

The Back Reaction Effect in Particle Creation in Curved Spacetime*

Robert M. Wald**

Enrico Fermi Institute, University of Chicago, Chicago, IL 60637, USA

Abstract. The problem of determining the changes in the gravitational field caused by particle creation is investigated in the context of the semiclassical approximation, where the gravitational field (i.e., spacetime geometry) is treated classically and an effective stress energy is assigned to the created particles which acts as a source of the gravitational field. An axiomatic approach is taken. We list five conditions which the renormalized stress-energy operator $T_{\mu\nu}$ should satisfy in order to give a reasonable semiclassical theory. It is proven that these conditions uniquely determine $T_{\mu\nu}$, i.e. there is at most one renormalized stress-energy operator which satisfies all the conditions. We investigate existence by examining an explicit “point-splitting” type prescription for renormalizing $T_{\mu\nu}$. Modulo some standard assumptions which are made in defining the prescription for $T_{\mu\nu}$, it is shown that this prescription satisfies at least four of the five axioms.

I. Introduction

In the past several years, a considerable amount of progress has been made in our understanding of quantum processes occurring in a strong gravitational field. A satisfactory quantum theory of the gravitational field itself still does not exist [1]. However, the framework of a semiclassical theory describing other quantum fields present in a strong gravitational field does exist and has been used to investigate particle creation effects. In this theory the gravitational field is described in an entirely classical manner as curvature in the geometry of spacetime, in accordance with the notions of general relativity. The fields (e.g., a scalar, Dirac, or Maxwell field) which are present in spacetime are described in accordance with the principles of quantum field theory. It is not believed that this theoretical framework can provide an exact description of nature, since it cannot be entirely consistent to have

* Supported in part by the National Science Foundation under grant MPS 74-17456-A01 and by the Sloan Foundation

** Sloan Foundation fellow

quantum fields (described in probabilistic terms) interact with a classical gravitational field (with definite, determined values). Rather, this semiclassical theory is viewed as an approximation to the true—as yet unknown—quantum theory of gravitation interacting with other fields. Such a semiclassical framework is analogous to the situation in atomic physics where, for the description of a wide range of phenomena, it is a good approximation to describe the electromagnetic field in an entirely classical manner while treating the electrons quantum mechanically. On dimensional grounds it is generally believed that quantum effects of gravity should be important at least when the spacetime curvature becomes comparable to the Planck length $(\hbar G/c^3)^{1/2} \approx 10^{-33}$ cm. However, for less extreme spacetime curvature, one hopes that the semiclassical approximation will be valid at least in many situations.

If the gravitational field has suitable asymptotic behavior in the past and future, a description of the quantum fields in terms of particles will be possible in these asymptotic regimes. One may then ask about particle creation: If the field is initially in the vacuum state, how many particles will be present at late times? More generally, what is the S -matrix? It turns out that a few simple assumptions within the semiclassical framework described above uniquely lead to an expression for the S -matrix in a manner which is very nearly free of any mathematical difficulties [2]. Thus, one can make well defined, unambiguous predictions concerning particle creation in a strong gravitational field.

The most remarkable application of these ideas is, of course, Hawking's discovery [3] that particle creation near a Schwarzschild black hole will result in a steady rate of emission of particles with an *exactly* thermal spectrum [2, 4]. This result is particularly striking in view of the analogies that had previously been discovered between black hole physics and thermodynamics [5, 6]. In the absence of any experimental or observational confirmation of the predictions of the semiclassical theory, it is the beauty of Hawking's result as well as the simplicity, naturalness, and good mathematical behavior of the theory which gives one confidence that this approach is on the right track.

In the particle creation calculations referred to above, the spacetime geometry (i.e., gravitational field) is taken to be that appropriate to some classical physical situation, e.g., the gravitational collapse of a body to form a black hole. The particle creation is then calculated in this fixed spacetime geometry. However, on physical grounds it is clear that the quantum particle creation must have some "back reaction" effect on the spacetime geometry. In particular, for the case of a black hole, the particle creation calculations show a flux of energy coming from the black hole. By conservation of energy one would expect this energy flux to be balanced by a decrease in the mass of the black hole (i.e., a decrease in the energy of the gravitational field). The determination of the nature and magnitude of the "back reaction" effect is of great interest and importance in its own right, particularly in the cosmological context where the "back reaction" of particle creation may have an important effect on the dynamics of the universe. It is also needed to check the validity of the particle creation calculations, since if the effect of the "back reaction" is large, it must be taken into account in these calculations.

In what framework can one analyze this "back reaction" effect? It is conceivable that one will need a complete quantum theory of gravitation in order to describe it,

since to describe it in the semiclassical framework involves having a quantum field act as a source for a classical field. If, for example, the quantum field source has a probability of $\frac{1}{2}$ of being “very small” and a probability of $\frac{1}{2}$ of being “very large”, it would not seem reasonable to try to describe the gravitational field to which it gives rise as a “medium sized” classical field. Thus, in particular, Hawking has expressed the view that “back reaction” can be described only in the context of quantum gravity. However, there is clearly *some* domain of validity to associating a classical gravitational field to a quantum source. After all, ordinary matter is, of course, in reality of a quantum nature but it certainly makes sense to assign it a classical gravitational field. More generally, if the gravitational field is not so strong that the effects of quantum gravity should be of direct importance, it seems reasonable to expect the approximation of a classical gravitational field to be valid whenever the expected quantum fluctuations in the source are negligible compared with the expected value of the source itself. It is, of course, not obvious whether this domain of validity extends to cases where particle creation effects are important. However, the example of particle creation near a black hole suggests that this may be the case since at least at large distances from the black hole the created particles are thermally distributed and hence should satisfy the above criterion. Further indication that the semiclassical approximation may have a wide range of validity in treating problems of an essentially quantum nature comes from the example of atomic physics: In atoms with many electrons it is a very good approximation to treat the electric field classically even though the electrons which give rise to this field must be treated quantum mechanically; in general, the “radiative corrections” to this approximation are negligible. For the remainder of this paper, we shall assume that the semiclassical approach to the “back reaction” effect has a nontrivial range of validity.

In classical general relativity, the source of the gravitational field is the stress-energy tensor $T_{\mu\nu}$ of the fields present in spacetime. The gravitation field (described by the spacetime metric) is related to $T_{\mu\nu}$ via Einstein’s equation,

$$G_{\mu\nu} = 8\pi T_{\mu\nu} \quad (1.1)$$

where $G_{\mu\nu}$ is the Einstein tensor. In quantum theory, observables are described as operators acting on the Hilbert space of states of the system. Hence, in quantum theory, the stress-energy tensor should become an operator. A natural procedure for treating the “back reaction” effect in the semiclassical approximation then suggests itself: we require that the classical Einstein tensor be set equal to the expected value of the stress energy tensor in the given quantum state,

$$G_{\mu\nu} = 8\pi \langle T_{\mu\nu} \rangle. \quad (1.2)$$

More precisely, the structure of the theory is as follows: For each (suitably well behaved) classical spacetime geometry, there should exist a stress-energy operator for each field of interest. A spacetime together with a quantum state of the field which satisfies Equation (1.2) is considered to be a solution of this semiclassical Einstein theory. This solution is to be taken seriously if the characteristic radii of curvature of the spacetime are much greater than the Planck length and if the expected fluctuations in $T_{\mu\nu}$ in this state are negligible compared with $\langle T_{\mu\nu} \rangle$. In the

limit where the field can be described classically (i.e., a large number of appropriately distributed particles and negligible particle creation), the theory should reduce to classical general relativity.

