# Thermodynamics of spacetime in generally covariant theories of gravitation

Christopher Eling *
*Department of Physics, University of Maryland,*
*College Park, MD 20742-4111, USA*
draft of a paper for publication with Ted Jacobson and Raf Guedens

May 24, 2005

**Abstract**

It has been shown that the Einstein equation can be derived from the requirement that the First Law $\delta Q = TdS$ holds for all local Rindler causal horizons through each spacetime point. Here $\delta Q$ and $T$ are the energy flux and Unruh temperature seen by an accelerating observer just inside the horizon, and the entropy $S$ is one quarter the horizon area in Planck units. In this paper we show that there are problems extending this thermodynamic derivation to include generic diffeomorphism invariant theories of gravitation, where the entropy is the integral of the Noether charge over the horizon. The requirement of thermodynamic equilibrium is demonstrated to be incompatible with the local stationarity of the horizon. In scalar-tensor theories this obstruction can be remedied with a conformal transformation of the metric, but this method fails to work in higher curvature theories. We conclude by using effective field theory to argue that the thermodynamic derivation should only apply in the lowest order regimes of the effective gravitational action.

## 1 Introduction

The deep connection between gravitation, thermodynamics, and quantum mechanics was first revealed by the discovery in the early 1970's of the four laws of black hole mechanics.[1] At an early stage theorists realized that these laws, which are derived from the Einstein field equation, are essentially analogous to the well known four laws of thermodynamics. The discovery of the Bekenstein-Hawking entropy and Hawking radiation [2, 3] dispelled the notion that this was only an analogy because it demonstrated that black holes do behave like thermodynamic objects with a physical temperature and entropy. However, the questions of how and why this identity relating two different branches of physics

*cteling@physics.umd.edu

exists have been the source of many investigations over the past 30 years. Several years ago, one of us proposed [4] an answer by reversing the logic and using the proportionality of entropy and horizon area along with the First law of thermodynamics to derive the Einstein equation as an equation of state. This derivation strongly supports the idea that if we know the entropy from the statistical mechanics of some yet undiscovered theory of the microscopic structure of spacetime then gravitation on a macroscopic scale must be a manifestation of thermodynamics.

In [4] it was briefly argued that changing the assumed entropy functional would simply change the field equations implied by the first law $\delta Q = TdS$. As a simple example, if the entropy was given by a polynomial in the Ricci scalar $\alpha_0 + \alpha_1 R + \alpha_2 R^2 + \cdots$ then the implied field equation would be consistent with those found by the variation of a Lagrangian polynomial in the Ricci scalar. In the first section of this paper we begin to examine this claim in detail by briefly reviewing the original thermodynamical field equation argument. In the following section we consider if the argument can be extended to include more general diffeomorphism invariant theories of gravitation. In [5, 6] Wald and Iyer showed that the entropy of a stationary black hole in any generally covariant theory is given by the integral of its Noether charge of diffeomorphisms over the (Killing) horizon. Using this as a prescription for the assumed entropy functional we attempt to run the argument for the specific cases of actions with an $f(R)$ term and scalar-tensor theories. In the case where the entropy is just proportional to the area, the condition for thermodynamic equilibrium is given by vanishing of the expansion and shear of the local Rindler horizon at a point. However, for these more general functionals, the expansion must be chosen to cancel off extra terms such as the change of $R$ or the scalar field along the horizon. This requirement turns out to be generally inconsistent with the field equations

In the third section we consider a couple of ways to possibly repair the argument. Perhaps the Wald prescription for the entropy is not correct in this non-stationary situation. For example, in the case of a dynamical horizon the ambiguities inherent in the Noether charge formalism become important. However, this point of view seems flawed since the concept of a local Rindler horizon appears to be essential. Second, one could make a conformal transformation that converts the entropy to just an area in the new metric. Although the equilibrium condition is again trivial, we can obtain the scalar-tensor field equations (in the "Einstein frame"). The similar procedure for the $f(R)$ theory turns out to be somewhat unsatisfactory and it cannot be generalized to other higher curvature terms. Thus, we conclude that the thermodynamical argument only applies to lowest order regimes of the effective gravitational action. From the viewpoint of effective field theory higher order generally covariant terms are suppressed by a cutoff (the Planck scale). We argue that an accelerating observer measuring entropy in the local Rindler frame is incapable of probing to the length scales required to detect additional curvature terms.

