Gravitational instability of an extreme Kerr black hole

James Lucietti[1,*] and Harvey S. Reall[2,†]

[1]*School of Mathematics and Maxwell Institute of Mathematical Sciences, University of Edinburgh, King's Buildings, Edinburgh EH9 3JZ, United Kingdom*
[2]*Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, United Kingdom*

Aretakis has proved the existence of an instability of a massless scalar field at the horizon of an extreme Kerr or Reissner-Nordström black hole: for generic initial data, a transverse derivative of the scalar field at the horizon does not decay, and higher transverse derivatives blow up. We show that a similar instability occurs for linearized gravitational, and electromagnetic, perturbations of an extreme Kerr black hole. We show also that the massless scalar field instability occurs for extreme black hole solutions of a large class of theories in various spacetime dimensions.

I. INTRODUCTION

Extreme (zero temperature) black holes are of special interest because they do not emit Hawking radiation. Hence, they are expected to have a simpler description in any candidate theory of quantum gravity. This expectation has been realized within string theory, which has been used to give a statistical mechanics derivation of the Bekenstein-Hawking entropy for certain supersymmetric (therefore extreme) black hole solutions to various supergravity theories [1]. Recently, there has been considerable interest in the proposal that an extreme Kerr black hole can be described by a conformal field theory [2].

Given their importance, it is natural to ask: are extreme black holes stable? We will say that an extreme black hole is stable if any perturbation that is small initially remains small for all time and, at late time, "settles down" to a stationary perturbation corresponding to a small variation of parameters within the family of stationary black hole solutions to which the extreme black hole belongs. (Such a variation of parameters generically makes the black hole slightly nonextreme.)

A heuristic argument suggests that extreme black holes might be classically unstable [3]. Near-extreme black holes usually possess an inner horizon which is believed to be unstable. In the extreme limit, the inner and outer horizons coincide, which suggests that the outer (i.e., event) horizon might be unstable in this limit.[1]

Before discussing the stability of extreme black holes, we will review briefly some stability results for nonextreme black holes. Consider a massless scalar field ψ in the Schwarzschild spacetime. The scalar field can be regarded as a toy model for the more interesting case of linearized gravitational perturbations. Pick a spacelike hypersurface Σ_0 which intersects the future horizon $\mathcal{H}^+$ and extends to null or spacelike infinity. Prescribe initial data for the scalar field on Σ_0 which vanishes at an appropriate rate at infinity. Let Σ_τ denote the surface obtained by translating Σ_0 into the future a parameter distance τ along the orbits of the timelike Killing vector field. It has been proved (see Ref. [4] for a review) that the scalar field decays outside $\mathcal{H}^+$ and also in a neighbourhood of $\mathcal{H}^+$. In particular, along the horizon, ψ and all its derivatives decay at least as fast as certain negative powers of τ. Similar stability results have been achieved for a massless scalar field in a nonextreme Reissner-Nordström [5] or nonextreme Kerr [6] spacetime.

Consider now the case of an extreme black hole. Recently, strong evidence for the existence of a classical *instability* has been obtained by Aretakis. He has considered the evolution of a massless scalar field ψ in the background of an extreme Reissner-Nordström black hole. He proved that, for arbitrary initial data specified on a spacelike surface Σ_0 intersecting the future even horizon $\mathcal{H}^+$, ψ decays on and outside $\mathcal{H}^+$ [7]. However, transverse derivatives of ψ do not decay on $\mathcal{H}^+$: if (v, r, θ, ϕ) denote advanced Eddington-Finkelstein coordinates then, for generic initial data, $\partial_r \psi$ does not decay on $\mathcal{H}^+$ and $\partial_r^k \psi$ blows-up as v^{k-1} for large v [8].

Aretakis has also investigated the case of a massless scalar field ψ in an extreme Kerr spacetime. He has proved decay of axisymmetric solutions ψ, on and outside $\mathcal{H}^+$ [9]. However, just as in the Reissner-Nordström case, he finds that, for generic axisymmetric initial data, derivatives of ψ transverse to $\mathcal{H}^+$ do not decay, and higher order transverse derivatives blow-up along $\mathcal{H}^+$ [10].

In this paper, we will consider linearized *gravitational* perturbations of an extreme Kerr black hole. Aretakis's results suggest that such perturbations might exhibit instabilities in extreme black hole spacetimes. We will prove in Sec. II that this is indeed the case. We do this by showing that

---

*j.lucietti@ed.ac.uk
†hsr1000@cam.ac.uk
[1]An extreme rotating black hole also has a quantum mechanical instability involving spontaneous emission of superradiant quanta. We will discuss only classical stability.

104030-1

Aretakis's arguments can be applied to the Teukolsky equation governing linearized gravitational (or electromagnetic) perturbations of Kerr. Our result implies that small linearized gravitational perturbations of an extreme Kerr black hole generically do not settle down to the stationary perturbation corresponding to a small variation of parameters within the Kerr family of solutions. Hence, an extreme Kerr black hole exhibits a linearized gravitational instability.[2]

Section III presents generalizations of Aretakis's work on massless scalar field instabilities. We prove that his nondecay result can be extended to *any* extreme black hole and that his blow-up result extends to extreme black hole solutions of a large class of theories in various dimensions.

II. GRAVITATIONAL INSTABILITY OF EXTREME KERR

A. Naive instability

Before we introduce our generalization of Aretakis's work, we will discuss a more obvious candidate instability of an extreme black hole.

