# Causal structure of black holes in shift-symmetric Horndeski theories

Robert Benkel,[1] Nicola Franchini,[1] Mehdi Saravani,[1] and Thomas P. Sotiriou[1, 2, 3]

[1]*School of Mathematical Sciences, University of Nottingham,*
*University Park, Nottingham NG7 2RD, United Kingdom*
[2]*School of Physics and Astronomy, University of Nottingham,*
*University Park, Nottingham NG7 2RD, United Kingdom*
[3]*Centro de Astrofísíca e Gravitação–CENTRA, Instituto Superior Técnico–IST,*
*Universidade de Lisboa–UL, Avenida Rovisco Pais 1, Lisboa 1049-001, Portugal*

In theories with derivative (self-)interactions, the propagation of perturbations on nontrivial field configurations is determined by effective metrics. Generalized scalar-tensor theories belong in this class and this implies that the matter fields and gravitational perturbations do not necessarily experience the same causal structure. Motivated by this, we explore the causal structure of black holes as perceived by scalar fields in the Horndeski class. We consider linearized perturbations on a *fixed* background metric that describes a generic black hole. The effective metric that determines the propagation of these perturbations does not generally coincide with the background metric (to which matter fields couple minimally). Assuming that the metric and the scalar respect stationarity and that the surface gravity of the horizon is constant, we prove that Killing horizons of the background metric are always Killing horizons of the effective metric as well. Hence, scalar perturbations cannot escape the region that matter fields perceive as the interior of the black hole. This result does not depend on asymptotics but only on local considerations and does not make any reference to no-hair theorems. We then demonstrate that, when one relaxes the stationarity assumption for the scalar, solutions where the horizons of the effective and the background metrics do not match can be found in the decoupling limit.

## I. INTRODUCTION

Black holes are among the simplest objects in the Universe and general relativity (GR) predicts that their defining feature, the event horizon, acts as a causal boundary for all fields. Black holes are the perfect probes of the strong-field regime, especially in the era of gravitational wave astronomy. This regime is also where alternative theories of gravity are expected to yield new predictions in the description of compact astrophysical objects. When new fundamental fields become a part of the picture, the simplicity of black holes in GR, as expressed by the no-hair conjecture [1–3], is often replaced by solutions with a more complex structure.

Even the simplest extensions of GR, scalar-tensor theories which postulate that a scalar field takes part in mediating the gravitational interaction, can introduce significant deviations [4, 5]. Consider the most general scalar-tensor theory that leads to second order equations upon direct variation, described by the Horndeski action [6–8]

$$\mathcal{L}_{\text{Horndeski}} = \mathcal{L}_2 + \mathcal{L}_3 + \mathcal{L}_4 + \mathcal{L}_5 \tag{1}$$

where

$$\mathcal{L}_2 = G_2 \,, \tag{2}$$

$$\mathcal{L}_3 = -G_3 \Box \Phi \,, \tag{3}$$

$$\mathcal{L}_4 = G_4 R + G_{4X} \left[ (\Box \Phi)^2 - (\nabla_a \nabla_b \Phi)^2 \right] \,, \tag{4}$$

$$\mathcal{L}_5 = G_5 G_{ab} \nabla^a \nabla^b \Phi \tag{5}$$
$$- \frac{1}{6} G_{5X} \left[ (\Box \Phi)^3 - 3 \Box \Phi (\nabla_a \nabla_b \Phi)^2 + 2 (\nabla_a \nabla_b \Phi)^3 \right] \,, \tag{6}$$

$X = -\frac{1}{2} \nabla_a \Phi \nabla^a \Phi$, $G_i = G_i(\Phi, X)$, and $G_{iX} = \partial_X G_i$. No-hair theorems do exist for broad subclasses of the Horndeski action under certain assumptions: when the scalar does not exhibit derivative self-couplings, $G_2 = G_2(\Phi)$, $G_4 = G_4(\Phi)$, $G_3 = G_5 = 0$, and the configuration is stationary and asymptotically flat [9–12]; when the theory respects shift symmetry, $\Phi \to \Phi+$constant, $G_i = G_i(X)$, assuming staticity and spherical symmetry [13] or slow rotation [14], and provided that the scalar does not couple to the Gauss-Bonnet invariant, $\mathcal{G} = R_{abcd}R^{abcd} - 4R_{ab}R^{ab} + R^2$ [14, 15]. However, there is no known no-hair theorem that covers the complete action. On the contrary, it is known to admit hairy solutions even for simple static, spherically symmetric and asymptotically flat configurations, *e.g.* [14–20]. Moreover, in certain cases it is possible to obtain stationary hairy black hole spacetimes by relaxing the stationarity assumption for the scalar only [21–24].

The existence of hairy black hole solutions in theories where the scalar exhibits derivative self-interactions might have implications for causality. Noncanonical kinetic terms allow for perturbations of the scalar field to propagate superluminally as long as the background field is nontrivial, i.e. has nonvanishing derivatives [25–27]. Indeed, as a simple approximation, consider the Lagrangian (1) as describing a scalar field on a *fixed* curved background. Expanding to second order in scalar perturbations $\pi$ on top of a background $\varphi$, we get the second order Lagrangian for scalar perturbations as follows

$$\mathcal{L}^{(2)}_{\text{Horndeski}} = -\frac{1}{2} f^{ab}[\varphi, g] \partial_a \pi \partial_b \pi. \tag{7}$$

The explicit form of $f^{ab}$ is presented in Appendix A for the special case where $\Phi$ respects shift symmetry. The

propagation of linear scalar perturbations is determined by effective metric $f^{ab}$. Hence, this metric, or its inverse $(f^{-1})_{ab}$, where $f^{ab}(f^{-1})_{bc} = \delta^a_c$, define causality for such excitations. The matter fields are assumed to couple minimally to $g_{ab}$ and hence massless matter perturbations will follow null trajectories of that metric. For a generic background scalar field $\varphi$, $f^{-1}_{ab}$ does not coincide with $g_{ab}$. This suggests that the causal structure of black holes in certain classes of theories described by the Horndeski action should be particularly intriguing. If $f^{-1}_{ab}$ and $g_{ab}$ are not conformally related then they have different null cones, so massless excitations that follow null geodesics of $f^{-1}_{ab}$ can be superluminal or subluminal.

One can take this a step further: for stationary metrics, event horizons will be Killing horizons, so one can ask under which conditions the Killing horizons of $f^{-1}_{ab}$ and $g_{ab}$ coincide and what is the causal structure when they do not. These are precisely the questions that we explore below. We restrict ourselves to a setup where the background metric is fixed and consider only linear scalar perturbations, as in the discussion above. This has obvious limitations: it only probes the local counterpart of the causal structure as perceived by scalar excitations and ignores nonlinear effects. It also does not take into account the role of metric perturbations, which will generally also propagate along null geodesics of a different effective metric. Nonetheless, this setup makes calculation tractable and already provides very interesting insights into the causal structure as perceived by the scalar.

The causal structure of black holes in Horndeski theories has been studied before in [28] and [29] (see [30] and [31] for similar approaches) using the method of characteristics. In Ref. [28] both the metric $g_{ab}$ and the scalar field $\varphi$ were taken to be dynamical and it was claimed that if a surface $\Sigma$ is a Killing horizon of $g_{ab}$, it is also a characteristic surface for all degrees of freedom when the metric and the scalar are stationary. It was then suggested that this implies that $\Sigma$ effectively acts as a horizon for all excitations. This last statement entails some implicit assumptions not spelled out in Ref. [28]. Our approach follows a different path and it provides a rigorous proof about when Killing horizons for different excitations coincide. Though less general due to the decoupling approximation, it is also more transparent physically. Hence, it simplifies the interpretation of the results and highlights certain subtleties and assumptions in the characteristics approach.