It is natural to postulate that the stress-energy operator is given in terms of the quantum field operator by the same formula by which the classical stress-energy tensor is related to the classical field. It is here, however, that a serious problem arises. The quantum field operator does not exist as an operator defined at each point of spacetime; only “smeared” fields make sense mathematically, i.e., the field is an operator valued distribution on spacetime. When one performs operations which are linear in the field, this distinction is basically just a technical point; for example, it leads to no difficulties in the derivation of the S -matrix [2]. However, nonlinear operations on distributions, such as taking products, have no obvious mathematical meaning. Since the stress-energy is quadratic in the field, the formula for the stress-energy operator involves a product of distributions and hence must be viewed as only a formal expression. Therefore, it is not surprising that when one naively attempts to calculate expectation values of the stress-energy operator, one gets infinite answers. Thus, some sort of renormalization prescription must be given.

In flat spacetime there is a completely satisfactory solution to this renormalization problem: normal ordering. One can view this prescription as renormalizing the energy of the vacuum state to zero. However, in curved spacetime, when particle creation takes place there is no invariant vacuum state and thus there is no natural analogue of normal ordering. Furthermore, even if no particle creation occurs (e.g., in a stationary spacetime) it is not at all clear that normal ordering is correct, since vacuum polarization effects may cause the stress-energy of the vacuum state to be nonzero. Thus, the problem of renormalizing the stress-energy tensor in curved spacetime is a nontrivial one, as has been further demonstrated by the considerable amount of effort that has gone into attempts to solve it.

A number of proposals for renormalizing $T_{\mu\nu}$ are discussed by DeWitt [7]. More recently, dimensional regularization [8] and zeta-function [9] techniques have been developed and further work has been done on the “point-splitting” method [10, 11]. However, all the prescriptions that have been given thus far are either applicable only to a very restricted class of spacetimes (e.g., the “adiabatic regularization” scheme of Parker and Fulling [12]) or have ambiguities (e.g., the direction dependent terms in the “point-splitting” approach), or, at the very least, have features which are ad hoc. The extent to which the different procedures agree or disagree has not been fully investigated.

In this paper, we shall take a different approach to the stress-energy renormalization problem. Rather than develop a particular renormalization scheme, we shall list conditions that the renormalized stress-energy operator should satisfy in order to give a reasonable semiclassical theory of “back reaction” within the framework described above. These five axioms are stated in Section II. In the absence of any likelihood of experimental or observational verifications of “back reaction” predictions, conditions such as these are the only available criteria for deciding whether any given renormalization prescription is likely to be the correct one.

The main result of the paper is the following: the axioms stated in Section II uniquely determine a renormalized stress-energy operator. In other words, there can be at most one stress-energy operator which satisfies all the axioms. This result is proved in Section III. However, the proof does not establish existence of such an operator.

The question of existence is investigated in Section IV by considering an explicit “point-splitting” type prescription for renormalizing $T_{\mu\nu}$. The discussion here is based on several technical assumptions for which complete proofs have not been given, so the results have more the flavor of plausibility arguments than theorems. We show that this prescription yields a stress-energy operator which satisfies at least the first four axioms. However, the last axiom is very difficult to check since it requires a detailed knowledge of how certain quantities change under variations of the spacetime geometry. Thus, while existence of the renormalized stress-energy operator is not shown, we do establish (modulo the technical assumptions) that the first four axioms are self-consistent. Some concluding remarks are made in Section V.

II. Axioms for $T_{\mu\nu}$

In this section we shall state and discuss the precise mathematical conditions which the renormalized stress-energy operator should satisfy. Our first task is to make more precise the class of spacetimes on which $T_{\mu\nu}$ should be defined and what type of mathematical object $T_{\mu\nu}$ is.

We shall take the class of spacetimes to consist of C^∞ spacetimes which are sufficiently well behaved asymptotically in the past and future to admit asymptotic notions of positive and negative frequency solutions¹. This permits one to characterize the states of the quantum field in terms of “in” and “out” Fock spaces. We shall further require that the spacetime be such that the condition, $\text{tr}(E^\dagger E) < \infty$, described in Reference [2], is satisfied, so that the S -matrix relating the “in” and “out” states really does exist. We do not mean to suggest that this is the *only* class of spacetimes on which $T_{\mu\nu}$ should be definable; however, $T_{\mu\nu}$ should be defined *at least* on this class.

What type of mathematical object should $T_{\mu\nu}$ be? It is clearly demanding too much to require that for each point p in spacetime, $T_{\mu\nu}(p)$ be an operator mapping the Hilbert space \mathcal{F} of quantum states into itself; neither the field operator nor the normal ordered stress-energy operator in flat spacetime exist in this sense. For the purposes of this paper, it will be important that $T_{\mu\nu}$ be defined at each point of spacetime but it will only be necessary that the matrix elements of $T_{\mu\nu}$ be well defined. Thus, it will suffice to require that the renormalized $T_{\mu\nu}$ be a bilinear map

$$T_{\mu\nu} : \bar{\mathcal{F}} \times \mathcal{F} \rightarrow \mathcal{T}(2, 0)$$

defined on a dense domain of vectors in $\bar{\mathcal{F}} \times \mathcal{F}$. (Here $\bar{\mathcal{F}}$ denotes the dual of the Hilbert space of states \mathcal{F} and $\mathcal{T}(2, 0)$ denotes the vector space of 2-covariant index symmetric tensor fields on spacetime.) This is a reasonable requirement since both

¹ This includes many nonasymptotically flat spacetimes where the Feynman propagator constructions of e.g., Rumpf [13], Hartle and Hawking [14], or Candelas and Raine [15] are available

the normal ordered stress-energy operator and the field operator do exist in this sense. For $\Phi, \Psi \in \mathcal{F}$ we will denote $T_{\mu\nu}(\Phi, \Psi)$ as $\langle \Phi | T_{\mu\nu} | \Psi \rangle$ and refer to this quantity as a matrix element. We further require that all diagonal matrix elements $\langle \Psi | T_{\mu\nu} | \Psi \rangle$ (i.e., all expectation values) be real and that the expectation values in, say, the “in” or “out” vacuum states be smooth (C^∞) tensor fields.

As mentioned in Section I above, the formal expression for $T_{\mu\nu}$ is ill-defined mathematically and naive calculations yield infinite expectation values. However, as we shall now illustrate for the case of a scalar field, the formal expression does yield unambiguous, finite results when used to calculate the matrix element between orthogonal states, i.e., $\langle \Phi | T_{\mu\nu} | \Psi \rangle$ for $\langle \Phi | \Psi \rangle = 0$. Classically, the stress-energy tensor of a massless Klein-Gordon field ϕ is given by,

$$T_{\mu\nu} = \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} g_{\mu\nu} \nabla_\sigma \phi \nabla^\sigma \phi \quad (2.1)$$

where ∇_μ denotes the covariant derivative. The formal expression for the quantum stress-energy operator is,

$$T_{\mu\nu} = \frac{1}{2} (\nabla_\mu \phi \nabla_\nu \phi + \nabla_\nu \phi \nabla_\mu \phi - g_{\mu\nu} \nabla_\sigma \phi \nabla^\sigma \phi) \quad (2.2)$$

where ϕ is the quantum field operator. We can express ϕ in terms of the “in” annihilation and creation operators a_i, a_i^\dagger by,

