Black-hole evaporation and ultrashort distances

Theodore Jacobson
Department of Physics, University of Maryland, College Park, Maryland 20742
(Received 28 March 1991)

The role played by ultrahigh frequencies or ultrashort distances in the usual derivations of the Hawking effect is discussed and criticized. The question "would a black hole radiate if there were a Planck scale cutoff in the rest frame of the hole?" is posed. Guidance is sought from Unruh's fluid-flow analogue of black-hole radiation, by taking into account the atomic nature of the fluid. Two arguments for black-hole radiation are given which assume a Planck length cutoff. One involves the response of static accelerated detectors outside the horizon, and the other involves conservation of the expectation value of the stress tensor. Neither argument is conclusive, but they do strongly suggest that, in spite of reasonable doubt about the usual derivations of black-hole radiation, a "safe" derivation which avoids our ignorance of ultrashort-distance physics can likely be formulated. Remaining open questions are discussed.

I. INTRODUCTION

The extrapolation of quantum field theory from flat to curved background spacetimes is physically justified by the equivalence principle. In a freely falling reference frame, quantum fields should behave as they do in flat space, provided only length scales less than the radius of curvature are involved. Although global constructs such as a preferred vacuum state or single-particle states do not generally carry over from flat to curved space, one can sometimes identify physically reasonable boundary conditions which allow truly curvature-dependent predictions to be made. Hawking's discovery [1] that black holes emit thermal radiation is one such prediction.

The existence of black-hole radiation is interesting not only because it brings about primordial black-hole explosions with potentially observable consequences, but also because of the way in which it extends the analogy between thermodynamics and classical black-hole mechanics [2] to a remarkable interplay of gravity, quantum field theory and thermodynamics that culminates in the generalized second law of thermodynamics [3, 4].

$$\delta(S_{\text{outside}} + \tfrac{1}{4}A_{\text{horizon}}) \geq 0 \; . \tag{1}$$

The first term is the entropy of matter outside the horizon and the second term is one fourth the area of the horizon, measured in square Planck lengths. This generalized second law (GSL) holds only if one takes seriously the state of the quantum fields around the black hole that give rise to the Hawking radiation [4]. Without the Hawking effect the GSL would *fail*, because one could slowly lower a box containing entropy to the horizon, hold it there, and then drop it in the hole. The entropy of the outside would go down, but the horizon area would remain unchanged, because all of the mass-energy of the box would have been converted to work on whatever was lowering the rope from far away. The Hawking radiation (in combination with the acceleration radiation seen by the static box along its accelerated world line) alters the situation, because it produces a buoyancy force requiring the box to be actually pushed towards the horizon.

Mostly because of the tie-in with thermodynamics, Hawking's prediction of black-hole radiation is considered by many to have a certain ring of truth. In fact, it seems there is very little doubt in the community that black holes do in principle evaporate. Yet there is some cause for doubt, due to the role apparently played by ultrahigh frequencies or ultrashort distances in the derivation. In Sec. II of this paper I will elaborate on this cause for doubt, discuss the cluster of issues that it raises, and pose the question: "Would a black hole radiate if there were a Planck scale cutoff in the rest frame of the hole?" In Sec. III guidance is sought from a fluid-flow analogue of black-hole radiation introduced by Unruh, by analyzing the effects of the atomic nature of the fluid. In Sec. IV two arguments for black-hole radiation are given which assume a Planck length cutoff. One involves the response of static accelerated detectors outside the horizon, and the other involves conservation of the expectation value of the stress tensor. Neither argument is conclusive, but they do strongly suggest that, in spite of reasonable doubt about the usual derivations of black-hole radiation, a "safe" derivation that avoids our ignorance of ultrashort-distance physics can likely be formulated. In the final section I discuss some of the remaining questions, and describe some approaches to resolving them.

II. HAWKING RADIATION AND SHORT DISTANCES

According to Hawking's analysis [1], if an object collapses to form a nonrotating black hole of mass M, it will radiate to infinity as if it were a hot body at a temperature

$$T_H = \hbar c^3/8\pi kGM \simeq 10^{-7}\ \mathrm{K}\,(1.5\ \mathrm{km}/r_s) \simeq 1\ \mathrm{MeV}/k\,(10^{16}\mathrm{g}/M)\ , \qquad (2)$$

where r_s is the Schwarzschild radius $2GM/c^2$. This is a result of linear quantum field theory in the time-dependent curved spacetime of the hole, and it is widely expected to occur for interacting fields as well. The initial vacuum state of the quantum fields evolves to a state that is not the vacuum state far from the hole in the future. After Hawking's original derivation, other derivations of this effect were given, often in the analytically extended vacuum Schwarzschild spacetime of an "eternal black hole." Since my goal is to understand the physical justification of the derivation, I prefer to stick to the realistic case of the collapsing body.

The essential feature, for our purposes, of the spacetime of a collapsing object is the infinite time dilation effect: an observer at rest with respect to the hole and far away will measure a finite time interval between the passage of any ingoing null geodesic and the passage of a later ingoing null geodesic that forms a generator of the horizon, whereas after propagating through the object the corresponding outgoing geodesics will be separated by an infinite time interval, as measured by another such observer (since the geodesic on the horizon is *never* seen). One finds that in order for a wave packet of a linear massless field to emerge from the hole with a fixed frequency ω_{out} at large radius at time t, it must begin its journey into the collapsing matter with a blueshifted frequency ω_{in} which grows exponentially with time as $\exp(t/4M)$, where M is the mass of the hole. (Here and hereafter we use units with $c = \hbar = k = G = 1$, unless otherwise indicated, so in particular $2M$ is the time it takes for light to travel one Schwarzschild radius.)

Thus, for example, at a time t since the hole formed, an outgoing mode with frequency equal to the Hawking temperature $1/8\pi M$ originated as an ingoing mode with frequency above the Planck frequency if t is greater than $4M\ \ln(8\pi M/M_{\mathrm{Planck}})$. If there were a Planck frequency cutoff on the ingoing modes, the Hawking radiation would seemingly be extinguished on this relatively short time scale [5].

