Energy extremality in the presence of a black hole[1]

Rafael D. Sorkin [a,b] and Madhavan Varadarajan [a,c]

[a]Department of Physics, Syracuse University,
Syracuse, New York 13244-1130, U.S.A.
[b]Instituto de Ciencias Nucleares, UNAM
A. Postal 70-543, México D.F. 04510, México
[c]Department of Physics, University of Utah,
Salt Lake City, Utah 84112, U.S.A.

ABSTRACT

We derive the so-called first law of black hole mechanics for variations about stationary black hole solutions to the Einstein–Maxwell equations in the absence of sources. That is, we prove that $\delta M = \kappa\delta A + \omega\delta J + V dQ$ where the black hole parameters $M, \kappa, A, \omega, J, V$ and Q denote mass, surface gravity, horizon area, angular velocity of the horizon, angular momentum, electric potential of the horizon and charge respectively. The unvaried fields are those of a stationary, charged, rotating black hole and the variation is to an arbitrary 'nearby' black hole which is not necessarily stationary. Our approach is 4-dimensional in spirit and uses techniques involving Action variations and Noether operators. We show that the above formula holds on any asymptotically flat spatial 3-slice which extends from an arbitrary cross-section of the (future) horizon to spatial infinity. (Thus, the existence of a bifurcation surface is irrelevant to our demonstration. On the other hand, the derivation assumes without proof that the horizon possesses at least one of the following two (related) properties: (i) it cannot be destroyed by arbitrarily small perturbations of the metric and other fields which may be present, (ii) the expansion of the null geodesic generators of the perturbed horizon goes to zero in the distant future.)

[1]Published in *Class. Quant. Gravity* **13**: 1949-1970 (1996). E-print archive: gr-qc/9510031

I Introduction

Working directly with the Action principle which defines a field theory typically lends clarity and unity to the basic formal relationships of the theory. Thus, in the case of symmetries and conservation laws, the first variation of the Action S, leads directly to the conserved "charges" themselves [1], while its second variation leads to conclusions of which the following is typical: every time-independent solution of the field equations is an extremum (critical-point) of the total energy E [2].

Two directions of generalization of this last theorem suggest themselves as natural. On one hand, one can inquire [2] into the behavior of the *second* variation, or "Hessian" of E (corresponding to third variation of S itself); on the other hand, one can allow for the presence of a spacetime boundary, such as will occur naturally if one narrows one's attention to the region outside a black hole horizon. Clearly, both directions of generalization are relevant to the stability and (more generally) the thermodynamics of black holes; but the latter kind of generalization seems to be more readily accomplished, as well as being a necessary prelude to the former kind in the black hole case. This second direction is pursued in the present paper, and leads—via steps which we now sketch in advance—directly to the so-called first law of black hole mechanics (or thermodynamics).

A central concept in connection with local transformation groups is that of the Noether operator, an object which acts (in general as a differential operator) on an infinitesimal group-generator ξ to give a spacetime current $\mathcal{J}^a$ corresponding to the variation of S induced by ξ. (See equations (1) and (3) below.) When the group is the diffeomorphism group, ξ is a vectorfield ξ^a, and the Noether operator is correspondingly a tensor (density of weight 1), with the current being formed as $\mathcal{T}^a_b \cdot \xi^b$. Under appropriate conditions of asymptotic flatness, a Poincaré quotient-group of diffeomorphisms serves as a global symmetry group of the theory, and—for solutions of the field equations—the Noether operator yields a corresponding family of conserved charges when applied to the Poincaré generators and integrated over an asymptotically flat spacelike hypersurface or "slice".[2]

[2]From a physical point of view, it seems best to interpret such a global charge, say the linear momentum, as the change in S due to an infinitesimal variation in which the system under consideration is translated *relative to the environment whose presence is implicit in the use of asymptotically flat boundary conditions*. Certainly, genuinely asymptotically flat metrics do not occur in nature, and the translation or rotation of a subsystem cannot in practice be meaningfully extended beyond some large but finite radius. By taking the existence of the environment more explicitly into account, it ought to be possible to interpret the conserved charge as the change in the *total* S, with the various divergence terms one customarily adds to the Lagrangian serving only to allow a nominal splitting of S, and therefore of δS, into separate parts associated with the approximately isolated subsystem and with the environment. The same, "more realistic" approach is also important for correctly interpreting the divergent or conditionally convergent integrals which occur in connection with the angular momentum and boost conservation laws. (For the elements of such an approach see [1]).

The extremality theorem cited above then follows directly from taking an additional variation of the defining equation (1) for the Noether operator: the integral identity that results from the variation yields the theorem when evaluated on a spacetime region whose fields are chosen to interpolate between the unvaried and the varied configurations. In its maximal generality, this theorem asserts that any field-configuration which both solves the equations of motion and is invariant under a given infinitesimal symmetry, is an extremum of the corresponding conserved quantity; and that in fact any two of the properties, "solution", "symmetric", "extremum" imply the third.[3]

What happens then if we apply the same considerations, not to the entire spacetime manifold, but just to the region outside an event horizon?

But first, why is it necessary to exclude the interior region at all? It is not that the general theorem ceases to apply just because a horizon is present, but that, for subtle reasons, it does not furnish the kind of information one might expect in that case. With the Schwarzschild metric, for example, there are two possibilities for what is inside the horizon. There might be the familiar second asymptotic region joined to the first by a "throat", or there might be no such region if the internal topology is such that the "throat" leads one back out to the same asymptotic region from which one came (the latter case resulting from the former by suitable identifications.) In the former case the quantity 'E' which occurs in the theorem turns out to be the *difference* of the energies seen in the two asymptotic regions, rather than the effective mass of the black hole, while in the latter case, the meaning of E is physically appropriate, but the solution is no longer globally stationary technically, because the non-simple-connectivity of spacetime obstructs the extension of the timelike Killing vector to the interior region as a single-valued field. In neither case does one learn anything directly about variations of the physical black hole energy. However, if we truncate the manifold at the horizon then the remaining, external portion of the spacetime possesses both a single asymptotic region and a globally defined Killing vector, and we can expect that a suitable generalization of our theorem will furnish the kind of information we desire.

Consider then, an arbitrary variation of the metric in the neighborhood of an asymptotically flat slice, Σ which meets the horizon in an arbitrary two-dimensional cross-section. (The freedom to choose the cross-section freely is important, because the ability to push such a section forward in time appears to be crucial to an understanding of the *second* law of black hole thermodynamics; see the further remarks in the conclusion section below.) As long as the variation stays away from the horizon, the general theorem applies exactly as before, and we conclude that *the energy of*

[3]Similar relationships can be derived naturally via Hamilton's equations, but the Lagrangian approach has the advantage of leading more directly to objects with a spacetime meaning, such as conserved currents and fluxes (cf.[3] for example). And of course, a spacetime approach is particularly well suited to dealing with black holes, whose very definition involves the global causal structure.

a time-independent black hole is an extremum against variations of the fields which vanish in a neighborhood of the horizon. From this alone, it follows immediately that for arbitrary variations (including non-stationary ones), *the change in the total energy can depend only on the behavior of the variation on the horizon itself*, or more precisely, on its behavior near the 2-surface where the horizon meets the slice on which the energy is being evaluated.

We have come this far on the basis of reasoning valid for any field theory. To discover specifically which horizon-variables determine δE, we will need to use the concrete form of the Action S. (One might think that, in addition to horizon variables, the choice of spacelike hypersurface might enter into δE, but that cannot be true, because the energy of the varied configuration does not depend on the slice on which it is evaluated—assuming that the variation itself does not lead out[4] of the space of solutions.) For simplicity, consider the special case of a static, uncharged black hole in pure gravity, and let ξ^a be the asymptotically timelike Killing vector of the unvaried solution. Then, the integral for δE that evaluated to zero in the absence of a boundary, becomes with a boundary present, the net flux of a fictitious energy-current (that of the Einstein tensor of the interpolating metric) across the horizon. This flux can be evaluated entirely straightforwardly using the Raychaudhuri equation, and yields the familiar expression $\kappa\delta A$, where A is the horizon area and κ its surface gravity. We thus arrive at the first law for non-rotating uncharged black holes.

The more general case of a charged and/or rotating black hole is equally straightforward, and needs no special elaboration in this introduction, except to mention that, in dealing with angular momentum, we will need the analog for rotation of the "asymptotic patching" lemma employed in [2] for the case of translation. This is derived in an appendix (in a version strengthened significantly over that of [4], the improvement being made possible by our imposition of the so-called "parity conditions" on the asymptotic fields in conjunction with the use of an improved Noether operator over that used in [4]).

The only sightly tricky point in the derivation will be locating the horizon of the varied metric, a task which in general would demand knowledge of the entire future of the varied solution. In fact, what we will need for our derivation is only that the expansion θ be of second order in the perturbation when evaluated on the correctly identified varied horizon. Although "local" in itself, this assertion refers implicitly to the entire future, and we do not prove it herein. Instead, we reduce it to either of two assumptions asserting the stability (in a certain sense) of the black hole horizon. Such assumptions are common in discussions of black hole dynamics,

[4]The Noether operator allows us to define a generalized energy, even for field-configurations which do not satisfy the Einstein equation, and we will need to use this feature in our derivation of δE. Thus, it may be of some interest that hypersurface independence for this more general type of variation still follows from the identity on which the proof is based (i.e. δE can depend on the choice of slice only insofar as this influences the cross-section in which the slice meets the horizon).

and may perhaps be viewed as special cases of "cosmic censorship". Were they to fail, black hole spacetimes would have to be understood as very special types of mathematical solutions without relevance to reality.

