9

Gravitational collapse and black holes

In this chapter, we shall show that stars of more than about $1\frac{1}{2}$ times the solar mass should collapse when they have exhausted their nuclear fuel. If the initial conditions are not too asymmetric, the conditions of theorem 2 should be satisfied and so there should be a singularity. This singularity is however probably hidden from the view of an external observer who sees only a 'black hole' where the star once was. We derive a number of properties of such black holes, and show that they probably settle down finally to a Kerr solution.

In § 9.1 we discuss stellar collapse, showing how one would expect a closed trapped surface to form around any sufficiently large spherical star at a late stage in its evolution. In § 9.2 we discuss the event horizon which seems likely to form around such a collapsing body. In § 9.3 we consider the final stationary state to which the solution outside the horizon settles down. This seems to be likely to be one of the Kerr family of solutions. Assuming that this is the case, one can place certain limits on the amount of energy which can be extracted from such solutions.

For further reading on black holes, see the 1972 Les Houches summer school proceedings, edited by B. S. de Witt, to be published by Gordon and Breach.

9.1 Stellar collapse

Outside a static spherically symmetric body such as a star, the solution of Einstein's equations is necessarily that part of one of the asymptotically flat regions of the Schwarzschild solution for which r is greater than some value r_0 corresponding to the surface of the star. This will be joined, for $r < r_0$, onto a solution which depends in detail on the radial distribution of density and pressure in the star. In fact even if the star is not static, providing it remains spherically symmetric the solution outside will still be part of the Schwarzschild solution cut off by the surface of the star. (This is Birkhoff's theorem, proof of which is given in appendix B.) If the star is static then r_0 must be

https://doi.org/10.1017/CBO9780511524646.010 Published online by Cambridge University Press

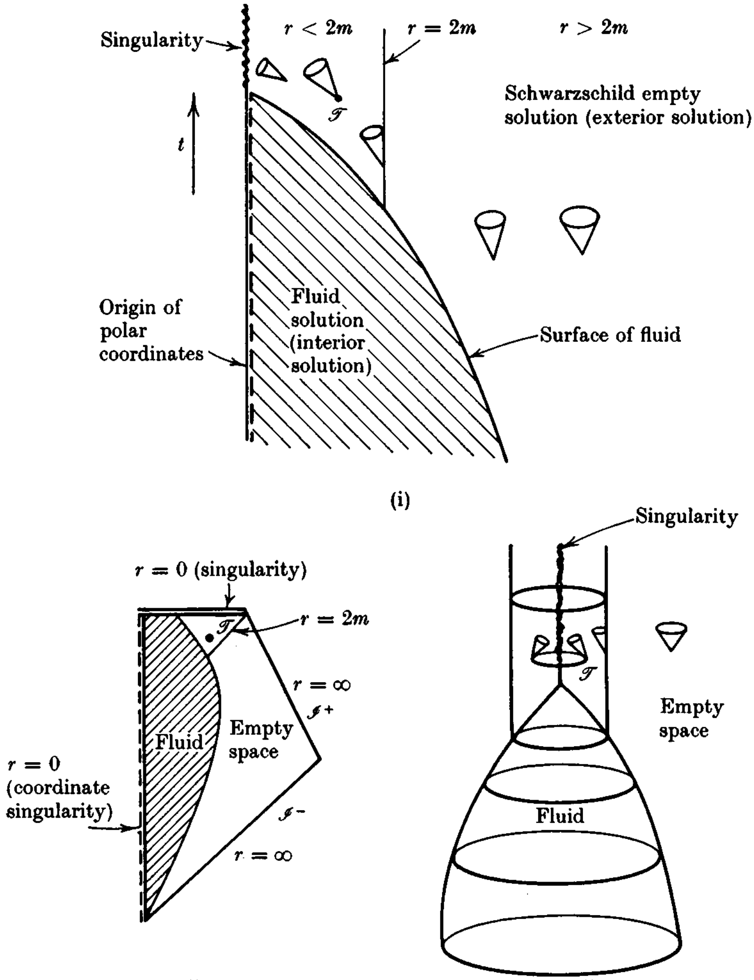

FIGURE 54. Collapse of a spherical star.

(i) Finkelstein diagram ((r, t) plane) of a collapsing spherically symmetric fluid ball. Each point represents a two-sphere.

(ii) Penrose diagram of the collapsing fluid ball.

(iii) Diagram of the collapse with only one spatial dimension suppressed.

greater than $2m$ (the 'Schwarzschild radius'). This follows because the surface of a static star must correspond to the orbit of a timelike Killing vector, and in the Schwarzschild solution there is a timelike Killing vector only where $r > 2m$. If r_0 were less than $2m$, the surface of the star would be expanding or contracting. To get an idea of the magnitude of the Schwarzschild radius, we note that the Schwarzschild radius of the earth is 1.0 cm and that of the sun is 3.0 Km;

the ratios of the Schwarzschild radius to the radius of the earth and the sun are 7×10^{-10} and 2×10^{-6} respectively. Thus normal stars are a long way from their Schwarzschild radii.

The life of a typical star will consist of a long ($\sim 10^9$ years) quasi-static phase in which it is burning nuclear fuel and supporting itself against gravity by thermal and radiation pressure. However when the nuclear fuel is exhausted, the star will cool, the pressure will be reduced, and so it will contract. Now suppose that this contraction cannot be halted by the pressure before the radius becomes less than the Schwarzschild radius (we shall see below that this seems likely for stars of greater than a certain mass). Then since the solution outside the star is the Schwarzschild solution, there will be a closed trapped surface $\mathscr{T}$ around the star (see figure 54), and so, by theorem 2, a singularity will occur provided that causality is not violated and the appropriate energy condition holds. Of course in this case, because the exterior solution is the Schwarzschild solution, it is obvious (see figure 54) that there must be a singularity. However the point is that even if the star is not exactly spherically symmetric, a closed trapped surface will still occur providing the departures from spherical symmetry are not too great. This follows from the stability of the Cauchy development proved in §7.5; for one can regard the solution as developing from a partial Cauchy surface $\mathscr{H}$ (figure 55). Now if one changes the initial data by a sufficiently small amount on the compact region $J^-(\mathscr{T}) \cap \mathscr{H}$, the new development of $\mathscr{H}$ will still be sufficiently near the old in the compact region $J^+(\mathscr{H}) \cap J^-(\mathscr{T})$ that there will still be a closed trapped surface around the star in the perturbed solution. Thus we have shown that there is a non-zero measure set of initial conditions which lead to a closed trapped surface and hence to a singularity by theorem 2.

The two principal reasons why a star may depart from spherical symmetry are that it may be rotating or may have a magnetic field. One may get some idea of how large the rotation may be without preventing the occurrence of a trapped surface by considering the Kerr solution. This solution can be thought of as representing the exterior solution for a body with mass m and angular momentum $L = am$. If a is less than m there are closed trapped surfaces, but if a is greater than m they do not occur. Thus one might expect that if the angular momentum of the star were greater than the square of its mass, it would be able to halt the contraction of the star before a closed trapped surface developed. Another way of seeing this is that if $L = m^2$ and angular momentum is conserved during the collapse, then the velocity

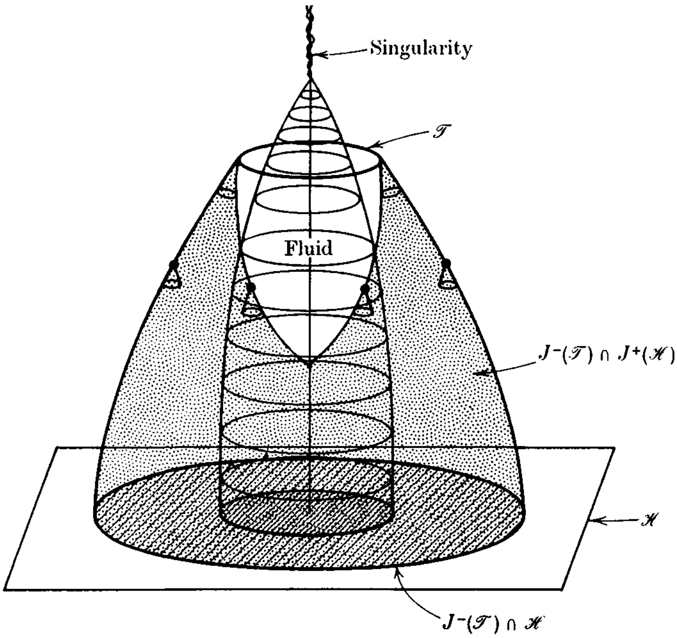

FIGURE 55. Collapse of a spherical star as in figure 54 (iii) showing a partial Cauchy surface $\mathscr{H}$. It is the initial data on the compact region $J^-(\mathscr{T}) \cap \mathscr{H}$ of $\mathscr{H}$ which leads to the occurrence of the closed trapped surface $\mathscr{T}$ in the compact region $J^-(\mathscr{T}) \cap J^+(\mathscr{H})$.

of the surface of the star would be about the velocity of light when the star was at its Schwarzschild radius. Now many stars have an angular momentum greater than the square of their mass (for the sun, $L \sim m^2$). However it seems reasonable to expect some loss of angular momentum during the collapse because of braking by magnetic fields and because of gravitational radiation. The situation is therefore that in some stars, and probably most, angular momentum would not prevent occurrence of closed trapped surfaces, and hence a singularity.

In a nearly spherical collapse a magnetic field **B** which is frozen into a star will increase as the matter density ρ to the $\frac{2}{3}$ power. Thus the magnetic pressure is proportional to $\rho^{\frac{4}{3}}$. This rate of increase is so slow that if the magnetic pressure is not important initially in supporting the star, then it will never be strong enough to have a significant effect on the collapse.

To see why a burnt-out star of more than a certain mass cannot support itself against gravity, we shall give a qualitative discussion (based on unpublished work by Carter) of the zero temperature equation of state for matter.

In hot matter there is pressure produced by the thermal motions of the atoms and by the radiation present. However in cold matter at densities lower than that of nuclear matter ($\sim 10^{14}\,\text{gm cm}^{-3}$), the only significant pressure will arise from the quantum mechanical exclusion principle. To estimate this, consider a number density n of fermions of mass m. By the exclusion principle, each fermion will effectively occupy a volume of n^{-1}. Thus by the uncertainty principle, it will have a spatial component of momentum of order $\hbar n^{\frac{1}{3}}$. If the fermions are non-relativistic, i.e. if $\hbar n^{\frac{1}{3}}$ is less than m, the velocity of the fermions will be of order $\hbar n^{\frac{1}{3}}/m$, while if the fermions are relativistic (i.e. $\hbar n^{\frac{1}{3}}$ is greater than m) then the velocity will be practically one (the speed of light). The pressure will be of order (momentum) $\times$ (velocity) $\times$ (number density), and so will be $\sim \hbar^2 n^{\frac{5}{3}} m^{-1}$ if $\hbar n^{\frac{1}{3}} < m$, and will be $\sim \hbar n^{\frac{4}{3}}$ if $\hbar n^{\frac{1}{3}} > m$. When the matter is non-relativistic, the principal contribution to the degeneracy pressure comes from the electrons, since m^{-1} for them is bigger than it is for baryons. However at high densities, when the particles become relativistic, the pressure is independent of the mass of the particles producing it and depends simply on their number density.

For small cold bodies, self-gravity can be neglected and the degeneracy pressure will be balanced by attractive electrostatic forces between nearest neighbour particles arranged in some sort of lattice. (We assume that there are equal numbers of positive and negative charges and approximately equal numbers of electrons and baryons.) These forces will produce a negative pressure of order $e^2 n^{\frac{4}{3}}$. Thus the mass density of a small cold body will be of order

$$e^6 m_e{}^3 m_n \hbar^{-6} \quad (\sim 1\,\text{gm cm}^{-3}), \tag{9.1}$$

where m_e is the electron rest-mass and m_n is the nucleon rest-mass.

For larger bodies self-gravity will be important, and will compress the matter against the degeneracy pressure. To obtain an exact solution would involve a detailed integration of Einstein's equations. However the important qualitative features can be seen more easily from a simple Newtonian order of magnitude argument. In a star of mass M and radius r_0, the gravitational force on a typical unit volume is of the order $(M/r_0{}^2)\, n m_n$, where $n m_n \simeq M/r_0{}^3$ is the mass density. The gravitational force will be balanced by a pressure gradient of order P/r_0, where P is the average pressure in the star. Thus

$$P = M^2/r_0{}^4 \simeq M^{\frac{2}{3}} n^{\frac{4}{3}} m_n{}^{\frac{4}{3}}.$$

If the density is sufficiently low that the main contribution to the pressure is from the degeneracy of non-relativistic electrons,

$$P = \hbar^2 n^{\frac{5}{3}} m_e^{-1} = M^{\frac{2}{3}} n^{\frac{4}{3}} m_n^{\frac{4}{3}},$$

so

$$n = M^2 m_n^{4} m_e^{3} \hbar^{-6}.$$

This will be the correct formula for bodies for which it yields a value of n greater than (9.1) and less than $m_e^{3}\hbar^{-3}$, i.e. for $e^3 m_n^{-2} < M < \hbar^{\frac{3}{2}} m_n^{-2}$. Such stars are known as white dwarfs.

If the density is so high that the electrons are relativistic, i.e. $n > m_e^{3}\hbar^{-3}$, then the pressure will be given by the relativistic formula; so $P = \hbar n^{\frac{4}{3}} = M^{\frac{2}{3}} n^{\frac{4}{3}} m_n^{\frac{4}{3}}$. Now n cancels out of this equation. Thus apparently one obtains a star of mass

$$M_L = \hbar^{\frac{3}{2}} m_n^{-2} \simeq 1.5\, M_\odot,$$

which can have any density greater than $m_e^{3} m_n \hbar^{-3}$, i.e. any radius less than $\hbar^{\frac{3}{2}} m_n^{-1} m_e^{-1}$. Stars of mass greater than M_L simply cannot be supported by the degeneracy pressure of electrons.

In fact, when the electrons become relativistic they tend to induce inverse beta decay with the protons, producing neutrons:

$$e^- + p \rightarrow \nu_e + n.$$

This denudes the electrons and hence reduces their degeneracy pressure, thereby causing the star to contract and making the electrons more relativistic. This is an unstable situation, and the process will continue until nearly all the electrons and protons have been converted into neutrons. At this stage, equilibrium is again possible with the star supported by the degeneracy pressure of the neutrons. Such a body is called a neutron star. If the neutrons are non-relativistic, one finds

$$n = M^2 m_n^{7} \hbar^{-6}.$$

If the neutrons are relativistic, the star must again have a mass M_L and a radius less than or equal to $\hbar^{\frac{3}{2}} m_n^{-2}$. However $M_L/\hbar^{\frac{3}{2}} m_n^{-2} = 1$ and so such a star is near the General Relativity limit $M_L/R \approx 2$.

The conclusion is that a cold star of mass greater than M_L cannot be supported by either electron or neutron degeneracy pressure. To show this rigorously, consider the Newtonian equation of support:

$$\mathrm{d}p/\mathrm{d}r = -\rho M(r)\, r^{-2}, \tag{9.2}$$

where

$$M(r) = 4\pi \int_0^r \rho r^2\, \mathrm{d}r$$

is the mass within radius r. Multiply both sides of (9.2) by r^4 and integrate by parts from 0 to r_0. This gives

$$4\int_0^{r_0} pr^3\,\mathrm{d}r = (M(r_0))^2/8\pi,$$

since $p = 0$ at $r = r_0$. On the other hand,

$$\frac{\mathrm{d}}{\mathrm{d}r}\left(\int_0^r pr'^3\,\mathrm{d}r'\right)^{\frac{3}{4}} = \frac{3}{4}\left(\int_0^r pr'^3\,\mathrm{d}r'\right)^{-\frac{1}{4}} pr^3$$

$$= \frac{3}{4}\left(\tfrac{1}{4}pr^4 - \frac{1}{4}\int_0^r \frac{\mathrm{d}p}{\mathrm{d}r'}r'^4\,\mathrm{d}r'\right)^{-\frac{1}{4}} pr^3 < \frac{3\sqrt{2}}{4}p^{\frac{3}{4}}r^2,$$

since $\mathrm{d}p/\mathrm{d}r$ is never positive. As p is never greater than $\hbar n^{\frac{4}{3}}$, this shows that

$$\int_0^{r_0} pr^3\,\mathrm{d}r < \hbar\left(\int_0^{r_0} nr^2\,\mathrm{d}r\right)^{\frac{4}{3}} = \hbar(M(r_0))^{\frac{4}{3}}\,(4\pi m_\mathrm{n})^{-\frac{4}{3}}.$$

Therefore $M(r_0)$ must be less than $(8\hbar)^{\frac{3}{2}}\,(4\pi)^{-\frac{1}{2}}\,m_\mathrm{n}{}^{-2}$, i.e.

$$M(r_0) < 8\hbar^{\frac{3}{2}}m_\mathrm{n}{}^{-2}.$$

We summarize these results in figure 56. In this diagram we plot the average nucleon density n against the total mass M of the body. The solid line shows the approximate equilibrium configuration of a cold body. In a hot body there will be thermal and radiation pressure in addition to degeneracy pressure and so such bodies may be in equilibrium above the solid line. The heavy dashed line on the right indicates where M/r_0 (which is $M^{\frac{2}{3}}n^{\frac{1}{3}}m_\mathrm{p}{}^{\frac{1}{3}}$) is equal to two. The region to the right of this line contains no equilibrium states, and corresponds to a star being within its Schwarzschild radius. Far away from this line to the left, the difference between Newtonian theory and General Relativity may be neglected. Near this line, one has to take into account General Relativistic effects. For a static spherically symmetric body composed of a perfect fluid, the Einstein field equations can be reduced to (see appendix B)

$$\frac{\mathrm{d}p}{\mathrm{d}r} = -\frac{(\mu+p)\,(\hat{M}(r)+4\pi r^3 p)}{r(r-2\hat{M}(r))}, \tag{9.3}$$

where the radial coordinate is such that the area of the two-surface $\{r = \text{constant},\ t = \text{constant}\}$ is $4\pi r^2$. $\hat{M}(r)$ is now defined as

$$\int_0^r 4\pi r^2 \mu\,\mathrm{d}r,$$

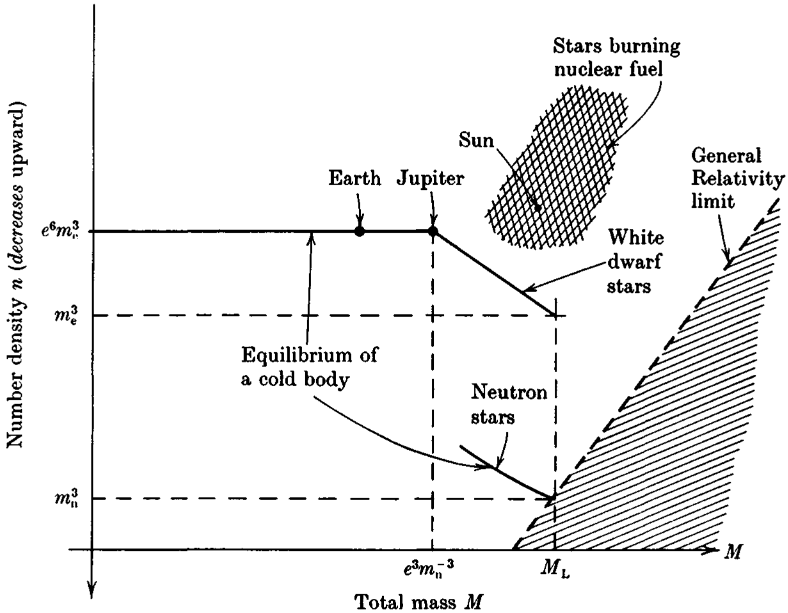

FIGURE 56. Nucleon number density n plotted against total mass of a static body M. The heavy line shows the equilibrium of cold bodies; hot bodies at suitable temperatures can be in equilibrium above this line. General Relativity forbids any bodies in the shaded region from being static.

where $\mu = \rho(1+\epsilon)$ is the total energy density, ρ is nm_n, and ϵ is the relativistic increase of mass associated with the momentum of the fermions. $\hat{M}(r_0)$ is equal to the Schwarzschild mass $\hat{M}$ of the exterior Schwarzschild solution for $r > r_0$. For a bound star this will be less than the conserved mass

$$\tilde{M} = \int_0^{r_0} \frac{4\pi\rho r^2 \,\mathrm{d}r}{(1-2M/r)^{\frac{1}{2}}} = Nm_n,$$

where N is the total number of nucleons in the star, because the difference $(\tilde{M}-\hat{M})$ represents the amount of energy radiated to infinity since the formation of the star from dispersed matter initially at rest. In practice this difference is never more than a few percent and in no case can it exceed $2\hat{M}$, since Bondi (1964) has shown that $(1-2\hat{M}/r)^{\frac{1}{2}}$ cannot be less than $\frac{1}{3}$ provided μ and p are positive and that μ decreases outwards, and cannot be less than $\frac{1}{2}$ if p is less than or equal to μ. Therefore $\hat{M} < \tilde{M} < 3\hat{M}$.

