Some properties of the Noether charge and a proposal for dynamical black hole entropy

Vivek Iyer and Robert M. Wald
Enrico Fermi Institute and Department of Physics, University of Chicago, 5640 South Ellis Avenue, Chicago, Illinois 60637
(Received 17 March 1994)

We consider a general, classical theory of gravity with arbitrary matter fields in n dimensions, arising from a diffeomorphism-invariant Lagrangian $\mathbf{L}$. We first show that $\mathbf{L}$ always can be written in a "manifestly covariant" form. We then show that the symplectic potential current $(n-1)$-form $\boldsymbol{\Theta}$ and the symplectic current $(n-1)$-form $\boldsymbol{\omega}$ for the theory always can be globally defined in a covariant manner. Associated with any infinitesimal diffeomorphism is a Noether current $(n-1)$-form $\mathbf{J}$ and corresponding Noether charge $(n-2)$-form $\mathbf{Q}$. We derive a general "decomposition formula" for $\mathbf{Q}$. Using this formula for the Noether charge, we prove that the first law of black hole mechanics holds for arbitrary perturbations of a stationary black hole. (For higher derivative theories, previous arguments had established this law only for stationary perturbations.) Finally, we propose a local, geometrical prescription for the entropy $S_{\rm dyn}$ of a dynamical black hole. This prescription agrees with the Noether charge formula for stationary black holes and their perturbations, and is independent of all ambiguities associated with the choices of $\mathbf{L}$, $\boldsymbol{\Theta}$, and $\mathbf{Q}$. However, the issue of whether this dynamical entropy in general obeys a "second law" of black hole mechanics remains open. In an appendix, we apply some of our results to theories with a nondynamical metric and also briefly develop the theory of stress-energy pseudotensors.

PACS number(s): 04.20.Fy, 97.60.Lf

I. INTRODUCTION

Recently, many authors have investigated the validity of the first law of black hole mechanics and the definition of the entropy of a black hole in a wide class of theories derivable from a Hamiltonian or Lagrangian [1–10]. In particular, in [6] the first law was proven to hold in an arbitrary theory of gravity derived from a diffeomorphism invariant Lagrangian, and the quantity playing the role of the entropy of the black hole was identified as the integral over the horizon of the Noether charge associated with the horizon Killing vector field. Although some key issues concerning the validity of the first law and the definition of black hole entropy in a general theory of gravity were thereby resolved, the analysis of [6], nevertheless, was deficient in the following ways. (1) It was not recognized that a diffeomorphism covariant choice of the symplectic potential current form always can be made. Consequently, several steps in the arguments were made in an unnecessarily awkward manner. (2) While a completely general proof of the first law of black hole mechanics was given for perturbations to nearby stationary black holes, a proof of the first law for nonstationary perturbations was given only for theories in which the Noether charge takes a particular, simple form. (3) A proposal was made for defining the entropy of a dynamical black hole. However, this proposal made use of a rather arbitrary choice of algorithm for defining the symplectic potential current form, and it turns out to possess the undesirable feature that the addition of an exact form to the Lagrangian (which has no effect upon the equations of motion of the theory) can induce a nontrivial change in this proposed formula for the entropy of a dynamical black hole [11].

The main purposes of this paper are to remedy all of the above deficiencies, and, in addition, develop further the theory of Noether currents and charges in diffeomorphism invariant theories. We shall show, first, that the Lagrangian of a diffeomorphism invariant theory always can be expressed in a manifestly covariant form. This will enable us to give globally defined, covariant definitions of the symplectic potential current form $\boldsymbol{\Theta}$ and symplectic current form $\boldsymbol{\omega}$ in an arbitrary diffeomorphism invariant theory. Furthermore, results on the general form of $\boldsymbol{\Theta}$ in an arbitrary theory will be obtained, from which it will follow that the Noether charge form $\mathbf{Q}$ always has a particular, simple structure. As a consequence of this structure of $\mathbf{Q}$, the first law of black hole mechanics will be proven to hold for nonstationary perturbations in an arbitrary theory of gravity derived from a diffeomorphism covariant Lagrangian. We thereby obtain a formula for the entropy S of a stationary black hole, which generalizes formulas of [7] and [8] to theories of gravity of arbitrarily high derivative order and matter couplings. We then shall propose a definition of the entropy $S_{\rm dyn}$ of an arbitrary cross section of a nonstationary black hole, wherein $S_{\rm dyn}$ is given by an integral over the horizon of a local, geometrical quantity. Our proposed definition agrees with the known answer (as determined by the first law) for stationary black holes and their perturbations, and is independent of all ambiguities associated with the choices of $\mathbf{L}$, $\boldsymbol{\Theta}$, and $\mathbf{Q}$. However, it is not known whether our $S_{\rm dyn}$ obeys a "second law" in general theories of gravity. The paper concludes with an Appendix in which some of our results are applied to theories with a nondynamical metric, and some results on stress-energy pseudotensors

are obtained.

We shall follow the notation and conventions of [12]. All spacetimes, tensor fields, and surfaces considered in this paper will be assumed to be smooth (C^∞).

II. THE FORM OF THE LAGRANGIAN FOR DIFFEOMORPHISM INVARIANT THEORIES

We wish to consider, here, Lagrangian theories on an n-dimensional, oriented manifold M, with the dynamical fields consisting of a Lorentz signature metric g_{ab}, and other fields ψ. For simplicity and definiteness, we shall restrict consideration to the case where ψ is a collection of tensor fields on M (with arbitrary index structure). However, we foresee no essential difficulty in extending our analysis and results to the case where ψ is a section of an arbitrary vector bundle which possesses a connection uniquely determined by g_{ab}.

We start with the general form of a Lagrangian postulated in [15] and [6]: Specifically, we introduce an arbitrary, fixed, globally defined, derivative operator $\overset{\circ}{\nabla}$ and take the Lagrangian to be a function of the quantities g_{ab}, ψ, and finitely many of their symmetrized derivatives with respect to $\overset{\circ}{\nabla}$. In addition, the Lagrangian is permitted to depend on additional "background fields" $\overset{\circ}{\gamma}$ – which, like $\overset{\circ}{\nabla}$, do not change under variation of the dynamical fields; a good example of such a background field upon which the Lagrangian could depend is the curvature $\overset{\circ}{R}_{bcd}{}^{e}$ of $\overset{\circ}{\nabla}$. Thus, we take the Lagrangian to be an n-form locally constructed, in the precise sense explained in [14], out of the quantities

$$\mathbf{L} = \mathbf{L}\left(g_{ab}, \overset{\circ}{\nabla}_{a_1} g_{ab}, ..., \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_k)} g_{ab}, \psi, \overset{\circ}{\nabla}_{a_1} \psi, ..., \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_l)} \psi, \overset{\circ}{\gamma}\right). \tag{1}$$

Here and in what follows, we use boldface letters to denote differential forms on spacetime, and we shall, in general, suppress their tensor indices. In the following, we also shall collectively refer to the dynamical fields "ψ and g" as "ϕ."

We shall be concerned here only with diffeomorphism invariant theories; i.e., the Lagrangian will be assumed to be diffeomorphism covariant in the sense that

$$\mathbf{L}(f^*(\phi)) = f^*\mathbf{L}(\phi), \tag{2}$$

where f^* is the action induced on the fields by a diffeomorphism $f : M \to M$. Note that on the left side of this equation f^* does not act on $\overset{\circ}{\nabla}$ or the background fields $\overset{\circ}{\gamma}$.

The main result to be established in this section is that any Lagrangian $\mathbf{L}$, which is diffeomorphism covariant in the sense of Eq. (2), always can be written in a manifestly covariant form. More precisely, we have the following lemma, which is closely related to the "Thomas replacement theorem" [16].

Lemma 2.1. If $\mathbf{L}$ *as given in (1) is diffeomorphism covariant in the sense of (2) then L can be reexpressed as*

$$\mathbf{L} = \mathbf{L}\left(g_{ab}, \nabla_{a_1} R_{bcde}, ..., \nabla_{(a_1} \cdots \nabla_{a_m)} R_{bcde}, \psi, \nabla_{a_1}\psi, \nabla_{(a_1} \cdots \nabla_{a_l)}\psi\right), \tag{3}$$

where ∇ *denotes the derivative operator associated with* g_{ab}, $m = \max(k-2, l-2)$, R_{abcd} *denotes the curvature of* g_{ab}, *and the absence of any dependence on "background fields" in (3) should be noted.*

Proof. We begin by using the relation (written here schematically)

$$\overset{\circ}{\nabla}\alpha = \nabla\alpha + \alpha \times \text{terms linear in } \overset{\circ}{\nabla} g$$

for any tensor field α to rewrite all of the $\overset{\circ}{\nabla}$ derivatives of the matter fields ψ in terms of ∇ derivatives of ψ, where ∇ is the derivative operator associated with g, together with terms involving the $\overset{\circ}{\nabla}$ derivatives of g. Next, we rewrite the ∇ derivatives of ψ in terms of symmetrized ∇ derivatives and the curvature of g and its derivatives. Then we rewrite the curvature of g and its derivatives in terms of $\overset{\circ}{\nabla}$ derivatives of g and the curvature of $\overset{\circ}{\nabla}$ and its $\overset{\circ}{\nabla}$ derivatives. Finally, we write all of the $\overset{\circ}{\nabla}$ derivatives of g in terms of symmetrized $\overset{\circ}{\nabla}$ derivatives of g and the curvature of $\overset{\circ}{\nabla}$ and its $\overset{\circ}{\nabla}$ derivatives. We thereby obtain

$$\mathbf{L} = \mathbf{L}\left(g, \overset{\circ}{\nabla}_{a_1} g_{ab}, ..., \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_s)} g_{ab}, \psi, \nabla_{a_1}\psi, ..., \nabla_{(a_1} \cdots \nabla_{a_l)}\psi, \overset{\circ}{\gamma}{}'\right), \tag{4}$$

where $s = \max(k, l)$ and $\overset{\circ}{\gamma}{}'$ is comprised by $\overset{\circ}{\gamma}$ together with the curvature of $\overset{\circ}{\nabla}$ and (finitely many of) its $\overset{\circ}{\nabla}$ derivatives. Next we eliminate $\overset{\circ}{\nabla}_a g_{bc}$ and its higher $\overset{\circ}{\nabla}$ derivatives in favor of

$$C^e{}_{cd} = \tfrac{1}{2} g^{ef}(\overset{\circ}{\nabla}_c g_{fd} + \overset{\circ}{\nabla}_d g_{fc} - \overset{\circ}{\nabla}_f g_{cd}) \tag{5}$$

and its $\overset{\circ}{\nabla}$ derivatives via the substitution

$$\overset{\circ}{\nabla}_a g_{bc} = g_{ec} C^e{}_{ab} + g_{be} C^e{}_{ac}. \tag{6}$$

Again, we express all $\overset{\circ}{\nabla}$ derivatives of $C^e{}_{cd}$ in terms of symmetrized $\overset{\circ}{\nabla}$ derivatives and the curvature of $\overset{\circ}{\nabla}$. We thereby obtain

$$\mathbf{L} = \mathbf{L}\left(g, C^e{}_{cd}, \overset{\circ}{\nabla}_{a_1} C^e{}_{cd}, ..., \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_{s-1})} C^e{}_{cd}, \psi, \nabla_{a_1}\psi, \nabla_{(a_1} \cdots \nabla_{a_l)}\psi, \overset{\circ}{\gamma}{}'\right). \tag{7}$$

It is tedious but straightforward to check that the symmetrized derivatives of C can be rewritten as

$$\begin{aligned}
\overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_p)} C^e{}_{cd} = {} & \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_p} C^e{}_{cd)} + \frac{p+3}{4(p+1)(p+2)} \sum_i \nabla_{(a_1} \cdots \hat{\nabla}_{a_i} \cdots \nabla_{a_p)} \left(R_{ca_id}{}^e + R_{da_ic}{}^e\right) \\
& + \frac{3p+4}{8p(p+1)(p+2)} \sum_{i \neq j} \Big(\nabla_{(d} \nabla_{a_1} \cdots \hat{\nabla}_{a_i} \hat{\nabla}_{a_j} \cdots \nabla_{a_p)} R_{a_ica_j}{}^e \\
& + \nabla_{(c} \nabla_{a_1} \cdots \hat{\nabla}_{a_i} \hat{\nabla}_{a_j} \cdots \nabla_{a_p)} R_{a_ida_j}{}^e \Big) \\
& + \text{terms involving no more than } (p-1) \overset{\circ}{\nabla} \text{ derivatives of } C,
\end{aligned} \tag{8}$$

where $\hat{\nabla}_{a_i}$ means the omission of this derivative operator in the sequence. By repeatedly making this substitution in sequence, starting with $p = s - 1$, then $p = s - 2$, etc., and, at each step, writing multiple derivatives in terms of symmetrized derivatives and curvatures, we can express the Lagrangian as

$$\begin{aligned}
\mathbf{L} = \mathbf{L}\Big(& g_{ab}, C^e{}_{cd}, \overset{\circ}{\nabla}_{(a_1} C^e{}_{cd)}, ..., \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_{s-1}} C^e{}_{cd)}, R_{bcde}, \nabla_{a_1} R_{bcde}, ..., \nabla_{(a_1} \cdots \nabla_{a_{s-2})} R_{bcde}, \\
& \psi, \nabla_{a_1}\psi, ..., \nabla_{(a_1} \cdots \nabla_{a_l)}\psi, \overset{\circ}{\gamma}{}'\Big).
\end{aligned} \tag{9}$$

The infinitesimal version of the diffeomorphism covariance condition (2) is

$$\mathcal{L}_\xi L(\phi) = \frac{\partial L}{\partial \phi} \mathcal{L}_\xi \phi. \tag{10}$$

