Commun. math. Phys. 43, 199—220 (1975)
© by Springer-Verlag 1975

Particle Creation by Black Holes

S. W. Hawking

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, England

Received April 12, 1975

Abstract. In the classical theory black holes can only absorb and not emit particles. However it is shown that quantum mechanical effects cause black holes to create and emit particles as if they were hot bodies with temperature $\frac{\hbar\kappa}{2\pi k} \approx 10^{-6}\left(\frac{M_\odot}{M}\right)°\mathrm{K}$ where κ is the surface gravity of the black hole. This thermal emission leads to a slow decrease in the mass of the black hole and to its eventual disappearance: any primordial black hole of mass less than about 10^{15} g would have evaporated by now. Although these quantum effects violate the classical law that the area of the event horizon of a black hole cannot decrease, there remains a Generalized Second Law: $S+\frac{1}{4}A$ never decreases where S is the entropy of matter outside black holes and A is the sum of the surface areas of the event horizons. This shows that gravitational collapse converts the baryons and leptons in the collapsing body into entropy. It is tempting to speculate that this might be the reason why the Universe contains so much entropy per baryon.

1.

Although there has been a lot of work in the last fifteen years (see [1, 2] for recent reviews), I think it would be fair to say that we do not yet have a fully satisfactory and consistent quantum theory of gravity. At the moment classical General Relativity still provides the most successful description of gravity. In classical General Relativity one has a classical metric which obeys the Einstein equations, the right hand side of which is supposed to be the energy momentum tensor of the classical matter fields. However, although it may be reasonable to ignore quantum gravitational effects on the grounds that these are likely to be small, we know that quantum mechanics plays a vital role in the behaviour of the matter fields. One therefore has the problem of defining a consistent scheme in which the space-time metric is treated classically but is coupled to the matter fields which are treated quantum mechanically. Presumably such a scheme would be only an approximation to a deeper theory (still to be found) in which space-time itself was quantized. However one would hope that it would be a very good approximation for most purposes except near space-time singularities.

The approximation I shall use in this paper is that the matter fields, such as scalar, electro-magnetic, or neutrino fields, obey the usual wave equations with the Minkowski metric replaced by a classical space-time metric g_{ab}. This metric satisfies the Einstein equations where the source on the right hand side is taken to be the expectation value of some suitably defined energy momentum operator for the matter fields. In this theory of quantum mechanics in curved space-time there is a problem in interpreting the field operators in terms of annihilation and creation operators. In flat space-time the standard procedure is to decompose

the field into positive and negative frequency components. For example, if ϕ is a massless Hermitian scalar field obeying the equation $\phi_{;ab}\eta^{ab}=0$ one expresses ϕ as

$$\phi=\sum_i\{f_i\boldsymbol{a}_i+\bar{f}_i\boldsymbol{a}_i^\dagger\} \tag{1.1}$$

where the $\{f_i\}$ are a complete orthonormal family of complex valued solutions of the wave equation $f_{i;ab}\eta^{ab}=0$ which contain only positive frequencies with respect to the usual Minkowski time coordinate. The operators $\boldsymbol{a}_i$ and $\boldsymbol{a}_i^\dagger$ are interpreted as the annihilation and creation operators respectively for particles in the ith state. The vacuum state $|0\rangle$ is defined to be the state from which one cannot annihilate any particles, i.e.

$$\boldsymbol{a}_i|0\rangle=0 \quad \text{for all } i\,.$$

In curved space-time one can also consider a Hermitian scalar field operator ϕ which obeys the covariant wave equation $\phi_{;ab}g^{ab}=0$. However one cannot decompose into its positive and negative frequency parts as positive and negative frequencies have no invariant meaning in curved space-time. One could still require that the $\{f_i\}$ and the $\{\bar{f}_i\}$ together formed a complete basis for solutions of the wave equations with

$$\tfrac{1}{2}i\int_S(f_i\bar{f}_{j;a}-\bar{f}_jf_{i;a})d\Sigma^a=\delta_{ij} \tag{1.2}$$

where S is a suitable surface. However condition (1.2) does not uniquely fix the subspace of the space of all solutions which is spanned by the $\{f_i\}$ and therefore does not determine the splitting of the operator ϕ into annihilation and creation parts. In a region of space-time which was flat or asymptotically flat, the appropriate criterion for choosing the $\{f_i\}$ is that they should contain only positive frequencies with respect to the Minkowski time coordinate. However if one has a space-time which contains an initial flat region (1) followed by a region of curvature (2) then a final flat region (3), the basis $\{f_{1i}\}$ which contains only positive frequencies on region (1) will not be the same as the basis $\{f_{3i}\}$ which contains only positive frequencies on region (3). This means that the initial vacuum state $|0_1\rangle$, the state which satisfies $\boldsymbol{a}_{1i}|0_1\rangle=0$ for each initial annihilation operator $\boldsymbol{a}_{1i}$, will not be the same as the final vacuum state $|0_3\rangle$ i.e. $\boldsymbol{a}_{3i}|0_1\rangle\neq0$. One can interpret this as implying that the time dependent metric or gravitational field has caused the creation of a certain number of particles of the scalar field.

Although it is obvious what the subspace spanned by the $\{f_i\}$ is for an asymptotically flat region, it is not uniquely defined for a general point of a curved space-time. Consider an observer with velocity vector v^a at a point p. Let B be the least upper bound $|R_{abcd}|$ in any orthonormal tetrad whose timelike vector coincides with v^a. In a neighbourhood U of p the observer can set up a local inertial coordinate system (such as normal coordinates) with coordinate radius of the order of $B^{-\frac{1}{2}}$. He can then choose a family $\{f_i\}$ which satisfy equation (1.2) and which in the neighbourhood U are approximately positive frequency with respect to the time coordinate in U. For modes f_i whose characteristic frequency ω is high compared to $B^{\frac{1}{2}}$, this leaves an indeterminacy between f_i and its complex conjugate $\bar{f}_i$ of the order of the exponential of some multiple of $-\omega B^{-\frac{1}{2}}$. The indeterminacy between the annihilation operator $\boldsymbol{a}_i$ and the creation operator $\boldsymbol{a}_i^\dagger$ for the

mode is thus exponentially small. However, the ambiguity between the a_i and the $a_i^\dagger$ is virtually complete for modes for which $\omega < B^{\frac{1}{2}}$. This ambiguity introduces an uncertainty of $\pm\frac{1}{2}$ in the number operator $a_i^\dagger a_i$ for the mode. The density of modes per unit volume in the frequency interval ω to $\omega + d\omega$ is of the order of $\omega^2 d\omega$ for ω greater than the rest mass m of the field in question. Thus the uncertainty in the local energy density caused by the ambiguity in defining modes of wavelength longer than the local radius of curvature $B^{-\frac{1}{2}}$, is of order B^2 in units in which $G=c=\hbar=1$. Because the ambiguity is exponentially small for wavelengths short compared to the radius of curvature $B^{-\frac{1}{2}}$, the total uncertainty in the local energy density is of order B^2. This uncertainty can be thought of as corresponding to the local energy density of particles created by the gravitational field. The uncertainty in the curvature produced via the Einstein equations by this uncertainty in the energy density is small compared to the total curvature of space-time provided that B is small compared to one, i.e. the radius of curvature $B^{-\frac{1}{2}}$ is large compared to the Planck length 10^{-33} cm. One would therefore expect that the scheme of treating the matter fields quantum mechanically on a classical curved space-time background would be a good approximation, except in regions where the curvature was comparable to the Planck value of 10^{66} cm^{-2}. From the classical singularity theorems [3–6], one would expect such high curvatures to occur in collapsing stars and, in the past, at the beginning of the present expansion phase of the universe. In the former case, one would expect the regions of high curvature to be hidden from us by an event horizon [7]. Thus, as far as we are concerned, the classical geometry–quantum matter treatment should be valid apart from the first 10^{-43} s of the universe. The view is sometimes expressed that this treatment will break down when the radius of curvature is comparable to the Compton wavelength $\sim 10^{-13}$ cm of an elementary particle such as a proton. However the Compton wavelength of a zero rest mass particle such as a photon or a neutrino is infinite, but we do not have any problem in dealing with electromagnetic or neutrino radiation in curved space-time. All that happens when the radius of curvature of space-time is smaller than the Compton wavelength of a given species of particle is that one gets an indeterminacy in the particle number or, in other words, particle creation. However, as was shown above, the energy density of the created particles is small locally compared to the curvature which created them.

