The laws of black hole mechanics and thermodynamics

Marek Liška

Institute of Theoretical Physics, Faculty of Mathematics and Physics, Charles University,
V Holešovičkách 2, 180 00 Prague 8, Czech Republic and
School of Theoretical Physics, Dublin Institute for Advanced Studies,
10 Burlington Road, Dublin 4, Ireland.[*]

INTRODUCTION

This is a short study text accompanying my lectures given as part of the subject Black hole thermodynamics. If you find anything in these notes unclear, chances are it is an error (factual or pedagogical) on my side. In that case please do not hesitate to contact me. The same applies if you want to hear more about any topic discussed (or even just hinted at) in this text.

Let me also include a brief note in regards to the exams. While we discuss both the covariant phase space formalism and the Euclidean grand-canonical ensemble construction in some detail because of their usefulness, we appreciate that the related mathematics takes some time to get used to. We will certainly not require that you reproduce the calculations presented in this text during the exam. Instead, we will ask general question aimed at understanding the logic behind the calculations, e.g. "What steps do you need to take to obtain a covariant prescription for the symplectic form for the given Lagrangian?"

I. COVARIANT PHASE SPACE FORMALISM

Reputedly, the entire field of black hole thermodynamics began with the question "What happens when you pour a cup of tea into a black hole?". In 1972, John Wheeler asked this his student Jacob Bekenstein, wondering about the entropy of the hot tea that apparently just disappears. This of course violates the second law of thermodynamics, thus posing a very serious problem for the plausibility of black holes as physical objects. Of course, there

[*] liska.mk@seznam.cz

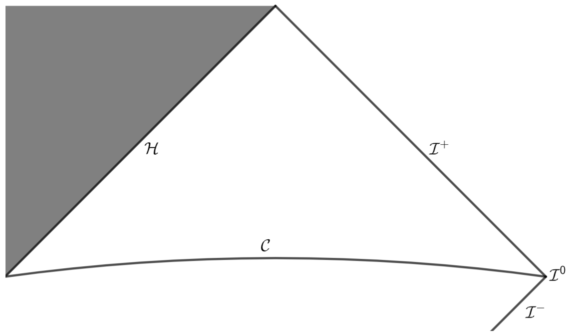

FIG. 1. A Penrose diagram of the stationary black hole spacetime we analyse. $\mathcal{H}$ denotes the horizon, $\mathcal{I}^+$ and $\mathcal{I}^-$ are the future and past null infinity, and $\mathcal{I}^0$ the spatial infinity. We draw the Cauchy surface $\mathcal{C}$ for the exterior region as the oblique line extended from $\mathcal{H}$ to $\mathcal{I}^+$. The grey region represents the interior of the black hole, whose structure we do not specify.

is a way out and Bekenstein succeeded in finding it. In the following, we will take a rather long (but hopefully rewarding) road to this answer, first treating the pouring of tea inside a black hole as a perturbation problem within the context of classical general relativity.

Our starting point is a stationary, asymptotically flat spacetime with a single black hole, whose Penrose diagram is shown in figure 1. As a consequence of stationarity, the spacetime possesses a Killing vector that is timelike outside the black hole horizon $\mathcal{H}$ and spacelike inside it. Hence, $\mathcal{H}$ is a Killing horizon. We are interested in the exterior region of the spacetime accessible to external observers, whose inner boundary is $\mathcal{H}$ and outer boundary the future null infinity $\mathcal{I}^+$. We do not care about the precise nature of the black hole's interior, e.g. whether there is a collapsing star inside and if there exists an inner horizon. We simply consider this part of the spacetime inaccessible. The exterior region can be fully described by giving initial data on a spacelike surface $\mathcal{C}$ which intersects the horizon and the spatial infinity $\mathcal{I}^0$ and then evolving them according to Einstein equations[1]. In other words, $\mathcal{C}$ is a Cauchy surface for the exterior region.

Now we pour the tea into the black hole (or perform any other small perturbation). We assume that the black hole settles down after a while, and again becomes stationary, albeit with slightly different properties (e.g., mass, angular momentum, etc.) than before. We are interested in mathematically describing this change of properties. While we might choose a

[1] This is no longer true when one introduces the Hawking radiation, since the exterior region is then also affected by its flux across $\mathcal{H}$. However, we remain fully in the realm of classical physics and treat $\mathcal{H}$ as intraversable from inside out.

specific black hole metric (e.g. Kerr-Newman) and perform its small perturbation, we can treat the problem more generally. For any stationary system in theoretical mechanics, we could find an expression for its Hamiltonian. Then, we recall that Hamiltonian of a stationary system is conserved. Hence, its perturbation vanishes, $\delta H = 0$, and this constrains the behaviour of the perturbations of quantities on which the Hamiltonian depends, giving us the information we wanted. This is the usual way to obtain the first law of thermodynamics for a mechanical system. It turns out that the same works for black holes. Specifically, for a stationary black hole, we can obtain a covariant expression for the perturbation of the Hamiltonian defined on the spacelike Cauchy surface. We will see that equating this expression to zero directly yields the first law of black hole mechanics. Moreover, the procedure easily generalises beyond general relativity to a wide class of alternative theories of gravity, as well as to other spacetimes with different global symmetries, making it a very useful and flexible computational tool.

A. Motivation from theoretical mechanics

Before going to black holes, let us practice on something simpler. We will just try to get the Hamiltonian for one nonrelativistic point particle of mass m in external potential $V(x)$. Our starting point is the action

$$S_{(1)} = \int_{t_1}^{t_2} L_{(1)} \mathrm{d}t = \int_{t_1}^{t_2} \left[\frac{1}{2} m \dot{x}_i \dot{x}^i - V(x)\right] \mathrm{d}t, \tag{1}$$

where $L_{(1)}$ is the Lagrangian, overdot denotes a time derivative, and i goes from 1 to D, being the dimension of space. To get Hamiltonian, we can just define the canonical momentum as $p_i = \partial L_{(1)}/\partial \dot{x}^i = m\dot{x}^i$ and use the Legendre transform, $H_{(1)} = p_i \dot{x}^i - L_{(1)} = p_i p^i/(2m) + V(x)$. We could even do that in general relativity (I refer you to chapter 21 of the eternal classic [1], where the ADM formalism is explained), but there it relies on fixing the direction of time first. We would rather have something covariant. A way to get it is by obtaining the symplectic structure of the theory in question. For a free particle, we can start by performing a small variation of S_{free}

$$\delta S_{(1)} = \int_{t_1}^{t_2} \left[m\dot{x}_i \frac{\mathrm{d}}{\mathrm{d}t}\left(\delta x^i\right) - V_{,i}\delta x^i\right] \mathrm{d}t = -\int_{t_1}^{t_2} \left(m\ddot{x}_i - V_{,i}\right)\delta x^i \mathrm{d}t + \int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}t}\left(m\dot{x}_i \delta x^i\right) \mathrm{d}t. \tag{2}$$

The first term is proportional to the equations of motion of the particle $m\ddot{x}_i = V_{,i}$. The second one is clearly a boundary term and one usually discards it. However, let us take a look at the expression inside the derivative, $\theta_{(1)}\left[\delta\right] = m\dot{x}_i\delta x^i = p_i\delta x^i$, where we identified the canonical momentum. For reasons that will become apparent soon, we call $\theta_{(1)}\left[\delta\right]$ the symplectic potential.

Now suppose that we consider two independent variations δ_1 and δ_2 and compute $\delta_1\theta_{(1)}\left[\delta_2\right]$ We easily obtain

$$\delta_1\theta_{(1)}\left[\delta_2\right] = \delta_1 p_i\delta_2 x^i + p_i\delta_1\delta_2 x^i. \tag{3}$$

A difference between $\delta_1\theta_{(1)}\left[\delta_2\right]$ and the expression with the order of variations flipped, $\delta_2\theta_{(1)}\left[\delta_1\right]$, reads

$$\delta_1\theta_{(1)}\left[\delta_2\right] - \delta_1\theta_{(1)}\left[\delta_2\right] = \delta_1 p_i\delta_2 x^i + p_i\delta_1\delta_2 x^i - \delta_2 p_i\delta_1 x^i - p_i\delta_2\delta_1 x^i = \delta_1 p_i\delta_2 x^i - \delta_2 p_i\delta_1 x^i, \tag{4}$$

where we used that the independent variations commute, i.e., $p_i\delta_1\delta_2 x^i = p_i\delta_2\delta_1 x^i$. Let us introduce a joint notation for coordinates and momenta, z^i so that $z^1 - z^D$ correspond to $x^1 - x^D$ and $z^{D+1} - z^{2D}$ to $p^1 - p^D$. Then, we can rewrite expression (4)

$$\delta_1\theta_{(1)}\left[\delta_2\right] - \delta_2\theta_{(1)}\left[\delta_1\right] = \Omega_{(1),ij}\delta_1 z^i\delta_2 z^j, \tag{5}$$

where $\Omega_{(1),ij}$ are components of a matrix with the following structure (for simplicity, we show it for $D = 3$)

$$\Omega_{(1)} = \begin{pmatrix} 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{6}$$

This matrix is clearly antisymmetric and nondegenerate (determinant is $(-1)^D$, i.e., nonzero). If we see $\Omega_{(1),ij}$ as a differential 2-form $\boldsymbol{\Omega}_{(1)} = \Omega_{(1),ij}\mathrm{d}z^i \wedge \mathrm{d}z^j$, it is clearly closed, $\mathrm{d}\boldsymbol{\Omega}_{(1)} = 0$. Nondegeneracy and closedness are the defining properties of a symplectic form. Hence, $\boldsymbol{\Omega}_{(1)}$ is a symplectic form on the phase space of our theory. Defining similarly a symplectic potential 1-form $\boldsymbol{\theta}_{(1)} = p_i\mathrm{d}x^i$, we have $\mathrm{d}\boldsymbol{\theta}_{(1)} = \boldsymbol{\Omega}_{(1)}$.

