathbf{k}||\mathbf{v}| < \omega$ and $E > E_0$, the argument of the δ-function is always > 0 and the transition amplitude vanishes. The transition is forbidden on energy conservation grounds – a direct consequence of Poincaré invariance.

If, on the other hand, instead of (3.52) we had chosen a more complicated trajectory, the integral in (3.50) would not have yielded a δ-function and the result would be nonzero. In such a case it is of interest to calculate the transition probability to *all* possible E and ψ, obtained by squaring the modulus of (3.50), and summing over E and the complete set ψ, to obtain

$$c^2 \sum_E |\langle E|m(0)|E_0\rangle|^2 \mathscr{F}(E - E_0), \tag{3.54}$$

where

$$\mathscr{F}(E) = \int_{-\infty}^{\infty} \mathrm{d}\tau \int_{-\infty}^{\infty} \mathrm{d}\tau' \mathrm{e}^{-\mathrm{i}E(\tau - \tau')} G^+(x(\tau), x(\tau')). \tag{3.55}$$

The detector *response function* $\mathscr{F}(E)$, is independent of the details of the detector, and is determined by the positive frequency Wightman Green function G^+ defined by (2.68). It represents the bath of 'particles' that the detector effectively experiences as a result of its motion. The remaining factor in (3.54) represents the *selectivity* of the detector to this bath, and clearly depends on the internal structure of the detector itself.

In the cases of detector trajectories in Minkowski space for which

$$G^+(x(\tau), x(\tau')) = g(\Delta\tau) \tag{3.56}$$

$$\Delta\tau \equiv \tau - \tau' \tag{3.57}$$

for some function g, the system is invariant under time translations in the reference frame of the detector ($\tau \to \tau +$ constant). This means that the detector is in equilibrium with the ϕ field, so that the number of quanta absorbed by the detector per unit τ is constant. If this rate is nonzero, the transition probability will diverge, as the transition amplitude (3.54) is computed for an infinite proper time interval. This can be seen immediately from (3.55), because the Wightman function will be a function of $\tau - \tau'$ only. Thus the double integration reduces to a Fourier transform of the two-point function multiplied by an infinite time integral.

This sort of circumstance frequently arises in quantum theory, and may be dealt with by adiabatically switching off the coupling as $\tau \to \pm\infty$, or considering instead the transition probability per unit proper time:

$$c^2 \sum_E |\langle E|m(0)|E_0\rangle|^2 \int_{-\infty}^{\infty} \mathrm{d}(\Delta\tau) \mathrm{e}^{-\mathrm{i}(E - E_0)\Delta\tau} G^+(\Delta\tau). \tag{3.58}$$

To simplify the following examples we now restrict attention to a massless scalar field ϕ. Then the positive frequency Wightman function is easily evaluated from (2.76), using the appropriate contour in fig. 3, to be

$$D^+(x, x') = -1/4\pi^2[(t - t' - \mathrm{i}\varepsilon)^2 - |\mathbf{x} - \mathbf{x}'|^2], \tag{3.59}$$

where the small imaginary part $i\varepsilon$, $\varepsilon > 0$, can be interpreted using the equation

$$1/(x \mp i\varepsilon) = (P/x) \pm i\pi\delta(x). \tag{3.60}$$

In the case of the inertial trajectory (3.52), equation (3.59) becomes

$$D^+(\Delta\tau) = -1/4\pi^2(\Delta\tau - i\varepsilon)^2, \tag{3.61}$$

(where we have absorbed a positive factor $(1 - v^2)^{-\frac{1}{2}}$ into ε) and the integral in (3.58) can be calculated as a contour integral, closing the contour in an infinite semicircle in the lower-half $\Delta\tau$ plane, since $E - E_0 > 0$. However the pole in the integrand is at $\Delta\tau = i\varepsilon$, which is in the upper-half plane, so the result is zero, as expected. No particles are detected.

As another example of this special equilibrium case, consider that the detector moves along a hyperbolic trajectory in the (t, z) plane:

$$x = y = 0, \quad z = (t^2 + \alpha^2)^{\frac{1}{2}} \quad \alpha = \text{constant}. \tag{3.62}$$

This represents a detector that accelerates uniformly with acceleration α^{-1} in the frame of the detector (see, for example, Rindler (1969)). The detector's proper time τ is related to t by

$$t = \alpha \sinh(\tau/\alpha) \tag{3.63}$$

so, from (3.59), we find

$$D^+(\Delta\tau) = -\left[16\pi^2\alpha^2 \sinh^2\left(\frac{\tau - \tau'}{2\alpha} - \frac{i\varepsilon}{\alpha}\right)\right]^{-1} \tag{3.64}$$

where we have absorbed a positive function of τ, τ' into ε. Using the identity

$$\operatorname{cosec}^2 \pi x = \pi^{-2} \sum_{k=-\infty}^{\infty} (x - k)^{-2} \tag{3.65}$$

we can write (3.64) as

$$D^+(\Delta\tau) = -(4\pi^2)^{-1} \sum_{k=-\infty}^{\infty} (\Delta\tau - 2i\varepsilon + 2\pi i\alpha k)^{-2}. \tag{3.66}$$

Substituting this into (3.58) and performing the Fourier transform with the help of a contour integral yields

$$\frac{c^2}{2\pi} \sum_E \frac{(E - E_0)|\langle E|m(0)|E_0\rangle|^2}{e^{2\pi(E - E_0)\alpha} - 1}. \tag{3.67}$$

The appearance of the Planck factor $[e^{2\pi(E-E_0)\alpha} - 1]^{-1}$ in (3.67) indicates that the equilibrium between the accelerated detector and the ϕ field in the state $|0_M\rangle$ is the same as that which would have been achieved had the detector remained unaccelerated, but immersed in a bath of thermal radiation at the temperature

$$T = 1/2\pi\alpha k_B = \text{acceleration}/2\pi k_B \tag{3.68}$$

where k_B is Boltzmann's constant.

The same conclusion can also be drawn from an examination of the *thermal* Green function for an *inertial* detector (Dowker 1977). This we have already derived for the case of $G_\beta^{(1)}$ in (2.111) (a similar derivation for G_β^+ is more complicated). Putting $\mathbf{x} = \mathbf{x}' = 0$, $t = \tau$, $t' = \tau'$ in (2.111) and using (2.79) gives, for a massless field,

$$D_\beta^{(1)}(\Delta\tau) = -(2\pi^2)^{-1} \sum_{k=-\infty}^{\infty} (\Delta\tau + ik\beta)^{-2}, \tag{3.69}$$

where the principal value is implied.

Let us compare this result with the $D^{(1)}$ function for the accelerated observer moving through a quantum field in the Minkowski *vacuum* state. It can be calculated from D^+ by making use of (2.67) in the form

$$D^{(1)}(\Delta\tau) = D^+(\Delta\tau) + D^+(-\Delta\tau).$$

Applying this identity to (3.66), noting (3.60), one observes that

$$D^{(1)}(\Delta\tau) = D_\beta^{(1)}(\Delta\tau)$$

where $\beta = 1/k_B T$ and T is given by (3.68). We conclude that the vacuum Green function for a uniformly accelerated detector is the same as the thermal Green function for an inertial detector.

What does this mean physically? It is often stated that a uniformly accelerated observer will 'see' thermal radiation (Davies 1975, Unruh 1976) even though the field ϕ is in the vacuum state $|0_M\rangle$ and, as far as inertial observers are concerned, no particles are detected whatever. Certainly the accelerated detector absorbs energy and makes transitions to excited states, just as if it were bathed in thermal radiation. However, as implied in §2.4, $\langle 0_M|:T_{\mu\nu}:|0_M\rangle = 0$. Transforming to the accelerated frame using the usual tensor transformation law gives $\langle 0_M|:T'_{\mu\nu}:|0_M\rangle = 0$ also, so both accelerated and unaccelerated observers agree that the stress–energy–momentum of the ϕ field vanishes. This has led to the descriptions 'quasi' or 'fictitious' particles for the quanta that excite the accelerated detector, but really the phenomenon is more an indication that the traditional quantum particle concept is applicable only in very restrictive circumstances.

If the state $|0_M\rangle$ cannot supply the energy to excite the detector, how can we reconcile its excitation with the principle of energy conservation? Moreover, the transition that elevates the detector from energy level E_0 to E is accompanied by the appearance of a quantum in the ϕ field ($|0_M\rangle \to |1_k\rangle$). This means that *both* the detector and the field gain energy.

The explanation comes from a consideration of the agency that brings about the acceleration of the detector in the first place. As the detector accelerates, its coupling to the ϕ field causes the *emission* of quanta, which produces a resistance against the accelerating force. The work done by the external force to overcome this resistance supplies the missing energy that feeds into the field via the quanta emitted from the detector, and also into the detector which simultaneously makes upward transitions. But as far as the detector is concerned, the net effect is the absorption of thermally distributed quanta.

Returning to the question of a suitable definition of a quantum vacuum state and Fock space could a case be made that such definitions based on the accelerated system are equally valid contenders as the traditional Minkowski-based constructs? In response it may be objected that basing one's treatment of these concepts on the considerations of accelerated observers is a fraud, because inertial observers occupy a special status in most physical theory. Hence, as far as Minkowski space is concerned, the vacuum $|0_M\rangle$ is a strong candidate for the 'correct' or 'physical' vacuum – the experiences of the accelerated observers being 'distorted' by the effects of their non-uniform motion. The trouble is that when gravitational fields are present, inertial observers become free-falling observers, and in general no two free-falling detectors will agree on a choice of vacuum. Only in exceptional cases of high symmetry will a set of detector trajectories exist that all register no particles, and this set may not even be free falling (e.g. the observers who accelerate to remain stationary close to a spherical star – see §8.4).

It is a simple matter to generalize the experiences of the detector to curved spacetime by replacing G^+ by its curved spacetime counterpart (see page 48) and the vacuum state $|0_M\rangle$ by some more general vacuum. Examples will be given in §§3.6 and 3.7.

It is also interesting to examine the case when the quantum field is not in a vacuum state but a many-particle state (2.22) (or its curved space counterpart). Then G^+ is replaced by

$$\langle {}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, {}^jn_{\mathbf{k}_j}|\phi(x)\phi(x')|{}^1n_{\mathbf{k}_1}, {}^2n_{\mathbf{k}_2}, \ldots, {}^jn_{\mathbf{k}_j}\rangle$$

$$= G^+(x,x') + \sum_i {}^inu_{\mathbf{k}_i}(x)u^*_{\mathbf{k}_i}(x') + \sum_i {}^inu^*_{\mathbf{k}_i}(x)u_{\mathbf{k}_i}(x'). \quad (3.70)$$

Passing to the continuum limit, the right-hand side of (3.70) is replaced by

$$G^+(x,x') + \int d^{n-1}k\, n_{\mathbf{k}} u_{\mathbf{k}}(x) u_{\mathbf{k}}^*(x') + \int d^{n-1}k\, n_{\mathbf{k}} u_{\mathbf{k}}^*(x) u_{\mathbf{k}}(x') \quad (3.71)$$

where $n_{\mathbf{k}}$ is the number density of quanta in k-space.

For an inertial detector moving along the trajectory (3.52) in n-dimensional Minkowski space, only the last term of (3.71) yields a contribution to the detector response function (3.55):

$$\frac{\mathscr{F}(E)}{T} = (2\pi)^{1-n} \int_{-\infty}^{\infty} d(\Delta\tau) e^{-iE\Delta\tau} \int \frac{d^{n-1}k}{2\omega} \exp\left[i(\omega - \mathbf{k}\cdot\mathbf{v})\Delta\tau(1-v^2)^{-\frac{1}{2}}\right] n_{\mathbf{k}} \quad (3.72)$$

where T is the total duration for which the detector is switched on. If $\mathbf{v} = 0$, the $\Delta\tau$ integration may be performed to yield $2\pi\delta(E - \omega)$. If in addition the quanta are distributed isotropically, then $n_{\mathbf{k}} = n_k$ and the $d^{n-1}k$ integral may be performed to yield

$$\frac{\mathscr{F}(E)}{T} = \frac{2^{2-n}\pi^{(3-n)/2}}{\Gamma((n-1)/2)} (E^2 - m^2)^{(n-3)/2} n_{(E^2-m^2)^{\frac{1}{2}}} \theta(E - m). \quad (3.73)$$

Substituting (3.73) into (3.54), one notices that the presence of the function $\theta(E - E_0 - m)$ shows that the absorption of a single quantum of mass m by the detector will not occur unless the energy level spacing $E - E_0$ in the detector is at least equal to the particle rest energy m. Further examination of (3.73) reveals that the transition response rate of the detector to the bath of quanta is proportional to the number of quanta in the mode of interest, as would be expected on physical grounds. The energy dependence however, is related in a complicated way to the particle number spectrum, and will be additionally complicated by the selectivity of the detector (energy dependence of the matrix element in (3.54)).

For $\mathbf{v} \neq 0$, the angular integration in (3.72) must be performed first. The τ integration then yields the difference between two θ-functions (rather than a δ-function) which delimits the range of the dk integration. In the massless four-dimensional case one obtains

$$\frac{\mathscr{F}(E)}{T} = \frac{1}{4\pi}\left(\frac{1-v^2}{v^2}\right)^{\frac{1}{2}} \int_{E^-}^{E^+} n_k\, dk, \quad (3.74)$$

where $E^{\pm} = E[(1 \pm v)/(1 \mp v)]^{\frac{1}{2}}$. This expression is readily understood physically. As the detector moves through the isotropic bath of radiation, a particular transition with energy $E - E_0$ will not select quanta from only one mode, but a whole range, varying from blueshifted lower energy modes in the forward direction to redshifted higher energy modes in the backward direction. The factors $[(1+v)/(1-v)]^{\frac{1}{2}}$ and $[(1-v)/(1+v)]^{\frac{1}{2}}$ are recognized as the usual Doppler blue- and redshift factors respectively, and one observes that the response function is proportional to the total number of quanta within the energy range made accessible by the Doppler shifting.

Given the vagueness of the particle concept in more general situations, one might wonder what meaning should be attached to the Bogolubov coefficients α and β introduced in the last section, and in particular what is meant by the expectation value for the number of quanta in a mode i, given by (3.42).

In general there is *no* simple relation between $\langle N_i \rangle$ and the particle number as measured by a detector, even if it is freely falling. However, in one special case a simple relation does exist. Consider a spacetime that is asymptotically static in the remote past and future, and construct the vacuum state in the in region associated with the standard plane wave modes of the form (2.11). This is the usual, physical, no-particle state, and an inertial detector in the in region would certainly register no quanta.

Of course, if the spacetime is only asymptotically static, the upper limits of the integrals in (3.55) are no longer strictly speaking infinity. We make the assumption that the monopole detector interaction is adiabatically switched off before the detector enters the non-static region. Similarly in the out region, the detector will be adiabatically switched on after the spacetime motion has ceased. This enables us to continue to use $\pm\infty$ in the limits as though the spacetime were always static. As already remarked, this type of assumption – of restricting an interaction to a finite duration using adiabatic switching – is routine in quantum theory, and so long as the duration of interaction is very much greater than $(E - E_0)^{-1}$, the spurious excitation of the detector by the switching process itself will be negligible.

In the out region the in modes will no longer in general be plane waves of fixed frequency, like (2.11). What, therefore, will a particle detector make of the *in* vacuum state in the *out* region? Clearly, an inertial detector in the out region will not register quanta if the quantum field is in the *out* vacuum state associated with modes that take the form (2.11) in the *out* region. However, it will, as we shall see, generally register the presence of some quanta if the field is in the *in* vacuum state.

To find out what the detector does register, one must analyse (in connection with (3.55)) the Wightman function G^+ (constructed using the in vacuum) evaluated in the out region. This will generally be a complicated function of the position variables $\mathbf{x}, \mathbf{x}'$. However, in the simple case of a homogeneous universe, such as the asymptotically static, spatially flat, Robertson–Walker model, G^+ will be invariant under spatial translations and rotations. We shall restrict discussion to this simple case.

If the in modes are denoted by $u_{\mathbf{k}}^{\mathrm{in}}$ and the in vacuum by $|0, \mathrm{in}\rangle$ then we must consider

$$G_{\mathrm{in}}^{+}(x,x') = \langle \mathrm{in}, 0|\phi(x)\phi(x')|0, \mathrm{in}\rangle = \int u_{\mathbf{k}}^{\mathrm{in}}(x)u_{\mathbf{k}}^{\mathrm{in}*}(x')\mathrm{d}^{n-1}k \qquad (3.75)$$

where x and x' are both situated in the out region. Denoting the out modes (i.e., the modes that reduce to the standard plane wave form (2.11) in the out region) by $u_{\mathbf{k}}^{\mathrm{out}}$, we may use (3.34) to expand $u_{\mathbf{k}}^{\mathrm{in}}$ in terms of $u_{\mathbf{k}}^{\mathrm{out}}$. Because of the spatial homogeneity, both sets of modes, $u_{\mathbf{k}}^{\mathrm{out}}$ and $u_{\mathbf{k}}^{\mathrm{in}}$, will remain as plane waves (i.e., $\propto e^{i\mathbf{k}\cdot\mathbf{x}}$) throughout, but with altered time dependence. The scalar products $(u_{\mathbf{k}}^{\mathrm{in}}, u_{\mathbf{k}'}^{\mathrm{out}})$ and $(u_{\mathbf{k}}^{\mathrm{in}}, u_{\mathbf{k}'}^{\mathrm{out}*})$ will therefore be proportional to $\delta_{\mathbf{k}\mathbf{k}'}$ and $\delta_{-\mathbf{k}\mathbf{k}'}$, respectively (see (3.28)). It follows from (3.36) that the Bogolubov transformation will be both diagonai and isotropic, i.e., will have the form

$$\alpha_{\mathbf{k}\mathbf{k}'} = \alpha_k \delta_{\mathbf{k}\mathbf{k}'} \qquad (3.76)$$

$$\beta_{\mathbf{k}\mathbf{k}'} = \beta_k \delta_{-\mathbf{k}\mathbf{k}'} \qquad (3.77)$$

where $k = |\mathbf{k}|$, so that

$$u_{\mathbf{k}}^{\mathrm{in}}(x) = \alpha_k u_{\mathbf{k}}^{\mathrm{out}}(x) + \beta_k u_{-\mathbf{k}}^{\mathrm{out}*}(x). \qquad (3.78)$$

Substituting (3.78) into (3.75) yields

$$G_{\mathrm{in}}^{+} = \int \mathrm{d}^{n-1}k[|\alpha_k|^2 u_{\mathbf{k}}^{\mathrm{out}}(x)u_{\mathbf{k}}^{\mathrm{out}*}(x') + \alpha_k\beta_k^* u_{\mathbf{k}}^{\mathrm{out}}(x)u_{-\mathbf{k}}^{\mathrm{out}}(x')$$
$$+ \beta_k\alpha_k^* u_{-\mathbf{k}}^{\mathrm{out}*}(x)u_{\mathbf{k}}^{\mathrm{out}*}(x') + |\beta_k|^2 u_{-\mathbf{k}}^{\mathrm{out}*}(x)u_{-\mathbf{k}}^{\mathrm{out}}(x')]. \qquad (3.79)$$

Inspection of (3.79) shows that, because of the $e^{i\mathbf{k}\cdot\mathbf{x}}$ dependence of $u_{\mathbf{k}}$, G^+ depends on position only through $\mathbf{x} - \mathbf{x}'$, i.e., it is invariant under spatial translations as expected. It follows immediately from this that when $\mathbf{x}$ and $\mathbf{x}'$ are restricted to the inertial detector trajectories (3.52), the position dependence disappears from G^+. The detector response (3.55) is therefore location-independent, as expected on grounds of homogeneity.

In the out region, the $u_{\mathbf{k}}^{\mathrm{out}}$ modes are simply standard plane waves, so the τ and τ' integrations in (3.55) are immediate, and yield δ-functions. The only non-vanishing term to survive from (3.79) is the final one.

The detector response function per unit time is

$$\frac{\mathscr{F}(E)}{T}=\frac{2^{2-n}\pi^{(3-n)/2}}{\Gamma((n-1)/2)}\int_0^\infty \frac{\mathrm{d}k\,k^{n-2}}{(k^2+m^2)^{\frac{1}{2}}}|\beta_k|^2\delta(E-(k^2+m^2)^{\frac{1}{2}})$$

$$=\frac{2^{2-n}\pi^{(3-n)/2}}{\Gamma((n-1)/2)}(E^2-m^2)^{(n-3)/2}|\beta_{(E^2-m^2)^{\frac{1}{2}}}|^2\theta(E-m) \qquad (3.80)$$

where for simplicity we have restricted the detector trajectory to the $\mathbf{v}=0$ case, and chosen the Robertson–Walker scale factor to be unity in the out region. This response function is identical to that associated with an isotropic bath of quanta with $|\beta_k|^2$ particles in mode k in a permanently static spacetime, as revealed by comparison of (3.80) with (3.73).

The simple form of this result implies the interpretation of $|\beta_k|^2$ as the number of quanta created in mode k by the cosmological motion, and is completely convincing. Of course, the use of a detector is not the only criterion by which the physical reasonableness of this interpretation can be based. Evaluation of $\langle T_{\mu\nu}\rangle$ would also reveal, in the out region, a bath of energy identical to that of $|\beta_k|^2$ particles in mode k in a permanently static spacetime.

3.4 Cosmological particle creation: a simple example

To see how, in practice, particle creation can occur in a spacetime with Minkowskian in and out regions, we shall consider a simple example. A suitable spacetime is a two-dimensional Robertson–Walker universe with line element

$$\mathrm{d}s^2=\mathrm{d}t^2-a^2(t)\mathrm{d}x^2 \qquad (3.81)$$

where the spatial sections expand (or contract) uniformly as described by the single scalar function $a(t)$. Introducing the new time parameter η (the so-called conformal time) defined by $\mathrm{d}\eta=\mathrm{d}t/a$, whence

$$t=\int^t \mathrm{d}t'=\int^\eta a(\eta')\mathrm{d}\eta', \qquad (3.82)$$

(3.81) may be recast as

$$\mathrm{d}s^2=a^2(\eta)(\mathrm{d}\eta^2-\mathrm{d}x^2)=C(\eta)(\mathrm{d}\eta^2-\mathrm{d}x^2), \qquad (3.83)$$

where we have defined the 'conformal scale factor' $C(\eta)=a^2(\eta)$. This form of the line element is manifestly conformal to Minkowski space (see §3.1).

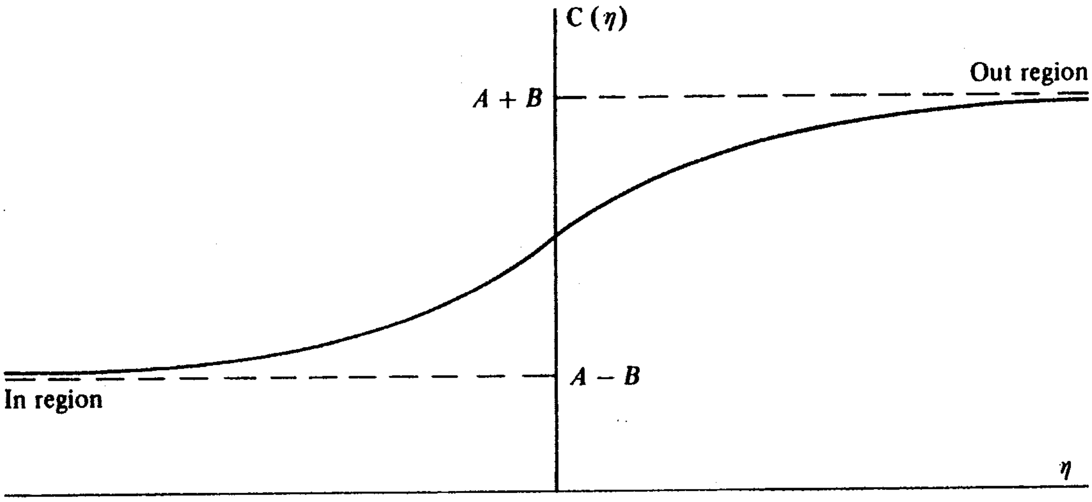


Fig. 7. The conformal scale factor $C(\eta) = A + B \tanh \rho\eta$ represents an asymptotically static universe that undergoes a period of smooth expansion.

Suppose that

$$C(\eta) = A + B \tanh \rho\eta, \quad A, B, \rho \text{ constants,} \tag{3.84}$$

then in the far past and future the spacetime becomes Minkowskian since

$$C(\eta) \to A \pm B, \quad \eta \to \pm\infty$$

(see fig. 7). We consider the production of massive, minimally coupled scalar particles in this spacetime; an investigation first carried out by Bernard & Duncan (1977). Note that in two dimensions minimal and conformal coupling are equivalent (see (3.27)).

Since $C(\eta)$ is not a function of x (the spatial coordinate) spatial translation invariance is still a symmetry in this spacetime, so we can separate the variables in the scalar mode functions appearing in (3.30):

$$u_k(\eta, x) = (2\pi)^{-\frac{1}{2}} \mathrm{e}^{\mathrm{i}kx} \chi_k(\eta). \tag{3.85}$$

Substituting (3.85) in place of ϕ into the scalar field equation (3.26), with $\xi = 0$ and the metric given by (3.83), one obtains an ordinary differential equation for $\chi_k(\eta)$:

$$\frac{\mathrm{d}^2}{\mathrm{d}\eta^2}\chi_k(\eta) + (k^2 + C(\eta)m^2)\chi_k(\eta) = 0. \tag{3.86}$$

This equation can be solved in terms of hypergeometric functions. The normalized modes which behave like the positive frequency Minkowski space modes (2.11) in the remote past ($\eta, t \to -\infty$) are

$$u_k^{\text{in}}(\eta, x) = (4\pi\omega_{\text{in}})^{-\frac{1}{2}} \exp\{ikx - i\omega_+\eta - (i\omega_-/\rho)\ln[2\cosh(\rho\eta)]\}$$
$$\times\, {}_2F_1(1 + (i\omega_-/\rho), i\omega_-/\rho; 1 - (i\omega_{\text{in}}/\rho); \tfrac{1}{2}(1 + \tanh\rho\eta))$$
$$\xrightarrow[\eta \to -\infty]{} (4\pi\omega_{\text{in}})^{-\frac{1}{2}} e^{ikx - i\omega_{\text{in}}\eta}, \tag{3.87}$$

where

$$\left.\begin{aligned} \omega_{\text{in}} &= [k^2 + m^2(A - B)]^{\frac{1}{2}} \\ \omega_{\text{out}} &= [k^2 + m^2(A + B)]^{\frac{1}{2}} \\ \omega_{\pm} &= \tfrac{1}{2}(\omega_{\text{out}} \pm \omega_{\text{in}}). \end{aligned}\right\} \tag{3.88}$$

On the other hand, the modes which behave like positive frequency Minkowski modes in the out region as $\eta \to +\infty$ are found to be

$$u_k^{\text{out}}(\eta, x) = (4\pi\omega_{\text{out}})^{-\frac{1}{2}} \exp\{ikx - i\omega_+\eta - (i\omega_-/\rho)\ln[2\cosh(\rho\eta)]\}$$
$$\times\, {}_2F_1(1 + (i\omega_-/\rho), i\omega_-/\rho; 1 + (i\omega_{\text{out}}/\rho); \tfrac{1}{2}(1 - \tanh\rho\eta))$$
$$\xrightarrow[\eta \to +\infty]{} (4\pi\omega_{\text{out}})^{-\frac{1}{2}} e^{ikx - i\omega_{\text{out}}\eta}. \tag{3.89}$$

Clearly u_k^{in} and u_k^{out} are not equal, which means that the β Bogolubov coefficient in (3.34) must be non-vanishing. To see this explicitly we can use the linear transformation properties of hypergeometric functions (see, for example, Abramowitz & Stegun 1965, Equations (15.3.6), (15.3.3)) to write u_k^{in} in terms of u_k^{out} as (cf. (3.78))

$$u_k^{\text{in}}(\eta, x) = \alpha_k u_k^{\text{out}}(\eta, x) + \beta_k u_{-k}^{\text{out}*}(\eta, x) \tag{3.90}$$

where

$$\alpha_k = \left(\frac{\omega_{\text{out}}}{\omega_{\text{in}}}\right)^{\frac{1}{2}} \frac{\Gamma(1 - (i\omega_{\text{in}}/\rho))\Gamma(-i\omega_{\text{out}}/\rho)}{\Gamma(-i\omega_+/\rho)\Gamma(1 - (i\omega_+/\rho))} \tag{3.91}$$

$$\beta_k = \left(\frac{\omega_{\text{out}}}{\omega_{\text{in}}}\right)^{\frac{1}{2}} \frac{\Gamma(1 - (i\omega_{\text{in}}/\rho))\Gamma(i\omega_{\text{out}}/\rho)}{\Gamma(i\omega_-/\rho)\Gamma(1 + (i\omega_-/\rho))}. \tag{3.92}$$

Comparison of (3.90) with (3.34) reveals that the Bogolubov coefficients are given by (cf. (3.76), (3.77))

$$\alpha_{kk'} = \alpha_k \delta_{kk'}, \quad \beta_{kk'} = \beta_k \delta_{-kk'}. \tag{3.93}$$

From (3.91) and (3.92) one obtains

$$|\alpha_k|^2 = \frac{\sinh^2(\pi\omega_+/\rho)}{\sinh(\pi\omega_{\text{in}}/\rho)\sinh(\pi\omega_{\text{out}}/\rho)} \tag{3.94}$$

$$|\beta_k|^2 = \frac{\sinh^2(\pi\omega_-/\rho)}{\sinh(\pi\omega_{\text{in}}/\rho)\sinh(\pi\omega_{\text{out}}/\rho)} \tag{3.95}$$

from which the normalization condition

$$|\alpha_k|^2 - |\beta_k|^2 = 1 \tag{3.96}$$

follows immediately (see (3.39)).

Consider the case that the quantum field resides in the state $|0,\text{in}\rangle$, defined by (3.33) in terms of the in region modes u_k^{in} (which we associate with the modes $\bar{u}_k$ of §3.2). In the remote past, where the spacetime is Minkowskian, all inertial particle detectors will register an absence of particles, so that unaccelerated observers there would identify the quantum state with a physical vacuum.

In the out region ($\eta \to +\infty$), the spacetime is also Minkowskian and the quantum field is also in the state $|0,\text{in}\rangle$ (as we are working in the Heisenberg picture), but in contrast to the situation in the in region, $|0,\text{in}\rangle$ is not regarded by inertial observers in the out region as the physical vacuum, this rôle being reserved for the state $|0,\text{out}\rangle$, defined in terms of modes u_k^{out}. Indeed, unaccelerated particle detectors there will register the presence of quanta (see (3.80)). In the mode k, the expected number of detected quanta is given by (3.95). We can therefore describe this quantum development as the creation of particles into the mode k as a consequence of the cosmic expansion.

3.5 Adiabatic vacuum

The results of the previous section have a readily visualizable physical description. In the massless limit, $\omega_- \to 0$, and the right-hand side of (3.95) vanishes: no particle production occurs. This is an example of a conformally trivial situation, i.e., a conformally invariant field propagating in a spacetime that is conformal to Minkowski space (see §3.1). Particle production takes place only when the conformal symmetry is broken by the presence of a mass, which provides a length scale for the theory. The production process can be regarded as caused by the coupling of the spacetime expansion to the quantum field via the mass. The changing 'gravitational field' feeds energy into the perturbed scalar field modes.

When attributing the production of field quanta to the changing gravitational field, it seems natural to describe the particles themselves as being produced *during* the period of expansion. Certainly no particles are produced in the asymptotically static regions. Moreover, if (3.95) represents a final density of particles which were not present in the in region, one would naively expect that a measurement taking place at an intermediate time, during the period of expansion, would reveal a particle density somewhere between zero and the value given by (3.95).

Unfortunately, these naive ideas do not stand up to scrutiny. As explained in §3.3, when spacetime is curved, no natural definition of particles is generally available. In spite of this, because of the special symmetry of the Robertson–Walker spacetime, one can identify a privileged class of observers, i.e., the comoving observers, who see the universe as expanding precisely isotropically. One might then wish to identify particles in the expansion region with the excitation of comoving particle detectors. This aspect will be studied in the next section.

Even if, for reasons of symmetry, a particular definition of particle is achieved as suggested above, the particle number will not be constant, a fact which makes its measurement inherently uncertain. If the average particle creation rate over an interval Δt is A, then to achieve a precise measurement of particle number, one must choose Δt such that $|A|\Delta t \ll 1$. However, there is also an uncertainty $(m\Delta t)^{-1}$ in the number of particles due to the Heisenberg energy–time uncertainty relation. Thus the total uncertainty in the particle number over a time interval Δt is (Parker 1969)

$$\Delta N \gtrsim (m\Delta t)^{-1} + |A|\Delta t,$$

which has a minimum value $2(|A|/m)^{\frac{1}{2}}$ when $\Delta t = (m|A|)^{-\frac{1}{2}}$. So long as $|A| \neq 0$ or $m \neq \infty$, this inherent uncertainty in N is non-vanishing.

In spite of this, we know from the success of standard Minkowski space quantum field theory that there must exist some sort of approximation to the curved space theory for which the particle number is 'almost meaningful'; we do, after all, inhabit an expanding universe. The above physical argument suggests that if the creation rate of particles is low, or the mass of the particles is high, then the notion of a well-defined particle number becomes a useful concept.

How can these ideas be made more precise?

The density and rate of particle production will obviously depend on the vigour of the expansion motion. In the limit of very weak expansion we would expect the creation rate to fall smoothly to zero, thereby recovering the Minkowski space theory. Inspection of (3.95) reveals that $|\beta_k|^2 \propto B^2 \rightarrow 0$

as the total *amount* of expansion approaches zero, and the Minkowski space limit is achieved. However, the particle creation falls much more sharply to zero if the expansion *rate* approaches zero. The rate is here parametrized by ρ, and for $\rho \to 0$ one obtains an exponential decline

$$|\beta_k|^2 \to e^{-2\pi\omega_{in}/\rho} \to 0. \tag{3.97}$$

The relevant 'slowness' parameter is ρ/ω_{in} which becomes small if $\rho \ll k$ or m. This condition is easy to understand physically. One expects the expansion motion to excite modes of the field for which $\omega \lesssim$ the expansion rate. For ω much greater than this, the particle production is exponentially suppressed. Thus, the high k modes are only excited very inefficiently. Similarly, the production of high-mass particles is exponentially small because of the large amount of energy which must emerge from the changing gravitational field to supply the particles' rest mass.

The fact that the creation of high k or m quanta is so strongly suppressed implies that, in the out region ($\eta \to \infty$) inertial particle detectors will register quanta in these energetic modes only extremely infrequently. The field which began vacuous in all modes at $\eta \to -\infty$, ends up 'almost' vacuous in the high energy modes. The slower the rate of expansion during the intermediate phase, the greater the probability that a given mode remains empty of quanta. As the probability declines exponentially in energy at high k and m, the approximation of neglecting any quanta that may have been created as a result of the expansion improves rapidly with energy.

If a high energy 'in' mode remains vacuous to high probability in the out region, it seems obvious that it must also be vacuous in the intermediate region, during the period of expansion. This idea can only be made meaningful, however, by specifying the motion of the particle detector. In the next section we shall show that a *comoving* detector will indeed almost certainly fail to register quanta in the high energy modes during the intermediate phase.

Although the above remarks are well illustrated by the example of the previous section, they apply completely generally to any Robertson–Walker spacetime with a smooth (C^∞) scale factor $C(\eta)$. In particular, the rapid decline in quanta as m or $k \to \infty$ is a completely general feature.

In the case of a Robertson–Walker model universe with static in and out regions, the situation is clear. If either the in or out vacuum states are chosen as the state of the quantum field, then a comoving particle detector will, over its entire world line, almost certainly fail to detect quanta in the high energy modes. So long as the mode frequency is much greater than the expansion rate, the probability of no detector response will remain very close to unity. However, for the lower modes, there will be quanta registered, signalling a breakdown of the meaningful approximation to a vacuum state. (Similar remarks apply if the field is chosen to be in a many-particle in or out state.) Moreover, because the in and out vacuum states are equally good in this respect, any linear combination of them will also lead to the above physical detector characteristics.

If there are no static in or out regions, an approximate definition of particles cannot be based on the above construction. Instead, a method must be found of selecting those exact mode solutions of the field equation that come in some sense 'closest' to the Minkowski space limit. Physically this might be envisaged as a construction that 'least disturbs' the field by the expansion, i.e., results in a definition of particles for which there is *minimal* particle production by the changing geometry. Such a construction has been given by Parker (1966, 1968, 1969, 1971, 1972) and has been subsequently developed by Parker & Fulling (1974), Fulling, Parker & Hu (1974), Fulling (1979), Bunch, Christensen & Fulling (1978), Birrell (1978), Hu (1978, 1979) and Bunch (1980*a*).

To arrive at a precise mathematical description of the above ideas will evidently entail some sort of high-mass expansion of the field modes. We give here a treatment valid for conformally coupled scalar fields in spatially flat Robertson–Walker spacetimes. Generalizations of this situation are given in the references cited above.

The line element for the spacetime is

$$ds^2 = C(\eta)[d\eta^2 - \sum_i (dx^i)^2] \tag{3.98}$$

where we take $C(\eta)$ to be a C^∞ function of the conformal time η. Because of the homogeneity of the spatial sections, the mode solutions of the wave equation are separable

$$u_{\mathbf{k}} = (2\pi)^{(1-n)/2} C^{(2-n)/4}(\eta) e^{i\mathbf{k}\cdot\mathbf{x}} \chi_k(\eta) \tag{3.99}$$

where $k = |\mathbf{k}|$. For a conformally coupled field χ_k satisfies the equation

$$\frac{d^2}{d\eta^2}\chi_k(\eta) + \omega_k^{\,2}(\eta)\chi_k(\eta) = 0 \tag{3.100}$$

where

$$\omega_k^{\,2}(\eta) = k^2 + C(\eta)m^2. \tag{3.101}$$

Equation (3.100) is reminiscent of the classical equation of motion for a harmonic oscillator with time-dependent frequency; for example, a simple pendulum whose length is slowly shortened, thereby decreasing the period. This problem was important in the formulation of quantum theory, because it appears that the energy E of one quantum of oscillation ($h\nu$) is insufficient to make up a whole quantum as the frequency ν rises. However, Einstein showed that so long as the pendulum length is decreased infinitely slowly, E/ν is an adiabatic invariant, and the number of quanta is conserved, irrespective of how great is the change in pendulum length (see, for example, Chandrasekhar 1958).

In the cosmological problem, we shall find that, similarly, the number of quanta (i.e., particle number) is an adiabatic invariant, irrespective of the total *quantity* of cosmological expansion, so long as the expansion *rate* is infinitely slow.

Equation (3.100) possesses the formal WKB-type solutions

$$\chi_k = (2W_k)^{-\frac{1}{2}} \exp\left[-\mathrm{i}\int^{\eta} W_k(\eta')\mathrm{d}\eta'\right] \tag{3.102}$$

where W_k satisfies the nonlinear equation

$$W_k^2(\eta) = \omega_k^2(\eta) - \frac{1}{2}\left(\frac{\ddot{W}_k}{W_k} - \frac{3}{2}\frac{\dot{W}_k^2}{W_k^2}\right). \tag{3.103}$$

There is, of course, an arbitrary phase factor implicit in (3.102) that can be specified by giving a lower limit to the integral.

If the spacetime is slowly varying, then the derivative terms in (3.103) will be small compared to ω_k^2, so a zeroth order approximation is to substitute

$$W_k^{(0)}(\eta) \equiv \omega_k(\eta) \tag{3.104}$$

into the integrand of (3.102). This solution obviously reduces to the standard Minkowski space modes as $C(\eta) \to$ constant.

Solutions to (3.103) may be approximated by iteration, using $W_k^{(0)}$ as the lowest order (Bunch 1980*a* – this method is equivalent to that of Liouville 1837, Chakroborty 1973). To clarify the 'slowness' property, it is helpful to introduce a parameter T known as the adiabatic parameter. If η is temporarily replaced by η/T (we take $T = 1$ at the end of the calculation) then the adiabatic limit of slow expansion may be examined by investigating what happens if $T \to \infty$.

With this device (3.100) may be rewritten

$$\frac{\mathrm{d}^2\chi(\eta_1)}{\mathrm{d}\eta_1^2} + T^2\omega_k^2(\eta_1)\chi_k(\eta_1) = 0 \tag{3.105}$$

where $\eta_1 = \eta/T$. Evidently

$$\frac{\mathrm{d}}{\mathrm{d}\eta}C(\eta/T) = \frac{1}{T}\frac{\mathrm{d}}{\mathrm{d}\eta_1}C(\eta_1) \tag{3.106}$$

so that in the limit $T \to \infty$, $C(\eta_1)$ and all its derivatives with respect to η vary infinitely slowly. Thus we can reproduce the effects of a slowly varying $C(\eta)$ by treating instead a large T approximation.

When an expansion in inverse powers of T is performed, the term of order T^{-n} will be called the nth adiabatic order. From (3.106) it is clear that the adiabatic order is in that case equivalent to the number of derivatives of C. It follows on dimensional grounds that if a quantity has dimensions m^d, a term of adiabatic order A in its expansion will contain $A - d$ powers of m^{-1} and k^{-1}.

The next iteration of (3.103) yields

$$(W_k^{(2)})^2 = \omega_k^{\,2} - \frac{1}{2}\left(\frac{\ddot{\omega}_k}{\omega_k} - \frac{3\dot{\omega}_k^{\,2}}{2\omega_k^{\,2}}\right) \tag{3.107}$$

which involves two derivatives of ω_k, hence C, so is of second adiabatic order in the slowness approximation. The Ath iterate yields a term of $2A$th adiabatic order. (When inserting $W_k^{(2)}$ in (3.102) to calculate the expansion to order two, the square root of (3.107) need be expanded only to terms of adiabatic order two.) In what follows we shall denote the Ath order adiabatic approximation to χ_k by $\chi_k^{(A)}$, and the associated modes (3.99) by $u_{\mathbf{k}}^{(A)}$.

Suppose that instead of using the exact solution χ_k given by (3.102) one used instead the zeroth order adiabatic approximation obtained by replacing W_k by $W_k^{(0)}$. In an in region, where the universe is static, both expressions yield the usual Minkowski solutions, with constant frequency. As the universe expands, however, the exact and approximate expressions will begin to differ, but only by terms of adiabatic order higher than zero, and this remains true however much the universe expands.

As an illustration, consider the case

$$C(\eta) = 1 + \mathrm{e}^{a\eta}, \quad a = \text{constant}, \quad \eta < 0, \tag{3.108}$$

which has an asymptotically static in region, followed by a period of expansion. The exact solution, which reduces to the usual positive frequency Minkowski space solution in the remote past is (Birrell 1979*a*)

$$\chi_k = \frac{\Gamma(1-(2\mathrm{i}\omega_k^-/a))}{(2\omega_k^-)^{\frac{1}{2}}}\left(\frac{m}{a}\right)^{2\mathrm{i}\omega_k^-/a} J_{-2\mathrm{i}\omega_k^-/a}(\mathrm{e}^{a\eta/2}) \tag{3.109}$$

where J is a Bessel function and $\omega_k^- = \omega_k(-\infty) = (k^2 + m^2)^{\frac{1}{2}}$. On the other hand, the zeroth order adiabatic solution (which also reduces to the standard Minkowski space positive frequency modes in the in region) is

$$\begin{aligned}\chi_k^{(0)} &= 2^{-\frac{1}{2}}(k^2 + m^2 + m^2 e^{a\eta})^{-1/4} \exp\left[-\mathrm{i}\int (k^2 + m^2 + m^2 e^{a\eta})^{\frac{1}{2}} \mathrm{d}\eta\right] \\ &= 2^{-\frac{1}{2}}(k^2 + m^2 + m^2 e^{a\eta})^{-1/4} \exp\left\{-\frac{2\mathrm{i}}{a}\left[(k^2 + m^2 + m^2 e^{a\eta})^{\frac{1}{2}}\right.\right. \\ &\quad \left.\left. - (k^2 + m^2)^{\frac{1}{2}} \tanh^{-1}\left(\frac{k^2 + m^2}{k^2 + m^2 + m^2 e^{a\eta}}\right)^{\frac{1}{2}}\right]\right\}. \end{aligned} \tag{3.110}$$

Clearly the exact and zeroth order adiabatic solutions agree in the remote past where the expansion is infinitely slow. Also, by using the standard asymptotic expansion of (3.109), for large m, it is easily verified that, to within an irrelevant arbitrary phase factor (associated with the indefinite lower limit of the integral in (3.102)), (3.109) reduces to (3.110) plus terms of adiabatic order greater than zero, regardless of the value of η.

In general there exists the following relation

$$u_{\mathbf{k}} = \alpha_k^{(A)}(\eta) u_{\mathbf{k}}^{(A)} + \beta_k^{(A)}(\eta) u_{\mathbf{k}}^{(A)*}, \tag{3.111}$$

defining an *exact* field mode (i.e., an exact solution of the field equation) in terms of the adiabatic approximation $u_{\mathbf{k}}^{(A)}$. Clearly $\alpha_k^{(A)}$ and $\beta_k^{(A)}$ must be constant to order A, because $u_{\mathbf{k}}^{(A)}$ and $u_{-\mathbf{k}}^{(A)}$ are solutions of the field equation to this order. Suppose that we make the particular choice

$$\begin{aligned} \alpha_k^{(A)}(\eta_0) &= 1 + O(T^{-(A+1)}) \\ \beta_k^{(A)}(\eta_0) &= 0 + O(T^{-(A+1)}) \end{aligned} \tag{3.112}$$

for some fixed time η_0. Then it follows that α and β are given by (3.112) for all time. The modes $u_{\mathbf{k}}$ defined by (3.111), (3.112) are said to be 'adiabatic positive frequency modes' to adiabatic order A. It is important to note that, this appellation notwithstanding, the modes $u_{\mathbf{k}}$ are not adiabatic approximations – they are exact. However, they are not defined uniquely by the prescription (3.111) and (3.112). There exist an infinite number of sets of such modes corresponding to different choices of η_0.

Special importance attaches to the exact modes $u_{\mathbf{k}}^{\text{in}}$ that reduce to the standard plane wave positive frequency exponential solutions in a static in region. Another set, $u_{\mathbf{k}}^{\text{out}}$, reduce to standard form in a static out region. In these static regions, all terms of adiabatic order greater than zero in

$u_{\mathbf{k}}^{(A)}$ vanish, so that such exact modes are of adiabatic positive frequency to infinite order. Thus, from the discussion associated with (3.112), the β Bogolubov coefficient connecting the two sets of modes $u_{\mathbf{k}}^{\mathrm{in}}$ and $u_{\mathbf{k}}^{\mathrm{out}}$, must fall off faster than any inverse power of T in the adiabatic limit $T \to \infty$. This was indeed found to be the case in the example of the previous section. It implies that the particle number associated with the quantization of either of these sets of modes is an adiabatic invariant during the cosmic expansion, in direct analogy to the problem of the pendulum with varying length.

If instead of using a set of exact field modes that reduce to standard form in, say, the in region, one used instead exact solutions that match on to Ath order *approximate* adiabatic modes at some later moment η_0, as in (3.111), (i.e., $u_{\mathbf{k}}$ and $\partial u_{\mathbf{k}}/\partial t$ are equated to adiabatic order A to $u_{\mathbf{k}}^{(A)}$ and $\partial u_{\mathbf{k}}^{(A)}/\partial t$ at $\eta = \eta_0$), then in the in region, these matched exact solutions will no longer reduce to standard form, and will in general be a linear combination of positive and negative frequency plane wave modes. A vacuum state constructed by quantizing these 'distorted' modes will not, therefore, be the same as the usual physical vacuum in the in region. That is, an inertial particle detector would register a bath of quanta when the field is in this 'distorted' vacuum. Nevertheless, the number spectrum of these quanta will generally fall off at large energy like $k^{-(A+1)}$ (or $m^{-(A+1)}$), reflecting the fact that the field modes have been matched on to approximate modes that differ from the standard in modes only by terms of adiabatic order $A + 1$ and higher. Thus, the 'distorted' vacuum may be regarded as an *adiabatic approximation* to the in vacuum. This adiabatic vacuum will with high probability leave the high energy modes vacuous, so that a comoving detector will, with high probability, register zero quanta in these modes. The probability that any given mode is empty of quanta for all time will not, however, be as high as in the case of the in or out vacuum states, with their associated more rapid decline at high energy (faster than any power of k or m).

Two features of the adiabatic vacuum concept should be properly understood. First, as remarked, the adiabatic vacuum is not some sort of approximate state based on approximate field modes. The adiabatic-approximate modes are themselves, of course, merely approximate solutions (to order A) of the field equation, but they are only used as a mathematical template against which to match exact mode solutions at some $\eta = \eta_0$. It is the *exact* modes themselves that become quantized. The associated vacuum state (i.e., the Ath order adiabatic vacuum) of these exact modes is a perfectly good contender for a vacuum state. True, it may not represent the experiences of a *comoving* detector as well as, say, the $(A + 1)$th adiabatic, or the in, vacuum states, but as far as the quantum field theory is concerned, it is respectable enough.

Secondly, there is no *unique* Ath order adiabatic vacuum, as the matching procedure may take place at any η_0. The associated exact modes will differ, for all time, only by terms of higher adiabatic order, so all are candidates for quantization and for the construction of an Ath order adiabatic vacuum. All such vacuum states will have comparable high energy behaviour, but will generally differ in the structure of the low energy modes.

Although an adiabatic vacuum is less specific than a vacuum state associated with a static in or out region, its representation of physical particles (in the sense of the experiences of a comoving particle detector) is nevertheless the best that is available if the spacetime *has no static* in and out regions. Current cosmological theory suggests that this is the case in the real universe.

As an illustration of the use of the adiabatic approximation in an asymptotically *non-static* four-dimensional cosmological model, consider the scale factor

$$C(\eta) = a^2 + b^2\eta^2, \quad -\infty < \eta < \infty \tag{3.113}$$

where a and b are constant (Audretsch & Schäfer 1978*a*). In the asymptotic regions, $\eta \to \pm\infty$, the model approaches the radiation-dominated Friedmann cosmology with

$$a(t) \equiv C^{\frac{1}{2}}(t) \propto t^{\frac{1}{2}}. \tag{3.114}$$

The time symmetry inherent in (3.113) indicates that the space contracts to a minimum scale factor at $\eta = 0$, 'bounces', and re-expands (see fig. 8).

Although not asymptotically static, the spacetime is nevertheless slowly varying as $\eta \to \pm\infty$, since

$$\frac{\mathrm{d}^l}{\mathrm{d}\eta^l}(\dot{C}/C) \to 0, \quad l \geq 0. \tag{3.115}$$

Thus, in the limit, the expansion rate vanishes and the adiabatic approximation becomes exact, so all adiabatic orders yield identical results as $\eta \to \pm\infty$.

More generally, the zeroth order adiabatic approximation becomes good whenever

$$\omega_k(\eta) = (k^2 + m^2a^2 + m^2b^2\eta^2)^{\frac{1}{2}}$$

becomes large compared with the derivatives on the left-hand side of (3.115). To see when this occurs, one studies (3.105) in which the factor

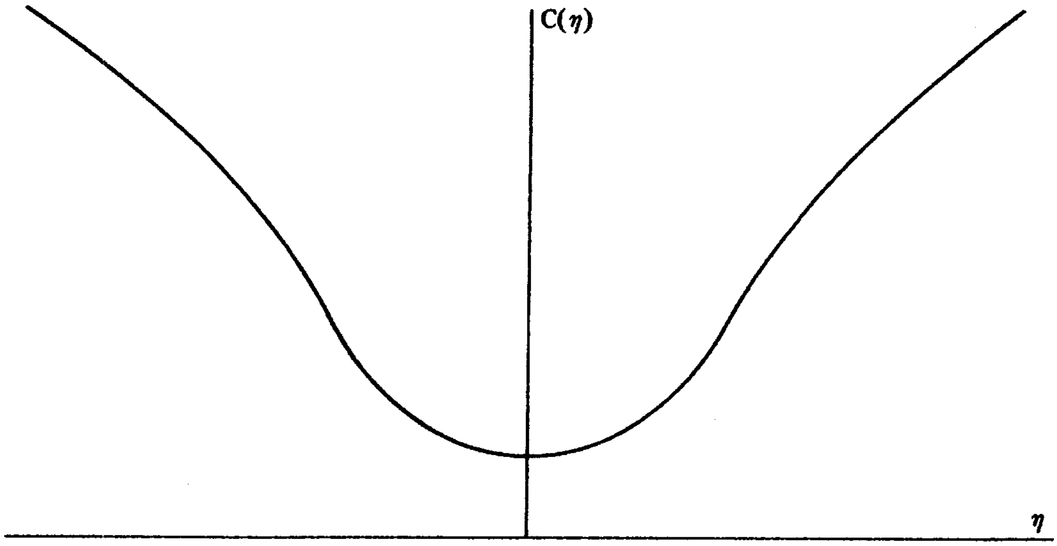


Fig. 8. The conformal scale factor $C(\eta) = a^2 + b^2\eta^2$ represents a universe that contracts to a minimum 'size' then re-expands symmetrically. At large η it behaves like a radiation-dominated Friedmann model.

$T^2\omega_k^2(\eta_1)$ can in this case be written as

$$T^2\omega_k^2(\eta_1) = mbT^2\lambda + m^2b^2\eta_1^2T^4$$

where

$$\lambda \equiv (ma^2/b) + (k^2/mb). \tag{3.116}$$

As the derivative $\dot{C}/C$ behaves like T^{-1}, one can now see that, for fixed $\eta_1 = \eta/T$, the zeroth order adiabatic approximation will be a good one for large λ, or large η, or large mb, or for any number of these quantities becoming large together.

Let us consider the limit of large λ with mb and η fixed. Substituting

$$W_k^{(0)} = \omega_k(\eta) = (mb\lambda)^{\frac{1}{2}} + O(T^{-2})$$

into (3.102) yields

$$\chi_k^{(0)}(\eta) \xrightarrow{\lambda\to\infty} (2mb\lambda)^{-\frac{1}{4}} \exp[-\mathrm{i}(mb\lambda)^{\frac{1}{2}}\eta], \quad \eta \text{ fixed}. \tag{3.117}$$

For comparison, one may now construct the exact solutions of (3.100) which reduce to (3.117) plus terms of $O(T^{-1})$(i.e., $O(\lambda^{-\frac{1}{2}})$) in the limit of large λ:

$$\chi_k^{\text{in}}(\eta) = (2mb)^{-1/4}\mathrm{e}^{-\pi\lambda/8}D_{-(1-\mathrm{i}\lambda)/2}[(\mathrm{i}-1)(mb)^{\frac{1}{2}}\eta], \quad \text{for } \eta < 0 \tag{3.118}$$

and

$$\chi_k^{\text{out}}(\eta) = \chi_k^{\text{in}}(-\eta)^*, \quad \text{for } \eta > 0. \tag{3.119}$$

(This can be checked by writing the parabolic cylinder function D in terms of a Whittaker function W and then using the asymptotic expansion of Whittaker functions for large index. See, for example, Bucholz 1969, page 99, equation (19*a*).)

One can also check that the large $|\eta|$ limit of (3.102), i.e.,

$$\chi_k^{(0)}(\eta) \xrightarrow{\eta \to \pm\infty} (2mb|\eta|)^{-\frac{1}{2}} \mathrm{e}^{\mp imb\eta^2/2} \tag{3.120}$$

is equal to the limit of χ_k^{in} as $\eta \to -\infty$, and the limit of χ_k^{out} as $\eta \to +\infty$.

Substituting (3.118) into (3.99) to obtain the full adiabatic positive frequency in modes $u_{\mathbf{k}}^{\text{in}}(x)$, the field ϕ may be expanded

$$\phi = \sum_{\mathbf{k}} (a_{\mathbf{k}}^{\text{in}} u_{\mathbf{k}}^{\text{in}} + a_{\mathbf{k}}^{\text{in}\dagger} u_{\mathbf{k}}^{\text{in}*}) \tag{3.121}$$

and the associated adiabatic vacuum state $|0_{\text{in}}^A\rangle$ defined by

$$a_{\mathbf{k}}^{\text{in}}|0_{\text{in}}^A\rangle = 0. \tag{3.122}$$

Similarly one may use (3.119) to define another adiabatic vacuum state of the same adiabatic order

$$a_{\mathbf{k}}^{\text{out}}|0_{\text{out}}^A\rangle = 0. \tag{3.123}$$

Because χ_k^{in} is matched to $\chi_k^{(0)}$ at $\eta = -\infty$, where all higher order adiabatic corrections vanish by virtue of (3.115), it is matched to $\chi_k^{(A)}$ for arbitrary A at $\eta = -\infty$. Thus the vacuum $|0_{\text{in}}^A\rangle$ defines an adiabatic in vacuum of infinite adiabatic order. Similarly $|0_{\text{out}}^A\rangle$ defines an adiabatic out vacuum of infinite adiabatic order. To see this explicitly, one may evaluate the Bogolubov transformation between $u_{\mathbf{k}}^{\text{in}}$ and $u_{\mathbf{k}}^{\text{out}}$:

$$u_k^{\text{in}} = \frac{\mathrm{i}(2\pi)^{\frac{1}{2}}\,\mathrm{e}^{-\pi\lambda/4}}{\Gamma(\frac{1}{2}(1-\mathrm{i}\lambda))} u_{\mathbf{k}}^{\text{out}} - \mathrm{i}\mathrm{e}^{-\pi\lambda/2} u_{\mathbf{k}}^{\text{out}*} \tag{3.124}$$

(see, for example, Gradshteyn & Ryzhik 1965, equation (9.248(3))). Thus, if the quantum state is chosen to be $|0_{\text{in}}^A\rangle$, a comoving particle detector in the out region will detect a spectrum

$$|\beta_k|^2 = \exp\{-\pi[(k^2/mb) + (ma^2/b)]\} \tag{3.125}$$

which does indeed fall off faster than any inverse power of k or m at high energies. It also falls faster than any positive power of b (the 'slowness' parameter) as $b \to 0$. Notice also that (3.125) remains finite as $a \to 0$, which corresponds to a model that intersects a singularity at $\eta = 0$.

As pointed out by Audretsch & Schäfer (1978*a*), the spectrum (3.125) is the same as that for a non-relativistic thermal gas of particles with momentum $kC^{-\frac{1}{2}}(\eta)$ at a chemical potential $-\frac{1}{2}ma^2C^{-1}(\eta)$ and temperature $b/(2\pi Ck_B)$, where k_B is, as before, Boltzmann's constant.

An alternative formulation of particle creation by a gravitational field, called Hamiltonian diagonalization, attempts to define particle states at each instant in analogy with the Minkowskian definition (see, for example, Imamura 1960, Grib & Mamaev 1969, 1971, Berger 1975, Castagnino, Verbeure & Weder 1974, 1975, Grib, Mamaev & Mostepanenko 1976, 1980*a*, *b*). Generally this procedure predicts vastly more creation than the methods employed here, and does not lead to the above-noted rapid decline in particle number at large k. Schäfer & Dehnen (1977), for example, predict the present particle creation rate to be 16 particles per km^3 per year – very many orders of magnitude in excess of what would be expected on physical grounds. Hamiltonian diagonalization has been strongly criticised, most notably by Fulling (1979) (see also Raine & Winlove 1975).

Analyses of the conceptual foundations of the particle concept in curved spacetime have been given by Ashtekar & Magnon (1975*a*, *b*), Hájíček (1976), Volovich, Zagrebnov & Frolov (1977), and Martellini, Sodano & Vitiello (1978).

3.6 Adiabatic expansion of Green functions

In the previous section it was shown that the high frequency behaviour of a massive scalar field is relatively insensitive to the long term time-dependence of the background Robertson–Walker spacetime. This is because the high frequency components of the field only probe the geometry in the immediate vicinity of the spacetime point of interest, and in this restricted neighbourhood the metric only changes by a small amount. In contrast the long wavelength, low frequency modes probe the entire manifold, and their structure is sensitive to the geometry and hence the particular adiabatic construction.

In some applications, e.g., the regularization of ultraviolet divergences, only the high frequency field behaviour is of interest. It is then only necessary to deal with high frequency approximations. Because the high frequencies probe only the short distances, one is led to examine short distance approximations.

Special interest attaches to the short distance behaviour of the Green functions, such as $G_F(x, x')$ in the limit $x \to x'$. We follow here a treatment of Bunch & Parker (1979) who obtained an adiabatic expansion of G_F.

Introducing Riemann normal coordinates y^μ for the point x, with origin at the point x'(e.g., Kreyszig 1968, Petrov 1969), one may expand

$$g_{\mu\nu}(x) = \eta_{\mu\nu} + \tfrac{1}{3}R_{\mu\alpha\nu\beta}y^\alpha y^\beta - \tfrac{1}{6}R_{\mu\alpha\nu\beta;\gamma}y^\alpha y^\beta y^\gamma$$
$$+ [\tfrac{1}{20}R_{\mu\alpha\nu\beta;\gamma\delta} + \tfrac{2}{45}R_{\alpha\mu\beta\lambda}R^\lambda{}_{\gamma\nu\delta}]y^\alpha y^\beta y^\gamma y^\delta + \ldots$$

where $\eta_{\mu\nu}$ is the Minkowski metric tensor, and the coefficients are all evaluated at $y=0$.

Defining

$$\mathscr{G}_F(x, x') = (-g(x))^{\frac{1}{4}} G_F(x, x') \tag{3.126}$$

and its Fourier transform by

$$\mathscr{G}_F(x, x') = (2\pi)^{-n}\int d^n k\, e^{-iky}\mathscr{G}_F(k) \tag{3.127}$$

where $ky = \eta^{\alpha\beta}k_\alpha y_\beta$, one can work in a sort of localized momentum space. Expanding (3.49) in normal coordinates and converting to k-space, $\mathscr{G}_F(k)$ can readily be solved by iteration to any adiabatic order. The result to adiabatic order four (i.e., four derivatives of the metric) is

$$\begin{aligned}\mathscr{G}_F(k) \approx{}& (k^2 - m^2)^{-1} - (\tfrac{1}{6} - \xi)R(k^2 - m^2)^{-2} \\ &+ \tfrac{1}{2}i(\tfrac{1}{6} - \xi)R_{;\alpha}\partial^\alpha(k^2 - m^2)^{-2} \\ &- \tfrac{1}{3}a_{\alpha\beta}\partial^\alpha\partial^\beta(k^2 - m^2)^{-2} + [(\tfrac{1}{6} - \xi)^2R^2 + \tfrac{2}{3}a^\lambda{}_\lambda](k^2 - m^2)^{-3}\end{aligned} \tag{3.128}$$

where $\partial_\alpha = \partial/\partial k^\alpha$,

$$\begin{aligned}a_{\alpha\beta} ={}& \tfrac{1}{2}(\xi - \tfrac{1}{6})R_{;\alpha\beta} + \tfrac{1}{120}R_{;\alpha\beta} - \tfrac{1}{40}R_{\alpha\beta;\lambda}{}^\lambda - \tfrac{1}{30}R_\alpha{}^\lambda R_{\lambda\beta} \\ &+ \tfrac{1}{60}R^\kappa{}_\alpha{}^\lambda{}_\beta R_{\kappa\lambda} + \tfrac{1}{60}R^{\lambda\mu\kappa}{}_\alpha\, R_{\lambda\mu\kappa\beta},\end{aligned} \tag{3.129}$$

and we are using the symbol $\approx$ to indicate that this is an asymptotic expansion. One ensures that (3.127) represents a time-ordered product (as in (3.48)) by performing the k^0 integral along the appropriate (Feynman) contour in fig. 3. This is equivalent to replacing m^2 by $m^2 - i\varepsilon$. Similarly, the adiabatic expansions of other Green functions can be obtained by using the other contours in fig. 3.

Substituting (3.128) into (3.127) yields

$$\begin{aligned}\mathscr{G}_F(x, x') \approx \int \frac{d^n k}{(2\pi)^n} e^{-iky}\bigg[& a_0(x, x') + a_1(x, x')\left(-\frac{\partial}{\partial m^2}\right) \\ &+ a_2(x, x')\left(\frac{\partial}{\partial m^2}\right)^2\bigg](k^2 - m^2)^{-1}\end{aligned} \tag{3.130}$$

where

$$a_0(x, x') \equiv 1 \tag{3.131}$$

and, to adiabatic order 4,

$$a_1(x, x') = (\tfrac{1}{6} - \xi)R - \tfrac{1}{2}(\tfrac{1}{6} - \xi)R_{;\alpha}y^\alpha - \tfrac{1}{3}a_{\alpha\beta}y^\alpha y^\beta \tag{3.132}$$

$$a_2(x, x') = \tfrac{1}{2}(\tfrac{1}{6} - \xi)^2R^2 + \tfrac{1}{3}a^\lambda{}_\lambda \tag{3.133}$$

with all geometric quantities on the right-hand side of (3.132) and (3.133) evaluated at x'.

If one uses the integral representation

$$(k^2 - m^2 + \mathrm{i}\varepsilon)^{-1} = -\mathrm{i}\int_0^\infty \mathrm{d}s\, \mathrm{e}^{\mathrm{i}s(k^2 - m^2 + \mathrm{i}\varepsilon)} \tag{3.134}$$

in (3.130), then the d^nk integration may be interchanged with the $\mathrm{d}s$ integration, and performed explicitly to yield (dropping the $\mathrm{i}\varepsilon$)

$$\mathscr{G}_\mathrm{F}(x, x') = -\mathrm{i}(4\pi)^{-n/2}\int_0^\infty \mathrm{i}\,\mathrm{d}s\,(\mathrm{i}s)^{-n/2}\exp[-\mathrm{i}m^2s + (\sigma/2\mathrm{i}s)]F(x, x'; \mathrm{i}s). \tag{3.135}$$

In (3.135) the function σ is defined by

$$\sigma(x, x') = \tfrac{1}{2}y_\alpha y^\alpha, \tag{3.136}$$

which is one-half of the square of the proper distance between x and x', while the function F has the following asymptotic adiabatic expansion

$$F(x, x'; \mathrm{i}s) \approx a_0(x, x') + a_1(x, x')\mathrm{i}s + a_2(x, x')(\mathrm{i}s)^2 + \ldots \tag{3.137}$$

Using (3.126), equation (3.135) gives a representation of $G_\mathrm{F}(x, x')$ originally derived by DeWitt (1965, 1975), following the work of Schwinger (1951*a*) (and hence labelled DS, see also Fock 1937, Nambu 1950)

$$G_\mathrm{F}^\mathrm{DS}(x, x') = -\mathrm{i}\Delta^{\frac{1}{2}}(x, x')(4\pi)^{-n/2}\int_0^\infty \mathrm{i}\mathrm{d}s\,(\mathrm{i}s)^{-n/2}$$
$$\exp[-\mathrm{i}m^2s + (\sigma/2\mathrm{i}s)]F(x, x'; \mathrm{i}s) \tag{3.138}$$

where Δ is the Van Vleck determinant (Van Vleck 1928)

$$\Delta(x, x') = -\det[\partial_\mu\partial_\nu\sigma(x, x')][g(x)g(x')]^{-\frac{1}{2}} \tag{3.139}$$

In the normal coordinates about x' that we are currently using, Δ reduces to $[-g(x)]^{-\frac{1}{2}}$. In the treatment of DeWitt, the extension of the asymptotic expansion (3.137) of F to all adiabatic orders is written as

$$F(x, x'; \mathrm{i}s) \approx \sum_{j=0}^{\infty} a_j(x, x')(\mathrm{i}s)^j \tag{3.140}$$

with $a_0(x, x') = 1$, the other a_j being given by recursion relations which enable their adiabatic expansions to be obtained (Christensen 1976). It is important to note that the so-called DeWitt–Schwinger 'proper time' representation, (3.138), is intended to be an exact representation of the Feynman propagator. The expansions (3.137) and (3.140) are, however, only asymptotic approximations in the limit of large adiabatic parameter T.

If (3.140) is substituted into (3.138) the integral can be performed to give the adiabatic expansion of the Feynman propagator in coordinate space:

$$G_{\mathrm{F}}^{\mathrm{DS}}(x, x') \approx \frac{-\mathrm{i}\pi\Delta^{\frac{1}{2}}(x, x')}{(4\pi\mathrm{i})^{n/2}} \sum_{j=0}^{\infty} a_j(x, x')\left(-\frac{\partial}{\partial m^2}\right)^j$$
$$\times \left[\left(\frac{2m^2}{-\sigma}\right)^{(n-2)/4} H^{(2)}_{(n-2)/2}((2m^2\sigma)^{\frac{1}{2}})\right] \tag{3.141}$$

in which, strictly, a small imaginary part iε should be subtracted from σ.

Since we have not imposed global boundary conditions on the Green function solution of (3.49), the expansion (3.141) does not determine the particular vacuum state in (3.48). In particular, the 'iε' in the expansion of G_{F} only ensures that (3.141) represents the expectation value, in some set of states, of a time-ordered product of fields. Under some circumstances the use of 'iε' in the exact representation (3.138) may give additional information concerning the global nature of the states – see, for example, the discussion on page 48. Expressed differently, the same high frequency behaviour of (3.48) results from almost all choices of vacuum state, a fact which will turn out to be of considerable importance.

In a Robertson–Walker spacetime, the Feynman Green function calculated as an expectation value in an adiabatic vacuum $|0^A\rangle$

$$\mathrm{i}G_{\mathrm{F}}^{A}(x, x') = \langle 0^A | T(\phi(x)\phi(x')) | 0^A \rangle, \tag{3.142}$$

should have an expansion of the form (3.138) and (3.140). The vacuum $|0^A\rangle$ is defined in terms of a set of adiabatic positive frequency modes $u_{\mathbf{k}}$ given by (3.111), (3.112) for some η_0, and so one can write (3.142) as

$$\mathrm{i}G_{\mathrm{F}}^{\mathrm{A}}(x, x') = \theta(x^0 - x^{0\prime}) \int \mathrm{d}^{n-1}k\; u_{\mathbf{k}}(x) u_{\mathbf{k}}^*(x')$$
$$+ \theta(x^{0\prime} - x^0) \int \mathrm{d}^{n-1}k\; u_{\mathbf{k}}^*(x) u_{\mathbf{k}}(x'). \tag{3.143}$$

To obtain an expansion of (3.143) to adiabatic order A, it is only necessary to use the expansion of $u_{\mathbf{k}}$ to this order; i.e., to calculate (3.143) with $u_{\mathbf{k}}$ replaced by $u_{\mathbf{k}}^{A}$. Doing this, one can explicitly verify that the expansions of (3.142) and (3.143) agree to order A (Birrell 1978, Bunch, Christensen & Fulling 1978, Bunch & Parker 1979). In particular, one can check that the '$i\varepsilon$' is necessary to guarantee the time ordering in (3.142) and (3.143).

We are now in a position to discuss the response of a particle detector to an Ath order adiabatic vacuum. To do this, one must substitute

$$G_{A}^{+}(x(\tau), x(\tau')) \equiv \langle 0^{A}|\phi(x(\tau))\phi(x(\tau'))|0^{A}\rangle \tag{3.144}$$

into (3.55), to obtain the detector response function along the world line $x(\tau)$. If we only wish to consider the contribution to the probability of terms in this exact two-point function up to adiabatic order A, then, for any spacetime, we can calculate these terms using the adiabatic expansion $u_{\mathbf{k}}^{(A)}$ for the modes. As in the case of G_{F}, the expansion of G_{A}^{+} will be equal to the expansion to order A of the DeWitt–Schwinger representation G_{DS}^{+}. The expansion of G_{DS}^{+} can be calculated in momentum space as

$$\mathrm{i}G_{\mathrm{DS}}^{+}(x, x') = [-g(x)]^{-1/4}\mathscr{G}^{+}(x, x')$$

where, to adiabatic order four, $\mathscr{G}^{+}$ is given by the right-hand side of (3.130) with the k^0 integral performed along the contour shown in fig. 9. Using the expansion for $[-g(x)]^{-1/4}$ in normal coordinates about x' one observes that G^{+} (to arbitrary adiabatic order) will be a sum of terms with the general

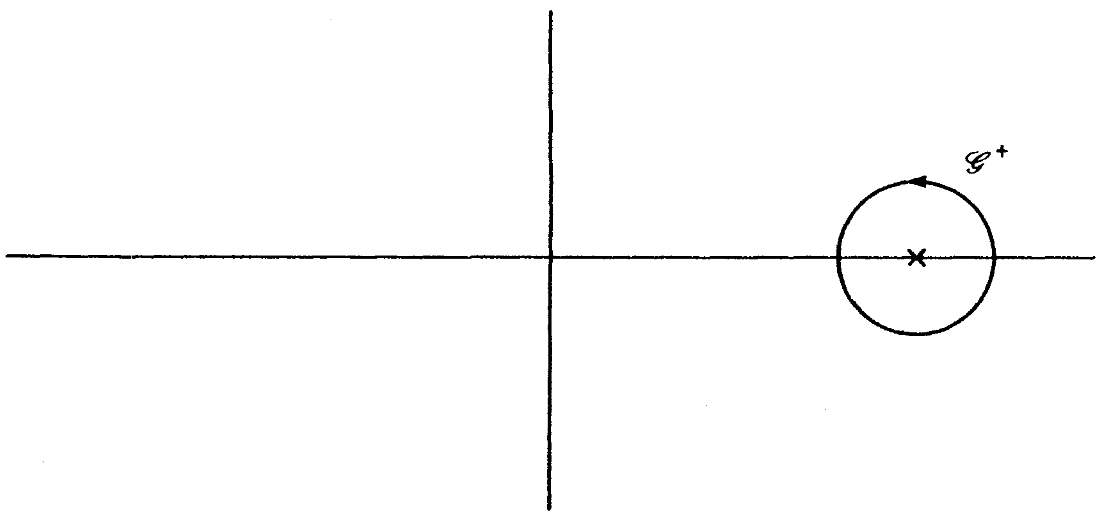


Fig. 9. The contour in the complex k^0 plane to be used in the evaluation of the integral giving $\mathscr{G}^{+}$. The cross indicates the pole at $k^0 = (|\mathbf{k}|^2 + m^2)^{\frac{1}{2}}$.

structure

$$\int \frac{\mathrm{d}^n k}{(2\pi)^n} \frac{\mathrm{e}^{-iky}}{(k^2 - m^2)^p} S_{\mu\nu\ldots\lambda}(x') y^\mu y^\nu \ldots y^\lambda \tag{3.145}$$

where S is a geometrical tensor and, as before, y^μ are the normal coordinates about x' of the point x. The k^0 integral can be performed around the contour shown in fig. 9 to give a linear combination of terms having the form

$$\int \frac{\mathrm{d}^{n-1} k}{(2\pi)^{n-1}} \frac{\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{y} - \mathrm{i}\omega y^0}}{(2\omega)^r} S_{\mu\nu\ldots\lambda}(x') (y^0)^q y^\mu y^\nu \ldots y^\lambda \tag{3.146}$$

where r and q are integers.

We now restrict attention to detectors that are comoving, i.e., with world lines $x(\tau)$ such that $x^i = \text{constant}$ and $\tau = t$ (the Robertson–Walker cosmic time). Using the properties of normal coordinates, $G^+_{\mathrm{DS}}(x(\tau), x(\tau'))$ will be given by a linear combination of terms like (3.146) with $\mathbf{y} = 0, y^0 = t$. Substitution of such a term into (3.55) yields integrals of the form

$$\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}(\omega + E - E_0)\tau} (\tau)^l \mathrm{d}\tau, \quad (l \text{ an integer}).$$

Because $\omega + E - E_0 > 0$, all such integrals vanish as they are derivatives of $\delta(\omega + E - E_0)$.

In conclusion, we may say that an adiabatic vacuum of order A is a state for which a comoving particle detector remains unexcited with a probability that differs from unity only by terms of order $A + 1$. That is, the probability for detection of a particle of energy ω will fall to zero faster than the Ath inverse power of ω in the ultraviolet limit, or as the Ath power of some suitable 'slowness' parameter. This proves Parker's (1969, §E) conjecture that adiabatic particles (although not named as such – see Parker & Fulling 1974) satisfy the criterion that they should 'be measured in a slowly expanding universe by essentially the same apparatus as in the static case'. This has also been demonstrated in an entirely different way by Parker (1966).

The higher the order of A chosen for the construction, the more probably are the high frequency modes found to be vacuous. Of course, there will still be particles detected in the low frequency modes; the above high frequency expansion will not reveal the details of this. Also, if the detector is not comoving, there will generally be particle detection in the high frequency modes that falls to zero slower than T^{-A}.

3.7. Conformal vacuum

Despite the appealing properties of adiabatic vacuum states, the fact remains that in a curved spacetime a particular set of mode solutions of the field equation and the corresponding vacuum and many-particle states do not *in general* have direct physical significance. In particular, a 'vacuum' state may not necessarily be measured as devoid of quanta, even by a freely-falling detector. Nevertheless, if there exist geometrical symmetries in the spacetime of interest, it may be that a particular set of modes and particle states emerge as in some sense 'natural'.

One case of special interest to which we shall frequently return is that of *conformal* triviality, i.e., a conformally invariant field propagating in a conformally flat spacetime (a spacetime conformal to Minkowski space). The symmetry of such spacetimes is manifested by the existence of a conformal Killing vector satisfying (3.23). Examples of conformally flat spacetimes are all two-dimensional spacetimes and the spatially flat Robertson–Walker cosmological models. Their metric tensors may always be cast in the form

$$g_{\mu\nu}(x) = \Omega^2(x)\eta_{\mu\nu} \tag{3.147}$$

where $\eta_{\mu\nu}$ is the Minkowski metric tensor.

The conformally invariant scalar wave equation requires $m = 0$ and the choice of ξ given by (3.27):

$$[\Box + \tfrac{1}{4}(n-2)R/(n-1)]\phi = 0. \tag{3.148}$$

Under the conformal transformation

$$g_{\mu\nu} \to \Omega^{-2} g_{\mu\nu} = \eta_{\mu\nu} \tag{3.149}$$

one obtains from (3.1), (3.5) and (3.148)

$$\bar{\Box}\bar{\phi} \equiv \eta^{\mu\nu}\partial_\mu\partial_\nu(\Omega^{(n-2)/2}\phi) = 0, \tag{3.150}$$

as $\bar{R} = 0$ in Minkowski space. The power $(n-2)/2$ of Ω is known as the conformal weight of the scalar field. Equation (3.150) possesses the familiar Minkowski space mode solutions for $\bar{\phi}$ (cf. (2.11))

$$\bar{u}_{\mathbf{k}}(x) = [2\omega(2\pi)^{n-1}]^{-\frac{1}{2}} e^{-ik\cdot x}, \quad k^0 = \omega. \tag{3.151}$$

These modes are positive frequency with respect to the timelike conformal Killing vector ∂_η. That is

$$\mathscr{L}_{\partial_\eta}\bar{u}_{\mathbf{k}}(x) = -i\omega\bar{u}_{\mathbf{k}}(x), \quad \omega > 0. \tag{3.152}$$

Noting from (3.150) that $\phi = \Omega^{(2-n)/2}\,\bar{\phi}$, the mode decomposition (3.30) for ϕ can in this case be written

$$\phi(x) = \Omega^{(2-n)/2}(x)\sum_{\mathbf{k}}\left[a_{\mathbf{k}}\bar{u}_{\mathbf{k}}(x) + a_{\mathbf{k}}^{\dagger}\bar{u}_{\mathbf{k}}^{*}(x)\right] \tag{3.153}$$

with $\bar{u}_{\mathbf{k}}$ given by (3.151). The vacuum state associated with these modes, defined by $a_{\mathbf{k}}|0\rangle = 0$, is called the *conformal vacuum*. Similar states exist for the massless spin $\frac{1}{2}$ field, and the electromagnetic fields (see §3.8).

The Green function equation (3.49) may be treated similarly. Using (3.5),

$$[\Box_x + \tfrac{1}{4}(n-2)R(x)/(n-1)]D_F(x,x') = -[-g(x)]^{-\frac{1}{2}}\delta^n(x-x')$$

becomes

$$\Omega^{-(n+2)/2}(x)\eta^{\mu\nu}\partial_\mu\partial_\nu(\Omega^{(n-2)/2}(x)D_F(x,x')) = -\Omega^{-n}(x)\delta^n(x-x').$$

Thus

$$\begin{aligned}\eta^{\mu\nu}\partial_\mu\partial_\nu(\Omega^{(n-2)/2}(x)D_F(x,x')) &= -\Omega^{(2-n)/2}(x)\delta^n(x-x')\\ &= -\Omega^{(2-n)/2}(x')\delta^n(x-x')\end{aligned}$$

which implies

$$D_F(x,x') = \Omega^{(2-n)/2}(x)\tilde{D}_F(x,x')\Omega^{(2-n)/2}(x') \tag{3.154}$$

where $\tilde{D}_F(x,x')$ is the massless version of the Minkowski space Feynman Green function satisfying (2.74). This relation can also be derived using the definition (3.48) of the Green function in terms of the field. Similar relations hold for the other scalar Green functions. For massless spin $\frac{1}{2}$ and spin 1 fields, analogous relations hold, with the power of Ω determined by the conformal weight appropriate to the field (see §3.8).

We shall now use these results to analyse the behaviour of a comoving particle detector in a Robertson–Walker universe described by the line element (3.98), to see how it responds to the conformal vacuum state. We work in four dimensions.

In Robertson–Walker spacetime the comoving geodesics $\mathbf{x} = \text{constant}$ map under the conformal transformation (3.149) into the geodesics $\mathbf{x} = \text{constant}$ in Minkowski space. However the proper time τ along these geodesics, which coincides with cosmic time t, does not coincide with the conformal time parameter $x^0 = \eta$ of (3.98) (see also(3.82)). Along the comoving geodesics the Green function $D^+(x,x')$ reduces to

$$D^+(\eta,\eta') = -\frac{C^{-\frac{1}{2}}(\eta)C^{-\frac{1}{2}}(\eta')}{4\pi^2(\eta-\eta'-\mathrm{i}\varepsilon)^2} \tag{3.155}$$

where we have used (3.59) and (3.154) with G_F replaced by D^+ and $\Omega^2 = C$. Substituting this into the response function (3.55) and using (3.82) to change from integrals with respect to τ to integrals with respect to η gives

$$\mathscr{F}(E) = -\frac{1}{4\pi^2}\int d\eta \int d\eta' \frac{\exp\left[-iE\int_{\eta'}^{\eta} C^{\frac{1}{2}}(\eta'')d\eta''\right]}{(\eta - \eta' - i\varepsilon)^2} \qquad (3.156)$$

which will *not* in general vanish, owing to the presence of the $C^{\frac{1}{2}}$ factor in the exponent. Explicit examples of this are given in §§5.3, 5.4.

Thus, even a comoving particle detector – the nearest analogue in Robertson–Walker space to an inertial observer in Minkowski space – will in general register the presence of particles in the conformal vacuum. On the other hand, inspection of (3.152) reveals that the modes which are positive frequency with respect to the conformal vacuum at one time remain so for all time. Thus, if a field satisfying a conformally invariant wave equation in a conformally flat spacetime is in the conformal vacuum at one time it will remain so for all time, and there will be no particle production (Parker 1966, 1968, 1969, 1971, 1973; see also page 189). This will be the case for neutrinos (if they are massless) and photons, though not for gravitons (Grishchuk 1974, 1975). The absence of particle production is to be interpreted in the following sense. If the cosmological expansion is allowed to cease (smoothly), then an inertial particle detector adiabatically switched on after the expansion has ceased will register no particles. The positive response of the detector that is switched on during the expansion might therefore be described as spurious, but it is really only another illustration that in regions of nonzero spacetime curvature the particle concept loses much of its intuitive meaning.

The conformal vacua of a slightly more general class of spacetimes will be discussed in §5.2 (see also Candelas & Dowker 1979).

3.8 Fields of arbitrary spin in curved spacetime

In Minkowski space field theory, the spin of a field can be classified according to the field's properties under infinitesimal Lorentz transformations

$$x^\alpha \to \bar{x}^\alpha = \Lambda^\alpha{}_\beta x^\beta = (\delta^\alpha{}_\beta + \omega^\alpha{}_\beta)x^\beta \qquad (3.157)$$

$$\omega_{\alpha\beta} = -\omega_{\beta\alpha} \qquad (3.158)$$

$$|\omega^\alpha{}_\beta| \ll 1.$$

(For the time being we restrict attention to four-dimensional Minkowski space.)

A general multicomponent field $T^{\gamma\delta\cdots\lambda}(x)$ transforms to

$$[D(\Lambda)]^{\gamma'\delta'\ldots\lambda'}{}_{\gamma\delta\ldots\lambda}T^{\gamma\delta\ldots\lambda} \tag{3.159}$$

where

$$D(\Lambda) = 1 + \tfrac{1}{2}\omega^{\alpha\beta}\Sigma_{\alpha\beta}. \tag{3.160}$$

In order that Lorentz transformations form a group, the antisymmetric $\Sigma_{\alpha\beta}$ (called the group generators) are constrained to satisfy the commutation identities

$$[\Sigma_{\alpha\beta}, \Sigma_{\gamma\delta}] = \eta_{\gamma\beta}\Sigma_{\alpha\delta} - \eta_{\alpha\gamma}\Sigma_{\beta\delta} + \eta_{\delta\beta}\Sigma_{\gamma\alpha} - \eta_{\delta\alpha}\Sigma_{\gamma\beta}, \tag{3.161}$$

or, equivalently,

$$\left.\begin{aligned} \mathbf{a} \times \mathbf{a} &= \mathrm{i}\mathbf{a} \\ \mathbf{b} \times \mathbf{b} &= \mathrm{i}\mathbf{b} \\ [a_i, b_i] &= 0 \end{aligned}\right\} \tag{3.162}$$

where

$$\left.\begin{aligned} a_i &= \tfrac{1}{2}(-\mathrm{i}\varepsilon_i{}^{kl}\Sigma_{kl} + \Sigma_{i0}) \\ b_j &= \tfrac{1}{2}(-\mathrm{i}\varepsilon_j{}^{kl}\Sigma_{kl} - \Sigma_{j0}). \end{aligned}\right\} \tag{3.163}$$

Equations (3.162) are recognized as the commutation relations for two independent angular momentum operators **a** and **b**. Each can be represented by an infinite matrix, which can be reduced to submatrices labelled by an integer or half integer A or B, where

$$\mathbf{a}^2 = A(A+1), \quad \mathbf{b}^2 = B(B+1), \tag{3.164}$$

acting in a vector space with $2A + 1$ or $2B + 1$ dimensions. These separately irreducible representations are combined as a direct product to form representations of the Lorentz group, labelled now by the two integers or half integers (A, B), acting in a vector space of $(2A + 1)(2B + 1)$ dimensions.

For example, a vector field T^γ transforms to $\Lambda^{\gamma'}{}_\gamma T^\gamma$, so that, using (3.157) and (3.160),

$$[\Sigma_{\alpha\beta}]^\gamma{}_\delta = \delta_\alpha{}^\gamma\eta_{\beta\delta} - \delta_\beta{}^\gamma\eta_{\alpha\delta} \tag{3.165}$$

which yields, from (3.164), $A = \tfrac{1}{2}, B = \tfrac{1}{2}$. Thus, the vector field can be classified with the $(\frac{1}{2},\frac{1}{2})$ irreducible representation of the Lorentz group. This field is therefore a spin $\frac{1}{2}+\frac{1}{2}=$ spin 1 field.

Similarly, a scalar field clearly has $\Sigma_{\alpha\beta}=0$, thus being identified with the (0, 0) irreducible representation, i.e., spin 0. A second rank tensor field transforms to $\Lambda^{\gamma'}{}_{\gamma}\Lambda^{\delta'}{}_{\delta}T^{\gamma\delta}$, so the $D(\Lambda)$ is simply a product of two vector field $D(\Lambda)$. The product representation $(\frac{1}{2},\frac{1}{2})\otimes(\frac{1}{2},\frac{1}{2})$ reduces to four irreducible representations: (1, 1), (1, 0), (0, 1) and (0, 0). The tensor field therefore contains components with spins 2, 1 and 0.

Finally, for the Dirac spinor field one must choose

$$\Sigma_{\alpha\beta}=\tfrac{1}{4}[\gamma_\alpha,\gamma_\beta] \tag{3.166}$$

where γ are Dirac matrices, associated with the $(\frac{1}{2},0)$ and $(0,\frac{1}{2})$, i.e., spin $\frac{1}{2}$, irreducible representations.

We wish to generalize these considerations to curved spacetime without losing the connection with the Lorentz group. This can be achieved by employing the so-called tetrad, or vierbein formalism (see, for example, Synge 1960, Weinberg 1972), which can easily be extended to n dimensions (n-beins). The essence of this approach is to erect normal coordinates y^α_X at each spacetime point X. In terms of y^α_X the metric at X is then simply $\eta_{\alpha\beta}$. In terms of a more general coordinate system, however, the metric tensor will be more complicated, but is related to $\eta_{\alpha\beta}$ by

$$g_{\mu\nu}(x)=V^\alpha_\mu(x)V^\beta_\nu(x)\eta_{\alpha\beta} \tag{3.167}$$

where

$$V^\alpha_\mu(X)=\left(\frac{\partial y^\alpha_X}{\partial x^\mu}\right)_{x=X},\quad \alpha=0,1,2,3$$

is called a vierbein. Note that the label α refers to the local inertial frame associated with the normal coordinates y^α_X at X, while μ is associated with the general coordinate system x^μ. We adopt the convention in this section that labels from the beginning of the Greek alphabet refer to the former, and those from the end refer to the latter.

As the general coordinate system is arbitrary, we can consider the effect of changing the x^μ while leaving the y^α_X fixed. Then V^α_μ transforms as a covariant vector

$$V^\alpha_\mu\to\frac{\partial x^\nu}{\partial x'^\mu}V^\alpha_\nu. \tag{3.168}$$

Clearly, it is also possible to Lorentz transform the y^α_X arbitrarily at each point X:

$$y^\alpha{}_X \to y'^\alpha{}_X = \Lambda^\alpha{}_\beta(X) y_X{}^\beta. \tag{3.169}$$

In this case, V^α_μ transforms as a Lorentz contravariant vector

$$V^\alpha{}_\mu(X) \to \Lambda^\alpha{}_\beta(X) V^\beta{}_\mu(X), \tag{3.170}$$

which obviously leaves the metric (3.167) invariant.

If a generally covariant vector A_μ is contracted into $V_\alpha{}^\mu$, the resulting object

$$A_\alpha = V_\alpha{}^\mu A_\mu$$

transforms as a collection of four scalars under general coordinate transformations, while under the local Lorentz transformations (3.169) it behaves as a vector. Thus, by use of vierbeins, one can convert general tensors into local, Lorentz-transforming tensors, shifting the additional spacetime dependence into the vierbeins.

If expression (3.159) is written schematically as $D(\Lambda)\psi$, where ψ is a tensor field, then the derivative of ψ, $\partial_\alpha\psi$, will also be a tensor field in Minkowski space. Under Lorentz transformations, $\partial_\alpha\psi$ will become $\Lambda_\alpha{}^\beta D(\Lambda)\partial_\beta\psi$. When passing to curved spacetime, we wish to generalize the derivative ∂_α to a covariant derivative ∇_α, but retaining this simple transformation property for arbitrary local Lorentz transformations at each spacetime point:

$$\nabla_\alpha\psi \to \Lambda_\alpha{}^\beta(x) D(\Lambda(x)) \nabla_\beta\psi(x). \tag{3.171}$$

This may be achieved by defining

$$\nabla_\alpha = V_\alpha{}^\mu(\partial_\mu + \Gamma_\mu) \tag{3.172}$$

where the connection

$$\Gamma_\mu(x) = \tfrac{1}{2}\Sigma^{\alpha\beta} V_\alpha{}^\nu(x)\Big(\nabla_\mu V_{\beta\nu}(x)\Big), \tag{3.173}$$

$\Sigma^{\alpha\beta}$ being the generator of the Lorentz group associated with the particular representation $D(\Lambda)$ under which ψ transforms, and $V_{\beta\nu} = g_{\mu\nu} V_\beta{}^\mu$.

The utility of the property (3.171) is that any function of ψ and $\nabla_\alpha\psi$ that is a scalar under Lorentz transformations in Minkowski space, remains a scalar under local changes in the vierbein, as well as under general coordinate transformations. Thus, the Lagrangian of the field may be generalized to curved spacetime by replacing all derivatives ∂_α by ∇_α and contracting all vectors, tensors, etc. into n-beins ($A_\alpha \to V_\alpha^\mu A_\mu$, etc.).

For example, for a scalar field ϕ, the spin is zero and $\Sigma_{\alpha\beta} = 0$, so (3.172) reveals that $\nabla_\alpha = \partial_\alpha$. The Lagrangian (2.2) therefore becomes

$$\mathscr{L}(x) = \tfrac{1}{2}(-g)^{\frac{1}{2}}(\eta^{\alpha\beta} V_\alpha{}^\mu \partial_\mu \phi V^\nu{}_\beta \partial_\nu \phi - m^2 \phi^2). \tag{3.174}$$

The factor $[-g(x)]^{\frac{1}{2}} = \det(V^\alpha{}_\mu)$ has been included to make $\mathscr{L}(x)$ a scalar density, and hence to make the action

$$S = \int \mathscr{L}(x) \mathrm{d}^4 x \tag{3.175}$$

a scalar. Using (3.167), equation (3.174) reduces to the expression quoted in (3.24), with $\xi = 0$. The term ξR, which vanishes in flat spacetime, is the only additional geometric scalar that can be added to the Lagrangian.

The Lagrangian density for a spin $\frac{1}{2}$ field in Minkowski space is given by (2.45). In curved spacetime this becomes, using (3.166) for the Σs in (3.173),

$$\begin{aligned} \mathscr{L}(x) &= \det V\{\tfrac{1}{2}\mathrm{i}[\bar{\psi}\gamma^\alpha V_\alpha{}^\mu \nabla_\mu \psi - V_\alpha{}^\mu(\nabla_\mu \bar{\psi})\gamma^\alpha \psi] - m\bar{\psi}\psi\} \\ &= \det V\{\tfrac{1}{2}\mathrm{i}[\bar{\psi}\gamma^\mu \nabla_\mu \psi - (\nabla_\mu \bar{\psi})\gamma^\mu \psi] - m\bar{\psi}\psi\} \end{aligned} \tag{3.176}$$

where $\gamma^\mu = V_\alpha{}^\mu \gamma^\alpha$ are the curved space counterparts of the Dirac γ matrices and which, from (3.167), clearly satisfy

$$\{\gamma^\mu, \gamma^\nu\} = 2g^{\mu\nu}, \tag{3.177}$$

i.e., the curved space generalization of (2.46).

The Lagrangian (3.176), which can also be used in n dimensions, is conformally invariant in the massless limit so long as ψ transforms as

$$\psi \to \Omega^{(1-n)/2}(x)\psi \tag{3.178}$$

under the transformations (3.1).

Variation of the action S with respect to $\bar{\psi}$ yields the covariant Dirac equation

$$\mathrm{i}\gamma^\mu \nabla_\mu \psi - m\psi = 0 \tag{3.179}$$

(Fock 1929, Fock & Ivanenko 1929, Bargmann 1932, Schrödinger 1932).

Next, the electromagnetic field $F_{\mu\nu}$ is a spin 1 field. The Minkowski space expression (2.54), together with the gauge-breaking and ghost Lagrangians, must be generalized by replacing A_α by $V_\alpha{}^\mu A_\mu$ and ∂_α by ∇_α with Γ_μ given by (3.173). In this case, $\Sigma_{\alpha\beta}$ is given by (3.165) for the $(\frac{1}{2}, \frac{1}{2})$ representation of the Lorentz group. The result is, for the Maxwell term

$$\mathscr{L}(x) = -\tfrac{1}{4}(-g)^{\frac{1}{2}} F_{\mu\nu} F^{\mu\nu} \tag{3.180}$$

where

$$F_{\mu\nu} = A_{\mu;\nu} - A_{\nu;\mu} = A_{\mu,\nu} - A_{\nu,\mu} \tag{3.181}$$

(the connection terms cancelling), and for the gauge-breaking term

$$\mathscr{L}_G = -\tfrac{1}{2}\zeta^{-1}(A^{\mu}{}_{;\mu})^2. \tag{3.182}$$

The ghost term is dealt with under the scalar case, and results in the Lagrangian density (see (2.141))

$$\mathscr{L}_{\text{ghost}} = g^{\mu\nu}\partial_\mu c \partial_\nu c^*. \tag{3.183}$$

Variation of $S = \int(\mathscr{L} + \mathscr{L}_G)\, d^4x$ with respect to A_μ yields

$$F_{\mu\nu;}{}^{\nu} + \zeta^{-1}(A^{\nu}{}_{;\nu})_{;\mu} = 0 \tag{3.184}$$

or

$$A_{\mu;\nu}{}^{\nu} + R_{\mu}{}^{\rho}A_{\rho} - (1-\zeta^{-1})A_{\nu;}{}^{\nu}{}_{\mu} = 0. \tag{3.185}$$

It is important to note that this theory is only conformally invariant in four spacetime dimensions.

Detailed discussions of spin 1 fields in curved spacetime have been given, in particular, by Schrödinger (1939), Parker (1972) and Mashoon (1973).

The method used above for transforming spin 0, $\frac{1}{2}$ and 1 theories to curved spacetime can equally well be used for theories of arbitrary spin fields. Such theories have been investigated in flat spacetime by, for example, Dirac (1936), Fierz (1939), Fierz & Pauli (1939), Wichmann (1962), Barut, Muzinich & Williams (1963), Weinberg (1964*a*, *b*, *c*) and Dowker (1967*a*, *b*) (see also the review of Mohan 1968).

The consequences of arbitrary spin theories in curved spacetimes have, for example, been studied by Belifante (1940), Hatalkar (1954), Duan' (1956), Penrose (1965), Dowker & Dowker (1966*a*, *b*), Dowker (1972), Grensing (1977), Christensen & Duff (1978*a*, 1979) and Birrell (1979*b*). We shall not pursue this interesting topic any further here.

As in the case of scalar fields, one can write down curved spacetime generalizations of the Green functions and Green function equations for fields of nonzero spin. For example, the Feynman propagators for spins $\frac{1}{2}$ and 1 still have the form (2.80) and (2.86) respectively, but now satisfy the equations (cf. (2.82), (2.87))

$$[\mathrm{i}\gamma^{\mu}(x)\nabla_{\mu}^{x} - m]S_F(x, x') = [-g(x)]^{-\frac{1}{2}}\delta^n(x - x'), \tag{3.186}$$

and

$$[g_{\mu\rho}(x)\Box_x + R_{\mu\rho}(x) - (1-\zeta^{-1})\nabla_\mu^x\nabla_\rho^x]D_F^{\rho\nu}(x,x')$$
$$= [-g(x)]^{-\frac{1}{2}}\delta_\mu^\nu\delta^n(x-x'), \tag{3.187}$$

which are derived using (3.179) and (3.185) respectively. In the spin $\frac{1}{2}$ case, the curved spacetime generalization of (2.84) is

$$S_F(x,x') = [i\gamma^\mu(x)\nabla_\mu^x + m]G_F(x,x') \tag{3.188}$$

where G_F is now a bi-spinor satisfying (3.49) with $\xi = \frac{1}{4}$ and $\Box_x = g^{\mu\nu}(x)\nabla_\mu^x\nabla_\nu^x$, as can be seen by substitution of (3.188) into (3.186).

Note that as in the scalar case the differential equations for the propagators do not determine the vacuum states appearing in the curved spacetime versions of (2.80) or (2.86), without the imposition of boundary conditions. These boundary conditions will depend on the particular physical system under study. Also, as in the scalar case one can obtain adiabatic (DeWitt–Schwinger) expansions for the propagator. Such an expansion has been given for spin $\frac{1}{2}$ by Bunch & Parker (1979) using the momentum space construction of the previous section, while the results for spin $\frac{1}{2}$ and 1 have been obtained using the DeWitt–Schwinger proper time method by Christensen (1978) (see also DeWitt & Brehme 1960, DeWitt 1965, Adler, Lieberman & Ng 1977). Particular terms (those having relevance to the conformal anomaly – see §6.3) in the adiabatic expansion of the propagator for high-spin fields have been obtained by Christensen & Duff (1978*a*, 1979) and Birrell (1979*b*).

