Cosmological singularity theorems and black holes

Alexander Vilenkin¹ and Aron C. Wall²

¹ Institute of Cosmology, Department of Physics and Astronomy,

Tufts University, Medford, MA 02155, USA

² Department of Physics, University of California,

Santa Barbara, California 93106, USA

Abstract

An extension of Penrose's singularity theorem is proved for spacetimes where black holes are allowed to form from non-singular initial data. With standard assumptions about the spacetime, and assuming the existence of a trapped surface which lies outside of black hole horizons and is not completely surrounded by horizons, we show that the spacetime region outside (or on) the horizons must contain singularities. If the trapped surface is surrounded by horizons, we show that the horizons divide spacetime into causally disconnected pieces. Unlike the original Penrose's theorem, our theorems provide some information about the location of singularities. We illustrate how they can be used to rule out some cosmological scenarios.

I. INTRODUCTION

The singularity theorems of General Relativity demonstrate that spacetimes having certain properties must necessarily be singular. The assumptions of the theorems are rather general and mild; in this sense these are very powerful results. Their main weakness is that they do not say much about the nature of the singularities.

Consider for example the question that motivated some of the early work on the subject. It has been known since Friedmann and Lemaitre that contracting open homogeneous and isotropic (FLRW) models collapse to a big crunch. But could it be that the singularity is an artifact of the symmetry of these models? If the contracting universe is made up of clumps, could the clumps miss one another and fly apart, so that the contraction is followed by an expansion? This picture was championed by Khalatnikov and Lifshitz (KL) [1], who argued that the most general solution of Einstein's equations would describe a non-singular contracting and re-expanding universe. Penrose's singularity theorem [2] (see also Hawking [3]) dealt a fatal blow to this idea. The key new concept introduced by Penrose was that of a trapped surface—a compact, 2-dimensional surface \mathcal{T} such that both outward and inward directed systems of null geodesics emanating orthogonally from \mathcal{T} toward the future are converging at all points on \mathcal{T} . The existence of such a surface, combined with mild assumptions about the global structure of spacetime and about its matter content (the null energy condition), makes singularities in the future of \mathcal{T} unavoidable. More precisely, at least one future-directed null geodesic emanating orthogonally from \mathcal{T} must terminate at a singularity. Trapped surfaces do exist in all contracting open FLRW models (any sphere of sufficiently large radius is a trapped surface), and it appears that such surfaces must exist even if the assumption of homogeneity and isotropy is lifted and matter is distributed in clumps. Thus, barring nontrivial global properties or exotic matter content, a contracting infinite universe must be singular and the KL picture cannot be right.

A time-reversed version of this argument can be applied to an expanding infinite universe. Such a universe necessarily contains an anti-trapped surface on which all past-directed null geodesics are converging. Then at least one of these geodesics must terminate at a singularity in the past.

¹ More generally, in a D dimensional spacetime, a trapped surface will be D-2 dimensional.

On the other hand, the Penrose theorem tells us very little about the location or the nature of the singularities. Suppose, for example, that the KL picture of a clumpy contracting and re-expanding universe is basically correct, except some of the clumps merge in the course of contraction and turn into black holes. Some future-directed null geodesics will then enter black holes and terminate at singularities, so this modified KL scenario is consistent with the theorem and cannot be excluded.

Another interesting example is the spontaneous inflation scenario of Carroll and Chen [4] (CC), which starts with a typical² initial state on an infinite spacelike Cauchy surface $S.^3$ Assuming a stable vacuum with a positive energy density, it is then argued that the resulting spacetime will approach a vacuum de Sitter state in both time directions. The universe will be contracting in the asymptotic past and expanding in the asymptotic future, with the thermodynamic arrow of time pointing away from S. The asymptotic de Sitter regions in the past (future) of S contain trapped (anti-trapped) surfaces, and the singularity theorem requires that some future and past-directed null geodesics must be incomplete. But once again a contradiction with the theorem is avoided if some black holes are formed in the future of S and some time-reversed black holes (i.e. white holes) are formed in the past of S. The resulting spacetime structure is illustrated in Fig. 1.

The purpose of the present paper is to extend the Penrose singularity theorem to cosmological spacetimes where black holes are allowed to form from non-singular initial data. Assuming the cosmic censorship hypothesis [7], for generic initial data, all singularities are hidden behind event horizons. There is also a stronger version of the hypothesis [8] which says that singularities cannot even lie on the horizon; this rules out "thunderbolt"-type situations where a null singularity expands outward at the speed of light. For our purposes, we will consider all of these to be "naked" singularities; a more precise definition will be given in section III.

