

General relativity from a thermodynamic perspective

T. Padmanabhan

Received: 19 December 2013 / Accepted: 27 January 2014 / Published online: 13 February 2014
© Springer Science+Business Media New York 2014

Abstract I show that the gravitational dynamics in a bulk region of space can be connected to a thermodynamic description in the boundary of that region, thereby providing clear physical interpretations of several mathematical features of classical general relativity: (1) The Noether charge contained in a bulk region, associated with a specific time evolution vector field, has a direct thermodynamic interpretation as the gravitational heat content of the boundary surface. (2) This result, in turn, shows that all static spacetimes maintain holographic equipartition in the following sense: In these spacetimes, the number of degrees of freedom in the boundary is equal to the number of degrees of freedom in the bulk. (3) In a general, evolving spacetime, the rate of change of gravitational momentum is related to the difference between the number of bulk and boundary degrees of freedom. It is this departure from the holographic equipartition which drives the time evolution of the spacetime. (4) When the equations of motion hold, the (naturally defined) total energy of the gravity plus matter within a bulk region, will be equal to the boundary heat content. (5) After motivating the need for an alternate description of gravity (if we have to solve the cosmological constant problem), I describe a thermodynamic variational principle based on null surfaces to achieve this goal. The concept of gravitational heat density of the null surfaces arises naturally from the Noether charge associated with the null congruence. The variational principle, in fact, extremises the total heat content of the matter plus gravity system. Several variations on this theme and implications are described.

Keywords Horizon entropy · Emergent gravity · Entropic gravity · Holography · Equipartition · Black hole · Noether current

T. Padmanabhan (✉)
IUCAA, Pune University Campus, Ganeshkhind, Pune 411 007, India
e-mail: paddy@iucaa.ernet.in

1 Introduction and summary of the results

The uncanny similarity between the laws of black hole dynamics and classical thermodynamics, discovered in the seventies (e.g., [1–5]; for a review, see [6]), suggested a possible connection between gravitational dynamics and horizon thermodynamics. This was further strengthened by the realization that black holes are in no way special and that thermodynamic parameters like temperature will be attributed to *any* null surface by a class of observers who perceive the null surface as a horizon due to their state of motion, even in flat spacetime [7,8]. This allows one to introduce the concept of local Rindler observers around any event in a spacetime [9,10] just as one introduces freely falling observers around any event. The fact, that observers in different states of motion will attribute different thermodynamic features to null surfaces, introduces a new level of observer dependence into the physical theory (see Sect. 4 of [11], Sect. 4.4 of [12,13]).

Several further investigations regarding the dynamics of gravity have shown that this connection is far deeper than was originally suspected. In particular, we now know that the relation between horizon thermodynamics and gravitational dynamics transcends Einstein’s theory of gravity and holds for a much wider class of models like, e.g., Lanczos–Lovelock models [14–16]. Such investigations have led to the emergent gravity paradigm in which the gravitational field equations have the same status as the equations of elasticity or fluid mechanics (for reviews, see e.g., [12,13,17]). The main purpose of this paper is to provide significant additional evidence for this paradigm within the context of general relativity. (I plan to address the generalization to Lanczos–Lovelock models of gravity in a future work). In addition to new results, I will also paraphrase some known features of gravity in the language of the emergent paradigm because it provides better insights into these results. In fact, one of the significant fall-outs of the emergent gravity program was that it highlighted several peculiar features of general relativity which had previously not attracted the attention they deserved [17]. Some of the results in this paper continue this tradition.

Since this is a somewhat long paper touching upon diverse aspects of gravitational dynamics, I will first provide a summary of the key results of the paper.

I begin in the next section by essentially “running a commercial” for describing gravity using $f^{ab} \equiv \sqrt{-g}g^{ab}$ and the corresponding canonical momenta N_{ab}^c . These variables are, of course, not new (and, in fact, I will christen them as “old variables” [18–20]) and have appeared in literature from time to time (like e.g., in [21,22]). Recently, we provided a fairly comprehensive discussion of these variables [23] and demonstrated how the description of gravity takes a simple and elegant form in terms of these variables. In Sect. 2, I will provide an ab-initio route to gravitational dynamics based on these variables. In addition to setting the stage, this discussion serves the following important purpose: It introduces the notion of the momentum density of the gravitational field and relates its variation to the Ricci tensor and—through it—to the dynamics. The key equation in this context is ¹

¹ I use the $(-+++)$ signature and (unless otherwise specified) units with $c = 1$, $\hbar = 1$, $k_B = 1$, $G = 1$, so that Einstein’s equations reduce to $G_{ab} = 8\pi T_{ab}$. In some contexts, especially while dealing with action functionals, I use the convention $16\pi G = 1$ (so that Einstein’s equations reduce to $G_{ab} = (1/2)T_{ab}$)

$$\mathbb{L}_q R_{ab} = -\nabla_c (\mathbb{L}_q N_{ab}^c) = 8\pi \mathbb{L}_q \mathcal{F}_{ab}; \quad \mathcal{F}_{ab} \equiv T_{ab} - \frac{1}{2} g_{ab} T \quad (1)$$

This exact equation relates the variation of the gravitational momentum density $\mathbb{L}_q N_{ab}^c$ (in the form of a Lie derivative along an arbitrary vector field q^a) to the corresponding variation in the Ricci tensor and—through gravitational dynamics—with the variation of the energy momentum tensor. As we shall see, many other results in the paper arise from the fact that the above equation allows us to relate $f^{ab} \delta R_{ab}$ to $f^{ab} \delta N_{ab}^l$. The former quantity occurs in the variation of the surface term of the gravitational action and thus plays a crucial role in: (i) the symplectic structure of gravity, (ii) the form of the Noether currents and (iii) the physics of the boundary surfaces, while the latter expression will occur ubiquitously throughout this paper.

I also review briefly the connection between the variations of the dynamical variables ($f^{ab} \delta N_{ab}^c$, $N_{ab}^c \delta f^{ab}$) and the variations of the thermodynamic variables ($S \delta T$, $T \delta S$). Previous work [23] has established that there is a one-to-one correspondence between these two, when evaluated on the null surfaces for a specific class of metric variations which preserve the null surface geometry. In the rest of the paper, we repeatedly encounter these variations—in particular, $f^{ab} \delta N_{ab}^c$ —and it is important to keep in mind the thermodynamic correspondence between this variation and $S \delta T$.

The fact that two such variations— $S \delta T$ and $T \delta S$ —exist and can be related to each other is not often emphasized and it is worthwhile to discuss this briefly in a familiar context. Consider the standard black hole thermodynamics in which the addition of a mass δM changes the energy of the black hole by the standard relation $\delta M = \delta E = T \delta S$. While this relation is often taken for granted as “natural” without a second thought, it does contain a very peculiar feature: The absence of a $S \delta T$ term makes it appear as though T is kept constant while the process takes place which, of course, is not true since the change in M changes both T and S of the black hole. In fact, the Schwarzschild black hole (with horizon area A) satisfies the relation (with $L_P^2 = (G\hbar/c^3)$ being the Planck area):

$$E = 2TS = \frac{1}{2} T \left(\frac{A}{L_P^2} \right) \quad (2)$$

with a crucial factor of 2 [and $(1/2)$] in the two equalities above. The second relation tells us that, if we attribute $N_{\text{sur}} = A/L_P^2$ degrees of freedom to the horizon area A , then each degree of freedom carries exactly $(1/2)k_B T$ amount of energy. This, in turn, tells us that if we attribute $N_{\text{bulk}} \equiv [E/(1/2)T]$ degrees of freedom with the bulk gravitational energy inside the horizon (“equipartition”), then the black hole is in holographic equipartition with $N_{\text{sur}} = N_{\text{bulk}}$. From Eq. (2), it follows that

Footnote 1 continued

after explicitly alerting the reader. The Latin letters run through 0–3 while the Greek letters run through 1–3. I define (...) and [...] for (anti)symmetrization of tensor indices *without* a factor $(1/2)$. I also use the convention that δ (something) = $-\mathbb{L}$ (something) with a *relative minus sign* when the variation of (something) is produced by a diffeomorphism represented by Lie differentiation along a vector field.

$$\delta E = 2S\delta T + 2T\delta S \quad (3)$$

with both the variations contributing to the change in energy. However, there is an additional relation $S \propto M^2 \propto T^{-2}$ which is maintained during the variation that comes to our rescue, allowing us to express δE either in terms of δS or in terms of δT alone:

$$\delta E = T\delta S = -2S\delta T = -\frac{1}{2}\frac{A}{L_P^2}\delta T = -\frac{1}{2}N_{\text{sur}}\delta T \quad (4)$$

While the first equality is so widely discussed, the (equally valid) remaining relations are seldom emphasized in the literature! Equation (4) tells us that addition of energy can also be thought of as resulting in an increase in the temperature while the number of surface degrees of freedom is held fixed, with the crucial minus sign indicating the negative specific heat of the gravitating system. While, in reality, both S and T change in the physical processes involving the horizon, they could be described in a complementary fashion, as either $T\delta S$ or as $-(1/2)(A/L_P^2)\delta T$. (The latter is, in fact, similar to the description of, say, heating a mono-atomic ideal gas with $\delta E = (3N/2)\delta T$; we do not change N while heating a fixed amount of gas). In the general context of null surfaces and other boundaries, we will find that it is the latter interpretation involving δT (which corresponds to δN_{bc}^a) that provides a more natural description. As a result, we will continuously encounter this variation in different guises while studying the boundary thermodynamics. Roughly speaking, f^{ab} acts like an extensive variable in thermodynamics, while N_{ab}^c acts like an intensive variable.

A closely related quantity is the total variation $\delta(f^{ab}N_{ab}^c)$ which corresponds to $\delta(TS)$. This expression arises when we study the contribution of the boundary term in the Einstein–Hilbert action on a local Rindler horizon, and it occurs without the factor 2 noted earlier. I have discussed the physical significance of this term (and the difference between TS and $2TS$) in my previous papers [24, 25]. The TS essentially corresponds to the enthalpy (or heat content)² of the gravitational field which measures the difference between the energy and free energy of a finite temperature system, while $2TS$ measures the equipartition energy obtained by attributing $(1/2)k_B T$ to each of the (A/L_P^2) degrees of freedom in an area A . The basic result in the context of black holes, written in two forms:

$$2(TS) = 2(M/2) = \frac{1}{2}T(A/L_P^2); \quad 2(\text{heat content}) = (\text{equipartition energy}) \quad (5)$$

will keep appearing in our discussions.

In Sects. 3–5, I will discuss the relationship between the above concepts, Noether currents and the gravitational dynamics. To do this consistently, I will first show (Sect. 3) how Noether currents can be thought of as arising purely from some, actually rather trivial, mathematical identities in differential geometry. Introducing them in

² In usual thermodynamics, enthalpy H is defined as $H = E + PV$ and satisfies the identity $E + PV - TS = \mu N$ where μ is the chemical potential. So for systems with $\mu = 0$, we have the result $H = TS$, which is what we can call the heat content of the system. In this paper, I will use the term enthalpy to refer to just TS .

such a manner allows certain flexibility and brings sharply into focus their connection with dynamics which I introduce later on. (On the other hand, if I obtain Noether currents in the more conventional manner, from the diffeomorphism invariance of the action, then it is hardly surprising that they possess dynamical content since the action functional already knows about the dynamics of the system). While one can associate a conserved current with *any* vector field in spacetime, the vector fields related to the time evolution are special. If we foliate the spacetime in the usual manner with u_a denoting the unit normal to the $t = \text{constant}$ surfaces, then the two vector fields which are closely related to time evolution are the following:

$$\xi^a \equiv Nu^a; \quad \zeta^a \equiv Nu^a + N^a \quad (6)$$

where N and $N^a \equiv h_b^a \zeta^b$ are the lapse and shift functions and $h_b^a = \delta_b^a + u^a u_b$ is the projection tensor. The first one (ξ^a) is related to the flow of proper time normal to the $t = \text{constant}$ surfaces and is parallel to the velocity vector u^a of the fundamental observers. The second one (ζ^a) also satisfies the condition $\zeta^a \nabla_a t = 1$ and takes into account the shift of spatial coordinates between two spacelike hypersurfaces. In the past literature, ζ^a has been used and investigated in many different contexts quite widely. I will, however, show that one can obtain quite elegant and physically pleasing results from the Noether current and charge corresponding to the vector ξ^a . To begin with, I will show (Sect. 4.1) that the total Noether charge (associated with ξ^a) in any bulk region \mathcal{V} bounded by a constant lapse surface $\partial\mathcal{V}$, is equal to the heat content of the boundary surface. Further, twice the Noether charge gives the equipartition energy of the surface:

$$\begin{aligned} \int_{\mathcal{V}} \sqrt{h} d^3x u_a J^a[\xi] &= \int_{\partial\mathcal{V}} d^2x T_S; \\ 2 \int_{\mathcal{V}} \sqrt{h} d^3x u_a J^a[\xi] &= \int_{\partial\mathcal{V}} \frac{\sqrt{\sigma} d^2x}{L_P^2} \frac{1}{2} (k_B T); \end{aligned} \quad (7)$$

The temperature and entropy density of a patch of area are defined using local Rindler observers (described later around Eq. (55)). This result holds in arbitrary, time evolving spacetimes and provides a simple and direct interpretation of the Noether charge associated with one of the time development vectors. (For the sake of completeness, I also discuss briefly in Sect. 7.3 the corresponding results for ζ^a . Since there is a relative boost between the observers moving along ζ^a and the fundamental observers moving along ξ^a , there are additional contributions related to the boost energy which mar the simple thermodynamic interpretation in the case of ζ^a).

The above relation also allows one to study the concept of holographic equipartition [26–28] in static spacetimes and—more importantly—relate the time evolution of the spacetime geometry to the departure from holographic equipartition. I first re-derive (Sect. 4.2) the holographic equipartition law in the form $N_{\text{sur}} = N_{\text{bulk}}$ for any static spacetime with natural definitions for the surface and bulk degrees of freedom. Further, I show that the time evolution of the spacetime geometry can be described in a rather elegant manner by the equation:

$$\int_{\mathcal{V}} \frac{d^3x}{8\pi} \sqrt{h} u_a g^{ij} \xi_{\xi} N_{ij}^a = - \int_{\mathcal{V}} \frac{d^3x}{8\pi} h_{ab} \xi_{\xi} p^{ab} = \frac{1}{2} k_B T_{\text{avg}} (N_{\text{sur}} - N_{\text{bulk}}) \quad (8)$$

where h_{ab} is the induced metric on the $t = \text{constant}$ surfaces, p^{ab} is its conjugate momentum and T_{avg} is the average Davies–Unruh temperature of the boundary of \mathcal{V} . On the left hand side, we again see the rate of change of gravitational momentum which, in thermodynamic language, is related to the change in the temperature. The above equation shows that this change is driven by the departure from holographic equipartition, and that time evolution will cease when $N_{\text{sur}} = N_{\text{bulk}}$. I stress that all these concepts depend on the choice of observers and their accelerations, etc. arising through the spacetime foliation. While the equations are generally covariant, they encode observer dependence through the choice of the foliation and non-geometrical fields like the velocity of the observers. *I emphasize that this is a feature and not a bug.* As I mentioned earlier, attribution of thermodynamic variables to a null surface *is* observer dependent. Therefore, if we relate the dynamical evolution to the thermodynamical concepts, different observers using different coordinate systems *must* perceive, for example, the time evolution of the metric, differently. We know that the spacetime metric describing a black hole is time independent in the Schwarzschild coordinates but not in, say, the synchronous coordinates; what we see here is fundamentally no different from this fact.

A closely related result pertains to the definition and evaluation of the energy associated with gravity in a bulk region of space. Defining energy for gravity is a notorious activity that evokes strong reactions from the cognoscenti. I will, nevertheless, show that the description of gravity in terms of (f^{ab}, N_{ab}^c) provides a natural way of associating a gravitational four-momentum current with every vector field. This current is, in fact, what is usually called the Noether current for the vector field in the literature except that we will now: (i) choose it for the specific time development vector ξ^a and (ii) motivate it somewhat more physically. More importantly, we will show that the total amount of gravitational energy contained in a bulk region is then exactly equal to the surface heat content when the equations of motion hold. That is, we will show that for any self-gravitating system, which could be time-dependent and dynamically evolving, the total energy contained in a region \mathcal{R} bounded by a constant lapse surface $\partial\mathcal{R}$, is given by

$$\int_{\mathcal{R}} d^3x \sqrt{h} u_a [P^a(\xi) + N T_b^a u^b] = \int_{\partial\mathcal{R}} d^2x T s \quad (9)$$

On the left hand side, the first term is the suitably defined gravitational contribution to the energy density and the second term is the contribution from matter. While it is well known in the literature that the quasi-local energies (for a review, see e.g., [29]) for gravity are usually given by boundary integrals, the precise thermodynamic meaning we obtain is important in revealing the holographic nature of gravity.

I stress that the validity of the above result is *not* restricted to static spacetimes, which—in turn—implies that this energy and the boundary heat content will evolve in time due to physical processes. By adopting the usual techniques used to obtain

the symplectic form of gravity from Noether currents, one can easily determine the variation of this energy with respect to the chosen time development vector. I will show that this is given by

$$\mathfrak{L}_\xi H_{\text{grav}} = \int_{\partial\mathcal{R}} d^2x \sqrt{\sigma} N r_a \left(T^{ab} \xi_b + g^{lm} \mathfrak{L}_\xi N_{lm}^a \right) \quad (10)$$

where r_a is the normal to $\partial\mathcal{R}$. Of the two terms in the integrand, the first one is due to the matter energy flux across the boundary. The second term, representing the gravitational sector, again has the variation of N_{lm}^a , which reinforces the idea that this term can be thought of as the change in the gravitational heat due to the processes at the boundary. For example, in a matter-free spacetime (describing e.g., gravitational waves), we will have

$$\mathfrak{L}_\xi H_{\text{grav}} = \int_{\partial\mathcal{R}} d^2x \sqrt{\sigma} N r_a g^{lm} \mathfrak{L}_\xi N_{lm}^a \quad (11)$$

showing that the flow of energy of gravitational waves in a bulk region is described by the surface term. The physical meaning is again thermodynamic and is related to the change in the heat content of the boundary.

Up to this point in the paper, I have been faithful to the party line and have only attempted to connect the conventional description of bulk gravitational dynamics with boundary thermodynamics. I, however, believe that the *so called cosmological constant problem is a clear indication that we are ignoring the most significant clue we have about the nature of gravitational dynamics*: Gravity does not couple to changes in the bulk energy density arising from the addition of a constant to the matter Lagrangian. I begin Sect. 6.1 by showing that one cannot solve the cosmological constant problem in any generally covariant theory in which the metric is varied as a dynamical variable, in an unrestricted manner in a local action principle. A possible alternative is to obtain the field equations of gravity as a consistency condition for the spacetime from a variational principle involving some other dynamical variable. This is indeed possible (not only for Einstein's gravity but even for all Lanczos–Lovelock theories [30,31]) by defining a thermodynamic functional on every null surface and demanding that this functional should be extremised simultaneously on all null surfaces. I connect up this formalism with the heat density of the null surfaces and relate it again to a Noether charge. In this case, the Noether current is evaluated for the null vector defining the congruence, which is analogous to the time development vector in the case of spacelike surfaces discussed earlier. It turns out that the field equations for gravity in the absence of matter can be obtained by extremising the following functional on all null surfaces simultaneously:

$$Q \equiv \frac{1}{16\pi} \int_{\lambda_1}^{\lambda_2} d\lambda \, d^2x \sqrt{\sigma} \left[g^{ij} \ell_a \mathfrak{L}_\ell N_{ij}^a \right] \quad (12)$$

Here, ℓ^a is the affinely parametrized null vector defining the congruence. The occurrence of $g^{ij}\mathbb{E}_\ell N_{ij}^a$, which—as we said before, is related to $S\delta T$ —suggests that this extremum principle can be given a thermodynamic interpretation. This strengthens the physical interpretation of the variational principle and relates it to the corresponding concepts on spacelike and timelike boundaries.

Right at the end (in Sect. 7) I discuss several other variations on the basic theme presented here. Among them, the possibility of using different Lagrangians, related to the Einstein–Hilbert Lagrangian by total divergences, and the possibility of using other time development vector fields (in particular ζ^a) are of some interest. I show how the Noether potential gets modified by the addition of total divergences to the Einstein–Hilbert Lagrangian. In particular, it turns out that the results related to a spacelike surface, obtained with the time development vector ξ^a , do *not* change when we use the Lagrangian with the addition of the $2K$ term at the boundary. (This does not happen for ζ^a , which is possibly yet another reason to use ξ^a in the formalism). Corrections do appear when one uses the Γ^2 Lagrangian which I will describe briefly. Finally, I describe the results pertaining to the use of ζ^a rather than ξ^a . As mentioned earlier, this introduces an additional boost relative to the fundamental observers which complicates the interpretation. To a limited extent, this can be understood in terms of a modified acceleration $\zeta^a\nabla_a u^b$ instead of $Nu^a\nabla_a u^b$, but the fact that ζ^a in general does not have good foliation properties—in contrast to u^a —makes the mathematics inelegant and the interpretations rather contrived. I believe it is the use of ζ^a rather than ξ^a in the past literature which has led to the missing of some of these results, which, as we will see, are mathematically rather simple.

2 Introducing the dramatis personae: a route to gravity

Recent research [23] motivated by the emergent paradigm of gravity shows that the description of classical gravity simplifies significantly if we use the variables $f^{ab} \equiv \sqrt{-g}g^{ab}$ and the corresponding canonical momentum N_{ab}^c . These ‘old variables’ were studied [18–20] in the early days of general relativity (and were used sporadically later on in the literature, like, e.g., in [21, 22]) but have not acquired the popularity they deserve. My first task is to advertise the virtues of this description.

Classical mechanics of a single degree of freedom $q(t)$ can be obtained from an action principle based on the Lagrangian $L_q \equiv p\dot{q} - H(p, q)$. Varying p and q independently, one sees that: (a) To get sensible equations of motion, we need $\delta q = 0$ at the boundary while δp is arbitrary. (b) When $\delta q = 0$ at the boundary, the equations of motion are $\partial_t q = (\partial H/\partial p)$ and $\partial_t p = -(\partial H/\partial q)$. The subscript q on L_q is to remind us that q is kept fixed at the end points to obtain the equations of motion. We can ask whether there is another Lagrangian which will lead to the *same equations of motion* when p is kept fixed. There indeed exists one, which is given by

$$L_p \equiv L_q - \frac{d(pq)}{dt} = -q\dot{p} - H(p, q) \quad (13)$$

It is easy to see that if we vary L_p , treating again q and p as independent, we can get the same equation of motion with $\delta p = 0$ at the boundary and δq arbitrary. Note that $L_q = L_q(q, \dot{q}, p)$ while $L_p = L_p(q, p, \dot{p})$.

We also know that we could have got the same equations of motion by treating L_q as just a function of q, \dot{q} and identifying $p = p(q, \dot{q})$ as $p = \partial L_q / \partial \dot{q}$ without treating it as an independent variable. In such an approach, there is a significant difference between L_p and L_q . When $L_q = L_q(q, \dot{q})$ depends only on up to first derivatives of q , the $L_p = L_p(q, \dot{q}, \ddot{q})$ will (in general) depend on second derivatives of q through the term $d[q(\partial L / \partial \dot{q})] / dt$. Normally, if a Lagrangian depends on the second derivatives of q , the equations of motion can be third order in q . But *with this specific kind of dependence*, the equations of motion will still be second order in q when we keep p fixed at the end points. Thus, it is easy³ to construct, even in classical mechanics, a Lagrangian containing \ddot{q} , which only leads to second order equations of motion.⁴

What is not stressed in textbooks is that we can do the same thing in field theory. The field equations, $\square \phi = -V'(\phi)$ for a scalar field ϕ , say, can be obtained using either of the two Lagrangians:

$$L_\phi = p^a \partial_a \phi - H; \quad L_p = -\phi \partial_a p^a - H = L_\phi - \partial_a (\phi p^a); \quad H = (1/2) p^a p_a + V(\phi). \quad (14)$$

These Lagrangians, again, differ by a total divergence. In contrast to the text book description, we are here treating H as a *Lorentz scalar* and not as the time component of a four-vector; further, full Lorentz invariance is maintained without any (1+3) split. Note that, while the $p^a \partial_a \phi$ term does not involve a metric, $p^a p_a = \eta_{ab} p^a p^b$ needs a background metric to ‘lower the index’ on p^a . We can also work out the dynamics of a vector field and, in particular, the $U(1)$ gauge field in an analogous manner. It is straightforward to generalize the description to a curved spacetime by the usual prescriptions.

