Evaporation of Two-Dimensional Black Holes

S. W. Hawking
*California Institute of Technology, Pasadena, California 91125
and Department of Applied Mathematics and Theoretical Physics, University of Cambridge,
Silver Street, Cambridge CB3 9EW, United Kingdom*
(Received 23 March 1992)

An interesting two-dimensional model theory has been proposed that allows one to consider black-hole evaporation in the semiclassical approximation. The semiclassical equations will give a singularity where the dilaton field reaches a certain critical value. This singularity will be hidden behind a horizon. As the evaporation proceeds, the dilaton field on the horizon will approach the critical value but the temperature and rate of emission will remain finite. These results indicate either that there is a naked singularity, or (more likely) that the semiclassical approximation breaks down.

PACS numbers: 97.60.Lf, 04.20.Cv, 04.60.+n

Callan, Giddings, Harvey, and Strominger (CGHS) [1] have suggested an interesting two-dimensional theory with a metric coupled to a dilaton field and N minimal scalar fields. The Lagrangian is

$$L=\frac{1}{2\pi}\sqrt{-g}\left[e^{-2\phi}[R+4(\nabla\phi)^2+4\lambda^2]-\frac{1}{2}\sum_{i=1}^{N}(\nabla f_i)^2\right].$$

If one writes the metric in the form

$$ds^2=e^{2\rho}dx_+dx_-,$$

the classical field equations are

$$\partial_+\partial_-f_i=0,$$

$$2\partial_+\partial_-\phi-2\partial_+\phi\,\partial_-\phi-\tfrac{1}{2}\lambda^2e^{2\rho}=\partial_+\partial_-\rho,$$

$$\partial_+\partial_-\phi-2\partial_+\phi\,\partial_-\phi-\tfrac{1}{2}\lambda^2e^{2\rho}=0.$$

These equations have a solution

$$\phi=-b\ln(-x_+x_-)-c-\ln\lambda,$$

$$\rho=-\tfrac{1}{2}\ln(-x_+x_-)+\ln(2b/\lambda),$$

where b and c are constants and b can be taken to be positive without loss of generality. A change of coordinates

$$u_\pm=\pm(2b/\lambda)\ln(\pm x_\pm)\pm(1/\lambda)(c+\ln\lambda)$$

gives a flat metric and a linear dilaton field

$$\rho=0,$$

$$\phi=-\tfrac{1}{2}\lambda(u_+-u_-).$$

This solution is known as the linear dilaton. The solution is independent of the constants b and c which correspond to freedom in the choice of coordinates. Normally b is taken to have the value $\frac{1}{2}$.

These equations also admit a solution

$$\phi=\rho-c=-\tfrac{1}{2}\ln(M\lambda^{-1}-\lambda^2e^{2c}x_+x_-).$$

This represents a two-dimensional black hole with horizons at $x_\pm=0$ and singularities at $x_+x_-=M\lambda^{-2}e^{-2c}$. Note that there is still freedom to shift the ρ field on the horizon by a constant and compensate by rescaling the coordinates $x_\pm$, but there is nothing corresponding to the freedom to choose the constant b. In terms of the coordinates $u_\pm$ defined as before with $b=\frac{1}{2}$,

$$\rho=-\tfrac{1}{2}\ln(1-M\lambda^{-1}e^{-\lambda(u_+-u_-)}),$$

$$\phi=-\tfrac{1}{2}\lambda(u_+-u_-)-\tfrac{1}{2}\ln(1-M\lambda^{-1}e^{-\lambda(u_+-u_-)}).$$

