Black hole thermodynamics and information loss in two dimensions

Thomas M. Fiola,[1] John Preskill,[2] Andrew Strominger,[1] and Sandip P. Trivedi[2]
[1]*Department of Physics, University of California at Santa Barbara, Santa Barbara, California 93106-9530*
[2]*Lauritsen Laboratory of High Energy Physics, California Institute of Technology, Pasadena, California 91125*
(Received 23 March 1994)

Black hole evaporation is investigated in a $(1+1)$-dimensional model of quantum gravity. Quantum corrections to the black hole entropy are computed, and the fine-grained entropy of the Hawking radiation is studied. A generalized second law of thermodynamics is formulated, and shown to be valid under suitable conditions. It is also shown that, in this model, a black hole can consume an arbitrarily large amount of information.

PACS number(s): 04.70.Dy, 04.60.Kz

I. INTRODUCTION

Hawking's discovery of black hole radiance [1] established a deep and satisfying link connecting gravitation, thermodynamics, and quantum theory. But it also raised some disturbing puzzles. Foremost among these are the mystery of black hole entropy and the paradox of information loss. These two puzzles are closely related. Together they comprise a crisis in fundamental physics.

Black hole thermodynamics has a compelling beauty. Bekenstein's bold conjecture [2] that a generalized second law of thermodynamics applies to processes involving black holes, combined with Hawking's explicit calculation of the black hole temperature, led to the remarkable result that a black hole has an intrinsic entropy given by $\frac{1}{4}$ the area of the event horizon (in Planck units). But previous efforts to verify the generalized second law [3,4] have been limited to quasistationary processes, and to the leading semiclassical approximation. In this paper we will study black hole thermodynamics in two-dimensional spacetime. For the special case of two dimensions we are able to go substantially further than previous analyses, by considering processes that are not quasistationary, and by taking explicit account of quantum-mechanical back reaction effects. We will propose a precise statement of the generalized second law, and will demonstrate that it is valid in a particular two-dimensional model, under suitable conditions.

In Hawking's semiclassical theory of black hole evaporation [1], the radiation emitted by the black hole was found to be exactly thermal [5]. Thus, in the leading semiclassical approximation, the radiation carries no information about the initial quantum state of the object that collapsed to form the black hole. This property of the radiation led Hawking to assert [6] that quantum-mechanical information can be destroyed when a black hole forms and then subsequently evaporates completely. Although the semiclassical approximation is not exact, it is highly plausible that more accurate calculations would still support the conclusion that the outgoing radiation carries very little information; the key point is that once it has fallen past the global horizon, the collapsing body is out of causal contact with the radiation emitted from the black hole. Still, no complete analysis of the microscopic state of the radiation has ever been carried out. In this paper we study black hole evaporation in a two-dimensional model, taking into account quantum-mechanical gravitational back reaction effects. We find that the microscopic state of the emitted radiation carries essentially no information, as in the leading semiclassical calculations. Thus, loss of information really seems to occur in this model. (Or, perhaps, the information about the initial quantum state is retained inside a stable or long-lived black hole remnant [7].)

It was emphasized in Ref. [8] that two-dimensional models of quantum gravity can serve as a theoretical laboratory for investigating the fundamental issue of information loss. A further motivation for studying the Callan-Giddings-Harvey-Strominger (CGHS) model introduced in Ref. [8] is that it can be viewed as the low-energy effective field theory that governs the S-wave modes propagating on the background of a magnetically charged dilaton black hole in four dimensions. The (four-dimensional) dilaton black hole is of particular interest because it is a classical solution to a field theory that arises as a low energy approximation to string theory [9].

Although the CGHS model is far simpler than four-dimensional gravity, the full quantum theory of the model is still difficult to analyze. Therefore, CGHS studied a particular limit in which the model simplifies further. In this limit, the number N of matter field species tends to infinity, with $N\hbar$ held fixed. Then, to leading order in an expansion in $1/N$, but all orders in $N\hbar$, the quantum fluctuations of the dilaton and metric may be ignored, and only the fluctuations of the matter fields need be retained. Later, Russo, Susskind, and Thorlacius (RST) [10] showed (expanding on ideas introduced in Ref. [11]) that the model can be simplified still further by introducing a suitably chosen finite local counterterm. Our calculations in this paper will be carried out in the RST model, to leading order in $1/N$. We will review the RST model in Sec. II.

The generalized second law of thermodynamics states that the *total* entropy is nondecreasing, where the total entropy is the sum of the intrinsic entropy of the black hole and the thermodynamic entropy of the matter outside the black hole. To investigate the validity of the second law we will carry out a three-step program. First,

we must define precisely what is meant by the entropy due to the matter "outside" the black hole, and we must calculate this entropy. Second, we must find the correct expression for the black hole entropy in the RST model, including corrections to all orders in $N\hbar$ (but to leading order in $1/N$). Third, we must consider how the total entropy evolves, for a variety of initial conditions satisfied by the "collapsing" matter.

To obtain an expression for the entropy outside the black hole we erect a sharp boundary at the apparent horizon, and then trace over the matter field degrees of freedom behind the horizon to obtain a density matrix $\rho_{\rm out}$ for the matter fields outside. We then calculate the "fine-grained" entropy $S_{\rm FG} = -{\rm tr}(\rho_{\rm out}\ln\rho_{\rm out})$ of this density matrix. The fine-grained entropy quantifies the degree of entanglement of the quantum fields outside the horizon with those inside. We will see that this quantity can also be interpreted as the thermodynamic entropy of the matter outside the black hole. (Actually this is not quite the whole story. For a black hole formed from collapse, we will need to add to the fine-grained entropy another term, the "Boltzmann entropy" of the infalling matter. This will be explained in Sec. VI.)

Our calculations of the fine-grained entropy are performed in Sec. III. The method that we use is a generalization of the technique introduced by Unruh [12] in his analysis of the thermal bath seen by a uniformly accelerated observer, later extended to other cases by Holzhey [13]. These calculations are of some intrinsic interest apart from the relevance of the results to black hole physics, and we therefore discuss them in detail. As we will see, the fine-grained entropy has an ultraviolet divergence that arises from the entanglement of very-short-wavelength field fluctuations just inside and just outside the boundary. We regulate the divergence by introducing a short-distance cutoff (or, equivalently, by smoothing the boundary). One way to introduce this cutoff is to foliate the spacetime with spacelike slices; then on each slice we assign to the boundary at the apparent horizon a "thickness" of proper length δ. The resulting expression for the fine-grained entropy depends on this length δ, but it does not depend on the choice of the foliation, or on the coordinates used on each slice. In particular, two slices that cross the apparent horizon at the same point, but with a relative boost, yield the same value of the fine-grained entropy. As the black hole evolves, the proper length δ is held fixed.

In two-dimensional spacetime, the ultraviolet divergence is logarithmic, and the cutoff-dependent term in the entropy is merely a numerical constant. (At least, it is a constant from the time of formation of the black hole until its ultimate disappearance.) Thus, the divergence does not prevent us from making statements about the *change* in the entropy that are free from cutoff dependence.[1] The situation is rather different in four dimensions. Then the divergence is quadratic, and proportional to the area of the horizon [14]. To obtain a cutoff-independent expression for the entropy in four dimensions, we must absorb this divergence into the renormalization of Newton's gravitational constant G as described in Ref. [25].

The second step in our program, finding the corrected expression for the black hole entropy in the RST model, is carried out in Sec. V. We find a finite correction to the entropy computed in the leading semiclassical theory; the correction arises from the back reaction on the geometry when the black hole accretes or emits a small amount of radiation. We regard the black hole entropy as finite, and attribute the ultraviolet divergence in the total entropy to the matter fields surrounding the black hole. This is really a matter of convention, as our calculations fix the black hole entropy only up to an additive constant. We have chosen to fix the constant by demanding that the intrinsic entropy of the black hole vanishes as its mass goes to zero.

We assemble our expression for the total entropy in the RST model in Sec. VI, and analyze the evolution of the entropy in Secs. VI and VII. Section VII contains our analysis of the generalized second law of thermal dynamics. To prove the second law we need to make some additional assumptions. Most notably, we assume that the state of the matter that collapses to form the black hole is of a particular type—it is a coherent state built on the asymptotic inertial vacuum. Some such assumption seems to be necessary. It is possible to construct strange quantum states that pack a lot of entropy into a region at a very low cost in energy [13,15], or states with negative energy density (though this is not possible for coherent states). By preparing one of these strange states and dropping it into a black hole, the generalized second law that we have formulated *can* be violated, at least for a while. It would certainly be of interest to find a modified formulation of the generalized second law with more general validity and/or a concise characterization of how and when our formulation breaks down.

Our expression for the fine-grained entropy also enables us to address the question of information loss. We can imagine sustaining a black hole for an arbitrarily long time by feeding it mass to compensate for the Hawking radiation that it emits. It was emphasized in Ref. [16] that, if we draw a suitable spacelike slice through the geometry of this black hole, the amount of information stored in the portion of the slice that is behind the global horizon can be arbitrarily large. Thus one may argue that the number of internal quantum states of a black hole is not limited by its intrinsic Bekenstein-Hawking entropy. In Sec. IV, we analyze this sustained black hole (in the RST model) from the viewpoint of an observer who remains outside the horizon. We show that the fine-grained entropy outside the horizon can increase by an arbitrarily large amount. In accord with the conclusion of Ref. [16], then, we find that there is no consistent way to regard the density matrix $\rho_{\rm out}$ as arising from the entanglement of the degrees of freedom outside the horizon with a *finite* number of internal degrees of freedom of the black hole. Unless there are stable black hole remnants with an infinite number of internal degrees

[1]However, we will see that the change in the entropy (as we define it) at the moment of black hole formation, as well as the total entropy produced by the entire formation and evaporation process, do depend significantly on the cutoff.

of freedom [7], information is inevitably lost in the RST model.

In fact, the amount of lost information is even larger than one might have naively expected. The evaporation of a warm black hole into cold empty space is a thermodynamically irreversible process—the increase in the thermodynamic entropy of the emitted radiation is larger than the decrease in the entropy of the black hole [17,18]. (In one spatial dimension, it is larger by a factor of 2.) We find in Sec. VI that the *fine-grained* entropy outside the horizon behaves like the thermodynamic entropy. This means that the number of bits of lost information exceeds the number of bits needed to describe the initial quantum state of the collapsing matter, by a factor of (approximately) 2. Thus, the Bekenstein-Hawking entropy of the black hole formed in the initial collapse does not correctly quantify the amount of information that is ultimately lost.

The fine-grained entropy can increase indefinitely because the field modes localized close to the horizon are subjected to a redshift that increases exponentially as the black hole evolves. We introduced a short-distance cutoff of fixed proper length at the apparent horizon. But it follows that this cutoff, when expressed in terms of the asymptotically inertial coordinates used to define the quantum vacuum (or, equivalently, in terms of the wavelength measured at past null infinity), decreases exponentially along the horizon. As shorter and shorter wavelengths come into play, the degree of entanglement between the fields inside and outside the horizon increases correspondingly. It is this feature of the quantum state outside the horizon that is responsible for both the thermal character of the outgoing radiation and for the loss of an indefinite amount of information in the RST model.

It is evident that the conclusion that information is lost is predicated on assumptions about how *extreme* Lorentz boosts act on the matter degrees of freedom. (This point has been especially emphasized by 't Hooft [19], Jacobson [20], Susskind [21], and the Verlindes [22].) While loss of information apparently occurs in the RST model, it might be avoided in a different model with different physics at *very* short distances. In such a model it may be possible to attribute the fine-grained entropy to entanglement with a finite number of microscopic internal degrees of freedom of the black hole, and to interpret the Bekenstein-Hawking entropy of the black hole in terms of these internal degrees of freedom. The explicit construction of a model with these properties would be of great interest.

The content of this paper overlaps with that of several other references that have appeared while our work was being completed. In particular, Keski-Vakkuri and Mathur [23] have also analyzed the fine-grained entropy outside the horizon of an evaporating black hole. Where there is overlap, our conclusions are in agreement with theirs. Calculations of the fine-grained entropy for moving-mirror spacetimes (which closely resemble black hole spacetimes) have been discussed by Holzhey, Larsen, and Wilczek [24]. Quantum corrections to the black hole entropy have been considered recently by Susskind and Uglum [25], Callan and Wilczek [26], Kabat and Strassler [27], and Dowker [28].

II. REVIEW OF THE RST MODEL

An elegant model for two-dimensional black hole evaporation was introduced by Russo, Susskind and Thorlacius [10], expanding on ideas introduced in [11]. The RST model differs from the original CGHS model [8] by a finite counterterm that is fine-tuned to preserve a global symmetry. The counterterm makes it possible to solve the model exactly in the large-N limit, where N is the number of scalar matter fields. Numerical analyses [29,30] of the CGHS model indicate that it is qualitatively similar to the RST model, despite the fine-tuning.

The original CGHS model [8] of two-dimensional dilaton gravity has the classical action

$$S_{\text{classical}}=\frac{1}{2\pi}\int d^2x\sqrt{-g}\left[e^{-2\phi}[R+4(\nabla\phi)^2+4\lambda^2]-\frac{1}{2}\sum_{i=1}^{N}(\nabla f_i)^2\right], \tag{1}$$

where g is the metric, R is the curvature scalar, ϕ is the dilaton field, and the f_i are the N scalar matter fields. This model can be regarded as the low-energy effective action that governs the radial modes propagating on the near-extreme magnetically charged black hole of four-dimensional dilaton gravity. The length scale λ^{-1} is proportional to the magnetic charge of the four-dimensional black hole.

Two-dimensional dilaton gravity has classical black hole solutions. The mass of a black hole can be expressed in terms of the value ϕ_H of the dilaton field at the event horizon as

$$M_{\text{BH}}=\frac{\lambda}{\pi}e^{-2\phi_H}. \tag{2}$$

We may also interpret Eq. (2) as the deviation from the extremal limit of the mass of a four-dimensional black hole. Semiclassically, the two-dimensional black hole has a nonzero Hawking temperature. This can be computed from the periodicity of the black hole solution in Euclidean time [8] or from the Bogolubov transformation that relates the asymptotic incoming modes of the matter fields to the asymptotic outgoing modes [31]. The temperature is

$$T_{\text{BH}}=\frac{\lambda}{2\pi}, \tag{3}$$

which is independent of the black hole mass. Thus the two-dimensional black hole has an infinite specific heat. The four-dimensional magnetically charged dilaton black hole also has this property [9]. We obtain an expression for the black hole entropy S_{BH} by integrating the thermodynamic identity $dS=dM/T$; it is

$$S_{\text{BH}}=\frac{M_{\text{BH}}}{T_{\text{BH}}}=2e^{-2\phi_H}, \tag{4}$$

where we have fixed the constant of integration by demanding that $S_{\text{BH}}\to 0$ as $M_{\text{BH}}\to 0$. We may interpret Eq. (4) as $\frac{1}{4}$ the area of the event horizon of the classical four-dimensional dilaton black hole.

CGHS considered the semiclassical corrections to this classical theory, including the back reaction of the Hawking radiation on the geometry. To make the analysis tractable, they assumed that the number N of scalar matter fields is very large, and calculated the back reaction to leading order in an expansion in $1/N$. In leading order, the quantum fluctuations of the dilaton and metric can be ignored, and we need only include the one-loop correction to the energy momentum tensor of the scalars. This correction can be computed from the conformal anomaly. Equivalently we add to the classical action Eq. (1) the Polyakov-Liouville term [32]

$$S_{\text{Liouville}}=-\frac{N}{96\pi}\int d^2x\sqrt{-g(x)} \times\int d^2x'\sqrt{-g(x')}R(x)G(x,x')R(x') , \tag{5}$$

where G is a Green function of the operator ∇^2. This term expresses the dependence on the background geometry of the functional measure for the scalar fields. The field equations derived from the action $S_{\text{classical}}+S_{\text{Liouville}}$ have been studied numerically [33,29,30], but analytic solutions have not been obtained. However, RST (following Ref. [11]) found that the model can be solved exactly if a local counterterm

$$S_{\text{ct}}=-\frac{N}{48\pi}\int d^2x\sqrt{-g}\ \phi R \tag{6}$$

is added to the action.

To solve the model including (6) we introduce null coordinates $x^{\pm}=x^0\pm x^1$ and invoke the conformal gauge condition

$$g_{+-}=g_{-+}=-\tfrac{1}{2}e^{2\rho} , \quad g_{--}=g_{++}=0 . \tag{7}$$

We then have

$$S_{\text{classical}}=\frac{1}{\pi}\int d^2x\left[2e^{-2\phi}\partial_+\partial_-\rho +e^{-2\phi}(\lambda^2e^{2\rho}-4\partial_+\phi\partial_-\phi) +\frac{1}{2}\sum_{i=1}^{N}\partial_+f_i\partial_-f_i\right] ,$$

$$S_{\text{ct}}=-\frac{N}{12\pi}\int d^2x\ \phi\partial_+\partial_-\rho , \tag{8}$$

$$S_{\text{Liouville}}=-\frac{N}{12\pi}\int d^2x\ \partial_+\rho\partial_-\rho .$$

We now perform the field redefinition[2]

$$\Omega=\frac{12}{N}e^{-2\phi}+\frac{\phi}{2}+\frac{1}{4}\ln\frac{N}{48} , \tag{9}$$

$$\chi=\frac{12}{N}e^{-2\phi}+\rho-\frac{\phi}{2}-\frac{1}{4}\ln\frac{N}{3} . \tag{10}$$

In the large-N limit, with χ and Ω held fixed, the quantum effective action is then

$$S_{\text{eff}}=\frac{1}{\pi}\int d^2x\left[\frac{N}{12}(-\partial_-\chi\partial_+\chi+\partial_+\Omega\partial_-\Omega+\lambda^2e^{2\chi-2\Omega}) +\frac{1}{2}\sum_{i=1}^{N}\partial_+f_i\partial_-f_i\right] . \tag{11}$$

(The effects of ghosts may be ignored in the large-N limit.)

There is a residual conformal gauge invariance in (11). We fix this by the "Kruskal gauge" choice

$$\chi=\Omega , \tag{12}$$

which implies

$$\rho=\phi+\frac{1}{2}\ln\frac{N}{12} . \tag{13}$$

In Kruskal gauge the equations of motion are simply

$$\partial_+\partial_-\Omega=-\lambda^2 ; \tag{14}$$

the constraints can be expressed as

$$\partial_\pm^2\Omega=-T^f_{\pm\pm}-t_\pm . \tag{15}$$

Appearing on the right-hand side of Eq. (15) is the expectation field of the scalar field energy-momentum tensor, which we have separated into two terms. The first term T^f is the "classical" piece that can be obtained by varying the matter action with respect to the metric, except that, in order to simplify Eq. (15), we have chosen an unconventional normalization, namely,

$$T^f_{\pm\pm}=\frac{12\pi}{N}(T^f_{\pm\pm})_{\text{conv}}=\frac{6}{N}\sum_{i=1}^{N}\partial_\pm f_i\partial_\pm f_i . \tag{16}$$

In particular, since "Newton's constant" is of order $1/N$, we have scaled $T^f_{\pm\pm}$ by a factor of $1/N$, so that $T^f_{\pm\pm}$ of order one produces a back reaction of order one. Fluctuations of the energy-momentum tensor about its expectation value are suppressed by $1/N$, so the energy-momentum may be treated as a classical quantity to leading order.

The functions $t_\pm(x^\pm)$ in Eq. (15) arise because the constraints in Kruskal gauge are governed by the energy-momentum tensor normal ordered with respect to the "Kruskal vacuum" state—the state that contains no quanta that are positive frequency with respect to Kruskal time. The quantum state of the scalar fields can be expressed in terms of f creation operators acting on the f-vacuum state. If this f vacuum differs from the Kruskal vacuum, there is a finite normal ordering correction to the energy momentum tensor, in addition to the "classical" term T^f. In effect, this term arises because we must subtract a ρ-dependent piece of the vacuum energy from both sides of the constraint equation in order to express the left-hand side of Eq. (15) in terms of Ω. It is important to recognize that Eq. (15) holds only in the Kruskal gauge. On the right-hand side of this equation, $T^f_{\pm\pm}$ transforms as a tensor, but $t_\pm$ does not.