The situation becomes really interesting when we consider the theory of a symmetric second rank tensor field described by a matrix f with elements denoted by f^{ab} . The “momenta” corresponding to f^{ab} will be denoted by N_{ab}^c which, in turn, can be thought of as elements of four matrices N^c . As in the previous cases, we can consider two possible Lagrangians (which differ from each other by a four-divergence) to describe the theory:

$$L_f = N_{ab}^c \partial_c f^{ab} - \mathcal{H}_g \left(f^{ij}, N_{lm}^k \right); \quad L_N = -f^{ab} \partial_c N_{ab}^c - \mathcal{H}_g = L_f - \partial_c \left(f^{ab} N_{ab}^c \right) \quad (15)$$

with a very specific choice:

$$\mathcal{H}_g = f^{ab} \left(N_{ad}^c N_{bc}^d - \frac{1}{3} N_{ac}^c N_{bd}^d \right) \quad (16)$$

³ In fact, given $L_q(q, \dot{q})$ one can construct a Lagrangian which leads to the same equations of motion when some arbitrary function $C(q, \dot{q})$ is kept fixed at the boundary; see [32] for details.

⁴ The Lagrangian $L_p = L_q - d(qp)/dt$ is almost never used in classical mechanics or field theory. For a free particle, $L_p(q, \dot{q}, \ddot{q}) = -(1/2)\dot{q}^2 - q\ddot{q}$; this should not be confused with the Lagrangian $-(1/2)q(d^2/dt^2)q$ which is sometimes used.

Note that, unlike the scalar and vector cases, we do *not* need to use $\sqrt{-g}d^4x$ as integration measure; the action is obtained in *any* coordinate system by integrating over just $d^4x = dx^0 dx^1 dx^2 dx^3$. This makes the action a polynomial function of the basic variables (f^{ab}, N_{jk}^i) , unlike what happens when we use g_{ab} due to the presence of the $\sqrt{-g}$ term. Further, these Lagrangians are defined without the use of any background metric (in contrast to the scalar/vector field theories) and the contractions in Eqs. (15) and (16) are purely combinatoric operations involving only δ_j^i . In fact, instead of thinking of f^{ab} as a field in spacetime etc., we can think of these Lagrangians as describing the (abstract) dynamics of five 4×4 matrices (f, N^c) , the elements of each of which depend on four parameters q^i . The matrix action is determined by integrating the L_f or L_N over d^4q . The variational principle using L_f (with f fixed at the boundary) or with L_N (with N^c fixed at the boundary) will lead to the dynamical equations for the matrices. (This fact allows a different way of approaching gravitational dynamics which I will comment on in Sect. 2.2).

Incredibly enough, the *resulting equations are identical to those of Einstein's gravity* (without sources; it can be easily extended to take care of matter sources [23]) if we identify the arbitrary curved spacetime coordinates x^i with the parameters q^i and set $f^{ab} \equiv \sqrt{-g}g^{ab}$. The equations then imply the further identification:

$$N_{bc}^a = -\Gamma_{bc}^a + \frac{1}{2} \left(\Gamma_{bd}^d \delta_c^a + \Gamma_{cd}^d \delta_b^a \right) = Q_{be}^{ad} \Gamma_{cd}^e + Q_{ce}^{ad} \Gamma_{bd}^e \quad (17)$$

where $Q_{cd}^{ab} = (1/2)[\delta_c^a \delta_d^b - \delta_d^a \delta_c^b]$ is the determinant tensor. (For a detailed proof, see [23]). Thus, gravitational dynamics in an *arbitrary* coordinate system with labels x^i and metric functions $g_{ab}(x^i)$ can be obtained from the dynamics of five matrices (f, N^c) .

Another surprise is that the momentum space Lagrangian L_N is numerically just $\sqrt{-g}R$ where R is the curvature scalar! That is, treating N_{jk}^i as a function of f^{ab} given by Eq. (17), one can show [23] that

$$L_N = L_f - \partial_c \left(f^{ab} N_{ab}^c \right) = -f^{ab} \partial_c N_{ab}^c - \mathcal{H}_g = \sqrt{-g}R \quad (18)$$

where R is the standard curvature scalar of differential geometry. The reason the Einstein–Hilbert action leads to second order equations is now clear: It is a momentum space Lagrangian in terms of the variables (f, N^a) .

As we shall see, the variables (f, N^a) will play an extremely important role in the rest of the paper. In view of this fact, I will next describe several important properties of these variables and connect them up with the standard description to enhance familiarity. (For more details see, [23]).

2.1 Description of gravity in terms of (f^{ab}, N_{ab}^c)

To begin with, the standard Einstein–Hilbert Lagrangian

$$(16\pi) A_{\text{EH}} = \int_{\mathcal{V}} d^4x \mathcal{L}_{\text{EH}} = \int_{\mathcal{V}} d^4x \sqrt{-g} R = \int_{\mathcal{V}} d^4x \sqrt{-g} Q_{cd}^{ab} R_{ab}^{cd} \quad (19)$$

can be decomposed into a quadratic Lagrangian (called the Γ^2 Lagrangian or the Einstein–Schrodinger Lagrangian) and a divergence term as: $\mathcal{L}_{\text{EH}} = \mathcal{L}_{\text{quad}} + \mathcal{L}_{\text{sur}}$, with

$$\mathcal{L}_{\text{quad}} \equiv 2\sqrt{-g} Q_a^{bcd} \Gamma_{dk}^a \Gamma_{bc}^k; \quad \mathcal{L}_{\text{sur}} \equiv 2\partial_c \left[\sqrt{-g} Q_a^{bcd} \Gamma_{bd}^a \right]. \quad (20)$$

These expressions take simpler forms when we use the variables f^{ab} and N_{bc}^a treating N_{bc}^a as a specified function of f^{ab} . We again have $\mathcal{L}_{\text{EH}} = \mathcal{L}_{\text{quad}} + \mathcal{L}_{\text{sur}}$ with

$$\mathcal{L}_{\text{quad}} = \frac{1}{2} N_{bc}^a \partial_a f^{bc}; \quad \mathcal{L}_{\text{sur}} = -\partial_c \left(f^{ab} N_{ab}^c \right) \quad (21)$$

The Ricci tensor can be expressed in terms of N_{bc}^a as:

$$R_{ab} = - \left(\partial_c N_{ab}^c + N_{ad}^c N_{bc}^d - \frac{1}{3} N_{ac}^c N_{bd}^d \right) \quad (22)$$

Though N_{bc}^a is not a tensor, its variation δN_{bc}^a is a tensor which can be related to the variation of the Ricci tensor, δR_{ab} , again, in a remarkably simple form:

$$\delta R_{ab} = -\nabla_c (\delta N_{ab}^c) \quad (23)$$

(Incidentally the result in Eq. (23) allows us to formulate the gravitational field equations in a novel manner which I have commented upon in Sect. 2.2). As a corollary, when the variations are induced by a Lie derivative along an arbitrary vector field q^a , Eq. (23) becomes

$$\mathcal{L}_q R_{ab} = -\nabla_c (\mathcal{L}_q N_{ab}^c) \quad (24)$$

The Lie derivative of non-tensorial objects like Γ_{bc}^a can be defined from first principles and the explicit form of $\mathcal{L}_q \Gamma_{bc}^a$ is given later in Eq. (47); the Lie derivative of N_{bc}^a follows on using Eq. (17). Since N_{bc}^a is the momentum density, Eq. (24) shows that the Ricci tensor is closely related to the variation of gravitational momentum when Lie transported along any vector field.

Multiplying Eq. (23) by $f^{ab} = \sqrt{-g} g^{ab}$ we immediately get:

$$f^{ab} \delta R_{ab} = \delta(\sqrt{-g} R) - R_{ab} \delta f^{ab} = -\partial_c \left[f^{ik} \delta N_{ik}^c \right], \quad (25)$$

which, in turn, gives the variation of the Einstein–Hilbert action in just one step:

$$\delta(\sqrt{-g} R) = R_{ab} \delta f^{ab} - \partial_c \left[f^{ik} \delta N_{ik}^c \right] = \sqrt{-g} \left[G_{ab} \delta g^{ab} - \nabla_c \left(g^{ik} \delta N_{ik}^c \right) \right] \quad (26)$$

To get the second equality, we have used the fact that for any two index object X_{ab} , we have $X_{ab} \delta f^{ab} = (X_{ab} - (1/2) g_{ab} X) \delta g^{ab}$. If the variations in Eq. (26) arise due to a Lie

differentiation along a vector field q , then on multiplying Eq. (24) by $f^{ab} = \sqrt{-g} g^{ab}$, identical algebra gives [on using $\mathcal{L}_q(\sqrt{-g}R) = \sqrt{-g} \nabla_a(Rq^a)$, $G_{ab}\mathcal{L}_q g^{ab} = -2\nabla_a(G^a_b q^b)$] the conservation law $\nabla_a J^a = 0$ for the current:

$$(16\pi)J^a[q] = 2G^a_b q^b + Rq^a + g^{ik}\mathcal{L}_q N^a_{ik} = 2R^a_b q^b + g^{ik}\mathcal{L}_q N^a_{ik} \quad (27)$$

(The factor 16π is introduced because we are using units with $G = 1$; this factor will be absent if we use a more natural choice of units with $16\pi G = 1$). This current will again play a vital role in our discussion throughout the paper.

Further, we can define a vector $P^a[q]$ associated with any vector field q^a by rewriting Eq. (27) as

$$(16\pi)P^a[q] \equiv (16\pi)J^a[q] - 2G^a_b q^b = g^{ik}\mathcal{L}_q N^a_{ik} + L_H q^a \quad (28)$$

where $L_H = R$ is the Einstein–Hilbert Lagrangian. This vector is same as J^a on-shell, when the equation of motion $G^a_b = 0$ is satisfied and is conserved on-shell. To see the additional significance of this vector, note that the Hamiltonian for gravity we started with in Eq. (15) has the form:

$$\mathcal{H}_g = -\sqrt{-g} [g^{ab}\partial_c N^c_{ab} + L_H] \quad (29)$$

where we have used $L_N = R\sqrt{-g} = L_H\sqrt{-g}$. This, of course, is not a scalar density because of the ∂_c in the expression. But the form of \mathcal{H}_g suggests the replacement of $g^{ab}\partial_c N^c_{ab}$ by $g^{ab}\mathcal{L}_q N^c_{ab}$ along any vector field to obtain a scalar density for this term. To make the indices match, we replace L_H by $L_H q^a$. This is precisely the vector $P^a[q]$, with the sign chosen for future convenience.⁵

In particular, if we chose $q^a = \xi^a \equiv Nu^a$ where u_a is the normal to $t = \text{constant}$ surfaces, we can interpret $P^a[\xi]$ as the gravitational four-momentum flux. When the equations of motion hold $2G^a_b \xi^b = (16\pi)T^a_b \xi^b$ can be thought of as the corresponding four-momentum flux of matter. The definition in Eq. (28) tells us that:

$$(16\pi)J^a[\xi] = (16\pi)P^a[\xi] + 2G^a_b \xi^b = (16\pi)(P^a[\xi] + T^a_b \xi^b) \equiv (16\pi)P^a_{total}[\xi] \quad (30)$$

allowing us to interpret the conserved current $J^a[\xi]$ as the *total* four-momentum flux. This interpretation will be made more precise and strengthened in Sect. 5.

These two results in Eqs. (26) and (27) and the interpretation of Eq. (28) highlight the importance of the combination $g^{ik}\mathcal{L}_q N^a_{ik}$ which, as we saw, can be related to gravitational momentum flow. We also know from Eq. (26) that $g^{ik}\delta N^a_{ik}$ is essentially

⁵ With our choice of signature, in flat spacetime, say, $u_a = -\delta^0_a$ so that $u^a = \delta^a_0$ is future pointing. This means that in the dot products $u_a v^a = -v^0$, we pick up the time component of v^a with an extra minus sign rather than just pick up v^0 (which would have happened if we were using the opposite signature). This is not of concern most of the time and we will not worry about it. Occasionally, our sign choices are decided by this fact.

the surface term in the variation of the gravitational action:

$$(16\pi)\delta A_{\text{sur}} = - \int_{\mathcal{V}} d^4x \partial_c [f^{ab} \delta N_{ab}^c] = - \int_{\partial\mathcal{V}} d^3x \sqrt{h} n_c g^{ab} \delta N_{ab}^c \quad (31)$$

It turns out that this variation can be given a simple thermodynamic interpretation. Consider a null surface with temperature $T = \kappa/2\pi$ and entropy density $s = \sqrt{\sigma}/4$ attributed to it by local Rindler observers who perceive it as a horizon with κ being the surface gravity defined using a suitable null congruence and $\sqrt{\sigma}$ being the area element of the local horizon surface: Then, (see [23] for details) the following results hold:

- The boundary term in the action, evaluated over a null surface, can be interpreted [25] in terms of its heat content Ts (or enthalpy density; see the footnote on page 4); that is:

$$\frac{1}{16\pi} \int d^3\Sigma_c (N_{ab}^c f^{ab}) = \int d\lambda d^2x \left(\frac{\kappa}{2\pi} \right) \left(\frac{\sqrt{\sigma}}{4} \right) = \int d\lambda d^2x T s \quad (32)$$

- More remarkably, the variations $f\delta N$ and $N\delta f$ have corresponding thermodynamic interpretations for a class of variations which preserve the null surface:

$$\frac{1}{16\pi} \int d^3\Sigma_c (N_{ab}^c \delta f^{ab}) = \int d\lambda d^2x \left(\frac{\kappa}{2\pi} \right) \delta \left(\frac{\sqrt{\sigma}}{4} \right) = \int d\lambda d^2x T \delta s; \quad (33)$$

$$\frac{1}{16\pi} \int d^3\Sigma_c (f^{ab} \delta N_{ab}^c) = \int d\lambda d^2x \left(\frac{\sqrt{\sigma}}{4} \right) \delta \left(\frac{\kappa}{2\pi} \right) = \int d\lambda d^2x s \delta T \quad (34)$$

We see that (f, N^c) are not only *dynamically conjugate* variables but their variations $(\delta f, \delta N^c)$ exhibit *thermodynamic conjugacy* in terms of corresponding variations in $(\delta T, \delta S)$. Of these, we can think of f as an extensive variable and N^a as intensive variables, just as in conventional thermodynamics. Later discussion will fortify this idea.

If we consider the gravitational action principle in a region of spacetime bound by null surfaces (the ‘causal diamond’), then the boundary condition on the surface corresponds to isothermality ($\delta T = 0$) with respect to the local Rindler observers who perceive the null surfaces as local Rindler horizons. It is obvious how thermodynamic considerations enhance our understanding of standard gravitational dynamics. We will see many more such examples in the sequel.

2.2 Aside: some additional comments

There are a few comments regarding this formalism which I would like to mention here as an aside, though it is not relevant directly to the rest of the paper. Those interested in the main theme can directly proceed to Sect. 3.

First, in the standard approach, we use g_{ab} rather than $f^{ab} \equiv \sqrt{-g} g^{ab}$ as the dynamical variables and one may wonder whether similar results can be obtained with g_{ab} . This requires using the action:

$$(16\pi)A = \int d^4x \sqrt{-g} L; \quad L = M^{cab} \partial_c g_{ab} - H \quad (35)$$

where

$$M^{abc} = g^{bd} g^{ce} \Gamma_{de}^a - \frac{1}{2} g^{bd} g^{ac} \Gamma_{de}^e - \frac{1}{2} g^{cd} g^{ab} \Gamma_{de}^e - \frac{1}{2} g^{bc} V^a. \quad (36)$$

with

$$V^c \equiv -g_{ab} M^{cab} = g^{ik} \Gamma_{ik}^c - g^{ck} \Gamma_{km}^m = 2Q_a{}^{bcd} \Gamma_{bd}^a = 2Q^{ijkc} \partial_i g_{jk} = -\frac{1}{g} \partial_b (g g^{bc}). \quad (37)$$

and

$$H = g^{bd} M_{dk}^c M_{bc}^k - \frac{2}{3} g^{ir} M_{ir}^k M_{dk}^d - \frac{1}{3} g^{bk} M_{bi}^i M_{dk}^d + \frac{1}{6} g_{dk} g^{ir} M_{ir}^d g^{xy} M_{xy}^k \quad (38)$$

where $M_{ab}^c \equiv g_{ai} g_{bj} M^{cij}$ etc. We see that: (i) We now have to define the action with a $\sqrt{-g} d^4x$ measure rather than with just d^4x introducing a factor which is *non-polynomial* in the dynamical variables unlike in the case of (f^{ab}, N_{bc}^a) . (ii) The expressions are a lot more complicated compared to when we use (f^{ab}, N_{bc}^a) as our variables. We will see on several future occasions that the use of the (f^{ab}, N_{bc}^a) pair simplifies the expressions.

Second, the matrix dynamics in terms of the five 4×4 matrices (f, N^a) achieves “general covariance without general covariance” in the following sense. Consider a 4×4 matrix with elements denoted by f^{ab} and another set of four 4×4 matrices $[N_{bc}^0, N_{bc}^1, N_{bc}^2, N_{bc}^3]$ with elements denoted by N_{bc}^a . We assume that all these elements $[f^{ab}(q^i), N_{bc}^a(q^i)]$ depend on four parameters q^i . We can now construct the index-free function $L_f(q^i)$ using Eqs. (15) and (16). This does not require any metric and the index placements take care of summation with just Kronecker deltas. (However, the definition of \mathcal{H}_g is not a pure matrix operation but involves the components). We can integrate $L_f(q^i)$ over a range of parameters q^i with the Cartesian measure d^4q to define an abstract matrix action. The extremisation of this action when the matrices (f, N^a) are varied, will lead to a set of equations which determine the relation between $\partial_c f$ and N^a as well as their dependence on q^i . Now consider the real spacetime manifold described by the metric functions $g_{ab}(x^i)$ in some specific coordinate system. If we identify the coordinates x^i with our parameters q^i and the combination $\sqrt{-g} g^{ab}$ with f^{ab} , then the matrix equations we obtained determine the spacetime metric functions in these coordinates. If we now choose another set of coordinates in the spacetime, the metric functions will change. *But we do not have to change anything as regards the abstract matrix description; in particular we do not have to multiply d^4q by a Jacobian.* We simply identify the new coordinates with q^i and the new $\sqrt{-g} g^{ab}$ with f^{ab} and everything will work out quite trivially. This separation of kinematics (coordinates,

explicit form of metric functions, their change under coordinate transformations, etc.) from the dynamics (described once and for all by the equations satisfied by the matrices (f, N^a)) is an attractive conceptual feature of this approach.

Third, one can easily introduce more ‘pre-geometric’ variables λ_A , say, and treat the matrix element $f^{ab}(q^i)$ as some kind of coarse-grained object after averaging over λ_A , like $f^{ab}(q^i) = \langle f^{ab}(q^i; \lambda_A) \rangle$ where the averaging is over the pre-geometric variables λ_A with some measure. This shows that the current approach may be useful for introducing and investigating pre-geometric structures.

Fourth, the nature of gravity is completely contained in the form of the Hamiltonian \mathcal{H}_g . On general grounds, one would expect it to be a quadratic in N_{jk}^i . The choice of the specific form of the quadratic function depends on the symmetries we want to impose on the theory. There is a natural, covariant, separation of any metric into a conformal form $g_{ab} = \Omega^2 \bar{g}_{ab}$ with \bar{g}_{ab} being a metric of unit determinant. This allows reformulation of the theory in terms of the variables (Ω, \bar{f}^{ab}) , which has some interesting features that I hope to discuss in a future publication.

Fifth, the relation Eq. (23) allows a ‘variational’ formulation of gravitational dynamics. Suppose we have a source $\mathcal{F}_{ab}(s) \equiv (T_{ab} - (1/2)g_{ab}T)$ where $T_{ab}(s)$ is the stress-tensor which depends on some parameter s (mass, charge, cosmological constant, anything....). For every value of s , we can, in principle, solve the field equations and obtain $g_{ab}(s)$, $f_{ab}(s)$, $N_{ab}^c(s)$, etc. Let an overdot denote the derivative of any of these with respect to s , like $\dot{\mathcal{F}}_{ab} = d\mathcal{F}_{ab}/ds$, etc. Then the gravitational field equations can be written simply as:

$$\nabla_c(\dot{N}_{ab}^c) = -8\pi \dot{\mathcal{F}}_{ab} \quad (39)$$

which tells us how the geometry *changes* when we *change* the source. Given $\dot{\mathcal{F}}_{ab}$, we can solve Eq. (39) for $\dot{N}_{bc}^a(s)$, and find $N_{bc}^a(s)$ by integrating $\dot{N}_{bc}^a(s)$ over s with flat spacetime as the initial condition, say. We again see, from Eq. (39), the importance of N_{bc}^a ; the field equations directly determine how it changes when the source changes. Such a description relating *changes* in the geometry to *changes* in the source may be closer in spirit to thermodynamics. These aspects open up interesting avenues for further work.

3 Conserved currents from identities in differential geometry

In the later sections, I will extensively use certain conserved currents (usually called Noether currents) to describe the gravitational dynamics. Conventionally, there is a tendency in literature to link the existence of such conserved currents with the diffeomorphism invariance of an action. If one follows the route of deriving the conserved currents from the invariance properties of a given action functional, then it is somewhat of a circuitous reasoning—though interesting—to redescribe dynamics in terms of these currents. (After all, complete dynamics can be obtained if the form of the action is known.) Fortunately, this approach is unnecessary and—to a great

extent—complicates matters. I will show how the variables (f, N^a) are closely related to conserved (Noether) currents for vector fields purely because of some—in fact, rather trivial—identities in differential geometry. *This delinks the existence of conserved (Noether) currents from the invariance properties of any action principle or such dynamical considerations.*

There are (at least) two ways of obtaining the standard Noether current associated with a vector field v^a *without ever mentioning the action principle for gravity or its diffeomorphism invariance!* These involve just (i) using either the antisymmetric part of $\nabla_i v_j$ and an identity for $R_j^i v^j$ or (ii) contracting the Lie derivative of Γ suitably. This shows that the resulting conservation law $\nabla_a J^a = 0$, for example, is just a differential geometric identity and *has nothing to do with the gravitational dynamics.* Delinking the *form* of the Noether current from any action principle allows us to reinterpret gravitational dynamics later on using Noether currents, without introducing any circuitous reasoning.

