

NEAR-EXTREMAL BLACK HOLES AT LATE TIMES, BACKREACTED

Shahar Hadar

*Center for the Fundamental Laws of Nature, Harvard University,
Cambridge, MA 02138, USA*

and

*Max Planck Institute for Gravitational Physics (Albert Einstein Institute),
Am Mühlenberg 1, 14476 Potsdam-Golm, Germany*

Abstract

Black holes display universal behavior near extremality. One such feature is the late-time blowup of derivatives of linearized perturbations across the horizon. For generic initial data, this instability is regulated by backreaction, and the final state is a near-extremal black hole. The aim of this paper is to study the late time behavior of such black holes analytically using the weakly broken conformal symmetry of their near-horizon region. In particular, gravitational backreaction is accounted for within the Jackiw-Teitelboim model for near-horizon, near-extremal dynamics coupled to bulk matter.

Contents

1	Introduction	1
2	Extremal behavior from symmetry	4
3	Near-extremal behavior from symmetry	10
4	Backreacted solution within Jackiw-Teitelboim theory	15

1 Introduction

Black holes (BHs) in the limit of vanishing Hawking temperature exhibit behavior reminiscent of critical phenomena, and this has surfaced in various different contexts in BH physics, e.g. [1],[2],[3]. In many-body systems near criticality, such behavior is explained by the emergence of a conformal symmetry. In the gravitational context, the BH's surface gravity vanishes in the extremal limit and the conformal symmetry is geometrically realized near the horizon [4],[5]. Such BHs are called near-extremal since they nearly saturate bounds on certain combinations of the mass, angular momentum and charges, and they bear particular significance, both theoretically and observationally. Astrophysically, neutral BHs with angular momentum close to the Kerr bound are believed to be common in the sky, and some possible observational signatures of these high-spin BHs, relevant for both gravitational-wave and optical observatories, have recently been proposed, for example [6],[7],[8],[9].

There has been much recent progress on understanding the late-time behavior of perturbed extremal BHs. In particular, it was discovered by Aretakis in a series of works [10],[11],[12],[13] that linearized perturbations of a massless scalar field on an extremal Reissner-Nordström (RN) BH background display both stable and unstable properties: perturbations outside the horizon decay at late times, but on the horizon transverse derivatives generally do not decay and may even blow up. This behavior is closely related to Aretakis' discovery of an infinite number of conserved quantities on the horizon, one per mode (ℓ, m) on the transverse sphere, or conversely a conserved function on the sphere. The spherically symmetric mode ϕ_0 of such perturbations, for example, behaves at late times ($v \rightarrow \infty$) on the horizon as

$$\partial_r^k \phi_0 \sim v^{k-1}, \tag{1.1}$$

for smooth initial data supported on the future horizon, where $\{v, r\}$ are ingoing Eddington-Finkelstein coordinates. Several subsequent works on the subject, using different techniques, studied late-time behavior on the horizon for different types of extremal BHs and different

types of perturbation fields [14],[15],[16],[17], and in particular in [18] it was shown that a massless scalar displays Aretakis behavior on *any* extreme BH background. The picture arising from these studies is that the Aretakis behavior is a general feature of extremal BHs, in the following sense: the k -th transverse-to-the-horizon derivative of the perturbation, at late times on the horizon, behaves (no worse than¹)

$$\sim v^{k+\Delta_-}, \quad (1.2)$$

where Δ_- is a negative k -independent, fixed non-universal constant.²

A convenient way to repackage (1.2) (assuming an expansion in the radial coordinate) is to use the following ansatz for the field (assume a scalar field for simplicity):

$$\phi \simeq v^{\Delta_-} F[v(r-M)], \quad (1.3)$$

where $F(z)$ is an arbitrary function which is smooth at $z=0$. It is important to note that this type of behavior is consistent with the results of [20], which showed that for generic initial data in extremal RN the decay and blowup rates (1.2) are saturated (again, modulo the skip phenomenon described in footnote 1). Following [21] (see also [3]), plugging (1.3) into the 1+1 dimensional wave equation for each mode and considering the near-horizon, late-time limit reduces it to an ordinary differential equation for $F(z)$, which yields a *universal* behavior for the field at late times near the horizon. For example, for the spherically symmetric mode ($\Delta_- = -1$) of the above discussed neutral massless scalar around extreme RN:³

$$\phi_0 \simeq \frac{H_0}{v(2+v(r-M)/M^2)}, \quad (1.4)$$

at late times, where H_0 is interpreted as the $\ell=0$ Aretakis constant.

Since the Aretakis analysis concerns linearized perturbations and involves blowup, one must ask whether and how *gravitational backreaction* becomes important. This was addressed in [22] numerically for spherically symmetric perturbations of a massless, neutral scalar field. The conclusions were that the instability persists under backreaction, and that for generic initial data, the BH approaches a new, *non-extremal* solution at late times, and this allows perturbations to exponentially decay at times of the order of the inverse temperature. This

¹in fact, for generic initial data, (1.2) is the precise decay/blowup rate modulo the so-called “skip phenomenon” [19],[20], which in some cases of integer Δ_- , for some values of k , suppresses (1.2) by an additional factor of v^{-1} —causing faster decay or slower blowup.

²which depends on: which mode is being considered (say in a spherical/spheroidal harmonic decomposition), the field’s spin, whether or not the initial data has support on the horizon and the type of extremal BH under study.

³Another universal piece $\sim 1/v$ arises in the case of non-vanishing $\ell=0$ Newman-Penrose constant; here we assume it does vanish.

means that even if the BH is perturbed with a small amplitude, certain quantities (with enough radial derivatives) can grow to parametrically larger values, and in particular not remain small, before eventually decaying.⁴ This makes sense from an outside observer’s point of view: energy is thrown down the hole, and this can raise the (effective) temperature, allowing dissipation; a picture which connects well also with [16] which studied the “transient instability” behavior of linearized perturbations to near-extremal Kerr BHs, finding Aretakis-like growth at intermediate times and eventually exponential decay at later times.

The universal behavior of extreme BHs has a unifying explanation, and as stated above, in gravity as in many-body systems, it is the emergence of a conformal symmetry at criticality. From the gravitational point of view it is realized geometrically, and this can be seen by zooming in on the near-horizon region, where a large degree of symmetry is found: a global $SL(2, \mathbb{R})$ isometry which is, at least in some cases, understood to be enhanced to a local infinite-dimensional asymptotic symmetry group [4],[5],[23]. There is a strong connection between the behavior discovered by Aretakis and this symmetry: the $SL(2, \mathbb{R})$ isometry alone is enough to show that (1.2) had to be the late-time behavior. This was argued in [14] for scalar perturbations in AdS_2 , then elaborated upon and extended to nonaxisymmetric modes with nonvanishing Aretakis constants in extremal Kerr in [21], and further generalized in [3].

For near-extremal BHs, the conformal symmetry is broken, but only weakly. This was understood recently in a very different context, that of the Sachdev-Ye-Kitaev (SYK) model [24],[25], a certain 0+1 dimensional many-body fermionic quantum mechanics with all-to-all coupling, which is solvable in the limit of a large number of degrees of freedom. SYK at low energies has substantial overlap with a seemingly very different theory—the Jackiw-Teitelboim (JT) model [26],[27], a 1+1 dimensional theory of dilaton gravity. As part of this line of work, in [23] the symmetry breaking pattern of JT and its corresponding near- AdS_2 solutions were understood. In particular it was shown that the conformal symmetry, which is exact in the extreme limit, is weakly broken for near-extreme BHs; both explicitly by the small temperature and spontaneously by the choice of reparametrization relating the near-horizon time coordinate to the boundary (proper) time, and this structure practically determines the low-energy effective action. Moreover, and importantly for this work, it was argued in [28], [23] and shown in detail in some specific cases [29],[30],[31] that the JT model provides a universal description of the near-horizon gravitational dynamics of near-extremal BHs, via dimensional reduction.

In this paper, the late-time dynamics of near-extremal BHs is studied analytically, from

⁴Note also that, as shown in [22], one may fine-tune the initial data such that the BH remains extremal, and the instability continues to evolve forever.

the near-horizon perspective. The JT model allows for the description of backreaction effects while remaining in near-AdS₂, as was pointed out in [28]. This enables one to both exploit the (weakly broken) symmetries, and to account for the out-of-equilibrium nature of the BH’s evolution during the process; and this is an *essential* effect to take into account if the perturbation’s energy is comparable to the energy above extremality. From a technical point of view, a crucial point is that the backreaction can be monitored by promoting the boundary to a degree of freedom characterized by a worldline on AdS₂; and the bulk fields, which propagate on the corresponding—locally AdS₂—cutout geometry, couple to gravity only through a local interaction at the boundary. This simplified picture is used here to compute the late-time behavior of perturbations on the horizon and outside it.

We use a boundary-bulk or gravity-matter coupling at the level of the equations of motion which was derived in [23]. It could be interesting to carry out the computation with a coupling at the level of the action, which would presumably require to consider the two-sided BH (which should be equivalent to a Schwinger-Keldysh field doubling of the one-sided case). It could be interesting to compare the boundary behavior derived here to expectations from holography. Including backreaction means taking into account the variation in the BH’s temperature during the evolution: a non-equilibrium process in a putative field theory dual, in which quenching pumps into the system an energy of the order of the initial thermal energy. Finally, it would be very interesting to work out in detail the analogous process in extreme Kerr and its corresponding near-NHEK geometry and spell out in detail the similarities and differences from the spherically symmetric charged case.

