

Adaptive analytical ray tracing of black hole photon rings

Alejandro Cárdenas-Avendaño^{1,2,*}, Alexandru Lupsasca^{1,3}, and Hengrui Zhu¹

¹*Princeton Gravity Initiative, Princeton University, Princeton, New Jersey 08544, USA*

²*Programa de Matemática, Fundación Universitaria Konrad Lorenz, 110231 Bogotá, Colombia*

³*Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37212, USA*



(Received 18 November 2022; accepted 18 January 2023; published 22 February 2023)

Recent interferometric observations by the Event Horizon Telescope have resolved the horizon-scale emission from sources in the vicinity of nearby supermassive black holes. Future space-based interferometers promise to measure the “photon ring”—a narrow, ring-shaped, lensed feature predicted by general relativity, but not yet observed—and thereby open a new window into strong gravity. Here we present AART: an adaptive analytical ray-tracing code that exploits the integrability of light propagation in the Kerr spacetime to rapidly compute high-resolution simulated black hole images, together with the corresponding radio visibility accessible on very long space-ground baselines. The code samples images on a nonuniform adaptive grid that is specially tailored to the lensing behavior of the Kerr geometry and is therefore particularly well-suited to studying photon rings. This numerical approach guarantees that interferometric signatures are correctly computed on long baselines, and the modularity of the code allows for detailed studies of equatorial sources with complex emission profiles and time variability. To demonstrate its capabilities, we use AART to simulate a black hole movie of a stochastic, nonstationary, nonaxisymmetric equatorial source; by time-averaging the visibility amplitude of each snapshot, we are able to extract the projected diameter of the photon ring and recover the shape predicted by general relativity.

DOI: [10.1103/PhysRevD.107.043030](https://doi.org/10.1103/PhysRevD.107.043030)

I. INTRODUCTION

According to general relativity, black holes display unique lensing behavior: for instance, any two spatial points outside the event horizon are connected by infinitely many light rays, each of which executes a different number of orbits around the black hole under its extreme gravitational pull [1–5]. As a result, images of a black hole surrounded by a nonspherical, optically thin emission region decompose into a sequence of superimposed layers indexed by photon half-orbit number n , with each layer consisting of a full lensed image of the main emission [5–11]. The direct ($n = 0$) layer typically displays a central dark area—the black hole—encircled by the weakly lensed, primary image of the accretion flow onto the hole, whose details depend sensitively on astrophysical conditions. On the other hand, the higher- n layers arise from photons on highly bent trajectories that are strongly lensed to form a series of narrow “photon rings.”¹ These rings are usually stacked on top of the broader $n = 0$ emission and their shape rapidly converges (exponentially fast in n) to that of

the “Kerr critical curve”² [15]: a theoretical curve in the observer sky corresponding to the apparent image of asymptotically bound photon orbits. In contrast to the astrophysics-dependent $n = 0$ image, this “ $n \rightarrow \infty$ photon ring” is completely determined by the Kerr black hole—depending only on its mass, spin, and inclination—and delineates its cross-sectional area in the sky.

Recently, 1.3 mm interferometric observations by the Event Horizon Telescope have resolved the horizon-scale emission from sources in the immediate vicinity of two nearby supermassive black holes: M87* [16], the central compact object at the core of our neighboring galaxy Messier 87, and Sgr A* [17], our own black hole at the center of the Milky Way. Their reconstructed images display a central brightness depression within a thick ring consistent with theoretical expectations for the direct image of the surrounding accretion flow [18–22].

However, these observations have thus far not provided any evidence for the presence of a lensed photon ring [23]. They are instead dominated by $n = 0$ photons [6,7], which form the image layer that is more sensitive to the astrophysics of the flow than to purely relativistic effects [24–26]. As a

*Corresponding author.
cardenas-avendano@princeton.edu

¹For an animation of this effect, see <https://youtu.be/4-DvyMPs-gA>.

²Spherical emission creates a “shadow” inside of this curve [11–13], which is often called the “shadow edge” even when its interior is not dark [14].

result, the bounds placed on possible deviations from general relativity are on the order of several percent [27,28], and are comparable to the constraints derived from gravitational-wave observations of stellar-mass, binary black holes [29], or x-ray spectroscopy measurements of low-mass binaries and active galactic nuclei [30]. By contrast, future detections of orbiting ($n \geq 1$) photons will open a new window into strong gravity and enable higher-precision probes of the Kerr geometry, since it is this orbiting light which forms the part of the image—the photon ring—that belongs to the black hole itself, rather than to its plasma.

There are three major obstacles to measuring a photon ring. First, since the photon rings are exponentially narrow (in n) features, resolving them requires interferometric observations on exponentially long baselines [7]. At the current observing frequency of 230 GHz, even Earth-spanning baselines are too short to resolve the first ring, but it should become accessible to next-generation space-based interferometers. In particular, the Single Aperture Large Telescope for Universe Studies (SALTUS) is a bold proposal—currently a contender for NASA’s upcoming Probe mission—to launch within the next decade a spacecraft far enough to access the first two rings of M87*.

Optical depth poses a second hurdle: even though photons could in principle circumnavigate the black hole indefinitely (albeit unstably), in practice, those that traverse its emission region multiple times are eventually reabsorbed by the matter they intersect—an effect that cuts off image layers past some $n > 0$. Nevertheless, since absorptivity decreases with photon energy, the first few rings still ought to be present in images taken at sufficiently high frequencies. Simple models suggest that at 230 GHz, the $n = 1$ ring is always visible while the $n = 2$ ring may be only marginally observable, whereas both rings should be clearly visible at 345 GHz [11]. State-of-the-art simulations of general-relativistic magnetohydrodynamic (GRMHD) flows [31] also confirm these expectations [32]. For this reason, SALTUS is slated to simultaneously observe at both frequencies. M87* makes for a particularly exciting prospective target because a measurement of its $n = 2$ ring diameter could deliver a stringent test of the Kerr hypothesis, which predicts a definite shape for its higher- n rings: photons orbiting just outside the horizon of a black hole can probe its extreme gravity and carry away information about its spacetime geometry, encoded in the observable shape of the rings that these photons produce in their observer’s sky [33–35].

Time variability introduces a third significant complication. While time-averaged GRMHD-simulated movies have shown that the photon rings are persistent, sharp features that come to dominate observations with very-long-baseline interferometry (VLBI) after averaging over sufficiently long timescales [7], it remains to be understood how clearly visible the rings will be to a realistic, near-future, space-VLBI mission like SALTUS, which will be limited in the number of snapshots it can collect. In each snapshot, such an interferometer—with a single space

leg—can only sample the radio visibility on one space-ground baseline, thereby only measuring the projected diameter d_φ of the photon ring at one angle φ in the image.³ To compensate for its sparse baseline coverage, the instrument can observe at regular intervals along its orbit around the Earth, eventually filling in every angle φ in the Fourier domain, with each d_φ thus sampled twice per orbit. The baseline lengths over which the ring signature dominates the signal fix the orbital radius (about lunar distance for the $n = 2$ ring of M87*) and hence the orbital period (~ 1 month), which in turn sets the cadence of these snapshots: $\sim 40M_{\text{M87*}}$, or roughly every two weeks. As it is evidently impractical to maintain coherence over such timescales, the snapshots must be incoherently time-averaged; moreover, only ~ 24 snapshots of d_φ can be sampled per year.

In a single snapshot, the “clean” interferometric signature of the ring—a periodic ringing in the visibility amplitude—is typically “polluted” by noise from both the instrument and from astrophysical fluctuations (plasma flares, emission ropes, or other ring mimickers), which can obscure the signal. The key question is then: *Can the interferometric signature of a photon ring—and hence its projected diameter—be recovered from an incoherent time-average over $N \approx 20$ snapshots of its visibility amplitude on very long space-ground baselines?*

An affirmative answer to this question would open the door to a consistency test of the Kerr hypothesis—a cornerstone of general relativity (GR) in the strong-field regime—via space-VLBI measurements of the photon ring shape. The paper [35] (henceforth GLM) took the first steps toward establishing the viability of such a test for M87*. GLM examined a range of models of stationary, axisymmetric, equatorial disks that reproduce the time-averaged observational appearance of GRMHD-simulated flows, and found that the observable shape of their $n = 2$ ring always follows a specific functional form, independent of the source model. They concluded that this ring shape is a robust prediction of strong-field GR.

In other words, observations of the $n = 2$ photon ring can *in principle* disentangle gravitational and astrophysical effects that are otherwise commingled in the direct image. Moreover, GLM simulated interferometric data of the kind that could be collected by a mission such as SALTUS, and were able to extract this ring shape from the visibility amplitude on space-ground baselines. Their experimental forecast achieved a subpercent level of precision for the resulting test of the Kerr hypothesis, suggesting that M87* holds exceptional promise as a target for such a test *in practice*. This analysis was recently reviewed in depth and extended to an even larger selection of equatorial disk models, supporting these conclusions [10].

³The angle in the image corresponds to the space element’s baseline angle in the Fourier plane [33–35], while the baseline length determines the index n of the subring whose interferometric signature dominates the signal [7,10].

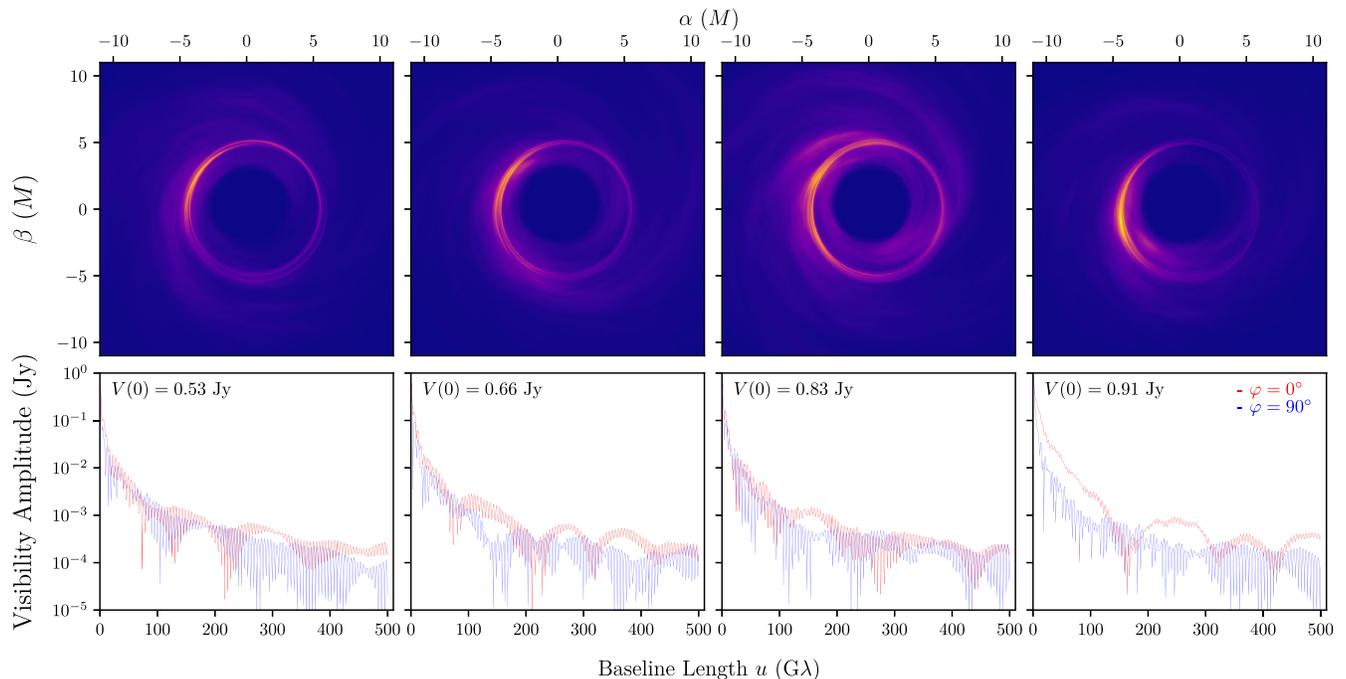


FIG. 1. Top: Snapshots of a nonstationary, nonaxisymmetric, equatorial source around a Kerr black hole ray-traced with AART. Most image features in these snapshots arise from fluctuations in the source, with the exception of the strikingly bright and narrow photon ring: a persistent, sharp feature that comes to dominate the image formed by time-averaging many of these snapshots (Fig. 2). These images were produced using the parameters listed in Table II at regular time intervals of $250M$. Bottom: The corresponding visibility amplitudes for spin-aligned (blue) and spin-perpendicular (red) cuts across each of the above images. The black hole spin is $a/M = 94\%$ and the observer inclination is $\theta_o = 17^\circ$.

While these early results are encouraging, demonstrating the feasibility of a ring shape measurement requires further theoretical work. Crucially, even though the GLM analysis did include realistic instrument noise, it only considered time-averaged images of equatorial disks. The latter limitation was recently tackled with a study of geometric thick-disk models [11], but to date a detailed investigation of source fluctuations and time variability has yet to be carried out.

The present work is the first attempt to remedy this lacuna. The main obstruction is technical: as high-order photon rings are exponentially narrow compared to the overall structure of a black hole image, resolving them in the image plane incurs a large computational cost. More precisely, their presence in the image introduces a large separation of scales between the pixel grid size (which must be large enough to capture the entire field of view) and the pixel spacing (the grid must achieve a sufficiently fine resolution to see the narrow rings). While a brute-force approach—pumping millions of pixels in the grid—can overcome this scale separation for a handful of images, it becomes intractable when dealing with a black hole *movie* consisting of several hundred snapshots, in which case adaptive ray tracing is necessary [36,37] (and less wasteful).

Here, we present a numerical tool⁴ designed to efficiently compute high-resolution “slow-light” movies of generic

(nonstationary and nonaxisymmetric) equatorial sources around a Kerr black hole, together with their associated radio visibility on very long baselines; we present example outputs in Fig. 1.

The code was developed with the intent to maximize speed while still guaranteeing the accuracy of its output, particularly the radio visibility on very long baselines, which encodes the high-frequency components of a snapshot’s Fourier transform and is therefore extremely sensitive to its most minute image features. The code’s structure is highly modular, and most of its individual components reproduce preexisting capabilities; its main novelty is arguably to combine all these routines into one single and convenient-to-use (we hope) package.

One new and important technique from which AART derives much of its power deserves special mention: it turns the very feature of photon rings that makes them so difficult to fully resolve—namely, their thinness—to its advantage. It does so by decomposing the full image into multiple layers labeled by half-orbit number n , and ray tracing in each layer the image of the n th photon ring with exponentially high (in n) resolution.

More precisely, the lensing behavior of a black hole forces the n th photon ring to lie within an exponentially small (in n) region of the image plane, dubbed the n th *lensing band*, with each band completely fixed by the Kerr geometry [10,35]. A key innovation of AART is to first

⁴The code is publicly available at <https://github.com/iAART/aart>.

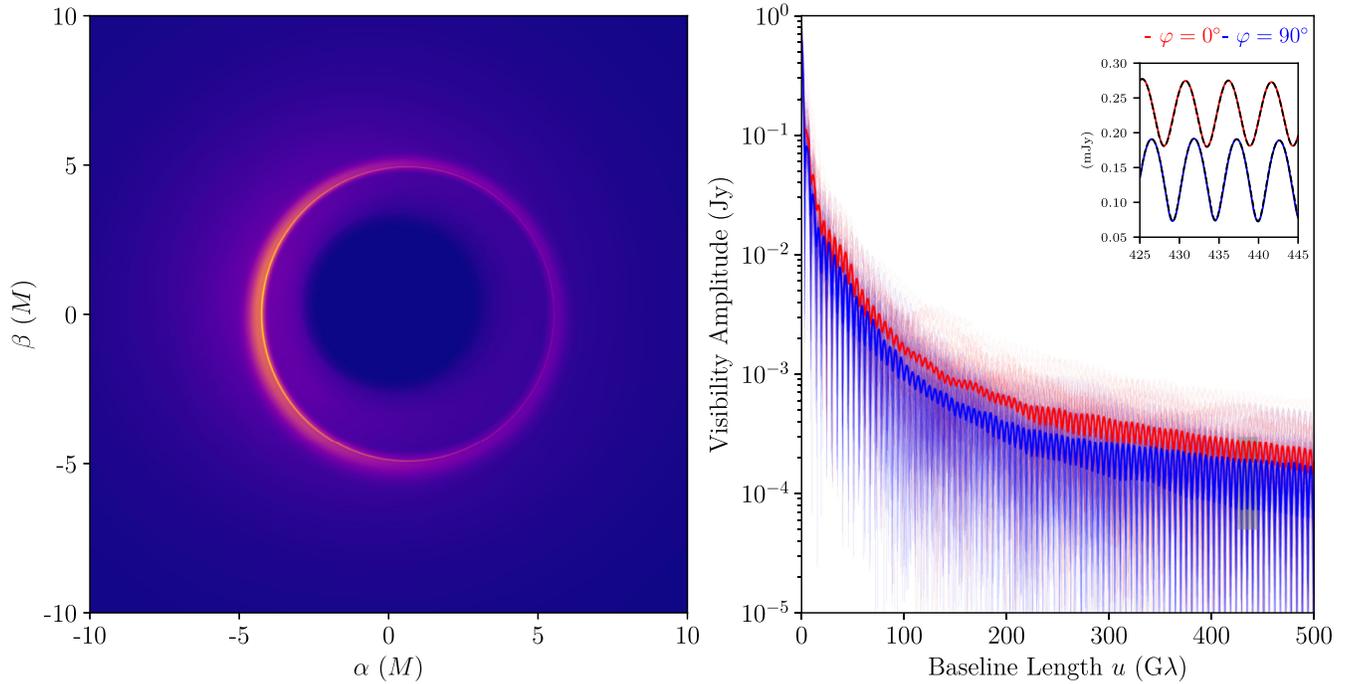


FIG. 2. Left: Time average of 100 ray-traced snapshots of INOISY equatorial profiles uniformly sampled over a time interval of 1000M. The INOISY parameters take the “best-guess” values for M87* presented in Table II. Hence, this image is directly comparable to the time-averaged image of M87* presented in Fig. 1 of Ref. [7], for which the underlying GRMHD simulation’s parameters were chosen to be consistent with the 2017 Event Horizon Telescope observations of M87*. Right: Visibility amplitudes of all 100 snapshots along cuts parallel ($\varphi = 0^\circ$, blue) and perpendicular ($\varphi = 90^\circ$, red) to the black hole spin axis. The solid lines correspond to the incoherently time-averaged amplitudes. Embedded is a panel zooming into the average visibility amplitude in the baseline range $u \in [425, 445] G\lambda$, with the best-fit ring signature overlaid (black).

compute (once and for all) the lensing bands associated with a given black hole spin and inclination, and then to ray trace, for the n th layer, only pixels lying in the n th lensing band. Since each lensing band contains an exponentially demagnified image of the main emission, by also increasing the resolution in each band exponentially,⁵ the code is guaranteed to resolve the source with roughly the same effective resolution in every layer. In other words, AART adapts its ray-tracing resolution (grid spacing) to each layer, but also adjusts the ray-tracing region (grid size), so as to resolve the increasingly fine features present in higher layers using only a fixed number of pixels per layer.

Other ray tracers use adaptive mesh refinement to increase pixel density in regions where they detect long geodesic path lengths [36] or large gradients [37,38]. This results in a new, nonuniform grid for every new snapshot, which is recursively refined until a desired criterion is met, or else the number of recursions exceeds a hard-set limit. Therefore, small image features can sometimes be missed when this limit is hit.

⁵Successive subrings are demagnified by an analytically known, angle-and-spin-dependent factor $e^{-\gamma(\varphi)}$, where $\gamma(\varphi(\tilde{r}))$ is the Lyapunov exponent that governs the instability of nearly bound photons at orbital radius \tilde{r} [5,7].

By contrast, the adaptiveness provided by the lensing bands is determined by the Kerr geometry alone and results in the same, uniform grid for every image of a given black hole spin and inclination. These grids therefore need be computed only once, and provided that their resolution increases at the proper rate set by the demagnification factor, they cannot miss any feature that is already resolved in the direct $n = 0$ image.

Finally, high- n layers comprise photons that execute many orbits in the Kerr photon shell [39] where their radial potential almost develops a double root and geodesic integrals diverge logarithmically, leading to a growing risk of numerical error. To minimize error, AART performs analytical ray tracing using an exact solution of the Kerr null geodesic equation recently given in terms of Legendre elliptic integrals and Jacobi elliptic functions [40], similar in spirit to previous implementations based on Carlson symmetric forms [41,42].

To summarize: AART uses lensing bands to prevent a large separation of scales (between the grid size and its spacing) and thereby maintain its speed as it increases the pixel density in higher- n photon rings, which is necessary to ray trace the fine image features that (due to the nonlocal character of the Fourier transform) influence the visibility on long baselines; moreover, the rings are ray traced analytically to avoid errors.

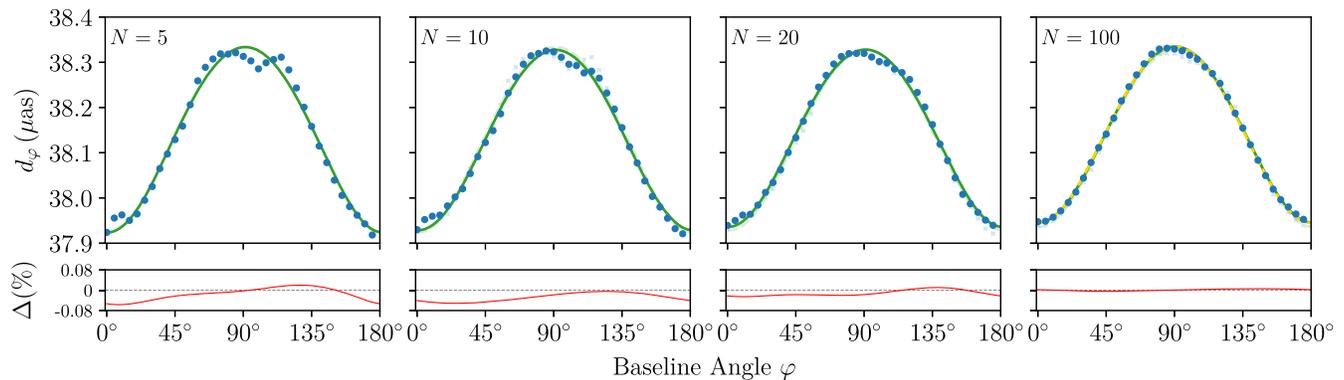


FIG. 3. Top: The projected diameter d_φ of the $n = 2$ photon ring inferred from an incoherent time average over N snapshots of the source. The green curves are best fits of the GR-predicted shape [Eq. (127)] to the synthetic data. To show how the fits improve with increasing N , each panel displays the data of its predecessor in the background. Bottom: The difference between the best-fitting curve at a given N and the “true” shape of the ring (solid gold line in last panel) to which the signal converges as $N \rightarrow \infty$. The parameters of the model are listed in Table II.

These features of AART are specially tailored to the photon ring and its interferometric signature; with this tool in hand, it becomes feasible to investigate the effects of time variability on measurements of the $n = 2$ ring shape, and we can begin to answer the key experimental question posed above.

To study the effects of source fluctuations on the observed visibility amplitude, we call upon both AART and INOISY [43]: a code that can rapidly generate realizations of a 2D Gaussian random field with Matérn covariance. Such a field can provide a simple statistical model for a generic (nonstationary and nonaxisymmetric) stochastic source in the equatorial plane of a Kerr black hole, with a prescribed two-point function. Since a “realistic” choice of autocorrelation structure is not yet known (and will require additional research into the plasma physics of the fluid), our goal will be to vary the statistics of the model within a wide range of “reasonable” possibilities (informed by GRMHD simulations), so as to parametrize our uncertainty in the expected variability of signals from sources such as M87*.

Completing such a parameter survey is a large undertaking beyond the scope of this first paper. Here, we will be content with a proof-of-concept demonstration that AART is up to the task for such a study. To showcase its capabilities, we ray trace 100 snapshots of an INOISY source with statistics set by the “best-guess” parameters for M87* (listed in Table II). Sample snapshots and their associated visibility amplitude are shown in Fig. 1, while Fig. 2 includes all of the snapshots, with the left panel displaying their time-averaged image and the right panel all the individual visibility amplitudes together with their incoherent time average (solid lines). As expected, we find that the astrophysical fluctuations wash out from the time average, leaving an image that is visibly dominated by a prominent photon ring with clear $n = 1$ and $n = 2$ subrings.

Correspondingly, the incoherently time-averaged visibility amplitude is dominated on long baselines by the

perfectly clean interferometric signature of the $n = 2$ ring. In particular, we find that the signal in the range $u \in [425, 445] \text{ G}\lambda$ —which a satellite at lunar distance from the Earth could access with 345 GHz observations—exactly follows the periodic ringing pattern predicted for a thin ring (black overlay in panel inset).

As we show in Fig. 3, the projected diameter d_φ of the $n = 2$ ring can then be extracted from the periodicity of this ringing in the visibility amplitude $|V(u, \varphi)|$. The top panels display the ring diameter d_φ inferred from an average over N snapshots, with $N = 5, 10, 20,$ and finally 100 , by which time the smooth shape of the ring has emerged. The bottom panels display the relative deviation from this fiducial shape that is induced by astrophysical fluctuations, whose noise clearly averages out of the image; it is encouraging to see this noise is also beat down in measurements of the ring shape using only a few snapshots.

Of course, whether these conclusions are likely to hold for M87* has yet to be established, and a systematic investigation of astrophysical fluctuations remains to be done. In a soon-to-be-released paper, we will initiate such a study by repeating this analysis for multiple models of M87*, varying the parameters both of the black hole (its spin and inclination) and of the source (the INOISY model). We also hope to report on the signature of the $n = 1$ ring and how it may encode the spin.

To guide the reader, we now give a summary of the rest of the paper, which—like AART—is written in a modular way.

A. Summary

In Sec. II, we review the problem of light propagation in the Kerr spacetime. We write down the null geodesic equation and its exact analytical solution as it is implemented in AART, and describe the key concept of lensing bands (Fig. 4). We then illustrate the lensing behavior of the black hole by plotting its “transfer functions”: mappings of directions in the observer sky to the spacetime points where

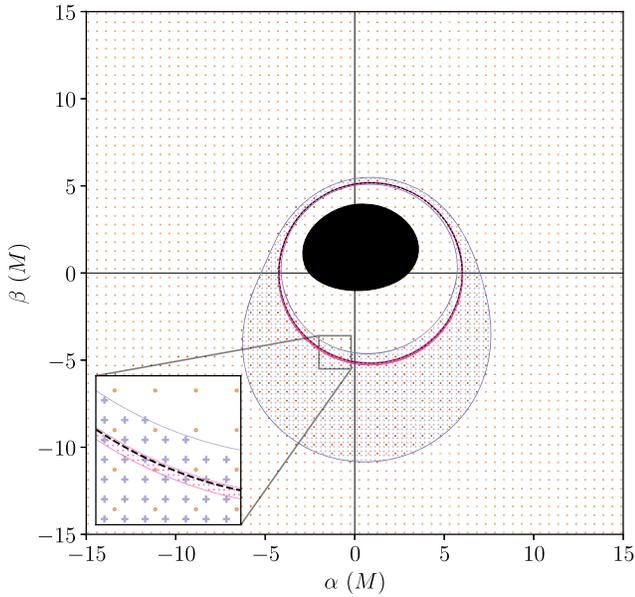


FIG. 4. Illustration of lensing bands with different grid resolutions, for black hole spin $a/M = 50\%$ and observer inclination $\theta_o = 60^\circ$. Orange, purple, and magenta marks depict light rays that intersect the equatorial plane once ($n = 0$ band), twice ($n = 1$ band), and thrice ($n = 2$ band), respectively. The black dashed line is the critical curve (21). Embedded is a panel showcasing the increasing grid resolution in higher- n bands. The region inside the lensed $n = 0$ equatorial horizon, shown in black, is always excluded from our calculations.

the corresponding light rays intersect the equatorial plane. The transfer functions for polar coordinates (r_s, ϕ_s) in the plane are shown in Figs. 5, 6, and 9 for the $n = 0$, $n = 1$, and $n = 2$ images, respectively. Likewise, the time lapse Δt between source and observer is also shown in Figs. 7 and 8 for the $n = 0$ and $n = 1$ images, respectively. The n th image always fills out the n th lensing band, as expected.

We also give an approximate formula for light bending that was derived by Beloborodov [44] for nonrotating black holes in the weak-deflection regime. We express his result in a very simple form [Eq. (43)] that proves to be remarkably accurate for the computation of $n = 0$ images of axisymmetric sources, *for most black hole spins and inclinations*. This observation, which is illustrated in Fig. 10, highlights the fact that $n = 0$ photons barely carry an imprint of the black hole spin, as they do not spend enough time near it to be strongly affected by its gravitational field. Beloborodov’s approximation can thus be regarded as a generalization to all inclinations of the “just add one” prescription for the transfer function of a spin-aligned observer, for whom the impact parameter ρ is simply related to emission radius r_s by adding one: $\rho \approx r_s + M$ [5,45]. AART can also parallel transport linear polarization. We check that Beloborodov’s approximation is adequate for ray tracing $n = 0$ polarimetric images (Fig. 11), which can be done using simple algebraic equations that we write down explicitly.

In Sec. III, we describe our model of equatorial emission. First, we review stationary and axisymmetric equatorial disk profiles that can reproduce the time-averaged observational appearance of M87* in GRMHD simulations [9,35]. Then, we add in astrophysical fluctuations with prescribed statistics. We model these fluctuations using Gaussian random fields and describe their statistical properties in great detail. In Sec. IV, we use INOISY to simulate a variable source. With AART, we ray trace its instantaneous snapshots (Fig. 12) and compute its light curve (Fig. 13). Great care must be taken in the choice of resolution and field of view used in each layer, and Sec. V discusses these issues in depth (Figs. 14–16).

We can then produce movies of a stochastic, nonstationary, nonaxisymmetric source. In Sec. VI, we compute movies of the associated visibility amplitude and use this synthetic data to reconstruct the projected diameter d_ϕ of the $n = 2$ ring, before concluding with a brief discussion of future prospects in Sec. VII. We relegate some details to Apps. A, B, and C. Throughout the paper, we work in $(-, +, +, +)$ metric signature with geometric units in which $G_N = c = 1$. Our conventions for Legendre elliptic integrals are listed in Appendix A of Ref. [46].

II. THEORETICAL FRAMEWORK

The special integrability properties of the Kerr space-time reduce its geodesic equation to a problem of quadratures [47], resulting in elliptic integrals that are expressible in Legendre normal form [40,48–50]. Modern computers can evaluate the Legendre elliptic integrals very fast and to arbitrary precision, reducing the computational cost of ray-tracing in Kerr. In this section, we review the exact solution of the Kerr null geodesic equation in Legendre form [5,40] as well as the approximate solution derived in Schwarzschild by Beloborodov [44], and we use them to plot the transfer functions mapping Bardeen’s coordinates in the observer sky [15] to the equatorial plane.

A. Null geodesics of the Kerr exterior

Astrophysical, rotating black holes of mass M and angular momentum $J = aM$ are subject to the Kerr bound $|a| \leq M$ and are described by the Kerr geometry.