It is easy to check that we can write the Hamilton equations of motion in terms of the symplectic structure[2], i.e.,

$$\frac{\partial H_{(1)}}{\partial z^i} = \Omega_{(1),ij}\dot{z}^j. \tag{7}$$

[2] For more details on the symplectic formalism in theoretical mechanics, see [2]

For a small variation of the Hamiltonian, we can then write (assuming the equations of motion are satisfied)

$$\delta H_{(1)} = \frac{\partial H_{(1)}}{\partial z^i} \delta z^j = \Omega_{(1),ij} \dot{z}^i \delta z^j. \tag{8}$$

In this way, we can compute a small variation of the Hamiltonian (and, in typical situations, guess the full structure of the Hamiltonian) once we are given the symplectic structure.

B. Covariant phase space formalism for general relativity

We have seen how to obtain a symplectic structure from a one particle Lagrangian. We now simply try to do the same thing for general relativity and assume it will work (you can actually show it will, but it is painful [3]). We take the Lagrangian density for the vacuum general relativity, without a cosmological constant[3],

$$L = \frac{1}{16\pi} R \sqrt{-g}, \tag{9}$$

and vary it with respect to $g^{\mu\nu}$. We get

$$\delta L = \frac{1}{16\pi} \sqrt{-g} \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right) \delta g^{\mu\nu} + \frac{1}{16\pi} \sqrt{-g} g^{\mu\nu} \delta R_{\mu\nu}. \tag{10}$$

The first term is of course proportional to the vacuum Einstein equations and vanishes if they are satisfied. The second term can be written as a total derivative

$$\frac{1}{16\pi} \sqrt{-g} g^{\mu\nu} \delta R_{\mu\nu} = \nabla_\mu \left[\frac{1}{16\pi} \left(g^{\mu\lambda} g_{\nu\rho} - \delta^\mu_\nu \delta^\lambda_\rho \right) \nabla_\lambda \delta g^{\nu\rho} \right]. \tag{11}$$

According to the analogy with theoretical mechanics, we identify this as the divergence of the symplectic potential corresponding to variation δ

$$\theta^\mu \left[\delta \right] = \frac{\sqrt{-g}}{16\pi} \left(g^{\mu\lambda} g_{\nu\rho} - \delta^\mu_\nu \delta^\lambda_\rho \right) \nabla_\lambda \delta g^{\nu\rho}. \tag{12}$$

In other words, we can write the variation of the Lagrangian (9) as

$$\delta L = \frac{1}{16\pi} \sqrt{-g} \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right) \delta g^{\mu\nu} + \nabla_\mu \theta^\mu \left[\delta \right]. \tag{13}$$

[3] Rather than working with tensor densities, the entire formalism is more often presented in terms of differential forms. In that case, we simply start with the Lagrangian 4-form, $\boldsymbol{L} = R\boldsymbol{\varepsilon} / (16\pi)$, with $\boldsymbol{\varepsilon}$ being the spacetime volume 4-form (i.e., the Levi-Civita tensor), and otherwise proceed in the same way as in these notes. Here, I do not follow this road as I find the differential form language a little less familiar than tensor densities.

Following the analogy, we then obtain the symplectic current corresponding to two independent variations δ_1, δ_2 as[4]

$$\Omega^\mu\left[\delta_1,\delta_2\right]=\delta_1\theta^\mu\left[\delta_2\right]-\delta_2\theta^\mu\left[\delta_1\right]. \tag{14}$$

There is an extra step compared to the case of one particle Lagrangian, in which the previous equation already yields the symplectic form. However, in any field theory we have to deal with the fact that our starting point is a Lagrangian density rather than a Lagrangian. Then, to get the symplectic form, we must carry out one additional integration over a suitably defined spacelike surface. In particular, we need this surface to be Cauchy. For our stationary black hole, we have the surface $\mathcal{C}$ in figure 1 (thanks to stationarity, you can show that the result does not depend on the particular choice of the Cauchy surface). So, integral of the symplectic current $\Omega^\mu\left[\delta_1,\delta_2\right]$ over $\mathcal{C}$ gives us the symplectic form

$$\Omega\left[\delta_1,\delta_2\right]=\int_{\mathcal{C}}\Omega^\mu\left[\delta_1,\delta_2\right]\mathrm{d}\mathcal{C}_\mu. \tag{15}$$

Since the symplectic form yields the perturbation of the Hamiltonian by virtue of the Hamilton equation of motion, it might seem that we are done. However, in the cases (like ours) when the spacetime possesses symmetries we can do even more. The Noether theorem teaches us that to symmetries correspond conserved currents and charges. As we will see, we can express the perturbation of the Hamiltonian (and the Hamiltonian itself) entirely in terms of such charges.

To see this, we start by looking at the local symmetries of general relativity. It is well known that it is invariant under arbitrary diffeomorphisms. Since we work with small perturbations, we are interested in infinitesimal diffeomorphisms, which can always be written as being generated by some vector field ξ^μ. The transformation of the metric is then given by a Lie derivative along this vector field, i.e.,

$$\delta_\xi g_{\mu\nu}=\pounds_\xi g_{\mu\nu}=\nabla_\mu\xi_\nu+\nabla_\nu\xi_\mu. \tag{16}$$

The Noether current corresponding to an infinitesimal diffeomorphism is in general defined as

$$j^\mu_\xi=\theta^\mu\left[\pounds_\xi\right]-L\xi^\mu. \tag{17}$$

[4] The explicit expression is somewhat lengthy and can be found, e.g. in [4].

To be a proper Noether current, the divergence of j^μ_ξ must vanish. We have, using equation (13) to express $\nabla_\mu \theta^\mu\left[\pounds_\xi\right]$,

$$\nabla_\mu j^\mu_\xi = \nabla_\mu \theta^\mu\left[\pounds_\xi\right] - \nabla_\mu\left(L\xi^\mu\right) = \pounds_\xi L - \frac{1}{16\pi}\sqrt{-g}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right)\pounds_\xi g^{\mu\nu} - \xi^\mu\nabla_\mu L - L\nabla_\mu\xi^\mu. \tag{18}$$

Now, we need to evaluate Lie derivative of the Lagrangian

$$\pounds_\xi L = \frac{1}{16\pi}\sqrt{-g}\pounds_\xi R + \frac{1}{16\pi}R\pounds_\xi\sqrt{-g}, \tag{19}$$

where we used the Leibniz rule. Lie derivative of the scalar curvature yields simply $\pounds_\xi R = \xi^\mu\nabla_\mu R$. However, $\sqrt{-g}$ is not a tensor but rather a tensor density of weight $w = 1$. For its Lie derivative, we find

$$\pounds_\xi\sqrt{-g} = -\frac{1}{2\sqrt{-g}}\pounds_\xi g = \frac{1}{2\sqrt{-g}}\left(-g\right)g^{\mu\nu}\pounds_\xi g_{\mu\nu} = \frac{\sqrt{-g}}{2}g^{\mu\nu}\left(\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu\right) = \sqrt{-g}\nabla_\mu\xi^\mu. \tag{20}$$

where we used a well-known relation for a small variation of the metric determinant $\delta g = gg^{\mu\nu}\delta g_{\mu\nu}$ for $\delta = \pounds_\xi$. Plugging the expression for $\pounds_\xi L$ to equation (18) and using that $\nabla_\mu g = 0$[5] yields

$$\begin{aligned}\nabla_\mu j^\mu_\xi =& L\nabla_\mu\xi^\mu + \xi^\mu\nabla_\mu L - \frac{1}{16\pi}\sqrt{-g}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right)\pounds_\xi g^{\mu\nu} - \xi^\mu\nabla_\mu L - L\nabla_\mu\xi^\mu\\ =& \frac{1}{8\pi}\sqrt{-g}\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}\right)\nabla^\mu\xi^\nu.\end{aligned} \tag{21}$$

If the vacuum equations of motion are satisfied, we indeed have $\nabla_\mu j^\mu_\xi = 0$ as required of a Noether current[6]. It follows that the Noether current can be written as a term proportional to the equations of motion plus some term whose divergence vanishes identically. The most general such term is a divergence of some rank two antisymmetric tensor density $Q^{\nu\mu}_\xi = -Q^{\mu\nu}_\xi$, i.e.,

$$j^\mu_\xi = \frac{1}{8\pi}\sqrt{-g}\left(R_\nu{}^\mu - \frac{1}{2}R\delta^\mu_\nu\right)\xi^\nu + \nabla_\nu Q^{\nu\mu}_\xi. \tag{22}$$

That j^μ_ξ must take this form is hopefully intuitively clear, but if one desires, it can be derived fully rigorously [5]. Notice that, when the equations of motion are satisfied, an integral of

[5] To see this, we can express the determinant as $g = \left[\alpha\beta\gamma\delta\right]g_{\alpha\gamma}g_{\beta\delta}$, where $\left[\alpha\beta\gamma\delta\right]$ is the antisymmetrisation symbol, $[0123] = 1$ and fully antisymmetric. Since ∇_μ is a Levi-Civita covariant derivative, i.e., torsion-free and metric compatible, we have $\nabla_\mu g_{\alpha\gamma} = 0$ and $\nabla_\mu\left[\alpha\beta\gamma\delta\right] = 0$. Therefore, $\nabla_\mu g = 0$.