In section 2, we review the derivation of late time behavior at extremality, including the Aretakis instability and boundary decay from the near-horizon conformal symmetry. In section 3 we derive in detail the near-extremal characteristic behavior from the (weakly broken) symmetry, including the exponential decay at late time via quasinormal modes and transient behavior at intermediate times. Finally in section 4 we treat the backreacted late-time behavior from the near-horizon point of view using JT gravity.

2 Extremal behavior from symmetry

In this section we will review the argument that the decay and blow up rates of a massless scalar field and its derivatives on the horizon at late times follow from an analysis of the symmetries of the near-horizon region. We focus on this simple case for clarity, but generalizations will be commented upon.

Begin with the Reissner-Nordström solution⁵

$$\begin{aligned} ds^2 &= - \left(1 - \frac{2M}{\hat{r}} + \frac{Q^2}{\hat{r}^2}\right) d\hat{t}^2 + \left(1 - \frac{2M}{\hat{r}} + \frac{Q^2}{\hat{r}^2}\right)^{-1} d\hat{r}^2 + \hat{r}^2 d\Omega^2, \\ \hat{A}_{\hat{t}} &= -\frac{Q}{\hat{r}}. \end{aligned} \quad (2.1)$$

Now define $\hat{\kappa} := \sqrt{1 - Q^2/M^2}$, which describes the BH's deviation from extremality: near-extremal BHs are defined by $\hat{\kappa} \ll 1$, and then $\hat{\kappa}$ is proportional to the BH's dimensionless surface gravity/temperature. Exactly extremal, zero temperature BHs satisfy $\hat{\kappa} = 0$.

In this section we will consider the maximally electrically charged ERN solution, $\hat{\kappa} = 0$ or $Q = M$,

$$\begin{aligned} ds^2 &= - \left(1 - \frac{M}{\hat{r}}\right)^2 d\hat{t}^2 + \left(1 - \frac{M}{\hat{r}}\right)^{-2} d\hat{r}^2 + \hat{r}^2 d\Omega^2, \\ A &= -\frac{M}{\hat{r}} d\hat{t}, \end{aligned} \quad (2.2)$$

where $d\Omega^2$ is the line element on the 2-sphere. A nondegenerate metric describing the near-horizon region can be obtained by introducing a 1-parameter family of coordinate and gauge transformations (see also recent discussion in [32],[33])

$$R = \frac{\hat{r} - M}{\epsilon M}, \quad T = \frac{\epsilon \hat{t}}{M}, \quad A = \hat{A} + d\hat{t}, \quad (2.3)$$

and taking the $\epsilon \rightarrow 0$ limit. The resulting near-horizon solution is

$$\frac{ds^2}{M^2} = -R^2 dT^2 + R^{-2} dR^2 + d\Omega^2, \quad (2.4)$$

$$\frac{A}{M} = -RdT. \quad (2.5)$$

The geometry (2.4) is referred to as the Robinson-Bertotti universe or $\text{AdS}_2 \times \text{S}^2$. In what follows, we will sometimes find it useful to work with different coordinate systems on its AdS_2 factor. Introduce ingoing Eddington-Finkelstein coordinates $\{V, R\}$ by

$$R_* = \int \frac{dR}{R^2}; \quad V := T + R_* = T - 1/R, \quad (2.6)$$

where on the right hand side of the second equation above we made a specific choice for the integration constant. The metric then becomes

$$\frac{ds^2}{M^2} = -R^2 dV^2 + 2dVdR + d\Omega^2. \quad (2.7)$$

⁵We use here gravitational units $G_N = c = 1$. Note also that from here on, and differently from section 1, hatted coordinates and quantities will be related to the full (extreme or near-extreme) RN geometry (2.1),(2.2), while unhatted coordinates and quantities will be related to near-horizon geometries (2.4),(3.2).

Yet another useful coordinate system is the so-called double-null or lightcone coordinates $\{V, U\}$, where we take

$$U := T - R_* = T + 1/R, \quad (2.8)$$

yielding the metric

$$\frac{ds^2}{M^2} = -4 \frac{dV dU}{(V - U)^2} + d\Omega^2. \quad (2.9)$$

The solution (2.4),(2.7) is invariant under the 1-parameter family of coordinate transformations (isometries)

$$\begin{aligned} T - T_\infty &= -\frac{\tilde{R}^2 \tilde{T}}{\tilde{R}^2 \tilde{T}^2 - 1} = -\frac{1 + \tilde{V} \tilde{R}}{\tilde{V}(2 + \tilde{V} \tilde{R})}, \\ R &= \frac{\tilde{R}^2 \tilde{T}^2 - 1}{\tilde{R}} = \tilde{V}(2 + \tilde{V} \tilde{R}), \end{aligned} \quad (2.10)$$

where $\{\tilde{T}, \tilde{R}\}, \{\tilde{V}, \tilde{R}\}$ are the new coordinates with metrics (2.4), (2.7) respectively and T_∞ is an arbitrary parameter.⁶ The transformation (2.10) may be thought of as a composition of an (exponentiated/large) AdS_2 global time translation and Poincaré time translation.

Importantly, even though (2.10) is an isometry, and therefore preserves the metric locally, it may modify the physical situation under consideration since the two coordinate systems in (2.10) cover different patches of the maximal analytic extension of this geometry (see figure 1), and in particular different portions of the boundary. Notably, this transformation will modify the physical situation whenever the AdS_2 spacetime is thought of as a near-horizon limit of a higher dimensional extremal BH, in which case the boundary is thought of as a surface upon which matching to the far-region of the extreme RN geometry is performed, and this defines a preferred choice of boundary.

Now consider a massless scalar ϕ propagating on this geometry. Its equation of motion is $\square\phi = 0$. Decomposing into spherical harmonics

$$\phi = \sum_{\ell m} \phi_{\ell m}(T, R) Y_{\ell m}(\Omega), \quad (2.11)$$

it can be written as

$$(R^{-2} \partial_T^2 + \partial_R R^2 \partial_R - \ell(\ell + 1)) \phi_{\ell m}(T, R) = 0. \quad (2.12)$$

⁶Supplemented by the gauge transformation $a = A + d(-2 \tanh^{-1}(\tilde{R} \tilde{T}))$, the electric field (2.5) is preserved as well.

Assuming separation of variables, or time dependence of the form $\phi \sim e^{-i\omega T}$, shows that for $R \rightarrow \infty$ the modes behave as $\phi_{\ell m \omega} \sim R^{\Delta_{\pm}}$, where

$$\Delta_{\pm} = -1/2 \pm (\ell + 1/2). \quad (2.13)$$

Let us first analyze the $\ell = 0$ case. The d'Alambertian in AdS₂ (2.12) is proportional to the 1 + 1 dimensional Minkowski space wave equation, written in terms of the spatial coordinate $1/R$,

$$\square = (1/R)^2 (-\partial_T^2 + \partial_{(1/R)}^2), \quad (2.14)$$

and therefore the general solution can be written as

$$\phi = f(U) + g(V), \quad (2.15)$$

a superposition of arbitrary rightmoving and leftmoving waves determined by the functions f, g which we will assume are smooth. Using (2.10) gives

$$\phi = f\left(T_{\infty} - \frac{\tilde{R}}{\tilde{V}\tilde{R} + 2}\right) + g\left(T_{\infty} - 1/\tilde{V}\right). \quad (2.16)$$

Imposing the physical requirement that the field decays at late times outside the horizon, $\phi \rightarrow 0$ for $\tilde{T} \rightarrow \infty$, we get $f(T_{\infty}) = -g(T_{\infty})$. This means that, on the horizon as well, $\phi \rightarrow 0$ as $\tilde{V} \rightarrow \infty$. The behavior of radial derivatives on the horizon can be inferred by directly operating on (2.16) with $\partial_{\tilde{R}}^k$ for $k \geq 1$, evaluating at $\tilde{R} = 0$ and taking the $\tilde{V} \rightarrow \infty$ limit:

$$\partial_{\tilde{R}}^k \phi|_{\text{horizon}} \approx f^{(k)}(T_{\infty}) \frac{(-1)^k k!}{2^k} \tilde{V}^{k-1}, \quad (2.17)$$

where $f^{(k)}$ stands for the k -th derivative of f . Hence the Aretakis behavior is recovered from near-horizon conformal symmetry.

Now assume $\ell > 0$. A general solution can be written for large R as

$$\phi = f_+(T) R^{\Delta_+} (1 + \mathcal{O}(R^{-1})) + f_-(T) R^{\Delta_-} (1 + \mathcal{O}(R^{-1})), \quad (2.18)$$

for some smooth f_{\pm} . Imagine we prescribe some initial data on, say, an ingoing null surface as depicted by the dashed line in figure 1. The behavior near the boundary at some future time $T = T_{\infty}$ will be given by the leading order piece (in $1/R$) of equation (2.18). Now, as in the $\ell = 0$ case, we can use the diffeomorphism (2.10) to place the point $(T = T_{\infty}, R = \infty)$ (encircled in figure 1 by the dotted line) at future timelike infinity of the patch described by the coordinates $\{\tilde{V}, \tilde{R}\}$. We will think of this patch as the physical one, i.e. that which describes the near-horizon region of a higher-dimensional extreme BH geometry.