In our analysis of black hole formation and evaporation we will typically be interested in incoming quantum states that are coherent states built on the "σ vacuum."

[2]Our conventions differ slightly from [10] and agree with [34]. They are chosen so that χ and Ω are held fixed as N is taken to infinity.

The $\sigma^{\pm}$ coordinates are related to the Kruskal coordinates $x^{\pm}$ by

$$\lambda x^{+}=e^{\lambda\sigma^{+}}\ ,\quad \lambda x^{-}=-e^{-\lambda\sigma^{-}}\ . \tag{17}$$

These coincide with the inertial coordinates on $\mathcal{I}^{-}$; thus, the σ vacuum state $|0,\sigma\rangle$ is the state that appears to contain no quanta according to inertial asymptotic observers in the past. A left-moving coherent state can be build on this vacuum at $\mathcal{I}^{-}$, of the form

$$|f^{c},\sigma\rangle = A:\exp\left[\frac{i}{\pi}\sum_{i=1}^{N}\int d\sigma^{+}\partial_{+}f_{i}^{c}(\sigma^{+})\hat{f}_{i}(\sigma^{+})\right]:_{\sigma}|0,\sigma\rangle\ , \tag{18}$$

where the normal ordering is with respect to the σ vacuum, and A is a normalization constant. In Eq. (18), $\hat{f}$ denotes the quantum field, and f^{c} is its expectation value:

$$\langle f^{c},\sigma|\hat{f}_{i}(\sigma^{+})|f^{c},\sigma\rangle=f_{i}^{c}(\sigma^{+})\ . \tag{19}$$

For the energy-momentum tensor $:\hat{T}_{++}(x^{+}):_{K}$ normal ordered with respect to the Kruskal vacuum $|0,K\rangle$ we then have

$$\langle f^{c},\sigma|:\hat{T}_{++}:_{K}|f^{c},\sigma\rangle=T^{f^{c}}_{++}+\langle 0,\sigma|:\hat{T}_{++}:_{K}|0,\sigma\rangle\ ; \tag{20}$$

thus t_{+} in Eq. (15) can be expressed as

$$t_{+}=\langle 0,\sigma|:\hat{T}_{++}(x^{+}):_{K}|0,\sigma\rangle\ , \tag{21}$$

where it is understood that $\hat{T}_{++}$ has the unusual normalization in Eq. (16), and that $\langle\hat{T}_{++}(x^{+})\rangle$ is to be evaluated in the Kruskal gauge.

In flat space with the metric

$$ds^{2}=-d\sigma^{+}d\sigma^{-}=-\frac{dx^{+}dx^{-}}{\lambda^{2}x^{+}x^{-}}\ , \tag{22}$$

we may use standard methods [35] to compute

$$t_{\pm}^{0}(x^{\pm})=\langle 0,\sigma|:T_{\pm\pm}:_{K}|0,\sigma\rangle = -\frac{12\pi}{N}\frac{N}{48\pi(x^{\pm})^{2}}=-\frac{1}{4(x^{\pm})^{2}}\ . \tag{23}$$

The solution to Eqs. (14) and (15) then becomes

$$\Omega=-\lambda^{2}x^{+}x^{-}-\tfrac{1}{4}\ln[-4\lambda^{2}x^{+}x^{-}] \tag{24}$$

or

$$\phi=-\frac{1}{2}\ln\left[\frac{-\lambda^{2}Nx^{+}x^{-}}{12}\right]=-\lambda\sigma^{1}-\frac{1}{2}\ln\left[\frac{N}{12}\right]\ ; \tag{25}$$

this is the "linear dilaton vacuum" solution, so called because ϕ is a linear function of $\sigma^{1}=\frac{1}{2}(\sigma^{+}-\sigma^{-})$. The solution corresponding to general incoming matter from $\mathcal{I}^{-}$ is (in Kruskal gauge)

$$\chi(x^{+},x^{-})=\Omega(x^{+},x^{-}) = -\lambda^{2}x^{+}\left[x^{-}+\frac{1}{\lambda^{2}}P_{+}(x^{+})\right]+\frac{1}{\lambda}M(x^{+}) -\frac{1}{4}\ln[-4\lambda^{2}x^{+}x^{-}]\ , \tag{26}$$

where

$$M(x^{+})=\lambda\int^{x^{+}}d\bar{x}^{+}\bar{x}^{+}T^{f}_{++}(\bar{x}^{+})\ , \tag{27}$$

$$P_{+}(x^{+})=\int^{x^{+}}d\bar{x}^{+}T^{f}_{++}(\bar{x}^{+})\ . \tag{28}$$

(We have chosen the origin of the Kruskal coordinate system so as to remove possible terms linear in x^{+} and x^{-}.) Here $P_{+}(x_{+})$ is the total "Kruskal momentum" that has flowed in from $\mathcal{I}^{-}$ up to advanced time x^{+}. If we express M in terms of the energy momentum in the σ gauge,

$$\mathcal{E}(\sigma^{+})=T^{f}_{++}(\sigma^{+})\ , \tag{29}$$

and recall that the σ coordinates coincide with inertial coordinates on $\mathcal{I}^{-}$, we see that

$$M(x^{+})=\int^{\sigma^{+}}d\bar{\sigma}\ \mathcal{E}(\bar{\sigma}^{+}) \tag{30}$$

is the total "energy at infinity" that has flowed in from $\mathcal{I}^{-}$ up to advanced time x^{+}.

If the incoming energy flux $\mathcal{E}(\sigma^{+})$ satisfies suitable conditions (described below), this solution describes a black hole that forms and evaporates. To make sense of this statement, we must explain what is meant by a "black hole" in this two-dimensional model. Since, in four-dimensional dilaton gravity, Ω plays the role of the area of a two-sphere (as defined by the canonical metric) we refer to the points with $\partial_{+}\Omega<0$ and $\partial_{-}\Omega<0$ as "trapped points"; the "area" necessarily decreases in the forward light cone of these points. The boundary of the region of trapped points, where $\partial_{+}\Omega=0$, is the apparent horizon of a black hole. From a two-dimensional viewpoint, the significance of the apparent horizon is that Ω^{-1} is a coupling constant that controls the higher-order quantum corrections in the model. Thus, observers inside the apparent horizon are ineluctably drawn more deeply into the strong-coupling region of the spacetime (at least for a while).

Viewed as a function of ϕ,Ω has a minimum at

$$\phi_{\rm cr}=-\frac{1}{2}\ln\frac{N}{48}\ ,\quad \Omega_{\rm cr}=\frac{1}{4}\ . \tag{31}$$

There is no real value of ϕ corresponding to $\Omega<\Omega_{\rm cr}$. This singular behavior occurs deep inside the strong-coupling region, where a semiclassical analysis is no longer trustworthy. Nevertheless, RST suggested that a simple "phenomenological" description of this strong-coupling physics might be possible. They advocated that $\Omega=\Omega_{\rm cr}$ should be regarded as the analogue of the origin of radial coordinates; it is a boundary of spacetime, and one should not continue to negative "radius." Instead, as long as the boundary is timelike, reflecting boundary conditions (consistent with energy conservation) can be im-

posed. Thus, RST propose

$$f_i|_{\Omega=\Omega_{cr}}=0\ . \tag{32}$$

RST also imposed boundary conditions on Ω. Using these boundary conditions, one can determine the dynamical motion of the line $\Omega=\Omega_{cr}$ in the (x^+,x^-) plane. However, it turns out to be a delicate matter to impose quantum-mechanically consistent boundary conditions. Fully consistent boundary conditions will be discussed in Ref. [36], but we need not be concerned with such subtleties in this paper.

If the energy flux $\mathcal{E}$ of the incoming matter is at all times less than the critical flux $\mathcal{E}_{cr}=\frac{1}{4}\lambda^2$, then the boundary remains timelike, and the incoming matter is benignly reflected to future null infinity $\mathcal{I}^+$ without any "loss of information." However, when $\mathcal{E}$ exceeds $\mathcal{E}_{cr}$, an apparent horizon appears and a black hole forms. Furthermore, behind the apparent horizon, the boundary becomes spacelike, and the scalar curvature R diverges on the spacelike portion of the boundary. It is no longer sensible to impose boundary conditions on the fields when the boundary becomes spacelike. Figure 1 depicts the spacetime of a black hole that forms from an initial incoming pulse of matter. After it forms, the black hole emits Hawking radiation, and the apparent horizon recedes along a timelike trajectory. The global event horizon is the boundary of the region in which all forward-directed timelike and null trajectories eventually meet the spacelike singularity. Of course, this singularity occurs deep within the strongly coupled region, and so might be absent in the full quantum theory. But observers inside the global event horizon are inevitably drawn to the strongly-coupled region where semiclassical methods are inapplicable.

If the value Ω at the global horizon is large when the black hole first forms, then semiclassical methods can be reliably used to analyze the evolution of the geometry and of the quantum matter fields *outside* the global horizon. This remains true until just before the apparent horizon meets the singularity at the "end point" shown in Fig. 1(a). The behavior of the spacetime in the future of this end point cannot be unambiguously predicted using semiclassical methods. RST argued that, after the end point, the boundary of the spacetime is again timelike, the matter fields again obey the boundary condition Eq. (32), and the quantum state of the matter fields returns to the vacuum state. In their scenario, information about the quantum-mechanical state of the original incoming matter is forever lost to asymptotic observers. For most of our analysis of the evolving black hole we need not enter into speculation about what happens beyond the end point. It will suffice to analyze the quantum state of the matter fields outside the horizon, without leaving the domain of validity of semiclassical methods.

It will sometimes be convenient to consider an incoming quantum state that is a coherent state built on the Kruskal vacuum state. Then $t_\pm$ in Eq. (15) vanish, and the general solution, in Kruskal gauge, is

$$\chi(x^+,x^-)=\Omega(x^+,x^-)$$
$$=-\lambda^2x^+\left[x^-+\frac{1}{\lambda^2}P_+(x^+)\right]+\frac{1}{\lambda}M(x^+)\ , \tag{33}$$

with M and P again given by Eqs. (27) and (28). The (static) vacuum solution with $P=0$ and constant M describes a black hole in equilibrium with a thermal radia-

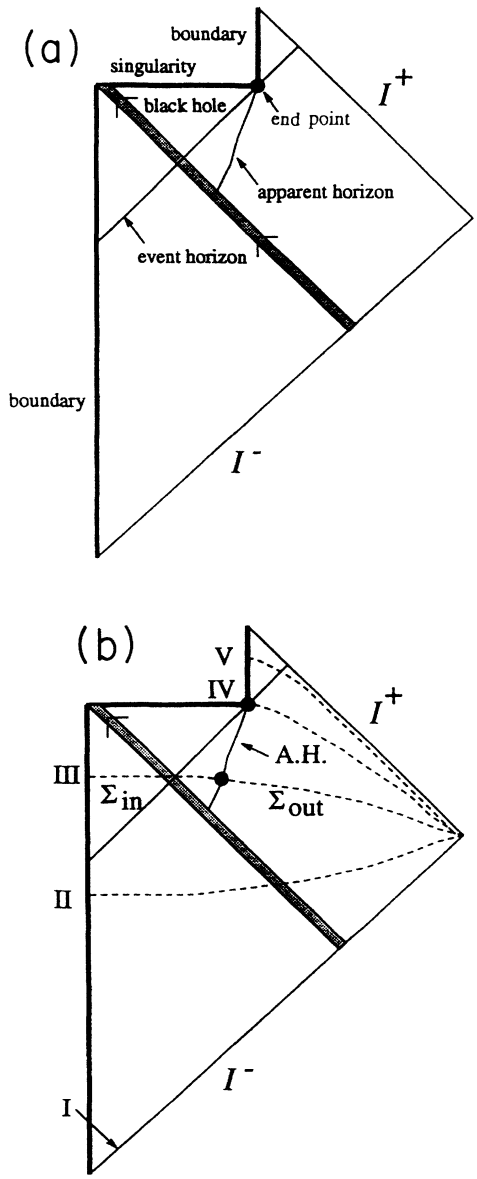

FIG. 1. (a) The two-dimensional spacetime of a black hole that forms due to the collapse of a shock wave, and then evaporates completely. After the black hole forms, the apparent horizon recedes along a timelike trajectory, eventually meeting the singularity at the "end point." The timelike boundary and the spacelike singularity are in the strongly coupled region. RST boundary conditions are imposed where the boundary is timelike. (b) Five spacelike slices through the spacetime, referred to in the text.

tion bath. Calculating the energy-momentum tensor of the matter fields in the asymptotic region we find that the incoming and outgoing energy flux are both given by $\mathcal{E}_{cr}$. From the normalization condition Eq. (16) we see that this corresponds to the conventionally normalized flux $N\lambda^2/48\pi$, which is the thermal flux for N scalar fields at temperature $T=\lambda/2\pi$. Thus, we see that back reaction effects do not modify the black hole temperature, to leading order in $1/N$.

The semiclassical field equations enable us to determine the evolution of the expectation values of Ω, χ, and the f_i's from specified initial conditions (though of course we must fix the gauge to determine χ). However, in our analysis of black hole thermodynamics we will need to keep track of the entropy of the matter fields outside the apparent horizon of the black hole. For this purpose, it is not sufficient to know expectation values; we must know the quantum states themselves.

Fluctuations of the energy-momentum tensor about its mean value will induce correlations between the quantum state of the matter and the quantum state of the dilaton field and of the geometry. Fortunately, this entanglement of the state of the matter with the state of the geometry is subdominant in the large-N limit and can be neglected to leading order. Thus, the large-N limit drastically simplifies the evolution of the quantum states. To leading order in $1/N$ we may regard the geometry and the dilaton field as a classical background, dynamically determined by the expectation value of the energy-momentum tensor, as prescribed by the semiclassical equations. Evolving the coherent state of a free massless scalar field on this background is easy; we need only choose the mean value f_i^c in Eq. (18) to be a solution to the classical field equation.

The quantum states also depend on the position of the boundary through the boundary condition Eq. (32). If the incoming energy flux never exceeds $\mathcal{E}_{cr}$, then the boundary remains timelike, and the incoming matter is reflected off the boundary to $\mathcal{I}^+$. Knowing the geometry and the dynamically determined trajectory of the boundary we can perform a Bogolubov transformation and express the reflected state in terms of Fock space states built on the inertial vacuum at $\mathcal{I}^+$. (The state $|f^c,\sigma\rangle$ will not, in general, be a simple coherent state in this natural asymptotic Fock basis on $\mathcal{I}^+$.) Thus, we can compute a unitary S matrix that relates the incoming and outgoing quantum states.

If the incoming energy flux ever exceeds $\mathcal{E}_{cr}$, then a black hole forms, and the boundary becomes spacelike. Nevertheless, we can determine the quantum state on a slice [such as slice III in Fig. 1(b)] that penetrates inside the black hole but avoids the spacelike singularity. To do so we must again know the dynamically determined trajectory of the boundary. But in our calculations in this paper we will make the simplifying assumption that no incoming matter meets the boundary before the global event horizon. The trajectory $x_B^-(x_B^+)$ of the boundary outside the global horizon is then determined by setting $\Omega=\Omega_{cr}=\frac{1}{4}$ in the vacuum solution Eq. (24); we find (in Kruskal coordinates)

$$x_B^+ x_B^- = -\frac{1}{4\lambda^2} . \tag{34}$$

From this boundary trajectory and the semiclassically determined geometry, the quantum state outside the global horizon can be completely determined to leading order in $1/N$. Our assumption that no matter meets the boundary before the global horizon not only simplifies our calculations; it also enables us to obtain results that are insensitive to any ambiguities concerning the proper choice of the boundary conditions satisfied by Ω.

In principle, we could carry out the Bogolubov transformation and express the outgoing quantum state in terms of the natural outgoing Fock basis. We will see in Sec. III, however, that the detailed form of this Bogolubov transformation will not be needed in our calculation of the entropy of the quantum state outside the apparent horizon of the black hole.

III. FINE-GRAINED ENTROPY

In our analysis of the formation and evaporation of a black hole in the RST model we will need to study the density matrix for the quantized matter fields outside the apparent horizon of the black hole. For a specified quantum state of the matter fields, this density matrix ρ is obtained by tracing over the field degrees of freedom behind the horizon. In this section we will derive a formula for the "fine-grained entropy" $S_{FG}=-\mathrm{tr}\rho\ln\rho$ of this density matrix. We will assume that the matter fields are free massless scalar fields.

Our derivation will proceed in several steps. First, we will consider a flat two-dimensional spacetime, and suppose that the quantum state is the Minkowski vacuum. We imagine that a finite spatial region R is inaccessible to an observer. The information accessible to this observer can therefore be encoded in a density matrix ρ that is obtained by tracing over the field degrees of freedom inside region R. We will calculate the entropy of this density matrix. (Our analytic formula for the entropy agrees with a numerical calculation by Srednicki [37]. This formula was obtained earlier by Holzhey [13], whose methods we follow closely.) We then proceed to generalize the entropy formula to more general "vacuum" states, and to curved spacetime.

In the RST model, scalar field modes are reflected by the boundary of the spacetime; this reflection induces correlations between left-moving and right-moving modes, which must be taken into account in the computation of the entropy. Thus, we consider a spacetime with a moving mirror, and derive a formula for the entropy of the density matrix that is obtained by tracing over a region that contains the mirror, when the quantum fields are in a "vacuum" state. The curved-spacetime generalization of this formula can be directly applied to the RST model.

Finally, in Appendix A, we consider more general quantum states, namely, coherent states built upon a specified "vacuum." We show (somewhat surprisingly) that the fine-grained entropy for any such coherent state takes the same value as for the corresponding "vacuum."

Thus, the quantum fields inside and outside of region R are no more entangled in an arbitrary coherent state than in the vacuum.

A. Minkowski vacuum

We begin with the case of the Minkowski vacuum in flat two-dimensional spacetime. Let us imagine that the only observables that we can measure have support outside of a finite spatial region R. In the vacuum state, the fields inside R are correlated with the fields outside R. Thus, even though the state of the whole system is pure, the density matrix ρ obtained by tracing over the inaccessible degrees of freedom inside R is mixed. We wish to calculate the entropy

$$S_{\rm FG} = -\mathrm{tr}\rho \ln\rho \tag{35}$$

of this density matrix, which we will refer to as the fine-grained entropy of the state outside R. Note that we could just as well imagine that we are able to measure only observables *inside* R. The two density matrices obtained by tracing over degrees of freedom inside or outside the region have the same nonzero eigenvalues, and hence the same entropy.

For massless free fields in two dimensions, the right-moving and left-moving modes are uncoupled, so it is sufficient to consider, say, the right movers alone. It is convenient to use the null coordinates

$$U = t - x, \quad V = t + x \ ; \tag{36}$$

for the right movers, we may specify the region R as the interval $[U_1, U_2]$ in null coordinates. To proceed with the entropy calculation we must construct a complete set of (right-moving) modes localized inside this interval, and a complete set of modes localized outside. Then we must decompose the Minkowski vacuum state in a basis consisting of states that are tensor products of states localized inside with states localized outside. Finally, we trace over the degrees of freedom outside R to obtain $\rho_{\rm inside}$, and compute $S_{\rm FG}$.

This seems a daunting task at first but upon reflection we recognize that we already know how to do the calculation when the region R is the half line. The right-moving modes with $U<0$ are those that are accessible to a (Rindler) observer who accelerates uniformly to the right. (See Fig. 2.) The density matrix seen by the Rindler observer was computed long ago by Unruh [12]. We need only generalize Unruh's calculation to the case where the region R is the finite interval $U_1 \le U \le U_2$ rather than the half line $U<0$.