The paper is organized as follows: in the next section, we provide a proof that the Killing horizons of the fixed background metric $g_{ab}$ will also be Killing horizons of $f^{-1}_{ab}$, provided that both the metric and the scalar respect stationarity. In Sec. III we relax the assumption of stationarity for the scalar and we show that, in the zero backreaction limit, it is rather straightforward to construct solutions where the horizons of these two metrics do not coincide. Though it is not clear if this feature will survive once backreaction is taken into account, our results certainly motivate further work in this direction.

Sec. IV contains a discussion of our results and their potential implications.

## II. NO-GO RESULTS FOR BLACK HOLES WITH MULTIPLE HORIZONS

As argued above $f^{-1}_{ab}$ and $g_{ab}$ do not coincide for a generic background. In fact, interestingly, they do not coincide even for the trivial background, $\varphi = $ constant. However, we will prove below that the black hole Killing horizon $\mathcal{H}$ perceived by the metric $g_{ab}$ acts as a Killing horizon for $f^{-1}_{ab}$ metric as well. In order to do so, we assume that the black hole spacetime has a Killing vector field $\xi^a$ which is timelike outside the black hole region, is null and orthogonal to $\mathcal{H}$ on the horizon $\mathcal{H}$ and the background scalar field satisfies this symmetry, $\xi^a \nabla_a \varphi = 0$. Moreover, we assume those derivatives of $G_i$'s appearing in the expression for $f^{ab}$ in Appendix A are finite.

Let us start with the easiest case, where the scalar background is trivial, $\varphi = $ constant. The effective metric for linear perturbations around this solution is given by (see Appendix A)

$$f^{ab} = c_1 g^{ab} + c_2 G^{ab} \qquad (8)$$

where $c_1$ and $c_2$ are constants. Consider a spacetime where no extra matter field is present. Then, the variation of the action w.r.t. $g^{ab}$ yields the vacuum Einstein equations $G_{ab} + \lambda g_{ab} = 0$. As a result, $f^{ab}$ and $g^{ab}$ are related by a conformal factor and their causal structure is the same.

In general, when the scalar field background is trivial, the $g_{ab}$ metric satisfies GR field equations, and thus describes a GR black hole. The most general GR black hole solution is described by a Kerr-Newman metric. Though not obvious, we have verified by direct calculation that the Killing horizon of the Kerr-Newman black hole acts as a horizon for the $f^{-1}_{ab}$ metric as well.[*] In the argument above, we have exploited the field equation for the metric to derive our result. In what follows, we provide a proof to show that the Killing horizon of $g_{ab}$ is a Killing horizon of the effective metric $f^{-1}_{ab}$ without any reference to the field equation for $g_{ab}$ (or to the equation for the background scalar field $\varphi$).

$f^{-1}_{ab}$ is a composite metric constructed from the spacetime metric and the scalar field derivatives. Consequently, it satisfies stationarity, i.e. $\mathcal{L}_\xi f^{-1}_{ab} = 0$ where $\mathcal{L}_\xi$ is the Lie derivative along $\xi^a$. The Killing horizon $\mathcal{H}$ is a Killing horizon for the effective metric, iff

$$\mathrm{I}: f^{ab}\xi_a\xi_b \overset{\mathcal{H}}{=} 0, \qquad (9)$$

$$\mathrm{II}: f^{ab}\xi_a \overset{\mathcal{H}}{=} \aleph \xi^b, \qquad (10)$$

---

[*] Presumably this statement can be proven using the properties of the electromagnetic field, but we have not attempted that as it is covered by the more general proof given below.

where $\stackrel{\mathcal{H}}{=}$ means equality holds on $\mathcal{H}$ and $\aleph$ is a scalar. We ignore $\aleph = 0$ which corresponds to a degenerate perturbation metric, signalling the instability of background scalar configuration. Condition I states that the Killing horizon $\mathcal{H}$ is a null surface w.r.t. the effective metric. Condition II requires that the Killing vector field be hypersruface orthogonal on $\mathcal{H}$ - a necessary condition for the hypersurface to be a Killing horizon. Indeed, multiplying both sides of Eq. (10) with $f_{ab}^{-1}$, we get

$$\xi_a \stackrel{\mathcal{H}}{=} \aleph f_{ab}^{-1}\xi^b. \tag{11}$$

$\xi_a$ is the normal to $\mathcal{H}$, thus this equation implies that $f_{ab}^{-1}\xi^b$ is normal to $\mathcal{H}$.

In Appendix B, we prove both conditions. For condition I we only need to assume that the background metric and scalar configuration satisfy stationarity. To prove condition II without having to resort to global considerations, we need to make the additional assumption that *the surface gravity of the horizon is constant*. Overall, we have proved that the spacetime metric horizon $\mathcal{H}$ is a Killing horizon for the perturbation metric. Hence scalar perturbations cannot escape the region that matter fields perceive as a black hole. Note that if the scalar perturbations are subluminal, it is conceivable that $f_{ab}^{-1}$ could have another horizon outside $\mathcal{H}$.

Clearly, condition II implies condition I. Multiplying both sides of the Eq. (11) with $\xi^a$ yields

$$f_{ab}^{-1}\xi^a\xi^b \stackrel{\mathcal{H}}{=} 0 \tag{12}$$

which means $\xi^a$ is null w.r.t. the perturbation metric on $\mathcal{H}$. We have chosen to consider it as separate condition to make better contact with the results of Ref. [28]. Condition I also identifies the characteristics for scalar perturbations, as is discussed in more detail in Appendix C. As shown there, if one applies the approach of Ref. [28] to the decoupling limit and considers perturbation around a stationary background configuration, one does indeed obtain condition I.

The characteristics can be thought of as maximal speed of propagation surfaces and the boundary of the future of a *given* subset of the manifold is a characteristic. Nonetheless, characteristics are clearly not horizons in general (any null surface in Minkowski space is a characteristic for the wave equation) and condition I does not imply condition II without further assumptions. So, proving the former condition is not sufficient to claim that the Killing horizon of $g_{ab}$ will also be a Killing horizon of $f_{ab}^{-1}$ given that the two metrics are distinct. One can argue around this technical obstruction. Condition I does show that scalar excitations can cross the Killing horizon of $g_{ab}$ in a single direction only, which depends on the orientation of the characteristic. Hence, assuming that both $g_{ab}$ and $f_{ab}^{-1}$ are asymptotically flat and identifying their asymptotic regions should suffice to claim that scalar perturbations are trapped by the Killing horizon of $g_{ab}$ and cannot reach infinity, as claimed in Ref. [28].

It is hard to imagine a physical solution that would not satisfy these assumptions and yet it is interesting that they are not sufficient to prove that $\mathcal{H}$ is a Killing horizon of $f_{ab}^{-1}$ without further assuming that the surface gravity is constant. Carter has proven without resorting to field equations that Killing horizons in axisymmetric spacetimes that satisfy "$t$–$\phi$ orthogonality" have constant surface gravity [32].[†] The is also trivially true in spherical symmetry. Indeed, in what follows we present a complementary proof that is valid in spherical symmetry and does not make use of this assumption.

It is also worth mentioning that, if one is willing to assume that the Killing vector is timelike w.r.t. $f_{ab}^{-1}$ in the exterior of $\mathcal{H}$, then there seems to be an alternative way to argue that condition II has to hold. The Killing flow cannot pierce $\mathcal{H}$, as the latter is a Killing horizon for $g_{ab}$. Hence, $\xi^a$ either has to be orthogonal to $\mathcal{H}$ and null w.r.t. $f_{ab}^{-1}$ as well, or is has to reside in $\mathcal{H}$ and be spacelike w.r.t. $f_{ab}^{-1}$. However, the latter case is excluded by continuity if $\xi^a$ is timelike immediately outside $\mathcal{H}$.