We finally note that the stress-tensor for a field theory of arbitrary spin in curved spacetime can be obtained by variation of the action with respect to the metric:

$$T_{\mu\nu}(x) = \frac{2}{[-g(x)]^{\frac{1}{2}}}\frac{\delta S}{\delta g^{\mu\nu}(x)} = \frac{V_{\alpha\mu}(x)}{\det[V(x)]}\frac{\delta S}{\delta V_\alpha^\nu(x)} \tag{3.189}$$

(the factor of $(-g)^{-\frac{1}{2}}$ being inserted to give a tensor rather than a tensor density). In particular, the stress-tensors for spin $0, \frac{1}{2}$, and 1 are found to be

$$T_{\mu\nu}(s=0) = (1-2\xi)\phi_{;\mu}\phi_{;\nu} + (2\xi-\tfrac{1}{2})g_{\mu\nu}g^{\rho\sigma}\phi_{;\rho}\phi_{;\sigma} - 2\xi\phi_{;\mu\nu}\phi$$
$$+\frac{2}{n}\xi g_{\mu\nu}\phi\Box\phi - \xi\left[R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \frac{2(n-1)}{n}\xi Rg_{\mu\nu}\right]\phi^2$$
$$+2\left[\tfrac{1}{4} - \left(1-\frac{1}{n}\right)\xi\right]m^2 g_{\mu\nu}\phi^2 \tag{3.190}$$

$$T_{\mu\nu}(s=\tfrac{1}{2})=\tfrac{1}{2}\mathrm{i}[\bar{\psi}\gamma_{(\mu}\nabla_{\nu)}\psi-(\nabla_{(\mu}\bar{\psi})\gamma_{\nu)}\psi] \tag{3.191}$$

$$T_{\mu\nu}(s=1)=T_{\mu\nu}^{\mathrm{Maxwell}}+T_{\mu\nu}^{\mathrm{G}}+T_{\mu\nu}^{\mathrm{ghost}} \tag{3.192}$$

$$T_{\mu\nu}^{\mathrm{Maxwell}}=\tfrac{1}{4}g_{\mu\nu}F^{\rho\sigma}F_{\rho\sigma}-F_{\mu}{}^{\rho}F_{\rho\nu} \tag{3.193}$$

$$T_{\mu\nu}^{\mathrm{G}}=\zeta^{-1}\{A_{\mu}A^{\rho}{}_{;\rho\nu}+A^{\rho}{}_{;\rho\mu}A_{\nu}-g_{\mu\nu}[A^{\rho}A^{\sigma}{}_{;\sigma\rho}+\tfrac{1}{2}(A^{\rho}{}_{;\rho})^{2}]\} \tag{3.194}$$

$$T_{\mu\nu}^{\mathrm{ghost}}=-c^{*}_{;\mu}c_{;\nu}-c^{*}_{;\nu}c_{;\mu}-g_{\mu\nu}g^{\rho\sigma}c^{*}_{;\rho}c_{;\sigma}. \tag{3.195}$$

These equations are obtained using (3.189) with the Lagrangian densities (3.24) and (3.176) for $s=0$ and $s=\frac{1}{2}$ respectively, while the spin 1 result is obtained using the sum of the Lagrangian densities (3.180), (3.182) and (3.183). In obtaining the spin 0 and $\frac{1}{2}$ results in the symmetrized form given, the relevant field equations have been applied, and in the cases of spin $\frac{1}{2}$ and 1 integration by parts has been used. We also note the following variational formulae which are useful for the derivation of these results:

$$\left.\begin{aligned}\delta g^{\mu\nu}&=-g^{\mu\rho}g^{\nu\sigma}\delta g_{\rho\sigma}\\ \delta(-g)^{\frac{1}{2}}&=\tfrac{1}{2}(-g)^{\frac{1}{2}}g^{\mu\nu}\delta g_{\mu\nu}\\ \delta R&=-R^{\mu\nu}\delta g_{\mu\nu}+g^{\rho\sigma}g^{\mu\nu}(\delta g_{\rho\sigma;\mu\nu}+\delta g_{\rho\mu;\sigma\nu})\\ \delta g_{\mu\nu}&=-[g_{\mu\rho}V^{\alpha}{}_{\nu}+g_{\nu\rho}V^{\alpha}{}_{\mu})\delta V_{\alpha}{}^{\rho}.\end{aligned}\right\} \tag{3.196}$$

4

Flat spacetime examples

Having invested so much effort in mastering curved space quantum field theory, the reader may be dismayed to return to the topic of flat spacetime. Flat spacetime does not, however, imply Minkowski space quantum field theory.

We consider three main topics in which the general curved spacetime formalism must be applied to achieve sensible results, even though the geometry is flat. This enables some non-trivial geometrical effects to be explored within the considerable simplification afforded by a flat geometry. In particular, we are able to discuss $\langle T_{\mu\nu} \rangle$ in some special cases without employing the full theory of curved space regularization and renormalization to be developed in chapter 6.

The first case examines the effects of a non-trivial topology. We do not treat particle creation at this stage, but limit the discussion to $\langle T_{\mu\nu} \rangle$, which is nonzero even for the vacuum. This topic is one of the few in our subject which makes contact with laboratory physics, for the disturbance to the electromagnetic vacuum induced by the presence of two parallel conducting plates is actually observable. The force of attraction that appears is called the Casimir effect, and has been extensively discussed in the literature.

The treatment of boundary surfaces leads naturally to a very simple, yet extremely illuminating, system that is well worth studying in detail. This is the case of the 'moving mirror', in which a boundary at which the quantum field is constrained moves about. The ensuing creation of particles provides a heuristic model of more complicated systems to be treated later. In particular, the evaporating black hole enjoys a very close analogue in a certain type of accelerating mirror arrangement. The reader is especially recommended to follow the details leading to the evaluation of the Bogolubov transformation (4.60), as these are identical to the black hole case, and so are not repeated in chapter 8.

The third topic in this chapter is also closely related to the black hole case, but is of considerable intrinsic interest. It concerns the experiences of an observer (particle detector) that accelerates uniformly through the

Minkowski vacuum state. The result – that the accelerating observer perceives a bath of radiation with an apparently thermal spectrum – is an intriguing one that casts light on the relationship between event horizons, quantum field theory and entropy. Once again, attention is especially directed to the discussion of the Bogolubov transformation on pages 114–116. Although somewhat technical, the treatment (particularly of the analyticity properties of the field modes) is directly relevant to the black hole system, and will repay careful study.

4.1 Cylindrical two-dimensional spacetime

The simplest generalization of Minkowski space quantum field theory is the introduction of non-trivial topological structures in a locally flat spacetime. The easiest such generalization is the $R^1 \times S^1$ two-dimensional spacetime with compactified (closed) spatial sections. This spacetime has the two-dimensional Minkowski space line element (3.8) or (3.10), but the spatial points x and $x + L$ are identified, where L is the periodicity length ('circumference of the universe'). This spacetime is shown in fig. 10.

The effect of the space closure is to restrict the field modes (2.11) to a discrete set (cf. (2.13)):

$$u_k = (2L\omega)^{-\frac{1}{2}} e^{i(kx - \omega t)} \tag{4.1}$$

where $k = 2\pi n/L, n = 0, \pm 1, \pm 2, \pm 3, \ldots$ Restricting attention to the massless case, $\omega = |k|$, the modes labelled by positive values of n have the form

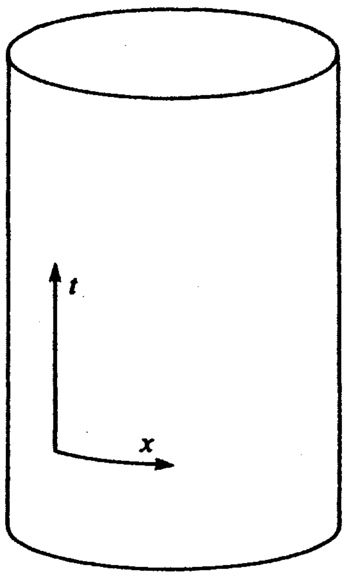


Fig. 10. Two-dimensional spacetime with compact spatial sections ($R^1 \times S^1$). The circumference of the cylinder is L.

$\exp[ik(x-t)]$ and represent waves that move from left to right, while negative values of n give $\exp[ik(x+t)]$, which represent left-moving waves.

In arriving at (4.1) we have imposed periodic boundary conditions on u_k (i.e., $u_k(t,x)=u_k(t,x+nL)$). One can also consider imposing antiperiodic boundary conditions $u_k(t,x)=(-1)^n u_k(t,x+nL)$, in which case the modes are given by (4.1) but with $k=2\pi(n+\frac{1}{2})/L, n=0, \pm 1, \pm 2, \pm 3, \ldots$ In the latter case the scalar field is to be regarded as a section through a non-product bundle and is sometimes referred to as a *twisted* field (see Isham 1978*b*).

Because the field modes are forced into a discrete set the field energy will be disturbed. In this two-dimensional case one finds from (2.26) the Cartesian components of the stress-tensor operator to be

$$T_{tt}=T_{xx}=\frac{1}{2}\left(\frac{\partial\phi}{\partial t}\right)^2+\frac{1}{2}\left(\frac{\partial\phi}{\partial x}\right)^2 \tag{4.2}$$

$$T_{tx}=T_{xt}=\frac{\partial\phi}{\partial t}\frac{\partial\phi}{\partial x}. \tag{4.3}$$

We shall evaluate $\langle 0_L|T_{\mu\nu}|0_L\rangle$, where $|0_L\rangle$ is the vacuum associated with the discrete modes (4.1). This state clearly has the property: $|0_L\rangle \to |0\rangle$ as $L\to\infty$, $|0\rangle$ being the usual Minkowski space vacuum. (See Fulling 1973 for a discussion of the inequivalence of $|0_L\rangle$ and $|0_{L'}\rangle$ for $L\neq L'$.)

Using (2.43) with the modes (4.1) (and hence the vacuum $|0_L\rangle$) one obtains

$$\langle 0_L|T_{tt}|0_L\rangle=(1/2L)\sum_{n=-\infty}^{\infty}|k|=(2\pi/L^2)\sum_{n=0}^{\infty}n \tag{4.4}$$

which is clearly infinite. This was expected, as the $R^1\times S^1$ system suffers from the same ultraviolet divergence properties as Minkowski space. The compactified spatial sections can modify the long wavelength modes, but the ultraviolet behaviour is unchanged.

In the case of the Minkowski space vacuum energy calculation given in §2.4, the ultraviolet divergence was removed by normal ordering with respect to the creation and annihilation operators of the Fock space associated with the modes (2.11). In the case of a general state $|\psi\rangle$ in this Fock space normal ordering reduces to

$$\langle\psi|:T_{\alpha\beta}:|\psi\rangle=\langle\psi|T_{\alpha\beta}|\psi\rangle-\langle 0|T_{\alpha\beta}|0\rangle, \tag{4.5}$$

which, in particular, guarantees $\langle 0|:T_{\alpha\beta}:|0\rangle=0$. If we consider Minkowski space as the covering space of $R^1\times S^1$, then $|0_L\rangle$ can be considered as a state in the above Fock space (Fulling 1973) and one can remove the divergence in $\langle 0_L|T_{\alpha\beta}|0_L\rangle$ by applying (4.5). (For a rigorous discussion of this step see Kay 1979.) In particular

$$\begin{aligned}\langle 0_L|:T_{tt}:|0_L\rangle &\equiv \langle 0_L|T_{tt}|0_L\rangle - \langle 0|T_{tt}|0\rangle \\ &= \langle 0_L|T_{tt}|0_L\rangle - \lim_{L'\to\infty} \langle 0_{L'}|T_{tt}|0_{L'}\rangle. \end{aligned} \tag{4.6}$$

Because both terms on the right-hand side of (4.6) are individually divergent, they cannot be subtracted without careful analysis. We defer the detailed discussion of this issue until chapter 6. Here we follow the simple procedure of introducing the cut-off factor $e^{-\alpha|k|}$ into the divergent sums of the type (4.4), and let $\alpha\to 0$ at the end of the calculation. Although *ad hoc*, this step is justified by the rigorous treatment given later.

With the cut-off factor, the sum (4.4) is finite and is readily performed:

$$\langle 0_L|T_{tt}|0_L\rangle = (2\pi/L^2)\sum_{n=0}^{\infty} n e^{-2\pi\alpha n/L} = (2\pi/L^2)e^{2\pi\alpha/L}(e^{2\pi\alpha/L}-1)^{-2}, \tag{4.7}$$

which may be expanded about $\alpha = 0$

$$\langle 0_L|T_{tt}|0_L\rangle = (1/2\pi\alpha^2) - (\pi/6L^2) + O(\alpha^3)$$

with a similar expansion for $\langle 0_{L'}|T_{tt}|0_{L'}\rangle$. Thus,

$$\lim_{L'\to\infty} \langle 0_{L'}|T_{tt}|0_{L'}\rangle = 1/2\pi\alpha^2.$$

Substituting these results into (4.6) and taking $\alpha\to 0$ one finds

$$\langle 0_L|:T_{tt}:|0_L\rangle = -\pi/6L^2.$$

This procedure can also be used to show that $\langle 0_L|:T_{tx}:|0_L\rangle = 0$.

Although $\langle T_{\alpha\beta}\rangle$ diverges when evaluated for both states $|0\rangle$ and $|0_L\rangle$ the difference between the two results is finite. Thus, if we require that $\langle 0|:T_{\alpha\beta}:|0\rangle = 0$, then the state $|0_L\rangle$ contains a finite, negative energy density

$$\rho = \langle 0_L|:T_{tt}:|0_L\rangle = -\pi/6L^2 \tag{4.8}$$

and pressure

$$p = \langle 0_L|:T_{xx}:|0_L\rangle = -\pi/6L^2. \tag{4.9}$$

The cloud of negative vacuum energy is distributed uniformly throughout the $R^1\times S^1$ universe with total energy $-\pi/6L$.

In the case of twisted fields the vacuum energy can be calculated

(Isham 1978*b*) along similar lines. Instead of (4.7) one now has

$$\langle 0_L|T_{tt}|0_L\rangle = (\pi/2L^2)\sum_{n=-\infty}^{\infty}|2n+1|\mathrm{e}^{-\pi\alpha|2n+1|/L}$$
$$= (\pi/2L^2)\left[\sum_{n=-\infty}^{\infty}|n|\mathrm{e}^{-\pi\alpha|n|/L} - \sum_{n=-\infty}^{\infty}2|n|\mathrm{e}^{-2\pi\alpha|n|/L}\right]$$
$$= 2[2\pi/(2L)^2]\sum_{n=0}^{\infty} n\mathrm{e}^{-2\pi\alpha n/(2L)} - (2\pi/L^2)\sum_{n=0}^{\infty} n\mathrm{e}^{-2\pi\alpha n/L}, \quad (4.10)$$

the second line resulting from the fact that the sum over odd n is the same as the sum over all n minus the sum over even n. The final term of (4.10) is precisely minus the sum in (4.7), while the preceding term is twice the sum in (4.7) with L replaced by $2L$. One can therefore read off the finite part of ρ from (4.8) as

$$\rho = -2\pi/6(2L)^2 + \pi/6L^2 = \pi/12L^2. \quad (4.11)$$

The vacuum energy of the twisted scalar field is $-\frac{1}{2}$ of that for untwisted fields.

Finally, for a massless spin $\frac{1}{2}$ field, use of (2.51) and (3.191) specialized to flat spacetime yields (Davies & Unruh 1977)

$$\langle 0_L|T_{\mu\nu}|0_L\rangle = \tfrac{1}{2}\mathrm{i}\sum_{\pm s}\sum_k[\bar{v}_{k,s}\gamma_{(\mu}\partial_{\nu)}v_{k,s} - \partial_{(\nu}\bar{v}_{k,s}\gamma_{\mu)}v_{k,s}], \quad (4.12)$$

from which, using (2.48) and (2.50), one obtains

$$\langle 0_L|T_{tt}|0_L\rangle = -(1/2L)\sum_{\pm s}\sum_k v^\dagger(k,s)v(k,s)$$
$$= -(2/L)\sum_k|k| = \begin{cases} -(8\pi/L^2)\sum_{n=0}^{\infty} n & \text{untwisted} \\ -(2\pi/L^2)\sum_{n=-\infty}^{\infty}|2n+1| & \text{twisted} \end{cases} \quad (4.13)$$

In both cases the result is simply minus four times the corresponding scalar field result.

Twisted spinor fields can be introduced in precisely the same way as twisted scalar fields (Isham 1978*c*, Avis & Isham 1979*a*, *b*) and their vacuum energy has been calculated in various topologies by DeWitt, Hart & Isham (1979) and Banach & Dowker (1979). Furthermore, Avis & Isham (1979*a*, *b*) have argued that, in most spacetimes with non-trivial topology, if the vacuum generating functional $Z(0,0)$ (see (2.129)) is to be invariant under

Lorentz transformations (3.170) of the vierbein, then one *must* include both twisted and untwisted spinors in one's calculations. Thus, twisted fields should not be considered as a mathematical curiosity, but rather as being equally as important as untwisted fields.

4.2 Use of Green functions

Rather than working directly with the stress-tensor $T_{\mu\nu}$ as given by (4.2) and (4.3), and using an arbitrary cut-off function in the mode integrals, the results of the previous section can be obtained more elegantly by working instead with the Green functions defined in §2.7.

It is also convenient to work in the null coordinates u and v defined in (3.9). In terms of u and v the stress-tensor components $T_{\mu\nu}(u, v)$ for a massless scalar field in two dimensions are simply

$$T_{uu} = (\partial_u\phi)^2 \tag{4.14}$$

$$T_{vv} = (\partial_v\phi)^2 \tag{4.15}$$

$$T_{uv} = T_{vu} = \tfrac{1}{2}\partial_u\phi\partial_v\phi. \tag{4.16}$$

Also

$$T_{tt} = T_{uu} + T_{vv} + 2T_{uv} \tag{4.17}$$

$$T_{xx} = T_{uu} + T_{vv} - 2T_{uv} \tag{4.18}$$

$$T_{tx} = T_{xt} = T_{vv} - T_{uu}. \tag{4.19}$$

It follows from (4.14) and (2.66) that

$$\langle 0_L|T_{uu}(u, v)|0_L\rangle = \lim_{v'', v' \to v} \lim_{u'', u' \to u} \partial_{u''}\partial_{u'}\tfrac{1}{2}D^{(1)}_L(u'', v''; u', v') \tag{4.20}$$

where we have affixed the subscript L to $D^{(1)}$ to indicate that it is calculated in the $R^1 \times S^1$ vacuum $|0_L\rangle$ of the previous section. (Recall that a Green function is denoted by D if it is that of a massless field.) We have also chosen to write (4.20) in a form which is symmetric under interchange of (u'', v'') with (u', v'). In addition, one may treat a thermal state $|\beta_L\rangle$ at temperature $T = (k_B\beta)^{-1}$ by using the thermal Green function $D^{(1)}_{L,\beta}$ (see (2.111), generalizing it to $R^1 \times S^1$) in place of $D^{(1)}_L$:

$$\langle \beta_L|T_{uu}|\beta_L\rangle = \lim_{v'', v' \to v} \lim_{u'', u' \to u} \partial_{u''}\partial_{u'}\tfrac{1}{2}D^{(1)}_{L,\beta}(u'', v''; u', v'), \tag{4.21}$$

where (4.21) reduces to (4.20) as $\beta \to \infty$ $(T \to 0)$.

We first construct the zero-temperature Green function using the modes

(4.1). We have

$$\begin{aligned}
D_L^{(1)}(u'',v'';u',v') &= \langle 0_L|\{\phi(u'',v''),\phi(u',v')\}|0_L\rangle \\
&= \sum_{n=-\infty}^{\infty} [u_k(u'',v'')u_k^*(u',v') + \text{c.c.}] \\
&= (1/2\pi)\sum_{n=1}^{\infty} n^{-1}(\mathrm{e}^{(-2\pi n i/L)\Delta u} + \mathrm{e}^{(-2\pi n i/L)\Delta v}) + \text{c.c.}
\end{aligned} \tag{4.22}$$

where $\Delta u = u'' - u'$, $\Delta v = v'' - v'$. In arriving at (4.22) we have discarded the infinite term that arises from $n = 0$ in the above summation. This infrared (zero frequency) divergence is a feature of two-dimensional spacetime massless quantum field theory. The discarded term would in any case have disappeared had the differentiation in (4.21) been performed before passing to the massless limit.

Because the exponents in (4.22) are purely imaginary, the summations are not absolutely convergent. The Green function $D_L^{(1)}$ must be defined carefully in the distributional sense, and some attention to this topic has been given by Bunch, Christensen & Fulling (1978). Here we shall simply insert into the exponent an infinitesimal real part of the appropriate sign to make the summations converge absolutely. This will always be done in the similar cases which follow.

Performing the summations in (4.22) yields

$$D_L^{(1)}(u'',v'';u',v') = -(1/4\pi)\ln[16\sin^2(\pi\Delta u/L)\sin^2(\pi\Delta v/L)]. \tag{4.23}$$

If instead we consider a twisted scalar field, then $D_L^{(1)}$ is given by the second line of (4.22) with n replaced by $n + \frac{1}{2}$. (Note that there is no infrared divergent term in this case.) One obtains

$$D_L^{(1)}(u'',v'';u',v') = -(1/4\pi)\ln[\tan^2(\pi\Delta u/2L)\tan^2(\pi\Delta v/2L)]. \tag{4.24}$$

The thermal Green function is now obtained from $D_L^{(1)}$ by taking the infinite image sum in (2.111):

$$D_{L,\beta}^{(1)} = -(1/4\pi)\sum_{m=-\infty}^{\infty}\ln\{16\sin^2[\pi(\Delta u + im\beta)/L]\sin^2[\pi(\Delta v + im\beta)/L]\} \qquad \text{(untwisted)} \tag{4.25}$$

$$D_{L,\beta}^{(1)} = -(1/4\pi)\sum_{m=-\infty}^{\infty}\ln\{\tan^2[\pi(\Delta u + im\beta)/2L]\tan^2[\pi(\Delta v + im\beta)/2L]\} \qquad \text{(twisted).} \tag{4.26}$$

The $m = 0$ term is the zero-temperature Green function $D_L^{(1)}$.

The stress-tensor now follows using (4.21):

$$\begin{aligned}\langle \beta_L | T_{uu} | \beta_L \rangle = & \lim_{\Delta u \to 0} -(\pi/4L^2)\operatorname{cosec}^2(\pi\Delta u/L) \\ & + (\pi/2L^2)\sum_{m=1}^{\infty} \operatorname{cosech}^2(\pi m\beta/L), \quad \text{(untwisted)} \\ = & \lim_{\Delta u \to 0} (\pi/16L^2)[1 - \operatorname{cosec}^2(\pi\Delta u/2L)] \\ & - (\pi/8L^2)\sum_{m=1}^{\infty} [\operatorname{sech}^2(\pi m\beta/2L) \\ & + \operatorname{cosech}^2(\pi m\beta/2L)] \quad \text{(twisted).}\end{aligned}$$

In the zero-temperature limit, $\beta \to \infty$, only the first (m-independent) terms survive. As expected, these diverge like $(\Delta u)^{-2}$ in the limit $\Delta u \to 0$. This is the infinite vacuum energy. To renormalize, we subtract

$$\langle 0 | T_{uu} | 0 \rangle = \lim_{L \to \infty} \langle 0_L | T_{uu} | 0_L \rangle = -1/4\pi\Delta u^2.$$

We may then take the limit $\Delta u \to 0$ and arrive at the finite result

$$-(\pi/12L^2) + (\pi/2L^2)\sum_{m=1}^{\infty} \operatorname{cosech}^2(\pi m\beta/L) \quad \text{(untwisted)}$$

$$(\pi/24L^2) + (\pi/8L^2)\sum_{m=1}^{\infty} [\operatorname{sech}^2(\pi m\beta/2L) + \operatorname{cosech}^2(\pi m\beta/2L)] \quad \text{(twisted)} \tag{4.27}$$

Because of the symmetry of $D^{(1)}_{L,\beta}$ under interchange of u and v it readily follows that

$$\langle \beta_L | T_{vv} | \beta_L \rangle = \langle \beta_L | T_{uu} | \beta_L \rangle. \tag{4.28}$$

Also, the fact that $D^{(1)}_{L,\beta}$ can be written as a v-independent function plus a u-independent function gives from (4.16)

$$\langle \beta_L | T_{uv} | \beta_L \rangle = \langle \beta_L | T_{vu} | \beta_L \rangle = 0. \tag{4.29}$$

Then, from (4.17), the energy densities are simply twice the quantities in (4.27), which agree in the zero-temperature limit ($\beta \to \infty$) with (4.8) and (4.11) respectively.

4.3 Boundary effects

So far we have restricted attention to manifolds without boundaries. Even if spacetime itself is unbounded, the quantum field may still be constrained by the presence of material boundaries. For example, an electromagnetic field will be modified in the presence of conducting surfaces. This possibility offers a valuable opportunity to test in the laboratory some of the geometrical effects that underlie curved space quantum field theory.

In the previous section it was pointed out that the vacuum expectation value of the stress-tensor $T_{\mu\nu}$ formally diverges, even in flat spacetime. If the topology is non-trivial, then the difference in vacuum stress between the non-trivial and trivial spacetimes is finite and nonzero. Similarly, conducting surfaces alter the topology of the field configuration and can lead to a nonzero vacuum stress.

To investigate these possibilities we first consider the simple case of an infinite plane in unbounded four-dimensional Minkowski space (DeWitt 1975, 1979), and a massless scalar field constrained to vanish at the plane's surface (Dirichlet boundary conditions). The field modes will no longer have the form (2.5) because the field reflects from the boundary, which we take to lie along the plane $x_3 = 0$. Instead one must work with modes of the form

$$\sin|k_3|x_3 e^{ik_1x_1 + ik_2x_2 - i\omega t} \tag{4.30}$$

which vanish at $x_3 = 0$. The corresponding vacuum state will also be different.

The Green function will no longer be given by (2.79). Its form may be found using the method of images

$$D_{\mathrm{B}}^{(1)}(x, x') = \frac{1}{2\pi^2}\left(\frac{1}{(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 - x_3')^2 - (t - t')^2} - \frac{1}{(x_1 - x_1')^2 + (x_2 - x_2')^2 + (x_3 + x_3')^2 - (t - t')^2}\right) \tag{4.31}$$

which vanishes, by construction, for $x_3 = 0$ and $x_3' = 0$. The first term of this expression is identical to that of unbounded Minkowski space, and diverges quadratically as expected when $x \to x'$; when differentiated to form $\langle 0|T_{\mu\nu}|0\rangle$, as explained in the previous section, (4.31) will yield a quartically divergent expression. The effect of the presence of the boundary can be computed by subtracting (2.79), the Green function for unbounded Minkowski space, from (4.31); i.e., discarding the first term. The remaining term is finite in the limit $x \to x'$. This must now be substituted into the four-dimensional analogue of (4.20). For example, from (2.27) one obtains $(m = 0)$

$$\langle 0|T_{tt}|0\rangle_{\mathrm{B}} = \lim_{\substack{t',x'_1,x'_2,x'_3\to t,x_1,x_2,x_3\\ t'',x''_1,x''_2,x''_3\to t,x_1,x_2,x_3}} \tfrac{1}{4}(\partial_{t''}\partial_{t'} + \partial_{x''_1}\partial_{x'_1} + \partial_{x''_2}\partial_{x'_2}$$

$$+ \partial_{x''_3}\partial_{x'_3})[D_{\mathrm{B}}^{(1)}(x'',x') - D^{(1)}(x'',x')]$$

$$= -1/16\pi^2 x_3^{\,4}. \tag{4.32}$$

Similarly $\langle 0|T_{ii}|0\rangle_{\mathrm{B}} = +(16\pi^2 x_3^{\,4})^{-1}$, all other components being zero.

Far from the boundary ($x_3 \to \infty$) the vacuum stress vanishes as expected. However, the presence of the boundary clearly modifies the vacuum stress in its vicinity. Moreover, this stress actually diverges as the surface is approached ($x_3 \to 0$). Integrating over all space yields an infinite vacuum energy per unit area of the boundary surface, even though we have already subtracted from $\langle T_{\mu\nu}\rangle$ the infinite vacuum energy of unbounded Minkowski space.

The origin of this infinite surface energy is not hard to find. The Green function (4.31) is deliberately constructed to vanish at $x_3 = 0$, even when $x \to x'$. On the other hand, the function $D^{(1)}(x, x')$ given by (2.79) clearly diverges on the boundary when $x \to x'$. So when the Green function is rendered finite by subtracting the latter from the former, the difference will also diverge on the boundary. Inspection of (4.31) shows that the second term taken on its own does indeed diverge at $x_3 = 0$ as $x \to x'$. Thus, the simple device of subtracting the infinite vacuum effects associated with unbounded Minkowski space to remove the quartically divergent vacuum stress works at all spacetime points except those on the boundary, where the quartic divergence persists, in the form of an x_3^{-4} term.

It is clear from this argument that an infinite surface energy will arise quite generally when the field is constrained to vanish on a boundary of some arbitrary shape. However, although $D_{\mathrm{B}}^{(1)} - D^{(1)}$ will diverge on the boundary, it does not necessarily follow that $\langle T_{\mu\nu}\rangle$ will do so, as the latter quantity is constructed from the former by a complicated formula.

To investigate this, consider the general form that $\langle T_{\mu\nu}\rangle$ might have in the vicinity of a single plane boundary at $x_3 = 0$ (DeWitt 1979). On grounds of symmetry, this tensor can only be built out of $\eta_{\mu\nu}$ and $\hat{x}_3{}^{\mu}\hat{x}_3{}^{\nu}$ where $\hat{x}_3{}^{\mu}$ is the unit normal vector to the boundary. Moreover, it can only be a function of x_3. Consequently

$$\langle T^{\mu\nu}\rangle = f(x_3)\eta^{\mu\nu} + g(x_3)\hat{x}_3{}^{\mu}\hat{x}_3{}^{\nu}. \tag{4.33}$$

If the covariant conservation condition is imposed (see §6.3)

$$\partial_\mu \langle T^{\mu\nu}\rangle = 0 \tag{4.34}$$

it is concluded that f and g can differ only by a constant. The right-hand side of (4.33) therefore must have the form

$$g(x_3)(\eta^{\mu\nu} + \hat{x}_3{}^{\mu}\hat{x}_3{}^{\nu}) + \alpha\eta^{\mu\nu}$$

where α is a constant. The trace of this quantity is $3g(x_3) + 4\alpha$. If the stress-tensor is required to be traceless, this forces $g(x_3) = -4\alpha/3 =$ constant. As the renormalized vacuum stress must approach zero far from the boundary, the constant must be zero in this case. Hence the renormalized vacuum stress will vanish for a field with a traceless stress-tensor. This is the case for the electromagnetic and neutrino fields, but not for the massless scalar field described by $T_{\mu\nu}$ as given by (2.26).

In §3.2 it was pointed out that in curved space, the scalar wave equation can acquire an additional term ξR. The corresponding stress-tensor operator is given by (3.190). Even in the flat space limit, (3.190) does not reduce to (2.26) if $\xi \neq 0$. (This curious fact is less surprising if it is remembered that $T_{\mu\nu}$ is obtained by varying the metric $g_{\mu\nu}$ in the field Lagrangian. Thus, even if one takes the limit $g_{\mu\nu} \to \eta_{\mu\nu}$ at the end of the calculation, one retains contributions to $T_{\mu\nu}$ associated with the $\xi R\phi^2$ term in (3.24).) The choice $\xi = \frac{1}{6}$ corresponds to a conformally invariant scalar field equation, which implies a traceless $T_{\mu\nu}$:

$$T_{\mu\nu}(\xi = \tfrac{1}{6}) = \tfrac{2}{3}\phi_{,\mu}\phi_{,\nu} - \tfrac{1}{6}\eta_{\mu\nu}\eta^{\sigma\rho}\phi_{,\sigma}\phi_{,\rho} - \tfrac{1}{3}\phi\phi_{;\mu\nu} + \tfrac{1}{12}\eta_{\mu\nu}\phi\Box\phi, \quad (4.35)$$

$$T^{\mu}{}_{\mu} = 0. \quad (4.36)$$

Sometimes (4.35) is called the 'new improved stress-tensor' (Chernikov & Tagirov 1968, Callan, Coleman & Jackiw 1970). For such a scalar field $\langle 0|T_{\mu\nu}|0\rangle = 0$ near a plane boundary.

Although the above argument implies that the infinite surface stress on a plane boundary will not be present for a conformally invariant field, the reasoning depended crucially on the symmetries associated with a flat surface. If the boundary is curved, the divergent surface energy reappears. It may be shown that, in general, as the surface is approached (Deutsch & Candelas 1979, Kennedy, Critchley & Dowker 1980)

$$\langle T_{\mu\nu}\rangle \propto \varepsilon^{-3}\chi_{\mu\nu} + O(\varepsilon^{-2})$$

for a conformally invariant field, where ε is the (small) distance from the surface and $\chi_{\mu\nu}$ the second fundamental form of the boundary. We shall return to this topic, and the physical significance of the infinite surface energy, in §6.6.

Restricting attention for now to plane boundaries and conformally invariant fields, it is easy to generalize the problem to the case where more than one boundary is present. Casimir (1948) considered the vacuum energy associated with the electromagnetic field in the region between two parallel reflecting planes. The restrictions of conservation and tracelessness require that

$$\langle T^{\mu\nu}\rangle = A(\tfrac{1}{4}\eta^{\mu\nu} + \hat{x}_3{}^{\mu}\hat{x}_3{}^{\nu})$$

where A is constant and the planes are again orthogonal to $\hat{x}_3$. On dimensional grounds $A \propto a^{-4}$ where a is the separation distance between the two planes. Calculation (Casimir 1948) shows that

$$A = -\pi^2/180a^4 \tag{4.37}$$

in the vacuum case.

The reason why A is nonzero is because the boundary planes constrain the electromagnetic field modes in the x_3 direction to form a discrete set. In this respect, the field suffers a topological distortion similar to the case of the $R^1 \times S^1$ model considered in §4.1, and the constant A may be calculated by following the method of that section. Alternatively, the $D_{\mathrm{B}}^{(1)}$ Green function may be computed as an infinite image sum (infinite reflections between the planes).

In the *scalar* field case

$$D_{\mathrm{B}}^{(1)}(x,x') = \frac{1}{2\pi^2}\sum_{n=-\infty}^{\infty}\left(\frac{1}{(x_1-x_1')^2+(x_2-x_2')^2+(x_3-x_3'-an)^2-(t-t')^2} - \frac{1}{(x_1-x_1')^2+(x_2-x_2')^2+(x_3+x_3'-an)^2-(t-t')^2}\right) \tag{4.38}$$

which vanishes at x_3 or $x_3' = 0$ and x_3 or $x_3' = a$, as required. The infinite vacuum divergence may be removed by discarding the $n = 0$ term. Use of (4.35) then yields

$$\langle 0|T_{\mu\nu}|0\rangle_{\mathrm{B}} = \frac{-\pi^2}{1440a^4}\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix} \tag{4.39}$$

in Cartesian coordinates. The coefficient of (4.39) is one-half the value of the electromagnetic case on account of the fact that the scalar field has only half as many modes (i.e., one rather than two polarization states).

If the quantum field is not in a vacuum state, but at finite temperature T, then we may replace $D_B^{(1)}$ in (4.38) by a thermal Green function, i.e., replace $t - t'$ by $t - t' - im\beta$ (where $\beta = 1/k_B T$) and sum over all m from $-\infty$ to $+\infty$. The corresponding energy density between the planes is

$$-kTa^{-3}f'(\xi) + a^{-4}f(\xi)$$

where

$$f(\xi) = -\frac{1}{8\pi^2}\sum_{m,n=1}^{\infty}\frac{(2\xi)^4}{[n^2 + 4\xi^2 m^2]^2}$$

and $\xi = k_B Ta$. In the limit of small Ta, the temperature correction to the energy density is

$$(kT)^3\zeta(3)/2\pi a$$

where ζ is a zeta-function, while for large Ta it is

$$\tfrac{1}{30}\pi^2(kT)^4.$$

The electromagnetic values are twice these; a detailed treatment has been given by Brown & Maclay (1969).

The presence of a vacuum stress between parallel boundary planes implies that there will be a force of attraction between two electrically neutral conducting surfaces. Casimir's calculation (see (4.37)) indicates a force per unit area of the surfaces given by

$$F = -\frac{\partial E}{\partial a} = \frac{-\pi^2}{240a^4}.$$

Forces arising in a manner similar to this have been detected experimentally (see, for example, Sparnaay 1958 and Tabor & Winterton 1969). The 'Casimir effect' has since been the subject of much study, extremely detailed analyses having been given by Lifschitz (1955) and Balian & Duplantier (1977, 1978).

Another plane boundary problem has been investigated by Dowker & Kennedy (1978), and Deutsch & Candelas (1979). They find for the renormalized conformal scalar field vacuum stress-tensor in the wedge between two inclined planes

$$\langle T_{\mu\nu}\rangle = \frac{1}{1440r^4\alpha^2}\left(\frac{\pi^2}{\alpha^2} - \frac{\alpha^2}{\pi^2}\right)\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

where α is the angle between the planes, r is the distance from the axis of intersection and cylindrical polar coordinates are used for the spacelike part of the tensor.

The electromagnetic case yields (Deutsch & Candelas 1979)

$$\frac{1}{720\pi^2 r^4}\left(\frac{\pi^2}{\alpha^2}+11\right)\left(\frac{\pi^2}{\alpha^2}-1\right)\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 3 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}.$$

In both cases the finite answer may be obtained by the usual subtraction of the infinite vacuum stress associated with unbounded Minkowski space.

4.4 Moving mirrors

Until now, the flat spacetime examples have all involved static situations, i.e., the disturbance of the quantum state induces vacuum energy and stress, but no particles are created. If the parallel plates discussed in the preceding section in connection with the Casimir effect were moved rapidly around, the irreversible production of entropy would occur and new quanta would appear between the plates. Even the motion of a single reflecting boundary (mirror) can create particles.

We shall treat this problem in detail for a two-dimensional spacetime where the conformal triviality can be exploited. Explicit examples for four dimensions are rather sparse (Candelas & Raine 1976, 1977*b*). We follow the treatment of Fulling & Davies (1976) and Davies & Fulling (1977*a*).

Suppose the mirror (which in two dimensions degenerates to a reflecting point) moves along the trajectory

$$x = z(t), \quad |\dot{z}(t)| < 1, \tag{4.40}$$

$$z(t) = 0, \quad t < 0.$$

A massless scalar field ϕ, satisfying the field equation

$$\Box\phi = \frac{\partial^2\phi}{\partial u \partial v} = 0 \tag{4.41}$$

with reflection boundary condition

$$\phi(t, z(t)) = 0 \tag{4.42}$$

has a set of mode solutions

$$u_k^{\mathrm{in}}(u, v) = \mathrm{i}(4\pi\omega)^{-\frac{1}{2}}(\mathrm{e}^{-\mathrm{i}\omega v} - \mathrm{e}^{-\mathrm{i}\omega(2\tau_u - u)}) \tag{4.43}$$

where $\omega = |k|$ and τ_u is determined implicitly by the trajectory (4.40) through

$$\tau_u - z(\tau_u) = u. \tag{4.44}$$

The solutions (4.43) apply to the right of the mirror. The incoming (left-moving) waves $e^{-i\omega v}$ correspond to standard exponential modes all the way from $\mathscr{I}^-$ (see fig. 11) to the mirror surface, but the right-moving (reflected) waves are complicated because of the Doppler shift suffered during the reflection from the moving mirror. This asymmetry between u and v is chosen to correspond to the usual retarded boundary condition (no incoming field quanta from $\mathscr{I}^-$).

In the region to the right of the mirror the field ϕ, constrained by (4.42), can be expanded in terms of the modes (4.43) as

$$\phi = \sum_{k>0} [a_k u_k^{\text{in}} + a_k^\dagger (u_k^{\text{in}})^*] \tag{4.45}$$

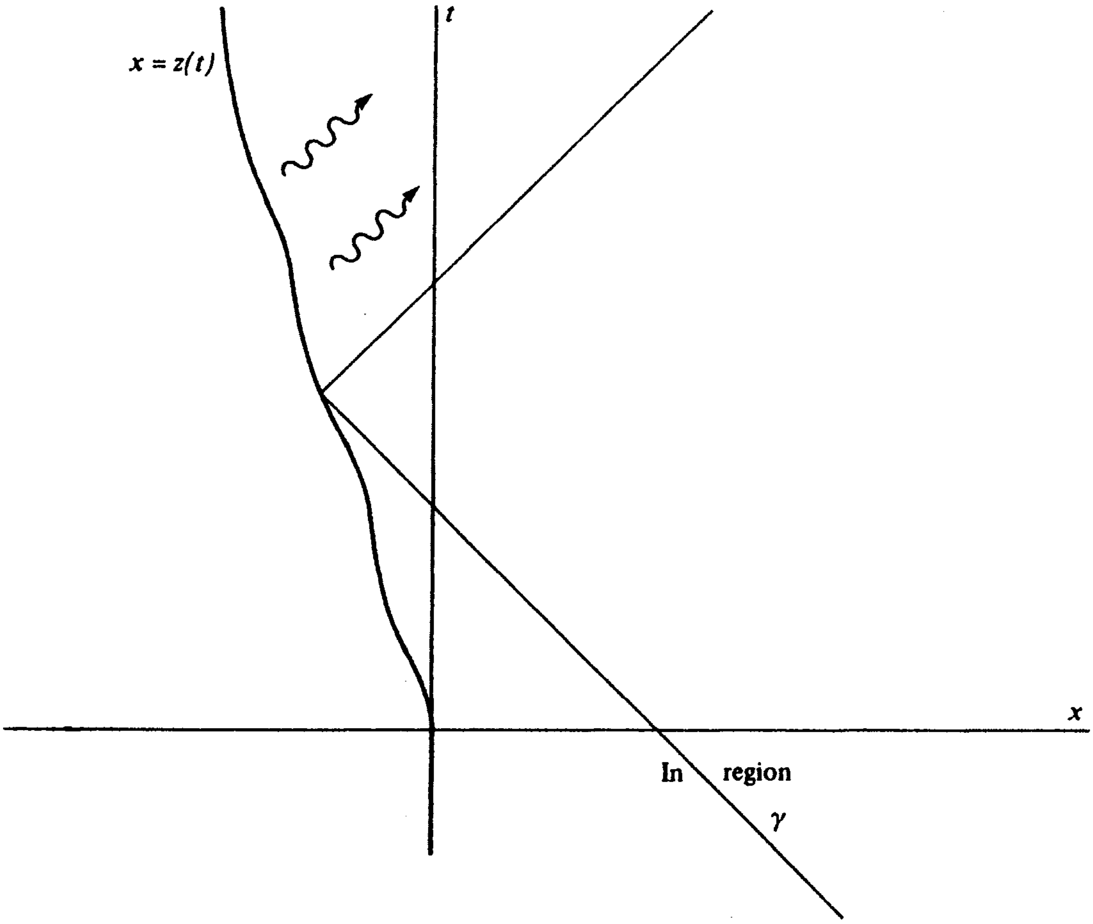


Fig. 11. Radiation by a moving mirror. The scalar field is constrained to vanish on the boundary $x = z(t)$. Null rays incoming from $\mathscr{I}^-$, such as γ, reflect from the mirror to $\mathscr{I}^+$. Wave modes reflected after $t = 0$ suffer a Doppler shift in frequency due to the 'mirror' motion.

and an 'in' vacuum defined by

$$a_k|0,\text{in}\rangle = 0. \tag{4.46}$$

We have labelled this state and the modes (4.43) by 'in' since in the in region, $t \le 0$, the mirror is at rest at $z = 0$ in the (u, v) frame. Then $\tau_u = u$ and (4.43) reduces to

$$u_k^{\text{in}} = \mathrm{i}(4\pi\omega)^{-\frac{1}{2}}(\mathrm{e}^{-\mathrm{i}\omega v} - \mathrm{e}^{-\mathrm{i}\omega u}) = (\pi\omega)^{-\frac{1}{2}} \sin\omega x\, \mathrm{e}^{-\mathrm{i}\omega t}, \quad t \le 0, \tag{4.47}$$

which is a positive frequency mode with respect to Minkowski time t.

The state $|0,\text{in}\rangle$ may be considered to be void of particles for $t < 0$. Indeed, the Wightman function in the in region is

$$\begin{aligned}&\langle \text{in}, 0|\phi(x)\phi(x')|0,\text{in}\rangle \\ &\quad = -(1/4\pi)\ln[(\Delta u - \mathrm{i}\varepsilon)(\Delta v - \mathrm{i}\varepsilon)/(v - u' - \mathrm{i}\varepsilon)(u - v' - \mathrm{i}\varepsilon)]\end{aligned} \tag{4.48}$$

where $\Delta u = u - u'$, $\Delta v = v - v'$ and $t, t' < 0$. It is easily verified by substituting (4.48) into (3.55) that in spite of the more complicated argument of the logarithm an inertial particle detector, with trajectory (3.52), which is adiabatically switched off outside the in region, records no particles. The presence of the mirror does not excite the detector, even if it is in uniform relative motion with respect to the detector.

If the mirror undergoes a period of acceleration when $t > 0$, the field modes will suffer distortion for $u > 0$, from the regular form (4.47) to the general form (4.43). The moving mirror therefore plays the same rôle as a time-dependent background geometry (i.e., gravitational field). However, notice that the function $2\tau_u - u$ in (4.43) is unchanged all along the null ray $u =$ constant from the mirror surface right out to $\mathscr{I}^+$. The distortion of the modes occurs suddenly (upon reflection) rather than gradually during an extended period of geometrical disruption, as in the gravitational case.

Consider the right-moving piece of (4.43), i.e., $\exp[-\mathrm{i}\omega(2\tau_u - u)]$. This represents a wave that reduces in the region $u < 0$ to the standard form for a right-moving wave, $\mathrm{e}^{-\mathrm{i}\omega u}$, and is therefore associated for $t < 0$ with the usual physical vacuum state defined by (4.46). On the other hand it does not reduce to standard right-moving waves in the region $u > 0$. Thus $|0,\text{in}\rangle$ no longer represents the physical vacuum in this region, but rather it will generally be a state containing particles. That is, the Doppler distortion described by the complicated exponential in (4.43) excites the field modes and causes particles to appear. Physically this is described by saying that the moving mirror creates particles, which stream away to the right along the null rays $u =$ constant.

To confirm the physical reality of these quanta, one can examine the experiences of an inertial particle detector in the region $u > 0$. The Wightman function in the in vacuum is

$$D^+(u, v; u', v')$$
$$= -(1/4\pi)\ln[(p(u) - p(u') - i\varepsilon)(v - v' - i\varepsilon)/(v - p(u') - i\varepsilon)(p(u) - v' - i\varepsilon)] \tag{4.49}$$

where we have defined

$$p(u) = 2\tau_u - u. \tag{4.50}$$

For a general $x(t)$, hence $p(u)$, the Fourier transform of (4.49) will be nonzero, so when D^+ is substituted in (3.55) it predicts a nonzero response from the detector.

As an illustration, consider mirror trajectories with the asymptotic form

$$z(t) \to -t - Ae^{-2\kappa t} + B \text{ as } t \to \infty \tag{4.51}$$

where A, B, κ are constants > 0. The behaviour at earlier times will be irrelevant as we shall only consider the particle flux in the asymptotic region as $t \to \infty$. However, one example of a trajectory with asymptotic form (4.51), which joins smoothly (C^1) onto a static trajectory for $t < 0$, is $z(t) = -\ln(\cosh \kappa t)$.

The class (4.51) of trajectories is of special geometrical interest because only the null rays with $v < B$ can reflect from the mirror. All rays with $v > B$ pass undisturbed to the left-hand portion of $\mathscr{I}^+$ (see fig. 12). The ray $v = B$ therefore acts as a sort of horizon, and equispaced lines of constant u, when traced back, through reflection, to $\mathscr{I}^-$ crowd up along $v = B$. In chapter 8 we shall see that an identical situation occurs in the collapse of a star to form a black hole.

From (4.44) and (4.51) one obtains

$$p(u) \equiv 2\tau_u - u \to B - Ae^{-\kappa(u+B)} \text{ as } u \to \infty. \tag{4.52}$$

We assume that the detector moves along the trajectory

$$x = x_0 + wt, \quad w \text{ constant.} \tag{4.53}$$

The expression (4.49) for D^+ may be written as a sum of four terms by factorizing the argument of the logarithm. When $p(u)$ is chosen to be (4.52) and the resulting D^+ is used in (3.55) for the detector response function (assuming the detector to be switched off adiabatically at early times), three of the terms give vanishing contributions. The remaining term, involving

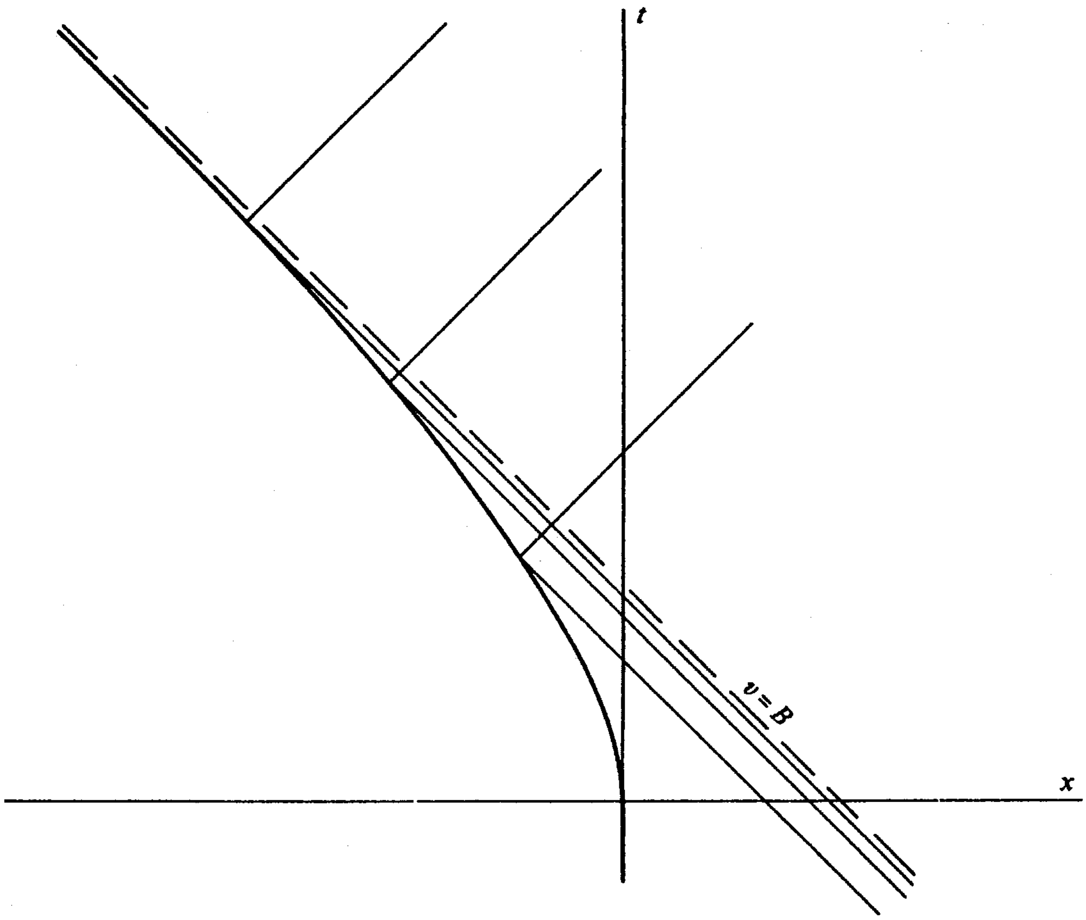


Fig. 12. The mirror trajectory, $z(t) = -t - Ae^{-2\kappa t} + B$, is asymptotic to the null ray, $v = B$. Late time surfaces of constant phase from $\mathscr{I}^+$ (equispaced null rays u = constant) crowd up along $v = B$ when reflected back off the mirror. Advanced (left-moving) rays later than $v = B$ do not intersect the mirror at all. The radiation excited from the vacuum by this mirror trajectory has a thermal spectrum at late times.

$p(u') - p(u)$, yields

$\mathscr{F}(E)$/unit time

$$= -(1/4\pi)\int_{-\infty}^{\infty} e^{-iE\Delta\tau}\ln\{\sinh[\tfrac{1}{2}\kappa((1-w)/(1+w))^{\frac{1}{2}}\Delta\tau - i\varepsilon]\}d\Delta\tau. \tag{4.54}$$

Using the identities

$$\sinh x = x\prod_{m=1}^{\infty}(m\pi - ix)(m\pi + ix)/(m\pi)^2$$

and

$$\int_{\sigma}^{\infty}\frac{e^{-i\omega x}}{\omega(e^{\beta\omega}-1)}d\omega \underset{\sigma\to 0}{\to} -\ln\left[\prod_{m=1}^{\infty}\sigma e^{\gamma}(\beta m + ix)\right]$$

where γ is Euler's constant, the right-hand side of (4.54) may be reduced to

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iE\Delta\tau}\int_{0}^{\infty}\frac{\cos\{\frac{1}{2}\omega\kappa\Delta\tau[(1-w)/(1+w)]^{\frac{1}{2}}\}}{\omega(e^{\pi\omega}-1)}d\omega d\Delta\tau.$$

Interchanging the order of integration gives a sum of two δ-functions, only one of which contributes to the ω integral giving

$$\mathscr{F}(E)/\text{unit time} = \frac{1}{E(e^{E/k_BT}-1)}, \tag{4.55}$$

where

$$k_BT = (\kappa/2\pi)[(1-w)/(1+w)]^{\frac{1}{2}}. \tag{4.56}$$

Inspection of this result shows that the detector responds to a flux of particles from the mirror that is constant in time and has the spectrum of thermal radiation travelling in the u-direction with a temperature given by (4.56). The factor $[(1-w)/(1+w)]^{\frac{1}{2}}$ is the Doppler shift due to the motion of the detector at velocity w relative to the radiation, resulting in a red- or blue-shifting of the temperature. If the detector is at rest relative to the privileged frame defined by the mirror in its initially static phase (i.e., $w = 0$) then

$$k_BT = \kappa/2\pi. \tag{4.57}$$

The thermal nature of the flux of radiation from a mirror moving along the trajectory (4.51) can also be deduced by explicitly evaluating the Bogolubov transformation between the in and the out modes. The modes u_k^{in} given by (4.43) assume a simple form on $\mathscr{I}^-$, but after reflection from the mirror they are generally complicated functions on $\mathscr{I}^+$. Conversely one can define modes u_k^{out} that are simple complex exponentials on $\mathscr{I}^+$, but complicated functions in the in region, i.e., on $\mathscr{I}^-$. The Bogolubov transformation between u_k^{in} and u_k^{out} can be evaluated on any spacelike surface; in the case of the asymptotically null trajectory (4.51) it is convenient to evaluate the transformation at $t = 0$, in the in region. The modes u_k^{in} then assume their simple form (4.47), but u_k^{out} will be complicated.

To determine the form of u_k^{out} at $t = 0$ we note that standard right-moving exponential waves $e^{-i\omega u}$ on $\mathscr{I}^+$ will, when traced back in time and reflected off the mirror into the in region, assume the form of complicated left-moving waves, $e^{-i\omega f(v)}$. Portions of the waves $e^{-i\omega u}$ at late time u will correspond to left-moving waves on $\mathscr{I}^-$ that crowd up along the null asymptote $v = B$. Thus $f(v)$ will be a rapidly varying function of v in this region. By symmetry, the function f will be the inverse of the function p defined by (4.52); in this case

$$f(v) \sim -\kappa^{-1}\ln[(B-v)/A] - B, \quad v < B, \tag{4.58}$$

in the region $v \to B$. We are not interested here in the left-moving modes in the region $v > B$, as these merely continue undisturbed to the left as far as $\mathscr{I}^+$, and do not contribute to the production of the thermal radiation from the surface of the mirror. Hence we put $u_k^{\text{out}} = 0$ for $v \geq B$.

We therefore wish to evaluate the Bogolubov transformation between u_k^{in} and that portion of u_k^{out} that corresponds to right-moving waves $e^{-i\omega u}$ travelling towards $\mathscr{I}^+$, i.e., the portion of $e^{-i\omega f(v)}$ in the region $v < B$. We use the scalar products (3.36) (with $\bar{u} = u^{\text{in}}$ and $u = u^{\text{out}}$), integrating along $t = 0$ as far as $v = B$ only. After an integration by parts, one obtains

$$\left.\begin{matrix}\alpha_{\omega'\omega}\\ \beta_{\omega'\omega}\end{matrix}\right\} = \pm(2\pi)^{-1}i(\omega'/\omega)^{\frac{1}{2}}\int_0^B e^{\pm i\omega f(x) - i\omega' x}dx. \tag{4.59}$$

Because $f(x)$ is rapidly varying near $x = B$, most of the contribution to the integral comes from this region, so we have discarded the boundary term in (4.59) at $x = 0$. We may also approximate (4.59) by using the asymptotic form (4.58) for all x. The resulting Bogolubov coefficients will then be good approximations in describing the radiation flux at late time (large u) in the out region. The integration may now be performed explicitly in terms of incomplete Γ-functions. However, the rapidly varying waveform near $v = B$ represents very high frequencies ω', so in the above approximation we may let $\omega' \to \infty$ and recover an ordinary Γ-function:

$$\left.\begin{matrix}\alpha_{\omega'\omega}\\ \beta_{\omega'\omega}\end{matrix}\right\} = \mp(4\pi^2\omega\omega')^{-\frac{1}{2}}e^{\pm\pi\omega/2\kappa}e^{\pm i\omega D - i\omega' B}(\omega')^{\pm i\omega/\kappa}\Gamma(1 \mp i\omega/\kappa) \tag{4.60}$$

where $D = \kappa^{-1}\ln A - B$, from which it follows that

$$|\beta_{\omega'\omega}|^2 = \frac{1}{2\pi\kappa\omega'}\left(\frac{1}{e^{\omega/k_BT} - 1}\right) \tag{4.61}$$

with k_BT given by (4.57). This result is in agreement with the spectrum displayed in (4.55).

Note that

$$\langle 0, \text{in}|N_\omega|\text{in}, 0\rangle = \int_0^\infty |\beta_{\omega'\omega}|^2 d\omega' \tag{4.62}$$

diverges logarithmically. This is because if the mirror continues to accelerate for all time then the steady flux of radiation will accumulate an infinite number of quanta per mode. The result may be converted into a number of quanta per $d\omega$ per unit time by constructing finite wave packets rather than plane wave modes. This topic will be discussed further in chapter 8.

As another illustration of the computation of a Bogolubov transformation in the moving mirror system we shall investigate the hyperbolic trajectory of a mirror which recedes with uniform acceleration. We take the trajectory to be

$$\begin{aligned} z(t) &= B - (B^2 + t^2)^{\frac{1}{2}}, \quad t > 0 \\ &= 0, \qquad\qquad\qquad\quad t < 0 \end{aligned} \tag{4.63}$$

where B is a constant. This is qualitatively similar to the case shown in fig. 12, also having a null asymptote along $v = B$. Again the out modes, when traced back to $t = 0$, have the form $e^{-i\omega f(v)}$ for $v < B$. Thus (4.59) still gives the Bogolubov coefficients, but with $f(x)$ appropriate to the trajectory (4.63). One finds $f(x) = Bx/(B - x)$, from which it follows that

$$\alpha_{\omega'\omega} = i(B/\pi)e^{-i(\omega + \omega')B}K_1(2iB(\omega\omega')^{\frac{1}{2}}) \tag{4.64}$$

$$\beta_{\omega'\omega} = (B/\pi)e^{i(\omega - \omega')B}K_1(2B(\omega\omega')^{\frac{1}{2}}). \tag{4.65}$$

where K_1 is a modified Bessel function. We shall return to this example in §7.1.

It is also possible to treat in detail particle production in the region between two moving mirrors (Moore 1970, Fulling & Davies 1976).

4.5 Quantum field theory in Rindler space

Two examples have now been given in which a detector responds to radiation with a thermal spectrum. In §3.3 it was shown that a uniformly accelerating detector perceives the usual Minkowski vacuum state to be a thermal bath of radiation, whereas in the previous section we saw how an inertial detector responds to a thermal flux of radiation streaming away from a mirror that recedes along the non-uniformly accelerating trajectory (4.51). These two situations are physically rather distinct, and it is usual to regard the thermal radiation in the latter case as in some sense more physical than in the former. This point of view is supported by a computation of $\langle T_{\mu\nu} \rangle$ in the two cases (see §7.1). In spite of their rather different physical status, the two examples are closely related geometrically. The conformal transformation that 'straightens' out the trajectory (4.51) simultaneously converts the straight (i.e., inertial) detector trajectory to a hyperbola, i.e., that of a uniformly accelerating detector (Davies & Fulling 1977*a*).

The thermal radiation detected by a uniformly accelerated system may be deduced by an entirely different analysis that throws important light on the intimate association between the quantum particle concept and the causal and topological structure of spacetime that will prove useful when we come to discuss quantum black holes. (For further details on this section see, for example, Fulling 1973, 1977, Davies 1975, Unruh 1976, Candelas & Deutsch 1977, Dowker 1977, Troost & van Dam 1977 and Horibe 1979.)

Consider two-dimensional Minkowski space with metric (3.8) or (3.10) i.e.,

$$ds^2 = \mathrm{d}\bar{u}\,\mathrm{d}\bar{v} = \mathrm{d}t^2 - \mathrm{d}x^2. \tag{4.66}$$

(A bar has been appended to the Minkowski space null coordinates $\bar{u}$, $\bar{v}$ for ease of comparison with later results.) Under the following coordinate transformation

$$t = a^{-1}\mathrm{e}^{a\xi}\sinh a\eta \tag{4.67}$$

$$x = a^{-1}\mathrm{e}^{a\xi}\cosh a\eta, \tag{4.68}$$

$a = \text{constant} > 0$ and $-\infty < \eta,\ \xi < \infty$, or equivalently

$$\bar{u} = -a^{-1}\mathrm{e}^{-au} \tag{4.69}$$

$$\bar{v} = a^{-1}\mathrm{e}^{av} \tag{4.70}$$

where $u = \eta - \xi$, $v = \eta + \xi$, (4.66) becomes

$$ds^2 = \mathrm{e}^{2a\xi}\mathrm{d}u\,\mathrm{d}v = \mathrm{e}^{2a\xi}(\mathrm{d}\eta^2 - \mathrm{d}\xi^2). \tag{4.71}$$

The coordinates (η, ξ) cover only a quadrant of Minkowski space, namely the wedge $x > |t|$ shown in fig. 13. Lines of constant η are straight ($x \propto t$) while lines of constant ξ are hyperbolae

$$x^2 - t^2 = a^{-2}\mathrm{e}^{2a\xi} = \text{constant}. \tag{4.72}$$

They therefore represent the world lines of uniformly accelerated observers treated in §3.3. Comparison of (3.62) with (4.72) shows that

$$a\mathrm{e}^{-a\xi} = \alpha^{-1} = \text{proper acceleration}. \tag{4.73}$$

Thus, lines of large positive ξ (far from $x = t = 0$) represent weakly accelerated observers, while the hyperbolae that closely approach $x = t = 0$ have large negative ξ and hence a high proper acceleration. All the

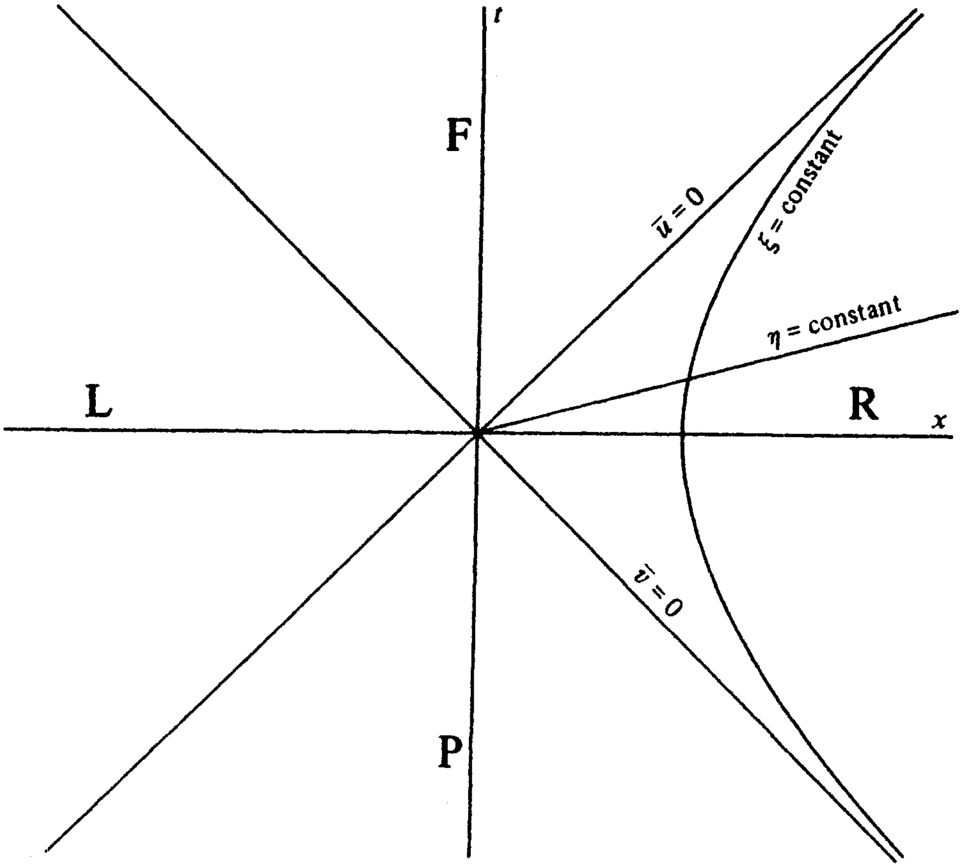


Fig. 13. Rindler coordinatization of Minkowski space. In R and L, time coordinates η = constant are straight lines through the origin, space coordinates ξ = constant are hyperbolae (corresponding to the world lines of uniformly accelerated observers) with null asymptotes $\bar{u} = 0, \bar{v} = 0$, which act as event horizons. The four regions R, L, F and P must be covered by separate coordinate patches. Rindler coordinates are non-analytic across $\bar{u} = 0$ and $\bar{v} = 0$.

hyperbolae are asymptotic to the null rays $\bar{u} = 0, \bar{v} = 0$ (or $u = \infty, v = -\infty$), which means that the accelerated observers approach the speed of light as $\eta \rightarrow \pm\infty$. These observers' proper time τ is related to ξ and η by

$$\tau = e^{a\xi}\eta. \tag{4.74}$$

The system (η, ξ) is known as the Rindler coordinate system (Rindler 1966), and the portion $x > |t|$ of Minkowski space is called the Rindler wedge. Uniformly accelerated observers are sometimes referred to as Rindler observers.

A second Rindler wedge $x < |t|$ may be obtained by reflecting the first in the t- and then the x-axis. This is achieved by changing the signs of the right-hand sides of the transformation equations (4.67)–(4.70). We label the left- and right-hand wedges by L and R respectively. Note that the sign reversals in L mean that, crudely speaking, the direction of time there is reflected, i.e., increasing t corresponds to decreasing η.

The causal structure of the Rindler wedge is interesting. Because the Rindler observers (with constant spatial coordinate ξ) approach but do not cross the null rays $u = \infty$, $v = -\infty$, these rays act as event horizons. For example, no events in L can be witnessed in R and vice versa, as events in L can only be connected with events in R by a somewhere spacelike line. Regions L and R therefore represent two causally disjoint universes. We have also marked the remaining future (F) and past (P) regions on fig. 13. Events in both P and F can be connected by null rays to both L and R.

This causal structure also appears on the Penrose conformal diagram shown in fig. 14. The Rindler observers intersect $\mathscr{I}^{\pm}$, rather than $i^{\pm}$ as do asymptotically inertial observers.Thus the null ray $u = \infty$ acts as a future event horizon, and events in the portion marked F cannot causally influence the diamond shaped R region.

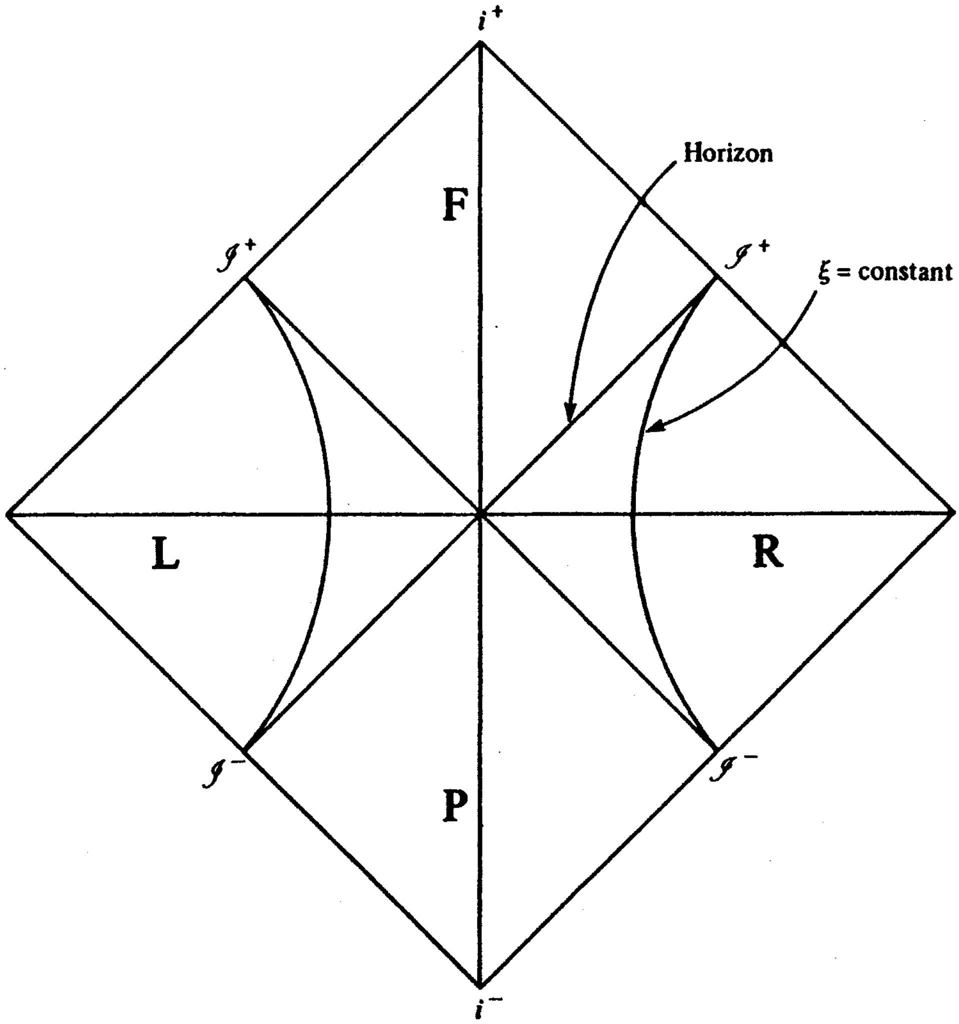


Fig. 14. Conformal diagram of Rindler system. The regions R, L, F and P are represented by diamond shaped regions. The timelike lines $\xi = \text{constant}$ intersect $\mathscr{I}^{\pm}$, not $\mathrm{i}^{\pm}$. Clearly, events in F cannot be witnessed in R, so the null ray $\bar{u} = 0$ ($u = \infty$) acts as an event horizon.

Consider the quantization of a massless scalar field ϕ in two-dimensional Minkowski spacetime. The wave equation

$$\Box\phi \equiv \left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right)\phi \equiv \frac{\partial^2\phi}{\partial\bar{u}\partial\bar{v}} = 0 \tag{4.75}$$

possesses standard orthonormal mode solutions

$$\bar{u}_k = (4\pi\omega)^{-\frac{1}{2}} e^{ikx - i\omega t} \tag{4.76}$$

(i.e., (2.11) with $n = 2$) where $\omega = |k| > 0$ and $-\infty < k < \infty$. These modes are positive frequency with respect to the timelike Killing vector ∂_t, in that they satisfy

$$\mathscr{L}_{\partial_t}\bar{u}_k = -i\omega\bar{u}_k. \tag{4.77}$$

The modes with $k > 0$ consist of right-moving waves

$$(4\pi\omega)^{-\frac{1}{2}} e^{-i\omega\bar{u}} \tag{4.78}$$

along the rays $\bar{u} =$ constant, while for $k < 0$ one has left-moving waves along $\bar{v} =$ constant

$$(4\pi\omega)^{-\frac{1}{2}} e^{-i\omega\bar{v}}\,. \tag{4.79}$$

The Minkowski vacuum state $|0_M\rangle$ and associated Fock space are constructed by expanding ϕ in terms of $\bar{u}_k$ as explained in §2.2.

In the Rindler regions R and L one may adopt an alternative quantization prescription, based not on the modes $\bar{u}_k$, but their Rindler counterparts u_k. The metric (4.71) is conformal to the whole of Minkowski space, for under the conformal transformation $g_{\mu\nu} \to e^{-2a\xi} g_{\mu\nu}$, (4.71) reduces to $d\eta^2 - d\xi^2$ with $-\infty < \eta, \xi < \infty$. Because the wave equation is conformally invariant, we can write it in Rindler coordinates as (see (3.150))

$$e^{2a\xi}\Box\phi = \left(\frac{\partial^2}{\partial\eta^2} - \frac{\partial^2}{\partial\xi^2}\right)\phi = \frac{\partial^2\phi}{\partial u\partial v} = 0, \tag{4.80}$$

for which there exist mode solutions

$$u_k = (4\pi\omega)^{-\frac{1}{2}} e^{ik\xi \pm i\omega\eta} \tag{4.81}$$

$$\omega = |k| > 0, \quad -\infty < k < \infty.$$

The upper sign in (4.81) applies in region L, the lower sign in region R. The presence of the sign change can either be regarded as due to the 'time reversal' in L, or due to the fact that a right-moving wave in R moves towards increasing values of ξ, while in L it moves towards decreasing values of ξ. In any case, the modes (4.81) are those which satisfy the normalization condition (3.29). They are positive frequency with respect to the timelike Killing vector $+\partial_\eta$ in R and $-\partial_\eta$ in L, satisfying

$$\mathscr{L}_{\pm\partial_\eta} u_k = -\mathrm{i}\omega u_k \tag{4.82}$$

(in R and L respectively) in place of (4.77).

The fact that (4.81) has the same functional form as (4.76) is a consequence of the conformal triviality of the system (i.e., Rindler space is conformal to Minkowski space and the wave equation (2.1) is conformally invariant).

Define

$$\left.\begin{aligned} {}^{\mathrm{R}}u_k &= (4\pi\omega)^{-\frac{1}{2}} \mathrm{e}^{\mathrm{i}k\xi - \mathrm{i}\omega\eta}, && \text{in R} \\ &= 0, && \text{in L} \end{aligned}\right\} \tag{4.83}$$

$$\left.\begin{aligned} {}^{\mathrm{L}}u_k &= (4\pi\omega)^{-\frac{1}{2}} \mathrm{e}^{\mathrm{i}k\xi + \mathrm{i}\omega\eta}, && \text{in L} \\ &= 0, && \text{in R.} \end{aligned}\right\} \tag{4.84}$$

The set (4.83) is complete in the Rindler region R, while (4.84) is complete in L, but neither set separately is complete on all of Minkowski space. However, both sets together are so complete, and lines $\eta = \text{constant}$ taken across both R and L are Cauchy surfaces for the whole spacetime. Therefore, the modes (4.83) and (4.84) can also be analytically continued (Boulware 1975*a, b*) into regions F and P (a becomes imaginary). Thus, these Rindler modes are every bit as good a basis for quantizing the ϕ field as the Minkowski space basis (4.76).

The field may be expanded in either set

$$\phi = \sum_{k=-\infty}^{\infty} (a_k \bar{u}_k + a_k^\dagger \bar{u}_k^*) \tag{4.85}$$

(cf. (2.17)), or

$$\phi = \sum_{k=-\infty}^{\infty} (b_k^{(1)}{}^{\mathrm{L}}u_k + b_k^{(1)\dagger}{}^{\mathrm{L}}u_k^* + b_k^{(2)}{}^{\mathrm{R}}u_k + b_k^{(2)\dagger}{}^{\mathrm{R}}u_k^*), \tag{4.86}$$

yielding two alternative Fock spaces, and two vacuum states, $|0_{\mathrm{M}}\rangle$ or $|0_{\mathrm{R}}\rangle$ (the subscripts M and R standing for 'Minkowski' and 'Rindler' respectively) defined by

$$a_k|0_{\mathrm{M}}\rangle = 0 \tag{4.87}$$

or

$$b_k^{(1)}|0_{\mathrm{R}}\rangle = b_k^{(2)}|0_{\mathrm{R}}\rangle = 0. \tag{4.88}$$

The fact that these vacuum states are not equivalent is obvious by inspection of the structure of the Rindler modes (4.81). Because of the sign change in the exponent at $\bar{u} = \bar{v} = 0$ (the 'crossover point' between L and R), the functions ${}^{\mathrm{R}}u_k$ do not go over smoothly to ${}^{\mathrm{L}}u_k$ as one passes from R to L. This means that as one passes from $\bar{u} < 0$ to $\bar{u} > 0$ (or $\bar{v} < 0$ to $\bar{v} > 0$) the right- (or left-) moving modes are non-analytic at this point. In contrast, the positive frequency Minkowski modes (4.78) and (4.79) are analytic not only on the real $\bar{u}$ (or $\bar{v}$) axis, but they are also analytic and bounded in the entire lower half of the complex $\bar{u}$ (or $\bar{v}$) planes. This analyticity property remains true of *any* pure positive frequency functions, i.e., any linear superposition of these positive frequency Minkowski modes. Hence the Rindler modes, by virtue of their non-analyticity at $\bar{u} = \bar{v} = 0$, cannot be a combination of pure positive frequency Minkowski modes, but must also contain negative frequencies. It will be recalled from §3.2 that the mixing of positive and negative frequencies implies that the vacuum states cannot be the same, i.e., the vacuum associated with one set of modes contains particles associated with the other set of modes.

To determine what Rindler particles are present in the Minkowski vacuum, we must determine the Bogolubov transformation between the two sets of modes. This may be achieved using (3.36), which are essentially Fourier transforms of the Rindler modes. An alternative, more elegant, method due to Unruh (1976), is to note that although ${}^{\mathrm{L}}u_k$ and ${}^{\mathrm{R}}u_k$ are non-analytic, the two (un-normalized) combinations

$$ {}^{\mathrm{R}}u_k + \mathrm{e}^{-\pi\omega/a}\,{}^{\mathrm{L}}u^*_{-k} \tag{4.89} $$

and

$$ {}^{\mathrm{R}}u^*_{-k} + \mathrm{e}^{\pi\omega/a}\,{}^{\mathrm{L}}u_k \tag{4.90} $$

are analytic and bounded, both for all real $\bar{u}$, $\bar{v}$, and everywhere in the lower half complex $\bar{u}$ and $\bar{v}$ planes.

To see this, one can write (4.89) and (4.90) in Minkowski coordinates, where they are proportional to

$$ \begin{array}{ll} \bar{u}^{\mathrm{i}\omega/a}, & k > 0 \\ \bar{v}^{-\mathrm{i}\omega/a}, & k < 0 \end{array} \tag{4.91} $$

and

$$ \begin{array}{ll} \bar{v}^{\mathrm{i}\omega/a}, & k > 0 \\ \bar{u}^{-\mathrm{i}\omega/a}, & k < 0 \end{array} \tag{4.92} $$

($\omega = |k|$) respectively, for all $\bar{u}$, $\bar{v}$ in the range $-\infty$ to ∞ (i.e., in both L and

R); they are clearly analytic across $\bar{u}=\bar{v}=0$. They are also analytic in the lower-half complex $\bar{u}$ and $\bar{v}$ planes if the branch cut of the complex powers in (4.91) and (4.92) is taken to lie in the upper-half plane (i.e., $\ln(-1)=-i\pi$, which fixes the signs of the exponents $e^{\pm\pi\omega/a}$ in (4.89) and (4.90)).

Because the modes (4.89) and (4.90) share the positive frequency analyticity properties of the Minkowski modes $\bar{u}_k$, they must also share a common vacuum state $|0_M\rangle$ (see remark on page 47). Thus, instead of (4.85) we can expand ϕ in terms of (4.89) and (4.90) as

$$\phi=\sum_{k=-\infty}^{\infty}[2\sinh(\pi\omega/a)]^{-\frac{1}{2}}[d_k^{(1)}(e^{\pi\omega/2a}\,{}^{R}u_k+e^{-\pi\omega/2a}\,{}^{L}u_{-k}^{*})$$

$$+d_k^{(2)}(e^{-\pi\omega/2a}\,{}^{R}u_{-k}^{*}+e^{\pi\omega/2a}\,{}^{L}u_k)]+\text{h.c.} \tag{4.93}$$

where now

$$d_k^{(1)}|0_M\rangle=d_k^{(2)}|0_M\rangle=0, \tag{4.94}$$

and we have also introduced a normalization factor. The operators $b_k^{(1,2)}$ can be related to $d_k^{(1,2)}$ by taking the inner products $(\phi,{}^{R}u_k)$, $(\phi,{}^{L}u_k)$, first with ϕ given by (4.86) and then by (4.93). One obtains

$$b_k^{(1)}=[2\sinh(\pi\omega/a)]^{-\frac{1}{2}}[e^{\pi\omega/2a}d_k^{(2)}+e^{-\pi\omega/2a}d_{-k}^{(1)\dagger}] \tag{4.95}$$

$$b_k^{(2)}=[2\sinh(\pi\omega/a)]^{-\frac{1}{2}}[e^{\pi\omega/2a}d_k^{(1)}+e^{-\pi\omega/2a}d_{-k}^{(2)\dagger}]. \tag{4.96}$$

These Bogolubov transformations provide the required relation between the states $|0_R\rangle$ and $|0_M\rangle$.

Now consider an accelerating Rindler observer at $\xi=$ constant. From (4.74) one sees that such an observer's proper time is proportional to η. One therefore expects that the vacuum for Rindler observers will be $|0_R\rangle$ as this is the state associated with modes which are positive frequency with respect to η. Thus, according to (4.86), a Rindler observer in L (respectively R) will detect particles counted by the number operator $b_k^{(1)\dagger}b_k^{(1)}$ (respectively $b_k^{(2)\dagger}b_k^{(2)}$). If the field is in the state $|0_M\rangle$ (that is, it is devoid of usual Minkowski space particles) then, using (4.95) and (4.96) it may be deduced that a Rindler observer will detect

$$\langle 0_M|b_k^{(1,2)\dagger}b_k^{(1,2)}|0_M\rangle=e^{-\pi\omega/a}/[2\sinh(\pi\omega/a)]=(e^{2\pi\omega/a}-1)^{-1} \tag{4.97}$$

particles in mode k. This is precisely the Planck spectrum for radiation at temperature $T_0=a/2\pi k_B$. The temperature T as seen by the accelerated observer is given by the Tolman relation

$$T=(g_{00})^{-\frac{1}{2}}T_0 \tag{4.98}$$

(see, for example, Tolman 1934, Landau & Lifshitz, 1958, §27, or Balazs 1958). Here g_{00} is obtained from the metric (4.71), in terms of which the accelerated observer has constant spatial coordinate. Using (4.73) and (4.98) one obtains $T = 1/(2\pi\alpha k_{\mathrm{B}})$, which is in exact agreement with (3.68) obtained for an accelerated detector in Minkowski space. The particles detected by such an observer are often called *Rindler particles*. The quantum field theory of accelerated observers and associated thermal effects have been considered in great generality by Sanchez (1979). Possible physical effects have been examined by Barshay & Troost (1978) and Hosoya (1979).

Here we predict thermal particles seen by an accelerating observer in flat space, but under a conformal transformation we could obtain a thermal bath seen by an inertial observer in curved space. This is the so-called Hawking effect (see chapter 8).

5
Curved spacetime examples

This chapter is devoted to a direct application of the curved spacetime quantum field theory developed in chapter 3. We treat particle creation by time-dependent gravitational fields by examing a variety of expanding and contracting cosmological models. Most of the models are special cases of the Robertson–Walker homogeneous isotropic spacetimes, chosen either for their simplicity, or special interest in illuminating certain aspects of the formalism.

All the main cases that have appeared in the literature are collected here. The Milne universe (technically flat spacetime) and de Sitter space are especially useful for illustrating the role of adiabaticity in assessing the physical reasonableness of a quantum state. De Sitter space also enjoys the advantage of being the only time-dependent cosmological model for which both the particle creation effects and the vacuum stress (deferred until §6.4) have been explicitly evaluated by all known techniques.

A small but important section, §5.5, presents a classification scheme that relates the vacuum states in conformally-related spacetimes. This topic too has a 'thermal' aspect to it. It will turn out to be of relevance for the computation of $\langle T_{\mu\nu} \rangle$ in Robertson–Walker spacetimes in chapter 6 and chapter 7.

The final section is an attempt to go beyond the simple Robertson–Walker models and treat the subject of anisotropy in cosmology. This is an issue of central importance in modern cosmological theory, because the observed high degree of isotropy in the universe is without adequate explanation. Here, once again, our theory makes contact with the real world, albeit in a rather model-dependent and modest way. It seems probable that quantum particle creation played a large part in determining the constituents and condition of the primeval cosmological fluid. Moreover, the back-reaction of quantum effects on the gravitational dynamics, to be discussed in chapters 6 and 7, certainly played a role in the dissipation of primeval anisotropy. The resulting entropy, in the form of quantum particles, remains with us today as an observational constraint on the quantum theory of the very early universe.

5.1 Robertson–Walker spacetimes

Most of the examples to be considered in this chapter will involve fields propagating in particular Robertson–Walker spacetimes. Such spacetimes are especially important in providing cosmological models which are in good agreement with observation. Robertson–Walker spacetimes with flat spatial sections were discussed in §§3.4–3.6. We start by summarizing the essential formulae, and then extend them to Robertson–Walker spacetimes with hyperbolic and spherical spatial sections. Attention is restricted to scalar fields, the extension to fields of higher spins following along the lines of §3.8.

The line element for n-dimensional Robertson–Walker spacetimes with flat spatial sections is

$$ds^2 = C(\eta)\left[d\eta^2 - \sum_{i=1}^{n-1} (dx^i)^2\right]. \tag{5.1}$$

The mode decomposition (3.30) for the field ϕ is

$$\phi(x) = \int d^{n-1}k[a_{\mathbf{k}}u_{\mathbf{k}}(x) + a_{\mathbf{k}}^{\dagger}u_{\mathbf{k}}^{*}(x)]. \tag{5.2}$$

The modes $u_{\mathbf{k}}$ can be written in a separated form as

$$u_{\mathbf{k}}(x) = (2\pi)^{(1-n)/2}e^{i\mathbf{k}\cdot\mathbf{x}}C^{(2-n)/4}(\eta)\chi_k(\eta) \tag{5.3}$$

where $k = |\mathbf{k}| = (\sum_{i=1}^{n-1} k_i^2)^{\frac{1}{2}}$ and χ_k satisfies

$$\frac{d^2\chi_k}{d\eta^2} + \{k^2 + C(\eta)[m^2 + (\xi - \xi(n))R(\eta)]\}\chi_k = 0 \tag{5.4}$$

with

$$\xi(n) \equiv \tfrac{1}{4}[(n-2)/(n-1)]. \tag{5.5}$$

The normalization condition (3.29) reduces to a condition on the Wronskian of the solutions χ_k:

$$\chi_k\partial_\eta\chi_k^* - \chi_k^*\partial_\eta\chi_k = i. \tag{5.6}$$

These equations can easily be extended to Robertson–Walker spacetimes with curved spatial sections. For simplicity we confine our attention to four-dimensional spacetimes; equations which allow the generalization to arbitrary dimensions can be found in Bander & Itzykson (1966). The general Robertson–Walker line element is

$$ds^2 = dt^2 - a^2(t) \sum_{i,j=1}^{3} h_{ij} dx^i dx^j \tag{5.7}$$

where

$$\sum_{i,j=1}^{3} h_{ij} dx^i dx^j = (1 - Kr^2)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta \, d\phi^2)$$
$$= d\chi^2 + f^2(\chi)(d\theta^2 + \sin^2\theta \, d\phi^2) \tag{5.8}$$

and

$$f(\chi) = r = \begin{cases} \sin\chi, & 0 \le \chi \le 2\pi, \quad K = +1 \\ \chi, & 0 \le \chi < \infty, \quad K = 0 \\ \sinh\chi, & 0 \le \chi < \infty, \quad K = -1. \end{cases} \tag{5.9}$$

Equation (5.8) gives the line element on the spatial sections, which are hyperbolic, flat or closed depending on whether $K = -1, 0, 1$ respectively. Writing $C(\eta) = a^2(t)$, with the conformal time parameter η given by

$$\eta = \int^t a^{-1}(t') dt', \tag{5.10}$$

the line element (5.7) can be recast in the form

$$ds^2 = C(\eta)\left(d\eta^2 - \sum_{i,j=1}^{3} h_{ij} dx^i dx^j \right). \tag{5.11}$$

Defining $\Upsilon = (1 - Kr^2)^{-1}$ and $D = \dot{C}/C$, where a dot denotes differentiation with respect to η, the nonzero Christoffel symbols for the metric (5.11) are (indices 0, 1, 2, 3 corresponding to η, r, θ, ϕ respectively)

$$\Gamma^0_{00} = \Gamma^1_{01} = \Gamma^2_{02} = \Gamma^3_{03} = \tfrac{1}{2}D, \quad \Gamma^0_{11} = \tfrac{1}{2}D\Upsilon, \quad \Gamma^0_{22} = \tfrac{1}{2}Dr^2,$$
$$\Gamma^0_{33} = \tfrac{1}{2}Dr^2 \sin^2\theta, \quad \Gamma^1_{11} = Kr\Upsilon \quad \Gamma^1_{22} = -r\Upsilon^{-1},$$
$$\Gamma^1_{33} = -r\Upsilon^{-1}\sin^2\theta, \quad \Gamma^2_{12} = \Gamma^3_{13} = 1/r,$$
$$\Gamma^2_{33} = -\sin\theta\cos\theta, \quad \Gamma^3_{23} = \cot\theta. \tag{5.12}$$

From (5.12) the nonzero components of the Ricci tensor, and hence the Ricci scalar, can be computed:

$$R_{00} = \tfrac{3}{2}\dot{D}, \quad R_{11} = -\tfrac{1}{2}(\dot{D} + D^2)\Upsilon - 2K\Upsilon,$$
$$R_{22} = \Upsilon^{-1} r^2 R_{11}, \quad R_{33} = \sin^2\theta \, R_{22}, \tag{5.13}$$

$$R = C^{-1}[3\dot{D} + \tfrac{3}{2}D^2 + 6K]. \tag{5.14}$$

The scalar field ϕ satisfies (3.26) with R given by (5.14). Its mode decomposition can now be written as

$$\phi(x) = \int \mathrm{d}\tilde{\mu}(k)[a_{\mathbf{k}} u_{\mathbf{k}}(x) + a_{\mathbf{k}}^\dagger u_{\mathbf{k}}^*(x)] \tag{5.15}$$

where the measure $\tilde{\mu}(k)$ will be defined shortly. The modes are separated as

$$u_{\mathbf{k}}(x) = C^{-\frac{1}{2}}(\eta)\mathscr{Y}_{\mathbf{k}}(\mathbf{x})\chi_k(\eta) \tag{5.16}$$

with $\mathbf{x} = (r, \theta, \phi)$ or (χ, θ, ϕ) and $\mathscr{Y}_{\mathbf{k}}(\mathbf{x})$ a solution of

$$\Delta^{(3)}\mathscr{Y}_{\mathbf{k}}(\mathbf{x}) = -(k^2 - K)\mathscr{Y}_{\mathbf{k}}(\mathbf{x}). \tag{5.17}$$

In (5.17) $\Delta^{(3)}$ is the Laplacian associated with the spatial metric h_{ij}:

$$\Delta^{(3)}\mathscr{Y}_{\mathbf{k}} \equiv h^{-\frac{1}{2}}\partial_i\left(h^{\frac{1}{2}} h^{ij}\partial_j \mathscr{Y}_{\mathbf{k}}\right), \tag{5.18}$$

$h = \det(h_{ij})$. With this separation χ_k satisfies (5.4) with $n = 4$, $(\xi(4) = \frac{1}{6})$. Further, if the functions $\mathscr{Y}_{\mathbf{k}}$ are normalized such that

$$\int \mathrm{d}^3x\, h^{\frac{1}{2}}\mathscr{Y}_{\mathbf{k}}(\mathbf{x})\mathscr{Y}^*_{\mathbf{k}'}(\mathbf{x}) = \delta(\mathbf{k}, \mathbf{k}') \tag{5.19}$$

where $\delta(\mathbf{k}, \mathbf{k}')$ is the δ-function with respect to the measure $\tilde{\mu}$:

$$\int \mathrm{d}\tilde{\mu}(k') f(\mathbf{k}')\delta(\mathbf{k}, \mathbf{k}') = f(\mathbf{k}), \tag{5.20}$$

then the normalization condition (3.29) once again reduces to (5.6).

The eigenfunctions $\mathscr{Y}_{\mathbf{k}}$ of the three-dimensional Laplacian are (Parker & Fulling 1974)

$$\mathscr{Y}_{\mathbf{k}}(\mathbf{x}) = \left\{\begin{array}{lll} (2\pi)^{-\frac{3}{2}}\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}, & \mathbf{k} = (k_1, k_2, k_3), & (K = 0) \\ \Pi^{(\pm)}_{kJ}(\chi)Y^M_J(\theta, \phi), & k = (k, J, M), & (K = \pm 1) \end{array}\right\} \tag{5.21}$$

where

$$-\infty < k_i < \infty; \quad k = |\mathbf{k}|, \quad (K = 0)$$

$$M = -J, -J+1, \ldots, J; \left\{\begin{array}{lll} J = 0, 1, \ldots, k-1; & k = 1, 2, \ldots, & (K = 1) \\ J = 0, 1, \ldots; & 0 < k < \infty, & (K = -1) \end{array}\right\} \tag{5.22}$$

The Y_J^M are spherical harmonics. The functions $\Pi^{(-)}$ are defined by (see, for example, Bander & Itzykson (1966), or Dolginov & Toptygin (1959)):

$$\Pi_{kJ}^{(-)}(\chi) = [\tfrac{1}{2}\pi k^2(k^2+1)\dots(k^2+J^2)]^{-\frac{1}{2}} \sinh^J \chi \left(\frac{\mathrm{d}}{\mathrm{d}\cosh\chi}\right)^{1+J} \cos k\chi. \tag{5.23}$$

One can obtain the functions $\Pi_{kJ}^{(+)}(\chi)$ from $\Pi_{kJ}^{(-)}(\chi)$ by replacing k by $-ik$ and χ by $-i\chi$ in the latter (see, for example, Lifshitz & Khalatnikov (1963)).

With these definitions, the measure $\tilde{\mu}(k)$ is defined by

$$\int \mathrm{d}\tilde{\mu}(k) = \begin{cases} \int \mathrm{d}^3k, & (K=0) \\ \sum\limits_{k,J,M}, & (K=1) \\ \int_0^\infty \mathrm{d}k \sum\limits_{J,M}, & (K=-1) \end{cases} \tag{5.24}$$

Finally, we note that it is also possible to write the spatial part of the mode decomposition in terms of Gegenbauer polynomials rather than the $\Pi^{(+)}$ functions (see, for example, Ford 1976).

5.2 Static Robertson–Walker spacetimes

The simplest cases in which the preceding analysis can be applied are the static spacetimes with $C(\eta) = c = a^2 = \text{constant}$. The static spacetime with flat spatial sections ($K = 0$) is, of course, Minkowski space, which we shall not consider again. Rather, we discuss the closed ($K = 1$) Einstein universe (Einstein 1917) and the static spacetime with hyperbolic spatial sections ($K = -1$).

In these cases the scalar curvature (5.14) reduces to

$$R = 6K/c, \tag{5.25}$$

which permits normalized solutions of (5.4) in four dimensions to be written down immediately:

$$\chi_k(\eta) = (2\omega_k)^{-\frac{1}{2}} e^{-i\omega_k\eta} \tag{5.26}$$

where

$$\begin{aligned} \omega_k^2 &= k^2 + \mu^2 \\ &= k^2 + cm^2 + (6\xi - 1)K. \end{aligned} \tag{5.27}$$

Note from (5.10) that $\eta = t/a$, so that the solutions (5.26) are positive frequency with respect to the Killing vectors ∂_η and ∂_t. Since the spacetimes are static and admit these global, timelike Killing vectors, the definition of particles here is no more ambiguous than the definition of particles in Minkowski space. In particular, using (5.16) and (5.26) to construct positive frequency modes

$$u_{\mathbf{k}}(x) = (2c\omega_k)^{-\frac{1}{2}}\mathscr{Y}_{\mathbf{k}}(\mathbf{x})e^{-i\omega_k\eta} \tag{5.28}$$

for the expansion (5.15), a vacuum state $|0\rangle$ is defined by

$$a_{\mathbf{k}}|0\rangle = 0. \tag{5.29}$$

One expects this vacuum to be as physically reasonable as the standard Minkowski vacuum, a conjecture which can be verified by an investigation of particle detectors. The Wightman function in the Einstein universe has been calculated by Dowker (1971) (see also Critchley 1976, Dowker & Critchley 1977*a* and Bunch & Davies 1977*a*) who finds

$$G^+(x,x') = \frac{i\mu}{8\pi c\sin(\Delta\chi)} \times \sum_{n=-\infty}^{\infty} \frac{(\Delta\chi + 2\pi n)H_1^{(2)}\{\mu[(\Delta\eta - i\varepsilon)^2 - (\Delta\chi + 2n\pi)^2]^{\frac{1}{2}}\}}{[(\Delta\eta - i\varepsilon)^2 - (\Delta\chi + 2\pi n)^2]^{\frac{1}{2}}}, \quad (K = +1) \tag{5.30}$$

where H is a Hankel function. We have defined as usual

$$\Delta\chi = \chi - \chi', \quad \Delta\eta = \eta - \eta' \tag{5.31}$$

and, without loss of generality, have used the isotropy of the spacetime to orient the axes such that $\phi = \phi'$, $\theta = \theta'$. The form of the Wightman function is that of a sum of Minkowski space positive frequency Wightman functions (cf. (2.77)). Using this fact it is not difficult to show using (3.55) that a particle detector at rest in the (χ, η) frame registers no particles.

We note for later use that in the massless, conformally coupled limit ($m \to 0$, $\xi \to \frac{1}{6}$, hence $\mu \to 0$) (5.30) reduces to

$$D^+(x,x') = \frac{1}{8\pi^2 c[\cos(\Delta\eta - i\varepsilon) - \cos(\Delta\chi)]}, \quad (K = +1). \tag{5.32}$$

The massless, conformally coupled Wightman function for a scalar field in a static, hyperbolic spacetime has been calculated by Bunch (1978*a*), who finds

$$D^+(x,x') = \frac{\Delta\chi}{4\pi^2 c\sinh(\Delta\chi)[\Delta\chi^2 - (\Delta\eta - i\varepsilon)^2]}, \quad (K = -1) \tag{5.33}$$

while for a massive field and arbitrary ξ one obtains

$$G^+(x,x')$$

$$= -\frac{\mathrm{i}\mu}{8\pi c\sinh(\Delta\chi)}\frac{\Delta\chi H_1^{(2)}\{\mu[(\Delta\eta-\mathrm{i}\varepsilon)^2-(\Delta\chi)^2]^{\frac{1}{2}}\}}{[(\Delta\eta-\mathrm{i}\varepsilon)^2-(\Delta\chi)^2]^{\frac{1}{2}}}, \quad (K=-1). \qquad (5.34)$$

Once again the similarity to the Minkowski space Wightman function guarantees that a detector at rest in the (χ,η) frame will register no particles.

If the spacetime is not static then in general solutions to (5.4) are very difficult to find and it is only in special cases or using approximation methods that progress can be made. However, for a massless, conformally coupled field, the conformal invariance of the field equation immediately allows one to write down the modes; they are given by (5.28) with c replaced by $C(\eta)$. This is analogous to the spatially flat case discussed in §3.7, and the modes so obtained are, as in §3.7, positive frequency with respect to the conformal Killing vector ∂_η. From these modes the Green functions in the conformal vacuum are readily obtained. For example, for $K=\pm 1$, G^+ in the conformal vacuum is obtained by replacing c by $C^{\frac{1}{2}}(\eta)C^{\frac{1}{2}}(\eta')$ in (5.32) and (5.33) respectively.

5.3 The Milne universe

The case of Robertson–Walker spacetimes with $a(t)=t$ is particularly interesting, not only because the conformally non-trivial $(m\neq 0)$ field equation can be solved exactly, but because it sheds light upon the nature of the relationship between the conformal and the adiabatic vacua (Fulling, Parker & Hu, 1974, Bunch 1977, 1978*a*, Davies & Fulling 1977*b*, Bunch, Christensen & Fulling 1978).

The four-dimensional hyperbolic spacetime with $a(t)=t$ (known as the Milne universe; Milne 1932) and its two-dimensional counterpart, are in fact merely unconventional coordinatizations of flat spacetime, analogous to the Rindler system (see §4.5). Nevertheless, as in the Rindler case, the Milne model leads to non-trivial quantum effects, and can also cast light on the uniformly expanding cosmological models with $K=0$ and $+1$ (which are not merely flat spacetime in disguise).

The two-dimensional Milne universe has line element

$$\begin{aligned} \mathrm{d}s^2 &= \mathrm{d}t^2 - a^2t^2\mathrm{d}x^2 \\ &= \mathrm{e}^{2a\eta}(\mathrm{d}\eta^2-\mathrm{d}x^2) \end{aligned} \qquad (5.35)$$

where $|t| = a^{-1}e^{a\eta}$ and a is a constant. Under the coordinate transformation

$$y^0 = a^{-1}e^{a\eta}\cosh ax, \quad y^1 = a^{-1}e^{a\eta}\sinh ax \tag{5.36}$$

(5.35) reduces to

$$ds^2 = (dy^0)^2 - (dy^1)^2, \quad 0 < y^0 < \infty \quad -\infty < y^1 < \infty \tag{5.37}$$

(see fig. 15).

