PART II General Theory of Stationary Black Hole States

1 Introduction

The necessity of facing up to the problem of catastrophic gravitational collapse in astrophysics was first appreciated by Chandrasekhar in 1931 when he discovered the upper mass limit for a spherical equilibrium configuration of a cold sphere of ordinary degenerate matter in Newtonian gravitational theory. Subsequent work by Chandrasekhar, Landau, Oppenheimer and others established fairly clearly by the end of the nineteen thirties that the upper mass limit would not be greatly affected by taking into account the existence of exotic forms of high density matter, such as a degenerate neutron fluid, nor by the corrections resulting from the use of Einstein's theory of gravity, according to which an object above this limit must ultimately disappear from sight within its Schwarzschild horizon—which occurs when the circumferential radius r is diminished to twice the conserved mass M (where here, as throughout this work, we use units in which Newton's constant G and the speed of light c are set equal to unity)—with the subsequent formation of a singularity.

It is true of course that an arbitrarily large mass can exist in quasi-equilibrium in a sufficiently extended differentially rotating disc-like configuration. However, it would be rather surprising if nature contrived to avoid the formation of at least some large stars with sufficiently small angular momentum for its effects to be negligible even after contraction to the Schwarzschild radius. Moreover there are many diverse effects (including viscosity, magnetic fields, and gravitational radiation) which would all tend to transfer angular momentum outwards, except in the special case of uniformly rotating axially symmetric configurations, for which there is in any case an upper mass limit, somewhat larger than the spherical upper mass limit, but of the same order of magnitude. (See for example the work of Ostriker *et al.*) Hence although its onset may in some cases be postponed or prevented, the basic issue of catastrophic collapse cannot be avoided merely by taking into account natural deviations from spherical symmetry.

Despite these considerations there was very little work in the field until the mid nineteen-sixties. The first signs of renewed interest were a pioneering attempt by Regge and Wheeler (1957) to determine the stability of the Schwarzschild horizon, and an equally important pioneering work by Lifshitz and Khalatnikov (1961) in which it was suggested that the ultimate singularity which occurs in the spherical collapse case might be avoided in more general situations. These two works forshadowed the subsequent subdivision of the field of gravitational collapse investigations into two main branches, the first—with which we shall be concerned here—being concerned primarily with horizons and astronomically observable effects, and the second being primarily concerned with the fundamental physical question of the nature of the singularities and the breakdown of the classical Einstein theory. Another event which was of key importance in stimulating interest in the field was the accidental discovery by Kerr (1963) of a rotating generalization of the spherical Schwarzschild solution of Einstein's vacuum equations—accidental in the sense that Kerr's investigations were not directly motivated by the astrophysical collapse problem.

However, the event which laid the foundations of the modern mathematical theory of gravitational collapse was the publication by Penrose (1965) of the now famous singularity theorem which conclusively refuted the suggestion that singularities are merely a consequence of the spherical idealization used in the early work. The actual conclusion of the Penrose theorem was rather restricted, being essentially negative in nature, and although its range of application has been very greatly extended by the subsequent work of Stephen Hawking, there is so far very little positive information about the nature of the singularities whose existence is predicted. The real importance of the Penrose theorem lay rather in the wealth of new techniques and concepts which were introduced in its proof, particularly the notion of a trapped surface and the idea of treating boundary horizons as dynamic entities in their own right. These developments lead directly to the realization that the outcome of a gravitational collapse—in so far as it can be followed up in terms of the classical Einstein theory—would have to be either (a) the formation of a well behaved event horizon separating a well behaved *domain of outer communication* (within which light signals can not only be received from but also sent to arbitrarily large distances) from hidden regions (for which the term *black holes* was subsequently introduced by Wheeler) within which the singularities would be located, and from which no light could escape to large distances, or (b) the formation of what have come to be known as *naked singularities*, i.e. singularities which can be approached arbitrarily closely by time-like curves or light rays which subsequently escape to infinity. If the latter situation were indeed to occur it would mean that the classical Einstein theory would be essentially useless for predicting the subsequent astronomically observable phenomena, so that the development of a more sophisticated gravitational theory would be an urgent practical necessity. However there are a number of features of the situation as we understand it at present which encourage belief in the conjecture which Penrose has termed the *cosmic censorship hypothesis*, which postulates that naked singularities do not in fact occur, i.e. that the only possible outcome of a gravitational collapse is the formation of black holes. It is probably fair to say that the verification or refutation of this conjecture is the most important unresolved mathematical problem in General Relativity theory today. If the cosmic censorship conjecture is correct then there will be no theoretical reason why Einstein's theory of gravity should not be adequate for all foreseeable astronomical purposes, except for dealing with primordial singularities (sometimes referred to as white holes) such as the cosmological big bang. (It will still of course be possible that the theory may fail empirically by conflict with observation.)

Having arrived at the idea of a black hole as a possible outcome, if not an inevitable outcome, of a gravitational collapse, it is natural to conjecture that it should settle down asymptotically in time towards a stationary final equilibrium state—any oscillations being damped out by gravitational radiation—as indicated in Figure 1. It was after all the non-existence of such a stationary final equilibrium state for an ordinary massive star which led to the introduction of the alternative black hole concept in the first place. With these considerations in mind, and the recently discovered Kerr solution as an example, several workers including myself, and notably Bob Boyer, began a systematic study of the properties of stationary black hole states during the years 1965-1966. Too much concentration on the stationary equilibrium states could not of course have been justified if the settling down process took place over astronomically long timescales, but it has subsequently become fairly clear from the dynamical perturbation investigations of the Schwarzschild solution by Doroshkeviteh, Zel'dovich and Novikov (1966), Vishveshwara (1970), Price (1972) and others that (as was suspected, by dimensional considerations, from the outset) any dynamic variations of a black hole can be expected to die away over timescales of the order of the time for light to cross a distance equal to the Schwarzschild radius. These timescales are in fact extremely short by astronomical timescales—of the order of milliseconds for the collapse of an ordinary massive star, and at most hours or days for the collapse of an entire galactic nucleus.

This early work served to clarify many of the elementary properties of the black hole event horizons in the stationary case, and also their relationship with other geometrically defined features such as what has subsequently come to be known as the ergosurface, which are described in detail in sections 2 to 5. This phase of elementary investigation was almost complete by the time of Boyer's tragic death in the summer of 1966, although the results were not actually published until somewhat later, (Boyer (1969), Carter (1969), Vishveshwara (1968)). The results described in section 8 were overlooked in this earlier period and were obtained quite recently (by Hartle, Hawking and myself) while the rigorous proofs in section 9 were actually worked out at Les Houches by Bardeen, Hawking and myself.

An analogous first phase of investigation of the elementary properties of black hole even horizons in the more general *dynamic* case has been carried out more recently almost exclusively by Stephen Hawking, using some of the techniques developed by Penrose, Geroch and himself for proving singularity theorems; the most important of these elementary properties is that the area of the intersection of a constant time hypersurface with the event horizon (which is of course constant in the stationary case) can only increase but never decrease with time in the dynamic case (Hawking 1971). An off-shoot of this work has been a much more sophisticated second phase of investigations of stationary black hole event

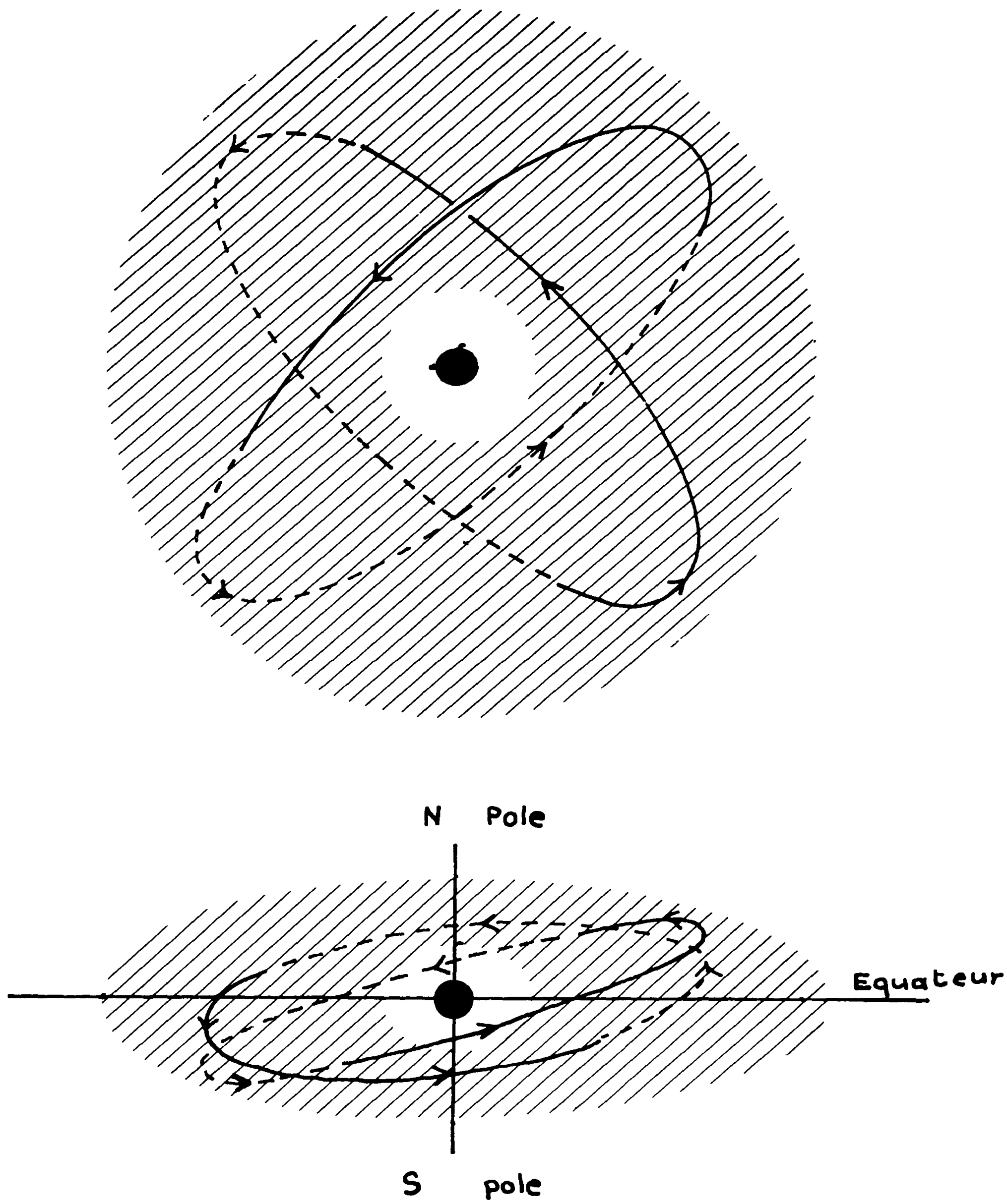

Figure 2. Equatorial and polar sections of a conceivable axially and equatorially symmetric galactic nucleus are illustrated. The nucleus is supposed to consist of a stellar system (whose intersections with the equatorial and polar planes are represented by the shaded regions) with a massive central black hole. (The hole will be non-Kerr-like due to the distorting influence of the stellar mass distribution.) Projections onto the equatorial and polar planes of two stellar orbits (related by a 90 degree axisymmetry rotation) are marked, using dotted lines where the orbits lie behind the relevant planes. These would be typical orbits in a system which is approximately stationary (in that mass redistribution as the stars fall into the hole takes place over timescales long compared with the orbital periods) but in which the circularity condition is violated due to a net inflow of matter at high latitudes compensated by a net outflow near the equator. (In order to obtain this circularity violation it is necessary not only that the orbits be eccentric but also that they should have a consistent average tilt.) The Papapetrou theorem is sufficiently powerful to ensure that the spacetime will be stationary-circular in the empty region in the neighbourhood of the hole despite the non-circularity resulting from the quasi-convective matter flow in the region occupied by the stars. (*See section 7.*)

horizons, in which (using the techniques originally developed by Newman and Penrose for dealing with gravitational radiation problems) Hawking has shown that many of the simplifying assumptions which were taken for granted in the earlier work on stationary event horizons, particularly that of axisymmetry, can in fact be rigorously justified. I shall use what I shall refer to as Hawking's *strong rigidity theorem* as the starting point of section 3 of this course. The actual derivation is discussed by Stephen Hawking himself in the accompanying lecture course, but the full technical details of the proof are beyond the scope of the present proceedings (see Hawking (1972), Hawking and Ellis (1973)). The main content of sections 6 and 7 will be the demonstration that this theorem can be combined with suitable generalizations (which I shall present in full detail) of earlier theorems, originally derived outside the context of the black hole problem, by Lichnerowicz (1945) and Papapetrou (1966), to justify the principle simplifying conditions namely staticity and circularity, which are involved at various stages in section 2 and in the subsequent sections.

The existence of the kinds of the results described in sections 2 to 7 were in most cases at least conjectured before 1967, even though many of the detailed formulae and rigorous proofs which are now available, particularly those due to Hawking, are of more recent origin. However in that year a new result was announced by Israel which took me, and as far as I know everyone else in the field, by surprise, despite the fact that there were already many hints which might have suggested it to us. This was the theorem which could be taken, subject to some fairly obvious assumptions which have since been fully justified, as implying that the Schwarzschild solution is the only possible pure vacuum exterior solution for a stationary black hole which is non-rotating (in the sense made precise in section 4). Until then I think that we had all thought of a black hole even in the stationary limit, as potentially a fairly complicated object. I know that I had always imagined it should have many internal degrees of freedom representing the vestigial multipole structure of the star or other object from whose collapse it had arisen. Yet here was a theorem which implies that if rotational effects were excluded, there could be no internal degrees of freedom at all, apart from the mass itself. My immediate reaction—and as far as I can remember this was how Israel himself saw the situation—was to suppose that this meant that only stars starting off with artificially restricted multipole structures, and in the non-rotating case only exactly spherical stars, could form black holes with well behaved event horizons, so that the natural outcome of a gravitational collapse under physically realistic conditions would always be the formation of a naked singularity. In short it seemed that the cosmic censorship conjecture was utterly erroneous, and that our work on black hole theory had been, from a physical point of view, rather a waste of time. This state of alarm did not last very long, however, since in the discussions that followed, in which Roger Penrose took the leading part, it soon became clear that there was an almost diametrically opposite and far more plausible alternative interpretation which could be made. This was that far from being an essentially unstable phenomenon, the formation of a black hole is actually a stabilizing effect, which enables an object collapsing from one of a very wide range of initial configurations, not only to form a well behaved event horizon, but also to settle down (as far as the part of space-time outside the event horizon is concerned) towards one of a very restricted range of possible stationary equilibrium states, the excess multipoles being lost in the form of gravitational radiation. Thus on this interpretation, an uncharged non-rotating body (if it collapses in a part of space sufficiently far from external perturbing influence) should give rise ultimately to a Schwarzschild black hole. The subsequent investigation of the decay of perturbations in a Schwarzschild background space by Vishveshwara, Price, and others, provides strong support for this way of viewing the situation, which is now almost universally accepted as correct, at least for cases where the initial deviations from spherical symmetry are not too large.

Having reached this point of view, it was natural to conjecture that the only possible pure vacuum exterior solutions for a *rotating black hole* were those which were already known, that is to say the Kerr solutions, which until then we had thought of as merely simple examples from a potentially much wider class. The idea that the multipole moments could settle down to preordained values by gravitational radiation made it seem particularly natural to conjecture that in the pure vacuum case the only multipole moments which would remain as independent degrees of freedom would be those for which there were no corresponding degrees of freedom in the asymptotic gravitational radiation field, that is to say the monopole and dipole moments, i.e. the moments corresponding to mass and angular momentum (which cannot be radiated away directly, although they can of course be varied indirectly in consequence of the non-linearity of the field equations). These are in fact just the two degrees of freedom which actually are possessed by the Kerr solutions. It is not yet known for certain whether the Kerr solutions are absolutely the only possible stationary black hole exterior solutions, but it can be established conclusively (using appropriate global conditions, and with the aid of Israel's theorem and the strong rigidity theorem of Stephen Hawking, which in conjunction make it possible to take axisymmetry for granted) that all solutions must indeed fall into discrete classes within which there are at most two degrees of freedom, as the intuitive physical reasoning described above would suggest. (It is this property of having no degrees of freedom except those corresponding to multipoles which can not be directly radiated away which Wheeler has described by the statement that "a black hole has no hair".) It can also be shown that in the unlikely event that there does exist any other family than the Kerr family, it must be somewhat anomalous, in so much as the range of variation of the angular momentum within the family cannot include zero.

In addition to the strong rigidity theorem, Stephen Hawking has also derived a very valuable theorem which he will describe in the accompanying lecture course, stating that the constant time sections of stationary black hole boundaries must be topologically spherical. Using these two theorems as the starting point and making the further highly plausible but not yet rigorously justified assumption that there cannot be more than one topologically spherical black hole component in an asymptotically flat pure vacuum stationary exterior space, I shall give a fairly complete derivation of the basic no-hair theorem whose content I have just described. I shall also demonstrate, subject to the further *assumption* of axisymmetry that the Schwarzschild black holes are the only ones with stationary pure vacuum exteriors and with zero angular momentum; I hope that the treatment given here will also be useful as an introduction to the study of the full scale Israel theorem, in which axisymmetry is not assumed. Israel's original version of the theorem (1967) contained an alternative simplifying assumption which has since been removed, at considerable cost in technical complexity, by Muller Zum Hagen, Robinson, and Siefert (1972). Both these versions of the theorem are incomplete in that they assume various boundary conditions on what is presumed to be the surface of the black hole without reference to a proper global definition of the black hole event horizon; it should however be clear from the present treatment that these boundary conditions can in fact be justified without too much difficulty by means of the lemma described in section 4.

The net effect of the mathematical results which I have just outlined is to make it virtually certain that for practical applications a stationary pure vacuum black hole can be taken to be a Kerr black hole. However it is by no means certain that the whole range of Kerr black hole states is physically attainable. The recent work by Chandrasekhar (1969), (1970), and Bardeen (1971) on equilibrium configurations for a self gravitating liquid in General relativity, suggests by analogy the likelihood that although the Kerr solutions are almost certainly stable for sufficiently small values of the dimensionless parameter J^2/M^4, (where M is the mass and J the angular momentum in natural units) there may be an eigenvalue of this parameter (below the cosmic censorship limit $J^2/M^4 = 1$) above which they would become unstable. The existence of such an eigenvalue might be revealed in a purely stationary analysis by the presence of a first order perturbation solution bifurcating from the Kerr solutions at the corresponding value of J^2/M^4. I have shown in the proof of the no-hair theorem that there is no axisymmetric first order bifurcating solution, but the theorem of Hawking, which excludes the existence of *exact* non-axisymmetric perturbations, does not rule out the possibility of first order non-axisymmetric perturbations. This suggests that if there is an instability in the higher angular momentum Kerr black holes it is likely that it will lead to the formation of *non*-axisymmetric deformations. If the cosmic censorship hypothesis is correct the deformed black hole would presumably get rid of its excess angular momentum by gravitational radiation, and ultimately settle down towards one of the stable Kerr black holes below the critical eigenvalue. The recent developments in dynamic perturbation theory, in which very diverse techniques have been introduced by groups such as Chandrasekhar and Friedman, Hawking and Hartle, and Press and Teukolsky, encourage the hope that the question of the stability of the upper part of the Kerr black hole sequence will be solved conclusively, one way or the other, in the not too distant future. Although it appears that the most exciting future development in black hole theory will be concerned with dynamic aspects, there remains a great deal to be done in stationary black hole theory, particularly in relation with non-vacuum black holes. To start with there is the question of how the results would be affected by the presence of an electromagnetic field. I have been able to allow for the presence of such a field in all the results described in sections 2 to 11, but although Israel was able by an impressive tour de force (Israel 1968) to generalize his theorem to cover the electromagnetic case, I have not yet been able to do the same for the no-hair-theorem (given in section 12). One would expect (as it is discussed in section 13) that a stationary black hole solution of the source-free Einstein–Maxwell equations would have just two extra degrees of freedom, (in addition to the mass and angular momentum) which would correspond to the conserved electric and magnetic monopole charges. In particular this means that an uncharged black hole should have no magnetic dipole moment.) While no one has yet been able to exclude the existence of a solution family branching off from the Kerr-Newman solution family (which has just the four degrees of freedom one would expect), it is not so difficult (as has been shown by Wald, and also independently by Ipser) to prove that no such bifurcating solution family can start off from the pure vacuum Kerr subfamily.

For practical astrophysical applications the question of electromagnetic generalizations is probably not of very great importance, since magnetic monopoles appear not to exist, and due to the fact that charges can be carried by electrons and positrons (which are very light compared with the baryons which normally give the principal contribution to the mass), it is very easy for electric charges to be neutralized by ionization or pair creation processes, so that in an actual collapse situation it is hard to see how any excess charge carried down into the black hole could ever be more than a small fraction of a per cent (in natural units) of the mass. A potentially more important generalization is to the case where the black hole is distorted from the simple Kerr form by the presence of some sort of fluid. In the case of a black hole formed from the collapse of an ordinary star such effects would probably be negligible, since in order to produce significant deviation effects it is necessary that the perturbing mass be not too small compared with the mass of the hole, and also that it be within a distance not too large compared with the Schwarzschild radius of the hole. In other words the perturbing material must be condensed to a density not too far below that determined by the dimensions of the hole. In the case of an ordinary star sized black hole this could only arise if the perturbing material itself consisted of a neutron star or another black hole in a close binary configuration, and due to its lack of overall axisymmetry such a system would lost energy rapidly by emission of gravitational radiation, and could survive for at most a matter of seconds. However it seems much more likely that a supermassive black hole of the kind whose existence in galactic nuclei was originally suggested by Lynden-Bell (1969) could deviate significantly from the Kerr form, at least during an initial active period, lasting perhaps for a million years or so after its formation. The sort of object I have in mind would be an axisymmetric system consisting of a central black hole of something like 10^9 solar masses, surrounded at a distance of a few Schwarzschild radii, by an equally massive cloud consisting of 10^9 ordinary individual stellar mass Kerr black holes. To minimize the rate of collisions (which would in any case be extremely small in terms of timescales comparable with the several hour long orbital periods) one could make the further not implausible supposition that the small Kerr black holes move in fairly coherent orbits in a Saturn-like disc configuration. Being out of phase, the gravitational radiation contributions of the individual Kerr black holes would almost completely cancel out, and as far as calculations of the large scale properties of the central non-Kerr black hole were concerned, the cloud of small black holes could be treated approximately as a (suitably non-isotropic) fluid. Except in the final sections 10 to 13 which are specialized to the vacuum cases, all the results which follow will be valid for general stationary but non-Kerr black hole configurations of the kind which I have just described. A more detailed treatment of many aspects of non-vacuum black hole theory is contained in the following course given by Jim Bardeen.

2 The Domain of Outer Communications and the Global Horizon

Throughout this course we shall be dealing with a space-time manifold $\mathscr{M}$ which is *asymptotically flat* and *pseudo-stationary*. By *asymptotically flat* we mean that $\mathscr{M}$ is weakly asymptotically simple, in the sense of Penrose (1967), and by *pseudo-stationary* we mean that $\mathscr{M}$ is invariant under a one-parameter isometry group action π^s:R(1) x $\mathscr{M}\rightarrow\mathscr{M}$ of the one-parameter translation group R(1), and that the trajectories of this action π^s (i.e. the sets of points $\{x \in \mathscr{M}: x = \pi^s(t, x_0), t \in R(1)\}$ for fixed $x_0 \in \mathscr{M}$) are *timelike* curves at least at sufficiently large asymptotic distances. If the trajectories of π^s were timelike curves everywhere, $\mathscr{M}$ would be said to be stationary in the strict sense. The maximal connected asymptotically flat subdomain $\mathscr{S}$ of $\mathscr{M}$ that is stationary in this strict sense will be referred to as the *outer stationary domain* of $\mathscr{M}$.

Our attention in this course will be almost exclusively restricted to the *domain of outer communications* $\ll\mathscr{I}\gg$ of $\mathscr{M}$ which is defined as the set of points from which there exist both future and past directed timelike curves extending to arbitrarily large asymptotic distances. In the terminology of Penrose (1967), $\ll\mathscr{I}\gg$ is the intersection of the chronological past of $\mathscr{I}^+$ with the chronological future of $\mathscr{I}^-$. It is evident that $\ll\mathscr{I}\gg$ (like $\mathscr{S}$) is an open subset of $\mathscr{M}$, and can therefore be regarded as an asymptotically flat space-time manifold in its own right. Clearly the boundary $\ll\mathscr{I}\gg^\bullet$ of $\ll\mathscr{I}\gg$ in $\mathscr{M}$, which we shall denote more briefly by $\mathscr{H}$, can be thought of as a union of the form $\mathscr{H} = \mathscr{H}^+ \cup \mathscr{H}^-$, where the subset of points $\mathscr{H}^+$ forms part of the boundary of the chronological past of $\mathscr{I}^+$, (i.e. the boundary of the subdomain of $\mathscr{M}$ from which there are future directed timelike curves extending to arbitrarily large asymptotic distances) and where similarly the subset $\mathscr{H}^-$ forms part of the boundary of the chronological future of $\mathscr{I}^-$. Using the lemmas of Penrose and Hawking (Penrose (1965), Hawking (1966)) it is easy to see that $\mathscr{H}^+$ must be a *global past horizon* in the sense (Carter 1971) that it is an achronal hypersurface (i.e. a not necessarily differentiable hypersurface with the property that no two points on it can be connected by a timelike curve) such that from any point on it there is a null geodesic which can be extended without bound towards the future (but not necessarily towards the past) lying entirely in $\mathscr{H}^+$. Similarly $\mathscr{H}^-$ is a future global horizon.

These definitions and properties of $\ll\mathscr{I}\gg$, $\mathscr{H}^+$ and $\mathscr{H}^-$ do not depend in any way on the pseudo-stationary property. In a dynamic collapse starting from well behaved intial conditions $\mathscr{H}^-$ would not exist, and $\mathscr{H}^+$ could be regarded as being either the *past* boundary of the *black hole* (i.e. the region of space-time outside $\ll\mathscr{I}\gg$ or equivalently as the *future* boundary of the domain of outer communications (see Figure 1). The example of spherical collapse shows however that one cannot expect that the corresponding pseudo-stationary limit space will, when extrapolated back in time by the group action, have well behaved initial conditions, although one can at least demand that the closure of the domain of outer communications should be geodesically complete (in the sense that any inextensible geodesic of finite length will have an end point on the boundary) if a horizon $\mathscr{H}^-$ is allowed for in $\mathscr{M}$. The region lying to the future of $\mathscr{H}^+$ can still be quite appropriately described as a *black hole* since, by its definition, no light from within it can escape to an external observer. The analogous region to the past of $\mathscr{H}^-$ (if it exists), which has the property that it can never be reached *by* any signal or probe from outside, has been referred to as a white hole—a rather misleading term, since light of any colour might in principle emerge from it. Personally I should prefer to refer to any such region (i.e. a region lying outside the future of $\mathscr{I}^-$) as a *primordial hole.* As far as the present course is concerned it will not be necessary to worry about these distinctions, and I shall simply refer to the whole of the region outside the domain of outer communications as a *hole*† without further qualification. Since this work is motivated by the problem of gravitational collapse, our attention will be entirely devoted to the domain of outer communications $\ll\mathscr{I}\gg$ and its future boundary, the horizon $\mathscr{H}^+$. It will at no stage be of any particular importance whether the horizon $\mathscr{H}^-$ exists, nor what its properties may be if it does.

Since the domain of outer communications and its bounding horizon $\mathscr{H}^+$ are globally defined subsets, it is not always easy to ascertain their position in relation

† A consensus at this school has agreed that the most appropriate French translation of the term black hole is "piège noir". For more general purposes it has been suggested that the term hole should be translated as "poche"; thus primordial hole would be translated as "poche primordiale".

to locally defined structures on $\mathcal{M}$. Under favourable conditions however it is possible to determine the position of $\mathcal{H}^+$ locally, in relation to the symmetry group structure, as will be shown in the next section. In establishing such relationships, I have found the following lemma, which gives an alternative global characterization of the domain of outer communications, extremely useful.

LEMMA 2 $\langle\!\langle \mathscr{I} \rangle\!\rangle$ is the maximal connected asymptotically flat domain of $\mathcal{M}$ with the property that the trajectory $\pi^s(x_0)$ of the pseudo-stationary action π^s through any point x_0 in the domain will, if extended sufficiently far in the forward direction, enter and remain in the chronological future of x_0 (i.e. in the set of points which can be reached by a future directed time-like curve from x_0).

Proof The proof consists of two parts: Part I in which it is shown that any connected asymptotically flat domain $\mathscr{D}$ of $\mathcal{M}$, with the property that the trajectory $\pi^s(x_0)$ of π^s through any point $x_0 \in \mathscr{D}$ will enter the chronological future of x_0 if extended sufficiently far in the forward direction, must lie within $\langle\!\langle \mathscr{I} \rangle\!\rangle$; and Part II in which it is shown that the trajectory of π^s through any point x_0 of $\langle\!\langle \mathscr{I} \rangle\!\rangle$ will enter and remain in the chronological future of x_0 if extended sufficiently far in the forward direction.

Part I: The required result follows at once from the fact that any point $x_0 \in \mathscr{D}$ can be connected to some point y' on the trajectory $\pi^s(y)$ of the action π^s through any point $y \in \mathscr{D}$ by a future directed timelike curve λ in $\mathscr{D}$. The existence of λ is established as follows. Since $\mathscr{D}$ is connected there certainly exists some curve γ from x_0 to y. By the defining property of $\mathscr{D}$, there must exist some point x_0' on the trajectory $\pi^s(x_0)$ such that x_0 lies in the chronological future of x_0'. Therefore any point x_1 on γ sufficiently close to x_0 will lie on some future directed timelike curve λ_1 from x_0'. Since γ is compact, it is possible by repeating this process to obtain a finite sequence of points x_i on γ starting with x_0 and ending with y, such that each point x_i lies on a future directed timelike curve λ_i which starts from some point x_{i-1}' on the trajectory $\pi^s(x_{i-1})$ of π^s through the preceding point x_{i-1} of the sequence. By transporting the timelike segments λ_i suitably under the group action π^s one can clearly construct a sequence of image segments which connect up end to end to form a future directed timelike curve λ from x_0 to some point on $\pi^s(y)$.