Even though the effects of particle creation may be negligible locally, I shall show in this paper that they can add up to have a significant influence on black holes over the lifetime of the universe $\sim 10^{17}$ s or 10^{60} units of Planck time. It seems that the gravitational field of a black hole will create particles and emit them to infinity at just the rate that one would expect if the black hole were an ordinary body with a temperature in geometric units of $\kappa/2\pi$, where κ is the "surface gravity" of the black hole [8]. In ordinary units this temperature is of the order of $10^{26} M^{-1}$ °K, where M is the mass, in grams of the black hole. For a black hole of solar mass (10^{33} g) this temperature is much lower than the 3 °K temperature of the cosmic microwave background. Thus black holes of this size would be absorbing radiation faster than they emitted it and would be increasing in mass. However, in addition to black holes formed by stellar collapse, there might also be much smaller black holes which were formed by density fluctua-

tions in the early universe [9, 10]. These small black holes, being at a higher temperature, would radiate more than they absorbed. They would therefore presumably decrease in mass. As they got smaller, they would get hotter and so would radiate faster. As the temperature rose, it would exceed the rest mass of particles such as the electron and the muon and the black hole would begin to emit them also. When the temperature got up to about 10^{12} °K or when the mass got down to about 10^{14} g the number of different species of particles being emitted might be so great [11] that the black hole radiated away all its remaining rest mass on a strong interaction time scale of the order of 10^{-23} s. This would produce an explosion with an energy of 10^{35} ergs. Even if the number of species of particle emitted did not increase very much, the black hole would radiate away all its mass in the order of $10^{-28}M^3$ s. In the last tenth of a second the energy released would be of the order of 10^{30} ergs.

As the mass of the black hole decreased, the area of the event horizon would have to go down, thus violating the law that, classically, the area cannot decrease [7, 12]. This violation must, presumably, be caused by a flux of negative energy across the event horizon which balances the positive energy flux emitted to infinity. One might picture this negative energy flux in the following way. Just outside the event horizon there will be virtual pairs of particles, one with negative energy and one with positive energy. The negative particle is in a region which is classically forbidden but it can tunnel through the event horizon to the region inside the black hole where the Killing vector which represents time translations is spacelike. In this region the particle can exist as a real particle with a timelike momentum vector even though its energy relative to infinity as measured by the time translation Killing vector is negative. The other particle of the pair, having a positive energy, can escape to infinity where it constitutes a part of the thermal emission described above. The probability of the negative energy particle tunnelling through the horizon is governed by the surface gravity κ since this quantity measures the gradient of the magnitude of the Killing vector or, in other words, how fast the Killing vector is becoming spacelike. Instead of thinking of negative energy particles tunnelling through the horizon in the positive sense of time one could regard them as positive energy particles crossing the horizon on past-directed world-lines and then being scattered on to future-directed world-lines by the gravitational field. It should be emphasized that these pictures of the mechanism responsible for the thermal emission and area decrease are heuristic only and should not be taken too literally. It should not be thought unreasonable that a black hole, which is an excited state of the gravitational field, should decay quantum mechanically and that, because of quantum fluctuation of the metric, energy should be able to tunnel out of the potential well of a black hole. This particle creation is directly analogous to that caused by a deep potential well in flat space-time [18]. The real justification of the thermal emission is the mathematical derivation given in Section (2) for the case of an uncharged non-rotating black hole. The effects of angular momentum and charge are considered in Section (3). In Section (4) it is shown that any renormalization of the energy-momentum tensor with suitable properties must give a negative energy flow down the black hole and consequent decrease in the area of the event horizon. This negative energy flow is non-observable locally.

The decrease in area of the event horizon is caused by a violation of the weak energy condition [5–7, 12] which arises from the indeterminacy of particle number and energy density in a curved space-time. However, as was shown above, this indeterminacy is small, being of the order of B^2 where B is the magnitude of the curvature tensor. Thus it can have a diverging effection a null surface like the event horizon which has very small convergence or divergence but it can not untrap a strongly converging trapped surface until B becomes of the order of one. Therefore one would not expect the negative energy density to cause a breakdown of the classical singularity theorems until the radius of curvature of space-time became 10^{-33} cm.

Perhaps the strongest reason for believing that black holes can create and emit particles at a steady rate is that the predicted rate is just that of the thermal emission of a body with the temperature $\kappa/2\pi$. There are independent, thermodynamic, grounds for regarding some multiple of the surface gravity as having a close relation to temperature. There is an obvious analogy with the second law of thermodynamics in the law that, classically, the area of the event horizon can never decrease and that when two black holes collide and merge together, the area of the final event horizon is greater than the sum of the areas of the two original horizons [7, 12]. There is also an analogy to the first law of thermodynamics in the result that two neighbouring black hole equilibrium states are related by [8]

$$dM = \frac{\kappa}{8\pi} dA + \Omega dJ$$

where M, Ω, and J are respectively the mass, angular velocity and angular momentum of the black hole and A is the area of the event horizon. Comparing this to

$$dU = TdS + pdV$$

one sees that if some multiple of A is regarded as being analogous to entropy, then some multiple of κ is analogous to temperature. The surface gravity is also analogous to temperature in that it is constant over the event horizon in equilibrium. Beckenstein [19] suggested that A and κ were not merely analogous to entropy and temperature respectively but that, in some sense, they actually were the entropy and temperature of the black hole. Although the ordinary second law of thermodynamics is transcended in that entropy can be lost down black holes, the flow of entropy across the event horizon would always cause some increase in the area of the horizon. Beckenstein therefore suggested [20] a Generalized Second Law: Entropy + some multiple (unspecified) of A never decreases. However he did not suggest that a black hole could emit particles as well as absorb them. Without such emission the Generalized Second Law would be violated by for example, a black hole immersed in black body radiation at a lower temperature than that of the black hole. On the other hand, if one accepts that black holes do emit particles at a steady rate, the identification of $\kappa/2\pi$ with temperature and $\frac{1}{4}A$ with entropy is established and a Generalized Second Law confirmed.

2. Gravitational Collapse

It is now generally believed that, according to classical theory, a gravitational collapse will produce a black hole which will settle down rapidly to a stationary axisymmetric equilibrium state characterized by its mass, angular momentum and electric charge [7, 13]. The Kerr-Newman solution represent one such family of black hole equilibrium states and it seems unlikely that there are any others. It has therefore become a common practice to ignore the collapse phase and to represent a black hole simply by one of these solutions. Because these solutions are stationary there will not be any mixing of positive and negative frequencies and so one would not expect to obtain any particle creation. However there is a classical phenomenon called superradiance [14–17] in which waves incident in certain modes on a rotating or charged black hole are scattered with increased amplitude [see Section (3)]. On a particle description this amplification must correspond to an increase in the number of particles and therefore to stimulated emission of particles. One would therefore expect on general grounds that there would also be a steady rate of spontaneous emission in these superradiant modes which would tend to carry away the angular momentum or charge of the black hole [16]. To understand how the particle creation can arise from mixing of positive and negative frequencies, it is essential to consider not only the quasi-stationary final state of the black hole but also the time-dependent formation phase. One would hope that, in the spirit of the "no hair" theorems, the rate of emission would not depend on details of the collapse process except through the mass, angular momentum and charge of the resulting black hole. I shall show that this is indeed the case but that, in addition to the emission in the superradiant modes, there is a steady rate of emission in all modes at the rate one would expect if the black hole were an ordinary body with temperature $\kappa/2\pi$.

I shall consider first of all the simplest case of a non-rotating uncharged black hole. The final stationary state for such a black hole is represented by the Schwarzschild solution with metric

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2+\left(1-\frac{2M}{r}\right)^{-1}dr^2+r^2(d\theta^2+\sin^2\theta d\phi^2)\,. \tag{2.1}$$

As is now well known, the apparent singularities at $r=2M$ are fictitious, arising merely from a bad choice of coordinates. The global structure of the analytically extended Schwarzschild solution can be described in a simple manner by a Penrose diagram of the r-t plane (Fig. 1) [6, 13]. In this diagram null geodesics in the r-t plane are at $\pm 45°$ to the vertical. Each point of the diagram represents a 2-sphere of area $4\pi r^2$. A conformal transformation has been applied to bring infinity to a finite distance: infinity is represented by the two diagonal lines (really null surfaces) labelled $\mathscr{I}^+$ and $\mathscr{I}^-$, and the points I^+, I^-, and I^0. The two horizontal lines $r=0$ are curvature singularities and the two diagonal lines $r=2M$ (really null surfaces) are the future and past event horizons which divide the solution up into regions from which one cannot escape to $\mathscr{I}^+$ and $\mathscr{I}^-$. On the left of the diagram there is another infinity and asymptotically flat region.