## 2 Local Rindler Horizon construction in GR

As discussed in [4], the first fundamental steps of the thermodynamic derivation are to define heat, temperature, and the equilibrium conditions in the context of spacetime dynamics. A natural definition for heat is the stress-energy that flows across a causal horizon. In order to establish generality, in this paper the idea of a causal horizon is not exclusively associated with black holes, but is instead simply the boundary of the past of any set $O$. Thus, it is a null hypersurface with the null geodesic horizon generators emanating backward in time from the set $O$. We want to consider the local gravitational thermodynamics associated with these causal horizons, where the system is made up of the degrees of freedom just in front of the horizon. The temperature of this system is defined via the Unruh effect, [7] which states that the quantum fluctuations of the vacuum have thermal characteristics when observed by a uniformly accelerating observer. Since different accelerations will be associated with different temperatures, we choose the thermodynamics of the accelerated observer whose worldline approaches the horizon in the limit in order to make the arguments as local as possible.

Because the system in question is generally not static, meaning that the horizon is generally expanding and shearing, we must define what we mean by equilibrium. We use the Equivalence Principle in order to consider the small neighborhood around any spacetime point $p$ as a piece of flat spacetime. There exists a spacelike 2-surface $B$ containing $p$ such that at least one congruence of null geodesics orthogonal to $B$ has vanishing expansion and shear at $p$. This congruence spans the local Rindler horizon (LRH) at $p$, which defines the part of spacetime beyond the horizon to be the system in local equilibrium. In this small region of flat spacetime there is an approximate Killing field $\chi^a$ associated with boost symmetries. The Unruh effect predicts that the quantum fluctuations of the vacuum have a temperature of

$$T_u = \frac{\kappa}{2\pi} \tag{1}$$

in units where $\hbar = c = 1$ and the surface gravity $\kappa$ is the acceleration of the Killing orbit on which $\chi^a$ is normalized to unity. In addition, the conserved stress-energy current $F^a = T^{ab}\chi_b$ defines the heat flow, where $T_{ab}$ is the matter stress-energy tensor. Letting $\chi^a$ generate the local Rindler horizon with the direction chosen to be future pointing to the inside past of $p$, we can write the heat flux across the horizon to be

$$\delta Q = -\int T^{ab}\chi_b \, d\Sigma_a. \tag{2}$$

If we assume $k^a$ is a tangent vector to the horizon generators then the LRH can be parameterized by the usual affine parameter $\lambda$. On the initial surface $B$ we choose $\lambda = 0$ and coordinates $\{x^1, x^2\}$, leading to the coordinate system $\{x, \lambda\}$ on the LRH. We can use the equation $\chi^a = -\kappa\lambda k^a$ implied by the equation for surface gravity [8] and $d\Sigma_a = k_a d\lambda \; d^2x$ to write the heat flux to the past of $p$

as

$$\delta Q = \int_0^\epsilon d\lambda \int d^2x\sqrt{h(\lambda)}\; F_a k^a \tag{3}$$

$$= 2\pi T_u \int d^2x\sqrt{h(0)}\; T_{ab}k^a k^b(0)\frac{\epsilon^2}{2} + O(\epsilon^3) \tag{4}$$

where $h$ is the determinant of the induced two metric and we have truncated the LRH with a second surface $B(\epsilon)$, where $\epsilon$ is small.

