Black Hole Entropy is Noether Charge

Robert M. Wald

University of Chicago

Enrico Fermi Institute and Department of Physics

5640 S. Ellis Avenue

Chicago, Illinois 60637-1433

Abstract

We consider a general, classical theory of gravity in n dimensions, arising from a diffeomorphism invariant Lagrangian. In any such theory, to each vector field, ξ^a, on spacetime one can associate a local symmetry and, hence, a Noether current $(n-1)$-form, $\mathbf{j}$, and (for solutions to the field equations) a Noether charge $(n-2)$-form, $\mathbf{Q}$, both of which are locally constructed from ξ^a and the the fields appearing in the Lagrangian. Assuming only that the theory admits stationary black hole solutions with a bifurcate Killing horizon (with bifurcation surface Σ), and that the canonical mass and angular momentum of solutions are well defined at infinity, we show that the first law of black hole mechanics always holds for perturbations to nearby stationary black hole solutions. The quantity playing the role of black hole entropy in this formula is simply 2π times the integral over Σ of the Noether charge $(n-2)$-form associated with the horizon Killing field (i.e., the Killing field which vanishes on Σ), normalized so as to have unit surface gravity. Furthermore, we show that this black hole entropy always is given by a local geometrical expression on the horizon of the black hole. We thereby obtain a natural candidate for the entropy of a dynamical black hole in a general theory of gravity. Our results show that the validity of the "second law" of black hole mechanics in dynamical evolution from an initially stationary black hole to a final stationary state is equivalent to the positivity of a total Noether flux, and thus may be intimately related to the positive energy properties of the theory. The relationship between the derivation of our formula for black hole entropy and the derivation via "Euclidean methods" also is explained.

PACS #: 04.20.-q, 97.60.Lf

One of the most remarkable developments in the theory of black holes in classical general relativity was the discovery of a close mathematical analogy between certain laws of "black hole mechanics" and the ordinary laws of thermodynamics. When the effects of quantum particle creation by black holes [1] were taken into account, this analogy was seen to be of a physical nature, and it has given rise to some deep insights into phenomena which may be expected to occur in a quantum theory of gravity.

The original derivation of the laws of black hole mechanics in classical general relativity [2] used many detailed properties of the Einstein field equations, and, thus, appeared to be very special to general relativity. However, recently it has become clear that at least some of the laws of classical black hole mechanics hold in a much more general context. In particular, it has been shown that a version of the first law of black hole mechanics holds in any theory of gravity derivable from a Hamiltonian [3]. (For the cases of $(1+1)$-dimensional theories of gravity [4] and Lovelock gravity [5], the explicit forms of this law have been given.) Furthermore, analogs of all of the classical laws of black hole mechanics have been shown to hold in $(1+1)$-dimensional theories [4].

However, despite the very general nature of the Hamiltonian derivation [3] of the first law of black hole mechanics, there remains one unsatisfactory aspect of the status of the first law in a general theory of gravity [6]: Although the derivation shows that for a perturbation of a stationary black hole, a surface integral at the black hole horizon (involving the unperturbed metric and its variation) is equal to terms involving the variation of mass and angular momentum (and possibly other asymptotic quantities) at infinity, the derivation does not show that this surface term at the horizon can be expressed as $\kappa/2\pi$ (where κ denotes the unperturbed surface gravity) times the variation of a surface integral of the form $S = \int_\Sigma F$, where F is locally constructed out of the metric and other dynamical fields appearing in the theory. It is necessary that the horizon surface term be expressible in this form in order to be able to identify a local, geometrical quantity, S, as playing the role of the entropy of the black hole.

