

# Horizon Instability of the Extremal BTZ Black Hole

Samuel E. Gralla,<sup>1,\*</sup> Arun Ravishankar,<sup>1,†</sup> and Peter Zimmerman<sup>2,‡</sup>

<sup>1</sup>*Department of Physics, University of Arizona*

<sup>2</sup>*Max Planck Institute for Gravitational Physics (Albert Einstein Institute)  
Am Mühlenberg 1, 14476 Potsdam, Germany*

(Dated: June 4, 2020)

We study real-time propagation of a massive scalar field on the extremal BTZ black hole space-time, focusing on the Aretakis instability of the event horizon. We obtain a simple time-domain expression for the  $\text{AdS}_3$  retarded Green function with Dirichlet boundary conditions and construct the corresponding time-domain BTZ retarded Green function using the method of images. The field decays at different rates on and off the horizon, indicating that transverse derivatives grow with time on the horizon (Aretakis instability). We solve the null geodesic equation in full generality and show that the instability is associated with a class of null geodesics that orbit near the event horizon arbitrarily many times before falling in. In an appendix we also treat the problem in the frequency domain, finding consistency between the methods.

---

\* [sgralla@email.arizona.edu](mailto:sgralla@email.arizona.edu)

† [arunravishankar@email.arizona.edu](mailto:arunravishankar@email.arizona.edu)

‡ [peterzimmerman@aei.mpg.de](mailto:peterzimmerman@aei.mpg.de)

# CONTENTS

|                                                          |    |
|----------------------------------------------------------|----|
| I. Introduction                                          | 3  |
| II. Metric                                               | 4  |
| A. Global $\text{AdS}_3$                                 | 4  |
| B. Poincaré coordinates and patch                        | 4  |
| C. Extremal coordinates and patch                        | 5  |
| D. Horizon coordinates                                   | 6  |
| E. Extremal Killing field and BTZ                        | 6  |
| III. Scalar field                                        | 6  |
| A. $\text{AdS}_3$ Green function                         | 7  |
| B. Extremal BTZ Green function                           | 8  |
| IV. Late-time Decay                                      | 10 |
| A. $\text{AdS}_3$                                        | 10 |
| B. BTZ                                                   | 11 |
| 1. Both points outside the horizon                       | 11 |
| 2. Field point on the horizon and source point outside   | 12 |
| 3. Both points on the horizon                            | 13 |
| V. Null geodesics                                        | 13 |
| Acknowledgements                                         | 17 |
| A. $\text{AdS}_3$ Green function                         | 17 |
| B. Massless axisymmetric perturbations: Aretakis' method | 19 |
| C. Mode approach                                         | 20 |
| 1. Preliminaries                                         | 20 |
| 2. Modes at late times near the horizon                  | 21 |
| 3. Non-periodic limit - $\text{AdS}_3$                   | 22 |
| References                                               | 23 |

## I. INTRODUCTION

In the decade since Aretakis’ initial study of massless scalar fields in the extremal Reissner-Nördstrom space-time [1, 2], it has become clear that extremal horizons generically exhibit weak derivative instabilities [3–5]. Independent of the type of extremal black hole (and even of the type of field that perturbs it!), the evidence suggests that sufficiently high-order transverse derivatives *always* grow at least polynomially in advanced time along the event horizon. This has the physical consequence that infalling observers experience large field gradients [4, 6, 7].

This general picture has emerged from a variety of techniques, whose domains of validity are largely non-overlapping. The original technique of Aretakis [3, 8, 9] involves a conserved quantity along the horizon that provides an obstruction to decay. This technique has been used only in what we call the “discrete case” (e.g., axisymmetric massless perturbations of Kerr [5]), and mainly when initial data extends to the horizon. We ourselves have been involved with a different technique based on frequency-domain analysis, in which the instability is associated with a singular branch point in each mode of the retarded Green function [6, 10–12]. This technique has been used mainly in the *non*-discrete case (e.g., non-axisymmetric mode perturbations of Kerr [6]), and only when initial data is confined *away* from the horizon. Other techniques rely on discrete inversion symmetries present only for some black holes [4, 13–15]. Analytic arguments involving the near-horizon geometry [5, 6, 16, 17] together with numerical work in a variety of settings [4, 18, 19], have provided complimentary insight into these various regimes. However, a fully general understanding of the instability remains elusive.

There is a dissonance here: a universal phenomenon should be simple at its core, yet the state of the art presents a sprawling complexity. One suspects that a key unifying idea has not yet been identified. In such a situation, it pays to study the effect in the simplest possible setting, where the nature of the phenomenon might be revealed with a minimum of distraction. Buoyed by this hope, in this paper we initiate the study of the Aretakis instability in the exceptionally simple setting of a 2+1-dimensional black hole—the Bañados-Teitelboim-Zanelli (BTZ) spacetime [20, 21]. As this black hole has negative cosmological constant, the results may also help understand the holographic implications of the instability [22, 23].

The BTZ black hole allows exceptional analytic control because it is locally isometric to three-dimensional Anti-de Sitter space ( $\text{AdS}_3$ ), a maximally symmetric spacetime. We focus on the retarded Green function of a massive scalar field satisfying Dirichlet boundary conditions. Using the results of [24] (and correcting some minor computational errors), we write the  $\text{AdS}_3$  retarded Green function in a very simple form. We then use the method of images [25] to construct the BTZ Green function.

The Aretakis instability emerges from this image sum in a beautiful way. Each term in the sum decays at the same rate, fixed by the  $\text{AdS}_3$  spacetime. For a field point off the horizon in BTZ, the high-order images are unimportant, and the full Green function (and hence field) also decays at this  $\text{AdS}_3$  rate. However, when the field point is chosen on the BTZ horizon, the high-order images become important and the infinite sum decays at precisely *half* the  $\text{AdS}_3$  rate,<sup>1</sup> as expected on general grounds [5]. Since the field decays at different rates on and off the horizon, sufficiently high-order transverse derivatives on the horizon must grow—the Aretakis instability.

By solving the null geodesic equation in BTZ, we are able to identify the late-time-relevant terms in the image sum with the arrival of wavefronts that have orbited the event horizon arbitrarily many times before falling in. The number of orbits increases linearly as the geodesic conserved quantity approaches that of the event horizon. We conjecture that this behavior is part of the following larger pattern for black holes: The surface gravity of the event horizon functions as a Lyapunov exponent for the deviation of nearby null geodesics; and when this exponent vanishes for an extremal black hole, the exponential deviation goes over to a power law. One can then say that wavefronts linger far longer near the event horizon of extremal black holes, as compared to analogous non-extremal black holes, providing an interpretation of the instability of extremal horizons.

---

<sup>1</sup> We demonstrate this fact numerically and give heuristic analytical arguments for the behavior, but do not provide a rigorous proof. Furthermore, this analysis assumes that the source point is outside the horizon, or equivalently that the initial data does not extend to the horizon.

Our results offer a new way to think about the Aretakis instability, but they do not solve the problem of unifying the previous different approaches. To provide a point of comparison, in the appendices we apply two other methods to the extremal BTZ spacetime: the conserved charge method (App. B) and the mode sum method (App. C). We cannot compare with the conserved charged method because our analysis is restricted to the non-discrete case, with initial data not extending to the horizon. We can compare with the decay rate of individual modes in the frequency domain analysis (and we find agreement), but this method does not give the behavior of the full field (see further discussion in App. C). Thus there remains much mystery surrounding the Aretakis instability of extremal horizons.

This paper is organized as follows. In Sec. II we review the  $\text{AdS}_3$  and extremal BTZ spacetimes, introducing notation. In Sec. III we consider a massive scalar field and construct the retarded Green function by the method of images. In Sec. IV we discuss late-time decay, and in Sec. V we interpret the instability in terms of null geodesics. Our signature is  $(-++)$  and we set the AdS radius to one.

## II. METRIC

We now discuss the  $\text{AdS}_3$  metric and its periodic identification to produce the extremal BTZ black hole. We introduce a few different coordinate patches that are useful in the analysis that follows.

### A. Global $\text{AdS}_3$

The maximally extended  $\text{AdS}_3$  metric is given in “global coordinates” by

$$ds^2 = \frac{1}{\cos^2 \chi} (-d\tau^2 + d\chi^2 + \sin^2 \chi d\Omega^2), \quad (1)$$

where  $\tau \in (-\infty, \infty)$ ,  $\chi \in [0, \pi/2)$ , and  $\Omega \sim \Omega + 2\pi$ . The spacetime can be conformally completed by a cylinder at  $\chi = \pi/2$ , which we refer to as the boundary. We will generally use global coordinates to visualize the spacetime, with  $(\tau, \chi, \Omega)$  treated as cylindrical coordinates representing height, radius, and angle, respectively.

### B. Poincaré coordinates and patch

“Poincaré coordinates” for  $\text{AdS}_3$  are given by

$$t = \frac{\sin \tau}{\cos \tau - \cos \Omega \sin \chi}, \quad z = \frac{\cos \chi}{\cos \tau - \cos \Omega \sin \chi}, \quad x = \frac{\sin \Omega \sin \chi}{\cos \tau - \cos \Omega \sin \chi}, \quad (2)$$

where the metric becomes

$$ds^2 = \frac{1}{z^2} (-dt^2 + dz^2 + dx^2). \quad (3)$$

These coordinates cover only a “Poincaré patch” where  $(\cos \tau - \cos \Omega \sin \chi)$  has a definite sign. We choose the region  $\tau \in (\tau_-, \tau_+)$ , with

$$\tau_{\pm} := \pm \arccos(\cos \Omega \sin \chi). \quad (4)$$

The bounding null surfaces  $\tau = \tau_{\pm}$  are called Poincaré horizons (Fig. 1 left). The interior has  $z > 0$  with  $t$  and  $x$  unbounded. The boundary is at  $z = 0$ , and all coordinates become large as the Poincaré horizons are approached.

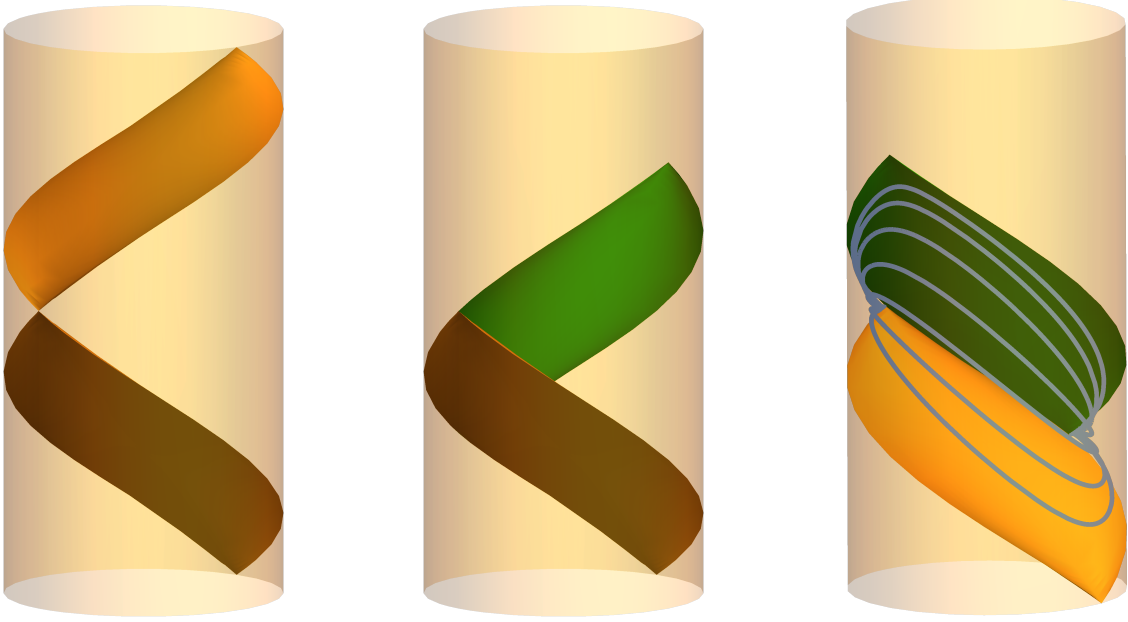


FIG. 1.  $\text{AdS}_3$  and its patches, plotted using global coordinates  $(\tau, \chi, \Omega)$  as cylindrical coordinates with height  $\tau$ , radius  $\chi$ , and angle  $\Omega$ . The maximally extended spacetime has a cylindrical conformal boundary at  $\chi = \pi/2$ , shown here in translucent orange. On the left, we show the Poincaré horizons  $\tau_{\pm}$  that bound the Poincaré patch. In the middle, we show  $\tau_-$  along with the future horizon  $\tau_H$  (which is also a Poincaré horizon); these bound the “extremal patch” that becomes the BTZ exterior after identification. Finally, on the right we show the same plot from a different perspective, along with a selection of integral curves of the Killing field  $\xi = \partial_X$ . The extremal BTZ black hole is obtained by identifying discretely along these curves.

