Gedanken experiments to destroy a black hole. II. Kerr-Newman black holes cannot be overcharged or overspun

Jonathan Sorce
Department of Physics, University of Chicago, Chicago, Illinois 60637, USA

Robert M. Wald
Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago, Illinois 60637, USA

(Received 21 July 2017; published 13 November 2017)

We consider gedanken experiments to destroy an extremal or nearly extremal Kerr-Newman black hole by causing it to absorb matter with sufficient charge and/or angular momentum as compared with energy that it cannot remain a black hole. It was previously shown by one of us that such gedanken experiments cannot succeed for test particle matter entering an extremal Kerr-Newman black hole. We generalize this result here to arbitrary matter entering an extremal Kerr-Newman black hole, provided only that the nonelectromagnetic contribution to the stress-energy tensor of the matter satisfies the null energy condition. We then analyze the gedanken experiments proposed by Hubeny and others to overcharge and/or overspin an initially slightly nonextremal Kerr-Newman black hole. Analysis of such gedanken experiments requires that we calculate all effects on the final mass of the black hole that are second-order in the charge and angular momentum carried into the black hole, including all self-force effects. We obtain a general formula for the full second order correction to mass, $\delta^2 M$, which allows us to prove that no gedanken experiments of the generalized Hubeny type can ever succeed in overcharging and/or overspinning a Kerr-Newman black hole, provided only that the nonelectromagnetic stress-energy tensor satisfies the null energy condition. Our analysis is based upon Lagrangian methods, and our formula for the second-order correction to mass is obtained by generalizing the canonical energy analysis of Hollands and Wald to the Einstein-Maxwell case. Remarkably, we obtain our formula for $\delta^2 M$ without having to explicitly compute self-force or finite size effects. Indeed, in an appendix, we show explicitly that our formula incorporates both the self-force and finite size effects for the special case of a charged body slowly lowered into an uncharged black hole.

DOI: 10.1103/PhysRevD.96.104014

I. INTRODUCTION

The Kerr-Newman family of metrics are the unique stationary, asymptotically flat black hole solutions of the Einstein-Maxwell equations in 4 spacetime dimensions. The Kerr-Newman metrics comprise a 3-parameter family of solutions parametrized by mass M, charge Q, and angular momentum $J = Ma$. However, these solutions describe black holes only for a limited region of this parameter space, characterized by the inequality

$$M^2 \geq (J/M)^2 + Q^2. \tag{1}$$

When this inequality is not satisfied, the spacetime contains a naked singularity, i.e., the singularity is visible from infinity.

The above facts give rise to a possible means of testing the weak cosmic censorship conjecture [1,2], which states that all singularities arising from gravitational collapse must be hidden within black holes, so that no physical process can give rise to a naked singularity. Suppose that we start with a Kerr-Newman black hole satisfying (1). Now throw/drop matter into the black hole carrying energy E, angular momentum, ℓ, and charge q, so that the final state will have mass $M + E$, angular momentum $J + \ell$, and charge $Q + q$. Then if ℓ and/or q can be made sufficiently large compared with E, the inequality (1) will be violated, resulting in a contradiction with the final state being a black hole.

The most obvious case to consider for an attempt to destroy a black hole in this manner would be to start with an extremal black hole, satisfying $M^2 = (J/M)^2 + Q^2$, and to throw in particle matter. This case was analyzed in 1974 by one of us in paper I of this series [3]. It was shown in paper I that no violations of (1) can occur by throwing particle matter into an extremal Kerr-Newman black hole. The nature of this result is well illustrated by considering the special case of attempting to "overcharge" an extremal Reissner-Nordstrom ($Q = M$) black hole. Let ξ^a denote the horizon Killing field, which, for a Reissner-Nordstrom black hole, coincides with the static Killing field $(\partial/\partial t)^a$. A test particle with mass m and charge q in this spacetime has energy given by

$$E = -(mu_a + qA_a)\xi^a, \tag{2}$$

where u_a is the four-velocity of the particle and A_a is the vector potential of the black hole's electromagnetic field.

© 2017 American Physical Society

Since ξ^a is null on the horizon, the first term $-mu_a\xi^a$ is non-negative on the horizon, although it can be made arbitrarily small. Thus, the energy of a particle that crosses the horizon is bounded below by the electromagnetic potential energy term

$$E \geq q\Phi_H, \tag{3}$$

where $\Phi_H = (-A_a\xi^a)|_H$ is the electromagnetic potential evaluated on the horizon. However, $\Phi_H = 1$ for an extremal Reissner-Nordstrom black hole, so any particle that enters the black hole must satisfy

$$E \geq q. \tag{4}$$

Consequently, we have $M + E \geq Q + q$, so (1) holds. In other words, any particle with sufficiently large charge q as compared with E to produce a violation of (1) for the final state would be repelled by the electric field of the black hole and thus cannot enter it. As shown in paper I [3], similar results hold for attempting to overcharge and/or overspin a general extremal Kerr-Newman black hole using particle matter.

Nevertheless, in 1999 Hubeny [4] proposed that violations of (1) might still occur if one suitably added matter to a slightly nonextremal black hole. To see this, consider a slightly nonextremal Reissner-Nordstrom black hole. It is useful to introduce the dimensionless parameter

$$\epsilon = \frac{\sqrt{M^2 - Q^2}}{M}, \tag{5}$$

so that $\epsilon \to 0$ in the extremal limit. For $\epsilon \ll 1$, we have

$$\Phi_H = Q/r_+ \approx 1 - \epsilon, \tag{6}$$

where $r_+ = M + \sqrt{M^2 - Q^2}$ is the horizon radius. In place of (4) we now obtain

$$E \geq q(1 - \epsilon). \tag{7}$$

Consequently, for this lower bound for E, we have

$$(M + E) - (Q + q) \approx -\epsilon q + \frac{M\epsilon^2}{2}. \tag{8}$$

Thus, it might appear that we can obtain a violation of (1) by taking $q > \epsilon M/2$ (but still keeping $q \ll Q$).

The main difficulty with Hubeny's argument is that for $q \sim \epsilon M$, the violation of (1) given by (8) is of order $\epsilon q \sim q^2/M$. Consequently, to determine if one truly can obtain a violation of (1), the quantities appearing in (8) must all be calculated consistently to the appropriate order. Specifically, the energy, E, of the matter must be calculated to order q^2. However, formula (2) applies only to "test matter" and is valid only to linear order in q; it does not take into account the contributions of electromagnetic self-energy (which require consideration of bodies of finite size) or the energy contributed by self-force effects, both of which enter at order q^2. In particular, it is possible that self-force effects could contribute to a repulsion of the body from the black hole, requiring that the body be given additional energy at order q^2 in order to enter the black hole.

Similar potential violations of (1) have been found for Reissner-Nordström black holes absorbing angular momentum [5], Kerr black holes absorbing charge or angular momentum [6–8], and for generic Kerr-Newman black holes [9,10]. However, just as in Hubeny's argument, in order to determine whether these potential violations actually occur, one needs to calculate all contributions to energy that are quadratic order in the relevant parameters of the particle. This would appear to require a complete analysis of self-force effects as well as finite size effects and any other effects that might enter at this order.

Unfortunately, the analytic computation of electromagnetic and gravitational self-force effects on the motion of bodies near a Kerr-Newman black hole is well beyond present capabilities. Thus, the main results that have been obtained thus far have come from numerical simulations. Numerical work has indicated that the self-force on particles falling into black holes may suffice to prevent Hubeny-type violations from occurring in the specific cases of overcharging a nearly extremal Reissner-Nordström black hole [11] and overspinning a nearly extremal Kerr black hole [12–15]. However, even for these special cases, no general analysis has been given of the second order corrections to energy. As such, there is no general proof that the cosmic censorship inequality (1) holds at quadratic order for processes involving matter that falls into nearly extremal Kerr-Newman black holes.

The main purpose of this paper is to give a complete analysis—valid to second order—of the contributions to the mass of a black hole for arbitrary matter that enters a black hole. At linear order, we derive a general expression—first obtained in [16]—that expresses δM in terms of the flux of charge and angular momentum carried into the black hole together with the nonelectromagnetic energy flux. Assuming only that the nonelectromagnetic contribution to the stress energy tensor satisfies the null energy condition, we will prove that for arbitrary processes involving matter falling into an exactly extremal Kerr-Newman black hole, no violation of (1) can occur at linear order in the perturbation. This result, which was previously obtained for charged scalar matter in [17] and generalized in [18], generalizes the results derived for particle matter in paper I [3] to completely general matter.

We then consider the possible Hubeny-type violations that might occur for slightly nonextremal black holes. Our general formula for δM shows that the linear order process obeys a generalization of (7), thus allowing the possibility of a violation of (1) but requiring an analysis of the second order effects on energy. We will perform this analysis by expressing the second order change in mass, $\delta^2 M$, of the black hole in terms of the canonical energy of the first order perturbation. We will then make the additional assumption that the *nonextremal* black hole is stable under *linear* perturbations, so that the first order perturbation decays to a stationary final state. This will allow us to evaluate the canonical energy in terms of a positive flux contribution through the horizon and a contribution from the final stationary perturbation. The resulting formula gives rise to an inequality on $\delta^2 M$, and we will see that this inequality is just what is needed to prove that no violations of the Hubeny type can ever occur. Remarkably, we are able to derive this inequality—which automatically takes account of all self-force and finite size effects—without having to explicitly calculate these effects themselves. We will show by explicit calculation in the Appendix that for the special case of lowering a charged body into an uncharged black hole, our general formula corresponds precisely to taking these effects into account.

Our analysis differs from most previous analyses—including that of paper I [3]—in the following three key respects: (1) We consider completely general matter rather than particle matter. Of course, "particle matter" makes sense in general relativity only when considered to be a limiting case of general matter as described in [19,20], so the general results derived in this paper also automatically hold for physically realizable particle matter. (2) Rather than analyzing the motion of bodies to determine what trajectories will or will not enter the black hole, we simply restrict consideration to the case where all matter that is initially present enters the black hole, and we compute the second order variation of the mass for this case. This allows us to derive the desired inequality without having to calculate the motion of bodies. (3) Most importantly, we obtain an exact expression for the full second order effects on the mass of a black hole. This allows us to obtain the above-mentioned inequality on $\delta^2 M$.

In Sec. II, we obtain the general variational formulas that we will need, including the generalization of the notion of canonical energy introduced in [21] for vacuum perturbations of vacuum black holes to the Einstein-Maxwell case. The gedanken experiments to destroy an extremal black hole are analyzed in Sec. III. We consider a perturbation of the black hole involving matter with charge and angular momentum such that the black hole is initially unperturbed in a neighborhood of the horizon and such that all of the matter eventually falls into the black hole. We obtain a general expression for δM that was first derived in [16]. We show that this expression yields an inequality that is sufficient to show that no violations can occur at linear order for extremal black holes, as previously found in [18]. This generalizes the results of paper I to completely general matter whose nonelectromagnetic stress-energy satisfies the null energy condition. The Hubeny-type gedanken experiments to destroy a slightly nonextremal black hole are considered in Sec. IV. We consider a process that is optimal at first order so that the first order perturbation saturates our lower bound on δM. We obtain an expression for $\delta^2 M$ involving the canonical energy of the first order perturbation. Assuming that the first order perturbation of the nonextremal black hole becomes stationary at late times (i.e., that the nonextremal black hole is linearly stable), we obtain a lower bound on $\delta^2 M$ that is sufficient to prove that no violations of (1) can occur. A simple pictorial representation of our results is presented in Sec. V. The relationship between our results and the electromagnetic self-force and self-energy is detailed in the Appendix for the case of a charged body lowered into an uncharged black hole.

Our metric signature, curvature, and abstract index conventions follow [22]. In many instances, we will suppress the indices on differential forms, in which case they will be denoted with boldface letters.