This black-hole solution is periodic in the imaginary time with period $2\pi\lambda^{-1}$. One would therefore expect it to have a temperature

$$T=\lambda/2\pi$$

and to emit thermal radiation [2]. This is confirmed by CGHS. They considered a black hole formed by sending in a thin shock wave of one of the f_i fields from the weak-coupling region (large negative ϕ) of the linear dilaton. One can calculate the energy-momentum tensors of the f_i fields, using the conservation and trace anomaly equations. If one imposes the boundary condition that there is no incoming energy momentum apart from the shock wave, one finds that at late retarded times u_- there is a steady flow of energy in each f_i field at the mass-independent rate

$$\lambda^2/48.$$

If this radiation continued indefinitely, the black hole would radiate an infinite amount of energy, which seems absurd. One might therefore expect that the backreaction would modify the emission and cause it to stop when the black hole had radiated away its initial mass. A fully quantum treatment of the backreaction seems very difficult even in this two-dimensional theory. But CGHS suggested that in the limit of a large number N of scalar fields f_i, one could neglect the quantum fluctuations of the dilaton and the metric and treat the backreaction of the radiation in the f_i fields semiclassically by adding to the action a trace anomaly term

$$\tfrac{1}{12}N\partial_+\partial_-\rho.$$

The evolution equations that result from this action are

$$\partial_+\partial_-\phi=(1-\tfrac{1}{24}Ne^{2\phi})\partial_+\partial_-\rho\,,$$

$$2(1-\tfrac{1}{12}Ne^{2\phi})\partial_+\partial_-\phi=(1-\tfrac{1}{24}Ne^{2\phi})\times(4\partial_+\phi\,\partial_-\phi+\lambda^2e^{2\rho})\,.$$

In addition, there are two equations that can be regarded as constraints on the data on characteristic surfaces of constant $x_\pm$,

$$\partial_+^2\phi-2\partial_+\rho\,\partial_+\phi=\tfrac{1}{24}Ne^{2\phi}[\partial_+^2\rho-\partial_+\rho\,\partial_+\rho-t_+(x^+)]\,,$$

$$\partial_-^2\phi-2\partial_-\rho\,\partial_-\phi=\tfrac{1}{24}Ne^{2\phi}[\partial_-^2\rho-\partial_-\rho\,\partial_-\rho-t_-(x^-)]\,,$$

where $t_\pm(x^\pm)$ are determined by the boundary conditions in a manner that will be explained later.

Even these semiclassical equations seem too difficult to solve in closed form. CGHS suggested that a black hole formed from an f wave would evaporate completely without there being any singularity. The solution would approach the linear dilaton at late retarded times u_- and there would be no horizons. They therefore claimed that there would be no loss of quantum coherence in the formation and evaporation of a two-dimensional black hole: The radiation would be in a pure quantum state, rather than in a mixed state.

In [3,4] it was shown that this scenario could not be correct. The solution would develop a singularity on the incoming f wave at the point where the dilaton field reached the critical value

$$\phi_0=-\tfrac{1}{2}\ln(N/12)\,.$$

This singularity will be spacelike near the f wave [4]. Thus at least part of the final quantum state will end up on the singularity, which implies that the radiation at infinity in the weak-coupling region will not be in a pure quantum state.

The outstanding question is: How does the spacetime evolve to the future of the f wave? There seem to be two main possibilities: (1) The singularity remains hidden behind an event horizon. One can continue an infinite distance into the future on a line of constant $\phi<\phi_0$ without ever seeing the singularity. If this were the case, the rate of radiation would have to go to zero. (2) The singularity is naked. That is, it is visible from a line of constant ϕ at a finite time to the future of the f wave. Any evolution of the solution after this would not be uniquely determined by the semiclassical equations and the initial data. Indeed, it is likely that the point at which the singularity became visible was itself singular and that the solution could not be evolved to the future for more than a finite time.

In what follows I shall present evidence that suggests the semiclassical equations lead to possibility (2). This probably indicates that the semiclassical approximation breaks down as the dilaton field on the horizon approaches the critical value.