Applying this to Eq. (9) we obtain

$$\begin{aligned}
& \frac{\partial \mathbf{L}}{\partial C^e{}_{cd}} \mathcal{L}_\xi C^e{}_{cd} + \frac{\partial \mathbf{L}}{\partial\, \overset{\circ}{\nabla}_{(a_1} C^e{}_{cd)}} \mathcal{L}_\xi \overset{\circ}{\nabla}_{(a_1} C^e{}_{cd)} + \cdots + \frac{\partial \mathbf{L}}{\partial\, \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_{s-1}} C^e{}_{cd)}} \mathcal{L}_\xi \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_{s-1}} C^e{}_{cd)} + \frac{\partial \mathbf{L}}{\partial\, \overset{\circ}{\gamma}{}'} \mathcal{L}_\xi \overset{\circ}{\gamma}{}' \\
& = \frac{\partial \mathbf{L}}{\partial C^e{}_{cd}} \delta C^e{}_{cd} + \frac{\partial \mathbf{L}}{\partial\, \overset{\circ}{\nabla}_{(a_1} C^e{}_{cd)}} \overset{\circ}{\nabla}_{(a_1} \delta C^e{}_{cd)} + \cdots + \frac{\partial \mathbf{L}}{\partial\, \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_{s-1}} C^e{}_{cd)}} \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_{s-1}} \delta C^e{}_{cd)},
\end{aligned} \tag{11}$$

where

$$\delta C^a{}_{bc} = g^{ad}\left(\overset{\circ}{\nabla}_{(b} \overset{\circ}{\nabla}_{c)} \xi_d - \overset{\circ}{R}_{d(bc)}{}^e \xi_e\right) - 2\, \overset{\circ}{\nabla}{}^{(a} \xi^{d)} g_{de} C^e{}_{bc} \tag{12}$$

is the variation of $C^a{}_{bc}$ arising from the metric variation $\delta g_{ab} = \mathcal{L}_\xi g_{ab} = 2\nabla_{(a}\xi_{b)}$. [Note that the terms in Eq. (10) arising from the variations of ψ, g, and the curvature R of g cancel and were therefore omitted when writing Eq. (11).] The dependence of the terms in Eq. (11) on ξ^a and its symmetrized $\overset{\circ}{\nabla}$ derivatives should be noted: no more than one derivative of ξ^a appears on the left side of this equation, but the right side contains terms with as many

as $(s+1)$ symmetrized derivatives of ξ^a. Since, at any given point in M, ξ^a and its symmetrized derivatives can be chosen independently, it follows directly that a necessary condition for Eq. (11) to hold for all ξ^a is

$$\frac{\partial \mathbf{L}}{\partial \overset{\circ}{\nabla}_{(a_1} \cdots \overset{\circ}{\nabla}_{a_i} C^e{}_{cd)}} = 0 \quad \text{for } i = 0, ..., (s-1). \tag{13}$$

This reduces the Lagrangian to the form

$$\mathbf{L} = \mathbf{L}\left(g_{ab}, R_{bcde}, \nabla_{a_1} R_{bcde}, ..., \nabla_{(a_1} \cdots \nabla_{a_m)} R_{bcde}, \psi, \nabla_{a_1}\psi, ..., \nabla_{(a_1} \cdots \nabla_{a_l)}\psi, \overset{\circ}{\gamma}{}'\right), \tag{14}$$

where $m = s-2 = \max(k-2, l-2)$. The diffeomorphism invariance condition (11) yields one more relation: namely,

$$\frac{\partial \mathbf{L}}{\partial \overset{\circ}{\gamma}{}'} \mathcal{L}_\xi \overset{\circ}{\gamma}{}' = 0, \tag{15}$$

where a sum over the fields $\overset{\circ}{\gamma}{}'$ should be understood. To show that this implies that $\mathbf{L}$ has no essential dependence on $\overset{\circ}{\gamma}{}'$, we proceed by introducing a local coordinate system $x^1, ..., x^n$, and viewing $\mathbf{L}$ as a function of the coordinate components of the dynamical fields and $\overset{\circ}{\gamma}{}'$. We then view the components of $\overset{\circ}{\gamma}{}'$ as given functions of x^μ. In this way, we may view $\mathbf{L}$ as

$$\mathbf{L} = \mathbf{L}\left(g, R_{bcde}, \nabla_{a_1} R_{bcde}, ..., \nabla_{(a_1} \cdots \nabla_{a_m)} R_{bcde}, \psi, \nabla_{a_1}\psi, ..., \nabla_{(a_1} \cdots \nabla_{a_l)}\psi, x^\mu\right); \tag{16}$$

i.e., we replace the dependence of $\mathbf{L}$ on $\overset{\circ}{\gamma}{}'$ by explicit dependence on coordinates. Condition (11) then implies that

$$\sum_\mu \frac{\partial \mathbf{L}}{\partial x^\mu} \mathcal{L}_\xi x^\mu = 0. \tag{17}$$

Clearly, this equation holds for all ξ^a if and only if

$$\frac{\partial \mathbf{L}}{\partial x^\mu} = 0. \tag{18}$$

We therefore see that (18) implies that any diffeomorphism invariant Lagrangian must be of the form

$$\mathbf{L} = \mathbf{L}\left(g_{ab}, R_{bcde}, \nabla_{a_1} R_{bcde}, ..., \nabla_{(a_1} \cdots \nabla_{a_m)} R_{bcde}, \psi, \nabla_{a_1}\psi, \nabla_{(a_1} \cdots \nabla_{a_l)}\psi\right), \tag{19}$$

as we desired to show. □

Note added. It should be noted that on account of differential identities satisfied by the curvature, the quantities $R_{abcd}, \nabla_e R_{abcd}, \nabla_{(f}\nabla_{e)} R_{abcd}, ...$, cannot be specified independently at a point. As a consequence, the choice of Lagrangian function of the form (3) is not unique. A natural way of fixing the Lagrangian would be to choose it to depend upon the the derivatives of the curvature only in the combination [13]

$$g_{ab} g_{cd} \nabla_{(e_1} \cdots \nabla_{e_m} R^b{}_g{}^d{}_{i)}$$

since, as proven recently by Anderson and Torre (unpublished), these quantities provide an appropriate set of "independent fields." In the following we shall assume that this choice of Lagrangian has been made, although all of the considerations of the paper would apply to any choice of Lagrangian of the form (3). We wish to thank Ian Anderson and Charles Torre for bringing this issue to our attention.

III. THE FORM OF THE SYMPLECTIC POTENTIAL AND SYMPLECTIC CURRENTS FOR DIFFEOMORPHISM INVARIANT THEORIES

As is well known (and as will be explicitly demonstrated in Lemma 3.1 below), if we vary the dynamical fields $\phi = (g_{ab}, \psi)$, then, by "integration by parts" manipulations of the terms involving the derivatives of $\delta\phi$, the first variation of the Lagrangian can always be expressed in the form

$$\delta \mathbf{L} = \mathbf{E}\delta\phi + d\mathbf{\Theta} \tag{20}$$

with

$$\mathbf{E}\delta\phi = (\mathbf{E}_g)^{ab}\delta g_{ab} + \mathbf{E}_\psi \delta\psi, \tag{21}$$

where a sum over the "matter fields" ψ is understood, and it is also understood that for each matter field, E_ψ

has tensor indices dual to ψ, and these indices are contracted with those of $\delta\psi$ in Eq. (21). Here $\mathbf{E}_g$ and $\mathbf{E}_\psi$ are locally constructed out of the dynamical fields ϕ and their derivatives, whereas $\mathbf{\Theta}$ is locally constructed out of ϕ, $\delta\phi$ and their derivatives and is linear in $\delta\phi$. The equations of motion of the theory are then taken to be

$$(\mathbf{E}_g)^{ab} = 0 \text{ , and } \mathbf{E}_\psi = 0. \tag{22}$$

The $(n-1)$-form $\mathbf{\Theta}$ defined by Eq. (20) is called the *symplectic potential form* (see below). However, although the equations of motion form $\mathbf{E}$ is uniquely determined by Eq. (20), this equation determines $\mathbf{\Theta}$ only up to the addition of a closed (and hence exact [14]) $(n-1)$-form. Thus, some arbitrariness is present in the choice of $\mathbf{\Theta}$. The principal result of this section is stated in the following lemma.

Lemma 3.1 Given a covariant Lagrangian of the form (19) one can always choose a covariant $\mathbf{\Theta}$ *satisfying (20). Moreover,* $\mathbf{\Theta}$ *can be chosen to have the form*

$$\mathbf{\Theta} = 2\mathbf{E}_R^{bcd}\nabla_d\delta g_{bc} + \mathbf{\Theta}', \tag{23}$$

where $\mathbf{\Theta}'$ *is of the form*

$$\mathbf{\Theta}' = \mathbf{S}^{ab}(\phi)\delta g_{ab} + \sum_{i=0}^{m-1}\mathbf{T}_i(\phi)^{abcda_1\cdots a_i}\delta\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd} + \sum_{i=0}^{l-1}\mathbf{U}_i(\phi)^{a_1\cdots a_i}\delta\nabla_{(a_1}\cdots\nabla_{a_i)}\psi. \tag{24}$$

In other words, in the expression for $\mathbf{\Theta}$*, the* δ*'s can be put to the left of derivatives of the dynamical fields everywhere except for the single term* $\mathbf{E}_R^{bcd}\nabla_d\delta g_{bc}$*. Finally,* $\mathbf{E}_R^{bcd}$ *is given by*

$$(\mathbf{E}_R^{bcd})_{b_2\cdots b_n} = E_R^{abcd}\boldsymbol{\epsilon}_{ab_2\cdots b_n}, \tag{25}$$

where $E_R^{abcd}\boldsymbol{\epsilon}_{b_1b_2\cdots b_n}$ *is the equation of motion form that would be obtained for* R_{abcd} *if it were viewed as an independent field in the Lagrangian (19) rather than a quantity determined by the metric.*

Proof. Given $\mathbf{L}$ in form (19) we write it as $\mathbf{L} = L\boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon}$ is the canonical volume form on M associated with g_{ab}. Computing a first variation, we obtain

$$\begin{aligned}\delta\mathbf{L} = \boldsymbol{\epsilon}\Bigg(&\frac{\partial L}{\partial g_{ab}}\delta g_{ab} + \frac{\partial L}{\partial R_{abcd}}\delta R_{abcd} + \frac{\partial L}{\partial\nabla_{a_1}R_{abcd}}\delta\nabla_{a_1}R_{abcd} + \cdots + \frac{\partial L}{\partial\nabla_{(a_1}\cdots\nabla_{a_m)}R_{abcd}}\delta\nabla_{(a_1}\cdots\nabla_{a_m)}R_{abcd}\\ &+ \frac{\partial L}{\partial\psi}\delta\psi + \frac{\partial L}{\partial\nabla_{a_1}\psi}\delta\nabla_{a_1}\psi + \cdots + \frac{\partial L}{\nabla_{(a_1}\cdots\nabla_{a_l)}\psi}\delta\nabla_{(a_1}\cdots\nabla_{a_l)}\psi\Bigg) + \frac{1}{2}g^{ab}\delta g_{ab}\mathbf{L}.\end{aligned} \tag{26}$$

(For tensors, such as δR_{abcd}, whose components are not algebraically independent at each point, we uniquely fix the partial derivative coefficients appearing in this equation by requiring them to have precisely the same tensor symmetries as the varied quantities.) In order to obtain the desired expression for $\mathbf{\Theta}$ we must suitably rewrite Eq. (26) in the form (20). To see how this can be done we focus attention on a typical term

$$\boldsymbol{\epsilon}\frac{\partial L}{\partial\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd}}\delta\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd} \tag{27}$$

and rewrite it as

$$\begin{aligned}\boldsymbol{\epsilon}\frac{\partial L}{\partial\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd}}\delta\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd} &= \boldsymbol{\epsilon}\left(\frac{\partial L}{\partial\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd}}\nabla_{a_1}(\delta\nabla_{a_2}\cdots\nabla_{a_i}R_{abcd})\right)\\ &\quad+\boldsymbol{\epsilon}\cdot\text{ terms proportional to }\nabla\delta g\\ &= \nabla_{a_1}\left(\boldsymbol{\epsilon}\frac{\partial L}{\partial\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd}}\delta\nabla_{a_2}\cdots\nabla_{a_i}R_{abcd}\right)\\ &\quad+\nabla_{a_1}[\boldsymbol{\epsilon}\cdot(\text{terms proportional to }\delta g)]\\ &\quad-\nabla_{a_1}\left(\boldsymbol{\epsilon}\frac{\partial L}{\partial\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd}}\right)\delta\nabla_{a_2}\cdots\nabla_{a_i}R_{abcd}\\ &\quad+\boldsymbol{\epsilon}\cdot(\text{terms proportional to }\delta g),\\ &= d\mathbf{V} - \boldsymbol{\epsilon}\nabla_{a_1}\left(\frac{\partial L}{\partial\nabla_{(a_1}\cdots\nabla_{a_i)}R_{abcd}}\right)\delta\nabla_{a_2}\cdots\nabla_{a_i}R_{abcd}\\ &\quad+\boldsymbol{\epsilon}\cdot(\text{ terms proportional to }\delta g),\end{aligned} \tag{28}$$

where the $(n-1)$-form $\mathbf{V}$ has the form

$$V_{b_2\cdots b_n} = \epsilon_{a_1 b_2\cdots b_n} \frac{\partial L}{\partial \nabla_{(a_1}\cdots\nabla_{a_i)} R_{abcd}} \delta\nabla_{a_2}\cdots\nabla_{a_i} R_{abcd} + \epsilon \times (\text{terms proportional to } \delta g). \tag{29}$$

This shows that we can rewrite our original term (27) as a sum of a similar term of lower differential order, the exact form $d\mathbf{V}$, and terms proportional to δg. By iterating this procedure and performing similar manipulations on all other terms in (26) containing derivatives of variations of the curvature or matter fields, we obtain

$$\delta\mathbf{L} = \epsilon\left(A_g^{ab}\delta g_{ab} + E_R^{abcd}\delta R_{abcd} + E_\psi \delta\psi\right) + d\tilde{\mathbf{\Theta}}, \tag{30}$$

where the $(n-1)$-form $\tilde{\mathbf{\Theta}}$ is covariant and has the same structure as the right side of Eq. (24). Note that in Eq. (30) ϵE_ψ is precisely the equations of motion form for the matter fields ψ and

$$\epsilon E_R^{abcd} = \epsilon\left(\frac{\partial L}{\partial R_{abcd}} - \nabla_{a_1}\frac{\partial L}{\partial \nabla_{a_1} R_{abcd}} + \cdots + (-1)^m \nabla_{(a_1}\cdots\nabla_{a_m)}\frac{\partial L}{\partial \nabla_{(a_1}\cdots\nabla_{a_m)} R_{abcd}}\right) \tag{31}$$

would be the equations of motion form for R_{abcd} *if* it were viewed as an independent field. In fact, however, R_{abcd} is not an independent field, and, taking account of the symmetries of E_R^{abcd}, we have

$$E_R^{abcd}\delta R_{abcd} = 2E_R^{abcd}\nabla_a\nabla_d\delta g_{bc} + E_R^{abcd} R_{abc}{}^e \delta g_{de}. \tag{32}$$

Making this substitution and integrating twice by parts, we obtain

$$\begin{aligned}\delta\mathbf{L} = {} & \epsilon(\tilde{A}_g^{bc}\delta g_{bc} + 2\nabla_a\nabla_d E_R^{abcd}\delta g_{bc} + E_\psi\delta\psi) \\ & + d(2\mathbf{E}^{bcd}\nabla_d g_{bc} - 2\nabla_d \mathbf{E}^{bcd}\delta g_{bc} + \tilde{\mathbf{\Theta}}),\end{aligned} \tag{33}$$

where

$$(\mathbf{E}^{bcd})_{b_2\cdots b_n} = E_R^{abcd}\epsilon_{ab_2\cdots b_n} \tag{34}$$

and

$$\tilde{A}_g^{bc} = A_g^{bc} + E_R^{pqrb} R_{pqr}{}^c. \tag{35}$$

Note that the equations of motion associated with g_{ab} are thus

$$(\mathbf{E}_g)^{bc} = \epsilon\left(\tilde{A}_g^{bc} + 2\nabla_a\nabla_d E_R^{abcd}\right). \tag{36}$$

Thus, by inspection (33) is of the form (20) with

$$\mathbf{\Theta} = 2\mathbf{E}_R^{bcd}\nabla_d\delta g_{bc} + \mathbf{\Theta}', \tag{37}$$

where

$$\mathbf{\Theta}' \equiv \tilde{\mathbf{\Theta}} - 2\nabla_d \mathbf{E}_R^{bcd}\delta g_{bc}. \tag{38}$$

This shows that $\mathbf{\Theta}$ is manifestly covariant and of the form claimed in the statement of the lemma. □

We comment, now, on the possible ambiguities in the choice of $\mathbf{\Theta}$ for a covariant Lagrangian. The above lemma proves that $\mathbf{\Theta}$ always can be chosen to be covariant. This appears to be a very natural requirement, and, in the following, we shall restrict consideration to covariant choices of $\mathbf{\Theta}$. The statement of the lemma also provides the canonical form (23) for $\mathbf{\Theta}$, which will play an important role in our analysis below. However, this general form does not uniquely determine $\mathbf{\Theta}$, since one could add to $\mathbf{\Theta}$ an exact $(n-1)$-form which has the structure of the right side of Eq. (24). The proof of the lemma does implicitly provide a particular algorithm which uniquely determines a particular $\mathbf{\Theta}$ from a given $\mathbf{L}$, but there does not appear to be any reason to prefer this algorithm over other possible ones. Thus, it appears most preferable to leave the choice of $\mathbf{\Theta}$ unspecified apart from the restriction of covariance. In fact, there exist two independent sources of ambiguity in $\mathbf{\Theta}$.

