QUANTUM FIELD THEORY IN CURVED SPACETIME

Bryce S. DeWITT

Center for Relativity, Department of Physics, University of Texas, Austin, Texas, USA

NORTH-HOLLAND PUBLISHING COMPANY – AMSTERDAM

PHYSICS REPORTS (Section C of Physics Letters) 19, no. 6 (1975) 295–357. NORTH-HOLLAND PUBLISHING COMPANY

QUANTUM FIELD THEORY IN CURVED SPACETIME*

Bryce S. DeWITT

Center for Relativity, Department of Physics, University of Texas, Austin, Texas, USA

Received February 1975

Abstract:

Quantum field theory predicts a number of unusual physical effects in non-Minkowskian manifolds (flat or curved) that have no immediate analogs in Minkowski spacetime. The following examples are reviewed: (1) The Casimir effect; (2) Radiation from accelerating conductors; (3) Particle production in manifolds with horizons, including both stationary black holes and black holes formed by collapse. In the latter examples curvature couples directly to matter through the stress tensor and induces the creation of real particles. However, it also induces serious divergences in the vacuum stress. These divergences are analyzed, and methods for handling them are reviewed.

Single orders for this issue

PHYSICS REPORTS (Section C of PHYSICS LETTERS) 19, no. 6 (1975) 295–357

Copies of this issue may be obtained at the price given below. All orders should be sent directly to the Publisher. Orders must be accompanied by check.

Single issue price Dfl. 20.–, postage included.

* This paper is an expanded version of a talk given at the meeting of the APS Division of Particles and Fields in Williamsburg, Virginia, September 1974. Its preparation was supported in part by the National Science Foundation and in part with the assistance of a NATO travel grant.

Contents:

1. Introduction

The existence of the Poincaré group as a local symmetry group for spacetime has been enormously important to particle physicists in helping them to sort out their ideas and to construct formalisms for describing experimental facts – formalisms that run the gamut from pure phenomenology through dispersion theory to axiomatic field theory. In fact, students are taught nowadays that elementary particles simply *are* certain representations of the Poincaré group.

An addiction of any kind ultimately extracts a penalty from the addict. Physicists learned this lesson well in the early decades of this century. Most of us are aware that quantum field theory cannot in the end be based on the Poincaré group. What is needed is a theory – or at least a framework – that respects the full general covariance of Einstein's view of spacetime as a Riemannian manifold.

It is not my purpose here to present such a theory; it does not yet exist, at least as a coherent discipline. What I shall do is describe several distinct but related examples of physical processes that involve the manifold structure of spacetime in an essential way and that show some of the important elements that must go into such a theory. These examples are chosen both for their pedagogical value and for their current interest, and I hope that they will convince the reader not only that a coherent theory can ultimately be built but that it will also be extremely beautiful.

The core of any theory of interacting fields is the set of currents that describe the interaction. The currents of general relativity theory are the components of the stress tensor. A fundamental task – I might even say *the* main problem – in developing a quantum field theory in curve spacetime is to understand the stress tensor. The stress tensor, like any current, is formally a bilinear product of operator-valued distributions (the field operators) and hence is meaningless. The problem is to give it meaning, by some subtraction process.

A subtraction, or regularization, procedure conventionally makes use of the vacuum state. Particle physicists know what the vacuum is: It is (modulo symmetry-breaking degeneracies) the trivial representation of the Poincaré group. General relativists are not so lucky. In the absence of geometrical symmetries they have many "vacua" to choose from.

1.1. Basis functions, vacuum states, and Bogoliubov transformations

Let φ be a linear free field propagating in curved spacetime. φ may be either a boson or fermion field. We suppress any indices it may bear and assume, without loss of generality., that it is real (Hermitian). (Any complex field can be split into its real and imaginary parts.) Its dynamical equations will have the form

$$F\varphi = 0 \tag{1}$$

where F is a self-adjoint differential operator in the sense that

$$\int \psi_1^*(F\psi_2)\, \mathrm{d}^4x = \int (F\psi_1)^*\psi_2\, \mathrm{d}^4x\,, \tag{2}$$

the integrals being taken over the (open) region of spacetime of interest and ψ_1 and ψ_2 being any two smooth complex functions having compact support in that region. The action functional for the field, which, under variation, yields equations (1), may be expressed in the form

$$S = \tfrac{1}{2}\int \varphi(F\varphi)\, \mathrm{d}^4x\,, \tag{3}$$

which I shall sometimes write more simply as

$$S = \tfrac{1}{2}\varphi F\varphi, \tag{4}$$

a further suppression of indices, namely the spacetime coordinate labels x^μ, and a summation-*integration* convention for the unwritten indices being understood.

Because F is self-adjoint there always exists a two-edged vector differential operator $\overleftrightarrow{f}^\mu$ related to F in the following way:

$$\int_{\Omega} [\psi_1^*(F\psi_2) - (F\psi_1)^* \psi_2] \mathrm{d}^4 x = \int_{\partial\Omega} \psi_1^* \overleftrightarrow{f}^\mu \psi_2 \, \mathrm{d}\Sigma_\mu \,, \tag{5}$$

where Ω is any compact region of spacetime with smooth boundary $\partial\Omega$, ψ_1, ψ_2 are any two smooth complex functions defined over an open region containing Ω, and $\mathrm{d}\Sigma_\mu$ is the outward directed surface element of $\partial\Omega$. Let u_1 and u_2 be any two complex solutions of the field equations (1) and let Σ be any complete Cauchy hypersurface for these equations. (We assume the region of spacetime of interest to be such that there *are* complete Cauchy hypersurfaces for it.) Then the operator $\overleftrightarrow{f}^\mu$ may be used to define an inner product for u_1 and u_2, which is invariant under smooth deformations and displacements of Σ:

$$\langle u_1, u_2 \rangle = -\mathrm{i} \int_{\Sigma} u_1^* \overleftrightarrow{f}^\mu u_2 \, \mathrm{d}\Sigma_\mu \,. \tag{6}$$

This inner product will not be positive definite for boson fields.

The game now is to introduce a complete (modulo gauge transformations, if any) set of conjugate pairs of solutions u_i, u_i^* of equations (1) satisfying the following orthonormality conditions*

$$\langle u_i, u_j \rangle = \delta_{ij}, \quad \langle u_i^*, u_j \rangle = 0 \,. \tag{7}$$

There will be an infinity of such sets. Choose one. Expand the field in the form**

$$\varphi = \sum_i (a_i u_i + a_i^* u_i^*) \,. \tag{8}$$

By using the canonical (anti)commutation relations, or, in a more elegant and manifestly covariant way, by using the Peierls [40] definition of the (anti)commutator, it is then easy to show that the operator coefficients in the expansion satisfy the (anti)commutation relations

$$[a_i, a_j^*]_\pm = \delta_{ij}, \quad [a_i, a_j]_\pm = 0. \tag{9}$$

This operator algebra serves in the traditional fashion to define a Fock space and a "vacuum" state:

$$a_i|\mathrm{vac}\rangle = 0. \tag{10}$$

Note that the curvature of spacetime does not interfere in any way with the above construction. Therefore we may proceed immediately to the (formal) computation of matrix elements of the stress tensor. The stress tensor is defined by functional differentiation of the action with respect to the metric tensor $g_{\mu\nu}$:

$$T^{\mu\nu}(\varphi, \varphi) \equiv 2 \frac{\delta S}{\delta g_{\mu\nu}} = \varphi \frac{\delta F}{\delta g_{\mu\nu}} \varphi \,. \tag{11}$$

* Some of the labels for which the indices i, j stand may be continuous. The symbol δ_{ij} is understood to include a δ-function for each such label.

** The asterisk, applied to an operator, denotes the Hermitian conjugate, to a c-number or matrix of c-numbers, the ordinary complex conjugate. The dagger will be applied only to matrices, having either c-numbers or operators as elements, and will indicate that a transposition of the matrix is to be effected in addition to complex (Hermitian) conjugation of its elements (cf. eq. (15)).

Actually this form is a tensor *density*, i.e., it includes the factor $g^{1/2}$ where $g \equiv -\det(g_{\mu\nu})$, and I shall always leave this factor in.

The simplest matrix element is the "vacuum" expectation value, which is immediately seen to be given by[☆]

$$\langle T^{\mu\nu}\rangle_{\text{vac}} = \sum_i T^{\mu\nu}(u_i, u_i^*) . \tag{12}$$

The only difficulty with this expression is that the sum diverges. The naive way out of the difficulty is to throw the divergence away and to "regularize" $T^{\mu\nu}$ via

$$T^{\mu\nu}_{\text{reg}} \equiv T^{\mu\nu} - \langle T^{\mu\nu}\rangle_{\text{vac}} , \tag{13}$$

with the subtraction being understood to be carried out *mode by mode.* This is equivalent to normal ordering the bilinear form $T^{\mu\nu}(\varphi, \varphi)$ relative to the decomposition (8). The trouble, of course, is that a different decomposition leads to a different, and generally inequivalent, normal ordering. For if $\overline{u}_i$ are the basis functions of an alternative set they will be related to the u_i by

$$\overline{u}_i = \sum_j (\alpha_{ij} u_j + \beta_{ij} u_j^*) , \tag{14}$$

where the coefficients α_{ij}, β_{ij} satisfy the matrix relation (indices suppressed)

$$\begin{pmatrix} \alpha & \beta \\ \beta^* & \alpha^* \end{pmatrix} \begin{pmatrix} \alpha^\dagger & \pm\beta^\sim \\ \pm\beta^\dagger & \alpha^\sim \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \tag{15}$$

the +(−) sign being taken for fermion (boson) fields and the tilde denoting matrix transportation. If the β_{ij} vanish the "vacuum" is left unchanged, but if the β_{ij} do not vanish we have a Bogoliubov transformation

$$\overline{a}_i = \sum_j (\alpha_{ij}^* a_j \pm \beta_{ij}^* a_j^*) , \tag{16}$$

with

$$\langle \overline{a}_i^* \overline{a}_i\rangle_{\text{vac}} = \sum_j |\beta_{ij}|^2 . \tag{17}$$

That is, the old "vacuum" contains new "particles." It may even contain an infinite number of new "particles", in which case the two Fock spaces cannot be related by a unitary transformation.

1.2. Killing vectors and positive-frequency functions

The alert reader will now object that an important criterion has been ignored in the above discussion. One must use basis functions that distinguish the *positive* frequency solutions from the *negative* frequency ones. Such a distinction can be made only if the concepts of positive and negative frequency have meaning in the spacetime under consideration. For these notions to have meaning the geometry must be stationary, or, in fancier language, spacetime must possess a global

[☆] Here a symmetrized form for $T^{\mu\nu}$ is always understood, so that $T^{\mu\nu}(u_1, u_2) = T^{\mu\nu}(u_2, u_1)$.

timelike Killing vector field. It may not admit the Poincaré group, but it must admit at least a one-parameter group of timelike motions.

Gary Gibbons [28] has given the following completely covariant account of the situation that exists when there is a Killing vector K^μ. First of all, the quantity

$$K \equiv -\int_\Sigma K_\mu T^{\mu\nu}\, \mathrm{d}\Sigma_\nu \,, \tag{18}$$

is conserved, i.e., is independent of the Cauchy hypersurface Σ. Secondly, although it is an ill defined operator, K possesses well defined commutation relations with the components of the field:

$$[\varphi, K]_- = \mathrm{i}\mathcal{L}_K \varphi \text{ (modulo gauge transformations, if any)}\,, \tag{19}$$

where $\mathcal{L}$ denotes the Lie derivative. Because the Lie algebra of a single Killing vector is Abelian, the group that it generates is obtainable by simple exponential, and one may choose basic functions u_i that satisfy

$$\mathcal{L}_K u_i = -\mathrm{i}\kappa_i u_i, \qquad \mathcal{L}_K u_i^* = \mathrm{i}\kappa_i u_i^*\,, \tag{20}$$

where the κ_i are constants. If K^μ is globally timelike one may introduce a coordinate t upon which the metric does not depend and with respect to which K^μ takes the form $(K^\mu) = (1, 0, 0, 0)$. Furthermore, K^μ may be scaled so that t gives directly the proper time measured by at least one clock (e.g., a clock at infinity in an asymptotically flat spacetime) whose 4-velocity always remains parallel to K^μ. In that case the functions u_i may be chosen in such a way that the constants κ_i are all positive, and κ_i is called the energy, *relative to that clock*, of a single particle in ith mode. From now on I shall use the symbol ϵ_i in place of κ_i to refer to the single-particle energy, and equations (20) will take the form

$$\partial u_i/\partial t = -\mathrm{i}\epsilon_i u_i, \qquad \partial u_i^*/\partial t = \mathrm{i}\epsilon_i u_i^*\,. \tag{21}$$

The u's and u^*'s are the positive and negative frequency solutions, or positive and negative energy solutions, respectively.

In terms of these basis functions and their associated operators a_i, a_i^*, one can now define a vacuum that *is* a vacuum. One can make the operator K well defined by normal ordering it. I shall denote the results by the symbol E for energy:

$$E = -\int_\Sigma K_\mu :T^{\mu\nu}: \mathrm{d}\Sigma_\nu\,. \tag{22}$$

The vacuum will then be the zero reference point for energy,

$$E|\mathrm{vac}\rangle = 0, \tag{23}$$

and the a_i^*, a_i will be energy-raising and lowering operators:

$$[a_i, E]_- = \epsilon_i a_i\,. \tag{24}$$

If there is another Killing vector L^μ that commutes with K^μ the basis functions may be chosen so as to satisfy also

$$\mathcal{L}_L u_i = -\mathrm{i}\lambda_i u_i\,, \tag{25}$$

where the λ_i are constants. The a_i^*, a_i then become also raising and lowering operators for the associated conserved quantity:

$$[a_i, L]_- = \lambda_i a_i \,, \tag{26}$$

$$L \equiv -\int_\Sigma L_\mu : T^{\mu\nu} : \mathrm{d}\Sigma_\nu \,. \tag{27}$$

More generally, if there is a set of independent Killing vectors generating a Lie algebra, the u_i may be selected to yield irreducible representations of that algebra.

1.3. Failure of conventional procedures

All this is just as in conventional particle physics. The only trouble with it is: *it's wrong*. It is not wrong in a technical mathematical sense. It simply provides a grossly inadequate foundation for the theory. Here are just some of the situations in which it fails:

1. There may be no Killing vector at all, timelike or spacelike. This is the generic situation. How to deal with it is unkown, except possibly when there is an approximate Killing vector that becomes exact asymptotically. It seems most unlikely that the particle picture will prove useful here, except approximately, in regions where quasi-adiabatic conditions hold (which, of course, are very important and typical regions in practice!).

2. There may be a global Killing vector, but it may not be everywhere timelike. In this case two options are available: (a) One may excise the non-timelike region from spacetime. This corresponds to the tacit imposition of a boundary condition. (b) One may retain the non-timelike region but attempt to define a meaningful vacuum by invoking strong physical arguments. I shall give examples of both procedures.

3. Spacetime may be stationary only in limited regions. If each region possesses complete Cauchy hypersurfaces then a local timelike Killing vector field may be set up in each and a vacuum defined for each. Suppose there are two such regions, causally connected. I shall call the earlier region the "in" region and the later region the "out" region and denote their respective vacua by $|\mathrm{in, vac}\rangle$ and $|\mathrm{out, vac}\rangle$. The question now arises: With respect to the basis functions of which region should the stress tensor be normal ordered? (Note that the basis functions once having been defined in each region, can be propagated throughout spacetime, although they will be pure positive or negative frequency functions only in their original domains.) Surely the answer, by the principle of relatively or democracy, or whatever, is *neither*. Neither region should be given preference. Moreover, it is not possible to define the stress tensor so that (a) it is normal ordered in both regions, (b) its matrix elements are smooth functions, and (c) it satisfies the divergence equation

$$T^{\mu\nu}{}_{;\nu} = 0 \,, \tag{28}$$

everywhere. Let us therefore agree here and now that *the stress tensor is always to be left in its un-normal-ordered form* and that we shall only try to regularize it by a subtraction process that respects equation (28). In the present case, for example, we could do the following. Suppose there are no particles present in the "in" region. Then the state vector of the system is $|\mathrm{in, vac}\rangle$. We can proceed to define the following tensor:

$$\langle \mathrm{in, vac}|T^{\mu\nu}|\mathrm{in, vac}\rangle - \langle \mathrm{out, vac}|T^{\mu\nu}|\mathrm{out, vac}\rangle \,. \tag{29}$$

(Here a mode-by-mode subtraction is again implied and a well-defined prescription for effecting it can be given.) This tensor describes the distribution and flow of energy of the particles in the "out" region that have been produced by the nonstationary geometry that lies between the two regions.

The last example illustrates very well the failure of the naive approach, but it also shows that none of the suggested procedures comes close to dealing with the really deep issues of the theory. Consider the tensor (29). Although it describes the physical situation in the "out" region it certainly does no such thing in the "in" region, for it fails to vanish there although no particles are present. In the "in" region it is equal to the *negative* of the tensor that describes the distribution and flow of energy of the particles that would have had to exist in the "in" region in order that the "out" region wind up particle-free. Surely *this* tensor cannot be regarded as the source of the gravitational field. Even in the "out" region it cannot be regarded as the true source, for it only describes the real particles and says nothing about the contribution from virtual particles. Surely there will be effects produced by curvature analogous to the vacuum polarization effects of quantum electrodynamics.

How then can we find the true source? What tensor, formally satisfying eq. (28), can we subtract from $T^{\mu\nu}$ to yield an operator that is mathematically well defined and at the same time describes both dispersive and reactive effects of the interaction between curvature and field? I shall indicate in the final section of this paper some of the proposals that have been made, but first I wish to describe a number of concrete physical examples. There is nothing better than a concrete example to help us get a feel for whether we are doing the right things.

2. The Casimir effect

2.1. A problem in vacuum energy

This well known effect, predicted and popularized by Casimir [10] and experimentally confirmed in the Philips laboratories, has at first sight nothing to do with curvature:

> Two extremely clean, neutral, parallel, microflat conducting surfaces, in a vacuum environment, attract one another by a very weak force that varies inversely as the fourth power of the distance between them.

However, just as curvature can be regarded as a cluttering up of spacetime with bumps, so can the Casimir apparatus be regarded as a cluttering-up of spacetime with neutral conductors. Although the effect was first computed as a kind of Van der Waals force, because the force turns out to be independent of the molecular details of the conductors Casimir quickly recognized that it could be computed as a problem in vacuum energy, and that is the way it is computed in the classroom today. It is true that the tiny energy involved is too small by many orders of magnitude to produce a gravitational field that anybody is going to detect, but one can easily construct Gedankenexperimente in which the law of conservation of energy is violated unless this energy is included in the source of the gravitational field. Relativists should note that the energy density involved is *negative*, and hence the stress tensor violates the classical energy theorems so crucial to black-hole theory*. Everybody should note that the Casimir energy is a pure vacuum energy; no real

* The negativity of the energy appears to be a function of conductor geometry. Boyer [4] and Davies [17] have shown that the vacuum energy inside a conducting sphere is positive.

particles are involved, only virtual ones. And experiments tell us that we have to take it seriously.

As far as I am aware the first person to calculate the actual energy density (i.e., the T^{00} component of the stress tensor), as opposed to the total energy between the conductors, was Larry Ford [23]. Ford's method can be applied to the computation of the other components of the stress tensor as well, and I wish to describe the very beautiful results. But before I do, it will be instructive to review what the formal situation is for the vacuum expectation value of the un-normal-ordered stress tensor in ordinary uncluttered Minkowski space.

2.2. *Regularizing the stress tensor*

The field involved in the present problem, is, of course, the electromagnetic field. The simplest basis functions u_i to introduce are running plane waves with linear polarization. The sum (12) for these waves diverges so we have to regularize it. A useful way from the point of view of axiomatic field theory, as well as heuristically, is to insert into the formal expression for $T^{\mu\nu}$ not the field operators themselves but operators that have been smeared out by means of a smooth function $s(x)$ of compact support:

$$\varphi_s(x) \equiv \int s(x-y)\, \varphi(y)\, \mathrm{d}^4 y \qquad \text{(Minkowski coordinates).} \tag{30}$$

The resulting operator is well defined and the behavior of its (finite) vacuum expectation value may be studied as the size of the support of $s(x)$ tends to zero. The procedure can also be applied in curved spacetime, but in that case the regularized $T^{\mu\nu}$ will not generally satisfy the divergence condition (28) except in the limit. In the present case eq. (28) is trivially satisfied because of the homogeneity of Minkowski space.

I wish to underscore the fact that this method of regularization is frame dependent. s cannot be a Lorentz invariant function of the interval $(x-y)$ and have compact support at the same time. To my mind this enhances its value. So-called covariant regularization schemes have as their only goal the technical elimination of the ambiguous parts of $T^{\mu\nu}$ and are too *ad hoc* to have any particular physical meaning.☆ A frame-dependent method is useful in that it emphasizes what is wrong with $T^{\mu\nu}$ and at the same time allows one to achieve a kind of down-to-earth or heuristic physical insight into the structure of the vacuum energy. Heuristic insights are always helpful when moving into new territory, and I shall emphasize frame dependence again later.

A regularization method equivalent to the smearing method but easier to apply in practice is simply to separate the points at which the two φ's in $T^{\mu\nu}$ are taken and then to examine the tensor as the points are brought together again. This method is obviously frame dependent because the separation interval introduces a preferred direction. I shall choose a timelike separation interval, parallel to the t (or x^0) axis. This is easily seen to be equivalent to introducing an oscillating factor of the form $\exp(\mathrm{i}\epsilon_i/\Lambda)$ into the summand of equation (12), Λ being the reciprocal of the length of the separation interval. Λ is in effect a high-energy cutoff, and the method is identical to the standard procedure for computing the Casimir energy.

☆ A once-popular covariant argument for disposing of $\langle T^{\mu\nu}\rangle_{\mathrm{vac}}$ runs as follows: $\langle T^{\mu\nu}\rangle_{\mathrm{vac}}$ must be a field-and-frame-independent object that transforms as a tensor under Lorentz transformations. The only such objects are multiples of the Minkowski metric $\eta^{\mu\nu}$. The multiplicative factor must vanish in the present case because the Maxwell stress tensor is traceless. Therefore $\langle T^{\mu\nu}\rangle_{\mathrm{vac}} = 0$ even before normal ordering! Clearly the idea of the electromagnetic field as a collection of harmonic oscillators has been totally abandoned here.

The sum (12), with the oscillating factor inserted, is easily evaluated. One finds

$$\langle T^{\mu\nu}\rangle_{\text{vac}} = \frac{3\Lambda^4}{\pi^2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix}. \tag{31}$$

This has exactly the same form as the stress tensor of a photon gas at rest (zero total 3-momentum) in the chosen frame.

Now introduce the parallel conductors and repeat the procedure. The use of a preferred frame is natural in this case because the conductors themselves provide it. (See, however, below.) The only difference from the preceding calculation is that the basis functions u_i are different. One may assume the conductors to be infinite planes. The u_i may then be taken to have the form of running waves parallel to the planes and standing waves in the perpendicular direction. The only tricky part is that one must be sure to impose the boundary conditions appropriate to electric and magnetic fields outside of perfect conductors and not to overlook any of the modes. The vacuum between the plates is no longer the vacuum of uncluttered Minkowski space, because the functions u_i are different. The right hand side of equation (12) reduces from a three-dimensional integral (plus the polarization sums) to a two-dimensional integral and a discrete infinite sum. The result is found to be (for large Λ)

$$\langle T^{\mu\nu}\rangle_{\text{vac}} = \frac{3\Lambda^4}{\pi^2}\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{3} \end{pmatrix} + \frac{\pi^2}{720a^4}\begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -3 \end{pmatrix}, \tag{32}$$

where a is the distance between the conducting surfaces and the x^3 direction is taken perpendicular to the surfaces. I should remark here that I am using units for which $\hbar = c = 1$ and a metric of positive signature: $(\eta_{\mu\nu}) = \text{diag}\,(-1, 1, 1, 1)$.

2.3. Properties of the Casimir stress tensor

Expression (32) has several remarkable properties:

1. The cutoff-dependent part of it is identical with expression (31) for the uncluttered vacuum. This part may therefore be identified, at least tentatively, as an irreducible core that will be found in all matrix elements of the stress tensor under all conditions. Indeed, we shall later find this part popping up in exactly the same form even when curvature is present. [See eq. (249).] Being universal it may be thrown away, leaving, in the present case, a finite remainder. An even better reason for throwing it away in the present case is that the finite part, and only the finite part, is what is observed in the laboratory.

2. The finite remainder is not merely cutoff-independent but also frame independent. To be sure, the conductors themselves determine a preferred x^3 axis, but they leave the x^0, x^1 and x^2

axes entirely arbitrary. The *finite part of* $\langle T_{\mu\nu}\rangle_{\mathrm{vac}}$ *remains unchanged under boosts of arbitrary magnitude in arbitrary directions parallel to the* (x^1, x^2) *plane.* Physically this means that a perfect plane conductor remains a perfect plane conductor in any state of motion parallel to its surface, and that the vacuum stresses in the vicinity of such a conductor look the same no matter how rapidly we are skimming over its surface, a result that would surely have pleased Einstein.

3. Both the finite and divergent parts of $\langle T^{\mu\nu}\rangle_{\mathrm{vac}}$ satisfy the trace condition $T^{\mu}_{\mu} = 0$.