On the other hand, the area of the LRH is given by

$$A(\lambda) = \int_{B(\lambda)} d^2x\sqrt{h(x,\lambda)}. \tag{5}$$

Under a small change we find that (expanding in a power series)

$$\Delta A = -\int d^2x\sqrt{h(0)}\;\frac{d\theta}{d\lambda}(0)\frac{\epsilon^2}{2} + O(\epsilon^3) \tag{6}$$

where $\theta = d(\ln\sqrt{(h)})/d\lambda$ is the expansion of the congruence. We will assume that $S = \eta A$, where $S$ is the entropy and $\eta$ is an unknown proportionality constant. Thus the 1st law $\delta Q = T_u\; \eta dA$ in the limit $\epsilon \to 0$ is equivalent to

$$-\frac{d\theta}{d\lambda}(0) = \frac{2\pi}{\eta}T_{ab}k^a k^b(0). \tag{7}$$

Using the Raychaudhuri equation

$$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \sigma^2 - R_{ab}k^a k^b \tag{8}$$

evaluated on the initial surface $B$, Eqn. 7 becomes

$$R_{ab}k^a k^b(0) = \frac{2\pi}{\eta}T_{ab}k^a k^b(0). \tag{9}$$

Since the First Law should hold for all LRHs through every spacetime point Eqn. 9 implies that

$$R_{ab} + f g_{ab} = \frac{2\pi}{\eta}T_{ab} \tag{10}$$

for some function f. Imposing local conservation of energy-momentum, $\nabla^a T_{ab} = 0$, determines f and yields the equation

$$G_{ab} + \Lambda g_{ab} = \frac{2\pi}{\eta}T_{ab}. \tag{11}$$

This is of course the Einstein equation plus cosmological constant if we fix $\eta = \frac{1}{4G}$.

# 3 More general theories

## 3.1 f(R) Theory

In this section we consider whether the thermodynamic argument can produce the field equations for any generally covariant theory of gravitation. This is not a far-fetched idea since it has been shown that a version of the First Law is valid for stationary black hole solutions in higher curvature gravity [9]. In this situation, the key question is what to use as the entropy functional since it will no longer be just the horizon area. Of course, in order to avoid circularity in the original argument we are forced to argue that horizons must be associated with entropy because they tend to hide information from the outside world. This "entanglement entropy" that involves relations between vacuum fluctuations inside and outside of the horizon must be proportional to the area of the horizon if there is some fundamental length scale (e.g. the Planck length). In a more general theory one could assume an observer in the local Rindler frame would detect the appropriate additional terms. As a first test we consider an entropy density proportional to $f(R)$. The change in this entropy functional when moving a small affine parameter $\epsilon$ along the local horizon reads

$$\Delta S = S(0) - S(\epsilon) = \eta\{\int_{B(0)} d^2x\sqrt{h(0)}f(R(0)) - \tag{12}$$

$$\int_{B(\epsilon)} d^2x\sqrt{h(\epsilon)}f(R(\epsilon))\}. \tag{13}$$

Expanding in a power series in $\epsilon$

$$\Delta S = -\eta\int_{B(0)} d^2x\sqrt{h(0)}\{(\theta(0)f(0) + \frac{df}{d\lambda}(0))\epsilon + \tag{14}$$

$$(\frac{d^2f}{d\lambda^2}(0) + 2\theta(0)\frac{df}{d\lambda}(0) + f(0)(\theta^2(0) + \frac{d\theta}{d\lambda}(0)))\frac{\epsilon^2}{2} + O(\epsilon^3)\}. \tag{15}$$

At $\lambda = 0$ vanishing shear and local stationarity of the entropy are two requirements for equilibrium. However, the vanishing of the 1st order term [1]

$$\theta f + \frac{df}{d\lambda} = 0, \tag{16}$$

corresponding to stationarity of the entropy, is not consistent with zero horizon expansion (unless f = constant). In addition, the energy flux term is still given by Eqn. 4, which at lowest order is quadratic in $\epsilon$. Despite this difficulty, we proceed and find that in order for the First law $\delta Q = T_u dS$ to be valid in the limit $\epsilon \to 0$ Eqn. 16 and the quadratic order condition

$$\frac{d^2f}{d\lambda^2} + 2\theta\frac{df}{d\lambda} + f(\theta^2 + \frac{d\theta}{d\lambda}) = -\frac{2\pi}{\eta}T_{ab}k^ak^b. \tag{17}$$