First, as noted above, Eq. (20) allows the freedom to alter $\mathbf{\Theta}$ by addition of an exact $(n-1)$-form

$$\mathbf{\Theta} \to \mathbf{\Theta} + d\mathbf{Y}(\phi, \delta\phi), \tag{39}$$

where the covariant $(n-2)$-form $\mathbf{Y}$ is linear in the varied fields. Second, if we alter the Lagrangian by addition of an exact n-form

$$\mathbf{L} \to \mathbf{L} + d\boldsymbol{\mu}, \tag{40}$$

then the equations of motion are unaffected, so we do not alter the dynamical content of the theory. Nevertheless, $\mathbf{\Theta}$ must be shifted by

$$\mathbf{\Theta} \to \mathbf{\Theta} + \delta\boldsymbol{\mu}. \tag{41}$$

(If $\mathbf{\Theta}$ is defined by the algorithm implicit in the proof of the above lemma, then an additional exact term also would be added to $\mathbf{\Theta}$.) Thus, $\mathbf{\Theta}$ is ambiguous up to the addition of two terms

$$\mathbf{\Theta} \to \mathbf{\Theta} + \delta\boldsymbol{\mu} + d\mathbf{Y}(\phi, \delta\phi). \tag{42}$$

The consequences of this ambiguity in $\mathbf{\Theta}$ for the Noether current and charge will be analyzed in the next section.

We conclude this section by briefly reviewing the definition of the symplectic form Ω in globally hyperbolic spacetimes and investigating its possible ambiguities for asymptotically flat solutions. This is of relevance here because Ω is used to define the notion of a Hamiltonian, which, in turn, gives rise to the notions of total energy and angular momentum. Thus, ambiguities in Ω could result in ambiguities in these notions.

Recall that the symplectic current $(n-1)$-form [15] is defined by

$$\boldsymbol{\omega}(\phi, \delta_1\phi, \delta_2\phi) = \delta_2\boldsymbol{\Theta}(\phi, \delta_1\phi) - \delta_1\boldsymbol{\Theta}(\phi, \delta_2\phi). \tag{43}$$

Let C be a Cauchy surface. We take the orientation of C to be given by $\tilde{\boldsymbol{\epsilon}}_{a_1\cdots a_{n-1}} = n^b\boldsymbol{\epsilon}_{ba_1\cdots a_{n-1}}$ where n^a is the future pointing normal to C and $\boldsymbol{\epsilon}_{ba_1\cdots a_{n-1}}$ is the positively oriented spacetime volume form. We define the symplectic form relative to C by

$$\Omega(\phi, \delta_1\phi, \delta_2\phi) = \int_C \boldsymbol{\omega}(\phi, \delta_1\phi, \delta_2\phi). \tag{44}$$

(More precisely, Eq. (44) defines a "presymplectic form" on field configuration space. As explained in detail in [15], the phase space then is obtained by factoring out by the degeneracy submanifolds of Ω, and Ω then gives rise to a symplectic form on phase space.) If C is noncompact, as we assume here, then some "asymptotic flatness" conditions must be imposed upon the dynamical fields ϕ (and, hence, on their variations) in order to assure convergence of the integral appearing in (44). One normally assumes that the metric g_{ab} approaches a flat metric η_{ab} and the matter fields ψ approach zero at some suitable rate. The precise asymptotic conditions appropriate for a given theory depend upon the details of the theory, and, thus, must be examined on a case-by-case basis, subject to the following general guidelines: the asymptotic fall-off rates of the dynamical fields should be sufficiently rapid that quantities of interest (such as Ω, energy, and angular momentum) are well defined, but not so rapid that a sufficiently wide class of solutions fails to exist. We shall not investigate this issue further here, but will merely assume that such suitable conditions have been imposed.

In principle, the definition of Ω depends upon the choice of C. However, since $d\boldsymbol{\omega} = 0$ whenever $\delta_1\phi$ and $\delta_2\phi$ satisfy the linearized equations of motion [15], the dependence of Ω on C when the equations of motion are imposed is given by an integral of $\boldsymbol{\omega}$ over a timelike surface near spatial infinity. If sufficiently strong asymptotic conditions at spatial infinity have been imposed on the dynamical fields to assure convergence of the integral appearing in (44), then the integral of $\boldsymbol{\omega}$ on this timelike surface typically will vanish, and, thus, the definition of Ω should be independent of C. Of course, if this were not automatically the case, one presumably would strengthen the asymptotic conditions imposed upon the dynamical fields in order to make Ω be independent of C.

As noted above, $\boldsymbol{\Theta}$ is ambiguous up to the terms given in (42). The term involving $\boldsymbol{\mu}$ does not contribute to $\boldsymbol{\omega}$ or Ω, so we find that the only ambiguity in Ω is

$$\Omega \to \Omega + \Delta\Omega \tag{45}$$

with

$$\Delta\Omega = \int_\infty [\delta_1\mathbf{Y}(\phi, \delta_2\phi) - \delta_2\mathbf{Y}(\phi, \delta_1\phi)], \tag{46}$$

where the integral is taken over an $(n-2)$-sphere at spatial infinity. It appears that the asymptotic conditions on the dynamical fields needed to ensure the vanishing of $\Delta\Omega$ typically will be weaker than the conditions needed to ensure that Ω, Eq. (44), is well defined for a given choice of $\boldsymbol{\Theta}$. In particular, taking account of the difficulties in constructing a covariant $(n-2)$-form, $\mathbf{Y}$, out of the metric and its first variation (as well as $\boldsymbol{\epsilon}$), we see that for a theory in spacetime dimension n in which no matter fields are present, the asymptotic conditions

$$g_{ab} \sim \eta_{ab} + o(r^{-\frac{(n-3)}{2}}) \tag{47}$$

together with faster fall-off conditions on derivatives of the metric, suffice to ensure that $\Delta\Omega = 0$. Thus, it does not appear that the ambiguity in $\boldsymbol{\Theta}$ will typically give rise to an ambiguity in the definition of Ω for suitable asymptotic conditions on the dynamical fields.

IV. THE FORM OF THE NOETHER CHARGE

In this section we will obtain an expression for the general structure of the Noether charge $(n-2)$-form, $\mathbf{Q}$, for a diffeomorphism invariant theory. We begin by reviewing the construction of the Noether charge given in [6] (see also [17]).

Let ξ^a be any smooth vector field on the spacetime manifold M (i.e., ξ^a is the infinitesimal generator of a diffeomorphism) and let ϕ be any field configuration. (ϕ is *not* required, at this stage, to be a solution of the equations of motion.) We associate to ξ^a and ϕ a Noether current $(n-1)$-form, defined by

$$\mathbf{J} = \boldsymbol{\Theta}(\phi, \mathcal{L}_\xi\phi) - \xi \cdot \mathbf{L}, \tag{48}$$

where $\boldsymbol{\Theta}(\phi, \mathcal{L}_\xi\phi)$ denotes the expression obtained by replacing $\delta\phi$ with $\mathcal{L}_\xi\phi$ in the expression for $\boldsymbol{\Theta}$, and the "centered dot" denotes the contraction of the vector field ξ^a into the first index of the differential form $\mathbf{L}$. A standard calculation (see, e.g., [15]) gives

$$d\mathbf{J} = -\mathbf{E}\mathcal{L}_\xi\phi \tag{49}$$

which shows $\mathbf{J}$ is closed (for all ξ^a) when the equations of motion are satisfied. Consequently [14] there is a $\mathbf{Q}$ locally constructed from ϕ and ξ^a such that whenever ϕ satisfies the equations of motion, $\mathbf{E} = 0$, we have

$$\mathbf{J} = d\mathbf{Q}. \tag{50}$$

We refer to $\mathbf{Q}$ as the Noether charge $(n-2)$-form. Note that for a given $\mathbf{J}$, Eq. (50) determines $\mathbf{Q}$ uniquely up to the addition of a closed (and, hence, exact [14]) $(n-2)$-form.

Proposition 4.1. *The Noether charge $(n-2)$-form can always be expressed in the form*

$$\mathbf{Q} = \mathbf{W}_c(\phi)\xi^c + \mathbf{X}^{cd}(\phi)\nabla_{[c}\xi_{d]} + \mathbf{Y}(\phi, \mathcal{L}_\xi\phi) + d\mathbf{Z}(\phi, \xi), \tag{51}$$

where $\mathbf{W}_c$, $\mathbf{X}^{ab}$, $\mathbf{Y}$, and $\mathbf{Z}$ are covariant quantities which are locally constructed from the indicated fields and their derivatives (with $\mathbf{Y}$ linear in $\mathcal{L}_\xi\phi$ and $\mathbf{Z}$ linear in ξ). This

decomposition of $\mathbf{Q}$ *is not unique in the sense that there are many different ways of writing* $\mathbf{Q}$ *in the form (51); i.e.,* $\mathbf{W}_c$, $\mathbf{X}^{ab}$, $\mathbf{Y}$, *and* $\mathbf{Z}$ *are not uniquely determined by* $\mathbf{Q}$ *(see below). However,* $\mathbf{X}^{ab}$ *may be chosen to be*

$$(\mathbf{X}^{cd})_{c_3\cdots c_n} = -E_R^{abcd}\epsilon_{abc_3\cdots c_n}, \tag{52}$$

where E_R^{abcd} *was defined by Eq. (31), and we may choose* $\mathbf{Y} = \mathbf{Z} = 0$.

Proof. We proceed by calculating $\mathbf{Q}$ using the choice of $\mathbf{\Theta}$ given in Lemma 3.1, and using the algorithm for calculating $\mathbf{Q}$ from $\mathbf{J}$ given in Lemma 1 of [14]. For the choice of $\mathbf{\Theta}$ given in Lemma 3.1 we have

$$\mathbf{J} = 2\mathbf{E}_R^{bcd}\nabla_d(\nabla_b\xi_c + \nabla_c\xi_b) + \mathbf{\Theta}'(\phi, \mathcal{L}_\xi\phi) - \xi\cdot\mathbf{L}. \tag{53}$$

Now the algorithm of Lemma 1 of [14] for obtaining $\mathbf{Q}$ from $\mathbf{J}$ reduces the highest number of derivatives of ξ^a appearing in the expression for $\mathbf{J}$ by one. Since $\mathbf{\Theta}'$ is linear in the quantities $(\delta g_{ab}, \delta R_{abcd}, \delta\nabla R_{abcd}, ...)$ and does not contain any terms involving derivatives of these quantities, it follows that $\mathbf{\Theta}'(\phi, \mathcal{L}_\xi\phi)$ is linear in $(\xi^a, \nabla_b\xi^a)$; i.e., it has no dependence on any derivatives of ξ^a higher than first. Since $\mathbf{E}_R^{bcd}$ is antisymmetric in c and d, the term $\mathbf{E}_R^{bcd}\nabla_d\nabla_c\xi_b$ has no dependence on derivatives of ξ^a. Thus, no derivatives of ξ^a higher than second appear in Eq. (53), and only the term $\mathbf{E}_R^{bcd}\nabla_d\nabla_b\xi_c$ involves second derivatives of ξ^a. The contribution of this latter term to $\mathbf{Q}$ is readily computed, and we find that with our choice of $\mathbf{\Theta}$ and algorithm for calculating $\mathbf{Q}$, we have

$$\mathbf{Q} = \mathbf{W}_c(\phi)\xi^c + \mathbf{X}^{cd}\nabla_{[c}\xi_{d]}, \tag{54}$$

where $\mathbf{W}$ is a covariant $(n-2)$-form locally constructed out of the dynamical fields, ϕ, and their derivatives, and where

$$(\mathbf{X}^{cd})_{c_3\cdots c_n} = -E_R^{abcd}\epsilon_{abc_3\cdots c_n}. \tag{55}$$

Equation (54) gives the general form of $\mathbf{Q}$ for our particular algorithm for choosing $\mathbf{\Theta}$ and obtaining $\mathbf{Q}$ from $\mathbf{J}$. Recall that $\mathbf{\Theta}$ had two ambiguities (42), one arising from the ambiguity in $\mathbf{L}$ and the other from its defining equation (20). Using the identity

$$\mathcal{L}_\xi\boldsymbol{\mu} = \xi\cdot d\boldsymbol{\mu} + d(\xi\cdot\boldsymbol{\mu}) \tag{56}$$

we see that the ambiguity in $\mathbf{\Theta}$ gives rise to the following ambiguity in $\mathbf{J}$:

$$\mathbf{J} \to \mathbf{J} + d(\xi\cdot\boldsymbol{\mu}) + d\mathbf{Y}(\phi, \mathcal{L}_\xi\phi). \tag{57}$$

Taking into account the additional ambiguity of addition of an exact form to $\mathbf{Q}$, we obtain the following ambiguity in $\mathbf{Q}$ [8]:

$$\mathbf{Q} \to \mathbf{Q} + \xi\cdot\boldsymbol{\mu} + \mathbf{Y}(\phi, \mathcal{L}_\xi\phi) + d\mathbf{Z}. \tag{58}$$

Thus, for any choice of $\mathbf{Q}$, we have

$$\mathbf{Q} = \mathbf{W}_c\xi^c + \mathbf{X}^{cd}\nabla_{[c}\xi_{d]} + \mathbf{Y}(\phi, \mathcal{L}_\xi\phi) + d\mathbf{Z}, \tag{59}$$

as we desired to show. □

As stated above, the decomposition (51) of $\mathbf{Q}$ is *not* unique. For example, it is clear that for any choice of $(n-2)$-form $\mathbf{U}_c(\phi)$, the quantity $d[\mathbf{U}_c(\phi)\xi^c]$ can be written as a sum of terms of the same form as the first three terms on the right side of (51), since we can write it as a sum of a term linear in ξ^c, a term linear in $\nabla_{[c}\xi_{d]}$, and a term linear in $2\nabla_{(c}\xi_{d)} = \mathcal{L}_\xi g_{cd}$. Thus, we can always add the term $\mathbf{U}_c(\phi)\xi^c$ to $\mathbf{Z}$ and make compensating changes in $\mathbf{W}$, $\mathbf{X}$, and $\mathbf{Y}$ without affecting $\mathbf{Q}$. One might be tempted to impose additional conditions to determine the terms $\mathbf{W}$, $\mathbf{X}$, $\mathbf{Y}$, and $\mathbf{Z}$ in Eq. (51). In particular, it might appear natural to fix the term $\mathbf{X}^{ab}$ (which plays a key role in the definition of black hole entropy below) by simply requiring it to be given by Eq. (55) above. However, this proposal suffers from the difficulty that a change of Lagrangian of the form (40), which should have no effect upon the physical content of the theory, would, in general produce in a change in $\mathbf{X}^{ab}$. For this reason we shall not attempt to give unique definitions of the individual terms in Eq. (51) but will derive the first law of black hole mechanics and give a proposal for defining the entropy of dynamical black holes based only upon the general form of $\mathbf{Q}$ given in Proposition 4.1.