Another way of looking at the role played by ultrashort distances in black-hole evaporation is provided by a recent paper of Fredenhagen and Haag [6], in which it is shown that the existence of Hawking radiation for free fields can be derived using only the form of the short-distance singularity of the two-point function $\langle\phi(x)\phi(y)\rangle$ at the sphere where the horizon exits the collapsing matter. That is, the *only* assumption needed concerning the quantum state of the fields is the assumption that, in an infinitesimally thin shell surrounding the horizon at one particular time, the strongest singularity in the two-point function has the same form as it has in the Minkowski vacuum, namely, σ^{-1}, with σ the square of the geodesic interval between x and y. This condition would rule out for example the Boulware vacuum, which has infinite stress at the horizon. The condition says essentially that the very-short-distance behavior of the vacuum fluctuations "appears to freely falling observers to be the same as in the Minkowski-space vacuum." This approach brings out very clearly the sense in which the Hawking effect seems to hinge on ultrashort-distance behavior of the fields.

In his original paper Hawking offered a physical picture, involving pair creation, for the origin of black-hole radiation. In this picture, a positive-energy particle is created outside the horizon and escapes to infinity. Its negative-energy partner tunnels inside the horizon, where it can exist as a real particle because the time translation Killing field (which defines the conserved energy) is spacelike there. Were the escaping particle to be created sufficiently far from the horizon, it would not suffer too much redshifting on the way out, and ultrahigh frequencies would not be involved in the process. However, since the tunneling probability dies off exponentially with distance, there seems to be a preference for the pair to be created "near" the horizon. But how near is "near"? Unfortunately, this picture of pair creation has not been made the basis of a solid derivation, so it is not possible to reach any definite conclusions from it.

It is important to determine whether the Hawking effect *requires* arbitrarily high-frequency modes and short distances, or whether their role can be eliminated in a more circumspect analysis. In particle physics, when there is ignorance about what is going on at short distances, one aims to extract predictions which are insensitive to the short-distance physics. In the same spirit, one should not be satisfied with a derivation of the Hawking effect unless it is independent of our ignorance about short distances.

A. Lorentz noninvariance

There is an essential flaw in the analogy however. In the particle-physics context, to speak of "short-distance physics" or a "short-distance cutoff" on the validity of an interacting theory in no way implies a lack of Lorentz invariance. The fact that the center-of-mass energy or momentum transfer of an interaction is large can be characterized by Lorentz-invariant scalar quantities. In contrast with this, the Hawking effect has nothing to do with interactions of the quantum fields. The high-frequency modes whose role is being questioned here can always be locally transformed to low frequency by an appropriate Lorentz transformation. In particular, no matter how high the frequency of an ingoing mode may be in the rest frame of the hole, the frequency in a frame that chases after it with sufficient speed is arbitrarily low.

If one is willing to *assume* exact local Lorentz invariance, then this cause for doubt about the Hawking effect is removed [7]. However such an assumption is unwise for at least two reasons. First, just as we have no experience with interactions at ultrahigh energy, we have no experience with physics in reference frames moving ultrafast relative to us. The fact that Lorentz-invariant *theories* agree with present observations serves only to place limits on possible deviations from Lorentz invariance [8, 9]. Although today it may seem almost paradoxical to imagine a preferred state of "rest," one should keep in mind the fact that the Universe as a whole does define such a rest frame: that of the microwave-background radiation.

It is not inconceivable that this cosmic rest frame also plays some role in local physics at short distances, even though nothing of the sort is true in currently accepted theory.

The second reason it is unwise to assume exact Lorentz invariance is that the assumption commits us to assigning an infinite number of degrees of freedom to the fields in any finite spatial volume. This idealization leads inexorably to the divergences in quantum field theory and the nonrenormalizability of quantum gravity. While it is possible that the solution to these problems will come in the form of a deus ex machina such as string theory, it is also possible that the solution lies in the removal of the offending assumption, exact Lorentz invariance.

B. Imposing a cutoff

Not knowing what form physics might take at ultrashort distances, and in order to explore the consequences of postulating only a finite number of degrees of freedom in a finite volume, let us frame our question about the Hawking effect this way: would a black-hole radiate if there were a mode cutoff at the Planck frequency in the rest frame of the hole? Although the rest frame is well defined only far from the hole, we will later extend it inward using radial timelike geodesics.

The first thing to notice is the fact that the presence of a Planck frequency cutoff on initial modes of a free field theory not only eliminates the Hawking radiation, according to the usual derivation, but it eliminates the corresponding outgoing modes themselves from the field degrees of freedom. That is, it is not just that the modes are not excited to the Hawking temperature, but that they are not present at all. A classical antenna attempting to radiate away from the hole would emit no waves at those frequencies. This observation poses a puzzle, however: it is implausible that just by a little thought one can arrive at the profound conclusion that either there is no cutoff on field modes at any scale, or there are bizarre effects due to missing modes outside black holes. It seems clear that there must be yet another possibility, namely, some mechanism, due to field interactions or perhaps quantum gravity effects, which could "regenerate" the mode frequencies that have been lost by redshifting. This idea of mode regeneration may at first seem hopelessly vague; however, there is in fact a model from condensed-matter physics that is analogous to the black-hole situation, and which exhibits just this behavior. We turn now to a discussion of this model, with the aim of clarifying the notion of mode regeneration and making it more plausible.

III. QUANTUM FLUID FLOW MODEL OF A BLACK HOLE

In a paper published in 1981, Unruh [10] invented a fluid-flow model of a black hole. His motivation was in part to eventually consider the issue of short-distance physics being discussed here, as well as the quantum back-reaction problem. Although in the paper the fluid is treated exclusively as a continuum, a very interesting result is obtained, namely, that "the same arguments which lead to black-hole evaporation also predict that a thermal spectrum of sound waves should be given out from the sonic horizon in transsonic fluid flow."