The layout of the remainder of this paper is as follows. In section 2 we recall the definition of the Noether operator for gravity and its relation to conserved quantities [1, 2, 3]. The rest of the main body of the paper is devoted to the proof of the first law of black hole mechanics for variations from stationary 'gravitoelectric' spacetimes to nearby (not necessarily stationary) gravitoelectric spacetimes (by 'gravitoelectric', we intend gravity coupled to the source-free electromagnetic field). In section 3 we sketch the idea for the proof based on an 'extremum' identity and set the stage for the proof with some technical remarks. Section 4 gives the proof for "vacuum" solutions and section 5 generalizes it to the gravitoelectric case. In the concluding section, we discuss the significance of the result obtained and consider its relation to earlier work. In the appendices, the asymptotic falloffs of the metric and the electromagnetic potential are specified, and a lemma crucial to our proofs is proved.

Since all variations will be about stationary solutions with vanishing spatial momentum $\vec{p}$, we will have $\delta E = \delta M$ to first order, where M is the mass of the black hole (compare the relation $E^2 = M^2 + \vec{p}^2$). From now on we will denote first order variations in energy by δM, and often call the energy itself M.

For a "minimalist" account of what follows, see the first part of reference [5].

II The Noether operator for gravity

For the geometrical invariances of classical field theories, the 'Noetherian' relation between symmetries and conservation laws can be codified in terms of a differential operator $\mathcal{T}^a_b$ which acts on an arbitrary diffeomorphism-generating vector field ξ^a to produce a conserved current $\mathcal{T}^a_b \cdot \xi^b$. One can associate such a Noether operator to any first order Lagrangian density $\mathcal{L}[Q]$ via the identity [1, 2]

$$\delta S \;=\; \int_\Omega (\delta\mathcal{L}/\delta Q)\delta Q d^n x \;+\; \int_{\partial\Omega} f\mathcal{T}^a_b \cdot \xi^b d\sigma_a \tag{1}$$

In this equation, the Q are the dynamical field variables, $S = \int_\Omega \mathcal{L} d^n x$ is the action, $(\delta\mathcal{L}/\delta Q)$ is its variational derivative, and δS is the result of a variation in which (i) the variables Q 'change in place' according to

$$\delta Q \;=\; -f\mathcal{L}_\xi Q \tag{2}$$

($\mathcal{L}_\xi$ being the Lie-derivative), and (ii) the region of integration shifts its boundary by an amount $f\xi^a$. Notice that the arbitrary scalar field f in (1) and (2) helps to reduce the ambiguity of $\mathcal{T}^a_b$, which, for $f \equiv 1$, would remain a solution of (1) after supplementation with an arbitrary total divergence formed from ξ and the fields Q. It is implicit in (1) that $\mathcal{T}^a_b = \mathcal{T}^a_b[Q]$ depends locally on Q; and (1) is required to hold for arbitrary ξ, f and Ω *whether or not Q solves the equations of motion.*

The expression $\mathcal{T}^a_b \cdot \xi^b$ can be integrated to produce conserved quantities or 'charges' associated with the asymptotic symmetries of an isolated system in a manner discussed in [1], wherein the interpretation of these global charges as variations of the total action is also explained. [5] Without going into further detail here, let us remark that the interpretation is valid when ξ is an exact symmetry of the background fields, and the presence of the function f in (2) is what allows the variations in question to remain compatible with the asymptotic flatness of the metric g_{ab} at large radii. If ξ^a generates time translation then $\int d\sigma_a \mathcal{T}^a_b \cdot \xi^b$ will be the total energy; if it generates a rotation it will be an angular momentum component, etcetera.

The form of S_{grav} which leads to the Noether operator we use for gravity is the first order form obtained from the covariant action $(1/2)\int RdV$ via integration by parts. To define it in an intrinsic manner we can introduce a background connection $\overset{\circ}{\nabla}_a$ (say torsion free) and let Γ^a_{bc} represent the difference therefrom of the metric's connection ∇_a. (In a specific coordinate system, if $\overset{\circ}{\nabla}_a$ is chosen as the coordinate derivative operator, then Γ^a_{bc} will just be the Christoffel symbol as usually defined.) The first order Action is then $S_{grav} = 1/(2k)\int dV(\Gamma^a_{bc}\Gamma^b_{ad} - \Gamma^a_{ab}\Gamma^b_{cd} + \overset{\circ}{R}_{cd})g^{cd}$, where $\overset{\circ}{R}_{cd}$ is the Ricci tensor of $\overset{\circ}{\nabla}_a$ and $k = 8\pi G$ (in 4 spacetime dimensions). The Noether operator which answers to S_{grav} turns out to be that acting as follows [1, 6]:

$$T^a_b \cdot \xi^b \;=\; -G^a_b\xi^b + \nabla_b(\nabla^{[a}\xi^{b]} + \omega^{[a}\xi^{b]}) \tag{3}$$

where $T^a_b = (-g)^{-1/2}\mathcal{T}^a_b$, $\omega^a = \Gamma^a_{bc}g^{bc} - \Gamma^b_{bc}g^{ac} = (g)^{-1}\overset{\circ}{\nabla}_b(-gg^{ab})$, and $x^{[a}y^{b]} = 1/2[x^a y^b - x^b y^a]$. We also define

$$W^{ab}_c \cdot \xi^c = (\nabla^{[a}\xi^{b]} + \omega^{[a}\xi^{b]}) \tag{3a}$$

Note on Notation: In the various integrals which appear, we shall incorporate all the density weights into the volume elements which will accordingly be scalars (of density weight 0). We write the various volume elements thusly: dV denotes the element of

[5] For ξ corresponding asymptotically to a rotation, the integral defining δS converges only conditionally and only to first order in the variation δ. This obstacle to giving angular momentum a precise meaning derives from the fact that S itself converges only conditionally in general. It can be overcome by redefining S as the Action inside a sufficiently large sphere, and taking limits as the size of the sphere goes to infinity

4-volume, dS_a denotes the volume element on a 3-dimensional submanifold and dS_{ab} that on a 2-dimensional submanifold (which will always be topologically a 2-sphere). We also absorb into the volume elements dS_a, etc. the orientations which implicitly occur in Stokes' Theorem, although in this case the received notation does not allow us to indicate which orientation is being taken. In general relations like equation (9) we will adhere to the rule that the boundary of a region is to be oriented *outward*. In integrals over separate spacelike or null hypersurfaces like the integrals over Σ in (13) and (14), we will orient the surfaces upward (as if they were regarded as the boundaries of their pasts). When the surface is a portion H of a black hole horizon, this has the consequence that H is oriented "inward" (as if regarded as the boundary of the region of spacetime outside the black hole). Other orientations will not be made explicit, but can be deduced from the context.

With these conventions, the energy E of a gravitational field evaluated on a spatial asymptotically flat slice Σ without boundary is given by

$$-E = (1/k) \int_\Sigma T^a_b \cdot \xi^b dS_a$$

where ξ is an asymptotic time translation and T^a_b is an ordinary tensor operator, not an object of density weight 1. Note that when the vacuum Einstein equations are satisfied, we have from (3) and (3a),

$$E = -(1/2) \int_{\partial\Sigma} W^{ab}_c \cdot \xi^c dS_{ab} \tag{4}$$

where $\partial\Sigma$ is the "2-sphere at spatial infinity" and $k = 8\pi G$ is set equal to 1.

The assumed fall off conditions on the metric are given in the Appendix.

III Preparation for the proof

For a more detailed exposition of the ideas in this section see [2, 7].

III.1 The "extremum identity" for variations about a stationary solution, and its application to black hole spacetimes

We notice that for a variation induced by a pure diffeomorphism generated by ξ^a [which corresponds to putting $f = 1$ in equation(1)] we have for the gravitational action in (3)

$$\delta S_{grav} := \delta_\xi S_{grav} = \int_\Omega dV (1/2) G^{ab} \mathcal{L}_\xi g_{ab} + \int_{\partial\Omega} T^a_b \cdot \xi^b dS_a \tag{5}$$

But S_{grav} is invariant under diffeomorphisms which preserve the background connection $\overset{\circ}{\nabla}$. Let us assume that ξ preserves the background connection; then

$$\delta_\xi S = 0 \tag{6}$$

$$\Rightarrow \int_\Omega dV (1/2) G^{ab} \mathcal{L}_\xi g_{ab} + \int_{\partial\Omega} T^a_b \cdot \xi^b dS_a = 0 \tag{7}$$

Note that this identity holds for arbitrary field configurations, which do not need to be solutions to the field equations. We now take the variation of (7) under an arbitrary variation δg_{ab} of the metric. We get

$$\int_\Omega (1/2)\delta(dV G^{ab})\mathcal{L}_\xi g_{ab} + \int_\Omega dV(1/2)G^{ab}\delta(\mathcal{L}_\xi g_{ab}) + \delta\int_{\partial\Omega} T^a_b \cdot \xi^b dS_a \;=\; 0 \tag{8}$$

Let us further assume that we are varying about a metric which is both a solution to the field equations ($G_{ab} = 0$) and stationary with respect to ξ^a ($\mathcal{L}_\xi g_{ab} = 0$). Then the above equation reduces to

$$\delta \int_{\partial\Omega} T^a_b \cdot \xi^b dS_a \;=\; 0 \tag{9}$$

which we shall call the *extremum identity.*

For the proof, we will take g_{ab} to be a stationary black hole solution to the vacuum field equations with killing vector field ξ^a, and we will choose Ω as the region bounded between the horizon '∂B' of the stationary black hole and two asymptotically flat spatial slices, Σ_i and Σ_f (Σ_f being to the future of Σ_i), each of which intersect the horizon at one "end" and go out to spatial infinity at the other "end". Also, we will choose ξ^a to be that Killing field of the stationary solution which is null on the horizon. (The general Kerr-Newman metric always admits such a Killing vector; the horizon "rotates rigidly".) Then

$$\xi^a \;=\; t^a \;+\; \omega\phi^a \tag{10}$$

where t^a and ϕ^a are the Killing vectors of the stationary solution which are asymptotically generators of time translation and rotation respectively; and ω is the angular velocity of the stationary horizon with respect to infinity. We aim to reproduce the first law with E, A, etc. evaluated on the slice Σ_f.