4. Both the finite and divergent parts are position-independent, i.e., constant and uniform. This property is not *a priori* necessary for the finite part and was a bit of a surprise when first discovered. Invariance of the physical set-up under displacements in the x^0, x^1, and x^2 directions guarantees, of course, that $\langle T^{\mu\nu}\rangle_{\mathrm{vac}}$ will not depend on these coordinates, but it could still depend on x^3. As a matter of fact, the quantities $\langle E^2\rangle_{\mathrm{vac}}$ and $\langle H^2\rangle_{\mathrm{vac}}$, where E and H are the electric and magnetic field vectors, do have, by themselves, an x^3-dependence, which, close to each conductor, takes the form*

$$\left.\begin{aligned} \langle E^2\rangle_{\mathrm{vac}} &= \frac{3\Lambda^4}{\pi^2} - \frac{3}{16\pi^2 z^4} \\ \langle H^2\rangle_{\mathrm{vac}} &= \frac{3\Lambda^4}{\pi^2} + \frac{3}{16\pi^2 z^4} \end{aligned}\right\} \quad z \ll a\,, \tag{33}$$

z being the distance from the conductor. It is only when E and H are put into the combinations in which they appear in the stress tensor that the x^3-dependence disappears. Incidentally, this does not mean that the x^3-dependence is unobservable. In principle it will lead to (very) small x^3-dependent shifts in the energy levels of an atom near a conductor (over and above the shifts due to the atom's image in the conductor!). But it will leave no imprint upon the gravitational field.

5. The relative magnitudes of the (0, 0) and (3, 3) components of the finite part of $\langle T^{\mu\nu}\rangle_{\mathrm{vac}}$, and the form of their dependence on a, are just what they would have to be if the vacuum were a gas confined in the space between the conductors, a gas, to be sure, with bizarre properties – negative energy density, negative pressure (tension) in the x^3 direction, positive pressure in the x^1 and x^2 directions – but a gas that satisfies the termodynamical law

$$\mathrm{d}E = T\mathrm{d}S - p\mathrm{d}V\,, \tag{34}$$

nevertheless. Thus, if one slowly ($\mathrm{d}S = 0$) pulls the conductors apart the work done against the tension shows up exactly as an increase in the vacuum energy. Maxwell would have been pleased with this result. It almost makes one believe in the ether!

If I had been cleverer (or if I had believed in the ether) I would have anticipated all these properties in advance, and then I would have known what form $\langle T^{\mu\nu}\rangle_{\mathrm{vac}}$ must have before I ever sat down to compute it. Symmetry considerations assure that the finite part of $\langle T^{\mu\nu}\rangle_{\mathrm{vac}}$ must be diagonal, with (1, 1) and (2, 2) components being equal. Property 2 requires that the (0, 0) and (1, 1) components be equal in magnitude but opposite in sign. Property 3 then yields the relative magnitude of the (3, 3) component, and, together with the divergence condition (28), implies property 4. The a-dependence finally follows from property 5. The a-dependence can also be obtained from dimensional arguments, provided one has heard of Planck's constant, because $\hbar$ and c

* In rationalized units.

are the only physical constants involved. The only thing left undetermined is the absolute magnitude, and *sign*, of any one of the nonvanishing components. This must be found by computation (or experiment).

One final remark: The vanishing of the finite part of $\langle T^{\mu\nu}\rangle_{\text{vac}}$ as $a \to \infty$ suggests that $\langle T^{\mu\nu}\rangle_{\text{vac}}$ reduces to expression (31) in the infinite halfspace on either side of a *single* plane conductor. This may, in fact, be verified directly by carrying out the sum (12) with basis functions appropriate to such a half-space. Because these basis functions differ from those of empty Minkowski space, the half-space vacuum is still not identical to the uncluttered vacuum. Equations (33), for example, continue to hold.

2.4. The Casimir effect as a problem in manifold structure. The massless scalar field

The method of computation in the above examples, in which we simply pick a set of basis functions appropriate to the desired boundary conditions, underscores the fact that even the Casimir effect is very much a problem of Riemannian manifold structure. In each case we pick a different Riemannian manifold – a slab, a half-space, or Minkowski space – and the properties of the vacuum depend on our choice. This prompts us to ask whether the properties we have found depend primarily on the manifold or are peculiar to the electromagnetic field. To answer this question in general would require the opening up of a whole new line of research. I can only report here on what I have found in the case of one other field, the massless scalar field.

What boundary conditions should one impose at the edges of a slab-manifold in the case of a scalar field? Setting the field equal to zero there would seem to be a natural procedure. And yet this leaves one with an uneasy feeling. What is the analog of a conductor in the case of a scalar field? In electromagnetic theory we know what a conductor is, both from years of experiment and years of model building. We do not hesitate to impose the standard boundary conditions for the electric and magnetic fields, because we know that the theory is consistent on many levels. Indeed, Boyer [5], in his study of the Casimir effect, has suggested that the electromagnetic field is unique – that there is no calculable analog of the Casimir effect for fields of other spin. Well, what are the facts?

To cut a short story even shorter (the calculation is easy) the facts are these: The vacuum expectation value of $T^{\mu\nu}$ inside a slab, with the field required to vanish on the boundary, has the form

$$\langle T^{\mu\nu}\rangle_{\text{vac}} = \frac{3\Lambda^4}{2\pi^2}\begin{pmatrix}1&0&0&0\\0&\frac{1}{3}&0&0\\0&0&\frac{1}{3}&0\\0&0&0&\frac{1}{3}\end{pmatrix} + \frac{\pi^2}{1440a^4}\begin{pmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&-3\end{pmatrix} + \frac{\pi^2}{48a^2}\,\frac{3-2\sin^2(\pi z/a)}{\sin^4(\pi z/a)}\begin{pmatrix}-1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&0\end{pmatrix} \tag{35}$$

where z is the distance from either boundary. Again we have the ubiquitous frame-and-cutoff-dependent term, reduced by a factor 2 now because there are only half as many modes. But instead of one frame-and-cutoff-independent term there are two, quite distinct. The first is just the uniform Casimir stress-energy (reduced by a factor 2), but the second is a new term, having a dependence on position. Both these terms are finite, so what is wrong?

The main thing wrong is that the last term diverges when integrated across the slab and so yields an infinite negative total energy (per unit area) in the slab. Boyer is right, at least in this case. The reason he is right is that it is not quite true that the scalar field has only half as many modes as the electromagnetic field. The electromagnetic field has some modes, in which the magnetic field is constant across the slab (for fixed x^1 and x^2), that have no analog for the scalar field. These modes conspire, in the sum (12), to cancel the z-dependent term in the electromagnetic case.

Well, how about going back to first principles, to decide what the vacuum stress tensor for the scalar field *should* look like, just as we did (after the fact) for the electromagnetic field? What key point in the electromagnetic argument is missing here? It is the fact that the condition $T^{\mu}_{\mu} = 0$ no longer holds. Aha! Then we should use the *conformally invariant* scalar field, whose stress tensor does satisfy this condition (see eqs. (232) and (233)). Indeed this does the trick. A straightforward calculation shows that for the conformally invariant scalar field the last term of equation (35) is missing. So Boyer is wrong after all.

But what about the mode-counting argument? In the absence of curvature the basis functions u_i are the same no matter which stress tensor we use. Moreover the two tensors differ from one another by a gradient and hence should yield the same total energy. But in point of fact they don't. The energy integrals differ by surface terms on the boundary, and these are what make the difference.

The success of the conformally invariant theory in this case, and the fact that it mimicks the electromagnetic results so well, gives one a measure of confidence in using it in more general problems, and in believing that the results obtained for such problems will, when spin dependent effects do not dominate, agree at least qualitatively, and very often quantitatively, with the results for the same problems using the electromagnetic field. Because the scalar field is so much easier to work with I shall stick with it from now on.

3. Accelerating conductors

3.1. Particle production by moving boundaries

The Casimir effect may be called a pre-curvature effect of manifold structure. Before going on to discuss true curvature effects let me follow Einstein's example by first discussing effects caused by acceleration. In applying the thermodynamical law (34) to the Casimir vacuum stress I required that the conductors be moved slowly. If I were to accelerate them appreciably they would emit photons, and the entropy in the slab region would be increased. It may seem surprising at first that by accelerating a *neutral* conductor one can produce photons, but then one quickly remembers that the surface layers of a real conductor carry currents. The free electrons near the surface react to the quantum fluctuations of the electromagnetic field just as they do to a classical field and produce currents of just the required amount to guarantee the standard boundary conditions. Because the boundary conditions suffice to determine the physics outside the conductors one need not refer to the currents, as such, at all.

To see how this works in practice consider, for simplicity, a massless scalar field in a flat spacetime of two dimensions. (In two dimensions this field is automatically conformally invariant.) Introduce Minkowski coordinates x and t. Suppose a conductor, or barrier, is present and that the

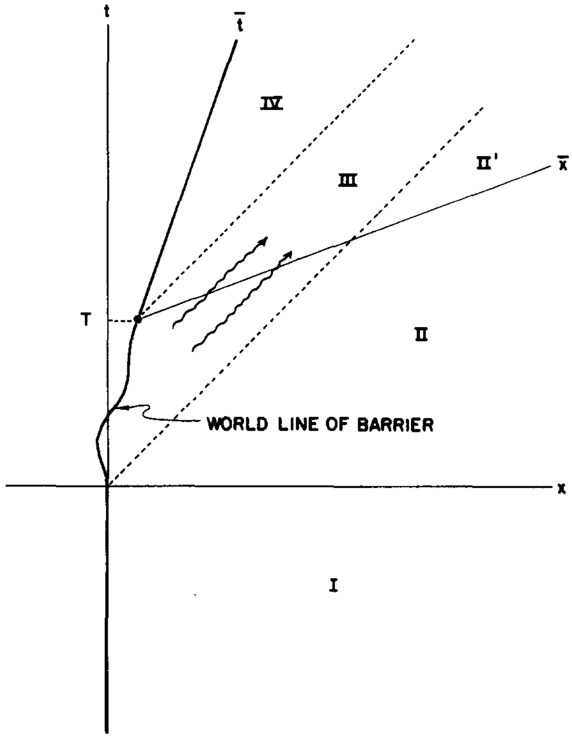


Fig. 1. Particle production by a moving barrier.

world line of its surface is given by a smooth function $z(t)$ with $|\dot{z}(t)| < 1$. Let the region to the "right" of the barrier be uncluttered. Suppose the surface (a point, actually) remains at rest at the origin until the time $t = 0$, then proceeds to undergo various accelerations until $t = T$, and remains thereafter in a state of uniform motion. The spacetime situation is depicted in fig. 1.

In regions I, II, and II′ the natural basis functions to use are

$$u(t, x|\epsilon) = \frac{1}{\sqrt{\pi\epsilon}} \sin(\epsilon x)\, e^{-i\epsilon t}, \qquad 0 < \epsilon < \infty. \tag{36}$$

In regions II′, III and IV the natural basis functions are

$$\bar{u}(t, x|\bar{\epsilon}) \equiv u(\bar{t}, \bar{x}|\bar{\epsilon}) = \frac{1}{\sqrt{\pi\bar{\epsilon}}} \sin(\bar{\epsilon}\bar{x})\, e^{-i\bar{\epsilon}\bar{t}}, \tag{37}$$

where $\bar{x}$ and $\bar{t}$ are related to x and t by a Lorentz transformation:

$$\begin{aligned} \bar{x} &= (1 - v^2)^{-1/2}[x - z(T) - v(t - T)] \\ \bar{t} &= (1 - v^2)^{-1/2}[t - T - v(x - z(T))] \end{aligned} \qquad v = \dot{z}(T). \tag{38}$$

Both sets of basis functions may be propagated to the remaining regions, but then they lose their simple forms. The propagation is very easily effected in two dimensions. For example, the first basis will have the general form

$$u(t, x|\epsilon) = f_\epsilon(x - t) + g_\epsilon(x + t)\,, \tag{39}$$

everywhere, and so will the second basis. The function g_ϵ is already completely determined by region I:

$$g_\epsilon(x) = -\frac{1}{2\mathrm{i}\sqrt{\pi\epsilon}}\,\mathrm{e}^{-\mathrm{i}\epsilon x}\,, \quad -\infty < x < \infty\,. \tag{40}$$

The function f_ϵ, however, is determined there (and in regions II and II′ as well) only for positive values of its argument. To get it for negative values of its argument one uses the boundary condition

$$0 = u(t, z(t)|\epsilon) = f_\epsilon(z(t) - t) + g_\epsilon(z(t) + t)\,, \tag{41}$$

which yields

$$f_\epsilon(z(t) - t) = \frac{1}{2\mathrm{i}\sqrt{\pi\epsilon}}\,\mathrm{e}^{-\mathrm{i}\epsilon[z(t)+t]}\,. \tag{42}$$

Suppose the system is initially in the vacuum state. Then the state is $|\mathrm{in, vac}\rangle$, satisfying

$$a(\epsilon)|\mathrm{in, vac}\rangle = 0 \quad \text{for all } \epsilon, \tag{43}$$

where $a(\epsilon)$ is the annihilation operator associated with the basis function $u(t, x|\epsilon)$. This is not the vacuum state vector relative to the basis functions $\bar{u}(t, x|\bar{\epsilon})$. The two bases are related by a Bogoliubov transformation which can be determined, for each function $z(t)$, by a straightforward (but tedious) computation making use in regions II′, III and IV, of equations (39), (40) and (42) and the orthonormality properties of the basis functions $\bar{u}$. The propagated basis functions u have a distinct form in each of the regions II′, III and IV. Region III is the region of "photon" production. It is where the positive and negative frequencies get mixed. In region IV equilibrium is reestablished and the new vacuum reigns; the functions u revert to their pure positive-frequency status, but each now carries two frequencies: the original frequency and the Doppler shifted frequency obtained by bouncing the primary wave off the moving barrier.

3.2. Constant acceleration

There is one particular accelerated motion that the barrier can execute for which a state of the field exists that remains in equilibrium at all times, namely, constant (absolute) acceleration forever. This case, which is conveniently studied in a Rindler-type coordinate system [42], was first analyzed by Fulling [24]. Let

$$t = \mathrm{e}^{\xi}\sinh\eta, \quad x = \mathrm{e}^{\xi}\cosh\eta. \tag{44}$$

Then the Minkowski line element may be rewritten in the form

$$\mathrm{d}s^2 = \mathrm{e}^{2\xi}(-\mathrm{d}\eta^2 + \mathrm{d}\xi^2)\,, \tag{45}$$

which is seen to be conformally related to the standard form. The new coordinates, however, span only the region of (two dimensional) Minkowski space for which $x > |t|$. The world line of the barrier will be given by $\xi = \zeta = \text{constant}$. The magnitude of its absolute acceleration is $\mathrm{e}^{-\zeta}$. More generally, any observer who remains at a constant fixed ξ will have an absolute acceleration equal to $\mathrm{e}^{-\xi}$.

The reason that an equilibrium state for the field can exist in this case is that the submanifold $x > |t|$ possesses a globally timelike Killing vector field parallel to the world line of the barrier, namely the (contravariant) vector $(1, 0)$ in the (η, ξ) system. This vector field is not globally timelike in the full Minkowski space but becomes null on the line $x = |t|$, $\xi = -\infty$.

The normalized basis functions for this system are identical in form with those for a barrier at rest (eq. (36)) in virtue of the conformal invariance of the theory. They are

$$u(\eta, \xi|\epsilon) = \frac{1}{\sqrt{\pi\epsilon}} \sin(\epsilon\xi)\, e^{-i\epsilon\eta}, \qquad 0 < \epsilon < \infty . \tag{46}$$

If one gives the barrier an infinite acceleration, by pushing it off to the edge of the manifold ($\xi = -\infty$), then running waves become appropriate:

$$u(\eta, \xi|p) = \frac{1}{2\sqrt{\pi\epsilon}}\, e^{i(p\xi - \epsilon\eta)}, \qquad -\infty < p < \infty, \qquad \epsilon = |p| . \tag{47}$$

This is because a barrier at the edge of the manifold, moving with the speed of light, can never bounce a wave back into the manifold. As Rindler [42] has emphasized, this property of the line $x = |t|$ is analogous to that of the event horizon in black-hole theory, and I shall be using just such running waves when I presently discuss black holes.

Fulling [24] has generalized the basis functions (47) to the case of massive particles and has computed the Bogoliubov transformation coefficients between these functions and the standard Minkowski plane-wave basis. The analysis is a little more complicated in the case of a barrier having finite acceleration, but in both cases the β coefficients are nonvanishing. These coefficients (or rather their analogs when the Minkowski basis is replaced by that appropriate to a uniformly moving barrier) become physically significant under the following conditions: Suppose the acceleration of the barrier suddenly drops to zero. Then the β's give directly the number of particles produced. An analogous but more homely situation is the following. Suppose a finite-temperature gas is allowed to come to equilibrium above a platform undergoing constant upward acceleration. If the acceleration is abruptly stopped there will suddenly be a lot of phonons around.

3.3. *The vacuum stress factor*

One may ask the question: What does the vacuum stress tensor look like above an accelerating platform? In the two dimensional case the question may be rephrased: What does the sum (12) give when we use the functions (46)? In the case of a barrier at rest the vacuum tensor reduces to that of uncluttered Minkowski space, just as in the 4-dimensional case. Because of the conformal invariance of the 2-dimensional theory one expects the same to be true for an accelerated barrier, and it is. But here we run into a new problem. If we insert an oscillating factor of the form $\exp(i\epsilon/\Lambda)$ into the summand of expression (12) we find, for the tensor *density*,

$$\langle T^{\mu\nu}\rangle_{\text{vac}} = e^{-2\xi}\, \frac{\Lambda^2}{2\pi} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{48}$$

This same form holds also in a local Lorentz frame, with time axis parallel to the lines of constant ξ. But this means that the vacuum stress vanishes as $\xi \to \infty$, under the Λ-regularization scheme, something it does *not* do in the unaccelerated case. The reason for the phenomenon is that ϵ in eq. (46) has the significance of a local particle energy only at $\xi = 0$. Anywhere else the local energy is $e^{-\xi}\epsilon$ because of the relation $d\eta = e^{-\xi}\, d\tau$ between η and the proper time τ at constant ξ. The cutoff Λ therefore refers not to a local energy but to a Doppler shifted one. If we agree to use a Λ that varies with position in such a way as to give always the same local cutoff energy, then equation (48) will be replaced by

$$\langle T^{\mu\nu}\rangle_{\text{vac}} = \frac{\Lambda^2}{2\pi}\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}, \tag{49}$$

and the ubiquitous zero-point energy will be recognized for what it is. Actually, in practice it does not matter which scheme we use as long as we are aware of the phenomenon. Some people might prefer expression (48) because it satisfies the divergence condition (28) which, in the present context takes the form

$$\partial T^{\xi\xi}/\partial\xi + T^{\xi\xi} + T^{\eta\eta} = 0\,. \tag{50}$$

I have not yet been able to compute successfully the form of the vacuum stress tensor above an accelerating barrier in the 4-dimensional case. This case is not conformally equivalent to that of a barrier at rest,* and hence there is no *a priori* reason to rule out a finite, and hence physically significant, addition to the usual divergent stress (31). The technical difficulty is that the basis functions become Bessel functions of a form already encountered by Fulling [24] in the two dimensional massive case, and there is discrete quantization in the ξ direction. The reason for the latter is that any photon, except one that is aimed vertically "upward", ultimately falls back to the barrier, and hence every orbit has a turning point of maximum ξ.

Note added. After this paper was written my attention was called to a valuable article by Moore [35] on the quantum theory of the electromagnetic field in a variable-length one-dimensional cavity. Moore studies the problem of *two* moving barriers and gives: (a) a careful statement of the mathematical structure of the corresponding quantum field theory, and (b) a method for finding a wide class of barrier motions admitting exact solutions of the problem, some of which are of considerable physical and conceptual interest.

4. The Kerr black hole

4.1. Geometrical preliminaries. Ergosphere and horizon

Now let us look at curved manifolds. I shall begin with a particularly exotic one, the Kerr black hole, because it illustrates well a great new range of problems. The line element is

$$ds^2 = -\frac{\Delta}{\rho^2}(dt - a\sin^2\theta\, d\phi)^2 + \frac{\sin^2\theta}{\rho^2}[(r^2+a^2)\,d\phi - a\,dt]^2 + \frac{\rho^2}{\Delta}dr^2 + \rho^2\, d\theta^2\,, \tag{51}$$

$$\rho \equiv \sqrt{r^2 + a^2\cos^2\theta}\,,\qquad \Delta \equiv r^2 - 2Mr + a^2\,, \tag{52}$$

* Conformal equivalence holds for the field around a single uniformly accelerating point source but not for the field above an entire accelerating barrier.

and I shall have to spend a moment taking note of some of its properties.* By examining it at $r \to \infty$ one finds that it corresponds to a source of mass M and spin angular momentum $J = Ma$. When the constant a is set equal to zero it reduces to the Schwarzschild line element. It is believed to be the unique metric that results, after gravitational radiation has died away, when gravitating matter undergoes catastrophic collapse through an event horizon. It is also believed that a can never be greater than M.

To locate the event horizon it is helpful first to note that the metric is independent of the coordinates t and ϕ. Hence there are two independent Killing vectors $(\xi_t^\mu) = (1, 0, 0, 0)$ and $(\xi_\phi^\mu) = (0, 0, 0, 1)$. (The coordinates are assumed to be numbered in the order t, r, θ, ϕ.) By direct computation one finds that they satisfy

$$\xi_t^2 = -(1 - 2Mr/\rho^2)\,, \tag{53}$$

$$(\xi_t \cdot \xi_\phi)^2 - \xi_t^2 \xi_\phi^2 = \Delta \sin^2 \theta\,. \tag{54}$$

ξ_t is evidently a timelike Killing vector over most of the manifold, and t coincides at $r \to \infty$ with a standard time coordinate. But ξ_t is not globally timelike. It becomes null on the surface of the so-called *ergosphere*, located at $r = M + \sqrt{M^2 - a^2 \cos^2 \theta}$, and is spacelike between this surface and another one located at $r = M - \sqrt{M^2 - a^2 \cos^2 \theta}$. Neither of these surfaces marks the horizon. Inside the outer surface, for example, there still exist timelike vectors that point in the direction of increasing r. In determining the boundary at which such vectors cease to exist it is sufficient to determine where timelike vectors having only t and ϕ components cease to exist. (A small positive r component can always be added to such a vector without destroying its timelike character.) Thus we consider combinations of ξ_t and ξ_ϕ of the form $\xi_t + \Omega\xi_\phi$, where Ω is a function of r, and possibly also of θ. In order that the combined vector be timelike Ω must lie in the range $\Omega_- < \Omega < \Omega_+$, where

$$\Omega_\pm = \xi_\phi^{-2}\left(-\xi_t \cdot \xi_\phi \pm \sqrt{(\xi_t \cdot \xi_\phi)^2 - \xi_t^2 \xi_\phi^2}\,\right). \tag{55}$$

Ω is the angular velocity that a spaceship would have (as seen from infinity) if its world line were parallel to the vector $\xi_t + \Omega\xi_\phi$. The lower bound, Ω_-, vanishes on the surface of the ergosphere where $\xi_t^2 = 0$. Inside this surface the spaceship cannot "sit still"; i.e., it cannot remain at a fixed r, θ and ϕ, but is forced, by the Lense–Thirring frame-dragging effect, to revolve about the black hole. Note that if Ω is taken, for the moment, independent of r and θ, then $\xi_t + \Omega\xi_\phi$ is a Killing vector. This means that the geometry still appears stationary to the crew of the spaceship.

If the spaceship reaches the surface where Ω_+ and Ω_- coincide, its cone of options (i.e., light cone) is narrowed down to nothing, at least from the point of view of the t, r, θ, ϕ coordinate system, which becomes singular on this surface. From then on it cannot escape but can only head on into regions of decreasing r, where the geometry is necessarily dynamic (no timelike Killing vector). This surface is the horizon. Its position is determined by the vanishing of the radical in eq. (55), which, in view of eq. (54), is equivalent to the vanishing of Δ. Δ has two roots, $r_\pm$, given by

$$r_\pm = M \pm \sqrt{M^2 - a^2}\,. \tag{56}$$

* For further details on black holes the reader may wish to consult a general reference, e.g., "Black Holes," eds. DeWitt and DeWitt (Gordon and Breach, New York, London, Paris, 1973).

It is the larger one that is relevant.

A crucial quantity in what follows is the value of Ω_+ and Ω_- when they coincide. This is known as the angular velocity of the horizon itself, and is found, after a little algebra to be given by

$$\Omega_{\mathrm{H}} = \frac{a}{2Mr_+} = \frac{a}{r_+^2 + a^2} \,. \tag{57}$$

Other useful quantities are the following:

$$g^{1/2} = \rho^2 \sin\theta, \tag{58}$$

$$g^{\mu\nu}\frac{\partial^2}{\partial x^\mu \partial x^\nu} = -\frac{1}{\Delta\rho^2}\left[(r^2+a^2)\frac{\partial}{\partial t} + a\frac{\partial}{\partial\phi}\right]^2 + \frac{1}{\rho^2 \sin^2\theta}\left(\frac{\partial}{\partial\phi} + a\sin^2\theta\,\frac{\partial}{\partial t}\right)^2 + \frac{\Delta}{\rho^2}\frac{\partial}{\partial r^2} + \frac{1}{\rho^2}\frac{\partial^2}{\partial\theta^2} \,. \tag{59}$$

The operator $\overleftrightarrow{f}^\mu$ for the scalar wave equation takes the form

$$g^{1/2}g^{\mu\nu}\frac{\overrightarrow{\partial}}{\partial x^\nu} - \frac{\overleftarrow{\partial}}{\partial x^\nu}g^{1/2}g^{\mu\nu} \,, \tag{60}$$

which, for a hypersurface t = constant in the Kerr geometry, becomes

$$\frac{\sin\theta}{\Delta}\left\{[(r^2+a^2) - \Delta\, a^2 \sin\theta]\frac{\overleftrightarrow{\partial}}{\partial t} + 2Mra\frac{\overleftrightarrow{\partial}}{\partial\phi}\right\} \,. \tag{61}$$

Only the region outside the horizon will be needed in the construction and normalization of our basis functions, and hence the domain of integration over this hypersurface is $r_+ < r < \infty$, $0 \leqslant \theta \leqslant \pi$, $0 \leqslant \phi \leqslant 2\pi$. Because the curvature scalar of the Kerr geometry vanishes (it is a major undertaking to verify this by direct computation!) there is no difference between the conformal and ordinary scalar wave equations. Nor shall we find any practical difference between the two stress tensors in the asymptotic region $r \to \infty$, which is the only region where we shall attempt to compute them.