The main purpose of this paper is to remedy this deficiency by showing that in a general theory of gravity derivable from a Lagrangian, the form of the first law of black hole mechanics for perturbations to nearby stationary black holes is such that the surface term at the horizon always takes the form $\frac{\kappa}{2\pi}\delta S$, where S is a local geometrical quantity, and is equal to 2π times the Noether charge at the horizon of the horizon Killing field (normalized so as to have unit surface gravity). The local, geometrical character of S suggests a possible generalization of the definition of entropy to dynamical black holes. The relationship between black hole entropy and Noether charge also suggests the possibility of a general relationship between the validity of the second law of black hole mechanics (i.e., increase of black hole entropy) and positive energy properties of a theory. An additional byproduct of our analysis is that it will enable us to make contact with the "Euclidean derivation" of formulas for black hole entropy, thereby demonstrating equivalence of that approach with other approaches – a fact that is not at all easy to see by a direct comparison of, say, references [3] and [7]. Our considerations in this paper will be limited to a general analysis of all the above issues; applications to particular theories will be given elsewhere [8]

Before presenting our new derivation of the first law, we comment upon the status of other "preliminary laws" of black hole mechanics in a general theory of gravity. We consider theories defined on an n-dimensional manifold M with dynamical fields consisting of a (Lorentzian) spacetime metric, g_{ab}, and possibly other matter fields, such that the equations of motion of the metric and other fields are derivable from a diffeomorphism invariant Lagrangian. (Our precise assumptions concerning the Lagrangian will be spelled out in more detail below.) We assume that a suitable notion of "asymptotic flatness" is defined in the theory. The *black hole* region of an asymptotically flat spacetime then is defined to be the complement of the past of the asymptotic region. In order to begin consideration of the classical laws of black hole mechanics, it is necessary that the event horizon of a stationary black hole be a Killing horizon, i.e., a null surface to which a Killing vector field is normal. This property is known to be true in general relativity by a nontrivial argument using the null initial value formulation [9], so it is not obvious that it would hold in more general theories of gravity. Nevertheless, this property automatically holds for all static black holes (since the static Killing field must be normal to the event horizon of the black hole), and, hence, it automatically holds for spherically symmetric black holes (in the $O(n-1)$ sense) and, thus, in particular, for all black holes in $(1+1)$-dimensional theories of gravity.

The *surface gravity*, κ, at any point, p, of a Killing horizon $\mathcal{H}$ is defined by (see, e.g., [10]),

$$\xi^a \nabla_a \xi^b = \kappa \xi^b \tag{1}$$

where ξ^a is the Killing field normal to $\mathcal{H}$. The zeroth law of black hole mechanics asserts that κ is constant over the event horizon of a stationary black hole. The proof of this law *does* make direct use of the specific form of the Einstein field equations [2], and, thus, does not appear likely to generalize to other theories of gravity [11]. Nevertheless, the zeroth law trivially holds for spherically symmetric black holes, and, in particular, in all $(1+1)$-dimensional theories.

It is worth noting that the validity of the zeroth law is, in essence, equivalent to the statement that – apart from the "degenerate case" of vanishing surface gravity – the event horizon of a stationary black hole must be of bifurcate type. Namely, it is easily proven that a bifurcate Killing horizon must have constant (and nonvanishing) surface gravity, whereas it can be shown [12] that any Killing horizon with constant, nonvanishing surface gravity can be locally extended (if necessary) to a bifurcate horizon.

It also should be noted that in an arbitrary theory of gravity, a black hole with constant surface gravity will "Hawking radiate" at temperature $\kappa/2\pi$ when quantum particle creation effects are taken into account, i.e., the Einstein field equations play no role in the derivation of the Hawking effect. Similarly, the theorems of [13] on the uniqueness and thermal properties of quantum states on black holes with bifurcate horizons hold in an arbitrary theory of gravity. Thus, $\kappa/2\pi$ always represents the physical temperature of a black hole.

We turn, now, to the presentation of a new derivation of the first law of black hole mechanics for stationary black holes with bifurcate horizons in a general theory of gravity in n-dimensions derived from a diffeomorphism invariant Lagrangian. We shall follow closely the framework of Lagrangian field theories developed in [14], with one small change: We shall view the Lagrangian as an n-form, $\mathbf{L}$, rather than as a scalar density; similarly, other tensor densities of [14] will appear here in their dualized version as differential forms. In order to define $\mathbf{L}$, it is necessary to introduce a fixed (i.e., "nondynamical") derivative operator, ∇_a, on spacetime. It also may be necessary to introduce other "non-dynamical, background fields", γ, such as the curvature of ∇_a (if ∇_a is non-flat); we shall assume, however, that any such additional fields, γ, are uniquely determined by ∇_a, and that γ changes by a diffeomorphism under the change induced in ∇_a by the action of that diffeomorphism. At each point p of spacetime, $\mathbf{L}$ then is required to be a function of the spacetime metric, g_{ab}, (or, alternatively, of a tetrad or soldering form) and finitely many of its (symmetrized) derivatives at p, as well as of other matter fields present in the theory and their (symmetrized) derivatives at p, and of γ at p. Note that no restriction is placed upon the number of derivatives of the metric or other fields upon which $\mathbf{L}$ can depend (other than that this number be finite), so "higher derivative" gravity theories are included in this framework.