### Special case: Spherical symmetry

For a static spherically symmetric spacetime and scalar field configuration, we make use of the $\{t, r, \theta, \phi\}$ coordinates. In this coordinate system, both spacetime and effective metrics are diagonal and given by

$$g_{ab}dx^a dx^b = -H(r)dt^2 + \frac{dr^2}{F(r)} + r^2 d\Omega^2, \tag{13}$$

$$(f^{-1})_{ab}dx^a dx^b = -\tilde{H}(r)dt^2 + \frac{dr^2}{\tilde{F}(r)} + \beta(r)^2 r^2 d\Omega^2, \tag{14}$$

where $d\Omega^2 = d\theta^2 + \sin\theta^2 d\phi^2$. The Killing vector is given by $\xi^a = (1, 0, 0, 0)$ and the Killing horizon of the effective metric is where $f_{tt}^{-1} = -\tilde{H}(r) = 0$, or equivalently $f^{tt} = \infty$.

The Killing horizon of the effective metric cannot be located away from the metric Killing horizon at $r = r_H$ where $H(r_H) = F(r_H) = 0$. The argument goes as follows: away from the metric Killing horizon, all components of the spacetime metric and its inverse, in $\{t, r, \theta, \phi\}$ coordinates, are finite. This means that all components of the Riemann tensor and field derivatives (with upper or lower indices) are finite. $f^{tt}$ is given by different combinations of the Riemann tensor and field derivatives contractions (see Appendix A), and thus it is finite and cannot satisfy the condition for the effective metric Killing horizon $f^{tt} = \infty$.

---

[†] By "$t$–$\phi$ orthogonality" one refers to the requirement that the $t$–$\phi$ plane be orthogonal to a family of 2-dimensional surfaces, which is a prerequisite for the line element to have $g_{t\phi}$ as the only off-diagonal component.

## III. TIME-DEPENDENT SCALARS AND BLACK HOLES WITH MULTIPLE HORIZONS

The theorem in the previous section ruled out the possibility of having multiple horizons for most stationary configurations in a large class of shift-symmetric Horndeski theories. However, as mentioned in the Introduction, static, spherically symmetric metric with time-dependent scalar configurations are known to exist in shift-symmetric Horndeski theories [22–24]. In particular, if $\Phi$ depends linearly on Killing time $t$, *i.e.* $\Phi = qt + \psi(r)$ in a suitable coordinate system and $q$ is a constant, then its derivatives respect staticity. Shift symmetry implies that $\Phi$ is present in the equations only through its derivatives, and hence the mismatch between the symmetries of the metric and the scalar does not necessarily lead to an inconsistency. This motivates the study of causality and the potential existence of multiple horizons in solutions with stationary metrics and time-dependent scalars. We will not attempt an exhaustive analysis here. Instead, we will just provide a simple example of a solution in which the Killing horizons of $g_{ab}$ and $f_{ab}^{-1}$ do not match, as evidence that this issue deserves further investigation.

One might actually expect the solutions of Refs. [22, 23] to readily provide such an example. Unfortunately, this is not the case within the confines of the zero-backreaction limit that we are employing here. This becomes apparent when one briefly examines the hairy black hole solutions presented in [22]. The action used there is a combination of $\mathcal{L}_2$ and $\mathcal{L}_4$

$$\mathcal{L} = \zeta R - 2\eta X + \beta G^{ab}\nabla_a\Phi\nabla_b\Phi - 2\Lambda\,, \qquad (15)$$

which corresponds to $G_2 = 2\eta X - 2\Lambda$ and $G_4 = \zeta + \beta X$. Shift symmetry implies that the equation for the scalar field can be written as a current conservation equation

$$\nabla_a J^a = 0\,, \qquad (16)$$

where

$$J^a = (\eta g^{ab} - \beta G^{ab})\partial_b\Phi\,. \qquad (17)$$

The field equation for $\Phi$ is linear in $\Phi$. Hence, one considers scalar perturbations in a *fixed* hairy background, consistent with our assumptions in the previous sections, $f^{ab} = 2\eta g^{ab} - 2\beta G^{ab}$. If that background is a solution of vacuum Einstein's equation with or without a cosmological constant, the $f_{ab}$ and $g_{ab}$ are conformally related and share their horizons.

Let us instead assume that $g_{ab}$ is one of the static, spherically symmetric metrics of the hairy solution reported in Ref. [22]. The following ansätze were used

$$ds^2 = -h(r)dt^2 + \frac{dr^2}{f(r)} + r^2 d\Omega^2\,, \qquad (18)$$

$$\Phi(t,r) = qt + \psi(r)\,, \qquad (19)$$

and all of the solutions found turn out to satisfy

$$\eta g^{rr} - \beta G^{rr} = 0 \qquad (20)$$

everywhere in the spacetime. However, Eq. (20) implies that $f^{rr} = 0$ everywhere and, hence, the effective metric for linear perturbations is degenerate throughout spacetime when backreaction is neglected. Clearly, in such solutions one cannot discuss in a meaningful way Killing horizons of $f^{ab}$ in the zero-backreaction limit. The degeneracy of $f^{ab}$ is clearly related to the choice of theory and the expression for the effective metric. Nonetheless, it does not actually imply that there is some general pathology of the solutions or the theory, as it appears only when one ignores the metric perturbations but still uses a background with a nontrivial scalar (*i.e.* when backreaction is neglected selectively at first order only.) See Refs. [26, 27, 33] for discussions on the linear stability of these solutions.

In order to provide an example of a solution with multiple horizons without taking into account backreaction, we turn to a different Lagrangian, known as k-essence,

$$\mathcal{L} = G_2[X]\,, \qquad (21)$$

where $G_{2X} > 0$. Varying this action with respect to $\Phi$, we arrive at the equation of motion

$$\left(G_{2X}g^{ab} - G_{2XX}\nabla^a\Phi\nabla^b\Phi\right)\nabla_a\nabla_b\Phi = 0\,. \qquad (22)$$

The causal structure of linear perturbations $\pi$ on top of a background field $\varphi$ is governed by

$$f^{ab} = G_{2X}g^{ab} - G_{2XX}\nabla^a\varphi\nabla^b\varphi\,. \qquad (23)$$

Inverting (23) yields

$$\left(f^{-1}\right)_{ab} = \frac{1}{G_{2X}}\left(g_{ab} + \frac{G_{2XX}}{G_{2X} + 2XG_{2XX}}\nabla_a\varphi\nabla_b\varphi\right) \qquad (24)$$

which determines how scalar perturbations propagate through the spacetime.

Let us assume that $g_{ab}$ describes a static spherically symmetric black hole spacetime with time-translational Killing vector $\xi^a$. The Killing horizon is located where $\xi^a$ becomes null according to the spacetime metric, i.e.

$$g_{ab}\xi^a\xi^b = 0\,. \qquad (25)$$

Scalar perturbations can evade the metric horizon provided that $\xi^a$ remains timelike with respect to the effective metric $(f^{-1})_{ab}$ on the Killing horizon. This happens when

$$\xi^a\nabla_a\varphi \neq 0 \qquad (26)$$

$$\frac{G_{2XX}}{G_{2X} + 2XG_{2XX}} < 0 \qquad (27)$$

on the metric horizon. The first condition expresses that the scalar field is time dependent and the second one

means that the perturbations are superluminal. It is important to note that the above conditions are local, i.e. they need to be satisfied only in a neighborhood of the Killing horizon.