The Kerr metric is written in Boyer-Lindquist coordinates (t, r, θ, ϕ) in terms of functions $\Sigma(r, \theta) = r^2 + a^2 \cos^2 \theta$ and $\Delta(r) = r^2 - 2Mr + a^2$. The roots of $\Delta(r)$ define the radii of the outer and inner event horizons:

$$r_{\pm} = M \pm \sqrt{M^2 - a^2}. \quad (1)$$

With respect to Mino time τ , a photon with four-momentum p^μ follows a null geodesic $x^\mu(\tau)$ obtained by solving

$$\frac{dx^\mu}{d\tau} = -\frac{\Sigma}{p_t} p^\mu. \quad (2)$$

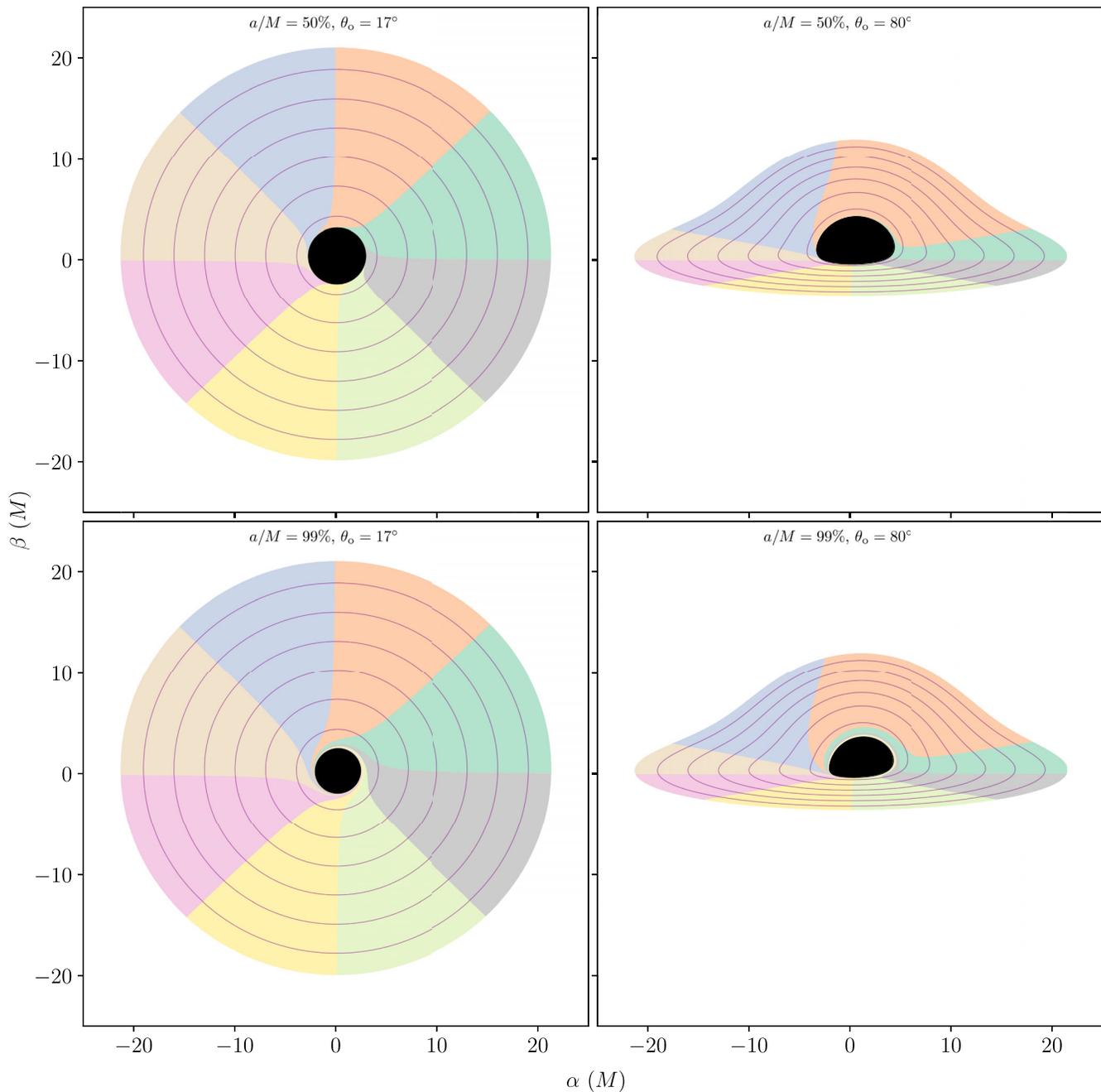


FIG. 5. Direct ($n = 0$) images of the equatorial plane $\theta_s = \pi/2$ for Kerr black holes with spins $a/M \in \{50\%, 99\%\}$ and observer inclinations $\theta_o \in \{17^\circ, 80^\circ\}$. The isoradial curves (purple) correspond to rings of constant Boyer-Lindquist radius $r_s/M \in \{3, 6, 9, 12, 15, 18\}$ between $r_s = r_+$ (the apparent image of the equatorial horizon, in black) and the cutoff at $r_s = 20M$. The colors change every 45° across curves of constant ϕ_s .

The resulting trajectory is independent of the photon energy $-p_t$ and can be parametrized by two conserved quantities: the energy-rescaled angular momentum and Carter constant

$$\lambda = -\frac{p_\phi}{p_t}, \quad \eta = \frac{p_\theta^2}{p_t^2} - a^2 \cos^2 \theta + \lambda^2 \cot^2 \theta. \quad (3)$$

We are interested in solving for the trajectories $x^\mu(\tau)$ that connect two spacetime events $(t_s, r_s, \theta_s, \phi_s)$ and

$(t_o, r_o, \theta_o, \phi_o)$, where the labels “s” and “o” stand for “source” and “observer,” respectively. We will specialize to distant observers ($r_o \gg M$) at nonzero inclination $\theta_o \in (0, \pi/2)$ above the equatorial plane $\theta_s = \pi/2$ where we place the source. (The measure-zero cases $\theta_o \in \{0, \pi/2\}$ technically require separate treatments [5], but in practice one can simply change the inclination slightly and use the same analytical expressions.) The symmetries of the Kerr

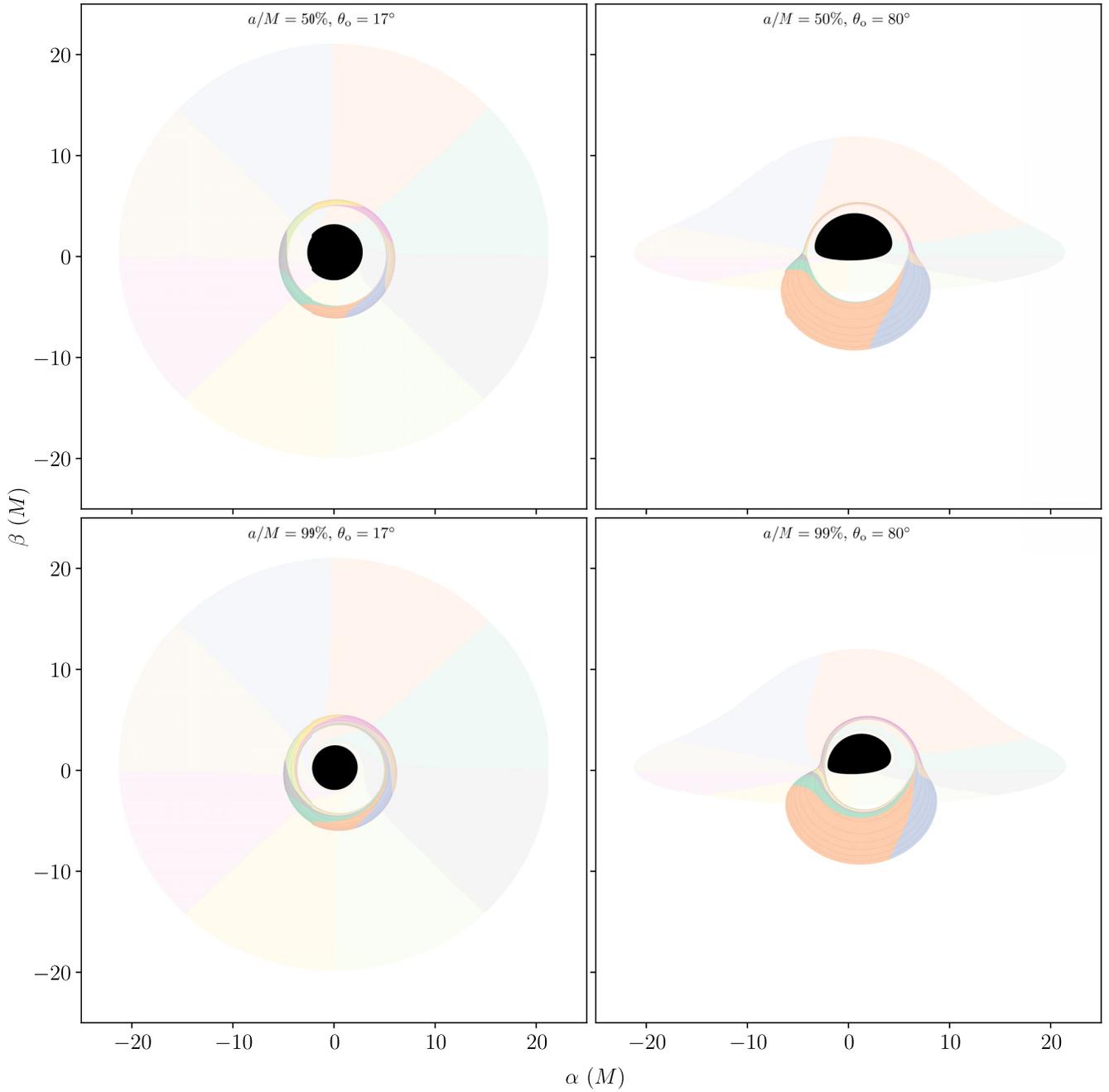


FIG. 6. Same as Fig. 5 for the $n = 1$ image, which is a demagnified and rotated copy of the direct $n = 0$ image (shown as a faded background).

geometry—its stationarity and axisymmetry—allow us to set $t_s = 0$ and $\phi_o = 0$ without loss of generality.

A photon that reaches an observer with four-momentum p_o^μ and conserved quantities (λ, η) appears in the sky at Cartesian position (α, β) given in terms of $\pm_o = \text{sign } p_o^\theta$ by Bardeen [15],

$$\alpha = -\frac{\lambda}{\sin \theta_o}, \quad \beta = \pm_o \sqrt{\eta + a^2 \cos^2 \theta_o - \lambda^2 \cot^2 \theta_o}, \quad (4)$$

for $r_o \rightarrow \infty$. If the photon carries a linear polarization, then its observed electric vector polarization angle (EVPA) is [51–54]

$$\chi = \arctan\left(\frac{\nu\kappa_1 - \beta\kappa_2}{\beta\kappa_1 + \nu\kappa_2}\right), \quad \nu = -(\alpha + a \sin \theta_o), \quad (5)$$

where κ denotes the complex-valued Penrose-Walker constant

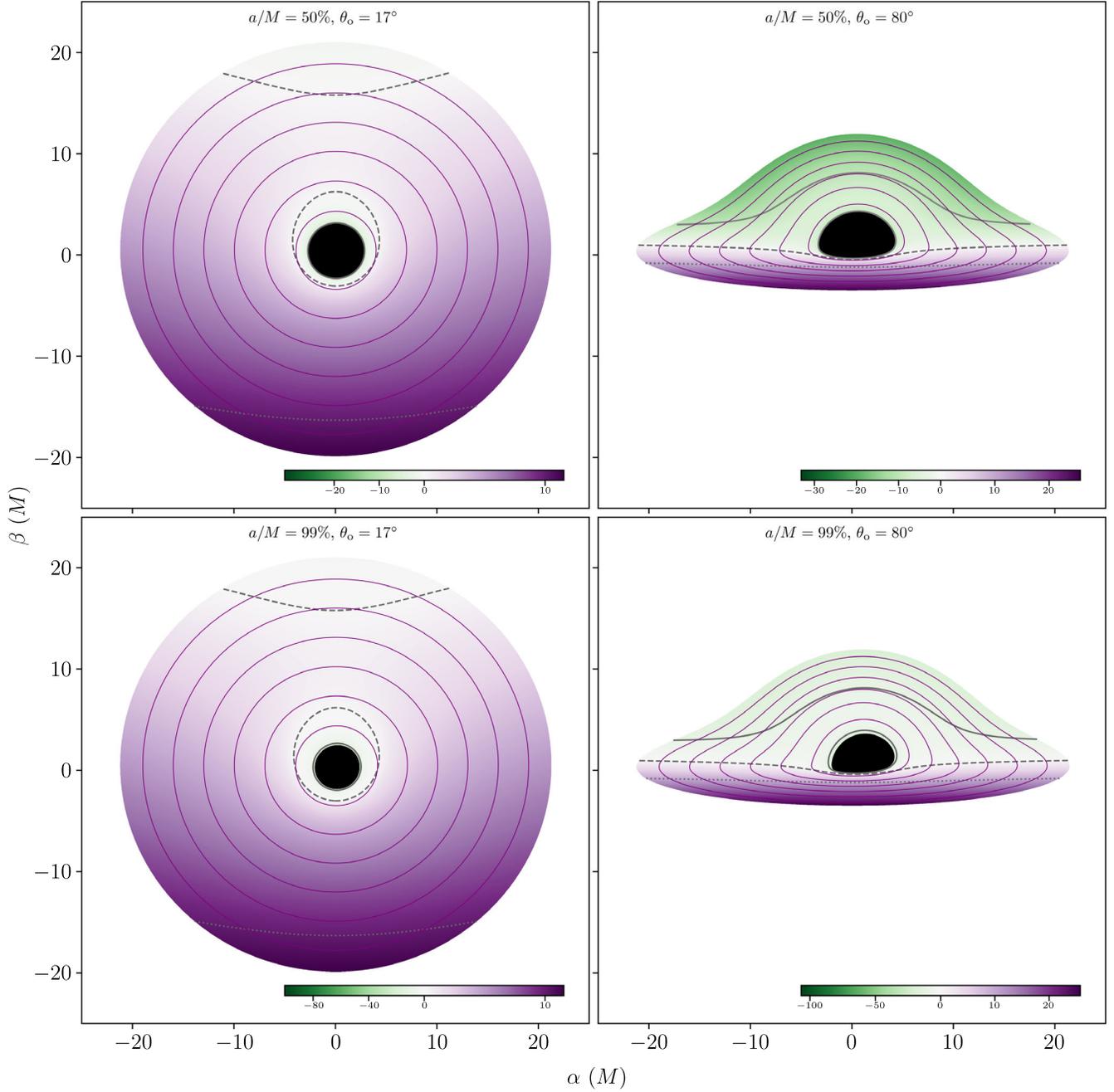


FIG. 7. Same as Fig. 5 for the renormalized time $\tilde{t}^{(0)}$ relative to an observer at $r_o = 10000M$ [Eq. (41)], with the same isoradial curves (purple). The solid, dashed, and dotted gray lines, respectively, denote contours of fixed $\tilde{t}/M \in \{-10, 0, 10\}$. Each panel has its own color scale for clarity.

$$\kappa = \kappa_1 + i\kappa_2 = (\mathcal{P}_A - i\mathcal{P}_B)(r - ia \cos \theta), \quad (6)$$

$$\mathcal{P}_A = (p^t f^r - p^r f^t) + a \sin^2 \theta (p^r f^\phi - p^\phi f^r), \quad (7)$$

$$\mathcal{P}_B = [(r^2 + a^2)(p^\phi f^\theta - p^\theta f^\phi) - a(p^t f^\theta - p^\theta f^t)] \sin \theta, \quad (8)$$

a quantity that is also conserved along null geodesics—thanks to the Petrov type D nature of the Kerr metric [55]—and can be used to algebraically solve the parallel transport

problem for a linear polarization vector f by evaluating κ at the source.

The separability of Kerr geodesic motion allows the r and θ trajectories to be decoupled [47]. These independent motions are then controlled by radial and angular geodesic potentials whose zeros give the turning points of their respective motion:

$$\mathcal{R}(r) = (r^2 + a^2 - a\lambda)^2 - \Delta(r)[\eta + (\lambda - a)^2], \quad (9)$$

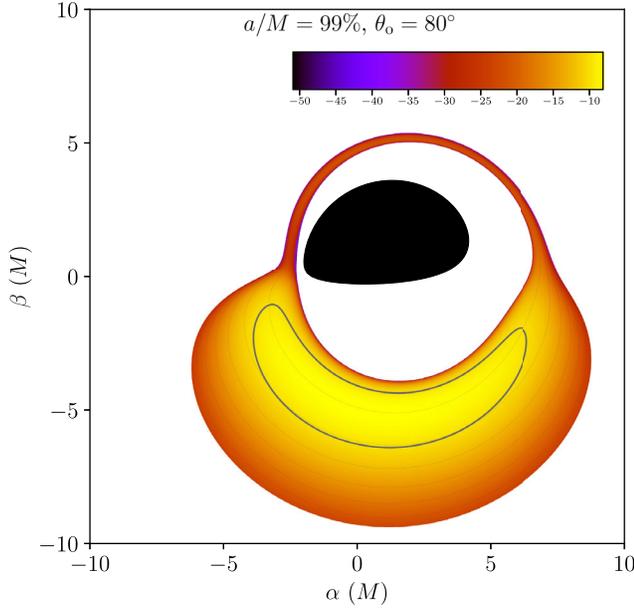


FIG. 8. Same as Fig. 7 for the $n = 1$ image. The solid gray line is an isochronous curve of fixed renormalized time $\tilde{t}^{(1)} = -10M$. The color scale is chosen to highlight the rapid variation in $\tilde{t}^{(1)}$ at the edges of the lensing band (which correspond to the horizon and infinity).

$$\Theta(\theta) = \eta + a^2 \cos^2 \theta - \lambda^2 \cot^2 \theta. \quad (10)$$

Both potentials have exactly four (not always real) roots

$$r_1 = -z - \sqrt{-\frac{\mathcal{A}}{2} - z^2 + \frac{\mathcal{B}}{4z}}, \quad \theta_1 = \arccos(\sqrt{u_+}), \quad (11)$$

$$r_2 = -z + \sqrt{-\frac{\mathcal{A}}{2} - z^2 + \frac{\mathcal{B}}{4z}}, \quad \theta_2 = \arccos(\sqrt{u_-}), \quad (12)$$

$$r_3 = z - \sqrt{-\frac{\mathcal{A}}{2} - z^2 - \frac{\mathcal{B}}{4z}}, \quad \theta_3 = \arccos(-\sqrt{u_-}), \quad (13)$$

$$r_4 = z + \sqrt{-\frac{\mathcal{A}}{2} - z^2 - \frac{\mathcal{B}}{4z}}, \quad \theta_4 = \arccos(-\sqrt{u_+}), \quad (14)$$

which depend only on the conserved quantities (λ, η) via⁶

$$z = \sqrt{\frac{\omega_+ + \omega_- - \mathcal{A}/3}{2}}, \quad \omega_{\pm} = \sqrt[3]{-\frac{\mathcal{Q}}{2} \pm \sqrt{\frac{\mathcal{P}^3}{27} + \frac{\mathcal{Q}^2}{4}}}, \quad (15)$$

⁶Here, $\sqrt[3]{x}$ denotes the real cube root of x if x is real, or else, the principal value of the function $x^{1/3}$ (that is, the cubic root with maximal real part).

$$\mathcal{A} = a^2 - \eta - \lambda^2, \quad \frac{\mathcal{B}}{2M} = \eta + (\lambda - a)^2, \quad \mathcal{C} = -a^2 \eta, \quad (16)$$

$$\mathcal{P} = -\frac{\mathcal{A}^2}{12} - \mathcal{C}, \quad \mathcal{Q} = -\frac{\mathcal{A}}{3} \left[\left(\frac{\mathcal{A}}{6} \right)^2 - \mathcal{C} \right] - \frac{\mathcal{B}^2}{8}, \quad (17)$$

$$u_{\pm} = \Delta_{\theta} \pm \sqrt{\Delta_{\theta}^2 + \frac{\eta}{a^2}}, \quad \Delta_{\theta} = \frac{1}{2} \left(1 - \frac{\eta + \lambda^2}{a^2} \right). \quad (18)$$

We will now restrict our attention to positive spins $0 \leq a \leq M$.

B. Kerr critical curve

The radial potential (9) develops a double root at $\tilde{r} \geq r_+$ [that is, $\mathcal{R}(\tilde{r}) = \mathcal{R}'(\tilde{r}) = 0$] if and only if [5,15]

$$\tilde{\lambda} = a + \frac{\tilde{r}}{a} \left[\tilde{r} - \frac{2\Delta(\tilde{r})}{\tilde{r} - M} \right], \quad \tilde{\eta} = \frac{\tilde{r}^3}{a^2} \left[\frac{4M\Delta(\tilde{r})}{(\tilde{r} - M)^2} - \tilde{r} \right], \quad (19)$$

where $\tilde{r} \in [\tilde{r}_-, \tilde{r}_+]$ with

$$\tilde{r}_{\pm} = 2M \left[1 + \cos \left(\frac{2}{3} \arccos \left(\pm \frac{a}{M} \right) \right) \right]. \quad (20)$$

A geodesic with critical conserved quantities $\lambda = \tilde{\lambda}$ and $\eta = \tilde{\eta}$ asymptotes to an unstably bound orbit at radius \tilde{r} in the Kerr photon shell. For a distant observer, the Kerr critical curve \mathcal{C} is the image in the sky of these asymptotically bound orbits:

$$\mathcal{C} = \{(\tilde{\alpha}, \tilde{\beta}) : (\lambda, \eta) = (\tilde{\lambda}, \tilde{\eta})\}. \quad (21)$$

Though this purely theoretical curve is not in itself observable, it does play a key role in the study of lensing by a Kerr black hole [56]. It is traced using Eq. (4) evaluated on Eq. (19) for all the values $\tilde{r} \in [\tilde{r}_-, \tilde{r}_+]$ such that $\tilde{\beta}^2 \geq 0$. It is always closed, convex, and reflection-symmetric about the α axis [34].

C. Analytical backwards ray tracing

The character (real or complex) and ordering of the radial roots $\{r_1, r_2, r_3, r_4\}$ (when they are real) lead to a classification of radial motion into four types [40,57,58]. We are, however, only interested in the rays that connect a distant observer to an equatorial source. This excludes two of the motion types, leaving only two others corresponding to rays that start from infinity at one end point before they either cross the horizon or return to infinity at the other.⁷

⁷The angular motion also has two possible behaviors according to whether $\eta \geq 0$. We ignore vortical rays with $\eta < 0$ as they cannot reach the equator [46]. Such rays always lie within the apparent image of the horizon.

The boundary between these two behaviors in the phase space of null geodesics constitutes the Kerr photon shell of asymptotically bound orbits, a special locus in phase space with emergent conformal symmetry [59]. We now describe how this separation is manifested in the sky.

Given a black hole spin and observer inclination (a, θ_o) , we can pick a direction (α, β) in the observer sky and shoot a light ray backwards into the geometry in that direction. Such a ray will cross the equatorial plane a total number of times $N(\alpha, \beta)$, which can be determined by computing the total Mino time τ elapsed along the entire development of the trajectory. Rays in the interior of the critical curve all fall into the black hole; that is, they encounter no radial turning point and terminate their motion across the Kerr exterior on the event horizon at $r = r_+$. Rays in the exterior of the critical curve are all deflected back to infinity; that is, their radial motion encounters a turning point at $r = r_4$, whereupon they bounce back toward $r \rightarrow \infty$. Thus, the critical curve (21) delineates the boundary between photon capture (its interior) and photon escape (its exterior), and may be regarded as the cross-sectional area of the hole. In-between rays that lie exactly on \mathcal{C} are trapped in the photon shell where they can in principle orbit forever; in practice, this never occurs because such orbits are unstable (or equivalently, because \mathcal{C} is infinitely thin).

Let $r_{ij} = r_i - r_j$ and define $A = \sqrt{r_{32}r_{42}}$, $B = \sqrt{r_{31}r_{41}}$, and

$$\mathcal{I}_0^{(2)} = \frac{2}{\sqrt{r_{31}r_{42}}} F\left(\arcsin \sqrt{\frac{r - r_4 r_{31}}{r - r_3 r_{41}} \left| \frac{r_{32}r_{41}}{r_{31}r_{42}} \right.}\right), \quad (22)$$

$$\mathcal{I}_0^{(3)} = \frac{1}{\sqrt{AB}} F\left(\arccos\left(\frac{1 - \frac{r-r_2}{r-r_1} \frac{B}{A}}{1 + \frac{r-r_2}{r-r_1} \frac{B}{A}}\right) \left| \frac{(A+B)^2 - r_{21}^2}{4AB} \right. \right), \quad (23)$$

where $F(\varphi|k)$ is an incomplete elliptic integral of the first kind. These functions are the antiderivatives (A10) and (A22), and all the radial geodesic integrals I_k are definite integrals that can be obtained from their respective anti-derivative \mathcal{I}_k as follows.

If a ray lies inside of \mathcal{C} , then its motion is, following the labeling introduced in Ref. [40], of type (2) when all roots are real (in which case $r_+ > r_- > r_4 > r_3 > r_2 > r_1$), or else of type (3) when $r_3 = \bar{r}_4$ are complex-conjugate roots with $r_+ > r_- > r_2 > r_1$.⁸ In either case, the definite integrals down to radius r_s on the ray are

$$I_k(r_s) = \mathcal{I}_k^{(2,3)}(r_o) - \mathcal{I}_k^{(2,3)}(r_s). \quad (24)$$

If a ray lies outside of \mathcal{C} , then its motion is always of type (2), and the definite integrals down to radius r_s on the ray are

$$I_k(r_s) = \mathcal{I}_k^{(2)}(r_o) \mp \mathcal{I}_k^{(2)}(r_s), \quad (25)$$

with sign $-/+$ before/after reaching the turning point at $r = r_4$.

The Mino time $\tau(r_s)$ elapsed along a ray inside \mathcal{C} is thus

$$\tau^-(r_s) = \int_{r_s}^{r_o} \frac{dr}{\sqrt{\mathcal{R}(r)}} \equiv I_0(r_s) = \mathcal{I}_0^{(2,3)}(r_o) - \mathcal{I}_0^{(2,3)}(r_s). \quad (26)$$

Hence, the total Mino time elapsed along the full light ray is

$$\tau_{\max}^- = \int_{r_+}^{r_o} \frac{dr}{\sqrt{\mathcal{R}(r)}} = \tau^-(r_+). \quad (27)$$

Likewise, the Mino time $\tau(r_s)$ elapsed along a ray outside \mathcal{C} is

$$\tau^+(r_s) = \left[\int_{r_s}^{r_o} + 2w \int_{r_4}^{r_s} \right] \frac{dr}{\sqrt{\mathcal{R}(r)}} = \mathcal{I}_0^{(2)}(r_o) \mp \mathcal{I}_0^{(2)}(r_s), \quad (28)$$

where the sign is $-$ before the turning point at $r = r_4$ ($w = 0$) and $+$ after the bounce ($w = 1$). Hence, the total Mino time elapsed along the full light ray is

$$\tau_{\max}^+ = 2 \int_{r_+}^{r_o} \frac{dr}{\sqrt{\mathcal{R}(r)}} = 2\mathcal{I}_0^{(2)}(r_o), \quad (29)$$

and its radial turn occurs at the ‘‘half-way’’ Mino time

$$\tau_4 = \int_{r_+}^{r_o} \frac{dr}{\sqrt{\mathcal{R}(r)}} = \mathcal{I}_0^{(2)}(r_o) = \frac{\tau_{\max}^+}{2}. \quad (30)$$

We will use τ_{\max} to denote the appropriate choice of τ_{\max}^\pm .

At last, the total number of equatorial crossings $N(\alpha, \beta)$ is

$$N = \left\lfloor \frac{\tau_{\max} \sqrt{-u_- a^2} + \text{sign}(\beta) F_o}{2K} \right\rfloor - H(\beta) + 1, \quad (31)$$

where $H(x)$ denotes the Heaviside function while K and F_o are defined in Eqs. (A4) and (5); see also Appendix A of Ref. [10].

A light ray crosses the equatorial plane for the $(n+1)$ th time at Mino time $\tau = \tau_s^{(n)}$, where

$$\tau_s^{(n)} = G_\theta^{(n)} \in [0, \tau_{\max}], \quad n \in \{0, \dots, N-1\}, \quad (32)$$

with $G_\theta^{(n)}$ given in Eq. (A1). This equatorial crossing occurs at

⁸There exist type (4) rays with both $r_1 = \bar{r}_2$ and $r_3 = \bar{r}_4$, but they are vortical.

$$r_s^{(n)} = r_s^{(2,3)}(\tau_s^{(n)}), \quad (33)$$

$$\phi_s^{(n)} = \phi_o - I_\phi^{(n)} - \lambda G_\phi^{(n)}, \quad (34)$$

$$t_s^{(n)} = t_o - I_t^{(n)} - a^2 G_t^{(n)}, \quad (35)$$

with $r_s^{(2,3)}(\tau)$ given in Eqs. (A20) and (A36) for geodesics of types (2) and (3), respectively, while the angular integrals $G_{t,\phi}^{(n)}$ are given in Eqs. (A2) and (A3). Meanwhile, the radial integrals

$$I_{t,\phi}^{(n)} = I_{t,\phi}(r_s^{(n)}) \quad (36)$$

decompose via Eqs. (A8) and (A9) into definite integrals that are to be evaluated via Eq. (24) for rays inside \mathcal{C} or via Eq. (25) for rays outside \mathcal{C} , using the antiderivatives $\mathcal{I}_k^{(2,3)}$ given by Eqs. (A10)–(A13) for type (2) geodesics or Eqs. (A22)–(A25) for type (3) geodesics. In practice, we take $r_o \gg M$ to be very large but not infinite, for reasons described in Eq. (41) below.

D. Kerr lensing bands

The functions $r_s^{(n)}$, $\phi_s^{(n)}$, and $t_s^{(n)}$ defined in Eqs. (33)–(35) are the “transfer functions” that map the equatorial plane to its $(n+1)$ th lensed image in the observer sky, where the index $n \geq 0$ may be thought of as a photon half-orbit number [5,7].

To describe the lensing behavior of a Kerr black hole, it is helpful to draw contour plots of these function—that is, level sets of fixed $r_s^{(n)}$, $\phi_s^{(n)}$, and $t_s^{(n)}$ within each image layer n . First, however, one must determine the regions of the image plane in which these functions have support.

This is a nontrivial problem because these functions always evaluate to some value: for instance, $r_s^{(n)}(\alpha, \beta)$ always returns some source radius, even when the ray shot back from (α, β) does not in fact intersect the equatorial plane $n+1$ times. In such cases, $r_s^{(n)}(\alpha, \beta)$ may sometimes be obviously unphysical (it could, for example, take a negative value), but not always.

We define the n th lensing band as the image-plane subregion consisting of those rays that cross the equatorial plane at least $n+1$ times after being shot back from the observer, before either terminating on the horizon (if they are shot back from inside the critical curve \mathcal{C}) or else returning to infinity (if they are shot back from outside \mathcal{C}) after an elapsed Mino time τ_{\max} .

By definition, the n th lensing band is the physical domain of the transfer functions $r_s^{(n)}$, $\phi_s^{(n)}$, and $t_s^{(n)}$. This definition also implies that it is the subregion of the image plane in which the function $N(\alpha, \beta)$ defined in Eq. (31) is precisely equal to $n+1$.

This last observation leads us to a formula that characterizes the boundary of the n th lensing band: it is the set of points (α, β) in the image plane for which

$$\frac{\tau_{\max} \sqrt{-u_- a^2 + \text{sign}(\beta) F_o} - H(\beta)}{2K} = n, \quad (37)$$

such that $N(\alpha, \beta)$ jumps by one across the boundary of each lensing band. This condition can equivalently be written as

$$\tau_{\max} = G_\theta^{(n)}, \quad (38)$$

with $G_\theta^{(n)}$ given in Eq. (A1). The $n \geq 1$ lensing bands have a different character than the $n=0$ layer—we now focus on the former and defer a discussion of the latter to the next section.