In order to reduce the number of symbols and indices appearing in formulas, I shall use the symbol "ϕ" to denote all of the dynamical fields, including the spacetime metric. We shall restrict attention to diffeomorphism invariant theories, by which we mean that for any diffeomorphism, $\psi : M \to M$, we have,

$$\mathbf{L}[\psi^*(\phi)] = \psi^* \mathbf{L}[\phi] \tag{2}$$

Note that on the left side of this equation, ψ^* is *not* applied to ∇_a or any other non-dynamical fields γ which may appear in $\mathbf{L}$. Equation (2) can be interpreted as stating that – although it may be necessary to introduce ∇_a and/or γ to define $\mathbf{L}$ – $\mathbf{L}$ actually depends only upon the dynamical fields ϕ.

Under a first order variation of the dynamical fields, the variation of $\mathbf{L}$ can be put in the form (see, e.g., [14]),

$$\delta \mathbf{L} = \mathbf{E} \delta\phi + d\mathbf{\Theta} \tag{3}$$

where summation over the dynamical fields (and contraction of their tensor indices with corresponding dual tensor indices of $\mathbf{E}$) is understood in the first term on the right side of this equation. The $(n-1)$-form, $\mathbf{\Theta}$, is locally constructed from ϕ and $\delta\phi$, but is determined by eq.(3) only up to addition of a closed (and, hence, exact [15]) form locally constructed from the fields appearing in $\mathbf{L}$; we shall adopt eq. (2.12) of [14] as our definition of $\mathbf{\Theta}$. The symplectic current $(n-1)$-form, $\mathbf{\Omega}$, is defined in terms of the variation of $\mathbf{\Theta}$ by,

$$\mathbf{\Omega}(\phi, \delta_1\phi, \delta_2\phi) = \delta_1[\mathbf{\Theta}(\phi, \delta_2\phi)] - \delta_2[\mathbf{\Theta}(\phi, \delta_1\phi)] \tag{4}$$

It should be noted that $\mathbf{\Theta}$ and $\mathbf{\Omega}$ will depend upon the choice of ∇_a in sufficiently high derivative theories [14] – although they change only by an exact form, i.e., a "surface term", under a change of derivative operator – and they need not be diffeomorphism invariant in the sense of eq.(2). Furthermore, $\mathbf{\Theta}$ and $\mathbf{\Omega}$ will change when an exact form is added to $\mathbf{L}$ – with the change in $\mathbf{\Omega}$ being given by an exact form – even though such a modification of $\mathbf{L}$ has no effect upon the equations of motion, $\mathbf{E} = 0$.

Now, let ξ^a be any vector field on M and consider the field variation $\hat{\delta}\phi = \mathcal{L}_\xi \phi$. The diffeomorphism invariance of $\mathbf{L}$ implies that under this variation, we have,

$$\hat{\delta}\mathbf{L} = \mathcal{L}_\xi \mathbf{L} = d(\xi \cdot \mathbf{L}) \tag{5}$$

where here and below, we make frequent use of the general identity

$$\mathcal{L}_\xi \mathbf{\Lambda} = \xi \cdot d\mathbf{\Lambda} + d(\xi \cdot \mathbf{\Lambda}) \tag{6}$$

holding for any differential form $\mathbf{\Lambda}$ and vector field ξ^a, where "$\cdot$" denotes the contraction of a vector field with the first index of a differential form. Equation (5) shows that the vector fields on M constitute a collection of infinitesimal local symmetries in the sense of [14]. Hence, to each ξ^a we may associate a Noether current $(n-1)$-form, $\mathbf{j}$, defined by

$$\mathbf{j} = \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \xi \cdot \mathbf{L} \tag{7}$$