We begin from the elementary fact that the derivative $\nabla_k v_j$ of any vector field v^j can be decomposed into the anti-symmetric and symmetric parts by

$$\nabla^j v^k + \nabla^k v^j \equiv \nabla^{(j} v^{k)} \equiv S^{jk}; \quad \nabla^j v^k - \nabla^k v^j \equiv \nabla^{[j} v^{k]} \equiv J^{jk} \quad (40)$$

[Recall that I define (...) and [...] *without* a factor $(1/2)$]. The antisymmetric part J^{lm} leads to a conserved current $J^i \equiv \nabla_k J^{ik}$; in other words, from every vector field v^k in the spacetime we can obtain a conserved current, fairly trivially.⁶

To find an explicit form for this current and explore its connection with N_{bc}^a , we proceed as follows: Substituting the decomposition, $\nabla^j v^k = (1/2)(J^{jk} + S^{jk})$, in the standard identity

$$\nabla_k (\nabla^j v^k) - \nabla^j (\nabla_k v^k) = R_l^j v^l \quad (41)$$

we get:

$$\nabla_k J^{jk} + \nabla_k (S^{jk} - g^{jk} S) = 2R_l^j v^l \quad (42)$$

The first term in the above equation is the conserved current J^k , while the physical meaning of the second term can be obtained by noting that the symmetric part S^{jk} gives the change in the metric tensor g^{jk} under the diffeomorphism $x^a \rightarrow x^a + v^a$. That is:

$$\mathcal{L}_v g^{ab} = -[\nabla^a v^b + \nabla^b v^a] = -S^{ab} \equiv -\delta g^{ab} \quad (43)$$

(Recall our convention that $\delta = -\mathcal{L}$). From the definition in Eq. (17) we can also obtain the explicit result:

$$g^{ab} \delta N_{ab}^c = \nabla_b (\delta g^{cb} - g^{cb} g_{ik} \delta g^{ik}) \quad (44)$$

⁶ In fact, the converse is also true; given a conserved current J^i one can always find the corresponding vector field v^k . Since J^{il} is invariant under the “gauge transformation” $v^k \rightarrow v^k + \partial^k f$, we in fact have an infinite number of such vectors associated with a given conserved current. This is just electromagnetism in disguise.

which is valid for arbitrary variations, not just diffeomorphism (see, e.g., p. 549 of Ref. [44]). In the case of diffeomorphisms (with $\delta N_{ab}^c = -\mathfrak{L}_v N_{ab}^c$, $\delta g^{ab} = S^{ab}$), we get:

$$g^{ab} \mathfrak{L}_v N_{ab}^c \equiv B^c[v] = -\nabla_b (S^{bc} - g^{cb} S); \quad \nabla_c (g^{ab} \mathfrak{L}_v N_{ab}^c) \equiv \nabla_c B^c \quad (45)$$

Comparing this with the form of the second term on the left hand side of Eq. (42), we see that the second term is just $(-g^{bc} \mathfrak{L}_v N_{bc}^j)$. This gives us the explicit form of the conserved current

$$J^a[v] = \nabla_b J^{ab}[v] = 2R_b^a v^b + g^{ij} \mathfrak{L}_v N_{ij}^a = 2R_b^a v^b + B^a[v] \quad (46)$$

which, of course, is the same one as obtained before in Eq. (27) (except for an irrelevant factor (16π) which is related to the choice of units), but now we see that its conservation follows trivially from the antisymmetry of $\nabla^{[j} v^{k]}$ which was far from obvious in the derivation leading to Eq. (27).

An alternate, faster, way of obtaining the same result is as follows: From the Lie derivative of the connection

$$\mathfrak{L}_v \Gamma_{bc}^a = \nabla_b \nabla_c v^a + R_{cmb}^a v^m \quad (47)$$

one can obtain, on using Eq. (17), the relation:

$$2Q^{adc} \mathfrak{L}_v \Gamma_{cd}^e = g^{bc} \mathfrak{L}_v N_{bc}^a = \nabla_b J^{ab} - 2R_b^a v^b \quad (48)$$

which is the same as the one obtained earlier.

We thus see that there are two routes to obtaining the form of the Noether current: (a) from the antisymmetric part of $\nabla_i v_j$ using the identity in Eq. (41) or (b) by contracting the Lie derivative of Γ suitably. In either procedure, the natural object we obtain is J^{ab} (called the Noether potential in general or the Komar superpotential in this particular case, which are just other names for $\nabla^{[j} v^{k]}$!) and J^a is *derived* from it. *This route is unique and natural.* In contrast, in the conventional approach one *first* obtains the Noether current J^a and *then* looks for a Noether potential J^{ab} thereby agonizing over its non-uniqueness.

Thus, with every vector field v^j we can associate a conserved current $J^i[v]$ given by Eq. (46). One could think of this as a trivial algebraic result arising from the antisymmetry of J^{ij} related to the vector field. On the other hand, it also *happens to be* the Noether current arising from the diffeomorphism invariance of the scalar R . Nobody knows of a clear *physical* reason as to why this happens; in other words, nobody knows how to guess this answer without doing the calculation. This fact, probably, has some physical significance and—in the discussions later—I will try to emphasize this to a certain extent.

4 Bulk gravitational dynamics and surface thermodynamics

For the rest of the paper, we will consider a spacetime foliated by a series of space-like hypersurfaces defined by constant values for a suitable scalar field $t(x)$. We define

the unit normal to $t = \text{constant}$ surfaces by $u_a = -N\nabla_a t$; if we choose our time coordinate to coincide with the hypersurface label, then $u_a = -N\delta_a^0$. We next define a ‘time-development’ vector ζ^a by the invariant condition $\zeta^a \nabla_a t = 1$. (In the preferred coordinate system, we can choose $\zeta^a = \delta_a^0$). In general, ζ^a and u^a will not be in the same direction and we have $\zeta^a = -(\zeta^b u_b)u^a + N^a$ where $u_a N^a = 0$ and $N^a = h_b^a \zeta^b$ where $h_b^a \equiv \delta_b^a + u^a u_b$ is the projection tensor. In component form, $\zeta^a = Nu^a + N^a$ with $N^a = (0, N^\alpha)$.

This decomposition also introduces the vector $\xi^a \equiv -(\zeta^b u_b)u^a = Nu^a$ which will turn out to be of considerable importance in what follows. When we impose the coordinate condition such that $g_{0\alpha} = 0$ in a local region, we will have $\zeta^a = \xi^a$ and, if the spacetime is static, we can identify ξ^a with the time-like Killing vector. Thus, in *any* spacetime, there exist two *natural* diffeomorphisms, with vector fields

$$\xi^a = Nu^a = -N^2 \nabla_a t; \quad \zeta^a = Nu^a + N^a \quad (49)$$

The latter one ζ^a has attracted a fair amount of attention in the past literature. I will, however, show that as far as Noether currents and related dynamics go, the vector ξ^a possesses more interesting properties. So, I will begin with a discussion of these properties and later comment on corresponding results vis-a-vis the vector field ζ^a in Sect. 7.3. (Incidentally, the Noether current for u^a is also *not* interesting compared to the one for ξ^a ; see the discussion in Appendix 9.2).

4.1 Noether charge as the surface heat content

We will begin by computing the Noether current for ξ^a . To calculate this current, it is convenient to use an identity connecting Noether currents for two vector fields q^a and $v^a \equiv f(x)q^a$. It can be shown that [see Appendix 9.1; Eq. (136)]

$$q_a J^a[fq] - f q_a J^a[q] = \nabla_b \left[(q^a q^b - q^2 g^{ab}) \nabla_a f \right] \quad (50)$$

which is particularly useful if $q_a = \nabla_a \phi$ so that $J^{ab}[q] = 0$. If we use Eq. (50) with $q_a = -\xi_a/N^2$, $f = -N^2$, we can obtain a nice result for the Noether charge density [see Appendix 9.2; Eq. (145)]:

$$(16\pi)u_a J^a(\xi) = 2D_\alpha(Na^\alpha) \quad (51)$$

where $a^i \equiv u^j \nabla_j u^i$ is the acceleration and $D_i a^i = D_\alpha a^\alpha = \nabla_i a^i - a^2$ where D_i is the covariant derivative on the $t = \text{constant}$ surface. The acceleration a_i , which will occur repeatedly in our discussion, has the explicit form [see Appendix 9.2; Eq. (138)]

$$Na_i = Nu^l \nabla_l u_i = h_i^j \nabla_j N. \quad (52)$$

The combination $Na_i = \xi^l \nabla_l u_i$ measures the change in velocity along our chosen time-development vector ξ^l .

Integrating Eq. (51) over $\sqrt{h}d^3x$ to obtain the total Noether charge, we find that the flux of the acceleration is essentially the total Noether charge contained inside a volume. Noting that we have set $G \equiv L_P^2 = 1$ and adding the correct proportionality constant, we get:

$$\int_{\mathcal{V}} \sqrt{h} d^3x u_a J^a[\xi] = \int_{\mathcal{V}} d\Sigma_a J^a[\xi] = \int_{\partial\mathcal{V}} \frac{\sqrt{\sigma} d^2x}{8\pi L_P^2} (N r_\alpha a^\alpha) \quad (53)$$

This result is valid for any region \mathcal{V} in any spacetime. Let us now choose the boundary to be a $N(t, \mathbf{x}) = \text{constant}$ surface within the $t = \text{constant}$ surface. In the above expression, r_α is then the normal to the $N(t, \mathbf{x}) = \text{constant}$ surface within the $t = \text{constant}$ surface. Therefore, one can write $r_\alpha \propto D_\alpha N$ or as $r_i \propto h_i^j \nabla_j N$ where $h_i^j = \delta_i^j + u^i u_j$ is the projection tensor to the $t = \text{constant}$ surface. Since $N a_i = h_i^j \nabla_j N$, it follows that r_i and a_i are in the same direction even in the most general case (non-static, $N_\alpha \neq 0$). Normalizing it, we can take $r_\alpha = a_\alpha/a$ where a is the magnitude of the acceleration. This leads to $N r_\alpha a^\alpha = Na = (h^{ij} \nabla_i N \nabla_j N)^{1/2}$. So, if we choose the boundary to be a surface with $N = \text{constant}$ (which is a generalization of the notion of an equipotential surface), we can interpret $T_{loc} = N r_\alpha a^\alpha / 2\pi = Na / 2\pi$ as the (Tolman redshifted) local Davies–Unruh temperature of the observers with four-velocity $u_a = -N \delta_a^0$. These observers, who are moving normal to the $t = \text{constant}$ hypersurfaces will have the acceleration a with respect to the local freely falling observers. The local vacuum of the freely falling frame will appear to be a thermal state with temperature $T_{loc} = Na / 2\pi$ to these observers. So we can write:

$$2 \int_{\mathcal{V}} \sqrt{h} d^3x u_a J^a[\xi] = \int_{\partial\mathcal{V}} \frac{\sqrt{\sigma} d^2x}{L_P^2} \left(\frac{1}{2} T_{loc} \right) \quad (54)$$

Thus, (twice) the Noether charge contained in a $N = \text{constant}$ surface is equal to the equipartition energy of the surface when we attribute $(1/L_P^2)$ degrees of freedom per unit area. (We will comment on the factor 2 in a moment.)

Another, equivalent interpretation emerges, if we think of $s = \sqrt{\sigma}/4L_P^2$ as the analogue of the entropy density. Then we get, directly from Eq. (53), the result:

$$\int_{\mathcal{V}} \sqrt{h} d^3x u_a J^a[\xi] = \int_{\partial\mathcal{V}} d^2x T s \quad (55)$$

which is the heat (enthalpy) density (TS/A) of the boundary surface. This interpretation of a quarter of the area as entropy is obvious in those limiting cases in which the boundary surface becomes a horizon (like e.g., the surface $t = \text{constant}$, $r = 2M + \epsilon$ in the Schwarzschild metric in the limit of $\epsilon \rightarrow 0$). But one can define this notion in a more general context, in terms of local Rindler observers, as follows: Any sufficiently small patch of the boundary can be thought of as a small section of the $y - z$ plane which acts as a cross-section of the local Rindler horizon for a suitably defined observer accelerating along the $x -$ axis with acceleration a . Then, one can again attribute the

entropy density $s = \sqrt{\sigma}/4$ and the temperature $T_{loc} = Na/2\pi$ to this patch of area from the point of view of the local Rindler observer.

Let me comment on the factor 2 on the left hand side of Eq. (54) and connect it to a result familiar to general relativists (also see [24]). The integral on the right of Eq. (54) gives $(1/2)TA = 2TS$ if we take (for the sake of illustration) $T = \text{constant}$ on the boundary and $S = A/4$. Therefore, the Noether charge Q is just the heat content (enthalpy) $Q = TS$, which is also clear from Eq. (55). Thus, the Noether charge is *half* of the thermal, equipartition, energy of the surface $(1/2)TA = 2TS$ if we attribute $(1/2)T$ per surface degree of freedom. In the case of the Schwarzschild geometry, say, the thermal, equipartition, energy of the surface is just the total mass $M = 2TS$. *But what the Noether charge measures is the heat content (enthalpy) $E - F = TS$ which is precisely $(M/2)$.* In fact, this is a well-known “problem” (see, e.g. [33]) when one tries to define total mass of a spacetime (which asymptotically tends to the Schwarzschild limit) using the so-called Komar integral. In this context, ξ^a will become the standard timelike Killing vector and the Noether potential will be the Komar potential. The integral one performs with the Killing vector ξ^a is identical to the computation of the Noether charge above and one gets $(M/2)$. In classical relativity, this was considered very puzzling because in classical general relativity we (at best!) only have a notion of energy but no notion of heat content (TS), free energy ($F = E - TS$), etc. The thermodynamic perspective—which requires \hbar to define the Davies–Unruh temperature $k_B T = (\hbar/c)(\kappa/2\pi)$ from an acceleration κ —tells us that the Noether charge is the heat content (enthalpy) TS and *not* the energy $2TS$, and that the result *must* be $M/2$ for consistency. In short, classical general relativity can only interpret M physically (as energy) while the thermodynamic considerations allow us to *also* interpret $M/2$ physically as the heat content TS . *This is yet another case of thermodynamic considerations throwing light on some puzzling features of classical general relativity.*

Getting back to our main theme, we see that the Noether charge has a delightfully simple interpretation as the surface heat content (or half the surface equipartition energy) in the most general context. I did not have to assume static nature, existence of Killing vectors, asymptotic behaviour, etc. As long as the boundary is an $N(t, \mathbf{x}) = \text{constant}$ surface in an otherwise general context, the Noether charge for the time-development vector, contained in a bulk region of space is equal to the boundary enthalpy. To do this, we have only used results from quantum field theory in curved spacetime and have *not* used the gravitational field equations.

I will now show how the gravitational field equations themselves acquire an interesting interpretation in this language.

4.2 Gravitational dynamics and holographic equipartition

Since the Noether charge is related to TS while δN_{bc}^a is related to δT , it is obvious that one can interpret the gravitational dynamics in terms of thermodynamic variables using Noether currents. To do this, we take the dot product of the Noether current $J^a[\xi]$ (given in Eq. (46) with $v^a = \xi^a$) with u_a and use the result in Eq. (51) to obtain:

$$u_a g^{ij} \xi_\xi N_{ij}^a = D_\alpha (2Na^\alpha) - 2N R_{ab} u^a u^b \quad (56)$$

We next integrate this result over a 3-dimensional region \mathcal{R} with the measure $\sqrt{h} d^3x$, resulting in:

$$\int_{\mathcal{R}} d^3x \sqrt{h} u_a g^{ij} \xi_\xi N_{ij}^a = \int_{\partial\mathcal{R}} d^2x \sqrt{\sigma} r_\alpha (2Na^\alpha) - \int_{\mathcal{R}} d^3x \sqrt{h} (2Nu^a u^b R_{ab}) \quad (57)$$

where r_α is the normal to the boundary of the 3-dimensional region. This result, again, has a simple physical meaning if we choose the boundary to be a $N(t, \mathbf{x}) = \text{constant}$ surface within the $t = \text{constant}$ surface. We introduce the gravitational dynamics through $R_{ab} = (8\pi L_P^2) \mathcal{F}_{ab}$ and divide the whole equation by $8\pi L_P^2$, to get:

$$\int_{\mathcal{R}} \frac{d^3x}{8\pi L_P^2} \sqrt{h} u_a g^{ij} \xi_\xi N_{ij}^a = \int_{\partial\mathcal{R}} \frac{d^2x \sqrt{\sigma}}{L_P^2} \left(\frac{Na_\alpha r^\alpha}{4\pi} \right) - \int_{\mathcal{R}} d^3x N \sqrt{h} (2u^a u^b \mathcal{F}_{ab}) \quad (58)$$

As we said before, even in the general, time-dependent case, r_i is in the direction of a_i and we can interpret $T_{loc} = Na_\alpha r^\alpha / 2\pi = Na / 2\pi$ as the Tolman redshifted Davies–Unruh temperature. So the first term on the right hand side can be interpreted as an integral over $(dA/L_P^2)(1/2)k_B T_{loc}$. In the second term, we identify $2N \mathcal{F}_{ab} u^a u^b = (\rho + 3p)N$ as the Komar energy density. (Note the factor 2 in the left hand side of this standard definition of the Komar energy density $(\rho + 3p)$. Algebraically, this is exactly the factor 2 which makes everything consistent when we proceed from the heat content to the equipartition energy.) Thus the result in Eq. (58) can be summarized in the form:

$$\frac{1}{8\pi L_P^2} \int_{\mathcal{R}} d^3x \sqrt{h} u_a g^{ij} \xi_\xi N_{ij}^a = \int_{\partial\mathcal{R}} \frac{d^2x \sqrt{\sigma}}{L_P^2} \left(\frac{1}{2} k_B T_{loc} \right) - \int_{\mathcal{R}} d^3x \sqrt{h} \rho_{\text{Komar}} \quad (59)$$

This result, again, has a remarkable physical meaning. If the spacetime is static and we choose the foliation such that ξ^a is the Killing vector, then $\xi_\xi N_{ij}^a = 0$ and the left hand side vanishes. The equality of two terms on the right hand side can be thought of as representing the *holographic equipartition* [26–28] if we define the bulk and surface degrees of freedom along the following lines: We take the number of surface degrees of freedom to be:

$$N_{\text{sur}} \equiv \frac{A}{L_P^2} = \int_{\partial\mathcal{R}} \frac{\sqrt{\sigma} d^2x}{L_P^2} \quad (60)$$

We next define an *average* temperature T_{avg} of the boundary surface $\partial\mathcal{R}$ by:

$$T_{\text{avg}} \equiv \frac{1}{A} \int_{\partial\mathcal{R}} \sqrt{\sigma} d^2x T_{loc} \quad (61)$$

Next, we will define the bulk degrees of freedom N_{bulk} by the following procedure: If the energy E in the region \mathcal{R} has reached equipartition at the average surface temperature T_{avg} , then $|E| = (1/2)N_{\text{bulk}}k_B T_{\text{avg}}$; that is, we define the number of bulk degrees of freedom by

$$N_{\text{bulk}} \equiv \frac{|E|}{(1/2)k_B T_{\text{avg}}} = \pm \frac{1}{(1/2)k_B T_{\text{avg}}} \int_{\mathcal{R}} \sqrt{h} d^3x \rho_{\text{Komar}} \quad (62)$$

where E is the total Komar energy in the bulk region \mathcal{R} contributing to gravity. (The \pm is to ensure that N_{bulk} remains positive even when the Komar energy becomes negative.) This is the relevant value of N_{bulk} if we assume equipartition holds for the energy E in the bulk region with the average surface temperature. Our result in Eq. (59) then says that *comoving observers in any static spacetimes* will indeed find:

$$N_{\text{sur}} = N_{\text{bulk}} \quad (\text{Holographic equipartition}) \quad (63)$$

That is, the equipartition is holographic in all static spacetimes. I have already discussed in Sect. 1 how this result leads to $E = 2TS$ for a Schwarzschild black hole where we have $N_{\text{sur}} = (A/L_P^2) = 4S$; $N_{\text{bulk}} = E/((1/2)T)$ so that $N_{\text{sur}} = N_{\text{bulk}}$ is the same as $E = 2TS$.

What is more, Eq. (59) suggests that *the discrepancy from holographic equipartition—resulting in a non-zero value for the right hand side—is what drives the dynamical evolution of the spacetime*. We can write Eq. (59) as:

$$\int \frac{d^3x}{8\pi L_P^2} \sqrt{h} u_a g^{ij} \xi_{\xi} N_{ij}^a = \frac{1}{2} k_B T_{\text{avg}} (N_{\text{sur}} - N_{\text{bulk}}) \quad (64)$$

Note that, even in a static spacetime non-static observers will perceive a departure from holographic equipartition because Eq. (59)—while being generally covariant—is foliation dependent through the normal u_i . This is very clear if one considers the de Sitter spacetime which can be expressed as a static spacetime in the Schwarzschild-like coordinates or as a time dependent spacetime in the Friedmann coordinates. Observers moving normal to the static equal time surfaces will see that the de Sitter spacetime maintains holographic equipartition with $N_{\text{sur}} = N_{\text{bulk}}$. The observers moving normal to the equal time surfaces in the Friedmann coordinates, on the other hand, are geodesic observers with $a_i = 0$. They will see that $N_{\text{sur}} \neq N_{\text{bulk}}$ and in fact will find that this discrepancy leads to the time dependence of the metric functions through Eq. (64). Similar phenomenon can occur in any spacetime if we compare synchronous coordinates with other coordinates. This should cause no more surprise than the fact that, de Sitter spacetime, say, can be viewed as static or time dependent in two different foliations.

There are two other ways of rewriting $u_a g^{ij} \xi_{\xi} N_{ij}^a$ in the left hand side of Eq. (64) by relating it to more familiar constructs in the Hamiltonian formulation of relativity [34]. A straightforward computation shows that the gravitational momentum flux

$\sqrt{h}u_a g^{ij} \xi_\xi N_{ij}^a$ has two other alternative descriptions which are physically illuminating. We can show [see Appendix 9.3, Eqs. (157) and (165)] that:

$$\sqrt{h}u_a g^{ij} \xi_\xi N_{ij}^a = -h_{ab} \xi_\xi p^{ab}; \quad p^{ab} \equiv \sqrt{h}(Kh^{ab} - K^{ab}) \quad (65)$$

and

$$\sqrt{h}u_a g^{ij} \xi_\xi N_{ij}^a = 2N\sqrt{h}[K_{ij}K^{ij} - u^a \nabla_a K] \quad (66)$$

So we can rewrite Eq. (64) either as

$$\int \frac{d^3x}{4\pi L_p^2} N\sqrt{h}[K_{ij}K^{ij} - u^a \nabla_a K] = \frac{1}{2}k_B T_{\text{avg}}(N_{\text{sur}} - N_{\text{bulk}}) \quad (67)$$

or, more simply, as:

$$- \int \frac{d^3x}{8\pi L_p^2} h_{ab} \xi_\xi p^{ab} = \frac{1}{2}k_B T_{\text{avg}}(N_{\text{sur}} - N_{\text{bulk}}) \quad (68)$$

which clearly shows how the departure from holographic equipartition drives the evolution of geometry [35], providing an interesting alternative description of spacetime dynamics.

These results provide a direct and simple relation between the dynamical evolution of the spacetime geometry and the departure from thermodynamic equipartition. It also encodes the holographic nature of gravity (noticed earlier in several contexts, like e.g., in [32, 36, 37]) by introducing the surface and bulk degrees of freedom of a region and their equality in static spacetimes. On the left hand side, we essentially have a measure of the change in the gravitational momentum. Given the correspondence between δN_{bc}^a and δT , one should be able to interpret the left hand side as the ‘heating of the spacetime’ [38, 40]. One also notices the obvious connection with the membrane paradigm [41–43] which brings in a viscous tensor and the concept of “dissipation without dissipation” (introduced in e.g., Sect. 2.3 of [38, 39]). Further work is required to make these connections precise.

It is also gratifying that the result involves the rate of change of p^{ab} which encodes the dynamical content of the field equations in the following sense: In the standard Hamiltonian formulation of general relativity, one has two constraint equations $\mathcal{H} = 0$, $\mathcal{H}_\alpha = 0$, one equation for ξh_{ab} which defines p_{ab} in terms of h_{ab} , and one dynamical equation involving ξp_{ab} . What we get here is the trace part of this dynamical equation. This is understandable, because to derive the result in Eq. (64) I only used the constraint equation $R_{ab}u^a u^b = (8\pi L_p^2) \mathcal{F}_{ab}u^a u^b$ and we cannot get out more information than we put in. But, as is well-known (see e.g., p. 259 of [44]), the validity of the constraint equation for all observers leads to the full set of field equations of gravity. In this sense, we only need the validity of Eq. (64) for all regions of space and all foliations in order to obtain the full set of field equations. It should be possible to “reverse-engineer” the above result and do this, but I believe that it is not going to add more insights into the situation. (I will describe in Sect. 6 how to obtain the gravitational field equations from a suitable thermodynamic variational principle properly).

As a description of the gravitational dynamics, we are essentially comparing Eq. (64) with the standard equation $R_{ab} = (8\pi L_P^2)\mathcal{F}_{ab}$. My main problem with the latter is that it is very difficult to attribute *physical* meaning to the left hand side. Further, there is no natural separation of time-dependent and static geometries and it is not obvious what drives the spacetime evolution. It is *not* T_{ab} because: (a) One can have time dependent solutions when $T_{ab} = 0$ and static solutions when $T_{ab} \neq 0$. (b) The notion of time independence, while can be expressed quite geometrically in terms of existence of a timelike Killing vector, is not easy to translate in terms of actual metric functions. As I have said several times, the natural coordinate system used by two different sets of observers can have a time independent metric in one case and a time-dependent metric in the other.