II. VARIATIONAL IDENTITIES AND CANONICAL ENERGY FOR EINSTEIN-MAXWELL THEORY

In this section, we generalize the canonical energy results obtained in [21] for vacuum perturbations of vacuum black holes to the Einstein-Maxwell case. It would be most natural to treat the electromagnetic field A_a as a connection on a principal $U(1)$-bundle and use the framework developed by Prabhu [23] for doing the Lagrangian analysis in the principal bundle. However, since this would require the introduction of considerable machinery and formalism, we will bypass this here and simply treat A_a as the one-form that one obtains on spacetime by making a choice of gauge. This leads to some awkwardness in that we will work—as is conventional—in a gauge such that, in the background black hole spacetime, A_a is stationary, $\pounds_\xi A_a = 0$, and $A_a \to 0$ at infinity, so the "horizon potential" $\Phi_H = -\xi^a A_a|_{\mathcal{H}}$ is nonvanishing, where ξ^a is the horizon Killing field and $\mathcal{H}$ denotes the future event horizon. Since $\xi^a = 0$ on the bifurcation surface, this implies that, in our gauge, A_a cannot be smooth at the bifurcation surface as a one-form on spacetime, which might be thought to cause difficulties. In fact, no such difficulties occur, as can be seen by performing the analysis in the principal bundle in the framework of Prabhu [23]. Namely, the connection, A_a, is smooth as a one-form in the bundle and this is consistent with the nonvanishing of Φ_H because the lift of ξ^a to the bundle has nonvanishing vertical part. Nevertheless, to keep our discussion simple, we will perform our analysis on spacetime and ignore the nonsmoothness of the background A_a, relying on the fact that the analysis could have been performed in the principal bundle, where all fields are smooth.

Although our interest is in 4-dimensional Kerr-Newman black holes in Einstein-Maxwell theory, we will consider general diffeomorphism covariant theories in n-dimensional spacetimes in subsections II A and II B. In II A, we review the derivation of a fundamental variational identity for theories derived from a diffeomorphism covariant Lagrangian. We define canonical energy in II B. The Einstein-Maxwell case in 4 spacetime dimensions is explicitly considered in II C. Gauge invariance issues are treated in II D.

A. The linear variational identity

The Lagrangian for a diffeomorphism-covariant theory on an n-dimensional spacetime is given by an n-form $\mathbf{L}$ on spacetime, which is a local function of the metric, g_{ab}, its curvature, and symmetrized covariant derivatives of the curvature, and which may also depend on other tensor fields, ψ, and their symmetrized covariant derivatives. We refer to the full field configuration as $\phi = (g_{ab}, \psi)$. We vary the Lagrangian by considering a one-parameter family of field configurations, $\phi(\lambda)$, and taking derivatives of $\mathbf{L}$ with respect to λ. Throughout this paper, the notation "δ" will be used to denote derivatives evaluated at $\lambda = 0$, e.g.,

$$\delta\mathbf{L} = \frac{d\mathbf{L}}{d\lambda}\bigg|_{\lambda=0}, \qquad \delta^2\mathbf{L} = \frac{d^2\mathbf{L}}{d\lambda^2}\bigg|_{\lambda=0}, \qquad \delta\phi = \frac{d\phi}{d\lambda}\bigg|_{\lambda=0}. \tag{9}$$

The first-order variation of the Lagrangian can be written as

$$\frac{d\mathbf{L}}{d\lambda} = \mathbf{E}(\phi)\cdot\frac{d\phi}{d\lambda} + d\boldsymbol{\theta}\left(\phi, \frac{d\phi}{d\lambda}\right), \tag{10}$$

where $\mathbf{E}$ is locally constructed from the fields ϕ and their derivatives, while $\boldsymbol{\theta}$ is locally constructed from ϕ, $d\phi/d\lambda$, and their derivatives; $\boldsymbol{\theta}$ corresponds to the "boundary term" one would obtain by putting the variation of $\mathbf{L}$ under an integral sign and integrating by parts to remove all spacetime derivatives from $d\phi/d\lambda$. The Euler-Lagrange equations of motion of the theory are simply

$$\mathbf{E}(\phi) = 0. \tag{11}$$

The *symplectic current* $(n-1)$-form $\boldsymbol{\omega}$ is defined in terms of a second variation of $\boldsymbol{\theta}$. For a two-parameter family of field configurations $\phi(\lambda_1, \lambda_2)$, we define

$$\boldsymbol{\omega}\left(\phi; \frac{\partial\phi}{\partial\lambda_1}, \frac{\partial\phi}{\partial\lambda_2}\right) = \frac{\partial}{\partial\lambda_1}\boldsymbol{\theta}\left(\phi, \frac{\partial\phi}{\partial\lambda_2}\right) - \frac{\partial}{\partial\lambda_2}\boldsymbol{\theta}\left(\phi, \frac{\partial\phi}{\partial\lambda_1}\right). \tag{12}$$

The symplectic current depends on the background field configuration ϕ, as well as on the perturbations $\partial\phi/\partial\lambda_1$ and $\partial\phi/\partial\lambda_2$. If both of these perturbations satisfy the linearized equations of motion $\frac{\partial}{\partial\lambda_1}E(\phi) = \frac{\partial}{\partial\lambda_2}E(\phi) = 0$, then it follows from Eq. (10) that

$$d\boldsymbol{\omega} = 0, \tag{13}$$

i.e., the symplectic current is conserved.

The Noether current associated with an arbitrary vector field X^a is defined as

$$\boldsymbol{\mathcal{J}}_X(\phi) = \boldsymbol{\theta}(\phi; \mathcal{L}_X\phi) - \iota_X\mathbf{L}(\phi), \tag{14}$$

where $\iota_X\mathbf{L}$ denotes contraction of X^a into the first index of the differential form $\mathbf{L}$. A simple calculation [24] shows that the first variation of $\boldsymbol{\mathcal{J}}_X$ can be written as

$$\frac{d\boldsymbol{\mathcal{J}}_X}{d\lambda} = -\iota_X\left(\mathbf{E}(\phi)\cdot\frac{d\phi}{d\lambda}\right) + \boldsymbol{\omega}\left(\phi; \frac{d\phi}{d\lambda}, \mathcal{L}_X\phi\right) + d\left[\iota_X\boldsymbol{\theta}\left(\phi, \frac{d\phi}{d\lambda}\right)\right]. \tag{15}$$

On the other hand, it was shown in [25] that the Noether current can be written in the form

$$\boldsymbol{\mathcal{J}}_X = \mathbf{C}_X + d\mathbf{Q}_X, \tag{16}$$

where $\mathbf{Q}_X$ is called the *Noether charge* and $\mathbf{C}_X \equiv X^a\mathbf{C}_a$ are the constraints of the theory, so that $\mathbf{C}_a = 0$ when the equations of motion are satisfied. In particular, $d\boldsymbol{\mathcal{J}} = 0$ when the equations of motion are satisfied, as can be shown directly from the definition (14) of $\boldsymbol{\mathcal{J}}$.

By differentiating.[1] Eq. (16) with respect to λ and comparing it to Eq. (15), we obtain the fundamental identity

$$d\left[\frac{d\mathbf{Q}_X}{d\lambda} - \iota_X\boldsymbol{\theta}\left(\phi, \frac{d\phi}{d\lambda}\right)\right] = \boldsymbol{\omega}\left(\phi; \frac{d\phi}{d\lambda}, \mathcal{L}_X\phi\right) - \frac{d\mathbf{C}_X}{d\lambda} - \iota_X\left(\mathbf{E}(\phi)\cdot\frac{d\phi}{d\lambda}\right). \tag{17}$$

This identity forms the basis for all calculations conducted in the remainder of this paper.

Now, assume that $\phi(\lambda)$ is globally hyperbolic with Cauchy surface Σ. Evaluating (17) at $\lambda = 0$ and integrating the resulting equation over Σ, we obtain

$$\int_{\partial\Sigma}[\delta\mathbf{Q}_X - \iota_X\boldsymbol{\theta}(\phi, \delta\phi)] = \int_\Sigma \boldsymbol{\omega}(\phi; \delta\phi, \mathcal{L}_X\phi) - \int_\Sigma \delta\mathbf{C}_X - \int_\Sigma \iota_X(\mathbf{E}(\phi)\cdot\delta\phi). \tag{18}$$

[1] Note that we take X^a to be λ-independent.

A Hamiltonian h_X associated with a vector field X^a is a functional of ϕ such that if and only if ϕ satisfies the equations of motion, then under all variations $\delta\phi$ we have

$$\delta h_X = \int_\Sigma \boldsymbol{\omega}(\phi; \delta\phi, \mathcal{L}_X\phi). \tag{19}$$

If the spacetime is asymptotically flat and there is no "interior boundary" to Σ, then a Hamiltonian, h_X, conjugate to X^a must satisfy

$$\delta h_X = \int_\infty [\delta \mathbf{Q}_X - \iota_X \boldsymbol{\theta}(\phi, \delta\phi)] + \int_\Sigma \delta \mathbf{C}_X, \tag{20}$$

where "$\int_\infty$" denotes the limit to spatial infinity of integration over a suitable family of spacelike $(n-2)$-spheres. This motivates the following definition[2] of the ADM conserved quantity H_X conjugate to an asymptotic symmetry X^a for asymptotically flat solutions: H_X (if it exists) is the quantity such that, for all one-parameter families of solutions, we have

$$\delta H_X = \int_\infty [\delta \mathbf{Q}_X - \iota_X \boldsymbol{\theta}(\phi, \delta\phi)]. \tag{21}$$

Finally, let us restrict consideration to the case where (i) $\phi_0 = \phi(\lambda = 0)$ is a globally hyperbolic, asymptotically flat solution of the equations of motion, $\mathbf{E} = 0$, and (ii) ϕ_0 possesses a Killing field ξ^a that is also a symmetry of the matter fields ψ, so that $\mathcal{L}_\xi \phi_0 = 0$. Then (18) yields

$$\int_{\partial\Sigma} [\delta \mathbf{Q}_\xi - \iota_\xi \boldsymbol{\theta}(\phi, \delta\phi)] = -\int_\Sigma \delta \mathbf{C}_\xi. \tag{22}$$

The case of greatest interest for us is where ϕ_0 represents the exterior of a stationary black hole, and ξ^a is the horizon Killing field

$$\xi^a = t^a + \Omega_H \varphi^a, \tag{23}$$

where t^a is the timelike Killing field of ϕ_0, φ^a is the axial Killing field of ϕ_0, and Ω_H is the angular velocity of the horizon. The contribution to the boundary integral from infinity is then just

$$\int_\infty [\delta \mathbf{Q}_\xi - \iota_\xi \boldsymbol{\theta}(\phi, \delta\phi)] = \delta H_\xi = \delta M - \Omega_H \delta J, \tag{24}$$

where M is the ADM mass and J is the ADM angular momentum. If the spacetime represents the exterior of a black hole, then there will be a contribution from the "internal boundary" as well. We will evaluate this internal boundary contribution for Einstein-Maxwell theory in subsection C below.

B. Second order variations and canonical energy

Let us now continue to restrict consideration to the case where $\phi_0 = \phi(\lambda = 0)$ is a globally hyperbolic solution of the equations of motion that possesses a Killing field ξ^a that is also a symmetry of the matter fields ψ, so that $\mathcal{L}_\xi \phi_0 = 0$. Again, we do *not* require that the perturbation $\delta\phi = (d\phi/d\lambda)|_{\lambda=0}$ satisfy the linearized equations of motion. Let Σ be a Cauchy surface. We define the *canonical energy* of the perturbation $\delta\phi$ on Σ by

$$\mathcal{E}_\Sigma(\phi; \delta\phi) \equiv \int_\Sigma \boldsymbol{\omega}(\phi; \delta\phi, \mathcal{L}_\xi \delta\phi). \tag{25}$$

We can obtain an extremely useful expression for canonical energy by differentiating (17) with respect to λ and evaluating the resulting expression at $\lambda = 0$. We obtain

$$d[\delta^2 \mathbf{Q}_\xi - \iota_\xi \delta\boldsymbol{\theta}(\phi, \delta\phi)] = \boldsymbol{\omega}(\phi; \delta\phi, \mathcal{L}_\xi \delta\phi) - \delta^2 \mathbf{C}_\xi - \iota_\xi(\delta \mathbf{E} \cdot \delta\phi), \tag{26}$$

Here, the meaning of the "δ's" in the expression $\delta\boldsymbol{\theta}(\phi, \delta\phi)$ is that both derivatives in this term are to be evaluated simultaneously, i.e.,

$$\delta\boldsymbol{\theta}(\phi, \delta\phi) \equiv \left[\frac{d}{d\lambda}\boldsymbol{\theta}\left(\phi, \frac{d\phi}{d\lambda}\right)\right]\Big|_{\lambda=0}. \tag{27}$$

Integrating (26) over Σ, we obtain

$$\mathcal{E}_\Sigma(\phi; \delta\phi) = \int_{\partial\Sigma} [\delta^2 \mathbf{Q}_\xi - \iota_\xi \delta\boldsymbol{\theta}(\phi, \delta\phi)] + \int_\Sigma \delta^2 \mathbf{C}_\xi + \int_\Sigma \iota_\xi(\delta \mathbf{E} \cdot \delta\phi). \tag{28}$$

The case we are most interested in here is one where ϕ_0 corresponds to a stationary black hole, ξ^a is the horizon Killing field,[3] and Σ is a Cauchy surface for the exterior of the black hole. In that case, it follows from (21) that the contribution to the boundary term in (28) from infinity is

$$\int_\infty [\delta^2 \mathbf{Q}_\xi - \iota_\xi \delta\boldsymbol{\theta}(\phi, \delta\phi)] = \delta^2 M - \Omega_H \delta^2 J. \tag{29}$$

We will evaluate the interior boundary term at the end of the next subsection.