The massive wave equation with $\xi = 0$ and $C = e^{2a\eta}$ is readily solved in terms of either Bessel or Hankel functions, from which two complete sets of normalized modes $\{\bar{u}_{\mathbf{k}}, \bar{u}_{\mathbf{k}}^*\}$ and $\{u_{\mathbf{k}}, u_{\mathbf{k}}^*\}$ can be constructed using (5.3) with

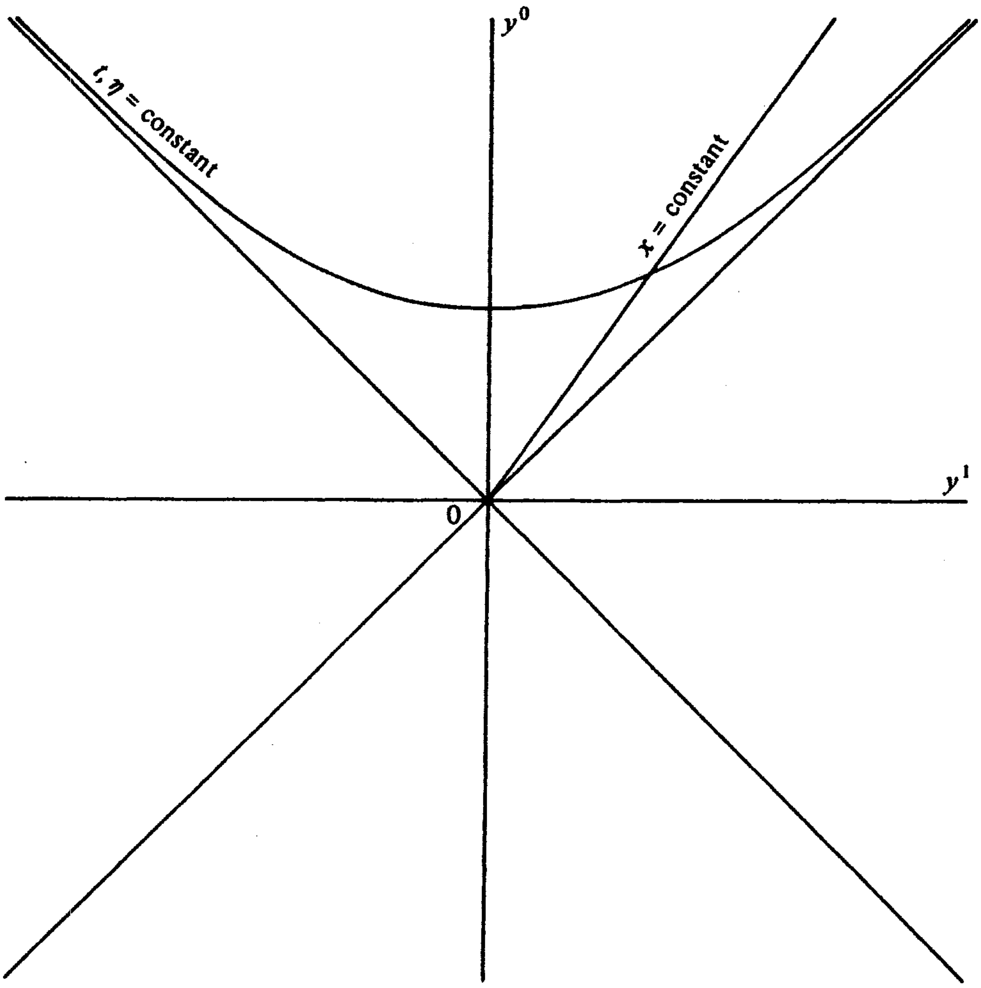


Fig. 15. Milne universe. The t, x coordinates cover the wedge of Minkowski space corresponding to region F in fig. 13. This region is therefore like the Rindler wedge on its side. The lines x = constant radiating from the origin represent the world lines of observers who perceive a universe expanding from a big bang origin at 0.

$$\bar{\chi}_k = [(2a/\pi)\sinh(\pi k/a)]^{-\frac{1}{2}} J_{-ik/a}(mt) \tag{5.38}$$

$$\chi_k = \tfrac{1}{2}(\pi/a)^{\frac{1}{2}} e^{\pi k/2a} H^{(2)}_{ik/a}(mt), \quad (k > 0). \tag{5.39}$$

These two complete sets of modes are related by a Bogolubov transformation:

$$\bar{\chi}_k = \alpha_k \chi_k + \beta_k \chi_k^* \tag{5.40}$$

with

$$\alpha_k = [e^{\pi k/a}/2\sinh(\pi k/a)]^{\frac{1}{2}}, \quad \beta_k = [e^{-\pi k/a}/2\sinh(\pi k/a)]^{\frac{1}{2}}, \tag{5.41}$$

from which the similarity with the Rindler case is manifest (compare (5.41) with (4.96)).

It is clear from (5.40) that the state $|\bar{0}\rangle$ defined by (3.32) and (3.33) with respect to the modes $\bar{u}_{\mathbf{k}}$ will be inequivalent to the state $|0\rangle$ defined with respect to the modes $u_{\mathbf{k}}$. This leads us to ask once again, whether either vacuum state is in any way privileged.

We can gain some insight into this question by taking the massless limit of (5.38):

$$\bar{\chi}_k \xrightarrow[m \to 0]{} e^{i\theta} e^{-ik\eta}/(2k)^{\frac{1}{2}}, \tag{5.42}$$

where θ is an η-independent phase. We see that in the massless limit $\bar{\chi}_k$ is positive frequency with respect to the conformal time η, so that $|\bar{0}\rangle$ can, in this limit, be identified with the conformal vacuum (see §3.7).

Using an integral representation for the Hankel function in (5.39), one can write (Fulling, Parker & Hu 1974)

$$u_{\mathbf{k}}(x) = (8\pi^2 a)^{-\frac{1}{2}} \int_{-\infty}^{\infty} d\rho \, e^{-i\omega(\rho)y^0} e^{ip(\rho)y^1} e^{-ik\rho/a} \tag{5.43}$$

where y^0 and y^1 are given by (5.36), and

$$p(\rho) = -m\sinh\rho, \quad \omega(\rho) = (p^2 + m^2)^{\frac{1}{2}}. \tag{5.44}$$

We see that $u_{\mathbf{k}}$ is a superposition of modes which are positive frequency with respect to Minkowski time y^0 (Sommerfield 1974, di Sessa 1974). Thus, if the spacetime (5.35) were embedded in two-dimensional Minkowski space, the vacuum $|0\rangle$ defined with respect to the modes $u_{\mathbf{k}}$ would be the usual Minkowski vacuum. Further, as a comoving observer has proper time $t \propto y^0$, such an observer will see no particles in this vacuum.

Another useful fact about the modes $u_{\mathbf{k}}$ can be obtained by comparing them with the positive frequency adiabatic modes based on (3.102) (which, for conformal coupling, also holds in two dimensions). To zeroth order

$$W_k^{(0)} = \omega_k = (k^2 + m^2 e^{2a\eta})^{\frac{1}{2}},$$

which when substituted into (3.105) yields a factor

$$T^2 \omega_k^2(\eta_1) = T^2 k^2 + T^2 m^2 e^{2a\eta_1 T}.$$

Clearly ω_k becomes large compared with the derivatives of $\dot{C}/C = O(T^{-1})$ (for fixed $\eta_1 = \eta/T$) as k, m or η become large either individually or together.

In the limit of large η

$$W_k \to \omega_k \simeq m e^{a\eta}$$

which when substituted in (3.102) gives

$$\chi_k^{(A)} \to (2me^{a\eta})^{-\frac{1}{2}} \exp(-ime^{a\eta}/a), \quad \eta \to \infty. \tag{5.45}$$

(A indicates that this limit holds to any adiabatic order. All higher order corrections vanish as $\eta \to \infty$.) On the other hand, using the asymptotic expansion of Hankel functions, one can easily show that (up to a constant phase factor) (5.39) reduces to (5.45) in the limit of large η (i.e., large $|t|$). One can also verify that in the limit of large k, or large k and m, (5.39) reduces to the corresponding limit of (3.102).

Thus the modes $u_{\mathbf{k}}$ are positive frequency with respect to the adiabatic definition given in §3.5, and the vacuum $|0\rangle$ is an adiabatic vacuum of infinite adiabatic order. So in this case the adiabatic definition of positive frequency agrees with the definition inherited from Minkowski space in a natural way. Further, not only does an inertial particle detector register no particles to any finite adiabatic order in this vacuum, it registers strictly no particles at all.

Since an inertial detector registers no particles in the adiabatic vacuum $|0\rangle$, it is clear from (5.40) that it must detect particles in the vacuum $|\bar{0}\rangle$. Confining attention to the massless case, this implies that a comoving detector will register particles in the Milne conformal vacuum as indeed it would in almost any other conformal vacuum (see §3.7).

The presence of particles in the conformal vacuum can be seen in another way: by calculating the conformal expectation value of the stress-tensor, $\langle \bar{0}|T_{\mu\nu}|\bar{0}\rangle$. To do this rigorously (Davies & Fulling 1977*b*) one must use the regularization and renormalization techniques described in the next chapter, and applied to this example in §7.1. However, even without regularization and renormalization, it is possible to calculate the finite difference between the expectation values of $T_{\mu\nu}$ in the $|0\rangle$ and $|\bar{0}\rangle$ states respectively. In the massless limit one finds (Bunch 1977, Bunch, Christensen & Fulling 1978)

$$\langle\bar{0}|T_{tt}|\bar{0}\rangle - \langle 0|T_{tt}|0\rangle = -(1/\pi a^2 t^2)\int_0^\infty k|\beta_k|^2 \mathrm{d}k$$

$$= -(1/\pi)\int_0^\infty q(\mathrm{e}^{2\pi q t} - 1)^{-1}\mathrm{d}q = -1/24\pi t^2, \qquad (5.46)$$

where we have used (5.41).

Next one can argue that because $|0\rangle$ is equivalent to the usual Minkowski space vacuum, the renormalized (or normal ordered) value of $\langle 0|T_{tt}|0\rangle$ must be zero, so that (5.46) gives the renormalized value of $\langle\bar{0}|T_{tt}|\bar{0}\rangle$. (These ideas will be made more precise in the next chapter.) Thus the conformal vacuum contains a negative energy density of radiation with a Planck spectrum at temperature $(2\pi t)^{-1}$. The relation between the two vacuum states is therefore closely similar to that between the Minkowski and Rindler definitions, even to the extent of the thermal association.

This explicit example nicely illustrates the physical differences between the adiabatic and conformal vacuum states. The difference is perhaps not surprising when it is remembered that the former is constructed from an analysis of the field behaviour at large mass, whereas the latter exploits the conformal symmetry of the massless case.

The situation is similar in the four-dimensional ($K = -1$) Milne universe. Expressions (5.38) and (5.39) still provide solutions of the field equation in the conformally coupled ($\xi = \frac{1}{6}$) case. The states $|0\rangle$ and $|\bar{0}\rangle$ are still adiabatic and (in the massless limit) conformal vacuum states, respectively. In fact, similar results hold even in the $K = 0$ and $+1$ models with $a(t) = t$, even though one can no longer appeal to the Minkowskian covering space to give special significance to $|0\rangle$.

The $K = 0$ model serves to illustrate a curious phenomenon concerning the massless limit of the adiabatic vacuum. Equation (5.4) reduces in this case to

$$\frac{\mathrm{d}^2\chi_k}{\mathrm{d}\eta^2} + [k^2 + (6\xi - 1)a^2]\chi_k = 0. \qquad (5.47)$$

The curvature term, $(6\xi - 1)a^2$, enters in the same way as would a mass. However, for $\xi < \frac{1}{6}$ it is negative (corresponding to a tachyonic mass), a state of affairs that is usually taken to imply that the vacuum state is unstable. If a self-interaction term such as $\lambda\phi^4$ is added to the field Lagrangian, spontaneous symmetry breaking can occur (Goldstone 1961). This feature also arises in other cosmological models (see §5.4) and has been exploited in the work of Frolov, Grib & Mostepanenko (1977, 1978).

The fact that particle states become ill-defined for $\xi < \frac{1}{6}$ in the above example may also be deduced by examining the adiabatically constructed massive modes in the $K = 0$ model:

$$\chi_k = \tfrac{1}{2}(\pi/a)^{\frac{1}{2}} \mathrm{e}^{-\mathrm{i}\pi\nu/2} H_\nu^{(2)}(mt) \tag{5.48}$$

where

$$\nu = [1 - 6\xi - (k/a)^2]^{\frac{1}{2}}.$$

The solutions (5.48), which reduce to (5.45) for large $|t|$, can be used to construct an adiabatic vacuum in the usual way. However, in the limit of small mass χ_k reduces to

$$\tfrac{1}{2}\mathrm{i}(\pi/a)^{\frac{1}{2}} \operatorname{cosec}(\pi\nu)\left(\frac{\mathrm{e}^{-\mathrm{i}\pi\nu/2}(\frac{1}{2}ma^{-1})^{-\nu}\mathrm{e}^{-\nu\eta}}{\Gamma(1-\nu)} - \frac{\mathrm{e}^{\mathrm{i}\pi\nu/2}(\frac{1}{2}ma^{-1})^{\nu}\mathrm{e}^{\nu\eta}}{\Gamma(1+\nu)}\right)$$

which diverges as $m \to 0$ if $\xi < \frac{1}{6}$, $k^2/a^2 < 1 - 6\xi$.

Particle creation in this model has been considered in detail by Chitre & Hartle (1977) using analytic continuation methods (see §8.5).

5.4 De Sitter space

De Sitter space (de Sitter 1917*a*, *b*) is the curved spacetime which has been most studied by quantum field theorists (see bibliography and references cited below). The reason for this special attention stems from the fact that de Sitter space is the unique maximally symmetric curved spacetime (see, for example, Weinberg 1972). It enjoys the same degree of symmetry as Minkowski space (ten Killing vectors), which greatly facilitates technical computations as far as quantum field theory is concerned. Even so, the presence of curvature, and non-trivial global properties, introduce new aspects to the quantization of fields in de Sitter space. We confine our attention in this section to scalar fields, although higher-spin fields have been treated in many of the papers cited; in particular, fields of arbitrary spin on de Sitter space have been discussed by Dowker & Critchley (1976*a*, *b*), Grensing (1977) and Birrell (1979*b*). See also §6.4.

Four-dimensional de Sitter space is most easily represented as the hyperboloid

$$z_0{}^2 - z_1{}^2 - z_2{}^2 - z_3{}^2 - z_4{}^2 = -\alpha^2 \tag{5.49}$$

embedded in five-dimensional Minkowski space with metric

$$\mathrm{d}s^2 = \mathrm{d}z_0{}^2 - \mathrm{d}z_1{}^2 - \mathrm{d}z_2{}^2 - \mathrm{d}z_3{}^2 - \mathrm{d}z_4{}^2 \tag{5.50}$$

(see, for example, Schrödinger 1956, Hawking & Ellis 1973 and, in particular, Fulling 1972, where the connection between the geometry of de Sitter space and quantum field theory is discussed).

From the form of (5.49) it is clear that the symmetry group of de Sitter space is the ten parameter group $SO(1,4)$ of homogeneous 'Lorentz transformations' in the five-dimensional embedding space (see for example, Gürsey 1964) known as the de Sitter group. Just as the Poincaré group plays a central role in the quantization of fields in Minkowski space, so the de Sitter group of symmetries on de Sitter space is fundamental to the discussion of quantization to be given here.

As usual, the first step in scalar field quantization is the solution of the field equation (3.26). To accomplish this, a coordinatization of de Sitter space must first be specified. From the discussions of the Rindler and Milne spacetimes already given it should be clear that the choice of a particular coordinate system can lead to what appears as an especially natural choice (or choices) of 'vacuum' state. For example, in the Rindler case, the (Rindler) vacuum defined by positive frequency modes (4.81) is the most natural for the coordinatization (4.71), while for the coordinatization (4.66), the Minkowski vacuum is the more natural choice. We are therefore led to consider in turn the three most widely used coordinatizations of the de Sitter space hyperboloid and to discuss the properties of the resulting solutions of the field equation.

Consider first the coordinates $(t, \mathbf{x})$ defined by

$$\left.\begin{aligned} z_0 &= \alpha \sinh(t/\alpha) + \tfrac{1}{2}\alpha^{-1} e^{t/\alpha}|\mathbf{x}|^2 \\ z_4 &= \alpha \cosh(t/\alpha) - \tfrac{1}{2}\alpha^{-1} e^{t/\alpha}|\mathbf{x}|^2 \\ z_i &= e^{t/\alpha} x_i, \quad i = 1, 2, 3, \quad -\infty < t, x_i < \infty. \end{aligned}\right\} \tag{5.51}$$

which cover the half of the de Sitter manifold with $z_0 + z_4 > 0$ (see fig. 16). In these coordinates the line element (5.50) becomes

$$ds^2 = dt^2 - e^{2t/\alpha} \sum_{i=1}^{3} (dx^i)^2, \tag{5.52}$$

which is the line element for the steady-state universe of Bondi & Gold (1948) and Hoyle (1948). In terms of the conformal time

$$\eta = -\alpha e^{-t/\alpha}, \quad -\infty < \eta < 0 \tag{5.53}$$

the line element (5.52) becomes

$$ds^2 = (\alpha^2/\eta^2)[d\eta^2 - \sum_{i=1}^{3} (dx^i)^2], \tag{5.54}$$

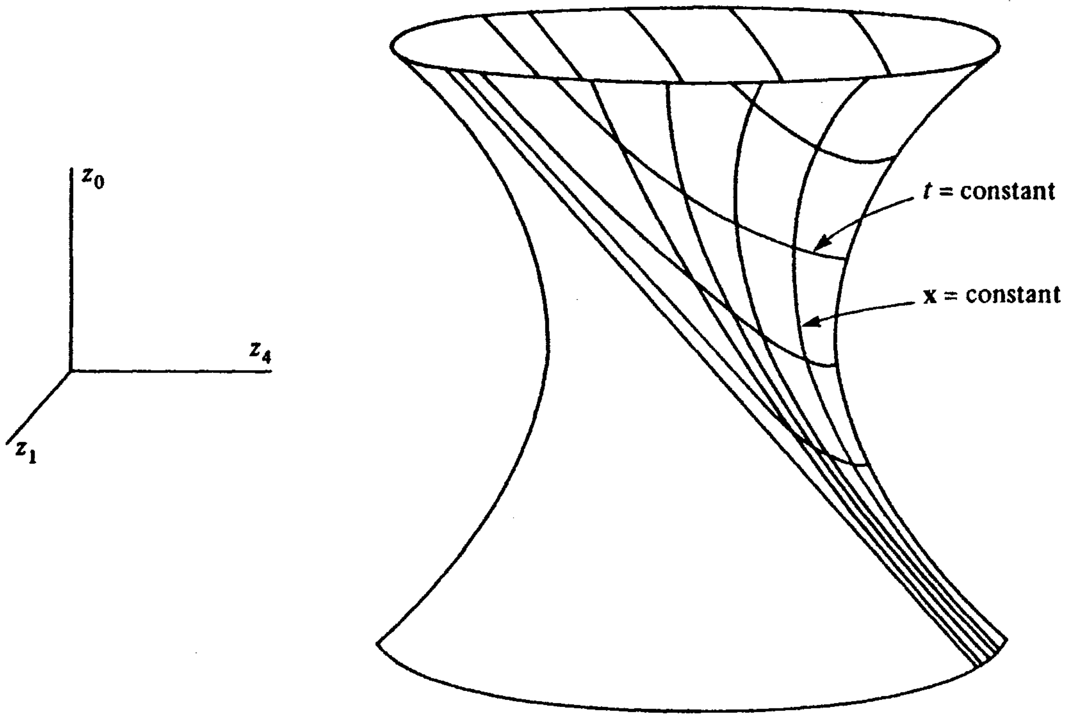


Fig. 16. De Sitter space represented as a hyperboloid embedded in five-dimensional flat spacetime (two dimensions are suppressed in the figure). The coordinates shown cover only half the space, called the steady-state universe, and are bounded by the null rays $z_0 + z_4 = 0$ ($t = -\infty$). The 3-spaces $t = \text{constant}$ are flat, and expand with increasing t.

revealing that this portion of de Sitter space is conformal to a portion of Minkowski space. The remaining half of de Sitter space can be coordinatized by reversing the signs of the right-hand sides in (5.51), or by allowing the conformal time η to range over all real numbers $-\infty < \eta < \infty$.

Since (5.54) has the form of a spatially flat Robertson–Walker line element (5.1) with $C(\eta) = (\alpha/\eta)^2$, we can work within the framework of §5.1. The Ricci scalar is calculated from (5.14) to be

$$R = 12\alpha^{-2}, \tag{5.55}$$

which, when substituted into (5.4) gives an equation for the η-dependent factor χ_k of the field modes. This equation can be solved in terms of Bessel or Hankel functions. The particular linear combination used to define the 'positive frequency' modes will, of course, determine the choice of vacuum.

Since

$$\frac{\mathrm{d}^l}{\mathrm{d}\eta^l}\left(\frac{\dot{C}}{C}\right) \xrightarrow[\eta \to \pm\infty]{} 0 \tag{5.56}$$

we expect an adiabatic vacuum to be a reasonable definition of a no-particle state as $\eta \to \pm\infty$ (see §3.5). This will be achieved if the exact modes are chosen equal (to good approximation) to the adiabatic modes (3.102) whenever

$$T^2\omega_k^2(\eta_1) = T^2k^2 + m^2\alpha^2/\eta_1^2 + 12(\xi - \tfrac{1}{6})/\eta_1{}^2 \tag{5.57}$$

is large compared with $\dot{C}/C = O(T^{-1})$ for fixed $\eta_1 = \eta/T$. That is, they must reduce to

$$\chi_k^{(A)} \xrightarrow[k,\eta\to\infty]{} (1/2k)^{\frac{1}{2}}\,\mathrm{e}^{-ik\eta} \tag{5.58}$$

for large k or η. It is easily verified that the correctly normalized exact solution having this property is

$$\chi_k(\eta) = \tfrac{1}{2}(\pi\eta)^{\frac{1}{2}}\,H_\nu^{(2)}(k\eta) \tag{5.59}$$

where H is a Hankel function and

$$\nu^2 = \tfrac{9}{4} - 12(m^2R^{-1} + \xi). \tag{5.60}$$

Note that (5.59) also reduces to (5.58) in the massless, conformally coupled ($\xi = \frac{1}{6}$) limit. It follows that, unlike in the Milne universe, here the conformal vacuum coincides with the massless limit of the adiabatic vacuum.

Since (5.59) has the asymptotic form of the adiabatic modes (5.58) as $\eta \to +\infty$ as well as when $\eta \to -\infty$, there is no particle production, even though the scale factor passes through a coordinate singularity at $\eta = 0$. This is expected of a vacuum which is invariant under the de Sitter group. However, just as Poincaré invariance of the Minkowski space vacuum does not guarantee its uniqueness, neither does de Sitter invariance define a unique vacuum in de Sitter space (Chernikov & Tagirov 1968). Thus it has been necessary to invoke the adiabatic prescription for defining positive frequency to choose between the various de Sitter invariant contenders for a vacuum state.

To see how well the chosen 'vacuum' accords with the 'no-particle' state of a comoving detector, one must obtain the positive frequency Wightman function G^+ for this vacuum. This has been calculated as a mode sum using (5.59) by several authors (for example, Schomblond & Spindel 1976, Critchley 1976, Bunch 1977, Bunch & Davies 1978*a*), who find

$$\begin{aligned} G^+(x,x') = {} & (16\pi\alpha^2)^{-1}(\tfrac{1}{4} - \nu^2)\sec\pi\nu \\ & \times F\left(\tfrac{3}{2} + \nu, \tfrac{3}{2} - \nu; 2; 1 + \frac{(\Delta\eta - i\varepsilon)^2 - |\Delta\mathbf{x}|^2}{4\eta\eta'}\right) \end{aligned} \tag{5.61}$$

where F is a hypergeometric function. In the massless, conformally coupled limit, (5.61) reduces to

$$D^+(x,x') = \frac{-\eta\eta'}{4\pi^2\alpha^2[(\Delta\eta - i\varepsilon)^2 - |\Delta\mathbf{x}|^2]} \tag{5.62}$$

which bears the expected conformal relation to the Minkowski space Wightman function (3.59) (see §3.7).

Substituting (5.62) into (3.55), and setting $\Delta\mathbf{x} = 0$, we can calculate the response function of a comoving observer with proper time t moving in the conformal (equivalently adiabatic) vacuum. Confining attention to the 'steady-state' part of de Sitter space defined by (5.53) we find

$$\begin{aligned}\mathscr{F}(E) &= \frac{1}{16\pi^2\alpha^2}\int_{-\infty}^{\infty} d\left(\frac{t+t'}{2}\right)\int_{-\infty}^{\infty} d\Delta t \frac{e^{-iE\Delta t}}{\sin^2[(i\Delta t/2\alpha)+\varepsilon]} \\ &= \frac{1}{16\pi^4\alpha^2}\int_{-\infty}^{\infty} d\left(\frac{t+t'}{2}\right)\sum_{n=-\infty}^{\infty}\int_{-\infty}^{\infty} d\Delta t \frac{e^{-iE\Delta t}}{[(i\Delta t/2\alpha\pi)+\varepsilon-n]^2}.\end{aligned}$$

Performing a contour integral for Δt, closing the contour in the lower-half complex Δt plane ($E > 0$), one obtains

$$\mathscr{F}(E)/(\text{unit time}) = (E/2\pi)(e^{2\alpha\pi E} - 1)^{-1}, \tag{5.63}$$

which is a thermal spectrum with temperature $T = 1/(2\pi\alpha k_B)$. Thus, although no particles are created, a comoving observer inhabiting this portion of de Sitter space will perceive a thermal bath of radiation. This is another example of the general phenomenon, discussed in §3.7, of the nonzero response of a comoving detector to a conformal vacuum.

The result obtained here is not in conflict with the fact that the conformal vacuum is also an adiabatic vacuum of arbitrary order, for it was shown in §3.6 that the probability for a comoving detector to detect a particle of energy E in such a vacuum should fall to zero faster than any inverse power of E in the limit of large E (see page 78). This is certainly the case with (5.63). Similarly it falls to zero faster than any inverse power of the adiabatic parameter α, which gives a measure of the magnitude of the curvature of the spacetime ($R = 12/\alpha^2$).

A second coordinate system (t, χ, θ, ϕ) frequently employed is defined by

$$\left.\begin{aligned} z_0 &= \alpha\sinh(t/\alpha), \\ z_1 &= \alpha\cosh(t/\alpha)\cos\chi, \\ z_2 &= \alpha\cosh(t/\alpha)\sin\chi\cos\theta, \\ z_3 &= \alpha\cosh(t/\alpha)\sin\chi\sin\theta\cos\phi, \\ z_4 &= \alpha\cosh(t/\alpha)\sin\chi\sin\theta\sin\phi, \end{aligned}\right\} \tag{5.64}$$

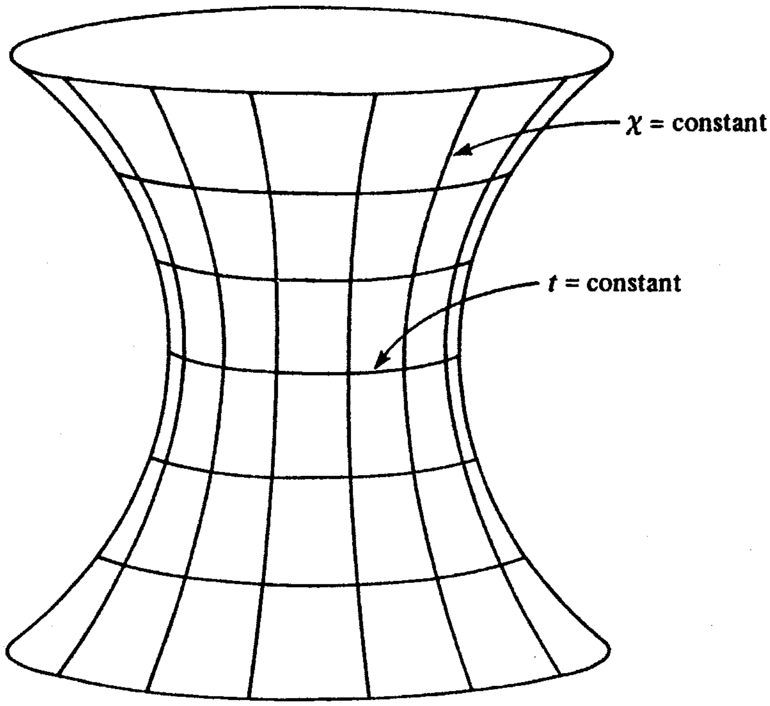


Fig. 17. De Sitter space represented as a contracting and re-expanding compact space with metric (5.65).

in which the line element (5.50) becomes

$$ds^2 = dt^2 - \alpha^2 \cosh^2(t/\alpha)[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)]. \quad (5.65)$$

If $-\infty < t < \infty$, $0 \le \chi \le \pi$, $0 \le \theta \le \pi$, $0 \le \phi \le 2\pi$ then the coordinates cover the whole de Sitter manifold. The metric (5.65) is that of a $K = +1$ (closed) Robertson–Walker spacetime (see fig. 17).

Introducing the conformal time

$$\eta = 2\tan^{-1}(e^{t/\alpha}), \quad 0 \le \eta < \pi, \quad (5.66)$$

the line element (5.65) can be cast into a form which is manifestly conformal to the Einstein universe:

$$ds^2 = \alpha^2 \sin^{-2}\eta[d\eta^2 - d\chi^2 - \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)]. \quad (5.67)$$

Using this line element, (5.4) reduces to

$$\frac{d^2\chi_k}{d\eta^2} + \{k^2 + \operatorname{cosec}^2\eta[m^2\alpha^2 + 12(\xi - \tfrac{1}{6})]\}\chi_k = 0, \quad (5.68)$$

which possesses solutions that can be written in terms of Legendre functions (Tagirov 1973, Critchley 1976):

$$\chi_k = \sin^{\frac{1}{2}}\eta[AP^{\nu}_{k-\frac{1}{2}}(-\cos\eta) + BQ^{\nu}_{k-\frac{1}{2}}(-\cos\eta)] \quad (5.69)$$

where, from (5.22), $k = 1, 2, 3, \ldots$ and ν is given by (5.60). The choice of constants A, B determine the vacuum. Unfortunately, because

$$\dot{C}/C = -2\cot\eta \to \pm\infty \text{ as } t \to \pm\infty,$$

the spacetime is not slowly expanding in these limits, and we cannot define physically reasonable adiabatic in and out vacua.

This difficulty is highlighted if one applies a Liouville transformation (Liouville 1837, see also Fulling 1979) to (5.68). In this case such a transformation can be effected by changing back from variable η to t and defining

$$\chi_1(t) = \sin^{-\frac{1}{2}}(\eta)\chi(\eta),$$

whence (5.68) becomes

$$\frac{d^2\chi_1}{dt^2} + \alpha^{-2}[\sin^2\eta(k^2 - \tfrac{1}{2}) - \tfrac{1}{4}\cos^2\eta + m^2\alpha^2 + 12(\xi - \tfrac{1}{6})]\chi_1 = 0. \qquad (5.70)$$

In the distant past and future, $t \to \pm\infty$, and this equation reduces to

$$\frac{d^2\chi_1}{dt^2} - (\nu/\alpha)^2\chi_1 = 0$$

with a solution

$$\chi_1 \propto e^{-\nu t/\alpha} \qquad (5.71)$$

which is positive frequency with respect to t (for $\nu^2 < 0, \operatorname{Im}\nu > 0$). Some authors (e.g., Gutzwiller 1956, Critchley 1976, Rumpf 1976*a*, *b*; see also Dowker & Critchley 1976*a*) have defined vacuum states by taking positive frequency solutions of (5.68) or (5.70) to be those which behave like (5.71) for large $|t|$. With such a definition they have found that the in and out vacua are not equivalent and that the Bogolubov transformation connecting them is constant (independent of frequency), implying an infinite amount of particle production. (Fulling, 1972, seems to have been the first person to remark on the inappropriate nature of this vacuum. See also Schäfer 1978.) This pathological result is explained by the fact that the regions $t \to \pm\infty$, and hence the above defined vacuum states, are not adiabatic in this case. However it follows from §3.5 that if one requires the Bogolubov coefficient β_k to fall off for large k sufficiently fast, so as to guarantee a finite total particle density, then it is necessary to use an adiabatic vacuum of order two or more.

Although adiabatic in and out regions do not exist, it is still possible to define adiabatic vacua as being those which are vacuous in the high momentum modes. Recall that the adiabatic modes (3.102) become good approximations to *exact* adiabatic positive frequency modes when

$$T^2\omega_k^2(\eta_1) = T^2k^2 + T^2\operatorname{cosec}^2(\eta_1 T)[m^2\alpha^2 + 12(\xi - \tfrac{1}{6})] \qquad (5.72)$$

becomes large with respect to the derivatives of $\dot{C}/C$ for fixed $\eta_1 = \eta/T$. This will clearly be the case for large k, m or α (for fixed η). It is not, however, the case for large $|t|$, i.e. $\eta \to 0$ or π, whence $T \to 0, \pi/\eta_1$. For example, as $\eta \to \pi$, $T \to \pi/\eta_1$, (5.72) becomes infinite at the same rate as $\mathrm{d}^l(\dot{C}/C)/\mathrm{d}\eta^l$ for $l = 1$, and more slowly for $l > 1$.

In the limit of large k (3.102) yields, to zeroth order

$$\chi_k^{(0)}(\eta) \xrightarrow[k\to\infty]{} \mathrm{e}^{-\mathrm{i}k\eta}/(2|k|)^{\frac{1}{2}}. \qquad (5.73)$$

Choosing constants A, B such that the correctly normalized solution (5.69) reduces to (5.73) for large k, one obtains (Chernikov & Tagirov 1968, Tagirov 1973)

$$\chi_k(\eta) = \sin^{\frac{1}{2}}\eta[\pi\Gamma(k + \tfrac{1}{2} - \nu)/4\Gamma(k + \tfrac{1}{2} + \nu)]^{\frac{1}{2}}\mathrm{e}^{+\mathrm{i}\nu\pi/2}$$
$$\times [P^{\nu}_{k-\frac{1}{2}}(-\cos\eta) - (2\mathrm{i}/\pi)Q^{\nu}_{k-\frac{1}{2}}(-\cos\eta)]. \qquad (5.74)$$

Since the solution (5.74) reduces to (5.73) in the limit of large k, regardless of the value of η, it defines an adiabatic vacuum for all time. Thus, as in the previous coordinatization of de Sitter space there is no 'particle' production; (5.74) defines a stable adiabatic vacuum. Further, in the massless conformally coupled limit the modes (5.74) reduce to the right-hand side of (5.73), so once again the conformal vacuum and the massless limit of the adiabatic vacuum coincide.

One can now use (5.74) to evaluate the positive frequency Wightman function for the adiabatic vacuum (Tagirov 1973). It turns out to be precisely the same function of the coordinates z_μ of the embedding space as (5.61). Thus the vacuum defined in the coordinatization (5.54) is equivalent to that defined in the coordinatization (5.67). Moreover, it turns out that a particle detector at rest in this coordinate system will perceive this vacuum state as a bath of thermal radiation with temperature $T = 1/(2\pi\alpha k_B)$, exactly as in the steady-state case (see (5.63)). This result was originally discovered by Gibbons & Hawking (1977*a*), who used it to argue for an observer-dependent formulation of the particle concept in quantum gravitational physics.

The final coordinates that we shall study is the static system defined by

$$\left.\begin{aligned} z_0 &= (\alpha^2 - r^2)^{\frac{1}{2}} \sinh(t/\alpha) \\ z_1 &= (\alpha^2 - r^2)^{\frac{1}{2}} \cosh(t/\alpha) \\ z_2 &= r\sin\theta\cos\phi \\ z_3 &= r\sin\theta\sin\phi \\ z_4 &= r\cos\theta, \quad 0 \le r < \alpha \end{aligned}\right\} \tag{5.75}$$

which only covers the half of the de Sitter manifold with $z_0 + z_1 > 0$ (see fig. 18). In these coordinates the line element (5.50) becomes

$$ds^2 = [1 - (r^2/\alpha^2)]dt^2 - [1 - (r^2/\alpha^2)]^{-1}dr^2 - r^2(d\theta^2 + \sin^2\theta\, d\phi^2). \quad (5.76)$$

The line element (5.76) possesses a coordinate singularity at $r = \alpha$, which is the event horizon for an observer situated at $r = 0$, following the trajectory of the Killing vector ∂_t. The structure of (5.76) is reminiscent of the Schwarzschild line element (3.18), and one may develop similar field quantizations in the two cases (Gibbons & Hawking 1977*a*, Lapedes 1978*b* and Lohiya & Panchapakesan 1978, 1979; see also Hájíček 1977).

With this formulation of quantum field theory in de Sitter space, the natural choice of field modes and vacuum state associated with the coordinate system (5.75) enjoys a different status from the previous two formulations that we have discussed as the vacuum state is not invariant

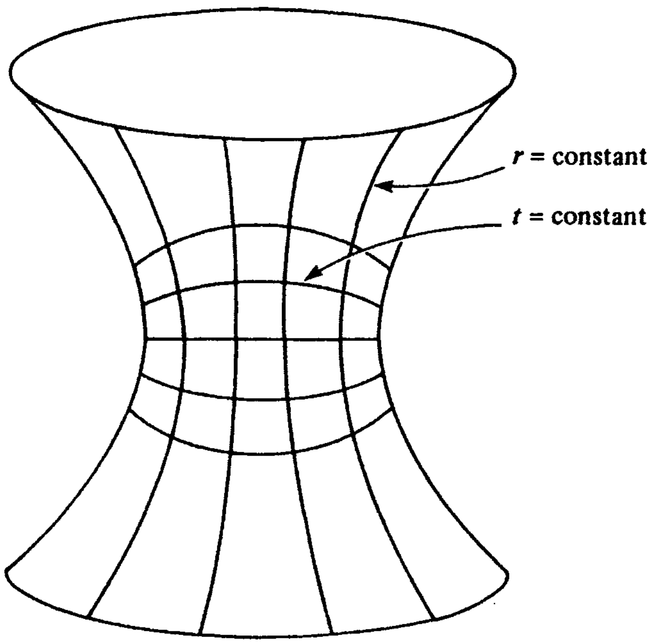


Fig. 18. De Sitter space as a static spacetime with metric (5.76).

under the de Sitter group (see Chernikov & Tagirov 1968). In fact, in §6.3 we shall show that the expectation value of the quantum field stress-tensor in this vacuum state diverges at the event horizon $r = \alpha$. As the location of the horizon is dependent on the origin of radial coordinates, each point in de Sitter space (i.e., each observer) must be associated with a different choice of vacuum state. Hence, this prescription leads to a vacuum that is not even translation invariant, and which comoving observers apparently perceive as a bath of thermal radiation. (See also the discussion of Schrödinger, 1956, §1.4, on the relation between the observer at $r = 0$ and the properties of the static coordinate system.)

5.5 Classification of conformal vacua

Before turning to a more general class of spacetimes, we can consolidate much of the information of the previous examples using a classification scheme proposed by Candelas & Dowker (1979). This approach will also prove to be very useful for discussing stress-tensors in various spacetimes.

The idea behind the scheme is to classify the conformal vacuum states associated with various spacetimes, and it proceeds as follows: Suppose that M is a curved spacetime which is conformally mapped into the flat spacetime $\tilde{M}$ (which need not necessarily be unbounded Minkowski space), and further suppose that Σ is a global Cauchy hypersurface of M which is mapped under the conformal transformation to a global Cauchy hypersurface $\tilde{\Sigma}$ of $\tilde{M}$, then for every globally timelike Killing vector in $\tilde{M}$, there exists a globally timelike conformal Killing vector in M. Thus, we can classify the conformal vacua defined with respect to the latter conformal Killing vectors by reference to the vacua defined in $\tilde{M}$. In particular if $\tilde{M}$ is Minkowski space, then because there exists only one timelike Killing vector field which is global, there is a unique vacuum defined with respect to the Killing vector ∂_t, while if $\tilde{M}$ is Rindler spacetime, then there are two vacua, one defined with respect to ∂_t and one with respect to ∂_η (see (4.82)).

The properties of global Cauchy hypersurfaces of the various Robertson–Walker spacetimes that we have dealt with are most easily discussed by studying the Penrose diagrams of the spacetimes (see §3.1 or Hawking & Ellis 1973, chapter 5), obtained by conformally mapping the spacetimes into the Einstein static universe (see §5.2), which has the line element

$$ds^2 = dt^2 - d\chi^2 - \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{5.77}$$

Suppressing the coordinates θ, ϕ, the Einstein universe can be represented

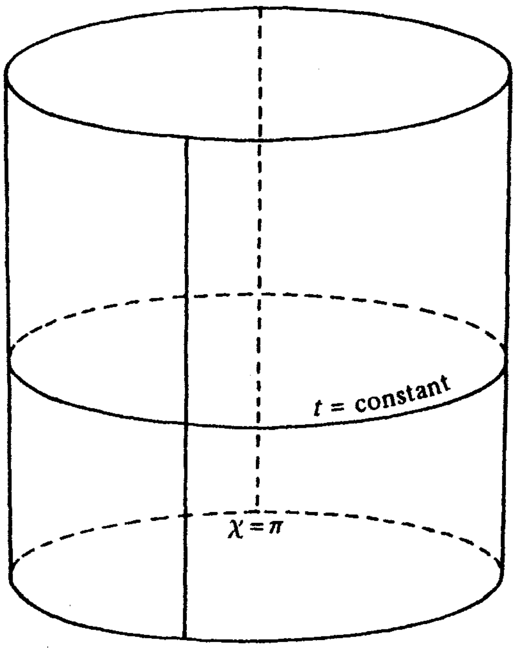


Fig. 19. Representation of the Einstein universe as a cylinder. The coordinates θ and ϕ are suppressed, t runs vertically from $-\infty$ to $+\infty$ and χ runs around the cylinder.

as the cylinder shown in fig. 19. Unwrapping this cylinder gives the region $-\infty < t < \infty$, $0 \leq \chi < \pi$ of the t, χ plane. The conformal images of other spacetimes on this region constitute the Penrose diagrams that we wish to study.

The Robertson–Walker cases are shown in fig. 20.

In the $K = 1, 0, -1$ Robertson–Walker spacetimes, figs. 20*b*, *d*, *f* respectively, the exact range of the conformal time will depend on the form of $C(\eta)$ appearing in (5.11). We have restricted the range to $\eta > 0$, but if this restriction is lifted and the spacetime is extended through what is usually a singularity at $\eta = 0$, then the bottom halves of the diagrams are filled in by reflection in the line $t = 0$. Similarly, if the steady-state universe in the form (5.54) is extended by allowing the range $-\infty < \eta < \infty$, then the top half of fig. 20*c* is filled in by reflection in $t = 0$.

The most important fact revealed by these diagrams is that the spacetimes represented by figs. 20*a*–*d* all have the surface $t = 0, 0 \leq \chi < \pi$ as the common conformal image of a global Cauchy hypersurface, while the spacetimes in figs. 20*e*, *f* have $t = \pi/2$, $0 \leq \chi < \pi/2$ as a common surface.

From this we can immediately classify the spacetimes of figs. 20 into two classes: (i) figs. 20*a*–*d*, only having conformal vacua defined with respect to the conformal Killing vector inherited by conformal transformation from the timelike Killing vector ∂_t of Minkowski space (t being Minkowski time);

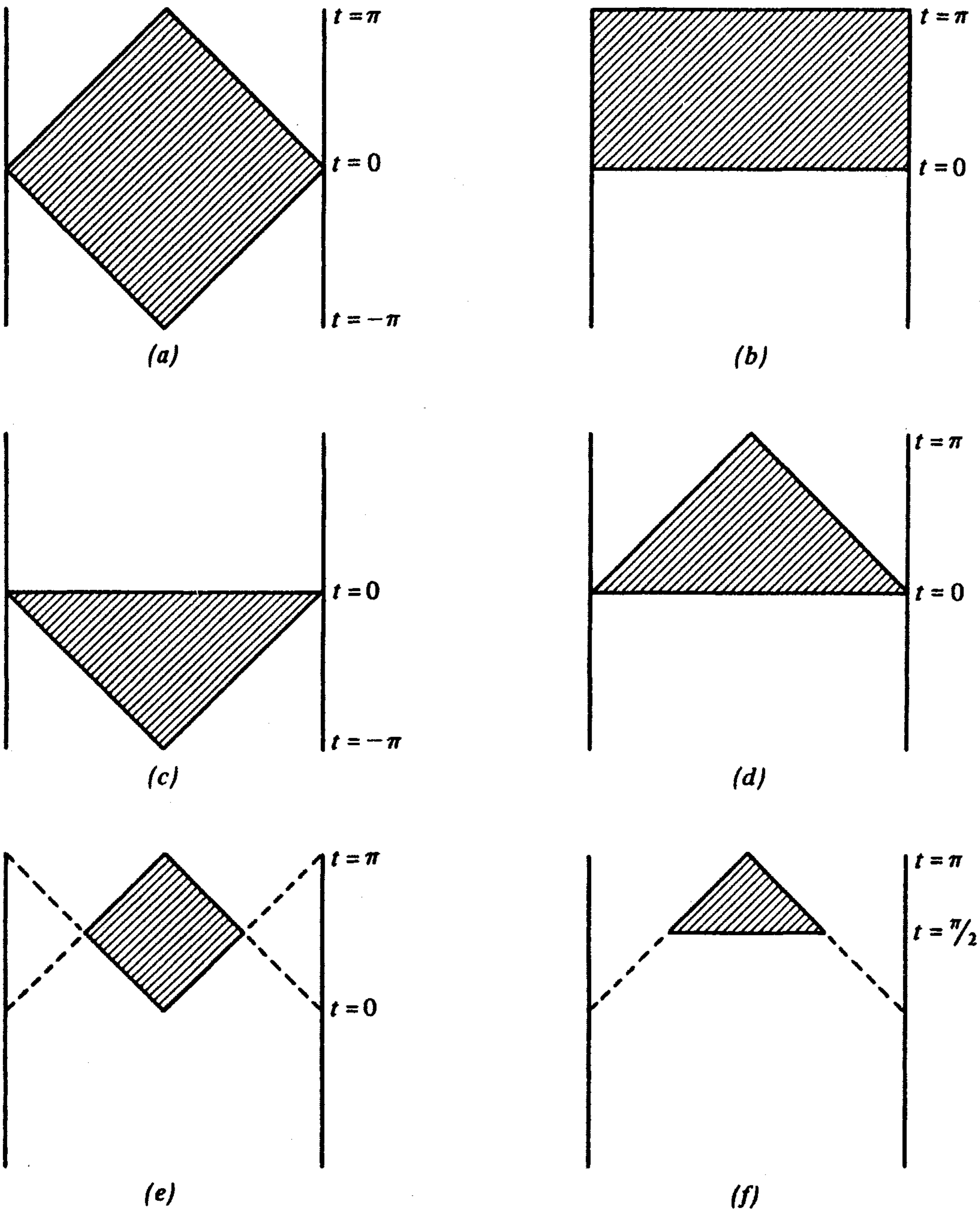


Fig. 20. The conformal images of various spacetimes in the Einstein cylinder. The coordinate t is the time coordinate in the Einstein cylinder (fig. 19). (*a*) Minkowski space, (*b*) de Sitter space and $K=1$ Robertson – Walker spacetime, (*c*) steady-state universe, (*d*) $K=0$ Robertson – Walker spacetime, (*e*) the Rindler wedge, static de Sitter space, the open Einstein universe and the Milne universe, (*f*) $K=-1$ Robertson – Walker spacetime. (After Candelas & Dowker 1979.)

(ii) figs. 20*e*, *f*, having two possible conformal vacua defined with respect to the conformal Killing vector inherited from the timelike Killing vector ∂_t and ∂_η of Rindler spacetime.

Since both the Minkowski and Rindler spacetimes are flat, the classification of conformal vacua has been reduced to purely topological considerations. Different topologies can be distinguished by the value of the topological invariant called the Euler–Poincaré characteristic (see, Gilkey 1975), which has been used by Christensen & Duff (1978*b*) to characterize the (Euclideanized) Minkowski and Rindler vacua. It takes the value 1 for Minkowski space (topology R^4) and 0 for Rindler spacetime (topology $R^3 \times S^1$).

What do these topological characterizations mean physically? In §4.5 we showed that the Minkowski vacuum contains a thermal spectrum of Rindler particles at temperature $a/2\pi k_B$. One can also demonstrate this by showing that the Green functions in the Minkowski vacuum are Rindler thermal Green functions (see §2.7). In a similar way one can relate the vacua of other spacetimes in fig. 20 which have similar geometries but different topologies. For example, static de Sitter space (5.76) and de Sitter space (5.65) have the same curvature, but static de Sitter space is a member of the Rindler class, fig. 20*e*, while de Sitter space is a member of the Minkowski class, fig. 20*b*. From this one can deduce that the de Sitter conformal vacuum is given by the thermalization of the static de Sitter ∂_η conformal vacuum at a temperature $1/2\pi\alpha k_B$. Similarly the conformal vacuum of the Einstein universe is the thermalization at imaginary temperature $T = 1/2\pi i k_B$ of the ∂_η conformal vacuum of the open Einstein universe (see Candelas & Dowker 1979 for further discussion).

These results are summarized in fig. 21, where the vertical arrows denote conformal transformations and the horizontal lines denote thermalization at the temperature shown. In addition to this, from our discussion of

$$
\begin{tikzcd}
\text{Rindler wedge} \arrow[r, "(2\pi a)^{-1}"] \arrow[d, leftrightarrow] & \text{Minkowski space} \arrow[d, leftrightarrow] \\
\text{Static de Sitter space} \arrow[r, "(2\pi\alpha)^{-1}"] \arrow[d, leftrightarrow] & \text{De Sitter space} \arrow[d, leftrightarrow] \\
\text{Open Einstein universe} \arrow[r, "(2\pi i)^{-1}"] \arrow[d, leftrightarrow] & \text{Einstein universe} \arrow[d, leftrightarrow] \\
K = -1 \text{ Robertson–Walker} \arrow[d, leftrightarrow] & K = 1 \text{ Robertson–Walker} \arrow[d, leftrightarrow] \\
\text{The Milne universe} & K = 0 \text{ Robertson–Walker} \arrow[d, leftrightarrow] \\
 & \text{Steady-state universe}
\end{tikzcd}
$$

Fig. 21. Thermal and conformal relationships between various spacetimes. The vertical arrows indicate that the spacetimes are related by conformal transformations, while horizontal arrows represent relationship by thermalization at the temperature given. (After Candelas & Dowker 1979).

particle detectors in the steady-state universe (see (5.63)), we know that the conformal vacuum in the steady-state universe can be obtained from the Minkowski vacuum by a thermalization at temperature $1/2\pi\alpha k_B$. That is, the vertical conformal relation in the diagram between the Minkowski and steady-state vacua is also a thermalization relation in this case. One can verify that a similar relation holds between the Einstein and the de Sitter (line element (5.65)) conformal vacua.

5.6 Bianchi I spacetimes and perturbation theory

So far, all of our examples have involved isotropic spacetimes. These are especially interesting because the universe today is observed to be to a high degree of approximation isotropic. The origin of this isotropy has long been a puzzle to cosmologists (see, for example, the review of MacCallum 1979). Amongst other speculations, it has been suggested (Zel'dovich 1970) that the universe may have begun anisotropic, but rapidly isotropized as a result of quantum effects in the primeval phase of the big bang. The investigation of this suggestion is technically difficult, because it involves the back-reaction of quantum field effects, induced by spacetime curvature on the dynamics of the gravitational field. We shall briefly return to this aspect at the end of chapter 7 after the stress–energy–momentum tensor has been discussed. Here we restrict attention to the quantum field theory.

Consider the four-dimensional (spatially flat) Bianchi type I spacetimes (Bianchi 1918) with line element

$$ds^2 = dt^2 - \sum_{i=1}^{3} a_i^2(t)dx_i^2. \tag{5.78}$$

(See Hu 1972, 1973, 1974, and Hu, Fulling & Parker 1973 for consideration of quantum field theory in closed, anisotropic, Bianchi IX spacetimes. See also Berger 1975.)

Define

$$[a(t)]^2 = C(t) \equiv (a_1a_2a_3)^{\frac{2}{3}} = (C_1C_2C_3)^{\frac{1}{3}}, \tag{5.79}$$

where $C_i \equiv a_i^2$. It proves convenient to introduce a new time parameter η by

$$\eta = \int^t a^{-1}(t')dt', \tag{5.80}$$

which reduces to the conformal time (5.10) in the isotropic limit $a_1 = a_2 = a_3 = a$. If we also define

$$d_i = \dot{C}_i/C_i \tag{5.81}$$

where a dot, as usual, denotes differentiation with respect to η, and

$$D \equiv \tfrac{1}{3}\sum_{i=1}^{3} d_i = \dot{C}/C, \tag{5.82}$$

$$Q \equiv \tfrac{1}{72}\sum_{i<j}(d_i - d_j)^2, \tag{5.83}$$

then the nonzero Christoffel symbols for the metric (5.78) are

$$\Gamma^{\eta}{}_{\eta\eta} = \tfrac{1}{2}D, \quad \Gamma^{\eta}{}_{ii} = \tfrac{1}{2}d_i C_i/C, \quad \Gamma^{i}{}_{i\eta} = \Gamma^{i}{}_{\eta i} = \tfrac{1}{2}d_i. \tag{5.84}$$

From these one may obtain the nonzero components of the Ricci tensor (see, for example, Fulling, Parker & Hu 1974):

$$R_{\eta\eta} = \tfrac{3}{2}\dot{D} + 6Q, \quad R_{ii} = -\tfrac{1}{2}C^{-1}C_i(\dot{d}_i + d_i D), \tag{5.85}$$

and the Ricci scalar:

$$R = C^{-1}[3\dot{D} + \tfrac{3}{2}D^2 + 6Q]. \tag{5.86}$$

Further, defining

$$\left.\begin{aligned} S &= \tfrac{1}{144}\sum_{\substack{i<j\\ k\neq i,j}}(d_k - D)(d_i - d_j)^2 \\ U &= \tfrac{1}{72}\sum_{i<j}(\dot{d}_i - \dot{d}_j)^2 \end{aligned}\right\} \tag{5.87}$$

the square of the Weyl tensor becomes

$$C^{\alpha\beta\gamma\delta}C_{\alpha\beta\gamma\delta} = 4C^{-2}(3U - 4\dot{S} + 12Q^2), \tag{5.88}$$

which gives a measure of the anisotropy of the spacetime. Because it is not conformally flat we are unable to define conformal vacuum states for field theories with conformally invariant Lagrangians, as we did in the case of the Robertson–Walker spacetimes. Adiabatic vacua still exist though (Fulling, Parker & Hu, 1974, Hu 1978, Fulling 1979).

The mode decomposition (3.30) with measure appropriate to flat spatial sections, is

$$\phi(x) = \int d^3k[a_{\mathbf{k}}u_{\mathbf{k}}(x) + a_{\mathbf{k}}^{\dagger}u_{\mathbf{k}}^{*}(x)]. \tag{5.89}$$

Because spatial translation invariance is still a symmetry of Bianchi I models, the modes can be separated (cf. (5.3)):

$$u_{\mathbf{k}}(x) = (2\pi)^{-\frac{3}{2}} e^{i\mathbf{k}\cdot\mathbf{x}} C^{-\frac{1}{2}}(\eta)\chi_{\mathbf{k}}(\eta). \tag{5.90}$$

Substituting this into (3.26) one obtains an equation for $\chi_{\mathbf{k}}$ (cf. (5.4)):

$$\frac{d^2\chi_{\mathbf{k}}}{d\eta^2} + \left\{C(\eta)\left(\sum_{i=1}^{3} \frac{k_i^2}{C_i(\eta)} + m^2 + (\xi - \tfrac{1}{6})R(\eta)\right) + Q(\eta)\right\}\chi_{\mathbf{k}} = 0. \tag{5.91}$$

From (3.28), the normalization condition becomes (cf. (5.6))

$$\chi_{\mathbf{k}}\partial_\eta\chi_{\mathbf{k}}^* - \chi_{\mathbf{k}}^*\partial_\eta\chi_{\mathbf{k}} = i. \tag{5.92}$$

As in §3.5, a positive frequency generalized WKB-type solution to (5.91) may be obtained from

$$\chi_{\mathbf{k}} = (2W_{\mathbf{k}})^{-\frac{1}{2}}\exp\left[-i\int^{\eta} W_{\mathbf{k}}(\eta')d\eta'\right], \tag{5.93}$$

where $W_{\mathbf{k}}$ is constructed to any adiabatic order A (to give $W_{\mathbf{k}}^{(A)}$, $\chi_{\mathbf{k}}^{(A)}$) by iteration from

$$W_{\mathbf{k}}^{(0)} = \omega_{\mathbf{k}} = \left\{C(\eta)\left(\sum_{i=1}^{3}\frac{k_i^2}{C_i} + m^2\right)\right\}^{\frac{1}{2}}. \tag{5.94}$$

One may now go on to define adiabatic vacuum states using the procedure already described for the Robertson–Walker models, i.e., by matching exact positive frequency solutions of (5.91) to the above approximate solutions at some instant of time.

Unfortunately it is extremely difficult to find exact solutions of (5.91). One simple case can, however, be found:

$$a_1(t) = t, \quad a_2 = a_3 = 1, \quad 0 < t < \infty, -\infty < x^i < \infty. \tag{5.95}$$

This spacetime is similar to the Milne universe of §5.3, as may be seen by employing the coordinate transformation

$$y^0 = t\cosh x^1, \quad y^1 = t\sinh x^1, \quad y^2 = x^2, \quad y^3 = x^3 \tag{5.96}$$

under which the line element (5.78) with (5.95) reduces to that for Minkowski space, with the restriction $y^0 > |y^1|$.

In this very simple case, (5.91) can be solved in terms of Bessel or Hankel functions (Fulling, Parker & Hu 1974, Nariai 1976, 1977*a*, *b*, *c*). Inspection then shows that the solution given in terms of the Hankel function $H^{(2)}$ is positive frequency with respect to the adiabatic definition. As in the Milne case, it also turns out to be positive frequency with respect to the usual definition in the covering Minkowski space.

In view of the difficulty in finding exact solutions of (5.91), it is natural to investigate approximation methods for solving it. One such method, which was originally exploited by Zel'dovich & Starobinsky (1971, 1977) and has subsequently been developed by Birrell & Davies (1980*a*), is the solution of (5.91) for small anisotropic perturbations about a Robertson–Walker spacetime (see also §7.3).

We give below an outline of their method. The line element (5.78) is first cast in the form

$$ds^2 = C(\eta)\{d\eta^2 - \sum_{i=1}^{3} [1 + h_i(\eta)](dx^i)^2\} \tag{5.97}$$

and attention restricted to cases in which

$$\max |h_i(\eta)| \ll 1 \tag{5.98}$$

and, for simplicity of the ensuing discussion,

$$\sum_{i=1}^{3} h_i(\eta) = 0. \tag{5.99}$$

We can now expand (5.91) to first order in h_i, to obtain

$$\frac{d^2\chi_{\mathbf{k}}}{d\eta^2} + \left[k^2 + m^2C(\eta) + (\xi - \tfrac{1}{6})C(\eta)R_{\mathrm{I}}(\eta) - \sum_{i=1}^{3} h_i(\eta)k_i^2\right]\chi_{\mathbf{k}} = 0, \tag{5.100}$$

where R_{I} is the Ricci scalar for the isotropic spacetime obtained by setting $h_i = 0$:

$$R_{\mathrm{I}} = C^{-1}[3\dot{D} + \tfrac{3}{2}D^2]. \tag{5.101}$$

Imposing the conditions

$$\left.\begin{array}{ll} \text{(i)} & h_i(\eta) \to 0 \text{ as } \eta \to \pm\infty \\ \text{(ii)} & C(\eta)R(\eta) \to 0 \text{ as } \eta \to \pm\infty \text{ if } \xi \neq \tfrac{1}{6} \\ \text{(iii)} & C(\eta) \to C(\infty) = C(-\infty) < \infty \text{ as } \eta \to \pm\infty \text{ if } m \neq 0, \end{array}\right\} \tag{5.102}$$

then the normalized positive frequency solution of (5.100) as $\eta \to -\infty$ is immediately given as

$$\chi_{\mathbf{k}}^{\mathrm{in}}(\eta) = (2\omega)^{-\frac{1}{2}} e^{-i\omega\eta}, \tag{5.103}$$

with

$$\omega^2 = k^2 + m^2C(\infty). \tag{5.104}$$

A method of relaxing the conditions (ii) and (iii) above using Liouville transformations has been given by Birrell (1979*a*, *c*).

With the initial condition (5.102), (5.100) can be written as an integral equation

$$\chi_{\mathbf{k}}(\eta) = \chi_{\mathbf{k}}^{\text{in}}(\eta) + \omega^{-1}\int_{-\infty}^{\eta} V_{\mathbf{k}}(\eta')\sin[\omega(\eta-\eta')]\chi_{\mathbf{k}}(\eta')\mathrm{d}\eta'. \quad (5.105)$$

where

$$V_{\mathbf{k}}(\eta) = \sum_i h_i(\eta)k_i^2 + m^2[C(\infty) - C(\eta)] - (\xi - \tfrac{1}{6})C(\eta)R_1(\eta). \quad (5.106)$$

In the late time region (5.105) possesses the solution

$$\chi_{\mathbf{k}}^{\text{out}}(\eta) = \alpha_{\mathbf{k}}\chi_{\mathbf{k}}^{\text{in}}(\eta) + \beta_{\mathbf{k}}\chi_{\mathbf{k}}^{\text{in}*}(\eta), \quad (5.107)$$

where the Bogolubov coefficients are given by

$$\alpha_{\mathbf{k}} = 1 + \mathrm{i}\int_{-\infty}^{\infty} \chi_{\mathbf{k}}^{\text{in}*}(\eta)V_{\mathbf{k}}(\eta)\chi_{\mathbf{k}}(\eta)\mathrm{d}\eta \quad (5.108)$$

$$\beta_{\mathbf{k}} = -\mathrm{i}\int_{-\infty}^{\infty} \chi_{\mathbf{k}}^{\text{in}}(\eta)V_{\mathbf{k}}(\eta)\chi_{\mathbf{k}}(\eta)\mathrm{d}\eta. \quad (5.109)$$

If we treat $V_{\mathbf{k}}(\eta)$ as 'small', then we can solve (5.105) by iteration. To lowest order in $V_{\mathbf{k}}$, one has $\chi_{\mathbf{k}}(\eta) = \chi_{\mathbf{k}}^{\text{in}}(\eta)$, which, when substituted in (5.108) and (5.109) gives the Bogolubov coefficients to first order in $V_{\mathbf{k}}$:

$$\alpha_{\mathbf{k}} = 1 + (\mathrm{i}/2\omega)\int_{-\infty}^{\infty} V_{\mathbf{k}}(\eta)\mathrm{d}\eta \quad (5.110)$$

$$\beta_{\mathbf{k}} = -(\mathrm{i}/2\omega)\int_{-\infty}^{\infty} \mathrm{e}^{-2\mathrm{i}\omega\eta}V_{\mathbf{k}}(\eta)\mathrm{d}\eta. \quad (5.111)$$

The formalism can also be used for cosmological models which expand from the singularity at $\eta = 0$, so long as $V_{\mathbf{k}}(\eta)$ vanishes as $\eta \to 0$. One simply replaces the lower limits of η integrations by zero.

The technique is really one of perturbation expansion about conformal triviality, which is broken by the anisotropy. We shall develop the generalized case, however, in which the presence of a (small) mass or (small) non-conformal coupling also contributes to the conformal symmetry breaking. Thus, for $m = 0$, $\xi = \frac{1}{6}$, the condition of small $V_{\mathbf{k}}$ corresponds simply to the small anisotropy condition (5.98). If $m \neq 0$, $\xi \neq \frac{1}{6}$, then the small $V_{\mathbf{k}}$ approximation will be valid provided m is small and $\xi \simeq \frac{1}{6}$. If $h_i \equiv 0$ these latter conditions imply a small deviation from conformal invariance in a Robertson–Walker spacetime with scale factor $a(\eta) = C^{\frac{1}{2}}(\eta)$.

If the quantum state chosen corresponds to the in vacuum, then in the out region ($\eta \to \infty$) the number density (per unit proper volume) is (cf (3.42))

$$n = (2\pi a)^{-3} \int |\beta_{\mathbf{k}}|^2 d^3k \tag{5.112}$$

and the energy density is

$$\rho = (2\pi)^{-3} a^{-4} \int |\beta_{\mathbf{k}}|^2 \omega \, d^3k. \tag{5.113}$$

Substituting (5.111) and (5.106) into (5.112), and performing the momentum integral (Birrell & Davies 1980*a*) one obtains for the second order (in $V_{\mathbf{k}}$) approximation to the number density

$$\begin{aligned} n = {} & (960\pi a^3)^{-1} \int_{-\infty}^{\infty} [C^2(\eta) C^{\alpha\beta\gamma\delta} C_{\alpha\beta\gamma\delta}(\eta) + 60 V^2(\eta)] d\eta \\ & + \bar{m}(960 a^3)^{-1} \int_{-\infty}^{\infty} d\eta_1 \int_{-\infty}^{\infty} d\eta_2 \Big\{ F(\bar{m}(\eta_1 - \eta_2)) \Big[-60\, V(\eta_1) V(\eta_2) \\ & + (8\bar{m}^4 - 6\bar{m}^2 \partial_{\eta_1} \partial_{\eta_2} + \tfrac{3}{2} \partial^2_{\eta_1} \partial^2_{\eta_2}) \sum_i h_i(\eta_1) h_i(\eta_2) \Big] \\ & - J_1(2\bar{m}(\eta_1 - \eta_2)) \Big[60 V(\eta_1) V(\eta_2) + \tfrac{1}{2} \sum_i \ddot{h}_i(\eta_1) \ddot{h}_i(\eta_2) \Big] \Big\} \end{aligned} \tag{5.114}$$

where

$$\bar{m}^2 \equiv C(\infty) m^2, \tag{5.115}$$

$$V(\eta) \equiv m^2[C(\infty) - C(\eta)] - (\xi - \tfrac{1}{6}) C(\eta) R_1(\eta), \tag{5.116}$$

$$F(x) \equiv -(1/\pi) + x[J_0(2x) \mathbb{H}_{-1}(2x) + \mathbb{H}_0(2x) J_1(2x)], \tag{5.117}$$

$\mathbb{H}$ denoting Struve functions and J Bessel functions. We have also used the fact that to second order in h_i, (5.88) reduces to $C^{\alpha\beta\gamma\delta} C_{\alpha\beta\gamma\delta} = \frac{1}{2} C^{-2} \sum_i (\ddot{h}_i)^2$. In the massless limit (5.114) reduces to

$$n = (960\pi a^3)^{-1} \int_{-\infty}^{\infty} C^2(\eta) [C^{\alpha\beta\gamma\delta} C_{\alpha\beta\gamma\delta} + 60(\xi - \tfrac{1}{6})^2 R_1^2] d\eta, \tag{5.118}$$

which is purely geometrical. In arriving at (5.114) and (5.118) we have, in addition to (5.102), imposed the condition

$$\dot{h}_i(\eta), \ddot{h}_i(\eta) \to 0 \quad \text{as}\ \eta \to \pm\infty. \tag{5.119}$$

If we further demand that

$$\dddot{h}_i(\eta) \to 0 \quad \text{as}\, \eta \to \pm\infty, \tag{5.120}$$

then the second order approximation to the energy density is calculated from (5.113) in a similar way:

$$\rho = (3840\pi^2 a^4)^{-1} \int_{-\infty}^{\infty} \mathrm{d}\eta_1 \int_{-\infty}^{\infty} \mathrm{d}\eta_2 \operatorname{Re} K_0(2\mathrm{i}\bar{m}(\eta_1 - \eta_2))(\partial_{\eta_1}\partial_{\eta_2} - 4\bar{m}^2)$$
$$\times \left[120V(\eta_1)V(\eta_2) + (\partial_{\eta_1}\partial_{\eta_2} - 4\bar{m}^2)^2 \sum_i h_i(\eta_1)h_i(\eta_2) \right], \tag{5.121}$$

where K_0 is a modified Bessel function.

In the massless limit (5.121) reduces to

$$\rho = -(3840\pi^2 a^4)^{-1} \int_{-\infty}^{\infty} \mathrm{d}\eta_1 \int_{-\infty}^{\infty} \mathrm{d}\eta_2 \ln[2\mathrm{i}\mu(\eta_1 - \eta_2)]$$
$$\times \left\{ 120\dot{V}(\eta_1)\dot{V}(\eta_2) + \sum_i \dddot{h}_i(\eta_1)\dddot{h}_i(\eta_2) \right\}. \tag{5.122}$$

In this expression μ is an arbitrary mass which can be changed without altering the value of ρ, because

$$\int_{-\infty}^{\infty} \dot{V}(\eta)\mathrm{d}\eta = \int_{-\infty}^{\infty} \dddot{h}_i(\eta)\mathrm{d}\eta = 0.$$

Unfortunately there are very few choices of h_i or V for which the integrals in (5.114) or (5.121) can be evaluated in terms of known functions, though these expressions are still of great value for numerical calculations. A related method of numerically (and sometimes analytically) calculating *exact* results has been given by Birrell (1979*a*, *c*).

A wider class of problems can be solved in closed form by returning to (5.111) to perform the η integration before calculating n or ρ from (5.112) and (5.113). As an example, consider the spacetime with line element (5.97) and

$$h_i(\eta) = \mathrm{e}^{-\alpha\eta^2} \cos(\beta\eta^2 + \delta_i) \tag{5.123}$$

where α, β, δ_i are constants, and (5.99) is satisfied by demanding that the δ_i differ from one another by $2\pi/3$.

From (5.106) and (5.111) the anisotropic contribution to the Bogolubov coefficient $\beta_{\mathbf{k}}$ is found to be

$$\beta_{\mathbf{k}} = -\frac{\mathrm{i}\pi^{\frac{1}{2}}}{2\omega} \sum_i k_i^2 \operatorname{Re}\left(\frac{\mathrm{e}^{-\omega 2/(\alpha + \mathrm{i}\beta)}}{(\alpha + \mathrm{i}\beta)^{\frac{1}{2}}} \mathrm{e}^{-\mathrm{i}\delta_i} \right). \tag{5.124}$$

Substituting this into (5.113) gives the following contribution to the energy density in the out region:

$$\rho = \frac{\bar{m}^2}{1536\pi^{\frac{1}{2}}a^4}\frac{(\alpha^2+\beta^2)^{\frac{3}{2}}}{\alpha^2}\mathrm{e}^{-3\alpha\bar{m}^2/(\alpha^2+\beta^2)}W_{-\frac{3}{2},\frac{3}{2}}\left(\frac{2\alpha\bar{m}^2}{\alpha^2+\beta^2}\right), \tag{5.125}$$

where W is a Whittaker function. In the limit $\xi = \frac{1}{6}$, $m = 0$, this is the only contribution to the energy density and it reduces to

$$\rho = \frac{1}{2880\pi}\frac{(\alpha^2+\beta^2)^{\frac{5}{2}}}{\alpha^3 a^4}. \tag{5.126}$$

Since only condition (i) of (5.102) applies in this limit, the energy density given by (5.126) is valid for any choice of scale factor $a(\eta) = C^{\frac{1}{2}}(\eta)$, allowing various cosmological hypotheses to be tested (Birrell & Davies 1980*a*, see also §7.4).

6
Stress-tensor renormalization

In previous chapters the production of quanta by a changing gravitational field was studied in detail. It was pointed out that only in exceptional circumstances does the particle concept in curved space quantum field theory correspond closely to the intuitive physical picture of a subatomic particle. In general, no natural definition of particle exists, and particle detectors will respond in a variety of ways that bear no simple relation to the usual conception of the quantity of matter present.

For some purposes it is advantageous to study the expectation values of other observables. Part of the problem with the particle concept concerns the fact that it is defined globally, in terms of field modes, and so is sensitive to the large scale structure of spacetime. In contrast, physical detectors are at least quasi-local in nature. It therefore seems worthwhile to investigate physical quantities that are defined locally, i.e., at a spacetime point, rather than globally. One such object of interest is the stress–energy–momentum (or stress) tensor, $T_{\mu\nu}(x)$, at the point x. In addition to describing part of the physical structure of the quantum field at x, the stress-tensor also acts as the source of gravity in Einstein's field equation. It therefore plays an important part in any attempt to model a self-consistent dynamics involving the gravitational field coupled to the quantum field. For many investigators, especially astrophysicists, it is this back-reaction of the quantum processes on the background geometry that is of primary interest.

In this chapter, we shall present the formalism necessary for the calculation of $\langle T_{\mu\nu} \rangle$ for a variety of quantum fields. It is a subject that involves some unusual subtleties, and requires delicate handling. Much of the chapter is devoted to a careful discussion of several procedures, known as regularization techniques, for computing a finite, renormalized $\langle T_{\mu\nu} \rangle$ from an apparently meaningless infinite quantity. While the reader is likely to encounter any one of these procedures in the published literature, with some notable exceptions, the end results are independent of which particular method of regularization is employed. We present all the major techniques, both for completeness and to demonstrate their mutual consistency and interrelation. It is not necessary, however, to master each individual method to understand the later sections, where we apply the results to some background spacetimes of special interest.

In much of our treatment we work with the action rather than $T_{\mu\nu}$, because this permits a more elegant approach, especially when combined with complex analytic methods. However, the treatment is very formal, and often involves the manipulation of quantities that are not obviously well-defined for the problem under study. The reader may well feel uneasy with such unfamiliar quantities as $\ln G_F$, and prefer the more 'nuts and bolts' approach involving $\langle T_{\mu\nu}\rangle$ directly, especially when combined with the non-analytic regularization techniques.

Along with chapter 3, this chapter is fundamental to the subject of the book. It is considerably longer and in some ways more technical than the other chapters. On a first reading we recommend that attention be mainly directed to the dimensional continuation method of regularization, and the topic of conformal anomalies. Then, after gaining familiarity with the application of dimensional regularization to renormalization, and having studied some specific examples, such as de Sitter space in §6.4, the reader will want to study the other techniques. In particular, point-splitting and adiabatic regularization offer the best starting point for those wishing to embark upon actual calculations involving explicit models. This chapter contains all the basic information necessary to begin such calculations.

The final section of the chapter can be read profitably even without a full understanding of regularization and renormalization theory.

Early discussion of regularization in curved space quantum field theory can be found in the work of Utiyama & DeWitt (1962), and Halpern (1967). Some physical conjectures were made by Sakharov (1967) and followed up by Grib & Mamaev (1969).

A somewhat different formalism from that described here has been given by Kibble & Randjbar-Daemi (1980) and Randjbar-Daemi, Kay & Kibble (1980) within the framework of Kibble's (1978) nonlinear generalization of quantum mechanics.

6.1 The fundamental problem

In §2.4 it was pointed out that the expectation value of H, even in the Minkowski vacuum state, is infinite. Again in §§4.1–4.3 we found that $\langle 0|T_{\mu\nu}|0\rangle$ was ultraviolet divergent. This behaviour is symptomatic of the problems that afflict any attempt to evaluate the expectation values of an operator that is quadratic in the field strength. For example, $\langle 0|\phi^2(x)|0\rangle$ may be obtained from the limit as $x'\to x$ of $G^{(1)}(x,x')$ (see (2.66)). But inspection of the DeWitt–Schwinger expansion (3.141) reveals that, quite generally, G_F (and hence $G^{(1)}$) diverges like $\sigma^{(2-n)/2}$ (or $\ln\sigma$ for $n=2$) as $\sigma\to 0$.

In Minkowski space quantum field theory, the divergence is simply discarded, for example by the use of normal ordering (see (2.41)). When the topology is non-trivial, but the geometry still flat, as in §§4.1.–4.2, one may employ an ultraviolet regulator function $e^{-\alpha|k|}$, to cut off the ultraviolet divergence, and then take the difference between $\langle T_{\mu\nu}\rangle$ in the topology of interest and its (cut-off) value in Minkowski space, letting $\alpha\to 0$ at the end of the calculation. Alternatively, in the Green function approach, the unbounded Minkowski space expression for $G^{(1)}(x,x')$ is first subtracted from the $G^{(1)}$ function evaluated in the topology of interest, and only after this manoeuvre is the limit $x\to x'$ taken.

There are two reasons why these simple devices cannot be trusted when spacetime is curved. The first concerns the rôle of $T_{\mu\nu}$ in gravity theory. In non-gravitational physics, only energy differences are observable, so that an infinite vacuum energy causes little embarrassment – one simply renormalizes the zero point by an infinite amount. When gravity is taken into account, however, this is not satisfactory. Energy is a source of gravity, and will bring about the very spacetime curvature whose effects we are trying to study. It cannot simply be thrown away; we are not free to rescale the zero point of energy. Instead, a more elaborate renormalization scheme is required involving the dynamics of the gravitational field.

A second reason that we run into trouble can be illustrated by a simple example. Consider the spatially flat Robertson–Walker spacetime described by the scale factor

$$a(t)=(1-A^2t^2)^{\frac{1}{2}},\quad A \text{ constant} \tag{6.1}$$

which has physical singularities at $t=\pm A^{-1}$.

In four dimensions the massless scalar field satisfying the wave equation

$$\Box\phi=0 \tag{6.2}$$

is not conformally coupled, but mode solutions are easily found (Bunch & Davies 1978*b*)

$$u_{\mathbf{k}}=(16\pi^3)^{-\frac{1}{2}}C^{-\frac{1}{2}}(\eta)(k^2+A^2)^{-\frac{1}{4}}\exp[\mathrm{i}\mathbf{k}\cdot\mathbf{x}-\mathrm{i}(k^2+A^2)^{\frac{1}{2}}\eta] \tag{6.3}$$

where $C(\eta)=a^2(\eta)=\cos^2 A\eta$ and η is the conformal time.

A Fock space may be constructed from the modes (6.3), and the vacuum state $|0\rangle$ used to evaluate $\langle 0|T_{\mu\nu}|0\rangle$. From (3.190) we find for the minimally coupled case $\xi = 0$

$$T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2}g_{\mu\nu}g^{\sigma\rho}\phi_{,\sigma}\phi_{,\rho} \tag{6.4}$$

which is of the same form as the flat space expression (2.26) with $m = 0$. Using the mode formula (2.43) (which is valid in curved space), one obtains for the energy density

$$\langle 0|T_0{}^0|0\rangle = (1/32\pi^3 C^2)\int d^3k[(k^2 + A^2)^{\frac{1}{2}} + (k^2 + \tfrac{1}{4}D^2)(k^2 + A^2)^{-\frac{1}{2}}] \tag{6.5}$$

where $D(\eta) = C^{-1}\partial C/\partial\eta$.

The integral diverges quartically as expected. If an ultraviolet cut-off factor $\exp[-\alpha(k^2 + A^2)^{\frac{1}{2}}]$ is introduced into (6.5), the integrals may be performed in terms of MacDonald functions. Expanding in powers of α, one readily obtains

$$\rho a^4 = (32\pi^2)^{-1}[48/\alpha^4 + (D^2 - 8A^2)/\alpha^2 + A^2(\tfrac{1}{2}D^2 - A^2)\ln\alpha] + O(\alpha^0) \tag{6.6}$$

where we have written the energy density as ρ and C^2 as a^4. The left-hand side of (6.6) represents the energy of massless radiation in a volume a^3, which redshifts as the universe expands.

To recover the Minkowski space result from (6.6) one may put $a = 1$ and $D = A = 0$. Only the first term on the right-hand side of (6.6) remains. It follows that in curved spacetime even if this term is discarded, ρa^4 still diverges in the limit $\alpha \to 0$. That is, the difference in zero-point energy between the Robertson–Walker universe and Minkowski space is still infinite. We cannot cure the divergence in $\langle 0|T_{00}|0\rangle$ simply by discarding a Minkowski-type term.

Inspection of (6.6) reveals that in addition to the expected quartic divergence, we have to contend with quadratic and logarithmic divergent terms also. Evidently the control of infinities in $\langle T_{\mu\nu}\rangle$ will involve a considerably more elaborate procedure than in the case of flat spacetime.

How are we to make physical sense out of a divergent result? To obtain a finite answer will obviously entail the subtraction of infinite quantities, but this can be done in an infinite variety of ways. Some additional criteria must be imposed in order that a unique answer be obtained. Divergences also plague Minkowski space quantum field theory, especially when field interactions are permitted. In quantum electrodynamics, infinite subtractions may be carried out systematically to yield finite results that are in good agreement with experiment, provided the subtractions are performed covariantly (see, for example, Schwinger 1951*a*, Valatin 1954*a, b, c*). Therefore we should endeavour to maintain general coordinate invariance when handling the divergences of $\langle T_{\mu\nu}\rangle$. In addition to maintaining general covariance, one might also require $\langle T_{\mu\nu}\rangle$ to possess a certain number of 'physically reasonable' properties. If enough such restrictions are imposed on $\langle T_{\mu\nu}\rangle$, then the subtraction procedure might be defined uniquely. In §6.6 we shall see that such an approach can indeed be implemented.

An alternative strategy is to treat the computation of $\langle T_{\mu\nu}\rangle$ as part of a wider dynamical theory involving gravity. In the semiclassical theory being considered in this book, the gravitational field is treated classically, while the matter fields (including the graviton to one-loop level – see chapter 1) are treated quantum mechanically. This is reminiscent of the successful semiclassical theory of electrodynamics, where the classical electromagnetic field is coupled to the *expectation value* of the electric current operator. By analogy, we seek a theory based on Einstein's field equation

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = -8\pi G T_{\mu\nu}, \tag{6.7}$$

but with the stress-tensor source regarded now as a quantum expectation value:

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \Lambda_{\mathrm{B}} g_{\mu\nu} = -8\pi G_{\mathrm{B}}\langle T_{\mu\nu}\rangle. \tag{6.8}$$

The reason for the subscripts B on the cosmological constant Λ and Newton's constant G, will be made clear in due course. In this chapter we shall explicitly incorporate factors of G, which elsewhere are set equal to one.

The classical Einstein equation (6.7) can be derived from the action

$$S = S_{\mathrm{g}} + S_{\mathrm{m}} \tag{6.9}$$

by the condition

$$\frac{2}{(-g)^{\frac{1}{2}}}\frac{\delta S}{\delta g^{\mu\nu}} = 0. \tag{6.10}$$

The first term on the right of (6.9) is the gravitational action

$$S_{\mathrm{g}} = \int L_{\mathrm{g}}(-g)^{\frac{1}{2}}\,\mathrm{d}^n x = \int (-g)^{\frac{1}{2}}(16\pi G_{\mathrm{B}})^{-1}(R - 2\Lambda)\mathrm{d}^n x \tag{6.11}$$

for which $2(-g)^{-\frac{1}{2}}\delta S_g/\delta g^{\mu\nu}$ yields the left-hand side of (6.7). The second term in (6.9) is the classical matter action, for which (cf. (3.189))

$$\frac{2}{(-g)^{\frac{1}{2}}}\frac{\delta S_{\mathrm{m}}}{\delta g^{\mu\nu}} = T_{\mu\nu}, \tag{6.12}$$

yields the right-hand side of (6.7).

How can this procedure work in the semiclassical case (6.8)? We seek a quantity W, called the *effective action* for the quantum matter fields, which, when functionally differentiated, yields $\langle T_{\mu\nu}\rangle$:

$$\frac{2}{(-g)^{\frac{1}{2}}}\frac{\delta W}{\delta g^{\mu\nu}} = \langle T_{\mu\nu}\rangle, \tag{6.13}$$

where the precise meaning of $\langle\ \rangle$ used here on $T_{\mu\nu}$ will be elucidated below.

To discover the structure of W, let us return to first principles, recalling the path-integral quantization procedure outlined in §2.8. Our notation will imply a treatment for the scalar field, but the formal manipulations are identical for fields of higher spins. In §2.8, the generating functional

$$Z[J] = \int \mathscr{D}[\phi]\exp\left\{\mathrm{i}S_{\mathrm{m}}[\phi] + \mathrm{i}\int J(x)\phi(x)\mathrm{d}^n x\right\} \tag{6.14}$$

was interpreted physically as the vacuum persistence amplitude $\langle \mathrm{out}, 0|0, \mathrm{in}\rangle$. The presence of the external current J can cause the initial vacuum state $|0, \mathrm{in}\rangle$ to be unstable, i.e., it can bring about the production of particles. In flat space, in the limit $J = 0$, no particles are produced, and we have the normalization condition

$$Z[0] \equiv \langle \mathrm{out}, 0|0, \mathrm{in}\rangle_{J=0} = \langle 0|0\rangle = 1. \tag{6.15}$$

However, when spacetime is curved, we have seen that, in general, $|0, \mathrm{out}\rangle \neq |0, \mathrm{in}\rangle$, even in the absence of source currents J. Hence (6.15) will no longer apply.

Path-integral quantization still works in curved spacetime; one simply treats S_{m} in (6.14) as the curved spacetime matter action, and $J(x)$ as a current density (a scalar density in the case of scalar fields). One can thus set $J = 0$ in (6.14) and examine the variation of $Z[0]$:

$$\begin{aligned}\delta Z[0] &= \mathrm{i}\int \mathscr{D}[\phi]\delta S_{\mathrm{m}}\mathrm{e}^{\mathrm{i}S_{\mathrm{m}}[\phi]}\\ &= \mathrm{i}\langle \mathrm{out}, 0|\delta S_{\mathrm{m}}|0, \mathrm{in}\rangle,\end{aligned} \tag{6.16}$$

which is a statement of Schwinger's variational principle (Schwinger 1951*b*). From (6.16) and (6.12) one immediately obtains

$$\frac{2}{(-g)^{\frac{1}{2}}}\frac{\delta Z[0]}{\delta g^{\mu\nu}} = \mathrm{i}\langle \mathrm{out}, 0|T_{\mu\nu}|0, \mathrm{in}\rangle. \tag{6.17}$$

Noting that the matter action S_{m} appears exponentiated in (6.14), we may therefore identify

$$Z[0] = e^{iW}, \tag{6.18}$$

whence,

$$W = -i \ln \langle \text{out}, 0 | 0, \text{in} \rangle \tag{6.19}$$

and, from (6.17),

$$\frac{2}{(-g)^{\frac{1}{2}}} \frac{\delta W}{\delta g^{\mu\nu}} = \frac{\langle \text{out}, 0 | T_{\mu\nu} | 0, \text{in} \rangle}{\langle \text{out}, 0 | 0, \text{in} \rangle} \tag{6.20}$$

The functional $Z[0]$ is evaluated in much the same way as in flat spacetime (see page 30). The main differences arise from (i) the replacement of the measure $d^n x$ by the covariant measure $d^n x[-g(x)]^{\frac{1}{2}}$; (ii) the replacement of the identity $\delta^n(x-y)$ by $\delta^n(x-y)[-g(y)]^{-\frac{1}{2}}$ for which

$$\int d^n x[-g(x)]^{\frac{1}{2}} \delta^n(x-y)[-g(y)]^{-\frac{1}{2}} = 1;$$

(iii) the replacement of (2.120) by

$$K_{xy} = (\Box_x + m^2 - i\varepsilon + \xi R)\delta^n(x-y)[-g(y)]^{-\frac{1}{2}} \tag{6.21}$$

which, using

$$\int d^n y[-g(y)]^{\frac{1}{2}} K_{xy} K_{yz}^{-1} = \delta(x-z)[-g(z)]^{-\frac{1}{2}} \tag{6.22}$$

and (3.49), implies

$$K_{xz}^{-1} = -G_F(x, z) \tag{6.23}$$

(cf. (2.123)). Following through steps similar to those on page 30 one finds (as in flat spacetime)

$$Z[0] \propto [\det(-G_F)]^{\frac{1}{2}} \tag{6.24}$$

where the proportionality constant is metric-independent and can be ignored. Thus we arrive at

$$W = -i \ln Z[0] = -\tfrac{1}{2} i \operatorname{tr}[\ln(-G_F)]. \tag{6.25}$$

In (6.25) G_F is to be interpreted as an operator which acts on a space of vectors $|x\rangle$, normalized by

$$\langle x | x' \rangle = \delta^n(x - x')[-g(x)]^{-\frac{1}{2}}, \tag{6.26}$$

in such a way that

$$G_F(x, x') = \langle x|G_F|x'\rangle. \tag{6.27}$$

The trace of an operator M which acts in this space, is defined by

$$\operatorname{tr} M = \int d^n x[-g(x)]^{\frac{1}{2}} M_{xx} = \int d^n x[-g(x)]^{\frac{1}{2}} \langle x|M|x\rangle. \tag{6.28}$$

To make sense of the formal expression (6.23) we must use a representation for the Feynman Green function G_F. We shall use the DeWitt–Schwinger representation given by the proper time integral (3.138). Writing the operator equivalent of (6.23) as

$$G_F = -K^{-1} = -\mathrm{i}\int_0^\infty e^{-\mathrm{i}Ks} ds, \tag{6.29}$$

we have from (3.138) the useful result

$$\langle x|e^{-\mathrm{i}Ks}|x'\rangle = \mathrm{i}(4\pi)^{-n/2}\Delta^{\frac{1}{2}}(x, x')e^{-\mathrm{i}m^2 s + \sigma/2\mathrm{i}s}F(x, x'; \mathrm{i}s)(\mathrm{i}s)^{-n/2}. \tag{6.30}$$

Now, assuming K to have a small negative imaginary part,

$$\int_\Lambda^\infty e^{-\mathrm{i}Ks}(\mathrm{i}s)^{-1}\mathrm{i}ds = -\mathrm{Ei}(-\mathrm{i}\Lambda K) \tag{6.31}$$

where Ei is the exponential integral function, having for small values of its argument the expansion

$$\mathrm{Ei}(x) = \gamma + \ln(-x) + O(x), \tag{6.32}$$

γ being Euler's constant. Substituting (6.32) into (6.31) and letting $\Lambda \to 0$ yields

$$\ln(-G_F) = -\ln(K) = \int_0^\infty e^{-\mathrm{i}Ks}(\mathrm{i}s)^{-1}\mathrm{i}ds, \tag{6.33}$$

which is correct up to the addition of a metric-independent (infinite) constant that can be ignored in what follows. Thus, in the DeWitt–Schwinger representation (6.30) (or (3.138))

$$\langle x|\ln(-G_F^{DS})|x'\rangle = -\int_{m^2}^\infty G_F^{DS}(x, x')dm^2, \tag{6.34}$$

where the integral with respect to m^2 brings down the extra power of $(\mathrm{i}s)^{-1}$ that appears in (6.33).

Returning to the expression (6.25) for W, (6.28) and (6.34) yield

$$W = \tfrac{1}{2}\mathrm{i}\int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}} \lim_{x'\to x} \int_{m^2}^{\infty} \mathrm{d}m^2 G_{\mathrm{F}}^{\mathrm{DS}}(x, x'). \tag{6.35}$$

Interchanging the order of integration and taking the limit $x' \to x$ one obtains

$$W = \tfrac{1}{2}\mathrm{i}\int_{m^2}^{\infty} \mathrm{d}m^2 \int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}} G_{\mathrm{F}}^{\mathrm{DS}}(x, x),$$

and the $\mathrm{d}^n x$ integral is seen to be precisely the expression for the one-loop Feynman diagram fig. 1*a* (see chapter 9 for Feynman rules). Thus W is known as the *one-loop effective action.* In the case of fermion effective actions, there would be a remaining trace over spinorial indices.

From (6.35) we may define an *effective Lagrangian density* $\mathscr{L}_{\mathrm{eff}}$, by

$$W = \int \mathscr{L}_{\mathrm{eff}}(x)\mathrm{d}^n x \equiv \int [-g(x)]^{\frac{1}{2}} L_{\mathrm{eff}}(x)\mathrm{d}^n x \tag{6.36}$$

whence

$$L_{\mathrm{eff}}(x) = [-g(x)]^{-\frac{1}{2}} \mathscr{L}_{\mathrm{eff}}(x) = \tfrac{1}{2}\mathrm{i} \lim_{x'\to x} \int_{m^2}^{\infty} \mathrm{d}m^2 G_{\mathrm{F}}^{\mathrm{DS}}(x, x'). \tag{6.37}$$

Inspection of (3.137) and (3.138) shows that L_{eff} diverges at the lower end of the s integral because the $\sigma/2s$ damping factor in the exponent vanishes in the limit $x' \to x$. (Convergence at the upper end is guaranteed by the $-\mathrm{i}\varepsilon$ that is implicitly added to m^2 in the DeWitt–Schwinger representation of G_{F}.) In four dimensions, the potentially divergent terms in the DeWitt–Schwinger expansion of L_{eff} are

$$L_{\mathrm{div}} = -\lim_{x'\to x} \frac{\Delta^{\frac{1}{2}}(x, x')}{32\pi^2} \int_0^{\infty} \frac{\mathrm{d}s}{s^3} \mathrm{e}^{-\mathrm{i}(m^2 s - \sigma/2s)} [a_0(x, x') + a_1(x, x')\mathrm{i}s + a_2(x, x')(\mathrm{i}s)^2] \tag{6.38}$$

where the coefficients a_0, a_1 and a_2 are given by (3.131)–(3.133). The remaining terms in this asymptotic expansion, involving a_3 and higher, are finite in the limit $x' \to x$.

The divergences in L_{eff} are, of course, the same as those that afflict $\langle T_{\mu\nu} \rangle$. Inspection of (3.131)–(3.133) in the limit $x' \to x$ reveals that the term in square brackets in (6.38) is entirely geometrical, i.e., it is built out of local tensors, $R_{\mu\nu\sigma\tau}$, and its contractions. This feature is readily understood. The divergences arise because of the ultraviolet behaviour of the field modes. These short wavelengths, however, only probe the local geometry in the neighbourhood of x – they are not sensitive to the global features of the spacetime, such as the topology or past behaviour. They are also independent of the quantum state employed, as we shall see below.

Because L_{div} is purely geometrical, it is better regarded as a contribution to the *gravitational* rather than the quantum matter Lagrangian. Although it arises from the action of the quantum matter field, it behaves as a quantity constructed solely from the gravitational field (the metric). Of course this will not be true of the remaining, finite portions of L_{eff}, which include the long wavelength part. This can probe the large scale structure of the manifold, and is also sensitive to the quantum state.

6.2 Renormalization in the effective action

Let us determine the precise form of the geometrical L_{div} terms, to compare them with the conventional gravitational Lagrangian L_g that appears in (6.11). This is a delicate matter because (6.38) is, of course, infinite. What we require is to display the divergent terms in the form $\infty \times$ geometrical object. This can be done in a variety of ways.

For example, in n dimensions, the asymptotic (adiabatic) expansion of L_{eff} is

$$L_{\text{eff}} \approx \lim_{x' \to x} \frac{\Delta^{\frac{1}{2}}(x,x')}{2(4\pi)^{n/2}} \sum_{j=0}^{\infty} a_j(x,x') \int_0^{\infty} (\mathrm{i}s)^{j-1-n/2} \mathrm{e}^{-\mathrm{i}(m^2 s - \sigma/2s)} \mathrm{i}\,\mathrm{d}s \tag{6.39}$$

of which the first $\frac{1}{2}n + 1$ terms are divergent as $\sigma \to 0$. If n is treated as a variable which can be analytically continued throughout the complex plane, then we may take the $x' \to x$ limit

$$L_{\text{eff}} \approx \tfrac{1}{2}(4\pi)^{-n/2} \sum_{j=0}^{\infty} a_j(x) \int_0^{\infty} (\mathrm{i}s)^{j-1-n/2} \mathrm{e}^{-\mathrm{i}m^2 s} \mathrm{i}\,\mathrm{d}s \tag{6.40}$$

$$= \tfrac{1}{2}(4\pi)^{-n/2} \sum_{j=0}^{\infty} a_j(x)(m^2)^{n/2-j}\Gamma(j-n/2) \tag{6.41}$$

where $a_j(x) \equiv a_j(x,x)$.

In what follows we shall wish to retain the units of L_{eff} as (length)$^{-4}$, even when $n \neq 4$. It is therefore necessary to introduce an arbitrary mass scale μ and to rewrite (6.41) as

$$L_{\text{eff}} \approx \tfrac{1}{2}(4\pi)^{-n/2}(m/\mu)^{n-4} \sum_{j=0}^{\infty} a_j(x) m^{4-2j}\Gamma(j-n/2). \tag{6.42}$$

As $n \to 4$, the first three terms of (6.42) diverge because of poles in the Γ-functions:

$$\left.\begin{aligned}\Gamma\left(-\frac{n}{2}\right) &= \frac{4}{n(n-2)}\left(\frac{2}{4-n}-\gamma\right)+O(n-4)\\ \Gamma\left(1-\frac{n}{2}\right) &= \frac{2}{2-n}\left(\frac{2}{4-n}-\gamma\right)+O(n-4)\\ \Gamma\left(2-\frac{n}{2}\right) &= \frac{2}{4-n}-\gamma+O(n-4).\end{aligned}\right\} \tag{6.43}$$

Calling these first three terms L_{div}, we have (see Bunch 1979)

$$L_{\text{div}} = -(4\pi)^{-n/2}\left\{\frac{1}{n-4}+\frac{1}{2}\left[(\gamma+\ln\left(\frac{m^2}{\mu^2}\right)\right]\right\}\left(\frac{4m^4a_0}{n(n-2)}-\frac{2m^2a_1}{n-2}+a_2\right) \tag{6.44}$$

where we have used the expansion

$$(m/\mu)^{n-4} = 1+\tfrac{1}{2}(n-4)\ln(m^2/\mu^2)+O((n-4)^2) \tag{6.45}$$

and dropped terms in (6.44) that vanish when $n \to 4$.

The functions a_0, a_1 and a_2 are given by taking the coincidence limits of (3.131)–(3.133) and (3.129):

$$a_0(x) = 1 \tag{6.46}$$

$$a_1(x) = (\tfrac{1}{6}-\xi)R \tag{6.47}$$

$$a_2(x) = \frac{1}{180}R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta} - \frac{1}{180}R^{\alpha\beta}R_{\alpha\beta} - \tfrac{1}{6}(\tfrac{1}{5}-\xi)\Box R + \tfrac{1}{2}(\tfrac{1}{6}-\xi)^2R^2. \tag{6.48}$$

It is now evident that L_{div} as given by (6.44) is a purely geometrical expression. At this stage we recall that L_{eff} is only part of the total Lagrangian. There is also a gravitational part. Because L_{div} is purely geometrical, we can try to absorb it into the gravitational Lagrangian.

Using (6.11), the total gravitational Lagrangian density now becomes $(-g)^{\frac{1}{2}}$ multiplied by

$$-\left(A+\frac{\Lambda_B}{8\pi G_B}\right)+\left(B+\frac{1}{16\pi G_B}\right)R-\frac{a_2(x)}{(4\pi)^{n/2}}\left\{\frac{1}{n-4}+\frac{1}{2}\left[\gamma+\ln\left(\frac{m^2}{\mu^2}\right)\right]\right\} \tag{6.49}$$

where

$$A = \frac{4m^4}{(4\pi)^{n/2}n(n-2)}\left\{\frac{1}{n-4}+\frac{1}{2}\left[\gamma+\ln\left(\frac{m^2}{\mu^2}\right)\right]\right\}$$

and

$$B = \frac{2m^2(\frac{1}{6} - \xi)}{(4\pi)^{n/2}(n-2)} \left\{ \frac{1}{n-4} + \frac{1}{2}\left[\gamma + \ln\left(\frac{m^2}{\mu^2}\right) \right] \right\}.$$

The first term in (6.49) is a constant. The contribution from L_{div}, i.e., the term A, is indistinguishable physically from Λ_{B}. This is the part of the gravitational Lagrangian that gives rise to the so-called cosmological term $\Lambda g_{\mu\nu}$ in the gravitational field equations (6.8). Hence, the effect of the scalar quantum field is to change, or renormalize, the cosmological constant from Λ_{B} to

$$\Lambda \equiv \Lambda_{\mathrm{B}} + \frac{32\pi m^4 G_{\mathrm{B}}}{(4\pi)^{n/2} n(n-2)} \left\{ \frac{1}{n-4} + \tfrac{1}{2}\left[\gamma + \ln\left(\frac{m^2}{\mu^2}\right) \right] \right\}. \tag{6.50}$$

Because a physical observation will only yield the renormalized value, Λ, we need not ask about the value of Λ_{B}, nor worry about the fact that, when we finally let $n \to 4$, the term in curly brackets in (6.50) diverges: we never see this term in isolation. This technique of absorbing an infinite quantity into a renormalized physical quantity is familiar in quantum field theory. For example, in quantum electrodynamics, an electron is 'dressed' in a cloud of virtual photons that contribute (infinitely) to the total mass of the electron. We only measure the renormalized electron mass (which is, of course, finite). The electron is inseparable from its photon cloud, so we never see the 'bare' mass. Adopting this terminology, we may say that the 'bare' cosmological constant, Λ_{B}, is never observed either. The appellation 'bare' prompts the use of the B subscript.

Turning to the second term in (6.49), we see that L_{div} also renormalizes Newton's gravitational constant, by changing G_{B} to

$$G = G_{\mathrm{B}}/(1 + 16\pi G_{\mathrm{B}} B). \tag{6.51}$$

The final term in (6.49) is not to be found in the usual Einstein Lagrangian. The factor $a_2(x)$ is of adiabatic order four, being of fourth order in derivatives of the metric (see (6.48)), and so represents a higher order correction to the general theory of relativity, which only contains terms with up to second derivatives of the metric. When this extra term is inserted into S_g, the left-hand side of the field equation becomes modified to

$$R_{\mu\nu} - \tfrac{1}{2}R g_{\mu\nu} + \Lambda g_{\mu\nu} + \alpha\,{}^{(1)}H_{\mu\nu} + \beta\,{}^{(2)}H_{\mu\nu} + \gamma H_{\mu\nu} \tag{6.52}$$

where

$${}^{(1)}H_{\mu\nu} \equiv \frac{1}{(-g)^{\frac{1}{2}}} \frac{\delta}{\delta g^{\mu\nu}} \int (-g)^{\frac{1}{2}} R^2 \mathrm{d}^n x = 2R_{;\mu\nu} - 2g_{\mu\nu}\Box R - \tfrac{1}{2}g_{\mu\nu}R^2 + 2RR_{\mu\nu} \tag{6.53}$$

$$
\begin{aligned}
{}^{(2)}H_{\mu\nu} &\equiv \frac{1}{(-g)^{\frac{1}{2}}}\frac{\delta}{\delta g^{\mu\nu}}\int(-g)^{\frac{1}{2}}R^{\alpha\beta}R_{\alpha\beta}\mathrm{d}^n x \\
&= R_{;\mu\nu} - \tfrac{1}{2}g_{\mu\nu}\Box R - \Box R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R^{\alpha\beta}R_{\alpha\beta} + 2R^{\alpha\beta}R_{\alpha\beta\mu\nu} \\
&= 2R_{\mu\ ;\nu\alpha}^{\ \alpha} - \Box R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}\Box R + 2R_\mu^{\ \alpha}R_{\alpha\nu} - \tfrac{1}{2}g_{\mu\nu}R^{\alpha\beta}R_{\alpha\beta}
\end{aligned}
\tag{6.54}
$$

$$
\begin{aligned}
H_{\mu\nu} &\equiv \frac{1}{(-g)^{\frac{1}{2}}}\frac{\delta}{\delta g^{\mu\nu}}\int(-g)^{\frac{1}{2}}R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta}\mathrm{d}^n x \\
&= -\tfrac{1}{2}g_{\mu\nu}R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} + 2R_{\mu\alpha\beta\gamma}R_\nu^{\ \alpha\beta\gamma} - 4\Box R_{\mu\nu} + 2R_{;\mu\nu} \\
&\quad - 4R_{\mu\alpha}R^\alpha_{\ \nu} + 4R^{\alpha\beta}R_{\alpha\mu\beta\nu}.
\end{aligned}
\tag{6.55}
$$

Note that in the special case $n = 4$, the generalized Gauss–Bonnet theorem (Chern 1955, 1962) states that

$$
\int \mathrm{d}^4x[-g(x)]^{\frac{1}{2}}(R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta} + R^2 - 4R_{\alpha\beta}R^{\alpha\beta}) \tag{6.56}
$$

is a topological invariant (called the Euler number), so that its metric variation will vanish identically. It then follows from (6.53)–(6.55) that

$$
H_{\mu\nu} = -{}^{(1)}H_{\mu\nu} + 4\,{}^{(2)}H_{\mu\nu}. \tag{6.57}
$$

The coefficients α, β and γ in (6.52) all contain $1/(n-4)$ and so diverge as n approaches the physical dimension 4. We must therefore introduce terms of adiabatic order 4 into the original gravitational Lagrangian with bare coefficients $a_{\mathrm{B}}, b_{\mathrm{B}}, c_{\mathrm{B}}$ into which the divergent terms involving α, β, γ can be absorbed to yield renormalized coefficients a, b, c. Because of (6.57), at the physical dimension $n = 4$, only two of these coefficients are independent, so we may choose $c = 0$. The values of a and b can only be determined by experiment. In order to avoid conflict with observation, it is necessary to assume that both a and b are very small numerically (see, for example, Stelle 1977, 1978, Horowitz & Wald 1978). In principle there is no reason why these renormalized quantities may not be set equal to zero, thus recovering Einstein's theory. Quantum field theory merely indicates that terms involving higher derivatives of the metric are *a priori* expected.