V. EXAMPLES OF LAGRANGIANS AND ASSOCIATED NOETHER CURRENTS AND CHARGES

In this section we shall give the symplectic potential $\mathbf{\Theta}$, the Noether current $\mathbf{J}$, and the Noether charge $\mathbf{Q}$ arising from three Lagrangians of interest. In giving these examples, we shall simply make convenient choices of $\mathbf{\Theta}$, $\mathbf{J}$, and $\mathbf{Q}$, but, of course, it should be kept in mind that the ambiguities (42), (57), and (58) remain present.

Our first example is general relativity. We have the Lagrangian four-form

$$\mathbf{L}_{abcd} = \frac{1}{16\pi}\epsilon_{abcd}R. \tag{60}$$

This yields a symplectic potential three-form

$$\mathbf{\Theta}_{abc} = \epsilon_{dabc}\frac{1}{16\pi}g^{de}g^{fh}\left(\nabla_f\delta g_{eh} - \nabla_e\delta g_{fh}\right). \tag{61}$$

From this we obtain the Noether current three-form

$$\mathbf{J}_{abc} = \frac{1}{8\pi}\epsilon_{dabc}\nabla_e\left(\nabla^{[e}\xi^{d]}\right) + \frac{1}{8\pi}\epsilon_{dabc}G^d{}_e\xi^e. \tag{62}$$

The second term on the right side of this equation vanishes when the field equations ($G_{ab} = 0$) hold. The corresponding Noether charge two-form is

$$\mathbf{Q}_{ab} = -\frac{1}{16\pi}\epsilon_{abcd}\nabla^c\xi^d. \tag{63}$$

Our second example is two-dimensional dilaton gravity (in the form given in [3]), with scalar field ϕ, coupling constant λ, and an additional "tachyon field" T. The Lagrangian two-form is

$$\mathbf{L}_{ab} = \tfrac{1}{2}\epsilon_{ab}e^\phi\left[R + (\nabla\phi)^2 - (\nabla T)^2 + \mu^2T^2 + \lambda\right]. \tag{64}$$

This yields the symplectic potential one-form

$$\Theta_a = \epsilon_{ab} e^{\phi} \left\{ (\nabla^b \phi)\delta\phi - (\nabla^b T)\delta T + \tfrac{1}{2} g^{bc} [\nabla^d (\delta g_{cd}) - g^{de} \nabla_c (\delta g_{de}) - (\nabla^d \phi)\delta g_{cd} + (\nabla_c \phi) g^{de} \delta g_{de}] \right\} . \tag{65}$$

From this we obtain (omitting terms proportional to the field equations) the Noether current one-form

$$\mathbf{J}_a = \epsilon_{ab} \nabla_c \left(e^{\phi} \nabla^{[c} \xi^{b]} + 2\xi^{[c} \nabla^{b]} e^{\phi} \right) , \tag{66}$$

which yields the Noether charge zero-form (i.e., function)

$$Q = -\tfrac{1}{2} \epsilon_{ab} \left(e^{\phi} \nabla^a \xi^b + 2\xi^a \nabla^b e^{\phi} \right) . \tag{67}$$

As our final example, we consider the special case of Lovelock gravity in n dimensions obtained by keeping only the terms in the Lagrangian up to quadratic order in the curvature (see [18]). The Lagrangian n-form is

$$\mathbf{L}_{a_1 \cdots a_n} = \epsilon_{a_1 \cdots a_n} \left(\frac{1}{16\pi} R + \alpha (R_{abcd} R^{abcd} - 4 R_{ab} R^{ab} + R^2) \right) . \tag{68}$$

This yields a symplectic potential $(n-1)$-form

$$\begin{aligned} \mathbf{\Theta}_{a_1 \cdots a_{n-1}} = \epsilon_{d a_1 \cdots a_{n-1}} \Bigg[& \left(\frac{1}{16\pi} + 2\alpha R \right) g^{de} g^{fh} (\nabla_f \delta g_{eh} - \nabla_e \delta g_{fh}) \\ & + \alpha [-2(\nabla^e R) g^{df} \delta g_{ef} + 4 R^{de} (\nabla_e \delta g_{fh}) g^{fh} + 4 R^{ef} (\nabla^d \delta g_{ef}) \\ & - 8 R^{ef} (\nabla_e \delta g_{fh}) g^{dh} - 4 (\nabla^e R^{df}) \delta g_{ef} + 4 R^{defh} \nabla_h \delta g_{ef}] \Bigg] . \end{aligned} \tag{69}$$

The corresponding Noether current $(n-1)$-form (again, omitting terms proportional to the field equations) is

$$\mathbf{J}_{a_1 \cdots a_{n-1}} = \epsilon_{d a_1 \cdots a_{n-1}} \nabla_e \left[\left(\frac{1}{8\pi} + 4\alpha R \right) \nabla^{[e} \xi^{d]} + 16\alpha (\nabla_f \xi^{[e}) R^{d]f} + 4\alpha R^{edfh} \nabla_f \xi_h \right] . \tag{70}$$

This yields the Noether charge $(n-2)$-form

$$\mathbf{Q}_{a_1 \cdots a_{n-2}} = -\epsilon_{dea_1 \cdots a_{n-2}} \left(\frac{1}{16\pi} \nabla^d \xi^e + 2\alpha (R \nabla^d \xi^e + 4 \nabla^{[f} \xi^{d]} R^e{}_f + R^{defh} \nabla_f \xi_h) \right) . \tag{71}$$

VI. THE FIRST LAW OF BLACK HOLE MECHANICS

In this section we will use Lemma 3.1 and Proposition 4.1 to improve upon the derivation of the first law of black hole mechanics given in [6]. We thereby will prove that the first law of black hole mechanics holds for non-stationary perturbations of a black hole in an arbitrary diffeomorphism covariant theory of gravity, without any restriction on the number of derivatives of fields which appear in the Lagrangian.

Let ϕ be any solution of the equations of motion, and let $\delta\phi$ be any variation of the dynamical fields (not necessarily satisfying the linearized equations of motion) about ϕ. Let ξ^a be an arbitrary, fixed vector field on M. We then have [6]

$$\begin{aligned} \delta \mathbf{J} &= \delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \xi \cdot \delta \mathbf{L} \\ &= \delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \xi \cdot d \mathbf{\Theta}(\phi, \delta\phi) \\ &= \delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \mathcal{L}_\xi \mathbf{\Theta}(\phi, \delta\phi) + d[\xi \cdot \mathbf{\Theta}(\phi, \delta\phi)], \end{aligned} \tag{72}$$

where Eq. (20) together with $\mathbf{E} = 0$ was used in the second line, and the identity (56) on Lie derivatives of forms was used in the last line. Since our choice of $\mathbf{\Theta}$ is covariant, $\mathcal{L}_\xi \mathbf{\Theta}$ is the same as the variation induced in $\mathbf{\Theta}$ by the field variation $\delta'\phi = \mathcal{L}_\xi \phi$. Consequently, we have

$$\delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \mathcal{L}_\xi \mathbf{\Theta}(\phi, \delta\phi) = \boldsymbol{\omega}(\phi, \delta\phi, \mathcal{L}_\xi \phi), \tag{73}$$

where $\boldsymbol{\omega}$ was defined by Eq. (43). We therefore obtain

$$\boldsymbol{\omega}(\phi, \delta\phi, \mathcal{L}_\xi \phi) = \delta \mathbf{J} - d(\xi \cdot \mathbf{\Theta}). \tag{74}$$

The fundamental identity which gives rise to the first law of black hole mechanics applies to the case where ξ^a is a symmetry of all of the dynamical fields, i.e, $\mathcal{L}_\xi \phi = 0$, and $\delta\phi$ satisfies the linearized equations of motion. When $\mathcal{L}_\xi \phi = 0$, the left side of Eq. (74) vanishes, and when $\delta\phi$ satisfies the linearized equations, we may replace $\delta \mathbf{J}$ by $\delta d \mathbf{Q} = d \delta \mathbf{Q}$ on the right side. Thus, we obtain

$$d \delta \mathbf{Q} - d(\xi \cdot \mathbf{\Theta}) = 0. \tag{75}$$

Integrating this equation over a hypersurface Ξ we obtain

$$\int_{\partial \Xi} \delta \mathbf{Q}[\xi] - \xi \cdot \mathbf{\Theta}(\phi, \delta\phi) = 0. \tag{76}$$

We emphasize that the only conditions needed for the validity of Eq. (76) are that ϕ be a solution to the equations of motion, $\mathbf{E} = 0$, satisfying $\mathcal{L}_\xi \phi = 0$, and $\delta\phi$ be a solution of the linearized equations (not necessarily satisfying $\mathcal{L}_\xi \delta\phi = 0$).

We shall be interested here in the case where Ξ is an asymptotically flat hypersurface in an asymptotically flat spacetime. In this case, a boundary term from an asymptotic $(n-2)$-sphere at infinity will contribute to Eq. (76). The following argument shows that this boundary term has the natural interpretation of being the variation of the "conserved quantity" canonically conjugate to the asymptotic symmetry generated by ξ^a.

Consider a solution, ϕ, corresponding to an asymptotically flat, globally hyperbolic spacetime, with Cauchy surface, C, having a single asymptotic region and a compact interior. We return to Eq. (74) but no longer impose the additional assumptions that $\mathcal{L}_\xi \phi = 0$ or that $\delta\phi$ satisfy the linearized equations of motion. We integrate Eq. (74) over C taking into account Eq. (44) and the fact that, by definition, Hamilton's equations of motion for the dynamics generated by the time evolution vector field ξ^a are

$$\delta H = \Omega(\phi, \delta\phi, \mathcal{L}_\xi \phi). \tag{77}$$

We thereby find that *if* a Hamiltonian H exists for the dynamics generated by ξ^a, then

$$\begin{aligned} \delta H &= \delta \int_C \mathbf{J} - \int_C d(\xi \cdot \boldsymbol{\Theta}) \\ &= \delta \int_C \mathbf{J} - \int_\infty \xi \cdot \boldsymbol{\Theta}. \end{aligned} \tag{78}$$

Thus, a Hamiltonian for the dynamics generated by ξ^a does exist if (and only if) we can find a (not necessarily diffeomorphism covariant) $(n-1)$-form $\mathbf{B}$ such that

$$\delta \int_\infty \xi \cdot \mathbf{B} = \int_\infty \xi \cdot \boldsymbol{\Theta} \tag{79}$$

in which case H is given by

$$H = \int_C \mathbf{J} - \int_\infty \xi \cdot \mathbf{B}. \tag{80}$$

Now evaluate H on solutions. We then may replace $\mathbf{J}$ by $d\mathbf{Q}$, whence H becomes

$$H = \int_\infty (\mathbf{Q} - \xi \cdot \mathbf{B}). \tag{81}$$

Thus, we have shown that in any theory arising from a diffeomorphism covariant Lagrangian, the Hamiltonian, if it exists, always is a pure "surface term" when evaluated "on shell." Similarly, for a closed universe (i.e., compact C), the Hamiltonian always vanishes "on shell."

We now shall assume that the asymptotic conditions on the dynamical fields have been specified in such a way that when ξ^a is an asymptotic time translation, $\mathbf{B}$ exists, and the surface integrals appearing in Eq. (81) approach a finite limit at infinity. We define the *canonical energy* $\mathcal{E}$ to be the value of the Hamiltonian: i.e.,

$$\mathcal{E} \equiv \int_\infty (\mathbf{Q}[t] - t \cdot \mathbf{B}), \tag{82}$$

where t^a is an asymptotic time translation. We then adopt Eq. (82) as the definition of the canonical energy associated to any asymptotically flat region of any solution, whether or not the spacetime is globally hyperbolic.

We illustrate this definition of canonical energy by evaluating $\mathcal{E}$ for vacuum general relativity. We consider spacetimes which are asymptotically flat in the sense that there exists a flat metric η_{ab} such that in a global inertial coordinate system of η_{ab} we have

$$g_{\mu\nu} = \eta_{\mu\nu} + O(1/r) \tag{83}$$

and

$$\frac{\partial g_{\mu\nu}}{\partial x^\alpha} = O(1/r^2). \tag{84}$$

Let t^a be the asymptotic time translation $(\partial/\partial t)^a$, and let the 2-sphere at infinity be the limit as $r \to \infty$ of the coordinate spheres r, t =const. Then from our previously calculated expression for $\mathbf{Q}_{ab}$, Eq. (63), we find that

$$\begin{aligned} \int_\infty \mathbf{Q}[t] &= -\frac{1}{16\pi} \int_\infty \epsilon_{abcd} \nabla^c t^d \\ &= -\frac{1}{16\pi} \int_\infty dS \left(\frac{\partial g_{tt}}{\partial r} - \frac{\partial g_{rt}}{\partial t} \right). \end{aligned} \tag{85}$$

Note that for a stationary spacetime with stationary Killing field t^a, the first line of Eq. (85) shows that $\int_\infty \mathbf{Q}[t]$ is precisely one-half of the Komar mass (see, e.g., [12]).