[2] We assume here that the matter fields fall off at infinity rapidly enough so as not to contribute to the surface integral on the right side of (21). Otherwise, these matter fields may make contributions of the form "potential times varied charge" that would need to be subtracted to obtain the conventional definition of ADM conserved quantities.

[3] Note that in [21], the canonical energy was defined with respect to the asymptotically timelike Killing field t^a rather than the horizon Killing field ξ^a. These quantities are equal to each other for axisymmetric perturbations, as considered in [21].

C. Einstein-Maxwell theory

We now consider Einstein-Maxwell theory in 4 spacetime dimensions and provide explicit expressions for many of the quantities appearing in the previous subsections. The Einstein-Maxwell Lagrangian is given by

$$\mathbf{L} = \frac{1}{16\pi}(R - F^{ab}F_{ab})\boldsymbol{\epsilon}, \tag{30}$$

where $\boldsymbol{\epsilon}$ is the volume element associated with the metric. For this Lagrangian, the field configuration consists of the metric and the vector potential, $\phi = (g_{ab}, A_a)$. As explained in the introductory paragraph to this section, we will treat A_a as a one-form on spacetime. The symplectic potential, Noether charge, equations of motion, and constraints for this Lagrangian were computed in [16]. The symplectic potential can be written as

$$\theta_{abc}\left(\phi, \frac{d\phi}{d\lambda}\right) = \theta^{GR}_{abc} + \theta^{\mathrm{EM}}_{abc}, \tag{31}$$

where

$$\theta^{GR}_{abc}\left(\phi, \frac{d\phi}{d\lambda}\right) = \frac{1}{16\pi}\epsilon_{dabc}g^{de}g^{fg} \times \left(\nabla_g \frac{dg_{ef}}{d\lambda} - \nabla_e \frac{dg_{fg}}{d\lambda}\right) \tag{32}$$

$$\theta^{\mathrm{EM}}_{abc}\left(\phi, \frac{d\phi}{d\lambda}\right) = -\frac{1}{4\pi}\epsilon_{dabc}F^{de}\frac{dA_e}{d\lambda}. \tag{33}$$

The Noether charge is given by

$$(Q_X)_{ab} = (Q^{GR}_X)_{ab} + (Q^{\mathrm{EM}}_X)_{ab}, \tag{34}$$

where

$$(Q^{GR}_X)_{ab} = -\frac{1}{16\pi}\epsilon_{abcd}\nabla^c X^d, \tag{35}$$

$$(Q^{\mathrm{EM}}_X)_{ab} = -\frac{1}{8\pi}\epsilon_{abcd}F^{cd}A_e X^e. \tag{36}$$

The equations of motion and constraints are given by

$$\mathbf{E}(\phi)\cdot\frac{d\phi}{d\lambda} = -\boldsymbol{\epsilon}\left[\frac{1}{2}T^{ab}\frac{dg_{ab}}{d\lambda} + j^a\frac{dA_a}{d\lambda}\right], \tag{37}$$

$$C_{bcda} = \epsilon_{ebcd}[T_a{}^e + A_a j^e]. \tag{38}$$

Here we have written $T_{ab} \equiv G_{ab} - 8\pi T^{\mathrm{EM}}_{ab}$—so that T_{ab} corresponds to the nonelectromagnetic part of the stress-energy tensor, and $j^a = (1/4\pi)\nabla_b F^{ab}$—so that j^a corresponds to the electromagnetic charge-current. Note that in the absence of sources, when both T_{ab} and j_a are zero, the constraints (38) vanish and the Euler-Lagrange equations of motion (37) are satisfied.

The symplectic current for the Einstein-Maxwell theory can be written in the form

$$\omega_{abc}\left(\phi; \frac{\partial\phi}{\partial\lambda_1}, \frac{\partial\phi}{\partial\lambda_2}\right) = \omega^{GR}_{abc} + \omega^{\mathrm{EM}}_{abc}, \tag{39}$$

where, from Eq. (31), we have

$$\omega^{GR}_{abc} = \frac{1}{16\pi}\epsilon_{dabc}w^d, \tag{40}$$

$$\omega^{\mathrm{EM}}_{abc} = \frac{1}{4\pi}\left[\frac{\partial}{\partial\lambda_2}(\epsilon_{dabc}F^{de})\frac{\partial A_e}{\partial\lambda_1} - \frac{\partial}{\partial\lambda_1}(\epsilon_{dabc}F^{de})\frac{\partial A_e}{\partial\lambda_2}\right], \tag{41}$$

where, in (40), we have

$$w^a = P^{abcdef}\left(\frac{\partial g_{bc}}{\partial\lambda_2}\nabla_d\frac{\partial g_{ef}}{\partial\lambda_1} - \frac{\partial g_{bc}}{\partial\lambda_1}\nabla_d\frac{\partial g_{ef}}{\partial\lambda_2}\right), \tag{42}$$

with

$$P^{abcdef} = g^{ae}g^{fb}g^{cd} - \frac{1}{2}g^{ad}g^{be}g^{fc} - \frac{1}{2}g^{ab}g^{cd}g^{ef} - \frac{1}{2}g^{bc}g^{ae}g^{fd} + \frac{1}{2}g^{bc}g^{ad}g^{ef}. \tag{43}$$

We now restrict attention to the case where $\phi_0 = \phi(\lambda = 0)$ is a stationary black hole solution to the Einstein-Maxwell equations (i.e., $T^{ab} = j^a = 0$ at $\lambda = 0$) with horizon Killing field ξ^a, and we let Σ be a Cauchy surface for the exterior region. In fact, by the black hole uniqueness theorems [22], ϕ_0 must be a Kerr-Newman solution, but we need not make use of this fact here. We work in a gauge where $\mathcal{L}_\xi A_a(\lambda = 0) = 0$ and $A_a(\lambda = 0) \to 0$ at infinity. As already discussed in the first paragraph of this section, in this gauge, $A_a(\lambda = 0)$ will, in general, be singular at the horizon, but this does not cause any difficulties. Furthermore, the variations δA_a and $\delta^2 A_a$ may be assumed to be smooth (as can be justified by working in the principal bundle framework of Prabhu [23]).

By definition, for a *nonextremal* black hole the horizon will be of bifurcate type, and Σ will terminate at the bifurcation surface B. For a nonextremal black hole, we now evaluate the boundary contribution to (22) arising from B. Since $\xi^a = 0$ on B, we have

$$\int_B [\delta\mathbf{Q}^{GR}_\xi - \iota_\xi\boldsymbol{\theta}^{GR}(\phi, \delta\phi)] = \int_B \delta\mathbf{Q}^{GR}_\xi = \frac{\kappa}{8\pi}\delta A_B, \tag{44}$$

where A_B is the area of B and κ is the surface gravity of the event horizon. To evaluate the electromagnetic contribution to the boundary term[4] at B, we note that by (33), $\boldsymbol{\theta}^{\mathrm{EM}}$ is smooth at B (since δA_a is smooth), so $\iota_\xi \boldsymbol{\theta}^{\mathrm{EM}} = 0$. However, by (36), we have

$$\delta \mathbf{Q}_\xi^{\mathrm{EM}} = -\frac{1}{8\pi}[\xi^e A_e \delta(\epsilon_{abcd}F^{cd}) + \xi^e(\delta A_e)\epsilon_{abcd}F^{cd})]. \tag{45}$$

Again, the second term vanishes at B on account of the smoothness of δA_a and the vanishing of ξ^a. However, the quantity

$$\Phi_H \equiv -[\xi^e A_e(\lambda)]|_{\mathcal{H}} \tag{46}$$

is, in general, nonvanishing at B. Since Φ_H must be constant on the horizon at $\lambda = 0$ [26] (see theorem 1 of [23] for a general proof for Yang-Mills fields), we find that the electromagnetic contribution to the boundary term at B is

$$\int_B [\delta \mathbf{Q}_\xi^{\mathrm{EM}} - \iota_\xi \boldsymbol{\theta}^{\mathrm{EM}}(\phi, \delta\phi)] = \frac{1}{8\pi}\Phi_H \int_B \delta(\epsilon_{abcd}F^{cd}) = \Phi_H \delta Q_B, \tag{47}$$

where Q_B is the electric charge flux integral over B.

The ingredients are now in place to write out (22) explicitly for a nonextremal black hole. We previously evaluated the boundary term from infinity in (24), and, in the previous paragraph, we have evaluated the boundary term from B. Using (38) and the fact that $T_{ab} = j^a = 0$ in the background spacetime (since ϕ_0 is a solution), we see that the remaining term $\delta \mathbf{C}_\xi$ takes the form

$$\delta C_{bcda}\xi^a = \epsilon_{ebcd}[\delta T_a{}^e + A_a \delta j^e]\xi^a \tag{48}$$

Thus, we see that (22) takes the explicit form

$$\delta M - \Omega_H \delta J - \frac{\kappa}{8\pi}\delta A_B - \Phi_H \delta Q_B = -\int_\Sigma \epsilon_{ebcd}[\delta T_a{}^e + A_a \delta j^e]\xi^a. \tag{49}$$

For source free perturbations, $\delta T_{ab} = \delta j_a = 0$, this yields the usual first law of black hole mechanics of Einstein-Maxwell theory.

It should be emphasized that (49) holds only for nonextremal black holes. In this paper, we will be concerned with both non-extremal and extremal black holes. However, it is clear from the derivation that (49) (with $\delta A_B = \delta Q_B = 0$) also holds for extremal black holes in the special case where Σ is not a Cauchy surface but rather an asymptotically flat hypersurface with one boundary at spatial infinity and the other boundary on the horizon at an early time such that the perturbation vanishes in a neighborhood of this internal boundary. In this case, there clearly will be no boundary contribution from the internal boundary of Σ. We will use (49) in this form for extremal black holes in Sec. III.

The canonical energy may also be split into gravitational and electromagnetic contributions

$$\mathcal{E}_\Sigma(\phi; \delta\phi) = \mathcal{E}_\Sigma^{GR} + \mathcal{E}_\Sigma^{\mathrm{EM}}. \tag{50}$$

Explicit formulas for these parts can be obtained from the definition (25), substituting from (40) and (41). These formulas are quite complicated and will not be written out explicitly here. Fortunately, we will need to evaluate the canonical energy integral only over (a portion of) the horizon (where its form simplifies considerably) and for stationary perturbations (where it can be evaluated straightforwardly).

We may now explicitly evaluate the terms appearing in (28) for Einstein-Maxwell theory, in exact parallel with our above evaluation of the terms appearing in (22). For a nonextremal black hole, we obtain[5]

$$\delta^2 M - \Omega_H \delta^2 J - \Phi_H \delta^2 Q_B - \frac{\kappa}{8\pi}\delta^2 A_B = \mathcal{E}_\Sigma(\phi; \delta\phi) - \int_\Sigma \iota_\xi(\delta \mathbf{E}(\phi)\cdot\delta\phi) - \int_\Sigma \delta^2 \mathbf{C}_\xi. \tag{51}$$

Again, this equation (with $\delta^2 A_B = \delta^2 Q_B = 0$) will hold for an extremal black hole if we restrict consideration to the case where both the first and second order perturbations vanish in a neighborhood of the horizon at the internal boundary of Σ. In Sec. IV, we will evaluate the right side of (51) in the context relevant to our calculations.

[4] We assume that $A_a t^a$ and $A_a \varphi^a$ fall off as $1/r$ and F_{ab} falls off as $1/r^2$ at infinity, so there is no electromagnetic contribution to the boundary term at infinity.

[5] It should be noted that since we take ξ^a to be fixed, the quantities Ω_H and κ do not vary. This means that if we perturb toward another stationary black with different values of Ω_H or κ, then ξ^a cannot be the horizon Killing field of the perturbed black hole. See [21] for further discussion.

