PHYSICAL REVIEW D **78,** 061501(R) (2008)

$\mathbf{f}(R)$ gravity equation of state

Emilio Elizalde[1,*] and Pedro J. Silva[2,+]

[1]*Institut de Ciències de l'Espai (CSIC) and Institut d'Estudis Espacials de Catalunya (IEEC/CSIC) UAB, Ciències, Torre C5-Parell-2a planta E-08193 Bellaterra (Barcelona) Spain*

[2]*Institut de Ciències de l'Espai (IEEC-CSIC) and Institut de Física d'Altes Energies UAB, E-08193 Bellaterra (Barcelona) Spain*

(Received 22 April 2008; published 25 September 2008)

The $\mathbf{f}(R)$ gravity field equations are derived as an equation of state of local space-time thermodynamics. Jacobson's arguments are nontrivially extended, by means of a more general definition of local entropy, for which Wald's definition of dynamic black hole entropy is used, as well as the concept of an effective Newton constant for graviton exchange, which recently appeared in the literature.

DOI: 10.1103/PhysRevD.78.061501 PACS numbers: 04.20.Cv, 04.62.+v, 04.70.Dy, 98.80.Jk

I. INTRODUCTION

In a key article [1], Jacobson derived Einstein's equations (Ee) from arguments based only on thermodynamics at equilibrium of local space-time. To derive this result, he first generalized black hole (Bh) thermodynamics to space-time thermodynamics as seen by a local observer. Then, he took the basic Clausius relation, to obtain a local differential equation (from the previously defined map between local space-time and thermodynamic observables), that turns out to be precisely Ee. From this we can say that Ee are derived as an "equation of state" from local space-time thermodynamics. He noted that his finding strongly suggests that, in a fundamental context, Ee are to be viewed as an equation of state and, therefore, they should probably not be taken as a basis for quantizing gravity. This is consistent with the idea that gravity is an emergent phenomenon of a more fundamental framework, like string theory (e.g., [2]). If this were true, not only general relativity, but presumably all generalized gravity theories, should be seen under this same light.

Modified gravity models constitute a very important dynamical alternative to ΛCDM cosmology, in that they have the capability to describe the current accelerated expansion of our Universe (dark energy epoch), but also the initial de Sitter phase and inflation, and even the galaxy rotation curves corresponding to dark (and ordinary) matter [3]. We will here prove that Jacobson's derivations can be generalized to cover these more complicated theories of gravity that are extensively used nowadays. First, we review Jacobson's arguments to introduce the basic notions and then derive the desired generalization. In particular, we will completely close the program for the so-called $\mathbf{f}(R)$ gravities (see, e.g., [3], and references therein), where the Lagrangian only depends on the Ricci scalar and its covariant derivatives, leaving the problem open for more general cases. In [4], the field equations for $\mathbf{f}(R)$ of polynomial form were derived using nonequilibrium thermodynamics arguments (see also [5]). Here, we propose an alternative approach where local thermodynamic equilibrium is maintained, using the idea of "local-boost-invariance" introduced in [6].

II. JACOBSON'S CONSTRUCTION IN BRIEF

Any free-falling local observer p has some gauge freedom to describe his local coordinate system. The equivalence principle can be used to describe space-time in a vicinity of p as flat. Then, we choose the local spacelike area element perpendicular to the worldline of p to have zero expansion rate θ and shear σ at a given point on the history of p, that we call p_0. In this setting, the past horizon of p_0 is called the "local Rindler horizon" at p_0. Since, locally, we have Poincaré symmetry, there is an approximate Killing field K generating boost at p_0, vanishing at p_0, which we take as the future pointing to the inside past of p_0.

Having this basic setting, we are ready to give precise meaning to the local thermodynamic definitions. First, note that local Rindler horizons are null and act as causal barriers. Therefore, we can associate entropy S to it, measuring the "many degrees of freedom outside," what presumably results in entanglement entropy just at the horizon. With this understanding, entropy is proportional to the area elements of the horizon, where a fundamental length has to be provided to give a UV cutoff. Heat Q is energy flow of microscopic degrees of freedom across the causal barrier, and is felt, therefore, via gravitational energy, where its source is undetectable. Last, the local temperature T is defined as "Unruh temperature," as seen by a local accelerated observer hovering just inside the horizon. Energy flow has to be measured by this same observer, for consistency.