Comparing (9.3) with (9.2), with μ in place of ρ and $\hat{M}$ in place of M, one sees that the extra terms on the right-hand side of (9.3) are all negative provided $\epsilon \geqslant 0$ and $p \geqslant 0$. Thus since in Newtonian theory a cold star of mass $M > M_L$ cannot support itself, neither can a cold star of Schwarzschild mass $\hat{M} > M_L$ in General Relativity. This means that a cold star which contains more than $3M_L/m_n$ nucleons cannot support itself. In practice, the extra terms in (9.3) mean that the limiting nucleon number is less than M_L/m_n.

In our discussion of neutron stars, we ignored the effects of nuclear forces. These will somewhat modify the position of the equilibrium line in figure 56 for such stars. For details, see Harrison, Thorne, Wakano and Wheeler (1965), Thorne (1966), Cameron (1970), and Tsuruta (1971). However they will not affect the important point that a star containing slightly more than M_L/m_n nucleons will not have any zero temperature equilibrium. This is because the point at which neutrons become relativistic in a star of mass M_L almost coincides with the General Relativity limit $M/R \approx 2$. Thus a star containing somewhat more than M_L/m_n nucleons will not reach nuclear densities until it is already inside its Schwarzschild radius.

The life history of a star will lie in a vertical line on figure 56, unless it manages to lose a significant amount of material by some process. The star will condense out of a cloud of gas. As it contracts, the temperature will rise due to the compression of the gas. If the mass is less than about $10^{-2}M_L$, the temperature will never rise sufficiently high to start nuclear reactions and the star will eventually radiate away its heat and settle down to a state in which gravity is balanced by degeneracy pressure of non-relativistic electrons. If the mass is greater than about $10^{-2}M_L$, the temperature will rise high enough to start the nuclear reaction which converts hydrogen to helium. The energy produced by this reaction will balance the energy lost by radiation and the star will spend a long period ($\sim 10^{10}(M_L/M)^2$ years) in quasi-static equilibrium. When the hydrogen in the core is exhausted, the core will contract and the temperature will rise. Further nuclear reactions may now take place, converting helium in the core into heavier elements. However the energy available from this conversion is not very great, and so the core cannot remain in this phase very long. If the mass is less than M_L, the star can settle down eventually to a white dwarf state in which it is supported by degeneracy pressure of non-relativistic electrons, or possibly to a neutron star state in which it is supported by neutron degeneracy pressure. However if the mass is more than slightly greater than M_L, there is no low temperature equilibrium state. Therefore the star must either pass within its Schwarzschild radius, or manage to eject sufficient matter that its mass is reduced to less than M_L.

Ejection of matter has been observed in supernovae and planetary nebulae, but the theory is not yet very well understood. What calculations there have been suggest that stars up to $20M_L$ may possibly be able to throw off most of their mass and leave a white dwarf or neutron star of mass less than M_L (see Weymann (1963), Colgate and White (1966), Arnett (1966), Le Blanc and Wilson (1970), and Zel'dovich and Novikov (1971)). However it is not really credible that a star of more than $20M_L$ could manage to lose more than 95 % of its matter, and so one would expect that the inner part of the star at any rate would collapse within its Schwarzschild radius. (Present calculations in fact indicate that stars of mass $M > 5M_L$ would not be able to eject sufficient mass to avoid a relativistic collapse.)

Going to larger masses, consider a body of about $10^8 M_L$. If this collapsed to its Schwarzschild radius, the density would only be of the order of 10^{-4} gm cm^{-3} (less than the density of air). If the matter were fairly cold initially, the temperature would not have risen sufficiently either to support the body or to ignite the nuclear fuel; thus there would be no possibility of mass loss, or uncertainty about the equation of state. This example also shows that the conditions when a body passes through its Schwarzschild radius need not be in any way extreme.

To summarize, it seems that certainly some, and probably most, bodies of mass $> M_L$ will eventually collapse within their Schwarzschild radius, and so give rise to a closed trapped surface. There are at least 10^9 stars more massive than M_L in our galaxy. Thus there are a large number of situations in which theorem 2 predicts the existence of singularities. We discuss the observable consequences of stellar collapse in the next sections.

9.2 Black holes

What would a collapsing body look like to an observer O who remained at a large distance from it? One can answer this if the collapse is exactly spherically symmetric, since then the solution outside the body will be the Schwarzschild solution. In this case, an observer O' on the surface of the star would pass within $r = 2m$ at some time, say 1 o'clock, as measured by his watch. He would not notice anything special at that time. However after he passes $r = 2m$ he will not be

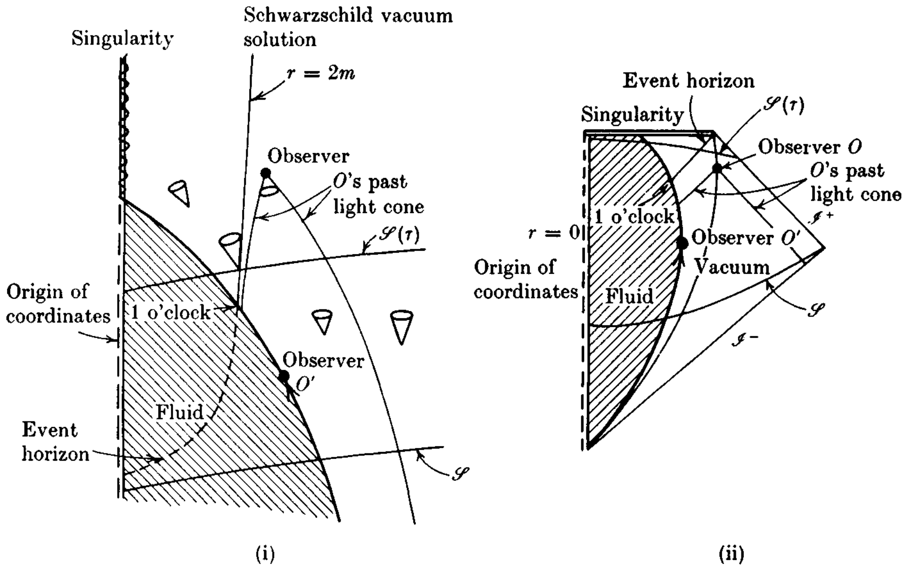

FIGURE 57. An observer O who never falls inside the collapsing fluid sphere never sees beyond a certain time (say, 1 o'clock) in the history of an observer O' on the surface of the collapsing fluid sphere.
(i) Finkelstein diagram; (ii) Penrose diagram.

visible to the observer O who remains outside $r = 2m$ (figure 57). However long the observer O waits, he will never see O' at a time later than 1 o'clock as measured by O''s watch. Instead he will see O''s watch apparently slow down and asymptotically approach 1 o'clock. This means that the light he receives from O' will have a greater and greater shift of frequency to the red and as a consequence a greater and greater decrease of intensity. Thus although the surface of the star never actually disappears from O's sight, it soon becomes so faint as to be invisible in practice. In fact O would first see the centre of the disc of the star become faint, and then this faint region would spread outwards to the limb (Ames and Thorne (1968)). The time scale for this diminution of intensity is of the order for light to travel a distance $2m$.

One would be left with an object which, for all practical purposes, is invisible. However it would still have the same Schwarzschild mass, and would still produce the same gravitational field, as it did before it collapsed. One might be able to detect its presence by its gravitational effects, for instance its effects on the orbits of nearby objects, or by the deflection of light passing near it. It is also possible that gas falling into such an object would set up a shock wave which might be a source of X-rays or radio waves.

The most striking feature of spherically symmetric collapse is that the singularity occurs within the region $r < 2m$, from which no light can escape to infinity. Thus if one remained outside $r = 2m$ one would never see the singularity predicted by theorem 2. Further the breakdown of physical theory which occurs at the singularity cannot affect one's ability to predict the future in the asymptotically flat region of space–time.

One can ask whether this is the case if the collapse is not exactly spherically symmetric. In the previous section we used the Cauchy stability theorem to show that small departures from spherical symmetry would not prevent the occurrence of closed trapped surfaces. However the Cauchy stability theorem in its present form says only that a sufficiently small perturbation in the initial data will produce a perturbation in the solution which is small on a compact region. One cannot argue from this that a perturbation of the solution will remain small at arbitrarily large times.

We expect that in general the occurrence of singularities will lead to Cauchy horizons (as in the Reissner–Nordström and Kerr solutions) and hence to a breakdown of one's ability to predict the future. However if the singularities are not visible from outside, one would still be able to predict in the exterior asymptotically flat region.

To make this precise, we shall suppose that $(\mathscr{M}, \mathbf{g})$ has a region which is asymptotically flat in the sense of being weakly asymptotically simple and empty (§ 6.9). There is then a space $(\tilde{\mathscr{M}}, \tilde{\mathbf{g}})$ into which $(\mathscr{M}, \mathbf{g})$ is conformally imbedded as a manifold with boundary $\bar{\mathscr{M}} = \mathscr{M} \cup \partial\mathscr{M}$, where the boundary $\partial\mathscr{M}$ of $\mathscr{M}$ in $\tilde{\mathscr{M}}$ consists of two null surfaces $\mathscr{I}^+$ and $\mathscr{I}^-$ which represent future and past null infinity respectively. Let $\mathscr{S}$ be a partial Cauchy surface in $\mathscr{M}$. We shall say that the space $(\mathscr{M}, \mathbf{g})$ is (*future*) *asymptotically predictable from* $\mathscr{S}$ if $\mathscr{I}^+$ is contained in the closure of $D^+(\mathscr{S})$ in the conformal manifold $\tilde{\mathscr{M}}$. Examples of spaces which are future asymptotically predictable from some surface $\mathscr{S}$ include Minkowski space, the Schwarzschild solution for $m \geqslant 0$, the Kerr solution for $m \geqslant 0$, $|a| \leqslant m$, and the Reissner–Nordström solution for $m \geqslant 0$, $|e| \leqslant m$. The Kerr solution with $|a| > m$ and the Reissner–Nordström solution with $|e| > m$ are not future asymptotically predictable, since for any partial Cauchy surface $\mathscr{S}$, there are past-inextendible non-spacelike curves from $\mathscr{I}^+$ which do not intersect $\mathscr{S}$ but approach a singularity. One can regard future asymptotic predictability as the condition that there should be no singularities to the future of $\mathscr{S}$ which are 'naked', i.e. which are visible from $\mathscr{I}^+$.

In a spherical collapse, one gets a space which is future asymptotically predictable. The question is whether this would still be the case for non-spherical collapse. We cannot answer this completely. Perturbation calculations by Doroshkevich, Zel'dovich and Novikov (1966) and Price (1971) seem to indicate that small perturbations from spherical symmetry do not give rise to naked singularities. In addition, Gibbons and Penrose (1972) have tried, and failed, to obtain contradictions which would show that in some situations the development of a future asymptotically predictable space was inconsistent. Their failure does not of course prove that asymptotic predictability will hold, but it does make it more plausible. If it does not hold, one cannot say anything definite about the evolution of any region of a space containing a collapsing star, as new information could come out of the singularity. We shall therefore proceed on the assumption that future asymptotic predictability holds at least for sufficiently small departures from spherical symmetry.

One would expect a particle on a closed trapped surface to be unable to escape to $\mathscr{I}^+$. However if one allowed arbitrary singularities one could always make suitable cuts and identifications to form an escape route for the particle. The following result shows that this is not possible in a future asymptotically predictable space.

Proposition 9.2.1

If

(*a*) $(\mathscr{M}, \mathbf{g})$ is future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$,

(*b*) $R_{ab}K^aK^b \geqslant 0$ for all null vectors K^a,

then a closed trapped surface $\mathscr{T}$ in $D^+(\mathscr{S})$ cannot intersect $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, i.e. cannot be seen from $\mathscr{I}^+$.

For suppose $\mathscr{T} \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ is non-empty. Then there would be a point $p \in \mathscr{I}^+$ in $J^+(\mathscr{T}, \bar{\mathscr{M}})$. Let $\mathscr{U}$ be the neighbourhood of $\mathscr{M}$ which is isometric to the neighbourhood $\mathscr{U}'$ of $\partial\mathscr{M}'$ in the conformal manifold $\tilde{\mathscr{M}}'$ of an asymptotically simple and empty space $(\mathscr{M}', \mathbf{g}')$. Let $\mathscr{S}'$ be a Cauchy surface in $\mathscr{M}'$, which coincides with $\mathscr{S}$ on $\mathscr{U}' \cap \mathscr{M}'$. Then $\mathscr{S}' - \mathscr{U}'$ is compact and so by lemma 6.9.3, every generator of $\mathscr{I}^+$ leaves $J^+(\mathscr{S}' - \mathscr{U}', \bar{\mathscr{M}}')$. This shows that if $\mathscr{W}$ is any compact set of $\mathscr{S}$, every generator of $\mathscr{I}^+$ leaves $J^+(\mathscr{W}, \bar{\mathscr{M}})$. From this it follows that every generator of $\mathscr{I}^+$ would leave $J^+(\mathscr{T}, \bar{\mathscr{M}})$, since this is contained in $J^+(J^-(\mathscr{T}) \cap \mathscr{S}, \bar{\mathscr{M}})$. Therefore a null geodesic generator μ of $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ would intersect $\mathscr{I}^+$. The generator μ must have past endpoint at $\mathscr{T}$, since otherwise it would intersect $I^-(\mathscr{S})$. Since μ meets $\mathscr{I}^+$ it would have infinite affine length. However by the condition (*b*) every null geodesic orthogonal to $\mathscr{T}$ would contain a point conjugate to $\mathscr{T}$ within a finite affine length. Thus it could not remain in $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ all the way out to $\mathscr{I}^+$. This shows that $\mathscr{T}$ cannot intersect $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. □

From the above it follows that a closed trapped surface in $D^+(\mathscr{S})$ in a future asymptotically predictable space must be contained in $\mathscr{M} - J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Therefore there must be a non-trivial (future) *event horizon* $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. This is the boundary of the region from which particles or photons can escape to infinity in the future direction. By § 6.3 the event horizon is an achronal boundary which is generated by null geodesic segments which may have past endpoints but which can have no future endpoints.

Lemma 9.2.2

If conditions (*a*), (*b*) of proposition 9.2.1 are satisfied and if there is a non-empty event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, then the expansion $\hat{\theta}$ of the null geodesic generators of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ is non-negative in

$$\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap D^+(\mathscr{S}).$$

Suppose there was an open set $\mathscr{U}$ such that $\hat{\theta} < 0$ in $\mathscr{U} \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Let $\mathscr{T}$ be a spacelike two-surface in $\mathscr{U} \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Then $\hat{\theta} = \chi_2{}^a{}_a < 0$. Let $\mathscr{V}$ be an open subset of $\mathscr{U}$ which intersects $\mathscr{T}$ and has compact closure contained in $\mathscr{U}$. One can vary $\mathscr{T}$ by a small amount in $\mathscr{V}$ so that $\chi_2{}^a{}_a$ is still negative but such that in $\mathscr{U}$, $\mathscr{T}$ intersects $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. As before, this leads to a contradiction since any generator of $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ in $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ would have past endpoint at $\mathscr{T}$ in $\mathscr{V}$, where it would be orthogonal to $\mathscr{T}$. However as $\chi_2{}^a{}_a < 0$ in $\mathscr{V}$, every outgoing null geodesic orthogonal to $\mathscr{T}$ in $\mathscr{V}$ would contain a point conjugate to $\mathscr{T}$ within a finite affine distance, and so could not remain in $\dot{J}^+(\mathscr{T}, \bar{\mathscr{M}})$ all the way out to $\mathscr{I}^+$. □

In a future asymptotically predictable space, $J^+(\mathscr{S}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ is contained in $D^+(\mathscr{S})$. If there were a point p on the event horizon in $J^+(\mathscr{S})$ which was not in $D^+(\mathscr{S})$, the smallest perturbation could lead to p being in $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, i.e. being visible from infinity, which would mean that the space was no longer asymptotically predictable. It therefore seems reasonable to slightly extend the definition of future asymptotically predictable, to say that space–time is *strongly future asymptotically predictable* from a partial Cauchy surface $\mathscr{S}$ if $\mathscr{I}^+$ is contained in the closure of $D^+(\mathscr{S})$ in $\bar{\mathscr{M}}$, and $J^+(\mathscr{S}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ is contained in $D^+(\mathscr{S})$. In other words, one can also predict a neighbourhood of the event horizon from $\mathscr{S}$.

