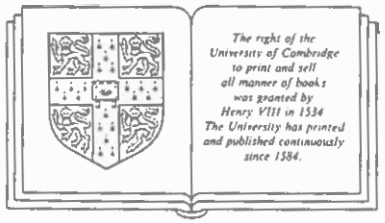
THE LARGE SCALE STRUCTURE OF SPACE-TIME

S. W. HAWKING, F.R.S.
Lucasian Professor of Mathematics in the University of Cambridge and Fellow of Gonville and Caius College

AND

G. F. R. ELLIS
Professor of Applied Mathematics, University of Cape Town



CAMBRIDGE UNIVERSITY PRESS
CAMBRIDGE
LONDON NEW YORK NEW ROCHELLE
MELBOURNE SYDNEY

where $\sigma < 0, x^4 < 0$, are spacelike surfaces which lie entirely inside the past null cone of the origin O, and so are not Cauchy surfaces (see figure 13). In fact the future Cauchy development of $\mathscr{S}_\sigma$ is the region bounded by $\mathscr{S}_\sigma$ and the past light cone of the origin. By lemma 4.5.2, the timelike geodesics through the origin O are orthogonal to the surfaces $\mathscr{S}_\sigma$. If $r \in D^+(\mathscr{S}_\sigma) \cup D^-(\mathscr{S}_\sigma)$ then the timelike geodesic through r and O is the longest timelike curve between r and $\mathscr{S}_\sigma$. If

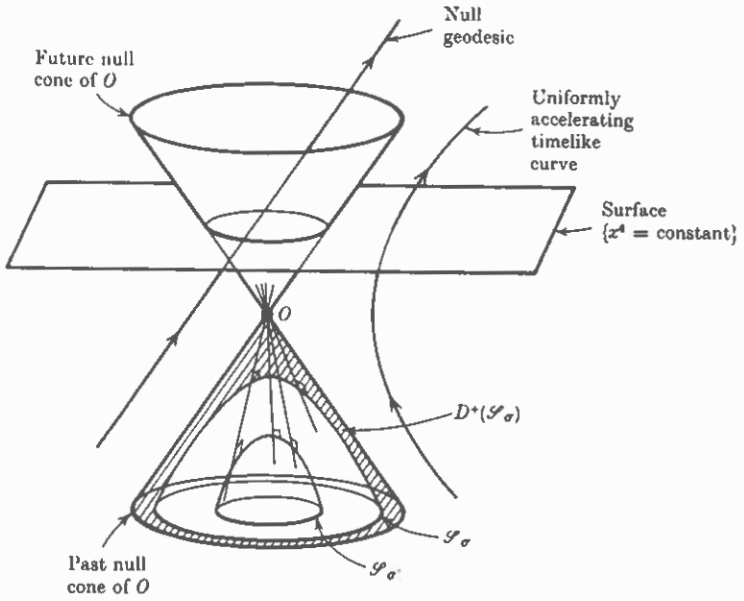


FIGURE 13. A Cauchy surface $\{x^4 = \text{constant}\}$ in Minkowski space–time, and spacelike surfaces $\mathscr{S}_\sigma, \mathscr{S}_{\sigma'}$ which are not Cauchy surfaces. The normal geodesics to the surfaces $\mathscr{S}_\sigma, \mathscr{S}_{\sigma'}$ all intersect at O.

however r does not lie in $D^+(\mathscr{S}_\sigma) \cup D^-(\mathscr{S}_\sigma)$ there is no longest timelike curve between r and $\mathscr{S}_\sigma$: either r lies in the region $\sigma \geqslant 0$, in which case there is no timelike geodesic through r orthogonal to $\mathscr{S}_\sigma$, or r lies in the region $\sigma < 0$, $x^4 \geqslant 0$, in which case there is a timelike geodesic through r orthogonal to $\mathscr{S}_\sigma$ but this geodesic is not the longest curve between r and $\mathscr{S}_\sigma$ as it contains a conjugate point to $\mathscr{S}_\sigma$ at O (cf. figure 13).

To study the structure of infinity in Minkowski space–time, we shall use the interesting representation of this space–time given by Penrose. From the null coordinates v, w, we define new null coordinates in

which the infinities of v, w have been transformed to finite values; thus we define p, q by $\tan p = v$, $\tan q = w$ where $-\frac{1}{2}\pi < p < \frac{1}{2}\pi$, $-\frac{1}{2}\pi < q < \frac{1}{2}\pi$ (and $p \geqslant q$). Then the metric of $(\mathscr{M}, \eta)$ takes the form

$$ds^2 = \sec^2 p \sec^2 q(-dp\,dq + \tfrac{1}{4}\sin^2(p-q)(d\theta^2 + \sin^2\theta\,d\phi^2)).$$

The physical metric η is therefore conformal to the metric $\bar{\mathbf{g}}$ given by

$$d\bar{s}^2 = -4dp\,dq + \sin^2(p-q)(d\theta^2 + \sin^2\theta\,d\phi^2). \tag{5.5}$$

This metric can be reduced to a more usual form by defining

$$t' = p+q, \quad r' = p-q,$$

where
$$-\pi < t'+r' < \pi, \quad -\pi < t'-r' < \pi, \quad r' \geqslant 0; \tag{5.6}$$

(5.5) is then

$$d\bar{s}^2 = -(dt')^2 + (dr')^2 + \sin^2 r'(d\theta^2 + \sin^2\theta\,d\phi^2). \tag{5.7}$$

Thus the whole of Minkowski space–time is given by the region (5.6) of the metric

$$ds^2 = \tfrac{1}{4}\sec^2(\tfrac{1}{2}(t'+r'))\sec^2(\tfrac{1}{2}(t'-r'))\,d\bar{s}^2$$

where $d\bar{s}^2$ is determined by (5.7); the coordinates t, r of (5.3) are related to t', r' by

$$2t = \tan(\tfrac{1}{2}(t'+r')) + \tan(\tfrac{1}{2}(t'-r')),$$

$$2r = \tan(\tfrac{1}{2}(t'+r')) - \tan(\tfrac{1}{2}(t'-r')).$$

Now the metric (5.7) is locally identical to that of the Einstein static universe (see § 5.3), which is a completely homogeneous space–time. One can analytically extend (5.7) to the whole of the Einstein static universe, that is one can extend the coordinates to cover the manifold $R^1 \times S^3$ where $-\infty < t' < \infty$ and r', θ, ϕ are regarded as coordinates on S^3 (with coordinate singularities at $r' = 0$, $r' = \pi$ and $\theta = 0$, $\theta = \pi$ similar to the coordinate singularities in (5.3); these singularities can be removed by transforming to other local coordinates in a neighbourhood of points where (5.7) is singular). On suppressing two dimensions, one can represent the Einstein static universe as the cylinder $x^2 + y^2 = 1$ imbedded in a three-dimensional Minkowski space with metric $ds^2 = -dt^2 + dx^2 + dy^2$ (the full Einstein static universe can be imbedded as the cylinder $x^2 + y^2 + z^2 + w^2 = 1$ in a five-dimensional Euclidean space with metric $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 + dw^2$, cf. Robertson (1933)).

One therefore has the situation: the whole of Minkowski space–time is conformal to the region (5.6) of the Einstein static universe, that is,

to the shaded area in figure 14. The boundary of this region may therefore be thought of as representing the conformal structure of infinity of Minkowski space–time. It consists of the null surfaces $p = \frac{1}{2}\pi$ (labelled $\mathscr{I}^+$) and $q = -\frac{1}{2}\pi$ (labelled $\mathscr{I}^-$) together with points $p = \frac{1}{2}\pi$, $q = \frac{1}{2}\pi$ (labelled i^+), $p = \frac{1}{2}\pi$, $q = -\frac{1}{2}\pi$ (labelled i^0) and $p = -\frac{1}{2}\pi$, $q = -\frac{1}{2}\pi$ (labelled i^-). Any future-directed timelike geodesic in

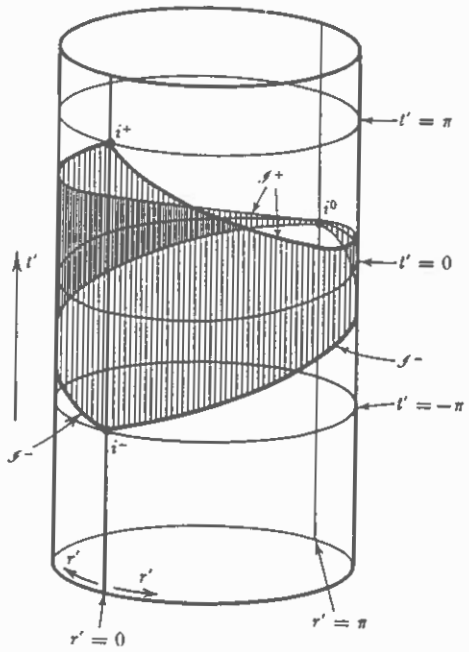


FIGURE 14. The Einstein static universe represented by an imbedded cylinder; the coordinates θ, ϕ have been suppressed. Each point represents one half of a two-sphere of area $4\pi\sin^2 r'$. The shaded region is conformal to the whole of Minkowski space–time; its boundary (part of the null cones of i^+, i^0 and i^-) may be regarded as the conformal infinity of Minkowski space–time.