### C. Extremal coordinates and patch

We also introduce “extremal coordinates” for  $\text{AdS}_3$ ,

$$t = \frac{1}{2} \left( T + X - \frac{R_H}{R^2 - R_H^2} - \frac{e^{2R_H(X-T)}}{2R_H} \right) \quad (5a)$$

$$z = \frac{e^{R_H(X-T)}}{\sqrt{R^2 - R_H^2}} \quad (5b)$$

$$x = \frac{1}{2} \left( T + X - \frac{R_H}{R^2 - R_H^2} + \frac{e^{2R_H(X-T)}}{2R_H} \right), \quad (5c)$$

bringing the metric into the form

$$ds^2 = -(R^2 - 2R_H^2)dT^2 + \frac{R^2}{(R^2 - R_H^2)^2}dR^2 - 2R_H^2dTdX + R^2dX^2. \quad (6)$$

These coordinates cover the region  $R > R_H$ , which is bounded by the past Poincaré horizon  $\tau_-$  and a globally translated and rotated Poincaré horizon  $\tau_H$  defined by<sup>2</sup>

$$\tau_H = \arcsin(\sin \Omega \sin \chi). \quad (7)$$

<sup>2</sup> In global coordinates,  $R > R_H$  corresponds to  $\alpha\beta > 0$  with  $\alpha = \sin \Omega \sin \chi - \sin \tau$  and  $\beta = \cos \tau - \cos \Omega \sin \chi$ . The zeros of  $\beta$  bound a set of Poincaré patches including our choice (4), while the zeros of  $\alpha$  bound an interleaving set related by the symmetries  $\tau \rightarrow \tau + \pi/2$  and  $\phi \rightarrow \phi + \pi/2$ . As our Poincaré patch (4) has  $\beta > 0$ , we require  $\alpha > 0$  as well, giving rise to  $\tau \in (\tau_-, \tau_H)$ .

We refer to this patch  $\tau \in (\tau_-, \tau_H)$  as the extremal patch (Fig. 1 middle and right). We call the bounding null surfaces  $\tau_-$  and  $\tau_H$  the past and future horizons (respectively), since these will become the event horizons of the extremal BTZ black hole after the identification  $X \sim X + 2\pi$ . These horizons are described by  $R = R_H$  in extremal coordinates, but the patch itself is independent of  $R_H$ .<sup>3</sup> The inverse transformation is given by

$$T = \frac{2x^2 - 2t^2 + z^2}{4(x-t)} - \frac{1}{4R_H} \log(2R_H(x-t)) \quad (8a)$$

$$R = R_H \sqrt{1 + \frac{2(x-t)}{R_H z^2}} \quad (8b)$$

$$X = \frac{2x^2 - 2t^2 + z^2}{4(x-t)} + \frac{1}{4R_H} \log(2R_H(x-t)). \quad (8c)$$

#### D. Horizon coordinates

Finally, it will be useful to consider “horizon coordinates”  $(V, R, \Phi)$  defined by

$$T = V + \frac{R}{2(R^2 - R_H^2)} - \frac{1}{4R_H} \log \frac{R - R_H}{R + R_H} \quad (9a)$$

$$X = \Phi + V + \frac{R}{2(R^2 - R_H^2)} + \frac{1}{4R_H} \log \frac{R - R_H}{R + R_H}, \quad (9b)$$

where the metric becomes

$$ds^2 = 2dRdV + 2(R^2 - R_H^2)d\Phi dV + R^2 d\Phi^2. \quad (10)$$

These coordinates cover the future horizon  $R = R_H$ , with the null generators given by  $\Phi = \text{const.}$  When  $\Phi \sim \Phi + 2\pi$ , these are ingoing, corotating coordinates for the extremal BTZ black hole. The horizon-generating Killing field is given by

$$\zeta = \frac{\partial}{\partial V} = \frac{\partial}{\partial T} + \frac{\partial}{\partial X}. \quad (11)$$

#### E. Extremal Killing field and BTZ

The extremal BTZ black hole is obtained by identifying points separated by parameter distance  $2\pi$  along the integral curves of the Killing field

$$\xi = \frac{\partial}{\partial X} = \frac{\partial}{\partial \Phi}. \quad (12)$$

This means we identify  $X \sim X + 2\pi$  in extremal coordinates (6) and (equivalently)  $\Phi \sim \Phi + 2\pi$  in horizon coordinates (9). We will call  $\xi$  the extremal Killing field. Its integral curves connect the two ends of the “conformal bifurcation line” where the future and past horizons meet at the boundary (Fig. 1).

### III. SCALAR FIELD

We consider a massive scalar field  $\psi$ ,

$$(\square - \mu^2)\psi = 0. \quad (13)$$

---

<sup>3</sup> Note that the original transformation given by BTZ does entail a different patch for each value of  $R_H$ ; these patches approach Poincaré patch as  $R_H \rightarrow 0$ . Our extremal patch remains distinct in the  $R_H \rightarrow 0$  limit.

The AdS<sub>3</sub>/BTZ spacetimes are not globally hyperbolic, and well-posed evolution requires specification of the behavior of the field on the boundary. We will consider so-called “Dirichlet” conditions, where  $\psi$  is required to vanish on the boundary. We define an associated retarded Green function by

$$(\square - \mu^2) G(x, x') = \delta_3(x, x'), \quad (14)$$

together with the requirements that  $G$  vanish when either point approaches the boundary or if  $x'$  is not in the causal past of  $x$ . (Here  $\delta_3$  is the covariant delta distribution, equal to  $1/\sqrt{-g}$  times the coordinate delta function). We assume  $\mu^2 \geq -1$  so that the Dirichlet dynamics are well-posed [26, 27]. This Green function can be used to construct the field from its initial value  $\psi_0(R', \Phi')$  on the null surface  $V' = 0$  via the Kirchhoff representation<sup>4</sup>

$$\psi(V, R, \Phi) = R_H^{-1} \int d\Phi' \int_{R_H}^{\infty} dR' R' (\psi_0 \partial_{R'} G - G \partial_{R'} \psi_0), \quad (15)$$

where we have assumed that the field point is on or outside the event horizon. In AdS<sub>3</sub>, the range of  $\Phi'$  is unbounded, while in BTZ it must be restricted to a fiducial range such as  $0 < \Phi' \leq 2\pi$ . In all other respects, Eqs. (13), (14), and (15) hold equally well in AdS<sub>3</sub> and external BTZ.

### A. AdS<sub>3</sub> Green function

The Dirichlet retarded Green function for AdS<sub>3</sub> was obtained in closed form by [24]. In App. A we review the derivation, correct some trivial errors, and obtain a new form of the result:

$$G_{\text{ret}}^{\text{AdS}_3} = \frac{\Theta(t - t')\Theta(\Sigma)}{\pi\sqrt{|\Sigma(\Sigma - 2)|}} \begin{cases} -\cos 2\nu \arccos(1 - \Sigma) & \text{if } \Sigma < 2 \\ \sin(2\pi\nu)e^{-2\nu \text{arccosh}(\Sigma - 1)} & \text{if } \Sigma > 2 \end{cases}. \quad (16)$$

Here  $\nu$  is a scaling dimension for the field defined by

$$\nu = \frac{1}{2} \sqrt{1 + \mu^2}, \quad (17)$$

while  $\Sigma$  is a biscalar on AdS<sub>3</sub> given by

$$\Sigma = \frac{(t - t')^2 - (x - x')^2 - (z - z')^2}{2zz'} \quad (18)$$

$$= 1 - \cos(\tau - \tau') \sec \chi \sec \chi' + \cos(\Omega - \Omega') \tan \chi \tan \chi'. \quad (19)$$

The Green function is singular on the surfaces  $\Sigma = 0$  and  $\Sigma = 2$ , whose union is the null wave front of the propagating field. This wave front bounces off the boundary as shown in Fig. 2. In the “exceptional” case where  $\nu$  is a half-integer, the Green function vanishes for  $\Sigma > 2$ , so that the returning wavefront appears to “cancel out” the propagating field. Although the wavefronts appear conical in the plots, they are not precisely equal to cones. The distinction is most graphically evident when the source point is near the boundary, in which case the wavefronts collapse to Poincaré horizons.

We now discuss the geometric interpretation of  $\Sigma$  and its role in the Green function. In a Poincaré patch there is always a unique geodesic connecting two points, and  $\Sigma$  is related to the world function  $\sigma$  (one-half the squared geodesic distance) by [29]

$$\Sigma = 2 \left( 1 - \cosh^2 \left( \frac{\sqrt{2}\sigma}{2} \right) \right) = -\sigma + \mathcal{O}(\sigma^2). \quad (20)$$

---

<sup>4</sup> To derive equation (15), we follow the steps of Ref. [28] using a volume bounded by the null surface  $V = 0$ , the future event horizon, and the boundary. The contribution at the boundary vanishes due to the Dirichlet condition.

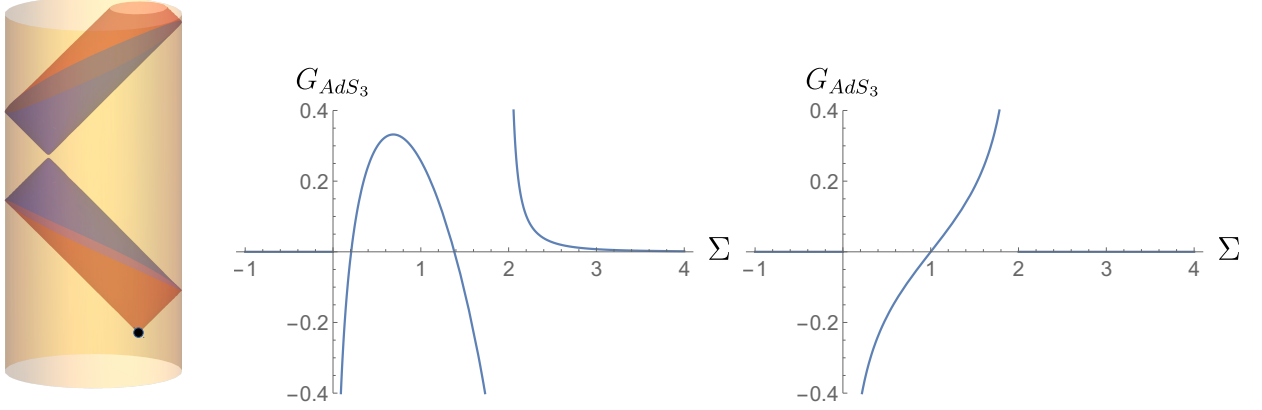


FIG. 2. Retarded propagation in  $\text{AdS}_3$  with Dirichlet boundary conditions. On the left, we show the wavefront of a point source (surface of singularity of the retarded Green function), which “bounces” off the boundary. The wavefront is described by  $\Sigma = 0$  (orange) and  $\Sigma = 2$  (blue). On the right, we show the behavior of the retarded green function as a function of  $\Sigma$ . Two qualitative behaviors exist: the “generic” case (middle; we have chosen  $\nu = 1.2$ ) and the “exceptional” case (we have chosen  $\nu = 0.5$ ). In the exceptional case, the Green function vanishes for  $\Sigma > 2$ .