The device of allowing n to be continued analytically away from the physical dimension enables us to render the formally divergent L_{div} temporarily finite, so that its component pieces may be manipulated meaningfully and absorbed into the gravitational Lagrangian. The technique of altering a formally divergent expression for such purposes is called *regularization.* The approach used above is known as dimensional regularization and was first used in interacting quantum field theories in Minkowski space (Ashmore 1972, Bollini & Giambiagi 1972, 't Hooft & Veltman 1972). At the end of the calculation, when renormalization of the bare constants has been achieved, the regularization may be relaxed, e.g., $n \to 4$.

Once the terms L_{div} have been removed from L_{eff}, the remainder is finite and will be called the renormalized effective Lagrangian:

$$L_{\text{ren}} \equiv L_{\text{eff}} - L_{\text{div}}. \tag{6.58}$$

In four dimensions the asymptotic expansion of L_{ren} will consist of all terms with $j \geq 3$ in (6.40). Putting $x' = x$ and $n = 4$ we can write this asymptotic expansion as

$$L_{\text{ren}} \approx \frac{1}{32\pi^2} \int_0^\infty \sum_{j=3}^{\infty} a_j(x)(\text{i}s)^{j-3} \text{e}^{-\text{i}m^2 s} \text{i} \, \text{d}s \tag{6.59}$$

which may be integrated three times by parts to yield (see (3.140))

$$-\frac{1}{64\pi^2} \int_0^\infty \ln(\text{i}s) \frac{\partial^3}{\partial(\text{i}s)^3} [F(x, x; \text{i}s)\text{e}^{-\text{i}sm^2}] \text{d}(\text{i}s)$$

$$+\frac{1}{64\pi^2} \int_0^\infty \ln(\text{i}s) \frac{\partial^3}{\partial(\text{i}s)^3} \{[(a_0 + a_1(\text{i}s) + a_2(\text{i}s)^2]\text{e}^{-\text{i}sm^2}\} \text{i} \, \text{d}s. \tag{6.60}$$

The latter, finite, term simply renormalizes Λ, G, a, b and c by finite amounts, i.e., it is of the same form as L_{div}, involving constant $\times \, a_j$, $j = 1, 2, 3$. Obviously the renormalized effective Lagrangian will always be ambiguous up to terms of this type – finite renormalization terms – so we can drop the second term in (6.60).

For the same reason we need not worry about the choice of the mass scale μ introduced in (6.42). Rescaling μ changes L_{div} by a finite amount, but only by altering the coefficients of the geometrical terms a_0, a_1 and a_2. In practice, one would choose a fixed value of μ and use the results of one's calculations with this value of μ to calibrate the instruments used to measure the constants Λ, G, a and b. Once these constants have been measured, further calculations using this same value of μ and the measured values of the constants can be used to make predictions about the outcome of experiments using the previously calibrated instruments. If the value of μ is changed one must either recalibrate one's instruments or else change the values of Λ, G, a and b. The effect of either of these changes will leave invariant the predictions made about the outcome of experiments. Such an analysis of the rescaling of μ in interacting field theories in Minkowski space leads to a renormalization group equation ('t Hooft 1973, see also §9.2). Therefore we may write in place of (6.60)

$$L_{\text{ren}} = -\frac{1}{64\pi^2}\int_0^\infty \ln(\text{i}s)\frac{\partial^3}{\partial(\text{i}s)^3}[F(x,x;\text{i}s)\text{e}^{-\text{i}sm^2}]\text{i}\,\text{d}s \tag{6.61}$$

where it is understood that any finite multiple of a_0, a_1 or a_2 may be added to this expression. Having been derived from an asymptotic expansion for F, (6.61) cannot be regarded as the complete Lagrangian associated with the physical, renormalized $\langle T_{\mu\nu}\rangle$, the construction of which is deferred until §6.4. We display L_{ren} in the above form for completeness only. It is worth noting, however, that in principle the complete renormalized Lagrangian could be computed from (6.61) if the exact expression for F were available.

Besides dimensional regularization, other regularization techniques are available. Consider an eigenfunction expansion of the operator K^{-1}:

$$K^{-1} = -G_{\text{F}} = \sum_m \frac{|m\rangle\langle m|}{\lambda_m} \tag{6.62}$$

where

$$K|m\rangle = \lambda_m|m\rangle, \tag{6.63}$$

$$\langle n|m\rangle = \delta_{nm},$$

and

$$\sum_m |m\rangle\langle m| = 1. \tag{6.64}$$

Then

$$K^\nu|m\rangle = \lambda_m^\nu|m\rangle \tag{6.65}$$

so

$$(-G_{\text{F}})^\nu = \sum_m \lambda_m^{-\nu}|m\rangle\langle m| \tag{6.66}$$

whence

$$\text{tr}(-G_{\text{F}})^\nu = \int \text{d}^4x[-g(x)]^{\frac{1}{2}}\sum_m \lambda_m^{-\nu}\langle x|m\rangle\langle m|x\rangle = \sum_m \lambda_m^{-\nu}, \tag{6.67}$$

where we have used the completeness relation

$$\int \text{d}^4x[-g(x)]^{\frac{1}{2}}|x\rangle\langle x| = 1. \tag{6.68}$$

The right-hand side of (6.67) is reminiscent of Riemann's ζ-function, $\sum_{m=1}^{\infty} m^{-\nu}$, and may be used to define a *generalized ζ-function* (Dowker & Critchley 1976*a*, *b*, 1977*b*, Hawking 1977*a*, Gibbons 1977*a*):

$$\operatorname{tr}(-G_F)^{\nu} = \sum_m \lambda_m^{-\nu} \equiv \zeta(\nu). \tag{6.69}$$

We now wish to obtain the effective action (6.25) in terms of the generalized ζ-function. First we note that the argument of the logarithm in (6.25) should strictly be dimensionless. As $G_F(x, x')$ has dimensions $(\text{mass})^{n-2}$ and $|x\rangle$, from (6.26), has dimensions of $(\text{mass})^{n/2}$, we see from (6.27) that the operator G_F has dimensions of $(\text{mass})^{-2}$. We can thus make the argument of the logarithm in (6.25) dimensionless for all n by inserting a factor μ^2, writing

$$W = -\tfrac{1}{2}\mathrm{i}\operatorname{tr}[\ln(-\mu^2 G_F)]. \tag{6.70}$$

We have, in fact, already adjusted (6.24) by just such a constant in going to (6.25). The parameter μ plays a similar role in ζ-function regularization to the μ introduced in (6.42) as part of dimensional regularization. As before, changes in μ result only in finite changes to the inherently ambiguous coefficients of the renormalization terms.

The effective action (6.70) can be written in terms of the generalized ζ-function as

$$\begin{aligned} W &= -\tfrac{1}{2}\mathrm{i} \lim_{\nu \to 0} \operatorname{tr} \frac{\mathrm{d}}{\mathrm{d}\nu}(-\mu^2 G_F)^{\nu} \\ &= \lim_{\nu \to 0} \{-\tfrac{1}{2}\mathrm{i}\mu^{2\nu}[\zeta'(\nu) + \zeta(\nu)\ln \mu^2]\}. \end{aligned} \tag{6.71}$$

The introduction of the generalized ζ-function must be understood as a purely formal operation. In general (6.69) will not converge for all values of ν. However, it may be defined by analytic continuation from regions where it does converge. We then find that $\zeta(0)$ and $\zeta'(0)$ are finite, allowing us to write

$$W = -\tfrac{1}{2}\mathrm{i}[\zeta'(0) + \zeta(0)\ln \mu^2]. \tag{6.72}$$

Thus, by this formal analytic process, the divergences in the effective action W have been eliminated. It is perhaps worth comparing this result with the corresponding result for the usual Riemann ζ-function: The series $\sum_{m=1}^{\infty} m^{-\nu}$ is clearly divergent if ν is set equal to zero, yet Riemann's ζ-function, defined by this series for $\nu > 1$, when analytically continued to $\nu = 0$, gives the finite value $-\tfrac{1}{2}$.

To show that $\zeta(0)$ and $\zeta'(0)$ appearing in (6.72) are finite, we shall evaluate them explicitly in terms of the DeWitt–Schwinger proper time representation. We first note that

$$\int_0^\infty (\mathrm{i}s)^{\nu-1}\mathrm{e}^{-\mathrm{i}Ks}\mathrm{d}(\mathrm{i}s) = K^{-\nu}\Gamma(\nu), \tag{6.73}$$

from which

$$(-G_\mathrm{F})^\nu = K^{-\nu} = [\Gamma(\nu)]^{-1}\int_0^\infty (\mathrm{i}s)^{\nu-1}\mathrm{e}^{-\mathrm{i}Ks}\mathrm{i}\,\mathrm{d}s. \tag{6.74}$$

Now using (6.69) and (6.30) one obtains in four dimensions

$$\zeta(\nu) = \mathrm{i}[\Gamma(\nu)]^{-1}(4\pi)^{-2}\int \mathrm{d}^4x[-g(x)]^{\frac{1}{2}}\int_0^\infty (\mathrm{i}s)^{\nu-3}\mathrm{e}^{-\mathrm{i}m^2s}F(x,x;\mathrm{i}s)\mathrm{i}\,\mathrm{d}s. \tag{6.75}$$

If Re $\nu > 2$, we can perform three integrations by parts. Recalling that m^2 is understood to mean $m^2 - \mathrm{i}\varepsilon$, one finds that the boundary terms vanish, to give

$$\zeta(\nu) = -\frac{\mathrm{i}(4\pi)^{-2}}{\Gamma(\nu+1)(\nu-1)(\nu-2)}\int \mathrm{d}^4x[-g(x)]^{\frac{1}{2}}\int_0^\infty (\mathrm{i}s)^\nu \frac{\partial^3}{\partial(\mathrm{i}s)^3} \times [F(x,x;\mathrm{i}s)\mathrm{e}^{-\mathrm{i}sm^2}]\mathrm{i}\,\mathrm{d}s, \tag{6.76}$$

from which one obtains

$$\zeta(0) = \mathrm{i}(4\pi)^{-2}\int \mathrm{d}^4x[-g(x)]^{\frac{1}{2}}[\tfrac{1}{2}m^4 - m^2a_1(x) + a_2(x)], \tag{6.77}$$

where we have used (3.140) to evaluate

$$\frac{\partial^2}{\partial(\mathrm{i}s)^2}[F(x,x;\mathrm{i}s)\mathrm{e}^{-\mathrm{i}sm^2}]_{s=0}.$$

Since $\zeta(0)$ only contains finite renormalization terms, it is now clear that changes of μ in (6.72) will only result in finite changes of the renormalized coefficients in the complete action. Thus, for our present purposes we can ignore the $\zeta(0)$ term in (6.72).

Differentiation of (6.76) with respect to ν yields

$$\zeta'(0) = \tfrac{1}{2}\mathrm{i}(4\pi)^{-2}\Bigg\{(\gamma - \tfrac{3}{2})\int \mathrm{d}^4x[-g(x)]^{\frac{1}{2}}[\tfrac{1}{2}m^4 - m^2a_1(x) + a_2(x)] - \int \mathrm{d}^4x[-g(x)]^{\frac{1}{2}}\int_0^\infty \ln(\mathrm{i}s)\frac{\partial^3}{\partial(\mathrm{i}s)^3}[F(x,x;\mathrm{i}s)\mathrm{e}^{-\mathrm{i}sm^2}]\mathrm{i}\,\mathrm{d}s\Bigg\}, \tag{6.78}$$

which is finite. The first integral in this expression only contains finite renormalization terms which once again can be ignored. Inserting the remaining term into (6.72) and using the definition (6.36), one obtains

$$L_{\mathrm{ren}} = -\frac{1}{64\pi^2}\int_0^\infty \ln(\mathrm{i}s)\frac{\partial^3}{\partial(\mathrm{i}s)^3}[F(x,x;\mathrm{i}s)\mathrm{e}^{-\mathrm{i}sm^2}]\mathrm{i}\,\mathrm{d}s, \tag{6.79}$$

$$= L_{\mathrm{eff}} + \text{finite renormalization terms}$$

which agrees with the result (6.61) obtained by dimensional regularization.

Note that the ζ-function technique described here does not require explicit infinite renormalization of coupling constants in the gravitational Lagrangian. (This is not the case, however, for the method adopted by Dowker & Critchley 1976*a*, *b*, 1977*b*.) The analytic continuation method converts a manifestly infinite series into a finite result (e.g., $\sum_{m=1}^{\infty} m^{-\nu}$ with $\nu = 0$ becomes $\frac{1}{2}$). Clearly, an infinite term has been tacitly discarded in this formal procedure.

It might therefore be supposed that variations of the technique would produce different answers. For example, if instead of (6.71), the logarithm in (6.70) were represented as follows

$$\lim_{\nu\to 0}[\nu^{-1}(-\mu^2 G_{\mathrm{F}})^\nu - \nu^{-1}]$$

then we would have

$$W = -\tfrac{1}{2}\mathrm{i}\lim_{\nu\to 0}\nu^{-1}\zeta(0) - \tfrac{1}{2}\mathrm{i}[\zeta'(0) + \zeta(0)\ln\mu^2] + c \tag{6.80}$$

where c is an infinite, metric-independent constant that can be safely discarded. Equation (6.80) differs from (6.72) only in the first term which, on inspection of (6.77), is seen to represent an infinite renormalization of the gravitational action. Thus, whether or not an infinite renormalization is carried out, the *renormalized* effective action is the same.

There is still an element of doubt about whether the renormalized effective action contains some ambiguity. Can one be sure that these renormalization procedures, involving questionable formal manipulations of divergent quantities, will always give the 'correct' answer? This issue can only be resolved by first deciding what physical criteria a 'correct' answer should satisfy, a subject to be investigated in §6.6. Secondly, a rigorous mathematical foundation for the formal manipulations given here is necessary. Such a treatment of flat space quantum field theory has been given by, for example, Taylor (1960, 1963) and Caianello (1973).

Another frequently employed regularization technique is called point-splitting, or point-separation. The basic idea is to return to (6.37), keeping x'

and x separated by an infinitesimal distance in a non-null direction. So long as $\sigma \neq 0$, L_{eff} remains finite, and the integral in (6.38) can be performed in terms of Hankel functions. (Alternatively, the first three terms of the sum in (3.138) can be substituted into (6.37) to give the same result). Expanding the Hankel functions in an asymptotic expansion in powers of σ and retaining only terms which do not vanish as $\sigma \to 0$, one obtains (Christensen 1978), in the four-dimensional case

$$\begin{aligned} L_{\text{div}} = \lim_{x' \to x} & (1/8\pi^2)\Delta^{\frac{1}{2}}(x,x')\{a_0(x,x')[\sigma^{-2} + \tfrac{1}{2}m^2\sigma^{-1} \\ & - \tfrac{1}{4}m^4 \times (\gamma + \tfrac{1}{2}\ln|\tfrac{1}{2}m^2\sigma|) + \tfrac{3}{16}m^4] \\ & - a_1(x,x')[\tfrac{1}{2}\sigma^{-1} - \tfrac{1}{2}m^2(\gamma + \tfrac{1}{2}\ln|\tfrac{1}{2}m^2\sigma|) + \tfrac{1}{4}m^2] \\ & - \tfrac{1}{2}a_2(x,x')[\gamma + \tfrac{1}{2}\ln|\tfrac{1}{2}m^2\sigma|]\}. \end{aligned} \tag{6.81}$$

The quantities $a_0(x,x')$, $a_1(x,x')$, $a_2(x,x')$ and $\Delta(x,x')$ may now be expanded in powers of σ. The leading term in the resulting expansion of (6.81) diverges like σ^{-2}. The coefficients in the expansion are all geometrical, involving the Riemann tensor and its derivatives, but a glance at (3.132) and (3.133) shows that when $x' \neq x$, these coefficients will contain vector quantities arising from the y^α factors. From (3.136), or general theory (DeWitt 1965), these vectors can be written in terms of σ as

$$y^\mu = \sigma^{;\mu} \equiv \sigma^\mu, \tag{6.82}$$

whence (3.136) reads

$$\sigma = \tfrac{1}{2}\sigma_\mu\sigma^\mu. \tag{6.83}$$

It is convenient to parametrize the strength of the divergences by one quantity ε, proportional to the geodesic distance between x' and x, and parametrize separately the (non-null) direction of splitting by a unit vector t^μ (see fig. 22). This is done by writing

$$\sigma^\mu = 2\varepsilon t^\mu \tag{6.84}$$

where

$$t^\mu t_\mu \equiv \Sigma = \pm 1, \tag{6.85}$$

depending on whether t^μ is timelike or spacelike respectively. Then (6.83) becomes

$$\sigma = 2\varepsilon^2\Sigma. \tag{6.86}$$

(The factor of 2 in (6.86) is introduced for later convenience.) In this

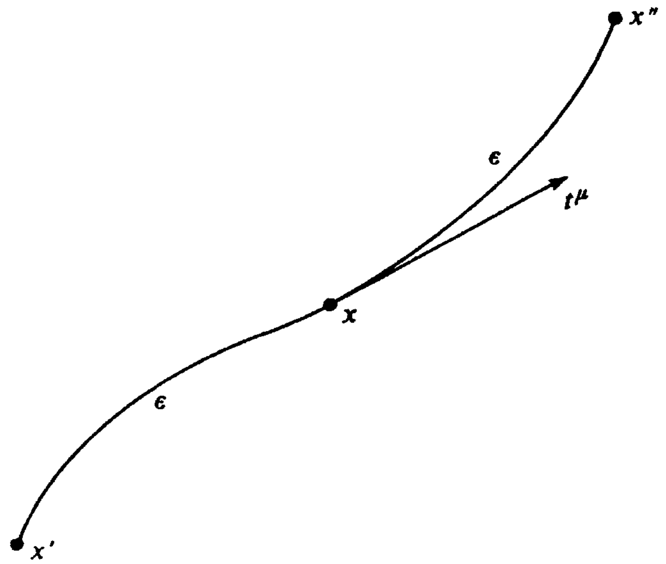


Fig. 22. The points x'', x' lie along a non-null geodesic through x, parametrized by t^μ, each a proper distance ε from x.

parametrization, the leading divergence in $L_{\rm div}$ behaves like ε^{-4}, and a typical direction-dependent coefficient will have the form

$$R_{\mu\nu}R_{\rho\sigma}t^\mu t^\nu t^\rho t^\sigma.$$

(Further details will be given in §6.4.) Such terms cannot be absorbed by renormalization, as the vector t^μ is a purely artificial construct and not part of the dynamics.

To rid the theory of direction-dependent terms, one may average over all directions using a suitable measure (Adler, Lieberman & Ng 1977). The troublesome terms then reduce to combinations of R, R^2, $R_{\alpha\beta}R^{\alpha\beta}$ and $R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta}$, which may be removed in the usual way by renormalization of Λ, G, a, b and c in the gravitational action. The finite remainder, $L_{\rm ren} = L_{\rm eff} - L_{\rm div}$, will obviously reproduce (6.59) and (6.61), and so agrees with both dimensional and ζ-function regularization. If we are interested in the field equations rather than the action, then no renormalization of c is necessary. This is because the point-splitting technique works throughout the computation in four dimensions, for which (6.56) is a topological invariant and hence does not contribute to the field equations when substituted into the functional derivative (6.10). It may therefore be used to eliminate one of the three constants a, b and c from the field equations.

It is instructive to examine further the close relation between the point separation and dimensional techniques of regularization. Both start from the DeWitt–Schwinger representation (6.39) in which the s integration can be performed (recall (3.141)). Consider the simplest case of all: Minkowski space. Then $a_0(x, x') = 1, a_j(x, x') = 0, j > 0$. Only the first term of the series contributes and one has from (3.141) and (6.37)

$$L_{\text{eff}} = \frac{\pi}{2(4\pi\mathrm{i})^{n/2}} \lim_{\sigma\to 0}\left(\frac{2}{-\sigma}\right)^{(n-2)/4} \int_{m^2}^{\infty} \mathrm{d}\bar{m}^2 (\bar{m}^2)^{(n-2)/4} H^{(2)}_{\frac{1}{2}n-1}[(2\bar{m}^2\sigma)^{\frac{1}{2}}]$$

$$= \lim_{\sigma\to 0} \frac{(-1)^{-(n+1)/2}\pi}{2(4\pi)^{n/2}} (2m^2)^{n/4} \sigma^{-n/4} H^{(2)}_{n/2}[(2m^2\sigma)^{\frac{1}{2}}]. \tag{6.87}$$

If dimensional regularization is used, then we want to be able to set $\sigma = 0$ in (6.87) and expand the result about $n = 4$. From the asymptotic forms

$$H^{(2)}_{\nu}(z) \xrightarrow[z\to 0]{} \begin{cases} (\mathrm{i}/\pi)\Gamma(\nu)(\frac{1}{2}z)^{-\nu}, & \operatorname{Re}\nu > 0 \quad (6.88a) \\ (\mathrm{i}/\pi)\mathrm{e}^{-\pi\nu\mathrm{i}}\Gamma(-\nu)(\frac{1}{2}z)^{\nu}, & \operatorname{Re}\nu < 0 \quad (6.88b) \end{cases}$$

it is clear that, for (6.87) to remain finite as $\sigma \to 0$, one must continue n to have negative real part, in which case

$$L_{\text{eff}} \xrightarrow[\sigma\to 0]{} \frac{(-1)^{-n}}{2(4\pi)^{n/2}} m^n \Gamma\left(-\frac{n}{2}\right) \xrightarrow[n\to 4]{} \frac{-m^4}{32\pi^2(n-4)}.$$

If, on the other hand, point-separation is used, then we want to be able to set $n = 4$ at the outset and use σ as the regulator. In this case (6.88*a*) must be used, and one obtains

$$L_{\text{eff}} \xrightarrow[\sigma\to 0]{} \frac{(-1)^{-n/2} 2^{\frac{1}{2}n-1}}{(4\pi)^{n/2}} \Gamma\left(\frac{n}{2}\right) \sigma^{-n/2} \xrightarrow[n\to 4]{} \frac{1}{8\pi^2\sigma^2}.$$

Note that in the massless case the divergent term vanishes completely in dimensional regularization and no renormalization at all is necessary. In contrast, in the point-separation expression, all m-dependence disappears so that L_{eff} still diverges like σ^{-2} in the massless limit. Renormalization is therefore necessary in this method.

In the case of two-dimensional spacetimes, inspection of (6.41) shows that for $n = 2$ only the $j = 0$ and 1 terms in the DeWitt–Schwinger series are potentially divergent. These terms may be removed by renormalizing Λ and G in the effective action as before. It should be noted that

$$\int \mathrm{d}^2x[-g(x)]^{\frac{1}{2}} R(x)$$

is a topological invariant in two dimensions, so that the variation of this term in the effective action gives no contribution to the two-dimensional 'Einstein equation'. Put another way, the two-dimensional Einstein tensor,

$R_{\mu\nu}-\frac{1}{2}Rg_{\mu\nu}$, vanishes identically. We must, however, include a term involving R in the effective action if we wish to remove all of the divergences by renormalization. The renormalized effective Lagrangian is (cf. (6.61))

$$L_{\text{ren}} = -\frac{1}{8\pi}\int_0^\infty \ln(\mathrm{i}s)\frac{\partial^2}{\partial(\mathrm{i}s)^2}\left[F(x,x;\mathrm{i}s)\mathrm{e}^{-\mathrm{i}sm^2}\right]\mathrm{d}(\mathrm{i}s). \tag{6.89}$$

In the case of higher-spin fields the above treatment goes through almost without change. For the spinor fields, the effective action can be written in terms of the bi-spinor G_{F} defined by (3.188):

$$W_{(\frac{1}{2})} = \tfrac{1}{2}\mathrm{i}\,\mathrm{tr}\left[\ln(-G_{\mathrm{F}})\right], \tag{6.90}$$

where the trace is now taken over spinor indices as well. The difference in sign between (6.90) and (6.25) is a direct result of the anticommuting nature of the spinor fields as opposed to the commuting properties of the scalar field. The case of the electromagnetic field is complicated by the presence of ghost fields (see §2.8). From the path-integral approach one finds

$$W_{\mathrm{EM}} = -\tfrac{1}{2}\mathrm{i}\,\mathrm{tr}\left[\ln(D_{\mathrm{F}})\right] + W_{\text{ghost}} \tag{6.91}$$

where D_{F} is defined by (3.187) and W_{ghost} is equal to -2 times the minimally coupled scalar effective action (the factor -2 being due to the fact that there are two, anticommuting, scalar ghost fields).

To evaluate the effective actions (6.90), (6.91) in terms of the DeWitt–Schwinger proper time expansions, one requires the appropriate coefficients a_i. These have been given by DeWitt (1965) and Christensen (1978), who also discusses in detail the point-separation renormalization of the effective actions. (In the electromagnetic case it is necessary to temporarily add a mass term to the Lagrangian for the DeWitt–Schwinger method to be useful.) We record here only the x, x' coincidence limits of the coefficients which occur in the divergent terms in four dimensions. For the spin $\frac{1}{2}$ case, the coefficients in the expansion of G_{F} in (3.188) and (6.90) are spinors:

$$\begin{aligned}
a_0(x) &= \mathbb{1}\\
a_1(x) &= -\tfrac{1}{12}R(x)\mathbb{1}\\
a_2(x) &= \left(\tfrac{1}{288}R^2 + \tfrac{1}{120}\Box R - \tfrac{1}{180}R^{\mu\nu}R_{\mu\nu} + \tfrac{1}{180}R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}\right)\mathbb{1}\\
&\quad + \tfrac{1}{48}\Sigma_{\mu\nu}\Sigma_{\rho\sigma}R^{\mu\nu\xi\lambda}R^{\rho\sigma}{}_{\xi\lambda},
\end{aligned} \tag{6.92}$$

where $\mathbb{1}$ is the unit spinor and

$$\Sigma_{\mu\nu} = \tfrac{1}{4}[\gamma_\mu, \gamma_\nu] = V_\mu{}^\alpha V_\nu{}^\beta \Sigma_{\alpha\beta},$$

γ_μ being the curved space gamma matrices defined on page 85, and $\Sigma_{\alpha\beta}$ being given by (3.166). For later use we record here the result of taking the trace over spinor indices of (6.92):

$$\begin{aligned}
\operatorname{tr} a_0(x) &= \operatorname{tr} \mathbb{1} \equiv \mathscr{s} \\
\operatorname{tr} a_1(x) &= -\tfrac{1}{12}\mathscr{s} R(x) \\
\operatorname{tr} a_2(x) &= \tfrac{1}{720}\mathscr{s}[\tfrac{5}{2}R^2 + 6\Box R - \tfrac{7}{2}R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} - 4R^{\mu\nu}R_{\mu\nu}],
\end{aligned} \tag{6.93}$$

where $\mathscr{s}$ is the number of spinor components (i.e., the dimension of the gamma matrices used). The coefficients in the expansions of $D_{F\mu\nu}$ of (3.187) and (6.91) are tensors, and are given in the Feynman gauge $\zeta = 1$ by

$$\begin{aligned}
a_{0\mu\nu}(x) &= g_{\mu\nu}(x) \\
a_{1\mu\nu}(x) &= \tfrac{1}{6}Rg_{\mu\nu} - R_{\mu\nu} \\
a_{2\mu\nu}(x) &= -\tfrac{1}{6}RR_{\mu\nu} + \tfrac{1}{6}\Box R_{\mu\nu} + \tfrac{1}{2}R_{\mu\rho}R^{\rho}{}_{\nu} - \tfrac{1}{12}R^{\lambda\sigma\rho}{}_{\mu}R_{\lambda\sigma\rho\nu} \\
&\quad + (\tfrac{1}{72}R^2 - \tfrac{1}{30}\Box R - \tfrac{1}{180}R^{\rho\sigma}R_{\rho\sigma} + \tfrac{1}{180}R^{\rho\sigma\lambda\omega}R_{\rho\sigma\lambda\omega})g_{\mu\nu}.
\end{aligned} \tag{6.94}$$

To summarize the conclusions of this rather technical section: The technique of renormalization enables us to work with an action for the coupled gravitational–quantum matter fields in the form

$$S = S_g + W$$

and to transfer divergent pieces of W into a suitably general S_g, absorbing the infinities into renormalized coupling constants. Thus

$$S = (S_g)_{\text{ren}} + W_{\text{ren}}$$

where $(S_g)_{\text{ren}}$ contains the renormalized, physical constants, and W_{ren} is now finite. Inserting S into (6.10) yields the semiclassical equation (in four dimensions)

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} + a^{(1)}H_{\mu\nu} + b^{(2)}H_{\mu\nu} = -8\pi G\frac{\langle \text{out}, 0|T_{\mu\nu}|0, \text{in}\rangle_{\text{ren}}}{\langle \text{out}, 0|0, \text{in}\rangle}, \tag{6.95}$$

where the right-hand side is now finite and Λ, a, b and G must be determined by measurement.

The renormalization did not make explicit use of $|0, \text{out}\rangle$ or $|0, \text{in}\rangle$. In particular, it is not necessary to assume that asymptotic in and out regions exist, where the vacuum concept is simple or corresponds closely to the concept of a physical vacuum. The in and out vacuum states enter here purely formally. The appearance of $\langle \text{out}, 0|T_{\mu\nu}|0, \text{in}\rangle$, rather than some other arrangement of vacuum states, is related to the boundary conditions implicit in the DeWitt–Schwinger representation of the Feynman propagator G_{F}. This is controlled by the use of the iε factor (see page 76). Frequently we are more interested in the quantity $\langle \text{in}, 0|T_{\mu\nu}|0, \text{in}\rangle$ (or perhaps $\langle \text{out}, 0|T_{\mu\nu}|0, \text{out}\rangle$). The effective action which yields these vacuum expectation values can be obtained by using (6.25) with G_{F} written in terms of the appropriate vacuum expectation value (e.g., for $\langle \text{in}, 0|T_{\mu\nu}|0, \text{in}\rangle$ use $G_{\mathrm{F}}(x, x') = -\mathrm{i}\langle \text{in}, 0|T(\phi(x)\phi(x'))|0, \text{in}\rangle$). Fortunately we do not need to embark upon a fresh discussion of renormalization for each case. The divergences arise from the short distance (high momentum) behaviour of the propagator (see §3.6), which is independent of global or state-dependent effects. Therefore one expects the divergences in all three above forms for $\langle T_{\mu\nu}\rangle$ to be the same (DeWitt 1975). This expectation is confirmed by using (3.45), (3.46) to write $|0, \text{out}\rangle$ in terms of many-particle in states. The quantity $\langle 1_{j_1}, 1_{j_2}, 1_{j_3}, \ldots, 1_{j_k}, \text{in}|T_{\mu\nu}|0, \text{in}\rangle$ may be evaluated by expanding $T_{\mu\nu}$ in creation and annihilation operators, and allowing them to act on $|0, \text{in}\rangle$. After some work one arrives at the result

$$\langle \text{in}, 0|T_{\mu\nu}|0, \text{in}\rangle = \frac{\langle \text{out}, 0|T_{\mu\nu}|0, \text{in}\rangle}{\langle \text{out}, 0|0, \text{in}\rangle} - \mathrm{i}\sum_{i,j} \Lambda_{ij} T_{\mu\nu}(u^*_{\text{in}, i}, u^*_{\text{in}, j}) \tag{6.96}$$

where $u_{\text{in}, i}$ are modes in the in region, and the notation in the final term is explained on page 17. A similar expression holds for $\langle \text{out}, 0|T_{\mu\nu}|0, \text{out}\rangle$ with Λ_{ij} replaced by V_{ij} from (3.46) and u_{in} replaced by u_{out}. The final term in (6.96) is finite. Hence the divergences present in $\langle \text{in}, 0|T_{\mu\nu}|0, \text{in}\rangle$ are the same as those present in $\langle \text{out}, 0|T_{\mu\nu}|0, \text{in}\rangle / \langle \text{out}, 0|0, \text{in}\rangle$.

6.3 Conformal anomalies and the massless case

Special interest attaches to field theories in which the classical action S is invariant under conformal transformations (see §3.1)

$$g_{\mu\nu}(x) \to \Omega^2(x) g_{\mu\nu}(x) = \bar{g}_{\mu\nu}(x). \tag{6.97}$$

From the definition of functional differentiation one has

$$S[\bar{g}_{\mu\nu}] = S[g_{\mu\nu}] + \int \frac{\delta S[\bar{g}_{\mu\nu}]}{\delta \bar{g}^{\rho\sigma}(x)} \delta \bar{g}^{\rho\sigma}(x)\, \mathrm{d}^n x, \tag{6.98}$$

which, using $\delta\bar{g}^{\mu\nu}(x) = -2\bar{g}^{\mu\nu}(x)\Omega^{-1}(x)\delta\Omega(x)$, and (6.12), gives

$$S[\bar{g}_{\mu\nu}] = S[g_{\mu\nu}] - \int[-\bar{g}(x)]^{\frac{1}{2}} T_\rho{}^\rho[\bar{g}_{\mu\nu}(x)]\Omega^{-1}(x)\delta\Omega(x)\mathrm{d}^n x. \tag{6.99}$$

From this equation one immediately obtains

$$T_\rho{}^\rho[g_{\mu\nu}(x)] = -\frac{\Omega(x)}{[-g(x)]^{\frac{1}{2}}}\left.\frac{\delta S[\bar{g}_{\mu\nu}]}{\delta\Omega(x)}\right|_{\Omega=1}, \tag{6.100}$$

and it is clear that if the classical action is invariant under the conformal transformations (6.97), then the classical stress-tensor is traceless. This is easily verified explicitly using (3.190), (3.191) and (3.193) for the massless scalar field with $\xi = \xi(n)$, the massless spin $\frac{1}{2}$ field and the four-dimensional electromagnetic field respectively. Because conformal transformations are essentially a rescaling of lengths at each spacetime point x, the presence of a mass and hence a fixed length scale in the theory will always break the conformal invariance. Therefore we are led to the massless limit of the regularization and renormalization procedures used in the previous section. This involves some delicate issues.

Although all the higher order ($j > 2$) terms in the DeWitt–Schwinger expansion of the effective Lagrangian (6.42) are infrared divergent at $n = 4$ as $m \to 0$, we can still use this expansion to yield the ultraviolet divergent terms arising from $j = 0$, 1, and 2 in the four-dimensional case. We may put $m = 0$ immediately in the $j = 0$ and 1 terms in the expansion, because they are of positive power for $n \sim 4$. These terms therefore vanish. The only non-vanishing potentially ultraviolet divergent term is therefore $j = 2$:

$$\tfrac{1}{2}(4\pi)^{-n/2}(m/\mu)^{n-4}a_2(x)\,\Gamma(2 - n/2), \tag{6.101}$$

which must be handled carefully.

Substituting for $a_2(x)$ with $\xi = \xi(n)$ from (6.48), and rearranging terms, we may write the divergent term in the effective action arising from (6.101) as follows

$$W_{\mathrm{div}} = \tfrac{1}{2}(4\pi)^{-n/2}(m/\mu)^{n-4}\Gamma(2 - n/2)\int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}}a_2(x) \tag{6.102}$$

$$= \tfrac{1}{2}(4\pi)^{-n/2}(m/\mu)^{n-4}\Gamma(2 - n/2)\int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}}[\alpha F(x)$$

$$+ \beta G(x)] + \mathrm{O}(n-4) \tag{6.103}$$

where

$$F = R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} - 2R^{\alpha\beta}R_{\alpha\beta} + \tfrac{1}{3}R^2 \tag{6.104}$$

and

$$G = R^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} - 4R^{\alpha\beta}R_{\alpha\beta} + R^2, \tag{6.105}$$

while the coefficients are

$$\alpha = \tfrac{1}{120}, \qquad \beta = -\tfrac{1}{360}. \tag{6.106}$$

In obtaining (6.103) we have dropped the $\Box R$ and R^2 terms from $a_2(x)$, the first because it is a total divergence, and so will not contribute to the action, the second because its coefficient is proportional to $(n-4)^2$ when the conformal coupling $\xi = \xi(n) = \frac{1}{4}(n-2)/(n-1)$ is inserted. In the limit $n \to 4$ this coefficient beats the $(n-4)^{-1}$ singularity from the Γ function in (6.43), causing this term to vanish.

The reason for decomposing a_2 into F and G is that in four dimensions (and only four) F is the square of the Weyl tensor $C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta}$. Moreover $\int(-g)^{\frac{1}{2}}G\,\mathrm{d}^4x$ is a topological invariant (see(6.56)). Both of these quantities remain invariant under conformal transformations. It follows that, at $n=4$, W_{div} in the massless conformally coupled limit is invariant under conformal transformations.

However, we must not relax the regularization and pass to $n=4$ before computing the physical quantities of interest, and away from $n=4$ W_{div} is *not* conformally invariant (though W is). We shall find that a vestige of this conformal breakdown survives in the physical quantities even when we put $n=4$ at the end of the calculation.

To see this, one may use the identities (Duff 1977)

$$\frac{2}{(-g)^{\frac{1}{2}}}g^{\mu\nu}\frac{\delta}{\delta g^{\mu\nu}}\int(-g)^{\frac{1}{2}}F\mathrm{d}^nx = -(n-4)(F - \tfrac{2}{3}\Box R) \tag{6.107}$$

$$\frac{2}{(-g)^{\frac{1}{2}}}g^{\mu\nu}\frac{\delta}{\delta g^{\mu\nu}}\int(-g)^{\frac{1}{2}}G\mathrm{d}^nx = -(n-4)G, \tag{6.108}$$

which enable the contribution of W_{div} to the trace of the stress-tensor to be evaluated immediately:

$$\langle T_\mu{}^\mu\rangle_{\mathrm{div}} = \frac{2}{(-g)^{\frac{1}{2}}}g^{\mu\nu}\frac{\delta W_{\mathrm{div}}}{\delta g^{\mu\nu}} = \tfrac{1}{2}(4\pi)^{-n/2}(m/\mu)^{n-4}(4-n)\Gamma(2-n/2)$$
$$\times[\alpha(F - \tfrac{2}{3}\Box R) + \beta G] + O(n-4). \tag{6.109}$$

From (6.43) we see that the $(n-4)$ factors arising from (6.107) and (6.108) cancel the $(n-4)^{-1}$ divergence from $\Gamma(2-n/2)$ to yield, as $n\to 4$.

$$\langle T_\mu{}^\mu\rangle_{\mathrm{div}} = (1/16\pi^2)[\alpha(F - \tfrac{2}{3}\Box R) + \beta G]. \tag{6.110}$$

Since this result is independent of m/μ, which has been retained in (6.109) essentially as an infrared cut-off, we can finally set $m = 0$, without changing the finite result (6.110). Note that because W_{div} is local, and independent of the state, so is $\langle T_\mu{}^\mu \rangle_{\text{div}}$. It depends only on the geometry at x.

Now because W is conformally invariant in the massless, conformally coupled limit, the expectation value of the trace of the *total* stress-tensor is zero:

$$\langle T_\mu{}^\mu \rangle \bigg|_{m=0,\,\xi=1/6} = -\frac{\Omega(x)}{[-g(x)]^{\frac{1}{2}}} \frac{\delta W[\bar{g}_{\mu\nu}]}{\delta\Omega(x)} \bigg|_{m=0,\,\xi=1/6,\,\Omega=1} = 0. \quad (6.111)$$

It follows that if the divergent portion $\langle T_{\mu\nu} \rangle_{\text{div}}$ has acquired the trace (6.110), then the finite, renormalized residue, $\langle T_{\mu\nu} \rangle_{\text{ren}}$, must also have a trace, i.e., the negative of (6.110):

$$\langle T_\mu{}^\mu \rangle_{\text{ren}} = -(1/16\pi^2)[\alpha(F - \tfrac{2}{3}\Box R) + \beta G] \quad (6.112)$$

$$= -a_2/16\pi^2 \quad (6.113)$$

$$= -(1/2880\pi^2)[R_{\alpha\beta\gamma\delta}R^{\alpha\beta\gamma\delta} - R_{\alpha\beta}R^{\alpha\beta} - \Box R] \quad (6.114)$$

$$= -(1/2880\pi^2)[C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta} + R_{\alpha\beta}R^{\alpha\beta} - \tfrac{1}{3}R^2 - \Box R] \quad (6.115)$$

The trace (6.112)–(6.115) has appeared in the theory even though the classical stress-tensor is traceless, and even though both W and W_{div} remain conformally invariant in four dimensions. It has arisen because the non-conformal nature of W_{div} (though not W) away from $n = 4$ leaves a finite imprint at $n = 4$ due to the $(n-4)^{-1}$ divergent nature of W_{div}. This result is known as a conformal, or trace, anomaly. A similar symmetry breaking, known as the axial vector anomaly, had previously been discovered (Adler 1969) in quantum electrodynamics.

The existence of an anomalous $\Box R$ term in the trace was originally found by Capper & Duff (1974, 1975). Following this, Deser, Duff & Isham (1976) demonstrated that the other terms might also arise, and in several subsequent investigations (see bibliography) their coefficients were established. Some of the early work was marred by mistakes (see Bunch 1979) and controversy. Fortunately, it is now known that all the regularization schemes predict the same conformal anomaly for the scalar field.

The appearance of the anomalous trace is closely associated with the scaling behaviour of the effective action and hence with the renormalization group (see chapter 9). To investigate this point we recall the technique of ζ-function regularization. Following along lines similar to those on page 80, one finds that under the conformal transformation

(6.97), the Feynman propagator for a massless, conformally coupled scalar field transforms to $\Omega^{-(n-2)/2}(x)G_F(x,x')\Omega^{-(n-2)/2}(x')$. Inspection of (6.26) reveals that $|x\rangle$ transforms to $\Omega^{-n/2}(x)|x\rangle$, so we see that the operator G_F, defined by (6.27) transforms to $\Omega^2 G_F$. Using this result in (6.70), one obtains

$$W[\bar{g}_{\mu\nu}] = -\tfrac{1}{2}\mathrm{i}\,\mathrm{tr}\,\{\ln[-(\mu\Omega)^2 G_F]\}, \tag{6.116}$$

in which Ω appears only in the combination $\mu\Omega$. Therefore, applying (6.100) to W, we may write

$$\langle T_\mu{}^\mu(x)\rangle = -\frac{\Omega(x)}{(-g)^{\frac{1}{2}}}\frac{\delta W[\bar{g}_{\mu\nu}]}{\delta\Omega(x)}\bigg|_{\Omega=1} = -\frac{\mu(x)}{(-g)^{\frac{1}{2}}}\frac{\delta W[\mu]}{\delta\mu(x)}\bigg|_{\mu=1}, \tag{6.117}$$

in which we are temporarily regarding μ as a function of x, and W to be a functional of μ. With this in mind we now employ (6.72) and (6.77) to write the massless limit of W as

$$W = -\tfrac{1}{2}\mathrm{i}\zeta'(0) + (1/32\pi^2)\int \mathrm{d}^4x[-g(x)]^{\frac{1}{2}}a_2(x)\ln[\mu^2(x)], \tag{6.118}$$

which when substituted in (6.117) yields

$$\langle T_\mu{}^\mu(x)\rangle_{\text{ren}} = -a_2(x)/16\pi^2. \tag{6.119}$$

Evidently the analytic continuation (in ν) used in the ζ-function method not only renders W finite, as was shown in the previous section, but also breaks the conformal invariance, producing the trace (6.119). This expression for the trace is in exact agreement with (6.113) obtained using dimensional regularization.

The calculation of the trace anomaly in spacetime of other than four dimensions is easily achieved (see, for example, Christensen 1978, Dowker & Kennedy 1978). First note that when n is odd, L_{eff}, given by (6.42), is finite. Hence one may conclude that there is no anomaly in odd-dimensional spacetimes.

For n even, and equal to n_0, only the first $1 + n_0/2$ terms in (6.42) are ultraviolet divergent, i.e., possess poles in their respective Γ functions at $n = n_0$. Of these, all but the $a_{n_0/2}$ term vanish at $n = n_0$ when $m \to 0$. The latter term contains the factor m^{n-n_0} and does not vanish. This term produces the anomalous trace

$$\langle T_\mu{}^\mu\rangle_{\text{ren}} = -a_{n_0/2}/(4\pi)^{n_0/2}, \tag{6.120}$$

a result which can be confirmed using ζ-function regularization.

One interesting case is $n_0 = 2$. We then have

$$\langle T_\mu{}^\mu \rangle_{\text{ren}} = -a_1/4\pi = -R/24\pi, \tag{6.121}$$

where we have used (6.47) with $\xi = 0$ (corresponding to conformal coupling in two dimensions).

It may be wondered whether the anomalous trace can be removed by adding suitable counterterms to the effective Lagrangian. There are no local geometrical quantities that, when varied in the effective action, yield a contribution to $\langle T_{\mu\nu} \rangle_{\text{ren}}$ whose trace will cancel the entire anomaly. However, it could be removed if one were to entertain a dynamical theory containing more complicated actions (Brown & Dutton 1978, Fradkin & Vilkovisky 1978). It is also possible to remove the $\Box R$ term from (6.114) or (6.115) by using the identity

$$\frac{2}{(-g)^{\frac{1}{2}}} g^{\mu\nu} \frac{\delta}{\delta g^{\mu\nu}} \int (-g)^{\frac{1}{2}} R^2 \mathrm{d}^4 x = -12 \Box R \tag{6.122}$$

(see (6.53)). By adding an R^2 term to L_{eff}(recall from (6.104)–(6.106) that $\alpha F + \beta G$ does not contain an R^2 term) the coefficient of $\Box R$ in the anomalous trace may be varied at will. It could, if desired, be made equal to zero. Of course, the introduction of an R^2 term in L_{eff} means that W_{eff} is no longer invariant under conformal transformations. Whether or not this is reasonable can only be answered by experiment. There seems to be no compelling theoretical reason why the conformal symmetry of the dynamics should be broken by hand, so we shall retain the $\Box R$ term in the anomaly in what follows (see, however, Horowitz & Wald 1978).

We now turn to the trace anomaly for higher-spin fields. So long as we require W to be conformally invariant, then the anomalous trace can only be of the form (6.112) for some α and β. However, if W is not conformally invariant (e.g., massless scalar field with $\xi \neq \xi(n)$), then the anomalous trace may contain additional $\Box R$ and R^2 terms. It will also, of course, contain an ordinary, i.e., non-anomalous, component that will depend on the quantum state. The latter contribution will not generally be of a geometrical, or even local form (see chapter 7 for some explicit examples). The anomalous portion of the trace will be the state-independent, local, geometrical piece.

As remarked above, the appearance of a conformal anomaly can be understood in the case when dimensional regularization is used as the fact that, although W is conformally invariant in n dimensions, W_{div} is only conformally invariant in four dimensions. When the dimensionality is analytically continued to n, the breakdown of conformal invariance away from $n = 4$ survives in $\langle T_\mu{}^\mu \rangle_{\text{div}}$ (and hence $\langle T_\mu{}^\mu \rangle_{\text{ren}}$) at $n = 4$. One may have to contend with theories in which not only W_{div}, but W itself is only conformally invariant at some particular dimension. Two examples of this are the scalar case, if one insists on fixing ξ independently of n, and the electromagnetic field, which is only conformally invariant in four dimensions. In these cases, dimensional regularization yields different results from the other regularization schemes. However, the difference is only in the coefficient of the $\Box R$ anomaly, which we have seen can anyway be changed by the addition of an R^2 counterterm to L_{eff}, and so must be determined experimentally.

The generalization of (6.119) to fields of arbitrary spin has been given by Christensen & Duff (1978*a*):

$$\langle T_\mu{}^\mu \rangle_{\text{ren}} = -\frac{(-1)^{2A+2B}}{16\pi^2}\operatorname{tr} a_2(A,B) \tag{6.123}$$

where (A, B) labels the representation of the Lorentz group under which the field concerned transforms (see (3.164)). This trace can be written in terms of four parameters as

$$\langle T_\mu{}^\mu \rangle_{\text{ren}} = (2880\pi^2)^{-1}\{aC_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta} + b(R_{\alpha\beta}R^{\alpha\beta} - \tfrac{1}{3}R^2) + c\Box R + \mathrm{d}R^2\}. \tag{6.124}$$

The coefficients a, b, c and d can all be expressed as simple polynomials in A and B. The general formulae are given by Christensen & Duff (1979) for the particular higher-spin field equations used by them. We list in table 1 some of the more important results for fields with spins ≤ 2.

Note that table 1 only gives the anomalous contribution. If the field is not conformally invariant, then there will be additional, non-anomalous

Table 1. *The coefficients appearing in the trace anomaly equation* (6.124) *for fields of various spins. The results for spin $\frac{1}{2}$ assume two component spinors, the results for four components being obtained by multiplication by two. The crosses indicate where consistency conditions for higher-spin fields require the corresponding geometrical object in the anomaly to vanish* (*Christensen & Duff* 1980).

(A,B)	a	b	c	d
$(0,0)$	-1	-1	$(6-30\xi)$	$-90(\xi-\frac{1}{6})^2$
$(\frac{1}{2},0)$	$-\frac{7}{4}$	$-\frac{11}{2}$	3	0
$(\frac{1}{2},\frac{1}{2})$	11	-64	-6	-5
$(1,0)$	-33	27	12	$-\frac{5}{2}$
$(1,\frac{1}{2})$	$\frac{291}{4}$	$\times$	$\times$	$\frac{61}{8}$
$(1,1)$	-189	$\times$	$\times$	$-\frac{747}{4}$

contributions, as already remarked. In particular, in the scalar case $\xi \neq \frac{1}{6}$, one has the non-anomalous contribution

$$6(\xi - \tfrac{1}{6})\{\langle \phi_{;\mu}\phi^{;\mu}\rangle + \xi R\langle \phi^2\rangle\}, \tag{6.125}$$

which depends on the quantum state chosen. The scalar, $(\frac{1}{2},0)$ spin $\frac{1}{2}$, and $(\frac{1}{2},\frac{1}{2})$ spin 1 results in the table can be obtained from (6.48), (6.93) and (6.94) respectively.

In general, a higher-spin physical field of interest will not correspond to a single representation (A,B) of the Lorentz group, but will be a linear combination of several such representations. For example, the electromagnetic field contains scalar ghost contributions, so to obtain the electromagnetic anomaly, one must subtract from the $(\frac{1}{2},\frac{1}{2})$ anomaly twice the $(0,0)$ anomaly. In table 2 we list these various physical combinations for the massless fields with spins ≤ 2. The results have been computed using ζ-function regularization. As remarked, use of dimensional regularization alters the coefficient of $\Box R$ in the electromagnetic case. Brown & Cassidy (1977*b*) find $c = 12$.

There is a useful consistency condition on the coefficients a, b, c and d that arises when dimensional regularization is applied to a theory which is conformally invariant in n-dimensions (Duff 1977). A comparison of (6.112) and (6.124) reveals that

$$a = -180(\alpha + \beta), \quad b = 360\beta, \quad c = 120\alpha, \quad d = 0 \tag{6.126}$$

from which one obtains the constraints

$$2a + b + 3c = 0, \quad d = 0. \tag{6.127}$$

Clearly only two of the four coefficients a, b, c and d remain to be determined.

The trace anomaly is especially important in the special case that the

Table 2. *The coefficients in the trace anomaly for the physical massless fields of spin* ≤ 2. *In the calculation of the spin* 1 *and spin* 2 *results* (0,0) *represents the minimally coupled scalar field* ($\xi = 0$). *The crosses have the same meaning as in table* 1.

Spin	(A,B)	a	b	c	d
0	$(0,0)$	-1	-1	$(6-30\xi)$	$-90(\xi-\frac{1}{6})^2$
$\frac{1}{2}$	$(\frac{1}{2},0)$	$-\frac{7}{4}$	$-\frac{11}{2}$	3	0
1	$(\frac{1}{2},\frac{1}{2})-2(0,0)$	13	-62	-18	0
$\frac{3}{2}$	$(1,\frac{1}{2})-2(\frac{1}{2},0)$	$\frac{233}{4}$	$\times$	$\times$	$\frac{61}{8}$
2	$(1,1,)+(0,0)-2(\frac{1}{2},\frac{1}{2})$	-212	$\times$	$\times$	$-\frac{717}{4}$

background spacetime is conformally flat. If the quantum field is also conformally invariant, then we have a conformally trivial situation (see §3.7). In this case, it turns out that the anomalous trace determines the entire stress-tensor once the quantum state has been specified (Brown & Cassidy 1977*a*, Bunch & Davies 1977*b*, Davies 1977*a*, *b*).

To see this we use (6.99) with S replaced by W_{ren}:

$$W_{\text{ren}}[\bar{g}_{\mu\nu}] = W_{\text{ren}}[g_{\mu\nu}] - \int [-\bar{g}(x)]^{\frac{1}{2}} \langle T_\rho{}^\rho[\bar{g}_{\mu\nu}(x)] \rangle_{\text{ren}} \Omega^{-1}(x) \delta\Omega(x) \mathrm{d}^n x. \tag{6.128}$$

Now using (6.13) and the fact that

$$\bar{g}^{\nu\sigma} \frac{\delta}{\delta \bar{g}^{\mu\sigma}} = g^{\nu\sigma} \frac{\delta}{\delta g^{\mu\sigma}},$$

one obtains

$$\begin{aligned} \langle T_\mu{}^\nu[\bar{g}_{\kappa\lambda}(x)] \rangle_{\text{ren}} &= (g/\bar{g})^{\frac{1}{2}} \langle T_\mu{}^\nu[g_{\kappa\lambda}(x)] \rangle_{\text{ren}} \\ &\quad - \frac{2}{[-\bar{g}(x)]^{\frac{1}{2}}} \bar{g}^{\nu\sigma}(x) \frac{\delta}{\delta \bar{g}^{\mu\sigma}} \int [-\bar{g}(x')]^{\frac{1}{2}} \\ &\quad \times \langle T_\rho{}^\rho[\bar{g}_{\kappa\lambda}(x')] \rangle_{\text{ren}} \Omega^{-1}(x') \delta\Omega(x') \mathrm{d}^n x'. \end{aligned} \tag{6.129}$$

For conformally invariant field theories, the trace in the integral on the right-hand side of this equation is purely anomalous, which we know to be local and state-independent. Thus, independently of the specification of the state on the left-hand side one may perform the variational integration (Brown & Cassidy 1977*a*, DeWitt 1979). Alternatively, within the framework of dimensional regularization, one may proceed as follows: From the discussion associated with (6.111)–(6.115), we have

$$\begin{aligned} \langle T_\rho{}^\rho[\bar{g}_{\kappa\lambda}(x)] \rangle_{\text{ren}} &= - \langle T_\rho{}^\rho[\bar{g}_{\kappa\lambda}(x)] \rangle_{\text{div}} \\ &= \frac{\Omega(x)}{[-\bar{g}(x)]^{\frac{1}{2}}} \frac{\delta W_{\text{div}}[\bar{g}_{\kappa\lambda}]}{\delta \Omega(x)}, \end{aligned} \tag{6.130}$$

which, when substituted in (6.129), permits the integration to be carried out immediately, giving

$$\begin{aligned} \langle T_\mu{}^\nu[\bar{g}_{\kappa\lambda}(x)] \rangle_{\text{ren}} &= (g/\bar{g})^{\frac{1}{2}} \langle T_\mu{}^\nu[g_{\kappa\lambda}(x)] \rangle_{\text{ren}} \\ &\quad - \frac{2}{[-\bar{g}(x)]^{\frac{1}{2}}} \bar{g}^{\nu\sigma}(x) \frac{\delta}{\delta \bar{g}^{\mu\sigma}(x)} W_{\text{div}}[\bar{g}_{\kappa\lambda}] \\ &\quad + \frac{2}{[-\bar{g}(x)]^{\frac{1}{2}}} g^{\nu\sigma} \frac{\delta}{\delta g^{\mu\sigma}(x)} W_{\text{div}}[g_{\kappa\lambda}]. \end{aligned} \tag{6.131}$$

We first apply this result to the two-dimensional case. Using (6.41) and (6.47) with $\xi = \xi(2) = 0$, and the fact that

$$\Gamma(1 - n/2) = 2/(2 - n) + O(1),$$

one has

$$W_{\text{div}}[g_{\kappa\lambda}] = -[1/4\pi(n-2)] \int [-g(x')]^{\frac{1}{2}} a_1 [g_{\kappa\lambda}(x')] \mathrm{d}^n x'$$

$$= -[1/24\pi(n-2)] \int [-g(x')]^{\frac{1}{2}} R(x') \mathrm{d}^n x', \qquad (6.132)$$

in which terms of order $n - 2$ have been dropped. Substituting (6.132) and its counterpart for $\bar{g}_{\kappa\lambda}$ into (6.131), one obtains

$$\begin{aligned} \langle T_\mu{}^\nu[\bar{g}_{\kappa\lambda}(x)]\rangle_{\text{ren}} = {} & (g/\bar{g})^{\frac{1}{2}} \langle T_\mu{}^\nu[g_{\kappa\lambda}(x)]\rangle_{\text{ren}} \\ & + [1/12\pi(n-2)][(\bar{R}_\mu{}^\nu - \tfrac{1}{2}\delta_\mu{}^\nu \bar{R}) \\ & - (R_\mu{}^\nu - \tfrac{1}{2}\delta_\mu{}^\nu R)], \end{aligned} \qquad (6.133)$$

which, using (3.3) and (3.4) for $\bar{R}_\mu{}^\nu$ and $\bar{R}$, gives

$$\begin{aligned} \langle T_\mu{}^\nu[\bar{g}_{\kappa\lambda}(x)]\rangle_{\text{ren}} = {} & (g/\bar{g})^{\frac{1}{2}} \langle T_\mu{}^\nu[g_{\kappa\lambda}(x)]\rangle_{\text{ren}} \\ & + (1/12\pi)[(\Omega^{-3}\Omega_{;\rho\mu} - 2\Omega^{-4}\Omega_{;\rho}\Omega_{;\mu})g^{\rho\nu} \\ & + \delta_\mu{}^\nu g^{\rho\sigma}(\tfrac{3}{2}\Omega^{-4}\Omega_{;\rho}\Omega_{;\sigma} - \Omega^{-3}\Omega_{;\rho\sigma})]. \end{aligned} \qquad (6.134)$$

It is not difficult to show using (6.93) that (6.132) and hence (6.134) also holds in the case of two-component spin $\frac{1}{2}$ spinor fields.

All two-dimensional spacetimes are conformally flat:

$$g_{\mu\nu} = C(x)\eta_{\mu\nu}.$$

Thus (6.134) with $g_{\kappa\lambda} = \eta_{\kappa\lambda}$, and $\Omega = C^{\frac{1}{2}}$, enables one to write the expectation value of the stress-tensor in any two-dimensional curved spacetime in terms of its expectation value in flat spacetime. This result takes a particularly simple form in null coordinates (3.9) in which

$$\mathrm{d}s^2 = C(u, v)\mathrm{d}u\mathrm{d}v, \qquad (6.135)$$

when (6.134) gives

$$\begin{aligned} \langle T_\mu{}^\nu[g_{\kappa\lambda}(x)]\rangle_{\text{ren}} = {} & (-g)^{-\frac{1}{2}} \langle T_\mu{}^\nu[\eta_{\kappa\lambda}(x)]\rangle_{\text{ren}} + \theta_\mu{}^\nu \\ & - (1/48\pi)R\delta_\mu{}^\nu, \end{aligned} \qquad (6.136)$$

(Davies 1977*b*) where

$$\left.\begin{aligned} \theta_{uu} &= -(1/12\pi)C^{\frac{1}{2}}\partial_u^2 C^{-\frac{1}{2}} \\ \theta_{vv} &= -(1/12\pi)C^{\frac{1}{2}}\partial_v^2 C^{-\frac{1}{2}} \\ \theta_{uv} &= \theta_{vu} = 0. \end{aligned}\right\} \tag{6.137}$$

If the state used in evaluating the expectation value in flat spacetime is a vacuum state, then the state appearing in the curved spacetime expectation value is a conformal vacuum. As discussed in §5.5, whether or not the flat spacetime vacuum is the usual Minkowski space vacuum depends on whether the curved spacetime is conformal to all, or only a part of Minkowski space. If it is conformal to all of Minkowski space, then the usual vacuum is indeed used, and the first term on the right-hand side of (6.136) vanishes. Otherwise this term will give a nonzero contribution. We shall pursue this point further below.

We now turn to the four-dimensional conformally trivial case. From (6.103)

$$W_{\text{div}} = -[1/16\pi^2(n-4)]\int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}}[\alpha F(x) + \beta G(x)] + O(1). \tag{6.138}$$

Substituting (6.138) in (6.131) with $\bar{g}_{\kappa\lambda}$ replaced by $g_{\kappa\lambda}$, and $g_{\kappa\lambda}$ replaced by $\bar{g}_{\kappa\lambda}$, a flat spacetime metric (e.g., $\eta_{\kappa\lambda}$), and performing the functional differentiations (Brown & Cassidy 1977*b*, Bunch 1979) one obtains

$$\langle T_\mu{}^\nu[g_{\kappa\lambda}]\rangle_{\text{ren}} = (\bar{g}/g)^{\frac{1}{2}}\langle T_\mu{}^\nu[\bar{g}_{\kappa\lambda}]\rangle_{\text{ren}} - (1/16\pi^2)\{\tfrac{1}{9}\alpha\,{}^{(1)}H_\mu{}^\nu + 2\beta\,{}^{(3)}H_\mu{}^\nu\}, \tag{6.139}$$

where ${}^{(1)}H_{\mu\nu}$ is given by (6.53), and

$$\begin{aligned} {}^{(3)}H_{\mu\nu} &\equiv \tfrac{1}{12}R^2 g_{\mu\nu} - R^{\rho\sigma}R_{\rho\mu\sigma\nu} \\ &= R_\mu{}^\rho R_{\rho\nu} - \tfrac{2}{3}RR_{\mu\nu} - \tfrac{1}{2}R_{\rho\sigma}R^{\rho\sigma}g_{\mu\nu} + \tfrac{1}{4}R^2 g_{\mu\nu}. \end{aligned} \tag{6.140}$$

In the scalar case, α and β are given by (6.106) and

$$\langle T_\mu{}^\nu[g_{\kappa\lambda}]\rangle_{\text{ren}} = (\bar{g}/g)^{\frac{1}{2}}\langle T_\mu{}^\nu[\bar{g}_{\kappa\lambda}]\rangle_{\text{ren}} - (1/2880\pi^2)[\tfrac{1}{6}{}^{(1)}H_\mu{}^\nu - {}^{(3)}H_\mu{}^\nu]. \tag{6.141}$$

For other conformally invariant field theories, α and β can be obtained from the anomaly coefficients a, b, c, d of (6.124), by using (6.126).

The result (6.139) has been derived in a less formal manner by Bunch &

Davies (1977*b*). As their method appeals more to the underlying physics of the situation we shall briefly outline here the main steps involved.

We first note that because $\langle T_{\mu\nu}\rangle_{\text{ren}}$ is derived from an effective action W_{ren}, it must automatically be covariantly conserved

$$\langle T_{\mu}{}^{\nu}\rangle_{\text{ren};\nu} = 0. \tag{6.142}$$

Next, we note that if the quantum state is a conformal vacuum, or any of the associated excited states, then $\langle T_{\mu}{}^{\nu}\rangle_{\text{ren}}$ must be a purely *local* tensor, i.e., it depends only on the geometry at x. This follows because G_{F} is purely local (it may be constructed by simple conformal scaling from Minkowski space where it is manifestly local; see, for example, (3.154)). Moreover, the differentiation of G_{F} required to produce $\langle T_{\mu}{}^{\nu}\rangle_{\text{ren}}$ and the renormalization procedure, are also both local (and covariant) processes.

Mindful of the above two restrictions we proceed to find the most general local conserved tensor with the required units of (length)$^{-4}$. From the relation

$$\begin{aligned} R_{\alpha\beta\gamma\delta} = \tfrac{1}{2}(g_{\alpha\gamma}R_{\beta\delta} - g_{\alpha\delta}R_{\beta\gamma} - g_{\beta\gamma}R_{\alpha\delta} + g_{\beta\delta}R_{\alpha\gamma}\\ + \tfrac{1}{6}R(g_{\alpha\delta}g_{\beta\gamma} - g_{\alpha\gamma}g_{\beta\delta}) + C_{\alpha\beta\gamma\delta}, \end{aligned} \tag{6.143}$$

it is clear that in a conformally flat spacetime, where $C_{\alpha\beta\gamma\delta} = 0$, the Riemann tensor is determined entirely in terms of the Ricci tensor and its contractions. Taking into account further degeneracy due to the conformal symmetry (Davies, Fulling, Christensen & Bunch 1977) one arrives at six independent geometrical quantities with the correct units:

$$R_{\mu}{}^{\alpha}R_{\alpha\nu}, RR_{\mu\nu}, R_{;\mu\nu}, R_{\alpha\beta}R^{\alpha\beta}g_{\mu\nu}, R^2 g_{\mu\nu}, g_{\mu\nu}\Box R. \tag{6.144}$$

We require linear combinations of these tensors that will be covariantly conserved. Two local conserved tensors are already known: ${}^{(1)}H_{\mu\nu}$ and ${}^{(2)}H_{\mu\nu}$. In conformally flat spacetimes, however, they are not independent:

$${}^{(2)}H_{\mu\nu} = \tfrac{1}{3}{}^{(1)}H_{\mu\nu}. \tag{6.145}$$

One readily verifies, using (6.142) that the only other local conserved tensor is ${}^{(3)}H_{\mu\nu}$ defined by (6.140). Although purely geometrical the latter tensor is only accidentally conserved in conformally flat spacetimes, i.e., it cannot be derived from variation of a geometrical term in the action, nor is it the limit of a tensor that is conserved in a non-conformally flat spacetime (Ginzburg, Kirzhnits & Lyubushin 1971).

In addition to these two geometrical tensors, there may be another conserved tensor that is local, but non-geometrical, i.e., a tensor determined entirely by the local geometry, but which cannot be expressed in terms of $R_{\alpha\beta}$ or R. Call this extra tensor ${}^{(4)}H_{\mu\nu}$. Thus

$$\langle T_\mu{}^\nu \rangle_{\text{ren}} = A\,{}^{(1)}H_\mu{}^\nu + B\,{}^{(3)}H_\mu{}^\nu + {}^{(4)}H_\mu{}^\nu \tag{6.146}$$

for some constants A and B. Taking the trace of (6.146) yields

$$\langle T_\mu{}^\mu \rangle_{\text{ren}} = -6A\,\Box R - B(R_{\alpha\beta}R^{\alpha\beta} - \tfrac{1}{3}R^2) + {}^{(4)}H_\mu{}^\mu, \tag{6.147}$$

which when compared with (6.112) (remembering $F = C_{\alpha\beta\gamma\delta}C^{\alpha\beta\gamma\delta} = 0$) gives

$$A = -\alpha/144\pi^2, \quad B = -\beta/8\pi^2, \quad {}^{(4)}H_\mu{}^\mu = 0. \tag{6.148}$$

The anomalous trace therefore determines $\langle T_{\mu\nu} \rangle_{\text{ren}}$ up to the local, conserved, traceless tensor ${}^{(4)}H_{\mu\nu}$, the value of which will in any case depend on the choice of quantum state. Comparison of (6.139) with (6.148) reveals exact agreement, with ${}^{(4)}H_{\mu\nu}$ identified as the flat spacetime 'boundary' term $(\tilde{g}/g)^{\frac{1}{2}}\langle T_\mu{}^\nu[\tilde{g}_{\kappa\lambda}] \rangle_{\text{ren}}$.

We now apply the form (6.139) to some specific examples, namely, the Robertson–Walker spacetimes discussed in §5.1. Being homogeneous and isotropic, any uniform distribution of matter (including the conformal vacuum state) in this class of spacetimes will share their geometrical symmetries. This restricts $\langle T_{\mu\nu} \rangle_{\text{ren}}$ in the following way

$$\left.\begin{aligned} &\langle T_1{}^1 \rangle = \langle T_2{}^2 \rangle = \langle T_3{}^3 \rangle \\ &\langle T_\mu{}^\nu \rangle_{\text{ren}} = 0, \quad \mu \neq \nu. \end{aligned}\right\} \tag{6.149}$$

Working in the coordinate system in which the metric takes the form (5.11) one can use (5.13) to obtain

$$\left.\begin{aligned} {}^{(1)}H_{00} &= C^{-1}[-9\ddot{D}D + \tfrac{9}{2}\dot{D}^2 + \tfrac{27}{8}D^4 + 9KD^2 - 18K^2] \\ {}^{(1)}H_{11} &= C^{-1}\Upsilon[6\dddot{D} - 3\ddot{D}D + \tfrac{3}{2}\dot{D}^2 - 9\dot{D}D^2 + \tfrac{9}{8}D^4 \\ &\quad - 12K\dot{D} + 3KD^2 - 6K^2] \end{aligned}\right\} \tag{6.150}$$

$$\left.\begin{aligned} {}^{(3)}H_{00} &= C^{-1}[\tfrac{3}{16}D^4 + \tfrac{3}{2}KD^2 + 3K^2] \\ {}^{(3)}H_{11} &= C^{-1}\Upsilon[-\tfrac{1}{2}\dot{D}D^2 + \tfrac{1}{16}D^4 - 2K\dot{D} + \tfrac{1}{2}KD^2 + K^2]. \end{aligned}\right\} \tag{6.151}$$

In the case of the spacetimes shown in figs. 20a–d the conformal vacuum is based on the Minkowski space vacuum, so the first term on the right-hand side of (6.139) vanishes, and the conformal vacuum expectation value of the stress-tensor is given in terms of (6.150) and (6.151) alone. In particular, for the Einstein universe, for which $C = a^2 = $ constant, $D = 0$ (see

§5.2), one obtains

$$\langle T_\mu{}^\nu \rangle_{\text{Einstein}} = \frac{p(s)}{2\pi^2 a^4} \operatorname{diag}(1, -\tfrac{1}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}), \tag{6.152}$$

where $p(s)$ is a spin-dependent coefficient which takes the values

$$p(0) = \tfrac{1}{240}, \quad p(\tfrac{1}{2}) = \tfrac{17}{960}, \quad p(1) = \tfrac{11}{120}.$$

(See Kennedy & Unwin 1980 for the case of an Einstein universe with points identified.) In the closed de Sitter spacetime (5.65), or the steady-state universe (5.52), one readily finds

$$\langle T_\mu{}^\nu \rangle_{\substack{\text{steady}\\ \text{state}}} = \frac{q(s)}{960\pi^2 \alpha^4} \delta_\mu{}^\nu, \tag{6.153}$$

where

$$q(0) = 1, \quad q(\tfrac{1}{2}) = \tfrac{11}{2}, \quad q(1) = 62.$$

For the spacetimes shown in figs. 20*e*, *f*, the conformal vacuum may be based on the Minkowskian or Rindler vacua (see §4.5). In the latter case the first term on the right-hand side of (6.139) no longer vanishes. The four-dimensional Rindler line element can be written as

$$ds^2 = \zeta^2 d\eta^2 - d\zeta^2 - dy^2 - dz^2, \quad (0 < \zeta < \infty) \tag{6.154}$$

which is the four-dimensional extension of (4.71), with $\zeta = a^{-1} e^{a\xi}$ and the η of (4.71) being replaced by $a^{-1}\eta$. Under the coordinate transformation

$$\begin{aligned} \zeta &= \Upsilon^{\frac{1}{2}}(1 - r\Upsilon^{\frac{1}{2}} \cos\theta)^{-1}, \\ y &= r\Upsilon^{\frac{1}{2}} \sin\theta \cos\phi (1 - r\Upsilon^{\frac{1}{2}} \cos\theta)^{-1}, \\ z &= r\Upsilon^{\frac{1}{2}} \sin\theta \sin\phi (1 - r\Upsilon^{\frac{1}{2}} \cos\theta)^{-1}, \end{aligned} \tag{6.155}$$

with Υ having its $K = -1$ value, $(1 + r^2)^{-1}$, (6.154) takes the form

$$ds^2 = \zeta^2 [d\eta^2 - \Upsilon dr^2 - r^2(d\theta^2 + \sin^2\theta \, d\phi^2)], \tag{6.156}$$

which clearly shows the conformal relation between the Rindler and $K = -1$ Robertson–Walker spacetimes. The Rindler vacuum expectation value of the stress-tensor has been evaluated by Candelas & Deutsch (1977, 1978), who obtain the result in the interesting form

$$\langle T_\mu{}^\nu [\eta_{\lambda\kappa}] \rangle_{\text{Rindler}} = \frac{h(s)}{2\pi^2 \zeta^4} \int_0^\infty \frac{d\nu \, \nu(\nu^2 + s^2)}{e^{2\pi\nu} - (-1)^{2s}} \operatorname{diag}(-1, \tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}) \tag{6.157}$$

where $h(s)$ is the number of helicity states of the spin s field.

It is now a straightforward matter to obtain $\langle T_\mu{}^\nu\rangle_{\mathrm{ren}}$ in the open static universe (§5.2). The geometrical contributions, ${}^{(1)}H_\mu{}^\nu$, ${}^{(3)}H_\mu{}^\nu$ will be the same as for the Einstein universe, so

$$\langle T_\mu{}^\nu\rangle_{\substack{\text{static}\\ \text{open}}} = (\zeta^4/a^4)\langle T_\mu{}^\nu\rangle_{\text{Rindler}} + \langle T_\mu{}^\nu\rangle_{\text{Einstein}} = 0. \tag{6.158}$$

The latter equality follows from the fact that $p(s)$ in (6.152) can be represented as

$$p(s) = h(s)\int_0^\infty \frac{\mathrm{d}v\; v(v^2+s^2)}{\mathrm{e}^{2\pi v} - (-1)^{2s}}. \tag{6.159}$$

Alternatively, one can independently obtain the result (6.158) (Bunch 1978*a*; see §7.2) and so arrive at (6.157). The thermal connection involved in these various vacuum states, represented schematically in fig. 20, is immediately evident from the Planckian-type integrals in (6.157) and (6.159).

Finally, the static form of de Sitter spacetime (5.76) provides another interesting example conformally related to the Rindler spacetime. The metric (5.76) written in the form

$$\mathrm{d}s^2 = [1-(r^2/\alpha^2)]\{\mathrm{d}t^2 - [1-(r^2/\alpha^2)]^{-2}\mathrm{d}r^2 - r^2[1-(r^2/\alpha^2)]^{-1}(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)\},$$

transforms, under the change of variable $r/\alpha = r'[1+(r')^2]^{-\frac{1}{2}}$, to

$$\mathrm{d}s^2 = (\alpha^2 - r^2)[(\mathrm{d}t^2/\alpha^2) - \Upsilon\mathrm{d}r'^2 - r'^2(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)] \tag{6.160}$$

with $\Upsilon = [1+(r')^2]^{-1}$, which is manifestly conformal to (6.156) if $\eta = t/\alpha$. The contribution from the geometrical terms of (6.139) is independent of coordinates, and so will be the same as the steady-state case (6.153). Thus, one immediately obtains for the stress-tensor in the static de Sitter (Rindler) conformal vacuum

$$\begin{aligned}\langle T_\mu{}^\nu\rangle_{\substack{\text{static}\\ \text{de Sitter}}} &= \zeta^4(\alpha^2 - r^2)^{-2}\langle T_\mu{}^\nu\rangle_{\text{Rindler}} + \langle T_\mu{}^\nu\rangle_{\substack{\text{steady}\\ \text{state}}}\\ &= \frac{-p(s)}{2\pi^2}(\alpha^2-r^2)^{-2}\,\mathrm{diag}\,(1, -\tfrac{1}{3}, -\tfrac{1}{3}, -\tfrac{1}{3})\\ &\quad + \frac{q(s)}{960\pi^2\alpha^4}\,\delta_\mu{}^\nu,\end{aligned} \tag{6.161}$$

which clearly diverges at the horizon ($r = \alpha$). This is the undesirable feature of the 'observer-dependent' de Sitter vacuum mentioned at the end of §5.4.

A useful relation between the trace anomaly and the complete stress-tensor in Robertson–Walker spacetimes has been given by Parker (1979). Robertson–Walker spacetimes possess a conformal Killing vector field ξ^μ (see §3.1)

$$\xi^\mu = a(t)\delta^\mu{}_t \tag{6.162}$$

where $a = C^{\frac{1}{2}}$, and the λ of (3.23) is here given by

$$\lambda = 2a'(t) \tag{6.163}$$

the prime denoting differentiation with respect to t. From the conservation equation (6.142), and (3.23), we conclude

$$\begin{aligned}[\langle T^{\mu\nu}\rangle_{\text{ren}}\xi_\nu]_{;\mu} &= \langle T^{\mu\nu}\rangle_{\text{ren}}\xi_{(\nu;\mu)} \\ &= \tfrac{1}{2}\lambda\langle T^{\mu\nu}\rangle_{\text{ren}}g_{\mu\nu}.\end{aligned} \tag{6.164}$$

Integrating (6.164) over a 4-volume bounded by constant time hyper-surfaces t_1 and t_2 gives

$$\begin{aligned}\tfrac{1}{2}\int \mathrm{d}^4x(-g)^{\frac{1}{2}}\lambda\langle T_\mu{}^\mu\rangle_{\text{ren}} = &\int_{t_2} \mathrm{d}^3x(-g)^{\frac{1}{2}}\langle T^{t\nu}\rangle_{\text{ren}}\xi_\nu \\ &- \int_{t_1} \mathrm{d}^3x(-g)^{\frac{1}{2}}\langle T^{t\nu}\rangle_{\text{ren}}\xi_\nu\end{aligned} \tag{6.165}$$

where we have used the divergence theorem. Using the fact that $\langle T_{\mu\nu}\rangle$ is a function of t alone, (6.165) reduces to

$$\rho(t_2)a^4(t_2) = \rho(t_1)a^4(t_1) - \int_{t_1}^{t_2} a^3a'\langle T_\mu{}^\mu\rangle_{\text{ren}}\mathrm{d}t, \tag{6.166}$$

where $\rho = \langle T^{tt}\rangle_{\text{ren}}$ is the proper energy density as measured by a comoving observer, and we have used (6.162) and (6.163).

If (6.166) is applied to ${}^{(4)}H^{\mu\nu}$ alone, then its tracelessness implies

$${}^{(4)}H_1{}^1 = -\tfrac{1}{3}{}^{(4)}H_t{}^t = -\tfrac{1}{3}\rho(t) \propto a^{-4}(t) \tag{6.167}$$

which expresses the fact that ${}^{(4)}H_{\mu\nu}$ describes massless radiation with classical behaviour, that redshifts in the usual way as the universe expands: the total energy in a comoving volume, ${}^{(4)}H_{tt}a^3$, decreases like a^{-1}. The conservation equation (6.142) reduces in this case to the simple form

$$p\mathrm{d}a^3 + \mathrm{d}(\rho a^3) = 0 \tag{6.168}$$

where the pressure $p = -{}^{(4)}H_1{}^1$.

If (6.166) is applied to the entire stress-tensor $\langle T_{\mu\nu}\rangle_{\text{ren}}$ given by (6.146), then we must take into account the non-vanishing trace (6.147). Using the formulae (5.13).

$$R_{\alpha\beta}R^{\alpha\beta} = C^{-2}[3\dot{D}^2 + \tfrac{3}{2}\dot{D}D^2 + \tfrac{3}{4}D^4 + 6K\dot{D} + 6KD^2 + 12K^2] \tag{6.169}$$

$$\Box R = C^{-2}[3\ddot{D} - \tfrac{9}{2}\dot{D}D^2 - 6K\dot{D}], \tag{6.170}$$

it is easily verified that the integrand of (6.166) is an exact time derivative. Integrating the trace term yields

$$3[B(a'^4 + 2Ka'^2) + 12A(-a^2a'a''' - aa'^2a'' + \tfrac{1}{2}a^2a''^2 + \tfrac{3}{2}a'^4 + Ka'^2)], \tag{6.171}$$

where A, B are given in (6.148).

In the case of an asymptotically static spacetime, (6.171) vanishes and (6.166) then yields

$$\rho(\text{in})a^4(\text{in}) = \rho(\text{out})a^4(\text{out}) \tag{6.172}$$

which expresses the fact that any energy present in the field originally (e.g., due to space curvature or the presence of quanta) merely redshifts like classical radiation into the out region. No augmentation or reduction of the field energy occurs as a result of the expansion. This is not surprising, perhaps, in a conformally trivial situation. Indeed, we already know from §3.7 that no quanta are created from the conformal vacuum by the expansion motion. Equation (6.172) is another expression of this.

6.4 Computing the renormalized stress-tensor

In the previous sections of this chapter it has been demonstrated how the formally divergent quantity $\langle T_{\mu\nu}\rangle$ can be rendered finite by renormalization of coupling constants in the gravitational action, i.e., by adding infinite counterterms that are purely geometrical and can be regarded as part of the gravitational dynamics. In practice, of course, interest centres not on the divergences themselves, but on the finite remainder, since this is supposed to be the physically relevant portion. In particular, $\langle T_{\mu\nu}\rangle_{\text{ren}}$ is the quantity that is intended to reside on the right-hand side of the generalized Einstein equation (6.95). Whilst some conformally trivial systems allow $\langle T_{\mu\nu}\rangle_{\text{ren}}$ to be computed entirely from a knowledge of the trace anomaly (see §6.3) in general such short cuts are not available, and laborious 'brute force' techniques are necessary. In this section we shall outline how such techniques can be employed.

First it has to be decided what quantity one wishes to compute. The standard development of the DeWitt–Schwinger representation yields

$$\frac{\langle \text{out}, 0 | T_{\mu\nu} | 0, \text{in} \rangle}{\langle \text{out}, 0 | 0, \text{in} \rangle}, \tag{6.173}$$

but this is not, obviously, a true expectation value. That role is reserved for a quantity of the sort $\langle \psi | T_{\mu\nu} | \psi \rangle$. Working in the Heisenberg picture, it is natural to specify the quantum state in the in region, e.g., $\langle \text{in}, 0 | T_{\mu\nu}(x) | 0, \text{in} \rangle$, though x may refer to any point in the spacetime. As shown at the end of §6.2, the divergences, hence renormalization, of the various vacuum expressions are identical. Only the finite remainder differs.

In principle, it is possible to compute $\langle \psi | T_{\mu\nu} | \psi \rangle$ using the finite renormalized effective Lagrangian L_{ren} combined with a computation of the A coefficients in (6.96). Besides being extremely complicated, however, the DeWitt–Schwinger representation is not trustworthy at long wavelengths if computed by an asymptotic expansion of the form used to obtain (6.61). Instead, it is better to work with $\langle \psi | T_{\mu\nu} | \psi \rangle$ from scratch, using the DeWitt–Schwinger representation solely for the purpose of renormalization.

Because $\langle T_{\mu\nu} \rangle_{\text{ren}}$ is determined by the low-energy, long-wavelength portion of $\langle T_{\mu\nu} \rangle$, it will be sensitive to the global structure of the spacetime manifold, and the quantum state chosen. It will not, therefore be a geometrical, or even a local, object in general. A variety of techniques is available for computing $\langle T_{\mu\nu} \rangle_{\text{ren}}$, depending on the regularization mechanism adopted.

Although our formal discussion of renormalization was based on the action functional, in a practical calculation it is not possible to follow this route. This is because in order to carry out the functional differentiation of W_{ren} with respect to $g_{\mu\nu}$ to form $\langle T_{\mu\nu} \rangle_{\text{ren}}$, it is generally necessary to know W_{ren} for all geometries $g_{\mu\nu}$. This is impossibly difficult. There is however one exception, namely the conformally trivial case, where the entire stress-tensor is determined by the conformal scaling behaviour alone. Apart from this case, it is necessary to work directly with $\langle T_{\mu\nu} \rangle$.

The number of explicit cases where analytic regularization techniques can be profitably employed is rather limited, owing to the paucity of spacetimes in which the relevant field equation can be solved in n dimensions, or divergent quantities cast into recognizable forms of generalized ζ-functions.

One interesting case in which dimensional regularization can be used to advantage, due to the high degree of geometrical symmetry, is de Sitter space. If a vacuum state is chosen that is invariant under the de Sitter group (see §5.4) then, as the only (maximally) form invariant, rank two tensor under the de Sitter group is $g_{\mu\nu}$ (see, for example, Weinberg 1972, §13.4), we must have

$$\langle T_{\mu\nu} \rangle = T g_{\mu\nu}/n \tag{6.174}$$

where T is the trace of the stress-tensor

$$T = \langle T_\mu{}^\mu \rangle. \tag{6.175}$$

For a massive scalar field, one obtains from (3.190) using the field equation (3.26)

$$T_\mu{}^\mu = m^2\phi^2 + (n-1)[\xi - \xi(n)]\Box\phi^2 \tag{6.176}$$

whence, from (2.66),

$$T = \langle T_\mu{}^\mu \rangle = \tfrac{1}{2}m^2 G^{(1)}(x,x) + \tfrac{1}{2}(n-1)[\xi - \xi(n)]\Box G^{(1)}(x,x). \tag{6.177}$$

Now $G^{(1)}(x,x)$ diverges at $n = 4$, but can be finite for $n \neq 4$. It has been obtained for all n by Candelas & Raine (1975)($\xi = 0$ only) and Dowker & Critchley (1976*b*). The result is

$$G^{(1)}(x,x) = \frac{2\alpha^2}{(4\pi\alpha^2)^{n/2}} \frac{\Gamma(\nu(n) - \frac{1}{2} + n/2)\Gamma(-\nu(n) - \frac{1}{2} + n/2)}{\Gamma(\frac{1}{2} + \nu(n))\Gamma(\frac{1}{2} - \nu(n))}\Gamma(1 - n/2) \tag{6.178}$$

where α is the radius of the de Sitter universe (see (5.49)) and

$$[\nu(n)]^2 = \tfrac{1}{4}(n-1)^2 - m^2\alpha^2 - \xi n(n-1) \tag{6.179}$$

cf. (5.60). Note that as $G^{(1)}$ is independent of x, only the first term in (6.177) will contribute to the trace:

$$T = \tfrac{1}{2}m^2 G^{(1)}(x,x). \tag{6.180}$$

The pole at $n = 4$ is manifest in (6.178). It must be removed by expanding $G^{(1)}$ about $n = 4$, and subtracting from it the adiabatic expansion of $G^{(1)}_{\text{DS}}$, truncated at an appropriate adiabatic order A, also expanded about $n = 4$. We shall denote the truncated DeWitt–Schwinger expansion by ${}^{(A)}G^{(1)}_{\text{DS}}$ in what follows. The rationale behind this step is that the low order adiabatic terms in the expansion of $G^{(1)}_{\text{DS}}$ are the ones which, through (6.37), can be used to form L_{div}. Subtraction of such terms from $G^{(1)}$ is then equivalent to renormalization of L_{eff} via (6.58).

To determine the adiabatic order A at which the expansion of $G^{(1)}_{\text{DS}}$ should be truncated, we note that fourth order adiabatic terms appear in L_{div} (in the form of $a_2(x)$). Hence we must retain in $G_{\mathrm{DS}}^{(1)}$ those terms which, when used to construct L_{div}, would yield terms up to fourth adiabatic order. In this respect one must remember that differentiation of a term will increase its adiabatic order. However, in the de Sitter space example, one sees from (6.177) and (6.180) that no derivatives are involved in computing $\langle T_{\mu\nu}\rangle$ from $G^{(1)}$. As the adiabatic order of $\langle T_{\mu\nu}\rangle$ is the same as that of L_{div}, this implies that the expansion of $G_{\mathrm{DS}}^{(1)}$ should be truncated at order 4 (in n dimensions, truncation is at adiabatic order n). Explicitly, using (3.135) and (3.137), and following the same steps as led to (6.41), one obtains

$$
\begin{aligned}
{}^{(4)}G_{\mathrm{DS}}^{(1)} = 2m^{n-4}(4\pi)^{-n/2}\{&a_0(x)m^2\Gamma(1-n/2) + a_1(x)\Gamma(2-n/2)\\
&+ a_2(x)m^{-2}\Gamma(3-n/2)\},
\end{aligned}
\tag{6.181}
$$

where, in n-dimensional de Sitter space

$$
a_0(x) = 1, \quad a_1(x) = (\tfrac{1}{6} - \xi)n(n-1)\alpha^{-2}
$$

and a_2, which is needed only in four dimensions, is given by

$$
a_2(x) = [2(1-6\xi)^2 - \tfrac{1}{15}]\alpha^{-4}, \quad (n=4).
$$

Subtracting (6.181) from (6.178), and expanding the result about $n=4$ gives

$$
\begin{aligned}
&G^{(1)}(x,x) - {}^{(4)}G_{\mathrm{DS}}^{(1)}(x,x)\\
&\quad = (1/8\pi^2\alpha^2)\{(m^2\alpha^2 + 12\xi - 2)[\psi(\tfrac{3}{2}+\nu) + \psi(\tfrac{3}{2}-\nu) - \ln(m^2\alpha^2) - 1]\\
&\qquad + m^2\alpha^2 - \tfrac{2}{3} - (\alpha/m)^2 a_2\} + O(n-4).
\end{aligned}
\tag{6.182}
$$

Now using (6.180) and (6.174), the renormalized stress-tensor at $n=4$ can be obtained (Dowker & Critchley 1976*b*):

$$
\begin{aligned}
\langle T_{\mu\nu}\rangle_{\mathrm{ren}} = (g_{\mu\nu}/64\pi^2)\{&m^2[m^2 + (\xi - \tfrac{1}{6})R][\psi(\tfrac{3}{2}+\nu)\\
&+ \psi(\tfrac{3}{2}-\nu) - \ln(12m^2R^{-1})] - m^2(\xi - \tfrac{1}{6})R\\
&- \tfrac{1}{18}m^2R - \tfrac{1}{2}(\xi - \tfrac{1}{6})^2R^2 + \tfrac{1}{2160}R^2\},
\end{aligned}
\tag{6.183}
$$

where we have used $R = 12\alpha^{-2}$. Note that in the massless, conformally coupled ($\xi - \frac{1}{6}$) limit, the trace of (6.183) as given by the renormalized version of (6.180) results entirely from the final term in (6.182):

$$
T_{\mathrm{ren}} = -a_2/16\pi^2,
$$

in agreement with (6.120).

Next we turn to an example of the use of ζ-function regularization (Dowker & Banach 1978). In §5.2 we obtained normal mode solutions of the conformally coupled massless scalar wave equation in the Einstein static universe (see (5.28)). Because the spatial sections are compact, the *total* renormalized vacuum energy due to the space curvature and non-trivial topology will be finite. Hence we may compute this quantity rather than the energy density, and then use the fact that the spatial homogeneity implies a uniform energy density, to compute $\langle 0|T_0{}^0|0\rangle$ by simply dividing out the total volume $(2\pi^2a^3)$ of space.

The total energy is given by

$$E = \int \mathrm{d}^3x\, h^{\frac{1}{2}} \langle 0|T_0{}^0|0\rangle, \tag{6.184}$$

which, for a conformally coupled, massless scalar field in the Einstein universe, is

$$E = \tfrac{1}{2}\int \mathrm{d}\mu(k)(k/a), \tag{6.185}$$

where we have used (3.190), (5.15), (5.28) and (5.19). This expression is analogous to the Minkowski space result (2.38); $\frac{1}{2}(k/a)$ being the proper energy eigenvalue of modes labelled by k. Using (5.24) we have

$$\begin{aligned} E &= \tfrac{1}{2}\sum_{k=1}^{\infty}\sum_{J=0}^{k-1}\sum_{M=-J}^{J}(k/a) \\ &= \tfrac{1}{2}\sum_{k=1}^{\infty} k^2(k/a), \end{aligned} \tag{6.186}$$

showing that the contribution to the energy from modes labelled by k has degeneracy k^2. The sum in (6.186) diverges like k^4. To regularize it we replace the energy eigenvalue $k/2a$ by $(k/2a)^{-s}$. Then

$$E = \lim_{s\to -1}\sum_{k=1}^{\infty} k^2(k/2a)^{-s}. \tag{6.187}$$

Now the sum in (6.187) is proportional to a Riemann ζ-function, namely

$$(2a)^s\zeta(s-2). \tag{6.188}$$

But this quantity, when analytically continued back to $s = -1$, is *finite*. So from (6.187)

$$E = \zeta(-3)/2a = 1/240a. \tag{6.189}$$

Dividing E by the proper volume of space, the energy density is simply

$$\rho = 1/480\pi^2 a^4, \tag{6.190}$$

and because the trace anomaly vanishes in the Einstein universe, the complete renormalized stress-tensor may be deduced from (6.190):

$$\langle 0|T_\mu{}^\nu|0\rangle_{\text{ren}} = \frac{1}{480\pi^2 a^4}\,\mathrm{diag}(1, -\tfrac{1}{3}, -\tfrac{1}{3}, -\tfrac{1}{3}). \tag{6.191}$$

This result was first obtained by Ford (1976) using a mode sum cut-off and is in agreement with (6.152) (see also Dowker & Critchley 1976*c*).

As a final example of the analytic techniques, consider the 'Casimir' energy density in flat $R^1 \times S^1$ spacetime due to a massless conformally coupled scalar field (Isham 1978*b*). The vacuum energy density is equal to (see §4.1)

$$(2\pi/L^2)\sum_{n=0}^{\infty} n \quad \text{or} \quad (\pi/L^2)\sum_{n=-\infty}^{\infty} |n+\tfrac{1}{2}| \tag{6.192}$$

for untwisted and twisted fields respectively. Replacing n and $|n+\frac{1}{2}|$ by n^{-s} and $|n+\frac{1}{2}|^{-s}$, (6.192) may be re-expressed as ζ-functions analytically continued back to $s = -1$:

$$(2\pi/L^2)\zeta(-1) = -\pi/6L^2 \tag{6.193}$$

and

$$(\pi/L^2)[\zeta(-1,\tfrac{1}{2}) + \zeta(-1,-\tfrac{1}{2}) + \tfrac{1}{2}] = \pi/12L^2, \tag{6.194}$$

respectively, where

$$\zeta(s,q) \equiv \sum_{n=0}^{\infty} (q+n)^{-s}, \quad \mathrm{Re}\, s > 1 \tag{6.195}$$

and $\zeta(-1,q)$ is given, for example, by Gradshteyn & Ryzhik (1965), (9.531). Equations (6.193) and (6.194) are in agreement with (4.8) and (4.11) obtained using cut-off methods.

Notice that in neither of the latter two examples has it been necessary to actually *renormalize* anything. The analytic continuation has simply discarded the appropriate divergent terms to yield a finite answer. Whether or not the continuation always contrives to discard exactly the right terms (i.e., those terms that, in another regularization scheme, would be absorbed into renormalization of coupling constants) has not been proved, though it is generally believed to be true.

From the point of view of practical computation, possibly the most efficient regularization technique is point-splitting. Once again, it is not in general possible to compute $\langle T_{\mu\nu}\rangle_{\text{ren}}$ by functionally differentiating a renormalized effective action. Instead, one works directly with $\langle T_{\mu\nu}\rangle$, or preferably with $G^{(1)}(x,x')$ (which is simpler). Renormalization may be carried out on $\langle T_{\mu\nu} \rangle$, by subtracting from it terms up to adiabatic order n in $\langle T_{\mu\nu} \rangle_{\text{DS}}$, which is formed by differentiation of $G^{(1)}_{\text{DS}}(x, x')$. These are the terms which arise from L_{div}, and their subtraction is equivalent to renormalization of constants in the generalized Einstein action. Alternatively, one can form $\langle T_{\mu\nu} \rangle_{\text{ren}}$ by operating on

$$G^{(1)}_{\text{ren}}(x, x') = G^{(1)}(x, x') - {}^{(n)}G^{(1)}_{\text{DS}}(x, x'), \tag{6.196}$$

with a differential operator obtained from (3.190) in a manner to be described shortly. In ${}^{(n)}G^{(1)}_{\text{DS}}$, only those terms which make a contribution to the stress-tensor of adiabatic order n or less are retained. These will be terms in $G^{(1)}_{\text{DS}}$ which are of order n or less, although care must be exercised because differentiation can increase the adiabatic order of a quantity. Thus, the procedure for computing the renormalized stress-tensor using point-splitting can be summarized as follows:

(1) Solve the field equation for a complete set of normal modes from which particle states may be defined.
(2) Construct $G^{(1)}(x, x')$ as a mode sum.
(3) Form $G^{(1)}_{\text{ren}}$ according to (6.196), truncating the expansion of $G^{(1)}_{\text{DS}}$ at order n.
(4) Operate on $G^{(1)}_{\text{ren}}$ to form $\langle 0 | T_{\mu\nu}(x, x') | 0 \rangle_{\text{ren}}$, discarding any terms of adiabatic order greater than n which have appeared from differentiation of terms in ${}^{(n)}G^{(1)}_{\text{DS}}$.
(5) Let $x' \to x$ and display the finite result $\langle 0 | T_{\mu\nu}(x) | 0 \rangle_{\text{ren}}$.

The state $|0\rangle$ will, of course, be dependent on the definition of 'positive frequency' modes at step (1).

The differentiation of $G^{(1)}(x, x')$ is generally a complicated procedure. Formally one has

$$\langle T_{\mu\nu}(x) \rangle = \lim_{x' \to x} \mathscr{D}_{\mu\nu}(x, x') G^{(1)}(x, x'). \tag{6.197}$$

The Green function $G^{(1)}(x, x')$ is not a scalar function of x, but a *bi-scalar* of the two spacetime points x and x'; that is, it transforms like a scalar at each point. (Higher spin fields involve bi-spinors, bi-vectors, etc.) Consequently the differential operator $\mathscr{D}_{\mu\nu}(x, x')$ is a non-local operator. For example, the first term of (3.190) gives rise to the expectation value

$$(1 - 2\xi) \langle 0 | \nabla_\mu \phi(x) \nabla_\nu \phi(x) | 0 \rangle, \tag{6.198}$$

which, using the point-splitting technique, is treated as

$$\lim_{x' \to x} \tfrac{1}{2}(1 - 2\xi)[\nabla_\mu \nabla_{\nu'} + \nabla_{\mu'} \nabla_\nu] \tfrac{1}{2} G^{(1)}(x, x'), \tag{6.199}$$

where the prime on a derivative indicates that it acts at x' rather than at x.

This means that the resulting object is not a tensor, but a bi-vector. (Expression (6.199) has been written in the most symmetrical way possible, although this is not essential.)

To construct a tensor from a bi-vector, and to maintain general covariance, it is necessary to parallel transport the derivative vector (spinors etc.) back to the same spacetime point, which could be the midpoint between x, x', one of the end points (e.g. x), or perhaps somewhere else. Differences between the parallel-transported and non-transported results will arise, even when the points x, x' are made to coincide, from a σ^{-1} factor in the expansion of $G^{(1)}(x, x')$ multiplying a σ-order transport correction. Fortunately, these complicated corrections have been worked out once and for all (for a general spacetime) by Christensen (1976, 1978), Davies, Fulling, Christensen & Bunch (1977), Bunch (1977) and Adler, Lieberman & Ng (1977, 1978), so we shall not go into the details here.

It should be noted, however, that if $G^{(1)}$ is renormalized first (i.e., before differentiation to form $\langle T_{\mu\nu}\rangle$) according to (6.196), then all σ^{-1} terms are in any case removed, so any transport corrections are of order σ and vanish when we let $\sigma \to 0$ at the end of the calculation. Thus, only if one insists on first constructing an unrenormalized stress-tensor will the effects of parallel transport need to be taken into account in a practical calculation.

We consider the symmetric case where one constructs $G^{(1)}(x'', x')$, the two points x'', x' lying at (small) equal proper distance from the spacetime point x of interest, on either side, along a non-null geodesic through x (see fig. 22). Symmetrization, while not essential, does simplify some of the expressions. At the end of the calculation we let both x'' and x' approach x. The (small) proper distance from x to x' is denoted ε and the direction of the geodesic is parametrized by the tangent vector t^μ at x. (If the separated points x'', x' remain in a normal neighbourhood of x, then this geodesic will be unique.)

To make use of this formalism, the first step is to convert $G^{(1)}$ into a function of ε and t^μ rather than x'' and x'. This involves solving the equation for the geodesic joining x'', x and x' in terms of ε and t^μ as a power series in ε up to order ε^{n+1}. The results are available in two important cases: two spacetime dimensions and four-dimensional Robertson–Walker spacetimes. For the two-dimensional case (Davies & Fulling 1977*b*)

$$\begin{aligned} u(\varepsilon) = u + \varepsilon t^u - \tfrac{1}{2} C^{-1} C_{,u} (t^u)^2 \varepsilon^2 \\ + \tfrac{1}{6} C[C^{-3}(3C_{,u}^2 - CC_{,uu}) t^u - \tfrac{1}{4} R t^v](t^u)^2 \varepsilon^3 + \ldots, \end{aligned} \tag{6.200}$$

where we use null coordinates u and v and the metric (6.135). A similar equation holds for $v(\varepsilon)$, with u and v interchanged. It is understood that C and its derivatives are all evaluated at the central point $x = (u, v)$. The end points x', x'' are given by $x' = (u(\varepsilon), v(\varepsilon)), x'' = (u(-\varepsilon), v(-\varepsilon))$.