D. Gauge invariance of canonical energy

In this subsection, we show that the canonical energy is gauge invariant when evaluated on linearized solutions to the Einstein-Maxwell equations, subject to the restrictions of Proposition 1 below. It should be noted that the symplectic form (i.e., the integral of $\boldsymbol{\omega}(\phi, \delta_1\phi, \delta_2\phi)$ over a Cauchy surface) is *not* gauge invariant, either in the sense of the Maxwell gauge transformations $\delta A_a \mapsto \delta A_a + \nabla_a \chi$ or the infinitesimal diffeomorphisms $\delta\phi \mapsto \delta\phi + \mathcal{L}_X \phi$, on account of boundary terms arising from the horizon.

For the purposes of analyzing gauge invariance, it is convenient to view the canonical energy as a bilinear form on the space of perturbations to a black hole background given by

$$\mathcal{E}_\Sigma(\phi; \delta_1\phi, \delta_2\phi) \equiv \int_\Sigma \boldsymbol{\omega}(\phi; \delta_1\phi, \mathcal{L}_\xi \delta_2\phi). \tag{52}$$

The canonical energy will be gauge invariant if and only if it vanishes whenever $\delta_1\phi$ or $\delta_2\phi$ is a pure gauge transformation.

If $\delta_1\phi$ and $\delta_2\phi$ are solutions, then, as shown in [21], $\mathcal{E}_\Sigma$ is symmetric. Namely, by the antisymmetry and bilinearity of the symplectic current, we have

$$\mathcal{E}_\Sigma(\phi; \delta_1\phi, \delta_2\phi) - \mathcal{E}_\Sigma(\phi; \delta_2\phi, \delta_1\phi) = \int_\Sigma \mathcal{L}_\xi \boldsymbol{\omega}(\phi; \delta_1\phi, \delta_2\phi). \tag{53}$$

Applying the Lie derivative identity $\mathcal{L}_\xi \boldsymbol{\omega} = \iota_\xi d\boldsymbol{\omega} + d(\iota_\xi \boldsymbol{\omega})$ and applying Stokes' theorem to the second term yields

$$\mathcal{E}_\Sigma(\phi; \delta_1\phi, \delta_2\phi) - \mathcal{E}_\Sigma(\phi; \delta_2\phi, \delta_1\phi) = \int_\Sigma \iota_\xi d\boldsymbol{\omega}(\phi; \delta_1\phi, \delta_2\phi) + \int_\infty \iota_\xi \boldsymbol{\omega}(\phi; \delta_1\phi, \delta_2\phi) - \int_B \iota_\xi \boldsymbol{\omega}(\phi; \delta_1\phi, \delta_2\phi). \tag{54}$$

The first term vanishes for solutions[6] by (13). The boundary term at infinity vanishes under the assumption that $\delta_1\phi$ and $\delta_2\phi$ are asymptotically flat perturbations with appropriate falloff conditions and the boundary term at the bifurcation surface vanishes since ξ^a vanishes on B, thus establishing that $\mathcal{E}_\Sigma$ is symmetric. This is convenient because it implies that to show gauge invariance of $\mathcal{E}_\Sigma$, we need only show that $\mathcal{E}_\Sigma$ vanishes when $\delta_2\phi$ is pure gauge in (52).

First let us consider a pure Maxwell gauge transformation given by $\delta g_{ab} = 0$, $\delta A_a = \nabla_a \chi$ for some smooth function χ. In analogy with (14), which defined the Noether current associated with a local diffeomorphism, we may define the Noether current associated with a Maxwell gauge transformation by

$$\mathcal{J}_\chi = \boldsymbol{\theta}(\phi, \nabla_a \chi). \tag{55}$$

Just as in (16), this Noether current can also be written in terms of a constraint and a charge as

$$\mathcal{J}_\chi = \boldsymbol{C}[\chi] + d\boldsymbol{Q}[\chi]. \tag{56}$$

A simple calculation shows that for the Einstein-Maxwell theory, the constraint and Noether charge are given by

$$(\mathcal{C}[\chi])_{abc} = \epsilon_{dabc} \chi j^d, \tag{57}$$

$$(\mathcal{Q}[\chi])_{ab} = -\frac{1}{8\pi} \epsilon_{cdab} \chi F^{cd}. \tag{58}$$

A calculation similar to that used to obtain (18) yields the identity

$$\int_{\partial\Sigma} \delta\boldsymbol{Q}[\chi] = \int_\Sigma \boldsymbol{\omega}(\phi; \delta\phi, \nabla_a \chi) - \int_\Sigma \delta\boldsymbol{C}, \tag{59}$$

i.e.,

$$W_\Sigma(\phi; \delta\phi, \nabla_a \chi) = \int_\infty \delta\boldsymbol{Q}[\chi] - \int_B \delta\boldsymbol{Q}[\chi] + \int_\Sigma \delta\boldsymbol{C}, \tag{60}$$

where $W_\Sigma(\phi; \delta_1\phi, \delta_2\phi) \equiv \int_\Sigma \boldsymbol{\omega}(\phi; \delta_1\phi, \delta_2\phi)$ is the symplectic form. The constraint term vanishes under the assumption that $\delta\phi$ satisfies the linearized equations of motion, so, using (58), we obtain,

$$W_\Sigma(\phi; \delta\phi, \nabla_a \chi) = -\frac{1}{8\pi} \int_\infty \chi \delta(\epsilon_{cdab} F^{cd}) + \frac{1}{8\pi} \int_B \chi \delta(\epsilon_{cdab} F^{cd}). \tag{61}$$

This expression is nonvanishing for generic perturbations and gauge transformations, since χ may be nonvanishing at infinity and at B. Thus, the symplectic form is not invariant under Maxwell gauge transformations. However, the gauge invariance of the canonical energy for Maxwell gauge transformation can be seen by replacing χ by $\mathcal{L}_\xi \chi = \xi^a \nabla_a \chi$ in (61). The resulting expression vanishes, since $\xi^a \nabla_a \chi$ goes to zero at infinity and vanishes at B. Thus, the Einstein-Maxwell canonical energy is indeed invariant under Maxwell gauge transformations, as we desired to show.

[6] The perturbations considered in Secs. III and IV do *not* satisfy the linearized equations of motion, since they have sources in the form of charged matter that is added to the black hole. However, the quantity $\int_\Sigma \iota_\xi d\boldsymbol{\omega}$ still vanishes for the particular surface Σ chosen in those sections (cf. Figs. 1 and 2), and so the gauge invariance established in this subsection still holds for that particular case.

We now analyze the gauge dependence of the canonical energy under smooth infinitesimal diffeomorphisms, $\delta\phi = \mathcal{L}_X\phi$, for which X^a is an asymptotic symmetry. The canonical energy of an infinitesimal diffeomorphism is given by

$$\begin{aligned}\mathcal{E}_\Sigma(\phi;\delta\phi,\mathcal{L}_X\phi) &= W_\Sigma(\phi;\delta\phi,\mathcal{L}_\xi\mathcal{L}_X\phi)\\ &= W_\Sigma(\phi;\delta\phi,\mathcal{L}_Y\phi),\end{aligned} \tag{62}$$

where $Y^a = [\xi, X]^a$ and we have used the fact that $\mathcal{L}_\xi\phi = 0$ at $\lambda = 0$. From (18) and (21), we have

$$\begin{aligned}\mathcal{E}_\Sigma(\phi;\delta\phi,\mathcal{L}_X\phi) &= W_\Sigma(\phi;\delta\phi,\mathcal{L}_Y\phi)\\ &= \delta H_Y - \int_B [\delta\mathbf{Q}_Y - \iota_Y\boldsymbol{\theta}(\phi,\delta\phi)],\end{aligned} \tag{63}$$

where we have used the assumptions that $\phi(\lambda = 0)$ and $\delta\phi$ satisfy the equations of motion and the linearized equations of motion, respectively.

It is easily seen that the right side of (63) cannot vanish unless some restrictions are placed on the allowed perturbations at the horizon and at infinity. These conditions are purely gauge conditions on the perturbations that do not restrict the physical perturbations we consider. First, following [21], we impose the gauge condition that the perturbed expansion of the horizon generators vanishes,

$$\delta\Theta|_{\mathcal{H}} = 0. \tag{64}$$

As shown in [21], this condition may always be imposed for nonextremal black holes. The infinitesimal diffeomorphisms X^a that preserve this condition are the ones that are tangent to the future horizon. This implies that $Y^a = \mathcal{L}_\xi X^a$ is normal to the horizon at B.

Second, we impose the condition

$$k^a\delta A_a|_{\mathcal{H}} = 0, \tag{65}$$

where k^a denotes an affinely parametrized tangent to the generators of the horizon. This condition always can be imposed by a Maxwell gauge transformation $\delta A_a \to \delta A'_a = \delta A_a - \nabla_a\chi$ with χ satisfying $k^a\nabla_a\chi = k^a\delta A_a$ on $\mathcal{H}$.

We now evaluate the terms appearing on the right side of (63), where $Y^a = \mathcal{L}_\xi X^a$. First, we evaluate the contribution to the boundary term at B arising from the symplectic potential. We split the symplectic potential into a gravitational and an electromagnetic part as in (31). As shown in [21], the gravitational part of the symplectic potential contribution yields

$$\int_B \iota_Y\boldsymbol{\theta}^{GR}(\phi,\delta\phi) = -\frac{1}{8\pi}\int_B f\delta\Theta\boldsymbol{\epsilon}, \tag{66}$$

where we have written $Y^a = fk^a$ on B with k^a normal to the horizon, since Y^a is normal to the horizon at B. This term vanishes as a consequence of our gauge condition (64).

As for the electromagnetic part of the symplectic potential, we have

$$\int_B \iota_X\boldsymbol{\theta}^{\mathrm{EM}}(\phi,\delta\phi) = -\frac{1}{4\pi}\int_B \epsilon_{dcab}Y^cF^{de}\delta A_e. \tag{67}$$

However, the assumption that the background spacetime is stationary restricts the form of F^{de}, since the flux of electromagnetic stress-energy

$$T^{\mathrm{EM}}_{ab} = \frac{1}{4\pi}\left[F_{ac}F_b{}^c - \frac{1}{4}g_{ab}F^{cd}F_{cd}\right] \tag{68}$$

through the horizon must vanish. For this flux to vanish, we must have $T^{\mathrm{EM}}_{ab}k^ak^b = 0$ on the horizon. The dominant energy condition (which is automatically satisfied by the electromagnetic field) then implies that $T^{\mathrm{EM}}_{ab}k^a$ must be proportional to k_b. This implies that on $\mathcal{H}$, F^{ab} must take the form

$$F^{ab} = v^{[a}k^{b]} + w^{ab}, \tag{69}$$

where w^{ab} is purely tangential to the horizon. From this, and from the assumption that X^a is tangent to the horizon generators on B, we find that the electromagnetic part of the symplectic potential can be written as

$$\int_B \iota_Y\boldsymbol{\theta}^{\mathrm{EM}}(\phi,\delta\phi) = -\frac{1}{8\pi}\int_B \epsilon_{dcab}Y^cv^dk^e\delta A_e, \tag{70}$$

where we have used the fact that the pullback to $\mathcal{H}$ of ϵ_{abcd} contracted into any vector tangent to $\mathcal{H}$ vanishes. The right side of (70) vanishes on account of our gauge condition (65).

Next, we consider the term δH_Y in (63). Since X^a is an asymptotic symmetry and $\xi^a = t^a + \Omega_H\varphi^a$ for a Kerr-Newman background, Y^a is a linear combination of an asymptotic space translation and an asymptotic rotation or boost in a direction orthogonal to the black hole's axis of rotation. So long as we restrict ourselves to perturbations with vanishing ADM linear momenta, $\delta P_i = 0$, and vanishing ADM angular momentum and center of mass in directions orthogonal to the axis of rotation, we have $\delta H_Y = 0$ for all suitable choices of infinitesimal diffeomorphism X^a. These conditions do not restrict the physical perturbation.