Part II: It is evident that a possible subdomain $\mathscr{D}$, with the properties specified above, is the outer stationary domain $\mathscr{S}$. It is clear from the definition of $\langle\!\langle \mathscr{I} \rangle\!\rangle$ that any point x_0 in $\langle\!\langle \mathscr{I} \rangle\!\rangle$ can be connected to some point x_1 in $\mathscr{S}$ by a future directed timelike curve λ_1, and that x_0 can be reached from some other point x_2 in $\mathscr{S}$ by a future directed timelike curve λ_3. Moreover by the properties of $\mathscr{D}$ discussed above, there must exist a future directed timelike curve λ_2 from x_1 to some point x_2' in the trajectory $\pi^s(x_2)$ of π^s through x_2. Therefore by transporting λ_3 suitably under the group action π^s, it is possible to construct an image segment λ_3' which links up with λ_1 and λ_2 to form a timelike curve λ from x_0 to some point x_0' on $\pi^s(x_0)$ via x_1 and x_2'. This establishes that $\pi^s(x_0)$ enters the chronological future of x_0. $\pi^S(x_0)$ may of course leave the future of x_0 after remaining within it only for a limited group parameter distance, but in this case (by repetition of the preceding argument) it will re-enter after a finite distance, and at the second entry must remain in the chronological future of x_0 for at least twice as long (as measured by the group parameter) as on the first occasion. It is easy to see that after at most a finite number of departures and re-entries, a point will be reached beyond which will never leave the chronological future of x_0. This completes the proof of the lemma.

In most of the work of this course, we shall need to employ the following postulate:

CAUSALITY AXIOM There are no compact (i.e. to be more explicit, topologically circular) causal (i.e. everywhere timelike or null) curves in $\mathscr{M}$.

This is almost the weakest possible global causality condition. Its necessity for any reasonably well behaved physical situation can hardly be question. The lemma which has just been established does not itself depend on the causality condition, but most of its applications do. An example is the following immediate corollary:

COROLLARY If the causality axiom holds in $\mathscr{M}$, then any degenerate trajectory of π^S (i.e. any fixed point of the action) and also any topologically circular trajectory of π^S, must lie outside the domain of outer communications $\ll\mathscr{I}\gg$.

Fixed points and circular trajectories *can* however lie on the boundary $\mathscr{H}$ of $\ll\mathscr{I}\gg$–indeed this is where one would normally expect them to turn up, as is shown by the example of the Kerr black holes. The causality axiom is satisfied in a Kerr black hole solution if, in defining $\mathscr{M}$, we exclude the region within the inner horizon at $r = r_-$. In this case there are in fact circular trajectories of the pseudo-stationary action on the Boyer–Kruskal axis $\mathscr{H}^+ \cap \mathscr{H}^-$, and also fixed points of the pseudo-stationary action where $\mathscr{H}^+ \cap \mathscr{H}^-$ intersects the rotation axis. In the more special case of a Schwarzschild black hole, $\mathscr{H}^+ \cap \mathscr{H}^-$ consists entirely of fixed points of the pseudo-stationary action.

The lemma which we have just proved can also be used, again in conjunction with the causality axiom, to show that the domain of outer communications of a pseudo-stationary space-time manifold is a fibre bundle over a well behaved three-dimensional base manifold, the fibres being the trajectories of the group action (Carter 1972b).

3 Axisymmetry and the Canonical Killing Vectors

Whenever it is convenient to do so, I shall suppose that the space-time manifold $\mathscr{M}$ under consideration is axisymmetric, i.e. that it is invariant under an action $\pi^A : SO(2) \times \mathscr{M} \rightarrow \mathscr{M}$ of the one-parameter rotation group SO(2), in addition to being invariant under the pseudo-stationary action π^S of R(1). In fact apart from the discussion of static ergosurfaces in the present section, the only stage at which

I shall *not* assume axisymmetry will be in the derivation of the Hawking–Lichnerowicz theorem (which would not in any way be simplified by assuming axisymmetry).

The original justification, (when the work described here was first undertaken) for postulating axisymmetry was simply that it is difficult to imagine any realistic astrophysical system to which the theory might be applied which could, as a good approximation be treated as pseudo-stationary without being able, as an equally good approximation, to be treated as axisymmetric (the idea being that any non-axisymmetric bulge would probably need to rotate to support itself, thereby radiating gravitationally). It is now possible to justify the axisymmetry condition as a mathematical necessity, at least in the case of *rotating* black holes, by appeal to Hawking's strong-rigidity theorem according to which, subject to very weak and general assumptions, which are described by Hawking in the accompanying notes, there exists an isometry group action $\pi^{s\dagger}$ of R(1) on $\mathscr{M}$ with the property that the null geodesic generators of $\mathscr{H}^+$ are trajectories of $\pi^{s\dagger}$. The black hole is said to be *non-rotating* if $\pi^{s\dagger}$ is the same (up to a scale factor) as π^s, and otherwise it is described as *rotating*. On account of the restrictions imposed by asymptotic flatness, it may be taken for granted that the axisymmetry action π^A whose existence I have postulated commutes with the pseudo-stationary action π^s (see Carter 1970; I should point out that in this reference I ought to have stated that the vector fields referred to in the statement of Proposition 5 and in the proof of Proposition 6 must be complete). Hawking has shown directly that the action $\pi^{s\dagger}$ predicted by his theorem commutes with π^s and that in the rotating case the two together generate a two parameter Abelian isometry group action π^{sA} with a subgroup action π^A of SO(2). In the non-rotating case it is possible to establish axisymmetry—and indeed spherical symmetry—by a rather different method provided that $\ll\mathscr{I}\gg$ is a pure or electromagnetic vacuum, by using first the generalized Hawking–Lichnerowicz theorem to be described in section 6, and then the theorems of Israel (1967), (1968). In the case of a non-rotating hole with external matter present, axisymmetry is *not* a mathematical necessity, but even in this case it seems unlikely that there would be any natural astrophysical applications which would not be axisymmetric in practice.

Even when $\mathscr{M}$ has a many parameter isometry group, resulting from axial or even spherical symmetry, there will still be only one unique pseudo-stationary subgroup action π^s (up to a scale factor), since any other subgroup action will have spacelike trajectories at large distances. We shall denote the Killing vector generator of π^s by

$$\frac{\partial}{\partial t} = k^a \frac{\partial}{\partial x^a} \tag{3.1}$$

where the x^a ($a = 0, 1, 2, 3$) are general local co-ordinates, and where t is a group parameter along the trajectories. (At a later stage we shall impose further restrictions in order that t shall be well defined as a canonical co-ordinate function.) We shall fix the scale factor by imposing the standard normalization condition, which consists of the requirement that the squared magnitude scalar

$$V = -k^a k_a \tag{3.2}$$

should satisfy

$$V \to 1 \tag{3.3}$$

in the asymptotic limit at large distances. Subject to this requirement, the vector k^a is uniquely determined. It will of course satisfy the Killing equations

$$k_{a;b} = k_{[a;b]} \tag{3.4}$$

Except in the spherically symmetric case, in which a choice must be exercised, the axisymmetry action π^A will also be uniquely determined, (up to a scale factor) since other one-parameter subgroup actions will in general have non-compact trajectories. We shall denote its generator by

$$\frac{\partial}{\partial\varphi} = m^a \frac{\partial}{\partial x^a} \tag{3.5}$$

where φ is a group parameter along the trajectories. The vector field m^a will be zero on a necessarily timelike two dimensional *rotation axis.* We shall fix the scale factor by imposing the standard normalization condition, which consists of the requirement that the squared magnitude scalar

$$X = m^a m_a \tag{3.6}$$

should satisfy

$$\frac{X_{,a}X^{,a}}{4X} \to 1 \tag{3.7}$$

in the limit on the rotation axis. This ensures that the group parameter φ has the standard periodicity 2π. Subject to this requirement m^a is uniquely determined in the spherical case. It will of course satisfy the Killing equations

$$m_{a;b} = m_{[a;b]} \tag{3.8}$$

We have already remarked that the actions π^S and π^A must commute, meaning that together they generate a 2-parameter action $\pi^{SA} = \pi^S \oplus \pi^A$ of the group R(1) x SO(2), defined by

$$\pi^{SA}(t, \varphi, x) = \pi^S(t, \pi^A(\varphi, x)) = \pi^A(\varphi, \pi^S(t, x))$$

for any $t \in \text{R}(1), \varphi \in \text{SO}(2), x \in \mathscr{M}$. It follows that the generators k^a and m^a satisfy the local commutation condition

$$m^a{}_{;b}k^b = k^a{}_{;b}m^b \tag{3.9}$$

The surfaces of transitivity of the action π^{sA} i.e. the sets of points

$$\{x \in \mathscr{M}: x = \pi^{sA}(t, \varphi, x_0), t \in R(1), \varphi \in SO(2)\}$$

will clearly be timelike in $\mathscr{S}$. We shall define the *outer stationary axisymmetric domain* $\mathscr{W}$ of $\mathscr{M}$ as the maximal asymptotically flat sub-domain of $\mathscr{M}$ in which these surfaces of transitivity are timelike, or equivalently in which the Killing bivector

$$\rho_{ab} = 2k_{[a}m_{b]} \tag{3.10}$$

is timelike. Thus $\mathscr{S}$ is the maximal connected asymptotically flat region in which $V > 0$ and $\mathscr{W}$ is the maximal connected asymptotically flat region in which $\sigma > 0$, where we introduce the notation

$$\sigma = -\tfrac{1}{2}\rho_{ab}\rho^{ab} = VX + W^2 \tag{3.11}$$

where

$$W = k^a m_a \tag{3.12}$$

We note that

$$W = X = \sigma = 0 \tag{3.13}$$

on the rotation axis, where m^a is zero. By a trivial application of the lemma of the previous section it is clear that $\mathscr{S}$ must lie within $\ll\mathscr{I}\gg$. By considering the locally intrinsically flat geometry of one of the cylindrical timelike 2-surfaces of transitivity of the action π^{sA} in $\mathscr{W}$, it is clear that any trajectory of the action through a point x_0 in the 2-cylinder must ultimately enter the chronological future of x_0 defined in relation to the intrinsic geometry on the 2-cylinder, and must hence, *a fortiori*, enter the chronological future of x_0 in the 4-dimensional space-time geometry of $\mathscr{M}$. Therefore by application of the lemma of the previous section we deduce that $\mathscr{W}$ must also lie in $\ll\mathscr{I}\gg$. Since we have already remarked that $\mathscr{S}$ must lie within $\mathscr{W}$ we see therefore that we must have

$$\mathscr{S} \subseteq \mathscr{W} \subseteq \ll\mathscr{I}\gg \tag{3.14}$$

whenever the relevant domains are defined.

We conclude this section by noting that while $\mathscr{S}$ is characterized by $V > 0$, and $\mathscr{W}$ is characterized by $\sigma > 0$, the *whole* of $\mathscr{M}$ except for the rotation axis on which $X = W = \sigma = 0$ must be characterized by

$$X > 0. \tag{3.15}$$

if the causality axiom holds.

4 Ergosurfaces, Rotosurfaces and the Horizon

In any space-time manifold with a pseudo-stationary isometry group action generated by a Killing vector k^a, the energy E of a particle moving on a timelike geodesic orbit is a constant given by $E = -mk^a v_a$ where m is the rest mass of the particle and v^a is the future oriented unit tangent vector of the orbit. It is evident that this energy must always be positive within the outer stationary domain $\mathscr{S}$. More generally E must always have a definite sign (depending on the time-orientation of k^a) in any region in which $V = -k^a k_a$ is positive, while on the other hand it may have either sign at any point where k^a is spacelike. We shall refer to any locus on which V is zero, i.e. any boundary separating regions where k^a is timelike from regions where it is spacelike, as *an ergosurface*, and more particularly, we shall refer to the boundary $\mathscr{S}^\bullet$ of $\mathscr{S}$ as *the outer ergosurface.*

When $\mathscr{M}$ is axisymmetric as well as pseudo-stationary we shall in an analogous manner refer to a locus on which σ is zero as a *rotosurface*, and more particularly we shall refer to the boundary $\mathscr{W}^\bullet$ of the outer stationary-axisymmetric domain $\mathscr{W}$ as the *outer rotosurface.*

Another way of thinking of these surfaces is to regard $\mathscr{S}^\bullet$ as a *staticity* limit, i.e. the boundary of the outer connected region within which it is kinematically possible for a timelike particle orbit to satisfy the condition of *staticity* i.e. for its tangent vector v^a to be parallel to k^a; similarly $\mathscr{W}^\bullet$ can be thought of as a *circularity* limit i.e. the boundary of the outer connected region within which it is kinematically possible for a timelike orbit to satisfy the condition of *circularity,* i.e. for its tangent vector to be a linear combination of k^a and m^a so that the orbit represents a uniform circular motion. [The ergosurface $\mathscr{S}^\bullet$ has sometimes been referred to in the literature as an "infinite red-shift surface", but this is misleading since the only physical particles whose observed light will suffer a genuine infinite red shift are those which cross the black hole horizon $\mathscr{H}^+$ itself.]

In consequence of (3.14) we see that the rotosurface $\mathscr{W}^\bullet$—when it is defined—must always lie outside or on the hole boundary $\mathscr{H}$ and similarly that the ergosurface $\mathscr{S}^\bullet$ must always lie outside or on both $\mathscr{W}^\bullet$ and $\mathscr{H}$. The main objective of this section is to show that there are natural conditions under which either $\mathscr{S}^\bullet$ or $\mathscr{W}^\bullet$ (or both) must actually coincide with $\mathscr{H}$. This will be of great value when we come on to the study of black hole uniqueness problems, since (being globally defined) the hole boundary may in general be difficult to locate without integrating the null geodesic equations. When $\mathscr{H}$ can be identified with $\mathscr{S}^\bullet$ or $\mathscr{W}^\bullet$ whose positions are locally determined in terms of the symmetry group structure, what would be an extremely intractable integro-partial differential problem reduces to a relatively straightforward (albeit non-linear) partial differential equation boundary problem.

The situations under which this simplification occurs, arise when either the pseudo stationary action π^S or the pseudo-stationary axisymmetry action $\pi^S \oplus \pi^A$

is *orthogonally transitive* meaning that the surface of transitivity (which will have dimensionality p equal respectively to 1 or 2 everywhere except on loci of degeneracy) are orthogonal to a family of surfaces of the conjugate dimension (i.e. of dimension $n - p$ which will be respectively 3 or 2). In the former case, i.e. when π^s is orthogonally transitive, the geometry is said to be *static*. In the latter case, i.e. when π^{sA} is orthogonally transitive, I shall describe the geometry as *circular* for reasons which will be made clear in section 7. By Frobenius theorem the necessary and sufficient condition for the geometry to be static in $\mathscr{S}$ (i.e. for the Killing vector k^a to be orthogonal to a family of hypersurfaces) is

$$k_{[a;b}k_{c]} = 0 \tag{4.1}$$

and similarly the necessary and sufficient condition for the geometry to be circular in $\mathscr{W}$, (i.e. for the Killing bivector ρ_{ab} to be orthogonal to a family of 2-surfaces) is

$$\begin{aligned} k_{[a;b}\rho_{cd]} &= 0 \\ m_{[a;b}\rho_{cd]} &= 0 \end{aligned} \tag{4.2}$$

It will be made clear in sections 6 and 7 that these conditions of staticity and circularity are not imposed gratuitously but that as a consequence of the Hawking strong rigidity theorem one or other can be expected to hold in the exterior of a stationary black hole under practically any naturally occurring conditions.

The purpose of the present section is to show that subject to suitable global conditions, the ergosurface (or staticity limit) $\mathscr{S}^\bullet$ coincides with $\mathscr{H}$ in the static case, while the rotosurface (or circularity limit) $\mathscr{W}^\bullet$ coincides with $\mathscr{H}$ in the circular case. The demonstration depends on the two following lemmas, of which the first was originally given (independently) by Vishveshwara (1968) and Carter (1969) and the second was originally given by Carter (1969).

LEMMA 4.1 If the staticity condition (4.1) is satisfied in $\mathscr{S}$, then the boundary $\mathscr{S}^\bullet$ consists of null hypersurface segments except at points of degeneracy of the action π^s (i.e. where k^a is zero).

LEMMA 4.2 If the circularity conditions (4.2) is satisfied in $\mathscr{W}$, then the boundary $\mathscr{W}^\bullet$ consists of null hypersurface segments except at points of degeneracy of the action $\pi^s \oplus \pi^A$ (i.e. where ρ^{ab} is zero).

Proof of Lemma 4.1 We start by using the Killing antisymmetry condition 3.3 to convert the Frobenius orthogonality condition 4.1 to the form

$$2k_{a;[b}k_{c]} = k_a k_{[b;c]} \tag{4.3}$$

from which, on contracting with k^a and using the definition (3.2) of V, we obtain

$$V_{,[b}k_{c]} = Vk_{[b;c]} \tag{4.4}$$

This tells us immediately that on the locus $\mathscr{S}^\bullet$ on which V is zero, the gradient $V_{,b}$ is parallel to k^a and hence null there, (except at points of degeneracy where k^a is actually zero). In the *non-degenerate* case, i.e. when the gradient $V_{,b}$ is non-zero, so that $V_{,b}$ determines the direction of the normal to the hypersurface on which $V = 0$, it follows immediately that this hypersurface is a null hypersurface as required. In the degenerate case where $V_{,b}$ is zero on the locus $V = 0$ more care is required since V might be zero not just on a hypersurface but over a domain of finite measure. However by more detailed consideration of the situation (cf. Carter (1969)) it can easily be verified that it will still be true that the actual boundary $\mathscr{S}^\bullet$ must be a null hypersurface.

Proof of Lemma 4.2 By analogy with the previous case, we use the Killing antisymmetry conditions (3.4) and (3.8) to convert the Frobenius orthogonality conditions (4.2) to the form

$$\left.\begin{aligned} k_{a;[b}\rho_{cd]} &= -k_a k_{[b;c}m_{d]} + m_a k_{[b;c}k_{d]} \\ m_{a;[b}\rho_{cd]} &= m_a m_{[b;c}m_{d]} - k_a m_{[b;c}m_{d]} \end{aligned}\right\} \tag{4.5}$$

from which we can directly obtain

$$2\rho_{ae;[b}\rho_{cd]} = \rho_{ae}\rho_{[cd;b]} \tag{4.6}$$

Contracting this with the Killing bivector ρ_{ab} we obtain

$$\sigma_{,[b}\rho_{cd]} = \sigma\rho_{[cd;b]} \tag{4.7}$$

This equation is analogous to (4.4), and tells us immediately that in the *non degenerate* case, i.e. when the gradient $\sigma_{,b}$ is non-zero on the boundary $\mathscr{W}^\bullet$ where σ is zero, the normal to $\mathscr{W}^\bullet$ which is parallel to $\sigma_{,b}$ lies in the plane of the Killing bivector ρ_{ab}. This is only possible (since the normal must also be orthogonal to this bivector) if the normal is null, i.e. if $\mathscr{W}^\bullet$ is a null hypersurface. As before it is still possible with rather more care (see Carter 1969) to deduce that the boundary $\mathscr{W}^\bullet$ is a null hypersurface even in the degenerate case where $\sigma_{,b}$ is zero, (except on the lower dimensional surfaces of degeneracy of the group action, where the Killing bivector ρ_{ab} is itself zero). This completes the proof.

The two preceding lemmas may be used respectively in conjunction with the lemma of section 2, to prove the two following theorems, which are of central importance in stationary black hole theory: (these results have been given in a somewhat more general mathematical context by Carter 1972b).

THEOREM 4.1 (Static Ergosurface Theorem) Let $\mathscr{M}$ be a pseudo-stationary asymptotically flat space-time manifold with a *simply-connected* domain of outer communications $\langle\!\langle \mathscr{I} \rangle\!\rangle$. Then if the (chronological) *causality axiom* and the *staticity condition* (4.1) are satisfied in $\langle\!\langle \mathscr{I} \rangle\!\rangle$ it follows that $\langle\!\langle \mathscr{I} \rangle\!\rangle = \mathscr{S}$, where $\mathscr{S}$ is the outer stationary domain and hence that *the outer ergosurface* $\mathscr{S}^\bullet$ *coincides with the hole boundary* $\mathscr{H}$.

THEOREM 4.2 (Rotosurface Theorem) Let $\mathscr{M}$ be a pseudo-stationary axisymmetric asymptotically flat space-time manifold with a *simply connected* domain of outer communications $\ll\mathscr{I}\gg$. Then if the *causality axiom* and the *circularity condition* (4.2) are satisfied in $\ll\mathscr{I}\gg$ it follows that $\ll\mathscr{I}\gg = \mathscr{W}$ where $\mathscr{W}$ is the outer stationary-axisymmetry domain, and hence that the *outer rotosurface* $\mathscr{W}^\bullet$ *coincides with the hole boundary* $\mathscr{H}$.

Proof of Theorem 4.1 We have already noted that $\mathscr{S}$ must always be entirely contained within $\ll\mathscr{I}\gg$. Let $\mathscr{D}$ be a connected component of the complement of $\mathscr{S}$ in $\ll\mathscr{I}\gg$. Since $\mathscr{S}$ is connected (by definition) so also is the complement of $\mathscr{D}$ in $\ll\mathscr{I}\gg$. Hence by the condition that $\ll\mathscr{I}\gg$ is *simply* connected, it follows (from elementary homotopy theory) that the boundary $\mathscr{D}^\bullet$ of $\mathscr{D}$, as restricted to $\ll\mathscr{I}\gg$, is *connected.*

Now (as we remarked at the end of section 2) it follows from the causality axiom by the corollary to Lemma 2 that the action π^S can never be degenerate (i.e. k^a can never be zero) in $\ll\mathscr{I}\gg$, and hence it follows from Lemma 4.1 that the boundary $\mathscr{D}^\bullet$ of $\mathscr{D}$, as restricted to $\ll\mathscr{I}\gg$, must consist entirely of one connected null hypersurface. It follows that the outgoing normal (from $\mathscr{D}$) of this hypersurface must therefore be everywhere future directed or everywhere past directed; in the former case no future directed timelike line in $\ll\mathscr{I}\gg$ could ever enter $\mathscr{D}$ from $\mathscr{S}$, and in the latter case no future directed timelike line in $\ll\mathscr{I}\gg$ could ever enter $\mathscr{S}$ from $\mathscr{D}$. Neither alternative is compatible with the condition that $\mathscr{D}$ lies in $\ll\mathscr{I}\gg$, unless $\mathscr{D}$ is empty. This establishes the required result.

Proof of Theorem 4.2 By analogy with the previous case we note that $\mathscr{W}$ must lie entirely within $\ll\mathscr{I}\gg$, and we choose $\mathscr{D}$ to be a connected component of the complement of $\mathscr{W}$ in $\ll\mathscr{I}\gg$. As before we see that the boundary $\mathscr{D}^\bullet$ of $\mathscr{D}$ as restricted to $\ll\mathscr{I}\gg$, is connected.

We now use the causality axiom to establish that the action $\pi^S \oplus \pi^A$ is nowhere degenerate (i.e. the Killing bivector ρ_{ab} is nowhere zero) in $\ll\mathscr{I}\gg$ except on the rotation axis where m^a is zero. This follows from the fact that k^a must be parallel to m^a at any point of degeneracy, so that the trajectories of the action π^S would be circles. Any such circular trajectory of π^S must lie outside $\ll\mathscr{I}\gg$ if the causality axiom holds, by the corollary to Lemma 2.

We can now use Lemma 4.2 to deduce that the boundary $\mathscr{D}^\bullet$ of $\mathscr{D}$ as restricted to $\ll\mathscr{I}\gg$ must consist of a null hypersurface everywhere, except perhaps at points on the rotation axis where m^a is zero. We can go on to deduce that the outgoing normal to this boundary must be everywhere future directed or everywhere past directed as before, despite the possibility of degeneracy on the rotation axis, since the rotation axis must be a *timelike* 2-surface everywhere (by local geometrical considerations) and as such could never form the boundary of a null hypersurface. Hence as in the previous case we deduce that $\mathscr{D}$ must be empty as required. This completes the proof.

The mathematically important qualification that the domain of outer communications should be simply connected does not restrict the practical application of these theorems, since Stephen Hawking has given arguments (see his discussion in the accompanying lecture course) which indicate that in any reasonably well behaved gravitational collapse situation the domain of outer communications of the resulting final equilibrium state should have the topology of the product of the Euclidean line R(1) with a 3-space which has the form of a Euclidean R(3) from which a number of three dimensional balls (solid spheres) have been removed, and which is therefore *necessarily* simply connected. Except in artificially contrived situations one would expect further that the 3-space would have the form of a Euclidean 3-space from which *only one* 3-dimensional ball has been removed, i.e. that it would have the topology S(2) x R(1) where S(2) is the 2-sphere, so that the domain of outer communications as a whole would have the topology S(2) x R(2). This latter stronger condition will be imposed in the theorems of sections 10 to 13.

We shall conclude this section by proving an important corollary to Theorem 4.2, which is as follows:

COROLLARY TO THEOREM 4.2 (Rigidity Theorem) Under the conditions of Theorem 4.2, it is possible to choose the normalization of the null tangent vector l^a of $\mathscr{H}^+$ in such a way that it has the form

$$l^a = k^a + \Omega^H m^a \tag{4.8}$$

where the scalar Ω^H is a *constant* over any connected component of ${}^{+}$.

Proof It is evident from (4.2) that l^a is a linear combination of k^a and m^a on the horizon, and hence can be expressed in the form (4.8) where Ω^H is some scalar which represents the local *angular velocity* of the horizon. The non trivial part of the proof is the demonstration that Ω^H is constant, i.e. that the rotation which it determines is *rigid*, which is a consequence of the commutation condition (3.8), which was not required for the proof of the basic Theorem 4.2. Since any vector in the horizon including m^a in particular, is orthogonal to l^a we find on contracting (4.8) with m^a that Ω^H must be given by

$$X\Omega^H = -W \tag{4.9}$$

where W and X are defined by (3.12) and (3.6). (On the rotation axis, where W and X are both zero, Ω^H must be defined by a limiting process). Using the relations

$$\left.\begin{aligned} X_{,a} &= 2m^b m_{b;a} \\ W_{,a} &= 2k^b m_{b;a} \end{aligned}\right\} \tag{4.10}$$

(of which the latter is a consequence of the commutation condition (3.8)) we find by differentiation of (4.9) that the gradient of Ω^H is given by

$$X^2 \Omega^H_{,a} = 2[Wm^b - Xk^b]m_{b;a} \tag{4.11}$$

It follows, from the second of the Frobenius equations (4.2), that we must have

$$X^2\Omega^H_{,[a}\rho_{bc]} = -4\sigma m_{[a;b}m_{c]} \tag{4.12}$$

where σ is defined by (3.11). Since we have already established that σ is zero on the horizon, this gives simply

$$\Omega^H_{,[a}\rho_{bc]} = 0 \tag{4.13}$$

This implies that the gradient of the coefficient Ω^H determined by (4.9) lies in the surface of transitivity of the group action π^{sA} at the horizon, which is impossible (since Ω^H is clearly a group invariant quantity) unless the gradient is actually zero on the horizon, i.e.

$$\Omega^H_{,a} = 0 \tag{4.14}$$

which is the required result. This completes the proof.

Having established that Ω^H as defined by (4.8) is constant on the horizon, we can use this constant in (4.8) to define a vector field l^a which will satisfy the Killing equations

$$l_{a;b} = l_{[a;b]} \tag{4.15}$$

everywhere in $\mathscr{M}$. Our demonstration of the existence of this Killing vector field l^a that is null on the horizon has been based on the assumption that either the staticity condition or the axisymmetric circularity condition is satisfied. The *strong* rigidity theorem of Hawking, to which we have already referred, establishes the existence of this Killing vector field without assuming either staticity or axisymmetry, subject only to very weak and general assumptions. If the hole is *rotating,* i.e. if l^a does not coincide with k^a then Hawking has shown, as an immediate corollary to his basic theorem, that there must exist an axisymmetry action π^A generated by a Killing vector field m^a such that l^a has the form (4.8). It then follows from the generalized Papapetrou theorem which will be described in section 7 that the circularity condition which was postulated in Theorem 4.2 must necessarily hold. On the other hand if the hole is non-rotating, i.e. if the Killing vector field l^a that is null on the horizon coincides with the pseudo-stationary Killing vector field k^a, it follows from the generalized Hawking–Lichnerowicz theorem which will be described in section 6 that the staticity condition which was postulated in Theorem 4.1 must necessarily hold (at least provided one is prepared to accept the as yet not rigorously proved supposition that the ergo-surface which coincides with the horizon in the non-rotating case is the same as *the outer* ergosurface).

The net effect of the results of this section is to establish that $\mathscr{H}^+$ must be a *Killing horizon* (in the sense defined by Carter 1969) that is to say it is *a null hyper-surface whose null tangent vector coincides (when suitably normalized) with the representative of some (fixed) Killing vector field*, this field being the l^a which has just been constructed. This makes it possible to carry out a precise analysis of the boundary conditions on $\mathscr{H}$ in the manner which will be described in the next section.

5 Properties of Killing Horizons

The horizon Killing vector l^a, which generates the action $\pi^{s\dagger}$ referred to in section 3, can be thought of as being in a certain sense complementary to the pseudo stationary Killing vector k^a. We shall consistently use a dagger to denote the complementary analogue, defined in terms of l^a, of a quantity originally defined in terms of k^a. Thus in particular we define

$$V^\dagger = -l^a l_a \tag{5.1}$$

$$W^\dagger = l^a m_a \tag{5.2}$$

The quantity X is self complementary, as also is σ which can be expressed in the form

$$\sigma = V^\dagger X + W^{\dagger 2} \tag{5.3}$$

Since l^a is orthogonal to any vector in the horizon, including m^a and l^a itself, it is immediately evident that $V^\dagger$ and $W^\dagger$ and hence also (as was established directly in Theorem 4.2) σ, are zero on the horizon, i.e. we have

$$V^\dagger = W^\dagger = \sigma = 0 \tag{5.4}$$

on $\mathscr{H}$, which is closely analogous to the condition (3.13) satisfied on the rotation axis. Whereas k^a is timelike at large asymptotic distances, (i.e. near $\mathscr{I}$) becoming spacelike between the ergosurface where $V = 0$ and the horizon (except in the non-rotating case where it is timelike right up to the horizon), on the other hand l^a is timelike just outside the horizon, becoming spacelike outside the *co-ergosurface,* where $V^\dagger = 0$ (except in the non-rotating case where l^a is timelike out to arbitrarily large distances). [We remark that an analogous co-ergosurface can be defined for any rigidly rotating body as the cylindrical surface at which a co-rotating frame is moving at the speed of light; the surface so defined is of importance in the Gold theory of pulsars.]