$$\phi = \sum_i (G_i a_i + \bar{G}_i a_i^\dagger) \quad (2.3)$$

where each G_i is a classical solution of the (curved space) scalar wave equation which is asymptotically positive frequency in the past (see, e.g. Ref. [2]). Substituting this into Equation (2.2), we can formally write $T_{\mu\nu}$ as a “mode sum”,

$$\begin{aligned} T_{\mu\nu} = & \frac{1}{2} \sum_{i,j} \{ [2\nabla_\mu G_i \nabla_\nu G_j - g_{\mu\nu} \nabla_\sigma G_i \nabla^\sigma G_j] a_i a_j \\ & + [\nabla_\mu \bar{G}_i \nabla_\nu G_j + \nabla_\nu \bar{G}_i \nabla_\mu G_j - g_{\mu\nu} \nabla_\sigma \bar{G}_i \nabla^\sigma G_j] a_i^\dagger a_j \\ & + [\nabla_\mu G_i \nabla_\nu \bar{G}_j + \nabla_\nu G_i \nabla_\mu \bar{G}_j - g_{\mu\nu} \nabla_\sigma G_i \nabla^\sigma \bar{G}_j] a_i a_j^\dagger \\ & + [2\nabla_\mu \bar{G}_i \nabla_\nu \bar{G}_j - g_{\mu\nu} \nabla_\sigma \bar{G}_i \nabla^\sigma \bar{G}_j] a_i^\dagger a_j^\dagger \} . \end{aligned} \quad (2.4)$$

Suppose we naively attempt to use this expression to calculate an expectation value, e.g. for the “in” vacuum state $|0_{\text{in}}\rangle$. In this case, only the $a_i a_i^\dagger$ terms contribute and we find,

$$\begin{aligned} \langle 0_{\text{in}} | T_{\mu\nu} | 0_{\text{in}} \rangle = & \frac{1}{2} \sum_i [\nabla_\mu G_i \nabla_\nu \bar{G}_i + \nabla_\nu G_i \nabla_\mu \bar{G}_i \\ & - g_{\mu\nu} \nabla_\sigma G_i \nabla^\sigma \bar{G}_i] . \end{aligned} \quad (2.5)$$

As is well known, this sum does not converge. However, if we use the formal expression to calculate a matrix element between orthogonal states, the $a_i a_j^\dagger$ type terms do not contribute infinite sums and we get unambiguous, finite results. For example, the matrix element of $T_{\mu\nu}$ between $|0_{\text{in}}\rangle$ and a 2-particle in-state consisting of particles in states k and l is simply,

$$\begin{aligned} \langle kl | T_{\mu\nu} | 0_{\text{in}} \rangle = & \frac{1}{2} [\nabla_\mu \bar{G}_k \nabla_\nu \bar{G}_l + \nabla_\nu \bar{G}_k \nabla_\mu \bar{G}_l \\ & - g_{\mu\nu} \nabla_\sigma \bar{G}_k \nabla^\sigma \bar{G}_l] . \end{aligned} \quad (2.6)$$

In a similar manner, it is not difficult to see that one will always get finite results for the matrix element between any two orthogonal states of the form of a finite linear combination of a finite number of “in” creation operators (for smooth 1-particle states) acting on $|0_{\text{in}}\rangle$. (States of this form are dense in \mathcal{F} .) The answer that one obtains does not depend on the choice of basis G_i nor on the expansion of ϕ in terms of “in” annihilation operators (i.e., one gets the same answer if one, say, expands ϕ in terms of the “out” operators).

If we require that the renormalized $T_{\mu\nu}$ must agree with the above mode sum expression for the matrix element between orthogonal states (see axiom 1 below), then the problem of finding all expectation values of the renormalized $T_{\mu\nu}$ reduces to the problem of finding a single expectation value, since the difference between two expectation values can be expressed as a matrix element between two orthogonal states. Thus, the problem of renormalizing $T_{\mu\nu}$ is essentially a “ c -number” problem².

We shall now list our five axioms for the renormalized $T_{\mu\nu}$. A discussion of each of these axioms is given below. In the following, $T_{\mu\nu}$ denotes the renormalized operator, while $\langle T_{\mu\nu} \rangle$ denotes its expectation value in some state.

1. The formal expression for $T_{\mu\nu}$ discussed above is valid for calculating the matrix element $\langle \Phi | T_{\mu\nu} | \Psi \rangle$ between any two orthogonal states, $\langle \Phi | \Psi \rangle = 0$.

2. Normal ordering is valid in Minkowski spacetime (i.e., flat spacetime with the standard \mathbb{R}^4 topology).

3. For any state, the expectation value $\langle T_{\mu\nu} \rangle$ is conserved, $\nabla^\mu \langle T_{\mu\nu} \rangle = 0$.

4. Causality: (a) For a fixed “in” state³, $\langle T_{\mu\nu} \rangle$ at a point p in spacetime depends only on the spacetime geometry to the causal past of p , i.e. changes in the metric to the future of p or at spacelike separations from p do not affect $\langle T_{\mu\nu} \rangle$ at p .

(b) For a fixed “out” state, $\langle T_{\mu\nu} \rangle$ at p depends only on the spacetime geometry to the causal future of p .

5. For fixed “in” or “out” state, $\langle T_{\mu\nu} \rangle$ varies continuously with the spacetime metric in a sufficiently strong manner to guarantee that the dynamics of the semiclassical theory, Equation (1.2), be of the same nature as the dynamics of classical general relativity Equation (1.1). The precise condition adopted here is the following (see Proposition 7.4.7 and Section 7.7 of Hawking and Ellis [17]): Consider a sequence $\{(g_{\mu\nu})_n\}$ of C^∞ spacetime metrics which agree outside a fixed compact region and are such that the components of $(g_{\mu\nu})_n$ and the derivatives of these components up to fourth order (in a fixed chart) converge uniformly to a C^∞ metric $g_{\mu\nu}$ and its derivatives up to fourth order, respectively. Then we require that for fixed “in” or “out” state $\{\langle T_{\mu\nu} \rangle_n\}$ and its derivatives up to third order converge (pointwise) to $\langle T_{\mu\nu} \rangle$ and its derivatives up to third order, respectively. (As discussed

² Though not always stated in this manner, the above remarks are well known (see, e.g. Wightman [16])

³ It is possible to consider changes in the spacetime geometry which leave the causal past of p unchanged but which behave badly enough in the asymptotic past (outside the lightcone of p) to destroy the notion of an “in” Fock space or, at least, obscure the identification of the “in” Fock space of the new spacetime with that of the old one. In such a case, the phrase “for fixed ‘in’ state” has no meaning. Thus, a more precise statement of condition 4(a) is: for variations of the spacetime geometry which vanish in the causal past of p and are sufficiently well behaved in the asymptotic past of the spacetime to permit identification of the new and old “in” states, $\langle T_{\mu\nu} \rangle$ at p is unchanged. In particular, for an asymptotically flat spacetime, $\langle T_{\mu\nu} \rangle$ at p must not change when the spacetime metric is varied in an arbitrary manner to the future of any asymptotically flat spacelike hypersurface passing through p

below, this condition should perhaps be weakened to requiring only that $\{\langle T_{\mu\nu}\rangle_n\}$ and its derivatives up to second order converge.)