The description in Eq. (64) does better as regards these issues. First, it brings in the observers through the foliation and—contrary to what is sometimes thought—*this is a very desirable feature*. All thermodynamics is observer dependent (see Sect. 4 of [11]) and any attempt to re-interpret the field equations in the thermodynamic language has to acknowledge this feature up front. [Of course, once we demand the validity of Eq. (64) for all foliations, we will get back $R_{ab} = (8\pi L_P^2)\mathcal{F}_{ab}$ which has no trace of foliation left in it. But the basic equation, viz. Eq. (64) has a foliation dependence].

Second, for a given observer, we now have a natural separation of static and dynamic spacetimes in terms of holographic equipartition. In stationary geometries, we can choose ξ^a to be the timelike Killing vector and we immediately get $N_{\text{sur}} = N_{\text{bulk}}$. When this condition does not hold, we get time evolution due to departure from holographic equipartition. This result also provides a physically transparent statement about dynamics. One should also mention in this context that the holographic thermodynamics necessarily involves a nonlocal description because of the involvement of a surface and bulk.

When there is no matter, our Eq. (59) reduces to an interesting form:

$$\frac{1}{8\pi L_P^2} \int d^3x \sqrt{h} u_a g^{ij} \xi_\xi N_{ij}^a = \int_{\partial\mathcal{R}} \frac{d^2x \sqrt{\sigma}}{L_P^2} \left(\frac{1}{2} k_B T_{\text{loc}} \right) \quad (69)$$

This equation holds, among other contexts, in the description of a gravitational wave spacetime. It shows that the momentum change in a region containing a source-free gravitational field is directly related to the surface heat energy. The comments made earlier regarding the membrane paradigm [41–43] and the concepts like ‘heating of spacetime’ [38, 40], “dissipation without dissipation” (introduced in Sect. 2.3 of [38, 39]) are again applicable in this context. I hope to get back to the implications of this description in a future work.

4.3 An explicit example

As a concrete example of the above ideas, let us consider a time dependent, spherically symmetric, metric of the form

$$ds^2 = -V dt^2 + W dr^2 + r^2 d\Omega^2 \quad (70)$$

where V and W are functions of r and t . In addition to standard black hole space-times and collapse scenarios, *this form of the metric also describes all the Friedmann universes*, since all of them are spherically symmetric and (in general) time dependent. The purpose of this short section is to show how the balance between the three terms in Eq. (59) is achieved in such a concrete case. My discussion will be brief and applications to cosmology will be discussed elsewhere.

The acceleration computed using $a^i = (g^{ij} + u^i u^j) \nabla_j \ln \sqrt{V}$ gives $a^r = (V'/2VW)$ with the magnitude $Na = (V'/2\sqrt{VW})$ so that

$$D_\alpha(2Na^\alpha) = \frac{1}{r^2 \sqrt{W}} \partial_r \left(r^2 \frac{V'}{\sqrt{VW}} \right) \quad (71)$$

(We use a prime to denote the derivative with respect to r and an overdot to denote the derivative with respect to t .) Another explicit computation gives

$$\begin{aligned} 2NR_{ab}u^a u^b &= \frac{2N}{N^2} R_{00} \\ &= \frac{2}{\sqrt{V}} \left(\frac{V''}{2W} - \frac{V'W'}{4W^2} + \frac{V'}{Wr} - \frac{(V')^2}{4VW} - \frac{\ddot{W}}{2W} + \frac{\dot{W}^2}{4W^2} + \frac{\dot{V}\dot{W}}{4VW} \right) \end{aligned} \quad (72)$$

We now see that the first four terms within the bracket in Eq. (72) will survive even in the static limit, while the remaining four terms have time derivatives. On the other hand, the expression in Eq. (71) has no time derivatives and we know that Eq. (71) and Eq. (72) *have to match* in the static limit. This is indeed the case because, on expanding the right hand side of Eq. (71), one can easily show that

$$D_\alpha(2Na^\alpha) = \frac{2}{\sqrt{V}} \left\{ \frac{1}{2} \frac{V''}{W} - \frac{1}{4} \frac{V'W'}{W^2} + \frac{V'}{rW} - \frac{1}{4} \frac{(V')^2}{VW} \right\} \quad (73)$$

Clearly these terms—which remain the same even in the time dependent case—match with the first four terms in Eq. (72), allowing us to write

$$2NR_{ab}u^a u^b = D_\alpha(2Na^\alpha) + \frac{2}{\sqrt{V}} \left\{ \frac{\dot{V}\dot{W}}{4VW} + \frac{\dot{W}^2}{4W^2} - \frac{\ddot{W}}{2W} \right\} \quad (74)$$

We therefore identify, from our general result, that

$$-u_a g^{ij} \xi_{ij} N_{ij}^a = \frac{2}{\sqrt{V}} \left\{ \frac{\dot{W}}{4W} \frac{(VW)'}{(VM)} - \frac{\ddot{W}}{2W} \right\} = -\frac{1}{\sqrt{W}} \frac{\partial}{\partial t} \left(\frac{\dot{W}}{\sqrt{VW}} \right) \quad (75)$$

which is an explicit expression for $u_a g^{ij} \mathfrak{L}_\xi N_{ij}^a$ in this simple context. For the sake of completeness, one can verify the last expression explicitly. In this particular case, we have $K_{\mu\nu} = -(1/2N)\partial_t h_{\mu\nu}$, leading to $K_{rr} = -(1/2\sqrt{V})\dot{W}$ and $K = -(1/2\sqrt{V})(\dot{W}/W)$. Working out the expression for $2N(K_{\mu\nu}K^{\mu\nu} - u^a\partial_a K)$, one can explicitly verify that it matches with $u_a g^{ij} \mathfrak{L}_\xi N_{ij}^a$. Thus, the static terms of $2NR_{ab}u^a u^b$ match with $D_\alpha(2Na^\alpha)$ while the time dependent terms match with $-u_a g^{ij} \mathfrak{L}_\xi N_{ij}^a$.

5 Gravitational energy and the surface thermodynamics

I will now demonstrate yet another curious connection between bulk gravitational dynamics and surface thermodynamics by showing that, when the field equations hold, there is a natural notion of total energy of matter plus gravity which is equal to the surface heat content. To do this, we first have to define a gravitational four momentum flux P^a such that its integral over $u_a \sqrt{h} d^3x$ gives a sensible notion of gravitational energy. The usual folklore is that such notions are either impossible to define or (when defined) fairly useless. We will see that the thermodynamic considerations change the picture quite a bit. In this section it is more convenient to use units with $16\pi G = 1$ so that the expressions for Noether current etc appears simpler.

5.1 Bulk total energy matches the surface heat energy

It is, of course, well-known that one cannot define a covariant, local, notion of gravitational energy density. Further, the very notion of energy or energy density, operationally speaking, has an implicit observer dependence. For example, a particle can have a covariant, observer independent, four-momentum \mathbf{p} ; but the energy attributed to this particle by an observer will be $E = -\mathbf{p} \cdot \mathbf{u}$ where \mathbf{u} is the four-velocity of the observer. Taking a cue from this we will use an *extra vector field* q^a to define the gravitational four momentum density $P_H^a[q]$ associated with it. (The subscript H is to remind us that we are using the Einstein–Hilbert Lagrangian; later on, in Sect. 7.2, we will discuss what happens when we use other forms of the Lagrangian. When no subscript is used, it means the Einstein–Hilbert Lagrangian.) Of course, such an entity depends on an extra vector field and not an intrinsic property of the spacetime geometry alone; however, we will see that it has a natural interpretation when the vector is chosen to be the time-development vector ξ^a . In the past such ideas have been explored with ζ^a (for a small sample see e.g., [29, 45–47] and references therein) but, again, we will see that the thermodynamic interpretation works better with ξ^a .

The exact form for $P_H^a[q]$ associated with a vector field q^a has already been motivated in Sect. 2.1 and we have defined the vector in Eq. (28) as

$$P^a(q) \equiv g^{ij} \mathfrak{L}_q N_{ij}^a + Lq^a \quad (76)$$

The Noether current approach allows us to study this P^a quite easily. To begin with, we have the relation:

$$\nabla_b J^{ab} = 2R_b^a q^b + g^{ij} \xi_q N_{ij}^a = 2G_b^a q^b + Lq^a + g^{ij} \xi_q N_{ij}^a \equiv 2G_b^a q^b + P^a \quad (77)$$

from which we get an alternative, and more useful, definition:

$$P^a(q) = g^{ij} \xi_q N_{ij}^a + Lq^a = \nabla_b J^{ab} - 2G_b^a q^b \quad (78)$$

It is also easy to see that

$$\nabla_a P^a = -2\nabla_a (G_b^a q^b) = -2G^{ab} \nabla_a q_b \quad (79)$$

which vanishes on-shell in pure gravity making P^a a conserved current in this context. In fact, P^a is what is usually called Noether current in the literature, obtained from our Noether current J^a by dropping terms that vanish on-shell. However, the idea that the Noether current is actually a measure of gravitational momentum flux does not seem to have been emphasized (and explored) in detail. We have seen enough evidence already to realize that J^a is directly related to energy (rather than, for example, entropy; see e.g., [24, 48]) and results in this section will reinforce this idea.

All this is true for general q^a . Let us now specialize to $q^a = \xi^a = Nu^a$ when we will be able to obtain several interesting results. Taking the dot product of $P^a(\xi)$ with u_a where $u_a = -N\delta_a^0$ and using Eq. (51), we get:

$$u_a P^a(\xi) = -2NG_{ab}u^a u^b + u_a J^a[\xi] = -2NG_{ab}u^a u^b + 2D_\alpha(Na^\alpha) \quad (80)$$

On the other hand, the corresponding matter energy density is given by $NT_{ab}u^a u^b$. Adding it to gravitational part, using the constraint equation $2G_{ab}u^a u^b = T_{ab}u^a u^b$, and integrating the expression over a three dimension region bounded by an $N =$ constant surface (as we did in the previous sections), we get a simple expression for the total (matter+gravity) energy in a bulk region on shell:

$$\int_{\mathcal{R}} d^3x \sqrt{h} u_a [P^a(\xi) + NT_b^a u^b] = \int_{\partial\mathcal{R}} d^2x T s \quad (81)$$

That is, when the gravitational field equations are satisfied, the total energy in the bulk region is exactly equal to the heat content of the bounding surface. This is yet another connection between gravitational dynamics, holography and surface thermodynamics. We can also say that this total energy is *half* of the thermal, equipartition, energy of the boundary obtained by attributing $(1/2)k_B T$ energy to each of the (A/L_P^2) degrees of freedom.

The expression for gravitational energy density itself has some interesting features. Note that the quantity

$$2G_{ab}u^a u^b = K^2 - K_{ab}K^{ab} + {}^3R = -\mathcal{H}_{\text{adm}} \quad (82)$$

is essentially the 00 bulk part of the ADM Hamiltonian defined by the second equality above (see e.g., Sect. 12.4.1 of [44]). It therefore follows that $u_a P^a[\xi]$ is essentially \mathcal{H}_{adm} plus an *important* total divergence:

$$u_a P^a[\xi] = N\mathcal{H}_{\text{adm}} + D_\alpha(2Na^\alpha) \quad (83)$$

The integral of this expression over a three surface with measure $\sqrt{h}d^3x$ will lead to the surface heat content from the second term and the bulk energy from the first term:

$$\int_{\mathcal{R}} d^3x \sqrt{h} u_a P^a(\xi) = \int_{\mathcal{R}} d^3x \sqrt{h} N \mathcal{H}_{\text{adm}} + \int_{\partial\mathcal{R}} d^2x T s \quad (84)$$

When the constraint equation holds, the sum of \mathcal{H}_{adm} and the matter energy density H vanishes, which leads to the surface heat content as the final result.

5.2 Variation of the gravitational energy

Since the gravitational energy defined through our P^a seems to lead to interesting results, it is useful to study its properties and—in particular—investigate how it changes due to processes operating on the boundary. In fact, the current $P^a(q)$ (for an arbitrary diffeomorphism with vector field q^a) also allows us to introduce a symplectic structure for gravity which is well-known in the literature (for a small sample, see e.g. [49–57]). I will show that, these results also have an interesting thermodynamic interpretation.

The symplectic structure can be obtained by varying $\sqrt{-g} P^a$ and manipulating the resulting terms. We then find that, for arbitrary variations with fixed q^a , we have the result [see Appendix 9.4, Eq. (174)]:

$$\delta(\sqrt{-g} P_H^a) - q^a R_{ij} \delta f^{ij} = \sqrt{-g} \omega^a + \partial_c \left(f^{lm} \delta N_{lm}^{[a} q^{c]} \right) \quad (85)$$

where we have defined the symplectic form as

$$\sqrt{-g} \omega^a(\delta, \mathbb{L}_q) \equiv \delta f^{lm} \mathbb{L}_q N_{lm}^a - (\mathbb{L}_q f^{lm}) \delta N_{lm}^a \quad (86)$$

This expression is completely general and ω^a involves one arbitrary variation δ and one Lie derivative \mathbb{L}_q .

We can use this result to obtain the change in the total gravitational energy as the system evolves, in terms of a Lie derivative along ξ^a . To do this, we need to compute $\mathbb{L}_\xi(\sqrt{h} u_a P_H^a(\xi))$ which is straightforward. First, we obtain from Eq. (85) an expression for the variation of $\delta(\sqrt{h} u_a P_H^a(\xi))$ (after setting $q^a = \xi^a$), either by using the fact that $u_a = -N\nabla_a t$ and taking the dot product or working it out from the basic definition. The result is [see Appendix 9.4, Eq. (183)]:

$$\delta(\sqrt{h} u_a P_H^a(\xi)) + R_{ab} \delta f^{ab} = \sqrt{h} \omega^a(\delta, \mathbb{L}_\xi) u_a + \partial_c \left[h_a^c f^{lm} \delta N_{lm}^a \right] \quad (87)$$

which is valid for arbitrary variations. Next we evaluate the above expression when the variation is due to a diffeomorphism along ξ^a so that we can replace δ by $-\mathcal{L}_\xi$. Because of the antisymmetry, the ω^a term drops out and we obtain [see Appendix 9.4, Eq. (184)]:

$$\mathcal{L}_\xi \left(\sqrt{h} u_a P_H^a(\xi) \right) = \partial_c \left[\left(2G_b^c \xi^b + h_a^c g^{lm} \mathcal{L}_\xi N_{lm}^a \right) \sqrt{-g} \right] \quad (88)$$

Integrating over a region of space, we get:

$$\mathcal{L}_\xi H_{\text{grav}} \equiv \mathcal{L}_\xi \int_{\mathcal{R}} d^3x \sqrt{h} u_a P^a = \int_{\partial\mathcal{R}} d^2x \sqrt{\sigma} N r_a \left(T^{ab} \xi_b + g^{lm} \mathcal{L}_\xi N_{lm}^a \right) \quad (89)$$

where we have also used the field equations $2G_{ab} = T_{ab}$. This expression shows how the energy in the bulk changes due to processes taking place at the boundary. (The result in Eq. (85) was obtained under the assumption $\delta q^a = 0$ but it will continue to hold if the variation δ is due to a diffeomorphism along the same vector field q^a because $\mathcal{L}_q q^a = 0$ will make all the additional terms vanish. This can be verified explicitly). Since we have already established that the energy is given by a boundary integral in Eq. (81), it is obvious that we will only get a surface contribution for $\mathcal{L}_\xi H_{\text{grav}}$. Of the two terms in Eq. (89), one is clearly due to the matter energy flux across the boundary. The second term, which is more interesting to us, is again the ubiquitous $f\delta N$ term. We once again see that this can be thought of as the change in the gravitational heat due to the processes at the boundary. In fact, if we consider pure gravity, this result becomes:

$$\mathcal{L}_\xi H_{\text{grav}} = \int_{\partial\mathcal{R}} d^2x \sqrt{\sigma} N r_a \left(g^{lm} \mathcal{L}_\xi N_{lm}^a \right) \quad (90)$$

which is directly applicable to, say, gravitational waves. We find that the energy associated with a gravitational wave metric in a bulk region changes due to surface processes involving $\mathcal{L}_\xi N_{lm}^a$. This is similar to the result we obtained earlier for the dynamical evolution of pure gravity in Eq. (69) and should have an interpretation as the heating of the surface [38, 40] in terms of the surface viscous tensor that arises in the membrane paradigm [41–43] and the “dissipation without dissipation” (Ref. [38], Sect. 2.3 of Ref. [39]). I have already commented about this connection earlier and the mathematical results in Eq. (69) and Eq. (90) essentially describe the same physics.

We will conclude this section after commenting on the relation of these results to some other well-known results in the literature.

The symplectic form and related constructs also arise in the approach to compute horizon entropy using the central charge and Virasoro algebra, pioneered by Carlip [57–60]. This approach requires us to start with $P_H^a(q)$ defined for a vector q^a , and construct its Lie derivative along another vector v^a to obtain $\mathcal{L}_v P^a[q]$. This allows a natural definition of the symmetric $(\mathcal{L}_v P^a[q] + \mathcal{L}_q P^a[v])$ and the antisymmetric $(\mathcal{L}_v P^a[q] - \mathcal{L}_q P^a[v])$ combinations of the double variations. From the above analysis, we see that one can do the same with the energy density expression. The symplectic form vanishes in the symmetric expression, leading essentially to the surface term.

It is this surface term which contributes to the central charge construction and to the horizon entropy [57–60]. It has been shown in a previous work [61] that one can obtain the same result by working essentially with the surface term in the action and performing the necessary algebra. The results obtained in this paper show that these are fairly generic features arising from the fundamental relation between the total energy and the boundary heat energy.

We also mention briefly the connection between these boundary terms and the ones usually encountered in the conventional Hamiltonian formulation of general relativity. Using the second relation in Eq. (78), we can also relate this variation of P_H^a to the variation of the Noether potential. A simple calculation gives [see Appendix 9.4, Eq. (178)], when the original spacetime is on-shell with $G_{ab} = 0$ but $\delta G_{ab} \neq 0$, the result:

$$\partial_b \left\{ \delta(\sqrt{-g} J_H^{ab}) - \sqrt{-g} g^{lm} \delta N_{lm}^{[a} q^{b]} \right\} = \sqrt{-g} \omega^a + 2\delta(G_b^a q^b \sqrt{-g}) \quad (91)$$

This allows us to make a connection with the usual Hamiltonian formulation of the theory. In the usual approach, one deals (see e.g., [62–64]) with the integral of the variation of $\delta(\mathcal{H}N + \mathcal{H}^\alpha N_\alpha)$ (where \mathcal{H} and \mathcal{H}^α are the constraints essentially corresponding to G_0^0 and G_α^0) which will lead to a bulk term *and a surface contribution* in the form:

$$\delta \int_{\mathcal{R}} d^3x (N\mathcal{H} + N_\alpha \mathcal{H}^\alpha) = \int_{\mathcal{R}} d^3x \delta B + \int_{\partial\mathcal{R}} d^2x \delta S \quad (92)$$

The variation in the left hand side is the same as $-\delta(2G_{ab}u^a q^b)$ if we choose $q^a = \zeta^a = Nu^a + N^\alpha$. Further, the bulk term in the above variation, on the right hand side, has to cancel the bulk term involving the integral of $u_a \omega^a$ when the equations of motion hold. (In fact, this is how one obtains the equations of motion in the symplectic approach.) The consistency of Eq. (91) then requires that the surface term in Eq. (91) *should* match with the surface term in Eq. (92). That is, we have the result:

$$\int_{\partial\mathcal{R}} d^2x \delta S = \int_{\partial\mathcal{R}} d^2x r_b u_a \left(\delta(\sqrt{-g} J^{ab}) - f^{lm} N_{lm}^{[a} \zeta^{b]} \right) \quad (93)$$

This allows a simple way to identify and interpret the surface term in the variation of the Hamiltonian. While this result is known previously in the literature, our approach suggests that even the conventional Hamiltonian formulation of gravity can be possibly recast in a thermodynamic language.

6 Where did we go wrong with gravity?

The results described above, as well as several other pieces of work in this area, suggest an intriguing connection between: (i) the thermodynamics attributed to null surfaces by local Rindler observers and (ii) dynamics of gravity. But in the description given above, I have only *re-interpreted* the standard gravitational dynamics—which, *I think, is actually flawed*—in the thermodynamic language. I will now show how it is

possible to provide a completely independent, stand-alone, derivation of the *correct* gravitational field equations from a thermodynamic perspective.

6.1 The single most important fact about the gravitational dynamics

To motivate this, I begin by stressing the *single most important fact* about gravitational dynamics which—purely because of historical accident—is completely ignored in formulating gravitational field equations: *Gravity does not couple to the bulk energy density arising from the addition of a constant to the matter Lagrangian.* Any attempt to describe gravity without incorporating this observed feature is bound to be wrong.

This fact, in turn, requires that the gravitational field equations *must be* invariant under the symmetry transformation of the matter sector equations:

$$L_{\text{matter}} \rightarrow L_{\text{matter}} + \text{constant}, \quad (94)$$

resulting in $T_b^a \rightarrow T_b^a + (\text{constant}) \delta_b^a$. (The electroweak symmetry breaking, for example, is equivalent to the shifting of the standard model Lagrangian by a large constant and we know that the evolution of the universe was unaffected by this transition.) The standard gravitational field equations, in contrast to matter field equations, are *not* invariant under the addition of a constant to the matter Lagrangian. The addition of the constant changes the energy-momentum tensor of the matter by

$$T_b^a \rightarrow T_b^a + (\text{constant}) \delta_b^a \quad (95)$$

Clearly, the gravitational field equations now become $\mathcal{G}_b^a = T_b^a + (\text{constant}) \delta_b^a$ which is equivalent to the introduction of a cosmological constant if one was not present originally, or changing its numerical value, if a cosmological constant was originally present in the gravitational Lagrangian. Obviously, this is the crucial problem related to the cosmological constant, viz., that its numerical value (either zero or non-zero) can be altered by the transformation in Eq. (94) which leaves the matter equations unchanged. A particle physicist interested in the standard model can choose the overall constant in the matter Lagrangian arbitrarily because the standard model does not care for this constant. But each choice for this constant will lead to a different value for the cosmological constant and a different geometry for the universe, many of which, of course, are observationally untenable.

Another way of stating this problem is as follows: Suppose we discover a fundamental principle which allows us to determine the numerical value of the cosmological constant (either zero or non-zero). Such a principle cannot help us if the gravitational field equations are not invariant under the transformations in Eqs. (94) or (95). The above discussion allows us to identify three ingredients which are *necessary* to solve the cosmological constant problem:

1. The gravitational field equations *must be* made invariant under the transformations in Eqs. (94) and (95) so that gravity is “protected” from the shift in the zero level of the energy densities.

2. At the same time, the solutions to the field equations *must* allow the cosmological constant to influence the geometry of the universe, because without it we cannot possibly explain the observed accelerated expansion of the universe.
3. We need a fundamental physical principle to determine the numerical value of the cosmological constant since it cannot be introduced as a low energy parameter in the Lagrangian if the theory is invariant under the transformation in Eq. (94).

The first two requirements might at first sight sound impossible to satisfy simultaneously, but it can be achieved! The trick is to construct a set of gravitational field equations which are invariant under the transformation in Eq. (95) but allow the inclusion of a cosmological constant as an *integration constant* in the solutions. As an example, consider a theory in which the field equations are given by the requirement:

$$(\mathcal{G}_b^a - T_b^a)n_{an}{}^b = 0 \quad (96)$$

for all *null* vectors n^a in the spacetime [30, 31]. Here $\mathcal{G}_b^a = 2G_b^a$ in Einstein's theory and could be some other geometrical tensor in alternate theories of gravity, but necessarily satisfying the generalized Bianchi identity $\nabla_a \mathcal{G}_b^a = 0$. The above equations can be solved by $\mathcal{G}_b^a - T_b^a = F(x)\delta_b^a$, but the generalized Bianchi identity ($\nabla_a \mathcal{G}_b^a = 0$) and the conservation of the energy-momentum tensor ($\nabla_a T_b^a = 0$) imply that $F(x)$ must be a constant. Therefore, Eq. (96) is equivalent to the standard gravitational field equations with an arbitrary cosmological constant appearing as an integration constant. Thus, if we can construct a theory of gravity in which the field equations reduce to those in Eq. (96), then we would have achieved the first two requirements in the list for solving the cosmological constant problem.