We are left with

$$\mathcal{E}_\Sigma(\phi;\delta\phi,\mathcal{L}_X\phi) = -\int_B \delta\mathbf{Q}_Y. \tag{71}$$

We split $\mathbf{Q}_Y$ into gravitational and electromagnetic parts as in (34). It was shown in the Appendix of [21] that since

Y^a is normal to the horizon, the pullback to B of $\delta \boldsymbol{Q}_Y^{GR}$ is given by

$$\delta \boldsymbol{Q}_Y^{GR} = -\frac{1}{16\pi}(\delta\epsilon_{abcd})\nabla^c Y^d. \tag{72}$$

The right side will be nonvanishing if and only if the quantity

$$U \equiv n_{cd}\nabla^c Y^d \tag{73}$$

is nonvanishing on B in the background spacetime, where $n_{ab} = n_{[ab]}$ is the binormal to B. We substitute $Y^a = \mathcal{L}_\xi X^a = \xi^b \nabla_b X^a - X^b \nabla_b \xi^a$ in this equation and expand using the Leibniz rule to get

$$\begin{aligned} U = n_{cd}[&\xi^b \nabla^c \nabla_b X^d + (\nabla^c \xi^b)\nabla_b X^d \\ &- X^b \nabla^c \nabla_b \xi^d - (\nabla^c X^b)\nabla_b \xi^d]. \end{aligned} \tag{74}$$

The first term vanishes since ξ^a vanishes on B. Since ξ^a is a Killing field, we have $\nabla_a \nabla_b \xi^c = R^c{}_{bad}\xi^d = 0$ on B, so the third term also vanishes on B. Finally, using the fact that $\nabla_a \xi_b \propto n_{ab}$ on B, the second and fourth terms can be seen to cancel. Thus, $U = 0$ on B and the contribution from $\delta \boldsymbol{Q}_Y^{GR}$ vanishes.

Remark.—In [21], the vanishing of the contribution from $\delta \boldsymbol{Q}_Y^{GR}$ was obtained by imposing the gauge condition $\delta\epsilon_{ab} = (\delta A/A)\epsilon_{ab}$ on the area element on B together with the restriction $\delta A = 0$ on the perturbation. The above calculation shows that it was not necessary to impose either this gauge condition or this restriction. In particular, the hypothesis that $\delta A = 0$ may be dropped from Proposition 3 of [21].

Finally, we evaluate the contribution from $\delta \boldsymbol{Q}_Y^{\mathrm{EM}}$. We obtain

$$\int_B \delta \boldsymbol{Q}_Y^{\mathrm{EM}} = -\int_B \frac{1}{8\pi}\delta(\epsilon_{abcd}F^{cd})A_e Y^e. \tag{75}$$

However, a diffeomorphism X^a will preserve our gauge condition (65) only if $\xi^a \mathcal{L}_X A_a = 0$ on the horizon,[7] which implies that $A_a Y^a$ vanishes at B. Thus, the contribution from $\delta \boldsymbol{Q}_Y^{\mathrm{EM}}$ also vanishes.

We summarize the results of this subsection in the following proposition:

Proposition 1.—Consider the subspace of perturbations, $\delta\phi$, that (i) satisfy the linearized equations of motion, $\delta E(\phi) = 0$, (ii) satisfy the gauge conditions (64) and (65) at the horizon, and (iii) have vanishing ADM linear momenta, $\delta P_i = 0$, and vanishing ADM angular momentum and center of mass in directions orthogonal to the axis of rotation of the unperturbed black hole. Then the Einstein-Maxwell canonical energy $\mathcal{E}_\Sigma(\phi; \delta_1\phi, \delta_2\phi)$ on this subspace is invariant under all infinitesimal diffeomorphisms $\delta\phi = \mathcal{L}_X \phi$ and Maxwell gauge transformations $\delta A_a = \nabla_a \chi$ (where it is understood that these transformations must preserve conditions (ii) and (iii)).

III. GEDANKEN EXPERIMENTS TO DESTROY AN EXTREMAL BLACK HOLE

Consider an extremal Kerr-Newman black hole,

$$M^2 = (J/M)^2 + Q^2. \tag{76}$$

We wish to see if we can cause the inequality (1) to be violated by throwing/dropping charged and/or rotating matter into the black hole. Specifically, (1) will be violated—and a contradiction with cosmic censorship obtained—if we can perturb the black hole so that

$$2M\delta M < 2(J/M)(M\delta J - J\delta M)/M^2 + 2Q\delta Q. \tag{77}$$

Writing $a = J/M$, we see that a violation will occur if the perturbation satisfies

$$\delta M < \frac{a}{M^2 + a^2}\delta J + \frac{QM}{M^2 + a^2}\delta Q. \tag{78}$$

To analyze whether it is possible to produce such a perturbation, let Σ_0 be an asymptotically flat hypersurface that terminates on the future horizon and extends to spatial infinity. We consider a perturbation $\delta\phi$ whose initial data on Σ_0 for the fields δg_{ab} and δA_a vanishes in a neighborhood of $\Sigma_0 \cap \mathcal{H}$, as shown in Fig. 1. We assume that the matter sources δT_{ab} and δj^a are nonvanishing only in a compact

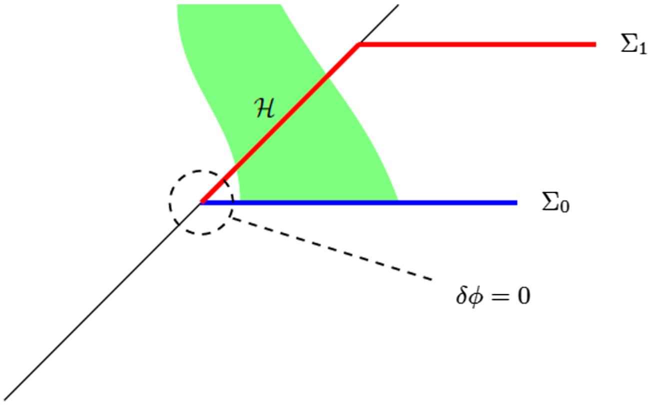

FIG. 1. Charged matter, occupying the shaded region, falls through the event horizon of an extremal black hole. The perturbed initial data on Σ_0 vanishes in a neighborhood of the horizon.

[7] Rather than restricting X^a so as to preserve the gauge condition (65), it would be more sensible to require that any X^a that violates (65) be accompanied by a Maxwell gauge transformation that restores (65). One would then get a nonvanishing contribution from (75) that would then be canceled by the contribution from the Maxwell gauge transformation.

region of Σ_0, as shown. Physically, this corresponds to considering a perturbation that is induced by bringing matter in from infinity in such a way that the disturbance to the black hole at very early advanced times is negligibly small. If we evolve the perturbation, in general, some of the matter will go into the black hole and some will go out to infinity or remain in orbit around the black hole. The matter that does not fall into the black hole is of no interest to us. Therefore, we can greatly simplify our analysis by restricting to the case where all of the matter goes into the black hole. Note that this also saves us the trouble of analyzing the motion of bodies outside of the black hole; we do not care about the details of how the matter managed to get into the black hole as long as it does get in.

Thus, we wish to consider a one-parameter family where δT_{ab} and δj^a are nonvanishing only in a region like the shaded region of Fig. 1. Let Σ be a hypersurface like that shown in Fig. 1 with the following characteristics: (a) It starts on the future event horizon in a region where the perturbation vanishes. (b) It continues up the future horizon until past the region where the matter sources are nonvanishing. (c) It then becomes spacelike and continues out towards infinity in an asymptotically flat manner. Let $\mathcal{H}$ denote the horizon portion of Σ, and let Σ_1 denote the spacelike portion (see Fig. 1) so that

$$\Sigma = \mathcal{H} \cup \Sigma_1. \tag{79}$$

We now use (49) (with $\delta A_B = \delta Q_B = 0$) for this choice of Σ. The integrand on the right side of (49) is nonvanishing only on $\mathcal{H}$. Thus, we obtain,

$$\delta M - \Omega_H \delta J = -\int_{\mathcal{H}} \epsilon_{ebcd} \xi_a \delta T^{ae} - \int_{\mathcal{H}} \xi_a A^a \delta(\epsilon_{ebcd} j^e). \tag{80}$$

Since $\Phi_H = -\xi^a A_a$ is constant on $\mathcal{H}$, we may pull it out of the integral. The integral $\int_{\mathcal{H}} \delta(\epsilon_{ebcd} j^e)$ is just the total flux of electromagnetic charge through the horizon, δQ_{flux}. Since all of the charge added to the spacetime falls through the horizon, this flux is just equal to the total perturbed charge of the black hole, $\delta Q_{\text{flux}} = \delta Q$. Combining these observations yields the following formula relating the perturbed parameters of the black hole spacetime:

$$\delta M - \Omega_H \delta J - \Phi_H \delta Q = -\int_{\mathcal{H}} \epsilon_{ebcd} \xi_a \delta T^{ae}. \tag{81}$$

This result was first derived in [16]. On the horizon, we may write

$$\epsilon_{ebcd} = -4 k_{[e} \tilde{\epsilon}_{bcd]}, \tag{82}$$

where k^a is the future-directed normal to the horizon and $\tilde{\epsilon}_{bcd}$ is the corresponding volume element on the horizon. The right side of (81) can be written as

$$-\int_{\mathcal{H}} \epsilon_{ebcd} \xi_a \delta T^{ae} = \int_{\mathcal{H}} \tilde{\epsilon}_{bcd} \delta T^{ae} \xi_a k_e. \tag{83}$$

Since $\xi^a \propto k^a$, the right side is non-negative provided only that the nonelectromagnetic stress energy tensor δT_{ab} satisfies the null energy condition, so that $\delta T_{ab} k^a k^b \geq 0$. Thus, (81) yields the inequality

$$\delta M - \Omega_H \delta J - \Phi_H \delta Q \geq 0, \tag{84}$$

which holds for all perturbations of an extremal Kerr-Newman black hole resulting from charged-matter entering the black hole.

For a general (not necessarily extremal) Kerr-Newman black hole, we have

$$\Omega_H = \frac{a}{r_+^2 + a^2} \tag{85}$$

and

$$\Phi_H = \frac{Q r_+}{r_+^2 + a^2}, \tag{86}$$

where r_+ is the horizon radius

$$r_+ = M + \sqrt{M^2 - (J/M)^2 - Q^2}. \tag{87}$$

For an extremal black hole, we have $r_+ = M$, so (84) yields

$$\delta M \geq \frac{a}{M^2 + a^2} \delta J + \frac{QM}{M^2 + a^2} \delta Q. \tag{88}$$

Thus, (78) cannot be satisfied, and an extremal black hole cannot be destroyed by dropping/throwing matter into it. This generalizes the results of paper I [3] to arbitrary matter, provided only that the nonelectromagnetic contribution to the stress-energy tensor satisfies the null energy condition. This argument that (81) implies that one cannot overcharge or overspin an extremal black hole was previously given in [18].

IV. GEDANKEN EXPERIMENTS TO DESTROY A SLIGHTLY NON-EXTREMAL BLACK HOLE

In the spirit of Hubeny [4], let us now repeat the gedanken experiment of the previous section starting with a slightly nonextremal Kerr-Newman black hole. The relevant spacetime diagram for this case is shown in Fig. 2, where the only significant difference is that Σ_0 and Σ are now taken to terminate at the bifurcation surface, B. This does not affect the analysis of the first order perturbation given in the previous section, since the perturbation is assumed to vanish on the horizon at sufficiently early advanced times. Since we will need to calculate second order effects in this section, we further assume that the

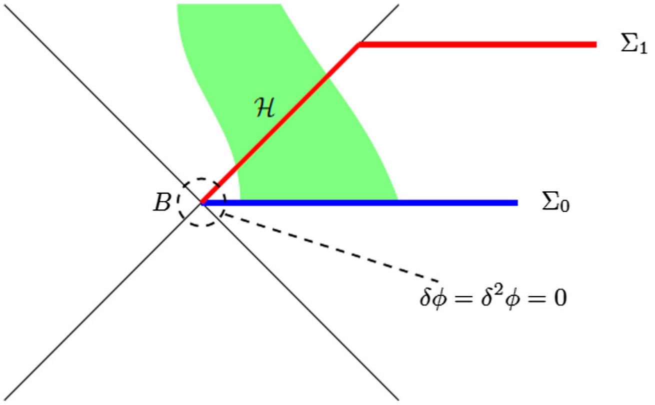

FIG. 2. A spacetime diagram showing charged matter falling into a black hole as in Fig. 1, but now shown for a nonextremal black hole. The surface Σ_0 is taken to pass through the bifurcation surface.

second order perturbation also vanishes in a neighborhood of B, and that all of the matter sources go into the black hole at second order, so that $\delta^2 T_{ab} = \delta^2 j^a = 0$ on Σ_1.

An exact repetition of the analysis of the previous section yields

$$\begin{aligned}\delta M &= \Omega_H \delta J + \Phi_H \delta Q - \int_{\mathcal{H}} \epsilon_{ebcd} \xi_a \delta T^{ae} \\ &\geq \Omega_H \delta J + \Phi_H \delta Q \\ &= \frac{a}{r_+^2 + a^2} \delta J + \frac{Q r_+}{r_+^2 + a^2} \delta Q. \end{aligned} \tag{89}$$

As already noted in the Introduction for the special case of a nearly extremal Reissner-Nordstrom black hole, this equation admits the possibility of violating (1). However, as discussed in the Introduction, in order to determine whether violations of (1) really occur, it is necessary to calculate the second order corrections, $\delta^2 M$, to the mass of the black hole.