Minkowski space approaches i^+ (i^-) for indefinitely large positive (negative) values of its affine parameter, so one can regard any timelike geodesic as originating at i^- and finishing at i^+ (cf. figure 15(*i*)). Similarly one can regard null geodesics as originating at $\mathscr{I}^-$ and ending at $\mathscr{I}^+$, while spacelike geodesics both originate and end at i^0. Thus one may regard i^+ and i^- as representing future and past timelike infinity, $\mathscr{I}^+$ and $\mathscr{I}^-$ as representing future and past null infinity, and i^0 as representing spacelike infinity. (However non-geodesic curves do not

obey these rules; e.g. non-geodesic timelike curves may start on $\mathscr{I}^-$ and end on $\mathscr{I}^+$.) Since any Cauchy surface intersects all timelike and null geodesics, it is clear that it will appear as a cross-section of the space everywhere reaching the boundary at i^0.

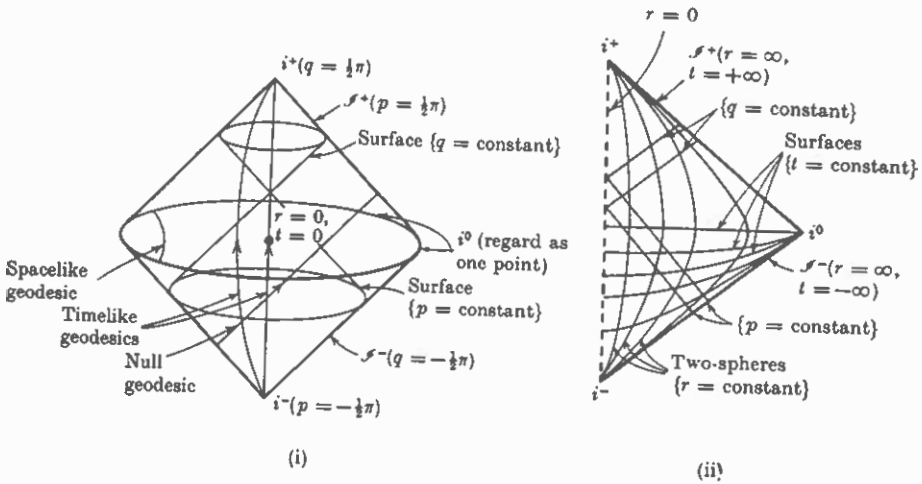


FIGURE 15

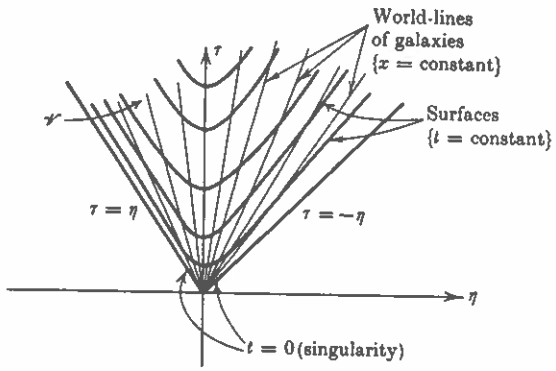
(i) The shaded region of figure 14, with only one coordinate suppressed, representing Minkowski space–time and its conformal infinity.

(ii) The Penrose diagram of Minkowski space–time; each point represents a two-sphere, except for i^+, i^0 and i^-, each of which is a single point, and points on the line $r = 0$ (where the polar coordinates are singular).

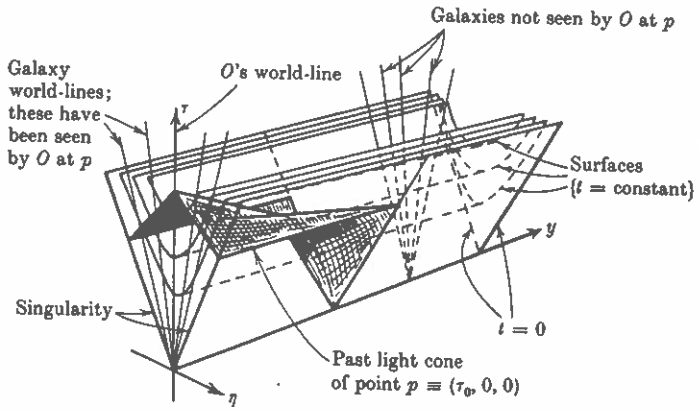
One can also represent the conformal structure of infinity by drawing a diagram of the (t', r') plane, see figure 15(ii). As in figure 12(ii), each point of this diagram represents a sphere S^2, and radial null geodesics are represented by straight lines at $\pm 45°$. In fact, the structure of infinity in any spherically symmetric space–time can be represented by a diagram of this sort, which we shall call a *Penrose diagram*. On such diagrams we shall represent infinity by single lines, the origin of polar coordinates by dotted lines, and irremovable singularities of the metric by double lines.

The conformal structure of Minkowski space we have described is what one would regard as the 'normal' behaviour of a space–time at infinity; we shall encounter different types of behaviour in later sections.

Finally, we mention that one can obtain spaces locally identical to $(\mathscr{M}, \eta)$ but with different (large scale) topological properties by identi-



(i)



(ii)

FIGURE 22. Dust-filled Bianchi I space with a pancake singularity.
(i) The (τ, η) plane; null lines are at $\pm 45°$.
(ii) A half-section of the space in (τ, η, y) coordinates (the z-coordinate is suppressed), showing the past light cone of the point $p = (\tau_0, 0, 0)$. There is a particle horizon in the y-direction but not in the x- (i.e. η) direction.

field equations equivalent to (5.16), (5.17) inside $\mathscr{V}$, and is a flat space–time outside $\mathscr{V}$. However the solution is not C^1 across the boundary of $\mathscr{V}$, and in fact the density of matter becomes infinite on this boundary (as $S \to 0$ there). Since the first derivatives are not square integrable, the Einstein field equations cannot be interpreted on the boundary even in a distributional sense (see § 8.4). While the

extension onto the boundary is unique, it is in no way unique beyond the boundary. We have carried out the extension in the case of dust; a similar extension could be carried out if one had a mixture of matter and radiation.

Let us now return to considering general non-empty spatially homogeneous models. The existence of a singularity in these models will follow directly from Raychaudhuri's equation if the motion of the matter is geodesic and without rotation (as must be the case, for example, if the world-lines are orthogonal to the surfaces of homogeneity) and the timelike convergence condition is satisfied; however there exist such spaces in which the matter accelerates and rotates, and either of these factors could possibly prevent the existence of a singularity. The following result, which is an improved version of a theorem of Hawking and Ellis (1965), shows that in fact neither acceleration nor rotation can prevent the existence of singularities in these models.

Theorem

$(\mathscr{M}, g)$ cannot be timelike geodesically complete if:

(1) $R_{ab}K^aK^b > 0$ for all timelike and null vectors K (this is true if the energy–momentum tensor is type I (§ 4.3) and $\mu + p_i > 0$, $\mu + \sum_i p_i - 4\pi\Lambda > 0$);

(2) there exist equations of motion for the matter fields such that the Cauchy problem has a unique solution (see chapter 7);

(3) the Cauchy data on some spacelike three-surface $\mathscr{H}$ is invariant under a group of diffeomorphisms of $\mathscr{H}$ which is transitive on $\mathscr{H}$.