Each $n \geq 1$ lensing band is annular in shape and foliated by contours of fixed $r_s^{(n)}$, with every source radius mapping onto a unique closed curve within the band. Moreover, this bijective map is order-preserving: moving radially outwards within a band, one crosses contours of monotonically increasing $r_s^{(n)}$. Thus, the inner edge of the n th lensing band is the $(n+1)$ th image of the equatorial event horizon (the contour of fixed $r_s^{(n)} = r_+$), while its outer edge is the $(n+1)$ th image of the equatorial circle at infinity (the contour of fixed $r_s^{(n)} = \infty$).

Rays that connect a distant observer to the event horizon cannot encounter a radial turning point along their trajectory, so the inner edge must always appear inside the critical curve. Conversely, rays that connect a distant observer to infinity have to make a turn along their radial trajectory, so the outer edge must always appear outside the critical curve. Hence, the $n \geq 1$ lensing bands are annuli that always straddle the critical curve. Since successive images of a source are demagnified and appear exponentially closer to the critical curve [5,7], the lensing bands form a stack of nested annuli, as seen in Fig. 4.

To summarize, the inner edge of the n th lensing band is the solution of Eq. (38) with $\tau_{\max} = \tau_{\max}^-$ given in Eq. (27), while its outer edge is the solution with $\tau_{\max} = \tau_{\max}^+$ given in Eq. (29). The two sides of Eq. (38) are continuous functions $G_\theta^{(n)}(\alpha, \beta)$ and $\tau_{\max}^\pm(\alpha, \beta)$ of the image-plane position (α, β) .

In practice, we use the following procedure to determine these edges. We first select a set of polar angles $\varphi_i \in [0, 2\pi]$ around the image plane, with associated positions $(\tilde{\alpha}_i, \tilde{\beta}_i)$ on the critical curve (21)—that is, $(\tilde{\alpha}_i, \tilde{\beta}_i) \in \mathcal{C}$ is the point where the critical curve is intersected by the ray emanating from the origin at polar angle φ_i . Next, we parametrize these rays as $(\alpha_i(\epsilon), \beta_i(\epsilon)) = \epsilon(\tilde{\alpha}_i, \tilde{\beta}_i)$. For each φ_i , we compute $G_\theta^{(n)}(\epsilon)$ along the corresponding ray over the range $\epsilon \in (0.5, 3)$. We also compute $\tau_{\max}^-(\epsilon)$ on the range $\epsilon \in (0.5, 1)$ and $\tau_{\max}^+(\epsilon)$ on the range $\epsilon \in (1, 3)$. We then

determine the unique parameters ϵ^\pm within each of these ranges such that $G_\theta^{(n)}(\epsilon^\pm) = \tau_{\max}^\pm(\epsilon^\pm)$. These values correspond to the points $(\alpha_i^\pm, \beta_i^\pm) = \epsilon^\pm(\tilde{\alpha}_i, \tilde{\beta}_i)$ along the ray of constant φ_i where it intersects the lensing band's outer and inner edges, respectively. Connecting the points $(\alpha_i^\pm, \beta_i^\pm)$ thus obtained at every angle φ_i traces out the edges of the lensing band, which is finally obtained as the annular region between these two closed curves. These boundaries—and hence, the lensing band—can be determined to arbitrarily high precision by sampling sufficiently many φ_i .

To check the computation, one can verify that for each i , $r_s^{(n)}(\alpha_i^-, \beta_i^-) \approx r_+$ and $r_s^{(n)}(\alpha_i^+, \beta_i^+) \approx \infty$ up to numerical error.⁹ Further discussion of lensing bands may be found in Ref. [10].

E. Apparent image of the horizon

We now return to the $n = 0$ lensing band, which (unlike its higher- n counterparts) is noncompact. This difference can be understood intuitively: if the black hole mass shrinks to zero, the $n \geq 1$ lensing bands must disappear with the hole, whereas the $n = 0$ band must extend to fill the entire image plane.

Mathematically, the equations in the previous section still hold. The inner and outer edges of the $n = 0$ band are still the contours of fixed $r_s^{(0)} = r_+$ and fixed $r_s^{(0)} = \infty$, respectively. These curves, respectively, correspond to the primary image of the equatorial event horizon (appearing inside \mathcal{C}) and of the circle of infinite radius (appearing outside \mathcal{C}). The inner edge can be determined via the same root-finding procedure outlined above (with $n = 0$), resulting in a closed curve \mathcal{H} contained within the critical curve \mathcal{C} . On the other hand, the outer edge is at infinity in the image plane (as it would be in flat space), so the $n = 0$ lensing band is no longer an annulus. Instead, it consists of the entire (unbounded) exterior of \mathcal{H} .

Meanwhile, the interior of \mathcal{H} may technically be viewed as the $n = -1$ band, since it consists of rays that never reach the equator before terminating on the horizon. In an equatorial disk model, this region would be completely dark, giving rise to an “inner shadow” feature [9], and may naturally be viewed as the apparent image of the horizon. In Fig. 4, we display this image as a shaded black region for the case of a black hole with spin $a/M = 50\%$ observed from an inclination $\theta_o = 60^\circ$.

⁹In principle, one could also determine the inner and outer boundaries of the n th lensing band as those contours of fixed $r_s^{(n)}(\alpha, \beta) = r_-$ (inside \mathcal{C}) or fixed $r_s^{(n)}(\alpha, \beta) = \infty$ (outside \mathcal{C}), respectively, which are closest to \mathcal{C} . However, this method is impractical because $r_s^{(n)}(\alpha, \beta)$ can vary significantly across these contours. Yet, it is this very behavior that makes this check effective.

F. Analytic grid adaptiveness

It has long been known that images of the equatorial plane of a Kerr black hole are lensed into increasingly demagnified, compact regions of the image plane [2,60]. More recently, this lensing behavior has been exploited to efficiently ray trace equatorial disk models in which the $n = 0, 1$, and 2 image components are all fully resolved [35].

AART is designed with this goal in mind and is built around the lensing behavior of a Kerr black hole. It ray traces images of an equatorial source layer-by-layer, using a nonuniform grid adapted to this layered structure. The n th image layer, which consists of the $(n + 1)$ th image of the source, is only ray traced on pixels within the n th lensing band, which is precisely the region of the image plane occupied by this image. As a result, AART avoids computing unneeded pixels in each layer, substantially improving its efficiency.

This layer-by-layer approach naturally allows for different resolutions to be used in each image layer. This is important because the exponential-in- n demagnification of successive images of a source requires the use of exponentially higher resolutions to resolve the source at the same level of detail in higher- n layers. AART increases the resolution in successive lensing bands by roughly the same demagnification factor e^γ that shrinks them (γ is a Lyapunov exponent governing the orbital instability of bound photon orbits [7]) such that every image of the source is resolved by roughly the same number of pixels—see Sec. V for a more detailed discussion of the resolution requirements in each layer. This approach enables ray tracing up to arbitrarily high n and ensures that all images of the source are equally well resolved, with a roughly fixed computational cost per layer.

To summarize, AART makes use of a nonuniform adaptive grid composed of multiple layers, each of which it ray traces at a resolution adapted to the lensing band that makes up the layer. Since the lensing bands are completely determined by the Kerr geometry via the analytic formula (38), this form of adaptiveness is analytic, with the ray tracing also carried out analytically using the exact transfer functions (33)–(35). This adaptiveness is purely geometrical and independent of image features or gradients (unlike, for example, the method in Refs. [36,37]).

In practice, to produce the $n = 0$ grid points, AART creates a Cartesian grid covering the desired area of the image plane, but excluding points in the apparent image of the horizon (that is, the interior of \mathcal{H}). In each $n \geq 1$ layer, the code produces a Cartesian grid of points in an area centered around the critical curve and then discards those points which do not lie within the n th lensing band. Numerically, the code only keeps those points which lie within the concave hull of the points (α_i^+, β_i^+) and outside the concave hull of the points (α_i^-, β_i^-) .

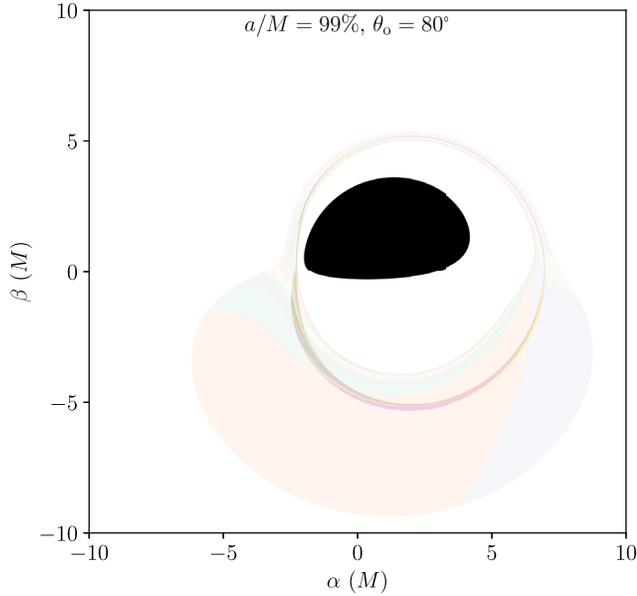


FIG. 9. Same as Figs. 5 and 6 for the $n = 2$ image, with the $n = 1$ image shown as a faded background. Despite a significant zoom, the extremely narrow $n = 2$ lensing band is still barely visible.

We caution the reader that the $n \geq 1$ bands are not always convex: at high inclinations and moderate to high spins, their edges may become nonconvex curves, as may be seen for instance in the right column of Fig. 6 for the $n = 1$ band or in Fig. 9 for the $n = 2$ band. In such cases, using the convex hull of the boundary points $(\alpha_i^\pm, \beta_i^\pm)$ to define the lensing band can produce an incorrect grid that overestimates the size of the band and includes points which do not really belong in it, which is why must use the concave hull instead.

In Fig. 4, we display an example of an AART grid (with a coarse resolution for visualization purposes), which includes points in the $n = 0$ (orange dots), $n = 1$ (purple crosses), and $n = 2$ (magenta dots) lensing bands.

G. Visualization of the equatorial transfer functions

Having defined the transfer functions (33)–(35) that map the equatorial plane of a Kerr black hole onto the image plane of a distant observer, we will now illustrate their behavior with the help of contour plots. Though we have yet to describe their range, we have already elucidated their domain of definition: as described in Sec. II D, the functions $r_s^{(n)}$, $\phi_s^{(n)}$, and $t_s^{(n)}$ are only physical within the corresponding n th lensing band.

Following Refs. [2,5], in Figs. 5 and 6, we display contours of fixed $r_s^{(0)}$ and $r_s^{(1)}$, which foliate the $n = 0$ and $n = 1$ lensing bands, respectively. Physically, these “isoradial” curves are the direct ($n = 0$) and first relativistic ($n = 1$) images of the rings of constant Boyer-Lindquist radius r_s in the equatorial plane $\theta_s = \pi/2$. Since the $n = 0$

lensing band is unbounded, $r_s^{(0)}$ has noncompact support, while $r_s^{(1)}$ maps the entire Kerr equatorial plane into a finite annulus: the $n = 1$ lensing band. In each band, $r_s^{(n)}$ spans the entire range $[r_+, \infty)$ once and only once, so every equatorial ring produces precisely one image in each lensing band. Moreover, the map $r_s^{(n)}$ is order-preserving: the source radius grows monotonically in the radial direction.

These two figures also illustrate the behavior of $\phi_s^{(0)}$ and $\phi_s^{(1)}$ by painting the Kerr equatorial plane with a color wheel, as in Ref. [61]. This wheel changes colors across “isopolar” curves of fixed ϕ_s , which are drawn at every 45° in the source plane. Their corresponding images—the curves of fixed $\phi_s^{(n)}$ —form a swirling pattern that illustrates the effects of frame-dragging. We emphasize that this swirl is a purely geometric effect, and that gravitational redshift is not yet included at this stage (its effects will be described in Sec. III below).

Unlike $r_s^{(n)}$, $\phi_s^{(n)}$ has a rather complicated range. While $\phi_s^{(0)}$ spans the entire range $[0, 2\pi)$ once and only once, so that every point source produces precisely one direct image (as in flat spacetime), as n grows large, the maps $\phi_s^{(n)}$ cover this range an increasing (linearly divergent) number of times [45,62,63].

As a result, the equatorial plane is unfolded increasingly many times in higher lensing bands, which therefore contain multiple images of a single point source. This surprising and intricate lensing behavior is connected to the development of caustics [64] and will be explored elsewhere.

While the transfer functions $r_s^{(n)}$ and $\phi_s^{(n)}$ suffice to describe the lensing of a static source, one still needs $t_s^{(n)}$ to describe a time-dependent source and account for its time-delay effects. Since light takes forever to reach a distant observer that is truly at asymptotic infinity, $t_s^{(n)}$ is technically infinite when $r_o \rightarrow \infty$. More precisely, the null geodesic equation in Kerr dictates that

$$\frac{dt}{dr} \Big|_{r \rightarrow \infty} \approx 1 + \frac{2M}{r} + \mathcal{O}\left(\frac{1}{r^2}\right). \quad (39)$$

Hence, the coordinate time elapsed along a geodesic grows as

$$t_s^{(n)r_o \rightarrow \infty} \approx r_o + 2M \log r_o + \mathcal{O}(1), \quad (40)$$

with the leading linear divergence arising from the first term in Eq. (39), and the subleading log-divergence from the second. Because of these divergences, $t_s^{(n)}$ is inherently dependent on the observer radius. However, subtracting these (“infrared”) divergences yields a “renormalized” time

$$\tilde{t}^{(n)} \equiv t_s^{(n)} - (r_o + 2M \log r_o), \quad (41)$$

which remains finite even as $r_o \rightarrow \infty$. Moreover, in this limit, the dependence on r_o drops out, and $\tilde{t}^{(n)}$ tends to a constant along any outward-propagating light ray. In the limit $M \rightarrow 0$, $\tilde{t}^{(0)} \rightarrow t_s^{(0)} - r_o$ and $\tilde{t}_s^{(0)} = 0$ is the future light cone of the origin. In other words, in flat space, a ray emitted radially outwards from $r_s = 0$ at $t_s = 0$ reaches null infinity $r_o \rightarrow \infty$ with $\tilde{t}^{(0)} = 0$.

We display density plots of $\tilde{t}^{(0)}$ and $\tilde{t}^{(1)}$ in Figs. 7 and 8, respectively. In practice, to take the limit $r_o \rightarrow \infty$, we evaluate Eq. (41) at a large value of $r_o \gg M$. The color scale in Fig. 7 ensures that $\tilde{t}^{(0)} = 0$ is white and sits in the middle of the scale. At higher spins and inclinations, the range of $\tilde{t}^{(0)}$ grows wider and more skewed toward negative values. Accordingly, we use a different color scale in each panel. We also display three contours of fixed time: $\tilde{t}^{(0)} = -10M$ (solid), $\tilde{t}^{(0)} = 0$ (dashed), and $\tilde{t}^{(0)} = 10M$ (dotted). These ‘‘isochronous’’ curves may be either open or closed, and in some cases, they may even be composed of multiple disconnected segments, highlighting the warped nature of the black hole spacetime.

The range of $\tilde{t}^{(n)}$ is always infinite in every lensing band. More precisely, it is unbounded below but bounded above: the reason is that if the entire equatorial plane emits light forever, then an observer can see light from arbitrarily far back in the past (emitted from regions that are either very close to the horizon or very far behind it) but it does take some minimum time for any light to reach an observer, even when it is emitted from the nearest point in the plane. In Fig. 7, we do not see arbitrarily large negative values of $\tilde{t}^{(0)}$ because we have a finite resolution (and therefore do not have pixels that resolve rays arbitrarily close to the horizon) and because we cut off the disk at $r_s = 20M$ (and therefore do not see rays emitted from farther behind the black hole).

The infinite range of $\tilde{t}^{(1)}$ is more readily apparent in Fig. 8, where we use a different color scale to highlight the divergent time elapsed along light rays at the edges of the lensing band: the inner edge consists of rays that asymptote to the horizon infinitely far back in the past, while the outer edge consists of rays that bounce back towards null infinity, incurring another time lapse that diverges as Eq. (40). The rapidly changing colors at the edges of the lensing band illustrate this behavior. Again, the finite resolution near the inner edge, together with the cutoff at $r_s = 20M$ near the outer edge, preclude us from resolving light rays emitted arbitrarily far back in the past. In Fig. 8, we chose a color scale spanning 99.9% of the range of time lapses sampled by the pixels making up the $n = 1$ grid.

Comparing Figs. 7 and 8, we see that successive images of the equatorial plane are delayed by a time of order $\tau \sim 15M$. Likewise, comparing the $n = 1$ image of the equatorial plane in Fig. 6 to its $n = 2$ image in Fig. 9 shows

that successive images are not only time-delayed but also demagnified and rotated. This lensing behavior results in a self-similar photon ring substructure that is governed by Kerr critical exponents γ , δ , and τ , which, respectively, control the demagnification, rotation, and time delay of successive subring images [5,7].

H. Analytic ray tracing with Beloborodov’s approximation

Direct ($n = 0$) light rays experience the smallest deflection. In Schwarzschild, Beloborodov used an ingenious expansion to derive an excellent analytic expression that approximates the trajectories of such rays [44]. His small-deflection-angle expansion is remarkably effective because its leading, linear term only receives its first subleading correction from a cubic term with a small coefficient. To describe this result, we define

$$b = \sqrt{\alpha^2 + \beta^2}, \quad \cos \psi = -\frac{\beta \tan \theta_o}{\sqrt{b^2 + \beta^2 \tan^2 \theta_o}} \in [-1, 1]. \quad (42)$$

In terms of these variables, a straightforward manipulation of Beloborodov’s formulas leads to the approximate expression

$$\frac{r_s^{(0)}}{M} \approx \sqrt{\left(\frac{1 - \cos \psi}{1 + \cos \psi}\right)^2 + \frac{b^2}{1 - \cos^2 \psi}} - \left(\frac{1 - \cos \psi}{1 + \cos \psi}\right), \quad (43)$$

which is highly accurate for low-to-moderate inclinations θ_o , as shown in Fig. 10. The agreement may seem surprising in the Kerr case, since this formula is derived for Schwarzschild. In practice, the approximation (43) is excellent for rays that never get within $\sim 4M$ of the black hole—for rays that come closer and experience greater deflection, it remains good along the portion of the ray up to the first equatorial crossing, but breaks down afterward, once the deflection angle grows large.

Intuitively, the reason Eq. (43) holds even for nonzero spins is that $n = 0$ rays are only weakly lensed and do not spend sufficient time orbiting the black hole to explore its geometry. Typically, $n = 0$ rays only come close enough to the black hole to probe the leading, monopole moment of its gravitational field (the mass), but are largely insensitive to its subleading, dipole moment (the spin), which contributes small corrections to the transfer function in an inverse-radius expansion.¹⁰

¹⁰This observation also underlies the ‘‘just add one’’ prescription $r_s^{(0)} \approx b - M$, which gives an excellent approximation to the transfer function for a polar observer at $\theta_o = 0$ and for any value of the spin, as first noticed in Ref. [5] and then derived, along with subleading $\mathcal{O}(M/b)$ corrections in Ref. [45].

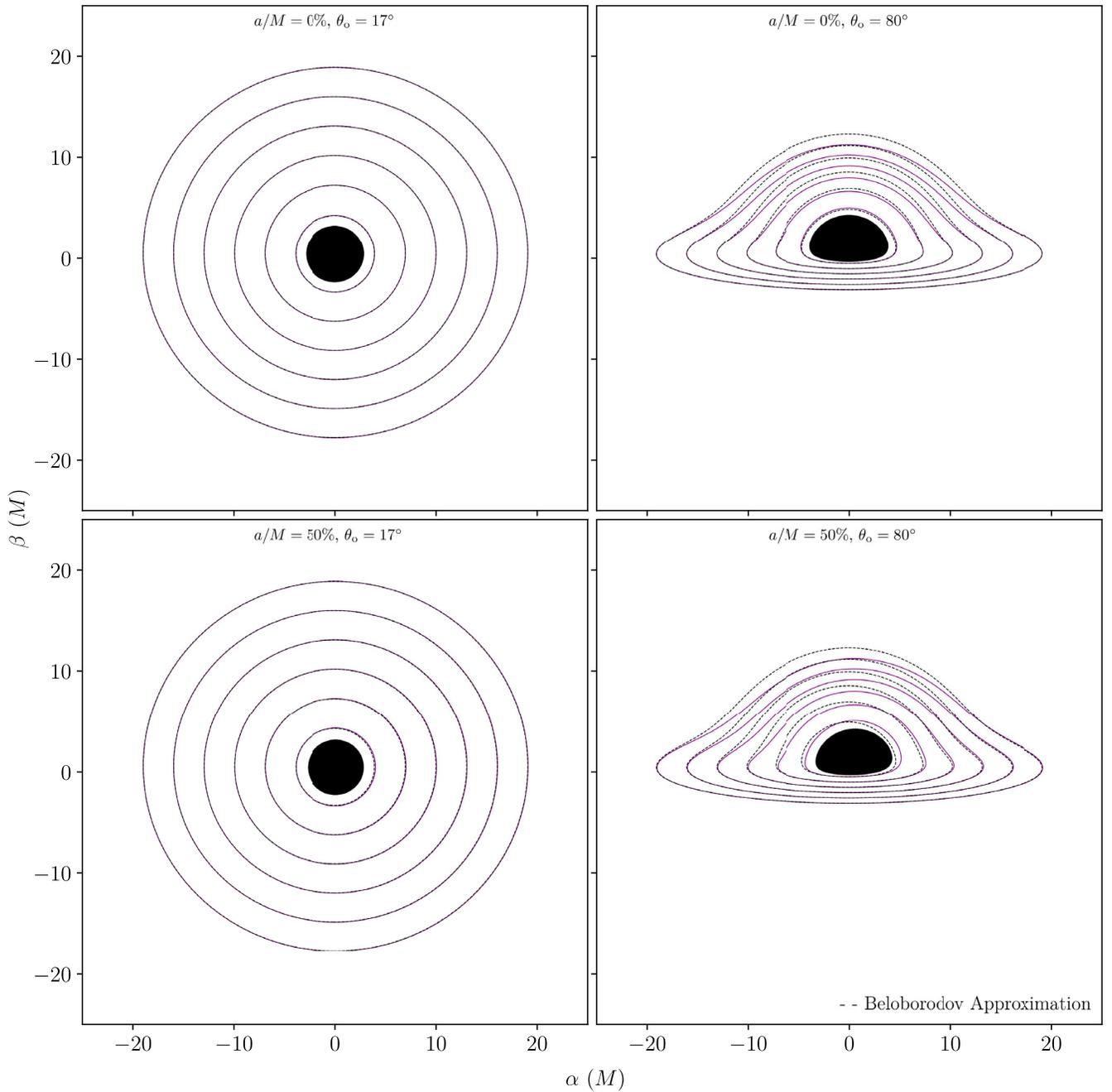


FIG. 10. Apparent positions of source rings of constant Boyer-Lindquist radius r_s in the equatorial plane ($\theta_s = \pi/2$) for different values of the spin and inclination within the $n = 0$ band, using the exact expression [Eq. (33)] and the Beloborodov approximation [Eq. (43)]. The Beloborodov approximation is still good even for moderate values of the spin and high inclination values.

With a little effort, under the Beloborodov approximation for the direct image, one can show that the sign of the quantity

$$\tau_B = \cos \psi - \frac{1}{1 - \frac{3\tilde{b}}{b} \cos[\frac{1}{3} \arccos(-\frac{\tilde{b}}{b})]}, \quad (44)$$

where $\tilde{b} = \sqrt{\tilde{\alpha}^2 + \tilde{\beta}^2} = 3\sqrt{3}M$ is the Schwarzschild critical impact parameter (apparent image radius of the critical

curve), determines whether a ray shot back from an image-plane position (α, β) with $b > \tilde{b}$, which must necessarily encounter a radial turning point, does so before ($\tau_B > 0$) or after ($\tau_B < 0$) first crossing the equatorial plane at radius (43).¹¹

¹¹Tsupko provides an analytic expression for the shape of $n \geq 2$ images [65].

I. Photon four-momentum at equatorial crossings

Last, we describe how to compute the source momentum of a photon that is loaded onto a ray as it crosses the equator.

The four-momentum $p = p_\mu dx^\mu$ of a Kerr photon is

$$p = E \left(-dt \pm_r \frac{\sqrt{\mathcal{R}(r)}}{\Delta(r)} dr \pm_\theta \sqrt{\Theta(\theta)} d\theta + \lambda d\phi \right), \quad (45)$$

where $E = -p_t$ is the photon energy, while $\mathcal{R}(r)$ and $\Theta(\theta)$ are the radial and angular geodesic potentials (9)–(10), which depend on the specific (energy-rescaled) angular momentum λ and Carter constant η of the photon defined in Eq. (3).

A ray shot backwards from image-plane position (α, β) has conserved quantities (λ, η) obtained by inverting Eq. (4). When such a ray crosses the equatorial plane for the $(n+1)$ th time at the radius $r_s^{(n)}$ given by Eq. (33), a photon of energy E loaded onto the ray will have a geodesic momentum given by Eq. (45) evaluated at $r = r_s^{(n)}$ and $\theta = \pi/2$ with those values of (λ, η) . This specifies the momentum up to discrete signs $\pm_{r,\theta}$. We now describe how to also determine these two signs.

First, $\pm_\theta = \text{sign } p_s^\theta$ is trivial to compute, since $n = 0$ photons must be emitted toward the observer, and so they always have $\pm_\theta = -\text{sign}(\cos \theta_o)$. This sign must flip at every crossing, so

$$\pm_\theta = (-1)^{n+1} \text{sign}(\cos \theta_o) \quad (46)$$

at the $(n+1)$ th equatorial crossing. As for $\pm_r = \text{sign } p_s^r$, two possibilities arise. Rays that appear inside the critical curve can never encounter a radial turning point and are therefore always outgoing; as such, for rays inside of \mathcal{C} , we always have

$$\text{Inside } \mathcal{C}: \quad \pm_r = +. \quad (47)$$

On the other hand, rays that appear outside the critical curve do encounter a radial turning point at radius $r = r_4$, which they reach at Mino time $\tau = \tau_4$ given in Eq. (30). Hence, for rays outside of \mathcal{C} , \pm_r depends on whether the Mino time $\tau_s^{(n)} = G_\theta^{(n)}$ of the $(n+1)$ th equatorial crossing, which is given in Eq. (A1), precedes or follows the radial turn at Mino time τ_4 ; that is,

$$\text{Outside } \mathcal{C}: \quad \pm_r = \text{sign}(\tau_4 - G_\theta^{(n)}). \quad (48)$$

When using the Beloborodov approximation from Sec. II H to ray trace $n = 0$ images, one merely replaces the exact $r_s^{(0)}$ in the preceding discussion by its approximation (43), which specifies the geodesic momentum at first equatorial crossing up to signs $\pm_{r,\theta}$. The angular sign is

still $\pm_\theta = -\text{sign}(\cos \theta_o)$. For the radial sign \pm_r , we must again distinguish between rays inside and outside \mathcal{C} , but this time the relevant critical curve is that of Schwarzschild: a circle of constant radius $b = \tilde{b}$, where $b = \sqrt{\alpha^2 + \beta^2}$ and $\tilde{b} = 3\sqrt{3}M$. Rays inside \mathcal{C} (with $b < \tilde{b}$) have $\pm_r = +$ as always, while rays outside \mathcal{C} with $(b > \tilde{b})$ have

$$\pm_r = \text{sign } \tau_B. \quad (49)$$

III. EQUATORIAL EMISSION MODEL

In the previous section, we described how Kerr black holes lens light. We will now introduce a simple, phenomenological model of electromagnetic emission from the equatorial plane and describe how to compute its observational appearance as seen by distant observers such as ourselves.

A. Intensity images

Our goal is to compute an observed intensity I_o at every image-plane position (α, β) . For each pixel in the image, we compute $I_o(\alpha, \beta)$ by tracing the corresponding ray backwards into the geometry; each time it passes through the emission region—in this case, the equatorial plane—we load additional photons onto the ray according to the local source intensity I_s . Since I/ν^3 is the invariant number of photons of frequency ν along a ray, it follows that

$$I_o(\alpha, \beta) = \sum_{n=0}^{N(\alpha, \beta)} \zeta_n g^3(r_s^{(n)}, \alpha, \beta) I_s(r_s^{(n)}, \phi_s^{(n)}, t_s^{(n)}), \quad (50)$$

which generalizes to nonstationary and nonaxisymmetric sources the analogous formula in Ref. [35]. In this expression, ζ_n is a geometric “fudge” factor and $g = \nu_o/\nu_s$ is the observed redshift, while the transfer functions $r_s^{(n)}$, $\phi_s^{(n)}$, and $t_s^{(n)}$, which are given in Eqs. (33)–(35), denote the spacetime position of the ray’s $(n+1)$ th equatorial crossing, with n ranging from 0 to the total number N of crossings given in Eq. (31).

The factor ζ_n is meant to account for the (neglected) effects of geometrical thickness, which we can mimic with [35]

$$\zeta_n = \begin{cases} 1 & n = 0, \\ \zeta & n > 0, \end{cases} \quad 0 < \zeta \leq 1. \quad (51)$$

The inclusion of this geometric factor significantly improves the agreement of this simplified equatorial model of emission with time-averaged radiative GRMHD simulations [9].

Finally, to derive the observed redshift g , we must prescribe a four-velocity for the accretion flow of radiating

TABLE I. Summary of the different accretion flows that we consider. The circular Keplerian flow \hat{u} corresponds to Cunningham’s geodesic prescription [67], while \bar{u} denotes the radial geodesic inflow. The nongeodesic, sub-Keplerian flow \hat{u} is the circular motion obtained by rescaling the Keplerian specific angular momentum $\hat{\ell} = -\hat{u}_\phi/\hat{u}_t$ by a factor $0 < \xi < 1$. The nongeodesic flow \tilde{u} is a general linear superposition of these purely circular and radial motions.

| Motion type | | Four-velocity | Definition |
|-------------|--------------------|---------------|--------------|
| Geodesic | Circular Keplerian | \hat{u} | Appendix B 1 |
| | Radial infall | \bar{u} | Appendix B 2 |
| Nongeodesic | Sub-Keplerian | \hat{u} | Appendix B 3 |
| | General flow | \tilde{u} | Appendix B 4 |

matter. We consider flows with a four-velocity of the general form

$$u = u^t(\partial_t - \iota\partial_r + \Omega\partial_\phi), \quad (52)$$

where the angular and radial-infall velocities are defined as

$$\Omega = \frac{u^\phi}{u^t}, \quad \iota = -\frac{u^r}{u^t}, \quad (53)$$

and u^μ are the contravariant components of the four-velocity. In Appendix B, we introduce two purely circular flows: a geodesic, Keplerian flow \hat{u} , together with a nongeodesic, sub-Keplerian flow \hat{u} obtained by rescaling the Keplerian angular momentum by a sub-Keplerianity parameter $0 < \xi \leq 1$, such that $\hat{u} \rightarrow \hat{u}$ as $\xi \rightarrow 1$. We also introduce the four-velocity \bar{u} of purely radial geodesic inflow. Then, following Refs. [11,66], we introduce a flow \tilde{u} given by a linear superposition of purely circular and purely radial motion. The resulting combined motion, which we present and derive in detail in Appendix B, is a nongeodesic family of flows with three parameters: the sub-Keplerianity factor $0 < \xi \leq 1$, and two other parameters $0 \leq \beta_r \leq 1$ and $0 \leq \beta_\phi \leq 1$ controlling the radial and angular components of the superposition, respectively. For $\beta_r = \beta_\phi = 0$, \tilde{u} reduces to the ξ -independent radial inflow \bar{u} . At the other extreme, when $\beta_r = \beta_\phi = 1$, \tilde{u} reduces to the sub-Keplerian flow \hat{u} , which for $\xi = 1$ recovers the circular-equatorial geodesic flow first used by Cunningham [67]. We summarize these flows in Table I.