so that $\mathbf{j}$ is locally constructed out of the fields appearing in $\mathbf{L}$ and ξ^a. A standard calculation [14] shows that

$$d\mathbf{j} = -\mathbf{E}\mathcal{L}_\xi \phi \tag{8}$$

so that $\mathbf{j}$ is closed whenever the equations of motion are satisfied. Since $\mathbf{j}$ is closed for all ξ^a, it follows [15] that there exists an $(n-2)$-form, $\mathbf{Q}$ – locally constructed out of the fields appearing in $\mathbf{L}$ and ξ^a – such that when evaluated on solutions to the equations of motion, we have,

$$\mathbf{j} = d\mathbf{Q} \tag{9}$$

Since $\mathbf{j}$ depends linearly on ξ^a, we adopt the explicit algorithm provided by lemma 1 of [15] to uniquely define $\mathbf{Q}$, from which it follows that $\mathbf{Q}$ depends on no more than $(k-1)$ derivatives of ξ^a, where k denotes the highest derivative of any dynamical field occurring in $\mathbf{L}$. (Note, however, that $\mathbf{Q}$ is unique up to addition of a closed – and, hence, exact [15] – $(n-2)$-form locally constructed from the fields appearing in $\mathbf{L}$ and from ξ^a, so the integral of $\mathbf{Q}$ over any closed $(n-2)$-dimensional surface, Σ, is uniquely defined by eq.(9) alone.) We shall refer to $\mathbf{Q}$ as the *Noether charge* $(n-2)$*-form* [16] relative to ξ^a, and its integral over a closed surface, Σ, will be referred to as the *Noether charge* of Σ relative to ξ^a.

The key identity upon which our derivation of the first law of black hole mechanics will be based is obtained by considering the variation of eq.(7) resulting from an arbitrary variation, $\delta\phi$, of the dynamical fields off of an arbitrary solution ϕ. We have,

$$\delta\mathbf{j} = \delta[\mathbf{\Theta}(\phi, \mathcal{L}_\xi\phi)] - \xi \cdot \delta\mathbf{L} \tag{10}$$

(Note that ξ^a is held fixed in this variation, i.e., we require that $\delta\xi^a = 0$.) However, by eq.(3), we have,

$$\begin{aligned} \xi \cdot \delta\mathbf{L} &= \xi \cdot [\mathbf{E}\delta\phi + d\mathbf{\Theta}] \\ &= \mathcal{L}_\xi\mathbf{\Theta} - d(\xi \cdot \mathbf{\Theta}) \end{aligned} \tag{11}$$

where the equations of motion, $\mathbf{E} = 0$, for ϕ and the identity (6) were used in the second line. Thus, we obtain,

$$\delta\mathbf{j} = \delta[\mathbf{\Theta}(\phi, \mathcal{L}_\xi\phi)] - \mathcal{L}_\xi[\mathbf{\Theta}(\phi, \delta\phi)] + d(\xi \cdot \mathbf{\Theta}) \tag{12}$$

Note that in eq.(12), no restrictions have been placed upon $\delta\phi$ or ξ^a.

Our next step is to identify certain "surface terms" appearing in eq.(12). First, we require ∇_a to be invariant under the diffeomorphisms generated by ξ^a. This requirement holds in the usual case where ξ^a is taken to be a coordinate vector field and ∇_a is taken to be the coordinate derivative operator of that coordinate system; it also will hold in our main application below where ∇_a will be taken to be the derivative operator of the unperturbed metric and ξ^a is a Killing field of that metric.) In that case, the first two terms on the right side of eq.(12) combine to yield

$$\delta[\boldsymbol{\Theta}(\phi, \mathcal{L}_\xi \phi)] - \mathcal{L}_\xi[\boldsymbol{\Theta}(\phi, \delta\phi)] = \boldsymbol{\Omega}(\phi, \delta\phi, \mathcal{L}_\xi \phi) \tag{13}$$

and eq.(12) becomes simply,

$$\delta \mathbf{j} = \boldsymbol{\Omega}(\phi, \delta\phi, \mathcal{L}_\xi \phi) + d(\xi \cdot \boldsymbol{\Theta}) \tag{14}$$