Since the intrinsic geometry of $\mathscr{H}$ is invariant under a transitive group of diffeomorphisms, these are isometries and $\mathscr{H}$ is complete, i.e. cannot have any boundary. It can be shown (see § 6.5) that if there is a non-spacelike curve which intersects $\mathscr{H}$ more than once, then there exists a covering manifold $\hat{\mathscr{M}}$ of $\mathscr{M}$ in which each connected component of the image of $\mathscr{H}$ will not intersect any non-spacelike curve more than once. We shall assume that $\hat{\mathscr{M}}$ is timelike geodesically complete, and show that this is inconsistent with conditions (1), (2) and (3).

Let $\hat{\mathscr{H}}$ be a connected component of the image of $\mathscr{H}$ in $\hat{\mathscr{M}}$. By (3), the Cauchy data on $\hat{\mathscr{H}}$ is homogeneous. Therefore by condition (2), the Cauchy development of any region of $\hat{\mathscr{H}}$ is isometric to the Cauchy development of any other similar region of $\hat{\mathscr{H}}$. This implies that the surfaces $\{s = \text{constant}\}$ are homogeneous if they lie within the Cauchy

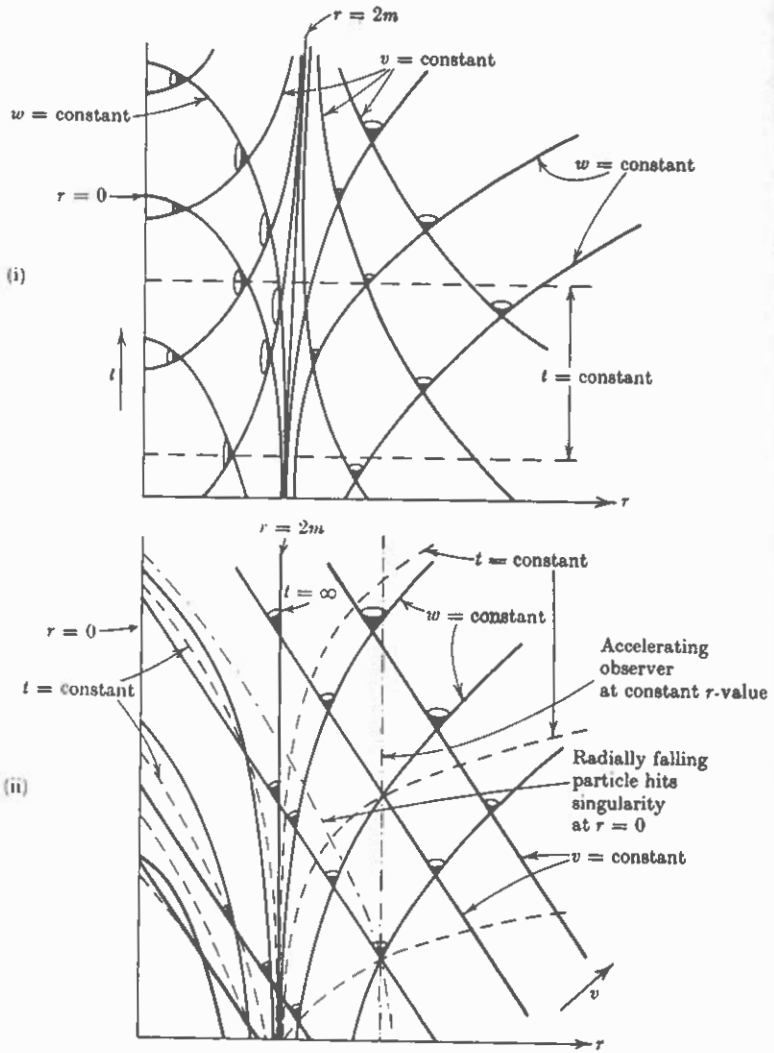


FIGURE 23. Section (θ, ϕ) constant of the Schwarzschild solution.
(i) Apparent singularity at $r = 2m$ when coordinates (t, r) are used.
(ii) Finkelstein diagram obtained by using coordinates (v, r) (lines at 45° are lines of constant v). Surface $r = 2m$ is a null surface on which $t = \infty$.

isometric to $(\mathscr{M}, \mathbf{g})$. A construction of this larger manifold has been given by Kruskal (1960). To obtain it, consider $(\mathscr{M}, \mathbf{g})$ in the coordinates (v, w, θ, ϕ); then the metric takes the form

$$ds^2 = -\left(1 - \frac{2m}{r}\right) dv\, dw + r^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

where r is determined by

$$\tfrac{1}{2}(v - w) = r + 2m \log(r - 2m).$$

This presents the two-space (θ, ϕ constant) in null conformally flat coordinates, as the space with metric $ds^2 = -dv\, dw$ is flat. The most general coordinate transformation which leaves this two-space expressed in such conformally flat double null coordinates is $v' = v'(v)$, $w' = w'(w)$ where v' and w' are arbitrary C^1 functions. The resulting metric is

$$ds^2 = -\left(1 - \frac{2m}{r}\right) \frac{dv}{dv'} \frac{dw}{dw'} dv'\, dw' + r^2(d\theta^2 + \sin^2\theta\, d\phi^2).$$

To reduce this to a form corresponding to that obtained earlier for Minkowski space-time, define

$$x' = \tfrac{1}{2}(v' - w'), \quad t' = \tfrac{1}{2}(v' + w').$$

The metric takes the final form

$$ds^2 = F^2(t', x')(-dt'^2 + dx'^2) + r^2(t', x')(d\theta^2 + \sin^2\theta\, d\phi^2). \tag{5.23}$$

The choice of the functions v', w' determines the precise form of the metric. Kruskal's choice was $v' = \exp(v/4m)$, $w' = -\exp(-w/4m)$. Then r is determined implicitly by the equation

$$(t')^2 - (x')^2 = -(r - 2m)\exp(r/2m) \tag{5.24}$$

and F is given by

$$F^2 = \exp(-r/2m) \,.\, 16m^2/r. \tag{5.25}$$

On the manifold $\mathscr{M}^*$ defined by the coordinates (t', x', θ, ϕ) for $(t')^2 - (x')^2 < 2m$, the functions r and F (defined by (5.24), (5.25)) are positive and analytic. Defining the metric $\mathbf{g}^*$ by (5.23), the region I of $(\mathscr{M}^*, \mathbf{g}^*)$ defined by $x' > |t'|$ is isometric to $(\mathscr{M}, \mathbf{g})$, the region of the Schwarzschild solution for which $r > 2m$. The region defined by $x' > -t'$ (regions I and II in figure 24) is isometric to the advanced Finkelstein extension $(\mathscr{M}', \mathbf{g}')$. Similarly the region defined by $x' > t'$ (regions I and II′ in figure 24) is isometric to the retarded Finkelstein extension $(\mathscr{M}'', \mathbf{g}'')$. There is also a region I′, defined by $x' < -|t'|$,