Consider a Kerr-Newman (KN) black hole in Einstein-Maxwell theory. Take an initial spacelike surface Σ_0 as described above, i.e., intersecting $\mathcal{H}^+$ and extending to infinity. We assume that Σ_0 extends a finite distance behind $\mathcal{H}^+$. Initial data specified on Σ_0 uniquely determines the black hole solution in the future domain of dependence $D^+(\Sigma_0)$. This region includes those parts of the black hole exterior and event horizon which lie to the future of Σ_0. If the black hole is nonextreme, it is believed that the solution is nonlinearly stable against arbitrary small perturbations of the initial data on Σ_0.

This does not seem to be the case for an extreme black hole. Consider a perturbation of the data on Σ_0 which corresponds simply to reducing the mass, remaining within the KN family. The effect of this perturbation is drastic: the resulting spacetime is a portion of a super-extreme KN solution, which does not possess an event horizon.

Is this an instability of the extreme KN solution? To answer this, we must decide what initial data is admissible on a surface such as Σ_0. In an extreme black hole, Σ_0 is necessarily geodesically incomplete, terminating either at the singularity or at an inner boundary behind $\mathcal{H}^+$. Usually one does not consider initial data on such a surface since it is not clear whether the incompleteness is physical. Incompleteness may not be a problem if the singularity is hidden behind a marginally outer trapped surface (MOTS), which is the case when perturbing a nonextreme black hole. But in the extreme case, the perturbed initial data we have just described does not contain a MOTS.

In the nonextreme case, we do not have to confront the problem of dealing with a perturbation specified on an incomplete surface; instead we could choose Σ_0 to be complete, either extending into a second asymptotically flat region, or intersecting the matter which collapses to form the black hole. But in the extreme case we have no choice: there are no complete spacelike surfaces Σ_0 which intersect $\mathcal{H}^+$.[3] So how are we to decide which kinds of initial data are admissible on Σ_0?

One possibility is to dictate that initial data with an incomplete Σ_0 is admissible only if the incompleteness is hidden behind a MOTS. Thus extreme KN initial data is admissible but superextreme KN initial data is not. This approach seems unsatisfactory because it simply "defines away" the possibility of a perturbation destroying the MOTS.

Alternatively, consider the case in which the extreme black hole forms by gravitational collapse. For example, it is possible to form an extreme Reissner-Nordström (RN) black hole by spherically symmetric gravitational collapse of charged matter (e.g., see Ref. [12] for collapse of charged shells). In this case, it is natural to impose initial conditions on a complete asymptotically flat hypersurface that does not intersect the horizon, corresponding to a time before the collapse has occurred. For suitable matter, such initial data will satisfy the mass-charge inequality $M \geq |Q|$ [13], which excludes the superextreme perturbation just discussed. This supports the view that this perturbation is not admissible for extreme RN. However, it does not appear possible to exclude the superextreme perturbation of extreme Kerr by this kind of argument.[4]

To summarise: we have observed that the question of stability of an extreme black hole involves subtleties not present in the nonextreme case. These prevent us from determining the admissibility of the superextreme perturbation of extreme Kerr. Nevertheless, in the next section, we will argue that generic admissible initial data will lead to a gravitational instability of extreme Kerr.

B. Teukolsky equation for extreme Kerr

Let $\{\ell, n, m, \bar{m}\}$ be a null tetrad. Using this we can define the Newman-Penrose Weyl scalars Ψ_A, $A = 0, 1, 2, 3, 4$. A transformation $m \rightarrow e^{i\alpha} m$ is called a *spin* and a quantity ψ has *spin-weight* s if $\psi \rightarrow e^{is\alpha} \psi$ under a spin. For example, Ψ_A has $s = 2\text{-}A$. The Kerr spacetime is type D, which means that we can choose the tetrad so that only Ψ_2 is

---

[2]We emphasize that nonextreme Kerr black holes are expected to be stable (at least in vacuum gravity), no matter how small the nonextremality. See Ref. [11] for a discussion of the Teukolsky equation *inside* a Kerr black hole.

[3]One might choose Σ_0 not to intersect the horizon but instead to contain the asymptotic "throat" region of the extreme black hole geometry. But then we still have the problem of deciding which initial data on Σ_0 are admissible, i.e., which boundary conditions should be imposed in the throat region.

[4]One problem is that the (vacuum) mass-angular momentum inequality $M \geq \sqrt{|J|}$ [14] requires axisymmetry, so this inequality cannot exclude the possibility of a spacetime containing a complete hypersurface on which the initial data is superextreme Kerr outside a compact set and nonaxisymmetric inside this set.

nonvanishing. Now consider a linearly perturbed Kerr spacetime. Take the tetrad to be an arbitrary linear perturbation of the one just discussed. Then $\delta\Psi_0$ and $\delta\Psi_4$ (the perturbations in Ψ_0 and Ψ_4) are invariant under infinitesimal diffeomorphisms and infinitesimal changes in the tetrad [15]. Teukolsky showed that the gauge-invariant quantities $\delta\Psi_0$ and $\delta\Psi_4$ each satisfies a second-order wave equation. These two equations take the same form if written in terms of $\delta\Psi_0$ or $\Psi_2^{-4/3}\delta\Psi_4$ respectively [15].