First we briefly recall Unruh's reasoning. The entropy does not depend on the bases that we use for the modes that are localized in $U<0$ and $U>0$, so we are free to choose these bases in any convenient way that simplifies the calculation. Unruh introduces Rindler coordinates u_R and u_L in the right and left Rindler wedges that are related to the Minkowski coordinates by

$$\begin{aligned} u_R &= -\ln(-U), \quad U<0 \ , \\ u_L &= -\ln(U) \ , \quad U>0 \ . \end{aligned} \tag{37}$$

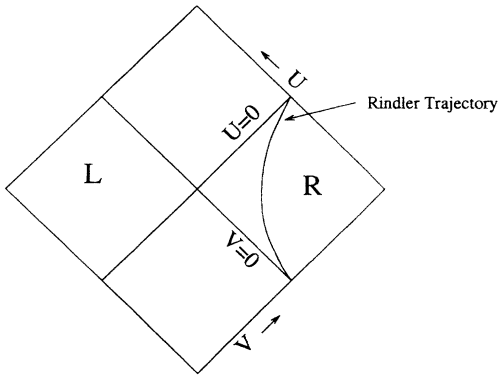

FIG. 2. Rindler spacetime. The "right wedge," with $U<0$ and $V>0$, is accessible to a "Rindler observer" that accelerates uniformly to the right. The "left wedge," with $U>0$ and $V<0$, is accessible to an observer that accelerates uniformly to the left.

The Rindler time defined by this transformation actually runs backward in the left wedge. Therefore, the modes

$$\begin{aligned} \phi_{R,\omega} &= \theta(-U) e^{-i\omega u_R} \ , \\ \phi_{L,\omega} &= \theta(U) e^{i\omega u_L} \ , \end{aligned} \tag{38}$$

($\omega>0$) are positive frequency modes with respect to Rindler time in the right and left wedges, respectively. Since the coordinate u_R covers the right wedge $U<0$ as u_R varies from $-\infty$ to ∞, arbitrary wave packets constructed from the modes $\phi_{R,\omega}$ are localized in the right wedge; similarly, wave packets constructed from the modes $\phi_{L,\omega}$ are localized in the left wedge.

If we choose as our basis these modes that have definite frequency with respect to Rindler time, then, as Unruh noted [12], it is easy to derive the Bogolubov coefficients that relate these modes to the modes that have positive frequency with respect to Minkowski time. We need only recall that a superposition of modes that are positive frequency with respect to Minkowski time will be an analytic function of the Minkowski null coordinate U in the lower U half plane. Thus, by analytically continuing the mode $\phi_{R\omega}$ to the left wedge, through the lower U half plane we obtain

$$\phi_{1,\omega} = N_\omega (\phi_{R,\omega} + e^{-\pi\omega} \phi^*_{L,\omega}) \ . \tag{39}$$

This combination of a positive frequency mode (with respect to Rindler time) in the right wedge and a negative frequency mode in the left wedge is a superposition of modes that have strictly positive frequency with respect to Minkowski time; N_ω is a normalization factor. Similarly, the combination

$$\phi_{2,\omega} = N_\omega (\phi_{L,\omega} + e^{-\pi\omega} \phi^*_{R,\omega}) \tag{40}$$

is also positive frequency with respect to Minkowski time.

Using the Bogolubov coefficients Eqs. (39) and (40), it is straightforward to express the Minowski vacuum state $|0_M\rangle$ in terms of Rindler Fock space states. (See Appendix A.) One finds

$$|0_M\rangle=\prod_j(1-e^{-2\pi\omega_j})^{1/2}\exp(e^{-\pi\omega_j}a^\dagger_{R,j}a^\dagger_{L,j})|0_R\rangle\otimes|0_L\rangle$$
$$=\prod_j(1-e^{-2\pi\omega_j})^{1/2}\sum_{n_j=0}^{\infty}e^{-\pi\omega_j n_j}|n_j,R\rangle\otimes|n_j,L\rangle\ . \tag{41}$$

Here $|0_R\rangle$ and $|0_L\rangle$ denote the Rindler vacuum states in the right and left wedges, and $|n_j,R\rangle,|n_j,L\rangle$ are the states containing n_j quanta with Rindler frequency ω_j.

We can now trace over the degrees of freedom in the left wedge to obtain the density matrix for the state in the right wedge; it is

$$\rho_R=\mathrm{tr}_L|0_M\rangle\langle 0_M|$$
$$=\prod_j\left[(1-e^{-2\pi\omega_j})\sum_{n_j}e^{-2\pi\omega_j n_j}|n_{j,R}\rangle\langle n_{j,R}|\right]\ . \tag{42}$$

This is evidently a thermal density matrix with temperature

$$T=\frac{1}{2\pi}\ . \tag{43}$$

The temperature is dimensionless because we have chosen to express the frequencies in terms of dimensionless Rindler time. If we reexpress the frequency in terms of the proper time measured by the uniformly accelerated Rindler observers, we find that $T=a/2\pi$, where a is the proper acceleration. Thus we obtain Unruh's result [12]: a uniformly accelerated observer in the Minkowski vacuum sees a thermal bath with temperature $a/2\pi$.

In one spatial dimension, the energy density of a (right-moving) ideal gas is

$$\mathcal{E}=\int_0^\infty\frac{d\omega}{2\pi}\frac{\omega}{e^{\omega/T}-1}=\frac{\pi}{12}T^2\ , \tag{44}$$

and the entropy density is obtained from the thermodynamic relation

$$\mathcal{S}=\int_0^T\frac{d\mathcal{E}}{T}=\frac{\pi}{6}T\ . \tag{45}$$

Integrating this entropy density over the half line gives an infinite result. We can obtain a finite answer by introducing ultraviolet and infrared cutoffs; then we find the fine-grained entropy

$$S_{\mathrm{FG}}\equiv-\mathrm{tr}\rho_R\ln\rho_R$$
$$=\frac{\pi}{6}T(u_{R,\max}-u_{R,\min})$$
$$=\frac{1}{12}\ln\left[\frac{U_{\max}}{U_{\min}}\right]\ . \tag{46}$$

Of course, including the left-moving modes would result in the additional term $\frac{1}{12}\ln(V_{\max}/V_{\min})$.

The logarithmic behavior of the fine-grained entropy is a consequence of the scale invariance of the vacuum fluctuations of a massless scalar field. Field modes of all wavelengths contribute to the entanglement of the quantum state in the right wedge with the quantum state in the left wedge. To exploit the scale invariance, it is convenient to construct a basis for the modes as follows: From the modes with wave number between k_0 and $2k_0$, we construct a basis of nonoverlapping wave packets, each with width of order k_0^{-1}. Among these modes, the one wave packet that overlaps the boundary between the two regions dominates the entanglement. Now complete the basis by replacing k_0 by 2^jk_0, for all integer j. For each value of j, a single wave packet dominates the entropy; on dimensional grounds, the contribution is a pure number of order one, and because of the scale invariance, the contribution is independent of j. Summing over all modes we thus obtain an expression for the fine-grained entropy that diverges logarithmically in both the ultraviolet and the infrared. The divergent behavior of Eq. (46) as $U_{\min}$ approaches zero arises because field modes that are localized just to the right of $U=0$ are entangled with the modes that are localized just to the left of $U=0$, in the Minkowski vacuum state. In three spatial dimensions, because of the enhanced density of states, the ultraviolet divergence becomes quadratic; the entropy is proportional to the transverse area [14,37,38], and is infrared finite.

We now want to generalize Unruh's procedure to the case where the inaccessible region is a finite interval $[U_1,U_2]$ rather than the half line. (This generalization was pioneered by Holzhey [13].) Again, the key idea is that, since the entropy is basis independent, we are free to introduce bases for the modes inside and outside the interval that make the computation of the entropy easy. Following Unruh, we seek handy coordinate systems that cover the inside and outside regions, which are related to one another by analytic continuation. We will also impose an infrared cutoff by restricting the null coordinate U to the range $[-L,L]$. Thus, we introduce the coordinate

$$u(U)=\ln\left|\frac{\sin\left[\frac{(U-U_1)\pi}{2L}\right]}{\sin\left[\frac{(U_2-U)\pi}{2L}\right]}\right|\ . \tag{47}$$

Here the vertical bars denote absolute value. Equation (47) really describes two distinct coordinate systems; one coordinate, which we call u_{in}, varies from $-\infty$ to ∞ as U varies from U_1 to U_2. The other coordinate u_{out} covers the region $[-L,L]$, *excluding* the interval $[U_1,U_2]$. This coordinate u_{out} approaches ∞ as U approaches U_2 (from above), and it approaches $-\infty$ as U approaches U_1 (from below). It also satisfies

$$u_{\mathrm{out}}(U=L)=u_{\mathrm{out}}(U=-L)\ . \tag{48}$$

Thus any wave packet constructed as a function of u_{out} automatically satisfies periodic boundary conditions as a function of U on the interval $[-L,L]$. The time coordinate defined by the transformation equation (47) runs *backward* in the region outside the interval $[U_1,U_2]$.

Now the modes of definite frequency with respect to u,

$$\phi_{\text{in},\omega}=\theta(U-U_1)\theta(U_2-U)e^{-i\omega u_{\text{in}}}\ ,$$
$$\phi_{\text{out},\omega}=[\theta(U_1-U)+\theta(U-U_2)]e^{i\omega u_{\text{out}}}\ , \tag{49}$$

are analogous to the Rindler modes Eq. (38). Following Unruh, we can calculate Bogolubov coefficients by analytically continuing these modes in the lower U half plane. We thus construct the mode

$$\phi_{1,\omega}=N_\omega(\phi_{\text{in},\omega}+e^{-\pi\omega}\phi^*_{\text{out},\omega})\ ; \tag{50}$$

this is a superposition of a positive frequency inside mode and a negative frequency outside mode that is positive frequency with respect to Minkowski time. Similarly, the superposition

$$\phi_{2,\omega}=N_\omega(\phi_{\text{out},\omega}+e^{-\pi\omega}\phi^*_{\text{in},\omega}) \tag{51}$$

is also positive frequency with respect to Minkowski time.

With our choice of coordinates, the Bogolubov coefficients Eqs. (50) and (51) are of just the same form as the Bogolubov coefficients Eqs. (39) and (40) for the Rindler case. Thus, the calculation of the density matrix obtained by tracing over the degrees of freedom inside the interval $[U_1,U_2]$ proceeds exactly as before—we obtain a thermal density matrix with temperature $T=1/2\pi$. We compute the entropy by integrating the thermal entropy density over the interval. As expected, the expression for the entropy has a logarithmic ultraviolet divergence at each endpoint of the interval, arising from the entanglement of the short-wavelength field fluctuations on either side of the end point. We can regulate the calculation by excluding the contribution due to the radiation bath within (affine) distance δ_2 of the upper end point and distance δ_1 of the lower end point. Then the result becomes

$$S_{\text{FG}}\equiv-\text{tr}\rho_{\text{in}}\ln\rho_{\text{in}}=\frac{1}{12}[u_{\text{in}}(U_2-\delta_2)-u_{\text{in}}(U_1+\delta_1)]$$
$$=\frac{1}{12}\ln\left[\frac{\sin\left[\frac{(U_2-U_1-\delta_2)\pi}{2L}\right]\sin\left[\frac{(U_2-U_1-\delta_1)\pi}{2L}\right]}{\sin\left[\frac{\delta_1\pi}{2L}\right]\sin\left[\frac{\delta_2\pi}{2L}\right]}\right]. \tag{52}$$

This is our expression for the fine-grained entropy (due to right movers only) of the density matrix that is obtained by tracing over the field degrees of freedom outside the interval $[U_1,U_2]$, in the Minkowski vacuum. Note that this expression is invariant if U_2-U_1 is replaced by $2L-(U_2-U_1)$; in other words we get the same entropy if we trace over the region outside the interval as if we trace over the region inside.

If we choose $U_1=-L$ and $U_2=L$, then our interval is the whole (periodically identified) box. Thus the density matrix ρ_{in} becomes pure, and the entropy should be zero. We readily see that Eq. (52) has this property. We also note that S_{FG} has a finite limit as the size of the box gets large; the entropy is infrared finite. (But see below.) If we take the limit $L\to\infty$ with the size of the interval held fixed we obtain

$$S_{\text{FG}}=\frac{1}{12}\ln\left[\frac{(U_2-U_1)^2}{\delta_1\delta_2}\right]. \tag{53}$$

Equation (53) was first derived by Holzhey [13]. Its curved space generalization will be used repeatedly in this paper.

Equation (52) has a simple interpretation. It is just the sum of two expressions of the form Eq. (46), one associated with each end point of the interval, and with the finite length of the interval acting as an infrared cutoff. However, there is an additional contribution to the fine-grained entropy that we have not yet included—the contribution due to the $\omega=0$ mode, the mode that is constant in $[U_1,U_2]$. This contribution to the entropy is formally infinite, because the zero-frequency mode has an infinite number of accessible quantum states.

If we were doing thermodynamics on the full line, rather than a finite interval, we could argue that different values of the constant mode of the field correspond to different superselection sectors of the quantum theory. Then it would be appropriate to project out a particular value of the zero mode, if we want to restrict our attention to one particular superselection sector. (Alternatively, we could impose boundary conditions, such as fixed end or antiperiodic boundary conditions, that remove the zero mode.) The infinite zero-mode entropy is associated with the existence of an infinite number of different superselection sectors, rather than an infinite contribution to the entropy in any particular sector.

However, if we are considering the fine-grained entropy on a finite interval, we do not have the option of projecting out the zero mode, or of removing it by a particular choice of boundary conditions. There are normalizable modes that are constant in the interval $[U_1,U_2]$, and decay outside the interval. These modes make a nonnegligible contribution to the entanglement of the fields inside and outside the interval.

It turns out that this additional term in the entropy will not be relevant to our discussion of black hole thermodynamics. But it is worthwhile to note that this term can be easily estimated. Suppose that we imagine using the nonoverlapping wave packet basis described following Eq. (46). In Eq. (53) we have included the contributions to the entropy due to wave packets that are narrow compared to U_2-U_1, and that straddle either the boundary at U_1 or the boundary at U_2. What we are missing is

the contribution due to the wave packets that are wide compared to U_2-U_1, and that straddle the whole interval.

The essential insight is that these broad wave packets produce a perfect correlation between the value of the constant mode of the scalar field in the interval $[U_1,U_2]$ and the entangled state of the long wavelength modes to the left and right of the interval. Thus, our calculation of the Rindler entropy can be used to find the degree of entanglement of the constant mode in the interval with the fields outside the interval. The Minkowski vacuum state has the form

$$|0_M\rangle = |\text{short}\rangle \otimes |\text{long}\rangle \otimes |\text{uncorrelated}\rangle \ , \tag{54}$$

where $|\text{short}\rangle$ represents the product over entangled modes with wavelength less than U_2-U_1, and $|\text{long}\rangle$ is the product over the entangled modes with wavelength greater than U_2-U_1; $|\text{uncorrelated}\rangle$ denotes the product over the modes that are well localized either entirely inside the interval or entirely outside, and so do not contribute significantly to the entanglement. Crudely speaking, the long-wavelength entangled state has the form (up to normalization)

$$|\text{long}\rangle \sim \prod_{j=0}^{j_{\max}} \sum_{n_j} |n_j,R\rangle \otimes |n_j,L\rangle \otimes |n_j,\text{inside}\rangle \ . \tag{55}$$

Here the jth factor is the contribution due to a wave packet mode of width $2^j(U_2-U_1)$, centered at the interval, and n_j labels the quantum state of that mode. The field fluctuations in this mode generate correlations between the quantum state $|n_j,R\rangle$ of the portion of the wave packet localized to the right of the interval and the quantum state $|n_j,L\rangle$ of the portion of the wavepacket that is localized to the left of the interval. Furthermore, these fluctuations are perfectly correlated with the quantum state $|n_j,\text{inside}\rangle$ of the constant mode inside the interval. We see that tracing over the state of the constant mode inside the interval, to obtain a density matrix for the state outside, produces just the same density matrix as if we traced over the left region to obtain a density matrix for the right region. Thus, we can use the Rindler entropy formula Eq. (46) to estimate the long-wavelength contribution to the fine-grained entropy for a finite interval, with the size of the interval playing the role of the ultraviolet cutoff. This contribution is

$$S_{\text{FG,long}} = \frac{1}{12}\ln\left[\frac{U_{\max}}{U_2-U_1}\right] . \tag{56}$$

Now, we can find the total fine-grained entropy outside of an interval of length L on a slice of fixed time. Combining the contributions of the right movers and left movers we obtain

$$S_{\text{FG}} = \frac{1}{3}\ln\left[\frac{L}{\delta}\right] + \frac{1}{6}\ln\left[\frac{L_{\max}}{L}\right] , \tag{57}$$

where δ is the short-distance cutoff at both ends of the interval,[3] and $L_{\max}$ is an infrared cutoff. The error in Eq. (56) should be a (nonuniversal) constant of order one that can be absorbed into δ in Eq. (57). The result Eq. (57) agrees with a numerical calculation (for antiperiodic boundary conditions) that was carried out by Srednicki [37].

B. Curved spacetime

So far, we have assumed that the state of the quantum field is the Minkowski vacuum. It is easy to extend the result to the case of a more general "vacuum state" in flat spacetime. Suppose that we introduce a new null coordinate $\hat{U}(U)$, and define a vacuum relative to this new coordinate; that is, we consider the state that contains no (right-moving) quanta that are positive frequency with respect to the coordinate $\hat{U}$. The same reasoning that we used above for the Minkowski vacuum applies just as well to this case. Thus, if the size of the interval $[\hat{U}_1,\hat{U}_2]$ is small compared to the infrared cutoff, the fine-grained entropy is again given by[4]

$$S_{\text{FG}} = \frac{1}{12}\ln\left[\frac{(\hat{U}_2-\hat{U}_1)^2}{\hat{\delta}_1\hat{\delta}_2}\right] . \tag{58}$$

The only new subtlety is that the short-distance cutoffs $\hat{\delta}_{1,2}$ are here expressed in terms of the new $\hat{U}$ coordinate. We can reexpress these cutoffs in terms of the Minkowski (affine) distances $\delta_{1,2}$ using the identities

$$\hat{\delta}_1 = \hat{U}_1'\delta_1, \quad \hat{\delta}_2 = \hat{U}_2'\delta_2 \ , \tag{59}$$

where the prime denotes a derivative with respect to U. When the cutoff is expressed in terms of the inertial coordinates, the entropy becomes [13]

$$S_{\text{FG}} = \frac{1}{12}\ln\left[\frac{(\hat{U}_2-\hat{U}_1)^2}{\hat{U}_1'\hat{U}_2'\delta_1\delta_2}\right] . \tag{60}$$

At this stage let us combine together the contributions to the entropy due to the right-moving and left-moving modes. Suppose that the left moving "vacuum" state is defined relative to the coordinate $\hat{V}(V)$. We consider a spacelike slice Σ, and a region on this slice bounded on the left by the point $(\hat{U}_2,\hat{V}_2)$ and on the right by the point $(\hat{U}_1,\hat{V}_1)$, as shown in Fig. 3. Tracing over the degrees of freedom inside this region yields a total fine-grained entropy

---

[3]The distance δ is actually $(\delta_R\delta_L)^{1/2}$, where δ_R and δ_L are cutoffs for the right movers and left movers, respectively. It can be interpreted as the invariant proper length over which the ends of the interval are smoothed out on the time slice.

[4]We are again neglecting the (infrared sensitive) contribution due to the mode that is constant in the interval. The contribution of this mode to the entropy must be considered separately.