Static black holes.—If the solution were to evolve without a naked singularity, it would presumably approach a static state in which a singularity was hidden behind an event horizon. This motivates a study of static black-hole solutions of the semiclassical equations. One could look for solutions in which ϕ and ρ were independent of the "time" coordinate $\tau=x_++x_-$ and depended only on a "radial" variable $\sigma=x_+-x_-$ but this has the disadvantage that the Killing vector $\partial/\partial\tau$ is timelike everywhere. This means the black-hole horizon is at $\sigma=-\infty$. Instead it seems better to choose the Killing vector to be that corresponding to boosts in the background two-dimensional Minkowski space. Then the past and future horizons will be the null lines $x_\pm=0$ intersecting at the origin. One can define a radial coordinate that is left invariant by the boost as

$$r^2=-x_+x_-\,.$$

It is straightforward to verify that r is regular on a spacelike surface through the origin and has nonzero gradient there if one chooses the positive square root on one side of the intersection of the horizons at $r=0$ and the negative root on the other. In the r coordinate the field equations for a static solution are

$$\phi''+\frac{1}{r}\phi'=\left[1-\frac{N}{24}e^{2\phi}\right]\left[\rho''+\frac{1}{r}\rho'\right],$$

$$\left[1-\frac{N}{12}e^{2\phi}\right]\left[\phi''+\frac{1}{r}\phi'\right]=2\left[1-\frac{N}{24}e^{2\phi}\right][(\phi')^2-\lambda^2e^{2\rho}]\,.$$

The boundary conditions for a regular horizon are

$$\phi'=\rho'=0\,.$$

A static black-hole solution is therefore determined by the values of ϕ and ρ on the horizon. The value of ρ, however, can be changed by a constant by rescaling the coordinates $x_\pm$. The physical distinct static solutions with a horizon are therefore characterized simply by ϕ_h, the value of the dilaton on the horizon.

If $\phi_h>\phi_0$, ϕ would increase away from the horizon and would always be greater than its horizon value. This shows that to get a static black-hole solution that is asymptotic to the weak-coupling region of the linear dilaton, ϕ_h must be less than the critical value ϕ_0. One can then show that both ϕ and ρ must decrease with increasing r. This means the backreaction terms proportional to N will become unimportant. For large r one can therefore approximate by putting $N=0$. This gives

$$\phi=\rho-(2b-1)\ln r-c\,,$$

$$\phi''+(1/r)\phi'=2\{[\rho'-(2b-1)r^{-1}]^2-\lambda^2e^{2\rho}\}\,.$$

Asymptotically these have the solution

$$\rho=-\ln r+\ln\frac{2b}{\lambda}-\frac{K+L\ln r}{r^{4b}}+\cdots\,,$$

where b, c, K, L are parameters that determine the solution. The parameters b and c correspond to the coordinate freedom in the linear dilaton that the solution approaches at large r. If $L=0$, the parameter K can be related to the Arnowitt-Deser-Misner (ADM) mass M of the solution. However, if $L\neq 0$, the ADM mass will be infinite. This is what one would expect for a static black hole in equilibrium with radiation at a nonzero temperature because there will be incoming and outgoing radiation all the way to infinity. Of course a solution formed by sending in an f wave to the linear dilaton will have a finite mass. But one might hope that it would settle down to a static black-hole solution which has finite mass because there is no incoming radiation (by boundary conditions) and no outgoing radiation (because the rate of radiation has gone to zero). Indeed this is what would have to happen if the singularity were to remain hidden for all time.

For $\phi_h \ll \phi_0$, the backreaction terms will be small at all values of r and the solutions of the semiclassical equations will be almost the same as the classical black holes. So

$$\phi_h = -\tfrac{1}{2}\ln(M/\lambda),$$

where M is the mass at a finite distance from the black hole.

