

# An entropy current and the second law in higher derivative theories of gravity

Sayantani Bhattacharyya,<sup>a</sup> Prateksh Dhivakar,<sup>b</sup> Anirban Dinda,<sup>a</sup> Nilay Kundu,<sup>b</sup> Milan Patra<sup>a</sup> and Shuvayu Roy<sup>a</sup>

<sup>a</sup>*School of Physical Sciences, National Institute of Science Education and Research, HBNI, Bhubaneswar, Khurda 752050, Odisha, India*

<sup>b</sup>*Department of Physics, Indian Institute of Technology Kanpur, Kalyanpur, Kanpur 208016, India*

*E-mail:* [sayanta@niser.ac.in](mailto:sayanta@niser.ac.in), [prateksh@iitk.ac.in](mailto:prateksh@iitk.ac.in),  
[anirban.dinda@niser.ac.in](mailto:anirban.dinda@niser.ac.in), [nilayhep@iitk.ac.in](mailto:nilayhep@iitk.ac.in),  
[milan.patra@niser.ac.in](mailto:milan.patra@niser.ac.in), [shuvayu.roy@niser.ac.in](mailto:shuvayu.roy@niser.ac.in)

**ABSTRACT:** We construct a proof of the second law of thermodynamics in an arbitrary diffeomorphism invariant theory of gravity working within the approximation of linearized dynamical fluctuations around stationary black holes. We achieve this by establishing the existence of an entropy current defined on the horizon of the dynamically perturbed black hole in such theories. By construction, this entropy current has non-negative divergence, suggestive of a mechanism for the dynamical black hole to approach a final equilibrium configuration via entropy production as well as the spatial flow of it on the null horizon. This enables us to argue for the second law in its strongest possible form, which has a manifest locality at each space-time point. We explicitly check that the form of the entropy current that we construct in this paper exactly matches with previously reported expressions computed considering specific four derivative theories of higher curvature gravity. Using the same set up we also provide an alternative proof of the physical process version of the first law applicable to arbitrary higher derivative theories of gravity.

**KEYWORDS:** Black Holes, Classical Theories of Gravity

ARXIV EPRINT: [2105.06455](https://arxiv.org/abs/2105.06455)

---

**Contents**

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction and summary</b>  | <b>1</b>  |
| 1.1      | Statement of the problem and summary of the final result   | 6         |
| <b>2</b> | <b>Basic setup and key conceptual elements</b>   | <b>9</b>  |
| 2.1      | The coordinate system adapted to the horizon   | 10        |
| 2.2      | Stationarity and small deviation from it by dynamical fluctuations   | 11        |
| 2.3      | Killing symmetry and boost weight of quantities  | 14        |
| 2.4      | Significance of the structure of certain components of the equations of motion for a proof of the second law | 16        |
| 2.5      | An important identity relating the equations of motion with the Noether charge for diffeomorphism            | 20        |
| <b>3</b> | <b>A proof of the existence of an entropy current</b>  | <b>21</b> |
| 3.1      | A brief and schematic sketch of the proof highlighting the basic strategy                                    | 22        |
| 3.2      | An elaborate analysis of the proof with technical details  | 25        |
| 3.2.1    | The generic structure of a covariant tensor with positive boost weight                                       | 26        |
| 3.2.2    | Analyzing the structure of the first term (i.e. $\Theta^r$ ) in $E_{vv}$                                     | 27        |
| 3.2.3    | Analyzing the structure of the second term (i.e. $D_\mu Q^{r\mu}$ ) in $E_{vv}$                              | 36        |
| 3.2.4    | Final analysis for the structure of $E_{vv}$   | 37        |
| <b>4</b> | <b>Verification of the proof for specific examples</b>   | <b>38</b> |
| <b>5</b> | <b>A proof of the physical process version of the first law</b>  | <b>42</b> |
| <b>6</b> | <b>Discussions and future directions</b>   | <b>45</b> |
| <b>A</b> | <b>Notations, conventions and useful definitions</b>   | <b>47</b> |
| <b>B</b> | <b>More details regarding the boost weight of a covariant tensor</b>   | <b>48</b> |
| B.1      | Equivalence between different ways of evaluating the Boost weights   | 48        |
| B.2      | Structure of covariant tensors in the stationary background solution   | 49        |
| <b>C</b> | <b>Arguments justifying eq. (3.8)</b>  | <b>50</b> |
| <b>D</b> | <b>Detailed proof of Result: 1 in eq. (3.14)</b>   | <b>51</b> |
| D.1      | The proof  | 51        |
| D.2      | Proof of the identity eq. (D.7)  | 54        |
| <b>E</b> | <b>Variation of a generic covariant tensor under diffeomorphism: a justification for eq. (3.23)</b>          | <b>54</b> |

---

## 1 Introduction and summary

In Einstein's classical theory of gravity, black holes are interesting solutions that behave like thermodynamic objects [1–4] and one can associate notions like temperature and entropy to them. It is believed that there is also a consistent statistical picture of black hole thermodynamics. This allows us to consider black holes as basically a thermodynamic ensemble of micro-states in an underlying quantum theory of gravity.

Any UV complete theory of gravity, in its low energy limit, will typically generate several higher derivative corrections to the classical two derivative theory of Einstein gravity. Black hole solutions should continue to exist in these theories, at least when the higher derivative coupling is treated perturbatively. The fact that black holes are indeed a collection of a large number of micro-states will remain so even with the higher derivative corrections being added to the two derivative gravity action. Therefore, they should also maintain the laws of thermodynamics provided we could correctly identify the thermodynamic properties with the geometric properties of black holes. This identification is well-known in two-derivative Einstein gravity. However, in higher derivative theories, we still do not have a complete understanding of how to do it for dynamical black holes.

More precisely, we know that in two-derivative theories of gravity, the area of a time-slice of the horizon plays the role of entropy of the black hole that satisfies both the first and the second law of thermodynamics. In particular, the second law follows from the area increase theorem for black holes [5, 6]. However, when one considers higher derivative theories of gravity, the same horizon area does not work as the definition of black hole entropy. We need to modify it to account for the higher derivative corrections. In [7, 8] a geometrical notion of black hole entropy, called the Wald entropy, was provided for an arbitrary diffeomorphism invariant theory of gravity (including theories with higher derivative corrections to Einstein's theory) in such a way that the first law of thermodynamics is satisfied. However, once a dynamical black hole solution is considered in general higher derivative theories of gravity, Wald entropy's construction becomes ambiguous, meaning a bunch of terms could be added to Wald entropy without affecting the first law [9–11]. These are generally known as the JKM ambiguities in the literature, and they are non-zero only for dynamical configurations.

The concept of an equilibrium black hole configuration in a theory of gravity is associated with a space-time metric admitting a Killing vector that becomes null on the event horizon. Such a geometric configuration is also known as the stationary black hole solution. As we know, the first law of thermodynamics is essentially a comparison between the black hole parameters (like mass, charge, etc.) for two slightly different but nearby stationary configurations. The second law of thermodynamics, however, necessarily refers to dynamics. It states that the black hole entropy compared between final and initial stationary configurations (not necessarily nearby) can never decrease, i.e., entropy is always produced in every dynamical process. Although the Wald entropy satisfies the first law by construction, there is no clear proof that it will obey the second law in any arbitrary diffeomorphism invariant theory of gravity. In other words, unlike the two-derivative theory, we still do not have a completely satisfactory geometric construction of entropy that satisfies

both the first and the second law of thermodynamics for dynamical black hole solutions in higher derivative theories of gravity.

Over the years, various attempts have been made to understand if indeed there is a notion of the second law for black hole thermodynamics in higher derivative theories of gravity, [8–16].<sup>1</sup> A natural approach is to start from Wald entropy as the equilibrium definition of black hole entropy and explore the possibilities of extending it to dynamical situations such that the second law is satisfied. It essentially means fixing the JKM ambiguities related to the Wald entropy for dynamical configurations by imposing the second law as a principle. This has so far been the basic theme of various approaches in this context.

In any such attempt, one would, however, naturally have to devise an algorithm to analyze dynamical black hole space-times, which by itself is very difficult to execute even in general relativity. Therefore we need to adapt a perturbative approach around the exactly known stationary solutions. A standard procedure, which we will also follow in our analysis in this paper, is to consider some stationary black hole space-time which then gets slightly perturbed due to an external source and then finally settles down to another stationary configuration. The amplitude of the dynamical fluctuation is small enough in the sense that we could always perform a perturbative expansion in its amplitude and work with only the linearized term in this expansion, ignoring all non-linear terms. We will call this approximation the linearized amplitude expansion.<sup>2</sup>

**The context and the backdrop of our current work.** Our analysis in this paper is significantly motivated and based on the important results that were reported recently in [14] and subsequently in a follow-up paper [22]. Therefore, to better understand the context of this current paper, it is useful and informative that we discuss them a priori.<sup>3</sup>

Working within the approximation of linearized amplitude expansion of small dynamical perturbation to a stationary black hole configuration, in [14], it was shown that these out-of-equilibrium JKM ambiguities of Wald entropy could uniquely be fixed up to a particular order by demanding that it must satisfy the linearized version of the second law. Additionally, in [14], the statement of the second law was satisfied in a more robust sense: not only the difference between the total Wald entropy of the final and the initial equilibrium configuration was non-negative; instead, the entropy was monotonically increasing at every instant of “time” throughout the evolution from the initial to the final stationary point. This work, thereby, provides an explicit construction of an entropy function — an out-of-equilibrium extension of the Wald entropy — satisfying the second law.

---

<sup>1</sup>See the recent reviews of black hole thermodynamics in higher derivative theories of gravity [17, 18] and references therein, for a detailed discussion on this topic.

<sup>2</sup>One should note that with this linearized amplitude expansion, we cannot study violent time-dependent processes like black hole mergers. Although we would have nothing to say about those situations within this paper’s premise, see [19–21] where a similar problem has been studied with exciting conclusions.

<sup>3</sup>For a detailed discussion the reader is requested to see section 2 of [22] where a comprehensive review of [14] is given.

The general strategy that was followed in [14] was to show that the “time” derivative of this entropy function is non-negative at every instant along the evolution.<sup>4</sup> This way, the basic working principle essentially resembles the same that is responsible for the area increase theorem in Einstein’s theory of gravity. After choosing a particular metric gauge suitable for the case at hand and working within the approximation of linearized amplitude expansion, a specific ‘time-time’ component of the rank-two equation of motion was studied. More precisely, a particular sub-class of the residual symmetry transformation of the chosen metric gauge (named as the “boost transformation”) was used to classify how different quantities transform with definite weights (termed as the “boost-weight”) under that symmetry of the stationary background. Eventually, using this classification based on boost-weights, a specific expression of that “time-time” component of the equation of motion was obtained, maintaining the second law up to linearized order in amplitude. Finally, it was argued that the zero boost-weight part of that expression determines the equilibrium Wald entropy. In contrast, the higher boost sector fixes the JKM ambiguities of the entropy (which is the same as the out-of-equilibrium part of it).

In essence, the importance of the work [14] lies in the fact that it not only provides an algorithm to fix the JKM ambiguities but also to reproduce the Wald entropy of the stationary configuration itself. As a check of its consistency, this entropy function defined locally in time should reduce to the known expression of Wald entropy for stationary configurations once the dynamics are switched off. However, in [22], a subtle technical issue of this construction regarding its implementation was pointed out. Actually, this constructive algorithm of [14] stumbles upon a road-block at the very leading order in amplitude expansion. The zero boost-weight sector of the “time-time” component of the equations of motion doesn’t acquire the desired form, which is responsible for reproducing the Wald entropy of the stationary black holes. However, this should not be a problem since, for stationary black holes, we already know that Wald’s construction works. In [14], this trickier set of terms were dealt with by taking recourse to the physical process version of the first law [11, 23–28]. It was argued that if the physical process version of the first law is to be valid, then things should work out nicely to reproduce the correct expression for the Wald entropy at equilibrium.

The main goal of the analysis carried out in [22] was to perform a brute force calculation focussing on only four derivative theories of gravity to check if the arguments presented in [14] are indeed true. By explicitly working out the “time-time” component of the equations of motion for those theories, it was shown that there are terms, in that zero boost-weight sector of it, which cannot be cast in the form required to identify them as the Wald entropy density. However, it was also noticed that the structure of every such anomalous term could be rearranged as the spatial divergence of a specific spatial vector, named the entropy current. In [14] the appearance of such additional terms in the particular components of the equations of motion with the specific structures mentioned above

---

<sup>4</sup>The reason we are putting “time” within the quotation mark is the fact that on horizon rather than having a ‘time’ coordinate as the component of a time-like vector, we will have a parameter running along the curve generated by a null vector, the generator of the horizon. The horizon being a null surface, we cannot actually have such a time like coordinate in true sense.

was not paid its due attention, but the physical process version of the first law was called for the rescue.

The physical process version of the first law is actually formulated along the same lines as the linearized version of the second law.<sup>5</sup> A stationary black hole is driven out of its equilibrium state by a small stress tensor associated with some matter falling in the black hole from asymptotic infinity, and finally, it settles down to another stationary configuration. A relation between the changes in the parameters (like entropy, mass, etc.) of the two equilibrium black holes follows accordingly. The presence of the additional anomalous terms in the form of a spatial divergence of a spatial current, which was pointed out in [22], didn't affect the statement of the physical process version of the first law. This is because, in the physical process version of the first law, one compares the total entropy integrated over the spatial slices of the horizon, and thereby the entire spatial divergence terms drop out. In turn, this also justifies the arguments presented in [14] that, once the dynamical evolution of a black hole between two equilibrium configurations satisfies the physical process version of the first law, the zero boost-weight sector of the “time-time” component of the equations of motion reproduces the correct expression of the Wald entropy for stationary black holes.

Although in [14] there was a temporal locality in the version of the second law, that the entropy was considered as an integrated expression over the spatial slices of the horizon (for the reasons mentioned above) signifies that there was no locality in the spatial directions. In other words, all the statements that were made in [14] involved an entropy function that was local in time but could only be associated to each globally complete spatial section of the horizon but not to any infinitesimal elements of it. On the other hand, the analysis of [22] explicitly showed that, although only for specific four derivative theories of gravity, introducing the notion of an entropy current, the second law for black holes could be formulated in its most potent possible form, i.e., in an ultra-local version which is local in both time and space. In fact, this statement has been recast even in a more robust sense as follows: in a general four derivative theory of gravity, working up to linear order in the amplitude of the dynamical fluctuations, the spatial components of the entropy current has to be accounted for if one aims to prove an ultra-local version of the second law (as described above).

The entropy current, constructed in [22], is a  $(d - 1)$ -vector defined on the horizon, with coordinates running along with the ‘time’ and the spatial coordinates that span the horizon. The ‘time’ component of the entropy current gives us the instantaneous entropy density, whereas the spatial components of the same account for a flow of entropy across different local segments of the spatial slice of the horizon. The existence of a non-negative divergence of this entropy current is definitely the statement that entropy is being produced locally at every point of the dynamical space-time. However, due to the non-zero spatial components of the entropy current, it also suggests that there is another simultaneous mechanism by which redistribution of entropy is taking place between different adjacent

---

<sup>5</sup>It should be noted that although the necessary formalism to address both of them is along the same lines, the implications of them are pretty different. One needs to fix the JKM ambiguities to prove a linearized version of the second law; however, they have vanishing contribution for the first law.

spatial regions.<sup>6</sup> This physical picture highlighting the significance of having an entropy current or the idea of an ultra-locality for the second law is undoubtedly consistent in Einstein’s theory of gravity. It is well known that the area increase theorem in general relativity is a local statement that works for every instance of time as well as at each spatial point of a time slice of the horizon. It is therefore not at all absurd to expect that a similar physical picture of the same kind would prevail even when higher derivative theories of gravity are considered, at least for dynamical fluctuations with small enough amplitudes. Actually, the results of [22] very strongly favour such an interpretation, at least for the theories explicitly studied there.

It is, therefore, of utmost importance to know whether the lessons that we have learned so far in the discussions above are applicable much more generally and to other generic theories of gravity as well. In this present work, our primary goal is to address this question. Rather than focussing on specific theories as was done in [22], here in this work, we will be considering a general diffeomorphism invariant theory of gravity. However, our methodology will still be the same. For example, we will extensively use the same boost-symmetry of the horizon geometry to constrain the possible structural form of the “time-time” component of the equations of motion.

Additionally, we will also restrict ourselves only to those dynamical situations where the approximation of a small amplitude of dynamical perturbation is good enough. In other words, a linearized expansion in the amplitude of the non-stationary fluctuations is allowed. Interestingly, our final result turns out to be an affirmative answer to the question we posed above: in an arbitrary diffeomorphism invariant theory of gravity, construction of an entropy current to account for the ultra-local version of the second law of thermodynamics, at least for dynamical processes with small amplitudes, is indeed possible.

It is worth highlighting that one crucial aspect of the construction of an entropy current in this paper has been to prove a linearized version of the second law without ever invoking the physical process version of the first law. It has been instrumental in enabling us to maintain strict locality throughout our work. As we have mentioned before, this is in contrast to [14], where the physical process version of the first law was used to take care of the zero boost-weight sector in specific components of equations of motion, and hence the spatial locality was lost. On the other hand, we have been able to design an independent proof of the first law’s physical process version, using the same basic formalism that was formulated to prove a linearized version of the second law.

Before we proceed, let us also mention that in the literature, the concept of an entropy current in the context of dynamical black holes has already been introduced; see [33–37]. Although they share the common primary goal in terms of constructing a local entropy current, the methodology and the working principles that we are using in our paper are very different from theirs. For example, in [33] the context was to interpret the field equations of gravity as an equation of state, the primary motivation behind [34, 35] has mainly been the celebrated fluid gravity correspondence [38]. On the other hand, a certain duality between

---

<sup>6</sup>The existence of such an entropy current is well known in theories where there is no dynamical gravity, e.g., in fluid dynamics [29–32].

membrane dynamics and black holes called the membrane-gravity duality had been the driving principle for [36, 37].

The outline of this paper is as follows. We start in section 1.1 with making a precise statement of the problem at hand and a summary of the main result of our paper. In section 2 we will then discuss the basic set up of our construction. This will include an extensive discussion of various basic concepts which play an important role in our analysis. After introducing the coordinate system that we will work with, the idea of a boost transformation and the weights of different quantities under this specific symmetry of the horizon geometry in our metric gauge will be explained. We will also discuss how these are related to the concepts of stationarity and a slight deviation from it.

Furthermore, an important identity relating the equations of motion and the Noether charges under diffeomorphism invariance will be established. We will end this section with a discussion of the strategy that will be followed. In the next section, i.e., section 3, we will present detailed proof of how to extract out the entropy current. This will be the central technical part of our paper. In the following section 4, we will show that the expression of the entropy current obtained from the abstract proof in this paper exactly reproduces the results known in four derivative theories of gravity which were reported in [22]. Next, in section 5 we use the same technical set up and present an independent proof of the physical process version of the first law for any arbitrary diffeomorphism invariant theory of gravity. Finally, in section 6, we conclude with a brief discussion of the consequences of our results and possible future directions. We have also summarised the notations, conventions, and useful definitions in appendix A for the convenience of the reader. The other appendices provide useful technical details for our computations.

## 1.1 Statement of the problem and summary of the final result

With the motivations and the context of our paper being discussed so far, in this subsection, we will write down a more precise and explicit statement of the problem at hand as well as the final result that we have been able to achieve. This will help us organize our presentation in the following sections in a much more coherent manner.

We will work with the most general diffeomorphism invariant theory of gravity in  $d$  space-time dimensions with coordinates denoted by  $x^\mu$ , with the action given by<sup>7</sup>

$$I = \frac{1}{4\pi} \int d^d x \sqrt{-g} (L_{\text{grav}} + L_{\text{mat}}) . \quad (1.1)$$

In eq. (1.1),  $L_{\text{grav}}$  is the Lagrangian for the degrees of freedom corresponding to the gravity part,<sup>8</sup>

$$L_{\text{grav}} = L_{\text{grav}}(g_{\alpha\beta}, R_{\alpha\beta\rho\sigma}, D_{\alpha_1} R_{\alpha\beta\rho\sigma}, \dots) . \quad (1.2)$$

<sup>7</sup>We are using a convention such that the Newton's constant  $G = 1/4$ , such that the area of the horizon gives the entropy of a black hole in Einstein's gravity.

<sup>8</sup>One can argue that, see section 2 of [8], any diffeomorphism invariant Lagrangian will have the specific functional dependence as mentioned here.



Moreover,  $L_{\text{mat}}$  is the Lagrangian for the matter fields present in the theory.<sup>9</sup> In this paper, we will be focussing on the gravity part of the Lagrangian  $L_{\text{grav}}$ . For the matter sector, all that we will need is to consider that it gives us a stress tensor satisfying the null energy condition. In what follows, we will make this more precise.

We will first make the following choice of gauge for the space-time metric with coordinates  $x^\mu = \{r, v, x^i\}$ , for  $i = 1, \dots, (d - 2)$ , that represents a dynamical black hole solution with a regular event horizon at  $r = 0$ ,

$$ds^2 = 2 dv dr - g_{vv}(r, v, x) dv^2 + 2g_{vi}(r, v, x) dv dx^i + h_{ij}(r, v, x) dx^i dx^j, \quad (1.3)$$

such that  $g_{rv} = 1, g_{rr} = 0, g_{ri} = 0, g_{vv}|_{r=0} = 0, \partial_r g_{vv}|_{r=0} = 0, g_{vi}|_{r=0} = 0$ .

The null hyper-surface of the event horizon, which is spanned by the coordinates  $\{v, x^i\}$ , will be denoted by  $\mathcal{H}$ . The constant  $v$ -slices of the horizon will be represented by  $\mathcal{H}_v$ . The coordinate  $v$  is an affine parameter corresponding to the null generator  $\chi^\mu \partial_\mu \equiv \partial_v$  of the horizon.<sup>10</sup> In section 2.1 we will explain in more detail the choice of our coordinates.

This paper will only focus on scenarios where the dynamics can be treated as small fluctuations around a stationary background. In operational terms, the full space-time metric can be decomposed as follows

$$g_{\mu\nu} = g_{\mu\nu}^{eq} + \epsilon \delta g_{\mu\nu}, \quad (1.4)$$

where  $g_{\mu\nu}^{eq}$  is the metric for the stationary black hole space-time, and  $\epsilon$  is a parameter denoting the amplitude of small fluctuations around that stationary background. The smallness of the fluctuation allows us to perform a perturbative expansion in the amplitude and work only up to linear order in that amplitude expansion, i.e., neglecting terms with  $\mathcal{O}(\epsilon^2)$  and higher orders.

As we have emphasized before, in this current work, the basic idea behind constructing a proof of the linearized version of the second law is to constrain the structure of a particular component of the equations of motion, namely the  $vv$ -component of it denoted by  $\mathcal{E}_{vv}$ . It should be noted that by  $\mathcal{E}_{\mu\nu}$  we are denoting the equations of motion which follow from varying the complete Lagrangian, including both the gravitational and the matter sector, with respect to the space-time metric. The part of  $\mathcal{E}_{\mu\nu}$  that is derived only from the gravitational part of the Lagrangian will be denoted by  $E_{\mu\nu}$ . Therefore, we will have the following relation

$$\mathcal{E}_{\mu\nu} = E_{\mu\nu} + T_{\mu\nu}, \quad (1.5)$$

where  $T_{\mu\nu}$  is the stress-energy tensor obtained from the matter part of the Lagrangian.<sup>11</sup>

---

<sup>9</sup>All possible couplings between the matter and the gravity sector are contained within the Lagrangian  $L_{\text{mat}}$ . It is also possible that  $L_{\text{mat}}$  has to contain higher spin fields to make the entire higher derivative theory consistent with causality concerning processes such as graviton scattering, as was pointed out in [39].

<sup>10</sup>For the reasons mentioned in footnote-4, we have been loosely calling the parameter  $v$  as the “time” coordinate, although it is actually a null direction.

<sup>11</sup>We are working with the following definition of the stress tensor. While obtaining the equations of motion  $\mathcal{E}_{\mu\nu}$  by varying the total Lagrangian with respect to the fluctuations in the metric, there will be one type of terms involving only the metric components and their derivatives. We denote them as  $E_{\mu\nu}$ . However, in  $\mathcal{E}_{\mu\nu}$ , there will be another type of terms that involve both the metric and the matter fields. They will all be included in  $T_{\mu\nu}$ .

In section 2.4 we have discussed in detail why the structure of this particular component of the equations of motion is so crucial to validate the second law.<sup>12</sup>

The primary content of our analysis in this paper can be precisely summarised as the following statement:

*Using arguments based on diffeomorphism invariance and a certain boost symmetry of the horizon geometry to classify possible structures that can appear in  $\mathcal{E}_{vv}$ , when evaluated on the horizon (i.e. at  $r = 0$ ), we have been able to show that*

$$\begin{aligned} \mathcal{E}_{vv}|_{r=0} &= \partial_v \left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \frac{1}{\sqrt{h}} \partial_i (\sqrt{h} \mathcal{J}^i) \right] + T_{vv} + \mathcal{O}(\epsilon^2) \\ &= \partial_v \left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right] + T_{vv} + \mathcal{O}(\epsilon^2), \end{aligned} \tag{1.6}$$

where  $h$  is the determinant of the induced spatial metric, i.e.  $h_{ij}$ , on the co-dimension two constant  $v$ -slices of the horizon, and  $\nabla_i$  is the covariant derivative compatible with  $h_{ij}$ . Also,  $\mathcal{J}^v$  and  $\mathcal{J}^i$  will get contribution only from the gravitational part of the Lagrangian and eq. (1.6) can equivalently be stated as

$$E_{vv}|_{r=0} = \partial_v \left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right] + \mathcal{O}(\epsilon^2), \tag{1.7}$$

which follows from eq. (1.5).

Furthermore, we will only focus on situations where the dynamics is initiated due to some perturbation coming from a matter stress tensor such that it satisfies the null energy condition. In our metric gauge this becomes  $T_{vv} \geq 0$ . Using this and also the fact that for on-shell processes the r.h.s. of eq. (1.6) should vanish, we immediately obtain

$$\partial_v \left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right] \leq 0. \tag{1.8}$$

Additionally, since we also require that the dynamics settles down to some stationary black hole configuration at late future as  $v \rightarrow \infty$ , where  $\partial_v$  becomes proportional to the Killing direction, we get<sup>13</sup>

$$\left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right] \rightarrow 0 \quad \text{as } v \rightarrow \infty. \tag{1.9}$$

Therefore, from eq. (1.8) and eq. (1.9) it follows that

$$\left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right] \geq 0 \quad \text{for all finite } v. \tag{1.10}$$

---

<sup>12</sup>Remembering that we are writing the components of equations of motion, we will always have  $\mathcal{E}_{vv} = 0$  when thinking of it in an on-shell manner. However, this is not how we would consider  $\mathcal{E}_{vv}$ ; rather, we would like to emphasize that our purpose here is to constrain the off-shell structure of  $\mathcal{E}_{vv}$ .

<sup>13</sup>The reason for this is the following: by construction  $\mathcal{J}^i$  includes at least one  $\partial_v$ -derivative and when evaluated on a stationary metric all the  $\partial_v$ 's should vanish.