Starting from the Kerr metric in Boyer-Lindquist coordinates (t, r, θ, ϕ), convert to Kerr coordinates (v, r, θ, χ) defined by

$$dv = dt + \frac{r^2 + a^2}{\Delta}dr, \qquad d\chi = d\phi + \frac{a}{\Delta}dr, \quad (1)$$

where $\Delta = r^2 - 2Mr + a^2$ (we will not assume extremality yet). This gives a coordinate chart regular across $\mathcal{H}^+$, which is at $\Delta = 0$. Choose the following tetrad for the background Kerr spacetime:

$$\ell = 2(r^2 + a^2)\frac{\partial}{\partial v} + 2a\frac{\partial}{\partial \chi} + \Delta\frac{\partial}{\partial r},$$

$$n = -\frac{1}{2(r^2 + a^2\cos^2\theta)}\frac{\partial}{\partial r}, \qquad (2)$$

$$m = \frac{1}{\sqrt{2}(r + ia\cos\theta)}\left(ia\sin\theta\frac{\partial}{\partial v} + \frac{\partial}{\partial \theta} + \frac{i}{\sin\theta}\frac{\partial}{\partial \chi}\right).$$

The vector fields ℓ and n coincide with the principal null directions, with ℓ tangential to $\mathcal{H}^+$. This tetrad is regular in a neighbourhood of $\mathcal{H}^+$ except at $\theta = 0, \pi$. By performing a spin one can introduce a new tetrad which is regular at either $\theta = 0$ or $\theta = \pi$, but it is not possible to define a tetrad which is globally regular with ℓ, n aligned with the principal null directions. Instead one has to work with different tetrads related by spins on coordinate chart overlaps (a spin with $\alpha = \pm\chi$ gives a tetrad regular at $\theta = 0, \pi$). This is not a problem because the Teukolsky equation can be written in a form which is manifestly covariant under spins [16] although we will not use this form here.

For the above choice of tetrad and coordinates, the Teukolsky equation is[5]

$$\frac{\partial}{\partial v}\left\{N(\psi) + 2a\frac{\partial \psi}{\partial \chi} + 2[(1 - 2s)r - ias\cos\theta]\psi\right\}$$

$$= \mathcal{O}\psi - \Delta\frac{\partial^2 \psi}{\partial r^2} - 2(r - M)(1 - s)\frac{\partial \psi}{\partial r} - 2a\frac{\partial^2 \psi}{\partial\chi\partial r}, \quad (3)$$

where we have introduced the smooth vector field

---

[5]Note that $\ell = \Delta\ell^K$, $n = \Delta^{-1}n^K$ where a superscript K refers to the Kinnersley tetrad used in Ref. [15]. This change of tetrad results in a corresponding change in the quantity occurring in the Teukolsky equation: $\psi = \Delta^s\psi^K$. So an easy way to obtain the Teukolsky equation in the tetrad and coordinates used here is to take the equation given in Ref. [15], substitute $\psi^K = \Delta^{-s}\psi$, multiply by Δ^s, and convert to Kerr coordinates.

$$N = 2(r^2 + a^2)\frac{\partial}{\partial r} + a^2\sin^2\theta\frac{\partial}{\partial v} \qquad (4)$$

and the operator

$$\mathcal{O}\psi = -\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) - \frac{1}{\sin^2\theta}\frac{\partial^2\psi}{\partial\chi^2}$$

$$- 2is\frac{\cos\theta}{\sin^2\theta}\frac{\partial\psi}{\partial\chi} + (s^2\cot^2\theta + s)\psi. \qquad (5)$$

Note that N is transverse to $\mathcal{H}^+$. The quantity ψ appearing in the above equation is determined by the value of s: $\psi = \delta\Psi_0$ if $s = 2$ and $\psi = \Psi_2^{-4/3}\delta\Psi_4$ if $s = -2$. For $s = 0$ this equation is just the massless scalar wave equation. Electromagnetic perturbations correspond to $s = \pm1$ [15].

The operator $\mathcal{O}$ appears in the theory of spin-weighted spherical harmonics. Using the notation of Ref. [17], we have $\mathcal{O} = -\eth\bar\eth = -\bar\eth\eth + 2s$. It is readily checked that, with respect to the standard measure on the unit sphere $d\Omega = \sin\theta d\theta \wedge d\chi$, the adjoint of $\eth$ is $-\bar\eth$ and, hence, $\mathcal{O}$ is a non-negative, self-adjoint operator. The eigenfunctions of $\mathcal{O}$ are the spin-weighted spherical harmonics $_sY_{jm}$. These are defined for $j = |s|, |s| + 1, \ldots$ and $|m| \leq j$ with eigenvalues

$$\mathcal{O}(_sY_{jm}) = [j(j + 1) - s(s - 1)](_sY_{jm}). \qquad (6)$$

We will assume that s is an integer; hence, so is j. The eigenspace with zero eigenvalue, which is given by $j = -s$, exists only for $s \leq 0$ and is equal to the kernel of $\bar\eth$. Note $\partial_\chi(_sY_{jm}) = im(_sY_{jm})$, so $_sY_{j0}$ is independent of the azimuthal angle χ.

So far, the discussion applies to any Kerr black hole but now we restrict to an extreme Kerr black hole: $M = a > 0$, with horizon at $r = a$. Let $H(v)$ denote a S^2 cross section of the future event horizon $\mathcal{H}^+$ i.e., a surface with $r = a$ and constant v. We will now follow reasoning similar to that of Aretakis [10] but with spherical harmonics replaced with spin-weighted harmonics.