In particular,  $\Sigma > 0$  is timelike separation,  $\Sigma < 0$  is spacelike separation, and  $\Sigma = 0$  is null separation. However, we have seen that  $\Sigma = 2$  also represents a null wavefront of the propagating field. This is not a contradiction, since the bouncing behavior of the wave fronts is *not* reflected in the behavior of geodesics, which simply continue indefinitely toward the boundary as the affine parameter increases. Thus, at least within the Poincaré patch, points for which  $\Sigma = 2$  are indeed connected by only a single geodesic, which is timelike, despite the fact that they are also connected by a bouncing wave front of the scalar field. The null character of the surface  $\Sigma = 2$  (fixing the prime point) can be seen directly by observing that it corresponds to  $\Sigma = 0$  for “conjugate points”  $\tau' \rightarrow \tau' + n\pi$ ,  $\Omega' \rightarrow \Omega' + n\pi$  for odd  $n$  (see Fig. 2). Finally, note that the wavefront preserves the inverse square root character of the singularity in the Green function inherited from the local Hadamard form. See e.g. [30] and references therein for rigorous results on the global propagation of singularities.

### B. Extremal BTZ Green function

As the BTZ black hole is a periodic identification of  $\text{AdS}_3$ , propagation in BTZ is equivalent to propagation on a portion of  $\text{AdS}_3$  subject to periodic boundary conditions. The retarded Green function can then be simply constructed from the non-periodic one by

$$G_{\text{ret}}^{\text{BTZ}}(x, x') = \sum_{n=-\infty}^{\infty} G_{\text{ret}}^{\text{AdS}_3}(x, e^{2\pi n \xi} x') = \sum_{n=-\infty}^{\infty} G_{\text{ret}}^{\text{AdS}_3}|_{\Phi' \rightarrow \Phi' + 2\pi n}, \quad (21)$$

where the last equation holds only in horizon coordinates. The notation  $e^{2\pi n \xi} x'$  indicates the spacetime point obtained by flowing along the integral curves of  $\xi$  for a parameter distance  $2\pi n$ ; when horizon coordinates are used, this simply increases  $\Phi'$  by  $2\pi n$ . It is straightforward to check that this expression satisfies all the conditions for the Dirichlet retarded Green function, together with the additional required periodicity. From the perspective of  $\text{AdS}_3$ , we place “image sources” at distances of  $2\pi n$  along the integral curves of  $\xi$  from the original source position (Fig. 3), all of which emit wave fronts. From the perspective of BTZ, the arriving fronts are interpreted as having circled the black hole  $|n|$  times. This method of constructing the Green function is generally called the method of images; it has been used extensively in Euclidean signature, but apparently not for retarded propagation.

It will be convenient to define a separate  $\Sigma_n$  for each image  $n$ . Letting

$$\delta V = V - V', \quad \delta \Phi_n = \Phi - \Phi' - 2\pi n, \quad (22)$$

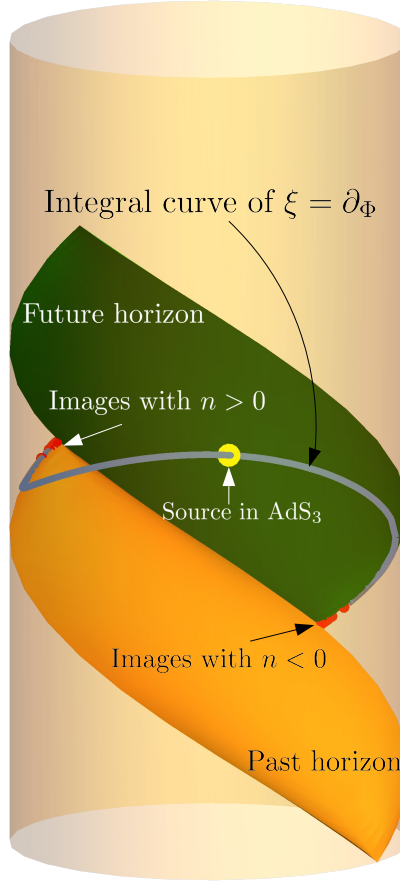


FIG. 3. Method of images for the extremal BTZ Green function. A source point (yellow) in the fiducial portion of  $\text{AdS}_3$  defines image points (red) appearing every  $2\pi$  along the integral curve of  $\xi = \partial_\Phi$ .

then (expressing in horizon coordinates) we have

$$\begin{aligned} \Sigma_n = 1 - \frac{e^{R_H \delta \Phi_n}}{2R_H} \left( \frac{R' - R}{2} + R_H + \left( \delta V + \frac{\delta \Phi_n}{2} \right) (R' + R_H)(R - R_H) \right) \\ + \frac{e^{-R_H \delta \Phi_n}}{2R_H} \left( \frac{R' - R}{2} - R_H + \left( \delta V + \frac{\delta \Phi_n}{2} \right) (R' - R_H)(R + R_H) \right). \end{aligned} \quad (23)$$

The other component of the  $\text{AdS}_3$  retarded Green function (16) is a theta function,  $\Theta(t - t')$ , that cuts out the past of  $t'$ . We will introduce the notation

$$s_n = t - t'|_{\Phi' \rightarrow \Phi' + 2\pi n}, \quad (24)$$

indicating that one is to express  $t$  and  $t'$  in terms of horizon coordinates and then send  $\Phi' \rightarrow \Phi' + 2\pi n$ . Explicitly, we have

$$s_n = \frac{1}{4(R + R_H)(R' + R_H)} \left[ e^{2R_H(\Phi' + 2\pi n)} (R + R_H)(R' - R_H) - e^{2R_H\Phi} (R' + R_H)(R - R_H) + 2R_H(R' - R + (R + R_H)(R' + R_H)(2\delta V + \delta \Phi_n)) \right]. \quad (25)$$

Putting everything together, the BTZ Dirichlet retarded Green function is written

$$G_{\text{ret}}^{\text{AdS}_3} = \sum_{n=-\infty}^{\infty} \left( \frac{\Theta(s_n)\Theta(\Sigma_n)}{\pi\sqrt{|\Sigma_n(\Sigma_n-2)|}} \begin{cases} -\cos 2\nu \arccos(1-\Sigma_n) & \text{if } \Sigma_n < 2 \\ \sin(2\pi\nu)e^{-2\nu\text{arccosh}(\Sigma_n-1)} & \text{if } \Sigma_n > 2 \end{cases} \right), \quad (26)$$

where  $\Sigma_n$  and  $s_n$  are given in Eqs. (23) and (25), respectively. The non-uniform behavior of  $\Sigma_n$  in the limits  $V \rightarrow \infty$ ,  $n \rightarrow \pm\infty$ ,  $R' \rightarrow R_H$  and  $R \rightarrow R_H$  is key to the phenomena discussed below.

#### IV. LATE-TIME DECAY

We now use the explicit retarded Green function to investigate late-time decay of field perturbations. We will discuss decay in null directions, varying  $V$  while fixing  $R$  and  $\Phi$ . When  $R = R_H$  this corresponds to decay along a horizon generator  $\Phi$ , measured by affine parameter  $V$ . When  $R > R_H$  the null ray remains outside the horizon.

##### A. AdS<sub>3</sub>

We begin with the case of AdS<sub>3</sub>. The AdS<sub>3</sub> Green function is simply the  $n = 0$  term in the sum (26). At large  $\delta V$  we have

$$s_0 \sim R_H \delta V, \quad \delta V \rightarrow \infty, \quad (27)$$

$$\Sigma_0 \sim C_0(R, R', \delta\Phi_0) \delta V, \quad \delta V \rightarrow \infty, \quad (28)$$

with

$$C_0 = \frac{1}{2R_H} \left[ e^{-R_H \delta\Phi_0} (R' - R_H)(R + R_H) - e^{R_H \delta\Phi_0} (R' + R_H)(R - R_H) \right], \quad (29)$$

where we exclude the measure-zero case  $C_0 = 0$ . If  $C_0 < 0$  then  $\Sigma_0$  is negative at late times and the Green function vanishes: the points of fixed  $R$  and  $\Phi$  with large  $V$  are out of causal contact with the source position  $R', \Phi', V'$ . When  $C_0 > 0$ , the Green function at late times is

$$G \sim \frac{4^{-\nu} C_0}{\pi} \sin(2\pi\nu) (\delta V)^{-(2\nu+1)}, \quad \delta V \rightarrow \infty. \quad (30)$$

If  $\nu$  is a half-integer then this expression vanishes. In fact the Green function (26) vanishes identically in this case,

$$G = 0, \quad \delta V > V_0, \quad (31)$$

where  $V_0$  is some constant. By the Kirchhoff integral (15), the field sourced by generic initial data will behave the same way,

$$\psi \sim C(R, \Phi) V^{-2h}, \quad V \rightarrow \infty \quad (2h \notin \mathbb{Z}^+) \quad (32)$$

$$\psi = 0, \quad V > V_0 \quad (2h \in \mathbb{Z}^+). \quad (33)$$

where  $C$  is some function determined by the initial data, and we introduced

$$h = \frac{1}{2} + \nu \quad (34)$$

in order to be consistent with previous work [5–7]. Notice that  $h \geq 1/2$  given our assumption on the mass  $\mu^2 \geq -1$ . We will refer to the case  $2h \notin \mathbb{Z}^+$  as “generic” and the case  $2h \in \mathbb{Z}^+$  as “exceptional”. This exceptional case includes the case  $h \in \mathbb{Z}^+$  previously called “discrete”.

