Region with trapped surfaces in spherical symmetry, its core, and their boundaries

Ingemar Bengtsson[1,*] and José M. M. Senovilla[2,†]

[1]*Stockholms Universitet, AlbaNova, Fysikum, S-106 91 Stockholm, Sweden*
[2]*Física Teórica, Universidad del País Vasco, Apartado 644, 48080 Bilbao, Spain*

We consider the region $\mathcal{T}$ in spacetime containing future-trapped closed surfaces and its boundary $\mathcal{B}$, and derive some of their general properties. We then concentrate on the case of spherical symmetry, but the methods we use are general and applicable to other situations. We argue that closed trapped surfaces have a nonlocal property, "clairvoyance", which is inherited by $\mathcal{B}$. We prove that $\mathcal{B}$ is not a marginally trapped tube in general, and that it can have portions in regions whose whole past is flat. For asymptotically flat black holes, we identify a general past barrier, well inside the event horizon, to the location of $\mathcal{B}$ under physically reasonable conditions. We also define the core $\mathcal{Z}$ of the trapped region as that part of $\mathcal{T}$ which is indispensable to sustain closed trapped surfaces. We prove that the unique spherically symmetric dynamical horizon is the boundary of such a core, and we argue that this may serve to single it out. To illustrate the results, some explicit examples are discussed, namely, Robertson-Walker geometries and the imploding Vaidya spacetime.

I. INTRODUCTION

If a black hole is a thing, does it have a boundary? If so, where is it? For stationary black holes, the event horizon seems to be the obvious answer [1], but for evolving black holes the situation is less clear. Obviously, all black holes must be formed (a dynamical process) and then undergo accretion and other evolutionary processes. Furthermore, as a matter of principle isolated black holes always evolve through Hawking radiation, and it has been argued that—strictly speaking—the event horizon may not even exist, due to quantum gravity effects close to the singularity [2]. In any case it has been argued that the event horizon is "unreasonably global" [3]. Thus there are exact solutions—such as the imploding Vaidya spacetime [4]—where an observer can cross the event horizon even though her entire causal past is a piece of flat Minkowski space [5]. In numerical relativity, this problem is posed sharply because one wants to identify the boundary of a black hole in an initial data set, not by inspection of the full solution.

It is then natural to turn to closed trapped surfaces, the hallmark of gravitational collapse [6]. Efficient algorithms to identify the region where they occur within a given initial data set do exist [7]. The boundary of such a spatial region is called the apparent horizon [8], and it is in itself a marginally outer trapped surface [9,10]. It "evolves" to form hypersurfaces foliated by such surfaces [11], which we refer to as apparent 3-horizons. In general, hypersurfaces foliated by marginally trapped surfaces are called marginally trapped tubes [3,12,13]. There are still problems, because the location of the horizons obtained in this way depends strongly on an arbitrary slicing into space and

time. In general there will be many marginally trapped tubes in a given black-hole spacetime, and they weave through each other in a complicated way [3].

Another possibility is to locate the boundary of the region through which closed trapped surfaces pass, or the boundary of the region through which outer trapped surfaces pass. For closed trapped surfaces, both null expansions are negative, while for outer trapped surfaces no condition is imposed on the inner expansion. It is important to make this distinction, because the two regions do not coincide in general. It was conjectured by Eardley [14] that the region defined by outer trapping coincides with the event horizon. Ben-Dov [15] proved this for the Vaidya spacetime, but he also proved that the region containing trapped surfaces with both expansions negative lies strictly inside it. Nevertheless it can still happen that an observer can cross a closed trapped surface even though her entire causal past is a piece of flat Minkowski spacetime, as we showed in an earlier paper [16]. This proves that closed trapped surfaces have highly nonlocal properties too, they are *clairvoyant*: they are "aware" of things that happen elsewhere, very far away with no causal connection.

We focus on trapped surfaces with both expansions negative, partly because marginally trapped surfaces form marginally trapped tubes—such as future outer trapping horizons [17,18] and dynamical horizons [5]—having particularly interesting properties with regard to energy fluxes and the like, and partly simply for definiteness. We concentrate on spherically symmetric imploding spacetimes. We originally thought that this restriction would enable us to fully characterize the boundary $\mathcal{B}$ of the trapped region—the region through which closed trapped surfaces pass—but in fact we did not succeed in this. Still we are able to give what we think is a quite coherent picture of the trapped regions that occur in spherically symmetric

*ingemar@physto.se
†josemm.senovilla@ehu.es

spacetimes, and, in particular, we identify a past barrier for the location of closed trapped surfaces, and of marginally trapped tubes. This boundary turns out to have some quite nonlocal properties, and may penetrate flat regions of the spacetime. We will further prove that in spherically symmetric spacetimes the boundary $\mathcal{B}$ can never in itself contain any marginally trapped surfaces, so that it cannot be a marginally trapped tube.

Once we have learned that the boundary $\mathcal{B}$ of the trapped region suffers from nonlocal properties which are related to those of the event horizon, and that it is not a marginally trapped tube, we put forward a novel idea that may allow us to determine a preferred marginally trapped tube. We define the core $\mathcal{Z}$ of the trapped region as the region which is indispensable to maintain closed trapped surfaces. This core turns out to be generically smaller than the trapped region, and its boundary may thus be used as a definition of the black hole. We will actually identify a particular core in spherically symmetric spacetimes, and we will prove that its boundary is the unique spherically symmetric marginally trapped tube. It remains as an interesting question to know if this is the unique marginally trapped tube which is the boundary of a core.

In outline, our paper is organized as follows: Section II contains reminders about trapped surfaces, the fauna of different cases that can occur, and how they can be characterized in a way convenient for our purposes by the mean curvature vector. Section III defines the trapped region and its boundary $\mathcal{B}$, and gives their basic properties. Throughout the paper we provide proofs of all statements that we make. Section IV gives some basic results on which we build the rest of the paper. The arguments are purely geometrical and based on the interplay between (causal) vector fields and surfaces with special properties of their mean curvature vectors. Many of these results are already known [19–21] but we do offer some sharpenings, such as Theorem 4.1 and Corollary 4.1. Section V is an interlude dealing with Robertson-Walker spacetimes. In this case the general results of Sec. IV are sufficient to pin down the boundary $\mathcal{B}$ exactly. Section VI introduces spherically symmetric spacetimes and their unique spherically symmetric apparent 3-horizon. In Sec. VII, we discuss perturbations of the resulting round marginally trapped tubes. We use the stability operator that describes how the outer expansion varies when a marginally outer trapped surface is deformed [22,23], and give a proof that one can always find trapped surfaces that extend to both sides of the spherically symmetric apparent 3-horizon. We also show that the region of the perturbed trapped sphere that is inside the apparent 3-horizon can be made arbitrarily small. Section VIII discusses imploding and asymptotically flat spherically symmetric spacetimes in general. In Sec. IX we identify and discuss a past barrier through which future trapped surfaces cannot pass. It is based on the presence of the hypersurface forming Kodama vector field [24], and

our restriction is a definite improvement on previous results [3,15]. We also prove that any trapped surface must lie at least partly inside the spherically symmetric apparent 3-horizon. In Sec. VII, we discuss the precise location and the properties of the boundary $\mathcal{B}$ in a generic spherically symmetric and imploding spacetime. We demonstrate that the boundary $\mathcal{B}$ of the trapped region cannot be a marginally trapped tube. Section XI raises a new issue: Given that trapped surfaces must be confined at least partly within the spherically symmetric apparent horizon, what is the minimal region that must be removed from spacetime in order for it to have no closed trapped surfaces at all? We call this the core of the trapped region. We show that the spherically symmetric apparent 3-horizon is the boundary of one such core. Finally, the important example provided by the Vaidya spacetime is treated in an Appendix.

II. PRELIMINARIES: THE TRAPPED-SURFACE FAMILY

Trapped surfaces are the basic objects to be studied in this paper. Thus, we start by providing their definition and their types, and by fixing our notation. Standard references are [6,8,25–27].

Let $(\mathcal{V}, g)$ be a four-dimensional causally orientable spacetime with metric $g_{\mu\nu}$ of signature $-, +, +, +$. Let S denote a connected two-dimensional surface with local intrinsic coordinates $\{\lambda^A\}$ $(A, B, \ldots = 2, 3)$ imbedded in $\mathcal{V}$ by the C^3 parametric equations

$$x^\alpha = \Phi^\alpha(\lambda^A) \qquad (\alpha, \beta, \ldots = 0, 1, 2, 3)$$

where $\{x^\alpha\}$ are local coordinates for $\mathcal{V}$. The tangent vectors $\vec{e}_A$ of S are locally given by

$$\vec{e}_A \equiv e_A^\mu \frac{\partial}{\partial x^\mu} \Big|_S \equiv \frac{\partial \Phi^\mu}{\partial \lambda^A} \frac{\partial}{\partial x^\mu} \Big|_S$$

so that the first fundamental form of S in $\mathcal{V}$ is

$$\gamma_{AB} \equiv g_{\mu\nu}|_S \frac{\partial \Phi^\mu}{\partial \lambda^A} \frac{\partial \Phi^\nu}{\partial \lambda^B}$$

which collects the scalar products $g(\vec{e}_A, \vec{e}_B)$. From now on, we shall assume that γ_{AB} is positive definite so that S is a *spacelike* surface. Then, the two linearly independent normal one-forms $k_\mu^\pm$ to S can be chosen to be null and future directed everywhere on S, so they satisfy

$$k_\mu^\pm e_A^\mu = 0, \qquad k_\mu^+ k^{+\mu} = 0,$$
$$k_\mu^- k^{-\mu} = 0, \qquad k_\mu^+ k^{-\mu} = -1,$$

where the last equality incorporates a condition of normalization. Obviously, there still remains the freedom

$$k_\mu^+ \rightarrow k_\mu'^+ = \sigma^2 k_\mu^+, \qquad k_\mu^- \rightarrow k_\mu'^- = \sigma^{-2} k_\mu^- \qquad (1)$$

where σ^2 is a positive function defined only on S.

The standard splitting into tangential and normal directions to S leads to a formula relating the covariant derivatives on $(\mathcal{V}, g)$ and on (S, γ) [26,28]:

$$\nabla_{\vec{e}_A}\vec{e}_B = \bar{\Gamma}^C_{AB}\vec{e}_C - \vec{K}_{AB}$$

where $\bar{\Gamma}^C_{AB}$ are the coefficients of the Levi-Civita connection $\bar{\nabla}$ of γ (i.e. $\bar{\nabla}\gamma = 0$), and $\vec{K}_{AB}$ is the shape tensor (also called second fundamental form vector, and extrinsic curvature vector) of S in $(\mathcal{V}, g)$. Note that $\vec{K}_{AB} = \vec{K}_{BA}$ is symmetric, and orthogonal to S, from where we deduce

$$\vec{K}_{AB} = -K^-_{AB}\vec{k}^+ - K^+_{AB}\vec{k}^-.$$

Here, $K^\pm_{AB}$ are two symmetric covariant tensor fields defined on S and called the two null (future) second fundamental forms of S in $(\mathcal{V}, g)$. They are explicitly defined by

$$K^\pm_{AB} \equiv -k^\pm_\mu e^\nu_A \nabla_\nu e^\mu_B = e^\mu_B e^\nu_A \nabla_\nu k^\pm_\mu.$$

The shape tensor gives the difference between the projection to S of the covariant derivative and the intrinsic derivative on S by means of the fundamental relation

$$e^\mu_A e^\nu_B \nabla_\mu v_\nu|_S = \bar{\nabla}_A \bar{v}_B + v_\mu|_S K^\mu_{AB} \qquad (2)$$

where, for all v_μ we denote by

$$\bar{v}_B \equiv v_\mu|_S e^\mu_B$$

its projection to S.

The mean curvature vector of S in $(\mathcal{V}, g)$ [26,28] is the trace of the shape tensor

$$\vec{H} \equiv \gamma^{AB}\vec{K}_{AB}$$

where γ^{AB} is the contravariant metric on S: $\gamma^{AC}\gamma_{CB} = \delta^A_B$. Observe that $\vec{H}$ is orthogonal to S and that

$$\vec{H} \equiv -\theta^-\vec{k}^+ - \theta^+\vec{k}^-$$

where

$$\theta^\pm \equiv \gamma^{AB}K^\pm_{AB} \qquad (3)$$

are the traces of the null second fundamental forms, usually called the (future) null expansions. Clearly, $\vec{H}$ is invariant under transformations (1).

The class of *weakly future-trapped* (f-trapped from now on) surfaces are characterized by having $\vec{H}$ pointing to the future everywhere on S, and similarly for weakly past trapped. There are three important subcases that deserve their own name: (i) the traditional f-trapped surfaces have $\vec{H} \neq \vec{0}$ timelike everywhere on S; (ii) marginally f-trapped surfaces have $\vec{H} \neq \vec{0}$ null everywhere on S; and (iii) minimal surfaces have $\vec{H} \equiv \vec{0}$ on S.

These conditions can be equivalently expressed in terms of the signs of the expansions as follows:

$\vec{H}$	Null expansions	Type of surface
causal future	$\theta^+ \leq 0, \theta^- \leq 0$	weakly f-trapped
Subcases:		
zero	$\theta^+ = \theta^- = 0$	minimal
null and future $\neq \vec{0}$	$\theta^+ = 0, \theta^- \leq 0$	marginally f-trapped
null and future $\neq \vec{0}$	$\theta^+ \leq 0, \theta^- = 0$	marginally f-trapped
timelike future	$\theta^+ < 0, \theta^- < 0$	f-trapped

This is to be compared with [3,8,27,29], as sometimes different names are given to the same objects, and vice versa. In particular, weakly f-trapped surfaces were called nearly f-trapped in [19,29]. Here we will follow the previous nomenclature which pretends to respect standard names as much as possible. See [29] for further details. In the case that $\vec{H}$ is proportional to one of the null normals, but realizing both causal orientations, the surface is said to be null dual, or marginally ($\pm$)-trapped, where the $\pm$ refers to the direction with vanishing expansion—as the definition is equivalent to having either θ^+ or θ^- vanishing. In the literature they are usually referred as MOTS (marginally outer trapped surfaces), by declaring the direction with vanishing expansion to be ''outer''.

For completeness and future reference, we also mention that a surface is called *untrapped* if the mean curvature vector is spacelike everywhere, or equivalently, if both expansions have opposite signs.

III. DEFINITION AND BASIC PROPERTIES OF $\mathcal{T}$

In this paper, we will be concerned with the following sets in $(\mathcal{V}, g)$, see also [17].

Definition 1.—The future-trapped region $\mathcal{T}$ is defined as the set of points $x \in \mathcal{V}$ such that x lies on a closed future-trapped surface.

Since the characterization of f-trapped surfaces is the negativity of both null expansions $\theta^\pm$ as defined in (3), the following general property follows easily.

Property 1.—The future-trapped region $\mathcal{T}$ is an open set.

Proof.—Take any $x \in \mathcal{T}$ and let $S \ni x$ be a closed f-trapped surface. We can perturb any such S along an arbitrary direction $\vec{q}$ defined on S, by moving any point $y \in S$ along $\vec{q}$ a distance $\epsilon > 0$. The deformed surface S_ϵ has future null expansions $\theta^\pm_\epsilon = \theta^\pm + \epsilon\delta_{\vec{q}}\theta^\pm + O(\epsilon^2)$, where $\theta^\pm$ are the null expansions on S and the variations $\delta_{\vec{q}}\theta^\pm$ are given by precise formulas involving $\vec{q}$, the future null normals to S and the geometric properties of S, see e.g. [9,11,17,22,23,30]. As $\theta^\pm < 0$ and $\delta_{\vec{q}}\theta^\pm$ are continuous on S, we can always choose $\epsilon(\vec{q})$ small enough such that $\theta^\pm_\epsilon < 0$ for *any* given direction $\vec{q}$. This implies that there exists a small neighborhood $U(x)$ of $x \in S$ such that $U(x) \subset \mathcal{T}$.

In general, $\mathcal{T}$ does not have to be connected. Any connected component of $\mathcal{T}$ will thus be termed as a ''connected f-trapped region''.

It is clear that $\mathcal{T}$ can be empty, or it can be the whole spacetime. An example of the former case $\mathcal{T} = \varnothing$ is provided by any globally static spacetime [19]. Examples of the latter case $\mathcal{T} = \mathcal{V}$ are de Sitter spacetime or some contracting Robertson-Walker geometries, see Sec. V. However, in asymptotically flat black-hole type spacetimes [1,8,27], neither of these cases will happen, because there are no trapped surfaces near spatial infinity, and there will appear future-trapped surfaces in the black-hole region. In these cases, $\mathcal{T}$ will have a boundary on $(\mathcal{V}, g)$.

Definition 2.—We denote by $\mathcal{B}$ the boundary of the future-trapped region $\mathcal{T}$:

$$\mathcal{B} \equiv \partial \mathcal{T}.$$

Property 2.—$\mathcal{B}$ being the boundary of an open set, it is itself a closed set without boundary. Moreover $\mathcal{B} \cap \mathcal{T} = \varnothing$.

Property 3.—$\mathcal{B}$ is also the boundary of the untrapped region defined by the set of points $x \notin \mathcal{T}$, that is, such that x does not lie on any closed f-trapped surface.

We remark that $\mathcal{B}$ is not necessarily connected.

$\mathcal{T}$ and $\mathcal{B}$ are genuine spacetime objects, independent of any foliations or initial Cauchy data sets. Therefore, $\mathcal{T}$ and $\mathcal{B}$ are different in nature from the trapped regions and their boundaries contained in given slices, as recently studied in [9].

The symmetries of the spacetimes respect $\mathcal{T}$ and $\mathcal{B}$. More precisely:

Result 3.1.—If G is the group of isometries of the spacetime $(\mathcal{V}, g)$, then $\mathcal{T}$ is invariant under the action of G, and the transitivity surfaces of G, relative to points of $\mathcal{B}$, remain in $\mathcal{B}$.

Proof.—Take any point $x \in \mathcal{T}$. Then there is a closed f-trapped surface S passing through x. By moving S via the motion group G, and as f-trapped surfaces are moved to f-trapped surfaces by isometries, it follows that the whole transitivity surface of the group passing through x lies in $\mathcal{T}$. Similarly, let $y \in \mathcal{B}$, so that any small neighborhood of y intersects $\mathcal{T}$. By moving one such small neighborhood using G, one similarly deduces that the transitivity surface of G relative to y is part of $\mathcal{B}$.

An implication of this result is that no globally defined Killing vector can be transversal to $\mathcal{B} = \partial \mathcal{T}$. Actually, this also holds for homothetic Killing vectors (vector fields $\vec{\xi}$ satisfying $(\mathcal{L}_{\vec{\xi}} g)_{\mu\nu} = 2c g_{\mu\nu}$), as they also respect f-trapped surfaces (if $c > 0$) [31].

Corollary 3.1.—Let G_n be the (global) continuous group of isometries of $(\mathcal{V}, g)$ where n is its dimension (i.e., the number of linearly independent Killing vectors), and let m be the dimension of its surfaces of transitivity.

(1) If $n \geq 4$ and $m = 4$, that is, if the spacetime is homogeneous, then $\mathcal{B} = \varnothing$, and either $\mathcal{T} = \varnothing$ or $\mathcal{T} = \mathcal{V}$.

(2) If $n \geq 3$ and $m = 3$, then either $\mathcal{B} = \varnothing$ or each connected component of $\mathcal{B}$ is one of the three-dimensional surfaces of transitivity.

(3) If $n \geq 2$ and $m = 2$, then either $\mathcal{B} = \varnothing$ or each connected component of $\mathcal{B}$ is a hypersurface without boundary foliated by the two-dimensional surfaces of transitivity. In particular, in arbitrary spherically symmetric spacetimes, $\mathcal{B}$ (if not empty) is a spherically symmetric hypersurface without boundary.

Proof.—Point 1 is immediate. Points 2 and 3 follow because any connected component of $\mathcal{B}$ cannot have a boundary and cannot be given by isolated two-dimensional surfaces of transitivity, as these would contradict its basic Properties 2 and 3.

What is the possible relevance of $\mathcal{B}$? Apart from answering natural questions such as "where can there be closed f-trapped surfaces?" or "is this event part of a closed f-trapped surface?," the location of $\mathcal{B}$ provides important physical information due to the fundamental relevance of closed trapped surfaces in the development of black holes and singularities [6,25,32,33]. More importantly, it provides a precise limit as to where dynamical horizons or marginally trapped tubes can develop. One could also hope that $\mathcal{B}$ is related to the surface of a dynamical black hole. We are going to see that this suffers from the same problems as other candidates.

IV. FUNDAMENTAL RESULTS

In this section, we present the main results that will allow us to put restrictions on the location of the region $\mathcal{T}$ containing the closed future-trapped surfaces of a spacetime and its boundary $\mathcal{B}$. These results are fully general, and can be obtained within the framework of the interplay between generalized symmetries and submanifolds with special properties of their mean curvature vector. The underlying ideas come from [19–21].

We start with the main formula to be used in what follows. Let $\vec{\xi}$ be an arbitrary C^1 vector field on $\mathcal{V}$ defined on a neighborhood of S. Recalling the identity $(\mathcal{L}_{\vec{\xi}} g)_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu$, where $\mathcal{L}_{\vec{\xi}}$ denotes the Lie derivative with respect to $\vec{\xi}$, one gets on using (2)

$$(\mathcal{L}_{\vec{\xi}} g|_S)_{\mu\nu} e_A^\mu e_B^\nu = \bar{\nabla}_A \bar{\xi}_B + \bar{\nabla}_B \bar{\xi}_A + 2 \xi_\mu|_S K_{AB}^\mu.$$

Contracting now with γ^{AB} we arrive at

$$\frac{1}{2} P^{\mu\nu} (\mathcal{L}_{\vec{\xi}} g|_S)_{\mu\nu} = \bar{\nabla}_C \bar{\xi}^C + \xi_\rho H^\rho \qquad (4)$$

where

$$P^{\mu\nu} \equiv \gamma^{AB} e_A^\mu e_B^\nu$$

is the orthogonal projector that projects any object to its part tangent to S.

The elementary formula (4) is very useful and permits one to obtain many interesting results about the existence of weakly trapped closed surfaces in the presence of generalized symmetries, see [19–21]. For instance:

1. If S is minimal ($\vec{H} = \vec{0}$), integrating (4) and using Gauss' theorem, one finds that the divergence term does not contribute *whenever* S is closed (that is, compact without boundary), ergo

$$\oint_S P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} = 0.$$

Observe that this relation must be satisfied for *all* possible vector fields $\vec{\xi}$. Therefore, closed minimal surfaces are very rare.