4.2. Absolute units

Before going farther it should be pointed out that we are now working in units for which $G = c = 1$, G being the gravity constant. When we presently start quantizing I shall add the condition $\hbar = 1$. Then we shall have unitless units, or absolute units. It will be useful to remember that the absolute units of length, time and mass respectively are 1.6×10^{-33} cm, 5×10^{-45} sec and 2×10^{-5} g. In these units the mass of the proton is 8×10^{-20}, the mass of the sun is 10^{38}, and the size, age and mass of the universe are 10^{62}.

4.3. Basis functions

It is a remarkable fact that the wave equation

$$g^{-1/2}\frac{\partial}{\partial x^\mu}g^{1/2}g^{\mu\nu}\frac{\partial}{\partial x^\nu}\varphi = 0, \tag{62}$$

is separable in the Kerr metric. The basis functions may be taken in the form

$$u(l, m, p|x) = N(p)(r^2 + a^2)^{-1/2} R_{lm}(p, a|r)\, S_{lm}(a\epsilon|\cos\theta)\, e^{im\phi}\, e^{-i\epsilon t}\,, \tag{63}$$

where N is a normalization constant and p is a certain function of ϵ (both presently to be determined). S_{lm} is a spheroidal harmonic satisfying the eigenvalue equation

$$\left[\frac{d}{d\xi}(1-\xi^2)\frac{d}{d\xi} - \frac{m^2}{1-\xi^2} + 2ma\epsilon - (a\epsilon)^2(1-\xi^2) + \lambda_{lm}(a\epsilon)\right] S_{lm}(a\epsilon|\xi) = 0\,, \tag{64}$$

the eigenvalue $\lambda_{lm}(a\epsilon)$ depending in an unfortunately nontrivial way on the integers $l(= 0, 1, 2...)$ and $m(= -l, -l+1 \,...\, l-1, l)$ and on its argument $a\epsilon$, but having the well known boundary value $\lambda_{lm}(0) = l(l+1)$. Before writing the differential equation satisfied by the "radial" function it will be convenient first to introduce the new coordinate

$$r^* = r + \frac{M}{\sqrt{M^2-a^2}}\left(r_+ \ln\frac{r-r_+}{r_+} - r_- \ln\frac{r-r_-}{r_-}\right)\,, \tag{65}$$

satisfying

$$dr^*/dr = (r^2 + a^2)/\Delta\,. \tag{66}$$

This new coordinate ranges over the entire real line, pushing the horizon off to minus infinity. In terms of it the "radial" equation takes the simple from

$$\left[\frac{d^2}{dr^{*2}} - V_{lm}(\epsilon, a|r)\right] R_{lm}(p, a|r) = 0\,, \tag{67}$$

where

$$V_{lm}(\epsilon, a|r) = -\left(\epsilon - m\frac{a}{r^2+a^2}\right)^2 + \lambda_{lm}(a\epsilon)\frac{\Delta}{(r^2+a^2)^2} + \frac{2(Mr-a^2)\Delta}{(r^2+a^2)^3} + \frac{3a^2\Delta^2}{(r^2+a^2)^4}\,. \tag{68}$$

We shall need the function R only in the asymptotic regions $r^* \to \pm\infty$. In these regions the function V reduces to

$$V_{lm}(\epsilon, a|r) \to \begin{cases} -(\epsilon - m\Omega_H)^2, & r^* \to -\infty \\ -\epsilon^2, & r^* \to \infty \end{cases} \tag{69}$$

In the intervening region V acts as a potential barrier, causing back-scattering. Because of the existence of the horizon a "radial" wave may originate from, propagate out of, and be scattered back into either asymptotic region. We therefore distinguish two classes of solutions of eq. (67), having, in virtue of (69), the asymptotic forms

$$\vec{R}_{lm}(p, a|r) \to \left\{\begin{matrix} \exp(ipr^*) + \vec{A}_{lm}(p, a)\exp(-ipr^*), & r^* \to -\infty \\ \vec{B}_{lm}(p, a)\exp\{i(p + m\Omega_H)r^*\}, & r^* \to \infty \end{matrix}\right\}, \quad p > 0, \quad \epsilon = p + m\Omega_H \tag{70}$$

$$\overleftarrow{R}_{lm}(p, a|r) \to \left\{\begin{matrix} \overleftarrow{B}_{lm}(p, a)\exp\{i(p - m\Omega_H)r^*\}\,, & r^* \to -\infty \\ \exp\{-ipr^*\} + \overleftarrow{A}_{lm}(p, a)\exp(ipr^*), & r^* \to \infty \end{matrix}\right\}, \quad p > 0, \quad \epsilon = p\,. \tag{71}$$

From now on I shall place arrows also over the corresponding basis functions.

The normalization of the basis functions, as is well known, is dominated by the asymptotic regions from which the waves originate. One may construct very broad wave packets confined to these regions at early times. Inserting such wave packets into the integral of eq. (6), using expression (61) and eq. (66), and passing to the limit of infinitely broad packets, one finds that the basis functions (63) satisfy the orthonormality relations

$$\langle \vec{u}(l, m, p), \vec{u}(l', m', p')\rangle = \langle \overleftarrow{u}(l, m, p), \overleftarrow{u}(l', m', p')\rangle = \delta_{ll'}\delta_{mm'}\delta(p - p') , \tag{72}$$

and their complex conjugates (all other inner products vanishing), provided one normalizes the spheroidal harmonics according to

$$\int_{-1}^{1} S_{lm}(a\epsilon|\xi)\, S_{l'm}(a\epsilon|\xi)\, \mathrm{d}\xi = \delta_{ll'} , \tag{73}$$

and chooses the normalization constant to be

$$N(p) = \frac{1}{2\pi\sqrt{2p}} . \tag{74}$$

4.4. Past and future horizons and the vacuum state

A few comments now about the role of the horizon: I made a statement earlier about a spaceship getting trapped inside and unable to get out, as if the horizon were a one-way membrane. This is only half of the story. Because of the invariance of the line element (51) under simultaneous inversion of t and ϕ there must be another horizon from which matter (or radiation) can only escape, without being able to return. This is known as the *past* horizon. The other is the *future* horizon. Equation (70) represents a wave that has originated in the past horizon. Part of it gets scattered into the future horizon and part of it escapes to infinity. Equation (71) represents a wave that has originated at infinity. Part of it gets scattered back to infinity and part of it winds up inside the future horizon.

I am going to choose for the "vacuum" state of the Kerr black hole the vacuum defined in the normal way relative to the basis functions (63), with (70) and (71) as my radial functions. In this "vacuum" there are no particles present that have originated from infinity, and there are none that have originated from the past horizon. That there should be no particles coming from infinity seems reasonable enough, because spacetime at infinity is ordinary familiar spacetime, and that is just what we should expect of a sensible vacuum. But that there should be no particles coming from the past horizon is a dubious assumption at best, at least as a model for a real black hole, for we believe that all real black holes (if any exist) were formed by a process of collapse, and for such black holes *there is no past horizon*. Indeed we shall see in the next section that taking the collapse process into account leads to quite different boundary conditions and to an important modification in our results. However, for the present I shall leave the basis functions and "vacuum" as is. The formalism then at least has the merit of looking like that of a standard scattering problem and hence is familiar.

4.5. Superradiance

I now wish to draw attention to a small but crucial point. Look at equation (70). If the azimuthal quantum number m satisfies $m\Omega_H < -p$ then ϵ is negative, and we appear to be dealing with a negative-energy wave! Does this not violate our basic principles? The answer is *no*. Our "principles" were founded upon the assumed existence of a globally timelike Killing vector field. We do not have such a field in the present case; ξ_t is spacelike in the ergosphere. The wave function (70) is, in fact, just what it should be. To an observer following a timelike world line near the horizon it appears as a positive-energy wave of angular frequency p. Two other facts should be noted: (a) The group velocity is radially outward for the initial and transmitted waves and inward for the reflected wave, as it should be; (b) An attempt to replace the basis function $\vec{u}$ with its complex conjugate would lead to a violation of the orthonormality relations (72).

However, the phenomenon should raise a warning flag in our minds. The mathematical situation is analogous to that which holds for a two dimensional harmonic oscillator with *negative* spring constant, carrying a charge and immersed in a uniform magnetic field. If the magnetic field is strong enough all the orbits will be stable. However, one of the two annihilation operators for the system is associated with a negative-frequency mode, and *there is no state of lowest energy.*

That there is likewise no state of lowest energy for a scalar field in the Kerr geometry emerges from the following analysis due to Misner [34] and Zel'dovich [50,51]. By making use of the constancy of the Wronskian

$$R_1 \frac{\mathrm{d}R_2}{\mathrm{d}r^*} - \frac{\mathrm{d}R_1}{\mathrm{d}r^*} R_2 \,, \tag{75}$$

for various combinations of the radial wave functions (70), (71), and their complex conjugates, one finds that the transmission and reflection amplitudes satisfy the following relations:

$$1 - |\vec{A}_{lm}(p,a)|^2 = \frac{p + m\Omega_H}{p} |\vec{B}_{lm}(p,a)|^2 \,, \tag{76}$$

$$1 - |\overleftarrow{A}_{lm}(p,a)|^2 = \frac{p - m\Omega_H}{p} |\overleftarrow{B}_{lm}(p,a)|^2 \,, \tag{77}$$

$$p\vec{A}^*_{lm}(p,a)\, \overleftarrow{B}_{lm}(p + m\Omega_H, a) = -(p + m\Omega_H)\, \vec{B}^*_{lm}(p,a)\, \overleftarrow{A}_{lm}(p + m\Omega_H, a) \,, \tag{78}$$

$$(p + m\Omega_H)\, \vec{B}_{lm}(p,a) = p\, \overleftarrow{B}_{lm}(p + m\Omega_H, a) \,, \tag{79}$$

from which we also obtain

$$|\vec{A}_{lm}(p,a)| = |\overleftarrow{A}_{lm}(p + m\Omega_H, a)| \,. \tag{80}$$

The important relations are the first two. If $m\Omega_H < -p$ then a wave originating in the past horizon is reflected back *with a greater amplitude than it had initially*. The same is true for a wave originating at infinity if $m\Omega_H > p$.

Misner has called this phenomenon *superradiance*. It is not a new phonomenon in physics. It was known already to an older generation of physicists, who called it the *Klein paradox*. Physically it corresponds to a process of stimulated emission, which suggests immediately the existence of a corresponding process of spontaneous emission. And indeed the latter process occurs. The Kerr "vacuum" is not the state of lowest energy. It spontaneously emits pairs, one particle of each pair going to infinity and the other into the black hole. As a result of this steady process the black hole must gradually lose its angular momentum. A classical phenomenon akin to this was first noted by Penrose [41] who pointed out that a particle that falls into the ergosphere can decay into two particles, one of which goes down the black hole while the other escapes from the ergosphere with greater energy than was possessed by its parent particle. In this way energy can be extracted from the black hole at the expense of its angular momentum (and, of course, of its mass). It was this process that first led to the coining of the word *ergosphere*.

4.6. Particle flux from a Kerr black hole

The first attempt to calculate the rate of particle flux from a Kerr black hole, by combining the transmission and reflection amplitudes with the idea of stimulated emission, was made by Starobinsky [45] (see also Starobinsky and Churilov [46]). The first full fledged computation, based on second quantization, was made by William Unruh [47]. Because Unruh makes crucial use of the stress tensor, it is particularly appropriate that we study his results.

We shall need the (r, t) and (r, ϕ) components of the stress tensor at infinity, as these yield the flux of energy and angular momentum there. I shall take $T^{\mu\nu}$ in its density form. For the conformally invariant theory we have

$$\langle T_{rt}\rangle_{\text{vac}} = \rho^2 \sin\theta \left[\tfrac{1}{3}\left\langle\left[\frac{\partial\varphi}{\partial r}, \frac{\partial\varphi}{\partial t}\right]_+\right\rangle_{\text{vac}} - \tfrac{1}{6}\left\langle\left[\varphi, \frac{\partial^2\varphi}{\partial r\,\partial t}\right]_+\right\rangle_{\text{vac}}\right] \text{ plus terms that vanish at infinity.} \tag{81}$$

The first and second terms inside the square brackets yield, apart from the numerical factors, identical contributions at infinity. Therefore we have

$$\langle T_{rt}\rangle_{\text{vac}} \xrightarrow[r\to\infty]{} \tfrac{1}{2} r^2 \sin\theta \left\langle\left[\frac{\partial\varphi}{\partial r}, \frac{\partial\varphi}{\partial t}\right]_+\right\rangle_{\text{vac}}, \tag{82}$$

which is just what the ordinary (non-conformally invariant) tensor gives. Similarly

$$\langle T_{r\phi}\rangle_{\text{vac}} \xrightarrow[r\to\infty]{} \tfrac{1}{2} r^2 \sin\theta \left\langle\left[\frac{\partial\varphi}{\partial r}, \frac{\partial\varphi}{\partial t}\right]_+\right\rangle_{\text{vac}}. \tag{83}$$

Each of these expressions immediately converts to a mode-sum, as in equation (12). Inserting the basis functions (63), (70), (71) into (82), for example, one finds

$$\langle T_{tr}\rangle_{\text{vac}} \xrightarrow[r\to\infty]{} \frac{\sin\theta}{8\pi^2}\sum_{l,m}\int_0^\infty \{-(p+m\Omega_{\text{H}})^2 |\vec{B}_{lm}(p,a)|^2 [S_{lm}(a\vec{\epsilon}|\cos\theta)]^2 + p^2(1-|\overleftarrow{A}_{lm}(p,a)|^2)[S_{lm}(a\overleftarrow{\epsilon}|\cos\theta)]^2\}\,\frac{\mathrm{d}p}{p}. \tag{84}$$

At this point we should regularize the tensor by inserting an oscillating factor. But it turns out that we do not have to if we recognize that, in virtue of equations (77) and (79), a mode-by-mode cancellation occurs between the first and second terms inside the curly brackets. The cancellation occurs for all modes except those in the first term for which $p < -m\Omega_{\text{H}}$ and those in the second term for which $p < m\Omega_{\text{H}}$.* These are just the superradiant modes. The $\rightarrow$ and $\leftarrow$ modes may be combined with the aid of eq. (80), and we are left with

* To see this easily make the shift $p \to p + m\Omega_{\text{H}}$ in the second term.

$$\langle T_{rt}\rangle_{\text{vac}} \underset{r\to\infty}{\longrightarrow} -\frac{\sin\theta}{4\pi^2} \sum_{\substack{l,m\\ m\Omega_{\mathrm{H}}>0}} \int_0^{m\Omega_{\mathrm{H}}} p(|\overleftarrow{A}_{lm}(p,a)|^2 - 1)[S_{lm}(a\overleftarrow{\epsilon}|\cos\theta)]^2\,\mathrm{d}p. \tag{85}$$

In a similar way one finds

$$\langle T_{r\phi}\rangle_{\text{vac}} \underset{r\to\infty}{\longrightarrow} \frac{\sin\theta}{4\pi^2} \sum_{\substack{l,m\\ m\Omega_{\mathrm{H}}>0}} m \int_0^{m\Omega_{\mathrm{H}}} (|\overleftarrow{A}_{lm}(p,a)]^2 - 1)[S_{lm}(a\overleftarrow{\epsilon}|\cos\theta)]^2\,\mathrm{d}p\,. \tag{86}$$

If we now equate the integrated fluxes at infinity to the rates at which the black hole loses mass and angular momentum respectively we find, upon making use of (73),

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \lim_{r\to\infty} \int_0^{\pi} \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi\,\langle T_{rt}\rangle_{\text{vac}} = -\frac{1}{2\pi} \sum_{\substack{l,m\\ m\Omega_{\mathrm{H}}>0}} \int_0^{m\Omega_{\mathrm{H}}} p(|\overleftarrow{A}_{lm}(p,a)|^2 - 1)\,\mathrm{d}p\,, \tag{87}$$

$$\frac{\mathrm{d}J}{\mathrm{d}t} = -\lim_{r\to\infty} \int_0^{\pi} \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi\langle T_{r\phi}\rangle_{\text{vac}} = -\frac{1}{2\pi} \sum_{\substack{l,m\\ m\Omega_{\mathrm{H}}>0}} m \int_0^{m\Omega_{\mathrm{H}}} (|\overleftarrow{A}_{lm}(p,a)|^2 - 1)\,\mathrm{d}p\,. \tag{88}$$

These are just the expressions one would use in a calculation that interprets the classical wave amplification as a stimulated emission process and relates it to the idea of spontaneous emission.

4.7. *Rate of decay. Critical mass*

To convert these expressions to numbers it is necessary to estimate the coefficients $\overleftarrow{A}$. Computer calculations have shown that the superradiance phenomenon is not very efficient (except for gravitational waves [45,46]). The reason for this is that when $p < |m\Omega_{\mathrm{H}}|$ the function V of eq. (68) presents a potential barrier having a height that goes roughly like l^2. Using a WKB approximation to estimate the barrier penetration factor one gets

$$|\overleftarrow{A}|^2 - 1 \sim \exp\left(-\int_{r_1^*}^{r_2^*} \sqrt{V}\,\mathrm{d}r^*\right) \sim \mathrm{e}^{-\zeta l}\,, \tag{89}$$

where r_1^* and r_2^* are the turning points and ζ is a number roughly of the order of unity. One sees that the dominant mode in equations (87) and (88) is $l = 1$, $|m| = 1$, and hence

$$\frac{\mathrm{d}M}{\mathrm{d}t} \sim -\frac{\mathrm{e}^{-\zeta}}{4\pi}\Omega_{\mathrm{H}}^2\,, \quad \frac{\mathrm{d}J}{\mathrm{d}t} \sim -\frac{\mathrm{e}^{-\zeta}}{2\pi}\Omega_{\mathrm{H}}\,. \tag{90}$$

At this point it is useful to introduce the so-called *irreducible mass.* It is defined by

$$M_{\mathrm{ir}}^2 = \tfrac{1}{2}Mr_+\,, \tag{91}$$

and has a value between $M/\sqrt{2}$ and M. With the aid of equations (56) and (57) one can easily show that it satisfies the relations (remember $J = Ma$)

$$M_{\mathrm{ir}}^2 + \frac{J^2}{4M_{\mathrm{ir}}^2} = M^2\,, \quad M_{\mathrm{ir}}^2 - \frac{J^2}{4M_{\mathrm{ir}}^2} = M\sqrt{M^2 - a^2}\,, \tag{92}$$

$$\Omega_{\mathrm{H}} = \frac{a}{4M_{\mathrm{ir}}^2} = \frac{J}{4MM_{\mathrm{ir}}^2} \,. \tag{93}$$

Inserting equation (93) into (90) we get

$$\frac{\mathrm{d}J}{\mathrm{d}t} \sim -\frac{\mathrm{e}^{-\zeta}}{8\pi MM_{\mathrm{ir}}^2} J \sim -\frac{\mathrm{e}^{-\zeta}}{8\pi M^3} J \,, \tag{94}$$

which tells us that the half-life for loss of angular momentum by spontaneous emission is

$$\tau \sim 8\pi\, \mathrm{e}^{\zeta}\, M^3 \,. \tag{95}$$

The age of the universe is 10^{62}. This means that for a black hole to have had its angular momentum significantly affected by this process since it was formed its mass must be less than☆

$$\left(\frac{10^{62}}{8\pi\, \mathrm{e}^{\zeta}}\right)^{1/3} \sim 10^{20} = 2 \times 10^{15}\ \mathrm{g} \,. \tag{96}$$

This is a typical asteroid mass. It would be compressed inside a radius smaller than 3×10^{-13} cm.

4.8. Consistency with Hawking's theorem

The importance of the irreducible mass is not to help with the algebra above. It lies in the following differential identity

$$\mathrm{d}M_{\mathrm{ir}} = \frac{M_{\mathrm{ir}}}{\sqrt{M^2 - a^2}} (\mathrm{d}M - \Omega_{\mathrm{H}}\, \mathrm{d}J) \,, \tag{97}$$

which may be derived from eqs. (92). By sending test particles into a black hole in all possible orbits that reach the horizon, and examining the increments $\mathrm{d}M$ and $\mathrm{d}J$ thereby imparted to the hole, Demetrios Christodoulou [14] (see also Chirstodoulou and Ruffini [15]) showed that $\mathrm{d}M_{\mathrm{ir}}$ can never be negative. Simultaneously Stephen Hawking [30] showed quite generally that the surface area of a black-hole (i.e., of the future horizon) can never decrease. The area A of a Kerr black hole (as computed directly from the line element (51)!) is $16\pi M_{\mathrm{ir}}^2$. Therefore Christodoulou's result is a special case of Hawking's theorem and can be restated in the form

$$\mathrm{d}A = \frac{2A}{\sqrt{M^2 - a^2}} (\mathrm{d}M - \Omega_{\mathrm{H}}\, \mathrm{d}J) \geqslant 0 \,. \tag{98}$$

Now consider one of the particles emitted to infinity in the spontaneous emission process. It removes from the black hole an amount of energy p and an amount of azimuthal angular momentum m. Its emission therefore produces the following change in the area of the black hole:

☆ Expression (95) is actually an upper bound to the half-life, for we have taken into account the quanta of only one field. Unruh [47] has shown that neutrinos are produced at a similar rate, and Starobinsky and Churilov [45,46] have shown that photons and gravitons are produced even more copiously. These particles alone already yield a half-life almost two orders of magnitude shorter. Moreover, massive particles will also be produced if their masses are less than $|\Omega_{\mathrm{H}}|$. For an extreme Kerr black hole ($a = M$) of mass 10^{20} the rotation frequency $|\Omega_{\mathrm{H}}|$ is equal to $1/4M = 2.5 \times 10^{-25} = 30$ MeV, so in this case electrons and positrons would be the only massive particles produced. For black holes having masses two or three orders of magnitude smaller, however, the number of particle varities subject to spontaneous emission might increase without limit, leading to explosive loss of energy and angular momentum.

$$\mathrm{d}A = -\frac{2A}{\sqrt{M^2 - a^2}}(p - m\Omega_{\mathrm{H}})\,. \tag{99}$$

But the only particles that get emitted to infinity are those in the ← superradiant modes, and for these $p - m\Omega_{\mathrm{H}}$ is always negative. Therefore the spontaneous emission process respects Hawking's theorem.

5. Exploding black holes

5.1. Late-time basis functions

I now wish to report on some astonishing recent work of Hawking [31,32] who has faced squarely up to the issue of the boundary conditions on the past horizon, by considering what happens in the case of a realistic black hole that is formed by collapse, when there is no past horizon. For simplicity I shall give the details only for the case of nonrotating black holes, for which $a = 0$, $J = 0$, $\Omega_{\mathrm{H}} = 0$, $r_+ = 2M$, and expression (51) reduces to the Schwarzschild line element. To establish continuity with what has gone before let us first note what the basis functions (63), (70), (71) look like in this case. One easily sees that they reduce to☆

$$u(l, m, p|x) = \frac{1}{2\pi\sqrt{2p}\,r} R_l(p|r)\, Y_{lm}(\cos\theta)\, \mathrm{e}^{\mathrm{i}m\phi}\, \mathrm{e}^{-\mathrm{i}\epsilon t}\,, \tag{100}$$

$$\vec{R}_l(p|r) \to \begin{cases} \mathrm{e}^{\mathrm{i}pr^*} + \vec{A}_l(p)\,\mathrm{e}^{-\mathrm{i}pr^*}, & r^* \to -\infty \\ B_l(p)\,\mathrm{e}^{\mathrm{i}pr^*}, & r^* \to \infty \end{cases} \tag{101}$$

$$\overleftarrow{R}_l(p|r) \to \begin{cases} B_l(p)\,\mathrm{e}^{-\mathrm{i}pr^*}, & r^* \to -\infty \\ \mathrm{e}^{-\mathrm{i}pr^*} + \overleftarrow{A}_l(p)\,\mathrm{e}^{\mathrm{i}pr^*}, & r^* \to \infty \end{cases} \tag{102}$$

where $\epsilon = p$ in all cases, and the coordinate r^* and the function V are given by

$$r^* = r + 2M \ln\left(\frac{r}{2M} - 1\right), \tag{103}$$

$$V_l(\epsilon|r) = -\epsilon^2 + \left(1 - \frac{2M}{r}\right)\left[\frac{l(l+1)}{r^2} + \frac{2M}{r^3}\right]. \tag{104}$$

There are no superradiant modes now (no ergosphere), and the identities (76) to (80) reduce to

$$\vec{B}_l(p) = \overleftarrow{B}_l(p) \equiv B_l(p)\,, \tag{105}$$

$$|\vec{A}_l(p)| = |\overleftarrow{A}_l(p)|, \tag{106}$$

$$1 - |\vec{A}_l(p)|^2 = 1 - |\overleftarrow{A}_l(p)|^2 = |B_l(p)|^2\,, \tag{107}$$

$$\vec{A}_l^*(p)\,B_l(p) = -B_l^*(p)\,\overleftarrow{A}_l(p)\,. \tag{108}$$

☆ The function denoted here by Y_{lm} is just $\lim_{a\to\infty} S_{lm}$. It satisfies the normalization condition (73) and hence differs by a constant factor from the function usually denoted by this symbol. Also it does not contain the factor $\exp(\mathrm{i}m\phi)$.