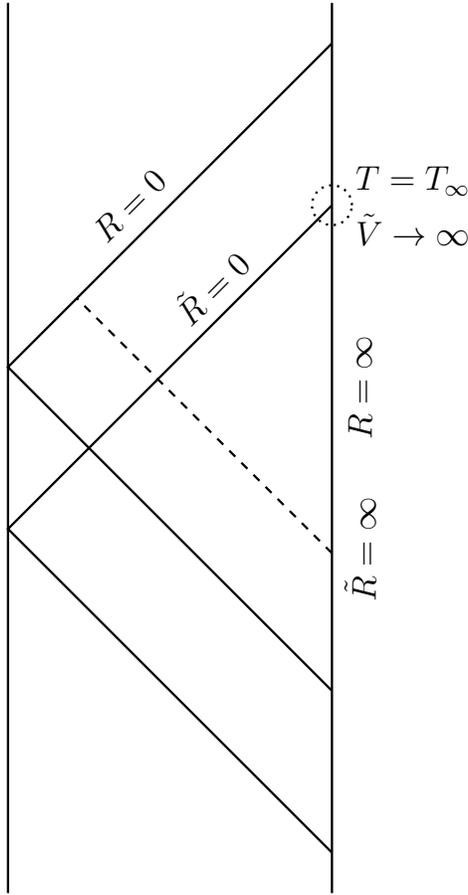


Figure 1: Penrose diagram displaying the isometry (2.10). The coordinates $\{T, R\}$ cover the upper Poincaré patch with metric (2.4), while the coordinates $\{\tilde{T}, \tilde{R}\}$ cover the lower Poincaré patch. The dashed null line represents a possible initial data surface, and the dotted circle highlights the neighborhood of $\tilde{V} \rightarrow \infty$ on which we are focusing. This point is the future endpoint of the future horizon and of the boundary of the patch $\{\tilde{T}, \tilde{R}\}$.

The late-time behavior is only affected by the boundary conditions close to T_∞ . In general they may be of mixed, Robin-type nature, and time-dependent, but we will *assume* that they are close enough to Dirichlet conditions of $\phi|_{\text{bdy}} = 0$ at late times—in particular, that $\partial_T^k f_+(T_\infty) = 0$ for $k < \Re(\Delta)_+ - \Re(\Delta_-) = 2\ell + 1$. It makes sense to pick boundary conditions which satisfy this condition (a much weaker condition than, for example, simply putting Dirichlet $\phi|_{\text{bdy}} = 0$ boundary conditions for any time) since this assumption yields decay rates for the field itself (before discussing the behavior of derivatives) which are consistent with known results on full geometries, both on and off the horizon, as will be seen below in the boundary analysis. The behavior of derivatives is inferred by the symmetry argument presented here.⁷

⁷Note that for cases when there are modes with complex scaling dimensions $\Delta_\pm = 1/2 \pm i\xi$ with real ξ , as is sometimes the case for extreme Kerr or charged perturbations on extreme RN, the only required

Assuming these boundary conditions and plugging (2.10) into (2.18), we find that for $\tilde{V} \rightarrow \infty$, $\tilde{R}\tilde{V}$ fixed, to leading order,

$$\phi = \tilde{V}^{-\ell-1} \left[-f_+^{(2\ell+1)}(T_\infty) \left(1 + \tilde{V}\tilde{R}\right)^{2\ell+1} + f_-(T_\infty) \right] \left(2 + \tilde{V}\tilde{R}\right)^{-\ell-1}, \quad (2.19)$$

which is of the form

$$\sim \tilde{V}^{\Delta_-} g(\tilde{R}\tilde{V}), \quad (2.20)$$

for a function $g(z)$ which is regular at $z = 0$, just as in (1.3). Taking k derivatives and evaluating at the horizon, therefore, one finds

$$\partial_{\tilde{R}}^k \phi \Big|_{\text{horizon}} \rightarrow \tilde{V}^{\Delta_- + k} \quad \text{as } \tilde{V} \rightarrow \infty, \quad (2.21)$$

the Aretakis decay/blow up rates for nonzero Aretakis constants. A similar argument also works more generally for perturbations around various extremal BHs, and different types of perturbations. Note that in the general case where $\Re(\Delta_+) - \Re(\Delta_-)$ is not an integer, the first term on the right hand side of equation (2.19) will be absent.

The significance of the above argument is that it gives insight into why the Aretakis behavior is a universal feature of extremal BHs. Moreover, under the assumption of the above boundary conditions, it gives a remarkably simple way to predict the decay/blow-up rates for any such BH: one only needs to determine the exponent Δ_- which is fixed by the large- R limit of the near-horizon wave equation. The argument has been used in [21] to derive for the first time the decay/blowup rates for extreme Kerr with nonzero Aretakis constants, yielding the a priori surprising result that for the “worst-behaving” non-axisymmetric modes (those with complex Δ_\pm), the rates are the same as in the case with vanishing Aretakis constants. For fields of nonzero spin, at least when perturbations separate and decouple, as in e.g. the case of the Teukolsky equations, a similar argument should go through. In this case, the diffeomorphism we use will need to be supplemented by a tetrad rotation (c.f. [6]).

The boundary late-time behavior may also be addressed using a similar symmetry argument, under a similar (one T -derivative stronger, as seen below) assumption on the boundary conditions. It allows to recover in detail the known late-time behavior (“tail”) [34],[35] of fields outside the horizon in the full extreme RN geometry. Again, it is natural to expect that an analogous assumption on the boundary conditions in similar near-horizon geometries such as NHEK [21], will recover the “tail” behavior from the symmetry argument presented here, for more general extreme BHs.

assumptions are that the field outside the horizon at late times decays and that the boundary conditions are given by some linear combination of the \pm modes at late times.

Using equation (2.10) with $\tilde{V} \rightarrow \infty$, $\tilde{V}\tilde{R} \gg 1$, places us on the boundary at late times. Imposing the condition that on the boundary $\phi(T_\infty, R = \infty) = \partial_T \phi(T_\infty, R = \infty) = 0$ (one T -derivative stronger than in the above analysis of horizon behavior, see also [14]), we can express the $\ell = 0$ solution for $T \rightarrow T_\infty$ as

$$\phi \simeq \frac{2f'(T_\infty)}{\tilde{T}^2 \tilde{R}} + (f^{(2)}(T_\infty) + g^{(2)}(T_\infty)) 1/\tilde{T}^2. \quad (2.22)$$

The condition $\phi(T_\infty, R = \infty) = \partial_T \phi(T_\infty, R = \infty) = 0$ enforces the decay rate (2.22) and is what we regard in this case to be boundary conditions that approach Dirichlet, $\phi|_{\text{bdy}} = 0$, “rapidly enough” at future timelike infinity to yield the known result in extreme RN ([34]). For example, requiring $\phi|_{\text{bdy}} = 0$ precisely from some $T < T_\infty$ onwards, implies the above assumption. In such a case, or in fact whenever $\partial_T^2 \phi(T_\infty, R = \infty) = 0$ is also imposed, the second term on the RHS of (2.22) will drop out and only the first one will survive.

For $\ell \geq 1$, we can use the form (2.18) and plug (2.10) into it. Taking the $\tilde{R} \rightarrow \infty$ limit yields, for large \tilde{T} ,

$$\phi \sim \tilde{R}^{\Delta-} \tilde{T}^{2\Delta-}, \quad (2.23)$$

for boundary conditions which are close enough to Dirichlet $\phi|_{\text{bdy}} = 0$ at late times. In the extremal RN case considered here, (2.23) is compatible with the result of [34]—namely that the late-time tails outside extreme RN behave as $\phi \sim \tilde{T}^{-(2\ell+2)}$.

3 Near-extremal behavior from symmetry

After reviewing the derivation of late-time behavior of linearized perturbations from a symmetry argument, in this section we will go further and show that the near-extremal behavior can also be deduced directly from the enhanced symmetry of the near-horizon geometry, which is only weakly broken for such BHs. We will recover the so called “transient instability” behavior [16] at intermediate times and the exponential decay at later times. Moreover, the decay outside the horizon at late times as a superposition of resonances with particular frequencies, or quasinormal modes (QNMs), will be recovered from the symmetry argument.