[1] dropping (0) notation for brevity

must be satisfied. Enforcing Eqn. 16 and applying it to Eqn. 17 gives

$$\frac{d^2 f}{d\lambda^2} - \theta^2 f + f\frac{d\theta}{d\lambda} = -\frac{2\pi}{\eta} T_{ab} k^a k^b. \tag{18}$$

We now use the Raychaudhuri equation evaluated at the initial surface and find that

$$\frac{d^2 f}{d\lambda^2} - f\frac{3}{2}\theta^2 + f R_{ab} k^a k^b = -\frac{2\pi}{\eta} T_{ab} k^a k^b. \tag{19}$$

where $\theta = -\frac{df/d\lambda}{f}$. Is this equation consistent with the equations of motion generated by varying the action associated with an entropy density of $f(R)$? To find the action we must look to the work of Wald and Iyer [5, 6]. For any diffeomorphism invariant theory admitting a stationary black hole solution, they have established a first law of black hole mechanics

$$T dS = \Delta M \tag{20}$$

where the entropy $S$ is given by the Noether charge of diffeomorphisms under the Killing vector that generates the horizon. Since the local Rindler horizon is approximately stationary, this appears to be a good prescription for the entropy. In the case of a fully dynamical horizon there is no accepted definition. For the higher curvature theory given by the action

$$S = \frac{1}{16\pi G} \int d^4x \sqrt{-g}\; g(R) + I_{matter} \tag{21}$$

the Noether charge formalism tells us that the entropy density is $g'(R)$, where the prime stands for $\partial/\partial R$. The equations of motion that follow from the variation of this action are

$$g' R_{ab} + (\Box g' - \frac{1}{2} g) g_{ab} - \nabla_a \nabla_b g' = 8\pi G T_{ab}. \tag{22}$$

Contraction with $k^a k^b$ and evaluating at $\lambda = 0$ produces

$$g' R_{ab} k^a k^b - \frac{d^2 g'}{d\lambda^2} = 8\pi G T_{ab} k^a k^b. \tag{23}$$

Thus, Eqn. 19 and Eqn. 23 are consistent up to a $\theta^2 f$ term, identifying $f = g'$ and again $\eta = \frac{1}{4G}$. The two equations are equivalent only if $\theta = 0$, meaning that all the original equilibrium conditions in Section 2 hold exactly. Unfortunately, in Eqn. 16 this in turn implies that

$$\frac{dg'}{d\lambda} = g'' \frac{dR}{d\lambda} = 0. \tag{24}$$

Since the First Law should apply for all LRHs through every spacetime point the scalar curvature must be a constant, which is generally far too restrictive. The only exact solutions permitted by the thermodynamical argument in these theories are those with constant curvature, which can exist only in vacuum. Note however that the $f\theta^2 = (df/d\lambda)^2/f$ term in Eqn. 19 is higher order relative to the other terms in an effective field theory expansion. Assume that

$$g'(R) = 1 + L_p^2 R + \cdots \tag{25}$$

where $L_p$ is the Planck length. Then $g(R)$ is given by

$$g(R) = R + L_p^2 R^2 + \cdots \tag{26}$$

Expanding the equation of motion (Eqn. 22) we find

$$G_{ab} + (RR_{ab} + (\Box R - \tfrac{1}{2}R)g_{ab} - \nabla_a \nabla_b R)L_p^2 + \cdots \tag{27}$$

However, expanding the offending $\theta^2 f$ term

$$\frac{1}{g'}\nabla_a g' \nabla_b g' \sim L_p^4 \frac{1}{1 + L_p^2 R}\nabla_a R \nabla_b R + \cdots \tag{28}$$

shows that it is of order $L_p^4$. Thus, although the thermodynamic derivation does not exactly reproduce the correct equations of motion, in the two lowest orders of the Planck length expansion there is agreement. This agreement at two orders surprising because we assumed non-zero horizon expansion. The agreement may be an artifact of the approximations involved in the LRH construction and will be a subject of future work.