From eq. (1.10), we can also conclude that our procedure of arriving at eq. (1.7) enables us to construct a local entropy current on the horizon with components  $\mathcal{J}^v$  and  $\mathcal{J}^i$ . Consequently, we can immediately identify its  $v$ -component given by  $\mathcal{J}^v$  as the local entropy density, which not only reproduces the equilibrium expression of Wald entropy but also determines the out-of-equilibrium extension of it in the form of fixing the JKM ambiguities. On the other hand, the spatial components, i.e.,  $\mathcal{J}^i$ , are identified with the entropy current density signifying a possible spatial flow of entropy locally at each point of the constant  $v$ -slices of the horizon. Also, we can see that with these identifications, eq. (1.10) signifies a local statement of entropy production at each space-time point on the horizon. Interestingly, we have also been able to show that when the stationary limit is taken by switching off the dynamical perturbations,  $\mathcal{J}^v$  reduces to the expression of Wald entropy known for equilibrium black holes and  $\mathcal{J}^i$  identically vanishes.<sup>14</sup> It is worth highlighting that obtaining eq. (1.7), therefore, not only justifies a consistent out-of-equilibrium extension of the Wald entropy but also provides us an algorithm to obtain  $\mathcal{J}^v$  and  $\mathcal{J}^i$  as entirely geometric quantities in the sense that they are constructed solely out of the metric components and their derivatives. However, an important point worth emphasizing is the fact that the choice of our gauge for the space-time metric, eq. (1.3), plays a vital role in the definitions of  $\mathcal{J}^v$  and  $\mathcal{J}^i$  via eq. (1.6) and eq. (1.7). Consequently, we must admit that our construction of the local entropy current is not a covariant one, and it relies on the choice of the constant  $v$ -slices of the horizon.

Finally, we conclude our introduction with the following comments regarding the novelty of the results reported in this paper. The key result of this paper is to prove that eq. (1.7) is valid for arbitrary diffeomorphism invariant theories of gravity, up to linearized order in the amplitude of the non-stationary perturbations. However, we must note that the rest of the arguments connected to the linearized version of the second law have been used before, e.g., see [14]. Even in the case of two derivative Einstein gravity, the area increase theorem for the second law of black hole mechanics is very similar. On the other hand, a case by case and explicit computation of  $E_{vv}$  in [22] has already revealed the appearance of the terms involving  $\mathcal{J}^i$  in eq. (1.7), but only for four derivative theories of gravity. In this context, the originality of our results, therefore, lies in the fact that we have been able to justify the ultra-local version of the second law via the entropy current on the horizon with an analysis that is based on general principles like Noether charge for diffeomorphism invariance and most importantly it makes a statement applicable to any higher derivative theory of gravity.

## 2 Basic setup and key conceptual elements

In this section, we will explain the basic setup of our problem. We will introduce a coordinate system adapted to the horizon. The choice of this coordinate system is akin to choosing a gauge for the metric, and we will work with this throughout this paper. Subsequently, we will also elaborate on the approximations that will be used. A specific symmetry, called the boost-invariance, will be introduced. By knowing how different terms transform under

<sup>14</sup>This actually justifies eq. (1.9) above.

this boost transformation, we will see how any generic tensor quantity, built out of metric functions and their derivatives, in a theory of gravity can be constrained. After we have discussed various elements of the basic setup,<sup>15</sup> we will also describe the main strategy, which will be followed in the subsequent section to prove the existence of an entropy current for a general diffeomorphism invariant theory of gravity.

## 2.1 The coordinate system adapted to the horizon

We are considering a classical theory of gravity, and black holes are specific solutions described by a space-time that is endowed with a horizon — a null hyper-surface separating the black hole singularity from the asymptotic infinity. Our goal is to choose a coordinate system to describe a space-time with dynamical black holes in this subsection. We will work in  $d$ -dimensional space-times, and let us suppose that on the horizon, a co-dimension one hyper-surface,  $\partial_v$  is the null generator with  $v$  being an affine parameter. We will choose  $v$  to be one of the coordinates along the horizon. Let  $\{x^i\}$ , for  $i = 1, 2, \dots, (d-2)$ , be the other  $(d-2)$  spatial coordinates along the horizon. Thus,  $\{v, x^i\}$  together construct a coordinate system on the horizon. Next, we consider null geodesics generated by  $\partial_r$ , emanating out of the horizon. We further demand that these geodesics, generated by  $\partial_r$ , make an angle such that, on the horizon, the inner product of  $\partial_r$  and  $\partial_v$  is one. Additionally, the inner product of  $\partial_r$  and  $\partial_i$  vanishes for every  $i$ . The origin of the  $r$  coordinate is chosen to lie on the horizon, i.e.,  $r = 0$  will always give the horizon's location. Finally, we also restrict the parameter  $r$  to be an affine parameter along these geodesics, denoting the coordinate away from the horizon. The full  $d$  dimensional coordinates will be denoted by  $x^\mu = \{r, v, x^i\}$ . In this coordinate system, the metric takes the following form

$$ds^2 = 2 dv dr - r^2 X(r, v, x) dv^2 + 2r \omega_i(r, v, x) dv dx^i + h_{ij}(r, v, x) dx^i dx^j. \quad (2.1)$$

The detailed arguments justifying that the most general form of the metric describing a black hole space-time can always be written as in eq. (2.1) are given in appendix A of [40].<sup>16</sup> It is interesting to note that the metric written in eq. (2.1) should not be considered as the global coordinates. However, the space-time sufficiently close to the horizon will always obtain the form mentioned above. It can be explicitly checked that for Schwarzschild black holes the full space-time can be written in the form given in eq. (2.1). On the other hand, for rotating Kerr black holes the near-horizon metric can be written in the form of eq. (2.1) but these coordinates cannot be used as global coordinates. For the arguments used in this paper it is sufficient that the local patch near the horizon can always be described by the metric in eq. (2.1).

However, one should note that the metric structure given in eq. (2.1) does not completely fix the coordinates on constant  $r$  and  $v$  slices. The following two types of coordinate redefinitions are still allowed, which preserve the gauge choice in eq. (2.1):

---

<sup>15</sup>We shall be very brief in explaining our setup here. For details, we refer the readers to [14, 22, 40].

<sup>16</sup>The particular form of the metric in which we have extracted out an explicit factor of  $r^2$  in  $g_{vv}$  and a factor of  $r$  in  $g_{vi}$  components of the metric has previously been used in [22].

|    |   |
|----|---|
| 1. | The metric coefficients: a scalar function $X(r, v, x)$ , a vector function $\omega_i(r, v, x)$ and a tensor function $h_{ij}(r, v, x)$ . |
| 2. | Three differential operators: $\{\partial_r, \partial_v, \nabla_i\}$ acting on $X(r, v, x)$ , $\omega_i(r, v, x)$ and $h_{ij}(r, v, x)$ . |

**Table 1.** The basic building blocks.

1. We are allowed to do a coordinate transformation as

$$v \rightarrow \tilde{v} = f_1(x^i)v + f_2(x^i), \quad (2.2)$$

along with an appropriate redefinition of the coordinate  $r$  such that the form of the metric in eq. (2.1) remains invariant. Here  $f_1$  and  $f_2$  are arbitrary functions of the  $x^i$  coordinates. This coordinate transform essentially redefines the constant  $v$ -slicing of the horizon and hence, will be important in determining the form of the entropy current. We will actually consider a special sub-class of this residual freedom of coordinate transformation in the following sub-section extensively for our analysis in this paper.

2. We can also perform a relabelling of the null generators of the horizon by

$$x^i \rightarrow \tilde{x}^i = g^i(x^j), \quad (2.3)$$

which also keeps our metric gauge in eq. (2.1) intact. Additionally, this does not change the constant  $v$ -slicing of the horizon. It rather mixes only the  $x^i$  coordinates among themselves. Therefore, it represents the diffeomorphism invariance for the metric component  $h_{ij}$ , the induced metric on a co-dimension two constant  $r$  and  $v$  hyper-surface. This would, for example, enable us to convert all partial derivatives with respect to  $x_i$  coordinates to covariant derivatives  $\nabla_i$  compatible with the metric  $h_{ij}$ .

We are interested in obtaining an expression for the entropy density and entropy current density, both of which are defined on the horizon (i.e., on the  $r = 0$  hyper-surface). Hence, the entropy density and entropy current should be a scalar and a vector, respectively, under the diffeomorphism that mixes only the  $x^i$  coordinates among themselves as in eq. (2.3). Also, in our chosen gauge eq. (2.1), both are constructed out of the metric functions  $X$ ,  $\omega_i$ ,  $h_{ij}$  and their derivatives. Therefore, in our attempt to construct an entropy current for a diffeomorphism invariant theory of gravity, we identify the following quantities as our basic building blocks listed in table 1. One should note that in the list given in table 1, the scalar, vector, and tensor properties of different quantities are determined with respect to the spatial diffeomorphism that mixes only the  $x^i$  coordinates among themselves.

## 2.2 Stationarity and small deviation from it by dynamical fluctuations

As we have mentioned before, our analysis in this paper is limited to the perturbative approximation of working only up to linear order in the amplitude of fluctuations around

a stationary black hole configuration. In other words, we should be able to decompose the full metric as a sum of a stationary metric and a dynamical fluctuation with a small amplitude around it.<sup>17</sup> Working with the horizon adapted coordinates  $\{r, v, x\}$  described above in section 2.1, in this subsection, our aim is to make these notions more precise, namely what it means for space-time to be stationary and also what a non-stationary dynamical perturbation is.

It is well known that for a stationary black hole, we can always find one Killing vector such that it is null on the event horizon, a co-dimension one null hyper-surface. In other words, the event horizon becomes a Killing horizon for that Killing vector.<sup>18</sup> With the knowledge of this, we can learn about what specific restrictions does the requirement of stationarity imposes on the space-time metric given in eq. (2.1).

As was discussed before, we consider a special sub-class of the residual coordinate transformation freedom mentioned in eq. (2.2) with  $f_1(x^i) = \lambda$ , an arbitrary constant parameter and  $f_2(x^i) = 0$ . To be more precise, we get the following rescaling of the  $v$  and  $r$  coordinate

$$r \rightarrow \tilde{r} = \lambda r, \quad v \rightarrow \tilde{v} = \frac{v}{\lambda}, \quad (2.4)$$

with  $\lambda$  being some arbitrary constant. Following [14] and [22], we will call this scaling transformation as the boost transformation in the rest of this paper. The infinitesimal version of this transformation is generated by the following vector<sup>19</sup>

$$\xi = \xi^\mu \partial_\mu = v \partial_v - r \partial_r. \quad (2.5)$$

Under this coordinate transformation, the metric eq. (2.1) remains almost invariant apart from the scaling of the arguments ( $r$  and  $v$ ) of the metric functions  $X$ ,  $\omega_i$  and  $h_{ij}$ ,

$$ds^2 = 2d\tilde{v} d\tilde{r} - \tilde{r}^2 \tilde{X}(\tilde{r}, \tilde{v}, x) d\tilde{v}^2 + 2\tilde{r} \tilde{\omega}_i(\tilde{r}, \tilde{v}, x) d\tilde{v} dx^i + \tilde{h}_{ij}(\tilde{r}, \tilde{v}, x) dx^i dx^j, \quad (2.6)$$

where  $\tilde{X}(\tilde{r}, \tilde{v}, x) = X\left(\frac{\tilde{r}}{\lambda}, \lambda\tilde{v}, x\right)$ ,  $\tilde{\omega}_i(\tilde{r}, \tilde{v}, x) = \omega_i\left(\frac{\tilde{r}}{\lambda}, \lambda\tilde{v}, x\right)$ ,  $\tilde{h}_{ij}(\tilde{r}, \tilde{v}, x) = h_{ij}\left(\frac{\tilde{r}}{\lambda}, \lambda\tilde{v}, x\right)$ . It is then obvious that the metric remains invariant if all the metric functions  $X$ ,  $\omega_i$  and  $h_{ij}$  depend on the coordinates  $r$  and  $v$  only through their product  $rv$ . More precisely, when evaluated on the horizon, the vector field  $\xi$ , which reduces to the null generator of the horizon, will be a Killing vector for any space-time with the following metric

$$ds^2 = 2dv dr - r^2 X(rv, x) dv^2 + 2r \omega_i(rv, x) dv dx^i + h_{ij}(rv, x) dx^i dx^j. \quad (2.7)$$

Therefore, we learn that in our horizon adapted coordinates  $\{r, v, x^i\}$  any metric of the form given above in eq. (2.7) describes a stationary space-time.<sup>20</sup> In [22] (see appendix A

<sup>17</sup>We only consider fluctuations that preserve our metric gauge eq. ((2.1)).

<sup>18</sup>However, for a static black hole, we have a stronger requirement, i.e., the Killing vector, which becomes null on the event horizon, is the time translations at the asymptotic infinity.

<sup>19</sup>The norm of  $\xi$  is given by  $\xi^\mu \xi^\nu g_{\mu\nu} = -2rv - r^2 v^2 X(r, v, x)$ . So,  $\xi$  becomes null at the horizon located at  $r = 0$ , where it is proportional to the null generator ( $\partial_v$ ) of the horizon.

<sup>20</sup>It might seem strange to see that the stationary metric in eq. (2.7) has  $v$  dependent components. As explained in detail in appendix A of [40], we should remember that  $v$  is an affine parameter along  $\partial_v$ , the null generator of the horizon. However, the metric functions for a stationary black hole should be independent of the Killing coordinate. The latter is generated along the Killing vector ( $v\partial_v - r\partial_r$ ). This is not exactly equal to, but rather proportional to the  $\partial_v$ , even at the horizon  $r = 0$ .

in it), it was shown that up to possible coordinate transformations, this is the most general form of stationary space-time with a Killing horizon.

Having understood the most general form of any stationary metric, our next job will be to quantify the departure from stationarity due to a dynamical perturbation with small amplitude. We remember that our basic building blocks are the metric functions  $X$ ,  $\omega_i$  and  $h_{ij}$  and their  $\partial_r$ ,  $\partial_v$  and  $\nabla_i$  derivatives given in table 1. Let us now consider a generic covariant tensor with an expression of the form

$$\mathcal{A} \sim (\partial_r)^{m_r} (\partial_v)^{m_v} \mathcal{B}, \tag{2.8}$$

where  $\mathcal{B}$  collectively denotes the metric coefficients:  $X$ ,  $\omega_i$ ,  $h_{ij}$  or any other covariant tensor constructed out of them using the action of  $\nabla_i$  only.<sup>21</sup> If we evaluate  $\mathcal{B}$  for any equilibrium or stationary space-time configuration given by the metric eq. (2.7), it will be a function of the product  $rv$  and  $x$ , i.e.  $\mathcal{B}(r, v, x)|_{equil.} \sim \mathcal{B}(rv, x)$ . Therefore, it can be argued that whenever  $m_v > m_r$ , the expression of  $\mathcal{A}$  given in eq. (2.8) will vanish on the horizon ( $r = 0$ ) as it will have  $(m_v - m_r)$  factors of  $r$ . However, this will not be true if we are evaluating  $\mathcal{A}$  for a general dynamical metric where the functional dependence of metric coefficients ( $X$ ,  $\omega_i$ ,  $h_{ij}$ ) on  $r$  and  $v$  is not restricted only through the product of them. Hence, we conclude that whenever any expression of the form given in eq. (2.8) with  $m_v > m_r$  evaluates to a non-zero value, the corresponding space-time metric is dynamical,

$$\begin{aligned} \text{Equilibrium configuration} &\rightarrow (\partial_r)^{m_r} (\partial_v)^{m_v} \mathcal{B}(rv, x)|_{r=0} = 0 \text{ (for } m_v > m_r), \\ \text{Non-equilibrium configuration} &\rightarrow (\partial_r)^{m_r} (\partial_v)^{m_v} \mathcal{B}(r, v, x)|_{r=0} \neq 0 \text{ (for } m_v > m_r). \end{aligned} \tag{2.9}$$

We will use this as a criterion for stationary and dynamical space-times.

To explain the linearity of dynamical fluctuations to a stationary configurations, we decompose the full space-time metric  $g_{\mu\nu}$  as a sum of two parts

$$g_{\mu\nu} = g_{\mu\nu}^{eq} + \epsilon \delta g_{\mu\nu}, \tag{2.10}$$

where  $g_{\mu\nu}^{eq}$  is a stationary metric and  $\delta g_{\mu\nu}$  captures the dynamics away from equilibrium. The small parameter  $\epsilon$  denotes the amplitude of the dynamical perturbation. In the coordinate system that we are working with, such a decomposition will imply that all the metric functions could also be written accordingly as a sum of two contributions

$$\begin{aligned} X &= X^{eq}(rv, x) + \epsilon \delta X(r, v, x), \\ \omega_i &= \omega_i^{eq}(rv, x) + \epsilon \delta \omega_i(r, v, x), \\ h_{ij} &= h_{ij}^{eq}(rv, x) + \epsilon \delta h_{ij}(r, v, x). \end{aligned} \tag{2.11}$$

Furthermore, as a consequence, we can also argue that for any  $\mathcal{B}$ , some covariant expression constructed out of the metric functions and their  $\nabla_i$  derivative(s), can always be written as

$$\mathcal{B}(r, v, x) = \mathcal{B}^{eq}(rv, x) + \epsilon \delta \mathcal{B}(r, v, x). \tag{2.12}$$

---

<sup>21</sup>It should be noted that the possible appearances of  $\partial_r$  and  $\partial_v$  in  $\mathcal{A}$  are explicitly shown in eq. (2.8), i.e.  $\mathcal{B}$  does not contain any factors involving  $\partial_r$  or  $\partial_v$  acting on  $X$ ,  $\omega_i$ ,  $h_{ij}$ . To avoid clutter of indices, we have also suppressed the components of the tensors  $\mathcal{A}$  and  $\mathcal{B}$ .

|    |   |
|----|---|
| 1. | Any expression that has more $\partial_v$ -derivatives than $\partial_r$ -derivatives could be non-zero solely because of dynamics.   |
| 2. | The expression will be linear in the amplitude of dynamics when the extra $\partial_v$ derivatives (i.e. those $\partial_v$ 's which are not compensated or paired with $\partial_r$ 's) are all acting on a single covariant expression of the form of $\mathcal{B}$ , eq. (2.13). |
| 3. | Whenever the extra $\partial_v$ derivatives are distributed between different factors, the resulting expression (e.g. $\mathcal{A}_1$ in eq. (2.15)) would be non-linear in the amplitude of dynamics.  |

**Table 2.** Distribution of  $\partial_v$ -derivatives and non-linearities in amplitude expansion.

Following our arguments leading to eq. (2.9), it is obvious that if we operate  $(\partial_r^m \partial_v^{m+k})$  (for  $k > 0$ ) on such a  $\mathcal{B}$  given in eq. (2.12) and evaluate it on the horizon, it is only  $\delta\mathcal{B}$  that will contribute, rendering the expression linear in dynamics

$$\mathcal{A}|_{r=0} \sim (\partial_r^m \partial_v^{m+k}) \mathcal{B}|_{r=0} = \epsilon (\partial_r^m \partial_v^{m+k}) \delta\mathcal{B}|_{r=0} \sim \mathcal{O}(\epsilon), \quad (\text{for } k > 0). \quad (2.13)$$

Now, let us consider an expression of the form

$$\mathcal{A}_1 \sim (\partial_r^{m_1} \partial_v^{m_1+k_1} \mathcal{B}_1) (\partial_r^{m_2} \partial_v^{m_2+k_2} \mathcal{B}_2), \quad (\text{for } k_1, k_2 > 0), \quad (2.14)$$

where both  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are two different but arbitrary covariant tensor expressions admitting similar decomposition as  $\mathcal{B}$  in eq. (2.12). Following the similar set of arguments presented above, we could now see that both of the factors in the r.h.s. of eq. (2.14) will have non-zero contributions only when  $\mathcal{B}_1$  and  $\mathcal{B}_2$  are replaced by  $\delta\mathcal{B}_1$  and  $\delta\mathcal{B}_2$  respectively,

$$\mathcal{A}_1|_{r=0} \sim (\partial_r^{m_1} \partial_v^{m_1+k_1}) \mathcal{B}_1 (\partial_r^{m_2} \partial_v^{m_2+k_2}) \mathcal{B}_2|_{r=0} \sim \mathcal{O}(\epsilon^2), \quad (\text{for } k_1, k_2 > 0), \quad (2.15)$$

leading us to the conclusion that the expression  $\mathcal{A}_1$  will be non-linear (quadratic) in the amplitude of dynamical fluctuations.

We will summarise the main lessons that we have learnt regarding the classification of generic tensor structures built out of the metric coefficient functions and their derivatives, in terms of stationarity and dynamical fluctuations (linear or non-linear) around it in table 2.

### 2.3 Killing symmetry and boost weight of quantities

As we have mentioned in the introduction, we would be analyzing the equation of motion for the metric, and our goal is to write its most general form in terms of our building blocks given in table 1. The equation of motion in a theory of gravity is a rank two tensor with respect to the full set of diffeomorphism that mixes all the space-time coordinates. In this subsection, we shall briefly describe how we can relate a particular component of a tensor (when evaluated on the horizon) with the number and distribution of  $\partial_r$  and  $\partial_v$  derivatives acting on different factors.

In our horizon adapted coordinate system  $\{r, v, x\}$ , the most general stationary black hole solution as in eq. (2.7), admits a Killing vector

$$\xi = \xi^\mu \partial_\mu = (v\partial_v - r\partial_r). \quad (2.16)$$



As was mentioned in the previous subsection (see eq. (2.5)), this is also the generator of the boost transformation eq. (2.4), or more precisely of the infinitesimal version of it.

In order to study small dynamical fluctuations away from a stationary black hole configuration, it will be useful to see how various quantities transform under this scaling or boost transformation that generates the Killing symmetry of the stationary background. Furthermore, since we are viewing every quantity to be built out of the basic building blocks (table 1), it is sufficient to know the transformation property of them.

Let us define the power of  $\lambda$  by which any quantity transform under the boost transformation eq. (2.4), as the boost weight of that quantity

$$\mathcal{A} \rightarrow \tilde{\mathcal{A}} = \lambda^w \mathcal{A}, \quad \text{under } \{r \rightarrow \tilde{r} = \lambda r, v \rightarrow \tilde{v} = \lambda^{-1} v\} \Rightarrow \text{boost weight of } \mathcal{A} \text{ is } w. \quad (2.17)$$

The first thing we should note following this definition, is that the coordinates  $r$  and  $v$  have boost weights  $+1$  and  $-1$  respectively. Similarly all the metric functions:  $X$ ,  $\omega_i$  and  $h_{ij}$  have zero boost weight and only  $\partial_v$  and  $\partial_r$  are the operators that transform non-trivially under eq. (2.4)

$$\partial_r \rightarrow \partial_{\tilde{r}} = \lambda^{-1} \partial_r, \quad \partial_v \rightarrow \partial_{\tilde{v}} = \lambda \partial_v. \quad (2.18)$$

Hence, the operators  $\partial_r$  and  $\partial_v$  have boost weights  $-1$  and  $+1$  respectively, and the operator  $\nabla_i$  has zero boost weight when operated on  $X$ ,  $\omega_i$  and  $h_{ij}$ .

In other words, since  $\xi$  given in eq. (2.16) generates Killing symmetry of the stationary background, the Lie derivative with respect to  $\xi$  (denoted as  $\mathcal{L}_\xi$ ) when operated on any covariant tensor that is constructed out of the stationary metric eq. (2.7), will vanish. Therefore, a non-zero value of this  $\mathcal{L}_\xi$  operator acting on any covariant tensor built out of metric components, is indicative of non-stationary or out-of-equilibrium dynamics. The action of  $\mathcal{L}_\xi$  on any covariant tensor  $S_{\alpha_1 \alpha_2 \dots \alpha_k}$  is given by

$$\begin{aligned} \mathcal{L}_\xi S_{\alpha_1 \alpha_2 \dots \alpha_k} &= \xi^\beta \partial_\beta S_{\alpha_1 \alpha_2 \dots \alpha_k} + \left( \partial_{\alpha_1} \xi^\beta \right) S_{\beta \alpha_2 \dots \alpha_k} + \left( \partial_{\alpha_2} \xi^\beta \right) S_{\alpha_1 \beta \alpha_3 \dots \alpha_k} + \dots \\ &+ \left( \partial_{\alpha_k} \xi^\beta \right) S_{\alpha_1 \alpha_2 \dots \beta}, \end{aligned} \quad (2.19)$$

and when we evaluate this in our coordinate system, eq. (2.1), with  $\xi = (v\partial_v - r\partial_r)$ , we will obtain the following general form

$$\mathcal{L}_\xi S_{\alpha_1 \alpha_2 \dots \alpha_k} = [w + (v\partial_v - r\partial_r)] S_{\alpha_1 \alpha_2 \dots \alpha_k} \quad (2.20)$$

where we identify the quantity  $w$  as the boost weight of the covariant tensor  $S_{\alpha_1 \alpha_2 \dots \alpha_k}$  as

$$\begin{aligned} w &\equiv \text{net boost weight of a covariant tensor (with all indices lowered)} \\ &= \text{number of lower } v \text{ indices} - \text{number of lower } r \text{ indices.} \end{aligned} \quad (2.21)$$

A more detailed explanation regarding the idea of boost weight of different quantities is provided in appendix B. Following eq. (2.21), we can check that under  $\mathcal{L}_\xi$ , the operators  $\partial_v$  and  $\partial_r$  transform like the  $v$  and  $r$  components of a lower-indexed vector  $\partial_\mu$  and therefore have boost weights  $+1$  and  $-1$  respectively, as expected. We also learn that if we evaluate a particular component of a covariant tensor on the horizon  $r = 0$ , the number of lower  $v$

|    |   |
|----|---|
| 1. | We have chosen a special coordinate system adapted to the dynamical null horizon. The coordinates on the horizon are $v$ (parameter along the null generator) and $\{x^i\}$ (all the spatial coordinates), whereas the coordinate away from the horizon is denoted by $r$ , see eq. (2.1). The coordinates $v$ and $r$ are also affinely parametrized.  |
| 2. | In this coordinate system, if we evaluate any component of a covariant tensor restricted on the horizon, schematically, it will have a structure with some number of $\partial_r$ , $\partial_v$ and $\nabla_i$ operators acting on the functions appearing in the metric eq. (2.1): $(X, \omega_i, \text{ and } h_{ij})$ or product of such structures.  |
| 3. | For an expression, we defined the boost weight $w$ of it to be the difference between the number of $\partial_v$ 's and $\partial_r$ 's, and then we have argued that for a tensor component, this is the same as the difference between the number of lower $v$ (or upper $r$ ) and the lower $r$ (or upper $v$ ) indices, see eq. (2.17) and eq. (2.21).                                      |
| 4. | Furthermore, the boost weight of a tensor component is nothing but the multiplicative factor it will have under the action of Lie derivative $\mathcal{L}_\xi$ along the vector $\xi (= v\partial_v - r\partial_r)$ , the general Killing vector field whenever the metric has a Killing horizon, see eq. (2.20).   |
| 5. | Finally, any expression with positive boost weight is non-zero only when evaluated on a dynamical, non-stationary metric. It is linear in the amplitude of the dynamical fluctuation provided all the excess $\partial_v$ 's are acting on a single function, and it is not a product of more than one expression, each having positive boost weight, see eq. (2.9), eq. (2.13) and eq. (2.15). |

**Table 3.** Summary of the important lessons so far that will serve as guidelines for later computations.

indices corresponds to the number of  $\partial_v$ 's (and the number of lower  $r$  indices corresponds to the number of  $\partial_r$ 's) acting on metric functions  $(X, \omega_i, \text{ and } h_{ij})$  with boost weight zero.<sup>22</sup>

Before we proceed, let us summarise the main points of this section in table 3, which we are going to use in the rest of the paper.