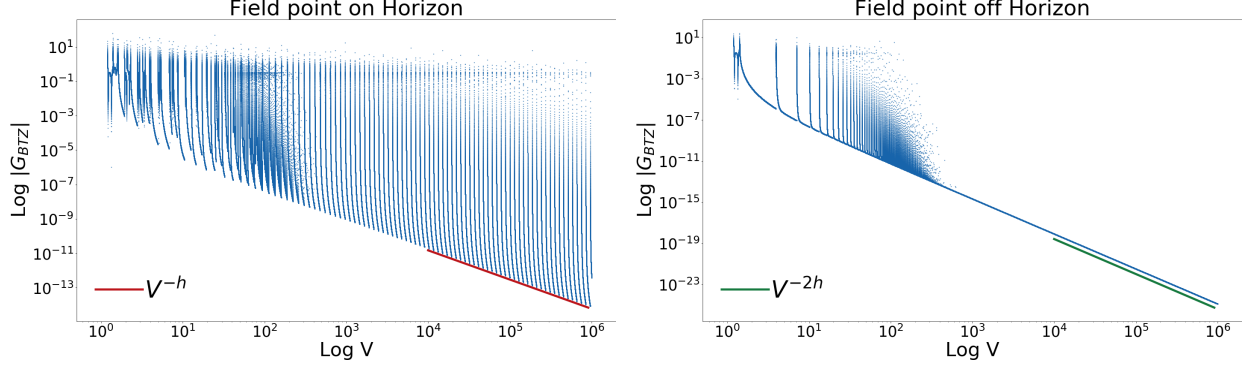


FIG. 4. The BTZ retarded Green function with field point off the horizon, sampled uniformly in  $\log V$ . The field decays at different rates depending on whether the field point is on the horizon (left plot,  $V^{-h}$ ) or off the horizon (right plot  $V^{-2h}$ ). The slower decay is due to spikes from wavefronts that orbit many times near the black hole and cross the horizon at arbitrarily late times. Spikes also arrive at arbitrarily late times from wavefronts that spend their time mainly far away from the black hole, but they are exponentially narrow and are not visible in the plots at late times (and do not affect the decay envelope). In these plots we have chosen  $R_H = 0.01$  and  $\nu = 1.2$  such that  $h = 1.7$ , with  $\delta\Phi_0 = 0$  and  $R' = 10$ ; in the left plot we have  $R = R_H = 0.01$ , while in the right plot we have  $R = 1 > R_H$ .

## B. BTZ

For a BTZ black hole, we must consider the infinite sum (26). Each term in the sum is an  $\text{AdS}_3$  Green function with a different source point, so each term behaves like  $\sin(2\pi\nu)V^{-2h}$  at late times as above. However, this does not guarantee that the full sum shares this behavior, and indeed we will see that it does not. To examine the full sum, it is helpful to rewrite (23) as

$$\Sigma_n = 1 + e^{2\pi n R_H} (A_+ + B_+(\delta V - \pi n)) - e^{-2\pi n R_H} (A_- + B_-(\delta V - \pi n)), \quad (35)$$

with

$$A_{\pm} = \frac{e^{\mp R_H(\Phi - \Phi')}}{2R_H} \left( \frac{R' - R}{2} \mp R_H \right) \quad (36a)$$

$$B_{\pm} = \frac{e^{\mp R_H(\Phi - \Phi')}}{2R_H} (R' \mp R_H)(R \pm R_H). \quad (36b)$$

To explore the deviation of the full sum from the behavior of an individual term we consider the regime of late times and large image numbers, i.e. large  $\delta V$  and large  $|n|$ . We restrict to the generic case  $2h \notin \mathbb{Z}^+$ .

### 1. Both points outside the horizon

Suppose now that both the source and field point are outside the horizon,  $R > R_H$  and  $R' > R_H$ . Then  $B_{\pm}$  are both non-zero, so in the regime of large  $\delta V$  and  $|n|$  we have

$$\Sigma_n \approx \begin{cases} B_+ e^{2\pi n R_H} (\delta V - \pi n), & n > 0 \\ -B_- e^{2\pi |n| R_H} (\delta V + \pi |n|), & n < 0 \end{cases}, \quad \delta V \gg 1, |n| \gg 1. \quad (37)$$

The contribution of the  $n^{\text{th}}$  term to the Green function is important when  $\Sigma_n$  is positive and of order unity (see Fig. 2), with the initial spike occurring when  $\Sigma_n = 0$ . Since  $B_{\pm}$  are positive, we see that the negative  $n$  terms do not contribute ( $\Sigma_n$  is negative), while the positive  $n$  terms are important for times of order the image number ( $\delta V \sim n$ ). The importance of the contribution can be assessed from the derivative

$$\frac{\partial \Sigma_n}{\partial V} \approx B_+ e^{2\pi n R_H}, \quad (n > 0). \quad (38)$$

Since this derivative is exponentially large, the width (in  $V$ ) of the region of importance is exponentially suppressed with image number. This suggests that the large- $n$  terms are not very important, and (aside from brief exponentially narrowing spikes) the full behavior of the sum will share the falloff  $V^{-2h}$  of each individual term. Plotting the Green function numerically supports this conclusion (Fig. 4 left).

The behavior of the field  $\psi$  follows from that of the Green function via the Kirchhoff integral (15). It is clear from the discussion above that the Green function decays as  $V^{-2h}$  apart from exponentially narrowing spikes that may be safely excluded from the integral. This late-time approximation is uniformly valid over any region of  $R'$  outside (but not including!) the event horizon, so the field will similarly decay as  $V^{-2h}$  provided the initial data is confined outside the horizon. The spikes in the Green function will be smoothed out by the integral into finite oscillations about this decay. That is, we claim that the generic outside-horizon, late-time behavior of fields sourced by initial data outside the horizon is

$$\psi \sim \psi_0 V^{-2h}, \quad R > R_H, V \rightarrow \infty \quad (39)$$

where  $\psi_0$  is an arbitrary function of  $R, \Phi$  and an  $O(1)$  function of  $V$ .

## 2. Field point on the horizon and source point outside

If the field point is now on the horizon ( $R = R_H$ ) while the source is still outside ( $R' > R_H$ ), then  $B_- = 0$  and at large  $\delta V$  and  $|n|$  we have

$$\Sigma_n \approx \begin{cases} B_+ e^{2\pi n R_H} (\delta V - \pi n), & n > 0 \\ B_+ e^{-2\pi |n| R_H} (\delta V - \pi n) - A_- e^{2\pi |n| R_H}, & n < 0 \end{cases}, \quad \delta V \gg 1, |n| \gg 1. \quad (40)$$

which may be compared with Eq. (37). The formula for the positive  $n$  terms is identical, and these terms again contribute at  $\delta V \sim n$ . However, now the negative  $n$  terms contribute as well, starting at even later times  $\delta V \sim e^{4\pi |n| R_H}$  (where  $\Sigma_n$  becomes positive). The derivatives in these regimes are given by

$$n > 0 : \quad \frac{\partial \Sigma_n}{\partial V} \approx B_+ e^{2\pi |n| R_H}, \quad (n \rightarrow \infty, \quad \delta V \gtrsim \pi n), \quad (41)$$

$$n < 0 : \quad \frac{\partial \Sigma_n}{\partial V} \approx B_+ e^{-2\pi |n| R_H} \quad (n \rightarrow -\infty, \quad \delta V \gtrsim \frac{A_-}{B_+} e^{4\pi |n| R_H}), \quad (42)$$

which may be compared to Eq. (38). The positive  $n$  terms again have a large derivative, indicating an exponentially narrow contribution to the Green function. However, the newly contributing negative  $n$  terms have a small derivative, indicating an exponentially *wide* spike in the Green function as a function of  $V$ . Thus the negative  $n$  spikes should be very important at late times, and there is no reason to expect that the sum has the same falloff as an individual spike. Plotting the Green function (Fig. 4 right) confirms the exponentially wide spikes and shows that the actual decay is  $V^{-h}$ .

We can understand this rate through the following heuristic argument. Let us label the exponentially wide spikes by  $m = -n$ , so that  $m$  is a positive integer. As discussed above, the  $m^{\text{th}}$  spike arrives at a time of order  $\delta V \sim e^{4\pi m R_H}$ . (In this discussion we drop the  $m$ -independent constants  $A_-$  and  $B_+$ .) At the precise time of arrival we have  $\Sigma_m = 0$ , but shortly thereafter (including the majority of the time before the arrival of the next spike) we have  $\Sigma_m \sim e^{2\pi m R_H} \gg 1$ . Thus we may approximate each image's contribution  $G_m$  to the green function as its large- $\Sigma$  behavior  $G_m \sim \Sigma_m^{-2h}$  [Eqs. (A14) and (34)]. Focusing on the period of order  $\delta V \sim e^{4\pi m R_H}$  before the next spike has arrived, we may equivalently express this in terms of  $m$  or  $\delta V$  as

$$G_m \sim \Sigma_m^{-2h} \sim e^{-4\pi h m R_H} \sim (\delta V)^{-h}, \quad \delta V \sim e^{4\pi m R_H}, \quad (43)$$

recalling that  $m$ -independent constants are dropped. This shows the  $(\delta V)^{-h}$  decay visible in Fig. 4. In essence, each image decays as  $C_m (\delta V)^{-2h}$ , but the effective amplitudes  $C_m$  grow as  $(\delta V)^h$ , resulting in an overall  $(\delta V)^{-h}$  decay.

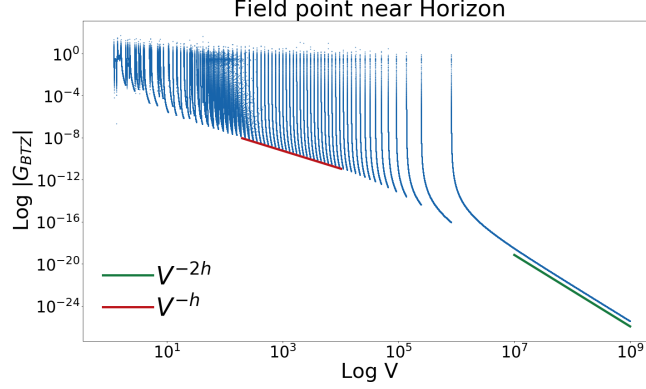


FIG. 5. Illustration of transient  $V^{-h}$  decay before ultimate  $V^{-2h}$  decay for field points near the horizon. The parameters are the same as Fig. 4 except we take  $(R - R_H)/R_H = 10^{-3}$ .

By similar arguments made for Eq. (39) above, we expect that the generic behavior of fields sourced by initial data outside the horizon is thus

$$\psi \sim \psi_H V^{-h}, \quad R = R_H, V \rightarrow \infty \quad (44)$$

where  $\psi_H$  is an arbitrary function of  $\Phi$  and an  $O(1)$  function of  $V$ .

For later comparison it will be helpful to have a more precise expression for the time of arrival of the  $n^{\text{th}}$  spike. This is obtained by letting  $\delta V \gg n$  and  $\Sigma_n = 0$  for  $\delta V$ , giving

$$\delta V \sim \pi n, \quad n \rightarrow \infty, \quad (45)$$

$$\delta V \sim \frac{e^{2R_H(\Phi - \Phi')}}{4R_H} \frac{R' + R_H}{R' - R_H} e^{4\pi|n|R_H}, \quad n \rightarrow -\infty. \quad (46)$$

If the field point is *near* the horizon but not exactly on it, one expects a transient period of  $V^{-h}$  decay followed by final decay of  $V^{-2h}$ . Plotting the Green function confirms this expectation (Fig. 5). The field  $\psi$  sourced by initial data outside the horizon will similarly show this transition from  $V^{-h}$  to  $V^{-2h}$  when evaluated near the horizon.

### 3. Both points on the horizon

If both the source and field point are on the horizon ( $R = R' = R_H$ ), then we have the simple expression

$$\Sigma_n = 1 - \cosh(R_H(\delta\Phi_0 - 2\pi n)). \quad (47)$$

In particular, the Green function is actually independent of  $V$  (apart from the causal factor  $\Theta(s_n)$ ). Of course, for a source point at any small distance off the horizon the Green function does decay, with the rate determined by whether the field point is on or off the horizon. That is, the late-time behavior of the Green function is highly non-uniform as the points approach the horizon. Determining the decay rate of fields with initial data that extends to the horizon would require a careful estimate of the Kirchhoff integral (15) using the full behavior of the Green function near the horizon. We leave this to future work.