## 3.2 Scalar-tensor theories

For a scalar-tensor theory the analysis is very similar. We assume that the observer in the local Rindler frame detects an entropy density proportional to some scalar field $\varphi$. Replacing $f(R)$ with $\varphi$ in the equations above and enforcing the equilibrium and quadratic conditions, we arrive at the equation

$$\frac{d^2\varphi}{d\lambda^2} - \frac{3}{2}\frac{(d\varphi/d\lambda)^2}{\varphi} + \varphi R_{ab}k^a k^b = -\frac{2\pi}{\eta}T_{ab}k^a k^b. \tag{29}$$

We now need to check whether this equation is consistent with the equation of motion derived from the theory associated with entropy density $\varphi$. A general scalar-tensor action

$$I = \int d^4x\sqrt{-g}\{\varphi R - \frac{\omega}{\varphi}\nabla^a\varphi\nabla_a\varphi + V(\varphi)\} + I_m(g_{ab}, \varphi) \tag{30}$$

has an entropy density proportional to $\varphi$ according to the Noether charge prescription. The metric field equation for this theory is

$$G_{ab} - \tfrac{1}{2}g_{ab}V(\varphi) = -\frac{1}{\varphi}T_{ab} - \frac{1}{\varphi}[\nabla_a\nabla_b - g_{ab}\Box]\varphi + \frac{\omega}{\varphi^2}\nabla_a\varphi\nabla_b\varphi - \tfrac{1}{2}\frac{\omega}{\varphi^2}g_{ab}(\nabla\varphi)^2 \tag{31}$$

Contracting with $k^a k^b$ and evaluating at $\lambda = 0$ yields

$$\varphi R_{ab}k^a k^b = -T_{ab}k^a k^b - d^2\varphi/d\lambda^2 + \frac{\omega}{\varphi}(d\varphi/d\lambda)^2, \tag{32}$$

which is consistent if $\omega = 3/2$. We will now show that the thermodynamic argument can derive the appropriate terms proportional to $g_{ab}$ in the equation motion. Since Eqn. 29 holds for all LRH's at every spacetime point, we have

$$\nabla_a\nabla_b\varphi - \frac{3}{2\varphi}\nabla_a\varphi\nabla_b\varphi + \varphi R_{ab} + f g_{ab} = -\frac{2\pi}{\eta}T_{ab}. \tag{33}$$

The reader may also be wondering how this type of derivation can produce the two equations of motion, one for the metric and one for the scalar field. The condition that helps to answer this question and allows us to solve for $f$ uniquely up to a constant is the local conservation of stress-energy. The matter action is invariant under infinitesimal changes of coordinates. Assuming that the matter fields satisfy the equations of motion $dS_{\text{matter}}/d\psi = 0$ "on-shell," the invariance of this action yields the generalized conservation law

$$\nabla^b T_{ab} = -\tfrac{1}{2}T_\varphi\nabla_a\varphi, \tag{34}$$

where $T_{ab} = \frac{-1}{\sqrt{(-g)}}\delta S_{\text{matter}}/\delta g_{ab}$ and $T_\varphi = \frac{1}{\sqrt{(-g)}}\delta S_{\text{matter}}/\delta\varphi$. Applying the conservation law, the commutator of covariant derivatives, and the contracted Bianchi identity $\nabla^b R_{ab} = \frac{1}{2}\nabla_a R$, we obtain the expression

$$\nabla_a f = \frac{\pi}{\eta}T_\varphi\nabla_a\varphi + [\nabla_a(\Box\varphi) - \frac{3}{2}\frac{1}{\varphi}\nabla^b(\nabla_a\varphi\nabla_b\varphi) + \frac{3}{2}\frac{1}{\varphi^2}(\nabla\varphi)^2\nabla_a\varphi - \tfrac{1}{2}\varphi\nabla_a R] \tag{35}$$

The goal of our future calculations will be to change all the terms in Eqn. 35 into either the gradient of a scalar or some function times the gradient of $\varphi$. Thus, after several lines of calculation, we find that

$$(\frac{\pi}{\eta}T_\varphi + \frac{3}{4}\frac{1}{\varphi^2}(\nabla\varphi)^2 - \frac{3}{2}\frac{1}{\varphi}(\Box\varphi) + \tfrac{1}{2}R)\nabla_a\varphi = \nabla_a(f + \tfrac{1}{2}\varphi R - \Box\varphi + \frac{3}{4\varphi}(\nabla\varphi)^2). \tag{36}$$