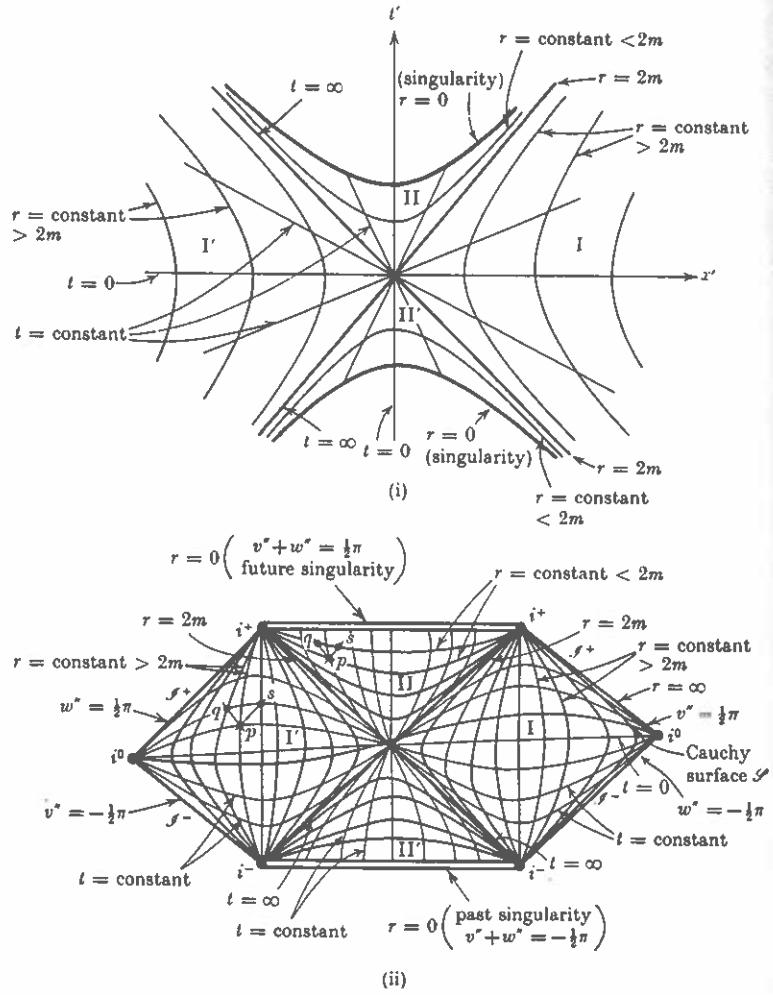


FIGURE 24. The maximal analytic Schwarzschild extension. The θ, ϕ coordinates are suppressed; null lines are at $\pm 45°$. Surfaces $\{r = \text{constant}\}$ are homogeneous.

(i) The Kruskal diagram, showing asymptotically flat regions I and I′ and regions II, II′ for which $r < 2m$.

(ii) Penrose diagram, showing conformal infinity as well as the two singularities.

which turns out to be again isometric with the exterior Schwarzschild solution $(\mathscr{M}, \mathbf{g})$. This can be regarded as another asymptotically flat universe on the other side of the Schwarzschild 'throat'. (Consider the section $t = 0$. The two-spheres $\{r = \text{constant}\}$ behave as in Euclidean

space, for large r; however for small r, they have an area which decreases to the minimum value $16\pi m^2$ and then increases again, as the two spheres expand into the other asymptotically flat three-space.) The regions I′ and II are isometric with the advanced Finkelstein extension of region I′, and similarly I′ and II′ are isometric with the retarded Finkelstein extension of I′, as can be seen from figure 24. There are no timelike or null curves which go from region I to region I′. All future-directed timelike or null curves which cross the part of the surface $r = 2m$ represented here by $t' = |x'|$ approach the singularity at $t' = (2m + (x')^2)^{\frac{1}{2}}$, where $r = 0$. Similarly past-directed timelike or null curves which cross $t' = -|x'|$ approach another singularity at $t' = -(2m + (x')^2)^{\frac{1}{2}}$, where again $r = 0$.

The Kruskal extension $(\mathscr{M}^*, \mathbf{g}^*)$ is the unique analytic and locally inextendible extension of the Schwarzschild solution. One can construct the Penrose diagram of the Kruskal extension by defining new advanced and retarded null coordinates

$$v'' = \arctan(v'(2m)^{-\frac{1}{2}}), \quad w'' = \arctan(w'(2m)^{-\frac{1}{2}})$$

for $\quad -\pi < v'' + w'' < \pi \quad$ and $\quad -\frac{1}{2}\pi < v'' < \frac{1}{2}\pi, \ -\frac{1}{2}\pi < w'' < \frac{1}{2}\pi$

(see figure 24 (ii)). This may be compared with the Penrose diagram for Minkowski space (figure 15 (ii)). One now has future, past and null infinities for each of the asymptotically flat regions I and I′. Unlike Minkowski space, the conformal metric is continuous but not differentiable at the points i^0.

If we consider the future light cone of any point outside $r = 2m$, the radial outwards geodesic reaches infinity but the inwards one reaches the future singularity; if the point lies inside $r = 2m$, both these geodesics hit the singularity, and the entire future of the point is ended by the singularity. Thus the singularity may be avoided by any particle outside $r = 2m$ (so it is not 'universal' as it is in the Robertson–Walker spaces), but once a particle has fallen inside $r = 2m$ (in region II) it cannot evade the singularity. This fact will turn out to be closely related to the following property: each point inside region II represents a two-sphere that is a closed trapped surface. This means the following: consider any two-sphere p (represented by a point in figure 24) and two two-spheres q, s formed by photons emitted radially outwards, inwards at one instant from p. The area of q (which is given by $4\pi r^2$) will be greater than the area of p, but the area of s will be less than the area of p, if all three lie in a region $r > 2m$. However if they all lie in the region II where $r < 2m$, then the areas of *both* q and s will be less

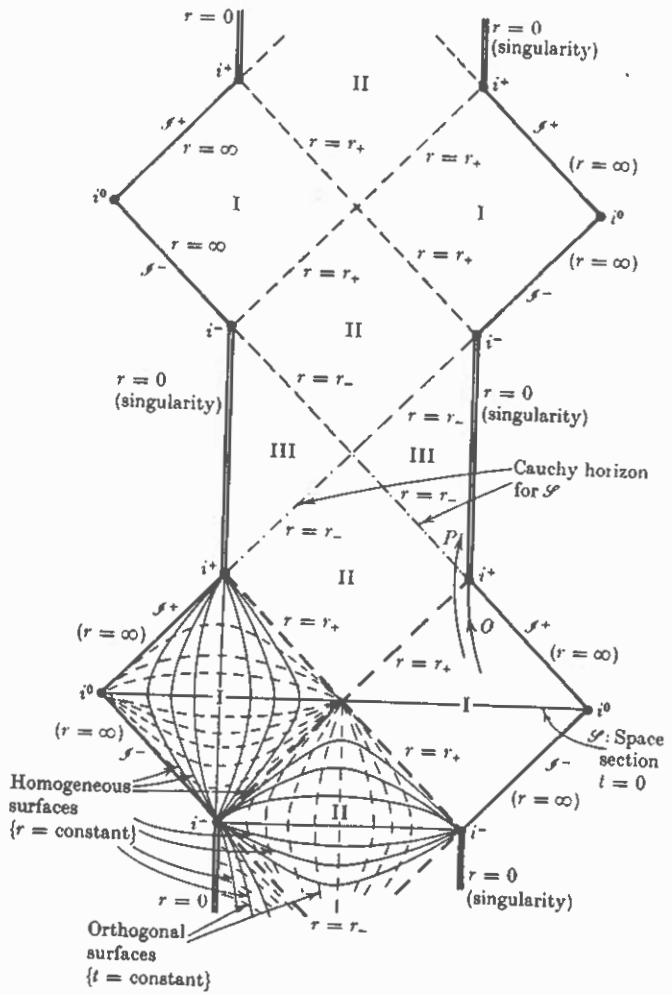


FIGURE 25. Penrose diagram for the maximally extended Reissner–Nordström solution ($e^2 < m^2$). An infinite chain of asymptotically flat regions I ($\infty > r > r_+$) are connected by regions II ($r_+ > r > r_-$) and III ($r_- > r > 0$); each region III is bounded by a timelike singularity at $r = 0$.

but unlike in the Schwarzschild solution, it is timelike and so can be avoided by a future-directed timelike curve from a region I which crosses $r = r_+$. Such a curve can pass through regions II, III and II and re-emerge into another asymptotically flat region I. This raises the intriguing possibility that one might be able to travel to other

universes by passing through the 'wormholes' made by charges. Unfortunately it seems that one would not be able to get back again to our universe to report what one had seen on the other side.