We now compute the contribution to $\mathcal{E}$ from the second term on the right side of Eq. (82). Using Eq. (61) we have

$$\begin{aligned} \int_\infty t^a \boldsymbol{\Theta}_{abc} &= -\frac{1}{16\pi} \int_\infty dS r_d g^{de} g^{fh} (\nabla_f \delta g_{eh} - \nabla_e \delta g_{fh}) \\ &= -\frac{1}{16\pi} \int_\infty dS g^{rr} \left[g^{tt} (\partial_t \delta g_{rt} - \partial_r \delta g_{tt}) + h^{ij} (\partial_i \delta h_{rj} - \partial_r \delta h_{ij}) \right] \\ &= -\frac{1}{16\pi} \delta \int_\infty dS \left[(\partial_r g_{tt} - \partial_t g_{rt}) + r^k h^{ij} (\partial_i h_{kj} - \partial_k h_{ij}) \right], \end{aligned} \tag{86}$$

where $r^a = (\partial/\partial r)^a$ and h_{ij} is the spatial metric. Thus, we see that Eq. (79) holds if $\mathbf{B}$ is chosen to be any three-form such that asymptotically at infinity, we have

$$t^a \mathbf{B}_{abc} = -\frac{1}{16\pi} \tilde{\epsilon}_{bc} \left[(\partial_r g_{tt} - \partial_t g_{rt}) + r^k h^{ij} (\partial_i h_{kj} - \partial_k h_{ij}) \right], \tag{87}$$

where $\tilde{\epsilon}_{bc}$ is the volume two-form for the sphere at infinity. Combining this with (85) we find that the canonical energy $\mathcal{E}$ for general relativity is

$$\begin{aligned}\mathcal{E} &= \int_\infty (\mathbf{Q}[t] - t\cdot\mathbf{B}) \\ &= \frac{1}{16\pi}\int_\infty dS r^k h^{ij}(\partial_i h_{kj} - \partial_k h_{ij}) \\ &= M_{\mathrm{ADM}},\end{aligned} \tag{88}$$

where M_{ADM} denotes the Arnowitt-Deser-Misner (ADM) mass. Thus, the term $t\cdot\mathbf{B}$ cancels the contribution to $\mathcal{E}$ from the term $\mathbf{Q}$, and, in addition, provides the term M_{ADM}, thereby making our definition of $\mathcal{E}$ in vacuum general relativity reduce to the standard, ADM, definition of energy. Note, however, that additional contributions to $\mathcal{E}$ in general relativity can occur when long-range matter fields are present; see [1] for an explicit evaluation of the contribution to $\mathcal{E}$ for Yang-Mills fields.

When ξ^a is an asymptotic rotation φ^a we may choose the surface at infinity to be everywhere tangent to φ^a, in which case the pullback of $\varphi\cdot\boldsymbol{\Theta}$ to that surface vanishes. Hence, we define the *canonical angular momentum*, $\mathcal{J}$ of any asymptotic region by

$$\mathcal{J} = -\int_\infty \mathbf{Q}[\varphi], \tag{89}$$

where it is assumed that the asymptotic conditions on the dynamical fields are such that this surface integral approaches a well-defined limit at infinity. [The relative sign difference occurring in the definitions (82) and (89) traces its origin to the Lorentz signature of the spacetime metric. The same relative sign difference occurs in the definitions, $E = -p_a t^a$ and $J = +p_a\varphi^a$, of the energy and angular momentum of a particle in special relativity.] In the axisymmetric case in vacuum general relativity, Eq. (89) is precisely the Komar formula for angular momentum. Thus, we see that in any theory, the Komar-type expression $-\int_\infty \mathbf{Q}[\varphi]$ always yields the angular momentum, but $\int_\infty \mathbf{Q}[t]$ does not, in general, yield the energy. Indeed, since the Komar and ADM masses agree for stationary solutions in general relativity [19], we see from our calculation above that $\int_\infty \mathbf{Q}[t]$ yields only half of the energy in that case. It is the presence of the "extra" $t\cdot\mathbf{B}$ term in Eq. (82) which accounts for this well-known "factor of 2" discrepancy in the Komar formulas for mass and angular momentum in general relativity.

We now are ready to apply Eq. (76) to the case of a stationary black hole solution with bifurcate Killing horizon. Let ξ^a be the Killing field which vanishes on the bifurcation $(n-2)$-surface Σ, normalized so that

$$\xi^a = t^a + \Omega_H^{(\mu)}\varphi^a_{(\mu)}, \tag{90}$$

where t^a is the stationary Killing field with unit norm at infinity, and summation over μ is understood. (This equation picks out a family of axial Killing fields, $\varphi^a_{(\mu)}$, acting in orthogonal planes, and also defines the "angular velocities of the horizon," $\Omega_H^{(\mu)}$. No summation is required when the spacetime dimension is less than five, and, of course, the second term on the right side is entirely absent in two dimensions.) Let Ξ be an asymptotically flat hypersurface having Σ as its only "interior boundary." Then, taking into account Eq. (90), the definitions of $\mathcal{E}$ and $\mathcal{J}$, and the fact that ξ vanishes on Σ, we obtain directly from Eq. (76) the result

$$\delta\int_\Sigma \mathbf{Q}[\xi] = \delta\mathcal{E} - \Omega_H^{(\mu)}\delta\mathcal{J}_{(\mu)}. \tag{91}$$

We now are ready to state and prove the first law of black hole mechanics in a form which strengthens the results of [6] by establishing the general validity of this law for nonstationary perturbations.

Theorem 6.1. Let ϕ be an asymptotically flat stationary black hole solution with a bifurcate Killing horizon, and let $\delta\phi$ be a (not necessarily stationary), asymptotically flat solution of the linearized equations about ϕ. Define S as,

$$S = 2\pi\int_\Sigma \mathbf{X}^{cd}\boldsymbol{\epsilon}_{cd}, \tag{92}$$

where $\mathbf{X}^{cd}$ is as given in Proposition 4.1, and the integral is taken over the bifurcation $(n-2)$-surface Σ, with $\boldsymbol{\epsilon}_{cd}$ denoting the binormal to Σ (i.e., $\boldsymbol{\epsilon}$ is the natural volume element on the tangent space perpendicular to Σ, oriented so that $\boldsymbol{\epsilon}_{cd}T^cR^d > 0$ when T^a is a future-directed timelike vector and the spacelike vector R^a points "toward infinity"). Then we have

$$\frac{\kappa}{2\pi}\delta S = \delta\mathcal{E} - \Omega_H^{(\mu)}\delta\mathcal{J}_{(\mu)}, \tag{93}$$

where κ is the surface gravity of the black hole.

Proof. The theorem will follow from Eq. (91) provided that we can show that

$$\delta\int_\Sigma \mathbf{Q}[\xi] = \frac{\kappa}{2\pi}\delta S. \tag{94}$$

To evaluate the left side of this equation, we appeal to Proposition 4.1 and examine the contribution of each of the four terms individually. Since ξ vanishes on Σ, it is clear that the term $\mathbf{W}_c\xi^c$ contributes neither to $\mathbf{Q}$ nor to its variation. Similarly, the term $d\mathbf{Z}$ clearly also makes no contribution to the left side of Eq. (94). Since $\mathcal{L}_\xi\phi = 0$, the term $\mathbf{Y}$ vanishes in the stationary background, and its first variation is given by

$$\begin{aligned}\delta\mathbf{Y}(\phi, \mathcal{L}_\xi\phi) &= \mathbf{Y}(\phi, \mathcal{L}_\xi\delta\phi) \\ &= \mathcal{L}_\xi\mathbf{Y}(\phi,\delta\phi) \\ &= \xi\cdot d\mathbf{Y} + d(\xi\cdot\mathbf{Y}),\end{aligned} \tag{95}$$

where the Lie derivative identity (56) was used in the last line. It follows immediately that the term $\mathbf{Y}$ also makes no contribution to the left side of Eq. (94). Thus, we have

$$\delta\int_\Sigma \mathbf{Q}[\xi] = \delta\int_\Sigma \mathbf{X}^{cd}(\phi)\nabla_{[c}\xi_{d]}. \tag{96}$$

Now, in the stationary background, we have, on Σ,

$$\nabla_c \xi_d = \kappa \epsilon_{cd}. \tag{97}$$

Furthermore, since $\xi^a = 0$ on Σ, and $\delta\xi^a = 0$ everywhere, we have

$$\delta \nabla_c \xi^d = 0 \tag{98}$$

on Σ. Consider, now, the variation, $\delta\epsilon_c{}^d$ of the binormal, $\epsilon_c{}^d$, with an index raised. Clearly, $s^c \delta\epsilon_c{}^d = 0$ for all s^c tangent to Σ, so $\delta\epsilon_c{}^d$ has no "tangential-tangential" piece. However, since $\epsilon_c{}^d \epsilon_d{}^c$ does not vary as the metric is changed, it follows that $g_{a[c}\delta\epsilon_{d]}{}^a$ has no "normal-normal" piece with respect to the background metric. Thus, writing

$$w_{cd} = \nabla_{[c}\xi_{d]} - \kappa\epsilon_{cd} \tag{99}$$

we have that w_{cd} vanishes in the stationary background, and

$$\begin{aligned} \delta w_{cd} &= \delta\left[g_{a[d}(\nabla_{c]}\xi^a - \kappa\epsilon_{c]}{}^a)\right] \\ &= -\kappa g_{a[d}\delta\epsilon_{c]}{}^a \end{aligned} \tag{100}$$

so that δw_{cd} has only a "normal-tangential" piece with respect to the background metric. Thus, substituting in Eq. (96) we find

$$\begin{aligned} \delta \int_\Sigma \mathbf{Q}[\xi] &= \delta \int_\Sigma \mathbf{X}^{cd}(\phi)[\kappa\epsilon_{cd} + w_{cd}] \\ &= \frac{\kappa}{2\pi}\delta S + \int_\Sigma \mathbf{X}^{cd}\delta w_{cd}. \end{aligned} \tag{101}$$

Finally, we note that since $\mathcal{L}_\xi \phi = 0$, we have $\mathcal{L}_\xi \mathbf{X}^{cd} = 0$, and, hence, by Lemma 2.3 of [20], at each point of Σ, $\mathbf{X}^{cd}$ must be invariant under "reflections" about Σ; i.e., $\mathbf{X}^{cd}$ must be invariant under the map of the tangent space which reverses the normal directions to Σ but keeps the tangential directions unchanged. On the other hand, since δw_{cd} is purely "normal-tangential," it reverses sign under reflections about Σ. However, the pullback of $\mathbf{X}^{cd}\delta w_{cd}$ to Σ is purely tangential, and, hence, invariant under reflections. Consequently, the pullback of $\mathbf{X}^{cd}\delta w_{cd}$ to Σ must vanish, so the second term on the right side of Eq. (101) does not contribute. □

Note that Eq. (92) corresponds to the following simple algorithm for determining the entropy of a stationary black hole in an arbitrary theory of gravity: Start with the Lagrangian n-form (3) and contract it with $(-1/n!)\epsilon^{a_1\cdots a_n}$ to obtain a scalar L. Take the functional derivative of L with respect to R_{abcd} (viewing it as a field independent of g_{ab}) to obtain the tensor field E_R^{abcd}. Then we have

$$S = -2\pi \int_\Sigma E_R^{abcd}\epsilon_{ab}\epsilon_{cd}, \tag{102}$$

where ϵ_{ab} again denotes the binormal to Σ, and the integral is taken with respect to the natural, induced volume element on Σ.

It should be noted that in the above discussion, Σ was explicitly chosen to be the bifurcation surface of a bifurcate Killing horizon. However, as pointed out in [8], for a stationary black hole with bifurcate horizon, the integral of $\mathbf{Q}$ is independent of the choice of horizon cross section. Namely, the difference between the integrals of $\mathbf{Q}$ over cross sections Σ and Σ' is given by an integral of $\mathbf{J}$ over the intervening portion of the horizon. However, by Eq. (48), the pullback of $\mathbf{J}$ to the horizon vanishes, since $\mathcal{L}_\xi\phi = 0$ and the pullback of $\xi \cdot \mathbf{L}$ vanishes since ξ^a is tangent to the horizon.

Furthermore, if we define the entropy, S, for an arbitrary horizon cross section, Σ', of a stationary black hole by

$$S[\Sigma'] = 2\pi \int_{\Sigma'} \mathbf{X}^{cd}\epsilon'_{cd}, \tag{103}$$

where ϵ'_{cd} denotes the binormal to Σ', then S also is independent of the choice of Σ' [8]. To prove this, we note that since $\mathbf{X}^{cd}$ is invariant under the one-parameter group of isometries, χ_t, generated by ξ^a, it follows immediately that $S[\chi_t(\Sigma')] = S[\Sigma']$. However, as $t \to -\infty$, $\chi_t(\Sigma')$ continuously approaches the bifurcation surface Σ and (since $\mathbf{X}^{cd}$ is smooth) we thus obtain $S[\Sigma'] = S[\Sigma]$, as we desired to show. It follows immediately that for stationary perturbations, the first law of black hole mechanics (93) holds with S taken to be the entropy of an arbitrary cross section of the horizon. However, when nonstationary perturbations are considered, it is essential for the validity of Eq. (93) that S be evaluated on the bifurcation surface Σ.

As emphasized at the end of Sec. IV, the decomposition of $\mathbf{Q}$ given by Eq. (51) does not uniquely determine $\mathbf{X}^{cd}$. Nevertheless, Theorem 6.1 and its proof show that all of the different possible choices of $\mathbf{X}^{cd}$ yield the same value of the entropy, S, for a stationary black hole. Furthermore, even for nonstationary perturbations, the first variation δS of S on Σ is independent of the choice of $\mathbf{X}^{cd}$. However, for nonstationary perturbations, δS will, in general, depend upon the choice of $\mathbf{X}^{cd}$ when evaluated on an arbitrary cross section Σ' of the horizon, and the dependence of S upon the choice of $\mathbf{X}^{cd}$ becomes even more severe if we attempt to generalize the notion of entropy to an arbitrary cross section of a nonstationary black hole via Eq. (103). We turn now to an analysis of the definition of entropy for nonstationary black holes.

VII. A PRESCRIPTION FOR DYNAMICAL BLACK HOLE ENTROPY

In this section we will suggest a definition of the entropy $S_{\rm dyn}$ for a "dynamical" (i.e., nonstationary) black hole. We seek a formula of the general type

$$S_{\rm dyn}[\mathcal{C}] = \int_{\mathcal{C}} \tilde{\mathbf{X}}^{cd}(\phi)\epsilon_{cd}, \tag{104}$$

where $\mathcal{C}$ is an arbitrary cross section of the event horizon of a dynamical black hole, and $\tilde{\mathbf{X}}^{cd}$ is a diffeomorphism covariant $(n-2)$-form locally constructed out of the dynamical fields, ϕ, and their derivatives by an algorithm whose sole input is the Lagrangian, $\mathbf{L}$. There are four basic criteria which our definition of $S_{\rm dyn}$ must satisfy.