For the Robertson–Walker spacetime with line element (5.7), one can choose $\theta'' = \theta' = \theta, \phi'' = \phi' = \phi$ without loss of generality, because the spacetime is isotropic. Then one requires only $\eta(\varepsilon)$ and $r(\varepsilon)$, which are given by (Bunch 1977, Bunch & Davies 1977*a*)

$$\left.\begin{aligned} \eta(\varepsilon) &= \eta + \varepsilon t^\eta + \frac{1}{2!}\varepsilon^2 t_2^\eta + \frac{1}{3!}\varepsilon^3 t_3^\eta + \frac{1}{4!}\varepsilon^4 t_4^\eta + \ldots \\ r(\varepsilon) &= r + \varepsilon t^r + \frac{1}{2!}\varepsilon^2 t_2^r + \frac{1}{3!}\varepsilon^3 t_3^r + \frac{1}{4!}\varepsilon^4 t_4^r + \ldots \end{aligned}\right\} \tag{6.201}$$

where

$$\begin{aligned}
t_2^\eta &= -\tfrac{1}{2}D(t^\eta)^2 - \tfrac{1}{2}D\Upsilon(t^r)^2 \\
t_2^r &= -Dt^\eta t^r - Kr\Upsilon(t^r)^2 \\
t_3^\eta &= (-\tfrac{1}{2}\dot{D} + \tfrac{1}{2}D^2)(t^\eta)^3 + (-\tfrac{1}{2}\dot{D} + \tfrac{3}{2}D^2)\Upsilon t^\eta(t^r)^2 \\
t_3^r &= (-\dot{D} + \tfrac{3}{2}D^2)(t^\eta)^2 t^r + 3DKr\Upsilon t^\eta(t^r)^2 + (\tfrac{1}{2}D^2 - K)\Upsilon(t^r)^3 \\
t_4^\eta &= (-\tfrac{1}{2}\ddot{D} + \tfrac{7}{4}\dot{D}D - \tfrac{3}{4}D^3)(t^\eta)^4 + (-\tfrac{1}{2}\ddot{D} + 5\dot{D}D - \tfrac{9}{2}D^3)\Upsilon(t^\eta t^r)^2 \\
&\quad + (\tfrac{1}{4}\dot{D}D - \tfrac{3}{4}D^3)\Upsilon^2(t^r)^4 \\
t_4^r &= (-\ddot{D} + 5\dot{D}D - 3D^3)(t^\eta)^3 t^r + (4\dot{D}Kr - 9D^2Kr)\Upsilon(t^\eta t^r)^2 \\
&\quad + (2\dot{D}D - 3D^3 + 6DK)\Upsilon t^\eta(t^r)^3 + (-2D^2Kr + K^2r)\,\Upsilon^2(t^r)^4 \\
t_5^\eta &= (-\tfrac{1}{2}\dddot{D} + \tfrac{11}{4}\ddot{D}D + \tfrac{7}{4}\dot{D}^2 - \tfrac{23}{4}\dot{D}D^2 + \tfrac{3}{2}D^4)(t^\eta)^5 \\
&\quad + (-\tfrac{1}{2}\dddot{D} + \tfrac{15}{2}\ddot{D}D + 5\dot{D}^2 - 32\dot{D}D^2 + 15D^4)\Upsilon(t^\eta)^3(t^r)^2 \\
&\quad + (\tfrac{3}{4}\ddot{D}D + \tfrac{1}{4}\dot{D}^2 - \tfrac{33}{4}\dot{D}D^2 + \tfrac{15}{2}D^4)\Upsilon^2 t^\eta(t^r)^4 \\
t_5^r &= (-\dddot{D} + \tfrac{15}{2}\ddot{D}D + 5\dot{D}^2 - \tfrac{43}{2}\dot{D}D^2 + \tfrac{15}{2}D^4)(t^\eta)^4 t^r \\
&\quad + (5\ddot{D}Kr - 35\dot{D}DKr + 30D^3Kr)\Upsilon(t^\eta)^3(t^r)^2 \\
&\quad + (\tfrac{7}{2}\ddot{D}D + 2\dot{D}^2 - \tfrac{47}{2}\dot{D}D^2 + 15D^4 + 10\dot{D}K - 30D^2K)\Upsilon(t^\eta)^2(t^r)^3 \\
&\quad + (-10\dot{D}DKr + 20D^3Kr - 10DK^2r)\Upsilon^2 t^\eta(t^r)^4 \\
&\quad + (-\dot{D}D^2 + \tfrac{3}{2}D^4 - 5D^2K + K^2)\Upsilon^2(t^r)^5.
\end{aligned} \tag{6.202}$$

The notation here is the same as in (5.12) and we have written $t^\eta = t^0, t^r = t^1$ etc.

Once $G^{(1)}$ has been cast into ε, t^{μ} notation, a useful check on the algebra is to ensure that the divergent terms agree with those of $G^{(1)}_{\mathrm{DS}}(x, x')$. The leading terms in $G^{(1)}_{\mathrm{DS}}$ for the scalar case are (Bunch, Christensen & Fulling 1978, Bunch 1977)

$$\begin{aligned} G^{(1)}_{\mathrm{DS}}(\varepsilon, t^{\mu}) = &-(1/\pi)\{[1-(m^2+\xi R)\varepsilon^2\,\Sigma](\gamma+\tfrac{1}{2}\ln|m^2\varepsilon^2|) \\ &+\tfrac{1}{2}(\xi-\tfrac{1}{6})m^{-2}R+\varepsilon^2\Sigma[m^2+\tfrac{1}{2}(\xi-\tfrac{1}{6})R]\} \\ &+\mathrm{O}(\varepsilon^2T^{-2})+\mathrm{O}(\varepsilon^4)+\mathrm{O}(T^{-4}), \end{aligned} \tag{6.203}$$

in two dimensions, and (Bunch 1977, Bunch & Davies 1978*b*)

$$\begin{aligned} G^{(1)}_{\mathrm{DS}}(\varepsilon, t^{\mu}) = &-\frac{1}{8\pi^2\varepsilon^2\Sigma}+\frac{m^2+(\xi-\frac{1}{6})R}{4\pi^2}(\tfrac{1}{2}\ln|m^2\varepsilon^2|+\gamma) \\ &+\frac{1}{24\pi^2}R_{\alpha\beta}\frac{t^{\alpha}t^{\beta}}{\Sigma}-\frac{m^2}{8\pi^2}+\frac{\varepsilon^2\Sigma}{24\pi^2}(\tfrac{1}{2}\ln|m^2\varepsilon^2|+\gamma)\Bigg\{-3m^4 \\ &-2m^2R_{\alpha\beta}\frac{t^{\alpha}t^{\beta}}{\Sigma}-6(\xi-\tfrac{1}{6})m^2R+(\xi-\tfrac{1}{6})\Bigg[-2RR_{\alpha\beta}\frac{t^{\alpha}t^{\beta}}{\Sigma} \\ &+R_{;\alpha\beta}\frac{t^{\alpha}t^{\beta}}{\Sigma}-\Box R-3(\xi-\tfrac{1}{6})R^2\Bigg]\Bigg\} \\ &+\frac{\varepsilon^2\Sigma}{1440\pi^2}\Bigg[225m^4+360m^2(\xi-\tfrac{1}{6})R+60m^2R_{\alpha\beta}\frac{t^{\alpha}t^{\beta}}{\Sigma} \\ &+6R_{\alpha\beta;\gamma\delta}\frac{t^{\alpha}t^{\beta}t^{\gamma}t^{\delta}}{\Sigma^2}-14R_{\alpha\beta}R_{\gamma\delta}\frac{t^{\alpha}t^{\beta}t^{\gamma}t^{\delta}}{\Sigma^2} \\ &+4R_{\alpha\rho}R^{\rho}{}_{\beta}\frac{t^{\alpha}t^{\beta}}{\Sigma}-\tfrac{4}{3}RR_{\alpha\beta}\frac{t^{\alpha}t^{\beta}}{\Sigma}-\Box R-R^{\alpha\beta}R_{\alpha\beta} \\ &+\tfrac{1}{3}R^2+30(\xi-\tfrac{1}{6})\Box R+90(\xi-\tfrac{1}{6})^2R^2\Bigg] \\ &+\frac{1}{1440\pi^2m^2}[R^{\alpha\beta}R_{\alpha\beta}-\tfrac{1}{3}R^2-\Box R+30(\xi-\tfrac{1}{6})\Box R \\ &+90(\xi-\tfrac{1}{6})^2R^2]+\mathrm{O}(\varepsilon^4)+\mathrm{O}(T^{-6}) \end{aligned} \tag{6.204}$$

in conformally flat four-dimensional spacetimes. In these expressions, t^{μ} is normalized according to (6.85), and, as previously, $\mathrm{O}(T^{-p})$ indicates terms of adiabatic order p (i.e., p derivatives of the metric).

Renormalization of $G^{(1)}(x'', x')$ is now immediate. Writing it as a function of ε, t^{μ}, one simply subtracts from it all of (6.203) for two dimensions or

(6.204) for four dimensions. However, one must not yet take the limit $\varepsilon \to 0$, even though $G^{(1)}_{\text{ren}}(x'', x')$ remains finite as $x'', x' \to x$. First it is necessary to differentiate to form $\langle 0|T_{\mu\nu}(x'', x')|0\rangle_{\text{ren}}$, discarding any terms of higher adiabatic order than n that result from differentiation of terms from (6.203) or (6.204).

For the general form

$$G^{(1)} = c + \varepsilon^2 \Sigma [e_{\alpha\beta}(t^\alpha t^\beta/\Sigma) + f] + \varepsilon^2 \Sigma [q_{\alpha\beta}(t^\alpha t^\beta/\Sigma) + r](\gamma + \tfrac{1}{2}\ln|\alpha^2\varepsilon^2|) \tag{6.205}$$

where α and γ are constants, and $c, f, r, e_{\alpha\beta}$ and $q_{\alpha\beta}$ can be functions of x, the resulting stress-tensor is (Bunch 1977; see also Davies, Fulling, Christensen & Bunch 1977 and Bunch & Davies 1978*b*) in the coincidence limit

$$\begin{aligned}
\langle T_{\mu\nu}\rangle = {} & (\tfrac{1}{2} - \xi)T^{(1)}_{\mu\nu} + \left[\left(\frac{n-1}{n}\right)(\xi - \xi(n)) - \frac{1}{n}(\tfrac{1}{2} - \xi)\right] g_{\mu\nu} T^{(1)\sigma}_{\ \ \ \sigma} \\
& + \xi T^{(2)}_{\mu\nu} - \frac{1}{n} g_{\mu\nu} \xi T^{(2)\sigma}_{\ \ \ \sigma} \\
& - \tfrac{1}{2}c\xi \left[R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} - \left(1 - \frac{n}{2}\right)\frac{1}{n}Rg_{\mu\nu} \right. \\
& \left. + \frac{2(n-1)}{n}(\xi - \xi(n))Rg_{\mu\nu} \right] \\
& + \left[\frac{1}{2n} - \left(\frac{n-1}{n}\right)(\xi - \xi(n))\right] m^2 g_{\mu\nu} c
\end{aligned} \tag{6.206}$$

where

$$\begin{aligned}
T^{(1)}_{\mu\nu} \equiv {} & \langle 0|\{\phi_{;\mu}, \phi_{;\nu}\}|0\rangle \\
= {} & -\tfrac{1}{2}(q_{\mu\nu} + rg_{\mu\nu})[\gamma + \tfrac{1}{2}\ln|\alpha^2 m^{-2}|] \\
& - \tfrac{1}{4}q_{\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma}\left(g_{\mu\nu} - \frac{2t_\mu t_\nu}{\Sigma}\right) - q_{(\mu\alpha}t^\alpha t_{\nu)}\Sigma^{-1} \\
& - \tfrac{1}{4}r\left(g_{\mu\nu} + \frac{2t_\mu t_\nu}{\Sigma}\right) + \tfrac{1}{4}c_{;\mu\nu} - \tfrac{1}{2}e_{\mu\nu} - \tfrac{1}{2}fg_{\mu\nu}
\end{aligned} \tag{6.207}$$

and

$$T^{(2)}_{\mu\nu} = -\langle 0|\{\phi, \phi_{;\mu\nu}\}|0\rangle = T^{(1)}_{\mu\nu} - \tfrac{1}{2}c_{;\mu\nu}. \tag{6.208}$$

In the conformally coupled case, $\xi = \xi(n)$, (6.206) manifestly has the trace $\frac{1}{2}m^2 c$. Note that terms of order ε^2 in $G^{(1)}$ make a finite contribution, when differentiated, to $\langle T_{\mu\nu}\rangle$.

Inspection of (6.206) shows that the only terms in (6.203) and (6.204) which will produce a contribution to $\langle T_{\mu\nu}\rangle_{\rm ren}$ of adiabatic order greater than n are those of order m^{-2}. The only contribution from these terms which is of sufficiently low adiabatic order not to be discarded by the above prescription comes from their substitution as c in the final term of (6.206). Thus the only contribution of the terms of order m^{-2} to $\langle T_{\mu\nu}\rangle_{\rm ren}$ are of order m^0, and there is no difficulty in taking the massless limit of $\langle T_{\mu\nu}\rangle_{\rm ren}$ if this is desired. The mass must be kept nonzero until after differentiation to avoid infrared divergences at intermediate steps of the calculation.

As an illustration of this rather cumbersome procedure let us evaluate $\langle T_{\mu\nu}\rangle_{\rm ren}$ for a massless, conformally coupled scalar field in two-dimensional spacetime in the conformal vacuum state. (Several other examples will be given in the next chapter.) This conformally trivial case has already been solved in §6.3 (see (6.136)). So as to provide a case in which the first term on the right-hand side of (6.136) is nonzero, we shall include compactification of the spatial sections. The Green functions are given by (4.23) and (4.24) for the twisted and untwisted fields respectively:

$$D_{\rm L}^{(1)}(x'',x') = \begin{cases} -(1/4\pi)\ln[16\sin^2(\pi\Delta u/L)\sin^2(\pi\Delta v/L)] & \text{untwisted} \\ -(1/4\pi)\ln[\tan^2(\pi\Delta u/2L)\tan^2(\pi\Delta v/2L)] & \text{twisted} \end{cases}$$

where $\Delta u = u(\varepsilon) - u(-\varepsilon)$, $\Delta v = v(\varepsilon) - v(-\varepsilon)$. Although when used in §4.2. these Green functions referred to flat spacetime, they remain the same in the present example (see (3.154)) because (i) the spacetime is conformally flat and (ii) the conformal weight for the scalar field in two dimensions is zero.

Using (6.200) one immediately obtains power series expansions in ε for $D_{\rm L}^{(1)}(x'',x')$:

$$\begin{aligned} &\text{constant} - (1/2\pi)\ln|\varepsilon^2 C^{-1}| + (\varepsilon^2/2\pi)\{\tfrac{1}{12}R\Sigma + \tfrac{2}{3}(\alpha\pi^2/L^2)[(t^u)^2 + (t^v)^2] \\ &\qquad + \tfrac{1}{6}C^{-2}(CC_{,uu} - 3C_{,u}^2)(t^u)^2 \\ &\qquad + \tfrac{1}{6}C^{-2}(CC_{,vv} - 3C_{,v}^2)(t^v)^2\} + O(\varepsilon^4), \end{aligned}$$

where $\alpha = 1$ for the untwisted field and $-\frac{1}{2}$ for the twisted field, and we have used the normalization condition (6.85), which in the metric (6.135) has the form

$$Ct^u t^v = \Sigma. \tag{6.209}$$

To renormalize, we now subtract (6.203), setting $\xi = 0$ (conformal coupling), and $m = 0$ except in the $m^{-2}R$ term (see remarks above). The remainder, aside from the $m^{-2}R$ term, is

$$\begin{aligned}\text{constant} &+ (1/2\pi)\ln C + \varepsilon^2\Sigma\{(\pi\alpha/3L^2)\Sigma^{-1}[(t^u)^2 + (t^v)^2] \\ &+ (1/12\pi)\Sigma^{-1}C^{-2}(CC_{,uu} - 3C_{,u}^2)(t^u)^2 \\ &+ (1/12\pi)\Sigma^{-1}C^{-2}(CC_{,vv} - 3C_{,v}^2)(t^v)^2 \\ &- (1/24\pi)R\} + O(\varepsilon^2).\end{aligned}$$

One may identify scalar functions c and f and a traceless tensor $e_{\alpha\beta}$ where

$$\begin{aligned} e_{uu} &= (\pi\alpha/3L^2) + \tfrac{1}{12\pi}C^{-2}(CC_{,uu} - 3C_{,u}^2) \\ e_{vv} &= (\pi\alpha/3L^2) + \tfrac{1}{12\pi}C^{-2}(CC_{,vv} - 3C_{,v}^2) \\ e_{uv} &= e_{vu} = 0 \end{aligned}$$

to be inserted in the general formula (6.205). All other terms are zero.

Reading off the relevant terms from (6.206)–(6.208), with $n = 2$, one obtains

$$\langle 0|T_{uu}|0\rangle_{\text{ren}} = -\frac{\pi\alpha}{12L^2} + \frac{1}{24\pi}\left[\frac{C_{,uu}}{C} - \frac{3C_{,u}^2}{2C^2}\right] \tag{6.210}$$

$$\langle 0|T_{vv}|0\rangle_{\text{ren}} = -\frac{\pi\alpha}{12L^2} + \frac{1}{24\pi}\left[\frac{C_{,vv}}{C} - \frac{3}{2}\frac{C_{,v}^2}{C^2}\right]. \tag{6.211}$$

The $-R/12\pi m^2$ term does not contribute to these stress-tensor components when inserted as c in the final term in (6.206) (since $g_{uu} = g_{vv} = 0$) but it does give rise to components

$$\begin{aligned}\langle 0|T_{uv}|0\rangle_{\text{ren}} = \langle 0|T_{vu}|0\rangle_{\text{ren}} &= \tfrac{1}{4}C\langle 0|T_{\alpha}^{\alpha}|0\rangle_{\text{ren}} \\ &= -RC/96\pi. \end{aligned} \tag{6.212}$$

Clearly the m^{-2} term gives rise to the trace anomaly. These results are in complete agreement with (6.136), when the zero temperature (leading) term of (4.27)–(4.29) is used for $\langle T_{\mu}^{\nu}[\eta_{\kappa\lambda}]\rangle_{\text{ren}}$.

In the above we have discussed how $\langle T_{\mu\nu}\rangle_{\text{ren}}$ can be obtained by forming $\langle T_{\mu\nu}\rangle$, suitably regularized, and subtracting from it $\langle T_{\mu\nu}\rangle_{\text{DS}}$, formed from terms in the DeWitt–Schwinger expansion. In practice this was achieved by first constructing $G^{(1)}$, suitably regularized, subtracting from it terms in $G^{(1)}_{\text{DS}}$ and then differentiating to form the stress-tensor. The terms that were subtracted were chosen to be equivalent to renormalization of coefficients in the generalized Einstein *action*, which resulted in the subtraction of all terms in $\langle T_{\mu\nu}\rangle_{\text{DS}}$ up to adiabatic order n (in n dimensions).

It may appear perverse to compute a renormalized $\langle T_{\mu\nu}\rangle$ by consideration of $G^{(1)}$ and L_{eff}, rather than the gravitational field equations themselves, in which $\langle T_{\mu\nu}\rangle$ appears on the right-hand side. One could, after all, attempt to absorb the divergent pieces of $\langle T_{\mu\nu}\rangle$ directly into geometrical terms on the left-hand side of the generalized Einstein equation, given by (6.95), by renormalizing the constants G, Λ, a and b. There is, however, a subtlety involved in proceeding this way, the illumination of which is useful for understanding the origin of the trace anomaly.

First consider dimensional regularization. Repeating the considerations of §6.2 for an arbitrary field having

$$\left.\begin{aligned} \operatorname{tr} a_0 &= s \\ \operatorname{tr} a_1 &= zR \\ \operatorname{tr} a_2 &= wR^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta} + xR^{\alpha\beta}R_{\alpha\beta} + yR^2 \end{aligned}\right\} \tag{6.213}$$

one obtains

$$\begin{aligned} \langle T_{\mu\nu}\rangle_{\text{div}} = -\frac{1}{(4\pi)^{-n/2}}&\left[\frac{1}{n-4} + \frac{1}{2}\left(\gamma + \ln\left(\frac{m^2}{\mu^2}\right)\right)\right] \\ &\times\left\{-\frac{4sm^4}{n(n-2)}g_{\mu\nu} - \frac{4m^2 z}{(n-2)}(R_{\mu\nu} - \tfrac{1}{2}g_{\mu\nu}R)\right. \\ &\left.\quad + 2wH_{\mu\nu} + 2x\,{}^{(2)}H_{\mu\nu} + 2y\,{}^{(1)}H_{\mu\nu}\right\}. \end{aligned} \tag{6.214}$$

These terms may be removed from $\langle T_{\mu\nu}\rangle$ by absorbing them into terms in generalized Einstein equation in n dimensions, the left-hand side of which is given by (6.52). Comparison of (6.213) with (6.123) and (6.124) gives w, x, y in terms of the anomaly coefficients a, b, c, d for a field transforming under the (A, B) representation of the Lorentz group:

$$\left.\begin{aligned} w &= -(180)^{-1}(-1)^{2A+2B}a \\ x &= -(180)^{-1}(-1)^{2A+2B}(b-2a) \\ y &= -(180)^{-1}(-1)^{2A+2B}(d - \tfrac{1}{3}b + \tfrac{1}{3}a). \end{aligned}\right\} \tag{6.215}$$

It is convenient to rewrite (6.214) as

$$\langle T_{\mu\nu}\rangle_{\rm div} = -\frac{1}{(4\pi)^{n/2}}\left[\frac{1}{n-4}+\frac{1}{2}\left(\gamma+\ln\left(\frac{m^2}{\mu^2}\right)\right)\right]$$
$$\times\left\{-\frac{4sm^4}{n(n-2)}g_{\mu\nu} - wC^{\alpha\beta\gamma\delta}C_{\alpha\beta\gamma\delta}g_{\mu\nu} + 4wC_{\mu\alpha\beta\gamma}C^{\alpha\beta\gamma}{}_{\nu}\right.$$
$$-\frac{4zm^2}{(n-2)}(R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R) + 2(4w+x)A_{\mu\nu}$$
$$\left.+{}^{(1)}H_{\mu\nu}[y+\tfrac{1}{3}(w+x)]\right\} - (4\pi)^{-n/2}(4w+x)$$
$$\times\left\{-\frac{{}^{(1)}H_{\mu\nu}}{6(n-1)} + 2{}^{(3)}H_{\mu\nu}\left[\frac{(n-2)}{(n-3)}-\frac{w}{4w+x}\right]\right\}, \tag{6.216}$$

(Bunch 1979) where the *traceless tensor* $A_{\mu\nu}$ is defined by

$$A_{\mu\nu} = -\frac{n}{4(n-1)}{}^{(1)}H_{\mu\nu} + {}^{(2)}H_{\mu\nu} - \frac{(n-2)(n-4)}{4(n-3)}{}^{(3)}H_{\mu\nu} \tag{6.217}$$

and ${}^{(3)}H_{\mu\nu}$ is the generalization to n-dimensional, non-conformally flat spacetimes of (6.140):

$${}^{(3)}H_{\mu\nu} = \frac{4(n-3)}{(n-2)^2}R^{\rho}_{\mu}R_{\rho\nu} - \frac{2n(n-3)}{(n-1)(n-2)^2}RR_{\mu\nu}$$
$$-\frac{2(n-3)}{(n-2)^2}R_{\rho\sigma}R^{\rho\sigma}g_{\mu\nu}$$
$$+\frac{(n+2)(n-3)}{2(n-1)(n-2)^2}R^2g_{\mu\nu} + \frac{4}{(n-2)}C_{\rho\mu\sigma\nu}R^{\rho\sigma}. \tag{6.218}$$

In obtaining (6.217) we have also used the identity

$$C_{\alpha\beta\gamma\delta} = R_{\alpha\beta\gamma\delta} - (n-2)^{-1}(g_{\alpha\delta}R_{\beta\gamma} + g_{\beta\gamma}R_{\alpha\delta} - g_{\alpha\gamma}R_{\beta\delta} - g_{\beta\delta}R_{\alpha\gamma})$$
$$+ (n-1)^{-1}(n-2)^{-1}(g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma})R. \tag{6.219}$$

In the conformally invariant case (6.216) simplifies because we may use the relations (6.127). This yields $w+x = -3y$, causing the divergent term proportional to ${}^{(1)}H_{\mu\nu}$ in (6.216) to vanish. If we also pass to the massless limit, the only surviving terms that diverge as $n\to 4$ are the $A_{\mu\nu}$ and Weyl tensor terms. The former is traceless, while the trace of the latter is $O(n-4)$.

This combines with the factor $(n-4)^{-1}$ to yield a finite trace. Adding it to the traces of the (finite) ${}^{(1)}H_{\mu\nu}$ and ${}^{(3)}H_{\mu\nu}$ terms in (6.216) yields the negative of the familiar trace anomaly. It follows that when $\langle T_{\mu\nu}\rangle_{\text{div}}$ is subtracted from the traceless $\langle T_{\mu\nu}\rangle$ to yield $\langle T_{\mu\nu}\rangle_{\text{ren}}$, the latter will acquire an anomalous trace. There is thus no obstacle to dimensionally renormalizing $\langle T_{\mu\nu}\rangle$ in the gravitational field equations themselves.

The situation is different, however, if one tries to follow this route using point-splitting regularization. In four dimensions the left-hand side of the generalized Einstein equation is

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} + \alpha\,{}^{(1)}H_{\mu\nu} + \beta\,{}^{(2)}H_{\mu\nu}. \tag{6.220}$$

The tensor $H_{\mu\nu}$ is absent in (6.220) because in four dimensions the relation (6.57) can be employed. Using (6.81) to form the point-split regularized effective action, and functionally differentiating to form $\langle T_{\mu\nu}\rangle_{\text{div}}$, one finds that all the divergent terms can indeed be removed by renormalization of coefficients in the generalized Einstein equation with the left-hand side (6.220). There is, however, a problem. In the massless, conformally coupled limit, one obtains (Christensen 1978)

$$\begin{aligned}\langle T_{\mu\nu}\rangle_{\text{div}} = {} & -(1/480\pi^2)({}^{(2)}H_{\mu\nu} - \tfrac{1}{3}{}^{(1)}H_{\mu\nu})(\gamma + \tfrac{1}{2}\ln|\tfrac{1}{2}\mu^2\sigma|) \\ & + (\text{geometrical terms that remain finite as } \sigma \to 0),\end{aligned} \tag{6.221}$$

where we have once again averaged over splitting vector directions and μ is an arbitrary infrared cut-off parameter.

The first term in (6.221) may be removed by renormalization of α and β in (6.220), but one finds that the remaining (finite) terms may not. However, from (6.53) and (6.54), the trace of the divergent terms in (6.221) is proportional to $\Box R$ and so its removal cannot give the full trace anomaly (6.113)–(6.115). One finds that the finite geometrical terms in (6.221), which cannot be removed by renormalization, are precisely those which make up the remainder of the anomaly. In contrast, if the *action* is renormalized, then the whole of W_{div}, including the parts which give the finite terms in (6.221), is removed from W, leading to the usual trace anomaly. Thus the resulting $\langle T_{\mu\nu}\rangle_{\text{ren}}$ depends crucially on whether renormalization is carried out before or after differentiation to obtain the stress-tensor.

Closer examination shows why this is so. The renormalized gravitational action is

$$\begin{aligned}S_{\text{grav,ren}} = \int & (-g)^{\frac{1}{2}}\{(16\pi G)^{-1}(R - 2\Lambda) + aR^2 \\ & + bR^{\alpha\beta}R_{\alpha\beta} + cR^{\alpha\beta\gamma\delta}R_{\alpha\beta\gamma\delta}\}\,\mathrm{d}^4x.\end{aligned}$$

However, a, b, and c (which should not be confused with the coefficients in (6.124)) are not strictly constants in the point-splitting regularization scheme. They contain σ-dependent divergences, such as σ^{-2} and $\ln \sigma$, that arise from (6.81), and σ is both a function of x and a functional of the metric. If the regularization is relaxed so that the points come together and these terms diverge, then this x-dependence is removed. That is what happens when renormalization of the action is undertaken. On the other hand, if renormalization of the gravitational field equation is required then, as with dimensional regularization, the regularization can only be relaxed after the renormalization has been effected, otherwise terms will be lost. Thus, it is necessary to functionally differentiate W with $\sigma \neq 0$. Consequently W is a non-local functional of the geometry, because σ is a bi-scalar $\sigma(x,x')$. Additional finite terms come from the differentiation of the σ-dependence, these being precisely the geometrical terms in (6.221) (in the massless, conformally coupled scalar case) and supplying the missing terms in the anomalous trace.

To see this explicitly, consider the portion of the massless effective action that comes from the final term of L_{div} in (6.81):

$$-(1/32\pi^2)\int(-g)^{\frac{1}{2}}\, a_2(x,x')\ln|\tfrac{1}{2}m^2\sigma(x,x')|\,\mathrm{d}^4x. \tag{6.222}$$

Here m is kept nonzero to act as an infrared cut-off. If the action is renormalized, and $\sigma\to 0$, this term is removed from W, a crucial step that yields an anomalous trace proportional to $a_2(x,x)$. However, if we functionally differentiate first, we obtain a finite contribution to $\langle T_{\mu\nu}\rangle$ from σ:

$$(1/64\pi^2)a_2(x)g_{\mu\nu}, \tag{6.223}$$

where we have used (Christensen 1978)

$$\frac{\delta\sigma(x,x')}{\delta g^{\mu\nu}} = -\tfrac{1}{2}\sigma_\mu\sigma_\nu + \mathrm{O}(\sigma^3{}_\mu)$$

and averaged over the σ_μ. If it is argued that, being finite, and not having the form of any of the tensors in (6.220), this term should be retained in $\langle T_{\mu\nu}\rangle_{\text{ren}}$, then $\langle T_{\mu\nu}\rangle_{\text{ren}}$ will remain traceless. If, however, it is regarded as part of the divergences (having come from the functional differentiation of a divergent term in L_{div}), then it must simply be discarded, yielding the correct trace $\langle T_{\mu\nu}\rangle_{\text{ren}}$. This is the procedure advocated in the steps on page 195, since the term (6.223) is of adiabatic order four and, by steps (3) and (4), is to be subtracted.

Notice that if one is prepared to contemplate a non-local action even *after* regularization has been removed, then it is possible, as the above calculation indicates, to add any multiple of $a_2(x)$ to the trace $\langle T_{\mu\nu}\rangle_{\mathrm{ren}}$. As remarked, Brown & Dutton (1978) have argued that a non-local action should be used to remove the conformal anomaly entirely.

6.5 Other regularization methods

Although point-splitting and analytic methods are the most developed regularization schemes, a number of other methods have been proposed. In Robertson–Walker spacetimes these methods are all related to the so-called adiabatic regularization scheme of Parker & Fulling (1974), which we shall consider first.

Adiabatic regularization is in fact a misnomer for a *subtraction* scheme which can encompass point-splitting, dimensional regularization or any other regularization method. Instead of subtracting the adiabatic expansion of $G^{(1)}_{\mathrm{DS}}$ to form a renormalized Green function, as in the previous section, adiabatic regularization works with subtractions based on the adiabatic expansion of the *modes*. The possibility of doing this arises from the relation, discussed in §3.6,

$$^{(A)}G^{(1)}_{\mathrm{DS}}(x'', x') = \langle 0^A|\{\phi(x''), \phi(x')\}|0^A\rangle|^{(A)}, \tag{6.224}$$

where the left-hand side is the expansion to adiabatic order A of $G^{(1)}_{\mathrm{DS}}$, and the right-hand side is the expansion to order A of the expectation value of a field in an Ath order adiabatic vacuum. Since the right-hand side is only required to order A, it is possible to replace the exact field ϕ by its expansion to order A, $\phi^{(A)}$, which can be formed as a mode expansion in terms of the modes $u^A_{\mathbf{k}}$ of §3.5. Thus

$$\begin{aligned} ^{(A)}G^{(1)}_{\mathrm{DS}}(x'',x') &= \langle 0^A|\{\phi^{(A)}(x''), \phi^{(A)}(x')\}|0^A\rangle|^{(A)} \\ &= \int d\tilde{\mu}(k)\{u^{(A)}_{\mathbf{k}}(x'')u^{(A)*}_{\mathbf{k}}(x') + u^{(A)*}_{\mathbf{k}}(x'')u^{(A)}_{\mathbf{k}}(x')\}|^{(A)}, \end{aligned} \tag{6.225}$$

in which the measure is given by (5.24). In the case of $K = \pm 1$ Robertson–Walker spacetimes with spatial curvature, the adiabatic formalism goes through with only obvious changes from the $K = 0$ case discussed in §3.5. The symbol $|^{(A)}$ in (6.225) indicates that the cross terms from the field products that are of adiabatic order greater than A are to be discarded.

As remarked above, point-splitting regularization and renormalization consists of subtracting what is really an adiabatic expansion, either from $\mathscr{L}_{\text{eff}}$, or $\langle T_{\mu\nu} \rangle$, or from $G^{(1)}$. In the latter case, the left-hand side of (6.225), with $A = n$, is expressed in the form of a DeWitt–Schwinger series, and subtracted from $G^{(1)}(x'', x')$. However, one could work instead with the right-hand side, i.e., the adiabatic mode sum, and subtract this from $G^{(1)}$ instead. The advantage of working directly with adiabatic modes (rather than an adiabatic expansion of an exact $G^{(1)}$) is that in some cases it may be possible to arrange for explicit cancellation of divergences *before* the mode integral is even performed, in which case potentially divergent terms never appear in the calculation; only finite integrals are involved. Thus, no actual regularization at all will be necessary. The points x'' and x' can be brought together *ab initio.* (If $\langle T_{\mu\nu} \rangle$ is required then differentiation of $G^{(1)}(x'', x')$ must be performed before x'', $x' \to x$.)

To illustrate this so called 'adiabatic regularization' technique, consider a massless, conformally coupled scalar field in two-dimensional Robertson–Walker spacetime with R^2 topology (Bunch 1977, 1978*b*). Following the treatment of §§4.1, 4.2, using the conformal vacuum, one obtains

$$\langle 0|T_{uu}|0\rangle = \langle 0|T_{vv}|0\rangle = \int \frac{\partial u_k}{\partial u}\frac{\partial u_k^*}{\partial u}\mathrm{d}k = \frac{1}{4\pi}\int_0^\infty k\,\mathrm{d}k \tag{6.226}$$

$$\langle 0|T_{uv}|0\rangle = \langle 0|T_{vu}|0\rangle = 0, \tag{6.227}$$

in which we have already put the points together in anticipation of ultimately having a finite mode integral to perform.

In two dimensions the adiabatic expansion of the modes is given by (see e.g. (3.86))

$$u_k^{(A)}(x) = (2\pi)^{-\frac{1}{2}}\mathrm{e}^{ikx}\chi_k^{(A)}(\eta), \tag{6.228}$$

where $\chi_k^{(A)}$ is given by (3.102) expanded to order A. Using (6.228) in (6.225), and differentiating to find the stress-tensor, one has

$$\langle 0^A|T_{uu}|0^A\rangle\Big|^{(A)} = \frac{1}{16\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{W_k^{(A)}}\left[(W_k^{(A)}+k)^2 + \tfrac{1}{4}\left(\frac{\dot{W}_k^{(A)}}{W_k^{(A)}}\right)^2\right]\Bigg|^{(A)} \tag{6.229}$$

$$\langle 0^A|T_{vv}|0^A\rangle\Big|^{(A)} = \frac{1}{16\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{W_k^{(A)}}\left[(W_k^{(A)}-k)^2 + \tfrac{1}{4}\left(\frac{\dot{W}_k^{(A)}}{W_k^{(A)}}\right)^2\right]\Bigg|^{(A)} \tag{6.230}$$

$$\langle 0^A|T_{uv}|0^A\rangle\Big|^{(A)} = \langle 0^A|T_{vu}|0^A\rangle\Big|^{(A)} = \frac{Cm^2}{16\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{W_k^{(A)}}\Bigg|^{(A)} \tag{6.231}$$

According to our adiabatic subtraction prescription (which is motivated by the desire to achieve equivalence with a result obtained by renormalization of the effective Lagrangian), we subtract (6.229)–(6.231) with $A=2$, from the respective quantities in (6.226) and (6.227), obtaining $W_k^{(2)}$ by iteration as described on page 67. One finds

$$W_k^{(2)} = \omega_k - \frac{\ddot{C}m^2}{8\omega_k^3} + \frac{5\dot{C}^2m^4}{32\omega_k^5}, \tag{6.232}$$

which, when substituted in (6.229) and (6.230) yields

$$\langle 0^{A=2}|T_{uu}|0^{A=2}\rangle|^{(2)} = \langle 0^{A=2}|T_{vv}|0^{A=2}\rangle|^{(2)}$$
$$= \frac{1}{8\pi}\int_0^\infty \left[\omega_k + \frac{k^2}{\omega_k} - \frac{1}{8}\frac{\ddot{C}m^2}{\omega_k^3} + \frac{1}{8}\frac{\ddot{C}k^2m^2}{\omega_k^5} + \frac{7}{32}\frac{\dot{C}^2m^4}{\omega_k^5} - \frac{5}{32}\frac{\dot{C}^2k^2m^4}{\omega_k^7}\right]\mathrm{d}k. \tag{6.233}$$

Subtracting this from (6.226) one obtains

$$\langle T_{uu}\rangle_{\text{ren}} = \langle T_{vv}\rangle_{\text{ren}} = -\frac{1}{8\pi}\int_0^\infty \frac{(\omega_k - k)^2}{\omega_k}\mathrm{d}k$$
$$+\frac{1}{64\pi}\int_0^\infty \left(\frac{\ddot{C}m^2}{\omega_k^3} - \frac{\ddot{C}k^2m^2}{\omega_k^5} - \frac{7}{4}\frac{\dot{C}^2m^4}{\omega_k^5} + \frac{5}{4}\frac{\dot{C}^2k^2m^4}{\omega_k^7}\right)\mathrm{d}k, \tag{6.234}$$

in which the limit $m \to 0$ is to be taken at the end. As expected, both integrals are *finite*. The first vanishes as $m \to 0$, while the second integral yields

$$\langle T_{uu}\rangle_{\text{ren}} = \langle T_{vv}\rangle_{\text{ren}} = \frac{1}{96\pi}\left(\frac{\ddot{C}}{C} - \frac{3\dot{C}^2}{2C^2}\right) = \theta_{uu} = \theta_{vv}, \tag{6.235}$$

with $\theta_{\mu\nu}$ given by (6.137).

Similarly, one obtains

$$\langle T_{uv}\rangle_{\text{ren}} = \langle T_{vu}\rangle_{\text{ren}} = -\frac{C\ddot{C}m^4}{64\pi}\int_0^\infty \frac{\mathrm{d}k}{\omega_k^5} + \frac{5C\dot{C}^2m^6}{256\pi}\int_0^\infty \frac{\mathrm{d}k}{\omega_k^7}$$
$$= -\frac{1}{96\pi}\left(\frac{\ddot{C}}{C} - \frac{\dot{C}^2}{C^2}\right) = -\frac{R}{48\pi}g_{\mu\nu}. \tag{6.236}$$

Hence,

$$\langle T_{\mu\nu}\rangle_{\text{ren}} = \theta_{\mu\nu} - (48\pi)^{-1}Rg_{\mu\nu}, \tag{6.237}$$

which is a special case of (6.136) and (6.210)–(6.212) that were obtained by different methods. The analogous four-dimensional calculation has also been performed by Bunch (1977, 1978*b*), giving agreement with (6.141).

In most cases the exact mode solutions will be too complicated to allow cancellation of potentially divergent terms before the mode sum is performed. In such cases a true regularization scheme must be used. If point-splitting or dimensional regularization is employed, then the method is precisely the same as we have used in previous sections. However, the relation (6.225) has the advantage of allowing other regularizations to be used, with the knowledge that the final result will be identical to that obtained by renormalization of constants in the generalized Einstein equation. The regularization is merely employed as a convenience in the manipulation of a quantity which is known already to be finite. In particular, one can use regularizations which are not manifestly covariant, but which are calculationally more convenient than the covariant schemes. For example, Birrell (1978) has used an exponential cut-off $e^{-\alpha k}$ to greatly facilitate calculations, while Bunch (1980*b*) has used a (non-covariant) form of dimensional regularization to examine the divergences in $\langle 0^A | T_{\mu\nu} | 0^A \rangle$. Alternatively, Hu (1979) has exploited the form of (6.225) as a mode sum to formally parametrize $\langle 0^A | T_{\mu\nu} | 0^A \rangle$ in terms of a single simple logarithmically divergent integral, a method that he has also used in Bianchi I spacetimes (Hu 1978) using the formulation of Parker, Fulling & Hu (1974).

Finally, the adiabatic regularization scheme is particularly useful for numerical calculations (Birrell 1978), where one may proceed as follows: Differentiating (6.225) to form the stress-tensor, and letting the points come together, one has (e.g., (6.229)–(6.231))

$$\langle 0^A | T_{\mu\nu}(x) | 0^A \rangle |^{(A)} = \int \mathrm{d}\tilde{\mu}(k) T^{(A)}_{\mu\nu}(\mathbf{k}; x) \tag{6.238}$$

where $T^{(A)}_{\mu\nu}(\mathbf{k}, x)$ is given in terms of the modes $u^{(A)}_{\mathbf{k}}(x)$. Similarly, the expectation value of the exact stress-tensor in the spacetime of interest can be written as (e.g., (6.226), (6.227))

$$\langle T_{\mu\nu}(x) \rangle = \int \mathrm{d}\tilde{\mu}(k) T_{\mu\nu}(\mathbf{k}; x), \tag{6.239}$$

where $T_{\mu\nu}(\mathbf{k}; x)$ is given in terms of the exact modes $u_{\mathbf{k}}(x)$. Then the renormalized expectation value is given by

$$\begin{aligned} \langle T_{\mu\nu} \rangle_{\mathrm{ren}} &= \langle T_{\mu\nu}(x) \rangle - \langle 0^A | T_{\mu\nu}(x) | 0^A \rangle |^{(A)} \\ &= \int \mathrm{d}\tilde{\mu}(k) [T_{\mu\nu}(\mathbf{k}, x) - T^{(A)}_{\mu\nu}(\mathbf{k}, x)] \end{aligned} \tag{6.240}$$

where $A = n$ in an n-dimensional spacetime. The portion $T_{\mu\nu}(\mathbf{k}, x)$ can be calculated numerically without difficulty given the modes and their derivatives (either from analytic expressions or from a numerical solution of the field equation), while $T_{\mu\nu}^{(A)}$ with $A = 2$ or 4 is trivially evaluated. Since the potentially divergent terms in (6.240) must cancel by construction, the integral is finite, and can be computed using straightforward quadrature methods. Numerical calculation is impossible using point-splitting or dimensional regularization because of the inability to take numerically the limit $x'' \to x'$ or $n \to n_0$ (the physical dimension) respectively.

Closely related to adiabatic regularization are the n-wave and Pauli–Villars regularization schemes. The n-wave regularization method was introduced by Zel'dovich & Starobinsky (1971) as a variant of the Pauli–Villars scheme, and has been shown by Parker & Fulling (1974) to be equivalent to adiabatic regularization. Here, we shall only discuss the Pauli–Villars method proper.

This method has long been used in Minkowski space interacting quantum field theory as a regularization technique (Pauli & Villars 1949). The basic idea is to augment the physical field Lagrangian of interest with contributions from additional fictitious fields, and arrange the parameters of the theory such that the divergent terms from the fictitious fields exactly cancel those of the physical field thus artificially rendering the theory finite. Several fictitious fields may be necessary to remove all of the divergences. To avoid the appearance of fictitious particles being created (and unitarity breaking down) and to remove the regularization, the masses of these field quanta are allowed to go to infinity at the end of the calculation. The infinities that would reappear in this limit are removed by renormalization, as in the other regularization schemes. In this case the potentially divergent terms are functions of the fictitious particle masses. By allowing some of the fictitious scalar fields to anticommute, rather than commute (or in the case of spinor fields, commute rather than anticommute), they can be arranged to contribute negatively to the stress-tensor, and so cancel the divergences of the opposite sign from the physical field.

This technique has been employed by Vilenkin (1978) to obtain the trace anomaly for the conformally coupled scalar field, and by Bernard & Duncan (1977) who give a detailed treatment of stress-tensor regularization. The idea of using a high-mass approximation to isolate and subtract the purely geometrical divergent terms in $\langle T_{\mu\nu} \rangle$ is essentially the same as both the adiabatic approach and the use of the DeWitt–Schwinger expansion in powers m^{-1}. Thus, Pauli–Villars regularization is basically the same as these other techniques in its approach to renormalization (Bunch, Christensen & Fulling 1978).

The DeWitt–Schwinger expansion (6.39) in the coincidence limit

$x = x'(\sigma = 0)$ yields

$$L_{\text{eff}} = \tfrac{1}{2}\mathrm{i}(4\pi)^{-n/2} \sum_{j=0}^{\infty} a_j(x) m^{n-2j} \int_0^{\infty} (\mathrm{i}\bar{s})^{j-1-n/2} \mathrm{e}^{-\mathrm{i}\bar{s}} \mathrm{d}\bar{s} \tag{6.241}$$

where we have changed variables: $\bar{s} = m^2 s$. The first $n/2$ terms are ultraviolet divergent. If we add to L_{eff} the contributions from the fictitious scalar regulator fields in an appropriate combination, these divergences can be formally cancelled. For example, in two dimensions, three additional fields are required: a commuting field with mass $(2M^2 - m^2)^{\frac{1}{2}}$ and two anticommuting fields, each with mass M. (At the end of the calculation we let $M \to \infty$.) The coefficient of the leading order (quadratic) divergent integral in (6.241) is then

$$a_0 m^2 + a_0(2M^2 - m^2) - 2a_0 M^2 = 0$$

while that of the next order (logarithmic) divergence is likewise

$$a_0 + a_0 - 2a_0 = 0.$$

In four dimensions, five extra scalar fields are necessary: for example, two anticommuting fields each with mass $(M^2 + m^2)^{\frac{1}{2}}$, one more with mass $(4M^2 + m^2)^{\frac{1}{2}}$ and two commuting fields with mass $(3M^2 + m^2)^{\frac{1}{2}}$. It is easy to check that this arrangement exactly cancels the a_0, a_1 and a_2 terms of (6.241).

Although the above procedure formally eliminates the ultraviolet divergent terms of (6.241), the change of variables $\bar{s} = m^2 s$ is not strictly legitimate for terms with $j = 0, 1$, which are divergent. Indeed, had we retained the s variable, the complete integral, including contributions from regulator fields, for $n = 2$, would be

$$\begin{aligned} &\int_0^{\infty} (\mathrm{i}s)^{j-2} [\mathrm{e}^{-\mathrm{i}m^2 s} + \mathrm{e}^{-\mathrm{i}(2M^2 - m^2)s} - 2\mathrm{e}^{-\mathrm{i}M^2 s}] \mathrm{d}s \\ &= -4 \int_0^{\infty} (\mathrm{i}s)^{j-2} \mathrm{e}^{-\mathrm{i}M^2 s} \sin^2 [\tfrac{1}{2}(M^2 - m^2)s] \mathrm{d}s. \end{aligned} \tag{6.242}$$

The integral is easily evaluated. For $j > 1$ it vanishes when $M \to \infty$. For $j = 0$, 1 it diverges in the limit $M \to \infty$. These divergences were missed by the change of variables to $\bar{s}$. They are the divergences which we expect to appear when the regularization is removed and can be eliminated by renormalization of the generalized Einstein action (or the gravitational field equations) in the usual way. At the end of this section we shall give a practical computation where these M-dependent divergences will be explicitly displayed.

First, however, it is instructive to see in detail how the trace anomaly arises in the Pauli–Villars method. The trace of the classical expression for $T_{\mu\nu}$ in the case of conformal coupling ($\xi = \xi(n)$) is $m^2\phi^2$, which clearly vanishes for $m = 0$. However $\langle T_\mu{}^\mu \rangle$ does not vanish because one must include in this quantity contributions from the regulator fields for which $M \to \infty$ rather than zero. Thus, in two dimensions

$$\langle 0|T_\mu^\mu|0\rangle = M^2 G^{(1)}_{2M^2}(x,x) - M^2 G^{(1)}_{M^2}(x,x) \qquad (6.243)$$

where we have put $m = 0$ and used $G^{(1)}(x,x) = 2\langle 0|\phi^2(x)|0\rangle$. Using the DeWitt–Schwinger expansion for $G^{(1)}$

$$\langle 0|T_\mu^\mu|0\rangle = (M^2/2\pi)\sum_{j=0}^{\infty} a_j(x)\int_0^\infty (\mathrm{i}s)^{j-1}[\mathrm{e}^{-2\mathrm{i}M^2 s} - \mathrm{e}^{-\mathrm{i}M^2 s}]\mathrm{i}\,\mathrm{d}s. \qquad (6.244)$$

Treating the integral in a similar way to (6.242), one sees that only the a_0 term is divergent as $M \to \infty$. This will be removed in the renormalization of Λ in (6.220). Evaluating the other integrals, one finds that all terms vanish as $M \to \infty$ except that involving a_1, which yields

$$\langle 0|T_\mu^\mu|0\rangle_{\mathrm{ren}} \xrightarrow[M\to\infty]{} -a_1(x)/4\pi, \qquad (6.245)$$

in agreement with (6.121).

In four dimensions, one similarly finds that the only surviving term after renormalization and relaxation of the regularization ($M \to \infty$) comes from the term proportional to m^{-2} in $G^{(1)}$, leaving $-a_2/16\pi^2$, the familiar result. In odd spacetime dimensions, there is no term proportional to m^{-2} in $G^{(1)}_{\mathrm{DS}}$, so the trace anomaly vanishes.

To illustrate the use of the Pauli–Villars method in a practical calculation, we shall follow the example given by Bernard & Duncan (1977). This example has already been discussed in §3.4 in connection with particle creation.

Consider a massive, conformally coupled scalar field in two-dimensional Robertson–Walker spacetime with metric given by (3.83) and (3.84). Near the region $\eta \to -\infty$, the deviation from flat spacetime may be considered small. If we choose $A = 1 + b/2, B = b/2$, then the conformal factor

$$\begin{aligned} C(\eta) &= 1 + \tfrac{1}{2}b(1 + \tanh \rho\eta) \\ &\simeq 1 + b\mathrm{e}^{2\rho\eta} \end{aligned} \qquad (6.246)$$

for $\rho\eta \to -\infty$. The quantity ρ^{-1} may be envisaged as equivalent to the adiabatic parameter T of §3.5.

To first order in $be^{2\rho\eta}$, the modes (3.87) reduce to

$$u_k^{\text{in}} \approx (4\pi\omega_{\text{in}})^{-\frac{1}{2}} e^{ikx - i\omega_{\text{in}}\eta}\left[1 - \frac{m^2 be^{2\rho\eta}}{4\rho^2(1 - i\omega_{\text{in}}/\rho)}\right] \tag{6.247}$$

where now $\omega_{\text{in}}^2 = k^2 + m^2$.

From (3.190), putting $n = 2$, $\xi = 0$

$$T_{\mu\nu} = \phi_{;\mu}\phi_{;\nu} - \tfrac{1}{2}g_{\mu\nu}g^{\sigma\rho}\phi_{;\sigma}\phi_{;\rho} + \tfrac{1}{2}m^2 g_{\mu\nu}\phi^2. \tag{6.248}$$

Then using (2.43) (which remains valid in curved spacetime) with modes (6.247), one readily obtains, to first order,

$$\langle \text{in}, 0|T_{\eta\eta}|0, \text{in}\rangle = \frac{1}{4\pi}\int_{-\infty}^{\infty}(k^2 + m^2)^{\frac{1}{2}}\mathrm{d}k + \frac{m^2 be^{2\rho\eta}}{8\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{(k^2 + m^2)^{\frac{1}{2}}} \tag{6.249}$$

$$\begin{aligned}\langle \text{in}, 0|T_{xx}|0, \text{in}\rangle &= \frac{1}{4\pi}\int_{-\infty}^{\infty}(k^2 + m^2)^{\frac{1}{2}}\mathrm{d}k - \frac{m^2}{4\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{(k^2 + m^2)^{\frac{1}{2}}}\\ &\quad - \frac{m^2 be^{2\rho\eta}}{8\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{(k^2 + m^2)^{\frac{1}{2}}} + \frac{m^4 be^{2\rho\eta}}{8\pi}\int_{-\infty}^{\infty}\frac{\mathrm{d}k}{(\rho^2 + k^2 + m^2)(k^2 + m^2)^{\frac{1}{2}}}.\end{aligned} \tag{6.250}$$

These expressions contain the usual quadratic and logarithmic divergences that are to be cancelled by the regulator fields. When these are added to the above expressions one has

$$\begin{aligned}&\int_{-\infty}^{\infty}[(k^2 + m^2)^{\frac{1}{2}} + (k^2 + 2M^2 - m^2)^{\frac{1}{2}} - 2(k^2 + M^2)^{\frac{1}{2}}]\mathrm{d}k\\ &\quad = \tfrac{1}{2}m^2\ln[(2M^2 - m^2)/m^2] + M^2\ln[M^2/(2M^2 - m^2)]\end{aligned} \tag{6.251}$$

and

$$\begin{aligned}&\int_{-\infty}^{\infty}[m^2(k^2 + m^2)^{-\frac{1}{2}} + (2M^2 - m^2)(k^2 + 2M^2 - m^2)^{-\frac{1}{2}}\\ &\qquad - 2M^2(k^2 + M^2)^{-\frac{1}{2}}]\mathrm{d}k\\ &\quad = m^2\ln[(2M^2 - m^2)/m^2] + 2M^2\ln[M^2/(2M^2 - m^2)],\end{aligned} \tag{6.252}$$

for the previously quadratic and logarithmically divergent terms, respectively. Substituting (6.251) and (6.252) into (6.249) and (6.250) (plus their regulator counterparts) yields, to the order considered here

$$\begin{aligned}\langle \text{in}, 0|T_{\eta\eta}|0, \text{in}\rangle &= (C/8\pi)\{m^2\ln[(2M^2 - m^2)/m^2]\\ &\quad + 2M^2\ln[M^2/(2M^2 - m^2)]\}\end{aligned} \tag{6.253}$$

$$\langle \text{in}, 0|T_{xx}|0, \text{in}\rangle = -(C/8\pi)\{m^2 \ln[(2M^2 - m^2)/m^2] + 2M^2 \ln[M^2/(2M^2 - m^2)]\} + f(m^2) + f(2M^2 - m^2) - 2f(M^2) \tag{6.254}$$

where

$$f(m^2) = \frac{m^4 b e^{2\rho\eta}}{8\pi} \int_{-\infty}^{\infty} \frac{\mathrm{d}k}{(k^2 + \rho^2 + m^2)(k^2 + m^2)^{\frac{1}{2}}} = \frac{m^4 b e^{2\rho\eta}}{8\pi\rho(m^2 + \rho^2)^{\frac{1}{2}}} \ln\left[\frac{(m^2 + \rho^2)^{\frac{1}{2}} + \rho}{(m^2 + \rho^2)^{\frac{1}{2}} - \rho}\right]. \tag{6.255}$$

Noting also that $\langle T_{\eta x}\rangle = \langle T_{x\eta}\rangle = 0$, these expressions can be written

$$\langle \text{in}, 0|T_{\mu\nu}|0, \text{in}\rangle = (g_{\mu\nu}/8\pi)\{m^2 \ln[(2M^2 - m^2)/m^2] + 2M^2 \ln[M^2/(2M^2 - m^2)]\} + \eta_{\mu 1}\eta_{\nu 1}[f(m^2) + f(2M^2 - m^2) - 2f(M^2)]. \tag{6.256}$$

The first term on the right-hand side of (6.256) can be absorbed in renormalization of the cosmological constant Λ in the Einstein equation (it diverges as $M \to \infty$). There is no renormalization associated with a_1 in the effective Lagrangian, because this gives rise to a multiple of the Einstein tensor in the Einstein equation and this vanishes identically in two dimensions. The remaining term in (6.256) is finite as $M \to \infty$, and, taking the limit, is easily evaluated. One has

$$\langle \text{in}, 0 | T_{\mu\nu}|0, \text{in}\rangle_{\text{ren}} = -\eta_{\mu 1}\eta_{\nu 1} \frac{b e^{2\rho\eta}}{4\pi}\left[m^2 - \tfrac{2}{3}\rho^2 - \frac{m^4}{2\rho(m^2 + \rho^2)^{\frac{1}{2}}} \ln\left(\frac{(m^2 + \rho^2)^{\frac{1}{2}} + \rho}{(m^2 + \rho^2)^{\frac{1}{2}} - \rho}\right)\right]. \tag{6.257}$$

Note that, for $m = 0$, the trace of (6.257) gives, to lowest order in $b e^{2\rho\eta}$, the correct conformal anomaly:

$$\langle \text{in}, 0|T_\mu{}^\mu|0, \text{in}\rangle_{\text{ren}} = -b e^{2\rho\eta}\rho^2/6\pi \simeq -R/24\pi. \tag{6.258}$$

One disadvantage of the use of Pauli–Villars regularization is that one must always be able to work with massive fields (even if in some approximation). In many calculations, the massless field equation is often considerably simpler than its massive counterpart.

6.6 Physical significance of the stress-tensor

The enormous amount of effort that has been invested in developing techniques for computing $\langle T_{\mu\nu}\rangle_{\text{ren}}$ calls for a searching appraisal of the physical basis that underlies this quantity. In the previous sections it has been shown how, by a variety of complicated mathematical devices, one may extract from a meaningless formal quantity, $\langle T_{\mu\nu}\rangle$, a residue that, at least in the cases explicitly investigated, must be considered a serious contender for the quantity to be placed on the right-hand side of the gravitational field equation.

Not withstanding the physical reasonableness of the answers obtained, the construction of $\langle T_{\mu\nu}\rangle_{\text{ren}}$ by infinite renormalization of gravitational coupling constants is open to criticism. The DeWitt–Schwinger proper time integral, the inverse mass expansion of G_{F}, and many of the more formal manipulations of §§6.1 and 6.2 (such as those involving tr ln G_{F}) are at best ill-defined for hyperbolic operators, and may not even exist in some cases. Few rigorous results have been proved.

In addition to these mathematical problems, there is the question of whether the semiclassical theory makes sense at all. How would one measure $\langle T_{\mu\nu}\rangle$? When is the semiclassical approximation valid? If $\langle T_{\mu\nu}\rangle$ can only be measured gravitationally (i.e., through its entry into the field equation) can one consistently neglect higher order (i.e., two-loop graviton, etc.) contributions? What is the relation between higher order corrections and the heuristic expectation that $\langle T^2_{\mu\nu}\rangle$ should be $\ll \langle T_{\mu\nu}\rangle$ if the latter quantity is to accurately approximate some average distribution of quantum stress–energy–momentum? When, if ever, will the 'back-reaction' (i.e. gravitational dynamics modified by gravitationally induced $\langle T_{\mu\nu}\rangle$) be approximately determined by $\langle T_{\mu\nu}\rangle_{\text{ren}}$ computed at the one-loop level? Misgivings about these issues have been expressed by a number of authors (for example Duff 1981). Many of them might be resolved if a full theory of quantum gravity were available, to which one could claim that the semiclassical theory is some sort of approximation.

One approach to the physical significance of $\langle T_{\mu\nu}\rangle$ is to abandon renormalization altogether, and to ask, simply, that if the semiclassical theory is to make physical sense, what criteria might one wish $\langle T_{\mu\nu}\rangle$ to satisfy? If these criteria are too restrictive, no such object may exist; too loose, and it may not be unique.

The approach of attempting to define a unique, existing $\langle T_{\mu\nu}\rangle$ purely by imposing physical criteria ('axioms') was instigated by Christensen (1975) and has been pursued with great success by Wald (1977, 1978*a*, *b*), whose work has considerably strengthened the results of the renormalization programme. In the weaker system of axioms, Wald proposes that any physically meaningful $\langle T_{\mu\nu}\rangle$ ought at least to satisfy four eminently reasonable conditions:

(1) covariant conservation
(2) causality
(3) standard results for 'off-diagonal' elements
(4) standard results in Minkowski space.

The first condition is simply relation (6.142) which is necessary if $\langle T_{\mu\nu} \rangle$ is to appear on the right-hand side of the gravitational field equations, as the left-hand side is divergenceless. The causality axiom is rather subtle. The precise statement is:

'For a fixed 'in' state, $\langle T_{\mu\nu} \rangle$ at a point p in spacetime depends only on the spacetime geometry to the causal past of p'. By this it is meant that changes in the metric structure of the spacetime outside the past null cone through p ought not to effect $\langle T_{\mu\nu} \rangle$, so long as the state of the quantum field in the remote past is unaltered. (One can conceive of alterations outside the past null cone that nevertheless modify the Fock space based on the 'in' modes. These are not permitted.) A time-reversed statement then applies to fixed 'out' states and changes in the geometry outside the future null cone.

Condition (3) is simply the observation that as $\langle \Phi | T_{\mu\nu} | \Psi \rangle$ is in any case finite for orthogonal states, $\langle \Phi | \Psi \rangle = 0$, the value of this quantity ought to be the usual (i.e., formal) one. By condition (4), we mean that the normal ordering procedure in Minkowski space (see §2.4) should be valid.

It is now straightforward to prove a remarkable result: if $\langle T_{\mu\nu} \rangle$ satisfies the first three of the four conditions above, then it is unique to within a local conserved tensor.

The proof runs as follows (see Wald, 1977, for a detailed exposition). If $T_{\mu\nu}$ and $\tilde{T}_{\mu\nu}$ are two (renormalized) stress-tensor operators satisfying conditions (1)–(3), then our aim is to show that the expectation value of

$$U_{\mu\nu} \equiv T_{\mu\nu} - \tilde{T}_{\mu\nu} \tag{6.259}$$

is a local, conserved tensor.

First note that by condition (3), the matrix elements of $U_{\mu\nu}$ between orthogonal states must vanish (because $\langle \Phi | T_{\mu\nu} | \Psi \rangle = \langle \Phi | \tilde{T}_{\mu\nu} | \Psi \rangle$). Furthermore, putting $|\Pi_\pm\rangle = 2^{-\frac{1}{2}}(|\Psi\rangle \pm |\Phi\rangle)$, then

$$\langle \Pi_+ | U_{\mu\nu} | \Pi_- \rangle = 0$$

so

$$\langle \Psi | U_{\mu\nu} | \Psi \rangle - \langle \Phi | U_{\mu\nu} | \Phi \rangle = 0, \quad \forall \Psi, \Phi, \tag{6.260}$$

i.e., all the diagonal elements (expectation values) are equal. Thus $U_{\mu\nu}$ is a multiple of the identity operator:

$$U_{\mu\nu} = u_{\mu\nu} I, \tag{6.261}$$

where $u_{\mu\nu}$ is an ordinary c-number tensor field.

We can now see that $u_{\mu\nu}$ must be a local tensor, for if we take the expectation value of $U_{\mu\nu}$ in some normalized 'in' state, we have

$$\langle \mathrm{in} | U_{\mu\nu}(p) | \mathrm{in} \rangle = u_{\mu\nu}(p). \tag{6.262}$$

But this would also be our answer if we took a normalized 'out' state

$$\langle \mathrm{out} | U_{\mu\nu}(p) | \mathrm{out} \rangle = u_{\mu\nu}(p). \tag{6.263}$$

As condition (2) requires that $u_{\mu\nu}(p)$ in (6.262) can only depend on the geometry in the causal past of p, while $u_{\mu\nu}(p)$ in (6.263) is similarly restricted by the geometry in the causal future, these two objects can only be equal if they depend solely on the geometry in the intersection of the past and future null cones, i.e., at p. Hence $u_{\mu\nu}(p)$ is a local tensor at p.

Finally, by condition (1), $u_{\mu\nu}$ must be conserved:

$$u^{\mu\nu}{}_{;\nu} = 0. \tag{6.264}$$

Thus, $\langle T_{\mu\nu} \rangle$ is unique to within a local, conserved tensor. However, any conserved tensor that is a function solely of the local geometry more properly belongs to the left-hand side of the gravitational field equations anyway, i.e., it is more reasonable to regard it as part of the gravitational dynamics than of the quantum field.

Even in renormalization theory, we are always free to take a local conserved tensor (e.g., ${}^{(1)}H_{\mu\nu}$) and place it on the right-hand side of the field equations if we wish. A physical measurement can resolve this ambiguity by determining the coefficient of such a term. Alternatively, one might attempt to advance arguments why no such term ought to be present on physical grounds (i.e., require that its coefficient be zero). For example, Wald (1977) suggested a fifth condition based on stability criteria for the gravitational dynamics, which, together with condition (4), uniquely fixes $u_{\mu\nu} = 0$. This fifth condition is now considered to be of dubious value as we shall discuss shortly.

One would like to know whether or not the renormalization prescriptions described in the earlier sections of this chapter satisfy Wald's conditions. If they do, then we have good reason to believe the results, whether the removal of the divergences is regarded as a legitimate renormalization of gravitational coupling constants, or merely as an *ad hoc* ansatz. This is especially important in the case of point-splitting, because one is unable to renormalize $\langle T_{\mu\nu} \rangle$ in the gravitational field equations anyway, it being necessary (guided by renormalization in the effective action) to discard a term in order to achieve the full conformal anomaly (see discussion on page 205). Moreover, the procedure of averaging over directions of the splitting vector t_μ prior to renormalization is open to the objection that no unique measure on the space of directions exists; one merely chooses a 'natural' one.

The point-splitting prescription starts with the DeWitt–Schwinger expansion of $G^{(1)}(x, x')$. If this expansion exists, then, in four dimensions, one can see by inspection of (3.141) that it will have the general form

$$S(x, x') = (U/\sigma) + V \ln \sigma + W \tag{6.265}$$

where $U = U(x, x')$ and

$$V(x, x') = \sum_{l=0}^{\infty} V_l(x, x')\sigma^l, \quad W(x, x') = \sum_{l=0}^{\infty} W_l(x, x')\sigma^l. \tag{6.266}$$

Expressed in this form, a Green function is called a Hadamard elementary solution, after the extensive work by Hadamard (1923) on the singularity structure of second order elliptic and hyperbolic equations. The question of when $G^{(1)}(x, x') = \langle 0|\{\phi(x), \phi(x')\}|0\rangle$ has the form of a Hadamard elementary solution (i.e., in what spacetimes and for what states $|0\rangle$) is still open, but it is generally believed to possess this form for a wide class of spacetimes and boundary conditions in which we are interested. In particular, Fulling, Sweeny & Wald (1978) have proved that if $G^{(1)}$ has the Hadamard singularity structure in an open neighbourhood of a Cauchy surface, then it has this form everywhere. This has as a corollary the result that, for a spacetime which is flat in the past of some Cauchy surface, $G^{(1)}$ calculated in the in vacuum will have the Hadamard form everywhere because it manifestly has that form in ordinary flat space quantum field theory.

The coefficients in (6.266) may be found by substituting (6.265) into the field equation and solving recursively (DeWitt & Brehme 1960, Garabedian 1964, Adler, Lieberman & Ng 1977). This procedure uniquely determines V. Moreover, W is uniquely determined once $W_0(x, x')$ is specified. Its specification may be regarded as the imposition of a boundary condition on the field, and it uniquely characterizes $S(x, x')$. An overall normalization then fixes U.

It is convenient to change the form of the Hadamard solution slightly and write it as

$$S(x, x') = [2/(4\pi)^2]\Delta^{\frac{1}{2}}[-(2/\sigma) + v \ln \sigma + w], \tag{6.267}$$

where $\Delta(x, x')$ is defined by (3.139) and series analogous to (6.266) exist for v and w. When (6.267) is expanded in inverse powers of mass, it reproduces the DeWitt–Schwinger Green function expansion $G^{(1)}_{\text{DS}}$, given in terms of the expansion (3.141), which was used as the basis of our renormalization programme. In four dimensions, this involved subtracting from $G^{(1)}(x, x')$ the first three terms (terms up to adiabatic order four) of the inverse mass expansion of $G^{(1)}_{\mathrm{DS}}$. However, to investigate consistency with the Wald axioms, it is preferable to work, not with a truncated inverse mass expansion, but with the exact Hadamard form (6.267). This is especially true in the massless limit, where the w terms of the DeWitt–Schwinger expansion diverge (being an expansion in m^{-1}). Shortly we shall discover the relation between renormalization based on $S(x, x')$ as compared with $G^{(1)}_{\mathrm{DS}}(x, x')$.

The Green function $G^{(1)}(x, x')$ is the basic object from which $\langle 0 | T_{\mu\nu} | 0 \rangle$ is constructed by differentiation. We assume that $G^{(1)}$ has the Hadamard form in a practical calculation. Now $G^{(1)}(x, x')$ is manifestly symmetric in x, x' and hence satisfies the field equation in both of these variables. In the case of the Hadamard solution (6.267), one can prove (Fulling, Sweeny & Wald 1978) that Δ and v are symmetric for all Hadamard solutions. However, the same conclusion cannot be drawn for w, as this non-singular part is only specified uniquely once $w_0(x, x')$ is given, and can vary from solution to solution, depending on boundary conditions. Thus, although $w(x, x')$ is symmetric for $G^{(1)}(x, x')$ by construction, it will not be so in general. We may therefore conclude that $S(x, x')$ will not, in general, satisfy the field equation in x', even though it will, by definition, satisfy it in x.

We wish to discover a way of removing the pole terms from $G^{(1)}(x, x')$ in such a fashion as to respect the Wald axioms and hence achieve an (almost) unique answer when the residue is differentiated to give $\langle T_{\mu\nu}(x) \rangle$. One likely procedure is to subtract from $G^{(1)}(x, x')$ a Hadamard solution $S(x, x')$ (Wald 1977, 1978*a*, Adler, Lieberman & Ng 1977). To make the procedure unambiguous, however, one must decide on the boundary conditions for S, i.e., specify $w_0(x, x')$. Since for massless fields in Minkowski space $D^{(1)}$ is given by (2.79), i.e., $w = 0$ ($w =$ constant in the massive case), for consistency with condition (4) we assume that $w_0(x, x') = 0$ in S. From our previous result we know that any other w_0 which is consistent with all four conditions, can yield a result which differs at most by a local conserved tensor from that obtained using $w_0 = 0$.

With this choice, $w_0 = 0$, we have

$$G^{\mathrm{B}} = G^{(1)} - S \tag{6.268}$$

where G^{B} (B standing for 'boundary-condition-dependent part') is a smooth, uniquely determined function of x and x'. From it, one may construct $\langle T^{\mathrm{B}}_{\mu\nu} \rangle$ by differentiation (see §6.4 for details of this step).

The result clearly satisfies Wald's condition (2), because the two pieces,

$G^{(1)}$ and S, from which it is constructed, are both causal, $G^{(1)}$ because the ϕ field from which it is computed propagates causally, and S because it is a purely local object, so does not depend on the geometry outside the null cones. When $x' \to x$, G^{B} depends only on the geometry on and inside the null cone through x.

Condition (3) can also easily be shown to be satisfied. From (6.260) we can express the matrix elements of $T^{\mathrm{B}}_{\mu\nu}$ between orthogonal states in the form

$$\begin{aligned}\langle \Pi_+ | T^{\mathrm{B}}_{\mu\nu} | \Pi_- \rangle &= \tfrac{1}{2}\langle \Psi | T^{\mathrm{B}}_{\mu\nu} | \Psi \rangle - \tfrac{1}{2}\langle \Phi | T^{\mathrm{B}}_{\mu\nu} | \Phi \rangle \\ &= \tfrac{1}{2} \lim_{x' \to x} \mathscr{D}_{\mu\nu}(G^{\mathrm{B}}_{\Psi} - G^{\mathrm{B}}_{\Phi}) \\ &= \tfrac{1}{2} \lim_{x' \to x} \mathscr{D}_{\mu\nu}(G_{\Psi} - G_{\Phi}) \end{aligned} \tag{6.269}$$

where $\mathscr{D}_{\mu\nu}$ is the operator discussed on page 195 and G_Ψ, etc., denote $\langle \Psi | \{\phi(x), \phi(x')\} | \Psi \rangle$, etc. The final equality in (6.269) follows from the fact that the same $S(x, x')$ is subtracted from both G_Ψ and G_Φ, by the 'renormalization' ansatz (6.268). According to condition (3), (6.269) is supposed to be the usual (formal, 'unrenormalized') expression, which it clearly is, since it is independent of S.

Unfortunately, the ansatz (6.268) does not satisfy Wald's condition (1). As remarked above, $S(x, x')$ will not in general satisfy the field equation in both x and x'. For the boundary condition $w_0 = 0$, it does not, in fact, satisfy this equation in x'. Calculation shows that, in the massless case

$$\nabla^\nu \langle T^{\mathrm{B}}_{\mu\nu} \rangle = \tfrac{1}{4} \lim_{x' \to x} \nabla_\mu [\Box_{x'} + \tfrac{1}{6} R(x')] G^{\mathrm{B}}(x, x'). \tag{6.270}$$

so that $\langle T^{\mathrm{B}}_{\mu\nu} \rangle$ fails to be conserved to the extent that it fails to satisfy the field equation in x'. Wald (1978*a*) evaluates the right-hand side of (6.270) and obtains

$$\nabla_\mu a_2(x)/64\pi^2 \tag{6.271}$$

where $a_2(x)$ is the usual coefficient in the DeWitt–Schwinger expansion. Thus, to construct a conserved $\langle T_{\mu\nu} \rangle$ that is still consistent with the conditions (2)–(4) one must take

$$\langle T_{\mu\nu}(x) \rangle = \langle T^{\mathrm{B}}_{\mu\nu}(x) \rangle - a_2(x) g_{\mu\nu}(x)/64\pi^2. \tag{6.272}$$

Whereas $g^{\mu\nu} \langle T^{\mathrm{B}}_{\mu\nu} \rangle = 0$ by construction,

$$g^{\mu\nu} \langle T_{\mu\nu} \rangle = -a_2(x)/16\pi^2 \tag{6.273}$$

in agreement with the conformal anomaly (6.119). As our result is unique up to a local conserved tensor, and as there exists no such tensor with the trace (6.273), we may conclude that the conformal anomaly is an inevitable consequence of a local, semiclassical theory. Of course one does have the freedom to adjust the $\Box R$ part of the anomaly by adding multiples of the local conserved tensors ${}^{(1)}H_{\mu\nu}$ and ${}^{(2)}H_{\mu\nu}$.

Wald originally suggested as a fifth condition, based on stability arguments, that $\langle T_{\mu\nu}\rangle$ should contain no terms of adiabatic order greater than three (such as terms giving rise to $\Box R$ in the trace). However, one can show that the stress-tensor will in general contain non-local terms involving higher adiabatic order contributions which cannot be removed by addition of (local) counterterms proportional to ${}^{(1)}H_{\mu\nu}$ and ${}^{(2)}H_{\mu\nu}$. Horowitz (1980) has suggested that, for massless fields, the presence of such non-local, higher derivative terms will lead to instability about flat spacetime. The full implication of such results for the semiclassical theory is still under investigation.

What is the relationship between $S(x, x')$ as used here to render $G^{(1)}$ finite, and the DeWitt–Schwinger $G^{(1)}_{\mathrm{DS}}(x, x')$ used in the previous sections to renormalize $G^{(1)}$? First note that (in four dimensions) we are only interested in terms in the expansion of $G^{(1)}_{\mathrm{DS}}$ up to adiabatic order four, and up to order σ. Terms of higher order in σ do not contribute to $\langle T_{\mu\nu}\rangle$ as $\sigma \to 0$. From (3.141) these terms are

$$\begin{aligned} G^{(1)}(x, x') = (\Delta^{\frac{1}{2}}/4\pi^2)\{&a_0[-(1/\sigma) + m^2 L(1 - \tfrac{1}{4}m^2\sigma) - \tfrac{1}{2}m^2 + \tfrac{5}{16}m^2\sigma] \\ &- a_1[L(1 - \tfrac{1}{2}m^2\sigma) + \tfrac{1}{2}m^2\sigma] \\ &- a_2\sigma[\tfrac{1}{2}L - \tfrac{1}{4}] + (1/2m^2)a_2 + O(T^{-6}) + O(\sigma^2)\}, \end{aligned} \qquad (6.274)$$

where $L \equiv \gamma + \frac{1}{2}\ln|\frac{1}{2}m^2\sigma|$, γ being Euler's constant.

Clearly the divergent and logarithmic terms in (6.274) have the Hadamard form (6.267), and, by the uniqueness property, must agree with the corresponding terms in $S(x, x')$. Moreover, because $G^{(1)}_{\mathrm{DS}}$ satisfies the field equation at each adiabatic order and at each power of σ, the determination of the coefficients of the terms $a_0m^2, a_0m^4\sigma, a_1m^2\sigma$ and $a_2\sigma$ are independent of the m^{-2} terms and higher. Hence they will be the same as in the Hadamard $S(x, x')$. This is confirmed by explicit calculation (Adler, Lieberman & Ng 1977). Thus the difference between $G^{(1)}_{\mathrm{DS}}$ and S is the 'anomaly term', i.e., the ultimate term in (6.274), which as $x' \to x$ is $a_2/(8\pi^2 m^2)$. In a practical calculation, one renormalizes $\langle T_{\mu\nu}\rangle$ by subtracting (6.274) from $G^{(1)}$, and after differentiation dropping all terms resulting from (6.274) of adiabatic order greater than four. As discussed on page 205 this procedure is equivalent to subtracting the final term in (6.274) only in its contribution made via the term $2\{\frac{1}{4}-[1-(1/n)]\xi\}m^2g_{\mu\nu}\phi^2$ in (3.190). Thus, in the four-dimensional, conformally-coupled, massless limit, the final term in (6.274) makes the single contribution $-a_2g_{\mu\nu}/(64\pi^2)$ to $\langle T_{\mu\nu}\rangle$, and so this is the only difference resulting in $\langle T_{\mu\nu}\rangle$ from the subtraction of $G^{(1)}_{\mathrm{DS}}$ rather than S. But in the method in which S is subtracted, precisely this difference term is added on in (6.272) to maintain covariant conservation. (The necessity of including this term can be seen from the fact that $G^{(1)}$ satisfies the field equation in x, order by order in the adiabatic parameter. The conservation of $\langle T_{\mu\nu}\rangle$ depends on this property. Thus, when computing a conserved $\langle T_{\mu\nu}\rangle$, one must work to a consistent adiabatic order. The term $m^2\phi^2$, because it contains no derivatives, must involve two higher adiabatic orders in the expansion (6.274) than do the other terms (containing two derivatives).) We may therefore conclude that the point-splitting renormalization prescription given in §§6.2 and 6.4 is equivalent to the Wald subtraction method based on the Hadamard solution, which we know to yield an (almost) unique answer (see also Bunch, Christensen & Fulling 1978 for the two-dimensional case). Thus, the point-splitting renormalization method and the other renormalization techniques that we have shown are closely related to it, can be used with the confidence that the results they produce are consistent with the physically very reasonable conditions (1)–(4), and are (to within local conserved tensors) uniquely determined. This implies that the resulting $\langle T_{\mu\nu}\rangle_{\mathrm{ren}}$ must be the physically correct answer if the semiclassical theory is to make sense at all.

Finally, we turn to another aspect of the physical interpretation of $\langle T_{\mu\nu}\rangle$ concerning boundaries. It is usually supposed spacetime is unbounded. However, material surfaces can act as effective boundaries, e.g., the presence of conductors constrains the electromagnetic field to vanish, at least in some approximation (ignoring a 'skin depth'). The general mathemetical analysis (Minakshisundaram & Pleijel 1949, McKean & Singer 1967, Greiner 1971, Stewartson & Waechter 1971, Waechter 1972, Gilkey 1975, Kennedy 1978) of elliptic operators in Riemannian space and, by association, of the wave operator in pseudo-Riemannian spacetime, incorporates the effects of a boundary ∂M to a manifold M in the DeWitt–Schwinger series or its elliptic equivalent. One finds that when a boundary is present, the effective action given by (6.36) and (6.40) is augmented by a surface effective action (see, for example, Kennedy, Critchley & Dowker 1980)

$$W_{\mathrm{S}}=\int_{\partial M}\mathrm{d}^{n-1}x(\pm h)^{\frac{1}{2}}L_{\mathrm{S}}(x), \tag{6.275}$$

where h is the determinant of the metric h_{ij} induced on the boundary by the spacetime metric $g_{\mu\nu}$ ($\pm$ according to whether ∂M is timelike or spacelike respectively) and the asymptotic expansion of L_S (comparable to (6.40)) is

$$L_S \approx \tfrac{1}{2}(4\pi)^{-n/2} \sum_{j=0}^{\infty} b_{(j+1)/2}(x) \int_0^{\infty} (is)^{(2j-1-n)/2} e^{-im^2 s} i\,ds. \qquad (6.276)$$

The precise form of the coefficients b depends on the boundary conditions imposed on the field on ∂M. For Dirichlet boundary conditions imposed on a scalar field ϕ,

$$\phi(x) = 0, \quad x \in \partial M, \qquad (6.277)$$

one obtains for the first few coefficients

$$b_{\frac{1}{2}} = -\tfrac{1}{2}\pi^{\frac{1}{2}}$$

$$b_1 = \tfrac{1}{3}\chi$$

$$b_{\frac{3}{2}} = \tfrac{1}{192}\pi^{\frac{1}{2}}[3(3-32\xi)\chi^2 + 6(16\xi - 1)\chi_{\mu\nu}\chi^{\mu\nu} - 16(1-6\xi)\hat{R} - 24(8\xi - 1)R_{\mu\nu}n^\mu n^\nu], \qquad (6.278)$$

where n_μ is the inward pointing unit vector on ∂M, $\chi_{\mu\nu}$ and $\chi = \chi_\mu{}^\mu$ are respectively the second fundamental form of ∂M and its trace, and $\hat{R}$ is the Ricci scalar of the induced metric $h_{\mu\nu}$ (see, for example, Hawking & Ellis 1973, §2.7). For Robin boundary conditions

$$[\psi(x) + n^\mu \nabla_\mu]\phi(x) = 0, \qquad (6.279)$$

one obtains

$$b_{\frac{1}{2}} = \tfrac{1}{2}\pi^{\frac{1}{2}}$$

$$b_1 = \tfrac{1}{3}(\chi - 6\psi)$$

$$b_{\frac{3}{2}} = \tfrac{1}{192}\pi^{\frac{1}{2}}[192\psi^2 + 96\psi\chi + 3(32\xi - 1)\chi^2 + 6(3 - 16\xi)\chi_{\mu\nu}\chi^{\mu\nu} + 16(1-6\xi)\hat{R} - 24(1 - 8\xi)R_{\mu\nu}n^\mu n^\nu]. \qquad (6.280)$$

It is clear that the terms in (6.276) with $j \leq (n-1)/2$ are divergent. In the foregoing sections, divergent terms have been dealt with by first introducing a regularization scheme, and then removing the divergent terms by renormalization. The surface effective action (6.275) can be regularized using the methods introduced previously, and one must consider whether one can remove the divergent terms by renormalization of constants in the gravitational action.

The generalized gravitational action, which was renormalized in §6.2, makes no provision for surface divergences, so it appears at first sight that the presence of a boundary ∂M, on which boundary conditions on the field are imposed, gives rise to truly infinite vacuum stress there. It has been argued by Deutsch & Candelas (1979) that this surface divergence represents a real physical effect of the quantum field. A boundary to spacetime is in any case such a pathological feature that the occurrence of a divergence in $\langle T_{\mu\nu}\rangle$ there is not especially surprising. In the case of material surfaces, one must take into account that treating them as manifold boundaries is an idealization. For example, in the electromagnetic case, the finite conductivity of a real conductor would render the material transparent at very high frequencies, thus providing an ultraviolet cut-off to the mode integrals. (The cut-off parameter used in §4.1 as a mathematical device has, in this case, some foundation in physics.) Consequently, the apparent divergences at a material surface are really only very large, but finite, contributions to $\langle T_{\mu\nu}\rangle$. Deutsch & Candelas point out that any change in the conductivity of the material would lead to a large, and possibly measurable shift in $\langle T_{\mu\nu}\rangle$.

On the other hand, if one allows the addition to the generalized Einstein action of a surface action involving terms appearing in (6.278) and (6.280), then one can remove the divergences arising from the matter fields' surface effective action by renormalization. This approach has been advocated by Kennedy, Critchley & Dowker (1980).

The necessity of adding a surface action to the conventional Einstein action S_g given by (6.11) has been pointed out in another context by Gibbons & Hawking (1977*b*). They note that the action (6.11) is no longer appropriate if one requires that under all variations of the metric that vanish on ∂M, stationarity of S_g leads to the Einstein equation. The reason is that the variation of terms in S_g that arise from terms in R that are linear in the second derivatives of the metric can be converted by integration by parts into an integral over ∂M that involves the normal derivative of the metric at the surface. Unless the normal derivative is required to vanish at ∂M, one must augment (6.11) by a term that will cancel this normal derivative surface integral. Gibbons & Hawking find such a term to be

$$(-1/8\pi G)\int_{\partial M}\chi(\pm h)^{\frac{1}{2}}\,\mathrm{d}^{n-1}x + C, \tag{6.281}$$

where C depends only on h, not on g. If the boundary can be embedded in flat space, with second fundamental form $\chi^0_{\mu\nu}$, a natural choice of C is

$$(1/8\pi G)\int_{\partial M}\chi^0(\pm h)^{\frac{1}{2}}\mathrm{d}^{n-1}x, \tag{6.282}$$

so that the surface action vanishes when the spacetime is flat. The importance of the inclusion of (6.281) in considerations of quantum gravity has been stressed by Hawking (1979) (see also §8.5).

7
Applications of renormalization techniques

This short chapter presents some explicit examples of the theory of regularization and renormalization discussed in chapter 6. The number of spacetimes for which one may compute $\langle T_{\mu\nu} \rangle$ in terms of simple functions is extremely limited, and we think it probable that all such cases have been included either here, in chapter 6, or in our references.

Special importance is attached to the Robertson–Walker models, both because of their cosmological significance, and also because, being conformally flat, they provide a good illustration of conformal anomalies at work. However, precisely because of their simplicity, these models do not display the full non-local structure of the stress-tensor, and in §7.3 we turn briefly to the less elegant but more realistic example of an anisotropic, homogeneous cosmological model.

Although the primary subject of this book is the theory of quantum fields propagating in a prescribed background spacetime, the motivation for much of this work rests with its possible application to cosmological and astrophysical situations, where the gravitational dynamics must be taken into account. Many cosmologists, for example, believe that the back-reaction of quantum effects induced by the background gravitational field could have a profound effect on the dynamical evolution of the early universe, such as bringing about isotropization. We do not dwell in detail on this important extension of the theory, but note that the results presented here constitute the starting point for such investigations. A short discussion of the wider cosmological implications is given in §7.4.

7.1 Two-dimensional examples

The renormalized stress-tensor for the conformally trivial case of a massless scalar field with $\xi = 0$ propagating in two-dimensional spacetime has been derived in §§6.3 and 6.4.

We work with the line element

$$ds^2 = C(u, v)\,du\,dv \tag{7.1}$$

which has scalar curvature

$$R = \Box \ln C. \tag{7.2}$$

In flat spacetime, $R = 0$, so

$$\Box \ln C = 4C^{-1}\partial_u \partial_v \ln C = 0. \tag{7.3}$$

Thus

$$C = F(u)G(v) \tag{7.4}$$

where F and G are arbitrary differentiable functions.

One solution of (7.4) is $F = G = 1$, which recovers the familiar Minkowski space metric:

$$ds^2 = du\,dv. \tag{7.5}$$

Another can be obtained from (7.5) by a coordinate transformation $u \to \bar{u}$, $v \to \bar{v}$, such as

$$\left.\begin{aligned} u &= f(\bar{u}) \\ v &= \bar{v} \end{aligned}\right\}, \tag{7.6}$$

which changes (7.5) to

$$ds^2 = f'(\bar{u})\, d\bar{u}\, d\bar{v}. \tag{7.7}$$

Here a prime is used to denote differentiation of a function with respect to its argument.

If we choose the function f as follows

$$\tfrac{1}{2}[\bar{t} - f(\bar{t})] = z\{\tfrac{1}{2}[\bar{t} + f(\bar{t})]\}, \tag{7.8}$$

where

$$\bar{t} = \tfrac{1}{2}(\bar{u} + \bar{v}) \tag{7.9}$$

is the time coordinate in the $\bar{u}$, $\bar{v}$ system, then (7.8) is the restriction to $\bar{x} = 0$ of

$$\tfrac{1}{2}[\bar{v} - f(\bar{u})] = z\{\tfrac{1}{2}[\bar{v} + f(\bar{u})]\}, \tag{7.10}$$

where

$$\bar{x} = \tfrac{1}{2}(\bar{v} - \bar{u}) \tag{7.11}$$

is the space coordinate in the $\bar{u}$, $\bar{v}$ system.

In the (u, v) system, (7.10) reduces to

$$x = z(t), \tag{7.12}$$

which can be taken as the trajectory of the moving mirror investigated in §4.4. In the $(\bar{u}, \bar{v})$ coordinate system the mirror remains at $\bar{x} = 0$. Because of (7.7), the moving mirror system is conformally related to a static mirror system, so we may employ the general relation (6.136) which connects $\langle T_{\mu\nu} \rangle$ evaluated in conformally related vacuum states. In §4.3 we showed that, in four dimensions with a conformally coupled massless field, $\langle T_{\mu\nu} \rangle$ in the half space bounded by a static plate vanishes. The same result is readily recovered for the the two-dimensional case. Hence we may set to zero the first term of the right of (6.136). The relevant conformal factor C is found from (7.7) to be $f'(\bar{u})$, so recalling that $R = 0$, (6.137) yields (Davies & Fulling 1977*b*) the expectation value in the 'conformal vacuum' state:

$$\langle T_{\bar{u}\bar{u}} \rangle_{\text{ren}} = -(1/12\pi)(f')^{\frac{1}{2}} \partial_{\bar{u}}^2 (f')^{-\frac{1}{2}}$$

$$\langle T_{\bar{u}\bar{v}} \rangle_{\text{ren}} = \langle T_{\bar{v}\bar{u}} \rangle_{\text{ren}} = \langle T_{\bar{v}\bar{v}} \rangle_{\text{ren}} = 0.$$

Transforming to the (u, v) coordinate system, this gives

$$\langle T_{uu} \rangle_{\text{ren}} = -(1/12\pi)(f')^{-\frac{3}{2}} \partial_{\bar{u}}^2 (f')^{-\frac{1}{2}} \tag{7.13}$$

$$\langle T_{uv} \rangle_{\text{ren}} = \langle T_{vu} \rangle_{\text{ren}} = \langle T_{vv} \rangle_{\text{ren}} = 0. \tag{7.14}$$

In (7.13) the argument of f is

$$\bar{u} = f^{-1}(u) = 2\tau_u - u \equiv p(u) \tag{7.15}$$

(see (4.44) and (4.50)). Writing (7.13) in terms of $p(u)$ one obtains

$$\langle T_{uu} \rangle_{\text{ren}} = (1/12\pi)(p')^{\frac{1}{2}} \partial_u^2 (p')^{-\frac{1}{2}}. \tag{7.16}$$

As $\langle T_{uu} \rangle_{\text{ren}}$ is a function of u only, it is constant along the retarded null rays $u = \text{constant}$ from $\mathscr{I}^+$ back to the surface of the mirror. Thus energy is created at the mirror itself and flows away undiminished to the right. Notice that although one cannot pin down where the individual quanta are created, the source of energy is unambiguous. There is no energy flux from right to left. This asymmetry results from the choice of retarded boundary conditions implied through (7.6). In a situation in which static in and out regions exist, such a choice is equivalent to taking $|0, \text{in}\rangle$ rather than $|0, \text{out}\rangle$ as the quantum state. To see this, suppose that $z(t) = \text{constant}$ for $t < t_0$, thereby defining an in region ($t < t_0$). Then from (7.8) $f'(\bar{u}) = 1$ for $t < t_0$ and it is clear that the 'conformal vacuum' defined from (7.7) agrees with the in vacuum, which is the conventional vacuum state for a static system. A similar relation between the advanced boundary condition and the out vacuum can be achieved by reversing the roles of u and v in (7.6) (Fulling & Davies 1976).

Transforming (7.14) and (7.16) to (t, x) coordinates, and using (4.44) and (7.15) to write the result in terms of the mirror trajectory $z(\tau_u)$, one obtains

$$\langle T_{tt}\rangle_{\text{ren}} = \langle T_{xx}\rangle_{\text{ren}} = -\langle T_{xt}\rangle_{\text{ren}} = -\langle T_{tx}\rangle_{\text{ren}}$$

$$= -\frac{1}{12\pi}\frac{(1-\dot{z}^2)^{\frac{1}{2}}}{(1-\dot{z})^2}\frac{\mathrm{d}}{\mathrm{d}\tau_u}\left[\frac{\ddot{z}}{(1-\dot{z}^2)^{\frac{3}{2}}}\right]. \tag{7.17}$$

Here $z = z(\tau_u)$ and the dot denotes differentiation with respect to τ_u. The parameter τ_u (defined by (4.44)) is, in fact, the time coordinate of the mirror when its trajectory intersects the retarded null ray u. Note that for non-relativistic motion, (7.17) is simply $-(12\pi)^{-1}\dddot{z}$. The right-hand side of (7.17) may be written

$$-\frac{1}{12\pi}\frac{(1-V^2)^{\frac{1}{2}}}{(1-V)^2}\frac{\mathrm{d}\alpha}{\mathrm{d}\tau_u} \tag{7.18}$$

where V is the mirror velocity $\dot{z}$ and α is the proper acceleration. As $|V| < 1$, (7.18) changes sign according to whether α is increasing or decreasing in time. Thus, the mirror may radiate *negative* energy if its acceleration is increasing to the right, or decreasing to the left. The emission of negative energy is a purely quantum phenomenon. It opens up the possibility of unusual new physical processes not encountered in classical theory. We shall see in the next chapter that negative energy fluxes play a role in the quantum evaporation of black holes. It might be supposed that a moving mirror could be used to violate the second law of thermodynamics by radiating negative energy into a hot body, thereby cooling it. However, if (7.17) is integrated over time between periods of mirror stasis, it is always positive definite, i.e., the negative flux is restricted to finite intervals (Fulling & Davies 1976). Ford (1978) has shown that negative energy fluxes cannot be sustained for long enough to reduce the entropy of a hot body by more than would be expected on the basis of ordinary thermal fluctuations.

One trajectory of interest occurs if α = constant, i.e., uniform proper acceleration, corresponding to a hyperbolic trajectory. From (7.18) it follows that $\langle T_{\mu\nu}\rangle_{\text{ren}} = 0$ which implies that no energy at all is radiated during intervals when the mirror acceleration is uniform. However in §4.4 we treated a special case of uniform acceleration and computed the non-trivial Bogolubov transformation (4.64) and (4.65). Clearly, the mirror emits *particles*, and from the discussion associated with (3.80) it is also clear that a particle detector would respond to the presence of quanta. Nevertheless no *energy* is radiated. This example beautifully illustrates the looseness of the relation between particles and energy–momentum. The presence of quanta need not imply the presence of energy.

There appears to be a paradox concerning how the particle detector can, in the absence of field energy, absorb quanta and make a transition to an excited state. The resolution is that in so doing, the detector emits negative energy into the field to compensate. The emission of negative energy by mirrors and detectors is not without precedent in quantum field theory. It is possible to construct many-particle states with negative or zero fluxes even in the absence of mirrors or detectors (Epstein, Gaser & Jaffe 1965; see also Appendix A of Davies & Fulling 1977*a*).

Another mirror trajectory of interest is (4.51) which gives rise to a thermal flux at late times. Applying (7.16) to (4.52) yields

$$\langle T_{tt}\rangle_{\text{ren}} = \frac{\kappa^2}{48\pi}, \quad t \to \infty \tag{7.19}$$

which is the energy of a thermal flux of radiation with temperature $\kappa/2\pi$. This result is in complete agreement with the computation of the thermal spectrum given in §4.4. Indeed, using the thermal spectrum (4.61)

$$\langle T_{tt}\rangle = \frac{1}{2\pi}\int_0^\infty \frac{\omega}{e^{2\pi\omega/\kappa} - 1}\mathrm{d}\omega = \frac{\kappa^2}{48\pi}, \tag{7.20}$$

which shows that the rate of creation of particles in the mode ω is $(1/2\pi)(e^{2\pi\omega/\kappa} - 1)^{-1}$. (This result is deduced from (4.61) by constructing wave packets to convert the particle *number* per mode, $|\beta_{\omega\omega'}|^2$, to a number rate. See §8.1.)

An interesting conclusion from (7.20) is that, in the case of thermal emission, the energy emitted is given by the particle number per mode, weighted by the energy of one quantum ω, integrated over all modes. We have seen that in general there is no simple relation between the number of quanta present and the field energy. For thermal radiation, however, the naive relation of '$\hbar\omega$ per quantum' survives. This can be traced to the complete absence of correlations between the modes in the thermal case (see also chapter 8). It will not be true for all mirror trajectories.

It is also possible to treat a problem of two mirrors, one static and one moving, which involves a 'disturbed Casimir' contribution from the first term on the right-hand side of (6.136). The results (Fulling & Davies 1976) are

$$\left.\begin{aligned}\langle T_{tt}\rangle_{\mathrm{ren}} &= \langle T_{xx}\rangle_{\mathrm{ren}} = \Lambda(u)+\Lambda(v)\\ \langle T_{tx}\rangle_{\mathrm{ren}} &= \langle T_{xt}\rangle_{\mathrm{ren}} = \Lambda(v)-\Lambda(u)\end{aligned}\right\} \tag{7.21}$$

where

$$\Lambda = -(1/24\pi)[(R'''/R') - \tfrac{3}{2}(R''/R') + \pi^2(R')^2/2L^2]. \tag{7.22}$$

The prime denotes differentiation with respect to the argument and the function R is determined by the equation

$$R[t+z(t)] = R[t-z(t)] + 2L \tag{7.23}$$

in terms of the moving mirror trajectory $z(t)$. The static mirror is at $x = 0$, and before $t = 0$ the other mirror is static at $x = L$. The vacuum is defined in this 'in' region ($t < 0$). Note that (7.21) represent a right- and left-moving flux superimposed. This arises because radiation created by the moving mirror reflects from the static mirror. In the special case $z(t) = L$ for all t, (7.23) yields $R(u) = u$ and (7.21) gives $\langle T_{tt}\rangle = -\pi/(24L^2)$, which is the Casimir energy for plate separation L. This result is one-quarter of (4.8), the difference arising due to the use of vanishing, rather than periodic, boundary conditions.

Another interesting case occurs for $F(x) = G(x) = \exp(ax)$, a = constant, in (7.4). Once again this is flat spacetime, but with metric

$$ds^2 = e^{2a\eta}(d\eta^2 - dx^2)$$

i.e., the Milne universe discussed in §5.3. Substituting $C = e^{2a\eta}$ into the general formula (6.136) and, as the two-dimensional Milne universe is conformal to the whole of two-dimensional Minkowski space, setting the first term on the right-hand side equal to zero, gives

$$\langle T_t{}^t\rangle_{\mathrm{ren}} = -\langle T_x{}^x\rangle_{\mathrm{ren}} = -1/24\pi t^2, \tag{7.24}$$

in agreement with (5.46).

Conformally non-trivial two-dimensional exactly soluble examples are few in number. One case is de Sitter space, which can be treated in the same way as its four-dimensional counterpart considered in §6.4. Another is the Robertson–Walker spacetime with conformal factor

$$C(\eta) \propto e^{\alpha\eta^2}, \quad \alpha \text{ constant}. \tag{7.25}$$

With the choice (7.25), $RC = 2\alpha$ and the normalized massless mode solutions of (5.4), with $\xi(n) = \xi(2) = 0$, and arbitrary ξ are

$$(4\pi\omega)^{-\frac{1}{2}} e^{i(kx - \omega\eta)} \tag{7.26}$$

where

$$\omega^2 = k^2 + \beta^2 \tag{7.27}$$

and

$$\beta^2 \equiv 2\alpha\xi.$$

Thus, using a vacuum state based on positive frequency modes (7.26) we have

$$G^{(1)}(x'', x') = (1/2\pi)\,\mathrm{Re} \int_{-\infty}^{\infty} \omega^{-1} e^{ik\Delta x - i\omega\Delta\eta} \mathrm{d}k \tag{7.28}$$

$$= (1/\pi)\,\mathrm{Re}\, K_0[i\beta(\Delta u \Delta v)^{\frac{1}{2}}] \tag{7.29}$$

where $\Delta u = (\eta'' - \eta') - (x'' - x')$, $\Delta v = (\eta'' - \eta') + (x'' - x')$. Then expanding (7.29) in powers of $\Delta u \Delta v$ yields

$$\begin{aligned} G^{(1)}(x'', x') = (1/\pi)\{ &-\gamma - \tfrac{1}{2}\ln|\tfrac{1}{4}\beta^2 \Delta u \Delta v| \\ &+ \tfrac{1}{4}\beta^2 \Delta u \Delta v[\gamma + \tfrac{1}{2}\ln|\tfrac{1}{4}\beta^2 \Delta u \Delta v| + 1] + \ldots\}. \end{aligned} \tag{7.30}$$

We may now use the expansion (6.200), and the corresponding expansion with u replaced by v, to obtain in explicitly geometrical form,

$$\begin{aligned} G^{(1)}(x'', x') = -(1/\pi)\{&(1 - \varepsilon^2 \Sigma \xi R)[\gamma + \tfrac{1}{2}\ln|\varepsilon^2 \xi R|] \\ &+ \varepsilon^2 \Sigma (A_{\alpha\beta} t^\alpha t^\beta \Sigma^{-1} - \xi R)\} \end{aligned} \tag{7.31}$$

where $A_{\alpha\beta}$ is a tensor with components

$$\left.\begin{aligned} A_{uu} &= A_{vv} = \tfrac{1}{48}(-\dot{D} + 2D^2) \\ A_{uv} &= A_{vu} = -\tfrac{1}{24}\dot{D}. \end{aligned}\right\} \tag{7.32}$$

According to our established renormalization procedure, it is now necessary to subtract from (7.31) terms up to adiabatic order two in the DeWitt–Schwinger expansion, i.e., the terms displayed in (6.203). The logarithmically divergent terms cancel, leaving

$$\begin{aligned} &-(1/\pi)\{\tfrac{1}{2}(1 - \varepsilon^2 \Sigma \xi R)\ln|\xi R m^{-2}| \\ &+ \varepsilon^2 \Sigma (A_{\alpha\beta} t^\alpha t^\beta \Sigma^{-1} - \tfrac{3}{2}\xi R + \tfrac{1}{12}R) \\ &- \tfrac{1}{2}(\xi - \tfrac{1}{6}) m^{-2} R + O(m^2)\}. \end{aligned} \tag{7.33}$$

Comparison of (7.33) with (6.205) leads to the following identifications, dropping terms of $O(m^2)$,

$$\begin{aligned} c &= -(1/2\pi)[\ln|\xi R m^{-2}| - (\xi - \tfrac{1}{6})m^{-2}R] \\ e_{\alpha\beta} &= -(1/\pi)A_{\alpha\beta} \\ f &= (1/2\pi)(3\xi - \tfrac{1}{6})R, \quad q = r = 0, \end{aligned}$$

so (6.206) yields the following renormalized stress-tensor in the massless limit

$$\begin{aligned} \langle 0|T_{\mu\nu}|0\rangle_{\text{ren}} = &-(1/48\pi)Rg_{\mu\nu} + \theta_{\mu\nu} + (\xi/4\pi)[(R_{;\mu\nu}/R) \\ &-(R_{;\mu}R_{;\nu}/R^2) + \tfrac{3}{2}Rg_{\mu\nu}], \end{aligned} \tag{7.34}$$

where $\theta_{\mu\nu}$ is the traceless tensor

$$\theta_{\mu\nu} = (1/4\pi)\{A_{\mu\nu} - \tfrac{1}{2}A_\alpha{}^\alpha g_{\mu\nu} - \tfrac{1}{4}[(R_{;\mu\nu}/R) - (R_{;\mu}R_{;\nu}/R^2)] - \tfrac{1}{8}Rg_{\mu\nu}\} \tag{7.35}$$

and we have dropped terms of adiabatic order greater than two coming from differentiation of terms in the DeWitt–Schwinger expansion as explained in chapter 6. (This eliminates a term proportional to m^{-2}.) In arriving at (7.34) we have also used the fact that $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 0$ in two-dimensional spacetime and that $\Box \ln R = -\Box \ln C = -R$ in this example. Note that for $\xi = 0$, the trace of (7.34) is $-R/24\pi$ as expected, while $\theta_{\mu\nu}$ in (7.35) is, of course, identical to $\theta_{\mu\nu}$ given by (6.137). Thus the term proportional to ξ in (7.34) may be regarded as a correction to (6.136) due to non-conformal coupling.