We should note the absence in the above list of any local positive energy condition on $\langle T_{\mu\nu}\rangle$. For most fields of interest, a local positive energy condition is satisfied by the classical stress energy. This condition is a major hypothesis of the singularity theorems of classical general relativity (see e.g., [17]). However, it is not a reasonable condition to impose on expectation values of the renormalized quantum stress-energy operator, since it is not even satisfied by the normal ordered stress energy in flat spacetime⁴. Namely, if, in flat spacetime, we consider a state consisting of a superposition of the vacuum state and a number ε times a two particle state, then for sufficiently small ε the expected normal ordered stress-energy will be dominated by the cross-term involving the matrix element of $T_{\mu\nu}$ between the two particle state and the vacuum. However, the local energy density of this term need not be positive [see Eq. (2.6) above], as is already clear from the fact that the sign of this term changes if we reverse the sign of ε . (On the other hand the total energy of this cross term vanishes and the expected total energy of the state is positive.) Since the state vector describing particle creation in a weak gravitational field is precisely of the form of the vacuum state plus a small amplitude times a two particle state (see e.g. [2, 19]) it is clear that a local positive energy condition should not be satisfied when particle creation occurs⁵.

We have already discussed above the meaning of the first axiom. Since the formal expression certainly should bear some relation to the renormalized $T_{\mu\nu}$ and since the matrix element between orthogonal states of the formal expression gives finite results in a straightforward, unambiguous manner, it seems reasonable to expect it to be the “right answer”.

The second axiom also requires little further discussion. Normal ordering gives a completely sensible, physically reasonable, and generally accepted prescription for $T_{\mu\nu}$ in Minkowski spacetime. Furthermore, assuming axiom 1, it is equivalent to postulating that the stress-energy of the vacuum state in Minkowski spacetime is zero. Note that axiom 2 applies only to globally Minkowski spacetime; we do not rule out Casimir-type effects [7] for topologies other than \mathbb{R}^4 , and indeed we shall see at the end of Section IV that such effects must be present.

Axiom 3 is, of course, satisfied by the classical stress-energy tensor; it expresses the local conservation of non-gravitational energy. If it were not automatically satisfied for every spacetime in classical theory, then the Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$ would impose this condition on solutions of Einstein’s Equation (1.1), which presumably would lead to stringent consistency relations; “very few” solutions would exist. [Thus, for this reason, the classical equation $R_{\mu\nu} = 4\pi T_{\mu\nu}$ i.e., $G_{\mu\nu} = 4\pi(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu})$ is not acceptable.] Similarly, it seems difficult to imagine that using Equation (1.2) one could obtain a reasonable semiclassical theory (which reduces to general relativity in the appropriate limit) unless axiom 3 is satisfied.

Axiom 4 basically says that if one creates particles to the future of a point p (or at spacelike separations from p) the effects of this particle creation should not register

⁴ This fact has also been noted by Davies and Fulling [18]

⁵ Hawking [17] has shown what is essentially the converse of this statement: The dominant energy condition together with conservation (axiom 3) imply that if the stress energy is zero initially, it must remain zero, i.e. no particle creation can occur

at p . Since, in the semiclassical framework one still assumes the spacetime notions of general relativity, it is reasonable to demand that the causal properties of spacetime be compatible with relativity. One frequently describes phenomena in quantum field theory in acausal language, e.g. “particles propagating backward in time”. However, all physical effects that have thus far been investigated are causal. In particular, the matrix elements of the field operator itself (which are perfectly well defined and require no renormalization) satisfy axiom 4.

Axiom 5 is a rather technical condition. In order to motivate it, we shall make a few heuristic remarks concerning classical dynamical evolution in general relativity. It is well known that in a fixed chart one can view the vacuum Einstein equation

$$G_{\mu\nu} = 0 \tag{2.7}$$

as a second order hyperbolic system of equations for the components of the metric. Consider, now, the Einstein Equation (1.1) with the classical stress-energy tensor $T_{\mu\nu}$ of some field acting as a source. If we fix initial data for the field on some Cauchy surface, we can view the field and hence this classical $T_{\mu\nu}$ as a (nonlocal) functional of the spacetime metric. Suppose, however, that this nonlocal functional were to include a “local curvature piece” i.e., a term whose value at point p depends on derivatives of the metric components higher than first order at p , e.g., a fourth order term like $\nabla_\mu \nabla_\nu R$, where R is the scalar curvature. In that case, the character of the dynamical evolution of Einstein’s equation with source would be entirely different than that of Einstein’s equation in vacuum. In the example just quoted, the evolution would have the character of a fourth order system

$$\nabla_\mu \nabla_\nu R = -G_{\mu\nu} + (\text{nonlocal part of } T_{\mu\nu})$$

rather than that of the second order system $G_{\mu\nu} = 0$. Note that this is true even if the “local curvature” term is “small”, i.e., if one adds to a differential equation a term containing higher order derivatives than originally appeared in the equation, the character of the equation becomes entirely different, even if the new term enters multiplied by a small coefficient⁶.

However, the phenomenon described above does not occur in classical general relativity; the dynamics of Einstein’s equation with a source has the same character as Einstein equation in vacuum. The reason why this is so is that the classical stress-energy tensor satisfies a classical version of axiom 5. Axiom 5 is a precise mathematical condition which expresses the intuitive notion that the stress energy contains no “local curvature term”. A “local curvature term” depending on second order or higher derivatives of the metric components would not vary continuously with the metric in the manner required by axiom 5.

The proof of Proposition 7.4.7 of Hawking and Ellis [17] shows that the classical field and its derivatives up to fourth order vary continuously (in a weak sense) under the variations of the metric considered in axiom 5. The classical version of axiom 5 then follows if we assume that the classical stress energy is a polynomial

⁶ A simple illustration of this point is provided by the ordinary differential equation $\dot{f} + f = 0$. The solution of this equation is, of course, Ce^{-t} , i.e., an exponentially decreasing function. However, if one adds the “small” term $-\epsilon \dot{f}$ to this equation, then for $0 < \epsilon \ll 1$ one picks up a solution which grows exponentially as $e^{t/\epsilon}$. Thus, this “small” term completely changes the stability properties of the original solutions and the effect of this term actually becomes more and more drastic as $\epsilon \rightarrow 0$.

in the field, its first derivative, and the spacetime metric, and if we ignore distinctions between weak convergence and pointwise convergence. (The proper convergence properties could probably be proven by the methods described on page 223 of Garabedian [20].) The arguments outlined in Section 7.7 of Hawking and Ellis [17] show that if axiom 5 is satisfied, the Cauchy problem for the Einstein equation with source is well posed and the dynamical evolution is of the same character as for the vacuum Einstein equation.

The motivation for imposing axiom 5 on the quantum stress-energy tensor should now be clear. As described in the introduction, we want the semiclassical theory, Equation (1.2), to reduce to general relativity in the classical limit. However, if the quantum stress energy does not satisfy a condition like axiom 5, it is difficult to imagine how this can happen. As previously illustrated in Footnote 6, even in cases where the quantum corrections are “small”, violation of axiom 5 would imply that these correction terms could drastically alter the character of Einstein’s equation. In particular, the stability properties of solutions would surely be quite different from that of classical general relativity. Thus, it appears that axiom 5 is necessary for the semiclassical theory to look anything like general relativity for small but finite \hbar .

In making the above statements, we should be a bit more careful about the required orders of differentiability. As mentioned above, the classical stress-energy tensor of most fields satisfies the classical version of axiom 5. On this ground it seems most reasonable to impose on the quantum stress energy the strong version of axiom 5 given above. However, if axiom 5 were weakened slightly by requiring only $\{\langle T_{\mu\nu}\rangle_n\}$ and its derivatives up to second order (rather than third order) to converge, this would, roughly speaking, still only permit “second derivative type terms in the metric” in $\langle T_{\mu\nu}\rangle$. It is quite plausible that the semiclassical Equation (1.2) could still have the basic character of a second order hyperbolic system which reduces to Einstein’s equation in the classical limit. (On the other hand, further weakening of axiom 5 in this manner would not be acceptable.) Thus, it is possible that only this weakened version of axiom 5 should be imposed. In particular for the case of the conformally invariant scalar field, even the classical stress energy satisfies only the weakened version of axiom 5, since it contains second-order derivatives of the field as well as Ricci tensor terms. Thus, at least for the case of this field, only the weakened version of axiom 5 should be imposed on its quantum stress energy.