Most of the Penrose diagram is not in fact relevant to a black hole formed by gravitational collapse since the metric is that of the Schwarzchild solution

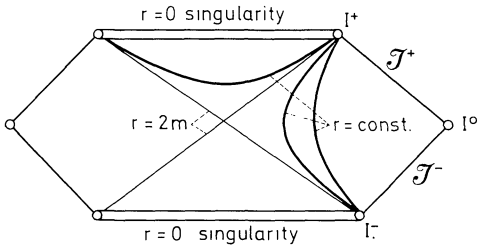


Fig. 1. The Penrose diagram for the analytically extended Schwarzschild solution

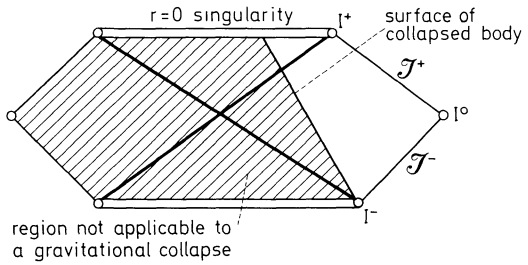


Fig. 2. Only the region of the Schwarzschild solution outside the collapsing body is relevant for a black hole formed by gravitational collapse. Inside the body the solution is completely different

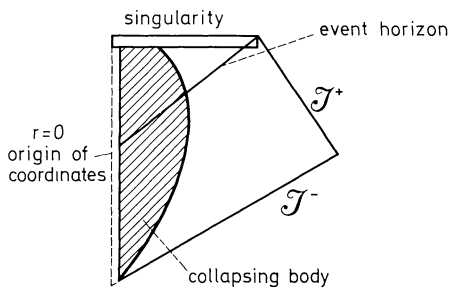


Fig. 3. The Penrose diagram of a spherically symmetric collapsing body producing a black hole. The vertical dotted line on the left represents the non-singular centre of the body

only in the region outside the collapsing matter and only in the asymptotic future. In the case of exactly spherical collapse, which I shall consider for simplicity, the metric is exactly the Schwarzschild metric everywhere outside the surface of the collapsing object which is represented by a timelike geodesic in the Penrose diagram (Fig. 2). Inside the object the metric is completely different, the past event horizon, the past $r=0$ singularity and the other asymptotically flat region do not exist and are replaced by a time-like curve representing the origin of polar coordinates. The appropriate Penrose diagram is shown in Fig. 3 where the conformal freedom has been used to make the origin of polar coordinates into a vertical line.

In this space-time consider (again for simplicity) a massless Hermitian scalar field operator ϕ obeying the wave equation

$$\phi_{;ab}g^{ab}=0\,. \tag{2.2}$$

(The results obtained would be the same if one used the conformally invariant wave equation:

$$\phi_{;ab}g^{ab}+\tfrac{1}{6}R\phi=0\,.)$$

The operator ϕ can be expressed as

$$\phi=\sum_i\{f_i\boldsymbol{a}_i+\bar{f}_i\boldsymbol{a}_i^\dagger\}\,. \tag{2.3}$$

The solutions $\{f_i\}$ of the wave equation $f_{i;ab}g^{ab}=0$ can be chosen so that on past null infinity $\mathscr{I}^-$ they form a complete family satisfying the orthonormality conditions (1.2) where the surface S is $\mathscr{I}^-$ and so that they contain only positive frequencies with respect to the canonical affine parameter on $\mathscr{I}^-$. (This last condition of positive frequency can be uniquely defined despite the existence of "supertranslations" in the Bondi-Metzner-Sachs asymptotic symmetry group [21, 22].) The operators $\boldsymbol{a}_i$ and $\boldsymbol{a}_i^\dagger$ have the natural interpretation as the annihilation and creation operators for ingoing particles i.e. for particles at past null infinity $\mathscr{I}^-$. Because massless fields are completely determined by their data on $\mathscr{I}^-$, the operator ϕ can be expressed in the form (2.3) everywhere. In the region outside the event horizon one can also determine massless fields by their data on the event horizon and on future null infinity $\mathscr{I}^+$. Thus one can also express ϕ in the form

$$\phi=\sum_i\{p_i\boldsymbol{b}_i+\bar{p}_i\boldsymbol{b}_i^\dagger+q_i\boldsymbol{c}_i+\bar{q}_i\boldsymbol{c}_i^\dagger\}\,. \tag{2.4}$$

Here the $\{p_i\}$ are solutions of the wave equation which are purely outgoing, i.e. they have zero Cauchy data on the event horizon and the $\{q_i\}$ are solutions which contain no outgoing component, i.e. they have zero Cauchy data on $\mathscr{I}^+$. The $\{p_i\}$ and $\{q_i\}$ are required to be complete families satisfying the orthonormality conditions (1.2) where the surface S is taken to be $\mathscr{I}^+$ and the event horizon respectively. In addition the $\{p_i\}$ are required to contain only positive frequencies with respect to the canonical affine parameter along the null geodesic generators of $\mathscr{I}^+$. With the positive frequency condition on $\{p_i\}$, the operators $\{\boldsymbol{b}_i\}$ and $\{\boldsymbol{b}_i^\dagger\}$ can be interpreted as the annihilation and creation operators for outgoing particles, i.e. for particles on $\mathscr{I}^+$. It is not clear whether one should impose some positive frequency condition on the $\{q_i\}$ and if so with respect to what. The choice of the $\{q_i\}$ does not affect the calculation of the emission of particles to $\mathscr{I}^+$. I shall return to the question in Section (4).

Because massless fields are completely determined by their data on $\mathscr{I}^-$ one can express $\{p_i\}$ and $\{q_i\}$ as linear combinations of the $\{f_i\}$ and $\{\bar{f}_i\}$:

$$p_i=\sum_j(\alpha_{ij}f_j+\beta_{ij}\bar{f}_j)\,, \tag{2.5}$$

$$q_i=\sum_j(\gamma_{ij}f_j+\eta_{ij}\bar{f}_j)\,. \tag{2.6}$$

These relations lead to corresponding relations between the operators

$$\boldsymbol{b}_i=\sum_j(\bar{\alpha}_{ij}\boldsymbol{a}_j-\bar{\beta}_{ij}\boldsymbol{a}_j^\dagger)\,, \tag{2.7}$$

$$\boldsymbol{c}_i=\sum_j(\bar{\gamma}_{ij}\boldsymbol{a}_j-\bar{\eta}_{ij}\boldsymbol{a}_j^\dagger)\,. \tag{2.8}$$

The initial vacuum state $|0\rangle$, the state containing no incoming particles, i.e. no particles on $\mathscr{I}^-$, is defined by

$$\boldsymbol{a}_i|0\rangle = 0 \quad \text{for all } i\,. \tag{2.9}$$

However, because the coefficients β_{ij} will not be zero in general, the initial vacuum state will not appear to be a vacuum state to an observer at $\mathscr{I}^+$. Instead he will find that the expectation value of the number operator for the ith outgoing mode is

$$\langle 0_-|\boldsymbol{b}_i^\dagger \boldsymbol{b}_i|0_-\rangle = \sum_j |\beta_{ij}|^2\,. \tag{2.10}$$

Thus in order to determine the number of particles created by the gravitational field and emitted to infinity one simply has to calculate the coefficients β_{ij}. One would expect this calculation to be very messy and to depend on the detailed nature of the gravitational collapse. However, as I shall show, one can derive an asymptotic form for the β_{ij} which depends only on the surface gravity of the resulting black hole. There will be a certain finite amount of particle creation which depends on the details of the collapse. These particles will disperse and at late retarded times on $\mathscr{I}^+$ there will be a steady flux of particles determined by the asymptotic form of β_{ij}.

In order to calculate this asymptotic form it is more convenient to decompose the ingoing and outgoing solutions of the wave equation into their Fourier components with respect to advanced or retarded time and use the continuum normalization. The finite normalization solutions can then be recovered by adding Fourier components to form wave packets. Because the space-time is spherically symmetric, one can also decompose the incoming and outgoing solutions into spherical harmonics. Thus, in the region outside the collapsing body, one can write the incoming and outgoing solutions as

$$f_{\omega' lm} = (2\pi)^{-\frac{1}{2}} r^{-1} (\omega')^{-\frac{1}{2}} F_{\omega'}(r) e^{i\omega' v} Y_{lm}(\theta, \phi)\,, \tag{2.11}$$

$$p_{\omega lm} = (2\pi)^{-\frac{1}{2}} r^{-1} \omega^{-\frac{1}{2}} P_\omega(r) e^{i\omega u} Y_{lm}(\theta, \phi)\,, \tag{2.12}$$

where v and u are the usual advanced and retarded coordinates defined by

$$v = t + r + 2M \log\left|\frac{r}{2M} - 1\right|, \tag{2.13}$$

$$u = t - r - 2M \log\left|\frac{r}{2M} - 1\right|. \tag{2.14}$$

Each solution $p_{\omega lm}$ can be expressed as an integral with respect to ω' over solutions $f_{\omega' lm}$ and $\bar{f}_{\omega' lm}$ with the same values of l and $|m|$ (from now on I shall drop the suffices l, m):

$$p_\omega = \int_0^\infty (\alpha_{\omega\omega'} f_{\omega'} + \beta_{\omega\omega'} \bar{f}_{\omega'}) d\omega'\,. \tag{2.15}$$

To calculate the coefficients $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$, consider a solution p_ω propagating backwards from $\mathscr{I}^+$ with zero Cauchy data on the event horizon. A part $p_\omega^{(1)}$ of the solution p_ω will be scattered by the static Schwarzchild field outside the collapsing body and will end up on $\mathscr{I}^-$ with the same frequency ω. This will give a $\delta(\omega' - \omega)$ term in $\alpha_{\omega\omega'}$. The remainder $p_\omega^{(2)}$ of p_ω will enter the collapsing body

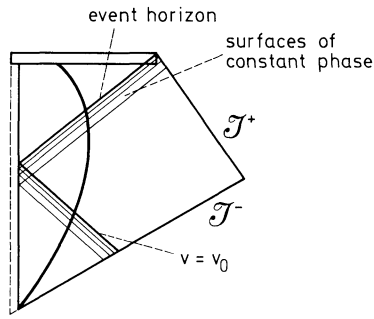


Fig. 4. The solution p_ω of the wave equation has an infinite number of cycles near the event horizon and near the surface $v=v_0$

where it will be partly scattered and partly reflected through the centre, eventually emerging to $\mathscr{I}^-$. It is this part $p_\omega^{(2)}$ which produces the interesting effects. Because the retarded time coordinate u goes to infinity on the event horizon, the surfaces of constant phase of the solution p_ω will pile up near the event horizon (Fig. 4). To an observer on the collapsing body the wave would seem to have a very large blue-shift. Because its effective frequency was very high, the wave would propagate by geometric optics through the centre of the body and out on $\mathscr{I}^-$. On $\mathscr{I}^-$ $p_\omega^{(2)}$ would have an infinite number of cycles just before the advanced time $v=v_0$ where v_0 is the latest time that a null geodesic could leave $\mathscr{I}^-$, pass through the centre of the body and escape to $\mathscr{I}^+$ before being trapped by the event horizon. One can estimate the form of $p_\omega^{(2)}$ on $\mathscr{I}^-$ near $v=v_0$ in the following way. Let x be a point on the event horizon outside the matter and let l^a be a null vector tangent to the horizon. Let n^a be the future-directed null vector at x which is directed radially inwards and normalized so that $l^a n_a = -1$. The vector $-\varepsilon n^a$ (ε small and positive) will connect the point x on the event horizon with a nearby null surface of constant retarded time u and therefore with a surface of constant phase of the solution $p_\omega^{(2)}$. If the vectors l^a and n^a are parallelly transported along the null geodesic γ through x which generates the horizon, the vector $-\varepsilon n^a$ will always connect the event horizon with the same surface of constant phase of $p_\omega^{(2)}$. To see what the relation between ε and the phase of $p_\omega^{(2)}$ is, imagine in Fig. 2 that the collapsing body did not exist but one analytically continued the empty space Schwarzchild solution back to cover the whole Penrose diagram. One could then transport the pair (l^a, n^a) back along to the point where future and past event horizons intersected. The vector $-\varepsilon n^a$ would then lie along the past event horizon. Let λ be the affine parameter along the past event horizon which is such that at the point of intersection of the two horizons, $\lambda=0$ and $\dfrac{dx^a}{d\lambda}=n^a$. The affine parameter λ is related to the retarded time u on the past horizon by

$$\lambda = -Ce^{-\kappa u} \tag{2.16}$$

where C is constant and κ is the surface gravity of the black hole defined by $K^a_{;b}K^b = -\kappa K^a$ on the horizon where K^a is the time translation Killing vector.

$\left(\text{For a Schwarzchild black hole } \kappa = \frac{1}{4M}\right)$. It follows from this that the vector $-\varepsilon n^a$ connects the future event horizon with the surface of constant phase $-\frac{\omega}{\kappa}(\log\varepsilon - \log C)$ of the solution $p^{(2)}_\omega$. This result will also hold in the real space-time (including the collapsing body) in the region outside the body. Near the event horizon the solution $p^{(2)}_\omega$ will obey the geometric optics approximation as it passes through the body because its effective frequency will be very high. This means that if one extends the null geodesic γ back past the end-point of the event horizon and out onto $\mathscr{I}^-$ at $v = v_0$ and parallelly transports n^a along γ, the vector $-\varepsilon n^a$ will still connect γ to a surface of constant phase of the solution $p^{(2)}_\omega$. On $\mathscr{I}^-$ n^a will be parallel to the Killing vector K^a which is tangent to the null geodesic generators of $\mathscr{I}^-$:

$$n^a = DK^a .$$

Thus on $\mathscr{I}^-$ for $v_0 - v$ small and positive, the phase of the solution will be

$$-\frac{\omega}{\kappa}(\log(v_0 - v) - \log D - \log C) . \tag{2.17}$$

Thus on $\mathscr{I}^-$ $p^{(2)}_\omega$ will be zero for $v > v_0$ and for $v < v_0$

$$p^{(2)}_\omega \sim (2\pi)^{-\frac{1}{2}} \omega^{-\frac{1}{2}} r^{-1} P^-_\omega \exp\left(-i\frac{\omega}{\kappa}\left(\log\left(\frac{v_0 - v}{CD}\right)\right)\right) \tag{2.18}$$

where $P^-_\omega \equiv P_\omega(2M)$ is the value of the radial function for P_ω on the past event horizon in the analytically continued Schwarzchild solution. The expression (2.18) for $p^{(2)}_\omega$ is valid only for $v_0 - v$ small and positive. At earlier advanced times the amplitude will be different and the frequency measured with respect to v, will approach the original frequency ω.

By Fourier transforming $p^{(2)}_\omega$ one can evaluate its contributions to $\alpha_{\omega\omega'}$ and $\beta_{\omega\omega'}$. For large values of ω' these will be determined by the asymptotic form (2.18). Thus for large ω'

$$\alpha^{(2)}_{\omega\omega'} \approx (2\pi)^{-1} P^-_\omega (CD)^{\frac{i\omega}{\kappa}} \exp(i(\omega - \omega')v_0)\left(\frac{\omega'}{\omega}\right)^{\frac{1}{2}} \Gamma\left(1 - \frac{i\omega}{\kappa}\right)(-i\omega')^{-1+\frac{i\omega}{\kappa}} , \tag{2.19}$$

$$\beta^{(2)}_{\omega\omega'} \approx -i\alpha^{(2)}_{\omega(-\omega')} . \tag{2.20}$$

The solution $p^{(2)}_\omega$ is zero on $\mathscr{I}^-$ for large values of v. This means that its Fourier transform is analytic in the upper half ω' plane and that $p^{(2)}_\omega$ will be correctly represented by a Fourier integral in which the contour has been displaced into the upper half ω' plane. The Fourier transform of $p^{(2)}_\omega$ contains a factor $(-i\omega')^{-1+\frac{i\omega}{\kappa}}$ which has a logarithmic singularity at $\omega' = 0$. To obtain $\beta^{(2)}_{\omega\omega'}$ from $\alpha^{(2)}_{\omega\omega'}$ by (2.20) one has to analytically continue $\alpha^{(2)}_{\omega\ \omega'}$ anticlockwise round this singularity. This means that

$$|\alpha^{(2)}_{\omega\omega'}| = \exp\left(\frac{\pi\omega}{\kappa}\right)|\beta^{(2)}_{\omega\omega'}| . \tag{2.21}$$

Actually, the fact that $p_\omega^{(2)}$ is not given by (2.18) at early advanced times means that the singularity in $\alpha_{\omega\omega'}$ occurs at $\omega'=\omega$ and not at $\omega'=0$. However the relation (2.21) is still valid for large ω'.

The expectation value of the total number of created particles at $\mathscr{I}^+$ in the frequency range ω to $\omega+d\omega$ is $d\omega\int_0^\infty|\beta_{\omega\omega'}|^2d\omega'$. Because $|\beta_{\omega\omega'}|$ goes like $(\omega')^{-\frac{1}{2}}$ at large ω' this integral diverges. This infinite total number of created particles corresponds to a finite steady rate of emission continuing for an infinite time as can be seen by building up a complete orthonormal family of wave packets from the Fourier components p_ω. Let

$$p_{jn}=\varepsilon^{-\frac{1}{2}}\int_{j\varepsilon}^{(j+1)\varepsilon}e^{-2\pi in\varepsilon^{-1}\omega}p_\omega d\omega \tag{2.22}$$

where j and n are integers, $j\geqq 0$, $\varepsilon>0$. For ε small these wave packets will have frequency $j\varepsilon$ and will be peaked around retarded time $u=2\pi n\varepsilon^{-1}$ with width ε^{-1}. One can expand $\{p_{jn}\}$ in terms of the $\{f_\omega\}$

$$p_{jn}=\int_0^\infty(\alpha_{jn\omega'}f_{\omega'}+\beta_{jn\omega'}\bar{f}_{\omega'})d\omega' \tag{2.23}$$

where

$$\alpha_{jn\omega'}=\varepsilon^{-\frac{1}{2}}\int_{j\varepsilon}^{(j+1)\varepsilon}e^{-2\pi in\varepsilon^{-1}\omega}\alpha_{\omega\omega'}d\omega \qquad \text{etc.} \tag{2.24}$$

For $j\gg\varepsilon$, $n\gg\varepsilon$

$$\begin{aligned}|\alpha_{jn\omega'}|&=\left|(2\pi)^{-1}P_\omega^-\omega^{-\frac{1}{2}}\Gamma\left(1-\frac{i\omega}{\kappa}\right)\varepsilon^{-\frac{1}{2}}(\omega')^{-\frac{1}{2}}\right.\\&\qquad\left.\cdot\int_{j\varepsilon}^{(j+1)\varepsilon}\exp i\omega''(-2\pi n\varepsilon^{-1}+\kappa^{-1}\log\omega')d\omega''\right|\\&=\left|\pi^{-1}P_\omega^-\omega^{-\frac{1}{2}}\Gamma\left(1-\frac{i\omega}{\kappa}\right)\varepsilon^{-\frac{1}{2}}(\omega')^{-\frac{1}{2}}z^{-1}\sin\tfrac{1}{2}\varepsilon z\right|\end{aligned} \tag{2.25}$$

where $\omega=j\varepsilon$ and $z=\kappa^{-1}\log\omega'-2\pi n\varepsilon^{-1}$. For wave-packets which reach $\mathscr{I}^+$ at late retarded times, i.e. those with large values of n, the main contribution to $\alpha_{jn\omega'}$ and $\beta_{jn\omega'}$ come from very high frequencies ω' of the order of $\exp(2\pi n\kappa\varepsilon^{-1})$. This means that these coefficients are governed only by the asymptotic forms (2.19, 2.20) for high ω' which are independent of the details of the collapse.

The expectation value of the number of particles created and emitted to infinity $\mathscr{I}^+$ in the wave-packet mode p_{jn} is

$$\int_0^\infty|\beta_{jn\omega'}|^2d\omega'\,. \tag{2.26}$$

One can evaluate this as follows. Consider the wave-packet p_{jn} propagating backwards from $\mathscr{I}^+$. A fraction $1-\Gamma_{jn}$ of the wave-packet will be scattered by the static Schwarzchild field and a fraction Γ_{jn} will enter the collapsing body.

$$\Gamma_{jn}=\int_0^\infty(|\alpha_{jn\omega'}^{(2)}|^2-|\beta_{jn\omega'}^{(2)}|^2)d\omega' \tag{2.27}$$

where $\alpha_{jn\omega'}^{(2)}$ and $\beta_{jn\omega'}^{(2)}$, are calculated using (2.19, 2.20) from the part $p_{jn}^{(2)}$ of the wave-packet which enters the star. The minus sign in front of the second term on the right of (2.27) occurs because the negative frequency components of $p_{jn}^{(2)}$ make a negative contribution to the flux into the collapsing body. By (2.21)

$$|\alpha_{jn\omega'}^{(2)}|=\exp(\pi\omega\kappa^{-1})|\beta_{jn\omega'}^{(2)}|\,. \tag{2.28}$$

Thus the total number of particles created in the mode p_{jn} is

$$\Gamma_{jn}(\exp(2\pi\omega\kappa^{-1})-1)^{-1}\,. \tag{2.29}$$

But for wave-packets at late retarded times, the fraction Γ_{jn} which enters the collapsing body is almost the same as the fraction of the wave-packet that would have crossed the past event horizon had the collapsing body not been there but the exterior Schwarzchild solution had been analytically continued. Thus this factor Γ_{jn} is also the same as the fraction of a similar wave-packet coming from $\mathscr{I}^-$ which would have crossed the future event horizon and have been absorbed by the black hole. The relation between emission and absorption cross-section is therefore exactly that for a body with a temperature, in geometric units, of $\kappa/2\pi$.

Similar results hold for the electromagnetic and linearised gravitational fields. The fields produced on $\mathscr{I}^-$ by positive frequency waves from $\mathscr{I}^+$ have the same asymptotic form as (2.18) but with an extra blue shift factor in the amplitude. This extra factor cancels out in the definition of the scalar product so that the asymptotic forms of the coefficients α and β are the same as in the Eqs. (2.19) and (2.20). Thus one would expect the black hole also to radiate photons and gravitons thermally. For massless fermions such as neutrinos one again gets similar results except that the negative frequency components given by the coefficients β now make a positive contribution to the probability flux into the collapsing body. This means that the term $|\beta|^2$ in (2.27) now has the opposite sign. From this it follows that the number of particles emitted in any outgoing wave packet mode is $(\exp(2\pi\omega\kappa^{-1})+1)^{-1}$ times the fraction of that wave packet that would have been absorbed by the black hole had it been incident from $\mathscr{I}^-$. This is again exactly what one would expect for thermal emission of particles obeying Fermi-Dirac statistics.

Fields of non-zero rest mass do not reach $\mathscr{I}^-$ and $\mathscr{I}^+$. One therefore has to describe ingoing and outgoing states for these fields in terms of some concept such as the projective infinity of Eardley and Sachs [23] and Schmidt [24]. However, if the initial and final states are asymptotically Schwarzchild or Kerr solutions, one can describe the ingoing and outgoing states in a simple manner by separation of variables and one can define positive frequencies with respect to the time translation Killing vectors of these initial and final asymptotic space-times. In the asymptotic future there will be no bound states: any particle will either fall through the event horizon or escape to infinity. Thus the unbound outgoing states and the event horizon states together form a complete basis for solutions of the wave equation in the region outside the event horizon. In the asymptotic past there could be bound states if the body that collapses had had a bounded radius for an infinite time. However one could equally well assume that the body had collapsed from an infinite radius in which case there would be no bound states. The possible existence of bound states in the past does not affect the rate of particle emission in the asymptotic future which will again be that of a body with temperature $\kappa/2\pi$. The only difference from the zero rest mass case is that the frequency ω in the thermal factor $(\exp(2\pi\omega\kappa^{-1})\mp 1)^{-1}$ now includes the rest mass energy of the particle. Thus there will not be much emission of particles of rest mass m unless the temperature $\kappa/2\pi$ is greater than m.

One can show that these results on thermal emission do not depend on spherical symmetry. Consider an asymmetric collapse which produced a black hole which settled to a non-rotating uncharged Schwarzchild solution (angular momentum and charge will be considered in the next section). The fact that the final state is asymptotically quasi-stationary means that there is a preferred Bondi coordinate system [25] on $\mathscr{I}^+$ with respect to which one can decompose the Cauchy data for the outgoing states into positive frequencies and spherical harmonics. On $\mathscr{I}^-$ there may or may not be a preferred coordinate system but if there is not one can pick an arbitrary Bondi coordinate system and decompose the Cauchy data for the ingoing states in a similar manner. Now consider one of the $\mathscr{I}^+$ states $p_{\omega lm}$ propagating backwards through this space-time into the collapsing body and out again onto $\mathscr{I}^-$. Take a null geodesic generator γ of the event horizon and extend it backwards beyond its past end-point to intersect $\mathscr{I}^-$ at a point y on a null geodesic generator λ of $\mathscr{I}^-$. Choose a pair of null vectors $(l^a, \hat{n}^a)$ at y with l^a tangent to γ and $\hat{n}^a$ tangent to λ. Parallelly propagate $l^a, \hat{n}^a$ along γ to a point x in the region of space-time where the metric is almost that of the final Schwarzchild solution. At x $\hat{n}^a$ will be some linear combination of l^a and the radial inward directed null vector n^a. This means that the vector $-\varepsilon\hat{n}^a$ will connect x to a surface of phase $-\omega/\kappa\ (\log\varepsilon - \log E)$ of the solution $p_{\omega lm}$ where E is some constant. As before, by the geometric optics approximation, the vector $-\varepsilon\hat{n}^a$ at y will connect y to a surface of phase $-\omega/\kappa\ (\log\varepsilon - \log E)$ of $p^{(2)}_{\omega lm}$ where $p^{(2)}_{\omega lm}$ is the part of $p_{\omega lm}$ which enters the collapsing body. Thus on the null geodesic generator λ of $\mathscr{I}^-$, the phase of $p^{(2)}_{\omega lm}$ will be

$$-\frac{i\omega}{\kappa}\left(\log(v_0 - v) - \log H\right) \tag{2.30}$$

where v is an affine parameter on λ with value v_0 at y and H is a constant. By the geometrical optics approximation, the value of $p^{(2)}_{\omega lm}$ on λ will be

$$L\exp\left\{-\frac{i\omega}{\kappa}\left[\log(v_0 - v) - \log H\right]\right\} \tag{2.31}$$

for $v_0 - v_v$ small and positive and zero for $v > v_0$ where L is a constant. On each null geodesic generator of $\mathscr{I}^- p^{(2)}_{\omega lm}$ will have the form (2.31) with different values of L, v_0, and H. The lack of spherical symmetry during the collapse will cause $p^{(2)}_{\omega lm}$ on $\mathscr{I}^-$ to contain components of spherical harmonics with indices (l', m') different from (l, m). This means that one now has to express $p^{(2)}_{\omega lm}$ in the form

$$p^{(2)}_{\omega lm} = \sum_{l'm'} \int_0^\infty \{\alpha^{(2)}_{\omega lm\omega' l'm'} f_{\omega' l'm'} + \beta^{(2)}_{\omega lm\omega' l'm'} \bar{f}_{\omega' l'm'}\} d\omega' . \tag{2.32}$$

Because of (2.31), the coefficients $\alpha^{(2)}$ and $\beta^{(2)}$ will have the same ω' dependence as in (2.19) and (2.20). Thus one still has the same relation as (2.21):

$$|\alpha^{(2)}_{\omega lm\omega' l'm'}| = \exp(\pi\omega\kappa^{-1})\,|\beta^{(2)}_{\omega lm\omega' l'm'}| . \tag{2.33}$$

As before, for each (l, m), one can make up wave packets p_{jnlm}. The number of particles emitted in such a wave packet mode is

$$\sum_{l',m'} \int_0^\infty |\beta_{jnlm\omega' l'm'}|^2 d\omega' . \tag{2.34}$$

Similarly, the fraction Γ_{jnlm} of the wave packet that enters the collapsing body is

$$\Gamma_{jnlm}=\sum_{l',m'}\int_0^\infty \{|\alpha^{(2)}_{jnlm\omega' l' m'}|^2-|\beta^{(2)}_{jnlm\omega' l' m'}|^2\}d\omega' . \tag{2.35}$$

Again, Γ_{jnlm} is equal to the fraction of a similar wave packet coming from $\mathscr{I}^-$ that would have been absorbed by the black hole. Thus, using (2.33), one finds that the emission is just that of a body of temperature $\kappa/2\pi$: the emission at late retarded times depends only on the final quasi-stationary state of the black hole and not on the details of the gravitational collapse.

3. Angular Momentum and Charge

If the collapsing body was rotating or electrically charged, the resulting black hole would settle down to a stationary state which was described, not by the Schwarzchild solution, but by a charged Kerr solution characterised by the mass M, the angular momentum J, and the charge Q. As these solutions are stationary and axisymmetric, one can separate solutions of the wave equations in them into a factor $e^{i\omega u}$ or $e^{i\omega v}$ times $e^{-im\phi}$ times a function of r and θ. In the case of the scalar wave equation one can separate this last expression into a function of r times a function of θ [26]. One can also completely separate any wave equation in the non-rotating charged case and Teukolsky [27] has obtained completely separable wave equations for neutrino, electromagnetic and linearised gravitational fields in the uncharged rotating case.

Consider a wave packet of a classical field of charge e with frequency ω and axial quantum number m incident from infinity on a Kerr black hole. The change in mass dM of the black hole caused by the partial absorption of the wave packet will be related to the change in area, angular momentum and charge by the classical first law of black holes:

$$dM=\frac{\kappa}{8\pi}dA+\Omega dJ+\Phi dQ \tag{3.1}$$

where Ω and Φ are the angular frequency and electrostatic potential respectively of the black hole [13]. The fluxes of energy, angular momentum and charge in the wave packet will be in the ratio $\omega:m:e$. Thus the changes in the mass, angular momentum and charge of the black hole will also be in this ratio. Therefore

$$dM(1-\Omega m\omega^{-1}-e\Phi\omega^{-1})=\frac{\kappa}{8\pi}dA . \tag{3.2}$$

A wave packet of a classical Boson field will obey the weak energy condition: the local energy density for any observer is non-negative. It follows from this [7, 12] that the change in area dA induced by the wave-packet will be non-negative. Thus if

$$\omega<m\Omega+e\Phi \tag{3.3}$$

the change in mass dM of the black hole must be negative. In other words, the black hole will lose energy to the wave packet which will therefore be scattered with the same frequency but increased amplitude. This is the phenomenon known as "superradiance".

For classical fields of half-integer spin, detailed calculations [28] show that there is no superradiance. The reason for this is that the scalar product for half-integer spin fields is positive definite unlike that for integer spins. This means that the probability flux across the event horizon is positive and therefore, by conservation of probability, the probability flux in the scattered wave packet must be less than that in the incident wave packet. The reason that the above argument based on the first law breaks down is that the energy-momentum tensor for a classical half-integer spin field does not obey the weak energy condition. On a quantum, particle level one can understand the absence of superradiance for fermion fields as a consequence of the fact that the Exclusion Principle does not allow more than one particle in each outgoing wave packet mode and therefore does not allow the scattered wave-packet to be stronger than the incident wave-packet.

Passing now to the quantum theory, consider first the case of an unchanged, rotating black hole. One can as before pick an arbitrary Bondi coordinate frame on $\mathscr{I}^-$ and decompose the operator ϕ in terms of a family $\{f_{\omega lm}\}$ of incoming solutions where the indices ω, l, and m refer to the advanced time and angular dependence of f on $\mathscr{I}^-$ in the given coordinate system. On $\mathscr{I}^+$ the final quasi-stationary state of the black hole defines a preferred Bondi coordinate system using which one can define a family $\{p_{\omega lm}\}$ of outgoing solutions. The index l in this case labels the spheroidal harmonics in terms of which the wave equation is separable. One proceeds as before to calculate the asymptotic form of $p^{(2)}_{\omega lm}$ on $\mathscr{I}^-$. The only difference is that because the horizon is rotating with angular velocity Ω with respect to $\mathscr{I}^+$, the effective frequency near a generator of the event horizon is not ω but $\omega - m\Omega$. This means that the number of particles emitted in the wave-packet mode p_{jnlm} is

$$\{\exp(2\pi\kappa^{-1}(\omega - m\Omega)) \mp 1\}^{-1} \Gamma_{jnlm} . \tag{3.4}$$

The effect of this is to cause the rate of emission of particles with positive angular momentum m to be higher than that of particles with the same frequency ω and quantum number l but with negative angular momentum $-m$. Thus the particle emission tends to carry away the angular momentum. For Boson fields, the factor in curly brackets in (3.4) is negative for $\omega < m\Omega$. However the fraction Γ_{jnlm} of the wave-packet that would have been absorbed by the black hole is also negative in this case because $\omega < m\Omega$ is the condition for superradiance. In the limit that the temperature $\kappa/2\pi$ is very low, the only particle emission occurs is an amount $\mp \Gamma_{jnlm}$ in the modes for which $\omega < m\Omega$. This amount of particle creation is equal to that calculated by Starobinski [16] and Unruh [29], who considered only the final stationary Kerr solution and ignored the gravitational collapse.

One can treat a charged non-rotating black hole in a rather similar way. The behaviour of fields like the electromagnetic or gravitational fields which do not carry an electric charge will be the same as before except that the charge on the black will reduce the surface gravity k and hence the temperature of the black hole. Consider now the simple case of a massless charged scalar field ϕ which obeys the minimally coupled wave equation

$$g^{ab}(\nabla_a - ieA_a)(\nabla_b - ieA_b)\phi = 0 . \tag{3.5}$$

The phase of a solution p_ω of the wave equation (3.5) is not gauge-invariant but the propagation vector $ik_a = \nabla_a(\log p_\omega) - ieA_a$ is. In the geometric optics or WKB limit the vector k_a is null and propagates according to

$$k_{a;b}k^b = -eF_{ab}k^b . \tag{3.6}$$

An infinitessimal vector z^a will connect points with a "guage invariant" phase difference of $ik_a z^a$. If z^a is propagated along the integral curves of k^a according to

$$z^a_{;b}k^b = -eF^a_b z^b \tag{3.7}$$

z^a will connect surfaces of constant guage invariant phase difference.

In the final stationary region one can choose a guage such that the electromagnetic potential A_a is stationary and vanishes on $\mathscr{I}^+$. In this guage the field equation (3.5) is separable and has solutions p_ω with retarded time dependence $e^{i\omega u}$. Let x be a point on the event horizon in the final stationary region and let l^a and n^a be a pair of null vectors at x. As before, the vector $-\varepsilon n^a$ will connect the event horizon with the surface of actual phase $-\omega/\kappa\ (\log\varepsilon - \log C)$ of the solution p_ω. However the guage invariant phase will be $-\kappa^{-1}(\omega - e\Phi)(\log\varepsilon - \log C)$ where $\Phi = K^a A_a$ is the electrostatic potential on the horizon and K^a is the time-translation Killing vector. Now propagate l^a like k^a in Eq. (3.6) back until it intersects a generator λ of $\mathscr{I}^-$ at a point y and propagate n^a like z^a in Eq. (3.7) along the integral curve of l^a. With this propagation law, the vector $-\varepsilon n^a$ will connect surfaces of constant guage invariant phase. Near $\mathscr{I}^-$ one can use a different electromagnetic guage such that A^a is zero on $\mathscr{I}^-$. In this guage the phase of $p^{(2)}_\omega$ along each generator of $\mathscr{I}^-$ will have the form

$$-(\omega - e\phi)\kappa^{-1}\{\log(v_0 - v) - \log H\} \tag{3.8}$$

where H is a constant along each generator. This phase dependence gives the same thermal emission as before but with ω replaced by $\omega - e\Phi$. Similar remarks apply about charge loss and superradiance. In the case that the black hole is both rotating and charged one can simply combine the above results.

4. The Back-Reaction on the Metric

I now come to the difficult problem of the back-reaction of the particle creation on the metric and the consequent slow decrease of the mass of the black hole. At first sight it might seem that since all the time dependence of the metric in Fig. 4 is in the collapsing phase, all the particle creation must take place in the collapsing body just before the formation of the event horizon, and that an infinite number of created particles would hover just outside the event horizon, escaping to $\mathscr{I}^+$ at a steady rate. This does not seem reasonable because it would involve the collapsing body knowing just when it was about to fall through the event horizon whereas the position of the event horizon is determined by the whole future history of the black hole and may be someway outside the apparent horizon, which is the only thing that can be determined locally [7].

Consider an observer falling through the horizon at some time after the collapse. He can set up a local inertial coordinate patch of radius $\sim M$ centred

on the point where he crosses the horizon. He can pick a complete family $\{h_\omega\}$ of solutions of the wave equations which obey the condition:

$$\tfrac{1}{2}i\int_S (h_{\omega_1}\bar{h}_{\omega_2;a}-\bar{h}_{\omega_2}h_{\omega_1;a})d\Sigma^a=\delta(\omega_1-\omega_2) \tag{4.1}$$

(where S is a Cauchy surface) and which have the approximate coordinate dependence $e^{i\omega t}$ in the coordinate patch. This last condition determines the splitting into positive and negative frequencies and hence the annihilation and creation operators fairly well for modes h_ω with $\omega>M$ but not for those with $\omega<M$. Because the $\{h_\omega\}$, unlike the $\{p_\omega\}$, are continuous across the event horizon, they will also be continuous on $\mathscr{I}^-$. It is the discontinuity in the $\{p_\omega\}$ on $\mathscr{I}^-$ at $v=v_0$ which is responsible for creating an infinite total number of particles in each mode. p_ω by producing an $(\omega')^{-1}$ tail in the Fourier transforms of the $\{p_\omega\}$ at large negative frequencies ω'. On the other hand, the $\{h_\omega\}$ for $\omega>M$ will have very small negative frequency components on $\mathscr{I}^-$. This means that the observer at the event horizon will see few particles with $\omega>M$. He will not be able to detect particles with $\omega<M$ because they will have a wavelength bigger than his particle detector which must be smaller than M. As described in the introduction, there will be an indeterminacy in the energy density of order M^{-4} corresponding to the indeterminacy in the particle number for these modes.

The above discussion shows that the particle creation is really a global process and is not localised in the collapse: an observer falling through the event horizon would not see an infinite number of particles coming out from the collapsing body. Because it is a non-local process, it is probably not reasonable to expect to be able to form a local energy-momentum tensor to describe the back-reaction of the particle creation on the metric. Rather, the negative energy density needed to account for the decrease in the area of the horizon, should be thought of as arising from the indeterminacy of order of M^{-4} of the local energy density at the horizon. Equivalently, one can think of the area decrease as resulting from the fact that quantum fluctuations of the metric will cause the position and the very concept of the event horizon to be somewhat indeterminate.

Although it is probably not meaningful to talk about the local energy-momentum of the created particles, one may still be able to define the total energy flux over a suitably large surface. The problem is rather analogous to that of defining gravitational energy in classical general relativity: there are a number of different energy-momentum pseudo-tensors, none of which have any invariant local significance, but which all agree when integrated over a sufficiently large surface. In the particle case there are similarly a number of different expressions one can use for the renormalised energy-momentum tensor. The energy-momentum tensor for a classical field ϕ is

$$T_{ab}=\phi_{;a}\phi_{;b}-\tfrac{1}{2}g_{ab}g^{cd}\phi_{;c}\phi_{;d}\,. \tag{4.2}$$

If one takes this expression over into the quantum theory and regards the ϕ's as operators one obtains a divergent result because there is a creation operator for each mode to the right of an annihilation operator. One therefore has to subtract out the divergence in some way. Various methods have been proposed for this (e.g. [30]) but they all seem a bit ad hoc. However, on the analogy of the pseudo-tensor, one would hope that the different renormalisations would all give the

same integrated fluxes. This is indeed the case in the final quasi-stationary region: all renormalised energy-momentum operators T_{ab} which obey the conservation equations $T^{ab}_{;b}=0$, which are stationary i.e. which have zero Lie derivative with respect to the time translation Killing vector K^a and which agree near $\mathscr{I}^+$ will give the same fluxes of energy and angular momentum over any surface of constant r outside the event horizon. It is therefore sufficient to evaluate the energy flux near $\mathscr{I}^+$: by the conservation equations this will be equal to the energy flux out from the event horizon. Near $\mathscr{I}^+$ the obvious way to renormalise the energy-momentum operator is to normal order the expression (4.2) with respect to positive and negative frequencies defined by the time-translation Killing vector K^a of the final quasi-stationary state. Near the event horizon normal ordering with respect to K^a cannot be the correct way to renormalise the energy-momentum operator since the normal-ordered operator diverges at the horizon. However it still gives the same energy outflow across any surface of constant r. A renormalised operator which was regular at the horizon would have to violate the weak energy condition by having negative energy density. This negative energy density is not observable locally.

In order to evaluate the normal ordered operator one wants to choose the $\{q_i\}$ which describe waves crossing the event horizon, to be positive frequency with respect to the time parameter defined by K^a along the generators of the horizon in the final quasi-stationary state. The condition on the $\{q_i\}$ in the time-dependent collapse phase is not determined but this should not affect wave packets on the horizon at late times. If one makes up wave-packets $\{q_{jn}\}$ like the $\{p_{jn}\}$, one finds that a fraction Γ_{jn} penetrates through the potential barrier around the black hole and gets out to $\mathscr{I}^-$ with the same frequency ω that it had on the horizon. This produces a $\delta(\omega-\omega')$ behaviour in $\gamma_{jn\omega'}$. The remaining fraction $1-\Gamma_{jn}$ of the wave-packet is reflected back by the potential barrier and passes through the collapsing body and out onto $\mathscr{I}^-$. Here it will have a similar form to $p^{(2)}_{jn}$. Thus for large ω',

$$|\gamma^{(2)}_{jn\omega'}|=\exp(\pi\omega\kappa^{-1})|\eta^{(2)}_{jn\omega'}|\,. \tag{4.3}$$

By a similar argument to that used in Section (2) one would conclude that the number of particles crossing the event horizon in a wave-packet mode peaked at late times would be

$$(1-\Gamma_{jn})\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}\,. \tag{4.4}$$

For a given frequency ω, i.e. a given value of j, the absorption fraction Γ_{jn} goes to zero as the angular quantum number l increases because of the centrifugal barrier. Thus at first sight it might seem that each wave-packet mode of high l value would contain

$$\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}$$

particles and that the total rate of particles and energy crossing the event horizon would be infinite. This calculation would, of course, be inconsistent with the result obtained above that an observer crossing the event horizon would see only a finite small energy density of order M^{-4}. The reason for this discrepancy seems to be that the wave-packets $\{p_{jn}\}$ and $\{q_{jn}\}$ provide a complete basis for solutions

of the wave equation only in the region outside the event horizon and not actually on the event horizon itself. In order to calculate the particle flux over the horizon one therefore has to calculate the flux over some surface just outside the horizon and take the limit as the surface approaches the horizon.

To perform this calculation it is convenient to define new wave-packets $x_{jn}=p^{(2)}_{jn}+q^{(2)}_{jn}$ which represent the part of p_{jn} and q_{jn} which passes through the collapsing body and $y_{jn}=p^{(1)}_{jn}+q^{(1)}_{jn}$ which represents the part of p_{jn} and q_{jn} which propagates out to $\mathscr{I}^-$ through the quasi-stationary metric of the final black hole. In the initial vacuum state the $\{y_{jn}\}$ modes will not contain any particles but each x_{jn} mode will contain $\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}$ particles. These particles will appear to leave the collapsing body just outside the event horizon and will propagate radially outwards. A fraction Γ_{jn} will penetrate through the potential barrier peaked at $r=3M$ and will escape to $\mathscr{I}^+$ where they will constitute the thermal emission of the black hole. The remaining fraction $1-\Gamma_{jn}$ will be reflected back by the potential barrier and will cross the event horizon. Thus the net particle flux across a surface of constant r just outside the horizon will be Γ_{jn} directed outwards.

I shall now show that using the normal ordered energy momentum operator, the average energy flux across a surface of constant r between retarded times u_1 and u_2

$$(u_2-u_1)^{-1}\int_{u_1}^{u_2}\langle 0_-|T_{ab}|0_-\rangle K^a d\Sigma^b \tag{4.5}$$

is directed outwards and is equal to the energy flux for the thermal emission from a hot body. Because the $\{y_{jn}\}$ contain no negative frequencies on $\mathscr{I}^-$, they will not make any contribution to the expectation value (4.5) of the normal ordered energy-momentum operator. Let

$$x_{jn}=\int_0^\infty(\zeta_{jn\omega'}f_{\omega'}+\xi_{jn\omega'}\bar{f}_{\omega'})d\omega'\,. \tag{4.6}$$

Near $\mathscr{I}^+$

$$x_{jn}=(\Gamma_{jn})^{\frac{1}{2}}p_{jn}\,. \tag{4.7}$$

Thus

$$\begin{aligned}(4.5)=(u_2-u_1)^{-1}\,\mathrm{Re}\{&\textstyle\sum_{j,n}\sum_{j'',n''}\int_0^\infty\int_{u_1}^{u_2}\omega\omega''\Gamma^{\frac{1}{2}}_{jn}p_{jn}\bar{\xi}_{jn\omega'}\\&\cdot(\bar{\Gamma}^{\frac{1}{2}}_{j''n''}\,\bar{p}_{j''n''}\xi_{j''n''\omega'}-\Gamma^{\frac{1}{2}}_{j''n''}p_{j''n''}\zeta_{j''n''\omega'})d\omega'\,du\}\end{aligned} \tag{4.8}$$

where ω and ω'' are the frequencies of the wave-packets p_{jn} and $p_{j''n''}$ respectively. In the limit u_2-u_1 tends to infinity, the second term in the integrand in (4.8) will integrate out and the first term will contribute only for $(j'',n'')=(j,n)$. By arguments similar to those used in Section 2,

$$\int_0^\infty|\xi_{jn\omega'}|^2d\omega'=\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}\,. \tag{4.9}$$

Therefore

$$(4.5)=\int_0^\infty\Gamma_\omega\omega\{\exp(2\pi\omega\kappa^{-1})-1\}^{-1}d\omega \tag{4.10}$$

where $\Gamma_\omega=\lim_{n\to\infty}\Gamma_{jn}$ is the fraction of wave-packet of frequency that would be absorbed by the black hole. The energy flux (4.10) corresponds exactly to the rate of thermal emission calculated in Section 2. Any renormalised energy momentum

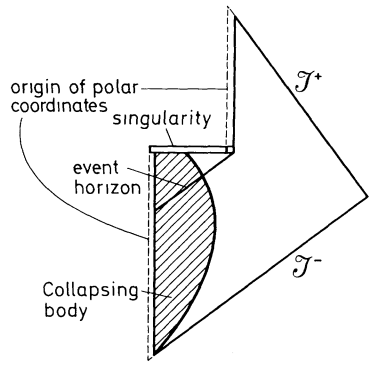


Fig. 5. The Penrose diagram for a gravitational collapse followed by the slow evaporation and eventual disappearance of the black hole, leaving empty space with no singularity at the origin

operator which agrees with the normal ordered operator near $\mathscr{I}^+$, which obeys the conservation equations, and which is stationary in the final quasi-stationary region will give the same energy flux over any surface of constant r. Thus it will give positive energy flux out across the event horizon or, equivalently, a negative energy flux in across the event horizon.

This negative energy flux will cause the area of the event horizon to decrease and so the black hole will not, in fact, be in a stationary state. However, as long as the mass of the black hole is large compared to the Planck mass 10^{-5} g, the rate of evolution of the black hole will be very slow compared to the characteristic time for light to cross the Schwarzchild radius. Thus it is a reasonable approximation to describe the black hole by a sequence of stationary solutions and to calculate the rate of particle emission in each solution. Eventually, when the mass of the black hole is reduced to 10^{-5} g, the quasi-stationary approximation will break down. At this point, one cannot continue to use the concept of a classical metric. However, the total mass or energy remaining in the system is very small. Thus, provided the black hole does not evolve into a negative mass naked singularity there is not much it can do except disappear altogether. The baryons or leptons that formed the original collapsing body cannot reappear because all their rest mass energy has been carried away by the thermal radiation. It is tempting to speculate that this might be the reason why the universe now contains so few baryons compared to photons: the universe might have started out with baryons only, and no radiation. Most of the baryons might have fallen into small black holes which then evaporated giving back the rest mass energy of baryons in the form of radiation, but not the baryons themselves.

The Penrose diagram of a black hole which evaporates and leaves only empty space is shown in Fig. 5. The horizontal line marked "singularity" is really a region where the radius of curvature is of the order the Planck length. The matter that runs into this region might reemerge in another universe or it might even reemerge in our universe through the upper vertical line thus creating a naked singularity of negative mass.

References

1. Isham, C. J.: Preprint (1973)
2. Ashtekar, A., Geroch, R. P.: Quantum theory of gravity (preprint 1973)
3. Penrose, R.: Phys. Rev. Lett. **14**, 57—59 (1965)
4. Hawking, S. W.: Proc. Roy. Soc. Lond. **A 300**, 187—20 (1967)
5. Hawking, S. W., Penrose, R.: Proc. Roy. Soc. Lond. **A 314**, 529—548 (1970)
6. Hawking, S. W., Ellis, G. F. R.: The large scale structure of space-time. London: Cambridge University Press 1973
7. Hawking, S. W.: The event horizon. In: Black holes. Ed. C. M. DeWitt, B. S. DeWitt. New York: Gordon and Breach 1973
8. Bardeen, J. M., Carter, B., Hawking, S. W.: Commun. math. Phys. **31**, 161—170 (1973)
9. Hawking, S. W.: Mon, Not. Roy. astr. Soc. **152**, 75—78 (1971)
10. Carr, B. J., Hawking, S. W.: Monthly Notices Roy. Astron. Soc. **168**, 399—415 (1974)
11. Hagedorn, R.: Astron. Astrophys. **5**, 184 (1970)
12. Hawking, S. W.: Commun. math. Phys. **25**, 152—166 (1972)
13. Carter, B.: Black hole equilibrium states. In: Black holes. Ed. C. M. DeWitt, B. S. DeWitt. New York: Gordon and Breach 1973
14. Misner, C. W.: Bull. Amer. Phys. Soc. **17**, 472 (1972)
15. Press, W. M., Teukolsky, S. A.: Nature **238**, 211 (1972)
16. Starobinsky, A. A.: Zh. E.T.F. **64**, 48 (1973)
17. Starobinsky, A. A., Churilov, S. M.: Zh. E.T.F. **65**, 3 (1973)
18. Bjorken, T. D., Drell, S. D.: Relativistic quantum mechanics. New York: McGraw Hill 1965
19. Beckenstein, J. D.: Phys. Rev. D. **7**, 2333—2346 (1973)
20. Beckenstein, J. D.: Phys. Rev. D. **9**,
21. Penrose, R.: Phys. Rev. Lett. **10**, 66—68 (1963)
22. Sachs, R. K.: Proc. Roy. Soc. Lond. **A 270**, 103 (1962)
23. Eardley, D., Sachs, R. K.: J. Math. Phys. **14** (1973)
24. Schmidt, B. G.: Commun. Math. Phys. **36**, 73—90 (1974)
25. Bondi, H., van der Burg, M. G. J., Metzner, A. W. K.: Proc. Roy. Soc. Lond. **A 269**, 21 (1962)
26. Carter, B.: Commun. math. Phys. **10**, 280—310 (1968)
27. Teukolsky, S. A.: Ap. J. **185**, 635—647 (1973)
28. Unruh, W.: Phys. Rev. Lett. **31**, 1265 (1973)
29. Unruh, W.: Phys. Rev. D. **10**, 3194—3205 (1974)
30. Zeldovich, Ya. B., Starobinsky, A. A.: Zh. E.T.F. **61**, 2161 (1971), JETP **34**, 1159 (1972)

Communicated by J. Ehlers

S. W. Hawking
California Institute of Technology
W. K. Kellogg Radiation Lab. 106-38
Pasadena, California 91125, USA