In order to proceed further with our analysis of the second order corrections to mass, we will make the following additional assumption:

> *Additional Assumption*: The (slightly) nonextremal, unperturbed Kerr-Newman black hole we are considering is linearly stable to perturbations, i.e., any source-free[8] solution to the linearized Einstein-Maxwell equations approaches a perturbation towards another Kerr-Newman black hole at sufficiently late times.

It should be emphasized that this linear stability assumption is entirely compatible with having an instability associated with overcharging or overspinning the black hole, i.e., we are not assuming what we wish to show. Since we are considering a nonextremal black hole (i.e., $M^2 > (J/M)^2 + Q^2$), a *finite* perturbation is required to overcharge or overspin it. A linear perturbation of a nonextremal black hole always can be scaled down so as to not violate (1). Thus, the presence of a linear instability of a nonextremal black hole would represent an instability that is independent of overcharging or overspinning. If a nonextremal black hole were linearly unstable, there would be no need to attempt to overcharge or overspin it in order to destroy it.

In view of this assumption, we may choose Σ in Fig. 2 so the horizon portion, $\mathcal{H}$, extends to sufficiently late times that it enters the late time stationary era of the perturbation. We may then take Σ_1 so that it extends far[9] from the black hole while remaining in the stationary region. The quantities $\delta^2 M$ and $\delta^2 J$ arising in the boundary term (91) on Σ will then have the interpretation of being the second order corrections to the mass and angular momentum of the perturbed black hole.[10]

We now consider our gedanken experiment to destroy the slightly non-extremal black hole. We assume that our first order perturbation has been done optimally [see (89)], so that

$$\delta M = \Omega_H \delta J + \Phi_H \delta Q = \frac{a}{r_+^2 + a^2} \delta J + \frac{Q r_+}{r_+^2 + a^2} \delta Q. \tag{90}$$

As can be seen from (89), this requires vanishing non-electromagnetic energy flux through the horizon, i.e., $\delta T_{ab} k^a k^b = 0$, as should be (nearly) achievable if the matter is lowered (nearly) to the horizon or is (nearly) at a turning point of its orbit just before entering the black hole.

The second order change in mass is given by (51) with $\delta^2 Q_B = \delta^2 A_B = 0$ (since the second order perturbation has been assumed to vanish in a neighborhood of B). We have

$$\delta^2 M - \Omega_H \delta^2 J = \mathcal{E}_\Sigma(\phi; \delta\phi) - \int_{\mathcal{H}} \iota_\xi (\delta \mathbf{E}(\phi) \cdot \delta\phi) - \int_{\mathcal{H}} \delta^2 \mathbf{C}_\xi. \tag{91}$$

Here, the integrals in the last two terms extend only over $\mathcal{H}$ rather than over all of $\Sigma = \mathcal{H} \cup \Sigma_1$ because $\delta \mathbf{E}$ and $\delta^2 \mathbf{C}_\xi$ vanish on Σ_1 by the assumption that there are no sources outside the black hole at late times.

We now evaluate the last two terms appearing on the right side of (91). From (37), we have

[8] Our perturbations are, in general, not source-free. However, we will only need to apply this assumption on the late-time surface Σ_1 sketched in Fig. 2, long after all sources have fallen into the black hole.

[9] If we wish to take Σ_1 to extend infinitely far from the black hole, we would have to take it to null infinity rather than spatial infinity.

[10] Note that since mass and angular momentum cannot be radiated away at linear order, we did not need to be careful in our specification of Σ_1 in our first order analysis in order for δM and δJ to represent the perturbed mass and angular momentum of the final black hole.

$$(\iota_\xi(\delta\mathbf{E}(\phi)\cdot\delta\phi))_{abc} = -\xi^d\epsilon_{dabc}\left[\frac{1}{2}\delta T^{ef}\delta g_{ef} + \delta j^e\delta A_e\right]. \tag{92}$$

Since ξ^a is tangent to the horizon, the pullback to $\mathcal{H}$ of this term vanishes, so it does not contribute to (91). From (38), we have

$$(\delta^2\mathbf{C}_\xi)_{abc} = \delta^2(\epsilon_{eabc}T_d{}^e\xi^d) + \delta^2(\epsilon_{eabc}A_d j^e\xi^d). \tag{93}$$

Using our gauge condition $\xi^a\delta A_a = 0$ on $\mathcal{H}$ [see (65) and the discussion of subsection II D], we see that on $\mathcal{H}$, the second term is

$$\delta^2(\epsilon_{eabc}A_d j^e\xi^d) = -\Phi_H\delta^2(\epsilon_{eabc}j^e), \tag{94}$$

and therefore

$$\delta^2\left[\int_{\mathcal{H}}\xi_a A^a\epsilon_{ebcd}j^e\right] = -\Phi_H\delta^2 Q_{\text{flux}} = -\Phi_H\delta^2 Q, \tag{95}$$

where $\delta^2 Q$ is the second-order change in charge of the black hole. On the other hand, using our assumption that the first order process was done optimally and thus there was vanishing nonelectromagnetic stress-energy flux through the horizon at first order, we have

$$\delta^2(\epsilon_{eabc}T_d{}^e\xi^d) = \epsilon_{eabc}\xi^d\delta^2 T_d{}^e. \tag{96}$$

Putting this together, we obtain

$$\delta^2 M - \Omega_H\delta^2 J - \Phi_H\delta^2 Q = \mathcal{E}_\Sigma(\phi;\delta\phi) - \int_{\mathcal{H}}\xi^a\epsilon_{ebcd}\delta^2 T_a{}^e. \tag{97}$$

The last term in this equation is positive provided that the nonelectromagnetic stress-energy tensor satisfies the null energy condition.

It remains to compute the canonical energy $\mathcal{E}_\Sigma(\phi;\delta\phi)$. Since $\Sigma = \mathcal{H}\cup\Sigma_1$, we have

$$\mathcal{E}_\Sigma(\phi;\delta\phi) = \int_{\mathcal{H}}\boldsymbol{\omega}(\phi,\delta\phi,\mathcal{L}_\xi\delta\phi) + \int_{\Sigma_1}\boldsymbol{\omega}(\phi,\delta\phi,\mathcal{L}_\xi\delta\phi). \tag{98}$$

Let us calculate first calculate the horizon contribution. We have

$$\int_{\mathcal{H}}\boldsymbol{\omega} = \int_{\mathcal{H}}\boldsymbol{\omega}^{GR} + \int_{\mathcal{H}}\boldsymbol{\omega}^{\text{EM}}, \tag{99}$$

where the gravitational and electromagnetic parts, $\boldsymbol{\omega}^{GR}$ and $\boldsymbol{\omega}^{\text{EM}}$, are given, respectively, by (40) and (41) above. The integral over $\mathcal{H}$ of the gravitational part of the canonical energy density was computed in [21], and is given by[11]

$$\int_{\mathcal{H}}\boldsymbol{\omega}^{GR}(g;\delta g,\mathcal{L}_\xi\delta g) = \frac{1}{4\pi}\int_{\mathcal{H}}(\xi^a\nabla_a u)\delta\sigma_{bc}\delta\sigma^{bc}\boldsymbol{\epsilon} + \frac{1}{16\pi}\int_S(\xi^a\nabla_a u)\delta g^{bc}\delta\sigma_{bc}\boldsymbol{\epsilon} \tag{100}$$

where $\delta\sigma_{ab}$ denotes the perturbed shear of the horizon generators, u is an affine parameter along the future horizon, and $S = \mathcal{H}\cap\Sigma_1$ is the 2-surface formed by the intersection of $\mathcal{H}$ and Σ_1. By our additional assumption above, the perturbation is physically stationary at S, so $\delta\sigma_{ab} = 0$ on S. Thus, we obtain

$$\int_{\mathcal{H}}\boldsymbol{\omega}^{GR}(\phi;\delta\phi,\mathcal{L}_\xi\delta\phi) = \frac{1}{4\pi}\int_{\mathcal{H}}(\xi^a\nabla_a u)\delta\sigma_{bc}\delta\sigma^{bc}\boldsymbol{\epsilon} \geq 0. \tag{101}$$

We may interpret this horizon flux contribution from $\boldsymbol{\omega}^{GR}$ as representing the total flux of gravitational wave energy into the black hole.

Next, we calculate the horizon flux contribution from $\boldsymbol{\omega}^{\text{EM}}$. From (41), we have

$$(\omega^{\text{EM}})_{abc}(\phi;\delta\phi,\mathcal{L}_\xi\phi) = \frac{1}{4\pi}\epsilon_{dabc}[\delta A_e\mathcal{L}_\xi\delta F^{de} - \delta F^{de}\mathcal{L}_\xi\delta A_e] + \frac{1}{4\pi}[(\mathcal{L}_\xi\delta\epsilon_{dabc})F^{de}\delta A_e - (\delta\epsilon_{dabc})F^{de}\mathcal{L}_\xi\delta A_e]. \tag{102}$$

The last two terms on the right side of this equation involve the background electromagnetic field strength F^{ab}. However, by (69) together with our gauge condition $\xi^a\delta A_a = 0$ on $\mathcal{H}$, it can be seen that the last two terms in (102) vanish. The first term in (102) can be written as

$$\epsilon_{dabc}\delta A_e\mathcal{L}_\xi\delta F^{de} = \mathcal{L}_\xi[\epsilon_{dabc}\delta A_e\delta F^{de}] - \epsilon_{dabc}\delta F^{de}\mathcal{L}_\xi\delta A_e. \tag{103}$$

When pulled back to $\mathcal{H}$, $\epsilon_{dabc}\delta A_e\delta F^{de}$ is a 3-form $\boldsymbol{\eta}$, on a 3-dimensional surface, so when pulled back to $\mathcal{H}$, we have

$$\mathcal{L}_\xi\boldsymbol{\eta} = \iota_\xi d\boldsymbol{\eta} + d(\iota_\xi\boldsymbol{\eta}) = d(\iota_\xi\boldsymbol{\eta}), \tag{104}$$

where the pullback of $\iota_\xi d\boldsymbol{\eta}$ vanishes since ξ^a is tangent to $\mathcal{H}$. Thus, the integral over $\mathcal{H}$ of the first term on the right side of (103) will merely contribute a boundary term at $S = \mathcal{H}\cap\Sigma_1$. However, since the perturbation is assumed to be stationary at S, the electromagnetic energy flux must

[11] Eq. (100) assumes that $\delta\Theta = 0$ on $\mathcal{H}$ (see [21]). This condition can be imposed in the present case because we assumed that the first order process was done optimally [see (90)], so $\delta T_{ab}k^a k^b = 0$.

vanish there, so δF_{ab} must be of the form (69). Using this fact together with our gauge condition $\xi^a \delta A_a = 0$ on $\mathcal{H}$, it can be seen that this boundary term vanishes. Finally, the second term on the right side of (103) combines with the second term of (102). This term can be further simplified by noting that

$$\mathcal{L}_\xi \delta \boldsymbol{A} = \iota_\xi d\delta \boldsymbol{A} + d(\iota_\xi \delta \boldsymbol{A}). \tag{105}$$

Under our gauge condition $\xi^a \delta A_a|_{\mathcal{H}} = 0$, the second term of (105) is normal to the horizon, and hence proportional to the horizon normal k^a. By the antisymmetry of δF_{ab}, $\delta F^{ab} k_b$ is orthogonal to k^a and hence tangent to the horizon. As this term only appears in (102) when contracted into the volume element on the horizon, it makes no contribution to the canonical energy integral. Putting everything together, we find that

$$\int_{\mathcal{H}} \boldsymbol{\omega}^{\mathrm{EM}}(\phi; \delta\phi, \mathcal{L}_\xi \delta\phi) = -\frac{1}{2\pi} \int_{\mathcal{H}} \epsilon_{dabc} \xi^e \delta F^{df} \delta F_{ef}. \tag{106}$$

The right side of this equation is nonnegative and can be interpreted as the total flux of electromagnetic energy into the black hole.