7.2 Robertson – Walker models

Much attention has been devoted to computing stress-tensors for quantum fields propagating in Robertson–Walker background spacetimes. Interest is due to the high degree of symmetry present in these models, as well as their cosmological relevance.

As a first illustration, consider the four-dimensional static, hyperbolic ($K = -1$) Robertson–Walker spacetime, which has been discussed in §5.2. We treat a massless conformally coupled scalar field (Bunch 1978*a*). The $D^{(1)}$ Green function is given from (5.33) by deleting the $i\varepsilon$ and multiplying by 2. Expanding $\sinh \Delta\chi$ to order $(\Delta\chi)^6$ and using the expansions (6.201) one readily obtains

$$D^{(1)}(x'',x') = -\frac{1}{(8\pi^2\varepsilon^2\Sigma)} + \frac{1}{(12\pi^2\Sigma)}(t^1)^2 - \frac{7}{180\pi^2\Sigma^2}\varepsilon^2(t^1)^4 + O(\varepsilon^4). \tag{7.36}$$

With the help of (5.13) we find $R_{\eta\eta} = 0$, $R_{\chi\chi} = 2$ and so may rewrite (7.36) in manifestly geometrical form

$$D^{(1)}(x'',x') = -\frac{1}{8\pi^2\varepsilon^2\Sigma} + \frac{1}{24\pi^2}R_{\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma} - \frac{7\varepsilon^2}{720\pi^2}R_{\alpha\beta}R_{\gamma\delta}\frac{t^\alpha t^\beta t^\gamma t^\delta}{\Sigma^2} + O(\varepsilon^4). \tag{7.37}$$

Comparison of (7.37) with (6.204) for the static hyperbolic model reveals that their massless terms are identical. Moreover, the coefficient of the m^{-2} (final) term in (6.204) vanishes for this spacetime, so there is no conformal anomaly. Hence, when $G^{(1)}$ is subtracted from $D^{(1)}$ to yield the renormalized Green function, one obtains precisely zero.

We therefore have the result

$$\langle 0|T_{\mu\nu}|0\rangle_{\text{ren}} = 0, \tag{7.38}$$

where the vacuum state here is the conformal vacuum, defined by positive frequency modes (5.28), and is associated with the conformal Killing vector related to the ∂_η Killing vector on Rindler spacetime (see §5.5), This result was arrived at indirectly, using the Rindler stress-tensor, in §6.3 (6.158). We could alternatively have used the above calculation together with (6.158) to determine the Rindler stress-tensor (6.157).

As a second illustration, consider the $K = 0$ and $+1$ models, but with C chosen to be a function of time. The $D^{(1)}$ function in the conformal vacuum is given by conformal scaling of (2.79) and (5.32) respectively (putting $\varepsilon = 0$ and doubling the latter). The two results may be written

$$D^{(1)}(x'',x') = \frac{KC^{-\frac{1}{2}}(\eta'')C^{-\frac{1}{2}}(\eta')}{4\pi^2[\cos(K^{\frac{1}{2}}\Delta\eta) - \cos(K^{\frac{1}{2}}\Delta\chi)]}, \tag{7.39}$$

in which, without loss of generality, we have chosen the separation between x'' and x' to lie in the (η,χ) plane.

Using the expansions (6.201), and the result

$$C^{-\frac{1}{2}}(\eta'')C^{-\frac{1}{2}}(\eta') = C^{-1}\{1+\varepsilon^2[(-\tfrac{1}{2}\dot{D}+\tfrac{1}{4}D^2)(t_1^0)^2+\tfrac{1}{4}D^2\Upsilon(t_1^1)^2]$$
$$+\varepsilon^4[(-\tfrac{1}{24}\dddot{D}+\tfrac{7}{48}\ddot{D}D+\tfrac{5}{24}\dot{D}^2-\tfrac{5}{16}\dot{D}D^2+\tfrac{1}{16}D^4)(t_1^0)^4$$
$$+(\tfrac{7}{48}\ddot{D}D+\tfrac{1}{12}\dot{D}^2-\tfrac{31}{48}\dot{D}D^2+\tfrac{1}{4}D^4)\Upsilon(t_1^0t_1^1)^2$$
$$+(-\tfrac{1}{24}\dot{D}D^2+\tfrac{1}{16}D^4)\Upsilon^2(t_1^1)^4]\}+\mathrm{O}(\varepsilon^6), \tag{7.40}$$

which is valid in any Robertson–Walker spacetime, one obtains

$$D^{(1)}(x'',x') = -(1/8\pi^2\varepsilon^2\Sigma)\{1+\varepsilon^2[(-\tfrac{1}{3}\dot{D}+\tfrac{1}{12}D^2+\tfrac{1}{3}K)(t_1^0)^2$$
$$+(\tfrac{1}{12}D^2+\tfrac{1}{3}K)\Upsilon(t_1^1)^2]+\varepsilon^4[(-\tfrac{1}{30}\dddot{D}+\tfrac{1}{10}\ddot{D}D$$
$$+\tfrac{7}{60}\dot{D}^2-\tfrac{2}{15}\dot{D}D^2+\tfrac{1}{60}D^4-\tfrac{1}{6}\dot{D}K+\tfrac{1}{12}D^2K+\tfrac{3}{45}K^2)(t_1^0)^4$$
$$+(\tfrac{1}{10}\ddot{D}D+\tfrac{11}{180}\dot{D}^2-\tfrac{23}{72}\dot{D}D^2+\tfrac{31}{360}D^4-\tfrac{5}{18}\dot{D}K+\tfrac{7}{18}D^2K$$
$$+\tfrac{8}{45}K^2)\Upsilon(t_1^0t_1^1)^2+(-\tfrac{1}{40}\dot{D}D^2+\tfrac{1}{60}D^4+\tfrac{1}{12}D^2K$$
$$+\tfrac{3}{45}K^2)(t_1^1)^4\Upsilon^2]\}. \tag{7.41}$$

If one now takes linear combinations of geometrical tensors with the appropriate adiabatic order, it is possible to express (7.41) entirely in terms of geometrical tensors:

$$D^{(1)}(x'',x') = -\frac{1}{8\pi^2\varepsilon^2\Sigma}+\frac{1}{24\pi^2}\left[R_{\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma}-\tfrac{1}{6}R\right]$$
$$+\frac{\varepsilon^2\Sigma}{1440\pi^2}\left[2R_\alpha{}^\lambda R_{\lambda\beta}\frac{t^\alpha t^\beta}{\Sigma}+4RR_{\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma}-\tfrac{1}{3}R^2\right.$$
$$\left.-R_{;\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma}-14R_{\alpha\beta}R_{\gamma\delta}\frac{t^\alpha t^\beta t^\gamma t^\delta}{\Sigma^2}+6R_{\alpha\beta;\gamma\delta}\frac{t^\alpha t^\beta t^\gamma t^\delta}{\Sigma^2}\right]. \tag{7.42}$$

Note that the two other available terms $R^{\alpha\beta}R_{\alpha\beta}$ and $\Box R$ do not occur here. The fact that (7.42) is purely geometrical is a consequence of the conformal triviality of this example.

To renormalize, we now subtract terms up to adiabatic order four in the DeWitt–Schwinger expansion (6.204). The divergences and the terms involving four t^α vectors cancel, leaving

$$-\frac{R}{144\pi^2}+\frac{\varepsilon^2\Sigma}{1440\pi^2}\left[(-2R_\alpha{}^\lambda R_{\lambda\beta}+\tfrac{16}{3}RR_{\alpha\beta}-R_{;\alpha\beta})\frac{t^\alpha t^\beta}{\Sigma}\right.$$
$$\left.-\tfrac{2}{3}R^2+\Box R+R^{\alpha\beta}R_{\alpha\beta}\right]-\frac{m^{-2}}{1440\pi^2}(R^{\alpha\beta}R_{\alpha\beta}-\tfrac{1}{3}R^2-\Box R). \tag{7.43}$$

Comparison with (6.205) reveals that we have terms of the form c, $e_{\alpha\beta}$ and f, so (6.206) yields for the renormalized stress-tensor (after dropping terms of order greater than four coming from differentiation of $G^{(1)}_{\mathrm{DS}}$)

$$\begin{aligned}\langle 0|T_{\mu\nu}|0\rangle_{\mathrm{ren}} &= (1/2880\pi^2)[(-\tfrac{1}{3}R_{;\mu\nu} + R_\mu{}^\rho R_{\rho\nu} - RR_{\mu\nu}) \\ &\quad + g_{\mu\nu}(\tfrac{1}{3}\Box R - \tfrac{1}{2}R^{\rho\tau}R_{\rho\tau} + \tfrac{1}{3}R^2)] \\ &= -(1/2880\pi^2)[\tfrac{1}{6}{}^{(1)}H_{\mu\nu} - {}^{(3)}H_{\mu\nu}],\end{aligned} \tag{7.44}$$

which is in agreement with (6.141), with the first term on the right-hand side set to zero because the $K = 0$ and $+1$ spacetimes have a conformal vacuum based on the Minkowski vacuum (see §5.5 and fig. 20). The result (7.44) was originally obtained by the method used above for $K = 0$ by Davies, Fulling, Christensen & Bunch (1977) and for $K = 0$ and $+1$ by Bunch & Davies (1977*a*).

A similar calculation may be carried out for the $K = -1$ case in the same conformal vacuum as in (7.38). The purely geometrical $D^{(1)}$ in (7.42) is found to be augmented by a local, but non-geometrical term, which, when substituted in (6.206), yields a term $(\zeta/a(\eta))^4\langle T_{\mu\nu}\rangle_{\mathrm{Rindler}}$ that appears in (6.141) in this case (see (6.158) and accompanying discussion).

More ambitious calculations have been given by Bunch & Davies (1978*b*), involving a non-conformally coupled massless scalar field in Robertson–Walker spacetime with special choices of the scale factor that enable the Green function $D^{(1)}$ to be computed in terms of known functions. For example, with $K = 0$ and

$$a(t) = \alpha t^c \tag{7.45}$$

(α, c constant) one has

$$C(\eta) = \alpha^{2/(1-c)}(1-c)^{2c/(1-c)}\eta^{2c/(1-c)} \tag{7.46}$$

and the *minimally* coupled field equation $\Box\phi = 0$ possesses mode solutions (5.3) with $n = 4$ and (Ford & Parker 1977)

$$\chi_k(\eta) = C^{\frac{1}{4}}(\eta)(|b|\eta/a_0^2)^{1/2b}[c_1 H^{(1)}_\nu(k\eta) + c_2 H^{(2)}_\nu(k\eta)], \tag{7.47}$$

where $H^{(1)}, H^{(2)}$ are Hankel functions, $k = |\mathbf{k}|$ and

$$\left.\begin{aligned} b &= (1-c)/(1-3c) \\ \nu &= 1/(2|b|) \\ a_0 &= \alpha[\alpha^3(1-3c)]^{c/(1-3c)}. \end{aligned}\right\} \tag{7.48}$$

The coefficients c_1 and c_2 are complex numbers subject to the Wronskian condition (5.6), which here reduces to

$$|c_2|^2 - |c_1|^2 = \pi/4b. \qquad (7.49)$$

A minimally coupled scalar field can be used to describe linearized gravitons in a Robertson–Walker model universe (Grishchuk 1974, 1975).

We choose the vacuum state defined by $c_1 = 0$, as this reduces to the standard Minkowski space vacuum in the limit $c \to 0$. The Green function $D^{(1)}(x'', x')$ is easily evaluated as a mode integral

$$\begin{aligned} D^{(1)}(x'', x') &= \int [u_{\mathbf{k}}(x'')u_{\mathbf{k}}^*(x') + u_{\mathbf{k}}^*(x'')u_{\mathbf{k}}(x')]\,\mathrm{d}^3k \\ &= (1/8\pi\eta''\eta')C^{-\frac{1}{2}}(\eta'')C^{-\frac{1}{2}}(\eta')(\tfrac{1}{4} - \nu^2)\sec \pi\nu \\ &\quad \times F(\tfrac{3}{2} + \nu, \tfrac{3}{2} - \nu; 2 : 1 + (\Delta\eta^2 - \Delta z^2)/4\eta''\eta') \end{aligned} \qquad (7.50)$$

where F is a hypergeometric function and we have separated the points x'', x' in the (η, z) plane, obtaining $\Delta\eta = \eta'' - \eta'$, $\Delta z = z'' - z' = x^{3''} - x^{3'}$.

It is now necessary to expand (7.50) in powers of ε up to order ε^2 using (6.201), with r replaced by z throughout, and using (7.40). By taking linear combinations of all available geometrical tensors of correct adiabatic order, the result can be cast in the following form

$$\begin{aligned} D^{(1)}(x'', x') = &-\frac{C^{-\frac{1}{2}}(\eta'')C^{-\frac{1}{2}}(\eta')}{2\pi^2(\Delta\eta^2 - \Delta z^2)} + \left[\tfrac{1}{2}\ln\left|\frac{\varepsilon^2}{C\eta^2}\right| + \gamma + \tfrac{1}{2}\psi(\tfrac{3}{2} + \nu) + \tfrac{1}{2}\psi(\tfrac{3}{2} - \nu)\right] \\ &\times \left[-\frac{R}{24\pi^2} + \frac{\varepsilon^2\Sigma}{288\pi^2}\left(4RR_{\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma} - 2R_{;\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma} + 2\Box R - R^2\right)\right] \\ &+ \frac{R}{48\pi^2} - \frac{\varepsilon^2\Sigma}{144\pi^2}\left[RR_{\alpha\beta}\frac{t^\alpha t^\beta}{\Sigma} - \tfrac{19}{24}R^2 + \frac{3R}{C\eta^2}\right] + \mathrm{O}(\varepsilon^4). \end{aligned} \qquad (7.51)$$

The first term on the right is the expression for $D^{(1)}(x'', x')$ in the conformally coupled case ($\xi = 1/6$), so we see that the effect of putting $\xi = 0$ is to add a correction term. Although still entirely a local bi-scalar, $D^{(1)}$ does now contain *non-geometrical* terms, e.g., the η^{-2} term, which arises because of the departure from conformal triviality. Note that when $R = 0$, the correction term vanishes. This is the case where there is no distinction between conformal and minimal coupling.

Renormalization is effected by subtracting the DeWitt–Schwinger terms (6.204) with $\xi = 0$ from (7.51). Because we already know the result of subtracting (6.204) with $\xi = 1/6$ from the *conformal* $D^{(1)}(x'', x')$, i.e., the first term of (7.51), the present result is simply (7.44) plus the correction terms resulting from the remainder of (7.51), and from the difference between the conformal and minimal $G_{\mathrm{DS}}^{(1)}$. The result is

$$\begin{aligned}\langle 0|T_{\mu\nu}|0\rangle_{\text{ren}} = & -(1/2880\pi^2)[\tfrac{1}{6}{}^{(1)}H_{\mu\nu} - {}^{(3)}H_{\mu\nu}] \\ & -(1/1152\pi^2){}^{(1)}H_{\mu\nu}[\ln(R/m^2) + \psi(\tfrac{3}{2}+\nu) + \psi(\tfrac{3}{2}-\nu) + \tfrac{4}{3}] \\ & +(1/13824\pi^2)[24\Box R g_{\mu\nu} + 24RR_{\mu\nu} + 3R^2 g_{\mu\nu}] \\ & - Rg_{\mu\nu}/192\pi^2 C\eta^2. \end{aligned} \tag{7.52}$$

The logarithmic term, which always arises when conform triviality is broken, contains an arbitrary mass scale m. Rescaling m merely adds multiples of the conserved tensor ${}^{(1)}H_{\mu\nu}$ to (7.52). However, as pointed out in §6.2 all renormalized stress-tensors are ambiguous up to multiples of ${}^{(1)}H_{\mu\nu}$, because such a term appears on the *left*-hand side of the gravitational field equations. Thus, one could remove all ${}^{(1)}H_{\mu\nu}$ terms from (7.52) by renormalizing α in (6.220). There is then no problem in the massless limit $(m \to 0)$; one merely has an additional (infrared this time) renormalization of α. Note that one could absorb the non-geometrical ψ-function terms and the factor $\frac{4}{3}$ into a rescaling of m.

The remaining terms of (7.52) are all local, some being 'accidently' geometrical, due to the high degree of symmetry in the model, and the final term being non-geometrical. There is no possibility of placing all the extra geometrical terms on the left-hand side of the gravitational field equations, as they are not conserved without the ln R and η^{-2} terms.

In a more general, or less symmetric example, one would expect not only non-geometrical contributions to $\langle T_{\mu\nu}\rangle_{\text{ren}}$, but *non-local* terms as well; for example, had a more complicated state been chosen, or if the spacetime were anisotropic. In §7.3 an example will be discussed where non-local terms appear explicitly.

7.3 Perturbation calculation of the stress-tensor

Exactly soluble models, whilst invaluable for pedagogic purposes, are of little value in practical calculations. To proceed beyond these models, one must turn either to numerical computation, or, as in §5.6, to approximation methods.

Indeed, the perturbation technique discussed in §5.6 is particularly useful in the calculation of $\langle T_{\mu\nu}\rangle_{\text{ren}}$. One simply iterates (5.105) to form approximate functions $\chi_{\mathbf{k}}$, which reduce in the in region to functions $\chi_{\mathbf{k}}^{\text{in}}$, these being positive frequency with respect to the conformal Killing vector ∂_η in the in region (see (5.103)). Then one can use the associated modes to construct $G^{(1)}(x'',x') = \langle \text{in},0|\{\phi(x''),\phi(x')\}|0,\text{in}\rangle$, where $|0,\text{in}\rangle$ is the conformal vacuum in the in region, which is renormalized and differentiated to form the vacuum expectation value of the stress-tensor in the usual way. Davies & Unruh (1979) have computed $\langle \text{in},0|T_{\mu\nu}|0,\text{in}\rangle_{\text{ren}}$ for $h_i = 0, m = 0$, up to second order in $(\xi - \frac{1}{6})$. They find

$$\begin{aligned}
&\langle \text{in},0|T_{\mu\nu}|0,\text{in}\rangle_{\text{ren}} \\
&\quad = (1/2880\pi^2)\{-\tfrac{1}{6}{}^{(1)}H_{\mu\nu} + {}^{(3)}H_{\mu\nu} + 10(\xi - \tfrac{1}{6}){}^{(1)}H_{\mu\nu} \\
&\qquad + 180(\xi - \tfrac{1}{6})^2[{}^{(1)}H_{\mu\nu}(1 + \ln C^{\frac{1}{2}}) + C_{\mu\nu} \\
&\qquad + \mathscr{H}_{\mu\nu}(C^{-1}\int_{-\infty}^{\eta} \tilde{V}'(\eta_1)\ln(\mu|\eta - \eta_1|)\mathrm{d}\eta_1 \\
&\qquad + \tfrac{1}{2}e_{\mu\nu}C^{-1}\tilde{V}(\eta)\int_{-\infty}^{\eta} \tilde{V}'(\eta_1)\ln(\mu|\eta - \eta_1|)\mathrm{d}\eta_1 \\
&\qquad - \tfrac{1}{2}e_{\mu\nu}C^{-1}\int_{-\infty}^{\eta}\mathrm{d}\eta_1\int_{-\infty}^{\eta}\mathrm{d}\eta_2\tilde{V}'(\eta_1)\tilde{V}'(\eta_2)\ln(\mu|\eta_1 - \eta_2|)]\},
\end{aligned} \tag{7.53}$$

where $\tilde{V}(\eta) = R(\eta)C(\eta)$, μ is an arbitrary mass scale, $e_{\mu\nu}$ is a tensor, diagonal in Cartesian coordinates, with constant components $e_{00} = 1$, $e_{ii} = \frac{1}{3}$, and primes denote differentiation of a function with respect to its argument. The tensor operator $\mathscr{H}_{\mu\nu}$ is defined by

$$2(\nabla_\mu\nabla_\nu - g_{\mu\nu}\Box + R_{\mu\nu} - \tfrac{1}{4}Rg_{\mu\nu})$$

so that $\mathscr{H}_{\mu\nu}[R(\eta)] = {}^{(1)}H_{\mu\nu}(\eta)$ (see (6.53)). Finally, $C_{\mu\nu}$ is a local, but non-geometrical tensor with components

$$\begin{aligned}
C_{\eta\eta} &= C^{-1}(-\tfrac{9}{2}\dot{D}D^2 - \tfrac{9}{4}D^4) \\
C_{xx} = C_{yy} = C_{zz} &= C^{-1}(6\ddot{D}D + \tfrac{9}{2}\dot{D}^2 + \tfrac{9}{2}\dot{D}D^2 - \tfrac{15}{8}D^4).
\end{aligned}$$

Expression (7.53), which is non-local only at second order in $(\xi - \frac{1}{6})$, reduces to (5.122) (with $h_i = m = 0$) in the out region $\eta \to \infty$, if one uses (5.102) and $\rho = \langle T_0{}^0\rangle = C^{-1}\langle T_{00}\rangle$.

One advantage of the perturbation technique is that, with each higher order of perturbation, an additional factor of ω^{-1} appears multiplying the correction to χ_k (see (5.105)). This means that divergences only occur in $\langle T_{\mu\nu}\rangle$ up to the second order of the perturbation. The higher orders all yield finite additional contributions, without the need for regularization, making them particularly amenable to numerical calculation.

An alternative approach to perturbation theory is to work with the effective action (Hartle 1977, Fischetti, Hartle & Hu 1979, Hartle & Hu 1979, Hu 1980, Hartle 1981). Consider the slightly more general metric

$$ds^2 = C(\eta)[d\eta^2 - (e^{2\beta(\eta)})_{ij} dx^i dx^j] \tag{7.54}$$

where $\beta(\eta)$ is a symmetric, traceless, 3×3 matrix. Note that to first order $2\beta_{ii} = h_i$.

For this spacetime the Feynman Green function for a massless, conformally coupled field satisfies an equation of the form

$$(\Box_x + \tfrac{1}{6}R_1)D_F(x, x') = -[-g(x)]^{-\frac{1}{2}}\delta^4(x - x') - VD_F(x, x') \tag{7.55}$$

where $\Box_x$ is the wave operator and R_1 is the Ricci scalar for the case of exact isotropy, and the operator V symbolizes the (small) correction to these quantities due to the anisotropy.

Equation (7.55) possesses the formal solution

$$D_F = D_F^{(0)} + D_F^{(0)}VD_F \tag{7.56}$$

where $D_F^{(0)}$ is the Feynman propagator in the conformal vacuum (i.e., (3.154) with $n = 4$); this being chosen as the natural propagator in the limit of exact isotropy. This equation for G_F may be approximated iteratively

$$D_F = D_F^{(0)} + D_F^{(0)}VD_F^{(0)} + D_F^{(0)}VD_F^{(0)}VD_F^{(0)} + \dots \tag{7.57}$$

and used in (6.25) to expand $\ln(-D_F)$, giving a perturbation series for the effective action

$$W = \sum_{i=0}^{\infty} W^{(i)} \tag{7.58}$$

where

$$W^{(0)} = -\tfrac{1}{2}i\,\mathrm{tr}[\ln(-D_F^{(0)})] \tag{7.59}$$

$$W^{(1)} = -\tfrac{1}{2}i\,\mathrm{tr}(VD_F^{(0)}) \tag{7.60}$$

$$W^{(2)} = -\tfrac{1}{4}i\,\mathrm{tr}(VD_F^{(0)}VD_F^{(0)}). \tag{7.61}$$

The zeroth order term $W^{(0)}$ was studied in §6.3, and gave rise to the stress-tensor (6.141). The dimensionally regularized $D_F^{(0)}(x, x)$ vanishes (see page 170), so that $W^{(1)} = 0$. However, $W^{(2)}$, being quadratic in $D_F^{(0)}$, has a nonzero dimensionally regularized value. Writing out the part of the dimensionally regularized $W^{(2)}$ which is of second order in β (there are also third and fourth order terms in (7.61)), one obtains (Hartle & Hu 1979)

$$W^{(2)} = \int d^n x \int d^n x' \beta^{ij}(\eta) K_{ijkl}(x - x') \beta^{kl}(\eta') \tag{7.62}$$

where

$$K_{ijkl}(x) = (2\pi)^{-n} \int d^n k\, e^{i\mathbf{k}\cdot\mathbf{x} - ik_0\eta_0} \hat{K}_{ijkl}(k) \tag{7.63}$$

and

$$\begin{aligned}\hat{K}_{ijkl}(k) = & -(1/1920\pi^2)[\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}](\mathbf{k}^2 - k_0^2)^2 \\ & \times [(1/(n-4) + \tfrac{1}{2}\ln(\mathbf{k}^2 - k_0^2) - \text{constant} + O(n-4)] \\ & + (\text{terms which give vanishing contribution to } W^{(2)}).\end{aligned} \tag{7.64}$$

In arriving at (7.64), the explicit forms of $G_F^{(0)}$ (in n dimensions) and V have been used, and the result has been expanded about $n = 4$. The pole term can be evaluated as

$$-[1/1920\pi^2(n-4)] \int d^n x (-g)^{\frac{1}{2}} C^{\alpha\beta\gamma\delta} C_{\alpha\beta\gamma\delta}$$

in which $(-g)^{\frac{1}{2}}$ and $C_{\alpha\beta\gamma\delta}$ are four-dimensional quantities. This can be rewritten as (cf. (6.138))

$$-[1/16\pi^2(n-4)] \int d^n x (-g)^{\frac{1}{2}} \alpha F(x) + O(1), \tag{7.65}$$

with $(-g)^{\frac{1}{2}}$ and F (see (6.104)) constructed from n-dimensional quantities and α being given by (6.106). The pole term in (7.65) can be removed by renormalization as discussed in §§6.2 and 6.3, leaving a finite $W^{(2)}$ with contributions from the logarithm in (7.64) and the finite (O(1)) terms in (7.65). This can be calculated to be

$$\begin{aligned}W^{(2)}_{\text{ren}} = (V/2880\pi^2) \Bigg\{ & \int_{-\infty}^{\infty} d\eta [-(\ddot{a}/a + \dot{a}^2/a^2)\dot{\beta}_{ij}\dot{\beta}^{ij} + 3[\tfrac{1}{2}i\pi + \ln a]\ddot{\beta}_{ij}\ddot{\beta}^{ij}] \\ & - 3\int_{-\infty}^{\infty} d\eta \int_{-\infty}^{\infty} d\eta' \ddot{\beta}_{ij}(\eta) K(\eta - \eta') \ddot{\beta}^{ij}(\eta') \Bigg\}\end{aligned} \tag{7.66}$$

where V is the volume of space, $a^2 = C(\eta)$ and

$$K(\eta) = (1/\pi) \int_0^{\infty} \cos(\omega\eta) \ln(\omega/\mu)\, d\omega,$$

μ being the arbitrary mass scale introduced in renormalization.

Functional differentiation of the perturbation series for W can be used to give the vacuum expectation value of the stress-tensor. This will consist of the usual isotropic anomaly terms (6.141) from $W^{(0)}$, together with anisotropy correction terms to this anomalous piece arising from (7.65), and finally the non-anomalous terms from (7.66), which may be regarded as representing particle creation induced by the perturbation. The explicit results have been given by Hartle & Hu (1979).

As yet we have not discussed the boundary conditions that have been built into the formal (integral) equation (7.56) by the choice of $D_F^{(0)}$ as the Feynman propagator in the conformal vacuum. This boundary condition determines which vacuum expectation value of the stress-tensor will be arrived at by variation of (7.58)

To examine this equation let us suppose that the anisotropy vanishes in the distant past and future, so that there exist in and out conformally flat regions and associated in and out conformal vacua. To establish the boundary conditions built into (7.56) we derive the equation in terms of field operators $\phi(x)$ rather than Green functions.

Using Gauss' theorem one can write

$$
\begin{aligned}
-\int \mathrm{d}^n y[-g(y)]^{\frac{1}{2}}\,[K_y D_F^{(0)}(x,y)]\,T(\phi(y)\phi(x'))
&= -\int \mathrm{d}^n y[-g(y)]^{\frac{1}{2}}\,D_F^{(0)}(x,y) K_y T(\phi(y)\phi(x')) \\
&\quad + \int_{\substack{\Sigma \\ y^0 \to \infty}} \mathrm{d}\Sigma_y^{\mu}[-g(y)]^{\frac{1}{2}}\,D_F^{(0)}(x,y)\,\overleftrightarrow{\nabla}_{\mu}{}^{y}(\phi(y)\phi(x')) \\
&\quad - \int_{\substack{\Sigma \\ y^0 \to -\infty}} \mathrm{d}\Sigma_y^{\mu}[-g(y)]^{\frac{1}{2}}\,D_F^{(0)}(x,y)\,\overleftrightarrow{\nabla}_{\mu}{}^{y}(\phi x')\phi(y))
\end{aligned}
\tag{7.67}
$$

(for notation see (3.28)) where for generality we work in n dimensions, and

$$K_y \equiv \Box_y + \tfrac{1}{4}[(n-2)/(n-1)]R(y). \tag{7.68}$$

Thus $D_F^{(0)}$, defined by (3.154), satisfies

$$K_y D_F^{(0)}(x,y) = -[-g(y)]^{-\frac{1}{2}}\delta^n(x-y). \tag{7.69}$$

We are also assuming, for simplicity, that the fields vanish at spacelike infinity, so that spatial boundary terms may be ignored.

In the conformally flat in and out regions, the field ϕ will have mode decompositions (3.153), i.e.,

$$\phi(y) \xrightarrow[y^0 \to -\infty]{} \sum_{\mathbf{k}} (a_{\mathbf{k}} u_{\mathbf{k}}^{(0)} + a_{\mathbf{k}}^{\dagger} u_{\mathbf{k}}^{(0)*}) \tag{7.70}$$

$$\phi(y) \xrightarrow[y^0 \to +\infty]{} \sum_{\mathbf{k}} (b_{\mathbf{k}} u_{\mathbf{k}}^{(0)} + b_{\mathbf{k}}^{\dagger} u_{\mathbf{k}}^{(0)*}) \tag{7.71}$$

where the modes $u_{\mathbf{k}}^{(0)}$ are positive frequency with respect to conformal time η and are conformally related to the flat spacetime modes (3.151)

$$u_{\mathbf{k}}^{(0)} = C^{(2-n)/4} \bar{u}_{\mathbf{k}}. \tag{7.72}$$

Because of the anisotropy between the in and out regions, $a_{\mathbf{k}}$ and $b_{\mathbf{k}}$ will not be equal. If they are used to define in and out conformal vacua

$$a_{\mathbf{k}}|0, \mathrm{in}\rangle = 0, \quad b_{\mathbf{k}}|0, \mathrm{out}\rangle = 0, \tag{7.73}$$

then these will not be the same state.

Using the mode decomposition for the flat space Feynman propagator, (3.154) can be written as

$$\left.\begin{aligned} \mathrm{i}D_{\mathrm{F}}^{(0)}(x, y) &= \theta(x^0 - y^0) \sum_{\mathbf{k}} u_{\mathbf{k}}^{(0)}(x) u_{\mathbf{k}}^{(0)*}(y) \\ &\quad + \theta(y^0 - x^0) \sum_{\mathbf{k}} u_{\mathbf{k}}^{(0)*}(x) u_{\mathbf{k}}^{(0)}(y) \\ &\xrightarrow[y^0 \to -\infty]{} \sum_{\mathbf{k}} u_{\mathbf{k}}^{(0)}(x) u_{\mathbf{k}}^{(0)*}(y) \\ &\xrightarrow[y^0 \to +\infty]{} \sum_{\mathbf{k}} u_{\mathbf{k}}^{(0)*}(x) u_{\mathbf{k}}^{(0)}(y). \end{aligned}\right\} \tag{7.74}$$

Employing (7.70), (7.71) and (7.74) in the surface terms of (7.67), applying the orthonormality conditions (3.29), and substituting (7.69) into the left-hand side of (7.67), one obtains

$$\begin{aligned} T(\phi(x)\phi(x')) = &- \int \mathrm{d}^n y [-g(y)]^{\frac{1}{2}} D_{\mathrm{F}}^{(0)}(x, y) K_y T(\phi(y)\phi(x')) \\ &+ \sum_{\mathbf{k}} u_{\mathbf{k}}^{(0)*}(x) b_{\mathbf{k}}^{\dagger} \phi(x') + \sum_{\mathbf{k}} u_{\mathbf{k}}^{(0)}(x) \phi(x') a_{\mathbf{k}}. \end{aligned} \tag{7.75}$$

If one now takes the $\langle \mathrm{out}, 0|\ldots|0, \mathrm{in}\rangle$ vacuum expectation value of (7.75), the final two terms vanish by virtue of (7.73), and one is left with

$$D_{\mathrm{F}}(x, x') = - \int \mathrm{d}^n y [-g(y)]^{\frac{1}{2}} D_{\mathrm{F}}^{(0)}(x, y) K_y D_{\mathrm{F}}(y, x'), \tag{7.76}$$

where

$$D_F(x, x') = \langle \text{out}, 0|T(\phi(x)\,\phi(x'))|0, \text{in}\rangle. \tag{7.77}$$

This will satisfy the n-dimensional generalization of (7.55), namely

$$K_x D_F(x, x') = -[-g(x)]^{-\frac{1}{2}}\delta^n(x - x') - VD_F(x, x'). \tag{7.78}$$

Application of (7.78) to (7.76) finally yields

$$D_F(x, x') = D_F^{(0)}(x, x') + \int d^n y[-g(y)]^{\frac{1}{2}} D_F^{(0)}(x, y)\, VD_F(y, x'), \tag{7.79}$$

which is precisely the explicit form of (7.56). Had we taken the $\langle$in, 0$|\ldots|$0, in$\rangle$ or $\langle$out, 0$|\ldots|$0, out$\rangle$ expectation values of (7.75), then one of the two surface terms would have survived and the resulting equation would differ from (7.56).

The above argument can be generalized to more complicated spacetimes without changing the basic result, i.e., that an equation of the form (7.56) determines a propagator that is of the $\langle$out, 0$|\ldots|$0, in$\rangle$ variety. The reason can be traced to the use of the Feynman propagator, which is formally achieved by the addition of a $(-i\varepsilon)$ term to the left-hand side of (7.55), (see the discussion on page 76).

Methods for solving the integral equations for either $\langle$out$|\ldots|$in$\rangle$ or $\langle$in$|\ldots|$in$\rangle$ propagators using momentum space techniques have been devised by Birrell (1979*c*) (see also Birrell 1979*a* for $\langle$out$|\ldots|$in$\rangle$ only).

Finally, a completely different approach to perturbation theory has been given by Horowitz (1980), who has attempted to use the Wald axioms of §6.6 to restrict the structure of the first order (general) perturbation about Minkowski space of the stress-tensor for a massless, conformally coupled scalar field. If it is assumed that $\langle T_{\mu\nu}\rangle$ does not contain terms of adiabatic order four or greater (e.g., terms of the $\Box Rg_{\mu\nu}$ type) then a unique, non-local result is obtained in terms of an integral over the past null cone through the spacetime point of interest. This perturbation result for $\langle T_{\mu\nu}\rangle$ had previously been derived in momentum space by Capper, Duff & Halpern (1974), and has been extended to cover arbitrary perturbations about conformally flat spacetimes by Horowitz & Wald (1980).

7.4 Cosmological considerations

As remarked in chapter 1, gravitational effects in quantum field theory are likely to be of observational significance only close to microscopic black holes, or in the early phases of the primeval universe. In the absence of observational data concerning black holes, we shall restrict attention to the consequences for physical cosmology of quantum field theory in curved spacetime.

Most cosmologists today assume that the universe had a singular origin about 15 billion years ago, and that the epoch immediately after about one second was characterized by isotropic expansion and thermodynamic equilibrium. As we have seen in previous sections, quantum field energy densities induced by the cosmological gravitational field are characterized by terms of the sort R^2, which become comparable with the gravitational terms in Einstein's equation only in the region of the Planck era (10^{-43} s). Reliable information about the primeval universe does not extend back before about one second after the hypothesized singular origin, so that the effects of interest in this book are unlikely to ever be directly observable. Nevertheless, there are indirect ways in which some of the model predictions can be tested.

Quantum gravitational effects in the primeval universe will result in the production of entropy through particle creation. In addition, back-reaction of the gravitationally induced stress-tensor will modify the cosmological dynamics. In particular, any initial anisotropy and inhomogeneity is likely to result in prolific particle creation (see §5.6) and vacuum stress (see §7.3). Back-reaction would then be expected to result in strong damping of the initial turbulence and irregularity. The fact that the presently observed universe is highly uniform on the large scale suggests that either initial anisotropy and inhomogeneity has been efficiently damped away, or that the universe began with a degree of uniformity that is *a priori* exceedingly improbable (Penrose 1979).

It has been pointed out by Barrow & Matzner (1977) that, on quite general grounds, the entropy per baryon produced by cosmological anisotropy dissipation is highly sensitive to the epoch of dissipation. Anisotropy perturbations can be regarded as behaving like a cosmological fluid with an energy density that varies with the cosmological scale factor as a^{-6} (Misner 1968). On the other hand, the energy density of radiation varies like a^{-4}. Consequently, the radiation entropy resulting in the damping of a given amount of anisotropy varies like a^{-2}. The earlier the dissipation epoch, the more the entropy one obtains.

As explained, the efficiency of quantum gravitational processes rises as the initial singularity is approached ($a \to 0$), which suggests that these effects will result in prolific entropy production at very early times. (For estimates see, for example, Mamaev, Mostepanenko & Starobinsky 1976, Frolov, Mamaev & Mostepanenko 1976.) As the second law of thermodynamics forbids this cosmic entropy from subsequently declining, we may use the currently observed entropy per baryon ratio in the universe to constrain the primeval anisotropy. (We are assuming that the universe contains a fixed nonzero baryon number. Certain recent gauge theories suggest that baryon number may not be conserved under the conditions found in the primeval universe. In this case the entropy per baryon ceases to be a meaningful parameter for judging the initial anisotropy.)

To obtain some idea of the numbers involved, we can use the results of the perturbation analysis of §5.6. Equation (5.122) gives the energy density of created massless particles in the out region where the anisotropy h_i and the non-conformal coupling V have fallen to zero. If h_i and V are appreciable only in some brief interval at conformal time η_0, we may neglect the slow logarithmic factor in the integrand and integrate (5.122) to obtain

$$\rho a^4 = \text{constant} \times [60(\xi - \tfrac{1}{6})^2 R^2(\eta_0) + C^{\alpha\beta\gamma\delta}(\eta_0) C_{\alpha\beta\gamma\delta}(\eta_0)],$$

where the constant is some appropriate numerical factor (not $\gg 1$).

From elementary considerations it is hard to see how a spacetime emerging from the Planck era could avoid attaining values of R^2 and $C^{\alpha\beta\gamma\delta}C_{\alpha\beta\gamma\delta}$ comparable with the Planck values, purely as a result of quantum metric fluctuations. Thus, even ignoring subsequent conformal symmetry breaking, one expects ρ to be comparable with the Planck energy density at the Planck time, redshifted to the present epoch (at least for gravitons; see Berger 1974, Grishchuk 1977). Expressed in thermodynamic language, the Planck temperature (10^{32} K) at the Planck time, redshifted to now, yields a thermal background of a few degrees K (Parker 1976), which is what is actually observed in the case of the photon background. We may therefore conclude that the observed cosmic entropy per baryon is not greatly in excess of the minimum value that is consistent with quantum metric fluctuations in the Planck era.

Further confirmation that large scale irregularities did not survive much beyond the Planck era comes from the calculations of a number of authors on damping due to back-reaction effects (Zel'dovich & Starobinsky 1971, Lukash & Starobinsky 1974, Lukash, Novikov & Starobinsky 1975, Lukash, Novikov, Starobinsky & Zel'dovich 1976, Hu & Parker 1978, Hartle & Hu 1980; see also Hu 1980 and Hartle 1981). Their work on model anisotropic spacetimes indicates that the damping occurs extremely rapidly over less than one Planck time.

Sometimes quantum field effects are invoked as an attempt to find a mechanism whereby the universe might avoid an initial singularity (Ruzmaikina & Ruzmaikin 1970, Nariai 1971, Parker & Fulling 1973, Davies 1977*a*, Melnikov & Orlov 1979, Parnovsky 1979). Of course, such a universe would then be of infinite age, with all the associated thermodynamic problems that implies (see, for example Davies 1974). The Hawking–Penrose singularity theorems (for a review and references see Hawking & Ellis 1973) depend crucially upon so-called energy conditions. In Robertson–Walker spacetimes one of these conditions reduces to requirement that

$$\rho + 3p > 0.$$

However, as we have seen, quantum field effects permit p and ρ to become negative under some circumstances, so the possibility arises that the universe may 'bounce' at some highly dense epoch, rather than encounter a singularity. Indeed we considered such models in §§3.4, 3.5.

In the absence of another length scale in the theory, it seems probable that any 'bounce' would occur close to the Planck regime. Unfortunately, this is precisely where the one-loop approximation is no longer reliable. Attempts have been made to solve the gravitational field equation with a source term given by $\langle T_{\mu\nu} \rangle$ for a massless conformally invariant field given in §6.3 (Davies 1977*a*, Fischetti, Hartle & Hu 1979, Starobinsky 1980). The analysis is greatly complicated by the presence of terms of the form ${}^{(1)}H_{\mu\nu}$ and ${}^{(2)}H_{\mu\nu}$.

If the theory contains a fundamental length scale in addition to the Planck length, then it may be that a 'bounce' can occur well away from the Planck regime. At first sight it appears that a nonzero field mass (typically $\sim 10^{-24}$ gm corresponding to a Compton time of 10^{-23}s) might produce a profound modification of the gravitational dynamics over length and time scales many orders of magnitude in excess of the Planck values. Indeed, early work (Parker & Fulling 1973) with specially contrived quantum states shows that this is a possibility. However, in general, the effect of a mass is negligible compared with other mechanisms for breaking conformal invariance (Birrell & Davies 1980*a*). Basically, this is because terms in $\langle T_{\mu\nu} \rangle$ such as $m^2 R$ only exceed terms like R^2 when $m^2 > R \sim t^{-2}$, i.e. for t greater than the particle Compton time. But at this late epoch, quantum gravitational effects are negligible anyway. There is thus a sort of resonance effect near $t \sim m^{-1}$.

This resonance effect can be seen explicitly by using the perturbation method of §5.6 to treat the simple model of a massive, conformally coupled scalar field in a spatially flat Robertson–Walker spacetime with scale factor

$$a(\eta) = 1 - \tfrac{1}{2}\alpha^2/(\alpha^2 + \eta^2), \quad \alpha \text{ constant} \tag{7.80}$$

The spacetime contracts to a small value of the scale factor, 'bounces' and then expands again, undergoing a period of most rapid expansion (with rate $\sim \alpha^{-1}$) when $\eta \sim \alpha$. Using (5.106) with $C = a^2(h_i = 0, \xi = \frac{1}{6})$ in (5.111) to calculate the Bogolubov coefficient β and subsequently the energy density, one obtains (Birrell & Davies 1980*a*)

$$\rho = \frac{\alpha m^5}{2048}\left\{\tfrac{1}{4}\alpha^3 \frac{\partial^2}{\partial \alpha^2}[\alpha^{-1}K_1(4m\alpha)] + 7\frac{\partial}{\partial \alpha}[\alpha K_1(4m\alpha)] + 49K_1(4m\alpha)\right\}. \quad (7.81)$$

The graph of $\alpha^4\rho$ against $m\alpha$ is shown in fig. 23 and exhibits a distinct resonance about $m\alpha \sim 1$. That is, the final energy density is greatest when the time of most rapid expansion (i.e., most particle production) occurs around the particle Compton time.

An alternative way in which a fundamental length can enter the theory is if one allows for non-gravitational interactions. For example, in the phenomenological theory of the decay of the neutral pi-meson, a coupling

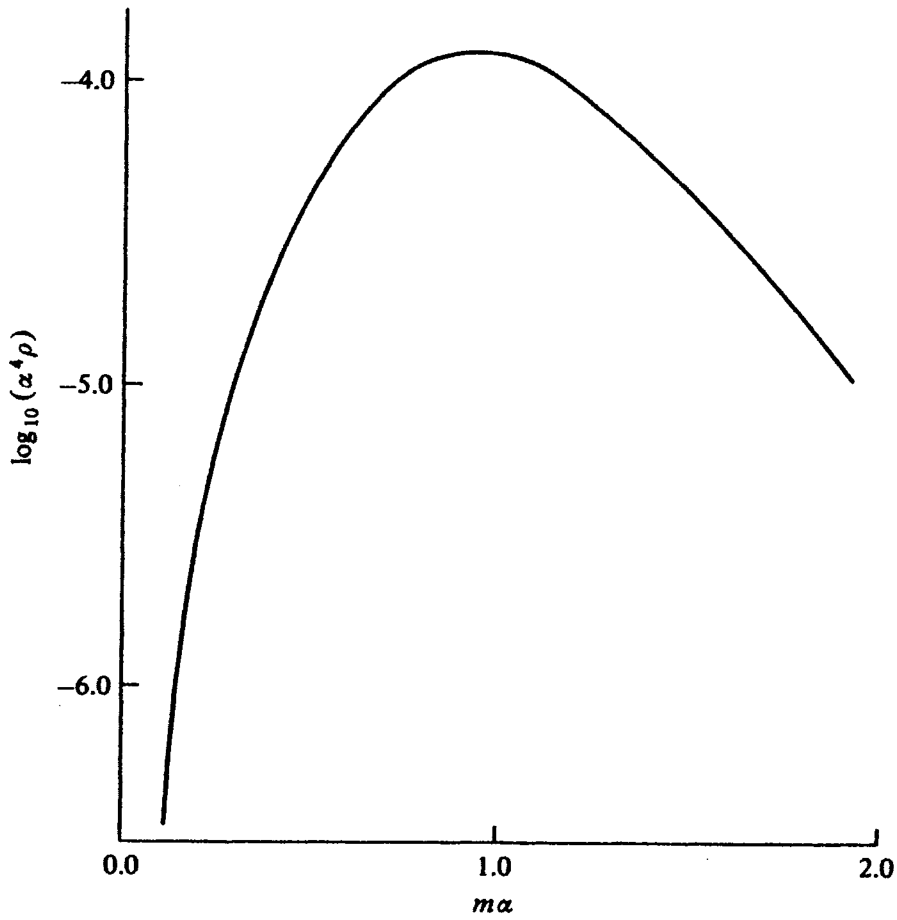


Fig. 23. The energy density of conformally coupled scalar particles created in a cosmological model with time dependence given by (7.80), as a function of particle mass.

constant with units of length is introduced. The back-reaction effects on the gravitational field of such interactions might become large at this characteristic length scale rather than the Planck length (Birrell, Davies & Ford 1980).

To a certain extent the magnitude of quantum effects in the early universe will depend on the actual quantum state, which we do not know. In the absence of any observational guide one is obliged to resort to mathematical criteria (see, for example, Chitre & Hartle 1977, who use analytic continuation techniques to select a particular quantum state).

At the present epoch particle creation and vacuum effects are utterly negligible (Parker 1968, 1969, 1971).

8
Quantum black holes

In January 1974, Hawking (1974) announced his celebrated result that black holes are not, after all, completely black, but emit radiation with a thermal spectrum due to quantum effects. This announcement proved to be a pivotal event in the development of the theory of quantum fields in curved spacetime, and greatly increased the attention given to this subject by other workers. In devoting an entire chapter to the topic of quantum black holes, we are reflecting the widespread interest in Hawking's remarkable discovery.

With the presentation of all the major aspects of free quantum field theory in curved spacetime complete, we here deploy all the various techniques described in the foregoing chapters. The basic result – that the gravitational disturbance produced by a collapsing star induces the creation of an outgoing thermal flux of radiation – is not hard to reproduce. The wavelength of radiation leaving the surface of a star undergoing gravitational collapse to form a black hole is well known to increase exponentially. It therefore seems plausible that the standard incoming complex exponential field modes should, after passing through the interior of the collapsing star and emerging on the remote side, also be exponentially redshifted. It is then a simple matter to demonstrate that the Bogolubov transformation between these exponentially redshifted modes and standard outgoing complex exponential modes is Planckian in structure. This implies that the 'in vacuum' state contains a thermal flux of outgoing particles.

Unfortunately, solutions of the wave equation in the background metric of a collapsing star cannot be written down in terms of simple functions, so we establish the exponential redshift in the outgoing modes using two simplified models. The first is a two-dimensional analogue of gravitational collapse, in which a metric is chosen to correspond to a spherically symmetric ball of matter imploding across its event horizon in an arbitrary way. The second is by ignoring the effects of backscattering in the four-dimensional picture.

The two-dimensional model has the added advantage of permitting a complete solution of the renormalized stress-tensor at all spacetime points, which greatly assists our investigations of the physics of the Hawking effect close to, and inside, the black hole. It also leads to a curious connection with the conformal anomaly. Some space is devoted to issues such as where the particles are created and how the black hole loses mass.

We also present a more elegant treatment of the quantum black hole which omits the collapse phase and replaces it with appropriate boundary conditions on the past horizon of a maximally extended manifold. This enables the intimate relation between the event horizons and the thermal properties of black holes to be more readily discerned, and also leads to a close connection between the black hole and the Rindler system discussed in chapter 4.

Although much of the interest in quantum black holes rests with their astrophysical and thermodynamic implications, we have restricted discussion of these topics to some brief remarks. Extensive references are, however, given in the text.

8.1 Particle creation by a collapsing spherical body

Consider a spherically symmetric ball of matter surrounded by empty space. In the exterior region the unique spherically symmetric solution of Einstein's equation is the Schwarzschild spacetime, described by the metric (3.18). We do not worry here about the interior metric, as it will turn out to be unimportant.

It is known (see, for example, Misner, Thorne & Wheeler 1973, chapter 31) that when sufficiently compact, the ball will implode catastrophically to form a Schwarzschild black hole. The exterior metric remains undisturbed by the collapse, but the modes of any quantum field propagating through the *interior* of the ball will be severely disrupted. Consequently, we expect particle production to take place.

If it is assumed that in the remote past the ball is sufficiently distended that the spacetime is approximately flat, then one may construct the standard Minkowski space quantum vacuum state. After the collapse, the spacetime will have the Schwarzschild form and, in this out region, the vacuum will no longer correspond to the Minkowski space vacuum constructed in the in region. To calculate the particle production, one must compute the Bogolubov transformation between the in and out vacuum states in the usual way.

Our treatment follows that of Hawking (1975) and Parker (1975, 1977);

we restrict our attention to a massless scalar field. As $R=0$ for Schwarzschild spacetime (and as the results do not depend on the detailed metric inside the ball), no distinction need be made between conformal and minimal coupling in the computation of Bogolubov coefficients.

Mode solutions of the wave equation $\Box\phi=0$ in the Schwarzschild spacetime have the form

$$r^{-1}R_{\omega l}(r)Y_{lm}(\theta,\phi)\mathrm{e}^{-\mathrm{i}\omega t} \tag{8.1}$$

where Y_{lm} is a spherical harmonic, and the radial function R satisfies the equation

$$\frac{\mathrm{d}^2R_{\omega l}}{\mathrm{d}r^{*2}}+\{\omega^2-[l(l+1)r^{-2}+2Mr^{-3}][1-2Mr^{-1}]\}R_{\omega l}=0 \tag{8.2}$$

(r^* is defined on page 41). In the asymptotic region $r\to\infty$, (8.2) possesses the solutions $\mathrm{e}^{\pm\mathrm{i}\omega r}$ and (8.1) reduces to

$$r^{-1}Y_{lm}\mathrm{e}^{-\mathrm{i}\omega u} \tag{8.3}$$

and

$$r^{-1}Y_{lm}\mathrm{e}^{-\mathrm{i}\omega v} \tag{8.4}$$

in terms of the null coordinates $u=t-r^*, v=t+r^*$.

Because of the 'potential' term in square brackets in (8.2), the standard incoming waves (8.4) will partially scatter back off the gravitational field to become a superposition of incoming and outgoing waves. Fortunately, it is not necessary to solve (8.2) in detail, so long as attention is restricted to observations made in the asymptotic ($r\to\infty$) region.

Following the general theory of chapter 3 we decompose ϕ into a complete set of positive frequency modes denoted $f_{\omega lm}$:

$$\phi=\sum_{l,m}\int\mathrm{d}\omega(a_{\omega lm}f_{\omega lm}+a^\dagger_{\omega lm}f^*{}_{\omega lm}) \tag{8.5}$$

where $f_{\omega lm}$ are normalized according to the condition (3.29), i.e.,

$$(f_{\omega_1l_1m_1},f_{\omega_2l_2m_2})=\delta(\omega_1-\omega_2)\delta_{l_1l_2}\delta_{m_1m_2}, \tag{8.6}$$

and are chosen to reduce to the incoming spherical modes (8.4) in the remote past. The quantum state is chosen to be the in vacuum, defined by

$$a_{\omega lm}|0\rangle=0,\quad\forall\omega,l,m, \tag{8.7}$$

which corresponds to the absence of incoming (advanced) radiation from $\mathscr{I}^-$.

To find the form of $f_{\omega lm}$ in the remote future, we first note that the incoming waves (8.4) will converge on the centre of the ball, where they will pass on through to become outgoing spherical waves. As the incoming waves approach the surface of the ball they will suffer a blueshift, but when they re-emerge after passing through the ball, there will be a redshift. If the ball is static, these two effects exactly compensate and the waves would reach $\mathscr{I}^+$ with the form (8.3). However, if the ball is collapsing, during the time that the waves spend in transit through the ball, the ball will shrink somewhat, thereby raising its surface gravity. The emerging waves will therefore suffer a redshift that is in excess of the blueshift acquired during the infall. For a ball undergoing complete gravitational collapse to a black hole, the shrinkage timescale is comparable to the light transit time across the ball, and the net redshift becomes appreciable. We shall find below that it increases exponentially with an e-folding time comparable to the transit time.

The situation is depicted schematically in fig. 24, which shows the ball collapsing to form a singularity. Some ingoing null rays are shown passing through the centre of the collapsing ball and out the other side. There exists a latest advanced ray marked γ, that just manages to penetrate the ball and reach $\mathscr{I}^+$ on the far side. This null ray forms the event horizon around the black hole. Later rays pass across the horizon and do not reach $\mathscr{I}^+$, but fall into the singularity instead. It is important to note that the direct interaction between the quantum field and the collapsing matter is being ignored. The presence of the matter in the model is used simply to produce an appropriate gravitational field.

To compute the form of the redshifted modes reaching $\mathscr{I}^+$, we shall suppress the angular variables and work with a two-dimensional model of a collapsing ball that will prove useful later. Outside the ball, the line element is taken to be

$$ds^2 = C(r)\,du\,dv \tag{8.8}$$

where now

$$\left.\begin{aligned} u &= t - r^* + R_0^* \\ v &= t + r^* - R_0^* \end{aligned}\right\} \tag{8.9}$$

$$r^* = \int C^{-1}dr, \tag{8.10}$$

$R_0^* = \text{constant}$. Inside the ball, we take the general line element

$$ds^2 = A(U, V)\,dU\,dV, \tag{8.11}$$

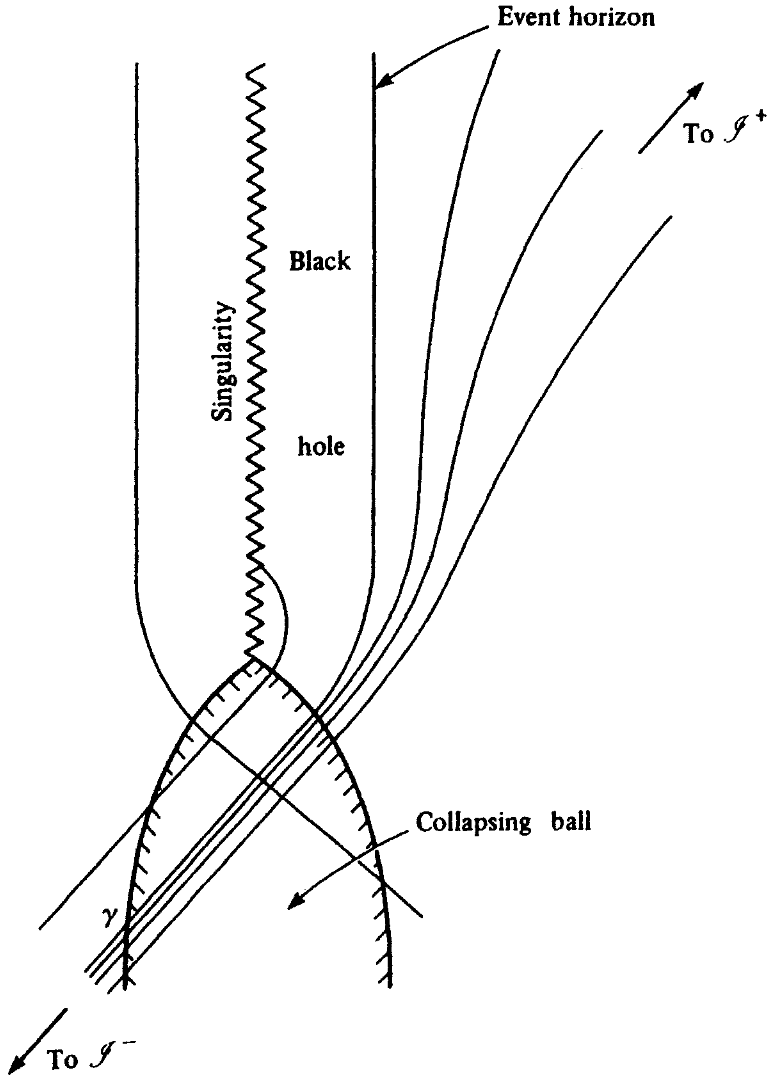


Fig. 24. As the ball collapses to a singularity, null rays converging on the ball's centre from $\mathscr{I}^-$ are distorted. One such ray, labelled, γ, forms the event horizon that marks the boundary between those rays that precede it and reach $\mathscr{I}^+$, and later rays which are trapped by the black hole and drawn into the singularity. Rays which are equispaced along $\mathscr{I}^+$ at late times crowd up along γ on $\mathscr{I}^-$.

where A is an arbitrary smooth, non-singular function and

$$\left.\begin{aligned} U &= \tau - r + R_0 \\ V &= \tau + r - R_0. \end{aligned}\right\} \tag{8.12}$$

The relation between R_0 and R_0^* is the same as that between r and r^* given by (8.10). The exterior metric (8.8) is chosen so that $C \to 1$ and $\partial C/\partial r \to 0$ as $r \to \infty$. There is an event horizon at some value of r for which $C = 0$ (we assume the spacetime to be non-singular outside the horizon). For example $C = 1 - 2Mr^{-1}$ has an event horizon at $r = 2M$, and models in two dimensions the Schwarzschild black hole. Similarly, $C = 1 - 2Mr^{-1} + e^2r^{-2}$

models the Reissner–Nordstrom black hole. However, it is not necessary to specify $C(r)$ explicitly at this stage.

Before $\tau = 0$ the ball is at rest with its surface at $r = R_0$. For $\tau > 0$ we assume that the surface shrinks along the world line $r = R(\tau)$. We shall find that for late times at $\mathscr{I}^+$ (i.e., large u), the precise form of both $A(U, V)$ and $R(\tau)$ are irrelevant. The coordinates (8.9) and (8.12) have been chosen so that at $\tau = t = 0$, the onset of collapse, $u = U = v = V = 0$ at the surface of the ball.

To model the spherical symmetry of the four-dimensional situation, we could reflect the metric (8.8) and (8.11) in the origin of spatial coordinates $r = 0$. Alternatively, we can restrict the treatment to the region $r \geq 0$, and reflect the null rays at $r = 0$. This reproduces the effect of radially incoming rays propagating through the centre of the ball and out again. Such reflection can be achieved by imposing the boundary condition $\phi = 0$ at $r = 0$.

Denote the transformation equations between the interior and exterior coordinates by

$$U = \alpha(u) \tag{8.13}$$

and

$$v = \beta(V) \tag{8.14}$$

where we have ignored any reflection at the surface of the ball. The precise form of α and β will be determined in due course. The centre of radial coordinates is the line

$$V = U - 2R_0. \tag{8.15}$$

We desire solutions of the two-dimensional wave equation

$$\Box\phi = 0$$

that vanish along (8.15) and reduce to standard exponentials on $\mathscr{I}^-$. Conformal symmetry facilitates the solution. Noting from (8.15) that at $r = 0$

$$v = \beta(V) = \beta(U - 2R_0) = \beta[\alpha(u) - 2R_0],$$

one is led to the mode solutions (cf. (4.43))

$$\mathrm{i}(4\pi\omega)^{-\frac{1}{2}}(\mathrm{e}^{-\mathrm{i}\omega v} - \mathrm{e}^{-\mathrm{i}\omega\beta[\alpha(u) - 2R_0]}). \tag{8.16}$$

Thus we see explicitly that the simple 'incoming' (left-moving) wave $\mathrm{e}^{-\mathrm{i}\omega v}$ is converted by the collapsing ball to the complicated 'outgoing' (right-moving) wave $\exp\{-\mathrm{i}\omega\beta[\alpha(u)-2R_0]\}$.

On physical grounds, one expects the complicated phase factor $\beta[\alpha(u)-2R_0]$ to reduce to a steady escalating redshift as the surface of the ball approaches the event horizon. To determine the form of this redshift factor, we have to match the interior and exterior metrics across the collapsing surface $r=R(\tau)$. This yields

$$\alpha'(u)=\frac{\mathrm{d}U}{\mathrm{d}u}=(1-\dot{R})C\{[AC(1-\dot{R}^2)+\dot{R}^2]^{\frac{1}{2}}-\dot{R}\}^{-1} \tag{8.17}$$

$$\beta'(V)=\frac{\mathrm{d}v}{\mathrm{d}V}=C^{-1}(1+\dot{R})^{-1}\{[AC(1-\dot{R}^2)+\dot{R}^2]^{\frac{1}{2}}+\dot{R}\} \tag{8.18}$$

where $\dot{R}$ denotes $\mathrm{d}R/\mathrm{d}\tau$, and U, V and C are here evaluated at $r=R(\tau)$. Note $\dot{R}<0$ for a collapsing surface, so $(\dot{R}^2)^{\frac{1}{2}}=-\dot{R}$.

As the surface of the ball approaches the event horizon, $C=0$, (8.17) and (8.18) simplify to

$$\frac{\mathrm{d}U}{\mathrm{d}u}\sim\frac{(\dot{R}-1)}{2\dot{R}}C(R) \tag{8.19}$$

$$\frac{\mathrm{d}v}{\mathrm{d}V}\sim\frac{A(1-\dot{R})}{2\dot{R}}. \tag{8.20}$$

Close to $C=0$ we may expand $R(\tau)$:

$$R(\tau)=R_{\mathrm{h}}+v(\tau_{\mathrm{h}}-\tau)+\mathrm{O}((\tau_{\mathrm{h}}-\tau)^2) \tag{8.21}$$

where $R=R_{\mathrm{h}}$ at the horizon, $\tau=\tau_{\mathrm{h}}$ when $R(\tau)=R_{\mathrm{h}}$, and $-v=\dot{R}(\tau_{\mathrm{h}})$. Then (8.19) integrates to give, to $\mathrm{O}(\tau_{\mathrm{h}}-\tau)$,

$$\kappa u=-\ln|U+R_{\mathrm{h}}-R_0-\tau_{\mathrm{h}}|+\text{constant} \tag{8.22}$$

where the quantity

$$\kappa=\frac{1}{2}\frac{\partial C}{\partial r}\bigg|_{r=R_{\mathrm{h}}} \tag{8.23}$$

is defined to be the surface gravity of the black hole (Carter 1975).

From (8.22) we see that as $U\to\tau_{\mathrm{h}}+R_0-R_{\mathrm{h}}$, $u\to\infty$. Inverting this relation yields

$$U\propto\mathrm{e}^{-\kappa u}+\text{constant} \tag{8.24}$$

for late times on $\mathscr{I}^+$.

The situation is represented in the Penrose diagram (fig. 25). The null ray

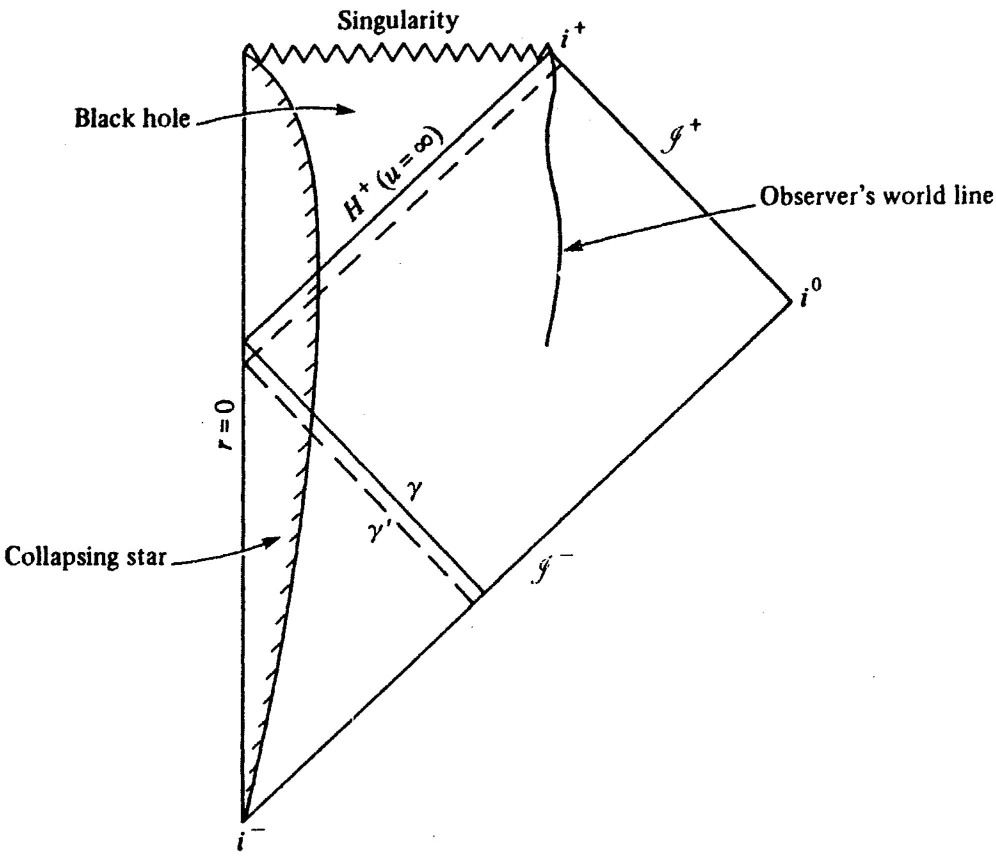


Fig. 25. Penrose diagram of a star that collapses to form a black hole. The exterior region is a fragment of fig. 6, including portions of regions I and III. The null ray γ passes through the centre of the star and emerges to form the event horizon ($u = \infty$), marked H^+, as in fig. 24. A ray γ' (broken line) immediately prior to γ on $\mathscr{I}^-$ reaches $\mathscr{I}^+$ at a finite value of u. Recalling the conformal compression of $\mathscr{I}$, one sees that the whole of the infinite future of the observer outside the hole corresponds to the narrow strip of null rays between γ and γ', representing a brief duration on $\mathscr{I}^-$. Hence the infinity of equispaced late time null rays $u =$ constant, when extended back through the star to $\mathscr{I}^-$, pile up next to γ inside this narrow strip.

$u = \infty$, that marks the horizon of the black hole, corresponds within the collapsing ball to the null ray $U = \tau_h + R_0 - R_h$. The associated 'incoming' ray from $\mathscr{I}^-$, marked γ, is the latest ray that can rebound off $r = 0$ and still reach $\mathscr{I}^+$. Later rays strike the singularity. The null rays $u =$ constant, for large values of u, when traced backwards in time and reflected out to $\mathscr{I}^-$, pile up densely along the latest advanced ray γ (this is reminiscent of fig. 12). Conversely, a very narrow range of values for v and V correspond to the whole late time asymptotic region of $\mathscr{I}^+$. Consequently, to compute the experiences of an asymptotic observer at late time u, we may treat A as approximately constant in (8.20), and integrate to give

$$v \sim \text{constant} - AV(1 + \nu)/2\nu. \qquad (8.25)$$

Substituting (8.24) and (8.25) into (8.16), one obtains the late time asymptotic modes

$$i(4\pi\omega)^{-\frac{1}{2}}(e^{-i\omega v} - e^{i\omega(ce^{-\kappa u}+d)}) \tag{8.26}$$

where c and d are constants. Evidently the 'outgoing' null rays suffer an exponentially increasing redshift, with an e-folding time of κ^{-1}, which is precisely the same as the redshift of the surface luminosity of the collapsing ball (see, for example, Misner, Thorne & Wheeler 1973, §32.3).

The result (8.26) has the same form as that obtained by substituting (4.52) into (4.43), i.e., for a moving mirror receding on a trajectory with asymptotic form (4.51). The Bogolubov transformations in the two cases are thus essentially identical. The reason for this can be traced to the fact that the above analysis in the black hole case is based on geometric optics. The propagation of the field modes through the interior of a collapsing 'star' is similar geometrically to the reflection of the modes from a receding mirror. In the latter case the Doppler shift reproduces the same effect that the gravitational redshift here produces on the field modes. Moreover, the functional form ln cosh, which gives a mirror trajectory with asymptotic behaviour (4.51), is precisely the form of the trajectory in (r^*, t) coordinates of a particle falling across the black hole event horizon (Davies & Fulling 1977*a*).

It follows from the general theory of chapter 3 that because the field modes reduce to standard form at $\mathscr{I}^-$, an inertial particle detector at $\mathscr{I}^-$ will register no particles for the vacuum state based on these modes (i.e., the in vacuum). However, because of the complicated form (8.26), an inertial detector at $\mathscr{I}^+$, in the out region, will register particles for this state. Moreover, because only the u-dependent part of the field modes involves the complicated redshift factor, these particles will be outgoing (i.e., right-moving in the two-dimensional model) corresponding to a flux of particles leaving the vicinity of the black hole and streaming outwards.

The spectrum of this flux may be determined in the usual way by computing the Bogolubov transformation between the modes (8.26) and standard exponential modes in the out region. This was first done by Hawking (1974, 1975). We do not need to carry out this calculation here, as it is mathematically identical to the Bogolubov transformation computed in §4.4. in connection with the receding mirror problem (see remarks above). From (4.61) we find that the expected spectrum is Planckian, corresponding to a thermal spectrum from a black body of temperature

$$T = \kappa/(2\pi k_B). \tag{8.27}$$

For the purpose of computing the Bogolubov transformation in §4.4, it was found to be convenient to take the surface of integration to lie in the in region. This meant inverting the function $p(u)$ to give the function $f(v)$ in (4.58). Physically this corresponds to selecting modes that are standard outgoing waves at $\mathscr{I}^+$, but which become complicated functions of v on $\mathscr{I}^-$. Functionally inverting (8.26) (to within a phase factor), one obtains for these latter modes

$$\mathrm{i}(4\pi\omega)^{-\frac{1}{2}}\{\mathrm{e}^{\mathrm{i}\omega\kappa^{-1}\ln[(v_0-v)/c]}-\mathrm{e}^{-\mathrm{i}\omega u}\},\quad v<v_0, \tag{8.28}$$

where v_0 is a constant corresponding to the final null ray γ. The pileup of advanced rays along γ is manifested in the rapid variation of phase in (8.28) as $v\to v_0$.

The four-dimensional calculation, given in Hawking's paper (Hawking 1975), is essentially the same as for the two-dimensional model described here. The central result – a flux of particles from the vicinity of the hole with a thermal spectrum corresponding to the temperature given by (8.27) – is the same. There are, however, some technical complications in the four-dimensional case, and we shall briefly summarize it here. It is not possible to write the solutions $R_{\omega l}(r)$ to the radial equation (8.2) in terms of known functions, though the properties of the solutions have been extensively investigated (Press & Teukolsky 1972, Starobinsky & Churilov 1973, Teukolsky 1973, Teukolsky & Press 1974, Boulware 1975*a*,*b*, Page 1976*a*, Rowan & Stephenson 1976, Candelas 1980: see also Boulware 1975*b* for the spin 1/2 case).

Equation (8.2) has the form of a one-dimensional wave equation with a potential term (square brackets). Physically this term will give rise to reflected waves which may be envisaged as backscattering of the field modes from the spacetime curvature. In the case of the collapsing ball, some of the ingoing field disturbance will be converted to outgoing disturbance as a result of backscattering rather than from passage through the interior of the ball to the opposite side. As the interesting thermal effects arise only from the latter contribution, we shall temporarily remove the backscattering by artificially deleting the potential term in (8.2). The radial functions then reduce to ordinary exponentials, and the normalized field modes become simply

$$\frac{Y_{lm}(\theta,\phi)}{(8\pi^2\omega)^{\frac{1}{2}}r}\times\begin{cases}\mathrm{e}^{-\mathrm{i}\omega u} & (8.29)\\ \mathrm{e}^{-\mathrm{i}\omega v} & (8.30)\end{cases}$$

which reduce to the usual flat space form at large r, where

$$u\equiv t-r^*\to t-r,\quad v\equiv t+r^*\to t+r.$$

We are interested in that particular linear combination of modes (8.30) that corresponds to standard modes at $\mathscr{I}^+$. Tracing these modes backwards in time (Hawking 1975) through the collapsing ball and out along the advanced null ray to $\mathscr{I}^-$, the mode which has the form (8.29) on $\mathscr{I}^+$ looks like

$$\begin{array}{ll} \dfrac{Y_{lm}(\theta,\phi)}{(8\pi^2\omega)^{\frac{1}{2}}r}\exp\{4Mi\omega\ln[(v_0-v)/c]\}, & v<v_0 \\ 0 & v>v_0 \end{array} \tag{8.31}$$

(c constant) close to the latest advanced ray γ in fig. 24. The similarity between (8.31) and (8.28) is obvious.

The ordinary in vacuum is defined with respect to modes (8.30). The Bogolubov coefficients relating (8.30) and (8.31) are given by

$$\left.\begin{array}{l}\alpha_{\omega\omega'}\\ \beta_{\omega\omega'}\end{array}\right\} = (1/2\pi)\int_{-\infty}^{v_0} dv(\omega'/\omega)^{\frac{1}{2}}e^{\pm i\omega' v}\exp\{4Mi\omega\ln[(v_0-v)/c]\}, \tag{8.32}$$

which is readily evaluated in terms of Γ-functions (Hawking 1975). The answer is essentially the same as (4.60). As in that case, the factor $(\omega')^{-\frac{1}{2}}$ implies that

$$\int|\beta_{\omega\omega'}|^2 d\omega'$$

diverges logarithmically. There are an infinite number of particles in each mode at $\mathscr{I}^+$. The divergence here is connected with the normalization of the continuous modes (8.29)–(8.31). The collapsing ball produces a steady flux of radiation to $\mathscr{I}^+$, so the total flux for all time is infinite. Of greater interest is the number of particles emitted per unit time. This may be found by localizing the modes in some way; e.g., by confining the system to a box with periodic boundary conditions so that the modes become discrete (Page 1976*a*).

In this case, one may use the Wronskian condition (3.39) to write

$$\sum_{\omega}(|\alpha_{\omega\omega'}|^2 - |\beta_{\omega\omega'}|^2) = 1. \tag{8.33}$$

From the evaluation of (8.32) (see (4.60)) one notes that

$$|\alpha_{\omega\omega'}|^2 = e^{8\pi M\omega}|\beta_{\omega\omega'}|^2 \tag{8.34}$$

(the analyticity argument leading to the factor $e^{8\pi M\omega}$ is similar to that discussed on page 116).

To compute the particle flux going to $\mathscr{I}^+$ at late times we note that the density of states inside a sphere of radius R centred on the collapsing ball is

$R\mathrm{d}\omega/2\pi$ (for fixed l, m). The particle number per mode is, from (8.33) and (8.34)

$$N_{\omega lm} = \sum_{\omega'} |\beta_{\omega\omega'}|^2 = 1/(\mathrm{e}^{8\pi M\omega} - 1) \tag{8.35}$$

so the number of particles per unit time in the frequency range ω to $\omega + \mathrm{d}\omega$, passing out through the surface of the sphere is

$$(\mathrm{d}\omega/2\pi)(\mathrm{e}^{8\pi M\omega} - 1)^{-1}, \tag{8.36}$$

where we have used the fact that a particle takes a time R to reach the surface of the sphere. This is a Planck (black body) spectrum with temperature given by (8.27), exactly as in the two-dimensional case. Alternatively, one can construct localized wave packets to use as a basis in place of (8.29)–(8.31) (Hawking 1975).

At this stage we must take into account the neglect of backscattering. Equation (8.33) can be interpreted as the conservation of probability. The effect of backscattering is to deplete the outgoing flux by a factor $1 - \Gamma_\omega$ representing reflection back down the black hole. Thus the right-hand side of (8.33) must be replaced by Γ_ω, which introduces a factor Γ_ω into (8.36). Because this backscattering is a function of ω, the spectrum is not precisely Planckian. Nevertheless, it may still be regarded as 'thermal' in the following sense. If the black hole is immersed in a heat bath at the temperature (8.27), then the same fraction $1 - \Gamma_\omega$ of incoming radiation will be backscattered by the hole as was removed from the outgoing flux by backscattering. Thus only a fraction Γ_ω of the incoming radiation in the mode ω is absorbed by the hole. It follows that the ratio of emission to absorption per mode is independent of Γ_ω and identical to that of a black body replacing the black hole. The hole therefore remains in thermal equilibrium with the surrounding heat bath in spite of the spectral distortion.

The total luminosity of the hole is given by integrating (8.36) over all modes

$$L = (1/2\pi) \sum_{l=0}^{\infty} (2l+1) \int_0^\infty \mathrm{d}\omega\, \omega \Gamma_{\omega l}/(\mathrm{e}^{8\pi M\omega} - 1), \tag{8.37}$$

where we have introduced a Γ factor, as well as l-dependence.

If we were dealing with fermions rather than bosons, because of their anticommuting nature, the normalization condition (8.33) would require a + sign in front of $|\beta_{\omega\omega'}|^2$ (see, for example, Parker 1971, DeWitt (1975)). One would therefore obtain a factor $\mathrm{e}^{8\pi M\omega} + 1$ in the expression for the spectrum, which is the appropriate Planck factor for Fermi statistics.

The computation of L for a realistic black hole model involves consideration of real quantum fields. Although the discussion so far has been restricted to a massless scalar field, the essential argument applies to any field. In particular, quanta of the electromagnetic, neutrino and linearized graviton fields will all be radiated thermally with the temperature (8.27). The chief difference lies in their respective Γ factors, which are rather sensitive to the spin of the field. Detailed results require a study of the radial equations such as (8.2) and numerical computations.

Page (1976*a*) (see also Page 1976*b*, 1977) has estimated

$$L = (3.4 \times 10^{46})(M/1\,\mathrm{gm})^{-2}\,\mathrm{erg\,s}^{-1} \tag{8.38}$$

for a Schwarzchild black hole of mass $\gg 10^{17}$ gm, which consists of 81% neutrinos (four types), 17% photons and 2% gravitons. (The suppression of higher spins is connected with a greater angular momentum barrier.)

The temperature (8.27) may be written

$$T = (1.2 \times 10^{26}\ \mathrm{K})(1\ \mathrm{gm}/M), \tag{8.39}$$

which for a solar mass object gives 6×10^{-8} K. For such an object, only massless quantum emission is relevant. However, for $M \lesssim 10^{17}$ gm, $T \gtrsim 10^{9}$ K, and the creation of thermal electron–positron pairs becomes possible. At lower masses, other species of elementary particles will be emitted. The details of this high temperature regime will be complicated by the presence of interactions between the particles. Note that a hole of mass 10^{15} gm has a radius of about one fermi, i.e., within the range of the strong interaction. However, although these complications preclude the analysis of the rate of energy emission, one may still appeal to the fundamentally thermodynamic nature of quantum particle production.

So far we have restricted attention to the spherically symmetric case. If rotation is permitted, however, the black hole belongs to the axisymmetric Kerr family, with a metric parametrized by the angular speed of the event horizon, Ω.

The solution of the wave equation in the Kerr background is more complicated than for the Schwarzschild case (see, for example, DeWitt 1975 for a detailed discussion). The spherical harmonics Y_{lm} of (8.1) are replaced by axisymmetric spheroidal harmonics, while the potential term in the radial equation (8.2) acquires some complicated new terms. However, at large r, the effect is merely to replace ω by $\omega - m\Omega$, where m is the azimuthal quantum number of the spheroidal harmonics (playing the role of m in Y_{lm}). Thus, as far as the analysis of radiation at $\mathscr{I}^+$ is concerned, one merely replaces ω by $\omega - m\Omega$ in the Planck factor in (8.36) or (8.37):

$$1/\{\exp[2\pi\kappa^{-1}(\omega - m\Omega)] \pm 1\}. \tag{8.40}$$

The rotation therefore enters into the thermal spectrum in much the same way as a chemical potential. The effect of this alteration on the luminosity of the hole has also been computed by Page (1976*b*), who finds that the rotation of the hole greatly enhances the emission of higher-spin particles.

Because the emission probability depends on the azimuthal quantum number m, it will be asymmetric around the hole. The factor (8.40) is larger for positive m than negative, thus favouring the emission of quanta with angular momenta oriented towards that of the hole rather than away. This implies that the emission will steadily deplete the hole's angular momentum, causing its rotation rate to slow.

In the boson case ($-$ sign in (8.40)), when $\omega < m\Omega$, (8.40) is negative. Moreover, even in the limit $M \to \infty$ $(T \to 0)$, when the Hawking thermal emission due to the collapsing body dies away, (8.40) remains finite, and equal to -1, for $\omega < m\Omega$ (it vanishes for $\omega > m\Omega$, and hence for all ω in the Schwarzschild case, $\Omega = 0$). The meaning of this negative flux is the following: when a low frequency, classical wave (with positive m) impinges on a rotating black hole, the effect of rotation is to amplify the wave, causing more energy to emerge than went in originally – a phenomenon known as super-radiance (Misner 1972, Press & Teukolsky 1972, Zel'dovich 1971, 1972). In quantum language the hole induces stimulated emission. Thus, the absorption probability of the hole for these modes is negative. Super-radiance is unconnected with the Hawking mechanism (which depends on mode propagation through the collapsing body) and even occurs around ordinary rotating stars (Ashtekar & Magnon 1975*b*). Quantum super-radiance was calculated before the Hawking effect by Starobinsky (1973) and Unruh (1974) (see also Ford 1975).

Also of interest is the effect of electric charge on the hole. First consider the case of an electrically neutral field propagating in the background of a Reissner–Nordstrom (non-rotating) hole with charge e. This has the line element

$$\begin{aligned} ds^2 = {} & [1 - (2M/r) + (e^2/r^2)]\, dt^2 - [1 - (2M/r) + (e^2/r^2)]^{-1}\, dr^2 \\ & - r^2(d\theta^2 + \sin^2\theta\, d\phi^2), \end{aligned} \tag{8.41}$$

which possesses an event horizon at

$$r = r_+ = M + (M^2 - e^2)^{1/2}. \tag{8.42}$$

Using (8.42) for R_h in (8.23) one obtains the surface gravity κ for the metric

(8.41). Equation (8.27) then yields the temperature of the charged black hole:

$$T = (1/8\pi k_B M)(1 - 16\pi^2 e^4/\mathscr{A}^2) \tag{8.43}$$

where $\mathscr{A} = 4\pi r_+^2$ is the area of the event horizon.

It is clear from (8.43) that the presence of a charge depresses the temperature of the hole. In the extreme case $e^2 = M^2$ (maximally charged hole), $T = 0$. Inspection of (8.41) and (8.42) shows that for $e^2 > M^2$ no event horizon exists: there is a naked singularity at $r = 0$. Thus, the cosmic censorship hypothesis (Penrose 1969) that naked singularities cannot form from gravitational collapse can be upheld by requiring $T > 0$, which may be interpreted as the third law of thermodynamics applied to black holes (Carter 1975). However, closer inspection (Davies 1978, Farrugia & Hájíček 1980) leaves open the question as to whether $T = 0$ could in principle be achieved.

If the black hole is sufficiently small, it will be hot enough to produce electron–positron pairs. The creation of charged particles in the field of a Reissner–Nordstrom black hole is complicated by the presence of the electric field of the hole itself. A detailed treatment has been given by Gibbons (1975).

It has long been known that the presence of a background electric field opens up the possibility of the spontaneous creation of charged particle pairs even in the absence of a background gravitational field. This may be seen even in Dirac's old semiclassical model of pair creation. The positive and negative energy states of a free Dirac field are separated by a gap $2m$. For particles of charge q, the presence of an external electric field of strength E will lead to a potential energy gradient qE. If a virtual particle–antiparticle pair is created, the two particles will accelerate apart under the action of the electric field. If their separation Δx becomes sufficiently large before their expected annihilation, then the energy gained from the field will be enough to promote the pair to real particles. This will happen if $qE\Delta x \simeq 2m$. This may equivalently be regarded as particles tunnelling through the energy gap $2m$, to appear as ordinary particles accompanied by antiparticles. This phenomenon became known as the Klein paradox (Klein 1929) and led to controversy concerning the stability of the vacuum.

When the gravitational field of a black hole is also present, the situation is more complicated, especially if it is rotating (Gibbons 1975, Deruelle & Ruffini 1976, Damour 1977). There will be a tendency for the hole to discharge itself rather quickly by preferentially emitting charged particles of the same sign as the charge of the hole, rather than the oppositely charged antiparticles. There will also be a 'charge super-radiance' phenomenon, similar to the effect of the hole's rotation, for particles with $\omega < m\Omega + q\Phi$, where Φ is the electric potential at the horizon (Hawking 1975). Thus, electric charge also contributes to the 'chemical potential' of the hole.

Because of the presence of the electric field, which encourages the production of particle pairs, electron–position emission can occur even for very massive black holes with very low temperatures. In this regime, the electric (Klein paradox) effect dominates. Only for $M \gtrsim 10^5 M_\odot$ is the charged particle creation suppressed. At the other extreme, for $M \lesssim 10^{15}$ gm, the gravitational (Hawking) process dominates over the electrical effects.

8.2 Physical aspects of black hole emission

At first sight, black hole radiance seems paradoxical, for nothing can apparently escape from within the event horizon. However, inspection of (8.36) shows that the average wavelength of the emitted quanta is $\sim M$, i.e., comparable with the size of the hole. As it is not possible to localize a quantum to within one wavelength, it is therefore meaningless to trace the origin of the particles to any particular region near the horizon. The particle concept, which is basically global, is only useful near $\mathscr{I}^+$. In the vicinity of the hole, the spacetime curvature is comparable with the radiation wavelength in the energy range of interest, and the concept of locally-defined particles breaks down.

Heuristically, one can envisage the emergent quanta as 'tunnelling' out through the event horizon (Hawking 1977*b*). Alternatively, the continuous, spontaneous creation of virtual particle–antiparticle pairs around the black hole can be used to explain the Hawking radiation. Virtual particle pairs created with wavelength λ separate temporarily to a distance $\simeq \lambda$. For $\lambda \simeq M$, the size of the hole, strong tidal forces operate to prevent re-annihilation. One particle escapes to infinity with positive energy to contribute to the Hawking flux, while its corresponding antiparticle enters the black hole trapped by the deep gravitational potential well on a timelike path of negative energy relative to infinity (see, for example, Gibbons 1979). Thus the hole radiates quanta with wavelength $\simeq M$.

In spite of the ill-defined nature of particles near the horizon, it is clear that the thermal emission will carry away energy to $\mathscr{I}^+$, and the question arises as to the source of this energy. It can only come from the gravitational field itself, which must lose mass as a consequence. To study this steady depletion of mass–energy, we can investigate the stress-tensor expectation value, $\langle T_{\mu\nu}(x)\rangle$, in the vicinity of the hole. Unlike the particle concept, the stress-tensor is a *local* object, and hence may be used to probe the physics close to, and within, the black hole itself.

In a two-dimensional model, $\langle T_{\mu\nu}\rangle$ may be evaluated explicitly for a conformally invariant quantum field (see §6.3). Consider a static ball described by the general static metric (8.8). The wave equation may be solved in terms of the modes $e^{-i\omega u}$ and $e^{-i\omega v}$ (with u and v given by (8.9)), which are of positive frequency with respect to the global Killing vector ∂_t. The vacuum state constructed with respect to these modes is stable; no particles are produced and the system is time-symmetric.

There will, however, be a nonzero vacuum 'polarization' stress due to spacetime curvature, given by substituting $C(r)$ into (6.136):

$$\langle 0|T_{uu}|0\rangle = \langle 0|T_{vv}|0\rangle = -F_u(C) = (1/192\pi)[2CC'' - C'^2] \quad (8.44)$$

$$\langle 0|T_{uv}|0\rangle = (1/96\pi)CC'', \quad (8.45)$$

where the functional F is defined by

$$F_x(y) = \frac{1}{12\pi} y^{\frac{1}{2}} \frac{\partial^2}{\partial x^2}(y^{-\frac{1}{2}}), \quad (8.46)$$

and a prime denotes differentiation with respect to r.

To take a specific case, the two-dimensional analogue of the Reissner–Nordstrom spacetime (8.41) is described by $C(r) = (1 - 2Mr^{-1} + e^2 r^{-2})$, from which (8.44) gives

$$\langle 0|T_{uu}|0\rangle = \langle 0|T_{vv}|0\rangle = \frac{1}{24\pi}\left[-\frac{M}{r^3} + \frac{3}{2}\frac{M^2}{r^4} + \frac{3e^2}{2r^4} - \frac{3Me^2}{r^5} + \frac{e^4}{r^6}\right] \quad (8.47)$$

which describes a static cloud of energy, singular at $r = 0$, and *negative* outside the event horizon $r = r_+$.

Suppose now that the 'star' undergoes gravitational collapse from radius R_0 in the fashion described on page 254. Although the metric (8.8) still correctly describes the geometry in the exterior of the 'star', the coordinates r^* and t (or u and v) are no longer appropriate for the (simple, exponential) solution of the wave equation, because 'outgoing' modes $e^{-i\omega u}$ degenerate to the complicated function (8.28) on $\mathscr{I}^-$. Instead, we wish to choose coordinates such that the 'incoming' modes, and hence the vacuum state (denoted by $|\hat{0}\rangle$), are of the standard Minkowski form on $\mathscr{I}^-$ (i.e., the usual 'in' vacuum state). As explained in the previous section, this means using, in place of u and v as defined above, the coordinates

$$\hat{u} = \beta[\alpha(u) - 2R_0] \tag{8.48}$$

$$\hat{v} = v, \tag{8.49}$$

where α and β are defined by (8.13) and (8.14). The metric (8.8) in these coordinates becomes

$$ds^2 = C(\hat{u}, \hat{v})d\hat{u}d\hat{v} \tag{8.50}$$

with

$$\hat{C}(\hat{u}, \hat{v}) = C(r)\frac{du}{d\hat{u}}\frac{dv}{d\hat{v}}. \tag{8.51}$$

Using (8.17) and (8.18) to evaluate $du/d\hat{u}$ and (8.49) to give $dv/d\hat{v}$, and then substituting into (6.136) one obtains (Davies 1976; see also Davies, Fulling & Unruh 1976, Hiscock 1977*a*, *b*, 1979, 1980)

$$\langle \hat{0}|T_{uu}|\hat{0}\rangle = (8.44) + (\alpha')^2 F_U(\beta') + F_u(\alpha'), \tag{8.52}$$

where $\alpha = \alpha(u)$, $\beta = \beta(U - 2R_0)$ and primes denote differentiation with respect to the argument of a function. The expressions for $\langle \hat{0}|T_{vv}|\hat{0}\rangle$ and $\langle \hat{0}|T_{uv}|\hat{0}\rangle$ are the same as (8.44) and (8.45) respectively.

Equation (8.52) applies to the region outside the collapsing body. In the interior region, described by the metric (8.11), we have

$$\langle \hat{0}|T_{UU}|\hat{0}\rangle = F_U(\beta') - F_U(A) \tag{8.53}$$

$$\langle \hat{0}|T_{VV}|\hat{0}\rangle = F_V(\beta') - F_V(A), \tag{8.54}$$

where $\beta = \beta(U - 2R_0)$ and $\beta = \beta(V)$, respectively.

The general formula (6.136) yields $\langle \hat{0}|T_{\mu\nu}|\hat{0}\rangle$ in the $\hat{u}, \hat{v}$ coordinates (8.48) and (8.49). In arriving at (8.52)–(8.54), we have converted to u, v, or U, V coordinates by using the ordinary tensor transformation relations. The results show that the effect of collapse is to augment the static vacuum energy (8.44) with an outgoing (retarded) flux of radiation that is constant along the retarded null rays u, or U (there is no backscattering here). In the interior of the body, both ingoing and outgoing radiation fluxes occur that depend in a complicated way upon the interior metric A and the collapse trajectory $R(\tau)$. This represents particles created inside the material of the collapsing 'star', and at its surface. For example, $F_V(\beta')$ in (8.54) describes radiation created at the surface that propagates towards $r = 0$, while $F_U(\beta')$ in (8.53) simply describes the same radiation propagating outward again after passage through the centre of the body. When this outgoing flux emerges (unchanged) from the body, it is described by the term proportional to $(\alpha')^2$ in (8.52). In addition to this 'rebounded' outgoing flux, there will be the contribution described by the final term of (8.52), which can also be traced back to the surface of the collapsing body, but which is *outgoing* from the outset.

If the function $\alpha'(u)$ is expressed in terms of τ as a function $\gamma(\tau)$, then we have from (8.17) the following limits as the surface of the body approaches the horizon ($R \to R_h$, $C \to 0$, $u \to \infty$)

$$\gamma \to 0 \text{ as } e^{-\kappa u}$$

$$\dot{\gamma} \to -(1-\dot{R})\kappa$$

$$\ddot{\gamma}\gamma \to 0$$

(here a dot denotes ∂_τ), from which it follows that the middle term on the right of (8.52) vanishes exponentially fast as measured by a distant observer using u, v coordinates. This term therefore represents a sort of prompt radiation emitted as a transient pulse, and which decays exponentially in an identical fashion to the surface luminosity of the collapsing 'star'.

The final term of (8.52) may be written

$$\frac{1}{24\pi}\left[\frac{-\gamma\ddot{\gamma}}{(1-\dot{R})^2} - \frac{\dot{\gamma}\gamma\ddot{R}}{(1-\dot{R})^3} + \frac{\frac{1}{2}\dot{\gamma}^2}{(1-\dot{R})^2}\right] \to \frac{\kappa^2}{48\pi}, \tag{8.55}$$

which is precisely the flux expected (in two spacetime dimensions) from a thermal radiator with temperature $T = \kappa/2\pi k_B$ (see (4.27) for $L \to \infty$), which is the Hawking temperature (8.27). This term therefore represents the Hawking radiation. Note that it is independent of the details of the collapse, depending only on the single parameter κ, the surface gravity of the final black hole. The collapse-dependent radiation is all contained in the exponentially decaying middle term of (8.52).

For late times, in the asymptotic region, $r \to \infty$, the vacuum polarization term (8.47) vanishes, and the stress-tensor reduces to

$$\langle \hat{0}| T_{uu} |\hat{0}\rangle = \kappa^2/48\pi, \quad \langle \hat{0}| T_{vv} |\hat{0}\rangle = \langle \hat{0}| T_{uv} |\hat{0}\rangle = 0. \tag{8.56}$$

The manifest time asymmetry present in (8.56) (retarded radiation only) reflects the irreversible nature of collapse across an event horizon.

Although the various terms of (8.52)–(8.54) have readily-identifiable origins, it is important not to attach too much physical significance to them individually. For instance, it would be wrong to conclude that the Hawking flux arises from the surface of the collapsing body. It must be remembered that an experimenter could not distinguish between the various terms operationally; only the total stress-tensor components can be measured. The decomposition given here is peculiar to the special u, v (or U, V) coordinate system. However, this system is singular at the event horizon.

To determine what an observer would actually measure, the world line of the apparatus must be specified. If the observer's instantaneous velocity vector is u^μ, then his measuring apparatus records the energy density $\langle T_{\mu\nu}\rangle u^\mu u^\nu$ and energy flux $\langle T_{\mu\nu}\rangle u^\mu n^\nu$, where $u^\mu n_\mu = 0$. For example, an observer constrained to fixed r would have 2-velocity $C^{-\frac{1}{2}}(1,0)$ (in(t, r^*) coordinates) and measure the energy density $C^{-1}\langle\hat{0}|T_{tt}|\hat{0}\rangle$, where T_{tt} is given using (4.17). This diverges at the horizon, where $C = 0$, which is a reflection of the fact that to remain at fixed r, the observer's world line deviates ever further from a free-falling world line as the horizon is approached.

A more realistic procedure would be to consider a freely-falling observer. In the Schwarzschild case, the Kruskal coordinates (which are non-singular at the horizon – see §3.1) would then be more appropriate. Using the transformation equations (3.19) and the conformal factor

$$\bar{C} = 2Mr^{-1}e^{-r/2M}$$

from (3.20), one finds that an observer moving along a line of constant Kruskal position with two-velocity $\bar{C}^{-\frac{1}{2}}(1,0)$ (in coordinates $(\bar{t}, \bar{r})$; $\bar{t} = \frac{1}{2}(\bar{v}+\bar{u})$, $\bar{r} = \frac{1}{2}(\bar{v}-\bar{u})$), would measure the energy density

$$\bar{C}^{-1}\langle\hat{0}|T_{tt}|\hat{0}\rangle = \langle\hat{0}|T_{\bar{u}}{}^{\bar{v}} + T_{\bar{v}}{}^{\bar{u}} + 2T_{\bar{u}}{}^{\bar{u}}|\hat{0}\rangle, \tag{8.57}$$

where

$$\langle\hat{0}|T_{\bar{u}}{}^{\bar{v}}|\hat{0}\rangle = \frac{(24\pi)^{-1}(\bar{v})^2 e^{-r/2M}}{8Mr}\left(1 + \frac{4M}{r} + \frac{12M^2}{r^2}\right), \tag{8.58}$$

$$\langle\hat{0}|T_{\bar{v}}{}^{\bar{u}}|\hat{0}\rangle = \frac{(24\pi)^{-1}e^{r/2M}}{\bar{v}^2}\left(\frac{1}{r^2} - \frac{3M}{2r^3}\right), \tag{8.59}$$

$$\langle\hat{0}|T_{\bar{u}}{}^{\bar{u}}|\hat{0}\rangle = -\frac{R}{48\pi} = -\frac{M}{12\pi r^3}. \tag{8.60}$$

It is clear that (8.57) is finite as $\bar{u} \to \infty$, because $\bar{v}$ remains finite on the future horizon. Hence we conclude that an observer who crosses the event horizon along a constant Kruskal position line measures a finite energy density.

These considerations resolve an apparent paradox concerning the Hawking effect. The proper time for a freely-falling observer to reach the event horizon from a finite distance is finite, yet the free-fall time as measured at infinity (in u, v coordinates) is infinite. Ignoring back-reaction, the black hole will emit an infinite amount of radiation during the time that the falling observer is seen, from a distance, to reach the event horizon. Hence it would appear that, in the falling frame, the observer should encounter an infinite amount of radiation in a finite time, and so be destroyed. On the other hand, the event horizon is a global construct, and has no local significance (Hawking 1973), so it is absurd to conclude that it acts as a physical barrier to the falling observer.

The paradox is resolved when a careful distinction is made between particle number and energy density. When the observer approaches the horizon, the notion of a well-defined particle number loses its meaning at the wavelengths of interest in the Hawking radiation (see page 264); the observer is 'inside' the particles. We need not, therefore, worry about the observer encountering an infinite quantity of particles. On the other hand, energy does have a local significance. In this case, however, although the Hawking flux does diverge as the horizon is approached, so does the static vacuum polarization, and the latter is *negative*. The falling observer cannot distinguish operationally between the energy flux due to the oncoming Hawking radiation and that due to the fact that he is sweeping through the cloud of vacuum polarization. The net effect is to cancel the divergence on the event horizon, and yield a finite result, such as that given above by (8.57)–(8.60). (The precise value depends on the infall trajectory.)

The present analysis also solves the mystery of how the black hole can lose mass without matter crossing from the interior of the black hole into the outside universe. Inspection of $\langle \hat{0}|T_{vv}|\hat{0}\rangle$ at the event horizon shows that it is given by

$$-\frac{1}{192\pi}\left(\frac{\partial C}{\partial r}\right)^2\bigg|_{r=R_h} = -\frac{\kappa^2}{48\pi},$$

which is always negative, and equal to minus the Hawking flux at infinity. This is necessarily true because covariant conservation was built into the construction of $\langle T_{\mu\nu}\rangle$. As $\langle T_{vv}\rangle$ represents a null flux crossing the event horizon, one can see that the steady loss of mass–energy by the Hawking flux is balanced by an equal negative energy flux crossing into the black hole from outside. The hole therefore loses mass, not by emitting quanta, but by absorbing negative energy. The idea of a flux of negative energy has already been encountered in connection with moving mirrors in §7.1.