Since the right hand side is the gradient of a scalar, the left hand side clearly must be as well. This implies that for some scalar function $V$,

$$\nabla_a(f + \tfrac{1}{2}\varphi R - \Box\varphi + \frac{3}{4\varphi}(\nabla\varphi)^2) = \tfrac{1}{2}\nabla_a V, \tag{37}$$

which gives $f = \frac{1}{2}V - \frac{1}{2}\varphi R - \frac{3}{4\varphi}(\nabla\varphi)^2 + \Box\varphi + \Lambda$, where $\Lambda$ is the integration constant and plays the role of the cosmological constant if we do not wish to absorb it into $V$. The resulting metric equation of motion is

$$\frac{2\pi}{\eta}T_{ab} = \varphi R_{ab} + \frac{3}{2\varphi}\nabla_a\varphi\nabla_b\varphi - \nabla_a\nabla_b\varphi + \tfrac{1}{2}g_{ab}V(\varphi) - \tfrac{1}{2}g_{ab}\varphi R - \frac{3}{4\varphi}g_{ab}(\nabla\varphi)^2 + g_{ab}(\Box\varphi) \tag{38}$$

Note that our equation

$$\nabla_a V = (\frac{2\pi}{\eta}T_\varphi + \frac{3}{2}\frac{1}{\varphi^2}(\nabla\varphi)^2 - \frac{3}{\varphi}(\Box\varphi) + R)\nabla_a\varphi \qquad (39)$$

has the general form $\nabla_a V = q\nabla_a\varphi$ for some arbitrary function $q$. In differential forms notation, this equation can be written as $dV = qd\varphi$ and one can show that locally $V$ is only a function of $\varphi$. [2] Thus, we have the scalar field equation of motion

$$V'(\varphi) = \frac{2\pi}{\eta}T_\varphi + \frac{3}{2}\frac{1}{\varphi^2}(\nabla\varphi)^2 - \frac{3}{\varphi}(\Box\varphi) + R \qquad (40)$$

appearing as an auxiliary condition on the potential $V(\varphi)$. Eqns. 38 and 40 are the equations of motion that result when one varies the scalar-tensor action in Eqn. 30 with $\omega = 3/2$ with respect to $\varphi$ and the metric.

Although a more general entropy density does produce equations that agree with a very special case of a scalar-tensor theories and are consistent to lowest order in the $f(R)$ theory, generically it appears there is an inconsistency. We also assumed a non-zero expansion initially to maintain stationarity of the entropy, which is again inconsistent with the LRH construction. If we had fixed $\theta = 0$ the thermodynamic argument would have (like the $f(R)$ case) restricted the theory to special solutions with $\varphi = const$. In general, it seems that this obstruction exists for any form of the entropy as a Noether charge since one would have to balance the expansion against the change of the entropy density functional along the horizon.

# 4 Conformal Transformations and other possible remedies

One immediate objection to the analysis of the previous section is the simple formal replacement of the area with the Wald prescription for the entropy of a stationary black hole. The Noether charge formalism contains several ambiguities. If the Lagrangian for the theory is given by an $n$-form $\mathbf{L}_{abcd} = L(\psi_m, \nabla_a\psi_m, g_{ab}, R_{abcd}, \nabla_e R_{abcd})\epsilon_{abcd}$ then there is a $(n-2)$-form Noether charge of diffeomorphisms $\mathbf{Q}_{ab} = Q^{cd}\epsilon_{abcd}$. In [6] the Noether charge (n-2)-form can be expressed in the following way

$$\mathbf{Q} = \mathbf{W}_c\xi^c + \mathbf{X}^{cd}(\varphi)\nabla_{[c}\xi_{d]} + \mathbf{Y}(\varphi, \mathcal{L}_\xi\varphi) + \mathbf{dZ}(\varphi, \xi) \qquad (41)$$

where $\xi$ is the generator of diffeomorphisms, $\varphi$ is some field configuration and indices have been suppressed on the differential forms. There are ambiguities in $\mathbf{W}_c$,$\mathbf{X}^{cd}$,$\mathbf{Y}$,and $\mathbf{dZ}$ arising from the fact that one can add an exact form to the Lagrangian without changing the equations of motion and that the Noether current and charge are only defined up to the addition of a closed form (see [10], for example). Fortunately, it has been shown that only the $\mathbf{X}^{cd}$ term