This turns out to be an extremely strong demand and has important consequences often overlooked in attempts to “solve” the cosmological constant problem. To see this, consider any theory of gravity interacting with matter satisfying the following three conditions:

1. The theory is generally covariant so that the matter action is constructed by integrating a scalar Lagrangian $L_m(g_{ab}, \phi_A)$ over the measure $\sqrt{-g}d^4x$.
2. The matter equations of motion are invariant under the transformation $L \rightarrow L + C$ where C is a scalar constant.
3. The gravitational field equations are obtained by an *unrestricted* variation of the metric tensor g_{ab} in the total action obtained by integrating a local Lagrangian over the spacetime.

It is easy to see that we *cannot* solve the cosmological constant problem in any theory satisfying the above three requirements. (This was clearly emphasized in Sect. 4 of ref. [65]). In particular, one cannot obtain the gravitational field equations of the form in Eq. (96) in any theory which satisfies the above three criteria.

We thus see that even though all the three criteria stated above seem very reasonable, they together will prevent us from solving the cosmological constant problem; so we need to give up at least one of them. Assuming we do not want to give up general covariance of the theory or the freedom to add a constant to the matter Lagrangian, we can only tinker with the third requirement.

One simple way of obtaining Eq. (96) is to postulate that the gravitational field equations are obtained by varying the metric but keeping $\sqrt{-g} = \text{constant}$. Such theories, called unimodular theories of gravity—involving only a restricted variation of the metric to bypass the condition (3) above—have been studied in the literature in the past [66–69]. Unfortunately, the motivation to keep $\sqrt{-g} = \text{constant}$ is at best weak and at worst non-existent.

It is also possible to obtain Eq. (96) from an alternative perspective of gravity which treats gravity as an emergent phenomenon (see ref. [30,31]). In this approach, one associates thermodynamic potentials with all null vector fields in a spacetime. Maximization of the relevant thermodynamic potential (entropy, free energy, ...) associated with *all* null vectors simultaneously will then lead to Eq. (96). The maximization involves varying the null vector fields rather than the metric and hence it bypasses the third requirement in our list. The metric is not varied at all to obtain the field equations. In such an approach, the original variational principle itself (not just the field equations) is invariant under the transformation in Eq. (95).⁷

While this idea has been developed in Ref. [30,31], I will revisit it here for three reasons: (a) In the original work [30,31], we considered a functional obtained by integrating a scalar over *spacetime* with the measure $\sqrt{-g} d^4x$. It would be nicer to reformulate this idea using an integration over a *null surface* rather than over spacetime. (b) The ideas developed in the previous sections of this paper suggest a connection between this formalism and the Noether current, which I want to demonstrate. (c) The connection with the Noether current, in turn, allows us to provide an interpretation of the functional by relating it again to a $f\delta N$ type structure which we have seen repeatedly.

The first task of obtaining a variational principle based on a null surface is easy. A starting point for such a variational principle will be the following: Given a null surface with an integration measure $d\lambda d^2x \sqrt{\sigma}$ and a null congruence ℓ^a , let us construct the functional:

$$\mathcal{H} \equiv \int_{\lambda_1}^{\lambda_2} \frac{d\lambda d^2x}{16\pi} \sqrt{\sigma} [-2R_{ab} + 16\pi T_{ab}] \ell^a \ell^b \quad (97)$$

It is then obvious that extremising this functional with respect to *all* ℓ^a , subject to the constraint $\ell^2 = 0$, will lead to $R_b^a - 8\pi T_b^a = f(x)\delta_b^a$ which, on using $\nabla_a G_b^a = 0 = \nabla_a T_b^a$ leads to Einstein's equations with an undetermined cosmological constant term arising as an integration constant.

A nicer version of the above variational principle can be obtained by noting that for a null congruence ℓ^a on a null surface we have [see Appendix 9.5, Eq. (185)]: $R_{ab}\ell^a\ell^b = -\nabla_i(\theta\ell^i) - \mathcal{S}$ where $\nabla_i\ell^i = \theta + \kappa$ and we have defined

$$\mathcal{S} \equiv [\nabla_i\ell^j\nabla_j\ell^i - (\nabla_i\ell^i)^2] \quad (98)$$

⁷ For a complete solution to the cosmological constant problem, we need, in addition to this, a physical principle to determine its numerical value. Such a principle is described elsewhere [70,71] and I will not discuss it here.

which we shall later identify with the heat (enthalpy) density associated with the null surface. While integrating over a null surface with the measure $d\lambda d^2x \sqrt{\sigma}$, we can ignore terms of the kind $\nabla_i(\phi \ell^i)$ (for any scalar ϕ) since they produce only boundary contributions. (see Appendix 9.5, Eq. (188)). It follows that we now have an alternative variational principle (in which we vary ℓ^a) based on the expression:

$$Q \equiv \int_{\lambda_1}^{\lambda_2} \frac{d\lambda d^2x}{16\pi} \sqrt{\sigma} [2S + 16\pi T_{ab} \ell^a \ell^b] \quad (99)$$

Since $T_{ab} \ell^a \ell^b$ can be thought of as the heat (enthalpy) density $(\rho + p) = Ts = TS/V$ of matter, we can again think of $(S/8\pi)$ as essentially the heat density of the null surface.

We will now show that these results can again be expressed in terms of appropriate Noether charges which lead to the heat density of the null surface in terms of $\mathcal{E}_\ell N_{ab}^c$. This provides the generalization of the earlier results to null surfaces.

6.2 Heat density of the null surfaces

Let ℓ_a be a null congruence defining a null surface which may not be affinely parametrized. If we take $\ell_a = A(x) \nabla_a B(x)$, then it is easy to prove that $\ell^i \nabla_i \ell_j = \kappa \ell_j$ where $\kappa = \nabla_i A \nabla^i B = \ell^a \nabla_a \ln A$. Just as we computed the Noether current for ξ^a while dealing with spacelike surfaces, we will compute the Noether current for ℓ_a in the context of null surfaces. We get [see Appendix 9.5, Eqs. (192) and (194)]:

$$\ell_a J^a(\ell) = \nabla_b(\kappa \ell^b) - \kappa^2 = \mathcal{D}_a(\kappa \ell^a) + \frac{d\kappa}{d\lambda} \quad (100)$$

Note that the quantity $\nabla_b(\kappa \ell^b) - \kappa^2$ in the case of a null surface is analogous to the right hand side of $D_i a^i = D_\alpha a^\alpha = \nabla_i a^i - a^2$ which arises when we deal with spacelike surfaces. Unfortunately the projection to a surface ‘orthogonal’ to a null vector is not well defined for us to introduce a covariant derivative. Instead, we have to work with a co-null vector k_a defined such that $k^2 = 0$ and $k_a \ell^b = -1$. Then the projection to the 2-surface is provided by the projection vector $q_b^a = \delta_b^a + k^a \ell_b + k_b \ell^a$ and \mathcal{D}_a is the covariant derivative defined using q_{ab} . We thus get the final result for the Noether charge corresponding to the null congruence as

$$\ell_a J^a(\ell) = 2R_{ab} \ell^a \ell^b + \ell_a g^{ij} \mathcal{E}_\ell N_{ij}^a = \mathcal{D}_a(\kappa \ell^a) + \frac{d\kappa}{d\lambda} \quad (101)$$

If we integrate this over the null surface with the measure $d\lambda d^2x \sqrt{\sigma}$ and ignore the pure boundary contribution, we get :

$$\int d\Sigma_a J^a(\ell) = \int d\lambda d^2x \sqrt{\sigma} \ell_a J^a(\ell) = \int d\lambda d^2x \sqrt{\sigma} \frac{d\kappa}{d\lambda} \quad (102)$$

This result is analogous to our result in Eq. (55) and shows that—in the case of null surfaces—the Noether charge is again related to the ‘heating’ of the boundary surface because $\kappa \propto T$.

The real importance of this result lies in the fact that it can be used in the variational principle introduced earlier based on the null surfaces. To do this, we re-introduce the $16\pi G$ factor with $G = 1$ and re-write Eq. (101) as

$$-\frac{1}{8\pi}R_{ab}\ell^a\ell^b = \frac{1}{16\pi}\ell_a g^{ij}\mathfrak{L}_\ell N_{ij}^a - \frac{1}{16\pi}\left[\mathcal{D}_a(\kappa\ell^a) + \frac{d\kappa}{d\lambda}\right] \quad (103)$$

Adding the term $T_{ab}\ell^a\ell^b$ to both sides, integrating over a null surface with the measure $d\lambda d^2x\sqrt{\sigma}$ and ignoring the surface contributions, we find that our variational principle can be based on the functional:

$$\mathcal{Q} \equiv \int_{\lambda_1}^{\lambda_2} d\lambda d^2x \sqrt{\sigma} \left[\frac{1}{16\pi} \left(g^{ij}\ell_a \mathfrak{L}_\ell N_{ij}^a - \frac{d\kappa}{d\lambda} \right) + T_{ab}\ell^a\ell^b \right] \quad (104)$$

In other words, we can obtain the field equations with an arbitrary cosmological constant by varying ℓ_a (with the constraint $\ell^2 = 0$) in the above functional and assuming that variations vanish at the boundaries ($\lambda = \lambda_1, \lambda_2$). If we integrate the last pair of equations in Eq. (100) over the null surface with $\sqrt{\sigma}d^2x d\lambda$ and ignore boundary and end point contributions, we get the result:

$$\int_{\lambda_1}^{\lambda_2} d\lambda d^2x \sqrt{\sigma} \left[\frac{d\kappa}{d\lambda} + \kappa^2 \right] = 0 \quad (105)$$

Therefore in our variational principle we can replace $-(d\kappa/d\lambda)$ by κ^2 everywhere and obtain:

$$\begin{aligned} \mathcal{Q} &\equiv \int_{\lambda_1}^{\lambda_2} d\lambda d^2x \sqrt{\sigma} \left[\frac{1}{16\pi} \left(g^{ij}\ell_a \mathfrak{L}_\ell N_{ij}^a - \mathfrak{L}_\ell \kappa \right) + T_{ab}\ell^a\ell^b \right] \\ &= \int_{\lambda_1}^{\lambda_2} d\lambda d^2x \sqrt{\sigma} \left[\frac{1}{16\pi} \left(g^{ij}\ell_a \mathfrak{L}_\ell N_{ij}^a + \kappa^2 \right) + T_{ab}\ell^a\ell^b \right] \end{aligned} \quad (106)$$

Since $T_{ab}\ell^a\ell^b$ can be thought of as the heat (enthalpy) density $TS/V = Ts$ of matter, we can think of the rest as essentially the heat density of the null surface.

When ℓ_a is affinely parametrized with $\kappa = 0$ (which is a choice we can always make), then the variational principle can be based on the integral

$$Q_1 \equiv \int_{\lambda_1}^{\lambda_2} d\lambda \, d^2x \, \sqrt{\sigma} \left[\frac{1}{16\pi} g^{ij} \ell_a \mathbb{E}_\ell N_{ij}^a + T_{ab} \ell^a \ell^b \right] \quad (107)$$

This again shows that the quantity $g^{ij} \ell_a \mathbb{E}_\ell N_{ij}^a$ plays a vital role even in the derivation of the field equations from an alternative extremum principle. In fact, the field equations in the absence of matter can be obtained by extremising the expression

$$Q_1 \equiv \frac{1}{16\pi} \int_{\lambda_1}^{\lambda_2} d\lambda \, d^2x \, \sqrt{\sigma} \left[g^{ij} \ell_a \mathbb{E}_\ell N_{ij}^a \right] \quad (108)$$

over all null surfaces simultaneously. We have seen earlier ([23]; see Eq. (33)) that the integral of this term on a null surface has a very simple physical meaning in terms of the heat content of the null surface. So, at least for pure gravity, we can obtain a variational principle which has an interpretation in thermodynamic language.

7 Variations on the basic theme

So far we have worked with (i) the Einstein–Hilbert Lagrangian and (ii) the vector field $\xi^a = Nu^a$ to obtain our results (except in the last section in which we used ℓ^a). The ideas can be generalized in two obvious directions.

First, one can consider Lagrangians related to the Einstein–Hilbert Lagrangian by total divergences. There are, again, two obvious choices: First is obtained by using the so called Γ^2 Lagrangian which is obtained by adding a total divergence to cancel the surface term in Eq. (20) while the second is obtained by adding a $2K$ term at each boundary [72, 73]. This is equivalent to adding a term $\nabla_a(Ku^i)$ in the bulk (with similar terms for other timelike boundaries when required).

Second, one can consider vectors other than ξ^a to study the properties of the Noether charge, the Hamiltonian etc. with the obvious alternative being $\zeta^a = Nu^a + N^a$. The purpose of this section is to briefly describe what happens in these contexts.

7.1 Noether currents for related lagrangians

We begin by considering the Noether current associated with L_{quad} which, as we know is not a scalar. But when two Lagrangians differ by a total divergence, their infinitesimal variations will also differ by a total divergence. Therefore, even if one of them is not a scalar, it will still be possible to obtain a conserved current associated with it. The strategy is to obtain the form of $\mathbb{E}_q(\sqrt{-g} L_{\text{quad}})$ in two different ways—by explicit variation of Γ s and by using the functional variation with respect to the metric—and equate the two expressions. The first calculation is considerably simplified by writing

$\sqrt{-g} L_{\text{quad}} = \sqrt{-g} L_H - L_{\text{sur}}$. The only care that is needed is in the manipulation of the Lie derivatives of non-tensorial objects [discussed in Appendix 9.6]. The explicit variation leads to the result [see Appendix 9.6, Eq. (204)]:

$$\mathcal{L}_q (\sqrt{-g} L_{\text{quad}}) = \partial_a [\sqrt{-g} L_{\text{quad}} q^a] - \partial_a (\sqrt{-g} K^a) \quad (109)$$

where

$$K^a = g^{lm} \partial_l \partial_m q^a - g^{al} \partial_l \partial_m q^m \quad (110)$$

On the other hand, performing the variation of $\sqrt{-g} L_{\text{quad}}$ by treating it as a functional of the metric, we have the result:

$$\mathcal{L}_q (\sqrt{-g} L_{\text{quad}}) = \partial_a (-2\sqrt{-g} G_b^a q^b + N_{lm}^a \mathcal{L}_q f^{lm}) \quad (111)$$

Equating the two, we get the conservation law $\partial_a (\sqrt{-g} J_{\text{quad}}^a) = 0$ where the Noether current for the quadratic Lagrangian may be defined as:

$$\sqrt{-g} J_{\text{quad}}^a = 2\sqrt{-g} G_b^a q^b + \sqrt{-g} L_{\text{quad}} q^a - N_{lm}^a \mathcal{L}_q f^{lm} - \sqrt{-g} K^a \quad (112)$$

Using the standard result for the Noether current from Einstein–Hilbert action J_H^a this can be expressed as [see Appendix 9.6, Eq. (207)]:

$$J_{\text{quad}}^a = J_H^a + \nabla_b (V^{[a} q^{b]}) = J_H^a + \nabla_b (g^{lm} N_{lm}^{[b} q^{a]}) \quad (113)$$

where we have used the result $\sqrt{-g} V^a = -f^{lm} N_{lm}^a$. If we write $J_H^a \equiv J_{\text{quad}}^a + J_{\text{sur}}^a$, thereby defining a Noether current associated with L_{sur} , then:

$$J_{\text{sur}}^a = -\nabla_b (V^{[a} q^{b]}) = \nabla_b (V^{[b} q^{a]}) = \nabla_b J_{\text{sur}}^{ab}, \quad J_{\text{sur}}^{ab} \equiv V^{[b} q^{a]} = q^a V^b - q^b V^a \quad (114)$$

We see that the final result is remarkably simple [57, 61]: adding the divergence of an entity v^a leads to an extra term in the Noether potential which is $(v^{[b} q^{a]})$. To understand this, at least when v^a is a vector, let us consider the following question [61]: Suppose we have a scalar Lagrangian $L = \nabla_a W^a$ which is total divergence of a vector W^a so that the action is a pure surface term. What is the corresponding Noether current? We note that, treated as a scalar density,

$$\mathcal{L}_q (\sqrt{-g} L) = \sqrt{-g} \nabla_a (L q^a) = \partial_a (\sqrt{-g} L q^a) \quad (115)$$

On the other hand, expanding out the Lie derivative, we have:

$$\mathcal{L}_q (\sqrt{-g} L) = \mathcal{L}_q (\partial_a [\sqrt{-g} W^a]) = \partial_a [\mathcal{L}_q (\sqrt{-g} W^a)] \quad (116)$$

where we have used the fact the Lie derivatives commute with partial derivatives since they are local variations. Equating the two, we get a conservation law $\partial_a (\sqrt{-g} J^a) =$

0, where:

$$\begin{aligned}\sqrt{-g} J^a &= \sqrt{-g} L q^a - \mathfrak{L}_q (\sqrt{-g} W^a) \\ &= q^a \partial_b (\sqrt{-g} W^b) - \left[q^b \partial_b (\sqrt{-g} W^a) - \sqrt{-g} W^b \partial_b q^a + \sqrt{-g} W^a \partial_b q^b \right]\end{aligned}\quad (117)$$

The first and the third terms combine to give $\partial_b (\sqrt{-g} q^a W^b)$ while the second and the fourth combine to give $-\partial_b (\sqrt{-g} q^b W^a)$. So we get [61] the Noether current and potential to be:

$$\sqrt{-g} J^a = \partial_b (\sqrt{-g} q^{[a} W^{b]}) = \sqrt{-g} \nabla_b (q^{[a} W^{b]}); \quad J^{ab} = q^a W^b - q^b W^a \quad (118)$$

So if V^c was a vector we could have written down our result in Eq. (113) using this fact; but in our case V^c is not a vector. Nevertheless, we see that the non-vectorial part involving K^a appears only *additively* (because of the special form of V^c) and thus cancels in the evaluation of Eq. (117).

The result obtained above allows us to tackle the second case, namely the one in which we have a $2K$ term on the boundary, with ease. Since this term arises from integrating $\nabla_a (K u^a)$ in the bulk, it will lead to a Noether potential $J^{ab} = K (q^a u^b - q^b u^a)$ for the diffeomorphism along a vector field q^a . (With more terms of similar nature included if there are other, say, timelike, boundaries).

This fact—in turn—leads to a remarkable conclusion: *The extra term for the space-like boundary vanishes for our special choice of vector $q^a = \xi^a$.* Therefore, we could have thought of results in the previous sections, obtained for J_H^a , as holding also for the $R + 2K$ action. (This does not happen for $q^a = \zeta^a$).

The situation regarding L_{quad} is different. Here, the Noether potential is $J_{\text{sur}}^{ab}(u) = u^a V^b - u^b V^a$. While $2K = V^c u_c$ for a foliation with $N_\alpha = 0$ (see p.250 of [44]), in general V^c can have a component orthogonal to u^a which can contribute to $u^a V^b - u^b V^a$. This would bring in a nontrivial difference between V^c and K as far as Noether currents are concerned. Also, note that while the Noether potential for K term is generally covariant, the one for V^c is not.

7.2 Gravitational hamiltonian from L_{quad} and its variation

The use of L_{quad} suggests an associated definition for gravitational momentum flux, which is sometimes used in the literature. This expression, analogous to $p\dot{q} - L$ can be defined as

$$P_{\text{quad}}^a = N_{lm}^a \mathfrak{L}_q f^{lm} - q^a \sqrt{-g} L_{\text{quad}} \quad (119)$$

It can be motivated exactly the way we did for P_H^a (leading to Eq. (28), but using L_{quad} instead of L_H . (The choice of overall sign is conventional and we could have switched it, as we did earlier). We will not repeat this argument.

The obvious trouble which makes the argument and motivation rather weak in this case, of course, is that P_{quad}^a —in contrast to P_H^a which we worked with all along—is,

noncovariant due to the presence of N_{bc}^a and L_{quad} . In the literature, when it is used, one usually concentrates on its variation (which takes care of the noncovariance of L_{quad}) and performs a background subtraction on N_{bc}^a in its definition to make it covariant. These features make it somewhat less attractive than P_H^a , but we will briefly discuss it for the sake of completeness.

We can easily relate P_H^a and P_{quad}^a . Using the notation $\mathcal{V}^a = \sqrt{-g} V^a$, one can show that [see Appendix 9.7, Eq. (210)]:

$$\sqrt{-g} P_H^a = -P_{\text{quad}}^a - \left[\sqrt{-g} K^a + \partial_c (q^{[c} \mathcal{V}^{a]}) \right] \quad (120)$$

where K^a is defined by Eq. (110). This gives, on combining with the on-shell version of Eq. (85), the on-shell result:

$$\begin{aligned} \delta P_{\text{quad}}^a &= -\delta(\sqrt{-g} P_H^a) - \delta(\sqrt{-g} K^a) + \partial_c \left(q^{[c} \delta N_{lm}^a] f^{lm} + q^{[c} N_{lm}^a] \delta f^{lm} \right) \\ &= -\sqrt{-g} \omega^a - \delta(\sqrt{-g} K^a) + \partial_c \left(q^{[c} N_{lm}^a] \delta f^{lm} \right) \end{aligned} \quad (121)$$

The extra term $\delta(\sqrt{-g} K^a)$ arises because of the non-tensorial character of P_{quad}^a . As mentioned before, one way to avoid this is to redefine P_{quad}^a by changing N_{lm}^a to $N_{lm}^a - [N_{lm}^a]_{\text{flat}}$ in Eq. (119) and use the result in Appendix 9.6, [Eq. (197)] to make everything covariant. This will result in $\delta(\sqrt{-g} K^a)$ disappearing in the above expression at the price of requiring a background subtraction.

We can also obtain corresponding expressions for $\delta(\sqrt{-g} J_{\text{quad}}^a)$ by relating it to $\delta(\sqrt{-g} J_H^a)$ using Eq. (113). Then we get [see Appendix 9.7, Eq. (212)]

$$\delta(\sqrt{-g} J_{\text{quad}}^a) = \sqrt{-g} \omega^a + \partial_b \left(\delta f^{lm} N_{lm}^{[b} q^{a]} \right) + E^a \quad (122)$$

where the equations of motion terms are:

$$E^a \equiv 2q^b \delta(\sqrt{-g} G_b^a) + q^a \sqrt{-g} G_{ij} \delta g^{ij} \quad (123)$$

We have mentioned these results only for the sake of completeness and to show why P_H^a , J_H^a etc. are, in fact, the better mathematical constructs. We shall not pursue P_{quad}^a and J_{quad}^a because of their noncovariant nature and associated difficulties.

7.3 Noether charge for ζ^a

Another possible variation of our theme is to use the vector $\zeta^a = \xi^a + N^a$ rather than ξ^a as the time development vector. This is indeed a possible choice (because of the $\zeta^a \nabla_a t = 1$ condition) and, in fact, has been explored in detail in the past literature. We, however, find that the thermodynamic interpretation is clearer and simpler in terms of ξ^a rather than ζ^a , except when they coincide. Mathematically, the main reason for this is the fact that u_a is hypersurface orthogonal. Physically, observers traveling on the integral curves of ζ^a are boosted with respect to the fundamental observers

moving orthogonal to the hypersurface. This boost affects their view of physics in terms of modified acceleration, energy density measurements etc and the extra terms cloud the simple interpretation in terms of the fundamental observers. This is why I have not bothered to discuss this case in detail. However, given the utility of ζ^a in certain contexts and the popularity it enjoys in the literature, I will make some brief comments about the Noether charge associated with this vector.