We now proceed to construct an explicit solution for the scalar field in a *Schwarzschild background* with mass $M$, and demonstrate that the horizon $f_{ab}^{-1}$ does not coincide with that usual horizon. To do so, we make a specific choice of k-essence theory (21), with

$$G_2 = X + \frac{1}{2}\alpha X^2 , \qquad (28)$$

where $\alpha > 0$ ($\alpha < 0$) corresponds to subluminal (superluminal) propagation. According to Eq. (26), time dependence in $\varphi$ is necessary to have multiple horizons. We choose the following ansatz

$$\varphi(v,r) = qv + \psi(r), \qquad (29)$$

where $v = t + r_*$ is the ingoing Eddington-Finkelstein coordinate, and $r_*$ is the tortoise coordinate defined as $dr_* = \frac{dr}{1-2M/r}$. We further assume that our solution is continuously connected to $\varphi =$ constant as $q \to 0$, and we expand $\psi(r)$ as follows,

$$\psi(r) = \sum_{n=1}^{\infty} \psi_n q^n . \qquad (30)$$

We then solve the equations perturbatively in $q$. At first order in $q$, Eq. (22) reads

$$\frac{2(M-r)\psi_1' - r[2 + (r - 2M)\psi_1'']}{r^2} = 0, \qquad (31)$$

whose solution, after imposing regularity on the horizon, is

$$\psi_1'(r) = -\left(1 + \frac{2M}{r}\right). \qquad (32)$$

With this solution and ignoring the conformal factor $G_{2X}$ which is irrelevant for the causal structure, the effective metric (24) yields

$$\left(f^{-1}\right)_{ab} dx^a dx^b \propto \left(\alpha q^2 - 1 + \frac{2M}{r}\right) dv^2$$
$$+ 2\left[1 - \alpha q^2\left(1 + \frac{2M}{r}\right)\right] dr dv$$
$$+ \alpha q^2\left(1 + \frac{2M}{r}\right)^2 dr^2 + r^2 d\Omega^2. \quad (33)$$

We want to express metric (33) in Eddington-Finkelstein-like coordinates, thus we perform the following coordinate transformation

$$dv = dv'\left(1 + \frac{1}{2}\alpha q^2\right) - dr\frac{\alpha q^2}{2}\left(1 + \frac{2M}{r}\right)^2. \quad (34)$$

Defining $M' = M\left(1 + \alpha q^2\right)$, up to second order in $q$ the metric will read

$$\left(f^{-1}\right)_{ab} dx^a dx^b \propto - \left(1 - \frac{2M'}{r}\right) dv'^2$$
$$+ 2\left[1 - \frac{\alpha q^2 M'(4M'^2 + 2M'r + r^2)}{r^3}\right] dr dv' + r^2 d\Omega^2. \quad (35)$$

Starting from $r > 2M' = 2M(1+\alpha q^2)$, outgoing radial null rays of the effective metric can escape to infinity. For $\alpha < 0$, corresponding to the superluminal case, this region extends to both sides of the Killing horizon of the background metric $g_{ab}$.

Our solution clearly provides an example of an effective metric whose horizon does not coincide with that of the background metric, albeit is a much simplified setup where there is no backreaction and where the background is that of GR. It can be seen as having been obtained at the decoupling limit, where there is no backreaction at all orders in perturbation theory. The potential caveats are: (i) once backreaction is taken into account, the causal structure of the effective metric can change; (ii) it is not guaranteed that beyond decoupling, a static metric with a time-dependent scalar actually exists. On the other hand, one can argue that if the scalar field configuration evolves very slowly in time compared to the characteristic timescale defined by the mass of the black hole, then treating the black hole as quasistationary might be a good approximation that justifies our treatment. In any case, our sole purpose here is to demonstrate that the possibility of having multiple horizons deserves further investigation when it comes to time-dependent scalars.

## IV. DISCUSSION

The causal structure of black holes in Horndeski theories can deviate from the causal structure dictated by the spacetime metric. Due to derivative couplings of the scalar field and metric, the perturbation metrics are not conformal to the spacetime metric. Here we have focused for simplicity on the causal structure as seen by the linear perturbations of the scalar field on a *fixed* background metric.

The main result of this paper is the following: although null cones of the linear perturbation metric ($f_{ab}^{-1}$) and the spacetime metric ($g_{ab}$) are generally different, we have shown that for stationary black holes and scalar fields, and provided that the surface gravity of the horizon is constant, a Killing horizon of $g_{ab}$ $\mathcal{H}$ is also a Killing horizon for $f_{ab}^{-1}$. This means that the black hole region as perceived by matter fields, which are minimally coupled to $g_{ab}$, is a subset of or the same as the one defined by linear perturbations of the scalar field.

Our results agree with previous results in the literature that used the method of characteristics [28] when there is overlap, but they go a bit further to prove that

$\mathcal{H}$ is actually a Killing horizon for $f_{ab}^{-1}$. They also help interpret previous result in the intuitive physical picture of effective metrics and pin down certain subtleties of the characteristics approach. Remarkably, to rigorously complete the proof that $\mathcal{H}$ is a Killing horizon for $f_{ab}^{-1}$ for a generic stationary configuration one needs the local assumption of constant surface gravity. Alternatively, one can argue that $\mathcal{H}$ is a causal boundary of all excitations simply by being a characteristic, and hence is expected to be a Killing horizon for $f_{ab}^{-1}$. This correspondence of assumptions begs the conjecture that Killing horizons have constant surface gravity for stationary, asymptotically flat black hole configurations in Horndeski theories. Though there is no known proof of that statement (or a theory-independent proof), it is worth mentioning the following. Spherical black holes always possess this property and Carter has proven that axisymmetric black holes will also have this property if they satisfy "$t$–$\phi$ orthogonality" [32]. The latter is a necessary condition for having a line element with $g_{t\phi}$ as the only off-diagonal component (the usual ansatz for stationary, axisymmetric solutions). Finally, the zeroth law of black hole thermodynamics is related to the fact that surface gravity is constant. Hence, finding black hole solutions that do not satisfy this property would indeed be remarkable.

It is important to clarify the complementarity between our proof and no-hair theorems. In our general proof we do not impose that the background configuration satisfies any specific field equations and, hence, we did not need to make any reference to no-hair theorems as it applies equally well to hairy solutions. As has been discussed above, for backgrounds with a trivial scalar configuration the proof becomes rather trivial. This might give the false impression that the physical significance of our result is limited. To clarify this point, we would like to emphasize that stationary hairy solutions do exist in shift-symmetric Horndeski theories and that our proof does not rely on most of the assumptions employed in no-hair theorems. In particular, we bring the following points to the reader's attention:

(i) Our proof uses only local properties of Killing horizons and makes no reference to the asymptotics. Hence, it applies to black holes with nonflat asymptotics or matter in their vicinity, which are not covered by no-hair theorems.

(ii) The known no-hair theorems for shift-symmetric Horndeski theories require the assumptions of staticity and spherical symmetry [13] or slow rotation [14]. We only assume stationarity.

(iii) Interactions between the scalar and other fields, including the metric, could provide a nontrivial configuration for the scalar field. However, as long as the new interaction terms do not spoil the structure of the perturbation metric our proof holds. That is, the interaction term only needs to respect shift symmetry upon linearization.

(iv) Even with the shift-symmetric Horndeski class, there exists a term that inevitably gives rise to scalar hair [14, 15], $\Phi\mathcal{G}$, where $\mathcal{G} = R_{abcd}R^{abcd} - 4R_{ab}R^{ab} + R^2$ is the Gauss-Bonnet invariant.