The metric (5.28) is analytic everywhere except at $r = r_-$ where it is degenerate but one can define different coordinates v''' and w''' by

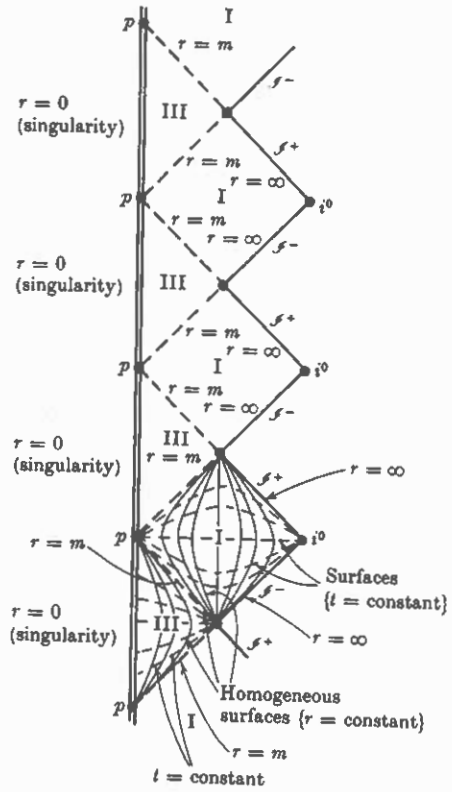
$$v''' = \arctan\left(\exp\left(\frac{r_+ - r_-}{2nr_-^2}v\right)\right),$$

$$w''' = \arctan\left(-\exp\left(\frac{-r_+ + r_-}{2nr_-^2}w\right)\right),$$

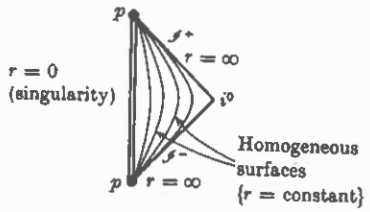
where n is an integer $\geqslant 2(r_+)^2(r_-)^{-2}$. In these coordinates, the metric is analytic everywhere except at $r = r_+$ where it is degenerate. The coordinates v''' and w''' are analytic functions of v'' and w'' for $r \neq r_+$ or r_-. Thus the manifold $\mathscr{M}^*$ can be covered by an analytic atlas, consisting of local coordinate neighbourhoods defined by coordinates v'' and w'' for $r \neq r_-$ and by local coordinate neighbourhoods defined by v''' and w''' for $r \neq r_+$. The metric is analytic in this atlas.

The case $e^2 = m^2$ can be extended similarly; the case $e^2 > m^2$ is already inextendible in the original coordinates. The Penrose diagrams of these two cases are given in figure 26.

In all these cases, the singularity is timelike. This means that, unlike in the Schwarzschild solution, timelike and null curves can always avoid hitting the singularities. In fact the singularities appear to be repulsive: no timelike geodesic hits them, though non-geodesic timelike curves and radial null geodesics can. The spaces are thus timelike (though not null) geodesically complete. The timelike character of the singularity also means that there are no Cauchy surfaces in these spaces: given any spacelike surface, one can find timelike or null curves which run into the singularity and do not cross the surface. For example in the case $e^2 < m^2$, one can find a spacelike surface $\mathscr{S}$ which crosses two asymptotically flat regions I (figure 25). This is a Cauchy surface for the two regions I and the two neighbouring regions II. However in the neighbouring regions III to the future there are past-directed inextendible timelike and null curves which approach the singularity and do not cross the surface $r = r_-$. This surface is therefore said to be the future Cauchy horizon for $\mathscr{S}$. The continuation of the solution beyond $r = r_-$ is not determined by the Cauchy data on $\mathscr{S}$. The continuation we have given is the only locally inextendible analytic one, but there will be other non-analytic C^∞ continuations which satisfy the Einstein–Maxwell equations.



(i)



(ii)

FIGURE 26. Penrose diagrams for the maximally extended Reissner–Nordström solutions:

(i) $e^2 = m^2$, (ii) $e^2 > m^2$.

In the first case there is an infinite chain or regions I ($\infty > r > m$) connected by regions III ($m > r > 0$). The points p are not part of the singularity at $r = 0$, but are really exceptional points at infinity.

A particle P crossing the surface $r = r_+$ would appear to have infinite redshift to an observer O whose world-line remains outside $r = r_+$ and approaches the future infinity i^+ (figure 25). In the region II between $r = r_+$ and $r = r_-$, the surfaces of constant r are spacelike and so each point of the figure represents a two-sphere which is a closed trapped surface. An observer P crossing the surface $r = r_-$ would see the whole of the history of one of the asymptotically flat regions I in a finite time. Objects in this region would therefore appear to be infinitely blue-shifted as they approached i^+. This suggests that the surface $r = r_-$ would be unstable against small perturbations in the initial data on the spacelike surface $\mathscr{S}$, and that such perturbations would in general lead to singularities on $r = r_-$.

5.6 The Kerr solution

In general, astronomical bodies are rotating and so one would not expect the solution outside them to be exactly spherically symmetric. The Kerr solutions are the only known family of exact solutions which could represent the stationary axisymmetric asymptotically flat field outside a rotating massive object. They will be the exterior solutions only for massive rotating bodies with a particular combination of multipole moments; bodies with different combinations of moments will have other exterior solutions. The Kerr solutions do however appear to be the only possible exterior solutions for black holes (see §9.2 and §9.3).

The solutions can be given in Boyer and Lindquist coordinates (r, θ, ϕ, t) in which the metric takes the form

$$ds^2 = \rho^2\left(\frac{dr^2}{\Delta} + d\theta^2\right) + (r^2 + a^2)\sin^2\theta\, d\phi^2 - dt^2 + \frac{2mr}{\rho^2}(a\sin^2\theta\, d\phi - dt)^2, \tag{5.29}$$

where $\rho^2(r,\theta) \equiv r^2 + a^2\cos^2\theta$ and $\Delta(r) \equiv r^2 - 2mr + a^2$.

m and a are constants, m representing the mass and ma the angular momentum as measured from infinity (Boyer and Price (1965)); when $a = 0$ the solution reduces to the Schwarzschild solution. This metric form is clearly invariant under simultaneous inversion of t and ϕ, i.e. under the transformation $t \to -t$, $\phi \to -\phi$, although it is not invariant under inversion of t alone (except when $a = 0$). This is what one would expect, since time inversion of a rotating object produces an object rotating in the opposite direction.

When $a^2 > m^2$, $\Delta > 0$ and the above metric is singular only when $r = 0$. The singularity at $r = 0$ is not in fact a point but a ring, as can be seen by transforming to Kerr–Schild coordinates $(x, y, z, \bar{t})$, where

$$x + iy = (r + ia)\sin\theta \exp i\int (d\phi + a\Delta^{-1}\,dr),$$

$$z = r\cos\theta, \quad \bar{t} = \int (dt + (r^2 + a^2)\Delta^{-1}\,dr) - r.$$

In these coordinates, the metric takes the form

$$ds^2 = dx^2 + dy^2 + dz^2 - d\bar{t}^2 + \frac{2mr^3}{r^4 + a^2z^2}\left(\frac{r(x\,dx + y\,dy) - a(x\,dy - y\,dx)}{r^2 + a^2} + \frac{z\,dz}{r} + d\bar{t}\right)^2, \quad (5.30)$$

where r is determined implicitly, up to a sign, in terms of x, y, z by

$$r^4 - (x^2 + y^2 + z^2 - a^2)r^2 - a^2z^2 = 0.$$

For $r \neq 0$, the surfaces $\{r = \text{constant}\}$ are confocal ellipsoids in the (x, y, z) plane, which degenerate for $r = 0$ to the disc $z^2 + y^2 \leqslant a^2$, $z = 0$. The ring $x^2 + y^2 = a^2$, $z = 0$ which is the boundary of this disc, is a real curvature singularity as the scalar polynomial $R_{abcd}R^{abcd}$ diverges there. However no scalar polynomial diverges on the disc except at the boundary ring. The function r can in fact be analytically continued from positive to negative values through the interior of the disc $x^2 + y^2 < a^2$, $z = 0$, to obtain a maximal analytic extension of the solution.

To do this, one attaches another plane defined by coordinates (x', y', z') where a point on the top side of the disc $x^2 + y^2 < a^2$, $z = 0$ in the (x, y, z) plane is identified with a point with the same x and y coordinates on the bottom side of the corresponding disc in the (x', y', z') plane. Similarly a point on the bottom side of the disc in the (x, y, z) plane is identified with a point on the top side of the disc in the (x', y', z') plane (see figure 27). The metric (5.30) extends in the obvious way to this larger manifold. The metric on the (x', y', z') region is again of the form (5.29), but with negative rather than positive values of r. At large negative values of r, the space is again asymptotically flat but this time with negative mass. For small negative values of r near the ring singularity, the vector $\partial/\partial\phi$ is timelike, so the circles (t = constant, r = constant, θ = constant) are closed timelike curves. These closed timelike curves can be deformed to pass through any point of the extended space (Carter (1968a)). This solution is geodesically incomplete at the ring singularity. However the only timelike and null geodesics which reach this singularity are those in the equatorial plane on the positive r side (Carter (1968a)).