Just to see the complications which arise, let us attempt a direct computation of the Noether charge for ζ^a . The, by now familiar, Noether current expression

$$u_a \nabla_b J^{ab}(\zeta) = 2R_b^a \zeta^b u_a + g^{ij} u_a \mathfrak{L}_\zeta N_{ij}^a \quad (124)$$

can be re-written in the form

$$g^{ij} u_a \mathfrak{L}_\zeta N_{ij}^a = D_\alpha (N J^{\alpha 0}) - 2R_b^a \zeta^b u_a \quad (125)$$

This requires us to compute $N J^{\alpha 0}$, the integral of which over the boundary will give the Noether charge. Direct computation in the adopted coordinate system gives a rather complicated result:

$$N J^{\alpha 0} = 2N a^\alpha - \frac{2}{N} N_\beta (D^\alpha N^\beta) + \frac{1}{N} h^{\alpha\beta} \partial_0 N_\beta \quad (126)$$

When we work in the gauge $g_{0\alpha} = N_\alpha = 0$ so that $N^a = 0$ and $\zeta^a = \xi^a$, we get back the previous results. Also note that the Komar energy measured by observers moving on the integral curves of ζ^a will involve the factor $R_{ab} \zeta^a \zeta^b$ and not $R_{ab} \zeta^a u^b$.

It is obvious that a more geometrical approach is required for the interpretation of the results in this context. From the standard Gauss theorem

$$\int_{\mathcal{R}} d^3x u_a J^a \sqrt{h} = \int_{\partial\mathcal{R}} d^2x \sqrt{\sigma} u_a r_b J^{ab} \quad (127)$$

we know that we need to compute the expression $r_a u_b J^{ba}$ to calculate the Noether charge. This is straightforward to do and we find that [see Appendix 9.2, Eq. (153)]:

$$r_a u_b J^{ba} = 2r_a \zeta^b \nabla_b u^a - (r_a \mathfrak{L}_\zeta u^a - u_a \mathfrak{L}_\zeta r^a) = 2r_a \zeta^b \nabla_b u^a - r_a \mathfrak{L}_\zeta u^a \quad (128)$$

where the last equality arises if we choose the foliation to satisfy the natural condition $u_a \mathfrak{L}_\zeta r^a = 0$. Of these two terms in Eq. (128), the first term can indeed be interpreted as arising due to an “acceleration” which measures the variation of the velocity u^a along the direction of ζ^a . This, in turn, has two components:

$$2A^i \equiv 2\zeta^b \nabla_b u^i = 2N a^i + 2N^b \nabla_b u^i \quad (129)$$

The first term is what we have been working with all along and is related to the standard acceleration. The second term arises from the spatial drift of the observers moving along the integral curves of ζ^a , with respect to the fundamental observers.

We can now go ahead and re-compute all the expressions which we originally calculated with ξ^a . We will find everywhere that there are extra terms arising from this boost. We will also encounter the following further difficulties in the interpretation:

- If we consider regions bounded by $N = \text{constant}$ surfaces, then the standard interpretation of the Davies–Unruh temperature requires the matching of the *magnitude* $a \equiv (a_i a^i)^{1/2}$ of the acceleration with its normal component so that one can write $r_\alpha a^\alpha = a$. (The Davies–Unruh temperature is associated with the magnitude of the acceleration, not with any given component; so we need $r_\alpha a^\alpha$ on the surface—which arises from the Gauss theorem, on integration—to be equal to a for the interpretation to work). This, obviously, will not work if we use A^i as the relevant acceleration and we again need to add boost dependent correction terms to the thermodynamic quantities to make the interpretation work.
- Second, in relating $\mathcal{L}P^{ab}$ to $\mathcal{L}N_{bc}^a$ to obtain, say, Eq. (65), we use the relation Eq. (162) of Appendix 9.3, which contains an extra spatial derivative term. This term identically vanishes for ξ^a but not for ζ^a . One needs to include this contribution in the holographic equipartition and elsewhere if we use ζ^a .
- We saw in Sect. 7.1 that the addition of the $2K$ term at the boundary does not change the Noether potential if we use ξ^a , but it picks up an additional contribution if we use ζ^a . So the results with the Einstein–Hilbert Lagrangian and the results for the Lagrangian with an additional $2K$ term will be different when we use ζ^a , but they are the same if we use ξ^a .

As I said before, all these are possible to handle, but I find the description based on ξ^a , a lot more natural and elegant.

8 Discussion

Since I have already summarized the results in Sect. 1 (which the reader is invited to glance through again now!) I will confine myself here to some comments about the overall perspective. This will be useful since I have attempted to gather together at least two distinct themes in this paper:

- Many results in standard general relativity, including the evolution equation for spacetime, can be given a thermodynamic interpretation (Sects. 3–5 and 7). In particular, one can interpret the time evolution as being driven by the departure from holographic equipartition.
- The standard formalism ignores the most important *observational fact* about gravitational dynamics and hence is bound to be incorrect. The thermodynamic perspective suggests an alternative paradigm which is closer to observational facts about gravity (Sect. 6).

The connecting thread between the two themes is the role played by the combination $f^{ab}\mathcal{L}_\xi N_{ab}^c$ throughout the paper. This quantity (1) reinforces the utility of the variables (f^{ab}, N_{ab}^c) (emphasized in [23]), (2) allows us to introduce gravity through a momentum space Lagrangian, (3) motivates the choice for the gravitational momentum flux, (4) relates to the surface term in the action principle, (5) connects up various

properties of the Noether current and (6) assumes the center-stage in the equation for the evolution of spacetime in terms of the departure from holographic equipartition. Obviously, one needs to understand the physical significance of this term from different angles (and find it a good name!).

We also found that the interpretations are simpler and more elegant when we use $\xi^a = Nu^a$ as our time development vector. This is a new feature since most of the discussion in the past literature has concentrated on $\zeta^a = \xi^a + N^a$. I hope to do a more detailed comparison in a future work to clarify their inter-relationship, but I believe ξ^a is the better choice. Using this vector, we could immediately interpret the Noether charge in the thermodynamic language as the surface heat energy (rather than as entropy). We also could clarify further the role played by the equipartition energy $2TS$ vis-a-vis the heat content TS , amplifying on the work in [24]. (As an off-shoot, we also know that, not only M but also $M/2$, have a clear physical meaning and there is no factor 2 problem in the Komar integrals!).

The *mathematics* behind many of the results works because of the following facts: (a) There is a well-known identity (see e.g., page 541 of [44]) for $R_{ab}u^au^b$ given by Eq. (155) which relates it to $\nabla_i a^i$ and the time evolution terms through K_{ab} or p^{ab} . If we take the field equations to be $R_{ab}u^au^b = (8\pi)\mathcal{F}_{ab}u^au^b$, and integrate this identity over all space, we can relate the time-evolution terms to the difference between the flux of acceleration and the Komar energy. (b) Further, the same combination $R_{ab}u^au^b$ occurs in the expression for the Noether charge density $u_i J^i$, allowing us to introduce the boundary variation term of the action (which occurs in the Noether charge) into the dynamics.

But this is an overly simplistic description of the *physics*! When one works out the expressions, there are several other structural features of gravity which combine together to produce these elegant results:

First is the fact that the Einstein–Hilbert action is a momentum space action, which is why we could introduce (f^{ab}, N_{ab}^c) and express the boundary variation term of the action—which occurs in the expression for the Noether current—in terms of the momentum flow.

Second, the idea almost works, but not quite—see the comments towards the end of Appendix 9.2—with u^a and we *need* to use $\xi^a = Nu^a$. This has to do with N being finite on real horizons where $N = 0$, in contrast to a . (It just does not work with ζ^a as nicely.)

Third, the flux of acceleration means nothing in classical general relativity and we need to introduce local Rindler observers and the Davies–Unruh temperature to bring in the thermodynamic perspective. This, of course, has been the corner-stone of the emergent gravity program all along. In addition to the *algebra* of rewriting $(a\sqrt{\sigma}/8\pi)$ as Ts , there is an important *conceptual* angle which appears in this approach. In the standard interpretation of gravity, a field equation like $\mathbf{G} = (8\pi)\mathbf{T}$, is a geometrical statement independent of any observer. But this equation has the same mathematical content as $R_{ab}u^au^b = (8\pi)\mathcal{F}_{ab}u^au^b$, if we insist that the latter should hold for all observers with four-velocities u^i . It is this equation $R_{ab}u^au^b = (8\pi)\mathcal{F}_{ab}u^au^b$ —or, more precisely, its left hand side—with an extra four-velocity field in it, that is amenable to the thermodynamic interpretation. This is because, by its very nature, the Davies–Unruh temperature has an observer dependence and one cannot use it to

interpret a purely geometrical equation like $\mathbf{G} = (8\pi)\mathbf{T}$. But when we demand the validity of the thermodynamic interpretation for *all* observers, we *are* demanding the validity of $R_{ab}u^a u^b = (8\pi)\mathcal{F}_{ab}u^a u^b$ for *all* observers, thereby leading to $\mathbf{G} = (8\pi)\mathbf{T}$. The algebra is the same, but the physics is quite different.

As regards the second theme of the paper, I have already presented detailed arguments as to why we have no hope of solving the cosmological constant problem in the conventional approach. Ignoring the *observed* fact that gravity does not respond to shifts in the matter Lagrangian by a constant is as bad as ignoring the principle of equivalence and trying to describe gravity. This fact also tells us that the cosmological constant problem exists even at the tree-level of quantum field theory and issues like the energy of the vacuum, etc. are red herrings. In the alternative approach, we again use a thermodynamically motivated variational principle (with $\int \mathcal{L}_\xi N_{ab}^c$ playing a crucial role), and demand the validity of $(R_{ab} - (8\pi)T_{ab})\ell^a \ell^b = 0$ for all null vectors. This is very similar in spirit to the idea of demanding the validity of $R_{ab}u^a u^b = (8\pi)\mathcal{F}_{ab}u^a u^b$ for all time-like unit vectors to reproduce standard general relativity. As I explained above, the thermodynamic interpretation requires such an approach. The complete story regarding the cosmological constant also requires a principle to fix its numerical value, which is described elsewhere [70,71].

Finally, let me comment on the broader ideas of the holographic principle and the emergent paradigm within the context of this work.

Let us first consider the idea of holography in the context of this work. I have shown in this paper that there is a clear correspondence between spacetime dynamics in a bulk region of space and the thermal physics at the boundary of this region. In static spacetimes, this correspondence leads to an *equality* between properly defined degrees of freedom in the bulk and on the surface, while—in the more general spacetime—the time evolution of the spacetime is related to the difference between the number of surface and bulk degrees of freedom. Such a correspondence between bulk and surface physics certainly deserves the adjective “holographic” when the term is used in the broad sense of its meaning, possibly close to the original idea [74]. In published literature, the term “holographic” is used in different contexts with different connotations, often with a more precise (but restricted) meaning in the context of AdS/CFT correspondence in string theory (see, for example, [75,76]). In this context, the degrees of freedom pertaining to two *different* theories living in the bulk and in the boundary of a region in AdS space show a correspondence. I stress that the usage in this paper is more general and, of course, has no direct correspondence with the usage in the AdS/CFT context.

The term ‘emergent’ is also used with very different connotations in the literature, even as regards the description of gravity (see, for e.g., [77]). For my purpose, I will divide the usage into three broad categories: (a) There are models (like, for example, analogue models) in which one attempts to introduce certain fundamental degrees of freedom and some sort of ‘environment’ and tries to obtain something analogous to gravity through the interaction between the degrees of freedom and the environment [78–81]. By and large, these type of models tend to be somewhat ad-hoc and it is difficult to judge them in the current backdrop of fundamental physics. (b) Quantum gravity models like string theory, loop quantum gravity etc. start with the dynamics

of something more fundamental than the metric tensor and try to obtain the spacetime itself as an emergent construct along with the metric tensor and the geometrical structure. These, of course, constitute the really fundamental approach to the problem but, for the same reason, are the most difficult to tackle. (c) The third category assumes the existence of spacetime, metric structure, etc. but tries to obtain the *dynamics* of the spacetime as an emergent phenomenon. This is clearly a “top-down” approach in length scales. Here, one probes the structure of classical gravity (with a single quantum mechanical input, viz., the Davies–Unruh temperature which brings in the Planck constant) and tries to discover certain generic properties of the microscopic physics. This “top-down” approach attempts to do something similar to what Boltzmann did, viz. to guess from the thermodynamical results that matter *must* have microscopic degrees of freedom and relate the notions of entropy, temperature etc. to these degrees of freedom. The “bottom-up” approach (string theory, LQG,), on the other hand, is analogous to deriving the thermodynamics of a system from its statistical mechanics. (There are also approaches like, for example, the ones based on causal sets [82] which do not belong to any of these categories but inherit different features from different categories).

The approach taken in this paper is along the lines of (c) above in which one assumes the existence of a spacetime manifold, etc. but obtains the dynamics as emergent. As shown in Sect. 6 one can indeed obtain the field equations from a variational principle which has a thermodynamic interpretation. I have given a list of several useful lessons one can learn from such a “top-down” approach elsewhere [17], and the hope is that a more fundamental “bottom-up” approach will connect up with this approach in the appropriate limit.

Acknowledgments I thank Bibhas Majhi, Krishnamohan Parattu and Dawood Kothawala for extensive discussions and comments on the earlier drafts. I thank Sunu Engineer for discussions. My work is partially supported by the J.C.Bose research grant of the Department of Science and Technology, Government of India.

9 Appendices: Calculational details

Several calculational details and background results are collected together in these appendices in order not to distract the flow of ideas in the main paper. Many of them exist in the literature but are compiled together for the sake of completeness. I have given a fair amount of details (and sometimes alternative derivations of the results) in the hope that these will be useful. Some of these results are easier to obtain in the index-free differential form language but I have presented everything in the more familiar index language.

9.1 An identity relating the noether currents

It is obvious that $J^{ab}[q] = 0$ identically, if $q_a = \nabla\phi$ is a pure gradient. Very often, we will need to find the Noether current for vectors of the form $v_a = f(x)\nabla_a\phi$. With this motivation, we will first prove a general relation between Noether currents for

two vector fields q^a and $v^a \equiv f(x)q^a$. Computing the expanded form of

$$J^a(v) = \nabla_b J^{ab}(v) = \nabla_b \left[\nabla^a(fq^b) - \nabla^b(fq^a) \right] \quad (130)$$

one can easily show that

$$\begin{aligned} J^a(v) &= \nabla_b J^{ab}(v) = \nabla_b [f J^{ab}(q) + q^b f^a - q^a f^b] = f J^a(q) + J^{ab}(q) f_b + \nabla_b A^{ba} \\ &= 2R_b^a f q^b + g^{ij} \mathfrak{L}_v N_{ij}^a = g^{ij} \mathfrak{L}_v N_{ij}^a + f(J^a(q) - g^{ij} \mathfrak{L}_q N_{ij}^a) \end{aligned} \quad (131)$$

where we use the notation $f_a \equiv \nabla_a f$ and $A_{ij} \equiv q_i f_j - q_j f_i$. This allows us to obtain the relation we need, viz.,

$$J^a(fq) - f J^a(q) = g^{ij} \mathfrak{L}_{fq} N_{ij}^a - f g^{ij} \mathfrak{L}_q N_{ij}^a = f_b J^{ab}(q) + \nabla_b A^{ba} \quad (132)$$

We take the dot product of this equation with q_a to obtain

$$q_a J^a(fq) - f q_a J^a(q) = q_a f_b J^{ab}(q) + q_a \nabla_b A^{ba} \quad (133)$$

Writing the last term as

$$\begin{aligned} q_a \nabla_b A^{ba} &= \nabla_b (q_a A^{ba}) - \frac{1}{2} A^{ba} J_{ba}(q) = \nabla_b (q_a A^{ba}) - \frac{1}{2} 2(q^b f^a) J_{ba}(q) \\ &= \nabla_b (q_a A^{ba}) - (q^b f^a) J_{ba}(q) \end{aligned} \quad (134)$$

we see that the last term in Eq. (134) cancels with first term in the right hand side of Eq. (133). Therefore, using the definition $A_{ij} \equiv q_i f_j - q_j f_i$ to simplify $\nabla_b (q_a A^{ba})$, we get:

$$q_a J^a(fq) - f q_a J^a(q) = \nabla_b (q_a A^{ba}) = \nabla_b \left([q^a q^b - q^2 g^{ab}] \nabla_a f \right) \quad (135)$$

In other words, we have the simple result

$$q_a J^a(fq) - f q_a J^a(q) = \begin{cases} -\nabla_b (q^2 \mathcal{P}^{ab}(q) \nabla_a f) & \text{when } q^2 \neq 0 \\ +\nabla_b (q^b q^a \nabla_a f) & \text{when } q^2 = 0 \end{cases} \quad (136)$$

where $\mathcal{P}^{ab}(q) = g^{ab} - (q^a q^b / q^2)$ is the projection tensor orthogonal to q^a . This result provides one way of computing the Noether current for vector fields.

9.2 Noether currents for u^a and ξ^a

In the paper, we extensively used the Noether current associated with the vector $\xi_a = Nu_a = -N^2 \nabla_a t$. Here we will compute the Noether current for *both* the vector fields

$u_a = -N\nabla_a t$ and $\xi_a = -N^2\nabla_a t$ because their comparison shows some structural differences which are of interest (and also because they can be computed without extra effort!). To calculate these currents, it is convenient to use the identity in Eq. (136). We start with $u_a = -N\nabla_a t$ and use Eq. (136) with $q_a = -u_a/N$, $f = -NJ^a[q] = 0$ to get

$$-\frac{u_a}{N}J^a(u) = -\nabla_b\left(-\frac{1}{N^2}h^{ab}(-\nabla_a N)\right) = -\nabla_b(a^b/N) \quad (137)$$

where $h^{ab} = g^{ab} + u^a u^b$ is the spatial projection tensor and $h^{ab}\nabla_a N = Na^b$ where a^b is the acceleration of u^a . To verify that this is indeed the acceleration, even in a general, time-dependent situation, let us compute it directly from the definition $a_k = u^l\nabla_l u_k$. We get:

$$a_k = u^l\nabla_l(-N\nabla_k t) = \left(u^l\nabla_l N\right)\frac{u_k}{N} + Nu^l\nabla_k\left(\frac{u_l}{N}\right) = u_k u^l\frac{\nabla_l N}{N} + \frac{\nabla_k N}{N} = h_k^l\frac{\nabla_l N}{N} \quad (138)$$

where we have used $\nabla_k\nabla_l t = \nabla_l\nabla_k t$ to arrive at the second term in the second equality. This shows that $Na_i = h_i^j\nabla_j N$ which is a useful result. Expanding out $\nabla_b(a^b/N)$ in Eq. (137), we get

$$\frac{u^a}{N}J^a(u) = \nabla_b\left(a^b/N\right) = \frac{1}{N}\nabla_b a^b - \frac{1}{N^2}\left(Na^2\right) = \frac{1}{N}\left(\nabla_i a^i - a^2\right) \quad (139)$$

To proceed further, we note that, if v^i is any spatial vector that satisfies the condition $u_i v^i = 0$ and D_i is the covariant derivative on the $t = \text{constant}$ surface, we have the result that

$$D_i v^i = D_\alpha v^\alpha \equiv \left(g^{ij} + u^i u^j\right)\nabla_i v_j = \nabla_i v^i - v_j a^j \quad (140)$$

(Note that, in the adopted coordinates we must have $v^0 = 0$, since $u_i = -N\delta_i^0$.) Applying this to $v^i = a^i$, we have

$$D_i a^i = D_\alpha a^\alpha = \nabla_i a^i - a^2 \quad (141)$$

Using Eq. (141) in Eq. (139), we get the Noether charge density to be:

$$u_a J^a[u] = D_\alpha a^\alpha \quad (142)$$

From the standard expression for the Noether current we also have:

$$D_i a^i = 2R_{ab}u^a u^b + g^{ij}u_a \mathcal{E}_u N_{ij}^a \quad (143)$$

Let us next consider the Noether current for ξ^a . Using again Eq. (136) with, say, $q^a = u^a$, $f = N$, we find that

$$u_a J^a(Nu) - Nu_a J^a(u) = \nabla_b(h^{ab}\nabla_a N) = \nabla_b(Na^b) \quad (144)$$

Using Eq. (142) we get:

$$\begin{aligned} u_a J^a(\xi) &= N u_a J^a(u) + \nabla_j (N a^j) = N (D_\alpha a^\alpha + \nabla_j a^j + a^2) \\ &= 2N \nabla_j a^j = 2D_\alpha (N a^\alpha) \end{aligned} \quad (145)$$

which is the final result we are after. Of course, in this particular case, one can also obtain this result by direct computation without using Eq. (136). The direct proof of this relation will proceed as follows:

$$u_b J^b(\xi) = \nabla_i \left(u_b \left(\nabla^b \xi^i - \nabla^i \xi^b \right) \right) - \frac{1}{2} (\nabla_i u_b - \nabla_b u_i) J^{bi} \quad (146)$$

$$\begin{aligned} &= \nabla_i \left[u_b (N \nabla^b u^i + u^i \nabla^b N - (i \leftrightarrow b)) \right] \\ &\quad - (a_i u_b - a_b u_i) \left[(\nabla^b N) u^i + N \nabla^b u^i \right] \end{aligned} \quad (147)$$

$$= \nabla_i \left[N a^i + u^i u^b \nabla_b N + \nabla^i N \right] - N a^2 - a_b \nabla^b N \quad (148)$$

$$= \nabla_i \left[N a^i + (g^{ib} + u^i u^b) \nabla_b N \right] - N a^2 - a_b \nabla^b N \quad (149)$$

$$= \nabla_i (2N a^i) - N a^2 - a_b \nabla^b N = 2N \nabla_i a^i + a^i \nabla_i N - N a^2 \quad (150)$$

where we have used $\nabla_{[i} u_{b]} = a_{[i} u_{b]}$ to get Eq. (147). Note that the last two terms, when combined as follows, will cancel each other:

$$\begin{aligned} N a^i [\nabla_i \ln N - a_i] &= h^{ij} \nabla_j N [\nabla_i \ln N - a_i] \\ &= (\nabla_j N) \left[\frac{1}{N} h^{ij} \nabla_i N - h^{ij} a_i \right] = 0 \end{aligned} \quad (151)$$

From Eq. (141), we also have

$$\frac{1}{2N} D_\alpha (2N a^\alpha) = D_\alpha a^\alpha + a^\alpha D_\alpha (\ln N) = D_\alpha a^\alpha + a^2 = \nabla_i a^i \quad (152)$$

so that $D_\alpha (2N a^\alpha) = 2N \nabla_i a^i$. This gives the result we are after.

Comparing with the Noether charge densities for u^i (in Eq. (142)) and ξ^a (in Eq. (145)) we see that a^α has been replaced by $2N a^\alpha$ which turns out to be important for the thermodynamic interpretation for two reasons. (i) The factor 2 is crucial in giving the correct expression for the equipartition result through Eq. (55). (ii) More importantly, the combination $N a$ leads to a finite result on the horizon—leading to standard surface gravity—while the magnitude a diverges on an $N = 0$ surface on static spacetimes. This is also seen from the fact that, the correct local temperatures are always defined with the Tolman redshift factor N .

Finally, we give the steps involved in proving the expression for the Noether charge associated with ζ^a which we needed in Sect. 7.3. It is slightly more convenient to expand the right hand side of Eq. (128) and show that it is equal to the left hand side.

The computation is straightforward:

$$\begin{aligned}
 2r_a \zeta^b \nabla_b u^a - (r_a \xi_\zeta u^a - u_a \xi_\zeta r^a) &= 2r_a \zeta^b \nabla_b u^a - r_a (\zeta^b \nabla_b u^a - u^b \nabla_b \zeta^a) \\
 &\quad + u_a (\zeta^b \nabla_b r^a - r^b \nabla_b \zeta^a) \\
 &= r_a u_b \nabla^b \zeta^a - u_a r_b \nabla^b \zeta^a = r_a u_b (\nabla^b \zeta^a - \nabla^a \zeta^b) \\
 &= r_a u_b J^{ba} \quad (153)
 \end{aligned}$$

where we have used the result $r_a u^a = 0$ to obtain the second equality. This is the result quoted in the text.