## V. NULL GEODESICS

We now study the null geodesics of the BTZ spacetime in order to understand the behavior of the Green function that gives rise to the instability. Let  $p^\mu$  denote the four-momentum of the null geodesic, where  $p^V > 0$

is our time orientation. Using the null condition and the two Killing fields  $\zeta = \partial_V$  (horizon Killing field) and  $\xi = \partial_\Phi$  (axial Killing field), we have three conserved quantities,

$$g_{\mu\nu}\xi^\mu p^\nu = L, \quad g_{\mu\nu}\zeta^\mu p^\nu = L - E, \quad g_{\mu\nu}p^\mu p^\nu = 0. \quad (48)$$

The quantity  $L$  is interpreted as the angular momentum, while  $E$  is the energy according to the “static” Killing field  $\partial_T = \zeta - \xi = \partial_V - \partial_\Phi$ . Note that  $E$  can be negative for trajectories inside the “ergoregion”  $R < \sqrt{2}R_H$ , where  $\partial_T$  is spacelike [31].

The special case  $E = L = 0$  corresponds to the horizon generators,

$$E = L = 0 : \quad p \propto \partial_V, \quad R = R_H \quad (\text{horizon generators}). \quad (49)$$

Henceforth we will consider the region outside and including the horizon,

$$R \geq R_H > 0. \quad (50)$$

We will further assume  $E \neq 0$  and introduce

$$b = \frac{L}{E}, \quad k^\mu = \frac{dx^\mu}{d\lambda} = \frac{p^\mu}{|E|}, \quad (51)$$

which introduces an energy-rescaled affine parameter  $\lambda$  that increases toward the future. The case  $E = 0, L \neq 0$  can be handled straightforwardly by the limit  $b \rightarrow \infty$ , and we shall see that the case  $E = L = 0$  is recovered (more subtly) by  $b \rightarrow 1$ . Solving Eqs. (48) assuming  $R \geq R_H > 0$  and  $E \neq 0$ , we find

$$k^V = \frac{s(R^2 - bR_H^2) + R^2 k^R}{(R^2 - R_H^2)^2}, \quad (52a)$$

$$k^R = \pm \frac{\sqrt{(1-b)(-2bR_H^2 + (1+b)R^2)}}{R}, \quad (52b)$$

$$k^\Phi = -\frac{s(1-b) + k^R}{(R^2 - R_H^2)}, \quad (52c)$$

where  $s = \text{sgn}(E)$ .

The requirement that  $k^R$  be real imposes an allowed region of  $R$  for each value of  $b$ . Taking into account the assumption (50), we find

$$R_H \leq R \leq R_{\text{turn}}, \quad |b| > 1 \quad (53a)$$

$$R_H \leq R < \infty, \quad |b| < 1, \quad (53b)$$

with turning point radius (for  $|b| > 1$  only)

$$R_{\text{turn}} = \sqrt{\frac{2b}{b+1}} R_H. \quad (54)$$

Thus there are three kind of trajectories outside the horizon: transits from the (white hole) horizon to the boundary ( $|b| < 1$ ), transits from the boundary to the (black hole) horizon (also with  $|b| < 1$ ), and transits from the white hole to the black hole ( $|b| > 1$ ). There are no trajectories that begin at the boundary, turn, and end at the boundary.

The sign  $s = \text{sgn}(E)$  is fixed by the time orientation  $k^V > 0$  (equivalently  $p^V > 0$ ) and may be expressed directly in terms of  $b$ , as follows. Since  $|-bR_H^2 + R^2| \geq |R^2 k^R|$  for  $R \geq R_H$ , we have  $\text{sgn}(k^V) = s \times \text{sgn}(R^2 - bR_H^2)$  so that  $s = \text{sgn}(R^2 - bR_H^2)$  to ensure  $k^V > 0$ . Given the allowed ranges (53) of  $R$ , we find that

$$s = \text{sgn}(1 - b). \quad (55)$$

In the case  $b = 1$  the sign is indeterminate, with the limits  $b \rightarrow 1^\pm$  having qualitatively different behavior.

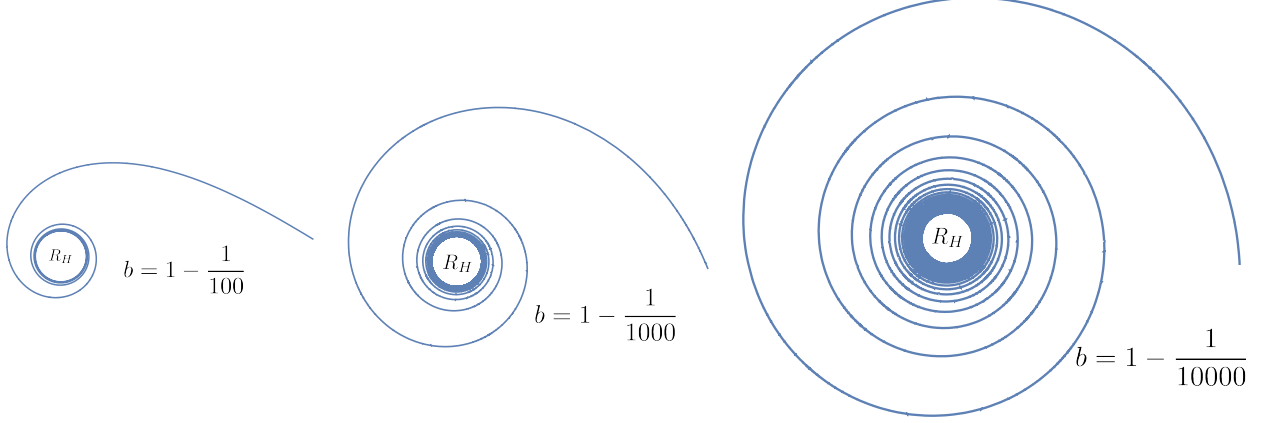


FIG. 6. Trajectories of the  $b \rightarrow 1^-$  null geodesics that account for the Aretakis instability. The angle  $\Phi$  used in the text co-rotates with the black hole, so we instead use  $\varphi = \Phi + V$  for these plots, treating  $(R, \varphi)$  as polar coordinates in the plane. The curves begin at a fixed  $R = R_0 > R_H$  and are terminated when the particle enters the horizon at  $R = R_H$ . The number of orbits scales as  $1/(1 - b)$ .

It is possible to solve Eqs. (52a) entirely for the trajectory  $x^\mu(\lambda)$ . Choosing the integration constants for simplicity, we find

$$V(\lambda) = \frac{1}{2} \left( \frac{1+b}{1-b} \right) \frac{s(1+b)\lambda \pm \text{sgn}((1-b^2)\lambda)R(\lambda)}{(R_H^2 - (1+b)^2\lambda^2)} - \frac{1}{2}\Phi_\pm(\lambda) \quad (56a)$$

$$R(\lambda) = \sqrt{\frac{2bR_H^2 + (1-b)(1+b)^2\lambda^2}{1+b}} \quad (56b)$$

$$\Phi(\lambda) = \mp \frac{1}{R_H} \log \left| \sqrt{\left| \frac{1+b}{1-b} \right|} \frac{R(\lambda) - \text{sgn}((1-b^2)\lambda)R_H}{R_H \pm s(1+b)\lambda} \right|, \quad (56c)$$

where  $\pm$  is the sign of  $dR/d\lambda$  (i.e.  $+$  when outgoing and  $-$  when ingoing). The three integration constants may be restored by shifting  $V$ ,  $\Phi$ , and  $\lambda$  by (separate) constant values. With the above choices, the range of  $\lambda$  is

$$-\infty < \lambda \leq R_H/(1+b) \quad |b| < 1, \text{ ingoing} \quad (57a)$$

$$R_H/(1+b) < \lambda < \infty \quad |b| < 1, \text{ outgoing} \quad (57b)$$

$$-R_H/(1+b) < \lambda \leq R_H/(1+b), \quad |b| > 1, \text{ turning at } \lambda = 0 \quad (57c)$$

As  $\lambda \rightarrow \pm R_H/(1+b)$  the particle approaches the horizon  $R \rightarrow R_H$ , whereas as  $\lambda \rightarrow \pm\infty$  the particle approaches the boundary  $R \rightarrow \infty$ .

We now fix a starting radius  $R_0 > R_H$  and compute the total change in time ( $\Delta V = V - V_0$ ) and angle ( $\Delta\Phi = \Phi - \Phi_0$ ) before the particle enters the horizon. These quantities also depends on the initial radial direction (ingoing or outgoing) when  $|b| > 1$ , since the particle will encounter a turning point if initially directed outwards. Considering only trajectories that end at the horizon, We will label initially ingoing trajectories (present for all  $b$ ) with “nt” (for “no turning point”) and initially outgoing trajectories (present only for  $|b| > 1$ )

with “t” for “has a turning point”. Eqs. (56) show that the lapse in time and angle is

$$\Delta V_{\text{nt}} = \frac{b}{2R_H(1-b)} - \frac{R_a - \text{sgn}(1-b)R_0}{2(R_0^2 - R_H^2)} - \frac{1}{2}\Delta\Phi_{\text{nt}}, \quad \forall b \quad (58a)$$

$$\Delta\Phi_{\text{nt}} = -\frac{\text{sgn}(1-b)}{R_H} \log \left| \frac{R_0 + R_H}{R_H + \text{sgn}(1-b)R_a} \right|, \quad \forall b \quad (58b)$$

$$\Delta V_t = \frac{b}{2R_H(1-b)} - \frac{-R_a + \text{sgn}(b)R_0}{2(R_0^2 - R_H^2)} - \frac{1}{2}\Delta\Phi_t, \quad |b| > 1 \quad (58c)$$

$$\Delta\Phi_t = -\frac{\text{sgn}(b)}{R_H} \log \left| \left( \frac{1+b}{1-b} \right) \frac{R_0 - R_H}{R_H - \text{sgn}(b)R_a} \right|, \quad |b| > 1 \quad (58d)$$

where

$$R_a = \sqrt{\frac{R_0^2(1+b) - 2bR_H^2}{1-b}}. \quad (59)$$

If the quantity  $R_a$  is not real, then there are no geodesics linking  $R_0$  to the horizon. This occurs when  $|b| < 1$  and the associated turning point is smaller than  $R_0$ .

Eqs. (58) display divergences as  $b \rightarrow 1$ , corresponding to geodesics that circle the horizon many times before falling in. However, as  $b \rightarrow 1^+$  (i.e. from above) the turning point moves to the horizon and there are no longer any trajectories linking  $R_0 > R_H$  to the horizon (also seen by  $R_a$  becoming imaginary). These  $b \rightarrow 1^+$  trajectories emerge from the past horizon and orbit arbitrarily many times near the horizon radius before falling in, and are not relevant to the late-time behavior of fields from initial data confined outside the horizon, for which  $R_0$  can be considered fixed.<sup>5</sup> For the relevant trajectories  $b \rightarrow 1^-$  (i.e. from below) that originate from some fixed  $R_0 > R_H$ , we may expand to obtain (dropping the label “nt”)

$$\Delta V = \frac{1}{2R_H(1-b)} - \frac{1}{\sqrt{1-b}} \frac{1}{\sqrt{2(R_0^2 - R_H^2)}} + \frac{1}{4R_H} \log \left| \frac{1-b}{2} \left( \frac{R_0 + R_H}{R_0 - R_H} \right) \right| + \mathcal{O}(1) \quad (60a)$$

$$\Delta\Phi = -\frac{1}{2R_H} \log \left| \frac{1-b}{2} \left( \frac{R_0 + R_H}{R_0 - R_H} \right) \right| + \mathcal{O}(1). \quad (60b)$$

To leading order we thus have

$$\Delta V \sim \frac{R_0 + R_H}{R_0 - R_H} \frac{e^{2R_H\Delta\Phi}}{4R_H}, \quad b \rightarrow 1^-. \quad (61)$$

We now relate this result to the BTZ Green function between points  $(V', R', \Phi')$  and  $(V, R_H, \Phi)$ . Recall that the BTZ Green function consists of a sum over image charges, such that the spikes in the Green function correspond to geodesics with initial values at the image charges. These initial values are given by

$$V_0 = V', \quad R_0 = R', \quad \Phi_0 = \Phi' + 2\pi n, \quad (62)$$

where  $n$  is any integer. These equations also imply  $\Delta V = \delta V$  and  $\Delta\Phi = \Phi - \Phi' - 2\pi n$ . Substituting in, we see that (61) agrees exactly with the arrival times of the  $n \rightarrow -\infty$  BTZ spikes (46), confirming that the  $b \rightarrow 1^-$  geodesics are “responsible” for the Aretakis instability. In Fig. 6 we plot a selection of these trajectories.