(1) For an arbitrary cross section Σ' of a stationary

black hole we must have

$$S_{\rm dyn}[\Sigma'] = S[\Sigma'] = 2\pi \int_{\Sigma'} \mathbf{X}^{cd}\boldsymbol{\epsilon}'_{cd} \tag{105}$$

[see Eq. (103) above].

(2) For an arbitrary (nonstationary) perturbation of a stationary black hole, on the bifurcation surface Σ we must have

$$\delta S_{\rm dyn}[\Sigma] = \delta S = 2\pi\delta \int_{\Sigma} \mathbf{X}^{cd}\boldsymbol{\epsilon}_{cd} \tag{106}$$

[see Eq. (92) above].

(3) If we alter the Lagrangian by the addition of an exact n-form

$$\mathbf{L} \to \mathbf{L} + d\boldsymbol{\mu} \tag{107}$$

then the definition of $S_{\rm dyn}$ should not change, since there is no change in the dynamical content of the theory.

(4) At least for an appropriate class of theories, $S_{\rm dyn}$ should obey a "second law"; i.e., $S_{\rm dyn}$ should be a nondecreasing quantity when evaluated on successively "later" cross sections of the horizon of a dynamical black hole.

The last of these criteria is by far the most interesting and important. Unfortunately, it also is the most difficult to analyze in a general theory of gravity for at least the following two reasons: First, it seems clear that, unlike the "first law," any proof of the second law would need to make detailed use of the equations of motion of the theory. Second, it seems clear that the "second law" should hold only for the case of theories which satisfy certain physically reasonable criteria, likely examples of which are the existence of a well-posed initial value formulation, cosmic censorship, and the property of having positive total energy. For example, even for general relativity, the second law can fail if matter is present which fails to satisfy the weak energy condition. However, it is far from clear as to precisely what conditions should be imposed upon a theory for the validity of the second law to hold, and, even if these conditions were known, it undoubtedly would be highly nontrivial to determine whether a given theory satisfied them.

Despite these two difficulties, there are some hints that it may be possible to prove some general results pertaining to the second law. In particular, we saw in the previous section that the entropy S of a stationary black hole is just its Noether charge with respect to the horizon Killing field ξ^a. Thus, the change in entropy between cross sections $\mathcal{C}$ and $\mathcal{C}'$ of a stationary black hole is given by the flux of the corresponding Noether current through the horizon between $\mathcal{C}$ and $\mathcal{C}'$. For a stationary black hole, this flux, of course, vanishes. However, if $S_{\rm dyn}$ could similarly be identified as the Noether charge of an appropriate vector field, one might be able to establish a relationship between the "second law" and positive energy (i.e., positive net Noether flux) properties of the theory. Another suggestive fact is that the quantity $\mathbf{X}^{cd}$ which plays a key role in the definition of entropy for stationary black holes can be chosen to be very simply related to E_R^{abcd} [see Eq. (52) above], and E_R^{abcd}, in turn, is a term in the equations of motion [see Eq. (36) above]. Thus, there is a hint that it may be possible to define $S_{\rm dyn}$ in such a way that its dynamical properties may be directly related to the equations of motion of the theory. Unfortunately, we have not, as yet, succeeded in developing either of these hints into any results regarding proposed definitions of $S_{\rm dyn}$. Thus, for the remainder of this section, we shall not consider criterion (4) further, and will merely seek a definition of $S_{\rm dyn}$ which satisfies conditions (1)–(3).

An obvious first try at defining $S_{\rm dyn}$ via an equation of the form (104) would be to simply set $\tilde{\mathbf{X}}^{cd} = \mathbf{X}^{cd}$, with $\mathbf{X}^{cd}$ given by the decomposition (59) of $\mathbf{Q}$. However, we already emphasized above that this decomposition is not unique. Although, as discussed at the end of the previous section, this ambiguity does not affect the evaluation of S on an arbitrary cross section of a stationary black hole horizon or the evaluation of δS on the bifurcation surface of a stationary black hole, this ambiguity in $\mathbf{X}^{cd}$ is of importance for a dynamical black hole.

An obvious try at circumventing this difficulty would be to continue to set $\tilde{\mathbf{X}}^{cd} = \mathbf{X}^{cd}$ and simply fix $\mathbf{X}^{cd}$ by some definite algorithm. In particular, the choice

$$\mathbf{X}^{cd}_{a_3\cdots a_n} = -E_R^{abcd}\boldsymbol{\epsilon}_{aba_3\cdots a_n} \tag{108}$$

[see Eq. (52) above] appears to be particularly simple and natural. This proposed definition of $S_{\rm dyn}$ clearly satisfies conditions (1) and (2) above. However, it is not difficult to verify that it fails [11] to satisfy condition (3): By adding an exact form to $\mathbf{L}$ which has suitable dependence upon the curvature, we can alter E_R^{abcd} in such a way as to produce nonvanishing changes in $S_{\rm dyn}$ for nonstationary black holes. We feel that it is unlikely that any other simple algorithm for fixing $\mathbf{X}^{cd}$ for a given $\mathbf{L}$ will fare any better in this regard.

Thus, it is a nontrivial challenge to find *any* prescription for $S_{\rm dyn}$ of the form (104) which satisfies conditions (1)–(3). We now shall demonstrate that such a prescription does exist. The basic idea will be to construct new dynamical fields relative to a cross section $\mathcal{C}$, which make $\mathcal{C}$ "look like" a bifurcation surface of a stationary black hole. We then shall define $S_{\rm dyn}[\mathcal{C}]$ to be the entropy of this stationary black hole. Before giving a precise statement of our prescription, we give the following two definitions.

Definition 7.1 *Let $\mathcal{C}$ be an $(n-2)$-dimensional spacelike surface in an n-dimensional spacetime, and let $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ be a (spacetime) tensor field defined on $\mathcal{C}$. Then $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ is said to be* boost invariant *on $\mathcal{C}$ if, for each $p \in \mathcal{C}$, $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ is invariant under Lorentz boosts in the tangent space at p in the two-dimensional timelike plane orthogonal to $\mathcal{C}$.*

The following simple criterion can be used to check if a tensor field $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ is boost invariant on $\mathcal{C}$. At each point $p \in \mathcal{C}$, choose a null tetrad with null vectors l^a and n^a orthogonal to $\mathcal{C}$, and spacelike vectors s^a_μ tangent to $\mathcal{C}$. Expand $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ in this basis. Then it is easy to verify that $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ is boost invariant if and only if its basis expansion is "balanced" with respect to l^a and n^a, i.e., if the basis expansion coefficients are nonvanishing only for terms involving equal numbers of

l^a's and n^a's. This motivates the following definition.

Definition 7.2 *Let $\mathcal{C}$ be an $(n-2)$-dimensional spacelike surface in an n-dimensional spacetime, and let $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ be a (spacetime) tensor field defined on $\mathcal{C}$. We define the* boost invariant part *of $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ to be the tensor field on $\mathcal{C}$ obtained by keeping only the terms which are balanced with respect to l^a and n^a in a null tetrad basis expansion.*

It is easily seen that the boost invariant part of $M^{a_1\cdots a_k}{}_{b_1\cdots b_l}$ does not depend upon the choice of null tetrad appearing in the definition.

Note that the spacetime metric g_{ab} on $\mathcal{C}$ is automatically boost invariant. However, the curvature of g_{ab} and its derivatives need not be. Nevertheless, we may define a notion of the boost invariant part (up to order q), $g^{I_q}_{ab}$, of the spacetime metric in a neighborhood of $\mathcal{C}$. The curvature of $g^{I_q}_{ab}$ and its covariant derivatives up to order $(q-2)$ then will automatically be boost invariant on $\mathcal{C}$. This construction of $g^{I_q}_{ab}$ will lead directly to a proposal for defining $S_{\rm dyn}$.

To define $g^{I_q}_{ab}$, it is convenient to introduce a coordinate system in a neighborhood of $\mathcal{C}$ as follows [20]. Define a null tetrad l^a, n^a, s^a_μ on $\mathcal{C}$ as above, with $l^a n_a = -1$. Let $\mathcal{O}$ be any neighborhood of $\mathcal{C}$ sufficiently small that each point $x \in \mathcal{O}$ lies on a unique geodesic orthogonal to $\mathcal{C}$. Given $x \in \mathcal{O}$ we find the point $p \in \mathcal{C}$ and the geodesic tangent v^a in the 2-plane normal to $\mathcal{C}$ such that x lies at unit affine parameter along the geodesic determined by p and v^a. We assign the coordinates $(U, V, s_1, \ldots s_{n-2})$ to $x \in \mathcal{O}$ by taking (U, V) to be the components of v^a along l^a and n^a, respectively, and taking s_i to be (arbitrarily chosen) coordinates of p on $\mathcal{C}$. We denote by ∂_a the flat derivative operator associated with these coordinates. Note that a change in tetrad, $l^a \to \alpha l^a$, $n^a \to \alpha^{-1} n^a$, at $p \in \mathcal{C}$ (corresponding to a Lorentz boost in the tangent space in the plane orthogonal to $\mathcal{C}$) induces the linear change in coordinates, $U \to \alpha^{-1}U$, $V \to \alpha V$, $s_i \to s_i$. Since linearly related coordinate systems define the same "ordinary derivative operator," it follows that ∂_a does not depend upon the choice of l^a and n^a, and so is invariant under the action of Lorentz boosts in the plane orthogonal to $\mathcal{C}$.

Now consider the first q terms in the Taylor series expansion of g_{ab} around $\mathcal{C}$ in U and V:

$$g^{(q)}_{ab}(x^\mu) = \sum_{n,m=0}^{q} \frac{U^m V^n}{m!n!} \sum_{\alpha\beta} \frac{\partial^{m+n} g_{\alpha\beta}}{\partial^m U \partial^n V}(s_i)\bigg|_{U=V=0} (dx^\alpha)_a (dx^\beta)_b. \tag{109}$$

The coefficients appearing in this expansion are just components of the tensors $\partial_{c_1}\cdots\partial_{c_r} g_{ab}$ on $\mathcal{C}$: namely,

$$\sum_{\alpha\beta} \frac{\partial^{m+n} g_{\alpha\beta}}{\partial^m U \partial^n V}(s_i)\bigg|_{U=V=0} (dx^\alpha)_a (dx^\beta)_b = l^{c_1}\cdots l^{c_m} n^{c_1}\cdots n^{c_n} \partial_{c_1}\cdots\partial_{c_{m+n}} g_{ab}. \tag{110}$$

We define $g^{I_q}_{ab}$ by replacing each tensor, $\partial_{c_1}\cdots\partial_{c_r} g_{ab}$, appearing in the expansion of $g^{(q)}_{ab}$ by its boost invariant part. In other words, we alter g_{ab} by extracting the boost invariant part of the coefficients of the first q terms of its Taylor expansion in U and V.

The nature of $g^{I_q}_{ab}$ can be best elucidated in the case where g_{ab} is analytic, in which case we may set $q = \infty$ and write g^I_{ab} for $g^{I_\infty}_{ab}$. It then follows that the vector field

$$\xi^a = U\left(\frac{\partial}{\partial U}\right)^a - V\left(\frac{\partial}{\partial V}\right)^a \tag{111}$$

(which induces Lorentz boosts of the coordinates) is a Killing field of the metric g^I_{ab}, that is,

$$\mathcal{L}_\xi g^I_{ab} = 0. \tag{112}$$

Furthermore, ξ^a vanishes on $\mathcal{C}$. Thus, our construction of g^I_{ab} has, in effect, created a new spacetime (which is *not* necessarily a solution of the field equations) in which $\mathcal{C}$ is the bifurcation surface of a bifurcate Killing horizon.

In an exactly similar manner, we define the boost invariant part, ψ^{I_q} of the matter fields ψ (up to order q) by extracting the boost invariant part of the coefficients of the first q terms of the Taylor expansion of ψ in U and V about $\mathcal{C}$. It then follows that ξ^a also Lie derives ψ^{I_q} up to order q.

Our proposal for defining $S_{\rm dyn}$ is the following: Choose q to be larger than the highest derivative of any dynamical field appearing in the decomposition of $\mathbf{Q}$ given in Proposition 4.1. Given a cross section $\mathcal{C}$ of the horizon of a black hole, we replace g_{ab} by $g^{I_q}_{ab}$ and ψ by ψ^{I_q} in a neighborhood of $\mathcal{C}$. Define $\tilde{\mathbf{Q}}[\xi]$ on $\mathcal{C}$ to be the Noether charge $(n-2)$-form of the dynamical fields $\phi^{I_q} = (g^{I_q}_{ab}, \psi^{I_q})$ for the vector field ξ^a defined by Eq. (111) above. Define $S_{\rm dyn}$ at "time" $\mathcal{C}$ by

$$S_{\rm dyn}[\mathcal{C}] = 2\pi \int_{\mathcal{C}} \tilde{\mathbf{Q}}[\xi]. \tag{113}$$

Equivalently, by Proposition 4.1 we have

$$S_{\rm dyn}[\mathcal{C}] = 2\pi \int_{\mathcal{C}} \tilde{\mathbf{X}}^{cd} \boldsymbol{\epsilon}_{cd}, \tag{114}$$

where

$$\tilde{\mathbf{X}}^{cd}(\phi) \equiv \mathbf{X}^{cd}(\phi^{I_q}). \tag{115}$$

Equation (114) shows that $S_{\rm dyn}$ is of the desired general form (104), and the equivalence of Eqs. (113) and (114) shows that the right side of (114) does not depend upon the choice of $\mathbf{X}^{cd}$ in the decomposition of Proposition 4.1. Note, incidentally, that since $\mathbf{X}^{cd}$ is a nonlinear function of the dynamical fields ϕ, the tensor field $\tilde{\mathbf{X}}^{cd}$ is *not* necessarily equal to the "boost invariant part" of the tensor field $\mathbf{X}^{cd}(\phi)$. [Use of the boost invariant part of $\mathbf{X}^{ab}(\phi)$ would not yield a satisfactory prescription for $S_{\rm dyn}$ since it would, in general, fail to satisfy condition (3) above.] More generally, for a nonlinear tensor function β of the dynamical fields ϕ, we have, in general, $[\beta(\phi)]^{I_q} \neq \beta(\phi^{I_q})$. On the other hand, if β is linear in ϕ, then $[\beta(\phi)]^{I_q} = \beta(\phi^{I_q})$.

We now may verify that our definition of $S_{\rm dyn}$ satisfies conditions (1)–(3) above. First, if $\mathcal{C}$ is taken to be the bifurcation surface Σ of a stationary black hole, then $\phi^{I_q} = \phi$, so, clearly, $S_{\rm dyn}[\Sigma] = S[\Sigma]$. On the other hand, if Σ' is an arbitrary cross section of a stationary black hole, then since our prescription for defining $S_{\rm dyn}$ is a "local, geometrical" one, by isometry invariance we clearly have $S_{\rm dyn}[\chi_t(\Sigma')] = S_{\rm dyn}[\Sigma']$. But it also is clear that our prescription for defining $S_{\rm dyn}$ is such that $S_{\rm dyn}[\mathcal{C}]$ varies continuously with $\mathcal{C}$. From these facts, it follows immediately by the same argument as given below Eq. (103) that $S_{\rm dyn}[\Sigma'] = S_{\rm dyn}[\Sigma]$. Thus, we have

$$S_{\rm dyn}[\Sigma'] = S_{\rm dyn}[\Sigma] = S[\Sigma] = S[\Sigma']; \tag{116}$$

i.e., condition (1) is satisfied.