9.3 Different expressions for $u_a g^{ij} \xi_\xi N_{ij}^a$

We will obtain here the two other equivalent expressions for $u_a g^{ij} \xi_\xi N_{ij}^a$ used in the text. Using $D_\alpha(2Na^\alpha) = 2N \nabla_i a^i$ in Eq. (56) we get

$$u_a g^{ij} \xi_\xi N_{ij}^a = 2N(\nabla_i a^i - R_{ab} u^a u^b) \quad (154)$$

On the other hand, from the standard identity for $R_{ab} u^a u^b$ (see e.g., page 541 of [44]) we have:

$$R_{ab} u^a u^b = u^a (\nabla_l \nabla_a u^l - \nabla_a \nabla_l u^l) = \nabla_l a^l - K_l^a K_a^l + \nabla_a (K u^a) + K^2 \quad (155)$$

where $K_{ij} = -\nabla_i u_j - u_i a_j$ is the extrinsic curvature. So

$$\nabla_i a^i - R_{ab} u^a u^b = K_{ij} K^{ij} - K^2 - \nabla_a (K u^a) \quad (156)$$

Substituting into Eq. (154) we get:

$$\begin{aligned}
 \frac{1}{2N} u_a g^{ij} \xi_\xi N_{ij}^a &= \nabla_i a^i - R_{ab} u^a u^b = K_{ij} K^{ij} - K^2 - \nabla_a (K u^a) \\
 &= K_{ij} K^{ij} - u^a \nabla_a K \quad (157)
 \end{aligned}$$

This allows a simple physical interpretation for the combination $(K_{ij} K^{ij} - K^2)$ in terms of the Lie derivative of N_{bc}^a .

This result can also be obtained directly from the definitions of various quantities. To do this, we first note that, from the definition of extrinsic curvature, $K_{lm} = -\nabla_l u_m - a_m u_l$, we have the result

$$K_{lm} K^{ml} = (\nabla_l u_m + a_m u_l) (\nabla^m u^l + a^l u^m) = (\nabla_l u_m) (\nabla^m u^l) \quad (158)$$

Further, using the notation $\alpha_i \equiv \nabla_i \ln N$, we can write:

$$\begin{aligned}\nabla_i(Nu_j) &= N\alpha_i u_j + N(-K_{ij} - u_i a_j) = -NK_{ij} + N(\alpha_i u_j - u_i a_j) \\ &= -NK_{ij} + N(\alpha_i u_j - h_j^k \alpha_k u_i) \\ &= -NK_{ij} + N(\alpha_i u_j - \alpha_j u_i) - u_i u_j u^k \nabla_k N\end{aligned}\quad (159)$$

Taking the symmetric and anti-symmetric parts of this expression we get

$$J_{ab}(\xi) = 2N(\alpha_a u_b - \alpha_b u_a); \quad S_{ab}(\xi) = -2N(K_{ab} + u_a u_b u^k \alpha_k) \quad (160)$$

The rest of the calculation proceeds directly as follows:

$$\begin{aligned}u_b \left(\nabla^b S - \nabla_i S^{bi} \right) &= \nabla_b \left(S u^b \right) - \nabla_i \left(u_b S^{bi} \right) - \left(\nabla_b u^b \right) S + \left(\nabla_i u_b \right) S^{bi} \\ &= \nabla_b \left(u^b \left(-2NK + 2Nu^k \alpha_k \right) \right) - \nabla_i \left(2Nu^i u^k \alpha_k \right) \\ &\quad + K \left(-2NK + 2Nu^k \alpha_k \right) + S^{bi} \left(-K_{ib} - a_b u_i \right) \\ &= -2\nabla_b \left(K N u^b \right) - 2N \left(K^2 - K u^k \alpha_k \right) \\ &\quad + 2N \left(K_{ib} + a_b u_i \right) \left(K^{bi} + u^b u^i u^j \alpha_j \right) \\ &= -2\nabla_b \left(N K u^b \right) + 2N \left(K_{ib} K^{bi} - K^2 \right) + 2N K u^k \alpha_k \\ &= 2N \left(K_{ij} K^{ji} - K^2 \right) - 2K N u^b \alpha_k - 2N \nabla_b \left(K u^b \right) + 2K N u^b \alpha_k \\ &= 2N \left[K_{ij} K^{ji} - K^2 - \nabla_b (K u^b) \right] = 2N \left[K_{ij} K^{ji} - u^b \nabla_b K \right]\end{aligned}\quad (161)$$

which agrees with Eq. (157) when we use the fact that the left hand side is $u_a g^{ij} \mathfrak{L}_\xi N_{ij}^a$.

In the text we also related $u_a g^{ij} \mathfrak{L}_\xi N_{ij}^a$ to $h_{ab} \mathfrak{L}_\xi p^{ab}$ where $p^{ab} \equiv -\sqrt{h}(K^{ab} - h^{ab}K)$ is the momentum conjugate to h_{ab} . This result can be obtained as follows. We start with the standard result (see page 550 of [44]) used in the variation of the action functional with the $2K$ term:

$$-g^{ab} u_m \delta N_{ab}^m = 2\delta K + K^{ab} \delta h_{ab} + D_a (h_b^a \delta u^b) \quad (162)$$

which expresses the variational term that appears in the Einstein–Hilbert action in terms of the variation of the extrinsic curvature. The troublesome term is the one involving the spatial derivative $D_a (h_b^a \delta u^b)$, which, fortunately, vanishes when the variation is due to a diffeomorphism along ξ^a . This is easy to see from:

$$D_a \left(h_b^a \mathfrak{L}_\xi u^b \right) = D_a \left(h_b^a \mathfrak{L}_\xi \left(\frac{\xi^b}{N} \right) \right) = D_a \left(h_b^a \left[\left(\frac{1}{N} \right) \mathfrak{L}_\xi (\xi^b) + \xi^b \mathfrak{L}_\xi \left(\frac{1}{N} \right) \right] \right) = 0 \quad (163)$$

because $\mathfrak{L}_\xi(\xi^b) = 0$ and $h_b^a \xi^b = 0$. As regards the rest of the terms in Eq. (162), an explicit calculation of $h_{ab} \delta p^{ab}$ gives

$$\begin{aligned} h_{ab} \delta \left[(h^{ab} K - K^{ab}) \sqrt{h} \right] &= K \sqrt{h} h^{ab} \delta h_{ab} + h_{ab} \sqrt{h} (-\delta K^{ab} + K \delta h^{ab} + h^{ab} \delta K) \\ &= 3\sqrt{h} (h_{ab} \delta K^{ab} + K^{ab} \delta h_{ab}) - \sqrt{h} h_{ab} \delta K^{ab} \\ &= 2\sqrt{h} h_{ab} \delta K^{ab} + 3\sqrt{h} K^{ab} \delta h_{ab} \\ &= 2\sqrt{h} \delta K + \sqrt{h} K^{ab} \delta h_{ab} \\ &= \sqrt{h} \left\{ 2\delta K + K^{ab} \delta h_{ab} \right\} \end{aligned} \quad (164)$$

It follows that:

$$\sqrt{h} u_c g^{ik} \mathfrak{L}_\xi N_{ik}^c = -h_{ab} \mathfrak{L}_\xi p^{ab} \quad (165)$$

This result does not hold for the diffeomorphism along the vector ζ^a and one will obtain an extra surface contribution in that case.

For the variations arising from the diffeomorphism along ξ^a , it is also possible to prove the result in Eq. (165) directly (without using Eq. (162)) as follows:

$$\begin{aligned} h_{ab} \mathfrak{L}_\xi p^{ab} &= \mathfrak{L}_\xi \left(h_{ab} \sqrt{h} (K h^{ab} - K^{ab}) \right) - p^{ab} \mathfrak{L}_\xi h_{ab} \\ &= \mathfrak{L}_\xi (\sqrt{h} 2K) - p^{ab} (-2N K_{ab}) \\ &= 2\sqrt{h} N u^a \partial_a K + 2K \frac{1}{2} \sqrt{h} h^{ab} (-2N K_{ab}) + 2N p^{ab} K_{ab} \\ &= 2\sqrt{h} N u^a \partial_a K - 2N \sqrt{h} K^2 + 2N \sqrt{h} (K^2 - K_{ab} K^{ab}) \\ &= 2N \sqrt{h} (u^a \partial_a K - K_{ab} K^{ab}) \end{aligned} \quad (166)$$

In the first equality, we have used the definition of $p_{ab} = \sqrt{h}(K h_{ab} - K_{ab})$; to get the second equality, we have used $\mathfrak{L}_\xi h_{ab} = -2N K_{ab}$. This result, when combined with Eq. (157), leads to the desired expression. We can therefore write the Noether current expression as

$$\sqrt{h} u_a g^{ij} \mathfrak{L}_\xi N_{ij}^a = \sqrt{h} D_\alpha (2N a^\alpha) - 2N \sqrt{h} R_{ab} u^a u^b = -h_{ab} (\mathfrak{L}_\xi p^{ab}) \quad (167)$$

This was the result used in the text.

9.4 Variation of the gravitational momentum flux P_H^a

We will derive the expression for $\delta(\sqrt{-g} P_H^a)$, discussed in the text, from first principles. We start with the definition for P^a based on the Einstein–Hilbert action for an arbitrary but fixed q^a :

$$P_H^a \equiv L_H q^a + g^{lm} \mathfrak{L}_q N_{lm}^a = J_H^a - 2G_b^a q^b \quad (168)$$

We next introduce an arbitrary variation of the metric, keeping q^a fixed. (The q_a will change but we will not need it.). The variation of $\sqrt{-g} P_H^a$ gives

$$\begin{aligned}\delta(\sqrt{-g} P_H^a) &= q^a \delta(R\sqrt{-g}) + \delta(f^{lm} \mathbb{E}_q N_{lm}^a) \\ &= q^a (R_{ij} \delta f^{ij} - \partial_c (f^{lm} \delta N_{lm}^c)) + \delta(f^{lm} \mathbb{E}_q N_{lm}^a) + \mathbb{E}_q (f^{lm} \delta N_{lm}^a) \\ &\quad - \mathbb{E}_q (f^{lm} \delta N_{lm}^a) \end{aligned} \quad (169)$$

where, in the second line, we have explicitly added and subtracted $\mathbb{E}_q (f^{lm} \delta N_{lm}^a)$ which is a trick to get the symplectic structure. This gives

$$\begin{aligned}\delta(\sqrt{-g} P_H^a) &= \delta(f^{lm} \mathbb{E}_q N_{lm}^a) - \mathbb{E}_q (f^{lm} \delta N_{lm}^a) + q^a (R_{ij} \delta f^{ij}) + \mathbb{E}_q (f^{lm} \delta N_{lm}^a) \\ &\quad - q^a \partial_c (f^{lm} \delta N_{lm}^c) \end{aligned} \quad (170)$$

The combination $Q^a \equiv f^{lm} \delta N_{lm}^a = \sqrt{-g} g^{lm} \delta N_{lm}^a$ which appears here is a tensor density of weight one and hence

$$\begin{aligned}\mathbb{E}_q (f^{lm} \delta N_{lm}^a) - q^a \partial_c (f^{lm} \delta N_{lm}^c) &= \mathbb{E}_q Q^a - q^a \partial_c Q^c = \partial_c (q^c Q^a - q^a Q^c) \\ &= \partial_c (f^{lm} \delta N_{lm}^a q^c) \end{aligned} \quad (171)$$

Therefore,

$$\begin{aligned}\delta(\sqrt{-g} P_H^a) - q^a R_{ij} \delta f^{ij} &= \delta(f^{lm} \mathbb{E}_q N_{lm}^a) - \mathbb{E}_q (f^{lm} \delta N_{lm}^a) + \partial_c (f^{lm} \delta N_{lm}^a q^c) \\ &= \delta f^{lm} \mathbb{E}_q N_{lm}^a - (\mathbb{E}_q f^{lm}) \delta N_{lm}^a + \partial_c (f^{lm} \delta N_{lm}^a q^c) \end{aligned} \quad (172)$$

where we have used the fact that δ and \mathbb{E}_q operations on N_{lm}^a commute when $\delta q^a = 0$ (see Appendix 9.6). Defining the symplectic form

$$\sqrt{-g} \omega^a(\delta, \mathbb{E}_q) \equiv \delta f^{lm} \mathbb{E}_q N_{lm}^a - (\mathbb{E}_q f^{lm}) \delta N_{lm}^a \quad (173)$$

we can write Eq. (172) as:

$$\delta(\sqrt{-g} P_H^a) - q^a R_{ij} \delta f^{ij} = \sqrt{-g} \omega^a + \partial_c (f^{lm} \delta N_{lm}^a q^c) \quad (174)$$

If we consider variations with the background being on-shell (i.e., $G_{ab} = 0$ but $\delta G_{ab} \neq 0$), then the second term on the left hand side of Eq. (174) vanishes and we get

$$\delta(\sqrt{-g} P_H^a) = \sqrt{-g} \omega^a + \partial_c (f^{lm} \delta N_{lm}^a q^c) \quad (175)$$

This is a completely covariant equation because P_H^a is a covariant object. On using Eq. (168), we also find:

$$\delta \left(\sqrt{-g} [J_H^a - 2G_b^a q^b] \right) = \sqrt{-g} \omega^a + \partial_c \left(f^{lm} \delta N_{lm}^{[a} q^{c]} \right) + q^a R_{ij} \delta f^{ij} \quad (176)$$

A slightly different, but equivalent, expression is:

$$\begin{aligned} \partial_b \left\{ \delta \left(\sqrt{-g} J_H^{ab} \right) - \sqrt{-g} g^{lm} \delta N_{lm}^{[a} q^{b]} \right\} &= \sqrt{-g} \omega^a + 2\delta \left(G_b^a q^b \sqrt{-g} \right) \\ &+ q^a \sqrt{-g} G_{lb} \delta g^{lb} \end{aligned} \quad (177)$$

which is valid off-shell. If we now assume that the original spacetime is on-shell with $G_{ab} = 0$ but $\delta G_{ab} \neq 0$, then:

$$\partial_b \left\{ \delta \left(\sqrt{-g} J_H^{ab} \right) - \sqrt{-g} g^{lm} \delta N_{lm}^{[a} q^{b]} \right\} = \sqrt{-g} \omega^a + 2\delta \left(G_b^a q^b \sqrt{-g} \right) \quad (178)$$

We next compute the Lie derivative of the gravitational energy density. To do this from first principles, we again start from the definition of P_H^a in Eq. (168) (for $q^a = \xi^a$) and obtain from it the result

$$\sqrt{h} u_a P_H^a[\xi] = -t_a f^{lm} \xi_\xi N_{lm}^a - \sqrt{-g} L_H \quad (179)$$

where $t_a \equiv -(u_a/N) = \delta_a^0$ in the adopted coordinates. The variation of this expression, on using the standard result for $\delta(\sqrt{-g} L_H)$, gives

$$\delta \left(\sqrt{h} u_a P_H^a \right) = -t_a \delta \left(f^{lm} \xi_\xi N_{lm}^a \right) - \sqrt{-g} G_{ab} \delta g^{ab} + \partial_c \left(f^{lm} \delta N_{lm}^c \right) \quad (180)$$

Using the usual trick for obtaining the symplectic structure and introducing a factor $t_a \xi^a = 1$ in one of the terms, we get

$$\begin{aligned} \delta \left(\sqrt{h} u_a P_H^a \right) &= -t_a \left[\delta \left(f^{lm} \xi_\xi N_{lm}^a \right) - \xi_\xi \left(f^{lm} \delta N_{lm}^a \right) \right] - t_a \xi_\xi \left(f^{lm} \delta N_{lm}^a \right) \\ &+ (t_a \xi^a) \partial_c \left(f^{lm} \delta N_{lm}^c \right) - \sqrt{-g} G_{ab} \delta g^{ab} \end{aligned} \quad (181)$$

Defining the symplectic form ω^a as in Eq. (173) and using the fact that, for any tensor density Q^a we have the result $\xi_\xi Q^a - \xi^a \partial_c Q^c = \partial_c (\xi^{[c} Q^{a]})$, we get

$$\delta \left(\sqrt{h} u_a P_H^a \right) + R_{ab} \delta f^{ab} = \sqrt{h} u_a \omega^a - \partial_c \left[u_a f^{lm} \delta N_{lm}^{[c} u^{a]} \right] \quad (182)$$

Simplifying the second term on the right hand side, we find that $u_a f^{lm} \delta N_{lm}^{[c} u^{a]} = -h_a^c f^{lm} \delta N_{lm}^a$ leading to the result:

$$\delta \left(\sqrt{h} u_a P_H^a \right) + R_{ab} \delta f^{ab} = \sqrt{h} u_a \omega^a + \partial_c \left[h_a^c f^{lm} \delta N_{lm}^a \right] \quad (183)$$

Further, using $R_{ab}\delta f^{ab} = \sqrt{-g} G_{ab}\delta g^{ab} = -\sqrt{-g} G_{ab}\xi_\xi g^{ab}$ and the Bianchi identities, we can simplify the second term on the left. These together lead to the result

$$\xi_\xi \left(\sqrt{h} u_a P_H^a(\xi) \right) = \partial_c \left[\left(2G_b^c \xi^b + h_a^c g^{lm} \xi_\xi N_{lm}^a \right) \sqrt{-g} \right] \quad (184)$$

This is the result used earlier.

9.5 Variational formulation based on the null surfaces

We used in the text the following result:

$$\begin{aligned} R_{ab}\ell^a\ell^b &= \ell^j(\nabla_i\nabla_j - \nabla_j\nabla_i)\ell^i = \nabla_i(\kappa\ell^i) - \nabla_i\ell^j\nabla_j\ell^i - \nabla_j(\ell^j(\kappa + \theta)) + (\nabla_i\ell^i)^2 \\ &= -\nabla_i(\theta\ell^i) - [\nabla_i\ell^j\nabla_j\ell^i - (\nabla_i\ell^i)^2] \equiv -\nabla_i(\theta\ell^i) - \mathcal{S} \end{aligned} \quad (185)$$

where we have used the standard result $\nabla_i\ell^i = \theta + \kappa$ (with κ defined through $\ell^i\nabla_i\ell^j = \kappa\ell^j$) and defined \mathcal{S} by,

$$\mathcal{S} \equiv [\nabla_i\ell^j\nabla_j\ell^i - (\nabla_i\ell^i)^2] \quad (186)$$

When we integrate expressions over a null surface with the measure $d\lambda d^2x\sqrt{\sigma}$, we can ignore terms of the kind $\nabla_i(\phi\ell^i)$ for any scalar ϕ since they produce only boundary contributions. To see this we only need to note that

$$\nabla_i(\phi\ell^i) = \frac{d\phi}{d\lambda} + \phi \frac{d}{d\lambda}(\ln\sqrt{\sigma}) = \frac{1}{\sqrt{\sigma}} \frac{d}{d\lambda}(\sqrt{\sigma}\phi) \quad (187)$$

where we have used $\ell^i\nabla_i = d/d\lambda$ and $\nabla_i\ell^i = d(\ln\sqrt{\sigma})/d\lambda$. So the integral over the null surface of the first term in Eq. (185) is given by

$$\int_{\lambda_1}^{\lambda_2} d\lambda d^2x \sqrt{\sigma} \nabla_i(\theta\ell^i) = \int_{\lambda_1}^{\lambda_2} d^2x \sqrt{\sigma} \theta \Big|_{\lambda_1}^{\lambda_2} \quad (188)$$

which is just a boundary contribution which can be ignored. We therefore conclude that

$$\int_{\lambda_1}^{\lambda_2} d\lambda d^2x \sqrt{\sigma} [-2R_{ab} + T_{ab}] \ell^a\ell^b = \int_{\lambda_1}^{\lambda_2} d\lambda d^2x \sqrt{\sigma} [2\mathcal{S} + T_{ab}\ell^a\ell^b] \quad (189)$$

when we consider variations of the null vectors ℓ^a which vanish at the boundaries ($\lambda = \lambda_1, \lambda_2$). It follows from the arguments given in the text that the field equations can be obtained by varying ℓ^a in the expression:

$$Q \equiv \int_{\lambda_1}^{\lambda_2} \frac{d\lambda}{16\pi} \frac{d^2x}{\sqrt{\sigma}} [2S + 16\pi T_{ab} \ell^a \ell^b] \quad (190)$$

where we have reintroduced $16\pi G$ with $G = 1$.

We next compute the Noether current for the null vector ℓ^a used in the text. Let ℓ_a be a null congruence defining a null surface which may not be affinely parametrized. If we take $\ell_a = A(x) \nabla_a B(x)$, then $\ell^i \nabla_i \ell_j = \kappa \ell_j$ where $\kappa = \nabla_i A \nabla^i B = \ell^a \nabla_a \ln A$. We can compute the Noether current for ℓ_a using our Eq. (136) and noting that the Noether current for $q_a = \ell_a/A$ is zero. This gives

$$\begin{aligned} \frac{\ell_a}{A} J^a(\ell) &= \nabla_b \left(\frac{\ell^b}{A} \frac{\ell^a}{A} \nabla_a A \right) = \nabla_b \left(\frac{\ell^b}{A} \kappa \right) = \nabla_b (\kappa \ell^b) \frac{1}{A} - \kappa \ell^b \frac{1}{A^2} \nabla_b A \\ &= \frac{1}{A} \left\{ \nabla_b (\kappa \ell^b) - \kappa^2 \right\} \end{aligned} \quad (191)$$

It follows that

$$\ell_a J^a(\ell) = \nabla_b (\kappa \ell^b) - \kappa^2 = \ell^b \nabla_b \kappa + \kappa^2 + \kappa \theta - \kappa^2 = \frac{d\kappa}{d\lambda} + \theta \kappa \quad (192)$$

where we have used $\nabla_a \ell^a = \theta + \kappa$ and $\ell^a \nabla_a = d/d\lambda$. The quantity $\nabla_b (\kappa \ell^b) - \kappa^2$ is analogous to the right hand side of Eq. (141) in the case of a null vector. But the projection to a surface ‘orthogonal’ to a null vector is not well defined for us to introduce a covariant derivative. Instead, we have to work with a co-null vector k_a defined such that $k^2 = 0$ and $k_a \ell^a = -1$. Then the projection to the 2-surface is provided by the projection vector $q_b^a = \delta_b^a + k^a \ell_b + k_b \ell^a$ and we can relate $\mathcal{D}_a(\kappa \ell^a) \equiv q^{ab} \nabla_a (\kappa \ell_b)$ to $\nabla_a (\kappa \ell^a)$. In this computation, it is useful to note the identity $\ell_b \nabla_a (\phi \ell^b) = 0$ for any scalar ϕ . This leads to the result

$$\mathcal{D}_a(\kappa \ell^a) = \nabla_a (\kappa \ell^a) + k^b [\ell_b (\ell^a \nabla_a \kappa) + \kappa \ell^a \nabla_a \ell_b] = \nabla_a (\kappa \ell^a) - \left(\frac{d\kappa}{d\lambda} + \kappa^2 \right) \quad (193)$$

Therefore,

$$\nabla_a (\kappa \ell^a) - \kappa^2 = \mathcal{D}_a(\kappa \ell^a) + \frac{d\kappa}{d\lambda} \quad (194)$$

These results lead to the expressions used in the text.

9.6 Lie derivatives of expressions involving connections

We derive a series of results related to Lie derivatives, which will be useful in several derivations in the paper. The Lie derivative $\mathcal{L}_q T$ of some tensor (or tensor density) T has the structure of terms involving $q \partial T$ plus a series of $T \partial q$. So, a variation $\delta(\mathcal{L}_q T)$ will involve the sum of terms each having either $T \delta q$ or $q \delta T$. The sum of all terms

involving $q\delta T$ will lead to $\mathbb{L}_q(\delta T)$ while the sum of terms with $T\delta q$ will be $\mathbb{L}_{\delta q}T$. Therefore,

$$\delta(\mathbb{L}_q T) = \mathbb{L}_q(\delta T) + \mathbb{L}_{\delta q}T. \quad (195)$$

So, for any tensorial object T , the operations δ and \mathbb{L}_q commute if $\delta q = 0$.