Proposition 9.2.3

If $(\mathscr{M}, \mathbf{g})$ is strongly future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$, there is a homeomorphism

$$\alpha: (0, \infty) \times \mathscr{S} \to D^+(\mathscr{S}) - \mathscr{S}$$

such that for each $\tau \in (0, \infty)$, $\mathscr{S}(\tau) \equiv (\{\tau\} \times \mathscr{S})$ is a partial Cauchy surface such that:

(*a*) for $\tau_2 > \tau_1$, $\mathscr{S}(\tau_2) \subset I^+(\mathscr{S}(\tau_1))$;

(*b*) for each τ, the edge of $\mathscr{S}(\tau)$ in the conformal manifold $\bar{\mathscr{M}}$ is a spacelike two-sphere $\mathscr{Q}(\tau)$ in $\mathscr{I}^+$ such that for $\tau_2 > \tau_1$, $\mathscr{Q}(\tau_2)$ is strictly to the future of $\mathscr{Q}(\tau_1)$,

(*c*) for each τ, $\mathscr{S}(\tau) \cup \{\mathscr{I}^+ \cap J^-(\mathscr{Q}(\tau), \bar{\mathscr{M}})\}$ is a Cauchy surface in $\bar{\mathscr{M}}$ for $D(\mathscr{S})$.

In other words, $\mathscr{S}(\tau)$ is a family of spacelike surfaces homeomorphic to $\mathscr{S}$ which cover $D^+(\mathscr{S}) - \mathscr{S}$ and intersect $\mathscr{I}^+$ (see figure 58). One could regard them as surfaces of constant time in the asymptotically predictable region. We choose them to intersect $\mathscr{I}^+$ so that the mass measured on them at infinity will decrease when the emission of gravitational or other forms of radiation takes place.

The construction for $\mathscr{S}(\tau)$ is rather similar to that of proposition 6.4.9. Choose a continuous family $\mathscr{Q}(\tau)$ $(\infty > \tau > 0)$ of spacelike two-spheres which cover $\mathscr{I}^+$, such that for $\tau_2 > \tau_1$, $\mathscr{Q}(\tau_2)$ is strictly to the future of $\mathscr{Q}(\tau_1)$. Put a volume measure on $\mathscr{M}$ such that the total volume of $\mathscr{M}$ in this measure is finite. We first prove:

Lemma 9.2.4

$k(\tau)$, the volume of the set $I^-(\mathscr{Q}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S})$ is a continuous function of τ.

Let $\mathscr{V}$ be any open set with compact closure contained in

$$I^-(\mathscr{Q}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S}).$$

Then there are timelike curves from every point of $\mathscr{V}$ to $\mathscr{Q}(\tau)$, which can be deformed to give timelike curves to $\mathscr{Q}(\tau - \delta)$ for some $\delta > 0$.

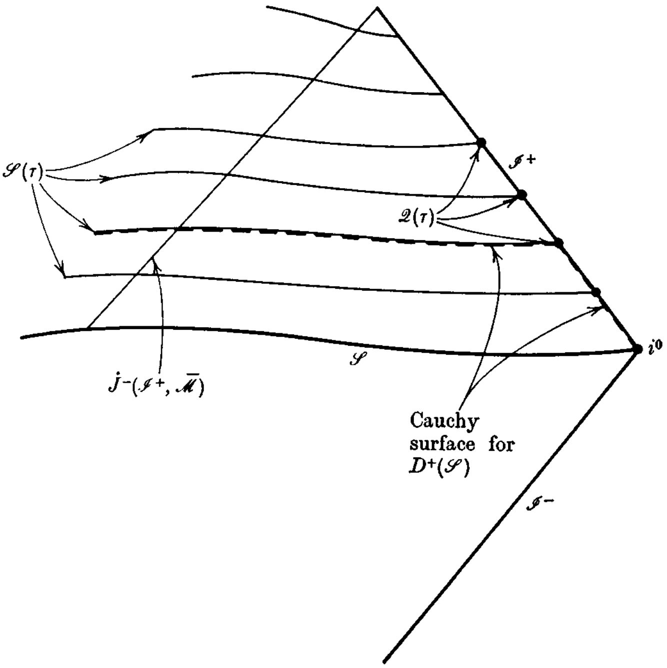

FIGURE 58. A space $(\mathscr{M}, \mathbf{g})$ which is strongly future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$, showing a family $\mathscr{S}(\tau)$ of spacelike surfaces which cover $D^+(\mathscr{S})-\mathscr{S}$ and intersect $\mathscr{I}^+$ in a family of two-spheres $\mathscr{D}(\tau)$.

Given any $\epsilon > 0$, one can find a $\mathscr{V}$ whose volume is $> k(\tau)-\epsilon$. Thus there is a $\delta > 0$ such that $k(\tau-\delta) > k(\tau)-\epsilon$. On the other hand, suppose there were an open set $\mathscr{W}$ which did not intersect $I^-(\mathscr{D}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S})$ but which was contained in $I^-(\mathscr{D}(\tau'), \bar{\mathscr{M}}) \cap D^+(\mathscr{S})$ for any $\tau' > \tau$. Then if $p \in \mathscr{W}$, there would be past-directed timelike curves $\lambda_{\tau'}$ from each $\mathscr{D}(\tau')$ to p. As the region of $\mathscr{I}^+$ between $\mathscr{D}(\tau)$ and $\mathscr{D}(\tau_1)$ is compact for any $\tau_1 > \tau$, there would be a past-directed non-spacelike curve λ from $\mathscr{D}(\tau)$ which was the limit curve of the $\{\lambda_{\tau'}\}$. Since the $\{\lambda_{\tau'}\}$ did not intersect $I^-(\mathscr{D}(\tau), \bar{\mathscr{M}})$, λ would not either, and so it would be a null geodesic and would lie in $\dot{I}^-(\mathscr{D}(\tau), \bar{\mathscr{M}})$. It would enter $\mathscr{M}$ and so it would either have a past endpoint at p, or would intersect $\mathscr{S}$. The former is impossible as it would imply that $\mathscr{W}$ intersected $I^-(\mathscr{D}(\tau), \bar{\mathscr{M}})$, and the latter is impossible as $p \in I^+(\mathscr{S})$. This shows that there is no open set which is in $I^-(\mathscr{D}(\tau'), \bar{\mathscr{M}})$ for every $\tau' > \tau$, but which is not in $I^-(\mathscr{D}(\tau), \bar{\mathscr{M}}) \cap D^+(\mathscr{S})$. Thus given ϵ, there is a δ such that

$$k(\tau+\delta) < k(\tau)+\epsilon.$$

Therefore $k(\tau)$ is continuous. □

Proof of proposition 9.2.3. Define functions $f(p)$ and $h(p,\tau)$, $p\in D^+(\mathscr{S})$, which are volumes of $I^+(p)$ and $I^-(p)-\overline{I^-}(\mathscr{Q}(\tau),\bar{\mathscr{M}})$. As in proposition 6.4.9, the function $f(p)$ is continuous on the globally hyperbolic region $D^+(\mathscr{S})-\mathscr{S}$, and goes to zero on every future-inextendible non-spacelike curve. Since $\overline{I^-}(\mathscr{Q}(\tau),\bar{\mathscr{M}})\cap\mathscr{M}$ is a past set,

$$D^+(\mathscr{S})-\overline{I^-}(\mathscr{Q}(\tau),\bar{\mathscr{M}})-\mathscr{S}$$

is globally hyperbolic. Thus for each τ, $h(p,\tau)$ is continuous on $D^+(\mathscr{S})-\mathscr{S}$. This means that given any $\epsilon>0$, one can find a neighbourhood $\mathscr{U}$ of p such that $|h(q,\tau)-h(p,\tau)|<\frac{1}{2}\epsilon$ for any $q\in\mathscr{U}$. By lemma 9.2.4, one can find a $\delta>0$ such that $|k(\tau')-k(\tau)|<\frac{1}{2}\epsilon$ for $|\tau'-\tau|<\delta$. Then $|h(q,\tau')-h(p,\tau)|<\epsilon$, which shows that $h(p,\tau)$ is continuous on $(D^+(\mathscr{S})-\mathscr{S})\times(0,\infty)$. The surfaces $\mathscr{S}(\tau)$ can then be defined as the set of points $p\in D^+(\mathscr{S})-\mathscr{S}$ such that $h(p,\tau)=\tau f(p)$. Clearly these are spacelike surfaces which cover $D^+(\mathscr{S})-\mathscr{S}$ and satisfy properties (*a*)–(*c*).

To define the homeomorphism α, one needs a timelike vector field on $D^+(\mathscr{S})-\mathscr{S}$ which intersects each surface $\mathscr{S}(\tau)$. We construct such a vector field as follows. Let $\mathscr{V}$ be a neighbourhood of $\mathscr{I}^+$ in the conformal manifold $\tilde{\mathscr{M}}$, let $\mathbf{X}_1$ be a non-spacelike vector field on $\mathscr{V}$ which is tangent to the generators of $\mathscr{I}^+$ on $\mathscr{I}^+$, and let $x_1\geqslant 0$ be a C^2 function which vanishes outside $\mathscr{V}$ and is non-zero on $\mathscr{I}^+$. Let $\mathbf{X}_2$ be a timelike vector field on $\mathscr{M}$, and let $x_2\geqslant 0$ be a C^2 function on $\tilde{\mathscr{M}}$ which is non-zero on $\mathscr{M}$ and is zero on $\mathscr{I}^+$. Then the vector field $\mathbf{X}=x_1\mathbf{X}_1+x_2\mathbf{X}_2$ has the required property. The homeomorphism $\alpha\colon D^+(\mathscr{S})-\mathscr{S}\to(0,\infty)\times\mathscr{S}$ then maps a point $p\in D^+(\mathscr{S})-\mathscr{S}$ to (τ,q) where τ is such that $p\in\mathscr{S}(\tau)$, and the integral curve of $\mathbf{X}$ through p intersects $\mathscr{S}$ at q. □

If there is an event horizon $\dot{J}^-(\mathscr{I}^+,\bar{\mathscr{M}})$ in the region $D^+(\mathscr{S})$ of a future asymptotically predictable space, then it follows from property (*b*) of proposition 9.2.3 that for sufficiently large τ, the surfaces $\mathscr{S}(\tau)$ will intersect it. We define a *black hole* on the surface $\mathscr{S}(\tau)$ to be a connected component of the set $\mathscr{B}(\tau)\equiv\mathscr{S}(\tau)-J^-(\mathscr{I}^+,\bar{\mathscr{M}})$. In other words, it is a region of $\mathscr{S}(\tau)$ from which particles or photons cannot escape to $\mathscr{I}^+$.

As τ increases, black holes can merge together, and new black holes can form as the result of further bodies collapsing. However, the following result shows that black holes can never bifurcate.

Proposition 9.2.5

Let $\mathscr{B}_1(\tau_1)$ be a black hole on $\mathscr{S}(\tau_1)$. Let $\mathscr{B}_2(\tau_2)$ and $\mathscr{B}_3(\tau_2)$ be black holes on a later surface $\mathscr{S}(\tau_2)$. If $\mathscr{B}_2(\tau_2)$ and $\mathscr{B}_3(\tau_2)$ both intersect $J^+(\mathscr{B}_1(\tau_1))$, then $\mathscr{B}_2(\tau_2) = \mathscr{B}_3(\tau_2)$.

By property (*c*) of proposition 9.2.3, every future-directed inextendible timelike curve from $\mathscr{B}_1(\tau_1)$ will intersect $\mathscr{S}(\tau_2)$. Thus

$$J^+(\mathscr{B}_1(\tau_1)) \cap \mathscr{S}(\tau_2)$$

is connected, and will be contained in a connected component of $\mathscr{B}(\tau_2)$. □

For physical applications, one is interested primarily in black holes which form as the result of gravitational collapse from an initially non-singular state. To make this notion precise, we shall say that the partial Cauchy surface $\mathscr{S}$ has an *asymptotically simple past* if $J^-(\mathscr{S})$ is isometric to the region $J^-(\mathscr{S}')$ of some asymptotically simple and empty space–time $(\mathscr{M}', \mathbf{g}')$, where $\mathscr{S}'$ is a Cauchy surface for $(\mathscr{M}', \mathbf{g}')$. By proposition 6.9.4, the surface $\mathscr{S}'$ has the topology R^3 and so $\mathscr{S}$ also has this topology. Proposition 9.2.3 therefore shows that if $(\mathscr{M}, \mathbf{g})$ is strongly future asymptotically predictable from a surface $\mathscr{S}$ with an asymptotically simple past, then each surface $\mathscr{S}(\tau)$ has the topology R^3, and the union of $\mathscr{S}(\tau)$ with the boundary two-sphere $\mathscr{Q}(\tau)$ on $\mathscr{I}^+$ is homeomorphic to the unit cube I^3.

Although one is primarily interested in spaces which have asymptotically simple pasts it will in the next section be convenient to consider future asymptotically predictable spaces which do not have this property, but which at large times may closely approximate to spaces which do. An example of this is the spherically symmetric collapse we considered at the beginning of this section. Once the surface of the star has passed inside the event horizon, the metric of the exterior region is that of the Schwarzschild solution, and is unaffected by the fate of the star. When studying the asymptotic behaviour it is therefore convenient simply to forget about the star, and consider the empty Schwarzschild solution as a space which is strongly future asymptotically predictable from a surface $\mathscr{S}$ such as that shown in figure 24 on p. 154. This surface does not have an asymptotically simple past, and its topology is $S^2 \times R^1$ instead of R^3. However the portion of $\mathscr{S}$ outside the event horizon in region I has the same topology as the region outside the event horizon of the surface $\mathscr{S}(\tau)$ in figure 57. We want to consider spaces which are strongly future asymptotically predictable from a surface $\mathscr{S}$, and are such that the portion of $\mathscr{S}$ outside the event horizon has the same topology as some surface $\mathscr{S}(\tau)$ in a space with an asymptotically simple past. Of course in more complicated cases there may be several components of $\mathscr{B}(\tau)$, corresponding to the collapse of several bodies. We shall therefore consider spaces which are strongly future asymptotically predictable from a surface $\mathscr{S}$, and with the property:

(α) $\mathscr{S} \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ is homeomorphic to R^3 – (an open set with compact closure).

(Note that this open set may not be connected.) It will also be convenient to demand the property:

(β) $\mathscr{S}$ is simply connected.

Proposition 9.2.6

Let $(\mathscr{M}, \mathbf{g})$ be a space which is strongly future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$ which satisfies (α), (β). Then:

(1) the surfaces $\mathscr{S}(\tau)$ also satisfy (α), (β);

(2) for each τ, $\partial\mathscr{B}_1(\tau)$, the boundary in $\mathscr{S}(\tau)$ of a black hole $\mathscr{B}_1(\tau)$, is compact and connected.

Since the surfaces $\mathscr{S}(\tau)$ are homeomorphic to $\mathscr{S}$, they satisfy property (β). One can define an injective map

$$\gamma: \mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \to \mathscr{S} \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$$

by mapping each point of $\mathscr{S}(\tau)$ down the integral curves of the vector field of $\mathbf{X}$ proposition 9.2.3. Since $(\mathscr{M}, \mathbf{g})$ is weakly asymptotically simple, one can find a two-sphere $\mathscr{P}$ near $\mathscr{I}^+$ in $\mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$. The portion of $\mathscr{S}(\tau)$ outside $\mathscr{P}$ will map into the region of $\mathscr{S}$ outside the two-sphere $\gamma(\mathscr{P})$. This shows that the region of $\mathscr{S} \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ which is not in $\gamma(\mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}))$ must have compact closure. Therefore $\gamma(\mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}))$ will be homeomorphic to R^3 – (an open set with compact closure). Since $\mathscr{S}(\tau)$ is homeomorphic to $R^3 - \mathscr{V}$ where $\mathscr{V}$ is an open subset of R^3 with compact closure, $\partial\mathscr{B}(\tau)$ will be homeomorphic to $\partial\mathscr{V}$ and so will be compact. $\partial\mathscr{B}_1(\tau)$ being a closed subset of $\partial\mathscr{B}(\tau)$ will be compact.

Suppose that $\partial\mathscr{B}_1(\tau)$ consisted of two disconnected components $\partial\mathscr{B}_1{}^1(\tau)$ and $\partial\mathscr{B}_1{}^2(\tau)$. One could find curves λ_1 and λ_2 in $\mathscr{S}(\tau) - \mathscr{B}(\tau)$ from $\mathscr{Q}(\tau)$ to $\partial\mathscr{B}_1{}^1(\tau)$ and $\partial\mathscr{B}_1{}^2(\tau)$ respectively. One could also find a curve μ in $\operatorname{int}\mathscr{B}_1(\tau)$ from $\partial\mathscr{B}_1{}^1(\tau)$ to $\partial\mathscr{B}_1{}^2(\tau)$. Joining these together one would obtain a closed curve in $\mathscr{S}(\tau)$ which crossed $\partial\mathscr{B}_1{}^1(\tau)$ only once. This cannot be deformed to zero in $\mathscr{S}(\tau)$, contradicting the fact that $\mathscr{S}(\tau)$ is simply connected. □

We are only interested in black holes that one can actually fall into, i.e. ones in which the boundary $\partial\mathscr{B}(\tau)$ is contained in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. We shall therefore add to properties (α), (β) the requirement:

(γ) for sufficiently large τ, $\mathscr{S}(\tau) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ is contained in $\overline{J^+}(\mathscr{I}^-, \bar{\mathscr{M}})$.

We shall say that $(\mathscr{M}, \mathbf{g})$ is a *regular predictable space* if it is strongly future asymptotically predictable from a partial Cauchy surface $\mathscr{S}$ and if properties (α), (β), (γ) are satisfied. All the spaces mentioned at the beginning of this section as being future asymptotically predictable are in fact also regular predictable spaces. Proposition 9.2.6 shows that when one is dealing with regular predictable spaces developing from a partial Cauchy surface $\mathscr{S}$, there is a one–one correspondence between black holes $\mathscr{B}_i(\tau)$ and their boundaries $\partial\mathscr{B}_i(\tau)$ in $\mathscr{S}(\tau)$. One could therefore in such a situation give an equivalent definition of a black hole as a connected component of $\mathscr{S}(\tau) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$.

The next result gives a property of the boundaries of black holes which will be important in the next section.