All that remains now is to calculate the contribution to canonical energy from Σ_1

$$\mathcal{E}_{\Sigma_1}(\phi; \delta\phi) = \int_{\Sigma_1} \boldsymbol{\omega}(\phi, \delta\phi, \mathcal{L}_\xi \delta\phi). \tag{107}$$

Since we have assumed that the perturbation is stationary on Σ_1, it might be thought that $\mathcal{L}_\xi \delta\phi = 0$ on Σ_1 and thus this contribution to the canonical energy vanishes. However, this is not the case because our conditions $\delta\xi^a = 0$ as well as our gauge condition $\xi^a \delta A_a = 0$ on $\mathcal{H}$ preclude our writing the perturbation in a gauge where $\mathcal{L}_\xi \delta g_{ab} = 0$ and $\mathcal{L}_\xi \delta A_a = 0$; see [21] for further discussion. Nevertheless, we can calculate $\mathcal{E}_{\Sigma_1}(\phi; \delta\phi)$ as follows. First, since, by assumption, $\delta\phi$ is equal to a perturbation $\delta\phi^{KN}$ to another Kerr-Newman black hole on Σ_1, we obviously may replace $\delta\phi$ by $\delta\phi^{KN}$ (written in our gauge) on the right side of (107)

$$\mathcal{E}_{\Sigma_1}(\phi; \delta\phi) = \mathcal{E}_{\Sigma_1}(\phi; \delta\phi^{KN}) = \int_{\Sigma_1} \boldsymbol{\omega}(\phi, \delta\phi^{KN}, \mathcal{L}_\xi \delta\phi^{KN}). \tag{108}$$

However, as can be seen from our analysis above, $\delta\phi^{KN}$ has no flux of canonical energy through $\mathcal{H}$, i.e., there is no flux of gravitational or electromagnetic energy through the horizon for a Kerr-Newman perturbation. Thus, we may replace Σ_1 by Σ in (108). Finally, we may evaluate $\mathcal{E}_\Sigma(\phi; \delta\phi^{KN})$ using (51). Consider the one-parameter family, $\phi^{KN}(\alpha)$, where each field configuration in the family is a Kerr-Newman black hole with parameters given by

$$M^{KN}(\alpha) = M + \alpha \delta M, \tag{109}$$

$$Q^{KN}(\alpha) = Q + \alpha \delta Q, \tag{110}$$

$$J^{KN}(\alpha) = J + \alpha \delta J, \tag{111}$$

where δM, δQ, and δJ are chosen to agree with the corresponding values for our first-order perturbation $\phi(\lambda)$. Then, for this family, we have $\delta^2 M = \delta^2 J = \delta^2 Q_B = 0$, as well as $\delta \boldsymbol{E} = \delta^2 \boldsymbol{C}_\xi = 0$. Thus, we obtain

$$\mathcal{E}_\Sigma(\phi; \delta\phi^{KN}) = -\frac{\kappa}{8\pi} \delta^2 A_B^{KN}, \tag{112}$$

where $\delta^2 A_B^{KN}$ denotes the second order change in the area of the horizon for the one-parameter family (109)–(111). This quantity can be evaluated by taking two variations of the area formula $A_B = 4\pi(r_+^2 + (J/M)^2)$, and is given explicitly as follows:

$$\begin{aligned}\delta^2 A_B^{KN} = &-\frac{8\pi}{M^8 \epsilon^3}[(\delta M)^2(J^4 + (2+\epsilon^2)J^2M^4 - M^8(1+\epsilon)(-1+\epsilon+2\epsilon^2)) \\ &+ (\delta Q)^2(M^6Q^2 + M^8(1+\epsilon)\epsilon^2) + (\delta J)^2(J^2M^2 + M^6\epsilon^2) \\ &+ \delta M \delta J(-2J^3M - 2JM^5(1+\epsilon^2)) + \delta J \delta Q(2JM^4Q) \\ &+ \delta M \delta Q(-2J^2M^3Q + 2M^7Q(-1+\epsilon^2))].\end{aligned} \tag{113}$$

Here we have introduced the parameter

$$\epsilon = r_+/M - 1 = \frac{\sqrt{M^2 - Q^2 - (J/M)^2}}{M} \tag{114}$$

[thereby generalizing (5) to the case where the black hole is rotating as well as charged] in order that we can keep better track of the extremal limit, $\epsilon \to 0$. However, we have not assumed that ϵ is small in (113).

We have now computed all of the terms appearing in (91). Using the positivity of the gravitational, electromagnetic, and nonelectromagnetic stress-energy fluxes through the horizon, we have thereby derived the following inequality involving the second order change of the mass of the black hole

$$\delta^2 M - \Omega_H \delta^2 J - \Phi_H \delta^2 Q \geq -\frac{\kappa}{8\pi}\delta^2 A_B^{KN}. \tag{115}$$

The surface gravity of a Kerr-Newman black hole is given by

$$\kappa = \frac{M^3}{M^4(1+\epsilon)^2 + J^2}\epsilon. \tag{116}$$

Expanding the right side of (115) to lowest order in ϵ, we obtain

$$\delta^2 M - \Omega_H \delta^2 J - \Phi_H \delta^2 Q \geq \frac{M}{(M^4+J^2)^2}[M^4(\delta J)^2 + (M^6 + J^2Q^2 + M^2J^2)(\delta Q)^2 - 2JM^2Q\delta J\delta Q] + O(\epsilon), \tag{117}$$

where we have used $\delta M = \Omega_H \delta J + \Phi_H \delta Q$ [see (90)] to eliminate δM from the expression.

We now show that this inequality is precisely what is needed to show that gedanken experiments of the Hubeny type can never succeed in overcharging or overspinning the black hole. Consider a one-parameter family, $\phi(\lambda)$, of the type we have been considering, where $\phi(0)$ is a nearly extremal Kerr-Newman black hole, $\epsilon \ll 1$. Define

$$f(\lambda) = M(\lambda)^2 - Q(\lambda)^2 - J(\lambda)^2/M(\lambda)^2. \tag{118}$$

Then, to second order in λ, we have

$$\begin{aligned} f(\lambda) = {}& \left(M^2 - Q^2 - \frac{J^2}{M^2}\right) + 2\lambda\left(\frac{M^4+J^2}{M^3}\delta M - \frac{J}{M^2}\delta J - Q\delta Q\right) \\ & + \lambda^2\left[\left(\frac{J^2+M^4}{M^3}\right)\delta^2 M - \frac{J}{M^2}\delta^2 J - Q\delta^2 Q + \frac{4J}{M^3}\delta J\delta M \right. \\ & \left. - \frac{1}{M^2}(\delta J)^2 + \left(\frac{M^4-3J^2}{M^4}\right)(\delta M)^2 - (\delta Q)^2\right]. \end{aligned} \tag{119}$$

We wish to know if, for small, λ, we can make $f < 0$. If we took into account only effects linear in λ, the inequality (89) would constrain f by

$$f(\lambda) \geq M^2\epsilon^2 + \frac{2}{M^4+J^2}((J^2 - M^4)Q\delta Q - 2JM^2\delta J)\lambda\epsilon + O(\lambda^2, \epsilon^3, \epsilon^2\lambda). \tag{120}$$

If the $O(\lambda^2)$ term and the higher order terms are neglected, then it is easy to see that it is possible to make $f(\lambda) < 0$, suggesting that the black hole could be over-charged or over-spun. However, when our calculation of the $O(\lambda^2)$ term given by inequality (117) is taken into account, we have shown that for an optimal first-order process with $\delta M = \Omega_H \delta J + \Phi_H \delta Q$, we have

$$\begin{aligned} f(\lambda) \geq {}& M^2\epsilon^2 + \frac{2}{M^4+J^2}((J^2 - M^4)Q\delta Q - 2JM^2\delta J)\lambda\epsilon \\ & + \frac{1}{M^2(M^4+J^2)^2}((J^2 - M^4)Q\delta Q - 2JM^2\delta J)^2\lambda^2 \\ & + O(\lambda^3, \epsilon^3, \epsilon^2\lambda, \epsilon\lambda^2). \end{aligned} \tag{121}$$

This expression can be rewritten as a perfect square,

$$\begin{aligned} f(\lambda) \geq {}& \left(\frac{(J^2 - M^4)Q\delta Q - 2JM^2\delta J}{M(M^4+J^2)}\lambda + M\epsilon\right)^2 \\ & + O(\lambda^3, \ldots). \end{aligned} \tag{122}$$

Thus, $f \geq 0$, and no violations of (1) can occur.

V. DISCUSSION

The Kerr-Newman parameter space $(M, Q, a = J/M)$ is shown in Fig. 3. In this parameter space, black holes lie within the "future light cone" $M > 0$, $M^2 - Q^2 - a^2 \geq 0$. Kerr-Newman solutions outside this cone correspond to

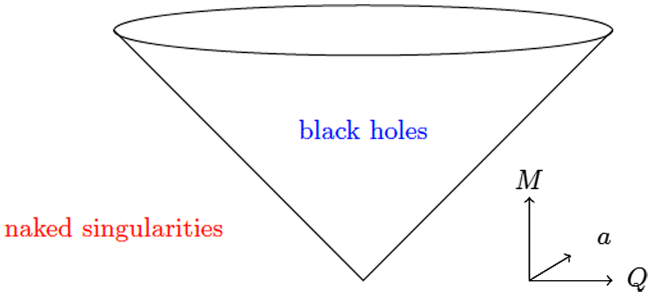

FIG. 3. The parameter space of Kerr-Newman black holes.

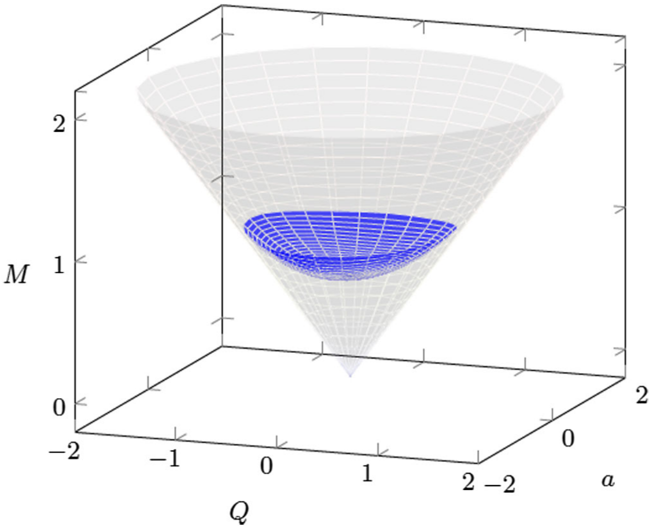

FIG. 4. A surface of constant area for Kerr-Newman black holes.

naked singularities. Extremal black holes live on the boundary of the cone, $M = \sqrt{Q^2 + a^2}$. The gedanken experiments to destroy an extremal black hole discussed in Sec. III correspond to analyzing whether, starting at the boundary, one can perturb the spacetime so as to move outside the cone. The gedanken experiments to destroy a slightly non-extremal black hole discussed in Sec. IV correspond to analyzing whether one can move out of the cone starting near (but not on) the boundary of the cone.

Within this cone, one can draw surfaces of constant area for the Kerr-Newman black holes. One such surface is shown in Fig. 4. It is important to note that the surfaces of constant area meet the boundary tangentially.

To linear order, the change in the parameters (M, Q, a) resulting from dropping matter into a Kerr-Newman black hole corresponds to a tangent vector in parameter space. Equation (89) shows precisely that for an arbitrary Kerr-Newman black hole, to linear order, any perturbation resulting from matter entering a black hole cannot decrease the area of the black hole[12] Thus, the tangent to the surface of constant area provides a lower bound to the slope of any tangent vector representing a physically achievable perturbation. In particular, for an extremal black hole, the best one can do is move tangentially to the cone. Thus, as we found in Sec. III, to first order it is impossible to escape from the cone into the naked singularity region of parameter space starting at the boundary of the cone.

The Hubeny argument for possibly escaping from the cone is illustrated in Fig. 5. For simplicity in the drawing, we have set $J = 0$ and thus show only the parameter space of Reissner-Nordstrom solutions. As is illustrated in this figure, except at the boundary, the tangent to the curve of

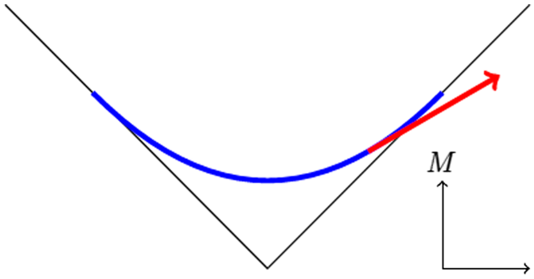

FIG. 5. The tangent to a curve of constant area for a slightly non-extremal Reissner-Nordstrom black hole.

constant area has a slope strictly less than one. Thus, a straight line tangent to such a curve will exit the cone. This means that if the linear approximation were valid for a finite perturbation, it would be possible to add charged matter to a slightly nonextremal Reissner-Nordstrom black hole so as to overcharge the black hole, as originally argued by Hubeny.