[2] note that this may not be true globally

contributes when the Noether charge is integrated over a stationary horizon. It can be chosen to take the form

$$S = -2\pi \oint (Y^{abcd} - \nabla_e Z^{e:abcd})\epsilon_{ab}\epsilon_{cd} \tag{42}$$

where $\epsilon_{ab}$ is the binormal to the cross-section, $Y^{abcd} = \partial L/\partial R_{abcd}$ and $Z^{e:abcd} = \partial L/\partial \nabla_e R_{abcd}$ [10]. In the case where the horizon is fully dynamical there is no accepted way to eliminate the ambiguities since there is no longer a Killing horizon. However, since the Wald proof does hold for stationary horizons there is good reason to expect that it could be extended in some way to the *locally* stationary LRH, which is generated by the approximate Killing field $\chi^a$ generating boosts in some direction. In the general case we showed in Section 3 that local thermodynamic equilibrium does not even coincide with local stationarity of the horizon. Although this gives correct field equations for the scalar-tensor theory and the $f(R)$ theory at lowest order, without local stationarity the LRH construction is not possible. Thus, there is no approximate Killing field and without this quantity clear definitions of heat flow and temperature do not appear to be possible.

Another possible way to restore consistency is to make a conformal transformation of the metric such that the horizon entropy is given by the area in the new metric. Let us consider the scalar-tensor theory Eqn. 30 whose entropy density is given by $\varphi$. Since there exists a conformal frame freedom in general scalar-tensor theories [11] it is likely that the transformation will have useful results. Making the conformal transformation

$$\widetilde{g}_{ab} = \varphi\, g_{ab} \tag{43}$$

absorbs the $A(\varphi)$ term in the new entropy

$$S' = \frac{1}{4G}\int \sqrt{\widetilde{h}}\; d^2x \tag{44}$$

Now local stationarity of the entropy functional again coincides with the local stationarity of the horizon and we have effectively transformed to the "Einstein frame" of the theory. Running the argument yields

$$\widetilde{R}_{ab}\widetilde{k}^a\widetilde{k}^b \tag{45}$$

on one side of the First Law equation. Any $\varphi$ dependent terms now have to be considered a part of the heat flow across the horizon. An observer in the local Rindler frame may have knowledge of the matter action. Assuming the scalar field and other matter fields couple minimally to the metric (any curvature couplings would contribute to the entropy), $I_m = I_m(\widetilde{g}_{ab}, \varphi, \psi_m)$. Note that $\widetilde{g}_{ab}$ is not yet a dynamical quantity. Therefore the argument produces

$$\widetilde{G}_{ab} = 8\pi G(\widetilde{T}^{\varphi}{}_{ab} + \widetilde{T}^{m}{}_{ab}) \tag{46}$$

where $T^{\varphi}{}_{ab}$ is the stress tensor of a minimally coupled scalar field while $T^{m}{}_{ab}$ corresponds to other types of matter fields. With this knowledge of the matter action we also have access to the scalar field and matter equations of motion. Thus, the First Law equation produced by the thermodynamical argument is consistent with the Einstein frame equations of motion.