To that end, a suitable choice of "gauge" for δg will be very helpful. The gauge-freedom arises in the following, well-known manner. If g is a metric on a manifold

N and $\overline{g}$ is a metric on $\overline{N}$, then the pairs (N, g) and $(\overline{N}, \overline{g})$ may be said to be "infinitesimally close" iff there exists a diffeomorphism $\Phi : N \to \overline{N}$, such that $\Phi^* \overline{g} = g + \delta g$ where δg is infinitesimal.[6] Such a diffeomorphism Φ (call it 'allowable') is not unique: if Π is allowable and Ψ is any diffeomorphism infinitesimally close to the identity, $\Pi \circ \Psi$ is also allowable. We will use this freedom in the specification of Φ in (b), (c) and (d) below.

We now divide the 4-manifold outside ∂B into 3 parts:

- Σ_i and its past
- Σ_f and its future
- The region Ω between Σ_i and Σ_f

and we contemplate a preliminary choice of δg_{ab} as follows.

(a) On Σ_i and to its past, $\delta g_{ab} = 0$

(b) On Σ_f and to its future $(g_{ab} + \delta g_{ab})$ is a black hole solution to the linearized field equations with mass $M + \delta M$ and angular momentum $J + \delta J$. We assume, without loss of generality, that the horizons of the varied solution and of the stationary unvaried solution coincide in this region.

(c) In Ω: we require that the part of ∂B between Σ_i and Σ_f be a null surface for the metric $(g_{ab} + \delta g_{ab})$. We shall henceforth refer to this portion of the horizon of $g + \delta g$ as H.

(d) On ∂B it is possible to "line up" the null geodesic generators k^a, for g_{ab} with those for $g_{ab} + \delta g_{ab}$, so that δk^a is proportional to k^a. This alone does not imply that the affine parameters of the unvaried and varied null geodesic fields agree on H. In general they will differ by infinitesimal amounts. But it is always possible to use the infinitesimal diffeomorphism freedom whose existence we emphasized above, to identify slices in the unvaried spacetime with slices in the varied spacetime, in such a way that the affine parameters actually do agree. Thus, on H, we can always arrange for $\delta k^a = 0$, without in any way constraining the location of Σ_f with respect to the varied metric.

Based on our choice of δg, we shall end up proving the following statement (for the vacuum case): Given a stationary vacuum black hole spacetime N with metric g and an infinitesimally nearby pair $(\overline{N}, \overline{g})$, there exists an allowable diffeomorphism Φ_0, which maps $\overline{g}$ to $g + \delta g$ and $\overline{\Sigma}$ (a slice in the exterior of the horizon of the varied spacetime) to an "infinitesimally nearby" slice Σ such that the first law holds on Σ with regard to the infinitesimal variation δg. But, since *all* slices in the stationary spacetime have the same horizon area, we will have actually proved the first law in the context of an arbitrary allowable diffeomorphism, not just Φ_0. Similar considerations apply to the gravito-electric case.

[6] The notion of infinitesimal can be made precise in the usual way: replace Φ and $\overline{g}$ by smooth parameterized families $\Phi(\lambda)$ and $\overline{g}(\lambda)$ with $\Phi(0)^* \overline{g}(0) = g$, and define $\delta g := \partial[\Phi(\lambda)^* \overline{g}(\lambda)]/\partial\lambda|_{\lambda=0} d\lambda$. Similarly, we interpret (b) below (for example) to mean that the horizons coincide for the members of some 1-parameter family of solutions whose derivative is δg_{ab}

We choose our background connection $\overset{\circ}{\nabla}$, so that it is compatible near spatial infinity with the flat metric η_{ab} which we will use to define our fall off conditions. We also choose it so that it is globally Lie derived by ξ. (For example we could use $\overset{\circ}{\nabla} = (1-\lambda)\overset{\circ}{\nabla}_{flat} + \lambda\nabla_{stationary}$ with $\lambda = 1$ on H, $\lambda = 0$ at spatial infinity and $\mathcal{L}_\xi \lambda = 0$.)

Note that the requirement that $g + \delta g$ on and to the future of Σ_f be a solution with mass $M + \delta M$ and angular momentum $J + \delta J$ different in general from the mass M and angular momentum J of the stationary solution implies that *in the region* Ω, $g + \delta g$ *cannot be a solution to the field equations.*

III.2 An important technical point: asymptotic patching

Although we would like to make a variation satisfying the conditions (a), (b) (c) and (d) of section III.1 above, we notice that such a variation would violate the asymptotic fall off conditions on the metric (for these conditions see the appendix). We can see this in the following way: The mass-information for a solution to the field equations is in the '$1/r$ part' of the metric. Now, the mass of the initial and final field configurations on the slices Σ_i and Σ_f are M and $M + \delta M$ respectively. So in between these slices near spatial infinity the $1/r$ part has to be time dependent. But this means that $\partial_t(g+\delta g)$ will fall off as $1/r$ and not $1/r^2$ as required. Similar considerations apply to 'angular momentum information'. Hence, the fall off conditions in the appendix would have to be violated. If this happened, the variations of the action S_{grav} would be ill defined, meaning that we could not use the variation outlined in section III.1. This is the technical problem. To avoid it, we employ the following trick, along the lines of the trick performed in [2]

First recall that our aim is to produce from the extremum identity the explicit horizon term to which we know that δM must reduce. Consider, then, an asymptotically flat spatial slice $\Sigma \subset \Omega$ extending from H to spatial infinity. For an arbitrary black hole *solution* to the field equations, the energy M evaluated on the spatial slice Σ is given by equation (4) as $M = -(1/2)\int_{S^2_\infty} W_c^{ab} \cdot t^c dS_{ab}$ (where S^2_∞ is the 2-sphere at spatial infinity and t^a is the asymptotic time translation vector field).

We can extend the definition of the energy to *non-solutions* by setting

$$\begin{aligned} -M &= \int_{S^2_\infty} \frac{1}{2} W_c^{ab} \cdot t^c dS_{ab} - \int_\Sigma G^a_b t^b dS_a \\ &= \Big(\int_{S^2_\infty} - \int_{S^2_{\Sigma\cap H}}\Big)\frac{1}{2} W_c^{ab} \cdot t^c dS_{ab} + \int_{S^2_{\Sigma\cap H}} \frac{1}{2} W_c^{ab} \cdot t^c dS_{ab} - \int_\Sigma G^a_b t^b dS_a \qquad (11) \end{aligned}$$

where $S^2_{\Sigma\cap H}$ denotes the 2-sphere in which Σ meets H. Of course this agrees with the energy *on* a solution because the Einstein tensor vanishes on a solution. Similarly we can extend the definition of the angular momentum to non-solutions by setting

$$J \;=\; \int_{S^2_\infty} \frac{1}{2} W^{ab}_c \cdot \phi^c dS_{ab} \;-\; \int_\Sigma G^a_b \phi^b dS_a \tag{12}$$

(where ϕ^a is the asymptotic rotational vector field). In view of Stokes theorem and eq. (3), our definitions of energy and angular momentum extended to non-solutions are, respectively:

$$-M \;=\; \int_\Sigma T^a_b \cdot t^b dS_a \;+\; \int_{S^2_{\Sigma\cap H}} \frac{1}{2} W^{ab}_c \cdot t^c dS_{ab} \tag{13}$$

and

$$J \;=\; \int_\Sigma T^a_b \cdot \phi^b dS_a \;+\; \int_{S^2_{\Sigma\cap H}} \frac{1}{2} W^{ab}_c \cdot \phi^c dS_{ab} \tag{14}$$

Relative to a given choice of background connection $\overset{\circ}{\nabla}$, equation (13) can be interpreted to say that the total energy, M, of the spacetime is the sum of the energy outside the horizon (the first term on the right hand side of (13)) and the energy of the horizon itself (the second term on the right hand side of (13)). The terms in equation (14) can be interpreted in a similar way. Equations (13) and (14) will provide the second key ingredient for our proof.

From the lemma in the appendix and lemma 3.1 of [2], we know that one can construct a field configuration $g_{ab} + \Delta g_{ab}$ in a neighborhood of Σ_f such that:
1. its mass as given by equation (13) is still $M + \delta M$ and its angular momentum as given by equation (14) is still $J + \delta J$, but asymptotically it agrees exactly with g_{ab}.
2. near the horizon on Σ_f , $g_{ab} + \Delta g_{ab}$ agrees exactly with $g_{ab} + \delta g_{ab}$.
(Notice that $g + \Delta g$ then has to violate the field equations in some region far from the horizon.)