Once a specific four-velocity (52) has been prescribed, the observed redshift is then given by [Eq. (54)]

$$g = \frac{E}{-p_\mu u^\mu} = \left[u^t \left(1 \pm_r \frac{\sqrt{\mathcal{R}(r)}}{\Delta(r)} \iota - \lambda\Omega \right) \right]^{-1}, \quad (54)$$

where the conserved quantities (λ, η) of the photon trajectory are related to its apparent position (α, β) via Eq. (4), while $\pm_r = \text{sign}p_s^r$ denotes the sign of the photon radial

momentum at its source whose computation is described in Sec. III above.

Given a black hole spin a , an observer inclination θ_o , some prescribed accretion flow u , and an equatorial emission profile I_s , we now have everything needed to compute an observed image (50). To summarize, the procedure is the following:

- (1) determine the lensing bands as described in Sec. II D;
- (2) in each lensing band, define a regular Cartesian grid of appropriate resolution—the resulting grids (indexed by n) form the different image layers described in Sec. II F;
- (3) for each pixel in a given layer, trace the corresponding ray back into the geometry and compute its equatorial crossings $x_s^{(n)}$ using the transfer functions in Sec. II C;
- (4) use the prescribed flow u to compute the redshift (54); and
- (5) finally, for each pixel in the image, use the equatorial crossings from 3, the redshift from 4, and the prescribed source profile I_s to compute the observed intensity (50).

A key observation is that the source profile I_s only enters this procedure at the very last step. In particular, steps 1 through 3 depend only on the black hole spin and observer inclination, and can therefore be computed once and for all for each choice of (a, θ_o) . The output can then be reused for each new choice of accretion flow u or source profile I_s , offering a significant time advantage (particularly when ray tracing *movie* frames).

Some comments are in order about step 2. First, while there is no reason to favor regular grids in principle, they do offer computational advantages in practice, since many numerical algorithms (such as grid interpolation) are optimized for such grids. Second, there is no real reason to favor Cartesian grids, and AART only relies on them for simplicity—it may be, however, that for extremely high n , regular but non-Cartesian grids become more computationally efficient. Third, the exact choice of “appropriate resolution” adapted to each grid will be described in detail in Sec. V.

B. Polarimetric images

In principle, besides the observed intensity (Stokes I), we could also ray trace the observed linear polarization (Stokes $P = Q + iU$) to obtain a *polarimetric image*. In practice, rather than computing the Stokes parameters Q and U separately, we will instead simply express the observed linear polarization as

$$P_o = mI_o e^{2i\chi}, \quad (55)$$

where $0 \leq m \leq 1$ is a fractional degree of polarization and χ denotes the EVPA (5). In a realistic model, m itself would

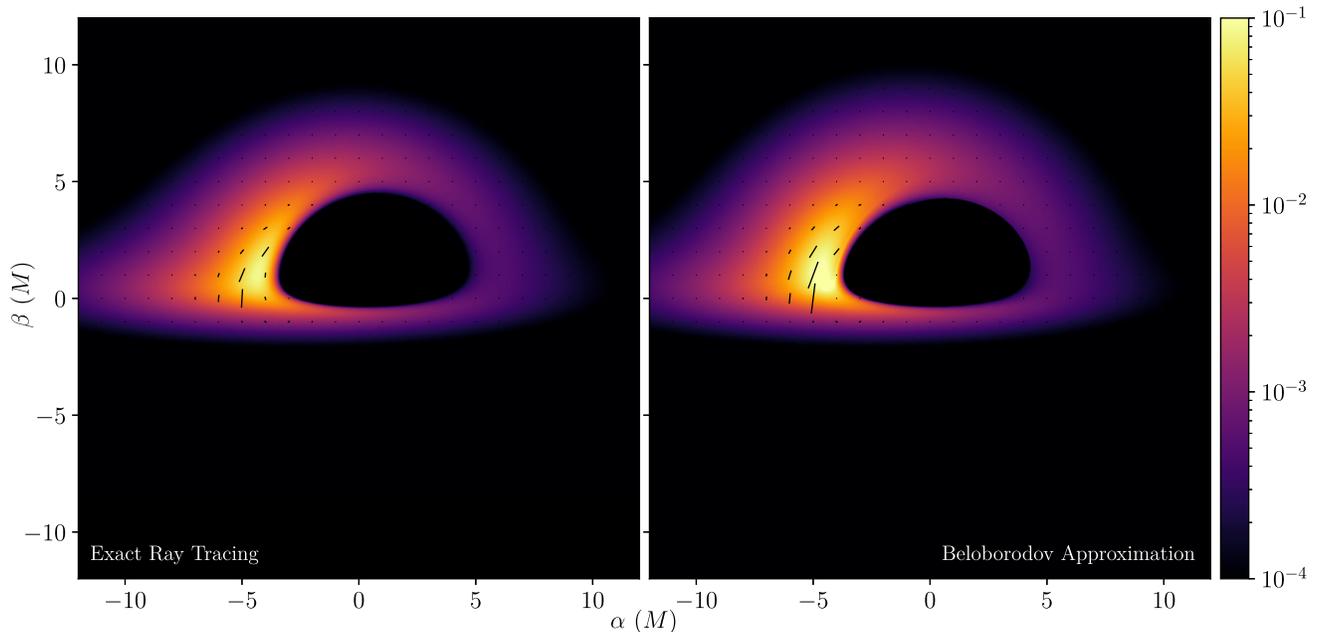


FIG. 11. Observed intensity I_o (in logarithmic scale) and linear polarization P_o corresponding to the stationary and axisymmetric source profile (58) with parameters $\mu = r_-$, $\vartheta = M/4$, and $\gamma = -1$. The linear polarization is represented by ticks (56) that are aligned with the plane of polarization of the light—given by the EVPA [Eq. (5)]—and whose length encodes the magnitude of Stokes P . Left: Polarimetric $n = 0$ image ray traced using the exact analytic transfer functions in Sec. II C. Right: Same image ray traced using the Beloborodov approximation (43) for the source radius $r_s^{(0)}$. The black hole spin $a/M = 50\%$ and observer inclination $\theta_o = 80^\circ$ correspond to the case from Fig. 10 with the largest difference between the exact and approximate isoradial curves, yet the resulting images appear almost identical. Here, the accretion flow follows the Cunningham prescription [67] corresponding to geodesic circular motion, i.e., $u = \hat{u}$ in Appendix B 1, and we took the source magnetic field to have purely radial spacetime components, i.e., only B'_s is nonzero.

vary across the image, since the degree of polarization of a photon usually depends on the angle at which the corresponding light ray intersects the local magnetic field at the source.

Here, we will be content to set m to a constant, resulting in sufficiently realistic polarimetric images for our purposes. In effect, this amounts to assuming that our astrophysical source emits isotropically, which is a convenient choice that allows us to continue treating the emission profile I_s as a scalar, rather than a directed quantity. With this choice (redefining the color scale to absorb the proportionality factor m), the magnitude $|P_o| = mI_o$ of the observed linear polarization is identical to the image intensity I_o , leaving only its direction—the EVPA—as the sole new feature to be displayed in polarimetric images. We indicate the orientation of the plane of polarization using “polarimetric ticks”: these are polarization vectors [53,68]

$$f_o = mI_o(-\sin\chi\partial_\alpha + \cos\chi\partial_\beta), \quad (56)$$

where m is chosen to make the ticks legible and χ is computed from a prescribed source polarization profile (described in the next section) that we parallel transport

along light rays using the conservation of the Penrose-Walker constant κ defined in Eq. (6).¹² The length $|f_o| = |P_o| \propto I_o$ of the ticks also encodes the magnitude of the polarization, which improves readability. We present an example of a polarimetric image in Fig. 11.

In the remainder of this section, we describe the equatorial emission profiles implemented in AART. The modularity of the code makes it easy to include any quantity of interest in the ray tracing, but we limit ourselves to Stokes I and P in this paper.

C. Stationary and axisymmetric source profiles

State-of-the-art time-averaged GRMHD-simulated images of realistic sources can be mimicked by ray tracing images of stationary, axisymmetric, equatorial emission profiles [9]. Such models therefore offer a computationally cheap way to study either time-independent or long-time-averaged sources.

¹²A slightly more realistic synchrotron emission model would set $m \propto \sin^2 \zeta$, where ζ denotes the angle of emission relative to the local magnetic field, but including this nonisotropy does not produce a large effect [54].

Following [10,35], we model such sources by setting

$$I_s(r_s, \phi_s, t_s) = J(r_s), \quad (57)$$

where we take the radial profile $J(r)$ to be the analytic function

$$J_{\text{SU}}(r; \mu, \vartheta, \gamma) \equiv \frac{e^{-\frac{1}{2}[\gamma + \text{arcsinh}(\frac{r-\mu}{\vartheta})]^2}}{\sqrt{(r-\mu)^2 + \vartheta^2}}, \quad (58)$$

which is derived from Johnson's Standard Unbounded (SU) distribution and can be rapidly computed to arbitrary precision everywhere. The three parameters μ , ϑ , and γ , respectively, control the location of the profile's peak, its width, and the profile asymmetry. We invite the reader to consult Sec. 5 of Ref. [10] for a more in-depth description of these parameters and their interpretation.

To illustrate the computation of a polarimetric image and the validity of Beloborodov's approximation in Sec. II H, we now consider a black hole of spin $a/M = 50\%$ viewed from an inclination $\theta_o = 80^\circ$. This is the case in Fig. 10 for which the differences between the exact and approximate $r_s^{(0)}$ are largest. We focus on the $n = 0$ image layer, which we ray trace with both the exact transfer function (33) and its approximation (43). We adopt Cunningham's prescription $u = \hat{u}$ (described in Appendix B 1) for the accretion flow, resulting in the redshift \hat{g} given by Eq. (B22) for sources outside the ISCO (B16) and by Eq. (B29) for sources within. This completes steps 1 through 4 above. In step 5, we set $J(r_s) = J_{\text{SU}}(r_s)$ with $\mu = r_-$ [Eq. (1)], $\vartheta = M/4$, and $\gamma = -1$, so the emission peaks past the horizon.

Finally, to prescribe the source polarization f_s , we impose

$$f_s \cdot p_s = 0, \quad f_s \cdot u_s = 0, \quad f_s \cdot B_s = 0. \quad (59)$$

The first condition states that the polarization is perpendicular to the direction of light propagation—this must by definition be true everywhere, including at the source. Meanwhile, the second condition implies that the polarization vector is purely spatial in the emitter frame, since

$$f^{(t)} = -f_{(t)} = -f_\mu \mathbf{e}^\mu_{(t)} = -f_s \cdot u_s, \quad (60)$$

where Latin indices (a) label tensor components in the local orthonormal frame $\mathbf{e}^\mu_{(a)}$ of the emitter, with time leg $\mathbf{e}^\mu_{(t)} = u_s^\mu$, while Greek indices μ label spacetime components. Imposing this condition does not lead to any loss of generality. Rather, it is simply a way to fix gauge: under gauge transformations, the polarization undergoes a gauge shift $f_s \rightarrow f_s + c p_s$ [53] that leaves the first condition invariant but shifts $f_s \cdot u_s$ by $c p_s \cdot u_s$. Thus, setting

$c = -(f'_s \cdot u_s)/(f_s \cdot u_s)$ results in a gauge-fixed f_s that obeys the second condition by construction. Last, the third condition asserts that the polarization is perpendicular to the spacetime vector B_s . This is the only physical assumption imposed by the relations (59). A suitable choice for modeling synchrotron emission is to let B_s be the local magnetic field.¹³

Together, the three conditions (59) determine three out of four components f_s^μ of the source polarization, which suffices to fix its spacetime orientation and hence the observed EVPA. The remaining component essentially controls the magnitude of the observed polarization, which in our model is not fixed at the source, but rather at the observer where we peg it to the observed intensity I_o . Mathematically, this prescription fixes the Penrose-Walker constant (6) up to an overall scale factor that drops out of the formula (5) for the EVPA χ . More precisely, solving (59) yields, in terms of $b^{\mu\nu} = B_s^\mu p_s^\nu - B_s^\nu p_s^\mu$,

$$f_r = \frac{b^{\theta\phi} - b^{t\theta}\Omega}{b^{\theta\phi}_t + b^{r\phi}\Omega} f_t, \quad (61)$$

$$f_\theta = \frac{b^{r\phi} + b^{t\phi}_t + b^{tr}\Omega}{b^{\theta\phi}_t + b^{r\phi}\Omega} f_t, \quad (62)$$

$$f_\phi = -\frac{b^{r\theta} + b^{t\theta}_t}{b^{\theta\phi}_t + b^{r\phi}\Omega} f_t, \quad (63)$$

where, for clarity, we have temporarily dropped the subscripts “s.” Lowering indices in Eq. (6) and evaluating it at $\theta = \pi/2$ then gives the source Penrose-Walker constant $\kappa_s = \kappa_1 + i\kappa_2$:

$$r\kappa_1 = a(p_r f_\phi - p_\phi f_r) - (r^2 + a^2)(p_t f_r - f_t p_r), \quad (64)$$

$$r\kappa_2 = (p_\theta f_\phi - p_\phi f_\theta) - a(p_t f_\theta - p_\theta f_t). \quad (65)$$

The computation of the photon momentum p_s at the source is described in Sec. II I, while the parameters (t, Ω) of the flow \hat{u} are given in Appendix B 1. The component f_t that controls the magnitude of the polarization remains undetermined, but it factors out of both κ_1 and κ_2 , and therefore drops out of the ratio in the formula (5) for the EVPA χ , as mentioned above. The resulting explicit expressions are implemented in AART.

¹³As a technical aside, we emphasize here that the condition $f_s \cdot u_s = 0$ is also crucial for the following reason. Suppose that we start with a polarization f'_s which only obeys $f'_s \cdot p_s = 0$, while $f'_s \cdot u_s \neq 0$ and $f'_s \cdot B_s \neq 0$. Then a gauge shift $f'_s \rightarrow f_s = f'_s + c p_s$ with $c = -(f'_s \cdot B_s)/(p_s \cdot B_s)$ results in a physically equivalent polarization vector that obeys $f_s \cdot B_s = 0$. Thus, one can always make the initial polarization perpendicular to the magnetic field by applying a gauge transformation that leaves the observed polarization invariant. The gauge-fixing condition in Eq. (59) is therefore essential.

All that is left to specify are the spacetime components B_s^μ of the local magnetic field. The decomposition of the electromagnetic field strength $F_{\mu\nu}$ into electric and magnetic fields is frame dependent. In a realistic synchrotron emission model, the polarization would be perpendicular to the local magnetic field in the emitter frame, with components $B_s^\mu = -\star F^{\mu\nu} u_\nu$.

For our example, we simply choose the spacetime field B_s to be purely radial, keeping only a nonzero B_s^r component whose magnitude also scales out of the EVPA χ . We display the resulting $n = 0$ images, ray traced using both the exact and approximate $r_s^{(0)}$ in the left and right panels of Fig. 11, respectively. Even though the corresponding isoradial curves (plotted in the right column of Fig. 10) differ noticeably, the ray traced, images in Fig. 11 nonetheless look very similar and would be indistinguishable with near-term observations. The Beloborodov approximation is thus excellent for ray tracing direct images of axisymmetric source configurations.

Having showcased the polarimetric capabilities of AART, we now leave polarization aside and turn to variable sources.

D. Modeling variable sources with Gaussian random fields

Variable (nonstationary and nonaxisymmetric) sources can broadly be divided into two classes. On the one hand, one can consider specific physical phenomena that are governed by deterministic equations. A commonly studied example is that of a “hot spot”: a localized source of enhanced emissivity that usually orbits around the black hole and thereby produces a characteristic pattern of light echoes (see, e.g., Refs. [62,69–73]). Adding such a source to our model is straightforward in AART: one merely needs to insert the deterministic source to the right-hand side (RHS) of Eq. (57) before carrying out the usual ray tracing.

On the other hand, one can also consider sources subject to statistical fluctuations. Hot spots are transient and unlikely to be present in any given observation of M87*, but we certainly expect its surrounding plasma to flare and produce emission ropes (or other photon ring mimickers), which in the absence of a definite physical model we can still represent as random fluctuations in the astrophysical source. A widely used tool to model such astrophysical noise is the Gaussian random field (GRF); see, e.g., Ref. [74] for applications to cosmology. A random field \mathcal{G} is a function on a space X such that $\mathcal{G}(\mathbf{x})$ is a random variable for every $\mathbf{x} \in X$. A GRF is a random field that is completely determined by its mean μ and covariance C ,

$$\mu(\mathbf{x}) = \langle \mathcal{G}(\mathbf{x}) \rangle, \quad C(\Delta\mathbf{x}) = \langle \mathcal{G}(\mathbf{x})\mathcal{G}(\mathbf{x} + \Delta\mathbf{x}) \rangle \geq 0. \quad (66)$$

More precisely, we define a random field $\mathcal{G}(\mathbf{x})$ with mean $\mu(\mathbf{x})$ and covariance $C(\Delta\mathbf{x})$ to be a GRF if it satisfies the property

$$\langle e^{i \sum_{\ell=1}^k s_\ell \mathcal{G}(\mathbf{x}_\ell)} \rangle = e^{-\frac{1}{2} \sum_{\ell,j=1}^k C(\Delta\mathbf{x}_{\ell j}) s_\ell s_j + i \sum_{\ell=1}^k \mu(\mathbf{x}_\ell) s_\ell}, \quad (67)$$

which states that its joint probability distribution on any set of k points \mathbf{x}_ℓ , with $\ell \in \{1, \dots, k\}$, is a k -dimensional multivariate Gaussian distribution with mean vector $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_k\}$ and covariance matrix $\mathbf{C}_{\ell j} = C(\Delta\mathbf{x}_{\ell j})$, where $\Delta\mathbf{x}_{\ell j} = |\mathbf{x}_\ell - \mathbf{x}_j|$.

The following discussion is technical and readers interested only in applications may skip to Sec. III H.

The mean and covariance are also known as the one-point and two-point (or autocorrelation) function, respectively. As a result of the defining relation (67), these low-point correlation functions determine all higher k -point functions of a GRF via Wick’s theorem (also known as the Isserlis theorem), which may be derived by differentiation with respect to multiple s_ℓ .

For instance, letting \mathcal{G}_ℓ denote $\mathcal{G}(\mathbf{x}_\ell)$ for a zero-mean GRF, it is immediately seen by differentiating Eq. (67) with respect to s_1, s_2 , and s_3 that $\mathcal{G}(\mathbf{x})$ must have a vanishing three-point function, $\langle \mathcal{G}_1 \mathcal{G}_2 \mathcal{G}_3 \rangle = 0$. An additional derivative with respect to s_4 implies that the four-point function $\langle \mathcal{G}_1 \mathcal{G}_2 \mathcal{G}_3 \mathcal{G}_4 \rangle$ is

$$\langle \mathcal{G}_1 \mathcal{G}_2 \rangle \langle \mathcal{G}_3 \mathcal{G}_4 \rangle + \langle \mathcal{G}_1 \mathcal{G}_3 \rangle \langle \mathcal{G}_2 \mathcal{G}_4 \rangle + \langle \mathcal{G}_1 \mathcal{G}_4 \rangle \langle \mathcal{G}_2 \mathcal{G}_3 \rangle. \quad (68)$$

In general, for a zero-mean GRF, all k -point functions vanish for k odd, whereas for k even they are given by all the possible “contractions” of the two-point functions $\langle \mathcal{G}_\ell \mathcal{G}_j \rangle = C(\Delta\mathbf{x}_{\ell j})$. This is Wick’s theorem in a nutshell, and it makes precise the idea that a GRF is a “simple” random field whose correlation structure is fully determined by two functions $\mu(\mathbf{x})$ and $C(\Delta\mathbf{x})$.

This property is very convenient for our modeling purposes, since only two functions need to be specified to prescribe all of the statistics of our astrophysical source. For example, Wick’s theorem also fixes all the even moments of a zero-mean GRF:

$$\langle \mathcal{G}^{2k}(\mathbf{x}) \rangle = (2k - 1)!! [C(0)]^k, \quad \langle \mathcal{G}^{2k+1}(\mathbf{x}) \rangle = 0, \quad (69)$$

where the $k = 2$ case is compatible with Eq. (68) evaluated at coincident points. (Of course, the odd moments all vanish.)

We now specialize to GRFs in d -dimensional Euclidean space $X = \mathbb{R}^d$. A GRF is homogeneous if it is invariant under translations, so $\mathcal{G}(\mathbf{x} + \mathbf{a}) = \mathcal{G}(\mathbf{x})$ for $\mathbf{a} \in \mathbb{R}^d$, and it is isotropic if it looks the same in all directions, so $\mathcal{G}(\mathbf{x}) = \mathcal{G}(x)$ with $x = |\mathbf{x}|$.

A real-space GRF $\mathcal{G}(\mathbf{x})$ defines a (generally complex) GRF $\tilde{\mathcal{G}}(\mathbf{k})$ in momentum space via the Fourier transform

$$\tilde{\mathcal{G}}(\mathbf{k}) = \int \mathcal{G}(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d^d\mathbf{x}, \quad \mathcal{G}(\mathbf{x}) = \int \frac{\tilde{\mathcal{G}}(\mathbf{k})}{(2\pi)^d} e^{i\mathbf{k}\cdot\mathbf{x}} d^d\mathbf{k}. \quad (70)$$

If $\mathcal{G}(\mathbf{x})$ is homogeneous, then we let $\mathbf{x}' = \mathbf{x} + \Delta\mathbf{x}$ and Fourier transform its autocorrelation to derive the covariance of $\tilde{\mathcal{G}}(\mathbf{k})$:

$$\langle \tilde{\mathcal{G}}^*(\mathbf{k}) \tilde{\mathcal{G}}(\mathbf{k}') \rangle = \iint \langle \mathcal{G}(\mathbf{x}) \mathcal{G}(\mathbf{x}') \rangle e^{i\mathbf{k}\cdot\mathbf{x} - i\mathbf{k}'\cdot\mathbf{x}'} d^d\mathbf{x} d^d\mathbf{x}' \quad (71)$$

$$= \iint C(\Delta\mathbf{x}) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} e^{-i\mathbf{k}'\cdot\Delta\mathbf{x}} d^d\mathbf{x} d^d\Delta\mathbf{x} \quad (72)$$

$$= \int e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{x}} d^d\mathbf{x} \int C(\Delta\mathbf{x}) e^{-i\mathbf{k}'\cdot\Delta\mathbf{x}} d^d\Delta\mathbf{x} \quad (73)$$

$$= (2\pi)^d \delta^{(d)}(\mathbf{k} - \mathbf{k}') \tilde{C}(\mathbf{k}'). \quad (74)$$

Similarly, higher-point functions of $\tilde{\mathcal{G}}(\mathbf{k})$ may also be obtained by Fourier transforming those of $\mathcal{G}(\mathbf{x})$. One finds that $\tilde{\mathcal{G}}(\mathbf{k})$ is a GRF obeying Eq. (67) with mean $\hat{\mu}(\mathbf{k})$ and covariance (74), which is also referred to as the power spectrum $P(\mathbf{k})$ of $\mathcal{G}(\mathbf{x})$. When $\mathcal{G}(\mathbf{x})$ is also isotropic, so is $P(\mathbf{k}) = P(k)$ with $k = |\mathbf{k}|$. Like covariances, power spectra are always non-negative.

The most ubiquitous GRF is the white noise process $\mathcal{W}(\mathbf{x})$. It is the standard, homogeneous, and isotropic GRF defined by a flat power spectrum with uniform $\tilde{C}(\mathbf{k}') = 1$, so that

$$\langle \tilde{\mathcal{W}}^*(\mathbf{k}) \tilde{\mathcal{W}}(\mathbf{k}') \rangle = (2\pi)^d \delta^{(d)}(\mathbf{k} - \mathbf{k}'), \quad (75)$$

$$\langle \mathcal{W}(\mathbf{x}) \mathcal{W}(\mathbf{x}') \rangle = \delta^{(d)}(\mathbf{x} - \mathbf{x}'). \quad (76)$$

With Eq. (67), this delta-function autocorrelation implies that $\mathcal{W}(\mathbf{x})$ consists of independent Gaussian random variables at every $\mathbf{x} \in \mathbb{R}^d$, and likewise in momentum space for $\tilde{\mathcal{W}}(\mathbf{k})$. Thus, creating realizations of white noise is trivial: it suffices to draw from independent normal distributions at every point.

In fact, Eq. (74) provides a straightforward way to produce realizations of any homogeneous GRF $\mathcal{G}(\mathbf{x})$: first, one creates a realization $\tilde{\mathcal{W}}(\mathbf{k})$ of white noise; next, one multiplies it by $\sqrt{\tilde{C}(\mathbf{k})}$ to obtain a realization of $\hat{\mathcal{G}}(\mathbf{k}) = \sqrt{\tilde{C}(\mathbf{k})} \tilde{\mathcal{W}}(\mathbf{k})$; finally, inverse Fourier transforming $\hat{\mathcal{G}}(\mathbf{k})$ yields a realization of $\mathcal{G}(\mathbf{x})$. [We note that this procedure is well-defined since $\tilde{C}(\mathbf{k}) \geq 0$.]

Just as a standard Gaussian distribution has zero mean and unit variance, we define a standard GRF as a zero-mean GRF with unit covariance at the origin: $\mu(\mathbf{x}) = 0$ and $C(0) = 1$.

We now consider the Matérn field $\mathcal{F}_\nu(\mathbf{x})$ of order ν , which is another standard, homogeneous, and isotropic

GRF that is well-known to statisticians, who use it to model a wide range of processes; see, e.g., Ref. [75] for its various applications. This zero-mean GRF obeys the Matérn covariance of order ν ,

$$C_\nu(\mathbf{x}) = \frac{1}{2^{\nu-1} \Gamma(\nu)} \left(\frac{x}{\lambda}\right)^\nu K_\nu\left(\frac{x}{\lambda}\right), \quad (77)$$

where Γ denotes the Gamma function and $K_\nu(x)$ the modified Bessel function of the second kind of order ν . Here, λ is a correlation length, while ν is a differentiability parameter that we will always take to be $\nu = n - d/2$ for some positive integer n . At short distances, $C_\nu(\mathbf{x}) \approx 1 + c_\nu x^{2\nu}$ for some constant c_ν , so it is best (though not strictly necessary) to require $\nu > 0$.

The Matérn field $\mathcal{F}_\nu(\mathbf{x})$ is also ubiquitous in physics: though its position-space covariance (77) may seem unfamiliar, in momentum space it becomes¹⁴ (see Appendix C for a derivation)

$$\tilde{C}_\nu(\mathbf{k}) = \frac{\mathcal{N}^2 \lambda^d}{(1 + \lambda^2 k^2)^{\nu + \frac{d}{2}}}, \quad \mathcal{N}^2 = (4\pi)^{\frac{d}{2}} \frac{\Gamma(\nu + \frac{d}{2})}{\Gamma(\nu)}, \quad (78)$$

which for $n = 1$ (or $\nu = 1 - d/2$), we recognize as the quantum propagator for a free scalar field of mass $m = \lambda^{-1}$ in Euclidean signature. To the best of our knowledge, this link has not been stated explicitly before, and we explore it in detail in Appendix C.

The Matérn field $\mathcal{F}_\nu(\mathbf{x})$ enjoys a special connection to linear stochastic partial differential equations (SPDE). Realizations of its Fourier transform $\tilde{\mathcal{F}}_\nu(\mathbf{k})$ are related to white noise via

$$\tilde{\mathcal{F}}_\nu(\mathbf{k}) = \sqrt{\tilde{C}_\nu(\mathbf{k})} \tilde{\mathcal{W}}(\mathbf{k}). \quad (79)$$

Such a relation holds for any homogeneous GRF, but it takes a particularly simple form for the Matérn covariance (78):

$$(1 + \lambda^2 k^2)^{\frac{d}{2} + \nu} \tilde{\mathcal{F}}_\nu(\mathbf{k}) = \mathcal{N} \lambda^{\frac{d}{2}} \tilde{\mathcal{W}}(\mathbf{k}). \quad (80)$$

Returning to position space gives the linear (fractional) SPDE

$$(1 - \lambda^2 \nabla^2)^{\frac{d}{2} + \nu} \mathcal{F}_\nu(\mathbf{x}) = \mathcal{N} \lambda^{\frac{d}{2}} \mathcal{W}(\mathbf{x}), \quad (81)$$

which exactly matches Eq. (2) of Ref. [76] since $\mathcal{N}^2 \sigma^2 = \lambda^{2\nu}$. As we show in Appendix C, for $n = 1$ and $\nu = 1 - d/2$, this SPDE is compatible with the identification $\mathcal{F}_\nu(\mathbf{x}) \equiv \Phi(\mathbf{x})$, where $\Phi(\mathbf{x})$ is a free Euclidean scalar field of mass $m = \lambda^{-1}$, as expected.

¹⁴Equivalently, $\tilde{C}_\nu(\mathbf{k}) = \frac{\sigma^{-2}}{(k^2 + \lambda^{-2})^{\nu + \frac{d}{2}}}$ with $\sigma^2 = \frac{\lambda^{2\nu}}{(4\pi)^{\frac{d}{2}}} \frac{\Gamma(\nu)}{\Gamma(\nu + \frac{d}{2})}$ as in Ref. [76].

If instead, $n = 2$ and $\nu = 2 - d/2$, then the resulting Matérn field, which we will simply denote $\mathcal{F}(\mathbf{x})$, obeys a linear SPDE

$$(1 - \lambda^2 \nabla^2) \mathcal{F}(\mathbf{x}) = \mathcal{N} \lambda^{\frac{d}{2}} \mathcal{W}(\mathbf{x}), \quad (82)$$

which has no fractional derivative and is thus simpler to solve numerically, as explained in Ref. [43] below Eq. (3) therein.

This connection between the Matérn field $\mathcal{F}(\mathbf{x})$ and a linear SPDE is especially useful once we introduce inhomogeneities. Inhomogeneous GRFs are hard to generate via other means than solving the associated SPDE; in particular, the trick (79) for generating realizations of $\tilde{\mathcal{F}}(\mathbf{k})$ breaks down because the Fourier modes are no longer delta-correlated as in Eq. (74).

E. Inhomogeneous and anisotropic Matérn fields

Before describing how to include inhomogeneity, we first discuss how to model homogeneous anisotropies, following closely the excellent treatment in Ref. [43].

Still in d dimensions, we introduce d orthonormal vectors \mathbf{u}_ℓ , d correlation lengths $\lambda_\ell > 0$, with $\ell \in \{1, \dots, d\}$, and define

$$\mathbf{\Lambda} = \sum_{\ell=1}^d \lambda_\ell^2 \mathbf{u}_\ell \mathbf{u}_\ell^T, \quad |\mathbf{\Lambda}| \equiv \det \mathbf{\Lambda} = \prod_{\ell=1}^d \lambda_\ell^2. \quad (83)$$

We will also allow the unit vectors \mathbf{u}_ℓ to not be orthogonal, as long as $|\mathbf{\Lambda}|$ retains the same form. The matrix $\mathbf{\Lambda}$ is invertible, and its inverse $\mathbf{\Lambda}^{-1}$ defines a metric on \mathbb{R}^d with line element

$$ds^2(\Delta \mathbf{x}) = \Delta \mathbf{x} \cdot \mathbf{\Lambda}^{-1} \Delta \mathbf{x} = \sum_{\ell=1}^d \left(\frac{\Delta \mathbf{x} \cdot \mathbf{u}_\ell}{\lambda_\ell} \right)^2. \quad (84)$$

We now use this to define the generalized Matérn covariance

$$C_\nu(\mathbf{x}) = \frac{1}{2^{\nu-1} \Gamma(\nu)} ds^\nu(\mathbf{x}) K_\nu(ds(\mathbf{x})). \quad (85)$$

When $\lambda_1 = \dots = \lambda_n = \lambda$, so that $\mathbf{\Lambda} = \lambda^2 \mathbf{I}$, this reproduces the homogeneous, isotropic Matérn covariance (77), as desired.