2. If $\vec{\xi}$ is a Killing vector, then the left-hand side of (4) vanishes. Integrating the right-hand side on S we get for closed S

$$\oint_S \xi_\rho H^\rho = 0.$$

Therefore, if the Killing vector $\vec{\xi}$ is timelike on S, then S cannot be weakly trapped (neither future nor past), unless it is minimal.

The following is an important consequence of formula (4).

Lemma 4.1.—Let $\vec{\xi}$ be a vector field which is future pointing on a region $\mathcal{R} \subset \mathcal{V}$, and let S be a surface contained in $\mathcal{R}$ such that $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} \geq 0$. Then, S cannot be closed and weakly f-trapped unless $\xi_\mu H^\mu = 0$ and $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} = 0$.

Proof.—Integrating (4) on the closed S, the divergence term integrates to zero and we get

$$\oint_S \xi_\rho H^\rho = \frac{1}{2}\oint_S P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} \geq 0$$

which implies that $\vec{H}$ cannot be future pointing all over S, unless $\xi_\mu H^\mu = P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} = 0$.

Remarks:

(i) For nonminimal S, the exceptional case $\xi_\mu H^\mu = 0$ implies that S is marginally f-trapped and that $\vec{\xi}|_S$ is null and proportional to $\vec{H}$.

(ii) Notice that only the averaged condition $\oint_S P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} \geq 0$ is needed here, so that $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu}$ can be negative somewhere on S as long as the averaged formula holds.

Important instances where Lemma 4.1 can be applied are given by the *conformal Killing vectors* (see e.g. [34]) and the *Kerr-Schild vector fields* [35]. The former satisfy

$$(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu} = 2\psi g_{\mu\nu} \qquad (5)$$

for some function ψ, so that $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} = 4\psi|_S$. Thus, the condition on Lemma 4.1 requires simply that $\psi|_S \geq 0$. On the other hand, Kerr-Schild vector fields are characterized by

$$(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu} = 2h\ell_\mu\ell_\nu, \qquad (\mathcal{L}_{\vec{\xi}}\ell)_\mu = b\ell_\mu \qquad (6)$$

for some functions h and b, where ℓ_μ is a fixed *null* 1-form field ($\ell_\mu \ell^\mu = 0$). Therefore, $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} = 2h\bar{\ell}_A\bar{\ell}^A$ and the condition on the Lemma 4.1 requires now $h|_S \geq 0$.

Stronger results can be found for the case where $\vec{\xi}$ is a hypersurface-orthogonal vector field, that is to say,

$$\xi_{[\mu}\nabla_\nu\xi_{\rho]} = 0 \Leftrightarrow \xi_\mu = -F\partial_\mu\tau$$

for some local functions $F > 0$ and τ. Then, $\vec{\xi}$ is orthogonal to the hypersurfaces $\tau = $ const, which are called the level hypersurfaces. In this case we have the following fundamental result.

Theorem 4.1.—Let $\vec{\xi}$ be a vector field which is future-pointing and hypersurface-orthogonal on a region $\mathcal{R} \subset \mathcal{V}$ with level hypersurfaces $\tau = $ const and let S be a f-trapped surface. Then, S cannot have a local minimum of τ at any point $q \in \mathcal{R}$ where $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu}|_q \geq 0$.

In the case that S is a weakly f-trapped surface, the conclusion still holds unless

$$\left.\frac{\partial^2\bar{\tau}}{\partial\lambda^A\partial\lambda^B}\right|_q = 0 \quad \text{and} \quad P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu}|_q = 0 \quad \text{and}$$

$$\xi_\rho H^\rho|_q = 0. \qquad (7)$$

Proof.—Let $q \in S \cap \mathcal{R}$ be a point where S has a local extreme of τ. Then, $\vec{\xi}$ is orthogonal to S at q, that is to say, $\bar{\xi}_A|_q = 0$. Another way of stating the same is that

$$\left.\frac{\partial\bar{\tau}}{\partial\lambda^A}\right|_q = 0$$

where $\tau = \bar{\tau}(\lambda^A)$ is the local parametric expression of τ in terms of the local coordinates $\{\lambda^A\}$ of S (note that $\bar{\xi}_A = -\bar{F}\partial\bar{\tau}/\partial\lambda^A$ with $\bar{F} \equiv F|_S$). Using the previous expression, we can now compute the divergence $\bar{\nabla}_C\bar{\xi}^C$ at q:

$$\bar{\nabla}_A\bar{\xi}^A|_q = \gamma^{AB}\bar{\nabla}_A\left(-\bar{F}\frac{\partial\bar{\tau}}{\partial\lambda^B}\right)\bigg|_q = -\bar{F}\gamma^{AB}\frac{\partial^2\bar{\tau}}{\partial\lambda^A\partial\lambda^B}\bigg|_q.$$

Introducing this in formula (4) we get at q

$$\bar{F}\gamma^{AB}\frac{\partial^2\bar{\tau}}{\partial\lambda^A\partial\lambda^B}\bigg|_q = -\frac{1}{2}P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu}\bigg|_q + \xi_\rho H^\rho|_q \leq \xi_\rho H^\rho|_q$$

so that, if S has a future pointing (possibly vanishing) $\vec{H}|_q$, we deduce

$$\gamma^{AB}\frac{\partial^2\bar{\tau}}{\partial\lambda^A\partial\lambda^B}\bigg|_q \leq 0$$

which, given that γ^{AB} is positive definite, implies that $\partial^2\bar{\tau}/\partial\lambda^A\partial\lambda^B|_q$ cannot be positive (semi)-definite. Therefore, $\bar{\tau}$ cannot have a local minimum at q.

Remarks:

(i) Observe that S does not need to be compact, nor contained in $\mathcal{R}$.

(ii) Notice that it is enough to assume $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu}|_q \geq$ 0 only at the points q that are local extrema of τ on S. The sign of $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu}$ can thus be left arbitrary everywhere else on S where $\bar{\xi}_A \neq 0$.

(iii) Let us stress that the possibility with a positive semidefinite $\partial^2\bar{\tau}/\partial\lambda^A\partial\lambda^B|_q$ is also excluded, so that $\bar{\tau}$ cannot even be constant along one direction at q. The only exceptional possibility is given by the case identified in the theorem satisfying (7). If $\vec{\xi}|_q$ is timelike, the last in (7) implies that $\vec{H}|_q = \vec{0}$. Therefore, letting aside this exceptional possibility, τ will always decrease at least along one tangent direction in T_qS. It follows that, under the conditions of the theorem, starting from any point $x \in S \cap \mathcal{R}$, one can always follow a connected path along $S \cap \mathcal{R}$ with decreasing τ.

(iv) Theorem 4.1 applies, in particular, (but not only) to (a) *static* Killing vectors of course, (b) hypersurface-orthogonal conformal Killing vectors (5) with $\psi \geq 0$, and (c) hypersurface-orthogonal Kerr-Schild vector fields (6) with $h \geq 0$.

(v) The results in section IV of [15] —that any f-trapped surface penetrating a flat portion of the spacetime cannot have a minimum of "inertial time" there, so that they have to "bend down in time"—are simple consequences of the more general Theorem 4.1, which applies not only to flat spacetimes but in general to any static region, and to much more general cases as remarked above.

Another important result for hypersurface-orthogonal vector fields is (see also [19–21]):

Theorem 4.2.—Let $\vec{\xi}$ be a vector field which is future-pointing and hypersurface-orthogonal on a region $\mathcal{R} \subset \mathcal{V}$. Then, all spacelike surfaces S (compact or not) contained in one of the level hypersurfaces $\tau = $ const within $\mathcal{R}$ have

$$2\xi_\rho H^\rho = P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu}. \tag{8}$$

In particular, at any point $x \in S$ such that $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu}|_x \geq 0$, S has a mean curvature vector which is not timelike future-pointing, and it can be future-pointing null or zero only if $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu}|_x = \xi_\mu H^\mu|_x = 0$.

Proof.—Let $S \cap \mathcal{R}$ be a (portion of a) surface contained in one of the hypersurfaces $\tau = $ const. Then $\bar{\xi}_A = 0$ all over $S \cap \mathcal{R}$ so that from (4) we deduce (8). This immediately implies the rest of results.

From Theorems 4.2 and 4.1, we deduce the following general property.

Corollary 4.1.—No f-trapped surface (closed or not) can touch a spacelike hypersurface to its past at a single point, or have a two-dimensional portion contained in the hypersurface, if the latter has a positive semidefinite second fundamental form.
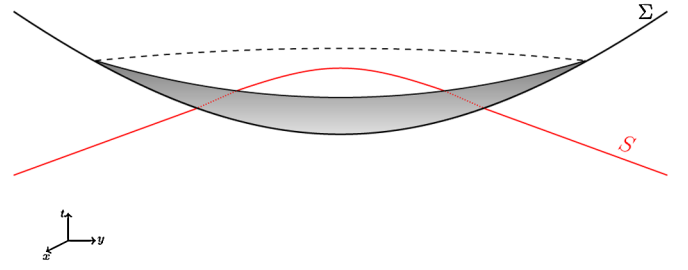

FIG. 1 (color online). Schematic diagram of the situation relevant to Corollary 4.1. For simplicity the picture is drawn in flat spacetime, one dimension suppressed. The trapped surface S, represented here by a red line, has to bend down in time t. Σ represents a spacelike hypersurface with positive definite second fundamental form.

Proof.—As the hypersurface is spacelike, its normal vector, say $\vec{\xi}$, is timelike, and can be extended to be hypersurface-orthogonal with level function τ and can be chosen to be future-pointing. Then, the projection of $(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu}$ to the hypersurface is proportional to its second fundamental form. Thus, if this were positive semidefinite, it would follow that $P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu} \geq 0$, where $P^{\mu\nu}$ is the projector to any such surface S, and the result follows.

An intuitive explanation of this result is presented in Fig. 1.

V. APPLICATION TO ROBERTSON-WALKER GEOMETRIES

Take the Robertson-Walker spacetimes, with line-element given by [27,34]

$$ds^2 = -dt^2 + a^2(t)d\Sigma_k^2 \tag{9}$$

where $d\Sigma_k^2$ is the standard metric on a maximally symmetric three-dimensional space with normalized sectional curvature $k = \pm 1, 0$, and $a(t)$ is a function of the preferred time t called the scale factor. The future is defined by increasing values of t. The above line-element possesses a timelike conformal Killing vector given by $\vec{\xi} = a(t)\partial_t$ such that ψ in (5) is simply $\dot{a} \equiv da/dt$. Furthermore, $\vec{\xi}$ is obviously hypersurface-orthogonal

$$\xi_\mu dx^\mu = -a(t)dt,$$

with level surfaces given by the preferred hypersurfaces $t = $ const. $\vec{\xi}$ has been used to derive results on closed weakly f-trapped surfaces in [19,20] and more recently on MOTS in [31].

Result 5.1.—No closed weakly f-trapped surface S can be fully contained in the region $\dot{a} \geq 0$ of a Robertson-Walker spacetime, unless $k = 1$ and S is a minimal surface imbedded in a 3-sphere $t = t_0$ with $\dot{a}(t_0) = 0$.

Proof.—Direct application of Lemma 4.1 implies [19,20] that the only open possibility is given by $\vec{H} = \vec{0}$ and $\dot{a}|_S = 0$. Given that the hypersurface $t = t_0$ has

vanishing second fundamental form, any such minimal surface must be minimal *within* the hypersurface $t = t_0$. This immediately rules out the cases $k = 0, -1$, as there are no compact minimal surfaces imbedded in Euclidean or hyperbolic spaces [36,37].

Similarly, from Theorem 4.1 follows that

Result 5.2.—No closed weakly f-trapped surface can have a local minimum of t at the region with $\dot{a} \geq 0$. Thus, the minimum of t must be attained at a hypersurface $t = \tilde{t}$ with $\dot{a}|_{\tilde{t}} < 0$, if they exist.

As a consequence, f-trapped closed surfaces are absent in generic *expanding* Robertson-Walker models. They can only be present in models with contracting phases. Actually, it is well-known—e.g. [8,33]—that there are closed trapped round spheres at any $t = $ const slice with $\dot{a} \neq 0$ and $\dot{a}^2 + k > 0$. This last condition is simply the positivity of $\varrho + \Lambda$, where ϱ is the energy density relative to the preferred observer ∂_t and Λ the cosmological constant. Given the homogeneity of the maximally symmetric slices, this implies that there are closed trapped round spheres through *any* point of the slices with $\dot{a} \neq 0$ and $\dot{a}^2 + k > 0$. These spheres are, of course, past-trapped if $\dot{a} > 0$ and f-trapped if $\dot{a} < 0$. The only remaining possibility—keeping $\dot{a}^2 + k > 0$—is that of three-sphere slices ($k = 1$) with $\dot{a} = 0$, in which case all round spheres in the slice are untrapped except for the equatorial one, which is a minimal surface.

From Corollary 3.1, we already know that the boundary $\mathcal{B}$ must be constituted by $t = $ const slices. Thus, $\mathcal{B}$ splits into two different disconnected sets, $\mathcal{B}^-$ and $\mathcal{B}^+$, according to whether the closed f-trapped surfaces lie locally to the future or past of $\mathcal{B}$, respectively. Combining this with the arguments in the previous paragraph we can deduce the following.

Result 5.3.—On a general Robertson-Walker spacetime with line-element (9) and $\dot{a}^2 + k \geq 0$

(i) If $\dot{a} \geq 0$ everywhere, then $\mathcal{T} = \varnothing$ and thus $\mathcal{B} = \varnothing$.
(ii) If $\dot{a} \leq 0$ everywhere, then $\mathcal{T} = \mathcal{V}$ and thus $\mathcal{B} = \varnothing$.
(iii) In the case that $\dot{a}$ changes sign, the past boundary $\mathcal{B}^-$, if nonempty, is necessarily contained in the region with $\dot{a} = 0$ and $\ddot{a} < 0$:

$$\mathcal{B}^- \subset \{\dot{a} = 0\} \cap \{\ddot{a} < 0\}.$$

Proof.—The first two points are direct consequences of Results 5.1 and 5.2. To prove the third point, suppose that an f-trapped surface S enters into a region with $\dot{a} > 0$. Then, S cannot have a minimum of t there due to Result 5.2. If S is closed, S must attain the minimum of t, hence S has to extend to the past for all values of t until it crosses the first slice with $\dot{a} = 0$ and $\ddot{a} > 0$, entering into a region with $\dot{a} < 0$. But we know that there are closed f-trapped round spheres through every point of such a region, so the boundary $\mathcal{B}^-$ cannot be there. Proceeding towards the past, we either encounter another slice with $\dot{a} = 0$ and $\ddot{a} < 0$ or

not. In the second possibility, there is no boundary $\mathcal{B}^-$ for the connected component of $\mathcal{T} \supset S$. In the first case, either that slice is such a boundary $\mathcal{B}^-$, or else there are closed f-trapped surfaces crossing the slice towards the past, that is, entering into another region with $\dot{a} > 0$. We can then repeat the argument from the beginning until one of the slices with $\dot{a} = 0$ and $\ddot{a} < 0$ is the past boundary of the mentioned connected f-trapped region or there is no such a boundary.

The last point in the previous result can be rephrased as saying that the past boundary $\mathcal{B}^-$ is constituted by preferred slices with $\dot{a} = 0$, which have a vanishing second fundamental form and therefore are maximal and totally geodesic, and such that the spacetime starts to recollapse there.

Corollary 5.1.—If the Robertson-Walker geometry (with $\dot{a}^2 + k \geq 0$) has an initial expanding epoch, then either $\mathcal{B} = \varnothing$—if it is expanding forever—, or the first recollapsing time is always part of $\mathcal{B}^-$.

There is the open question whether the future boundary $\mathcal{B}^+$ can be nonempty. If so, one easily obtains that it is necessarily contained in the region with $\dot{a} > 0$. (Actually, one can further prove that $\mathcal{B}^+ \subset \{\dot{a} > 0\} \cap \{\ddot{a} \geq 0\}$, but this will not be necessary in what follows.) The question of whether there can be weakly f-trapped surfaces intersecting both expanding and contracting regions is answered in the positive, as follows from de Sitter spacetime or in more general cases from the results in [38].

An important consequence of all the above is the following.

Result 5.4.—On a general Robertson-Walker spacetime with line-element (9) and $\dot{a}^2 + k \geq 0$, the boundary $\mathcal{B}$ of the future-trapped region $\mathcal{T}$ does not contain any non-minimal weakly f-trapped-surface (closed or not).

Proof.—Take any slice $t = t_0$ which is part of $\mathcal{B}$, so that $\dot{a}(t_0) \geq 0$, and choose any surface S, closed or not, imbedded into $t = t_0$. Then $\vec{\xi}$ is normal to S so that $\bar{\xi}_A = 0$ and Theorem 4.2 implies that $\xi_\rho|_S H^\rho = \dot{a}(t_0) \geq 0$. Thus, $\vec{H}$ has to be spacelike, possibly zero, or past-pointing everywhere on S (the past-pointing possibility is forbidden if $\dot{a}(t_0) = 0$, i.e., within $\mathcal{B}^-$.)

We will see that this turns out to be a rather general property in spherical symmetry, so that $\mathcal{B}$ will never contain closed weakly f-trapped surfaces.

Observe that, in the case $k = 0$ (say; a similar reasoning works for the case $k = -1$), if $\mathcal{B}^-$ is nonempty, there are f-trapped round spheres as close to $\mathcal{B}^-$ as we like. What happens if we try to approach $\mathcal{B}^-$ following a sequence of such f-trapped round spheres? The answer is simply that they get larger and larger, and if we go to the limit approaching $\mathcal{B}^-$, they actually break and become noncompact minimal planes.

As we have seen, not all the $\dot{a} = 0$ hypersurfaces will be part of the boundary in general. An illustrative example is given by de Sitter spacetime, which has $k = 1$ and

$a(t) = \cosh t$, so that $\dot{a}$ is negative, zero, or positive for $t < 0$, $t = 0$, or $t > 0$, respectively. Thus, de Sitter spacetime has f-trapped round spheres in any slice with $t < 0$. However, de Sitter spacetime is maximally symmetric, and therefore all of its points are fully equivalent, so that there are f-trapped round spheres through every point. Hence, $\mathcal{B} = \varnothing$ and $\mathcal{T} = \mathcal{V}$ in de Sitter spacetime.

A different, more standard example, is provided by the closed Friedman model for dust and $\Lambda = 0$, defined by $k = 1$ with the following parametric form for $a(t)$:

$$a(\gamma) = C(1 - \cos\gamma), \quad t(\gamma) = C(\gamma - \sin\gamma), \quad \gamma \in (0, 2\pi)$$

where C is a constant. Obviously $\dot{a} = \sin\gamma/(1 - \cos\gamma)$. Now there are closed f-trapped round spheres through all

points of any slice with $t > C\pi$ ($\gamma > \pi$), as then $\dot{a} < 0$. However, no closed f-trapped surface can enter the region with $t < C\pi$ ($\gamma < \pi$), as otherwise they would reach their minimum value of t at a slice with $\dot{a} > 0$. Thus, in this case $\mathcal{B} = \mathcal{B}^- = \{t = C\pi\}$. The Penrose diagram of this situation is depicted in Fig. 2.

Obviously, the same holds for arbitrary models with any value of k (but keeping $\dot{a}^2 + k \geq 0$), such that $a(t)$ has a unique maximum value, that is, for models with one expanding phase, a unique recollapsing time, and one contracting phase.

Observe that, in accordance with our general results, the slice $t = 0$ in de Sitter spacetime has $\dot{a} = 0$ but $\ddot{a} > 0$, while in the previous models with a unique recollapsing time, such as the closed Robertson-Walker metric dust model of Fig. 2, $\ddot{a} < 0$ everywhere.

VI. SPHERICALLY SYMMETRIC SPACETIMES AND THE APPARENT 3-HORIZON (AH)

A spacetime is spherically symmetric if it admits an SO(3) group of isometries. This group acts transitively on round spheres embedded in spacetime. Let $d\Omega^2$ denote the standard metric on the unit round spheres, then the line-element can always be written as

$$ds^2 = dL^2 + r^2 d\Omega^2$$

for some Lorentzian two-dimensional metric dL^2 and where r is constant on each round sphere. We assume from now on that $dr \neq 0$, hence r can be chosen as one of the coordinates for the metric dL^2. It is called the area coordinate because the preferred round spheres have an area equal to $4\pi r^2$. We shall use the invariantly defined *mass function*

$$m \equiv r(1 - \nabla_\mu r \nabla^\mu r)/2.$$

By choosing v as a coordinate labeling the incoming radial null geodesics, the spherically symmetric line-element can then be written as

$$ds^2 = -e^{2\beta}\left(1 - \frac{2m(v, r)}{r}\right)dv^2 + 2e^\beta dv dr + r^2 d\Omega^2$$

$$(10)$$

with $\beta(v, r)$. The future is defined by increasing values of v. These are called the advanced Eddington-Finkelstein coordinates. They may not be globally defined in some spacetimes, but they are well adapted to our purposes: describing the cases with incoming matter and radiation.