Being orthogonal to the horizon, l^a must satisfy the Frobenius orthogonality condition

$$l_{[a;b}l_{c]} = 0 \tag{5.5}$$

on $\mathscr{H}$. It is a well known consequence of this condition that the null tangent vector of any null hypersurface must satisfy the geodesic equation. To see this one uses the fact that the squared magnitude $-V^\dagger$ of l^a is a constant (namely zero) on the horizon, so that its gradient must be orthogonal to the horizon, and hence satisfies

$$V^{\dagger}_{,[a}l_{b]} = 0 \tag{5.6}$$

on the horizon. Contracting (5.5) with l^c and using (5.6) together with the condition $V^{\dagger} = 0$, one immediately obtains the geodesic equation

$$l_{[a}l_{b];c}l^c = 0 \tag{5.7}$$

This is equivalent to the condition that there exists a scalar κ which is in fact the positive root of the equation

$$\kappa^2 = \tfrac{1}{2}l_{a;b}l^{b;a} \tag{5.8}$$

such that

$$l^a{}_{;b}l^b = \kappa l^a \tag{5.9}$$

on $\mathscr{H}^+$. (On $\mathscr{H}^-$, l^a will satisfy an equation of the same form, except that κ is replaced by $-\kappa$). The scalar κ measures the deviation from affine parametrization of the null geodesic. If v is a group parameter on one of the null geodesic generators of $\mathscr{H}^+$, i.e. a parameter such that

$$l^a = \frac{dx^a}{dv} \tag{5.10}$$

where $x^a = x^a(v)$ is the equation of the geodesic, and if v^a is a renormalized tangent vector given by

$$v^a = \frac{dx^a}{dw} = \frac{dv}{dw}\, l^a \tag{5.11}$$

where w is an affine parameter, i.e. a function of v chosen so that v^a will satisfy the simple affine geodesic equation

$$v^a{}_{;b}v^b = 0, \tag{5.12}$$

then it follows that they will be related by

$$\frac{d}{dv}\left(\ln \frac{dw}{dv}\right) = \kappa \tag{5.13}$$

[The expression (5.8) for κ is obtained by contracting (5.5) with $l^{b;a}$ and using the Killing equation (4.15) with the defining relation (5.9).]

The scalar κ defined at each point on the horizon in this way, will turn up repeatedly in different contexts throughout the remainder of this course. (It also plays an important role in the accompanying course of Stephen Hawking; in terms of the Newman-Penrose notation used by Hawking, it is given by $\kappa = \epsilon + \bar{\epsilon}$ where ϵ is one of the standard spin coefficients for l^a, defined with respect to a null-tetrad which is Lie propagated by the field l^a.)

This scalar was first discussed explicitly by Boyer in some work which was written up posthumously by Ehlers and Stachel (Boyer 1968). To start with, Boyer pointed out that the condition $\kappa = 0$ is a criterion for *degeneracy* of the horizon, in the sense in which we have previously used this description, as meaning that the gradient of σ, or in the non-rotating case, of V, is zero. To see this we use (4.15) to convert (5.9) to the form

$$V^{\dagger}_{,a} = 2\kappa l_a \tag{5.14}$$

which makes it clear that the vanishing of κ is necessary and sufficient for the gradient of $V^{\dagger}$ to be zero on the horizon, and hence for the gradient of V to be zero on the horizon in the non-rotating case, and (since $W^{\dagger}$ is always zero on the horizon) for the gradient of σ to be zero on the horizon in the rotating case.

Next, Boyer pointed out that since κ is clearly constant on each null geodesic of the horizon, i.e.

$$\kappa_{,a} l^a = 0, \tag{5.15}$$

(this being evident from the fact that l^a, in terms of which κ is completely determined by (5.8) or (5.9), is a Killing vector) it follows that (5.13) can be integrated explicitly to give the affine parameter w (with a suitable choice of scale and origin) directly in terms of the group parameter v by

$$w = e^{\kappa v} \tag{5.16}$$

except in the degenerate case, i.e. when κ is zero, in which case it will be possible simply to take $w = v$. This shows that in the non-degenerate case, the affine parameter w varies only from 0 to ∞ as the group parameter ranges from $-\infty$ to ∞, and hence that the horizon $\mathscr{H}^+$ is incomplete in the past. Moreover the vector l^a itself tends to zero, as measured against the affinely parameterized tangent vector v^a by (5.11) in the limit as w tends to 0, and hence unless the space-time manifold $\mathscr{M}$ itself is incomplete along the null geodesics of $\mathscr{H}^+$, there must exist a *fixed point* of the subgroup action $\pi^{s\dagger}$ generated by l^a at the end point, with limiting parameter value $w = 0$, of each null geodesic of $\mathscr{H}^+$, and also a continuation of each null geodesic beyond the past boundary of $\mathscr{H}^+$ to negative values of the affine parameter w. By continuity, such fixed points cannot be isolated in space-time, but must clearly form spacelike 2-sections of the null hypersurface formed by continuing the null geodesics of $\mathscr{H}^+$ past their endpoints as subgroup trajectories. (The null geodesics cannot intersect each other, since their spacelike separation must remain invariant under the subgroup action.) Such a 2-surface, being invariant under the group action, will obviously determine two null hypersurfaces which intersect on it, forming the boundaries of its past and future, these null hypersurfaces being themselves also invariant under the subgroup action. Furthermore since in the present case the spacelike 2-surface is pointwise invariant under the subgroup action, it is clear that each individual null geodesic of the two null hypersurfaces will be invariant under the subgroup action, so that *both* null hypersurfaces are Killing horizons on which $V^\dagger$ is zero. [Both Boyer himself in his original unpublished work, and also Ehlers and Stachel in their rather different edited version (Boyer 1968) made rather heavier work than necessary of the proof of this last point.] Since $\mathscr{H}^+$, from which we started this construction process, forms the part of one of these null hypersurfaces lying to the future of the fixed point 2-surface, it can be seen that the part of *the other* null hypersurface lying to *the past* of the fixed point 2-surface must also lie on the boundary of $\ll\mathscr{I}\gg$, and hence must form part of $\mathscr{H}^-$. We could have made a similar construction starting from $\mathscr{H}^-$. We thus arrive at the following conclusion:

THEOREM 5.1 (Boyer's Theorem) Let the conditions of Theorem 4.1 or Theorem 4.2 be satisfied. Then if the horizon $\mathscr{H}$ is non-degenerate, and if the closure $\overline{\ll\mathscr{I}\gg}$ of $\ll\mathscr{I}\gg$ is geodesically complete, then both $\mathscr{H}^-$ and $\mathscr{H}^+$ exist and they intersect on a spacelike 2-surface (on which the Killing vector l^a is zero) which contains a future endpoint for every null geodesic of $\mathscr{H}^-$, and a past endpoint for every null geodesic of $\mathscr{H}^+$. On the other hand if $\mathscr{H}$ is degenerate, then $\mathscr{H}^+$ contains null geodesics which can be extended to arbitrary affine distance not only towards the future (as always) but also towards the past.

Since a null geodesic on $\mathscr{H}^+$ cannot approach $\mathscr{I}$ when it is extended towards the *past* (since it lies on the boundary of the *past* of $\mathscr{I}$) it follows that *in the degenerate case* $\ll\mathscr{I}\gg$ *possesses an internal infinity, or in other words the hole is bottomless*, in the manner we became familiar with in Part I of this course, in the limiting cases with $a^2 + Q^2 + P^2 = M^2$ of the generalized Kerr solution black holes. This fact lends support to the conjecture, which is associated with the cosmic censorship hypothesis, that degenerate black holes represent physically unattainable limits, in much the same way that the speed of light represents an unattainable limit for the speed of a massive particle, or (to introduce an analogy which we shall see later can be pushed considerably further) in the same way that the absolute zero of temperature represents an unattainable limit in thermodynamics.

We shall continue this section with the description of some further important boundary condition properties of Killing horizons. Substituting l^a into the Raychaudhuri identity

$$(u^a{}_{;a})_{;b}u^b = (u^a{}_{;b}u^b)_{;a} - u^a{}_{;b}u^b{}_{;a} - R_{ab}u^a u^b \tag{5.17}$$

which is satisfied by any vector field u^a whatsoever, where R_{ab} is the Ricci tensor, and using (4.15), we obtain the identity

$$V^{\dagger;a}_{;a} = 2l^a{}_{;b}l^b{}_{;a} + 2R_{ab}l^a l^b \tag{5.18}$$

which is satisfied by any Killing vector field. Using the further identity

$$V^\dagger l^a{}_{;b}l^b{}_{;a} = \tfrac{1}{2}V^\dagger_{;a}V^{\dagger;a} - 2\omega^\dagger_a\omega^{\dagger a} \tag{5.19}$$

which also follows directly from the Killing equation, where we have introduced the rotation vector $\omega^{\dagger a}$ of the field given by

$$\omega^{\dagger}_{a} = \tfrac{1}{2}\epsilon_{abcd}l^{b}l^{d;c} \tag{5.20}$$

(where ϵ_{abcd} is the alternating tensor) (5.19) can be expressed as

$$V^{\dagger\,;a}{}_{;a} = \frac{V^{\dagger}_{;a}V^{\dagger;a} - 4\omega^{\dagger}_{a}\omega^{\dagger a}}{V^{\dagger}} + 2R_{ab}l^{a}l^{b} \tag{5.21}$$

Evaluating the various terms on the horizon, we deduce directly from (4.15) and (5.9) that the Laplacian of $V^{\dagger}$ takes the value

$$V^{\dagger\,;a}{}_{;a} = 4\kappa^{2} \tag{5.22}$$

on $\mathscr{H}$, and hence from (5.8) and (5.18) that

$$R_{ab}l^{a}l^{b} = 0 \tag{5.23}$$

on $\mathscr{H}$. Since, by (5.5) the rotation vector $\omega^{\dagger}_{a}$ is zero on $\mathscr{H}$ it is clear in the non-degenerate case, and can be verified to be true even in the degenerate limit case, that the ratio $\omega^{\dagger}_{a}\omega^{\dagger a}/V^{\dagger}$ tends to zero on $\mathscr{H}$; it therefore follows that

$$\frac{V^{\dagger}_{;a}V^{\dagger;a}}{4V^{\dagger}} \to \kappa^{2} \tag{5.24}$$

in the limit on $\mathscr{H}$. This last equation gives rise to another interpretation of κ. The acceleration a^{a} of an observer rigidly co-rotating with the hole, i.e. following one of the trajectories of the field l^{a}, is given by

$$V^{\dagger}a^{a} = l^{a}{}_{;b}l^{b} \tag{5.25}$$

Introducing the convention (which will be employed frequently in the following sections) of using a bar to denote quantities associated with orbits which have been renormalized by multiplication by the time dilatation factor of the orbit (which in this case is $V^{\dagger\,1/2}$), we define

$$\bar{a}^{a} = V^{\dagger\,1/2}a^{a} \tag{5.26}$$

In terms of this renormalized acceleration vector, we see that (5.24) has the form

$$\bar{a}^{a}\bar{a}_{a} \to \kappa^{2} \tag{5.27}$$

on $\mathscr{H}$, i.e. κ represents the limiting magnitude of the renormalized acceleration of a co-rotating observer.

Since ordinary matter will always give positive contributions to $R_{ab}l^{a}l^{b}$, the equation (5.23) implies that the immediate neighbourhood of the hole boundary must always be empty, except possibly for the presence of an electromagnetic field, which can be contrived so as to give zero contribution to the term $R_{ab}l^{a}l^{b}$. The restrictions required to satisfy this requirement are rather stringent however, and place well defined boundary conditions on the field. Introducing the Maxwell energy tensor T_F^{ab}, which will be proportional to the Ricci tensor under these conditions, by

$$T_F^{ab} = \frac{1}{4\pi}\,[F^{ac}F_c^{\ b} - \tfrac{1}{4}F^{cd}F_{cd}g^{ab}] \tag{5.28}$$

where F_{ab} is the electromagnetic field tensor, we obtain

$$T_{Fab}l^a l^b = \frac{1}{8\pi}\,[\bar{E}_a^\dagger \bar{E}^{\dagger a} + \bar{B}_a^\dagger \bar{B}^{\dagger a}] \tag{5.29}$$

where we have introduced the co-rotating electric and magnetic field vectors $\bar{E}_a^\dagger$ and $\bar{B}_a^\dagger$ defined by

$$\bar{E}_a^\dagger = F_{ab}l^b \tag{5.30}$$

$$\bar{B}_a^\dagger = \tfrac{1}{2}\epsilon_{abcd}F^{cd}l^b \tag{5.31}$$

Since they are both orthogonal to l^a, neither can be timelike within the closure of $\langle\!\langle \mathscr{I} \rangle\!\rangle$, so that (5.29) is a sum of non-negative terms. Thus $R_{ab}l^a l^b$ and $T_{Fab}l^a l^b$ can only be zero on $\mathscr{H}$ if $E_a^\dagger$ and $B_a^\dagger$ are both null on $\mathscr{H}$, and hence (by the orthogonality) parallel to l_a on $\mathscr{H}$. In other words the electromagnetic field must satisfy the boundary conditions

$$\bar{E}_{[a}^\dagger l_{b]} = 0 \tag{5.32}$$

$$\bar{B}_{[a}^\dagger l_{b]} = 0 \tag{5.33}$$

on $\mathscr{H}$.

6 Stationarity, Staticity, and the Hawking-Lichnerowicz Theorem

The terms stationary and static are used widely in different physical contexts. Generally speaking a system is said to be *stationary* if it is invariant under a continuous transformation group which maps earlier events into later ones. In General Relativity this condition can be more precisely defined in terms of the requirement that space-time be invariant under the action of a continuous one-parameter isometry group generated by a Killing vector field which is either timelike everywhere (in which case the space-time is said to be stationary in the *strict* sense) or else timelike at least somewhere, e.g. in sufficiently distant asymptotically flat region, (in which case I shall refer to the space-time as *pseudo-stationary*). It is not quite so easy to give a definition of what is meant by the statement that a general physical system is *static*, without referring to the specific context, but broadly speaking a system is said to be *static* if it is not only stationary in the strict sense, but also such that there is no motion (i.e. no material flow or current) relative to the stationary reference system. How this is to be interpreted will of course depend on which kinds of flow or current are relevant. For example a river through which water is flowing at a steady rate could be said to be stationary, but not static; on the other hand a stagnant pond might be describable as static in so far as only liquid flow was being taken into account, while being non-static from the point of view of a more detailed analysis in which perhaps the presence of a steady downward heat flux might need to be considered. In general relativity the requirement that a flow represented by a current vector field be static means that the vector field must be parallel to a timelike Killing vector.

The most fundamental generally defined flow vector in General Relativity is the timelike eigenvector of the energy-momentum tensor, which by Einstein's equations is the same as the timelike eigenvector of the Ricci Tensor. Thus we are led to the idea of a static Ricci tensor as one which satisfies the equation.

$$k^c R_{c[a} k_{b]} = 0 \tag{6.1}$$

where the timelike eigenvector k^a is a Killing vector. When electromagnetism is present another fundamental flow vector is the electric current vector j^a. The electric current vector is said to be static if it satisfies

$$j_{[a} k_{b]} = 0 \tag{6.2}$$

i.e. if it only has a component (representing a fixed electric charge density) parallel to the Killing vector.

Now although the basic physical idea of what is meant by a system being static depends essentially on the behaviour of current flows, the concept can be often extended by means of the field equations, to apply to the fields of which the relevant currents are the sources. In particular the definition of the term static can be extended to apply both to the *space-time metric tensor* g_{ab} and to the *electromagnetic field tensor* F_{ab}. The space-time metric tensor (or simply the space-time itself) is said to be static if the Killing vector trajectories are everywhere orthogonal to a family of spacelike hypersurfaces, i.e. (by Frobenius theorem) if

$$k_{[a;b} k_{c]} = 0 \tag{6.3}$$

and the electromagnetic field is said to be static if it satisfies

$$F_{[ab} k_{c]} = 0 \tag{6.4}$$

The justification for these definitions is that, by the field equations, they are not only *sufficient* for the stacity condition (6.1) and (6.2) to be satisfied by the corresponding source vectors, but also, at least under suitable global conditions, *necessary*.

To prove that (6.3) and (6.4) are respectively sufficient conditions for (6.1) and (6.2) to hold it is merely necessary to differentiate and use the appropriate field equations locally. Using the Killing equations

$$k_{a;b} = k_{[a;b]} \tag{6.5}$$

together with the definition of the Riemann tensor, by which

$$k_{a;[b;c]} = \tfrac{1}{2}R^{d}_{abc}k_d\,, \tag{6.6}$$

and using the Riemann tensor symmetries, we obtain the equation

$$k_{a;b;c} = R^{d}_{abc}k_d \tag{6.7}$$

for the second derivatives of the Killing vector, and hence, on contracting, the condition

$$k_{a;c}{}^{;c} = R_{ac}k^c \tag{6.8}$$

for the Dalembertian of the Killing vector. Thus again using (6.5), we obtain the identity

$$\{k_{[a;b}k_{c]}\}^{;c} = \tfrac{2}{3}k^c R_{c[a}k_{b]} \tag{6.9}$$

from which it is evident that the metric staticity condition (6.3) is *sufficient* for the Ricci staticity condition (6.1) to hold.

To obtain (2) from (4) we must use the antisymmetry condition

$$F_{ab} = F_{[ab]} \tag{6.10}$$

together with the condition that F_{ab} is invariant under the action generated by k_a, i.e.

$$\mathscr{L}_k\,[F_{ab}] \equiv F_{ab;c}k^c - 2F_{c[a}k^c{}_{;b]} = 0 \tag{6.11}$$

This, in conjunction with the Killing equation (6.5), leads to the identity

$$\{F_{[ab}k_{c]}\}^{;c} = \tfrac{2}{3}k_{[a}F_{b]c}{}^{;c} \tag{6.12}$$

from which, using second of the Maxwell equations

$$F_{[ab;c]} = 0 \tag{6.13}$$

$$F^{ab}{}_{;b} = 4\pi j^a \tag{6.14}$$

it is evident that the electromagnetic field staticity condition (6.4) is a sufficient condition for the current staticity condition (6.2) to hold.

If the field staticity condition (6.4) is satisfied, then it follows automatically that the Maxwell energy tensor T_F^{ab} defined by 4.12 will satisfy the staticity condition

$$k_a T_F^{a[b}k^{c]} = 0 \tag{6.15}$$

If the Einstein equations

$$R_{ab} - \tfrac{1}{2}Rg_{ab} = 8\pi T_{ab} \tag{6.16}$$

hold, with the total energy tensor T_{ab} given by

$$T^{ab} = T_F^{ab} + T_M^{ab} \tag{6.17}$$

where T_M^{ab} is the matter contribution, then the Ricci staticity condition (6.1) will be satisfied at the same time as the field staticity condition 6.4 if and only if the analogous *matter staticity condition*

$$k_a T_M^{a[b} k^{c]} = 0 \tag{6.18}$$

is satisfied. If the matter tensor is expressed in the canonical form

$$T_M^{ab} = \rho u^a u^b + p^{ab} \tag{6.19}$$

where the velocity u^a and the pressure tensor p^{ab} are required to satisfy the normalization and orthogonality conditions

$$u_a u^a = -1 \qquad p^{ab} u_b = 0 \tag{6.20}$$

and where the eigenvalue ρ represents the mass density, the staticity condition (6.18) can be seen to be equivalent to

$$u_{[a} k_{b]} = 0 \tag{6.21}$$

The basic idea of theorems of the Lichnerowicz type is to establish converses to these results, i.e. to find global conditions under which the basic flow staticity conditions (6.2) and (6.21) are *sufficient* as well as necessary for the metric and field staticity conditions to hold. It is a well known result of classical electromagnetic theory that (6.2) is a sufficient condition for (6.4) to hold in asymptotically source free flat space. The first examination of this question in General Relativity theory was made by Lichnerowicz (1939). It was shown by Lichnerowicz and Choquet-Bruhat (see Lichnerowicz (1955)) that in a space-time which is asymptotically Minkowskian and topologically Euclidean, (6.2) is a consequence of (6.1) and hence, in the non-electromagnetic case, of (2.1). When (6.31) has been established it is easy to show under the same conditions, that (6.4) is a consequence of (6.2), but the demonstration that (6.3) itself is a consequence of (6.21) and (6.2) involves certain technical difficulties which I did not succeed in resolving until a few days before the beginning of this school. I have not had time to undertake a thorough search of the literature, but as far as I know this section contains the first published account of the full electromagnetic generalization of the Lichnerowicz theorem.

The global topology conditions originally assumed by Lichnerowicz and Choquet-Bruhat were such as to explicitly exclude the presence of a central black hole. However Hawking has recently shown that the original theorem can be extended to cover the case of a *non-rotating* black hole, (i.e. one whose horizon is an ergosurface) provided it is assumed that the horizon is in fact the *outer-ergosurface.* In the present section I shall show that Hawking's result can also be extended to the case where an electromagnetic field is present, at least in the non-degenerate case. (At the time of writing I have not had time to verify that this result still holds in the degenerate limit.)

Formally the results which will be established in this section may be stated as follows:

THEOREM 6.1 (Generalized Lichnerowicz Theorem If $\mathscr{M}$ is topologically Minkowskian and both asymptotically flat, and asymptotically source free, if it is stationary in the *strict* sense and if the electromagnetic and material current staticity conditions (6.2) and (6.21) are satisfied everywhere, then the electromagnetic field staticity condition (6.4) and the metric staticity condition (6.3) will be satisfied everywhere.

THEOREM 6.2 (Generalized Hawking–Lichnerowicz Theorem) If $\mathscr{M}$ is asymptotically flat and asymptotically source free, and the asymptotic magnetic monopole moment is zero, if the domain of outer communications $\ll\mathscr{I}\gg$ is stationary in the *strict* sense (which implies by Lemma 2 that $\mathscr{H}$ is the outer ergosurface) and *simply* connected, and if the electromagnetic and material current staticity conditions (6.2) and (6.21) are satisfied everywhere, then the electromagnetic field staticity condition (6.4) and the metric staticity condition (6.3) will be satisfied in $\ll\mathscr{I}\gg$, subject (provisionally) to the requirement that the horizon be non-degenerate.

Proof of Theorem 6.1 We start by introducing the rotation vector ω^a of the stationary Killing vector k^a by the definition

$$\omega_a = \tfrac{1}{2}\epsilon_{abcd}k^b k^{d;c} \tag{6.22}$$

It is clear that the metric staticity condition (6.3) that we wish to establish, has the form $\omega_a = 0$. By differentiation of (6.22), is easy to verify using the identity (6.7) that ω_a must satisfy

$$\omega_{[a;b]} = \tfrac{1}{2}\epsilon_{abcd}k^c R^{de}k_e \tag{6.23}$$

We introduce electric and magnetic field vectors $\bar{E}^a$ and $\bar{B}^a$ defined by

$$\bar{E}_a = -F_{ab}k^b \tag{6.24}$$

$$\bar{B}_a = \tfrac{1}{2}\epsilon_{abcd}k^b F^{dc}, \tag{6.25}$$

in terms of which the electromagnetic field tensor can be given by

$$VF_{ab} = -2k_{[a}\bar{E}_{b]} + \epsilon_{abcd}k^c\bar{B}^d. \tag{6.26}$$

It is clear that the field staticity condition (6.4) that we wish to establish has the form $\bar{B}_a = 0$.

Evaluating (6.23) using the Einstein equations, we obtain

$$\omega_{[a;b]} = -2\bar{E}_{[a}\bar{B}_{b]} - 4\pi\epsilon_{abcd}k^c T_M^{de}k_e \tag{6.27}$$

From Maxwell's equations (6.13) and the field invariance condition (6.11) it is clear that we shall always have

$$\bar{E}_{[a;b]} = 0 \tag{6.28}$$

When the electric current staticity condition (6.4) is satisfied, we shall also have

$$\bar{B}_{[a;b]} = 0 \tag{6.29}$$

Since we are postulating that $\ll\mathscr{I}\gg$ (which is the same as $\mathscr{M}$ under the conditions of Theorem 6.1) is simply connected, these two equations imply the existence of globally well behaved scalars Φ and Ψ such that

$$\bar{E}_a = \Phi_{,a} \tag{6.30}$$

$$\bar{B}_a = \Psi_{,a} \tag{6.31}$$

When the matter staticity condition ((6.18) which follows from (6.21)) is also true, (6.37) reduces to

$$U_{[a;b]} = 0 \tag{6.32}$$

where we have used the abbreviation

$$U_a = \tfrac{1}{2}\omega_a + \Phi\bar{B}_a - \Psi\bar{E}_a \tag{6.33}$$

As in the case of (6.21) and (6.29), this equation implies the existence of a globally well behaved scalar U such that

$$U_a = U_{,a} \tag{6.34}$$

Now by their definition the scalars Φ, Ψ and U must satisfy the divergence identity

$$\left\{\frac{(U + 2\Phi\Psi)\omega^a}{V^2} + \frac{2\Psi\bar{B}^a}{V}\right\}_{;a} = 2\,\frac{\omega^a\omega_a}{V^2} + 2\,\frac{\bar{B}^a\bar{B}_a}{V}$$
$$+ (U + 2\Phi\Psi)\left\{\frac{\omega_a}{V^2}\right\}^{;a} + 2\Psi\left\{\left(\frac{\bar{B}_a}{V}\right)^{;a} + 2\bar{E}^a\omega_a\right\} \tag{6.35}$$

Moreover it follows directly from the definition (6.22) that ω_a must satisfy the identity

$$\left\{\frac{\omega_a}{V^2}\right\}^{;a} = 0 \tag{6.36}$$

and it follows from the Maxwell equations (6.13) that $\bar{B}^a$ must satisfy

$$\left(\frac{\bar{B}_a}{V}\right)^{;a} + 2\bar{E}^a\omega_a = 0 \tag{6.37}$$

Hence the right hand side of (6.35) reduces to a sum of two terms which are non-negative in $\ll\mathscr{I}\gg$, since under the postulated conditions V is positive and k^a (to which ω^a and $\bar{B}^a$ are orthogonal by their definitions) is timelike in $\ll\mathscr{I}\gg$.

We can convert the identity to which (6.35) reduces, from a scalar divergence equation to a vector divergence equation by multiplication by k^a, using the group invariance conditions. In this way we obtain

$$\left\{\frac{(U+2\Phi\Psi)\omega^{[a}k^{b]}}{V}+\frac{2\Psi\bar{B}^{[a}k^{b]}}{V}\right\}_{;a}=\left\{\frac{\omega^a\omega_a}{V}+\frac{\bar{B}^a\bar{B}_a}{V}\right\}k^b \tag{6.39}$$

We now construct an asymptotically flat spacelike hypersurface Σ properly imbedded in $\langle\!\langle\mathscr{I}\rangle\!\rangle$ so that its only edge consists of a 2-surface where it intersects $\mathscr{H}^+$ (under the conditions of Theorem 4.1, Σ will have no edge at all). According to Stoke's theorem, the integral of the divergence of any antisymmetric quantity V^{ab} over a hypersurface with metric normal element $d\Sigma^a$ must satisfy

$$\int_\Sigma V^{ab}{}_{;b}\,d\Sigma_a=\oint_{\partial\Sigma} V^{ab}\,dS_{ab} \tag{6.40}$$

where dS_{ab} is the 2-dimensional antisymmetric normal element to the edge $\partial\Sigma$ of Σ. Applying this to 6.39 we obtain

$$\begin{aligned}\int_\Sigma\left\{\frac{\omega^a\omega_a}{V^2}+\frac{\bar{B}^a\bar{B}_a}{V}\right\}k^a\,d\Sigma_a&=\oint_\infty dS_{ab}\left\{\frac{(U+2\Phi\Psi)\omega^{[a}k^{b]}}{V^2}+2\Psi\frac{\bar{B}^{[a}k^{b]}}{V}\right\}\\&\quad-\oint_H dS_{ab}\left\{\frac{(U+2\Phi\Psi)\omega^{[a}k^{b]}}{V^2}+2\Psi\frac{\bar{B}^{[a}k^{b]}}{V}\right\}\end{aligned} \tag{6.41}$$

where $\oint_\infty dS_{ab}$ indicates the limit of large distance of the integral over a topologically spherical 2-surface in Σ, and $\oint_H dS_{ab}$ indicates the integral over the edge of Σ where it meets $\mathscr{H}^+$.

The product $-k^a\,d\Sigma_a$ is always positive since k^a and $d\Sigma^a$ are both timelike in $\langle\!\langle\mathscr{I}\rangle\!\rangle$, and as we have already remarked, the terms $\omega^a\omega_a/V^2$ and $\bar{B}^a\bar{B}_a/V$ are non-negative in $\langle\!\langle\mathscr{I}\rangle\!\rangle$. Hence we shall obtain the required result, i.e. that the spacelike vectors ω^a and B^a must be zero everywhere in $\langle\!\langle\mathscr{I}\rangle\!\rangle$ if we can establish that the surface integral terms on the right hand side of (6.41) must be zero.

Now by the asymptotic boundary conditions, ω_a must diminish like the inverse cube, and $\bar{E}_a$ and $\bar{B}_a$ like the inverse square of the radial distance, which makes it possible to choose U, Φ and Ψ in such a way that their asymptotic limits are all zero. With this choice, U will diminish like the inverse square, and Φ and Ψ like the inverse first power of the radial distance, which will lead at once to the condition

$$\oint_\infty dS_{ab}\left\{\frac{(U+2\Phi\Psi)}{V^2}\,\omega^{[a}k^{b]}+\frac{2\Psi\bar{B}^{[a}k^{b]}}{V}\right\}=0 \tag{6.42}$$

(since by (3.3) the asymptotic limit of V is 1). This is sufficient to establish Theorem 6.1.

When a black hole is present, the situation is rather more complicated. The integration on $\mathscr{H}^+$ is rather more delicate than the integration in the asymptotic limit, since the quantity V in the denominators is zero. However the analysis is not too difficult if the degenerate case, when the gradient of V is zero, is excluded.

Since, under the conditions of Theorem 6.2, the horizon Killing vector l^a discussed in the previous section is the same as k^a, it follows from the boundary conditions (5.32) and (5.33) that $\bar{E}_a$ and $\bar{B}_a$ will be orthogonal to $\mathscr{H}^+$, and hence that the scalars Φ and Ψ must be *constant* on $\mathscr{H}^+$, which makes it possible to choose Φ and Ψ to satisfy

$$\Phi = \Psi = 0 \tag{6.43}$$

on $\mathscr{H}^+$. With this choice, it is clear at least in the non-degenerate case that Φ/V and Ψ/V will have finite limits on $\mathscr{H}^+$. Since under the conditions assumed for Theorem 6.2 k^a is null on $\mathscr{H}^+$ and hence orthogonal to $\mathscr{H}^+$, it follows that the Frobenius orthogonality condition $\omega_a = 0$ will hold at least on $\mathscr{H}^+$, which implies that subject to the choice (6.43) we shall have $U_a = 0$ on the horizon, which will make it possible to choose

$$U = 0 \tag{6.44}$$

on $\mathscr{H}^+$. With this choice it is clear (at least in the non-degenerate case) that not only U/V but even U/V^2 will have a finite limit on $\mathscr{H}^+$, so that we shall obtain

$$\oint_H dS_{ab} \left\{ \frac{(U + 2\Phi\Psi)\omega^{[a}k^{b]}}{V^2} + \frac{2\Psi\bar{B}^{[a}k^{b]}}{V} \right\} = 0 \tag{6.45}$$

This is still not sufficient to establish Theorem 6.2 since the choice $\Phi = \Psi = U = 0$ on $\mathscr{H}^+$ is not compatible with the choice of Φ, Ψ and U in such a way that their asymptotic large distance limits are zero, which was necessary in our previous derivation of (6.42). In order to obtain 6.42 at the same time as (6.45) we must use the additional assumption that the asymptotic magnetic monopole moment is zero, which means that $\bar{B}_a$ must diminish as the cube (not just the square) of the asymptotic radial distance. In this case it is clear that 6.42 will be true independently of the choice of U, Ψ or Φ. This enables us to deduce that

$$\bar{B}_a = 0 \tag{6.46}$$

$$\omega_a = 0 \tag{6.47}$$

everywhere in $\ll\mathscr{I}\gg$ as required, thus completing the proof.