We shall assume that black hole horizons are sufficiently sparse that they do not completely surround the trapped surface \mathcal{T} .⁴ As an example of a situation where \mathcal{T} is completely

² CC use the word "generic", but since this word also has some more precise meanings in general relativity proofs [5, 6], we substitute the word typical.

³ The assumption that the Cauchy surface is infinite is stated on p.19 of Ref. [4].

⁴ For asymptotically flat spacetimes which are also "asymptotically predictable", trapped surfaces \mathcal{T} are required to lie completely inside of a black hole horizon [5, 6]. However, this theorem does not necessarily apply to cosmological settings.

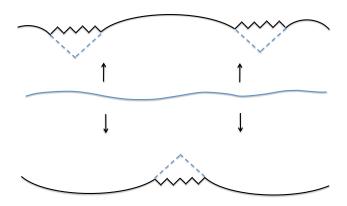
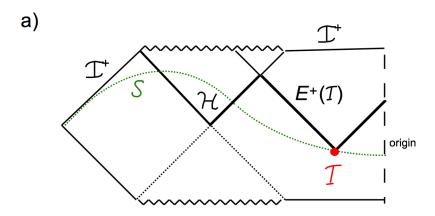


FIG. 1: Spacetime diagram representing the spontaneous inflation scenario. The Cauchy surface is shown by a blue curved line. Black hole (and time-reversed black hole) singularities, shown by zigzag lines, are hidden behind horizons (dashed lines). The arcs represent spacelike future and past infinities of the inflating regions. Thermodynamic arrows of time are indicated by arrows.

surrounded by the horizon, consider the "bag of gold" spacetime [9], where a contracting and re-expanding asymptotically de Sitter space is connected to an asymptotically Minkowski space by an Einstein-Rosen bridge. The causal diagram for this spacetime is shown in Fig. 2(a). The spacelike surface \mathcal{S} indicated by a dotted green line is a Cauchy surface for this spacetime. This surface has the geometry of a plane which is connected by a neck to a sphere. As this initial data is evolved either to the past or to the future, a black hole is formed in the neck region. The horizon bounding this black hole has two components, facing the de Sitter and Minkowski sides, respectively. The black hole thus has the topology of a hollow sphere (a black 2-brane). Now, the contracting part of de Sitter contains trapped surfaces; one such surface is indicated by a red dot in the figure. This surface \mathcal{T} , which is shown in Fig. 2(b) as it appears in a spacelike slice \mathcal{S} containing \mathcal{T} , is completely surrounded by the horizon.

On the other hand, a set of isolated compact black holes, each with a connected horizon, cannot completely surround a surface. One could also imagine a more complicated "black



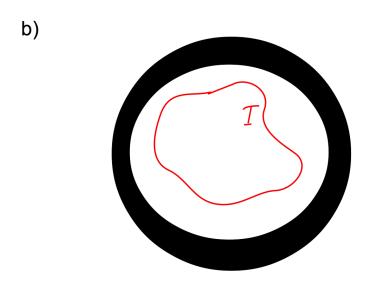


FIG. 2: (a) A causal diagram of the "bag of gold" spacetime, where asymptotically de Sitter and Minkowski regions are connected by an Einstein-Rosen bridge. The green dotted line indicates a spacetime slice \mathcal{S} , which is a Cauchy surface for this spacetime. The black hole horizon \mathcal{H} and the future light cone of the trapped surface \mathcal{T} are joined to form the surface \mathcal{E} , which is shown by a thick black line. (b) The trapped surface \mathcal{T} as it appears in a spacelike slice \mathcal{S} containing \mathcal{T} . The black ring represents the black hole, with its inner and outer boundaries representing horizon components on the de Sitter and Minkowski sides, respectively. The surface \mathcal{T} is completely surrounded by the black hole horizon.

system" consisting of e.g. a network of black strings, which would still not completely surround \mathcal{T} if there are "holes" through which a timelike worldline could escape without falling across the horizon. In situations like these, we will show that there must be additional (i.e. naked) singularities besides those which are inside of the black holes. If we stipulate that there are no naked singularities, this is a contradiction. Thus we are forced to drop one of the other assumptions, concluding that either (i) the universe is actually spatially compact, or (ii) the singularities are so extensive that each trapped surface \mathcal{T} is either inside, or else completely surrounded by, an event horizon.