Proposition 9.2.7

Let $(\mathscr{M}, \mathbf{g})$ be a regular predictable space developing from a partial Cauchy surface $\mathscr{S}$, in which $R_{ab}K^aK^b \geqslant 0$ for every null vector K^a. Let $\mathscr{B}_1(\tau)$ be a black hole on the surface $\mathscr{S}(\tau)$, and let $\{\mathscr{B}_i(\tau')\}$ ($i = 1$ to N) be the black holes on an earlier surface $\mathscr{S}(\tau')$ which are such that $J^+(\mathscr{B}_i(\tau')) \cap \mathscr{B}_1(\tau) \neq \varnothing$. Then the area $A_1(\tau)$ of $\partial\mathscr{B}_1(\tau)$ is greater than or equal to the sum of the areas $A_i(\tau')$ of $\partial\mathscr{B}_i(\tau')$; the equality can hold only if $N = 1$.

In other words, the area of the boundary of a black hole cannot decrease with time, and if two or more black holes merge to form a single black hole, the area of its boundary will be greater than the areas of the boundaries of the original black holes.

Since the event horizon is the boundary of the past of $\mathscr{I}^+$, its null geodesic generators would have future endpoints only if they intersected $\mathscr{I}^+$. However this is impossible, as the null geodesic generators of $\mathscr{I}^+$ have no future endpoints. Thus the null generators of the event horizon have no future endpoints. By lemma 9.2.2, their expansion $\hat{\theta}$ is non-negative. Thus the area of a two-dimensional cross-section of the generators cannot decrease with τ. By property (c) of proposition 9.2.3, and by proposition 9.2.5, all the null geodesic generators of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ which intersect $\mathscr{S}(\tau')$ in any of the $\partial\mathscr{B}_i(\tau')$ must intersect $\mathscr{S}(\tau)$ in $\partial\mathscr{B}_1(\tau)$. Thus the area of $\partial\mathscr{B}_1(\tau)$ is greater than or equal to the sum of the areas of the $\{\mathscr{B}_i(\tau')\}$. When $N > 1$, $\partial\mathscr{B}_1(\tau)$ will contain N disjoint closed subsets which correspond to the generators of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ which intersect each $\partial\mathscr{B}_i(\tau')$. Since $\partial\mathscr{B}_1(\tau)$ is connected, it must contain an open set of generators which do not intersect any $\partial\mathscr{B}_i(\tau')$, but have past endpoints between $\mathscr{S}(\tau)$ and $\mathscr{S}(\tau')$. □

It has been convenient to define black holes in terms of the event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, because this is a null hypersurface with a number of nice properties. However this definition depends on the whole future behaviour of the solution; given the partial Cauchy surface $\mathscr{S}(\tau)$, one cannot find where the event horizon is without solving the Cauchy problem for the whole future development of the surface. It is therefore useful to define a different sort of horizon which depends only on the properties of space–time on the surface $\mathscr{S}(\tau)$.

One knows from proposition 9.2.1 that any closed trapped surface on $\mathscr{S}(\tau)$ in a regular predictable space developing from a partial Cauchy surface $\mathscr{S}$ must be in $\mathscr{B}(\tau)$. This result depends only on the fact that the outgoing null geodesics orthogonal to the two-surface are converging. It does not matter whether the ingoing null geodesics are converging or not. We shall therefore say that an orientable compact spacelike two-surface in $D^+(\mathscr{S})$ is an *outer trapped surface* if the expansion $\hat{\theta}$ of the outgoing null geodesics orthogonal to it is non-positive. (We include the case $\hat{\theta} = 0$ for convenience.) In order to define which is the outgoing family of null geodesics we make use of property (β) of the partial Cauchy surfaces $\mathscr{S}(\tau)$. Let $\mathbf{X}$ be the timelike vector field of proposition 9.2.3. Then any compact orientable spacelike two-surface $\mathscr{P}$ in $D^+(\mathscr{S})$ can be mapped by the integral curves of $\mathbf{X}$ into a compact orientable two-surface $\mathscr{P}'$ in $\mathscr{S}(\tau)$, for any given value of τ. Let λ be a curve in $\mathscr{S}(\tau) \cup \mathscr{Q}(\tau)$ from $\mathscr{Q}(\tau)$ to $\mathscr{P}'$ which intersects $\mathscr{P}'$ only at its endpoint. Then one can define the outgoing direction on $\mathscr{P}'$ in $\mathscr{S}(\tau)$ as the direction for which λ approaches $\mathscr{P}'$. As $\mathscr{S}(\tau)$ is simply connected, this definition is unique. The outgoing family of null geodesics orthogonal to $\mathscr{P}$ is then that family which is mapped by $\mathbf{X}$ onto curves in $\mathscr{S}(\tau)$ which are outgoing for $\mathscr{P}'$.

Knowing the solution on the surface $\mathscr{S}(\tau)$, one can find all the outer trapped surfaces $\mathscr{P}$ which lie in $\mathscr{S}(\tau)$. We shall define the *trapped* *region* $\mathcal{T}(\tau)$ in the surface $\mathcal{S}(\tau)$ as the set of all points $q \in \mathcal{S}(\tau)$ such that there is an outer trapped surface $\mathcal{P}$ lying in $\mathcal{S}(\tau)$, through q. As is shown by the following result, the existence of the trapped region $\mathcal{T}(\tau)$ implies the existence of a black hole $\mathcal{B}(\tau)$, and in fact $\mathcal{T}(\tau)$ lies in $\mathcal{B}(\tau)$ for each value of τ.

Proposition 9.2.8

Let $(\mathcal{M}, \mathbf{g})$ be a regular predictable space developing from a partial Cauchy surface $\mathcal{S}$, in which $R_{ab}K^aK^b \geqslant 0$ for any null vector K^a. Then an outer trapped surface $\mathcal{P}$ in $D^+(\mathcal{S})$ does not intersect $J^-(\mathcal{I}^+, \bar{\mathcal{M}})$.

The proof is similar to that of proposition 9.2.1. Suppose $\mathcal{P}$ intersects $J^-(\mathcal{I}^+, \bar{\mathcal{M}})$. Then $\dot{J}^+(\mathcal{P}, \bar{\mathcal{M}})$ would intersect $\mathcal{I}^+$. To each point of $\mathcal{I}^+ \cap \dot{J}^+(\mathcal{P}, \bar{\mathcal{M}})$ there would be a past-directed null geodesic generator of $\dot{J}^+(\mathcal{P}, \bar{\mathcal{M}})$ which had past endpoint on $\mathcal{P}$, and which contained no point conjugate to $\mathcal{P}$. By (4.35) the expansion $\hat{\theta}$ of these generators would be non-positive, as it is non-positive at $\mathcal{P}$ and as $R_{ab}K^aK^b \geqslant 0$. Thus the area of a two-dimensional cross-section of the generators would always be less than or equal to the area of $\mathcal{P}$. This establishes a contradiction, as the area of $\mathcal{I}^+ \cap \dot{J}^+(\mathcal{P}, \bar{\mathcal{M}})$ is infinite, as it is at infinity. □

We shall call the outer boundary $\partial\mathcal{T}_1(\tau)$ of a connected component $\mathcal{T}_1(\tau)$ of the trapped region $\mathcal{T}(\tau)$, an *apparent horizon*. By the previous result, the existence of an apparent horizon $\partial\mathcal{T}_1(\tau)$ implies the existence of a component $\partial\mathcal{B}_1(\tau)$ of the event horizon outside it, or coinciding with it. However the converse is not necessarily true: there may not be outer trapped surfaces within an event horizon.

On the other hand, there may be more than one connected component of $\mathcal{T}(\tau)$ within one component $\partial\mathcal{B}_1(\tau)$ of the event horizon. These possibilities are illustrated in figure 59. A similar situation arises when one considers the collision and merger of two black holes. On an initial surface $\mathcal{S}(\tau_1)$, one would have two separate trapped regions $\mathcal{T}_1(\tau_1)$ and $\mathcal{T}_2(\tau_1)$ contained in black holes $\mathcal{B}_1(\tau)_1$ and $\mathcal{B}_2(\tau_1)$ respectively. As they approached each other, the two components $\partial\mathcal{B}_1(\tau)$ and $\partial\mathcal{B}_2(\tau)$ of the event horizon would amalgamate to form a single black hole $\mathcal{B}_3(\tau_2)$ on a later surface $\mathcal{S}(\tau_2)$. The apparent horizons $\partial\mathcal{T}_1(\tau)$ and $\partial\mathcal{T}_2(\tau)$ would however not join up immediately. Instead what would happen is that a third trapped region $\mathcal{T}_3(\tau)$ would develop surrounding

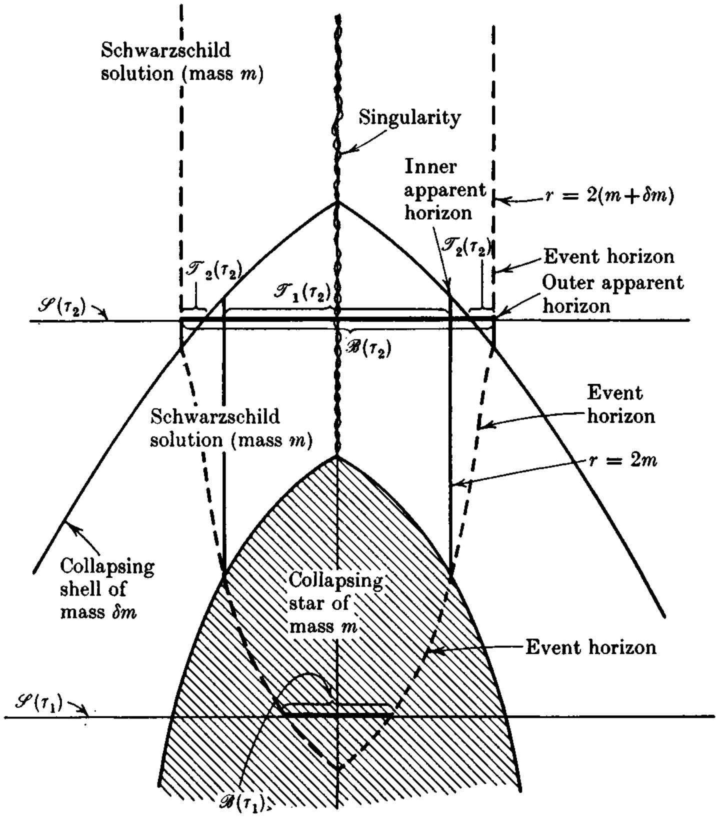

FIGURE 59. The spherical collapse of a star of mass m, followed by the spherical collapse of a shell of matter of mass δm; the exterior solution will be a Schwarzschild solution of mass m after the collapse of the star, and a Schwarzschild solution of mass $m+\delta m$ after the collapse of the shell. At time τ_1 there is an event horizon but no apparent event horizon; at time τ_2 there are two apparent horizons within the event horizon.

them both (figure 60). At some later time, $\mathscr{T}_1$, $\mathscr{T}_2$ and $\mathscr{T}_3$ might merge together.

We shall only outline the proofs of the principal properties of the apparent horizon. First of all one has:

Proposition 9.2.9

Each component of $\partial\mathscr{T}(\tau)$ is a two-surface such that the outgoing orthogonal null geodesics have zero convergence $\hat{\theta}$ on $\partial\mathscr{T}(\tau)$. (We shall call such a surface, a *marginally outer trapped surface.*)

If $\hat{\theta}$ were positive in a neighbourhood in $\partial\mathscr{T}(\tau)$ of a point $p \in \partial\mathscr{T}(\tau)$, then there would be a neighbourhood $\mathscr{U}$ of p such that any outer

https://doi.org/10.1017/CBO9780511524646.010 Published online by Cambridge University Press

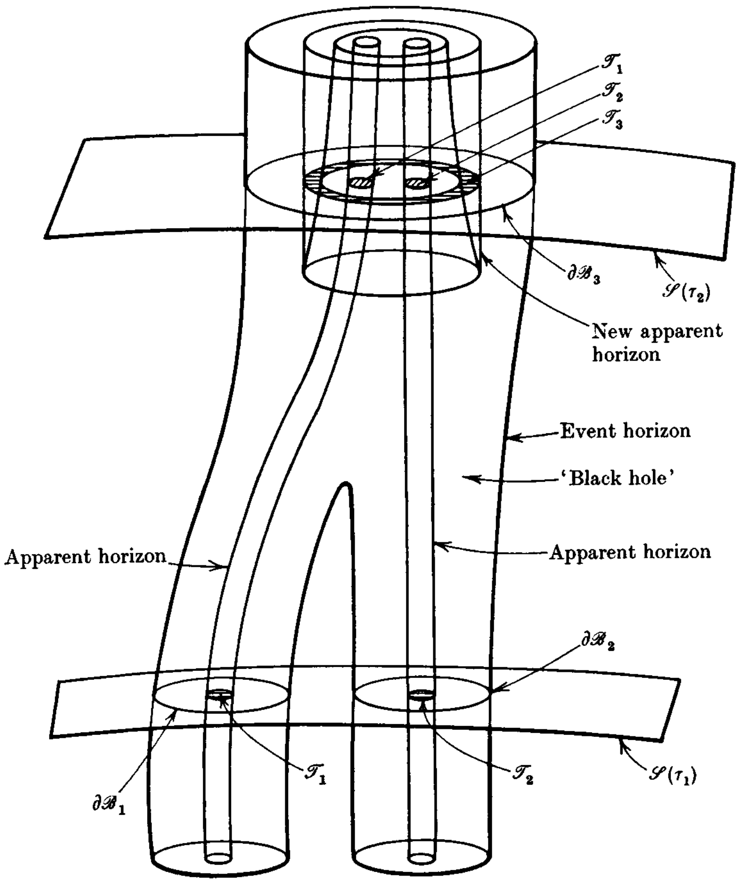

FIGURE 60. The collision and merging of two black holes. At time τ_1, there are apparent horizons $\partial\mathscr{T}_1$, $\partial\mathscr{T}_2$ inside the event horizons $\partial\mathscr{B}_1$, $\partial\mathscr{B}_2$ respectively. By time τ_2, the event horizons have merged to form a single event horizon; a third apparent horizon has now formed surrounding both the previous apparent horizons.

trapped surface in $\mathscr{S}(\tau)$ which intersected $\mathscr{U}$ would also intersect $\partial\mathscr{T}(\tau)$. Thus $\hat{\theta} \leqslant 0$ on $\partial\mathscr{T}(\tau)$.

If $\hat{\theta}$ were negative in a neighbourhood in $\partial\mathscr{T}(\tau)$ of a point $p \in \partial\mathscr{T}(\tau)$, one could deform $\partial\mathscr{T}(\tau)$ outwards in $\mathscr{S}(\tau)$ to obtain an outer trapped surface outside $\partial\mathscr{T}(\tau)$. □

The null geodesics orthogonal to the apparent horizon $\partial\mathscr{T}(\tau)$ on a surface $\mathscr{S}(\tau)$ will therefore start out with zero convergence. However if they encounter any matter or any Weyl tensor satisfying the generality condition (§4.4), they will start converging, and so their intersection with a later surface $\mathscr{S}(\tau')$ will lie inside the apparent horizon $\partial\mathscr{T}(\tau')$. In other words, the apparent horizon moves outwards at least as fast as light; and moves out faster than light if any matter or radiation falls through it. As the example above shows, the apparent horizon can also jump outwards discontinuously. This makes it harder to work with than the event horizon, which always moves in a continuous manner. We shall show in the next section that the event and apparent horizons coincide when the solution is stationary. One would therefore expect them to be very close together if the solution is nearly stationary for a long time. In particular, one would expect their areas to be almost the same under such circumstances. If one has a solution which passes from an initial nearly stationary state through some non-stationary period to a final nearly stationary state, one can employ proposition 9.2.7 to relate the areas of the initial and final horizons.

9.3 The final state of black holes

In the last section, we assumed that one could predict the future far away from a collapsing star. We showed that this implied that the star passed inside an event horizon which hid the singularities from an outside observer. Matter and energy which crossed the event horizon would be lost for ever from the outside world. One would therefore expect that there would be a limited amount of energy available to be radiated to infinity in the form of gravitational waves. Once most of this energy had been emitted, one would expect the solution outside the horizon to approach a stationary state. In this section we shall therefore study black hole solutions which are exactly stationary, in the expectation that the exterior regions will closely represent the final states of solutions outside collapsed objects.

More precisely, we shall consider spaces $(\mathscr{M}, \mathbf{g})$ which satisfy the following conditions:

(1) $(\mathscr{M}, \mathbf{g})$ is a regular predictable space developing from a partial Cauchy surface $\mathscr{S}$.

(2) There exists an isometry group $\theta_t: \mathscr{M} \rightarrow \mathscr{M}$ whose Killing vector $\mathbf{K}$ is timelike near $\mathscr{I}^+$ and $\mathscr{I}^-$.

(3) $(\mathscr{M}, \mathbf{g})$ is empty or contains fields like the electromagnetic field or scalar field which obey well-behaved hyperbolic equations, and satisfy the dominant energy condition: $T_{ab}N^aL^b \geqslant 0$ for future-directed timelike vectors $\mathbf{N}, \mathbf{L}$.

We shall call a space satisfying these conditions, a *stationary regular predictable space*. We expect that for large values of τ, the region $J^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{S}(\tau))$ of a regular predictable space containing collapsing stars will be almost isometric to a similar region of a stationary regular predictable space.

The justification for condition (3) is that one would expect any non-zero rest-mass matter eventually to fall through the horizon. Only long range fields like the electromagnetic field would be left. Conditions (2) and (3) imply that $(\mathscr{M}, \mathbf{g})$ is analytic in the region near infinity where the Killing vector field **K** is timelike (Müller zum Hagen (1970)). We shall take the solution elsewhere to be the analytic continuation of this outer region. The stationary solutions we are considering here will not have asymptotically simple pasts, as they represent only the final state of the system and not the earlier dynamical stage. However we shall be concerned only with the future properties of these solutions, and not their past properties. These might not be the same, as there is no *a priori* reason why they should be time reversible, though in fact it will be a consequence of the results we shall prove that they are time reversible.

In a stationary regular predictable space, the area of a two-section of the horizon will be time independent. This gives the following fundamental result:

Proposition 9.3.1

Let $(\mathscr{M}, \mathbf{g})$ be a stationary, regular predictable space–time. Then the generators of the future event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ have no past end-points in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Let $Y_1{}^a$ be the future-directed tangent vectors to these generators; then in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$, $Y_1{}^a$ has zero shear $\hat{\sigma}$ and expansion $\hat{\theta}$, and satisfies

$$R_{ab} Y_1{}^a Y_1{}^b = 0 = Y_{1[e} C_{a]bc[d} Y_{1f]} Y_1{}^b Y_1{}^c.$$

In order not to break up the discussion we shall defer the proof of this and other results to the end of this section. This proposition shows that in a stationary space–time, the apparent horizon coincides with the event horizon.