First consider $s \leq 0$. Restrict (3) to $r = a$ and project onto $_sY_{jm}$ with $j = -s$ and $m = 0$. The terms on the rhs give zero contribution, showing that the quantity

$$I_0^{(s)} = \int_{H(v)} d\Omega(_sY_{-s0})^*\{N(\psi) + 2a[(1 - 2s) - is\cos\theta]\psi\} \qquad (7)$$

is independent of v, i.e., it is conserved along $\mathcal{H}^+$. For $s = 0$ this agrees with the conserved quantity found by Aretakis [10]. Let Σ_0 be a spacelike surface whose intersection with $\mathcal{H}^+$ is $H(v_0)$. We are free to specify initial data for ψ on Σ_0. A generic perturbation will have initial data for which $I_0^{(s)}$ is nonzero. Since $I_0^{(s)}$ is conserved, it remains nonzero for all $v > v_0$. It follows that ψ and the $j = -s$ component of its transverse derivative $N(\psi)$ do not both decay along $\mathcal{H}^+$ as $v \to \infty$.

One might question the assertion that generic initial data give nonzero $I_0^{(s)}$. Above, we discussed the difficulties associated to defining data on an incomplete surface Σ_0. Perhaps admissible initial data on Σ_0 always has vanishing $I_0^{(s)}$. To see why not, consider, for simplicity, the case of a massless scalar in extreme RN, instead of extreme Kerr. The results of Ref. [10] (or Sec. III of the present paper) show that, for a massless scalar in extreme RN, there is a conserved quantity I exactly analogous to $I_0^{(0)}$. As discussed above, one can form an extreme RN black hole by spherically symmetric gravitational collapse of charged matter. In this case, one can take Σ_0 to be a *complete* surface which intersects $\mathcal{H}^+$ after the matter has fallen through it and intersects the collapsing matter behind $\mathcal{H}^+$. Let Σ_* be a complete asymptotically flat spacelike surface which does not intersect $\mathcal{H}^+$, i.e., it corresponds to a time before the black hole has formed. It is uncontroversial that we are free to prescribe arbitrary smooth initial data for ψ on Σ_* subject to appropriate boundary conditions at infinity. Cauchy evolution gives a one-to-one correspondence between data on Σ_* and data on Σ_0. Hence, we are free to specify arbitrary data on Σ_0. Such data generically has nonvanishing I. This is for extreme RN but there is no reason why extreme Kerr should be any different. Hence, generic admissible data has nonvanishing $I_0^{(s)}$.

Now, still with $s \leq 0$, act on (3) with the vector field N, set $r = M = a$ and again project onto $_sY_{jm}$ with $j = -s$, $m = 0$. This gives

$$\partial_v J_0^{(s)} = -2(1-s) \int_{H(v)} d\Omega (_sY_{-s0})^* N(\psi), \qquad (8)$$

where

$$
\begin{aligned}
J_0^{(s)}(v) &= \int_{H(v)} d\Omega (_sY_{-s0})^* \{ N(N(\psi)) \\
&\quad + 2a[(1-2s) - is\cos\theta] N(\psi) - a^2\sin^2\theta \mathcal{O}\psi \\
&\quad + 2a^2[4(1-2s) - (1-s)\sin^2\theta]\psi \} \\
&= \int_{H(v)} d\Omega (_sY_{-s0})^* \{ N(N(\psi)) \\
&\quad + 2a[(1-2s) - is\cos\theta] N(\psi) \\
&\quad + 2a^2[2(3-5s) - (4-3s)\sin^2\theta]\psi \} \qquad (9)
\end{aligned}
$$

and the second equality follows from integration by parts and using the identity

$$\mathcal{O}[\sin^2\theta(_sY_{-s0})] = 2[2(s-1) + (3-2s)\sin^2\theta](_sY_{-s0}). \qquad (10)$$

Consider the case in which $\psi \to 0$ along $\mathcal{H}^+$ as $v \to \infty$. Conservation of $I_0^{(s)}$ implies

$$\int_{H(v)} d\Omega (_sY_{-s0})^* N(\psi) \to I_0^{(s)} \qquad (11)$$

as $v \to \infty$ and, therefore,

$$\partial_v J_0^{(s)} \to -2(1-s)I_0^{(s)}. \qquad (12)$$

For generic initial data, $I_0^{(s)} \neq 0$ and, hence, $J_0^{(s)}$ blows up linearly:

$$J_0^{(s)} \sim -(2(1-s)I_0^{(s)})v. \qquad (13)$$

Inspecting $J_0^{(s)}$ it follows that, if $\psi \to 0$ then either $N(\psi)$ or the $j = -s$ component of $N(N(\psi))$ must blow up at least as fast as v as $v \to \infty$ on $\mathcal{H}^+$.

In summary, we have proved that for an axisymmetric[6] perturbation ψ, if $s \leq 0$ then ψ and the $j = -s$ component of its transverse derivative $N(\psi)$ cannot both decay along $\mathcal{H}^+$ as $v \to \infty$. For $s = 0$, it is known that ψ does decay [9], and this seems likely also for $s \leq 0$ (although proving this would involve a detailed global analysis). In this case, the $j = -s$ component of $N(\psi)$ cannot decay and $N(\psi)$ or the $j = -s$ component of $N(N(\psi))$ must blow up at least as fast as v along $\mathcal{H}^+$ as $v \to \infty$. Presumably, one could extend the above analysis to prove that higher derivatives of ψ blow up even faster: for $s = 0$ Aretakis states that the kth transverse derivative blows up as v^{k-1}.