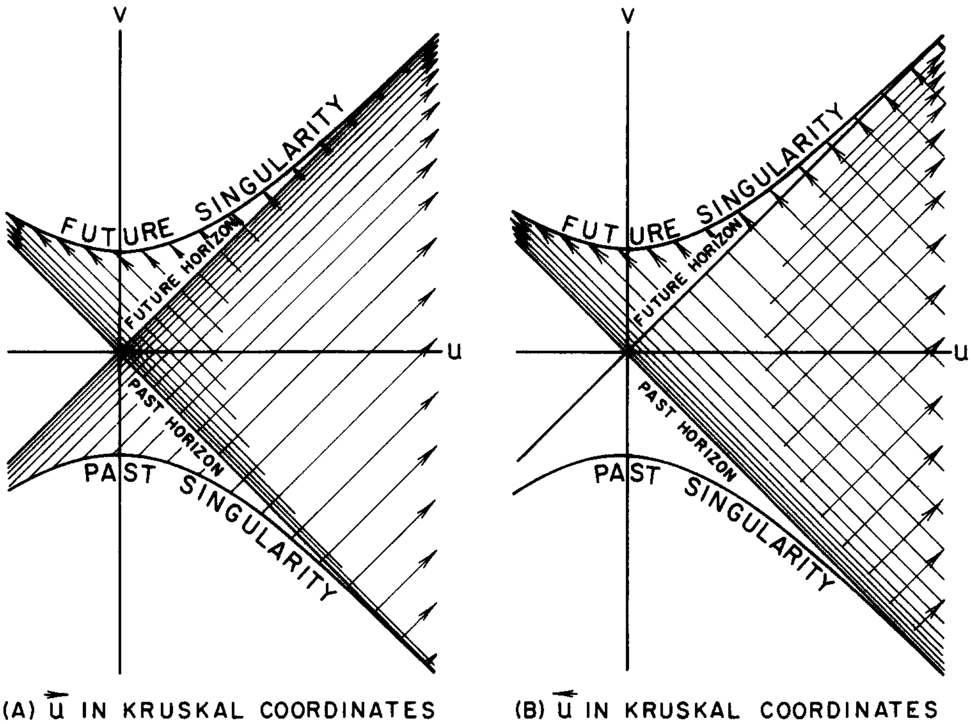


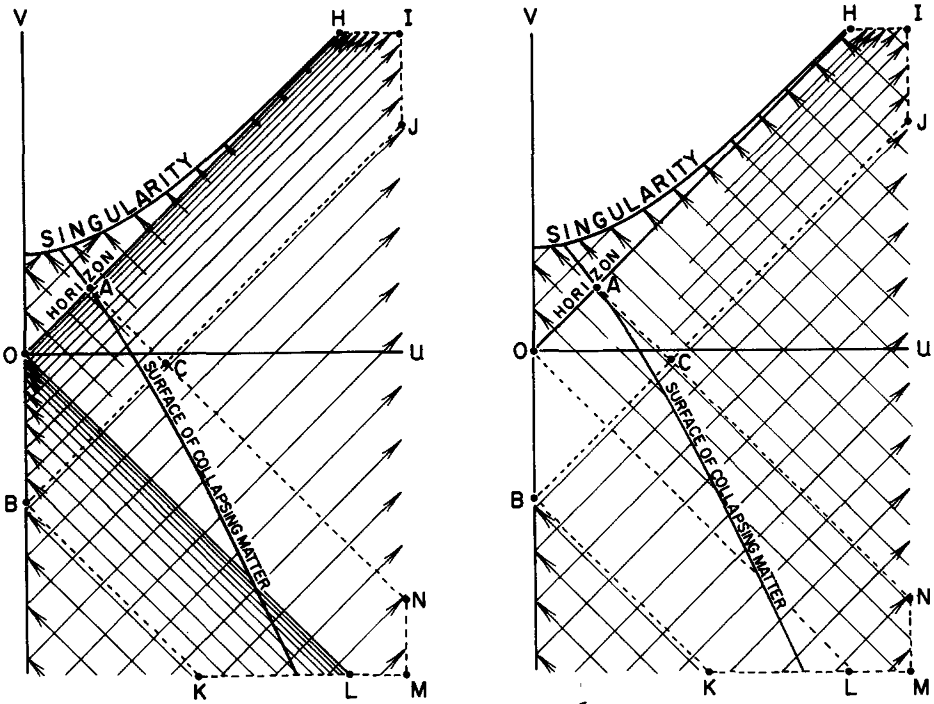
Fig. 2.

Although the above basis functions are no longer valid for early times (because there is no past horizon) they will still be useful to us as $t \to \infty$, after the collapse has been completed and a quiescent black hole has formed. Figure 2 shows what the spacetime behavior of the radial part of these functions would be like if there *were* a past horizon. The figure is drawn using coordinates u and v for which radial null directions are at 45° and which, near the horizons, are related to the standard Schwarzschild coordinates by the Kruskal transformation

$$\begin{aligned} u &= \left(\frac{r}{2M} - 1\right)^{1/2} \mathrm{e}^{r/4M} \cosh\,(t/4M) \\ v &= \left(\frac{r}{2M} - 1\right)^{1/2} \mathrm{e}^{r/4M} \sinh\,(t/4M). \end{aligned} \tag{109}$$

The Kruskal transformation provides a "maximal analytic extension" of the Schwarzschild line element, which has the virtue of keeping the metric tensor well behaved (nonsingular) on the horizons ($r = 2M$, $t = \pm\infty$). Strictly speaking, equations (109) hold only in the right hand quadrant (outside the horizons) and must be replaced by similar expressions in the other quadrants. The quadrant omitted in each picture may be regarded as another universe joined to our own through a "wormhole", but in the collapse situation it does not exist and hence has no relevance for the present discussion.

Each point on the diagram represents a 2-sphere of radius r, and lines of constant r are hyperbolae. Lines at 45° bearing arrowheads are lines of constant phase (wave crests) for the various



(A) $\overrightarrow{u}$ IN COLLAPSE SITUATION (B) $\overleftarrow{u}$ IN COLLAPSE SITUATION

Fig. 3.

components of the basis functions. It will be noted that they crowd together infinitely densely at the horizons. This is an expression of the gravitational red shift: The nearer to the horizon a wavelet finds itself the shorter must be its local wavelength in order that it have (or have had) a preassigned fixed (monochromatic) frequency at infinity.

5.2. Global behavior of the late-time basis functions

Figure 3 shows the actual behavior of the basis functions in the collapse situation. Here there is only one horizon, a future horizon, which has been formed by the catastrophic in-fall of a spherical distribution of matter. The coordinates, labeled u and v as before, are again chosen so that radial null directions are at 45°. Each point again represents a 2-sphere, except for the v axis itself, which represents the world line of the center of the mass distribution. Points for which $v < 0$ are now missing. As the collapse proceeds a light cone is eventually reached, from the inside of which nothing can escape to infinity. This is the horizon. Its apex (the birth event of the horizon) is located at the origin of the u, v coordinates. The point A, at which the surface of the collapsing matter crosses the horizon, marks the birth of the black hole itself.

Above the dotted line ACN and outside the horizon in each picture the new coordinates u and v, and the basis functions $\vec{u}$, $\overleftarrow{u}$, coincide with those of fig. 2. Below the dotted line there are significant differences. Consider first the function $\overleftarrow{u}$. Above the line ACN the incoming waves of this function originate at infinity with unit amplitude. They maintain this amplitude until they arrive in the region where the function V_l, eq. (104), begins to assume significant values. In the diagram the outer boundary of this region would be marked roughly by an r = constant line passing through the point C. As the incoming waves traverse this region their amplitude decreases, until it reaches the value $|B_l|$ which they carry as they plunge across the horizon. Outgoing (scattered) waves are born in the same region. These escape to infinity across the line HIJ, carrying an amplitude $|\overleftarrow{A}_l|$.

Below the line BCJ the outgoing waves still carry the amplitude $|\overleftarrow{A}_l|$ to infinity. However, the scattering process from which they originate differs significantly from that above the line. As one follows these waves backwards in time one encounters the collapsing matter well outside the horizon, which implies a weakened function V_l. Below the line OL, moreover, there is no longer a horizon to absorb the unscattered incoming waves. The result is that the amplitude of the incoming waves rapidly decreases as one traverses the region between the lines ACN and OL, and drops virtually to zero in the region OLKB. Below the line BK the amplitude picks up again, finally stabilizing, at early times, at the value $|\overleftarrow{A}_l|$. These early incoming waves, as one follows their progress forwards in time, ultimately become transformed completely into outgoing waves, partly by a process of back-scattering off of the curvature of spacetime and partly by passing completely through the center of the collapsing matter. They therefore must carry the same amplitude as the outgoing waves.

A word must be inserted about possible non-gravitational interactions between the scalar field (or any other field that one may be quantizing) and the collapsing matter. If there is a moderate or strong coupling between the two one may ask why we omit it from consideration in the description of the basis functions $\vec{u}$, $\overleftarrow{u}$, particularly as these functions propagate into and through the matter below the line ACN. The answer is that we shall be considering a vacuum problem. There are no field quanta present at early times. The collapsing matter therefore interacts initially only with the vacuum fluctuations of the field, and the issue becomes one of computing the corrections to the physical properties of the matter arising from such interactions, and of making any renormalizations that may be necessary in the observable physical parameters of the matter. If we assume these corrections already to be included in our description of the matter, we do not have to consider them a second time. As for the real quanta that get produced during the collapse process, they do indeed interact directly with the matter when coupling is present. But I shall defer discussion of their behaviour until later.

Let us consider next the function $\vec{u}$, the behavior of which differs markedly from that of $\overleftarrow{u}$. The most significant feature of this function is the crowding of an infinite number of outgoing waves into the region just outside the horizon. Consider the outgoing waves contained in the region OHIJB. These waves carry the amplitude $|B_l|$ to infinity, across the line HIJ. But they cross the line OL with unit amplitude. This means that the incoming waves in the region OLKB, which give rise to them, must carry from infinity (across the line KL) at least as big an amplitude. Near the line OL the amplitude of these incoming waves must, in fact, be exactly unity. That is because the number of waves near OL is *infinite*, and hence short wavelengths (high frequencies) dominate. Such waves propagate according to geometrical optics, without becoming weakened by scattering.

It is not difficult to determine the form of $\overleftarrow{u}$ near OL at infinity. First note that

$$e^{r^*/2M} = \left(\frac{r}{2M} - 1\right) e^{r/2M} \quad , \tag{110}$$

which, when substituted into (109), yields

$$r^* - t = 4M \ln(u - v)\,, \tag{111}$$

and hence

$$\vec{u}(l, m, p\,x) \xrightarrow[\substack{r^* \to -\infty \\ t \to \infty}]{} \frac{1}{2\pi\sqrt{2p}\,r}\, Y_{lm}\,(\cos\theta)\, e^{im\phi}\, [\exp\{\,i4Mp\ln(u - v)\} + \vec{A}_l(p)\exp\{-ip(r^* + t)\}]. \tag{112}$$

The form that $\vec{u}$ takes near the line OL is obtained from (112) by bouncing the first term inside the brackets off of the v axis (geometrical optics) and replacing the second term by expression (101) for $r^* \to \infty$:

$$\vec{u}(l, m, p|x) \xrightarrow[\substack{r^* \to \infty \\ \text{near OL}}]{} \frac{1}{2\pi\sqrt{2p}\,r}\, Y_{lm}(\cos\theta)\, e^{im\phi}\, [\theta(-u-v)\exp\{i4Mp\ln(-u-v)\} + B_l(p)\exp\{ip(r^* - t)\}]. \tag{113}$$

The step function in the bounce term effects the sudden switch-off of this term that is evident from fig. 3(A). To avoid problems of indeterminate phase in future dealings with the bounce term, it will be useful to give p an infinitesimal negative imaginary part so that this term actually vanishes *on* the line OL.

To get expression (113) into a form that can be used one must replace u and v by r^* and t. (In terms of r^* and t the metric retains its standard Schwarzschild form everywhere outside the matter.) The connection between the two sets of coordinates is obtained by the following argument due to Hawking:

Let k be a small contravariant null inward-pointing displacement vector close to the horizon in the upper part of the (u, v) plane, having the components $(-\epsilon, \epsilon)$ in the (u, v) coordinate system. Suppose k intersects N wave crests of the function $\vec{u}$. Let k be displaced in a parallel fashion into the past along the constant phase lines (which are null geodesics), through the center of the collapsing matter, and out again to infinity. The displaced k will still be null and will still intersect N wave crests (of the bounce term), but it will now be outward pointing. Therefore it must now have the components (ϵ, ϵ) (in the (u, v) coordinate system). In the (r^*, t) coordinate system, which has the line element $ds^2 = dr^{*2} - dt^2$ near infinity, its components must evidently be of the form $(D\epsilon, D\epsilon)$ where D is some factor that is constant (for large r^*) along the line OL. From this it follows that

$$-u - v \xrightarrow[r^* \to \infty]{} D(t_0 - r^* - t) \quad \text{near OL} \tag{114}$$

for some t_0. This relation may be used directly in eq. (113).

5.3. *The steady-state component and its scaling property*

Far from the line OL expression (113), with the replacement (114), is no longer exactly correct. First of all, the factor D does not remain exactly constant. Secondly, the amplitude of the bounce term decreases. Below the line BK, for example, the amplitude of the incoming component must tend to $|B_l|$ to match that of the outgoing waves below the line BCJ. These changes, however, affect only the transient behavior of the quantized field. We are presently going to look at the stress tensor at large values of r^* *above* the line BCJ. In this region the field has settled to a steady state, and the behavior of $\vec{u}$ there would remain unchanged if it were replaced by a function whose form at early times were given *exactly* by (113). I shall call the modified function obtained this way the *steady-state component* and denote it by the boldface symbol $\vec{\boldsymbol{u}}$:

$$\vec{\boldsymbol{u}}(l,m,p|x) \xrightarrow[\substack{r^*\to\infty\\ t\to-\infty}]{} \frac{1}{2\pi\sqrt{2p}\,r}\,[\theta(t_0-r^*-t)\exp\{\mathrm{i}4Mp[\ln(t_0-r^*-t)+\ln D]\}$$
$$+B_l(p)\exp\{\mathrm{i}p(r^*-t)\}]\;Y_{lm}(\cos\theta)\,\mathrm{e}^{\mathrm{i}\phi}\,. \tag{115}$$

The only uncertain quantity in expression (115) is the constant D, whose value depends on how the geometry behaves in the region occupied by matter and hence on the details of the collapse. But it turns out that we do not need to know it. This is because D only occurs in a logarithm and hence contributes only an irrelevant phase factor☆. The logarithmic occurrence is an expression of an important scaling property of the wave crests near the horizon and near the line OL: The spacetime distribution of these crests, in particular their infinite crowding, *looks the same under all magnifications.*

5.4. Early-time basis functions. The Bogoliubov transformation

Because the metric varies with time during the collapse process we may expect particle production, arising purely from geometry, to occur. In order to compute its rate we need to define an inital state. Let us assume that the geometry and distribution of matter are static at early times, before the onset of the collapse process. The basis functions that are useful in this regime have the form

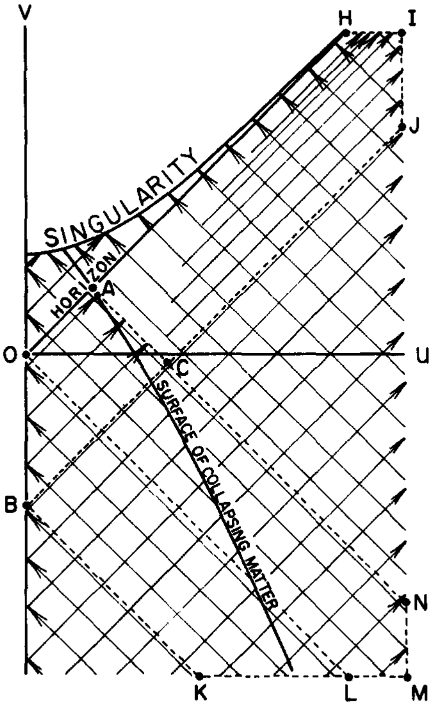
$$f(l,m,p|x)=\frac{1}{2\pi\sqrt{2p}\,r}\,F_l(p|r)\,Y_{lm}(\cos\theta)\,\mathrm{e}^{\mathrm{i}m\phi}\,\mathrm{e}^{-\mathrm{i}\epsilon t}\,,\qquad p>0,$$
$$F_l(p|r)\xrightarrow[r^*\to\infty]{}\exp(-\mathrm{i}pr^*)-(-1)^l\exp\{\mathrm{i}(pr^*+2\delta_l(p))\} \tag{116}$$

where $\epsilon=p$ and $\delta_l(p)$ is the familiar S-matrix phase shift. Here we do not have two sets of basis functions, bearing arrows $\rightarrow$ and $\leftarrow$, but only one. The birth of the horizon appears to effect a sudden doubling of the number of degrees of freedom of the field! This is purely a manner of speaking, however, as spacetime may be completely covered (up to the singularity) by a sequence of spacelike hypersurfaces, and the Cauchy problem is well posed on every member of the sequence, whether that member crosses the horizon or not. This means that the functions $\vec{u}$, $\overleftarrow{u}$ are related to the functions f by an invertible Bogoliubov transformation:

$$\left.\begin{aligned}\vec{u}(l,m,p|x)&=\int_0^\infty[\vec{\alpha}_l(p,p')f(l,m,p'|x)+\vec{\beta}_l(p,p')f^*(l,-m,p'|x)]\,\mathrm{d}p'\\ \overleftarrow{u}(l,m,p|x)&=\int_0^\infty[\overleftarrow{\alpha}_l(p,p')f(l,m,p'|x)+\overleftarrow{\beta}_l(p,p')f^*(l,-m,p'|x)]\,\mathrm{d}p'\,.\end{aligned}\right\} \tag{117}$$

The invertibility may be expressed by the matrix relation (labels suppressed, cf. eq. (15))

☆ For this same reason our final results, eqs. (135) and (136), would remain unchanged if the inital matter distribution were aspherical, provided only its total spin angular momentum vanishes. (See Hawking [32].)



THE BASIS FUNCTION f

Fig. 4.

$$\begin{pmatrix} \overrightarrow{\alpha} & \overrightarrow{\beta} \\ \overrightarrow{\beta}^* & \overrightarrow{\alpha}^* \\ \overleftarrow{\alpha} & \overleftarrow{\beta} \\ \overleftarrow{\beta}^* & \overleftarrow{\alpha}^* \end{pmatrix} \begin{pmatrix} \overrightarrow{\alpha}^\dagger & -\overrightarrow{\beta}^\sim & \overleftarrow{\alpha}^\dagger & -\overleftarrow{\beta}^\sim \\ -\overrightarrow{\beta}^\dagger & \overrightarrow{\alpha}^\sim & -\overleftarrow{\beta}^\dagger & \overleftarrow{\alpha}^\sim \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{118}$$

("1" denoting a delta function), which implies

$$\begin{aligned} f(l, m, p|x) = \int_0^\infty [&\overrightarrow{\alpha}_l^*(p', p)\, \overrightarrow{u}(l, m, p'|x) - \overrightarrow{\beta}_l(p', p)\, \overrightarrow{u}^*(l, -m, p'|x) \\ &+ \overleftarrow{\alpha}_l^*(p', p)\, \overleftarrow{u}(l, m, p'|x) - \overleftarrow{\beta}_l(p', p)\, \overleftarrow{u}^*(l, -m, p'|x)]\, \mathrm{d}p' \,. \end{aligned} \tag{119}$$

It is not possible to calculate the α and β coefficients exactly. However, we need only those parts of these coefficients that relate to the steady-state regime at late times. To determine these parts we look at the spacetime behavior of the function f, as depicted in fig. 4. All incoming waves, over the entire figure, begin (at infinity) with unit amplitude. As they propagate inward, those lying above the line OL split into two components, a back-scattered component and a component which continues on across the horizon. In the region AHIJC these two components make a total contribution to f that is *identical* to the function $\overleftarrow{u}$ in the same region. This is the steady-state region. Therefore, using boldface to denote the steady-state component of f, we have

$$f(l, m, p|x) = \overleftarrow{u}(l, m, p|x) + \int_0^\infty [\vec{\boldsymbol{\alpha}}_l^*(p', p)\,\vec{\boldsymbol{u}}(l, m, p'|x) - \vec{\boldsymbol{\beta}}_l(p', p)\,\vec{\boldsymbol{u}}^*(l, -m, p'|x)]\,\mathrm{d}p'. \tag{120}$$

The integral in this expression gives rise to the incoming waves in region OLKB of fig. 4 and to the outgoing waves in region OACB, which are shown terminating on the line AC but which actually continue on into the steady-state region and superimpose themselves on the other outgoing waves there. The integral expresses these waves as a linear combination of the steady-state functions $\vec{\boldsymbol{u}}$ and $\vec{\boldsymbol{u}}^*$.

5.5. Computation of the steady-state transformation coefficients

The coefficients of the linear combination may be determined by computing the inner products

$$\vec{\boldsymbol{\alpha}}_l(p', p) = \langle f(l, m, p), \vec{\boldsymbol{u}}(l, m, p')\rangle \tag{121}$$

$$\vec{\boldsymbol{\beta}}_l(p', p) = -\langle f^*(l, -m, p), \vec{\boldsymbol{u}}(l, m, p')\rangle\,, \tag{122}$$

at early times. At early times the exact form of the inner product [see eq. (6)] is not known because we do not know the metric inside the matter. But this difficulty may be circumvented by replacing the functions (116) with broad wave packets, going to very early times, and afterward passing to the limit as the packets become arbitrarily large. The functions F_l then become effectively $\exp(-\mathrm{i}pr^*)$ and the inner product may be taken in the form

$$\begin{aligned}\langle u_1, u_2\rangle &= \mathrm{i}\int_{2M}^\infty \mathrm{d}r \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \left(1 - \frac{2M}{r}\right)^{-1} r^2 \sin\theta\, u_1^* \frac{\overleftrightarrow{\partial}}{\partial t} u_2 \\ &= \mathrm{i}\int_{-\infty}^\infty \mathrm{d}r^* \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi\, r^2 \sin\theta\, u_1^* \frac{\overleftrightarrow{\partial}}{\partial t} u_2\,,\end{aligned} \tag{123}$$

in which the two-edged operator is just (61) with a taken equal to zero.

We may illustrate the procedure by computing $\vec{\boldsymbol{\alpha}}$:

$$\begin{aligned}\vec{\boldsymbol{\alpha}}_l(p', p) &= \frac{1}{8\pi^2\sqrt{pp'}} \int_{-\infty}^\infty \mathrm{d}r^* \int_0^\pi \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \sin\theta \exp\{\mathrm{i}p(r^* + t)\}\,[Y_{lm}(\cos\theta)]^2 \\ &\times \left\{\theta(t_0 - r^* - t)\left(\frac{4Mp'}{t_0 - r^* - t} + p\right) \exp\{\mathrm{i}4Mp'[\ln(t_0 - r^* - t) + \ln D]\} + B_l(p)(p' + p)\exp\{\mathrm{i}p'(r^* - t)\}\right\} \\ &= \frac{1}{4\pi\sqrt{pp'}} D^{\mathrm{i}4Mp'} \exp(\mathrm{i}pt_0) \int_0^\infty (4Mp'\, x^{-1+\mathrm{i}4Mp'} + px^{\mathrm{i}4Mp'})\,\mathrm{e}^{-\mathrm{i}px}\,\mathrm{d}x\end{aligned}$$

where $x = t_0 - r^* - t$. Since $p > 0$ the integration contour may be rotated into the negative imaginary axis. Setting $x = -\mathrm{i}\xi/p$ one immediately sees that the integral may be expressed in terms of gamma functions. With the aid of the well known relation $z\,\Gamma(z) = \Gamma(z + 1)$ one gets

$$\vec{\boldsymbol{\alpha}}_l(p', p) = -\frac{\mathrm{i}e^{2\pi M p'}}{2\pi\sqrt{pp'}} D^{\mathrm{i}4Mp'} \exp(\mathrm{i}pt_0)\, p^{-\mathrm{i}4Mp'}\Gamma(1 + \mathrm{i}4Mp'). \tag{124}$$

In a similar manner one finds

$$\vec{\boldsymbol{\beta}}_l(p', p) = \frac{\mathrm{i}e^{-2\pi M p'}}{2\pi\sqrt{pp'}} D^{\mathrm{i}4Mp'} \exp(-\mathrm{i}pt_0)\, p^{-\mathrm{i}4Mp'}\Gamma(1 + \mathrm{i}4Mp'). \tag{125}$$

By making use of the representation of the delta function

$$\delta(x) = (2\pi)^{-1} \int_0^\infty p^{-1 \pm \mathrm{i}x}\, \mathrm{d}p \tag{126}$$

and the identity

$$|\Gamma(1 + \mathrm{i}x)|^2 = \frac{\pi x}{\sinh \pi x}, \qquad x \text{ real}, \tag{127}$$

and remembering that p' and p'' are always positive, one easily verifies that the functions $\vec{\boldsymbol{\alpha}}_l$ and $\vec{\boldsymbol{\beta}}_l$ satisfy the following orthonormality relations☆:

$$\int_0^\infty [\vec{\boldsymbol{\alpha}}_l^*(p'', p)\vec{\boldsymbol{\alpha}}_l(p', p) - \vec{\boldsymbol{\beta}}_l^*(p'', p)\vec{\boldsymbol{\beta}}_l(p', p)]\, \mathrm{d}p = \delta(p' - p''), \tag{128}$$

$$\int_0^\infty \vec{\boldsymbol{\alpha}}_l(p'', p)\vec{\boldsymbol{\beta}}_l(p', p)\, \mathrm{d}p = 0. \tag{129}$$

5.6. Particle production. The Planckian spectrum. Temperature of a Schwarzschild black hole

If now the quantum state at early times is taken to be the vacuum state relative to the basis functions f, then the expectation value of the stress tensor in the steady-state region is given by

$$\begin{aligned}
\langle T^{\mu\nu}\rangle &= \sum_{l,m} \int_0^\infty T^{\mu\nu}(f(l, m, p|x), f^*(l, m, p|x))\, \mathrm{d}p \\
&= \sum_{l,m} \Bigg\{ \int_0^\infty T^{\mu\nu}(\overleftarrow{u}(l, m, p|x), \overleftarrow{u}^*(l, m, p|x))\, \mathrm{d}p \\
&\quad + \int_0^\infty \mathrm{d}p \int_0^\infty \mathrm{d}p' \int_0^\infty \mathrm{d}p'' [\vec{\boldsymbol{\alpha}}_l^*(p'', p)\vec{\boldsymbol{\alpha}}_l(p', p) + \vec{\boldsymbol{\beta}}_l^*(p'', p)\vec{\boldsymbol{\beta}}_l(p', p)]\, T^{\mu\nu}(\vec{\boldsymbol{u}}(l,m,p''|x), \vec{\boldsymbol{u}}^*(l,m,p'|x)) \\
&\quad + \int_0^\infty \mathrm{d}p \int_0^\infty \mathrm{d}p' [\vec{\boldsymbol{\alpha}}_l^*(p', p)\, T^{\mu\nu}(\vec{\boldsymbol{u}}(l,m,p'|x), \overleftarrow{u}^*(l,m,p|x)) + \vec{\boldsymbol{\alpha}}_l(p',p)\, T^{\mu\nu}(\overleftarrow{u}(l,m,p|x), \vec{\boldsymbol{u}}^*(l,m,p'|x)) \\
&\quad - \vec{\boldsymbol{\beta}}_l^*(p', p)\, T^{\mu\nu}(\vec{\boldsymbol{u}}(l,m,p|x), \overleftarrow{u}(l,-m,p'|x)) - \vec{\boldsymbol{\beta}}_l(p',p)\, T^{\mu\nu}(\vec{\boldsymbol{u}}^*(l,m,p'|x), \overleftarrow{u}^*(l,-m,p|x))]\Bigg\}.
\end{aligned} \tag{130}$$

☆ The functions $\vec{\alpha}_l$ and $\vec{\beta}_l$, of which $\vec{\boldsymbol{\alpha}}_l$ and $\vec{\boldsymbol{\beta}}_l$ are the steady-state components, also satisfy the relation (128), as well as a relation obtained by antisymmetrizing (129) in p' and p'' [see eq. (118)].