We shall now present some results which indicate that the Kerr family of solutions (§ 5.6) are probably the only empty stationary regular predictable space–times. We shall not give the proofs of the theorems of Israel and Carter here, but shall refer to the literature. The other results will be proved at the end of this section. Because of these results, we expect that the solution outside an uncharged collapsed object will settle down to a Kerr solution. If the collapsed body had a net electric charge, we would expect the solution to approach one of the charged Kerr solutions.

Proposition 9.3.2

Each connected component in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ of the horizon $\partial\mathscr{B}(\tau)$ in a stationary regular predictable space is homeomorphic to a two-sphere.

It is possible that there could be several connected components of $\partial\mathscr{B}(\tau)$ representing several black holes at constant distances from each other. This situation can occur in the limiting case where the black holes have charge e equal to their mass m, and are non-rotating (Hartle and Hawking (1972*a*)). It seems probable that this is the only case in which one can get a sufficiently strong repulsive force to balance the gravitational attraction between the black holes. We shall therefore consider solutions where $\partial\mathscr{B}(\tau)$ has only one connected component.

Proposition 9.3.3

Let $(\mathscr{M}, \mathbf{g})$ be a stationary regular predictable space. Then the Killing vector K^a is non-zero in $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, which is simply connected. Let τ_0 be such that $\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ is contained in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. If $\partial\mathscr{B}(\tau_0)$ has only one connected component, then $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M}$ is homeomorphic to $[0, 1) \times S^2 \times R^1$.

The discussion now takes one of two possible courses, depending on whether or not the Killing vector K^a has zero curl, $K_{a;b}K_c\eta^{abcd}$, everywhere. If the curl is zero, the solution is said to be a *static regular predictable space–time*. Roughly speaking, one would expect the solution to be static if the black hole is not rotating in some sense.

Proposition 9.3.4

In a static regular predictable space–time, the Killing vector $\mathbf{K}$ is timelike in the exterior region $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ and is non-zero and directed along the null generators of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ on

$$\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}}).$$

Since the curl of $\mathbf{K}$ vanishes, it is hypersurface orthogonal, i.e. there is a function ξ such that K_a is proportional to $\xi_{;a}$. One can then decompose the metric in the exterior region in the form $g_{ab} = f^{-1}K_aK_b + h_{ab}$ where $f \equiv K^aK_a$ and h_{ab} is the induced metric in the surfaces

$\{\xi = \text{constant}\}$ and represents the separation of the integral curves of K^a. The exterior region therefore admits an isometry which sends a point on a surface ξ to the point on the surface $-\xi$ on the same integral curve of $\mathbf{K}$. This isometry reverses the direction of time, and a space admitting such an isometry will be said to be *time symmetric*. Thus if the analytic extension of the exterior region contains a future event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$, it will also contain a past event horizon $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}})$. These event horizons may or may not intersect; the Schwarzschild solution and the Reissner–Nordström solution with $e^2 < m^2$ are examples where they do intersect, and the Reissner–Nordström solution with $e^2 = m^2$ is an example where they do not. The gradient of f is zero on the horizon in the latter case, but not in the former cases. The significance of this comes from the fact that on the future horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$, $K_{a;b}K^b = \frac{1}{2}f_{;a} = \beta K_a$, where $\beta \geqslant 0$ is constant along the null geodesic generators of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Let v be a future-directed affine parameter along such a generator. Then $\mathbf{K} = \alpha\, \partial/\partial v$ where α is a function along the generator which obeys $\mathrm{d}\alpha/\mathrm{d}v = \beta$. If $\beta \neq 0$ and the generator is geodesically complete in the past direction, α and the Killing vector $\mathbf{K}$ will be zero at some point. This point cannot lie in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$, and so will be a point of intersection of the future event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ and the past event horizon $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}})$ (Boyer (1969)). If $\beta = 0$, $\mathbf{K}$ will always be non-zero and there will be no such point where the horizon bifurcates.

Israel (1967) has shown that a static regular predictable space–time must be a Schwarzschild solution if:

(*a*) $T_{ab} = 0$;

(*b*) the magnitude $f \equiv K^a K_a$ of the Killing vector has non-zero gradient everywhere in $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$;

(*c*) the past event horizon $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}})$ intersects the future event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ in a compact two-surface $\mathscr{F}$.

(It follows from (*c*) and proposition 9.3.2 that $\mathscr{F}$ is connected and has the topology of a two-sphere. Israel did not give the conditions in this precise form, but these are equivalent.) Israel (1968) has further shown that the solution must be a Reissner–Nordström solution if the empty space condition (*a*) is replaced by the requirement that the energy–momentum tensor is that of an electromagnetic field. Müller zum Hagen, Robinson and Seifert (1973) have removed condition (*b*) in the vacuum case.

From these results we expect that if the final state of the solution outside the event horizon is static, then the metric in the exterior region will be that of a Schwarzschild solution.

We shall now consider the case where the final state of the exterior solution is stationary but not static. We would expect this to be the case when the object that collapsed was rotating initially.

Proposition 9.3.5

In an empty stationary regular predictable space which is not static, the Killing vector K^a is spacelike in part of the exterior region $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$.

The region of $\overline{J^+}(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ on which K^a is spacelike, is called the *ergosphere*. From proposition 9.3.4 it follows that there is no ergosphere if the solution is static. The significance of the ergosphere is that in it, it is impossible for a particle to move on an integral curve of the Killing vector K^a, i.e. to remain at rest as viewed from infinity. Since the ergosphere is outside the horizon it is still possible for such a particle to escape to infinity. An example of a stationary non-static regular predictable space with an ergosphere is the Kerr solution for $a^2 \leqslant m^2$ (§ 5.6).

Penrose (1969), Penrose and Floyd (1971) have pointed out that one can extract a certain amount of energy from a black hole with an ergosphere, by throwing a particle from infinity into the ergosphere. Since the particle moves on a geodesic, $E_0 \equiv -p_0{}^a K_a > 0$ is constant along its trajectory

$$((p_0{}^a K_a)_{;b} p_0{}^b = (p_0{}^a{}_{;b} p_0{}^b) K_a + p_0{}^a K_{a;b} p_0{}^b = 0,$$

as $p_0{}^a$ is a geodesic vector and K^a is a Killing vector), where $p_0{}^a = mv_0{}^a$ is the momentum vector of the particle, m is its rest-mass and $\mathbf{v}_0$ is the unit tangent to the particle world-line. The particle is then supposed to split into two particles with momentum vectors $p_1{}^a$ and $p_2{}^a$, where $p_0{}^a = p_1{}^a + p_2{}^a$. Since K^a is spacelike, it is possible to choose $p_1{}^a$ to be a future pointing timelike vector such that $E_1 \equiv -p_1{}^a K_a < 0$. Then $E_2 \equiv -p_2{}^a K_a$ will be greater than E_0. This means that the second particle can escape to infinity where it will have more energy than the original particle that was thrown in. One has thus extracted a certain amount of energy from the black hole.

The particle with negative energy cannot escape to infinity, but must remain in the region where K^a is spacelike. Suppose that the ergosphere did not intersect the event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Then the particle would have to remain in the exterior region. By repeating the process, one could continue to extract energy from the solution. As one did this, one would expect the solution to change gradually. However the ergosphere cannot shrink to zero, as there has to be somewhere for these negative energy particles to exist. It therefore appears that either one could extract an infinite amount of energy (which seems improbable), or that the ergosphere would eventually have to intersect the horizon. We shall show that in the latter case the solution would spontaneously become either axisymmetric or static without any further extraction of energy by the Penrose process. Either the possibility of the extraction of an infinite amount of energy or the occurrence of a spontaneous change would seem to indicate that the original state of the black hole was unstable. It therefore seems reasonable to assume that in any realistic black hole situation the ergosphere intersects the horizon.

Hajicek (1973) has shown that the stationary limit surface, which is the outer boundary of the ergosphere, will contain at least two integral null geodesic curves of K^a. If the gradient of f is non-zero on these curves, and if they are geodesically complete in the past, they will contain points where K^a is zero. However there can be no such points in the exterior region (see proposition 9.3.3), so the ergosphere must intersect the horizon in this case. However although it might be reasonable to assume that the integral curves of K^a were complete in the future, it does not seem reasonable to assume that they are complete in the past, since that would be to assume something about the past region of the solution which, as we said before, is not physically significant. In the static case one could show that the solution was time symmetric, but there is no *a priori* reason why a stationary non-static solution should be time symmetric. For this reason we shall rely on the energy extraction argument above rather than on Hajicek's results, to justify our assumption that the ergosphere intersects the horizon.

One can explain the significance of the ergosphere touching the horizon as follows. Let $\mathscr{Q}_1$ be one connected component of

$$\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$$

and let $\mathscr{G}_1$ be the quotient of $\mathscr{Q}_1$ by its generators. By propositions 9.3.1 and 9.3.2, this will be homeomorphic to a two-sphere. By proposition 9.3.1, the spatial separation of two neighbouring generators is constant along the generators, and so can be represented by an induced metric $\mathbf{h}$ on $\mathscr{G}_1$. The isometry θ_t moves generators into generators, and so acts as an isometry group of $(\mathscr{G}_1, \mathbf{h})$. If the ergosphere intersects the horizon, K^a will be spacelike somewhere on the horizon and the action of θ_t on $(\mathscr{G}_1, \mathbf{h})$ is non-trivial. Therefore it must correspond to a rotation of the sphere $\mathscr{G}_1$ around an axis, and the orbits of the group in $\mathscr{G}_1$ will be two points, corresponding to the poles, and a family of circles. A particle moving along one of the generators of the horizon would therefore appear to be moving relative to the frame defined by K^a which is stationary at infinity. One could therefore say that the horizon was *rotating* with respect to infinity.

The next result shows that a rotating black hole must be axisymmetric.

Proposition 9.3.6

Let $(\mathscr{M}, \mathbf{g})$ be a stationary non-static regular predictable space, in which the ergosphere intersects $\dot{J}^-(\mathscr{I}^+, \overline{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \overline{\mathscr{M}})$. Then there is a one-parameter cyclic isometry group $\tilde{\theta}_\phi$ $(0 \leqslant \phi \leqslant 2\pi)$ of $(\mathscr{M}, \mathbf{g})$ which commutes with θ_t, and whose orbits are spacelike near $\mathscr{I}^+$ and $\mathscr{I}^-$.

The method of proof of proposition 9.3.6 is to use the analyticity of the metric $\mathbf{g}$ to show that there is an isometry $\tilde{\theta}_\phi$ in a neighbourhood of the horizon. One then extends the isometry by analytic continuation. The method would therefore work even if the metric were not analytic in isolated regions away from the horizon, for example if there were a ring of matter or a frame of rods around the black hole. This leads to an apparent paradox. Consider a rotating star surrounded by a stationary square frame of rods. Suppose that the star collapsed to form a rotating black hole. If the black hole approached a stationary state, it would follow from proposition 9.3.6 that the metric $\mathbf{g}$ was axisymmetric except where it was non-analytic at the rods. However the gravitational effect of the rods would prevent the metric being axisymmetric. The resolution of the paradox seems to be that the black hole would not be in a stationary state while it was rotating. What would happen is that the gravitational effect of the rods would distort the black hole slightly. The back reaction on the frame would cause it to start rotating and so to radiate angular momentum. Eventually the rotation of both the black hole and the frame would be damped out and the solution would approach a static state. A static black hole need not be axisymmetric if the space outside it is not empty, i.e. if condition (*a*) of Israel's theorem is not satisfied.

The above discussion indicates that a realistic black hole will never be exactly stationary while it is rotating, as the universe will not be exactly axisymmetric about it. However in most circumstances, the rate of slowing down of the rotation of the black hole is extremely slow (Press (1972), Hartle and Hawking (1972*b*)). Thus it is a good approximation to neglect the small asymmetries produced by matter at a distance from the black hole, and to regard the rotating black hole as being in a stationary state. We shall therefore now consider the properties of a rotating axisymmetric black hole.

The following result of Papapetrou (1966), generalized by Carter (1969), shows that the Killing vectors K^a corresponding to the time translation θ_t and $\tilde{K}^a$ corresponding to the angular rotation $\tilde{\theta}_\phi$ are both orthogonal to families of two-surfaces.

Proposition 9.3.7

Let $(\mathscr{M}, \mathbf{g})$ be a space–time which admits a two-parameter abelian isometry group with Killing vectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. Let $\mathscr{V}$ be a connected open set of $\mathscr{M}$, and let $w_{ab} \equiv \xi_{1[a}\xi_{2b]}$. If

(*a*) $w_{ab}R^b{}_c\eta^{cdef}w_{ef} = 0$ on $\mathscr{V}$,

(*b*) $w_{ab} = 0$ at some point of $\mathscr{V}$,

then $w_{[ab;c}w_{d]e} = 0$ on $\mathscr{V}$.

Condition (*b*) is satisfied in a stationary axisymmetric space–time on the axis of axisymmetry, i.e. the set of points where $\tilde{K}^a = 0$. Condition (*a*) is satisfied in empty space, and when the energy–momentum tensor is that of a source-free electromagnetic field (Carter (1969)). By Frobenius' theorem (Schouten (1954)), the vanishing of $w_{[ab;c}w_{d]e}$ is, when $w_{ab} \neq 0$, the condition that there should exist locally a family of two-surfaces which are orthogonal to w_{ab}, i.e. to any linear combination of $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. In the case of a stationary axisymmetric space–time, this means that one can locally introduce coordinates (t, ϕ, x^1, x^2) such that $\mathbf{K} = \partial/\partial t$, $\tilde{\mathbf{K}} = \partial/\partial\phi$, and $K^a x^m{}_{;a} = 0 = \tilde{K}^a x^m{}_{;a}$ for $m = 1, 2$. The metric then locally admits the isometry $(t, \phi, x^1, x^2) \rightarrow (-t, -\phi, x^1, x^2)$, which reverses the direction of time, i.e. it is time-symmetric. Thus if the analytic extension of metric near infinity of an empty stationary regular predictable space–time contains a future event horizon, it will also contain a past event horizon.

In analogy with proposition 9.3.4, one has

Proposition 9.3.8 (*cf. Carter* (1971*b*))

Let $(\mathscr{M}, \mathbf{g})$ be a stationary axisymmetric regular predictable space-time in which $w_{[ab;c}w_{d]e} = 0$, where $w_{ab} \equiv K_{[a}\tilde{K}_{b]}$. Then at any point in the exterior region $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ off the axis $\tilde{\mathbf{K}} = 0$, $h \equiv w_{ab}w^{ab}$ is negative. On the horizons $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ and $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, h is zero but $w_{ab} \neq 0$ except on the axis.

This shows that at each point off the axis in the exterior region, there is some linear combination of the Killing vectors K^a and $\tilde{K}^a$ which is timelike. Outside the ergosphere, K^a itself is timelike, but between the stationary limit surface and the horizon one has to add a multiple of $\tilde{K}^a$ to obtain a timelike Killing vector. On the horizon there is no linear combination which is timelike, but there is a linear combination which is null, and is directed along the null generators of the horizon. Off the axis $\tilde{\mathbf{K}} = 0$, one can locally characterize the horizon as the set of points on which $h \equiv w_{ab}w^{ab} = 0$.

We now come to the theorem of Carter (1971*b*) which indicates that the Kerr solutions are probably the only empty stationary black holes. He considered stationary regular predictable spaces which satisfy:

(*a*) $T_{ab} = 0$,

(*b*) they are axisymmetric,

(*c*) the past event horizon $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}})$ intersects the future event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ in a compact connected two-surface $\mathscr{F}_1$.

(By proposition 9.3.2, this will be a two-sphere.) He showed that such solutions fall into disjoint families, each of which depends only on two parameters. The two parameters can be taken to be the mass m and angular momentum L as measured from infinity. One such family is known, namely the Kerr solutions for $m \geqslant 0$, $a^2 \leqslant m^2$, where $a = L/m$. (The Kerr solutions with $a^2 > m^2$ contain naked singularities and so are not regular predictable spaces.) It seems unlikely that there are any other disjoint families. It has been conjectured, therefore, that the solution outside an uncharged collapsed object will settle down to a Kerr solution with $a^2 \leqslant m^2$. This conjecture is supported by analyses of linear perturbations from a spherical collapse by Regge and Wheeler (1957), Doroshkevich, Zel'dovich and Novikov (1966), Vishveshwara (1970), and Price (1972).