When integrated over a Cauchy surface, $\mathcal{C}$ of the unperturbed solution, eq.(14) corresponds to eq.(3.22) of [14], but eq.(14) contains vital additional information concerning the "surface term", $d(\xi \cdot \boldsymbol{\Theta})$, which did not appear in [14], since attention there was restricted to the case of compact $\mathcal{C}$. Comparison of eq.(14) with Hamilton's equations of motion shows that if a Hamiltonian, H, corresponding to evolution by ξ^a exists on phase space, then H must satisfy,

$$\delta H = \delta \int_{\mathcal{C}} \mathbf{j} - \int_{\mathcal{C}} d(\xi \cdot \boldsymbol{\Theta}) \tag{15}$$

where, in this equation, projection of the right side to phase space (in the manner discussed in [14]) should be understood. This shows that apart from the "surface term" $d(\xi \cdot \boldsymbol{\Theta})$, the Noether current, $\mathbf{j}$, acts as a Hamiltonian density.

We now further restrict attention to the case where $\delta\phi$ satisfies the linearized equations of motion, so that both ϕ and its variation are solutions. Then we may replace $\mathbf{j}$ and its variation by $d\mathbf{Q}$ in eqs.(14) and (15). It then can be seen immediately from eq.(15) that the Hamiltonian – if it exists – is purely a "surface term". In an asymptotically flat spacetime, it is natural to associate the value of the surface contribution to the Hamiltonian from infinity with the corresponding "conserved quantity" associated with ξ^a in the manner of [17]. In other words, *if* the theory admits a suitable definition of the "canonical energy", $\mathcal{E}$, associated with an asymptotic time translation, t^a, and of the "canonical angular momentum" $\mathcal{J}$, associated with an asymptotic rotation, φ^a, the variations of these quantities should be given by the formulas

$$\delta \mathcal{E} = \int_\infty (\delta \mathbf{Q}[t] - t \cdot \boldsymbol{\Theta}) \tag{16}$$

$$\delta \mathcal{J} = - \int_\infty \delta \mathbf{Q}[\varphi] \tag{17}$$

where the integrals are taken over an $(n-2)$-dimensional sphere at infinity and the term $\varphi \cdot \mathbf{\Theta}$ does not appear in eq.(17) because φ^a is assumed to be tangent to this sphere. Thus, if one can find an $(n-1)$-form, $\mathbf{B}$, such that

$$\delta \int_\infty t \cdot \mathbf{B} = \int_\infty t \cdot \mathbf{\Theta} \tag{18}$$

the canonical energy and angular momentum can be defined by,

$$\mathcal{E} = \int_\infty (\mathbf{Q}[t] - t \cdot \mathbf{B}) \tag{19}$$

$$\mathcal{J} = -\int_\infty \mathbf{Q}[\varphi] \tag{20}$$

Note that $\mathcal{E}$ corresponds to the "ADM mass" of general relativity plus possible additional contributions from any long-range matter fields that may be present; see [3] for explicit discussion of the case of the Yang-Mills field. Note also that for the Hilbert Lagrangian of general relativity, the expressions $\int_\infty \mathbf{Q}[t]$ and $-\int_\infty \mathbf{Q}[\varphi]$ correspond – up to numerical factors – to the Komar expressions for mass and angular momentum. The presence of the "extra term" $t \cdot \mathbf{B}$ in eq.(19) accounts for why different relative numerical factors must be chosen in the Komar formulas for these quantities.) It is, of course, a nontrivial condition on a theory that it admit a notion of asymptotic flatness such that $\mathcal{E}$ and $\mathcal{J}$ are well defined. In the following, I shall assume that this is the case, and derive the first law of black hole mechanics for such a theory.