[6] We can even define Noether current that is divergence-free identically, even when the equations of motion are not satisfied. Using the contracted Bianchi identities, we can easily prove that the Einstein tensor is divergence-free, i.e., $\nabla^\mu\left(R_{\mu\nu} - Rg_{\mu\nu}/2\right) = 0$. Then, we have $\nabla_\mu j^\mu_\xi = \nabla_\mu\left[\sqrt{-g}\left(R_\nu{}^\mu - R\delta^\mu_\nu/2\right)\xi^\nu/\left(8\pi\right)\right]$ and a new Noether current $J^\mu_\xi = j^\mu_\xi - \sqrt{-g}\left(R_\nu{}^\mu - R\delta^\mu_\nu/2\right)\xi^\nu/\left(8\pi\right)$ satisfies $\nabla_\mu J^\mu_\xi = 0$ identically.

j^μ_ξ over a Cauchy surface $\mathcal{C}$ may be rewritten using the Stokes' theorem

$$\int_{\mathcal{C}} j^\mu_\xi \mathrm{d}\mathcal{C}_\mu = \int_{\mathcal{C}} \nabla_\nu Q^{\nu\mu}_\xi \mathrm{d}\mathcal{C}_\mu = \int_{\partial\mathcal{C}} Q^{\nu\mu}_\xi \mathrm{d}\mathcal{C}_{\mu\nu}, \tag{23}$$

to an integral of the antisymmetric tensor $Q^{\nu\mu}_\xi$ over its boundary $\partial\mathcal{C}$. In other words, $Q^{\nu\mu}_\xi$ plays the role of the Noether charge corresponding to the Noether current j^μ_ξ.

We can find the explicit form of the Noether charge $Q^{\nu\mu}_\xi$ by splitting the Noether current defined by equation (17) into a part proportional to the vacuum equations of motion and some remainder, which must have the form $\nabla_\nu Q^{\nu\mu}_\xi$. By plugging in the expressions for the symplectic potential $\theta^\mu \left[\pounds_\xi\right]$ (12) (using expression (16) for the Lie derivative of the metric) and for the Lagrangian L (9), we obtain

$$\begin{aligned} j^\mu_\xi =& \frac{\sqrt{-g}}{16\pi} \left(g^{\mu\lambda} g_{\nu\rho} - \delta^\mu_\nu \delta^\lambda_\rho\right) \nabla_\lambda \left(\nabla_\nu \xi_\rho + \nabla_\rho \xi_\nu\right) - \frac{\sqrt{-g}}{16\pi} R \xi^\mu \\ =& \frac{\sqrt{-g}}{16\pi} \left(\nabla_\nu \nabla^\nu \xi^\mu + \nabla_\nu \nabla^\mu \xi^\nu - 2\nabla^\mu \nabla_\nu \xi^\nu - R\xi^\mu\right). \end{aligned} \tag{24}$$

To get the vacuum Einstein equations, we need some Ricci tensors. For this, we use the definition of the Riemann tensor

$$R_{\sigma\mu\nu\rho} \xi^\sigma = \nabla_\rho \nabla_\nu \xi_\mu - \nabla_\nu \nabla_\rho \xi_\mu, \tag{25}$$

which in particular implies[7]

$$\nabla_\nu \nabla^\mu \xi^\nu - \nabla^\mu \nabla_\nu \xi^\nu = R_\nu{}^\mu \xi^\nu, \tag{26}$$

$$-\nabla^\mu \nabla_\nu \xi^\nu = -\nabla_\nu \nabla^\mu \xi^\nu + R_\nu{}^\mu \xi^\nu. \tag{27}$$

Using these identities, we obtain for the Noether current

$$j^\mu_\xi = \frac{\sqrt{-g}}{16\pi} \left(2R_\nu{}^\mu \xi^\nu - R\xi^\mu + \nabla_\nu \nabla^\nu \xi^\mu - \nabla_\nu \nabla^\mu \xi^\nu\right). \tag{28}$$

As it must be, the first two terms inside the bracket correspond to twice the vacuum Einstein equations contracted with ξ^μ, while the other two are a divergence of an antisymmetric tensor, $\nabla^\nu \xi^\mu - \nabla^\mu \xi^\nu$. Hence, we can identify the second part with the Noether charge

$$Q^{\nu\mu}_\xi = \frac{\sqrt{-g}}{16\pi} \left(\nabla^\nu \xi^\mu - \nabla^\mu \xi^\nu\right). \tag{29}$$

[7] A cautionary remark: there exist stronger versions of these statements for Killing vectors. However, we cannot use them, since ξ^μ is, at this stage, an arbitrary vector field.

We now wish to relate the symplectic form with the Noether current. The road to this result is rather indirect but nevertheless interesting. Consider any metric that solves the vacuum Einstein equation and introduce its small perturbation. We want the perturbed spacetime to again be a solution of the Einstein equations (in other words, the perturbation must solve the linearised Einstein equations). This perturbation leads to a change in the Noether current corresponding to some vector field ξ^{μ} (this vector field is by definition unaffected by the perturbation). On the one side, starting from definition (17), we find

$$\delta j^{\mu}_{\xi} = \delta\theta^{\mu}\left[\pounds_{\xi}\right] - \xi^{\mu}\delta L = \delta\theta^{\mu}\left[\pounds_{\xi}\right] - \xi^{\mu}\nabla_{\nu}\theta^{\nu}\left[\delta\right]. \tag{30}$$

We used that, since the equations of motion are satisfied, the perturbation of the Lagrangian simplifies to $\delta L = \nabla_{\nu}\theta^{\nu}\left[\delta\right]$.

On the other side, we know that the Noether current can be written as a term proportional to equations of motion (which we can disregard, since they are satisfied by both the original and the perturbed spacetime) and divergence of the Noether charge. Therefore, it holds $\delta j^{\mu}_{\xi} = \nabla_{\nu}\delta Q^{\nu\mu}_{\xi}$[8], and equation (30) becomes

$$\nabla_{\nu}\delta Q^{\nu\mu}_{\xi} = \delta\theta^{\mu}\left[\pounds_{\xi}\right] - \xi^{\mu}\nabla_{\nu}\theta^{\nu}\left[\delta\right]. \tag{31}$$

The first term on the right hand side is already one half of the symplectic current $\Omega^{\mu}\left[\delta, \pounds_{\xi}\right]$ given by equation (14). Let us look at the second half

$$-\pounds_{\xi}\theta^{\mu}\left[\delta\right] = -\xi^{\nu}\nabla_{\nu}\theta^{\mu}\left[\delta\right] + \theta^{\nu}\left[\delta\right]\nabla_{\nu}\xi^{\mu} + \theta^{\mu}\left[\delta\right]\nabla_{\nu}\xi^{\nu}, \tag{32}$$

where we used the standard expression for the Lie derivative of a vector field and the previously computed Lie derivative of the metric determinant (20). Then, adding and subtracting from equation (31) expression $-\pounds_{\xi}\theta^{\mu}\left[\delta\right]$, once expanded as in equation (32), yields

$$\nabla_{\nu}\delta Q^{\nu\mu}_{\xi} = \delta\theta^{\mu}\left[\pounds_{\xi}\right] - \pounds_{\xi}\theta^{\mu}\left[\delta\right] + 2\nabla_{\nu}\left(\xi^{[\nu}\theta^{\mu]}\right) = \Omega^{\mu}\left[\delta, \pounds_{\xi}\right] + 2\nabla_{\nu}\left(\xi^{[\nu}\theta^{\mu]}\left[\delta\right]\right). \tag{33}$$

At this point, we are done. Integrating the previous equation with respect to some Cauchy surface $\mathcal{C}$ gives us the symplectic form

$$\Omega\left[\delta, \pounds_{\xi}\right] = \int_{\mathcal{C}}\Omega^{\mu}\left[\delta, \pounds_{\xi}\right]\mathrm{d}\mathcal{C}_{\mu} = \int_{\mathcal{C}}\nabla_{\nu}\left(\delta Q^{\nu\mu}_{\xi} - 2\xi^{[\nu}\theta^{\mu]}\left[\delta\right]\right)\mathrm{d}\mathcal{C}_{\mu} = \int_{\partial\mathcal{C}}\left(\delta Q^{\nu\mu}_{\xi} - 2\xi^{[\nu}\theta^{\mu]}\left[\delta\right]\right)\mathrm{d}\mathcal{C}_{\mu\nu}, \tag{34}$$

[8] A small trick occurs here: you can check that $\nabla_{\nu}Q^{\nu\mu}_{\xi} = \partial_{\nu}Q^{\nu\mu}_{\xi}$ (This follows from $Q^{\nu\mu}_{\xi}$ being an antisymmetric tensor density). Then, we have $\delta\nabla_{\nu}Q^{\nu\mu}_{\xi} = \delta\partial_{\nu}Q^{\nu\mu}_{\xi} = \partial_{\nu}\delta Q^{\nu\mu}_{\xi} = \nabla_{\nu}\delta Q^{\nu\mu}_{\xi}$, since the perturbation commutes with a partial derivative.