As we shall show in Section III below, the strong version of axiom 5 together with the other axioms uniquely determines a stress-energy operator, i.e., not more than one can exist which satisfies the above axioms. If we impose the weaker version of axiom 5, the corresponding result is that the stress-energy operator is uniquely determined up to a multiple of the classical Einstein tensor $G_{\mu\nu}$ times the identity operator. (This remaining ambiguity in $T_{\mu\nu}$ can then be eliminated by imposing some additional scaling assumptions (i.e., by dimensional arguments) as we shall see in more detail in Section III.) We should emphasize that axiom 5 is used in the uniqueness proof only to exclude the addition of local curvature terms to the stress-energy tensor. Thus, the uniqueness proof will remain valid under any restatement of axiom 5 which preserves the intuitive meaning described above.

The “conformal anomaly” discussed recently by a number of authors [21, 8, 9] provides a good illustration of the implications of axiom 5. The term “conformal

anomaly” refers to the claim that the trace of the quantum stress-energy tensor of a conformally invariant field may be nonzero, even though the trace of the classical stress energy vanishes identically. In two-dimensional model spacetimes studied by Davies et al. [22] and others, one can have a conformally flat spacetime—all two-dimensional spacetimes are conformally flat—which is flat outside a compact region and yet has the property that particles of a conformally invariant field are created. This implies that the quantum stress-energy tensor cannot be conformally invariant, even though the classical stress-energy tensor is conformally invariant. Since the tracelessness of the classical stress energy can be viewed as a consequence of its conformal invariance, it is not surprising that the quantum stress energy should lose its tracelessness as well, as Davies et al. [22] found. However, it is not at all clear that similar behavior should occur in four-dimensional spacetimes. It is well known that two-dimensional manifolds have anomalous conformal properties, and the above phenomenon may be merely a reflection of this fact. In particular, it is not difficult to show that in four dimensions no particle production can occur for a conformally invariant field in a conformally flat spacetime which is flat outside a compact region. Thus, the type of particle creation effect which occurs in two dimensions and strongly suggests a conformal anomaly in the trace of the stress energy does not occur in four dimensions.

The above axioms do not say anything directly about the existence of non-existence of conformal anomalies. However, the form of the conformal anomaly in the trace of the stress-energy tensor of a conformally invariant scalar field proposed by a number of authors [21, 8, 9] violates axiom 5, since, according to these results, it is a purely local curvature term containing derivatives of the metric up to fourth order⁷. As discussed above, such a violation of axiom 5 would imply that the semiclassical Equation (1.2) is a fourth order (or worse) system, and it is very difficult to see how general relativity could arise as its classical limit. In particular, it is not at all obvious that this theory will even contain Newtonian gravitation as a limiting case, since as discussed above the stability properties of the solutions may have an entirely different character.

III. Uniqueness

We shall now show that the five axioms given in the previous section uniquely determine a renormalized stress-energy operator. Let $T_{\mu\nu}$ and $\tilde{T}_{\mu\nu}$ be two stress-energy operators defined on the general class of spacetimes discussed at the beginning of Section II and satisfying axioms 1–5. We set

$$U_{\mu\nu} = T_{\mu\nu} - \tilde{T}_{\mu\nu}. \quad (3.1)$$

Our aim is to show $U_{\mu\nu} = 0$.

By axiom 1, the matrix element between any two orthogonal states of $T_{\mu\nu}$ is equal to that of $\tilde{T}_{\mu\nu}$. This means that the matrix element of $U_{\mu\nu}$ between any two orthogonal states must vanish. But this implies that $U_{\mu\nu}$ is a multiple of the identity operator, i.e., it is of the form,

$$U_{\mu\nu} = u_{\mu\nu} I \quad (3.2)$$

⁷ Interestingly, in two-dimensional spacetimes [22] this trace contains only derivatives of the metric up to second order and thus is consistent with the weakened form of axiom 5

where $u_{\mu\nu}$ is an ordinary (“ c -number”) tensor field on spacetime and I denotes the identity operator. (This is most easily shown by considering the matrix elements of $U_{\mu\nu}$ in an orthonormal basis. The off-diagonal matrix elements must vanish; the diagonal elements must all be equal, since if any two were unequal one could take linear combinations of them and construct orthogonal states for which the matrix element of $U_{\mu\nu}$ is nonvanishing.) $u_{\mu\nu}$ is, of course, a functional of the spacetime geometry alone.

Next, we apply the causality condition, axiom 4. Since $T_{\mu\nu}$ and $\tilde{T}_{\mu\nu}$ satisfy axiom 4(a), so does $U_{\mu\nu}$, i.e., for fixed “in” state $\langle U_{\mu\nu} \rangle$ at p depends only on the spacetime geometry to the causal past of p . But by Equation (3.2), this expectation value (for any normalized “in” state) is just $u_{\mu\nu}(p)$. Thus, $u_{\mu\nu}(p)$ is a functional of the spacetime geometry to the past of p alone. But by exactly the same argument, Condition 4(b) implies that $u_{\mu\nu}(p)$ depends only on the spacetime geometry to the future of p ! Hence, $u_{\mu\nu}(p)$ can depend only on spacetime geometry at p itself, i.e., it is a completely local object. Thus, we have proven that if $g_{\rho\sigma}$ and $g'_{\rho\sigma}$ are two spacetime metrics which agree in an open neighborhood of p , then $u_{\mu\nu}(p) = u'_{\mu\nu}(p)$.

We now use axiom 5 to show that $u_{\mu\nu}(p)$, in fact, can only depend on the components of the metric and their first derivatives (in some fixed chart) at p . The proof of this result is similar to the proof that a distribution with support consisting of a single point must be a sum of delta functions and derivatives of delta functions at that point (see, e.g., Courant and Hilbert [23], page 785). We first establish that $u_{\mu\nu}(p)$ depends only on the metric and its first four derivatives at p .

Let $g_{\rho\sigma}$ and $g'_{\rho\sigma}$ be two spacetime metrics whose components at p and derivatives of components up to fourth order are equal at p . We prove that $u_{\mu\nu}(p) = u'_{\mu\nu}(p)$ as follows: By smoothly joining the metric obtained from the fourth order “Taylor series” polynomial in the coordinates—determined by the components of $g_{\rho\sigma}$ and their derivatives at p —to the actual spacetime metric $g_{\rho\sigma}$ one can (by accomplishing this joining closer and closer to p) explicitly construct a sequence of spacetime metrics $\{(g_{\rho\sigma})_n\}$ which are fourth degree polynomials in the coordinates in a neighborhood of p and converge to $g_{\rho\sigma}$ in the manner specified in the hypothesis of axiom 5. One can also construct a similar sequence $\{(g'_{\rho\sigma})_n\}$ converging to $g'_{\rho\sigma}$. Since $g_{\rho\sigma}$ and $g'_{\rho\sigma}$ agree up to fourth derivatives at p , these sequences will have the property that for each n , $(g_{\rho\sigma})_n = (g'_{\rho\sigma})_n$ in a (sufficiently small) neighborhood of p . From the results we have already established using axiom 4, it follows that at point p $(u_{\mu\nu})_n = (u'_{\mu\nu})_n$ since these quantities are sensitive only to the spacetime geometry at p . But axiom 5 implies that the limit of the sequence $\{(u_{\mu\nu})_n\}$ is $u_{\mu\nu}$ and the limit of $\{(u'_{\mu\nu})_n\}$ is $u'_{\mu\nu}$. Hence, $u_{\mu\nu}(p) = u'_{\mu\nu}(p)$.