The future-directed radial null geodesic vector fields are given by

$$\vec{\ell} = -e^{-\beta}\partial_r, \qquad \vec{k} = \partial_v + \frac{1}{2}\left(1 - \frac{2m}{r}\right)e^\beta \partial_r \quad (11)$$
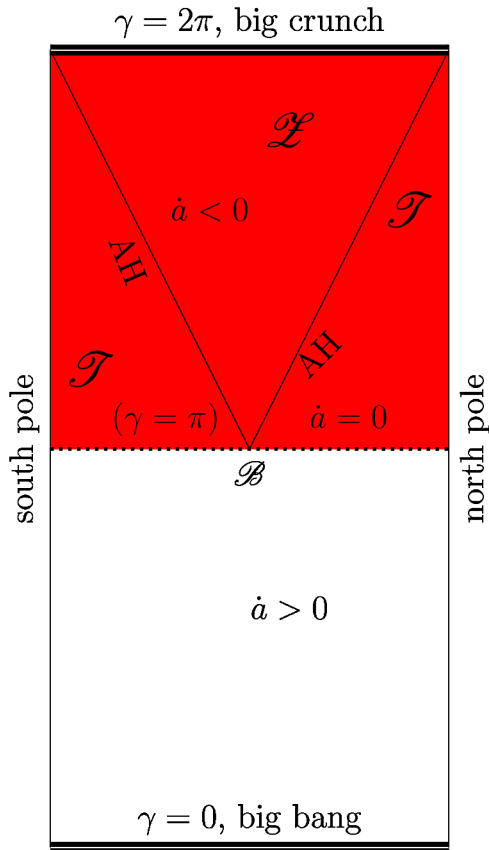
so that

FIG. 2 (color online). Conformal diagram of the Robertson-Walker spacetime (9) for $k = 1$ and dust. As usual, null radial lines are drawn at 45° and the future direction is upwards. Closed f-trapped surfaces exist in the colored region (in red, this is $\mathcal{T}$), defined by $\dot{a} < 0$. The boundary $\mathcal{B}$ is the maximal hypersurface with $\dot{a} = 0$, the recollapsing time, shown here as a dotted horizontal line. A marginally trapped tube—corresponding to the apparent 3-horizon AH to be defined in Sec. VI—foliated by marginally f-trapped round spheres in also shown. The region $\mathcal{Z}$ to the future of the AH is a core of the trapped region, as defined in Sec. XI. If $\mathcal{Z}$ is removed from the spacetime, then there are no remaining closed f-trapped surfaces.

$$\ell_\mu dx^\mu = -dv, \qquad k_\mu dx^\mu = e^\beta dr - \frac{1}{2}\left(1 - \frac{2m}{r}\right)e^{2\beta}dv.$$

They satisfy $k_\mu \ell^\mu = -1$.

The mean curvature vector for each round sphere $\vec{H}_{\mathrm{sph}}$ can be easily computed

$$\vec{H}_{\mathrm{sph}} = \frac{2}{r}\left(e^{-\beta}\partial_v + \left(1 - \frac{2m}{r}\right)\partial_r\right). \qquad (12)$$

Taking into account that for each round sphere the two future-pointing null normals are $\vec{k}^+_{\mathrm{sph}} = \vec{k}$ and $\vec{k}^-_{\mathrm{sph}} = \vec{\ell}$, the null expansions can be read off from (12):

$$\theta^+_{\mathrm{sph}} = \frac{e^\beta}{r}\left(1 - \frac{2m}{r}\right), \qquad \theta^-_{\mathrm{sph}} = -\frac{2e^{-\beta}}{r}. \qquad (13)$$

Thus, these round spheres are f-trapped if and only if $r < 2m$, and they are marginally f-trapped if $r = 2m$. The set defined by

$$\mathrm{AH}: \; r - 2m(v, r) = 0$$

is formed by hypersurfaces foliated by marginally f-trapped round spheres. Thus, each of these hypersurfaces is a "marginally trapped tube" [3]. They are called the spherically symmetric "apparent 3-horizons", as each of its marginally f-trapped round spheres is an apparent horizon—in the sense of [8], see also [10]—for a time slice that respects the symmetry. And it is unique with these properties:

Result 6.1.—AH are the unique spherically symmetric hypersurfaces—with respect to the given SO(3) group—foliated by marginally f-trapped surfaces.

Remarks:

(i) There are Lorentzian manifolds with several SO(3) groups of isometries, such as flat, de Sitter, or Robertson-Walker spacetimes (simply change the origin of coordinates). This is why we have to fix the group defining the spherical symmetry. However, in generic situations the isometry group SO(3) is unique, and then the AH is unique in absolute sense.

(ii) The *foliation* hypothesis is crucial here, as there can be marginally trapped surfaces different from round spheres, and actually of any genus, imbedded in spherically symmetric hypersurfaces. Explicit examples are given in [38] for Robertson-Walker spacetimes with $k = 1$, where marginally trapped surfaces of any genus are imbedded in $t = \mathrm{const}$ slices.

(iii) The assumption of spherical symmetry for the hypersurface is essential here, as there are nonspherically-symmetric marginally trapped tubes in spherically symmetric spacetimes. This is a general property, as we will show in Sec. VII B. Explicit examples for closed ($k = 1$) Robertson-Walker spacetimes are given in [38].

Proof.—That AH is the unique spherically symmetric set constituted by marginally f-trapped round spheres is obvious by construction. Suppose then that there is a spherically symmetric hypersurface Σ foliated by marginally f-trapped closed surfaces $\{S_\lambda\}$. Pick up any such surface, say S_0. By using the SO(3) group of isometries, move S_0 to obtain a new, necessarily marginally f-trapped, closed surface S'_0 in Σ. This surface must be tangent at some point x to one of the foliating surfaces, say S_{λ_0}. But then, using the maximum principle as in the proof of Proposition 3.1 of [3], one can deduce that $S'_0 = S_{\lambda_0}$, and *a fortiori*, that $S_0 = S'_0 = S_{\lambda_0}$, so that S_0 must be tangent to the SO(3) Killing vectors. The proof in [3] assumed that Σ was spacelike, however this is not necessary and the reasoning works equally well for timelike Σ. Actually, it can be seen that the result holds for null Σ as long as its null generator does not point along the null direction with vanishing expansion of S_λ. In the remaining case in which the null generator of Σ points along the direction of vanishing null expansion of the S_λ, all possible cross sections of Σ have the corresponding null expansion vanishing—in particular, the round spheres in Σ are marginally f-trapped—, and therefore the result holds true too.

The exceptional situation that has arisen in the previous proof is only possible in cases where AH happens to have portions of isolated-horizon type [5], that is, null portions such that their null generators point along $\vec{k}^+_{\mathrm{sph}} = \vec{k}$. To understand when this occurs, note that the normal 1-form to AH is

$$n_\mu dx^\mu = \left(1 - 2\frac{\partial m}{\partial r}\right)dr - 2\frac{\partial m}{\partial v}dv$$

whose norm is

$$n_\mu n^\mu\big|_{\mathrm{AH}} = -4e^{-\beta}\frac{\partial m}{\partial v}\left(1 - 2\frac{\partial m}{\partial r}\right)\bigg|_{\mathrm{AH}}$$

so that AH is null at any point $x \in \mathrm{AH}$ with $\partial m/\partial v|_x = 0$. Moreover, AH can be null at points where $\partial m/\partial r = 1/2$, and it can also be timelike or spacelike. Therefore, AH is a dynamical horizon [5,39] (a spacelike marginally trapped tube) on the region where it is spacelike, and an isolated horizon [5] on any open region where $m = m(r)$. This isolated-horizon portion of AH, if nonempty, is characterized also by:

$$G_{\mu\nu}k^\mu k^\nu\big|_{\mathrm{AH}} = 0 \Leftrightarrow \text{isolated-horizon portion of AH}$$

$$\equiv \mathrm{AH}^{\mathrm{iso}}$$

where $G_{\mu\nu}$ is the Einstein tensor of the spacetime. This will be relevant for the perturbations of AH to be studied in the next section. On the other hand, the dynamical horizon portion of AH is *generic* in the sense of [3], and therefore it is actually an outer trapping horizon in the sense of Hayward [17]. In an open region with $m = r/2 + f(v)$, the AH is null, but it is *not* an isolated horizon as the null

normal to AH is not the null normal with vanishing expansion.

A simple illustrative example of these different possibilities is provided by the Robertson-Walker spacetimes: fixing an origin of coordinates arbitrarily, the corresponding spherically symmetric apparent 3-horizon is a marginally trapped tube [3] with the property of being spacelike, null or timelike if $(\varrho + p)(\varrho - 3p + 4\Lambda)$ is less, equal or greater than zero, see pp. 779–780 in [33]—see also [13]—, where ϱ and p are the energy density and pressure, respectively, $(\varrho + \Lambda = 3(\dot{a}^2 + k)/a^2, p - \Lambda = -2\ddot{a}/a - (\dot{a}^2 + k)/a^2$.) A timelike AH is explicitly shown in Fig. 2. An example of a null AH which is not an isolated horizon is provided by the Robertson-Walker spacetime with $p = \varrho/3$ and $\Lambda = 0$ [40]. These results are obviously related to the instability of MOTS in Robertson-Walker geometries proven in [31] if $\varrho + \Lambda > 0$, $\varrho + p \geq 0$, and $\varrho - 3p + 4\Lambda \geq 0$.

Observe that AH can be empty (e.g. in flat spacetime), but this will only happen if there are no marginally f-trapped round spheres on the entire spacetime. As our aim is to study the region with closed f-trapped surfaces and its boundary, we will assume that they certainly exist. In this situation, AH cannot be empty. Under reasonable hypotheses, and for general asymptotically flat initial data sets (so that the cosmological constant $\Lambda = 0$), one then knows [41] that there is a regular complete future null infinity $\mathcal{J}^+$ (so that close to infinity the round spheres are untrapped). The event horizon (EH) is defined as the boundary of the causal past of future null infinity: $\partial J^-(\mathcal{J}^+)$ [1,8,27], hence it is a null hypersurface by definition. In our case, it is also spherically symmetric.

The apparent 3-horizon AH does not need to be connected, and it can have as many connected components as desired, even if $m(\upsilon, r)$ is bounded everywhere by a finite positive mass M. As an elementary example, take the case with $m = m(r)$, $m(r) \leq M < \infty$, so that AH is null everywhere. The connected components of AH are given by the null hypersurfaces $r = r_i$ where r_i are the positive roots of the equation $r - 2m(r) = 0$. In general, we will only be interested in the particular connected component of AH which is related to the event horizon (EH) of an asymptotically flat end (in the example above, the one with the largest r_i). This particular connected component of AH will be denoted by AH_1:

$$\mathrm{AH}_1 \equiv \{\text{Portion of AH that merges with}$$

$$\text{or is asymptotic to the EH}\}$$

The region where the round spheres are untrapped will be denoted by

$$\mathcal{R}_0: r - 2m(\upsilon, r) > 0$$

and we also use the notation

$$\mathcal{R} = \mathcal{R}_0 \cup \mathrm{AH} \ (\Leftrightarrow \mathcal{R}: r - 2m(\upsilon, r) \geq 0).$$

Note that $\mathcal{R}_0$ can actually be empty. For instance, in cases where all the round spheres are f-trapped (this happens, for example, in the Kantowski-Sachs models [34]). However, $\mathcal{R}_0$ can never be empty in the asymptotically flat cases considered herein, because then the round spheres close to spacelike infinity must be untrapped. Notice, similarly, that the whole spacetime may sometimes coincide with $\mathcal{R}$, so that no round sphere is f-trapped, but they still can be marginally f-trapped. An example is provided by the extreme Reissner-Nordström solution which has a degenerate horizon [8,34]. We avoid this situation due to the assumption of the existence of f-trapped spheres.

VII. PERTURBATIONS OF ROUND SPHERES ON AH

We are going to use the stability operator [11,22] to probe the possible perturbations of marginally f-trapped round spheres with the aim, in particular, of ascertaining if there can be closed f-trapped surfaces traversing AH [42]. We will also find a characterization of the AH in terms of the Einstein tensor, as well as other interesting results. Related conclusions were derived in full generality, by the same method of perturbation, in [43]. We will not restrict the causal character of AH, though, and we will also prove that the deformed surfaces can be made genuinely f-trapped while extending to both sides of $\mathrm{AH}\backslash\mathrm{AH}^{\mathrm{iso}}$.

Choose any connected component of AH, so that this is a spherically symmetric marginally f-trapped tube: a hypersurface foliated by marginally f-trapped round spheres (defined by constant values of r and υ). As explained in the previous section, the future null normals are given by (11) with $\vec{k}_{\mathrm{sph}}^- = \vec{\ell}$ and $\vec{k}_{\mathrm{sph}}^+ = \vec{k}$, so that the corresponding expansions were presented in (13), which restricted to AH are:

$$\theta_{\mathrm{sph}}^- = -e^{-\beta}\frac{2}{r} < 0, \qquad \theta_{\mathrm{sph}}^+ = 0.$$

We can now perturb any such marginally f-trapped round sphere, say ς, along a direction $f\vec{n}$ defined on ς and orthogonal to it. Following [22], f is any function defined on ς and we describe $\vec{n}$ by means of the outward vector field

$$\vec{n} = -\vec{\ell} + \frac{n_\mu n^\mu}{2}\vec{k}|_\varsigma \tag{14}$$

which is normalized with respect to the fixed null directions such that

$$k_\mu n^\mu = 1.$$

Observe that the causal character of $\vec{n}$ is unrestricted.

Deform the round sphere ς by going orthogonally to ς along $f\vec{n}$ a distance ϵ. The formula for the variation of the expansion θ^+ can be found in many references, e.g. in [9,11,17,22,23,30,43], but the version better adapted to our purposes is that appearing in [22] because it keeps the

norm of $\vec{n}$ free. Given the marginal character of ς and its spherical symmetry most of the terms in the variation formula vanish and the outer expansion of the perturbed surface ς_ϵ is given by

$$\theta_\epsilon^+ = \epsilon \delta_{f\tilde{n}}\theta^+ + O(\epsilon^2),$$

$$\delta_{f\tilde{n}}\theta^+ = -\Delta_\varsigma f + f\left(\frac{1}{r_\varsigma^2} - G_{\mu\nu}k^\mu u^\nu|_\varsigma\right)$$

(15)

where r_ς is the constant value of r on ς, Δ_ς is the Laplacian on ς, and $\vec{u}$ is the following vector field orthogonal to ς and $\vec{n}$:

$$\vec{u} = \vec{\ell} + \frac{n_\mu n^\mu}{2}\vec{k}, \qquad u_\mu n^\mu = 0, \qquad u_\mu u^\mu = -n^\mu n_\mu.$$

Observe, by the way, that selecting $f = \text{const}$ (15) informs us that the vector $\vec{u}$ such that $1/r_\varsigma^2 - G_{\mu\nu}k^\mu u^\nu|_\varsigma = 0$ produces no variation on θ^+, so that the corresponding $\vec{n}$ is tangent to the AH simply leading to other marginally f-trapped round spheres on AH. Let us call such a vector field $\vec{m}$, so that

$$\frac{1}{r_\varsigma^2} - G_{\mu\nu}k^\mu\ell^\nu|_\varsigma - \frac{m_\rho m^\rho}{2}G_{\mu\nu}k^\mu k^\nu|_\varsigma = 0 \qquad (16)$$

together with

$$\vec{m} = -\vec{\ell} + \frac{m_\mu m^\mu}{2}\vec{k}|_\varsigma$$

characterizes $\text{AH}\backslash\text{AH}^{\text{iso}}$, since $\vec{m}$ is the unique spherically symmetric direction tangent to it. The exceptional isolated-horizon portion AH^{iso} has the null $\vec{k}$ as the tangent vector field. Observe that the only forbidden direction—due to the normalization used—for $\vec{n}$ (and $\vec{m}$) is that defined by $\vec{k}$ (which would correspond to $n_\mu n^\mu \to \pm\infty$). The situation is depicted in Fig. 3.

Whether $\vec{m}$ is spacelike, null or timelike (and accordingly for AH) depends on the magnitude of $G_{\mu\nu}k^\mu\ell^\nu|_\varsigma$ and the sign of $G_{\mu\nu}k^\mu k^\nu|_\varsigma$. This sign is non-negative if the null energy condition is assumed. In the next section, we will actually assume the stronger dominant energy condition, so that for most of our purposes, the positive sign must be kept in mind. Recall, though, that the condition $G_{\mu\nu}k^\mu k^\nu = 0$ defines the portions of AH which are isolated horizons. We have to consider both cases separately.

A. Perturbations on AH^{iso}

Assume that $G_{\mu\nu}k^\mu k^\nu = 0$ holds on a region so that we are dealing with AH^{iso}. This can be seen equivalent to the condition $\partial m/\partial v = 0$. From the variation formula (15) we deduce that

$$\delta_{f\tilde{n}}\theta^+ = -\Delta_\varsigma f + f\left(\frac{1}{r_\varsigma^2} - G_{\mu\nu}k^\mu\ell^\nu|_\varsigma\right)$$

so that the perturbed expansion is *independent of the direction* of deformation $\vec{n}$. One can check that
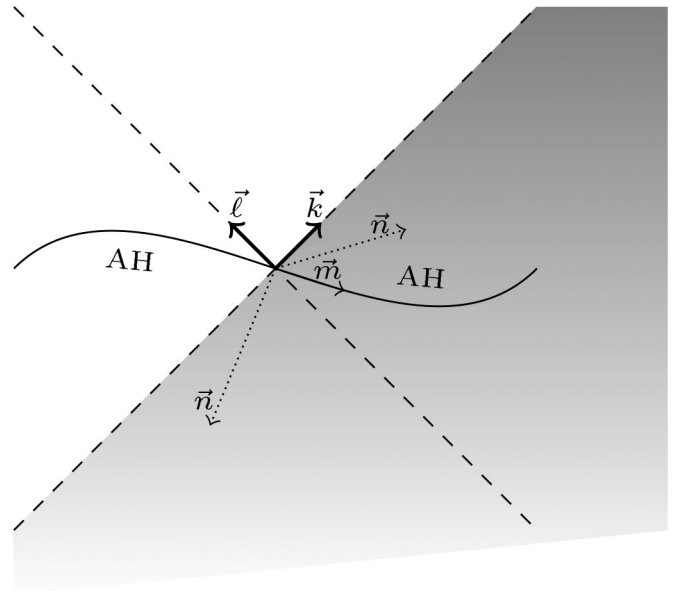

FIG. 3. This is the scheme for the vector fields involved in the perturbation analysis. The spherically symmetric null directions are represented at $45°$, defined by the radial null vector fields $\vec{\ell}$ and $\vec{k}$ given in (11). The vertex at the center represents a marginally f-trapped round sphere in the apparent 3-horizon AH, which is also drawn nearby. The vector $\vec{m}$, as defined in the main text, is tangent to AH. Thus, perturbing the initial round sphere in the direction of $c\vec{m}$ with constant c leads to another marginally f-trapped round sphere in AH. By letting $n_\mu n^\mu \in (-\infty, \infty)$, the general vector $\vec{n}$ defined in (14) always points towards the shadowed region, though its direction may depend on the point of the sphere, and it can be spacelike, null or timelike. The AH can also have any causal character.

$G_{\mu\nu}k^\mu\ell^\nu|_\varsigma = (2/r_\varsigma^2)\partial m/\partial r|_\varsigma$ and the previous relation can be rewritten as

$$\Delta_\varsigma f - f\frac{1}{r_\varsigma^2}\left(1 - 2\frac{\partial m}{\partial r}\bigg|_\varsigma\right) = -\delta_{f\tilde{n}}\theta^+. \qquad (17)$$

Notice that the term between round brackets will generally be positive—for instance if $m = \text{const}$—, and it will certainly be so for any part of AH^{iso} related to an asymptotically flat end, because then $r - 2m(r)$ changes from negative to positive values. Equation (17) can be seen as an equation $L(f) = -\delta_{f\tilde{n}}\theta^+$ where $L = \Delta_\varsigma - (1/r_\varsigma^2)(1 - 2\partial m/\partial r|_\varsigma)$ is an elliptic operator on ς, and thus it is adapted for direct application of the maximum principle, see e.g. [44] or Appendix 3 in [43]. In particular, if $\delta_{f\tilde{n}}\theta^+$ is nonpositive everywhere, it follows that f must be negative everywhere on ς.

Combining this with the known fact that arbitrary perturbations along the null generator $\vec{k}$ of the isolated horizon AH^{iso} produce marginally f-trapped surfaces [43], we obtain the following theorem.

Theorem 7.1.—On any isolated-horizon portion AH^{iso} of AH arbitrary deformations of its round spheres within

AH^{iso} lead to marginally f-trapped surfaces. Moreover, if AH^{iso} is such that $1 \geq 2\partial m/\partial r$, any other possible perturbation leading to weakly f-trapped surfaces has $f < 0$, so that the deformed surfaces lie strictly outside the region $\mathcal{R}$.

Observe that from (17)

$$\left(1 - 2\frac{\partial m}{\partial r}\right)\Big|_{\varsigma}\oint_{\varsigma} f = -\oint_{\varsigma} \delta_{f\vec{n}}\theta^+$$

and, given that the right-hand side can be chosen as small as desired, the minimum of the nonpositive function f can be made as small in magnitude as needed. In concrete situations, one can even use the freedom of choosing the variation vector $\vec{n}$ if this helps.

B. Perturbations on AH\AH^{iso}

Let us now consider the parts of AH with $G_{\mu\nu}k^\mu k^\nu > 0$. From Fig. 3, we deduce that the perturbation along $f\vec{n}$ will enter into the region with f-trapped round spheres (which is trivially part of $\mathcal{T}$ and can be identified because $\vec{k}$ always points into it) at points with

$$f(n_\mu n^\mu - m_\mu m^\mu) > 0.$$

For easy control of these signs we note that, according to (15) and (16),

$$(G_{\mu\nu}k^\mu k^\nu|_{\varsigma})f(n_\mu n^\mu - m_\mu m^\mu) = -2(\Delta_{\varsigma}f + \delta_{f\vec{n}}\theta^+).$$
(18)

An interesting conclusion arises by integrating this equality on ς

$$(G_{\mu\nu}k^\mu k^\nu|_{\varsigma})\oint_{\varsigma} f(n_\mu n^\mu - m_\mu m^\mu) = -2\oint_{\varsigma}\delta_{f\vec{n}}\theta^+$$

from where we deduce the following facts:
  (i) Spherically symmetric deformations, defined by having constant f and $n_\mu n^\mu$, are uninteresting because they only lead to untrapped round spheres in the region $\mathcal{R}_0$ if $f(n_\mu n^\mu - m_\mu m^\mu) < 0$ and to f-trapped round spheres outside $\mathcal{R}$ if $f(n_\mu n^\mu - m_\mu m^\mu) > 0$.
  (ii) The deformed surface can be f-trapped –with a negative definite sign of the variation—only if $f(n_\mu n^\mu - m_\mu m^\mu)$ is positive somewhere. Hence, an f-trapped surface (obtained in this way) must lie at least partially in the region where the round spheres are f-trapped. This will turn out to be a fully general result (Corollary 9.1).
  (iii) The deformed surface can be untrapped only if $f(n_\mu n^\mu - m_\mu m^\mu)$ is somewhere negative.
  (iv) If the deformed surface lies entirely within $\mathcal{R}_0$—so that $f(n_\mu n^\mu - m_\mu m^\mu) < 0$ everywhere—, then $\delta_{f\vec{n}}\theta^+$ must be positive somewhere.
  (v) If the deformed surface lies entirely outside $\mathcal{R}$, then $\delta_{f\vec{n}}\theta^+$ must be negative somewhere.