In more detail, different accelerated observers would measure different energy flows and temperature, both diverging at the horizon but with constant ratio, and this is just what will be used. We have also imposed $\theta = \sigma = 0$ at p_0, to give a sort of "local definition of equilibrium" since,

*elizalde@ieec.uab.es
elizalde@math.mit.edu
+psilva@ifae.es

in general, causal horizons change in time as they expand and twist. In this construction, locally and at the instant p_0, there is not such a deformation, and the space is "at equilibrium."

A. Accelerated observer and approximations

On the above point p_0 with its associated local Rindler horizon $\mathcal{H}$, take an accelerated observer hovering just inside the horizon χ. By the above construction, χ is an approximate local boost Killing field future directed to the past of p_0. Then, the variation of heat (caused by energy flow across the horizon), measured by χ, is $\delta Q = \int_{\mathcal{H}} T_{ab}\chi^b d\Sigma^a$, where the integration is over a pencil of generators of $\mathcal{H}$ at p_0. If K is a tangent vector to the generators of $\mathcal{H}$, with affine parameter λ such that $\lambda = 0$ at p_0, we have that $\chi^a = -k\lambda K^a + O(\lambda^2)$, where k is the acceleration of χ. Therefore, $d\Sigma^a = K^a d\lambda dA$, where dA is the cross section area element of $\mathcal{H}$. Thus, the final expression for the variation of heat, at leading order, is

$$\delta Q = -\int_{\mathcal{H}} k\lambda T_{ab}K^a K^b d\lambda dA. \tag{1}$$

Note that for χ the Unruh temperature T is set to be

$$T = \hbar k/2\pi. \tag{2}$$

On the other hand, Jacobson uses that the variation of the entropy is proportional to the variation of the horizon area $\mathcal{A}$, i.e., $\delta S = \eta\delta\mathcal{A}$, with η an unknown proportionality constant. Here, $\delta\mathcal{A}$ measures the change of the area as we approach the point p_0, and therefore is given by $\delta\mathcal{A} = \int_{\mathcal{H}} \theta(\lambda)d\lambda dA$. Next, we use Raychaudhuri's equation to integrate θ near p_0. In this coordinate system, at leading order in λ, we obtain $\theta = -\lambda R_{ab}K^a K^b + O(\lambda^2)$, and the relevant expression for the entropy variation to this order is

$$\delta S = -\eta\int_{\mathcal{H}} \lambda R_{ab}K^a K^b d\lambda dA. \tag{3}$$

B. Thermodynamic relations

To derive information of thermodynamic systems, like the equation of state, we need just the basic thermodynamic relation

$$\delta Q = T\delta S, \tag{4}$$

and the functional dependence of S with respect to the energy and size of the system. In our case we have Eqs. (1)–(3) at our disposal, to get the beautiful relation

$$T_{ab}K^a K^b = \frac{1}{2\pi}\eta R_{ab}K^a K^b. \tag{5}$$

Since K is an arbitrary null vector on $\mathcal{H}$, we can write the unprojected equation $T_{ab} = \frac{1}{2\pi}\eta R_{ab} + g_{ab}h$, with h an unknown function, arbitrary as of now. Using then that the left-hand side is divergence-free, plus the Bianchi identities for the Ricci tensor, we get the integrability conditions, $\frac{1}{4\pi}\eta\nabla_a R = -\nabla_a h$, and therefore the final form of the thermodynamic relation is

$$\left(\frac{2\pi}{\eta}\right)T_{ab} = \left(R_{ab} - \frac{R}{2}g_{ab}\right) + \Lambda g_{ab}, \tag{6}$$

where Λ is an integration constant.

To summarize, we have here obtained the Ee as an equation of state for a local free-falling observer. To deduce the above, we have used the following critical assumptions: (i) Measurements are done in a vicinity of a general point p_0. (ii) Our local coordinate system is at equilibrium, in the sense that θ, $\sigma = 0$ at p_0. (iii) The accelerated observer χ tends to K, a null vector generator of the causal horizon. (iv) We always restrict ourselves to the leading-order approximation in the affine parameter λ.