Using these facts we can construct the new variation Δg in Ω under which we choose to evaluate equation (9). Δg is such that:
[i] $\Delta g = 0$ at Σ_i
[ii] Δg agrees with 1. and 2. above; i.e. $g + \Delta g$ has mass $M + \delta M$ and angular momentum $J + \delta J$ on Σ_f, and Δg near Σ_f vanishes near spatial infinity.
[iii] Δg within Ω vanishes in a neighborhood of spatial infinity. In other words Δg is of compact support within Ω. Also, H remains a null surface for the metric $g_{ab}+\Delta g_{ab}$, and $\Delta g_{ab} = \delta g_{ab}$ near H.

From [iii] just above and (d) of section III.1, we have that on H

$$\Delta k^a = 0 \tag{15}$$

From this property of the variation in conjunction with its other properties one observes the following two useful facts: Since H is still a null surface its null normals must be proportional to $g_{ab}k^b = k_a$. So

$$\Delta k_a = ck_a \tag{16}$$

for some real c. And since ξ is null on H,

$$\xi^a = \alpha k^a \tag{17}$$

for some coefficient of proportionality α.
To summarize: The trick is to replace δg by Δg. This allows the variation to be of compact support, freeing us from having to evaluate variations near spatial infinity.

As a final remark, we note that the variations δ and Δ are identical near H and we shall continue to denote by δ, the variations of quantities near H.

IV The proof for gravity alone

We would like to emphasize that what we have described in the preceding sections is just the necessary background to do the actual proof of the first law. Once this background is assumed and absorbed, the proof itself is extremely simple (cf. [5]).

The extremum identity, equation (9), written for the Δ variation is
$\Delta \int_{\partial\Omega} T^a_b \cdot \xi^b dS_a = 0$. But $\Delta g = 0$ near spatial infinity and on Σ_i, whence

$$\Delta \int_{\Sigma_f} T^a_b \cdot \xi^b dS_a \; + \; \delta \int_H T^a_b \cdot \xi^b dS_a \; = \; 0 \tag{18}$$

Taking the variation of equations (13) and (14) with $\Sigma = \Sigma_f$ we get

$$-\delta M \; = \; \Delta \int_{\Sigma_f} T^a_b \cdot t^b dS_a \; + \; \delta \int_{S^2_{\Sigma_f \cap H}} \frac{1}{2} W^{ab}_c \cdot t^c dS_{ab} \tag{19}$$

and

$$\delta J \; = \; \Delta \int_{\Sigma_f} T^a_b \cdot \phi^b dS_a \; + \; \delta \int_{S^2_{\Sigma_f \cap H}} \frac{1}{2} W^{ab}_c \cdot \phi^c dS_{ab} \tag{20}$$

Equations (10), (18), (19) and (20) yield then

$$-\delta M \; + \; \omega \delta J \; = \; \delta \int_{S^2_{\Sigma_f \cap H}} \frac{1}{2} W^{ab}_c \cdot \xi^c dS_{ab} \; - \; \delta \int_H T^a_b \cdot \xi^b dS_a \tag{21}$$

Using the definition of the Noether operator for gravity and Stokes theorem (and recalling that $\delta g = 0$ on $\Sigma_i \cap H$) we get

$$-\delta M \;+\; \omega \delta J \;=\; \delta \int_H G^a_b \xi^b dS_a \tag{22}$$

a relation of some interest in its own right. Notice that (22) can be read as equating the change in energy or angular momentum to what would be the corresponding "energy flux" or "angular momentum flux" through the horizon if δG_{ab} were interpreted as the stress energy of some fictitious matter field.

We also note that this relation shows in another way that the left hand side of (22) must reduce to a surface integral on $\Sigma_f \cap H$, because on the right hand side we can confine δ to a neighborhood of $\Sigma_f \cap H$ on H, as in [1], and in the limit can put $\delta g_{ab} \propto \theta(\lambda - \lambda_f)$, where θ is the step function. Yet another way to show this, perhaps the most systematic of all, would be to use the theorem in [7] to derive a "potential" for $\delta(G^a_b \xi^b)$ from the identity $\partial_a \delta(\sqrt{-g} G^a_b \xi^b) \equiv 0$, see the discussion in [5].

In evaluating the right hand side of (22), let λ be an affine parameter for the null geodesic generators of H so that $k^a = dx^a/d\lambda$. Then

$$dS_a = -k_a d\lambda d^2 A \tag{23}$$

where d^2A is the area element of the 2-sphere cross sections of H (and λ increases toward the future). Moreover, by differentiating equation (17), it follows immediately (in view of the fact that k^a is the tangent to an affinely parameterized null geodesic, whence $k^a \nabla_a k^b = 0$) that

$$d\alpha/d\lambda = \kappa \tag{24}$$

where κ is the surface gravity of the stationary black hole, defined by

$$\xi^a \nabla_a \xi^b = \kappa \xi^b. \tag{25}$$

In the extremum identity, ξ is not to be varied, i.e. $\Delta\xi = 0$. Then since $\delta k^a = 0$ as well, the right hand side of (22) becomes

$$-\int_H \alpha k^a k^b (\delta(G_{ab} d^2 A)) d\lambda \tag{26}$$

Using $G_{ab} = 0$ for the stationary solution and the fact that k^a remains null under the variation, we get

$$-\oint d^2 A \int d\lambda \; \alpha k^a k^b (\delta R_{ab}) \tag{27}$$

To evaluate $(\delta R_{ab})k^a k^b$ we use the Raychaudhuri equation for the rate of expansion of a congruence of null geodesics (see for example [8]):

$$d\theta/d\lambda = -(1/2)\theta^2 - \sigma_{ab}\sigma^{ab} + \omega_{ab}\omega^{ab} - R_{ab}k^a k^b \tag{28}$$

where θ is the expansion, σ the shear and ω the twist of the null geodesic congruence with k^a parameterized by affine parameter λ. We apply this equation to the null geodesic congruence on H (for which anyway $\omega = 0$) and take its δ variation. Using the fact that the expansion and shear of the stationary solution vanish we get:

$$d(\delta\theta)/d\lambda = -(\delta R_{ab})k^a k^b \tag{29}$$

Inserting this into (27) yields

$$\delta \int_H G^a_b \xi^b dS_a \;=\; \oint d^2A \int d\lambda \alpha d(\delta\theta)/d\lambda \tag{30}$$

$$=\; (\oint d^2A \alpha \delta\theta)|_i^f - \oint d^2A \int d\lambda \frac{d\alpha}{d\lambda}\delta\theta \tag{31}$$

where i, f refer to the 2-spheres which are the intersections of H with Σ_i and Σ_f respectively.

But $\delta\theta_i = 0$ by construction, and $\delta\theta_f = 0$ under the assumptions of Lemma 1 (which is given at the end of this section). Also, $\theta = (1/d^2A)\; d(d^2A)/d\lambda$ by definition. Hence,

$$\delta \int_H G^a_b \xi^b dS_a = -\oint d^2A \int d\lambda \frac{d\alpha}{d\lambda}\, \delta[\frac{1}{d^2A}\; \frac{d(d^2A)}{d\lambda}] \tag{32}$$

Using (24) together with the fact that for the stationary solution $d(d^2A)/d\lambda = 0$, we get for (32)

$$-\oint d^2A \int d\lambda\, \kappa\, \frac{1}{d^2A}\, \delta \frac{d(d^2A)}{d\lambda} \tag{33}$$

But $\delta\lambda = 0$, so

$$-\delta \int_H G^a_b \xi^b dS_a \;=\; \kappa \oint \int d\lambda \frac{d(\delta d^2A)}{d\lambda} \tag{34}$$

$$=\; \kappa\, \delta(\oint d^2A)|_i^f \tag{35}$$

$$=\; \kappa\, \delta(\oint d^2A)|_f \tag{36}$$

where in the last line we have used that the variation δ vanishes on Σ_i. Hence we get by (22)

$$\delta M \;=\; \kappa\delta A \;+\; \omega\delta J \tag{37}$$

where everything is evaluated on Σ_f.

To complete our proof, we have to justify setting $\delta\theta_f = 0$, which we do via the following lemma.

Lemma 1: Let $(g_{ab}, \{\phi\})$ be a stationary black hole solution to the field equations, where $\{\phi\}$ denotes all fields other than the gravitational field, and let the stress-energy tensor T^{ab} of $\{\phi\}$ satisfy the positivity condition that $T^{ab}k_a k_b \geq 0$ for every null k^a.

Consider the spacetime region Ω' to the future of some spatial slice Σ_f. Denote the part of the horizon in this region by H'. Let $(g_{ab}+\delta g_{ab}, \{\phi+\delta\phi\})$ be an infinitesimally differing solution, with the relevant part of its horizon coinciding with H'.

Then the variation $\delta\theta$ of the expansion evaluated on Σ_f vanishes if either of the following (rather weak) stability conditions holds:

(i) It is not the case that arbitrarily small perturbations of the metric and the other fields can destroy the horizon; or

(ii) $\delta\theta$ approaches zero in the distant future along the horizon.

Proof:

(i) $(g_{ab} + \delta g_{ab}, \{\phi + \delta\phi\})$ is a solution to the linearized field equations. Therefore $(g_{ab} - \delta g_{ab}, \{\phi - \delta\phi\})$ is also a solution to the linearized equations. Thus, if for the former solution, the variation of the expansion on H' is $\delta\theta$ then for the latter solution, the variation in the expansion will be $-\delta\theta$.