It turns out that the linearized perturbations of an irrotational flow can be described by a massless scalar field propagating in a curved background spacetime whose metric is determined by the background fluid velocity field. For a spherically symmetric, static convergent flow which exceeds the speed of sound at some radius, the metric has approximately the form of a black-hole metric. Quantizing the sound field, and *assuming* the field is in the comoving ground state near the horizon for all times, Unruh patched on to the Hawking argument outside the horizon to conclude that there is a thermal flux of phonons at the temperature

$$T = (\hbar/2\pi k)(\partial v/\partial r) \simeq 10^{-7}\ \mathrm{K}\left(\frac{\partial v}{\partial r}\bigg/\frac{100\,\mathrm{m/s}}{1\,\mathrm{mm}}\right), \quad (3)$$

where $\partial v/\partial r$ is the gradient of the background velocity field at the horizon. Comparison with (2) indicates that $\partial v/\partial r$ plays the role of the black-hole surface gravity $\kappa/c = c/2r_s$.

Taking into account the atomic nature of the fluid, one has a model in which there is both a preferred comoving local rest frame, and a high-frequency–short-wavelength cutoff on the phonon field modes. Let us see how, in this model, the mode regeneration effect referred to in the previous section can occur. While it may be that the sonic horizon radiation is a general effect which would occur in *any* atomic fluid, it is helpful to examine the behavior of a particular physical fluid, since we do not yet know which features (if any) of the mode behavior near the cutoff are generic. Thus, to make the model concrete, the fluid will be taken here to be helium-4 at zero temperature and pressure. The cutoff arises because the phonons are really collective excitations of the helium atoms, and when the wavelength becomes much smaller than the average interatomic spacing, the field-theoretic description of the excitations is no longer valid.

An elementary excitation is a stable state of the fluid with definite energy and momentum. The energy ε and momentum $\mathbf{p}$ of the excitations for helium-4 satisfy a particular dispersion relation $\varepsilon = \varepsilon(p)$, shown in Fig. 1 for $T = 1.1$ K at the corresponding vapor pressure [11]. States lying on this curve cannot decay to combinations of other states on this curve, as can be seen from energy and momentum conservation. For wavelengths longer than about 10 Å ($p/\hbar = 2\pi/\lambda \sim 0.6$ Å^{-1}), the dispersion relation for these excitations is approximately linear, with slope equal to the speed of sound, $s = 238$ m/s. These are the true "phonons." For shorter wavelengths the slope of the dispersion curve varies (and even becomes negative for a while), so the group velocity $v_g = d\varepsilon/dp$ deviates from the sound velocity s. Excitations near the minimum are called rotons, and the spectrum cuts off at a wavelength of about 2.3 Å. The cutoff occurs when the single excitation description fails, presumably because the corresponding state can decay into

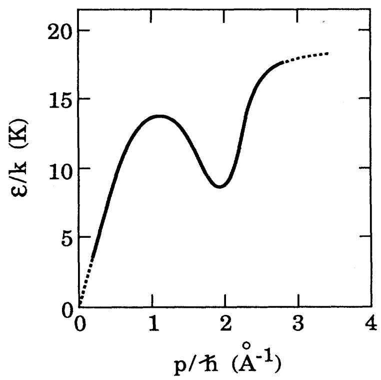

FIG. 1. Dispersion curve for elementary excitations of liquid-helium-4 at $T = 1.1$ K and saturated vapor pressure.

a pair of lower-energy rotons or has some other decay modes [12].

In Unruh's black-hole model, the fluid has a spherically symmetric convergent flow which is faster towards the center and exceeds the speed of sound at some radius, forming a sonic horizon [13]. Now excitations that are propagating upstream loose energy as they climb away from the sonic horizon, just as in the black-hole case. It is the energy ε in the comoving rest frame that is decreasing, while the energy ε_0 with respect to the asymptotic rest frame is constant [14]. The relation between the two is given by

$$\varepsilon_0 = \varepsilon + \mathbf{p} \cdot \mathbf{v} = \varepsilon - pv \; , \tag{4}$$

where $\mathbf{v}$ is the velocity of the fluid, and in the second equality we have assumed $\mathbf{p}$ is antiparallel to $\mathbf{v}$.

Consider an excitation propagating away from the horizon upstream in the fluid. Having swum against the flow gradient, the excitation must have had a shorter wavelength the closer it was to the horizon. If the cutoff, which is at fixed wavelength in the comoving frame, is reached before we can extrapolate all the way back to the horizon, then we can infer that some sort of mode reconstruction must have taken place to produce this excitation. It may seem that this will certainly happen; however, one must take into account the fact that the location of what we might call the "effective horizon" depends on the group velocity v_g and therefore on the momentum.

To determine whether the cutoff is reached before the horizon, consider the energy $\varepsilon(v)$ and momentum $p(v)$ as functions of the local velocity of the fluid, and require that they simultaneously satisfy the dispersion relation Fig. 1 and the energy relation (4), with the lab energy ε_0 held fixed. To solve graphically, examine the intersection between the dispersion curve and the straight line $\varepsilon = \varepsilon_0 + pv$. This point of intersection gives the energy and momentum of the propagating excitation as a function of v. First consider a long-wavelength phonon far from the horizon where $v = 0$, with momentum less than that of the first maximum on the dispersion curve. Then extrapolating backwards, v increases and the point slides up the dispersion curve toward the maximum. Before it reaches the maximum, the straight line becomes tangent to the dispersion curve. At this point, the phonon is at rest with respect to the lab frame, which means that the effective horizon has been reached *before* the cutoff.

Now consider an excitation with momentum greater than that of the roton minimum when far from the horizon. Extrapolating backwards, as v increases the corresponding point slides towards the cutoff along the dispersion curve. If at some v the straight line is tangent to the dispersion curve, then the effective horizon is reached before the cutoff. However, as long as v_g does not vanish as the cutoff is approached [12], outgoing excitations with energy greater than some threshold lab energy will, when propagated backwards, reach the cutoff before reaching the effective horizon. More precisely, if v_g at the cutoff is some $v_c > 0$, then the threshold lab energy is given by $\varepsilon_c - p_c v_c$, where ε_c and p_c are the energy and momentum at the cutoff.