However, our work shows that at second order, there are corrections to the straight line, as illustrated in Fig. 6. Consider a one-parameter family of solutions corresponding to adding charged matter to the black hole. As we have noted above, the curve representing the final state parameters has a tangent whose slope is bounded below by the tangent to the curve of constant area. In addition, however, if its slope is the minimum possible, we have proven in Sec. IV that the second derivative of the curve must be greater than the second derivative of the curve of constant area. The quadratic approximation to this curve thus coincides with the curve of constant area and does not exit the cone. The linear approximation is simply not an adequate approximation. Second order effects do not allow one to exit from the cone.

Finally, it is worth noting that there is a discontinuity in our lower bound on $\delta^2 M$ in the extremal limit. Consider, for simplicity, the case of adding charged matter with no angular momentum to a Reissner-Nordstrom black hole, so $J = \delta J = \delta^2 J = 0$. Without loss of generality, we also may take $\delta^2 Q = 0$. Then, for $\epsilon > 0$, for an optimal perturbation with $\delta M = \Phi_H \delta Q$, it follows from (117) that

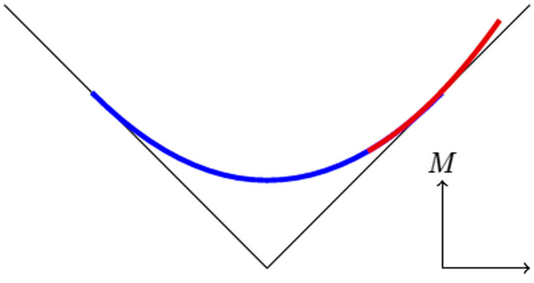

FIG. 6. The quadratic approximation to the curve of final state parameters obtained by adding charged matter to a slightly non-extremal Reissner-Nordstrom black hole.

[12] This result was first obtained for particle matter by Christodoulou [27].

$$\delta^2 M \geq \frac{(\delta Q)^2}{M} + O(\epsilon). \tag{123}$$

Thus, as $\epsilon \to 0$, the right side approaches $(\delta Q)^2/M$. Now consider adding charged matter to an exactly extremal black hole, $\epsilon = 0$. As shown in Sec. III, the optimal perturbation satisfies $\delta M = \Phi_H \delta Q = \delta Q$, so optimally, the perturbation moves one tangent to the cone. However, the derivation of (117) does *not* apply to this case—even if we assume that the linearized perturbation becomes stationary at late times—because our evaluation of $\mathcal{E}_{\Sigma_1}$ is valid only for nonextremal black holes. Nevertheless, if the perturbation *decreases* the charge of the black hole (i.e., if δQ has sign opposite that of Q) then one would expect that $\delta^2 M \geq (\delta Q)^2/M$, so that, optimally, at second order the area of the black hole will remain constant. On the other hand, if δQ *increases* the charge, then there is no reason why this bound need be satisfied since the area of the black hole will increase in any case. Our expectation is that

$$\delta^2 M \geq 0, \tag{124}$$

so that, optimally, the black hole will remain extremal at second order. Indeed, the explicit example of adding a charged shell of matter shows that the lower bound (124) can, in fact, be achieved. Thus, there is a discontinuity between (123) and (124) when $\epsilon \to 0$. It would be interesting to derive (124) from first principles and to see if it is related to other discontinuous behavior as $\epsilon \to 0$, such as the Aretakis instability [28].

ACKNOWLEDGMENTS

We thank Kartik Prabhu for helpful discussions on the gauge invariance of canonical energy. We also thank Hongbao Zhang for numerous comments on the original version of this manuscript, including pointing out several errors. This research was supported in part by NSF Grant No. PHY 15-05124 to the University of Chicago.

APPENDIX: SELF-FORCE ENERGY AND FINITE SIZE EFFECTS

The second-order correction to the mass of a black hole given by Eq. (117) gives a lower bound on the energy of any matter that enters a black hole that is valid to quadratic order in the charge and angular momentum of the body. Since particlelike matter in general relativity must be described as a limiting case of general continuum matter (see [19,20]), this formula applies to particle matter as well. At second order, self-force effects contribute to the energy of a particle. In addition, at second order, a charged body will have an electromagnetic self-energy that diverges when the size of the body is taken to zero, so the size of the body must be finite. However, the finite size of the body may prevent one from lowering the body all the way to the horizon. Our bound (115) must implicitly take into account all of these effects. The purpose of this Appendix is to show explicitly that this is the case for the special case of a charged, particle-like body that enters an uncharged, nonextremal Kerr black hole along the black hole's symmetry axis. The self-force effects in this case were previously calculated by Leaute and Linet [29], while self-energy and finite size effects in this case were previously obtained by Hod [6].

It is particularly easy to evaluate our lower bound on $\delta^2 M$ for the case of a charged body entering a Kerr black hole along the symmetry axis, since $Q = 0$ and $\delta J = \delta^2 J = 0$. An optimal process therefore has $\delta M = 0$ at first order. Thus, (113) reduces to[13]

$$\delta^2 A_B^{KN} = -\frac{8\pi}{\epsilon}(1+\epsilon)(\delta Q)^2 \tag{A1}$$

Hence, (115) yields

$$\delta^2 M \geq -\frac{1}{8\pi}\kappa \delta^2 A_B^{KN} = \frac{r_+}{r_+^2 + a^2}(\delta Q)^2 \tag{A2}$$

where we have used the expression (116) for κ and have used (114) to replace ϵ by r_+. Since $Q = 0$, we have $r_+^2 + a^2 = 2Mr_+$, and so (A2) may be written as

$$\delta^2 M \geq \frac{1}{2M}(\delta Q)^2. \tag{A3}$$

Taking into account the Taylor coefficient of $1/2$, this means that any charged matter with no angular momentum that enters an uncharged black hole must carry an energy

$$E \geq \frac{1}{4M}(\delta Q)^2. \tag{A4}$$

into the black hole. This bound holds for any Kerr black hole with $a < M$.

On the other hand, Leaute and Linet's expression [29] for the (proper, locally measured) self-force on a charged particle on the symmetry axis of Kerr is repulsive and has magnitude

$$f(r) = \frac{Mr}{(r^2 + a^2)^2}(\delta Q)^2. \tag{A5}$$

The force exerted at infinity when lowering the charged body is reduced from this by the redshift factor $(-g_{tt})^{1/2}$ (see, e.g., [30]). However, the infinitesimal proper distance traversed when lowering is given by $dl = (g_{rr})^{1/2} dr$. The factors $(-g_{tt})^{1/2}$ and $(g_{rr})^{1/2}$ cancel on the symmetry axis of Kerr. Thus, we find that the work done at infinity in overcoming the self-force when lowering the charge from infinity to the horizon is

[13] Note that this is an exact expression, i.e., we have not assumed that ϵ is small.

$$E_{SF} = \int_{r_+}^{\infty} f(r)dr = \frac{M}{2(r_+^2 + a^2)}(\delta Q)^2. \qquad \text{(A6)}$$

Note that $E_{SF} < E_{\text{min}}$ for a nonextremal black hole, with E_{min} given by the right side of (A4).

However, the self-force expression is only valid for a small body that is roughly spherical in shape. For such a body, there will be potentially important self-energy and finite size effects, which can be calculated as follows. For a charged spherical body of radius R and charge δQ, the electromagnetic contribution to the rest mass of the body is minimized for a thin shell and is given by

$$m_{\text{EM}} = \frac{1}{2}\frac{(\delta Q)^2}{R}. \qquad \text{(A7)}$$

If the body is dropped into the black hole from a proper distance l from the horizon, its electromagnetic self-energy will contribute an energy

$$E_{\text{self}} = m_{\text{EM}} V(l) \qquad \text{(A8)}$$

to the black hole, where $V(l)$ is the redshift factor at the dropping point. However, near the black hole, we have

$$V(l) = \kappa l, \qquad \text{(A9)}$$

where κ is the surface gravity of the black hole. Since we must have $l \geq R$, we obtain

$$E_{\text{self}} \geq \frac{\kappa}{2}(\delta Q)^2. \qquad \text{(A10)}$$

Substituting for κ from (116) and adding these two contributions yields a minimal total added energy of

$$E_{\text{self}} + E_{SF} = \frac{(\delta Q)^2}{4M}, \qquad \text{(A11)}$$

in exact agreement[14] with (A4). Thus, we see explicitly in this example how our general bound (A4) incorporates both self-force effects and self-energy/finite size effects.

One could attempt to evade our bound by making E_{self} smaller by choosing, instead of a small spherical shell, a body that has radial extent much smaller than its angular extent. Such a body could be lowered arbitrarily close to the black hole without making its self-energy arbitrarily large. However, choosing such a shape for the body would result in other second-order corrections to the energy (such as self-repulsion effects) that would inevitably have to reproduce our bound (A4). As an extreme example of this, one can consider a thin spherical shell of charge collapsing around a Schwarzchild black hole, which experiences a large self-repulsion but for which the (redshifted) electromagnetic self-energy can be made exactly zero. Using the methods of Boulware [31], it is straightforward to show that such a shell still adds a minimal energy of $(\delta Q)^2/4M$ to the black hole. This illustrates, again, that our bound automatically takes all effects on energy into account.

[14] For a nearly extremal Kerr black hole, this is sufficient to prevent overextremizing the black hole, as previously found by Hod [6].

[1] R. Penrose, Riv. Nuovo Cimento **1**, 252 (1969).
[2] R. M. Wald, arXiv:gr-qc/9710068.
[3] R. M. Wald, Ann. Phys. (N.Y.) **82**, 548 (1974).
[4] V. E. Hubeny, Phys. Rev. D **59**, 064013 (1999).
[5] F. de Felice and Y. Yu, Classical Quantum Gravity **18**, 1235 (2001).
[6] S. Hod, Phys. Rev. D **66**, 024016 (2002).
[7] T. Jacobson and T. Sotiriou, Phys. Rev. Lett. **103**, 141101 (2009).
[8] G. Chirco, S. Liberati, and T. P. Sotiriou, Phys. Rev. D **82**, 104015 (2010).
[9] A. Saa and R. Santarelli, Phys. Rev. D **84**, 027501 (2011).
[10] S. Gao and Y. Zhang, Phys. Rev. D **87**, 044028 (2013).
[11] P. Zimmerman, I. Vega, E. Poisson, and R. Haas, Phys. Rev. D **87**, 041501 (2013).
[12] E. Barausse, V. Cardoso, and G. Khanna, Phys. Rev. Lett. **105**, 261102 (2010).
[13] E. Barausse, V. Cardoso, and G. Khanna, Phys. Rev. D **84**, 104006 (2011).
[14] M. Colleoni and L. Barack, Phys. Rev. D **91**, 104024 (2015).
[15] M. Colleoni, L. Barack, A. G. Shah, and M. van de Meent, Phys. Rev. D **92**, 084044 (2015).
[16] S. Gao and R. M. Wald, Phys. Rev. D **64**, 084020 (2001).
[17] G. Z. Tóth, Gen. Relativ. Gravit. **44**, 2019 (2012).
[18] J. Natário, L. Queimada, and R. Vicente, Classical Quantum Gravity **33**, 175002 (2016).
[19] S. E. Gralla and R. M. Wald, Classical Quantum Gravity **28**, 159501 (2011).
[20] S. E. Gralla, A. I. Harte, and R. M. Wald, Phys. Rev. D **80**, 024031 (2009).
[21] S. Hollands and R. M. Wald, Commun. Math. Phys. **321**, 629 (2013).
[22] R. M. Wald, *General Relativity* (University of Chicago Press, Chicago, 1984).
[23] K. Prabhu, Classical Quantum Gravity **34**, 035011 (2017).
[24] V. Iyer and R. M. Wald, Phys. Rev. D **50**, 846 (1994).
[25] V. Iyer and R. M. Wald, Phys. Rev. D **52**, 4430 (1995).

[26] B. Carter, in *Les Houches 1972*, edited by C. Dewitt and B. S. Dewitt (Gordon and Breach, New York, 1973).

[27] D. Christodoulou, Phys. Rev. Lett. **25**, 1596 (1970).

[28] S. Aretakis, Adv. Theor. Math. Phys. **19**, 507 (2015).

[29] B. Léauté and B. Linet, J. Phys. A **15**, 1821 (1982).

[30] W. G. Unruh and R. M. Wald, Phys. Rev. D **25**, 942 (1982).

[31] D. G. Boulware, Phys. Rev. D **8**, 2363 (1973).