We shall also prove a second version of this theorem, where instead of assuming that \mathcal{T} is not surrounded by the horizon, we require that all black holes originate at some finite time and have horizons with just one connected component. In this case we again get a contradiction with the existence of a trapped surface.

These theorems place restrictions on the KL and CC scenarios, even when modified to include isolated black hole singularities. They show that the singularities appearing in a spatially noncompact universe must be of a more "cosmological" character than black hole singularities. We cannot show that every causal curve encounters a singularity—a counterexample is the "bag of gold" spacetime of Fig. 2. However, if you start inside a trapped surface, then successful evasion of the singularity requires entering a compact bubble universe which is completely surrounded by event horizons.

In the next Section, we briefly outline the proof of the standard Penrose singularity theorem; the proofs of our theorems follow the same pattern. Our extended theorems are formulated and proved in Section III. Finally, our results are summarized and discussed in Section IV.

We use standard notation in which $I^{\pm}(X)$ is the chronological future/past of a set X, $J^{\pm}(X)$ is the causal future/past of X (which also includes lightlike separated points), and ∂X is the boundary of X.

II. THE PENROSE THEOREM

In this Section we shall outline the standard proof of the Penrose theorem. A variety of singularity theorems have been proved in the literature, but the number of basic ideas and methods of proof is rather limited. For a comprehensive review of the subject, see Refs. [5, 6].

The Penrose theorem says that spacetime which (a) has a non-compact, connected Cauchy surface \mathcal{S} , (b) obeys the null convergence condition, and (c) contains a trapped surface \mathcal{T} must be null incomplete to the future. Specifically, at least one of the future-directed null geodesics emanating orthogonally from \mathcal{T} must be incomplete.

The existence of a Cauchy surface implies a rather simple causal structure of spacetime, namely that spacetime is globally hyperbolic.⁵ Here, we are not going to be interested in causality violations, so this is a reasonable assumption to adopt.

The null convergence condition (NCC) requires that $R_{\mu\nu}N^{\mu}N^{\nu} \geq 0$ for all null vectors N^{μ} , where $R_{\mu\nu}$ is the Ricci tensor. Combined with Einstein's equations, NCC is equivalent to the null energy condition (NEC), requiring that $T_{\mu\nu}N^{\mu}N^{\nu} \geq 0$ for all null N^{μ} , where $T_{\mu\nu}$ is the energy-momentum tensor. NEC is normally satisfied for reasonable classical matter models that are minimally coupled to the metric. It can be violated by quantum effects, but such violations are localized and restricted by "quantum inequalities" (see, e.g. Ref. [11] and references therein).⁶ We shall disregard possible violations of NCC in what follows.

Although the Penrose theorem is normally regarded as a "singularity theorem", by focusing on premise (a) one can instead cast it as a theorem which requires the universe to be spatially compact given certain assumptions. If one assumes that our universe was singularity-free to the past, then the time-reverse of the "open universe" theorem implies that the universe must in fact be spatially compact. Note that if a cosmology admits a spatially compact but nonsimply connected Cauchy surface \mathcal{S} , we can move to the universal cover of \mathcal{S} to apply the theorem [12]. Thus the theorem places restrictions on open or flat cosmologies, even if the universe has nontrivial topology. However, it does not violate the theorem to include a seemingly open universe inside of an expanding bubble within a spatially closed universe.

Globally hyperbolic spacetimes are those for which hyperbolic wave equations have well-defined initial value problems. They satisfy two requirements: (i) there are no closed causal curves, and (ii) for any two points p and q, the region $J^+(p) \cap J^-(q)$ is compact. (Traditionally, the definition of global hyperbolicity included a stronger assumption than (i), known as "strong causality", but these two definitions of global hyperbolicity have been proven to be equivalent [10].) Condition (i) forbids time travel paradoxes, and condition (ii) ensures that no information leaks in or out of the spacetime. In particular, it rules out "timelike" singularities, which are forbidden by certain versions of the cosmic censorship hypothesis [8].

⁶ Many of the results based on NCC can also be derived using weaker conditions, such as the integral convergence condition [12, 13], repeated integral convergence condition [14], or the generalized second law [15].