The integrand of the last integral inside the curly brackets oscillates very rapidly when r^* is large, and hence this integral vanishes in the limit $r^* \to \infty$. The second integral inside the curly brackets (the triple integral) can be reduced to a single integral with the aid of the same identities (eqs. (126) and (127)) as were used to obtain the orthonormality relations (128) and (129). One then finds

$$\langle T^{\mu\nu}\rangle \underset{r^* \to \infty}{\Longrightarrow} \sum_{l,\, m} \int_0^\infty [T^{\mu\nu}(\overleftarrow{u}(l,m,p|x), \overleftarrow{u}^*(l,m,p|x)) + \operatorname{ctnh}(4\pi Mp)\, T^{\mu\nu}(\vec{u}(l,m,p|x)\, \vec{u}^*(l,m,p|x))]\, \mathrm{d}p. \tag{131}$$

To obtain the steady-state particle production rate we need the (r, t) component of this expression above the line BCJ. Using the explicit form of the stress tensor one finds, for both the normal scalar field and the conformally invariant scalar field☆,

$$T_{rt}(\vec{u}(l,m,p|x), \vec{u}^*(l,m,p|x)) \xrightarrow[\text{above BCJ}]{r^* \to \infty} -\frac{\sin\theta}{8\pi^2}\, [Y_{lm}(\cos\theta)]^2 p|B_l|^2\ , \tag{132}$$

$$T_{rt}(\overleftarrow{u}(l,m,p|x), \overleftarrow{u}^*(l,m,p|x)) \xrightarrow[\text{above BCJ}]{r^* \to \infty} \frac{\sin\theta}{8\pi^2}\, [Y_{lm}(\cos\theta)]^2\, p(1 - |\overleftarrow{A}_l|^2)\ , \tag{133}$$

and hence, using (107),

$$\langle T_{rt}\rangle \xrightarrow[\text{above BCJ}]{r^* \to \infty} -\frac{\sin\theta}{4\pi^2} \sum_{l,\, m} [Y_{lm}(\cos\theta)]^2 \int_0^\infty |B_l(p)|^2\, \frac{p}{\mathrm{e}^{8\pi Mp} - 1}\, \mathrm{d}p\ . \tag{134}$$

One immediately sees that the spectrum of the emitted radiation is Planckian! The black hole looks like a black body having the temperature

$$T = 1/8\pi kM \tag{135}$$

seen through a filter, the filter being represented by the transmission probability factor $|B_l|^2$. This association of a temperature with a black hole is Hawking's astonishing discovery.

5.7. Consistency with thermodynamical equilibrium principles

It will be noted that the temperature depends only on the mass of the black hole and not on the details of its formation. The temperature is also independent of the strength of the coupling (if any) between the scalar field and the collapsing matter. This can be seen as follows. Suppose first that there is no coupling. Then the collapsing matter, even if it is originally very hot, will appear to an observer at infinity to have a temperature that decreases (red shift) with exponential rapidity as it approaches the horizon. The geometrically generated radiation, however, will continue to arrive at infinity with the temperature T. Now suppose the coupling is switched on. It

☆ In these equations the stress tensor is taken in its density form.

then becomes important to ask: From where does the radiation originate? Although much work remains to be done before this question can be tackled with complete assurance, there seem to be only two reasonable answers. The radiation is either generated at a steady rate close to the horizon along its entire length, or else the bulk of it is generated inside the region OACB (figs. 3 and 4), which is mostly inside the matter. In the former case there is no problem; coupling between matter and radiation will not affect the steady-state temperature. In the latter case the following argument holds: Because the radiation is (a) thermal, and (b) inexhaustible (until the black hole itself decays, see below) any coupling between it and the matter can only serve to bring the matter to the same temperature that it has. The quanta that reach an observer at infinity will in this case mostly have been emitted by the matter (reradiation) but they will still carry the temperature T. That is, the observed temperature of the matter will stabilize at the value T instead of dropping exponentially to zero. Note that this implies an extremely high *local* temperature for the matter (infinite if the radiation energy were truly inexhaustible) near the point A of figs. 3 and 4, in order to compensate for the red shift.

Hawking deduced the existence of the thermal radiation by means of wave packet arguments not based on the stress tensor, and he did not give an exact expression for the total luminosity of the black hole. With the stress tensor before us it is now easy to obtain such an expression (cf. eq. (87)):

$$\frac{\mathrm{d}M}{\mathrm{d}t} = \lim_{\substack{r\to\infty \\ \text{above BCJ}}} \int_0^{\pi} \mathrm{d}\theta \int_0^{2\pi} \mathrm{d}\phi \langle T_{rt}\rangle = -\frac{1}{2\pi}\sum_{l=0}^{\infty}\int_0^{\infty} (2l+1)|B_l(p)|^2 \frac{p}{\mathrm{e}^{p/kT}-1}\,\mathrm{d}p. \tag{136}$$

If the temperature idea is to be consistent with thermodynamical principles there should be no mass loss if the black hole is removed from isolation and immersed in a radiation bath at its own temperature. A state of equilibrium should then exist, with the black hole absorbing as much radiation as it emits. This may be checked by direct calculation of the absorption rate. The density of scalar photons at temperature T, having momenta in the momentum-space volume element $\mathrm{d}^3\boldsymbol{p}$ around $\boldsymbol{p}$ is

$$\frac{\mathrm{d}^3\boldsymbol{p}}{(2\pi)^3(\mathrm{e}^{p/kT}-1)}\,. \tag{137}$$

The photons of momentum $\boldsymbol{p}$ that have impact parameters corresponding to angular momentum l are contained in a coaxial tube centered on the black hole, having an annular cross section of area $(2l+1)\pi/p^2$. The energy absorption rate is obtained by multiplying this area by $p|\overleftarrow{B}_l(p)|^2$ times expression (137), where $\overleftarrow{B}_l$ is the transmission amplitude (into the black hole) for incoming radiation, and then integrating over $\boldsymbol{p}$ and summing over l. In virtue of eq. (105) one finds immediately that the absorption rate is equal to the emission rate $|\mathrm{d}M/\mathrm{d}t|$.*

5.8. *Temperature of a Kerr black hole*

The temperature idea can be extended to Kerr black holes. It is not difficult to show (for ex-

* It has been pointed out by Larry Ford and Stephen Hawking (private communication) that the equilibrium between a black hole and a radiation bath at the same temperature is actually an unstable one. The more energy a back hole emits the hotter it gets [see eq. (146)], the more it absorbs the cooler it gets. Some stars behave this way too, at certain points in their evolutionary cycles.

ample, by examining the line element (51) on the equator: $\theta = \pi/2$) that a set of Kruskal-like coordinates for the Kerr metric is obtained, in the neighbourhood of the horizon, by replacing the $4M$ in equations (109) by $4Mr_+/(r_+ - r_-)$. This has the consequence that the effective temperature of a Kerr black hole is

$$T = \frac{r_+ - r_-}{8\pi k M r_+}\,, \tag{138}$$

and equation (134) gets replaced by

$$\langle T_{rt}\rangle \xrightarrow[\text{above BCJ}]{r^* \to \infty} -\frac{\sin\theta}{4\pi^2} \sum_{l,m} \int_0^\infty (1 - |\overleftarrow{A}_{lm}(p,a)|^2)[S_{lm}(a\overleftarrow{\epsilon}|\cos\theta)]^2 \frac{p}{\exp\{(p - m\Omega_{\rm H})/kT\} - 1}\,\mathrm{d}p. \tag{139}$$

If the mass of the black hole is very large, so that the temperature is very small, the Planck factor in (139) is negligible for $p > m\Omega_{\rm H}$ and is practically equal to -1 for $p < m\Omega_{\rm H}$. In this limit, therefore, eq. (139) reduces to (85). That is to say, if the mass is very large the thermal radiation is negligible, and the energy flux reduces to the spontaneous emission of Starobinsky and Unruh.

5.9. *Entropy of a black hole. The generalized second law of thermodynamics*

If one assigns a temperature to a black hole then, in order to complete the thermodynamical picture, one must consider assigning also an entropy. The idea that one might be able to assign an entropy to a black hole was first suggested by Bekenstein [1], using the analogy of Hawking's area theorem (eq. (98)) to the second law of thermodynamics. This theorem led Bekenstein to define the entropy as being proportional to the area, and he tried to estimate the proportionality constant by the following reasoning: If matter falls into a black hole, the entropy it carries will be effectively lost to the world outside, and a violation of the second law will occur, unless the entropy of the black hole increases by at least as much. The simplest object that can fall into a black hole is an elementary particle. The least information that one can possess about a particle is whether or not it exists; that is one *bit* of information. The maximum entropy that a particle can carry, therefore, is $k \ln 2$, yielding a corresponding change $\Delta S = k \ln 2$, in the entropy of the black hole. Now Chrisodoulou and Ruffini [15] showed that there is only one type of orbit by which an idealized point particle can cross the horizon of a black hole without increasing the area of the hole, namely an orbit for which the particle crosses tangentially. If the particle has a radius r and mass m, Bekenstein showed that the area increases by an amount $\Delta A = 2mr$ even for this type of orbit. In the quantum theory all particles have an effective radius of the order of the Compton wavelength. That is, $r = \xi/m$, and hence $\Delta A = 2\xi$, where ξ is some number of the order of unity. The expressions for ΔS and ΔA will be consistent if we define

$$S = k\,\frac{\ln 2}{2\xi}\,A\,. \tag{140}$$

Bekenstein examined this definition in several other physical contexts as well, and found it to be consistent in every case.

Hawking's discovery allows us to fix the value of the constant ξ. We rewrite equation (98) in the form

$$\mathrm{d}M = \frac{r_+ - r_-}{32\pi M r_+}\,\mathrm{d}A + \Omega_\mathrm{H}\,\mathrm{d}J = \tfrac{1}{4}kT\,\mathrm{d}A + \Omega_\mathrm{H}\,\mathrm{d}J\,, \tag{141}$$

and invoke the analogy of this to eq. (34). We are led at once to the assignment

$$S = \tfrac{1}{4}kA = 4\pi k M_\mathrm{ir}^2\,, \tag{142}$$

and hence

$$\xi = 2\ln 2. \tag{143}$$

Because we lose all knowledge of the state of order or disorder of the matter that falls into a black hole, the entropy of a black hole must be the *maximum* entropy that a body of fixed mass and angular momentum can have. Indeed Bekenstein has pointed out that on an astrophysical scale the entropy given by eq. (142) is exceptionally large. For example, the entropy of a black hole of solar mass is 5×10^{60} erg/deg, whereas the entropy of the sun itself is only about 10^{42} erg/deg.

With the establishment of a generalized second law of thermodynamics, in which only the total entropy of a given astrophysical system, including the entropy of any black holes it contains, is required to be nondecreasing, Hawking's area theorem is transcended. The area of a black hole may now decrease, and indeed it will if it is not immersed in a radiation bath of at least equal temperature or else subjected to some other form of bombardment.

5.10. The mass decay law. Critical mass

It is of interest to examine the mass decay law implied by eq. (136). For simplicity I shall approximate the integral on the right hand side by assuming that the value of the transmission amplitude $B_l(p)$ is determined for all wavelengths, even long ones, by geometrical optics. In the geometrical optics approximation if the function V_l of eq. (104) becomes positive for some value of r then there will be no transmission. It is not difficult to verify that if $Mp \gg 1$ then transmission ceases when l exceeds $\sqrt{27}\,Mp$, and that the cut-off point does not differ greatly from this even when Mp is small. Therefore I shall take

$$B_l(p) \sim \theta(\sqrt{27}\,Mp - l)\,. \tag{144}$$

Inserting this into (136) and making use of

$$\sum_l (2l+1)\,\theta(\sqrt{27}\,Mp - l) \approx 27M^2p^2\,, \tag{145}$$

we find

$$\frac{\mathrm{d}M}{\mathrm{d}t} \sim -\frac{27\,M^2}{2\pi}\int_0^\infty \frac{p^3}{\mathrm{e}^{8\pi Mp} - 1}\,\mathrm{d}p = -\frac{9}{10\pi \times 8^4 \times M^2}\,. \tag{146}$$

The solution of this differential equation is

$$M^3 \sim M_0^3 - \frac{27}{10\pi \times 8^4}\,t\,, \tag{147}$$

where M_0 is the value of M at $t = 0$. The lifetime of a black hole is therfore given by

$$\tau \sim \frac{10\pi \times 8^4}{27} M^3 , \tag{148}$$

which may be compared with eq. (95). The approximation (148) may be improved by dividing it by the number of distinct massless quanta that exist in Nature and by taking into account also the emission of massive particles.

The decay law (147) is not exponential, as is the law (94), but explosive. In the last tenth of a second of its life a black hole releases of the order of 10^{30} erg. In order for a black hole to have survived for the age of the universe its mass must exceed 10^{15} g. If there were any black holes less massive than this, created in the early universe, they no longer exist.

5.11. Charged black holes

Gibbons [28] and Damour and Ruffini [16] have studied the problem of particle production by charged black holes. Here, in addition to emission of energy and angular momentum by the mechanisms already described, there is a net emission of *charge*, the electrostatic field of the black hole rather than the grativational field being responsible. For black holes of mass greater than 10^{15} g it is a true Klein-paradox process that dominates, and the emission rate is governed by a Schwinger-type formula (for pair production in a constant electric field). For black holes of smaller mass Hawking's thermal process dominates, and the charge flux is analogous to a thermionic current. In both regimes the black hole tends to discharge itself and, unless it is supermassive or carries an unrealistically large charge, does so quite rapidly☆. It is therefore very unlikely that small black holes ($M \sim 10^{17}$ g) will ever be seen in a bubble chamber, even if there are enough of them around to be statistically significant (which in itself is extremely unlikely, as may be inferred by considering the known limits on the mass density of the universe).

5.12. The naked singularity

If the area of a black hole goes to zero in a finite time then the event that marks its disappearance is a *naked singularity*. This follows from the necessarily noncausal structure of spacetime in the vicinity of a vanishing event horizon. The quantum theory therefore appears to lead not only to a violation of the Hawking area theorem but also to a violation of the *cosmic censorship principle* (i.e., the principle that singularities other than the original Big Bang are always hidden inside event horizons). It consequently becomes of great interest to examine the geometry of a decaying black hole in finer detail.

Unfortunately Hawking's derivation of the thermal emission phenomenon, by concentrating mainly on the radiation at infinity, yields little or no information about changes taking place near the horizon. The thermal emission, as we have seen, is insensitive to the details of the collapse process. The question arises: can we get more information by constructing idealized models of collapse, which permit us to study the behavior of the basis functions during the collapse itself? Hawking, in fact, first discovered the thermal emission effect *via* this route. But he was dissatisfied

☆ Ruffini and Wilson (quoted in Damour and Ruffini [16]) have nevertheless suggested that black holes of mass larger than $M_\odot$ (= 2×10^{33} g) may carry a large charge if surrounded by an appropriate plasma.

with the derivations because they did not explain the insensitivity of the effect to the model. He did not publish his discovery until, as he put it (private communication), he "found a better way".

Ulrich Gerlach [27] and David Boulware [3] have recently rederived the Hawking result using a collapsing spherical dust-shell model. The following picture *appears* to emerge from their analysis: The thermal radiation originates in the shell itself. The shell always remains outside the horizon. The radiation draws kinetic energy from the shell, and the area of the horizon, which is proportional to square of the sum of the rest, kinetic and gravitational binding energies of the shell, steadily decreases. The area of the shell and the area of the horizon vanish simultaneously.

Many issues remain to be settled, however, before this picture can be accepted as firmly established. First of all, it is unlikely that the collapse of the shell can be meaningfully followed beyond the point at which its radius reaches the absolute unit $\sim 10^{-33}$ cm (the so-called, "Planck length"). At this point the concept of a classical "background geometry" breaks down because of quantum fluctuations in the gravitational field itself. What lies beyond is anybody's guess.

Secondly, there is private disagreement at present among the experts concerning the "local reality" of the thermal radiation. Everyone agrees that the radiation at infinity is real. The disagreements are over its reality near the horizon. Hawking [31] believes that it is meaningless to speak of the radiation as originating in any given region. Because the horizon is a global construct and because an observer traversing the horizon in free fall sees nothing unusual about the geometry there, Hawking in particular repudiates the notion that the radiation originates close to the horizon. The issue is complicated by the difficulty of defining "particle" in the absence of a timelike Killing vector and by the insistence of some workers that what an observer in free fall calls a particle differs from what an accelerated observer (e.g., one "at rest" in a Schwarzschild field) calls it.

If only to stop nonsense discussions, it seems fairly urgent to attempt to settle the latter issue by building explicit models of "particle detectors" and computing how they perform under conditions of acceleration and free fall. But it is surely equally urgent to turn one's back (at least momentarily) on the issue and ask instead: What is the quantum expectation value of the stress tensor near the horizon? What modification is induced in the spacetime geometry (both inside and outside of the horizon) by the action of this expectation value? As I have previously remarked, a proper definition of the effective stress tensor must take into account not only the real particles produced but the virtual particles as well, and hence the effective stress tensor will not be well defined until we have a well defined and consistent way of regularizing and renormalizing $T^{\mu\nu}$.

6. The divergences

6.1. Resumé of proposals

The methods that have been proposed so far for dealing with the divergent parts of $T^{\mu\nu}$ fall into two classes: (1) covariant schemes designed to show that the divergences can be cancelled by adding counter terms to the gravitational action functional, and (2) frame-dependent techniques, introduced within the context of specific physical problems. Among the names associated with the first class are 't Hooft and Veltman [33], Capper and Ramón-Medrano [7] (see also Capper, Leibbrandt and Ramón-Medrano [8] and Capper, Duff and Halpern [9]), Deser and Van Nieuwenhuizen [18], and Utiyama and DeWitt [48]; names associated with the second class are Zel'dovich and Starobinsky [52] (see also Zel'dovich, Lukash and Starobinsky [53]), and Parker and Fulling [39] (see also Fulling and Parker [25], and Fulling, Parker and Hu [26]).

The authors of the class-2 schemes are primarily concerned with particle production in the early universe and during its final collapse. One of their aims is to determine whether curvature-induced particle creation and the virtual processes associated with it could have damped any initial anisotropies that were present in the universe and have led to a cosmology compatible with present observations. For this they need a finite stress tensor, from which the infinities have been subtracted, to use as a source that reacts back on the spacetime geometry. Zel'dovich and his associates have chosen, somewhat pragmatically, a particularly simple algorithm for subtracting these infinities within the context of a spatially flat Kasner model, and find that the resulting stress tensor leads to a rapid isotropization of the universe. (See the references for the details.) Parker and Fulling, more concerned with the theoretical aspects of the problem, attempt to go beyond this algorithm by extending it to cases in which 3-space is curved and showing that the subtractions may be effected by suitable counter terms in the gravitational action functional. They are thus trying to provide a solid physical justification (or at least a consistency proof) for the algorithm and to establish contact with the covariant schemes of class 1. They have achieved only partial success to date; some puzzling finite terms remain that cannot be identified with tensor components of the required type.

A major stumbling block in investigations of this kind is the difficulty of controlling the *a priori* geometrical (tensorial) character of selected parts of divergent mode-sums when the basis functions are those appropriate only to very restricted types of geometries. Unfortunately the covariant schemes of class 1 referred to above are of no help, for two reasons. Firstly, they are directly applicable only to weak (linearized) gravitational fields in flat spacetime*, and secondly, they are usable only if the whole problem can be set up in a Lorentz covariant manner. There is however, a scheme, known as "the method of the background field", that bypasses Lorentz covariance and achieves true general covariance in a straightforward manner. It is a technique that I have been trying to sell for a number of years [19, 20]. Workers in effective-potential theory and in weak-interaction theories of the Yang–Mills type have found it useful, and were it not for the effort required to learn it (on top of all the other things one must learn in order to handle quantum field theory effectively) more workers in general relatively would have found it useful by now. At any rate it is the only method of which I am aware that seems capable of resolving the difficulties at the present time, and in this final section I shall use it as a vehicle to display the basic idea of the Zel'dovich–Strobinsky scheme and show how the goals of Parker and Fulling may be reached.

6.2. *"In" and "out" regions, Bogoliubov coefficients, and the S-matrix*

In order to fix ideas let us assume that spacetime has two causally connected stationary regions, an "in" region and an "out" region, each possessing complete Cauchy hypersurfaces and a timelike Killing vector, as described in section 1. 3-space may be either finite or infinite, with arbitrary

* The authors of the class-1 schemes habitually display counter terms having full general covariance, but they are cheating. All they directly verify, with their Lorentz-covariant momentum-space closed-loop calculations, are the quadratic parts of these terms. The full counter terms are obtained by *invoking* general covariance and/or the so-called Ward–Slavnov identities.

connectivity. I shall argue later that the assumed existence of the "in" and "out" regions is inessential to the final results. The subtraction procedure chosen will remain valid in the limit as the volumes of these regions go to zero.

Let $\{u_{\text{in}\,i}\}$ and $\{u_{\text{out}\,i}\}$ be complete sets of normalized basis functions that contain only positive frequencies in the "in" and "out" regions respectively. They will be connected by a Bogoliubov transformation (cf. eq. (14)),

$$u_{\text{out}\,i} = \sum_j (\alpha_{ij} u_{\text{in}\,j} + \beta_{ij} u_{\text{in}\,j}{}^*) \,, \tag{149}$$

where the transformation coefficients satisfy the relation (15). For simplicity I shall assume that the field being quantized is a scalar field, but I shall not insist that it be massless or conformally invariant. The field equations will have the form

$$F\varphi \equiv g^{1/2}(\varphi_{;\mu}{}^{\mu} - \xi R\varphi - m^2\varphi) = 0 \,, \tag{150}$$

where m is the mass, R is the curvature scalar, and ξ is a numerical constant. (For the conformally invariant field $m = 0$ and $\xi = \frac{1}{6}$.) The methods and qualitative results of this section will be equally applicable to fields with spin, both fermion and boson.

The vacuum state vectors in the "in" and "out" regions are defined by

$$a_{\text{in}\,i}|\text{in, vac}\rangle = 0, \qquad a_{\text{out}\,i}|\text{out, vac}\rangle = 0, \tag{151}$$

for all i, where

$$\varphi = \sum_i (a_{\text{in}\,i} u_{\text{in}\,i} + a_{\text{in}\,i^*} u_{\text{in}\,i^*}) = \sum_i (a_{\text{out}\,i} u_{\text{out}\,i} + a_{\text{out}\,i^*} u_{\text{out}\,i^*}). \tag{152}$$

The annihilation operators in the "in" and "out" regions are related by (cf. eq. (16))

$$a_{\text{out}\,i} = \sum_j (\alpha_{ij}{}^* a_{\text{in}\,j} - \beta_{ij}{}^* a_{\text{in}\,j}{}^*), \qquad a_{\text{in}\,i} = \sum_j (\alpha_{ji} a_{\text{out}\,j} + \beta_{ji}{}^* a_{\text{out}\,j}{}^*). \tag{153}$$

These relations allow one to construct the S-matrix in terms of the Bogoliubov coefficients.