Consider a near-horizon limit of (2.1) as follows. Define a 1-parameter family of solutions, coordinate systems (and gauges) by taking (2.1) with $\hat{\kappa} = \epsilon\kappa$, and

$$r = \frac{\hat{r} - r_+}{\epsilon r_+}, \quad t = \frac{\epsilon \hat{t}}{M}, \quad a = \hat{A} + d\hat{t}, \quad (3.1)$$

where $r_+ = M(1 + \hat{\kappa})$ is the location of the outer horizon, and take the $\epsilon \rightarrow 0$ limit. The resulting solution is

$$\frac{ds^2}{M^2} = -r(r + 2\kappa)dt^2 + \frac{dr^2}{r(r + 2\kappa)} + d\Omega^2, \quad a = -M(r + \kappa)dt. \quad (3.2)$$

This metric is sometimes referred to as the near-AdS₂ metric. A fact we will use below is that even though the metrics (3.2) and (2.4) describe the same geometry locally, they cover different patches of the maximally extended geometry (which would be conveniently covered by global AdS₂ coordinates). An explicit example for a mapping between a patch described by (3.2) into a part of a patch described by (2.4) is the family of large diffeomorphisms (see figure 2 for a corresponding Penrose diagram)⁸

$$\begin{aligned} T - T_\infty &= -e^{-\kappa t} \frac{r + \kappa}{\sqrt{r(r + 2\kappa)}}, \\ R &= \frac{1}{\kappa} e^{\kappa t} \sqrt{r(r + 2\kappa)}, \end{aligned} \quad (3.3)$$

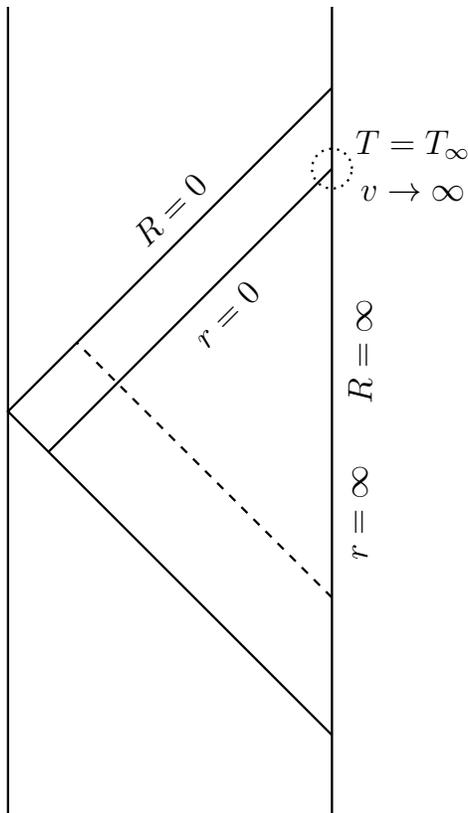


Figure 2: Penrose diagram displaying the mapping (3.3). The coordinates $\{T, R\}$ cover the Poincaré patch $0 < R < \infty$ with metric (2.4), while the coordinates $\{t, r\}$ cover the Rindler patch $0 < r < \infty$ with metric (3.2). The dashed null line represents a possible initial data surface.

We will now use this mapping in order to derive the late-time behavior of scalar perturbations on the future horizon of near-extremal BHs. The transformation (3.3) places the point

⁸One can supplement this coordinate transformation with a gauge transformation $a = A + d \left(\ln \sqrt{\frac{r}{r+2\kappa}} \right)$ in order that the electric field in (2.5) transforms into the form in (3.2).

$\{T = T_\infty, R = \infty\}$ at the future boundary of both the future horizon and the boundary $r \rightarrow \infty$. Defining a tortoise coordinate

$$r_* = \frac{1}{2\kappa} \ln \frac{r}{r + 2\kappa}, \quad (3.4)$$

and an outgoing Eddington-Finkelstein coordinate $v := t + r_*$, we can write

$$\begin{aligned} V &= -e^{-\kappa v} + T_\infty, \\ U &= -e^{-\kappa v} \frac{r}{r + 2\kappa} + T_\infty. \end{aligned} \quad (3.5)$$

Consider first the $\ell = 0$ case and a solution of the form (2.15). Similarly to (2.16), the solution here can be written as

$$\phi = f \left(T_\infty - e^{-\kappa v} \frac{r}{r + 2\kappa} \right) + g \left(T_\infty - e^{-\kappa v} \right), \quad (3.6)$$

where f, g are arbitrary smooth functions. Expanding around T_∞ gives

$$\phi = \sum_{n=1}^{\infty} (-)^n e^{-n\kappa v} \left[f^{(n)}(T_\infty) \left(\frac{r}{r + 2\kappa} \right)^n + g^{(n)}(T_\infty) \right]. \quad (3.7)$$

The relation between $f^{(n)}$ and $g^{(n)}$ is determined by the boundary conditions. Specifically, since we would like the field to decay on the boundary at late times we set $f^{(0)}(T_\infty) + g^{(0)}(T_\infty) = 0$ and therefore have omitted the $n = 0$ element of the sum.

Expression (3.7) describes a familiar behavior. For any value of the radial coordinate, including the horizon $r = 0$ and the boundary $r = \infty$, it is a sum over decaying exponentials. These describe nothing but a superposition of *quasinormal modes* (QNMs). We identify the QNM frequencies to be

$$\omega_n = -i\kappa n, \quad (3.8)$$

or in terms of far-region frequency, i.e. that which is conjugate to the far-region time coordinate \hat{t} ,

$$\hat{\omega}_n = -i\hat{\kappa}n. \quad (3.9)$$

As mentioned above, the $n = 0$ term is absent, by boundary conditions, which means that $n = 1, 2, \dots$ ⁹ The QNM's amplitudes are the coefficients in front of the decaying exponentials in (3.7). This constitutes a derivation of $\ell = 0$ near-extremal QNMs directly from the near-horizon conformal symmetry.

⁹The $n = 0$ mode is redundant in this context. It is a zero mode with no spatial or time dependence—and simply corresponds to the shift symmetry $\phi \rightarrow \phi + \text{const.}$ of the spherically symmetric sector of the field we are considering.

For $\ell > 0$, we will use

$$\begin{aligned} T - T_\infty &= -e^{-\kappa v} \frac{r + \kappa}{r + 2\kappa}, \\ R &= e^{\kappa v} \frac{r + 2\kappa}{\kappa}, \end{aligned} \quad (3.10)$$

and the expansion near the boundary point $\{T = T_\infty, R = \infty\}$,

$$\phi = f_+(T) R^{\Delta_+} \left(1 + \sum_{k=1}^{\infty} a_k R^{-k} \right) + f_-(T) R^{\Delta_-} \left(1 + \sum_{k=1}^{\infty} b_k R^{-k} \right), \quad (3.11)$$

where we assume that f_\pm are smooth at $T = T_\infty$ and may be expanded near $T = T_\infty$ as

$$f_\pm(T) = \sum_{n=0}^{\infty} f_\pm^{(n)}(T_\infty) (T - T_\infty)^n. \quad (3.12)$$

Note that the sums in the brackets in (3.11) do not contain logarithmic terms [36].

Now, let us for simplicity assume that

$$f_+^{(n)}(T_\infty) = 0, \quad \forall n, \quad (3.13)$$

enforcing boundary conditions which are “close enough” to Dirichlet, $\phi|_{\text{bdy}} = 0$, at late times (this condition can be relaxed to an assumption on only a finite number of derivatives, as in section 2). We can now rewrite (3.11) as

$$\phi = \sum_{n=1}^{\infty} A_{n\ell} e^{-\kappa(-\Delta_- + n-1)v}, \quad (3.14)$$

with

$$A_{n\ell} := \sum_{j=0}^{n-1} f_-^{(j)}(T_\infty) b_{n-1-j} (-1)^j \frac{\kappa^{n-1-j-\Delta_-} (r + \kappa)^j}{(r + 2\kappa)^{n-1-\Delta_-}}. \quad (3.15)$$

Similarly to the $\ell = 0$ case, we interpret this expression as a sum over QNMs with frequencies

$$\omega_{n\ell} = -i\kappa(-1 - \Delta_- + n) = -i\kappa(\ell + n), \quad (3.16)$$

or far region frequency

$$\hat{\omega}_{n\ell} = -i\hat{\kappa}(-1 - \Delta_- + n) = -i\hat{\kappa}(\ell + n). \quad (3.17)$$

for $n = 1, 2, \dots$. The boundary behavior can be extracted by taking the $r \rightarrow \infty$ limit in (3.14).

Note that the above family of QNMs (3.17) are usually referred to as “weakly damped QNMs”. These and related modes have been extensively discussed recently (for example, in [37]) following the suggestion of [38] that their existence in RN-de Sitter spacetimes implies a violation of the strong cosmic censorship conjecture. Note also that matching to the far region will determine the form of the QNM wavefunctions (3.15) but will not modify (to leading order in $\hat{\kappa}$) the frequencies (3.17).

It is natural to ask whether the Aretakis instability occurs for precisely extremal BHs only, or if it has some trace in the near-extremal limit. This was considered in [22] within an analytical “toy model” attempting to explain the behavior seen in the full nonlinear numerical evolution, and in [16], for extremal Kerr with vanishing Aretakis constants, where a frequency-domain analysis was used. These analyses find qualitatively similar behavior—a “transient instability”: for moderately late times $1 \ll \hat{v} \ll M\hat{\kappa}^{-1}$ an Aretakis-like behavior with power-law blow-ups, while for later times $\hat{v} > M\hat{\kappa}^{-1}$ the exponential decay (3.7), (3.14) is restored. Here we will show, again using a large diffeomorphism, that this behavior too can be recovered using the symmetry argument.

Consider the boundary to bulk retarded propagator $G(V, R; V')$ in AdS_2 with Dirichlet $\phi|_{\text{bdy}} = 0$ boundary conditions. It constitutes a solution to the wave equation $\square\phi = 0$ that arises from turning on an instantaneous source term on the boundary, at some time that we will arbitrarily take to be $V' = 0$. This propagator was recently computed in [3]. It is given, in ingoing Eddington-Finkelstein coordinates, by

$$G(V, R; 0) = a_0 \Theta(V) V^{\Delta_-} (2 + VR)^{\Delta_-}, \quad (3.18)$$

where a_0 is a constant amplitude and $\Theta(V)$ is the Heaviside step function which enforces causality, or equivalently the condition of no incoming radiation from the past horizon. Note that (3.18) is of the universal form (2.20).