To verify that condition (2) holds, we note that since we have $\phi^{I_q} = \phi$ on the bifurcation surface Σ of a stationary black hole, and since $\delta[\mathbf{X}^{cd}\boldsymbol{\epsilon}_{cd}]$ clearly is linear in $\delta\phi$, it follows that $\delta[(\tilde{\mathbf{X}}^{cd} - \mathbf{X}^{cd})\boldsymbol{\epsilon}_{cd}]$ has no boost invariant part. However, this immediately implies that the pullback of this differential form to Σ vanishes, from which it follows that $\delta S_{\rm dyn}[\Sigma] = \delta S[\Sigma]$, as desired.

Finally, the complete ambiguity in $\mathbf{Q}$ (including that arising from the change in Lagrangian $\mathbf{L} \to \mathbf{L} + d\boldsymbol{\mu}$) is given by Eq. (58). It is manifest that none of these ambiguous terms can contribute to $\int_{\mathcal{C}} \tilde{\mathbf{Q}}[\xi]$. Consequently, we see from Eq. (113) that condition (3) holds.

Thus, we have proven the existence of a definition of $S_{\rm dyn}$ which satisfies conditions (1)–(3). These conditions do *not* uniquely determine $S_{\rm dyn}$. Nevertheless, we have been unable to come up with any "natural" alternative definitions of $S_{\rm dyn}$. Thus, we believe that our definition of $S_{\rm dyn}$ is a serious candidate for the definition of the entropy of a nonstationary black hole in a general theory of gravity. *See the Note added to the end of the Section.*

We conclude this section by evaluating $S_{\rm dyn}$ for the three theories considered in Sec. V. Consider, first, vacuum general relativity. Let $\mathcal{C}$ be an arbitrary cross section of a black hole, let $\boldsymbol{\epsilon}_{ab}$ be the binormal to $\mathcal{C}$, and let $\tilde{\boldsymbol{\epsilon}}_{ab}$ denote the volume element on $\mathcal{C}$. Comparing Eqs. (51) and (63) we see that the two-form $\mathbf{X}^{cd}$ is given simply by

$$(\mathbf{X}^{cd})_{ab} = -\frac{1}{16\pi}\boldsymbol{\epsilon}_{ab}{}^{cd}. \tag{117}$$

Since $\mathbf{X}^{cd}$ does not depend upon any derivatives of g_{ab}, it is clear that it is unaffected when g_{ab} is replaced by its boost invariant part. Thus, we obtain

$$\begin{aligned} S_{\rm dyn}[\mathcal{C}] &= -\frac{1}{8}\int_{\mathcal{C}} \boldsymbol{\epsilon}_{ab}{}^{cd}\boldsymbol{\epsilon}_{cd} \\ &= \frac{1}{4}\int_{\mathcal{C}} \tilde{\boldsymbol{\epsilon}}_{ab} \\ &= \frac{\text{Area}[\mathcal{C}]}{4} \end{aligned} \tag{118}$$

in agreement with the usual formula for the entropy of a dynamical black hole in general relativity. By the area theorem, this definition of $S_{\rm dyn}$ satisfies the "second law" (assuming that the cosmic censor hypothesis is valid). Note that if we add to the Lagrangian "matter terms" which have no explicit dependence upon the curvature, then $\mathbf{X}^{cd}$ does not change [see Eq. (52)], so Eq. (118) also holds for general relativity with matter present, provided only that the matter does not have an explicit coupling to the curvature in the Lagrangian.

The calculation of $S_{\rm dyn}$ for dilaton gravity with Lagrangian (64) in two spacetime dimensions proceeds similarly. We see from Eq. (67) that the 0-form X^{cd} is given by

$$\mathbf{X}^{cd} = -\tfrac{1}{2}e^{\phi}\boldsymbol{\epsilon}^{cd}. \tag{119}$$

Again $\mathbf{X}^{cd}$ does not depend upon any derivatives of the dynamical fields, and is unchanged when they are replaced by their boost invariant parts. In this case, a cross section $\mathcal{C}$ of the horizon is a point, and we obtain

$$S_{\rm dyn}[\mathcal{C}] = 2\pi e^{\phi}\big|_{\mathcal{C}}. \tag{120}$$

It is known that this definition of $S_{\rm dyn}$ also satisfies the second law [3].

Lovelock gravity provides a more interesting illustration of our prescription, since it can be seen from Eq. (71) that $\mathbf{X}^{cd}$ contains terms involving the curvature

$$(\mathbf{X}^{cd})_{a_1\cdots a_{n-2}} = -\boldsymbol{\epsilon}^{cd}{}_{a_1\cdots a_{n-2}}\left(\frac{1}{16\pi} + 2\alpha R\right) - 8\alpha\boldsymbol{\epsilon}^{[d}{}_{f a_1\cdots a_{n-2}}R^{c]f} - 2\alpha\boldsymbol{\epsilon}_{fha_1\cdots a_{n-2}}R^{cdfh} \tag{121}$$

and the replacement of the metric by its boost invariant part will have a nontrivial effect. Indeed, since, after this replacement is made, both extrinsic curvatures of $\mathcal{C}$ embedded in M vanish, we see (using a "Gauss-Codazzi" equation, see, e.g., [12]) that the curvature of the boost invariant part of the metric satisfies

$${}^{(n-2)}R = R - 2t^{ab}R_{ab} + t^{ac}t^{bd}R_{abcd}, \tag{122}$$

where $t_{ab} = -n_a n_b + r_a r_b$ is the metric for the subspace

orthogonal to C (spanned by the unit timelike and spacelike normals, n^a and r^a, respectively) and ${}^{(n-2)}R$ is the scalar curvature of C. From Eqs. (121) and (122) we obtain

$$\epsilon_{cd}\tilde{\mathbf{X}}^{cd}{}_{a_1\cdots a_n} = \left(\frac{1}{8\pi} + 4\alpha\, {}^{(n-2)}R[g^{I_q}]\right)\tilde{\epsilon}_{a_1\cdots a_{n-2}}, \quad (123)$$

where ${}^{(n-2)}R[g^{I_q}]$ is the $(n-2)$-scalar curvature of C computed with the boost invariant part of the metric and $\tilde{\epsilon}_{a_1\cdots a_{n-2}}$ is the volume form C. However, we clearly have ${}^{(n-2)}R[g^{I_q}] = {}^{(n-2)}R[g]$. Hence, we obtain

$$S_{\rm dyn} = \tfrac{1}{4}\mathrm{Area}[C] + 8\pi\alpha\int_C {}^{(n-2)}R. \quad (124)$$

(In particular, Eq. (124) yields the entropy of a stationary black hole in Lovelock gravity, in agreement with [4].) Note that this formula differs from what would be obtained from simply substituting the expression (121) into Eq. (92). It is not known whether this definition of $S_{\rm dyn}$ satisfies the second law.

Note added. After this paper was submitted, it came to our attention that a fifth condition could be added to the requirements for the definition of $S_{\rm dyn}$: (5) $S_{\rm dyn}$ should not change under a local, nonderivative field redefinition $\phi \to F(\phi)$. Since, in general, we have $[F(\phi)]^{I_q} \neq F(\phi^{I_q})$, our proposed definition does *not* satisfy condition (5) for arbitrary theories of gravity, although it does satisfy this condition for the three theories explicitly considered above. Thus, it appears that the definition of $S_{\rm dyn}$ in an arbitrary theory of gravity remains an open problem.

ACKNOWLEDGMENTS

This research was supported in part by NSF Grant No. PHY-9220644 to the University of Chicago.

APPENDIX: APPLICATIONS TO THEORIES WITH A NONDYNAMICAL METRIC

In the body of this paper we have considered theories which are diffeomorphism covariant in the sense of Eq. (2). It was seen in Sec. II that this condition implies the absence of "nondynamical fields" in the Lagrangian. In particular, the diffeomorphism covariance condition excludes the case of theories with a nondynamical metric, such as theories of fields in flat spacetime. Nevertheless, a number of formulas and results derived in the body of this paper continue to hold for theories with a Lagrangian locally constructed out of a metric, g_{ab}, and matter fields ψ, of the form (3), i.e., for the Lagrangian

$$\mathbf{L} = \mathbf{L}\left(g_{ab}, \nabla_{a_1}R_{bcde}, ..., \nabla_{(a_1}\cdots\nabla_{a_m)}R_{bcde}, \psi, \nabla_{a_1}\psi, \nabla_{(a_1}\cdots\nabla_{a_l)}\psi\right) \quad (A1)$$

but where the metric g_{ab} is now treated as a fixed, nondynamical entity, so that, in particular, the equations of motion, $\mathbf{E}_g = 0$, no longer are imposed. The purpose of this appendix is to present simple, unified derivations of some formulas and results (most of which are "well known") for such theories with a nondynamical metric.

In a theory with Lagrangian of the form (A1) but with nondynamical metric, we define the *stress-energy tensor* $T^{ab} = T^{(ab)}$ of the matter fields by

$$T^{ab}\boldsymbol{\epsilon} = 2(\mathbf{E}_g)^{ab}. \quad (A2)$$

For each vector field, ξ^a, we again define the Noether current $\mathbf{J}$ by Eq. (48) above. However, the (matter) equations of motion no longer imply that $\mathbf{J}$ is closed. Indeed, by Eq. (49), we see that when $\mathbf{E}_\psi = 0$, we have

$$\begin{aligned} d\mathbf{J} &= -(\mathbf{E}_g)^{ab}\mathcal{L}_\xi g_{ab} \\ &= -T^{ab}\nabla_{(a}\xi_{b)}\boldsymbol{\epsilon} \\ &= -\nabla_a[T^{ab}\xi_b]\boldsymbol{\epsilon} + \xi_b\nabla_a[T^{ab}]\boldsymbol{\epsilon} \\ &= -d(k\cdot\boldsymbol{\epsilon}) + \xi_b\nabla_a[T^{ab}]\boldsymbol{\epsilon}, \end{aligned} \quad (A3)$$

where

$$k^a \equiv T^{ab}\xi_b. \quad (A4)$$

By inspection of Eq. (A3), we see that the n-form $\xi_b\nabla_a[T^{ab}]\boldsymbol{\epsilon}$ is exact for all ξ^a. However, since ξ^a is arbitrary, this is impossible unless

$$\nabla_a T^{ab} = 0 \quad (A5)$$

which shows that the stress-energy tensor is covariantly conserved whenever the matter equations of motion hold. Note that Eq. (A3) then yields simply

$$d(\mathbf{J} + k\cdot\boldsymbol{\epsilon}) = 0 \quad (A6)$$

from which it follows immediately [14] that

$$\mathbf{J} + k\cdot\boldsymbol{\epsilon} = d\mathbf{K}, \quad (A7)$$

where $\mathbf{K}$ is locally constructed out of g_{ab}, ψ, ξ^a, and their derivatives. In other words, we have shown that, apart from a "surface term," the Noether current is equivalent to the stress-energy current $-T^{ab}\xi_b$.

It is important to note that the equations of motion for g_{ab} were not used anywhere in the derivation of Eqs. (72)–(74) or Eqs. (77)–(80). Thus, these equations remain valid in the case of a theory with a nondynamical metric. In particular, if a Hamiltonian exists for a time translation vector field, t^a, on a globally hyperbolic, asymptotically flat spacetime, then it is natural to define the canonical energy at "time" C by

$$\mathcal{E} = \int_C \mathbf{J} - \int_\infty t\cdot\mathbf{B} \quad (A8)$$

[see Eqs. (80) and (82) above]. In other words, apart from the possible "surface term" $\int_\infty t\cdot\mathbf{B}$ (which vanishes

in most of the commonly considered theories of matter fields in a background spacetime), the canonical energy is simply the integral of the Noether current, $\mathbf{J}$, over a Cauchy surface. However, since we no longer have $\mathbf{J} = d\mathbf{Q}$, this volume integral no longer can be converted into a surface integral. In particular, $\mathcal{E}$ depends upon the choice of t^a in the interior of the spacetime, not just upon its asymptotic value at infinity. Note also that since $\mathbf{J}$ need not be closed [see Eq. (A3)], $\mathcal{E}$ need not be conserved, i.e., independent of C.

Using Eq. (A7) we find

$$\begin{aligned}\mathcal{E} &= -\int_C k\cdot\boldsymbol{\epsilon} + \int_\infty (\mathbf{K} - t\cdot\mathbf{B}) \\ &= \int_C T_{ab}n^a t^b\tilde{\boldsymbol{\epsilon}} + \int_\infty (\mathbf{K} - t\cdot\mathbf{B}),\end{aligned} \tag{A9}$$

where n^a denotes the future-directed unit normal to C, and $\tilde{\boldsymbol{\epsilon}}_{b_1\cdots b_{n-1}} = n^a\boldsymbol{\epsilon}_{ab_1\cdots b_{n-1}}$ is the natural volume element on C. Thus, apart from some possible surface term contributions which can arise from both $\mathbf{K}$ and $\mathbf{B}$, the canonical energy is given by the usual formula involving an integral of the stress-energy tensor over C.

As noted above, in general $\mathcal{E}$ is not conserved, i.e., independent of choice of Cauchy surface, C. However, if the spacetime metric is stationary, $\mathcal{L}_t g_{ab} = 0$ (but stationarity need *not* be imposed upon the matter fields), then the first line of Eq. (A3) shows that $\mathbf{J}$ is closed. Equation (A7) then immediately implies that the stress-energy current form $-k\cdot\boldsymbol{\epsilon}$ also is closed (as also could easily be verified directly). Equation (A8) then implies that $\mathcal{E}$ does not change when the Cauchy surface, C, undergoes variations of compact support. In the usual case where $t\cdot\mathbf{B}$ vanishes at infinity and $\mathbf{J}$ goes to zero suitably rapidly at infinity, $\mathcal{E}$ will take the same value for all asymptotically flat Cauchy surfaces.