Note that (15) is also adapted for direct application of the maximum principle, as it takes the form $L(f) = -\delta_{f\vec{n}}\theta^+$ where now the elliptic operator $L = \Delta_{\varsigma} + (G_{\mu\nu}k^\mu k^\nu|_{\varsigma}) \times (n_\mu n^\mu - m_\mu m^\mu)/2$. Thus, we also have
  (i) All possible perturbations with $n_\mu n^\mu - m_\mu m^\mu \leq 0$ and leading to $\delta_{f\vec{n}}\theta^+ \leq 0$ everywhere are such that f is negative everywhere. Thus, all perturbed weakly f-trapped surfaces with $n_\mu n^\mu \leq m_\mu m^\mu$ are strictly outside $\mathcal{R}_0$.

In order to construct examples of f-trapped deformed surfaces which lie partly in $\mathcal{R}_0$, we choose to consider perturbations such that

$$n_\mu n^\mu - m_\mu m^\mu > 0.$$

For this choice, the deformed surface enters the region with f-trapped round spheres if $f > 0$, and it enters $\mathcal{R}_0$ if $f < 0$. We introduce a constant a_0. We will aim for f-trapped surfaces for which

$$(G_{\mu\nu}k^\mu k^\nu|_{\varsigma})a_0(n_\mu n^\mu - m_\mu m^\mu) + 2\delta_{f\vec{n}}\theta^+ = 0.$$

By our assumptions, this implies that $\delta_{f\vec{n}}\theta^+ < 0$ if $a_0 > 0$, so that the deformed surface is f-trapped. Next we set

$$f \equiv a_0 + \tilde{f}$$

for some as yet undetermined function $\tilde{f}$. Equation (18) becomes

$$(G_{\mu\nu}k^\mu k^\nu|_{\varsigma})(n_\mu n^\mu - m_\mu m^\mu)\tilde{f} + 2\Delta_{\varsigma}\tilde{f} = 0. \quad (19)$$

We conclude that our assumptions require that

$$\frac{1}{2}(G_{\mu\nu}k^\mu k^\nu|_{\varsigma})(n_\mu n^\mu - m_\mu m^\mu) = -\frac{\Delta_{\varsigma}\tilde{f}}{\tilde{f}} > 0. \quad (20)$$

This is a (mild) restriction on the function $\tilde{f}$. A simple solution is to choose $\tilde{f}$ to be an eigenfunction of the Laplacian Δ_{ς}, say $\tilde{f} = \sum_{m=-l}^{l} c_{lm}Y_l^m$ for a fixed $l \in \mathbb{N}$ and constants c_{lm}, where Y_l^m are the spherical harmonics. Then, on using $\Delta_{\varsigma}Y_l^m = -\frac{l(l+1)}{r_{\varsigma}^2}Y_l^m$, the deformation direction $\vec{n}$ is determined by

$$n_\mu n^\mu - m_\mu m^\mu = \frac{2}{G_{\mu\nu}k^\mu k^\nu|_{\varsigma}}\frac{l(l+1)}{r_{\varsigma}^2} > 0$$

and the variation of the expansion then reads

$$\delta_{f\vec{n}}\theta^+ = -a_0\frac{l(l+1)}{r_{\varsigma}^2}$$

which is negative (resp. positive) for all $a_0 > 0$ (resp. $a_0 < 0$.) As the other expansion was initially negative, by choosing ϵ very small we can always achieve that $\theta_\epsilon^- = \theta_{sph}^- + \epsilon\delta_{f\vec{n}}\theta^- + O(\epsilon^2)$ is also negative, and therefore the deformed ς_ϵ is f-trapped (resp. untrapped). Throughout we assume that the deformation is small enough so that the we can rely on the first order perturbation.

It only remains to check that f realizes all signs, so that the deformed surface crisscrosses AH. Given that

$$f = a_0 + \sum_{m=-l}^{l} c_{lm} Y_l^m,$$

it is enough to adjust the constants c_{lm} to achieve this goal. For instance, the choice $c_{lm} = 0$ for $m \neq 0$ and $c_{l0} < -a_0$ if $a_0 > 0$ (or $c_{l0} > -a_0$ if $a_0 < 0$) will do, so that f has the sign of a_0 at the region where $Y_l^0 = P_l \leq 0$, and the opposite sign around the north pole of ς where $P_l > 0$ (P_l are the Legendre polynomials).[1]

Thus, we have proven the following theorem.

Theorem 7.2.—In arbitrary spherically symmetric space-times there are closed f-trapped, as well as untrapped, surfaces (topological spheres) penetrating both sides of the apparent 3-horizon AH at any region where $G_{\mu\nu}k^\mu k^\nu|_{\mathrm{AH}} > 0$.

Therefore, any part of AH which is not an isolated-horizon belongs to the f-trapped region $\mathcal{T}$, so that these parts of AH never belong to the boundary $\mathcal{B}$.

We remark that the previous reasoning is independent of the causal character of AH, which can be spacelike, null, or timelike. The only restriction is that $G_{\mu\nu}k^\mu k^\nu|_{\varsigma} > 0$. Observe also that the original round sphere has a positive Gaussian curvature, and thus the deformed f-trapped surfaces penetrating both sides of AH will also have, for sufficiently small ϵ, positive Gaussian curvature. This disproves a conjecture by Hayward [45]. Actually, explicit examples of the same kind but going far away from AH were presented in [46].

We can now address the nonuniqueness of dynamical horizons. The perturbation argument tells us that there are f-trapped surfaces penetrating into both sides of AH\AH$^{\mathrm{iso}}$. We also know that there are untrapped round spheres lying just outside it. If AH is spacelike, this means that we can find a spacelike hypersurface having such an outer trapped sphere as its inner boundary and an untrapped round sphere as its outer boundary, and such that it contains a path connecting the boundaries and lying entirely outside AH (that is, inside $\mathcal{R}_0$). There is a theorem that ensures that such a spacelike hypersurface necessarily contains a marginally (outer) trapped surface [9]. By construction, it has a part lying inside $\mathcal{R}_0$, and we know that it must penetrate outside $\mathcal{R}$. Moreover, generically such a surface "evolves" into a marginally outer trapped tube [11]. As long as we stay sufficiently close to AH, all the marginally outer trapped surfaces in the argument will be inner trapped as well. Thus we have obtained:

Corollary 7.1.—In arbitrary spherically symmetric spacetimes there are marginally trapped tubes penetrating

---

[1]The idea behind the argument just given was communicated to us by R. M. Wald, who informed us that it arose in conversations with G. Galloway [42].

both sides of the apparent 3-horizon AH at any region where $G_{\mu\nu}k^\mu k^\nu|_{\mathrm{AH}} > 0$.

Explicit examples in Robertson-Walker spacetimes can be found in [38], and in the Vaidya spacetime in [47].

As a final question, we wonder how small the fraction of the closed f-trapped-surface that extends outside $\mathcal{R}$ can be made. This will be relevant in Sec. XI, when we will ask the question of whether or not the complement of $\mathcal{R}$ is the optimal set to be removed from spacetime in order to get rid of all closed f-trapped surfaces. With the assumptions used in the proof of Theorem 7.2, we see that this means that we must produce a C^2 function $\tilde{f}$ defined on the sphere and obeying the inequality (20), and which is positive only in a region that we can make arbitrarily small. If we choose a sufficiently small constant a_0, the last requirement implies that the region where the surface extends outside $\mathcal{R}$ can be made arbitrarily small. To find such a function it is convenient to introduce stereographic coordinates $\{\rho, \phi\}$ on the sphere, so that the Laplacian takes the form

$$\Delta_\mathrm{s} = \Omega^{-1}\left(\partial_\rho^2 + \frac{1}{\rho}\partial_\rho + \frac{1}{\rho^2}\partial_\phi^2\right), \qquad \Omega = \frac{4r_\mathrm{s}^2}{(1+\rho^2)^2}.$$

A solution to the problem as stated is

$$\tilde{f}(\rho) = \begin{cases} c_1(e^{(1/2a)(2a-\rho^2)} - 1) & \rho^2 < 4a \\ \frac{8c_1 a}{e}\frac{1}{\rho^2} - c_1(1 + e^{-1}) & \rho^2 > 4a. \end{cases} \quad (21)$$

This function is C^2 (and can be further smoothed if necessary), and it is positive only if $\rho^2 < 2a$, that is on a disk surrounding the origin whose size can be chosen at will. The function obeys

$$-\frac{\Delta_\mathrm{s}\tilde{f}}{\tilde{f}} = \begin{cases} \frac{\Omega^{-1}}{a^2}\frac{2a-\rho^2}{1-e^{-(1/2a)(2a-\rho^2)}} & \rho^2 < 4a \\ \frac{32a\Omega^{-1}}{\rho^4}\frac{\rho^2}{(e+1)\rho^2 - 8a}, & \rho^2 > 4a. \end{cases}$$

This is always larger than zero. Thus we have proven the following important result.

Theorem 7.3.—In spherically symmetric spacetimes, there are closed f-trapped surfaces (topological spheres) penetrating both sides of the apparent 3-horizon AH\AH$^{\mathrm{iso}}$ with arbitrarily small portions outside the region $\mathcal{R}$.

VIII. GENERAL IMPLODING, ASYMPTOTICALLY FLAT, SPHERICALLY SYMMETRIC SPACETIME

In this section, we present the restrictions on the mass function in order to describe the case of inflow of matter and radiation, satisfying the dominant energy condition, entering into an initially flat spacetime and leading to the formation of a black hole.

If the Einstein field equations hold (with vanishing cosmological constant), the dominant energy condition [8] requires, among other restrictions, that the following inequalities hold, e.g. [41,48]:

$$\frac{\partial \beta}{\partial r} \geq 0, \tag{22}$$

$$\frac{\partial m}{\partial r} \geq \frac{r}{2}\left(1 - \frac{2m}{r}\right)\frac{\partial \beta}{\partial r}, \tag{23}$$

$$\frac{\partial m}{\partial v} \geq -e^{\beta}\frac{\partial \beta}{\partial r}\frac{r}{4}\left(1 - \frac{2m}{r}\right)^2. \tag{24}$$

From (22) and (23) one can deduce (see, e.g., [48])

$$\frac{\partial(me^{\beta})}{\partial r} \geq \frac{r}{2}\frac{\partial e^{\beta}}{\partial r} \geq 0$$

which implies that, at any null hypersurface $v = v_c = $ const, the mass function satisfies

$$e^{\beta(v_c, r)}m(v_c, r) \geq e^{\beta(v_c, r_0)}m(v_c, r_0) \quad \forall r \geq r_0$$

so that, if the mass function happens to be positive at any round sphere (v_c, r_0), then it is positive for all round spheres (v_c, r) with $r \geq r_0$. In particular, if the mass function is non-negative at $r = 0$, then it is non-negative everywhere. Note also that, using (22) and (23) implies that

$$\frac{\partial m}{\partial r} \geq 0 \quad \text{on} \quad \mathcal{R}.$$

Similarly, from (22)–(24), we deduce

$$\frac{\partial m}{\partial v} + \frac{1}{2}\left(1 - \frac{2m}{r}\right)e^{\beta}\frac{\partial m}{\partial r} \geq 0$$

at the same region $\mathcal{R}$. These last two expressions can be physically reinterpreted if we note that they are equivalent to

$$\vec{\ell}(m) \leq 0, \qquad \vec{k}(m) \geq 0 \quad \text{on} \quad \mathcal{R} \tag{25}$$

where the null vector fields $\vec{\ell}$ and $\vec{k}$ are given in (11).

In other words, the dominant energy condition implies that the mass function must be nonincreasing (respectively, nondecreasing) along any future-pointing ingoing (resp. outgoing) radial null geodesic. Observe also that the mass function must be nondecreasing along any spacelike outward direction on $\mathcal{R}$ (see e.g. [49]), as follows from

$$a\vec{k}(m) - b\vec{\ell}(m) \geq 0, \qquad \forall a, b > 0 \quad \text{on} \quad \mathcal{R}.$$

Yet another implication of the above conditions is that the hypersurfaces $m(v, r) = $ const are nonspacelike everywhere on $\mathcal{R}$.

Observe also that, on AH, the dominant energy condition (24) implies

$$\left.\frac{\partial m}{\partial v}\right|_{\text{AH}} \geq 0. \tag{26}$$

Only continuous piecewise differentiable mass functions will be considered, so that distributional singularities on the curvature tensor—such as shells of matter or radiation—are avoided. We will restrict ourselves to mass functions bounded by a finite least upper bound $M > 0$, so that $m(v, r) \leq M$ for all v, r, and [41] there is a regular complete future null infinity $\mathcal{J}^+$ for an asymptotically flat end. We shall also restrict ourselves to the physical case where the mass-energy starts flowing in from past infinity at a given advanced time, so that previous to that instant the spacetime has no mass-energy and is flat. The value of v when the mass inflow starts will be chosen as $v = 0$. Then, the mass function satisfies

$$m(v, r) = 0 \quad \forall v < 0;$$

$$\forall v > 0 \begin{cases} 0 \leq m(v, r) \leq M < \infty \\ m(v, r) \not\equiv 0. \end{cases} \tag{27}$$

We will not assume in general, however, that the energy-mass travels at the speed of light, so that the infalling mass can be composed of massive dust particles or more general matter. Therefore, the hypersurface σ separating the flat portion and the rest of the spacetime can be timelike or null. The mass function can actually reach the value M or not. In the former case, given that we are assuming that there is a regular future null infinity $\mathcal{J}^+$ for an asymptotically flat end, there exists a value v_1 of v such that $m(v, r) = M$ for all $v > v_1$. This implies that $m(v_1, r) = M$ for all r. Note that charged cases (such as those with an asymptotic, and static, Reissner-Nordström region) are included in the other case characterized by $m < M$ everywhere.

Whether or not the spacetime becomes singular when the incoming matter reaches the center depends on the particular properties of the falling matter and energy. The intersection of σ with $r = 0$ will *not* be a curvature singularity—so that there will be a regular center $r = 0$ at a portion of the nonflat region—if m and β satisfy there (e.g. [48]):

$$\frac{\partial \beta}{\partial r}(v, 0) = 0;$$

$$m(v, 0) = \frac{\partial m}{\partial r}(v, 0) = \frac{\partial^2 m}{\partial r^2}(v, 0) = 0. \tag{28}$$

In this case, some later singularities can develop. If (28) do not hold, then the singularity appears already at $\sigma(r \to 0)$. At this general level, and in any of the previous cases, one cannot know if the singularity will be spacelike, timelike, or null. Thus, we will not prejudge this, and leave the future evolution of the spacetime open, not showing it in some of the Penrose diagrams. These are depicted, for the essentially different possibilities of interest herein, in Figs. 4–6.

Summarizing, we consider spacetimes with line-element (10) satisfying the dominant energy condition and subject to (27) so that there is an asymptotically flat end with regular $\mathcal{J}^+$ and a nondegenerate EH, and such that AH_1 is the connected component of AH associated to this EH. The actual position of EH depends on the particular properties of the mass function $m(v, r)$. Generically, AH_1 separates the region $\mathcal{R}_1$, defined as the connected subset
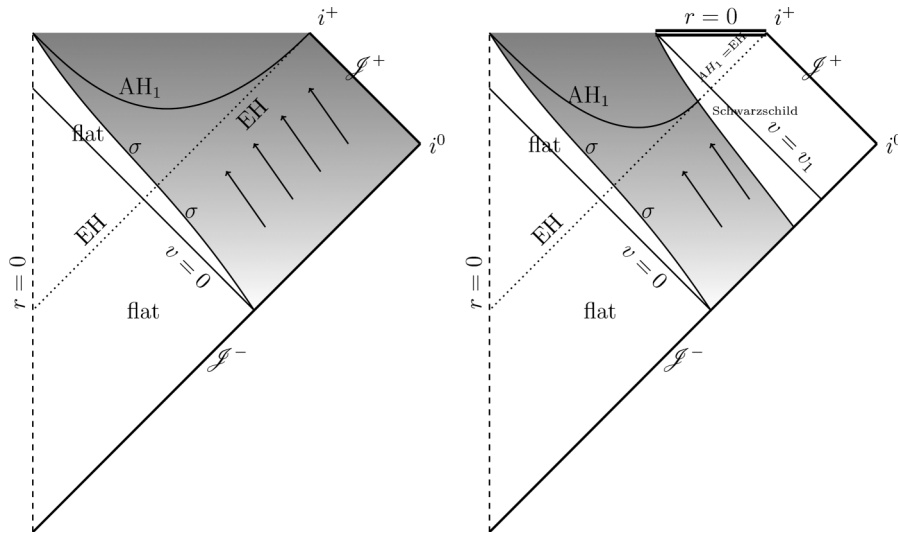
FIG. 4.   These are conformal diagrams of (10) with (29) and (27) when (i) $m(v, r) < M$ everywhere (left) and (ii) $m(v, r) = M$ in some open asymptotic region (right). As usual, null radial lines are drawn at 45° and the future direction is upwards. The discontinuous line marked as $r = 0$ is the origin of coordinates. The spacetime is flat until energy starts falling to the center in a spherical manner from past infinity at $v = 0$. The hypersurface σ separating the flat region from the rest is nonspacelike everywhere, so that material particles travel causally. (The particular case with a null σ is depicted in Fig. 5). Thus, the shaded regions are nonflat spherically symmetric spacetimes with nonvanishing energy-momentum. In the figure on the right, the spacetime becomes Schwarzschild with mass M at some other nonspacelike hypersurface (and therefore also for all $v > v_1$, where v_1 is the supremum of v on that hypersurface). The connected component of the apparent 3-horizon AH_1 approaches the event horizon EH either asymptotically (left) or at some finite value of $v \leq v_1$ and $r = 2M$ (right). In the latter case, AH_1 and EH merge and remain together for $v \geq v_1$ and $r = 2M$. In both cases, EH starts developing in the flat region. In the two cases, AH_1 is spacelike (ergo a dynamical or future outer trapping horizon) when approaching the EH, but this is not necessarily so in other regions. These possibilities are depicted in Fig. 6. Whether or not the initial collapsing shell σ leads to the formation of a singularity when focusing at the center $r = 0$ depends on the particular properties of the mass function $m(v, r)$, see the main text. Thus, we have left open the future evolution of the spacetime for the shaded regions.
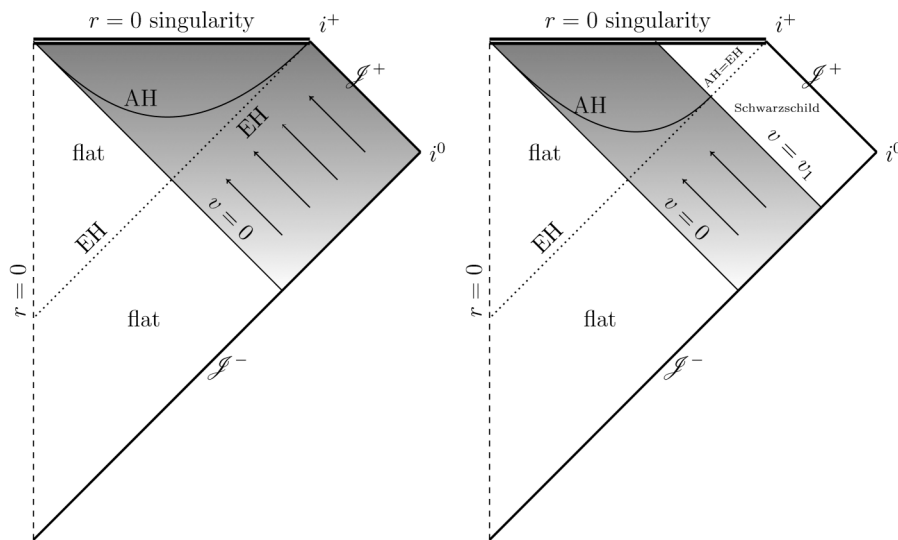

FIG. 5.   These are conformal diagrams similar to those of Fig. 4 but with the energy flowing in along null incoming radial hypersurfaces, so that σ is now the null hypersurface $v = 0$. A particular case of this situation is given by the Vaidya imploding spacetime, analyzed thoroughly in the Appendix. As argued there, the generic situation now leads to the formation of a singularity in the future with $r = 0$, as shown, and therefore there is a unique connected component of AH. The case on the left has $m(v, r) < M$ everywhere and that on the right has $m(v, r) = M$ for all $v > v_1$. In the picture on the left, the apparent 3-horizon AH approaches the event horizon EH asymptotically, while in that on the right, AH and EH merge together at $v = v_1$ and $r = 2M$. In both cases, EH starts developing in the flat region. Note that the part of flat spacetime that lies inside the event horizon is the intersection of the interiors of two light cones; it is shown without conformal distortion.
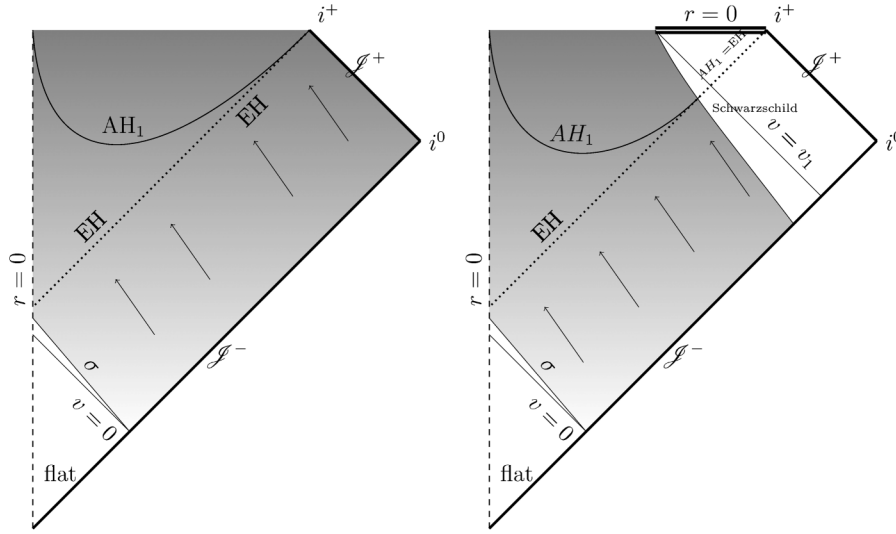
FIG. 6. Once again, these are conformal diagrams of (10) with (29) and (27) when (i) $m(v, r) < M$ everywhere (left) and (ii) $m(v, r) = M$ in some open asymptotic region (right). All features are essentially the same as in Fig. 4 except that EH does not intersect the flat region and that the connected component AH_1 of the apparent 3-horizon may be timelike in some region (in the figures, close to the left upper corner). An explicit example of the latter behavior is provided by the matching of a $p = \gamma\varrho$ Robertson-Walker metric with $\Lambda = 0$ and $\gamma < 1/3$ to the outwardly radiating Vaidya spacetime, see [40] (time reversal of Fig. 7 therein). See also [13].

of $\mathcal{R}_0$ which contains the flat region of the spacetime, from a region containing f-trapped round spheres. Under these assumptions, AH_1 will eventually be spacelike (actually achronal) and asymptotic (probably merging) to the EH, see [50]. (The recent counterexamples presented in [51] violate some of our assumptions.) Thus, AH_1 has a portion that is a spherically symmetric, regular, dynamical horizon. Nevertheless, in general, AH_1 can have timelike and null portions, see e.g. [13,33,40]. This has been represented in the Penrose diagrams of Fig. 6.