III. THE GENERAL CASE OF MODIFIED GRAVITY

We apply the above construction to more general theories of gravity. Following Iyer and Wald [6], we just assume that our Lagrangian is diffeomorphism invariant, in an n-dimensional oriented manifold $\mathcal{M}$, being the dynamical fields a Lorentz signature metric g_{ab} and other matter fields ψ. The most general Lagrangian is

$$\begin{aligned}\mathbf{L} &= \mathbf{L}(g_{ab}, R_{cdef}, \nabla_{a_1}R_{cdef}, \ldots, \nabla_{(a_1}\ldots\nabla_{a_n)}R_{cdef},\\ &\qquad \psi, \nabla_{a_1}\psi, \ldots, \nabla_{(a_1}\ldots\nabla_{a_n)}\psi).\end{aligned} \tag{7}$$

The corresponding field equations can be found by a variational procedure on (g_{ab}, ψ), so that we get

$$\delta\mathbf{L} = \epsilon(\mathbf{E}_g^{ab}\delta g_{ab} + \mathbf{E}_\psi\delta\psi) + d\mathbf{\Theta}, \tag{8}$$

where ϵ is the volume element and $\mathbf{\Theta}$ a $(n-1)$-form. Hence, the field equations of the theory are $\mathbf{E}_g^{ab} = 0$, $\mathbf{E}_\psi = 0$. In [6], it was found how to write them from a variation of the (g_{ab}, R_{cdef}), as if they were independent variables, so that we get, after the corresponding identifications,

$$\mathbf{E}_g^{ab} = A_g^{ab} + E_R^{pqra}R_{pqr}{}^b + 2\nabla_p\nabla_q E_R^{pabq}, \tag{9}$$

where (A_g^{ab}, E_R^{pabq}) are the variations of $\mathbf{L}$ with respect to (g_{ab}, R_{pabq}) in each case, taken as independent variables. In the above expressions, if the derivatives of R_{cdef} occur in the Lagrangian, one integrates by parts and then takes its variation, to obtain E_R^{pabq}.

This form of the field equations is useful due to its relation to Bh thermodynamics. Basically, it has been known for a while now [7], that in the case when we have a stationary Bh solution, the entropy S can be calculated as a Noether charge evaluated at the bifurcation $(n-2)$-surface of the event-horizon Σ. In these cases, the entropy is given by

$$S = -2\pi \int_\Sigma E_R^{abpq} \epsilon_{ab} \epsilon_{pq}, \tag{10}$$

where ϵ_{ab} is the binormal vector of Σ.

Less understood is the case of dynamical Bh entropy. There, the event-horizon is not bifurcated and the above formula does not hold. On the other hand, in [6] a prescription that passes the tests of consistency for the entropy is presented. The idea is to approximate the metric g, in a vicinity of a given point p of the event-horizon, by a boost-invariant metric g^{I_q} (q boost parameter, see below). This is done by altering the original Taylor expansion of the metric around p, so that the new metric is boost-invariant up to some order q, that defines the size of the vicinity where our approximation is valid. Then, for this boost-invariant metric, there is a Killing vector field that, on the horizon, is null, and vanishes at p. We thus have created an approximated bifurcation surface of order q and can use the same expression as before for the entropy, only that the integration is done on the boost-invariant variables:

$$S_{\rm dyn}(\Sigma_p) = -2\pi \int_{\Sigma_p} \hat{E}_R^{abpq} \epsilon_{ab} \epsilon_{pq}, \tag{11}$$

where $\hat{E}_R^{abpq} = E_R^{abpq}(g^{I_q})$ (see [6] for details).

Having understood these modifications for calculating the Bh entropy, we are almost ready to continue. Still, we need some information on the geometrical meaning of Eq. (10). In [8] it was noticed that, for a static Bh, entropy can always be reexpressed as the area of the bifurcation $(n-2)$-surface A divided by 4 in units of an effective Newton constant $G_{\rm eff}$, i.e.,

$$S = \frac{A}{4G_{\rm eff}} \quad \text{where} \quad \frac{1}{8\pi G_{\rm eff}} = E_R^{abpq} \epsilon_{ab} \epsilon_{pq}. \tag{12}$$

The above result has been checked for some string theory cases where it was found that $G_{\rm eff}$ is indeed constant on the bifurcation surface. This result has to be supplemented with the key observation that the above effective Newton constant plays also the role of an effective gravitational coupling for graviton exchange. In other words, the kinetic term of the n-dimensional graviton, obtained from the general Lagrangian $\mathbf{L}$, is precisely of the form $\frac{1}{4} E_R^{abpq}(\nabla_r h_{bq} \nabla^r h_{ap} + \ldots)$ and, hence, E_R^{abpq} can be thought of as the strength of the graviton interaction in all possible polarizations. In retrospective, $G_{\rm eff}$ corresponds to the strength of the gravitational interaction in the particular polarizations defined by the binormal of the bifurcation $(n-2)$-surface A.