COROLLARY TO THEOREMS 6.1 AND 6.2 Under the conditions of Theorem 6.1 and 6.2 the field tensor can be derived by the standard equation

$$F_{ab} = 2A_{[a;b]} \tag{6.48}$$

from a well behaved vector potential A_a in $\ll\mathscr{I}\gg$ which satisfies both the invariance condition

$$\underset{k}{\mathscr{L}}[A_a] \equiv A_{a;b}k^b + A_b m^b_{;a} = 0 \tag{6.49}$$

and also the *vector potential staticity condition*

$$A_{[a}k_{b]} = 0 \tag{6.50}$$

Proof It can easily be checked that the vector A_a given in terms of the scalar potential Φ, (whose existence has already been established) by

$$VA_a = \Phi k_a \tag{6.51}$$

satisfies all the conditions (6.48), (6.49), (6.50). (It is to be noted that it is possible to choose Φ *either* so that A_a has a finite limit on $\mathscr{H}$, or so that A_a tends asymptotically to zero at large radial distance, but not both at once.)

The Frobenius staticity condition implies (taking into account that $\ll\mathscr{I}\gg$ is required to be simply connected) that there exist a globally well behaved time co-ordinate function t on $\ll\mathscr{I}\gg$ defined up to an additive constant, in terms of which the metric $ds^2 = g_{ab}\, dx^a\, dx^b$ can be expressed in the form

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu - V\, dt^2 \tag{6.52}$$

where μ and ν run only from 1 to 3. The electromagnetic potential form $A = A_a\, dx^a$ will take the form

$$A = \Phi\, dt \tag{6.53}$$

7 Stationary-Axisymmetry, Circularity, and the Papapetrou Theorem

Just as a flow vector is said to be static if it is everywhere parallel to the stationary Killing vector k^a, so more generally it will be said to be *circular* or equivalently *non-convective* if it is everywhere parallel to a general linear combination of the two independent Killing vectors k^a and m^a, i.e. if each integral curve of the flow field lies in one of the surfaces of transitivity of the 2-parameter group generated by k^a and m^a, and therefore coincides with an integral curve of one of Killing vector fields generating the group. (If furthermore the flow vector coincides with a linear combination of k^a and m^a with everywhere *constant* coefficients, i.e. if it coincides with a *fixed* Killing vector field generator of the group, then the flow is said to be *rigid.*) Thus the *circularity* condition requires that each flow trajectory be invariant under some one-parameter transformation sub-group, which is consistent with steady circular motion relative to a non-rotating frame at infinity, whereas the stronger *staticity* condition is incompatible with any motion at all relative to such a frame. (A reasonably realistic example of a stationary-axisymmetric but *non-circular* flow would be the case of a stationary-axisymmetric star containing two convective zones, one in the northern hemisphere and one in the southern, with fluid rising towards the surface at the equator, and descending towards the centre at the poles.)

As in the previous section the basic flow vectors we shall wish to consider are the timelike eigenvector of the energy momentum tensor, or equivalently, of the Ricci tensor, and the electric current vector. When the Ricci tensor eigenvector is circular the Ricci tensor will be referred to as *invertible* (meaning that it is invariant under the tangent space isometry transformation in which directions in the plane of the Killing vectors are reversed while orthogonal directions are left unchanged). The condition for this, i.e. the necessary and sufficient condition for the timelike eigenvector of the Ricci tensor to be invertible, is that the equations

$$\begin{aligned} k^a R_{d[a} k_b m_{c]} &= 0 \\ m^d R_{d[a} k_b m_{c]} &= 0 \end{aligned} \tag{7.1}$$

be satisfied. This circularity condition will play a role in the present section analogous to that of the Ricci staticity condition (6.1) in the previous section. The corresponding circularity condition for the electric current vector, i.e. the analogue of (6.2), is simply

$$j_{[a} k_b m_{c]} = 0 \tag{7.2}$$

We can pursue the analogy with staticity by defining corresponding circularity conditions for the space-time metric tensor and the electromagnetic field tensor. The circularity condition for the metric is the condition that there should exist a family of 2-surfaces everywhere orthogonal to the surfaces of transitivity of the two-parameter group action, i.e. everywhere orthogonal to the plane of the Killing vectors k^a and m^a. This *orthogonal transitivity* condition is the precise analogue of the (one-parameter) orthogonal transitivity condition for metric staticity. By Frobenius theorem the necessary and sufficient condition for orthogonal transivity is

$$\begin{aligned} k_{[a;b} k_c m_{d]} &= 0 \\ m_{[a;b} k_c m_{d]} &= 0 \end{aligned} \tag{7.3}$$

these equations jointly being the analogue of (6.3). The circularity condition for the electromagnetic field tensor is that it should satisfy the equations

$$\begin{aligned} F_{ab} k^a m^b &= 0 \\ F_{[ab} k_c m_{d]} &= 0 \end{aligned} \tag{7.4}$$

these being jointly the analogue of (6.4).

As for the staticity conditions of previous section, so also here we justify the description of the conditions (7.3) and (7.4) as circularity conditions by showing not only that they are sufficient for the respective source flow circularity conditions (7.1) and (7.2) to hold, but also, under global conditions much weaker than those which were necessary in the previous case, that they are necessary. To prove that (7.3) is a sufficient condition for (7.1) to hold we simply use Killing equations

$$m_{a;b} = m_{[a;b]} \tag{7.5}$$

and the consequent identity

$$\{m_{[a;b}m_{c]}\}^{;c} = \tfrac{2}{3}k^c R_{c[a}k_{b]} \tag{7.6}$$

together with the analogous equations (6.5) and (6.6) for k^a, and the commutation condition

$$m^a{}_{;b}k^b - k^a{}_{;b}m^b = 0 \tag{7.7}$$

to derive the identities

$$\begin{aligned} \{k_{[a;b}k_c m_{d]}\}^{;d} &= -\tfrac{1}{2}k^d R_{d[a}k_b m_{c]} \\ \{m_{[a;b}k_c m_{d]}\}^{;d} &= -\tfrac{1}{2}m^d R_{d[a}k_b m_{c]} \end{aligned} \tag{7.8}$$

It is immediately clear from these identities that the Ricci invertibility condition, (7.1), is a consequence of the orthogonal transitivity condition (7.3).

To prove that (7.4) is sufficient for (7.2) to hold we use the group invariance condition

$$\underset{m}{\mathscr{L}}[F_{ab}] \equiv F_{ab;c}m^c - 2F_{c[a}m^c_{;b]} = 0 \tag{7.9}$$

together with the analogous equation (6.11) and the commutation condition (7.7) to derive the identities

$$\begin{aligned} \{F_{ab}k^a m^b\}_{;c} &= 3F_{[ab;c]}k^a m^b \\ \{F_{[ab}k_c m_{d]}\}^{;d} &= \tfrac{1}{4}k_{[a}m_b F_{c]}^{;d}{}_d \end{aligned} \tag{7.10}$$

from which, using the Maxwell equations (6.13) and (6.14), it is clear that the current circularity condition (7.2) will hold whenever the field circularity conditions (7.4) are satisfied.

Continuing the analogy with the previous section, we check that it follows automatically from the electromagnetic field, circularity conditions that the Maxwell energy tensor T_F^{ab} satisfies the staticity condition

$$\left.\begin{aligned} k_a T_F^{a[b}k^c m^{d]} &= 0 \\ m_a T_F^{a[b}k^c m^{d]} &= 0 \end{aligned}\right\} \tag{7.11}$$

from which it follows, if the Einstein equations hold, that the Ricci circularity condition (7.1) will hold at the same time as the field staticity condition (7.4) if and only if the analogous *matter circularity condition*

$$\left.\begin{aligned} k_a T_M^{a[b}k^c m^{d]} &= 0 \\ m_a T_M^{a[b}k^c m^{d]} &= 0 \end{aligned}\right\} \tag{7.12}$$

are satisfied, i.e. if and only if

$$u^{[a}k^b m^{c]} = 0 \tag{7.13}$$

The basic converse theorem is due to Papapetrou (see Papapetrou 1966, Kundt and Trumper 1966, Carter 1969). The electromagnetic generalization (Carter 1969) is quite straightforward, and does not require any special trick such as was needed for the electromagnetic generalization of the Lichnerowicz theorem. Moreover the topological and boundary conditions required for the Papapetrou theorem are very much weaker than for the Lichnerowicz theorem, so that the original theorem can be applied directly to the case where a hole is present.

The general result is as follows:

THEOREM 7 (Generalized Papapetrou Theorem) If $\mathscr{M}$ is pseudo-stationary, asymptotically flat and asymptotically source free, and if the electromagnetic and material current circularity conditions (7.2) and (7.13) are satisfied in a *connected* subdomain $\mathscr{D}$ of $\mathscr{M}$ which intersects the rotation axes (or is otherwise known to contain points at which (7.4) and (7.3) are satisfied) then the electromagnetic field circularity condition (7.4) and the metric circularity condition (7.3) will be satisfied everywhere in $\mathscr{D}$.

Proof We prove the basic Papapetrou theorem by introducing a second twist vector ψ^a, analogous to the vector ω^a introduced (by equation 6.22) in the previous section, defined by

$$\psi_a = \tfrac{1}{2}\epsilon_{abcd}m^b m^{c;d} \tag{7.14}$$

It is evident that the required metric circularity condition 7.3 will be satisfied if and only if the twist scalars $m^c\omega_c$ and $k^c\psi_c$ are both zero. Now the identities (7.7) can be converted to the equivalent dual forms

$$\left.\begin{aligned}(m^c\omega_c)_{;a} &= \epsilon_{abcd}k^b m^c R^{dl}k_l\\ (k^c\psi_c)_{;a} &= \epsilon_{abcd}k^b m^c R^{dl}m_l\end{aligned}\right\} \tag{7.15}$$

respectively. Hence it is clear that the Ricci circularity condition (7.1) will be satisfied in the domain $\mathscr{D}$ if and only if the *twist scalars* are *constant* i.e.

$$\left.\begin{aligned}(m^c\omega_c)_{,a} &= 0\\ (k^c\psi_c)_{,a} &= 0\end{aligned}\right\} \tag{7.16}$$

in $\mathscr{D}$. Since m^c, and hence also ψ^c, are zero on the rotation axis, the twist scalars are also zero on the rotation axis, and hence by conditions of the theorem, at some points of $\mathscr{D}$. Thus (7.16) implies that we shall have

$$\left.\begin{aligned}m^c\omega_c &= 0\\ k^c\psi_c &= 0\end{aligned}\right\} \tag{7.17}$$

at all points of $\mathscr{D}$. This completes the proof of the basic theorem establishing that (7.3) is a consequence of (7.1), and hence in the non-electromagnetic case of (7.13).

To cover the electromagnetic case, we proceed from the fact that using the Maxwell equations (6.13) and (6.14), the identities (7.10) can be reduced to the form

$$\left.\begin{aligned}(\bar{E}_b m^b)_{,a} &= 0\\ (\bar{B}_b m^b)_{,a} &= 4\pi\epsilon_{abcd}k^b m^c j^d\end{aligned}\right\} \tag{7.18}$$

where $\bar{E}_b$ and $\bar{B}_b$ are as defined by (6.24) and (6.25). It is clear therefore that the correct statisticity condition (7.4) is sufficient to ensure that the scalar $\bar{B}_b m^b$ is constant in $\mathscr{D}$, while $\bar{E}_b m^b$ will be constant in any case, i.e. we shall have

$$\left.\begin{aligned}(\bar{E}_b m^b)_{,a} &= 0\\ (\bar{B}_b m^b)_{,a} &= 0\end{aligned}\right\} \tag{7.19}$$

in $\mathscr{D}$. Now the required electromagnetic field circularity condition (7.4) is clearly equivalent to the requirement that the scalars $\bar{E}_b m^b$ and $\bar{B}_b m^b$ should be zero. These scalars will obviously be zero on the rotation axis where m^b is zero and hence at some points of $\mathscr{D}$, and therefore it follows from (7.19) that we shall have

$$\left.\begin{aligned}\bar{E}_b m^b &= 0\\ \bar{B}_b m^b &= 0\end{aligned}\right\} \tag{7.19}$$

which is equivalent to (7.4), at all points of $\mathscr{D}$. We have already remarked that this is sufficient for (7.11) and hence, subject to (7.13), for (7.1) to hold. Therefore by the first part of the proof (7.3) will still hold in the electromagnetic case. This completes the proof of the theorem.

COROLLARY TO THEOREM 7 If the conditions of Theorem 7.1 are satisfied, and in addition the domain $\mathscr{D}$ is *simply* connected, then (except on the rotation axis) the electromagnetic field tensor F_{ab} is derivable, via (6.48) from a vector potential A_a which satisfies the group invariance condition

$$\underset{m}{\mathscr{L}}\,[A_a] \equiv A_{a;b}m^b + A_b m^b_{;a} = 0 \tag{7.20}$$

and its analogue (6.49), and also the *electromagnetic potential circularity condition*

$$A_{[a}k_b m_{c]} = 0 \tag{7.21}$$

(The vector potential will be well behaved on the rotation axis only if the magnetic flux

$$4\pi P = \tfrac{1}{2}\oint_S \epsilon^{abcd}F_{ab}dS_{cd} \tag{7.22}$$

over any compact 2-surface S is zero, i.e. only if there are no magnetic monopoles.)

Proof It is clear that the electromagnetic field circularity condition (7.4) implies that (except on the rotation axis) the field can be expressed in the form

$$F_{ab} = 2e_{[a}k_{b]} + 2f_{[a}m_{b]} \tag{7.23}$$

where e_a and f_a are vectors which satisfy the orthogonality conditions

$$\left.\begin{aligned} e_a k^a &= e_a m^a = 0 \\ f_a k^a &= f_a m^a = 0 \end{aligned}\right\} \tag{7.24}$$

Moreover by the group invariance (7.8) and (6.11) e_a and f_a can be chosen to satisfy the corresponding group invariance conditions

$$\left.\begin{aligned} e_{a;b}k^b + e_a k^b_{;a} = 0 \qquad & e_{a;b}m^b + e_a m^b_{;a} = 0 \\ f_{a;b}k^b + f_a k^b_{;a} = 0 \qquad & f_{a;b}m^b + f_a m^b_{;a} = 0 \end{aligned}\right\} \tag{7.25}$$

Now it can be verified by a little algebra that the Maxwell equation (6.13) implies that the vectors Φ_a and B_a defined by

$$\left.\begin{aligned} \Phi_a &= Ve_a - Wf_a \\ B_a &= Xf_a + We_a \end{aligned}\right\} \tag{7.26}$$

will satisfy

$$\left.\begin{aligned} \Phi_{[a;b]} &= 0 \\ B_{[a;b]} &= 0 \end{aligned}\right\} \tag{7.27}$$

It follows from the simple connectivity condition that there will exist scalars Φ and B everywhere on $\mathscr{D}$ (including the rotation axis) such that

$$\left.\begin{aligned} \Phi_a &= \Phi_{,a} \\ B_a &= B_{,a} \end{aligned}\right\} \tag{7.28}$$

It can now be checked that the vector A_a given by

$$A_a = -\frac{X\Phi + WB}{\sigma} k_a + \frac{W\Phi + VB}{\sigma} m_a \tag{7.29}$$

will satisfy the conditions (6.49), (7.20), (7.21). (However it follows from the conditions (3.13) that this vector potential will be singular on the rotation axis unless B can be chosen to be zero on the rotation axis, and it can be verified by Stokes' theorem that this will be possible only if the no magnetic monopole condition is satisfied.) This completes the proof of the corollary.

Except where the Killing bivector $\rho_{ab} = 2k_{[a}m_{b]}$ is null or degenerate (and hence, by Theorem 4.2, everywhere in $\ll\mathscr{I}\gg$ except on the rotation axis) the Frobenius orthogonality condition 4.3 implies that it is possible to choose locally well behaved functions t and φ that are constant on the 2-surfaces orthogonal to the surfaces of transitivity of the action $\pi^{\mathscr{S}A}$, where t is also constant on the trajectories of the action π^A generated by m^a, and φ is also constant on the trajectories of the action π^s generated by k^a. In other words t and φ can be chosen so as to satisfy

$$t_{,[a}k_b m_{c]} = 0 \tag{7.30}$$

$$\varphi_{,[a}k_b m_{c]} = 0 \tag{7.31}$$

$$t_{,a}m^a = 0 \tag{7.32}$$

$$\varphi_{,a}k^a = 0 \tag{7.33}$$

They will be determined uniquely up to an arbitrary additive constant if we also impose the standard normalization conditions

$$t_{,a}k^a = 1 \tag{7.34}$$

$$\varphi_{,a}m^a = 1 \tag{7.35}$$

Under these conditions the metric form $ds^2 = g_{ab}\, dx^a\, dx^b$ on $\ll\mathscr{I}\gg$ can be expressed (except on the rotation axis) in the form

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta + X\, d\varphi^2 + 2W\, d\varphi\, dt - V\, dt^2 \tag{7.36}$$

with X, W, V as defined in section 3, where α, β run from 1 to 2, and where the co-ordinates x^α are constant on the surfaces of transitivity of the action π^{sA}. For many purposes it will be convenient to work not with φ but with a complementary angle co-ordinate function $\varphi^\dagger$ defined analogously to φ except for the requirement that it be constant on the trajectories of the action $\pi^{s\dagger}$ instead of π^s, i.e. $\varphi^\dagger$ is defined by the requirements

$$\varphi^\dagger_{,[a}l_b m_{c]} = 0 \tag{7.37}$$

$$\varphi^\dagger_{,a}l^a = 0 \tag{7.38}$$

$$\varphi^\dagger_{,a}m^a = 1 \tag{7.39}$$

which determines it uniquely up to an additive constant, which may be chosen in such a way that $\varphi^\dagger$ is related to φ by

$$\varphi^\dagger = \varphi - \Omega^H t \tag{7.40}$$

In terms of $\varphi^\dagger$ the metric form (7.36) may be rewritten in the complementary form

$$ds^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta + X\, d\varphi^{\dagger 2} + 2W^\dagger\, d\varphi^\dagger\, dt - V^\dagger\, dt^2 \tag{7.41}$$

It can easily be seen that if $\ll\mathscr{I}\gg$ is *simply* connected t can be taken to be a globally well behaved function on $\ll\mathscr{I}\gg$, and φ and $\varphi^\dagger$ can be taken to be well behaved angle co-ordinates, defined modulo 2π, on $\ll\mathscr{I}\gg$.

In terms of such a co-ordinate system, the electromagnetic field form $A = A_a\, dx^a$

can be expressed in terms of the scalars Φ and B introduced in the proof of the corollary to theorem 7, as

$$A = \Phi\, dt + B\, d\varphi \tag{7.42}$$

For many purposes it is convenient to use the complementary expression

$$A = \Phi^\dagger\, dt + B\, d\varphi^\dagger \tag{7.43}$$

where

$$\Phi^\dagger = \Phi + \Omega^H B. \tag{7.44}$$

8 The Four Laws of Black Hole Mechanics

For many purposes it is useful to analyse the neighbourhood of the Killing Horizon $\mathscr{H}^+$ in terms of a canonical null co-ordinate system constructed as follows. One first chooses a null hypersurface cutting across the Killing horizon $\mathscr{H}^+$. One can then construct a local null co-ordinate function v which is defined by the requirement that it should be zero on the chosen null hypersurface, and that it should satisfy $v_{,a} l^a = 1$, so that it will be constant on a family of null hypersurfaces congruent to the original one. By requiring that the other co-ordinate functions, r^*, x^2, x^3 should be constant on the trajectories of the action $\pi^{s\dagger}$ generated by l^a we ensure that v will be an ignorable co-ordinate. We choose the co-ordinates x^2, x^3 so as to be constant on each of the null geodesic generators of the null hyper-surfaces on which v is constant. The co-ordinate r^* is then specified uniquely by the requirement that it be zero on the Killing horizon $\mathscr{H}^+$ and that it varies along the null geodesics, on which x^2 and x^3 are required to be constant, in such a way that the metric takes the standard form

$$ds^2 = -V^\dagger\, dv^2 + 2\, dv\, dr^* + 2l_i\, dv\, dx^i + g_{ij}\, dx^i\, dx^j \tag{8.1}$$

where i, j run over 2, 3 and where $V^\dagger$, l_2 and l_3 are functions of r^*, x^2, x^3 only. It is clear that in addition to the familiar condition $V^\dagger = 0$, we shall also have

$$l_i = 0 \tag{8.2}$$

on the Killing horizon $\mathscr{H}^+$ at $r^* = 0$, and it follows further from (5.14) that in terms of such a co-ordinate system we shall have

$$\frac{\partial V^\dagger}{\partial r^*} = 2\kappa \tag{8.3}$$

on $\mathscr{H}^+$. Now it is evident that when the conditions of the Hawking-Lichnerowicz Theorem 6.2 or the Papapetrou Theorem 7 are satisfied, there must exist a local co-ordinate transformation of the form

$$v = t + \tfrac{1}{2}F(r^*, x^i) \tag{8.4}$$

relating the form (8.1) *either* to the form (6.52) with $x^1 = r^*$ (in the non-rotating case) *or* to the form (7.41) with $x^1 = r^*, \varphi^\dagger = x^3$ (in the rotating case), where F is a function of r^*, x^2, x^3 only. It is clear that this requires

$$V^\dagger \frac{\partial F}{\partial r^*} = 2 \tag{8.5}$$

$$V^\dagger \frac{\partial F}{\partial x^2} = l_2 \tag{8.6}$$

$$V^\dagger \frac{\partial F}{\partial x^3} = l_3 - W^\dagger \tag{8.7}$$

(provided we adopt the convection that $W^\dagger$ is zero in the non-rotating case). Since $V^\dagger$ is a well behaved function which lends to zero on $\mathscr{H}^+$, it follows from (8.3) that in the non-degenerate case (i.e. when κ is non-zero) we must have

$$F(r^*, x^i) = \frac{\ln r^*}{\kappa} + G(r^*, x^i) \tag{8.8}$$

if (8.5) is to be satisfied, where $G(r^*, x^i)$ is a function of r^* and x^i which is well behaved in the limit as $r^* \to 0$ on $\mathscr{H}^+$. Now since l_2 and l_3 are all well behaved functions which tend to zero on $\mathscr{H}^+$, (8.6) and (8.7) imply that the quantities $V^\dagger(\partial F/\partial x^i)$ must also be well behaved functions which tend to zero on $\mathscr{H}^+$. This will clearly be compatible with (8.4) only if the scalar κ satisfies

$$\frac{\partial \kappa}{\partial x^i} = 0. \tag{8.9}$$

Thus we arrive at the following conclusion

THEOREM 8 Under the conditions of Theorem 6.2 or Theorem 7, the quantity κ defined on the horizon $\mathscr{H}^+$ by (5.9) is *constant* over $\mathscr{H}^+$. It is an immediate consequence of this theorem that a connected component of $\mathscr{H}^+$ is either degenerate everywhere or not at all.

The result contained in Theorem 8 was first noticed, in the particular context of the Kerr solutions, by Boyer and Lindquist (1967) when they discovered that transformation from the original co-ordinate system (which had the form (8.1)) in which the solution was discovered by Kerr (1963) to the now standard Boyer-Lindquist co-ordinate system (which has the form 7.41). I cam across the more general result given by Theorem 8 in the course of an examination of the necessary and sufficient boundary conditions required for the black hole uniqueness problem (Carter 1971, Carter 1972). This result was discovered independently under even more general conditions by Hawking (1972) as a lemma in the proof of the strong rigidity theorem. However this result has recently acquired a much greater significance (which is the reason why I am giving it special attention at the present stage) from the work of Hartle and Hawking (1972), to be described in the accompanying course by Hawking. As a result it is now clear that Theorem 8 (which Hawking and I had previously regarded as a minor lemma) deserves to be dignified as the *zeroth law of black hole mechanics* for reasons which will be explained in this section.

Before proceeding it is worth remarking that the fact that the limit (5.24) must be constant on the horizon is closely analogous to the fact that the limit (3.7) used in normalizing the axisymmetry Killing vector is constant on the rotation axis. The analogy can be seen more clearly by considering the fixed point axis $\mathscr{H}^+ \cap \mathscr{H}^-$ predicted by Boyer's Theorem 5—such an axis bearing the same relation to a Lorentz rotation as the rotation axis does to an ordinary space rotation.

The main content of this section will be to describe the extension to black holes of very general heuristic argument originally due to Thorne (1969) and Zeldovich (Zledovich and Novikov (1971)) relating variations in the equilibrium mass of an isolated self gravitating system to corresponding variations in angular momentum, chemical composition and entropy. Before doing so I shall rapidly run through the basic argument, including its generalization (Carter 1972) to include the effect of variation of electric charge.

The generalized Thorne-Zeldovich formula is derived as follows. We consider a reversible change in the equilibrium state of the system, which to start with we shall think of as a rotating star, assumed to be both stationary and axisymmetric, which interacts with a freely falling particle of rest mass m, charge e, and unit velocity vector v^a which is sent in from infinity. The energy

$$E = -k^a p_a \tag{8.6}$$

and angular momentum

$$L_z = m^a p_a \tag{8.7}$$

of the particle are conserved during the free motion, where the momentum vector p_a is defined by

$$p_a = mv_a - eA_a \tag{8.8}$$

where A_a is the electromagnetic vector potential. Following Thorne and Zeldovich, we suppose that the particle interacts with the matter of the star at some point, transferring some of its matter and momentum in the process, and that it is then ejected back to infinity. We suppose that the material motion of the star is purely circular, in the sense of section 7, so that at any point the unit velocity vector u^a of the material is a linear combination of the Killing vectors, i.e.

$$u^a = -(\bar{u}_c\bar{u}^c)^{-1/2}\bar{u}^c \tag{8.9}$$

where the renormalized flow vector $\bar{u}^a$ has the form

$$\bar{u}^a = k^a + \Omega m^a \tag{8.10}$$

and where the quantity Ω defined by this equation is the local angular velocity at the point under consideration. (In the case of a star with *rigid* motion Ω will be independent of position and the renormalized flow vector $\bar{u}^a$ will itself be a Killing vector.) [We shall consistently use a bar to denote any quantity which has been renormalized by multiplication by the *time dilatation* factor $(-\bar{u}_a\bar{u}^a)^{1/2}$.] In the local rest frame, the energy δU transferred to the material of the star will be

$$\begin{aligned}\delta U &= u_a\, d(mv^a)\\ &= u^a(dp_a + A_a\, de)\\ &= (-\bar{u}_a\bar{u}^a)^{-1/2}\{-dE + \Omega\, dL_2 + \bar{u}^a A_a\, de\}\end{aligned} \tag{8.11}$$

Provided A_a is chosen so as to tend to zero in the limit of large distances, the contributions to the charge in total mass M, angular momentum J and charge Q of the star will be given by

$$\delta M = -dE \tag{8.12}$$

$$\delta J = -dL_z \tag{8.13}$$

$$\delta Q = -de \tag{8.14}$$

and hence we obtain

$$\delta M - \Omega\delta J - (\bar{u}^a A_a)\delta Q = (-\bar{u}_a\bar{u}^a)^{1/2}\delta U \tag{8.15}$$

It the star is initially in local thermal and chemical equilibrium with a well defined temperature Θ and well defined chemical potentials $\mu^{(i)}$ associated with various kinds of exchanged particles which are conserved in the interaction process, and in the particular case of a *thermodynamically reversible* process (which will only be possible, even in principle, if hysteresis effects can be neglected) the local energy transfer will be given by

$$\delta U = \mu^{(i)}\delta N_i + \Theta\delta S \tag{8.16}$$

where δN_i are the numbers of the various kinds of conserved particles which are transferred, and δS is the entropy transferred. Thus introducing the renormalized *effective temperature* $\bar{\Theta}$ and effective chemical potentials $\bar{\mu}^{(i)}$ defined by

$$\bar{\Theta} = (-\bar{u}^a\bar{u}_a)^{1/2}\Theta \tag{8.17}$$

$$\bar{\mu}^{(i)} = (-\bar{u}^a\bar{u}_a)^{1/2}\mu^{(i)} \tag{8.18}$$

we are led to the basic formula

$$\delta M = \Omega\delta J + \bar{u}^a A_a\delta Q + \bar{\mu}^{(i)}\delta N_{(i)} + \bar{\Theta}\,\delta S \tag{8.19}$$

for the change in mass of the star. Of course after this process the star will no longer be exactly in mechanical equilibrium. However the formula (8.19) should still be valid for the change in mass after the star has settled down to a new equilibrium (by radiation and other damping mechanisms) in the *small perturbation limit* provided that it can be argued that the energy corrections due to the initial departure from mechanical equilibrium are of second order. If this is the case, we may evaluate the total first order change dM in the mass between the mechanical equilibrium states due to a sequence of such transfer operations by integrating (8.19) in the form

$$dM = \int \Omega \delta J + \int \bar{u}^a A_a \delta Q + \int \bar{u}^{(i)} \delta N_{(i)} + \int \bar{\Theta}\, \delta S \tag{8.20}$$