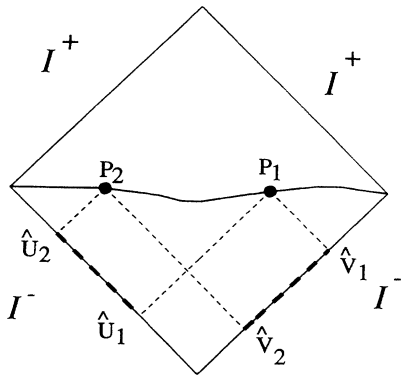

FIG. 3. A spacelike slice through flat spacetime. By tracing over the field degrees of freedom on the portion of the slice in the region between the points $P_1=(\hat{U}_1,\hat{V}_1)$ and $P_2=(\hat{U}_2,\hat{V}_2)$, we obtain a density matrix $\rho_{\rm out}$ for the fields on the portion of the slice outside that region.

$$S_{\rm FG}=\frac{1}{12}\ln\left[\frac{(\hat{U}_2-\hat{U}_1)^2}{\hat{U}_1'\hat{U}_2'\delta_{1,R}\delta_{2,R}}\right]+\frac{1}{12}\ln\left[\frac{(\hat{V}_2-\hat{V}_1)^2}{\hat{V}_1'\hat{V}_2'\delta_{1,L}\delta_{2,L}}\right], \tag{61}$$

where, e.g., $\delta_{1,R}$ denotes the short-distance cutoff, in inertial coordinates, on the wavelength of the right-moving modes at end point 1. By combining together the contributions of the right movers and the left movers we thus obtain an expression that is invariant under Lorentz boosts, for the product $\delta_R\delta_L$ of the cutoffs on the right-moving and left-moving modes is boost invariant. This quantity is just (the square of) a proper length measured on the slice Σ.

When expressed in terms of the new $(\hat{U},\hat{V})$ coordinates, the Minkowski spacetime metric $ds^2=-dU\,dV$ becomes

$$ds^2=-e^{2\rho}d\hat{U}\,d\hat{V}\ , \tag{62}$$

where

$$e^{-2\rho}=\hat{U}'\hat{V}'\ . \tag{63}$$

In terms of this metric, the expression Eq. (61) for the entropy becomes

$$S_{\rm FG}=\frac{1}{6}(\rho_1+\rho_2)+\frac{1}{12}\ln\left[\frac{(\hat{U}_2-\hat{U}_1)^2}{\delta_{1,R}\delta_{2,R}}\right]+\frac{1}{12}\ln\left[\frac{(\hat{V}_2-\hat{V}_1)^2}{\delta_{1,L}\delta_{2,L}}\right]. \tag{64}$$

This formula has the advantage that it can be applied to curved spacetime as well. In curved spacetime, there is no global inertial frame. But we are free to introduce coordinates $(\hat{U},\hat{V})$, and to consider the "vacuum" state defined by these coordinates—the state that contains no quanta that are positive frequency with respect to $\hat{U}$ and $\hat{V}$. If the spacetime metric has the form Eq. (62) in terms of these coordinates, then Eq. (64) gives the fine-grained entropy that results if we trace over the field degrees of freedom contained in a finite interval of a spacelike slice. The cutoffs in Eq. (64) are expressed in terms of the locally flat coordinates (U,V) at the end points of the interval, for which the metric takes the form $ds^2=-dU\,dV$. As noted above, the entropy is unchanged by the local Lorentz transformations that preserve this metric. Since our cutoff is in effect smeared over a region with width of order δ, it is implicit in Eq. (64) that ρ does not vary appreciably over this region.

We should also remark that, for a given "vacuum" state, the coordinates $(\hat{U},\hat{V})$ are not uniquely defined. We have the freedom to perform an SL(2,C) transformation on the coordinates without changing the vacuum. It is easy to check that Eq. (64) is SL(2,C) invariant. As expected, then, the conformal transformations that preserve the quantum state of the fields also preserve our expression for the fine-grained entropy.

Finally, we note that our expression Eq. (46) for the entropy on the half line can also be easily generalized to curved spacetime. Combining the contributions of the right movers and the left movers, and expressing the short-distance cutoffs δ_R,δ_L in terms of locally inertial coordinates at the boundary, we obtain

$$S_{\rm FG}=\frac{1}{6}\rho_P+\frac{1}{12}\ln\left[\frac{-\hat{U}_{\rm max}\hat{V}_{\rm max}}{\delta_R\delta_L}\right]. \tag{65}$$

Here, again, the vacuum is defined with respect to the $(\hat{U},\hat{V})$ coordinates, and ρ_P is the conformal factor in these coordinates at the point P that divides the space in half; $\hat{U}_{\rm max}$ and $\hat{V}_{\rm max}$ are the infrared cutoffs.

C. Moving mirror

In a space without a boundary, the right-moving and left-moving modes of a free massless scalar field are completely uncoupled, and the quantum states of the right movers and left movers can be regarded as independent. But if spacetime has a reflecting boundary (as in the RST model) then correlations between the right-moving and left-moving quantum states are induced. These correlations must be taken into account in the computation of the fine-grained entropy.

Suppose, then, that space is bounded on the left by a perfectly reflecting mirror, as shown in Fig. 4. We suppose that the mirror moves on some timelike trajectory. Then we can express the quantum state of the field as a left-moving state at $\mathcal{I}^-$ (since there are no right movers at $\mathcal{I}^-$). In particular, we can introduce a null coordinate $\hat{V}$, and consider the "vacuum" state defined on $\mathcal{I}^-$ in terms of the $\hat{V}$ coordinate. Then we may define a $\hat{U}$ coordinate by demanding $\hat{U}=\hat{V}$ at the boundary, the position of the mirror.

Now consider a spacelike slice Σ, and an interval on the slice bounded by a point P_1 with coordinates $(\hat{U}_1,\hat{V}_1)$ and a point P_2 with coordinates $(\hat{U}_2,\hat{V}_2)$. As a warm-up for our analysis of black holes (where the interval will correspond to the black hole interior), we would like to trace over the field degrees of freedom inside this interval, and obtain a density matrix for the state on the slice outside the interval. The right-moving and left-moving modes in the interval are correlated. In fact, as Fig. 4

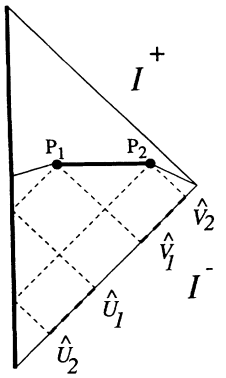

FIG. 4. A spacelike slice through the moving mirror spacetime. Coordinates have been chosen so that the trajectory of the mirror is $\hat{V}(\hat{U})=\hat{U}$. By tracing over the field degrees of freedom on the portion of the slice in the region between the points $P_1=(\hat{U}_1,\hat{V}_1)$ and $P_2=(\hat{U}_2,\hat{V}_2)$, we obtain a density matrix $\rho_{\rm out}$ for the fields on the portion of the slice outside that region.

shows, the right-moving modes in the interval are the same as the left-moving modes on $\mathcal{I}^-$, in an interval $\hat{U}_2<\hat{V}<\hat{U}_1$. Thus, tracing over the left movers and right movers inside the interval bounded by P_1 and P_2 on the slice Σ is (almost) equivalent to tracing over the left-movers only on $\mathcal{I}^-$, in the *union* of the two intervals $\hat{U}_2<\hat{V}<\hat{U}_1$ and $\hat{V}_1<\hat{V}<\hat{V}_2$. (But see the caveat below.)

Tracing over the field degrees of freedom in a union of two disjoint intervals is a bit complicated, but a simpler problem turns out to be adequate for our purposes. We consider a point P with coordinates $(\hat{U}_P,\hat{V}_P)$ on the slice Σ, and we trace over the field degrees of freedom on Σ between P and the mirror. As shown in Fig. 5(a), this is (almost) equivalent to tracing over the interval $\hat{U}_P<\hat{V}<\hat{V}_P$ on $\mathcal{I}^-$ (recalling that the $\hat{U}$ coordinate is defined by the condition that $\hat{U}=\hat{V}$ at the boundary). We may now appeal to Eq. (53) to conclude that

$$S_{\rm FG}=\frac{1}{12}\ln\left[\frac{(\hat{V}_P-\hat{U}_P)^2}{\hat{\delta}_R\hat{\delta}_L}\right]. \tag{66}$$

Here $\hat{\delta}_R$ is the short-distance cutoff on the right-moving modes at the point P, expressed in $\hat{U}$ coordinates; because of the way the $\hat{U}$ coordinate has been defined, this is the same as the cutoff at $\hat{V}=\hat{U}_P$ on $\mathcal{I}^-$, expressed in terms of $\hat{V}$ coordinates. We may introduce "locally inertial" coordinates U and V such that the metric in the vicinity of the point P is $ds^2=-dUdV$; if δ_R and δ_L are the cutoffs expressed in terms of these coordinates, then Eq. (66) becomes

$$S_{\rm FG}=\frac{1}{6}\rho_P+\frac{1}{12}\ln\left[\frac{(\hat{V}_P-\hat{U}_P)^2}{\delta_R\delta_L}\right], \tag{67}$$

where ρ_P is the conformal factor of the metric Eq. (62) at the point P. The same derivation will of course apply if we choose the $\hat{U}$ coordinate so that $\hat{V}-\hat{U}$ is a nonzero constant, except that we will now have

$$S_{\rm FG}=\frac{1}{6}\rho_P+\frac{1}{12}\ln\left[\frac{(\hat{V}_P-\hat{V}_B)^2}{\delta_R\delta_L}\right]; \tag{68}$$

here $\hat{V}_B$ is defined as the value of $\hat{V}$ at the point on the boundary that is contained in a null line through P, as shown in Fig. 5(b).

If we had imposed Neumann boundary conditions at the mirror, the model would be equivalent to a model with left movers only and no boundary. Then Eq. (66) would be the exact expression for the entropy due to the modes that are not constant on the interval between the point P and the mirror. In addition, there would be an infrared divergent contribution to the entropy of the form Eq. (56), arising from modes that are constant between P and the mirror, and decay outside of P. The situation with Dirichlet boundary conditions is a bit different. The condition that the fields vanish at the mirror removes the mode that is constant behind P, and as a result the entropy is infrared finite. To understand why the entropy of the left movers defined at $\mathcal{I}^-$ is not exactly the same as the entropy of the left movers and right movers on the spacelike slice, consider two nonoverlapping wave packet modes at $\mathcal{I}^-$, both localized inside the interval $[\hat{V}_B,\hat{V}_P]$, and both entangled with modes out-

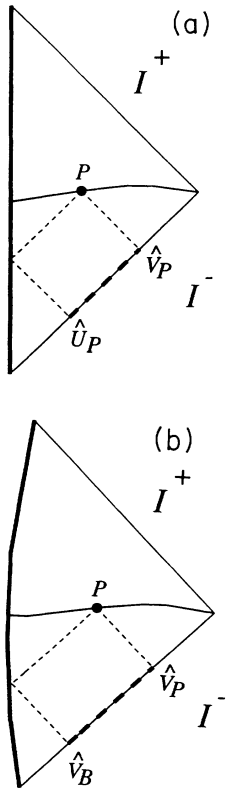

FIG. 5. (a) A spacelike slice through the moving mirror spacetime. Coordinates have been chosen so that the trajectory of the mirror is $\hat{V}(\hat{U})=\hat{U}$. By tracing over the field degrees of freedom on the portion of the slice between the point $P=(\hat{U}_P,\hat{V}_P)$ and the mirror, we obtain a density matrix $\rho_{\rm out}$ for the fields on the portion of the slice to the right of the point P. (b) If coordinates are not chosen so that $\hat{V}=\hat{U}$ at the mirror, we define V_B as the advanced time of an incoming null ray that reflects off the mirror and then passes through P.

side. Suppose that one of these wave packets reflects from the mirror prior to the slice Σ, and that the two wave packets then interfere destructively on Σ. Thus, although each of the two modes is entangled with the fields outside the interval at $\mathcal{I}^-$, their coherent sum (namely, zero) is not entangled with the fields on Σ outside of the point P.

While this error is quite small for the modes with wavelength much less than the width of the interval, it is significant for modes of long wavelength. However, on dimensional grounds, the total error in our estimate of the entropy is a constant of order one. (The error is dimensionless, and does not depend on the ultraviolet or infrared cutoffs.) This constant can be absorbed into the ultraviolet cutoff in Eqs. (66), (67), and (68).

Equation (68) is our main result for the fine-grained entropy in the moving mirror spacetime. To summarize, the quantum state is the "vacuum" defined with respect to the $\hat{V}$ coordinate on $\mathcal{I}^-$, and $S_{\rm FG}$ is the entropy of the density matrix that is obtained by tracing over the field degrees of freedom on a spacelike interval between the point P and the mirror. The $\delta_{R,L}$ are the cutoff wavelengths for left and right movers at the point P, expressed in terms of the locally inertial coordinates U, V (such that the metric at P has the form $ds^2=-dU\,dV$); ρ_P is the value of the conformal factor at the point P for the metric $ds^2=-e^{2\rho}d\hat{U}\,d\hat{V}$, where the $\hat{U}$ coordinate is defined by the condition $\hat{V}-\hat{U}=$ constant at the mirror. (It is also assumed that ρ can be regarded as constant over a region with width comparable to the cutoff length scale.) We recall that the product $\delta_R\delta_L$ (a proper length squared on the spacelike slice) is invariant under local Lorentz boosts, and that $S_{\rm FG}$ is unchanged by the SL(2,C) transformations that modify the $\hat{V}$ coordinate without altering the vacuum state. We also emphasize again that this formula for $S_{\rm FG}$ applies in curved two-dimensional spacetime, as well as in flat spacetime.

D. Black hole

The application of Eq. (68) to the RST model is immediate. In the spacetime of a black hole that forms due to infalling matter, there is a timelike boundary, up until the formation of the spacelike singularity. We consider a spacelike slice Σ (as in Fig. 6) that passes through the apparent horizon at the point P, and meets the timelike boundary behind the horizon. Let the quantum state be the vacuum defined by the coordinate σ^+—this is the state that appears to contain no quanta to the inertial observers at $\mathcal{I}^-$. Construct a density matrix $\rho_{\rm out}$ outside the apparent horizon by tracing over the field degrees of freedom behind the apparent horizon. Recalling that the RST model contains N species of free massless scalar field, Eq. (68) becomes

$$S_{\rm FG}\equiv-{\rm tr}(\rho_{\rm out}\ln\rho_{\rm out})=\frac{N}{6}\left[\rho_{H,\sigma}+\ln\left[\frac{\sigma_H^+-\sigma_B^+}{\delta}\right]\right], \tag{69}$$

where $\rho_{H,\sigma}=\rho(\sigma_H^-,\sigma_H^+)$ is the conformal factor of the metric (in σ coordinates) at the point P where the slice

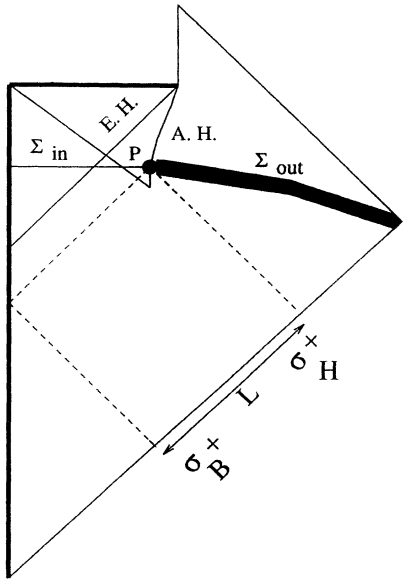

FIG. 6. A spacelike slice Σ through the black hole spacetime. The slice crosses the apparent horizon at the point $P=(\sigma_H^-,\sigma_H^+)$. We define σ_B^+ as the advanced time of an incoming null ray that reflects off the boundary and then passes through P. Incoming null rays with advanced time between σ_B^+ and σ_H^+ cross Σ inside the apparent horizon.

crosses the apparent horizon. Here σ_B^+ is the value of σ^+ at the point where the null line through P meets the boundary, as indicated in Fig. 6. The cutoff δ is the proper length $(\delta_R\delta_L)^{1/2}$; alternatively we may choose the local Lorentz frame at P so that $\delta_R=\delta_L\equiv\delta$.

Note that, in defining the conformal factor ρ in Eq. (69), we have implicitly used a σ^- coordinate that satisfies

$$\sigma^-=\sigma^+ +\text{const} \tag{70}$$

on the timelike boundary. This σ^- coordinate does not necessarily coincide with the σ^- coordinate that is defined in terms of the Kruskal coordinate x^- by Eq. (17). However, we saw in Eq. (34) that, in the linear dilaton vacuum, the σ coordinates defined by Eq. (17) satisfy

$$\lambda(\sigma_B^+-\sigma_B^-)=-2\ln 2 \tag{71}$$

at the boundary; thus, Eq. (70) is satisfied, and the two definitions of σ^- *do* agree. The same holds true if no infalling matter has reached the boundary before the advanced time $\sigma^+=\sigma_B^+$. In our analysis of the thermodynamics of a black hole formed from collapse, we will find it convenient to assume that this condition holds, so that Eqs. (70) and (17) are both valid.

Under this assumption we can reexpress Eq. (69) in terms of the value ϕ_H of the dilaton field at the apparent horizon. First, we see from Eq. (17) that the conformal factor ρ_σ in σ gauge is related to the conformal factor ρ_K in Kruskal gauge by

$$ds^2=-e^{(2\rho_\sigma)}d\sigma^+d\sigma^-=-e^{(2\rho_K)}dx^+dx^-=-e^{(2\rho_K)}e^{\lambda(\sigma^+-\sigma^-)}d\sigma^+d\sigma^-\,, \tag{72}$$

or

$$\rho_\sigma=\rho_K+\frac{\lambda}{2}(\sigma^+-\sigma^-)\ . \tag{73}$$

For the point on the boundary with the same retarded time as the apparent horizon (as in Fig. 6) we have $\sigma_B^-=\sigma_H^-$; thus, combining Eqs. (71), (13), and (31) we find that the value $\rho_{H,\sigma}$ of the conformal factor at the apparent horizon, in σ gauge, is

$$\rho_{H,\sigma}=\phi_H-\phi_{\rm cr}+\tfrac{1}{2}\lambda(\sigma_H^+-\sigma_B^+)\ . \tag{74}$$

Our expression for the fine-grained entropy outside the apparent horizon then becomes

$$S_{\rm FG}=\frac{N}{6}\left[\phi_H-\phi_{\rm cr}+\frac{1}{2}\lambda L+\ln\frac{L}{\delta}\right], \tag{75}$$

where we have defined

$$L=\sigma_H^+-\sigma_B^+\ . \tag{76}$$

Roughly speaking, L is the affine volume (in σ coordinates) behind the horizon at advanced time σ_H^+ (as shown in Fig. 6).

We derived Eqs. (69) and (75) under the assumption that the quantum state at $\mathcal{I}^-$ is the inertial vacuum. However, we will show in Appendix A that Eq. (69) and (75) still hold if the incoming state is a coherent state built on this vacuum. (Coherent states are a natural basis to use in the present context because, in the large-N limit, they are orthogonal and have a simple evolution law.) *If* we assume that the infalling matter is in a coherent state of this type, and that no incoming matter reaches the boundary prior to the global horizon, then Eq. (75) is the correct expression for the fine-grained entropy of the matter fields outside the apparent horizon of the black hole.

IV. EVAPORATION AND INFORMATION

When a black hole forms from collapsing matter, some of the information about the initial quantum state of the matter becomes encoded in the correlations of the quantum fields outside the horizon with the fields inside the horizon. This information remains inaccessible to an observer who remains outside the horizon at all times. Our expression Eq. (69) for the fine-grained entropy quantifies the amount of this missing information. Thus, by studying the behavior of $S_{\rm FG}$ as the black hole evolves, we can track the information content of the Hawking radiation that is emitted.

The simplest case to consider is that in which the black hole remains "critically illuminated" for a long time. That is, we imagine that the incoming energy flux $\mathcal{E}(\sigma^+)$ matches the outgoing thermal flux $\mathcal{E}_{\rm cr}=\frac{1}{4}\lambda^2$ due to the Hawking radiation. During the period of critical illumination, the black hole mass, and the value ϕ_H of the dilaton field at the horizon, remain unchanged. If the quantum state of the infalling matter is a coherent state built on the asymptotic inertial vacuum, we may then use Eq. (75) to find the change in the fine-grained entropy of the matter fields outside the horizon during the process; it is

$$\Delta S_{\rm FG}=\frac{N}{6}\left[\frac{\lambda}{2}(L_f-L_i)+\ln\frac{L_f}{L_i}\right], \tag{77}$$

where L_i and L_f denote the values of L at the beginning and end of the critical illumination. (Note that, though our expression for the fine-grained entropy depends on a short-distance cutoff, this entropy *change* is cutoff independent.) During critical illumination, the horizon is null and σ_B^+ is fixed, so that $dL/d\sigma_H^+=1$. It is clear then, that if the critical illumination lasts long enough, the increase in the fine-grained entropy may be as large as desired. We conclude that there is no limit to the amount of information that can be destroyed by the black hole, or in other words, no limit to the degree of entanglement of the fields outside the global horizon with those inside. It was argued in Ref. [16] that an arbitrary amount of information can be stored on a slice inside the horizon of a black hole. Eq. (77) is the other side of the coin—there is no limit to the amount of information that can be *missing* from the region outside the horizon.