In the above analysis, we have assumed that the quantum state is the conventional vacuum state in the in region. It might be wondered to what extent the presence of quanta initially will change the Hawking effect (Wald

1976). From (2.44) (which is also valid in curved spacetime), one finds the contribution to $\langle T_{\mu\nu} \rangle$ in the out region due to the initial presence of n quanta in the mode σ. In the case of bosons, as a result of stimulated emission, there will be additional energy present in the out region (Bekenstein & Meisels 1977). The form of the outgoing modes in the out region is given by the second term of (8.26), from which one readily sees that the final term of (2.44) decays exponentially in u with e-folding time $(2\kappa)^{-1}$, i.e., the collapse timescale. Thus, the effect of initial quanta fades out exponentially on the same timescale as any surface luminosity, and the black hole soon settles down to thermal equilibrium, having 'forgotten' the details of the initial state. We may therefore conclude that the Hawking effect is extremely general, and independent of any physically reasonable initial quantum state.

These general features of black hole radiance persist in the full four-dimensional treatment.

The fact that the black hole loses mass raises the question of the back-reaction on the structure of the hole itself. Assuming that one may continue to treat the background metric as classical, then the stream of negative energy across the event horizon will cause the area of the horizon to shrink. This is shown schematically in fig. 26 for a Schwarzschild hole. As the area decreases, so does the mass, implying that the temperature (8.39) and luminosity (8.38) rise. Schwarzschild black holes therefore have a negative specific heat – they radiate and get hotter, behaviour which is typical of self-gravitating systems (Lynden-Bell & Wood 1967).

Having an expression for $\langle T_{\mu\nu} \rangle$ at the event horizon enables the features of the back-reaction to be examined in more detail, by attempting to integrate the Einstein equations. This provides a description of the evolution of the event horizon. In particular, one expects that, in the adiabatic approximation that the black hole's shrinkage is ignored in computing $\langle T_{\mu\nu} \rangle$, the horizon should shrink at a steady rate appropriate to the loss of mass in the Hawking radiation. Unfortunately the horizon evolution equation is nonlinear, and has not been solved. If some higher order terms are neglected, however, the rate of shrinkage is indeed what would be naively expected (Candelas 1980). Nevertheless, it has yet to be determined whether the effect of the higher order terms is in fact negligible, and whether the shape of the shrinking horizon in fig. 26 is correct (Tipler 1980, Hájíček & Israel 1980).

As a hole gets hotter, so it will radiate subatomic particles of greater mass. Once above about 10^{10} K, electrons and positrons will emerge, and any residual charge on the hole will rapidly disappear. Moreover, super-

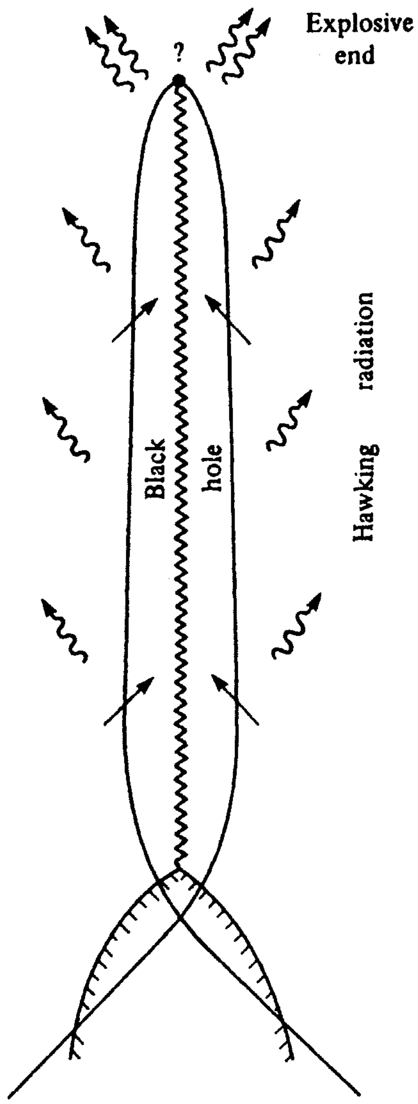


Fig. 26. Black hole evaporation. The lower portion of the diagram corresponds to fig. 24. Gradually the flux of negative energy crossing the horizon into the hole (straight arrows) causes the horizon area to shrink. The process escalates until the horizon collapses rapidly onto the singularity amid an explosive radiation of quanta.

radiance effects will tend to deplete the angular momentum, so the black hole has a tendency to slowly approach the Schwarzschild form.

The continuation of the Hawking process seems to imply that the hole will evaporate away ever faster. Its ultimate fate cannot be decided within the context of the present theory, for when

$$\frac{1}{M}\frac{\mathrm{d}M}{\mathrm{d}t} \simeq k_{\mathrm{B}}T \simeq M^{-1} \tag{8.61}$$

the hole is shrinking at a rate comparable with the frequency of the radiation. Thermal equilibrium no longer applies, nor is the notion of a fixed background spacetime a good approximation. Condition (8.61) is reached when

$$\frac{\mathrm{d}M}{\mathrm{d}t} \sim (kT)^4 \mathscr{A} \sim 1$$

or

$$M^{-4} \times M^2 \sim 1,$$

i.e., when $M \sim 1$ (the Planck mass $\sim 10^{-5}$ gm), where we have used Stefan's radiation law and the fact that the area $\mathscr{A}$ of the hole is $16\pi M^2$ in the Schwarzschild case. At this stage, the hole is of Planck dimensions (10^{-33} cm) and higher order quantum gravity effects will undoubtedly be important.

It has been conjectured that the end result of the Hawking evaporation process is explosive disappearance (Hawking 1977*a*) or a naked singularity (DeWitt 1975, Penrose 1979), or perhaps a Planck mass object. Fig. 27 shows a possible Penrose diagram for these situations, where the dot

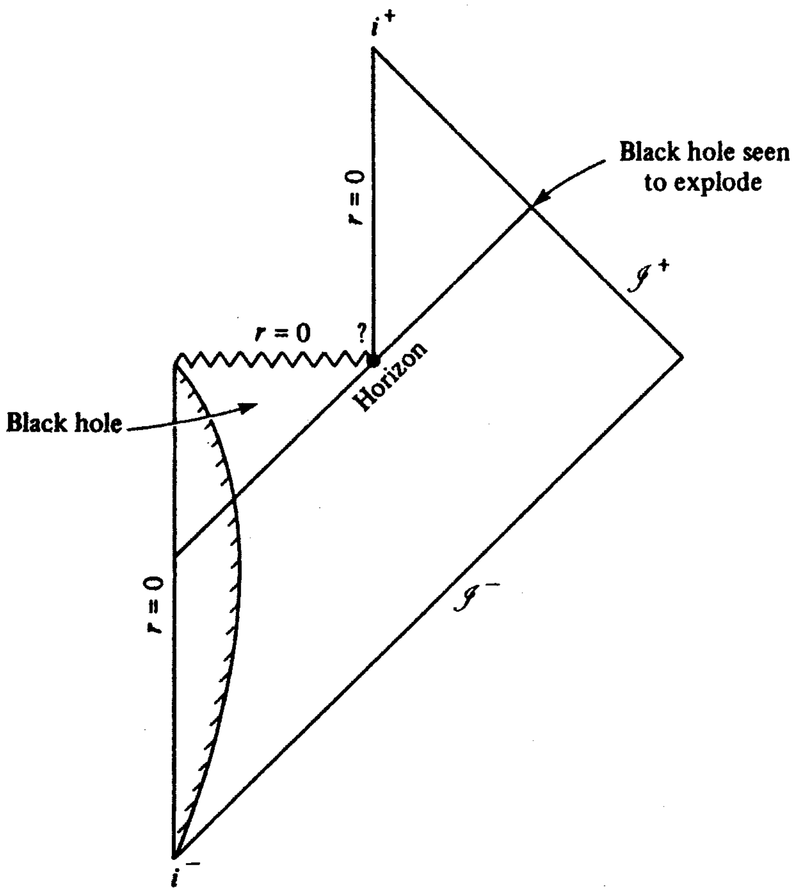


Fig. 27. Penrose diagram for an evaporating black hole. The lower portion of the diagram corresponds to fig. 25. However, when back-reaction is included, the hole evaporates, and the horizon intersects $\mathscr{I}^+$ at a finite time, after passing through a naked singularity (marked ?) Presumably the region above this is Minkowski space.

represents one of the above three alternatives. It should be noted, however, that Gerlach (1976) claims the horizon will not form at all. Whatever the outcome, the detailed behaviour of the hole in the final stages will depend on the nature of subatomic particles at high energies. For example, if there are a small number of truly elementary particles, then the emission rate does not escalate as fast as if the number of particle species rises rapidly at high masses as suggested by Hagedorn (1965). These could affect the observational consequences (Blandford 1977, Rees 1977). Thus, a study of black hole evaporation could provide a unique opportunity for us to probe the physics of ultra-high energy particles.

During the final tenth of a second of its life, a black hole will emit in excess of 10^{30} ergs, or the equivalent of 10^6 megatonne thermonuclear bombs, a significant fraction of which will be in the form of γ-radiation. Existing γ-ray telescopes have failed to reveal evidence for such bursts, although the flux necessary for their detection with present equipment is far greater than expected on other grounds (Blandford & Thorne 1979). More promising is a search for radio bursts caused by the explosive ejection of plasma from the hole into the interstellar magnetic field (Rees 1977). The fact that such bursts have not been detected with current equipment enables an upper limit to be placed on the black hole explosion rate of about $10^{-5}\,\mathrm{pc}^{-3}\,\mathrm{yr}^{-1}$ (Meikle, 1977).

Assuming that the higher order gravitational terms can be neglected, the lifetime of an evaporating hole can be computed from the luminosity (8.38). One obtains $10^{-26}\,(M/1\,\mathrm{gm})^3$ sec, so for a black hole of fermi size (10^{-13} cm), about 10^{15} gm, the lifetime is comparable to the age of the universe. As it is most unlikely that such mini-holes would form anywhere other than the primeval universe, this implies that black holes with an initial mass less than 10^{15}gm would have evaporated away by now.

In spite of the fact that a quantum black hole creates elementary particles and antiparticles in pairs, some of the subatomic conservation laws are violated. For example, a hole that forms from the collapse of a star swallows up mainly baryons, but emits mainly neutrinos and photons, as for the greater part of its life its temperature is too low for massive particle emission. Thus, the law of baryon number conservation is transcended. (It must be remembered that in spite of the evaporation of the hole there is still a singularity present at which baryons may leave spacetime.)

The fact that the quantum black hole enables some of the laws of particle physics to be transcended means that the presence of a hole allows certain subatomic processes to occur that would otherwise be forbidden. One may even suppose that reactions could take place via virtual black hole intermediate states. For example, Hawking (1981) has discussed

$$\mu^{\pm} \rightarrow \text{black hole} \rightarrow e^{\pm} + \gamma$$

(the reaction $\mu \rightarrow e + \gamma$ is otherwise forbidden by conservation of muon lepton number).

Undoubtedly the most persuasive evidence that the Hawking black hole radiance should be taken seriously is the strong connection that it provides between black holes and thermodynamics. Even before Hawking's paper it was appreciated that black holes conform to four laws closely analogous to the four laws of classical thermodynamics (Bardeen, Carter & Hawking 1973). The existence of a surface gravity parameter κ, constant over the event horizon, is reminiscent of the zeroth law of thermodynamics that requires the existence of a temperature parameter constant throughout a system in thermal equilibrium. The conservation of energy during black hole encounters and processes such as the Penrose mechanism (see page 262) and super-radiance, where energy is extracted from, or delivered to, a black hole, is equivalent to the first law of thermodynamics. Moreover, it has already been mentioned (page 263) that a kind of third law, forbidding approach to an extreme Kerr–Newman black hole, was known.

The second law of thermodynamics, requiring the irreversible increase of entropy, finds a natural analogue in Hawking's area theorem (Hawking 1972), which requires the event horizon area $\mathscr{A}$ to be non-decreasing

$$\mathrm{d}\mathscr{A} \geq 0 \tag{8.62}$$

in all black hole processes for which the weak energy condition (Hawking & Penrose 1970) is satisfied. This strongly suggests the identification of $\mathscr{A}$ with entropy $\mathscr{S}$.

The existence of black hole entropy is also implied by the well known connection between entropy and information (Shannon & Weaver 1949; see also the review of Wehrl 1978). When a star implodes to form a black hole, all information about the internal microstates of the star is hidden by the event horizon. Assuming roughly one bit of information per subatomic particle, the total information lost down the hole is about M/m, where M is the mass of the hole and m is the mass of a typical subatomic constituent of the sacrificed body. The associated entropy is

$$\mathscr{S} \sim Mk_{\mathrm{B}}/m.$$

It may appear that, in principle, there is no lower limit to m, and $\mathscr{S}$ should be unbounded. (This accords with the fact that a classical black hole, being black, has zero temperature.) However, as originally pointed out by Bekenstein (1972, 1973) there is a lower bound on m due to the fact that, in order to 'fit' into the hole, the Compton wavelength of the constituent particles should be $\lesssim M$ (the black hole radius). Hence, the maximum information loss is $\sim M^2/h$ and so

$$\mathscr{S} \sim M^2 k_B/\hbar = M^2 k_B \propto k_B \mathscr{A}$$

in our units. Clearly $\mathscr{S}$ becomes infinite in the classical limit ($\hbar \to 0$).

Hawking's work places Bekenstein's conjecture on a firm foundation, and supplies the precise relation (see, for example, DeWitt 1975)

$$\mathscr{S} = \tfrac{1}{4} k_B \mathscr{A}. \tag{8.63}$$

The area law (8.62) is thus seen to be merely a special case of the second law of thermodynamics, $d\mathscr{S} \geq 0$.

The evaporation of the hole, during which the area shrinks, is in violation of (8.62). This is because the existence of negative energy vacuum stress violates the weak energy condition on which the theorem is based. There is, however, no violation of the second law of thermodynamics, because account must also be taken of the increase of entropy in the environment of the hole brought about by the emission of thermal radiation. The *total* entropy, therefore, still increases. Nor does it appear possible to violate the second law by deliberately shooting negative energy (such as from a moving mirror – see §7.1) down a black hole (Ford 1978).

It is possible to develop a complete theory of black hole thermodynamics (Davies 1977*c*, 1978, Hawking 1976*a*, Hut 1977) including features such as phase transitions, Carnot cycles, stability analysis, etc., and even extend the theory to non-equilibrium situations (Candelas & Sciama 1977, Sciama 1976; see also Zurek 1980). Some problems remain, however: What is the relation between information loss due to the imploded star, and the internal microstates of an eternal black hole, such as described by the Kruskal solution (see the next section) that is everywhere a *vacuum* solution of Einstein's equation (Bekenstein 1975)? Can the notion of black hole entropy be extended to arbitrary gravitational fields (Davies 1974, 1981, Penrose 1979)? Will the Hawking radiation always be precisely thermal, even in the presence of interactions (see §9.3) and recoil of the hole (Page 1980)? Does the Hawking process imply time-reversal symmetry violation in quantum gravity (Wald 1980)? Many of these questions are still under active investigation.

8.3 Eternal black holes

In the previous sections we have considered the behaviour of a quantum field in the background spacetime of a collapsing body. However, the features of the Hawking effect turned out to be independent of the details of collapse, which suggests that the effect is more a consequence of the causal and topological structure of spacetime than the specific geometry. This indeed turns out to be the case. One is prompted to dispense entirely with the collapsing body and examine quantum field theory on the maximally extended manifold which is everywhere a solution of the vacuum Einstein equation (see §3.1).

We begin by treating the two-dimensional model of a Schwarzschild black hole, by putting $C=(1-2Mr^{-1})$ in (8.8). The results can easily be extended to the Reissner–Nordstrom case (Davies 1978). The Penrose diagram for this manifold is shown in fig. 6. In terms of Kruskal coordinates $\bar{u}, \bar{v}$ defined by (3.19), the line element is

$$ds^2 = 2Mr^{-1}e^{-r/2M}d\bar{u}\,d\bar{v}, \tag{8.64}$$

which is regular everywhere except at the physical singularity $r=0$. Though this spacetime is symmetric under time reversal, the quantum state imposed on it need not be.

Two natural basis modes for the massless scalar wave equation exist, being either proportional to $e^{-i\omega u}, e^{-i\omega v}$, in terms of Schwarzschild null coordinates u and v, or $e^{-i\omega\bar{u}}$, $e^{-i\omega\bar{v}}$ in terms of Kruskal null coordinates (3.19). The former set oscillate infinitely rapidly on the event horizon, while the latter set are regular on the entire manifold. Each will have its associated vacuum, which we denote by $|0_S\rangle$ and $|0_K\rangle$ respectively. The scalar Green functions in each case will be given in terms of the Minkowski space Green function (4.23) with $L\to\infty$ by a relation of the form (3.154) with $n=2$. Discarding, as usual, an infinite constant (infrared divergence) one obtains

$$D_S^{(1)}(x'',x') = -(1/2\pi)\ln\Delta u\,\Delta v \tag{8.65}$$

$$D_K^{(1)}(x'',x') = -(1/2\pi)\ln\Delta\bar{u}\,\Delta\bar{v}. \tag{8.66}$$

If we transform $D_K^{(1)}$ into Schwarzschild coordinates we obtain

$$\begin{aligned} D_K^{(1)}(x'',x') = &-(1/2\pi)\ln[\cosh\kappa(t''-t') - \cosh\kappa(r^{*''}-r^{*'})] \\ &+ \text{function of } (r'',r') \end{aligned} \tag{8.67}$$

which is manifestly invariant under the transformation $t''\to t''+2\pi i n/\kappa$ (n integer), where here the surface gravity $\kappa=(4M)^{-1}$. That is, $D_K^{(1)}$ is periodic in imaginary Schwarzschild time, with period $2\pi/\kappa$. As explained in §2.7 (see (2.104)), this is a feature characteristic of thermal Green functions. In this case the temperature is $\kappa/2\pi k_B$, the Hawking temperature (8.27). A similar periodicity is found in four dimensions, even when the black hole is rotating

(Hartle & Hawking 1976). For example, for modes with azimuthal quantum number m

$$D_{\mathrm{K}}^{(1)}(t'' - t' + 2\pi \mathrm{i} n/\kappa; r'', r') = \mathrm{e}^{2\pi m\Omega/\kappa} D_{\mathrm{K}}^{(1)}(t'' - t'; r'', r') \qquad (8.68)$$

confirming that $m\Omega$ behaves like a chemical potential (see remark following (2.104)). These concepts also generalize readily to the case of electrically charged holes and quanta (Gibbons & Perry 1976, 1978).

Far from the hole, the Green function (8.65) reduces to the usual $D^{(1)}$ for flat space quantum field theory; in this region $|0_{\mathrm{S}}\rangle$ is the conventional vacuum state. Evidently, an observer in this region would regard $|0_{\mathrm{K}}\rangle$ as a thermal state at temperature $\kappa/2\pi k_{\mathrm{B}}$ and $D_{\mathrm{K}}^{(1)}$ as a thermal Green function, i.e., the black hole is immersed in a bath of thermal radiation at the Hawking temperature. Indeed, careful analysis (Birrell & Davies 1978*a*) shows that far from the hole $D_{\mathrm{K}}^{(1)}$ can be written as an infinite sum (2.111) of Green functions $D_{\mathrm{S}}^{(1)}$, confirming that it is precisely a thermal Green function. Clearly (8.67) is invariant under time reversal, so the thermal bath represents a flux of thermal radiation passing into the hole from $\mathscr{I}^-$ at a rate equal to that of the Hawking flux reaching $\mathscr{I}^+$. This may be readily confirmed by using $D_{\mathrm{K}}^{(1)}$ to compute $\langle 0_{\mathrm{K}}| T_{\mu\nu} |0_{\mathrm{K}}\rangle$ in the standard way, to find an energy density at $\mathscr{I}^{\pm}$ of $\kappa^2/24\pi$. The vacuum $|0_{\mathrm{K}}\rangle$ therefore describes a steady-state thermal equilibrium between the black hole and its surroundings, such as would be obtained by confining the hole to the interior of a perfectly reflecting cavity. (If the cavity is too large, the equilibrium is unstable – see for example, Davies (1978)). The state $|0_{\mathrm{K}}\rangle$ is known as the Hartle–Hawking, or Israel vacuum (Hartle & Hawking 1976, Israel 1976).

A comparison of (3.19) with (4.69) and (4.70), putting $a = \kappa = (4M)^{-1}$, reveals that the state $|0_{\mathrm{S}}\rangle$ is closely analogous to the state $|0_{\mathrm{R}}\rangle$ associated with an accelerated observer in the Rindler wedge of Minkowski space. Likewise $|0_{\mathrm{K}}\rangle$ corresponds to $|0_{\mathrm{M}}\rangle$. This is no surprise, as the horizon structures of the Rindler wedge and the Schwarzschild black hole are identical. Comparison of figs. 6 and 14 shows that the region R of Minkowski space possesses the same causal relationship to L as the region I (representing the universe external to the black hole) has to II (the 'mirror' universe). In the same way as in §4.5, we may here relate the modes defined on the extended manifold to those defined in regions I and II only (Unruh 1976). Because the mathematical relation between the coordinates u, v and $\bar{u}, \bar{v}$ is identical in both examples, we do not need to repeat the analysis here, but may write down straight away the Bogolubov transformations connecting both sets of modes: these will be formally identical to (4.95) and (4.96), where $b_k^{(2)}$ now represents an annihilation operator for excitation of the Schwarzschild modes in region I (annihilation of a particle in 'our' universe outside the hole) and $b_k^{(1)}$ represents an annihilation operator for quanta in region II. Similarly the d_k operators are associated with the Kruskal modes.

The Bogolubov transformation leads, as demonstrated in §4.5, to a thermal spectrum (4.97), with temperature $a/2\pi k_B = \kappa/2\pi k_B$, as expected. However, the transformation contains more information than this. Putting $\tanh\phi_\omega = e^{-\pi\omega/a}$, (4.95) and (4.96) give

$$b_k^{(1)} = e^{iJ} d_k^{(1)} e^{-iJ} \tag{8.69}$$

where $\omega = |k|$,

$$J = \sum_k i\phi_\omega (b_{-k}^{(1)\dagger} b_k^{(2)\dagger} - b_{-k}^{(1)} b_k^{(2)}) \tag{8.70}$$

and we have used the commutation relations.

From the definitions

$$b_k^{(1)}|0_S\rangle = b_k^{(2)}|0_S\rangle = 0 \tag{8.71}$$

$$d_k^{(1)}|0_K\rangle = d_k^{(2)}|0_K\rangle = 0 \tag{8.72}$$

and $$e^{-iJ} b_k^{(1)}|0_S\rangle = d_k^{(1)} e^{-iJ}|0_S\rangle,$$

which follows from (8.69), one has

$$|0_K\rangle = e^{-iJ}|0_S\rangle. \tag{8.73}$$

This can be expanded and rearranged to give

$$\begin{aligned} |0_K\rangle &= \exp\{\sum_k [-\ln\cosh\phi_\omega + \tanh\phi_\omega b_k^{(1)\dagger} b_k^{(2)\dagger}]\}|0_S\rangle \\ &= \prod_k (\cosh\phi_\omega)^{-1} \sum_{n_k=0}^{\infty} e^{-n_k\pi\omega/\kappa} |n_k^{(1)}\rangle |n_k^{(2)}\rangle \end{aligned} \tag{8.74}$$

in terms of states with n_k quanta in region I and n_k quanta in region II (Unruh 1976, Israel 1976).

If an observer is restricted to the region I (outside the black hole) then he will not have access to the modes in region II and will not be able to measure $|n_k^{(1)}\rangle$. When this observer measures an observable A, with associated operator $\hat{A}$, in the quantum state $|0_K\rangle$, then the $|n_k^{(1)}\rangle$ part of (8.74) just factors out:

$$\begin{aligned} \langle 0_K|\hat{A}|0_K\rangle &= \sum_{n_k} \prod_k \langle n_k^{(2)}|\hat{A}|n_k^{(2)}\rangle \exp(-2n_k\pi\omega/\kappa) \\ &\quad \times [1 - \exp(-2\pi\omega/\kappa)] \\ &= \mathrm{tr}(\hat{A}\rho) \end{aligned} \tag{8.75}$$

where ρ can be written

$$\sum_n \prod_k \left\{ \frac{e^{-\beta E_n}}{\sum_{m=0}^{\infty} e^{-\beta E_m}} \right\} |n_k\rangle\langle n_k| \qquad (8.76)$$

with $E_n = n\omega$, $\beta = 2\pi/\kappa$, and regarded as a density matrix corresponding to a *thermal* average (see §2.7) at temperature $\kappa/2\pi k_B$. In particular, if $\hat{A}$ is the particle number operator for mode σ

$$\langle 0_K | N_\sigma | 0_K \rangle = \sum_{n=0}^{\infty} n_\sigma e^{-\beta E_n} \Big/ \sum_{m=0}^{\infty} e^{-\beta E_m}$$

$$= 1/(e^{\beta\sigma} - 1), \qquad (8.77)$$

i.e., a Planck spectrum.

The fact that the pure state $|0_K\rangle$, defined on the whole manifold, appears to an observer confined to region I as a mixed state to be described by a density matrix (8.76), is no surprise. The presence of the event horizon hides the information about the modes in region II, and this loss of information is naturally associated with a nonzero entropy in region I (the entropy of a pure state is zero). Indeed, not only does (8.76) imply a Planck spectrum, it also implies completely thermal radiation, i.e., the complete absence of correlations between the emitted quanta (Parker 1975, Wald 1975, Hawking 1976*b*). For example, the probability that $|0_K\rangle$ contains ${}^1n_{k_1}$ quanta in mode k_1, ${}^2n_{k_2}$ quanta in mode k_2, etc. in region I, is

$$|\langle {}^1n_{k_1}, {}^2n_{k_2}, \ldots, | 0_K \rangle|^2 = \prod_j \exp(-2\pi\, {}^jn_{k_j} |k_j|/k)[1 - \exp(-2\pi/k_j|/\kappa)]$$

$$= \prod_j P({}^jn_{k_j}), \qquad (8.78)$$

where $P({}^jn_{k_j})$ is the probability that the mode k_j contains ${}^jn$ particles. The fact that all $P({}^jn_{k_j})$ are multiplied shows that these individual probabilities are independent of each other.

The details concerning the density matrix and thermal states are immediately generalized to four dimensions and also apply to the case of uniformly accelerated observers. Attempts have also been made (Gibbons & Hawking 1977*a*, Lapedes 1978*b*, Lohiya & Panchapakesan 1978, 1979, Denardo & Spallucci 1980*a*, *b*) to extend them to non-black hole horizons, such as de Sitter space, though in the latter case the physical significance of the thermal structure is not so clear.

In arriving at these results we have neglected the back-reaction of the particle production on the background spacetime. The recoil of emission from the black hole will necessarily introduce some correlations on the emitted quanta when only a small fraction of the total energy has been radiated (Page 1980). This, of course, would occur with any hot body, and is a consequence of the hole not being in thermodynamic equilibrium with its environment.

Although the above treatment has been based on the maximally extended Kruskal manifold, the thermal nature of the radiation and the absence of correlations is also a feature of the more realistic model involving collapse. In the case of a massless field, $\mathscr{I}^-$ is a Cauchy surface for the collapse spacetime, but $\mathscr{I}^+$ is not; one must take into account the propagation of the field across the future event horizon H^+. The union of $\mathscr{I}^+$ and H^+ does form a Cauchy surface (see fig. 25). Thus, to construct a complete set of modes in the out region, it is necessary to augment the standard exponential modes on $\mathscr{I}^+$ with modes on H^+. An explicit construction of such a complete set has been given by Wald (1975). However, the results of measurements made on $\mathscr{I}^+$ are independent of the detailed form of the horizon modes. Restricting measurement to $\mathscr{I}^+$ and forsaking information about the modes on the horizon once again introduces a density matrix of the form (8.76). One may envisage the horizon modes as representing quanta that cross into the black hole. Thus, one arrives at the model of the Hawking process mentioned on page 264: particle pairs are created close to the horizon, one particle of which travels to $\mathscr{I}^+$ (ignoring back scattering) and the other of which enters the hole. By relinquishing information about the latter quanta, the former must be described by a mixed state (i.e., using a density matrix). Thus, whether the Hawking effect is examined on the full Kruskal manifold or for the collapsing body, measurements at $\mathscr{I}^+$ will acquire a thermal nature.

We have so far discussed two 'vacuum' states, $|0_K\rangle$ and $|0_S\rangle$, the former corresponding to a bath of thermal radiation, the latter reducing to the conventional Minkowski vacuum state at large distance from the hole. As always in quantum theory, additional physical criteria are necessary to decide which quantum state corresponds to the physical situation of interest. If the universe contains an eternal black hole, only observation can reveal what quantum state is actually realized. However, the state $|0_K\rangle$ clearly reproduces the features outside the hole that would be present if a black hole that was formed from a collapsing body were subsequently confined in a box and allowed to come into thermal equilibrium.

It is possible to construct yet another 'vacuum' state on the maximally extended manifold, that will reproduce the effects of a collapsing body, i.e., yield a time-asymmetric thermal *flux* from the hole, rather than a thermal bath. This state was discovered by Unruh (1976).

Note that $|0_S\rangle$ is associated with modes that are positive frequency with respect to the Killing vector ∂_t (cf. 4.77)), while $|0_K\rangle$ is defined using modes which are positive frequency with respect to the vector $\partial_{\bar{t}}$, where $\bar{t}$ is the Kruskal time ($\bar{t} = \frac{1}{2}(\bar{u} + \bar{v})$) (cf. (4.82)). Inspection of fig. 6 shows that the null rays $u =$ constant, that mark the surfaces of constant phase of the standard (Minkowski space) exponential modes ($e^{-i\omega u}$) at $\mathscr{I}^+$, when followed backwards in time would, if the collapsing body had been absent, cross the *past* horizon, H^-, of the extended Kruskal manifold. In the collapse picture (fig. 25) the null ray γ (the latest advanced ray to travel through the collapsing body, and emerge to form the future horizon H^+) plays a role analogous to H^- in fig. 6. Incoming rays immediately prior to γ reflect off the centre of coordinates and pass across γ on their way to $\mathscr{I}^+$, close to H^+. These rays represent waves that are positive frequency with respect to the affine parameter along γ (Hawking 1975; see also Parker 1977, Gibbons 1980). For rays that emerge very close to H^+, this affine parameter coincides to good approximation with the affine parameter $\bar{u}$ along the past horizon in the extended manifold, fig. 6. Hence, the geometrical effects of the collapse may be mocked up by choosing the in modes to be of the form $e^{-i\omega\bar{u}}$ along H^- and $e^{-i\omega v}$ along $\mathscr{I}^-$. The modes $e^{-i\omega\bar{u}}$ are of positive frequency with respect to the vector $\partial_{\bar{u}}$, which is a Killing vector on H^- (Unruh 1976). The collapsing body may then be dispensed with, and we may work with the full Kruskal manifold. The vacuum state associated with this choice of modes in the past is called the Unruh vacuum, and we denote it by $|0_U\rangle$. It corresponds to a thermal flux of particles leaving the region of the black hole. The time reversed state, with modes $e^{-i\omega\bar{v}}$ along H^+ and $e^{-i\omega u}$ along $\mathscr{I}^+$, would describe a steady flux of thermal radiation going into the hole.

Further insight into the properties of the vacua $|0_S\rangle$ (sometimes called the Boulware vacuum (Boulware 1975*a*, *b*)), $|0_K\rangle$ and $|0_U\rangle$ can be obtained by studying the experiences of a model particle detector at a fixed Schwarzschild radius $r = R$. For massless fields in two dimensions, the response function (3.55) can be evaluated explicitly using the Wightman functions

$$D_S^+(x, x') = -(1/4\pi)\ln[(\Delta u - i\varepsilon)(\Delta v - i\varepsilon)] \tag{8.79}$$

$$D_K^+(x, x') = -(1/4\pi)\ln[(\Delta\bar{u} - i\varepsilon)(\Delta\bar{v} - i\varepsilon)] \tag{8.80}$$

$$D_U^+(x, x') = -(1/4\pi)\ln[(\Delta\bar{u} - i\varepsilon)(\Delta v - i\varepsilon)]. \tag{8.81}$$

For a detector at $r = R$, the proper time is given by

$$\mathrm{d}\tau = (1 - 2M/R)^{\frac{1}{2}}\mathrm{d}t, \tag{8.82}$$

while $\Delta u = \Delta v = \Delta t$ and, from (3.19),

$$\Delta\bar{u} = -4M\mathrm{e}^{R^*/4M}(\mathrm{e}^{-t/4M} - \mathrm{e}^{-t'/4M}), \tag{8.83}$$

$$\Delta\bar{v} = 4M\mathrm{e}^{R^*/4M}(\mathrm{e}^{t/4M} - \mathrm{e}^{t'/4M}). \tag{8.84}$$

For the Boulware vacuum, substituting (8.79) into the response function (3.55), gives an expression which is essentially the same as for a detector at rest in two-dimensional Minkowski space in the usual vacuum. As in that case one obtains $\mathscr{F}(E) = 0$ for $E > 0$, i.e., the detector registers no particles in $|0_S\rangle$.

For the Unruh vacuum, substitution of (8.81) into (3.55), noting (8.82) and (8.84), gives a response function per unit proper time which is identical to (4.54) with $w = 0$ and $\kappa = [16M^2(1 - 2M/R)]^{-\frac{1}{2}}$. Thus,

$$\mathscr{F}_U(E)/\text{unit proper time} = 1/E(\mathrm{e}^{E/kT} - 1) \tag{8.85}$$

where $kT = [64\pi^2M^2(1 - 2M/R)]^{-\frac{1}{2}}$. As in the case of the accelerating mirror, which led to (4.54), the detector registers a flux of particles at apparent temperature $T_0 = 1/8\pi k_B M$ given by the Tolman relation (4.98). This is in agreement with (8.27) (noting $\kappa = (4M)^{-1}$ for a Schwarzschild black hole). As the detector approaches the horizon ($R \to 2M$), the temperature of the flux determined by the detector diverges. This is due to the fact that the detector must be non-inertial to maintain a fixed distance from the black hole. The magnitude of the acceleration relative to the local freely-falling frame is $M/[R^2(1 - 2M/R)^{\frac{1}{2}}]$. Such acceleration gives rise to the detection of additional particles, as in Minkowski space (see §3.3). As the horizon is approached, the acceleration diverges, as does the temperature (cf. (3.68)).

The calculation for the vacuum $|0_K\rangle$ is similar, and one obtains twice the result (8.85), corresponding to a bath of radiation at apparent temperature T.

The approximate response for a particle detector in four dimensions has been studied by Unruh (1976) and Candelas (1980). In particular, Candelas has evaluated the response function as $R \to 2M$ and $R \to \infty$ for each of the three vacua studied above, and obtains results consistent with those in two dimensions.

An extensive analysis of the properties of $|0_S\rangle, |0_K\rangle$ and $|0_U\rangle$ has been given by Fulling (1977).

8.4 Analysis of the stress-tensor

It is a straightforward matter to compute $\langle T_{\mu\nu}\rangle$ in the above 'vacuum' states using the theory of chapter 6. In the two-dimensional case the results may be obtained in closed form. However, it is instructive to carry out a general analysis of $\langle T_{\mu\nu}\rangle$ in the spirit of the Wald axioms (see §6.6). This was done by Christensen & Fulling (1977).

We begin by treating first the two-dimensional 'Schwarzschild' case. Putting $C = (1 - 2M/r)$ in (8.8) the covariant conservation equation (6.142) reduces to

$$\langle T^r_t\rangle = \text{constant} \tag{8.86}$$

$$\frac{\partial}{\partial r}\left[\left(1 - \frac{2M}{r}\right)\langle T_r{}^r\rangle\right] = \frac{M}{r^2}\mathscr{T}, \tag{8.87}$$

where $\mathscr{T}$ is the trace $\langle T_\mu{}^\mu\rangle$ and we assume that all the components of $\langle T_{\mu\nu}\rangle$ are independent of time. Integrating (8.87) and using $\langle T_t{}^t\rangle = \mathscr{T} - \langle T_r{}^r\rangle$ yields

$$\langle T_\mu{}^\nu\rangle = \langle T^{(1)}{}_\mu{}^\nu\rangle + \langle T^{(2)}{}_\mu{}^\nu\rangle + \langle T^{(3)}{}_\mu{}^\nu\rangle, \tag{8.88}$$

where, in (t, r^*) coordinates,

$$\langle T^{(1)\nu}_\mu\rangle = \begin{bmatrix} \dfrac{-1}{(1-2M/r)}H(r) + \mathscr{T}(r) & 0 \\ 0 & \dfrac{1}{(1-2M/r)}H(r) \end{bmatrix} \tag{8.89}$$

$$\langle T^{(2)\nu}_\mu\rangle = \frac{K}{M^2}\frac{1}{(1-2M/r)}\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \tag{8.90}$$

$$\langle T^{(3)\nu}_\mu\rangle = \frac{Q}{M^2}\frac{1}{(1-2M/r)}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{8.91}$$

and

$$H(r) = M\int_{2M}^{r} \mathscr{T}(\rho)\rho^{-2}\mathrm{d}\rho, \tag{8.92}$$

while K and Q are constants to be determined according to which quantum state is chosen.

Consider first the state $|0_S\rangle$. This has been investigated in detail by Blum (1973) and Boulware (1975*a*, 1976). At $\mathscr{I}^\pm$ it coincides with the conventional Minkowski space vacuum $|0_M\rangle$ so there can be no radiation at infinity. Hence we must choose $K=0$ and $Q=-M^2H(\infty)$, so that (8.88) vanishes at large r. Using (6.121) and (7.2) to give

$$\mathscr{T}(r) = -M/6\pi r^3, \tag{8.93}$$

one obtains from (8.92)

$$H(r) = -(1/384\pi M^2) + (M^2/24\pi r^4). \tag{8.94}$$

Thus, for the state $|0_S\rangle$, $Q = 1/384\pi$ and $\langle T_\mu{}^\nu\rangle$ reduces to the expression given by (8.47) (with $e=0$). Thus, $|0_S\rangle$ is the state appropriate to a vacuum around a static star.

It cannot represent the state of a black hole, however, as may be seen from the following analysis: Noting that (cf. (4.17)–(4.19))

$$T_{uu} = \tfrac{1}{4}(T_{tt} + T_{r^*r^*} - 2T_{tr^*}), \tag{8.95}$$

one finds from (8.88)–(8.91) that

$$\begin{aligned}\langle T_{uu}\rangle &= -\tfrac{1}{2}(H + Q/M^2) + \tfrac{1}{4}(1-2M/r)\mathscr{T}\\ &\to -Q/2M^2 \quad \text{as} \quad r\to 2M.\end{aligned} \tag{8.96}$$

Now the Schwarzschild coordinates are singular at the horizon. To investigate the behaviour of the stress-tensor there, we transform to Kruskal coordinates (3.19), which are regular at the horizon, and obtain

$$\langle T_{\bar{u}\bar{u}}\rangle = -(Q/32M^4)e^{-r/M}\bar{v}^2(1-r/2M)^{-2}. \tag{8.97}$$

This diverges as $r\to 2M$ unless $Q=0$. The quantity $\langle T_{\bar{u}\bar{u}}\rangle$ is more indicative of the physical situation at the horizon, i.e., to what a freely-falling observer would measure (Fulling 1977). Hence the state $|0_S\rangle$ is unphysical if the metric has the Schwarzschild form as far as $r=2M$. In practice, if one attempted to set up such a quantum state around a black hole, the back-reaction near $r=2M$ would greatly modify the gravitational field.

Turning to another choice of constants consider the time symmetric case, where $\langle T_\mu{}^\nu\rangle$ is required to be regular on both the past and future horizons. An analysis of $\langle T_{vv}\rangle$ as $r\to 2M$ shows that one requires both $K=Q=0$. Hence $\langle T_\mu{}^\nu\rangle = \langle T^{(1)}{}_\mu{}^\nu\rangle$, and at $r\to\infty$ this reduces to

$$\tfrac{1}{12}\pi(k_B T)^2\begin{pmatrix}2 & 0\\ 0 & -2\end{pmatrix} \tag{8.98}$$

with $T = 1/8\pi k_B M = \kappa/2\pi k_B$. This is the stress-tensor for a *bath* of thermal radiation at temperature T, and therefore reproduces the properties of the Israel–Hartle–Hawking vacuum $|0_K\rangle$.

Finally, to describe the Hawking evaporation process one requires an outward flux of thermal radiation only, for which the stress-tensor at large r is

$$\tfrac{1}{12}\pi(k_B T)^2\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}, \tag{8.99}$$

with $T = 1/8\pi k_B M$. Thus the energy density and flux are numerically equal, which demands that (with $Q = 0$)

$$\begin{aligned} K &= \tfrac{1}{2}M^2[H(\infty) - \mathscr{T}(\infty)] \\ &= \tfrac{1}{2}M^3\int_{2M}^{\infty}\mathscr{T}(\rho)\rho^{-2}\mathrm{d}\rho, \end{aligned} \tag{8.100}$$

noting that $\mathscr{T}(\infty) = 0$. This is a remarkable relation, giving the Hawking flux at $r = \infty$ (i.e., K/M^2) in terms of an integral over the trace of the stress-tensor. The fact that we know from the thermal nature of the Bogolubov transformation that $K \neq 0$ (there must be an energy flux if the spectrum is thermal) proves that there exists an anomalous trace, $\mathscr{T}$. In chapter 6 the existence of a trace anomaly was deduced from subtle arguments involving renormalization of $\langle T_{\mu\nu}\rangle$. Now, following Christensen & Fulling (1977), we have derived its existence by an entirely independent route, and one which makes no reference to renormalization.

One cannot, of course, obtain the form of $\mathscr{T}$ from (8.100), although, given that it must be geometrical, one immediately concludes $\mathscr{T} \propto R$, as R is the only geometrical scalar with dimensions (length)$^{-2}$ in two dimensions. Hence, knowledge of K fixes the anomaly coefficient or vice versa. From either (8.99) or (8.93), we arrive at $K = -(768\pi)^{-1}$ and $\mathscr{T} = -R/24\pi$. We have demanded that $Q = 0$ so that the stress-tensor is regular on the future horizon H^+ and the situation describes the Unruh vacuum $|0_U\rangle$. The stress-tensor is not, however, regular on H^-.

Note that, in the case of a massless fermion field, the alteration of the Planck factor to $(e^{\omega/k_B T} + 1)$ introduces a factor $\frac{1}{2}$ into (8.98) and (8.99). When we take into account the two helicity states, the results are identical to the scalar case. Thus, we deduce that, in two dimensions, the conformal anomaly and the Hawking flux are the same for spins 0 and $\frac{1}{2}$, which confirms the results of chapter 6.

Turning to the four-dimensional case, the requirements of spherical symmetry, time-independence and covariant conservation do not suffice to uniquely relate the trace to the Hawking flux; there is an additional arbitrariness in the angular components $\langle T_\theta{}^\theta \rangle = \langle T_\phi{}^\phi \rangle$. Christensen & Fulling (1977) have analysed this case in detail and present arguments that place qualitative restrictions on $\langle T_\theta{}^\theta \rangle$. Their conjectures on the asymptotic form of the stress-tensor as $r \to \infty$ or $2M$ in each of the three 'vacuum' states have been confirmed by the explicit calculations of Candelas (1980).

We shall finish this section by briefly returning to the two-dimensional Reissner–Nordstrom black hole with $C = (1 - 2Mr^{-1} + e^2 r^{-2})$. This spacetime has two horizons, r_+, defined by (8.42), and another defined by

$$r_- = M - (M^2 - e^2)^{\frac{1}{2}}, \tag{8.101}$$

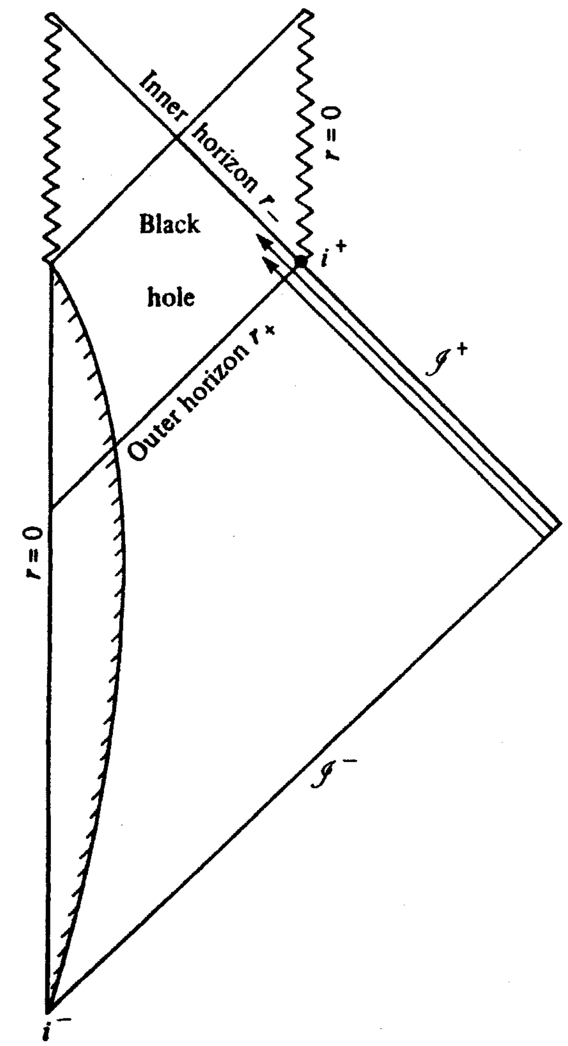


Fig. 28. Penrose diagram for a spherical star that collapses to form a charged, Reissner–Nordstrom, black hole. (Compare fig. 25. The manifold is drawn incomplete and can be extended vertically. This extension is irrelevant for the present discussion.) The singularity is timelike. Late time, infalling null rays from $\mathscr{I}^-$ crowd up, blueshifted, along r_-.

at which C also vanishes. This inner horizon is a Cauchy horizon, but not an event horizon for the spacetime (see, for example, Hawking & Ellis 1973, chapter 5). A conformal diagram is shown in fig. 28, from which one may see at a glance, recalling the conformal compression of $\mathscr{I}$, that r_- represents a surface of infinite blueshift, because null rays from $\mathscr{I}^-$ crowd along r_- inside the black hole.

It is a simple matter to evaluate $\langle T_\mu{}^\nu \rangle$ for the interior of the hole in the states $|0_U\rangle$ and $|0_K\rangle$ (Davies 1978). One finds (Hiscock 1977*a*, *b*, 1979) that $\langle T_\mu{}^\nu \rangle$ diverges on r_- even in a coordinate system that is regular there. General arguments (Birrell & Davies 1978*b*, Hiscock 1979) suggest that this result remains true in four dimensions. One concludes that the back-reaction effects of this quantum stress would disrupt the interior geometry of the hole and presumably give rise to a singularity along the Cauchy horizon, that would prevent analytic extension of the manifold to other asymptotically flat spacetime regions (see also Simpson & Penrose 1973, McNamara 1978).

Finally, it has been shown by Curir & Francaviglia (1978) that in the case of a rotating black hole, the inner horizon may be interpreted as a sort of negative temperature surface (in the Hawking sense), satisfying a type of area theorem similar to (8.62), and bearing some relation to the concept of spin temperature. (See also Calvani & Francaviglia 1978.)

8.5 Further developments

The detailed analysis to which the quantum black hole has been subjected, and the remarkable consistency of the results from widely different theoretical approaches, has encouraged a number of authors to use the quantum black hole as the starting point for a more searching analysis of quantum field theory in curved space, and even quantum gravity. Hawking and coworkers, in particular, have developed an extensive programme in which thermal aspects of black hole physics play a central role (for a review see Hawking 1979).

Much of this further work depends on the observation that the Green function which describes a black hole in the Israel–Hartle–Hawking vacuum state $|0_K\rangle$ (or its generalization to other spacetimes) is periodic in imaginary Schwarzschild time (see page 276). This suggests defining $\tau = it$ and considering, in place of the Schwarzschild spacetime, a related Riemannian space (i.e., negative definite metric, as opposed to pseudo-Riemannian, with indefinite metric) with a Kruskal-like line element

$$ds^2 = -(2M/r)e^{r/2M}(dX^2 + dY^2) - r^2(d\theta^2 + \sin^2\theta \, d\phi^2) \quad (8.102)$$

where

$$X = 4M[r/2M - 1]^{\frac{1}{2}}\, e^{r/4M} \cos \kappa\tau$$

$$Y = 4M[r/2M - 1]^{\frac{1}{2}}\, e^{r/4M} \sin \kappa\tau.$$

This space is Ricci flat and has topology $R^2 \times S^2$. The Killing vector ∂_τ generates rotations about the fixed point $X = Y = 0$. This rotational symmetry endows τ with the properties of an angular coordinate, thus building in the periodicity requirement; the points whose τ coordinates differ by $2\pi/\kappa$ are identified. The origin, $X = Y = 0$, corresponds to $r = 2M$, i.e., the event horizon. The region corresponding to the inside of the black hole, and in particular the singularity at $r = 0$, are absent from this Riemannian space. Hartle & Hawking (1976) use this as a starting point for defining a natural Green function which, when analytically continued back to the pseudo-Riemannian Schwarzschild spacetime, coincides with the Green function associated with $|0_K\rangle$. There is thus an elegant connection established between the geometrical symmetry of the Riemannian manifold, as implemented by rotations in the $X - Y$ plane, and the thermal character of the Hawking radiation. Generalization to Kerr–Newman black holes is straightforward (Gibbons & Perry 1976, 1978).

The replacement of t by $\mathrm{i}\tau$ in the above is reminiscent of the way in which the Feynman propagator is constructed in ordinary Minkowski space quantum field theory (see page 24). In §2.8 it was explained how the convergence problems of the path-integral approach to quantization are alleviated by the formal device of replacing t by $-\mathrm{i}\tau$ in the functional integral (2.115). The exponent then becomes $-\hat{S}$ in place of $\mathrm{i}S$, where $\hat{S}$ is the action in Euclidean spacetime. This treatment can be extended to the Riemannian case.

If the action $\hat{S}$ of the Riemannian manifold is used to construct the generating functional $Z[J]$, as the functional integral over fields which are periodic in the variable τ with period β, the Green functions obtained using (2.117) will not be the usual vacuum Green functions, but rather thermal Green functions at temperature $T = 1/k_B\beta$.

The connection between the path-integral approach and thermodynamics is strengthened by noting that the amplitude for field configuration ϕ_1 at time t_1 to propagate to ϕ_2 at time t_2 is given by

$$\langle \phi_2, t_2 | \phi_1, t_1 \rangle = \int \mathscr{D}[\phi]\, e^{\mathrm{i}S[\phi]}, \tag{8.103}$$

where the integral is over all fields satisfying the given initial and final conditions. In the Schrödinger picture this can be written as

$$\langle \phi_2 | \exp[-\mathrm{i}H(t_2 - t_1)] | \phi_1 \rangle,$$

where H is the Hamiltonian. Putting $t_2 - t_1 = -\mathrm{i}\beta$ and $\phi_2 = \phi_1$, and summing over a complete set of field configurations gives

$$Z = \sum_n \mathrm{e}^{-\beta E_n},$$

where E_n is the energy of the field configuration ϕ_n. From (2.95), Z is identified as the thermodynamic partition function at temperature $T = 1/k_B\beta$ and zero chemical potential. On the other hand, from (8.103), we have

$$Z = \int \mathscr{D}[\phi] \mathrm{e}^{-\hat{S}[\phi]}, \tag{8.104}$$

where the integral is over all fields which are periodic in τ with periodicity β. Thus, using this path integral on the Riemannian manifold as a representation of the partition function, the thermodynamics of the system may be evaluated. (For a rigorous discussion of the connection between thermodynamics and path integrals in Minkowski space, see, for example, Ginibre 1971).

Gibbons & Hawking (1977*a*) have applied this result to the gravitational field itself (see also Hawking 1978). If the quantization of the gravitational field is carried out using the background field method (see chapter 1) the metric is written as

$$g_{\mu\nu} = g_{c\mu\nu} + \bar{g}_{\mu\nu}$$

where g_c is a solution of the classical Einstein equation and $\bar{g}$ is a field giving the quantum fluctuations about this background. Expanding the action in a functional Taylor series about the classical background as

$$\hat{S}[g] = \hat{S}[g_c] + S_2[\bar{g}] + \text{ higher order terms,}$$

where S_2 is quadratic in $\bar{g}$, the partition function is given by

$$\ln Z = -\hat{S}[g_c] + \ln \int \mathscr{D}[\bar{g}] \exp(-S_2[\bar{g}]) + \text{higher order terms.} \tag{8.105}$$

The first term on the right-hand side is the contribution of the classical gravitational field, while the second term gives the 'one-loop' correction due to the quantization of gravitons, and can be treated in the manner discussed for matter fields in chapter 6. The higher order terms give the contribution of graviton Feynman diagrams with more than one loop, and their meaningful evaluation is barred by the non-renormalizable nature of quantum gravity.

Ignoring for now all but the first (classical) term on the right-hand side of (8.105), the action is given by (see §6.6, setting $G = 1$ in the units of this chapter)

$$\hat{S} = -(1/16\pi)\int(R - 2\Lambda)g^{\frac{1}{2}}\,d^4x - (1/8\pi)\int(\chi - \chi^0)(h)^{\frac{1}{2}}d^3x, \qquad (8.106)$$

where g is now the determinant of a Riemannian metric, h is the determinant of the induced metric on the boundary of the manifold, and χ and χ^0 are defined on page 224. Taking the background spacetime to be Schwarzschild spacetime, with Riemannianized metric (8.102), then $R = \Lambda = 0$ and $\hat{S}[g_c]$ reduces to $4\pi M^2$ (Gibbons & Hawking 1977*b*; see also Gibbons 1977*b*). Thus the classical contribution to the partition function is given by

$$\ln Z = -4\pi M^2, \qquad (8.107)$$

from which the average value at temperature $T = 1/k_B\beta = 1/8\pi k_B M$ of the energy, $\langle E\rangle_\beta$, and the entropy, $\mathscr{S}$, are computed by the standard procedure (see, for example, Isihara 1971, §3.2)

$$\langle E\rangle_\beta = -\frac{\partial}{\partial\beta}\ln Z = M$$

$$S = k_B\beta\langle E\rangle_\beta + k_B\ln Z = 4\pi k_B M^2 = \tfrac{1}{4}k_B\mathscr{A},$$

which is identical to (8.63).

This result is all the more remarkable because it has been derived by appeal to the action of the classical *gravitational* field, not the quantum matter fields as before. Indeed, the fact that $\mathscr{S}$ is a purely *geometrical* quantity (the event horizon area) rather than dependent on the types of matter fields present, as is usual for entropy, indicates that the concept of black hole entropy is more fundamental than quantum field theory in curved spacetime, and is really an *intrinsic* quality of the black hole, a feature confirmed by the present treatment. This is the first time that the notion of objective, instrinsic entropy has appeared in physics. In the traditional treatment, entropy has an element of subjectivity associated with coarse-graining and our inability to distinguish between members of various classes of microstates.

The development of Riemannianized quantization of the gravitational field has led to a connection between black holes and the subject of instantons (see, for example, Hawking 1977*c*, Charap & Duff 1977, Gibbons & Hawking 1979; see also Gibbons & Pope 1978 for the extension of these ideas to other spacetimes), and a better understanding of the relationship between spacetime topology and quantum field theory (Hawking 1979). Similar techniques to those discussed in this section have been applied to de Sitter space (Gibbons & Hawking 1977*b*) and Robinson–Bertotti spacetime (Lapedes 1978*a*) where thermal effects also arise. Further discussion of these topics is beyond the scope of this book, and we refer the reader to the literature.

Some attention has also been given to the creation of particles near spacetime singularities. The intense quantum effects that accompany escalating spacetime curvature might be expected to produce a strong back-reaction that could be relevant to the question of cosmic censorship. Ford & Parker (1978) have investigated whether quantum back-reaction will disrupt the formation of an idealized naked singularity brought about by the implosion of a highly charged ($e^2 > M^2$) body. Quantum field theory near 'white holes' has also been discussed by Zel'dovich, Novikov & Starobinsky (1974) and Wald & Ramaswamy (1980).

If a singularity is allowed to form, the question of particle creation in its vicinity is complicated by the breakdown of predictability (Hawking & Ellis 1973). In particular, one does not know what boundary conditions to place on the quantum fields at the singularity. This problem occurs also in the cosmological case, as discussed in §7.4.

Hawking (1976*b*) has attempted to relate the randomness of black hole thermal emission to the boundary conditions at singularities by arguing that the creation of a particle pair just outside the horizon, with one particle going to infinity and its antiparticle 'tunnelling' into the hole eventually to strike the singularity, can be viewed in the spirit of Feynman's picture that treats an antiparticle as a particle travelling 'backwards in time'. The emitted particle can thereby be envisaged as originating on the singularity, travelling backwards in time to just outside the horizon, then scattering off the gravitational field into the future direction, to become part of the Hawking flux. In this way Hawking attributes the thermal character of the radiation to the singularity itself. Extending this idea to all singularities led to a new 'principle of ignorance' according to which the boundary conditions at a singularity should correspond to completely random influences emerging therefrom. This idea could have implications for cosmology as well as black holes.

9
Interacting fields

Once the theory of free quantum fields in curved spacetime had been worked out, the most natural extension was to include the effects of non-gravitational self and mutual interactions. Although this topic is still being developed, the basic framework is well established, and in this final chapter we outline the formal steps necessary for the computation of particle creation effects and the renormalization of $\langle T_{\mu\nu} \rangle$.

Two questions immediately spring to mind once interactions are included. The first is to what extent interactions can stimulate or inhibit particle creation by gravity over and above the free field case. Of course, interactions can lead to non-gravitational creation too, but we are more interested in processes that would be forbidden in Minkowski space, such as the simultaneous creation of a photon with an electron–positron pair.

The second question concerns renormalization theory. Will a field theory (e.g. Q.E.D.) that is renormalizable in Minkowski space remain so when the spacetime has a non-trivial topology or curvature? This question is of vital importance, for if a field theory is likely to lose its predictive power as soon as a small gravitational perturbation occurs, then its physical utility is suspect. It turns out to be remarkably difficult to establish general renormalizability, and significant progress has so far been limited to the so-called $\lambda\phi^4$ theory.

A third issue of great interest concerns black hole radiance. Is the Hawking flux precisely thermal even in the presence of field interactions? If not, a violation of the second law of thermodynamics seems possible. We discuss a particular model calculation in which the thermal character does indeed survive.

This chapter is intended only to introduce the reader to the topic of field interactions in curved space. No attempt is made at a comprehensive coverage, and only a few concrete examples are given.

9.1 Calculation of S-matrix elements

We consider a general interacting field theory defined by the Lagrangian density

$$\mathscr{L} = \mathscr{L}_0 + \mathscr{L}_I, \tag{9.1}$$

where $\mathscr{L}_0$ is a free field Lagrangian density as, for example, discussed in §§3.2 and 3.8, and $\mathscr{L}_I$ is an interaction Lagrangian density, containing field products of higher than quadratic order. We assume, as before, that the spacetime under consideration is globally hyperbolic, and that a choice of time parameter x^0 has been made.

It is further assumed that the interaction in (9.1) is switched off adiabatically in the distant past and future and that the fields, which will generically be denoted as ϕ, reduce to free fields ϕ_{in} and ϕ_{out} respectively in these regions:

$$\lim_{x^0 \to -\infty} \phi(x) = \phi_{\text{in}}(x), \tag{9.2}$$

$$\lim_{x^0 \to +\infty} \phi(x) = \phi_{\text{out}}(x), \tag{9.3}$$

where the limit is defined in the sense of weak operator convergence. These asymptotic conditions, which are necessary if particle states are to be defined, are justified in Minkowski space quantum field theory by consideration of a typical scattering situation in which the particles are initially and finally well separated, outside the effective range of interaction. It is only at intermediate times, as the particles approach one another, that the interaction becomes important. Such an argument can frequently be extended to curved spacetime. However, it is not difficult to envisage spacetimes in which the asymptotic conditions (9.2) and (9.3) cannot be justified. For example, in a spacetime with closed spatial sections, the particles can never be infinitely far apart, and the asymptotic conditions must be regarded as only approximate. In addition, it may be that the spacetime model of interest contains singularities which bound the space in the past or future, so that asymptotic time regions do no exist. In this case one may try appealing to other physical or mathematical criteria, or attempt an alternative formulation of quantum field theory (see, for example Kay 1980)

Since ϕ_{in} and ϕ_{out} are free fields, they can be treated in exactly the manner described in §3.2. In particular, they can be expanded in terms of complete sets of modes as in (3.30). As discussed extensively in chapter 3, the definition of positive frequency modes in the distant past and future need not necessarily agree, so let us assume that $u_{\mathbf{k}}^{\text{in}}$ and $u_{\mathbf{k}}^{\text{out}}$ are positive frequency modes in the far past and future, respectively, chosen, for example, using the adiabatic definition of positive frequency. Then ϕ_{in} and ϕ_{out} can be expanded in terms of either set of modes:

$$\phi_{in}(x) = \sum_i (a_i^{in} u_i^{in}(x) + a_i^{in\dagger} u_i^{in*}(x)) \tag{9.4a}$$

$$= \sum_i (\bar{a}_i^{in} u_i^{out}(x) + \bar{a}_i^{in\dagger} u_i^{out*}(x)) \tag{9.4b}$$

$$\phi_{out}(x) = \sum_i (a_i^{out} u_i^{in}(x) + a_i^{out\dagger} u_i^{in*}(x)) \tag{9.5a}$$

$$= \sum_i (\bar{a}_i^{out} u_i^{out}(x) + \bar{a}_i^{out\dagger} u_i^{out*}(x)). \tag{9.5b}$$

Since, in general, u_i^{in} will not be equal to u_i^{out}, but rather will be related by a Bogolubov transformation of the form (3.34), so too will a_i^{in} be related to $\bar{a}_i^{in}$, and a_i^{out} to $\bar{a}_i^{out}$ by Bogolubov transformations of the form (3.37). Thus, there will in general exist four inequivalent vacua defined by (cf. (2.19), (3.33))

$$a_i^{in}|0, \mathrm{in}\rangle = 0, \quad \forall i, \tag{9.6}$$

$$\bar{a}_i^{in}|\bar{0}, \mathrm{in}\rangle = 0, \quad \forall i, \tag{9.7}$$

$$a_i^{out}|0, \mathrm{out}\rangle = 0, \quad \forall i, \tag{9.8}$$

$$\bar{a}_i^{out}|\bar{0}, \mathrm{out}\rangle = 0, \quad \forall i, \tag{9.9}$$

each with associated Fock spaces. The Fock space based on $|0, \mathrm{in}\rangle$ will be related to that based on $|\bar{0}, \mathrm{in}\rangle$ by S-matrix elements such as those in (3.45), which can be calculated purely in terms of the Bogolubov coefficients. The $|0, \mathrm{out}\rangle$ and $|\bar{0}, \mathrm{out}\rangle$ based Fock spaces will also be related in such a manner. However the relationship between the Fock spaces based on $|0, \mathrm{in}\rangle$ and $|0, \mathrm{out}\rangle$ will depend on the interaction. If it is removed, these Fock spaces will be equivalent.

Since the vacuum $|0, \mathrm{in}\rangle$ is defined with respect to modes that are positive frequency in the distant past where the full interacting field has the form ϕ_{in}, it defines the 'physical' Fock space in the far past. Similarly, the vacuum $|\bar{0}, \mathrm{out}\rangle$ being defined with respect to modes which are positive frequency in the far future, where ϕ reduces to ϕ_{out}, defines the 'physical' Fock space in this region. It is the S-matrix elements relating the Fock space based on $|0, \mathrm{in}\rangle$ to that based on $|\bar{0}, \mathrm{out}\rangle$ that will describe the physics of the interactions of 'particles' in curved spacetime.

Notice that one is still faced with the problems discussed in chapter 3 of the meaning of particle states in curved spacetime, and it will only be in special cases, such as when the adiabatic definition of positive frequency, or conformal symmetry can be exploited, that the choice of u_i^{in} and u_i^{out} will be unambiguous.

Assuming that a definition of particle states has been chosen, then we are interested in computing the S-matrix elements (scattering amplitudes) between in and out states (see §3.2). An in state such as $|{}^1n_{i_1}, {}^2n_{i_2}, \ldots, \text{in}\rangle$ is likely to develop into an out state $|{}^1\bar{m}_{i_1}, {}^2\bar{m}_{i_2}, \ldots, \text{out}\rangle$ as a result of spacetime curvature and the interaction. It proves helpful to separate these two effects in the formalism that we shall give. This can be achieved by expanding the state $|{}^1\bar{m}_{i_1}, {}^2\bar{m}_{i_2}, \ldots, \text{out}\rangle$ in terms of a complete set of unbarred out states, after the fashion of (3.44). Then the relevant S-matrix element is

$$\begin{aligned}
&\langle \text{out}, \ldots, {}^2\bar{m}_{i_2}, {}^1\bar{m}_{i_1} | {}^1n_{i_1}, {}^2n_{i_2}, \ldots, \text{in}\rangle \\
&= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j_1 \ldots j_k} \langle \text{out}, \ldots, {}^2\bar{m}_{i_2}, {}^1\bar{m}_{i_1} | 1_{j_1}, 1_{j_2}, \ldots, 1_{j_k}, \text{out}\rangle \\
&\quad \times \langle \text{out}, 1_{j_k}, \ldots, 1_{j_2}, 1_{j_1} | {}^1n_{i_1}, {}^2n_{i_2}, \ldots, \text{in}\rangle. \qquad (9.10)
\end{aligned}$$

Wald (1979a, b) discusses the existence of the S-matrix in curved spacetime. The probability for a transition between these particular in and out states is given by

$$|\langle \text{out}, \ldots, {}^2\bar{m}_{i_2}, {}^1\bar{m}_{i_1} | {}^1n_{i_1}, {}^2n_{i_2}, \ldots, \text{in}\rangle|^2. \qquad (9.11)$$

The amplitude $\langle \text{out}, \ldots, {}^2\bar{m}_{i_2}, {}^1\bar{m}_{i_1} | 1_{j_1}, 1_{j_2}, \ldots, 1_{j_k}, \text{out}\rangle$ is given entirely in terms of Bogolubov coefficients (see (3.44)), and is independent of details of the interaction. On the other hand, the amplitude $\langle \text{out}, 1_{j_k}, \ldots, 1_{j_2}, 1_{j_1} | {}^1n_{i_1}, {}^2n_{i_2}, \ldots, \text{in}\rangle$ depends on the interaction, but is an S-matrix element formed from elements of two Fock spaces defined using the same definition (u_i) of positive frequency modes. It can thus be calculated using methods familiar in Minkowski space quantum field theory. We shall now discuss two such methods.

One technique is the so-called LSZ (Lehmann, Symanzik & Zimmermann 1955) method of reducing S-matrix elements to expressions given in terms of Green functions. The derivation of these reduction formulae proceeds essentially as in Minkowski space (see, for example Bjorken & Drell 1965, §16.7). In particular, for a self-interacting scalar field with $\mathscr{L}_0$ given by (3.24), one obtains (cf. Bjorken & Drell, equation (16.81))

$$\begin{aligned}
&\langle \text{out}, 1_{p_1}, \ldots, 1_{p_l} | 1_{q_1}, \ldots, 1_{q_m}, \text{in}\rangle / \langle \text{out}, 0 | 0, \text{in}\rangle \\
&= \mathrm{i}^{m+l} \prod_{i=1}^{m} \int \mathrm{d}^n x_i [-g(x_i)]^{\frac{1}{2}} \prod_{j=1}^{l} \int \mathrm{d}^n y_i [-g(y_i)]^{\frac{1}{2}} u_{q_i}(x_i) u^*_{p_j}(y_j) \\
&\quad \times [\Box_{x_i} + m^2 + \xi R(x_i)][\Box_{y_j} + m^2 + \xi R(y_j)] \tau(y_1 \ldots y_l, x_1 \ldots x_m), \qquad (9.12)
\end{aligned}$$

where it has been assumed that $p_i \neq q_j, \forall i,j$, and we have defined the Green function

$$\tau(x_1, x_2 \ldots x_m) = \frac{\langle \text{out}, 0|T(\phi(x_1)\phi(x_2)\ldots\phi(x_m))|0, \text{in}\rangle}{\langle \text{out}, 0|0, \text{in}\rangle}. \tag{9.13}$$

Details of the derivation of equations such as (9.12) are given in Birrell (1979*c*) and Birrell & Taylor (1980), where reduction formulae for the full amplitude (9.11) in terms of $\langle \text{out}, \bar{0}|\ldots|0, \text{in}\rangle$ Green functions and Bogolubov coefficients have been given.

There are two points in the derivation of (9.12) which warrant special mention. The first is that, unlike the case of Minkowski space where $|0, \text{out}\rangle = |0, \text{in}\rangle$ (up to a phase factor) the vacuum $|0, \text{in}\rangle$ in curved spacetime will not in general be stable: $\langle \text{out}, 0|0, \text{in}\rangle \neq 1$. This possible instability is not due to different definitions of positive frequency ($|0, \text{out}\rangle$ and $|0, \text{in}\rangle$ are based on the same definition), but rather to the lack of Poincaré invariance in curved spacetime, which can give rise to particle production additional to that discussed in chapter 3. This topic will form the subject of §9.3.

The second comment to be made on the derivation of (9.12) concerns the fact that in arriving at this equation Gauss' theorem has been used, with surface terms at spacelike infinity being discarded. If the spacetime has spacelike boundaries, it will be necessary to retain these surface terms and apply physically motivated boundary conditions to them.

The Green functions (9.13) must now be computed. Exact expressions are almost impossible to obtain in a practical calculation, and some approximation scheme, such as perturbation theory, must be used. This may be implemented by first constructing the so-called evolution matrix U (Dyson 1949; see, for example, Bjorken & Drell 1965, chapter 17), which leads to a perturbation series for S which is essentially the same as the interaction picture approach to be described shortly.

An alternative technique is to use the path-integral formulation (§2.8), and work with the Green function generating functional which in the interacting case is given by

$$Z[J] = \int \mathscr{D}[\phi] \exp\left\{\mathrm{i}\int \mathscr{L}_{\mathrm{I}}[\phi]\mathrm{d}^n x\right\} \exp\left\{\mathrm{i}\int [\mathscr{L}_0[\phi] + J\phi]\mathrm{d}^n x\right\}. \tag{9.14}$$

If the interaction Lagrangian density is a polynomial in the fields, then (9.14) can be rewritten as

$$Z[J] = \exp\left\{\mathrm{i}\int \mathscr{L}_{\mathrm{I}}\left[\frac{1}{\mathrm{i}}\frac{\delta}{\delta J}\right]\mathrm{d}^n x\right\}\int \mathscr{D}[\phi] \exp\left\{\mathrm{i}\int [\mathscr{L}_0[\phi] + J\phi]\mathrm{d}^n x\right\}. \tag{9.15}$$

The functional integral in (9.15) is now simply the free generating functional, which we shall denote by $Z_0[J]$. Thus

$$Z[J] = \exp\left\{i \int \mathscr{L}_I\left[\frac{1}{i}\frac{\delta}{\delta J}\right] d^n x\right\} Z_0[J]. \tag{9.16}$$

For scalar fields, Z_0 is given by (2.126), which also holds in curved spacetime. Hence, in this case,

$$Z[J] \propto [\det(-G_F)]^{\frac{1}{2}} \exp\left\{i \int \mathscr{L}_I\left[\frac{1}{i}\frac{\delta}{\delta J(x)}\right] d^n x\right\}$$

$$\times \exp\left\{-\tfrac{1}{2}i \int J(y) G_F(y, z) J(z) d^n y\, d^n z\right\} \tag{9.17}$$

where the proportionality constant is independent of the metric and J, and does not affect the Green functions. By expanding the exponentials in (9.17) and performing the functional differentiations one obtains a perturbation expansion for Z entirely in terms of the free field Feynman propagator. This expansion can then be used in (2.117) to generate expansions for the Green functions.

It is important to note that the boundary conditions that are encoded in the interacting Green functions generated by (9.17) will depend on the choice of Feynman propagator G_F. Since we wish to generate the Green functions (9.13), which are defined with respect to only one definition of positive frequency modes u_i, the associated G_F must be calculated from these modes:

$$iG_F(x, y) = \theta(x^0 - y^0)\sum_i u_i(x)u_i^*(y) + \theta(y^0 - x^0)\sum_i u_i^*(x)u_i(y) \tag{9.18}$$

(cf. (7.74)). With this choice we have

$$\tau_c(x_1, x_2, \ldots, x_m) = i^{-m}\left[\frac{\delta^m \ln Z[J]}{\delta J(x_1)\delta J(x_2)\ldots\delta J(x_m)}\right]_{J=0}, \tag{9.19}$$

where the subscript 'c' on τ_c means that only connected Feynman diagrams will be generated by the use of the perturbation expansion (9.17) in (9.19), as may be explicitly verified for specific interaction Lagrangians. We shall consider a particular example in the next section.

One method of studying interacting quantum fields in curved spacetime which has been used for some time (Utiyama 1962, Freedman & Pi 1975), is the expansion of the generating functional (or Green functions) in a functional Taylor series (Volterra series) in the metric about Minkowski space. Each term in this series can then be examined using ordinary Minkowski space perturbation theory in the interaction coupling constant.

Rather than expand $Z[J]$ itself, it is more convenient, in practice, to expand $\ln Z$, which appears in (9.19). Writing (cf. (6.25))

$$W[J;\, g_{\mu\nu}] = -\mathrm{i}\ln Z[J;g_{\mu\nu}], \tag{9.20}$$

where

$$g_{\mu\nu}(x) = n_{\mu\nu} + h_{\mu\nu}(x) \tag{9.21}$$

we can expand W as a Volterra series (see, for example, Rzewuski 1969, §I, 2):

$$W[J;g_{\mu\nu}] = \sum_{m=0}^{\infty} \frac{1}{m!}\left(\prod_{i=1}^{m} \mathrm{d}^n x_i h^{\mu_i\nu_i}(x_i) W_{\mu_1\nu_1,\mu_2\nu_2,\ldots,\mu_m\nu_m}[J;x_1,x_2,\ldots,x_m]\right), \tag{9.22}$$

with

$$W_{\mu_1\nu_1,\mu_2\nu_2,\ldots,\mu_m\nu_m}[J;x_1,x_2,\ldots,x_m] = \left[\frac{\delta^m W[J;g_{\mu\nu}]}{\delta g^{\mu_1\nu_1}(x_1)\delta g^{\mu_2\nu_2}(x_2)\ldots\delta g^{\mu_m\nu_m}(x_m)}\right]_{g^{\mu_i\nu_i} = \eta^{\mu_i\nu_i}}. \tag{9.23}$$

Setting $J = 0$ in (9.22) gives an expansion for the effective action for the interacting theory, and we see from (6.13) that the first order term in this expansion involves

$$W_{\mu_1\nu_1}[0] = \langle 0|T_{\mu_1\nu_1}|0\rangle, \tag{9.24}$$

the vacuum expectation value of the Minkowski space stress-tensor for the interacting theory. This quantity has been studied by several authors (Callan, Coleman & Jackiw 1970, Freedman, Muzinich & Weinberg 1974, Freedman & Weinberg 1974, Collins 1976). The higher order terms are more complicated, involving not only terms such as $\langle 0|T_{\mu_1\nu_1}T_{\mu_2\nu_2}\cdots T_{\mu_m\nu_m}|0\rangle$, but also the so-called 'seagull' contributions arising from the functional differentiation of $T_{\mu\nu}[g_{\mu\nu}]$ (Freedman & Pi 1975).

The disadvantages of treating interacting theories in curved spacetime by this method are three-fold: (i) Calculationally it is only useful for small deviations from Minkowski space. (ii) It gives no information about the effects of spacetime topology (see, for example, Drummond & Hathrell 1980). (iii) It is not possible, without great difficulty, to prescribe the vacuum state which is used in the calculation. We shall not, therefore, pursue this approach here, but, when using the path-integral formulation, shall consider only perturbation series in the interaction coupling strength, as given by (9.17).

As already mentioned, there is a second technique, which can easily be adapted from Minkowski space, for the calculation of the amplitudes $\langle \text{out}, {}^k j_k, \ldots, {}^2 j_2, {}^1 j_1 | {}^1 n_{i_1}, {}^2 n_{i_2}, \ldots, \text{in} \rangle$ appearing in (9.11). This is the interaction picture approach in which the fields satisfy the free equations, and the dynamical information is carried by the states of the system. (See, for example, Roman 1969, chapter 4, for a covariant treatment in Minkowski space. Formulation of the interaction picture method in curved spacetime has been given by Birrell & Ford 1979, and Bunch, Panangaden & Parker 1980.)

In the interaction picture, the states $|\psi\rangle$ satisfy the Schrödinger equation

$$\mathscr{H}_{\text{I}}(x)|\psi[\Sigma]\rangle = \mathrm{i}\frac{\delta|\psi[\Sigma]\rangle}{\delta\Sigma(x)}, \tag{9.25}$$

where $\mathscr{H}_{\text{I}}$ is the interaction Hamiltonian density, which, for non-derivative coupling, is given by

$$\mathscr{H}_{\text{I}} = -\mathscr{L}_{\text{I}}, \tag{9.26}$$

and $\Sigma(x)$ is a spacelike Cauchy hypersurface through x. (For constant time slicing, (9.25) becomes

$$\mathscr{H}_{\text{I}}(x)|\psi(x^0)\rangle = \mathrm{i}\frac{\partial|\psi(x^0)\rangle}{\partial x^0}.) \tag{9.27}$$

The solution to equation (9.25) is determined in terms of a unitary operator U, defined by

$$|\psi[\Sigma]\rangle = U[\Sigma, \Sigma_0]|\psi[\Sigma_0]\rangle, \tag{9.28}$$

which, using (9.25), is seen to satisfy the so-called Tomonaga–Schwinger equation (Tomonaga 1946, Schwinger 1948, 1949*a*, *b*)

$$\mathscr{H}_{\text{I}}(x)U[\Sigma, \Sigma_0] = \mathrm{i}\frac{\delta U[\Sigma, \Sigma_0]}{\delta\Sigma(x)}, \tag{9.29a}$$

with initial condition

$$U[\Sigma_0, \Sigma_0] = 1. \tag{9.29b}$$

Writing (9.29) as an integral equation

$$U[\Sigma, \Sigma_0] = 1 - \mathrm{i}\int_{\Sigma_0}^{\Sigma} \mathscr{H}_{\text{I}}(x')U[\Sigma', \Sigma_0]\mathrm{d}^n x', \tag{9.30}$$

the following closed form solution is obtained by iteration:

$$U[\Sigma, \Sigma_0] = P\exp\left[-\mathrm{i}\int_{\Sigma_0}^{\Sigma} \mathscr{H}_{\mathrm{I}}(x')\mathrm{d}^n x'\right], \tag{9.31}$$

where the ordering symbol P is the same as the time-ordering symbol T, except that no sign changes are made under transpositions in the case of fermion fields.

A theorem due to Haag (1955) implies that U cannot be a well-defined operator on the Hilbert space of states (see, for example, Roman 1969, §8.4), so we consider (9.31) as purely a formal expression for the purpose of computing the S-matrix operator defined by

$$S = U[\Sigma^{\mathrm{out}}, \Sigma^{\mathrm{in}}]. \tag{9.32}$$

The surface Σ^{out} lies in the out region in the infinitely far future, while Σ^{in} is a surface in the in region in the infinitely remote past. Then, from (9.28), one has

$$|\psi[\Sigma^{\mathrm{out}}]\rangle = S|\psi[\Sigma^{\mathrm{in}}]\rangle, \tag{9.33}$$

and, from (9.31), S possesses the perturbation expansion

$$S = \sum_{m=0}^{\infty} S^{(m)} \tag{9.34}$$

with

$$S^{(0)} = 1 \tag{9.35}$$

$$S^{(m)} = \frac{(-\mathrm{i})^m}{m!}\int P(\mathscr{H}_{\mathrm{I}}(x_1)\mathscr{H}_{\mathrm{I}}(x_2)\ldots\mathscr{H}_{\mathrm{I}}(x_m))\,\mathrm{d}^n x_1\,\mathrm{d}^n x_2\ldots\mathrm{d}^n x_m, \tag{9.36}$$

Let us now choose $|\psi[\Sigma^{\mathrm{in}}]\rangle$ to be a member of the Fock space based on the Heisenberg vacuum $|0, \mathrm{in}\rangle$ (see(9.6)). This immediately implies that on the surface Σ^{in}, the interaction picture field ϕ, and the Heisenberg picture field ϕ must be equal. But on Σ^{in}, the Heisenberg field reduces to the free field ϕ_{in} (see(9.2)). So therefore does the interaction picture field. Since the interaction picture field obeys the free field equation, it follows that it will be equal to ϕ_{in} for all time.

Consider, in particular, the choice of state

$$|\psi[\Sigma^{\mathrm{in}}]\rangle = |1_{j_1}, 1_{j_2}, \ldots, 1_{j_k}, \mathrm{in}\rangle. \tag{9.37}$$

Then (9.33) becomes

$$|1_{j_1}, 1_{j_2}, \ldots, 1_{j_k}, \mathrm{out}\rangle = S|1_{j_1}, 1_{j_2} \ldots . 1_{j_k}, \mathrm{in}\rangle. \tag{9.38}$$

Note that, because S represents only the effects of interaction, $|\psi[\Sigma^{\text{out}}]\rangle$ as given by the left-hand side of (9.38), is defined in terms of the unbarred out states. We therefore possess a second means of computing that part of the amplitude (9.10) independent of the Bogolubov transformation. From (9.38):

$$\langle \text{out}, 1_{j_k}, \dots, 1_{j_2}, 1_{j_1} | {}^1n_{i_1}, {}^2n_{i_2}, \dots, \text{in}\rangle$$
$$= \langle \text{in}, 1_{j_k}, \dots 1_{j_2}, 1_{j_1} | S^\dagger | {}^1n_{i_1}, {}^2n_{i_2}, \dots, \text{in}\rangle. \tag{9.39}$$

The right-hand side of (9.39) can be calculated to any order of perturbation theory, by using (9.34), with $\mathscr{H}_\text{I}$ in (9.36) constructed from ϕ_{in} as given by (9.4).

Whether one uses the interaction picture approach or reduction formulae and path-integral quantization, as in Minkowski space, one encounters infinities which must be removed by renormalization. The application of the computational methods defined above, and the regularization and renormalization of the resulting infinite quantities are best described by a specific example, which we shall now give.

9.2 Self-interacting scalar field in curved spacetime

We consider a self-interacting scalar field theory with Lagrangian density (9.1); $\mathscr{L}_0$ being given by (3.24), and $\mathscr{L}_\text{I}$ by

$$\mathscr{L}_\text{I} = -\frac{1}{4!}(-g)^{\frac{1}{2}}\lambda\phi^4. \tag{9.40}$$

This is the so-called $\lambda\phi^4$ theory.

Of particular interest are amplitudes representing scattering from an initial vacuum to a final many-particle state, since these processes cannot occur in Minkowski space. The examination of such amplitudes is sufficient to demonstrate the renormalization techniques required in the calculation of other amplitudes. Setting $|{}^1n_{i_1}, {}^2n_{i_2} \dots, \text{in}\rangle = |0, \text{in}\rangle$ in (9.10), one obtains

$$\langle \text{out}, \dots, {}^2\bar{m}_{i_2}, {}^1\bar{m}_{i_1} | 0, \text{in}\rangle$$
$$= \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{j_1 \dots j_k} \langle \text{out}, \dots, {}^2\bar{m}_{i_2}, {}^1\bar{m}_{i_1} | 1_{j_1}, 1_{j_2}, \dots, 1_{j_k}, \text{out}\rangle$$
$$\times \langle \text{out}, 1_{j_k}, \dots, 1_{j_2}, 1_{j_1} | 0, \text{in}\rangle. \tag{9.41}$$

The first amplitude on the right-hand side is easily constructed from Bogolubov coefficients and, in particular, is finite. We thus consider the first few contributions from the second amplitude.

To illustrate in detail the two calculational methods described in the previous section, we shall first describe the perturbation theory calculation of $\langle \text{out}, 1_{p_1}, 1_{p_2} | 0, \text{in} \rangle$. Starting with the LSZ method, we note that the reduction formula (9.12) gives for this amplitude

$$\frac{\langle \text{out}, 1_{p_1}, 1_{p_2} | 0, \text{in} \rangle}{\langle \text{out}, 0 | 0, \text{in} \rangle}$$

$$= \int \mathrm{d}^n y_1 \, \mathrm{d}^n y_2 [-g(y_1)]^{\frac{1}{2}} [-g(y_2)]^{\frac{1}{2}} u^*_{p_1}(y_1) u^*_{p_2}(y_2)$$

$$\times \mathrm{i}[\Box_{y_1} + m^2 + \xi R(y_1)] \mathrm{i}[\Box_{y_2} + m^2 + \xi R(y_2)] \tau(y_1, y_2). \quad (9.42)$$

We thus need to calculate the Green function $\tau(y_1, y_2)$, which is also known as the complete propagator (for the interacting theory).

Substituting (9.40) into (9.17), and expanding the exponentials, it is not difficult to compute, to a given order in λ, the terms which will contribute only two factors of J to the expression of $\ln Z[J]$. It is these terms which will give non-vanishing contributions to

$$\tau_c(y_1, y_2) = -\left[\frac{\delta^2 \ln Z[J]}{\delta J(y_1) J(y_2)}\right]_{J=0} \quad (9.43)$$

(see (9.19)). Disconnected contributions to τ can be calculated by simply combining together connected components, and we shall not discuss them further except to say that the disconnected components will contribute to the total amplitude.

To first order in λ, use of (9.17) and (9.43) gives

$$\tau_c(y_1, y_2) = \mathrm{i} G_F(y_1, y_2)$$

$$-\tfrac{1}{2}\lambda \int G_F(y_1, x) G_F(x, x) G_F(x, y_2) [-g(x)]^{\frac{1}{2}} \mathrm{d}^n x \quad (9.44)$$

which reduces to i $G_F(y_1, y_2)$ when $\lambda = 0$, as one would expect (cf. (9.13) and (2.69)). It is convenient to introduce Feynman diagrams to represent expressions such as (9.44). Representing λ by a vertex in the diagram, and i$G_F(x, y)$ by a line, the diagrammatic form of (9.44) is given in fig. 29, where integration over the point at a vertex is understood.

The quantity involving the Green function that actually appears in the reduction formula (9.42) is

$$\gamma(y_1, y_2) \equiv \mathrm{i}[\Box_{y_1} + m^2 + \xi R(y_1)] \mathrm{i}[\Box_{y_2} + m^2 + \xi R(y_2)] \tau(y_1, y_2). \quad (9.45)$$

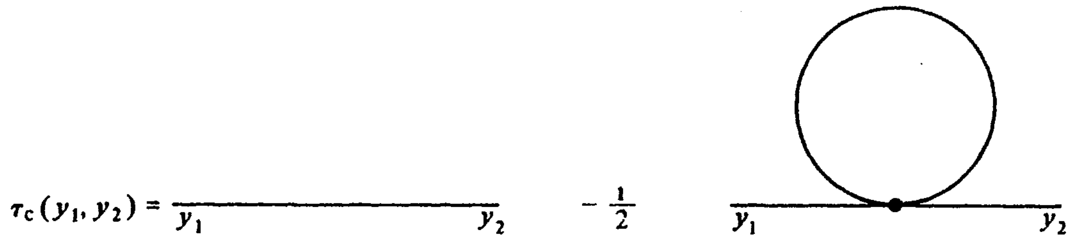


Fig. 29. Feynman diagram of the zeroth and first order contributions to the complete, connected propagator $\tau_c(y_1, y_2)$ for $\lambda\phi^4$ theory.

Substituting (9.44) into (9.45) and using (3.49), one obtains

$$\gamma_c(y_1, y_2) = \mathrm{i}K_{y_1 y_2} + \tfrac{1}{2}\lambda G_F(y_1, y_1)\delta(y_1 - y_2)/[-g(y_1)]^{\frac{1}{2}}, \quad (9.46)$$

where K is defined by (6.21) or (6.23). The function γ is known as the *amputated Green function*, because its diagrammatic representation is obtained from that of τ_c in fig. 29 by amputation of the external 'legs'. Finally, substituting (9.46) into (9.42), one finds

$$\frac{\langle \text{out}, 1_{p_1}, 1_{p_2}|0, \text{in}\rangle}{\langle \text{out}, 0|0, \text{in}\rangle} = \tfrac{1}{2}\lambda \int u^*_{p_1}(y)u^*_{p_2}(y)G_F(y, y)[-g(y)]^{\frac{1}{2}}\mathrm{d}^n y + O(\lambda^2). \quad (9.47)$$

The first term in (9.46) does not contribute to (9.47) because the modes satisfy the free field equation.

Before discussing this result, higher order corrections and other amplitudes, we show how it can be reproduced using the interaction picture approach. One starts from (9.39), which in this case gives

$$\langle \text{out}, 1_{p_1}, 1_{p_2}|0, \text{in}\rangle = \langle \text{in}, 1_{p_1}, 1_{p_2}|S^\dagger|0, \text{in}\rangle, \quad (9.48)$$

and, to obtain a result to first order in λ, one substitutes the first two terms of (9.34) for S. The first term does not contribute, and one is left with (using (9.26) and (9.40))

$$\langle \text{out}, 1_{p_1}, 1_{p_2}|0, \text{in}\rangle$$
$$= -\frac{\mathrm{i}\lambda}{4!}\int \langle \text{in}, 1_{p_1}, 1_{p_2}|T(\phi^4(x))|0, \text{in}\rangle[-g(x)]^{\frac{1}{2}}\mathrm{d}^n x + O(\lambda^2). \quad (9.49)$$

The matrix element in the integrand of (9.49) can be evaluated by substituting (9.4*a*) for ϕ (recall $\phi = \phi_{\text{in}}$ in the interaction picture), or, alternatively, using Wick's theorem (Wick 1950; see, for example, Roman 1969, chapter 4). Either way one obtains

$$\langle \text{in}, 1_{p_1}, 1_{p_2}|T(\phi^4(x))|0, \text{in}\rangle = 12u^*_{p_1}(x)u^*_{p_2}(x)\mathrm{i}G_F(x, x)\langle \text{in}, 0|0, \text{in}\rangle, \quad (9.50)$$

where G_F is given by (9.18). Substituting this into (9.49) and noting that

$$\langle \text{out}, 0|0, \text{in}\rangle = \langle \text{in}, 0|0, \text{in}\rangle + O(\lambda), \tag{9.51}$$

agreement with (9.47) is obtained. This method of calculating amplitudes generates (at higher orders) disconnected as well as connected diagrams. Dividing through by (9.51) removes the disconnected 'bubble diagrams' with no external legs (see, for example, Bjorken & Drell 1965, §17.6; or Schweber 1961, §14b).

Inspection of (9.47) immediately reveals a problem; $G_F(y, y)$ is infinite as $n \to 4$. Indeed, we have kept n arbitrary in (9.47) in anticipation of the need to regularize the expression; a procedure most easily carried out using dimensional regularization. The nature of the pole terms can be determined using the DeWitt–Schwinger expansion for G_F. From (3.138), or (6.29), one obtains

$$G_F(y, y) \approx -\frac{i}{(4\pi)^{n/2}} \sum_{j=0}^{\infty} (m^2)^{\frac{1}{2}n - j - 1} a_j(y) \Gamma(j - \tfrac{1}{2}n + 1)$$

$$= -\frac{2i}{(4\pi)^2} \frac{[m^2 + (\xi - \frac{1}{6})R]}{(n-4)} + G_F^{\text{finite}}(y, y) \tag{9.52}$$

where $G_F^{\text{finite}}(y, y)$ is finite as $n \to 4$. The pole terms must be removed by renormalization of constants in the Lagrangian, as in Minkowski space theory. However, there is an evident difference, i.e., the appearance of a pole term proportional to the Ricci scalar R. We shall see that this extra pole term can be removed by renormalization of the constant ξ, which is absent in Minkowski space theory.

A method of systematically performing renormalization in dimensional regularization has been devised by 't Hooft (1973) and studied by Collins & Macfarlane (1974), and Collins (1974), who carries out the explicit second order renormalization of $\lambda\phi^4$ in Minkowski space. (See also the textbook treatment of Nash 1978.) The total Lagrangian (9.1) is

$$\mathscr{L} = \tfrac{1}{2}(-g)^{\frac{1}{2}}[g^{\mu\nu}\phi_{,\mu}\phi_{,\nu} - (m_R^2 + \xi_R R(x))\phi_R^2]$$

$$-(-g)^{\frac{1}{2}}\left[\frac{1}{4!}\lambda\phi^4 + \tfrac{1}{2}(\delta m^2 + \delta\xi R)\phi^2\right], \tag{9.53}$$

where m_R and ξ_R are 'renormalized constants' related to the 'bare constants' m and ξ by

$$m_R^2 = m^2 - \delta m^2 \tag{9.54}$$

$$\xi_{\mathrm{R}} = \xi - \delta\xi. \tag{9.55}$$

For convenience we have not used a subscript B for the bare quantities. We now treat the first term in (9.53) involving renormalized quantities as a free Lagrangian density $\mathscr{L}_0$, and the second term as a new interaction Lagrangian density. Our aim is to cancel pole terms appearing in the S-matrix by appropriate choices of δm^2 and $\delta\xi$. In anticipation of this we write

$$\delta m^2 = m_{\mathrm{R}}^2 \sum_{\nu=1}^{\infty} \sum_{j=\nu}^{\infty} b_{\nu j} \lambda_{\mathrm{R}}^j (n-4)^{-\nu} \equiv m_{\mathrm{R}}^2 \sum_{\nu=1}^{\infty} b_\nu(\lambda_{\mathrm{R}})(n-4)^{-\nu} \tag{9.56}$$

$$\delta\xi = \sum_{\nu=1}^{\infty} \sum_{j=\nu}^{\infty} d_{\nu j} \lambda_{\mathrm{R}}^j (n-4)^{-\nu} \equiv \sum_{\nu=1}^{\infty} d_\nu(\lambda_{\mathrm{R}})(n-4)^{-\nu} \tag{9.57}$$

where we also allow for the possibility of coupling constant renormalization by writing

$$\begin{aligned} \lambda = \lambda_{\mathrm{R}} + \delta\lambda &= \mu^{4-n}\left[\lambda_{\mathrm{R}} + \sum_{\nu=1}^{\infty} \sum_{j=\nu}^{\infty} a_{\nu j} \lambda_{\mathrm{R}}^j (n-4)^{-\nu}\right] \\ &= \mu^{4-n}\left[\lambda_{\mathrm{R}} + \sum_{\nu=1}^{\infty} a_\nu(\lambda_{\mathrm{R}})(n-4)^{-\nu}\right]. \end{aligned} \tag{9.58}$$

As in §6.2, an arbitrary mass scale μ has been introduced into (9.58) so as to maintain the dimensionless nature of the total action.

One must also take into account a possible renormalization of the fields ('wavefunction' renormalization) in which

$$\phi \rightarrow \phi_{\mathrm{R}} = Z^{-\frac{1}{2}} \phi, \tag{9.59}$$

where

$$Z = 1 + \sum_{\nu=1}^{\infty} \sum_{j=\nu}^{\infty} c_{\nu j} \lambda_{\mathrm{R}}^j (n-4)^{-\nu} \equiv \sum_{\nu=1}^{\infty} c_\nu(\lambda_{\mathrm{R}})(n-4)^{-\nu}. \tag{9.60}$$

The theory is renormalizable if all pole terms in S-matrix elements can be removed by appropriate choices of $a_{\nu j}, b_{\nu j}, c_{\nu j}$ and $d_{\nu j}$.

If the second term in (9.53) is used as the interaction Lagrangian i.e.,

$$\mathscr{L}_1 = -(-g)^{\frac{1}{2}}\left[\frac{1}{4!}\lambda\phi^4 + \tfrac{1}{2}(\delta m^2 + \delta\xi R)\phi^2\right], \tag{9.61}$$

then the amputated Green function γ_c is easily computed to be

$$\gamma_c(y_1, y_2) = \mathrm{i}K_{y_1 y_1} + \lambda_R \left\{ \tfrac{1}{2}\mu^{4-n} G_F(y_1, y_1) \right.$$

$$\left. - \mathrm{i} \sum_{\nu=1}^{\infty} [m_R^2 b_{\nu 1} + d_{\nu 1} R(y_1)](n-4)^{-\nu} \right\}$$

$$\times [-g(y_1)^{-\frac{1}{2}} \delta^n(y_1 - y_2)] + O(\lambda_R^2), \tag{9.62}$$

in place of (9.46). In obtaining this expression we have used (9.56)–(9.58), and have only retained terms up to order λ_R. All quantities in (9.62) are expressed in terms of the renormalized constants m_R^2 and ξ_R, rather than the bare parameters as before. Since $\tau_c(y_1, y_2)$ is the vacuum expectation value of two fields ϕ (see (9.13)), we expect to have to renormalize the fields by multiplying τ_c by Z^{-2}. However, in forming the amputated Green function, two factors of G_F, and thus a factor Z^{-4}, are removed. It follows that (9.62) should be multiplied by Z^2. The only contribution at order λ_R to S-matrix elements that this multiplication will make comes from the first term on the right in (9.62), resulting in

$$\lambda_R \sum_{\nu=1}^{\infty} c_{\nu 1}(n-4)^{-\nu} \mathrm{i} K_{y_1 y_2}$$

$$= \mathrm{i}\lambda_R \sum_{\nu=1}^{\infty} c_{\nu 1}(n-4)^{-\nu} [\Box_{y_1} + m_R^2 + \xi_R R(y_1)][-g(y_1)]^{-\frac{1}{2}} \delta^n(y_1 - y_2). \tag{9.63}$$

Substituting (9.52) (with the replacements $m \to m_R$, $\xi \to \xi_R$) into (9.62), one observes that all of the pole terms in $\gamma_c(y_1, y_2)$ can be removed by the choice

$$\left. \begin{aligned} &c_{\nu 1} = 0, \forall \nu; \quad d_{\nu 1} = b_{\nu 1} = 0, \ \forall \nu \neq 1; \\ &b_{11} = -1/16\pi^2; \quad d_{11} = -(\xi_R - \tfrac{1}{6})/16\pi^2. \end{aligned} \right\} \tag{9.64}$$

With this choice, the use of (9.62) in the reduction formulae now yields a finite amplitude as $n \to 4$:

$$\left[\frac{\langle \mathrm{out}, 1_{p_1}, 1_{p_2} | 0, \mathrm{in} \rangle}{\langle \mathrm{out}, 0 | 0, \mathrm{in} \rangle} \right]_{\mathrm{ren}}$$

$$= \tfrac{1}{2}\lambda_R \int u_{p_1}^*(y) u_{p_2}^*(y) \{ G_F^{\mathrm{finite}}(y, y)$$

$$+ (\mathrm{i}/8\pi^2)[m_R^2 + (\xi_R - \tfrac{1}{6})R(y)] \ln \mu \} [-g(y)]^{\frac{1}{2}} \mathrm{d}^4 y. \tag{9.65}$$

The appearance of the arbitrary mass scale μ in this amplitude expresses the renormalization ambiguity inherent in any renormalization scheme. Rescaling of μ readjusts the relationship between the bare and renormalized constants, and reflects the fact that measurements must be made to fix the values of λ_R, m_R and ξ_R. (See the discussion on page 163, where this point was discussed in connection with stress-tensor renormalization. We shall return to this issue below.)

Evaluation of other many-particle amplitudes proceeds similarly (Birrell 1980, Birrell & Ford 1980, Bunch, Panangaden & Parker 1980, Bunch & Panangaden 1980, Bunch & Parker 1979, Bunch 1980*b*). Of these only the four-particle amplitude is divergent, and its renormalization along the above lines fixes the coefficients a_{vj} in (9.58). Therefore, the renormalizations of the two- and four-particle amplitudes have fixed all the coefficients in (9.56)–(9.58) and (9.60), and hence in the scalar field Lagrangian. However, as we shall see, it turns out that there exists a divergence in the vacuum-to-vacuum amplitude also.

To see how this further divergence can be removed by renormalization, recall that the vacuum-to-vacuum amplitude is related to the effective action $W[g_{\mu\nu}] = W[0; g_{\mu\nu}]$(see (9.20)) by

$$W[g_{\mu\nu}] = -\mathrm{i}\ln(\langle \mathrm{out}, 0|0, \mathrm{in}\rangle) = -\mathrm{i}\ln Z[0]. \tag{9.66}$$

Hence one expects that this additional divergence can be removed by renormalization of constants in the generalized Einstein action, just as in the case of the free field effective action discussed in §6.2. That this is indeed the case has been verified to second order in λ_R by Bunch & Panangaden (1980), while similar renormalizations have been found necessary in massless quantum electrodynamics in de Sitter space (Shore 1980*a*, *b*; see also Drummond & Hathrell 1980, Brown & Collins 1980).

We treat here the renormalization of the vacuum-to-vacuum amplitude to first order in λ_R. Working in the interaction picture, we have

$$\begin{aligned}\langle \mathrm{out}, 0|0, \mathrm{in}\rangle &= \langle \mathrm{in}, 0|S|0, \mathrm{in}\rangle \\ &= \langle \mathrm{in}, 0|0, \mathrm{in}\rangle + \langle \mathrm{in}, 0|S^{(1)}|0, \mathrm{in}\rangle + O(\lambda_R^2),\end{aligned} \tag{9.67}$$

where we have used (9.34). Substituting this into (9.66) and expanding the logarithm gives

$$W[g_{\mu\nu}] = -\mathrm{i}\ln(\langle \mathrm{in}, 0|0, \mathrm{in}\rangle) - \mathrm{i}\frac{\langle \mathrm{in}, 0|S^{(1)}|0, \mathrm{in}\rangle}{\langle \mathrm{in}, 0|0, \mathrm{in}\rangle} + O(\lambda_R^2). \tag{9.68}$$

Notice that we have not assumed that $\langle \mathrm{in}, 0|0, \mathrm{in}\rangle$ is normalized to unity. This is merely a convenience; if it is taken to have its un-normalized value

$$\langle \mathrm{in}, 0|0, \mathrm{in}\rangle = [\det(-G_F)] \,, \tag{9.69}$$

given by (9.17) (with no interaction), then

$$W_0 \equiv -\mathrm{i}\ln(\langle \mathrm{in}, 0|0, \mathrm{in}\rangle) \tag{9.70}$$

and when used in (6.13) W_0 will yield $\langle \text{in}, 0|T_{\mu\nu}|0, \text{in}\rangle / \langle \text{in}, 0|0, \text{in}\rangle$.