Following Aretakis, we may go further and derive an infinite set of higher order conserved quantities for any s. First differentiate (3) p times with respect to r,[7] set $r = M = a$ and project onto $_sY_{jm}$. The result is

$$
\begin{aligned}
\frac{\partial}{\partial v} \int_{H(v)} &d\Omega (_sY_{jm})^* \frac{\partial^p}{\partial r^p} \{ N(\psi) + 2[(1-2s)r \\
&- ias\cos\theta + ima]\psi \} \\
&= \int_{H(v)} d\Omega (_sY_{jm})^* \left[(j+p+1-s)(j-p+s)\frac{\partial^p \psi}{\partial r^p} \right. \\
&\quad \left. - 2ima\frac{\partial^{p+1}\psi}{\partial r^{p+1}} \right]. \qquad (14)
\end{aligned}
$$

Note the surprising simplification of the rhs with the first three terms on the rhs of (3) reducing to a single term. Now set $m = 0$ and $j = p - s$. Since $j \geq |s|$, we must take $p \geq \max(0, 2s)$. The rhs above is now zero and, hence, we have an infinite set of conserved quantities:

$$
\begin{aligned}
I_p^{(s)} = \int_{H(v)} &d\Omega (_sY_{p-s0})^* \frac{\partial^p}{\partial r^p} \{ N(\psi) + 2[(1-2s)r \\
&- ias\cos\theta]\psi \}. \qquad (15)
\end{aligned}
$$

Note that for $s \leq 0$ we may take $p = 0$ which reduces to our earlier conserved quantity (7).

---

[6]Projecting to $m = 0$ eigenspaces is equivalent to considering axisymmetric perturbations.

[7]We could act p times with N rather than with $\partial/\partial r$. This also leads to a conserved quantity, but it is harder to write down explicitly.

To obtain higher derivative analogues of $J_0^{(s)}$, use Eq. (14) with $p \to p + 1$, $j = p - s$ and $m = 0$, which gives

$$\partial_v J_p^{(s)} = -2(p + 1 - s) \int_{H(v)} d\Omega (_sY_{p-s0})^* N\left(\frac{\partial^p \psi}{\partial r^p}\right), \quad (16)$$

where

$$J_p^{(s)}(v) = \int_{H(v)} d\Omega (_sY_{p-s0})^* \left(4a^2 \frac{\partial^{p+1}}{\partial r^{p+1}} \{N(\psi)\right.$$
$$+ 2[(1 - 2s)r - ias\cos\theta]\psi\}$$
$$\left. - 2(p + 1 - s)a^2\sin^2\theta \frac{\partial^p \psi}{\partial r^p}\right). \quad (17)$$

Note that for $s \leq 0$ and $p = 0$, this again agrees with our earlier formulas (8) and (9).

Now consider $s > 0$. The smallest permitted value of p in $I_p^{(s)}$ is $p = 2s$, so the argument starts from the conserved quantity $I_{2s}^{(s)}$. Generically this will be nonzero, from which it follows that at least one of the following quantities cannot decay along $\mathcal{H}^+$: $\partial_r^{2s-1}\psi$, $\partial_r^{2s}\psi$ and the $j = s$ component of $\partial_r^{2s}(N(\psi))$. Hence, the best one can hope for is decay of ψ and its first $2s$ derivatives and nondecay of $\partial_r^{2s+1}\psi$. In this case, using $[N, \partial_r^p] = -4pr\partial_r^p - 2p(p - 1)\partial_r^{p-1}$, implies that $\partial_v J_{2s}^{(s)} \to -2(s + 1)I_{2s}^{(s)}$, and, hence,

$$J_{2s}^{(s)} \sim -(2(s + 1)I_{2s}^{(s)})v \quad (18)$$

as $v \to \infty$, which implies that $\partial_r^{2s+2}\psi$ must blow up.

Let us now apply these results to linearized gravitational perturbations ($s = \pm 2$). If the extreme Kerr black hole were stable then an arbitrary initial perturbation would settle down to a stationary perturbation corresponding to a small variation of parameters within the Kerr family of solutions. Such a perturbation preserves the type D condition and so has $\delta\Psi_0 = \delta\Psi_4 \equiv 0$. Hence, if the black hole were stable, we could evaluate $I_p^{(s)}$ at large v to deduce $I_p^{(s)} = 0$. It follows that initial data for which one of the $I_p^{(s)} \neq 0$, cannot settle down to such a stationary perturbation and, hence, the extreme Kerr solution has a linearized gravitational instability.

Furthermore, we learn that if $\delta\Psi_4$ decays, then a transverse derivative of $\delta\Psi_4$ generically does not decay along $\mathcal{H}^+$ and certain second transverse derivatives will blow up along $\mathcal{H}^+$. If $\delta\Psi_0$ and its first four derivatives decay, then a fifth transverse derivative generically will not decay, and a sixth transverse derivative will blow up. It appears that the Weyl component perturbation $\delta\Psi_4$ exhibits worse behavior than $\delta\Psi_0$. Note that the former involves two factors of the transverse basis vector field n^a in its definition ($\Psi_4 = n^a \bar{m}^b n^c \bar{m}^d C_{abcd}$), whereas the latter involves only tangential basis vector fields ($\Psi_0 = \ell^a m^b \ell^c m^d C_{abcd}$). This means that Ψ_4 corresponds to the most tangential components of the Weyl tensor ($C \sim \ell m \ell m$) and Ψ_0 to the most transverse ($C \sim n\bar{m}n\bar{m}$). The former is usually

associated with outgoing radiation and the latter with ingoing radiation.