In order for it to be practically useful it is necessary that this formula should be able to be interpreted as a space integral of locally well defined quantities. With any metric normal element $d\Sigma_a$ associated with an element $d\Sigma$ of a space-like hypersurface Σ, there will be associated flux elements of, charge, particle numbers, and entropy given by

$$dQ = j^a\, d\Sigma_a \tag{8.21}$$

$$dN_{(i)} = n^a_{(i)}\, d\Sigma_a \tag{8.22}$$

$$dS = s^a\, d\Sigma_a \tag{8.23}$$

where j^a, $n^a_{(i)}$ and s^a are the current vectors of charge particle numbers and entropy, which satisfy the conservation laws

$$j^a_{;a} = 0 \tag{8.24}$$

$$n^a_{(i);a} = 0 \tag{8.25}$$

and in the case of entropy, the semi-conservation law

$$s^a_{;a} \geqslant 0 \tag{8.26}$$

the latter being a strict equality for the reversible processes under consideration here. There is also a well defined angular momentum flux element given by

$$dJ = T^{ab} m_b\, d\Sigma_a \tag{8.27}$$

associated with an angular momentum current $T^{ab} m_b$ which satisfies the conservation law

$$[T^{ab} m_b]_{;a} = 0 \tag{8.28}$$

in consequence of the conservation law

$$T^{ab}{}_{;b} = 0 \tag{8.29}$$

of the total energy momentum tensor T^{ab} and of the Killing equations (7.5) satisfied by m^a. [It is not possible to give an equally meaningful local definition of conserved mass since the time symmetry generated by k^a is necessarily violated during any alteration process.] With these definitions, and the interpretation $\delta \equiv d(d)$ where the (first d refers to the alteration and the second d refers to the differential element in the integration) the terms in the formula (8.20) become well defined space integrals. In the non-electromagnetic case such an interpretation will be perfectly valid, but in the electromagnetic case it must be born in mind that the angular momentum transferred in an interaction is not well localized at the point of exchange where Ω is measured, since only part of it goes into the matter, the rest being located elsewhere in the electromagnetic field. Thus the value of Ω to be associated with a contribution $\delta J = d(T^{ab} m_b \, d\Sigma_a)$ in the angular momentum integral in (8.20) is not the locally measured value but some weighted space average of Ω over nearby and to a lesser extent more distant regions. Thus the straightforward interpretation will only be correct either when there is no electromagnetic field (the case originally treated by Thorne and Zeldovich) or when there is no differential rotation, at least of the electromagnetically interacting parts of the system. These are also conditions under which one could expect that the deviations from mechanical equilibrium caused by the variation process will be of second order in the perturbation, as required for the above formula to be applicable. In particular one can be confident that these deviations will be of second order in the particular case when the star is not only in crude mechanical equilibrium but also in equilibrium with respect to all relevant perturbation processes in the sense that there can be no energy release by internal transfer processes; clearly from (8.20) this requires (1) that the star be in thermal equilibrium in the sense that the effective temperature $\bar{\Theta}$ is constant, (2) that it be in chemical equilibrium in the sense that the effective Gibbs potentials $\bar{\mu}^{(i)}$ are constant, (3) that it be in rotational equilibrium (i.e. rigid) in the sense that Ω is constant, and (4) that it be in electrical equilibrium in the sense that the comoving electrical potential

$$\Phi^S = \bar{u}^a A_a \tag{8.30}$$

be constant within the star. It is to be noted that this last condition is equivalent, subject to rigidity, to the requirement that the locally measured electric field E_a in the star be zero, since E_a is given by

$$E_a = (-\bar{u}_c \bar{u}^c)^{-1/2} \bar{E}_a \tag{8.31}$$

where

$$\bar{E}_a = \Phi^S_{,a} \tag{8.32}$$

in the rigid case. When *all* these equilibrium conditions are satisfied, (8.20) can be integrated explicitly to give the change in mass in a reversible variation directly in terms of the *total* changes in angular momentum, charge, particle numbers and entropy in the form

$$dM = \Omega \, dJ + \Phi^S \, dQ + \bar{\mu}^{(i)} \, dN_{(i)} + \bar{\Theta} \, dS \tag{8.33}$$

Under conditions when there are effective thermodynamic restraints which allow the star to exist in equilibrium with variable Ω, Φ^S, $\bar{\mu}^{(i)}$, $\bar{\Theta}$ more care is needed to verify that the energy deviations from the mechanical equilibrium value after the alteration process is really of second order. In fact non-uniform temperature $\bar{\Theta}$ and chemical potentials $\bar{\mu}^{(i)}$ cause no difficulties. In the non-electromagnetic case non-uniform angular velocity Ω (which requires zero viscosity) causes no difficulty either since it is always possible consistently with the corresponding restraint (that of local conservation of angular momentum in the individual matter rings) to minimize the energy by appropriate expansions and displacements of the matter rings. However in the electromagnetic case, when the zero viscosity condition is extended to apply to the electric current, there will be additional restraints (corresponding to conservation of magnetic flux through the matter rings) which may be incompatible with the adjustments of the current rings which would be required for energy minimization. Thus we are again led back to the requirement that the electromagnetically interacting parts of the system should be rigid if the formula (8.20) is to be applicable. Thus the most general practically applicable formula which we can obtain by this line of reasoning has the form

$$dM = \Omega_F \int \delta J_F + \int \Omega\, \delta J_M + \int \Phi^s\, \delta Q + \int \bar{\mu}^{(i)}\, \delta N_{(i)} + \int \bar{\Theta}\, \delta S \tag{8.34}$$

where we have separated the angular momentum contributions of the electromagnetic field, given by the flux element

$$dJ_F = T_F^{ab} m_b\, d\Sigma_a \tag{8.35}$$

from those of the matter field, given by

$$dJ_M = T_M^{ab} m_b\, d\Sigma_a \tag{8.36}$$

in terms of the separate electromagnetic and matter energy tensors defined by (5.28) and (6.19), and where we have been obliged to assume that there is a well defined uniform angular velocity Ω_F associated with the electromagnetic contributions.

Let us now consider the extension of this formula to cover variations, in the case where instead of or in addition to the material system, there is a central black hole. It is obvious (Carter 1972) that in the neighbourhood of the hole we should set

$$\bar{u}^a = l^a \tag{8.37}$$

where l^a is the horizon Killing vector field determined by the rigid angular velocity Ω^H of the hole as described in section 4. There is no alternative to this choice of $\bar{u}^a$ in the black hole limit, since *any* vector of the form (8.7) must approach the null tangent vector l^a of the horizon in the limit if it is to remain timelike up to the horizon. The potential corresponding to Φ^S as defined by (8.30) in the case of a star will be

$$\Phi^\dagger = l^a A_a \tag{8.38}$$

which is clearly the same as the quantity introduced by (7.43). The electric field vector $\bar{E}_a^\dagger$ introduced by (5.30) is clearly given by

$$\bar{E}_a^\dagger = \Phi^\dagger_{,a} \tag{8.39}$$

and it therefore follows from the boundary conditions (5.32) that is *constant* on the horizon. We shall denote the value of this constant by Φ^H i.e. we set

$$\Phi^\dagger = \Phi^H \tag{8.40}$$

on $\mathscr{H}^+$, where Φ^H is uniquely defined by the gauge condition, which we have been using throughout this section, that A^a (and hence also $\Phi^\dagger$) tends to zero in the large distance asymptotic limit. Thus we see that a black hole is analogous to an ordinary body (with finite viscosity and electrical conductivity) in rigid electrical equilibrium.

In extrapolating the formula (8.34) to a black hole, it is clear that the entropy and particle number contributions associated with the hole will be zero, since the time-dilation factor $(-\bar{u}^a\bar{u}_a)^{1/2}$ in the definitions (8.17) and (8.18) of the effective temperature and chemical potentials, tends to zero on the horizon. In other words it is clear that the effective temperature $\bar{\Theta}^H$ and the effective chemical potentials $\bar{u}^{(i)H}$ of the hole must all be taken to be zero i.e.

$$\bar{\Theta}^H = 0 \tag{8.41}$$

$$\bar{\mu}^{(i)H} = 0 \tag{8.42}$$

This is an expression of the so called "transcendence" of ordinary particle conservation laws and the second law of thermodynamics by the black hole.

It is to be noted that though particles which go down a black hole are neither extractable nor even externally detectable, such a process is not necessarily irreversible from a thermodynamic point of view since it is always possible to send in a corresponding number of antiparticles. Thus by considering the black hole limit of Thorne-Zledovich type interaction processes, we are lead to the heuristic deduction that a *reversible* variation of the equilibrium state of a system consisting of an axisymmetric black hole with surrounding matter rings should satisfy

$$dM = \Omega^H(dJ_H + \int \delta J_F) + \int \Omega\, dd_M + \Phi^H dQ_H + \int \Phi^S \delta Q + \int \bar{\Theta}\, \delta S + \int \bar{\mu}^{(i)} \delta N_{(i)} \tag{8.43}$$

where J_H and Q_H are the angular momentum and charge of the hole itself, and where we have supposed that the angular velocity of the electromagneticly interacting parts of the system is the same as that of the hole.

In the particular case of the Kerr-Newman vacuum solutions Φ^H and Ω^H as defined here are known functions of M, J and of Q (which in this case is the same as Q_H) given by

$$\Omega^H = JM^{-1}(2Mr_+ - Q^2)^{-1} \tag{8.44}$$

$$\Phi^H = Qr_+ (2Mr_+ - Q^2)^{-1} \tag{8.45}$$

where

$$r_+ = M + (M^2 - J^2M^{-2} - Q^2)^{1/2}. \tag{8.46}$$

With these values, (8.43) reduces in this case to the exact differential of the mass formula of Christodoulou and Ruffini (1970, 1971) which takes the form

$$M = \{(M_0 + \tfrac{1}{4}Q^2M_0^{-1})^2 + \tfrac{1}{4}J^2M^{-2}M_0^{-2}\}^{1/2} \tag{8.47}$$

where M_0 is an integration constant, given by the expression

$$M_0^2 = \frac{1}{16\pi}\mathscr{A} = \frac{1}{4}\{2M^2 - Q^2 + 2(M^4 - M^2Q^2 - J^2)^{1/2}\} \tag{8.48}$$

where $\mathscr{A}$ is the surface area of the black hole (i.e. the integral of surface area over a spacelike 2-dimensional section of $\mathscr{H}^+$).

Although the basic variation formula for a (8.34) for a star was derived by considering *reversible* processes, we can immediately deduce that it will hold for *any* process if we assume that the equilibrium configuration is a well defined function of the distribution of J, Q, S and $N^{(i)}$ over the rotating matter rings, which will be the case under a wide range of natural conditions. Now in the case when a central black hole is present our experience with the Kerr solutions leads to the *generalized no-hair conjecture* according to which the system as a whole should have just *two additional degrees of freedom in the non-electromagnetic case, and three in the electromagnetic case* (leaving a mathematically conceivable magnetic monopole moment out of account) *in addition to the degrees of freedom* (determining the distribution of J, Q, S, $N^{(i)}$) *associated directly with the external matter rings.*

Now the formula (8.43) derived by considering reversible variations, involves just one additional degree of freedom, namely J_H, associated with the hole in the non-electromagnetic case, and just two, namely J_H and Q_H in the electromagnetic case. However it has been shown by Hawking (1971) (in the manner which he describes in the accompanying course) that the result discovered by Christodoulou and Ruffini in the special case of the Kerr-Newman black holes must be true in general, i.e. the surface area $\mathscr{A}$ of the hole must *always remain constant in any transformation which is reversible.* Taking $\mathscr{A}$ itself as the additional degree of freedom, it therefore follows from the generalized no hair conjecture that the mass variation in a completely general (not necessarily reversible) change between neighbouring black hole equilibrium states should be given by

$$\begin{aligned} dM = \mathscr{T}d\mathscr{A} &+ \Omega_H(dJ_H + \int \delta J_F) + \int \Omega\, dJ_M \\ &+ \Phi^H dQ_H + \int \Phi^S \delta Q + \int \bar{\Theta}\, \delta S + \int \bar{\mu}^{(i)} \delta N_{(i)} \end{aligned} \tag{8.49}$$

where the form of the coefficient $\mathscr{T}$ (which has the dimensions of surface tension) remains to be determined. Since the $\mathscr{T}d\mathscr{A}$ contribution (unlike all the other terms) can only be produced by a non-reversible transformation there is no hope of evaluating it by considering a Thorne-Zeldovich type process. In the particular case of the Kerr-Newman solutions, where M is a known function of J, Q and $\mathscr{A}$ only, it is of course possible to write down a differential formula of the form (8.49) and read off the value of the co-efficient $\mathscr{T}$ empirically, as has been done by Beckenstein (1972), but this does not give any insight into the general form of the coefficient $\mathscr{T}$. However a simple and elegant solution to the problem of evaluating $\mathscr{T}$ has come out of the recent work of Hartle and Hawking (1972). By applying the Newman-Penrose equations to the perturbation effect on the horizon of a test field which represents uncharged matter falling from infinity through the horizon $\mathscr{H}^+$, without interacting with any external matter which may be present (so that there should be no contribution either to dQ_H or to the variations δJ, δQ, $\delta N_{(i)}$, δS in the matter) and using a limit in which there should be no gravitational or electromagnetic radiation to infinity, Hawking and Hartle obtained the formula

$$dM = \frac{\kappa}{8\pi} d\,\mathscr{A} + \Omega^H\, dJ \tag{8.50}$$

(where κ is the constant whose existence was established by Theorem 8) in the manner described by Hawking in his accompanying course. If the generalized no hair conjecture is correct, the coefficient of $d\mathscr{A}$ should be the same for a general variation as in the special kind of variation considered by Hawking and Hartle, and therefore we are led from (8.50) to the conclusion that the coefficient $\mathscr{T}$ in the general mass variation formula (8.49) must be given by

$$\mathscr{T} = \frac{\kappa}{8\pi} \tag{8.51}$$

Since both the generalized Thorne-Zeldovich argument and the Hartle-Hawking discussion (not to mention the generalized no-hair conjecture itself), on which the present derivation of the general black hole mass variation formula (8.49) is based, are essentially heuristic, it is obviously desirable to have a mathematical proof of the validity of the formula (8.49). Such a proof has in fact been constructed, during the course of the present summer school, by Bardeen, Hawking and myself in collaboration. The existence of this proof, which is described in the next section, provides strong evidence in favour of the generalized no-hair conjecture.

Hawking (1971) showed not only that the surface area $\mathscr{A}$ of a black hole must remain constant in any reversible variation, but also that it must increase in any irreversible transformation. This result immediately suggested to many people an analogy between the role played by the surface area $\mathscr{A}$ in black hole mechanics and the role played by entropy in what is traditionally known as thermodynamics (although the term thermal equilibrium mechanics would be more appropriate). The results of this section show that this analogy can be carried very much further, with the locally defined scalar κ playing the role analogous to that of temperature. Thus we are led to formulate the following *four laws of black hole equilibrium mechanics*, which are closely analogous to the four standard laws of thermodynamics.

The zeroth law of black hole mechanics will obviously be the result proved in Theorem 8, that the scalar κ is constant over the horizon $\mathscr{H}^+$.

The *first law of black hole mechanics* will be the mass variation formula (8.49) whose heuristic derivation has just been described, and which will be proved rigorously in the next section.

The *second law of black hole mechanics* will of course be the rule

$$d\mathscr{A} \geqslant 0 \tag{8.52}$$

whose derivation and applications are described by Hawking in the accompanying course.

Continuing the analogy, I suggest that the *third law of black hole mechanics* should be the statement that it is impossible by any procedure, no matter how idealized, to reduce the constant κ of a black hole to zero by a finite sequence of operations. In short, degenerate black hole states represent physically unattainable limits.

Unlike the other three laws which are based on rigorous proofs, this third law is still essentially conjectural. However the evidence provided by our knowledge of the extreme (bottomless) Kerr and Reissner-Nordstrom black hole spaces (as described in Part I of the present source and in the accompanying course of Bardeen) provides strong evidence in favour of this conjecture, which would appear to be a fairly direct consequence of the cosmic censorship hypothesis. Conversely by showing that all degenerated black hole equilibrium states are essentially bottomless, Boyer's Theorem 5 suggests directly that degenerate black hole states are unattainable and this in turn provides support for the cosmic censorship hypothesis. Thus it seems that the third law as stated above is virtually equivalent to the cosmic censorship hypothesis in the sense that they will stand or fall together.

The four laws collected together above are clearly of fundamental importance in their own right. Although they correspond closely to the classical laws of thermodynamics, it is to be emphasized that this is only an analogy whose significance should not be exaggerated. Although they are analogous, $\mathscr{T}$ and $\mathscr{A}$ play a quite distinct role from the temperature and entropy with which they should not be confused (and which enter separately into the first law equation (8.43)). The real effective temperature $\overline{\Theta}^H$ of a black hole is well defined and unambiguously zero, as also are its chemical potentials $\overline{\mu}^{(i)H}$. The ordinary particle conservation laws, and the ordinary second law of thermodynamics are unquestionably transcended by a black hole, in the sense that particles and entropy can be lost without trace from an external point of view. It is not possible to mitigate this trandescension by somehow relating the amount of entropy, (or the number of particles) which have gone in to the subsequent increase in surface area $\mathscr{A}$.

9 Generalized Smarr Formula and the General Mass Variation Formula

The results described in this section were worked out by Bardeen, Hawking and myself at Les Houches. This work originated as a search for a general black hole mass formula of the kind discovered empirically by Smarr (1972) in the particular case of the Kerr and Kerr-Newman solutions. This work leads us on naturally to a generalization of the previous mass variation formulae of Hartle and Sharp (1967) Bardeen (1970) and Carter (1972) so as to obtain a rigorous derivation of the first law of black hole dynamics as given by the formula (8.49).

It follows from the asymptotic boundary conditions that the total mass M and angular momentum J of a general stationary axisymmetric system can be expressed in the Komar form

$$M = -\frac{1}{4\pi}\oint_{\infty} k^{a;b}\, dS_{ab} \tag{9.1}$$

and

$$J = \frac{1}{8\pi}\oint m^{a;b}\, dS_{ab} \tag{9.2}$$

where the integrals are taken over a spacelike 2-sphere with metric normal element dS_{ab}, surrounding the system at large distance. The fact that $k^{a;b}$ and $m^{a;b}$ are antisymmetric makes it possible to apply the generalized Stokes theorem to obtain

$$M = -\frac{1}{4\pi}\int k^{a;b}{}_{;b}\, d\Sigma_a - \frac{1}{4\pi}\oint_H k^{a;b}\, dS_{ab} \tag{9.3}$$

and

$$J = \frac{1}{8\pi}\int m^{a;b}{}_{;b}\, d\Sigma_a + \frac{1}{8\pi}\oint_H m^{a;b}\, dS_{ab} \tag{9.4}$$

where the suffix is used to denote a boundary integral over a 2-sphere on the surface of the central black hole (if there is one) and $d\Sigma_a$ is a metric normal element of a spacelike hypersurface Σ extending from the boundary of the hole out to infinity.

Using the standard identities

$$k^{a;b}{}_{;b} = -R^a{}_b k^b \tag{9.5}$$

$$m^{a;b}{}_{;b} = -R^a{}_b m^b \tag{9.6}$$

(derived in sections 6 and 7) which hold for any Killing vectors, and making the obvious definitions

$$M_H = -\frac{1}{4\pi} \oint_H k^{a;b}\, dS_{ab} \tag{9.7}$$

and

$$J_H = \frac{1}{8\pi} \oint_H m^{a;b}\, dS_{ab} \tag{9.8}$$

we can relate the boundary integrals to the Ricci tensor R_{ab} in the intervening space by

$$M = \frac{1}{4\pi} \int R^a_b k^b\, d\Sigma_a + M_H \tag{9.9}$$

and

$$J = -\frac{1}{8\pi} \int R^a_b m^b\, d\Sigma_a + J_H \tag{9.10}$$

These formulae differ from the standard formulae for an ordinary star only through the presence of the black hole boundary terms. For the Kerr solutions on the other hand, these boundary terms will be the only ones that remain. The neat mass formula recently discovered by Larry Smarr for the Kerr solution prompts us to examine these terms more closely. Introducing the angular velocity Ω^H of the hole, and the rigidly co-rotating Killing vector

$$l^a = k^a + \Omega^H m^a \tag{9.11}$$

we obtain the formula

$$\tfrac{1}{2}M_H = \Omega^H J_H - \frac{1}{8\pi} \oint_H l^{a;b}\, dS_{ab} \tag{9.12}$$

Introducing a second null vector n^a orthogonal to the 2-sphere on the horizon, with the normalization condition $l^a n_a = 1$, we can express the normal element in the form $dS_{ab} = l_{[a}n_{b]}\, dS$ where dS is the element of surface area, and noting that the gravitation constant κ discussed in the previous section can be expressed in the form

$$\kappa = l^a{}_{;b} l^b n_a \tag{9.13}$$

we thus obtain

$$\tfrac{1}{2}M_H = \Omega_H J_H + \frac{\kappa}{8\pi} \mathscr{A} \tag{9.14}$$

We can use this expression to eliminate M_H from the mass formula (9), which leads to the basic generalized Smarr formula

$$\tfrac{1}{2}M = \frac{1}{8\pi}\int R^a_b k^b \, d\Sigma_a + \Omega^H J_H + \mathscr{T}\mathscr{A} \,, \tag{9.15}$$

where $\mathscr{T}$ is given by the expression (8.45), which reduces to the original Smarr formula, as obtained for the Kerr solutions, in the pure vacuum case when the Ricci terms on the right hand side are zero.

The expression is not merely elegant. It is also extremely convenient as a starting point for the rigorous derivation of the general mass variation formula. Before starting on the variational calculation, we shall carry out a further decomposition of the unperturbed mass formula, in order to separate the electromagnetic field contributions in the manner suggested by the original Smarr formula for the charged Kerr-Newman solutions. Thus we split up the total energy momentum tensor T^{ab} appearing in the Einstein equations

$$R^{ab} - \tfrac{1}{2}Rg^{ab} = 8\pi\, T^{ab} \tag{9.16}$$

in the form

$$T^{ab} = T^{ab}_M + T^{ab}_F \tag{9.17}$$

where T^{ab}_M is given by (6.19) and T^{ab}_F is given by (5.28).

The field is assumed to have the form $F_{ab} = 2A_{[a;b]}$ where the electromagnetic potential satisfies the group invariance conditions

$$\begin{aligned} &A_{a;b}k^b + A_b k^b{}_{;a} = 0 \\ &A_{a;b}m^b + A_b m^b{}_{;a} = 0 \end{aligned} \tag{9.18}$$

The electric current vector j^a is defined as usual by

$$F^{ab}{}_{;b} = 4\pi j^a \tag{9.19}$$

We can decompose the angular momentum into matter, field, and hole contributions, the two former being defined by

$$J_M = \int T^a_{Mb} m^b \, d\Sigma_a \tag{9.20}$$

and

$$\begin{aligned} J_F &= \int T^a_{Fb} m^b \, d\Sigma_a \\ &= \int m^c A_c j^a \, d\Sigma_a + \frac{1}{4\pi} \oint_H m^c A_c F^{ab} \, dS_{ab} \end{aligned} \tag{9.21}$$

where in deriving the last formula we have used the asymptotic boundary conditions to eliminate a surface integral contribution. (We have also used the fact—which will enable us to drop out many angular momentum contributions—that Σ can be chosen in such a way that $m^a\, d\Sigma_a = 0$.) In terms of these the total angular momentum takes the form

$$J = J_M + J_F + J_H \tag{9.22}$$

Before giving the corresponding subdivision of the mass term, we introduce the total charge Q of the system, defined analogously to the mass and angular momentum by a boundary integral at infinity as

$$Q = -\frac{1}{4\pi} \oint_{\infty} F^{ab}\, dS_{ab} \tag{9.23}$$

Using (9.19) we can write this as

$$Q = -\int j^a\, d\Sigma_a + Q_H \tag{9.24}$$

with the obvious definition

$$Q_H = -\frac{1}{4\pi} \oint_{H} F^{ab}\, dS_{ab} \tag{9.25}$$

It was shown in the previous section that

$$\Phi^{\dagger} = l^a A_a \tag{9.26}$$

is constant on the boundary of the hole. Using this to evaluate the boundary integrals

$$\oint_{H} l^c A_c F^{ab}\, dS_{ab} = -4\pi \Phi^H Q_H \tag{9.27}$$

and (with a little more work)

$$\oint_{H} A_c F^{ca} l^b\, dS_{ab} = 2\Phi^H Q_H \tag{9.28}$$

we obtain the basic generalization of the electromagnetic Smarr formula in the form

$$\begin{aligned} \tfrac{1}{2}M = &\int [T^a_{Mb} k^b - \tfrac{1}{2} T_M k^a]\, d\Sigma_a - \Omega^H J_M \\ &- \tfrac{1}{2} \int l^c A_c j^a\, d\Sigma_a + \int A_b j^{[b} l^{a]}\, d\Sigma_a \\ &+ \Omega^H J + \mathscr{T}\mathscr{A} + \tfrac{1}{2}\Phi^H Q_H \end{aligned} \tag{9.29}$$

In the source free case, when the electric current and the matter contributions are zero, so that only the last three terms are left, this reduces to the original Smarr formula as given for the Kerr-Newman solutions.

Although the expression (9.29) shows the connection with the original Smarr formula most clearly, it is rather more convenient for starting the variational calculation to recast it in a form in which the Lagrangian densities R and $F_{cd}F^{cd}$ for the gravitational and electromagnetic fields are brought into evidence. Thus (again using (9.28)) we choose to set it out as

$$M = \int [T^a_{Mb}k^b - l^c A_c j^a]\, d\Sigma_a + \frac{1}{16\pi}\int (R + F^d_c F^c_d) k^a\, d\Sigma_a$$
$$+ \tfrac{1}{2}M + \Omega^H J_F + \Omega^H J_H + \Phi^H Q_H + \mathscr{T}\mathscr{A} \qquad (9.30)$$

where the term $\frac{1}{2}M$ on the right hand side is to be interpreted as being given directly by (9.7).

Up to this point we have been able to work with a fully general matter tensor (subject only to group invariance) but in order to carry out the variational calculation we shall now specialize to the case when the circularity conditions discussed in section 7 are satisfied, so that the vector $\bar{u}^a$ introduced in the previous section is well defined.

We now begin the actual variational calculation. In any variational calculation there is a certain freedom of choice in the way in which one identify parts of the manifold before and after the variation. When matter is present a naturally convenient choice is to identify particular particle world lines, i.e. trajectories of u^a or equivalently l^a before and after the variation. In the present cases however an even more important consideration, to which we shall give priority, is the preservation of the invariance group properties. Thus we shall require that the Killing vectors (in their natural contravariant form) be left invariant by the variation, i.e.

$$dk^a = 0 \qquad dm^a = 0 \qquad (9.31)$$

Unfortunately (except in the restrictive case when the angular velocities are left invariant by the variation, i.e. when $d\Omega = 0$) this will not be compatible with preservation of the particle world lines taken as a whole. This is not a very serious problem however, since we are only interested in quantities evaluated on a particular spacelike hypersurface Σ or on its boundary, and we can, and therefore shall, require that the points at which particle world lines cross Σ be left invariant. In particular we require that the boundary points on which particular null generators of the horizon meet the boundary of Σ remain the same, even though the canonical null tangent vector itself will have a variation given by

$$dl^a = m^a\, d\Omega^H \qquad (9.32)$$

Introducing the metric variation tensor h_{ab} given by

$$dg_{ab} = h_{ab}; \qquad dg^{ab} = -h^{ab} \qquad (9.33)$$

we can express the variation of the covariant form of l_a as

$$dl_a = h_{ab} l^b + m_a \, d\Omega^H \tag{9.34}$$

Since l_a remains normal to the horizon, which itself remains invariant, this vector must also satisfy the restriction

$$l_{[a} \, dl_{b]} = 0 \tag{9.35}$$

Using this, together with the group invariance condition

$$l^b(dl_a)_{;b} + l^b{}_{;a} \, dl_b = 0 \tag{9.36}$$

it is easy to verify that the differential of the black hole surface gravity constant κ is given by

$$\begin{aligned} d\kappa &= -\tfrac{1}{2}(dl_a)^{;a} - m_{a;b} l^a n^b \, d\Omega^H \\ &= \tfrac{1}{2} h_{a;b}{}^{;b} l^a - m_{a;b} l^a n^b \, d\Omega^H \end{aligned} \tag{9.37}$$

The corresponding formula for the differential of the black hole surface potential ϕ^H is simply

$$d\overset{H}{\phi} = l^a \, dA_a + m^a A_a \, d\Omega^H \tag{9.38}$$

We can now proceed with the evaluation of the variations of the integrals in (9.30), noting that for any integrand

$$d[\qquad d\Sigma_a] = (-g)^{-1/2} \, d[(-g)^{1/2} \qquad] \, d\Sigma_a \tag{9.39}$$

and that

$$d(-g)^{1/2} = \tfrac{1}{2}(-g^{1/2}) h^a_a \tag{9.40}$$

It is evident that

$$d\left\{ \frac{1}{16\pi} F^d{}_c F^c{}_d (-g)^{1/2} \right\} = \left\{ \tfrac{1}{2} T^{cd}_F h_{cd} + \frac{1}{4\pi} F^{cd}(dA_d)_{;c} \right\} (-g)^{1/2} \tag{9.41}$$

It is well known (see any good standard textbook, such as Landau and Lifshits, whose sign conventions I am following here) that

$$d\left\{ \frac{1}{16\pi} R(-g)^{1/2} \right\} = -\left\{ \frac{1}{16\pi}(R^{cd} - \tfrac{1}{2} R g^{cd}) h_{cd} + \frac{1}{8\pi} h_c{}^{[c;b]}{}_b \right\} (-g)^{1/2} \tag{9.41}$$

It is also true (see Carter 1972) that in a perfectly elastic variation of a solid or perfect fluid

$$\begin{aligned} d\{T^a_{Mb} k^b (-g)^{1/2}\} \, d\Sigma_a &= -\Omega d\{T^a_{Mb} m^b (-g)^{1/2}\} \, d\Sigma_a \\ &\quad - l^a \, d\{\rho(-g)^{1/2}\}|_\Omega \, d\Sigma_a \end{aligned} \tag{9.42}$$

where $d\rho|_\Omega$ denotes the variation in ρ calculated as it would be if $d\Omega$ were held equal to zero. (It is not true that $\partial\rho/\partial\Omega$ is itself zero, as would be the case in the

Newtonian limit, but the contribution to which it gives rise cancels out.) In a purely elastic variation, in which entropy and particle numbers are conserved, the variation will be given by $d\rho|_\Omega = (\partial\rho/\partial g_{ab})h_{ab}$ where

$$\frac{\partial\rho}{\partial g_{ab}} = -\tfrac{1}{2}(p^{ab} + \rho g^{ab} + \rho u^a u^b) \tag{9.43}$$

(see Carter 1972). In a perfectly elastic case the entropy and particle number flux vectors can have no components transverse to the flow since it is necessary to be able to regard the system as being in local thermal and chemical equilibrium, i.e. we must have

$$S^a = Su^a \tag{9.44}$$

$$n^a_{(i)} = n_{(i)}u^a \tag{9.55}$$

Under these conditions it is easy to allow for more general variations in which the local thermal and chemical equilibrium is altered, i.e. variations which do not preserve entropy and particle numbers, provided that ρ is a well defined function not only of the geometry but also of the entropy and particle number density scalars s and $n_{(i)}$ defined by (9.44) and (9.45) so that the local temperature Θ and Gibbs chemical potentials $\mu^{(i)}$ are well defined by

$$\frac{\partial\rho}{\partial s} = \Theta \tag{9.46}$$

$$\frac{\partial p}{\partial n_{(i)}} = \mu^{(i)} \tag{9.47}$$

It is to be noted that these quantities are not all independent, being related by the identity

$$\rho + p = \Theta s + \mu^{(i)} n_{(i)} \tag{9.48}$$

where we define

$$p = \tfrac{1}{3}p^a_a \tag{9.49}$$

When these more general variations are allowed, (9.43) will still be valid provided we substitute

$$d\rho|_\Omega = -\tfrac{1}{2}(p^{ab} + \rho u^a u^b + \rho g^{ab}) + \Theta\, \text{đ}s + \mu^{(i)}\, \text{đ}n_{(i)} \tag{9.50}$$

where $\text{đ}s$ and $\text{đ}n_{(i)}$ are not the total variations in s and the $n_{(i)}$ but only the contributions due to non-conservation of entropy and particle numbers in the local matter rings, i.e. they are given by

$$d(su^a\, d\Sigma_a) = \text{đ}s u^a\, d\Sigma_a \tag{9.51}$$

$$d(n_{(i)} u^a\, d\Sigma_a) = \text{đ}n_{(i)} u^a\, d\Sigma_a \tag{9.52}$$

Under these conditions (9.42) becomes

$$d\{T^a_{Mb}k^b(-g)^{1/2}\}\,d\Sigma_a = \Omega d\{T^a_{Mb}m^b(-g)^{1/2}\}\,d\Sigma_a$$
$$+\, l^a T^{cd}_M h_{cd}(-g)^{1/2}\,d\Sigma_a - \{\bar\Theta\,ds + \bar\mu^{(i)}\,dn_{(i)}\}\,u^a(-g)^{1/2}\,d\Sigma_a \qquad (9.53)$$

where we have introduced the effective temperature and chemical potentials $\bar\Theta$ and $\bar\mu^{(i)}$ defined by (8.17) and (8.18).