However, if one wishes to use W for a purpose other than the evaluation of the stress-tensor expectation value, one can normalize $\langle \text{in}, 0|0, \text{in}\rangle$ to have value unity, in which case the first term in (9.68) vanishes.

Using (9.26), (9.36) and the interaction Lagrangian density (9.61) one obtains

$$\frac{\langle \text{in}, 0|S^{(1)}|0, \text{in}\rangle}{\langle \text{in}, 0|0, \text{in}\rangle}$$

$$= \mathrm{i}\int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}}\left\{\frac{\lambda}{8}G_{\mathrm{F}}^2(x,x) - \frac{\mathrm{i}}{2}(\delta m^2 + \delta\xi R)G_{\mathrm{F}}(x,x)\right\}$$

$$= \mathrm{i}\int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}}\left\{\frac{\lambda_{\mathrm{R}}}{8}G_{\mathrm{F}}^2(x,x) + \frac{\mathrm{i}\lambda_{\mathrm{R}}}{32\pi^2}\frac{[m_{\mathrm{R}}^2 + (\xi_{\mathrm{R}} - \frac{1}{6})R]}{(n-4)}G_{\mathrm{F}}(x,x)\right\} + \mathrm{O}(\lambda_{\mathrm{R}}^2), \tag{9.71}$$

where in the second term we have used (9.56)–(9.58) with the coefficients (9.64). Now substituting (9.52) for G_{F} gives

$$\frac{\langle \text{in}, 0|S^{(1)}|0, \text{in}\rangle}{\langle \text{in}, 0|0, \text{in}\rangle} = \mathrm{i}\int \mathrm{d}^n x[-g(x)]^{\frac{1}{2}}\left\{\frac{\lambda_{\mathrm{R}}}{8}[G_{\mathrm{F}}^{\text{finite}}(x,x)]^2 + \frac{\lambda_{\mathrm{R}}}{512\pi^4}\frac{[m_{\mathrm{R}}^2 + (\xi_{\mathrm{R}} - \frac{1}{6})R]^2}{(n-4)^2}\right\}. \tag{9.72}$$

The pole terms proportional to $(n-4)^{-1}$ have cancelled from this expression because of the presence of the m and ξ renormalizations. There is still, however, a double pole, which has not been removed in this way, and which will give a double pole term in the effective action (9.68). This remaining divergent term can be removed by renormalization of constants in the generalized Einstein Lagrangian. In this case, renormalization of Newton's constant, the cosmological constant and the constant multiplying R^2 will be required.

The proof that the renormalization scheme outlined above will remove all the poles that arise in $\lambda\phi^4$ theory, to all orders of perturbation, is already non-trivial in Minkowski space (see, for example, Breitenlohner & Maison 1977 and Manoukian 1979, who shows that the subtractions prescribed in methods such as Breitenlohner & Maison's, are equivalent to renormalization of constants in the action). In curved spacetime, the construction of such proofs is much more difficult and is beset by technical as well as more fundamental problems (Birrell & Taylor 1980). In particular, arguments involving Poincaré invariance, which can be used to good effect in demonstrating renormalizability in Minkowski space (see, for example, Roman 1969, §§4.2 & 5.1), can no longer be applied in curved spacetime. One immediate consequence of this already encountered above is the appearance of pole terms proportional to the Ricci scalar, and the consequent necessity of renormalizing ξ. However, R is not the only new scalar that could appear. For example a pole term proportional of G_F^{finite} arises from the term G_F^2 in (9.71). This scalar quantity will in general be an extremely complicated non-local function of the spacetime geometry. (Of course, the asymptotic expansion of $G_F^{\text{finite}}(y, y)$ in the first line of (9.52) is local.)

In terms of Feynman diagrams, the G_F^2 term in (9.71) is represented by fig. 30. The pole proportional to G_F^{finite} arises because the pole associated with one of the loops multiplies the finite part of the other loop. This is known as an 'overlapping' divergence, and its cancellation is vital for the renormalizability of the effective action W in terms of purely local quantities. In the case of (9.71), the overlapping divergence was cancelled by a term arising from m and ξ renormalization. Cancellation of overlapping divergences is also necessary in the two-particle S-matrix elements if they are to be made finite by (local) m and ξ renormalization. Bunch (1980) has proved that such cancellations do occur, and that $\lambda\phi^4$ theory remains renormalizable in curved spacetime.

The fact that the presence of an interaction will, in general, require additional λ-dependent renormalizations of constants in the generalized Einstein action, immediately suggests that there will be a λ-dependent contribution to the conformal anomaly. To study this it proves convenient to use renormalization group methods as adapted to dimensional

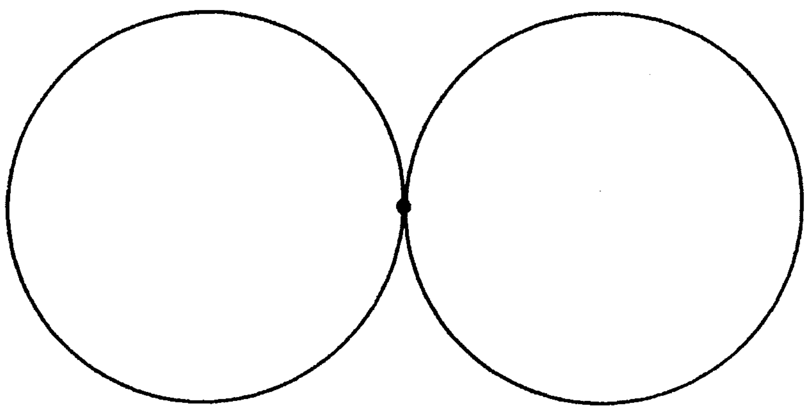


Fig. 30. Feynman diagram for the first order contribution to the vacuum-to-vacuum amplitude in $\lambda\phi^4$ theory.

regularization by 't Hooft (1973) (see also Collins & Macfarlane 1974).

The essential idea has already been encountered in §6.2. The theory can only predict precise S-matrix elements for a fixed value of the arbitrary mass scale μ. This introduces an inherent ambiguity into the values of the renormalized constants, which cannot be determined in practice by experiment. If μ is varied, it is necessary for the values of the renormalized constants to vary also in such a way that the S-matrix elements retain their physical, observed values.

Following 't Hooft, such changes in, for example, λ_R and m_R (the treatment of ξ_R and any other constant is similar to that of m_R) can be deduced by expressing the (fixed) bare constants λ and m given by (9.54), (9.55) and (9.58) in terms of a new mass scale μ', related to μ by the infinitesimal transformation

$$\mu' = \mu(1 + \varepsilon). \tag{9.73}$$

One obtains from (9.58) (to first order in ε)

$$\begin{aligned}\lambda &= (\mu')^{4-n}[1 + \varepsilon(n-4)]\{\lambda_R + \sum_{\nu=1}^{\infty} a_\nu(\lambda_R)(n-4)^{-\nu}\} \\ &= (\mu')^{4-n}\{\varepsilon(n-4)\lambda_R + \lambda_R + \varepsilon a_1(\lambda_R) \\ &\quad + \sum_{\nu=1}^{\infty} [a_\nu(\lambda_R) + \varepsilon a_{\nu+1}(\lambda_R)](n-4)^{-\nu}\}.\end{aligned} \tag{9.74}$$

Because this expansion contains a term of order $(n-4)$, it is not of the same form as the original expansion (9.58). One is always free to add such terms to the expansions of the bare constants in terms of the renormalized ones, because they vanish at the physical dimension $n = 4$; they reflect the usual renormalization ambiguity. However, if we had started the renormalization process with the mass scale μ' instead of μ, we should not have had terms of order $(n-4)$, as they do not appear in the basic ansatz (9.58). We thus remove the term proportional to $(n-4)$ in (9.74) by the transformation

$$\lambda_R = \tilde{\lambda}_R - \varepsilon(n-4)\tilde{\lambda}_R, \tag{9.75}$$

which, once again, is only non-trivial away from $n = 4$. With this transformation, (9.74) becomes

$$\begin{aligned}\lambda = (\mu')^{4-n}\Bigg[&\tilde{\lambda}_R + \varepsilon a_1(\tilde{\lambda}_R) - \varepsilon\tilde{\lambda}_R \frac{\partial a_1(\tilde{\lambda}_R)}{\partial \tilde{\lambda}_R} \\ &+ \sum_{\nu=1}^{\infty} (n-4)^{-\nu}\left(a_\nu(\tilde{\lambda}_R) + \varepsilon a_{\nu+1}(\tilde{\lambda}_R) - \varepsilon\tilde{\lambda}_R \frac{\partial a_{\nu+1}(\tilde{\lambda}_R)}{\partial \tilde{\lambda}_R}\right)\Bigg] + O(\varepsilon^2).\end{aligned} \tag{9.76}$$

Also, making the transformation (9.75) in (9.56) gives a new expansion for m:

$$m^2 = m_R^2 - \varepsilon m_R^2 \tilde{\lambda}_R \frac{\partial b_1(\tilde{\lambda}_R)}{\partial \tilde{\lambda}_R}$$

$$+ m_R^2 \sum_{\nu=1}^{\infty} (n-4)^{-\nu} \left(b_\nu(\tilde{\lambda}_R) - \varepsilon \tilde{\lambda}_R \frac{\partial b_{\nu+1}(\tilde{\lambda}_R)}{\partial \tilde{\lambda}_R} \right) + O(\varepsilon^2). \quad (9.77)$$

To recast (9.76) and (9.77) into the standard forms (9.56) and (9.58) (with μ replaced by μ') which we would have obtained had we used the mass scale μ' *ab initio*, define

$$\lambda'_R = \lambda_R + \varepsilon \left(a_1(\tilde{\lambda}_R) - \tilde{\lambda}_R \frac{\partial a_1(\tilde{\lambda}_R)}{\partial \tilde{\lambda}_R} \right) \quad (9.78)$$

and

$$(m'_R)^2 = m_R^2 - \varepsilon m_R^2 \tilde{\lambda}_R \frac{\partial b_1(\tilde{\lambda}_R)}{\partial \tilde{\lambda}_R} \quad (9.79)$$

respectively. Since λ and m are considered fixed, if the renormalized S-matrix elements are to have the same value when the mass scale μ' is used as when μ is used, then λ_R and m_R must undergo the transformations (9.78) and (9.79), which reduce at $n = 4$ to

$$\lambda'_R = \lambda_R + \varepsilon \left(a_1(\lambda_R) - \lambda_R \frac{\partial a_1(\lambda_R)}{\partial \lambda_R} \right), \quad (9.80)$$

$$(m'_R)^2 = m_R^2 - \varepsilon m_R^2 \lambda_R \frac{\partial b_1(\lambda_R)}{\partial \lambda_R}. \quad (9.81)$$

From these infinitesimal transformations, differential equations for the change of λ_R and m_R^2 with respect to μ are obtained:

$$\mu \frac{\partial \lambda_R}{\partial \mu} = a_1(\lambda_R) - \lambda_R \frac{\partial a_1(\lambda_R)}{\partial \lambda_R} \equiv \beta(\lambda_R), \quad (9.82)$$

$$\mu \frac{\partial m_R^2}{\partial \mu} = -m_R^2 \lambda_R \frac{\partial b_1(\lambda_R)}{\partial \lambda_R} \equiv -m_R^2 \gamma_m(\lambda_R). \quad (9.83)$$

The functions β and γ_m, are sometimes called the Callan–Symanzik β- and γ-functions, respectively, because of the role that they play in the renormalization group equations originally derived by Callan (1970) and Symanzik (1970). We shall now see that they play a similar role in connection with conformal anomalies (Drummond & Shore 1979, Birrell & Davies 1980*b*, Brown & Collins 1980).

The free scalar Lagrangian density (3.24) is invariant in four dimensions under the conformal transformations (3.1) and (3.7), provided $m = 0$ and $\xi = \frac{1}{6}$. The interaction Lagrangian density (9.40) is also invariant under these transformations in four dimensions. As discussed in §6.3, such invariance formally guarantees the vanishing of the trace of the vacuum expectation value of the stress-tensor. We have already seen that renormalization of constants in generalized Einstein action breaks this conformal invariance, and gives rise, even in the free field case, to an anomalous trace. In fact the free field trace anomaly can be separated from any interacting contributions, because the effective action (9.66) for the interacting field can, from (9.17), be written as

$$W = W_0 + W_1, \tag{9.84}$$

where W_0 is the free field effective action (6.25), and W_1 is the interaction-dependent part:

$$W_1 = \mathrm{i}\ln\left[\left\{\exp \mathrm{i}\int \mathrm{d}^n x \mathscr{L}_1\left(\frac{1}{\mathrm{i}}\frac{\delta}{\delta J(x)}\right)\right\}\right.$$
$$\left.\times \exp\left[-\tfrac{1}{2}\mathrm{i}\int J(y) G_{\mathrm{F}}(y,z) J(z)\mathrm{d}^n y\,\mathrm{d}^n z\right]\right]_{J=0} \tag{9.85}$$

Thus, using (cf. (6.117))

$$\langle T_\mu{}^\mu(x)\rangle = -\frac{1}{[-g(x)]^{\frac{1}{2}}}\left.\frac{\delta W[\Omega^2 g_{\mu\nu}]}{\delta\Omega(x)}\right|_{\Omega=1}, \tag{9.86}$$

one obtains,

$$\langle T_\mu{}^\mu\rangle = \langle T_\mu{}^\mu\rangle_0 + \langle T_\mu{}^\mu\rangle_1 \tag{9.87}$$

where $\langle T_\mu{}^\mu\rangle_0$ is the free field anomaly already discussed, and $\langle T_\mu{}^\mu\rangle_1$ is an interaction-dependent part obtained by substituting W_1 for W in (9.86).

In the case of interacting fields, it is not generally possible to maintain the conformal invariance of the action, let alone the effective action. The reason for this is that it is necessary to renormalize m and ξ, so that they cannot maintain the values $m = 0$ and $\xi = \frac{1}{6}$ needed for conformal invariance. Rather, these constants become indeterminate and it is only the renormalized constants m_{R} and ξ_{R} whose values can be determined by experiment. In a given regularization scheme, the bare and renormalized constants can be related by (9.54) and (9.55), where δm^2 and $\delta\xi$ have well-defined values. For example, in 't Hooft's scheme, provided one can show that the ansatz (9.56) successfully implements mass renormalization, then $m_R = 0$ implies $m = 0$. However, in general, δm^2 and $\delta\xi$ are totally arbitrary, and, in particular, if a different regularization scheme were used, $m_R = 0$ would not necessarily imply $\delta m = 0$. This is the case if point-splitting is used (Birrell & Ford 1979).

Since $m_R = 0$ implies $\delta m^2 = 0$ in the 't Hooft scheme, one might ask whether there is also a choice of ξ_R for which $\delta\xi = 0$. We see from (9.64) that if $\xi_R = \frac{1}{6}$, then $\delta\xi = 0$ to first order in λ_R. However, lengthy calculations (Birrell 1980, Bunch & Panangaden 1980) show that this relation breaks down at second order. If one is prepared to abandon the strict 't Hooft renormalization ansatz by allowing ξ_R to be a function of n, then Collins (1976) and Brown & Collins (1980) have shown that the choice $\xi_R = \xi(n)$ (see (3.27)) gives $\delta\xi = 0$ up to third order in λ_R, but not for higher orders. This choice of $\xi_R = \xi(n)$ is equivalent to taking $\xi_R = \frac{1}{6}$ in (9.57) and adding terms of order $n - 4$, thereby exploiting a form of the renormalization ambiguity discussed previously.

Since $\delta\xi \neq 0$ at fourth and higher order and hence the action is not conformally invariant, one does not expect the trace of the stress-tensor vacuum expectation value to be even formally traceless. However, the resulting trace can be divided into an anomalous and a non-anomalous part. The anomalous part comes from three sources: (i) The free field trace anomaly, which we have shown can be treated separately. (ii) Coupling-constant (λ_R) dependent renormalization of constants in the generalized Einstein action. There will also be a non-anomalous part arising from this source, but the anomalous part comes about in a similar fashion to the free field anomaly. (iii) The third source of the anomalous part of the trace is perhaps the most interesting, and arises from the fact that (9.40) is only conformally invariant in four dimensions. In n dimensions, under the transformations (3.1) and (3.7), (9.40) transforms to

$$\begin{aligned}\mathscr{L}_I[\bar{g}_{\mu\nu}] &= \Omega^{4-n}\mathscr{L}_I[g_{\mu\nu}] \\ &= -[-g(x)]^{\frac{1}{2}}(\mu\Omega)^{4-n}\frac{1}{4!}[\lambda_R + \sum_{\nu=1}^{\infty} a_\nu(\lambda_R)(n-4)^{-\nu}]\phi^4,\end{aligned} \tag{9.88}$$

where we have used (9.58). Because $\Omega(x)$ only enters (9.88) in the combination $\mu\Omega$ we may replace the variation of Ω by the variation of μ in (9.86), treating μ temporarily as a function of x:

$$\langle T_\mu{}^\mu(x)\rangle_I = -\frac{1}{[-g(x)]^{\frac{1}{2}}}\mu\frac{\delta W_I[\Omega^2 g_{\mu\nu}]}{\delta\mu(x)}\bigg|_{\substack{\mu(x)=\mu\\ \Omega=1}} \tag{9.89}$$

We can calculate (9.89) by using our knowledge of how λ_R and ξ_R must vary in order to exactly compensate for changes in μ. Suppose that $\lambda_R(x)$ and $\xi_R(x)$ have been chosen to compensate for changes in $\mu(x)$. Then

$$\left\{\left[\mu\frac{\delta}{\delta\mu(x)}+\mu\frac{\partial\lambda_R}{\partial\mu}\frac{\delta}{\delta\lambda_R(x)}+\mu\frac{\partial\xi_R}{\partial\mu}\frac{\delta}{\delta\xi_R(x)}\right]W_1[\Omega^2 g_{\mu\nu}]\right\}_{\Omega=1}=0, \quad (9.90)$$

where $\Omega = 1$ also signifies $\mu(x) = \mu$, $\lambda_R(x) = \lambda_R$, $\xi_R(x) = \xi_R$. Using (9.82), and additionally defining in analogy to (9.83)

$$\gamma_\xi(\lambda_R) \equiv -\mu\frac{\partial\xi_R}{\delta\mu} = \lambda_R\frac{\partial d_1(\lambda_R)}{\partial\lambda_R}, \quad (9.91)$$

(9.90) becomes

$$\left\{\left[\mu\frac{\delta}{\delta\mu(x)}+\beta(\lambda_R)\frac{\delta}{\delta\lambda_R(x)}-\gamma_\xi(\lambda_R)\frac{\delta}{\delta\xi_R(x)}\right]W_1[\Omega^2 g_{\mu\nu}]\right\}_{\Omega=1}=0, \quad (9.92)$$

which is similar in form to the usual renormalization group equation (see, for example, Collins & Macfarlane 1974, equation (11)). Employing (9.92) in (9.89), one obtains for this particular contribution to $\langle T_\mu{}^\mu\rangle_I$

$$\left\{\frac{1}{[-g(x)]^{\frac{1}{2}}}\left[\beta(\lambda_R)\frac{\delta}{\delta\lambda_R(x)}-\gamma_\xi(\lambda_R)\frac{\delta}{\delta\xi_R(x)}\right]W_1[\lambda_R(x),\xi_R(x)]\right\}_{\substack{\lambda_R(x)=\lambda_R\\ \xi_R(x)=\xi_R}} \quad (9.93)$$

In practice, one would not calculate the trace in a piecewise manner as above, but rather would endeavour to calculate the entire quantity. Equation (9.93) does, however, illustrate a very important difference between the free and interacting 'anomalous' traces: While the free field anomaly is state-independent and local, the interacting trace depends on the entire interaction part of the effective action, which will in general be both state-dependent and non-local. In particular, the use of $\langle T_\mu{}^\mu\rangle_I$ in (6.166) will generally give rise to an increase in energy density with time, representing particle production, unlike the case of the free field anomaly discussed on page 189. Particle production due to interactions has been discussed using this approach by Birrell & Davies (1980*b*). In the next section we consider the effects of interactions on particle production from an alternative point of view.

9.3 Particle production due to interaction

When interactions are present, particle production can occur for two reasons. Firstly, as in the free field case, the definition of positive and negative frequency differs between the in and out regions. Secondly, lack of Poincaré invariance permits energy and momentum to appear from the vacuum. Using the results of the previous two sections, it is not difficult to calculate the number of particles created by these effects.

If we work in the Heisenberg picture, then the state of the system is fixed for all time. Let us choose it to be the state which is the physical vacuum at early times, i.e., $|0, \text{in}\rangle$. In this picture, the operators carry the dynamics, and, in particular, the number operators at early and late times will be different. At early times, the operator representing the number of particles in the mode i, associated with the positive frequency mode function u_i^{in} (see (9.4a)) is

$$N_i^{\text{in}} = a_i^{\dagger\text{in}} a_i^{\text{in}}, \tag{9.94}$$

and $\langle \text{in}, 0|N_i^{\text{in}}|0, \text{in}\rangle = 0$. While at late times, the field has the mode decomposition (9.5), and the modes u_i^{out} are positive frequency, so the physical number operator is

$$\bar{N}_i^{\text{out}} = \bar{a}_i^{\text{out}\dagger} \bar{a}_i^{\text{out}}. \tag{9.95}$$

The number of particles which have been created from the $|0, \text{in}\rangle$ vacuum is thus

$$\langle \text{in}, 0|\bar{N}_i^{\text{out}}|0, \text{in}\rangle. \tag{9.96}$$

We can evaluate this expectation value in several ways. Firstly, we could write

$$|0, \text{in}\rangle = \sum_{l=0}^{\infty} \frac{1}{l!} \sum_{j_1, \ldots, j_l} |\bar{1}_{j_1}, \ldots, \bar{1}_{j_l}, \text{out}\rangle\langle \text{out}, \bar{1}_{j_l}, \ldots, \bar{1}_{j_1}|0, \text{in}\rangle, \tag{9.97}$$

which, when inserted in (9.96), gives products of expectation values

$$\langle \text{out}, \bar{1}_{j_l}, \ldots, \bar{1}_{j_1}|\bar{N}_i^{\text{out}}|\bar{1}_{j'_1}, \ldots, \bar{1}_{j'_l}, \text{out}\rangle \tag{9.98}$$

with amplitudes of the form (9.41). The expectation value (9.98) is trivially evaluated, since the barred out-Fock space is defined with respect to the $\bar{a}_i^{\text{out}}$ operators appearing in (9.95), and all of the interest lies in the amplitude (9.41). In (9.41), the two sources (i) and (ii) of particle production are distinctly separated; all the effects of mixing of positive and negative frequencies being confined to the first amplitude on the right-hand side, while it is only because of the lack of Poincaré invariance that the second amplitude is nonzero in more than just the trivial vacuum-to-vacuum case.

Since the amplitudes formed from the Bogolubov coefficients are fairly complicated expressions, in practice it is more convenient to follow a different procedure, expanding

$$|0,\mathrm{in}\rangle = \sum_{l=0}^{\infty} \frac{1}{l!} \sum_{j_1,\ldots,j_l} |1_{j_1},\ldots,1_{j_l},\mathrm{out}\rangle\langle \mathrm{out},1_{j_l},\ldots,1_{j_1}|0,\mathrm{in}\rangle, \qquad (9.99)$$

whence

$$\begin{aligned}\langle \mathrm{in},0|\bar{N}_i^{\mathrm{out}}|0,\mathrm{in}\rangle = \sum_{l=0}^{\infty}\frac{1}{l!}\sum_{m=0}^{\infty}\frac{1}{m!}\sum_{j_1,\ldots,j_l}\sum_{k_1,\ldots,k_m}&\langle \mathrm{in},0|1_{k_1},\ldots,1_{k_m},\mathrm{out}\rangle\\ &\times\langle \mathrm{out},1_{j_l},\ldots,1_{j_1}|0,\mathrm{in}\rangle\\ &\times\langle \mathrm{out},1_{k_m},\ldots,1_{k_1}|\bar{N}_i^{\mathrm{out}}|1_{j_1},\ldots,1_{j_l},\mathrm{out}\rangle. \qquad (9.100)\end{aligned}$$

Now, the first two amplitudes on the right-hand side are, in general, nonzero because of the lack of Poincaré invariance, while it is the final amplitude which carries information about particle creation due to mixing of positive and negative frequencies. This final amplitude is most easily evaluated by using the Bogolubov transformation (3.38) to write (9.95) in terms of the operators a_i^{out}, with respect to which the unbarred out-Fock space is defined. For simplicity, we consider only spatially flat Robertson–Walker spacetimes, in which the Bogolubov coefficients have the form (3.76), (3.77). Then (3.38) gives

$$\bar{a}_i^{\mathrm{out}} = \alpha_i^* a_i^{\mathrm{out}} - \beta_i^* a_{-i}^{\mathrm{out}\dagger}, \qquad (9.101)$$

and thus

$$\begin{aligned}\bar{N}_i^{\mathrm{out}} = |\alpha_i|^2 N_i^{\mathrm{out}} &+ |\beta_i|^2 a_{-i}^{\mathrm{out}} a_{-i}^{\mathrm{out}\dagger}\\ &- \alpha_i\beta_i^* a_i^{\mathrm{out}\dagger} a_{-i}^{\mathrm{out}\dagger} - \alpha_i^*\beta_i\, a_{-i}^{\mathrm{out}}\, a_i^{\mathrm{out}}, \qquad (9.102)\end{aligned}$$

with

$$N_i^{\mathrm{out}} = a_i^{\mathrm{out}\dagger} a_i^{\mathrm{out}}. \qquad (9.103)$$

Substituting (9.102) in (9.100), and including only those terms which give a nonzero contribution to the sum, one obtains

$$\begin{aligned}&\langle \mathrm{in},0|\bar{N}_i^{\mathrm{out}}|0,\mathrm{in}\rangle\\ &= |\beta_i|^2 + \sum_{m=0}^{\infty}\frac{1}{m!}\sum_{k_1,\ldots,k_m}\{|\alpha_i\langle \mathrm{out},1_{k_m},\ldots,1_{k_1}|0,\mathrm{in}\rangle|^2\\ &\quad\times\langle \mathrm{out},1_{k_m},\ldots,1_{k_1}|N_i^{\mathrm{out}}|1_{k_1},\ldots,1_{k_m},\mathrm{out}\rangle\\ &\quad - 2\mathrm{Re}[\alpha_i\beta_i^*\langle \mathrm{in},0|1_i,1_{-i},1_{k_1},\ldots,1_{k_m},\mathrm{out}\rangle\langle \mathrm{out},1_{k_m},\ldots,1_{k_1}|0,\mathrm{in}\rangle\\ &\quad\times\langle \mathrm{out},1_{k_m},\ldots,1_{k_1},1_{-i},1_i|a_i^{\mathrm{out}\dagger}a_{-i}^{\mathrm{out}\dagger}|1_{k_1},\ldots,1_{k_m},\mathrm{out}\rangle]\}. \qquad (9.104)\end{aligned}$$

In arriving at the first term on the right-hand side we have assumed that $\langle \text{in}, 0|0, \text{in}\rangle$ is normalized to unity (see page 307). This first term is precisely the number of particles produced in the absence of interaction (cf. (3.42)).

The remaining terms arise from the interaction, and must be calculated in perturbation theory. We shall give expressions only to first order in λ_R. The term involving $|\langle \text{out}, 1_{k_m}, \ldots, 1_{k_1}|0, \text{in}\rangle|^2$ in (9.104) does not contribute until order λ_R^2 so we shall ignore it. On the other hand, the term involving the β Bogolubov coefficient gives an order λ_R contribution coming from the $m = 0$ term in the sum:

$$
\begin{aligned}
&-2\mathrm{Re}[\alpha_i\beta_i^*\langle \text{in}, 0|1_i, 1_{-i}, \text{out}\rangle\langle \text{out}, 0|0, \text{in}\rangle\langle \text{out}, 1_{-i}, 1_i|a_i^{\text{out}\dagger}a_{-i}^{\text{out}\dagger}|0, \text{out}\rangle] \\
&\quad = -2\mathrm{Re}[\alpha_i\beta_i^*\langle \text{in}, 0|1_i, 1_{-i}, \text{out}\rangle\langle \text{out}, 0|0, \text{in}\rangle]. \qquad (9.105)
\end{aligned}
$$

We have seen in the previous section that the first amplitude in (9.105) is of order λ_R, while $\langle \text{out}, 0|0, \text{in}\rangle = 1 + O(\lambda_R)$ (with the normalization assumed above). Hence, to first order in λ_R, use of (9.65) in (9.105) gives the renormalized expectation value

$$
\begin{aligned}
&\langle \text{in}, 0|\bar{N}_i^{\text{out}}|0, \text{in}\rangle_{\text{ren}} \\
&\quad = |\beta_i|^2 - \mathrm{Re}\{\alpha_i\beta_i^*\lambda_R\int u_i(y)u_{-i}(y)[G_F^{\text{finite}*}(y, y) \\
&\qquad - (\mathrm{i}/8\pi^2)(m_R^2 + (\xi_R - \tfrac{1}{6})R(y))\ln\mu][-g(y)]^{\frac{1}{2}}\mathrm{d}^4y\}. \qquad (9.106)
\end{aligned}
$$

Note that the first order contribution comes about because of the combined effect of the Bogolubov transformation and the interaction. In particular, if $\beta_i = 0$, then there is no particle production to first order.

The occurrence of particle production to first order suggests that, even for fairly weak coupling (i.e., small λ_R), the contribution of the interaction to the total number of particles will be of importance compared with the free field production resulting from $|\beta_i|^2$. The existence of such first order terms has been noted by Lotze (1978) and Birrell & Ford (1979), and by Bunch, Panangaden & Parker (1980), who calculate the expectation value of the stress-tensor in the out region. Numerical examples of particle production due to interactions have been given by Birrell & Ford (1979) and Birrell, Davies & Ford (1980).

9.4 Other effects of interactions

The formalism developed in the preceding section can be used in the case of an eternal black hole with little difficulty (Birrell & Taylor 1980), but it does not immediately indicate whether the Hawking radiation will still be thermal in the presence of interactions. In view of the importance of the thermal nature of the Hawking radiation to the second law of thermodynamics (see §8.2), it is desirable to verify that interactions do not destroy this property.

An argument that interacting Hawking radiation is indeed thermal has been given by Gibbons & Perry (1976, 1978) and developed by Hawking (1981). Recall from §8.3 that, in the case of the free field, the scalar Green function G evaluated in the Kruskal vacuum is the same as a thermal Green function associated with the Schwarzschild coordinates. In particular, it is periodic in imaginary Schwarzschild time. Gibbons & Perry note that, in perturbation theory, the corresponding interacting Green functions are constructed from the free field Green functions, and so the Kruskal-associated one has this same time periodicity automatically built into it. Hence, as the time periodicity argument on page 26 made no reference to whether H and ϕ are free or interacting, this latter Green function would appear to share the thermal properties of the free field case.

If one wishes to forsake perturbation theory then formidable technical problems arise. Even in Minkowski space, the properties of exact (non-perturbative) quantum field theories can only be discussed in general terms (e.g., using spectral representations for the propagators). In curved spacetime, without even Poincaré invariance as a guide, progress is still harder.

Fortunately, some very special model field theories exist that can be extended to curved spacetime without use of perturbative or approximation techniques. One of these is the so-called massless Thirring model (Thirring 1958*a*,*b*), which has been extensively studied in the Minkowski space case (see, for example, Klaiber 1968). This model possesses exact solutions that can readily be extended to curved spacetime (Scarf 1962, Birrell & Davies 1978*a*).

The Thirring model is a theory of a massless, self-interacting, spin-$\frac{1}{2}$ field in two-dimensional spacetime. The theory is transcribed from Minkowski space to curved spacetime by applying the method used for free spinor fields in §3.8. One obtains for the Lagrangian density the sum of the free Lagrangian density (3.176) with m set to zero, and an interaction term

$$\mathscr{L}_\mathrm{I} = (\det V)\lambda J^\mu J_\mu, \tag{9.107}$$

where

$$J^\mu = \bar{\psi}\gamma^\mu\psi \tag{9.108}$$

(γ^μ being the curved spacetime Dirac gamma matrices; see page 85).

The theory is easily solved in two-dimensional curved spacetime because it is conformally invariant, with conformal weight $\frac{1}{2}$. Thus, the operator solution ψ of the field equation obtained by variation of the action is given in the line element (7.1) by

$$\psi = C^{-\frac{1}{4}}\tilde{\psi}, \tag{9.109}$$

where $\tilde{\psi}$ is the solution in Minkowski space (see Klaiber 1968).

Using (9.109), it is a simple matter to write down the Green functions in the spacetime of an eternal black hole in terms of the Minkowski spacetime Green functions given by Klaiber. In particular, using Kruskal coordinates to evaluate the Green functions in the vacuum $|0_K\rangle$ (as in the free field case, see page 276), one finds (Birrell & Davies 1978*a*) that, far from the black hole, these Green functions agree, as functions of the Schwarzschild coordinates, with the Thirring model thermal Green functions constructed by Dubin (1976). This confirms in the case of the Thirring model that interactions do not destroy the thermal nature of the Hawking radiation.

The Thirring model is exceptional for an interacting field theory in curved spacetime because it can be solved exactly. In general, it is extremely difficult to obtain solutions even using perturbation theory, because of the complicated nature of Feynman propagators in a general curved spacetime. However, the effects of topology on interacting fields can be readily investigated in the case where the geometry remains flat. For example, in §4.1, free quantum field theory in two-dimensional, flat spacetime with topology $R^1 \times S^1$ was studied. It is not difficult to extend that study to the perturbation solution of interacting theories, by using momentum space in much the same way as in Minkowski space theory.

To demonstrate this we consider flat spacetime with topology $R^3 \times S^1$. The Feynman propagator in this spacetime has the same form as in Minkowski space, namely $(p^2 - m^2 + i\varepsilon)^{-1}$, but the momentum component p^3 (corresponding to the dimension with S^1 topology) takes on discrete values

$$p^3 = \begin{cases} 2\pi j/L, & \text{untwisted fields} \\ 2\pi(j+\frac{1}{2})/L, & \text{twisted fields} \end{cases} \qquad j = 0, \pm 1, \pm 2 \ldots \tag{9.110}$$

where L is the periodicity length (as in §4.1). Thus, the evaluation of Feynman diagrams can be carried out in momentum space using techniques (such as Feynman parametrization) which are familiar from Minkowski space theory.

Consider, for example, the evaluation of the first order contribution to the amputated Green function (9.62) using dimensional regularization. To calculate $G_F(x,x)$ in n dimensions, we assume that the additional $n-4$

dimensions have topology R^1, i.e., we compute the Feynman propagator in the flat spacetime with topology $R^3 \times S^1 \times R^{n-4}$. There is no fundamental reason why we must add dimensions with topology R^1; it is equally as natural to add dimensions with S^1 topology. This is, of course, always arbitrary in dimensional regularization. However, the difference in the topology of the extra dimensions will only make a difference of order $(n-4)$ in $G_F(x, x)$, which leaves (9.62) unchanged. At higher orders of perturbation theory, one can show that the difference will only result in additions to terms which must in any case be renormalized, i.e., to terms which must be measured by experiment anyway (the usual renormalization ambiguity–see page 310).

With the above choice of n-dimensional topology, one has

$$G_F(x,x) = \frac{1}{L}\sum_{j=-\infty}^{\infty}\int\frac{d^{n-1}p}{(2\pi)^{n-1}}(p^2 - m_R^2 + i\varepsilon)^{-1}. \tag{9.111}$$

The integral can be evaluated using the useful formula ('t Hooft & Veltman 1972)

$$\int d^n p(m^2 - i\varepsilon - 2pk - p^2)^{-\alpha} = \frac{i\pi^{n/2}}{(m^2+k^2)^{\alpha-n/2}}\frac{\Gamma(\alpha-\frac{1}{2}n)}{\Gamma(\alpha)}. \tag{9.112}$$

The sum can then be simplified by various techniques (see, for example, Ford 1980*a*, Appendix B). In the limit in which $m_R = 0$, after performing the integral, (9.111) becomes

$$\frac{i\pi^{(n-5)/2}}{2L^{n-2}}\Gamma\left(\frac{3-n}{2}\right)\times\begin{cases}\sum_{j=-\infty}^{\infty}|j|^{n-3}, & \text{untwisted fields}\\ \sum_{j=-\infty}^{\infty}|j+\frac{1}{2}|^{n-3}, & \text{twisted fields.}\end{cases}$$

In this case, the sum can be evaluated in terms of Riemann's ζ-function, giving

$$G_F(x,x) = -\frac{i\pi^{(n-5)/2}}{2L^{n-2}}\Gamma\left(\frac{3-n}{2}\right)\zeta(3-n)\times\begin{cases}1, & \text{untwisted fields}\\ (2^{3-n}-1), & \text{twisted fields,}\end{cases} \tag{9.113}$$

which have finite limits as $n \to 4$:

$$G_F(x,x)\xrightarrow[n\to 4]{}\begin{cases}-i/(12L^2), & \text{untwisted fields}\\ i/(24L^2), & \text{twisted fields.}\end{cases} \tag{9.114}$$

The finite nature of the result is characteristic of the massless theory, and is anticipated by the fact that the pole term involving $b_{\nu 1}$ in (9.62) vanishes in this case (R is, of course, also zero in this case).

Substituting (9.114) into (9.62) (and setting $n = 4$), one observes that the effect is exactly the same as that of including a finite addition to δm^2 (recall (9.56)). In the case of untwisted fields this addition is $\lambda_R/(24L^2)$, while, for twisted fields the addition is $-\lambda_R/(48L^2)$. Thus, the field, which is massless at the zero-loop level, has developed an effective mass m_L at the one-loop level, due to the effect of the topology:

$$\left.\begin{aligned} m_L^2 &= \lambda_R/(24L^2), && \text{untwisted} \\ m_L^2 &= -\lambda_R/(48L^2), && \text{twisted.} \end{aligned}\right\} \tag{9.115}$$

This mass generation phenomenon in $\lambda\phi^4$ theory has been noted and studied by Ford & Yoshimura (1979), Birrell & Ford (1980), Denardo & Spallucci (1980*a*) and Toms (1980*a*, *b*, *c*), while similar phenomena in other theories have been investigated by Denardo & Spallucci (1980*b*), Ford (1980*a*), Omero & Percacci (1980) and Toms (1980*d*). It is closely related to the first order correction to the Casimir energy (Ford 1979, Kay 1979).

For $\lambda_R > 0$, the mass generated for untwisted fields is real, while that generated for twisted fields is imaginary (tachyonic). The presence of a tachyonic mass often means that the theory is unstable (Ford 1980*b*) and that spontaneous symmetry breaking will occur (see, for example, Goldstone 1961).

In the case of untwisted fields in flat spacetime with $R^3 \times S^1$ topology, the effects of the topology are very similar to those of nonzero temperature. In particular nonzero temperature $\lambda\phi^4$ theory leads to temperature-dependent mass generation and to restoration, at sufficiently high temperature, of broken symmetries (Kirzhnits & Linde 1972, 1976, Dolan & Jackiw 1974, Weinberg 1974, Kislinger & Morley 1976).

In the case of twisted fields, there is no simple relation between topological and nonzero temperature effects, and the study of spontaneous symmetry breaking is considerably complicated by the fact that the classical solutions to the field equation with a tachyonic mass cannot be constant (as in the untwisted case) because of the need to maintain antiperiodic boundary conditions (Avis & Isham 1978).

The combined effect of topology, nonzero temperature, and curvature on theories involving spontaneously broken symmetries is likely to have been of considerable importance in the early stages of the evolution of the universe. Some possible consequences are discussed in, for example, the papers of Dreitlein (1974), Kobsarev, Okun & Zel'dovich (1974), Domokos, Janson & Kovesi-Domokos (1975), Bludman & Ruderman (1977), Canuto & Lee (1977), Frolov, Grib & Mostepanenko (1977, 1978) and Gibbons (1978), within which references to earlier works can be found (see also page 128).

References*

Abers, E.S. and Lee, B.W. (1973) *Phys. Rep.*, **9C**, 1.
Abramowitz, M. and Stegun, I.A. (eds.) (1965) *Handbook of Mathematical Functions* (New York: Dover).
Abrikosov, A.A., Gorkov, L.P. and Dzyaloskinskii, I.Ye. (1963) *Quantum Field Theory Methods in Statistical Physics* (Englewood Cliffs: Prentice-Hall).
Adler, S.L. (1969) *Phys. Rev.*, **177**, 2426.
Adler, S.L., Lieberman, J. and Ng, Y.J. (1977) *Ann. Phys.* (NY), **106**, 279.
Adler, S.L., Lieberman, J. and Ng, Y.J. (1978) *Ann. Phys.* (NY), **113**, 294.
Ashmore, J.F. (1972) *Lett. Nuovo Cimento*, **4**, 289.
Ashtekar, A. and Magnon, A. (1975*a*) *Proc. R. Soc. London*, **A346**, 375.
Ashtekar, A. and Magnon, A. (1975*b*) *C.R. Acad. Sci. Ser. A*, **281**, 875.
Audretsch, J. and Schäfer, G. (1978*a*) *Phys. Lett.*, **66A**, 459.
Audretsch, J. and Schäfer, G. (1978*b*) *J. Phys. A: Gen. Phys.*, **11**, 1583.
Avis, S.J. and Isham, C.J. (1978) *Proc. R. Soc. London*, **A363**, 581.
Avis, S.J. and Isham, C.J. (1979*a*) *Nucl. Phys.*, **B156**, 441.
Avis, S.J. and Isham, C.J. (1979*b*) 'Quantum field theory and fibre bundles in a general spacetime' in *Recent Developments in Gravitation–Cargese 1978*, eds. M. Levy and S. Deser (New York: Plenum).
Avis, S.J. and Isham, C.J. (1980) *Commun. Math. Phys.*, **72**, 103.
Avis, S.J., Isham, C.J. and Storey, D. (1978) *Phys. Rev. D*, **18**, 3565.
Balazs N.L. (1958) *Astrophys. J.*, **128**, 398.
Balian, R. and Duplantier, B. (1977) *Ann. Phys.* (NY), **104**, 300.
Balian, R. and Duplantier, B. (1978) *Ann. Phys.* (NY), **112**, 165.
Banach, R. and Dowker, J.S. (1979) *J. Phys. A: Gen. Phys.*, **12**, 2527, 2545.
Bander, M. and Itzykson, C. (1966) *Rev. Mod. Phys.*, **38**, 346.
Bardeen, J.W., Carter, B. and Hawking, S.W. (1973) *Commun. Math. Phys.*, **31**, 161.
Bargmann, V. (1932) *Sitz. Preuss. Akad. Wiss.*, 346.
Barrow, J.D. and Matzner, R.A. (1977) *Mon. Not. R. Astron. Soc.*, **181**, 719.
Barshay, S. and Troost, W. (1978) *Phys. Lett.*, **73B**, 437.
Barut, A.O., Muzinich, I.J. and Williams, D.N. (1963) *Phys. Rev.*, **130**, 442.
Bekenstein, J.D. (1972) *Lett. Nuovo Cimento*, **4**, 7371.
Bekenstein, J.D. (1973) *Phys. Rev. D*, **7**, 2333.
Bekenstein, J.D. (1975) *Phys. Rev. D*, **12**, 3077.
Bekenstein, J.D. and Meisels, A. (1977) *Phys. Rev. D*, **15**, 2775.
Belifante, F.J. (1940) *Physica* (Utrecht), **7**, 305.
Berger, B. (1974) *Ann. Phys.* (NY), **83**, 458.
Berger, B. (1975) *Phys. Rev. D*, **12**, 368.

*In what follows we adhere to the American Physical Society approved style, as detailed in the APS *Style Manual* (3rd edition, 1978), reproduced in the *Bulletin of the American Physical Society* (1979), volume 24.

Bernard, C. and Duncan, A. (1977) *Ann. Phys.* (NY), **107**, 201.
Bianchi, L. (1918) *Lezioni Sulla Teoria dei Gruppi Continui Finiti Transformazioni* (Pisa: Spoerri).
Birrell, N.D. (1978) *Proc. R. Soc. London*, **A361**, 513.
Birrell, N.D. (1979*a*) *Proc. R. Soc. London*, **A367**, 123.
Birrell, N.D. (1979*b*) *J. Phys. A: Gen. Phys.*, **12**, 337.
Birrell, N.D. (1979*c*) Ph.D. Thesis, King's College, London (unpublished).
Birrell, N.D. (1980) *J. Phys. A: Gen. Phys.*, **13**, 569.
Birrell, N.D. and Davies, P.C.W. (1978*a*) *Phys. Rev. D*, **18**, 4408.
Birrell, N.D. and Davies, P.C.W. (1978*b*) *Nature* (London), **272**, 35.
Birrell, N.D. and Davies, P.C.W. (1980*a*) *J. Phys. A: Gen. Phys.*, **13**, 2109.
Birrell, N.D. and Davies, P.C.W. (1980*b*) *Phys. Rev. D*, **22**, 322.
Birrell, N.D., Davies, P.C.W. and Ford, L.H. (1980) *J. Phys. A: Gen. Phys.*, **13**, 961.
Birrell, N.D. and Ford, L.H. (1979) *Ann. Phys.* (NY), **122**, 1.
Birrell, N.D. and Ford, L.H. (1980) *Phys. Rev. D*, **22**, 330.
Birrell, N.D. and Taylor, J.G. (1980) *J. Math. Phys.* (NY), **21**, 1740.
Bjorken, J.D. and Drell, S.D. (1965) *Relativistic Quantum Mechanics* (Chs. 1–10) and *Relativistic Quantum Fields* (Chs. 11–19) (New York: McGraw-Hill).
Blandford, R.D. (1977) *Mon. Not. R. Astron. Soc.*, **181**, 489.
Blandford, R.D. and Thorne, K.S. (1979) in *General Relativity: An Einstein Centenary Survey*, eds. S.W. Hawking and W. Israel (Cambridge: Cambridge University Press).
Bleuler, K. (1950) *Helv. Phys. Acta*, **23**, 567.
Bludman, S.A. and Ruderman, M.A. (1977) *Phys. Rev. Lett.*, **38**, 255.
Blum, B.S. (1973) Ph.D. Thesis, Brandeis University (unpublished).
Bogolubov, N.N. (1958) *Zh. Eksp. Teor. Fiz.*, **34**, 58 (*Sov. Phys. JETP*, **7**, 51 (1958)).
Bogolubov, N.N. and Shirkov, D.V. (1959) *Introduction to the Theory of Quantized Fields* (New York: Interscience).
Bollini, C.G. and Giambiagi, J.J. (1972) *Phys. Lett.*, **40B**, 566.
Bondi, H. and Gold, T. (1948) *Mon. Not. R. Astron. Soc.*, **108**, 252.
Boulware, D.G. (1975*a*) *Phys. Rev. D*, **11**, 1404.
Boulware, D.G. (1975*b*) *Phys. Rev. D*, **12**, 350.
Boulware, D.G. (1976) *Phys. Rev. D*, **13**, 2169.
Breitenlohner, P. and Maison, D. (1977) *Commun. Math. Phys.*, **52**, 11.
Brown, L.S. and Cassidy, J.P. (1977*a*) *Phys. Rev. D*, **16**, 1712.
Brown, L.S. and Cassidy, J.P. (1977*b*) *Phys. Rev. D*, **15**, 2810.
Brown, L.S. and Collins, J.C. (1980) *Ann. Phys.* (NY), **130**, 215.
Brown, L.S. and Maclay, G.J. (1969) *Phys. Rev.*, **184**, 1272.
Brown, M.R. and Dutton, C.R. (1978) *Phys. Rev. D*, **18**, 4422.
Bucholz, H. (1969) *The Confluent Hypergeometric Function* (Berlin: Springer-Verlag).
Bunch, T.S. (1977) Ph.D. Thesis, King's College, London (unpublished).
Bunch, T.S. (1978*a*) *Phys. Rev. D*, **18**, 1844.
Bunch, T.S. (1978*b*) *J. Phys. A: Gen. Phys.*, **11**, 603.
Bunch, T.S. (1979) *J. Phys. A: Gen. Phys.*, **12**, 517.
Bunch, T.S. (1980) *J. Phys. A: Gen. Phys.*, **13**, 1297.
Bunch, T.S. (1981*a*) *Gen. Relativ. Gravit.*
Bunch, T.S. (1981*b*) *Ann. Phys.* (NY), **131**, 118.
Bunch, T.S., Christensen, S.M. and Fulling, S.A. (1978) *Phys. Rev. D*, **18**, 4435.

Bunch, T.S. and Davies, P.C.W. (1977*a*) *Proc. R. Soc. London*, **A357**, 381.
Bunch, T.S. and Davies, P.C.W. (1977*b*) *Proc. R. Soc. London*, **A356**, 569.
Bunch, T.S. and Davies, P.C.W. (1978*a*) *Proc. R. Soc. London*, **A360**, 117.
Bunch, T.S. and Davies, P.C.W. (1978*b*) *J. Phys. A :Gen. Phys.*, **11**, 1315.
Bunch, T.S. and Panangaden, P. (1980) *J. Phys. A :Gen. Phys.*, **13**, 919.
Bunch, T.S., Panangaden, P. and Parker, L. (1980) *J. Phys. A :Gen. Phys.*, **13**, 901.
Bunch, T.S. and Parker, L. (1979) *Phys. Rev. D*, **20**, 2499.
Caianello, E.R. (1973) *Combinatorics and Renormalization in Quantum Field Theory* (New York :W.A. Benjamin).
Callan, C.G. (1970) *Phys. Rev. D*, **2**, 1541.
Callan, C.G., Coleman, S. and Jackiw, R. (1970) *Ann. Phys.* (NY), **59**, 42.
Calvani, M. and Francaviglia, M. (1978) *Acta Phys. Pol.*, **B9**, 11.
Candelas, P. (1980) *Phys. Rev. D*, **21**, 2185.
Candelas, P. and Deutsch, D. (1977) *Proc. R. Soc. London*, **A354**, 79.
Candelas, P. and Deutsch, D. (1978) *Proc. R. Soc. London*, **A362**, 251.
Candelas, P. and Dowker, J.S. (1979) *Phys. Rev. D*, **19**, 2902.
Candelas, P. and Raine, D.J. (1975) *Phys. Rev. D*, **12**, 965.
Candelas, P. and Raine, D.J. (1976) *J. Math. Phys.* (NY) **17**, 2101.
Candelas, P. and Raine, D.J. (1977*a*) *Phys. Rev. D*, **15**, 1494.
Candelas, P. and Raine, D.J. (1977*b*) *Proc. R. Soc. London*, **A354**, 79.
Candelas, P. and Sciama, D.W. (1977) *Phys. Rev. Lett.*, **38**, 1372.
Canuto, V. and Lee, J.F. (1977) *Phys. Lett.*, **B72**, 281.
Capper, D.M. and Duff, M.J. (1974) *Nuovo Cimento*, **A23**, 173.
Capper, D.M. and Duff, M.J. (1975) *Phys. Lett.*, **53A**, 361.
Capper, D.M., Duff, M.J. and Halpern, L. (1974) *Phys. Rev. D*, **10**, 461.
Carter, B. (1975) in *Black Holes*, eds. C. DeWitt and B.S. DeWitt (New York :Gordon & Breach).
Casimir, H.B.G. (1948) *Proc. Kon. Ned. Akad. Wet.*, **51**, 793.
Castagnino, M., Verbeure, A. and Weder, R.A. (1974) *Phys. Lett.*, **8A**, 99.
Castagnino, M., Verbeure, A. and Weder, R.A. (1975) *Nuovo Cimento*, **26B**, 396.
Chakraborty, B. (1973) *J. Math. Phys.* (NY), **14**, 188.
Chandrasekhar, S. (1958) in *The Plasma in a Magnetic Field*, ed. R.K.M. Landshoff (Stanford, California :Stanford University Press).
Charap, J.M. and Duff, M.J. (1977) *Phys. Lett.*, **69B**, 445; **71B**, 219.
Chern, S.S. (1955) *Hamburg Abh.*, **20**, 177.
Chern, S.S. (1962) *J. Soc. Indust. Appl. Math.*, **10**, 751.
Chernikov, N.A. and Tagirov, E.A. (1968) *Ann. Inst. Henri Poincaré*, **9A**, 109.
Chitre, D.M. and Hartle, J.B. (1977) *Phys. Rev. D*, **16**, 251.
Christensen, S.M. (1975) *Bull. Am. Phys. Soc.*, **20**, 99.
Christensen, S.M. (1976) *Phys. Rev. D*, **14**, 2490.
Christensen, S.M. (1978) *Phys. Rev. D*, **17**, 946.
Christensen, S.M. and Duff, M.J. (1978*a*) *Phys. Lett.*, **76B**, 571.
Christensen, S.M. and Duff, M.J. (1978*b*) *Nucl. Phys.*, **B146**, 11.
Christensen, S.M. and Duff, M.J. (1979) *Nucl. Phys.*, **B154**, 301.
Christensen, S.M. and Duff, M.J. (1980) *Nucl. Phys.*, **B170** [FSI], 480.
Christensen, S.M. and Fulling, S.A. (1977) *Phys. Rev. D*, **15**, 2088.
Cline, D.B. and Mills, F.E. (eds.) (1978) *Unification of Elementary Forces and Gauge Theories* (London :Harwood Academic Publ.).
Collins, J.C. (1974) *Phys. Rev. D*, **10**, 1213.
Collins, J.C. (1976) *Phys. Rev. D*, **14**, 1965.
Collins, J.C. and Macfarlane, A.J. (1974) *Phys. Rev. D*, **10**, 1201.

Critchley, R. (1976) Ph.D. Thesis, University of Manchester (unpublished).
Curir, A. and Francaviglia, M. (1978) *Acta Phys. Pol.*, **B9**, 3.
Damour, T. (1977) in *Recent Developments in the Fundamentals of General Relativity*, ed. R. Ruffini (Amsterdam :North Holland).
Davies, P.C.W. (1974) *The Physics of Time Asymmetry* (London : Surrey University Press; Berkeley & Los Angeles : University of California Press).
Davies, P.C.W. (1975) *J. Phys. A :Gen. Phys.*, **8**, 609.
Davies, P.C.W. (1976) *Proc. R. Soc. London*, **A351**, 129.
Davies, P.C.W. (1977*a*) *Phys. Lett.*, **68B**, 402.
Davies, P.C.W. (1977*b*) *Proc. R. Soc. London*, **A354**, 529.
Davies, P.C.W. (1977*c*) *Proc. R. Soc. London*, **A353**, 499.
Davies, P.C.W. (1978) *Rep. Prog. Phys.*, **41**, 1313.
Davies, P.C.W. (1981) in *Quantum Gravity II :A Second Oxford Symposium*, eds. C.J. Isham, R. Penrose and D.W. Sciama (Oxford :Clarendon).
Davies, P.C.W. and Fulling, S.A. (1977*a*) *Proc. R. Soc. London*, **A356**, 237.
Davies, P.C.W. and Fulling, S.A. (1977*b*) *Proc. R. Soc. London*, **A354**, 59.
Davies, P.C.W., Fulling, S.A., Christensen, S.M. and Bunch, T.S. (1977) *Ann. Phys.* (NY), **109**, 108.
Davies, P.C.W., Fulling, S.A. and Unruh, W.G. (1976) *Phys. Rev. D.*, **13**, 2720.
Davies, P.C.W. and Unruh, W.G. (1977) *Proc. R. Soc. London*, **A356**, 569.
Davies, P.C.W. and Unruh, W.G. (1979) *Phys. Rev. D.*, **20**, 388.
Delbourgo, R. and Prasad, V.B. (1974) *Nuovo Cimento*, **21A**, 32.
Denardo, G. and Spallucci, E. (1979) *Nuovo Cimento*, **53B**, 334.
Denardo, G. and Spallucci, E. (1980*a*) *Nuovo Cimento*, **55B**, 97.
Denardo, G. and Spallucci, E. (1980*b*) *Nucl. Phys.*, **B169**, 514.
Denardo, G. and Spallucci, E. (1980*c*) *Nuovo Cimento*, **59A**, 1.
Deruelle, N. and Ruffini, R. (1976) *Phys. Lett.*, **57B**, 248.
Deser, S., Duff, M.J. and Isham, C.J. (1976) *Nucl. Phys.*, **B111**, 45.
Deser, S. and Zumino, B. (1976) *Phys. Lett.*, **62B**, 335.
de Sitter, W. (1917*a*) *Proc. Kon. Ned. Akad. Wet.*, **19**, 1217.
de Sitter, W. (1917*b*) *Proc. Kon. Ned. Akad. Wet.*, **20**, 229.
Deutsch, D. & Candelas, P. (1979) *Phys. Rev. D*, **20**, 3063.
DeWitt, B.S. (1953) *Phys. Rev.*, **90**, 357.
DeWitt, B.S. (1965) 'The dynamical theory of groups and fields' in *Relativity, groups and Topology*, eds. B.S. DeWitt and C. DeWitt (New York: Gordon & Breach).
DeWitt, B.S. (1967*a*) *Phys. Rev.*, **160**, 1113.
DeWitt, B.S. (1967*b*) *Phys. Rev.*, **162**, 1195, 1239.
DeWitt, B.S. (1975) *Phys. Rep.*, **19C**, 297.
DeWitt, B.S. (1979) in *General Relativity*, eds. S.W. Hawking and W. Israel (Cambridge: Cambridge University Press).
DeWitt, B.S. and Brehme, R.W. (1960) *Ann. Phys.* (NY), **9**, 220.
DeWitt, B.S., Hart, C.F. and Isham, C.J. (1979) in *Themes in Contemporary Physics*, ed. S. Deser (Amsterdam: North Holland).
Dirac, P.A.M. (1936) *Proc. R. Soc. London*, **A155**, 447.
di Sessa, A. (1974) *J. Math. Phys.* (NY), **15**, 1892.
Dolan, L. and Jackiw, R. (1974) *Phys. Rev. D*, **9**, 3320.
Dolginov, A.Z. and Toptygin, I.N. (1959) *Zh. Eksp. Teor. Fiz.*, **37**, 1411 (*Sov. Phys. JETP*, **10**, 1022 (1960)).
Domokos, G., Janson, M.M. and Kovesi-Domokos, S. (1975) *Nature* (London), **257**, 203.
Dowker, J.S. (1967*a*) *Proc. R. Soc. London*, **A297**, 351.
Dowker, J.S. (1967*b*) *Supp. Nuovo Cimento*, **5**, 734.

Dowker, J.S. (1971) *Ann. Phys.* (NY), **62**, 361.
Dowker, J.S. (1972) *Ann. Phys.* (NY), **71**, 577.
Dowker, J.S. (1977) *J. Phys. A: Gen. Phys.*, **10**, 115.
Dowker, J.S. and Banach, R. (1978) *J. Phys. A: Gen. Phys.*, **11**, 2255.
Dowker, J.S. and Critchley, R. (1976*a*) *Phys. Rev. D*, **13**, 224.
Dowker, J.S. and Critchley, R. (1976*b*) *Phys. Rev. D*, **13**, 3224.
Dowker, J.S. and Critchley, R. (1976*c*) *J. Phys. A: Gen. Phys.*, **9**, 535.
Dowker, J.S. and Critchley, R. (1977*a*) *Phys. Rev. D*, **15**, 1484.
Dowker, J.S. and Critchley, R. (1977*b*) *Phys. Rev. D*, **16**, 3390.
Dowker, J.S. and Dowker, Y.P. (1966*a*) *Proc. R. Soc. London*, **A294**, 175.
Dowker, J.S. and Dowker, Y.P. (1966*b*) *Proc. R. Soc. London*, **87**, 65.
Dowker, J.S. and Kennedy, G. (1978) *J. Phys. A: Gen. Phys.*, **11**, 895.
Dreitlein, J. (1974) *Phys. Rev. Lett.*, **33**, 1243.
Drummond, I.T. and Hathrell, S.J. (1980) *Phys. Rev D*, **21**, 958.
Drummond, I.T. and Shore, G.M. (1979) *Phys. Rev. D*, **19**, 1134.
Duan', I.S. (1956) *Zh. Eksp. Teor. Fiz.*, **31**, 1098 (*Sov. Phys. JETP*, **7**, 437 (1956)).
Dubin, D.A. (1976) *Ann. Phys.* (NY), **102**, 71.
Duff, M.J. (1975) in *Quantum Gravity: An Oxford Symposium*, eds. C.J. Isham, R. Penrose, and D.W. Sciama (Oxford: Clarendon).
Duff, M.J. (1977). *Nucl. Phys.*, **B125**, 334.
Duff, M.J. (1981) in *Quantum Gravity II: A Second Oxford Symposium*, eds. C.J. Isham, R. Penrose & D.W. Sciama (Oxford: Clarendon).
Dyson, F.J. (1949) *Phys. Rev.*, **75**, 486, 1736.
Einstein, A. (1917) *Sitz. Preuss. Akad. Wiss.*, 142. For an English translation, see *The Principle of Relativity* (New York: Dover 1952).
Epstein, H., Gaser, V. and Jaffe, A. (1965) *Nuovo Cimento*, **36**, 1016.
Fadeev, L.D. and Popov, V.N. (1967) *Phys. Lett.*, **25B**, 29.
Farrugia, C.J. and Hájicek, P. (1980) *Commun. Math. Phys.*, **68**, 291.
Fetter, A.L. and Walecka, J.D. (1971) *Quantum Theory of Many-Particle Systems* (New York: McGraw-Hill).
Feynman, R.P. (1963) *Acta Phys. Pol.*, **24**, 697.
Feynman, R.P. and Hibbs, A.R. (1965) *Quantum Mechanics and Path Integrals* (New York: McGraw-Hill).
Fierz, M. (1939) *Helv. Phys. Acta*, **12**, 3.
Fierz, M. and Pauli, W. (1939) *Proc. R. Soc. London*, **A173**, 211.
Fischetti, M.V., Hartle, J.B. and Hu, B.L. (1979) *Phys. Rev. D.*, **20**, 1757.
Fock, V.A. (1929) *Z. Phys.*, **57**, 261.
Fock, V.A. (1937) *Phys. Z. Sowjetunion*, **12**, 404.
Fock, V.A. and Ivanenko, D. (1929) *Z. Phys.*, **54**, 798.
Ford, L.H. (1975) *Phys. Rev. D.*, **12**, 2963.
Ford, L.H. (1976) *Phys. Rev. D*, **14**, 3304.
Ford, L.H. (1978) *Proc. R. Soc. London*, **A364**, 227.
Ford, L.H. (1979) *Proc. R. Soc. London*, **A368**, 305.
Ford, L.H. (1980*a*) *Phys. Rev. D*, **21**, 933.
Ford, L.H. (1980*b*) *Phys. Rev. D*, **22**, 3003.
Ford, L.H. and Parker, L. (1977) *Phys. Rev. D*, **16**, 245.
Ford, L.H. and Parker, L. (1978) *Phys. Rev. D*, **17**, 1485.
Ford, L.H. and Yoshimura T. (1979) *Phys. Lett.*, **70A**, 89.
Fradkin, E.S. and Vilkovisky, G.S. (1978) *Phys. Lett.*, **73B**, 209.
Frampton, P.H. (1977) *Lectures on Gauge Field Theories. Part Two: Quantization* University of California at Los Angeles publication UCLA/77/TEP/21.

Freedman, D.Z., Muzinich, I.J. and Weinberg, E.J. (1974) *Ann. Phys.* (NY), **87**, 95.

Freedman, D.Z. and Pi, S-Y. (1975) *Ann. Phys.* (NY), **91**, 442.

Freedman, D.Z., van Nieuwenhuizen, P. and Ferrara, S. (1976) *Phys. Rev. D*, **13**, 3214.

Freedman, D.Z. and Weinberg, E.J. (1974) *Ann. Phys.* (NY) **87**, 354.

Friedlander, F.G. (1975) *The Wave Equation in Curved Space-Time* (Cambridge: Cambridge University Press).

Frolov, V.M., Grib, A.A. and Mostepanenko, V.M. (1977) *Teor. & Math. Fiz.* (USSR), **33**, 42 (*Theor. & Math. Phys.*, **33**, 869 (1977)).

Frolov, V.M., Grib, A.A. and Mostepanenko, V.M. (1978) *Phys. Lett.*, **65A**, 282.

Frolov, V.M., Mamaev, S.G. and Mostepanenko, V.M. (1976) *Phys. Lett.*, **55A**, 389.

Fulling, S.A. (1972) Ph.D. Thesis, Princeton University (unpublished).

Fulling, S.A. (1973) *Phys. Rev. D*, **7**, 2850.

Fulling, S.A. (1977) *J. Phys. A: Gen. Phys.*, **10**, 917.

Fulling, S.A. (1979) *Gen. Relativ. Gravit.*, **10**, 807.

Fulling, S.A. and Davies, P.C.W. (1976) *Proc. R. Soc. London*, **A348**, 393.

Fulling, S.A., Parker, L. and Hu, B.L. (1974) *Phys. Rev. D*, **10**, 3905; erratum *ibid* **11**, 1714.

Fulling, S.A., Sweeny, M. and Wald, R.M. (1978) *Commun. Math. Phys.* **63**, 257.

Garabedian, P.R. (1964) *Partial Differential Equations* (New York: Wiley).

Georgi, H. and Glashow, S.L. (1974) *Phys. Rev. Lett.*, **32**, 438.

Gerlach, U. (1976) *Phys. Rev. D*, **14**, 1479.

Gibbons, G.W. (1975) *Commun. Math. Phys.*, **44**, 245.

Gibbons, G.W. (1977*a*) *Phys. Lett.*, **60A**, 385.

Gibbons, G.W. (1977*b*) *Phys. Lett.*, **61A**, 3.

Gibbons, G.W. (1978) *J. Phys. A:Gen. Phys.*, **11**, 1341.

Gibbons, G.W. (1979) in *General Relativity: An Einstein Centenary Survey*, eds. S.W. Hawking and W. Israel (Cambridge: Cambridge University Press).

Gibbons, G.W. (1980) *Ann. Phys.* (NY), **125**, 98.

Gibbons, G.W. and Hawking, S.W. (1977*a*) *Phys. Rev. D*, **15**, 2738.

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

Gibbons, G.W. and Hawking, S.W. (1979) *Commun. Math. Phys.*, **66**, 291.

Gibbons, G.W. and Perry, M.J. (1976) *Phys. Rev. Lett.*, **36**, 985.

Gibbons, G.W. and Perry, M.J. (1978) *Proc. R. Soc. London*, **A358**, 467.

Gibbons, G.W. and Pope, C.N. (1978) *Commun. Math. Phys.*, **61**, 239.

Gilkey, P.B. (1975) *Adv. Math.*, **15**, 334.

Ginibre, J. (1971) in *Statistical Mechanics and Quantum Field Theory*, eds. C. DeWitt and R. Stora (New York: Gordon and Breach).

Ginzburg, V.L., Kirzhnits, D.A. and Lyubushin, A.A. (1971) *Zh. Eksp. Teor. Fiz.*, **60**, 451 (*Sov. Phys. JETP*, **33**, 242 (1971)).

Gitman, D.M. (1977) *J. Phys. A: Gen. Phys.*, **10**, 2007.

Goldstone, J. (1961) *Nuovo Cimento*, **19**, 154.

Gradshteyn, I.S. and Ryzhik, I.M. (1965) *Tables of Integrals, Series and Products* (New York: Academic Press).

Greiner, P. (1971) *Archs. Ration. Mech. Analysis*, **41**, 163.

Grensing, G. (1977) *J. Phys. A: Gen. Phys.*, **10**, 1687.

Grib, A.A. and Mamaev, S.G. (1969) *Yad. Fiz.*, **10**, 1276 (*Sov. J. Nucl. Phys.*, **10**, 722 (1970)).

Grib, A.A. and Mamaev, S.G. (1971) *Yad. Fiz.*, **14**, 800 (*Sov. J. Nucl. Phys.*, **14**, 450 (1972)).

Grib, A.A., Mamaev, S.G. and Mostepanenko, V.M. (1976) *Gen. Relativ. Gravit*,. **7**, 535.
Grib, A.A., Mamaev, S.G. and Mostepanenko, V.M. (1980*a*) *J. Phys. A: Gen: Phys.*, **13**, 2057.
Grib, A.A., Mamaev, S.G. and Mostepanenko, V.M. (1980*b*) *Quantum Effects in Strong External Fields* (Moscow: Atomizdat). In Russian.
Grishchuk, L.P. (1974) *Zh. Eksp. Teor. Fiz.*, **67**, 825 (*Sov. Phys. JETP*, **40**, 409 (1975)).
Grishchuk, L.P. (1975) *Lett. Nuovo Cimento*, **12**, 60.
Grishchuk, L.P. (1977) 'Graviton creation in the early universe' in *Eighth Texas Symposium on Relativistic Astrophysics*, ed. M.D. Papagiannis (New York: New York Academy of Sciences).
Gupta, S.N. (1950) *Proc. Phys. Soc.*, (London) **A63**, 681.
Gürsey, F. (1964) 'Introduction to group theory' in *Relativity, Groups and Topology*, eds. B.S. DeWitt and C. DeWitt (New York: Gordon and Breach).
Gutzwiller, M. (1956) *Helv. Phys. Acta*, **29**, 313.
Haag, R. (1955) *Kgl. Danske Videnskab Selskab, Mat.-Fys. Medd.*, **29**, No. 12.
Hadamard, J. (1923) *Lectures on Cauchy's Problem in Linear Partial Differential Equations* (New Haven: Yale University Press).
Hagedorn, R. (1965) *Supp. Nuovo Cimento*, **3**, 147.
Hájiček, P. (1976) *Nuovo Cimento*, **33B**, 597.
Hájiček, P. (1977) *Phys. Rev. D*, **15**, 2757.
Hájiček, P. and Israel, W. (1980) *Phys. Lett.*, **80A**, 9.
Halpern, L. (1967) *Ark. Fys.*, **34**, 539.
Hartle, J.B. (1977) *Phys. Rev. Lett.*, **39**, 1373.
Hartle, J.B. (1981) in *Quantum Gravity II: A Second Oxford Symposium*, eds. C.J. Isham, R. Penrose and D.W. Sciama (Oxford: Clarendon).
Hartle, J.B. and Hawking, S.W. (1976) *Phys. Rev. D*, **13**, 2188.
Hartle, J.B. and Hu, B.L. (1979) *Phys. Rev. D*, **20**, 1772.
Hartle, J.B. and Hu, B.L. (1980) *Phys. Rev. D*, **21**, 2756.
Hatalkar, M.M. (1954) *Phys. Rev.*, **94**, 1472.
Hawking, S.W. (1970) *Commun. Math. Phys.*, **18**, 301.
Hawking, S.W. (1972) *Commun. Math. Phys.*, **25**, 152.
Hawking, S.W. (1973) 'The event horizon' in *Black Holes*, eds. C. DeWitt and B.S. DeWitt (New York: Gordon and Breach).
Hawking, S.W. (1974) *Nature* (London), **248**, 30.
Hawking, S.W. (1975) *Commun. Math. Phys.*, **43**, 199.
Hawking, S.W. (1976*a*) *Phys. Rev. D*, **13**, 191.
Hawking, S.W. (1976*b*) *Phys. Rev. D*, **14**, 2460.
Hawking, S.W. (1977*a*) *Commun. Math. Phys.*, **55**, 133.
Hawking, S.W. (1977*b*) *Sci. Am.*, **236**, 34.
Hawking, S.W. (1977*c*) *Phys. Lett.*, **60A**, 81.
Hawking, S.W. (1978) *Phys. Rev. D*, **18**, 1747.
Hawking, S.W. (1979) 'The path-integral approach to quantum gravity' in *General Relativity: An Einstein Centenary Survey*, eds. S.W. Hawking and W. Israel (Cambridge: Cambridge University Press).
Hawking, S.W. (1981) 'Acausal propagation in quantum gravity' in *Quantum Gravity II: A Second Oxford Symposium*, eds. C.J. Isham, R. Penrose and D. W. Sciama (Oxford: Clarendon).
Hawking, S.W. and Ellis, G.F.R. (1973) *The Large Scale Structure of Space-Time* (Cambridge: Cambridge University Press).
Hawking, S.W. and Penrose, R. (1970) *Proc. R. Soc. London*, **A314**, 529.
Hiscock, W.A. (1977*a*) *Phys. Rev. D*, **15**, 3054.

Hiscock, W.A. (1977*b*) *Phys. Rev. D*, **16**, 2673.
Hiscock, W.A. (1979) Ph.D. Thesis, University of Maryland (unpublished).
Hiscock, W.A. (1980) *Phys. Rev. D*, **21**, 2063.
Horibe, M. (1979) *Prog. Theor. Phys.*, **61**, 661.
Horowitz, G.T. (1980) *Phys. Rev. D*, **21**, 1445.
Horowitz, G.T. and Wald, R.M. (1978) *Phys. Rev. D*, **17**, 414.
Horowitz, G.T. and Wald, R.M. (1980) *Phys. Rev. D*, **21**, 1462.
Hosoya, A. (1979) *Prog. Theor. Phys.*, **61**, 280.
Hoyle, F. (1948) *Mon. Not. R. Astron. Soc.*, **108**, 372.
Hu, B.L. (1972) Ph.D. Thesis, Princeton University (unpublished).
Hu, B.L. (1973) *Phys. Rev. D*, **8**, 1048.
Hu, B.L. (1974) *Phys. Rev. D*, **9**, 3263.
Hu, B.L. (1978) *Phys. Rev. D*, **18**, 4460.
Hu, B.L. (1979) *Phys. Lett.*, **71A**, 169.
Hu, B.L. (1980) 'Quantum field theories and relativistic cosmology' in *Recent Developments in General Relativity*, ed. R. Ruffini, (Amsterdam: North Holland).
Hu, B.L., Fulling, S.A. and Parker, L. (1973) *Phys. Rev. D*, **8**, 2377.
Hu, B.L. and Parker, L. (1978) *Phys. Rev. D*, **17**, 933.
Hut, P. (1977) *Mon. Not. R. Astron. Soc.*, **180**, 379.
Imamura, T. (1960) *Phys. Rev.*, **118**, 1430.
Isham, C.J. (1975) 'An introduction to quantum gravity' in *Quantum Gravity: An Oxford Symposium*, eds. C.J. Isham, R. Penrose and D.W. Sciama (Oxford: Clarendon).
Isham, C.J. (1978*a*) 'Quantum field theory in a curved spacetime – a general mathematical framework' in *Proceedings of the Bonn Conference on Differential Geometrical Methods in Mathematical Physics*, eds. K. Bleuler, H.R. Petry and A. Reetz (New York: Springer Verlag).
Isham, C.J. (1978*b*) *Proc. R. Soc. London*, **A362**, 383.
Isham, C.J. (1978*c*) *Proc. R. Soc. London*, **A364**, 591.
Isham, C.J. (1979) *Proc. R. Soc. London*, **A368**, 33.
Isham, C.J. (1981) 'Quantum gravity – an overview' in *Quantum Gravity II: A Second Oxford Symposium*, eds. C.J. Isham, R. Penrose and D.W. Sciama (Oxford: Clarendon).
Isihara, A. (1971) *Statistical Physics* (New York: Academic Press).
Israel, W. (1976) *Phys. Lett.*, **57A**, 107.
Itzykson, C. and Zuber, J.-C. (1980) *Quantum Field Theory* (New York: McGraw-Hill).
Jauch, J.M. and Rohrlich, F. (1955) *The Theory of Photons and Electrons* (Reading, Mass.: Addison-Wesley).
Kadanoff, L.P. and Baym, G. (1962) *Quantum Statistical Mechanics* (Menlo Park, California: Benjamin).
Kay, B.S. (1979) *Phys. Rev. D*, **20**, 3052.
Kay, B.S. (1980) *Commun. Math. Phys.*, **71**, 29.
Kennedy, G. (1978) *J. Phys. A: Gen. Phys.*, **11**, L173.
Kennedy, G., Critchley, R. and Dowker, J.S. (1980) *Ann. Phys.* (NY), **125**, 346.
Kennedy, G. and Unwin, S.D. (1980) *J. Phys. A: Gen. Phys.*, **13**, L253.
Kibble, T.W.B. (1978) *Commun. Math. Phys.*, **64**, 73.
Kibble, T.W.B. and Randjbar-Daemi, S. (1980) *J. Phys. A: Gen. Phys.*, **13**, 141.
Kirzhnits, D.A. and Linde, A.D. (1972) *Phys. Lett.*, **42B**, 471.

Kirzhnits, D.A. and Linde, A.D. (1976) *Ann. Phys.*, (NY), **101**, 195.
Kislinger, M.B. and Morley, P.D. (1976) *Phys. Rev. D*, **13**, 2771.
Klaiber, B. (1968) in *Lectures in Theoretical Physics*, Vol. XA, eds. A.O. Barut and W.E. Britten (New York: Gordon and Breach).
Klein, O. (1929) *Z. Phys.*, **53**, 157.
Kobsarev, I.Yu., Okun, L.B. and Zel'dovich, Ya.B. (1974) *Phys. Lett.*, **50B**, 340.
Kreyszig, E. (1968) *Introduction to Differential Geometry and Riemannian Geometry* (Toronto: University of Toronto Press).
Kruskal, M.D. (1960) *Phys. Rev.*, **119**, 1743.
Landau, L.D. and Lifshitz, E.M. (1958) *Statistical Physics* (London: Pergamon).
Lapedes, A.S. (1978*a*) *Phys. Rev. D*, **17**, 2556.
Lapedes, A.S. (1978*b*) *J. Math. Phys.* (NY), **19**, 2289.
Lehmann, H., Symanzik, K. and Zimmermann, W. (1955) *Nuovo Cimento*, **1**, 1425.
Lifshitz, E.M. (1955) *Zh. Eksp. Teor. Fiz.*, **29**, 94 (*Sov. Phys. JETP*, **2**, 73 (1956)).
Lifshitz, E.M. and Khalatnikov, I.M. (1963) *Adv. Phys.*, **12**, 185.
Liouville, J. (1837) *Journal de Math.*, **2**, 24.
Lohiya, D. and Panchapakesan, N. (1978) *J. Phys. A: Gen. Phys.*, **11**, 1963.
Lohiya, D. and Panchapakesan, N. (1979) *J. Phys. A: Gen. Phys.*, **12**, 523.
Lotze, K.-H. (1978) *Acta Phys. Pol.*, **B9**, 665, 677.
Lukash, V.N., Novikov, I.D. and Starobinsky, A.A. (1975) *Zh, Eksp. Teor. Fiz.*, **69**, 1484 (*Sov. Phys. JETP*, **42**, 751 (1976)).
Lukash, V.N., Novikov, I.D., Starobinsky, A.A. and Zel'dovich, Ya.B. (1976) *Nuovo Cimento*, **35B**, 293.
Lukash, V.N. and Starobinsky, A.A. (1974) *Zh. Eksp. Teor. Fiz.*, **66**, 515 (*Sov. Phys. JETP*, **39**, 742 (1974)).
Lynden-Bell, D. and Wood, R. (1967) *Mon. Not. R. Astron. Soc.*, **136**, 107.
MacCallum, M.A.H. (1979) 'Anisotropic and inhomogenous relativistic cosmologies' in *General Relativity: An Einstein Centenary Survey*, eds. S.W. Hawking and W. Israel (Cambridge: Cambridge University Press).
Mamaev, S.G., Mostepanenko, V.M. and Starobinsky, A.A. (1976) *Zh. Eksp. Teor. Fiz.*, **70**, 1577 (*Sov. Phys. JETP*, **43**, 823 (1976)).
Manoukian, E.B. (1979) *Nuovo Cimento*, **53A**, 345.
Martellini, M., Sodano, P. and Vitiello, G. (1978) *Nuovo Cimento*, **48A**, 341.
Mashhoon, B. (1973) *Phys. Rev. D*, **8**, 4297.
Mattuck, R.D. (1967) *A Guide to Feynman Diagrams in the Many-Body Problem* (London: McGraw-Hill).
McKean, H.P. and Singer, I.M. (1967) *J. Diff. Geom.*, **1**, 43.
McNamara, J.M. (1978) *Proc. R. Soc. London*, **A358**, 499.
Meikle, W.P.S. (1977) *Nature* (London), **269**, 41.
Melnikov, V.N. and Orlov, S.V. (1979) *Phys. Lett.*, **70A**, 263.
Milne, E.A. (1932) *Nature* (London), **130**, 9.
Minakshisundaram, S. and Pleijel, A. (1949) *Can. J. Math.*, **1**, 242.
Misner, C.W. (1968) *Astrophys. J.*, **151**, 431.
Misner, C.W. (1972) *Phys. Rev. Lett.*, **28**, 994.
Misner, C.W., Thorne, K.S. and Wheeler, J.A. (1973) *Gravitation* (San Francisco: Freeman).
Mohan, G. (1968) in *Lectures in Theoretical Physics*, Vol. XB, eds. A.O. Barut and W.E. Britten (New York: Gordon and Breach).
Moore, G.T. (1970) *J. Math. Phys.* (NY), **9**, 2679.
Nambu, Y. (1950) *Prog. Theor. Phys.*, **5**, 82.

Nambu, Y. (1966) *Phys. Lett.*, **26B**, 626.
Nariai, H. (1971) *Prog. Theor. Phys.*, **46**, 433 and 776.
Nariai, H. (1976) *Nuovo Cimento*, **35B**, 259.
Nariai, H. (1977*a*) *Prog. Theor. Phys.*, **57**, 67.
Nariai, H. (1977*b*) *Prog. Theor. Phys.*, **58**, 560.
Nariai, H. (1977*c*) *Prog. Theor. Phys.*, **58**, 842.
Nash, C. (1978) *Relativistic Quantum Fields* (London: Academic Press).
Omero, C. and Percacci, R. (1980) *Nucl. Phys.*, **B165**, 351.
Page, D.N. (1976*a*) *Phys. Rev. D*, **13**, 198.
Page, D.N. (1976*b*) *Phys. Rev. D*, **14**, 3260.
Page, D.N. (1977) *Phys. Rev., D*, **16**, 2402.
Page, D.N. (1980) *Phys. Rev. Lett.*, **44**, 301.
Parker, L. (1966) 'The Creation of Particles in an Expanding Universe', Ph.D. Thesis, Harvard University (available from: University Microfilms Library Service, Xerox Corp., Ann Arbor, Michigan, USA).
Parker, L. (1968) *Phys. Rev. Lett.*, **21**, 562.
Parker, L. (1969) *Phys. Rev.*, **183**, 1057.
Parker, L. (1971) *Phys. Rev. D*, **3**, 346.
Parker, L. (1972) *Phys. Rev. D*, **5**, 2905.
Parker, L. (1973) *Phys. Rev. D*, **7**, 976.
Parker, L. (1975) *Phys. Rev. D*, **12**, 1519.
Parker, L. (1976) *Nature* (London), **261**, 20.
Parker, L. (1977) 'The production of elementary particles by strong gravitational fields' in *Asymptotic Structure of Space-Time*, eds. F.P. Esposito and L. Witten (New York: Plenum).
Parker, L. (1979) 'Aspects of quantum field theory in curved spacetime: effective action and energy–momentum tensor' in *Recent Developments in Gravitation*, eds. S. Deser and M. Levy (New York: Plenum).
Parker, L. and Fulling, S.A. (1973) *Phys. Rev. D*, **7**, 2357.
Parker, L. and Fulling, S.A. (1974) *Phys. Rev. D*, **9**, 341.
Parker, L., Fulling and Hu, B.L. (1974) *Phys. Rev. D*, **10**, 3905.
Parnovsky, S.L. (1979) *Phys. Lett.*, **73A**, 153.
Pauli, W. and Villars, F. (1949) *Rev. Mod. Phys.*, **21**, 434.
Penrose, R. (1964)'Conformal treatment of infinity' in *Relativity, Groups and Topology*, eds. B.S. DeWitt and C. DeWitt, New York: Gordon and Breach).
Penrose, R. (1965) *Proc. R. Soc. London*, **A284**, 159.
Penrose, R. (1969) *Rev. Nuovo Cimento*, **1**, 252.
Penrose, R. (1979) 'Singularities and time-asymmetry' in *General Relativity: An Einstein Centenary Survey*, eds. S.W. Hawking and W. Israel (Cambridge: Cambridge University Press).
Petrov, A.Z. (1969) *Einstein Spaces* (Oxford: Pergamon).
Planck, M.K.E.L. (1899) *Sitz. Deut. Akad. Wiss. Berlin, Kl. Math. -Phys Tech.*, 440.
Press, W.H. and Teukolsky, S.A. (1972) *Nature* (London), **238**, 211.
Raine, D.J. and Winlove, C.P. (1975) *Phys. Rev. D*, **12**, 946.
Randjbar-Daemi, S., Kay, B.S. and Kibble, T.W.B. (1980) *Phys. Lett.*, **91B**, 417.
Rees, M.J. (1977) *Nature* (London), **266**, 333.
Rindler, W. (1966) *Am. J. Phys.*, **34**, 1174. Many features of this work were anticipated in a paper by Max Born (1909) *Ann. Phys.* (Leipzig), **30**, 1.
Rindler, W. (1969) *Essential Relativity* (New York: Van Nostrand).
Roman, P. (1969) *Introduction to Quantum Field Theory* (New York: Wiley).

Rowan, D.J. and Stephenson, G. (1976) *J. Phys. A: Gen. Phys.*, **9**, 1631.
Rumpf, H. (1976*a*) *Phys. Lett.*, **61B**, 272.
Rumpf, H. (1976*b*) *Nuovo Cimento*, **35B**, 321.
Ruzmaikina, T.V. and Ruzmaikin, A.A. (1970) *Zh. Eksp. Teor. Fiz.*, **57**, 680 (*Sov. Phys. JETP*, **30**, 372 (1970)).
Rzewuski, J. (1969) *Field Theory*, Vol. II (London: Iliffe Books).
Sakharov, A.D. (1967) *Dok. Akad. Nauk. SSR*, **177**, 70 (*Sov. Phys. Dok.*, **12**, 1040 (1968)).
Salam, A. (1968) 'Weak and electromagnetic interactions' in *Elementary Particle Theory*, ed. N. Svartholm, (Stockholm: Almquist, Forlag AB), p. 367.
Sanchez, N. (1979) *Phys. Lett.*, **87B**, 212.
Scarf, F.L. (1962) 'A soluble quantum field theory in curved space' in *Les Théories Relativistes de la Gravitation* (Paris: Centre National de Recherche Scientifique).
Schäfer, G. (1978) *J. Phys. A: Gen. Phys.*, **11**, L179.
Schäfer, G. and Dehnen, H. (1977) *Astron. Astrophys.*, **54**, 823.
Schiff, L.I. (1949) *Quantum Mechanics* (New York: McGraw-Hill).
Schomblond, C. and Spindel, P. (1976) *Ann. Inst. Henri Poincaré*, **25A**, 67.
Schrödinger, E. (1932) *Sitz. Preuss. Akad. Wiss.*, 105.
Schrödinger, E. (1939) *Physica* (Utrecht), **6**, 899.
Schrödinger, E. (1956) *Expanding Universes* (Cambridge: Cambridge University Press).
Schweber, S. (1961) *An Introduction to Relativistic Quantum Field Theory* (New York: Harper & Row).
Schwinger, J. (1948) *Phys. Rev.*, **74**, 1439.
Schwinger, J. (1949*a*) *Phys. Rev.*, **75**, 651.
Schwinger, J. (1949*b*) *Phys. Rev.*, **76**, 790.
Schwinger, J. (1951*a*) *Phys. Rev.*, **82**, 664.
Schwinger, J. (1951*b*) *Proc. Nat. Acad. Sci.* (USA), **37**, 452.
Sciama, D.W. (1976) *Vistas in Astron.*, **19**, 385.
Segal, I.E. (1967) in *Applications of Mathematics to Problems in Theoretical Physics*, ed. F. Luciat (New York: Gordon and Breach).
Sexl. R.U. and Urbantke, H.K. (1967) *Acta Phys. Austriaca*, **26**, 339.
Sexl, R.U. and Urbantke, H.K. (1969) *Phys. Rev.*, **179**, 1247.
Shannon, C. and Weaver, W. (1949) *The Mathematical Theory of Communication*, (Urbana: University of Illinois Press).
Shore, G.M. (1980*a*) *Phys. Rev. D*, **21**, 2226.
Shore, G.M. (1980*b*) *Ann. Phys.* (NY), **128**, 376.
Simpson, M. and Penrose, R. (1973) *Int. J. Theor. Phys.*, **7**, 183.
Sommerfield, C.M. (1974) *Ann. Phys.* (NY), **84**, 285.
Sparnaay, M.J. (1958) *Physica* (Utrecht), **24**, 751.
Starobinsky, A.A. (1973) *Zh. Eksp. Teor. Fiz.*, **64**, 48 (*Sov. Phys. JETP*, **37**, 28(1973))
Starobinsky, A.A. (1980) *Phys. Lett.*, **91B**, 99.
Starobinsky, A.A. and Churilov, S.M. (1973) *Zh. Eksp. Teor. Fiz.*, **65**, 3 (*Sov. Phys. JETP*, **38**, 1 (1974)).
Stelle, K.S. (1977) *Phys. Rev. D*, **16**, 953.
Stelle, K.S. (1978) *Gen. Relativ. Gravit.*, **9**, 353.
Stewartson, K. and Waechter, R.T. (1971) *Proc. Camb. Phil. Soc.*, **69**, 353.
Symanzik, K. (1970) *Commun. Math. Phys.*, **18**, 48.
Synge, J.L. (1960) *Relativity: The General Theory* (Amsterdam: North Holland).
Tabor, D. and Winterton, R.H.S. (1969) *Proc. R. Soc. London*, **A312**, 435.

Tagirov, E.A. (1973) *Ann. Phys.* (NY), **76**, 561.
Takahashi, Y. and Umezawa, H. (1957) *Nuovo Cimento*, **6**, 1324.
Taylor, J.G. (1960) *Nuovo Cimento*, **17**, 695.
Taylor, J.G. (1963) *Supp. Nuovo Cimento*, **1**, 857.
Taylor, J.C. (1976) *Gauge Theories of Weak Interactions* (Cambridge: Cambridge University Press).
Teukolsky, S.A. (1973) *Astrophys. J.*, **185**, 635.
Teukolsky, S.A. and Press, W.H. (1974) *Astrophys. J.*, **193**, 443.
Thirring, W.E. (1958*a*) *Ann. Phys.* (NY), **3**, 91.
Thirring, W.E. (1958*b*) *Nuovo Cimento*, **9**, 1007.
't Hooft, G. (1973) *Nucl. Phys.*, **B61**, 455.
't Hooft, G. and Veltman, M. (1972) *Nucl. Phys*, **B44**, 189.
Tipler, F. (1980) *Phys. Rev. Lett.*, **45**, 949.
Tolman, R.C. (1934) *Relativity, Thermodynamics, and Cosmology* (Oxford: Clarendon).
Tomonaga, S. (1946) *Prog. Theor. Phys.*, **1**, 27.
Toms, D.J. (1980*a*) *Phys. Rev. D*, **21**, 928.
Toms, D.J. (1980*b*) *Phys. Rev. D*, **21**, 2805.
Toms, D.J. (1980*c*) *Ann. Phys.* (NY), **129**, 334.
Toms, D.J. (1980*d*) *Phys. Lett.*, **77A**, 303.
Troost, W. and van Dam H. (1977) *Phys. Lett.*, **71B**, 149.
Unruh, W.G. (1974) *Phys. Rev. D*, **10**, 3194.
Unruh, W.G. (1976) *Phys. Rev. D*, **14**, 870.
Urbantke, H.K. (1969) *Nuovo Cimento*, **63B**, 203.
Utiyama, R. (1962) *Phys. Rev.*, **125**, 1727.
Utiyama, R. and DeWitt, B.S. (1962) *J. Math. Phys.* (NY), **3**, 608.
Valatin, J.G. (1954*a*) *Proc. R. Soc. London*, **A222**, 93.
Valatin, J.G. (1954*b*) *Proc. R. Soc. London*, **A222**, 228.
Valatin, J.G. (1954*c*) *Proc. R. Soc. London*, **A225**, 535.
van Nieuwenhuizen, P. and Freedman, D.Z. (1979) (eds.) *Supergravity* (Amsterdam: North Holland).
Van Vleck (1928) *Proc. Nat. Acad. Sci.* (USA), **14**, 178.
Vilenkin, A. (1978) *Nuovo Cimento*, **44A**, 441.
Volovich, I.V., Zagrebnov, V.A. and Frolov, V.P. (1977) *Teor. & Math. Fiz.* (USSR), **33**, 3 (*Theor. & Math. Phys.*, **33**, 843 (1977)).
Waechter, R.T. (1972) *Proc. Camb. Phil. Soc.*, **72**, 439.
Wald, R.M. (1975) *Commun. Math. Phys.*, **45**, 9.
Wald, R.M. (1976) *Phys. Rev. D*, **13**, 3176.
Wald, R.M. (1977) *Commun. Math. Phys.*, **54**, 1.
Wald, R.M. (1978*a*) *Phys. Rev. D*, **17**, 1477.
Wald, R.M. (1978*b*) *Ann. Phys.* (NY), **110**, 472.
Wald, R.M. (1979*a*) *Ann. Phys.* (NY), **118**, 490.
Wald, R.M. (1979*b*) *Commun. Math. Phys.*, **70**, 221.
Wald, R.M. (1980) *Phys. Rev. D*, **21**, 2742.
Wald, R.M. and Ramaswamy, S. (1980) *Phys. Rev. D*, **21**, 2736.
Wehrl, A. (1978) *Rev. Mod. Phys.*, **50**, 221.
Weinberg, S. (1964*a*) *Phys. Rev.*, **133**, B1318.
Weinberg, S. (1964*b*) *Phys. Rev.*, **134**, B882.
Weinberg, S. (1964*c*) 'The quantum theory of massless particles' in *Lectures on Particles and Field Theory*. Vol. II, eds. K.A. Johnson and others (Englewood Cliffs, N.J.: Prentice-Hall).
Weinberg, S. (1967) *Phys. Rev. Lett.*, **19**, 1264.

Weinberg, S. (1972) *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* (New York: Wiley).
Weinberg, S. (1974) *Phys. Rev. D*, **9**, 3357.
Wichmann, E.H. (1962) *Selected Topics in the Theory of Particles and Fields* (Copenhagen: Nordita).
Wick, G.C. (1950) *Phys. Rev.*, **80**, 268.
Zel'dovich, Ya,B. (1970) *Pis'ma Zh. Eksp. Teor. Fiz.*, **12**, 443 (*JETP Lett.*, **12**, 307(1970)).
Zel'dovich, Ya.B. (1971) *Pis'ma Zh. Eksp. Teor. Fiz.*, **14**, 270 (*JETP Lett.*, **14**, 180(1971)).
Zel'dovich, Ya.B. (1972) *Zh. Eksp. Teor. Fiz.*, **62**, 2076 (*Sov. Phys. JETP*, **35**, 1085 (1972)).
Zel'dovich, Ya.B., Novikov, I.D. and Starobinsky, A.A. (1974) *Zh. Eksp. Teor. Fiz.*, **66**, 1897 (*Sov. Phys. JETP*, **39**, 933 (1974)).
Zel'dovich, Ya.B. and Pitaevsky, L.P. (1971) *Commun. Math. Phys.*, **23**, 185.
Zel'dovich, Ya.B. and Starobinsky, A.A. (1971) *Zh. Eksp. Teor. Fiz.*, **61**, 2161 (*Sov. Phys. JETP*, **34**, 1159 (1972)).
Zel'dovich, Ya.B. and Starobinsky, A.A. (1977) *Pis'ma Zh. Eksp. Teor. Fiz.*, **26**, 373 (*JETP Lett.*, **26**, 252 (1977)).
Zurek, W.H. (1980) *Phys. Lett.*, **77A**, 399.

CONFORMAL ANOMALIES

Adler, S.L., Lieberman, J. and Ng, Y.J. (1977) *Ann. Phys.* (NY), **106**, 279.
Bernard, C. and Duncan, A. (1977) *Ann. Phys.* (NY), **107**, 201.
Brown, L.S. (1977) *Phys. Rev. D*, **15**, 1469.
Brown, L.S. and Cassidy, J.P. (1977*a*) *Phys.Rev. D*, **15**, 2810.
Brown, L.S. and Cassidy, J.P. (1977*b*) *Phys. Rev. D*, **16**, 1712.
Bunch, T.S. (1978) *J. Phys.*, **A11**, 603.
Bunch, T.S. (1979) *J. Phys.*, **A12**, 517.
Capper, D.M. and Duff, M.J. (1974) *Nuovo Cimento*, **A23**, 173.
Capper, D.M. and Duff, M.J. (1975) *Phys. Lett.*, **53A**, 361.
Christensen, S.M. (1976) *Phys. Rev. D*, **14**, 2490.
Christensen, S.M. (1978) *Phys. Rev. D*, **17**, 946.
Christensen, S.M. and Duff, M.J. (1978*a*) *Phys. Lett.*, **76B**, 571.
Christensen, S.M. and Duff, M.J. (1978*b*) *Nucl. Phys.*, **B146**, 11.
Christensen, S.M. and Fulling, S.A. (1977) *Phys. Rev. D*, **15**, 2088.
Critchley, R. (1978*a*) *Phys. Rev. D*, **18** 1849.
Critchley, R. (1978*b*) *Phys. Lett.*, **78B**, 410.
Davies, P.C.W. (1977) 'Stress tensor calculations and conformal anomalies' in *Eighth Texas Symposium on Relativistic Astrophysics*, ed. M.D. Papagiannis (New York: New York Academy of Sciences).
Davies, P.C.W. and Fulling, S.A. (1977) *Proc. R. Soc. London*, **A354**, 59.
Davies, P.C.W., Fulling, S.A. and Unruh, W.G. (1976) *Phys. Rev. D*, **13**, 2720.
Deser, S., Duff, M.J. and Isham, C.J. (1976) *Nucl. Phys.*, **B111**, 45.
Dowker, J.S. and Critchley, R. (1976) *Phys. Rev. D*, **13**, 3224.
Dowker, J.S. and Critchley, R. (1977) *Phys. Rev. D*, **16**, 3390.
Duncan, A. (1977) *Phys. Lett.*, **66B**, 170.
Duff, M.J. (1978) *Nucl. Phys.*, **B125**, 334.
Eguchi, T. and Freund, P.G.O. (1976) *Phys. Rev. Lett.*, **37**, 125.
Fradkin, E.S. and Vilkovisky, G.A. (1978) *Phys. Lett.*, **73B**, 209.

Hawking, S.W. (1977) *Commun. Math. Phys.*, **55**, 133.
Hu, B.L. (1978) *Phys. Rev. D.*, **18**, 4460.
Hu, B.L. (1979) *Phys. Lett.*, **71A**, 169.
Perry, M.J. (1978) *Nucl. Phys.*, **B143**, 114.
Tsao, H.S. (1977) *Phys. Lett.*, **68B**, 79.
Wald, R.M. (1977) *Commun. Math. Phys.*, **54**, 1.
Wald, R.M. (1978a) *Ann. Phys.* (NY), **110**, 472.
Wald, R.M. (1978b) *Phys. Rev. D*, **17**, 1477.
Vilenkin, A. (1978) *Nuovo Cimento*, **44A**, 441.

DE SITTER SPACE

Adler, S.L. (1972) *Phys. Rev. D*, **6**, 3445.
Adler, S.L. (1973) *Phys. Rev. D*, **8**, 2400.
Birrell, N.D. (1978) *Proc. R. Soc. London*, **A361**, 513.
Birrell, N.D. (1979) *J. Phys. A: Gen. Phys.*, **12**, 337.
Borner, G. and Durr, H.P. (1969) *Nuovo Cimento*, **64**, 669.
Bunch, T.S. and Davies, P.C.W. (1978) *Proc. R. Soc. London*, **360**, 117.
Candelas, P. and Raine D.J. (1975) *Phys. Rev. D*, **12**, 965.
Chernikov, N.A. and Tagirov. E.A. (1968) *Ann. Inst. Henri Poincaré*, **9A**, 109.
Dirac, P.A.M. (1935) *Ann. Math.*, **36**, 657.
Dowker, J.S. and Critchley, R. (1976*a*) *Phys. Rev. D*, **13**, 224.
Dowker, J.S. and Critchley, R. (1976*b*) *Phys. Rev. D*, **13**, 3224.
Drummond, I.T. (1975) *Nucl. Phys.*, **B94**, 115.
Drummond, I.T. and Shore, G.M. (1979) *Phys. Rev. D*, **19**, 1134.
Figari, R., Hoegh-Krohn, R. and Nappi, C. (1975) *Commun. Math. Phys.*, **44**, 265.
Fronsdal, C. (1965) *Rev. Mod. Phys.*, **37**, 221.
Fronsdal, C. (1974) *Phys. Rev. D*, **10**, 589.
Fronsdal, C. and Haugen, R.B. (1975) *Phys. Rev. D*, **12**, 3810.
Fulling, S.A. (1972) Ph.D. Thesis, Princeton University (unpublished).
Geheniau, J. and Schomblond, C. (1968) *Acad. R. de Belgique, Bull. Cl. des Sciences*, **54**, 1147.
Gibbons, G.W. and Hawking, S.W. (1977) *Phys. Rev. D*, **15**, 2738.
Grensing, G. (1977) *J. Phys. A: Gen. Phys.*, **10**, 1687.
Lohiya, D. (1978) *J. Phys. A: Gen Phys.*, **11**, 1335.
Lohiya, D. and Panchapakesan, N. (1978) *J. Phys. A: Gen. Phys.*, **11**, 1963.
Lohiya, D. and Panchapakesan, N. (1979) *J. Phys. A: Gen. Phys.*, **12**, 533.
Nachtmann, O. (1967) *Commun. Math. Phys.*, **6**, 1.
Rumpf, H. (1979) *Gen. Relativ. Gravit.*, **10**, 647.
Schomblond, C. and Spindel, P. (1976) *Ann. Inst. Henri Poincaré*, **25A**, 67.
Shore, G.M. (1979) *Ann. Phys.* (NY), **117**, 89.
Shore, G.M. (1980) *Phys. Rev. D*, **21**, 2226.
Tagirov, E.A. (1973) *Ann. Phys.* (NY), **76** 561.

Index