Consider, now, a stationary black hole solution with a bifurcate Killing horizon, with bifurcation $(n-2)$-surface Σ. Choose ξ^a to be the Killing field which vanishes on Σ, normalized so that

$$\xi^a = t^a + \Omega_H^{(\mu)} \varphi^a_{(\mu)} \tag{21}$$

where t^a is the stationary Killing field (with unit norm at infinity) and summation over μ is understood. (This equation both picks out a particular family of axial Killing fields, $\varphi^a_{(\mu)}$, acting in orthogonal planes, and defines the "angular velocity of the horizon", $\Omega_H^{(\mu)}$.) Choose ∇_a to be the derivative operator of this solution, so that ∇_a is invariant under the isometries generated by ξ^a. Then eq.(13) holds, and, in addition, the right side now vanishes since $\mathcal{L}_\xi \phi = 0$. Let $\delta\phi$ be an arbitrary, asymptotically flat solution of the linearized equations. Then, the fundamental identity eq.(12) yields simply

$$d(\delta\mathbf{Q}) = d(\xi \cdot \mathbf{\Theta}) \tag{22}$$

Choose $\mathcal{C}$ be an asymptotically flat hypersurface with "interior boundary" Σ. Integrating eq.(22) over $\mathcal{C}$, taking into account eqs. (16), (17), and (21) together with the fact that ξ^a vanishes on Σ, we obtain

$$\delta \int_\Sigma \mathbf{Q} = \delta\mathcal{E} - \Omega_H^{(\mu)} \delta\mathcal{J}_{(\mu)} \tag{23}$$

Equation (23) corresponds precisely to the first law of black hole mechanics as derived by Hamiltonian methods [3]. However, eq.(23) has the advantage over this previous derivation that the surface term arising from the black hole has now been explicitly identified as the variation of the Noether charge of Σ.

Equation (23) still is not of the desired form in the sense that the left side of eq.(23) has not yet been written as κ times the variation of a local, geometrical quantity on Σ, since $\mathbf{Q}$ is locally constructed from ξ^a and its derivatives as well as from the fields appearing in the Lagrangian. However, we now will show that the desired form of the first law holds when we further restrict attention to the case where $\delta\phi$ describes a perturbation to a nearby stationary black hole. First, we note that any derivative, $\nabla_{a_1}...\nabla_{a_n}\xi^b$ of any Killing field ξ^a can be re-expressed in terms of a linear combination of ξ^a and its first derivative, $\nabla_a\xi_b$, with coefficients depending upon the Riemann curvature and its derivatives (see, e.g., eq.(C.3.6) of [10]). Next, we note that on Σ we have $\xi^a = 0$ and $\nabla_a\xi_b = \kappa\epsilon_{ab}$, where ϵ_{ab} denotes the bi-normal to Σ. Now define the $(n-2)$-form $\tilde{\mathbf{Q}}$ on Σ by the following algorithm: Express $\mathbf{Q}$ in terms of ξ^a and $\nabla_a\xi_b$ by eliminating the higher derivatives of ξ^a, as described above. Then set $\xi^a = 0$ and replace $\nabla_a\xi_b$ by ϵ_{ab}. Since any reference to ξ^a has been eliminated, we see that $\tilde{\mathbf{Q}}$ is locally constructed out of the fields appearing in $\mathbf{L}$. Furthermore, since $\tilde{\mathbf{Q}}$ on Σ is determined by a well defined algorithm whose only input is a Lagrangian $\mathbf{L}$ which is invariant under diffeomorphisms of the dynamical fields, ϕ, (see eq.(2) above) it follows that $\tilde{\mathbf{Q}}$ is similarly invariant under spacetime diffeomorphisms which map Σ into itself. Thus, $\tilde{\mathbf{Q}}$ is a "local, geometrical quantity" on Σ. Finally, it is worth noting that $\tilde{\mathbf{Q}}$ is just the Noether charge $(n-2)$- form associated with the Killing field $\tilde{\xi}^a = \kappa \xi^a$, i.e., $\tilde{\xi}^a$ is the horizon Killing field normalized so as to have unit surface gravity.

Now identify the unperturbed and perturbed stationary black hole spacetimes in such a way that the Killing horizons of the two spacetimes coincide, and the unit surface gravity horizon Killing fields $\tilde{\xi}^a$ coincide in a neighborhood of the horizons. (That this always can be done follows from the general formula for Kruskal-type coordinates given in [12]. Note, however, that for a perturbation which changes the surface gravity, we cannot identify the two spacetimes so that the two horizon Killing fields coincide on the horizon when normalized via eq.(21); in addition, since we take $\delta t^a = \delta \varphi^a_{(\mu)} = 0$ near infinity, for a perturbation which changes $\Omega_H^{(\mu)}$, the requirement that $\delta \xi^a = 0$ precludes us from choosing ξ^a even to be proportional to the horizon Killing field near infinity in the perturbed spacetime.) Then, on Σ we have,