where we used the Stokes theorem to change integral of a divergence into a boundary integral. Now, if the Hamiltonian corresponding to evolution along the vector field ξ^μ exists, its perturbation is given by the Hamilton equations of motion as being equal to the symplectic form, i.e.,

$$\delta H_\xi = \Omega\left[\delta, \pounds_\xi\right] = \int_{\partial\mathcal{C}} \left(\delta Q_\xi^{\nu\mu} - 2\xi^{[\nu}\theta^{\mu]}\left[\delta\right]\right) \mathrm{d}\mathcal{C}_{\mu\nu}. \tag{35}$$

For the Hamiltonian itself, we then have

$$H_\xi = \int_{\partial\mathcal{C}} \left(Q_\xi^{\nu\mu} - 2\xi^{[\nu}B^{\mu]}\right) \mathrm{d}\mathcal{C}_{\mu\nu}, \tag{36}$$

where B^μ is such that $\theta^\mu\left[\delta\right] = \delta B^\mu$ (B^μ need not be covariant). This establishes an important result: if the Hamiltonian exists, it can always be given as a boundary integral. One can even specify the necessary and sufficient condition under which the Hamiltonian does exist, which we state for an interested reader without a proof (it can be found in [6])

$$\int_{\partial\mathcal{C}} \Omega^\mu\left[\delta, \pounds_\xi\right]\xi^\nu \mathrm{d}\mathcal{C}_{\mu\nu} = 0, \tag{37}$$

for an arbitrary (equations of motion satisfying) perturbation δ.

Notice that nothing in our construction of the Hamiltonian really relied on the fact that we work in vacuum general relativity. In fact, equation (35) holds in any diffeomorphism invariant theory, although things get a little complicated when one includes derivatives of the Riemann tensor in the Lagrangian [7–9]. This is the main power of the formalism we have developed. It allows to identify conserved quantities such as mass, angular momentum or entropy in any diffeomorphism invariant theory of gravity (and even in some more general cases). Given the rate at which new proposals for modified theories of gravity appear, a systematic way to check the conserved quantities for them proves very useful. We will now demonstrate this approach on the example of black holes in general relativity.

C. First law of black hole mechanics

All we need to do is to apply equation (35) to our example of vacuum, stationary axisymmetric black hole spacetime in general relativity. We choose the vector field ξ^μ as the Killing vector field timelike everywhere in the exterior region (see figure 1), i.e.,

$$\xi^\mu = t^\mu + \Omega_{\mathcal{H}}\varphi^\mu. \tag{38}$$

Here, t^μ and φ^μ are the time translational and rotational Killing vector fields, respectively, and $\Omega_{\mathcal{H}}$ denotes the constant angular velocity of the horizon (the constancy of $\Omega_{\mathcal{H}}$ guarantee the rigidity theorems). Since ξ^μ is a Killing vector, we have $\pounds_\xi g_{\mu\nu} = 0$ (by definition). We choose the Cauchy surface $\mathcal{C}$ orthogonal to ξ^μ. Then, it is easy to prove that $\Omega^\mu\left[\delta, \pounds_\xi\right] = 0$ and, consequently, $\delta H_\xi = 0$. From equation (35) we have

$$\int_{\partial\mathcal{C}} \left(\delta Q_\xi^{\nu\mu} - 2\xi^{[\nu}\theta^{\mu]}\left[\delta\right]\right) \mathrm{d}\mathcal{C}_{\mu\nu} = 0. \tag{39}$$

The boundary $\partial\mathcal{C}$ consists of two components, the intersection of $\mathcal{C}$ with the spatial null infinity $\mathcal{I}^0$, $\mathcal{C}\cap\mathcal{I}^0$, and its intersection with the Killing horizon $\mathcal{H}$, $\mathcal{C}\cap\mathcal{H}$. We first look at the contribution of $\mathcal{C}\cap\mathcal{I}^0$, which we split into the time translational

$$\int_{\mathcal{C}\cap\mathcal{I}^0} \left(\delta Q_t^{\nu\mu} - 2t^{[\nu}\theta^{\mu]}\left[\delta\right]\right) \mathrm{d}\mathcal{C}_{\mu\nu}, \tag{40}$$

and the rotational part

$$\Omega_{\mathcal{H}} \int_{\mathcal{C}\cap\mathcal{I}^0} \left(\delta Q_\varphi^{\nu\mu} - 2\varphi^{[\nu}\theta^{\mu]}\left[\delta\right]\right) \mathrm{d}\mathcal{C}_{\mu\nu} = \Omega_{\mathcal{H}} \int_{\mathcal{C}\cap\mathcal{I}^+} \delta Q_\varphi^{\nu\mu} \mathrm{d}\mathcal{C}_{\mu\nu}, \tag{41}$$

where we use that φ^μ is tangent to the surface $\mathcal{C}\cap\mathcal{I}^0$. The charges at infinity related with the time translations and rotations are usually interpreted as total mass and total angular momentum of the spacetime, respectively. In particular, we have

$$\delta\mathcal{M} = \int_{\mathcal{C}\cap\mathcal{I}^+} \left(\delta Q_t^{\nu\mu} - 2t^{[\nu}\theta^{\mu]}\left[\delta\right]\right) \mathrm{d}\mathcal{C}_{\mu\nu}, \tag{42}$$

for the perturbation of the total mass, and

$$\delta\mathcal{J} = -\int_{\mathcal{C}\cap\mathcal{I}^+} \delta Q_\varphi^{\nu\mu} \mathrm{d}\mathcal{C}_{\mu\nu}, \tag{43}$$

for the perturbation of the total angular momentum. One can show that these reduce to the Arnowitt-Deser-Misner (ADM) prescriptions for mass and angular momentum, whenever both definitions are applicable [8]. And this is just about all the justification you can get for any expression for mass and angular momentum in general relativity. It all comes down to the fact that they relate to the already known expressions, make sense for the Kerr black hole and in some appropriate (post)-Newtonian limit.

Let us move on to the contribution of the internal boundary $\mathcal{C}\cap\mathcal{H}$ (the minus sign follows from the choice of the oriented area element $\mathrm{d}\mathcal{C}_{\mu\nu}$ which is explained below),

$$-\int_{\mathcal{C}\cap\mathcal{H}} \left(\delta Q_\xi^{\nu\mu} - 2\xi^{[\nu}\theta^{\mu]}\left[\delta\right]\right) \mathrm{d}\mathcal{C}_{\mu\nu}. \tag{44}$$

We can write $\mathrm{d}\mathcal{C}_{\mu\nu} = \epsilon_{\mu\nu}\mathrm{d}^2\mathcal{A}$, where $\mathrm{d}^2\mathcal{A}$ is the coordinate area element and $\epsilon_{\mu\nu}$ is the 2-dimensional antisymmetrisation symbol defined so that $\epsilon_{01} = 1$. Since ξ^μ is tangent to the horizon (and also normal to it, as it is a null surface), it can be shown that any term containing the Killing vector ξ^μ vanishes (in particular, the second term in the integrand (44), $-2\xi^{[\nu}\theta^{\mu]}[\delta]$ does not contribute). Moreover, the covariant derivative of ξ^μ on the horizon obeys $\sqrt{-g}\nabla^\nu\xi^\mu = \kappa\sqrt{h}\epsilon^{\nu\mu}$, where h is the determinant of the induced 2-metric on the spacelike surface $\mathcal{C}\cap\mathcal{H}$ (as an example, for a Kerr-Newman black hole in Boyer-Lindquist coordinates it holds $\sqrt{h} = r^2\sin\theta$). Then, we have for $\delta Q_\xi^{\nu\mu}$

$$-\delta Q_\xi^{\nu\mu} = \frac{1}{16\pi}\delta\left(\sqrt{-g}\nabla^{[\nu}\xi^{\mu]}\right) = -\frac{1}{16\pi}\delta\left(\kappa\sqrt{h}\right)\epsilon^{\nu\mu}. \tag{45}$$

We impose $\delta\kappa = 0$ (this is not necessarily the case for a generic perturbation, but one runs into problems without this assumption). Now, we finally obtain

$$-\int_{\mathcal{C}\cap\mathcal{H}}\left(\delta Q_\xi^{\nu\mu} - 2\xi^{[\nu}\theta^{\mu]}[\delta]\right)\mathrm{d}\mathcal{C}_{\mu\nu} = -\int_{\mathcal{C}\cap\mathcal{H}}\delta Q_\xi^{\nu\mu}\epsilon_{\mu\nu}\mathrm{d}^2\mathcal{A} = -\frac{1}{16\pi}\int_{\mathcal{C}\cap\mathcal{H}}\kappa\epsilon^{\nu\mu}\epsilon_{\mu\nu}\delta\sqrt{h}\mathrm{d}^2\mathcal{A}. \tag{46}$$

We can use that κ is constant on the horizon (the zeroth law of black hole mechanics) and that $\epsilon^{\nu\mu}\epsilon_{\mu\nu} = -2$ (very easy to show). We have

$$-\int_{\mathcal{C}\cap\mathcal{H}}\left(\delta Q_\xi^{\nu\mu} - 2\xi^{[\nu}\theta^{\mu]}[\delta]\right)\mathrm{d}\mathcal{C}_{\mu\nu} = -\frac{\kappa}{8\pi}\delta\mathcal{A}. \tag{47}$$

Putting together the contributions from the total mass perturbation (40), the angular momentum perturbation (41) and the horizon term (47) to equation (39), we get

$$\delta\mathcal{M} - \Omega_\mathcal{H}\delta\mathcal{J} - \frac{\kappa}{8\pi}\delta\mathcal{A} = 0. \tag{48}$$

This is the first law of black hole mechanics in vacuum. It relates infinitesimal changes in mass and angular momentum to the corresponding change of horizon area. We will discuss the physical content of this law in the following.