Using (3.3), (3.5), we write

$$\begin{aligned} V &= T_\infty (1 - e^{-\kappa(v-v_0)}), \\ R &= T_\infty^{-1} \frac{r + 2\kappa}{\kappa} e^{\kappa(v-v_0)}, \end{aligned} \quad (3.19)$$

where $T_\infty \equiv e^{-\kappa v_0}$. Using the mapping (3.19) gives

$$G(v, r; v_0) = a_0 T_\infty^{\Delta_-} \Theta(v - v_0) (1 - e^{-\kappa(v-v_0)})^{\Delta_-} \left(2 + (e^{\kappa(v-v_0)} - 1) \frac{r + 2\kappa}{\kappa} \right)^{\Delta_-}. \quad (3.20)$$

This is the boundary to bulk propagator in near- AdS_2 , obtained from the AdS_2 propagator by symmetry. Note that the no-incoming wave and boundary conditions are preserved by the transformation, and that here in the new near- AdS_2 coordinates the source is turned on

at $v = v_0 = (-1/\kappa) \ln T_\infty$. For $0 < \kappa(v - v_0) \ll 1$, we can expand (3.20) to leading order in $\kappa(v - v_0)$ and obtain

$$G(v, r; v_0) \approx a_0 (\kappa T_\infty)^{\Delta_-} (v - v_0)^{\Delta_-} (2 + (v - v_0)(r + 2\kappa))^{\Delta_-}. \quad (3.21)$$

The solution at these times exhibits “transient” Aretakis behavior, since

$$\partial_r^k G(v, r; v_0)|_{r=0} \propto (v - v_0)^{k+\Delta_-}, \quad (3.22)$$

to leading order in $\kappa(v - v_0)$. Thus, this constitutes a symmetry argument for the “transient” Aretakis instability for near-extreme RN BHs. A translation of the statement (3.22) to the full extreme RN geometry (eliminating v_0 by a time translation) would be that for advanced times $M^{-1}\hat{\kappa}\hat{v} \ll 1$ derivatives of the field behave as

$$\partial_{\hat{r}}^k \phi|_{\hat{r}=0} \propto \hat{v}^{k+\Delta_-}, \quad (3.23)$$

and for times $M^{-1}\hat{\kappa}\hat{v} > 1$ the field exhibits exponential decay through QNMs as shown in equations (3.7),(3.14).

For more general cases (such as massive/charged fields, axisymmetric modes of rotating BHs, and probably also higher spin fields), it seems likely that a similar argument would go through. For example, in extremal Kerr, for modes with complex $\Delta_- = 1/2 + ir$ with r real, at late times, the solution may be written as [39]

$$\phi \sim V^{\Delta_-} \sum_n V^{i\chi_n} f_n(VR), \quad (3.24)$$

where χ_n are some real phases and f_n are smooth. Thus if we transform to coordinates of the type (3.19) (near-NHEK coordinates, in this example) these modes will exhibit a “transient instability” behavior for $\kappa v \ll 1$.

4 Backreacted solution within Jackiw-Teitelboim theory

What “happens” to a perturbed extremal BH? What is the endpoint of the Aretakis instability and how does the system evolve toward it? This was addressed in [22] by numerically solving the coupled Einstein-scalar equations for spherically symmetric massless, neutral scalar perturbations around extreme RN. The process was understood there as follows: the instability persists nonlinearly, as explained in section 1; but for generic initial data, at some point backreaction kicks in, causing the BH to become slightly non-extremal, and thus small perturbations at late times eventually decay exponentially at a timescale corresponding to

the final inverse temperature. This makes sense from an outside observer’s point of view: perturbing a system at the absolute zero with some small but nonzero energy, it seems reasonable that its final state would be, at least in some effective sense, thermal. But how can this be connected to the near-horizon picture for the late-time behavior discussed in sections 2, 3? As argued in [4], pure AdS₂ is incompatible with nonsingular finite energy excitations, when backreaction is taken into account. However, as was argued in [28] and refined in [23], Jackiw-Teitelboim (JT) theory constitutes a universal model which allows the treatment of backreaction effects for *nearly*-AdS₂ spacetimes. There has been much recent interest in the JT model, a 1+1 dimensional theory of dilaton gravity which we will discuss in this section. This was spurred most prominently by the realization [40] that this theory captures the dynamics of the low-energy reparametrization mode of the Sachdev-Ye-Kitaev (SYK) model [24],[25] and thus by the prospect of finding a new, solvable model of holographic duality between a (0+1) dimensional boundary and a (1+1) dimensional bulk. It also constitutes a controllable toy model to address questions of principle in quantum gravity, c.f. [41],[42]. In our context, JT theory will be useful since it allows a remarkably simple description of *backreaction* effects in the near-horizon region of near-extremal BHs described by the nearly-AdS₂ geometry; thereby we will be able to study the nonlinear evolution of near-extremal BHs at late times analytically.

At least to some degree of universality, it has recently been argued that the JT model constitutes a general description of the low energy perturbations of extremal BHs. This has been explicitly shown for spherically symmetric perturbations of large extremal RN-AdS BHs in [29]¹⁰ (see also [31]). In this section we will be interested in the simplest case of spherically symmetric perturbations of a minimally coupled massless, neutral scalar field around extreme RN, and therefore we will study JT theory coupled to a massless neutral scalar in near-AdS₂.

The JT theory is given by the action

$$S = \frac{1}{2} \int \sqrt{-g} d^2x \Phi(R + 2) + \Phi_{\text{bdy}} \int_{\text{bdy}} \sqrt{h} (K - 1) du + S_{\text{matter}}, \quad (4.1)$$

where K is the extrinsic curvature of the boundary, \sqrt{h} is the induced metric on the boundary and u is the proper time along it, Φ is a usually referred to as the dilaton, and S_{matter} is the bulk matter action. Importantly, we adopt the assumption of [28],[23] that the bulk matter Lagrangian is independent of Φ , so there is no direct matter-dilaton coupling. Φ encodes the volume of the transverse S^2 of the original higher-dimensional geometry¹¹. At large radii,

¹⁰and non-spherically symmetric perturbations can presumably also be accounted for by a Kaluza-Klein reduction into massive fields in 2 dimensions, see recent discussion in [30].

¹¹In fact, a reduction from higher dimensions will result in another, topological, term $S_{\text{Topo}} =$

the dilaton generically diverges and therefore the solution needs to be cut off at some radial distance $\sim \epsilon^{-1}$, where ϵ is a small parameter controlling this cutoff. The physical boundary is taken to be the (1-dimensional) timelike hypersurface where the value of Φ is fixed to some value $\Phi_{bdy} := \Phi_r \epsilon^{-1}$, where Φ_r is a renormalized value for the dilaton, factoring out its scaling with ϵ .

Even though the theory (4.1) is already quite simple, it turns out that it admits an alternative formulation which looks even simpler, and is particularly useful for the present application [23],[41]. Φ is an auxiliary (nondynamical) field, and it is simply integrated out exactly, setting the Ricci Scalar $R = -2$. This fixes the geometry, locally, to AdS_2 . The only remaining degree of freedom, then, is the boundary trajectory which is assumed to reside (radially) far from the horizon, so that the AdS_2 throat is long. The gravitational part of the action then reduces to

$$S = \Phi_{bdy} \int_{bdy} \sqrt{h} (K - 1) du + S_{matter}. \quad (4.2)$$

Parameterizing the boundary worldline using Poincaré coordinates $\{T_b(u), R_b(u)\}$ where u is the proper time on the boundary, and using the relation $R_b^{-1}(u) \approx \epsilon \dot{T}_b(u)$ for large R_b , the purely gravitational part of the action can be put in the form

$$S_{grav} = -\Phi_r \int \{T_b(u), u\} du, \quad (4.3)$$

where

$$\{T_b(u), u\} = \frac{\ddot{T}_b}{\dot{T}_b} - \frac{3}{2} \left(\frac{\ddot{T}_b}{\dot{T}_b} \right)^2, \quad (4.4)$$

is the Schwarzian derivative, and $\dot{} = d/du$. (4.3) describes the purely gravitational (no matter fields of yet) dynamics of near- AdS_2 spacetimes. It is $\text{SL}(2, \mathbb{R})$ invariant; in particular in the vacuum case, a conserved energy may be defined:

$$-E = \Phi_r \{T_b(u), u\}. \quad (4.5)$$

Now we would like to explicitly add bulk matter to this picture. As mentioned above, in the JT model matter fields propagate (locally) freely on an AdS_2 background and are coupled to the gravitational degrees of freedom only through the boundary. We will assume for simplicity in this section Dirichlet boundary conditions which set the matter field to zero

$\frac{1}{2}\Phi_0 \left(\int R + 2 \int_{bdy} K \right)$ (where Φ_0 is a constant); but this term is just a constant in our case since the topology is fixed (and is actually equal to 0 since we consider the Lorentzian strip topology). $\Phi_0 + \Phi$ is the total volume of the transverse S^2 .

at the boundary, $\phi|_{\text{bdy}} = 0$ where ϕ is our bulk scalar field.¹² Following [23], a particularly simple way to take into account the influence of matter fields is by keeping track of how the total energy E is distributed between the BH and the matter¹³. This may be done by reinstating the dilaton and writing down the ADM energy in its terms of its value, and the value of its normal derivative on the boundary, using a relation between total energy and variation of the action with respect to the boundary metric [43]. Then one may use the EOM $\delta S/\delta g^{RR} = 0$ in order to relate these quantities to the pressure component of the matter field's stress-energy T_{RR} there. A detailed derivation (in Euclidean signature) is found in appendix A of [23]. The result is a generalization of (4.5) to non-vacuum solutions:

$$-E = \Phi_r \{T_b(u), u\} + \dot{T}_b R_b^3 T_{RR}, \quad (4.6)$$

which we find particularly useful for our application.