Because the fine-grained entropy can increase without bound, while the black hole mass remains fixed, it is not possible to attribute the fine-grained entropy to the entanglement of the degrees of freedom outside the black hole with a *finite* number of internal degrees of freedom of the black hole. Unless there is a stable black hole remnant with an *infinite* number of degrees of freedom [7], information is unavoidably lost.

It is useful to recall the origin of the two terms in Eq. (77) by referring to Eq. (69). The value of ρ at the apparent horizon in sigma gauge is related to the dilaton field ϕ by Eq. (74), or

$$\rho_{H,\sigma}=\phi_H-\phi_{\rm cr}+\tfrac{1}{2}\lambda L\ ; \tag{78}$$

the first term in Eq. (77) is just $(N/6)(\rho_{H,f}-\rho_{H,i})$. Equation (78) expresses the familiar property that the field modes that cling near to the horizon for a long while undergo an exponential redshift. We recall that the cutoff δ is a fixed proper length at the apparent horizon. This means that the cutoff in σ coordinates at the apparent horizon is shrinking exponentially, according to

$$\delta_\sigma^2=e^{-2\rho_\sigma}\delta^2\sim e^{-\lambda L}\delta^2\ . \tag{79}$$

[In the second equality we have neglected the correction in Eq. (78) due to the evolution of ϕ.] Since the σ coordinates are the inertial coordinates on $\mathcal{I}^-$, Eq. (79) says that, as the black hole evolves, shorter and shorter wavelength incoming modes, as measured on $\mathcal{I}^-$, are being included in the calculation of the fine-grained entropy. It is the *very-short-distance* correlations between these modes just inside and just outside the horizon that are responsible for the dominant contribution to the entropy in Eq. (77). The subdominant second term in Eq. (77) arises from the *long-distance* correlations between field modes inside and outside the horizon.

It may be appropriate to be somewhat more explicit about the connection between the cutoff expressed in σ coordinates and wavelengths measured at $\mathcal{I}^-$. In our

analysis of the fine-grained entropy in Sec. III, we really imposed two cutoffs, one on left-moving modes and one on right-moving modes. In the case of the black hole background, these can both be expressed in terms of the σ^+ coordinate—we see from Fig. 6 that there is a cutoff on left movers in the vicinity of σ_B^+, and another cutoff on left movers at σ_H^+. Let us denote these two cutoffs by $\delta\sigma_H^-$ and $\delta\sigma_H^+$. (Recall that $\sigma_H^-=\sigma_B^+ +\text{const.}$) Individually, the two cutoffs have no invariant significance; it is only their product $\delta\sigma_H^-\delta\sigma_H^+=\delta_\sigma^2$ that is determined by Eq. (79).

The individual cutoffs $\delta\sigma_H^-$ and $\delta\sigma_H^+$ depend on how we choose our time slices. However, there is a natural way to foliate the spacetime with spacelike slices. We fix a position far from the black hole, by specifying a value of the dilaton field ϕ. A family of observers, with their clocks initially synchronized, fall freely toward the black hole from this fixed position at regular intervals. The natural time slices are those on which all observers record the same proper time.[5] With this choice, the cutoff $\delta\sigma_H^+$ remains essentially constant along the apparent horizon, so that the other cutoff shrinks according to

$$\delta\sigma_H^- \propto e^{-2\rho_{H,\sigma}} . \tag{80}$$

Thus, as the black hole evolves, shorter and shorter wavelengths modes, as measured on $\mathcal{I}^-$ near $\sigma^+=\sigma_B^+$, are being included in the calculation of the fine-grained entropy. It is the very-short-distance correlations between the modes localized just inside and just outside the horizon that are responsible for the increase in the entropy.

Although there is a sense in which the dominant contribution to the entropy can be attributed to very-short-distance correlations, it is not correct to say that the entropy can be very well localized near the horizon. Since the cutoff is a fixed proper length at the horizon, an observer in the vicinity of the horizon would conclude that ultra-short-distance modes (with wavelength much less than δ) make no contribution to the entropy. It is only when these modes are followed backward to $\mathcal{I}^-$, where they are enormously blueshifted, that ultra-short-distances need be considered. In fact, on a spacelike slice, most of the fine-grained entropy is due to the entanglement of fields far outside the horizon with fields that are far inside.

How secure is our conclusion that information is lost in the RST model? One potential worry is that it is a subtle task to control the fine-grained entropy in a semiclassical calculation [39]. We have attempted to do so by appealing the $1/N$ expansion, so that we can neglect the quantum fluctuations about the background geometry. However, expanding the entropy in powers of $\hbar$ (as we are attempting to do here[6]) can be a tricky business, since the $\hbar\to 0$ limit of the entropy may be highly singular. For example, knowing each matrix element of the density matrix ρ_{out} to leading order in $1/N$ may not be sufficient to determine S_{FG} to leading order, since the *size* of the matrix grows as $N\to\infty$. We believe, though, that this criticism does not apply to our calculations. We have derived an expression for the S_{FG} itself, rather than the matrix elements of ρ, that is valid to leading order in $1/N$.

A second worry [19,22] arises due to the extreme redshifting of the field modes that are responsible for the emitted Hawking radiation. If information is *not* lost, then the fine-grained entropy of the Hawking radiation can be attributed to entanglement with the internal degrees of freedom of the black hole. The number of internal degrees of freedom would presumably be given by the Bekenstein number $e^{S_{\text{BH}}}$. Therefore, to argue persuasively that information is lost, we must follow the evaporation of the black hole long enough so that the increase of S_{FG} exceeds

$$S_{\text{BH}}=\frac{M_{\text{BH}}}{T_{\text{BH}}}=\frac{2\pi M_{\text{BH}}}{\lambda} . \tag{81}$$

We thus require

$$\lambda(L_f-L_i)\sim\frac{24\pi M_{\text{BH}}}{N\lambda} \tag{82}$$

[neglecting the logarithmic term in Eq. (77)]. It follows that the quanta that are emitted during the late stages of the critical illumination process are in modes that have been redshifted (relative to their frequency at $\mathcal{I}^-$) by the factor

$$e^{2(\rho_{H,f}-\rho_{H,i})}=\exp\left[\frac{24\pi M_{\text{BH}}}{N\lambda}\right] . \tag{83}$$

In the RST model, it is understood that the incoming and outgoing energy fluxes, and the mass of the black hole, are all quantities of order N. Thus, the argument of the exponential in Eq. (83) is formally of order one in the large-N limit. Still, $24\pi M_{\text{BH}}/N\lambda$ should be large in a well-controlled semiclassical calculation, so that this redshift factor is truly enormous. Because the Hawking radiation is being emitted in modes that have *very* large energy as measured at $\mathcal{I}^-$, one may wonder whether there are correspondingly large *fluctuations* in energy momentum. If so, the response of the geometry to these fluctuations should be included when the evolution of quantum states is studied.

In fact, we are not aware of any calculation that convincingly demonstrates that these large fluctuations occur in the RST model, or that they precipitate a breakdown of semiclassical methods. At any rate, even if they do occur, their effects are systematically suppressed in the $1/N$ expansion. We can always justify neglecting the

[5]This foliation might not be globally defined on a general spacetime. However, for our purposes it is sufficient to define time slices locally in the vicinity of a particular point on the apparent horizon. Also, we note that the time slices defined by the family of freely falling observers are *not* the same as the slices of constant "σ time" $\sigma^0=\frac{1}{2}(\sigma^+ +\sigma^-)$. On the σ^0 time slices, we have $\delta\sigma_H^-=\delta\sigma_H^+\propto e^{-\rho_{H,\sigma}}$.

[6]Because $N\hbar$ is of order one, corrections higher order in $1/N$ are equivalent to corrections higher order in $\hbar$.

response of the geometry to the fluctuations of a mode that is blueshifted by the factor Eq. (83), by allowing N to be sufficiently large. For example, suppose we want the energy measured at $\mathcal{I}^-$ of a typical mode to be less than some small fraction ϵ of the mass of the black hole. The typical quantum emitted in the Hawking radiation has an energy of order λ, so that the energy measured at $\mathcal{I}^-$ is of order λ times the blueshift factor. This blueshifted energy is less than $\epsilon M_{\rm BH}$ provided that

$$N > \frac{1}{\epsilon}\left[\frac{N\lambda}{M_{\rm BH}}\right]\exp\left[\frac{24\pi M_{\rm BH}}{N\lambda}\right]. \tag{84}$$

Since $M_{\rm BH}/N\lambda$ is a quantity of order one, this condition is satisfied for N sufficiently large (although the required value of N grows exponentially with the mass of the black hole).

While we believe that the above technical objections can be answered, our discussion of "loss of information" in black hole evaporation should still include some important caveats. We can follow the evolution of a black hole[7] far enough to exclude the scenario described by Page [39], in which the fine-grained entropy begins to decrease sharply after about half of the mass has been radiated away. But we cannot follow the evolution all the way up to the end point of the evaporation process (without additional assumptions about the behavior of Planckian black holes). It remains a logical possibility, therefore, that the "lost" information is finally recovered in the very late stages of the process, when the large-N approximation breaks down. (General arguments [40,41] indicate that, in this event, the final stage would have to take an exceedingly long time.)

We also note that implicit assumptions have been made about how physics in our toy model behaves under *extreme* boosts, and these assumptions might not be appropriate in the real world. We remark again that, since the redshift factor $\exp(24\pi M_{\rm BH}/N\lambda)$ is very large, the fine-grained entropy that we have computed is dominated by the contributions due to field modes that are of extraordinarily short wavelength on $\mathcal{I}^-$. As has been emphasized by 't Hooft [19], Jacobson [20], Susskind [21], and the Verlindes [22], loss of information could conceivably be avoided if ordinary relativistic field theory ceases to apply at sufficiently short distances, so that our calculation of the fine-grained entropy is invalidated. While loss of information appears to occur in the RST model (for sufficiently large N), it might not occur in a different model with different short-distance physics.

A related point is that we have made an assumption about the nature of the cutoff that arises in the definition of the entropy. This cutoff can be regarded as the proper length over which we have smeared the boundary between the region inside the black hole and the region outside. Our procedure has been to keep this proper length fixed as the black hole evolves. This procedure is the only reasonable one we could think of, but if some justification could be found for varying the cutoff along the horizon, our conclusions would be altered.

[7]The case of (nearly) complete evaporation, as opposed to critical illumination, will be further discussed in Sec. VI C.

V. BLACK HOLE ENTROPY

In Sec. III, we derived an expression for the (fine-grained) entropy of the matter fields outside the apparent horizon of a black hole. To do black hole thermodynamics, we will also need an expression for the intrinsic entropy of the black hole. In the leading semiclassical approximation (neglecting all gravitational back reaction) it is easy to find the black hole entropy. But in our analysis of the RST model, back reaction effects of order $N\hbar$ are included, and we will need to include a next-to-leading correction to the black hole entropy. In this section we will derive this correction.

The leading semiclassical expression for the black hole entropy can be obtained using thermodynamic reasoning, given the relation between the black hole mass and the temperature of the Hawking radiation. If we imagine that the black hole is in equilibrium with a thermal radiation bath in a (small) cavity, we may regard a process in which the black hole accretes or emits an infinitesimal amount of radiation as a reversible thermodynamic process. Integrating the identity $dS = dM/T$ then determines the black hole entropy up to an additive constant.

For the black hole in two-dimensional dilaton gravity (and the four-dimensional magnetically charged dilaton black hole to which it is intimately related), the temperature $T_{\rm BH} = \lambda/2\pi$ is independent of its mass. Because the specific heat of the black hole is actually infinite, there are very large fluctuations in thermal equilibrium; the black hole mass wanders randomly [42,43]. However, in the large-N limit, these fluctuations are suppressed, and may be ignored. (The characteristic *time scale* of the fluctuations increases as $\sqrt{N}$ as N increases.) Thus, the naive thermodynamic arguments are valid. The leading expression for the entropy becomes

$$S_{\rm BH} = M_{\rm BH}/T_{\rm BH} = 2e^{-2\phi_H}, \tag{85}$$

where ϕ_H denotes the value of the dilaton field ϕ at the apparent horizon.

To go beyond this leading calculation we wish to find the correction to the relation between the $M_{\rm BH}$ and ϕ_H, for a black hole in contact with a radiation bath. However, it is not even clear how to define $M_{\rm BH}$ for a black hole surrounded by radiation—the Arnowitt-Deser-Misner (ADM) mass, for example, includes both a contribution from the black hole and a contribution from the bath. We will therefore proceed in two steps. For a black hole surrounded by radiation in a (finite) cavity, we imagine adiabatically introducing a small amount of additional left-moving matter, which eventually crosses the apparent horizon and is accreted by the black hole. The first step is to find how the accretion process changes the value of ϕ_H (or equivalently Ω_H). Using thermodynamics we can then find the relation between the change in ϕ_H and the change in the total entropy contained in the cavity.

This first step is not quite the whole story, though, be-

cause the total entropy is the sum of the entropy of the black hole and the entropy of the bath, both of which change in this process. The temperature of the bath is unchanged, but when the black hole accretes the additional matter, the apparent horizon shifts outward, concealing some of the radiation behind the apparent horizon, and thus reducing the entropy of the bath. The second step is to find how the horizon shift changes the entropy of the radiation outside the apparent horizon. Only then can we infer the relation between the change in ϕ_H and the change in $S_{\rm BH}$.

To carry out the first step of the calculation we begin by noting that, for an eternal black hole in equilibrium with a radiation bath, the quantum state of the matter fields is the Kruskal vacuum, or "Hartle-Hawking state"—there are no quanta that are positive frequency with respect to the Kruskal coordinates $x^{\pm}$ [44]. Now we recall that if we build an arbitrary coherent state of left-moving matter on this vacuum, the general solution to the field equations in Kruskal gauge has the form

$$\Omega(x^+,x^-)=-\lambda^2 x^+\left[x^-+\frac{1}{\lambda^2}P_+(x^+)\right]+\frac{1}{\lambda}M(x^+)\ , \tag{86}$$

where P_+ is the total incoming Kruskal momentum up to advanced time x^+, and M is the total mass (at infinity) of the incoming matter. (We have chosen the origin of the Kruskal coordinate system to remove possible linear terms in x^+ and x^-.) If we assume that P_+ and M are constants, then the position of the apparent horizon, determined by the condition $\partial_+\Omega=0$, is

$$x_H^-(x^+)=-\frac{1}{\lambda^2}P_+(x^+)\ , \tag{87}$$

and the value of Ω at the horizon is

$$\Omega_H=\frac{1}{\lambda}M(x^+)\ . \tag{88}$$

Therefore, if a pulse of left-moving matter that carries Kruskal momentum ΔP_+ and mass ΔM is accreted by the black hole, then the horizon shifts outward according to

$$\Delta x_H^-(x^+)=-\frac{1}{\lambda^2}\Delta P_+(x^+)\ , \tag{89}$$

and Ω at the horizon changes according to

$$\Delta\Omega_H=\frac{1}{\lambda}\Delta M\ . \tag{90}$$

We must recall, though, that the energy-momentum used in the field equations has the unconventional normalization Eq. (16). In thermodynamics we should use the conventionally normalized mass $M_{\rm conv}=(N/12\pi)M$, so that

$$\Delta\Omega_H=\frac{12\pi}{N\lambda}\Delta M_{\rm conv}\ . \tag{91}$$

Now the identity $dS=dM/T$ becomes

$$\Delta S_{\rm tot}\equiv\Delta(S_{\rm BH}+S_{\rm matter})=\frac{1}{T}\Delta M_{\rm conv}=\frac{N}{6}\Delta\Omega_H\ . \tag{92}$$

We now proceed to the second step, which is to calculate $\Delta S_{\rm matter}$, so that $\Delta S_{\rm BH}$ can be extracted from Eq. (92). To carry out this step we need a precise definition of the entropy carried by the matter outside of the apparent horizon. Our proposal will be that $S_{\rm matter}$ is given by Eq. (35)—it is the fine-grained entropy of the matter fields outside of the apparent horizon.[8] It is not *a priori* obvious that this expression for $S_{\rm matter}$ is correct or appropriate. Ordinarily, the thermodynamic entropy is a coarse-grained entropy [46]. Surely, for a pure state, $S_{\rm FG}=0$ would be a very poor estimate of the thermodynamic entropy. We are proposing that the quantum fields inside and outside the horizon are so thoroughly entangled that it is reasonable to regard the fine-grained entropy outside the horizon as the thermodynamic entropy. In any event, it is hard to think of another way to give the notion of the "entropy outside the apparent horizon" any precise meaning.

For an eternal black hole, there is no reflecting boundary; the right-moving modes and left-moving modes are uncorrelated. The fine-grained entropy is given by our curved-space generalization of the formula for the entropy on the half line. If the quantum state of N scalar fields is a coherent state built on the Kruskal vacuum, Eq. (65) becomes

$$S_{\rm FG}=\frac{N}{6}\left[\rho_{H,K}+\frac{1}{2}\ln\left[\frac{-x^+_{\rm max}x^-_{\rm max}}{\delta^2}\right]\right]\ , \tag{93}$$

where $x^-_{\rm max}$ and $x^+_{\rm max}$ are infrared cutoffs (in Kruskal coordinates) for the right movers and left movers. Of course, the conformal factor ρ is gauge dependent; the subscript K in Eq. (93) indicates that $\rho_{H,K}$ is evaluated in the Kruskal gauge.

We can check that it is reasonable to interpret $S_{\rm FG}$ as the thermodynamic entropy of the radiation bath by evaluating the infrared divergent part of Eq. (93). The Kruskal coordinates $x^{\pm}$ are related to the $\sigma^{\pm}$ coordinates (which become inertial in the asymptotic region) by Eq. (17); thus the infrared divergent term in $S_{\rm FG}$ is

$$S_{\rm FG}\sim\frac{N}{12}\lambda(\sigma^+-\sigma^-)_{\rm max}=\frac{N}{6}\lambda\sigma^1_{\rm max}\ . \tag{94}$$

We may interpret $\sigma^1_{\rm max}$ as the size L of the cavity that contains the radiation. Thus, Eq. (94) agrees with the entropy

$$S=2\frac{\pi}{6}TL \tag{95}$$

of a thermal bath at temperature $T=\lambda/2\pi$, times a factor of N for the N species. (The factor of 2 arises because

[8]The fine-grained entropy outside the black hole horizon has also been discussed recently by Frolov and Novikov [45].

both left movers and right movers contribute to the entropy of the bath.)

When the black hole accretes some incoming matter, only the ρ_H term in Eq. (93) is affected by the shift of the horizon. Furthermore, since in Kruskal gauge we have $\rho=\phi+\text{const}$, we conclude that

$$\Delta S_{\text{matter}}=\frac{N}{6}\Delta\phi_H \ . \tag{96}$$

Combining with Eq. (92), we find that

$$\Delta S_{\text{BH}}=\frac{N}{6}(\Delta\Omega_H-\Delta\phi_H) \ . \tag{97}$$

We can fix the arbitrary constant of integration by demanding that the black hole entropy reaches zero when the apparent horizon meets the singularity, or when $\phi_H=\phi_{\text{cr}}=-\frac{1}{2}\ln(N/48)$; thus, from the expression Eq. (9) for Ω in terms of ϕ, we obtain

$$S_{\text{BH}}=2e^{-2\phi_H}-\frac{N}{12}\phi_H-\frac{N}{24}-\frac{N}{24}\ln\left[\frac{N}{48}\right] \ . \tag{98}$$

This is our corrected formula for the black hole entropy.