II. BEKENSTEIN ENTROPY

At last, we return to the question we began with: “What happens when you pour a cup of tea into a black hole?”. To answer it, we first need a physical description of the tea. We will actually consider a somewhat simpler situation of a perfect fluid rotating around a black hole. We assume the resulting spacetime is again stationary and asymptotically flat. To derive the first law of black hole mechanics in the Noether charge formalism in this case, we need to compute the symplectic potential, Noether current and Noether charge corresponding to the perfect fluid Lagrangian. As we have remarked, the formalism works for any diffeomorphism invariant theory, so there are no conceptual issues. However, the practical calculations are cumbersome (see [10] for the first law and [11] for Lagrangian description of relativistic perfect fluids). We only present the final result, i.e., the first law of black hole mechanics in the presence of a perfect fluid

$$\delta\mathcal{M}-\Omega_{\mathcal{H}}\delta\mathcal{J}_{\mathcal{H}}-\frac{1}{8\pi}\kappa\delta\mathcal{A}-\int_{\mathcal{C}}\mu\delta N^{\mu}\mathrm{d}\mathcal{C}_{\mu}-\int_{\mathcal{C}}\mathcal{T}\delta S^{\mu}\mathrm{d}\mathcal{C}_{\mu}-\int_{\mathcal{C}}\Omega\delta J^{\mu}\mathrm{d}\mathcal{C}_{\mu}=0. \tag{49}$$

Here, $\mathcal{M}$ is as before the total mass of the spacetime, $\mathcal{J}_{\mathcal{H}}$ angular momentum of the black hole, μ the chemical potential of the fluid (as measured at infinity), N^{μ} the particle flux, $\mathcal{T}$ the fluid temperature (measured at infinity), S^{μ} the entropy flux, Ω fluid's angular velocity (measured at infinity), and J^{μ} the angular momentum of the fluid.

It is important to appreciate the physical content of equation (49). We have the particles of the fluid falling into the black hole. That is no problem, number of particles needs not be conserved and the mass of the black hole increases correspondingly to the loss of mass in the exterior. Likewise, we have the transfer of angular momentum between the fluid and the black hole (in both directions). Since the black hole has a well defined angular momentum that also changes, there is no issue here. However, we have entropy of the perfect fluid disappearing inside the black hole. And there is no corresponding black hole entropy that can increase. Then, the total entropy of the spacetime seemingly decreases. Here, we see a very explicit violation of the second law of thermodynamics.

Let us look for a way out. There is still one term in the first law which we have not interpreted in any way,

$$-\frac{1}{8\pi}\kappa\delta\mathcal{A}. \tag{50}$$

First, since κ is the surface gravity of the horizon and $\mathcal{A}$ the horizon area, this contribution really comes from the presence of the horizon rather than from the fluid. Second, we know from curved spacetime quantum field theory that $T_{\mathrm{H}}=\kappa/(2\pi)$ is the famous Hawking temperature, at which black holes radiate[9]. Then, we can write

$$-\frac{1}{8\pi}\kappa\delta\mathcal{A}=-T_{\mathrm{H}}\frac{\delta\mathcal{A}}{4}. \tag{51}$$

[9] We are slightly cheating by bringing a quantum result into an otherwise fully classical formulation of the first law. The sad truth is that there is no known way to completely eliminate this shortcoming.

It is very tempting to interpret this term as the heat flux $-T\delta S$, i.e., identify the black hole entropy as

$$S_{\mathrm{B}} = \frac{\mathcal{A}}{4}, \tag{52}$$

or, restoring all the constants for a moment,

$$S_{\mathrm{B}} = k_{\mathrm{B}}\frac{\mathcal{A}}{4l_{\mathrm{P}}^2}, \tag{53}$$

with l_{P} being the Planck constant. This is the famous Bekenstein entropy of a black hole[10]. To compute it, one just needs to take a spacelike 2-dimensional cross-section of the Killing horizon orthogonal to the Killing vector field (in adapted coordinates a 2-surface defined by any constant t and r equal to the horizon radius) and divide by four. If the black hole does not posses a Killing horizon, we take a different notion of the horizon (typically an apparent horizon) and do the same thing (but then entropy depends on how we choose our spacelike slice and everything becomes messy). With this identification of entropy the first law of black hole mechanics becomes a genuine first law of thermodynamics

$$\delta\mathcal{M} - \Omega_{\mathcal{H}}\delta\mathcal{J}_{\mathcal{H}} - T_{\mathrm{H}}\delta S_{\mathrm{B}} - \int_{\mathcal{C}}\mu\delta N^{\mu}\mathrm{d}\mathcal{C}_{\mu} - \int_{\mathcal{C}}\mathcal{T}\delta S^{\mu}\mathrm{d}\mathcal{C}_{\mu} - \int_{\mathcal{C}}\Omega\delta J^{\mu}\mathrm{d}\mathcal{C}_{\mu} = 0. \tag{54}$$

Looking at the first law of black hole thermodynamics with a perfect fluid (54) and taking into account the existence of Hawking temperature (which is completely independent of gravitational dynamics), it should not be too hard to accept that black hole entropy is given by the Bekenstein prescription (52). Nevertheless, this definition of entropy has some peculiarities. Most notably, the heat capacity of the simplest black hole solution, a Schwarzschild black hole is negative

$$C = T\frac{\partial S}{\partial T} = -8\pi M^2 < 0. \tag{55}$$

In other words, by swallowing matter, the black hole simultaneously cools down and increases its entropy. Therefore, it cannot achieve a thermodynamic equilibrium with an external heat bath (unless one introduces some artificial way to make the heat capacity positive,

[10] A quick note on terminology and history: the often used name Bekenstein-Hawking entropy feels unfair, since Bekenstein was the one to propose the concept of entropy proportional to horizon area [12] (albeit partially inspired by Hawking's area increase theorem). Hawking was originally quite sceptical of the notion of black hole entropy, which is apparent in his paper discussing the four laws of black hole mechanics [13] (ostensibly not thermodynamics). His main contribution lies in the realisation that black holes radiate, which is reflected by naming the black hole temperature after him [14].

To further complicate things, the result of the calculation which we just carried out can be called Wald entropy. However, this name usually refers to entropy derived by the Noether charge approach in theories generalising general relativity.

e.g. a negative cosmological constant). Since the standard Clausius entropy is well defined only in the thermodynamic equilibrium, we need a different way to interpret Bekenstein entropy. As already noticed by Bekenstein, a natural option are definitions of entropy related to information, i.e., classical Shannon entropy and quantum von Neumann entropy. These entropies are defined without any reference to thermodynamic equilibrium. Moreover, since the black hole horizon hides information contained inside it from external observers, assigning to it entropy related to this lack of information is very natural.

To be a good definition of entropy, Bekenstein entropy must also satisfy the second law of thermodynamics. In other words, we require that the sum of Bekenstein entropy of the black hole and the total entropy of its exterior never decreases. This is known as the generalised second law of thermodynamics. This law has been shown to hold in very general settings [15]. Here, we will limit ourselves to showing the validity of the generalised second law in two interesting situations, already analysed in the Bekenstein's original paper [12].

First, consider a merger of two Schwarzschild black holes with masses M_1 and M_2. For simplicity, we assume that no gravitational waves are emitted and the final black hole has mass $M_3 = M_1 + M_2$. Since the horizon area of a Schwarzschild black hole of mass M is $\mathcal{A} = 4\pi r_{\mathrm{S}}^2 = 16\pi M^2$, we have

$$S_{\mathrm{B1}} + S_{\mathrm{B2}} = 4\pi \left(M_1^2 + M_2^2 \right) < S_{\mathrm{B3}} = 4\pi \left(M_1^2 + M_2^2 + 2M_1 M_2 \right) . \tag{56}$$

Hence, the final entropy is greater than the initial one and the generalised second law holds. Amazingly, the validity of the generalised second law in real life mergers has been even experimentally confirmed by gravitational waves observations [16].

Second, we let a quantum harmonic oscillator fall inside a Kerr-Newman black hole (because a free particle has no entropy and this is the next simplest thing we can do). We model the oscillator as two particles of mass $m/2$, joined by a spring with a spring constant k. We place this oscillator inside a box kept at temperature T such that $T \ll m$, so the thermal effects on the oscillator can be described nonrelativistically. The vibrational frequency of the oscillator is $\omega = 2\sqrt{k/m}$. The probability that the oscillator is in the quantum state with frequency $n\omega$, with n being any natural number, is given by the Maxwell-Boltzmann distribution

$$p_n = \frac{e^{-nx}}{1 - e^{-x}}, \tag{57}$$

where we denote $x = \omega/T$. For the expectation value of the oscillator's energy we then have $\langle E\rangle = \left[1/\left(e^x - 1\right) + 1/2\right]\omega$ and for its entropy $S_\mathrm{o} = -\ln\left(1 - e^{-x}\right) + x/\left(e^x - 1\right)$.