The standard proof of the Penrose theorem is by contradiction:

Proof. Suppose the conditions (a)-(c) are satisfied and the spacetime is null complete to the future. Let \mathcal{T} be a trapped surface. The future light cone of \mathcal{T} , $E^+(\mathcal{T})$, can be defined as the boundary of its future,

$$E^{+}(\mathcal{T}) = \partial I^{+}(\mathcal{T}). \tag{1}$$

In the vicinity of \mathcal{T} , the light cone is comprised of the two future-directed sheets of null geodesics emanating from \mathcal{T} . The null geodesics on $E^+(\mathcal{T})$ converge, and it follows from NCC and future null completeness that each of these geodesics comes to a conjugate point (that is, crosses nearby geodesics on $E^+(\mathcal{T})$) in a finite affine parameter time. After crossing, the geodesics do not stay on the light cone and enter its interior. Since this happens to all geodesics in a finite affine time, the light cone must be compact. And since $E^+(\mathcal{T})$ is a boundary of a set $(I^+(\mathcal{T}))$, it must have no boundary. It must also be achronal, which means that no two points on $E^+(\mathcal{T})$ can be connected by a timelike curve. (Otherwise, one of these points would be in the interior of $I^+(\mathcal{T})$.)

Now, the existence of a compact, edgeless, achronal hypersurface is inconsistent with a non-compact, connected Cauchy surface. In order to see this, consider a smooth timelike vector field $V^{\mu}(x)$ whose integral curves cross the Cauchy surface \mathcal{S} exactly once. (The existence of such a field follows from the fact that \mathcal{S} is a Cauchy surface.) Since the light cone $E^+(\mathcal{T})$ is achronal, the integral curves of V^{μ} can cross it no more than once. Thus they define a continuous one-to-one map $E^+(\mathcal{T}) \to \mathcal{S}$. Such a map, however, is possible only if \mathcal{S} is itself compact. Hence, with the assumptions (a)-(c) the spacetime cannot be future null-complete: at least one of the geodesics on $E^+(\mathcal{T})$ must terminate at a singularity before reaching a conjugate point.

III. INCLUDING BLACK HOLES

We will now show that the Penrose theorem can be extended to allow for black hole formation. Like the original Penrose theorem, our result can be applied in any spacetime dimension D > 2. We shall use a generalized notion of black holes, to include dynamically

⁷ This step uses the assumption that spacetime is connected, because otherwise \mathcal{S} could contain multiple components, only some of which are noncompact.

evolving black brane-like solutions with arbitrary dimension and topology.⁸ We will also include objects like the Einstein-Rosen bridge bounded by two or more disconnected horizons, although later we shall show that excluding such black holes leads to a stronger result.

Let us assume that there exist at least some future-infinite (timelike or null) worldlines, and thus a nonempty set of points at future infinity \mathcal{I}^+ . The union of the event horizons of all black holes can then be defined as

$$\mathcal{H} = \partial I^{-}(\mathcal{I}^{+}), \tag{2}$$

with each connected part of \mathcal{H} representing an individual black hole horizon. Similarly one can define past horizons using past infiniinfinityty \mathcal{I}^- .

In the proofs that follow, we will not be interested in the interior regions of the black holes. Consequently, when we assume the existence of a trapped surface \mathcal{T} , we will need to require that it lie at least partly outside of \mathcal{H} . More generally, we can define trapped surfaces so as to include compact D-2 surfaces with a boundary $\partial \mathcal{T} \in \mathcal{H}$.

Let us now discuss the types of singularities which can appear in the spacetime. We will define a singularity to be "naked" if a) it lies in $J^+(p)$ for some point p but b) it also lies in $J^-(\mathcal{I}^+)$, the region which can be seen by an observer outside of any horizons. Criterion (a) excludes garden variety initial singularities, such as appear in the interior of a Schwarzschild white hole, or at the Big Bang of an FLRW model.

In the theorems that follow, we will assume that the spacetime is globally hyperbolic, i.e. that it admits a Cauchy surface. We will also assume that there are no naked singularities which lie on the black hole horizon itself (because of the use of J^- in criterion (b), such singularities would still be naked, even though they are not ruled out by global hyperbolicity). We shall now prove two theorems extending the Penrose's result.

Theorem 1: Let \mathcal{M} be a spacetime that obeys NCC, has a non-compact, connected Cauchy surface \mathcal{S} , and contains some black holes. Let all singularities be in the interior of the event horizon \mathcal{H} , i.e. there are no naked singularities on or outside the horizon.