$$\delta \mathbf{Q} = \kappa \delta \tilde{\mathbf{Q}} \tag{24}$$

where κ is the surface gravity of the unperturbed black hole. Hence, for perturbations to nearby stationary black holes, the first law of black hole mechanics (23) takes the form

$$\frac{\kappa}{2\pi} \delta S = \delta \mathcal{E} - \Omega_H^{(\mu)} \delta \mathcal{J}_{(\mu)} \tag{25}$$

where the “black hole entropy”, S, is defined by

$$S = 2\pi \int_\Sigma \tilde{\mathbf{Q}} \tag{26}$$

Thus, we have established the existence and “local, geometrical character” of the notion of black hole entropy, S, in a general theory of gravity [18].

Our “local, geometrical” formula (26) for the entropy of a stationary black hole suggests the following generalization to the non-stationary case: For an arbitrary cross-section, Σ', of the horizon of a non-stationary black hole, construct $\tilde{\mathbf{Q}}$ by exactly the same mathematical algorithm as used above for the bifurcation surface, Σ, of a Killing horizon. Then 2π times the integral of $\tilde{\mathbf{Q}}$ over Σ' yields a candidate expression for the black hole entropy at “time” Σ'. The viability of this proposed definition is presently under investigation [8].

Note that eq.(25) has been derived only for the case of perturbations to nearby stationary black holes, even though eq.(23) holds in the more general case of non-stationary perturbations. However, since $\delta\xi^a = 0$ and, on Σ, we have $\xi^a = 0$, it follows that $\delta[\nabla_b \xi^a] = 0$ on Σ. From this, it follows that for non-stationary perturbations, we have, on Σ,

$$\delta[\nabla_{[a}\xi_{b]}] = \kappa \delta\epsilon_{ab} + w_{ab} \tag{27}$$

where, as before, ϵ_{ab} denotes the binormal to Σ, and w_{ab} is purely "normal-tangential", i.e., it vanishes when both of its indices are projected into Σ or both projected normal to Σ. It then follows from the existence of a reflection isometry about Σ (see lemma 2.3 of [13]) that the w_{ab}-term makes no contribution to the variation of $\mathbf{Q}$. It then can be seen that for sufficiently "low derivative" theories where $\mathbf{Q}$ depends only upon ξ^a and its first antisymmetrized derivative (as occurs, in particular, in general relativity and in $(1+1)$-dimensional theories), eq.(24) continues to hold for non-stationary perturbations. Thus, in such theories, the first law of black hole mechanics continues to hold in the form (25), with S given by (26). The nature of the first law for non-stationary perturbations in more general theories is presently under investigation [8].

The fact that for stationary black holes, S is just 2π times the Noether charge of the horizon Killing field (normalized to have unit surface gravity) implies that for an initially stationary black hole which undergoes a dynamical process and later "settles down" to a stationary final state, the net change in black hole entropy is just the total flux through the horizon of Noether current associated with a suitable "time translation" on the horizon. This suggests a possible relationship between the validity of the second law of black hole mechanics in a theory and positive energy properties of that theory. It also suggests some possible approaches toward establishing (or disproving) the second law in general theories of gravity. These issues also are under investigation [8].

Finally, we consider the relationship between the results of this paper and the formula for black hole entropy obtained via the "Euclidean approach" in the manner first given in [7]. We begin by noting that since $\mathcal{L}_\xi \phi = 0$, the Noether current (7) associated with the horizon Killing field, ξ^a, of a stationary black hole is simply

$$\mathbf{j} = -\xi \cdot \mathbf{L} \tag{28}$$

Let $\mathcal{C}$ be an asymptotically flat hypersurface with "interior boundary" Σ. Integrating eq.(28) over $\mathcal{C}$ and taking into account eqs.(9), (21), (19), and (20), we obtain,

$$\mathcal{E} - \Omega_H^{(\mu)} \mathcal{J}_{(\mu)} - \int_\Sigma \mathbf{Q} = -\int_\mathcal{C} \xi \cdot \mathbf{L} - \int_\infty t \cdot \mathbf{B} \tag{29}$$