#### 2.4 Significance of the structure of certain components of the equations of motion for a proof of the second law

Until now, we have summarized the basic concepts in relevance to our work. In the following, we will briefly review the arguments required to justify a local version of the second law of thermodynamics for linearized dynamical perturbations around a stationary black

<sup>22</sup>If we allow ourselves to go away from the horizon, positive boost weights could also be absorbed by factors of  $r$ . For example, the  $vv$  component of the metric,  $g_{vv}$ , which is one example of a rank-2 covariant tensor component with boost weight 2, has a factor of  $r^2$  multiplying a boost-invariant function  $X$ . The same is true for  $g_{vi}$  and  $g_{ij}$  the  $vi$  and  $ij$  components of the metric respectively. In fact, this is why all the metric functions  $(X, \omega_i \text{ and } h_{ij})$  are of zero boost weights.

hole configuration. As we have mentioned before, to achieve this, we will be looking at the structure of a particular component of the equations of motion, to be more specific, the ‘ $vv$ ’-component of it, i.e.,  $E_{vv}$ . Thereby, we will also explain how significant the role of the structure of  $E_{vv}$  is to establish the second law.

The notion of entropy for stationary black holes in arbitrary diffeomorphism invariant theory of gravity is associated with the Wald entropy defined as

$$S_W = -2\pi \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} E_R^{\alpha_1\alpha_2\alpha_3\alpha_4} \epsilon_{\alpha_1\alpha_2} \epsilon_{\alpha_3\alpha_4} = \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} \left(1 + s_w^{\text{HD}}\right), \quad (2.22)$$

where  $E_R^{\alpha_1\alpha_2\alpha_3\alpha_4}$  is given by eq. (3.18),  $\epsilon_{\alpha_1\alpha_2}$  are the bi-normal to the constant  $v$ -slices of the horizon,  $\sqrt{h}$  is the determinant of the induced metric  $h_{ij}$  on the horizon and  $s_w^{\text{HD}}$  is the contribution to the Wald entropy density from the higher derivative part of the Lagrangian leaving the Einstein-Hilbert part aside. This definition of the Wald entropy is, by construction, consistent with the first law of thermodynamics. From the last expression on the r.h.s. above it is clear that for Einstein’s gravity  $S_W$  reduces to the well known formula of computing area of the spatial slices of the horizon.

Starting from the Wald entropy formula in eq. (2.22), one can then write down an entropy function which is the out-of-equilibrium extension of it as the following

$$S_{\text{total}} = S_W + \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} s_{\text{cor}} = \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} \left(1 + s_w^{\text{HD}} + s_{\text{cor}}\right), \quad (2.23)$$

where  $s_{\text{cor}}$  are the corrections to the Wald entropy density due to non-stationary dynamics, and it certainly includes the JKM ambiguities that we mentioned before. Once the dynamics is switched off,  $s_{\text{cor}}$  vanishes and therefore  $S_{\text{total}}$  reduces to  $S_W$ , i.e.

$$s_{\text{cor}}|_{\text{equilibrium}} = 0.$$

The main goal of [14] was to constrain the form of  $s_{\text{cor}}$ , and hence to fix the JKM ambiguities, by requiring  $S_{\text{total}}$  to satisfy a linearized version of the second law. More precisely, the strategy would be to determine  $s_{\text{cor}}$  such that  $\partial_v S_{\text{total}} \geq 0$ . One defines a local expansion parameter  $\vartheta$ ,<sup>23</sup> as follows

$$\frac{\partial S_{\text{total}}}{\partial v} \equiv \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} \vartheta. \quad (2.24)$$

Now, one way to prove  $\partial_v S_{\text{total}} \geq 0$  is to show that  $\vartheta \geq 0$  for all  $v \geq 0$ . The trick one uses for this is to argue that  $\vartheta$  is a monotonically decreasing function of  $v$  (i.e.  $\partial_v \vartheta \leq 0$ ), along with the restriction that at asymptotic future it approaches to zero, meaning  $\vartheta \rightarrow 0$  for  $v \rightarrow \infty$ . This is actually motivated by the physical situation that we are examining here. After getting slightly perturbed by a small dynamical fluctuation, the initial stationary black hole solution finally settles down to another stationary black hole solution in the future.

---

<sup>23</sup>The corresponding quantity in Einstein’s gravity, say  $\vartheta_E$ , is the expansion parameter for the null congruence  $\partial_v$ . The evolution of this parameter  $\vartheta$  with  $v$  is governed by the Raychaudhuri equation and this plays a crucial role in proving the area increase theorem in general relativity.

Now calculating  $\partial_v \vartheta$  in our chosen metric gauge by using eq. (2.24) as the definition of  $\vartheta$  and  $S_{\text{total}}$  from eq. (2.23), we obtain<sup>24</sup>

$$\partial_v \vartheta = -R_{vv} + \partial_v \left( \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \left( s_w^{\text{HD}} + s_{\text{cor}} \right) \right) \right) + \mathcal{O}(\epsilon^2). \quad (2.25)$$

Next, we should remember that the  $vv$ -component of the equation of motion  $E_{\mu\nu}$  can be written as

$$-E_{vv} = R_{vv} + E_{vv}^{\text{HD}} = T_{vv} \quad (2.26)$$

where  $E_{vv}$  is obtained from just the gravity part of the Lagrangian,<sup>25</sup>  $E_{vv}^{\text{HD}}$  only contains the contribution from the higher derivative part of the gravity Lagrangian excluding the Einstein-Hilbert term,  $T_{vv}$  is the  $vv$ -component of the stress energy tensor  $T_{\mu\nu}$  coming from the matter sector. Substituting this for  $R_{vv}$  in eq. (2.25) we get

$$\partial_v \vartheta = E_{vv}^{\text{HD}} + \partial_v \left( \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \left( s_w^{\text{HD}} + s_{\text{cor}} \right) \right) \right) - T_{vv} + \mathcal{O}(\epsilon^2). \quad (2.27)$$

Furthermore, we also assume that the stress-energy tensor, for the matter part, satisfies the null energy condition, which, when translated to our choice of metric gauge and  $\xi^\mu$  eq. (2.16), takes the form of  $T_{vv} \geq 0$ . Thus from eq. (2.27) we see that the imposition of null energy condition on the stress-energy tensor for the matter perturbation actually contributes in favor of making  $\partial_v \vartheta \leq 0$ . In that sense, with the assumption of null energy condition,  $T_{vv}$  trivially drops out from our analysis.

We are considering dynamical perturbations to a stationary black hole configuration such that the metric of the equilibrium background gets corrected at  $\mathcal{O}(\epsilon)$ , see eq. (2.10). Consequently, we must also take into consideration the fact that terms linear in the amplitude of the dynamical fluctuation can come with both signs depending on the sign of  $\epsilon$ . On the other hand, on equilibrium configurations, i.e.  $\epsilon \rightarrow 0$ , we must get  $\partial_v \vartheta \rightarrow 0$ . Therefore, the only way in which  $\partial_v \vartheta \leq 0$  can be met while working up to linear order  $\mathcal{O}(\epsilon)$ , is to show that

$$\partial_v \vartheta = \mathcal{O}(\epsilon^2), \quad (2.28)$$

or in other words,  $\partial_v \vartheta$  should vanish up to linear order in the amplitude expansion.

As a result it is obvious from eq. (2.27) that eq. (2.28) will be satisfied if the following holds true

$$E_{vv}^{\text{HD}} = -\partial_v \left( \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \left( s_w^{\text{HD}} + s_{\text{cor}} \right) \right) \right) + \mathcal{O}(\epsilon^2). \quad (2.29)$$

<sup>24</sup>We have used here the fact that  $\partial_v \vartheta_E = -R_{vv} + \mathcal{O}(\epsilon^2)$ , where  $\vartheta_E$  is the corresponding quantity for  $\vartheta$  when computed from just the Einstein-Hilbert term in the action.

<sup>25</sup>The minus sign on the l.h.s. is there to make it consistent with the convention used in eq. (1.5). With this choice of convention, for Einstein gravity we get

$$E_{vv} = -R_{vv} = -T_{vv},$$

and this also defines the sign convention of the stress tensor such that the null energy condition works out as  $T_{vv} \geq 0$ . We should also remember that while writing the  $vv$ -component of the equations of motion in the form of eq. (2.26), we have evaluated it on the horizon which is located at  $r = 0$ . For example, from our choice of metric given in eq. (2.1) it is clear that  $g_{vv}|_{r=0} = 0$ .

This relation in eq. (2.29) highlights the significance of  $E_{vv}^{\text{HD}}$  in arriving at a proof of the linearized version of the second law presented in the form of eq. (2.28).

At this point, it is worth making a comment regarding the dependance of  $T_{vv}$  on the order of  $\epsilon$  and its significance for eq. (2.28). As mentioned before in section 2.2 (see eq. (2.10)), the correction to the stationary background metric due to dynamical perturbations is of  $\mathcal{O}(\epsilon)$ . Let the matter fields that induce the perturbation are of order  $\epsilon$ . Now, following our assumptions the stress tensor due to the matter perturbation satisfies null energy condition  $T_{vv} \geq 0$ , and vanishes in the equilibrium limit  $\epsilon \rightarrow 0$ . As a consequence, based on general arguments given above for  $\partial_v \vartheta$ , we learn that  $T_{vv} \sim \mathcal{O}(\epsilon^2)$ ; see section (2.1) and appendix B of [22] for a detailed discussion on this point. Therefore, up to our working precision of  $\mathcal{O}(\epsilon)$ ,  $T_{vv}$  can be dropped from the r.h.s. of eq. (2.27) and the equality in eq. (2.28) is justified once the restriction on  $E_{vv}$  in eq. (2.29) is met.<sup>26</sup>

In other words, we can view eq. (2.29) as a constraint on the off-shell structure of  $E_{vv}$ , which can be summarised as follows:

*In any diffeomorphism invariant theory of gravity, working in the chosen metric gauge eq. (2.1) and up to linear order in the amplitude expansion, if we can show that the structure of the  $vv$ -component of equations of motion can always be written as the r.h.s. of eq. (2.29), we will be able to construct a proof of the linearized version of the second law by satisfying eq. (2.28).*

In view of the statement made above, we are now convinced that the primary goal of this current paper is indeed justified. Once we are able to show that in any arbitrary higher derivative theory of gravity  $E_{vv}$  can always be written as eq. (1.7), which we write here again for convenience

$$E_{vv}|_{r=0} = \partial_v \left[ \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \mathcal{J}^v \right) + \nabla_i \mathcal{J}^i \right] + \mathcal{O}(\epsilon^2). \tag{2.30}$$

Once this is compared with eq. (2.29), one can straightforwardly obtain the components of entropy current (i.e.  $\mathcal{J}^v$  and  $\mathcal{J}^i$ ). Furthermore, we are also convinced that when the equilibrium limit is taken  $\mathcal{J}^v$  should reproduce the equilibrium Wald entropy density<sup>27</sup>

$$\mathcal{J}^v|_{\text{equilibrium}} = (1 + s_w^{\text{HD}}). \tag{2.31}$$

---

<sup>26</sup>We can instead have a situation, possibly a more physically appealing one, where the corrections to the space-time geometry are generated due to backreactions from a  $\mathcal{O}(\epsilon^2)$  stress tensor representing the non-stationary perturbation. The metric which is sourced by this stress tensor would then receive its first correction at  $\mathcal{O}(\epsilon^2)$  and not at  $\mathcal{O}(\epsilon)$ . The corrections to the metric at  $\mathcal{O}(\epsilon)$  would trivially vanish. In other words, the dynamical fluctuations of the metric will get non-trivial contributions starting at  $\mathcal{O}(\epsilon^2)$ . This is certainly true for minimal coupling of the matter sector, which we will be considering in this paper. However, as it has been discussed in section (2.1) of [22], even in that situation, the statement made in eq. (2.29) would be valid. In that case eq. (2.28) would become trivial at  $\mathcal{O}(\epsilon)$ . The first non-trivial contribution for this would occur at  $\mathcal{O}(\epsilon^2)$  as the following:  $\partial_v \vartheta|_{\mathcal{O}(\epsilon^2)} \leq 0$ .

<sup>27</sup>Since  $E_{vv}$  also includes the contributions from the Einstein-Hilbert term, in  $\mathcal{J}^v$  we get an extra additive factor of one in addition to  $s_w^{\text{HD}}$ . This extra factor of one is actually the Area of the horizon piece.

## 2.5 An important identity relating the equations of motion with the Noether charge for diffeomorphism

In the previous sub-sections, we have established the importance of a certain desired structure of the  $vv$ -component of the equations of motion, denoted by  $E_{vv}$ , for our work. In the following, we will write down a key equation that will help us construct the entropy current from  $E_{vv}$  by relating the latter to the Noether charge under diffeomorphism. This relation is, in fact, an identity that is true for any diffeomorphism invariant theory of gravity. We will follow the construction of Noether currents and charges corresponding to the diffeomorphism invariance in such a theory of gravity as prescribed in [8].<sup>28</sup>

In a diffeomorphism invariant theory of gravity with an action  $S = \int \sqrt{-g} L$ , the variation of the Lagrangian due to an arbitrary variation  $\delta g_{\mu\nu}$  of the metric, is given by

$$\delta [\sqrt{-g} L] = \sqrt{-g} E^{\mu\nu} \delta g_{\mu\nu} + \sqrt{-g} D_\mu \Theta^\mu[\delta g], \quad (2.32)$$

where  $E^{\mu\nu}$  is the equation of motion,  $\Theta^\mu[\delta g]$  is the total derivative term generated in the variation of the Lagrangian and following our convention in this paper,  $D_\mu$  is the covariant derivative with respect to the metric  $g_{\mu\nu}$ .

Next we consider that  $\delta g_{\mu\nu}$  is produced due to an infinitesimal coordinate transformation generating the diffeomorphism  $x^\mu \rightarrow x^\mu + \zeta^\mu$ ,

$$\delta g_{\mu\nu} = D_\mu \zeta_\nu + D_\nu \zeta_\mu, \quad (2.33)$$

for which we obtain

$$\delta [\sqrt{-g} L] = \sqrt{-g} D_\mu (\zeta^\mu L). \quad (2.34)$$

Substituting this back in eq. (2.32) and cancelling the  $\sqrt{-g}$  factor on both sides we get

$$D_\mu (\zeta^\mu L - 2E^{\mu\nu} \zeta_\nu - \Theta^\mu[\partial\zeta]) = -2\zeta_\nu D_\mu E^{\mu\nu}. \quad (2.35)$$

Now we make a choice for  $\zeta^\mu$  such that it is non-zero only in a small region  $\mathcal{R}$ . With this, if one integrates the both sides of eq. (2.35) over full space-time, the l.h.s. will vanish since it would integrate to a pure boundary term at infinity where  $\zeta$  vanishes. Hence, we obtain

$$\int_{\text{full space-time}} \zeta_\nu D_\mu E^{\mu\nu} = 0, \quad (2.36)$$

which is true for any  $\zeta$  as long as it has a non-zero support over a finite region  $\mathcal{R}$ . Therefore, we see that the following relation identically holds

$$D_\mu E^{\mu\nu} = 0. \quad (2.37)$$

We should note that this is essentially the Bianchi identity corresponding to any diffeomorphism invariant theory.

---

<sup>28</sup>Before we proceed, let us emphasize that this is a well-known identity in gravitational theories and here we are just stating and reproducing it for the sake of completeness.

Let us now substitute the identity eq. (2.37) in eq. (2.35) and allow  $\zeta^\mu$  to be any vector without any restriction on its support. We obtain that the vector  $\zeta^\mu L - 2E^{\mu\nu}\zeta_\nu - \Theta^\mu(\partial\zeta)$  is identically conserved on any geometry for any  $\zeta$

$$D_\mu (\zeta^\mu L - 2E^{\mu\nu}\zeta_\nu - \Theta^\mu[\partial\zeta]) = 0. \quad (2.38)$$

Next, we use the result that an identically conserved vector can itself be always written as the divergence of an arbitrary antisymmetric rank-2 tensor. Hence one can argue that locally it is always true that

$$\zeta^\mu L - 2E^{\mu\nu}\zeta_\nu - \Theta^\mu = -D_\nu Q^{\mu\nu}, \quad (2.39)$$

for some arbitrary antisymmetric tensor  $Q^{\mu\nu}$ .<sup>29</sup>

We can contract the free index on both sides of eq. (2.39) with another  $\zeta_\mu$  to get

$$\zeta^2 L - \Theta^\mu \zeta_\mu + \zeta_\mu D_\nu Q^{\mu\nu} = 2\zeta^\mu \zeta^\nu E_{\mu\nu}. \quad (2.40)$$

Although eq. (2.40) is true for any  $\zeta^\mu$ , for our purpose, we will choose  $\zeta^\mu$  to be  $\xi = v\partial_v - r\partial_r$  and evaluate both sides of eq. (2.40) on the horizon  $r = 0$ ,

$$\begin{aligned} (2\xi^\mu \xi^\nu E_{\mu\nu})|_{r=0} &= 2v^2 E_{vv}|_{r=0} \\ (\xi^2 L - \Theta \cdot \xi + \xi_\mu D_\nu Q^{\mu\nu})|_{r=0} &= v(-\Theta^r + D_\nu Q^{r\nu})|_{r=0}. \end{aligned} \quad (2.41)$$

Finally, cancelling one factor of  $v$  from both sides, we arrive at

$$2v E_{vv}|_{r=0} = (-\Theta^r + D_\nu Q^{r\nu})|_{r=0}. \quad (2.42)$$

Equation (2.42) constitutes the main relation that we will use in section 3 to prove the existence of an entropy current in a diffeomorphism invariant theory of gravity.

### 3 A proof of the existence of an entropy current

Following the fundamental concepts developed in section 2, in this section, we will carry out an extensive analysis to derive an expression of the entropy current for dynamical black holes in any diffeomorphism invariant theory of gravity up to linearized order in the amplitude of the out-of-equilibrium fluctuations. In practice, we aim to establish that a specific component of the equations of motion, namely  $E_{vv}$ , in such a theory can always be recast in a form similar to the r.h.s. of eq. (1.7), and thereby obtaining the form of  $\mathcal{J}^v$  and  $\mathcal{J}^i$ . As we have already discussed in section 2.4, showing eq. (1.7) is equivalent to argue for a linearized version of the second law involving black hole dynamics.

The starting point of our calculation would be to consider the important relation, eq. (2.42), that was derived in section 2.5. Here we re-write that equation again for the sake of convenience,

$$2v E_{vv}|_{r=0} = (-\Theta^r + D_\nu Q^{r\nu})|_{r=0}. \quad (3.1)$$

---

<sup>29</sup>The negative sign on the r.h.s. of eq. (2.39) is due to our convention and could be understood as a part of the definition of  $Q^{\mu\nu}$  through this equation.

With this at hand, our main job now is to establish that the expression for  $E_{vv}$  on the r.h.s. of eq. (3.1) can always be rearranged as the following

$$E_{vv}|_{r=0} = \partial_v \left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right] + \mathcal{O}(\epsilon^2), \quad (3.2)$$

which is the same as eq. (1.7).

For the sake of convenience of the reader, in the following, we will separate the material to be presented into two sub-sections. To start with, in section 3.1 we will provide a brief schematic overview of the proof. The primary purpose of this would be to convey the essential ingredients of the analysis postponing the details of technical computations to the following section 3.2.

### 3.1 A brief and schematic sketch of the proof highlighting the basic strategy

In this sub-section, our primary goal is to provide an illustrative sketch of the arguments justifying the construction of a local entropy current in an arbitrary higher derivative theory of gravity. Before we dive into the technical computations presented in the following sub-section, it will be beneficial to develop an overall and intuitive understanding of the line of logical sequence.

**Basic strategy.** We start with a brief explanation of the basic strategy as follows. Since a particular coordinate system, as in eq. (2.1), has been chosen for our analysis, it is worth noting the explicit appearance of the coordinate  $v$  on the l.h.s. of eq. (3.1). The off-shell structure of  $E_{vv}$  cannot involve any factor of  $v$  except through the metric functions ( $X, \omega_i, h_{ij}$ ) or their derivatives. Therefore, our primary task going ahead would be to figure out the explicit  $v$ -dependence in each of the two terms appearing on the r.h.s. of eq. (3.1) so that we can compare the terms with equal powers of  $v$  on both sides of it. This, in turn, would produce the desired structure of  $E_{vv}$  that we are looking for.

We notice that on the r.h.s. of eq. (3.1) two quantities determine the structure of  $E_{vv}$  when evaluated on the horizon. One of them is the  $r$ -component of  $\theta^\mu$ , which is generated as a total derivative term in the variation of the Lagrangian under a diffeomorphism, eq. (2.32). The second term involves  $Q^{\mu\nu}$ , which is defined in eq. (2.39). In order to extract out the explicit  $v$ -dependence in both of these terms, we will use their corresponding definitions outlined in the construction of the Noether charge for diffeomorphism invariant gravity Lagrangians, [8]. These definitions are given in eq. (3.16) and eq. (3.17) for  $\Theta^\mu$ , and in eq. (3.60) for  $Q^{\mu\nu}$ .

Furthermore, we will also need to consider the Killing symmetry of the stationary background geometry, eq. (2.7), generated by the null generator of the horizon  $\xi^\mu$  eq. (2.5), and its breaking by a non-stationary perturbation in the linearised amplitude approximation (see eq. (2.11)). Since  $\xi^\mu$  also generates the boost transformation eq. (2.4), we will have to figure out how each of the two terms on the r.h.s. of eq. (3.1) transforms under it, i.e., the boost weights for each of them.

One crucial element for our computations would be determining the general structure of an arbitrary covariant tensor quantity with a given boost weight. At this point, it should



be emphasized that every tensorial quantity appearing in our analysis should be considered as if they are built out of the basic building blocks as given in table 1. In section 2.3 we have already learned about how to assign boost weights for the basic building blocks as well as for any generic tensor built out of them by counting the number of  $r$  and  $v$  components of that particular tensor, see eq. (2.21).

Once we know how to assign boost weight for a generic covariant tensor, we can easily determine its generic structure. For that, we need to analyze the distribution of  $\partial_v$ -derivatives carefully. The fact that we will be working in the linearised amplitude approximation will play a crucial role in executing this. We should also remember another related and equally important aspect, that is, eq. (3.1) has to be evaluated on the horizon located at  $r = 0$ . The guiding principles which are important in this context are summarised in table 2 and table 3. Following these guidelines, we can compute the generic structure of any covariant tensor with a positive boost weight, and the result is given in eq. (3.14). The detailed arguments for deriving this relation would be discussed in section 3.2.1. We will call this relation as “Result: 1” and it will be repeatedly used in our calculations.

Next, we would directly implement the strategy outlined above for both the terms on the r.h.s. of eq. (3.1) one at a time. Using the definitions of  $\Theta^\mu$  and  $Q^{\mu\nu}$  as given in eq. (3.16) and eq. (3.60) respectively, we will analyze every term that contributes to each of them. Finally, substituting the resulting expressions in eq. (2.42), we will derive the desired structure for  $E_{vv}$  as mentioned in eq. (1.7).

**Schematic illustration of the proof.** As written above in eq. (3.2), the principal quantity that we are going to examine is  $E_{vv}$ , the  $vv$  component of the equation of motion. Following the criteria set in eq. (2.21) and table 3,  $E_{vv}$  is a tensor component with boost weight +2. Following the arguments presented in [14] and also reviewed in section (2.2.2) of [22], we have seen that  $E_{vv}$  being a term with boost weight +2 could be expressed, up to linear order in the amplitude of the dynamical fluctuations, as

$$E_{vv} \sim \partial_v^2 C_{(0)} + A_{(0)} \partial_v^2 B_{(0)} + \mathcal{O}(\epsilon^2). \tag{3.3}$$

Here we have used the convention that the subscripts in  $A_{(0)}$ ,  $B_{(0)}$  and  $C_{(0)}$  denote the boost weights of these quantities, and this will be followed throughout in this paper. It is also essential to keep in mind that the purpose of writing eq. (3.3) is to highlight the structural form on the r.h.s., ignoring specific numerical factors, possible index contraction, and factors of  $\sqrt{\hbar}$  to avoid the clutter.

Furthermore, for a more elaborate derivation of eq. (3.3), one should compare it with eq. (2.22) in [22]. This would also emphasise an important difference in the structures of the two terms on the r.h.s. of eq. (3.3). In general, for a given theory of higher derivative gravity the quantities  $A_{(0)}$ ,  $B_{(0)}$  and  $C_{(0)}$  should be thought of as they are built out of the metric functions  $(X, \omega_i, h_{ij})$  or their derivatives with possible index contractions such that the resulting boost weights are zero for all of them. However,  $C_{(0)}$  is always a product of two terms with non-zero boost weights equal in magnitude but opposite in signs as shown below

$$C_{(0)} \sim \sum_k \tilde{T}_{(-k)} \partial_v^{(k)} T_{(0)}. \tag{3.4}$$

It is obvious from this equation that  $C_{(0)}$  has boost weight zero, but it is product of two quantities  $\tilde{T}_{(-k)}$  and  $\partial_v^{(k)}T_{(0)}$  which have individual boost weights as  $-k$  and  $+k$  respectively. However, the other term on the r.h.s. of eq. (3.3) involves  $A_{(0)}$  and  $B_{(0)}$  which could not be written like this. We name the terms of the type  $C_{(0)}$  as ‘JKM-type’ terms and the terms of the form  $A_{(0)}\partial_v^2B_{(0)}$  as the ‘zero-boost’ terms.<sup>30</sup> Therefore, it follows from eq. (3.3) that any tensor component with boost weight  $+2$ , such as  $E_{vv}$ , must have a set of such ‘zero-boost’ terms plus  $\partial_v^2$  acting on ‘JKM type’ terms. The statements mentioned above actually follows more generally from the important result that determines the general structure of any covariant tensor quantity with positive boost-weight, as given in eq. (3.14) named as “Result: 1”.

Our goal in this paper is to show that the ‘zero-boost’ terms in  $E_{vv}$  could also be written as<sup>31</sup>

$$A_{(0)}\partial_v^2B_{(0)} \sim \partial_v^2\mathcal{Q}_{(0)} + \partial_v\left(\nabla_i\mathcal{Q}_{(1)}^i\right) + \mathcal{O}\left(\epsilon^2\right), \tag{3.5}$$

where  $\mathcal{Q}_{(0)}$  is a quantity with vanishing boost weight and  $\mathcal{Q}_{(1)}^i$  is a quantity with boost weight  $+1$  with a spatial index  $i = 1, 2, \dots, D - 2$ . Also,  $\nabla_i$  is the covariant derivative associated with the induced metric  $h_{ij}$  on the co-dimension two spatial slices of the horizon. With this eq. (3.3) becomes

$$E_{vv} \sim \partial_v^2(C_{(0)} + \mathcal{Q}_{(0)}) + \partial_v\left(\nabla_i\mathcal{Q}_{(1)}^i\right) + \mathcal{O}(\epsilon^2). \tag{3.6}$$

In the following sub-section, we will carry out a detailed analysis to arrive at eq. (3.72) which is actually a more precise version of eq. (3.6). It should also be noted that the presence of the first term on the r.h.s. of eq. (3.6) was already explained in [14]. The actual new result in this current paper is to argue that in addition to the first term on the r.h.s., the second term on the r.h.s. can also, in general, be there in  $E_{vv}$ . This new term was identified with an entropy current in [22].