Consider a situation in which a black hole of large mass $(M \gg N\lambda/12)$ is created by sending in an f wave. One could approximate the subsequent evolution by a sequence of static black-hole solutions with a steadily increasing value of ϕ on the horizon. However, when the value of ϕ on the horizon approaches the critical value ϕ_0, the backreaction will become important and will change the black-hole solutions significantly. Let

$$\phi = \phi_0 + \bar{\phi}, \quad \rho = \ln\lambda + \bar{\rho}.$$

Then N and λ disappear and the equations for static black holes become

$$\phi'' + \frac{1}{r}\bar{\phi}' = \tfrac{1}{2}(2 - e^{2\bar{\phi}})\left(\bar{\rho}'' + \frac{1}{r}\bar{\rho}'\right),$$

$$(1 - e^{2\bar{\phi}})\left(\bar{\phi}'' + \frac{1}{r}\bar{\phi}'\right) = (2 - e^{2\bar{\phi}})[(\bar{\phi}')^2 - e^{2\bar{\rho}}].$$

As the dilaton field on the horizon approaches the critical value ϕ_0, the term $1 - e^{2\bar{\phi}}$ will approach 2ϵ, where $\epsilon = \phi_0 - \phi_h$. This will cause the second derivative of $\bar{\phi}$ to be very large until $\bar{\phi}'$ approaches $-e^{\bar{\rho}_h}$ in a coordinate distance Δr of order 4ϵ. By the above equations, ρ' approaches $-2e^{\bar{\rho}_h}$ in the same distance. A power series solution and numerical calculations carried out by Jonathan Brenchley confirm that in the limit as ϵ tends to zero, the solution tends to a limiting form $\bar{\phi}_c, \bar{\rho}_c$.

The limiting black hole is regular everywhere outside the horizon, but has a fairly mild singularity on the horizon with R diverging like r^{-1}. At large values of r, the solution will tend to the linear dilaton in the manner of the asymptotic expansion given before. One or both of the constants K and L must be nonzero, because the solution is not exactly the linear dilaton. Fitting to the asymptotic expansion gives a value

$$b_c \approx 0.4.$$

If the singularity inside the black hole were to remain hidden at all times, as in possibility (1) above, one might expect that the temperature and rate of evolution of the black hole would approach zero as the dilaton field on the horizon approached the critical value. However, this is not what happens. The fact that the black holes tend to the limiting solution $\bar{\phi}_c, \bar{\rho}_c$ means that the period in imaginary time will tend to $4\pi b_c/\lambda$. Thus the temperature will be

$$T_c = \lambda/4\pi b_c.$$

The energy-momentum tensor of one of the f_i fields can be calculated from the conservation equations. In the $x_\pm$ coordinates, they are

$$\langle T^f_{++}\rangle = -\tfrac{1}{12}[\partial_+\bar{\rho}\,\partial_+\bar{\rho} - \partial_+^2\bar{\rho} + t_+(x_+)],$$

$$\langle T^f_{--}\rangle = -\tfrac{1}{12}[\partial_-\bar{\rho}\,\partial_-\bar{\rho} - \partial_-^2\bar{\rho} + t_-(x_-)],$$

where $t_\pm(x_\pm)$ are chosen to satisfy the boundary conditions on the energy-momentum tensor. In the case of a black hole formed by sending in an f wave, the boundary condition is that the incoming flux $\langle T^f_{++}\rangle$ should be zero at large r. This would imply that

$$t_+ = 1/4x_+^2.$$

The energy-momentum tensor would not be regular on the past horizon, but this does not matter as the physical spacetime would not have a past horizon but would be different before the f wave.

On the other hand, the energy-momentum tensor should be regular on the future horizon. This would imply that $t_-(x_-)$ should be regular at $x_- = 0$. Converting to the coordinates $u_\pm$, one then would obtain a steady rate

$$\lambda^2/192 b_c^2$$

of energy outflow in each f field at late retarded times u_-.

In conclusion, the fact that the temperature and rate of emission of the limiting black hole do not go to zero establishes a contradiction with the idea that the black hole settles down to a stable state. Of course, this does not tell us what the semiclassical equations will predict, but it makes it very plausible that they will lead either to a naked singularity or to a singularity that spreads out to infinity at some finite retarded time.