The formula Eq. (98) for the black hole entropy has a satisfying interpretation. The action of two-dimensional dilaton gravity can be obtained by spherical reduction of the four-dimensional action for a near-extreme magnetically charged dilaton black hole. When this reduction is carried out, the area of the sphere of constant radius in four-dimensions becomes the ϕ-dependent prefactor of the Ricci scalar in the classical two-dimensional action [8]. Now in the RST model, an extra term is added to this prefactor. The modified prefactor has just the form of the black hole entropy in Eq. (98). Thus, loosely speaking, the relation $S_{\text{BH}}=\frac{1}{4}A$ is satisfied by our corrected entropy formula, but where A is the *corrected* "area" of the RST model.

It may help to clarify the nature of the correction that we have found to Eq. (4) if we restore the factors of $\hbar$ and "Newton's constant" G that have been suppressed until now. In the classical action Eq. (1), there is a factor G^{-1} multiplying the term $[R+4(\nabla\phi)^2+4\lambda^2]$, where $\hbar G$ is dimensionless. Thus the dilaton field ϕ is dimensionless, and λ^{-1} has the dimensions of length. The leading term in the black hole entropy is then

$$S_{\text{BH},0}=\frac{2\pi M_{\text{BH}}}{\hbar\lambda} \ , \tag{99}$$

and the correction is

$$S_{\text{BH},1}=-\frac{N}{12}\phi_H \ . \tag{100}$$

Relative to the leading term, then, the correction is suppressed by

$$\frac{S_{\text{BH},1}}{S_{\text{BH},0}}=-\frac{N\hbar}{24\pi}\frac{\lambda\phi_H}{M_{\text{BH}}}=\frac{N\hbar}{48\pi}\frac{\lambda}{M_{\text{BH}}}\ln\left[\frac{\pi GM_{\text{BH}}}{\lambda}\right] \ . \tag{101}$$

Thus, the correction is higher order in $\hbar$, but cannot be neglected in the large-N limit. It is also suppressed, for a very massive black hole, by the factor $\ln(M_{\text{BH}})/M_{\text{BH}}$.

In the RST model, it is possible to obtain a simple analytic expression for the value Ω_H of Ω at the apparent horizon, on a general time-dependent background. There is not such simple expression for ϕ_H, as Ω and ϕ are related by the transcendental Eq. (9). Thus, we cannot write down an analytic formula for S_{BH} or S_{FG} on a general background. However, when these quantities are added together, a notable simplification occurs. Combining Eqs. (98) and (69), and comparing with Eq. (10), we see that

$$S_{\text{BH}}+S_{\text{FG}}=\frac{N}{6}\left[\chi_{H,\sigma}-\frac{1}{4}+\ln 2+\ln\frac{L}{\delta}\right] \tag{102}$$

can be expressed in terms of χ, which obeys a simple field equation in the RST model. [Here $L=\sigma_H^+-\sigma_B^+$, as in Eq. (76).] Of course, the value χ_H of χ at the apparent horizon is gauge dependent, while $S_{\text{BH}}+S_{\text{FG}}$ is not. This formula is valid if χ_H is evaluated in the same coordinate system used to define the vacuum, in other words, in the "σ gauge." Recall that σ^+ is the null coordinate with respect to which the incoming vacuum state is defined, and σ^- must be chosen so that $\sigma^+-\sigma^-=\text{const}$ at the boundary of the spacetime, as we explained in Sec. III. We should emphasize that Eq. (102) is a general formula that applies under the above conditions. In particular, it need not be assumed that the $\sigma^\pm$ coordinates are related to the Kruskal coordinates by Eq. (17).

We will discuss the evolution of $S_{\text{BH}}+S_{\text{FG}}$ in the next section. For now, we remark that Eq. (102), like Eq. (98), has an intriguing interpretation. We observe that χ is proportional to the coefficient of the scalar curvature R in the *quantum-corrected* effective action of the (large-N) RST model. Thus, if we neglect the logarithmic term in Eq. (102), we find that the sum $S_{\text{BH}}+S_{\text{FG}}$ is related to the quantum-corrected Newton's constant just as S_{BH} is related to the classical Newton's constant of the model. This remark makes contact with the observations in Ref. [25], where a connection between entropy and the renormalization of Newton's constant is proposed.

[One is tempted to go further, and regard Eq. (102) as a hint that the proper way to define the fine-grained entropy is to use the "χ metric" $ds^2=-e^{2\chi_\sigma}d\sigma^+d\sigma^-$ when implementing the short-distance cutoff. Then Eq. (102) could be interpreted as wholly due to the entropy of entanglement between the regions outside and inside the black hole—there would be no need to add in a separate Bekenstein-Hawking term.]

VI. EVAPORATION AND THERMODYNAMICS

Equipped now with our formulas for the black hole entropy S_{BH} and the fine-grained entropy S_{FG} outside the apparent horizon, we are prepared to study the thermodynamics of a process in which a black hole forms from infalling matter and then evaporates, as in Fig. 1. We wish to find the time dependence of the total entropy in this process. We will assume that the incoming matter state is a coherent state built on the inertial σ^+ vacuum at $\mathcal{I}^-$. For such states we know how to evolve the

geometry using the RST equations, and we know how to calculate the fine-grained entropy. We will also make the further assumption that none of the infalling matter reaches the reflecting boundary of the spacetime before the appearance of the global event horizon. This assumption simplifies the calculation of $S_{\rm FG}$, as we explained in Sec. III.

In their analysis of the model, RST noted that the boundary condition Eq. (32) can be reimposed at the endpoint of black hole evaporation (when the singularity meets the apparent horizon), and that the final quantum state can be chosen to be the vacuum. This prescription results in the emission of a thunderpop. Furthermore, the information about the quantum state of the initial incoming matter is lost *by assumption*. But we wish to emphasize that the time dependence of the entropy up until the apparent horizon meets the singularity is insensitive to the RST prescription for continuing past this point, and is not affected by the thunderpop. It will be of interest to see how the fine-grained entropy outside the horizon behaves as the black hole approaches its demise.

We have seen that the fine-grained entropy depends on an arbitrary ultraviolet cutoff. However, the ultraviolet divergence is logarithmic, and the cutoff-dependent term is a time-independent additive constant. Thus, the sensitivity to the cutoff does not prevent us from making definite statements about how the entropy outside the black hole *changes* during its evolution, or about the change in the intrinsic entropy of the black hole itself.

A. Boltzmann entropy

In the previous section we argued that, in the Hartle-Hawking vacuum state, the fine-grained entropy $S_{\rm FG}$ could be regarded as the thermodynamic entropy outside the event horizon of the black hole. But for the black hole formed from infalling matter, this assignment must be modified. To see why, cover the spacetime of Fig. 1(a) with a sequence of spacelike slices, as depicted in Fig. 1(b). Slices I and II in the figure represent times prior to the formation of the black hole. Since there is no apparent horizon, the quantum state "outside" the horizon on these slices is a pure coherent state, which has $S_{\rm FG}=0$.

But even though the incoming matter is in a pure state, it surely carries thermodynamic entropy. We can assign a nonzero entropy to this state by performing a coarse-graining procedure. Our coherent state carries the *left-moving* energy density

$$\mathcal{E}(\sigma^+)\equiv T^f_{++}(\sigma^+)\ . \tag{103}$$

We may regard $\mathcal{E}$ as a measurable macroscopic quantity. Given the energy-density profile $\mathcal{E}$ of the incoming state, we assign an entropy by counting the number of microscopic quantum states with this energy profile—the entropy is the logarithm of the number of states. We will refer to the entropy defined by this procedure as $S_{\rm Boltz}$, the Boltzmann entropy of the incoming coherent state.

The spacetime is asymptotically flat, so we may use standard flat-space thermodynamics on $\mathcal{I}^-$. We may then appeal to the equivalence of the microcanonical and canonical ensembles in the thermodynamic limit, and express both the entropy density and the energy density in terms of a locally measured temperature. Fluctuations of the entropy and energy densities about these values are suppressed in the large-N limit. If the energy density is conventionally normalized, we can express the energy density $\mathcal{E}_{\rm conv}$ and entropy density $\mathcal{S}$ for N left-moving massless free scalar fields in terms of the temperature T as

$$\mathcal{E}_{\rm conv}=N\frac{\pi}{12}T^2\ ,\quad \mathcal{S}=N\frac{\pi}{6}T\ , \tag{104}$$

so that the entropy and energy densities are related by

$$\mathcal{S}=\left[\frac{\pi}{3}N\mathcal{E}_{\rm conv}\right]^{1/2}\ . \tag{105}$$

The energy density in Eq. (103) has the unconventional normalization

$$\mathcal{E}=\frac{12\pi}{N}\mathcal{E}_{\rm conv}\ , \tag{106}$$

so that the Boltzmann entropy can be written

$$S_{\rm Boltz}=\frac{N}{6}\int_{\mathcal{I}^-}d\sigma^+\sqrt{\mathcal{E}(\sigma^+)}\ . \tag{107}$$

We can now evolve the incoming matter state from slice I of Fig. 1(b) to slice II, which is still prior to the formation of the black hole. In general, $S_{\rm Boltz}$ can change under unitary evolution, but for a free field it is invariant as a consequence of the curved space generalization of Liouville's theorem [47]. In the present context, this is simply the statement that the energy profile $\mathcal{E}(\sigma^+)$ is unchanged.

The black hole is finally encountered on slice III. Liouville's theorem continues to apply here, so that $S_{\rm Boltz}$ is still unchanged. However, we are interested in the entropy of the matter outside the black hole. Therefore, we divide slice III into two segments, $\Sigma_{\rm in}$ and $\Sigma_{\rm out}$, inside and outside the apparent horizon. The Boltzmann entropy $S_{\rm BO}$ outside the apparent horizon is

$$S_{\rm BO}=\frac{N}{6}\int_{\Sigma_{\rm out}}d\sigma^+\sqrt{\mathcal{E}(\sigma^+)}\ . \tag{108}$$

In defining $S_{\rm BO}$ we have chosen to divide the slice Σ at the *apparent* horizon. We made the same choice when we defined the fine-grained entropy $S_{\rm FG}$ outside the black hole in Sec. III B. Furthermore, our formula Eq. (98) for the black hole entropy $S_{\rm BH}$ has been expressed in terms of the value of the dilaton field at the apparent horizon. These choices deserve some explanation. If we are adopting the viewpoint of an observer who remains outside the black hole, it may seem more logical to divide the slice at the *global* event horizon instead. After all, it is possible for the observer to cross the apparent horizon (very carefully) and return to tell about it. However, we find it more appropriate to define $S_{\rm BO}$, $S_{\rm FG}$, and $S_{\rm BH}$ using the apparent horizon, for several reasons. First of all, the position of the apparent horizon can be determined locally in time, without any required information about the global properties of the spacetime. Our observer on a time

slice can readily identify the apparent horizon as the location where $\partial_+\Omega$ vanishes. Second, because the position of the apparent horizon is determined by this local condition, it is easy to compute the trajectory of the apparent horizon using the RST equations. Third, if we use the global horizon to define the entropy, the resulting expressions do not seem to have a nice thermodynamic interpretation. In particular, the would-be second law is easily violated by sending in a very sharp pulse with large entropy and energy density but small total entropy and energy. The essential point is that the value of the dilaton at the global horizon responds less sensitively to the incoming pulse than does the dilaton at the apparent horizon.

B. Total entropy

Once the black hole forms, matter entropy can become concealed behind the horizon, and the left-moving Boltzmann entropy Eq. (108) can decrease. If physics perceived by an observer outside the black hole is to respect the second law of thermodynamics, then (as Bekenstein argued [2]) we must attribute entropy to the black hole. Furthermore, we must not neglect the entropy carried by the outgoing Hawking radiation.

We propose to adopt, as our definition of the total thermodynamic entropy

$$S_{\rm tot}\equiv S_{\rm BH}+S_{\rm BO}+S_{\rm FG}\ . \tag{109}$$

The fine-grained entropy $S_{\rm FG}$ outside the apparent horizon is dominated by the entanglement of the right-moving modes outside the horizon with the right-moving modes just inside the horizon. It roughly corresponds to the thermodynamic entropy of the outgoing Hawking radiation, while $S_{\rm BO}$ is the entropy of the incoming matter. We have seen that the fine-grained entropy does not include the entropy of the incoming matter—an incoming coherent state has the same $S_{\rm FG}$ as the vacuum state—so $S_{\rm BO}$ must be added on.

While the expression Eq. (109) may appear (and indeed, is) somewhat strange, we believe it to be a precise two-dimensional analogue of the notion of "total entropy" used implicitly in discussions of four-dimensional black hole thermodynamics. This prescription might be interpreted as follows. We may consider, instead of a pure initial state, the mixed initial state ρ that maximizes $-{\rm tr}\rho\ln\rho$, subject to the constraint that the energy density is given by the specified function $\mathcal{E}(\sigma^+)$. For this mixed initial state we have $S_{\rm Boltz}=-{\rm tr}\rho\ln\rho$. What we are adding to $S_{\rm BH}$ in Eq. (109) is the fine-grained entropy outside the horizon for this particular mixed initial state.[9] In any event we have not been able to find any other reasonable and precise alternative to Eq. (109) that obeys a generalized second law.

As we noted at the end of Sec. V the sum $S_{\rm BH}+S_{\rm FG}$ can be expressed in terms of the field χ at the apparent horizon (in σ gauge), for which we can find an analytic expression. Alternatively, we may combine Eqs. (98) and (75), to obtain directly an expression in terms of the gauge invariant quantity Ω_H. We obtain

$$S_{\rm BH}+S_{\rm FG}=\frac{N}{6}\left[\Omega_H-\frac{1}{4}+\frac{1}{2}\lambda L+\ln\frac{L}{\delta}\right], \tag{110}$$

where $L=\sigma_H^+-\sigma_B^+$, as in Eq. (76). Now we may use the general solution Eq. (26) to the field equations in Kruskal gauge, which applies if the state of the matter is a coherent state built on the sigma vacuum. Recalling that the apparent horizon is defined by the condition $\partial_+\Omega=0$, we deduce from Eq. (26) that

$$\begin{aligned}\Omega_H&=\frac{1}{4}+\frac{1}{\lambda}M(x_H^+)-\frac{1}{4}\ln(-4\lambda^2x_H^+x_H^-)\\&=\frac{1}{4}+\frac{1}{\lambda}M(\sigma_H^+)-\frac{1}{4}\lambda(\sigma_H^+-\sigma_H^-)-\frac{1}{2}\ln 2\ ,\end{aligned} \tag{111}$$

where

$$M(\sigma_H^+)=\int^{\sigma_H^+}d\sigma^+\mathcal{E}(\sigma^+) \tag{112}$$

is the total mass flowing in from $\mathcal{I}^-$ up until advanced time σ_H^+.

Next, we express Ω_H in terms of the quantity $L=\sigma_H^+-\sigma_B^+$. Under the assumption that there is no infalling matter up until advanced time x_B^+, the position of the boundary defined by $\Omega=\Omega_{\rm cr}=\frac{1}{4}$ is given by Eq. (71). For the point on the boundary with the same retarded time as the apparent horizon (as in Fig. 6), we have $\sigma_B^-=\sigma_H^-$. Combining Eq. (71) with (111) we find

$$\Omega_H=\frac{1}{4}+\frac{1}{\lambda}M-\frac{1}{4}\lambda L\ . \tag{113}$$

Inserting into Eq. (110) now yields

$$S_{\rm BH}+S_{\rm FG}=\frac{N}{6}\left[\frac{1}{\lambda}M(\sigma_H^+)+\frac{1}{4}\lambda L+\ln\frac{L}{\delta}\right]. \tag{114}$$

Adding the Boltzmann entropy Eq. (108) outside the black hole we find

$$\begin{aligned}S_{\rm tot}&\equiv S_{\rm BH}+S_{\rm FG}+S_{\rm BO}\\&=\frac{N}{6}\left[\frac{1}{\lambda}M(\sigma_H^+)+\frac{1}{4}\lambda L+\ln\frac{L}{\delta}\right.\\&\qquad\left.+\int_{\sigma_H^+}^{\infty}d\sigma^+\sqrt{\mathcal{E}(\sigma^+)}\right],\end{aligned} \tag{115}$$

our final expression for the total entropy.

It is instruction to compare $S_{\rm tot}$ and $S_{\rm Boltz}$ on the same time slice, or equivalently, to compare $S_{\rm BH}+S_{\rm FG}$ with the Boltzmann entropy $S_{\rm BI}$ *inside* the apparent horizon. Since we assume that there is no incoming energy density before the advanced time $\sigma^+=\sigma_B^+$, we can choose the lower limit of integration in Eq. (112) to be σ_B^+, and we then have

[9]Note that we have not really established that this interpretation is correct. In particular, our expression for $S_{\rm FG}$ has been derived only for *coherent* incoming states, and may not apply for arbitrary states.

$$S_{\rm BH}+S_{\rm FG}-S_{\rm BI}=\frac{N}{6}\left[\int_{\sigma_B^+}^{\sigma_H^+}d\sigma^+\frac{1}{\lambda}\left[\sqrt{\mathcal{E}(\sigma^+)}-\frac{\lambda}{2}\right]^2+\ln\frac{L}{\delta}\right]. \tag{116}$$

This expression is always positive, so that $S_{\rm tot}$ is always greater than $S_{\rm Boltz}$. In particular, the total entropy $S_{\rm tot}$ always jumps by a (cutoff-dependent) positive amount when the apparent horizon first appears.

The first term in Eq. (116) is minimized if we choose $\mathcal{E}=\lambda^2/4$. This incoming energy flux is the critical flux $\mathcal{E}_{\rm cr}$ that matches the flux of the outgoing Hawking radiation. [From Eq. (106) we see that $\mathcal{E}_{\rm cr}$ corresponds to the conventionally normalized thermal flux $\mathcal{E}_{\rm conv}=N\pi T^2/12$, where $T=\lambda/2\pi$.] We see from[10] Eq. (116) that, even when the black hole is critically illuminated, the total entropy continues to grow like $(N/6)\ln L$. This increasing term arises from the *long-distance* correlations of the quantum fields outside the black hole with the fields in the region behind the horizon. The existence of this term is a bit of a surprise, as one might have expected the critical illumination of the black hole to be a thermodynamically reversible process. Indeed, one might say that the result Eq. (116) calls into question our proposal to identify $S_{\rm tot}$ with the thermodynamic entropy—an expression without the $\ln L$ term would look more plausible. However, we will see in Sec. VII that the second law can be (mildly) violated for an appropriately chosen energy density profile $\mathcal{E}(\sigma^+)$, if the $\ln L$ term is absent.

Note that for a very long-lived black hole, the $\ln L$ term becomes very slowly varying, so that the total entropy of a critically illuminated black hole does become very nearly constant. This is how Eq. (116) becomes reconciled with our calculation of the black hole entropy in Sec. V, where we *did* assume that the emission of radiation by a black hole in a thermal bath is thermodynamically reversible, so that the total entropy remains unchanged. In other words (and not so surprisingly), the process in which a black hole immersed in a thermal bath accretes or emits a small net amount of radiation becomes reversible only when it is carried out arbitrarily slowly.

C. Complete evaporation

Let us now consider a process in which a black hole forms from infalling matter and eventually evaporates completely. Our semiclassical approximations actually break down at the very end of this process, but we can still make definite statements about how the total entropy behaves as the end point of the process approaches.

The end point occurs when the apparent horizon and the singularity coincide, or when $\Omega_H=\Omega_{\rm cr}=\frac{1}{4}$. From Eq. (113), we see that at the end point

$$M=\tfrac{1}{4}\lambda^2 L=\mathcal{E}_{\rm cr}L\ . \tag{117}$$

Equation (117) simply says that, at the end point, the total energy M that has propagated in matches the total energy $\mathcal{E}_{\rm cr}L$ of the Hawking radiation that has been emitted.[11] The relation between M and L is independent of the energy profile of the incoming matter, because the temperature of the black hole is independent of its mass.