⁸ In four dimensions, this would in principle allow black brane-like objects of dimension 1 and 2. Note however, that in four dimensional general relativity, unlike in higher dimensions, there are no such stationary brane solutions in asymptotically flat spacetime [16]. A further constraint on 1-branes comes from the topological censorship theorem [17], which in certain contexts requires the topologies of horizons to be simply connected [18].

⁹ Note that in order for \mathcal{T} to be compact, it must be a closed set in the sense of including its own boundary.

Suppose also that \mathcal{M} contains a trapped surface \mathcal{T} outside of the black hole horizons. (In the case where a trapped surface is partly inside and partly outside of \mathcal{H} , we can without loss of generality consider \mathcal{T} to include only the part which is outside.) Then \mathcal{T} must be completely surrounded by an event horizon.

Since event horizons can come into existence, whether or not \mathcal{T} is completely surrounded by an event horizon might depend on the choice of time slice. Therefore, we define "completely surrounded" more precisely to mean that there exists some edgeless achronal slice \mathcal{F} , such that \mathcal{T} is contained in a compact subregion \mathcal{R} of \mathcal{F} , and the boundary of \mathcal{R} is part of \mathcal{H} . This indicates that if one waits long enough, a black hole horizon will form outside of \mathcal{T} . Moreover, because \mathcal{F} is achronal, no causal curve which starts on \mathcal{T} and moves to the future can avoid being enclosed by the horizon. The same applies if you start on any compact spatial region whose edge is \mathcal{T} .

Proof. We again consider the future light cone of the trapped surface, $E^+(\mathcal{T})$. The null geodesics on $E^+(\mathcal{T})$ converge and either reach their conjugate points outside of black hole horizons, or else cross one of the horizons. (Inside of the horizon, a geodesic may either have a conjugate point or else terminate on a singularity in the interior. Since there are no naked singularities, it cannot terminate on a singularity at the event horizon itself.) The part of $E^+(\mathcal{T})$ lying outside of black hole horizons is an achronal compact hypersurface, but now it can have edges at the intersections of $E^+(\mathcal{T})$ with the horizons.

To remedy this situation, we construct a new surface \mathcal{E} by joining $E^+(\mathcal{T})$ with black hole horizons along the intersections, including only the part of $E^+(\mathcal{T})$ which is to the past of \mathcal{H} and vice versa (see Fig. 2). This new surface can also be defined as a boundary of a set,

$$\mathcal{E} = \partial[(\mathcal{M} - I^{+}(\mathcal{T})) \cap I^{-}(\mathcal{I}^{+})]. \tag{3}$$

It follows that \mathcal{E} is edgeless (i.e. it has no boundary, except possibly at infinity or at a singularity). Furthermore, the definitions (1), (2) imply that the surfaces $E^+(\mathcal{T})$ and \mathcal{H} are achronal. In order to show that the combined surface \mathcal{E} is also achronal, we have to make sure that no point on the relevant parts of \mathcal{H} can be connected to the part of $E^+(\mathcal{T})$ outside of horizons by a timelike curve. This is indeed the case, since all future-directed timelike curves originating on \mathcal{H} are contained entirely inside the horizons, and all future-directed timelike curves originating on $E^+(\mathcal{T})$ are contained in $I^+(\mathcal{T})$.

If some black holes never cross $E^+(\mathcal{T})$, then \mathcal{E} will contain multiple connected components. In this case we will only be interested in the connected component of \mathcal{E} which includes $E^+(\mathcal{T})$. Let us call this connected component $\tilde{\mathcal{E}}$, and let $\tilde{\mathcal{H}}$ be the part of \mathcal{H} which is in $\tilde{\mathcal{E}}$.

Although $\tilde{\mathcal{E}}$ is edgeless, it cannot be compact. For the existence of an edgeless, compact, achronal hypersurface is inconsistent with a non-compact Cauchy surface \mathcal{S} , as discussed in section II. Therefore, $\tilde{\mathcal{E}}$ must have a "boundary" at infinity, or perhaps at an initial singularity, if \mathcal{H} extends back that far in time.

Nevertheless, $E^+(\mathcal{T}) \cap I^-(\mathcal{I}^+)$ is still compact, and its edge is contained in $\tilde{\mathcal{H}}$. By making the identification¹⁰ $\tilde{\mathcal{E}} = \mathcal{F}$, we see that \mathcal{T} is completely surrounded by the event horizon $\tilde{\mathcal{H}}$.