The details of S_{matter} are not universal—different matter content and other (symmetry) assumptions in the higher-dimensional parent theory, will generically yield different S_{matter} . We will focus for simplicity in this paper on the case of spherically symmetric perturbations of a neutral, minimally coupled massless scalar field which will correspond to taking

$$S_{\text{matter}} = -\frac{1}{2} \int \sqrt{-g} d^2x (\nabla\phi)^2. \quad (4.7)$$

Allowing, for example, non-spherically symmetric scalar perturbations will introduce a tower of massive fields. As noted above, a general feature of JT theory which crucially simplifies the analysis here is that Φ does not appear in S_{matter} , and therefore does not couple directly to the matter fields at leading order in the strength of the perturbations; at leading order, gravity-matter coupling occurs only through the boundary degree of freedom.

We wish to solve equation (4.6) in the presence of a massless scalar field propagating in the bulk with Dirichlet boundary conditions. Assume that initial data is chosen at $V = 0$ so that the outgoing component of the field is given, for $U > 0$, by

$$\phi_{\text{outgoing}} = H_0 f(U), \quad (4.8)$$

where f is some arbitrary function determining the profile of the outgoing wave. We have factored out the amplitude of the perturbation H_0 (we will later assume it is small). The

¹²It is possible to modify the assumed boundary conditions. This may be desired since when reducing from a generic perturbed asymptotically flat BH we expect the field (and its derivative) to have some time dependent value on the boundary. The realistic, precise boundary conditions would be of Robin type, and time dependent; however Dirichlet, $\phi = 0$ boundary conditions seem to be a good approximation for the low frequency modes which govern the late-time behavior. The solution found here, then, is expected to capture the late-time behavior even under more realistic boundary conditions.

¹³This energy is conserved with the boundary conditions of choice here; more generally one would need to keep track of the energy “leaking out” to infinity.

boundary condition $\phi = 0$ we impose at $R_b = \mathcal{O}(\epsilon)$, then, determines the solution for the scalar field to be of the form

$$\phi = f(U) - f(V) + \Delta\phi(V), \quad (4.9)$$

where $\Delta\phi = \mathcal{O}(\epsilon)$. Specifically, near the boundary, we can expand

$$T_{RR} = \frac{2H_0^2}{R^4} [(f'(T))^2 + \mathcal{O}(\epsilon)]. \quad (4.10)$$

Equation (4.6), then, is given to leading order in ϵ by

$$-E = \Phi_r \{T_b(u), u\} + 2\epsilon H_0^2 \dot{T}_b^2(u) (f'(T_b(u)))^2. \quad (4.11)$$

We wish to understand how the boundary trajectory is affected by radiation in the bulk, and how this affects the late-time behavior. To this end, we find it useful to revert at this stage to a Poincaré time parametrization $R_b(T)$, using $R_b^{-1} \approx \epsilon \dot{T}_b$ to rewrite (4.11) as

$$[\partial_T (\ln R_b)]^2 - 2\partial_T^2 (\ln R_b) + \frac{(2\pi)^2}{\beta^2} \epsilon^2 R_b^2 = -\frac{4\epsilon}{\Phi_r} H_0^2 (\partial_T f)^2, \quad (4.12)$$

where $\beta = 2\pi\sqrt{\Phi_r/(2E)}$. Equation (4.12) governs (together with the wave equation) the behavior of the BH+matter system for general near-horizon initial data. The choice of initial data enters through a specific choice of $f(T)$. In the vacuum case, the right hand side vanishes and (4.12) reduces to the statement of conservation of the Schwarzian $\{T_b, u\}$.

For small perturbations, (4.12) may be solved perturbatively in H_0 . Beginning by putting $H_0 = 0$, the solution has a particularly simple form:

$$R_{b,(0)}^{-1}(T) = \epsilon \left[1 - \left(\frac{\pi}{\beta} \right)^2 T^2 \right], \quad (4.13)$$

and we defined integration constants so as to get precisely $R_b^{min} = \epsilon^{-1}$, where R_b^{min} is the minimal value of the Poincaré radial coordinate along the boundary trajectory. There is also freedom to shift T by a constant; for convenience we choose $R_b^{(0)}(T=0) = R_b^{min}$. To perturb (4.13), it is convenient to use an ansatz of the form

$$R_b^{-1}(T) = R_{b,(0)}^{-1}(T) + \frac{\epsilon H_0^2}{\Phi_r} R_{b,(1)}^{-1}(T). \quad (4.14)$$

Plugging this into (4.12) and equating the $\mathcal{O}(\epsilon H_0^2/\Phi_r)$ terms on both sides, we obtain an equation for the perturbation:

$$\partial_T^2 R_{b,(1)}^{-1} + 2 \left[T^2 - \left(\frac{\beta}{\pi} \right)^2 \right]^{-1} \left(R_{b,(1)}^{-1} - T \partial_T R_{b,(1)}^{-1} \right) = \frac{2\epsilon\pi^2}{\beta^2} \left[T^2 - \left(\frac{\beta}{\pi} \right)^2 \right] (\partial_T f)^2. \quad (4.15)$$

This second order, non-homogeneous linear ordinary differential equation is integrable for arbitrary forcing term $f(T)$. The integration constants in the general solution are determined by the initial conditions, namely that for $V = 0$ the solution and its derivative identify with the unperturbed solution (4.13). Employing these conditions gives

$$R_{b,(1)}^{-1}(T) = 2\epsilon \left\{ \left(\frac{\pi}{\beta} T - 1 \right)^2 \int_{\epsilon}^T y (\partial_y f(y))^2 dy - T \int_{\epsilon}^T \left(\frac{\pi}{\beta} y - 1 \right)^2 (\partial_y f(y))^2 dy \right\}. \quad (4.16)$$

Notice that $R_{b,(1)}^{-1}$ is always negative in the relevant range: in our case, the boundary can only be pushed outwards, and the temperature increased.

To be more concrete, it will be interesting to consider the specific profile $f(T)$ which corresponds to the special—universal—near-horizon solution in the sense explained in preceding sections, namely

$$\phi \simeq \frac{H_0}{V(2 + VR)} = \frac{H_0}{2} (1/V - 1/U). \quad (4.17)$$

The $\propto 1/U$ piece in the above can be thought of as arising from near-horizon initial data corresponding to an outgoing wave and specified on, e.g., the surface $V = 0$, and the $\propto 1/V$ piece can be thought of as its reflection off the boundary, given the boundary conditions we impose. It is interesting to note that the computation we perform is relevant also to the scenario where the Aretakis constants vanish, and the BH is perturbed from afar. As discussed in section 3, this amounts to considering a solution for ϕ which is proportional to the boundary-to-bulk propagator, just with an amplitude suppressed by a factor $\sim \epsilon^{-\Delta_-} = \epsilon$ since the point of emission/perturbation resides at large R , and this propagator is of the exact same form (4.17). Interestingly, this particular suppression of the amplitude with ϵ will in fact pop up in the below analysis as a consistency condition for our perturbative analysis. Another important point to note is that, as in (4.9), there will be corrections to (4.17) due to the modification of the boundary trajectory—namely, the reflected part will be $\propto 1/V + \mathcal{O}(\epsilon)$. For the trajectory computation, at leading order we can neglect it, but for example for the field's late time behavior this will be important to keep in mind. The above discussion amounts to simply plugging $f(T) = -(2T)^{-1}$ in (4.16). The solution then is given, to leading order in ϵ , by

$$R_{b,(1)}^{-1}(T) = -\frac{\Theta(T - \epsilon)}{6\epsilon^2} \left(T - \frac{3\epsilon}{2} + \frac{\epsilon^3}{2T^2} \right), \quad (4.18)$$

and note that the two rightmost terms in the brackets above are important only at early times $T \sim \epsilon$. This result, combined with (4.14) is telling us that in order for the perturbative

computation to be consistent in the case (4.17), it is mandatory to take the amplitude of the scalar field ϕ to scale no stronger than ϵ . Physically we can interpret this as the incompatibility of the near-horizon picture with too strong an amplitude, realized in [4]: it is impossible to squeeze perturbations of higher amplitude in the near-horizon region and still keep them outside the event horizon.¹⁴ Finally, it is also interesting to note that the extra factor of ϵ we are forced to suppress the amplitude with arises naturally when considering the case of vanishing Aretakis constant—indeed, the far-near Green’s function (3.18) (in other words, the solution in the near-horizon region arising from an instantaneous, localized perturbation in the far region) is precisely of the form (1.4), only suppressed with a factor of ϵ as discussed above and in section 3.