However, if the constants λ_ℓ are unequal, then the resulting GRF remains homogeneous but becomes anisotropic along a direction set by the unit vectors \mathbf{u}_ℓ . An example with $d = 2$ is displayed in Fig. 1 of Ref. [43]. Following the steps that led to Eq. (73), the corresponding momentum-space covariance is

$$\tilde{C}_\nu(\mathbf{k}) = \frac{\mathcal{N}^2 \sqrt{|\mathbf{\Lambda}|}}{(1 + \mathbf{k} \cdot \mathbf{\Lambda} \mathbf{k})^{\nu + \frac{d}{2}}}. \quad (86)$$

In this case, we can still use Eq. (79) to obtain realizations of the anisotropic Matérn field. We can also follow the derivation of Eq. (81) to find that the anisotropic field obeys the SPDE

$$(1 - \nabla \cdot \mathbf{\Lambda} \nabla)^{\frac{d}{2} + \frac{1}{2}} \mathcal{F}_\nu(\mathbf{x}) = \mathcal{N} |\mathbf{\Lambda}|^{\frac{1}{4}} \mathcal{W}(\mathbf{x}). \quad (87)$$

Finally, to include inhomogeneity, we allow the unit vectors $\mathbf{u}_\ell \rightarrow \mathbf{u}_\ell(\mathbf{x})$ to become functions. Then $\mathbf{\Lambda} \rightarrow \mathbf{\Lambda}(\mathbf{x})$ becomes a function too—with the same definition (83)—but the rest of the previous discussion breaks down. Instead, we must reverse the logic and *define* the inhomogeneous, anisotropic Matérn field $\mathcal{F}_\nu(\mathbf{x})$ by the SPDE (87) with variable coefficients $\mathbf{\Lambda}(\mathbf{x})$.

We emphasize that the resulting field $\mathcal{F}_\nu(\mathbf{x})$ no longer obeys the covariance (86), which is not a function on momentum space once $\mathbf{\Lambda}(\mathbf{x})$ is position-dependent, but it is a well-defined GRF arising as a solution to the SPDE (87); see also Ref. [77].

As explained below Eq. (81), in practice, we will only use the field $\mathcal{F}(\mathbf{x})$ with $n = 2$ and $\nu = 2 - d/2$ as it obeys the SPDE

$$[1 - \nabla \cdot \mathbf{\Lambda}(\mathbf{x}) \nabla] \mathcal{F}(\mathbf{x}) = \mathcal{N} |\mathbf{\Lambda}(\mathbf{x})|^{\frac{1}{2}} \mathcal{W}(\mathbf{x}), \quad (88)$$

which has no fractional derivatives and is therefore more tractable numerically. Since we took our white noise to have unit variance, this SPDE exactly matches Eq. (5) of Ref. [43] (after setting $\sigma = 1$ therein), which is what INOISY solves.

One last comment is in order: in the inhomogeneous case, the Matérn field $\mathcal{F}(\mathbf{x})$ that INOISY produces by numerically integrating Eq. (88) does not follow the covariance (85); in particular, it may not be standard with $\mu(\mathbf{x}) = 0$ and $C(0) = 1$. To “standardize” it, we will always work with a rescaled field

$$\hat{\mathcal{F}}(\mathbf{x}) = \frac{\mathcal{F}(\mathbf{x}) - \langle \mathcal{F}(\mathbf{x}) \rangle}{\sqrt{\langle \mathcal{F}^2(\mathbf{x}) \rangle - \langle \mathcal{F}(\mathbf{x}) \rangle^2}}, \quad (89)$$

which also happens to be precisely what INOISY implements. For a homogeneous field, this makes no statistical difference.

F. Nonstationary and nonaxisymmetric source profiles

With the inhomogeneous, anisotropic Matérn field $\hat{\mathcal{F}}(\mathbf{x})$ in hand, we are now in a position to model the fluctuations of an equatorial accretion disk and produce INOISY simulations.

Before doing so, we first describe what we want to achieve intuitively, without being too rigorous. We let $\mathbf{x}_s = (r_s, \phi_s, t_s)$. As in Ref. [8], we want the total intensity (57) to consist of a background radial profile $J(r_s)$ with some fluctuations $\Delta J(\mathbf{x}_s)$:

$$I_s(\mathbf{x}_s) = J(r_s) + \Delta J(\mathbf{x}_s) = J(r_s)[1 + F(\mathbf{x}_s)]. \quad (90)$$

Here, $F \equiv \Delta J/J$ is the fractional variation in the source, which at first we assume to be fractionally small: $0 < |F(\mathbf{x}_s)| \ll 1$. We also want these fluctuations to wash out under averaging,

$$\langle F(\mathbf{x}_s) \rangle = 0, \quad (91)$$

so that after averaging over fluctuations using $\langle J(r_s) \rangle = J(r_s)$, we recover the stationary, axisymmetric profile (57):

$$\langle I_s(\mathbf{x}_s) \rangle = \langle J(r_s) \rangle + \langle J(r_s)F(\mathbf{x}_s) \rangle = J(r_s). \quad (92)$$

Eventually, we will also want to consider large fluctuations with $\Delta J \sim J$, or $F \sim \mathcal{O}(1)$, but this could pose a problem: if at some point \mathbf{x}_s , a fluctuation grew so large and negative that $F(\mathbf{x}_s) < -1$, then we would have $\Delta J(\mathbf{x}_s) < -J(r_s)$, and hence $I_s(\mathbf{x}_s) < 0$ at that point, which is nonsensical because an intensity is a (necessarily positive) count of photons—we must therefore prevent this from happening.

Of course, we could demand that the fluctuations always remain small, but there is a way to allow them to grow large while still ensuring they do not result in a negative intensity. The idea is to note that for small fluctuations $0 < |F(\mathbf{x}_s)| \ll 1$, $1 + F \approx 1 + F + \frac{1}{2!}F^2 + \frac{1}{3!}F^3 + \dots = e^F$, so we may well replace the RHS of Eq. (90) by a (necessarily positive) exponential:

$$I_s(\mathbf{x}_s) \cong J(r_s)e^{F(\mathbf{x}_s)}. \quad (93)$$

For small fluctuations, this seems completely equivalent, since

$$\frac{\Delta J(\mathbf{x}_s)}{J(r_s)} = \frac{I_s(\mathbf{x}_s) - J(r_s)}{J(r_s)} \cong e^{F(\mathbf{x}_s)} - 1 \stackrel{|F| \ll 1}{\approx} F(\mathbf{x}_s), \quad (94)$$

reproducing Eq. (90). The advantage of this definition is that it can now be freely extended to arbitrarily large $|F|$, even in the negative direction: a fluctuation $F(\mathbf{x}_s) \ll -1$ would just result in a small but still positive (and hence still physical) intensity, rather than an unphysical, negative emissivity. Hence, our new definition (93) for $I_s(\mathbf{x}_s)$ remains sensible even for large and negative field excursions, unlike our first attempt (90).

However, as we are about to see, this new definition suffers from a subtle issue: it does not actually satisfy Eq. (92). For now, we use “ \cong ” to remind us of this, and proceed naively.

The preceding intuitive discussion hinges on being able to produce a fluctuation field $F(\mathbf{x}_s)$ with some desired properties, such as Eq. (91). We now make this mathematically precise.

We replace $F(\mathbf{x}_s) \rightarrow \sigma \hat{\mathcal{F}}(\mathbf{x}_s)$, where $\hat{\mathcal{F}}(\mathbf{x}_s)$ is the standard, zero-mean Matérn field defined in Eq. (89) as the

solution to the SPDE (87). We then define an emissivity field [43]

$$I_s(\mathbf{x}_s) \equiv \mathcal{J}(\mathbf{x}_s) \cong J(r_s)e^{\sigma \hat{\mathcal{F}}(\mathbf{x}_s)}, \quad (95)$$

where the parameter σ controls the scale of the fluctuations, since $\hat{\mathcal{F}}(\mathbf{x}_s)$ has unit covariance $C(0) = 1$. Because $\hat{\mathcal{F}}(\mathbf{x}_s)$ has zero mean, Eq. (91) is satisfied. Let us now check Eq. (92):

$$\langle \mathcal{J}(\mathbf{x}_s) \rangle \cong J(r_s) \langle e^{\sigma \hat{\mathcal{F}}(\mathbf{x}_s)} \rangle = J(r_s) \sum_{n=0}^{\infty} \frac{\langle \sigma^n \hat{\mathcal{F}}^n(\mathbf{x}_s) \rangle}{n!}, \quad (96)$$

where we series-expanded the exponential and used linearity of the expectation value. Then by Eq. (69), relabeling $k = 2n$,

$$\frac{\langle \mathcal{J}(\mathbf{x}_s) \rangle}{J(r_s)} \cong \sum_{k=0}^{\infty} \frac{\sigma^{2k} \langle \hat{\mathcal{F}}^{2k}(\mathbf{x}_s) \rangle}{(2k)!} = \sum_{k=0}^{\infty} \frac{(2k-1)!!}{(2k)!} \sigma^{2k}. \quad (97)$$

Evaluating the sum results in

$$\langle \mathcal{J}(\mathbf{x}_s) \rangle \cong J(r_s)e^{\frac{1}{2}\sigma^2}, \quad (98)$$

and we now see why the definition (93) is not what we wanted, since it does not satisfy the requirement (92). Instead, we let

$$I_s(\mathbf{x}_s) = \mathcal{J}(\mathbf{x}_s) \equiv J(r_s)e^{\sigma \hat{\mathcal{F}}(\mathbf{x}_s) - \frac{1}{2}\sigma^2}, \quad (99)$$

which does have the desired property (92), and hence recovers the background profile (57) after averaging. We note that this can also be checked directly from the property (67) with $k = 1$,

$$\langle e^{is\mathcal{G}(\mathbf{x})} \rangle = e^{-\frac{1}{2}C(0)s^2}, \quad (100)$$

by taking $s = -i\sigma$ and $\mathcal{G} = \hat{\mathcal{F}}$ with $\mu = 0$ and $C(0) = 1$. This also shows that σ^2 effectively rescales the covariance $C(0)$. In particular, the fluctuations disappear as $s \rightarrow 0$ and we recover the pure envelope $J(r_s)$ of the background surface brightness.

G. Complete statistics of the variable source

We can also use the GRF property to explicitly write down the complete correlation structure of the random field $\mathcal{J}(\mathbf{x}_s)$. With $s_\ell = -i\sigma$, $\mathcal{G} = \hat{\mathcal{F}}$, $\mu = 0$, and $C(0) = 1$, Eq. (67) becomes

$$\langle e^{\sigma \sum_{\ell=1}^k \hat{\mathcal{F}}(\mathbf{x}_\ell)} \rangle = e^{\frac{1}{2}\sigma^2 \sum_{\ell,j=1}^k C(\Delta \mathbf{x}_{\ell j})}. \quad (101)$$

Since $\langle \mathcal{J}(\mathbf{x}_1) \cdots \mathcal{J}(\mathbf{x}_k) \rangle = J(r_1) \cdots J(r_k) \langle e^{\sigma \sum_{\ell=1}^k \hat{\mathcal{F}}(\mathbf{x}_\ell) - \frac{k}{2}\sigma^2} \rangle$,

$$\frac{\langle \mathcal{J}(\mathbf{x}_1) \cdots \mathcal{J}(\mathbf{x}_k) \rangle}{J(r_1) \cdots J(r_k)} = \exp \left[-\frac{k\sigma^2}{2} + \frac{\sigma^2}{2} \sum_{\ell, j=1}^k C(\Delta \mathbf{x}_{\ell j}) \right]. \quad (102)$$

This result characterizes all the statistics of the variable source $\mathcal{J}(\mathbf{x}_s)$ in terms of a single covariance function, as advertised.

Of course, this formula with $k = 1$ confirms that Eq. (92) holds. Meanwhile, for $k = 2$ the double sum in the exponential is $\sum_{\ell, j=1}^2 C(\Delta \mathbf{x}_{\ell j}) = C(\Delta \mathbf{x}_{11}) + C(\Delta \mathbf{x}_{12}) + C(\Delta \mathbf{x}_{21}) + C(\Delta \mathbf{x}_{22})$ so we obtain the two-point function

$$\frac{\langle \mathcal{J}(\mathbf{x}_1) \mathcal{J}(\mathbf{x}_2) \rangle}{J(r_1) J(r_2)} = e^{-\sigma^2 + \frac{\sigma^2}{2} [2C(0) + 2C(\Delta \mathbf{x}_{12})]} = e^{\sigma^2 C(\Delta \mathbf{x}_{12})}, \quad (103)$$

which is the exponential of the Matérn field's autocorrelation. Going further, for $k = 3$ the double sum in the exponential is $\sum_{\ell, j=1}^3 C(\Delta \mathbf{x}_{\ell j}) = 3C(0) + 2C(\Delta \mathbf{x}_{12}) + 2C(\Delta \mathbf{x}_{13}) + 2C(\Delta \mathbf{x}_{23})$ so

$$\frac{\langle \mathcal{J}(\mathbf{x}_1) \mathcal{J}(\mathbf{x}_2) \mathcal{J}(\mathbf{x}_3) \rangle}{J(r_1) J(r_2) J(r_3)} = e^{\sigma^2 [C(\Delta \mathbf{x}_{12}) + C(\Delta \mathbf{x}_{13}) + C(\Delta \mathbf{x}_{23})]}, \quad (104)$$

and one can check that $\mathcal{J}(\mathbf{x}_s)$ is a non-Gaussian random field. That is, it does not quite obey the GRF property (67), though its statistics (102) are still governed by a single covariance.

For some intuition, let us reexamine the fractional variation $F \equiv \Delta J / J$, whose statistics are encoded in the random field

$$\hat{\mathcal{J}}(\mathbf{x}_s) \equiv \frac{\mathcal{J}(\mathbf{x}_s) - J(r_s)}{J(r_s)} = e^{\sigma \hat{\mathcal{F}}(\mathbf{x}_s) - \frac{1}{2}\sigma^2} - 1. \quad (105)$$

Its one-point and two-point functions are

$$\langle \hat{\mathcal{J}}(\mathbf{x}_s) \rangle = 0, \quad \langle \hat{\mathcal{J}}(\mathbf{x}_s) \hat{\mathcal{J}}(\mathbf{x}'_s) \rangle = e^{\sigma^2 C(\Delta \mathbf{x})} - 1. \quad (106)$$

If the fluctuations are very small, then this is approximately

$$\langle \hat{\mathcal{J}}(\mathbf{x}_s) \hat{\mathcal{J}}(\mathbf{x}'_s) \rangle^{\sigma^2 \ll 1} \approx \sigma^2 C(\Delta \mathbf{x}), \quad (107)$$

which is the autocorrelation of a GRF with covariance σ^2 . This is also true for higher-point functions; using Eq. (102), we can show that the field $\hat{\mathcal{J}}(\mathbf{x}_s)$ is approximately Gaussian at leading order in σ . Thus, small fluctuations are Gaussian, but larger fluctuations introduce non-Gaussianities.

H. Generating realizations of an accretion flow using INOISY

To summarize, we model variable sources via the emissivity (99), which is defined in terms of the Matérn field (89) and displays the correlation structure (102). The zero-mean field $\hat{\mathcal{F}}(\mathbf{x}_s)$ and its statistics are fully determined by its covariance, which is controlled by a metric $\Lambda(\mathbf{x}_s)$ of the form (83). Given such a metric, a realization of $\hat{\mathcal{F}}(\mathbf{x}_s)$ is obtained by solving the linear SPDE (88) with some realization $\mathcal{W}(\mathbf{x}_s)$ of white noise. This is essentially what INOISY does, though in practice, the continuous SPDE (88) is actually discretized and its solution is a Gaussian Markov random field approximating $\hat{\mathcal{F}}(\mathbf{x}_s)$ [43]. As for the envelope $J(r_s)$, we use the radial profile (58).

Therefore, all we need to do to specify our model is to fix a choice of unit vectors $\mathbf{u}_\ell(\mathbf{x}_s)$ and correlation lengths λ_ℓ , where now $\ell \in \{0, 1, 2\}$. Such a choice will define $\Lambda(\mathbf{x}_s)$ and hence our nonstationary, nonaxisymmetric stochastic source profile.

Since we want to produce “realistic” images and visibility amplitudes, we will choose $\hat{\mathcal{F}}(\mathbf{x}_s)$ to be an inhomogeneous, anisotropic Matérn field whose power spectrum is comparable to that observed in GRMHD simulations [43], and therefore serves as a good phenomenological model.

Instead of $\mathbf{x}_s = (r_s, \phi_s, t_s)$, we use a regular Cartesian grid $\mathbf{x}_s = (t_s, x_s, y_s)$ in the Kerr equatorial plane, and INOISY solves the SPDE (88) on this grid with periodic boundary conditions. Following Eq. (83), we pick a position-dependent anisotropy

$$\Lambda(\mathbf{x}_s) = \sum_{\ell=0}^2 \lambda_\ell^2(r_s) \mathbf{u}_\ell(\mathbf{x}_s) \mathbf{u}_\ell^T(\mathbf{x}_s). \quad (108)$$

Under this prescription, $\mathbf{u}_0(\mathbf{x}_s)$ sets the temporal correlation of the flow, with characteristic correlation time $\lambda_0(r_s)$, while $\mathbf{u}_1(\mathbf{x}_s)$ and $\mathbf{u}_2(\mathbf{x}_s)$ determine its spatial structure, which at any given time exhibits correlations of characteristic lengths $\lambda_1(r_s)$ and $\lambda_2(r_s)$, respectively. Following Ref. [43], we take these three-dimensional (3D) unit vectors to have (t_s, x_s, y_s) components

$$\mathbf{u}_0(\mathbf{x}_s) = (1, v_x(\mathbf{x}_s), v_y(\mathbf{x}_s)), \quad (109)$$

$$\mathbf{u}_1(\mathbf{x}_s) = (0, \cos \theta(\mathbf{x}_s), \sin \theta(\mathbf{x}_s)), \quad (110)$$

$$\mathbf{u}_2(\mathbf{x}_s) = (0, -\sin \theta(\mathbf{x}_s), \cos \theta(\mathbf{x}_s)), \quad (111)$$

where we have yet to specify v_x , v_y , and θ . In practice, we will only let these functions depend on the spatial position (x_s, y_s) .

We note that $\mathbf{u}_1(\mathbf{x}_s) \cdot \mathbf{u}_2(\mathbf{x}_s) = 0$ are always orthogonal to each other, but not to $\mathbf{u}_0(\mathbf{x}_s)$. Nonetheless, the resulting $\Lambda(\mathbf{x}_s)$ still has a determinant of the form required by

Eq. (83), since $|\mathbf{\Lambda}(\mathbf{x}_s)| = \lambda_0^2 \lambda_1^2 \lambda_2^2$. Also, the sign of \mathbf{u}_1 and \mathbf{u}_2 is arbitrary, as it sets only the spatial correlation; the flow in time is set by \mathbf{u}_0 .

To be consistent with our choice of accretion flow (52), we must take the temporal correlations to have the same velocity:

$$\vec{v} \equiv \frac{dr}{dt} \partial_r + \frac{d\phi}{dt} \partial_\phi = -\iota \partial_r + \Omega \partial_\phi, \quad (112)$$

where Ω and ι , respectively, denote the angular and radial-infall velocities (53). In Cartesian coordinates this velocity becomes $\vec{v} = v_x \partial_x + v_y \partial_y$ with

$$v_x(\mathbf{x}_s) = -\frac{x_s}{r_s} \iota - y_s \Omega, \quad v_y(\mathbf{x}_s) = -\frac{y_s}{r_s} \iota + x_s \Omega. \quad (113)$$

As for the spatial correlations, we adopt the prescription [43]

$$\theta(\mathbf{x}_s) = \arctan(y_s, -x_s) + \theta_\perp. \quad (114)$$

This sets the major axis $\mathbf{u}_1(\mathbf{x}_s)$ of the spatial correlation tensor to lie at a constant angle θ_\perp relative to the equatorial circles of constant radius r_s . The resulting flow displays spiral features with opening angle θ_\perp . The choice $\theta_\perp \approx 20^\circ$ produces spiral arms broadly consistent with GRMHD simulations [43,78].

This completes the specification of our stochastic surface brightness. To summarize, the accretion flow model takes as input 12 parameters, each of which is listed in Table II, along with a short description and the default value used for the main examples in this paper. To compute images, we

TABLE II. Summary of the 15 parameters in our nonstationary and nonaxisymmetric stochastic source model, together with their default values used for the majority of our examples. The first three pertain to the geometry, while the other 12 prescribe the statistics of the equatorial surface brightness.

| Parameter | Default value | Description |
|----------------|----------------|--|
| ψ | 1.07 | BH mass-to-distance ratio |
| a/M | 94% | BH spin |
| θ_o | 17° | Observer inclination |
| ζ | 1.5 | Geometrical factor |
| σ | 0.4 | Fluctuation scale |
| ξ | 0.95 | Sub-Keplerian factor |
| β_r | 0.95 | Radial velocity factor |
| β_ϕ | 0.95 | Angular velocity factor |
| θ_\perp | 20° | Anisotropy direction |
| λ_0 | $2\pi/\Omega$ | Temporal correlation |
| λ_1 | $5r_s$ | Spatial correlation in the x_s direction |
| λ_2 | $0.1\lambda_1$ | Spatial correlation in the y_s direction |
| μ | r_- | Controls location of profile peak |
| ϑ | 1/2 | Controls profile width |
| γ | -3/2 | Controls profile asymmetry |

must also specify three more parameters: the black hole spin a , the observer inclination θ_o , and the mass-to-distance ratio

$$\psi = \frac{1}{3.62 \mu\text{as}} \frac{M}{r_o}, \quad (115)$$

which combines the black hole mass M and observer distance r_o , and which is defined relative to a fiducial value for M87*.

IV. APPLICATIONS

Having fully described our stochastic source model, we can now simulate it with INOISY and then use AART to ray trace its appearance and compute its visibility. We will present a whole suite of simulations in a follow-up paper, but for now we limit ourselves to one complete example from which Figs. 1–3 are derived. We first explain in detail how to obtain Figs. 1 and 2, while Fig. 3 will be the focus of Sec. VI.

For this simulation, we ran INOISY on a regular Cartesian grid (x_s, y_s, t_s) of size $2048 \times 2048 \times 512$. For each of the spatial coordinates (x_s, y_s) , we uniformly placed 2048 pixels within the range $[-50, 50]M$, resulting in a spacing of $0.05M$ between pixels, whereas for the time coordinate t_s , we placed 512 pixels uniformly within the range $[0, 1000]M$, resulting in a cadence of one frame every $1.95M$. We used the default values listed in Table II for the 15 parameters in the model.

The left panel in Fig. 12 shows an example of a snapshot extracted from the resulting INOISY run. The accretion flow appears qualitatively similar to GRMHD-simulated flows. In particular, we tuned the scale σ of fluctuations to ensure that the total observed flux (the blue light curve in Fig. 13) varies by a factor of ~ 3 between its minimum and maximum. This level of time-variability mimics the light curves obtained from GRMHD simulations: see, for example, Figs. 3 and 4 of Ref. [79], Fig. 8 of Ref. [80], or Fig. 5 of Ref. [31]. Ensuring that this agreement also holds quantitatively is an interesting challenge that we hope to tackle in the future.

A. Fast-light: Stationary nonaxisymmetric source

After producing our INOISY simulation, we use AART to ray trace it. A commonly adopted simplification in carrying out ray tracing is the so-called “fast-light” approximation. When staring at a variable source, a single frame $I_o(t_o, \alpha, \beta)$ of its observational appearance at a fixed moment in time is formed by photons that were emitted at different times $t_s^{(n)}(\alpha, \beta)$ in the history of the source. The reason is that photons that appear at different positions (α, β) in the image plane follow different paths in the geometry, thereby incurring different time delays $\Delta t = t_o - t_s^{(n)}$ on their way from source to observer, as shown in Figs. 7 and 8.

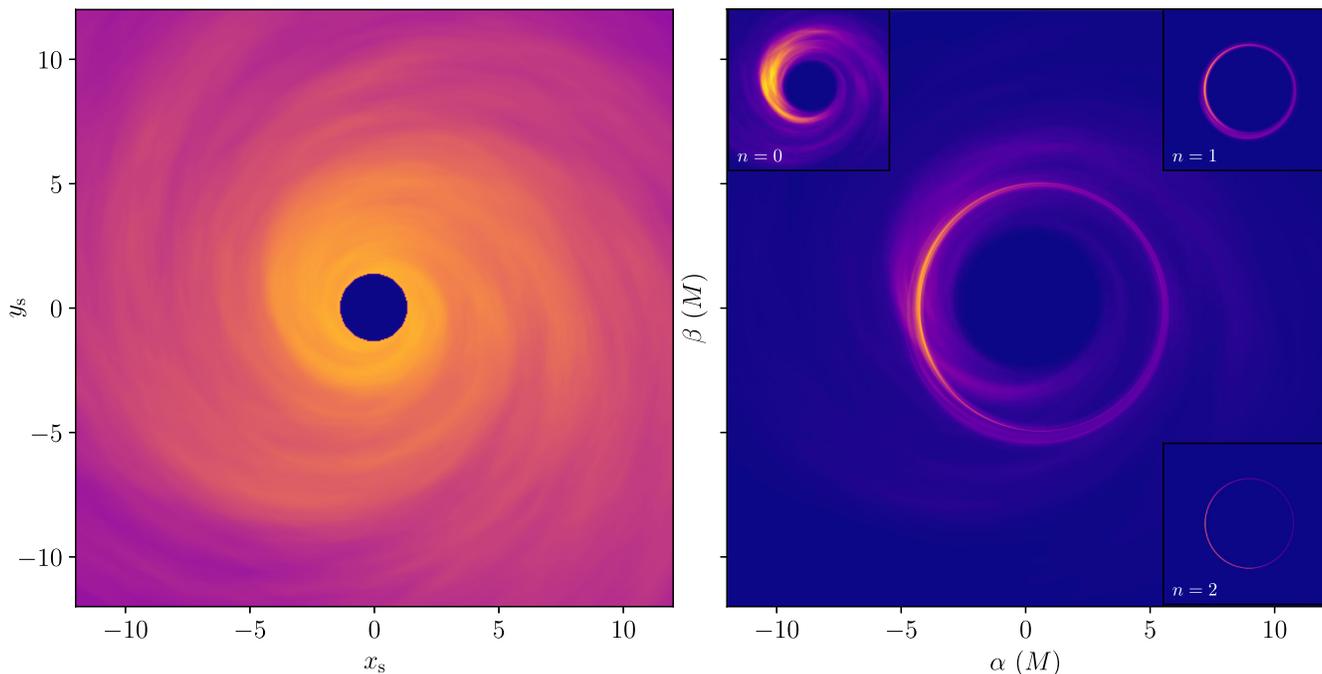


FIG. 12. Left: Intensity profile (in logarithmic scale) of a single snapshot from an INOISY simulation with the parameters in Table II. Right: Ray-traced image corresponding to this equatorial source profile, treating the realization of the random field on the left as static (the “fast-light” approximation). The inset panels decompose the image into layers: the direct $n = 0$ image and the first two ($n = 1$ and $n = 2$) photon rings.

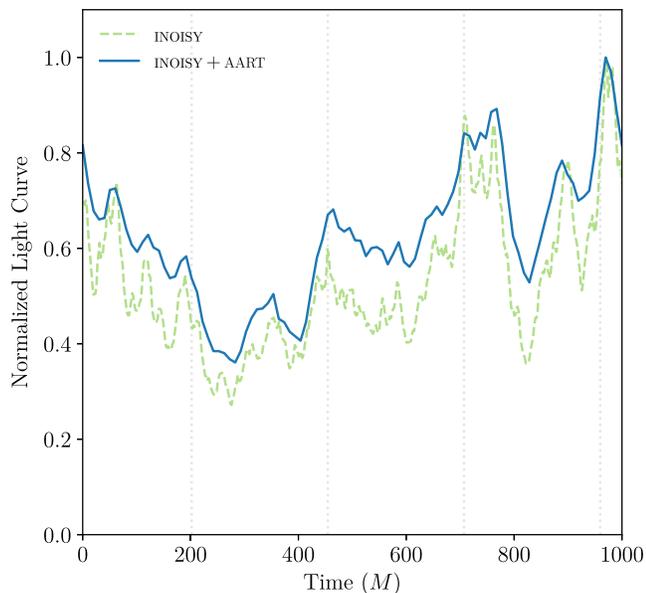


FIG. 13. Normalized light curves of the stochastic INOISY source with parameters in Table II. Green line: total emitted flux measured from INOISY snapshots. Blue line: total observed flux measured from ray-traced images. Relativistic effects (gravitational redshift, Doppler boosting, and light bending) make the observed flux smoother and less variable than the emitted flux. The vertical dotted lines indicate the times corresponding to the four snapshots displayed in Fig. 1.

The fast-light approximation simply ignores this variable time delay and maps the image plane (α, β) onto a single snapshot of the equatorial plane, using only the transfer functions $r_s^{(n)}$ and $\phi_s^{(n)}$ from Sec. II C and replacing Eq. (50) by

$$I_o(\alpha, \beta) = \sum_{n=0}^{N(\alpha, \beta)} \zeta_n g^3(r_s^{(n)}, \alpha, \beta) I_s(r_s^{(n)}, \phi_s^{(n)}). \quad (116)$$

Mathematically, this is equivalent to keeping the emission time t_s fixed while letting the observation time t_o vary across the image. Naturally, if the source is stationary, then this is a moot distinction since the time dependence drops out anyway, so this amounts to treating each individual INOISY snapshot as a stationary (though nonaxisymmetric) source. Thus, the fast-light approximation is exact for a stationary source, and it can offer a decent approximation as long as the source varies slowly relative to the “fast” light being traced, hence the name.

As an example, we take the first ($t_s = 0$) snapshot from our INOISY simulation and ray trace it as described in Sec. II C, assuming a black hole spin of $a/M = 94\%$ and an observer inclination of $\theta_o = 17^\circ$. In the $n = 0$ layer, we ray trace at the same spatial resolution as the underlying INOISY simulation; that is, we use a spacing of $\delta x^{(0)} = \delta x_{\text{INOISY}} = 0.05M$ between grid pixels. In the higher- n layers, we adaptively increase the resolution by a factor of

2 in each successive band, and ray trace on grids with spacings of $\delta x^{(n)} = 2^n \delta x^{(0)}$ between pixels.

The right panel of Fig. 12 displays the resulting image (in a logarithmic color scale to highlight the accretion flow), which looks qualitatively similar to the single snapshots produced with state-of-the-art GRMHD simulations [9,31]. The figure also displays a decomposition of the image into its $n \in \{0, 1, 2\}$ layers, which are shown individually in the inset panels. Each separate layer appears well-resolved, suggesting that our grid resolutions are sufficient to resolve the photon rings. Indeed, we have verified (using a convergence test detailed in Sec. V) that the results derived from the observable of interest (which in our case is the visibility amplitude, presented in Sec. IV C) do not change when the resolutions are doubled.