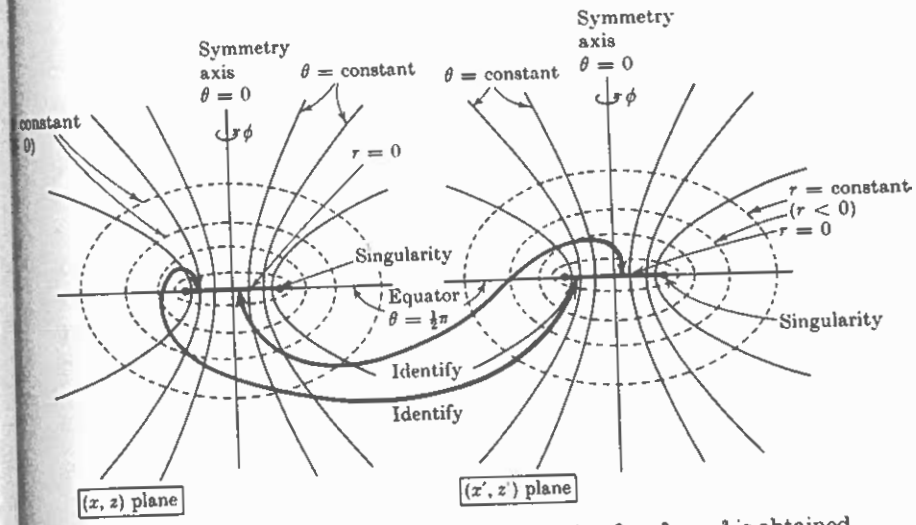


FIGURE 27. The maximal extension of the Kerr solution for $a^2 > m^2$ is obtained by identifying the top of the disc $x^2 + y^2 < a^2$, $z = 0$ in the (x, y, z) plane with the bottom of the corresponding disc in the (x', y', z') plane, and vice versa. The figure shows the sections $y = 0$, $y' = 0$ of these planes. On circling twice round the singularity at $x^2 + y^2 = a^2$, $z = 0$ one passes from the (x, y, z) plane to the (x', y', z') plane (where r is negative) and back to the (x, y, z) plane (where r is positive).

The extension in the case $a^2 < m^2$ is rather more complicated, because of the existence of the two values $r_+ = m + (m^2 - a^2)^{\frac{1}{2}}$ and $r_- = m - (m^2 - a^2)^{\frac{1}{2}}$ of r at which $\Delta(r)$ vanishes. These surfaces are similar to the surfaces $r = r_+$, $r = r_-$ in the Reissner–Nordström solution. To extend the metric across these surfaces, one transforms to the Kerr coordinates (r, θ, ϕ_+, u_+), where

$$du_+ = dt + (r^2 + a^2)\Delta^{-1}\,dr, \quad d\phi_+ = d\phi + a\Delta^{-1}\,dr.$$

The metric then takes the form

$$ds^2 = \rho^2\,d\theta^2 - 2a\sin^2\theta\,dr\,d\phi_+ + 2\,dr\,du_+ + \rho^{-2}[(r^2 + a^2)^2 - \Delta a^2\sin^2\theta]\sin^2\theta\,d\phi_+^2 - 4a\rho^{-2}mr\sin^2\theta\,d\phi_+\,du_+ - (1 - 2mr\rho^{-2})\,du_+^2 \quad (5.31)$$

on the manifold defined by these coordinates, and is analytic at $r = r_+$ and $r = r_-$. One again has a singularity at $r = 0$, which has the same ring form and geodesic structure as that described above. The metric can also be extended on the manifold defined by the coordinates (r, θ, ϕ_-, u_-) where

$$du_- = dt - (r^2 + a^2)\Delta^{-1}dr, \quad d\phi_- = d\phi - a\Delta^{-1}dr;$$

the metric again takes the form (5.31), with ϕ_+, u_+ replaced by $-\phi_-$, $-u_-$. The maximal analytic extension can be built up by a combination of these extensions, as in the Reissner–Nordström case (Boyer and Lindquist (1967), Carter (1968*a*)). The global structure is very similar to that of the Reissner–Nordström solution except that one can now continue through the ring to negative values of r. Figure 28 (i) shows the conformal structure of the solution along the symmetry axis. The regions I represent the asymptotically flat regions in which $r > r_+$. The regions II ($r_- < r < r_+$) contain closed trapped surfaces. The regions III ($-\infty < r < r_-$) contain the ring singularity; there are closed timelike curves through every point in a region III, but no causality violation occurs in the other two regions.

In the case $a^2 = m^2$, r_+ and r_- coincide and there is no region II. The maximal extension is similar to that of the Reissner–Nordström solution when $e^2 = m^2$. The conformal structure along the symmetry axis in this case is shown in figure 28 (ii).

The Kerr solutions, being stationary and axisymmetric, have a two-parameter group of isometries. This group is necessarily Abelian (Carter (1970)). There are thus two independent Killing vector fields which commute. There is a unique linear combination K^a of these Killing vector fields which is timelike at arbitrarily large positive and negative values of r. There is another unique linear combination $\tilde{K}^a$ of the Killing vector fields which is zero on the axis of symmetry. The orbits of the Killing vector K^a define the stationary frame, that is, an object moving along one of these orbits appears to be stationary with respect to infinity. The orbits of the Killing vector $\tilde{K}^a$ are closed curves, and correspond to the rotational symmetry of the solution.

In the Schwarzschild and Reissner–Nordström solutions, the Killing vector K^a which is timelike at large values of r is timelike everywhere in the region I, becoming null on the surfaces $r = 2m$ and $r = r_+$ respectively. These surfaces are null. This means that a particle which crosses one of these surfaces in the future direction cannot return again to the same region. They are the boundary of the region

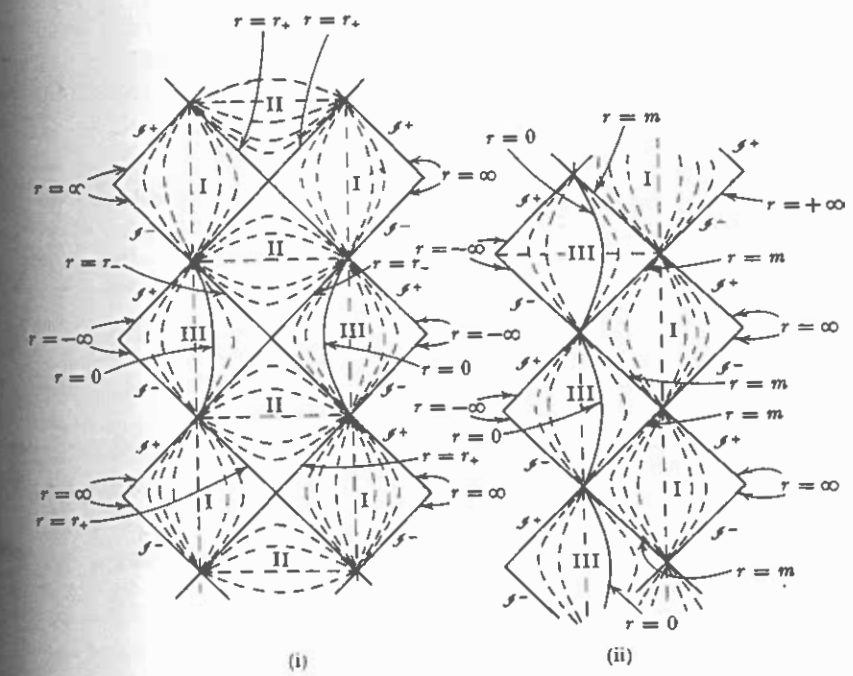


FIGURE 28. The conformal structure of the Kerr solutions along the axis of symmetry, (i) in the case $0 < a^2 < m^2$, (ii) in the case $a^2 = m^2$. The dotted lines are lines of constant r; the regions I, II and III in case (i) are divided by $r = r_+$ and $r = r_-$, and the regions I and III in case (ii) by $r = m$. In both cases, the structure of the space near the ring singularity is as in figure 27.

of the solution from which particles can escape to the infinity $\mathscr{I}^+$ of a particular region I, and are called the *event horizons* of that $\mathscr{I}^+$. (They are in fact the event horizon in the sense of § 5.2 for an observer moving on any of the orbits of the Killing vector K^a in the region I.)