The semiclassical evolution of these two-dimensional black holes is very similar to that of charged black holes in four dimensions with a dilaton field [5]. If one supposes that there are no fields in the theory that can carry away the charge, the steady loss of mass would suggest that the black hole would approach an extreme state. However, unlike the case of the Reissner-Nordström solutions, the extreme black holes with a dilaton have a finite temperature and rate of emission. So one obtains a similar contradiction. If the solution were to evolve to a state of lower mass but the same charge, the singularity would become naked.

There seems to be no way of avoiding a naked singularity in the context of the semiclassical theory. If spacetime is described by a semiclassical Lorentz metric, a black hole cannot disappear completely without there being some sort of naked singularity. But there seem to be zero-temperature nonradiating black holes only in a few cases, for example, charged black holes with no dilaton field and no fields to carry away the charge.

What seems to be happening is that the semiclassical approximation is breaking down in the strong-coupling regime. In conventional general relativity, this breakdown occurs only when the black hole gets down to the Planck mass. But in the two- and four-dimensional dilatonic theories, it can occur for macroscopic black holes when the dilaton field on the horizon approaches the critical value. When the coupling becomes strong, the semiclassical approximation will break down. Quantum fluctuations of the metric and the dilaton could no longer be neglected. One could imagine that this might lead to a tremendous explosion in which the remaining mass energy of the black hole was released. Such explosions might be detected as gamma-ray bursts.

Even though the semiclassical equations seem to lead to a naked singularity, one would hope that this would not happen in a full quantum treatment. Exactly what it means not to have naked singularities in a quantum theory of gravity is not immediately obvious. One possible interpretation is the no boundary condition [6]: Spacetime is nonsingular and without boundary in the Euclidean regime. If this proposal is correct, some sort of Euclidean wormhole would have to occur, which would carry away the particles that went in to form the black hole, and bring in the particles to be emitted. These wormholes could be in a coherent state described by alpha parameters [7]. These parameters might be determined by the minimization of the effective gravitational constant G [7–9]. In this case, there would be no loss of quantum coherence if a black hole were to evaporate and disappear completely or the alpha parameters might be different moments of a quantum field α on superspace [10]. In this case there would be effective loss of quantum coherence, but it might be possible to measure all the alpha parameters involved in the evaporation of a black hole of a given mass. In that case, there would be no further loss of quantum coherence when black holes of up to that mass evaporated.

I was greatly helped by talking to S. B. Giddings and A. Strominger who were working along similar lines. I also had useful discussions with G. Hayward, G. T. Horowitz, and J. Preskill. This work was carried out during a visit to Caltech as a Sherman Fairchild Distinguished Scholar. This work was supported in part by the U.S. Department of Energy under Contract No. DEAC-03-81ER40050.

[1] C. G. Callan, S. B. Giddings, J. A. Harvey, and A. Strominger, Phys. Rev. D **45**, R1005 (1992).
[2] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[3] T. Banks, A. Dabholkar, M. R. Douglas, and M. O'Loughlin, Phys. Rev. D **45**, 3607 (1992).
[4] J. G. Russo, L. Susskind, and L. Thorlacius, Report No. SU-ITP-92-4 (unpublished).
[5] D. Garfinkle, G. T. Horowitz, and A. Strominger, Phys. Rev. D **43**, 3140 (1991).
[6] J. B. Hartle and S. W. Hawking, Phys. Rev. D **28**, 2960 (1983).
[7] S. Coleman, Nucl. Phys. **B310**, 643 (1988).
[8] J. Preskill, Nucl. Phys. **B323**, 141 (1989).
[9] S. W. Hawking, Nucl. Phys. **B335**, 155 (1990).
[10] S. W. Hawking, Nucl. Phys. **B363**, 117 (1991).