Apart from the above, we will need a further assumption, given by

$$\frac{\partial m}{\partial v} \geq 0 \quad \text{on} \quad (\mathcal{R}_1 \cup AH_1) \cap J^+(\text{EH}) \qquad (29)$$

To justify this assumption, and to understand its reasonability, let us make the following considerations. The region $(\mathcal{R}_1 \cup AH_1) \cap J^+(\text{EH})$ may contain a portion of the flat region, where $m(v, r) = 0$ so that the assumption is trivial there, and it is bounded to the future by AH_1, where (29) holds due to (26). The rest of its boundary is given by a portion of EH, whose null generators are given by $\vec{k}$. Observe also that $\vec{k}|_{AH_1} = \partial_v|_{AH_1}$.

If there is part of the origin $r = 0$ in $(\mathcal{R}_1 \cup AH_1) \cap J^+(\text{EH})$, then from (24) we have $\partial m/\partial v|_{r=0} \geq 0$. And if there is a flat portion in $(\mathcal{R}_1 \cup AH_1) \cap J^+(\text{EH})$, this is separated from the rest by the spherically symmetric hypersurface σ, where $m|_\sigma = 0$ by continuity. Let us denote by

$$\vec{w} = \frac{\partial m}{\partial r}\partial_v - \frac{\partial m}{\partial v}\partial_r$$

the vector field tangent to the hypersurfaces $m(v, r) = $ const and orthogonal to the round spheres. Thus, we have

$$\vec{w}(m) = 0.$$

Note that $\vec{w}$ is future-pointing on $\mathcal{R}$. The hypersurface σ is imbedded in flat spacetime, so that the mass-energy is flowing in only if $\vec{w}(r)|_\sigma < 0$. But this implies that

$$\frac{\partial m}{\partial v}\bigg|_\sigma > 0.$$

All in all, the dominant energy condition always ensures that there is a region of $(\mathcal{R}_1 \cup AH_1) \cap J^+(\text{EH})$ where (29) is automatically satisfied, and this includes its future boundary and its frontier with the flat region if nonempty. In summary, the assumption (29) is equivalent to assuming that the mass function is nondecreasing to the future along any hypersurface $r = $ const on $(\mathcal{R}_1 \cup AH_1) \cap J^+(\text{EH})$.

IX. THE PAST BARRIER Σ

In this section, we are going to identify a past barrier for f-trapped surfaces in the spacetimes considered in the previous section. This barrier severely restricts the possible locations of marginally trapped tubes and dynamical horizons.

Consider the vector field

$$\vec{\xi} = e^{-\beta}\partial_v$$

which characterizes the spherically symmetric directions tangent to the hypersurfaces $r = $ const. These hypersurfaces are timelike everywhere on $\mathcal{R}_0$, and null at AH, while

they are spacelike outside $\mathcal{R}$. $\vec{\xi}$ is hypersurface orthogonal, with the level function τ defined by

$$\xi_\mu dx^\mu = -F d\tau = dr - e^\beta\left(1 - \frac{2m(v, r)}{r}\right)dv. \quad (30)$$

The hypersurfaces $\tau = \mathrm{const}$ are orthogonal to the hypersurfaces $r = \mathrm{const}$ everywhere. Put in another way, the expansion of the round spheres along $\vec{\xi}$ vanishes, that is, the mean curvature vector defined in (12) is such that

$$\xi_\mu H^\mu_{\mathrm{sph}} = 0,$$

hence $\vec{\xi}$ provides the invariantly defined direction in which the area of the round spheres remains constant.

We note that $\vec{\xi}$ is the Kodama vector field [24], which has been recently used in related investigations [52–54]. We have

$$\xi_\mu\xi^\mu = -\left(1 - \frac{2m(v, r)}{r}\right), \qquad \ell_\mu\xi^\mu = -e^{-\beta}$$

so that $\vec{\xi}$ is future-pointing timelike on the region $\mathcal{R}_0$, future-pointing null at AH, and spacelike outside $\mathcal{R}$. Therefore, τ can considered as a time function—"the Kodama time" [46,52]—in the whole region $\mathcal{R}_0$. Observe that

$$\vec{\xi} = e^{-\beta}\vec{k} + \frac{e^\beta}{2}\left(1 - \frac{2m(v, r)}{r}\right)\vec{\ell}.$$

The deformation of the metric along $\vec{\xi}$ can be easily computed

$$(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu} = e^\beta\frac{2}{r}\frac{\partial m}{\partial v}\ell_\mu\ell_\nu - \frac{\partial\beta}{\partial r}(\delta^r_\mu\xi_\nu + \delta^r_\nu\xi_\mu). \quad (31)$$

Thus, $\vec{\xi}$ is a Kerr-Schild vector field [35] on any open region with $\partial\beta/\partial r = 0$, an example is the Vaidya spacetime treated in the Appendix. $\vec{\xi}$ is a Killing vector in the situations with $\partial m/\partial v = \partial\beta/\partial r = 0$, such as the case of the Schwarzschild spacetime.

We can use the Kodama time to restrict the location of f-trapped surfaces.

Theorem 9.1.—Assume that the spacetime (10) satisfies (29). Then, no f-trapped surface S can have a local minimum of τ on $\mathcal{R}_0$, nor they can have an open portion with $\tau = \mathrm{const}$ there.

Proof.—Let $q \in S \cap \mathcal{R}$ be a point where S has a local minimum of τ, or belonging to an open portion of $S \cap \{\tau = \tau_0\}$ for some constant τ_0. Because of Theorems 4.1 and 4.2, and since $\vec{\xi}$ is future-pointing on $\mathcal{R}$, it is enough to show that on any such point

$$P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu}|_q \geq 0.$$

Projecting (31) onto S we get

$(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu}e^\mu_A e^\nu_B$

$$= e^\beta\frac{2}{r}\frac{\partial m}{\partial v}\bigg|_S \bar{\ell}_A\bar{\ell}_B - \frac{\partial\beta}{\partial r}\bigg|_S\left(\frac{\partial\bar{r}}{\partial\lambda_A}\bar{\xi}_B + \frac{\partial\bar{r}}{\partial\lambda_B}\bar{\xi}_A\right).$$

In particular, given that $\bar{\xi}_A|_q = 0$, we obtain

$$P^{\mu\nu}(\mathcal{L}_{\vec{\xi}}g|_S)_{\mu\nu}|_q = e^\beta\frac{2}{r}\frac{\partial m}{\partial v}\bar{\ell}_A\bar{\ell}^A\bigg|_q \geq 0 \quad (32)$$

whose non-negativity follows from (29).

By taking into account the third remark to Theorem 4.1, the same conclusion holds for weakly f-trapped surfaces unless the exceptional situation (7) occurs.

Corollary 9.1.—If the spacetime (10) satisfies (29), then no *closed* f-trapped surface can be contained in any connected component of $\mathcal{R}$. In particular, no *closed* f-trapped surface can be fully contained in the region $\mathcal{R}_1 \cup \mathrm{AH}_1$.

The only weakly f-trapped surfaces contained in $\mathcal{R}$ are the marginally f-trapped surfaces foliating AH. Combining this corollary with the standard result [8,27]—see [55] for a rigorous derivation—that no closed weakly f-trapped surface can penetrate outside the EH (actually, no *outer* f-trapped closed surface penetrates into this region [8,27]), we arrive at the following conclusion.

Corollary 9.2.—Letting aside the marginally f-trapped surfaces in AH_1, no closed weakly f-trapped surface can be fully contained in the region $(\mathcal{R}_1 \cup \mathrm{AH}_1) \cap J^+(\mathrm{EH})$, and thus they must penetrate outside $\mathcal{R}_1 \cup \mathrm{AH}_1$.

Observe that, in the cases when AH_1 is spacelike so that it is a dynamical horizon, and recalling that then EH is its past Cauchy horizon [8,27], $\mathrm{EH} = H^-(\mathrm{AH}_1)$, Corollary 9.2 can be rephrased as "no closed weakly f-trapped surface can be fully contained in the past Cauchy development $D^-(\mathrm{AH}_1)$", which is in agreement with theorem 4.1 in [3].

Suppose that $r < 2m(v, r)$ somewhere to one side of AH_1. We are naturally led to the question of what is the extension of the connected f-trapped region $\mathcal{T}_1$ containing the f-trapped round spheres to that side of AH_1. Equivalently, the question is what is the exact location of the connected component $\mathcal{B}_1$ of the boundary $\mathcal{B}$ which is to the past of AH_1. At first, one is tempted to think that AH_1 could actually be this boundary $\mathcal{B}_1$, so that all f-trapped closed surfaces remain outside of $\mathcal{R}_1 \cup \mathrm{AH}_1$, but we already know that this is not the case in general, as follows from Theorem 7.2. Actually, f-trapped surfaces with spherical topology were explicitly exhibited in [16] for the self-similar Vaidya spacetime (see Appendix) such that they penetrate $\mathcal{R}_1$ and even extend to the flat region of the spacetime. The example from [16] is shown in Fig. 8 below, other examples were constructed in [46]. Thus, the connected f-trapped region $\mathcal{T}_1$ will enter into $\mathcal{R}_1 \cap J^+(\mathrm{EH})$. One can then wonder if $\mathcal{T}_1$ will actually extend all the way down to EH. This was shown to be impossible in a particular Vaidya solution with a shell of null dust in [15]. We are going to prove in the following that this is a

general property by identifying a past barrier to the connected f-trapped region $\mathcal{T}_1$.

Put by definition

$$\Sigma \equiv \{\tau = \tau_\Sigma\}, \qquad \tau_\Sigma \equiv \inf_{x \in AH_1} \tau|_x.$$

Observe that τ_Σ is the least upper bound of τ on the event horizon EH. In other words, τ_Σ is either; (i) the constant value of τ which defines the portion of the EH in the Schwarzschild region of the spacetime in the case when $m(v, r) = M$ for all $v > v_1$, or (ii) the common limit of τ on both AH and EH when $v \to \infty$ in the case that $m(v, r) < M$ everywhere. Σ can be completely characterized as the *last* hypersurface orthogonal to $\vec{\xi}$ which is nontimelike everywhere.

Theorem 9.2.—Assume that the spacetime (10) satisfies the dominant energy condition, (27) and (29). Then, no closed f-trapped surface can penetrate into $I^-(\Sigma)$ (i.e. the region with $\tau < \tau_\Sigma$.)

Proof.—Consider the closed set

$$\mathcal{K} \equiv \overline{D^-(\tau = \tau_\Sigma)} = \overline{J^-(\tau = \tau_\Sigma) \cap J^+(EH)}$$

This set is contained in the region $\mathcal{R}_1 \subset \mathcal{R}$ where $\vec{\xi}$ is future pointing. $\mathcal{K}$ is bounded to the future by $\tau = \tau_\Sigma$ and to the past by EH\AH (see Fig. 7). Therefore, any compact surface S such that $S \cap \text{int} \mathcal{K} \neq \varnothing$ will reach a minimum on $\mathcal{K}$. This minimum cannot be on $\tau = \tau_\Sigma$, because this is the maximum value of τ on $\mathcal{K}$. Thus it will have to be either a nonlocal minimum on EH\AH or a local one
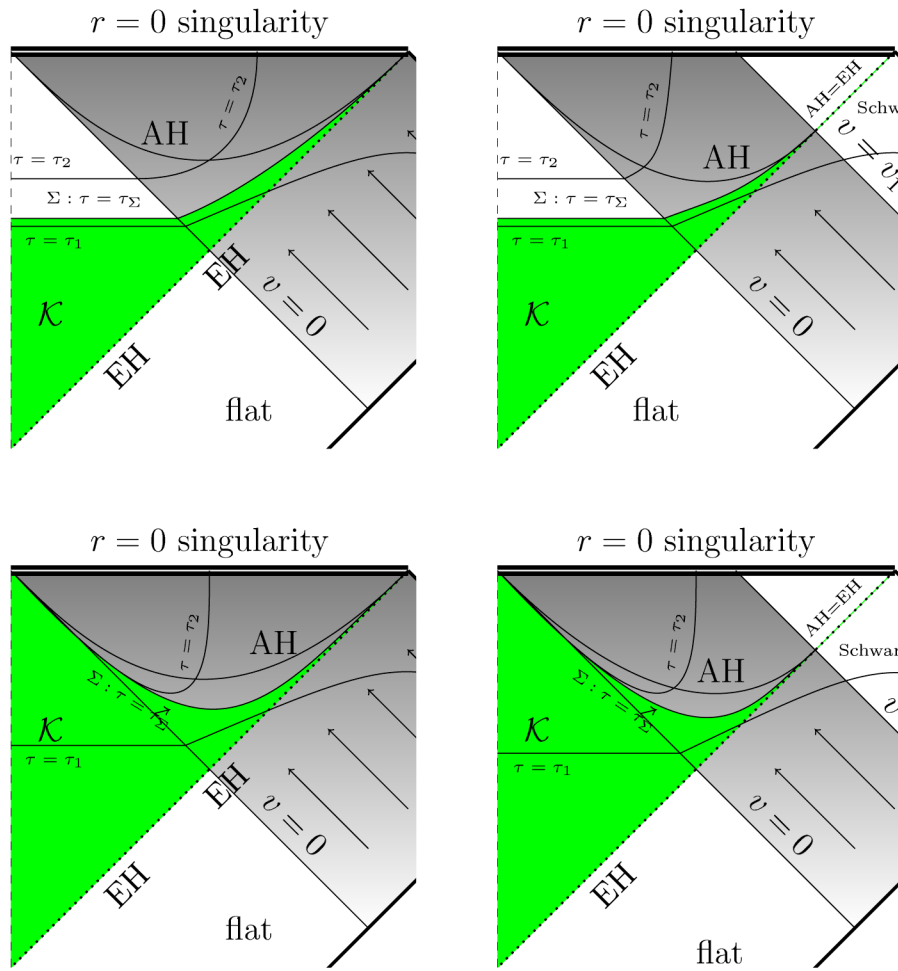

FIG. 7 (color online). These are enlargements of relevant regions in Fig. 5; similar drawings could be performed for the cases of Figs. 4 and 6. On the left, we picture the cases with $m(v, r) < M$ everywhere, and on the right with $m(v, r) = M$ for all $v > v_1$. The two pictures on top describe the cases when Σ penetrates the flat region, for instance if $8\dot{m}_0 > (1 - 2m_0')^2$, see Sec. IX A, while the two bottom pictures depict the cases where Σ never penetrates the flat region. Pertinent $\tau = $ const hypersurfaces are shown for all cases. Observe that these are spacelike everywhere (and approaching spacelike infinity i^0) if they have $\tau < \tau_\Sigma$ (example shown as $\tau = \tau_1$), while they are partly spacelike and partly timelike, becoming null at AH, if $\tau > \tau_\Sigma$ (example shown as $\tau = \tau_2$.) The hypersurface Σ separating these two cases is nontimelike everywhere, and spacelike where it differs from EH: thus, it is spacelike everywhere for the left drawings, while it becomes null and identical with EH for all $v > v_1$ for the cases on the right. The closed set $\mathcal{K}$ used in the proof of Theorem 9.2 limits to the future with Σ and to the past with EH, and is shown in green in all cases.

attained on $\mathcal{K} \cap \mathcal{R}_1$. However, these two possibilities forbid that S be f-trapped, because the latter would contradict Theorem 9.1, while the former would contradict the standard result [8,15,27,55] that closed f-trapped surfaces never touch EH.

Thus, the hypersurface Σ is a limit, to the past, for f-trapped closed surfaces. In fact, they cannot even touch Σ.

Theorem 9.3.—Under the assumptions of the previous theorem all closed f-trapped surfaces must be contained in the region $I^+(\Sigma)$ (defined by $\tau > \tau_\Sigma$) and penetrate outside $\mathcal{R}_1 \cup \mathrm{AH}_1$.

Proof.—The last part of the theorem states that all f-trapped closed surfaces have points outside $\mathcal{R}$, that is points with $r < 2m(v, r)$, but this is already known from Corollary 9.2. To prove the first part, observe that Theorem 9.2 ensures that $\tau \geq \tau_\Sigma$ on any such f-trapped closed surface S. Thus, we only need to show that τ cannot reach the value τ_Σ on the surface, so that S can never touch Σ. But S cannot touch the portion of Σ which coincides with EH (if any), so this could only happen on the part of Σ within $\mathcal{R}_1$, that is, with $r > 2m(v, r)$. But if there were a point $x \in S \cap \mathcal{R}_1 \cap \Sigma$, then Theorem 4.1 would imply that $\tau|_S$ cannot have a local minimum on x, and that there cannot be any one-dimensional line $L \subset S$ with $L \ni x$ such that $\tau|_L = \tau_\Sigma$; Theorem 4.1 would imply that no two-dimensional piece of S can be entirely contained in Σ. In summary, the existence of x would lead to the existence of points on S with $\tau < \tau_\Sigma$, in contradiction.

A. The location of the past barrier Σ

As we have shown, the hypersurface Σ provides a strict limitation to the extension, towards the past, of the connected f-trapped region $\mathcal{T}_1$, and thereby to the location of its boundary $\mathcal{B}_1$. Therefore, it is important to know the exact location of Σ. This is the goal of this subsection.

Recall that Σ is nontimelike everywhere, and actually spacelike on the entire portion where Σ does not coincide with the EH. The location of Σ depends, as is to be expected, on the particular properties of the mass function $m(v, r)$. However, one can deduce general properties of the hypersurfaces $\tau = \mathrm{const.}$ (One may also consult the Appendix where the specific case of the imploding Vaidya spacetime is treated in full, as then the equations can be explicitly integrated—for appropriate choices of the mass function.) In particular, we are going to answer partially the question of whether or not Σ can penetrate into the flat region of the spacetime. To that end, recall the definition (30) of τ, so that the $\tau = \mathrm{const}$ hypersurfaces are given by the solutions to the ordinary differential equation (ODE)

$$\frac{dv}{dr} = \frac{e^{-\beta}}{1 - 2m(v, r)/r}. \tag{33}$$

Let $v_0 \geq 0$ be the value of $v|_\Sigma$ at $r = 0$. The ODE (33) will not have any critical point at $(v = v_0^+, r = 0)$ whenever

$$\lim_{r \to 0} \frac{m(v_0, r)}{r} = 0. \tag{34}$$

In particular, this will be the case when there is no curvature singularity at $(v_0, 0)$, due to (28).

On the other hand, if (34) does not hold, the ODE (33) is equivalent to the autonomous system

$$\frac{dv}{du} = r, \qquad \frac{dr}{du} = e^\beta (r - 2m(v, r))$$

which has a critical point at $v = v_0$, $r = 0$. Its linear stability is ruled by the eigenvalues $\lambda_\pm$ and eigenvectors $\vec{z}_\pm$ of the corresponding matrix

$$\begin{pmatrix} 0 & 1 \\ -2\dot{m}_0 e^{\beta_0} & e^{\beta_0}(1 - 2m_0') \end{pmatrix},$$

where β_0, $\dot{m}_0$ and m_0' are the limits of β, $\partial m/\partial v$ and $\partial m/\partial r$ when approaching $(v_0^+, 0)$, respectively. These eigenvalues and eigenvectors are

$$\lambda_\pm = \frac{e^{\beta_0}}{2} \left(1 - 2m_0' \pm \sqrt{(1 - 2m_0')^2 - 8\dot{m}_0}\right),$$

$$\vec{z}_\pm = (1, \lambda_\pm).$$

The character of the critical point is different depending on the sign of $8\dot{m}_0 - (1 - 2m_0')^2$. The different possibilities are:

(i) If $8\dot{m}_0 > (1 - 2m_0')^2$, the critical point is a focus if $1 - 2m_0' \neq 0$—unstable or stable depending on whether $1 - 2m_0'$ is positive or negative. In this case, no solution reaches $(v_0, 0)$. The hypersurface Σ always penetrates the flat region in this case. See the illustrative explicit solution for the Vaidya case in the Appendix. If on the other hand $2m_0' = 1$, then the critical point is a center at the linear level, and can become a focus, a node, or remain as a center depending on the properties of $m(v, r)$.

(ii) If $0 < 8\dot{m}_0 < (1 - 2m_0')^2$, the critical point is a node, unstable or stable depending on whether $1 - 2m_0'$ is positive or negative. All possible solutions except one emerge from or approach $(v_0, 0)$ with the same tangent direction, given by the eigenvector $(1, \lambda_-)$ in the unstable case or by $(1, \lambda_+)$ in the stable one. The exception for each case is given by one solution emerging from or approaching $(v_0, 0)$ with the tangent direction of the other eigenvector. There exist three qualitatively different possibilities (with the same values of $\dot{m}_0$ and m_0'), depending on whether or not the special solutions $\tau = \tau_\pm$ —corresponding to the eigenvalues $\lambda_\pm$ at $(v_0, 0)$—eventually meet AH_1. This, in turn, depends on the specific properties of the mass function and on the total mass M. If at least one of the special solutions does *not* meet AH_1, then the hypersurface Σ cannot penetrate the flat region. On the other hand, if both special solutions meet AH_1, then Σ will have a portion in the flat region. Explicit illustrative cases are given in the Appendix for the Vaidya spacetime.

(iii) If $8\dot{m}_0 = (1 - 2m'_0)^2 \neq 0$, the critical point is a degenerate node (unstable or stable depending on whether $1 - 2m'_0$ is positive or negative) in the linear stability analysis, and it remains as such, or it may become an unstable focus or node, depending on the specific properties of the mass function. Its properties are once more analogous to those of the Vaidya example in the Appendix.

X. ON THE REGION $\mathcal{T}$ AND ITS BOUNDARY $\mathcal{B}$

In this section, we want to discuss the possible extension of the connected f-trapped region $\mathcal{T}_1$ associated to AH_1, and the relation between its boundary $\mathcal{B}_1$ as defined in definition 2 with marginally trapped tubes and closed weakly f-trapped surfaces.