It is natural to ask about the evolution of this linearized instability in the full nonlinear theory. One possibility is that a small initial perturbation becomes large but, nevertheless, the spacetime eventually settles down to a slightly nonextreme Kerr black hole. Another possibility is that the spacetime develops a null singularity instead of a horizon [3].

III. SCALAR FIELD INSTABILITY OF GENERAL EXTREME HORIZONS

In this section, we will extend Aretakis's argument for an instability of a massless scalar field in certain four-dimensional axisymmetric extreme black hole spacetimes [10]. We will show that his nondecay result can be generalized to *any* extreme black hole, and his blow-up result can be generalized to extreme black hole solutions of a large class of theories in various dimensions.

We will work in Gaussian null coordinates [18], which for convenience we now recall. Let (M, g) be a D-dimensional spacetime and $\mathcal{H}^+$ a smooth, degenerate, Killing horizon of a Killing vector field K. Let H_0 be a $D-2$ dimensional spacelike submanifold of $\mathcal{H}^+$ and assume that each orbit of K is isomorphic to $\mathbb{R}$ and intersects H_0 precisely once.[8] The manifold H_0 is called a cross section and below we will assume these are compact. The degeneracy condition means that the Killing vector K is tangent to *affinely* parametrized null generators of $\mathcal{H}^+$: let $\hat{V} \in \mathbb{R}$ be this affine parameter distance from H_0. Let $(\hat{x}^a)$ be coordinates on H_0 containing some point $p \in H_0$. This defines coordinates $(\hat{V}, \hat{x}^a)$ in a tubular neighbourhood of the integral curve of K through p in $\mathcal{H}^+$ (the $\hat{x}^a$ are extended into this neighbourhood by being taken to be constant along integral curves of K).

Now let U be the unique past-directed null vector field on $\mathcal{H}^+$ satisfying $U \cdot K = 1$ and $U \cdot \partial/\partial x^a = 0$. Assign coordinates $(\hat{V}, \hat{\lambda}, \hat{x}^a)$ to the point affine parameter distance $\hat{\lambda}$ along the null geodesic starting at the point on $\mathcal{H}^+$ with coordinates $(\hat{V}, \hat{x}^a)$ with tangent vector U there. These are called Gaussian null coordinates. In these coordinates $K = \partial/\partial\hat{V}$, $U = \partial/\partial\hat{\lambda}$ and it can be shown that the metric is

$$ds^2 = \hat{\lambda}^2 \hat{F} d\hat{V}^2 + 2d\hat{V}d\hat{\lambda} + 2\hat{\lambda}\hat{h}_a d\hat{V}d\hat{x}^a + \hat{\gamma}_{ab}d\hat{x}^a d\hat{x}^b, \quad (19)$$

where all components are smooth functions of $(\hat{\lambda}, \hat{x}^a)$. Degeneracy of the horizon is what implies $g_{\hat{V}\hat{V}} = \mathcal{O}(\hat{\lambda}^2)$. The coordinates $(\hat{x}^a)$ in the above construction are arbitrary; under a change of these coordinates, $\hat{h}_a$ and $\hat{\gamma}_{ab}$

---

[8]These assumptions are satisfied by all known stationary extreme black hole solutions. Such Killing horizons can also arise in nonblack hole spacetimes.

transform as the components of a one-form and a Riemannian metric on H_0.

It is convenient to generalize these coordinates slightly by using a different affine parameter along the geodesics. Define coordinates (V, λ, x^a) by $\hat{V} = V$, $\hat{\lambda} = \Gamma(x)\lambda$ and $\hat{x}^a = x^a$ where Γ is a smooth positive function. The metric becomes

$$ds^2 = \lambda^2 F dV^2 + 2\Gamma dV d\lambda + 2\lambda h_a dV dx^a + \gamma_{ab} dx^a dx^b, \tag{20}$$

where $F = \Gamma^2 \hat{F}$, $h_a = \Gamma \hat{h}_a + \partial_a \Gamma$, $\gamma_{ab} = \hat{\gamma}_{ab}$ are all smooth functions of (λ, x^a). Let $S(V, \lambda)$ denote a surface of constant (V, λ), and D_a the covariant derivative induced on $S(V, \lambda)$. Note that $H(V) \equiv S(V, 0)$ is a cross section of the horizon and $H(0) = H_0$. It turns out that there is a preferred choice for the function Γ[9]:

Lemma 0.—There exists a unique (up to scale), smooth, positive function Γ on H_0 such that $(D_a h^a)|_{\lambda=0} = 0$.