Since $\theta = 0$ for the stationary solution, there exists a solution to the full Einstein equations in a neighborhood of the stationary solution whose expansion at first order is $\delta\theta$, which, as just explained, can be arranged to be negative unless it is always zero. But it is known that a negative expansion implies a singular evolution of the horizon. Namely, by using ("to all orders") the Raychaudhuri equation (28), the Einstein equation and the positive energy condition above, one can conclude that if the expansion $\theta = \theta_0$ is negative at some $\lambda = \lambda_0$ (the geodesics are converging), then a conjugate point (caustic) will develop at or before some definite future instant λ_1, where λ_1 depends on θ_0. But the occurrence of a conjugate point contradicts the definition of the horizon, because no geodesic generator can remain on the horizon after reaching a conjugate point. Thus, the only possibility is that the generator in question terminates in a singularity, i.e. that the perturbation destroys the horizon, contradicting assumption (i).

(ii) Applying the variation δ to the Raychaudhuri equation and noticing as before that the expansion, shear and twist of the stationary solution vanish, we get from (29)

$$d(\delta\theta)/d\lambda = -(\delta R_{ab})k^a k^b = -\delta T_{ab}k^a k^b, \tag{38}$$

in view of the Einstein equation. But from (28) and the Einstein equations it follows that for the stationary solution

$$R_{ab}k^a k^b = T_{ab}k^a k^b = 0$$

Hence the positivity condition on the stress energy tensor implies that

$$\delta T_{ab}k^a k^b \geq 0$$

for arbitrary variations of the fields. But as δT_{ab} is linear in the infinitesimal variations $(\delta g_{ab}, \{\delta\phi\})$, this is impossible unless $\delta T_{ab}k^a k^b = 0$ for all variations, whence

$$d(\delta\theta)/d\lambda = 0 \tag{39}$$

for all variations. Hence $\delta\theta_f = \delta\theta_\infty$ where $\delta\theta_\infty$ is $\delta\theta$ evaluated at large affine parameter and $\delta\theta_f$ is $\delta\theta$ evaluated at Σ_f. But $\delta\theta_\infty = 0$ by assumption. Hence $\delta\theta_f$ vanishes as well. (Notice that if Σ_f intersected the horizon in a bifurcation surface as in [9] then, since by definition $\alpha = 0$ at such a surface, $\alpha\delta\theta_f = 0$ would be immediate and lemma 1 would not be needed.)

V The proof for gravito-electric spacetimes

V.1 The Electromagnetic Noether operator

In order to generalize (37) to gravito-electric spacetimes, we will apply the extremum identity (9) to field variations about a stationary, charged black hole solution of the Einstein-Maxwell equations. From the general argument in section 1, we know that δM and δJ will be expressible entirely in terms of "horizon variables", and since the total Action is now the sum of S_{grav} and a single electromagnetic term, we know that a single extra term will appear in (37). We can find this term by using the gravito-electric versions of equations (9), (13) and (14). To do this, we need expressions for the relevant Noether operators. We display them below (for details see [3]).

The Action for the electromagnetic field is $S_{em} = (-1/4)\int F_{ab}F^{ab}dV$ where $F_{ab} = \nabla_a A_b - \nabla_b A_a$ is the electromagnetic field tensor and A_b its potential. The Action for the combined gravitational and electromagnetic fields is $S_{grav} + S_{em}$ (S_{grav} has been defined in section II). The action of the gravito-electric Noether operator on a vector field ξ^b is:

$$\mathcal{T}^a_{total\ b} \cdot \xi^b := \sqrt{-g}\ T^a_{total\ b} \cdot \xi^b = \sqrt{-g}\ (T^a_{em\ b} + T^a_{grav\ b}) \cdot \xi^b \tag{40}$$

where $T^a_{grav\ b} \cdot \xi^b$ is given in equation (3) and $T_{em}{}^a_b$ is given [3] by

$$\begin{aligned} T^a_{em\ b} \cdot \xi^b &= F^{ab}\mathcal{L}_\xi A_b - (1/4)F_{cd}F^{cd}\xi^a & (41)\\ &= T^a_{S\ b}\xi^b - \nabla_b(\xi^c A_c)F^{ab}, & (42) \end{aligned}$$

T_S^{ab} being the stress energy tensor of the electromagnetic field,

$$T_S^{ab} = F^{ac}F^b{}_c - (1/4)g^{ab}F_{cd}F^{cd} \tag{43}$$

The electromagnetic analogue of (3) is therefore

$$T^a_{em\ b} \cdot \xi^b = (T^a_{S\ b} + (\nabla_c F^{ac})A_b)\xi^b - \nabla_b(F^{ab}A_c\xi^c) \tag{44}$$

and the total Noether current is—as in (3)—the sum of a divergence with a term which vanishes "on shell":

$$\begin{aligned} T^a_{total\ b} \cdot \xi^b \ &= \ [T^a_{S\ b} - G^a_b + (\nabla_c F^{ac})A_b]\xi^b \\ &\quad + \nabla_b(W_c^{ab} \cdot \xi^c - F^{ab}A_c\xi^c) \end{aligned} \tag{45}$$

with $W_c^{ab} \cdot \xi^c$ given by (3).

V.2 The extremum identity

The extremum identity reads now

$$\delta \int_{\partial\Omega} T^a_{total\ b} \cdot \xi^b dS_a \ = \ 0 \tag{46}$$

As before, we will apply this with $\xi^a = t^a + \omega\phi^a$ where t^a and ϕ^a are Killing fields which are asymptotically time translational and rotational respectively; ξ^a is that killing field which is null on the horizon and ω is the angular velocity of the black hole.

V.3 Some useful properties of the stationary solution

The variations will be made around a stationary black hole configuration, g_{ab}, A_c. The following properties of this stationary black hole solution will be of use:

(a) $\mathcal{L}_\xi A_a = 0$. (This amounts to a suitable gauge choice and was implicitly assumed when we wrote down the extremum identity above)

(b) Claim: On the horizon $F^{ab}\xi_a = \gamma\xi^b$ for some function γ .

Proof: Using (28) for the stationary solution on the horizon, the Einstein equations and the proportionality of the null geodesic horizon-generators to the Killing field ξ, one gets

$$T_{Sab}\xi^a\xi^b = 0$$

on the horizon. Substituting the explicit form of T_{Sab} from (43) into this equation, one concludes that $F^{ac}\xi_c$ is a null vector. From the antisymmetry of F^{ac}, one concludes that this null vector has to be proportional to ξ^a, since it is both null and orthogonal to ξ.

(c) The 'electric potential' $-\xi^a A_a$ is constant on the horizon. (Contract the identity $\xi^a F_{ab} = \mathcal{L}_\xi A_b - \nabla_b(\xi^a A_a)$ with any v^b tangent to the horizon and use (a) and (b).) Following the common notation, we will write $V = -\xi^a A_a$ for this constant.

These relationships express the fact that the horizon behaves like a conducting surface: in equilibrium (condition a) the electric potential must be constant (condition c) and hence the "electric field" must be perpendicular to the surface (condition b) (see, for example [10]).

V.4 The variation 'δ'

We use the same notation as in previous sections for spacetime regions, volume elements and anything involving only the metric.

The conditions on 'δ variations' are similar to those in section III.1. We choose $(\delta A_c, \delta g_{ab})$ as follows:
(a) On Σ_i and to its past, $\delta g_{ab} = \delta A_c = 0$ (we don't want to involve Σ_i at all)
(b) On Σ_f and to its future $(g_{ab} + \delta g_{ab},\ A_c + \delta A_c)$ is a black hole solution to the field equations with energy $M + \delta M$ angular momentum $J + \delta J$ and charge $Q + \delta Q$. We arrange that the horizons of the varied and unvaried solutions coincide.
(c), (d) These conditions are identical to conditions (c) and (d) of Section III.1.

We also choose our background connection, $\overset{\circ}{\nabla}$, as in III.1. As before, we have to deal with non-solutions in the region between Σ_i and Σ_f.

V.5 The variation 'Δ'

We face a similar technical problem as in Section III.2. To get around it, we repeat the steps followed there. We first define expressions for the energy M and angular momentum J of a field configuration on an asymptotically flat spatial slice Σ which reduce to the correct values when the source free Einstein-Maxwell equations are satisfied, but which are also useful for fields which are not solutions (these expressions follow directly from (45) and the fact that the total energy and angular momentum on a boundary-less slice can always be written as $\int_\Sigma T^b_{total\ a} \cdot \xi^a dS_b$ with ξ the relevant generator), to wit:

$$-M = \int_\Sigma T^b_{total\ a} \cdot t^a dS_b + \oint_{S^2_{\Sigma\cap H}} \frac{1}{2} F^{ab} A_c t^c dS_{ab} + \oint_{S^2_{\Sigma\cap H}} \frac{1}{2} W_c^{ab} \cdot t^c dS_{ab} \tag{47}$$

$$J = \int_\Sigma T^b_{total\ a} \cdot \phi^a dS_b + \oint_{S^2_{\Sigma\cap H}} \frac{1}{2} F^{ab} A_c \phi^c dS_{ab} + \oint_{S^2_{\Sigma\cap H}} \frac{1}{2} W_c^{ab} \cdot \phi^c dS_{ab} \tag{48}$$

Using arguments similar to those in section III.2 one can introduce variations of compact support of all the fields (i.e. A and g) such that the new field configuration on Σ_f has mass $M+\delta M$ and angular momentum $J+\delta J$, and is identical with $(A+\delta A, g+\delta g)$ near the horizon. We will not display the relevant patching arguments since they are similar to those of Appendix B. We simply note that in order to complete them, one has to use the asymptotic conditions on A_c as well. These conditions are listed in Appendix C.