B. Slow-light: Nonstationary nonaxisymmetric source

With the transfer functions (33)–(35) implemented in AART, it is also easy to perform a “slow-light” ray-tracing that takes into account varying photon emission times across the image. We simply use Eq. (50) and plug in all the transfer functions defined in Sec. II C. This requires us to use the entire series of snapshots generated by INOISY; that is, we must now use the full time evolution of the accretion flow.

Slow-light tracing does introduce two new complications. First, since the pixels in a single frame of a movie of the source can depend on a very wide range of emission times $t_s^{(n)}(\alpha, \beta)$, in order to produce a movie of some duration T_o , we typically need to simulate the source over a longer duration $T_s > T_o$. In practice, we can estimate how much longer T_s needs to be by examining isochronal curves in the $n = 0$ layer, such as those displayed in Fig. 7. At low inclinations, the range of $t_s^{(0)}(\alpha, \beta)$ over sampled pixels is about $50M$, but this range grows much larger at higher inclinations. In addition, the higher- n images are composed of photons emitted roughly a time $n\tau$ earlier, so in order to ray trace a movie containing n layers, we expect to need $T_s - T_o \gtrsim 50M + n\tau \approx 80M$. Some rays may occasionally require sampling the source even earlier than the start of the simulation, but this is not really an issue with INOISY thanks to its use of periodic boundary conditions (including in time).

Second, because $t_s^{(n)}(\alpha, \beta)$ varies smoothly across the image plane, the code must be able to sample the source $I_s(r_s, \phi_s, t_s)$ for continuous values of t_s , including at times in between the frames computed by INOISY. This problem can be dealt with using interpolation, but only so long as the underlying INOISY simulation has sufficiently high resolution to smoothly resolve the flow’s motion. This is the case in our example simulation, as the INOISY movie of the equatorial source has high enough cadence to be smoothly interpolated in time.

Thus, AART readily produces a movie of the source (whose parameters are listed in Table II). We display four snapshots from this movie, taken at intervals of $250M$, in the top row of Fig. 1. We also plot its light curves in Fig. 13: the total observed flux (measured in the image plane) and total emitted flux (measured in the equatorial plane) are shown in blue and green, respectively, with the former appearing smoother and less variable than the latter due to relativistic effects.

As in the stationary (fast-light) case, the snapshot images in Fig. 1 present bright transient features that are qualitatively similar to those seen in state-of-the-art snapshots of GRMHD simulations [9,31]. These transient features can obscure the photon ring in instantaneous snapshots, but they wash out of the average over 100 snapshots in Fig. 2, leaving the photon ring as the only prominent feature in the time-averaged image shown in the left panel. This image can be directly compared to the one shown in Fig. 1 of Ref. [7], which was also obtained by time-averaging over 100 uniformly spaced snapshots taken from a GRMHD simulation with a time range of $1000M$.

C. Visibility on long baselines

A radio interferometer such as the EHT samples a (complex) radio visibility $V(\mathbf{u})$. By the van Cittert–Zernike theorem, this is related to a sky brightness $I_o(\mathbf{x}_o)$ via a 2D Fourier transform

$$V(\mathbf{u}) = \int I_o(\mathbf{x}_o) e^{-2\pi i \mathbf{u} \cdot \mathbf{x}_o} d^2 \mathbf{x}_o. \quad (117)$$

Here, \mathbf{x}_o are dimensionless image-plane coordinates measured in radians, such as $(\alpha, \beta)/r_o$, while the dimensionless vector \mathbf{u} is a “baseline” projected onto the plane perpendicular to the line of sight and measured in units of the observation wavelength [68]. We use radio-astronomy conventions to make this Fourier transform 2π -symmetric, unlike the definition (70).

Therefore, to relate simulated images to actual observables, we must Fourier transform them to compute their visibility. In particular, the narrow image features—like the photon rings—that encode precise information about the spacetime dominate the interferometric signal on very long baselines $u = |\mathbf{u}| \gg 1$ [7,35], requiring us to compute Fourier transforms up to very high frequencies. To do so, we again exploit the layered image structure: since the observed intensity (50) decomposes into

$$I_o(\alpha, \beta) = \sum_{n=0}^{N(\alpha, \beta)} I_n(\alpha, \beta), \quad (118)$$

with $I_n(\alpha, \beta)$ denoting the n th image layer, we can likewise use the linearity of the Fourier transform (117) to decompose the total complex visibility into individual subring components:

$$V(\mathbf{u}) = \sum_{n=0}^{\infty} V_n(\mathbf{u}), \quad V_n(\mathbf{u}) = \int I_n(\mathbf{x}_o) e^{-2\pi i \mathbf{u} \cdot \mathbf{x}_o} d^2 \mathbf{x}_o. \quad (119)$$

The subring images $I_n(\mathbf{x}_o)$ with $n > 0$ consist of narrow rings of characteristic widths $e^{-\gamma n}$ (where γ is a Lyapunov exponent) and roughly equal intensities. Hence the total flux $V_n(0)$ of the n th image layer also scales like $e^{-\gamma n}$ [7]. Moreover, a narrow ring produces a characteristic, weak $u^{-1/2}$ power-law falloff in its Fourier transform: indeed one can show that in the regime

$$\frac{1}{d} \ll u \ll \frac{1}{w}, \quad (120)$$

the visibility (117) in polar coordinates $\mathbf{u} = (u, \varphi)$ of a ring of diameter $\sim d$ and width $\sim w$ takes the “universal” form [33,34]

$$\frac{1}{2} V(\mathbf{u}) = \frac{e^{-2\pi i C_\varphi u}}{\sqrt{u}} \left[\alpha_\varphi^L e^{-\frac{i\pi}{4}} e^{i\pi d_\varphi u} + \alpha_\varphi^R e^{\frac{i\pi}{4}} e^{-i\pi d_\varphi u} \right], \quad (121)$$

where $\alpha_\varphi^{L,R} = \alpha_{\varphi+\pi}^{R,L} > 0$ encodes the polar intensity profile around the ring, while d_φ and C_φ are its projected diameter and centroid displacement at angle φ in the image, respectively.

As first pointed out in Ref. [7], this suggests that the n th photon ring dominates the interferometric signal in the regime

$$\frac{1}{w_{n-1}} \ll u \ll \frac{1}{w_n}, \quad (122)$$

where $w_n \approx e^{-\gamma n} w_0$ is the width of the n th subring, producing a characteristic cascading structure of damped oscillations with periodicity encoding the diameter of successive subrings; see also Secs. 2 and 4 of Ref. [10] for a more detailed discussion.

Since we compute images layer-by-layer—as described in Sec. II F—with a different grid and resolution in each layer I_n , it is convenient to also compute the visibility layer-by-layer as well, and obtain each V_n as the Fourier transform (119) of I_n .

Instead of taking this 2D Fourier transform directly, as in Refs. [10,35], we make use of the projection-slice theorem to compute $V_n(u, \varphi)$ along slices of fixed polar angle φ in the Fourier plane. The procedure is as follows. For each angle φ , we first compute the Radon transform along the cut at angle φ across the image; that is, in each lensing band, we integrate the observed intensity I_n along lines perpendicular to the slice of constant φ . Then, we interpolate each Radon transform to the resolution of the highest-order lensing band. Last, we sum up the contributions from all the image layers I_n , and perform a one-dimensional Fourier transform to finally obtain $V(u, \varphi)$.

The bottom row of Fig. 1 presents the visibility amplitudes $|V(u, \varphi)|$ along slices of constant $\varphi = 0^\circ$ (in red) and $\varphi = 90^\circ$ (in blue) of the corresponding snapshots in the top row. Since these individual snapshots display a strong dependence on the variable emission profile, the visibility amplitudes also exhibit a large variability. Nevertheless, as we mentioned previously, these fluctuations wash out under time-averaging, leaving the characteristic ringing signature of the photon ring as the main persistent feature that dominates the visibility.

We also display the visibility amplitudes at baseline angles $\varphi = 0^\circ$ and $\varphi = 90^\circ$ (again, in red and blue, respectively) for our 100 snapshots in the right panel of Fig. 2. Their incoherent time-average is also shown with bold solid lines. “Incoherent” here means that the averaging is performed at the level of the visibility amplitudes $|V(u, \varphi)|$, as in a realistic experiment [35], rather than at the level of the complex visibilities themselves, before taking the amplitude—these operations (averaging and taking amplitudes) do not commute for complex quantities. (Experimentally, the latter type of “coherent” time-averaging would require tracking the visibility phase over the course of all observations, which is beyond our near-term capabilities.) These time-averaged amplitudes can be directly compared to those of a radial profile; see, e.g., Figs. 4 and 5 of Ref. [35].

V. RAY TRACING REQUIREMENTS

As described in Sec. II F, we decompose images into layers labeled by photon half-orbit number n and ray trace each layer separately. In principle, one can choose any grid resolution in each lensing band. In practice, however, these resolutions must be sufficiently high to resolve the image features present in each layer. Otherwise, the output is not a faithful image of the source, and the resulting visibility is likewise inaccurate. In addition, the $n = 0$ lensing band occupies all of the image plane, whereas the $n = 0$ grid must of course have a finite size. Hence, the direct image of the source may need to be truncated, which could in turn introduce errors in the visibility. In this section, we describe the requirements on grid resolution and size that must be met in order to ensure that both the image and the visibility of a source are correctly computed. We also strive to explain the types of errors encountered when these requirements are not met.

We first discuss grid resolution and then grid size. A finite resolution erases small-scale features and a finite field of view cuts off large-scale features, but these effects can be remedied with the use of interpolation and extrapolation, respectively.

A. Resolution requirements

1. Requirements on the underlying simulation of the source

When computing images of some source model $I_s(\mathbf{x}_s)$, we must be able to sample the source intensity I_s at

arbitrary \mathbf{x}_s within the range of the simulation. This is because the transfer functions (33)–(35) vary smoothly across the image plane, as discussed in Sec. II G and as illustrated in Figs. 5 through 9.

For a stationary and axisymmetric source profile of the form (57), this requirement is trivial, since we typically specify the radial profile $J(r_s)$ analytically, as we did in Eq. (58) with Johnson’s SU distribution (which can be computed anywhere to arbitrary precision). On the other hand, if the source profile $I_s(\mathbf{x}_s)$ is not known exactly, then we must define it everywhere in the range where it is to be sampled via interpolation.

In that case, it is important that the interpolation be done on a dense grid, so that we do not miss any features of the source. In practice, this means that if its smallest features are of size $\sim w$, then before interpolating the source, we must first sample it on a grid with pixel spacing $\delta x \lesssim w$. Otherwise, if the grid spacing is some $\delta x \gtrsim w$, then we may miss a feature of size w in between two successive grid points x_i and $x_i + \delta x$, leading to an error in both the computed image and its visibility.

Moreover, the grid density must be high enough to resolve all the features in the source with many points, so as to ensure that the resulting interpolation is smooth: if some features are underresolved (for instance, if a bump in a radial profile has only one point underneath it), then interpolation will produce spurious sharp features in the image that will have a significant impact on its Fourier transform, and hence the visibility. Thus, $\delta x \lesssim w$ is not enough—we must in fact require that $\delta x \ll w$.¹⁵

For the GRFs that we introduced in Sec. III to model astrophysical fluctuations, this means that the spacing between grid points $\mathbf{x}_s = (t_s, x_s, y_s)$ must be smaller than the characteristic size of fluctuations, which is set by the correlation lengths λ_ℓ with $\ell \in \{0, 1, 2\}$. That is, we require a grid spacing $\delta x \ll \lambda_\ell$.

We reiterate that interpolation is unavoidable here because the INOISY simulation grid does not coincide (barring some incredible fine-tuning) with the equatorial crossings of rays traced backwards from points in the image-plane grids.

2. Requirements on the image grids used in ray tracing

We now describe the resolutions needed in the grids used to ray trace within each lensing band, as described in Sec. II F.

For the direct $n = 0$ image, we must ray trace on a grid with spacing $\delta x^{(0)} \lesssim \delta x^{\text{model}}$ for precisely the same reasons

¹⁵This smooth interpolation effectively defines the source down to arbitrarily small length scales by decreeing that no new features appear below size w , which is the minimal (most conservative) “ultraviolet completion” of the model. Any other choice would modify the visibility on baselines $u \gtrsim 1/w$ and should then be specified as part of the model itself, rather than by fiat.

as listed above for the underlying model—provided we do so, we will not miss any image features and they will all be well resolved.

Likewise, in the higher- n layers, we must ray trace on grids with exponentially fine spacings $\delta x^{(n)} \approx e^{-\gamma n} \delta x^{(0)}$. The reason is simply that if the smallest feature in the $n = 0$ image is of size $w_0 \approx w$, then its n th mirror image in the n th lensing band will be roughly of size $w_n \approx e^{-\gamma n} w_0$. To avoid missing these features (or underresolving them), we must demand that $\delta x^{(n)} \ll w_n$, a condition that is automatically satisfied with this adaptive grid scaling tailored to the Kerr lensing behavior.

Provided that each image layer is ray traced at a sufficiently high resolution to ensure that no subgrid structure is omitted, we can safely interpolate the image layers $I_n(\mathbf{x}_o)$. In turn, we can evaluate their Radon transforms along any desired slice, and then take a 1D Fourier transform to obtain their correctly computed visibilities $V_n(\mathbf{u})$, as described in Sec. IV C.

Conversely, we note that if the resolution is not increased in each successive layer, then eventually some image layer $I_n(\mathbf{x}_o)$ will have lensed features smaller than its grid spacing. This unresolved subgrid structure will render the visibility $V_n(\mathbf{u})$, and thus the full visibility $V(\mathbf{u})$, inaccurate. Hence, adaptive ray tracing is necessary to guarantee accurate visibilities.

3. Convergence tests

To summarize, for both the emission model $I_s(\mathbf{x}_s)$ and its ray-traced image layers $I_n(\mathbf{x}_o)$, we must use sufficiently high resolutions to ensure that we do not truncate any subgrid structure (that is, structure on length scales smaller than the resolution). This then allows us to safely interpolate $I_s(\mathbf{x}_s)$ and $I_n(\mathbf{x}_o)$, without missing any features or introducing any spurious new ones in the interpolation procedure.

Interpolating $I_s(\mathbf{x}_s)$ is needed to obtain the layers $I_n(\mathbf{x}_o)$, as ray tracing requires us to compute the intensity loaded on rays that intersect the equatorial source at generic crossing points. Interpolating the image layers $I_n(\mathbf{x}_o)$ is also needed to compute their Fourier transforms $V_n(\mathbf{u})$, and hence the visibility (119). Any errors in either of these steps introduce errors in the final visibility $V(\mathbf{u})$, so we must carefully avoid them by using the appropriate resolutions, as explained above.

In practice, however, these requirements are only heuristic. For instance, the Lyapunov exponent is an angle-dependent function $\gamma(\varphi)$ around the image plane, so we cannot precisely realize the exponential scaling $\delta x^{(n)} \sim e^{-\gamma n} \delta x^{(0)}$ on a Cartesian grid. In addition, $e^{\gamma(\varphi)}$ can be as large as ~ 20 , but such an increase in pixel density is not always necessary. In fact, recall that for the example in Sec. IV, we simply set $\delta x^{(n)} = 2^n \delta x^{(0)}$, a much more modest but still adequate scaling of pixel density.

Inversely, a realization of an INOISY source realization may occasionally make a field excursion that creates a far

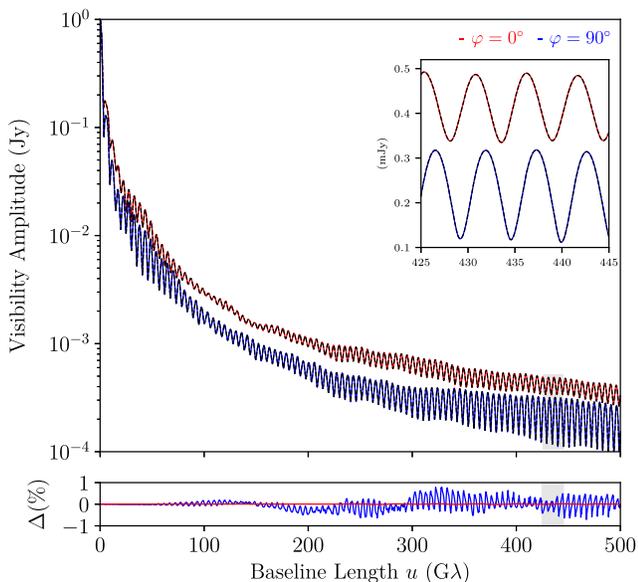


FIG. 14. Convergence test of the resolution needed to ray trace an INOISY snapshot. We plot its visibility amplitude at baseline angles $\varphi = 0^\circ$ (red lines) and $\varphi = 90^\circ$ (blue lines) computed using the grid resolutions described in the main text. The black dashed lines correspond to the visibility amplitudes obtained after doubling the nominal resolution. The inset panel zooms into the region where the fits were performed to infer the projected ring diameter as described in Sec. VI. Since the differences are consistently below 1%, we can validate the resolution.

smaller fluctuation than the minimum size w typically expected from its correlation matrix $\Lambda(\mathbf{x}_s)$. While statistically improbable, such events cannot be precluded from occurring, and could result in a too-large grid spacing $\delta x^{\text{model}} \gtrsim w$. Unfortunately, it is intractable to directly check that this does not happen.

Instead, it is best to numerically check for the absence of errors using a simple convergence test, in which one doubles all the resolutions (of the simulated profile and image grids) and computes again the observables (images and visibilities). In Fig. 14, we show an example of a convergence test for the example from Sec. IV: after halving the grid resolutions, the changes in the visibility amplitude remain small, confirming that the resolution scaling $\delta x^{(n)} = 2^n \delta x^{(0)}$ is adequate.

B. Field of view requirements

Being able to completely resolve an emission profile $I_s(\mathbf{x}_s)$ and a (central) part of its image $I_o(\mathbf{x}_o)$ is a necessary but not sufficient condition to correctly compute its visibility $V(\mathbf{u})$.

Another requirement for the accurate computation of a source model’s visibility is related to the size of its image’s field of view (FOV) and how fast the emission profile decays past it. Even when all the photon rings are fully resolved (that is, the resolution in each lensing band is

sufficiently high), it is still imperative to check that the FOV is sufficiently large for the image to include all of the features with significant flux.

If the FOV is too small and cuts out a portion of the source, then this truncation may create a sharp drop in the observed intensity at the edges of the image. Such an artificial transition would effectively introduce high-frequency components to the image, polluting its Fourier transform, and hence its visibility. This is the well-known Gibbs phenomenon.

Since we are interested in narrow features of the image, and in particular the interferometric signature of its $n = 2$ photon ring that is encoded in the visibility on very long baselines, we cannot avoid dealing with this potential effect.

1. Field of view of the direct image

We first focus on the direct image, as it is the most prone to the Gibbs phenomenon. This is because it occupies the $n = 0$ lensing band, which has noncompact support: as described in Sec. II E, it fills the entire image plane (except for the interior of the apparent equatorial horizon).

If the emission profile has compact support in the equatorial plane (r_s, ϕ_s) , however, its direct image $I_0(\mathbf{x}_o)$ will also have compact support within the $n = 0$ layer, and we can terminate our FOV just past the edge of this image without introducing any truncation errors or triggering the Gibbs phenomenon.

On the other hand, if the emission profile extends infinitely far in the equatorial plane (remaining nonzero as $r_s \rightarrow \infty$), or else if its support is too large to fully ray trace over in practice, then we must necessarily cut off its $n = 0$ image.¹⁶

For instance, consider the simple, stationary, axisymmetric source (57) with radial profile $J(r_s)$ given by Johnson’s SU distribution (58). While this emission profile does decay very fast as $r_s \rightarrow \infty$, it always remains finite. In Fig. 15, we plot it in logarithmic scale for two sets of values of its parameters. From now on, we focus on the red profile, with parameters $\mu = r_-$, $\vartheta = M/2$, and $\gamma = -3/2$. We first ray trace its direct $n = 0$ image on a grid of size $400M$, and then we compute the corresponding visibility amplitude $|V_0(u, 0^\circ)|$, which we plot as a black dashed curve in the top left panel of Fig. 16.

At the edge of our FOV, $J(r_s)$ has dropped to $\lesssim 10^{-11}$ times its maximum intensity. As such, the sudden drop introduced by our truncation is too small to produce a significant ringing effect on long baselines, and so it cannot meaningfully affect the signal. We will consider this large FOV and its associated visibility amplitude as the “correct” one. Next, we ray trace the same image with its FOV

¹⁶We saw in Sec. VA that a finite grid resolution introduced an “ultraviolet” cutoff; now we see a finite grid size as the corresponding “infrared” cutoff.

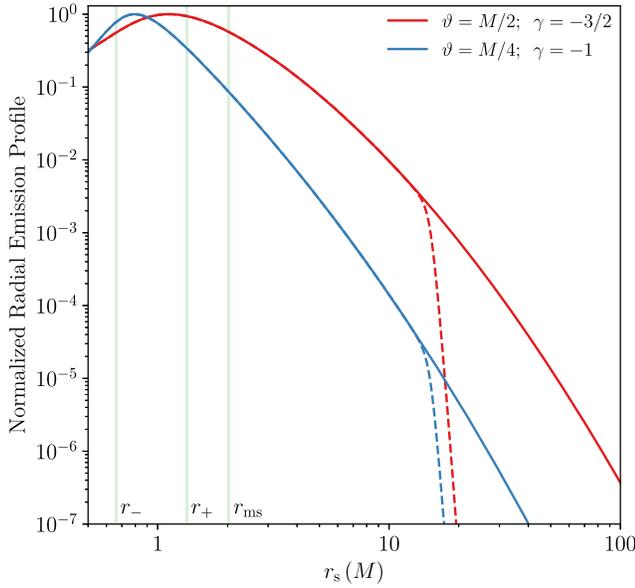


FIG. 15. Radial emission profile $J_{\text{SU}}(r; \mu, \vartheta, \gamma)$ [Eq. (58)] with $\mu = r_-$ to ensure that the profile peaks past the outer horizon at $r = r_+$. The values of the other parameters ϑ and γ are shown in the legend. The dashed lines depict the profile multiplied by the cutoff function (123) with $s_- = M$ and $r_- = 15M$. From left to right, the vertical lines indicate the inner and outer horizons $r = r_{\pm}$ and the ISCO radius r_{ms} .

truncated at $r_s \approx 15M$, a cutoff past which $J(r_s)$ drops to $\lesssim 10^{-4}$ times its maximum intensity. We then plot the resulting visibility amplitude as the solid blue curve in the top left panel of Fig. 16. Perhaps surprisingly, it differs quite significantly from the “correct” one. In fact, their relative difference—which we plot as a blue curve in the small panel below the main—shows a large and growing percentage error that exceeds 100% before even reaching $100 G\lambda$. Even though the total flux that we ignored is tiny, the sudden drop in intensity at the FOV’s edge triggered the Gibbs phenomenon.

The resulting error can affect the interferometric signature of the $n = 2$ photon ring and is therefore an obstacle that we need to overcome, especially since this very profile is roughly consistent with the 2017 EHT observations of M87* [35].

There are two approaches to mitigating this problem. The first is to increase the FOV until the profile decays enough at the edges that it “effectively vanishes”—in that case, these spurious ringing effects will not be noticeable. In effect, this is what we did when we chose an FOV of $400M$. However, it may not always be feasible to extend the FOV this much. For instance, if the emission profile is an INOISY source simulated on a grid that stretches out to some maximal r_s , then extending the source past this cutoff would require extrapolation. This is already subtle for smooth profiles, as it amounts to introducing *ad hoc* data to the model.

The second approach—referred to as “apodization”—has already been used to deal with this problem in Ref. [10]. This method involves the multiplication of the image by a suitable window (also known as tapering or apodization function) that smoothly tapers the intensity to precisely zero before the edge of the FOV. For our example, we will use the window function

$$C(s_-, r_-) = \frac{1 - \tanh[s_-(r - r_-)]}{2}, \quad (123)$$

with a cutoff value of $r_- = 15M$ (well within the “small” FOV of $50M$) and $s_- = M$, which results in the smooth but rapidly truncated profile depicted with a dashed red line in Fig. 15.

Ray tracing the direct image of this profile and computing its visibility amplitude now results in the solid red curve that we plot in the top left panel of Fig. 16. Its difference relative to the “correct” amplitude is much smaller, and the percentage difference remains fairly constant well past $100 G\lambda$. This can therefore be an effective method.

In practice, we prefer to always use emission profiles that smoothly terminate within the FOV of our direct image, which we can easily take to be $50M$ wide. In case our source profile extends beyond this FOV, we apply the window function (123) to it and take the resulting profile as our “true” source.

2. Field of view of higher layers

Since the $n > 0$ lensing bands all have compact support, it is straightforward to avoid the Gibbs phenomenon in higher- n image layers and visibilities: we just take a FOV that includes the totality of the n th photon ring (that is, all the flux in the n th image layer).

We briefly describe the effects that windowing the emission profile as discussed above has on higher visibility amplitudes. The top right panel of Fig. 16 shows that the windowed profile produces roughly the same visibility (solid red curve) as the original one (dashed black curve) up to baselines $u \sim 100 G\lambda$, which are necessary to resolve image features of size

$$100 G\lambda = \frac{10^{11}}{\text{rad}} \approx \frac{1}{2 \mu\text{as}}. \quad (124)$$

For M87*, whose photon ring has a diameter $\sim 10M \approx 40 \mu\text{as}$, this corresponds to a size of $0.5M$, or half the $n = 1$ ring width.

Thus, differences in the visibility amplitude $|V_1(u, 0^\circ)|$ only become apparent on baselines that start to resolve its width, as can be seen in the small plot of percentage deviation. On the other hand, truncating the FOV of the $n = 1$ image from $400M$ to $50M$ has no discernible effect on its visibility amplitude (solid blue curve) and results in a vanishingly small deviation.

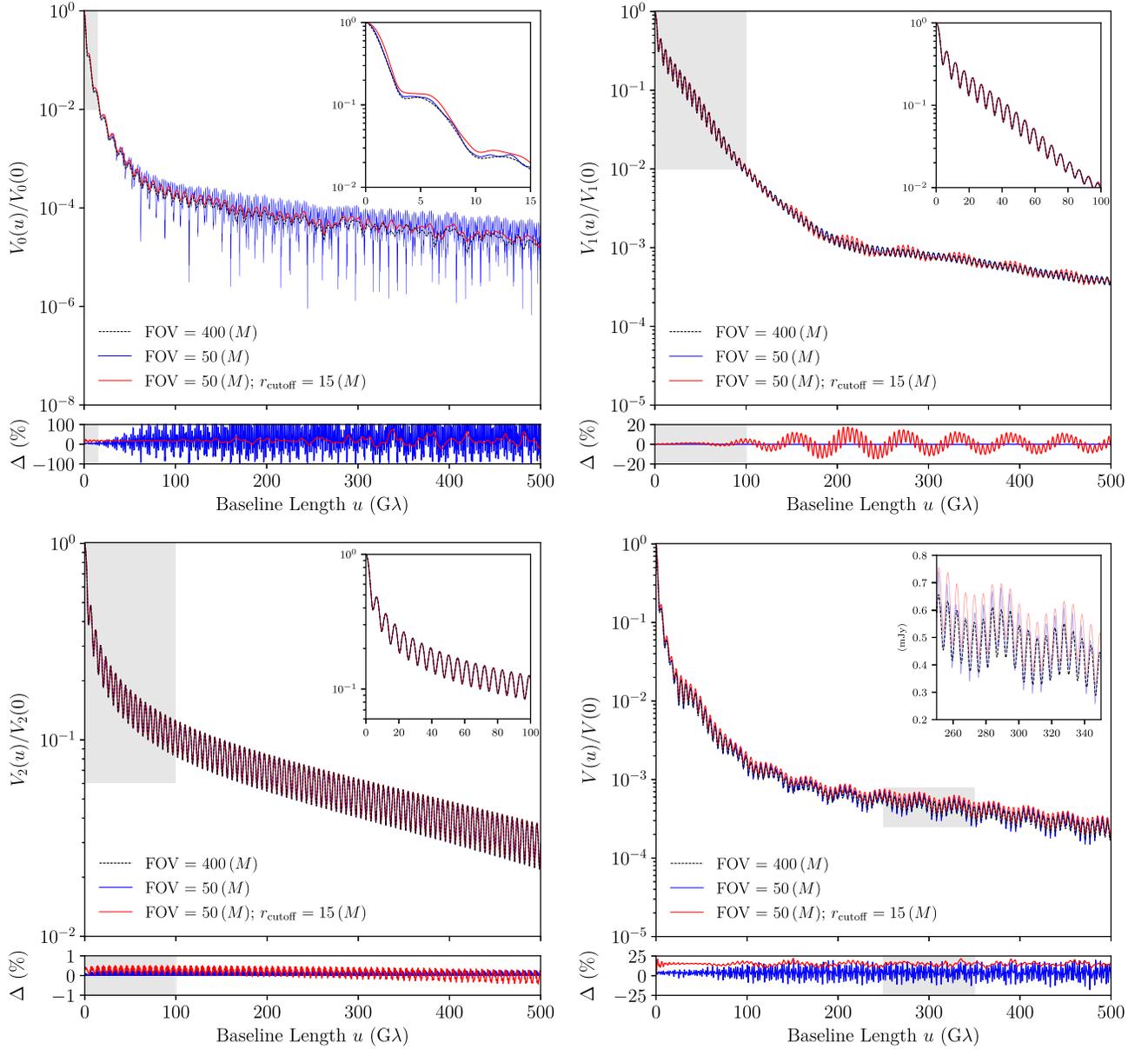


FIG. 16. Normalized visibility amplitudes $|V_n(u, 0^\circ)|$ corresponding to the image layers $I_n(\alpha, \beta)$ produced by the red radial emission profile in Fig. 15 (Johnson’s SU distribution with $\mu = r_-, \vartheta = M/2$, and $\gamma = -3/2$), for $n = 0$ (top left), $n = 1$ (top right), and $n = 2$ (bottom left). Bottom right: Normalized visibility amplitude $|V(u, 0^\circ)|$ corresponding to the full image $I_o(\alpha, \beta)$ obtained by summing over the preceding three layers. In each quadrant, the main panel shows the visibility amplitudes obtained using different values for the field of view (FOV): dashed black for a large $400M$ FOV, blue for a smaller $50M$ FOV, and red for the $50M$ FOV with a cutoff applied [Eq. (123) with $s_- = M$ and $r_- = 15M$]. Inset panels zoom into the shaded regions of each main plot to highlight the differences between the visibility amplitudes, while the smaller panels beneath the main ones plot percentage differences relative to the amplitude with large $400M$ FOV. In all these examples, we have assumed that the flow follows the Cunningham prescription [67], which in our notation corresponds to the Keplerian circular motion described in Sec. B 1.

Finally, we can repeat this exercise for the $n = 2$ ring. Again, windowing the profile has a negligible effect of $\lesssim 1\%$, as does the truncation. In fact, the latter deviation in this case ought to vanish entirely and is only nonzero due to numerical error introduced by the discretization of the Fourier transform: extending the FOV of the image to $400M$

by “zero padding” changes the sampling of the fast Fourier transform, resulting in a minute effect.

We note that these tiny errors are essentially negligible in the full visibilities, which we compare in the bottom right panel of Fig. 16, where most of the percentage error can be attributed to the direct image, as expected.

VI. MEASURING THE PHOTON RING SHAPE

Using black hole imaging to test general relativity in a regime where gravity is strong and yet nondynamical poses significant challenges: one must disentangle the astrophysical effects of the radiating source from the purely gravitational ones [26,29]. A promising approach is to focus on the $n = 2$ photon ring, whose interferometric signature is expected to dominate the time-averaged radio visibility on long baselines.