Apart from the "bag of gold" spacetime, another example of a situation where \mathcal{T} is completely surrounded by the horizon is an effectively (1+1)-dimensional spacetime, where all but one of the spatial dimensions are compactified. In this case, a sufficiently large black hole will always cut space into two disconnected pieces. After the black hole originates, its horizon also splits up into two disconnected, spatially compact pieces. A pair of black holes of this kind are capable of completely surrounding a trapped surface \mathcal{T} . Note that in both examples black hole horizons divide \mathcal{M} into causally disconnected parts.

We now turn to the second theorem, which makes a different assumption about the black holes. Instead of assuming that they do not completely surround the event horizon, one can also get a contradiction by assuming that each black hole horizon originates, not at an initial singularity, but at some "origin set" $\mathcal{O} \in \mathcal{M}$, and that after \mathcal{O} each black hole contains just one connected component.

Theorem 2: Let \mathcal{M} be a spacetime that obeys NCC, has a non-compact, connected Cauchy surface \mathcal{S} , and contains some black holes. Let all singularities be in the interior of the event horizon \mathcal{H} , i.e. there are no naked singularities on or outside the horizon. Suppose also that \mathcal{M} contains a trapped surface \mathcal{T} which is outside of the black holes. (Again, if a trapped surface is partly inside and partly outside, we only consider the piece of \mathcal{T} which is outside.) Then either (a) \mathcal{H} extends all the way back in time to \mathcal{I}^- or to an initial singularity, or else (b) there exist black holes whose horizons contain multiple connected components

¹⁰ We remind the reader that \mathcal{F} is an edgeless achronal slice which was introduced above in the definition of "completely surrounded".

(i.e. some connected component of \mathcal{H} breaks up into multiple connected components of $\mathcal{H} - \mathcal{O}$, as in the case of the Einstein-Rosen bridge discussed above.)

Proof. To prove this result, we must describe in greater detail the structure of the horizon. Each of the horizon generators in $\tilde{\mathcal{H}}$ can be labelled with an affine parameter λ , which we take to be increasing towards the future. Note that generators can enter a future horizon, but they can never leave it [5]. If (a) is false, then each generator must enter the horizon at a finite value of the affine parameter. The space of points in \mathcal{M} at which horizon generators enter the horizon is the origin set \mathcal{O} . \mathcal{O} is an acausal (i.e. strictly spacelike) set, which can contain pieces whose dimensionality ranges from 0 to D-2, D being the spacetime dimension. We now divide the null rays generating $\tilde{\mathcal{H}}$ into two classes:

- i. those which never intersect $E^+(\mathcal{T})$, and
- ii. those which eventually intersect $E^+(\mathcal{T})$ at some finite affine time $\lambda = \lambda_*$.

If we assume that all generators of $\tilde{\mathcal{H}}$ are intersected by the surface $E^+(\mathcal{T})$, as would happen if $\tilde{\mathcal{H}}$ consists entirely of isolated black hole horizons, then class (i) is empty and we can restrict attention to class (ii). But then $\tilde{\mathcal{H}}$ is compact, and the part of $E^+(\mathcal{T})$ outside of black holes is also compact; hence $\tilde{\mathcal{E}}$ is compact. We thus conclude that $\tilde{\mathcal{E}}$ is an edgeless, compact, achronal hypersurface. The existence of such a hypersurface is inconsistent with a non-compact Cauchy surface \mathcal{S} , so we obtain a contradiction.

The alternative is to suppose that class (i) is *not* empty. In this case $\tilde{\mathcal{E}}$ is noncompact, since each generator in class (i) extends infinitely far to the future. Although $\tilde{\mathcal{E}}$ is edgeless, it still has a "boundary at infinity", which is part of \mathcal{I}^+ . Hence $\tilde{\mathcal{E}}$ is not a *complete* achronal surface, i.e. not every inextendible causal curve will intersect it.

Note that because $E^+(\mathcal{T}) \cap I^-(\mathcal{I}^+)$ is compact, its edge must also be compact. This edge is the locus of points for which $\lambda = \lambda_*$. Since this edge is shared with generators in class (ii), it follows that the class (ii) piece of $\tilde{\mathcal{H}}$ is compactly generated.