At the end point, the black hole entropy goes to zero, so we readily find the fine-grained entropy to be

$$S_{\rm FG}=S_{\rm BH}+S_{\rm FG}=\frac{N}{6}\left[\frac{2M}{\lambda}+\ln\left[\frac{4M}{\lambda^2\delta}\right]\right]. \tag{118}$$

We may regard Eq. (118) as an expression for the amount of information that is destroyed due to the formation and complete evaporation of the black hole. It is not entirely clear how to interpret the ultraviolet divergence in this formula, since the amount of lost information should be finite. Presumably, in a complete description of the evaporation process, there will be some quantum fuzziness in the endpoint, and hence in the position of the global horizon. It then seems plausible that δ would be replaced by a (small) characteristic time scale for the final quantum-mechanical transition that returns the quantum fields to the vacuum state. Thus, we expect that the first term in Eq. (118) will actually dominate over the cutoff-dependent term, in the evaporation of a sufficiently large black hole.

It is easy to understand the origin of the two terms in Eq. (118), by referring to Eq. (69). From Eq. (78) we see that the first term is just $(N/6)\rho_{H,\sigma}$ evaluated at the end point (where $\phi=\phi_{\rm cr}$). As we have already discussed in Sec. IV, $e^{2\rho_{H,\sigma}}$ is the factor by which the modes emitted in the late stages of the process have been redshifted, relative to frequencies measured on $\mathcal{I}^-$. It is the *very-short-distance* correlations between these modes just inside and just outside the horizon that are responsible for the dominant contribution to the entropy in Eq. (118). The subdominant second term in Eq. (118) arises from the *long-distance* correlations between field modes inside and outside the horizon.

The first term in Eq. (118) also has an interpretation in terms of standard thermodynamics. Recalling the relation Eq. (106) between our normalization of energy and the conventional normalization we see that Eq. (118) can be reexpressed as

$$S_{\rm FG}=\frac{2M_{\rm conv}}{T}+\cdots\ , \tag{119}$$

in terms of the conventionally normalized mass that has been emitted by the black hole during its lifetime. The

[10]Since $S_{\rm Boltz}=S_{\rm BO}+S_{\rm BI}$ is conserved (by Liouville's theorem), the expression in Eq. (116) differs from the total entropy by an additive constant.

[11]Actually, this explanation does not exclude a possible extra additive term on the right-hand side of Eq. (117) that is subleading for large L, both because the Hawking flux takes a short while to turn on, and becomes the emitted radiation "overshoots" (resulting in the emission of a negative energy thunderpop at the end point). But it turns out that this potential subleading term is absent.

factor of 2 in Eq. (119) arises because the emission of thermal radiation into cold empty space is an irreversible process [17]. (This factor becomes $(D+1)/D$ in D-dimensional space; to compute it we observe that the entropy S of a relativistic ideal gas is related to its energy E by $S=[(D+1)/D]E/T$. In three dimensions, $\frac{4}{3}$ is modified by "grey-body factors" [18], but there are no such factors in the RST model.)

While this factor of 2 agrees with thermodynamic expectations, that it appears in the *fine-grained* entropy is nonetheless intriguing. We have found that if a black hole forms from collapse and then evaporates, the fine-grained entropy of the emitted radiation is (approximately) twice as large as the Bekenstein-Hawking entropy of the black hole that initially formed. We might have expected, instead, that the amount of quantum-mechanical information that is lost due to the collapse of a pure state is correctly quantified by $S_{\rm BH}$, as it is often presumed [2] that the number of distinct quantum states from which the black hole could have formed is $\exp(S_{\rm BH})$. Then the extra factor of two in the coarse-grained entropy of the emitted radiation would not be due to an intrinsic loss of information; the fine-grained entropy would be only half as large as the coarse-grained entropy, because of subtle correlations among the quanta. Evidently, the radiation outside the horizon is so thoroughly entangled with the degrees of freedom behind the horizon that virtually *all* of its thermodynamic entropy can be attributed to correlations with the fields behind the horizon, and hence to "lost information." Indeed, we can attribute all of the thermodynamic entropy to the exponential redshifting of the modes near the horizon, which, as we noted above, allows shorter and shorter wavelength modes to make a contribution to the fine-grained entropy as the black hole evolves.

Of course, we can make the mass M in Eq. (118) as large as we please by maintaining the black hole for a long time; we just send in a continuous flux of matter that compensates for the outgoing Hawking flux. And we can choose the infalling matter to be in a pure coherent state, with $S_{\rm FG}=0$. It is clear, then, that there is no limit to the amount of information that can be destroyed by the black hole, or in other words, no limit to the degree of entanglement of the fields outside the global horizon with those inside, a conclusion that was already stated in Sec. IV.

The subdominant logarithmic term in Eq. (118) arises from the long-distance correlations of the quantum fields outside the horizon with those inside. This term indicates that the amount of missing information is even greater than naive thermodynamic expectations can accommodate. It would be satisfying to find an interpretation of the logarithmic term in thermodynamic language, but we know no such interpretation.

VII. THE SECOND LAW

Bekenstein conjectured that a generalized second law of thermodynamics applies to processes involving black holes, so that the sum of the entropy outside the black hole and the intrinsic black hole entropy is always nondecreasing [2]. According to this conjecture, although entropy can disappear behind the horizon, the increase in the area of the horizon always compensates (and typically overcompensates) for the lost entropy. Similarly, the emission of Hawking radiation causes the horizon to shrink, but the decrease in horizon area is always compensated by the entropy of the emitted radiation.

We want to examine whether this conjecture holds in the RST model. To show that Bekenstein's conjecture is correct, we need to attach a precise meaning to the notion of the "entropy outside the black hole." Our proposal is that the entropy outside is $S_{\rm FG}+S_{\rm BO}$. Bekenstein's conjecture then becomes the statement that the quantity $S_{\rm tot}$ given by Eq. (115) is nondecreasing. This expression depends on the short-distance cutoff δ that we introduced by smoothing the apparent horizon. But since the cutoff-dependent term is just an additive constant, *changes* in the entropy are not sensitive to the cutoff, at times after the formation of the black hole and before the end point of its evaporation.

Our task is to determine whether there is any energy density profile of the incoming matter for which $S_{\rm tot}$ can decrease as the black hole evolves. We continue to assume, as in Sec. VI, that the incoming matter is in a coherent state built on the asymptotic vacuum state at $\mathcal{I}^-$, and that no infalling matter reaches the boundary of the spacetime before the global event horizon. Under these assumptions, we will show that the second law is valid.

To find the time evolution of $S_{\rm tot}$ in Eq. (115), we will need to know how $L=\sigma_H^+-\sigma_B^+$ evolves, and hence how the position (σ_H^+,σ_H^-) of the apparent horizon evolves. Since the apparent horizon is defined by the condition $\partial_+\Omega|_H=0$, the trajectory $x_H^-(x_H^+)$ of the apparent horizon in Kruskal coordinates satisfies

$$\frac{dx_H^-}{dx_H^+}=-\left.\frac{\partial_+^2\Omega}{\partial_-\partial_+\Omega}\right|_H =-\frac{1}{\lambda^2}\left[T^f_{++}(x_H^+)-\frac{1}{4(x_H^+)^2}\right]; \tag{120}$$

in the second equality we have used the Eq. (15) satisfied by Ω in the Kruskal gauge. Recalling that T^f_{++} transforms as a tensor, we may reexpress this condition in σ coordinates as

$$\frac{d\sigma_H^-}{d\sigma_H^+}=-\frac{1}{\lambda^2}e^{-\lambda(\sigma_H^+-\sigma_H^-)}\left[\mathcal{E}(\sigma_H^+)-\frac{1}{4}\lambda^2\right] =e^{-\lambda L}\left[1-\frac{\mathcal{E}(\sigma_H^+)}{\mathcal{E}_{\rm cr}}\right], \tag{121}$$

where we have used Eq. (71), and have expressed the result in terms of the critical (thermal) flux $\mathcal{E}_{\rm cr}=\frac{1}{4}\lambda^2$. We note that the trajectory of the apparent horizon is time-

like if the incoming flux is less than the outgoing flux due to Hawking radiation, and becomes null when the incoming and outgoing flux match.

If we regard the total entropy as a function of the advanced time σ_H^+ at the apparent horizon, then we may use

$$\frac{dL}{d\sigma_H^+} \equiv \frac{d}{d\sigma_H^+}(\sigma_H^+ - \sigma_B^+) = 1 + e^{-\lambda L}\left[\frac{\mathcal{E}}{\mathcal{E}_{\rm cr}} - 1\right] \tag{122}$$

and Eq. (112) to see that the total entropy given by Eq. (115) varies at the rate

$$\frac{d}{d\sigma_H^+} S_{\rm tot} = \frac{N\lambda}{24}\left[[\sqrt{\tilde{\mathcal{E}}(\sigma_H^+)} - 1]^2 + e^{-\lambda L}[\tilde{\mathcal{E}}(\sigma_H^+) - 1]\left[1 + \frac{4}{\lambda L}\right] + \frac{4}{\lambda L}\right], \tag{123}$$

which we have expressed in terms of

$$\tilde{\mathcal{E}}(\sigma_H^+) = \frac{\mathcal{E}(\sigma_H^+)}{\mathcal{E}_{\rm cr}}, \tag{124}$$

the ratio of the incoming flux to the thermal flux. As expected, the rate of change of the entropy does not depend on the short-distance cutoff δ.

It is not hard to check that Eq. (123) is *positive* for any $\tilde{\mathcal{E}} \geq 0$ and any finite $L > 0$. For a fixed L, $dS_{\rm tot}/d\sigma_H^+$ is minimized when the incoming flux is

$$\tilde{\mathcal{E}} = \left[1 + e^{-\lambda L} + \frac{4}{\lambda L} e^{-\lambda L}\right]^{-2}, \tag{125}$$

and the minimum value attained is

$$\left.\frac{d}{d\sigma_H^+} S_{\rm tot}\right|_{\rm min} = \frac{N\lambda}{24}\left[\frac{4}{\lambda L} - \frac{e^{-2\lambda L}\left[1 + \frac{4}{\lambda L}\right]^2}{1 + e^{-\lambda L}\left[1 + \frac{4}{\lambda L}\right]}\right]. \tag{126}$$

This expression is a monotonically decreasing function of λL that approaches zero as $\lambda L \to \infty$. Thus, we see that the total entropy is always increasing, in accord with the generalized second law.

If the black hole is critically illuminated ($\tilde{\mathcal{E}} = 1$), the mass radiated away is matched exactly by the incoming matter flux. We see from Eq. (123) that the total entropy nevertheless continues to increase for $L < \infty$ (as we already noted in Sec. VI). The entropy increase is due to the $\ln(L/\delta)$ term in $S_{\rm tot}$, the term arising from the long-distance correlations of the quantum fields outside the horizon with those inside. This term is consistent with the property that a black hole can reach thermal equilibrium with a radiation bath, because the rate of change of the entropy approaches zero as the age L of the black hole gets arbitrarily large. Still, since the $\ln(L/\delta)$ term has no clear thermodynamic interpretation, one is tempted to seek a reformulation of the second law in which the long-distance contribution to the fine-grained entropy is absent.

The obvious thing to try is to subtract the offending term away, and define a new total entropy

$$S_{\rm tot}^{\rm (new)} = S_{\rm tot} - \frac{N}{6}\ln\left[\frac{L}{\delta}\right]. \tag{127}$$

The rate of change of this entropy is

$$\frac{d}{d\sigma_H^+} S_{\rm tot}^{\rm (new)} = \frac{N\lambda}{24}\left[(1 + e^{-\lambda L})\left[\sqrt{\tilde{\mathcal{E}}} - \frac{1}{1 + e^{-\lambda L}}\right]^2 - \frac{e^{-2\lambda L}}{1 + e^{-\lambda L}}\right]. \tag{128}$$

We see that the new entropy does not strictly satisfy the second law. The entropy of a critically illuminated black hole is constant, but the entropy decreases slowly if the incoming flux is slightly below critical. On the other hand, for $\lambda L \gg 1$ the violations of the new second law are extremely mild, and occur only under very rare conditions. The entropy is nonincreasing unless the flux lies in the narrow range

$$1 > \tilde{\mathcal{E}} > \tanh^2\left[\frac{\lambda L}{2}\right] \approx 1 - 4e^{-\lambda L}. \tag{129}$$

Thus, for $\lambda L \gg 1$, the second law fails only when the flux is tuned to be exponentially close to critical, and even then the rate of decrease of the entropy is exponentially small.

We caution the reader again that our derivation of the second law applies only under special conditions. In particular, we have assumed that the incoming matter is in a coherent state built on the inertial vacuum at $\mathcal{I}^-$. When more general quantum states are considered, our proof breaks down. We will show in Appendix B that states can be constructed that carry, locally, a large amount of fine-grained entropy and a small amount of energy, or carry negative energy density without accompanying negative entropy [13,15]. (Neither of these pathologies occurs for the coherent states built on the inertial vacuum.) Thus, the second law, as we have formulated it here, can be violated at least for a while by tossing matter in such a state into the black hole. Such examples show that if there is a very general statement of the second law, our expression for the total entropy cannot apply in all situations.

Boltzmann's derivation of the macroscopic second law of thermodynamics from the microscopic laws of statistical mechanics is one of the most satisfying developments in the history of physics. We believe that there should be an equally satisfying derivation of Bekenstein's generalized second law. In this paper, beginning from the microscopic laws of a specific two-dimensional theory, we have given a derivation of Bekenstein's generalized

second law which is applicable to a wide range of processes. Yet we do not feel that our derivation has provided complete insight into *why* the generalized second law is (often) valid, because we relied mainly on explicit calculation, rather than general reasoning. Indeed, it is not evident from our derivation that the generalized second law will hold in variants of the RST model. Thus, while we have made some progress, the true nature of Bekenstein's generalized second law remains an outstanding enigma.

ACKNOWLEDGMENTS

We have benefited from discussions with S. Das, L. Thorlacius, and especially S. Mathur. This work was supported in part by the U.S. Department of Energy under Grant No. DOE-91ER40618 and Grant No. DE-FG03-92-ER40701.

APPENDIX A

In our calculations of the fine-grained entropy in Sec. III, we considered the quantum state of the scalar field to be either the inertial vacuum or a "vacuum" state that is conformally related to the inertial vacuum. In this appendix we will generalize the results to include the case of a coherent state built on such a "vacuum." We will show that the fine-grained entropy for the coherent state is the same as the fine-grained entropy of the vacuum state. Thus, if space is divided into two regions, building a coherent state on the vacuum does not affect the degree of entanglement of the quantum fields in the two regions.

To begin, we consider a toy problem that incorporates all of the essential features of the general case. Consider a system of two uncoupled harmonic oscillators, with associated annihilation operators a_1 and a_2. Perform a Bogolubov transformation of the form

$$\begin{aligned} a_1^\gamma &= \frac{1}{\sqrt{1-\gamma^2}}(a_1-\gamma a_2^\dagger)\ , \\ a_2^\gamma &= \frac{1}{\sqrt{1-\gamma^2}}(a_2-\gamma a_1^\dagger)\ , \end{aligned} \tag{A1}$$

where γ is real and $\gamma^2<1$. This is the most general Bogolubov transformation in which a_1^γ is a linear combination of an a_1 annihilation operator and an $a_2^\dagger$ creation operator, up to phases that can be removed by adjusting the phases of the a_1, a_2, a_1^γ, and a_2^γ.

We can now construct the "γ vacuum" that is annihilated by a_1^γ and a_2^γ; it is

$$\begin{aligned} |\gamma\rangle &= \sqrt{1-\gamma^2}\exp(\gamma a_1^\dagger a_2^\dagger)|0,1\rangle\otimes|0,2\rangle \\ &= \sqrt{1-\gamma^2}\sum_{n=0}^{\infty}\gamma^n|n,1\rangle\otimes|n,2\rangle\ , \end{aligned} \tag{A2}$$

where $|n,1\rangle$ and $|n,2\rangle$ denote the nth excitation of oscillators 1 and 2, respectively. The easiest way to verify the first equality in Eq. (A2) is to use the representation of the commutation relations with

$$a_1=\frac{\partial}{\partial a_1^\dagger}\ ,\quad a_2=\frac{\partial}{\partial a_2^\dagger}\ . \tag{A3}$$

The conditions $a_1^\gamma|\gamma\rangle=a_2^\gamma|\gamma\rangle=0$ become two coupled first-order differential equations satisfied by the coefficient of $|0,1\rangle\otimes|0,2\rangle$; the expression in Eq. (A2) is the unique solution that yields a normalized state.

If we now trace over the state of the second oscillator to find a density matrix for the first oscillator, we obtain

$$\begin{aligned} \rho_1^\gamma &\equiv \mathrm{tr}_2(|\gamma\rangle\langle\gamma|) \\ &= (1-\gamma^2)\sum_{n=0}^{\infty}\gamma^{2n}|n,1\rangle\langle n,1|\ . \end{aligned} \tag{A4}$$

This has the precise form of a thermal density matrix with inverse temperature β given by

$$\gamma^2=e^{-\beta\omega}\ , \tag{A5}$$

where ω is the frequency of oscillator 1. The calculation we have performed is just what is needed to proceed from Eqs. (39) and (40) to Eq. (42).

A general coherent state built "on top of" the state $|\gamma\rangle$ has the form

$$|\gamma,\alpha_1,\alpha_2\rangle = N_{\alpha_1\alpha_2}\exp[\alpha_1(a_1^\gamma)^\dagger]\exp[\alpha_2(a_2^\gamma)^\dagger]|\gamma\rangle\ , \tag{A6}$$

where $N_{\alpha_1\alpha_2}$ is a normalization constant. This is the unique normalized state that obeys the conditions

$$\begin{aligned} (a_1^\gamma-\alpha_1)|\gamma,\alpha_1,\alpha_2\rangle&=0\ , \\ (a_2^\gamma-\alpha_2)|\gamma,\alpha_1,\alpha_2\rangle&=0\ . \end{aligned} \tag{A7}$$

Thus, we may regard the coherent state as the "vacuum" state of the *shifted* annihilation operators

$$\hat{a}_1^\gamma=a_1^\gamma-\alpha_1,\quad \hat{a}_2^\gamma=a_2^\gamma-\alpha_2\ . \tag{A8}$$

If we also define shifted annihilation operators

$$\hat{a}_1=a_1-\left[\frac{\alpha_1+\gamma\alpha_2^*}{\sqrt{1-\gamma^2}}\right]\ ,\quad \hat{a}_2=a_2-\left[\frac{\alpha_2+\gamma\alpha_1^*}{\sqrt{1-\gamma^2}}\right]\ , \tag{A9}$$

then the Bogolubov transformation relating $\hat{a}_{1,2}^\gamma$ to $\hat{a}_{1,2}$ is

$$\begin{aligned} \hat{a}_1^\gamma &= \frac{1}{\sqrt{1-\gamma^2}}(\hat{a}_1-\gamma\hat{a}_2^\dagger)\ , \\ \hat{a}_2^\gamma &= \frac{1}{\sqrt{1-\gamma^2}}(\hat{a}_2-\gamma\hat{a}_1^\dagger)\ , \end{aligned} \tag{A10}$$

which has exactly the same form as Eq. (A1). Since the shifted operators obey the standard commutation relations, the same argument as before shows that the coherent state can be expressed as

$$|\gamma,\alpha_1,\alpha_2\rangle=\sqrt{1-\gamma^2}\exp(\gamma\hat{a}_1^\dagger\hat{a}_2^\dagger)|\hat{0},1\rangle\otimes|\hat{0},2\rangle\ , \tag{A11}$$

where $|\hat{0},1\rangle$ and $|\hat{0},2\rangle$ are the ground states of the *shifted* oscillators 1 and 2 (or, in other words, coherent states of the unshifted oscillators). We can trace over the second oscillator just as before and find

$$\begin{aligned} \rho_1^{\gamma\alpha_1\alpha_2} &\equiv \mathrm{tr}_2(|\gamma,\alpha_1,\alpha_2\rangle\langle\gamma,\alpha_1\alpha_2|) \\ &= (1-\gamma^2)\sum_{n=0}^{\infty}\gamma^{2n}|\hat{n},1\rangle\langle\hat{n},1|\ . \end{aligned} \tag{A12}$$

This density matrix has exactly the same form as Eq. (A4), except that we are now expanding in terms of the basis of states that have definite occupation number with respect to the shifted oscillators. The coherent state density matrix, then, has exactly the same eigenvalues as the vacuum density matrix, and it therefore also has exactly the same entropy. Note that it is not quite correct to describe Eq. (A12) as a "thermal density matrix," because the eigenstates of the shifted number operator $\hat{a}_1^\dagger \hat{a}_1$ are not eigenstates of the Hamiltonian $H=\omega a_1^\dagger a_1$.