There are therefore situations in which an elementary excitation propagating freely upstream could not have originated as an elementary excitation arbitrarily near to the horizon. This leads to a puzzle, the analog of which previously led us to doubt the consistency of imposing a cutoff on initial modes in the collapsing black-hole spacetime. The puzzle is, how did the outgoing excitation or mode come to exist?

The answer for the superfluid must be that the mode arises from an interaction of two (or more) lower-energy modes, for example in a process which is the time reverse of the decay to a roton pair. Even in the ground state at zero temperature, the effective-field-theory decription of helium is an interacting one. A detailed analysis of the issue of horizon radiation would require finding the time evolution of the quantum state of the interacting fluid as the inhomogeneous background flow is established and thereafter.

Fortunately it is not really necessary to understand the details of the corresponding process in superfluid helium to draw from the model an important lesson for the black-hole situation. We learn that the existence of a high-frequency cutoff on ingoing modes *can* be consistent with the existence of a full spectrum of outgoing modes (below the cutoff), provided there is some nonlinear process involving the short-distance physics that can "regenerate" the outgoing modes just outside the horizon. Moreover, we learn that one should not rely on any derivation of black-hole radiation that involves free propagation through this "zone of ignorance" outside the horizon, where some unknown physics is presumably taking place.

IV. DO BLACK HOLES EVAPORATE?

In view of our ignorance of physics at ultrashort distances, we seek a derivation of black-hole evaporation which does not assume "business as usual" outside the horizon in the "zone of ignorance" just mentioned. To get a feeling for the physical issues involved, I will discuss here two arguments which assume the existence of a cutoff at the Planck scale. Neither argument is conclusive, however they should provide guidance in the search for a bona fide derivation.

A. The cutoff

We are entertaining the hypothesis of a cutoff which, for reasons discussed earlier, must violate Lorentz invariance. It is therefore necessary to specify the reference frame in which the cutoff is applied. The least arbitrary assumption would be that in quasiflat regions of spacetime, the cutoff is uniform in the rest frame of the cosmos, i.e., that of the microwave background radiation. Now we must face the question of how the cutoff is to be extended into the spacetime region surrounding a black hole that is approximately at rest with respect to the cosmic rest frame.

As already discussed, it is probably *not* correct simply to impose the cutoff on the field degrees of freedom before the hole forms, and then propagate it using the free field equations, for this leads to an implausibly redshifted cutoff on outgoing modes. Instead, we must make an assumption for the nature of the cutoff in each region of spacetime that somehow takes into account whichever mode regeneration processes may have occurred. Lacking the fundamental cutoff theory, the only assumption that suggests itself is that the preferred rest frame is carried down into the hole by observers who fall freely from far away at rest. The resulting frame will be called the "falling frame." Note that in the fluid flow analog, the falling frame corresponds to the comoving rest frame of the fluid.

Of course this definition of the local rest frame is unambiguous only in the spherically symmetric case, and then only if one restricts to radial free fall trajectories that are always infalling, ignoring those which pass through the center of the collapsing matter and later fall back. In more general situations, the fundamental cutoff theory would be needed to even formulate the nature of the cutoff. But in order to explore the qualitative implications of a cutoff, it seems reasonable to stick to the spherically symmetric case and to impose the cutoff in the falling frame.

There is another frame that has a preferred status, namely, the frame of the static observers at constant radius. To interpret the consequences of the cutoff it is useful to know how the cutoff is viewed in this "static frame." This is determined from the relative velocity of the frames, by the Doppler shift factor.

The falling four-velocity u is the unit tangent of a radial timelike geodesic which starts from rest at infinity. Everywhere along the geodesic u has unit inner product with the time-translation Killing field $\xi = \partial/\partial t$, where t is the Schwarzschild time coordinate. The norm of ξ is given by $N := (\xi \cdot \xi)^{1/2} = (1 - 2M/r)^{1/2}$, where r is the Schwarzschild radial coordinate. The function N will also be called the "lapse," since it gives the relation $ds = N\,dt$ between proper time and dt at fixed radius. Note that $N \to 0$ at the horizon, and $N \to 1$ at spatial infinity.

The static four-velocity is parallel to the time translation Killing field; hence, it is the unit vector $\hat{\xi} = N^{-1}\xi$. The inner product of falling and static four-velocities at any point is therefore given by $u \cdot \hat{\xi} = N^{-1}$. From this one finds that the Doppler shift factor for massless modes is given by $N(1 + \sqrt{1 - N^2})^{-1}$ for outgoing modes, and by its inverse for ingoing modes. We shall consider only massless modes since in any case the frequencies that will be important in our discussion are much larger than any mass we might wish to consider.

If the cutoff on both outgoing and ingoing modes is at the Planck frequency in the falling frame, $\omega_c^{\text{fall}} = \omega_P$, then in the static frame near the horizon where $N \ll 1$ it will be given by

$$\begin{aligned} \omega_{c,\,\text{out}}^{\text{stat}} &\simeq \tfrac{1}{2} N \omega_P \ , \\ \omega_{c,\,\text{in}}^{\text{stat}} &\simeq 2N^{-1}\omega_P \ . \end{aligned} \qquad (5)$$

Note that the cutoff on outgoing modes approaches *zero* near the horizon in the static frame.

Recall that the Hawking radiation is in some sense related to the thermal excitation of static detectors at a temperature that diverges at the horizon. However, these detectors cannot be thermally excited at a temperature above the cutoff, so the fact that the cutoff $\omega_{c,\,\text{out}}^{\text{stat}}$ approaches zero near the horizon precludes their being excited and calls into question the existence of the Hawking effect itself. In order to get a feeling for the implications of this line of reasoning, we now review the relation between accelerated detector response and Hawking radiation, and then consider the effect of imposing a cutoff as just described.