6.3. Particle creation and annihilation amplitudes

Of particular importance to quantum cosmologists are the many-particle production and annihilation amplitudes:

$$\mathrm{i}^{n/2}\, V_{i_1 \ldots i_n} \equiv \mathrm{e}^{-\mathrm{i}W} \langle \text{out}, i_1 \ldots i_n | \text{in, vac}\rangle \tag{154}$$

$$\mathrm{i}^{n/2}\, \Lambda_{i_1 \ldots i_n} \equiv \mathrm{e}^{-\mathrm{i}W} \langle \text{out, vac} | \text{in}, i_1 \ldots i_n\rangle \,, \tag{155}$$

where

$$\mathrm{e}^{\mathrm{i}W} \equiv \langle \text{out, vac} | \text{in, vac}\rangle \,, \tag{156}$$

$$|\text{in}, i_1 \ldots i_n\rangle \equiv a_{\text{in}\,i_1}{}^* \ldots a_{\text{in}\,i_n}^* |\text{in, vac}\rangle \tag{157}$$

$$|\text{out}, i_1 \ldots i_n\rangle \equiv a_{\text{out}\,i_1}^* \ldots a_{\text{out}\,i_n}^* |\text{out, vac}\rangle \,. \tag{158}$$

Assuming unit normalization for the "in" and "out" vacuum state vectors, one may write

$$\left.\begin{aligned} |\mathrm{in, vac}\rangle &= \mathrm{e}^{\mathrm{i}W} \sum_{n=0}^{\infty} \frac{\mathrm{i}^{n/2}}{n!} \sum_{j_1 \ldots j_n} V_{j_1 \ldots j_n} |\mathrm{out}, j_1 \ldots j_n\rangle \\ |\mathrm{out, vac}\rangle &= \mathrm{e}^{-\mathrm{i}W} \sum_{n=0}^{\infty} \frac{(-\mathrm{i})^{n/2}}{n!} \sum_{j_1 \ldots j_n} \Lambda_{j_1 \ldots j_n}^{*} |\mathrm{in}, j_1 \ldots j_n\rangle \,. \end{aligned}\right\} \tag{159}$$

Insertion of these expressions into (151) and use of relations (153) yield

$$0 = \mathrm{e}^{\mathrm{i}W} \sum_{n=0}^{\infty} \frac{\mathrm{i}^{n/2}}{n!} \sum_{j_1 \ldots j_n, k} \{\mathrm{i}^{1/2} V_{j_1 \ldots j_n, k} \alpha_{ki} |\mathrm{out}, j_1 \ldots j_n\rangle + V_{j_1 \ldots j_n} \beta_{ki}^{*} |\mathrm{out}, k, j_1 \ldots j_n\rangle\} \tag{160}$$

$$0 = \mathrm{e}^{-\mathrm{i}W} \sum_{n=0}^{\infty} \frac{(-\mathrm{i})^{n/2}}{n!} \sum_{j_1 \ldots j_n, k} \{(-\mathrm{i})^{1/2} \Lambda_{j_1 \ldots j_n, k}{}^{*} \alpha_{ik} |\mathrm{in}, j_1 \ldots j_n\rangle - \Lambda_{j_1 \ldots j_n}^{*} \beta_{ik}^{*} |\mathrm{in}, k, j_1 \ldots j_n\rangle\} \tag{161}$$

from which one may infer

$$V_{ij} = \mathrm{i} \sum_k \beta_{ki}^{*} \alpha^{-1}{}_{jk} \,, \quad \Lambda_{ij} = -\mathrm{i} \sum_k \beta_{ki} \alpha^{-1}{}_{kj} \,, \tag{162}$$

$$V_{i_1 \ldots i_n} = \begin{cases} 0, & n \text{ odd} \\ \sum_p V_{i_1 i_2} \cdots V_{i_{n-1} i_n}, & n \text{ even} \end{cases} \tag{163}$$

$$\Lambda_{i_1 \ldots i_n} = \begin{cases} 0, & n \text{ odd} \\ \sum_p \Lambda_{i_1 i_2} \cdots \Lambda_{i_{n-1} i_n}, & n \text{ even} \end{cases} \tag{164}$$

where "Σ_p" denotes a summation over the $n!/2^{n/2}(n/2)!$ distinct pairings of the labels $i_1 \ldots i_n$.

Equations (163) and (164) reveal the particle production and annilation processes as composed of individual pair creation and annihilation events. The complete symmetry of the amplitudes (Bose statistics), in particular the symmetry

$$V_{ij} = V_{ji}, \quad \Lambda_{ij} = \Lambda_{ji} \,, \tag{165}$$

follows from eq. (15). (In the fermion case complete antisymmetry holds.) Existence of the inverse matrix $(\alpha^{-1}{}_{ij})$, appearing in eqs. (162), follows from

$$\alpha\alpha^{\dagger} = 1 + \beta\beta^{\dagger} \quad \text{(positive definite).} \tag{166}$$

6.4. *One-particle scattering amplitudes and the optical theorems*

The remaining structural elements of the S-matrix are the one-particle scattering amplitudes:

$$\delta_{ij} + \mathrm{i}\, I_{ij} \equiv \mathrm{e}^{-\mathrm{i}W} \langle \text{out}, i | \text{in}, j \rangle = \mathrm{e}^{-\mathrm{i}W} \langle \text{out}, i | a_{\text{in}\, j}{}^* | \text{in, vac} \rangle$$
$$= \sum_{n=0}^{\infty} \frac{\mathrm{i}^{n/2}}{n!} \sum_{k_1 \dots k_n,\, l} V_{k_1 \dots k_n} \langle \text{out}, i | (\alpha_{lj}{}^* a_{\text{out}\, l}{}^* + \beta_{lj} a_{\text{out}\, l}) | \text{out}, k_1 \dots k_n \rangle$$
$$= \alpha_{ij}{}^* + \mathrm{i} \sum_l V_{il} \beta_{lj} \,. \tag{167}$$

Suppressing labels and making use of the symmetry of V_{ij}, one may rewrite this in the form

$$1 + \mathrm{i}\, I = \alpha^* - (\beta^* \alpha^{-1})^{\sim} \beta = \alpha^* - \alpha^{-1\sim} \beta^\dagger \beta = \alpha^* - \alpha^{-1\sim} (\alpha^{\sim} \alpha^* - 1) = \alpha^{-1\sim} \,. \tag{168}$$

The content of eq. (15) is then contained in the following identities, which may be derived from it:

$$\begin{aligned} (1 + \mathrm{i}\, I)\ (1 - \mathrm{i}\, I^\dagger) &= 1 - V V^\dagger \\ (1 - \mathrm{i}\, I^\dagger)(1 + \mathrm{i}\, I) &= 1 - \Lambda^\dagger \Lambda \\ V^\dagger (1 + \mathrm{i}\, I) &= (1 - \mathrm{i}\, I^*) \Lambda \\ \Lambda (1 - \mathrm{i}\, I^\dagger) &= (1 + \mathrm{i}\, I^{\sim}) V^\dagger \,. \end{aligned} \tag{169}$$

These identities constitute the relativistic version of the well known *optical theorem* and, together with eq. (1972) below, guarantee the unitarity of the complete many-particle S-matrix. (See DeWitt [20]. For the fermion case see Schwinger [44].)

6.5. *Vacuum-to-vacuum amplitude. Relation of its divergences to those of $T^{\mu\nu}$*

The only piece missing from the above derivation is the vacuum-to-vacuum amplitude itself. This is obtained by imposing the condition of unit total probability for transition out of the initial vacuum state:

$$1 = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{i_1 \dots i_n} |\langle \text{out}, i_1 \dots i_n | \text{in, vac} \rangle|^2 = \mathrm{e}^{-2 \mathrm{Im}\, W} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{i_1 \dots i_n} |V_{i_1 \dots i_n}|^2$$
$$= \mathrm{e}^{-2 \mathrm{Im}\, W} \det(1 - V V^\dagger)^{-1/2} \,, \tag{170}$$

whence, in virtue of (169),

$$\mathrm{e}^{-2 \mathrm{Im}\, W} = \det(1 - V V^\dagger)^{1/2} = \det(1 - \Lambda^\dagger \Lambda)^{1/2} \,, \tag{171}$$

and, with a natural choice of phase,

$$\mathrm{e}^{\mathrm{i}W} = \det(1 + \mathrm{i}\, I)^{1/2} = (\det \alpha)^{-1/2} \,. \tag{172}$$

The determinants above are of the Fredholm type, and a variety of methods is in principle available for their evaluation. The determinants (171), which yield the imaginary part of W, are finite. But (172), which yields also the real part of W (and hence the natural phase), contains divergences, the very same divergences, in fact, as are contained in the stress tensor. To see this, imagine that the metric tensor suffers an infinitesimal change $\delta g_{\mu\nu}$, yielding a change δS in the action functional (4) for the field φ. If the support of $\delta g_{\mu\nu}$ is confined to the spacetime region between the "in" and "out" regions then, by a well known variational principle (which incorporates the natural phase), obtain

$$\delta W = -\mathrm{i}\mathrm{e}^{-\mathrm{i}W}\delta \mathrm{e}^{\mathrm{i}W} = -\mathrm{i}\mathrm{e}^{-\mathrm{i}W}\delta\langle\mathrm{out, vac}|\mathrm{in, vac}\rangle = \mathrm{e}^{-\mathrm{i}W}\langle\mathrm{out, vac}|\delta S|\mathrm{in, vac}\rangle\,, \tag{173}$$

which, in virtue of eq. (11), yields

$$2\frac{\delta W}{\delta g_{\mu\nu}} = \mathrm{e}^{-\mathrm{i}W}\langle\mathrm{out, vac}|T^{\mu\nu}(\varphi,\varphi)|\mathrm{in, vac}\rangle = \sum_{n=0}^{\infty}\frac{\mathrm{i}^{n/2}}{n!}\sum_{i_1\ldots i_n}\Lambda_{i_1\ldots i_n}\langle\mathrm{in}, i_1\ldots i_n|T^{\mu\nu}(\varphi,\varphi)|\mathrm{in, vac}\rangle$$

$$= \sum_i T^{\mu\nu}(u_{\mathrm{in}\,i}, u_{\mathrm{in}\,i}{}^*) + \mathrm{i}\sum_{i,j}\Lambda_{ij}T^{\mu\nu}(u_{\mathrm{in}\,i}{}^*, u_{\mathrm{in}\,j}{}^*)\,. \tag{174}$$

Suppose no particles are present in the "in" region. Then the first term on the right of eq. (174) (last line) is the expectation value of the stress tensor, and we have

$$\langle\mathrm{in, vac}|T^{\mu\nu}|\mathrm{in, vac}\rangle = 2\frac{\delta W}{\delta g_{\mu\nu}} - \mathrm{i}\sum_{i,j}\Lambda_{ij}T^{\mu\nu}(u_{\mathrm{in}\,i}{}^*, u_{\mathrm{in}\,j}{}^*). \tag{175}$$

Similarly, one finds

$$\langle\mathrm{out, vac}|T^{\mu\nu}|\mathrm{out, vac}\rangle = 2\frac{\delta W}{\delta g_{\mu\nu}} - \mathrm{i}\sum_{i,j}V_{ij}T^{\mu\nu}(u_{\mathrm{out}\,i}{}^*, u_{\mathrm{out}\,j}{}^*)\,, \tag{176}$$

and, more generally, analogous expressions hold for the expectation values of the stress tensor in other states as well. The second terms on the right of eqs. (175) and (176) (and analogous terms for the expectation values in more general states) are finite. In every case the infinities are contained in the term $2\delta W/\delta g_{\mu\nu}$. Therefore it suffices to study W to study the infinities.

6.6. *Green's-function analysis of W*

It is convenient to begin this study by rewriting the variational law (173) in the more explicit form

$$\delta W = \tfrac{1}{2}\,\mathrm{e}^{-\mathrm{i}W}\langle\mathrm{out, vac}|\varphi\,\delta F\,\varphi|\mathrm{in, vac}\rangle = \tfrac{1}{4}\,\mathrm{e}^{-\mathrm{i}W}\,\mathrm{Tr}(\delta F\langle\mathrm{out, vac}|[\varphi,\varphi]_+|\mathrm{in, vac}\rangle)\,, \tag{177}$$

where "Tr" indicates that an integration is to be carried out over the suppressed spacetime labels that δF and the two φ's bear. Introduction of the anticommutator is allowed because of the symmetry (reality) of the self-adjoint operator δF in the boson case. (In the fermion case the commutator would appear.) Now

$$\tfrac{1}{2}[\varphi(x),\varphi(x')]_+ - (\varphi(x)\varphi(x'))_+ = \tfrac{1}{2}[\theta(x',x) - \theta(x,x')][\varphi(x),\varphi(x')]_-\,, \tag{178}$$

where "$(\ )_+$" denotes the chronological product and $\theta(x,x')$ is the chronological step function:

$$\theta(x,x') = \begin{cases} 1 \text{ if } x \text{ lies to the future of a spacelike hypersurface through } x' \\ 0 \text{ otherwise.}\end{cases} \tag{179}$$

Because of the commutativity of the φ's for spacelike separations the choice of the hypersurface in (179) is immaterial. The commutator is well known (see Peierls [40]) to be given by

$$[\varphi(x),\varphi(x')]_- = \mathrm{i}[G_{\mathrm{adv}}(x,x') - G_{\mathrm{ret}}(x,x')]\,, \tag{180}$$

where G_{adv} and G_{ret} are the advanced and retarded Green's functions for the operator F. Therefore

$$\tfrac{1}{2}[\varphi(x),\varphi(x')]_+ - (\varphi(x)\varphi(x'))_+ = \tfrac{1}{2}\mathrm{i}\,[G_{\mathrm{adv}}(x,x') + G_{\mathrm{ret}}(x,x')] \equiv \mathrm{i}\bar{G}(x,x')\,. \tag{181}$$

It is also well known that*

$$\mathrm{e}^{-\mathrm{i}W}\langle \mathrm{out},\mathrm{vac}|(\varphi(x)\varphi(x'))_+|\mathrm{in},\mathrm{vac}\rangle = -\mathrm{i}\,G(x,x')\,, \tag{182}$$

where G is the *Feynman propagator* relative to the "in" and "out" regions: i.e., that Green's function of F which, regarded as a function of either of its arguments, has purely positive-frequency behavior in the "out" region and purely negative-frequency behavior in the "in" region. [Equation (182) is most easily derived by introducing an external source $J(x)$ between the "in" and "out" regions (coupled linearly to $\varphi(x)$ in S), taking the second variation of W with respect J, and then setting J equal to zero.] Therefore

$$\tfrac{1}{2}\,\mathrm{e}^{-\mathrm{i}W}\langle \mathrm{out},\mathrm{vac}|[\varphi(x),\varphi(x')]_+|\mathrm{in},\mathrm{vac}\rangle = -\mathrm{i}[G(x,x') - \bar{G}(x,x')] \equiv \tfrac{1}{2}G^{(1)}(x,x'). \tag{183}$$

$G^{(1)}$ is sometimes known as "Hadamard's elementary function".

In terms of the various Green's functions above the variational law (177) may be rewritten in the compact forms (labels suppressed)

$$\delta W = \tfrac{1}{4}\mathrm{Tr}(G^{(1)}\delta F) = -\tfrac{1}{2}\mathrm{i}\,\mathrm{Tr}(G\,\delta F - \bar{G}\,\delta F) = -\tfrac{1}{2}\mathrm{i}\,\mathrm{Tr}\,(G\,\delta F - G_{\substack{\mathrm{adv}\\ \mathrm{ret}}}\,\delta F)\,. \tag{184}$$

Because

$$FG = -1, \qquad FG_{\substack{\mathrm{adv}\\ \mathrm{ret}}} = -1$$

eq. (184) may be formally integrated, yielding

$$W = -\tfrac{1}{2}\mathrm{i}(\ln\det G - \ln\det G_{\substack{\mathrm{adv}\\ \mathrm{ret}}}) + \Phi \tag{186}$$

$$\mathrm{e}^{\mathrm{i}W} = \mathrm{e}^{\mathrm{i}\Phi}\,(\det G)^{1/2}/(\det G_{\substack{\mathrm{adv}\\ \mathrm{ret}}})^{1/2}\,, \tag{187}$$

where Φ is a (necessarily real) constant metric-independent phase.

6.7. *The Schwinger formalism*

Equation (184) shows that in order to compute the functional derivative of W with respect to the metric it suffices to know the Green's functions G and $\bar{G}$. A general knowledge of these functions is, of course, as difficult to obtain as the expectation value of the stress tensor itself. But because F (and hence δF) is a local differential operator we need to know how these functions behave only when the two points x and x' are close together. For this purpose a technique due to Schwinger [43] is particularly useful. One introduces a fictitious (i.e., non-quantum-mechanical) Hilbert space, a set of formal operators $\boldsymbol{x}^\mu, \boldsymbol{p}_\nu$ satisfying the commutation relations

$$[\boldsymbol{x}^\mu,\boldsymbol{x}^\nu] = 0, \qquad [\boldsymbol{x}^\mu,\boldsymbol{p}_\nu] = \mathrm{i}\,\delta^\mu{}_\nu, \qquad [\boldsymbol{p}_\mu,\boldsymbol{p}_\nu] = 0\,, \tag{188}$$

and a set of eigenvectors $|x\rangle$ of the $\boldsymbol{x}^\mu$, normalized according to

* Definition (182) for the Feynman propagator differs from the conventional one by a factor i. The choice (182) has the advantage that all the Green's functions then satisfy the same equation: $FG(x,x') = -\delta(x,x')$.

$$\langle x|x'\rangle = \delta(x, x') . \tag{189}$$

One then writes

$$\left.\begin{aligned} g^{-1/4}(x)\, F(x, x')\, g^{-1/4}(x') &= \langle x|g^{-1/4} F g^{-1/4}|x'\rangle \\ g^{1/4}(x)\, G(x, x')\, g^{1/4}(x') &= \langle x|g^{1/4} G g^{1/4}|x'\rangle \end{aligned}\right\} \tag{190}$$

where [see eq. (150)]

$$g^{-1/4} F g^{-1/4} = -g^{-1/4} p_\mu g^{1/2} g^{\mu\nu} p_\nu g^{-1/4} - \xi R - m^2 , \tag{191}$$

$$g \equiv -\det (g_{\mu\nu}), \quad g_{\mu\nu} \equiv g_{\mu\nu}(x), \text{ etc.}, \tag{192}$$

and

$$F G = -1 . \tag{193}$$

The basis vectors $|x\rangle$ transform as densities of weight $\frac{1}{2}$ under coordinate transformations, and the reason for affixing the factors $g^{1/4}$, $g^{-1/4}$ above is to obtain operators that leave these transformation properties intact.

6.8. *The Feynman propagator and the WKB expansion*

It turns out that a knowledge of the Feynman propagator automatically yields a knowledge of the Green's functions, G_{ret}, G_{adv} and $\bar{G}$. Hence we may confine our attention to the former. In flat spacetime it is well known that the boundary conditions on the Feynman propagator are automatically secured by giving m^2 an infinitesimal negative imaginary part or, what is the same thing, by giving the operator F an infinitesimal positive imaginary part. This simple rule may be verified in perturbation theory, as well as by more powerful techniques, and continues to hold in the present curved-spacetime context. The only restriction is that the points x and x' must be taken to lie either in the "in" region, the "out" region, or the region in between, and that all integrations (and variations) must be confined to this domain.

Accordingly we may write

$$g^{1/4} G g^{1/4} = -\frac{1}{g^{-1/4} F g^{-1/4} + \mathrm{i}0} = \mathrm{i} \int_0^\infty \exp (\mathrm{i}\, g^{-1/4} F g^{-1/4} s)\, \mathrm{d}s , \tag{194}$$

which yields

$$G(x, x') = \mathrm{i} \int_0^\infty g^{-1/4}(x)\, \langle x, s|x', 0\rangle\, g^{-1/4}(x')\, \mathrm{d}s , \tag{195}$$

where

$$\langle x, s|x', 0\rangle \equiv \langle x|\exp (\mathrm{i}\, g^{-1/4} F g^{-1/4} s)|x'\rangle . \tag{196}$$

The matrix element (196) may be regarded as the transition amplitude for a fictitious dynamical system. It satisfies the "Schrödinger equation"

$$\mathrm{i}\,\frac{\partial}{\partial s}\langle x, s|x', 0\rangle = (-\nabla_\mu\nabla^\mu + \xi R + m^2)\,\langle x, s|x', 0\rangle\,, \tag{197}$$

where "∇_μ" denotes the covariant derivative. For the study of this amplitude when x and x' are close together a WKB expansion suffices:

$$\langle x, s|x', 0\rangle = -\frac{\mathrm{i}}{(4\pi)^2}\,\frac{D^{1/2}(x, x')}{s^2}\,\exp\left[\mathrm{i}\,\frac{\sigma(x, x')}{2s} - \mathrm{i}m^2 s + \Omega(x, x', s)\right]. \tag{198}$$

Here Ω is a function, presently to be determined, that vanishes when $s = 0$, σ is one-half the square of the distance along the geodesic between x and x', and D is the 4×4 determinant

$$D \equiv -\det(-\partial^2\sigma/\partial x^\mu \partial x'^\nu)\,. \tag{199}$$

Apart from a factor $(16s^4)^{-1}$, D is the *Van Vleck–Morette determinant* (see Van Vleck [49] and Morette [36]) for a dynamical system with action function $\sigma/2s$ (i.e., for a particle of mass $\frac{1}{2}$ executing geodesic motion in a four-dimensional manifold of signature $(-, +, +, +)$). It satisfies the identity

$$D^{-1}(D\sigma_;{}^\mu)_{;\mu} = 4\,, \tag{200}$$

which may be derived from the "Hamilton–Jacobi equation"

$$\sigma = \tfrac{1}{2}\sigma_{;\mu}\sigma_;{}^\mu = \tfrac{1}{2}g^{\mu\nu}\sigma_{,\mu}\sigma_{,\nu}\,. \tag{201}$$

D and σ may be regarded as single-valued functions when x and x' are sufficiently close to one another. Strictly speaking, when there is more than one geodesic connecting x and x', expression (198) should be modified by the addition of extra terms, one for each geodesic*. In particular, when 3-space is compact there is an infinite number of such geodesics, and an infinite number of terms is needed (see Dowker [21, 22]). However, it is only the leading term that gives rise to the divergences of the theory, so I keep only it. The numerical factors in front are chosen so as to secure the normalization

$$\langle x, s|x', 0\rangle \xrightarrow[s \to 0]{} \delta(x, x')\,. \tag{202}$$

Substituting (198) into (197) and making use of (200) and (201), one finds that the function Ω must satisfy the differential equation

$$\mathrm{i}\partial\Omega/\partial s + D^{-1}(D\Omega_;{}^\mu)_{;\mu} + \Omega_{;\mu}\Omega_;{}^\mu + \mathrm{i}s^{-1}\sigma_;{}^\mu\Omega_{;\mu} = \xi R - D^{-1/2}D^{1/2}{}_{;\mu}{}^\mu\,. \tag{203}$$

The solution that vanishes as s goes to zero may be expressed as a power series,

$$\Omega(x, x', s) = \sum_{n=1}^{\infty} a_n(x, x')\,(\mathrm{i}s)^n\,, \tag{204}$$

where the coefficients are determined by the differential recursion relations

* If x and x' are almost-conjugate points along one of these geodesics then an Airy interpolation (or generalization thereof) of the WKB form must be used for the corresponding term.

$$\left.\begin{aligned} &\sigma_{;}{}^{\mu}a_{1;\mu} + a_1 = D^{-1/2}D^{1/2}{}_{;\mu}{}^{\mu} - \xi R \\ &\sigma_{;}{}^{\mu}a_{2;\mu} + 2a_2 = D^{-1}(Da_{1;}{}^{\mu})_{;\mu} \\ &\sigma_{;}{}^{\mu}a_{n;\mu} + na_n = D^{-1}(Da_{n-1;}{}^{\mu})_{;\mu} + \sum_{r=1}^{n-2} a_{r;\mu}a_{n-r-1;}{}^{\mu}, \quad n = 3,4\ldots \end{aligned}\right\} \tag{205}$$

These relations may in principle be solved, in succession, by integrating along the geodesics emanating from x'.

6.9. Series expansions

We now have

$$G(x,x') = \frac{\Delta^{1/2}}{(4\pi)^2}\int_0^\infty \frac{1}{s^2}\exp\{\mathrm{i}(\sigma/2s - m^2 s)\}\,\mathrm{e}^{\Omega(s)}\,\mathrm{d}s\,, \tag{206}$$

where

$$\Delta(x,x') \equiv g^{-1/2}(x)\,D(x,x')\,g^{-1/2}(x')\,. \tag{207}$$

The integral (206) may be evaluated as an asymptotic series in inverse powers of m^2 by expanding the factor $\exp\{\Omega(s)\}$ in a power series in s and integrating term by term:

$$\begin{aligned} G(x,x') &\sim \frac{\Delta^{1/2}}{(4\pi)^2}\exp\left[\sum_{n=1}^\infty a_n\left(-\frac{\partial}{\partial m^2}\right)^n\right]\int_0^\infty \frac{1}{s^2}\exp\{\mathrm{i}(\sigma/2s - m^2 s)\}\,\mathrm{d}s \\ &= -\frac{\Delta^{1/2}}{8\pi}\exp\left[\sum_{n=1}^\infty a_n\left(-\frac{\partial}{\partial m^2}\right)^n\right]\frac{m^2 H_1^{(2)}((-2m^2\sigma)^{1/2})}{(-2m^2\sigma)^{1/2}}\,. \end{aligned} \tag{208}$$

Here $H_1^{(2)}$ is the Hankel function of the second kind, of order 1. Explicitly

$$\begin{aligned} \frac{m^2 H_1^{(2)}((-2m^2\sigma)^{1/2})}{(-2m^2\sigma)^{1/2}} &= \frac{1}{\pi\mathrm{i}}\left(\frac{1}{\sigma+\mathrm{i}0} + 2m^2\left\{[\gamma + \tfrac{1}{2}\ln(m^2/2) + \tfrac{1}{2}\ln(\sigma+\mathrm{i}0)]\right.\right. \\ &\quad\left.\left.\times\left[\tfrac{1}{2} + \frac{2m^2\sigma}{2^2\cdot 4} + \frac{(2m^2\sigma)^2}{2^2\cdot 4^2\cdot 6} + \ldots\right] - \tfrac{1}{4} - \frac{2m^2\sigma}{2^2\cdot 4}(1+\tfrac{1}{4}) - \frac{(2m^2\sigma)^2}{2^2\cdot 4^2\cdot 6}(1+\tfrac{1}{2}+\tfrac{1}{6}) - \ldots\right\}\right)\,. \end{aligned} \tag{209}$$

The i0's appear because the Green's function G is actually a *distribution*, being the "boundary value" of a function analytic in the upper half σ-plane. With the aid of the relations

$$\frac{1}{\sigma+\mathrm{i}0} = \frac{1}{\sigma} - \pi\mathrm{i}\delta(\sigma), \qquad \ln(\sigma+\mathrm{i}0) = \ln|\sigma| + \pi\mathrm{i}\theta(-\sigma), \tag{210}$$

one easily separates G into its components $\bar{G}$ and $G^{(1)}$ (eq. (183)):

$$\begin{aligned} \bar{G}(x,x') &= \frac{\Delta^{1/2}}{8\pi}\exp\left[\sum_{n=1}^\infty a_n\left(-\frac{\partial}{\partial m^2}\right)^n\right]\left\{\delta(\sigma) - m^2\theta(-\sigma)\left[\tfrac{1}{2} + \frac{2m^2\sigma}{2^2\cdot 4} + \frac{(2m^2\sigma)^2}{2^2\cdot 4^2\cdot 6} + \ldots\right]\right\} \\ &= \frac{\Delta^{1/2}}{8\pi}\{\delta(\sigma) - m^2\theta(-\sigma)(\tfrac{1}{2} + \tfrac{1}{8}m^2\sigma + \ldots) + a_1\theta(-\sigma)(\tfrac{1}{2} + \tfrac{1}{4}m^2\sigma + \ldots) \\ &\quad - (\tfrac{1}{2}a_1^2 + a_2)\sigma\,\theta(-\sigma)(\tfrac{1}{4} + \tfrac{1}{16}m^2\sigma + \ldots) + \ldots\}\,, \end{aligned} \tag{211}$$

$$G^{(1)}(x,x') \sim \frac{\Delta^{1/2}}{4\pi^2}\left\{\frac{1}{\sigma} + m^2(\gamma + \tfrac{1}{2}\ln|m^2\sigma/2|)(1+\tfrac{1}{4}m^2\sigma+\ldots) - \tfrac{1}{2}m^2 - \tfrac{5}{16}m^4\sigma - \ldots\right.$$
$$- a_1[(\gamma+\tfrac{1}{2}\ln|m^2\sigma/2|)(1+\tfrac{1}{2}m^2\sigma+\ldots) - \tfrac{1}{2}m^2\sigma - \ldots]$$
$$+ (\tfrac{1}{2}a_1^2 + a_2)\sigma[(\gamma+\tfrac{1}{2}\ln|m^2\sigma/2|)(\tfrac{1}{2}+\tfrac{1}{8}m^2\sigma+\ldots) - \tfrac{1}{4} - \ldots] + \ldots$$
$$\left. + \frac{1}{2m^2}[\tfrac{1}{2}a_1^2 + a_2 + \mathrm{O}(\sigma)] + \frac{1}{2m^4}[\tfrac{1}{6}a_1^3 + a_1a_2 + a_3 + \mathrm{O}(\sigma)] + \ldots\right\}. \tag{212}$$

Although the Green's function $\bar{G}$ is real, the function $G^{(1)}$ is generally not. This fact is not revealed by eq. (212). An asymptotic expansion in inverse powers of m^2 is incapable of yielding the imaginary part, which is nonvanishing whenever particle production occurs.