We have now shown that $u_{\mu\nu}$ at p is a function of the metric and its first four derivatives at p alone. However, if $u_{\mu\nu}(p)$ had a nontrivial dependence on derivatives of the metric of second order or higher, then the derivatives of $u_{\mu\nu}$ up to third order would have a nontrivial dependence on derivatives of the metric of higher order than fourth. It is easy to show that this would contradict axiom 5. Thus, we obtain the desired result that $u_{\mu\nu}(p)$ depends only on the metric and its first derivative at p . Note that we are really only using axiom 5 to give a precise expression to the notion that neither $T_{\mu\nu}$ nor $\tilde{T}_{\mu\nu}$ should contain any “local curvature terms” involving derivatives of the metric of second order or higher.

We have now shown that $u_{\mu\nu}(p)$ is a function of the metric and its first derivative at p . However, it is well known that the most general tensor of rank $(2, 0)$ which can be constructed out of the components of the metric and their first derivatives in some chart at p is a constant multiple of the metric tensor itself. Thus, there exists a constant C such that $u_{\mu\nu}(p) = Cg_{\mu\nu}(p)$. C is, of course, independent of the spacetime geometry, since otherwise $u_{\mu\nu}(p)$ would not be a function of only the metric and its first derivatives at p . Putting all the above results together, we have shown that,

$$U_{\mu\nu} = Cg_{\mu\nu}I. \quad (3.3)$$

Consider, now, the case of Minkowski spacetime. By axiom 2, (normal ordering) $T_{\mu\nu}$ and $\tilde{T}_{\mu\nu}$ must agree in this case, so $U_{\mu\nu} = 0$ for Minkowski spacetime. Hence $C = 0$ and we obtain our final result,

$$U_{\mu\nu} = 0. \quad (3.4)$$

Note that axiom 3 ($\nabla^\mu \langle T_{\mu\nu} \rangle = 0$) was not used in this uniqueness proof.

If we had used the weakened form of axiom 5, the proof would have proceeded as above except that we would have found that $u_{\mu\nu}(p)$ is a function of the metric and its derivatives up to second order at p . Many tensors can be constructed out of the metric and its first and second derivatives. However, we now use axiom 3, which implies that $\nabla^\mu u_{\mu\nu} = 0$. According to a result of Lovelock [24], this condition restricts the most general form of $u_{\mu\nu}$ to

$$u_{\mu\nu} = Cg_{\mu\nu} + C'G_{\mu\nu} \quad (3.5)$$

where $G_{\mu\nu}$ is the Einstein tensor and C and C' are constants. Again axiom 2 implies $C = 0$, and we are left with the result that $T_{\mu\nu}$ and $\tilde{T}_{\mu\nu}$ can differ at most by a multiple of the Einstein tensor $G_{\mu\nu}$.

It is clear that this remaining ambiguity is of a trivial nature. Since $\tilde{T}_{\mu\nu} = T_{\mu\nu} - C'G_{\mu\nu}I$, if we write the semiclassical Equation (1.2) for $\tilde{T}_{\mu\nu}$ and bring the term in $G_{\mu\nu}$ to the left-hand side of the equation we see that it is equivalent to the semiclassical equation for $T_{\mu\nu}$ but with a different numerical value of the gravitational constant G . Intuitively it is clear that this arbitrariness should go away if we impose further requirements on the theory which imply that in the classical limit, the semiclassical theory should reduce to general relativity with the correct value of the gravitational constant, i.e., that the renormalized $T_{\mu\nu}$ already incorporates any renormalization of the gravitational constant which may occur. A suitable set of such requirements are the following scaling assumptions: The only fundamental lengths which occur in the semiclassical theory are the Planck length $(\hbar G/c^3)^{1/2}$ and (for a massive field) the Compton wavelength of the particles \hbar/mc . The stress-energy operator is independent of G . Furthermore, the expected stress energy of, say, the “in” or “out” vacua should go to zero (or, at last, not diverge) in the limit $\hbar \rightarrow 0$. Under these assumptions (which are frequently made implicitly if not explicitly) simple dimensional arguments show that $u_{\mu\nu}$ cannot equal $C'G_{\mu\nu}$ unless $C' = 0$. Thus, we again recover uniqueness.

It should be emphasized that the above uniqueness proofs show that there can be *at most* one stress-energy operator $T_{\mu\nu}$ which satisfies the axioms, but, of course, the proofs do not establish that one actually exists. This issue will be dealt with in the next section.

Finally, we remark that a close analogy exists between the phenomena of particle production by a gravitational field and the production of charged particles by an external electromagnetic field in flat spacetime. In the electromagnetic case, one would expect the particle creation to “back react” on the electromagnetic field via a current operator j_μ . One can write down analogues of axioms 1–5 for j_μ . The uniqueness proof given above for $T_{\mu\nu}$ then carries over to a uniqueness proof for j_μ in the electromagnetic case.

IV. Existence

In this section we shall attempt to demonstrate the existence of a renormalized stress-energy operator satisfying axioms 1–5. For concreteness we explicitly consider the case of the massless Hermitian scalar field,

$$\nabla_\mu \nabla^\mu \phi = 0. \quad (4.1)$$

We shall give a “point splitting” type prescription for $T_{\mu\nu}$, which, it will be argued, satisfies at least the first four axioms. However, one would need further detailed information on how certain quantities vary under changes of the metric in order to check axiom 5. Thus, we do not succeed in demonstrating existence, but we do show that the first four axioms are self-consistent. Since several key assumptions on which the prescription for $T_{\mu\nu}$ is based have not been rigorously proven (at least as far as I am aware), the results of this section must be considered as plausibility arguments rather than theorems.

We begin by recalling some results on second order hyperbolic wave equations originally proven by Hadamard [25]. If we consider an equation like Equation (4.1) in a 4-dimensional analytic spacetime, we may seek a solution $S(x)$ which has singularity structure near the point x' of the form

$$S(x, x') = U/\sigma + V \ln \sigma + W \quad (4.2)$$

where 2σ is the square of the geodesic distance between x and x' and U , V , and W are analytic functions of x . If we write V and W as a power series in σ ,

$$\begin{aligned} V(x, x') &= \sum_{l=0}^{\infty} V_l(x, x') \sigma^l, \\ W(x, x') &= \sum_{l=0}^{\infty} W_l(x, x') \sigma^l \end{aligned} \quad (4.3)$$

substitute this in Equation (4.1) and set the coefficient of each power of σ equal to zero, we get an equation for U and recursion relations for V_l and W_l . One can show [20] that the equation for U uniquely determines it up to normalization, and given U , each V_l is uniquely determined. $W_0(x, x')$ is completely arbitrary but given W_0 , all the remaining $W_l (l \geq 1)$ are uniquely determined by the recursion relations. One can then show [20] that the series expansions for V and W actually converge in a sufficiently small neighborhood of x' , so this procedure truly does yield a solution of Equation (4.1). We will refer to such a solution as a Hadamard elementary solution.