Let us, as before, call this new variation of compact support the Δ variation. Just as earlier, it satisfies the analogs of properties [i]–[iii] of Section III.2, as well as equations (15)–(17). And, we can again continue to denote Δ on the horizon by δ, as there is no difference between the two variations there.

V.6 The proof

We have the extremum identity:

$$\Delta \int_{\partial\Omega} T^{a}_{total\ b} \cdot \xi^b dS_a = 0 \tag{49}$$

$$\Rightarrow \Delta \int_{\Sigma_f} T^{a}_{total\ b} \cdot \xi^b dS_a + \delta \int_{H} T^{a}_{total\ b} \cdot \xi^b dS_a = 0 \tag{50}$$

Using the above expressions for M and J with $\Sigma = \Sigma_f$, and taking respectively $\xi = t$ and $\xi = \phi$ in (50) yields, in view of (40),

$$\begin{aligned} -\,\delta M = {} & \delta \int_{S^2_{\Sigma_f \cap H}} \frac{1}{2} W^{ab}_{c} \cdot t^c dS_{ab} \;-\; \delta \int_H T^{a}_{grav\ b} \cdot t^b dS_a \;-\; \delta \int_H T^{a}_{em\ b} \cdot t^b dS_a \\ & + \delta \int_{S^2_{\Sigma_f \cap H}} \frac{1}{2} F^{ab} A_c t^c dS_{ab} \end{aligned} \tag{51}$$

$$\begin{aligned} \delta J = {} & \delta \int_{S^2_{\Sigma_f \cap H}} \frac{1}{2} W^{ab}_{c} \cdot \phi^c dS_{ab} \;-\; \delta \int_H T^{a}_{grav\ b} \cdot \phi^b dS_a \;-\; \delta \int_H T^{a}_{em\ b} \cdot \phi^b dS_a \\ & + \delta \int_{S^2_{\Sigma_f \cap H}} \frac{1}{2} F^{ab} A_c \phi^c dS_{ab} \end{aligned} \tag{52}$$

Then the same manipulations of the gravitational terms in these equations as we performed in section IV produce here

$$-\,\delta M + \omega \delta J = (\mathrm{grav}) \;+\; (\mathrm{em}) \tag{53}$$

where

$$(\text{grav}) = \delta \int_H G^a_b \xi^b dS_a = -\kappa \delta A + \oint_S d^2 A \alpha \delta \theta \tag{54}$$

is the same expression as before (with ξ^a again denoting $t^a + \omega \phi^a$), and

$$(\text{em}) = \delta(-\int_H T^a_{em\ b} \cdot \xi^b dS_a + \frac{1}{2} \oint_S F^{ab} A_c \xi^c dS_{ab}) \tag{55}$$

Here A denotes the area of the 2-sphere $S := \Sigma_f \cap H$ and $\oint_S$ denotes integration over this 2-sphere. Notice that in (55) it is *not* convenient to convert the integral over $\Sigma_f \cap H$ to an integral over H as we did with the corresponding integral for gravity in equation (21).

As expected, our expression for $-\delta M + \omega \delta J$ is just the same as before, with the exception of the additional term '(em)' in equation (53). To convert this term to a surface integral over S (as we already know must be possible on general grounds), we proceed as follows. Direct substitution of eq. (42) into (55) yields

$$(\text{em}) = I_1 + I_2 + I_3 \tag{56}$$

where

$$\begin{aligned} I_1 &:= -\delta \textstyle\int_H F^{ab} \mathcal{L}_\xi A_b dS_a \\ I_2 &:= \delta \textstyle\int_H \frac{1}{4} F^{cd} F_{cd} \xi^a dS_a \\ I_3 &:= \delta \textstyle\oint_S A_c \xi^c \frac{1}{2} F^{ab} dS_{ab}. \end{aligned}$$

Now I_3 is already a surface integral over S, and I_2 is easily seen to vanish in light of (23), (16) and the fact that $\xi \cdot \xi$ and $\delta \xi^a$ both vanish. For I_1, we can transform its integrand using (V3.a) to get

$$F^{ab} \mathcal{L}_\xi \delta A_b = \mathcal{L}_\xi (F^{ab} \delta A_b) = \nabla_c Y^{ac} - \nabla_c (F^{ab} \delta A_b) \xi^a$$

where $Y^{ac} \equiv F^{ab} \delta A_b \xi^c - (a \leftrightarrow c)$. The final term does not contribute to the integral I_1 because $\xi^a dS_a = 0$ by (17) and (23), and the first term yields an integral which can be converted by Stokes' theorem to obtain

$$I_1 = -\oint_S F^{ab} \delta A_b \xi^c dS_{ac}$$

(remember that $\delta = 0$ on Σ_i). Returning to I_3 for a moment, we can write

$$I_3 = A_c \xi^c \delta \int \frac{1}{2} F^{ab} dS_{ab} + \int \frac{1}{2} dS_{ab} F^{ab} \delta (A_c \xi^c) \tag{57}$$

where we have used (V.3c) to take $-V = A\cdot\xi$ out from under the integral sign. It is then a matter of a few lines of straightforward algebra to show that the second integral in (57) cancels with I_1. (The algebra uses (23), (V.3b), and the fact that the surface element dS_{ab} of S can be written as

$$dS_{ab} = d^2A\,(k_a l_b - k_b l_a) \tag{58}$$

where l^a is any future-null vector orthogonal to S such that $k_a l^a = -1$.) Thus we are left with

$$(\text{em}) = I_1 + I_2 + I_3 = A_c \xi^c \delta \oint_S \frac{1}{2} F^{ab} dS_{ab}$$

or

$$(\text{em}) = -\ V\delta Q \tag{59}$$

because of the perfectly general relation

$$Q = \oint_S \frac{1}{2} F^{ab} dS_{ab} \tag{60}$$

Q being the charge of the black hole, evaluated as an (outward) flux through $S = H \cap \Sigma_f$.

Substituting this back into equation (53), we get

$$-\delta M + \omega\delta J = -\kappa\delta A + \oint_S d^2A\alpha\delta\theta\ -\ V\delta Q \tag{61}$$

Finally, Lemma 1 comes to our aid to tell us once again that $\delta\theta = 0$ on S, so, putting this back in equation (61), we get our generalized extremality theorem for the gravito-electric case:

$$\delta M\ =\ \kappa\delta A + \omega\delta J\ +\ V\delta Q \tag{62}$$

VI Conclusions

As we intimated in the Introduction, the primary impetus for writing this paper came from our curiosity about how the extremality theorem of reference [2] would be modified by adaptation to the spacetime region exterior to a black hole. Here we would like to comment more at length on the wider significance of the results obtained, and more generally on that of the first law itself. In the course of these comments we also will mention some further related work which either has been done already, or would seem to be worth doing.

As its name makes clear, the most important implications of the so-called first law pertain to the thermodynamic attributes of black holes; however the associated issue of stability is important in its own right (and arises already in the purely classical setting).

By virtue of being time-independent, a stationary black hole solution may be characterized as a state of equilibrium of a self-gravitating system. If this equilibrium is to be stable in the thermodynamic sense, then of course it must (at least locally) be a maximum of the entropy S (we use 'S' for entropy in this section). Conversely, if some configuration does maximize S, then the second law of thermodynamics implies that the equilibrium state in question is indeed a stable one. Assuming that S is a smooth function on the relevant space of "configurations" or "states", a more or less necessary and sufficient condition for equilibrium is thus that S have vanishing gradient and negative definite Hessian. More generally, one may just define an equilibrium configuration to be an extremum of S, and diagnose its (thermodynamic or "secular") stability by the behavior of the Hessian there.

Actually, the criteria we have just stated need to be qualified, as equilibria need not be unconditional maxima of S, but only so at fixed values of the relevant conserved quantities. In the black hole situation, the relevant quantities will usually be the energy E, the angular momentum J, and the electric charge Q. If we identify the area A with the entropy (up to a numerical factor), then a necessary condition for a black hole configuration to represent a thermodynamic equilibrium state is that a relation of the form (62) (i.e. $dA = \beta dE - \beta\omega dJ - \beta V dQ$) hold for *arbitrary* variations of the fields. (Clearly this equation entails extremality of S at fixed E, J and Q; and conversely, extremality implies a relation of this general form, although of course, it does not give us the specific expressions (24), (V.3c), etc. for the coefficients β, V, ω, etc. which appear in (62).) This relation to thermodynamic equilibrium is the main reason why it is important that (62) hold more generally than just for variations from one stationary solution to another.

Although the satisfaction of equation (62) qualifies a black hole solution as an equilibrium state in the above sense, it has nothing directly to say about stability. As we just pointed out, this concerns the second derivatives of the entropy, or in the classical limit, of the horizon area. In thus using A as the measure of classical stability, we of course presuppose the more general identification of area with entropy, which rests first of all on (62) itself. In the $\hbar \to 0$ limit, this geometrical contribution dominates all other sources of entropy, whence the latter can be ignored. The consequent fact that A can be used to diagnose classical stability is consistent with the interpretation of the classical law of area increase as expressing the second law of thermodynamics for black holes, but unfortunately the area law can be justified as an independent classical fact only to the extent that "cosmic censorship" holds—one of many examples of the close interconnection between cosmic censorship and black hole thermodynamics. If we could establish cosmic censorship, or prove the area law on some independent basis then area maximization could be used as a criterion of stability independently of any thermodynamic argument: all that matters is that horizon area be (classically) non-decreasing, and therefore a valid "Lyapunov functional".