Eqs. (58) also display weaker, logarithmic divergences as  $b \rightarrow -1^-$ , corresponding to geodesics that are initially outgoing and reach a turning point at very large radius. As opposed to the  $b \rightarrow 1^-$  that spend a lot of time near the black hole, these  $b \rightarrow -1^-$  geodesics spend a lot of time near the boundary. Expanding Eqs. (58a) and (58b) (and dropping the label “t”), we find

$$\Delta V = -\frac{1}{2R_H} \log \left| (1+b) \frac{R_0 - R_H}{4R_H} \right| + \mathcal{O}(1) \quad (63a)$$

$$\Delta\Phi = \frac{1}{R_H} \log \left| (1+b) \frac{R_0 - R_H}{4R_H} \right| + \mathcal{O}(1) \quad (63b)$$

---

<sup>5</sup> If we scale  $R_0 \rightarrow R_H$  at the same rate as  $b \rightarrow 1^+$  then we can recover the geodesics that turn arbitrarily close to the horizon. These are likely relevant to the case where initial data extends to the horizon.

Thus to leading order we have

$$\Delta V \sim -\frac{1}{2}\Delta\Phi_t, \quad b \rightarrow -1^- \quad (64)$$

Making the substitutions  $\Delta V = \delta V$  and  $\Delta\Phi = \Phi - \Phi' - 2\pi n$  (see Eq. (62)) now with  $n > 0$  to make  $\Delta V$  positive, Eq. (64) agrees precisely with the arrival times (45) of the weak late-time spikes  $n \rightarrow \infty$  in the BTZ Green function. Thus these weaker spikes are associated with the  $b \rightarrow -1^-$  geodesics.

As discussed in detail in Sec. III A, the Green function also contains spikes that are associated with wavefronts that bounce off the boundary. Each such spike is in effect associated with two geodesics (one outgoing and one ingoing), which must be glued together by some rule determined by solving the wave equation near the boundary. Since the Aretakis instability is visible from the above analysis of the ordinary geodesics alone, we do not study this phenomenon further.

To summarize, we have shown that geodesics originating from a fixed point  $R_0 > R_H$  spend arbitrarily long time outside the horizon only near the special limits  $b \rightarrow 1^-$  and  $b \rightarrow -1^-$ . The first limit  $b \rightarrow 1^-$  corresponds to geodesics that orbit many times near the horizon and contribute wide, important spikes to the BTZ Green function at late times. The second limit  $b \rightarrow -1^-$  corresponds to geodesics that spend a long time near the timelike boundary of the spacetime and contribute narrow, unimportant spikes at late times.

## ACKNOWLEDGEMENTS

We gratefully acknowledge Gary Horowitz, Don Marolf, Harvey Reall, and Bob Wald for helpful conversations. We also thank the anonymous referee for useful discussion that improved our understanding of the origin of the on-horizon decay rate. Portions of this work were completed at the Aspen Center for Physics, which is supported by NSF grant PHY-1607611. This work was supported in part by NSF grants PHY-1506027 and PHY-1752809 to the University of Arizona.

## Appendix A: AdS<sub>3</sub> Green function

Here, we review the derivation of the Dirichlet retarded Green function in AdS<sub>3</sub> given in [24], correcting some trivial errors and applying some identities to obtain a nice form. The Green function equation (14) in Poincaré coordinates (4) becomes

$$\left(\partial_z^2 - \frac{1}{z}\partial_z - \partial_t^2 + \partial_x^2 - \frac{\mu^2}{z^2}\right)G = z\delta(t-t')\delta(z-z')\delta(x-x'). \quad (A1)$$

Mode solutions to the homogeneous equation are given by

$$\psi_{k\omega}(z)e^{-i\omega t}e^{ikx}, \quad (A2)$$

where  $\psi_{k\omega}(z)$  are related to Bessel functions [32]. We satisfy the Dirichlet condition by choosing

$$\psi_{k\omega} = zJ_{2\nu}(qz), \quad \omega^2 = q^2 + k^2, \quad \nu = \frac{1}{2}\sqrt{1 + \mu^2}. \quad (A3)$$

We may use these to construct the inhomogeneous solution for  $G$  as follows. Integrating (A1) over small interval around  $t'$  replaces one of the delta functions with a jump condition,

$$\partial_t G|_{t \rightarrow t'+} - \partial_t G|_{t \rightarrow t'-} = -z\delta(x-x')\delta(z-z'). \quad (A4)$$

The retarded Green function vanishes for  $t < t'$  by definition, so (A4) reduces to an initial condition

$$\partial_t G|_{t \rightarrow t'+} = -z\delta(x-x')\delta(z-z'). \quad (A5)$$

We resolve the remaining delta functions using the relevant completeness relations

$$\delta(x - x') = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{ik(x-x')}, \quad \delta(z - z') = z' \int_0^{\infty} dq q J_{2\nu}(qz) J_{2\nu}(qz'), \quad (\text{A6})$$

giving

$$\partial_t G|_{t \rightarrow t'^+} = -\frac{zz'}{2\pi} \int_{-\infty}^{\infty} dk e^{ik(x-x')} \int_0^{\infty} dq q J_{2\nu}(qz) J_{2\nu}(qz'). \quad (\text{A7})$$

The retarded Green function must vanish for  $t < t'$ , satisfy (A7) at  $t = t'$ , and be a homogeneous solution of (A1) for  $t > t'$ . In light of the homogeneous mode solutions (A2)-(A3), we may satisfy the latter two conditions by multiplying the right-hand side of (A7) by  $-i\omega e^{-i\omega(t-t')}$ . Adjoining a step function  $\Theta(t - t')$  fulfills the first condition, giving

$$G = -\Theta(t - t') \frac{2zz'}{\pi} \int_0^{\infty} dk \int_0^{\infty} dq \frac{q}{\sqrt{q^2 + k^2}} \sin(\sqrt{q^2 + k^2}(t - t')) \cos(k(x - x')) J_{2\nu}(qz) J_{2\nu}(qz'), \quad (\text{A8})$$

where we also use that fact that the integrand is odd in  $k$  and even in  $q$ . Since the Green function vanishes as we approach the boundary  $z \rightarrow 0$  (or  $z' \rightarrow 0$ ), this is the Dirichlet retarded Green function as desired. Note that with our choice of Dirichlet boundary conditions,  $G$  is conformal to the retarded propagator on the upper-half plane of Minkowski space with a mirror at infinity [29] with conformal factor  $zz'$ .

We perform the  $k$ -integral using Eq. (3.876-1) of [33], giving

$$G = -\Theta(t - t') \Theta(t - t' - x + x') z z' \int_0^{\infty} dq q J_{2\nu}(qz) J_{2\nu}(qz') J_0(q\sqrt{(t - t')^2 - (x - x')^2}) \quad (\text{A9})$$

The remaining  $q$  integral is resolved with Eq. (6.578-8) of [33], giving the AdS<sub>3</sub> Green function as

$$G_{\text{ret}}^{\text{AdS}_3} = -\Theta(t - t') \Theta(\Sigma) \begin{cases} \frac{2}{\pi \sqrt{2\pi} \sqrt{\sin w}} P_{2\nu-1/2}^{1/2}(\cos w) & \text{if } \Sigma < 2 \\ \frac{-1}{\sqrt{2\pi}} \frac{\sin(2\nu\pi)}{\sqrt{\sinh u}} e^{-\frac{i\pi}{2}} Q_{2\nu-1/2}^{1/2}(\cosh u) & \text{if } \Sigma > 2 \end{cases} \quad (\text{A10})$$

where

$$\cosh u = -\cos w = 1 - \Sigma, \quad \Sigma = \frac{(t - t')^2 - (x - x')^2 - (z - z')^2}{2zz'}. \quad (\text{A11})$$

This is the form given in [24], correcting a couple of typos. We can simplify further using Eqs. (14.5.17), (14.3.10), and (14.5.11) in [34] to obtain

$$G_{\text{ret}}^{\text{AdS}_3} = \frac{\Theta(t - t') \Theta(\Sigma)}{\pi \sqrt{|\Sigma(\Sigma - 2)|}} \begin{cases} -\cos(2\nu \arccos(1 - \Sigma)) & \text{if } \Sigma < 2 \\ \sin(2\nu\pi) e^{-2\nu \text{arccosh}(\Sigma-1)} & \text{if } \Sigma > 2 \end{cases}. \quad (\text{A12})$$

This form explicitly shows the inverse square root singularity at the wavefronts  $\Sigma = 0, 2$ . Another form of the result is

$$G_{\text{ret}}^{\text{AdS}_3} = \frac{\Theta(t - t') \Theta(\Sigma)}{2\pi \sqrt{|\Sigma(\Sigma - 2)|}} \begin{cases} -(Q(\Sigma)^{2\nu} + Q(\Sigma)^{-2\nu}) & \text{if } \Sigma < 2 \\ 2 \sin(2\nu\pi) Q(\Sigma)^{-2\nu} & \text{if } \Sigma > 2 \end{cases}, \quad (\text{A13})$$

where  $Q(\Sigma) = |1 - \Sigma| + \sqrt{\Sigma(\Sigma - 2)}$ . The large- $\Sigma$  behavior is

$$G_{\text{ret}}^{\text{AdS}_3} \sim \sin(2\nu\pi) \frac{4^{-\nu}}{\pi} \Sigma^{-2\nu-1}, \quad \Sigma \rightarrow \infty. \quad (\text{A14})$$

## Appendix B: Massless axisymmetric perturbations: Aretakis' method

Using the technique employed originally by Aretakis for extreme black holes in four dimensions [1, 2], in this appendix we show that massless linear perturbations arising from axisymmetric data with support extending to the horizon conserve a charge on the extremal BTZ horizon. As in the four dimensional case, the associated conservation law prevents derivative decay, and higher-order derivatives grow polynomially. Assuming decay of the perturbation itself, we obtain rates for the derivatives by a hierarchy argument.