Having thus worked out the contributions to the variational integrands, we are now ready to perform the actual integration. Using the group invariance condition

$$h_{ab;c}k^c + h_{ac}k^c_{;b} + h_{bc}k^c_{;a} = 0 \qquad (9.54)$$

to cast the integrand $k^a h_c{}^{[c;b]}{}_{;b}$ as a divergence in the form

$$k^a h_c{}^{[c;b]}{}_{;b} = \{k^a h_c{}^{[c;b]} - k^b h_c{}^{[c;\ ]}\}_{;b} \qquad (9.55)$$

we obtain

$$\frac{1}{8\pi}\int k^a h^{[c;b]}_{c;b}\,d\Sigma_a = \frac{1}{4\pi}\oint_\infty k^a h_c^{[c;b]}\,dS_{ab} - \frac{1}{4\pi}\oint_H k^a h_c^{[c;b]}\,dS_{ab}$$
$$= \tfrac{1}{2}dM + \mathscr{A}d\,\mathscr{T} + J_H\,d\Omega^H \qquad (9.56)$$

where the standard contribution $\frac{1}{2}dM$ from the surface integral at infinity is easily obtained from the a symptotic boundary conditions, and where the hole terms $\mathscr{A}d\,\mathscr{T} + J_H\,d\Omega^H$ are obtained with the aid of the formula (9.36). In a similar manner we can cast the integrand $k^a F^{cd}(dA_d)_{;c}$ as a divergence in the form

$$k^a F^{cd}(dA_d)_{;c} = 2\{F^{c[a}k^{b]}\,dA_c\}_{;b} + 4\pi k^a j^c\,dA_c$$

and hence we obtain

$$\frac{1}{4\pi}\int k^a F^{cd}(dA_d)_{;c}\,d\Sigma_c = -Q\,d\Phi^H$$
$$-(J_F - \int j^a m^c A_c\,d\Sigma_a)\,d\Omega^H + \int l^a j^c\,dA_c\,d\Sigma_a \qquad (9.57)$$

with the aid of (9.37). Finally the variation of the second term in (9.30) gives simply

$$d\{\int l^c A_c j^a\,d\Sigma_a\} = \int l^c j^a\,dA_c\,d\Sigma_a$$
$$+\, d\Omega_H \int j^a m^c A_c\,d\Sigma_a - \int l^c A_c\,\delta Q \qquad (9.58)$$

where we have introduced the abbreviation

$$\delta Q = d[-j^a\,d\Sigma_a] \qquad (9.59)$$

Introducing the analogous abbreviations

$$\delta N_{(i)} = d[-n^a_{(i)}\, d\Sigma_a] \tag{9.60}$$

$$\delta S = d[-s^a\, d\Sigma_a] \tag{9.61}$$

$$\delta J_M = d[-T^a_{Mb} m^b\, d\Sigma_a] \tag{9.62}$$

combining (9.40), (9.41), (9.51) with the Einstein equations, and finally using (9.54), (9.55), (9.56) to simplify the residual integrals, we obtain the variation of (9.30) in the form

$$\begin{aligned} dM = {} & \Omega^H(dJ_H + dJ_F) + \int \Omega \delta J_M + \int \bar{\Theta}\, \delta S \\ & + \int \bar{\mu}^{(i)}\, \delta N_{(i)} + \Phi^H\, dQ_H + \int l^c A_c\, \delta Q \\ & + \frac{\kappa}{8\pi} d\,\mathscr{A} + 2\int j^{[c} l^{a]}\, dA_c\, d\Sigma_a \end{aligned} \tag{9.63}$$

This variation formula, which represents our final expression of the first law of black hole mechanics, reduces to the formula (8.49) when the electromagnetic rigidity condition

$$j^{[c} l^{a]} = 0 \tag{9.64}$$

is satisfied so that the last term drops out; this last term also drops out for any variation which satisfies the perfect conductivity condition that the magnetic flux (cf. equation (7.22)) through any comoving circuit be conserved.

10 Boundary Conditions for the Vacuum Black Hole Problem

From this point onward we shall restrict our attention to the case of black hole spaces in which the pure vacuum or source free Einstein-Maxwell equations are satisfied, excluding the degenerate limit case due to lack to time and space. (Further details of the degenerate limit case are given by Bardeen.) As has been explained in the preceding sections it has been established with certainty in the rotating case, and it is virtually certain in the non-rotating case also, that such a space will satisfy the conditions of Theorem 7 and Theorem 4.2 and it is also virtually certain that the horizon will be connected, with topologically spherical space sections. These various conditions can be summed up as follows.

Condition 10 $\mathscr{M}$ is time orientable asymptotically flat, pseudo-stationary and axisymmetric. The causality axiom is satisfied in $\mathscr{M}$. The domain of outer communications $\ll \mathscr{I} \gg$ of $\mathscr{M}$ is topologically the product of the Euclidean 2-plane and the 2-sphere, and the horizon $\mathscr{H}^+$ is non-degenerate and is topologically the product of a Euclidean line and a 2-sphere. The Einstein equations

$$R^{ab} = 0 \tag{10.1}$$

or the Einstein-Maxwell equations

$$R^{ab} = 8\pi T_F^{ab}$$
$$j^a = 0 \tag{10.2}$$

are satisfied in $\mathscr{M}$.

It follows from Condition 10, by Theorem 4.2 and Theorem 7, as we have seen in section 7, that the metric on the union $\ll \mathscr{I} \gg \cup \mathscr{H}^+$ can be expressed in the form

$$ds^2 = ds_{\mathrm{II}}^2 + X\, d\varphi^2 + 2W\, d\varphi\, dt - V\, dt^2 \tag{10.3}$$

where $ds_{\mathrm{II}}^2 = g_{\alpha\beta}\, dx^\alpha\, dx^\beta$ ($\alpha, \beta = 1, 2$) is the metric projected orthonally onto a 2-dimensional surface, $\overline{\mathscr{B}}$ say, which intersects each surface of transitivity within $\ll \mathscr{I} \gg \cup \mathscr{H}^+$ of the action π^{sA} once. The form (10.3) will be well behaved except on $\mathscr{H}^+$ and on the rotation axis, which form an edge to $\overline{\mathscr{B}}$. Since the topological conditions are such that the rotation axis will have two branches (extending respectively from the north and south poles of the hole to infinity) $\overline{\mathscr{B}}$ will have the topology of a square in Euclidean 2-space possessing three edges, but from which the remaining edge has been excluded, the included edges corresponding in order, to the intersection of $\overline{\mathscr{B}}$ with the southern rotation axis, the horizon $\mathscr{H}^+$, and the northern rotation axis.

The purpose of this section is to show that the system of Einstein equations and global conditions required by Condition 10 can be reduced to a comparatively simple set of partial differential equations with boundary conditions for V, W, X and the scalars B, Φ defined by (7.28) as functions on $\overline{\mathscr{B}}$.

We start by considering the metric form ds_{II}^2 on $\mathscr{B}$. In deriving the general form (10.3) or (7.36) for the metric and the form (7.42) for the field potential we have in effect made use of and satisfied only the Einstein-Maxwell equations involving the cross components in the Ricci-tensor between directions lying in and orthogonal to the surfaces of transitivity, but we have made no use of the equations for the components of the Ricci tensor lying wholly in or wholly orthogonal to the surfaces of transitivity. Now it is true generally that the source-free Einstein-Maxwell equations require that the Ricci tensor as a whole be trace-free; and in the case of a field satisfying the circularity conditions (7.4), it is easy to see further that the energy momentum tensor is such that the projections of the Ricci tensor into the orthogonal to the surfaces of transitivity must be trace free separately. Now it is well known from the work of Papapetrou and others that the condition that the projection of the Ricci tensor into the surfaces of transitivity be trace free, i.e. (in terms of the co-ordinate system of the form (10.3))

$$XR_{tt} - 2WR_{t\varphi} + VR_{\varphi\varphi} = 0 \tag{10.4}$$

is equivalent to the condition that the scalar ρ defined, (whenever it is real) as the non-negative root of the equation

$$\rho^2 = \sigma \tag{10.5}$$

must be *harmonic* i.e. its Laplacian, defined in terms of the two dimensional metric ds^2_{II}, is zero. We can use this fact to show that ρ *has no critical points in the interior of* $\overline{\mathscr{B}}$ (i.e. no points where its gradient vanishes) and hence that can ρ be used as a globally well behaved co-ordinate in $\ll \mathscr{I} \gg$ except on the axi-symmetry axis (Figure 10.1).

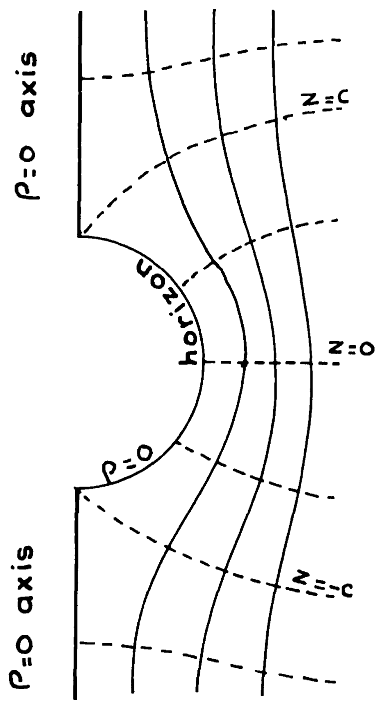

Figure 10.1. Plan of the 2-surface $\overline{\mathscr{B}}$. The continuous lines indicate locuses on which ρ is constant and the dotted lines indicate locuses on which z is constant.

To prove this we remark that since by Theorem 4.2, σ is strictly positive in $\ll \mathscr{I} \gg$ except on the rotation axis (which corresponds to the boundary of $\overline{\mathscr{B}}$), ρ is *strictly positive* in the interior of $\overline{\mathscr{B}}$. On the other hand ρ *is zero on the whole of the boundary of* $\overline{\mathscr{B}}$, since it is immediately clear that ρ is zero both on the rotation axis where X and W are both zero (because m^a is zero) and on the horizon $\mathscr{H}^+$, since, by Theorem 4.2, $\mathscr{H}^+$ lies on the rotosurface where σ is zero. Furthermore the asymptotic flatness boundary conditions at infinity ensure that ρ behaves in the limit at large distances like an ordinary cylindrical radial co-ordinate, and hence that it has no critical points at large distances. Under such well defined boundary conditions as these, ordinary Morse theory tells us that provided there are no *degenerate* critical points, (i.e. points where higher than first deviations of ρ are zero) the number of maxima plus the number of minima minus the number of saddle points of ρ in the interior of $\bar{\mathscr{B}}$ is an invariant of the differential topology. By considering the special case of a cylindrical co-ordinate in ordinary flat space it is obvious that in the present case this invariant is *zero*. Since a harmonic function can have no maxima or minima, it follows that in the present case there can be no saddle points either, provided there are no degenerate critical points.

Now the more specialized harmonic Morse theory and Heinz (1949) enables us to exclude degenerate critical points as well. The critical points of a harmonic function can be classified with a positive integral index number which is the order of the highest partial derivative which vanishes at the point under consideration, this index number being unity for a non-degenerate critical point. According

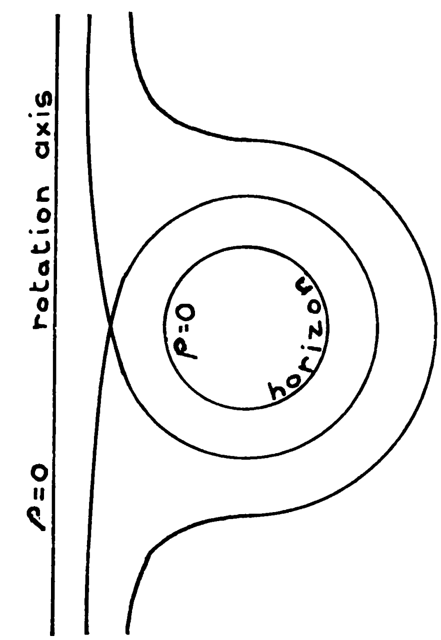

Figure 10.2. Plan of hypothetical alternative form of the 2-surface $\bar{\mathscr{B}}$ corresponding to the mathematically conceivable (but physically impossible) case of a topologically toroidal as opposed to a spherical hole. The continuous lines indicate locuses on which ρ is constant. In this case there is necessarily just one non-degenerate critical point.

to this theory, if degenerate critical parts are present, the sum of the indices of the critical points will be equal to the Morse invariant. In the present case this tells us unambiguously that *there are no critical points of ρ at all* degenerate or otherwise in $\bar{\mathscr{B}}$. (In the case where the topology of $\bar{\mathscr{B}}$ corresponded to that of a toroidal black hole, this theory would tell us equally unambiguously that there would be *one* non-degenerate critical point (Figure 10.2); in the case of two distinct toroidal black holes there would be an ambiguity, since there might be either two non-degenerate critical points or one degenerate critical point of index two.)

Having established that ρ has no critical points in the interior of $\mathscr{B}$ it follows that the curves on which it is constant have no intersections, and that they have everywhere the same topology, namely that at the Euclidean line. Hence not only can we take ρ itself to be a globally well defined co-ordinate on $\overline{\mathscr{B}}$, but we can also choose a globally well behaved scalar z without critical points, and constant on curves orthogonal to those on which z is constant, as a second globally well behaved co-ordinate. In terms of such a co-ordinate system, the metric ds_{II}^2 on $\overline{\mathscr{B}}$ will have the form

$$ds_{\text{II}}^2 = \Sigma(d\rho^2 + Zdz^2) \tag{10.6}$$

where Σ and Z are strictly positive functions in the interior of $\overline{\mathscr{B}}$. Now it follows at once by further application of the harmonicity condition that Z is a function of of z only, and hence that by a variable rescaling of the function z we can arrange to set Z equal to unity thus reducing the metric to

$$ds_{\text{II}}^2 = \Sigma\,(d\rho^2 + dz^2) \tag{10.7}$$

In this form the co-ordinate z is uniquely defined to an additive constant, and this constant may itself be specified uniquely by the requirement that z take equal and opposite values,

$$z = \pm c \tag{10.8}$$

say, where c is a positive constant, in the limit as the junction of the symmetry axis and the horizon is approached.

Thus we deduce finally that $\langle\!\langle \mathscr{I} \rangle\!\rangle$ may be covered globally apart from the degeneracy on the symmetry axis, by a Weyl-Papapetrou co-ordinate system ρ, z, φ, t in which the metric takes the canonical form

$$ds^2 = \Sigma\,(d\rho^2 + dz^2) + X\,d\varphi^2 + 2W\,d\varphi\,dt - V\,dt^2 \tag{10.9}$$

while the electromagnetic field will have the form

$$F = 2(B_{,\rho}\,d\rho + B_{,z}\,dz) \wedge d\varphi + 2(\Phi_{,\rho}\,d\rho + \Phi_{,z}\,dz) \wedge dt \tag{10.10}$$

where the co-ordinate φ is defined modulo 2π, the co-ordinate ρ ranges over the positive half of the real line, and where t and z range over the entire real line, and where these co-ordinates are defined uniquely apart from the possibility of adding a (clearly ignorable) constant to t or φ. The same applies to the complementary forms

$$ds^2 = \Sigma(d\rho^2 + dz^2) + X\,d\varphi^{\dagger 2} + 2W^{\dagger}\,d\varphi^{\dagger}\,dt - V^{\dagger}\,dt^2 \tag{10.11}$$

$$F = 2(B_{,\rho}\,d\rho + B_{,z}) \wedge d\varphi^{\dagger} + 2(\Phi^{\dagger}_{,\rho}\,d\rho + \Phi^{\dagger}_{,z}\,dz) \wedge dt \tag{10.12}$$

It is well known (see e.g. Ernst (1969) that for a metric of the simple form (10.9) with an electromagnetic field of the form (10.10), the source free Einstein-Maxwell field equations for the metric and potential components in the surfaces of transitivity, i.e. for the variables V, W, X and B, Φ decouple from the remaining field equations which are either redundant or serve to determine

Σ on $\overline{\mathscr{B}}$ uniquely, (up to a constant multiplicative factor fixed by the boundary conditions), by an explicit quadrature, as a function of V, W, X, B, Φ.

In this section we shall give boundary conditions on these five variables V, W, X, B, Φ and $\overline{\mathscr{B}}$ which we shall show to be both necessary and (subject to the field equations) sufficient for the axisymmetry axis and the horizon $\mathscr{H}^+$ in $\mathscr{M}$ to be well behaved.

We start by considering the more familiar case of the axisymmetry axis, on which m^a and hence also X and W, are zero, *leaving out of consideration for the time being the intersection* (on which V will be zero also) *of the axis with the horizon*. Since the Killing vector squared magnitude X must increase in proportion to the square of the orthogonal spacelike distance from the axis, the same must be true also of σ, and it follows that its square root ρ can be used as well behaved co-ordinate, with finite non-vanishing gradient not only in the interior of $\overline{\mathscr{B}}$ but also in the part of the boundary corresponding to the axisymmetry axis. It follows that the coefficient Σ in the metric form, must be finite on the rotation axis (in fact it must satisfy $\Sigma = V^{-1}$ there) and that the orthogonal co-ordinate z on $\overline{\mathscr{B}}$ is also well behaved there.

Now (by considering the affect of a rotation by an angle π about the axis) it is evident that in order to be well behaved on the axis the scalars V, W, X, Φ, B must all be *even* functions of ρ, and hence can be expressed as well behaved functions not only of ρ and z but also of σ and z. Thus we see at once that necessary boundary conditions on X, W, V are

$$V = V(\sigma, z) \tag{10.13}$$

$$W = \sigma W_1(\sigma, z) \tag{10.14}$$

$$X = \sigma X_1(\sigma, z) \tag{10.15}$$

where the functions W_1, X_1 and V are well behaved functions of σ and z, the two last being strictly positive on the axis where σ is zero. We can deduce corresponding conditions for Φ and B by considering the requirements for the regularity of (10.10). The form dt in $\ll \mathscr{I} \gg$ is well behaved, but the form $d\varphi$ is singular on the axis, and therefore B must be correspondingly restricted in order to ensure that $(B_{,\rho}\, d\rho + B_{,z}\, dz) \wedge d\varphi$ is well behaved. Now since we have $B_{,\rho}\, d\rho = B_{,\sigma}\, d\sigma$ and since the form $d\sigma$ tends to zero on the axis in such a way that $d\sigma \wedge d\varphi$ is well behaved, it follows that the restriction applies only to the partial derivative $B_{,z}$,which must itself tend to zero on the axis. Thus we obtain the conditions

$$\Phi = \Phi(\sigma, z) \tag{10.16}$$

$$B = B^A + \sigma B_1(\sigma, z) \tag{10.17}$$

where Φ and B_1 are well behaved functions of σ and z, and B^A is a constant, which will be the same on both the north and south branches of the axis only if the magnetic flux defined by (7.22) is zero.

The conditions (10.13), (10.14), (10.15), (10.16), (10.17) are not only necessary for the symmetry axis to be well behaved, but they are also sufficient, *provided* that the *Einstein-Maxwell field equations* are *satisfied*, since it follows from the conditions (10.16) and (10.17) that the ρ, z component of the energy-momentum tensor, and hence also at the Ricci tensor, in terms of the co-ordinate system of the metric form (10.9), must tend to zero on the symmetry axis. Now by analysing the explicit form of the Ricci tensor, using (10.13), (10.14), (10.15) it can be checked that satisfying the equation $R_{\rho z} = 0$ on the axis *automatically* ensures that the scalar function Σ in the metric satisfied the boundary condition

$$\Sigma = \epsilon^{-2} [X_1(\sigma, z) + \sigma \Sigma_1(\sigma, z)] \tag{10.18}$$

on the axis where ϵ is a non-zero multiplicative constant of integration, Σ_1, is a well behaved function of σ and z, and X_1 is the same function as was introduced in equation (10.15). It is now easy to see that the condition (10.13), (10.14), (10.15) together with (10.18) are sufficient for the cylindrical co-ordinate degeneracy of the form (10.7) on the symmetry axis $\rho = 0$ to be removable in the usual way be replacing ρ and φ by Cartesian type co-ordinates x, y defined by

$$\left.\begin{aligned} x &= \rho \cos \varphi \\ y &= \rho \sin \varphi \end{aligned}\right\} \tag{10.19}$$

provided that the constant ϵ has the value unity, which will in fact be the case under the assumed global topological conditions, since both disconnected components of the axisymmetry axis extend to asymptotically large distances, where the asymptotic flatness conditions ensure that ϵ is indeed unity. Under these conditions the use of (10.13), (10.14), (10.15), (10.18) in conjunction with the transformation (10.19) reduces the cylindrical co-ordinate form (10.9) of the metric to the Cartesian form

$$\begin{aligned} ds^2 = X_1 [dx^2 + dy^2 + dz^2] + \Sigma_1 [(x\, dx + y\, dy)^2 + (x^2 + y^2)\, dz^2] \\ + 2W_1(x\, dy - y\, dx)dt - V\, dt^2 \end{aligned} \tag{10.20}$$

which can easily be seen to be well behaved on the axisymmetry axis where $x = y = 0$. Similarly the use of (10.16) and (10.17) in conjuction with the transformation (10.19) reduces the cylindrical co-ordinate form (7.41) of the vector potential to the form

$$A = \Phi\, dt + B_1(y\, dx - x\, dy) + B^A\, d\varphi \tag{10.21}$$

Since B^A is a constant on each branch of the axis, the final singular term can be removed by a canonical form preserving gauge transformation whenever the magnetic monopole (7.22) is zero, and in any case, the electromagnetic field F will have the well behaved Cartesian form

$$F = 2\, d\Phi \wedge dt + 2dB_1 \wedge (y\, dx - x\, dy) + 4B_1\, dy \wedge dx \tag{10.22}$$

We now move on to consider the analogous boundary conditions required by the regularity of the horizon $\mathscr{H}^+$, again excluding for the time being the junction of the horizon with the axis. For analysing the horizon, on which as we have seen, σ and $V^\dagger$, and therefore also $W^\dagger$ are zero, it is convenient to work with the complementary form (10.11) of the metric instead of (10.9). In consequence the non-degeneracy condition the gradient of $V^\dagger$, and therefore also (since we are excluding the rotation axis where X is zero) the gradient of σ in $\mathscr{M}$ are non-zero on $\mathscr{H}^+$, and it follows this time that σ can be used as a well behaved co-ordinate with finite non-vanishing gradient not only in the interior of $\overline{\mathscr{B}}$ but also on the part of the boundary of $\overline{\mathscr{B}}$ corresponding to the horizon. Thus the square root, ρ, of σ will have a singular gradient on the horizon boundary of $\overline{\mathscr{B}}$ (in contrast with the situation on the axis boundary where ρ, but not σ, is well behaved as a co-ordinate). Now using (5.24) we can easily see that we must have

$$\Sigma^{-1} = \rho_{,\alpha}\rho^{,\alpha} = \kappa X \tag{10.23}$$

on $\mathscr{H}^+$. Since Σ thus tends to a finite limit on the horizon boundary of $\overline{\mathscr{B}}$, it follows that z is a well behaved co-ordinate on the horizon boundary. Hence we deduce that $V^\dagger$, $W^\dagger$, X, $\Phi^\dagger$, B will be well behaved functions of σ and z on the horizon boundary, and therefore that in the neighbourhood of the horizon boundary they will have the form

$$V^\dagger = \sigma V_1^\dagger(\sigma, z) \tag{10.24}$$

$$W^\dagger = \sigma W_1^\dagger(\sigma, z) \tag{10.25}$$

$$X = X(\sigma, z) \tag{10.26}$$

where the functions $W_1^\dagger$, $V_1^\dagger$ and X are well behaved functions of σ and z, the two last being strictly positive, and

$$\Phi^\dagger = \Phi^H + \sigma\Phi_1^\dagger(\sigma, z) \tag{10.27}$$

$$B = B(\sigma, z) \tag{10.28}$$

where $\Phi_1^\dagger$ and B are well behaved functions, and Φ^H is the *constant* previously introduced by (8.40).

As before it can be seen these conditions are also *sufficient* for the regularity of the horizon when the Einstein-Maxwell field equations are satisfied since as in the previous section the conditions (10.27), (10.28) imply that the Ricci component $R_{\rho z}$ in the form (10.11) must tend to zero on the horizon, and hence by (10.24), (10.25), (10.26) that the scalar functions Σ must have the form

$$\Sigma = \kappa^{-2}[V_1(\sigma, z) + \sigma\, \Sigma_1(\sigma, z)] \tag{10.29}$$

where Σ_1 is a well behaved function of σ and z, and κ is a strictly positive multiplicative *constant* or integration, which is clearly, by (10.23) the same as the constant κ with which we are already familiar. Thus the boundary conditions (10.24) to (10.28) are sufficient to ensure that the condition (whose necessity was shown in Theorem 8) that κ be constant will be satisfied. We can therefore remove the co-ordinate degeneracy on the horizon by a Finkelstein (1958) type co-ordinate transformation in which t and ρ are replaced in the form (10.11) by σ and a new ignoreable co-ordinate v defined by

$$v = t + \kappa^{-1} \ln |\rho| \tag{10.30}$$

which leads to the form

$$ds^2 = V_1 [\kappa^{-2} dz^2 - (\sigma\, dv - \tfrac{1}{2}\kappa^{-1} d\sigma) dv] + X\, d\varphi^{\dagger 2} + \kappa^{-2} \Sigma_1 [\sigma\, dz^2 + \tfrac{1}{4} d\sigma^2] + 2W_1(\sigma\, dv - \tfrac{1}{2}\kappa^{-1} d\sigma)\, d\varphi^\dagger \tag{10.31}$$

which, can easily be seen to be well behaved on the horizon where σ is zero, since V_1 and X are strictly positive there. The corresponding forms for the vector potential A and the electromagnetic field F are

$$A = B\, d\varphi^\dagger + \Phi_1^\dagger [\sigma\, dv - \tfrac{1}{2}\kappa^{-1} d\sigma] + \Phi^H \kappa^{-1} \rho^{-1} d\rho \tag{10.32}$$

which is non-singular except for the final term, which could be removed by a gauge change, and

$$F = 2dB \wedge d\varphi^\dagger + 2d\Phi_1^\dagger \wedge [\sigma\, dv - \tfrac{1}{2}\kappa^{-1} d\sigma] + 2\Phi_1^\dagger\, d\sigma \wedge dv \tag{10.33}$$

which is well behaved in any case.

The condition (10.24) to (10.28) are not only sufficient for the future bounding horizon $\mathscr{H}^+$of $\ll \mathscr{I} \gg$ to be well behaved, as we have just shown (still leaving aside the junction of the horizon with the axisymmetry axis) but they are also sufficient for there to exist a Kruskal type co-ordinate extension of $\ll \mathscr{I} \gg$ to cover both a corresponding past boundary horizon $\mathscr{H}^-$ of $\ll \mathscr{I} \gg$ and a crossover axis $\mathscr{H}^+ \cap \mathscr{H}^-$ on which l^a is zero, where the past and future horizons bounding $\ll \mathscr{I} \gg$ meet (whether or not the past boundary horizon and the crossover axis were included in $\mathscr{M}$ as it was originally specified). To see this we introduce co-ordinates w^+ and w^- in plane of t and ρ, by the equation

$$w^\pm = \rho\, e^{\pm \kappa t} \tag{10.34}$$

and thus obtain transform (10.11) to

$$ds^2 = \kappa^{-2} V_1 [dw^+ dw^- + dz^2] + X\, d\varphi^{\dagger 2} + \kappa^{-2} \Sigma_1 [\tfrac{1}{4}(w^- dw^+ + w^+ dw^-)^2 + w^+ w^- dz^2] + \kappa^{-1} W_1 [w^- dw^+ - w^+ dw^-]\, d\varphi^\dagger \tag{10.35}$$

which is well behaved both in the neighbourhood of the future bounding horizon, $\mathscr{H}^+$ which is represented by $w^+ = 0$, and on the past bounding horizon $\mathscr{H}^-$ represented by $w^- = 0$, including the crossover where they both meet. The corresponding values for the vector potential and field are

$$A = B\, d\varphi^\dagger + \tfrac{1}{2}\kappa^{-1} \Phi_1 [w^- dw^+ - w^+ dw^-] + \Phi^H dt \tag{10.37}$$

which is well behaved in the neighbourhood of the horizons, except for the final singular term which is removable by a gauge transformation, and

$$F = 2dB \wedge d\varphi^{\dagger} + \kappa^{-1} d\Phi_1 \wedge [w^- dw^+ - w^+ dw^-] + 2\kappa^{-1}\Phi_1 dw^- \wedge dw^+ \qquad (10.39)$$

which is well behaved near the horizons in any case.