The existence of Hadamard elementary solutions has been a powerful tool in analyzing solutions to second order equations. In the case of an elliptic equation of

second order, an analogous construction also works and one can prove [20] that $S(x, x')$ viewed as a function of x' (for fixed x) is a fundamental solution of the adjoint equation. This fact allows one to prove existence theorems for solutions. In the hyperbolic case, the Hadamard elementary solution has been used to prove that the Cauchy problem is well posed [20].

In the following we shall assume that the above construction also works for C^∞ spacetimes (with the series for V and W now interpreted as asymptotic series). We shall also use the following facts about Hadamard elementary solutions. First, as has already been noted above, with the normalization of U chosen, a Hadamard elementary solution is completely characterized by the arbitrary function $W_0(x, x')$. Furthermore, U and V do not depend on W_0 , i.e., W_0 enters only the recursion relations for $W_l (l \geq 1)$. For a self-adjoint equation like Equation (4.1), we shall assume⁸ that if $W_0(x, x')$ is chosen symmetric in x and x' , $W_0(x, x') = W_0(x', x)$, then so is the Hadamard solution, $S(x, x') = S(x', x)$. Finally, we note that the solution for U and the recursion formulae [20] for V_l and $W_l (l \geq 1)$ involve integrals along the geodesic joining x and x' and thus, given $W_0(x, x')$, the Hadamard solution in a (convex) neighborhood of x' is completely determined by the spacetime geometry in that neighborhood.

As is well known, the quantum field operator $\phi(x)$ in curved spacetime makes complete mathematical sense when interpreted as an operator valued distribution. On the other hand, as mentioned in Section I, no obvious mathematical meaning can be given to the products like $\phi(x)\phi(x)$ and, indeed, this is the origin of the stress-energy renormalization problem. However, a product like $\phi(x)\phi(x')$ (i.e., the product of two distributions at different spacetime points) makes perfect sense as a 2-point distribution, i.e., if we smear separately in x and x' we get a well-defined operator. In particular, the expectation value in any state

$$G(x, x') = \frac{1}{2} \langle \phi(x)\phi(x') + \phi(x')\phi(x) \rangle \tag{4.4}$$

is a well-defined distribution in x and x' . Since ϕ satisfies Equation (4.1) (in a distributional sense), $G(x, x')$ is also a (distributional) solution of Equation (4.1) in each variable. One may formally express the expected stress energy $\langle T_{\mu\nu} \rangle$ in terms of $G(x, x')$ as [10]

$$\langle T_{\mu\nu}(x) \rangle = \frac{1}{2} \lim_{x' \rightarrow x} \{ \nabla_{\mu'} \nabla_{\nu} G + \nabla_{\mu} \nabla_{\nu'} G - g_{\mu\nu} \nabla_{\alpha} \nabla^{\alpha'} G \} \tag{4.5}$$

where the primed indices denote vectors in the tangent space at x' so that, e.g., $\nabla_{\mu'} \nabla_{\nu} G$ denotes the bitensor formed by fixing x' and taking the derivative of G at x and then fixing x and taking the derivative of $\nabla_{\nu} G$ at x' . The limit in Equation (4.5), of course, does not exist. The basic idea of the ‘‘point splitting’’ renormalization prescription is to examine the singularity behavior of this limit, identify the divergent terms, subtract them off, and get a finite result.

The starting point for detailed study of this limit is the assertion that $G(x, x')$ is, in fact, a Hadamard elementary solution. (This may be verified directly for flat spacetime and it is certainly plausible that it should remain true in curved spacetime,

⁸ It should be possible to give a proof of this statement directly from the recursion relations for V_l and W_l . In the elliptic case the analogous result follows from the arguments on page 165 of Garabedian [20]

but as far as I am aware, no complete proof has been given.) Assuming this is the case, the quantity inside the brackets in Equation (4.5) is finite provided x and x' cannot be connected by a null geodesic (i.e., $\sigma(x, x') \neq 0$). One can explicitly compute the behavior of this quantity as x' approaches x along a timelike or spacelike geodesic. Some ambiguities arise in extracting the “finite piece”, however, from the fact that it depends on the direction of “point splitting” as well as the precise manner in which x' approaches x (e.g., even if x' is required to approach x along a geodesic one gets different results depending on whether one uses “endpoint expansion” or “midpoint expansion” [11]).

We propose here a slightly modified type of “point-splitting” prescription which is free of the above direction dependent and limit taking ambiguities. To renormalize a particular expectation value, we take $G(x, x')$ for that state, Equation (4.4), and subtract from it a Hadamard solution $S(x, x')$ constructed in the manner to be specified below. To compute $\langle T_{\mu\nu} \rangle$, we substitute this difference

$$F(x, x') = G(x, x') - S(x, x') \quad (4.6)$$

in place of G in Equation (4.5) i.e.,

$$\langle T_{\mu\nu}(x) \rangle = \frac{1}{2} \lim_{x' \rightarrow x} \{ \nabla_\mu \nabla_\nu F + \nabla_\mu \nabla_\nu F - g_{\mu\nu} \nabla_\alpha \nabla^\alpha F \}. \quad (4.7)$$

Since U and V are the same for all (normalized) Hadamard solutions, the singularities in G and S cancel and F is smooth. Hence, the coincidence limit Equation (4.7) will yield a well-defined finite answer, free of direction-dependent and other ambiguities. The expectation value $\langle T_{\mu\nu} \rangle$ for all states, of course, uniquely determines all matrix elements of $T_{\mu\nu}$, i.e., it determines the operator $T_{\mu\nu}$.

We construct $S(x, x')$ as follows: we examine $\frac{1}{2} \langle 0 | \phi(x) \phi(x') + \phi(x') \phi(x) | 0 \rangle$ in Minkowski spacetime and compute the function $W_0(x, x')$ for this Hadamard solution. (For the massless scalar field it is zero; for massive fields it is a nonzero constant.) Returning to curved spacetime, we define $S(x, x')$ to be the unique Hadamard solution (with the standard normalization of U) with W_0 chosen equal to this Minkowski space value.

Now that we have given our prescription for $T_{\mu\nu}$, we can ask if it satisfies the axioms. It is not difficult to verify that axiom 1 is satisfied. We can express the matrix elements between orthogonal states in terms of the difference of expectation values. By Equation (4.7) we can express such a difference in terms of $F_1 - F_2$ where F_1 is the quantity defined by Equation (4.6) for the first expectation value and F_2 is the similar quantity for the second. But $F_1 - F_2 = G_1 - G_2$ and when we substitute and take the coincidence limit we recover the formal expression, in accordance with axiom 1.

To verify axiom 2 (normal ordering in Minkowski spacetime) it is only necessary to show that $\langle 0 | T_{\mu\nu} | 0 \rangle = 0$. But this is obvious, since by construction of S , for the vacuum state in Minkowski spacetime we have $G = S$ so $F = 0$ and thus $\langle 0 | T_{\mu\nu} | 0 \rangle = 0$.

$F(x, x')$ is a well-behaved solution of Equation (4.1) in both x and x' . It follows from the same arguments as given in Appendix B of Ref. [11] that expectation values of $T_{\mu\nu}$ are conserved, $\nabla^\mu \langle T_{\mu\nu} \rangle = 0$. Thus axiom 3 is also satisfied.