B. Accelerated detectors

A detector with uniform acceleration a in the usual vacuum state of flat Minkowski space will be thermally excited [15–17] to a temperature $T_U = (\hbar/2\pi k)(a/c)$. This result is easily seen by considering the two-point function $\langle 0|\phi(x(s))\phi(0)|0\rangle$, where $x(s)$ is the accelerated world line and s is the proper time along it. In the Minkowski vacuum state the two-point function depends only on the invariant interval $x^2(s)$. For uniformly accelerated motion, $x(s)$ is the result of exponentiating a boost, so it is periodic in the translation of s by the imaginary quantity $i2\pi/a$, where a is the proper acceleration. Now a two-point function periodic in imaginary time is a thermal two-point function, and one concludes that the detector feels a temperature $T_U = a/2\pi$ [18].

Following Unruh [15] and DeWitt [16], this result can be applied to a static detector outside the horizon of a black hole. The acceleration of a world line at constant radius is given by $a = N^{-1}(1 - N^2)^2\,\kappa$, where $\kappa = 1/4M$ is the "surface gravity" of the hole. At the horizon $N \to 0$, so the acceleration diverges, and at infinity $N \to 1$, and the acceleration vanishes.

The equivalence principle suggests that the response of a static detector in Schwarzschild spacetime should be related somehow to that of a uniformly accelerated detector in Minkowski space. However, the response of a detector is determined by the two-point correlation function of field fluctuations, which depends on the state of the field, and not just on the acceleration of the detector. If the two-point function along the world line of the static detector were the same function of proper time as it is along an accelerated world line in the flat-space Minkowski vacuum, one could conclude that the detector is thermally excited at the Unruh temperature

$$T_U = a/2\pi = N^{-1}(1 - N^2)^2\, T_H \ , \qquad (6)$$

where $T_H = \kappa/2\pi$ is the Hawking temperature. As viewed from infinity, this temperature suffers a redshift given by the lapse function N, so at infinity it is given by

$$T_{U,\infty} = (1 - N^2)^2\, T_H \ , \qquad (7)$$

which agrees with the Hawking temperature of the hole provided the detector is placed just outside the horizon where $N \to 0$.

But why should the two-point function behave in this way, that is, can this behavior along the static world line of the detector be deduced from our assumptions about the initial state? Far from the hole, the acceleration vanishes, so any thermality must be attributed "by hand" to the state of the field, rather than being derived from the acceleration temperature effect. It is only when applied near the horizon that the above argument has any chance of being correct and having predictive power. Moreover, the state resulting from collapse is *not* the thermal equilibrium (Hartle-Hawking) state with both ingoing and outgoing modes thermally populated; instead, only the *outgoing* modes are thermally populated. I do not know any property of the state of the ingoing modes near the horizon that holds generically in the collapsing case other than regularity of the stress tensor or equivalent conditions, and this does not distinguish between the equilibrium and collapse states.

Only if we restrict attention to *outgoing* modes *near* the horizon does the argument appear to carry some weight. At late times, the modes that finally peel away from the horizon have been exponentially redshifted, so it is plausible that their quantum state is independent of the details of the initial state and the collapse process, and it is plausible that, in a thin shell just outside the horizon, that state is close to the Minkowski vacuum state. In any case, this is the best I can do to solidify the argument.

Now let us consider how the argument is affected by the presence of a Planck frequency cutoff in the falling frame. The essential point is that the static cutoff frequency for outgoing modes (5) goes to zero at the horizon, whereas the acceleration temperature (6) diverges. To apply the argument one must therefore stay far enough from the horizon that the acceleration temperature is below the cutoff. The question that arises is whether this is *so* far that the argument breaks down completely. In fact, we see from (5) and (6) that the outgoing static cutoff is equal to the acceleration temperature when the lapse function is given by

$$N_{\rm min} \simeq (T_H/\omega_P)^{1/2} \simeq (l_P/r_s)^{1/2} \ , \qquad (8)$$

provided $l_P \ll r_s$, so that the higher-order terms in N can be neglected. At this lapse, the temperature "at infinity" indicated by (7) differs by a term of order $O(l_P/r_s)$ from the Hawking temperature.

The above reasoning suggests that as long as the Schwarzschild radius r_s is much greater than the Planck length l_P, the cutoff will not make a significant difference for observations far from the hole. More specifically, it suggests that the spectrum of emitted radiation would differ from Hawking's thermal spectrum by terms of order $O(l_P/r_s)$. Another indication of the analysis is that for a given mass hole, there is a maximum local temperature to which a static detector would be excited, given by the acceleration temperature at the position where it is equal to the cutoff. From (5) and (8), this temperature is seen to be

$$T^{\rm local}_{\rm max} \simeq N_{\rm min}\omega_P \simeq (T_H\omega_P)^{1/2} \ . \qquad (9)$$

Since the existence of this limit violates the usual redshift relation for the variation of temperature in a static gravitational field, it may have some implications for the existence or properties of a thermal-equilibrium state surrounding a black hole.

C. Stress-energy tensor

In making the argument that a static detector just outside the horizon of a black hole is thermally excited, it was necessary to assume that the two-point function along the detector world line behaves like that along a uniformly accelerated world line in the Minkowski vacuum. A plausibility argument was advanced to support this assumption for the *outgoing* mode contribution; however, it is an admittedly weak one, especially because to actually *calculate* the two-point function from initial conditions before the collapse would involve whatever physics takes place in the zone of ignorance where the mode reconstruction process presumably occurs. We seek instead a derivation that uses assumptions which can be justified on general physical grounds alone. A promising strategy in this regard is based on conservation of the stress tensor.

Consider a conformally invariant field in the two-dimensional black-hole spacetime that results when the angular coordinates are dropped. Employing a slight modification of the analysis of Christensen and Fulling [19], it will be shown below that the outgoing flux of energy from the hole at late times is completely determined just by conservation and finiteness of the expectation value of the renormalized stress tensor, together with the value of the anomalous trace. The flux so determined agrees with the Hawking result for the net flux, and it is notable that the result is obtained without evaluating the Bogoliubov coefficients connecting the ingoing and outgoing modes, without specific assumptions regarding the initial state, and without assuming time independence of the final state. Moreover, the result applies to general interacting fields, as long as they are classically conformal invariant so that their trace is determined by the trace anomaly.