In the following we will present a rough sketch for the arguments to argue that eq. (3.6) is indeed true. For this we will use equation eq. (3.1) and analyze both the terms on the r.h.s. of this equation individually.

- From the construction of Noether’s charge for diffeomorphism as pointed out in [8], we know that both  $\Theta^r$  and  $Q^{r\mu}$  are linear in  $\xi^\mu\partial_\mu = v\partial_v - r\partial_r$ . The corresponding expressions justifying this can be found in eq. (3.24) for  $\Theta^r$  and eq. (3.60) for  $Q^{r\mu}$ .
- Since  $\xi$  is the only source of explicit dependence on  $v$ , both  $\Theta^r$  and  $Q^{r\mu}$ , being linear in  $\xi^\mu$ , will have the following structure

$$\Theta^r|_{r=0} = \Theta_{(1)} + v\Theta_{(2)}, \quad Q^{rv}|_{r=0} = Q_{(0)} + vQ_{(1)}, \quad Q^{ri}|_{r=0} = J_{(1)}^i + v\partial_v\tilde{J}_{(1)}^i. \tag{3.7}$$

---

<sup>30</sup>The fact that the second term on the r.h.s. of eq. (3.3) has been given the name ‘zero-boost’ terms should not be confused with the fact that this term  $A_{(0)}\partial_v^2B_{(0)}$  actually has a non-zero boost weight. To see the origin of this nomenclature we refer the reader to section (2.2.2) and section (2.2.3) in [22]. Both the terms  $A_{(0)}$  and  $B_{(0)}$  do not have any  $v$  or  $r$  derivatives in them which is not the case for  $C_{(0)}$ .

<sup>31</sup>This is very schematic, and there have to be relevant factors of  $\sqrt{h}$  to make everything consistent, but in this sub-section, since we are aiming at a heuristic argument we shall be ignoring all these subtleties.

Here also, the subscripts in the r.h.s. denote the boost weights of the corresponding quantities. In the last equation, we have also used the fact that any spatial current with boost weight +2 could always be expressed as  $\partial_v$  acting on a current term with boost weight +1, up to corrections non-linear in the amplitude of the dynamics. We have suppressed some index structures to avoid cluttering in the presentation.

- Using the fact that  $\Theta^\mu$  could be expressed in terms of  $\mathcal{L}_\xi$  (and a single derivative of it)<sup>32</sup> acting on covariant tensors (see the arguments leading to eq. (3.24) in section 3.2.2), one could give further structure to  $\Theta^r$ . Actually, as discussed in appendix C, one can show that

$$\Theta_{(2)} = \partial_v \Theta_{(1)} + \partial_v^2 \mathcal{M}_{(0)}, \tag{3.8}$$

where  $\mathcal{M}_{(0)}$  is a JKM-type term, i.e. a boost invariant term and also a product of two factors each of which has opposite boost weights.

- Now, substituting eq. (3.7) and eq. (3.8) in eq. (3.1) we find

$$2v E_{vv}|_{r=0} = \left[ -\Theta_{(1)} + Q_{(1)} + \partial_v Q_{(0)} + \nabla_i J_{(1)}^i \right] + v \partial_v \left[ -\Theta_{(1)} - \partial_v \mathcal{M}_{(0)} + Q_{(1)} + \nabla_i \tilde{J}_{(1)}^i \right] + \mathcal{O}(\epsilon^2). \tag{3.9}$$

Since the commutator of  $\partial_v$  and  $\nabla_i$  is of boost weight +1, it has been neglected in the last line where it is acting on  $\tilde{J}_{(1)}^i$ .

- It is important to note that the  $v$ -dependence in eq. (3.9) is manifest explicitly. In other words, it is an identity with no explicit factors of  $v$  in  $E_{vv}$ ,  $\Theta$ ,  $Q$  and  $\tilde{J}^i$ . Therefore, it follows that the coefficients of different powers of  $v$  must vanish independently

$$\Theta_{(1)} - Q_{(1)} - \partial_v Q_{(0)} - \nabla_i J_{(1)}^i = 0 \tag{3.10}$$

$$\partial_v \left[ -\Theta_{(1)} - \partial_v \mathcal{M}_{(0)} + Q_{(1)} + \nabla_i \tilde{J}_{(1)}^i \right] = 2E_{vv}|_{r=0} \tag{3.11}$$

- Finally, substituting eq. (3.10) in eq. (3.11) we get the desired result,

$$2E_{vv}|_{r=0} = -\partial_v \left[ \partial_v \left( Q_{(0)} + \mathcal{M}_{(0)} \right) + \nabla_i \left( J_{(1)}^i - \tilde{J}_{(1)}^i \right) \right] + \mathcal{O}(\epsilon^2). \tag{3.12}$$

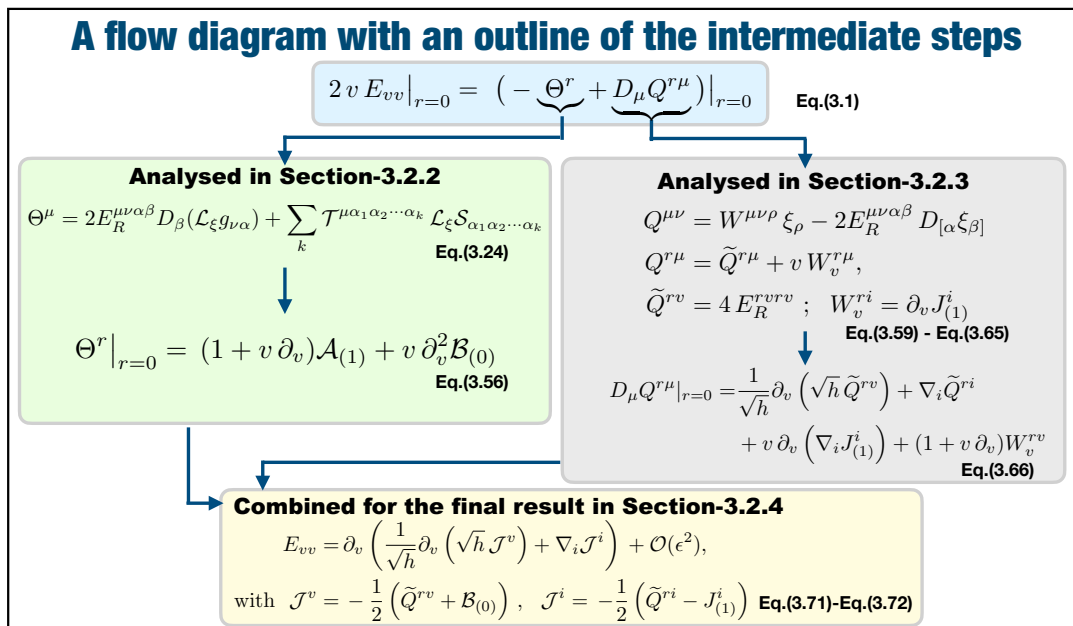
By comparison, we find this is to be of the same form as on the r.h.s. of eq. (3.2). In the following sub-section, we will derive this relation with much more technical rigour and arrive at eq. (3.72) with the expressions of  $\mathcal{J}^v$  and  $\mathcal{J}^i$  in eq. (3.73).

### 3.2 An elaborate analysis of the proof with technical details

In this sub-section, our main goal is to methodically discuss in more detail how eq. (3.1) can lead us to the construction of an entropy current in any higher derivative theory of gravity. Simultaneously one also obtains the components  $\mathcal{J}^v$  and  $\mathcal{J}^i$  as given in eq. (3.2)

---

<sup>32</sup>Here  $\mathcal{L}_\xi$  denotes Lie derivative along the vector  $\xi^\mu$ .



**Figure 1.** A flow diagram summarising the main technical result in each sub-section.

such that the linearized second law is satisfied. For a better organization of the technical computations to be presented in the following, we will separate them into several sub-parts. In section 3.2.1 we will start with obtaining an important result regarding the generic structure of a general covariant tensor with any positive boost weight. Using this result, in section 3.2.2 and section 3.2.3, the two terms on the r.h.s. of eq. (3.1) will be analyzed one at a time. Finally, we will collect them all together in section 3.2.4, to arrive at our final result for the structural form of  $E_{vv}$ . In figure 1, we provide a flow-diagram summarising the main results from each of the following sub-sections.

### 3.2.1 The generic structure of a covariant tensor with positive boost weight

We shall first write down an important result involving the boost weight of a covariant tensor and possible distributions of partial differential operator  $\partial_v$ , evaluated on the horizon. The typical structure of any covariant tensor component with positive boost weight  $w = a + 1 > 0$ , restricted to the horizon and up to linear order in the amplitude of the non-stationary perturbation, will be of the form

$$t_{(a+1)}^{(k)}|_{r=0} = \tilde{T}_{(-k)} \partial_v^{(k+a+1)} T_{(0)}|_{r=0} + \mathcal{O}(\epsilon^2). \quad (3.13)$$

Here the subscripts denote the boost weight of the corresponding quantity and for convenience we have suppressed the index structure for  $t_{(a+1)}^{(k)}$ ,  $\tilde{T}_{(-k)}$  and  $T_{(0)}$ . In writing eq. (3.13) we have followed the guidelines summarised in both table 2 and table 3. More specifically, we made use of the fact that one  $\partial_v$ -derivative operator carries a boost weight of (+1). The quantities  $\tilde{T}_{(-k)}$  and  $T_{(0)}$ , appearing on the r.h.s. of eq. (3.13) are arbitrary tensor quantities with boost weight ‘ $-k$ ’ and zero respectively. The superscript ‘ $k$ ’ on

$t_{(a+1)}^{(k)}$  signifies the fact that any covariant tensor with boost weight  $w = a + 1 > 0$ , can be represented by a member in an one parameter family of terms denoted by  $k = 0, 1, 2, \dots$ .

Once a typical covariant tensor with positive boost weight has been written as in eq. (3.13), one can rearrange the  $\partial_v$ -derivatives and recast it in the following form

$$\left. \begin{aligned}
 t_{(a+1)}^{(k)}|_{r=0} &= \tilde{T}_{(-k)} \partial_v^{(k+a+1)} T_{(0)} + \mathcal{O}(\epsilon^2) \\
 &= \partial_v^{(a+1)} \left[ \sum_{m=0}^{k-1} (-1)^m \left[ {}^{m+a}C_m \right] \tilde{T}_{(-k+m)} \partial_v^{(k-m)} T_{(0)} \right] \\
 &\quad + (-1)^k \left[ {}^{k+a}C_a \right] \tilde{T}_{(0)} \partial_v^{a+1} T_{(0)} + \mathcal{O}(\epsilon^2),
 \end{aligned} \right\} \text{Result: 1} \quad (3.14)$$

where  $\tilde{T}_{(-k+m)} = \partial_v^m \tilde{T}_{(-k)}$  and the numerical coefficient  $\left[ {}^{k+a}C_a \right] = \frac{(k+a)!}{k!a!}$ . The detailed arguments justifying that such a rearrangement is always possible allowing us to write  $t_{(a+1)}^{(k)}$  as the r.h.s. of eq. (3.14), is presented in appendix D.

This is a significant result which we will use quite extensively in the rest of this section. For that reason, we have also identified eq. (3.14) as ‘‘Result: 1’’ and will refer to it accordingly whenever we use it.

Before we proceed, let us mention a notable feature of the final expression in eq. (3.14). In the first term on the r.h.s., the quantity within the square bracket is boost invariant, however, it is product of two terms which are individually not boost invariant: one of them has positive boost weight and the other one has equal boost weight but of negative sign,

$$\text{The first term on the r.h.s. of eq. (3.14)} \sim \partial_v^{(a+1)} \left[ \tilde{T}_{(-k+m)} \partial_v^{(k-m)} T_{(0)} \right] \quad (3.15)$$

where  $\tilde{T}_{(-k+m)}$  and  $\partial_v^{(k-m)} T_{(0)}$  have boost weights equal to  $(-k + m)$  and  $(k - m)$  respectively. Following our set of guidelines presented in table 2, we can therefore conclude that boost invariant terms of this kind are non-zero only due to dynamics and will vanish when evaluated on stationary configurations.

### 3.2.2 Analyzing the structure of the first term (i.e. $\Theta^r$ ) in $E_{vv}$

In this subsection, our focus will be on studying the first term’s structure on the r.h.s. of eq. (3.1), i.e.,  $\Theta^r$ .

For any diffeomorphism invariant theory of gravity, the functional dependence of the Lagrangian is given in eq. (1.2). Under an arbitrary change of the space-time metric denoted by  $g_{\mu\nu} \rightarrow g_{\mu\nu} + \delta g_{\mu\nu}$ , the variation in the Lagrangian, written in eq. (2.32), contains a total derivative term, i.e.  $D_\mu \theta^\mu$ . Furthermore, following the Lemma 3.1 of [8], the general structure of  $\theta^\mu$  can be obtained as the following

$$\Theta^\mu [g_{\beta_1\beta_2}, \delta g_{\beta_1\beta_2}, \delta \phi] = 2E_R^{\mu\nu\alpha\beta} D_\beta \delta g_{\nu\alpha} + \Theta'^\mu, \quad (3.16)$$

where

$$\begin{aligned}
 \Theta'^\mu &= S^{\mu\alpha\beta} (g_{\beta_1\beta_2}, \phi) \delta g_{\alpha\beta} + \sum_{i=0}^{m-1} T_i^{\mu\alpha\beta\rho\sigma\alpha_1\dots\alpha_i} (g_{\beta_1\beta_2}, \phi) \delta D_{(\alpha_1} \dots D_{\alpha_i)} R_{\alpha\beta\rho\sigma} \\
 &\quad + \sum_{i=0}^{l-1} U_i^{\mu\alpha_1\dots\alpha_i} (g_{\beta_1\beta_2}, \phi) \delta D_{(\alpha_1} \dots D_{\alpha_i)} \phi,
 \end{aligned} \quad (3.17)$$

where  $E_R^{\mu\nu\alpha\beta}$  is given by

$$E_R^{\mu\nu\alpha\beta} = \frac{\partial L}{\partial R_{\mu\nu\alpha\beta}} - D_{\rho_1} \frac{\partial L}{\partial D_{\rho_1} R_{\mu\nu\alpha\beta}} + \cdots + (-1)^m D_{(\rho_1} \cdots D_{\rho_m)} \frac{\partial L}{\partial D_{(\rho_1} \cdots D_{\rho_m)} R_{\mu\nu\alpha\beta}}, \quad (3.18)$$

which is a covariant tensor with the same symmetry properties of  $R_{\mu\nu\rho\sigma}$  and  $\phi$  is the matter field that are also present in the Lagrangian.

It should be noted that in the expression of  $\theta^\mu$  on the r.h.s. of eq. (3.16), the variation that is denoted by  $\delta$ , can be put to the left of the covariant derivatives everywhere in  $\Theta^\mu$  except for the term  $E_R^{\mu\nu\alpha\beta} D_\beta \delta g_{\nu\alpha}$ . This distinction between the two terms on the r.h.s. of eq. (3.16) will be crucial in the following.

Let us rewrite eq. (3.16) schematically as the following

$$\Theta^\mu = 2E_R^{\mu\nu\alpha\beta} D_\beta \delta g_{\nu\alpha} + \sum_k \mathcal{T}^{\mu\alpha_1\alpha_2\cdots\alpha_k} \delta \mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_k}, \quad (3.19)$$

where on the r.h.s., the second term collectively includes all the contributions that are in  $\theta^\mu$ , eq. (3.17). Therefore, the quantity  $\mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_k}$  is actually a representative member of the family of covariant tensors as indicated below

$$\mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_k} : \{ g_{\alpha\beta}, R_{\alpha\beta\rho\sigma}, D_{\alpha_1} R_{\alpha\beta\rho\sigma}, \cdots, D_{(\alpha_1} \cdots D_{\alpha_m)} R_{\mu\nu\alpha\beta}, \cdots, \phi, D_{\alpha_1} \phi, \cdots, D_{(\alpha_1} \cdots D_{\alpha_i)} \phi, \cdots \}. \quad (3.20)$$

It should be noted that in our discussion so far, there has been neither any reference to the diffeomorphism transformation nor have we used any invariance of the theory under diffeomorphism. Hence, eq. (3.16) is true for arbitrary variations  $\delta g_{\alpha\beta}$ .

Now, if  $\delta g_{\alpha\beta}$  is generated due to a diffeomorphism given by  $x^\mu \rightarrow x^\mu + \zeta^\mu(x)$  where  $\zeta^\mu$  is some vector field, we can write

$$\delta g_{\alpha\beta} = \mathcal{L}_\zeta g_{\alpha\beta} = D_\alpha \zeta_\beta + D_\beta \zeta_\alpha, \quad (3.21)$$

where  $\mathcal{L}_\zeta$  is the Lie derivative along  $\zeta^\mu$ . If we substitute for  $\delta g_{\alpha\beta}$  from the equation above in eq. (3.16), the corresponding  $\Theta^\mu$  for diffeomorphism will be obtained, which we denote by  $\Theta^\mu[g, \mathcal{L}_\zeta g]$ . Implementing this in eq. (3.19) leads us to the following

$$\Theta^\mu = 2E_R^{\mu\nu\alpha\beta} D_\beta (\mathcal{L}_\zeta g_{\nu\alpha}) + \sum_k \mathcal{T}^{\mu\alpha_1\alpha_2\cdots\alpha_k} \delta \mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_k} [\delta g_{\alpha\beta} \rightarrow \mathcal{L}_\zeta g_{\alpha\beta}]. \quad (3.22)$$

In the second term on the r.h.s. of eq. (3.22), the meaning of the substitution  $\delta g_{\alpha\beta} \rightarrow \mathcal{L}_\zeta g_{\alpha\beta}$  in the argument of  $\delta \mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_k}$  is obvious in the following sense: for each possible terms in  $\delta \mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_k}$  listed in eq. (3.20), one should first reduce it to a form involving  $\delta g_{\alpha\beta}$  and arbitrary number of derivatives on it, and then substitute for  $\delta g_{\alpha\beta} \rightarrow \mathcal{L}_\zeta g_{\alpha\beta}$ .

Next, we adopt a specific choice for the vector field that generates diffeomorphism as  $\xi^\mu$  in place of  $\zeta^\mu$ . In our coordinate system  $\{r, v, x^i\}$  in eq. (2.1),  $\xi^\mu$  is given by  $\xi^\mu \partial_\mu = v \partial_v - r \partial_r$ . From our discussions both in section 2.2 and section 2.3, we immediately identify this  $\xi^\mu$  as the generator of the boost transformation given in eq. (2.4) as well as the Killing vector, eq. (2.16), for the most general stationary black holes in eq. (2.7).

Furthermore, with the input of diffeomorphism invariance of our theory, we can show that

$$\delta \mathcal{S}_{\alpha_1 \alpha_2 \dots \alpha_k} [\delta g_{\alpha\beta} \rightarrow \mathcal{L}_\xi g_{\alpha\beta}] = \mathcal{L}_\xi \mathcal{S}_{\alpha_1 \alpha_2 \dots \alpha_k}, \quad (3.23)$$

and thus obtain the following expression for  $\theta^\mu$

$$\Theta^\mu = 2E_R^{\mu\nu\alpha\beta} D_\beta (\mathcal{L}_\xi g_{\nu\alpha}) + \sum_k \mathcal{T}^{\mu\alpha_1 \alpha_2 \dots \alpha_k} \mathcal{L}_\xi \mathcal{S}_{\alpha_1 \alpha_2 \dots \alpha_k}. \quad (3.24)$$

We refer the readers to appendix E for a detailed clarification regarding eq. (3.23).

In what follows, we will consider each of the two terms on the r.h.s. of eq. (3.24) one at a time and try to obtain the general structure of them. As explained in section 3.1, we would like to know the explicit  $v$ -dependence of each of the two terms that determines  $\Theta^\mu$  in eq. (3.24). In both of these quantities, the source of  $v$ -dependence is solely from the factor of  $\xi^\mu$ . We know that  $\xi$  dependence in  $\Theta$  comes from  $\mathcal{L}_\xi$ , which is linear in  $\xi$  and therefore linear in  $v$ . From here, we learn that, when evaluated on the horizon, the  $v$ -dependence in  $\Theta^\mu$  can at most be linear in  $v$ , and schematically we can write

$$\Theta^\mu|_{r=0} = \Theta_1^\mu + v \Theta_2^\mu, \quad (3.25)$$

where  $\Theta_1^\mu$  and  $\Theta_2^\mu$  are, as of now, components of covariant tensors without any explicit  $v$  dependence. Focussing on  $\Theta^r$ , which is our primary object, and following our rules of boost weight counting, we learn that the corresponding weights of  $\Theta_{(1)}^r$  and  $\Theta_{(2)}^r$  are (+1) and (+2) respectively, as also denoted by their subscripts. With this knowledge of boost weights for  $\Theta_{(1)}^r$  and  $\Theta_{(2)}^r$ , we can use Result:1 from eq. (3.14) to predict the following schematic structures for each of them (ignoring the explicit constant coefficients for now)

$$\begin{aligned} \Theta_{(1)}^r|_{r=0} &\sim \partial_v \mathcal{X} + \tilde{\mathcal{X}}_{(0)} \partial_v \mathcal{X}_{(0)} \\ \Theta_{(2)}^r|_{r=0} &\sim \partial_v^2 \mathcal{Y} + \tilde{\mathcal{Y}}_{(0)} \partial_v^2 \mathcal{Y}_{(0)}, \end{aligned} \quad (3.26)$$

where  $\mathcal{X}$ ,  $\mathcal{X}_{(0)}$ ,  $\tilde{\mathcal{X}}_{(0)}$ ,  $\mathcal{Y}$ ,  $\mathcal{Y}_{(0)}$ ,  $\tilde{\mathcal{Y}}_{(0)}$  are all boost invariant quantities. Recall that according to Result:1 (in eq. (3.14)), although  $\mathcal{Y}$  and  $\mathcal{X}$  are both boost invariant terms, each of them could have pieces that are always product of two terms, one with strictly positive and the other with strictly negative boost weight and therefore vanish on stationary horizons (see the discussions around eq. (3.15)).

We now readily obtain the following form for  $\Theta^r$

$$\Theta^r|_{r=0} \sim \partial_v \mathcal{X} + \tilde{\mathcal{X}}_{(0)} \partial_v \mathcal{X}_{(0)} + v \left[ \partial_v^2 \mathcal{Y} + \tilde{\mathcal{Y}}_{(0)} \partial_v^2 \mathcal{Y}_{(0)} \right]. \quad (3.27)$$

Following our discussion so far, it might appear that  $\Theta_{(1)}^r$  and  $\Theta_{(2)}^r$  are two independent quantities and therefore their structures as written in eq. (3.26) are not related to each other. However, this is not true and due to the fact that both of them are actually generated by an action of the operator  $\mathcal{L}_\xi$  in eq. (3.24), we can constrain the structure of  $\Theta^r$  further. In particular we will explicitly show that one can relate the quantities appearing in eq. (3.26) up to non-linear order corrections in the amplitude of the fluctuations,

$$\tilde{\mathcal{Y}}_{(0)} \partial_v^2 \mathcal{Y}_{(0)} + \mathcal{O}(\epsilon^2) = \tilde{\mathcal{X}}_{(0)} \partial_v^2 \mathcal{X}_{(0)} + \mathcal{O}(\epsilon^2) = \partial_v \left( \tilde{\mathcal{X}}_{(0)} \partial_v \mathcal{X}_{(0)} \right) + \mathcal{O}(\epsilon^2). \quad (3.28)$$

Substituting this relation in equation eq. (3.27) we find the final schematic structure for  $\Theta^r$

$$\begin{aligned} \Theta^r|_{r=0} &= (1 + v \partial_v) \mathcal{A}_{(1)} + v \partial_v^2 \mathcal{B}_{(0)} + \mathcal{O}(\epsilon^2) \\ \text{such that, } \mathcal{A}_{(1)} &= \tilde{\mathcal{X}}_{(0)} \partial_v \mathcal{X}_{(0)} + \partial_v \mathcal{X}, \quad \text{and } \mathcal{B}_{(0)} = \mathcal{Y} - \mathcal{X}. \end{aligned} \quad (3.29)$$

Therefore, in the following, our job is to justify the critical relation eq. (3.28), for both the terms in the expression of  $\Theta^r$  as given in eq. (3.24).

**Analysis of the second term in  $\Theta^r$  from eq. (3.24).** For reasons that will be clear later, to start with, we focus first on the second term on the r.h.s. of (3.24). As we are considering the  $r$ -component of  $\Theta^\mu$ , it follows from our rule of assigning boost weights that  $\Theta^r$  must have a weight of (+1) and consequently both the terms on the r.h.s. of eq. (3.24) will have boost weights equal to (+1). With this knowledge let us now write down the schematic form of the second term in  $\Theta^r$  for our convenience

$$\sum_j \mathcal{T}^{r\alpha_1\alpha_2\cdots\alpha_j} \mathcal{L}_\xi \mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_j} = \sum_{k \geq 0} \tilde{T}_{(-k)} \mathcal{L}_\xi T_{(k+1)} + \sum_{k \geq 0} S_{(k+1)} \mathcal{L}_\xi \tilde{S}_{(-k)}, \quad (3.30)$$

where we have suppressed the details of the index structures of the covariant tensor components  $T$ ,  $\tilde{T}$ ,  $S$  and  $\tilde{S}$ , and the subscripts denote the boost weights for each of them. One can see that the two terms appearing on the r.h.s. of eq. (3.30), are two different cases where the Lie derivative  $\mathcal{L}_\xi$  is acting on quantities with positive and negative boost weight respectively. It should be noted that here we are just doing an operation which is in principle possible without stating anything specific about the structures of the quantities on the r.h.s. above. Thus all the quantities  $T$ ,  $\tilde{T}$ ,  $S$  and  $\tilde{S}$  on the r.h.s. of the equation above are covariant tensors which are otherwise arbitrary apart from having the specific boost weights assigned to them. We can then perform further manipulations as shown below to bring it to a form as given in the following

$$\begin{aligned} &\sum_j \mathcal{T}^{r\alpha_1\alpha_2\cdots\alpha_j} \mathcal{L}_\xi \mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_j} \\ &= \sum_{k \geq 0} \tilde{T}_{(-k)} \mathcal{L}_\xi T_{(k+1)} + \sum_{k \geq 0} S_{(k+1)} \mathcal{L}_\xi \tilde{S}_{(-k)} \\ &= \sum_{k \geq 0} \tilde{T}_{(-k)} \mathcal{L}_\xi T_{(k+1)} + \mathcal{L}_\xi \left( \sum_{k \geq 0} [S_{(k+1)} \tilde{S}_{(-k)}] \right) - \sum_{k \geq 0} \tilde{S}_{(-k)} \mathcal{L}_\xi S_{(k+1)} \\ &= (1 + v \partial_v) \left( \sum_{k \geq 0} [S_{(k+1)} \tilde{S}_{(-k)}] \right) + \sum_{k \geq 0} \tilde{T}_{(-k)} \mathcal{L}_\xi T_{(k+1)} - \sum_{k \geq 0} \tilde{S}_{(-k)} \mathcal{L}_\xi S_{(k+1)} \\ &= (1 + v \partial_v) \mathbb{S}_{(1)} + \sum_{k \geq 0} \tilde{\mathbb{T}}_{(-k)} \mathcal{L}_\xi \mathbb{T}_{(k+1)} \\ &= (1 + v \partial_v) \mathbb{S}_{(1)} + \sum_{k \geq 0} \tilde{\mathbb{T}}_{(-k)} (k + 1 + v \partial_v) \mathbb{T}_{(k+1)}. \end{aligned} \quad (3.31)$$

In the second last line we have defined new quantities for brevity

$$\mathbb{S}_{(1)} \equiv \sum_{k \geq 0} (S_{(k+1)} \tilde{S}_{(-k)}), \quad \text{and } \tilde{\mathbb{T}}_{(-k)} \mathcal{L}_\xi \mathbb{T}_{(k+1)} \equiv \tilde{T}_{(-k)} \mathcal{L}_\xi T_{(k+1)} - \tilde{S}_{(-k)} \mathcal{L}_\xi S_{(k+1)}. \quad (3.32)$$



Comparing with eq. (3.29), it is clear from the final expression on the r.h.s. of eq. (3.31) that the first term there is already of the desired form.