To elaborate on this last point, consider a theory that belongs to the class of theories given by the Lagrangian (1) and assume $\varphi = const$ is a solution to that theory. Then add $\alpha\Phi\mathcal{G}$, where $\alpha$ is a coupling constant, to the Lagrangian. The resulting theory is still in the class of shift-symmetric Horndeski theories, since $\mathcal{G}$ is a total derivative term in 4 dimensions. The new term in the Lagrangian adds $\alpha\mathcal{G}$ to the scalar field's equation of motion and $\mathcal{G}$ does not vanish in black hole spacetimes. Consequently, the new theory must admit a hairy solution only. However, the new term in the scalar field equation of motion vanishes from the equation of motion for scalar linear perturbations on a *fixed* background. Hence, it does not change the structure of the effective metric and our proof applies to this class of theories as well.

Clearly, our central simplifying assumption to consider scalar perturbations on fixed metric backgrounds is a very strong limitation. It is pertinent to revisit the problem taking metric perturbations into account. The perturbative analysis presented in Ref. [27], though focused on the role of causality for stability rather than exploring the horizon structure, provides some useful background for extending our analysis.

We also discussed briefly the possibility of having different horizons in configurations where the scalar does not respect stationarity and we provided a simple example in the decoupling limit, in order to motivate further work. Another crucial piece of motivation to look further into this issue is the following. As has been pointed out in the context of Lorentz-violating theories, where solutions with nested horizons for different excitations are common [34–36], the region between two such horizons can be seen as an ergoregion. In that region, the slowest of the two excitations can carry negative Killing energy as it is behind its Killing horizon and at the same time the faster excitation can still escape. If the two modes interact, then there is a possibility of energy extraction [37–39]. Assuming that such a process can take place [40], it could lead to thermodynamics conundrums [37–39], or perhaps just provide a natural decay channel from hairy to nonhairy solutions [40].

## V. ACKNOWLEDGEMENTS

[1] W. Israel, Phys. Rev. **164**, 1776 (1967).

[2] B. Carter, Phys. Rev. Lett. **26**, 331 (1971).

[3] D. C. Robinson, Phys. Rev. Lett. **34**, 905 (1975).

[4] T. P. Sotiriou, *Proceedings of the 7th Aegean Summer School : Beyond Einstein's theory of gravity. Modifications of Einstein's Theory of Gravity at Large Distances.: Paros, Greece, September 23-28, 2013*, Lect. Notes Phys. **892**, 3 (2015), arXiv:1404.2955 [gr-qc].

[5] T. P. Sotiriou, Class. Quant. Grav. **32**, 214002 (2015), arXiv:1505.00248 [gr-qc].

[6] G. W. Horndeski, Int. J. Theor. Phys. **10**, 363 (1974).

[7] C. Deffayet, S. Deser, and G. Esposito-Farèse, Phys. Rev. **D80**, 064015 (2009), arXiv:0906.1967 [gr-qc].

[8] T. Kobayashi, M. Yamaguchi, and J. Yokoyama, Prog. Theor. Phys. **126**, 511 (2011), arXiv:1105.5723 [hep-th].

[9] J. E. Chase, Communications in Mathematical Physics **19**, 276 (1970).

[10] S. W. Hawking, Commun. Math. Phys. **25**, 167 (1972).

[11] J. D. Bekenstein, Phys. Rev. **D51**, R6608 (1995).

[12] T. P. Sotiriou and V. Faraoni, Phys. Rev. Lett. **108**, 081103 (2012), arXiv:1109.6324 [gr-qc].

[13] L. Hui and A. Nicolis, Phys. Rev. Lett. **110**, 241104 (2013), arXiv:1202.1296 [hep-th].

[14] T. P. Sotiriou and S.-Y. Zhou, Phys. Rev. Lett. **112**, 251102 (2014), arXiv:1312.3622 [gr-qc].

[15] T. P. Sotiriou and S.-Y. Zhou, Phys. Rev. **D90**, 124063 (2014), arXiv:1408.1698 [gr-qc].

[16] P. Kanti, N. E. Mavromatos, J. Rizos, K. Tamvakis, and E. Winstanley, Phys. Rev. **D54**, 5049 (1996), arXiv:hep-th/9511071 [hep-th].

[17] N. Yunes and L. C. Stein, Phys. Rev. **D83**, 104002 (2011), arXiv:1101.2921 [gr-qc].

[18] H. O. Silva, J. Sakstein, L. Gualtieri, T. P. Sotiriou, and E. Berti, Phys. Rev. Lett. **120**, 131104 (2018), arXiv:1711.02080 [gr-qc].

[19] D. D. Doneva and S. S. Yazadjiev, Phys. Rev. Lett. **120**, 131103 (2018), arXiv:1711.01187 [gr-qc].

[20] G. Antoniou, A. Bakopoulos, and P. Kanti, Phys. Rev. Lett. **120**, 131102 (2018), arXiv:1711.03390 [hep-th].

[21] C. A. R. Herdeiro and E. Radu, Phys. Rev. Lett. **112**, 221101 (2014), arXiv:1403.2757 [gr-qc].

[22] E. Babichev and C. Charmousis, JHEP **08**, 106 (2014), arXiv:1312.3204 [gr-qc].

[23] C. Charmousis, T. Kolyvaris, E. Papantonopoulos, and M. Tsoukalas, JHEP **07**, 085 (2014), arXiv:1404.1024 [gr-qc].

[24] T. Kobayashi and N. Tanahashi, PTEP **2014**, 073E02 (2014), arXiv:1403.4364 [gr-qc].

[25] E. Babichev, V. Mukhanov, and A. Vikman, JHEP **02**, 101 (2008), arXiv:0708.0561 [hep-th].

[26] E. Babichev, C. Charmousis, G. Esposito-Farèse, and A. Lehébel, (2017), arXiv:1712.04398 [gr-qc].

[27] E. Babichev, C. Charmousis, G. Esposito-Farèse, and A. Lehébel, (2018), arXiv:1803.11444 [gr-qc].

[28] N. Tanahashi and S. Ohashi, Class. Quant. Grav. **34**, 215003 (2017), arXiv:1704.02757 [hep-th].

[29] M. Minamitsuji, Phys. Lett. **B743**, 272 (2015).

[30] H. Reall, N. Tanahashi, and B. Way, Class. Quant. Grav. **31**, 205005 (2014), arXiv:1406.3379 [hep-th].

[31] K. Izumi, Phys. Rev. **D90**, 044037 (2014), arXiv:1406.0677 [gr-qc].

[32] B. Carter, in *Proceedings, Ecole d'Eté de Physique Théorique: Les Astres Occlus: Les Houches, France, August, 1972* (1973) pp. 57–214.

[33] H. Ogawa, T. Kobayashi, and T. Suyama, Phys. Rev. **D93**, 064078 (2016), arXiv:1510.07400 [gr-qc].

[34] C. Eling and T. Jacobson, Class. Quant. Grav. **23**, 5643 (2006), [Erratum: Class. Quant. Grav. **27**, 049802 (2010)], arXiv:gr-qc/0604088 [gr-qc].

[35] E. Barausse, T. Jacobson, and T. P. Sotiriou, Phys. Rev. D **83**, 124043 (2011), arXiv:1104.2889 [gr-qc].

[36] E. Barausse and T. P. Sotiriou, Class. Quant. Grav. **30**, 244010 (2013), arXiv:1307.3359 [gr-qc].

[37] S. L. Dubovsky and S. M. Sibiryakov, Phys. Lett. **B638**, 509 (2006), arXiv:hep-th/0603158 [hep-th].

[38] C. Eling, B. Z. Foster, T. Jacobson, and A. C. Wall, Phys. Rev. D **75**, 101502 (2007), arXiv:hep-th/0702124 [hep-th].

[39] T. Jacobson and A. C. Wall, Found. Phys. **40**, 1076 (2010), arXiv:0804.2720 [hep-th].