Proof.—On H_0, write $-D_a h^a = -D^2 \Gamma - D_a(\hat{h}^a \Gamma) \equiv \mathcal{L}\Gamma$. We need to show existence of a positive solution of the elliptic partial differential equation $\mathcal{L}\Gamma = 0$. Any second-order smooth linear elliptic operator on a compact manifold possesses a principal eigenvalue μ (which is real and less that or equal to the real part of any other eigenvalue), whose associated eigenfunction ϕ is everywhere positive and unique up to scaling [19]. Integrating $\mathcal{L}\phi = \mu\phi$ over H_0 then implies $\mu \int_{H_0} \phi = 0$ and, hence, since $\phi > 0$ everywhere, $\mu = 0$. Therefore, $\mathcal{L}\phi = 0$ and, hence, taking Γ to be (up to scale) the principal eigenfunction of $\mathcal{L}$ gives the required function.

We will consider a massless scalar field ψ in the above geometry. Initial data is prescribed on the spacelike surface Σ_0 intersecting $\mathcal{H}^+$ and we assume that boundary conditions are imposed so that $\psi = 0$ at infinity. Hence, if stable, ψ should decay along $\mathcal{H}^+$.

Writing out the massless scalar wave equation in the above coordinates gives

$$0 = \Gamma\sqrt{\gamma}\Box\psi$$
$$= \partial_V\left[\sqrt{\gamma}\left(2\partial_\lambda\psi + \frac{\partial_\lambda\gamma}{2\gamma}\psi\right)\right] - \partial_\lambda[\lambda^2\sqrt{\gamma}A\partial_\lambda\psi]$$
$$\quad - \partial_\lambda(\lambda\sqrt{\gamma}h^a\partial_a\psi) - \lambda\partial_a(\sqrt{\gamma}h^a\partial_\lambda\psi)$$
$$\quad + \partial_a(\Gamma\sqrt{\gamma}\gamma^{ab}\partial_b\psi), \tag{21}$$

where $\gamma = \det\gamma_{ab}$, $h^a = \gamma^{ab}h_b$, γ^{ab} is the inverse of γ_{ab}, and we have defined the function

$$A = \frac{F - h_a h^a}{\Gamma}. \tag{22}$$

Integrate the above equation over $S(V, \lambda)$: the final two terms are total derivatives and so drop out, leaving

$$\partial_V \int_{S(V,\lambda)} \sqrt{\gamma}\left(2\partial_\lambda\psi + \frac{\partial_\lambda\gamma}{2\gamma}\psi\right)$$
$$= \partial_\lambda\left\{\lambda^2 \int_{S(V,\lambda)} \sqrt{\gamma}A\partial_\lambda\psi - \lambda \int_{S(V,\lambda)} \sqrt{\gamma}(D_a h^a)\psi\right\}, \tag{23}$$

where in the final term we have integrated by parts. We can now state the first main result of this section:

Lemma 1.—Choose Γ as in Lemma 0. Then the following quantity is a constant along $\mathcal{H}^+$ (i.e., independent of V):

$$I = \int_{H(V)} \sqrt{\gamma}\left(2\frac{\partial\psi}{\partial\lambda} + \frac{\partial_\lambda\gamma}{2\gamma}\psi\right). \tag{24}$$

Proof.—Evaluate (23) at $\lambda = 0$ and use $D_a h^a|_{\lambda=0} = 0$.

Note that $\partial_\lambda\gamma/(2\gamma) = \Gamma\nabla_\mu(\Gamma^{-1}(\partial/\partial\lambda)^\mu)$, where ∇_μ is the spacetime covariant derivative; hence, this is a smooth quantity. It is also worth noting that converting back to Gaussian null coordinates gives

$$I = \int_{H(\hat{V})} \sqrt{\hat{\gamma}}\Gamma[2U(\psi) + (\nabla_\mu U^\mu)\psi]. \tag{25}$$

It is easy to see this conserved quantity agrees with that for extreme RN [10]. We have also checked that it agrees with the conserved quantity (7) (with $s = 0$) for extreme Kerr [10].[10]

Corollary 1.—Generic initial data has $I \neq 0$ and, hence, for such data ψ and $\partial_\lambda\psi$ cannot both decay along $\mathcal{H}^+$ as $v \to \infty$.

This is a nondecay result that applies to *any* extreme black hole. To demonstrate blow-up, we need an extra assumption about the black hole, whose validity we will discuss at the end of this section.

Lemma 2.—Let Γ be a smooth positive function on H_0 as in Lemma 0. Suppose further that $A|_{\lambda=0} = A_0$ where $A_0 \neq 0$ is a constant. Let

$$J(V) \equiv \int_{H(V)} \partial_\lambda\left[\sqrt{\gamma}\left(2\partial_\lambda\psi + \frac{\partial_\lambda\gamma}{2\gamma}\psi\right)\right]. \tag{26}$$

If $\psi \to 0$ along $\mathcal{H}^+$ as $V \to \infty$ and $I \neq 0$, then $J(V)$ blows up linearly: $J(V) \sim A_0 IV$ along $\mathcal{H}^+$ as $V \to \infty$.

Proof.—Act on (23) with ∂_λ and evaluate at $\lambda = 0$ to obtain

$$\partial_V J(V) = 2 \int_{H(V)} \sqrt{\gamma}[A\partial_\lambda\psi - \partial_\lambda(D_a h^a)\psi]. \tag{27}$$

By assumption $\psi \to 0$, so the final term on the rhs of (27) decays and the first term on the rhs asymptotically

---

[9]In all examples known to us, this choice of Γ ensures that h^a is a Killing vector field on H_0. However, we will not assume this.