In the following we will process the second term in eq. (3.31). Schematically, this has two different structures in it, namely  $\tilde{\mathbb{T}}_{(-k)} \mathbb{T}_{(k+1)}$  and  $\tilde{\mathbb{T}}_{(-k)} \partial_v \mathbb{T}_{(k+1)}$ . Our strategy would be to use Result:1 from eq. (3.14) repeatedly for each of them. Firstly, we see that, according to Result:1,  $\mathbb{T}_{(k+1)}$  will have the following schematic structure (ignoring the coefficients)

$$\mathbb{T}_{(k+1)} = \partial_v^{k+1} C_{(0)} + A_{(0)} \partial_v^{(k+1)} B_{(0)}. \quad (3.33)$$

Using this we define the quantities  $t_{(1)}^{(k)}$  and  $t_{(2)}^{(k)}$  as given by

$$\begin{aligned} t_{(1)}^{(k)} &\equiv \tilde{\mathbb{T}}_{(-k)} \mathbb{T}_{(k+1)} = \tilde{\mathbb{T}}_{(-k)} \partial_v^{k+1} C_{(0)} + A_{(0)} \tilde{\mathbb{T}}_{(-k)} \partial_v^{(k+1)} B_{(0)}, \\ t_{(2)}^{(k)} &\equiv \tilde{\mathbb{T}}_{(-k)} \partial_v \mathbb{T}_{(k+1)} = \tilde{\mathbb{T}}_{(-k)} \partial_v^{k+2} C_{(0)} + A_{(0)} \tilde{\mathbb{T}}_{(-k)} \partial_v^{(k+2)} B_{(0)}, \end{aligned} \quad (3.34)$$

which then leads us to the following form of the second term in eq. (3.31)

$$\sum_{k \geq 0} \tilde{\mathbb{T}}_{(-k)} \mathcal{L}_\xi \mathbb{T}_{(k+1)} = \sum_{k \geq 0} \tilde{\mathbb{T}}_{(-k)} (k+1 + v \partial_v) \mathbb{T}_{(k+1)} = \sum_{k \geq 0} \left( (k+1) t_{(1)}^{(k)} + v t_{(2)}^{(k)} \right). \quad (3.35)$$

Furthermore, from eq. (3.34) we also learn that, in this case, both the terms that are present in  $t_{(1)}^{(k)}$  has the general structure as

$$t_{(1)}^{(k)} \sim X_{(-k)} \partial_v^{k+1} Y_{(0)}, \quad (3.36)$$

with the following possibilities

$$\text{either } X_{(-k)} \sim \tilde{\mathbb{T}}_{(-k)}, \quad Y_{(0)} \sim C_{(0)}; \quad \text{or } X_{(-k)} \sim A_{(0)} \tilde{\mathbb{T}}_{(-k)}, \quad Y_{(0)} \sim B_{(0)}. \quad (3.37)$$

However, for both the cases mentioned above  $t_{(2)}^{(k)}$  must also have the form given by

$$t_{(2)}^{(k)} \sim X_{(-k)} \partial_v^{k+2} Y_{(0)}. \quad (3.38)$$

We should remember that eq. (3.28) is the key equation that we need to show to recast  $\Theta^r$  into the desired structure given by eq. (3.29). In this case, for the choices of  $t_{(1)}^{(k)}$  and  $t_{(2)}^{(k)}$  given by eq. (3.36) and eq. (3.38) respectively in terms of  $X_{(-k)}$  and  $Y_{(0)}$ , eq. (3.28) results in the following statement:

*If we apply Result:1 to  $t_{(1)}^{(k)}$  in eq. (3.36), it will generate a term of the form  $X_{(0)} \partial_v Y_{(0)}$  and similarly  $t_{(2)}^{(k)}$  in eq. (3.38), will generate a term of the form  $X_{(0)} \partial_v^2 Y_{(0)}$ . The constant coefficient multiplying  $X_{(0)} \partial_v Y_{(0)}$  in  $t_{(1)}^{(k)}$  is  $(k+1)$  times the constant coefficient multiplying  $X_{(0)} \partial_v^2 Y_{(0)}$  in the expansion of  $t_{(2)}^{(k)}$ , see eq. (3.35).*

The italicized statement written above follows trivially, as we show explicitly below, once we apply Result:1 to  $t_{(1)}^{(k)}$  and  $t_{(2)}^{(k)}$  and keep track of the coefficients of these two particular

terms. This also recasts the term  $\tilde{\mathbb{T}}_{(-k)} \mathcal{L}_\xi \mathbb{T}_{(k+1)}$  in the following form,

$$\begin{aligned}
& \tilde{\mathbb{T}}_{(-k)} \mathcal{L}_\xi \mathbb{T}_{(k+1)} \\
&= (k+1) \left( X_{(-k)} \partial_v^{k+1} Y_{(0)} \right) + v \left( X_{(-k)} \partial_v^{k+2} Y_{(0)} \right) \\
&= (k+1) \left[ \partial_v \left( \sum_{m=0}^{k-1} (-1)^m X_{(-k+m)} \partial_v^{k-m} Y_{(0)} \right) + (-1)^k X_{(0)} \partial_v Y_{(0)} \right] \\
&\quad + v \partial_v^2 \left( \sum_{m=0}^{k-1} (-1)^m (m+1) X_{(-k+m)} \partial_v^{k-m} Y_{(0)} \right) \\
&\quad + v (-1)^k (k+1) \left( X_{(0)} \partial_v^2 Y_{(0)} \right) \\
&= (k+1) \left[ \partial_v \left( \sum_{m=0}^{k-1} (-1)^m X_{(-k+m)} \partial_v^{k-m} Y_{(0)} \right) + (-1)^k X_{(0)} \partial_v Y_{(0)} \right] \\
&\quad + v \partial_v^2 \left( \sum_{m=0}^{k-1} (-1)^m [(k+1) - (k-m)] X_{(-k+m)} \partial_v^{k-m} Y_{(0)} \right) \\
&\quad + v (-1)^k (k+1) \left( X_{(0)} \partial_v^2 Y_{(0)} \right) \\
&= (1+v\partial_v) \left[ (k+1) \left( \sum_{m=0}^{k-1} (-1)^m \partial_v \left[ X_{(-k+m)} \partial_v^{k-m} Y_{(0)} \right] + (-1)^k X_{(0)} \partial_v Y_{(0)} \right) \right] \\
&\quad - v \partial_v^2 \left( \sum_{m=0}^{k-1} (-1)^m (k-m) X_{(-k+m)} \partial_v^{k-m} Y_{(0)} \right) + \mathcal{O}(\epsilon^2) .
\end{aligned} \tag{3.39}$$

In the last line we have simply rearranged the terms and used the fact that

$$X_{(0)} \partial_v^2 Y_{(0)} = \partial_v \left( X_{(0)} \partial_v Y_{(0)} \right) + \mathcal{O}(\epsilon^2) .$$

Next, we substitute eq. (3.39) in eq. (3.31) and obtain the following form for the second term in  $\Theta^r$  eq. (3.24),

$$\begin{aligned}
\sum_j \mathcal{T}^{r\alpha_1\alpha_2\cdots\alpha_j} \mathcal{L}_\xi \mathcal{S}_{\alpha_1\alpha_2\cdots\alpha_j} &= (1+v\partial_v) \mathbb{S}_{(1)} + \sum_{k \geq 0} \tilde{\mathbb{T}}_{(-k)} \mathcal{L}_\xi \mathbb{T}_{(k+1)} + \mathcal{O}(\epsilon^2) \\
&= (1+v\partial_v) \mathcal{M}_{(1)} + v \partial_v^2 \mathcal{N}_{(0)} + \mathcal{O}(\epsilon^2) ,
\end{aligned} \tag{3.40}$$

such that

$$\begin{aligned}
\mathcal{M}_{(1)} &= \mathcal{S}_{(1)} + \sum_{k \geq 0} (k+1) \left( \sum_{m=0}^{k-1} (-1)^m \partial_v \left[ X_{(-k+m)} \partial_v^{k-m} Y_{(0)} \right] + (-1)^k X_{(0)} \partial_v Y_{(0)} \right) , \\
\text{and } \mathcal{N}_{(0)} &= \sum_{k \geq 0} \sum_{m=0}^{k-1} (-1)^{m+1} (k-m) X_{(-k+m)} \partial_v^{k-m} Y_{(0)} .
\end{aligned} \tag{3.41}$$

It should be noted that the form in eq. (3.40) is consistent with eq. (3.29).

Finally, we conclude this sub-section with the following comments regarding the structure of  $\mathcal{N}_{(0)}$  that we have obtained in eq. (3.41). It is evident that  $\mathcal{N}_{(0)}$  is a boost invariant

quantity of the type mentioned in eq. (3.15). From its expression, we note that it is a product of two terms that are not individually boost invariant, i.e.

$$\mathcal{N}_{(0)} \sim X_{(-k+m)} \partial_v^{k-m} Y_{(0)}, \quad (3.42)$$

where  $X_{(-k+m)}$  has boost weight of  $(-k+m)$  but  $\partial_v^{k-m} Y_{(0)}$  has a boost weight of  $(k-m)$ . Following our guidelines listed in table 2 (see Rule: 2) we can conclude that  $\mathcal{N}_{(0)}$  is evaluated to be linear in the amplitude of the dynamical fluctuations, i.e.  $\mathcal{N}_{(0)} \sim \mathcal{O}(\epsilon)$ . Furthermore, using Rule: 1 from the same table 2 we also learn that  $\mathcal{N}_{(0)}$  will vanish when the stationarity limit is taken by turning the dynamics off. From these properties it is evident that  $\mathcal{N}_{(0)}$  is a possible candidate to generate the JKM ambiguities associated with the Wald entropy, which will be used in our subsequent analysis.

**Analysis of the first term in  $\Theta^r$  from eq. (3.24).** Next, we consider, for the  $r$  component of  $\Theta^\mu$ , the first term on the r.h.s. of eq. (3.24)

$$2 E_R^{r\nu\rho\sigma} D_\sigma [\mathcal{L}_\xi g_{\nu\rho}] = 2 D_\sigma (E_R^{r\nu\rho\sigma} \mathcal{L}_\xi g_{\nu\rho}) - 2 D_\sigma E_R^{r\nu\rho\sigma} \mathcal{L}_\xi g_{\nu\rho}. \quad (3.43)$$

Now the second term on the r.h.s. of the above equation has the same structure as that of the second term in eq. (3.24). Therefore one can do a similar analysis for that term following the steps mentioned in section 3.2.2.

For the first term in eq. (3.43). In the following, we will only process the first term in eq. (3.43).

Let us define the following quantity

$$H^{\mu\sigma} \equiv E_R^{\mu\nu\rho\sigma} \mathcal{L}_\xi g_{\nu\rho}, \quad (3.44)$$

and using the symmetries of  $E_R^{\mu\nu\rho\sigma}$  under exchanging its indices, we note that  $H^{\mu\sigma}$  is a symmetric tensor,

$$H^{\sigma\mu} = E_R^{\sigma\nu\rho\mu} \mathcal{L}_\xi g_{\nu\rho} = E_R^{\sigma\rho\nu\mu} \mathcal{L}_\xi g_{\rho\nu} = E_R^{\nu\mu\sigma\rho} \mathcal{L}_\xi g_{\nu\rho} = E_R^{\mu\nu\rho\sigma} \mathcal{L}_\xi g_{\nu\rho} = H^{\mu\sigma}. \quad (3.45)$$

Next, we calculate  $D_\sigma H^{r\sigma}$  and evaluate that at the horizon to obtain the following expression<sup>33</sup>

$$\begin{aligned} D_\sigma H^{r\sigma}|_{r=0} &= g^{r\mu} D_\sigma H_\mu^\sigma|_{r=0} = D_\sigma H_v^\sigma|_{r=0} \\ &= \left[ \frac{1}{\sqrt{g}} \partial_\sigma [\sqrt{g} H_v^\sigma] - \frac{1}{2} H^{\mu\rho} \partial_v g_{\mu\rho} \right]_{r=0} \\ &= \left[ \frac{1}{\sqrt{g}} \partial_r [\sqrt{g} H_v^r] + \frac{1}{\sqrt{g}} \partial_v [\sqrt{g} H_v^v] + \frac{1}{\sqrt{g}} \partial_i [\sqrt{g} H_v^i] - \frac{1}{2} H^{ij} \partial_v g_{ij} \right]_{r=0} \\ &= \left[ \frac{1}{\sqrt{g}} \partial_r [\sqrt{g} H^{r\mu} g_{\mu v}] + \frac{1}{\sqrt{g}} \partial_v [\sqrt{g} H^{v\rho} g_{\rho v}] + \frac{1}{\sqrt{g}} \partial_i [\sqrt{g} H^{i\rho} g_{\rho v}] - \frac{1}{2} H^{ij} \partial_v g_{ij} \right]_{r=0} \\ &= \left[ \frac{1}{\sqrt{g}} \partial_r [\sqrt{g} H^{rr}] + \omega_i H^{ri} + \frac{1}{\sqrt{g}} \partial_v [\sqrt{g} H^{vr}] + \frac{1}{\sqrt{g}} \partial_i [\sqrt{g} H^{ir}] - \frac{1}{2} H^{ij} \partial_v g_{ij} \right]_{r=0}. \end{aligned} \quad (3.46)$$

<sup>33</sup>We have used the following formula to compute the divergence of a symmetric tensor

$$D_\sigma H_\mu^\sigma = \frac{1}{\sqrt{g}} \partial_\sigma [\sqrt{g} H_\mu^\sigma] - \frac{1}{2} H^{\sigma\nu} \partial_\mu g_{\sigma\nu}$$

To proceed from here we need to compute different components of  $H^{\mu\sigma}$  for our metric gauge and evaluate that at the horizon. First we define  $\Delta_{\mu\nu} \equiv \mathcal{L}_\xi g_{\mu\nu}$ , and compute its non-zero components as given below

$$\Delta_{vv} = r^2(v\partial_v - r\partial_r)X, \quad \Delta_{vi} = r(v\partial_v - r\partial_r)\omega_i, \quad \Delta_{ij} = (v\partial_v - r\partial_r)h_{ij}. \quad (3.47)$$

This enables us to obtain the following expressions for different components of  $H^{\mu\sigma}$

$$\begin{aligned} H^{rr} &= E_R^{rvvr} \Delta_{vv} + E_R^{rvir} \Delta_{vi} + E_R^{rijr} \Delta_{ij} = r v E_R^{rvir} \partial_v \omega_i + E_R^{rijr} (v\partial_v - r\partial_r)h_{ij} + \mathcal{O}(r^2) \\ H^{rv} &= H^{vr} = E_R^{rvjv} \Delta_{vj} + E_R^{rjkv} \Delta_{jk} = v E_R^{rjkv} \partial_v h_{jk} + \mathcal{O}(r) \\ H^{ri} &= H^{ir} = E_R^{rvvi} \Delta_{vv} + 2E_R^{rvji} \Delta_{vj} + E_R^{rjki} \Delta_{jk} = v E_R^{rjki} \partial_v h_{jk} + \mathcal{O}(r) \\ H^{ij} &= E_R^{ivvj} \Delta_{vv} + 2E_R^{ivkj} \Delta_{vk} + E_R^{ikk'j} \Delta_{kk'} = v E_R^{ikk'j} \partial_v h_{kk'} + \mathcal{O}(r) \end{aligned} \quad (3.48)$$

Remembering the rule of assigning boost weights, we note that  $H^{ri}|_{r=0} = \mathcal{O}(\epsilon^2)$  and  $H^{ij}|_{r=0} = \mathcal{O}(\epsilon)$ , and, consequently we conclude  $H^{ij}\partial_v g_{ij} = H^{ij}\partial_v h_{ij} = \mathcal{O}(\epsilon^2)$ . Implementing all of these results in eq. (3.46) we are left with

$$D_\sigma H^{r\sigma} = \left[ \frac{1}{\sqrt{g}} \partial_r [\sqrt{g} H^{rr}] + \frac{1}{\sqrt{g}} \partial_v [\sqrt{g} H^{vr}] \right]_{r=0} + \mathcal{O}(\epsilon^2), \quad (3.49)$$

with  $H^{rr}$ ,  $H^{vr}$  and  $H^{ir}$  being given in eq. (3.48). Next we evaluate the terms on the r.h.s. of eq. (3.49). The first term can be manipulated using eq. (3.48) as

$$\begin{aligned} \frac{1}{\sqrt{g}} \partial_r [\sqrt{g} H^{rr}] |_{r=0} &= E_R^{rijr} (v\partial_r \partial_v - \partial_r) h_{ij} + \mathcal{O}(\epsilon^2) = E_R^{rijr} (-1 + v\partial_v)(\partial_r h_{ij}) + \mathcal{O}(\epsilon^2) \\ &= \left[ E_R^{rijr} \mathcal{L}_\xi (\partial_r h_{ij}) \right]_{r=0} + \mathcal{O}(\epsilon^2) \end{aligned} \quad (3.50)$$

$$(3.51)$$

It should be noted that  $\partial_r h_{ij}$  cannot be identified as a component of any covariant tensor constructed out of metric. However, it has a definite boost weight ( $= -1$ ) and therefore, the action of  $\mathcal{L}_\xi$ , once it is expressed in terms of boost weight and  $(\xi \cdot \partial)$  as in eq. (2.20), is well-defined. We have used this in arriving at the last equality above in eq. (3.50).

We can clearly see from the r.h.s. of eq. (3.50) that it has the same schematic structure as the second term in equation eq. (3.24), analyzed in previous subsection section 3.2.2. Though  $(\partial_r h_{ij})$  is not a covariant tensor component, the analysis of the previous subsection will go through here since it uses only the boost weight defined through the eq. (2.20). Therefore following the same analysis as in section 3.2.2 we get

$$\begin{aligned} \frac{1}{\sqrt{g}} \partial_r (\sqrt{g} H^{rr}) |_{r=0} &= E_R^{rijr} \mathcal{L}_\xi (\partial_r h_{ij}) |_{r=0} \\ &= (1 + v\partial_v) \mathcal{C}_{(1)} + v\partial_v^2 \mathcal{D}_{(0)} + \mathcal{O}(\epsilon^2), \end{aligned} \quad (3.52)$$

where  $\mathcal{C}_{(1)}$  is a tensor quantity with boost weight  $(+1)$  and  $\mathcal{D}_{(0)}$  is a boost invariant quantity.

For the second term in eq. (3.49), using the corresponding relation from eq. (3.48), we obtain the following form

$$\begin{aligned} \frac{1}{\sqrt{g}} \partial_v (\sqrt{g} H^{rv}) &= \frac{1}{\sqrt{g}} \partial_v \left[ \sqrt{g} v E_R^{rjkv} \partial_v h_{jk} \right] = E_R^{rjkv} \partial_v h_{jk} + v \left( \frac{1}{\sqrt{h}} \partial_v \left[ \sqrt{h} E_R^{rjkv} \partial_v h_{jk} \right] \right) \\ &= (1 + v \partial_v) \mathcal{E}_{(1)} + \mathcal{O}(\epsilon^2), \end{aligned} \quad (3.53)$$

where

$$\mathcal{E}_{(1)} = E_R^{rjkv} \partial_v h_{jk}. \quad (3.54)$$

For the second term in eq. (3.43). As we have mentioned before the second term on the r.h.s. of eq. (3.43) is already in the form of the second term in eq. (3.24) which has been analyzed in detail in the previous section 3.2.2. Therefore following a similar method we can obtain the following form for this term as shown below

$$-2 D_\sigma E_R^{r\nu\rho\sigma} \mathcal{L}_\xi g_{\nu\rho} \Big|_{r=0} = (1 + v \partial_v) \mathcal{F}_{(1)} + v \partial_v^2 \mathcal{G}_{(0)} + \mathcal{O}(\epsilon^2), \quad (3.55)$$

where again,  $\mathcal{F}_{(1)}$  is a tensor quantity with boost weight (+1) and  $\mathcal{G}_{(0)}$  is a boost invariant quantity.

Finally, collecting all the results derived in parts so far from eq. (3.52), eq. (3.53) and eq. (3.55), we obtain the expression for the first term in eq. (3.24) as

$$\begin{aligned} 2 E_R^{r\nu\rho\sigma} D_\sigma [\mathcal{L}_\xi g_{\nu\rho}] \Big|_{r=0} &= (1 + v \partial_v) \mathcal{P}_{(1)} + v \partial_v^2 \mathcal{Q}_{(0)} + \mathcal{O}(\epsilon^2), \\ \text{such that } \mathcal{P}_{(1)} &= \mathcal{C}_{(1)} + \mathcal{E}_{(1)} + \mathcal{F}_{(1)}, \quad \text{and } \mathcal{Q}_{(0)} = \mathcal{D}_{(0)} + \mathcal{G}_{(0)}. \end{aligned} \quad (3.56)$$

which is, again, consistent with the expected form of eq. (3.29).

With all these results at our disposal, we are in a position to collect them together in eq. (3.24) for  $\Theta^r$  as predicted in eq. (3.29), to write down

$$\begin{aligned} \Theta^r \Big|_{r=0} &= (1 + v \partial_v) \mathcal{A}_{(1)} + v \partial_v^2 \mathcal{B}_{(0)} + \mathcal{O}(\epsilon^2), \\ \text{such that } \mathcal{A}_{(1)} &= \mathcal{M}_{(1)} + \mathcal{P}_{(1)}, \quad \text{and } \mathcal{B}_{(0)} = \mathcal{N}_{(0)} + \mathcal{Q}_{(0)}. \end{aligned} \quad (3.57)$$

Before we conclude, let us make the following comment regarding  $\mathcal{B}_{(0)}$ . The term  $\mathcal{Q}_{(0)}$  in eq. (3.56) has the same origin as the term  $\mathcal{N}_{(0)}$  in eq. (3.41). To be more precise, both of them are produced from the first term on the r.h.s. of eq. (3.14), i.e. terms of the kind as given in eq. (3.15). Therefore following the discussion around eq. (3.42), we can argue that  $\mathcal{B}_{(0)}$  is a boost invariant term, although it is product of two terms who are not individually boost invariant. Such terms will vanish in stationary configurations and are non-zero only because of dynamical fluctuations. Using the summary listed in table 2 we can also argue that

$$\mathcal{B}_{(0)} \sim \mathcal{O}(\epsilon) \quad (3.58)$$

As we will see later  $\mathcal{B}_{(0)}$  will, therefore, contribute to the JKM ambiguities of the Wald entropy.

### 3.2.3 Analyzing the structure of the second term (i.e. $D_\mu Q^{r\mu}$ ) in $E_{vv}$

Next we turn our focus to the second term on the r.h.s. of eq. (3.1) in the expression of  $E_{vv}$ . This term involves the quantity  $Q^{\mu\nu}$ , which was defined in eq. (2.39). This  $Q^{\mu\nu}$  is the Noether charge corresponding to diffeomorphism invariance. The general form for this  $Q^{\mu\nu}$  in a diffeomorphism invariant theory of gravity was constructed in [8]. Following the Proposition 4.1 in [8] we can obtain this as given below

$$Q^{\mu\nu} = W^{\mu\nu\rho} \xi_\rho - 2E_R^{\mu\nu\alpha\beta} D_{[\alpha}\xi_{\beta]} + Y^{\mu\nu} + D_\alpha Z^{\mu\nu\alpha}, \quad (3.59)$$

where we have suppressed the functional dependence of the various quantities, e.g.  $W^{\mu\nu\rho}(g_{\beta_1\beta_2}, \phi)$ , etc. and  $E_R^{\mu\nu\rho\sigma}$  is already defined in eq. (3.18). It should be noted that  $Y^{\mu\nu}$  and  $Z^{\mu\nu\alpha}$  are the possible ambiguities in the definition of the Noether charge. They are also present in the definition of  $\theta^\mu$ , but, interestingly, get cancelled in the combination  $(-\Theta^\mu + D_\nu Q^{\nu\mu})$  which is important for our calculations, see eq. (3.1). Therefore, in whatever follows we will neglect them, focussing only on

$$Q^{\mu\nu} = W^{\mu\nu\rho} \xi_\rho - 2E_R^{\mu\nu\alpha\beta} D_{[\alpha}\xi_{\beta]}. \quad (3.60)$$

It should be noted that the two terms present on the r.h.s. above are linear either in  $\xi^\mu$  or in  $D_\alpha \xi_\beta$ . Therefore, following the same argument as in the case of  $\Theta^r$ , we could decompose  $Q^{r\mu}$ , by using  $\xi^\mu$  from eq. (2.16) and explicitly working out the  $v$  dependence as

$$Q^{r\mu} = \tilde{Q}^{r\mu} + v W_v^{r\mu}, \quad (3.61)$$

where  $\tilde{Q}^{r\mu}$  and  $W_v^{r\mu}$  have no explicit  $v$  dependence.<sup>34</sup> In the process of deriving eq. (3.61) from eq. (3.60), we can also obtain the following expression for  $\tilde{Q}^{r\mu}$

$$\tilde{Q}^{rv} = 4 E_R^{rvrv}, \quad (3.62)$$

from which it can be concluded that  $\tilde{Q}^{r\mu}$  actually reproduces the equilibrium Wald entropy density. This will be important later in our analysis.