[40] R. Benkel, J. Bhattacharyya, J. Louko, D. Mattingly, and T. P. Sotiriou, (2018), arXiv:1803.01624 [gr-qc].

[41] E. Gourgoulhon and J. L. Jaramillo, Phys. Rept. **423**, 159 (2006), arXiv:gr-qc/0503113 [gr-qc].

## Appendix A: Effective metric in shift-symmetric Horndeski

In this section we show the exact form of the effective metric defined in Eq. (7). We split

$$f^{ab} = \sum_{n=2}^{5} f^{ab}_{(n)}, \qquad (A1)$$

where the subscript index $n$ refers to the contribution coming from $\mathcal{L}_n$, and use the simplifying notation $\nabla_a \varphi = \varphi_a$ and $\nabla_b \nabla_a \varphi = \varphi_{ab}$. We have explicitly

$$f_{(2)}^{ab} = G_{2X} g^{ab} - G_{2XX} \varphi^a \varphi^b \,, \tag{A2}$$

$$f_{(3)}^{ab} = 2G_{3X} \left( \varphi^{ab} - \Box\varphi g^{ab} \right) + G_{3XX} \left( \Box\varphi \varphi^a \varphi^b + \varphi_c \varphi_d \varphi^{cd} g^{ab} - \varphi^{ca} \varphi^b \varphi_c - \varphi^{cb} \varphi^a \varphi_c \right) \,, \tag{A3}$$

$$f_{(4)}^{ab} = G_{4X} f_{(4,1)}^{ab} + G_{4XX} f_{(4,2)}^{ab} + G_{4XXX} f_{(4,3)}^{ab} \,, \tag{A4}$$

$$f_{(5)}^{ab} = G_{5X} f_{(5,1)}^{ab} + G_{5XX} f_{(5,2)}^{ab} + G_{5XXX} f_{(5,3)}^{ab} \,, \tag{A5}$$

where

$$f_{(4,1)}^{ab} = -2G^{ab} \,, \tag{A6}$$

$$f_{(4,2)}^{ab} = -R\varphi^a \varphi^b + 3(\Box\varphi)^2 g^{ab} - 3(\varphi_{cd})^2 g^{ab} - 6\Box\varphi\varphi^{ab} - 2R_{cd}\varphi^c\varphi^d g^{ab} + 2R^{acbd}\varphi_c\varphi_d + 4R^{ac}\varphi_c\varphi^b + 6\varphi^{ac}\varphi^b{}_c \,, \tag{A7}$$

$$f_{(4,3)}^{ab} = -2\Box\varphi\varphi_{cd}\varphi^c\varphi^d g^{ab} - (\Box\varphi)^2\varphi^a\varphi^b + 2\varphi_{cd}\varphi^c\varphi^d\varphi^{ab} + (\varphi_{cd})^2\varphi^a\varphi^b + \tag{A8}$$
$$+ 2\varphi^{cd}\varphi_{ed}\varphi_c\varphi^e g^{ab} + 4\Box\varphi\varphi^{ac}\varphi_c\varphi^b - 2\varphi^{ac}\varphi^{bd}\varphi_c\varphi_d - 4\varphi_{cd}\varphi^b\varphi^d\varphi^{ac} \,,$$

$$f_{(5,1)}^{ab} = \Box\varphi G^{ab} + G_{cd}\varphi^{cd} g^{ab} - 2G^{ac}\varphi_c{}^b + \Box\varphi R^{ab} + R_{cd}\varphi^{cd} g^{ab} - 2\varphi^a{}_c R^{bc} + 2\varphi_{cd} R^{acdb} \,, \tag{A9}$$

$$f_{(5,2)}^{ab} = -G^{ab}\varphi_c\varphi_d\varphi^{cd} - G_{cd}\varphi^{cd}\varphi^a\varphi^b + 2G^{ac}\varphi^b\varphi_{cd}\varphi^d + \Box\varphi \left( (\varphi_{cd})^2 + \varphi^c\varphi^d R_{cd} \right) g^{ab} + (\Box\varphi)^2\varphi^{ab} + \tag{A10}$$
$$- 2\varphi_{cd}\varphi^d R^{ce}\varphi_e g^{ab} - 2\Box\varphi R^{ac}\varphi_c\varphi^b - \varphi^{ab} \left( 2(\varphi_{cd})^2 + \varphi^c\varphi^d R_{cd} \right) - 2\varphi_{cd}\varphi^d\varphi_e R^{beca} - 2\varphi_{cd}\varphi^b\varphi_e R^{deac} + $$
$$- g^{ab}\varphi^{cd} \left( \varphi_{ce}\varphi^e{}_d - \varphi^e\varphi^f R_{cefd} \right) - \Box\varphi (3\varphi^{ac}\varphi^b{}_c - \varphi_c\varphi_d R^{acdb}) + 2\varphi^{ac}\varphi^b R_{cd}\varphi^d + 2R^{ac}\varphi_c\varphi^{bd}\varphi_d + $$
$$+ 2\varphi^{cb} (2\varphi^{ad}\varphi_{cd} + \varphi^d\varphi_e R^{ae}{}_{cd}) - \frac{1}{6}g^{ab} \left( 4(\Box\varphi)^3 - 6\Box\varphi(\varphi_{cd})^2 + 2(\varphi_{cd})^3 \right) + $$
$$+ (\Box\varphi)^2\varphi^{ab} - \Box\varphi\varphi^{ac}\varphi^b{}_c \,,$$

$$f_{(5,3)}^{ab} = -(\Box\varphi)^2\varphi^{ac}\varphi_c\varphi^b - \Box\varphi\varphi_{cd}\varphi^d\varphi^{ce}\varphi_e g^{ab} + \varphi_{cd}\varphi^c\varphi^{de}\varphi_e\varphi^{ab} + (\varphi_{de})^2\varphi^{ac}\varphi_c\varphi^b + \varphi^{cd}\varphi_{ce}\varphi^e\varphi_{df}\varphi^f g^{ab} + \tag{A11}$$
$$+ \Box\varphi\varphi^{ac}\varphi_c\varphi^{bd}\varphi_d + 2\Box\varphi\varphi^{ac}\varphi^b\varphi_{cd}\varphi^d - 2\varphi^{ad}\varphi_d\varphi^{bc}\varphi_{ce}\varphi^e - 2\varphi^{ad}\varphi_{dc}\varphi^b\varphi^{ce}\varphi_e + $$
$$+ \frac{1}{6} \left( (\Box\varphi)^3 - 3\Box\varphi(\varphi_{cd})^2 + 2(\varphi_{cd})^3 \right) \varphi^a\varphi^b + \frac{1}{2}(\Box\varphi)^2\varphi_c\varphi_d\varphi^{cd} g^{ab} + $$
$$- \frac{1}{2}\varphi_c\varphi_d\varphi^{cd}(\varphi_{ef})^2 g^{ab} - \Box\varphi\varphi_c\varphi_d\varphi^{cd}\varphi^{ab} + \varphi_c\varphi_d\varphi^{cd}\varphi^{ae}\varphi^b{}_e \,.$$

## Appendix B: Killing horizon of $f_{ab}^{-1}$

Here, we prove Eqs. (9) and (10). For the proof of Eq. (9) we only use stationarity of the metric and the scalar field. In the proof of Eq. (10), we also assume that the Killing horizon has constant surface gravity. We use the notation introduced in Appendix A for simplicity, namely $\varphi_a = \nabla_a \varphi$ and $\varphi_{ab} = \nabla_b \nabla_a \varphi$.