Having identified the past barrier Σ for the connected f-trapped region $\mathcal{T}_1$, we can ask whether closed f-trapped surfaces can actually extend all the way down to Σ, in other words, if Σ coincides with the connected component $\mathcal{B}_1$ of the boundary. This turns out not to be the case (Corollary 10.2.)

We already know that the region outside $\mathcal{R}$, such that $r < 2m(v, r)$, belongs to $\mathcal{T}$. Now we collect some important properties of how closed f-trapped surfaces can cross AH_1 penetrating into $\mathcal{R}_1$.

Theorem 10.1.—Assume that the spacetime (10) satisfies the dominant energy condition, (27) and (29). Any closed f-trapped surface S crossing AH_1 attains the minimum value τ_m of τ outside $\mathcal{R}_1 \cup AH_1$, and has

$$\tau|_{S \cap \mathcal{R}_1} > \hat{\tau}_m > \tau_m > \tau_\Sigma, \qquad r|_{S \cap \mathcal{R}_1} < \hat{r}$$

where $\hat{\tau}_m$ is the minimum value of $\tau|_S$ on AH_1, and $\hat{r}$ is the value of r at the round sphere $\hat{\mathfrak{s}} \equiv \{\tau = \hat{\tau}_m\} \cap AH_1$. (Let us note that $\hat{\mathfrak{s}} \in AH_1 \backslash EH$.)

Proof.—As S is compact, it must attain a minimum of τ which is also a local minimum unless $\tau|_S = $ const. This last possibility is not feasible for S entering into $\mathcal{R}_1$ due to Theorem 4.2 and the non-negativity of $P^{\mu\nu}(\mathcal{L}_{\bar{\xi}}g|_S)_{\mu\nu}|_S$ (if $\bar{\xi}_A = 0$), as follows from (32). For the former possibility, Theorem 4.1 ensures that the local minimum cannot lie on $\mathcal{R}$, so that τ_m has to be attained outside $\mathcal{R}_1 \cup AH_1$.

Pick up any value $\tau_2 > \tau_m > \tau_\Sigma$. As in the proof of Theorem 9.2 consider the closed set

$$\mathcal{K}_2 \equiv \overline{J^+(EH) \cap \{\tau \leq \tau_2\} \cap (R_1 \cup AH_1)}.$$

$\mathcal{K}_2$ is bounded to the future partly by $\tau = \tau_2$ and partly by $AH|_{\tau \leq \tau_2}$, and to the past by $EH \backslash AH_1$, see Fig. 7. Therefore, if $S \cap \text{int}\,\mathcal{K}_2 \neq \varnothing$, $\tau|_{S \cap \mathcal{K}_2}$ will reach a minimum $\hat{\tau}_m$ on $\mathcal{K}_2$. As usual, this minimum cannot be on $EH \backslash AH_1$ due to the standard result [8,15,27,55] that S can never touch EH; it cannot be on $\tau = \tau_2$ because this is the maximum value of τ on $\mathcal{K}_2$; and it cannot be on $\mathcal{R}_1 \cap \mathcal{K}_2$ either, due to Theorem 4.1. Therefore, such a minimum has to be attained on AH_1. Besides, there cannot be any point $x \in S \cap \mathcal{R}_1$ such that $\tau|_x = \hat{\tau}_m$, because this would

contradict either Theorem 4.1 or Theorem 4.2. Thus, $\hat{\tau}_m < \tau|_{S \cap \mathcal{R}_1}$.

Consider now $\hat{r}$, and observe that it is the maximum value of r on $S \cap AH_1$, because $\tau|_{AH}$ is a monotonically decreasing function of r on AH as follows from the definition (30). To prove the result, recall that the hypersurfaces $\tau = $ const and $r = $ const are orthogonal everywhere, the former are spacelike and the latter are timelike on $\mathcal{R}_0$, and they both become null—and tangent—at AH. Therefore, the hypersurface $\tau = \hat{\tau}_m$ is to the future of all hypersurfaces $r = r_c \geq \hat{r}$ everywhere on $\mathcal{R}_1$. Thus, if S reached a value of $r \geq \hat{r}$ at a point $x \in S \cap \mathcal{R}_1$, then x would be to the past of $\tau = \hat{\tau}_m$, that is, $\tau|_x < \hat{\tau}_m$, which is impossible.

In the case that AH_1 is spacelike—a dynamical horizon—, Theorem 4.3 in [3] implies that no closed f-trapped S can penetrate into the region $J^-(AH|_{r > \hat{r}})$. Theorem 10.1 provides a stricter restriction, independently of the causal character of AH_1, since S cannot penetrate $J^-(\{\tau = \hat{\tau}_m\} \cap \mathcal{R}_1) = \{\tau \leq \hat{\tau}_m\} \cap \mathcal{R}_1$. This is graphically explained in Fig. 8.

Proposition 10.1.—Assume that the spacetime (10) satisfies the dominant energy condition, (29) and (27), and it has f-trapped round spheres to one side of AH_1. Then, the connected component $\mathcal{B}_1$ cannot have a positive minimum value of r, and furthermore

$$\tau_\mathcal{B} \equiv \inf_{x \in \mathcal{B}} \tau|_x = \inf_{x \in \mathcal{B}_1} \tau|_x = \tau_\Sigma.$$

Proof.—That $\tau_\mathcal{B} = \inf_{x \in \mathcal{B}_1} \tau|_x$ is obvious, as the spacetime is flat for $v < 0$ so that there cannot be f-trapped surfaces penetrating the past of $\mathcal{B}_1$. To see that $\tau_\mathcal{B} = \tau_\Sigma$, we first note that $\tau_\mathcal{B} < \tau_\Sigma$ is impossible, as otherwise there would be f-trapped closed surfaces penetrating the region to the past of Σ contradicting Theorem 9.2.

If $\tau_\mathcal{B} > \tau_\Sigma$, then $\mathcal{B}_1$ would be fully contained in the region $\tau \geq \tau_\mathcal{B}$. But this would mean, due to Properties 2 and 3, that the region defined by $\tau < \tau_\mathcal{B}$ would be either part of $\mathcal{T}$, or completely external to it. However, this is again impossible because the part of this region with $\tau \in (\tau_\Sigma, \tau_\mathcal{B})$ has f-trapped round spheres outside $\mathcal{R}_1 \cup AH_1$, and no closed f-trapped surface can penetrate its part with $\tau \leq \tau_\Sigma$.

The same reasoning serves to prove that

$$\inf_{x \in \mathcal{B}_1} r|_x = 0.$$

For, if this infimum were positive, say $r_a < 2M$, it would follow that $\mathcal{B}_1$ would be fully contained in the region $r \geq r_a > 0$. But there are f-trapped round spheres for *all* values of $r \in (0, 2M)$ outside $\mathcal{R}_1 \cup AH_1$.

Proposition 10.2.—Under the same assumptions

$$\mathcal{B}_1 \subset (\mathcal{R}_1 \cup AH_1) \cap J^+(\Sigma),$$

and $\mathcal{B}_1$ merges with, or approaches asymptotically, Σ, AH_1 and EH in such a way that $\mathcal{B}_1 \cap AH_1 = \varnothing$ at any portion of
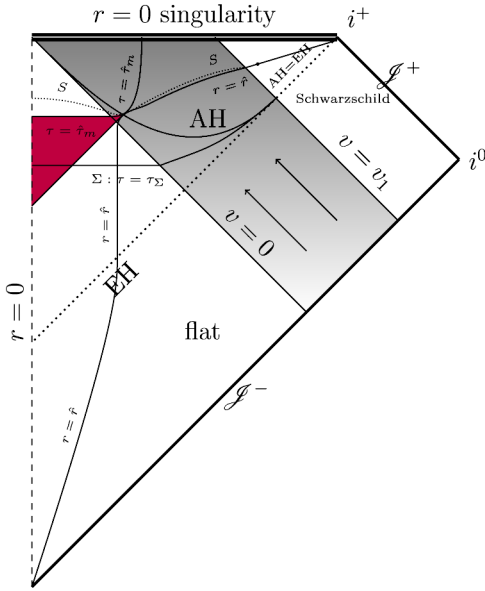
FIG. 8 (color online).   A Penrose diagram —corresponding to the case on the right of Fig. 5, but any other choice would be similar—showing a nonspherically symmetric closed f-trapped surface S penetrating into the flat region of the spacetime. Such f-trapped surfaces were explicitly constructed in [16]. S is shown as a dotted line emerging into the Schwarzschild part as a finite part of a line of constant r. The intersection of S with AH occurs at $r = \hat{r}$ and $\tau = \hat{\tau}_m$, using the notation of Theorem 10.1. The hypersurfaces $r = \hat{r}$ and $\tau = \hat{\tau}_m$ are explicitly shown. They meet at AH where both of them become null and then change their causal character. Theorem 4.3 in [3] implies that S cannot penetrate into the region $J^-(\text{AH}|_{r>\hat{r}})$. However, Theorem 10.1 provides an improvement on that restriction, since S cannot actually penetrate into the region with $\tau \leq \hat{\tau}_m$ below AH. Thus, the region shown in purple, allowed in principle by the former restriction, becomes a forbidden region for S.

AH$_1$ with $G_{\mu\nu}k^\mu k^\nu|_{\text{AH}_1} > 0$. Furthermore, $\mathcal{B}_1$ cannot be nonspacelike everywhere.

Remark.—In the case that v_1 is finite and the mass function is constant for $v > v_1$, the EH and AH$_1$ coincide for all $v > v_1$, and so does $\mathcal{B}_1$. This portion of AH is an isolated horizon with $m = M = $ constant, and $G_{\mu\nu}k^\mu k^\nu = 0$ on that portion. However, there can be other portions of AH$_1$ with $G_{\mu\nu}k^\mu k^\nu = 0$ such that they are isolated horizons. This happens if $m = m(r) < M$ (possibly constant) for some interval $v \in (v_2, v_3)$ with $v_3 < v_1$. Physically, this means that the inflow of matter and radiation stops between v_2 and v_3, and then it starts again. In principle, $\mathcal{B}_1$ may coincide with AH$_1$ on these particular portions of AH$_1$ of isolated-horizon type. This is represented in Fig. 9. It will be useful to have a name for these portions, so that we set:

$$\text{AH}_1^{(\text{iso})} \equiv \text{AH}_1 \cap \text{AH}^{(\text{iso})}.$$

Note that EH will belong to AH$_1^{(\text{iso})}$ if v_1 is finite and the mass function reaches the value M. Also, that τ is constant
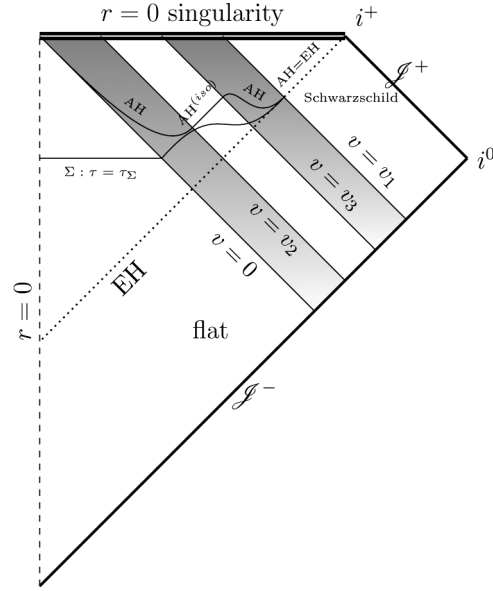

FIG. 9.   Conformal diagram for a particular example where the AH has a nonvanishing portion of isolated-horizon type different from EH. The energy flows in along null incoming radial hypersurfaces, starting at $v = 0$ but suddenly stopping at $v = v_2$. There is no inflow of energy between v_2 and v_3, and then the energy enters again from past null infinity until the radiation stops at $v = v_1$, and then forever. A particular case of this situation is given by the Vaidya imploding spacetime (see Appendix) with $m(v) = $ constant for $v \in (v_2, v_3)$. In the picture, the AH is spacelike for $v \in (0, v_2) \cup (v_3, v_1)$, but it has two portions which are isolated horizons, one given by part of EH (for $v \geq v_1$) and the other one for $v \in (v_2, v_3)$, represented by AH$^{(\text{iso})}$. The past barrier Σ merges with AH and EH as usual. The boundary $\mathcal{B}$ will also merge with them at $r = 2M$ and $v = v_1$, and it must be strictly below AH everywhere on AH\AH$^{(\text{iso})}$, but it is not guaranteed that it cannot touch, or even coincide with, AH$^{(\text{iso})}$.

on AH$_1^{(\text{iso})}$, as AH$_1^{(\text{iso})}$ is defined by portions of $r = $ constant hypersurfaces within AH$_1$ which are null. Observe that the condition $G_{\mu\nu}k^\mu k^\nu \neq 0$—required also in the perturbations of Sec. VII—becomes $G_{\mu\nu}k^\mu k^\nu > 0$, if the dominant energy condition holds.

Proof.—There are f-trapped closed surfaces outside $\mathcal{R}_1 \cup \text{AH}_1$, but due to Theorem 9.2 there are none penetrating $J^-(\Sigma)$. Thus, $\mathcal{B}_1 \subset (\mathcal{R}_1 \cup \text{AH}_1) \cap J^+(\Sigma)$. However, $\mathcal{B}_1$ cannot meet AH$_1$\AH$_1^{(\text{iso})}$ according to Theorem 7.2. As Σ and AH$_1$ merge together, or approach each other asymptotically, so does $\mathcal{B}_1$. Finally, consider the domain of dependence $D(\Sigma)$ of Σ. From the previous observations, $\mathcal{B}_1 \subset D^+(\Sigma)$. But $D(\Sigma)$ is globally hyperbolic with Σ as a Cauchy hypersurface, therefore if $\mathcal{B}_1$ were nonspacelike everywhere, it would have to cross Σ [8,27,32,33], in contradiction with the fact that $\mathcal{B}_1 \subset J^+(\Sigma)$.

As in the case of Robertson-Walker spacetimes (Result 5.4, Sec. V) we derive the following important result.

Theorem 10.2.—Under the same assumptions, $\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$ cannot be a marginally trapped tube, let alone a dynamical or trapping horizon. Furthermore, $\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$ does not contain any nonminimal closed weakly f-trapped surface.

Proof.—From the previous corollary, $\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$ is contained in the region $\mathcal{R}_1$. But there are no closed weakly f-trapped surfaces completely contained in $\mathcal{R}_1$ due to Lemma 4.1.

Thus, the only closed marginally f-trapped surfaces that can be contained in $\mathcal{B}_1$ are those which are actually on its part $\mathcal{B}_1 \cap \text{AH}_1^{(\text{iso})}$, if any. In fact, this property could have been deduced more easily from Result 6.1, as $\mathcal{B}_1$ is a spherically symmetric hypersurface. We have decided to include the alternative proof as it may probably be generalized to non-spherically-symmetric cases.

Theorem 10.2 implies that the notion of "limit section" in [17] (a spacelike 2-surface in $\mathcal{B}$ arising as the uniform limit of a sequence of trapped surfaces approaching $\mathcal{B}$, definition in p. 6473) is generically nonexistent or ill defined. Thus the assumptions of theorem 7 in [17] are very rarely met.

Proposition 10.3.—Assume that the space-time (10) satisfies the dominant energy condition, (29) and (27), and has f-trapped round spheres to one side of AH_1. Then, τ is a nonincreasing function of r on any portion of the connected component $\mathcal{B}_1$ which is locally to the past of $\mathcal{T}_1$. And it is actually strictly decreasing at least somewhere on $\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$.

Remark.—$\mathcal{B}_1$ is a connected component of the boundary $\mathcal{B}$ and, due to Property 3, $\mathcal{B}_1$ has two sides, one with and another without closed f-trapped surfaces. Thus, by "locally to the past" we mean that the Kodama vector field points towards $\mathcal{T}_1$ at $\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$—at $\mathcal{B}_1 \cap \text{AH}_1^{(\text{iso})}$ it is tangent.

Proof.—If $\mathcal{B}_1$ coincides partly with $\text{AH}_1^{(\text{iso})}$ the result is trivial there, as τ and r are constant on $\text{AH}_1^{(\text{iso})}$. Suppose then that $\tau|_{\mathcal{B}_1}$ were a nondecreasing function of r around a round sphere $\mathfrak{s} \subset \mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$ given by $\mathfrak{s} \equiv \{\tau = \tau_b, r = r_b\}$ for some constants $r_b > 0$ and $\tau_b \geq \tau_\Sigma$, and assume that the f-trapped region $\mathcal{T}_1$ is to the future of $\mathfrak{s}$ (see Fig. 10).
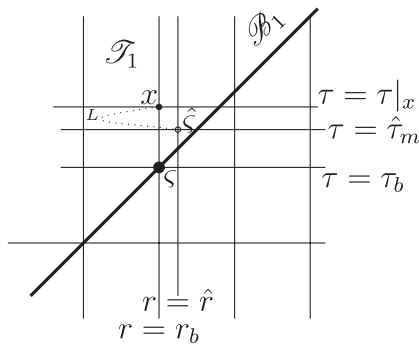

FIG. 10. This figure illustrates the proof of Proposition 10.3.

Then, any point $x \in \mathcal{R}_1$ lying on the round spheres with $r = r_b$ and $\tau > \tau_b$ but near enough τ_b would belong to at least one f-trapped closed surface. Pick up any such x and an f-trapped $S \ni x$, and let $\hat{\tau}_m$ and $\hat{r}$ be the minimum and the maximum values of $\tau|_S$ and $r|_S$ on AH_1, respectively. From Theorem 10.1 we know that $\hat{\tau}_m < \tau|_{S \cap \mathcal{R}_1}$ and $\hat{r} > r|_{S \cap \mathcal{R}_1}$. In particular, $\hat{\tau}_m < \tau|_x$, $\hat{r} > r|_x = r_b$. From Theorem 4.1 (see its third Remark) there should be a connected path $L \subset S \cap \mathcal{R}_1$ lying entirely on S starting at x and finishing on $\hat{\mathfrak{s}} \subset \text{AH}_1$, $\hat{\mathfrak{s}} \equiv \{\tau = \hat{\tau}_m\} \cap \{r = \hat{r}\}$, such that $\tau|_L$ is nonincreasing —and strictly decreasing somewhere. Thus, L would eventually cross all hypersurfaces $r = \text{const}$ with $r \in [r_b, \hat{r}]$, and each of the crossings would happen with a smaller value of τ. Because of the results in Corollary 3.1 and Proposition 10.1, and since S, and hence L, cannot intersect $\mathcal{B}_1$, this would mean that $\hat{\tau}_m$ and $\hat{r}$ should be such that $\tau_b < \hat{\tau}_m < \tau|_x$ and $r_b < \hat{r} < r|_{\mathcal{B}_1 \cap \{\tau = \hat{\tau}_m\}}$. But this leads to a contradiction, because if such a result held for *all* $x \in \{r = r_b\} \cap \{\tau > \tau_b\}$, by taking an appropriate sequence $\{x_n\}$ of such x approaching $\mathfrak{s}$, the sequence would have a limit on $\mathfrak{s}$, which would in turn produce a sequence of round spheres $\{\tau = \hat{\tau}_m(x_n), r = \hat{r}(x_n)\}$, all of them belonging to AH_1, and converging to $\mathfrak{s} \in \mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$. As AH_1 is closed, $\mathfrak{s}$ would belong to AH_1. But this is impossible as $(\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}) \cap \text{AH}_1 = \varnothing$ due to Proposition 10.2.

Corollary 10.1.—Under the same assumptions, $\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$ cannot be tangent to a $\tau = \text{const}$. hypersurface everywhere.

In particular, $\mathcal{B}_1$ never touches Σ before they merge together (and together with $\text{AH}_1 \cap \text{EH}$), that is to say, $(\mathcal{B}_1 \cap \Sigma|_{R_1}) = \varnothing$.

Corollary 10.2.—The past barrier $\Sigma \backslash EH$ is not part of the boundary $\mathcal{B}_1$.

Corollary 10.3.—Under the same assumptions, $\mathcal{B}_1$ is spacelike close enough to its merging (or asymptotic approaching) to EH, Σ and AH_1.

Proof.—We already know that $\mathcal{B}_1$ has to be spacelike somewhere. In the region of this corollary the f-trapped closed surfaces are to the future of $\mathcal{B}_1$, so that τ has to be a nonincreasing function of r. The limitation to the past by Σ and to the future by AH_1 then implies the result.

The combinations of all results obtained hitherto can be schematically represented as in Fig. 11. We can see that the boundary $\mathcal{B}_1$ is highly nonlocal too, and it can have portions in flat regions of spacetime whose whole past is also flat. This is surely not a good candidate for the surface of a dynamical black hole. Nevertheless, it remains as an interesting puzzle to find the exact location and the defining properties of $\mathcal{B}$. Some relevant results in this direction are collected in the remainder of this section.