We now come finally to the question of regularity on the north and south poles of the black holes i.e. on the junctions of the two rotation axis components

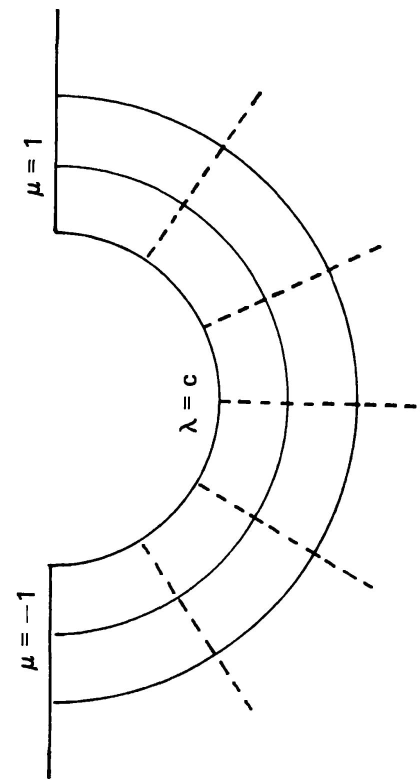

Figure 10.3. Plan of the 2-surface $\mathscr{B}$. The continuous lines indicate locuses on which λ is constant and the dotted lines indicate locuses on which μ is constant.

with the horizon. It is clear from symmetry considerations that the axis must intersect the horizons orthogonally, and therefore it will obviously be convenient to discuss the junction in terms of a new orthogonal co-ordinate system chosen transverse to the ρ, z system, in such a way that one co-ordinate is constant on $\mathscr{H}^+$ while the other is constant on the rotation axis. The simplest way to construct such a system is by introducing ellipsoidal polar type co-ordinates (Figure 10.3), μ running from -1 to $+1$ and λ running from c to ∞ (where $\pm c$ are the values of z on the junctions and where c is *strictly* greater than zero since as we have seen, z is a well behaved co-ordinate function of the horizon), these co-ordinates being defined on $\overline{\mathscr{B}}$ and hence on $\ll \mathscr{I} \gg \cup \mathscr{H}^+$ in terms of ρ and z by

$$\rho^2 = \sigma = (\lambda^2 - c^2)(1 - \mu^2) \tag{10.40}$$

$$z = \lambda\mu \tag{10.41}$$

In this system the two axisymmetry axis components are characterized by $\mu = \pm 1$ respectively, and the horizon boundary is characterized by $\lambda = c$. Corresponding to the conditions that ρ is well behaved on the axis while σ is well behaved on the horizon, we shall have the conditions that $(1 - \mu^2)^{1/2}$ is well behaved on the axis (and hence everywhere except on the equator where $\mu = 0$) and λ is well behaved on the horizon (and hence everywhere without exception).

It follows that we can replace the separate axis and horizon boundary condition that the co-ordinate functions X, etc., be well behaved as functions of σ and z, by the joint boundary conditions that they be well behaved as functions of λ and μ everywhere. Thus the necessary boundary conditions (10.13) to (10.17) and (10.24) and (10.25) can be replaced (noting that the differences between V, W, Φ and $V^\dagger$, $W^\dagger$, $\Phi^\dagger$ tend to zero on the axisymmetry axis) by the single set of necessary boundary conditions

$$X = (1 - \mu^2)\hat{X}(\mu, \lambda) \tag{10.42}$$

$$W^\dagger = (\lambda^2 - c^2)(1 - \mu^2)\hat{W}(\mu, \lambda) \tag{10.43}$$

$$V^\dagger = (\lambda^2 - c^2)\hat{V}(\mu, \lambda) \tag{10.44}$$

$$B = B^A + (1 - \mu^2)\hat{B}(\mu, \lambda) \tag{10.45}$$

$$\Phi^\dagger = \Phi^H + (\lambda^2 - c^2)\Phi(\mu, \lambda) \tag{10.46}$$

where $\hat{X}$, $\hat{W}$, $\hat{V}$, $\hat{B}$, $\hat{\Phi}$ are well behaved functions of μ, λ everywhere including both the axis where $\mu = \pm 1$ and on the horizon where $\lambda = c$ and where B^A and Φ^A are the gauge constants which have already been introduced (and where $\hat{X}$ and $\hat{V}$ are strictly positive in the neighbourhood of both the axis and the horizon).
positive in the neighbourhood of both the axis and the horizon).

It follows that the metric on $\overline{\mathscr{B}}$ takes the form

$$ds_{\text{II}}^2 = \Xi \left\{\frac{d\lambda^2}{\lambda^2 - c^2} + \frac{d\mu^2}{1 - \mu^2}\right\} \tag{10.47}$$

where the function Ξ, which is related to the function Σ by

$$\Xi = (\lambda^2 - c^2\mu^2)\,\Sigma, \tag{10.48}$$

is well behaved and non-zero everywhere including the north and south poles, provided that the field equations are satisfied, since (in consequence of the identities $X_1 = (\lambda^2 - c^2)^{-1}\hat{X}$ and $V_1^\dagger = (1 - \mu^2)^{-1}\hat{V}$) the consequential boundary conditions (10.18) and (10.24) for Σ imply that we can write

$$\Xi = \mu^2 [\hat{X} + (1 - \mu^2)\Xi_1] \tag{10.49}$$

and

$$\Xi = \lambda^2 \kappa^{-2} [\hat{V} + (\lambda^2 - c^2)\Xi_1^\dagger] \tag{10.50}$$

for suitably chosen functions Ξ_1, and $\Xi_1^\dagger$ which will be well behaved functions of λ and μ on the axis $\mu = \pm 1$ and the horizon $\lambda = c$ respectively, and hence both well behaved also on the poles where the axis and horizon meet.

We can now introduce co-ordinates $\hat{x}, \hat{y}$ (analogous to the previous cartesian co-ordinates x, y) and $\hat{w}^+$, $\hat{w}^-$ (analogous to the previous Kruskal type null co-ordinates w^+, w^-) simultaneously by the definitions

$$\left.\begin{aligned} \hat{x} &= \sqrt{1 - \mu^2}\, \sin \varphi^\dagger \\ \hat{y} &= \sqrt{1 - \mu^2}\, \cos \varphi^\dagger \\ \hat{w}^\pm &= \sqrt{\lambda^2 - c^2}\, e^{\pm \kappa t} \end{aligned}\right\} \tag{10.51}$$

which transforms (10.11) to the form

$$\begin{aligned} ds^2 = \Xi &\left\{ \frac{d\hat{w}^+ \, d\hat{w}^-}{\hat{w}^+ \hat{w}^- + c^2} + \frac{d\hat{x}^2 + d\hat{y}^2}{1 - (\hat{x}^2 + \hat{y}^2)} \right\} \\ &+ \kappa^{-1} \hat{W} [\hat{w}^- \, d\hat{w}^+ - \hat{w}^+ \, d\hat{w}^-] \, [\hat{x}\, dy - \hat{y}\, d\hat{x}] \\ &+ \tfrac{1}{4}\kappa^2 \, \Xi_1^\dagger (\hat{w}^- \, d\hat{w}^+ - \hat{w}^+ \, d\hat{w}^-)^2 - \Xi_1 [\hat{x}\, d\hat{y} - \hat{y}\, d\hat{x}]^2 \end{aligned} \tag{10.52}$$

which is well behaved everywhere in both the northern hemisphere (i.e. where $z > 0$) and in the southern hemisphere (i.e. where $z < 0$) including the horizon on which $\hat{w}^+ = 0$, (and it can also be extended over a past bounding horizon of $\ll \mathscr{I} \gg$ on which $\hat{w}^- = 0$, and over a Kruskal crossover axis $\hat{w}^+ = \hat{w}^- = 0$) although it is singular on the equator $z = 0$ since we have $\hat{x}^2 = \hat{y}^2 = 1$ there.

The field in these co-ordinates, is derived from the vector potential

$$\begin{aligned} A = \hat{B}[\hat{x}\, d\hat{y} - \hat{y}\, d\hat{x}] &+ \kappa^{-1} \, d\hat{\Phi} \wedge [\hat{w}^+ \, d\hat{w}^- - \hat{w}^- \, d\hat{w}^+] \\ &+ B^A \, d\varphi^\dagger + \Phi^H \, dt \end{aligned} \tag{10.53}$$

and therefore has the form

$$\begin{aligned} F = 2\, d\hat{B} \wedge [\hat{x}\, d\hat{y} - \hat{y}\, d\hat{x}] &+ \kappa^{-1} \, d\Phi \wedge [\hat{w}^+ \, d\hat{w}^- - \hat{w}^- \, d\hat{w}^+] \\ &+ 4\hat{B}\, d\hat{x} \wedge d\hat{y} + 2\kappa^{-1} \hat{\Phi}\, d\hat{w}^+ \wedge d\hat{w}^- \end{aligned} \tag{10.54}$$

which is also well behaved in each hemisphere, although not on the equator.

Since we have already covered the equator by the Kruskal type co-ordinate patch (10.35) (which is well behaved everywhere except at the poles), this completes the verification that (subject to the field equations) *the necessary conditions* (10.42) *to* (10.46) *are also completely sufficient for the whole of* $\ll \mathscr{I} \gg$ *and its future boundary horizon* $\mathscr{H}^+$ *to be regular*, and also that they imply the existence of a symmetric past boundary horizon $\mathscr{H}^-$, with a regular topologically two-spherical intersection $\mathscr{H}^+ \cap \mathscr{H}^-$ in a suitable extension of $\ll \mathscr{I} \gg \cup \mathscr{H}^+$ (although not necessarily in $\mathscr{M}$).

Before proceeding, it is worth commenting on the precise interpretation of the parameter c, which puzzled me when I originally worked out these results. The solution to this problem (as was pointed out to me by Bardeen) is provided by the generalized Smarr formula (9.29). It follows directly from (10.23) using the metric form (10.9) that the surface area $\mathscr{A}$ of the hole is given in terms of c by

$$\mathscr{A} = \frac{4\pi c}{\kappa} \tag{10.55}$$

Hence using the source free form of (9.29) we obtain

$$c = M - 2\Omega^H J - \Phi^H Q \tag{10.56}$$

11 Differential Equation Systems for the Vacuum Black Hole Problem

It is well known, from the work of Papapetrou and others, that the source free Einstein-Maxwell equations for the form (10.9) in conjunction with (7.41) can be reduced to a set of four independent equations for the four unknowns V, W, Φ, B and also that the asymptotic boundary conditions reduce to a corresponding set of four conditions on these four unknowns, in terms of the four corresponding conserved asymptotic quantities, namely the mass M, the angular momentum J, the electric charge Q, and the magnetic monopole moment P. What we have shown here so far is that the remaining regularity conditions for a system satisfying the conditions (10) can also be reduced to a set of four boundary conditions, namely (10.42), (10.43), (10.45), (10.46) for the four unknowns V, W, Φ, B only (since the fifth condition (10.44) for X is not independent but clearly a consequence of (10.42) and (10.43).

The tranditional form of the field equations, in terms of V, W, Φ, B is very convenient for studying properties at large asymptotic distances, but unfortunately these equations become singular on the ergosurface where V is zero. We can get over this difficulty by noticing that the metric form (10.9) and the electromagnetic field (7.41) are algebraicly invariant under a change in which V, W, Φ, B are replaced by $-X$, W, B, Φ respectively. By making this interchange we get over the singularity difficulty, since as has already been remarked, the causality condition ensures that X is never zero except of course in an entirely predictable way on the rotation axis.

The equations we obtain may be expressed conveniently in terms of the background metric

$$d\hat{s}_{\text{II}}^2 = \frac{d\lambda^2}{\lambda^2 - c^2} + \frac{d\mu^2}{1-\mu^2} \tag{11.1}$$

on $\overline{\mathscr{B}}$ as two Maxwell equations

$$\nabla\left\{\frac{X\nabla\phi - W\nabla B}{\rho}\right\} = 0 \tag{11.2}$$

$$\nabla\left\{\rho\frac{\nabla B}{X} + \frac{W}{X}\frac{[X\nabla\phi - W\nabla B]}{\rho}\right\} = 0 \tag{11.3}$$

and two Einstein equations

$$\nabla\left\{\frac{X\nabla W - W\nabla X}{\rho} + 4B\frac{[X\nabla\phi - W\nabla B]}{\rho}\right\} = 0 \tag{11.4}$$

$$\nabla\left\{\frac{\rho\nabla X}{X}\right\} + \frac{|X\nabla W - W\nabla X|^2}{\rho X^2} + 2\frac{|X\nabla\phi - W\nabla B|^2}{\rho X} + 2\rho\frac{|\nabla B|^2}{X} = 0 \tag{11.5}$$

Noting that the co-ordinates λ, μ can be related to co-ordinates r and θ which behave asymptotically like the traditional Schwarzschild spherical co-ordinates by the transformation $\lambda = r - M$ and $\mu = \cos\theta$, we can easily express the standard Papapetrou (1948) type boundary conditions in the terms of the requirement that W, B, Φ and $\lambda^{-2}X$ are well behaved functions of μ and λ^{-1} in the limit as $\lambda^{-1} \to 0$, and that they satisfy.

$$\Phi = Q\lambda^{-1} + 0(\lambda^{-2}) \tag{11.6}$$

$$B = -P\mu + 0(\lambda^{-1}) \tag{11.7}$$

$$W = -2J\lambda^{-1} + 0(\lambda^{-2}) \tag{11.8}$$

$$\lambda^{-2}X = (1+\mu^2)[1 + 2M\lambda^{-1}] + 0(\lambda^{-2}) \tag{11.9}$$

as $\lambda^{-1} \to 0$ (In imposing these conditions we have fixed the gauge of ϕ and B.)

The boundary conditions derived in the previous section can be expressed as

$$\frac{\partial\Phi}{\partial\mu} = 0(1) \qquad \frac{\partial\Phi}{\partial\lambda} = 0(1) \tag{11.10}$$

$$\frac{\partial B}{\partial\mu} = 0(1) \qquad \frac{\partial B}{\partial\lambda} = 0(1-\mu^2) \tag{11.11}$$

$$W = 0(1-\mu^2) \tag{11.12}$$

$$X = 0(1-\mu^2)\frac{1}{X}\frac{\partial X}{\partial \mu} = 1 + 0(1-\mu^2) \tag{11.13}$$

as $\mu \to \pm 1$, and

$$\frac{\partial \Phi}{\partial \lambda} = 0(1) \qquad \frac{\partial \Phi}{\partial \mu} + \Omega^H \frac{\partial B}{\partial \mu} = 0(\lambda^2 - c^2) \tag{11.14}$$

$$\frac{\partial B}{\partial \lambda} = 0(1) \qquad \frac{\partial B}{\partial \mu} = 0(1) \tag{11.15}$$

$$W + \Omega^H X = 0(\lambda^2 - c^2) \tag{11.16}$$

$$X = 0(1) \qquad \frac{1}{X} = 0(1) \tag{11.17}$$

as $\lambda \to c$.

Thus our results so far amount to a demonstration of the following key lemma:

To each domain of outer communications $\ll \mathscr{I} \gg$ in a spacetime in which the condition 10 of the previous section satisfied, there corresponds a cannonically defined solution of the system of equations (11.2) to (11.5) on the λ, μ plane in the co-ordinate range $-1 < \mu < 1, c < \lambda < \infty$ subject to the boundary condition (11.6) to (11.17), and conversely to each solution of this system there corresponds a canonically defined manifold $\ll \mathscr{I} \gg$ which can be extended to form a manifold $\mathscr{M}$ within which $\ll \mathscr{I} \gg$ is the domain of communications, and in which the conditions 10 are satisfied.

In short there is a one-one correspondence between source-free stationary axisymmetric black hole exterior solutions and solutions of the above system. The only known solutions of this system are the pure vacuum family of Kerr (1963), and its electromagnetic generalization given by Newman *et al.* (1965). These solutions are given by

$$\Phi = \frac{Qr - Pa\mu}{r^2 + a^2\mu^2} \tag{11.18}$$

$$B = \frac{P\mu(r^2 + a^2) - Qar(1-\mu^2)}{r^2 + a^2\mu^2} \tag{11.19}$$

$$W = \frac{-a(1-\mu^2)(2Mr - Q^2 - P^2)}{r^2 + a^2\mu^2} \tag{11.20}$$

$$X = (1-\mu^2)\left\{r^2 + a^2 + \frac{a^2(1-\mu^2)}{r^2 + a^2\mu^2}[2Mr - Q^2 - P^2]\right\} \tag{11.21}$$

where we have used the standard notation

$$r = \lambda + M \tag{11.22}$$

$$a = \frac{J}{M} \tag{11.23}$$

These solutions are uniquely specified by the values of Q, P, J, M (the gravitational part being invariant under a duality rotation in which Q and P are altered in such a way that the sum of their squares remains constant) which are restricted only by the condition that the boundary valve parameter c, which is given by

$$c^2 = M^2 - a^2 - P^2 - Q^2 \tag{11.24}$$

should remain real and positive. The other parameter involved in the specification of the boundary-conditions, namely the black hole rotation velocity, Ω^H, is given by

$$\Omega^H = \frac{a}{(M + c)^2 + a^2} \tag{11.25}$$

The fundamental conjecture which one would like to verify is that there exists a uniqueness theorem for the system according to which these should be the only solutions. However due to the essential non-linearity of the system it has not yet been possible to attain this objective, except in the static case. It is therefore worthwhile to start by investigating the truth of the weaker conjecture that there exists what has come to be known loosely as a *no-hair theorem* according to which a continuous variation (in a suitably defined sense) of a solution of this system should be uniquely determined by the corresponding continuous variation of the four conserved quantities Q, P, J, M, i.e. any solutions other than those above should also form discrete non-bifurcating families depending on at most these four parameters.

It is evident that if the no-hair conjecture is correct the parameters c, Ω^H and Φ^H which appear in the boundary conditions must be essentially redundant [part of this redundancy being made explicit by the relation (10.56)] merely duplicating the information on the rotation and scale of the black hole already given by M, J and Q. The scale parameter c is necessary to define the manifold on which we are working and therefore cannot easily be eliminated from the problem, so that it is more convenient instead to eliminate the mass parameter M which also controls the overall scale, and which governs only the higher order asymptotic corrections to X. However it turns out that Ω^H and Φ^H can be made to drop out of the problem altogether by recasting the problem in the manner described by Ernst (1967, 1969).

The Ernst method, whose basic purpose is to simplify the field equations, consists of taking advantage of the fact that the equation (11.2) can be interpreted as an integrability condition which is necessary and (under the present global conditions) sufficient for the existence of an electric pseudo-potential E satisfying

$$\left.\begin{aligned}(1-\mu^2)E_{,\mu} &= X\phi_{,\lambda} - WB_{,\lambda}\\ -(\lambda^2-c^2)E_{,\lambda} &= X\phi_{,\mu} - WB_{,\mu}\end{aligned}\right\} \tag{11.26}$$

and hence that (11.4) can be interpreted as an integrability condition which is necessary and sufficient for the existence of a *twist potential* Y satisfying

$$\begin{aligned}(1-\mu^2)Y_{,\mu} &= XW_{,\lambda} - WX_{,\lambda} + 2(1-\mu^2)(BE_{,\mu} - EB_{,\mu})\\ -(\lambda^2-c^2)Y_{,\lambda} &= XW_{,\mu} - WX_{,\mu} + 2(\lambda^2-c^2)(BE_{,\lambda} - EB_{,\lambda})\end{aligned} \tag{11.27}$$

The final stage in the Ernst procedure is to eliminate the variables ϕ and W in favour of the new variable E and Y. It was pointed out by Ernst (1967) that in the pure vacuum case (when E and B are zero) the resulting field equations can be derived from a very simple positive definite Lagrangian. It turns out that in the electromagnetic case the equations can still be derived from a positive definite (but not quite so simple) Lagrangian. The Lagrangian integral to be varied has the form

$$I = \int \mathscr{L} d\lambda\, d\mu \tag{11.28}$$

where the Lagrangian density is

$$\mathscr{L} = \frac{|\nabla X|^2 + |\nabla Y + 2(E\nabla B - B\nabla E)|^2}{2X^2} + 2\frac{|\nabla E|^2 + |\nabla B|^2}{X} \tag{11.29}$$

(the gradient contractions still being expressed in terms of the metric form given by (11.1)) which reduces to the Lagrangian given by Ernst when E and B are set equal to zero. It is to be remarked that this Lagrangian is invariant under a duality rotation in which E and B are replaced by $E\cos\alpha + B\sin\alpha$ and $B\cos\alpha - E\sin\alpha$ respectively, where α is a *constant* duality angle, as also is the resulting system of field equations. These equations consist of two Maxwell equations

$$E_B \equiv \nabla\left\{\frac{\rho\nabla B}{X}\right\} + \frac{\rho}{X^2}\nabla E[\nabla Y + 2(E\nabla B - B\nabla E)] = 0 \tag{11.30}$$

$$E_E \equiv \nabla\left\{\frac{\rho\nabla E}{X}\right\} - \frac{\rho}{X^2}\nabla B[\nabla Y + 2(E\nabla B - B\nabla E)] = 0 \tag{11.31}$$

and two Einstein equations

$$E_Y \equiv \nabla\left\{\frac{\rho}{X^2}[\nabla Y + 2(E\nabla B - B\nabla E)]\right\} = 0 \tag{11.32}$$

$$E_X \equiv \nabla\left\{\frac{\rho\nabla X}{X^2}\right\} + \rho\frac{|\nabla X|^2 + |\nabla Y + 2(E\nabla B - B\nabla E)|^2}{X^3}$$

$$+ 2\rho\frac{|\nabla B|^2 + |\nabla E|^2}{X^2} = 0 \qquad (11.33)$$

the two latter being obtained directly by variation with respect to Y and X respectively and the two former being obtained with the aid of (9.9), by variation with respect to B and E respectively.

The asymptotic boundary conditions, necessary and sufficient for asymptotic flatness, (but dropping the higher order restriction the explicity specifying the mass M in the condition on X) are that E, B, Y and $\lambda^{-2}X$, be well behaved function of λ^{-1} and μ in the limit as $\lambda^{-1} \to 0$, and that they satisfy

$$E = -Q\mu + 0(\lambda^{-1}) \qquad (11.34)$$

$$B = -P\mu + 0(\lambda^{-1}) \qquad (11.35)$$

$$Y = 2J\mu(3 - \mu^2) + 0(\lambda^{-1}) \qquad (11.36)$$

$$\lambda^{-2}X = (1 - \mu^2) + 0(\lambda^{-1}) \qquad (11.37)$$

as $\lambda^{-1} \to 0$, (with suitable choice of gauge for E and Y). The symmetry axis boundary conditions which are necessary and sufficient for (11.10), (11.11), (11.12), (11.13) to hold are that E, B, Y, X should be well behaved functions of μ and λ and that they satisfy the conditions

$$\frac{\partial E}{\partial \mu} = 0(1) \qquad \frac{\partial E}{\partial \lambda} = 0(1 - \mu^2) \qquad (11.39)$$

$$\frac{\partial B}{\partial \mu} = 0(1) \qquad \frac{\partial B}{\partial \lambda} = 0(1 - \mu^2)$$

$$\frac{\partial Y}{\partial \mu} + 2\left(E\frac{\partial B}{\partial \mu} - B\frac{\partial E}{\partial \mu}\right) = 0(1 - \mu^2) \qquad \frac{\partial Y}{\partial \lambda} = 0(1 - \mu^2) \qquad (11.40)$$

$$X = 0(1 - \mu^2) \qquad \frac{1}{X}\frac{\partial X}{\partial \mu} = 1 + 0(1 - \mu^2) \qquad (11.41)$$

as $\mu \to \pm 1$. The horizon boundary conditions are extremely simple in the present formulation, and are merely that E, B, Y, X be well behaved functions of λ and μ as $\lambda \to c$, with X non-vanishing, (the parameter Ω^H disappearing from the specification completely) i.e.

$$\frac{\partial E}{\partial \lambda} = 0(1) \qquad \frac{\partial E}{\partial \mu} = 0(1) \qquad (11.42)$$

$$\frac{\partial B}{\partial \lambda} = 0(1) \qquad \frac{\partial B}{\partial \mu} = 0(1) \qquad (11.43)$$

$$\frac{\partial Y}{\partial \lambda} = 0(1) \qquad \frac{\partial Y}{\partial \mu} = 0(1) \tag{11.44}$$

$$X = 0(1) \qquad X^{-1} = 0(1) \tag{11.45}$$

as $\lambda \to c$.

This Ernst formulation of the problem is completely equivalent to the previous formulation, i.e. *there is a one-one correspondence between solutions of (11.30) to (11.33) subject to (11.34) to (11.45), and solution of (11.2) to (11.5) subject to (11.6) to (11.17)* [and hence also a one one correspondence with domains of communications in manifolds satisfying the basic stationary axisymmetric source free black hole conditions (10)].

The values of the Ernst potentials E and Y in the Kerr-Newman solutions are

$$E = \frac{Q\mu(r^2 + a^2) - Par(1 - \mu^2)}{r^2 + a^2\mu^2} \tag{11.46}$$

$$Y = 2aM\mu(3 - \mu^2) - \frac{2a\mu(1 - \mu^2)}{r^2 + a^2\mu^2}[(Q^2 + P^2)r - Ma^2(1 - \mu^2)] \tag{11.47}$$

It is evident that the Ernst formulation of the problem is as effective in simplifying the boundary conditions as it is in simplifying the field equations. This simplification would not have occurred if we had worked with the complementary form of the equations using V instead of X. (On the other hand the actual solution is more complicated this way round.) The condition that the Lagrangian density is positive definite and well behaved also depends on using X, which is non-vanishing. The complementary form of the Lagrangian density, obtained by using $-V$ in place of X contains terms of opposite signs outside the ergosurface, and is singular on the ergosurface. Unfortunately even in the present case the positive definiteness of the Lagrangian cannot easily be used for drawing global conclusions, since the boundary conditions on X as $\mu \to \pm 1$ ensure that the integral I is divergent when taken over the whole domain.

The only case in which proper black hole uniqueness theorems (as opposed to no-hair theorems) are available at present are those in which the angular momentum is zero. In conjunction with theorem 4.1, the theorems of Israel (1969), (1968) and their refinements (cf. Muller zum Hagen, Robinson, Siefert (1972) give an almost complete demonstration that a black hole solution of given mass and charge is unique, (Schwarzschild or Riesmer-Nordstrom), provided that *staticity* is assumed. We shall complete this section by proving the following result.

Theorem 11 If the condition (10) is satisfied and if the angular momentum is zero, then the metric is necessarily static (in the sense that Y and hence also W

are zero) *and if furthermore the magnetic monopole moment is zero, then the electromagnetic field is also static* (in the sense that the magnetic potential B is zero).

Proof: We shall start by considering the special case when the magnetic monopole moment P is taken to be zero. The proof depends on the identity

$$\rho\frac{|\nabla Y + 2(E\nabla B - B\nabla E)|^2}{X^2} + 4\rho\frac{|\nabla B|^2}{X}$$

$$\equiv \nabla\left\{\frac{\rho(Y+2EB)[\nabla Y + 2(E\nabla B - B\nabla E)]}{X^2} + 4\rho\frac{B\nabla B}{X}\right\}$$

$$-(Y+2EB)E_Y - 4BE_B \tag{11.48}$$

When the field equations $E_Y = 0, E_B = 0$ are satisfied, the right hand side reduces to a divergence, and hence the integral of the left hand side over the entire domain $\overline{\mathscr{B}}$ can be expressed in terms of a boundary integral. Since the quantities Y, E, B, X^{-1} are all well behaved on the horizon, the presence of the factor ρ in the divergence ensures that the horizon gives no contribution to the boundary integral. The situation on the axis $\mu = \pm 1$ is more critical, since X^{-1} is singular there; however the boundary conditions ensure that Y and B are always constant on the axis, and hence in the particular case when J and P are zero Y and B are respectively zero also on the axis. Hence we see using (9.22) that the axis gives no contribution to the boundary integral when J and P are both zero, and it is also clear that under these conditions the same applies to the asymptotic boundary integral. It follows that each of the two non-negative terms on the left hand side must be zero everywhere. In conjunction with the boundary conditions, the vanishing of the second term implies that B is zero, and hence the vanishing of the first term, in conjunctions with the boundary conditions, implies that Y is zero also.

To see that Y must be zero whenever the angular momentum is zero, even when a magnetic monopole P (and hence also a non-zero magnetic contribution B to the field) is present, we have only to notice that it is always possible to reduce the magnetic monopole moment to zero by a duality rotation which of course leaves Y invariant. This completes the proof.

In conjunction with this result, Israel's theorems provide an almost complete proof of uniqueness for black holes with *zero angular momentum*, subject to the axisymmetry assumption.

[Approaching the same conclusions from a slightly different angle, the closely related generalised Hawking–Lichnerowicz Theorem 6.2, in conjunction with Israels' theorems, provides an almost complete proof of uniqueness for black holes with *zero angular velocity*, subject to the assumption that the hole boundary ergosurface is the outer ergosurface.]

12 The Pure Vacuum No-Hair Theorem

In order to make progress when non-zero angular momentum is present, we shall restrict our attention in this section to the purely gravitational case where there is no electromagnetic field present, i.e. when E and B are zero.

In this case the field equations reduce to

$$G_X(X, Y) \equiv \nabla\left\{\rho \frac{\nabla X}{X^2}\right\} + \rho \frac{|\nabla X|^2 + |\nabla Y|^2}{X^3} = 0 \qquad (12.1)$$

$$G_Y(X, Y) \equiv \nabla\left\{\rho \frac{\nabla Y}{X^2}\right\} = 0 \qquad (12.2)$$

and the boundary conditions are simply that as $\lambda^{-1} \to 0$, Y and $\lambda^{-2}X$ must be well behaved functions of λ^{-1} and μ with

$$\lambda^{-2}X = (1 - \mu^2)[1 + 0(\lambda^{-1})] \qquad (12.3)$$

$$Y = 3J\mu(1 - \mu^2) + 0(\lambda^{-1}) \qquad (12.4)$$

that as $\mu \to \pm 1$, X and Y must be well behaved functions of μ and λ with

$$X = 0(1 - \mu^2) \qquad X^{-1}\frac{\partial X}{\partial \mu} = 1 + 0(1 - \mu^2) \qquad (12.5)$$

$$\frac{\partial Y}{\partial X} = 0(1 - \mu^2) \qquad \frac{\partial Y}{\partial \mu} = 0(1 - \mu^2) \qquad (12.6)$$

and that as $\lambda \to c$, X and Y must be well behaved function of λ, μ with no other restrictions than

$$X = 0(1) \qquad X^{-1} = 0(1) \qquad (12.7)$$

$$\frac{\partial Y}{\partial \lambda} = 0(1) \qquad \frac{\partial Y}{\partial \mu} = 0(1) \qquad (12.8)$$

Simple as it is, this system remains essentially non-linear, and it is therefore not easy to obtain a solution of the full uniqueness problem. However it turns out that we can, without too much difficulty, prove the truth of the fundamental no-hair conjecture by showing that *continuous variations of these solutions are uniquely determined by the corresponding continuous variations of the scale parameter c and the angular momentum parameter J.* We use the term *continuous* in this statement to denote a restriction on solution families under discussion sufficient to ensure that if $\{X_1, Y_1\}(J, c)$ and $\{X_2, Y_2\}(J, c)$ are two *distinct* continuous families of solutions, which are functions over a set of values of the parameter pairs (J, c), and which tend (in some appropriate topology on the space of functions) to a common limit $\{X_0, Y_0\}$ as (J, c) tends to (J_0, c_0) then there exists at least one subset of values of (J, c) such that over this subset the difference $\{X_1 - X_2, Y_1 - Y_2\}(J, c)$ *has a well defined limit direction*, represented by a non-zero function-space tangent vector $\{\dot{X}, \dot{Y}\}$, as (J, c) tends to (J_0, c_0), (in the sense that there exists some parameter β say which is a function of (J, c) on the subset, tending to zero as (J, c) tends to (J_0, c_0), and such that $\beta^{-1}\{X_1 - X_2, Y_1 - Y_2\}$ tends to $\{\dot{X}, \dot{Y}\}$ as $\{X_1 - X_2, Y_1 - Y_2\}$ tends to zero).