6.10. The effective Lagrangian

Return now to equation (184) which, in the Schwinger formalism, may be rewritten in the form

$$\delta W = \tfrac{1}{4}\mathrm{Tr}\,[g^{1/4}G^{(1)}g^{1/4}\delta(g^{-1/4}Fg^{-1/4})]\,. \tag{213}$$

In virtue of the preceding analysis this is equivalent to

$$\delta W = -\tfrac{1}{2}\mathrm{i}\,\mathrm{Tr}\,[g^{1/4}Gg^{1/4}\,\delta(g^{-1/4}Fg^{-1/4})]\,, \tag{214}$$

provided we throw out the i0's that appear when G is expressed as a function of σ. Inserting eq. (194) into (214), we get

$$\delta W = \tfrac{1}{2}\mathrm{Tr}\int_0^\infty \exp(\mathrm{i}g^{-1/4}Fg^{-1/4}s)\,\delta(g^{-1/4}Fg^{-1/4})\,\mathrm{d}s$$
$$= -\frac{\mathrm{i}}{2}\delta\,\mathrm{Tr}\int_0^\infty s^{-1}\exp(\mathrm{i}g^{-1/4}Fg^{-1/4}s)\,\mathrm{d}s = \delta\int L\,\mathrm{d}^4x\,, \tag{215}$$

and hence

$$W = \int L\,\mathrm{d}^4x + \text{const.} \tag{216}$$

where the spacetime integral is over the domain between the "in" and "out" regions, and the function L is an *effective Lagrangian*, given (see eqs. (196) and (198)) by

$$L = -\frac{\mathrm{i}}{2}\int_0^\infty s^{-1}\langle x,s|x,0\rangle\,\mathrm{d}s$$
$$= -(32\pi^2)^{-1}g^{1/2}\int_0^\infty s^{-3}\exp[-\mathrm{i}m^2s + \Omega(x,x,s)]\,\mathrm{d}s + \text{terms arising from multiple geodesics (if any).} \tag{217}$$

In passing to the last line use has been made of the *coincidence limits*

$$\sigma \underset{x'\to x}{\longrightarrow} 0, \quad \sigma_{;\mu} \underset{x'\to x}{\longrightarrow} 0, \quad \sigma_{;\mu\nu} \underset{x'\to x}{\longrightarrow} g_{\mu\nu}, \quad \frac{\partial^2\sigma}{\partial x^\mu\,\partial x'^\nu} \underset{x'\to x}{\longrightarrow} -g_{\mu\nu}, \quad D \underset{x'\to x}{\longrightarrow} g\,. \tag{218}$$

The integral (217) diverges at the lower limit. Note, however, that it is formally a scalar density *constructed entirely out of spacetime geometry*. This means that W, given by eq. (216), is formally a coordinate-invariant geometrical quantity. If we can find a method for splitting off its infinities in a coordinate invariant and metric-independent way, the remainder, W_{reg}, will automatically yield a conserved contribution to the expectation value of the stress tensor:

$$2(\delta W_{\text{reg}}/\delta g_{\mu\nu})_{;\nu} = 0. \tag{219}$$

This means that the expectation value itself will be conserved in virtue of the differential equations satisfied by the basis functions.

6.11. The method of "background field". Identity of the single-loop and WKB approximations

Now is perhaps the time to explain why the formalism being presented here is called "the method of the background field". I remarked earlier that the method is applicable to fields of any spin. *This includes the gravitational field itself.* The geometry on which W depends is then a *classical background geometry* and the contributions to W, of the type considered in this section, come from the linear fluctuations away from this geometry implied by the quantum theory. It is sometimes erroneously believed that the method of the background field stops here, i.e., that it is *merely* the quantum theory of a linearized field on a classical background. Actually the method embraces the full gamut of self-interactions implied by the nonlinear character of the gravitational field. The computations of this section (or rather their equivalents for the pure gravitational field) yield only the first approximation to W, and hence to the vacuum-to-vacuum amplitude. This approximation is, in fact, the WKB approximation and is sometimes written by including the classical action in the phase, in the form

$$\langle \text{out, vac}|\text{in, vac}\rangle = \mathrm{e}^{\mathrm{i}(S+W)} \approx \frac{(\det G)^{1/2}}{(\det G_{\substack{\text{adv}\\ \text{ret}}})^{1/2}}\, \mathrm{e}^{\mathrm{i}S}\,, \tag{220}$$

where the constant Φ of eq. (187) has been absorbed into S.* The ratio of the two determinants in front may be shown to be precisely the Van Vleck–Morette determinant for the classical trajectory (history) followed by the background field.

In conventional terminology the WKB approximation in field theory is known as the *single-closed-loop approximation*. Higher order approximations can be put into one-to-one correspondence with diagrams having more than one closed loop, in the familiar manner. The only difference from textbook convention is that the propagators associated with internal lines are propagators for curved rather than flat spacetime. If several interacting fields are to be considered at once then eq. (220) still holds provided a combined background field is introduced and S and W are the *total* action and vacuum correction respectively. It is well known that the nth-order functional derivatives of $S + W$ with respect to the background fields are the total nth-order vertex functions of the theory, and the exact S-matrix is obtained by replacing S by $S + W$ and calculating all am-

* Here S is the action function for the full nonlinear gravitational field.

plitudes in the tree approximation only. Since the S-matrix ultimately governs the coherent scattering of large-amplitude (classical) waves, as well as of individual particles, it follows that $S + W$ and *not* S, is the effective action functional even for macroscopic fields. For this reason the proposal to take real and virtual quantum processes into account in general relativity, by writing Einstein's equation in the form[☆]

$$G^{\mu\nu} = 8\pi \langle T^{\mu\nu} \rangle, \tag{221}$$

is not merely of heuristic validity but is *the correct way*. For it is a modified version of

$$\delta(S+W)/\delta g_{\mu\nu} = 0, \tag{222}$$

which includes the vacuum-polarization part of W as well as the effects of real particle production. For the same reason it is absolutely correct to cancel the infinities of W by counter terms in S. Only under conditions of extreme energies or curvatures, when W develops a very large imaginary part and particle production is so excessive as to become meaningless, does the semiclassical conception embodied in eq. (221) provide an inadequate description of the physics.

6.12. Isolation of the divergences by Schwinger's method

Let us turn now to the divergences. I shall first describe briefly Schwinger's method [43] for handling them. He begins by rotating the integration contour of eq. (217) into the negative imaginary axis, which is equivalent to making the replacement $s = -\mathrm{i}\, t$. The function Ω is real on the imaginary axis. Therefore if it has poles in the lower half plane these must be symmetrically distributed in both quadrants, and the rotation process will pick up the residues from those on the right. These residues will generally make contributions to the imaginary part of W and hence the poles (or more complicated singularities such as branch points) *must* be there whenever real particle production occurs. Keeping only the real part, we have

$$\mathrm{Re} L = (32\pi^2)^{-1} g^{1/2} \int_0^\infty t^{-3} \exp\,[-m^2 t + \Omega(x, x, -\mathrm{i}t)]\, \mathrm{d}t + \text{terms arising from multiple geodesics.} \tag{223}$$

We see that the infinities of W are confined to $\mathrm{Re}W$.

Schwinger isolates the infinities simply by expanding e^{Ω} about $t = 0$:

$$L_{\mathrm{div}} = (32\pi^2)^{-1} g^{1/2} \left\{ \int_0^\infty t^{-3} \mathrm{e}^{-m^2 t} \mathrm{d}t + a_1(x, x) \int_0^\infty t^{-2} \mathrm{e}^{-m^2 t} \mathrm{d}t + [\tfrac{1}{2} a_1^2(x, x) + a_2(x, x)] \int_0^\infty t^{-1}\, \mathrm{e}^{-m^2 t} \mathrm{d}t \right\}. \tag{224}$$

The coefficients of the divergent integrals can be determined in a straightforward, if slightly tedious, manner by taking repeated covariant derivatives of eqs. (200), (201) and (205), and making use of the coincidence limits (218). One finds

$$a_1(x,x) = (\tfrac{1}{6} - \xi) R\,, \tag{225}$$

[☆] The $\langle T^{\mu\nu} \rangle$ on the right hand side of this equation includes contributions from gravitons.

$$a_2(x,x) = -\tfrac{1}{180} R_{\mu\nu}R^{\mu\nu} + \tfrac{1}{180} R_{\mu\nu\sigma\tau}R^{\mu\nu\sigma\tau} + \tfrac{1}{6}(\tfrac{1}{5} - \xi)\, R_{;\mu}{}^{\mu}\,, \tag{226}$$

where $R_{\mu\nu\sigma\tau}$ and $R_{\mu\nu}$ are the Riemann and Ricci tensors respectively. The terms in ξR come from the scalar field equation (150). All other terms arise simply from the failure of covariant differentiation to be commutatitve; hence they represent a purely geometrical contribution to the divergences. It is immediately apparent that cancellation of the first two terms inside the curly brackets of eq. (224) by counter terms in the gravitational action is equivalent to renormalization of the cosmological and gravitational constants. The third term is of a form not normally included in the classical action. Terms of these three types are familiar to all workers in quantum gravity.

6.13. Relation of Schwinger's method to other methods

What, now, is the relation of Schwinger's method to that of Zel'dovich and Starobinsky? The latter authors introduced their method in the context of a flat 3-space, so they were able to select a momentum variable $\boldsymbol{k}$ as a label for their basis functions. Their technique for regularizing the stress tensor may be summed up in the formula

$$\langle T^{\mu\nu}(x)\rangle_{\text{reg}} = \int T^{\mu\nu}_{\text{reg}}(\boldsymbol{k},x)\,\mathrm{d}^3\boldsymbol{k}\,, \tag{227}$$

where

$$T^{\mu\nu}_{\text{reg}}(\boldsymbol{k},x) \equiv \lim_{n\to\infty}\Big\{T^{\mu\nu}(u(m,\boldsymbol{k}|x), u^*(m,\boldsymbol{k}|x)) - \left[1 + \frac{\partial}{\partial(n^{-2})} + \frac{1}{2}\,\frac{\partial}{\partial(n^{-2})^2}\right]\frac{1}{n}\; T^{\mu\nu}(u(nm,n\boldsymbol{k}|x), u^*(nm,n\boldsymbol{k}|x))\Big\}\,, \tag{228}$$

the dependence of the basis functions on the rest mass m being explicitly indicated. Parker and Fulling were able to show that this formula is equivalent to a method of "adiabatic regularization", which they were able to generalize for application also to a class of curved 3-spaces. In the latter method each term in the mode-sum is compared with what it would be if the time rate of change of the metric were slowed down and the curvature of 3-space were correspondingly decreased. An expansion is carried out in inverse powers of the "slowness parameter" that measures this decrease, and the first three terms (of orders 0, −2, and −4) are thrown away. Parker and Fulling show that the role of the slowness parameter is identical with that of the parameter n in eq. (228).

In the geometries considered by the above authors a preferred set of spatial coordinates exist, and the regularization method depends on them. Furthermore, the regularization is performed mode by mode. What analogous method can we possibly adopt in the general case? What do we do when there is no obvious mode decomposition? The answer is to split the points at which the field operators appearing in $T^{\mu\nu}$ are evaluated. The infinities at once disappear (provided splitting in a null direction is avoided), and the regularization (228) may be carried out either before or after the integration (227) is performed. Furthermore, rescaling of the variable $\boldsymbol{k}$ is permitted at any stage. It is not difficult to see that the regularization (228) is then equally well effected by leaving $\boldsymbol{k}$ alone, multiplying $x - x'$ by n^{-1}, and affixing the factor n^{-4} instead of n^{-1} to the second term. This procedure translates immediately into the covariant language of eq. (217). We have only to reinsert the term $\mathrm{i}\sigma/2s$ into the exponent of the integrand and write (ignoring the multiple-geodesic terms)

$$L_{\text{reg}} \equiv -\lim_{x'\to x} (32\pi^2)^{-1} g^{1/2} \int_0^\infty \lim_{\lambda\to 0} \frac{1}{s^3} \Big[\exp\{\mathrm{i}(\sigma/2s - m^2 s) + \Omega(s)\}$$

$$-\left(1 + \frac{\partial}{\partial\lambda} + \frac{1}{2}\frac{\partial^2}{\partial\lambda^2}\right) \lambda^2 \exp\{\mathrm{i}(\lambda\sigma/2s - m^2 s/\lambda) + \Omega(s)\}\Big] \mathrm{d}s\,, \tag{229}$$

the symbol λ being now introduced in place of n^{-2}. As long as σ is different from zero the oscillatory behavior of the exponential prevents the integral from diverging at the lower limit. Furthermore s in the second term may be subjected to the rescaling $s \to \lambda s$, which then yields

$$L_{\text{reg}} = -\lim_{x'\to x} (32\pi^2)^{-1} g^{1/2} \int_0^\infty \lim_{\lambda\to 0} \frac{1}{s^3} \exp\{\mathrm{i}(\sigma/2s - m^2 s)\}\left[\mathrm{e}^{\Omega(s)} - \left(1 + \frac{\partial}{\partial\lambda} + \frac{1}{2}\frac{\partial^2}{\partial\lambda^2}\right)\mathrm{e}^{\Omega(\lambda s)}\right] \mathrm{d}s$$

$$= -(32\pi^2)^{-1} g^{1/2} \int_0^\infty \frac{1}{s^3} \exp(-\mathrm{i}m^2 s)[\mathrm{e}^{\Omega(s)} - 1 - \mathrm{i}a_1 s + (\tfrac{1}{2}a_1^2 + a_2)s^2]\,\mathrm{d}s = L - L_{\text{div}}. \tag{230}$$

The Zel'dovich–Starobinsky scheme, generalized in this manner, is seen to yield *exactly the same regularization as Schwinger's.*

Schwinger has pointed out that his scheme, which may be generalized also to multi-loop processes (see Bogoliubov and Shirkov [2]), is capable of regularizing anything, provided only that integration over the parameter s (which he calls a "proper-time" parameter) is reserved to the last. He has shown that his scheme may be regarded as a kind of ultimate extension of the Pauli–Villars method and claims that it is guaranteed to preserve every invariance (e.g., gauge invariance) that is present in the formal theory. It certainly preserves general coordinate invariance, as we have seen, and hence guarantees the conservation law (219). But there is a question whether it preserves conformal invariance.

6.14. Conformal invariance

The stress tensor associated with the field equation (150) is given by

$$T^{\mu\nu} = g^{1/2}\{\tfrac{1}{2}(1-2\xi)[\varphi_;{}^{\mu}, \varphi_;{}^{\nu}]_+ + (2\xi - \tfrac{1}{2}) g^{\mu\nu}\varphi_{;\sigma}\varphi_;{}^{\sigma} - \xi[\varphi, \varphi_;{}^{\mu\nu}]_+ + \xi g^{\mu\nu}[\varphi, \varphi_{;\sigma}{}^{\sigma}]_+$$
$$+ \xi(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R)\varphi^2 - \tfrac{1}{2}m^2 g^{\mu\nu}\varphi^2\}\,. \tag{231}$$

In the conformally invariant massless case (see Callan, Coleman and Jackiw [6]) this reduces to

$$T^{\mu\nu} = \tfrac{1}{6} g^{1/2}\{2[\varphi_;{}^{\mu}, \varphi_;{}^{\nu}]_+ - g^{\mu\nu}\varphi_{;\sigma}\varphi_;{}^{\sigma} - [\varphi, \varphi_;{}^{\mu\nu}]_+ + g^{\mu\nu}[\varphi, \varphi_{;\sigma}{}^{\sigma}]_+ + (R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R)\varphi^2\}\,, \tag{232}$$

which formally satisfies

$$T^{\mu}_{\mu} = 0\,, \tag{233}$$

in virtue of the field equations. Equation (233) also follows as a consequence of the conformal invariance of the action functional for this case: S remains invariant under the infinitesimal changes

$$\delta g_{\mu\nu} = g_{\mu\nu}\,\delta\lambda, \quad \delta\varphi = -\tfrac{1}{2}\varphi\,\delta\lambda\,, \tag{234}$$

where $\delta\lambda$ is an arbitrary infinitesimal scalar function. If we can show that W too is invariant under (234) it will follow that

$$g^{\mu\nu}(\delta W/\delta g_{\mu\nu}) \equiv 0 \,, \tag{235}$$

and hence

$$T_{\mathrm{reg}\,\mu}{}^{\mu} = 0 \,, \tag{236}$$

provided that the divergent part that is split off from W is also conformally invariant.

The conformal invariance of W may be proved formally as follows: One first verifies that the Riemann tensor and its contractions transform under (234) according to

$$\begin{aligned} \delta R_{\mu\nu\sigma\tau} &= R_{\mu\nu\sigma\tau}\,\delta\lambda - \tfrac{1}{2}(g_{\mu\sigma}\delta\lambda_{;\nu\tau} + g_{\nu\tau}\delta\lambda_{;\mu\sigma} - g_{\mu\tau}\delta\lambda_{;\nu\sigma} - g_{\nu\sigma}\delta\lambda_{;\mu\tau}), \\ \delta R_{\mu\nu} &= -\tfrac{1}{2}(2\delta\lambda_{;\mu\nu} + g_{\mu\nu}\delta\lambda_{;\sigma}{}^{\sigma}), \quad \delta R = -R\delta\lambda - 3\delta\lambda_{;\mu}{}^{\mu} \,. \end{aligned} \tag{237}$$

One then shows that when $m = 0$ and $\xi = \frac{1}{6}$ the operator (191) has the particularly simple transformation law

$$\delta(g^{-1/4}Fg^{-1/4}) = -\tfrac{1}{2}[g^{-1/4}Fg^{-1/4}, \delta\boldsymbol{\lambda}]_{+} \,. \tag{238}$$

In virtue of the fact that

$$FG^{(1)} = 0, \quad G^{(1)}F = 0, \tag{239}$$

it follows immediately from eq. (213) that

$$\delta W = 0. \tag{240}$$

6.15. The Weyl tensor and the generalized Gauss–Bonnet invariant

Now the only local geometrical conformal invariant that can be constructed from the metric tensor and its first and second derivatives is $g^{1/2}C_{\mu\nu\sigma\tau}C^{\mu\nu\sigma\tau}$ (or functions thereof) where $C_{\mu\nu\sigma\tau}$ is the *Weyl tensor*[☆]:

$$C_{\mu\nu\sigma\tau} \equiv R_{\mu\nu\sigma\tau} - \tfrac{1}{2}(g_{\mu\sigma}R_{\nu\tau} + g_{\nu\tau}R_{\mu\sigma} - g_{\mu\tau}R_{\nu\sigma} - g_{\nu\sigma}R_{\mu\tau}) + \tfrac{1}{6}(g_{\mu\sigma}g_{\nu\tau} - g_{\mu\tau}g_{\nu\sigma})R \,, \tag{241}$$

$$\delta C_{\mu\nu\sigma\tau} = C_{\mu\nu\sigma\tau}\delta\lambda \,. \tag{242}$$

This implies that the only counter term that can be used to cancel the divergences of W and at the same time preserve the conformal invariance of the theory must have the form

$$\text{const.} \times \int g^{1/2} C_{\mu\nu\sigma\tau}C^{\mu\nu\sigma\tau}\,\mathrm{d}^4x \equiv \text{const.} \times \int g^{1/2}(R_{\mu\nu\sigma\tau}R^{\mu\nu\sigma\tau} - 2R_{\mu\nu}R^{\mu\nu} + \tfrac{1}{3}R^2)\,\mathrm{d}^4x \,. \tag{243}$$

This integral may be simplified by making use of the well-known fact (see Chern [11, 12]) that the integral

$$\int g^{1/2}(R_{\mu\nu\sigma\tau}R^{\mu\nu\sigma\tau} - 4R_{\mu\nu}R^{\mu\nu} + R^2)\,\mathrm{d}^4x \,, \tag{244}$$

☆ The Weyl tensor vanishes whenever spacetime is conformally flat. It also satisfies $C_{\mu\sigma\nu}{}^{\sigma} \equiv 0$.

is a topological invariant, *independent of the geometry*, for any 4-manifold. (It is an analog of the well known Gauss–Bonnet invariant for 2-manifolds.) Thus

$$\text{const.} \times \int g^{1/2} C_{\mu\nu\sigma\tau} C^{\mu\nu\sigma\tau} \, \mathrm{d}^4 x = \text{const.} \times \int g^{1/2} (R_{\mu\nu} R^{\mu\nu} - \tfrac{1}{3} R^2) \, \mathrm{d}^4 x + \text{const.} \tag{245}$$

6.16. Apparent failure of conformal invariance in Schwinger's method. Resolution of the difficulty

Referring to eqs. (224), (225) and (226), we see that the three divergent terms isolated by Schwinger's method in the conformally invariant case take respectively the forms

$$\text{const.} \times \int g^{1/2} \, \mathrm{d}^4 x, \quad 0, \quad \text{and} \tag{246}$$
$$\text{const.} \times \int g^{1/2} (R_{\mu\nu\sigma\tau} R^{\mu\nu\sigma\tau} - R_{\mu\nu} R^{\mu\nu} + R_{;\mu}{}^{\mu}) \, \mathrm{d}^4 x = \text{const.} \times \int g^{1/2} (R_{\mu\nu} R^{\mu\nu} - \tfrac{1}{3} R^2) \, \mathrm{d}^4 x + \text{const.}$$

The last of these terms has exactly the form (245). The second one vanishes; hence there is no renormalization of the gravitation constant. But the first term, which renormalizes the cosmological constant, is *not* conformally invariant and does *not* vanish. How do we deal with this fact?

One way is to adopt the dimensional regularization method (see 't Hooft and Veltman [33]). In this method the cosmological term automatically vanishes. But it does so because a special convention is implicitly adopted in the analytic continuation to complex dimensions. Nouri-Moghadan and Taylor [37] have shown that as far as divergent terms are concerned, this analytic continuation is inherently ambiguous. Furthermore, in its present form the method is applicable only to Lorentz or de Sitter-invariant quasi-linearized theories☆.