Conservation and finiteness are assumptions that can be plausibly justified without regard to the short-distance physics; however, this is not so for the value of the trace. In this section I shall estimate the modification of the trace due to the existence of a Planck length cutoff, and then examine the implications for the outgoing energy flux in this two-dimensional model. Of course it would be preferable to have a four-dimensional analysis; however, in four dimensions it is not possible to derive the exact amount of energy flux in this manner, because the tangential stress remains undetermined [19]. Nevertheless, as discussed in Ref. [19], *some* information can be garnered from physical assumptions regarding the tangential stress, and it is possible that our argument in the presence of a cutoff can at least be partially extended to four dimensions.

The Schwarzschild metric with angular coordinates dropped can be expressed in null coordinates by

$$ds^2 = C\,du\,dv\,, \quad C = 1 - 2M/r\,,$$
$$e^{(v-u)/4M} = (r/2M - 1)e^{r/2M}\,. \tag{10}$$

At the horizon $C \to 0$ and $u \to \infty$, whereas at future null infinity $C \to 1$ and $v \to \infty$. Let $T_{\mu\nu} = \langle \hat{T}_{\mu\nu} \rangle_{\rm ren}$ be the expectation value of the renormalized stress tensor. Then $T_\mu{}^\nu{}_{;\nu} = 0$ implies

$$T_{uu}(u,v) = T_{uu}(u,v_0) - \frac{1}{4}\int_{v_0}^{v} C T_{,u}\,dv\,, \tag{11}$$

where $T = \langle \hat{T}^\mu_\mu \rangle_{\rm ren}$ is the trace. The T_{uu} component of the stress tensor signifies outgoing energy flux when evaluated at infinity.

Now the assumption of finiteness (in nonsingular coordinates) implies that $T_{uu}(u,v_0)$ must go to zero as u approaches ∞ at the horizon. To see why, suppose that $(\lambda(u,v),\ v)$ is a regular coordinate system at the horizon. Then $\partial\lambda/\partial u \to 0$ at the horizon, since λ covers a finite range in an infinite range of u. The tensor transformation law gives $T_{uu} = (\partial\lambda/\partial u)^2 T_{\lambda\lambda}$, which shows that T_{uu} vanishes at the horizon by the finiteness assumption. More precisely, taking λ to be an affine parameter along a line of constant v, we have $d\lambda = C\,du$, so $(\partial\lambda/\partial u)^2 = C^2 \simeq e^{(v-u)/2M}$ near the horizon. Thus conservation and finiteness alone imply

$$T_{uu}(u,v) = O(e^{-u/2M}) - \frac{1}{4}\int_{v_0}^{v} C T_{,u}\,dv\,. \tag{12}$$

At late times at fixed radius, the first factor falls exponentially with time, so the energy flux out at infinity is determined solely by the trace. If the trace is entirely due to the anomaly, (12) shows that this flux is *independent* of the initial state of the field, on account of the redshifting away of any incoming energy. Using the standard value of the trace anomaly for a scalar field, $T = -R/24\pi$, and letting $v \to \infty$, (12) yields

$$T_{uu}(u,\infty) = \frac{1}{768\pi M^2} + O(e^{-u/2M})\,, \tag{13}$$

which agrees with the Hawking flux as $u \to \infty$.

Now let us consider the above argument in the context of a cutoff theory. We assume there exists a quantity corresponding to $\langle \hat{T}_{\mu\nu} \rangle_{\rm ren}$, even if it is not calculated by the usual rules of quantum field theory but by some other theory which remains unknown. We assume moreover that it is conserved, and that it is regular at the horizon; however, we do not assume the usual trace anomaly. Nevertheless, to proceed from (12) we need to know something about the trace. Since it cannot be computed in an unknown theory, I propose simply to compute it using standard quantum-field theory with an *ad hoc* cutoff. Fujikawa's method [20] will be employed, because that method seems best suited to accommodate the effects of a cutoff. In other calculational schemes, the dimensionful scale comes in "through the back door," and I do not know how to relate it to a physical scale.

In Fujikawa's approach, the anomaly arises because although the action is conformal invariant, the measure in the functional integral is not. Evaluating the nontrivial Jacobian resulting from the change of measure under conformal transformation of the fields, one finds [20]

$$T = \langle \hat{T}^\mu_\mu(x) \rangle = \sum_n \phi_n(x)\phi_n(x)\,, \tag{14}$$

where $\{\phi_n\}$ is a complete set of eigenmodes of the differential operator $\mathcal{D}$ in the action, $\mathcal{D}\phi_n = \lambda_n{}^2\phi_n$. The sum is divergent, and to regulate it Fujikawa introduces a mass scale Λ and defines the sum by the limit $\Lambda \to \infty$ of

$$\sum_n \phi_n(x) e^{-\lambda_n^2/\Lambda^2}\phi_n(x)\,. \tag{15}$$

So defined, the sum is convergent only for Euclidean signature spacetime, because only then are the eigenvalues $\lambda_n{}^2$ of $\mathcal{D}$ all positive. Since it is the Lorentzian cutoff theory itself that interests us, the use of Euclidean signature cannot be justified. Unfortunately, however, I have nothing better to offer at this time, so the Euclidean definition will be used and the result will be formally continued to Lorentz signature. It remains an important problem for future work to calculate the trace directly in the Lorentzian cutoff theory. In such a calculation, a dependence on the frame in which the cutoff is applied is expected to occur.

Using the Euclidean method, in the two-dimensional case the sum (15) has the form [21]

$$O(\Lambda^2) - \frac{1}{24\pi}R + O(R^2/\Lambda^2)\,. \tag{16}$$

The $O(\Lambda^2)$ term can be taken to renormalize the cosmological constant, the $O(\Lambda^0)$ term is the standard trace anomaly, and the $O(R^2/\Lambda^2)$ term is a correction term we are after. Setting the cutoff Λ equal to the inverse Planck length, the correction ΔT to the trace is seen to be of the form $\Delta T = O(l_P^2 R)R$. For the path of integration $u = \text{const}$, $v_0 < v < \infty$ in (12), this correction to the trace is bounded by $\Delta T \sim (l_P/r_s)^2 R$. Therefore the result for $T_{uu}(u,\infty)$ will differ from that given in (13) by a term of relative order $(l_P/r_s)^2$.