Following our rules for determining boost weight that we have already declared in table 3 using index counting,  $Q^{rv}$  is boost invariant but  $Q^{ri}$  has boost weight +1. Next, substituting this decomposition for  $Q^{r\mu}$  from eq. (3.61) in the expression of  $D_\mu Q^{r\mu}$  we find

$$\begin{aligned} D_\mu Q^{r\mu} &= D_\mu \tilde{Q}^{r\mu} + W_v^{rv} + v D_\mu W_v^{r\mu} \\ &= \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \tilde{Q}^{rv}) + \nabla_i \tilde{Q}^{ri} + W_v^{rv} + v \left[ \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} W_v^{rv}) + \nabla_i W_v^{ri} \right]. \end{aligned} \quad (3.63)$$

Let us now process the last term on the r.h.s. above. It should be noted that  $W_v^{rv}$  has boost weight (+1) and therefore must be of order  $\mathcal{O}(\epsilon)$ . It follows that

$$\frac{1}{\sqrt{h}} \partial_v (\sqrt{h} W_v^{rv}) = \partial_v W_v^{rv} + \mathcal{O}(\epsilon^2). \quad (3.64)$$

<sup>34</sup>Since  $Q^{\nu\mu}$  is anti-symmetric in its two indices, for  $Q^{r\mu}$  we should note that possible entries for  $\mu$  are  $\mu = v, i, j, k \dots$ , except  $r$ .

Following the same logic, we learn that  $W_v^{ri}$  is a term of boost weight (+2). Therefore, applying Result: 1 from eq. (3.14) it could be written as<sup>35</sup>

$$W_v^{ri} = \partial_v J_{(1)}^i + \mathcal{O}(\epsilon^2) \tag{3.65}$$

such that  $J_{(1)}^i$  is some spatial vector with boost weight (+1).

It should be noted here that the form of  $W_v^{ri}$  written above is a simple representation of its possible structure. For an arbitrary covariant tensor with boost weight equal to +2, we will generally obtain a much more complicated structure. However, here our purpose is to justify the existence of such a term in general. We should also point out that the first term on the r.h.s. of Result: 1 (in eq. (3.14)) will always generate a term like  $\partial_v J_{(1)}^i$  and there is no obstruction to that in principle. As we will see later in this paper, this particular term will be crucial to give us the spatial components of an entropy current. Having said this, we should also mention that one might come up with a particular theory of gravity where such a term is not generated. Einstein's two derivative gravity or  $R^2$  theory of higher derivative gravity are such examples where  $J_{(1)}^i$  turns out to be zero.

From eq. (3.65) we can then derive

$$\nabla_i W_v^{ri} = \partial_v \left( \nabla_i J_{(1)}^i \right) + \mathcal{O}(\epsilon^2). \tag{3.66}$$

Finally, substituting eq. (3.64) and eq. (3.66) in equation eq. (3.63) we arrive at the following expression for  $D_\mu Q^{r\mu}$

$$D_\mu Q^{r\mu} = \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \tilde{Q}^{rv} \right) + \nabla_i \tilde{Q}^{ri} + v \partial_v \left( \nabla_i J_{(1)}^i \right) + (1 + v \partial_v) W_v^{rv} + \mathcal{O}(\epsilon^2). \tag{3.67}$$

For our purpose, we do not need to process this term any further.

### 3.2.4 Final analysis for the structure of $E_{vv}$

In the previous sub-sections, we have obtained the expressions for both  $\Theta^r$  and  $D_\mu Q^{r\mu}$ . For convenience, let us rewrite the final structures of them here again,

$$\begin{aligned} \Theta^r &= (1 + v \partial_v) \mathcal{A}_{(1)} + v \partial_v^2 \mathcal{B}_{(0)} + \mathcal{O}(\epsilon^2), \text{ and,} \\ D_\mu Q^{r\mu} &= \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \tilde{Q}^{rv} \right) + \nabla_i \tilde{Q}^{ri} + v \partial_v \left( \nabla_i J_{(1)}^i \right) + (1 + v \partial_v) W_v^{rv} + \mathcal{O}(\epsilon^2), \end{aligned} \tag{3.68}$$

where  $\mathcal{A}_{(1)}$  and  $\mathcal{B}_{(0)}$  are given in eq. (3.57). Next we collect them together and substitute on the r.h.s. of eq. (3.1) to obtain

$$\begin{aligned} 2v E_{vv}|_{r=0} &= (-\Theta^r + D_\mu Q^{r\mu})|_{r=0} \\ &= -\mathcal{A}_{(1)} + \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \tilde{Q}^{rv} \right) + \nabla_i \tilde{Q}^{ri} + W_v^{rv} \\ &\quad + v \partial_v \left[ -\mathcal{A}_{(1)} + W_v^{rv} + \nabla_i J_{(1)}^i - \partial_v \mathcal{B}_{(0)} \right] + \mathcal{O}(\epsilon^2). \end{aligned} \tag{3.69}$$

---

<sup>35</sup>We are not careful with the coefficients dictated by Result: 1.

Since the dependence on  $v$  is explicit in the equation above, we can compare the coefficients of  $v^0$  and  $v$  on both sides of equation eq. (3.69) to find the following two relations

$$\begin{aligned}\mathcal{A}_{(1)} &= \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \tilde{Q}^{rv} \right) + \nabla_i \tilde{Q}^{ri} + W_v^{rv}, \\ 2E_{vv} &= \partial_v \left[ -\mathcal{A}_{(1)} + W_v^{rv} + \nabla_i J_{(1)}^i - \partial_v \mathcal{B}_{(0)} \right].\end{aligned}\tag{3.70}$$

Substituting  $\mathcal{A}_{(1)}$  from the first relation into the second one of eq. (3.70), we find

$$2E_{vv} = \partial_v \left[ -\left( \frac{1}{\sqrt{h}} \partial_v \left( \sqrt{h} \tilde{Q}^{rv} \right) + \nabla_i \tilde{Q}^{ri} \right) + \nabla_i J_{(1)}^i - \partial_v \mathcal{B}_{(0)} \right] + \mathcal{O}(\epsilon^2).\tag{3.71}$$

As it was explained after eq. (3.57), we know that  $\mathcal{B}_{(0)}$  is of boost weight zero and it is always of order  $\mathcal{O}(\epsilon)$ , see eq. (3.58). Therefore, we learn that

$$\partial_v^2 \mathcal{B}_{(0)} = \frac{1}{\sqrt{h}} \partial_v \left[ \sqrt{h} \partial_v \mathcal{B}_{(0)} \right] + \mathcal{O}(\epsilon^2).$$

Substituting this in the expression of  $E_{vv}$  in eq. (3.71), we find

$$2E_{vv}|_{r=0} = -\partial_v \left( \frac{1}{\sqrt{h}} \partial_v \left[ \sqrt{h} \left( \tilde{Q}^{rv} + \mathcal{B}_{(0)} \right) \right] + \nabla_i \left[ \tilde{Q}^{ri} - J_{(1)}^i \right] \right) + \mathcal{O}(\epsilon^2).\tag{3.72}$$

This constitutes the main result of our paper. We have established that the general structure of  $E_{vv}$  as given above in eq. (3.72) exactly matches with required expression mentioned in eq. (1.7). By comparing them we obtain the following components of the entropy current

$$\mathcal{J}^v = -\frac{1}{2} \left( \tilde{Q}^{rv} + \mathcal{B}_{(0)} \right), \quad \text{and} \quad \mathcal{J}^i = -\frac{1}{2} \left( \tilde{Q}^{ri} - J_{(1)}^i \right).\tag{3.73}$$

In other words, we have also constructed an entropy current on the horizon, with components given by  $\mathcal{J}^v$  and  $\mathcal{J}^i$  for an arbitrary higher derivative theory of gravity.

We also note that  $\tilde{Q}_{rv}$ , as worked out in eq. (3.62), is the Wald entropy density. Additionally, as argued immediately after eq. (3.57),  $\mathcal{B}_{(0)}$  will contribute to the JKM ambiguities that are associated with Wald entropy, and it will vanish on stationary horizons. Therefore, eq. (3.72) also highlights how these ambiguous terms are fixed in order to argue for a linearized second law, which was already pointed out in [14].

## 4 Verification of the proof for specific examples

In section 3 we have constructed a proof of the existence of an entropy current in arbitrary higher derivative theories of gravity. In the following, we would like to verify that the expression of the entropy current suggested by the abstract proof, as given in eq. (3.72), indeed matches with the explicit calculation of the same in some specific model of higher derivative gravity, namely the four derivative theories. The entropy currents for four derivative theories of gravity have already been worked out in [22]. By comparing them with our result reported in this paper, we would, therefore, have a conclusive check of the entropy currents derived via general Noetherian principles and the boost symmetry of the



horizon. For the reader's convenience, we summarize the main equations that we would like to explicitly check. The Key equation that we work with is the following:

$$2v E_{vv}|_{r=0} = [-\Theta^r + D_\mu Q^{r\mu}]|_{r=0}, \quad (4.1)$$

where  $\Theta^r$  has been argued to have the following form, see eq. (3.57),

$$\Theta^r = (1 + v \partial_v) \mathcal{A}_{(0)} + v \partial_v^2 \mathcal{B}_{(0)}. \quad (4.2)$$

The Noether charge has the following form eq. (3.61):

$$Q^{r\mu} = \tilde{Q}^{r\mu} + v W_v^{r\mu} \quad (4.3)$$

Finally we have the expression for the entropy currents from  $E_{vv}$  eq. (3.72)

$$2E_{vv} = -\partial_v \left( \frac{1}{\sqrt{h}} \partial_v [\sqrt{h} (\tilde{Q}^{rv} + \mathcal{B}_{(0)})] + \nabla_i [\tilde{Q}^{ri} - J_{(1)}^i] \right) + \mathcal{O}(\epsilon^2). \quad (4.4)$$

From Proposition 4.1 of [8], we have the following general form for the Noether charge<sup>36</sup>

$$Q^{\mu\nu} = W^{\mu\nu\alpha} \xi_\alpha - 2E_R^{\mu\nu\alpha\beta} D_{[\alpha} \xi_{\beta]} \quad (4.5)$$

If we evaluate the above expression in our gauge eq. (2.1), which we note below for convenience

$$ds^2 = 2 dv dr - r^2 X(r, v, x) dv^2 + 2r \omega_i(r, v, x) dv dx^i + h_{ij}(r, v, x) dx^i dx^j. \quad (4.6)$$

and using the inverse metric components

$$\begin{aligned} g^{rr} &= r^2 X(r, v, x^i) + r^2 \omega^i(r, v, x^i) \omega_i(r, v, x^i) = r^2 X + r^2 \omega^2 \\ g^{ri} &= -r \omega^i & g^{rv} &= 1 & g^{ij} &= h^{ij} & g^{vv} &= g^{vi} = 0 \end{aligned} \quad (4.7)$$

we find the following quantities

$$\begin{aligned} \tilde{Q}^{rv} &= 4E_R^{rvrv}, & \tilde{Q}^{ri} &= 4E_R^{rirv}, \\ W_v^{ri} &= -2E_R^{rirj} \omega_j + W^{rir} = \partial_v J_{(1)}^i. \end{aligned} \quad (4.8)$$

We also define the following quantities related to the extrinsic curvature of the horizon, as the final results have these structures:

$$\begin{aligned} K_{ij} &= \frac{1}{2} \partial_v h_{ij} & K^{ij} &= -\frac{1}{2} \partial_v h^{ij} & \bar{K}_{ij} &= \frac{1}{2} \partial_r h_{ij} & \bar{K}^{ij} &= -\frac{1}{2} \partial_r h^{ij} \\ K &= \frac{1}{2} h^{ij} \partial_v h_{ij} = \frac{1}{\sqrt{h}} \partial_v \sqrt{h} & \bar{K} &= \frac{1}{2} h^{ij} \partial_r h_{ij} = \frac{1}{\sqrt{h}} \partial_r \sqrt{h} \end{aligned} \quad (4.9)$$

The results of the explicit calculations of the Noether charge  $Q^{\mu\nu}$  and  $\Theta^\mu$  for examples considered in [22] are given in tables 4, 5.

Using the above expressions of tables 4, 5, it is straightforward to evaluate  $\Theta^r$  and the various quantities of eq. (4.8) which are given in table 6. From the above results, it is clear that in examples considered,  $\Theta^r$  is really of the form given in eq. (3.57). Using the expressions of table 6, it is straightforward to evaluate the entropy currents  $\mathcal{J}^v, \mathcal{J}^i$  which are summarized in table 7.

<sup>36</sup>We have chosen not to consider the ambiguities of the Noether charge here.

|   |   |  |
|---|---|--|
| 1 | Einstein's theory:<br>$L = R$   | $\Theta^\mu = g^{\mu\alpha} D^\sigma \delta g_{\alpha\sigma} - g^{\rho\sigma} D^\mu \delta g_{\rho\sigma}$   |
| 2 | Ricci Scalar Squared Theory:<br>$L = R^2$   | $\Theta^\mu = 2 [R(g^{\mu\alpha} D^\nu \delta g_{\nu\alpha} - g^{\nu\alpha} D^\mu \delta g_{\nu\alpha}) - g^{\mu\alpha} (D^\beta R) \delta g_{\alpha\beta} + (D^\mu R) g^{\alpha\beta} \delta g_{\alpha\beta}]$  |
| 3 | Ricci Tensor Squared Theory:<br>$L = R_{\mu\nu} R^{\mu\nu}$                       | $\Theta^\mu = 2g^{\mu\alpha} R^{\sigma\nu} D_\sigma \delta g_{\nu\alpha} - 2(D^\alpha R^{\mu\nu}) \delta g_{\nu\alpha} - R^{\alpha\beta} D^\mu \delta g_{\alpha\beta} + (D^\mu R^{\alpha\beta}) \delta g_{\alpha\beta} - g^{\rho\sigma} R^{\mu\nu} D_\nu \delta g_{\rho\sigma} + \frac{1}{2} (D^\mu R) g^{\alpha\beta} \delta g_{\alpha\beta}$ |
| 4 | Riemann Tensor Squared Theory:<br>$L = R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$ | $\Theta^\mu = 4R^{\mu\rho\nu\sigma} D_\sigma \delta g_{\nu\rho} - 4(D_\sigma R^{\mu\nu\rho\sigma}) \delta g_{\nu\rho}$   |

**Table 4.** Symplectic Potential for Einstein gravity and other four derivative theories of gravity.

|   |   |   |
|---|---|---|
| 1 | Einstein's theory:<br>$L = R$   | $Q^{\mu\nu} = (D^\nu \xi^\mu - D^\mu \xi^\nu)$  |
| 2 | Ricci Scalar Squared Theory:<br>$L = R^2$   | $Q^{\mu\nu} = [4(D^\mu R) g^{\nu\alpha} - 4(D^\nu R) g^{\mu\alpha}] \xi_\alpha + 2R(D^\nu \xi^\mu - D^\mu \xi^\nu)$   |
| 3 | Ricci Tensor Squared Theory:<br>$L = R_{\mu\nu} R^{\mu\nu}$                       | $Q^{\mu\nu} = (2D^\mu R^{\nu\alpha} - 2D^\nu R^{\mu\alpha} + (D^\mu R) g^{\nu\alpha} - (D^\nu R) g^{\mu\alpha}) \xi_\alpha + 2R^{\nu\beta} D_\beta \xi^\mu - 2R^{\mu\beta} D_\beta \xi^\nu$ |
| 4 | Riemann Tensor Squared Theory:<br>$L = R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$ | $Q^{\mu\nu} = 4R^{\mu\nu\alpha\sigma} D_\sigma \xi_\alpha - 8(D_\sigma R^{\mu\nu\alpha\sigma}) \xi_\alpha$  |

**Table 5.** Noether Charges for Einstein gravity and other four derivative theories of gravity.

Finally, we provide the expressions for Gauss-Bonnet Theory:  $L = R^2 - 4R_{\mu\nu} R^{\mu\nu} + R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$ .  $\Theta^\mu$  is given by

$$\begin{aligned} \Theta^\mu &= 2R(g^{\mu\alpha} D^\sigma \delta g_{\alpha\sigma} - g^{\rho\sigma} D^\mu \delta g_{\rho\sigma}) - 2g^{\mu\nu} (D^\alpha R) \delta g_{\alpha\nu} \\ &\quad - 8g^{\mu\alpha} R^{\sigma\nu} D_\sigma \delta g_{\nu\alpha} + 4(D^\alpha R^{\mu\nu}) \delta g_{\nu\alpha} + 4R^{\alpha\beta} D^\mu \delta g_{\alpha\beta} \\ &\quad + 4g^{\rho\sigma} R^{\mu\nu} D_\nu \delta g_{\rho\sigma} + 4R^{\mu\nu\rho\sigma} D_\sigma \delta g_{\nu\rho} \end{aligned} \quad (4.10)$$

The Noether charge is given by

$$Q^{\mu\nu} = 2R(D^\nu \xi^\mu - D^\mu \xi^\nu) + 8(R^{\mu\beta} D_\beta \xi^\nu - R^{\nu\beta} D_\beta \xi^\mu) + 4R^{\mu\nu\alpha\sigma} D_\sigma \xi_\alpha \quad (4.11)$$

Thus,  $\Theta^r$  and the quantities in eq. (4.8) are straightforward to compute:

$$\begin{aligned} \Theta^r &= (1 + v\partial_v)[-4RK + 8R^{ij} K_{ij} + 8R_{rv} K - 8K\partial_v \bar{K} \\ &\quad + 8R_{vijr} K^{ij} + 8K_{ij} \partial_v \bar{K}^{ij} + 8\partial_v (K\bar{K} - 8K_{ij} \bar{K}^{ij})] \\ &\quad + v\partial_v^2 [-8K\bar{K} + 8\bar{K}^{ij} K_{ij}] \end{aligned} \quad (4.12)$$

$$\begin{aligned} \tilde{Q}^{rv} &= -4R + 16R_{rv} + 8R_{rvrv} \\ \tilde{Q}^{ri} &= 8\nabla_j K^{ij} - 8\nabla^i K - 4\omega^i K + 4\omega_j K^{ij} \\ J_{(1)}^i &= -4\omega^i K + 4\omega_j K^{ij} \end{aligned} \quad (4.13)$$

|   |  |   |
|---|--|---|
| 1 | Einstein's theory:<br>$L = R$  | $\Theta^r = (1 + v\partial_v)(-2K)$<br>$\tilde{Q}^{rv} = -2 \quad \tilde{Q}^{ri} = 0 \quad J_{(1)}^i = 0$   |
| 2 | Ricci Scalar Squared Theory:<br>$L = R^2$  | $\Theta^r = (1 + v\partial_v)(-4RK)$<br>$\tilde{Q}^{rv} = -4R \quad \tilde{Q}^{ri} = 0 \quad J_{(1)}^i = 0$   |
| 3 | Ricci Tensor Squared Theory:<br>$L = R_{\mu\nu}R^{\mu\nu}$                       | $\Theta^r = (1 + v\partial_v)[-2R^{ij}K_{ij} - 2R_{vr}K + 2K\partial_v\bar{K} - 2\partial_v(K\bar{K})] + v\partial_v^2[2K\bar{K}]$<br>$\tilde{Q}^{rv} = -4R_{rv}$<br>$\tilde{Q}^{ri} = h^{ij}\partial_v\omega_j - 2\nabla_n K^{in} + 2\nabla^i K + \omega^i K$<br>$J_{(1)}^i = \omega^i K - h^{ij}\partial_v\omega_j + 2\nabla_j K^{ij}$    |
| 4 | Riemann Tensor Squared Theory:<br>$L = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ | $\Theta^r = (1 + v\partial_v)[8R_{vijr}K^{ij} + 8K_{ij}\partial_v\bar{K}^{ij} - 8\partial_v(\bar{K}^{ij}K_{ij})] + v\partial_v^2(8\bar{K}^{ij}K_{ij})$<br>$\tilde{Q}^{rv} = 8R_{rvrv}$<br>$\tilde{Q}^{ri} = 4h^{ij}\partial_v\omega_j + 4\omega_k K^{ik}$<br>$J_{(1)}^i = -4h^{ij}\partial_v\omega_j + 4\omega_j K^{ij} + 8\nabla_j K^{ij}$ |

**Table 6.** Summary of the various quantities needed to compute the Entropy currents.

|   |  |   |
|---|--|---|
| 1 | Einstein's theory:<br>$L = R$  | $\mathcal{J}^v = 1 \quad \mathcal{J}^i = 0$   |
| 2 | Ricci Scalar Squared Theory:<br>$L = R^2$  | $\mathcal{J}^v = 2R \quad \mathcal{J}^i = 0$  |
| 3 | Ricci Tensor Squared Theory:<br>$L = R_{\mu\nu}R^{\mu\nu}$                       | $\mathcal{J}^v = 2R_{rv} - K\bar{K}$<br>$\mathcal{J}^i = -h^{ij}\partial_v\omega_j + 2\nabla_j K^{ij} - \nabla^i K$   |
| 4 | Riemann Tensor Squared Theory:<br>$L = R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ | $\mathcal{J}^v = -4R_{rvrv} - 4\bar{K}^{ij}K_{ij}$<br>$\mathcal{J}^i = -4h^{ij}\partial_v\omega_j + 4\nabla_j K^{ij}$ |

**Table 7.** Entropy currents for Einstein gravity and other four derivative theories of gravity.

The Entropy currents are therefore given by

$$\begin{aligned}
 \mathcal{J}^v &= 2R - 8R_{rv} - 4R_{rvrv} + 4K\bar{K} - 4\bar{K}^{ij}K_{ij} = 2\mathcal{R} \\
 \mathcal{J}^i &= -4\nabla_j K^{ij} + 4\nabla^i K
 \end{aligned}
 \tag{4.14}$$

where  $\mathcal{R}$  is the intrinsic Ricci scalar of the horizon.

These results exactly match with the explicit results of [22]. To see this, we must compare the expressions written above for  $\mathcal{J}^v$  and  $\mathcal{J}^i$  as listed in table 7 with section 3.1 in [22].<sup>37</sup> Thus, the results in this section are validation of the proof provided above.

<sup>37</sup>It must be noted that there is an overall negative sign in the equations of motion used in the calculations above ((i.e.)  $E_{vv}$ ) when compared to the expressions used in [22]. Due to an additional overall negative sign in definitions of the entropy currents, they exactly match with the results provided here.

## 5 A proof of the physical process version of the first law

So far in this paper, our focus has been to construct an entropy current with the motivation to produce a proof of the linearized version of the second law of thermodynamics for dynamical black holes. However, a similar framework could be used to construct an alternative proof of a particular version of the first law, namely, the physical process version of it, [11, 23–28]. Interestingly, one should note that in [14] the validity of this version of the first law was actually in a subtle way needed to prove the linearized version of the second law (see section (2.2.3) in [22] for a detailed discussion on this point).

The purpose of this section is to reorganize the setup that we have already developed in this paper in order to devise an alternative proof of the physical process version of the first law applicable to the most general higher derivative theories of gravity. The main ingredient that we will use to achieve this, is essentially the fact that  $\chi^\mu \xi^\nu E_{\mu\nu}$ ,<sup>38</sup> could always be expressed in the form of eq. (1.7).

Here we are restricting ourselves to a situation where the dynamical fluctuation around a stationary horizon is of small amplitude. It is initiated by a small in-falling matter perturbing a black hole configuration at equilibrium. By small matter, what we mean is more precisely the introduction of a perturbation in the system of equations by a small matter stress tensor denoted by  $\delta T_{\mu\nu}$ . The perturbed horizon eventually settles down to a new stationary black hole configuration with slightly shifted asymptotic parameters (identified with the total energy, angular momentum, etc.) and slightly altered geometry around the horizon (identified with the total entropy). Since the same stress tensor component generates all such infinitesimal shifts, they are related, and this relation is nothing but the statement of the first law.

Let us denote the change in entropy due to the evolution from the initial to the final equilibrium configuration as  $\delta S$ . The relevant component of the stress tensor which gets related to  $\delta S$  can be identified with  $\chi^\mu \xi^\nu \delta T_{\mu\nu}$  and that helps us writing down the first law more concretely as follows<sup>39</sup>

$$\delta S = 2\pi \int_{\mathcal{H}} dv d^{d-2}x \sqrt{h} \chi^\mu \xi^\nu \delta T_{\mu\nu} = -2\pi \int_{-\infty}^{+\infty} dv \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} v E_{vv} \quad (5.1)$$

where  $\mathcal{H}$  is the co-dimension one horizon at  $r = 0$ ,  $\mathcal{H}_v$  is the constant  $v$ -slice of the horizon. Also for the last equality above we have used the fact that the equation of motion is given by  $E_{\mu\nu} = -\delta T_{\mu\nu}$  (see eq. (2.26)), along with  $\chi^v = 1$  and  $\xi^v = v$  are the only non-zero components of  $\chi^\mu, \xi^\nu$  in our set up.

Next, let us recollect that the main result of section 3 was to obtain  $E_{vv}$  in the form given in eq. (3.72), which we write here again for convenience,

$$E_{vv} = \partial_v \left( \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right) + \mathcal{O}(\epsilon^2), \quad (5.2)$$

with  $\mathcal{J}^v = -\frac{1}{2} (\tilde{Q}^{rv} + \mathcal{B}_{(0)})$ ,  $\mathcal{J}^i = -\frac{1}{2} (\tilde{Q}^{ri} - J_{(1)}^i)$ .

<sup>38</sup>Recall  $\chi = \chi^a \partial_a = \partial_v$  is the affinely parametrized null generator of the horizon and  $\xi = \xi^\mu \partial_\mu = v \partial_v - r \partial_r$  is the Killing vector field.

<sup>39</sup>A detailed derivation of this can be found in section 2 of [11]. Also see section (2.2.3) of [22] for more discussions regarding the same.

In what follows, we will argue that the form of  $E_{vv}$  as given in eq. (5.2) is indeed consistent with eq. (5.1). Therefore, to show that eq. (5.2) is true in arbitrary higher derivative theories of gravity is equivalent to verify the physical process version of the first law in such theories.

We first substitute for  $E_{vv}$  from eq. (5.2) in eq. (5.1) followed by certain manipulations involving integrations by parts to move the overall  $\partial_v$  on the r.h.s. of eq. (5.2) in  $E_{vv}$  to obtain

$$\begin{aligned} \delta S &= - \int_{-\infty}^{\infty} dv \int_{\mathcal{H}_v} d^{d-2}x v \sqrt{h} \partial_v \left( \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} \mathcal{J}^v) + \nabla_i \mathcal{J}^i \right) \\ &= - \left[ \int_{\mathcal{H}_v} d^{d-2}x v \left( \partial_v (\sqrt{h} \mathcal{J}^v) + \partial_i (\sqrt{h} \mathcal{J}^i) \right) \right]_{v=-\infty}^{v=+\infty} \\ &\quad + \int_{-\infty}^{\infty} dv \int_{\mathcal{H}_v} d^{d-2}x \frac{\partial_v (v\sqrt{h})}{\sqrt{h}} \left( \partial_v (\sqrt{h} \mathcal{J}^v) + \partial_i (\sqrt{h} \mathcal{J}^i) \right). \end{aligned} \tag{5.3}$$

Note that the combination “ $\partial_v (\sqrt{h} \mathcal{J}^v) + \partial_i (\sqrt{h} \mathcal{J}^i)$ ” is of boost weight (+1) and therefore must vanish when evaluated on any stationary configuration. Due to this and the fact that at  $v = \pm\infty$  we have two stationary configurations at the beginning and the end of the dynamical process, the first term on the r.h.s. of eq. (5.3) (appearing on the second line) simply vanishes.