A more direct approach is clearly desirable, and the less sophisticated the better. As always, the point-splitting method suggests itself: If the term in $\mathrm{i}\sigma/2s$ is reinserted into the exponent of the integrand of eq. (217) we have (neglecting the multiple-geodesic corrections)

$$\begin{aligned}
L &= -\lim_{x' \to x} \frac{D^{1/2}}{32\pi^2} \int_0^\infty \frac{1}{s^3} \exp\{\mathrm{i}(\sigma/2s - m^2 s)\} \, \mathrm{e}^{\Omega(s)} \, \mathrm{d}s \\
&= \lim_{x' \to x} \mathrm{i} \frac{\partial}{\partial \sigma} [g^{1/4}(x) \, G(x, x') \, g^{1/4}(x')] \qquad \text{(see eq. (206))} \\
&= -\lim_{x' \to x} \tfrac{1}{2} \frac{\partial}{\partial \sigma} [g^{1/4}(x) \, G^{(1)}(x, x') \, g^{1/4}(x')] \ \text{(casting out the i0's)} \\
&= \lim_{x' \to x} \frac{D^{1/2}}{8\pi^2} \left\{ \frac{1}{\sigma^2} - \frac{m^2}{2\sigma} - \frac{m^4}{4} (\gamma + \tfrac{1}{2} \ln|m^2 \sigma/2|) + \tfrac{3}{16} m^4 \right. \\
&\quad \left. + a_1 \left[\frac{1}{2\sigma} + \frac{m^2}{2} (\gamma + \tfrac{1}{2} \ln|m^2 \sigma/2|) - \tfrac{1}{4} m^2 \right] - (\tfrac{1}{2} a_1^2 + a_2) [\tfrac{1}{2} (\gamma + \tfrac{1}{2} \ln|m^2 \sigma/2|) - \tfrac{1}{4}] + \ldots \right\} .
\end{aligned} \tag{247}$$

In the final expression, which is derived from eq. (212), only the terms that have counterparts in the divergent Lagrangian (224) are shown. The first line inside the curly brackets obviously cor-

☆ An extension of the method to arbitrary background geometries would certainly be of interest. *There* is a thesis topic for somebody. The extension to deSitter spacetime has only recently been achieved by Candelas and Raine (private communication).

responds to the cosmological term. It is the only term that survives in flat spacetime. Therefore it must yield the scalar version of the vacuum stress (31). The way it does so is instructive. Let $\epsilon^\mu = x^\mu - x'^\mu$ be a constant coordinate displacement. When x and x' are close together one may write $\sigma = \frac{1}{2} g_{\mu\nu}\epsilon^\mu \epsilon^\nu$. When $m = 0$ the "cosmological" part of W therefore takes the form

$$W_{\text{cosm}} = \frac{1}{2\pi^2}\int g^{1/2}(g_{\mu\nu}\epsilon^\mu\epsilon^\nu)^{-2}\,\mathrm{d}^4x\,. \tag{248}$$

This expression, depending as it does on the frame selected by ϵ^μ, is not coordinate invariant. It is however, *conformally invariant* and yields a traceless contribution to $\langle T^{\mu\nu}\rangle$:

$$2\frac{\delta W_{\text{cosm}}}{\delta g_{\mu\nu}} = \frac{g^{1/2}}{2\pi^2}\,[g^{\mu\nu}(g_{\alpha\beta}\epsilon^\alpha\epsilon^\beta)^{-2} - 4\epsilon^\mu\epsilon^\nu(g_{\alpha\beta}\epsilon^\alpha\epsilon^\beta)^{-3}]\,. \tag{249}$$

In a local Minkowski frame, with the choice $(\epsilon^\mu) = (\Lambda^{-1}, 0, 0, 0)$, expression (249) is identical, apart from the degrees-of-freedom-factor 2, with expression (31). Under the spatial-averaging method of Fulling and Parker [25], which yields an alternative regularization, one has effectively $(\epsilon^\mu\epsilon^\nu) = \mathrm{diag}\,(0, \frac{1}{3}\Lambda^{-2}, \frac{1}{3}\Lambda^{-2}, \frac{1}{3}\Lambda^{-2})$, producing an expression differing from (31) by a factor $-\frac{1}{6}$.

The lesson to be learned from this is that in dealing with the divergent parts of W one must adopt different procedures depending on whether one wants to display (formally) coordinate invariance or conformal invariance. In either case the finite W_{reg} that is left behind by Schwinger's method, after the infinities have been split off, is both coordinate invariant and conformally invariant. Explicit calculations [13] show that similar results hold for the massless spin-$\frac{1}{2}$ field and the electromagnetic field, both of which are conformally invariant. In each case the gravitation constant remains unrenormalized and the logarithmically divergent part of W takes the form (245). The only noteworthy difference between the fermion and boson cases is that the cosmological terms have opposite signs*. It should be mentioned that in order to obtain these results in the electromagnetic case one must include contributions from the fictitious quanta whose formal presence is necessary to secure gauge invariance (see DeWitt [20]). In calculations in Minkowski space (e.g., standard quantum electrodynamics) these quanta may be completely ignored.

6.17. Particle production and vacuum stress in conformally flat spacetime

Fulling, Parker and Hu [26] have studied the conformally invariant scalar field in cosmological models of the Kasner and Robertson–Walker types. Their analysis of the stress tensor in these models confirms a fact that had been noted earlier by Parker [38], namely, that there is no production of conformally invariant massless quanta in Robertson–Walker universes. This result may be understood formally as a consequence of eq. (240). Every Robertson–Walker universe is conformally equivalent to an Einstein universe of constant radius. The value of W for the two must therefore be the same. But an Einstein universe is static. Therefore $\mathrm{Im}\,W = 0$ (no particle production) for both.

Fulling, Parker and Hu go on to assert, however, that the vacuum stress tensor itself, i.e., $2\delta W/\delta g_{\mu\nu}$, vanishes in these universes. Such an assertion might seem to follow from the fact that

* This is a general rule for fermion versus boson fields. Zumino [54] has shown that in a supersymmetric theory the sum of all contributions to the cosmological term vanishes as a consequence of this rule. Zumino's result holds to all orders in the supersymmetric interactions.

these universes are conformally flat and that one expects $W_{\text{reg}} = 0$ when $R_{\mu\nu\sigma\tau} = 0$. In point of fact, however, W_{reg} *cannot be assumed to vanish merely because spacetime is flat.* For example, consider a static flat universe for which 3-space has the topology $R^2 \times S^1$. The analysis of the vacuum stress in this universe is almost identical with that of the Casimir effect. The vacuum stress is nonvanishing and depends on the circumference of the S^1-cycle. But the detailed derivations and arguments of the above authors can be paralleled, *mutatis mutandis*, in this case, leading to an opposite conclusion. Their conclusions when 3-space has the topology S^3 must therefore also be regarded as suspect☆. In general W_{reg} may be safely inferred to vanish only when spacetime is conformally flat, asymptotically flat, and homeomorphic to Minkowski space.

6.18. *The infrared problem*

Mention must be made of the infrared problem that arises when $m = 0$. When $m \neq 0$ the integrals inside the curly brackets of eq. (224) diverge only at the lower limit. This limit corresponds to the ultraviolet limit $\Lambda \to \infty$. When $m = 0$, however, the third integral diverges at the upper limit as well. This is an infrared divergence. As has been pointed out by Fulling and Parker [25] this divergence should not be included in L_{div}, for otherwise a corresponding divergence would be introduced into W_{reg}. What one must do is to introduce an upper cutoff, T. The precise value of T is arbitrary. If we change T to T' the result will be equivalent to changing the renormalized Lagrangian for the classical gravitational field by an amount equal to

$$\Delta L_{\text{ren}} = (32\pi^2)^{-1} \ln(T'/T)\, g^{1/2} [\tfrac{1}{2} a_1^2(x,x) + a_2(x,x)]. \tag{250}$$

In the conformally invariant case this corresponds to a change in the renormalized classical action given by

$$\Delta S_{\text{ren}} = \frac{\ln(T'/T)}{1920\pi^2} \int g^{1/2} (R_{\mu\nu}R^{\mu\nu} - \tfrac{1}{3}R^2)\, \mathrm{d}^4x = -\Delta W_{\text{reg}}. \tag{251}$$

What we are doing here is writing

$$S + W = S_{\text{ren}} + W_{\text{reg}}, \tag{252}$$

and then shifting integrals of the form (251) arbitrarily back and forth between S_{ren} and W_{reg}. The functional S_{ren} has to be determined by experiment. The arbitrariness (251) indicates that S_{ren} cannot be taken simply in the traditional classical form $(16\pi)^{-1}\int g^{1/2}R\,\mathrm{d}^4x$ (with $G = 1$) but must be assumed to be☆☆

$$S_{\text{ren}} = (16\pi)^{-1}\int g^{1/2}R\,\mathrm{d}^4x + (16\pi\mu^2)^{-1} \int g^{1/2}(R_{\mu\nu}R^{\mu\nu} - \tfrac{1}{3}R^2)\,\mathrm{d}^4x. \tag{253}$$

Experimental relativity is not complete until the value of the constant μ has been determined corresponding to a previously chosen cutoff T. A lower bound on μ may be obtained by considering the celestial mechanics of the solar system. The second term of (253) leads to fractional corrections in the motions of the planets of order $\mathrm{e}^{-\mu r}$ where r is the distance from the sun. The correction is biggest in the case of mercury, for which $r \sim 2 \times 10^{45}$ absolute units. Assuming that the

☆ *Note added.* Larry Ford and Ya. Zel'dovich (private communications) have recently shown in this case that $\langle T_{\text{reg}}{}^{\mu\nu}\rangle = (\hbar c/480\pi^2R^4)\,\mathrm{diag}\,(1, \tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3})$ in a local "rest" frame, where R is the radius of the universe.

☆☆ The renormalized cosmological constant is assumed to be zero.

motions of the planets are known to conform to the choice $\mu^{-1} = 0$ to an accuracy of somewhere between one part in 10^8 and one part in 10^{10} (the final result is quite insensitive to the precise value) we obtain $\mu \geqslant 2 \times 10^{-44}$ absolute units. Hence even if the coefficient in front of the second integral in (253) is as big as 10^{85} this term can have no possible effect on planetary astronomy. A cutoff T corresponds to a maximum wavelength of order $T^{1/2}$. If the sun is regarded as the source of the gravitational field the natural maximum to take is the radius of the sun ($\sim 10^{44}$ absolute units). But even if the maximum wavelength is merely assumed to lie somewhere between the radius of the proton ($\sim 10^{20}$) and the radius of the universe ($\sim 10^{62}$) the resulting uncertainty in S_{ren} (expression (251)) makes, at the level of the solar system, an utterly negligible contribution to the term in μ^2.

In the non-conformally-invariant massless case the infrared situation is somewhat different. In that case (which includes the quantized gravitational field) the scalar $a_1(x, x)$ does not vanish identically. Instead it plays a role analogous to m^2 in the integral (217) and can provide a natural infrared cutoff. It leads to terms of the form $a_2 \ln|R|$ in the effective Lagrangian L. Terms analogous to this, but not fully covariant, have been found by Fulling and Parker [25] and by Ginzburg, Kirzhnits and Lyubushin [29].

In the case $m \neq 0$ the renormalization procedure is more straightforward. One simply drops all the terms appearing in (247) from the effective Lagrangian L. The unwritten terms that remain, beginning with a term in $m^{-2}(\frac{1}{6}a_1^3 + a_1 a_2 + a_3)$, are finite and covariant. They are analogous to the Uehling and Euler–Heisenberg corrections to the Maxwell Lagrangian in quantum electrodynamics. Being the terms of an asymptotic expansion in inverse powers of m^2 they can at best yield only an approximation to W_{reg}. The approximation will be a good one provided the components of the Riemann tensor, in quasistationary orthonormal frames, are small compared to m^2. An approximation to $\langle T^{\mu\nu}_{\text{reg}}\rangle$ can then be obtained by functional differentiation.

6.19. Isolation of mode-sum divergences by the point-splitting method

When the components of the Riemann tensor are not small compared to m^2, or when $m = 0$, one must have recourse to other methods for computing $\langle T^{\mu\nu}_{\text{reg}}\rangle$. Usually this will involve isolating the infinities from the mode-sum (12) directly. Because the basis functions in the mode-sum can be computed only for a very restricted class of geometries, it will generally not be possible to compute $\langle T^{\mu\nu}_{\text{reg}}\rangle$ by taking the functional derivative of W_{reg} with respect to the metric. Therefore much of the formal apparatus of this section would seem to be not very practical. Fortunately this is not so. There is a direct way of *covariantly* identifying the divergent parts of $\langle T^{\mu\nu}\rangle$ that is applicable to any geometry, restricted or not.

The method is based upon the observation that, in virtue of eqs. (174), (183) and (231), one may write

$$2g^{-1/2}\frac{\delta W}{\delta g_{\mu\nu}} = \lim_{x'\to x}\, [\tfrac{1}{2}(1-2\xi)G^{(1)}{}_{;}{}^{\mu\nu'} + (\xi - \tfrac{1}{4})g^{\mu\nu}G^{(1)}{}_{;\sigma}{}^{\sigma'} - \xi G^{(1)}{}_{;}{}^{\mu'\nu'} + \xi g^{\mu\nu}G^{(1)}{}_{;\sigma'}{}^{\sigma'}$$
$$+ \tfrac{1}{2}\xi(R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R)\,G^{(1)} - \tfrac{1}{4}m^2 g^{\mu\nu}G^{(1)}]\,. \tag{254}$$

One may therefore insert eq. (212) into (254), carry out the indicated differentiations, set $\sigma_{;}^{\mu} = -\sigma_{;}{}^{\mu'} = \epsilon^{\mu}$, and proceed as in the analysis of expression (249). Before one does this, however, one must replace $\Delta^{1/2}$ and a_1 in (212) by their expansions

$$\Delta^{1/2}(x,x') = 1 + \tfrac{1}{12}R_{\mu\nu}\sigma^{\mu}_{;}\sigma^{\nu}_{;} - \tfrac{1}{24}R_{\mu\nu;\sigma}\sigma^{\mu}_{;}\sigma^{\nu}_{;}\sigma^{\sigma}_{;} + (\tfrac{1}{288}R_{\mu\nu}R_{\sigma\tau} + \tfrac{1}{360}R^{\alpha}{}_{\mu}{}^{\beta}{}_{\nu}R_{\alpha\sigma\beta\tau} + \tfrac{1}{80}R_{\mu\nu;\sigma\tau})\,\sigma^{\mu}_{;}\sigma^{\nu}_{;}\sigma^{\sigma}_{;}\sigma^{\tau}_{;} + \ldots \tag{255}$$

$$a_1(x,x') = (\tfrac{1}{6}-\xi)R - \tfrac{1}{2}(\tfrac{1}{6}-\xi)R_{;\mu}\sigma^{\mu}_{;} + [-\tfrac{1}{90}R_{\mu\alpha}R_{\nu}{}^{\alpha} + \tfrac{1}{180}R_{\mu\alpha\nu\beta}R^{\alpha\beta} + \tfrac{1}{180}R_{\alpha\beta\gamma\mu}R^{\alpha\beta\gamma}{}_{\nu} + \tfrac{1}{120}R_{\mu\nu;\alpha}{}^{\alpha} + (\tfrac{1}{40}-\tfrac{1}{6}\xi)R_{;\mu\nu}]\,\sigma^{\mu}_{;}\sigma^{\nu}_{;} + \ldots \tag{256}$$

which are derivable by the methods used to get (225) and (226). Then after differenting one must make use of the expansions

$$\sigma^{;\mu\nu}_{;} = g^{\mu\nu} - \tfrac{1}{3}R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta}\sigma^{\alpha}_{;}\sigma^{\beta}_{;} + \ldots \tag{257}$$

$$\sigma^{;\mu\nu'}_{;} = -g^{\mu\nu} - \tfrac{1}{6}R^{\mu}{}_{\alpha}{}^{\nu}{}_{\beta}\sigma^{\alpha}_{;}\sigma^{\beta}_{;} + \ldots \tag{258}$$

which are derived similarly.

The resulting expression will be given elsewhere, in connection with a study of the stress tensor near a black hole. It is both complicated and, because of the many ϵ^{μ}'s that appear in it, very frame dependent. However, the frame dependence is confined entirely to terms independent of the Riemann tensor, linear in the Riemann tensor or its first or second covariant derivatives, and quadratic in the Riemann tensor. These are precisely the terms arising from W_{div}. When all such terms are added together the result is $2g^{-1/2}\,\delta W_{\text{div}}/\delta g_{\mu\nu}$. When the geometry is given *a priori*, evaluation of the terms, component by component, is straightforward. Moreover, there is no particular difficulty in choosing a convenient point-splitting vector ϵ^{μ} for use both in these terms and in the mode-sum (12). Therefore the expression found for $2\delta W_{\text{div}}/\delta g_{\mu\nu}$ may be subtracted component by component from the mode-sum. The subtraction may be carried out in any convenient coordinate system. The result will be independent of the ϵ^{μ}'s and covariant.

Many of the ϵ^{μ}-dependent terms in $2\delta W_{\text{div}}/\delta g_{\mu\nu}$ are finite and hence analogous to the puzzling "non tensorial" terms found by Fulling and Parker [25]. It may very well prove possible to establish an identification between the terms found by the two methods and hence to resolve the difficulty that these terms have presented up to now. There is, however, a difference in the method of regularization adopted in the two cases that should be pointed out. Fulling and Parker use a momentum cutoff, whereas the point-splitting method adopted here corresponds (at least for timelike splittings) to an energy cutoff. If the two cutoffs are denoted by K and Λ respectively they will be related by $\Lambda^2 = K^2 + m^2$. This may account for the fact that the ϵ^{μ}-dependent terms found in the present method involve powers of m no higher than the fourth, whereas Fulling and Parker obtain also terms in m^6 and m^8.

6.20. Future outlook

With the computation of single-closed-loop processes more or less under control one may now ask what happens when multi-loop calculations are attempted. In the first place such calculations are exceedingly complicated. This is because when one goes after multi-loop corrections to the classical gravitational action one must, to be consistent, include gravitons, and the complexity of graviton-graviton vertices is such as to make strong men quail. But there is a more fundamental problem, which raises the most crucial issue in the quantum theory of gravity. This theory is not, by standard criteria, renormalizable. Amplitudes or matrix elements that involve more than one loop are more divergent than those involving single loops. Additional counter terms are required to dispose of the infinities, and these are of increasingly complicated types. Use of a counter term of the type (246), to dispose of the logarithmic divergences in the single-loop case, is already dangerous. Such a term in the classical action leads to field equations of the fourth differential order, and to all the ghost difficulties associated with such equations.

The painful facts are these: Physicists have come to believe, since Einstein, in the curvature of spacetime. Curvature leads to uncomfortable divergences already in the WKB approximation (single closed loops). Nevertheless, the WKB approximation must have *some* validity. The Casimir effect is an example, and it has been measured in the laboratory. At the same time we cannot believe that the WKB approximation is the end of the story. It is only an approximation.

It may be, as many have speculated, that quantum gravity contains its own cutoff – that it is actually finite. Heroic attempts have been made to prove this by summing infinite classes of multi-loop amplitudes. Unfortunately these attempts remain today both ambiguous and ultimately frame dependent (gauge dependent). At the present time only two possible procedures seem feasible, even if computational difficulties are ignored: (1) Accept the infinities as they come, order by order, but rigorously kill them all off by counter terms, save for the classical term $\int g^{1/2} R\, d^4x$ (with the modification embodied in eqs. (251) and (253) in the case $m = 0$); *or* (2) treat the renormalized graviton propagator obtained in the WKB approximation as the zeroth order propagator, and use *it* in the internal lines of all multi-loop graphs. Because it is less singular at small distance than the "bare" propagator it turns out that all higher-order amplitudes will diverge no worse than quartically in the energy cutoff Λ. Whether either method makes ultimate sense is for the future to determine.

A final comment is in order concerning the use of "in" and "out" regions. It should be clear by now that the divergences (at least in the WKB approximation) depend only locally on the spacetime geometry. They involve only the metric tensor and its first four derivatives. It does not matter where the "in" and "out" regions are located or how big they are; the form of the divergences is always the same. Hence it cannot matter if these regions simply disappear. In the total absence of Killing vectors it may, of course, be difficult to construct a meaningful or useful state-vector space. But the subtractions introduced here should serve to regularize any matrix element of $T^{\mu\nu}$ no matter how the states are defined.

I wish to express my appreciation to Professor Dennis Sciama for the hospitality extended to me by the Astrophysics Department of the University of Oxford where this paper was begun. I have profited by conversations with many people including especially Larry Ford, Stephen Fulling, Gary Gibbons, Stephen Hawking, Christopher Isham and William Unruh.

References

[1] J.D. Bekenstein, Phys. Rev. D7 (1973) 2333.
[2] N.N. Bogoluibov and D.V. Shirkov, Introduction to the Theory of Quantized Fields (Interscience Publishers, Inc., New York, 1959).
[3] D.G. Boulware, 1975, paper in course of preparation.
[4] T.H. Boyer, Phys. Rev. 174 (1968) 1764.
[5] T.H. Boyer, Phys. Rev. 185 (1969) 2039.
[6] C.G. Callan, S. Coleman and R. Jackiw, Ann. Phys. (N.Y.) 59 (1970) 42.

[7] D.M. Capper and M. Ramón-Medrano, Phys. Rev. D9 (1974) 1641.
[8] D.M. Capper, G. Leibbrandt and M. Ramón-Medrano, Phys. Rev. D8 (1974) 4320.
[9] D.M. Capper, M.J. Duff and L. Halpern, Phys. Rev. D10 (1974) 461.
[10] H.B.G. Casimir, Proc. Kon. Ned. Akad. Wetenschap. 51 (1948) 793. [See also T.H. Boyer, Ann. Phys. (N.Y.) 56 (1970) 474.]
[11] S.S. Chern, Hamburg. Abh. 20 (1955) 117.
[12] S.S. Chern, J. Soc. Indust. Appl. Math 10 (1962) 751.
[13] S. Christensen, 1974, private manuscript (to be published)
[14] D. Christodoulou, Phys. Rev. Lett. 25 (1970) 1596.
[15] D. Christodoulou and R. Ruffini, Phys. Rev. D4 (1971) 3552.
[16] T. Damour and R. Ruffini, Quantum Electrodynamical Effects in Kerr-Newman Geometries, Princeton preprint (1974).
[17] B. Davies, J. Math. Phys. 13 (1972) 1324.
[18] S. Deser and P. Van Nieuwenhuizen, Phys. Rev. D10 (1974) 401, 411.
[19] B.S. DeWitt, Dynamical Theory of Groups and Fields (Gordon and Breach, Inc., New York, 1965).
[20] B.S. DeWitt, Phys. Rev. 162 (1967) 1195, 1239.
[21] J.S. Dowker, J. Phys. (London) A3 (1970) 451.
[22] J.S. Dowker, Ann. Phys. (N.Y.) 62 (1971) 361.
[23] L. Ford, Ph.D. thesis, Princeton University (1974).
[24] S.A. Fulling, Phys. Rev. D7 (1973) 2850
[25] S.A. Fulling and L. Parker, 1974, Ann. Phys. (N.Y.), in press
[26] S.A. Fulling, L. Parker and B.L. Hu, Conformal Energy-Momentum Tensor in Curved Spacetime: Adiabatic Regularization and Renormalization, University of Wisconsin-Milwaukee preprint (1974).
[27] U.H. Gerlach, The Mechanism of Black Body Radiation from an Incipient Black Hole, Ohio State University preprint (1974).
[28] G.W. Gibbons, Vacuum Polarization and the Spontaneous Loss of Charge by Black Holes, University of Cambridge preprint (1974), to be published.
[29] V.L. Ginzburg, D.A. Kirzhnits and A.A. Lyubushin, Zh.E.T.F. 60 (1971) 451 [English transl. Sov. Phys. -J.E.T.P. 33 (1971) 242].
[30] S.W. Hawking, Phys. Rev. Lett. 26 (1971) 1344. [See also S.W. Hawking and G.F.R. Ellis, The Large Scale Structure of Space-time (Cambridge University Press, 1973).]
[31] S.W. Hawking, Nature 248 (1974) 30.
[32] S.W. Hawking, Particle Creation by Black Holes, University of Cambridge Preprint (1974), to appear in Commun. Math. Phys.
[33] G. 't Hooft and M. Veltmann, One Loop Divergences in the Theory of Gravitation, CERN Preprint TH 1723 (1974), to be published in Annales de l'Institut Henri Poincaré.
[34] C.W. Misner, 1972, private communication to W. Press and S. Teukolsky. [See Press and Teukolsky, Nature 238 (1972) 211, and J. Bekenstein, Phys. Rev. D7 (1973) 949.]
[35] G.T. Moore, J. Math. Phys. 11 (1970) 2679.
[36] C. Morette, Phys. Rev. 81 (1951) 848.
[37] M. Nouri-Moghadan and J.G. Taylor, The Ambiguities of Dimensional Regularization Scheme, King's College (London) preprint (1974).
[38] L. Parker, Phys. Rev. 183 (1969) 1057.
[39] L. Parker and S.A. Fulling, Phys. Rev. D9 (1974) 341.
[40] R.E. Peierls, Proc. Roy. Soc. (London) A214 (1952) 143.
[41] R. Penrose, Nuovo Cimento 1, special number (1969) 252.
[42] W. Rindler, Am. J. Phys. 34 (1966) 1174.
[43] J.S. Schwinger, Phys. Rev. 82 (1951) 664.
[44] J.S. Schwinger, Phys. Rev. 93 (1954) 615.
[45] A.A. Starobinsky, Zh.E.T.F. 64 (1973) 48 [English transl. Sov. Phys. -J.E.T.P. 37 (1973) 28].
[46] A.A. Starobinsky and S.M. Churilov, Zh.E.T.F. 65 (1973) 3 [English transl. Sov. Phys. -J.E.T.P. 38 (1973) 1].
[47] W.G. Unruh, Second Quantization in the Kerr Metric, University of California preprint (1974), to be published.
[48] R. Utiyama and B.S. DeWitt, J. Math. Phys. 3 (1962) 608.
[49] J.H. Van Vleck, Proc. Nat. Acad. Sci. U.S.A. 14 (1928) 178.
[49] Ya.B. Zel'dovich and A.A. Starobinsky, Zh.E.T.F. 61 (1971) 2161; English transl. Sov. Phys. -J.E.T.P. 34 (1972) 1159].
[50] Ya.B. Zel'dovich, Zh.E.T.F. Pis'ma v. Red. 14 (1971) 270 [English transl. J.E.T.P. Letters 14 (1971) 180].
[51] Ya.B. Zel'dovich, Zh.E.T.F. 62 (1972) 2076 [English transl. Sov. Phys. -J.E.T.P. 35 (1972) 1085].
[52] Ya.B. Zel'dovich and A.A. Starobinsky, Zh.E.T.F. 61 (1971) 2161 [English transl. Sov. Phys.-J.E.T.P. 34 (1972) 1159].
[53] Ya.V. Zel'dovich, V.N. Lukash and A.A. Starobinsky, Particle Creation in Gravitational Fields and Its Influence on Cosmological Expansion, preprint No. 23 (1974), Institute of Applied Mathematics, Academy of Sciences, U.S.S.R., Moscow (in Russian).
[54] B. Zumino, Supersymmetry and the Vacuum, CERN preprint No. TH-1942 (1974).