More precisely, Ref. [35] proposed a test of strong-field GR based on a shape measurement of the $n = 2$ photon ring around M87*, which has been recently reviewed in Ref. [10].

Here, we show in the context of our example from Sec. IV how AART may be used to simulate this test using a source that includes “realistic” astrophysical fluctuations (generated with INOISY, as described in Sec. III). We offer a brief overview of the key points in the test, and refer the reader to Refs. [10,35] for more details.

In the regime (120), a ring of projected diameter d_φ has a visibility that takes the universal form (121). Therefore, its amplitude can be well approximated in this regime by [33,34]

$$|V(u, \varphi)| = \sqrt{(\alpha_\varphi^L)^2 + (\alpha_\varphi^R)^2 + 2\alpha_\varphi^L \alpha_\varphi^R \sin(2\pi d_\varphi u)}, \quad (125)$$

where the functions

$$\alpha_\varphi^{L/R}(u) = \frac{e_{\text{upper}}(u) \pm e_{\text{lower}}(u)}{2} \quad (126)$$

encode the intensity profile around the ring image, while $e_{\text{upper/lower}}(u)$ are the upper/lower envelopes of the visibility amplitude, respectively.

The width w and diameter d of the $n = 1$ ring do not always satisfy the condition $w/d \ll 1$ needed for the universal regime (120) to exist, but those of the $n = 2$ ring do. Hence, we can fit Eq. (125) to the visibility amplitudes in the regime (122) dominated by the $n = 2$ ring after numerically computing the functions $\alpha_\varphi^{L/R}(u)$ (a simple cubic interpolation is enough).

The inset panel in Fig. 2 shows the visibility amplitude for two baseline angles $\varphi = 0^\circ$ and $\varphi = 90^\circ$ in the baseline range $u \in [425, 445]$ G λ (corresponding to an Earth-Moon baseline with an operation band of 332–344 GHz) with their best fit to Eq. (125) overplotted in black dashed lines.

According to GR, the shape of the $n = 2$ photon ring is the sum of a circle, with radius R_0 , and an ellipse centered at the origin, with radii R_1 and R_2 [35]. This “circlipse” shape has a projected diameter given by [34]

$$\frac{d_\varphi}{2} = R_0 + \sqrt{R_1^2 \sin^2(\varphi - \varphi_0) + R_2^2 \cos^2(\varphi - \varphi_0)}, \quad (127)$$

where φ_0 is an offset angle to account for the orientation of the ring within the image plane.

TABLE III. Best-fit values for the projected diameter d_φ [Eq. (127)] obtained after averaging N snapshots. The resulting fits are shown in Fig. 3. The normalized residuals RMSE are root-mean-square errors divided by the average value of d_φ .

| N | R_0 | R_1 | R_2 | φ_0 | RMSE |
|----------|--------|-------|-------|-------------|----------------------|
| 5 | 37.045 | 1.289 | 0.879 | 1.319° | 4.5×10^{-4} |
| 10 | 36.250 | 2.077 | 1.679 | 1.308° | 3.2×10^{-4} |
| 20 | 36.909 | 1.419 | 1.027 | 0.987° | 2.4×10^{-4} |
| 100 | 36.052 | 2.283 | 1.894 | 0.426° | 1.1×10^{-4} |
| ∞ | 36.115 | 2.219 | 1.830 | 0.202° | 4.2×10^{-5} |

Using the statistical model described in Sec. III for the emission profile with the parameters shown in Table II, we (i) average the visibility amplitudes of N snapshot images; (ii) fit Eq. (125) to the visibility amplitude for 36 angles $\varphi = \{0^\circ, 5^\circ, \dots, 175^\circ\}$ to obtain the projected diameters d_φ ; and (iii) fit the general GR prediction given in Eq. (127) to obtain $\{R_0, R_1, R_2, \varphi_0\}$. The best global fits of Eqs. (125) and (127) were found using a Markov chain Monte Carlo (MCMC) method.¹⁷

We report the best-fit values of the parameters in Eq. (127) and their normalized root-mean-square error (RMSE) for each N in Table III. The resulting curves, shown in Fig. 3, follow the GR prediction very closely, even when only five snapshots are (incoherently) averaged. The lower panels in Fig. 3 show the difference of the best-fit curve for each case with the best-fit curve for the purely radial profile consisting of the envelope of the emission model (that is, when there are no fluctuations).

The results of this example are very encouraging for future missions targeting measurements of the photon ring shape: with only a few snapshots, one may already start to check whether the data follow the GR prediction. A systematic study including instrumental noise is required to simulate a realistic experimental forecast—we leave this for future work.

VII. FUTURE OUTLOOK

Here, we have presented AART: a new, publicly available, numerical code designed for precision studies of a black hole’s photon rings. AART exploits the integrability of null geodesics in the Kerr geometry to ray trace images analytically, with no loss of numerical precision even for strongly lensed photons that orbit the black hole multiple times. The code decomposes the image plane into layers—lensing bands—with increasing grid resolution adapted to the lensing behavior of the hole.

¹⁷Since we are neglecting the phase of the oscillation, Eq. (125) allows for several values of d_φ separated by $\sim 1/u$ to provide a good fit [35]. However, the global maximum is not always the actual diameter d_φ of the ring as measured directly in the image domain. As explained in Ref. [10], in those cases, one must infer the diameter by fitting at multiple angles.

The modular structure of AART can accommodate any time-dependent and nonaxisymmetric equatorial emission profile, and the code could be extended to nonequatorial, anisotropic sources. Also, the components described herein can be used separately: AART is designed to individually export the lensing bands, critical curve, apparent horizon, and redshift factors, which can therefore serve as input for other studies.

As a prime application, we showed how the experimental forecast for the test of general relativity proposed in Ref. [35] could be further refined by including source fluctuations. We used INOISY [43] to simulate a stochastic model of equatorial emission, and we produced high-resolution synthetic images together with their corresponding visibility on long baselines. We then successfully extracted the GR-predicted shape for the projected diameter of the $n = 2$ ring from its interferometric signature. In a follow-up paper, we will use this framework to carry out a parameter estimation survey that also includes instrument response and noise—this will further validate the test proposed in Ref. [35] and bears relevance to SALTUS and other space-VLBI missions targeting the photon ring [81,82].

Although we only studied the Kerr geometry, our approach may serve as a useful guide to studying other theories. For instance, the Kerr lensing bands may offer a starting point for numerically finding those of slightly deformed spacetimes.

Finally, recent EHT observations of Sgr A* found that only $\sim 3.5\%$ of the EHT GRMHD models passed the light-curve variability constraint [22]. Given the uncertainty associated with the variability excess, and the possibility of missing physical ingredients in current astrophysical models, efforts to develop astrophysics-agnostic phenomenological approaches such as the one presented in this work are just as valuable as the improvement of accretion disk simulations.

ACKNOWLEDGMENTS

We thank Charles Gammie, Delilah Gates, Samuel Gralla, Aaron Held, Daeyoung Lee, Hadrien Pagnat, Eliot Quataert, Leo Stein, Frédéric Vincent, Maciek Wielgus, and George Wong for their useful comments. This work used resources provided by Princeton Research Computing, a consortium that includes PICSciE (the Princeton Institute for Computational Science and Engineering) as well as the Office of Information Technology's Research Computing division. A. C. A. and A. L. gratefully acknowledge support from Will and Kacie Snellings. A. C. A. also acknowledges support from the Simons Foundation.

APPENDIX A: EXPLICIT FORM OF NULL GEODESICS IN THE KERR EXTERIOR

This appendix lists all the formulas needed to compute the observational appearance of Kerr equatorial sources

in the sky of a distant observer. Readers are referred to Refs. [5,40] for derivations and further details. Throughout, $F(\varphi|k)$, $E(\varphi|k)$, and $\Pi(n;\varphi|k)$ denote the incomplete elliptic integrals of the first, second, and third kinds, respectively, defined according to the conventions listed in Appendix A of Ref. [46]. $K(k) \equiv F(\pi/2|k)$ denotes the complete integral of the first kind and we also let $E'(\varphi|k) \equiv \partial_k E(\varphi|k) = [E(\varphi|k) - F(\varphi|k)]/(2k)$.

1. Angular geodesic integrals

The angular motion of a Kerr photon can display two qualitatively different behaviors depending on the sign of its energy-rescaled Carter constant η . Since we only consider sources in the Kerr equatorial plane, we may ignore vortical motion with $\eta < 0$, which can never reach the equator [46]. We thus restrict our attention to ordinary motion with $\eta > 0$, in which case the angular geodesic integrals in Sec. II are [5]

$$G_\theta^{(n)} = \frac{1}{a\sqrt{-u_-}} [2m(n)K \mp_\circ F_\circ], \quad (\text{A1})$$

$$G_\phi^{(n)} = \frac{1}{a\sqrt{-u_-}} [2m(n)\Pi \mp_\circ \Pi_\circ], \quad (\text{A2})$$

$$G_t^{(n)} = -\frac{2u_+}{a\sqrt{-u_-}} [2m(n)E' \mp_\circ E'_\circ], \quad (\text{A3})$$

where $\pm_\circ = \text{sign}\beta$, $m(n) \equiv n + H(\beta)$ counts the number of angular turning points encountered along the trajectory, $H(x)$ denotes the Heaviside function, and we also introduced

$$K = K\left(\frac{u_+}{u_-}\right) = F\left(\frac{\pi}{2} \middle| \frac{u_+}{u_-}\right), \quad (\text{A4})$$

$$F_\circ = F\left(\arcsin\left(\frac{\cos\theta_\circ}{\sqrt{u_+}}\right) \middle| \frac{u_+}{u_-}\right), \quad (\text{A5})$$

$$E'_\circ = E'\left(\arcsin\left(\frac{\cos\theta_\circ}{\sqrt{u_+}}\right) \middle| \frac{u_+}{u_-}\right), \quad (\text{A6})$$

$$\Pi_\circ = \Pi\left(u_+; \arcsin\left(\frac{\cos\theta_\circ}{\sqrt{u_+}}\right) \middle| \frac{u_+}{u_-}\right). \quad (\text{A7})$$

2. Radial geodesic integrals

The radial integrals can be decomposed into [40]

$$I_\phi = \frac{2Ma}{r_+ - r_-} \left[\left(r_+ - \frac{a\lambda}{2M}\right) I_+ - \left(r_- - \frac{a\lambda}{2M}\right) I_- \right], \quad (\text{A8})$$

$$I_t = \frac{(2M)^2}{r_+ - r_-} \left[r_+ \left(r_+ - \frac{a\lambda}{2M}\right) I_+ - r_- \left(r_- - \frac{a\lambda}{2M}\right) I_- \right] + (2M)^2 I_0 + (2M) I_1 + I_2, \quad (\text{A9})$$

with the integrals I_0 , I_1 , I_2 , and I_{\pm} reducible to Legendre form.

Their precise form depends on the nature of the radial roots $\{r_1, r_2, r_3, r_4\}$: there are four different cases for null geodesics in the Kerr exterior, but only two of these—labeled type (2) and type (3)—arise for light rays that connect the equatorial plane to a distant observer; meanwhile, type (4) geodesics can reach distant observers but never the equator, since they are all vortical, whereas type (1) geodesics are all bound to the black hole and cannot reach a distant observer at infinity [40].

a. Type (2)

In this case, all the roots are real and the motion is restricted to $r \geq r_4 > r_3 > r_2 > r_1$.

The antiderivatives of the radial geodesic integrals take the manifestly real and smooth forms [40]

$$\mathcal{I}_0 = F^{(2)}(r), \quad (\text{A10})$$

$$\mathcal{I}_1 = r_3 F^{(2)}(r) + r_{43} \Pi_1^{(2)}(r), \quad (\text{A11})$$

$$\mathcal{I}_2 = \frac{\sqrt{\mathcal{R}(r)}}{r - r_3} - \frac{r_1 r_4 + r_2 r_3}{2} F^{(2)}(r) - E^{(2)}(r), \quad (\text{A12})$$

$$\mathcal{I}_{\pm} = -\Pi_{\pm}^{(2)}(r) - \frac{F^{(2)}(r)}{r_{\pm 3}}, \quad (\text{A13})$$

with (recall that $r_{ij} = r_i - r_j$)

$$F^{(2)}(r) = \frac{2}{\sqrt{r_{31} r_{42}}} F(\arcsin x_2(r) | k_2) \geq 0, \quad (\text{A14})$$

$$E^{(2)}(r) = \sqrt{r_{31} r_{42}} E(\arcsin x_2(r) | k_2) \geq 0, \quad (\text{A15})$$

$$\Pi_1^{(2)}(r) = \frac{2}{\sqrt{r_{31} r_{42}}} \Pi\left(\frac{r_{41}}{r_{31}}; \arcsin x_2(r) | k_2\right) \geq 0, \quad (\text{A16})$$

$$\Pi_{\pm}^{(2)}(r) = \frac{2r_{43}}{r_{\pm 3} r_{\pm 4} \sqrt{r_{31} r_{42}}} \Pi\left(\frac{r_{\pm 3} r_{41}}{r_{\pm 4} r_{31}}; \arcsin x_2(r) | k_2\right), \quad (\text{A17})$$

where the auxiliary function $x_2(r)$ is defined as

$$x_2(r) = \sqrt{\frac{r - r_4 r_{31}}{r - r_3 r_{41}}} \in \left[0, \sqrt{\frac{r_{31}}{r_{41}}}\right] \subset [0, 1), \quad (\text{A18})$$

while the elliptic modulus is

$$k_2 = \frac{r_{32} r_{41}}{r_{31} r_{42}} \in (0, 1). \quad (\text{A19})$$

Finally, the source radius is given in terms of the Jacobi elliptic sine function $sn(\varphi|k)$ by

$$r_s^{(2)}(\tau) = \frac{r_4 r_{31} - r_3 r_{41} \text{sn}^2(X_2(\tau) | k_2)}{r_{31} - r_{41} \text{sn}^2(X_2(\tau) | k_2)}, \quad (\text{A20})$$

$$X_2(\tau) = \frac{1}{2} \sqrt{r_{31} r_{42}} \tau - F(\arcsin x_2(r_0) | k_2). \quad (\text{A21})$$

b. Type (3)

In this case, only r_1 and r_2 are real roots while $r_3 = \bar{r}_4$ are complex-conjugate roots, and the range of radial motion is $r \geq r_+ > r_- > r_2 > r_1$.

The antiderivatives of the radial geodesic integrals take the manifestly real and smooth forms [40]

$$\mathcal{I}_0 = F^{(3)}(r), \quad (\text{A22})$$

$$\mathcal{I}_1 = \left(\frac{Ar_1 + Br_2}{A + B}\right) F^{(3)}(r) + \Pi_1^{(3)}(r), \quad (\text{A23})$$

$$\mathcal{I}_2 = \left(\frac{Ar_1 + Br_2}{A + B}\right)^2 F^{(3)}(r) + 2\left(\frac{Ar_1 + Br_2}{A + B}\right) \Pi_1^{(3)}(r) + \sqrt{AB} \Pi_2^{(3)}(r), \quad (\text{A24})$$

$$\mathcal{I}_{\pm} = -\frac{A + B}{Ar_{\pm 1} + Br_{\pm 2}} F^{(3)}(r) + \frac{2r_{21} \sqrt{AB}}{(Ar_{\pm 1})^2 - (Br_{\pm 2})^2} R_1(\alpha_{\pm}; \arccos x_3(r) | k_3), \quad (\text{A25})$$

with (recall that $r_{ij} = r_i - r_j$)

$$F^{(3)}(r) = \frac{1}{\sqrt{AB}} F(\arccos x_3(r) | k_3), \quad (\text{A26})$$

$$\Pi_{\ell}^{(3)}(r) = \left(\frac{2r_{21} \sqrt{AB}}{B^2 - A^2}\right)^{\ell} R_{\ell}(\alpha_0; \arccos x_3(r) | k_3), \quad (\text{A27})$$

$$R_1(\alpha; \varphi | j) = \frac{1}{1 - \alpha^2} \left[\Pi\left(\frac{\alpha^2}{\alpha^2 - 1}; \varphi | j\right) - \alpha f_1 \right], \quad (\text{A28})$$

$$R_2(\alpha; \varphi | j) = \frac{1}{\alpha^2 - 1} \left[F(\varphi | j) - \frac{\alpha^2 E(\varphi | j)}{j + (1 - j)\alpha^2} \right] + \frac{\alpha^3 \sin \varphi \sqrt{1 - j \sin^2 \varphi}}{(\alpha^2 - 1)[j + (1 - j)\alpha^2](1 + \alpha \cos \varphi)} + (2j - \frac{\alpha^2}{\alpha^2 - 1}) \frac{R_1(\alpha; \varphi | j)}{j + (1 - j)\alpha^2}, \quad (\text{A29})$$

$$f_1 = \frac{p_1}{2} \log \left| \frac{p_1 \sqrt{1 - j \sin^2 \varphi} + \sin \varphi}{p_1 \sqrt{1 - j \sin^2 \varphi} - \sin \varphi} \right|, \quad (\text{A30})$$

$$p_1 = \sqrt{\frac{\alpha^2 - 1}{j + (1 - j)\alpha^2}}, \quad (\text{A31})$$

where the auxiliary function $x_3(r)$ is defined as

$$x_3(r) = \frac{A(r-r_1) - B(r-r_2)}{A(r-r_1) + B(r-r_2)} \in \left(-\frac{1}{\alpha_0}, 1\right) \subset (-1, 1), \quad (\text{A32})$$

while the elliptic modulus is

$$k_3 = \frac{(A+B)^2 - r_{21}^2}{4AB} \in (0, 1), \quad (\text{A33})$$

and the other parameters are

$$\alpha_0 = \frac{B+A}{B-A} > 1, \quad \alpha_{\pm} = \frac{Br_{\pm 2} + Ar_{\pm 1}}{Br_{\pm 2} - Ar_{\pm 1}} = -\frac{1}{x_3(r_{\pm})}, \quad (\text{A34})$$

$$A = \sqrt{r_{32}r_{42}} > 0, \quad B = \sqrt{r_{31}r_{41}} > 0. \quad (\text{A35})$$

Finally, the source radius is given in terms of the Jacobi elliptic cosine function $\text{cn}(\varphi|k)$ by

$$r_s^{(3)}(\tau) = \frac{(Ar_1 - Br_2) - (Ar_1 + Br_2)\text{cn}(X_3(\tau)|k_3)}{(A-B) - (A+B)\text{cn}(X_3(\tau)|k_3)}, \quad (\text{A36})$$

$$X_3(\tau) = \sqrt{AB}\tau - F(\arccos x_3(r_0)|k_3). \quad (\text{A37})$$

APPENDIX B: ACCRETION FLOW FOUR-VELOCITIES

Let us consider an equatorial four-velocity

$$u = u^t \partial_t + u^r \partial_r + u^\phi \partial_\phi = u_t dt + u_r dr + u_\phi d\phi, \quad (\text{B1})$$

with u^μ and u_μ the contravariant and covariant components, respectively. The angular and radial-infall velocities are

$$\Omega = \frac{u^\phi}{u^t}, \quad \iota = -\frac{u^r}{u^t}. \quad (\text{B2})$$

Meanwhile, the energy and specific angular momentum are

$$\mathcal{E} = -u_t, \quad \ell = \frac{u_\phi}{\mathcal{E}}. \quad (\text{B3})$$

For geodesic motion, these quantities are conserved, whereas

$$\nu = \frac{u_r}{u_t} \quad (\text{B4})$$

is not. These variables parametrize the four-velocity as

$$u = u^t(\partial_t - \iota \partial_r + \Omega \partial_\phi) = \mathcal{E}(-dt - \nu dr + \ell d\phi). \quad (\text{B5})$$

Any two components of u determine it entirely, as the third can be recovered from the normalization condition $u \cdot u = -1$. We will parametrize the four-velocity using either the pair (ι, Ω) , the pair (ν, ℓ) , or the pair (\mathcal{E}, ℓ) , as follows. First, define

$$\Pi(r) = r^2 g_{\phi\phi}|_{\theta=\pi/2} = (r^2 + a^2)^2 - a^2 \Delta(r). \quad (\text{B6})$$

Given (ι, Ω) , unit-normalization fixes

$$u^t = \left[1 - (r^2 + a^2)\Omega^2 - \frac{2M}{r}(1 - a\Omega)^2 - \frac{r^2}{\Delta(r)} \iota^2 \right]^{-1/2}, \quad (\text{B7})$$

which is physically admissible provided the quantity in square brackets is positive. We took the positive square root to ensure $u^t > 0$ is future-directed. Lowering the index of u^μ yields

$$\mathcal{E} = \chi u^t, \quad \nu = \frac{r^2 \iota}{\Delta(r) \chi}, \quad \ell = \frac{\Pi(r)\Omega - 2aMr}{r^2 \chi}, \quad (\text{B8a})$$

$$\chi = 1 - \frac{2M}{r}(1 - a\Omega). \quad (\text{B8b})$$

Conversely, given (ν, ℓ) , unit-normalization fixes the energy to

$$\mathcal{E} = \sqrt{\frac{\Delta(r)}{\frac{\Pi(r)}{r^2} - \frac{4aM\ell}{r} - (1 - \frac{2M}{r})\ell^2 - [\frac{\Delta(r)}{r}\nu]^2}}, \quad (\text{B9})$$

which is physically admissible provided the quantity under the square root is positive. Raising the index of u_μ then yields

$$u^t = \frac{\mathcal{E}}{\chi}, \quad \iota = \frac{\Delta(r)}{r^2} \chi \nu, \quad \Omega = \frac{\chi}{r^2} (\ell + aH), \quad (\text{B10a})$$

$$\chi = \frac{1}{1 + \frac{2M}{r}(1 + H)}, \quad H = \frac{2Mr - a\ell}{\Delta(r)}. \quad (\text{B10b})$$

It is often convenient to express Ω and χ in terms of ℓ only as

$$\Omega = \frac{a + (1 - \frac{2M}{r})(\ell - a)}{\frac{\Pi(r)}{r^2} - \frac{2aM\ell}{r}}, \quad \chi = \frac{\Delta(r)}{\frac{\Pi(r)}{r^2} - \frac{2aM\ell}{r}}. \quad (\text{B11})$$

Last, given (\mathcal{E}, ℓ) , the full four-velocity can be recovered by solving the normalization condition $u \cdot u = -1$ [or equivalently, Eq. (B9)] for

$$\nu = \frac{r}{\Delta(r)} \sqrt{\frac{\Pi(r)}{r^2} - \frac{4aM\ell}{r} - \left(1 - \frac{2M}{r}\right)\ell^2 - \frac{\Delta(r)}{\mathcal{E}^2}}. \quad (\text{B12})$$

Finally, the observed redshift $g(\iota, \Omega)$ is

$$g = \frac{E}{-P_\mu u^\mu} = \frac{1}{u^t \left(1 \pm_r \frac{\sqrt{R(r)}}{\Delta(r)} \iota - \lambda \Omega\right)}. \quad (\text{B13})$$

Alternatively, the redshift can also be expressed in terms of (\mathcal{E}, ν, ℓ) as [Eq. (B10)]

$$g = \frac{\Delta(r)}{\mathcal{E} \left[G \pm_r \sqrt{\mathcal{R}(r)} \frac{\Delta(r)}{r^2} \nu \right]}, \quad (\text{B14a})$$

$$G = \frac{\Pi(r)}{r^2} - \left(1 - \frac{2M}{r} \right) \ell \lambda - \frac{2aM}{r} (\ell + \lambda). \quad (\text{B14b})$$

1. Keplerian circular orbits (geodesic motion)

We let $u = \dot{u}$ denote the four-velocity of timelike, circular-equatorial geodesics. These orbits define Keplerian motion in the Kerr spacetime; they are only stable if [Eq. (2.20) of [56]]

$$r^2 - 6Mr + 8a\sqrt{Mr} - 3a^2 \geq 0, \quad (\text{B15})$$

or equivalently, if $r \geq r_{\text{ms}}$, where r_{ms} denotes the radius of the (marginally stable) innermost stable circular orbit (ISCO),

$$r_{\text{ms}} = M[3 + Z_2 - \sqrt{(3 - Z_1)(3 + Z_1 + 2Z_2)}], \quad (\text{B16a})$$

$$Z_1 = 1 + \sqrt[3]{1 - a_*^2 [\sqrt[3]{1 + a_*} + \sqrt[3]{1 - a_*}]}, \quad (\text{B16b})$$

$$Z_2 = \sqrt{3a_*^2 + Z_1^2}, \quad a_* = \frac{a}{M}. \quad (\text{B16c})$$

Inside the ISCO, we must use an inspiraling geodesic motion, as orbits are unstable and particles must fall into the horizon.

a. Outside the ISCO radius

For $r \geq r_{\text{ms}}$, we can set $\dot{r} = 0$ and the Kerr geodesic equation determines the Keplerian angular velocity [Eq. (2.16) of [56]]:

$$\dot{\Omega} = \frac{\sqrt{M}}{r^{3/2} + a\sqrt{M}}. \quad (\text{B17})$$

The contravariant four-velocity is then [Eqs. (B5) and (B9)]

$$\dot{u}^t = \frac{r^{3/2} + a\sqrt{M}}{\sqrt{r^3 - 3Mr^2 + 2a\sqrt{M}r^{3/2}}}, \quad (\text{B18a})$$

$$\dot{u}^r = 0, \quad (\text{B18b})$$

$$\dot{u}^\phi = \frac{\sqrt{M}}{\sqrt{r^3 - 3Mr^2 + 2a\sqrt{M}r^{3/2}}}, \quad (\text{B18c})$$

and the Keplerian conserved quantities are [Eq. (B8)]

$$\dot{\mathcal{E}} = \frac{\sqrt{r}(r - 2M) + a\sqrt{M}}{\sqrt{r^3 - 3Mr^2 + 2a\sqrt{M}r^{3/2}}}, \quad (\text{B19a})$$

$$\dot{\mathcal{L}} = \frac{\sqrt{M}(r^2 + a^2 - 2a\sqrt{Mr})}{\sqrt{r}(r - 2M) + a\sqrt{M}}. \quad (\text{B19b})$$

Likewise [Eq. (B8)],

$$\dot{\nu} = 0, \quad \dot{\chi} = 1 - \frac{2M\sqrt{r}}{r^{3/2} + a\sqrt{M}}. \quad (\text{B20})$$

Hence, the covariant four-velocity is [Eq. (B5)]

$$\dot{u}_t = -\frac{\sqrt{r}(r - 2M) + a\sqrt{M}}{\sqrt{r^3 - 3Mr^2 + 2a\sqrt{M}r^{3/2}}}, \quad (\text{B21a})$$

$$\dot{u}_r = 0, \quad (\text{B21b})$$

$$\dot{u}_\phi = \frac{\sqrt{M}(r^2 - 2a\sqrt{Mr} + a^2)}{\sqrt{r^3 - 3Mr^2 + 2a\sqrt{M}r^{3/2}}}. \quad (\text{B21c})$$

Finally, the observed redshift is [Eq. (B13)]

$$\dot{g} = \frac{\sqrt{r^3 - 3Mr^2 + 2a\sqrt{M}r^{3/2}}}{r^{3/2} + \sqrt{M}(a - \lambda)}. \quad (\text{B22})$$

All of the above quantities are physically admissible provided $r^3 - 3Mr^2 + 2a\sqrt{M}r^{3/2} > 0$, which is the case for all $r \geq r_{\text{ms}}$.

b. Inside the ISCO radius

For $r \in [r_+, r_{\text{ms}}]$, we follow Cunningham's prescription [67] and smoothly extend the flow past the ISCO using geodesic inspirals with the conserved quantities of the ISCO:

$$\mathcal{E}_{\text{ms}} = \dot{\mathcal{E}}|_{r=r_{\text{ms}}} = \sqrt{1 - \frac{2M}{3r_{\text{ms}}}}, \quad (\text{B23a})$$

$$\ell_{\text{ms}} = \dot{\mathcal{L}}|_{r=r_{\text{ms}}} = \frac{\sqrt{M}(r_{\text{ms}}^2 - 2a\sqrt{Mr_{\text{ms}}} + a^2)}{\sqrt{r_{\text{ms}}}(r_{\text{ms}} - 2M) + a\sqrt{M}}. \quad (\text{B23b})$$

This choice results in a nonzero radial component [Eq. (B12)]

$$\dot{\nu} = \frac{r^2}{\Delta(r)\dot{\mathcal{E}}_{\text{ms}}} \sqrt{\frac{2M}{3r_{\text{ms}}}} \left(\frac{r_{\text{ms}}}{r} - 1 \right)^{3/2} > 0, \quad (\text{B24})$$

where we used the identity $r_{\text{ms}}^2 - 6Mr_{\text{ms}} + 8a\sqrt{Mr_{\text{ms}}} - 3a^2 = 0$ [Eq. (B15)] to simplify the expressions for both \mathcal{E}_{ms} and $\dot{\nu}$.

It then follows that [Eq. (B10)]

$$\dot{i} = \frac{\dot{\chi} \left(\frac{r_{\text{ms}}}{r} - 1 \right)^{3/2}}{\sqrt{\frac{3r_{\text{ms}}}{2M} - 1}}, \quad \dot{\Omega} = \frac{\dot{\chi}}{r^2} (\ell_{\text{ms}} + a\dot{H}), \quad (\text{B25a})$$

$$\dot{\chi} = \frac{1}{1 + \frac{2M}{r}(1 + \dot{H})}, \quad \dot{H} = \frac{2Mr - a\ell_{\text{ms}}}{\Delta(r)}, \quad (\text{B25b})$$

so that the covariant four-velocity is explicitly [Eq. (B5)]

$$\dot{u}_t = -\sqrt{1 - \frac{2M}{3r_{\text{ms}}}}, \quad (\text{B26a})$$

$$\dot{u}_r = -\frac{r^2}{\Delta(r)} \sqrt{\frac{2M}{3r_s} \left(\frac{r_{\text{ms}}}{r} - 1 \right)^{3/2}}, \quad (\text{B26b})$$

$$\dot{u}_\phi = \frac{\sqrt{M}(r_{\text{ms}}^2 - 2a\sqrt{Mr_{\text{ms}}} + a^2)}{\sqrt{r_{\text{ms}}^3 - 3Mr_{\text{ms}}^2 + 2a\sqrt{Mr_{\text{ms}}}^{3/2}}}. \quad (\text{B26c})$$

The contravariant four-velocity is then [Eqs. (B5) and (B10)]

$$\dot{u}^t = \sqrt{1 - \frac{2M}{3r_{\text{ms}}} \frac{\Pi(r) - 2aM\ell_{\text{ms}}r}{r^2\Delta(r)}}, \quad (\text{B27a})$$

$$\dot{u}^r = -\sqrt{\frac{2M}{3r_{\text{ms}}} \left(\frac{r_{\text{ms}}}{r} - 1 \right)^{3/2}}, \quad (\text{B27b})$$

$$\dot{u}^\phi = \sqrt{1 - \frac{2M}{3r_{\text{ms}}} \frac{a + (1 - \frac{2M}{r})(\ell_{\text{ms}} - a)}{\Delta(r)}}, \quad (\text{B27c})$$

The contravariant components may also be recast in the form

$$\dot{u}^t = \frac{\mathcal{E}_{\text{ms}}}{\dot{\chi}}, \quad \dot{u}^\phi = \frac{\mathcal{E}_{\text{ms}}}{r^2} (\ell_{\text{ms}} + a\dot{H}), \quad (\text{B28a})$$

which matches Eq. (A12) of [67] since $(\gamma_e, \lambda_e) = (\mathcal{E}_{\text{ms}}, \ell_{\text{ms}})$.