[10]In particular, we have checked $\Gamma U(\psi)|_{\hat{\lambda}=0} = (4a^2)^{-1}N(\psi)|_{r=a}$ for any axisymmetric function ψ, where N is the vector field (4), and $\Gamma = (1 + \cos^2\theta)/2$ can be read off from the near-horizon geometry (see, e.g., Ref. [20]). It then easily follows that $I = I_0^{(0)}$, where $I_0^{(0)}$ is the conserved quantity (7) with $s = 0$.

approaches $A_0 I$. Therefore, as $V \to \infty$, $\partial_V J(V) \to A_0 I$ and integrating this proves the result.

Corollary 2.—If $\psi \to 0$ along $\mathcal{H}^+$ as $V \to \infty$ for generic initial data then either $\partial_\lambda \psi$ or $\partial_\lambda^2 \psi$ diverges along $\mathcal{H}^+$ as $V \to \infty$ (and if $\partial_\lambda \psi$ diverges then it most do so consistently with constancy of I).

In summary, we have shown that for generic initial data, we must have one of the following possibilities: (i) ψ does not decay along $\mathcal{H}^+$ or (ii) ψ decays, $\partial_\lambda \psi$, does not decay and, subject to the assumption about A of Lemma 2, one of the quantities $\partial_\lambda \psi$, $\partial_\lambda^2 \psi$ blows up as $V \to \infty$ along $\mathcal{H}^+$. The "most stable" outcome consistent with our results is (ii) with $\partial_\lambda \psi$ nondecaying but bounded and $\partial_\lambda^2 \psi$ blowing up.

Let us return to the assumption in Lemma 2. Since this involves a quantity *intrinsic* to the horizon, it can be regarded as an assumption about the near-horizon geometry of the extreme black hole in question (defined by $V \to V/\epsilon$, $\lambda \to \epsilon \lambda$ and $\epsilon \to 0$, see, e.g., Ref. [21]).

This assumption is true for a large class of near-horizon geometries in various dimensions and theories. All extreme black holes solutions known to us satisfy this assumption. For many examples, it follows from the near-horizon AdS$_2$-symmetry theorems proved in Refs. [21,22]. The results of Ref. [21] imply that the assumption is valid (with $A_0 < 0$) for extreme black hole solutions of a class of theories in $D = 4, 5$ dimensions consisting of Einstein gravity coupled to arbitrarily many Abelian vectors and uncharged scalars, assuming that the black hole has D-3 commuting rotational symmetries and that the horizon topology is nontoroidal. Reference [22] determined the near-horizon geometries of extreme Myers-Perry black holes [23], and these also satisfy our assumption with $A_0 < 0$. In that work, it was also shown that the assumption is valid for $D > 5$ extreme vacuum black holes with cohomogeneity-one near-horizon geometries possessing certain non-Abelian rotational symmetry groups.

ACKNOWLEDGMENTS

[1] A. Strominger and C. Vafa, Phys. Lett. B **379**, 99 (1996).
[2] M. Guica, T. Hartman, W. Song, and A. Strominger, Phys. Rev. D **80**, 124008 (2009).
[3] D. Marolf, Gen. Relativ. Gravit. **42**, 2337 (2010).
[4] M. Dafermos and I. Rodnianski, arXiv:0811.0354.
[5] P. Blue and A. Soffer, J. Funct. Anal. **256**, 1 (2009).
[6] M. Dafermos and I. Rodnianski, *Proceedings of the 12th Marcel Grossmann meeting*, edited by T. Damour, R. T. Jantzen, and R. Ruffini (World Scientific, Singapore, 2012).
[7] S. Aretakis, Commun. Math. Phys. **307**, 17 (2011).
[8] S. Aretakis, Annales Henri Poincare **12**, 1491 (2011).
[9] S. Aretakis, J. Funct. Anal. **263**, 2770 (2012).
[10] S. Aretakis, arXiv:1206.6598.
[11] G. Dotti, R. J. Gleiser, and I. F. Ranea-Sandoval, Int. J. Mod. Phys. E **20**, 27 (2011).
[12] K. Kuchar, Czechoslovak Journal of Physics, Section B **18**, 435 (1968); Ch. J. Farrugia and P. Hajicek, Commun. Math. Phys. **68**, 291 (1979).
[13] G. W. Gibbons and C. M. Hull, Phys. Lett. **109B**, 190 (1982).
[14] S. Dain, J. Diff. Geom. **79**, 33 (2008).
[15] S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972); Astrophys. J. **185**, 635 (1973).
[16] J. M. Stewart and M. Walker, Proc. R. Soc. A **341**, 49 (1974).
[17] J. N. Goldberg, A. J. MacFarlane, E. T. Newman, F. Rohrlich, and E. C. G. Sudarshan, J. Math. Phys. (N.Y.) **8**, 2155 (1967).
[18] V. Moncrief and J. Isenberg, Commun. Math. Phys. **89**, 387 (1983).
[19] L. Andersson, M. Mars, and W. Simon, Adv. Theor. Math. Phys. **12**, 853 (2008).
[20] H. K. Kunduri and J. Lucietti, J. Math. Phys. (N.Y.) **50**, 082502 (2009).
[21] H. K. Kunduri, J. Lucietti, and H. S. Reall, Classical Quantum Gravity **24**, 4169 (2007).
[22] P. Figueras, H. K. Kunduri, J. Lucietti, and M. Rangamani, Phys. Rev. D **78**, 044042 (2008).
[23] R. C. Myers and M. J. Perry, Ann. Phys. (Berlin) **172**, 304 (1986).