For the $f(R)$ the situation becomes more complex. It is again possible to make a conformal transformation

$$\widetilde{g}_{ab} = f'(R) g_{ab} \tag{47}$$

so that the entropy becomes an area in the new metric. The usual procedure of course yields the contracted Einstein equation in the new metric. Re-writing this equation in the old metric we find (not unexpectedly)

$$R_{ab}k^a k^b - (df'^2/d\lambda^2)/f' + \frac{3}{2}(df'/d\lambda)^2/(f')^2 = 8\pi G T_{ab}k^a k^b(1/f') \tag{48}$$

which is exactly Eqn. 19. Thus, this equation cannot be made equal to the equation of motion for the $f(R)$ theory. In addition, we cannot legitimately assume the necessary correction terms are a part of the heat flow or in the matter action detected by the observer since they involve curvature. A possible loophole is to follow the procedure of (for example) Flanagan [12], which relates the action in Eqn. 21 to a scalar-tensor theory in the Einstein frame. This is done by introducing an auxiliary scalar field $\varphi$ such that on-shell $\varphi = R$. While this enables us to work with $f(R)$ terms in some sense, what about other higher curvature terms such as $R_{abcd}R^{abcd}$ that should appear in an effective action for gravity? (see [13]) In this case the entropy is

$$S = \oint R^{abcd}\epsilon_{ab}\epsilon_{cd}, \tag{49}$$

which depends on both the intrinsic and extrinsic geometry of the horizon. A conformal transformation therefore cannot bring this expression to just the horizon area, which involves only the intrinsic geometry.

One could attempt a series of field redefinitions and conformal transformations on individual terms, but this is not likely to succeed since there are terms cubic and higher in the Riemann tensor that appear in the general diffeomorphism invariant effective action. However, it is important to note that these terms are suppressed relative to the Einstein-Hilbert action by powers of the Planck length. Can an observer in the local Rindler frame ever detect such terms when measuring the entropy? Recall that in the original construction of the argument we invoked the equivalence principle to view a small patch of spacetime as flat with Rindler coordinates

$$g_{ab} = \eta_{ab} + O(R_{acbd}x^c x^d) \tag{50}$$

The Unruh temperature and heat flow are associated with the Killing vector of the flat spacetime region. In order to detect the effects of the suppressed curvature terms in the entropy the observer would have to probe to longer length scales. However, at these scales the approximation of a flat piece of spacetime, an associated LRH with vanishing expansion and shear, and an approximate Killing field begin to fail. Thus, it makes sense that the approximation of equilibrium thermodynamics begins to fail.

## 5 Discussion

In this paper we have shown that in general diffeomorphism invariant theories of gravity local stationarity of the entropy functional is not consistent with local stationarity of the associated horizon. In the case of higher curvature gravity we expect from effective field theory that the detection of additional Planck length suppressed terms in the entropy functional would require one to go beyond the conditions of equilibrium thermodynamics. Scalar-tensor theory appears to be a specialized case, where the use of a conformal transformation to the Einstein frame and the assumption that the heat flow includes the scalar field are legitimate procedures. In the future perhaps it will be possible to understand generally covariant theories in the context of non-equilibrium thermodynamics.

## References


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

[2] J.D. Bekenstein, Phys. Rev. D **7** 2333 (1973).

[3] S.W. Hawking, Comm. Math. Phys. **43** 199 (1975).

[4] T. Jacobson,"Thermodynamics of spacetime: The Einstein Equation of State," Phys. Rev. Lett. **75** 1260 (1995).

[5] R.M. Wald,"Black Hole Entropy is Noether Charge," Phys. Rev. **D**48 3427 (1993).

[6] R.M. Wald and V. Iyer,"Some Properties of Noether Charge and a Proposal for Dynamical Black Hole Entropy," Phys. Rev. **D**50 846 (1994)

[7] W.G. Unruh, Phys. Rev. D**14** 870 (1976).

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

[9] T. Jacobson, G. Kang, and R.C. Myers, Phys. Rev. D**52** 3518(1995).

[10] T. Jacobson, G. Kang, and R.C. Myers, "On Black Hole Entropy," Phys. Rev **49** 6587 (1994).

[11] E. Flanagan, "The Conformal Frame Freedom in theories of Gravitation," Class. Quant. Grav. **21** 3817 (2004).

[12] E. Flanagan, "Palatini Form of 1/R Gravity," Phys. Rev. Lett. **92** 071101 (2004).

[13] C.P. Burgess,"Quantum Gravity in Everyday Life: General Relativity as an Effective Field Theory," Living Rev. Rel. **7** 5 (2004).