In the Kerr solution on the other hand, the Killing vector K^a is spacelike in a region outside $r = r_+$, called the *ergosphere* (figure 29). The outer boundary of this region is the surface $r = m + (m^2 - a^2\cos^2\theta)^{\frac{1}{2}}$ on which K^a is null. This is called the *stationary limit surface* since it is the boundary of the region in which particles travelling on a timelike curve can travel on an orbit of the Killing vector K^a, and so remain at rest with respect to infinity. The stationary limit surface is a timelike surface except at the two points on the axis, where it is null (at these points it coincides with the surface $r = r_+$). Where it is timelike it can be crossed by particles in either the ingoing or the outgoing direction.

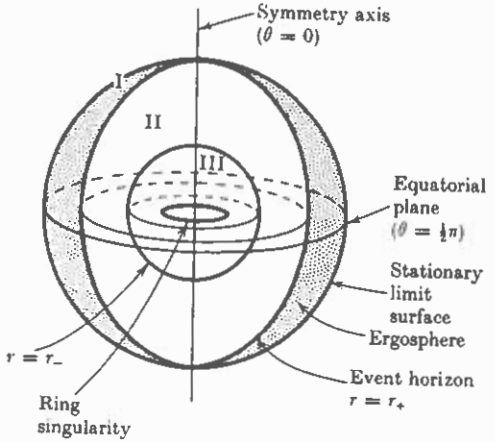


FIGURE 29. In the Kerr solution with $0 < a^2 < m^2$, the ergosphere lies between the stationary limit surface and the horizon at $r = r_+$. Particles can escape to infinity from region I (outside the event horizon $r = r_+$) but not from region II (between $r = r_+$ and $r = r_-$) and region III ($r < r_-$; this region contains the ring singularity).

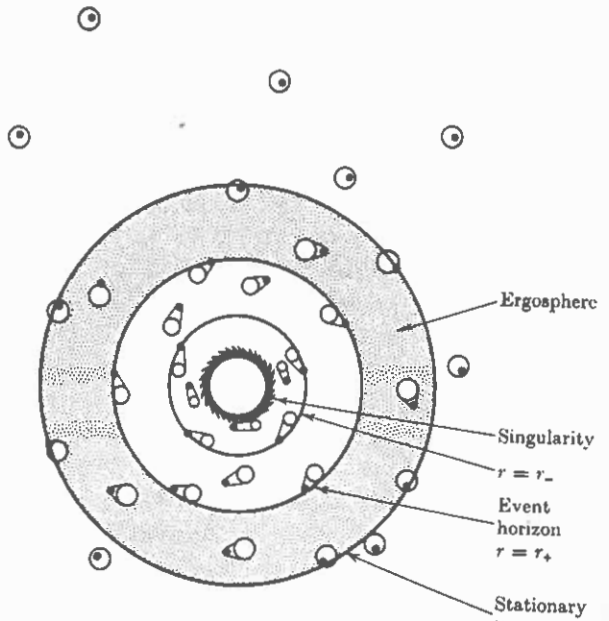


FIGURE 30. The equatorial plane of a Kerr solution with $m^2 > a^2$. The circles represent the position a short time later of flashes of light emitted by the points represented by heavy dots.

It is therefore not the event horizon for $\mathscr{I}^+$. In fact the event horizon is the surface $r = r_+ = m + (m^2 - a^2)^{\frac{1}{2}}$. Figure 30 shows why this is. It shows the equatorial plane $\theta = \frac{1}{2}\pi$; each point in this figure represents an orbit of the Killing vector K^a, i.e. it is stationary with respect to $\mathscr{I}^+$. The small circles represent the position a short time later of flashes of light emitted from the points represented by the heavy black dots. Outside the stationary limit the Killing vector K^a is timelike and so lies within the light cone. This means that the point in figure 30 representing the orbit of emission lies within the wavefront of the light.

On the stationary limit surface, K^a is null and so the point representing the orbit of emission lies on the wavefront. However the wavefront lies partly within and partly outside the stationary limit surface; it is therefore possible for a particle travelling along a timelike curve to escape to infinity from this surface. In the ergosphere between the stationary limit surface and $r = r_+$, the Killing vector K^a is spacelike and so the point representing the orbit of emission lies outside the wavefront. In this region it is impossible for a particle moving on a timelike or null curve to travel along an orbit of the Killing vector and so to remain at rest with respect to infinity. However the positions of the wavefronts are such that the particles can still escape across the stationary limit surface and so out to infinity. On the surface $r = r_+$, the Killing vector K^a is still spacelike. However the wavefront corresponding to a point on this surface lies entirely within the surface. This means that a particle travelling on a timelike curve from a point on or inside the surface cannot get outside the surface and so cannot get out to infinity. The surface $r = r_+$ is therefore the event horizon for $\mathscr{I}^+$ and is a null surface.

Although the Killing vector K^a is spacelike in the ergosphere, the magnitude $K^a\tilde{K}^b K_{[a}\tilde{K}_{b]}$ of the *Killing bivector* $K_{[a}\tilde{K}_{b]}$ is negative everywhere outside $r = r_+$, except on the axis $\tilde{K}^a = 0$ where it vanishes. Therefore K^a and $\tilde{K}^a$ span a timelike two-surface and so at each point outside $r = r_+$ off the axis there is a linear combination of K^a and $\tilde{K}^a$ which is timelike. In a sense, therefore, the solution in the ergosphere is locally stationary, although it is not stationary with respect to infinity. In fact there is no one linear combination of K^a and $\tilde{K}^a$ which is timelike everywhere outside $r = r_+$. The magnitude of the Killing bivector vanishes on $r = r_+$, and is positive just inside this surface. On $r = r_+$, both K^a and $\tilde{K}^a$ are spacelike but there is a linear combination which is null everywhere on $r = r_+$ (Carter (1969)).

The behaviour of the ergosphere and the horizon we have discussed will play an important part in our discussion of black holes in § 9.2 and § 9.3.

Just as the Reissner–Nordström solution can be thought of as a charged version of the Schwarzschild solution, so there is a family of charged Kerr solutions (Carter (1968*a*)). Their global properties are very similar to those of the uncharged Kerr solutions.

5.7 Gödel's universe

In 1949, Kurt Gödel published a paper (Gödel (1949)) which provided a considerable stimulus to investigation of exact solutions more complex than those examined so far. He gave an exact solution of Einstein's field equations in which the matter takes the form of a pressure-free perfect fluid ($T_{ab} = \rho u_a u_b$ where ρ is the matter density and u_a the normalized four-velocity vector). The manifold is R^4 and the metric can be given in the form

$$ds^2 = -dt^2 + dx^2 - \tfrac{1}{2}\exp(2(\sqrt{2})\omega x)\,dy^2 + dz^2 - 2\exp((\sqrt{2})\omega x)\,dt\,dy,$$

where $\omega > 0$ is a constant; the field equations are satisfied if $\mathbf{u} = \partial/\partial x^0$ (i.e. $u^a = \delta^a{}_0$) and

$$4\pi\rho = \omega^2 = -\Lambda.$$

The constant ω is in fact the magnitude of the vorticity of the flow vector u^a.

This space–time has a five-dimensional group of isometries which is transitive, i.e. it is a completely homogeneous space–time. (An action of a group is transitive on $\mathscr{M}$ if it can map any point of $\mathscr{M}$ into any other point of $\mathscr{M}$.) The metric is the direct sum of the metric $\mathbf{g}_1$ given by

$$ds_1^2 = -dt^2 + dx^2 - \tfrac{1}{2}\exp(2(\sqrt{2})\omega x)\,dy^2 - 2\exp((\sqrt{2})\omega x)\,dt\,dy$$

on the manifold $\mathscr{M}_1 = R^3$ defined by the coordinates (t, x, y), and the metric $\mathbf{g}_2$ given by

$$ds_2^2 = dz^2$$

on the manifold $\mathscr{M}_2 = R^1$ defined by the coordinate z. In order to describe the properties of the solution it is sufficient to consider only $(\mathscr{M}_1, \mathbf{g}_1)$.