### 1. Proof of condition I

The proof of condition I, $f^{ab}\xi_a\xi_b \stackrel{\mathcal{H}}{=} 0$, is rather straightforward, but it involves going through each term of $f^{ab}$ in Appendix A to show it vanishes. For an interested reader, the following identities are helpful to reproduce the result:

$$g^{ab}\xi_a\xi_b \stackrel{\mathcal{H}}{=} 0$$
$$\xi^a\varphi_a = 0$$
$$\varphi^{ab}\xi_a \stackrel{\mathcal{H}}{\propto} \xi^b$$
$$R^{ab}\xi_a\xi_b \stackrel{\mathcal{H}}{=} 0$$
$$R_{acbd}\xi^a\eta^c\xi^b\hat{\eta}^d \stackrel{\mathcal{H}}{=} 0,$$

where $\eta$ and $\hat{\eta}$ are tangential spacelike directions on the horizon.

### 2. Proof of condition II

Let us introduce the following definition:

$$p^a \stackrel{\mathcal{H}}{\equiv} q^a \iff p^a \stackrel{\mathcal{H}}{=} q^a + c\,\xi^a, \tag{B1}$$

i.e. two (co)vectors are equivalent if they differ by a multiple of $\xi^a$.

Then, we can express Eq. (10) as

$$f^{ab}\xi_a \stackrel{\mathcal{H}}{\equiv} 0. \tag{B2}$$

Given the expression for $f^{ab}$ (Appendix A), the following three relations

$$\xi_a \nabla^b\nabla^a\varphi \stackrel{\mathcal{H}}{\equiv} 0, \tag{B3}$$

$$\xi_a R^{ab} \stackrel{\mathcal{H}}{\equiv} 0, \tag{B4}$$

$$\xi_a R^{cbda} T_{cd} \stackrel{\mathcal{H}}{\equiv} 0, \tag{B5}$$

where $T_{cd}$ is any symmetric tensor satisfying $T_{cd}\xi^d \stackrel{\mathcal{H}}{\equiv} 0$, are sufficient for Eq. (B2) to be satisfied.

In what follows, we prove the relations in Eqs. (B3-B5). In the proof of (B4) and (B5) we assume that the surface gravity of the horizon $\mathcal{H}$ is constant.

$$a. \quad \xi_a \varphi^{ab} \stackrel{\mathcal{H}}{\equiv} 0.$$

The scalar field satisfies stationarity, i.e. $\xi^a \varphi_a = 0$. Taking a derivative of this condition, we get

$$\xi_a \varphi^{ab} = -\varphi_a \nabla^b \xi^a. \tag{B6}$$

We can use the expansion of $\nabla^b \xi^a$ on the horizon

$$\nabla^b \xi^a \stackrel{\mathcal{H}}{=} -\kappa(\xi^a k^b - \xi^b k^a) + \xi^a \eta^b - \xi^b \eta^a, \tag{B7}$$

where $k^a$ is a null direction transverse to $\mathcal{H}$ satisfying $\xi^a k_a = -1$, $\eta^a$ is a spacelike direction orthogonal to $\xi^a$ and $k^a$ and $\kappa$ is the surface gravity of $\mathcal{H}$. Equation (B6) yields

$$\xi_a \varphi^{ab} \stackrel{\mathcal{H}}{=} -\kappa k_a \varphi^a \xi^b + \eta_a \varphi^a \xi^b \stackrel{\mathcal{H}}{\equiv} 0, \tag{B8}$$

$$b. \quad \xi_a R^{ab} \stackrel{\mathcal{H}}{\equiv} 0.$$

The Killing horizon $\mathcal{H}$ has zero expansion, shear and vorticity. Thus applying Raychaudhuri equation on $\mathcal{H}$ yields

$$R_{ab} \xi^a \xi^b \stackrel{\mathcal{H}}{=} 0. \tag{B9}$$

Moreover, for any tangent spacelike direction $\hat{\eta}^a$ on $\mathcal{H}$

$$\kappa_{,a} \hat{\eta}^a \stackrel{\mathcal{H}}{=} -R_{ab} \xi^a \hat{\eta}^b. \tag{B10}$$

Thus for a constant surface gravity of $\mathcal{H}$

$$R_{ab} \xi^a \hat{\eta}^b \stackrel{\mathcal{H}}{=} 0. \tag{B11}$$

Eq. (B9) implies that $R^{ab}\xi_a$ is tangent to $\mathcal{H}$ and eq. (B11) implies that this vector is orthogonal to all spacelike directions on $\mathcal{H}$. The only vector on $\mathcal{H}$ satisfying these properties has to be a multiple of $\xi^b$. Consequently,

$$R^{ab} \xi_a \stackrel{\mathcal{H}}{\equiv} 0, \tag{B12}$$

$$c. \quad \xi_a R^{cbda} T_{cd} \stackrel{\mathcal{H}}{\equiv} 0.$$

In order to prove Eq. (B5), we consider a family of null hypersurfaces in an open neighborhood around $\mathcal{H}$ with tangent null direction $l^a$ such that

$$l^a \stackrel{\mathcal{H}}{=} \xi^a. \tag{B13}$$

As a result,

$$T^{cd} R_{cbda} \xi^a \stackrel{\mathcal{H}}{=} T^{cd} R_{cbda} l^a = T^{cd} \left( \nabla_c \nabla_b l_d - \nabla_b \nabla_c l_d \right). \tag{B14}$$

By the properties of null hypersurfaces,

$$\nabla_b l_d = \theta_{bd} + \omega_b l_d - l_b k^c \nabla_c l_d \tag{B15}$$

where $k^a$ is the null direction transverse to the null hypersurfaces satisfying $l.k = -1$, $\theta_{ab}$ is the projection of $\nabla_b l_a$ onto 2-dimensional spacelike submanifold of the null hypersurfaces (orthogonal to $l^a$ and $k^a$) and $\omega_a = l^b \nabla_b k_a$ (Eq. (5.20) of [41]).

Taking the derivative of Eq. (B15) and substituting in Eq. (B14), we arrive at

$$\begin{aligned} T^{cd} R_{cbda} \xi^a \stackrel{\mathcal{H}}{=}\ & T^{cd} \left( \nabla_c \theta_{bd} - \nabla_b \theta_{cd} \right) \\ & + T^{cd} \left( \nabla_c \omega_b - \nabla_b \omega_c \right) l_d \\ & + T^{cd} \left( \omega_b \nabla_c l_d - \omega_c \nabla_b l_d \right) \\ & - T^{cd} \left( \nabla_c l_b - \nabla_b l_c \right) k^e \nabla_e l_d \\ & - T^{cd} \left( l_b \nabla_c - l_c \nabla_b \right) \left( k^e \nabla_e l_d \right). \end{aligned} \tag{B16}$$

We show that each line in the above equation gives a contribution proportional to $\xi_b$ on $\mathcal{H}$.

(i) Let us define $u_b^{(1)} = T^{cd} \left( \nabla_c \theta_{bd} - \nabla_b \theta_{cd} \right)$. We prove $u_b^{(1)} \stackrel{\mathcal{H}}{\equiv} 0$ by showing $u_b^{(1)} l^b \stackrel{\mathcal{H}}{\equiv} 0$ and $u_b^{(1)} \eta^b \stackrel{\mathcal{H}}{\equiv} 0$ where $\eta^b$ is any tangent spacelike direction on $\mathcal{H}$.

Note that $l^b \theta_{ab} = 0$ and $\theta_{ab} \stackrel{\mathcal{H}}{=} 0$. Thus

$$l^b \nabla_b \theta_{cd} \stackrel{\mathcal{H}}{=} 0, \tag{B17}$$

$$l^b \nabla_c \theta_{bd} = - \left( \nabla_c l^b \right) \theta_{bd} \stackrel{\mathcal{H}}{=} 0, \tag{B18}$$

which yields $u_b^{(1)} l^b \stackrel{\mathcal{H}}{=} 0$.