Assuming the validity of this Carter–Israel conjecture, one would expect the area of the two-surface $\partial\mathscr{B}(\tau)$ in the event horizon to approach the area of a two-surface in the event horizon $r = r_+$ of a

Kerr solution with the same mass and angular momentum, as measured at $\mathscr{Q}(\tau)$ on $\mathscr{I}^+$. This area is $8\pi m(m+(m^2-a^2)^{\frac{1}{2}})$, where m is the mass of the Kerr solution and ma is the angular momentum. (If the collapsing body has a net electrical charge e one would expect the solution to settle down to a charged Kerr solution. The area of a two-surface in the event horizon of such a solution is

$$4\pi(2m^2-e^2+2m(m^2-a^2-e^2)^{\frac{1}{2}}).$$

Using this expression one can generalize our results to charged black holes.) Consider a collapse situation which by a surface $\mathscr{S}(\tau_1)$ has settled down to a Kerr solution with mass m_1 and angular momentum $m_1 a_1$. Suppose one now lets the black hole interact with particles or radiation for a finite time. The solution will eventually settle down, by a surface $\mathscr{S}(\tau_2)$, to a different Kerr solution with parameters m_2, a_2. From the discussion of § 9.2, the area of $\partial\mathscr{B}(\tau_2)$ must be greater than or equal to the area of $\partial\mathscr{B}(\tau_1)$. In fact it must be strictly greater than, since $\hat{\theta}$ can be zero only if no matter or radiation crosses the horizon. This then implies that

$$m_2(m_2+(m_2{}^2-a_2{}^2)^{\frac{1}{2}}) > m_1(m_1+(m_1{}^2-a_1{}^2)^{\frac{1}{2}}). \tag{9.4}$$

If $a_1 \neq 0$, then the inequality (9.4) allows m_2 to be less than m_1. Since there is a conservation law for total energy and momentum in an asymptotically flat space–time (Penrose (1963)), this would mean that one had extracted a certain amount of energy from the black hole. One way of doing this would be to construct a square frame of rods about the black hole and employ the torque exerted by the rotating black hole on the frame to do work. Alternatively, one could use Penrose's process of throwing a particle into the ergosphere, where it divides into two particles, one of which escapes to infinity with greater energy than the original particle. The other particle will fall through the event horizon and reduce the angular momentum of the solution. One can thus regard the process as extracting rotational energy from the black hole. Christodoulou (1970) has shown that one can achieve a result arbitrarily near the limit set by the inequality (9.4). In fact the maximum energy extraction occurs when $a_2 = 0$; then the available energy (m_1-m_2) is less than

$$m_1\left\{1-\frac{1}{\sqrt{2}}\left(1+\left(1-\frac{a_1{}^2}{m_1{}^2}\right)^{\frac{1}{2}}\right)^{\frac{1}{2}}\right\}.$$

Consider now a situation in which two stars a long way apart collapse to produce black holes. There is thus some τ' such that $\partial\mathscr{B}(\tau')$ consists of two separate two-spheres $\partial\mathscr{B}_1(\tau')$ and $\partial\mathscr{B}_2(\tau')$. Since these are a long way apart, one can neglect their interaction and assume that the solutions near each are close to Kerr solutions with parameters m_1, a_1 and m_2, a_2 respectively. Thus the areas of $\partial\mathscr{B}_1(\tau')$ and $\partial\mathscr{B}_2(\tau')$ will be approximately $8\pi m_1(m_1+(m_1{}^2-a_1{}^2)^{\frac{1}{2}})$ and $8\pi m_2(m_2+(m_2{}^2-a_2{}^2)^{\frac{1}{2}})$ respectively. Now suppose that these black holes fall towards each other, collide and coalesce. In such a collision a certain amount of gravitational radiation will be emitted. The system will eventually settle down by a surface $\mathscr{S}(\tau'')$ to resemble a single Kerr solution with parameters m_3, a_3. By the same argument as previously, the area of $\partial\mathscr{B}(\tau'')$ must be greater than the total area of $\partial\mathscr{B}(\tau')$, which is the sum of the areas $\partial\mathscr{B}_1(\tau')$ and $\partial\mathscr{B}_2(\tau')$. Thus

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

By the conservation law for asymptotically flat spaces, the amount of energy carried away to infinity by gravitational radiation is

$$m_1+m_2-m_3.$$

This is limited by the above inequality. The efficiency

$$\epsilon \equiv (m_1+m_2-m_3)\,(m_1+m_2)^{-1}$$

of conversion of mass to gravitational radiation is always less than $\frac{1}{2}$. If $a_1 = a_2 = 0$, then $\epsilon < 1-1/\sqrt{2}$. It should be stressed that these are upper limits; the actual efficiency might be much less, although the mere existence of a limit might suggest that one could attain an appreciable fraction of it.

We have shown that the fraction of mass which can be converted to gravitational radiation in the coalescence of one pair of black holes is limited. However if there were initially a large number of black holes, these could combine in pairs and then the resulting holes could combine, and so on. On dimensional grounds one would expect the efficiency to be the same at each stage. Thus one would eventually convert a very large fraction of the original mass to gravitational radiation. (This argument was suggested by C. W. Misner and M. J. Rees.) At each stage, the energy emitted in gravitational radiation would be larger. This might be able to explain Weber's recent observations of short bursts of gravitational radiation.

We now give the proofs of the propositions we have stated in this section. For convenience, we repeat the statements of the propositions.

Proposition 9.3.1

Let $(\mathscr{M}, \mathbf{g})$ be a stationary, regular predictable space–time. Then the generators of the future event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ have no past endpoints in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Let $Y_1{}^a$ be the future-directed tangent vector to these generators; then in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$, $Y_1{}^a$ has zero shear $\hat{\sigma}$ and expansion $\hat{\theta}$, and satisfies

$$R_{ab}Y_1{}^aY_1{}^b = 0 = Y_{1[e}C_{a]bc[d}Y_{1f]}Y_1{}^bY_1{}^c.$$

Let $\mathscr{C}$ be a spacelike two-sphere on $\mathscr{I}^-$. Then one can cover $\mathscr{I}^-$ by a family of two-spheres $\mathscr{C}(t)$ obtained by moving $\mathscr{C}$ up and down the generators of $\mathscr{I}^-$ under the action of θ_t, i.e. $\mathscr{C}(t) = \theta_t(\mathscr{C})$. We now define the function x at the point $p \in J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ to be the greatest value of t such that $p \in J^+(\mathscr{C}(t), \bar{\mathscr{M}})$. Let $\mathscr{U}$ be a neighbourhood of $\mathscr{I}^+$ and $\mathscr{I}^-$ which is isometric to a corresponding neighbourhood of an asymptotically simple space–time. Then x will be continuous and have some lower bound x' on $\mathscr{S} \cap \mathscr{U}$. From this it follows that x will be continuous in the region of $\overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ where it is greater than x'. Let $p \in J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Then under the isometry θ_t, p will be moved into the region of $\overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, where $x > x'$. However

$$x|_{\theta_t(p)} = x|_p + t.$$

Therefore x will be continuous at p.

Let $\tau_0 > 0$ be such that $\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ is contained in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Let λ be a generator of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ which intersects $\mathscr{S}(\tau_0)$. Suppose there were some finite upper bound x_0 to x on λ. Since the space is weakly asymptotically simple, $x \to \infty$ as one approaches $\mathscr{Q}(\tau_0)$ on $\mathscr{S}(\tau_0)$. Thus there will be some lower bound x_1 of x on

$$\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}).$$

Under the action of the group θ_t, λ is moved into another generator $\theta_t(\lambda)$. As the generators of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ have no future endpoints, the past extension of $\theta_t(\lambda)$ will still intersect $\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$. This leads to a contradiction, since the upper bound of x on $\theta_t(\lambda)$ would be less than x_1 if $t < x_1 - x_0$.

Let x_2 be the upper bound of x on $\mathscr{S}(\tau_0) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Then every generator λ of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ which intersects $\mathscr{S}(\tau_0)$ will intersect $\mathscr{F}(t) \equiv \dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}}) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ for $t \geqslant x_2$. Every generator of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ which intersects $\mathscr{F}(t')$ will intersect $\theta_t(\mathscr{S}(\tau_0))$ for $t \geqslant t' - x_1$. But $\theta_t(\mathscr{S}(\tau_0)) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) = \theta_t(\mathscr{S}(\tau_0) \cap \dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}))$ is compact. Thus $\mathscr{F}(t)$ is compact.

Now consider how the area of $\mathscr{F}(t)$ varies as t increases. Since $\hat{\theta} \geqslant 0$ the area cannot decrease. If $\hat{\theta}$ were > 0 on an open set, the area would increase. Also if the generators of the horizon had past endpoints on $\mathscr{F}(t)$ the area would increase. However as $\mathscr{F}(t)$ is moving under the isometry θ_t, the area must remain the same. Therefore $\hat{\theta} = 0$, and there are no past endpoints on the region of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ for which $x \geqslant x_2$. However since each point of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ can be moved by the isometry θ_t to where $x > x_2$, this result applies to the whole of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. From the propagation equations (4.35) and (4.36) one then finds $\hat{\sigma}_{mn} = 0$, $R_{ab}Y_1{}^aY_1{}^b = 0$ and $Y_{1[e}C_{a]bc[d}Y_{1f]}Y_1{}^bY_1{}^c = 0$, where $Y_1{}^a$ is the future-directed tangent vector to the null geodesic generators of the horizon. □

Proposition 9.3.2

Each connected component in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ of the horizon $\partial\mathscr{B}(\tau)$ in a stationary, regular predictable space is homeomorphic to a two-sphere.

Consider how the expansion of the outgoing null geodesics orthogonal to $\partial\mathscr{B}(\tau)$ behaves if one deforms $\partial\mathscr{B}(\tau)$ slightly outwards into $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Let $Y_2{}^a$ be the other future-directed null vector orthogonal to $\partial\mathscr{B}(\tau)$, normalized so that $Y_1{}^aY_{2a} = -1$. This leaves the freedom $\mathbf{Y}_1 \to \mathbf{Y}_1' = e^y\mathbf{Y}_1$, $\mathbf{Y}_2 \to \mathbf{Y}_2' = e^{-y}\mathbf{Y}_2$. The induced metric on the spacelike two-surface $\partial\mathscr{B}(\tau)$ is $\hat{h}_{ab} = g_{ab} + Y_{1a}Y_{2b} + Y_{2a}Y_{1b}$. Define a family of surfaces $\mathscr{F}(\tau, w)$ by moving each point of $\partial\mathscr{B}(\tau)$ a parameter distance w along the null geodesic curve with tangent vector $Y_2{}^a$. The vectors $Y_1{}^a$ will be orthogonal to $\mathscr{F}(\tau, w)$ if they propagate according to

$$\hat{h}_{ab}Y_1{}^b{}_{;c}Y_2{}^c = -\hat{h}_a{}^bY_{2c;b}Y_1{}^c \quad \text{and} \quad Y_1{}^aY_{2a} = -1.$$

Then

$$(Y_1{}^a{}_{;b}\hat{h}_a{}^c\hat{h}^b{}_d)_{;g}Y_2{}^g\hat{h}_c{}^s\hat{h}^d{}_t = \hat{h}^{sa}p_{a;b}\hat{h}^b{}_t + p^sp_t - \hat{h}^s{}_aY_1{}^a{}_{;g}\hat{h}^{ge}Y_{2e;b}\hat{h}^b{}_t + R^a{}_{ceb}Y_2{}^eY_1{}^ch_a{}^sh^b{}_t, \quad (9.5)$$

where $p^a \equiv -\hat{h}^{ba}Y_{2c;b}Y_1{}^c$. Contracting with $\hat{h}_s{}^t$, one obtains

$$\frac{d\hat{\theta}}{dw} = (Y_1{}^a{}_{;b}\hat{h}^b{}_a)_{;c}Y_2{}^c$$
$$= p_{b;d}\hat{h}^{bd} - R_{ac}Y_1{}^aY_2{}^c + R_{adcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b + p_ap^a - Y_1{}^a{}_{;c}\hat{h}^c{}_dY_2{}^d{}_{;b}\hat{h}^b{}_a.$$

On the horizon, $Y_1{}^a{}_{;c}\hat{h}^{cd}\hat{h}^b{}_a$ is zero, as the shear and divergence of the horizon are zero. Under a rescaling transformation $\mathbf{Y}_1' = e^y\mathbf{Y}_1$,

$\mathbf{Y}_2' = e^{-y}\mathbf{Y}_2$, the vector p^a changes to $p'^a = p^a + \hat{h}^{ab}y_{;b}$, and so $\mathrm{d}\hat{\theta}/\mathrm{d}w|_{w=0}$ changes to

$$\left.\frac{\mathrm{d}\hat{\theta}'}{\mathrm{d}w'}\right|_{w=0} = p_{b;d}\hat{h}^{bd} + y_{;bd}\hat{h}^{bd} - R_{ac}Y_1{}^aY_2{}^c + R_{adcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b + p'^a p'_a. \quad (9.6)$$

The term $y_{;bd}\hat{h}^{bd}$ is the Laplacian of y in the two-surface $\partial\mathscr{B}(\tau)$. By a theorem of Hodge (1952), one can choose y so that the sum of the first four terms on the right of (9.6) is a constant on $\partial\mathscr{B}(\tau)$. The sign of this constant will be determined by that of the integral of

$$(-R_{ac}Y_1{}^aY_2{}^c + R_{adcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b)$$

over $\partial\mathscr{B}(\tau)$ ($p_{b;d}\hat{h}^{bd}$, being a divergence, has zero integral). This integral can be evaluated using the Gauss–Codacci equations for the scalar curvature $\hat{R}$ of the two-surface with metric $\hat{h}$:

$$\hat{R} = R_{ijkl}\hat{h}^{ik}\hat{h}^{jl} = R - 2R_{ijkl}Y_1{}^iY_2{}^jY_1{}^kY_2{}^l + 4R_{ij}Y_1{}^iY_2{}^j,$$

since $\hat{\theta} = \hat{\sigma} = 0$ on $\partial\mathscr{B}(\tau)$. By the Gauss–Bonnet theorem (Kobayashi and Nomizu (1969))

$$\int_{\partial\mathscr{B}(\tau)} \hat{R}\,\mathrm{d}\hat{S} = 2\pi\chi,$$

where $\mathrm{d}\hat{S}$ is the surface area element of $\partial\mathscr{B}(\tau)$ and χ is the Euler number of $\partial\mathscr{B}(\tau)$. Thus

$$\int_{\partial\mathscr{B}(\tau)} (-R_{ab}Y_1{}^aY_2{}^b + R_{adcb}Y_1{}^dY_2{}^cY_2{}^aY_1{}^b)\,\mathrm{d}\hat{S} = -\pi\chi + \int_{\partial\mathscr{B}(\tau)} (\tfrac{1}{2}R + R_{ab}Y_1{}^aY_2{}^b)\,\mathrm{d}\hat{S}. \quad (9.7)$$

By the Einstein equations,

$$\tfrac{1}{2}R + R_{ab}Y_1{}^aY_2{}^b = 8\pi T_{ab}Y_1{}^aY_2{}^b,$$

which is $\geqslant 0$ by the dominant energy condition. The Euler number χ is $+2$ for the sphere, zero for the torus, and negative for any other compact orientable two-surface ($\partial\mathscr{B}(\tau)$ has to be orientable as it is a boundary). Hence the right-hand side of (9.7) can be negative only if $\partial\mathscr{B}(\tau)$ is a sphere.

Suppose that the right-hand side of (9.7) was positive. Then one could choose y so that $\mathrm{d}\hat{\theta}'/\mathrm{d}w'|_{w=0}$ was positive everywhere on $\partial\mathscr{B}(\tau)$. For small negative values of w' one would obtain a two-surface in $J^-(\mathscr{I}^+, \overline{\mathscr{M}})$ such that the outgoing null geodesics orthogonal to the surface were converging. This would contradict proposition 9.2.8.

Suppose now that χ was zero and that $T_{ab}Y_1{}^aY_2{}^b$ was zero on $\partial\mathscr{B}(\tau)$. Then one could choose y so that the sum of the first four terms on the right of (9.6) was zero on $\partial\mathscr{B}(\tau)$. Then

$$p'^a{}_{;b}\hat{h}^b{}_a + R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d = 0$$

on $\partial\mathscr{B}(\tau)$. If $R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d$ was non-zero somewhere on $\partial\mathscr{B}(\tau)$, then the term $p'^ap'_a$ in (9.6) would be non-zero somewhere and one could change y slightly so as to make $\mathrm{d}\hat{\theta}'/\mathrm{d}w'|_{w=0}$ positive everywhere. This would again lead to a contradiction.

Now suppose that $R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d$ and p'^a were zero everywhere on $\partial\mathscr{B}(\tau)$. One could move the two-surface $\partial\mathscr{B}(\tau)$ back along $Y_2{}^a$, choosing the rescaling parameter y at each stage so that

$$p'^a{}_{;b}\hat{h}^b{}_a + R_{abcd}Y_1{}^aY_2{}^bY_1{}^cY_2{}^d - \tfrac{1}{2}R - 2R_{ab}Y_1{}^aY_2{}^b = p'^a{}_{;b}\hat{h}^b{}_a - \tfrac{1}{2}\hat{R} = 0.$$

If $T_{ab}Y_1{}^aY_2{}^b$ or p'^a were non-zero for $w' < 0$ then one could adjust y to obtain a two-surface in $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ with $\hat{\theta} < 0$. This would contradict proposition 9.2.8. On the other hand if $T_{ab}Y_1{}^aY_2{}^b$ and p'^a were zero everywhere for $w' < 0$, one would obtain a two-surface in $J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ with $\hat{\theta} = 0$ which again contradicts proposition 9.2.8.

One avoids a contradiction only if $\chi = 2$, i.e. if $\partial\mathscr{B}(\tau)$ is a two-sphere. □

Proposition 9.3.3

Let $(\mathscr{M}, \mathbf{g})$ be a stationary regular predictable space–time. Then the Killing vector K^a is non-zero in $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, which is simply connected. Let τ_0 be such that $\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ is contained in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. If $\partial\mathscr{B}(\tau_0)$ has only one connected component, then $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M}$ is homeomorphic to $[0, 1) \times S^2 \times R^1$.

The function x defined in proposition 9.3.1 is continuous on $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, and has the property that $x|_{\theta_t(p)} = x|_p + t$. This shows that $\mathbf{K}$ cannot be zero in $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$. The integral curves of $\mathbf{K}$ establish a homeomorphism between two of the surfaces

$$\dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}}) \cap \bar{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M} \quad (-\infty < t < \infty).$$

The region $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M}$ is covered by these surfaces, and so is homeomorphic to $R^1 \times \dot{J}^+(\mathscr{C}(t'), \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M}$ for any t'. Choose t to be large enough that $\dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}})$ intersects

$\mathscr{S}(\tau_0)$ in the neighbourhood $\mathscr{U}$ of $\mathscr{I}^+$ which is isometric to a similar neighbourhood in an asymptotically simple space. The integral curves of $\mathbf{K}$ establish a homeomorphism between

$$\dot{J}^+(\mathscr{C}(t), \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M} \quad \text{and} \quad \mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}).$$

By property (α) and proposition 9.3.2, this is simply connected. If further $\partial\mathscr{B}(\tau)$ has only one connected component, then

$$\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$$

has the topology $[0, 1) \times S^2$. Thus $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}) \cap \mathscr{M}$ has the topology $[0, 1) \times S^2 \times R^1$. □

Proposition 9.3.4

In a static regular predictable space–time, the Killing vector $\mathbf{K}$ is timelike in the exterior region $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ and is non-zero and directed along the null generators of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ on

$$\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}}).$$

The event horizon $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ is mapped into itself by the isometry θ_t. Thus on $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$, $\mathbf{K}$ must be null or spacelike. Let τ_0 be such that $\mathscr{S}(\tau_0) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ is contained in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Then $f \equiv K^a K_a$ must be zero on some closed set $\mathscr{N}$ in

$$J^+(\mathscr{S}(\tau_0)) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}}).$$

From the fact that K^a is a Killing vector and curl $\mathbf{K} = 0$, it follows that

$$fK_{a;b} = K_{[a}f_{;b]}. \tag{9.8}$$

By proposition 9.3.3, K^a is non-zero on the simply connected set $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$. By Frobenius' theorem, it follows from the condition curl $\mathbf{K} = 0$, that there is a function ξ on this region such that $K_a = -\alpha\xi_{;a}$, where α is some positive function.