We conclude that if the stress tensor expectation value is conserved and finite, then in the two-dimensional model there is necessarily *some* flux of energy at infinity, provided only that the gradient of the trace is nonvanishing. If the trace is given by the usual trace anomaly, then the flux is precisely the Hawking flux. Our method of estimating the trace leads to the conclusion that the effect of a Planck-length cutoff is to modify the trace by a term of relative order $(l_P/r_s)^2$ near the hole, which leads to a change in the flux at infinity $\Delta T^{(\infty)}_{uu}$ of the same relative order:

$$\Delta T_{uu}^{(\infty)} = O(l_P^2/r_s^2)\, T_{uu}^{(\infty)} . \tag{17}$$

The analysis does not apply directly to the four-dimensional case, for which supplementary analysis regarding the modification of the tangential stress would be required. Supposing qualitatively similar results, we can conclude that the modification (17) is negligible for any but the tiniest black holes. In particular, it would not affect the order-of-magnitude estimate [1] of the energy released in a mini-black-hole explosion, because most of that energy would be released before the hole gets anywhere near the Planck size.

V. DISCUSSION

So, do black holes evaporate? The analysis presented here supports the viewpoint that black-hole evaporation is not really an ultrashort-distance effect, despite the role played by ultrashort distances in its usual derivations. Thus it seems that the existence of a Planck-scale cutoff or other unknown Planck-scale physics would probably not affect the existence of black-hole radiation. Nevertheless, many open questions remain.

The notion of mode regeneration was introduced in this paper to account for the compatibility of a short-distance cutoff with the infinite redshift effect of black holes. The implication is that nonlinear field propagation is crucial to a sound physical understanding of black-hole radiation. This is somewhat disturbing, since it leads one to question what happens if the regeneration process is not complete. Why wouldn't the final state depend upon some coupling constants, and how could this agree with the Hawking result? A possible answer goes like this. Since the regeneration process is reestablishing the full spectrum of vacuum fluctuations locally, and the Minkowski vacuum is statistically like a zero-temperature thermal state [22, 17], perhaps the regeneration is an equilibration-type process, that leads to a state which is independent of coupling constants provided enough time passes. This would suffice, since the "pure" Hawking radiation is, strictly speaking, what emerges from the hole at very late times after the collapse.

Another question is how can the process of mode regeneration be incorporated into quantum field theory? A changing dimension of the state space seems to be required; however, this would be something that lies entirely outside quantum field theory as we know it. The analogy with the atomic fluid suggests that the quantum dynamics of spacetime geometry must play a role. This issue arises also in the cosmological context, where one would need to understand how the cutoff could remain at the same proper scale in spite of the redshifting due to the expansion of the universe. One hypothesis is that, as the universe expands, new modes are always being added to the state space at the cutoff scale. The qunatum state of such modes must then be specified by an initial condition as they appear in the state space [23].

Lacking the fundamental short-distance theory, our present goal should perhaps be to identify the minimal and most physically sound assumptions needed to infer the existence of black-hole radiation. This was the motivation for the stress tensor argument of the preceding section. To improve on that, one needs to generalize from conformal invariant fields in two dimensions to any fields in four dimensions, but it is difficult to see how this can be done since the trace and tangential components of the stress are no longer determined by general considerations.

To avoid the need to understand the mode-regeneration process one might do the following. Instead of imposing a boundary condition on the state at past infinity (as in Hawking's original derivation [1]) or on the short-distance limit of the two-point function at the horizon at one time (as in the derivation of Fredenhagen and Haag [6]), one could impose a condition in a thin shell just *outside* the horizon at *every* time. Points in such a region are not causally independent, so one cannot impose an *initial* condition there; however, it may be possible to identify a condition which, if met, would yield the Hawking effect. Roughly speaking, the required condition would be that outgoing modes up to the cutoff frequency are in their "ground" state, as viewed by a freely falling observer [24]. The analysis of Sec. IV B suggests that this approach should work, as does a more careful analysis [23]. Further work could focus on whether or not this condition is met in any particular cutoff theory.

Finally, returning to the fluid flow model of a black hole, it is interesting to ask whether Unruh's conclusion that a sonic horizon will radiate thermally remains true when the fluid is treated not as a continuum but as the realistic superfluid, helium-4. If the temperature were high enough for something other than the long-wavelength phonons to emerge far from the horizon, then due to the form of the excitation spectrum Fig. 1, the effective horizon might occur at a speed substantially less than the speed of sound. Unfortunately this is impossible, for if the temperature were to be 1 K, then according to (3) the gradient of the velocity at the horizon would need to be 100 m/s per angström.

For long-wavelength phonons, we have seen that extrapolating backwards, the effective horizon is reached before the phonon reaches the first maximum of the dispersion curve Fig. 1, and therefore well *before* the cutoff is reached. Thus it seems that, due to the nonlinearity of the dispersion relation, the issue of the requisite modes being cut off does not arise for a prospective low-temperature sonic horizon. On the other hand, since the flow speed must be so high, the flow will be subject to various instabilities. If the instability to vortex creation at very low speeds could somehow be suppressed, there is still a critical speed [24], about 60 m/s, at which the superfluid is unstable to the appearance of a periodic roton condensate [25] which would complicate the flow pattern.

Ignoring such complications, Unruh's argument suggests that a low-temperature sonic horizon emits thermal phonons. However, we expect that each mode would be populated at a different temperature, determined by the gradient of the velocity field evaluated at the effective horizon for that mode. For modes whose wavelength is so short that mode regeneration must have occurred, the effective horizon would presumably be located where the mode was regenerated. It will be interesting to pursue this question, not only to determine what happens in helium-4, but also because it may be relevant to black-hole evaporation. After all, the dispersion relation for quantum field excitations might too be nonlinear at ultrahigh frequencies.