To begin we express the massless wave operator in “radially inverted” ingoing cororating coordinates  $(V, \rho, \Phi)$ , where  $\rho = R_H/R$ ,

$$\square\psi = (\rho^2 - 1)f[\psi] + \partial_V(1 - 2\rho\partial_\rho)\psi - \partial_\Phi\left((1 + \rho^2) - \frac{\rho}{R_H}\partial_\Phi\right)\psi. \quad (\text{B1})$$

Here,  $f$  is a 2nd-order differential operator having smooth coefficients in a neighborhood of  $\rho = 1$  (the horizon) whose precise form is irrelevant for the arguments to follow. Integrating the wave equation on a cylindrical section of the horizon  $\rho = 1$  extending from an initial time  $V_1$  to some final time  $V_2$  leads to the conservation law

$$\int_{S^1} \int_{V_1}^{V_2} dV \partial_V(1 - 2\rho\partial_\rho)\psi|_{\rho=1} = 0, \quad (\text{B2})$$

where we have used that  $\int_{S^1} \partial_\Phi\psi = 0$  for sufficiently smooth  $\psi$ . The conservation law is stated locally as

$$\partial_V H|_{\mathcal{H}} = 0, \quad (\text{B3})$$

where

$$H \equiv \int_{S^1} (1 - 2\rho\partial_\rho)\psi|_{\rho=1}. \quad (\text{B4})$$

The conservation law implies that if  $\psi$  decays to zero on the horizon as  $V \rightarrow \infty$ , the radial derivative must asymptote to a constant. Compactly,

$$[\psi_{\mathcal{H}} \rightarrow 0 \quad \text{and} \quad (\text{B3})] \implies \partial_\rho\psi_{\mathcal{H}} \rightarrow -\frac{H}{2}. \quad (\text{B5})$$

Furthermore, if both  $\psi$  and its tangential  $V$  derivatives along the horizon decay asymptotically as  $V \rightarrow \infty$ , then higher-order radial derivatives grow at rates

$$\partial_\rho^n \psi_{\mathcal{H}} \sim c_n(H) V^{n-1}, \quad V \rightarrow \infty, \quad (\text{B6})$$

where  $c_n$  is an order one constant depending on the specific choice initial data through  $H$ .

To demonstrate the blow-up of the first derivative, take the radial derivative of the wave equation  $\partial_\rho \square\psi = 0$ , pull back to the horizon, and average over the circular cross section:

$$0 = \int_{S^1} \left[ 8R_H \partial_\rho \psi + \partial_V(\psi - 4\partial_\rho \psi - 2\partial_\rho^2 \psi) \right]_{\mathcal{H}}. \quad (\text{B7})$$

Using our decay assumptions and results above for the horizon rates  $\psi \rightarrow 0$ ,  $\partial_V \psi \rightarrow 0$ , and  $\partial_\rho \psi \rightarrow -H/2$  as  $V \rightarrow \infty$ , integration of (B7) gives linear growth

$$\partial_\rho^2 \psi_{\mathcal{H}} \sim -4R_H H V, \quad V \rightarrow \infty. \quad (\text{B8})$$

The derivation of the rates for higher derivatives proceeds analogously, and straightforward induction yields (B6).

## Appendix C: Mode approach

### 1. Preliminaries

The starting point for our mode calculations is the introduction of corotating, constant proper volume coordinates given by

$$T = T, \quad y = \frac{2R_H^2}{R^2 - R_H^2}, \quad \eta = X - T. \quad (\text{C1})$$

In these coordinates the separated mode solutions resemble those found for perturbations of  $\text{AdS}_2$  with an electric field at small frequencies, whose properties have been exhaustively analysed in previous studies of the Aretakis instability [5]. Decomposing into modes, the field takes the form of a Fourier-Laplace integral

$$\psi = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{im\eta} \int_{-\infty}^{\infty} d\omega e^{-i\omega T} \psi_{m\omega}(y). \quad (\text{C2})$$

For Dirichlet data, the field modes  $\psi_{m\omega}$  are determined by a radial convolution integral containing mode decomposed initial data and an integral kernel called the “transfer function”. The range of the integral spans the support of the data, which we choose to be bounded away from the event horizon and infinity. Since we are considering generic mode-evolution of this data, we choose to work directly with the transfer function, denoted here as  $g_m(y, y'; \omega)$ . It satisfies the inhomogeneous equation

$$g_m''(y, y'; \omega) + \left( \frac{\omega^2}{16R_H^2} + \frac{\omega(\omega + 2m)}{8R_H^2 y} + \frac{h(1-h)}{y^2} \right) g_m(y, y'; \omega) = \frac{1}{4R_H^2} \delta(y - y'), \quad (\text{C3})$$

where prime denotes ordinary  $y$ -differentiation. Here again,  $h$  is as defined in (34).

The proceeding arguments may be understood directly from the BTZ wave equation and Kirchhoff representation of the mode decomposed retarded propagator

$$G = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{im(\eta - \eta')} \int_{-\infty}^{\infty} d\omega g_m(y, y'; \omega) e^{-i\omega(T - T')}. \quad (\text{C4})$$

To ensure causal evolution, we choose the transfer function corresponding to the retarded solution. This choice dictates our selection of homogeneous radial functions subject to boundary conditions corresponding to ingoing waves at the horizon  $R_{\text{in}}$  and Dirichlet falloff infinity  $R_D \sim y^h$  as  $y \rightarrow 0$ . These solutions exist provided  $\mu^2 > -1$  [27]. Imposing a  $C^0$  match of the two homogeneous solutions at the support of the delta function allows the transfer function to be written as

$$g_m(y, y'; \omega) = \frac{1}{4R_H^2} \frac{R_{\text{in}}(y_{>}) R_D(y_{<})}{\mathbf{W}[R_{\text{in}}(y'), R_D(y')]}, \quad (\text{C5})$$

where  $y_{<} = \min(y, y')$  and  $y_{>} = \max(y, y')$ . Here we have introduced the conserved Wronskian of the two solutions  $\mathbf{W}[R_{\text{in}}, R_D] = R_{\text{in}} R_D' - R_D R_{\text{in}}'$ . We express the homogeneous solutions in terms of Whittaker functions as

$$R_{\text{in}} = W_{iq, \nu} \left( \frac{-i\omega y}{2R_H} \right), \quad R_D = M_{-iq, \nu} \left( \frac{i\omega y}{2R_H} \right), \quad (\text{C6})$$

where we have defined

$$q = \frac{1}{2R_H} \left( m + \frac{\omega}{2} \right). \quad (\text{C7})$$

These solutions are linearly independent provided  $\nu$  is non-integer. This special case will not be considered here. Using Eq. (13.14.27) of Ref. [34] for the Wronskian of these two functions, we find that (C5) can be written as

$$g_m(y, y'; \omega) = \frac{(-1)^h}{2R_H} \frac{\Gamma(h - iq)}{i\omega \Gamma(2h)} W_{iq, \nu} \left( \frac{-i\omega y_{>}}{2R_H} \right) M_{-iq, \nu} \left( \frac{i\omega y_{<}}{2R_H} \right). \quad (\text{C8})$$

The coordinates  $(T, y, \eta)$  do not extend to the horizon of the black hole, and thus fail to characterize Aretakis' instability. For a suitable extension to  $R_H$  we now readopt the “horizon coordinates”  $V$  and  $\Phi$  (see Eq. (9)) and invert  $y$ ,

$$V = T + R_*, \quad Y = \frac{1}{y}, \quad \Phi = \eta + \Phi_\#.$$
 (C9)

Explicitly,  $R_*$  and  $\Phi_\#$  are given in terms of  $Y$  by

$$R_* = -\frac{1}{4R_H} \left( \frac{\sqrt{1+2Y}}{Y} + \log \left( \frac{1+Y+\sqrt{1+2Y}}{Y} \right) \right),$$
 (C10a)

$$\Phi_\# = \frac{1}{2R_H} \log \left( \frac{1+Y+\sqrt{1+2Y}}{Y} \right).$$
 (C10b)

In these coordinates the horizon sits at  $Y = 0$ . On the horizon,  $\partial_V$  is tangent and  $\Phi$  labels generators. While at infinity,  $V = T$  and  $\Phi = \eta$ . From the  $T - T'$  and  $\eta - \eta'$  dependence in the phases  $e^{-i\omega(T-T')}$  and  $e^{im(\eta-\eta')}$  of the mode decomposition (C4), we see that, under the coordinate transformation (C9), the radial Green function  $g_{m\omega}$  is transformed by

$$g_m \rightarrow e^{-im(\Phi_\# - \Phi'_\#) + i\omega(R_* - R'_*)} g_m.$$
 (C11)

Both the primed and unprimed phases may be determined from the expression

$$e^{i\omega R_* - im\Phi_\#} = \left( \frac{Y}{1+Y+\sqrt{1+2Y}} \right)^{iq} \exp \left( -\frac{i\omega}{4R_H} \frac{\sqrt{1+2Y}}{Y} \right)$$
 (C12)

by complex conjugating at the primed point.

## 2. Modes at late times near the horizon

To obtain the late-time behavior of a fixed  $m$  mode, we apply the asymptotic theory of Fourier-Laplace transforms [35] wherein the functional behavior as  $T \rightarrow \infty$  is determined by the leading-order non-analytic term in an asymptotic expansion of the Laplace transform about its uppermost singular point in the complex  $\omega$  plane.

Inspection of (C8) reveals that the transfer function has its uppermost singular point at  $\omega = 0$ . To determine the late-behavior near the horizon, we expand  $g_{m\omega}$  in an asymptotic series about  $\omega = 0$ , keeping  $\omega/Y$  held fixed. This limit motivates the definition of the near-far late-time transfer function

$$g_m^{\text{nf,lt}} := g_m(\omega \rightarrow 0) \quad \text{fixing } \omega/Y,$$
 (C13)

which excludes all terms analytic in  $\omega$  with the exception of those involving  $\omega/Y$ . In the late-time limit the  $M$  function simplifies via Eq. (13.14.14) of Ref. [34], and we find

$$g_m^{\text{nf,lt}} = -\frac{\Gamma(h-iq_0)}{4R_H^2 \Gamma(2h)} \left[ Y^{iq_0} Y'^{-h} e^{2iR_H q_0 \Phi'_\#} \right] (-i\omega/2R_H)^{h-1} W_{iq_0, \nu} \left( -\frac{i\omega}{2R_H Y} \right) e^{-\frac{i\omega}{4R_H Y}}.$$
 (C14)

Here we have defined

$$q_0 = q|_{\omega=0} = \frac{m}{2R_H}.$$
 (C15)

To perform the inverse transform, we use the identity Eq. (9) in Sec. 5.20, of Ref [36]. With this, the late-time result, stated terms of the shifted time coordinate

$$\delta V = (V - V')$$
 (C16)

introduced previously, is

$$g_m^{\text{nf,lt}}(Y, Y'; \delta V) = -\frac{Y'^{-h} e^{2iR_H q_0 \Phi'_\#} \Gamma(h - iq_0)}{4R_H^2 \Gamma(2h) \Gamma(1 - h - iq_0)} \left[ (2R_H \delta V)^{-h - iq_0} (1 + 2R_H Y \delta V)^{-h + iq_0} \right]. \quad (\text{C17})$$

We conclude this section with a few remarks regarding our result for  $g_m^{\text{nf,lt}}$ . On the horizon, the decay of a fixed  $m$ -mode is given by  $\delta V^{-h}$ . Higher transverse derivatives  $\partial_R^n|_{\mathcal{H}}$  exhibit the Aretakis instability, growing at rate  $\delta V^{-h+n}$ . These decay and instability rates are consistent with those determined from the sum over images using the method of images above in the main text. However, at large  $m$ , the magnitude of  $g_m^{\text{nf,lt}}$  grows as  $m^{2h-1}$ , indicating that the mode sum does not converge pointwise (recall that  $h \geq 1/2$ ). The high- $m$  regime of the transfer function probes the roughest part of the initial data's  $\Phi$ -dependence. If the data is sufficiently regular, with differentiability exceeding  $2h - 1$ , then its Fourier coefficients will decay in  $m$  at a faster rate than the  $m$ -growth exhibited by the coefficients in the transfer function, such that formal application of the Kirchhoff integral (15) provides a finite result for the field. However, the field will then be less regular than the initial data, in contradiction with rigorous results of Warnick [37]. Thus our results for each  $m$ -mode cannot be straightforwardly promoted to results about the whole field, indicating some nonuniformity in the late-time and large- $m$  limits. Nevertheless, the rates predicted by the mode expansion do agree with those of the image sum. We hope to explore these issues further in future work.