A. Field equation as equation of state

Now that we have the basic inputs for the possible geometrical interpretation of the Bh entropy S for generalized theories of gravity, we are ready to consider the problem of defining there a local version of Bh thermodynamics, following the steps of Jacobson. Note that all definitions regarding local observers (accelerated or not), local Rindler horizons, and so on, are based on differential geometry and the equivalence principle. We expect all these definitions to hold in the general setting and, thus, we leave them unchanged.

What makes the difference, being one of the key points in this generalization, is the definition of the local entropy. After the discussion above, it seems natural to relate entropy to the area of the causal horizon, only that now we replace the proportionality constant with a field-dependent effective constant. In other words, we state that the local variation of the entropy is still proportional to the variation of the area of the causal horizon, but in units of this effective Newton constant. Therefore, we write now [9]

$$\delta S = \delta(\eta_e A), \tag{13}$$

where η_e is, in general, a function of the metric and its derivatives to a given order $l+2$, i.e.,

$$\eta_e = \eta_e(g_{ab}, R_{cdef}, \nabla^{(l)} R_{pqrs}). \tag{14}$$

Using the above *ansatz*, we are ready to proceed with our derivation. Since we just change the definition of entropy variation, due to the energy flow across the local Rindler horizon, we get the modified expression

$$\delta S = -\int_{\mathcal{H}} \lambda(\eta_e R_{ab} - \nabla_a \nabla_b \eta_e) K^a K^b d\lambda dA + O(\lambda^2). \tag{15}$$

It is important to notice that, in this expression, $(\eta_e, k^a \nabla_a \eta_e)$ is to be evaluated at its *leading* contribution in λ. We have used its boost-invariant part at first order in lambda to effectively incorporate the boost-invariant notion of [6] creating an "approximated bifurcation point at first order in λ" at p_0.

The other part of the derivation is unaffected and gives the same result of (1), namely,

$$\delta Q = -\int_{\mathcal{H}} k\lambda T_{ab} K^a K^b d\lambda dA + O(\lambda^2). \tag{16}$$

Therefore, the thermodynamic relation (4) implies

$$T_{ab} K^a K^b = \frac{1}{2\pi}(\eta_e R_{ab} - \nabla_{(a} \nabla_{b)} \eta_e) K^a K^b. \tag{17}$$

At this point we consider the general differential equation, removing the contraction with K, thus

$$T^{ab} = \frac{\eta_e}{2\pi} R^{ab} - \nabla^{(a} \nabla^{b)} \frac{\eta_e}{2\pi} + g^{ab} H, \tag{18}$$

where the new terms are added based on the fact that K is a tangent vector of the null geodetics at p_0, generating local boost. Hence, we have a local equation with two unknown functions of the metric and its derivatives (η_e, H).

To find the form of these three functions, we use the integrability condition

$$\nabla_a\left(\frac{\eta_e}{2\pi}R^{ab} - \nabla^{(a}\nabla^{b)}\frac{\eta_e}{2\pi} + g^{ab}H\right) = 0, \qquad (19)$$

obtained from the observation that the right-hand side should be divergence-free. After some algebra,

$$0 = \frac{\eta_e}{4\pi}\nabla^b R + (\nabla_a\nabla^b\nabla^a - \nabla^b\nabla^2)\frac{\eta_e}{2\pi} + \nabla^b H - \frac{1}{2}(\nabla^2\nabla^b + \nabla_a\nabla^b\nabla^a)\frac{\eta_e}{2\pi}. \qquad (20)$$

At this point, with no lose of generality, we can set

$$H = h + \nabla^2\frac{\eta_e}{2\pi}, \qquad (21)$$

finally obtaining the expression

$$0 = \frac{\eta_e}{4\pi}\nabla^b R + \nabla^b h. \qquad (22)$$

B. The specific case of $f(R)$ gravities

Equation (19) can in principle be solved in many different ways. Here we will consider the simplest possibility, that eventually leads to the so-called $\mathbf{f}(R)$ gravities [3] (note the $\mathbf{f}$), a special—but phenomenologically very important—family of the general theories of modified gravity where only the Ricci scalar is involved, as well as their covariant derivatives to any order, i.e. $\mathbf{L} = \mathbf{f}(R, \nabla^n R)$.