We will explicitly scale the amplitude, then, as $H_0 := \epsilon h_0$ where $h_0 = \mathcal{O}(\epsilon^0)$. The full solution for the trajectory is now given by

$$R_b^{-1}(T) = \epsilon \left[1 - \left(\frac{\pi}{\beta} \right)^2 T^2 - \Theta(T - \epsilon) \frac{h_0^2}{6 \Phi_r} \left(T - \frac{3\epsilon}{2} + \frac{\epsilon^3}{2T^2} \right) \right]. \quad (4.19)$$

In particular, as a consequence of the backreaction there will be a shift in the position of future timelike infinity (located at the intersection of the physical boundary and the boundary of global AdS_2 —see figure 3). It will now be located at

$$T_\infty = \frac{\beta}{\pi} \left(\sqrt{1 + \xi^2} - \xi \right) \approx \frac{\beta}{\pi} (1 - \xi), \quad (4.20)$$

where

$$\xi = \frac{\beta h_0^2}{12\pi\Phi_r}. \quad (4.21)$$

In particular, T_∞ decreases as the perturbation’s amplitude h_0 increases. Importantly, the value of T_∞ determines the position of the event horizon as the latter is determined simply by the latest right-moving null ray reaching T_∞ . Figure (3) depicts schematically the backreacted near-horizon computation, and in particular the boundary’s motion and the event horizon.

Now that we have computed the boundary trajectory and the location of the backreacted event horizon, we can proceed to determine the late-time behavior of the scalar perturbations. It is worth stressing again that in the bulk we are simply solving the wave equation on (a patch of) a fixed AdS_2 geometry. We then only need to *evaluate* the field along the correct curves, which correspond to the backreacted event horizon and boundary. The problem of determining the late-time behavior on the horizon now reduces, thereby, to that already

¹⁴This criterion on the amplitude may be related to the distinction in [22] between “degenerate apparent horizon” and “first order mass perturbation” types of initial data considered there.

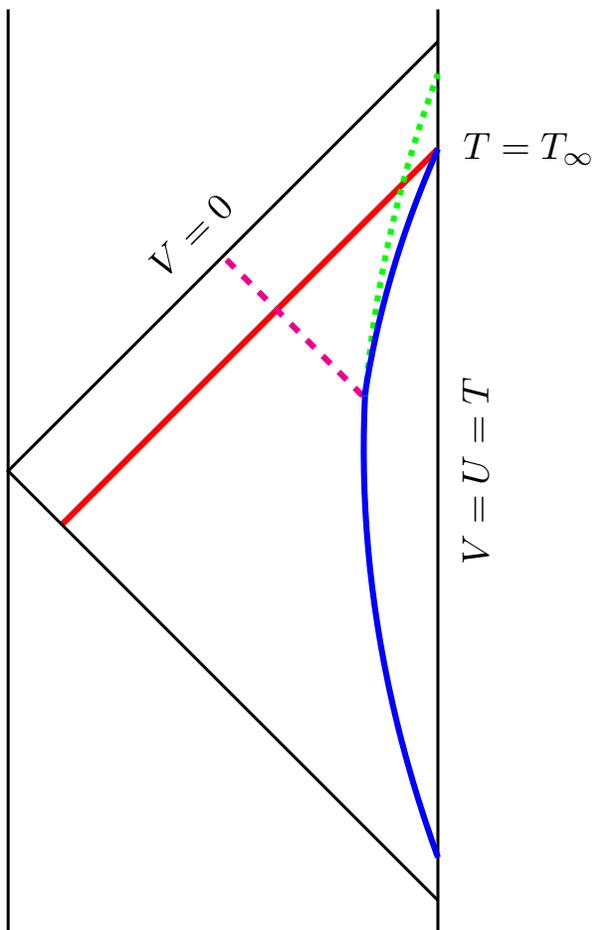


Figure 3: Penrose diagram illustrating the near-horizon computation of late-time behavior including the effects of backreaction. The physical boundary, in solid blue, is promoted to a degree of freedom encoding the black hole’s (small) deviation from extremality. The red solid line is the event horizon. The magenta dashed line depicts a possible initial data surface, or an advanced time in which the BH is perturbed from afar. The dotted green line shows the later part of the boundary trajectory had the scalar field not been excited. The late-time behavior of fields, as in fig. 2, is determined by analyzing the neighborhood of the point $T = T_\infty$.

solved in section 3. For simplicity and since it is especially important (due to its universality) we will restrict to the case (4.17), but in principle the same method will apply for the general solution, (4.16). In particular, the late-time behavior will be given by equation (4.9), where here we will need to include the correction $\Delta\phi(V)$ of equation (4.9) to get the correct value of the field near the boundary.¹⁵ The boundary conditions we choose to impose here give

$$\Delta\phi(V) = -\frac{H_0}{V^2 R_b(V)} \quad (4.22)$$

¹⁵This correction is negligible when computing the boundary trajectory, equations (4.6),(4.10).

and thus the solution is corrected to

$$\phi = \epsilon h_0 \left[\frac{1}{2V} - \frac{1}{2U} - \epsilon \left(\frac{\pi^2}{\beta^2} + \frac{h_0^2}{6\Phi_r} \frac{1}{V} - \frac{1}{V^2} \right) \right], \quad (4.23)$$

and boundary conditions are obeyed up to $\mathcal{O}(\epsilon^2)$. The field's boundary behavior at any time T may be extracted analytically by expanding (4.23) for large R and substituting $R_b(T)$ into (4.23), namely $V = V(R_b(T), T)$ and $U = U(R_b(T), T)$. Translating to boundary time, it may be interesting to compare this with expectations from holography. The (transient instability) behavior at intermediate times is given as well by the corresponding expression from section 3; in this case it is not essential to take into account the correction (4.22).

The above analysis gives the late-time behavior in terms of near-horizon quantities. In the higher-dimensional setting, we are considering a situation where a near-extremal BH is perturbed with energy comparable to its own energy above extremality, so in order to translate from near-region to far-region language we only need to determine what is the temperature/surface gravity of the new, resulting BH (we assume that the temperature of the old one is $\hat{\kappa}$ as in section 3). The change in temperature of the BH can be inferred from the shift in T_∞ , equation (4.20). The perturbed trajectory traverses a shorter Poincaré time than the original one, by a factor of $(1 - \frac{\beta h_0^2}{12\pi\Phi_r})$, but must traverse the same (very large) amount of boundary/proper time as the unperturbed trajectory. This means that $\frac{d}{du}$ must pick up an inverse factor to the above, and therefore the boundary energy, proportional to the Schwarzian, will pick up a factor of $\sim (1 + \frac{\beta h_0^2}{6\pi\Phi_r})$. But given that the surface gravity is $\sqrt{1 - Q^2/M^2}$ for $M \simeq Q$, the perturbed temperature/surface gravity will be enhanced by a factor of $(1 + \frac{\beta h_0^2}{12\pi\Phi_r})$.

It is tempting to go further and identify $(2\pi\epsilon)/\beta$ as the surface gravity $\hat{\kappa}$ and h_0^2/Φ_r as the amplitude squared A^2 of the spherically symmetric perturbation in appropriate units. This yields, for the surface gravity of the new BH,

$$\hat{\kappa}' \simeq \hat{\kappa} + \frac{A^2}{12M}. \quad (4.24)$$

In particular, the late-time behavior will be given by equation (3.7), but now translating to the far-region quantities one would need to replace $\hat{\kappa}$ with $\hat{\kappa}'$ of equation (4.24), yielding for the QNM frequencies governing the behavior of the new ringing-down BH,

$$\hat{\omega}_n = -i\hat{\kappa}'n = -i\left(\hat{\kappa} + \frac{A^2}{12M}\right)n. \quad (4.25)$$

It would be interesting to verify in detail this matching to far-region quantities.

Note here that sometimes it makes more practical sense to leave results in the resummed form (3.6) than expand to the QNM-like form (3.7). Indeed, in some situations one finds

that near-extremal QNMs are explicitly summable analytically, e.g. [6],[16], but this should be of no surprise since the QNM sum “originates” from a resummed expression as in (3.6) to begin with.

Acknowledgements

I am grateful to Joan Camps, Hadi Godazgar, Michal Heller, Achilleas Porfyriadis, and Andy Strominger for useful conversations and to Sam Gralla, Alex Lupsasca, and Harvey Reall for helpful comments on the manuscript. I gratefully acknowledge support from the Jacob Goldfield Postdoctoral Support Fund.

References

- [1] S. A. Hartnoll, “Lectures on holographic methods for condensed matter physics,” *Class. Quant. Grav.* **26**, 224002 (2009) doi:10.1088/0264-9381/26/22/224002 [arXiv:0903.3246 [hep-th]].
- [2] S. E. Gralla, A. Lupsasca and A. Strominger, “Near-horizon Kerr Magnetosphere,” *Phys. Rev. D* **93**, no. 10, 104041 (2016) doi:10.1103/PhysRevD.93.104041 [arXiv:1602.01833 [hep-th]].
- [3] S. E. Gralla and P. Zimmerman, “Scaling and Universality in Extremal Black Hole Perturbations,” *JHEP* **1806**, 061 (2018) doi:10.1007/JHEP06(2018)061 [arXiv:1804.04753 [gr-qc]].
- [4] J. M. Maldacena, J. Michelson and A. Strominger, “Anti-de Sitter fragmentation,” *JHEP* **9902**, 011 (1999) doi:10.1088/1126-6708/1999/02/011 [hep-th/9812073].
- [5] M. Guica, T. Hartman, W. Song and A. Strominger, “The Kerr/CFT Correspondence,” *Phys. Rev. D* **80**, 124008 (2009) doi:10.1103/PhysRevD.80.124008 [arXiv:0809.4266 [hep-th]].
- [6] S. Hadar, A. P. Porfyriadis and A. Strominger, “Gravity Waves from Extreme-Mass-Ratio Plunges into Kerr Black Holes,” *Phys. Rev. D* **90**, no. 6, 064045 (2014) doi:10.1103/PhysRevD.90.064045 [arXiv:1403.2797 [hep-th]].
- [7] S. E. Gralla, S. A. Hughes and N. Warburton, “Inspirals into Gargantua,” *Class. Quant. Grav.* **33**, no. 15, 155002 (2016) doi:10.1088/0264-9381/33/15/155002 [arXiv:1603.01221 [gr-qc]].