### 3. Non-periodic limit - AdS<sub>3</sub>

The late-time behavior (C17) of each angular mode displays growth of transverse derivatives, consistent with the full extremal BTZ spacetime possessing the Aretakis instability. However, the angular modes also apply to the AdS<sub>3</sub> spacetime, which has no such instability. From the perspective of the mode approach, the only difference between AdS<sub>3</sub> and BTZ is that the latter admits only a subset of modes compatible with the periodicity of  $\Phi$  (i.e.,  $m$  is continuous for AdS<sub>3</sub> and quantized for BTZ). That is, formally speaking we have

$$G_{\text{BTZ}}^{\text{nf,lt}} = \sum_{m=-\infty}^{\infty} e^{im(\Phi - \Phi')} g_m^{\text{nf,lt}}, \quad (\text{C18})$$

$$G_{\text{AdS}_3}^{\text{nf,lt}} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{im(\Phi - \Phi')} g_m^{\text{nf,lt}} dm. \quad (\text{C19})$$

We have already noted that the mode sum (C18) does not converge due to  $g_m^{\text{nf,lt}}$  behaving as  $m^{2h-1}$  at large  $m$ , and the integral (C19) is likewise divergent. However, the following formal manipulation assigns it a value that agrees with AdS<sub>3</sub> expectations. Using Eq. (C17) and expressing the integral (C19) in terms of  $q_0$  defined in Eq. (C15), we have

$$G_{\text{AdS}_3}^{\text{nf,lt}} = -\frac{[Y'(2R_H \delta V)(1 + 2R_H Y \delta V)]^{-h}}{2R_H \Gamma(2h)} \int_{-\infty}^{\infty} dq_0 \frac{\Gamma(h - iq_0)}{\Gamma(1 - h - iq_0)} e^{iq_0 a}, \quad (\text{C20})$$

where  $a = \left[ 2R_H(\Phi - \Phi' + \Phi'_\#) + \log \left( \frac{1 + 2R_H Y \delta V}{2R_H \delta V} \right) \right]$ .

Assuming the generic case  $2h \notin \mathbb{Z}^+$ , the integrand contains poles at  $q_0 = -i(h + n)$  with  $n \in \mathbb{Z}^+$ . By Cauchy's theorem we may express the integral as a sum over these poles, together with a contribution from a semicircular arc at large radius in the lower-half plane. This arc contribution does *not* vanish, and in fact it is infinite on account of the behavior  $q_0^{2h-1}$  of the integrand at large  $|q_0|$ . Ignoring this infinite contribution provides a natural regulator that leaves a finite result for the integral, namely the residue sum. This sum can be done in closed form using the formula

$$\sum_{n=0}^{\infty} \frac{(-1)^n e^{(n+h)a}}{\Gamma(n+1) \Gamma(1-2h-n)} = \frac{e^{ah} (1 - e^a)^{-2h}}{\Gamma(1-2h)}. \quad (\text{C21})$$

The result for  $G_{\text{AdS}_3}^{\text{nf,lt}}$ , regulated by dropping the arc contribution, is then

$$G_{\text{AdS}_3}^{\text{nf,lt}} = -\frac{Y'^{-h} e^{2R_H h(\Phi - \Phi' + \Phi'_\sharp)}}{2R_H \Gamma(1-2h) \Gamma(2h)} \left[ 1 - e^{2R_H(\Phi - \Phi' + \Phi'_\sharp)} \left( Y + \frac{1}{2R_H \delta V} \right) \right]^{-2h} (2R_H \delta V)^{-2h}. \quad (\text{C22})$$

The dependence on  $Y\delta V$  in the mode result (C17) has now disappeared, and correspondingly Eq. (C22) shows no Aretakis instability (all transverse derivatives decay). In fact, the second term in square brackets is subleading in the limit we consider ( $\delta V \rightarrow \infty$  fixing  $Y\delta V$ ),<sup>6</sup> so that to leading order we have simply

$$G_{\text{AdS}_3}^{\text{nf,lt}} = -\frac{Y'^{-h} e^{2R_H h(\Phi - \Phi' + \Phi'_\sharp)}}{2R_H \Gamma(1-2h) \Gamma(2h)} (2R_H \delta V)^{-2h}. \quad (\text{C23})$$

Although there are many unresolved issues with convergence and regulators, the discussion in this appendix supports the following general picture: Each angular mode in BTZ or  $\text{AdS}_3$  displays the Aretakis instability, but these are only promoted to an instability arising from compactly supported initial data in the BTZ spacetime, where a subset of modes is selected out by the periodic identification. This is consistent with the simple idea that a single angular mode in BTZ has compact spatial support, while a corresponding mode in the  $\text{AdS}_3$  spacetime does not.

- 
- [1] S. Aretakis, (2010), [arXiv:1006.0283 \[math.AP\]](#).
  - [2] S. Aretakis, *Annales Henri Poincaré* **12**, 1491 (2011), [arXiv:1110.2009 \[gr-qc\]](#).
  - [3] S. Aretakis, *Adv. Theor. Math. Phys.* **19**, 507 (2015), [arXiv:1206.6598 \[gr-qc\]](#).
  - [4] J. Lucietti, K. Murata, H. S. Reall, and N. Tanahashi, *JHEP* **03**, 035 (2013), [arXiv:1212.2557 \[gr-qc\]](#).
  - [5] S. E. Gralla and P. Zimmerman, *JHEP* **06**, 061 (2018), [arXiv:1804.04753 \[gr-qc\]](#).
  - [6] S. E. Gralla and P. Zimmerman, *Class. Quant. Grav.* **35**, 095002 (2018), [arXiv:1711.00855 \[gr-qc\]](#).
  - [7] S. E. Gralla, A. Zimmerman, and P. Zimmerman, *Phys. Rev.* **D94**, 084017 (2016), [arXiv:1608.04739 \[gr-qc\]](#).
  - [8] S. Aretakis, *Class. Quant. Grav.* **30**, 095010 (2013), [arXiv:1212.1103 \[gr-qc\]](#).
  - [9] S. Aretakis, (2013), [arXiv:1310.1365 \[gr-qc\]](#).
  - [10] E. W. Leaver, *Journal of Mathematical Physics* **27**, 1238 (1986).
  - [11] M. Casals, S. E. Gralla, and P. Zimmerman, *Phys. Rev.* **D94**, 064003 (2016), [arXiv:1606.08505 \[gr-qc\]](#).
  - [12] M. Casals and P. Zimmerman, (2018), [arXiv:1801.05830 \[gr-qc\]](#).
  - [13] P. Bizon and H. Friedrich, *Class. Quant. Grav.* **30**, 065001 (2013), [arXiv:1212.0729 \[gr-qc\]](#).
  - [14] H. Godazgar, M. Godazgar, and C. N. Pope, (2017), [arXiv:1707.09804 \[hep-th\]](#).
  - [15] S. Bhattacharjee, B. Chakrabarty, D. D. K. Chow, P. Paul, and A. Virmani, *Class. Quant. Grav.* **35**, 205002 (2018), [arXiv:1805.10655 \[gr-qc\]](#).
  - [16] J. Lucietti and H. S. Reall, *Phys. Rev.* **D86**, 104030 (2012), [arXiv:1208.1437 \[gr-qc\]](#).
  - [17] S. Hadar and H. S. Reall, *JHEP* **12**, 062 (2017), [arXiv:1709.09668 \[hep-th\]](#).
  - [18] K. Murata, H. S. Reall, and N. Tanahashi, *Class. Quant. Grav.* **30**, 235007 (2013), [arXiv:1307.6800 \[gr-qc\]](#).
  - [19] L. M. Burko and G. Khanna, *Phys. Rev.* **D97**, 061502 (2018), [arXiv:1709.10155 \[gr-qc\]](#).
  - [20] M. Bañados, M. Henneaux, C. Teitelboim, and J. Zanelli, *Phys. Rev.* **D48**, 1506 (1993), [Erratum: *Phys. Rev.* **D88**, 069902 (2013)], [arXiv:gr-qc/9302012 \[gr-qc\]](#).
  - [21] M. Bañados, C. Teitelboim, and J. Zanelli, *Phys. Rev. Lett.* **69**, 1849 (1992), [arXiv:hep-th/9204099 \[hep-th\]](#).
  - [22] S. E. Gralla, A. Ravishankar, and P. Zimmerman, *JHEP* **12**, 087 (2018), [arXiv:1808.07053 \[hep-th\]](#).
  - [23] S. Hadar, *JHEP* **01**, 214 (2019), [arXiv:1811.01022 \[hep-th\]](#).
  - [24] U. H. Danielsson, E. Keski-Vakkuri, and M. Kruczenski, *JHEP* **01**, 002 (1999), [arXiv:hep-th/9812007 \[hep-th\]](#).
  - [25] A. R. Steif, *Phys. Rev.* **D49**, 585 (1994), [arXiv:gr-qc/9308032 \[gr-qc\]](#).
  - [26] P. Breitenlohner and D. Z. Freedman, *Phys. Lett.* **B115**, 197 (1982).
  - [27] A. Ishibashi and R. M. Wald, *Class. Quant. Grav.* **21**, 2981 (2004), [arXiv:hep-th/0402184 \[hep-th\]](#).
  - [28] E. Poisson, A. Pound, and I. Vega, *Living Reviews in Relativity* **14** (2011), 10.1007/lrr-2011-7.

---

<sup>6</sup> The corresponding term appeared in a phase in Eq. (C20) and could not be dropped at that stage

- [29] C. Dappiaggi and H. R. C. Ferreira, *Phys. Rev.* **D94**, 125016 (2016), [arXiv:1610.01049 \[gr-qc\]](#).
- [30] B. S. Kay, M. J. Radzikowski, and R. M. Wald, *Commun. Math. Phys.* **183**, 533 (1997), [arXiv:gr-qc/9603012 \[gr-qc\]](#).
- [31] S. Carlip, *Class. Quant. Grav.* **12**, 2853 (1995), [arXiv:gr-qc/9506079 \[gr-qc\]](#).
- [32] V. Balasubramanian, P. Kraus, and A. E. Lawrence, *Phys. Rev.* **D59**, 046003 (1999), [arXiv:hep-th/9805171 \[hep-th\]](#).
- [33] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, Series, and Products*, seventh ed., edited by A. Jeffrey and D. Zwillinger (Academic Press, 2007).
- [34] DLMF, “[NIST Digital Library of Mathematical Functions](#),” <http://dlmf.nist.gov/>, Release 1.0.5 of 2012-10-01, online companion to [? ].
- [35] G. Doetsch, *Introduction to the Theory and Application of the Laplace Transformation* (Springer Berlin Heidelberg, Berlin, Heidelberg, 1974).
- [36] B. M. Project, H. Bateman, A. Erdélyi, and U. S. O. of Naval Research, *Tables of Integral Transforms: Based, in Part, on Notes Left by Harry Bateman*, Tables of Integral Transforms: Based, in Part, on Notes Left by Harry Bateman No. v. 2 (McGraw-Hill, 1954).
- [37] C. M. Warnick, *Commun. Math. Phys.* **321**, 85 (2013), [arXiv:1202.3445 \[gr-qc\]](#).