In any case, in order to investigate stability on the basis of horizon area, one should determine the negative-definiteness (or lack thereof) of the "reduced Hessian", $d^2A - \beta d^2E$, or more generally of $d^2A - \beta d^2E - \beta V d^2Q - \beta\omega d^2J$. One might hope to prove negative-definiteness in all directions except those corresponding to "super-radiant modes" in the rotating case. We have not studied this Hessian directly; however there exists an indirect way to investigate its eigensigns, namely the so-called turning point method commonly employed in studies of stellar stability and the stability of other astrophysical objects (cf.[11, 12, 13]). This method can prove instability, but can only make stability plausible. As applied to the Reissner-Nordstrom and Kerr-Newman families of black holes, it indicates stability, or rather it indicates that nowhere along the sequence does a new unstable mode come into being and remain for a finite range of parameter value [12]. Indirectly, this implies that the above reduced Hessian is indeed negative in the required sense. A direct confirmation would be of considerable interest.

We believe that our method of proving the "extended first law" for Einstein gravity is simpler than previous ones (cf. [5]), and that for this reason, it clarifies as well the origin of the "unextended first law", the one applying only to variations from one stationary solution to another. Aside from this possibly greater simplicity and the associated thoroughgoing spacetime character of the definitions and the derivation, the most important way in which our result differs technically from earlier ones is that (without making any reference to a possible "bifurcation surface") it allows the hypersurface Σ on which the energy, area, etc. are evaluated to intersect the horizon in an arbitrary manner, and in particular to intersect it arbitrarily far into the future. (The importance of this type of hypersurface freedom was stressed also in [14].) This freedom in the choice of Σ is mandated physically, since it is only some future-segment of a Schwarzschild metric (say) that is relevant to an astrophysically realistic black hole like one formed in stellar collapse. But the more important issue of principle, in our view, has to do with the second law itself.

For the theory of black hole thermodynamics to achieve a satisfactory status, it will be necessary not only to derive the equilibrium value of the entropy from first principles, but also to explain why the law of entropy increase continues to hold when black holes are present. One attempt to foresee how such a proof would go [15], presupposes that (say in a setting which is near enough to classical that an approximately well defined identification of individual spacelike hypersurfaces can be made) the total entropy in question is that of an effective quantum density-operator associated to the portion of each such hypersurface lying outside the black hole(s). The second law is then identified with the assertion that this entropy is weakly increasing as the hypersurface advances in time along the horizon (or at spatial infinity).

For consistency, then, it would be necessary that our earlier considerations on area extremization hold for such hypersurfaces, as well as for ones tied, for example, to a bifurcation 2-surface. In fact, the desire to prove extremality for such hypersurfaces was our second main reason for writing this paper.

Extremal theorems can also be approached from a Hamiltonian point of view, as we mentioned in the Introduction. Recently, such an approach was employed in [9], the result being a theorem similar to ours, except that the hypersurface on which the canonical variables were defined was assumed to meet the horizon in a bifurcation 2-surface. Other recent papers dealing with these same general issues are [16, 17, 18], the last two of which are rather similar in spirit to our own. Some of these are concerned with higher derivative or higher dimensional gravity. There is little about our methods which would not seem to apply in those cases as well, but that is a question we have not looked into.

In concluding, let us mention a final technical point and a final issue for further study. In our derivation we have defined the energy and angular momentum directly from the Noether operator expressions described in Section II. For completeness, one would like to confirm that these definitions agree in general with the ADM ones, given appropriate asymptotic falloffs. This has been done and will be reported by one of us in another place [19]. Related to this way of introducing conserved quantities, is the fact that E, for example, splits naturally into a sum of an "exterior energy"—a spatial volume integral—and a surface contribution from the horizon (see equations (13) and (47), or (14) and (48) for J). This raises the question whether (perhaps via a suitable choice of background connection $\overset{\circ}{\nabla}$) these separate terms (or their variations) might have any individual significance, so that one could in this way meaningfully disentangle the energy or angular momentum "of the black hole" from that "of the exterior matter".

The work of RDS was partially supported by NSF grant PHY 9307570.

Appendix

A The asymptotic fall off conditions on the metric

The following notations/conventions will be used in the rest of the Appendix. Fix an asymptotically Minkowskian coordinate system (t, x, y, z) in a neighborhood of spatial infinity. Greek indices will denote 4-d spacetime Cartesian components, Latin indices i, j, k... will denote 3-d spatial Cartesian components and Latin letters a, b, c... will be abstract indices. The matrix $\eta_{\mu\nu}$ representing the fixed flat metric at spatial infinity will be diagonal with components $(-1,1,1,1)$. We also choose our fixed background connection $\overset{\circ}{\nabla}_\mu$ to coincide with the Cartesian coordinate derivative ∂_μ in a neighborhood of spatial infinity.

We define the inversion map, $\mathcal{I}$, on a neighborhood of spatial infinity in the following way:

$$\mathcal{I}(x^\mu) = -x^\mu \tag{63}$$

The point-map $\mathcal{I}$ naturally induces a map on tensor fields in a neighborhood of spatial infinity which we shall also call $\mathcal{I}$. Our falloff conditions will entail that, to leading nontrivial order,

$$\begin{aligned} \mathcal{I} g_{ab} &= g_{ab} \\ \mathcal{I} \mathring{\nabla}_c g_{ab} &= \mathring{\nabla}_c g_{ab} \end{aligned} \tag{64}$$

The full set fall off conditions on the spacetime metric g_{ab} at spatial infinity in terms of $r = (x^2 + y^2 + z^2)^{1/2}$ are as follows (spatial infinity being approached as $r \to \infty$ at fixed t).
Let $h_{\mu\nu} := g_{\mu\nu} - \eta_{\mu\nu}$. Then
(a) $h_{\mu\nu} = \alpha_{\mu\nu}(x^\tau)/r + O(1/r^2)$,
where $\alpha_{\mu\nu}(x^\tau)$ is bounded and $\alpha_{\mu\nu}(-x^\tau) = \alpha_{\mu\nu}(x^\tau)$
(b) $\partial_\alpha h_{\mu\nu} = \beta_{\alpha\mu\nu}(x^\tau)/r^2 + O(1/r^3)$,
where $\beta_{\alpha\mu\nu}(x^\tau)$ is bounded and $\beta_{\alpha\mu\nu}(-x^\tau) = -\beta_{\alpha\mu\nu}(x^\tau)$
(c) $\partial_\alpha \partial_\beta h_{\mu\nu} = O(1/r^3)$

Now consider an asymptotically flat spatial slice, Σ, which is asymptotically a $t = a$ slice (a being some constant). Conditions (a) and (b) above relate fields on the $t = a$ slice to those on the $t = -a$ slice, but for our estimates, we would like conditions dealing exclusively with fields on a single slice. The fields at $(x^i, t = -a)$ are related to those at $(x^i, t = a)$ by an integration of their derivatives over a finite time interval ($\Delta t = 2a$). Using this fact we obtain, on a fixed slice Σ:
(a1) $h_{\mu\nu} = \alpha_{\mu\nu}(x^i)/r + O(1/r^2)$
(b1) $\partial_\alpha h_{\mu\nu} = \beta_{\alpha\mu\nu}(x^i)/r^2 + O(1/r^3)$
where $\alpha_{\mu\nu}$ and $\beta_{\alpha\mu\nu}$ in (a) and (b) have been restricted to Σ and are respectively even and odd functions of x^i for large r.

The evenness conditions expressed in (64), (a), (b), (a1) and (b1) are often called "parity conditions". In this appendix, we will employ the term more generally to denote our falloff conditions as a whole.

B Asymptotic patching and angular momentum

Lemma: Let g_{ab} and $\hat{g}_{ab}$ be two metrics which satisfy the parity conditions of part A of this appendix, and set $J(g_{ab}) := \int_\Sigma T_a^b \cdot \phi^a dS_b$ where T_b^a is the gravitational Noether operator and ϕ^a is an asymptotically Killing, rotational vector field which commutes with $\overset{\circ}{\nabla}$ in a neighborhood of infinity. In a neighborhood of Σ, let

$$\overline{g}_{ab} := \ bg_{ab} + (1-b)\hat{g}_{ab}$$

where the "patching function" $b(x)$ is defined as follows. Fix a smooth function f on $[0,\infty]$ such that $f(y) = 1$ for $y < 1$ and $f(y) = 0$ for $y > 2$. Choose a radius R large enough so that the asymptotic coordinate system is defined for $r > R/2$, and set $b(x) = f(r/R)$. (Strictly speaking, r might not be defined throughout the region enclosed by the $r = R/2$ two-sphere, but we still put $b \equiv 1$ on that region.) Then the difference $J(\overline{g}_{ab}) - J(g_{ab})$ can be made as small as desired by choosing R big enough.