Now we lower the oscillator inside the black hole. If it were a point-like particle, the analysis of the geodesic motion in Kerr-Newman spacetime show that it could be absorbed by the black hole without increasing its horizon area, provided that the radial velocity of the particle in the moment of crossing the horizon is zero. Then, Bekenstein entropy of the black hole would not decrease and since the entropy of the oscillator vanished beyond the horizon, the generalised second law would be violated. However, and this is important, our oscillator cannot be a point-like particle, because thermal fluctuations lead to motion of the balls forming the oscillator. Hence, we have nonzero oscillations in the distance y of the balls $\Delta y = y - \langle y\rangle$, with $\langle y\rangle$ being the mean distance. Clearly, radius b of the box containing the harmonic oscillator must be sufficient to accommodate these oscillations, i.e., $b \geq \Delta y/2$ (otherwise the balls will keep hitting the walls of the box, completely changing the dynamics of the system). The virial theorem then implies that the potential energy due to thermal oscillations $m\omega^2\Delta y^2/8$ is equal to $\langle E\rangle/2$. From this we have $b \geq \Delta y = \sqrt{\langle E\rangle/m}/\omega$.

It is natural to ask whether we can eliminate thermal fluctuations. However, this would require to cool the oscillator to zero temperature. Then, the third law of thermodynamics implies zero entropy. So, if the oscillator carries nonzero entropy, it must fluctuate and, hence, has finite size. This observation is absolutely crucial for dealing with entropy in the context of gravitational physics. We will return to this point later.

If we lower an object of radius b and mass-energy μ into a Kerr-Newman black hole, Bekenstein showed that the minimal corresponding increase of the horizon area is $\delta\mathcal{A} = 8\pi b\mu$ [12] (the argument is fairly technical and involves concepts such as the Carter constant, so we do not reproduce it here). This implies the change in Bekenstein entropy $\delta S_\mathrm{B} = 2\pi b\mu$, where $b \geq \sqrt{\langle E\rangle/m}/\omega$ and $\mu \geq m + \langle E\rangle$. Recalling that entropy of the oscillator equals $S_\mathrm{o} = -\ln\left(1 - e^{-x}\right) + x/\left(e^x - 1\right)$, the total change of entropy due to its fall into a black hole reads

$$\delta S_\mathrm{total} = \delta S_\mathrm{B} - S_\mathrm{o} \geq \left(2\pi/\omega\right)\left(m + \langle E\rangle\right)\sqrt{\langle E\rangle/m} + \ln\left(1 - e^{-x}\right) - x/\left(e^x - 1\right). \qquad (58)$$

The right hand side has a global minimum with respect to x given implicitly by

$$x_\mathrm{min} = \left(2\pi/\omega\right)\left(m/\langle E\rangle + 1\right)\sqrt{\langle E\rangle/m} \qquad (59)$$

(keep in mind that $\langle E\rangle$ depends on x). At $x = x_{\rm min}$, the change of entropy obeys

$$\delta S_{\rm total} \geq x_{\rm min}/2 + \ln\left(1 - e^{-x_{\rm min}}\right), \tag{60}$$

or, restoring $x_{\rm min} = (\omega/T)_{\rm min}$,

$$\delta S_{\rm total} \geq (\omega/T)_{\rm min}/2 + \ln\left(1 - e^{-(\omega/T)_{\rm min}}\right). \tag{61}$$

The first term is always positive (we have $\omega \geq 0$, $T \geq 0$). For a nonrelativistic oscillator it must hold $\langle E\rangle \ll m$, and it follows that also $T \ll \omega$. Hence, $e^{-(\omega/T)_{\rm min}} \ll 1$ and we can Taylor-expand the logarithm to the leading order

$$\delta S_{\rm total} \geq (\omega/T)_{\rm min}/2 - e^{-(\omega/T)_{\rm min}}, \tag{62}$$

which is clearly positive. Hence, the generalised second law is obeyed in this case, lending further support to Bekenstein expression for black hole entropy. Of course, the vast literature on the subject contains many further arguments in favour of Bekenstein entropy (and negligibly few against it), but we hope that the low-tech discussion we provided suffices to convince you.

To conclude, we return to our brief discussion of the size of any system with entropy. In fact, the generalised second law is only valid if the amount of entropy that can be contained in a small spacetime region is finite. In particular, we need $S \leq 2\pi RE$, where R is the radius of the system and E its total energy. This is known as the Bekenstein entropy bound. It is saturated for a Schwarzschild black hole ($R = 2M$, $E = M$), which can then be conjectured to be the maximally entropic state of the matter. In the context of flat spacetime quantum field theory, we even have an explicit proof of the Bekenstein bound [17]. However, the Bekenstein bound does not make much sense beyond weak gravity regime (because then we cannot even properly define E and R). Nevertheless, the covariant entropy bound later proposed by Bousso still applies and is sufficient to ensure the validity of the generalised second law of thermodynamics [18]. The Bousso bound is formulated in the following way. Consider any closed, spacelike 2-dimensional surface B of area $\mathcal{A}$ in a general curved spacetime. Four congruences of null geodesics cross B, two to its past and two to its future. The area of the spatial cross-section of a null geodesic congruence is in general not constant, but either decreases or increases along them. It can be shown (e.g. by using the Raychaudhuri equation) that at least one of the null geodesic congruences pointing to the future of 2-surface B has a decreasing (or constant, in flat spacetime) area of spatial cross-section. The total entropy S contained in this contracting null geodesic congruence is then bounded by $S \leq \mathcal{A}/4$, i.e., by the Bekenstein entropy corresponding to the 2-surface B. Under some reasonable assumptions such as the validity of the (quantum) null energy condition, there are no known counterexamples to the covariant entropy bound.

III. BLACK HOLE EUCLIDEAN GRAND-CANONICAL ENSEMBLE

The covariant phase space perspective on black hole thermodynamics, while elegant, suffers from serious drawbacks. Most notably, it fails to identify the black hole temperature, which must be heuristically fixed to the known Hawking value. The same issue actually occurs in regards to all the intensive thermodynamic potentials, such as the angular velocity or the electrostatic potential. When exploring more complicated black hole spacetimes (e.g., C-metric, Taub-NUT, solutions in scalar-tensor theories), the covariant phase space formalism quickly loses the power to provide a correct and unique thermodynamic interpretation of the first law.

To unambiguously define the intensive potentials, we need to construct a black hole grand-canonical ensemble. Then, the values of the intensive potentials describe the equilibrium configuration and we can obtain a full, unambiguous thermodynamic description (as we show in the following).

One standard way to construct a grand-canonical ensemble lies in considering a path integral of the theory with a Wick-rotated, imaginary time parameter $\tau = it$

$$Z = \int \mathcal{D}\psi \exp\left(-\int_{\tau_1}^{\tau_2} \mathrm{d}\tau L\left[\psi\right]\right), \tag{63}$$

where ψ collectively denotes the dynamical fields, $\mathcal{D}\psi$ represent a formal integration measure on the field configuration space and $L\left[\psi\right]$ denotes the Lagrangian. If we restrict the integration to configurations in thermodynamic equilibrium, it can be shown that Z has the interpretation of the (grand-)canonical partition function.

To understand how the equilibrium configurations behave, consider a generic τ-dependent observable in equilibrium

$$\hat{A}\left(\tau\right) = e^{\hat{H}\tau}\hat{A}\left(0\right)e^{-\hat{H}\tau}, \tag{64}$$

and its thermodynamic average

$$A\left(\tau\right)\equiv\mathrm{Tr}\left(\hat{\rho}\hat{A}\left(\tau\right)\right)=\frac{1}{Z}\mathrm{Tr}\left(e^{-\beta\hat{H}}\hat{A}\left(\tau\right)\right),\tag{65}$$

where β is the inverse temperature, $\hat{H}$ Hamiltonian operator, $\hat{\rho}=e^{-\beta\hat{H}}/Z$ the density operator and $Z=\mathrm{Tr}e^{-\beta\hat{H}}$ denotes the partition function. Let us now look at the value of the average at time $\tau+\beta$

$$A\left(\tau+\beta\right)=\frac{1}{Z}\mathrm{Tr}\left(e^{-\beta\hat{H}}e^{\beta\hat{H}}\hat{A}\left(\tau\right)e^{-\beta\hat{H}}\right)=\frac{1}{Z}\mathrm{Tr}\left(e^{-\beta\hat{H}}\hat{A}\left(\tau\right)\right)=A\left(\tau\right),\tag{66}$$

where we used that $e^{-\beta\hat{H}}e^{\beta\hat{H}}=\mathrm{Id}$ and the cyclicity of the trace. It follows that $A\left(\tau\right)$ is periodic in the Euclidean time with a period equal to the inverse temperature β. This periodicity is known as the Kubo-Martin-Schwinger (KMS) condition which serves to identify thermal states in quantum physics.