- [8] S. E. Gralla, A. Lupsasca and A. Strominger, “Observational Signature of High Spin at the Event Horizon Telescope,” *Mon. Not. Roy. Astron. Soc.* **475**, no. 3, 3829 (2018) doi:10.1093/mnras/sty039 [arXiv:1710.11112 [astro-ph.HE]].
- [9] D. Gates, D. Kapec, A. Lupsasca, Y. Shi and A. Strominger, “Polarization Whorls from M87 at the Event Horizon Telescope,” arXiv:1809.09092 [hep-th].
- [10] S. Aretakis, “Stability and Instability of Extreme Reissner-Nordström Black Hole Spacetimes for Linear Scalar Perturbations I,” *Commun. Math. Phys.* **307**, 17 (2011) doi:10.1007/s00220-011-1254-5 [arXiv:1110.2007 [gr-qc]].
- [11] S. Aretakis, “Stability and Instability of Extreme Reissner-Nordstrom Black Hole Spacetimes for Linear Scalar Perturbations II,” *Annales Henri Poincare* **12**, 1491 (2011) doi:10.1007/s00023-011-0110-7 [arXiv:1110.2009 [gr-qc]].
- [12] S. Aretakis, “Horizon Instability of Extremal Black Holes,” *Adv. Theor. Math. Phys.* **19**, 507 (2015) doi:10.4310/ATMP.2015.v19.n3.a1 [arXiv:1206.6598 [gr-qc]].
- [13] S. Aretakis, “A note on instabilities of extremal black holes under scalar perturbations from afar,” *Class. Quant. Grav.* **30**, 095010 (2013) doi:10.1088/0264-9381/30/9/095010 [arXiv:1212.1103 [gr-qc]].
- [14] J. Lucietti, K. Murata, H. S. Reall and N. Tanahashi, “On the horizon instability of an extreme Reissner-Nordström black hole,” *JHEP* **1303**, 035 (2013) doi:10.1007/JHEP03(2013)035 [arXiv:1212.2557 [gr-qc]].
- [15] M. Casals, S. E. Gralla and P. Zimmerman, “Horizon Instability of Extremal Kerr Black Holes: Nonaxisymmetric Modes and Enhanced Growth Rate,” *Phys. Rev. D* **94**, no. 6, 064003 (2016) doi:10.1103/PhysRevD.94.064003 [arXiv:1606.08505 [gr-qc]].
- [16] S. E. Gralla, A. Zimmerman and P. Zimmerman, “Transient Instability of Rapidly Rotating Black Holes,” *Phys. Rev. D* **94**, no. 8, 084017 (2016) doi:10.1103/PhysRevD.94.084017 [arXiv:1608.04739 [gr-qc]].
- [17] P. Zimmerman, “Horizon instability of extremal Reissner-Nordström black holes to charged perturbations,” *Phys. Rev. D* **95**, no. 12, 124032 (2017) doi:10.1103/PhysRevD.95.124032 [arXiv:1612.03172 [gr-qc]].
- [18] J. Lucietti and H. S. Reall, “Gravitational instability of an extreme Kerr black hole,” *Phys. Rev. D* **86**, 104030 (2012) doi:10.1103/PhysRevD.86.104030 [arXiv:1208.1437 [gr-qc]].

- [19] S. Aretakis, “The Wave Equation on Extreme Reissner-Nordstrom Black Hole Spacetimes: Stability and Instability Results,” arXiv:1006.0283 [math.AP].
- [20] Y. Angelopoulos, S. Aretakis and D. Gajic, “Late-time asymptotics for the wave equation on extremal Reissner-Nordström backgrounds,” arXiv:1807.03802 [gr-qc].
- [21] S. Hadar and H. S. Reall, “Is there a breakdown of effective field theory at the horizon of an extremal black hole?,” JHEP **1712**, 062 (2017) doi:10.1007/JHEP12(2017)062 [arXiv:1709.09668 [hep-th]].
- [22] K. Murata, H. S. Reall and N. Tanahashi, “What happens at the horizon(s) of an extreme black hole?,” Class. Quant. Grav. **30**, 235007 (2013) doi:10.1088/0264-9381/30/23/235007 [arXiv:1307.6800 [gr-qc]].
- [23] J. Maldacena, D. Stanford and Z. Yang, “Conformal symmetry and its breaking in two dimensional Nearly Anti-de-Sitter space,” PTEP **2016**, no. 12, 12C104 (2016) doi:10.1093/ptep/ptw124 [arXiv:1606.01857 [hep-th]].
- [24] A. Kitaev, A simple model of quantum holography. Seminar at KITP, 2015.
- [25] S. Sachdev and J. Ye, “Gapless spin fluid ground state in a random, quantum Heisenberg magnet,” Phys. Rev. Lett. **70**, 3339 (1993) doi:10.1103/PhysRevLett.70.3339 [cond-mat/9212030].
- [26] R. Jackiw, “Lower Dimensional Gravity,” Nucl. Phys. B **252**, 343 (1985). doi:10.1016/0550-3213(85)90448-1
- [27] C. Teitelboim, “Gravitation and Hamiltonian Structure in Two Space-Time Dimensions,” Phys. Lett. **126B**, 41 (1983). doi:10.1016/0370-2693(83)90012-6
- [28] A. Almheiri and J. Polchinski, “Models of AdS₂ backreaction and holography,” JHEP **1511**, 014 (2015) doi:10.1007/JHEP11(2015)014 [arXiv:1402.6334 [hep-th]].
- [29] P. Nayak, A. Shukla, R. M. Soni, S. P. Trivedi and V. Vishal, “On the Dynamics of Near-Extremal Black Holes,” arXiv:1802.09547 [hep-th].
- [30] U. Moitra, S. P. Trivedi and V. Vishal, “Near-Extremal Near-Horizons,” arXiv:1808.08239 [hep-th].
- [31] G. Srosi, “AdS₂ holography and the SYK model,” PoS Modave **2017**, 001 (2018) doi:10.22323/1.323.0001 [arXiv:1711.08482 [hep-th]].

- [32] A. P. Porfyriadis, “Scattering of gravitational and electromagnetic waves off $AdS_2 \times S^2$ in extreme Reissner-Nordstrom,” arXiv:1805.12409 [hep-th].
- [33] A. P. Porfyriadis, “near- AdS_2 perturbations and the connection with near-extreme Reissner-Nordstrom,” arXiv:1806.07097 [hep-th].
- [34] A. Ori, “Late-time tails in extremal Reissner-Nordstrom spacetime,” arXiv:1305.1564 [gr-qc].
- [35] S. Bhattacharjee, B. Chakrabarty, D. D. K. Chow, P. Paul and A. Virmani, “On late time tails in an extreme ReissnerNordstrm black hole: frequency domain analysis,” *Class. Quant. Grav.* **35**, no. 20, 205002 (2018) doi:10.1088/1361-6382/aade59 [arXiv:1805.10655 [gr-qc]].
- [36] K. Skenderis, “Lecture notes on holographic renormalization,” *Class. Quant. Grav.* **19**, 5849 (2002) doi:10.1088/0264-9381/19/22/306 [hep-th/0209067].
- [37] O. J. C. Dias, H. S. Reall and J. E. Santos, “Strong cosmic censorship for charged de Sitter black holes with a charged scalar field,” arXiv:1808.04832 [gr-qc].
- [38] V. Cardoso, J. L. Costa, K. Destounis, P. Hintz and A. Jansen, “Quasinormal modes and Strong Cosmic Censorship,” *Phys. Rev. Lett.* **120**, no. 3, 031103 (2018) doi:10.1103/PhysRevLett.120.031103 [arXiv:1711.10502 [gr-qc]].
- [39] S. E. Gralla and P. Zimmerman, “Critical Exponents of Extremal Kerr Perturbations,” *Class. Quant. Grav.* **35**, no. 9, 095002 (2018) doi:10.1088/1361-6382/aab140 [arXiv:1711.00855 [gr-qc]].
- [40] J. Maldacena and D. Stanford, “Remarks on the Sachdev-Ye-Kitaev model,” *Phys. Rev. D* **94**, no. 10, 106002 (2016) doi:10.1103/PhysRevD.94.106002 [arXiv:1604.07818 [hep-th]].
- [41] J. Maldacena, D. Stanford and Z. Yang, “Diving into traversable wormholes,” *Fortsch. Phys.* **65**, no. 5, 1700034 (2017) doi:10.1002/prop.201700034 [arXiv:1704.05333 [hep-th]].
- [42] D. Harlow and D. Jafferis, “The Factorization Problem in Jackiw-Teitelboim Gravity,” arXiv:1804.01081 [hep-th].
- [43] V. Balasubramanian and P. Kraus, “A Stress tensor for Anti-de Sitter gravity,” *Commun. Math. Phys.* **208**, 413 (1999) doi:10.1007/s002200050764 [hep-th/9902121].