$G(x, x')$ is constructed out of expectation values of the field operator ϕ . Consequently, it is causal, i.e., for a fixed “in” state $G(x, x')$ depends only on the

spacetime geometry in the union of the past of x and x' . On the other hand, S is locally constructed, i.e., $S(x, x')$ depends only on the spacetime geometry along the geodesic connecting x and x' . Thus, in particular, for a fixed “in” state $F(x, x')$ depends only on the spacetime geometry in the causal past of the geodesic connecting x and x' . Hence, when we take the coincidence limit, Equation (4.7), it follows that $\langle T_{\mu\nu} \rangle$ is causal in the manner required by axiom 4(a). Exactly similar arguments show that $T_{\mu\nu}$ also satisfies axiom 4(b).

It seems clear from the nature of the above prescription that $\langle T_{\mu\nu} \rangle$ will, in some sense, vary continuously under continuous changes of the metric. However, we have no strong reason to believe that the rather stringent condition required by axiom 5 will be satisfied. In order to check whether this condition is satisfied, we would need detailed information of how variations in the spacetime geometry affect the W_0 term in the Hadamard solution $G(x, x')$. This does not appear to be an easy task. Thus, unfortunately, it is not clear whether axiom 5 is satisfied.

However, the uniqueness proof of Section III shows that the first four axioms uniquely determine a stress-energy operator up to the addition of a conserved local curvature term which vanishes in flat spacetime. (By “local curvature term” we mean precisely a quantity whose value at a point p depends only upon the spacetime geometry at p .) Hence, we obtain the following criterion for the existence of a stress-energy operator satisfying all five axioms: If our renormalized stress-energy tensor defined above can be decomposed into the sum of a local curvature term and a term satisfying axiom 5, then this second term will satisfy all the axioms. On the other hand, if such a decomposition is not possible, then no stress-energy operator satisfying all the axioms can exist (since if it did exist we could obtain the required decomposition by taking the difference between it and our stress-energy operator).

Finally, we point out that since the difference of two stress-energy operators satisfying the first four axioms is a local curvature term which vanishes in flat spacetime, any two such stress-energy operators must agree in any flat region of a (non-flat) spacetime. Thus, even though our prescription for $T_{\mu\nu}$ may not satisfy axiom 5, it must agree with the true renormalized $T_{\mu\nu}$ (assuming one exists!) in any flat region. It is not difficult to check that in a flat region our prescription agrees with the “point-splitting” prescription used by DeWitt [7] in his investigation of the Casimir effect. Thus, the results of this paper may be viewed as justifying the original rather ad hoc derivation of this effect given by DeWitt.

V. Conclusions

In this paper, we have listed five axioms for the renormalized quantum stress-energy operator. As argued extensively in Section II, it appears that these axioms must be satisfied in order to yield a reasonable semiclassical theory of quantum fields interacting with gravity within the framework described in Section I. The main result of the paper, proven in Section III, is that there can be at most one stress-energy operator satisfying all the axioms. Unfortunately, we were unable to fully demonstrate the existence of such a stress-energy operator in Section IV. However, making a few assumptions stated in Section IV, we were able to give an explicit prescription for a renormalized $T_{\mu\nu}$ which satisfies at least the first four axioms. In particular, this demonstrates (modulo the assumptions) that these four axioms are

self-consistent. Furthermore, in terms of our prescription for $T_{\mu\nu}$ we gave a necessary and sufficient condition for the existence of a stress-energy operator satisfying all five axioms. The stumbling block toward further progress along these lines is axiom 5. In order to check whether a given prescription for $T_{\mu\nu}$ satisfies axiom 5, we need to know detailed information on how quantities vary under a general class of variations of the spacetime geometry. Even for classical fields this information is not easy to come by.

What if no stress-energy operator satisfying our five axioms exists? Of course, if a precise technical condition like axiom 5 is violated but another condition which has the same essential content is satisfied, one would merely modify the axioms. But if the five axioms are inconsistent in a nontrivial manner, then unless one can somehow evade the arguments of Section II one would be forced to conclude that “back reaction” effects cannot be treated within the context of the semiclassical approximation.

What if one succeeds in obtaining the renormalized $T_{\mu\nu}$? In the first place one will be able to proceed with calculations of many interesting effects such as the detailed nature of black hole evaporation. One would also be able to determine the effect of particle creation in the early universe on the dynamics of the universe. Of particular interest is the behavior of the semiclassical theory in situations where, in the context of classical general relativity, spacetime singularities are produced. As discussed in Section II, the quantum stress-energy tensor is not expected to satisfy a local energy condition. Hence the classical singularity theorems do not apply and it is possible that in many cases singularities may be avoided or at least take on a different character in the semiclassical theory than in general relativity. In the long run, perhaps the insight we will get from a complete semiclassical theory will point us in the right direction toward obtaining a quantum theory of gravitation.

Acknowledgements. I wish to thank all my colleagues, particularly Robert Geroch, for many valuable discussions throughout the course of this work.

References

1. Ashtekar, A., Geroch, R.: Rep. Prog. Phys. **37**, 1211 (1974)
2. Wald, R.: Commun. math. Phys. **45**, 9 (1975)
3. Hawking, S.W.: Commun. math. Phys. **43**, 199 (1975)
4. Hawking, S.W.: Phys. Rev. D **14**, 2460 (1976);
Parker, L.: Phys. Rev. D **12**, 1519 (1975)
5. Bardeen, J.M., Carter, B., Hawking, S.W.: Commun. math. Phys. **31**, 161 (1973)
6. Bekenstein, J.D.: Phys. Rev. D **9**, 3292 (1974)
7. DeWitt, B.: Phys. Rep. **19**, 295 (1975)
8. Brown, L.: Phys. Rev. D **15**, 1469 (1977)
9. Dowker, J.S., Critchley, R.: Phys. Rev. D **13**, 3224 (1976);
Dowker, J.S.: Conformal anomalies in quantum field theory (to be published);
Hawking, S.W.: Commun. math. Phys. (in press)
10. Christensen, S.M.: Phys. Rev. D **14**, 2490 (1976)
11. Adler, S., Lieberman, J., Ng, Y.J.: Ann. Phys. (in press)
12. Parker, L., Fulling, S.A.: Phys. Rev. D **9**, 341 (1974)
13. Rumpf, H.: Covariant treatment of particle creation in curved spacetime (to be published)
14. Hartle, J., Hawking, S.W.: Phys. Rev. D **13**, 2188 (1976)
15. Candelas, P., Raine, D.J.: Phys. Rev. D **15**, 1494 (1977)
16. Wightman, A.S.: The Dirac equation. In: Aspects of quantum theory (eds. A.Salam, E.Wigner).
Cambridge: University Press 1973

17. Hawking, S. W., Ellis, G. F. R.: The large scale structure of spacetime. Cambridge: University Press 1973
18. Davies, P. C. W., Fulling, S. A.: Radiation from moving mirrors and from black holes (to be published)
19. Woodhouse, N.: Phys. Rev. Lett. **36**, 999 (1976)
20. Garabedian, P. R.: Partial differential equations. New York: John Wiley and Sons, Inc. 1964
21. Christensen, S. M., Fulling, S. A.: Trace anomalies and the Hawking effect (to be published)
22. Davies, P. C. W., Fulling, S. A., Unruh, W. G.: Phys. Rev. D **13**, 2720 (1976)
23. Courant, R., Hilbert, D.: Methods of Mathematical Physics, Vol. II. New York: Interscience Publishers 1962
24. Lovelock, D.: J. Math. Phys. **13**, 874 (1972)
25. Hadamard, J.: Lectures on Cauchy's problem in linear partial differential equations. New Haven: Yale University Press 1923

Communicated by R. Geroch

Received January 25, 1977