Next we have to consider the way in which the class (ii) generators may be joined to the class (i) part of $\tilde{\mathcal{H}}$. These two parts of the horizon cannot be glued together along some subset G of horizon generators, because that would require that for generators g' near G on the class (ii) side, that $\lambda_*(g') \to +\infty$ as $g' \to g \in G$. But then the boundary at $\lambda = \lambda_*$ would be noncompact, which is disallowed.

In other words, if $E^+(\mathcal{T})$ intersects any generator of $\tilde{\mathcal{H}}$, it must also intersect every other generator in the same connected component of $\tilde{\mathcal{H}} - \mathcal{O}$. This means that whenever $E^+(\mathcal{T})$ hits a black hole whose horizon has just one connected component after it collapses, then it completely wraps around the black hole. It cannot intersect some of the generators but not others.

Hence, the class (i) generators must be joined to the class (ii) generators at \mathcal{O} , where they enter the horizon. Thus the set \mathcal{O} will generally divide $\tilde{\mathcal{H}}$ into a number of connected components, some of which are type (i) and others type (ii). Thus, on the assumption that (a) is false, we have shown that (b) is true.

We further note that if a simply connected spacetime has a black hole with two or more connected components (i.e. if it satisfies (b)), then after the formation of the black hole, the exterior of the horizon will be divided into multiple causally disconnected regions. That is, if two components $\tilde{\mathcal{H}}', \tilde{\mathcal{H}}'' \subset \tilde{\mathcal{H}} - \mathcal{O}$ join one another at $\mathcal{O}' \subset \mathcal{O}$, then the parts of \mathcal{M} facing different components of the horizon are causally disconnected from one another. More precisely, the region $\mathcal{M} - (J^+(\mathcal{H}) \cup J^-(\mathcal{O}'))$ must split into multiple components, none of whose points are connected by a causal curve in \mathcal{M} . This is because they are connected, but not causally connected, through the black hole horizon. Because the spacetime is simply connected, they cannot also be causally connected through some path which does not cross through the black hole horizon.¹¹

IV. SUMMARY AND DISCUSSION

The two theorems we proved in Section III are similar to Penrose's singularity theorem [2]. As in Penrose's theorem, we assume that the spacetime obeys the null convergence condition, has a non-compact connected Cauchy surface, and includes a trapped surface. Assuming in addition that the trapped surface lies (at least partly) outside the black hole horizons and is not completely surrounded by horizons, we proved in Theorem 1 that the spacetime must contain naked singularities on the horizon. Thus, unlike the Penrose's original theorem, our theorem provides some information about the nature and the location of the singularities.

¹¹ If spacetime is not simply connected, we can pass to the "universal cover" of the spacetime, which is simply connected. This universal cover must then have causally disconnected regions.

The examples we discussed to illustrate how a trapped surface \mathcal{T} can be surrounded by a horizon suggested that this can only happen when \mathcal{T} is located in a spatially compact region which is causally disconnected from the rest of spacetime by black hole horizons. This led us to Theorem 2, which shows that if the black hole horizon surrounding \mathcal{T} originates at some finite time (rather than at an initial singularity or past infinity), then it must have multiple connected components. But if the horizon does have multiple connected components, then on a simply connected spacetime, once the black hole appears it must divide the spacetime into causally disconnected pieces.

Another mechanism that can generate singularities out of non-singular initial data is the formation of negative-energy, Anti-de Sitter (AdS) bubbles. The interiors of such bubbles have the geometry of open (k = -1) FLRW universes, and the negative energy density of the vacuum causes them to contract and collapse to a big crunch [19]. The resulting singularity is contained within the future light cone of the nucleation center (the 'bubble cone'). Since these count as black hole horizons in (2), Theorems 1 and 2 already cover this case.

To illustrate how these theorems can be used to rule out cosmological models, we consider the Carroll-Chen (CC) scenario as an example (see also [20]). The starting point of this scenario is that some typical (or "generic"), regular boundary conditions are specified on an infinite spacelike Cauchy surface \mathcal{S} . This initial state is then evolved both to the past and to the future of \mathcal{S} . Assuming that the lowest-energy (true) vacuum has a positive energy density, CC argue that (i) the universe approaches an empty true-vacuum de Sitter configuration in both time directions, and that (ii) the resulting spacetime is non-singular, apart from isolated black holes that may be formed in the future and/or in the past of \mathcal{S} (see Fig. 1).