ACKNOWLEDGMENTS

For helpful discussions on various aspects of this research I would like to thank R. Ferrell, L. Ford, D. Garfinkle, B. Kay, L. Pitaevskii, J. Polchinski, D. Samuels, L. Smolin, and R. Wald. This research was supported by the National Science Foundation under Grant No. PHY89-10226 and by the University of Maryland. In addition I would like to thank the Relativity Group and the Department of Physics at the University of California at Santa Barbara for hospitality and support under NSF Grant No. PHY85-06686 while some of this research was being done.

[1] S.W. Hawking, Nature (London) **248**, 30 (1974); Commun. Math. Phys. **43**, 199 (1975).

[2] J.M. Bardeen, B. Carter, and S.W. Hawking, Commun. Math. Phys. **31**, 181 (1973).

[3] J.D. Bekenstein, Phys. Rev. D **7**, 2333 (1973); **9**, 3292 (1974).

[4] W.G. Unruh and R.M. Wald, Phys. Rev. D **25**, 942 (1982); **27**, 2271 (1983); correction: M.J. Radzikowski and W.G. Unruh, Phys. Rev. D **37**, 3059 (1988).

[5] G.W. Gibbons, in *Proceedings of the First Marcel Grossmann Meeting on General Relativity*, edited by R. Ruffini (North-Holland, Amsterdam, 1977). For related remarks including a careful discussion of what it means to say "a time t since the hole formed," see R.M. Wald, Phys. Rev. D **13**, 3176 (1976).

[6] K. Fredenhagen and R. Haag, Commun. Math. Phys. **127**, 273 (1990).

[7] Even assuming local Lorentz invariance, the role of the short-distance singularity of the two-point function in Ref. [6] might be further cause for doubt. I have not focused on this because I did not see how to impose a Lorentz-invariant cutoff while preserving the wave equation on which the derivation of [6] was based. Perhaps such a cutoff could be accommodated by modifying the derivation to apply to a quantum field satisfying a higher-derivative equation of motion. Of course this higher-derivative theory would then have its own problems however.

[8] D.I. Blokhintsev, Phys. Lett. **12**, 272 (1964).

[9] H.B. Nielsen and I. Picek, Nucl. Phys. **B211**, 269 (1983).

[10] W.G. Unruh, Phys. Rev. Lett. **46**, 1351 (1981).

[11] This graph is adapted from Fig. 9 of A.D.B. Woods and R.A. Cowley, Rep. Prog. Phys. **36**, 1135 (1973). For the current status of experiment and theory of liquid helium-4, see *Elementary Excitations in Quantum Fluids*, Proceedings of the Symposium, Hiroshima, Japan, 1987, edited by K. Ohbayashi and M. Watabe, Springer Series in Solid-State Sciences, Vol. 79 (Springer-Verlag, Berlin, 1989).

[12] The experimental information on the form of the dispersion relation at the highest momenta is somewhat cloudy. What is actually measured is neutron scattering cross sections, and to find the dispersion relation for elementary excitations one has to infer the energy of an excitation from the location of the peak it contributes to the cross section at fixed momentum transfer as a function of energy tranfer. Near the cutoff, this peak broadens due to the presence of decay modes which are not elementary excitations.

[13] One might instead consider a flow through a cylindrical capillary tube. To model the collapse to a black hole, one could begin with a uniform flow in the tube, and then contract the walls over some length, producing a thinner stretch of tube in which the speed of the fluid exceeds the speed of sound. Although perhaps simpler from an experimental point of view, the lower symmetry of this model would probably make a detailed theoretical analysis more complicated, except to the extent that it could be treated as a one-dimensional problem.

[14] The constancy of ε_0 is due to the fact that the excitation cannot exchange energy with the superfluid condensate. In a realistic situation this can be expected to be approximately true. An example of propagation of excitations in an inhomogeneous background superfluid flow is provided by the scattering of rotons by vortices [see, e.g., D.C. Samuels and R.J. Donnelly, Phys. Rev. Lett. **65**, 187 (1990)]. In this example conservation of ε_0 would fail to the extent that the motion of the vortex core can absorb energy. In the black-hole model, there is no such localized inhomogeneity, and conservation of ε_0 can be expected to hold more accurately. I would like to thank D.C. Samuels for a discussion of this point.

[15] W.G. Unruh, Phys. Rev. D **14**, 870 (1976).

[16] B.S. DeWitt, in *General Relativity, an Einstein Centenary Survey*, edited by S.W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1982).

[17] D.W. Sciama, P. Candelas, and D. Deutsch, Adv. Phys. **30**, 327 (1981).

[18] This argument is incomplete. If applied to the time-ordered (Feynman) two-point function $G_F(x(s))$, it is correct once the appropriate analytic behavior of $G_F(x)$ is invoked. However, the response of a linearly coupled detector is determined by the positive-frequency two-point function $G_+((s))$. In terms of G_+, thermality is expressed not by periodicity but rather by the KMS condition $G_+(s) = G_+(-s - i2\pi/a)$, which implies a detailed balance at temperature $a/2\pi$. It turns out that $G_+(x)$ does enjoy the analytic behavior required to deduce the KMS condition from the periodicity of $x(s)$ under translation of s by $i2\pi/a$. However, $G_+(x)$ is not, in general, periodic itself. See S. Takagi, Prog. Theor. Phys. Suppl. **88**, 1 (1986), for a thorough discussion of these issues.

[19] S.M. Christensen and S.A. Fulling, Phys. Rev. D **15**, 2088 (1977).

[20] K. Fujikawa, Phys. Rev. D **23**, 2262 (1981).

[21] K. Fujikawa, Phys. Rev. D **25**, 2584 (1982).

[22] D.W. Sciama, in *Quantum Gravity 2*, edited by C.J. Isham, R. Penrose, and D.W. Sciama (Clarendon, Oxford, England, 1981).

[23] T. Jacobson (work in progress).

[24] W. G. Unruh, Phys. Rev. D **15**, 365 (1977).

[25] L.D. Landau, Zh. Eksp. Teor. Fiz. **5**, 71 (1941).

[26] L.P. Pitaevskii, Pis'ma Zh. Eksp. Teor. Fiz. **39**, 423 (1984) [JETP Lett. **39**, 511 (1984)].