Proposition 10.3 informs us that $\mathcal{B}_1 \backslash \text{AH}_1^{(\text{iso})}$ has to bend down in the Kodama time τ. This has direct consequences on the extrinsic curvature of $\mathcal{B}_1$. Observe that, as follows from (31) and (29), we know that the level hypersurfaces
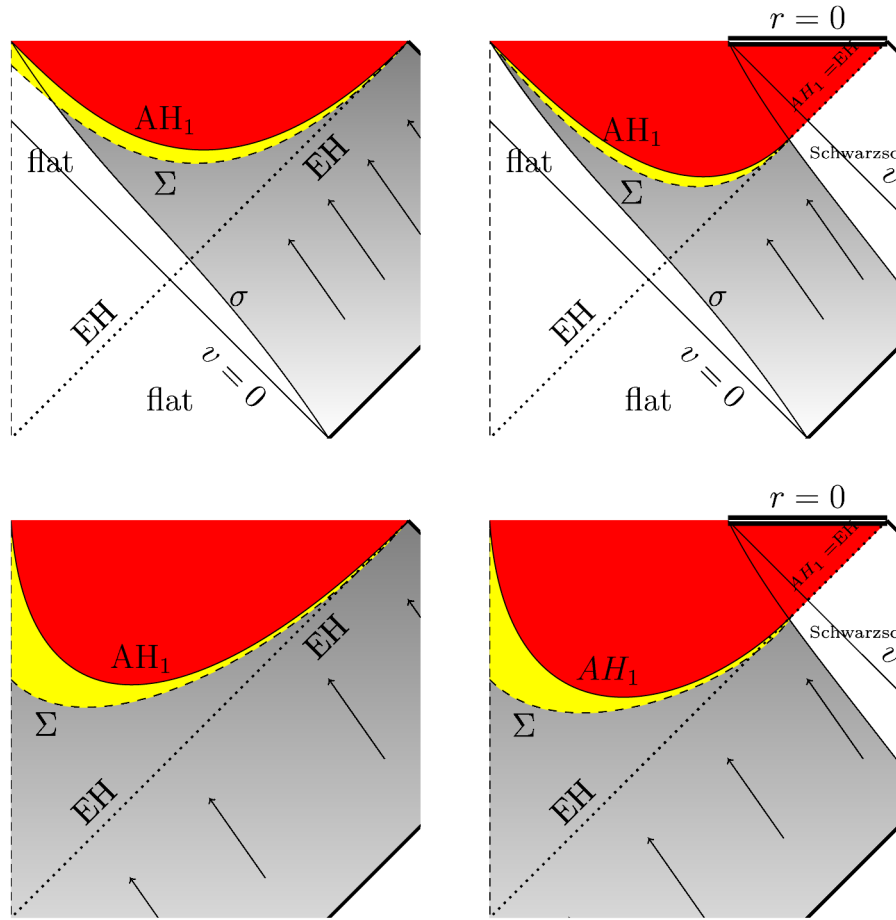
FIG. 11 (color online).   These are enlargements of some previous conformal diagrams showing the possible location of the boundary $\mathcal{B}_1$. The red regions are part of the f-trapped region $\mathcal{T}_1$, but this region actually extends further down and includes AH_1, as follows from Theorem 7.2. On the other hand, Corollary 10.2 informs us that $\mathcal{B}_1$ can never touch Σ outside EH. Thus, the yellow zones are the allowed regions for $\mathcal{B}_1$, keeping in mind that there is always a red zone below AH_1, and $\mathcal{B}_1$ must be placed strictly above Σ. In the left upper picture, Σ and also $\mathcal{B}_1$ extend to the flat region of the spacetime.

$\tau = $ const. have a non-negative semidefinite second fundamental form. Actually, they have two vanishing eigenvalues and the other one is proportional to $\partial m / \partial v$. Hence, one can prove that the boundary $\mathcal{B}_1$ must have a second fundamental form with a negative double eigenvalue.

Proposition 10.4.—Assume that the spacetime (10) satisfies the dominant energy condition, (29) and (27). Then, any portion of the connected component $\mathcal{B}_1$ which is locally to the past of $\mathcal{T}_1 \neq \varnothing$ has a second fundamental form with a non-positive (and strictly negative whenever $\mathcal{B}_1$ is not tangent to a $\tau = $ const hypersurface) double eigenvalue at any point where it is spacelike. In particular, it cannot have a positive semidefinite second fundamental form there.

Proof.—As $\mathcal{B}_1$ is part of the boundary, it has spherical symmetry. Take a portion $\hat{\mathcal{B}}_1$ of $\mathcal{B}_1$ where it is spacelike and to the past of $\mathcal{T}_1$, and include this portion in a local foliation by spherically symmetric spacelike hypersurfaces $t = $ const. Locally, the future-pointing vector field orthogonal to the foliation can be given by

$$\vec{\eta} = e^{-\beta} \partial_v + A \partial_r, \quad \eta_\mu dx^\mu = -F' dt$$

$$= dr - e^\beta (1 - \frac{2m}{r} - A) dv$$

for some function $A(v, r)$. The level hypersurfaces are spacelike ($\vec{\eta}$ is timelike) so that

$$1 - \frac{2m}{r} - 2A > 0. \tag{35}$$

Observe that $\vec{\eta} = \vec{\xi} - Ae^\beta \vec{\ell}$, and furthermore

$$\eta_\mu dx^\mu = -F' dt = \frac{A}{1 - 2m/r} dr - F(1 - \frac{A}{1 - 2m/r}) d\tau$$

by using the Kodama time τ on $\mathcal{R}_1 \supset \hat{\mathcal{B}}_1$. On $\hat{\mathcal{B}}_1$, we have that t is constant, so that

$$\frac{d\tau}{dr}\Big|_{\hat{\mathcal{B}}_1} = \frac{1}{F} \frac{A}{1 - 2m/r - A}\Big|_{\hat{\mathcal{B}}_1}.$$

However, Proposition 10.3 implies that this is strictly negative, which requires necessarily (if $A \neq 0$)

$$\frac{1 - 2m/r}{A}\Big|_{\hat{\mathcal{B}}_1} < 1.$$

Given that $1 - 2m/r > 0$ at $\hat{\mathcal{B}}_1 \subset \mathcal{R}_1$, this together with (35) implies that (if $A \neq 0$)

$$A < 0.$$

We are now going to show that A is essentially the double eigenvalue of the second fundamental form of $\hat{\mathcal{B}}_1$. A straightforward calculation provides the deformation of the metric along $\vec{\eta}$

$$(\mathcal{L}_{\vec{\eta}} g)_{\mu\nu} = -2 \frac{\partial \beta}{\partial r} \eta_\mu \eta_\nu + e^\beta \left[\frac{\partial \beta}{\partial r} \left(1 - \frac{2m}{r} - A \right) - \frac{\partial A}{\partial r} \right]$$
$$\times (\ell_\mu \eta_\nu + \ell_\nu \eta_\mu) + \left[e^\beta \frac{2}{r} \frac{\partial m}{\partial v} + 2 e^\beta \frac{\partial A}{\partial v} \right.$$
$$\left. + e^{2\beta} \frac{\partial A}{\partial r} (1 - 2m/r - A) - A \frac{\partial [e^{2\beta}(1 - 2m/r)]}{\partial r} \right] \ell_\mu \ell_\nu$$
$$+ 2r A d\Omega_{\mu\nu} \tag{36}$$

where $d\Omega_{\mu\nu}$ represents the angular part of the metric, that is, the metric on the unit round sphere. Of course, the projection to $\mathcal{B}_1$ of $\mathcal{L}_{\vec{\eta}} g$ is (proportional to) the second fundamental form $K_{\mu\nu}$ of $\mathcal{B}_1$ whenever it is spacelike, so that $K_{\mu\nu}|_{\hat{\mathcal{B}}_1}$ is proportional to the restriction to $\hat{\mathcal{B}}_1$ of the second and third lines in (36). In particular, the double eigenvalue is proportional to A.

As a matter of fact, one can try to do better and try to set restrictions also on the third eigenvalue of $K_{\mu\nu}|_{\hat{\mathcal{B}}_1}$. This is given by those terms in Eq. (36) that contain $\ell_\mu \ell_\nu$ (projected to $\hat{\mathcal{B}}_1$). The idea is based on theorem 4.1 and corollary 4.1. There are closed f-trapped surfaces passing through every point $x \in \mathcal{T}_1$ which is locally to the future of $\hat{\mathcal{B}}_1$. Choose a sufficiently smooth sequence of closed f-trapped surfaces approaching $\hat{\mathcal{B}}_1$. The smoothness is assumed such that one can take the limit and obtain a piece of a surface ζ touching $\hat{\mathcal{B}}_1$. The set obtained as limit of the surfaces may fail to be compact, e.g. the Robertson-Walker example mentioned after Result 5.4; or to be spacelike everywhere, or even to be connected. However, all these problems will be irrelevant in what follows as long as ζ exists. (This is the main problem when trying to turn this reasoning into a formal proof, as proving the existence of ζ encounters some technical difficulties from a mathematical viewpoint).

Given that all the surfaces in the sequence are f-trapped, they all have $t > t_{\hat{\mathcal{B}}_1}$, where $t_{\hat{\mathcal{B}}_1}$ is the constant value of t at $\hat{\mathcal{B}}_1$. Therefore, ζ is such that $t|_\zeta \geq t_{\hat{\mathcal{B}}_1}$, which implies that ζ has a local minimum of t at the intersection $\zeta \cap \hat{\mathcal{B}}_1$,

or that ζ has a two-dimensional portion within $\hat{\mathcal{B}}_1$. In particular, ζ is spacelike around that minimum because $\vec{\eta}$ is orthogonal to ζ there. The mean curvature vector of this spacelike portion of ζ (including $\zeta \cap \hat{\mathcal{B}}_1$) must be future-pointing or zero, given that ζ is the limit of the sequence of f-trapped surfaces. Therefore, we have $\bar{\nabla}_A \bar{\eta}^A|_{\zeta \cap \hat{\mathcal{B}}_1} \leq 0$ and also $\eta_\mu H^\mu|_{\zeta \cap \hat{\mathcal{B}}_1} \leq 0$. Now, we can apply the same reasoning as in Theorem 4.1, for which the compactness is not necessary. Formula (4) applied to the spacelike portion of ζ containing $\zeta \cap \hat{\mathcal{B}}_1$ gives

$$\bar{\nabla}_A \bar{\eta}^A|_{\zeta \cap \hat{\mathcal{B}}_1} + \eta_\mu H^\mu|_{\zeta \cap \hat{\mathcal{B}}_1} = \frac{1}{2} P^{\mu\nu} (\mathcal{L}_{\vec{\eta}} g)_{\mu\nu}|_{\zeta \cap \hat{\mathcal{B}}_1}$$

where $P^{\mu\nu}$ is the orthogonal projector to ζ. Thus, we deduce that

$$P^{\mu\nu} (\mathcal{L}_{\vec{\eta}} g)_{\mu\nu}|_{\zeta \cap \hat{\mathcal{B}}_1} = P^{\mu\nu} K_{\mu\nu}|_{\zeta \cap \hat{\mathcal{B}}_1} \leq 0.$$

Observe that, provided this argument can be promoted into a rigorous proof, it may restrict the third eigenvalue severely, because it has to hold for *all* projectors which are limits of the projectors to closed f-trapped surfaces close enough to $\hat{\mathcal{B}}_1$. If one could gain control on the variety of such projectors, then much more precise restrictions could be set on the boundary $\mathcal{B}$.

XI. THE CORE OF THE TRAPPED REGION AND AH

At this point we know that the EH is teleological, and also that closed f-trapped surfaces are clairvoyant: they are "aware" of things that happen elsewhere, with spacelike separation. For instance, they can have portions in a flat region of spacetime whose whole past is also flat in clairvoyance of energy that crosses them elsewhere to make their compactness and trapping feasible [16,46], see Fig. 8. This nonlocal property of trapped surfaces is inherited by everything which is based on them, such as marginally trapped tubes including dynamical horizons. In conjunction with the nonuniqueness of dynamical horizons, this poses a fundamental puzzle for the physics of black holes, a problem that has been recognized and discussed many times lately, see e.g. [3,5,12–14,43,45,47,56] and references therein.

Four possible solutions have been put forward [47]. First, one can rely on the old and well-defined event horizon. This encounters very serious problems because one needs to know the whole future evolution of the spacetime. The event horizon is unreasonably global [2,3]. Actually, the whole construction of trapping and dynamical horizons was developed to solve this problem and to have nice, local, definitions of the surface of a black hole [5,12]. Second, one can treat all possible horizons on equal footing. The problem is how to associate unique physical properties to the corresponding black hole, because each

dynamical horizon comes with its own set of magnitudes. And they do not agree. The third strategy consisted in finding the boundary $\mathcal{B}$ as defined in this paper. However, as we have shown, not only $\mathcal{B}$ will not be a marginally trapped tube in general, it also suffers from the nonlocal properties associated to f-trapped surfaces. For instance, we have seen that $\mathcal{B}$ can enter the flat regions of spacetime. Finally, the fourth approach consists in trying to define a preferred dynamical horizon. Hitherto, there has been no good definition for that.

In the following, we are going to pursue a novel strategy. The idea is based on the simple question: what part of the spacetime is absolutely indispensable for the existence of the black hole? We already know that, in the cases considered so far with flat regions and matter imploding so that a black hole eventually forms, the flat region is certainly not essential for the existence of the black hole. What is? By answering this question we might actually get a bonus and provide a positive, constructive solution to the fourth strategy mentioned before: it may happen that a unique dynamical horizon is selected.

From Corollary 9.2 and Theorem 9.3 it is clear that if the whole complement of $\mathcal{R}$ is removed from the spacetime, then no closed f-trapped surfaces remain. The question arises of whether or not proper subsets of that removed region suffice to achieve the same, that is, to short-circuit all closed f-trapped surfaces. To be precise, we give the following definition.

Definition 3.—A region $\mathcal{Z}_1 \subset \mathcal{V}$ is called the *core* of a connected component $\mathcal{T}_1$ of the f-trapped region $\mathcal{T}$ if it is a minimal closed connected set that needs to be removed from the spacetime in order to get rid of all closed f-trapped surfaces in $\mathcal{T}_1$, and such that any point on the boundary $\partial \mathcal{Z}_1$ is connected to $\mathcal{B}_1 = \partial \mathcal{T}_1$ in the closure of the remainder.

Here, "minimal" means that there is no other set $\mathcal{Z}'$ with the same properties and properly contained in $\mathcal{Z}_1$. The final condition states that the excised spacetime $(\mathcal{V} \backslash \mathcal{Z}_1, g)$, which no longer has a connected f-trapped region $\mathcal{T}_1$, has the property that furthermore each point in the closure $(\mathcal{V} \backslash \mathcal{Z}_1) \cup \partial \mathcal{Z}_1$ can be joined to the original boundary $\mathcal{B}_1$ by a continuous curve fully contained in $(\mathcal{V} \backslash \mathcal{Z}_1) \cup \partial \mathcal{Z}_1$. (This curve may have zero length at points $x \in \partial \mathcal{Z}_1 \cap \partial \mathcal{B}_1$). This is needed because one could identify a particular removable region to eliminate the f-trapped surfaces, excise it, but then put back a tiny but central isolated portion to make it smaller. However, this is not what one wants to cover with the definition.

Obviously, $\mathcal{Z}_1 \subset \mathcal{T}_1$, however, $\mathcal{Z}_1$ will be generally smaller than $\mathcal{T}_1$. As an example, take the Robertson-Walker spacetime of Fig. 2. There, the future-trapped region $\mathcal{T}$ is the whole future of the recollapsing time, shown in red. However, one only needs to remove the triangle to the future of the AH in order to get rid of all f-trapped surfaces. (Note that the boundary of this region is

an apparent 3-horizon AH). This example also proves that $\mathcal{Z}$ is not unique: one can choose any other region $\mathcal{Z}$ equivalent to the chosen one by moving all its points by the group of symmetries on each homogeneous slice.

Actually this kind of nonuniqueness is rather trivial, and is due to the existence of a high degree of symmetry. Nevertheless, even in less symmetric cases the uniqueness of the cores $\mathcal{Z}$ cannot be assumed beforehand. We are actually going to show that it does not hold in general.

We can use the results found in Sec. VII, especially Theorem 7.3, to identify one core of the f-trapped region in spherically symmetric spacetimes.

Theorem 11.1.—Assume that $AH^{iso} = \varnothing$ for simplicity. Then, the complement of $\mathcal{R}_0$ (i.e. the region $\mathcal{Z} \equiv \{r \le 2m(v, r)\}$) is the disjoint union of core f-trapped regions. Each of its connected components is the core of the corresponding connected components of $\mathcal{T}$.

Remark.—The cases with $AH^{iso} \ne \varnothing$ are technically more involved, but one expects that the result will hold true too.

Proof.—First of all, it is clear that every closed f-trapped-surface has points in $\mathcal{Z} \equiv \mathcal{V} \backslash \mathcal{R}_0$, as follows from Corollary 9.2 or Theorem 9.3. Hence, if we remove $\mathcal{Z}$ from the spacetime, all closed f-trapped surfaces disappear. To see that there is no proper subset of $\mathcal{Z}$ with this property, observe that its boundary is $\partial \mathcal{Z} = AH$. Take then any closed connected proper subset $\mathcal{Z}'$ of $\mathcal{Z}$ such that all points of $\partial \mathcal{Z}'$ are connected to its nearest part of the boundary $\mathcal{B}$. This implies that there is a curve from every $x \in \partial \mathcal{Z}'$ to such a part of the boundary $\mathcal{B}$, and all these curves must therefore cross AH. In summary, $\mathcal{Z} \backslash \mathcal{Z}'$ always contains an open region around AH and outside $\mathcal{R}_0$. But then, theorem 7.3 ensures that there are closed f-trapped surfaces fully contained in $\mathcal{V} \backslash \mathcal{Z}'$, so that $\mathcal{Z}'$ cannot be a core.

As a bonus, we have obtained that the boundary of the identified core $\mathcal{Z} = \{r \le 2m\}$ is formed by the marginally trapped tubes AH, in particular, by their dynamical horizon portions. One can wonder if this property selects these marginally trapped tubes in spherically symmetric spacetimes. Observe, however, that according to the results in [3], given any regular dynamical horizon H, there cannot be any closed weakly f-trapped surface fully contained in its past domain of dependence $D^-(H)$. Therefore, if we remove the appropriate future part of H, we also remove all possible closed f-trapped surfaces. The question now is whether or not these alternative would-be cores are actually optimal, or if they remove more than is needed from the spacetime to get rid of the f-trapped region. Independently of whether or not they are optimal, the result in [3] allows us to prove that there are nonspherically symmetric cores in spherically symmetric spacetimes.

First we show that $\mathcal{Z}$ are the unique spherically symmetric cores.

Proposition 11.1.—In spherically symmetric spacetimes, $\mathcal{Z} = \{r \le 2m\}$ are the only cores of $\mathcal{T}$ which are

invariant under the action of the corresponding SO(3) group of isometries. Therefore, $\partial \mathcal{Z} = $ AH are the only spherically symmetric boundaries of a core.

Proof.—Suppose there was another spherically symmetric core $\mathcal{Z}'$. Obviously $\mathcal{Z}'$ could not be a proper subset of $\mathcal{Z}$, nor vice versa, because both are cores. Thus $\mathcal{Z} \backslash \mathcal{Z}' \neq \varnothing$ and this set would be spherically symmetric. However, every round sphere in $\mathcal{Z}$ is f-trapped, and therefore there would be f-trapped round spheres having no intersection with $\mathcal{Z}'$, contradicting the hypothesis that $\mathcal{Z}'$ was a core of the trapped region.

Proposition 11.2.—There exist nonspherically symmetric cores of the f-trapped region in spherically symmetric spacetimes.

Proof.—Consider the case when the spacetime has only one connected component of AH, which is spacelike, such as, for example, the Vaidya spacetime in the Appendix. From Corollary 7.1, or the general results in [3], we know that there are nonspherically symmetric dynamical horizons interweaving the AH—see also the explicit constructions in [16,46,47]. Take any of these, say H, so that H lies partly to the future of AH (and partly to its past). From Theorem 4.1 in [3], no weakly f-trapped surface can be fully contained in the past domain of dependence of H. Consider then the causal future $J^+(H)$ of H. Removing $J^+(H)$ from the spacetime eliminates all closed f-trapped surfaces. Nevertheless, it may happen that $J^+(H)$ is not a core, because it is not minimal. In any case, there is a subset of $J^+(H)$ which is a core of the f-trapped region $\mathcal{T}$. This new core will never include those parts of the spacetime which are to the future of AH, but to the past of H. Thus, this core of $\mathcal{T}$ is not $\mathcal{Z}$, and due to Proposition 11.1, it cannot be spherically symmetric.

Still, the identified core $\mathcal{Z} = \{r \leq 2m\}$ may be unique in the sense that its boundary $\partial \mathcal{Z} = $ AH is a marginally trapped tube. This would happen if, for instance, in the example of the previous proof, any dynamical horizon H other than AH is such that $J^+(H)$ is *not* a core of the f-trapped region, the core being a proper subset of $J^+(H)$. If this is the case, then AH would be selected as the unique dynamical horizon which is the boundary of a core of the f-trapped region $\mathcal{T}$.

Whether or not this happens is a very interesting open question.

ACKNOWLEDGMENTS

APPENDIX A: THE IMPLODING VAIDYA SPACETIME

Consider the important case of the Vaidya spacetime with incoming radiation. The line-element reads [4,34]

$$ds^2 = -\left(1 - \frac{2m(v)}{r}\right)dv^2 + 2dv\,dr + r^2 d\Omega^2 \quad \text{(A1)}$$

so that this is the particular case of (10) with $\beta = 0$ and a mass function independent of r, $m(v) \geq 0$.

The Einstein tensor of (A1) is of pure radiation type

$$G_{\mu\nu} = \frac{2}{r^2}\frac{dm}{dv}\ell_\mu \ell_\nu,$$

and thus, if the Einstein field equations are assumed, the null convergence condition (which in this particular case implies the dominant energy condition) [8] requires that

$$\frac{dm}{dv} \geq 0 \quad \text{(A2)}$$

so that the mass function cannot decrease as a function of v. Thus, in this case, the condition (29) is guaranteed.

There is only one connected component of AH defined by

$$\text{AH:}\ r - 2m(v) = 0$$

and unique associated regions $\mathcal{R}_0$: $r > 2m(v)$ and $\mathcal{R}$: $r \geq 2m(v)$. AH is a spacelike hypersurface for all v such that $dm/dv > 0$, and it is null where $dm/dv = 0$. Therefore, AH is a dynamical horizon [5] on the region where it is spacelike, and an isolated horizon [5] on any open region where $m(v) = $ const.

We adopt all the assumptions as in the general case so that the mass function satisfies

$$m(v) = 0 \qquad \forall v < 0; \qquad m(v) \leq M < \infty \qquad \forall v > 0 \quad \text{(A3)}$$

together with (A2).

The spherically symmetric collapse of null radiation may lead to the formation of a naked singularity [57,58]. To avoid this possibility one has to assume that [57]

$$\lim_{v \to 0^+} \frac{m(v)}{v} > \frac{1}{16}.$$

A hidden, non-naked, curvature singularity is always present at $r = 0$, $v > 0$. This is a spacelike future singularity.

Under all the above conditions, the Penrose diagrams for the imploding Vaidya spacetime are depicted in Fig. 5, the first of them for the case with $m(v) < M$ everywhere, the second with $m(v) = M$ from $v = v_1$ on.