This result, however, is not true for non-tensorial objects involving connections, etc. Since we need to handle them, we will derive a series of results for a particular set of them. To begin with, we have the expression for the Lie derivative of the connection

$$\mathbb{L}_v \Gamma_{bc}^a = \nabla_b \nabla_c v^a + R_{cmb}^a v^m \quad (196)$$

which can be obtained from first principles, knowing the transformation properties of the connection or from its definition in terms of the metric. As an aside, we point out the following curious consequence. In flat spacetime, we get a non-zero result:

$$\mathbb{L}_v[\Gamma_{bc}^a]_{\text{flat}} = \nabla_b \nabla_c v^a \quad (197)$$

(In fact, if we are using Cartesian coordinates, this result tells us that the Lie derivative of zero is non-zero; i.e., $\mathbb{L}_v 0 = \partial_b \partial_c v^a$!). This fact is sometimes used to define a background subtraction scheme in order to make expressions involving connections covariant. The essential idea is to work with the difference $\Gamma_{bc}^a - \Gamma_{bc}^a|_{\text{flat}}$, which, of course, is a tensor. The result in Eq. (196) can be re-written as:

$$\mathbb{L}_q \Gamma_{bc}^a = [\mathbb{L}_q \Gamma_{bc}^a]_{\text{std}} + \partial_b \partial_c q^a \quad (198)$$

where $[\mathbb{L}_q \dots]_{\text{std}}$ stands for the Lie derivative computed treating the indexed object (...) as though it was a tensor or tensor density. From this, we get the results:

$$\mathbb{L}_q N_{bc}^a = [\mathbb{L}_q N_{bc}^a]_{\text{std}} - \partial_b \partial_c q^a + \frac{1}{2} \left(\delta_c^a \partial_b \partial_l q^l + \delta_b^a \partial_c \partial_l q^l \right) \quad (199)$$

$$\begin{aligned} \mathbb{L}_q(f^{bc} N_{bc}^a) &= [\mathbb{L}_q(f^{bc} N_{bc}^a)]_{\text{std}} - f^{bc} \partial_b \partial_c q^a + f^{ab} \partial_b \partial_l q^l \\ &= [\mathbb{L}_q(f^{bc} N_{bc}^a)]_{\text{std}} - \sqrt{-g} K^a \end{aligned} \quad (200)$$

where

$$K^a = g^{lm} \partial_l \partial_m q^a - g^{al} \partial_l \partial_m q^m \quad (201)$$

For $V^a \equiv -g^{bc} N_{bc}^a$, we have:

$$\mathbb{L}_q(\sqrt{-g} V^a) = [\mathbb{L}_q \sqrt{-g} V^a]_{\text{std}} + \sqrt{-g} K^a \quad (202)$$

Further, using $L_{\text{sur}} \equiv \partial_a(\sqrt{-g} V^a)$, we obtain the result:

$$\mathbb{L}_q L_{\text{sur}} = \mathbb{L}_q \partial_a(\sqrt{-g} V^a) = \mathbb{L}_q \partial_a(\sqrt{-g} V^a) \Big|_{\text{std}} + \partial_a(\sqrt{-g} K^a) \quad (203)$$

We are now in a position to work out the Lie derivative of $\sqrt{-g} L_{\text{quad}} \equiv \sqrt{-g} L_H - L_{\text{sur}}$ needed in the paper. We have

$$\begin{aligned}\mathfrak{L}_q(\sqrt{-g} L_{\text{quad}}) &= \mathfrak{L}_q(\sqrt{-g} L_H - L_{\text{sur}}) \Big|_{\text{std}} - \partial_a(\sqrt{-g} K^a) \\ &= \partial_a [\sqrt{-g} L_{\text{quad}} q^a] - \partial_a(\sqrt{-g} K^a)\end{aligned}\quad (204)$$

Further, let us note the following two useful facts: (i) Because of the specific structure of Eqs. (198) and (199), the operations δ and \mathfrak{L} still commute on Γ_{bc}^a and N_{bc}^a even though they are non-tensorial. (ii) Using the notation $\mathcal{V}^a \equiv \sqrt{-g} V^a$ and Eq. (202), we see that $\sqrt{-g} K^a$ disappears in the following combination (which occurs frequently):

$$\begin{aligned}\mathfrak{L}_q(\mathcal{V}^a) - q^a \partial_c(\mathcal{V}^c) - \sqrt{-g} K^a &= [\mathfrak{L}_q \mathcal{V}^a]_{\text{std}} - q^a \partial_c(\mathcal{V}^c) \\ &= q^c \partial_c \mathcal{V}^a - \mathcal{V}^c \partial_c q^a + \mathcal{V}^a \partial_c q^c - q^a \partial_c \mathcal{V}^c \\ &= \partial_c(q^c \mathcal{V}^a - q^a \mathcal{V}^c)\end{aligned}\quad (205)$$

These results allow us to relate the Noether currents for the two Lagrangians L_H and L_{quad} . Using the standard result for the Noether current for the Einstein–Hilbert action J_H^a , we can re-write the result in Eq. (112) in the form

$$\sqrt{-g} J_{\text{quad}}^a = \sqrt{-g} J_H^a - [\mathfrak{L}_q(-\sqrt{-g} V^a) + q^a \partial_c(\sqrt{-g} V^c) + \sqrt{-g} K^a] \quad (206)$$

Using the notation $\mathcal{V}^a \equiv \sqrt{-g} V^a$, we see that $\sqrt{-g} K^a$ disappears in the expression in square brackets because of Eq. (205), giving

$$J_{\text{quad}}^a = J_H^a + \nabla_b \left(V^{[a} q^{b]} \right) = J_H^a + \nabla_b \left(-\frac{f^{lm}}{\sqrt{-g}} N_{lm}^{[a} q^{b]} \right) = J_H^a + \nabla_b \left(g^{lm} N_{lm}^{[b} q^{a]} \right) \quad (207)$$

where we have used $\mathcal{V}^a = \sqrt{-g} V^a = -f^{lm} N_{lm}^a$. If we write $J_H^a \equiv J_{\text{quad}}^a + J_{\text{sur}}^a$, thereby defining a Noether current associated with L_{sur} then:

$$J_{\text{sur}}^a = -\nabla_b \left(V^{[a} q^{b]} \right) = \nabla_b \left(V^{[b} q^{a]} \right) = \nabla_b J_{\text{sur}}^{ab}, \quad J_{\text{sur}}^{ab} \equiv V^{[b} q^{a]} = q^a V^b - q^b V^a \quad (208)$$

9.7 Variation of the gravitational momentum flux P_{quad}^a

Here we give some of the algebraic details related to the discussion in Sect. 7.2. We begin by relating P_H^a and P_{quad}^a by

$$\begin{aligned}\sqrt{-g} P_H^a &= f^{lm} \mathfrak{L}_q N_{lm}^a + q^a L_H \sqrt{-g} \\ &= \mathfrak{L}_q \left(f^{lm} N_{lm}^a \right) - N_{lm}^a \mathfrak{L}_q f^{lm} + q^a \left(\sqrt{-g} L_q - \partial_c \left(f^{lm} N_{lm}^c \right) \right) \\ &= -P_{\text{quad}}^a + \mathfrak{L}_q \left(f^{lm} N_{lm}^a \right) - q^a \partial_c \left(f^{lm} N_{lm}^c \right)\end{aligned}\quad (209)$$

Using the notation $\mathcal{V}^a = \sqrt{-g} V^a = -f^{lm} N_{lm}^a$ and using Eq. (205), we get

$$\sqrt{-g} P_H^a = -P_{\text{quad}}^a - [\mathfrak{L}_q(\mathcal{V}^a) - q^a \partial_c \mathcal{V}^c] = -P_{\text{quad}}^a - \left[\sqrt{-g} K^a + \partial_c (q^{[c} \mathcal{V}^{a]}) \right] \quad (210)$$

This gives, on combining with the on-shell result in Eq. (175), the result:

$$\begin{aligned} \delta P_{\text{quad}}^a &= -\delta(\sqrt{-g} P_H^a) - \delta(\sqrt{-g} K^a) + \partial_c \left(q^{[c} \delta N_{lm}^{a]} f^{lm} + q^{[c} N_{lm}^{a]} \delta f^{lm} \right) \\ &= -\sqrt{-g} \omega^a - \delta(\sqrt{-g} K^a) + \partial_c \left(q^{[c} N_{lm}^{a]} \delta f^{lm} \right) \end{aligned} \quad (211)$$

The extra term $\delta(\sqrt{-g} K^a)$ arises because of the non-tensorial character of P_{quad}^a .

We can also obtain corresponding expressions for $\delta(\sqrt{-g} J_{\text{quad}}^a)$ by relating it to $\delta(\sqrt{-g} J_H^a)$ using Eq. (113). Then we get:

$$\begin{aligned} \delta(\sqrt{-g} J_{\text{quad}}^a) &= \delta(\sqrt{-g} J_H^a) + \partial_b \left(\delta f^{lm} N_{lm}^{[b} q^{a]} + f^{lm} \delta N_{lm}^{[b} q^{a]} \right) \\ &= \delta f^{lm} \mathfrak{L}_q N_{lm}^a - \mathfrak{L}_q f^{lm} \delta N_{lm}^a + \partial_c \left(f^{lm} \delta N_{lm}^{[a} q^{c]} \right) + 2q^b \delta(\sqrt{-g} G_b^a) \\ &\quad + q^a \sqrt{-g} G_{ij} \delta g^{ij} + \partial_b \left(\delta f^{lm} N_{lm}^{[b} q^{a]} + f^{lm} \delta N_{lm}^{[b} q^{a]} \right) \\ &= \sqrt{-g} \omega^a + \partial_b \left(\delta f^{lm} N_{lm}^{[b} q^{a]} \right) + E^a \end{aligned} \quad (212)$$

where the equations of motion terms are:

$$E^a \equiv 2q^b \delta(\sqrt{-g} G_b^a) + q^a \sqrt{-g} G_{ij} \delta g^{ij} \quad (213)$$

It is possible to obtain the above result by direct computation as well. This derivation is instructive in clarifying the origin of the K^a term. Let us start with the definition:

$$\sqrt{-g} P_{\text{quad}}^a = (N_{ij}^a \mathfrak{L}_q f^{ij} - L_{\text{quad}} q^a \sqrt{-g}) \quad (214)$$

and compute the on-shell variation δP_{quad}^a . To do this, we will first obtain a preliminary result for the on-shell variation of L_{quad} in a specific form. We begin with the standard on-shell result

$$q^d \delta(L_{\text{quad}} \sqrt{-g}) = q^d \partial_c (N_{ab}^c \delta f^{ab}) \quad (215)$$

and re-write it in the following way:

$$\begin{aligned} q^d \delta(L_{\text{quad}} \sqrt{-g}) &= q^d \partial_c (N_{ab}^c \delta f^{ab}) = \partial_c \left(q^d N_{ab}^c \delta f^{ab} \right) - \left(\partial_c q^d \right) N_{ab}^c \delta f^{ab} \\ &= \left[\partial_c \left(q^c N_{ab}^d \delta f^{ab} \right) - N_{ab}^c \left(\delta f^{ab} \right) \partial_c q^d \right] + \partial_c \left[\left(q^d N_{ab}^c - q^c N_{ab}^d \right) \delta f^{ab} \right] \end{aligned} \quad (216)$$

To arrive at the second line, we have performed the usual symplectic trick of adding and subtracting the quantity $\partial_c(q^c N_{ab}^d \delta f^{ab})$. Let $v^d \equiv N_{ab}^d \delta f^{ab}$, so that

$$q^d \delta (L_{\text{quad}} \sqrt{-g}) = \left[\partial_c (q^c v^d) - v^c \partial_c q^d \right] + \partial_c \left[q^{[d} N_{ab}^{c]} \delta f^{ab} \right] \quad (217)$$

Then it is easy to see that (with $\mathcal{V}^a = -f^{lm} N_{lm}^a$)

$$\begin{aligned} \mathbb{E}_q(v^d) &= \mathbb{E}_q \left[-\delta \mathcal{V}^a - f^{ab} \delta N_{ab}^d \right] = -\mathbb{E}_q [\delta \mathcal{V}^a] - \mathbb{E}_q [f^{ab} \delta N_{ab}^d] \\ &= \left[\mathbb{E}_q v^d \right]_{\text{std}} - \delta (\sqrt{-g} K^a) = \partial_c (q^c v^d) - v^c \partial_c q^d - \delta (\sqrt{-g} K^a) \\ &= \delta f^{ab} \mathbb{E}_q N_{ab}^d + N_{ab}^d \delta (\mathbb{E}_q f^{ab}) \end{aligned} \quad (218)$$

where we have used Eq. (202) to obtain the third equality. Therefore

$$\partial_c (q^c v^d) - v^c \partial_c q^d = \left[\delta f^{ab} \mathbb{E}_q N_{ab}^d + N_{ab}^d \delta (\mathbb{E}_q f^{ab}) \right] + \delta (\sqrt{-g} K^a) \quad (219)$$

where we have used the fact that when $\delta q^a = 0$ the Lie differentiation and variation commute giving $\mathbb{E}_q \delta (f^{ab}) = \delta \mathbb{E}_q (f^{ab})$. So

$$\delta (q^d L_{\text{quad}} \sqrt{-g}) = \left\{ \delta f^{ab} \mathbb{E}_q N_{ab}^d + N_{ab}^d \delta \mathbb{E}_q f^{ab} \right\} - \partial_c \left[q^{[c} N_{ab}^{d]} \delta f^{ab} \right] + \delta (\sqrt{-g} K^d) \quad (220)$$

Once we have the variation of L_{quad} in place, we can compute the variation δP_{quad}^a defined with L_{quad} . We get

$$\begin{aligned} \delta P_{\text{quad}}^d &= \left\{ \delta N_{ab}^d \mathbb{E}_q f^{ab} + N_{ab}^d \delta (\mathbb{E}_q f^{ab}) - \delta f^{ab} \mathbb{E}_q N_{ab}^d - N_{ab}^d \delta \mathbb{E}_q f^{ab} \right\} \\ &\quad - \delta (\sqrt{-g} K^a) + \partial_c \left[q^{[c} N_{ab}^{d]} \delta f^{ab} \right] \\ &= \left\{ \delta N_{ab}^d \mathbb{E}_q f^{ab} - \delta f^{ab} \mathbb{E}_q N_{ab}^d \right\} + \partial_c \left[q^{[c} N_{ab}^{d]} \delta f^{ab} \right] - \delta (\sqrt{-g} K^a) \end{aligned} \quad (221)$$

This again has a symplectic structure in the first term and an explicit surface term in the second.

References

1. Bekenstein, J.D.: Lettere Al Nuovo Cimento **4**, 737 (1972)
2. Bekenstein, J.D.: Phys. Rev. **D7**, 2333 (1973)
3. Bekenstein, J.D.: Phys. Rev. **D9**, 3292 (1974)
4. Hawking, S.W.: Commun. Math. Phys. **43**, 199 (1975)
5. Hawking, S.W.: Phys. Rev. **D13**, 191 (1976)
6. Wald, R.M.: Living Rev. Rel. **4**, 6 (2001). [gr-qc/9912119](https://arxiv.org/abs/gr-qc/9912119)
7. Davies, P.C.W.: J. Phys. **A 8**, 609 (1975)
8. Unruh, W.G.: Phys. Rev. **D 14**, 870 (1976)
9. Jacobson, T.: Phys. Rev. Lett. **75**, 1260 (1995). [gr-qc/9504004](https://arxiv.org/abs/gr-qc/9504004)

10. Padmanabhan, T.: AIP Conf. Proc. **1241**, 93 (2010). [arXiv:0911.1403](#)
11. Padmanabhan, T.: AIP Conf. Proc. **1483**, 212 (2012). [arXiv:1208.1375](#)
12. Padmanabhan, T.: Rept. Prog. Phys. **73**, 046901 (2010). [arXiv:0911.5004](#)
13. Padmanabhan, T.: Phys. Rept. **406**, 49 (2005). [gr-qc/0311036](#)
14. Padmanabhan, T., Kothawala, D.: Phys. Repts. **531**, 115 (2013). [arXiv:1302.2151](#)
15. Kothawala, D., Padmanabhan, T.: Phys. Rev. **D79**, 104020 (2009). [arXiv:0904.0215](#)
16. Kolekar, S., Kothawala, D., Padmanabhan, T.: Phys. Rev. **D85**, 064031 (2012). [arXiv:1111.0973](#)
17. Padmanabhan, T.: J. Phys. Conf. Ser. **306**, 012001 (2011). [arXiv:1012.4476](#)
18. Eddington, A.: The Mathematical Theory of Relativity, 2nd edn. Cambridge University Press, Cambridge (1924)
19. Schrodinger, E.: Space-Time Structure. Cambridge Science Classics. Cambridge University Press, Cambridge (1950)
20. Einstein, A., Kaufman, B.: Ann. Math. **62**, 128 (1955)
21. Babak, S., Grishchuk, L.: Phys. Rev. **D61**, 024038 (1999). [gr-qc/9907027](#)
22. Kijowski, J.: Gen. Relativ. Gravit **29**, 307 (1997)
23. Parattu, K., Majhi, B.R., Padmanabhan, T.: Phys. Rev. **D 87**, 124011 (2013). [arXiv:1303.1535](#)
24. Padmanabhan, T.: Gen. Relativ. Gravit **44**, 2681 (2012). [arXiv:1205.5683](#)
25. Majhi, B.R., Padmanabhan, T.: Eur. Phys. J. **C 73**, 2651 (2013). [arXiv:1302.1206](#)
26. Padmanabhan, T.: Class. Quantum Grav. **21**, 4485 (2004). [gr-qc/0308070](#)
27. Padmanabhan, T.: Mod. Phys. Lett. **A25**, 1129 (2010). [arXiv:0912.3165](#)
28. Padmanabhan, T.: Phys. Rev. **D81**, 124040 (2010). [arXiv:1003.5665](#)
29. Szabados, L.B.: Living Rev. Rel. **7**, 4 (2004)
30. Padmanabhan, T., Paranjape, A.: Phys. Rev. **D75**, 064004 (2007). [gr-qc/0701003](#)
31. Padmanabhan, T.: Gen. Relativ. Gravit **40**, 529 (2008). [arXiv:0705.2533](#)
32. Mukhopadhyay, A., Padmanabhan, T.: Phys. Rev. **D 74**, 124023 (2006). [hep-th/0608120](#)
33. Katz, J.: Class. Quantum Grav. **2**, 423 (1985)
34. Arnowitt, R.L., Deser, S., Misner, C.W.: [gr-qc/0405109](#)
35. Padmanabhan, T.: Res. Astron. Astrophys. **12**, 891 (2012). [arXiv:1207.0505](#)
36. Padmanabhan, T.: Braz. J. Phys. **35**, 362 (2005). [gr-qc/0412068](#)
37. Kolekar, S., Padmanabhan, T.: Phys. Rev. **D82**, 024036 (2010). [arXiv:1005.0619](#)
38. Padmanabhan, T.: Phys. Rev. **D83**, 044048 (2011). [arXiv:1012.0119](#)
39. Padmanabhan, T.: Int. J. Mod. Phys. **D20**, 2817 (2011)
40. Kolekar, S., Padmanabhan, T.: Phys. Rev. **D85**, 024004 (2012). [arXiv:1109.5353](#)
41. Damour, T.: Quelques proprietes mecaniques, electromagnetiques, thermodynamiques et quantiques des trous noirs (available at <http://www.ihes.fr/damour/Articles/>), These de doctorat d'Etat, Universite Paris 6 (1979)
42. Damour, T.: Proceedings of the Second Marcel Grossmann Meeting on General Relativity (1982)
43. Thorne, K.S., Price, R.H., MacDonald, D.A.: Black Holes: The Membrane Paradigm. Yale University Press, USA (1986)
44. Padmanabhan, T.: Gravitation: Foundations and Frontiers. Cambridge University Press, Cambridge (2010)
45. Kijowski, J.: Gen. Relativ. Gravit **29**, 307 (1997)
46. Chen, C.M., Nester, J.M.: Class. Quantum Grav. **16**, 1279 (1999). [gr-qc/9809020](#)
47. Chen, C.M., Nester, J.M.: Grav. Cosmol. **6**, 257 (2000). [gr-qc/0001088](#)
48. Ashworth, M.C., Hayward, S.A.: Phys. Rev. **D60**, 084004 (1999). [gr-qc/9811076](#)
49. Regge, T., Teitelboim, C.: Ann. Phys. **88**, 286 (1974)
50. Brown, J.D., Henneaux, M.: J. Math. Phys. **27**, 489 (1986)
51. Wald, R.M., et al.: Phys. Rev. **D61**, 084027 (2000). [gr-qc/9911095](#)
52. Iyer, V., et al.: Phys. Rev. **D50**, 846 (1994). [gr-qc/9403028](#)
53. Julia, B., Silva, S.: Class. Quantum Grav. **15**, 2173 (1998). [gr-qc/9804029](#)
54. Julia, B., Silva, S.: [hep-th/0205072](#)
55. Barnich, G., Brandt, F.: Nucl. Phys. B **633**, 3 (2002)
56. Francaviglia, M., Raiteri, M.: Class. Quantum Grav. **19**, 237 (2002). [gr-qc/0107074](#)
57. Majhi, B.R., Padmanabhan, T.: Phys. Rev. **D 85**, 084040 (2012). [arXiv:1111.1809](#)
58. Carlip, S.: Phys. Rev. Lett. **82**, 2828 (1999). [hep-th/9812013](#)
59. Carlip, S.: Class. Quantum Grav. **16**, 3327 (1999). [gr-qc/9906126](#)
60. Carlip, S.: Phys. Rev. Lett. **88**, 241301 (2002). [gr-qc/0203001](#)

61. Majhi, B.R., Padmanabhan, T.: Phys. Rev. D **86**, 101501 (2012). [arXiv:1204.1422](#)
62. Brown, J.D., York, J.W.: Phys. Rev. **D47**, 1407 (1993)
63. Brown, J.D., Lau, S.R., York, J.W.: [gr-qc/0010024](#)
64. Booth, I.: [gr-qc/0008030](#)
65. Padmanabhan, T.: Adv. Sci. Lett. **2**, 174 (2009). [arXiv:0807.2356](#)
66. Ellis, G.F.R.: [arXiv:1306.3021](#)
67. Anderson, J.L., Finkelstein, D.: Am. J. Phys. **39**, 901 (1971)
68. Finkelstein, D.R., Galiatdinov, A.A., Baugh, J.E.: J. Math. Phys. **42**, 340 (2001). [gr-qc/0009099](#)
69. Unruh, W.G.: Phys. Rev. D **40**, 1048 (1989)
70. Padmanabhan, H., Padmanabhan, T.: Int. J. Mod. Phys D **22**, 1342001 (2013). [arXiv:1302.3226](#)
71. Padmanabhan, T.: [arXiv:1210.4174](#)
72. Gibbons, G., Hawking, S.: Phys. Rev. **D15**, 2752 (1977)
73. York Jr, J.W.: In: Zurek, W.H., van der Merwe, A., Miller, W.A. (eds.) Boundary Terms in the Action Principles of General Relativity, p. 246. Princeton Univ Press, Princeton (1988)
74. t Hooft, G.: [arXiv:gr-qc/9310026](#)
75. Horowitz, G.T., Polchinski, J.: In: Oriti D. (eds.) Approaches to Quantum Gravity. Cambridge University Press, Cambridge (2009), [arXiv:gr-qc/0602037](#)
76. de Mello Koch, R., Murugan, J.: In: Ellis, G., Murugan, J., Weltman, A. (eds.) Foundations of Space and Time. Cambridge University Press, Cambridge (2012), [arXiv:0911.4817](#)
77. Carlip, S.: [arXiv:1207.2504](#)
78. Barcelo, C., Liberati, S., Visser, M.: Living Rev. Rel. **8**, 12 (2005). [arXiv:gr-qc/0505065](#)
79. Unruh, W.G.: Phys. Rev. **D51**, 2827 (1995)
80. Hu, B.L.: Int. J. Theor. Phys. **44**, 1785 (2005). [arXiv:gr-qc/0503067](#)
81. Volovik, G.E.: The Universe in a Helium Droplet. Oxford University Press, Oxford (2009)
82. See, for example, Sorkin, R. D.: [arXiv:gr-qc/0309009](#)