Now, the "Euclidean action", I, corresponds to

$$I = -\frac{2\pi}{\kappa}[\int_\mathcal{C} \xi \cdot \mathbf{L} + \int_\infty t \cdot \mathbf{B}] \tag{30}$$

More precisely, in the static case, the right side of eq.(30) equals what would be obtained by integrating the suitably analytically continued Lagrangian, $\mathbf{L}' = \mathbf{L} + d\mathbf{B}$, over a "Euclidean section", constructed by replacing the Killing parameter, t, by $\tau = it$, and then periodically identifying τ with period $2\pi/\kappa$ (see, e.g., [19] for further details). In the stationary but non-static case, there is no such thing as a "Euclidean section", but the right side of eq.(30) corresponds to what researchers mean by the "Euclidean action" in that case. Thus, we obtain the following formula for I,

$$\frac{\kappa}{2\pi} I = \mathcal{E} - \Omega_H^{(\mu)} \mathcal{J}_{(\mu)} - \int_\Sigma \mathbf{Q} \tag{31}$$

Now, in the Euclidean approach, $\frac{\kappa}{2\pi} I$ is identified as the thermodynamic potential of the black hole [7]. This leads immediately to the the following formula for black hole entropy,

$$\begin{aligned} \frac{\kappa}{2\pi} S &= \mathcal{E} - \Omega_H^{(\mu)} \mathcal{J}_{(\mu)} - \frac{\kappa}{2\pi} I \\ &= \int_\Sigma \mathbf{Q} \end{aligned} \tag{32}$$

which agrees with eq.(26). Thus, we have shown that the "Euclidean procedure" for obtaining black hole entropy gives the same result as obtained by our method.

This research was supported in part by National Science Foundation grant PHY-9220644 to the University of Chicago.

References

[1] S.W. Hawking, Commun. Math. Phys. **43**, 199 (1975).

[2] J.M. Bardeen, B. Carter, S.W. Hawking, Commun. Math. Phys. **31**, 161 (1973).

[3] D. Sudarsky and R.M. Wald, Phys. Rev. D **46**, 1453 (1992); R.M. Wald, " The first law of black hole mechanics" in *Directions in General Relativity*, vol. 1, ed. by B.L. Hu, M. Ryan, and C.V. Vishveshwara (Cambridge University Press, Cambridge, 1993). See also J.D. Brown and J.W. York Phys. Rev. **D47**, 1407 and 1420 (1993).

[4] V.P. Frolov, Phys. Rev. **D46**, 5383 (1992).

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

[6] I wish to thank Ted Jacobson for bringing this deficiency to my attention.

[7] G.W. Gibbons and S.W. Hawking, Phys. Rev. **D15**, 2752 (1977).

[8] V. Iyer and R.M. Wald, to be published.

[9] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973).

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

[11] It is proven in [12] that in an arbitrary theory of gravity, global nonsingularity of the event horizon of a stationary black hole implies that the zeroth law holds. However, there are no compelling physical grounds to impose such a global nonsingularity condition, since for a black hole formed by the collapse of matter, the parallelly propagated curvature singularity found in [12] would be "covered up" by the collapsing matter.

[12] I. Racz and R.M. Wald, Class and Quant. Grav., **9**, 2643 (1992).

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

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

[15] R.M. Wald, J. Math. Phys. **31**, 2378 (1990). The results of this reference also can be derived using the "free variational bicomplex"; see I.M. Anderson in *Mathematical Aspects of Classical Field Theory*, ed. by M. Gotay, J. Marsden, and V. Moncrief, Cont. Math. **132**, 51 (1992).

[16] This quantity corresponds to the "gravitational field strength" discussed in W. Simon, Gen. Rel. and Grav. **17**, 439 (1985).

[17] T. Regge and C. Teitelboim, Ann. Phys. **88**, 286 (1974).

[18] A much more indirect argument for this conclusion has been given independently by J. Simon and B. Whiting (to be published), based upon the Euclidean approach. See also, M. Visser (to be published).

[19] S.W. Hawking, "The Path Integral Approach to Quantum Gravity", in *General Relativity, an Einstein Centennary Survey*, ed. by S.W. Hawking and W. Israel (Cambridge University Press, Cambridge, 1979)