The Kodama vector field $\vec{\xi} = \partial_v$ defined in subsection 9 is actually a proper Kerr-Schild vector field [35] (KSVF from now on) of type (6) relative to the null direction $\vec{\ell}$ for the Vaidya spacetime. It is immediate to get

$$(\mathcal{L}_{\vec{\xi}}g)_{\mu\nu} = 2\frac{dm}{dv}\ell_\mu\ell_\nu, \qquad (\mathcal{L}_{\vec{\xi}}\ell)_\mu = 0$$

so that the function h in (6) is $h = dm/dv \geq 0$. This is one of the requirements of Theorem 4.1. The other requirements are satisfied as in the general case, so that $\vec{\xi}$ is future pointing on the region $\mathcal{R} = \mathcal{R}_0 \cup$ AH, and the level function τ is defined by

$$\xi_\mu dx^\mu = -Fd\tau = dr - \left(1 - \frac{2m(v)}{r}\right)dv. \qquad \text{(A4)}$$

Besides, $\vec{\xi}$ is timelike on $\mathcal{R}_0$ and null at the AH: $r = 2m(v)$.

Notice that the KSVF $\vec{\xi} = \partial_v$ coincides with the standard static Killing vector on the regions with $m(v) = $ const and, in particular, with a static Killing vector in the flat region $v < 0$ where $m(v) = 0$. Thus, all portions of AH that are isolated horizons are actually Killing horizons, including the portion of the EH with $v > v_1$ in the case that $m(v) = M$ for all $v > v_1$.

1. The level hypersurfaces $\tau = $ const and the past barrier Σ in the Vaidya spacetime

Now, we analyze the shape and position of the level hypersurfaces $\tau = $ const for the KSVF $\vec{\xi} = \partial_v$ in the Vaidya spacetime (A1) subject to (A2) and (A3). Observe that these hypersurfaces are characterized by being spherically symmetric and orthogonal to the hypersurfaces $r = $ const.

The definition of τ is (A4), so that the sought hypersurfaces are defined by the solution to the differential equation

$$\frac{dv}{dr} = \frac{1}{1 - 2m(v)/r}. \qquad \text{(A5)}$$

This is equivalent to the autonomous system

$$\frac{dv}{du} = r, \qquad \frac{dr}{du} = r - 2m(v)$$

so that its orbits on the phase plane $\{v, r\}$ provide the required hypersurfaces. Given that $m(0) = 0$, the ODE (A5) has a critical point at $v = 0, r = 0$. Its linear stability is ruled by the eigenvalues $\lambda_\pm$ and eigenvectors $\vec{z}_\pm$ of the corresponding matrix

$$\begin{pmatrix} 0 & 1 \\ -2m_0 & 1 \end{pmatrix}, \qquad m_0 \equiv \lim_{v\to 0^+}\frac{m(v)}{v}$$

which are given by

$$\lambda_\pm = \frac{1}{2}(1 \pm \sqrt{1 - 8m_0}), \qquad \vec{z}_\pm = (1, \lambda_\pm).$$

The character of the critical point is different depending on whether $m_0 < 1/8$ or not. If $m_0 > 1/8$, the critical point is an unstable focus, so that no solution actually reaches (0, 0). The schematic phase plane is represented in Fig. 12. As we can check, the hypersurface Σ always penetrates the flat region in this case.
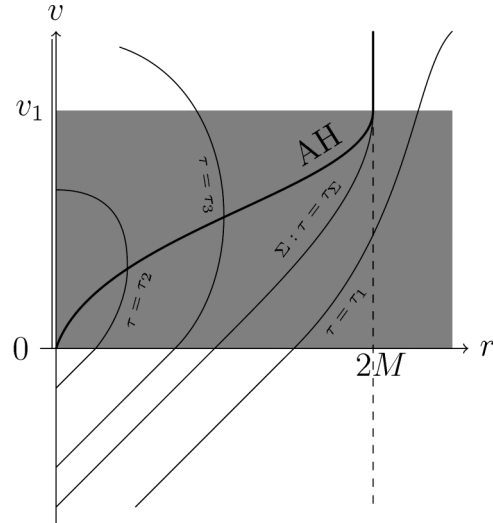

FIG. 12. The phase plane $\{v, r\}$ around the critical point (0, 0) for the Eq. (A5) in the spacetime (A1) subject to (A2) and (A3). Only the case with $m(v) = M$ for all $v \geq v_1$ is represented. It should be noticed that horizontal lines ($v = $ const) are null in the spacetime, while vertical lines ($r = $ const) are timelike below AH, null at AH, and spacelike above it. Several $\tau = $ const hypersurfaces are represented. They are spacelike while they remain below AH, null when they meet AH, if they do, and timelike above AH. Here, we have represented the case with $m_0 > 1/8$ (see main text), so that the origin is a *focus*, which implies that all $\tau = $ const hypersurfaces, and in particular Σ, never reach the origin.

If $m_0 < 1/8$, the critical point is an unstable node, with $\lambda_\pm > 0$ real and positive. In fact, one can see that

$$\lambda_+ \in \left(\frac{1}{2}, 1\right], \qquad \lambda_- \in \left[0, \frac{1}{2}\right), \qquad \lambda_+ + \lambda_- = 1.$$

In this case, all possible solutions except one emerge from (0, 0) with the same tangent direction given by the eigenvector $(1, \lambda_-)$. The exception is given by one solution emerging from (0, 0) with the tangent direction of the other eigenvector $(1, \lambda_+)$. Recall that $m_0 \geq 1/16$ is necessary to avoid (locally) naked singularities [57], and thus the allowed intervals for $\lambda_\pm$ can be further restricted if such singularities are to be avoided. There are three qualitatively different possibilities in this case (with the same value of m_0), depending on whether or not the special solutions $\tau = \tau_\pm$—corresponding to the eigenvalues $\lambda_\pm$ at (0, 0)—eventually meet the AH. This, in turn, will depend on the specific properties of the mass function $m(v)$ for $v \neq 0$ and on the total mass M. Loosely speaking, if there is a period when the mass function has a large derivative, then the chances are that the special solutions will meet the AH. If at least one of the special solutions does *not* meet the AH, then the hypersurface Σ cannot penetrate the flat region. On the other hand, if both special solutions meet AH then Σ will have a portion in the flat region. These possibilities are schematically represented in the Figs. 13 and 14.
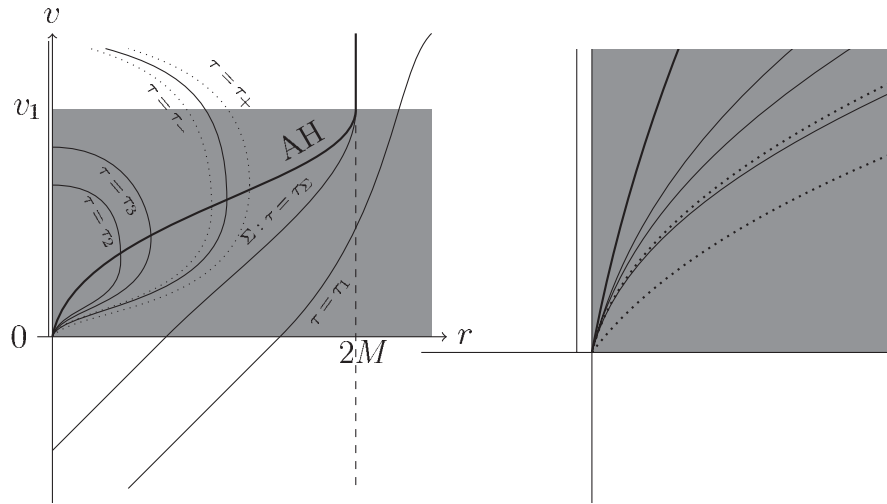
FIG. 13.   Same phase plane as in Fig. 12 but with $m_0 < 1/8$. The origin is now a *node*, with two special solutions given by $\tau = \tau_\pm$ and represented by dotted lines. All solutions but $\tau = \tau_+$ emerge from the origin with the same slope given by that of the solution $\tau = \tau_-$. This is represented in the right picture, which is a magnification of a small neighborhood of $(0, 0)$. If the mass function $m(v)$ is such that both solutions $\tau = \tau_\pm$ cross the AH, as depicted, then Σ will never reach the origin. The other possibilities are given in Fig. 14.
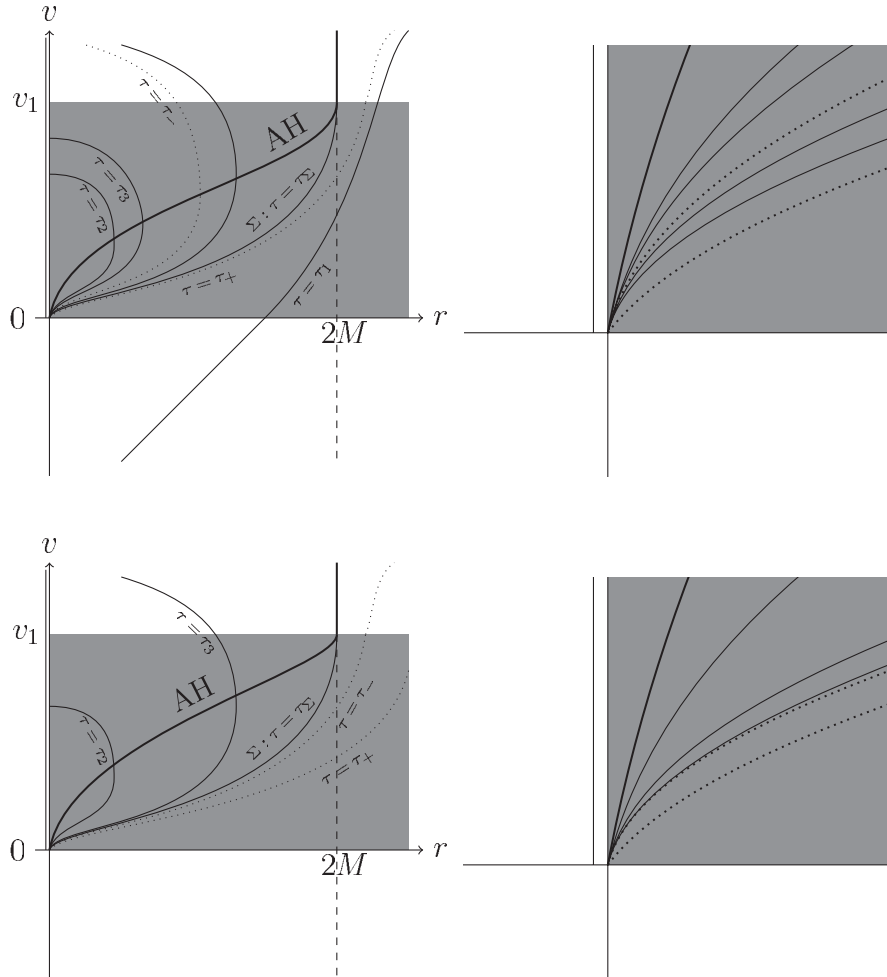

FIG. 14.   Same case $m_0 < 1/8$ as in Fig. 13 so that the origin is a *node*, with their corresponding magnifications on the right. If the mass function $m(v)$ is such that at least one of the solutions $\tau = \tau_\pm$ crosses the AH, then Σ will reach the origin tangent to $\tau = \tau_-$. This implies that Σ will never penetrate into the flat part of the spacetime, which is the half-plane $v < 0$. The upper figure represents the case when only one of the two solutions crosses AH, while the bottom does the same when both $\tau = \tau_\pm$ are spacelike everywhere, never crossing AH.

The limit case with $m_0 = 1/8$ has the character of a degenerate unstable node (only one universal direction through which all solutions emerge from the critical point) in the linear stability analysis, and it remains as such, or it may become an unstable focus or node, depending on the specific properties of the function $m(v)$ around $v = 0$. The schematic structure of the phase portraits for this case are thus analogous to those already shown, with the small difference that the solutions $\tau = \tau_+$ and $\tau = \tau_-$ coincide when $(0, 0)$ is a degenerate node for the full, nonlinear, system.

The corresponding Penrose diagrams, including the most relevant $\tau = $ const hypersurfaces, are presented in the next Fig. 15.

To illustrate the above, we present a particular example where the solutions of (A5) can be given explicitly in full. This is given by the self-similar Vaidya spacetime, with a linear mass function

$$m(v) = \begin{cases} M & v > v_1 = M/\mu \\ \mu v & 0 \le v \le v_1 \\ 0 & v < 0 \end{cases}$$

which admits the following homothetic Killing vector field for all $v < v_1$

$$\vec{\zeta} = v\partial_v + r\partial_r, \qquad (\mathcal{L}_{\vec{\zeta}} g)_{\mu\nu} = 2g_{\mu\nu}.$$

Observe that $m_0 = \mu$ in this case, so that $\lambda_\pm = (1 \pm \sqrt{1 - 8\mu})/2$. The solutions to the ODE (A5) provide the level hypersurfaces for $\vec{\xi}$. They are given by $\tau = v - r$ for all $v < 0$. For $v > 0$, we have:

(i) $\mu > 1/8$. The critical point is an unstable focus and the solutions are

$$\tau(v, r) = -\sqrt{r^2 - vr + 2\mu v^2}$$

$$\times \exp\left\{\frac{1}{\sqrt{8\mu - 1}} \arctan\left(\frac{1 - 4\mu v/r}{\sqrt{8\mu - 1}}\right)\right\}$$

(ii) $\mu = 1/8$. The critical point is a degenerate unstable node whose particular special solution is simply

$$\tau = \tau_+ = \tau_- \Leftrightarrow v = 2r$$

and the rest of solutions, all of them emanating from the origin tangent to the special solution $v = 2r$, are

$$\tau(v, r) = (v - 2r) \exp\left\{\frac{2r}{v - 2r}\right\}$$

(iii) $\mu < 1/8$. The critical point is an unstable node with the special solutions

$$\tau = \tau_+ \Leftrightarrow 2\mu v = \lambda_+ r,$$

$$\tau = \tau_- \Leftrightarrow 2\mu v = \lambda_- r$$

the first of them being the exceptional one. The rest of the solutions, all of them emanating from the origin tangent to the second special solution $2\mu v = \lambda_- r$, are

$$\tau(v, r) = (2\mu v - \lambda_- r)^{\lambda_+/\sqrt{1 - 8\mu}}$$

$$\times (2\mu v - \lambda_+ r)^{-(\lambda_-/\sqrt{1 - 8\mu})}.$$

In all three cases, the hypersurface Σ is given by $\tau = \tau_\Sigma$, where $\tau_\Sigma \equiv \tau(M/\mu, 2M)$. Observe that Σ does not enter
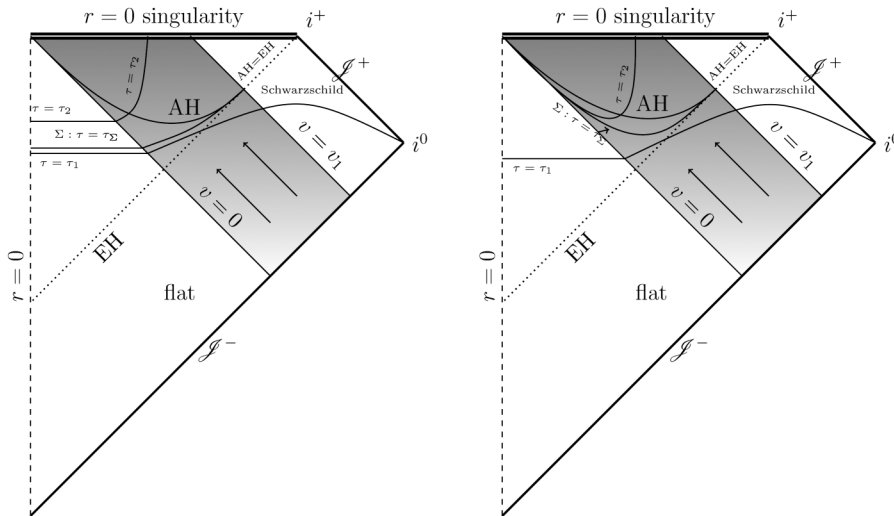

FIG. 15.   These are conformal diagrams of spacetime (A1) with (A2) and (A3) when $m(v) = M$ for all $v > v_1$. The notation is as in Fig. 5. Here, on the left picture we have represented the case in which the hypersurface Σ has a portion in the flat region of the spacetime (e.g. if $m_0 > 1/8$, but not only). On the right picture, the other possibility is represented, when Σ never touches the flat region. These two pictures are the global Penrose diagrams corresponding to the pictures on the right of Fig. 7.

the flat region if $\mu \leq 1/8$, and therefore no closed f-trapped surface can penetrate the flat region in these cases [16].

We note as a final remark that the AH in this particular Vaidya spacetime is an intrinsically flat hypersurface, and that the trace of its second fundamental form (its expansion) is proportional to $1 - 8\mu$. Therefore, the AH is nonexpanding exactly for the limit case with $\mu = 1/8$. In the cases where the past barrier Σ enters the flat portion of the spacetime, the AH is contracting.

---

[1] S. W. Hawking, Commun. Math. Phys. **25,** 152 (1972).

[2] P. Hajicek, Phys. Rev. D **36,** 1065 (1987).

[3] A. Ashtekar and G. J. Galloway, Adv. Theor. Math. Phys. **9,** 1 (2005).

[4] P. C. Vaidya, Proc. Indian Acad. Sci. A **33,** 264 (1951).

[5] A. Ashtekar and B. Krishnan, Living Rev. Relativity **7,** 10 (2004), http://www.livingreviews.org/lrr-2004-10.

[6] R. Penrose, Phys. Rev. Lett. **14,** 57 (1965).

[7] J. Thornburg, Living Rev. Relativity **10,** 3 (2007), http://www.livingreviews.org/lrr-2007-3.

[8] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973).

[9] L. Andersson and J. Metzger, Commun. Math. Phys. **290,** 941 (2009).

[10] M. Kriele and S. A. Hayward, J. Math. Phys. (N.Y.) **38,** 1593 (1997).

[11] L. Andersson, M. Mars, and W. Simon, Phys. Rev. Lett. **95,** 111102 (2005).

[12] I. Booth, Can. J. Phys. **83,** 1073 (2005).

[13] I. Booth, L. Brits, J. A. Gonzalez, and C. Van Den Broeck, Classical Quantum Gravity **23,** 413 (2006).

[14] D. M. Eardley, Phys. Rev. D **57,** 2299 (1998).

[15] I. Ben-Dov, Phys. Rev. D **75,** 064007 (2007).

[16] I. Bengtsson and J. M. M. Senovilla, Phys. Rev. D **79,** 024027 (2009).

[17] S. A. Hayward, Phys. Rev. D **49,** 6467 (1994).

[18] S. A. Hayward, Classical Quantum Gravity **11,** 3025 (1994).

[19] M. Mars and J. M. M. Senovilla, Classical Quantum Gravity **20,** L293 (2003).

[20] J. M. M. Senovilla, in arXiv:gr-qc/0311005v1;

[21] J. M. M. Senovilla, J. High Energy Phys. 11 (2003) 046.

[22] L. Andersson, M. Mars, and W. Simon, Adv. Theor. Math. Phys. **12,** 853 (2008).

[23] R. P. A. C. Newman, Classical Quantum Gravity **4,** 277 (1987).

[24] H. Kodama, Prog. Theor. Phys. **63,** 1217 (1980).

[25] S. W. Hawking and R. Penrose, Proc. R. Soc. A **314,** 529 (1970).

[26] M. Kriele, *Spacetime* (Springer, Berlin, 1999).

[27] R. M. Wald, General Relativity (The University of Chicago Press, Chicago, 1984).

[28] B. O'Neill, *Semi-Riemannian Geometry: With Applications to Relativity* (Academic Press, New York, 1983).

[29] J. M. M. Senovilla, Classical Quantum Gravity **24,** 3091 (2007).

[30] G. J. Galloway, Commun. Anal. Geom. **16,** 217 (2008).

[31] A. Carrasco and M. Mars, Classical Quantum Gravity **26,** 175002 (2009).

[32] R. Penrose, *Techniques of Differential Topology in Relativity*, Regional Conference Series in Applied Math. Vol. 7 (SIAM, Philadelphia, 1972).

[33] J. M. M. Senovilla, Gen. Relativ. Gravit. **30,** 701 (1998).

[34] H. Stephani, D. Kramer, M. A. H. MacCallum, C. Hoenselaers, and E. Herlt, *Exact Solutions to Einstein's Field Equations Second Edition* (Cambridge University Press, Cambridge, 2003).

[35] B. Coll, S. R. Hildebrandt, and J. M. M. Senovilla, Gen. Relativ. Gravit. **33,** 649 (2001).

[36] S. B. Myers, Trans. Am. Math. Soc. **71,** 211 (1951).

[37] R. Osserman, *A Survey of Minimal Surfaces* (Dover, New York, 1986).

[38] J. L. Flores, S. Haesen, and M. Ortega, Classical Quantum Gravity **27,** 145021 (2010).

[39] R. Bartnik and J. Isenberg, Classical Quantum Gravity **23,** 2559 (2006).

[40] F. Fayos, J. M. M. Senovilla, and R. Torres, Phys. Rev. D **54,** 4862 (1996).

[41] M. Dafermos, Classical Quantum Gravity **22,** 2221 (2005).

[42] G. J. Galloway and R. M. Wald (private communication).

[43] I. Booth and S. Fairhurst, Phys. Rev. D **75,** 084019 (2007).

[44] M. Spivak, *A Comprehensive Introduction to Differential Geometry* (Publish or Perish, Houston, 1999), 3rd ed., Vol. 5.

[45] S. A. Hayward, arXiv:0906.2528.

[46] J. E. Åman, I. Bengtsson, and J. M. M. Senovilla, J. Phys. Conf. Ser. **229,** 012004 (2010); arXiv:0912.3691.

[47] Alex B. Nielsen, Michael Jasiulek, Badri Krishnan, and Erik Schnetter, arXiv:1007.2990.

[48] F. Fayos, J. M. M. Senovilla, and R. Torres, Classical Quantum Gravity **20,** 2579 (2003).

[49] S. A. Hayward, Phys. Rev. Lett. **81,** 4557 (1998).

[50] C. Williams, Ann. Inst. Henri Poincaré **9,** 1029 (2008).

[51] C. Williams, arXiv:1005.5401.

[52] G. Abreu and M. Visser, Phys. Rev. D **82,** 044027 (2010).

[53] I. Racz, Classical Quantum Gravity **23,** 115 (2006).

[54] Roh-Suan Tung, Classical Quantum Gravity **25,** 085005 (2008).

[55] C.-M. Claudel, arXiv:gr-qc/0005031.

[56] B. Krishnan, Classical Quantum Gravity **25,** 114005 (2008).

[57] Y. Kuroda, Prog. Theor. Phys. **72,** 63 (1984).

[58] A. Papapetrou, in *A Random Walk in Relativity and Cosmology*, edited by N. Dadhich *et al.* (Wiley, New York, 1985).