Now, for the second term on the r.h.s. of eq. (5.3) (appearing on the third line), we note that it is a product of two factors, one of which is the same quantity with boost weight (+1) that we mentioned above and therefore this is already linear in the amplitude of the dynamics at  $\mathcal{O}(\epsilon)$ . So, from the second factor, we have to consider only that piece which is boost-invariant, neglecting the other  $\mathcal{O}(\epsilon)$  piece, i.e.

$$\partial_v (v\sqrt{h}) = \sqrt{h} + \underbrace{v \partial_v \sqrt{h}}_{\sim \mathcal{O}(\epsilon)},$$

ensuring that the r.h.s. of eq. (5.3) is calculated up to  $\mathcal{O}(\epsilon)$ . Therefore we obtain

$$\delta S \approx \int_{-\infty}^{\infty} dv \int_{\mathcal{H}_v} d^{d-2}x \left( \partial_v (\sqrt{h} \mathcal{J}^v) + \partial_i (\sqrt{h} \mathcal{J}^i) \right) + \mathcal{O}(\epsilon^2). \tag{5.4}$$

In eq. (5.4) the second term on the r.h.s. being a total spatial derivative, integrates to boundary terms on  $\mathcal{H}_v$ , the constant  $v$  slices of the horizon. Therefore, it vanishes if we assume that the horizon is compact. Neglecting this term, we finally arrive at

$$\delta S \approx \int_{-\infty}^{\infty} dv \partial_v \left[ \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} \mathcal{J}^v \right] + \mathcal{O}(\epsilon^2) = \left[ \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} \mathcal{J}^v \right]_{v=-\infty}^{v=+\infty} + \mathcal{O}(\epsilon^2) \tag{5.5}$$

From eq. (5.5) it is clear that the change in entropy during this process is given by the change in the quantity  $\mathcal{J}^v$  evaluated at the two stationary configurations at  $v = -\infty$  and  $v = +\infty$  respectively. We also obtain an expression that has to be identified with the total entropy on every constant  $v$  slices,

$$S \approx \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} \mathcal{J}^v = -\frac{1}{2} \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} \left( \tilde{Q}^{rv} + \mathcal{B}_{(0)} \right), \tag{5.6}$$

where we have substituted for the expression of  $\mathcal{J}^v$  from eq. (5.2).

From eq. (5.6) we immediately recognize  $\tilde{Q}^{rv}$  as the Wald entropy density for equilibrium black holes, which we have already argued in the previous sections. The other term on the r.h.s. of eq. (5.6), i.e.,  $\mathcal{B}_{(0)}$ , has also previously been identified with the JKM ambiguities present in the out-of-equilibrium extensions of Wald entropy; see the discussion after eq. (3.57). Since the JKM ambiguities vanish on stationary configurations, the term  $\mathcal{B}_{(0)}$  will not contribute to the r.h.s. of eq. (5.5), justifying that the physical process version of the first law is not affected by these terms. However, as we have already seen, they still play a crucial role in obtaining the linearized version of the second law.

It is worth highlighting that the crucial input for this derivation was the structure of  $E_{vv}$  as predicted in eq. (1.7). In other words, our proof of the second law within the approximation of linearized dynamics also allows us to construct the proof of the first law using the physical process version in the most general higher derivative theory of gravity.

Finally, we conclude this section with the following clarification regarding the implications of our derivation of this version of the first law. As it turns out, for this physical process version of the first law to be true in its most general form for an arbitrary higher derivative theory of gravity, the  $vv$ -component of the equations of motion, i.e.  $E_{vv}$ , has to satisfy the following restriction imposed on its off-shell structure, (see eq. (25) in [18]),

$$\int_{-\infty}^{\infty} dv \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} v \left( \partial_v^2 s_w - s_w R_{vv} + E_{vv}^{\text{HD}} \right) = \mathcal{O}(\epsilon^2), \quad (5.7)$$

where  $s_w = s_w^{\text{HD}} + s_{\text{cor}}$ , such that  $s_w^{\text{HD}}$  is the contribution to the Wald entropy density from higher derivative part of the Lagrangian defined in eq. (2.22), and  $s_{\text{cor}}$  is the possible out-of-equilibrium extension or correction to  $s_w^{\text{HD}}$  defined in eq. (2.23). Also  $E_{vv}^{\text{HD}}$  is the component of equations of motion getting contributions only from the higher derivative part of the Lagrangian; see eq. (2.26).<sup>40</sup> It has been shown that eq. (5.7) is true for  $f(R)$  theories [11], and for Gauss-Bonnet or Lovelock theories [12, 27, 28]. However, it is not yet known if this condition is true for arbitrary higher derivative theories of gravity.<sup>41</sup>

In this section, we have shown that the explicit form of  $E_{vv}$  given in eq. (5.2) indeed satisfies this condition up to linearised order in the amplitude expansion. To see this, we should compute the  $vv$ -component of the Ricci Tensor, i.e.,  $R_{vv}$  in our metric gauge, eq. (2.1), which yields

$$R_{vv} = -\partial_v \left( \frac{1}{\sqrt{h}} \partial_v (\sqrt{h}) \right) + \mathcal{O}(\epsilon^2). \quad (5.8)$$

Using eq. (5.8) one can show that the terms inside the parenthesis in the integrand on the l.h.s. of eq. (5.7), can be expressed as the following

$$\partial_v^2 s_w - s_w R_{vv} + E_{vv}^{\text{HD}} = \partial_v \left( \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} s_w) \right) + E_{vv}^{\text{HD}} + \mathcal{O}(\epsilon^2). \quad (5.9)$$

---

<sup>40</sup>The condition written in eq. (5.7) refers to the physical process version of the first law in a manner that is only determined by the geometry near the dynamical horizon and, most importantly, independent of the information at the asymptotic boundary of space-time.

<sup>41</sup>For the arguments to derive eq. (5.7) and a detailed discussion on this we refer the reader to section 2 of [26]. This is also discussed in section 3 of the review [18].

With eq. (5.9) at hand, the restriction on  $E_{vv}$  as written above in eq. (5.7) takes the following form

$$\int_{-\infty}^{\infty} dv \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} v \left( \partial_v \left( \frac{1}{\sqrt{h}} \partial_v (\sqrt{h} s_w) \right) + E_{vv}^{\text{HD}} \right) = \mathcal{O}(\epsilon^2). \quad (5.10)$$

It is now straightforward to check that eq. (5.10) follows trivially from eq. (2.29), which we have already encountered in section 2.4. Hence, it is also obvious that the form of  $E_{vv}$  given in eq. (5.2) with the specific identifications of  $\mathcal{J}^v$  and  $\mathcal{J}^i$ , satisfies the condition in eq. (5.7).

One crucial outcome of our analysis in this paper is to justify that certain total spatial derivatives in the form of  $\nabla_i \mathcal{J}^i$  are generically present in the structure of  $E_{vv}$  as in eq. (5.2). However, we now see that the presence of such total spatial divergences does not affect the arguments leading to this version of the first law. As we know, in the statement of the first law only the total entropies between two equilibrium configurations are compared and the expressions of this total entropy involve a spatial integration on the co-dimension two hyper-surface  $\mathcal{H}_v$ . Consequently, for compact horizons, it is easy to check that the spatial total derivatives integrate to zero,

$$\int_{-\infty}^{\infty} dv \int_{\mathcal{H}_v} d^{d-2}x \sqrt{h} v \partial_v \left( \nabla_i \mathcal{J}^i \right) \sim \mathcal{O}(\epsilon^2). \quad (5.11)$$

## 6 Discussions and future directions

In this paper, we have constructed a proof of the second law of black hole thermodynamics in arbitrary higher derivative theories of gravity working with a particular choice of coordinates. We have focussed on situations where the dynamical fluctuations connecting the two stationary black holes are of small amplitude such that a linearized amplitude expansion around the initial stationary background is possible. The dynamics are initiated by, e.g., throwing a small matter into the equilibrium black hole, and we have assumed a null energy condition for that matter stress tensor. In the course of designing the proof, we have also constructed an entropy current on the horizon. The divergence of this entropy current being non-negative signifies entropy production locally at each point of the space-time. The ‘time’ component of this entropy current produces the correct expression for the Wald entropy density and also fixes the associated JKM ambiguities as a result of imposing the local second law. The spatial components of the entropy current provide us a mechanism of entropy equilibration via the spatial flow of entropy on the constant ‘time’-slice of the horizon. This, in turn, provides us with a possible out-of-equilibrium extension of the equilibrium Wald entropy function, maintaining the second law in its strongest possible form. As a by-product of our analysis, we have also completed an alternative proof of the first law using physical processes like in-falling matter initiating black hole dynamics in the most general higher derivative theories of gravity.

Let us now comment on some possible future directions of our work:

**Extending beyond linear order in amplitude expansion.** From our analysis in this paper, it might appear that our results up to linear order in amplitude expansion are enough to prove the second law completely in higher derivative gravitational theory up to all orders. In other words, there is apparently no need to worry about the non-linear terms as long as we are treating the higher derivative terms as perturbative corrections to the leading two derivative Einstein gravity. The reason for this is as follows. When considering the terms that are non-linear in the amplitude of the dynamics, we could see entropy production at two derivative order itself. Furthermore, the higher derivative terms, being considered as a small perturbative correction, should always be suppressed compared to the two derivative Einstein-Hilbert terms in the gravity Lagrangian. Hence, they could never change the sign of the rate of entropy production. Therefore, one might conclude that the higher derivative corrections in the gravity Lagrangian have to be analyzed carefully only when the two derivative contributions to the entropy production vanish, namely at the linear order in the amplitude of the dynamics.

However, as it was shown in [40], there are situations where such an argument does not work, at least locally, and one needs to add further corrections to the entropy density (and also presumably to the current) that are manifestly non-linear in the amplitude. In view of this observation, the immediate question is whether we could extend our proof to non-linear orders in amplitude.

**Interpretation of the entropy current via fluid-gravity duality.** Another interesting future direction in our opinion is to explore the following question: what does this gravitational entropy density along with the spatial current translate to in the non-gravitational dual systems, whenever they exist, like fluid dynamics [29] and membrane dynamics [36, 37]. We already have algorithms to construct an entropy current on these dual systems that will maintain the second law locally even at nonlinear order in the amplitude of the dynamics [31, 32, 36, 37]. If we can cleanly work out the map between the coordinates of this current paper and the different gauges or coordinates used to construct these dual fluid or membrane systems, it would presumably tell us a lot about the non-linear completion of the gravitational entropy current.

**Exploring a “covariant version” of the entropy current.** In the dual systems that we have mentioned above, in any given coordinate system, the space and time components of the entropy current are always related by the constraints imposed by the freedom of coordinate transformation. However, our proof relies heavily on the horizon adapted coordinate system that we have chosen, and our effective ‘time’ is basically an affine parameter along a null direction. So naive mixing of time and space is not possible here. In other words, we have been able to justify the existence of an entropy current following the construction of [8], but only working in a particular choice of the space-time metric as in eq. (2.1). Nevertheless, we believe that the notion of non-negative divergence of an entropy current, if it exists, should be a robust feature in a covariant sense as well. It would be nice if we could obtain a proof supporting this in a more coordinate invariant fashion.



Additionally, the components of the entropy current that we have obtained in our analysis also crucially depend on the particular slicing of the null horizon. It would be interesting to understand how they transform under an arbitrary redefinition of the constant  $v$ -slices of the horizon, e.g., eq. (2.2). Such an attempt might lead to a covariant formula also for the spatial entropy current. One possible direction going ahead would be to explore the problem with a viewpoint of covariant phase space analysis along the lines of [41].

We hope to come back to these issues in our future works.

## Acknowledgments

We want to thank Jyotirmoy Bhattacharya for initial collaboration, useful discussions, and important inputs. We are grateful to Aron Wall for his efforts in carefully reading our manuscript and providing us with his valuable suggestions and several significant comments. We would also like to thank Shamik Banerjee, Parthajit Biswas, Arunabha Saha for helpful discussions. M.P. would like to thank Shankha Singha Mahapatra for valuable discussions. P.D. would like to duly acknowledge the Council of Scientific and Industrial Research (CSIR), New Delhi for financial assistance through the Junior Research Fellowship (JRF) scheme.

## A Notations, conventions and useful definitions

- We work in  $d$ -dimensional space-time with coordinates denoted by  $x^\mu$ .
- Lower-case Greek indices:  $\{\alpha, \beta, \dots, \mu, \nu, \dots, \alpha_1, \alpha_2, \dots\}$ , are used for space-time coordinates.
- The space-time metric is denoted by  $g_{\mu\nu}$  and the associated covariant derivative is denoted by the operator  $D_\mu$ .
- The horizon, a co-dimension one hyper-surface  $\mathcal{H}$ , is located at  $r = 0$ . Here  $r$  is the coordinate running along the affinely parametrized geodesics  $\partial_r$  that takes us away from the horizon.
- The affinely parametrized null generator on the horizon is given by  $\partial_v$  and  $v$  is the coordinate along those null generators.
- Constant  $v$ -slices of the horizon is a co-dimension two hyper-surface  $\mathcal{H}_v$  spanned by  $(d - 2)$  spatial coordinates  $x^i$ .
- Lower-case Latin indices:  $\{i, j, \dots, i_1, j_1, \dots\}$  are used for coordinates on  $\mathcal{H}_v$ .
- The full space-time coordinates are  $x^\mu = \{r, v, x^i\}$  and coordinates on the horizon  $\mathcal{H}$  are  $\{v, x^i\}$ .
- The intrinsic metric on the horizon is  $h_{ij}$  and the associated covariant derivative is  $\nabla_i$ .

## B More details regarding the boost weight of a covariant tensor

In this appendix, we will provide additional details regarding some of the results which were used in the paper.

### B.1 Equivalence between different ways of evaluating the Boost weights

We will first try to show the equivalence between the definitions of Boost Weight mentioned in section 2.3. We will collect them here for reference:

$$\mathcal{A} \rightarrow \tilde{\mathcal{A}} = \lambda^w \mathcal{A}, \text{ under } \{r \rightarrow \tilde{r} = \lambda r, v \rightarrow \tilde{v} = \lambda^{-1} v\} \Rightarrow \text{boost weight of } \mathcal{A} \text{ is } w. \quad (\text{B.1})$$

$$\mathcal{L}_\xi S_{\alpha_1 \alpha_2 \dots \alpha_k} = [w + (v \partial_v - r \partial_r)] S_{\alpha_1 \alpha_2 \dots \alpha_k} \quad (\text{B.2})$$

where  $\xi = v \partial_v - r \partial_r$  and we identify the quantity  $w$  as the boost weight of the covariant tensor  $S_{\alpha_1 \alpha_2 \dots \alpha_k}$  as

$$\begin{aligned} w &\equiv \text{net boost weight of a covariant tensor (with all indices lowered)} \\ &= \text{number of lower } v \text{ indices} - \text{number of lower } r \text{ indices.} \end{aligned} \quad (\text{B.3})$$

To prove the equivalence, we consider a generic tensor component with  $m_v$  lower  $v$  indices and  $m_r$  lower  $r$  indices:  $S_{v \dots v r \dots r}(r, v, x^i)$  where we have explicitly indicated the functional dependence on the space-time variables. Let us consider an infinitesimal Boost transformation:

$$\tilde{r} = (1 + \epsilon)r \quad \tilde{v} = (1 - \epsilon)v \quad \text{where } \epsilon \text{ is infinitesimal} \quad (\text{B.4})$$

Under the above transformation, the covariant tensor component  $S_{v \dots v r \dots r}$  transforms as follows:

$$\begin{aligned} S_{\tilde{v} \dots \tilde{v} \tilde{r} \dots \tilde{r}}(\tilde{r}, \tilde{v}, x^i) &= \left(\frac{\partial v}{\partial \tilde{v}}\right)^{m_v} \left(\frac{\partial r}{\partial \tilde{r}}\right)^{m_r} S_{v \dots v r \dots r}(r, v, x^i) \\ &= (1 + \epsilon)^{m_v} (1 - \epsilon)^{m_r} S_{v \dots v r \dots r}(r, v, x^i) \\ &= (1 + (m_v - m_r)\epsilon) S_{v \dots v r \dots r}(r, v, x^i) + O(\epsilon^2) \end{aligned} \quad (\text{B.5})$$

Now we will extract the Lie Derivative along the Integral curves of this particular coordinate transformation:

$$\begin{aligned} S_{\tilde{v} \dots \tilde{v} \tilde{r} \dots \tilde{r}}(r + \epsilon r, v - \epsilon v, x^i) &= (1 + (m_v - m_r)\epsilon) S_{v \dots v r \dots r}(r, v, x^i) + O(\epsilon^2) \\ S_{\tilde{v} \dots \tilde{v} \tilde{r} \dots \tilde{r}}(r, v, x^i) + \epsilon r \partial_r S_{\tilde{v} \dots \tilde{v} \tilde{r} \dots \tilde{r}}(r, v, x^i) - \epsilon v \partial_v S_{\tilde{v} \dots \tilde{v} \tilde{r} \dots \tilde{r}}(r, v, x^i) &+ O(\epsilon^2) \\ &= (1 + (m_v - m_r)\epsilon) S_{v \dots v r \dots r}(r, v, x^i) + O(\epsilon^2) \\ S_{\tilde{v} \dots \tilde{v} \tilde{r} \dots \tilde{r}}(r, v, x^i) - S_{v \dots v r \dots r}(r, v, x^i) &= (m_v - m_r)\epsilon S_{v \dots v r \dots r}(r, v, x^i) \\ &+ \epsilon v \partial_v S_{v \dots v r \dots r}(r, v, x^i) - \epsilon r \partial_r S_{v \dots v r \dots r}(r, v, x^i) + O(\epsilon^2) \\ \implies \mathcal{L}_\xi S_{v \dots v r \dots r}(r, v, x^i) &= (m_v - m_r) S_{v \dots v r \dots r}(r, v, x^i) \\ &+ v \partial_v S_{v \dots v r \dots r}(r, v, x^i) - r \partial_r S_{v \dots v r \dots r}(r, v, x^i) \end{aligned} \quad (\text{B.6})$$

where in the final step we have defined  $\xi^\mu \partial_\mu = v \partial_v - r \partial_r$  and the coordinate transformation is generated along the integral curves of this vector field. Thus, we have proved the

equivalence between the two definitions of Boost Weight of the tensor component  $S_{v\dots vr\dots r}$  which is given by  $w = m_v - m_r$ . The generalization of the above equivalence for tensors of the form  $S_{v\dots vij\dots}(r, v, x^i)$  and  $S_{r\dots rij\dots}(r, v, x^i)$  is straightforward once we realise that the Boost transformation doesn't change the coordinates  $x^i$ . Thus, for  $S_{v\dots vij\dots}(r, v, x^i)$ , we essentially have  $m_r = 0$  and for  $S_{r\dots rij\dots}(r, v, x^i)$ , we have  $m_v = 0$ .

## B.2 Structure of covariant tensors in the stationary background solution

We will now show that if  $\xi$  is a Killing vector for our space-time, then the Lie derivative of any covariant tensor evaluated on that space-time will vanish. We will consider the above generic tensor component  $S_{v\dots vr\dots r}$  which has  $m_v$  number of lower  $v$  indices and  $m_r$  number of lower  $r$  indices. Assume that once the tensor component is evaluated in our gauge, it has  $m_v^{(1)}$  factors of  $r$  and  $m_v^{(2)}$  factors of  $\partial_v$ 's such that  $m_v^{(1)} + m_v^{(2)} = m_v$ . Thus, a Boost weight of  $w = m_v - m_r$  implies that the tensor component takes the following form

$$S_{v\dots vr\dots r} = r^{m_v^{(1)}} (\partial_v)^{m_v^{(2)}} (\partial_r)^{m_r} Y(rv) \quad (\text{B.7})$$

where  $Y$  is a function of the product  $rv$  so that it has Boost weight zero. Using this, we further simplify the tensor component:

$$\begin{aligned} S_{v\dots vr\dots r} &= r^{m_v^{(1)}} (\partial_v)^{m_v^{(2)}} (\partial_r)^{m_r} Y(rv) \\ &= r^{m_v^{(1)}} (\partial_v)^{m_v^{(2)}} \left( v^{m_r} \left[ \left( \frac{d}{dz} \right)^{m_r} Y(z) \right]_{z=rv} \right) \\ &= r^{m_v^{(1)}} \sum_{m=0}^k {}^k C_m [(\partial_v)^m v^{m_r}] (\partial_v)^{m_v^{(2)}-m} \left[ \left( \frac{d}{dz} \right)^{m_r} Y(z) \right]_{z=rv} \quad \text{where } k = \text{Min}(m_r, m_v^{(2)}) \\ &= r^{m_v^{(1)}+m_v^{(2)}} \sum_{m=0}^k {}^k C_m [(\partial_v)^m v^{m_r}] r^{-m} \left[ \left( \frac{d}{dz} \right)^{m_r+m_v^{(2)}-m} Y(z) \right]_{z=rv} \\ &= r^{m_v} \sum_{m=0}^k {}^k C_m \left[ \frac{m_r!}{(m_r-m)!} \right] v^{m_r-m} r^{-m} \left[ \left( \frac{d}{dz} \right)^{m_r+m_v^{(2)}-m} Y(z) \right]_{z=rv} \\ &= r^{m_v} v^{m_r} \sum_{m=0}^k {}^k C_m \left[ \frac{m_r!}{(m_r-m)!} \right] (rv)^{-m} \left[ \left( \frac{d}{dz} \right)^{m_r+m_v^{(2)}-m} Y(z) \right]_{z=rv} \\ &= r^{m_v} v^{m_r} \tilde{Y}(rv) \end{aligned} \quad (\text{B.8})$$

where

$$\tilde{Y}(rv) = \sum_{m=0}^k {}^k C_m \left[ \frac{m_r!}{(m_r-m)!} \right] (rv)^{-m} \left[ \left( \frac{d}{dz} \right)^{m_r+m_v^{(2)}-m} Y(z) \right]_{z=rv}, \quad k = \text{Min}(m_r, m_v^{(2)})$$

We use the final result of (B.8) to evaluate the Lie derivative of the tensor component:

$$\begin{aligned} \mathcal{L}_\xi S_{v\dots vr\dots r} &= (m_v - m_r + v\partial_v - r\partial_r)(r^{m_v} v^{m_r} \tilde{Y}(rv)) \\ &= (m_v - m_r + m_r - m_v) r^{m_v} v^{m_r} \tilde{Y}(rv) = 0 \end{aligned} \quad (\text{B.9})$$

where in the final step we have used the fact that  $(v\partial_v - r\partial_r)\tilde{Y}(rv) = 0$ . This vanishing of the Lie derivative on our stationary space-time was possible only after we have explicitly worked out the  $\partial_r$ 's and  $\partial_v$ 's in terms of factors of  $r$ 's and  $v$ 's.

## C Arguments justifying eq. (3.8)

In this appendix we provide some justifications in support of eq. (3.8). According to the definition of  $\Theta^\mu$  in a general higher derivative theory of gravity one can write it as a sum of two terms, i.e.  $\Theta \sim \text{Term}_1 + \text{Term}_2$ . This will be argued in the section 3.2.2 (see eq. (3.24)). One of them, named  $\text{Term}_1$ , has the following generic structure

$$\text{Term}_1 \sim [\text{Tensor-1}] \mathcal{L}_\xi [\text{Tensor-2}] ,$$

with appropriate index contractions being implied on the r.h.s. above. On the other hand,  $\text{Term}_2$  has a little bit different structure than  $\text{Term}_1$ , since it involves a covariant derivative of  $L_\xi$ :

$$\text{Term}_2 \sim D_\nu [E^{r\mu\nu\rho} \mathcal{L}_\xi g_{\mu\rho}] |_{r=0} .$$

However, if we explicitly evaluate this term at the horizon ( $r = 0$ ) in our coordinate system we will find that it has the following form:

$$\text{Term}_2 = E^{rijr} (-1 + v\partial_v) (\partial_r h_{ij}) + (1 + v\partial_v) [E^{rjkv} (\partial_v h_{jk})] .$$

A more careful derivation of this can be found in section 3.2.2 (see eq. (3.49), eq. (3.50) and eq. (3.53)). Although note that the term  $\partial_r h_{ij}$  is not a covariant tensor component, the operator

$$\mathcal{L}_\xi = \text{boost weight} + v\partial_v$$

has a well-defined action on it and we could express this term as  $E^{rijr} \mathcal{L}_\xi (\partial_r h_{ij})$ . Since for our analysis we do not need any details other than boost weight and distribution of  $\partial_v$ s and  $\partial_r$ s, for our purpose, this term is also effectively of the same form as  $\text{Term}_1$ . We do not need to process the second term  $(1 + v\partial_v) [E^{rjkv} (\partial_v h_{jk})]$  since it is already of the structure of equation eq. (3.8) with  $\mathcal{M}_{(0)}$  set to zero.

So finally we have to analyze terms of the of the form of  $\text{Term}_1$ .

$$\begin{aligned} \text{A typical term in } \Theta^r \text{ generated from } \text{Term}_1 &\sim \tilde{t}_{(-k)} (k + 1 + v \partial_v) t_{(k+1)} \\ \Rightarrow \Theta_{(1)} &\sim (k + 1) \tilde{t}_{(-k)} t_{(k+1)}, \quad \Theta_{(2)} = \tilde{t}_{(-k)} \partial_v t_{(k+1)} \end{aligned} \quad (\text{C.1})$$

For these terms, eq. (3.8) reduces to

$$\tilde{t}_{(-k)} \partial_v t_{(k+1)} - \partial_v [(k + 1) \tilde{t}_{(-k)} t_{(k+1)}] = \partial_v^2 \text{ (JKM type terms)} \quad (\text{C.2})$$

Note that for  $k = 0$  this is trivially true with no JKM type terms.

Now extending the arguments in [14] (by transferring  $\partial_v$  s to negative boost weight terms after adding total  $\partial_v$  terms), we could show that a typical term of positive boost weight ( $n+1$ ) could always be expressed as (see eq. (3.14) in section 3.2.1 for more discussion on this)

$$\begin{aligned} \tilde{T}_{(-k)} \partial_v^{n+1+k} T_{(0)} &= \partial_v^{n+1} [\text{JKM type terms}] + (-1)^k [{}^{k+n}C_n] [\tilde{T}_{(0)} \partial_v^{n+1} T_{(0)}] \\ \text{where } \tilde{T}_{(0)} &= \partial_v^k \tilde{T}_{(-k)} \end{aligned} \quad (\text{C.3})$$

Applying the result eq. (C.3) to the two terms  $\tilde{t}_{(-k)} t_{(k+1)}$  and  $\tilde{t}_{(-k)} \partial_v t_{(k+1)}$  we find the following relations

$$\tilde{t}_{(-k)} t_{(k+1)} \sim \tilde{t}_{(-k)} \partial_v^{(k+1)} t_{(0)} = \partial_v [\text{JKM type terms}] + (-1)^k [\tilde{t}_{(0)} \partial_v t_{(0)}] \quad (\text{C.4})$$

by putting  $(n = 0)$  in eq. (C.3), and similarly

$$\tilde{t}_{(-k)} \partial_v t_{(k+1)} \sim \tilde{t}_{(-k)} \partial_v^{(k+2)} t_{(0)} = \partial_v^2 [\text{JKM type terms}] + (-1)^k (k+1) [\tilde{t}_{(0)} \partial_v t_{(0)}] \quad (\text{C.5})$$

by putting  $(n = 1)$  in eq. (C.3). Equation (C.2) follows immediately from equation eq. (C.4) and eq. (C.5).