Let p be a point of $\mathscr{N}$ and let $\lambda(v)$ be a curve through p lying in the surface of constant ξ through p. Then by (9.8),

$$\tfrac{1}{2}K^a \frac{\mathrm{d}}{\mathrm{d}v} \log f = \frac{\mathrm{D}}{\partial v} K^a.$$

If $\lambda(v)$ left $\mathscr{N}$, the left-hand side of this equation would be unbounded. However the right-hand side is continuous; therefore $\lambda(v)$ must lie in $\mathscr{N}$, so $\mathscr{N}$ must contain the surface $\xi = \xi|_p$. However f cannot be zero on an open neighbourhood of p, since it would then be zero everywhere. Thus the connected component of $\mathscr{N}$ through p is the three-surface $\xi = \xi|_p$. Suppose $p \in J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}})$. Then there would be a future-directed timelike curve $\gamma(u)$ from $\mathscr{I}^-$ through p to $\mathscr{I}^+$. On $\xi = \xi|_p$, K^a would be future-directed. Thus $(\partial/\partial u)_\gamma \xi > 0$ when $\xi = \xi|_p$. This leads to a contradiction as $\xi = \xi|_p$ cannot intersect $\mathscr{I}^+$ or $\mathscr{I}^-$ since K^a is timelike near infinity. Thus near $\mathscr{I}^+$ and $\mathscr{I}^-$, either ξ is greater than $\xi|_p$ or less than $\xi|_p$. □

Proposition 9.3.5

In an empty regular predictable space–time which is not static, the Killing vector K^a is spacelike in part of the exterior region

$$J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \overline{\mathscr{M}}).$$

The function x introduced in proposition 9.3.1 is continuous on $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \overline{\mathscr{M}})$, and is such that along each integral curve of K^a, $\partial x/\partial t = 1$. One can approximate the surface $x = 0$ in $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \overline{\mathscr{M}})$ by a smooth surface $\mathscr{K}$ which is nowhere tangent to K^a. One can then define a smooth function $\bar{x}$ on $J^+(\mathscr{I}^-, \overline{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \overline{\mathscr{M}})$ by specifying that $\bar{x} = 0$ on $\mathscr{K}$ and $\bar{x}_{;a} K^a = 1$. One can express the gradient of the Killing vector as

$$fK_{a;b} = \eta_{abcd} K^c \omega^d + K_{[a} f_{;b]},$$

where $f \equiv K^a K_a$ is the magnitude of the Killing vector, and

$$\omega^a \equiv \tfrac{1}{2} \eta^{abcd} K_b K_{c,d}.$$

The second derivatives of $\mathbf{K}$ satisfy

$$2K_{a;[bc]} = R_{dabc} K^d.$$

However $K_{a;bc} = K_{[a;b]c}$. Therefore

$$K_{a;bc} = R_{dcba} K^d$$

which implies

$$K^{a;b}{}_b = -R^a{}_d K^d. \qquad (9.9)$$

The vector $q_a \equiv f^{-1} K_a - \bar{x}_{;a}$ is orthogonal to K^a. Multiplying (9.9) by q_a and integrating over the region $\mathscr{L}$ of $J^-(\mathscr{I}^+, \overline{\mathscr{M}})$ bounded by the surfaces $\mathscr{N}_1$ and $\mathscr{N}_2$ defined by $\bar{x} = x_2 + 1$ and $x = x_2 + 2$, where x_2 is as in proposition 9.3.1, one finds

$$\int_{\mathscr{L}} R_{ab} K^a q^b \, \mathrm{d}v = -\int_{\mathscr{L}} (K^{a;b} q_a)_{;b} \, \mathrm{d}v + \int_{\mathscr{L}} K_{a;b} q^{a;b} \, \mathrm{d}v$$

$$= -\int_{\partial\mathscr{L}} K^{a;b} q_a \, \mathrm{d}\sigma_b - 2\int_{\mathscr{L}} f^{-2} \omega^a \omega_a \, \mathrm{d}v. \qquad (9.10)$$

The boundary $\partial\mathscr{L}$ of $\mathscr{L}$ consists of the surfaces $\partial\mathscr{L}_1 \equiv \mathscr{N}_1 \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, $\partial\mathscr{L}_2 \equiv \mathscr{N}_2 \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$, the portion $\partial\mathscr{L}_3$ of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$ between $\mathscr{N}_1$ and $\mathscr{N}_2$, and the portion $\partial\mathscr{L}_4$ of $\mathscr{I}^-$ between $\mathscr{N}_1$ and $\mathscr{N}_2$. The surface integral over $\partial\mathscr{L}_1$ is minus that over $\partial\mathscr{L}_2$, since these surfaces are carried into each other by the isometry θ_1.

Near $\mathscr{I}^-$, $f = -1 + (2m/r) + O(r^{-2})$ and $\omega^a\omega_a = O(r^{-6})$, where r is some suitable radial coordinate. Thus the surface integral over $\partial\mathscr{L}_4$ at $\mathscr{I}^-$ vanishes. Suppose now that K^a were timelike everywhere in $\mathscr{L}$, becoming null on the horizon. Then ω^a, being orthogonal to $\mathbf{K}$, would be spacelike everywhere in $\mathscr{L}$. Therefore if $\boldsymbol{\omega}$ is non-zero, i.e. the solution is non-static, the last term on the right of (9.10) will be negative. This leads to a contradiction if the space is empty and if the integral over $\partial\mathscr{L}_3$ is zero.

To evaluate this integral, one has to apply a limiting procedure. Let z be a function on the surface $\mathscr{N}_1$ which is zero on the horizon but such that the gradient of z in $\mathscr{N}_1$ is not zero on the horizon. The function z can be defined on $\bar{\mathscr{L}}$ by the condition $z_{;a}K^a = 0$. One can express the gradient of z as

$$z_{;a} = \bar{x}_{;b}z^{;b}(K_a + fR_a),$$

where R^a is a vector field tangent to the surfaces $\{\bar{x} = \text{constant}\}$ and normalized so that $R^aK_a = -1$. One now takes $\int K^{a;b}q_a\,\mathrm{d}\sigma_b$ over the surface $\{z = \text{constant}\}$ between $\mathscr{N}_1$ and $\mathscr{N}_2$. Then $\mathrm{d}\sigma_b = \mathrm{d}\sigma z_{;b}$, where $\mathrm{d}\sigma$ is some continuous measure. Thus

$$\int K^{a;b}q_a\,\mathrm{d}\sigma_b = \int (\tfrac{1}{2}\bar{x}_{;a}(f)^{;a} - \bar{x}_{;a}K^a{}_{;b}R^bf + \tfrac{1}{2}f_{;b}R^b)\,\bar{x}_{;b}z^{;b}\,\mathrm{d}\sigma.$$

Since the horizon was the surface $f = 0$ and since K^a was directed along the null generators of the horizon, $f_{;a}$ is proportional to K^a on the horizon. Therefore

$$\int_{\partial\mathscr{L}_3} K^{a;b}q_a\,\mathrm{d}\sigma_b = 0.$$

This gives a contradiction which shows that K^a must be spacelike somewhere in $\bar{\mathscr{L}}$ if the space is empty. □

Proposition 9.3.6

Let $(\mathscr{M}, \mathbf{g})$ be a stationary non-static regular predictable space–time in which the ergosphere intersects $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$. Then there is a one-parameter cyclic isometry group $\hat{\theta}_\phi$ $(0 \leqslant \phi \leqslant 2\pi)$ of $(\mathscr{M}, \mathbf{g})$ which commutes with θ_t, and whose orbits are spacelike near $\mathscr{I}^+$ and $\mathscr{I}^-$.

Let $\mathscr{Q}_1$ be one connected component of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$, and let $\mathscr{G}_1$ be the quotient of $\mathscr{Q}_1$ by its generators. Then the orbits of the isometry θ_t in the horizon $\mathscr{Q}_1$ will be spirals which repeatedly intersect the same generators. Let $t_1 > 0$ be such that θ_{t_1} is one rotation of $\mathscr{G}_1$. Then if $p \in \mathscr{Q}_1$, $\theta_{t_1}(p)$ will lie on the same generator of $\mathscr{Q}_1$. It will lie to the future of p, since

$$x|_{\theta_{t_1}(p)} = x|_p + t_1.$$

One can now choose the future-directed null vector $\mathbf{Y}_1$ to be directed along the generators, and scaled so that

(i) $Y_{1a;b} Y_1{}^b = 2\epsilon Y_{1a}$, where $\epsilon_{;a} Y_1{}^a = 0$,

(ii) if v is a parameter along the generators such that $\mathbf{Y}_1 = \partial/\partial v$, then

$$v|_{\theta_{t_1}(p)} = v|_p + t_1.$$

The vector field $\mathbf{Y}_1$ defined in this way is invariant under the isometry θ_t, i.e. $L_{\mathbf{K}} \mathbf{Y}_1 = 0$. One can now define a spacelike vector field $\mathbf{Y}_3$ in $\mathscr{Q}_1$ by $\mathbf{Y}_3 \equiv \mathbf{K} - \mathbf{Y}_1$; then $L_{\mathbf{K}} \mathbf{Y}_3 = 0$ and $L_{\mathbf{Y}_1} \mathbf{Y}_3 = 0$ (note that $\mathbf{Y}_3$ is not a unit vector, and in fact it will vanish on the generators γ_1 and γ_2 corresponding to the poles of $\mathscr{G}_1$). The integral curves of $\mathbf{Y}_3$ in $\mathscr{Q}_1$ will be circles which degenerate to points on γ_1 and γ_2.

Let μ be a curve in $\mathscr{Q}_1$ from γ_1 to γ_2 orthogonal to $\mathbf{Y}_1$ and $\mathbf{Y}_3$, and such that the orbits of $\mathbf{Y}_3$ which intersect μ form a smooth spacelike two-surface $\mathscr{P}$ in $\mathscr{Q}_1$. Let $\mathscr{P}(v)$ be the family of spacelike two-surfaces in $\mathscr{Q}_1$ obtained by moving each point of $\mathscr{P}$ a parameter distance v up the generators of $\mathscr{Q}_1$. $\mathscr{P}(v)$ is also equal to $\theta_v(\mathscr{P})$. Let $\mathbf{Y}_2$ be the other null vector orthogonal to $\mathscr{P}(v)$, normalized so that $Y_1{}^a Y_{2a} = -1$ (see figure 61); then $L_{\mathbf{K}} \mathbf{Y}_2 = 0$.

Let $\mathbf{Y}_4$ be a spacelike vector on μ, tangent to μ. Then one can define $\mathbf{Y}_4$ on $\mathscr{Q}_1$ by dragging it along by $\mathbf{K}$ and $\mathbf{Y}_1$, i.e. $L_{\mathbf{K}} \mathbf{Y}_4 = 0 = L_{\mathbf{Y}_1} \mathbf{Y}_4$. (These are compatible because $L_{\mathbf{K}} \mathbf{Y}_1 = 0$.) $\mathbf{Y}_4$ will be orthogonal to $\mathbf{Y}_1$ on $\mathscr{Q}_1$ because $L_{\mathbf{K}}(Y_4{}^a g_{ab} Y_1{}^b) = 0$, and

$$(Y_4{}^a Y_{1a})_{;b} Y_1{}^b = Y_1{}^a{}_{;b} Y_4{}^b Y_{1a} + Y_1{}^a{}_{;b} Y_{4a} Y_1{}^b.$$

The first term is zero because $\mathbf{Y}_1$ is null and the second term equals $2\epsilon Y_{1a} Y_4{}^a$. Thus $Y_{1a} Y_4{}^a$, being zero initially, remains zero. $\mathbf{Y}_4$ will be orthogonal to $\mathbf{Y}_2$ on $\mathscr{Q}_1$ because it lies in the surface $\mathscr{P}(v)$, and $\mathbf{Y}_2$ is normal to the surface. It will also be orthogonal to $\mathbf{Y}_3$ on $\mathscr{Q}_1$ because $L_{\mathbf{K}}(Y_3{}^a g_{ab} Y_4{}^b) = 0$, and

$$(Y_3{}^a Y_{4a})_{;b} Y_1{}^b = Y_1{}^a{}_{;b} Y_3{}^b Y_{4a} + Y_1{}^a{}_{;b} Y_4{}^b Y_{3a} = 0$$

since $Y_{1a;b} \hat{h}^{ac} \hat{h}^{bd} = 0$.

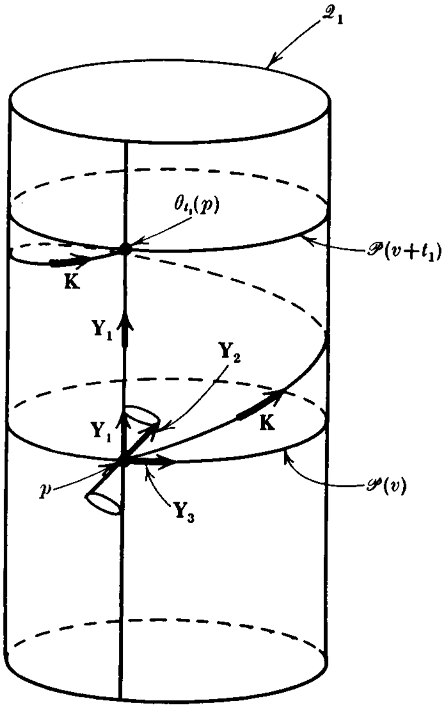

FIGURE 61. The isometry θ_{t_1} moves the point p and the surface $\mathscr{P}(v)$ into the point $\theta_{t_1}(p)$ and the surface $\mathscr{P}(v+t_1)$ in the horizon $\mathscr{Q}_1$. $\mathbf{Y}_1$ is tangent to a null geodesic generator of $\mathscr{Q}_1$, $\mathbf{Y}_2$ is a null vector orthogonal to $\mathscr{P}(v)$, and $\mathbf{Y}_3$ lies in $\mathscr{P}(v)$. **K** is the Killing vector field on $\mathscr{Q}_1$ which generates the isometry group θ_t.

In a neighbourhood of $\mathscr{Q}_1$, there will be a unique null geodesic λ orthogonal to a surface $\mathscr{P}(v)$ through a given point r. One can then define coordinates (v, w, θ, ϕ) for the point r, where w is the affine distance (as measured by $\mathbf{Y}_2$) along μ, and (v, θ, ϕ) have their values at $\mu \cap \mathscr{Q}_1$, where θ and ϕ are spherical polar coordinates for the generators of $\mathscr{Q}_1$ such that $Y_3{}^a\theta_{,a} = 0$, $Y_4{}^a\phi_{,a} = 0$. (In other words, we choose $\mathbf{Y}_3 = (2\pi/t_1)\,\partial/\partial\phi$ and $\mathbf{Y}_4 = \partial/\partial\theta$ on $\mathscr{Q}_1$.) We shall take the basis $\{\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \mathbf{Y}_4\}$ to be parallelly propagated along the null geodesics with tangent vector $\mathbf{Y}_2$. Then $\mathbf{Y}_2 = \partial/\partial w$. We define the vector $\hat{\mathbf{K}}$ to be $\partial/\partial v$. This means that the Lie derivative of $\hat{\mathbf{K}}$ by $\mathbf{Y}_2$ is zero. We define the vector Z^a to be

$$Z^a = \frac{1}{\sqrt{2}}\left\{\frac{Y_3{}^a}{(Y_3{}^b Y_{3b})^{\frac{1}{2}}} + \mathrm{i}\frac{Y_4{}^a}{(Y_4{}^b Y_{4b})^{\frac{1}{2}}}\right\}.$$

Then $$Z^a Z_a = 0, \quad Z^a \bar{Z}_a = 1, \quad \bar{Z}^a \bar{Z}_a = 0,$$

where $^-$ denotes the complex conjugate.

One can define on $\mathscr{Q}_1$ a family $\{\mathbf{g}_n\}$ of tensor fields, where

$$\mathbf{g}_0 = \mathbf{g} \quad \text{and} \quad \mathbf{g}_n = \underbrace{L_{\mathbf{Y}_2}(L_{\mathbf{Y}_2}(\ldots(L_{\mathbf{Y}_2}\mathbf{g})\ldots))}_{n \text{ terms}}.$$

In the coordinates given above, $g_{n\,ab} = \partial^n(g_{ab})/\partial w^n$. Since the solution is analytic, it is completely determined by the family $\mathbf{g}_n$ on $\mathscr{Q}_1$. We shall show that on $\mathscr{Q}_1$, the Lie derivatives with respect to $\hat{\mathbf{K}}$ of all the $\mathbf{g}_n$ vanish. Then the Lie derivative of the $\mathbf{g}_n$ with respect to $\tilde{\mathbf{K}} = \hat{\mathbf{K}} - \mathbf{K}$ will also vanish. This shows that the solution will admit a one-parameter group $\tilde{\theta}_\phi$ generated by $\tilde{\mathbf{K}}$. For simplicity we shall consider only the empty space case, but similar arguments hold in the presence of matter fields, like the electromagnetic or scalar fields, which obey well-behaved hyperbolic equations.

By our choice of coordinates, the components of $L_{\hat{\mathbf{K}}}\mathbf{g}$ are the partial derivatives with respect to v of the coordinate components g_{ab}. These are all constant on $\mathscr{Q}_1$, so $L_{\hat{\mathbf{K}}}\mathbf{g}|_{\mathscr{Q}_1} = 0$. We shall show below $L_{\hat{\mathbf{K}}}\mathbf{g}_1|_{\mathscr{Q}_1} = 0$, and then proceed by a method of induction. Suppose that

$$L_{\hat{\mathbf{K}}}\mathbf{g}_n|_{\mathscr{Q}_1} = 0, \quad n \geqslant 1.$$

It then follows from the construction of the basis that $L_{\hat{\mathbf{K}}}$ of the nth covariant derivatives of all the basis vectors $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Z}, \overline{\mathbf{Z}}$ are zero. Now

$$g_{n+1\,ab} = g_{n\,ab;c}Y_2{}^c + g_{n\,cb}Y_2{}^c{}_{;a} + g_{n\,ac}Y_2{}^c{}_{;b}.$$

The Lie derivative with respect to $\hat{\mathbf{K}}$ of the second and third terms on the right are zero. The first term involves covariant derivatives of $\mathbf{Y}_2$ of order $(n+1)$ and lower orders. The Lie derivative with respect to $\hat{\mathbf{K}}$ of all the lower order terms are zero. The terms involving $(n+1)$ covariant derivatives are

$$\begin{aligned}
&(Y_{2\,a;bef\ldots ghc} + Y_{2\,b;aef\ldots ghc})Y_2{}^eY_2{}^f\ldots Y_2{}^hY_2{}^c \\
&\quad = (Y_{2\,a;be}Y_2{}^e + Y_{2\,b;ae}Y_2{}^e)_{;f\ldots ghc}Y_2{}^f\ldots Y_2{}^c + \text{lower order terms} \\
&\quad = ((Y_{2\,a;e}Y_2{}^e)_{;b} + R_{pabe}Y_2{}^pY_2{}^e + (Y_{2\,b;e}Y_2{}^e)_{;a} + R_{pbae}Y_2{}^pY_2{}^e)_{;f\ldots gh} \\
&\qquad \times Y_2{}^f\ldots Y_2{}^c + \text{lower order terms.}
\end{aligned}$$

The Lie derivatives with respect to $\hat{\mathbf{K}}$ of this expression will be zero, if the Lie derivative with respect to $\hat{\mathbf{K}}$ of the Riemann tensor and its covariant derivatives to order $(n-1)$ vanish. Then $L_{\hat{\mathbf{K}}}\mathbf{g}_{n+1}|_{\mathscr{Q}_1}$ will be zero.