If (i) is true, then the asymptotic de Sitter region to the past of S necessarily contains trapped surfaces. For example, any spherical surface of radius greater than the de Sitter horizon is trapped. Then, assuming that NCC is satisfied, it follows from Theorems 1 and 2 that (ii) must be false—there are either naked singularities, or else the "black hole" singularities are sufficiently extensive to divide the spacetime into multiple compact regions.

In order to save the CC scenario, one must change one of the assumptions. For example, one could suppose that the typical state on S, rather than evolving to an isolated black hole surrounded by inflating regions, instead evolves mostly to singular configurations, with isolated regions which inflate. In that case, the isolated regions will be surrounded by one,

topologically complicated "black hole", and neither Theorem 1 nor Theorem 2 will apply.

Alternatively, one could suppose that the universe actually started with a spatially compact Cauchy surface S. In this case, there would be no guarantee that any inflating regions would exist in a typical state. If S were sufficiently large, it is conceivable that (for some definition of "typical" states) there would be a high probability of containing at least one inflating region. However, some arguments against this possibility have been given in [20].

Acknowledgements

This work was supported in part by the National Science Foundation grant PHY-1213888 and by the Templeton Foundation (A.V.), and by the Simons Foundation and the National Science Foundation grant PHY-1205500 (A.W.).

[1] E.M. Lifshitz and I.M. Khalatnikov, "Investigations in relativistic cosmology" Adv. Phys. 12, 185 (1963).

- [3] S.W. Hawking, "Occurrence of Singularities in Open Universes" Phys. Rev. Lett. **15**, 689 (1965).
- [4] S. Carroll and J. Chen, "Spontaneous inflation and the origin of the arrow of time", arXiv:0410270 [hep-th].
- [5] S.W. Hawking and G.F.R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge, 1973).
- [6] R.M. Wald, General Relativity (University of Chicago Press, Chicago, 1984).
- [7] R. Penrose, "Gravitational Collapse: The Role of General Relativity", Rivista del Nuovo Cimento, Numero Speziale I, 257 (1969), reprinted as General Relativity and Gravitation, 34, 1141 (2002).
- [8] R. Penrose, "The question of cosmic censorship", J. Astrophys. Astr., 20, 233 (1999), http://adsabs.harvard.edu/full/1999JApA...20..233P.

^[2] R. Penrose, "Gravitational Collapse and Space-Time Singularities", Phys. Rev. Lett. 14, 57 (1965).

- [9] J. A. Wheeler, in *Relativity, Groups, and Fields*, eds. B. S. DeWitt and C. M. DeWitt (Gordon and Breach, New York, 1964).
- [10] A.N. Bernal, M. Sanchez, "Globally hyperbolic spacetimes can be defined as 'causal' instead of 'strongly causal'", Class. Quant. Grav. 24, 745 (2007), arXiv:gr-qc/0611138v1.
- [11] L.H. Ford, "The Classical Singularity Theorems and their Quantum Loopholes", Int. J. Theor. Phys. 42, 1219 (2003).
- [12] F.J. Tipler, "Energy conditions and spacetime singularities" Phys. Rev. **D17**, 2521 (1978).
- [13] T. A. Roman, "Quantum Stress Energy Tensors and the Weak Energy Condition," Phys. Rev. D 33, 3526 (1986); "On the 'Averaged Weak Energy Condition' and Penrose's Singularity Theorem," Phys. Rev. D 37, 546 (1988).
- [14] A. Borde, "Geodesic Focusing, Energy Conditions and Singularities", Class. Quant. Grav. 4, 343 (1987).
- [15] A. Wall, "The generalized second law implies a quantum singularity theorem", Class. Quantum Grav. **30**, 165003 (2013), arXiv:1010.5513.
- [16] S. Hawking, "Black holes in general relativity", Comm. Math. Phys, 25, 152.
- [17] J.L. Friedman, K. Schleich, and D.M. Witt, "Topological censorship", Phys. Rev. Lett. 71, 1486 (1993).
- [18] T. Jacobson, S. Venkataramani, "Topology of Event Horizons and Topological Censorship", Class. Quant. Grav. 12, 1055 (1995), arXiv:gr-qc/9410023.
- [19] S. R. Coleman and F. De Luccia, "Gravitational Effects on and of Vacuum Decay," Phys. Rev. D21, 3305 (1980).
- [20] A. Vilenkin, "Arrows of time and the beginning of the universe" Phys. Rev. D88, 043516 (2013), arXiv:1305.3836.