Therefore, to obtain the partition function, we must restrict the path integral to configurations periodic in the imaginary time. To the leading order, its logarithm can then be approximated by

$$\ln Z\approx-\int_{\tau_0}^{\tau_0+\beta}\mathrm{d}\tau L\left[\psi\right]\bigg|_{\text{stationary}}.\tag{67}$$

where the inverse temperature β is chosen so that it corresponds to a stationary point of the action, i.e., $\partial\ln Z/\partial\psi=0$. The free energy then equals

$$F=-\frac{1}{\beta}\ln Z=\frac{1}{\beta}\int_{\tau_0}^{\tau_0+\beta}\mathrm{d}\tau L\left[\psi\right]\bigg|_{\text{stationary}}.\tag{68}$$

Now we simply need to apply this algorithm to a stationary black hole spacetime, following the off-shell method introduced by Braden, Brown, Whiting and York [19]. We focus on the class of static, spherically symmetric spacetimes in $4D$, which can be described by the following metric

$$\mathrm{d}s^2=-b^2\left(r\right)\mathrm{d}t^2+a^2\left(r\right)\mathrm{d}r^2+r^2\mathrm{d}\Omega_2,\tag{69}$$

where $b\left(r\right)$, $a\left(r\right)$ are functions of the radial coordinate r and $r^2\mathrm{d}\Omega_2$ denotes the area element on a 2-sphere. Wick-rotating the time coordinate yields

$$\mathrm{d}s^2=b^2\left(r\right)\mathrm{d}\tau^2+a^2\left(r\right)\mathrm{d}y^2+r^2\mathrm{d}\Omega_2,\tag{70}$$

where we choose the Euclidean time coordinate to be real (for consistency with the seminal paper [19]) and 2π-periodic, i.e., $\tau\in[0,2\pi]$.

It can be shown that the Euclidean black hole metric is only real for $r > r_+$ [20] and we have to limit our path integral to these values of r. Moreover, as we have discussed, the heat capacity of a black hole is negative, making it impossible to construct a grand-canonical ensemble for it. To circumvent this issue, one can introduce an artificial timelike boundary located at some finite radial distance $r = r_{\mathrm{b}}$ (known as York boundary). Specifying reflective (Dirichlet) boundary conditions, i.e., fixed metric at $r = r_{\mathrm{b}}$, allows us to construct a grand-canonical ensemble consisting of the black hole and the radiation contained between the horizon and York boundary.

In summary, we restrict our Euclidean metric to $r_+ \leq r \leq r_{\mathrm{b}}$. It becomes convenient to introduce a new coordinate $y \in [0,1]$ such that $r\,(y=0) = r_+$ and $r\,(y=1) = r_{\mathrm{b}}$. We then work with the following metric

$$\mathrm{d}s^2 = b^2\,(y)\,\mathrm{d}\tau^2 + a^2\,(y)\,\mathrm{d}y^2 + r^2\,(y)\,\mathrm{d}\Omega_2. \tag{71}$$

We impose the Dirichlet boundary conditions by fixing $r\,(1) = r_{\mathrm{b}}$ and $b\,(1)$. Since the black hole is a thermodynamic system obeying the KMS condition, the Euclidean time period on the boundary

$$\beta = \int_0^{2\pi} b\,(1)\,\mathrm{d}\tau = 2\pi b\,(1) \tag{72}$$

gives the inverse temperature measured by an observer who is stationary on York boundary. In this way, the Euclidean grand-canonical approach directly specifies the physical temperature of the black hole.

We further require that the metric is regular on the horizon, to avoid a conical singularity whose contribution to the partition function is difficult to analyse. Any hypersurface of constant r, including the horizon, has topology $S^1 \times S^2$, where S^1 corresponds to the periodic Euclidean time and S^2 is the 2-sphere of constant r, τ. At the horizon, the S^1 circle degenerates to a point, implying $b\,(0) = 0$. To make this point smooth, the $\tau - y$ plane near $y = 0$ must be isometric to a flat geometry (otherwise a conical singularity will appear). Using that $b\,(0) = 0$, the metric at this 2-plane reads

$$b^2\,(y)\,\mathrm{d}\tau^2 + a^2\,(y)\,\mathrm{d}y^2 \approx b'^2\,(0)\,y^2\mathrm{d}\tau^2 + a^2\,(y)\,\mathrm{d}y^2 = a^2\,(y)\left[\left(\frac{b'\,(0)}{a\,(y)}\right)^2 y^2\mathrm{d}\tau^2 + \mathrm{d}y^2\right]. \tag{73}$$

If it holds $[b'/a]_{y=0} = 1$ this metric is indeed conformal to a flat disc $\mathrm{d}s^2 = y^2\mathrm{d}\tau^2 + \mathrm{d}y^2$ and we have a smooth geometry. Lastly, the Euler characteristic of the horizon given by the

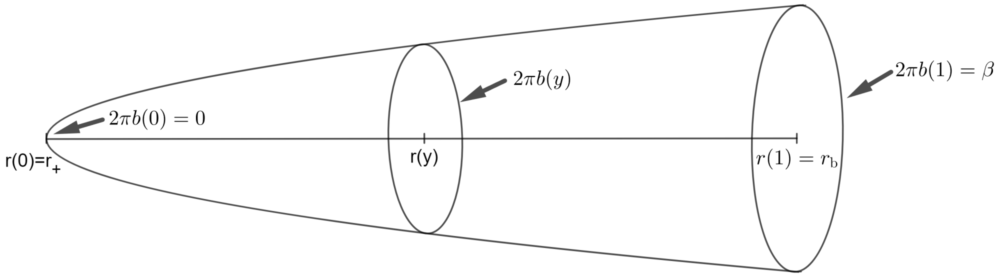

FIG. 2. The black hole geometry we consider. The radial coordinate y is measured along the horizontal axis, the periodic Euclidean time τ runs along the circles and we suppress the angular coordinates. The proper length of the circles is $2\pi b\,(y)$, going (smoothly) to 0 at the horizon $y = 0$ and corresponding to the inverse temperature β at York boundary $y = 1$.

Chern-Gauss-Bonnet formula must be $\chi = 2$ (as for S^2), i.e.,

$$\chi = \frac{1}{2\pi}\int\left\{\frac{1}{r_+^2}\frac{b'\,(0)}{a\,(0)}\left[1-\frac{r'^2\,(0)}{a^2\,(0)}\right]\right\}r_+^2\mathrm{d}\Omega_2 = 2, \tag{74}$$

where the term in the braces is the Gaussian curvature of the horizon. This condition holds if and only if $r'\,(0)\,/a\,(0) = 0$. In total, we need to fix the following regularity conditions on the horizon

$$b\,(0) = 0, \tag{75}$$

$$\left.\frac{b'}{a}\right|_{y=0} = 1, \tag{76}$$

$$\left.\frac{r'}{a}\right|_{y=0} = 0. \tag{77}$$

We sketch the Euclidean black hole geometry we work with in figure 2.

In principle, the grand-canonical ensemble we discuss works for any static, spherically symmetric and asymptotically flat black hole spacetimes. Nevertheless, for concreteness, we focus on electrovacuum spacetimes in general relativity (following the seminal paper [19]). Thence, we also need to discuss boundary conditions and horizon regularity of the electromagnetic field. To obtain a grand-canonical ensemble, we need to fix the electrostatic potential Φ on the boundary[11]. That can be accomplished by fixing the boundary electromagnetic potential to be $A_\mu\,(1) = (A_\tau\,(1)\,,0,0,0)$, where $A_\tau\,(1)$ is proportional to Φ (we fix

[11] Fixing the electric charge on the boundary yields a canonical ensemble.

the precise relation in the following). On the horizon, we require that the proper orthonormal frame components of the electromagnetic potential are finite, i.e., $\lim_{y\to 0} A_\tau/b < \infty$. Since $b\,(0) = 0$, it follows that $A_\tau\,(0) = 0$.

Our action consists of the Euclidean Einstein-Hilbert action for general relativity and the action for electromagnetic field. Moreover, we need to add boundary terms fixing the Dirichlet conditions for $g_{\mu\nu}$ and A_μ. The latter requires no boundary terms. The standard boundary term for the Einstein-Hilbert action reads

$$\frac{1}{8\pi}\oint \left(K - K^0\right)\sqrt{\gamma}\mathrm{d}^3x, \tag{78}$$

where K denotes the extrinsic curvature of the boundary, K^0 is the extrinsic curvature of the corresponding boundary in flat spacetime (in our case $K^0 = 2/r$) and γ the determinant of the boundary metric. Subtraction of K^0 removes the divergent flat spacetime contribution, ensuring finiteness of the action in the limit $r_\mathrm{b} \to \infty$. In total, our action reads

$$I = -\frac{1}{16\pi}\int \left(R - F_{\mu\nu}F^{\mu\nu}\right)\sqrt{g}\mathrm{d}^4x + \frac{1}{8\pi}\oint \left(K - K^0\right)\sqrt{\gamma}\mathrm{d}^3x. \tag{79}$$

Evaluating it for the static spherically symmetric metric ansatz (71) yields

$$I = -\int_0^{2\pi}\mathrm{d}\tau\int_0^1\mathrm{d}y\left[-\left(\frac{r^2b'}{2a}\right)' - \frac{ab}{2r'}\left(\frac{rr'^2}{a^2} - r\right)' + \frac{r^2}{ab}A_\tau'^2\right] + \int_0^{2\pi}\mathrm{d}\tau\left(-\frac{\left(br^2\right)'}{2a} + br\right)\Bigg|_{y=1}. \tag{80}$$

Integrating the first term by parts then leads to

$$I = -\int_0^{2\pi}\mathrm{d}\tau\int_0^1\mathrm{d}y\left[\frac{r'}{a}rb' + \frac{1}{2}ab\left(\frac{r'^2}{a^2} + 1\right) + \frac{r^2}{ab}A_\tau'^2\right] + \int_0^{2\pi}\mathrm{d}\tau br\Bigg|_{y=1} - \int_0^{2\pi}\mathrm{d}\tau\frac{\left(br^2\right)'}{2a}\Bigg|_{y=0}. \tag{81}$$

Next, we apply the kinematical constraint equations. These are the Hamiltonian constraint of general relativity obtained by varying the action with respect to b

$$\frac{a}{2r'}\left(\frac{rr'^2}{a^2} - r\right)' + \frac{r^2}{ab^2}A_\tau'^2 = 0, \tag{82}$$

and the Gauss law found by varying with respect to A_τ

$$\left(\frac{r^2}{ab}A_\tau'\right)' = 0. \tag{83}$$

Variations with respect to a and r yield evolution equations, which we do not impose, keeping our spacetime off-shell.