To show that the Lie derivatives with respect to $\hat{\mathbf{K}}$ of $\mathbf{g}_1$ and of the covariant derivatives of the Riemann tensor are zero, it is convenient to use some notation introduced by Newman and Penrose (1962).

This involves using a pseudo-orthonormal basis with the two spacelike vectors $\mathbf{Y}_3$ and $\mathbf{Y}_4$ combined to give a single complex null vector $\mathbf{Z}$, giving each component of the connection and the curvature tensor a separate symbol, and writing out all the Bianchi identities and the defining equations for the curvature tensor explicitly without summation. These relations are combined in pairs to form half the number of complex equations. The symbols for the connection components are:

$$\kappa = Y_{1a;b}Z^aY_1{}^b, \qquad \pi = -Y_{2a;b}\bar{Z}^aY_1{}^b,$$

$$\rho = Y_{1a;b}Z^a\bar{Z}^b, \qquad \lambda = -Y_{2a;b}\bar{Z}^a\bar{Z}^b,$$

$$\sigma = Y_{1a;b}Z^aZ^b, \qquad \mu = -Y_{2a;b}\bar{Z}^aZ^b,$$

$$\tau = Y_{1a;b}Z^aY_2{}^b, \qquad \nu = -Y_{2a;b}\bar{Z}^aY_2{}^b,$$

$$\epsilon = \tfrac{1}{2}(Y_{1a;b}Y_2{}^aY_1{}^b - Z_{a;b}\bar{Z}^aY_1{}^b), \quad \alpha = \tfrac{1}{2}(Y_{1a;b}Y_2{}^a\bar{Z}^b - Z_{a;b}\bar{Z}^a\bar{Z}^b),$$

$$\beta = \tfrac{1}{2}(Y_{1a;b}Y_2{}^aZ^b - Z_{a;b}\bar{Z}^aZ^b), \quad \gamma = \tfrac{1}{2}(Y_{1a;b}Y_2{}^aY_2{}^b - Z_{a;b}\bar{Z}^aY_2{}^b).$$

The symbols for the Weyl tensor are:

$$\Psi_0 = -C_{abcd}Y_1{}^aZ^bY_2{}^cZ^d,$$

$$\Psi_1 = -C_{abcd}Y_1{}^aY_2{}^bY_1{}^cZ^d,$$

$$\Psi_2 = -\tfrac{1}{2}C_{abcd}(Y_1{}^aY_2{}^bY_1{}^cY_2{}^d - Y_1{}^aY_2{}^bZ^c\bar{Z}^d),$$

$$\Psi_3 = C_{abcd}Y_1{}^aY_2{}^bY_2{}^c\bar{Z}^d$$

$$\Psi_4 = -C_{abcd}Y_2{}^a\bar{Z}^bY_2{}^c\bar{Z}^d.$$

We are considering empty space, so the Ricci tensor is zero (i.e. $\Phi_{AB} = 0 = \Lambda$ in the Newman–Penrose formalism). Since the basis is parallelly propagated along $\mathbf{Y}_2$, $\nu = \gamma = \tau = 0$. As $\mathbf{Y}_2$ is the gradient of the coordinate v, $\pi = \bar{\beta} + \alpha$ and $\mu = \bar{\mu}$. Furthermore on $\mathscr{Q}_1$, $\kappa = \rho = \sigma = 0$, $\epsilon = \bar{\epsilon}$, $Y_1(\epsilon) = 0$ and $\Psi_0 = 0$.

The equations we shall need are:

$$Y_1(\alpha) - \bar{Z}(\epsilon) = (\rho + \bar{\epsilon} - 2\epsilon)\alpha + \beta\bar{\sigma} - \bar{\beta}\epsilon - \kappa\lambda + (\epsilon + \rho)\pi, \tag{9.11a}$$

$$Y_1(\beta) - Z(\epsilon) = (\alpha + \pi)\sigma + (\bar{\rho} - \bar{\epsilon})\beta - \mu\kappa - (\bar{\alpha} - \bar{\pi})\epsilon + \Psi_1, \tag{9.11b}$$

$$Y_1(\lambda) - \bar{Z}(\pi) = \rho\lambda + \bar{\sigma}\mu + \pi^2 + (\alpha - \bar{\beta})\pi - (3\epsilon - \bar{\epsilon})\lambda, \tag{9.11c}$$

$$Y_1(\mu) - Z(\pi) = \bar{\rho}\mu + \sigma\lambda + \pi\bar{\pi} - (\epsilon + \bar{\epsilon})\mu - \pi(\bar{\alpha} - \beta) + \Psi_2, \tag{9.11d}$$

$$Z(\rho) - \bar{Z}(\sigma) = \rho(\bar{\alpha} + \beta) - \sigma(3\alpha - \bar{\beta}) - \Psi_1 \tag{9.11e}$$

(these are obtained from the Newman–Penrose equations (4.2)), and:

$$Y_1(\Psi_1) - \bar{Z}(\Psi_0) = -3\kappa\Psi_2 + (2\epsilon + 4\rho)\Psi_1 - (-\pi + 4\alpha)\Psi_0, \quad (9.12a)$$

$$Y_1(\Psi_2) - \bar{Z}(\Psi_1) = -2\kappa\Psi_3 + 3\rho\Psi_2 - (-2\pi + 2\alpha)\Psi_1 - \lambda\Psi_0, \quad (9.12b)$$

$$Y_1(\Psi_3) - \bar{Z}(\Psi_2) = -\kappa\Psi_4 - (2\epsilon - 2\rho)\Psi_3 + 3\pi\Psi_2 - 2\lambda\Psi_1, \quad (9.12c)$$

$$Y_1(\Psi_4) - \bar{Z}(\Psi_3) = -(4\epsilon - \rho)\Psi_4 + (4\pi + 2\alpha)\Psi_3 - 3\lambda\Psi_2, \quad (9.12d)$$

$$Y_2(\Psi_0) - Z(\Psi_1) = -\mu\Psi_0 - 2\beta\Psi_1 + 3\sigma\Psi_2 \quad (9.12e)$$

(these are obtained from the Newman–Penrose equations (4.5)).

From (9.11e), $\Psi_1 = 0$ on $\mathcal{Q}_1$. Then from (9.12b), $Y_1(\Psi_2) = \hat{K}(\Psi_2) = 0$ on $\mathcal{Q}_1$. Adding (9.11a) to the complex conjugate of (9.11b), one obtains

$$Y_1(\pi) = Y_1(\alpha + \bar{\beta}) = \bar{Z}(\epsilon) + Z(\bar{\epsilon}) + 2\pi\rho + 2\overline{\pi\sigma} - \pi(\epsilon - \bar{\epsilon}) - \kappa\lambda - \overline{\kappa\mu} + \bar{\Psi}_1.$$

On $\mathcal{Q}_1$, this becomes $\quad Y_1(\pi) = \bar{Z}(\epsilon) + Z(\bar{\epsilon}).$

Therefore $Y_1(Y_1(\pi)) = Y_1(\bar{Z}(\epsilon) + Z(\bar{\epsilon}))$ on $\mathcal{Q}_1$. But on $\mathcal{Q}_1$, $L_{\mathbf{Y}_1}\mathbf{Z} = 0$ and $Y_1(\epsilon) = 0$. Thus $Y_1(Y_1(\pi)) = 0$ on $\mathcal{Q}_1$. This shows that $\pi = A + Bv$ on $\mathcal{Q}_1$, where A and B are constant along a generator of $\mathcal{Q}_1$. However $\pi|_p = \pi|_{\theta_{t_1}(p)}$; therefore π is a constant along the generators of $\mathcal{Q}_1$. Subtracting the complex conjugate of (9.11b) from (9.11a), one finds that $(\alpha - \bar{\beta})$ is constant along the generators.

One now applies similar arguments to (9.11c) and (9.11d) to show that μ and λ are constant along the generators of $\mathcal{Q}_1$. Since π, μ and λ determine the covariant derivative of $\mathbf{Y}_2$, it follows that $L_{\hat{\mathbf{K}}} Y_2{}^a{}_{;b} = 0$ on $\mathcal{Q}_1$ and hence that $L_{\hat{\mathbf{K}}}\mathbf{g}_1 = 0$ on $\mathcal{Q}_1$.

One can also apply the above kind of argument to (9.12c) and (9.12d) to show that $Y_1(\Psi_3) = Y_1(\Psi_4) = 0$ on $\mathcal{Q}_1$. Thus $L_{\hat{\mathbf{K}}} R_{abcd} = 0$ on $\mathcal{Q}_1$ and so the Lie derivative with respect to $\hat{\mathbf{K}}$ of the second derivatives of the basis vectors are zero. In particular $\mathbf{Y}_1\mathbf{Y}_2$ acting on any of the components of the connection gives zero.

From (9.12e), $\hat{K}(Y_2(\Psi_0)) = Y_1 Y_2(\Psi_0) = 0$ on $\mathcal{Q}_1$. One now operates with $\mathbf{Y}_1\mathbf{Y}_2$ on (9.12a). The commutator $\mathbf{Y}_1\mathbf{Y}_2 - \mathbf{Y}_2\mathbf{Y}_1$ involves only the first covariant derivatives of the basis vectors. Thus

$$L_{\hat{\mathbf{K}}}(\mathbf{Y}_1\mathbf{Y}_2 - \mathbf{Y}_2\mathbf{Y}_1) = 0 \quad \text{on} \quad \mathcal{Q}_1.$$

From this it follows by an argument like that given above that

$$\hat{K}(Y_2(\Psi_1)) = Y_1(Y_2(\Psi_1)) = 0 \quad \text{on} \quad \mathcal{Q}_1.$$

One now repeats the argument for (9.10b), (9.10c) and (9.10d) to show that $\hat{K}(Y_2(\Psi_2)) = \hat{K}(Y_2(\Psi_3)) = \hat{K}(Y_2(\Psi_4)) = 0$ on $\mathcal{Q}_1$. This shows that the Lie derivatives with respect to $\hat{\mathbf{K}}$ of the first covariant derivatives of the Riemann tensor vanish. One then repeats the process, showing that $\hat{K}(Y_2(Y_2(\Psi_0))) = 0$ on $\mathscr{Q}_1$, and so on. □

Proposition 9.3.7

Let $(\mathscr{M}, \mathbf{g})$ be a space–time which admits a two-parameter abelian isometry group with Killing vectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$. Let $\mathscr{V}$ be a connected open set of $\mathscr{M}$, and let $w_{ab} = \xi_{1[a}\xi_{2b]}$. If

(*a*) $w_{ab}R^b{}_c\eta^{cdef}w_{ef} = 0$ on $\mathscr{V}$,

(*b*) $w_{ab} = 0$ at some point of $\mathscr{V}$,

then
$$w_{[ab;c}w_{d]e} = 0 \quad \text{on} \quad \mathscr{V}.$$

Let ${}_{(1)}\chi = \xi_{1a;b}w_{cd}\eta^{abcd}$, and ${}_{(2)}\chi = \xi_{2a;b}w_{cd}\eta^{abcd}$. Then

$$\begin{aligned}\eta^{abcd}{}_{(1)}\chi &= -4!\,\xi_1{}^{[a;b}\xi_1{}^c\xi_2{}^{d]}\\ &= 3!\,\xi_1{}^d\xi_2{}^{[a}\xi_1{}^{b;c]} - 3!\,\xi_2{}^d\xi_1{}^{[a}\xi_1{}^{b;c]} - 2\times 3!\,\xi_1{}^{[a}\xi_2{}^b\xi_1{}^{c];d}.\end{aligned}$$

Therefore

$$\begin{aligned}(3!)^{-1}\eta^{abcd}{}_{(1)}\chi_{;d} = {}& \xi_1{}^d{}_{;d}\xi_2{}^{[a}\xi_1{}^{b;c]} + \xi_1{}^d\xi_2{}^{[a}{}_{;d}\xi_1{}^{b;c]}\\ &+ \xi_1{}^d\xi_2{}^{[a}\xi_1{}^{b;c]}{}_{;d} - \xi_2{}^d{}_{;d}\xi_1{}^{[a}\xi_1{}^{b;c]} - \xi_2{}^d\xi_1{}^{[a}{}_{;d}\xi_1{}^{b;c]}\\ &- \xi_2{}^d\xi_1{}^{[a}\xi_1{}^{b;c]}{}_{;d} - 2\xi_1{}^{[a}{}_{;d}\xi_2{}^b\xi_1{}^{c];d}\\ &- 2\xi_1{}^{[a}\xi_2{}^b{}_{;d}\xi_1{}^{c];d} - 2\xi_1{}^{[a}\xi_2{}^b\xi_1{}^{c];d}{}_{;d}.\end{aligned} \tag{9.13}$$

The first and fourth terms vanish because $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are Killing vectors; the second and fifth terms cancel each other because $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ commute. Because $\boldsymbol{\xi}_1$ is a Killing vector, $L_{\boldsymbol{\xi}_1}\xi_{1a;b} = 0$. This implies that the third term vanishes. Similarly $L_{\boldsymbol{\xi}_2}\xi_{1a;b} = 0$ because $\boldsymbol{\xi}_2$ is a Killing vector which commutes with $\boldsymbol{\xi}_1$. This implies that the sixth and eighth terms cancel. The seventh term vanishes because $\xi_1{}^a{}_{;d}\xi_1{}^{c;d}$ is symmetric; and because of the relation $\xi_{a;bc} = R_{dcba}\xi^d$ satisfied by any Killing vector, $\xi^{a;d}{}_d = -R^a{}_b\xi^b$. Equation (9.13) is therefore

$$\eta^{abcd}{}_{(1)}\chi_{;d} = 2\,.\,3!\,\xi_1{}^{[a}\xi_2{}^bR^{c]}{}_d\xi_1{}^d.$$

By condition (*a*), the right-hand side of this equation vanishes on $\mathscr{V}$. Thus ${}_{(1)}\chi$ is a constant on $\mathscr{V}$; in fact it will be zero on $\mathscr{V}$ since it must vanish when w_{ab} does. Similarly ${}_{(2)}\chi$ will be zero on $\mathscr{V}$. However the vanishing of ${}_{(1)}\chi$ and ${}_{(2)}\chi$ is the necessary and sufficient condition that

$$w_{[ab;c}w_{d]e} = 0.$$ □

Proposition 9.3.8

Let $(\mathscr{M}, \mathbf{g})$ be a stationary axisymmetric regular predictable space–time in which $w_{[ab;c}w_{d]e} = 0$, where $w_{ab} \equiv K_{[a}\tilde{K}_{b]}$. Then at any point in the exterior region $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$ off the axis $\tilde{\mathbf{K}} = 0$, $h \equiv w_{ab} w^{ab}$ is negative. On the horizons $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ and $\dot{J}^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$, h is zero but $w_{ab} \neq 0$ except on the axis.

By proposition 9.3.3, K^a is non-zero in $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$. Let λ be an S^1 which is a non-zero integral curve of the vector field $\tilde{\mathbf{K}}$ in $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap J^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Under the isometry θ_t, λ can be moved into $D^+(\mathscr{S})$. As there are no closed non-spacelike curves in $D^+(\mathscr{S})$, λ must be a spacelike curve, and hence $\tilde{K}^a$ must be spacelike in

$$J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$$

except on the axis where it is zero. Suppose there were some point p at which $\tilde{K}^a$ and K^a were both non-zero and in the same direction. As $\tilde{K}^a$ and K^a commute, the integral curves of $\tilde{K}^a$ through p would coincide with those of K^a. However the former is closed while the latter is not. Thus $\tilde{K}^a$ and K^a are linearly independent where they are non-zero. Thus w_{ab} is non-zero in $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ except on the axis.

The axis will be a two-dimensional surface. Let $\mathscr{Y}$ be the set $J^+(\mathscr{I}^-, \bar{\mathscr{M}}) \cap \overline{J^-}(\mathscr{I}^+, \bar{\mathscr{M}})$ – (the axis), and let $\mathscr{X}$ be the quotient of $\mathscr{Y}$ by $\tilde{\theta}_\phi$. As the integral curves of K^a are closed and spacelike in $\mathscr{Y}$, the quotient $\mathscr{X}$ will be a Hausdorff manifold. On $\mathscr{X}$, there will be a Lorentz metric $\tilde{h}_{ab} = g_{ab} - (\tilde{K}^c \tilde{K}_c)^{-1} \tilde{K}_a \tilde{K}_b$. One can project the Killing vector K^a by $\tilde{h}_{ab}$ to obtain a non-zero vector field $\tilde{h}_{ab} K^b$ in $\mathscr{X}$ which is a Killing vector field for the metric $\tilde{h}_{ab}$. The condition $w_{[ab;c} w_{d]e} = 0$ in $\mathscr{M}$ implies that in $\mathscr{X}$, $(K^b \tilde{h}_{b[c})_{|d} \tilde{h}_{e]f} K^f = 0$, where $|$ denotes the covariant derivative with respect to $\tilde{h}$. This is just the condition that there should exist a function ξ on $\mathscr{X}$ such that $K^b \tilde{h}_{ba} = -\alpha \xi_{|a}$. The argument is then similar to that in proposition 9.3.4. One shows that if $K_a K_b \tilde{h}^{ab} = 0$ at a point $p \in \mathscr{X}$, then the surface $\xi = \xi|_p$ is a null surface in $\mathscr{X}$ with respect to the metric $\tilde{h}$. The function ξ on $\mathscr{X}$ induces a function ξ on $\mathscr{Y}$, with the property: $\xi_{;a} K^a = 0$. Thus $\xi = \xi|_p$ will be a null surface in $\mathscr{M}$ with respect to the metric $\mathbf{g}$.

Suppose p corresponded to an integral curve λ of $\tilde{K}^a$ which did not lie on $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}})$. Let $q \in \mathscr{M}$ be a point of λ. Then there would be a future-directed timelike curve $\gamma(v)$ from $\mathscr{I}^-$ through q to $\mathscr{I}^+$. If this curve intersected the axis, it could be deformed slightly to avoid it. One would then obtain a contradiction similar to that in proposition 9.3.4. □