Now we note that the case of two entangled oscillators described above is all that we need to deal with when we compute the fine-grained entropy for a free field. It follows from Eqs. (39) and (40) that

$$a_{1,\omega}=\frac{1}{\sqrt{1-e^{-2\pi\omega}}}(a_{R,\omega}-e^{-\pi\omega}a_{L,\omega}^\dagger)\ ,$$
$$a_{2,\omega}=\frac{1}{\sqrt{1-e^{-2\pi\omega}}}(a_{L,\omega}-e^{-\pi\omega}a_{R,\omega}^\dagger)\ , \tag{A13}$$

are operators that annihilate modes that are positive frequency with respect to Minkowski time; $a_{R,\omega}$ and $a_{L,\omega}$ denote the operators that annihilate the modes of Rindler frequency ω that are localized in the right and left wedges, respectively. [The minus signs in Eq. (A13) arise from the minus sign in the Klein-Gordon inner product of two negative frequency modes.] Thus, the expression Eq. (41) for the Minkowski vacuum is a tensor product of states that have just the form Eq. (A2), with $\gamma=e^{-\pi\omega}$. Each field mode in the right Rindler wedge is correlated with a particular mode in the left Rindler wedge; for each such pair of modes, the entanglement of the state of the right mode with the state of the left mode has exactly the same form as the entanglement of oscillator 1 with oscillator 2 in the above discussion.

Furthermore, a general coherent state built on the Minkowski vacuum also has the property that it can be factorized into a tensor product of correlated states for pairs of modes. The general coherent state can be expressed as

$$|\text{Minkowski coherent}\rangle = N\prod_j (e^{(\alpha_{1,j}a_{1,j}^\dagger)}e^{(\alpha_{2,j}a_{2,j}^\dagger)}|0_M,j\rangle)\ , \tag{A14}$$

where $|0_M,j\rangle$ denotes the state that is annihilated by the Minkowski annihilation operators $a_{1,j}$ and $a_{2,j}$. Equation (A14) is just a product of states of the form $|\gamma_j=e^{-\pi\omega_j},\alpha_{R,j},\alpha_{L,j}\rangle$. The evaluation of the density matrix ρ_R in the right Rindler wedge than proceeds as above, and we find that it has the same eigenvalues for the coherent state as for the Minkowski vacuum.

As our arguments in Sec. III show, the Minkowski vacuum still has the form Eq. (41) when expressed in terms of the modes that are localized inside and outside a finite region of space, and the general coherent state built on the Minkowski vacuum still has the form Eq. (A14). These statements remain true if we consider, not the Minkowski vacuum, but a state that is conformally related to it. Also, the form Eq. (A14) applies in curved space as well as in flat space.

We conclude, finally, that our formula Eq. (69) for the fine-grained entropy outside the apparent horizon of a black hole applies not just when the incoming quantum state of the matter fields is the asymptotic inertial vacuum, but also when the quantum state is an arbitrary coherent state built on the inertial vacuum.

APPENDIX B

In our derivation of the generalized second law in Sec. VII, we made some restrictive assumptions about the incoming matter. In particular, we assumed that the quantum state of the matter is a coherent state built on the asymptotic inertial vacuum state at $\mathcal{I}^-$. In this appendix we will examine what happens when this assumption is relaxed. We will show that if more general quantum states are allowed, the total entropy can decrease. Thus, the second law can be violated.

The crucial point is that quantum states can be constructed that pack a large positive density of (fine-grained) entropy without carrying a large energy density. We can prepare matter in such a state, and allow the matter to fall into a black hole. Then the fine-grained entropy decreases sharply, but without any compensating sharp increase in the black hole entropy. Hence, the total entropy decreases.

Alternately, we can make the total entropy decrease (momentarily) by simply sending negative energy into the black hole. It can be arranged that the black hole shrinks and loses entropy without a compensating increase in the fine-grained entropy.

To demonstrate the existence of such states, consider an initial state of left-moving matter than is in the "vacuum" state defined not with respect to the asymptotic inertial coordinate σ^+, but rather with respect to a different coordinate $\hat{x}^+(\sigma^+)$. In this quantum state, the incoming energy flux, expressed in the σ gauge, is [35]

$$\mathcal{E}(\sigma^+)\equiv\langle :\hat{T}_{++}(\sigma^+):\sigma\rangle$$
$$=-\left[\frac{d\hat{x}^+}{d\sigma^+}\right]^{3/2}\frac{d^2}{d(\hat{x}^+)^2}\left[\frac{d\hat{x}^+}{d\sigma^+}\right]^{1/2}$$
$$=\frac{3(h')^2}{4h^2}-\frac{h''}{2h}\ , \tag{B1}$$

where

$$h\equiv\frac{d\hat{x}^+}{d\sigma^+}\ , \tag{B2}$$

and the prime denotes differentiation with respect to σ^+. [Here we have used the normalization convention of Eq. (16), and have assumed that there are N massless scalar matter fields.] Note that the energy density is not necessarily positive. In this "vacuum" the equation for the trajectory of the apparent horizon, in Kruskal coordinates, is

$$\frac{dx_H^-}{dx_H^+}=-\left.\frac{\partial_+^2\Omega}{\partial_-\partial_+\Omega}\right|_H=-\frac{1}{\lambda^2}t_+(x_H^+)$$
$$=-\frac{1}{\lambda^4(x_H^+)^2}\left[\mathcal{E}-\frac{1}{4}\lambda^2\right]\ . \tag{B3}$$

Thus, as in our previous analysis of coherent states built on the σ vacuum, the condition for "critical illumination" is $\mathcal{E}=\mathcal{E}_{cr}\equiv\frac{1}{4}\lambda^2$; when this condition is satisfied, the incoming flux matches the flux of the outgoing Hawking radiation, and the apparent horizon is null.

For this state, the expression Eq. (66) for the fine-grained entropy outside of the apparent horizon becomes

$$S_{FG}=\frac{N}{6}\left[\rho_{H,\sigma}-\frac{1}{2}\ln\left[\frac{d\hat{x}_H^+}{d\sigma_H^+}\frac{d\hat{x}_H^-}{d\sigma_H^-}\right]+\ln\left[\frac{\hat{x}_H^+-\hat{x}_B^+}{\delta}\right]\right]. \tag{B4}$$

This formula differs from our old expression Eq. (69) in two respects. First, the affine volume in the argument of the logarithm in the third term is expressed in terms of the $\hat{x}^+$ coordinate that is used to define the vacuum, rather than the inertial σ^+ coordinate. Second, the term that enters when we reexpress the cutoff in terms of the inertial coordinates at the horizon is the conformal factor of the metric in $\hat{x}$ coordinates. This differs from the conformal factor in σ coordinates, which accounts for the second term in Eq. (B4).

Now let us suppose that the function $\hat{x}^+(\sigma^+)$ is chosen so that the black hole is critically illuminated at a particular advanced time σ_H^+. At that moment, x_H^- is instantaneously constant, as is the value Ω_H of Ω at the apparent horizon. Thus, it is easy to evaluate the rate at which the fine-grained entropy is changing. Using Eq. (74) we find

$$\frac{dS_{FG}}{d\sigma_H^+}=\frac{N}{6}\left[\frac{1}{2}\lambda-\frac{h'}{2h}+\frac{h}{\hat{x}_H^+-\hat{x}_B^+}\right]. \tag{B5}$$

It is clear from Eq. (B5) that we can make the rate of change of S_{FG} large and negative by choosing $\hat{x}^+(\sigma^+)$ so that h' is large and positive. Furthermore, we may simultaneously arrange that h'' is large, so that $\mathcal{E}(\sigma^+)$ in Eq. (B1) obeys the critical illumination condition. Finally, under critical illumination, the black hole entropy is constant, so that no increase in the black hole entropy compensates for the decrease in the fine grained entropy, and the Boltzman entropy outside the black hole is also decreasing. Hence, the total entropy decreases.

Another way to make the total entropy momentarily decrease is to throw negative energy into the black hole. Evidently this can be achieved by choosing $h'=0$ and $h''>0$ in Eq. (B1). The black hole will then shrink and decrease its entropy, but there will not in general be any compensating increase in S_{FG}. It is not clear, however, how an analog of the Boltzman entropy should be defined for these states that carry negative energy density.

A preliminary investigation of the properties of states with the above properties indicates that such an imbalance between entropy and energy cannot be sustained indefinitely [48]. We expect that there are fundamental limitations on the severity and duration of these violations of the generalized second law.

[1] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).

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

[3] W. G. Unruh and R. M. Wald, Phys. Rev. D **25**, 942 (1982); W. H. Zurek and K. S. Thorne, Phys. Rev. Lett. **54**, 2171 (1985); K. S. Thorne, W. H. Zurek, and R. H. Price, in *Black Holes: The Membrane Paradigm,* edited by K. S. Thorne, R. H. Price, and D. A. MacDonald (Yale University Press, New Haven, 1986), p. 280; R. M. Wald, in *Black Hole Physics,* edited by V. De Sabbata and Zhenjiu Zhang (Kluwer, Dordrecht, 1992), p. 55; I. D. Novikov and V. P. Frolov, *Physics of Black Holes* (Kluwer, Dordrecht, 1989), and references therein.

[4] V. P. Frolov and D. N. Page, Phys. Rev. Lett. **71**, 3902 (1993).

[5] R. M. Wald, Commun. Math. Phys. **45**, 9 (1975).

[6] S. W. Hawking, Phys. Rev. D **14**, 2460 (1976); Commun. Math. Phys. **87**, 395 (1982).

[7] Y. Aharonov, A. Casher, and S. Nussinov, Phys. Lett. B **191**, 51 (1987); T. Banks, A. Dabholkar, M. R. Douglas, and M. O'Loughlin, Phys. Rev. D **45**, 3607 (1992); T. Banks and M. O'Loughlin, *ibid.* **47**, 540 (1993); T. Banks, M. O'Loughlin, and A. Strominger, *ibid.* **47**, 4476 (1993); S. B. Giddings, *ibid.* **49**, 947 (1994).

[8] C. G. Callan, S. B. Giddings, J. A. Harvey, and A. Strominger, Phys. Rev. D **45**, R1005 (1992); for reviews see J. A. Harvey and A. Strominger, in *Recent Directions in Particle Theory—From Superstrings and Black Holes to the Standard Model,* Proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, 1992, edited by J. Harvey and J. Pochinski (World Scientific, Singapore, 1993); and S. B. Giddings, in *String Quantum Gravity and Physics at the Planck Energy Scale,* Proceedings of the International Workshop on Theoretical Physics, Erice, Italy, 1992, edited by N. Sanchez (World Scientific, Singapore, 1993).

[9] G. W. Gibbons and K. Maeda, Nucl. Phys. **B298**, 741 (1988); D. Garfinkle, G. Horowitz, and A. Strominger, Phys. Rev. D **43**, 3140 (1991).

[10] J. G. Russo, L. Susskind, and L. Thorlacius, Phys. Rev. D **46**, 3444 (1992); **47**, 533 (1993).

[11] A. Bilal and C. G. Callan, Nucl. Phys. **B394**, 73 (1993); S. P. de Alwis, Phys. Lett. B **289**, 278 (1992); **300**, 330 (1993); S. B. Giddings and A. Strominger, Phys. Rev. D **46**, 2454 (1993).

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

[13] C. Holzhey, Ph.D. thesis, Princeton University, 1993.

[14] G. 't Hooft, Nucl. Phys. **B256**, 727 (1985).

[15] F. Wilczek, in *Black Holes, Membranes, Wormholes, and Superstrings,* Proceedings of the International Symposium, Woodlands, Texas, 1992, edited by S. Kalara and D. V. Nanopoulos (World Scientific, Singapore, 1993), p. 1.

[16] A. Strominger and S. Trivedi, Phys. Rev. D **48**, 5778 (1993).

[17] W. H. Zurek, Phys. Rev. Lett. **49**, 1683 (1982).

[18] D. N. Page, Phys. Rev. D **14**, 3260 (1976); Phys. Rev. Lett. **50**, 1013 (1983).

[19] G. 't Hooft, Nucl. Phys. **B335**, 138 (1990), and references

therein.
[20] T. Jacobson, Phys. Rev. D **44**, 173 (1991); **48**, 728 (1993).
[21] L. Susskind, Phys. Rev. Lett. **71**, 2367 (1993); L. Susskind and L. Thorlacius, Phys. Rev. D **49**, 966 (1994); L. Susskind, *ibid.* **49**, 6606 (1994); "Some Speculations about Black Hole Entropy in String Theory," Rutgers Report No. RU-93-44, hep-th/9309145, 1993 (unpublished).
[22] E. Verlinde and H. Verlinde, "A Unitary *S*-Matrix for 2D Black Hole Formation and Evaporation," Princeton Report No. PUPT-1380, IASSNS-HEP-93/8, hep-th/9302022, 1993 (unpublished); K. Schoutens, E. Verlinde, and H. Verlinde, Phys. Rev. D **48**, 2670 (1993).
[23] E. Keski-Vakkuri and S. D. Mathur, Phys. Rev. D **50**, 917 (1994).
[24] C. Holzhey, F. Larsen, and F. Wilczek, "Geometric and Renormalized Entropy in Conformal Field Theory," Princeton Report No. PUTP 1454, hep-th/9403108, 1994 (unpublished).
[25] L. Susskind and J. Uglum, Phys. Rev. D **50**, 2700 (1994).
[26] C. Callan and F. Wilczek, "On Geometric Entropy," Institute for Advanced Study Report No. IASSNS-HEP-93/87, hep-th/9401072, 1994 (unpublished).
[27] D. Kabat and M. J. Strassler, "A Comment on Entropy and Area," Rutgers University Report No. RU-84-10, hep-th/9401125, 1994 (unpublished).
[28] J. S. Dowker, Class. Quantum Grav. **11**, L55 (1994).
[29] D. A. Lowe, Phys. Rev. D **47**, 2446 (1993).
[30] T. Piran and A. Strominger, Phys. Rev. D **48**, 4729 (1993).
[31] S. B. Giddings and W. M. Nelson, Phys. Rev. D **46**, 2486 (1992).
[32] A. M. Polyakov, Phys. Lett. **103B**, 207 (1981).
[33] S. W. Hawking, Phys. Rev. Lett. **69**, 406 (1992); B. Birnir, S. B. Giddings, J. A. Harvey, and A. Strominger, Phys. Rev. D **46**, 638 (1992); S. W. Hawking and J. M. Stewart, Nucl. Phys. **B400**, 393 (1993).
[34] A. Strominger, Phys. Rev. D **48**, 5769 (1993).
[35] S. M. Christensen and S. A. Fulling, Phys. Rev. D **15**, 2088 (1977).
[36] A. Strominger and L. Thorlacius, Phys. Rev. D (to be published); A. Strominger, L. Thorlacius, and S. Trivedi (unpublished).
[37] M. Srednicki, Phys. Rev. Lett. **71**, 666 (1993); and (private communication).
[38] L. Bombelli, R. K. Koul, J. Lee, and R. Sorkin, Phys. Rev. D **34**, 373 (1986).
[39] D. N. Page, Phys. Rev. Lett. **71**, 3743 (1993).
[40] R. D. Carlitz and R. S. Willey, Phys. Rev. D **36**, 2336 (1987).
[41] J. Preskill, in *Black Holes, Membranes, Wormholes, and Superstrings* [15], p. 22.
[42] L. Susskind, L. Thorlacius, and J. Uglum, Phys. Rev. D **48**, 3743 (1993).
[43] N. Seiberg, S. Shenker, L. Susskind, L. Thorlacius, and J. Tuttle (unpublished).
[44] J. B. Hartle and S. W. Hawking, Phys. Rev. D **13**, 2188 (1976).
[45] V. Frolov and I. Novikov, Phys. Rev. D **48**, 4545 (1993).
[46] For a cogent recent review, see J. L. Lebowitz, Physica A **194**, 1 (1993).
[47] C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Freeman, New York, 1973).
[48] J. Preskill and S. Trivedi (unpublished).

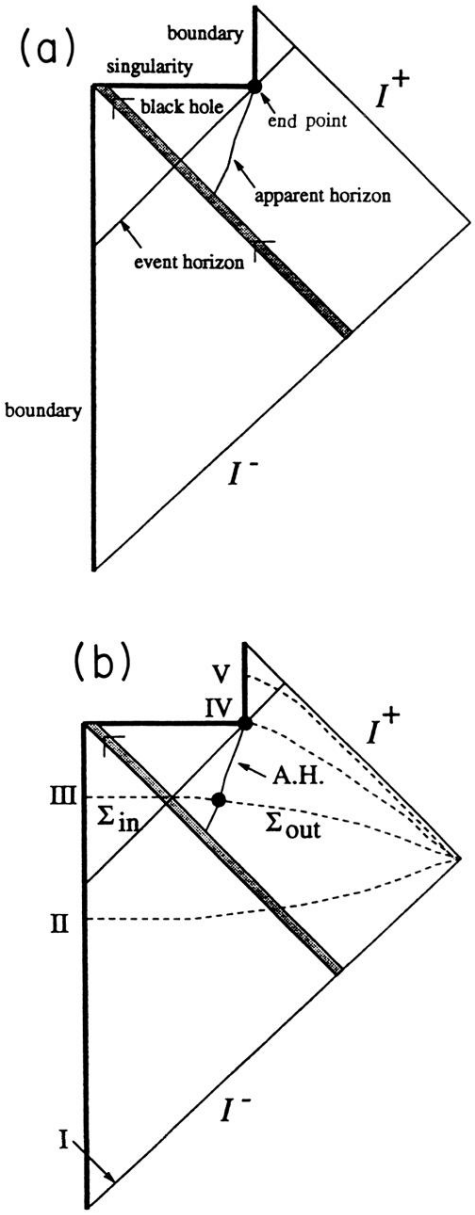

FIG. 1. (a) The two-dimensional spacetime of a black hole that forms due to the collapse of a shock wave, and then evaporates completely. After the black hole forms, the apparent horizon recedes along a timelike trajectory, eventually meeting the singularity at the "end point." The timelike boundary and the spacelike singularity are in the strongly coupled region. RST boundary conditions are imposed where the boundary is timelike. (b) Five spacelike slices through the spacetime, referred to in the text.

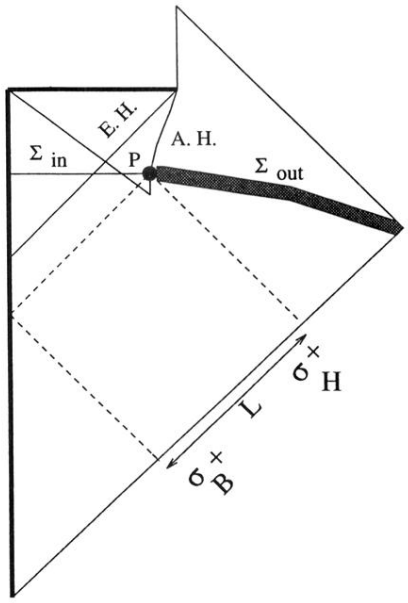

FIG. 6. A spacelike slice Σ through the black hole spacetime. The slice crosses the apparent horizon at the point $P=(\sigma_H^-,\sigma_H^+)$. We define σ_B^+ as the advanced time of an incoming null ray that reflects off the boundary and then passes through P. Incoming null rays with advanced time between σ_B^+ and σ_H^+ cross Σ inside the apparent horizon.