Integrating the Gauss law, we obtain

$$A_\tau' = -\frac{ab}{r^2} ie, \tag{84}$$

where we choose the integration constant e so that it has the interpretation of the electric charge in flat spacetime, when we fix the relation between A_τ and the electrostatic potential Φ on York boundary to be

$$A_\tau\left(1\right) = -i\frac{\beta\Phi}{2\pi}, \tag{85}$$

where β appears due to the transformation of A_τ to the proper orthonormal frame, $A_\tau \to A_\tau / b$.

Next, we plug the result of the Gauss law into the Hamiltonian constraint (82), simplify and integrate to find

$$\frac{r'}{a} = \sqrt{1-\frac{r_+}{r}}\sqrt{1-\frac{e^2}{rr_+}}. \tag{86}$$

Using equations (84) and (86), we can integrate the action, obtaining

$$I = \beta r_{\mathrm{b}}\left(1-\sqrt{1-\frac{r_+}{r_{\mathrm{b}}}}\sqrt{1-\frac{e^2}{r_{\mathrm{b}}r_+}}\right) - \beta\Phi e - \pi r_+^2. \tag{87}$$

We now need to find a value of this action corresponding to an equilibrium configuration. This configuration will correspond to a stationary point of I with respect to r_+ and e, i.e.,

$$\frac{\partial I}{\partial r_+} = 0, \tag{88}$$

$$\frac{\partial I}{\partial e} = 0. \tag{89}$$

These conditions imply for β and Φ

$$\beta = \frac{4\pi r_+}{1-\frac{e^2}{r_+^2}}\sqrt{1-\frac{r_+}{r_{\mathrm{b}}}}\sqrt{1-\frac{e^2}{r_+r_{\mathrm{b}}}}, \tag{90}$$

$$\Phi = \frac{Q}{r_+}\sqrt{\frac{1-\frac{r_+}{r_{\mathrm{b}}}}{1-\frac{e^2}{r_+r_{\mathrm{b}}}}}. \tag{91}$$

They correspond to the inverse Hawking temperature $T_{\mathrm{H}} = \left(1-e^2/r_+^2\right)/\left(4\pi r_+\right)$ and the electrostatic potential $\Phi = e/r_+$ of the Reissner-Nordström black hole blue-shifted to York boundary at $r = r_{\mathrm{b}}$. In the limit $r_{\mathrm{b}} \to \infty$ we recover the standard expressions. That these expressions are physically determined and even correctly include the blue shift factors represents the main advantage of the Euclidean approach over the covariant phase space formalism.

In the following, unless specified otherwise, we assume that we have removed York boundary by taking the limit $r_{\mathrm{b}} \to \infty$. As we discussed, the action in thermodynamic equilibrium provides a classical approximation for the grand-canonical free energy

$$F\left[\beta, \Phi, r_{\mathrm{b}}\right] = \frac{I}{\beta} = \frac{r_+}{2} + \frac{e^2}{2r_+} - \Phi e - \frac{\pi r_+^2}{\beta} \tag{92}$$

$$= \frac{\beta\left(1-\Phi^2\right)^2}{16\pi}, \tag{93}$$

where we expressed r_+, e in terms of β, Φ which are the thermodynamic potentials fixed in the grand-canonical ensemble.

The standard analysis of a grand-canonical ensemble now yields expressions for entropy

$$S = \beta^2 \left(\frac{\partial F}{\partial \beta}\right)_{\Phi} = \pi r_+^2 = \frac{\mathcal{A}}{4}, \tag{94}$$

and mean values of charge

$$\langle Q \rangle = -\left(\frac{\partial F}{\partial \Phi}\right)_{\beta} = e, \tag{95}$$

and of internal energy

$$\langle E \rangle = F + \frac{S}{\beta} + \Phi \langle Q \rangle = \frac{r_+}{2} + \frac{e^2}{2r_+} = M, \tag{96}$$

where M denotes the ADM mass of a Reissner-Nordström black hole. In summary, in the limit $r_{\mathrm{b}} \to \infty$ we recover the standard thermodynamic description of a Reissner-Nordström black hole.

One can also analyse the grand-canonical ensemble for finite r_{b}, i.e., in presence of York boundary. Then, one obtains [19]

$$S = \frac{\mathcal{A}}{4}, \tag{97}$$

$$\langle Q \rangle = e, \tag{98}$$

$$\langle E \rangle = r_{\mathrm{b}} \left(1 - \sqrt{1 - \frac{r_+}{r_{\mathrm{b}}}} \sqrt{1 - \frac{e^2}{r_{\mathrm{b}} r_+}}\right). \tag{99}$$

In this case, the first law of thermodynamics contains an extra term corresponding to variations of the area of York boundary, $\mathcal{A}_{\mathrm{b}} = \pi r_{\mathrm{b}}^2$ and reads

$$\delta \langle E \rangle = \frac{1}{\beta} \delta S + \Phi \delta \langle Q \rangle - \lambda \delta \mathcal{A}_{\mathrm{b}}, \tag{100}$$

where λ is the positive surface pressure on the York boundary

$$\lambda = \frac{1}{8\pi r_{\mathrm{b}}} \left[\frac{1 - \frac{r_+}{2r_{\mathrm{b}}}\left(1 + \frac{e^2}{r_+^2}\right)}{\sqrt{1 - \frac{r_+}{r_{\mathrm{b}}}}\sqrt{1 - \frac{e^2}{r_+ r_{\mathrm{b}}}}} - 1 \right] . \tag{101}$$

Of course, this extra term in the first law vanishes in the limit $r_{\mathrm{b}} \to \infty$. Nevertheless, it shows that the presence of York boundary does significantly affect thermodynamics of the system. With the boundary present, we do not really study a Reissner-Nordström black hole, but rather an equilibrium state of a black hole and its surrounding radiation, whose effects shows up in the presence of the surface pressure λ.

[1] C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Princeton University Press, Princeton, 2017).

[2] J. Podolský, Teoretická mechanika v jazyce diferenciální geometrie (2019).

[3] J. Lee and R. M. Wald, Local symmetries and constraints, J. Math. Phys. **31**, 725 (1990).

[4] T. Jacobson and M. R. Visser, Gravitational thermodynamics of causal diamonds in (A)dS, SciPost Phys. **7**, 079 (2019).

[5] R. M. Wald, On identically closed forms locally constructed from a field, Journal of Mathematical Physics **31**, 2378 (1990).

[6] R. M. Wald and A. Zoupas, A general definition of "conserved quantities" in general relativity and other theories of gravity, Phys. Rev. D **61**, 084027 (2000).

[7] R. M. Wald, Black hole entropy is Noether charge, Phys. Rev. D **48**, 3427 (1993).

[8] V. Iyer and R. M. Wald, Some properties of Noether charge and a proposal for dynamical black hole entropy, Phys. Rev. D **50**, 846 (1994).

[9] D. Harlow and J.-q. Wu, Covariant phase space with boundaries, J. High Energ. Phys. **2020** (10).

[10] V. Iyer, Lagrangian perfect fluids and black hole mechanics, Phys. Rev. D **55**, 3411 (1997).

[11] J. D. Brown, Action functionals for relativistic perfect fluids, Class. Quant. Grav. **10**, 1579 (1993).

[12] J. D. Bekenstein, Black Holes and Entropy, Phys. Rev. D **7**, 2333 (1973).

[13] J. D. Bardeen, B. Carter, and S. W. Hawking, Black holes and entropy, Phys. Rev. D **7**, 2333

(1973).

[14] S. W. Hawking, Particle creation by black holes, Commun. Math. Phys. **43**, 199 (1975).

[15] A. C. Wall, Proof of the generalized second law for rapidly changing fields and arbitrary horizon slices, Phys. Rev. D **85**, 104049 (2012).

[16] M. Cabero, C. D. Capano, O. Fischer-Birnholtz, B. Krishnan, A. B. Nielsen, A. H. Nitz, and C. M. Biwer, Observational tests of the black hole area increase law, Phys. Rev. D **97**, 124069 (2018).

[17] H. Casini, Relative entropy and the bekenstein bound, Class. Quant. Grav. **25**, 205021 (2008).

[18] R. Bousso, A covariant entropy conjecture, J. High Energ. Phys. **1999** (004).

[19] H. W. Braden, J. D. Brown, B. F. Whiting, and J. W. York Jr., Charged black hole in a grand canonical ensemble, Phys. Rev. D **42**, 3376 (1990).

[20] G. W. Gibbons and S. W. Hawking, Action integrals and partition functions in quantum gravity, Phys. Rev. D **15**, 2752 (1977).