In order to prove $u_b^{(1)} \eta^b \stackrel{\mathcal{H}}{=} 0$, notice that

$$\eta^b \nabla_b \theta_{cd} \stackrel{\mathcal{H}}{=} 0. \tag{B19}$$

As a result,

$$u_b^{(1)} \eta^b \stackrel{\mathcal{H}}{=} T^{cd} \eta^b \nabla_c \theta_{bd}. \tag{B20}$$

From $T^{cd} l_c \stackrel{\mathcal{H}}{\equiv} 0$, it follows

$$T^{cd} l_c \stackrel{\mathcal{H}}{=} \mathcal{T} l^d, \tag{B21}$$

where $\mathcal{T}$ is a scalar. Thus, we can express $T^{cd}$ as

$$T^{cd} \stackrel{\mathcal{H}}{=} \hat{T}^{cd} - \mathcal{T} k^c l^d. \tag{B22}$$

where $\hat{T}^{cd} l_c \stackrel{\mathcal{H}}{=} 0$. Consequently,

$$T^{cd} \nabla_c \theta_{bd} \stackrel{\mathcal{H}}{=} \hat{T}^{cd} \nabla_c \theta_{bd} - \mathcal{T} k^c l^d \nabla_c \theta_{bd} \tag{B23}$$

The index $c$ of $\hat{T}^{cd}$ has to be along a tangent direction to $\mathcal{H}$ which implies $\hat{T}^{cd} \nabla_c \theta_{bd} \stackrel{\mathcal{H}}{=} 0$. Moreover, $l^d \nabla_c \theta_{bd} \stackrel{\mathcal{H}}{=} 0$ and this proves $u_b^{(1)} \eta^b \stackrel{\mathcal{H}}{=} 0$.

(ii) If we define $u_b^{(2)} = T^{cd} \left( \nabla_c \omega_b - \nabla_b \omega_{c;b} \right) l_d$, by Eq. (B21)

$$u_b^{(2)} \overset{\mathcal{H}}{=} \mathcal{T} l^c (\nabla_c \omega_b - \nabla_b \omega_c). \tag{B24}$$

Again, one can show

$$u_b^{(2)} l^b \overset{\mathcal{H}}{=} \mathcal{T} l^b l^c (\nabla_c \omega_b - \nabla_b \omega_c) = 0, \tag{B25}$$

$$u_b^{(2)} \eta^b \overset{\mathcal{H}}{=} \mathcal{T} \eta^b l^c (\nabla_c \omega_b - \nabla_b \omega_c) \overset{\mathcal{H}}{=} R_{ab} l^a \eta^b \overset{\mathcal{H}}{=} 0, \tag{B26}$$

where the second equality is a result of Damour-Navier-Stokes equation (Eq. (6.14) in [41]) on $\mathcal{H}$. Equations (B25) and (B26) imply

$$u_b^{(2)} \overset{\mathcal{H}}{\equiv} 0. \tag{B27}$$

(iii) Defining $u_b^{(3)} = T^{cd} (\omega_b \nabla_c l_d - \omega_c \nabla_b l_d)$ and substituting $\nabla_b l_d$ from Eq. (B15), after straightforward calculations one gets

$$u_b^{(3)} \overset{\mathcal{H}}{\equiv} 0. \tag{B28}$$

(iv) $u_b^{(4)} = T^{cd} (\nabla_c l_b - \nabla_b l_c) k^e \nabla_e l_d \overset{\mathcal{H}}{\equiv} 0$ by substituting $\nabla_c l_b - \nabla_b l_c$ from Eq. (B15) and carrying out the calculations.

(v) Lastly, we define $u_b^{(5)} = -T^{cd} (l_b \nabla_c - l_c \nabla_b)(k^e \nabla_e l_d)$. We get the desired result as follows

$$\begin{aligned} u_b^{(5)} &\overset{\mathcal{H}}{\equiv} l_c T^{cd} \nabla_b (k^e \nabla_e l_d) \\ &\overset{\mathcal{H}}{=} \mathcal{T} l^d \nabla_b (k^e \nabla_e l_d) \\ &= \mathcal{T} \nabla_b (l^d k^e \nabla_e l_d) - \mathcal{T} k^e (\nabla_e l_d)(\nabla_b l^d) \\ &\overset{\mathcal{H}}{=} -\mathcal{T} k^e \nabla_e l_d (\omega_b l^d - l_b k^c \nabla_c l^d) \\ &\overset{\mathcal{H}}{=} -\mathcal{T} l^d (\nabla_e l_d) \omega_b k^e = 0, \end{aligned} \tag{B29}$$

where we have used $l_a l^a = 0$ and $\nabla_b l_a l^a = 0$.

We have shown that all five terms on the right-hand side of Eq. (B16) are proportional to $\xi^a$, thus the proof is complete.

### Appendix C: Comparison with ref. [28]

Here, we investigate the relation between our result and Ref. [28]. In particular, how the principal symbol $P$ is related to the effective metric for linear perturbations $f^{ab}$. To recap, the principal symbol $P$ of surface $\Sigma$ is defined as

$$P = \frac{\partial E_I}{\partial v_{J,ab}} n_a n_b \tag{C1}$$

where $E_I$'s are the eoms, $v_J$'s are the degrees of freedom and $n_a$ is the normal vector to $\Sigma$. The key assumption in this definition is that the eoms are quasilinear when expressed in a foliation.

In Ref. [28] $P$ is calculated without resorting to any perturbative treatment. It is then shown that for stationary spacetime metrics and scalar fields $P$ becomes degenerate on the horizon $\mathcal{H}$ and this is interpreted to mean that the horizon is a characteristic surface. Taking the stationary limit is rather subtle as, when one imposes stationarity for the whole configuration the equations are no longer hyperbolic and it becomes meaningless to refer to characteristics. One should rather impose stationarity on a background configuration and consider the principal symbol for linear perturbations, which actually coincides with $P$ in a quasilinear theory. Alternatively, as done in Ref. [28], one could consider a potentially nonlinear perturbation that is localised to the interior of the horizon while taking the stationary limit of the exterior.

In this paper, we take the background metric to be fixed, thus the only degree of freedom is the scalar field. We are interested in the properties of $P$ on the Killing horizon, so we set $n_a = \xi_a$. In our setup, this amounts to

$$P = \frac{\partial E_\Phi}{\partial \Phi_{ab}} \xi_a \xi_b \overset{\mathcal{H}}{=} 0, \tag{C2}$$

where $E_\Phi$ is the scalar field eom. We use the notation introduced in Appendix A for simplicity, namely $\varphi_a = \nabla_a \varphi$ and $\varphi_{ab} = \nabla_b \nabla_a \varphi$.

However, $\frac{\partial E_\Phi}{\partial \Phi_{ab}}$ is nothing but the effective metric $f^{ab}$. By definition, $f^{ab}$ is the coefficient of the second order field derivative in perturbation around a background configuration. Taking $E_\Phi = E_\Phi(\Phi_a, \Phi_{ab})$, perturbation $\pi$ around a background configuration $\varphi$ results into

$$E_\Phi (\varphi_a + \pi_a, \varphi_{ab} + \pi_{ab}) = E_\Phi (\varphi_a, \varphi_{ab}) + \frac{\partial E_\Phi}{\partial \Phi_a} \pi_a + \frac{\partial E_\Phi}{\partial \Phi_{ab}} \pi_{ab}, \tag{C3}$$

which gives

$$f^{ab} = \frac{\partial E_\Phi}{\partial \Phi_{ab}}. \tag{C4}$$

Consequently, the result of Ref. [28] can be translated to our notation as

$$f^{ab} \xi_a \xi_b \overset{\mathcal{H}}{=} 0 \tag{C5}$$

which is Eq. (9).