Finally, the observed redshift is [Eq. (B14)]

$$\dot{g} = \frac{\Delta(r)}{\sqrt{1 - \frac{2M}{3r_{\text{ms}}} \left[\dot{G} \pm_r \sqrt{\mathcal{R}(r)} \frac{(\frac{r_{\text{ms}}}{r} - 1)^{3/2}}{\sqrt{\frac{3M}{2} - 1}} \right]}}, \quad (\text{B29a})$$

$$\dot{G} = \frac{\Pi(r)}{r^2} - \left(1 - \frac{2M}{r} \right) \ell_{\text{ms}} \lambda - \frac{2aM}{r} (\ell_{\text{ms}} + \lambda). \quad (\text{B29b})$$

2. Radial infall (geodesic motion)

We let $u = \bar{u}$ denote the four-velocity of timelike, radially infalling equatorial geodesics. One class of such trajectories consists of particles that fall in from spatial infinity with zero initial velocity and vanishing (conserved) angular momentum:

$$\bar{\mathcal{E}} = 1, \quad \bar{\ell} = 0. \quad (\text{B30})$$

As particles fall in, they pick up a radial velocity [Eq. (B12)]

$$\bar{v} = \frac{\sqrt{2Mr(r^2 + a^2)}}{\Delta(r)} > 0, \quad (\text{B31})$$

and due to frame-dragging (the off-diagonal term $g_{t\phi} \neq 0$), they also acquire an angular velocity $\bar{\Omega} \neq 0$ [Eq. (B10)]:

$$\bar{v} = \frac{\Delta(r)}{\Pi(r)} \sqrt{2Mr(r^2 + a^2)}, \quad \bar{\Omega} = \frac{2aMr}{\Pi(r)}, \quad (\text{B32a})$$

$$\bar{\chi} = \frac{r^2\Delta(r)}{\Pi(r)}, \quad \bar{H} = \frac{2Mr}{\Delta(r)}. \quad (\text{B32b})$$

Explicitly, the covariant four-velocity is then [Eq. (B5)]

$$\bar{u}_t = -1, \quad \bar{u}_r = -\frac{\sqrt{2Mr(r^2 + a^2)}}{\Delta(r)}, \quad \bar{u}_\phi = 0, \quad (\text{B33})$$

while the contravariant four-velocity is [Eqs. (B5) and (B10)]

$$\bar{u}^t = \frac{\Pi(r)}{r^2\Delta(r)}, \quad (\text{B34a})$$

$$\bar{u}^r = -\frac{\sqrt{2Mr(r^2 + a^2)}}{r^2}, \quad (\text{B34b})$$

$$\bar{u}^\phi = \frac{2aM}{r\Delta(r)}. \quad (\text{B34c})$$

Finally, the observed redshift is [Eq. (B13)]

$$\bar{g} = \frac{r^2\Delta(r)}{\Pi(r) - 2aMr\lambda \pm_r \sqrt{\mathcal{R}(r)} \sqrt{2Mr(r^2 + a^2)}}. \quad (\text{B35})$$

3. Sub-Keplerian orbits (nongeodesic motion)

We now define a four-velocity $u = \hat{u}$ for timelike, equatorial sub-Keplerian orbits. Such trajectories cannot be geodesic; following [83], we introduce a ‘‘sub-Keplerianity’’ parameter $\xi \in (0, 1)$ and instead prescribe a non-geodesic motion via

$$\hat{\ell} = \xi \dot{\ell} = \xi \frac{\sqrt{M}(r^2 + a^2 - 2a\sqrt{Mr})}{\sqrt{r}(r - 2M) + a\sqrt{M}}, \quad (\text{B36})$$

where $\dot{\ell}$ denotes the Keplerian specific angular momentum (B19b). As for Keplerian motion (with $\xi = 1$), we must treat radii outside and inside of the ISCO radius r_{ms} separately.

a. Outside the ISCO radius

For $r \geq r_{\text{ms}}$, we demand that, as for Keplerian orbits,

$$\hat{v} = 0. \quad (\text{B37})$$

This fixes the sub-Keplerian orbital energy to be [Eq. (B9)]

$$\hat{\mathcal{E}} = \sqrt{\frac{\Delta(r)}{\frac{\Pi(r)}{r^2} - \frac{4aM\hat{\ell}}{r} - \left(1 - \frac{2M}{r}\right)\hat{\ell}^2}} > 0, \quad (\text{B38})$$

which is manifestly real for all $\xi \in (0, 1)$, since the quantity in the square root is strictly greater than when $\xi = 1$.

It then follows that [Eqs. (B10) and (B11)]

$$\hat{t} = 0, \quad \hat{\Omega} = \frac{a + \left(1 - \frac{2M}{r}\right)(\hat{\ell} - a)}{\frac{\Pi(r)}{r^2} - \frac{2aM\hat{\ell}}{r}}, \quad (\text{B39})$$

so that the covariant four-velocity is explicitly [Eq. (B5)]

$$\hat{u}_t = -\hat{\mathcal{E}}, \quad \hat{u}_r = 0, \quad \hat{u}_\phi = \hat{\ell} \hat{\mathcal{E}}. \quad (\text{B40})$$

The contravariant four-velocity is then [Eqs. (B5) and (B10)]

$$\hat{u}^t = \frac{\hat{\mathcal{E}}}{\Delta(r)} \left[\frac{\Pi(r)}{r^2} - \frac{2aM\hat{\ell}}{r} \right], \quad (\text{B41a})$$

$$\hat{u}^r = 0, \quad (\text{B41b})$$

$$\hat{u}^\phi = \frac{\hat{\mathcal{E}}}{\Delta(r)} \left[a + \left(1 - \frac{2M}{r}\right)(\hat{\ell} - a) \right]. \quad (\text{B41c})$$

Finally, the observed redshift is [Eq. (B14)]

$$\hat{g} = \frac{\Delta(r)}{\hat{\mathcal{E}} \left[\frac{\Pi(r)}{r^2} - \left(1 - \frac{2M}{r}\right)\hat{\ell} - \frac{2aM}{r}(\hat{\ell} + \lambda) \right]}. \quad (\text{B42})$$

b. Inside the ISCO radius

For $r \in [r_+, r_{\text{ms}}]$, we do not expect orbits to remain circular. Instead, we use Cunningham's prescription (B23) to smoothly extend the sub-Keplerian condition (B36) past the ISCO using the conserved quantities of the sub-Keplerian ISCO orbit,

$$\hat{\mathcal{E}}_{\text{ms}} = \hat{\mathcal{E}}|_{r=r_{\text{ms}}}, \quad \hat{\ell}_{\text{ms}} = \hat{\ell}|_{r=r_{\text{ms}}} = \xi \ell_{\text{ms}}, \quad (\text{B43})$$

or more explicitly,

$$\hat{\mathcal{E}}_{\text{ms}} = \sqrt{\frac{\Delta(r_{\text{ms}})}{\frac{\Pi(r_{\text{ms}})}{r_{\text{ms}}^2} - \frac{4aM\hat{\ell}_{\text{ms}}}{r_{\text{ms}}} - \left(1 - \frac{2M}{r_{\text{ms}}}\right)\hat{\ell}_{\text{ms}}^2}}, \quad (\text{B44a})$$

$$\hat{\ell}_{\text{ms}} = \xi \frac{\sqrt{M}(r_{\text{ms}}^2 - 2a\sqrt{Mr_{\text{ms}}} + a^2)}{\sqrt{r_{\text{ms}}}(r_{\text{ms}} - 2M) + a\sqrt{M}}. \quad (\text{B44b})$$

This choice results in a nonzero radial component [Eq. (B12)]

$$\hat{v} = \frac{r}{\Delta(r)} \sqrt{\frac{\Pi(r)}{r^2} - \frac{4aM\hat{\ell}_{\text{ms}}}{r} - \left(1 - \frac{2M}{r}\right)\hat{\ell}_{\text{ms}}^2 - \frac{\Delta(r)}{\hat{\mathcal{E}}_{\text{ms}}^2}}. \quad (\text{B45})$$

It then follows that [Eq. (B10)]

$$\hat{t} = \frac{[\Delta(r)]^2 \hat{v}}{\Pi(r) - 2aM\hat{\ell}_{\text{ms}}r}, \quad \hat{\Omega} = \frac{\hat{\chi}}{r^2} (\hat{\ell}_{\text{ms}} + a\hat{H}), \quad (\text{B46a})$$

$$\hat{\chi} = \frac{1}{1 + \frac{2M}{r}(1 + \hat{H})}, \quad \hat{H} = \frac{2Mr - a\ell_{\text{ms}}}{\Delta(r)}, \quad (\text{B46b})$$

so that the covariant four-velocity is explicitly [Eq. (B5)]

$$\hat{u}_t = -\hat{\mathcal{E}}_{\text{ms}}, \quad \hat{u}_r = -\hat{v}\hat{\mathcal{E}}_{\text{ms}}, \quad \hat{u}_\phi = \hat{\ell}_{\text{ms}}\hat{\mathcal{E}}_{\text{ms}}. \quad (\text{B47})$$

The contravariant four-velocity is then [Eqs. (B5) and (B10)]

$$\hat{u}^t = \frac{\hat{\mathcal{E}}_{\text{ms}}}{\Delta(r)} \left[\frac{\Pi(r)}{r^2} - \frac{2aM\hat{\ell}_{\text{ms}}}{r} \right], \quad (\text{B48a})$$

$$\hat{u}^r = -\frac{\Delta(r)}{r^2} \hat{v} \hat{\mathcal{E}}_{\text{ms}}, \quad (\text{B48b})$$

$$\hat{u}^\phi = \frac{\hat{\mathcal{E}}_{\text{ms}}}{\Delta(r)} \left[a + \left(1 - \frac{2M}{r}\right)(\hat{\ell}_{\text{ms}} - a) \right]. \quad (\text{B48c})$$

Finally, the observed redshift is [Eq. (B14)]

$$\hat{g} = \frac{\Delta(r)}{\hat{\mathcal{E}}_{\text{ms}} [\hat{G} \pm_r \sqrt{\mathcal{R}(r)} \frac{\Delta(r)}{r^2} \hat{v}]}, \quad (\text{B49a})$$

$$\hat{G} = \frac{\Pi(r)}{r^2} - \left(1 - \frac{2M}{r}\right)\hat{\ell}_{\text{ms}}\lambda - \frac{2aM}{r}(\hat{\ell}_{\text{ms}} + \lambda). \quad (\text{B49b})$$

4. General flow (nongeodesic motion)

We can linearly superpose the preceding four-velocities to obtain a general flow $u = \tilde{u}$ combining both circular motion and radial inflow. Following [11,66], we do so by defining

$$\tilde{u}^r = \hat{u}^r + (1 - \beta_r)(\bar{u}^r - \hat{u}^r), \quad (\text{B50})$$

$$\tilde{\Omega} = \hat{\Omega} + (1 - \beta_\phi)(\bar{\Omega} - \hat{\Omega}), \quad (\text{B51})$$

where \hat{u} is the four-velocity of the sub-Keplerian flow, given by Eq. (B41) for $r \geq r_{\text{ms}}$ (outside the ISCO) and by Eq. (B48) for $r_+ < r < r_{\text{ms}}$ (inside the ISCO), while \bar{u} is the radial inflow four-velocity (B34). Likewise, $\hat{\Omega} = \hat{u}^\phi / \hat{u}^t$ is the sub-Keplerian angular velocity, given by Eq. (B46) outside the ISCO and by Eq. (B46) inside the ISCO, while $\bar{\Omega} = \bar{u}^\phi / \bar{u}^t$ is the angular velocity (B32) of radial inflow.

Given \tilde{u}^r and $\tilde{\Omega} = \tilde{u}^\phi / \tilde{u}^t$, we can solve the normalization condition $u \cdot u = -1$ for \tilde{u}^t , the only missing component of the four-velocity, resulting in a simple modification of Eq. (B7):

$$\tilde{u}^t = \sqrt{\frac{1 + \frac{r^2}{\Delta(r)} (\tilde{u}^r)^2}{1 - (r^2 + a^2) \tilde{\Omega}^2 - \frac{2M}{r} (1 - a\tilde{\Omega})^2}}, \quad (\text{B52})$$

This fully specifies the general four-velocity \tilde{u} .

For generic values of the three parameters $(\xi, \beta_r, \beta_\phi)$, the flow \tilde{u} is not geodesic. However, when $\beta_r = \beta_\phi = 0$, \tilde{u} reduces to the radial inflow \bar{u} , which is ξ -independent and geodesic. At the opposite end, when $\beta_r = \beta_\phi = 1$, \tilde{u} reduces to the sub-Keplerian flow $\tilde{u}(\xi)$, which in turn reduces when $\xi = 1$ to geodesic Keplerian motion $\hat{u} = \tilde{u}(1)$.

APPENDIX C: MORE ON THE MATÉRN COVARIANCE

This appendix offers a field-theoretic interpretation of the Gaussian random field with Matérn covariance.

1. Classical field

A free, massive scalar field $\Phi(\mathbf{x})$ with mass m in Euclidean spacetime \mathbb{R}^d is described by the Lagrangian

$$\mathcal{L} = \frac{1}{2} \partial_\mu \Phi \partial^\mu \Phi + \frac{1}{2} m^2 \Phi^2, \quad (\text{C1})$$

and therefore obeys the Euler-Lagrange field equation

$$\frac{\partial \mathcal{L}}{\partial \Phi} = \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \Phi)} \right), \quad (\text{C2})$$

which in this case is the classical Klein-Gordon wave equation

$$(\nabla^2 - m^2) \Phi(\mathbf{x}) = 0, \quad (\text{C3})$$

where $\nabla^2 = \partial_\mu \partial^\mu$ is the scalar Laplacian. The Green function $\Delta_d(\mathbf{x})$ for this equation is defined by

$$(\nabla^2 - m^2) \Delta_d(\mathbf{x}) = -\delta^{(d)}(\mathbf{x}). \quad (\text{C4})$$

In momentum space, this reads $(k^2 + m^2) \tilde{\Delta}_d(\mathbf{k}) = 1$, so

$$\tilde{\Delta}_d(\mathbf{k}) = \frac{1}{k^2 + m^2}, \quad (\text{C5})$$

where $k^2 = \mathbf{k} \cdot \mathbf{k}$. The inverse Fourier transform (70) yields

$$\Delta_d(\mathbf{x}) = \int \frac{e^{i\mathbf{k} \cdot \mathbf{x}}}{\mathbf{k}^2 + m^2} \frac{d^d \mathbf{k}}{(2\pi)^d} = \frac{m^{d-2}}{(2\pi)^{\frac{d}{2}}} (mx)^{1-\frac{d}{2}} K_{1-\frac{d}{2}}(mx), \quad (\text{C6})$$

where $x^2 = \mathbf{x} \cdot \mathbf{x}$. We present a full derivation of this nontrivial identity in Eq. (C34) below.

2. Quantum field

Integrating the Lagrangian (C1) by parts recasts it as

$$\mathcal{L} = -\frac{1}{2} \Phi (\nabla^2 - m^2) \Phi. \quad (\text{C7})$$

The path integral over this Lagrangian defines the Euclidean quantum field theory of a free, massive scalar field $\Phi(\mathbf{x})$ with mass m in \mathbb{R}^d , which is fully characterized by its two-point function: the Feynman propagator (C6).

Letting $\nu = 1 - d/2$ and $\lambda = 1/m$, the propagator becomes

$$\Delta_d(\mathbf{x}) = \frac{m^{d-2}}{(2\pi)^{\frac{d}{2}}} \left(\frac{x}{\lambda} \right)^\nu K_\nu \left(\frac{x}{\lambda} \right), \quad (\text{C8})$$

which is (proportional to) a Matérn covariance.

We conclude that a d -dimensional Gaussian random field with Matérn covariance of order $\nu = 1 - d/2$ and correlation length λ is equivalent to a (Euclidean) quantum scalar field of Compton wavelength λ (and therefore mass $m = 1/\lambda$).

3. Higher-derivative fields

Consider now the higher-derivative theory

$$\mathcal{L} = -\frac{1}{2} \Phi (\nabla^2 - m^2)^n \Phi, \quad (\text{C9})$$

where n is an integer power. This is a free (Gaussian) theory because it remains quadratic in the field for any n . Classically, the field obeys the higher-derivative wave equation

$$(\nabla^2 - m^2)^n \Phi(\mathbf{x}) = 0, \quad (\text{C10})$$

whose Green function $\Delta_{d,n}(\mathbf{x})$ is defined by

$$(\nabla^2 - m^2)^n \Delta_{d,n}(\mathbf{x}) = -\delta^{(d)}(\mathbf{x}). \quad (\text{C11})$$

In momentum space, this reads $(k^2 + m^2)^n \tilde{\Delta}_{d,n}(\mathbf{k}) = 1$, so

$$\tilde{\Delta}_{d,n}(\mathbf{k}) = \frac{1}{(k^2 + m^2)^n}, \quad (\text{C12})$$

and the inverse Fourier transform (70) yields

$$\Delta_{d,n}(\mathbf{x}) = \int \frac{e^{i\mathbf{k} \cdot \mathbf{x}}}{(\mathbf{k}^2 + m^2)^n} \frac{d^d \mathbf{k}}{(2\pi)^d} \propto (mx)^{n-\frac{d}{2}} K_{n-\frac{d}{2}}(mx). \quad (\text{C13})$$

This nontrivial identity, including the proportionality factor, is derived in Eq. (C34).

Letting $\nu = n - d/2$ and $\lambda = 1/m$, the propagator becomes

$$\Delta_{d,n}(\mathbf{x}) = C \left(\frac{x}{\lambda} \right)^\nu K_\nu \left(\frac{x}{\lambda} \right), \quad (\text{C14})$$

which is (proportional to) a Matérn covariance.

We conclude that a d -dimensional Gaussian random field with Matérn covariance of order $\nu = n - d/2$ and correlation length λ is equivalent to a (Euclidean) quantum scalar field of Compton wavelength λ (and therefore mass $m = 1/\lambda$) with a kinetic term including $2n$ derivatives.

In Lorentzian signature, such a kinetic term would lead to problems with causality, but in Euclidean signature this seems like an acceptable statistical field.

4. Connection with the associated SPDE

A d -dimensional Gaussian random field $\Phi(\mathbf{x})$ with Matérn covariance of order $\nu = n - d/2$ and correlation length $\lambda = 1/m$ obeys the stochastic PDE

$$(m^2 - \nabla^2)^{\frac{n}{2}} \Phi(\mathbf{x}) = \mathcal{W}(\mathbf{x}), \quad (\text{C15})$$

where the pseudodifferential operator on the left-hand side (LHS) is defined via its spectral properties [76], and the Gaussian random field $\mathcal{W}(\mathbf{x})$ is the standard white noise process defined in Sec. III D.

We now wish to reconcile this statement with more familiar facts from quantum field theory, particularly in the case $n = 1$.

When $n = 1$, the Gaussian random field $\Phi(\mathbf{x})$ is a Euclidean scalar field. Yet, Eq. (C15) implies that it also obeys

$$\sqrt{m^2 - \nabla^2} \Phi(\mathbf{x}) = \mathcal{W}(\mathbf{x}), \quad (\text{C16})$$

which seems to impose yet another condition on the field. We now show how this statement is consistent with the propagator (C6), which fully characterizes the behavior of the field.

The key is to multiply two copies of this equation inserted at two different points \mathbf{x}_1 and \mathbf{x}_2 , keeping track of which position the derivatives in the Laplacian act upon:

$$\sqrt{m^2 - \nabla_1^2} \Phi(\mathbf{x}_1) \sqrt{m^2 - \nabla_2^2} \Phi(\mathbf{x}_2) = \mathcal{W}(\mathbf{x}_1) \mathcal{W}(\mathbf{x}_2). \quad (\text{C17})$$

We now massage the LHS as follows. First, we are free to move the derivatives to the left, since ∇_i^2 acts only on $\Phi(\mathbf{x}_i)$:

$$\sqrt{m^2 - \nabla_1^2} \sqrt{m^2 - \nabla_2^2} \Phi(\mathbf{x}_1) \Phi(\mathbf{x}_2) = \mathcal{W}(\mathbf{x}_1) \mathcal{W}(\mathbf{x}_2). \quad (\text{C18})$$

Next, we take expectation values of the fields on both sides:

$$\sqrt{(m^2 - \nabla_1^2)(m^2 - \nabla_2^2)} \langle \Phi(\mathbf{x}_1) \Phi(\mathbf{x}_2) \rangle = \langle \mathcal{W}(\mathbf{x}_1) \mathcal{W}(\mathbf{x}_2) \rangle. \quad (\text{C19})$$

By definition, the two-point function on the LHS is of course the Euclidean quantum propagator

$$\langle \Phi(\mathbf{x}_1) \Phi(\mathbf{x}_2) \rangle = \Delta_d(\mathbf{x}_1 - \mathbf{x}_2). \quad (\text{C20})$$

As for the two-point function on the RHS, it is none other than the autocorrelation function of Gaussian white noise,

$$\langle \mathcal{W}(\mathbf{x}_1) \mathcal{W}(\mathbf{x}_2) \rangle = \delta^{(d)}(\mathbf{x}_1 - \mathbf{x}_2), \quad (\text{C21})$$

which is by definition a delta function. We thus have

$$\sqrt{(m^2 - \nabla_1^2)(m^2 - \nabla_2^2)} \Delta_d(\mathbf{x}_1 - \mathbf{x}_2) = \delta^{(d)}(\mathbf{x}_1 - \mathbf{x}_2). \quad (\text{C22})$$

Since both two-point functions depend only on $\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2$, we have $\nabla_1^2 = \nabla_2^2 = \nabla^2$, and thus we are left with

$$(m^2 - \nabla^2) \Delta_d(\mathbf{x}) = \delta^{(d)}(\mathbf{x}). \quad (\text{C23})$$

This is exactly Eq. (C4), proving that it is consistent for the Euclidean quantum field $\Phi(\mathbf{x})$ to also obey the SPDE (C16). This is the quantum-statistical analog of the classical field equation (C3): the presence of the white noise source term $\mathcal{W}(\mathbf{x})$ is what lends the Euclidean field its statistical nature.

Finally, when $n = 2$, the Gaussian random field $\Phi(\mathbf{x})$ obeys

$$(m^2 - \nabla^2) \Phi(\mathbf{x}) = \mathcal{W}(\mathbf{x}), \quad (\text{C24})$$

which is a linear SPDE and therefore straightforward to solve. This is the reason why it is implemented in INOISY [43].

5. Derivation of the scalar propagator

Here, we derive Eqs. (C6) and (C13), which are equivalent to the position-space formula (77) for the Matérn covariance. In spherical coordinates,

$$\Delta_{d,n}(\mathbf{x}) = \frac{\Omega_{d-1}}{(2\pi)^d} \int_0^\infty \frac{k^{d-1}}{(k^2 + m^2)^n} \int_0^\pi e^{ikx \cos \theta} \sin^{d-2} \theta d\theta dk, \quad (\text{C25})$$

where Ω_n denotes the solid angle on the sphere S^n , which is

$$\Omega_n = \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}. \quad (\text{C26})$$

We now need two identities from Gradshteyn and Ryzik [84]: by 3.915-5 (p492), we have for $\text{Re}\nu > -1/2$

$$\int_0^\pi e^{i\beta \cos \theta} \sin^{2\nu} \theta d\theta = \sqrt{\pi} \left(\frac{2}{\beta}\right)^\nu \Gamma\left(\nu + \frac{1}{2}\right) J_\nu(\beta), \quad (\text{C27})$$

and by 6.565-4 (p678), we have for $-1 < \text{Re}\nu < \text{Re}(2\mu + 2/3)$, $r > 0$, and $m > 0$,

$$\int_0^\infty \frac{k^{\nu+1} J_\nu(kx)}{(k^2 + m^2)^{\mu+1}} dk = \frac{m^{\nu-\mu} x^\mu}{2^\mu \Gamma(\mu + 1)} K_{\nu-\mu}(mx). \quad (\text{C28})$$

Using the first formula with $\beta = kx$ and $\nu = d/2 - 1$, we find

$$\Delta_{d,n}(\mathbf{x}) = X \int_0^\infty \frac{k^{d-1}}{(k^2 + m^2)^n} \left(\frac{2}{kx}\right)^{\frac{d}{2}-1} J_{\frac{d}{2}-1}(kx) dk, \quad (\text{C29})$$

where the prefactor is

$$X = \frac{\Omega_{d-1}}{(2\pi)^d} \sqrt{\pi} \Gamma\left(\frac{d-1}{2}\right) = \frac{2\pi^{\frac{d-1}{2}}}{(2\pi)^d} \sqrt{\pi} = \frac{2^{1-\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}}. \quad (\text{C30})$$

We can thus rewrite

$$\Delta_{d,n}(\mathbf{x}) = \frac{x^{1-\frac{d}{2}}}{(2\pi)^{\frac{d}{2}}} \int_0^\infty \frac{k^{\frac{d}{2}}}{(k^2 + m^2)^n} J_{\frac{d}{2}-1}(kx) dk. \quad (\text{C31})$$

The second identity with $\nu = d/2 - 1$ and $\mu = n - 1$ then yields

$$\Delta_{d,n}(\mathbf{x}) = \frac{x^{1-\frac{d}{2}} m^{\frac{d}{2}-n} x^{n-1}}{(2\pi)^{\frac{d}{2}} 2^{n-1} \Gamma(n)} K_{\frac{d}{2}-n}(mx), \quad (\text{C32})$$

so we finally obtain

$$\Delta_{d,n}(\mathbf{x}) = \frac{2^{1-n}}{(2\pi)^{\frac{d}{2}} (n-1)!} \left(\frac{m}{x}\right)^{\frac{d}{2}-n} K_{\frac{d}{2}-n}(mx). \quad (\text{C33})$$

Since $K_{-n}(x) = K_n(x)$, this exactly reproduces Eq. (C13):

$$\Delta_{d,n}(\mathbf{x}) = \frac{2^{1-n} m^{d-2n}}{(2\pi)^{\frac{d}{2}} (n-1)!} (mx)^{n-\frac{d}{2}} K_{n-\frac{d}{2}}(mx), \quad (\text{C34})$$

and setting $n = 1$ recovers Eq. (C6).

-
- [1] C. Darwin, *Proc. R. Soc. A* **249**, 180 (1959).
[2] J. P. Luminet, *Astron. Astrophys.* **75**, 228 (1979).
[3] H. C. Ohanian, *Am. J. Phys.* **55**, 428 (1987).
[4] V. Bozza, *Gen. Relativ. Gravit.* **42**, 2269 (2010).
[5] S. E. Gralla and A. Lupsasca, *Phys. Rev. D* **101**, 044031 (2020).
[6] S. E. Gralla, D. E. Holz, and R. M. Wald, *Phys. Rev. D* **100**, 024018 (2019).
[7] M. D. Johnson, A. Lupsasca, A. Strominger, G. N. Wong, S. Hadar, D. Kapec, R. Narayan, A. Chael, C. F. Gammie, P. Galison, D. C. M. Palumbo, S. S. Doeleman, L. Blackburn, M. Wielgus, D. W. Pesce, J. R. Farah, and J. M. Moran, *Sci. Adv.* **6**, eaaz1310 (2020).
[8] S. Hadar, M. D. Johnson, A. Lupsasca, and G. N. Wong, *Phys. Rev. D* **103**, 104038 (2021).
[9] A. Chael, M. D. Johnson, and A. Lupsasca, *Astrophys. J.* **918**, 6 (2021).
[10] H. Pagnat, A. Lupsasca, F. Vincent, and M. Wielgus, *Astron. Astrophys.* **668**, A11 (2022).
[11] F. H. Vincent, S. Gralla, A. Lupsasca, and M. Wielgus, *Astron. Astrophys.* **667**, A170 (2022).
[12] H. Falcke, F. Melia, and E. Agol, *Astrophys. J. Lett.* **528**, L13 (2000).
[13] R. Narayan, M. D. Johnson, and C. F. Gammie, *Astrophys. J. Lett.* **885**, L33 (2019).
[14] V. Perlick and O. Y. Tsupko, *Phys. Rep.* **947**, 1 (2022).
[15] J. M. Bardeen, in *Black Holes (Les Astres Occlus)*, edited by C. Dewitt and B. S. Dewitt (Gordon and Breach Science Publishers, New York, 1973), pp. 215–239.
[16] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **875**, L1 (2019).
[17] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **930**, L12 (2022).
[18] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **875**, L5 (2019).
[19] P. Arras, P. Frank, P. Haim, J. Knollmüller, R. Leike, M. Reinecke, and T. Enßlin, *Nat. Astron.* **6**, 259 (2022).
[20] C. L. Carilli and N. Thyagarajan, *Astrophys. J.* **924**, 125 (2022).
[21] W. Lockhart and S. E. Gralla, *Mon. Not. R. Astron. Soc.* **509**, 3643 (2022).
[22] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **930**, L16 (2022).
[23] W. Lockhart and S. E. Gralla, *Mon. Not. R. Astron. Soc.* **517**, 2462 (2022).
[24] S. E. Gralla, *Phys. Rev. D* **103**, 024023 (2021).
[25] G. Lara, S. H. Völkel, and E. Barausse, *Phys. Rev. D* **104**, 124041 (2021).
[26] A. M. Bauer, A. Cárdenas-Avendaño, C. F. Gammie, and N. Yunes, *Astrophys. J.* **925**, 119 (2022).

- [27] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **875**, L6 (2019).
- [28] K. Akiyama *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J. Lett.* **930**, L17 (2022).
- [29] A. Cárdenas-Avenidaño, S. Nampalliwar, and N. Yunes, *Classical Quantum Gravity* **37**, 135008 (2020).
- [30] D. Ayzenberg and C. Bambi, [arXiv:2111.13918](https://arxiv.org/abs/2111.13918).
- [31] G. N. Wong, B. S. Prather, V. Dhruv, B. R. Ryan, M. Mościbrodzka, C.-K. Chan, A. V. Joshi, R. Yarza, A. Ricarte, H. Shiokawa, J. C. Dolence, S. C. Noble, J. C. McKinney, and C. F. Gammie, *Astrophys. J. Suppl. Ser.* **259**, 64 (2022).
- [32] G. N. Wong, V. Dhruv, and B. S. Prather, Characteristics of thermal synchrotron emission in radiatively inefficient accretion flows (to be published).
- [33] S. E. Gralla, *Phys. Rev. D* **102**, 044017 (2020).
- [34] S. E. Gralla and A. Lupsasca, *Phys. Rev. D* **102**, 124003 (2020).
- [35] S. E. Gralla, A. Lupsasca, and D. P. Marrone, *Phys. Rev. D* **102**, 124004 (2020).
- [36] G. N. Wong, *Astrophys. J.* **909**, 217 (2021).
- [37] Z. Gelles, B. S. Prather, D. C. M. Palumbo, M. D. Johnson, G. N. Wong, and B. Georgiev, *Astrophys. J.* **912**, 39 (2021).
- [38] C. J. White, *Astrophys. J. Suppl. Ser.* **262**, 28 (2022).
- [39] E. Teo, *Gen. Relativ. Gravit.* **53**, 10 (2021).
- [40] S. E. Gralla and A. Lupsasca, *Phys. Rev. D* **101**, 044032 (2020).
- [41] J. Dexter and E. Agol, *Astrophys. J.* **696**, 1616 (2009).
- [42] X.-L. Yang and J.-C. Wang, *Astron. Astrophys.* **561**, A127 (2014).
- [43] D. Lee and C. F. Gammie, *Astrophys. J.* **906**, 39 (2021).
- [44] A. M. Beloborodov, *Astrophys. J. Lett.* **566**, L85 (2002).
- [45] D. E. A. Gates, S. Hadar, and A. Lupsasca, *Phys. Rev. D* **102**, 104041 (2020).
- [46] D. Kapec and A. Lupsasca, *Classical Quantum Gravity* **37**, 015006 (2020).
- [47] B. Carter, *Phys. