BLACK HOLES

by

S.W. Hawking

Institute of Theoretical Astronomy
University of Cambridge
Cambridge, England

An essay written for the
Gravity Research Foundation awards 1971

Summary

It is assumed that the singularities which occur in gravitational collapse are not visible from outside but are hidden behind an event horizon. A black hole on a spacelike surface is defined to be a connected region on the surface bounded by the event horizon. As time increases, black holes may merge together but may never bifurcate. The areas of the boundaries of the black holes can never decrease. These areas are related by the Carter-Israel conjecture to the masses and angular momento of the black holes. Together with the previous results this gives upper bounds on the amount of energy that can be extracted from the black holes. In particular, if the bursts of gravitational radiation that Weber reports are produced by collisions of black holes, then the black holes must have masses of at least a hundred times that of the sun.

It has been known for some time that a nonrotating star of more than about two solar masses has no low temperature equilibrium configuration. This means that such a star must undergo catastrophic collapse when it has exhausted its nuclear fuel unless it has managed to eject sufficient matter to reduce its mass to less than twice that of the sun. If the collapse is exactly spherically symmetric, the metric is that of the Schwarzschild solution outside the star and has the following properties:

1) The surface of the star will pass inside the Schwarzschild radius $r = 2Gc^{-2}M$. After this has happened there will be closed trapped surfaces[1,2] around the star. A closed trapped surface is a spacelike 2-surface such that both the future directed families of null geodesics orthogonal to it are converging. In other words, it is in such a strong gravitational field that even the outgoing light from it is dragged inwards.

2) There is a space-time singularity.

3) The singularity is not visible to observers who remain outside the Schwarzschild radius. This means that the breakdown of our present physical theory which one expects to occur at a singularity cannot affect what happens outside the Schwarzschild radius and one can still predict the future in the exterior region from Cauchy data on a spacelike surface.

One can ask whether these three properties of spherical collapse are stable, i.e. whether they would still hold if the initial data for the collapse were perturbed slightly. This is vital because no real collapse situation will ever be exactly spherical. From the stability of the Cauchy problem in general relativity[3] one can show that a sufficiently small perturbation of the initial data on a spacelike surface will produce a perturbation of the solution which will remain small on a compact region in the Cauchy development of the surface. This shows that property (1) is stable, since there is a compact region in the Cauchy development of the initial surface which contains closed trapped surfaces. It then follows that property (2) is stable provided one makes certain reasonable assumptions such as that the energy density of matter is always positive. This is because the existence of a closed trapped surface implies the occurrence of a singularity under these conditions[4]. There remains the problem of the stability of property (3). Since the question of whether singularities are visible from outside depends on the solution at arbitrarily large times, one cannot appeal to the result on the stability of the Cauchy problem referred to above. Nevertheless it seems a reasonable conjecture that property (3) is indeed stable. If this is the case, we can still predict what happens outside collapsed objects, and we need not worry that something unexpected might occur every time a star in the galaxy collapsed. My belief in this conjecture is strengthened by the fact that Penrose[5] and I have tried and failed to obtain a contradiction to it, which would show that naked singularities must occur. I shall therefore assume that property (3) is stable. In Section 2a precise definition of a black hole is given in terms of an event horizon and it is shown that the area of a 2-dimensional section of this horizon cannot decrease with time. This area is related by the Carter-Israel conjecture in Section 3 to the mass and angular momentum of the black hole. This result can be used to place an upper bound on the amount of gravitational radiation that can be emitted when two black holes collide and coalesce. This may be of importance in connection with Weber's observations of short bursts of gravitational radiation.

2. The Event Horizon

In order to discuss the region outside a collapsed object one needs a precise notion of infinity in an asymptotically flat space-time. This is provided by Penrose's concept of a weakly asymptotically simple space[2]; the spacetime manifold M of such a space can be imbedded in a larger manifold $\tilde{M}$ on which there is a Lorentz metric $\tilde{g}_{ab}$ which is conformal to the space-time metric g_{ab} i.e. $\tilde{g}_{ab} = \Omega^2 g_{ab}$ where Ω is a smooth function which is zero and has non-fanishing gradient on the boundary of M in $\tilde{M}$. This boundary consists of two null hypersurfaces $\mathcal{J}^+$ and $\mathcal{J}^-$ which each have topology $S^2 \times R^1$ and which represent future and past null infinity respectively. One can then interpret property (3) as saying that it should be possible to predict events near $\mathcal{J}^+$. I shall therefore say that a weakly asymptotically simple space is (future) asymptotically predictable if there is a partial Cauchy surface S such that $\mathcal{J}^+$ lies in the closure in $\tilde{M}$ of $D^+(S)$ the future Cauchy development of S. (A partial Cauchy surface is a spacelike surface without edge which does not intersect any nonspacelike curve more than once. $D^+(S)$ is the set of all points p such that every past directed non-spacelike curve from p intersects S if extended far enough.)

Roughly speaking one would expect a space to be asymptotically predictable if there are no singularities in $J^+(S)$, the future of S, which are naked, i.e. which lie in $J^-(\mathcal{J}^+)$, the past of future null infinity. One can make this more precise. Consider an asymptotically predictable space in which there are no singularities to the past of S (Fig. 1). Suppose there is a closed trapped surface T in $D^+(S)$. Then there will be a singularity to the future of T , i.e. there will be a nonspacelike geodesic in $J^+(T)$ which is future incomplete. Can this geodesic be seen from $\mathscr{I}^+$? The answer is no. For suppose T intersected $J^-(\mathscr{I}^+)$. Then there would be a point $p \in \mathscr{I}^+$ in $J^+(T)$. The past directed null geodesic generator of $\mathscr{I}^+$ through p would eventually leave $J^+(T)$ and so would contain a point q of the boundary $\dot{J}^+(T)$. Now the boundary of the future of any closed set W is generated by null geodesic segments which either have no past end-points or have past end-points on W[2,4]. Since the generator λ of $\dot{J}^+(T)$ through q would enter $D^+(S)$ it would have to have an end point on T since otherwise it would intersect S and pass into the past of S which would be impossible, as T is to the future of S . The generator λ would intersect T orthogonally. However, as T is a closed trapped surface, the null geodesics orthogonal to T are converging. Together with the weak energy condition: $T_{ab}K^aK^b \geq 0$ for any timelike vector K^a , this implies that there will be a point conjugate to T within a finite affine length on any null geodesic orthogonal to T[4]. Points on such a geodesic beyond the conjugate point will lie in the interior of $J^+(T)$ and not on its boundary[2,4]. However the generator λ of $\dot{J}^+(T)$ would have infinite affine length from T to $\mathscr{I}^+$ since $\mathscr{I}^+$ is at infinity. This establishes a contradiction which shows that T does not intersect $J^-(\mathscr{I}^+)$. Thus the future incomplete geodesic in $J^+(T)$ is not visible from $\mathscr{I}^+$.

Since $J^-(\mathscr{I}^+)$ does not contain T , its boundary $\dot{J}^-(\mathscr{I}^+)$ must be non-empty. This is the <u>event horizon</u> for $\mathscr{I}^+$ and is the boundary of the region from which particles or photons can escape to infinity. It is generated by null geodesic segments which have no future end-points. The convergence ρ

of these generators cannot be positive. For suppose it were positive on some open set U of $\dot{J}^-(\mathscr{I}^+)$. Let F be a spacelike 2-surface in U. Then the outgoing null geodesics orthogonal to F would be converging. One could deform a small part of F so that it intersected $J^-(\mathscr{I}^+)$ but so that the outgoing null geodesics orthogonal to F were still converging. This again would lead to a contradiction since the null geodesics orthogonal to F could not remain in $\dot{J}^+(F)$ all the way out to $\mathscr{I}^+$.

If there were a point on the event horizon which was not in $D^+(S)$, the future Cauchy development of S, a small perturbation could result in there being points near $\mathscr{I}^+$ which were not in $D^+(S)$. Since I am assuming that asymptotic predictability is stable, I shall slightly extend the definition to exclude this kind of situation. In an asymptotically predictable space $J^+(S) \cap J^-(\mathscr{I}^+)$ is in $D^+(S)$. I shall say that such a space is <u>strongly</u> asymptotically predictable if in addition $J^+(S) \cap \dot{J}(\mathscr{I}^+)$ is in $D^+(S)$.

In such a space one can construct a family $S(t)$ $(t > 0)$ of partial Cauchy surfaces in $D^+(S)$ such that

(a) For $t_2 > t_1$, $S(t_2) \subset J^+(S(t_1))$.

(b) Each $S(t)$ intersects $\mathscr{I}^+$ in a 2-sphere $A(t)$.

(c) For each $t > 0$, $S(t) \cup [\mathscr{I}^+ \cap J^-(A(t))]$ is a Cauchy surface for $D^+(S)$.

The construction is as follows. Choose a suitable family $A(t)$ of 2-spheres on $\mathscr{I}^+$. Put a volume measure on M so that the total volume of M in this measure is finite[6]. Define the functions $f(p)$ and $h(p,t)$ $p \in D^+(S)$ as the volumes of $J^+(p) \cap D^+(S)$ and $[J^-(p) - J^-(A(t))] \cap D^+(S)$ respectively. They will be continuous in p and t. The surface $S(t)$ is then defined to be the set of points p such that $h(p,t) = tf(p)$.

For sufficiently large t , the surfaces $S(t)$ will intersect the event horizon and so $B(t)$ defined as $S(t) - \dot{J}^-(\mathscr{I}^+)$ will be nonempty. I shall define a <u>black hole</u> on the surface $S(t)$ to be a connected component of $B(t)$. In other words, it is a region of $S(t)$ from which there is no escape to $\mathscr{I}^+$. As time increases, black holes may merge together and new black holes may be created by further bodies collapsing but a black hole can never bifurcate. For suppose the black hole $B_1(t_1)$ on the surface $S(t_1)$ divided into two black holes $B_2(t_1)$ and $B_3(t_2)$ by a later surface $S(t_2)$. Then $B_2(t_2)$ and $B_3(t_2)$ would each have to contain points of $J^+(B_1(t_1))$. However every nonspacelike curve which intersected $B_1(t_1)$ would also intersect $S(t_2)$. Therefore $J^+(B_1(t_1)) \cap S(t_2)$ would be connected and would be contained in $B_2(t_2) \cup B_3(t_2)$.

Since the generators of $\dot{J}^-(\mathscr{I}^+)$ have no future end points and have convergence $\rho \leq 0$, the area of $\partial B_1(t)$ cannot decrease with t where $\partial B_1(t)$ is the boundary in $S(t)$ of a black hole $B_1(t)$. If two black holes $B_1(t_1)$ and $B_2(t_1)$ on the surface $S(t_1)$ merge to form a single black hole $B_3(t_2)$ on a later surface $S(t_2)$, then the area of $\partial B_3(t_1)$ must be at least one sum of the areas of $\partial B_1(t_1)$ and $\partial B_2(t_1)$. In fact it must be strictly greater than this sum because $\partial B_3(t_2)$ contains two disjoint closed sets which correspond to the generators of $\dot{J}^-(\mathscr{I}^+)$ which intersect $\partial B_1(t_1)$ and $\partial B_2(t_1)$. Since $\partial B_3(t_2)$ is connected, it must also contain an open set of points which correspond to generators which have past end points between t_2 and t_1 . These results will be used in the next section to place certain limits on the possible behaviour of black holes.

3. The Carter-Israel Conjecture

In a collapse that was strongly asymptotically predictable one would expect the solution outside the event horizon eventually to approach a stationary state. This has lead to a study of strongly asymptotically predictable spaces which are exactly stationary. Israel[7] has shown that the Schwarzschild solution is the only empty static such solution in which the surfaces of constant potential are topologically $S^2 \times R^1$. Carter[8] has shown that the empty stationary axisymmetric such solutions form two parameter families. The two parameters represent the mass m and angular momentum per unit mass $\frac{ac}{G}$ as measured on $\mathcal{J}^+$. One such family is known: namely, the Kerr solutions for $m \geq 0$ and $a \leq m$. It seems unlikely that there are any others. The Carter-Israel conjecture is therefore that the solution outside a black hole $B(t)$ will settle down to the Kerr solution with the same mass and angular momentum as that measured on the 2-surface $A(t)$ on $\mathcal{J}^+$. If this is the case the area of the 2-surface $\partial B(t)$ will approach the area of a two-section of the event horizon in the Kerr solution. This area is

$$8\pi G^2 c^{-4} m(m + (m^2 - a^2)^{\frac{1}{2}}) \tag{1}$$

Consider a situation in which a black hole $B(t_1)$ has settled down by a surface $S(t_1)$ to a Kerr solution with parameters m_1 and a_1. Suppose the black hole now interacts with various particles or fields and then settles down by a surface $S(t_2)$ to a Kerr solution with parameters m_2 and a_2. The area of $\partial B(t_2)$ must be greater than the area of $\partial B(t_1)$. Therefore

$$m_1(m_1 + (m_1^2 - a_1^2)^{\frac{1}{2}}) < m_2(m_2 + (m_2^2 - a_2^2)^{\frac{1}{2}}) \quad (2)$$

Using an idea of Penrose[9], Christodolou[10] has shown that one can get arbitrarily near the limit set by this inequality. In particular, if a_2 is less than a_1 then m_2 can be less than m_1. This means that by reducing the angular momentum of a black hole one can extract a certain amount of energy from it. One can regard this energy as the rotational energy of the black hole.

Consider now a situation in which two stars a long way apart collapse to form black holes $B_1(t_1)$ and $B_2(t_1)$ on a surface $S(t_1)$. One can neglect the interaction between them and take the areas of $\partial B_1(t_1)$ and $\partial B_2(t_1)$ to be given by formula (1) with the values of the parameters m_1, a_1 and m_2, a_2 respectively. Suppose the two black holes now collide and merge to form a single black hole $B_3(t_2)$ on a later surface $S(t_2)$. In this process a certain amount of gravitational radiation will be emitted. By the conservation law for weakly asymptotically simple space-times, the energy of this radiation will be $(m_1 + m_2 - m_3)c^2$. This is limited by the requirement that the area of $\partial B_3(t_2)$ must be greater than the sum of the areas of $\partial B_1(t_1)$ and $\partial B_2(t_1)$. This gives the inequality

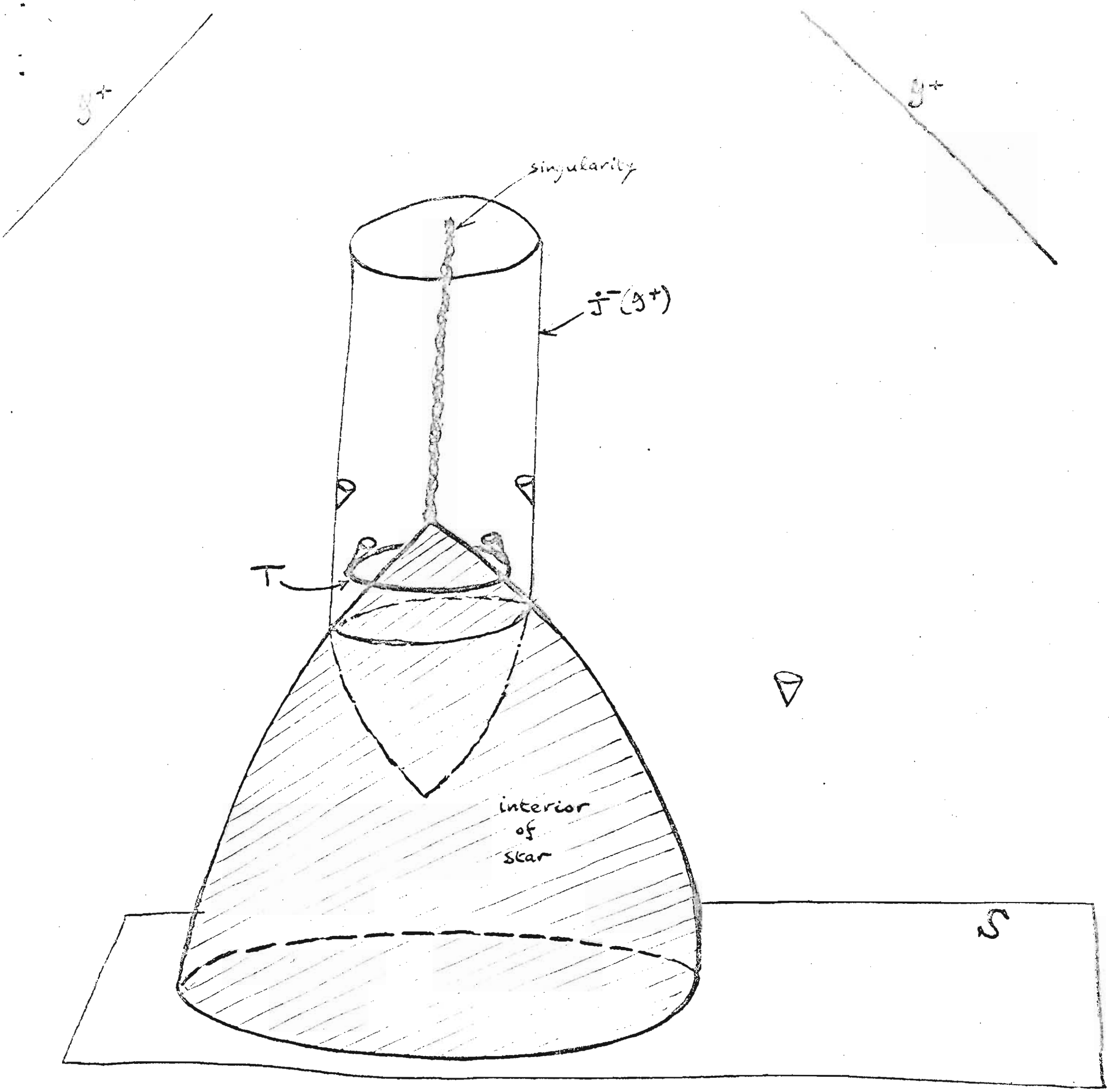
$$m_3(m_3 + (m_3^2 - a_3^2)^{\frac{1}{2}}) > m_1(m_1 + (m_1^2 - a_1^2)^{\frac{1}{2}}) + m_2(m_2 + (m_2^2 - a_2^2)^{\frac{1}{2}})$$

The efficiency $\epsilon = (m_1 + m_2)^{-1}(m_1 + m_2 - m_3)$ of conversion of rest mass energy into radiation is always less than $\frac{1}{2}$. If $a_1 = a_2 = 0$, then $\epsilon < 1 - 2^{-\frac{1}{2}}$. An interesting special case occurs when two black holes with angular momentum have their rotation axes aligned along the direction of their approach to each other. As this situation is axisymmetric, no angular momentum can be carried away by the gravitational radiation. If the angular momenta are in the same sense they will add up and the amount of radiation that can be emitted will be less than if they had opposite senses. This suggests that the attractive force between two black holes may depend on the relative orientation of the rotation axes.

These limits may be important in connection with Weber's recent reports of short bursts of gravitational radiation[11,12,13]. The energies in these bursts seem so large that one cannot account for them by nuclear reactions since these only release about one per cent of the rest mass energy. Two possibly more efficient processes are gravitational collapse or the capture of one black hole by another. In the former case, one would expect at least a small fraction of the energy to be emitted as electro-magnetic radiation or neutrinos. The fact that these have not been observed[14,15] favours the second process. In this case the limits above imply that the black holes must have masses of about one hundred times that of the sun. Such masses would produce bursts of gravitational radiation whose energies were mostly at about the frequency at which Weber observes

Acknowledgement

I am grateful to George Ellis and Roger Penrose for some useful discussions.



singularity
$\dot{J}^-(\mathcal{J}^+)$
$\mathcal{J}^+$
$\mathcal{J}^+$
T
interior of star
S

Figure 1.

References

1. Penrose, R. Phys. Rev. Lett. 14, 57 (1965).
2. Penrose, R. in Battelle Rencontres (1967) Ed. C.M. de Witt and J.A. Wheeler, published by Ben amin, New York 1968.
3. Hawking, S.W. and Ellis, G.F.R. Singularities, Causality, Cosmology, Cambridge University Press, to be published.
4. Penrose, R. and Hawking, S.W. Proc. Roy. Soc. A 314, 529 (1970).
5. Penrose, R. Seminar at Cambridge University, January 1971 (unpublished).
6. Geroch, R.P. Domain of Dependence. In J. Math. Phys. 11, 437, 1970.
7. Israel, W. Phys. Rev. 164, 1776 (1967).
8. Carter, B. Phys. Rev. Lett. (1971).
9. Penrose, R. Nuova Cimento Serie 1, 1, Numero speciale 252 (1969).
10. Christodolou, D. Phys. Rev. Lett. 25, 1596 (1970).
11. Weber, J. Phys. Rev. Lett. 22, 1320 (1969).
12. Weber, J. Phys. Rev. Lett. 24, 276 (1970).
13. Weber, J. Phys. Rev. Lett. 25, 180 (1970).
14. Jelley, J.V. Nature (1970).
15. Bahcall, J. Report at the 5th Texas Symposium on Relativistic Astrophysics, December 1970.

BIOGRAPHICAL SKETCH

I was born in 1942. I obtained a B.A. in Physics at Oxford in 1962. Since then I have been at Cambridge. My research is in the fields of general relativity and cosmology.

I am married and have two young children.

S.W. Hawking