In view of a comment made by Wald we emphasize that this continuity requirement has nothing to do with analyticity or even piece-wise analyticity; in fact in a *finite* dimensional vector space the continuity property postulated above would be an *automatic* consequence of the ordinary continuity condition that any neighbourhood of the point $\{X_0, Y_0\}$ contains members of the families $\{X_1, Y_1\}$ and $\{X_2, Y_2\}$ other than $\{X_0, Y_0\}$ (in consequence of the fact that in a finite dimensional vector space, as opposed to an infinite dimensional Banach space, the unit sphere is compact). We shall not attempt to investigate here the precise restrictivity of this continuity requirement in terms of explicitly defined Banach space structure of function space. For anyone who may be interested in following up such fine mathematical points, we refer to a discussion of such questions in the context of a rather simple kind of non-linear partial differential equation system by Berger (1969).

The continuity property we have postulated implies that if there are two distinct families $\{X_1, Y_1\}$, $\{X_2, Y_2\}$ of solutions bifurcating from some given solution $\{X_0, Y_0\}$ then there will exist some subset of values, parametrized as functions over some corresponding subset of values of a parameter β such that

$$X_2(J, c) = X_1(J, c) + \beta\dot{X} + 0(\beta) \tag{12.9}$$

$$Y_2(J, c) = Y_1(J, c) + \beta\dot{Y} + 0(\beta) \tag{12.10}$$

as $\beta \to 0$ over this subset for some (not everywhere vanishing) functions $\dot{X}$, $\dot{Y}$. [We emphasize again that this condition does not imply any assumption that $\{X_1, Y_1\}$ and $\{X_2, Y_2\}$ or the associated parameter values (J, c) are analytic or even differentiable functions of β, so that the present treatment automatically covers "higher order bifurcation" in which the derivatives of $\{X_2 - X_1, Y_2 - Y_1\}$ with respect to (J, c)—if indeed such derivatives exist—are zero at (J_0, c_0).]

On substitution of the above expression into the operators defining the field equations we obtain

$$G_X(X_2, Y_2) = G_X(X_1, Y_1) + \beta\dot{G}_X(X_1, Y_1; \dot{X}, \dot{Y}) + 0(\beta) \tag{12.11}$$

$$G_Y(X_2, Y_2) = G_Y(X_1, Y_1) + \beta\dot{G}_Y(X_1, Y_1; \dot{X}, \dot{Y}) + 0(\beta) \tag{12.12}$$

as $\beta \to 0$ where the linearized perturbation operators $\dot{G}_X(X, Y; \dot{X}, \dot{Y})$ and

$\dot{G}_Y(X, Y; \dot{X}, \dot{Y})$ are given by

$$\dot{G}_X \equiv \nabla\left\{\rho \frac{\nabla \dot{X}}{X^2} - 2\rho \frac{\dot{X}\nabla X}{X^3}\right\} + 2\rho\left\{\frac{\nabla X \nabla \dot{X} + \nabla Y \nabla \dot{Y}}{X^3}\right\} - 3\rho\left\{\frac{|\nabla X|^2 + |\nabla Y|^2}{X^4}\right\}\dot{X} \tag{12.13}$$

$$\dot{G}_Y \equiv \nabla\left\{\rho \frac{\nabla \dot{Y}}{X^2} - 2\rho \frac{\dot{Y}\nabla X}{X^3}\right\} \tag{12.14}$$

Using the fact that $\{X_2, Y_2\}$ and $\{X_1, Y_1\}$ are solutions of the exact field equations, $G_X(X_2, Y_2) = G_Y(Y_2, Y_2) = 0$ and $G_X(X_1, Y_1) = G_Y(X_1, Y_1) = 0$, and dividing by β, we obtain

$$\dot{G}_X(X_1, Y_1; \dot{X}, \dot{Y}) + 0(1) = 0 \tag{12.15}$$

$$\dot{G}_Y(X_1, Y_1; \dot{X}, \dot{Y}) + 0(1) = 0 \tag{12.16}$$

Hence we finally deduce that we must have

$$\dot{G}_X(X, Y; \dot{X}, \dot{Y}) = 0 \tag{12.17}$$

$$\dot{G}_Y(X, Y; \dot{X}, \dot{Y}) = 0 \tag{12.18}$$

when X, Y take the values X_0, Y_0 of the solutions at which the hypothetical bifurcation takes place.

Similarly by substituting (12.9), (12.10) into the boundary conditions (12.3), (12.4), (12.5), (12.6), (12.7), (12.8), dividing by β, and taking the limit, we obtain the corresponding linearized boundary conditions on $\dot{X}$, $\dot{Y}$ in the form of conditions that as $\lambda \to \infty = \dot{Y}$ and $\lambda^{-2}\dot{X}$ be well behaved functions of μ and λ^{-1} with

$$\lambda^{-2}\dot{X} = 0(\lambda^{-1}) \tag{12.19}$$

$$\dot{Y} = 0(\lambda^{-1}) \tag{12.20}$$

that as $\mu \to \pm 1$, $\dot{X}$ and $\dot{Y}$ must be well behaved functions of λ, μ with

$$\dot{X} = 0(1 - \mu^2) \tag{12.21}$$

$$\frac{\partial \dot{Y}}{\partial \lambda} = 0(1 - \mu^2) \qquad \frac{\partial \dot{Y}}{\partial \mu} = 0(1 - \mu^2) \tag{12.22}$$

and that as $\lambda \to c$, $\dot{X}$, $\dot{Y}$ must be well behaved functions of λ, μ with no other restrictions than

$$\dot{X} = 0(1) \tag{12.23}$$

$$\frac{\partial \dot{Y}}{\partial \lambda} = 0(1) \qquad \frac{\partial \dot{Y}}{\partial \mu} = 0(1) \tag{12.24}$$

We prove that the families $\{X_1, Y_1\}$ and $\{X_2, Y_2\}$ must be identical, i.e. that there can be no bifurcations by showing that if $\dot{X}$, $\dot{Y}$ satisfy the equations (12.17), (12.18) and the boundary conditions (12.19), (12.20), (12.21), (12.22), (12.23), (12.24) they must always be zero everywhere and so cannot determine a well defined direction in function space. We point out that the converse deduction could *not* be made i.e. although the *non-existence* of a non-zero linearized bifurcation solution *is* sufficient to rule out the existence of a bifurcating family of exact solutions, the *existence* of a non-zero linearized bifurcation solution would *not necessarily* imply the existence of a bifurcating family of solutions of the full non-linear system; it would only be in this latter case that higher order analysis might be relevant.

The proof depends on the identity

$$\rho\left|\nabla\left(\frac{\dot{X}}{X}\right)+\frac{\dot{Y}\nabla Y}{X^2}\right|^2+\rho\left|\nabla\left(\frac{\dot{Y}}{X}\right)-\frac{\dot{X}\nabla Y}{X^2}\right|^2+\rho\left|\frac{\dot{X}\nabla Y-\dot{Y}\nabla X}{X^2}\right|^2$$
$$\equiv\nabla\left\{\rho\left[\frac{\dot{X}}{X}\nabla\left(\frac{\dot{X}}{X}\right)+\frac{\dot{Y}}{X}\nabla\left(\frac{\dot{Y}}{X}\right)\right]\right\} \tag{12.23}$$
$$+\frac{2\dot{X}^2+\dot{Y}^2}{X}G_X+\frac{\dot{X}\dot{Y}}{X}G_Y-\dot{X}\dot{G}_X-\dot{Y}\dot{G}_Y$$

which is analogous to that of the previous section in that the right hand reduces to a divergence when the field equations $G_X = G_Y = 0$ and the linearized field equations $\dot{G}_X = \dot{G}_Y = 0$ are satisfied, so that the integral of the left hand side over the entire domain $\bar{\mathscr{B}}$ can be expressed in terms of a boundary integral, which can be seen to vanish in consequence of the boundary conditions (12.19), (12.20), (12.21), (12.22), (12.23), (12.24). Since each term on the left hand side is non-negative, it follows that they must all three be zero everywhere. Thus we obtain three linear first order differential equations for $\dot{X}$ and $\dot{Y}$, from which, by taking linear combinations we can obtain $\nabla\dot{Y} = \dot{Y}X^{-1}\nabla X$ which implies (by the boundary condition that $\dot{Y}$ is zero on the axis, together with the standard uniqueness theorem for the solution of a homogeneous gradient equation of this type) that $\dot{Y}$ is zero everywhere. The remaining first order differential equations then imply directly that $\dot{X}$ is zero also, which completes the proof of the following result:

THEOREM 12 (No-Hair Theorem) The mathematically possible domains of communications $\ll\mathscr{J}\gg$ of space-time manifolds $\mathscr{M}$ satisfying the conditions 10, fall in to discrete continuous families depending on at most the two parameters J and c.

COROLLARY There is at most one such family for which the angular momentum parameter J can take the value zero.

This family consists of course of the Kerr vacuum solution with $a^2 < M^2$ for which J can take any value without restriction for arbitrary positive valves of C, the mass parameter M being given as a function of C by $M^2 = \frac{1}{2}\{C^2 + (C^2 + 4J^2)^{1/2}\}$. The corollary follows immediately from the staticity Theorem 11 and the appropriate form of the Israel theorem establishing uniqueness in the static case.

We can give a simple explicit proof of the relevant restriction of Israel's theorem (i.e. the restriction to the axisymmetric pure vacuum case) by noticing that in the static case when Y is zero the system $G_X = G_Y = 0$ reduces simply to

$$\nabla\{\rho\nabla(\ln X)\} = 0 \tag{12.24}$$

which is *linear* in $\ln X$, and hence will also be satisfied if X is replaced by the quotient X_1/X_2 of any two different solutions X_1 and X_2. Hence using the further identity

$$\rho\,|\nabla(\ln X)|^2 \equiv \nabla\{\rho \ln X\,\nabla(\ln X)\} - (\ln X)\,\nabla\{\rho\nabla(\ln X)\} \tag{12.25}$$

we can deduce that the quotient of any two solutions of (12.24) satisfies

$$\rho\left|\nabla\left[\ln\left(\frac{X_1}{X_2}\right)\right]\right|^2 = \nabla\left\{\rho \ln\left(\frac{X_1}{X_2}\right)\nabla\left[\ln\left(\frac{X_1}{X_2}\right)\right]\right\} \tag{12.26}$$

By integrating over the whole domain $\overline{\mathscr{B}}$, we can again use the relevant boundary conditions (12.3), (12.5), (12.7) to deduce that the non-negative term on the left hand side is zero everywhere, which proves that $X_1 = X_2$ i.e. that the solution is unique, and must therefore be the relevant Schwarzschild solution for which $X = (\lambda + C)^2(1 - \mu^2)$.

We remark that although this proof is much more restricted than Israel's original (1967) demonstration in that it assumes axisymmetry at the outset, it has the advantage of being independent of Israel's assumption that the gradient of V is nowhere vanishing, instead making use of the condition that X is nowhere vanishing (except on the axis).

(In a *static* as opposed to merely stationary domain, it is clear that X must be non-zero except on the axis, whether or not the global causality requirement is satisfied in $\mathscr{M}$.)

It has in fact recently been shown by Muller-zum-Hager, Robinson and Siefert that Israel's postulate that the gradient of V be non-zero can in fact be dispensed with in the general (non-axisymmetric) vacuum Israel theorem, albeit at the expense of considerable technical complexity in the demonstration. It would be desirable to extend both the general demonstration of Muller-zum-Hagen, Robinson and Siefert, and the very much simpler restricted demonstration given here, to cover the electromagnetic case. Unfortunately, the demonstration given in this section depends essentially on the *linearization* of the system which can be achieved when both the angular momentum and the electromagnetic field are zero, but which fails when even a static electric field is present.

13 Unsolved Problems

Having succeeded in proving the truth of the no-hair conjecture in the pure vacuum case, we now consider the question of the electromagnetic generalization, according to which any two continuous families

$$\{X_1, Y_1, E_1, B_1\}(c, J, Q, P) \quad \text{and} \quad \{X_2, Y_2, E_2, B_2\}(c, J, Q, P)$$

of solutions $\{X, Y, E, B\}$ of the system (11.30) to (11.45), parametrized by the four quantities c, J, Q, P should coincide if they have any solution, $\{X_0, Y_0, E_0, B_0\}$ say, in common. As before we may assume that if they do not coincide there exist non-zero perturbation functions $\dot{X}, \dot{Y}, \dot{E}, \dot{B}$ such that for a subset of values of some parameter there exists a corresponding subset of values of c, J, Q, P for which the equations (12.9) and (12.10) and the further equations

$$E_2 = E_1 + \beta\dot{E} + 0(\beta) \tag{13.1}$$

$$B_2 = B_1 + \beta\dot{B} + 0(\beta) \tag{13.2}$$

as $\beta \to 0$ are satisfied, where $\{X_1, Y_1, E_1, B_1\}$ and $\{X_2, Y_2, E_2, B_2\}$ both tend to $\{X_0, Y_0, E_0, B_0\}$ as $\beta \to 0$. By the same reasoning as used before the perturbation functions $\dot{X}, \dot{Y}, \dot{E}, \dot{B}$ must satisfy the linearized equations

$$E_X(X, Y, E, B; \dot{X}, \dot{Y}, \dot{E}, \dot{B}) \tag{13.3}$$

$$E_Y(X, Y, E, B; \dot{X}, \dot{Y}, \dot{E}, \dot{B}) \tag{13.4}$$

$$E_E(X, Y, E, B; \dot{X}, \dot{Y}, \dot{E}, \dot{B}) \tag{13.5}$$

$$E_B(X, Y, E, B; \dot{X}, \dot{Y}, \dot{E}, \dot{B}) \tag{13.6}$$

when $\{X, Y, E, B\}$ takes the value $\{X_0, Y_0, E_0, B_0\}$, where

$$\begin{aligned} \dot{E}_X \equiv\ & \nabla\left\{\rho\frac{\nabla\dot{X}}{X^2} - 2\rho\frac{\dot{X}\nabla X}{X^2} - 3\frac{\dot{X}}{X^3}\left[|\nabla X|^2 + |\nabla Y + 2(E\nabla B - B\nabla E)|^2\right]\right\} \\ & + \frac{2}{X^2}\nabla \mathrm{X}\nabla\dot{X} + \frac{2}{X^2}[\nabla Y + 2(E\nabla B - B\nabla E)] \\ & \times [\nabla\dot{Y} + 2(E\nabla\dot{B} - B\nabla\dot{E} + \dot{E}\nabla B - \dot{B}\nabla E)] \\ & - \frac{2\dot{X}}{X^2}\left[|\nabla E|^2 + |\nabla B|^2\right] + \frac{2}{X}[\nabla E\nabla\dot{E} + \nabla B\nabla\dot{B}] \end{aligned} \tag{13.7}$$

$$\begin{aligned} \dot{E}_Y \equiv\ & \nabla\left\{\frac{\rho}{X^2}[\nabla\dot{Y} + 2(E\nabla\dot{B} - B\nabla\dot{E} + \dot{E}\nabla B - \dot{B}\nabla E)]\right. \\ & \left. - 2\rho\frac{\dot{X}}{X}[\nabla Y + 2(E\nabla B - B\nabla E)]\right\} \end{aligned} \tag{13.8}$$

$$\dot{E}_E \equiv \nabla\left\{\rho\frac{\nabla\dot{E}}{X} - \rho\frac{\dot{X}\nabla E}{X^2}\right\} - \rho\left(\frac{\nabla\dot{B}}{X^2} - 2\frac{\dot{X}\nabla B}{X^3}\right)[\nabla Y + 2(E\nabla B - B\nabla E)]$$

$$+\frac{\rho}{X^2}\nabla B[\nabla\dot{Y} + 2(E\nabla\dot{B} - B\nabla\dot{E} + \dot{E}\nabla B - \dot{B}\nabla E)] \qquad (13.9)$$

$$\dot{E}_B \equiv \nabla\left\{\rho\frac{\nabla\dot{B}}{X} - \rho\frac{\dot{X}\nabla B}{X^2}\right\} + \rho\left(\frac{\nabla\dot{E}}{X^2} - 2\frac{\dot{X}\nabla E}{X^3}\right)[\nabla Y + 2(E\nabla B - B\nabla E)]$$

$$-\frac{\rho}{X^2}\nabla E[\nabla\dot{Y} + 2(E\nabla\dot{B} - B\nabla\dot{E} + \dot{E}\nabla B - \dot{B}\nabla E)] \qquad (13.10)$$

The corresponding homogeneous boundary conditions, derived from (11.34) to (11.45) are that $\lambda^{-2}\dot{X}$, $\dot{Y}, \dot{E}, \dot{B}$ be well behaved functions of μ, λ^{-1} satisfying

$$\lambda^{-2}\dot{X} = 0(\lambda^{-1}) \qquad (13.11)$$

$$\dot{Y} = 0(\lambda^{-1}) \qquad (13.12)$$

$$\dot{E} = 0(\lambda^{-1}) \qquad (13.13)$$

$$\dot{B} = 0(\lambda^{-1}) \qquad (13.14)$$

as $\lambda^{-1} \to 0$, that $\dot{X}$, $\dot{Y}, \dot{E}, \dot{B}$ be well behaved function of μ, λ satisfying

$$\dot{X} = 0(1-\mu^2) \qquad (13.15)$$

$$\left.\begin{array}{l} \dfrac{\partial\dot{Y}}{\partial\lambda} = 0(1-\mu^2) \\ \dfrac{\partial\dot{Y}}{\partial\mu} + 2\left(E\dfrac{\partial\dot{B}}{\partial\mu} - B\dfrac{\partial\dot{E}}{\partial\mu} + \dot{E}\dfrac{\partial B}{\partial\mu} - \dot{B}\dfrac{\partial E}{\partial\mu}\right) = 0(1-\mu^2) \end{array}\right\} \qquad (13.16)$$

$$\frac{\partial\dot{E}}{\partial\lambda} = 0(1-\mu^2) \qquad \frac{\partial\dot{E}}{\partial\mu} = 0(1) \qquad (13.17)$$

$$\frac{\partial\dot{B}}{\partial\lambda} = 0(1-\mu^2) \qquad \frac{\partial\dot{B}}{\partial\mu} = 0(1) \qquad (13.18)$$

as $\mu \to \pm 1$, and that $\dot{X}$, $\dot{Y}, \dot{E}, \dot{B}$ be well behaved functions of λ, μ subject only to

$$\dot{X} = 0(1) \qquad (13.19)$$

$$\frac{\partial\dot{Y}}{\partial\lambda} = 0(1) \qquad \frac{\partial\dot{Y}}{\partial\mu} = 0(1) \qquad (13.20)$$

$$\frac{\partial\dot{E}}{\partial\lambda} = 0(1) \qquad \frac{\partial\dot{E}}{\partial\mu} = 0(1) \qquad (13.21)$$

$$\frac{\partial\dot{B}}{\partial\lambda} = 0(1) \qquad \frac{\partial\dot{B}}{\partial\mu} = 0(1) \qquad (13.22)$$

as $\lambda \to C$.

I have not found any way of determining whether or not this system can have non-zero solutions $\dot{X}$, $\dot{Y}$, $\dot{E}$, $\dot{B}$ for general values of X, Y, E, B subject to (11.30) to (11.45). If non-zero solutions do exist for some eigenvalues of c, J, Q, P it would suggest†, but would not prove that the solution family bifurcates at the corresponding solutions. There is however a special case which is much more tractable, namely the case where the electromagnetic field in the unperturbed solution is zero, i.e. when $E = B = 0$, since in this case the general form of the Einstein-Maxwell equations ensures (independently of the special symmetry conditions assumed in the present problem) that the linearized equations for the gravitational perturbation decouple from those for the electromagnetic perturbation, the latter reducing simply to a set of pure Maxwell equations in the curved space background. Advantage has been taken of this by Wald (1971), and also independently by Ipser (1971), who have shown that in the particular case of a pure (non-electromagnetic) Kerr solution background the perturbation solution of the Maxwell equations are indeed zero when the corresponding perturbed values of the charge Q and monopole moment P are zero, from which it follows that even if bifurcations from the Kerr-Newman family do exist, they cannot start from the pure vacuum members.

We shall conclude this course by presenting a generalization of this Wald-Ipser theorem showing that the conclusion holds for electromagnetic perturbations about *any* pure vacuum black hole solution family satisfying conditions (10) even if it is not the Kerr family. That is to say *it is true not only for the Kerr-Newman family but for any other family of electromagnetic black hole solutions of the system* (11.30) to (11.45) (if there are any others) *that no bifurcation can take place starting from the pure vacuum members.*

To prove this we use the fact that when E and B are zero, the full linearized equations $\dot{E}_E = 0, \dot{E}_B = 0$ reduce simply to the pure Maxwell equations

$$M_E \equiv \nabla\left\{\rho\frac{\nabla\dot{E}}{X}\right\} - \rho\frac{\nabla Y \nabla\dot{B}}{X^2} = 0 \tag{13.24}$$

$$M_B \equiv \nabla\left\{\rho\frac{\nabla\dot{B}}{X}\right\} + \rho\frac{\nabla Y \nabla\dot{B}}{X^2} = 0 \tag{13.25}$$

(the other two equations $\dot{E}_X = 0, \dot{E}_Y = 0$ reducing to the equations $\dot{G}_X = 0$, $\dot{G}_Y = 0$ which we have already studied).

To establish that the only solution of (13.24), (13.25) subject to (13.17) to (13.22) and (12.1) to (12.8) is $\dot{E} = \dot{B} = 0$, we use the identity

$$\frac{1}{X^3}\{|X\nabla\dot{E} - B\nabla\dot{Y}|^2 + |X\nabla\dot{B} - \dot{E}\nabla Y|^2\} + \frac{1}{X}\left\{\left|\nabla\left(\frac{\dot{E}}{X}\right)\right|^2 + \left|\nabla\left(\frac{\dot{B}}{X}\right)\right|^2\right\}$$
$$\equiv \nabla\left\{\rho\nabla\left(\frac{\dot{E}^2 + \dot{B}^2}{X}\right)\right\} + (\dot{E}^2 + \dot{B}^2)G_X - 2(\dot{E}M_E + \dot{B}M_B) \tag{13.26}$$

† The existence of a perturbation eigenvalue might be a symptom of the setting in of a dynamic instability even where no *exactly stationary* bifurcation exists.

When the Maxwell equations $M_E = 0, M_B = 0$ are satisfied, and when the unperturbed pure vacuum field equation $G_X = 0$ is satisfied, the right hand side reduces to a divergence whose integral over the whole domain is a boundary integral which can easily be seen to vanish by the boundary conditions. It follows that each of the non-negative terms on the left hand side is zero. Hence, with further use of the boundary conditions, it is clear that $\dot{E}$ and $\dot{B}$ themselves must be zero everywhere, as required.

By an examination (based on the work of Vishveshwara (1968)) of stationary perturbations of the Schwarzschild and Reissner Nordstrom solutions it has also been shown by Wald (1972) that there can be no bifurcation *from the spherical members* of the Kerr–Newman family of electromagnetic black hole solutions. (This work is a generalization of an earlier perturbation analysis, now superseded by the Vaccuum No Hair Theorem 12, which was carried out by Hartle and Thorne (1968).)

Although it represents an amusing mathematical challenge, the problem of generalizing these partial results to a complete electromagnetic no-hair theorem—or finding a counter-example—is less important (and probably less difficult) than the problem of generalizing the vacuum no-hair theorem to an absolute uniqueness theorem for vacuum black holes—or finding a counter example, i.e. a family of non-Kerr pure vacuum black holes with the pathological property that the angular momentum cannot be varied to zero. Neither of these problems is as pressing as that of determining the *stability* of the black hole equilibrium states which we have been discussing. (In the unlikely event that non-Kerr stationary pure vacuum black holes are discovered, the physical implications would be less startling if it were to turn out that they were all unstable.)

Bibliography

Bardeen, J. (1970) *Ap. J.* **162**, 71.
Bardeen, J. (1971) *Ap. J.* **167**, 425.
Bardeen, J., Carter, B. and Hawking, S. W. (1972) *Black Hole Mass Formulae.* Preprint Inst. of Astronomy, Cambridge.
Beckenstein, J. (1972) Phd. Thesis, Princeton.
Berger, M. (1969) *Bifurcation Theory and Nonlinear Eigenvalue Problems,* 113. Ed. B. Keller and S. Autman, (Benjamin).
Boyer, R. H. and Price, T. G. (1965) *Proc. Camb. Phil. Soc.* **62**, 531.
Boyer, R. H. and Lindquist, R. W. (1967) *J. Math. Phys.* **8**, 265.
Boyer, R. H. (1968) *Proc. Roy. Soc. A.*
Carter, B. (1968) *Phys. Rev.* **174**, 1559.
Carter, B. (1969) *J. Math. Phys.* **10**, 70.
Carter, B. (1970) *Commun. Math. Phys.* **17**, 233.
Carter, B. (1971a) *General Relativity and Gravitation* **1**, 349.
Carter, B. (1971b) *Phys. Rev. Lett.* **26**, 331.
Carter, B. (1972a) *Domains of Stationary Communication in Spacetime.* Preprint Inst. of Astronomy, Cambridge.
Carter, B. (1972b) *Nature.*
Carter, B. (1972c) *The Electrical Equilibrium of a Black Hole.* Preprint Inst. of Astronomy, Cambridge.

Carter, B. (1972d) *Variation Principle for a Rotating Solid Star.* Preprint Inst. of Astronomy, Cambridge.
Chandrasekhar, S. (1931) *Ap. J.* **74**, 81.
Chandrasekhar, S. (1935) *Mon. Not. R.A.S.* **95**, 207.
Chandrasekhar, S. (1969) *Ellipsoidal Figures of Equilibrium*, Yale.
Chandrasekhar, S. (1970) *Ap. J.* **161**, 561.
Chandrasekhar, S. and Friedman, J. (1972) *Ap. J.* **175**, 379.
Christodoulou D. (1970) *Phys. Rev. Lett.* **25**, 1596.
Christodoulou, D. and Ruffini, R. (1971) *Phys. Rev.* **4**, 2552.
Doroskevitch, A. G., Zel'dovich, Ya. B. and Novikov, L. D. (1966), *Sov. Phys. J.E.T.P.* **22**, 122.
Ernst, F. J. (1968a) *Phys. Rev.* **167**, 175.
Ernst, F. J. (1968b) *Phys. Rev.* **167**, 1175.
Geroch, R. P. (1970) *J. Math. Phys.* **11**, 437, 2580.
Hajicek, P. (1971) *General Theory of Ergospheres.* Preprint, Bern.
Hartle, J. B. and Sharp, D. H. (1967) *Ap. J.* **147**, 317.
Hartle, J. B. and Hawking, S. W. (1972) *Commun. Math. Phys.* **27**, 283.
Hartle, J. B. and Thorne, K. S. (1968), *Ap. J.* **153**, 807.
Hawking, S. W. (1967) *Proc. Roy. Soc. A.* **300**, 187. *Commun. Math. Phys.* **27**, 283.
Hawking, S. W. and Penrose, R. (1970) *Proc. Roy. Soc. A.* **314**, 529.
Hawking, S. W. (1971) *Phys. Rev. Lett.* **26**, 1344.
Hawking, S. W. (1972) *Commun. Math. Phys.* **25**, 152.
Ipser, J. R. (1971) *Electromagnetic Perturbations of a Kerr Metric Black Hole. Phys. Rev. Letters,* **27**, 529.
Israel, W. (1967) *Phys. Rev.* **164**, 1776.
Israel, W. (1968) *Commun. Math. Phys.* **8**, 245.
Kerr, R. (1963) *Phys. Rev. Lett.* **11**, 237.
Kundt, W. and Trumper, M. (1966) *Z. Physik* **192**, 419.
Landau, L. D. (1932) *Phys. Zs. Swjetunion* **1**, 285.
Landau, L. D. and Lidshitz (1962) *Classical Theory of Fields,* Addison Wesley.
Lewis, T. (1937) *Proc. Roy. Soc. A* **136**, 176.
Lichnerowicz, A. (1955) *Theories Relativistes de Gravitation et de l'Electromagnetism* (Paris: Mason).
Lynden-Bell, D. (1969) *Nature* **223**, 690.
Muller-Zum-Hagen, H., Robinson, P. and Seifert, H. J. (1972) *Black Holes in Static Vacuum Spacetimes.* Preprint, Kings College, London.
Oppenheimer, J. R., Serber, R. (1938) *Phys. Rev.* **54**, 530.
Oppenheimer, J. R., Volkoff, G. (1939) *Phys. Rev.* **55**, 370.
Ostriker, J. P., Bodenheimer, P. and Lynden Bell, D. (1966) *Phys. Rev. Letters* **17**, 816.
Ostriker, J. P. and Tassoul, J. L. (1969) *Ap. J.* **155**, 987.
Papapetrou, A. (1966) *Ann. Inst. H. Poincaré* **A4**, 83.
Penrose, R. (1965) *Phys. Rev. Lett.* **14**, 57.
Penrose, R. (1967) *Battelle Rencontres* Ed. C. M. De Witt and J. A. Wheeler (Benjamin).
Penrose, R. (1969) *Riv. Nuovo Cimento I,* **1**, 252.
Vishveshwara (1968) *J. Math. Phys.* **9**, 1319.
Vishveshwara (1970), *Phys. Rev.* **D1**, 2870.
Wald, R. (1971), *Static Axially Symmetric Electromagnetic Fields in Kerr Spacetime.* Preprint, Princeton University.
Wald, R. (1972), *J. Math. Phys.* **13**, 490.