Proof: A useful expression for $T^{\nu}_{\mu} \cdot \phi^{\mu}$ is:

$$\begin{aligned} T^{\tau}_{\sigma} \cdot \phi^{\sigma} \quad &= \quad -\frac{1}{2}(\nabla_{\eta}\phi_{\rho})[2\Gamma^{\tau}_{\sigma\alpha}g^{\eta\alpha}g^{\sigma\rho} - \Gamma^{\nu}_{\nu\sigma}(g^{\eta\tau}g^{\sigma\rho} + g^{\eta\sigma}g^{\tau\rho}) + \Gamma^{\nu}_{\nu\sigma}g^{\eta\rho}g^{\sigma\tau} - \Gamma^{\tau}_{\beta\alpha}g^{\beta\alpha}g^{\eta\rho}] \\ &\quad + \frac{1}{2}\phi^{\tau}[R - \nabla_{\rho}(\Gamma^{\rho}_{\beta\alpha}g^{\beta\alpha} - \Gamma^{\nu}_{\nu\sigma}g^{\sigma\rho})] \end{aligned} \tag{65}$$

(In this equation, 'R' stands for the Ricci scalar, of course; it is not the radius parameter involved in the definition of the patching function.) Since the second derivatives $\partial\partial g$ cancel out in the second term of (65), the main term to worry about is the first, which might in principle decay as slowly as $O(1/r^2)$ because $\partial\phi$ is $O(1)$ and Γ is $O(1/r^2)$. Thus the crucial fact about the first term is that the coefficient of $\nabla_{\eta}\phi_{\rho}$ is symmetric in η and ρ. Using this, and taking R big enough so that ϕ^{σ} is a Killing vector for the background flat metric $\eta_{\mu\nu}$ for $r > R$ one has symbolically (in the spirit of [2]),

$$I = J(\overline{g}) - J(g) = \int_{r>R} \partial\overline{g}\partial\overline{g}\phi \ + \ \int_{r>R} \partial\phi(\overline{g} - \eta)\partial\overline{g} \tag{66}$$

As a result, the contributions to I are of the following type (as could also have been seen from the footnote to equation (4) in reference [3]):
$I_1 = \int_R^{2R} bb(\partial h)(\partial h)\phi \quad I_2 = \int_R^{2R} b(\partial b)(\partial h)h\phi \quad I_3 = \int_R^{2R} hh(\partial b)(\partial b)\phi$
$I_4 = \int_R^{2R} (\partial\phi)hbb\partial h \quad I_5 = \int_R^{2R} (\partial\phi)hbh\partial b \quad I_6 = \int_R^{\infty} [(\partial h)(\partial h)\phi + (\partial\phi)h\partial h]$
$I_7 = \int_{2R}^{\infty} [(\partial h)(\partial h)\phi + (\partial\phi)h\partial h]$
(Note: In the above all indices have been suppressed and by h we mean either h or $\hat{h}$)

Now since $\partial b = O(1/r)$, none of these integrals can be worse than logarithmically divergent. In fact we now show that they all are bounded and vanish as $R \to \infty$. We use heavily that b and the $\alpha_{\mu\nu}$ are even functions on the 2-sphere at constant r and that the $\partial_\mu b, \beta_{\alpha\mu\nu}$ and ϕ^μ are odd functions on the 2-sphere at constant r. We denote the element of solid angle on the 2-sphere by $d\Omega$. Consider I_1 as a typical example. We have

$$\begin{aligned} I_1 &= \int_R^{2R} b\, b\, \partial h\, \partial h\, \phi \\ &\leq \int_R^{2R} bb(\beta/r^2)(\beta/r^2)\phi r^2 dr d\Omega \;+\; \int_R^{2R} (bbC_1/r^5)\phi r^2 dr d\Omega \end{aligned}$$

where for ease of writing we have omitted absolute value signs around the the members of the inequality. Using the parity conditions for b, β, ϕ and that $|\phi^\mu| < 2R,\ |b| < 1$ we get:

$$I_1 \leq \int_R^{2R} (odd\ function/r^3) r^2 dr d\Omega \;+\; 8\pi C_1/R \tag{67}$$

$\Rightarrow I_1 \leq A_1/R$ for some constant A_1. In a similar fashion, I_i for $i = 2$ to 5 are all bounded by some A_i/R. (To show this, one also has to use, in addition to parity arguments, the fact that $|\partial b| \leq constant/R$, which follows directly from its definition as $f(r/R)$.) Using parity arguments it is similarly easy to show that $I_6,\ I_7$ are also bounded by some A_6/R and A_7/R respectively. Therefore, as $R \to \infty$ all the above contributions to I vanish, and we have proved the lemma.

Remark: In relation to the case of a slice without boundary, we could also apply this lemma to a proof of an extremum theorem for angular momentum along the lines of the extremum theorem for mass in [2] (cf. [5]). Notice in this connection that, our asymptotic conditions are about as weak as one could expect to be physically relevant, and as such are less restrictive than those used in the proof of the extremum theorem for angular momentum given in [4].

C The asymptotic fall off conditions on A_a

We use the notation of section A above. As there, our fall off conditions will involve parity conditions under the inversion map $\mathcal{I}$, requiring to leading order in $(1/r)$ that

$$\mathcal{I} A_a = -A_a \tag{68}$$

More specifically our conditions are

(a) $A_\mu = \ \chi_\mu(x^\tau)/r \ + \ O(1/r^2)$ where $\chi_\mu(x^\tau)$ is bounded and $\chi_\mu(-x^\tau) = \chi_\mu(x^\tau)$
(b) $\partial_\nu A_\mu = \rho_{\mu\nu}(x^\tau)/r^2 + O(1/r^3)$ where $\rho_{\mu\nu}$ is bounded and $\rho_{\mu\nu}(-x^\tau) = -\rho_{\mu\nu}(x^\tau)$
(c) $\partial_\alpha\partial_\beta A_\mu = O(1/r^3)$

The same reasoning as in A shows that on a 't =constant' slice Σ, (a) and (b) imply:
(a1) $A_\mu = \ \chi_\mu(x^i)/r \ + \ O(1/r^2)$
(b1) $\partial_\nu A_\mu = \rho_{\mu\nu}(x^i)/r^2 + O(1/r^3)$
where χ_μ and $\rho_{\mu\nu}$ have been restricted to Σ and are respectively even and odd functions on the r =constant, t =constant 2-sphere for large r.

Finally, we may observe that all the arguments in the paper would go through if we imposed the opposite parity on A_a from that expressed in (68). Our choice of sign was made only to conform with the character of the field due to a static point charge at the origin of Minkowski spacetime. It means in effect that the spacetime inversion $\mathcal{I}$ has been chosen to act as PT rather than as CPT.

References

[1] R.D. Sorkin, "Conserved quantities as Action variations", in *Mathematics and general relativity*, ed. J.W. Isenberg, (pg. 23-37, Vol 71 in the AMS's Contemporary Mathematics series) (1988).

[2] B.F. Schutz and R.D. Sorkin, Ann. Phys. **107**, 1 (1977).

[3] R.D. Sorkin, Proc. R. Soc. Lond. A **435**, 635 (1991).

[4] E. Nahmad-Achar and B. Schutz, Class. Quantum Grav. **4**, 929 (1987).

[5] R.D. Sorkin, "Two Topics concerning Black Holes: Extremality of the Energy, Fractality of the Horizon", in S. Fulling (ed.), *Proceedings of the Conference on Heat Kernels and Quantum Gravity held Winnipeg, Canada, August, 1994* (University of Texas Press, to appear).

[6] J. Katz, "A Note on Komar's Anomalous Factor", *Class. Quant. Grav* **2**:423 (1985).

[7] R.D. Sorkin, "On Stress-Energy Tensors", *Gen. Rel. Grav.* **8**:437-449 (1977)

[8] R.M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).

[9] D. Sudarsky and R. Wald, Phys. Rev. D **46**, 1453 (1992).

[10] *Black Holes: The Membrane Paradigm* (edited by D.A. MacDonald, R.H. Price and K.S. Thorne, Yale University Press, New Haven, 1986).

[11] R.D. Sorkin, Ap. J. **249**:254-257 (1981).

[12] R.D. Sorkin, Ap. J. **257**:847-854 (1982).

[13] J.L. Friedman, J.R. Ipser and R.D. Sorkin, Ap. J. **325**, 722 (1988).

[14] T.D. Jacobson, G. Kang and R.C. Myers, "On Black Hole Entropy", *Phys. Rev. D***49**:6587 (1994), (gr-qc preprint no. 9312023).

[15] R.D.Sorkin Phys. Rev. Lett. **56**, 1885-1888 (1986). See also R.D. Sorkin, "On the Entropy of the Vacuum Outside a Horizon", in B. Bertotti, F. de Felice, Pascolini, A., (eds.), *Tenth International Conference on General Relativity and Gravitation (held Padova, 4-9 July, 1983), Contributed Papers*, vol. II, pp. 734-736 (Roma, Consiglio Nazionale Delle Ricerche, 1983) and Bombelli, L., Koul, R.K., Lee, J. and R.D. Sorkin, *Phys. Rev.* **D34**, 373-383 (1986).

[16] T.D. Jacobson and R.C. Myers, Phys. Rev. Lett., **70**, 3684 (1993).

[17] T.D. Jacobson, G. Kang and R.C. Myers, “On Black Hole Entropy”, *Phys. Rev. D*__49__:6587 (1994), (gr-qc preprint no. 9312023).

[18] R.M. Wald, Phys. Rev. D **48**, R3427 (1993).

[19] M. Varadarajan, In preparation.