Now, suppose that g_{ab} is stationary, i.e., $\mathcal{L}_t g_{ab} = 0$, and suppose that $\psi(\lambda)$ is a one-parameter family of solutions to the matter equations of motion (in the fixed metric g_{ab}) such that $\psi(0)$ is stationary, i.e., $\mathcal{L}_t\psi(0) = 0$. Let $\mathcal{E}(\lambda)$ denote the canonical energy of these solutions. Then, since $\mathcal{L}_t g_{ab} = 0$, by Eqs. (77) and (80), we have, for all λ,

$$\frac{d\mathcal{E}}{d\lambda}(\lambda) = \Omega\left(\psi(\lambda), \frac{d\psi}{d\lambda}, \mathcal{L}_t\psi(\lambda)\right). \tag{A10}$$

In particular, since $\mathcal{L}_t\psi$ vanishes at $\lambda = 0$, we see that the first variation of the canonical energy about a stationary solution vanishes:

$$\delta\mathcal{E} = 0. \tag{A11}$$

Now take the derivative of Eq. (A10) with respect to λ and evaluate the resulting equation at $\lambda = 0$. A nonzero contribution will occur on the right side only when the λ derivative acts on $\mathcal{L}_t\psi$. We thereby find that the second variation of canonical energy about a stationary solution ψ is given by

$$\delta^2\mathcal{E} \equiv \frac{1}{2}\frac{d^2\mathcal{E}}{d\lambda^2}\bigg|_{\lambda=0} = \frac{1}{2}\Omega(\psi, \delta\psi, \mathcal{L}_t\delta\psi), \tag{A12}$$

where

$$\delta\psi \equiv \frac{d\psi}{d\lambda}\bigg|_{\lambda=0}. \tag{A13}$$

Note, in particular, that $\delta^2\mathcal{E}$ depends only upon $\delta\psi$, and not upon $\delta^2\psi$.

Equation (A12) is one of the key results of this appendix. To elucidate its meaning, we note that if $\delta\psi$ satisfies the linearized equations of motion about a stationary solution, ψ, then so does $\mathcal{L}_t\delta\psi$. Consequently, the symplectic current form $\boldsymbol{\omega}(\psi, \delta\psi, \mathcal{L}_t\delta\psi)$, defined above by Eq. (43), is closed. Thus, its integral over a Cauchy surface, C, yields a conserved quantity for perturbations. Equation (A12) shows that, apart from a factor of 2, this conserved quantity is just the second order change in the canonical energy associated with this perturbation. By our previous results, we see that this conserved quantity is equivalent, up to possible "surface terms," to the conserved quantities $\int_C \delta^2\mathbf{J}$ and $\int_C \delta^2 T_{ab}n^a t^b\tilde{\boldsymbol{\epsilon}}$.

As a simple application of the above result, consider the theory of a linear field ψ in a stationary spacetime, where by "linear" we mean that $\mathbf{L}$ is quadratic in ψ, so that the equations of motion for ψ are linear. In this case, the equations of motion are the same as the linearized equations about $\psi = 0$, so we may choose the "unperturbed solution" to be $\psi = 0$, and we may write $\delta\psi = \psi$ in the above formulas. We also have $\delta^2\mathcal{E} = \mathcal{E}$ and $\delta^2 T_{ab} = T_{ab}$. Hence, we obtain, from Eqs. (A9) and (A12),

$$\Omega(\delta\psi, \mathcal{L}_t\delta\psi) = 2\int_C T_{ab}n^a t^b\tilde{\boldsymbol{\epsilon}} + 2\int_\infty (\mathbf{K} - t\cdot\mathbf{B}). \tag{A14}$$

Again, for the types of theories usually considered (such as a Klein-Gordon scalar field), the surface terms from infinity in Eq. (A14) vanish. The resulting relation plays an important role in defining a natural vacuum state for linear quantum fields in a stationary spacetime [21,22].

Consider, now, the case where the nondynamical metric is a flat metric, η_{ab}. We denote the (flat) derivative operator associated with η_{ab} by ∂_a. Let ξ^a be a translational Killing field of η_{ab}, so that $\partial_a\xi^b = 0$. Then, clearly, at each point of spacetime the Noether current $\mathbf{J}$ associated with ξ^a is linear in the value of ξ^a at that point. Hence, there exists a unique tensor field, $\mathcal{T}^a{}_b$, called the *canonical energy-momentum tensor*, such that

$$J_{a_1\cdots a_{n-1}} = -\mathcal{T}^a{}_b\xi^b\boldsymbol{\epsilon}_{aa_1\cdots a_{n-1}}. \tag{A15}$$

Conservation of $\mathbf{J}$ implies conservation of $\mathcal{T}^a{}_b$ in its first index: i.e.,

$$\partial_a\mathcal{T}^a{}_b = 0 \tag{A16}$$

However, $\mathcal{T}^{ab}$ need not be symmetric. Nevertheless, Eq. (A7) implies that there exists a tensor field $H^{abc} = H^{[ab]c}$, locally constructed out of η_{ab} and ψ, such that

$$\mathcal{T}^{ab} = T^{ab} + \partial_c H^{cab}. \tag{A17}$$

Thus, we have rederived the well-known fact that $\mathcal{T}_{ab}$

always can be "symmetrized" by the addition of an identically conserved tensor $\partial_c H^{cab}$.

Finally, we note that much of the theory of pseudotensors can be derived by applying the results of this appendix back to the case where the metric again is a dynamical variable in the Lagrangian (A1). For a diffeomorphism invariant theory of the type considered in the body of this paper, we may introduce a fixed flat metric, η_{ab}, on spacetime, and express the dynamical metric g_{ab} as

$$g_{ab} = \eta_{ab} + h_{ab}. \tag{A18}$$

We then may treat η_{ab} and h_{ab} as independent fields, and view our theory as a theory with Lagrangian of the form (A1) with the dynamical fields (h_{ab}, ψ) and a nondynamical metric η_{ab}. One of the (very few) advantages of doing this is that many more quantities qualify as "covariant" when η_{ab} and h_{ab} are viewed as independent fields. In particular, in general relativity, no diffeomorphism covariant $(n-1)$-form, $\mathbf{B}$, satisfying Eq. (79) can be constructed out of g_{ab}, but there is no difficulty in constructing a diffeomorphism covariant $\mathbf{B}$ out of the independent fields η_{ab} and h_{ab}. Thus, we may change the Lagrangian via

$$\mathbf{L} \to \mathbf{L}' = \mathbf{L} - d\mathbf{B} \tag{A19}$$

and still view $\mathbf{L}'$ as being of the general form (A1) (with η_{ab} and h_{ab} viewed as independent fields). Under the change of Lagrangian (A19), $\boldsymbol{\Theta}$ is modified by

$$\boldsymbol{\Theta} \to \boldsymbol{\Theta}' = \boldsymbol{\Theta} - \delta\mathbf{B} \tag{A20}$$

[see Eq. (41) above]. Consequently, we have $\int_\infty t\cdot\boldsymbol{\Theta}' = 0$, and the canonical energy now is given simply by

$$\mathcal{E} = \int_C \mathbf{J}'. \tag{A21}$$

Since the theory has been recast to have a Lagrangian of the form (A1) in a spacetime with a nondynamical flat metric η_{ab}, a canonical energy-momentum tensor can be defined by Eq. (A15). We denote this tensor as $t^a{}_b$, and refer to it as a *pseudotensor* because it depends upon the choice of flat metric, η_{ab} and thus is not covariant with respect to diffeomorphisms which act only upon the dynamical fields. For the case of vacuum general relativity with the Lagrangian $\mathbf{L}'$ of Eq. (A19) with an appropriate choice of $\mathbf{B}$, $t^a{}_b$ corresponds to the Einstein pseudotensor [23]. Note that Eq. (A21) can be rewritten in terms of $t^a{}_b$ as

$$\begin{aligned}\mathcal{E} &= \int_C t^a{}_b n_a t^b \tilde{\boldsymbol{\epsilon}} \\ &= \int t^0{}_0 d^3x,\end{aligned} \tag{A22}$$

where the last line holds when C is taken to be the hypersurface t =const in a global inertial coordinate system of η_{ab}.

In order to define a stress-energy tensor corresponding to (A2) we must specify the functional dependence of the Lagrangian on η_{ab} for general (nonflat) η_{ab}. One way to do this would be to take $\mathbf{L}$ for general η_{ab} to be given by the substitution (A18) in the original Lagrangian. In that case, $\mathbf{L}$ clearly depends upon η_{ab} and h_{ab} only in the combination $\eta_{ab} + h_{ab}$. Consequently, the equations of motion for η_{ab} will be satisfied whenever the equations of motion for the dynamical field h_{ab} hold. Note that $\mathbf{B}$ need not depend only on the combination $\eta_{ab} + h_{ab}$, so $\mathbf{L}'$, defined by Eq. (A19), need not depend only upon this combination. Nevertheless, since addition of an exact form to $\mathbf{L}$ does not alter the equations of motion, it remains true for $\mathbf{L}'$ that the equations of motion for η_{ab} will be satisfied whenever the equations of motion for the dynamical field h_{ab} hold. But, this implies that the energy-momentum tensor defined by (A2) vanishes by virtue of the equations of motion for the dynamical fields. Equation (A17) then yields

$$t^{ab} = \partial_c H^{cab}. \tag{A23}$$

This proves that, when the equations of motion are imposed, the pseudotensor $t^a{}_b$ always can be derived from a "superpotential" H^{cab}. Consequently, the volume integral (A22) always can be converted to a surface integral at infinity. This fact, of course, corresponds to our previous result, Eq. (82), which was derived in a much more simple and direct manner.

When recast in the form (A1), the Lagrangian obtained from $\mathbf{L}'$ by the simple substitution (A18) described above will, in general, have a nontrivial, explicit dependence upon the curvature of η_{ab}. However, an alternative procedure for defining a Lagrangian for nonflat η_{ab}, which clearly agrees with $\mathbf{L}'$ when η_{ab} is flat, would be to modify the Lagrangian of the previous paragraph by simply setting the terms in $\mathbf{L}'$ involving the curvature of η_{ab} to zero. If we do so, the stress-energy tensor defined by (A2) for this modified Lagrangian will be nonvanishing. We denote this stress-energy tensor by $\tilde{t}_{ab}$. If, as seems plausible, the surface term $\mathbf{K}$ does not contribute to Eq. (A9), then the symmetric pseudotensor $\tilde{t}_{ab}$ will be equivalent to t_{ab} insofar as the calculation of canonical energy is concerned, i.e., Eq. (A22) will hold with t_{ab} replaced by $\tilde{t}_{ab}$. Note that Eqs. (A9) and (A23) imply that $\tilde{t}_{ab}$ also is derivable from a superpotential. Other symmetric pseudotensors derivable from a superpotential can be explicitly constructed in the case of general relativity (see, e.g., [24]).

The dependence of pseudotensors such as t_{ab} or $\tilde{t}_{ab}$ on the choice of η_{ab} significantly limits their physical interpretation and utility. Indeed, it is difficult to imagine any use to which they could be put other than for the definition or calculation of the canonical energy and other asymptotic conserved quantities, and this can be accomplished much more straightforwardly by the methods described in the body of this paper. Nevertheless, the canonical energy can be correctly computed from a pseudotensor via Eq. (A22). In particular, if we consider perturbations of a stationary solution and choose the flat metric η_{ab} so that the Killing field, t^a, of the stationary background is a translational Killing field of η_{ab}, then

$$\delta^2\mathcal{E} = \int \delta^2 t^0{}_0 d^3x \qquad \text{(A24)}$$

is a nontrivial conserved quantity which depends only on the first order perturbation of the dynamical fields. For an arbitrary pseudotensor, this formula for $\delta^2\mathcal{E}$ is not very useful because to get an expression for the conserved quantity in terms of the first order perturbation, one must use the second order field equations to eliminate the terms in $\delta^2 t^0{}_0$ which involve the second order perturbation. However, as shown by Sorkin [23], for the Einstein pseudotensor, the terms in $\delta^2 t^0{}_0$ involving the second order perturbation are separately conserved (irrespective of the second order field equations), so the Einstein pseudotensor can be used to obtain a nontrivial conserved quantity constructed out of the first order perturbation of the dynamical fields. For perturbations of static, electrovac spacetimes in general relativity, this conserved quantity is equivalent to the conserved flux integral obtained by Chandrasekhar and Ferrari [25,26]. As was shown explicitly in [27], the Chandrasekhar-Ferrari conserved current also is equivalent to the symplectic current $(n-1)$-form $\boldsymbol{\omega}(\phi; \delta\phi, \mathcal{L}_t\delta\phi)$, which directly yields $\delta^2\mathcal{E}$ by Eq. (A12) above. Note that $\boldsymbol{\omega}$ is constructed entirely out of the dynamical fields and their perturbations—in particular, no background flat metric need be introduced—so it provides a covariant version of the conserved flux integral. However, $\boldsymbol{\omega}$ is not gauge invariant under infinitesimal gauge transformations of the perturbed dynamical fields, so it also does not provide a meaningful notion of the local energy density of the perturbation.

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

[2] R.M. Wald, in *Directions in General Relativity*, edited by B.L. Hu, M. Ryan, and C.V. Vishveshwara (Cambridge University Press, Cambridge, England, 1993), Vol. 1.

[3] V.P. Frolov, Phys. Rev. D **46**, 5383 (1992).

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

[5] J.D. Brown and J.W. York, Phys. Rev. D **47**, 1407 (1993); **47**, 1420 (1993).

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

[7] M. Visser, Phys. Rev. D **48**, 5697 (1993).

[8] T.A. Jacobson, G. Kang, and R.C. Myers, Phys. Rev. D **49**, 6587 (1994).

[9] M. Bañados, C. Teitelboim, and J. Zanelli, Phys. Rev. Lett. **72**, 957 (1994).

[10] S. Carlip and C. Teitelboim (unpublished).

[11] We wish to thank R. Myers and T. Jacobson for bringing this issue to our attention.

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

[13] R. Penrose, Ann. Phys. (N.Y.) **10**, 171 (1960).

[14] R. M. Wald, J. Math. Phys. **31**, 2378 (1993). The main results of this reference also can be derived using the "free variational bicomplex"; see I.M. Anderson, in *Mathematical Aspects of Classical Field Theory*, edited by M. Gotay, J. Marsden, and V. Moncrief [Cont. Math. **132**, 51 (1992)].

[15] J. Lee and R. M. Wald, J. Math. Phys. **31**, 725 (1990).

[16] T.Y. Thomas, *Differential Invariants of Generalised Spaces* (Cambridge University Press, Cambridge, England, 1934).

[17] D. Bak, D. Cangemi, and R. Jackiw, Phys. Rev. D **49**, 5173 (1994).

[18] D. Wiltshire, Phys. Lett. **169B**, 36 (1986).

[19] R. Beig, Phys. Lett. **69A**, 153 (1979); A. Ashtekar and A. Magnon, J. Math. Phys. **20**, 793 (1979).

[20] B.S. Kay and R.M. Wald, Phys. Rep. **207**, 49 (1991).

[21] A. Ashtekar and A. Magnon Proc. R. Soc. London **A346**, 375 (1975).

[22] B.S. Kay, Commun. Math. Phys. **62**, 55 (1978).

[23] R.D. Sorkin, Proc. R. Soc. London **A435**, 635 (1991).

[24] L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, England, 1962).

[25] S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London **A428**, 325 (1990).

[26] S. Chandrasekhar and V. Ferrari, Proc. R. Soc. London **A435**, 645 (1991).

[27] G. Burnett and R.M. Wald, Proc. R. Soc. London **A430**, 57 (1990).