The solution of Eq. (19) we are considering is the one where the first two terms cancel each other. The last can be easily integrated assuming η_e is a function of R only, thus $h = -\frac{1}{2}\mathbf{f}(R)$, with $\frac{\eta_e}{2\pi} = \frac{\partial \mathbf{f}}{\partial R}$. Therefore, the final form is

$$E^{ab} = \frac{\partial \mathbf{f}}{\partial R}R^{ab} - \nabla^{(a}\nabla^{b)}\frac{\partial \mathbf{f}}{\partial R} + g^{ab}\left(-\frac{1}{2}\mathbf{f} + \nabla^2\frac{\partial \mathbf{f}}{\partial R}\right), \qquad (23)$$

$\mathbf{f}$ a function of R and its covariant derivatives.

Equation (23) is in fact the correct field theory equation for $\mathbf{f}(R)$ gravities, provided we identify the function $\mathbf{f}(R)$ as the Lagrangian of the theory. Also, in this case the effective Newton constant of (12) is related to η_e, as is expected from the relation

$$\frac{1}{8\pi G_{\rm eff}} = E_R^{pqrs}\epsilon_{pq}\epsilon_{rs} = \frac{\partial \mathbf{f}}{\partial R}(g^{pr}g^{qs} - g^{qr}g^{ps})\epsilon_{pq}\epsilon_{rs} = \frac{\partial \mathbf{f}}{\partial R} = \frac{\eta_e}{2\pi}. \qquad (24)$$

Note that, for these theories, the different polarizations of the gravitons only enter in the definition of the effective Newton constant through the metric itself. This is an important simplification that, in turn, permits us to find the solution of the integrability condition (19). To summarize, we have succeeded in our thermodynamic derivation of $\mathbf{f}(R)$ gravities where, remarkably, exactly as in the case of Einstein gravity, *the local field equations can be thought of as an equation of state of equilibrium thermodynamics*.

It will be very interesting to see if this derivation can be extended to the more complicated cases, stemming from string theory, where the full Riemann tensor is involved in the Lagrangian. This seems to imply a sort of tetrad decomposition of the effective Newton constant such that one recovers, at the end, only the polarization normal to the causal barrier of the local Rindler horizon. Work along this line is in progress. As a last comment, in our derivation we have used the first law, but no information is given about the second law. In fact it is not known if the second law is present in generalized gravities (see [10]).

ACKNOWLEDGMENTS

We thank S. Odintsov and E. Verlinde for useful discussions. Work was partially funded by Ministerio de Educación y Ciencia, Spain, Project Nos. CICYT-FEDER-FPA2005-02211 and FIS2006-02842, CSIC under the I3P program, and AGAUR, Contract No. 2005SGR-00790.

[1] T. Jacobson, Phys. Rev. Lett. **75**, 1260 (1995).
[2] P. J. Silva, J. High Energy Phys. 11 (2005) 012.
[3] G. Cognola, E. Elizalde, S. Nojiri, S. D. Odintsov, L. Sebastiani, and S. Zerbini, Phys. Rev. D **77**, 046009 (2008); S. Capozziello, V. F. Cardone, and A. Troisi, Phys. Rev. D **71**, 043503 (2005); E. Elizalde, S. Nojiri, and S. D. Odintsov, Phys. Rev. D **70**, 043539 (2004); S. Capozziello, S. Carloni, and A. Troisi, Recent Res. Dev. Astron. Astrophys. **1**, 625 (2003).
[4] C. Eling, R. Guedens, and T. Jacobson, Phys. Rev. Lett. **96**, 121301 (2006).
[5] A. Paranjape, S. Sarkar, and T. Padmanabhan, Phys. Rev. D **74**, 104015 (2006).
[6] V. Iyer and R. M. Wald, Phys. Rev. D **50**, 846 (1994); H. Buchdahl, Acta Math. **85**, 63 (1964).
[7] R. M. Wald, Phys. Rev. D **48**, R3427 (1993).
[8] R. Brustein, D. Gorbonos, and M. Hadad, arXiv: 0712.3206.
[9] We thank T. Jacobson for pointing out that in a first version of this work, δS assumed η_e constant along K before using the boost-invariant truncation.
[10] T. Jacobson, G. Kang, and R. C. Myers, Phys. Rev. D **52**, 3518 (1995).