Defining new coordinates (t', r, ϕ) on $\mathscr{M}_1$ by

$$\exp((\sqrt{2})\omega x) = \cosh 2r + \cos\phi \sinh 2r,$$

$$\omega y \exp((\sqrt{2})\omega x) = \sin\phi \sinh 2r,$$

$$\tan\tfrac{1}{2}(\phi + \omega t - (\sqrt{2})t') = \exp(-2r)\tan\tfrac{1}{2}\phi,$$

the metric $\mathbf{g}_1$ takes the form

$$ds_1^2 = 2\omega^{-2}(-dt'^2 + dr^2 - (\sinh^4 r - \sinh^2 r)\,d\phi^2 + 2(\sqrt{2})\sinh^2 r\,d\phi\,dt),$$

where $-\infty < t < \infty$, $0 \leqslant r < \infty$, and $0 \leqslant \phi \leqslant 2\pi$, $\phi = 0$ being identified with $\phi = 2\pi$; the flow vector in these coordinates is $\mathbf{u} = (\omega/(\sqrt{2}))\,\partial/\partial t'$. This form exhibits the rotational symmetry of the solution about the axis $r = 0$. By a different choice of coordinates the axis could be chosen to lie on any flow line of the matter.

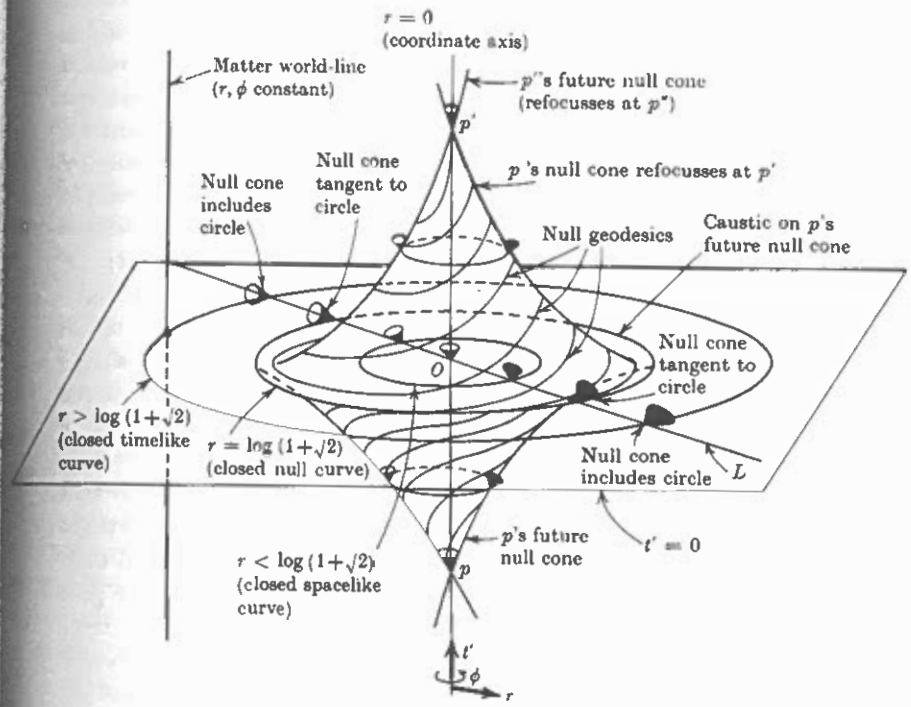


FIGURE 31. Gödel's universe with the irrelevant coordinate z suppressed. The space is rotationally symmetric about any point; the diagram represents correctly the rotational symmetry about the axis $r = 0$, and the time invariance. The light cone opens out and tips over as r increases (see line L) resulting in closed timelike curves. The diagram does not correctly represent the fact that *all* points are in fact equivalent.

The behaviour of $(\mathscr{M}_1, \mathbf{g}_1)$ is illustrated in figure 31. The light cones on the axis $r = 0$ contain the direction $\partial/\partial t'$ (the vertical direction on the diagram) but not the horizontal directions $\partial/\partial r$ and $\partial/\partial\phi$. As one moves away from the axis, the light cones open out and tilt in the

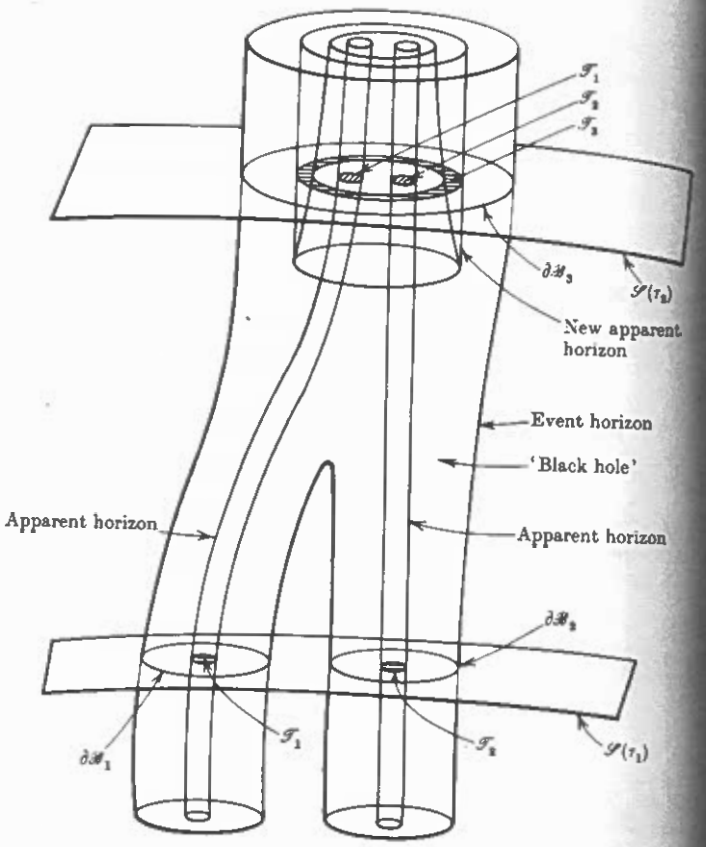


FIGURE 60. The collision and merging of two black holes. At time τ_1, there are apparent horizons $\partial\mathcal{T}_1$, $\partial\mathcal{T}_2$ inside the event horizons $\partial\mathcal{B}_1$, $\partial\mathcal{B}_2$ respectively. By time τ_2, the event horizons have merged to form a single event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

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

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

The null geodesics orthogonal to the apparent horizon $\partial\mathcal{T}(\tau)$ on a surface $\mathcal{S}(\tau)$ 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 $\mathcal{S}(\tau')$ will lie inside the apparent horizon $\partial\mathcal{T}(\tau')$. 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 $(\mathcal{M}, \mathbf{g})$ which satisfy the following conditions:

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

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

(3) $(\mathcal{M}, \mathbf{g})$ 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 $\mathbf{N}, \mathbf{L}$.

Proposition 9.3.1

Let $(\mathscr{M}, \mathbf{g})$ be a stationary, regular predictable space–time. Then if 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 vector these generators; then in $J^+(\mathscr{I}^-, \bar{\mathscr{M}})$, $Y_1{}^a$ has zero shear $\hat{\sigma}$ and expansion θ, and satisfies

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

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 $\mathscr{I}^+$ 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 $t \geqslant t' - x_2$. 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.

We 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 $x > x_2$. However since each point of $\dot{J}^-(\mathscr{I}^+, \bar{\mathscr{M}}) \cap J^+(\mathscr{I}^-, \bar{\mathscr{M}})$ can be mapped 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 $C_{abc[d}Y_{1f]}Y_1{}^bY_1{}^c = 0$, 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 $Y_1{}' = e^{\nu}Y_1$, $Y_2 \to Y_2{}' = e^{-\nu}Y_2$. The induced metric on the space-like 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

$$\hat{h}_a{}^c\hat{h}^b{}_d)_{;g}Y_2{}^g\hat{h}_c{}^s\hat{h}^d{}_t = \hat{h}^{sa}p_{a;b}\hat{h}^b{}_t + p^sp_t - \hat{h}^s{}_aY_1{}^a{}_{;g}\hat{h}^{gc}Y_{2c;b}\hat{h}^b{}_t + R^a{}_{ccb}Y_2{}^cY_1{}^ch_a{}^s\hat{h}^b{}_t, \quad (9.5)$$

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 generators are zero. Under a rescaling transformation $Y_1{}' = e^{\nu}Y_1$,