## D Detailed proof of Result: 1 in eq. (3.14)

A typical component of a covariant tensor with boost weight  $w = a + 1 > 0$  will have the following structure when restricted to the horizon

$$t_{(a+1)}^{(k)}|_{r=0} = \tilde{T}_{(-k)} \partial_v^{(k+a+1)} T_{(0)}|_{r=0} + \mathcal{O}(\epsilon^2). \quad (\text{D.1})$$

Here the subscripts denote the boost weight and as before, for convenience we have suppressed the index structure for  $\tilde{T}$  and  $T$ .

Our aim is to prove Result:1 which states that  $t_{(a+1)}^{(k)}$  could always be re-expressed as follows

$$\begin{aligned} t_{(a+1)}^{(k)} &= \tilde{T}_{(-k)} \partial_v^{(k+a+1)} T_{(0)} \\ &= \partial_v^{a+1} \left[ \sum_{m=0}^{k-1} (-1)^m \left[ {}^{m+a} C_m \right] \left( \tilde{T}_{(-k+m)} \partial_v^{k-m} T_{(0)} \right) \right] + (-1)^k \left[ {}^{k+a} C_a \right] \tilde{T}_{(0)} \partial_v^{a+1} T_{(0)} \end{aligned} \quad (\text{D.2})$$

where  $\tilde{T}_{(-k+m)} \equiv \partial_v^m \tilde{T}_{(-k)}$

### D.1 The proof

The basic idea in the proof is to carefully rearrange the  $v$  derivatives and to read off the binomial coefficients appropriately. We use the method of induction to prove our desired result.

We check that eq. (D.2) is true for  $a = 0$ :

$$\begin{aligned} t_{(1)}^{(k)} &= \tilde{T}_{(-k)} \partial_v^{(k+1)} T_{(0)} \\ &= \partial_v \left[ \tilde{T}_{(-k)} \partial_v^{(k)} T_{(0)} \right] - \tilde{T}_{(-k+1)} \partial_v^{(k)} T_{(0)} \\ &= \partial_v \left[ \tilde{T}_{(-k)} \partial_v^{(k)} T_{(0)} - \tilde{T}_{(-k+1)} \partial_v^{(k-1)} T_{(0)} \right] + \tilde{T}_{(-k+2)} \partial_v^{(k-1)} T_{(0)} \\ &= \dots \\ &= \partial_v \left[ \sum_{m=0}^{k-1} (-1)^m \tilde{T}_{(-k+m)} \partial_v^{k-m} T_{(0)} \right] + (-1)^k \tilde{T}_{(0)} \partial_v T_{(0)} \\ &= \partial_v \left[ \sum_{m=0}^{k-1} (-1)^m \left[ {}^{m+0} C_0 \right] \tilde{T}_{(-k+m)} \partial_v^{k-m} T_{(0)} \right] + (-1)^k \left[ {}^{k+0} C_0 \right] \tilde{T}_{(0)} \partial_v T_{(0)} \end{aligned} \quad (\text{D.3})$$

where the fourth step indicates that we must extract all the  $k$  derivatives in the form of total derivative. The final step indicates that  $t_{(1)}^{(k)}$  is of the form predicted by eq. (D.2). Next, we shall show that if the result eq. (D.2) is true for  $a = n$ , then it is also true for  $a = n + 1$ .

The statement for  $a = n$  which we assume to be true is given by:

$$\begin{aligned}
 t_{(n+1)}^{(k)} &= \tilde{T}_{(-k)} \partial_v^{(k+n+1)} T_{(0)} \\
 &= \partial_v^{n+1} \left[ \sum_{m=0}^{k-1} (-1)^m \binom{m+n}{n} C_n \left( \tilde{T}_{(-k+m)} \partial_v^{k-m} T_{(0)} \right) \right] + \binom{k+n}{n} C_n \left( \tilde{T}_{(0)} \partial_v^{n+1} T_{(0)} \right) \\
 \text{where } \tilde{T}_{(-k+m)} &= \partial_v^m \tilde{T}_{(-k)}
 \end{aligned} \tag{D.4}$$

We shall now process  $t_{(n+2)}^{(k)} = \tilde{T}_{(-k)} \partial_v^{(k+n+2)} T_{(0)}$

$$\begin{aligned}
 t_{(n+2)}^{(k)} &= \tilde{T}_{(-k)} \partial_v^{(k+n+2)} T_{(0)} \\
 &= \partial_v \left[ \tilde{T}_{(-k)} \partial_v^{(k+n+1)} T_{(0)} \right] - \tilde{T}_{(-k+1)} \partial_v^{(k+n+1)} T_{(0)} \\
 &= \dots \\
 &= \partial_v \left[ \sum_{m=0}^{k-1} (-1)^m \tilde{T}_{(-k+m)} \partial_v^{(k+n+1-m)} T_{(0)} \right] + (-1)^k \tilde{T}_{(0)} \partial_v^{(n+2)} T_{(0)}
 \end{aligned} \tag{D.5}$$

In the first term of the final step, each term within the summation is of the form  $t_{(n+1)}^{(k-m)}$ . So we shall apply our previous assumption eq. (D.4) here for  $a = n$ .

$$\begin{aligned}
 t_{(n+2)}^{(k)} &= \tilde{T}_{(-k)} \partial_v^{(k+n+2)} T_{(0)} \\
 &= \partial_v \left[ \sum_{m=0}^{k-1} (-1)^m \tilde{T}_{(-k+m)} \partial_v^{(k+n+1-m)} T_{(0)} \right] + (-1)^k \tilde{T}_{(0)} \partial_v^{(n+2)} T_{(0)} \\
 &= \partial_v \sum_{m=0}^{k-1} (-1)^m \left( \partial_v^{n+1} \left[ \sum_{p=0}^{k-m-1} (-1)^p \binom{n+p}{n} C_n \tilde{T}_{(-k+m+p)} \partial_v^{(k-m-p)} T_{(0)} \right] \right. \\
 &\quad \left. + (-1)^{k-m} \binom{k-m+n}{n} C_n \tilde{T}_{(0)} \partial_v^{n+1} T_{(0)} \right) + (-1)^k \tilde{T}_{(0)} \partial_v^{(n+2)} T_{(0)} \\
 &= \partial_v^{n+2} \left( \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} (-1)^{m+p} \binom{n+p}{n} C_n \left[ \tilde{T}_{(-k+m+p)} \partial_v^{(k-m-p)} T_{(0)} \right] \right) \\
 &\quad + (-1)^k \left[ 1 + \sum_{m=0}^{k-1} \binom{k-m+n}{n} C_n \right] \tilde{T}_{(0)} \partial_v^{n+2} T_{(0)} \\
 &= \partial_v^{n+2} \left( \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} (-1)^{m+p} \binom{n+p}{n} C_n \left[ \tilde{T}_{(-k+m+p)} \partial_v^{(k-m-p)} T_{(0)} \right] \right) \\
 &\quad + (-1)^k \left[ \sum_{m=0}^k \binom{k-m+n}{n} C_n \right] \tilde{T}_{(0)} \partial_v^{n+2} T_{(0)}
 \end{aligned} \tag{D.6}$$

The final result of eq. (D.6) is almost of the desired form in terms of the structure of  $\partial_v$  derivatives, but now we have to transform the coefficients to the desired form. For this purpose we use the following identity which is proved at the end of this section:

$$\sum_{m=0}^k \binom{k-m+n}{C_n} = {}^{k+n+1}C_{n+1} \tag{D.7}$$

Substituting in eq. (D.6) we find

$$\begin{aligned} t_{(n+2)}^{(k)} &= \tilde{T}_{(-k)} \partial_v^{(k+n+2)} T_{(0)} \\ &= \partial_v^{n+2} \left( \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} (-1)^{m+p} \binom{n+p}{C_n} \left[ \tilde{T}_{(-k+m+p)} \partial_v^{(k-m-p)} T_{(0)} \right] \right) \\ &\quad + (-1)^k \left[ \sum_{m=0}^k \binom{k-m+n}{C_n} \right] \left( \tilde{T}_{(0)} \partial_v^{n+2} T_0 \right) \\ &= \partial_v^{n+2} \left( \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} (-1)^{m+p} \binom{n+p}{C_n} \left[ \tilde{T}_{(-k+m+p)} \partial_v^{(k-m-p)} T_{(0)} \right] \right) \\ &\quad + (-1)^k \left[ {}^{k+n+1}C_{n+1} \right] \tilde{T}_{(0)} \partial_v^{n+2} T_0 \end{aligned} \tag{D.8}$$

So the second term in eq. (D.6) has been cast into the desired form. We will further process the double sum in the first term. Note that the sum has the following form

$$\sim \sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} g(p) f(m+p),$$

where  $g(p) \equiv \binom{n+p}{C_n}$  and  $f(m+p) \equiv (-1)^{m+p} \left[ \tilde{T}_{(-k+m+p)} \partial_v^{(k-m-p)} T_{(0)} \right]$ . Thus, we have

$$\begin{aligned} &\sum_{m=0}^{k-1} \sum_{p=0}^{k-m-1} g(p) f(m+p) \\ &= \sum_{m=0}^{k-1} \sum_{q=m}^{k-1} g(q-m) f(q) \quad m \text{ has been changed to } q = (m+p) \\ &= \sum_{q=0}^{k-1} \left[ \sum_{m=0}^q g(q-m) \right] f(q) = \sum_{q=0}^{k-1} \left[ \sum_{m=0}^q \binom{n+q-m}{C_n} \right] f(q) \\ &= \sum_{q=0}^{k-1} {}^{n+q+1}C_{n+1} f(q) \end{aligned} \tag{D.9}$$

Substituting eq. (D.9) in eq. (D.8) we get the required result eq. (D.2) that we set out to prove

$$\begin{aligned} t_{(n+2)}^{(k)} &= \partial_v^{n+2} \left( \sum_{m=0}^{k-1} (-1)^m \binom{n+1+m}{C_{n+1}} \left[ \tilde{T}_{(-k+m)} \partial_v^{(k-m)} T_{(0)} \right] \right) \\ &\quad + (-1)^k \left[ {}^{k+n+1}C_{n+1} \right] \left( \tilde{T}_{(0)} \partial_v^{n+2} T_0 \right) \end{aligned} \tag{D.10}$$

## D.2 Proof of the identity eq. (D.7)

Consider the finite sum

$$\begin{aligned}
 S(x) &= \sum_{m=0}^k (1+x)^{k+n-m} = \sum_{m=0}^k \sum_{p=0}^{k+n-m} \binom{k-m+n}{p} x^p \\
 &= (1+x)^n \sum_{m=0}^k (1+x)^{k-m} = (1+x)^n \left[ \frac{(1+x)^{k+1} - 1}{x} \right] \\
 &= \left[ \frac{(1+x)^{k+n+1} - (1+x)^n}{x} \right]
 \end{aligned} \tag{D.11}$$

Equating the coefficient of  $x^n$  on both sides of eq. (D.11), we get

$$\begin{aligned}
 \text{Coefficient of } x^n \text{ in } S(x) &= \text{Coefficient of } x^{n+1} \text{ in } \left[ (1+x)^{k+n+1} - (1+x)^n \right] \\
 \Rightarrow \sum_{m=0}^k \binom{k-m+n}{n} &= {}^{k+n+1}C_{n+1}
 \end{aligned} \tag{D.12}$$

Hence proved.

## E Variation of a generic covariant tensor under diffeomorphism: a justification for eq. (3.23)

Suppose we have a Diffeomorphism Covariant tensor  $S_{\alpha_1 \dots \alpha_k}$  which is a functional of various dynamical fields  $\{\psi\}$  (such as  $g_{\mu\nu}, R_{\mu\nu\rho\sigma}, D_{\alpha_1} R_{\mu\nu\rho\sigma}, \dots, \phi, D_{\alpha_1} \phi, \dots$ ), the notion of Diffeomorphism covariance under the vector field  $\xi^\mu \partial_\mu$  is expressed by the following equation:

$$\delta S_{\alpha_1 \dots \alpha_k} [\delta\psi \rightarrow \mathcal{L}_\xi \psi] = \mathcal{L}_\xi S_{\alpha_1 \dots \alpha_k} \tag{E.1}$$

where  $\xi^\mu \partial_\mu = v \partial_v - r \partial_r$  which is a killing vector of our background Space-time eq. (2.7). The reason for this specific choice will be clear in a moment. It should be emphasized that in the l.h.s. of eq. ((E.1)), the variation is induced by the variation of the dynamical fields. More precisely,

$$\delta S_{\alpha_1 \dots \alpha_k} [\delta\psi \rightarrow \mathcal{L}_\xi \psi] = \lim_{\epsilon \rightarrow 0} \frac{S_{\alpha_1 \dots \alpha_k} [\psi + \epsilon \mathcal{L}_\xi \psi] - S_{\alpha_1 \dots \alpha_k} [\psi]}{\epsilon} \tag{E.2}$$

Whereas, in the r.h.s. of eq. ((E.1)), the variation is induced by a space-time Lie derivative or, in other words, the variation induced by a change of co-ordinate charts along the flow of the integral curves of  $\xi^\mu \partial_\mu$  [6]. Thus, the variation in the l.h.s. affects only the dynamical fields, whereas the variation in the r.h.s. affects both the dynamical and background fields if any [41]. For the equality to hold, we must make sure that the variations of all the background fields are set to zero. Hence, the choice of  $\xi^\mu \partial_\mu$  is clear: it is a symmetry of the only background field present in the problem eq. (2.7) (i.e.)  $\mathcal{L}_\xi g_{\mu\nu}^{eq} = 0$  and hence the equality of eq. (E.1) holds. We provide a few examples which explicitly demonstrate this.



When one applies this notion of Diffeomorphism covariance to the Lagrangian of a theory, we get

$$\frac{\partial L}{\partial \psi} \mathcal{L}_\xi \psi = \mathcal{L}_\xi L[\psi], \quad (\text{E.3})$$

where the l.h.s. is the usual variation of the Lagrangian with respect to the fields with  $\delta\psi = \mathcal{L}_\xi \psi$ .

**A scalar field coupled to non-dynamical gravity.** To illustrate this Diffeomorphism covariance, let us consider a simple scalar field Lagrangian:

$$L = g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi \quad (\text{E.4})$$

The Lie derivative is given by

$$\mathcal{L}_\xi L = \xi^\mu \partial_\mu L \quad (\text{E.5})$$

The variation of the Lagrangian is however given by

$$\delta L = \delta g^{\mu\nu} \partial_\mu \varphi \partial_\nu \varphi + 2g^{\mu\nu} \partial_\mu \delta \varphi \partial_\nu \varphi \quad (\text{E.6})$$

Substituting  $\delta g_{\mu\nu} = \mathcal{L}_\xi g_{\mu\nu} = D_\mu \xi_\nu + D_\nu \xi_\mu$  and  $\delta \varphi = \mathcal{L}_\xi \varphi = \xi^\rho \partial_\rho \varphi$ , we get

$$\begin{aligned} \delta L[\delta\psi \rightarrow \mathcal{L}_\xi \psi] &= -(D^\mu \xi^\nu + D^\nu \xi^\mu) \partial_\mu \varphi \partial_\nu \varphi + 2g^{\mu\nu} \partial_\mu (\xi^\rho \partial_\rho \varphi) \partial_\nu \varphi \\ &= -(\partial^\mu \xi^\nu + \partial^\nu \xi^\mu + g^{\mu\rho} \Gamma_{\rho\sigma}^\nu \xi^\sigma + g^{\nu\rho} \Gamma_{\rho\sigma}^\mu \xi^\sigma) \partial_\mu \varphi \partial_\nu \varphi \\ &\quad + 2g^{\mu\nu} \partial_\mu \xi^\rho \partial_\rho \varphi \partial_\nu \varphi + 2g^{\mu\nu} \xi^\rho \partial_\rho (\partial_\mu \varphi) \partial_\nu \varphi \\ &= g^{\mu\nu} \xi^\rho D_\rho (\partial_\mu \varphi \partial_\nu \varphi) \\ &= \xi^\rho \partial_\rho L = \mathcal{L}_\xi L \end{aligned} \quad (\text{E.7})$$

where the end result is what you get by computing the Lie derivative directly.

**Einstein's theory of gravity.** We now consider a general diffeomorphism invariant theory. Assuming eq. (E.3), it has been shown [8] that such a Lagrangian can be re-expressed as

$$L = L(g_{\mu\nu}, R_{\mu\nu\rho\sigma}, D_{\alpha_1} R_{\mu\nu\rho\sigma}, \dots, D_{(\alpha_1} \dots D_{\alpha_m)} R_{\mu\nu\rho\sigma}, \phi, D_{\alpha_1} \phi, \dots, D_{(\alpha_1} \dots D_{\alpha_m)} \phi) \quad (\text{E.8})$$

where  $\phi$  collectively denote the set of matter fields in the theory. Hence, with this choice of Lagrangian, the variation of the metric  $g_{\mu\nu}$  and  $R_{\mu\nu\rho\sigma}$  are treated independently.

To again illustrate this diffeomorphism covariance for a more complicated theory, consider General relativity with  $L = R$ :

$$\begin{aligned} \delta L &= 2(\delta g^{\alpha\gamma}) g^{\beta\delta} R_{\alpha\beta\gamma\delta} + g^{\alpha\gamma} g^{\beta\delta} \delta R_{\alpha\beta\gamma\delta} \\ \implies \delta L[\delta g_{\mu\nu} \rightarrow \mathcal{L}_\xi g_{\mu\nu}] &= 2(\mathcal{L}_\xi g^{\alpha\gamma}) g^{\beta\delta} R_{\alpha\beta\gamma\delta} + g^{\alpha\gamma} g^{\beta\delta} \mathcal{L}_\xi R_{\alpha\beta\gamma\delta} \\ &= \mathcal{L}_\xi (g^{\alpha\gamma} g^{\beta\delta} R_{\alpha\beta\gamma\delta}) = \mathcal{L}_\xi R \end{aligned} \quad (\text{E.9})$$

This Diffeomorphism covariance eq. (E.1) has been used as an important definition [8] to prove the first law of black hole thermodynamics for any diffeomorphism invariant theory.

**Open Access.** This article is distributed under the terms of the Creative Commons Attribution License ([CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)), which permits any use, distribution and reproduction in any medium, provided the original author(s) and source are credited.

## References

- [1] S.W. Hawking, *Gravitational radiation from colliding black holes*, *Phys. Rev. Lett.* **26** (1971) 1344 [[INSPIRE](#)].
- [2] J.M. Bardeen, B. Carter and S.W. Hawking, *The Four laws of black hole mechanics*, *Commun. Math. Phys.* **31** (1973) 161 [[INSPIRE](#)].
- [3] J.D. Bekenstein, *Black holes and entropy*, *Phys. Rev. D* **7** (1973) 2333 [[INSPIRE](#)].
- [4] S.W. Hawking, *Particle Creation by Black Holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [[INSPIRE](#)].
- [5] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge Monographs on Mathematical Physics, Cambridge University Press (2011) [[DOI](#)] [[INSPIRE](#)].
- [6] R.M. Wald, *General relativity*, Chicago University Press, Chicago, IL (1984) [[DOI](#)] [[INSPIRE](#)].
- [7] R.M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev. D* **48** (1993) R3427 [[gr-qc/9307038](#)] [[INSPIRE](#)].
- [8] V. Iyer and R.M. Wald, *Some properties of Noether charge and a proposal for dynamical black hole entropy*, *Phys. Rev. D* **50** (1994) 846 [[gr-qc/9403028](#)] [[INSPIRE](#)].
- [9] T. Jacobson and R.C. Myers, *Black hole entropy and higher curvature interactions*, *Phys. Rev. Lett.* **70** (1993) 3684 [[hep-th/9305016](#)] [[INSPIRE](#)].
- [10] T. Jacobson, G. Kang and R.C. Myers, *On black hole entropy*, *Phys. Rev. D* **49** (1994) 6587 [[gr-qc/9312023](#)] [[INSPIRE](#)].
- [11] T. Jacobson, G. Kang and R.C. Myers, *Increase of black hole entropy in higher curvature gravity*, *Phys. Rev. D* **52** (1995) 3518 [[gr-qc/9503020](#)] [[INSPIRE](#)].
- [12] S. Sarkar and A.C. Wall, *Generalized second law at linear order for actions that are functions of Lovelock densities*, *Phys. Rev. D* **88** (2013) 044017 [[arXiv:1306.1623](#)] [[INSPIRE](#)].
- [13] S. Bhattacharjee, S. Sarkar and A.C. Wall, *Holographic entropy increases in quadratic curvature gravity*, *Phys. Rev. D* **92** (2015) 064006 [[arXiv:1504.04706](#)] [[INSPIRE](#)].
- [14] A.C. Wall, *A Second Law for Higher Curvature Gravity*, *Int. J. Mod. Phys. D* **24** (2015) 1544014 [[arXiv:1504.08040](#)] [[INSPIRE](#)].
- [15] S. Bhattacharjee, A. Bhattacharyya, S. Sarkar and A. Sinha, *Entropy functionals and c-theorems from the second law*, *Phys. Rev. D* **93** (2016) 104045 [[arXiv:1508.01658](#)] [[INSPIRE](#)].
- [16] A.C. Wall, *A proof of the generalized second law for rapidly changing fields and arbitrary horizon slices*, *Phys. Rev. D* **85** (2012) 104049 [*Erratum ibid.* **87** (2013) 069904] [[arXiv:1105.3445](#)] [[INSPIRE](#)].
- [17] A.C. Wall, *A Survey of Black Hole Thermodynamics*, [arXiv:1804.10610](#) [[INSPIRE](#)].
- [18] S. Sarkar, *Black Hole Thermodynamics: General Relativity and Beyond*, *Gen. Rel. Grav.* **51** (2019) 63 [[arXiv:1905.04466](#)] [[INSPIRE](#)].

- [19] T. Liko, *Topological deformation of isolated horizons*, *Phys. Rev. D* **77** (2008) 064004 [[arXiv:0705.1518](#)] [[INSPIRE](#)].
- [20] S. Sarkar and A.C. Wall, *Second Law Violations in Lovelock Gravity for Black Hole Mergers*, *Phys. Rev. D* **83** (2011) 124048 [[arXiv:1011.4988](#)] [[INSPIRE](#)].
- [21] S. Chatterjee and M. Parikh, *The second law in four-dimensional Einstein-Gauss-Bonnet gravity*, *Class. Quant. Grav.* **31** (2014) 155007 [[arXiv:1312.1323](#)] [[INSPIRE](#)].
- [22] J. Bhattacharya, S. Bhattacharyya, A. Dinda and N. Kundu, *An entropy current for dynamical black holes in four-derivative theories of gravity*, *JHEP* **06** (2020) 017 [[arXiv:1912.11030](#)] [[INSPIRE](#)].
- [23] S. Gao and R.M. Wald, *The ‘Physical process’ version of the first law and the generalized second law for charged and rotating black holes*, *Phys. Rev. D* **64** (2001) 084020 [[gr-qc/0106071](#)] [[INSPIRE](#)].
- [24] A.J. Amsel, D. Marolf and A. Virmani, *The Physical Process First Law for Bifurcate Killing Horizons*, *Phys. Rev. D* **77** (2008) 024011 [[arXiv:0708.2738](#)] [[INSPIRE](#)].
- [25] S. Bhattacharjee and S. Sarkar, *Physical process first law and caustic avoidance for Rindler horizons*, *Phys. Rev. D* **91** (2015) 024024 [[arXiv:1412.1287](#)] [[INSPIRE](#)].
- [26] A. Mishra, S. Chakraborty, A. Ghosh and S. Sarkar, *On the physical process first law for dynamical black holes*, *JHEP* **09** (2018) 034 [[arXiv:1709.08925](#)] [[INSPIRE](#)].
- [27] A. Chatterjee and S. Sarkar, *Physical process first law and increase of horizon entropy for black holes in Einstein-Gauss-Bonnet gravity*, *Phys. Rev. Lett.* **108** (2012) 091301 [[arXiv:1111.3021](#)] [[INSPIRE](#)].
- [28] S. Kolekar, T. Padmanabhan and S. Sarkar, *Entropy Increase during Physical Processes for Black Holes in Lanczos-Lovelock Gravity*, *Phys. Rev. D* **86** (2012) 021501 [[arXiv:1201.2947](#)] [[INSPIRE](#)].
- [29] S. Bhattacharyya et al., *Local Fluid Dynamical Entropy from Gravity*, *JHEP* **06** (2008) 055 [[arXiv:0803.2526](#)] [[INSPIRE](#)].
- [30] S. Bhattacharyya, *Constraints on the second order transport coefficients of an uncharged fluid*, *JHEP* **07** (2012) 104 [[arXiv:1201.4654](#)] [[INSPIRE](#)].
- [31] S. Bhattacharyya, *Entropy current and equilibrium partition function in fluid dynamics*, *JHEP* **08** (2014) 165 [[arXiv:1312.0220](#)] [[INSPIRE](#)].
- [32] S. Bhattacharyya, *Entropy Current from Partition Function: One Example*, *JHEP* **07** (2014) 139 [[arXiv:1403.7639](#)] [[INSPIRE](#)].
- [33] R. Guedens, T. Jacobson and S. Sarkar, *Horizon entropy and higher curvature equations of state*, *Phys. Rev. D* **85** (2012) 064017 [[arXiv:1112.6215](#)] [[INSPIRE](#)].
- [34] S. Chapman, Y. Neiman and Y. Oz, *Fluid/Gravity Correspondence, Local Wald Entropy Current and Gravitational Anomaly*, *JHEP* **07** (2012) 128 [[arXiv:1202.2469](#)] [[INSPIRE](#)].
- [35] C. Eling, A. Meyer and Y. Oz, *Local Entropy Current in Higher Curvature Gravity and Rindler Hydrodynamics*, *JHEP* **08** (2012) 088 [[arXiv:1205.4249](#)] [[INSPIRE](#)].
- [36] Y. Dandekar and A. Saha, *Large D membrane for Higher Derivative Gravity and Black Hole Second Law*, *JHEP* **02** (2020) 083 [[arXiv:1910.10964](#)] [[INSPIRE](#)].
- [37] A. Saha, *General Theory of Large D Membranes Consistent with Second Law of Thermodynamics*, *JHEP* **04** (2021) 152 [[arXiv:2012.12834](#)] [[INSPIRE](#)].

- [38] S. Bhattacharyya, V.E. Hubeny, S. Minwalla and M. Rangamani, *Nonlinear Fluid Dynamics from Gravity*, *JHEP* **02** (2008) 045 [[arXiv:0712.2456](#)] [[INSPIRE](#)].
- [39] X.O. Camanho, J.D. Edelstein, J. Maldacena and A. Zhiboedov, *Causality Constraints on Corrections to the Graviton Three-Point Coupling*, *JHEP* **02** (2016) 020 [[arXiv:1407.5597](#)] [[INSPIRE](#)].
- [40] S. Bhattacharyya, F.M. Haehl, N. Kundu, R. Loganayagam and M. Rangamani, *Towards a second law for Lovelock theories*, *JHEP* **03** (2017) 065 [[arXiv:1612.04024](#)] [[INSPIRE](#)].
- [41] D. Harlow and J.-Q. Wu, *Covariant phase space with boundaries*, *JHEP* **10** (2020) 146 [[arXiv:1906.08616](#)] [[INSPIRE](#)].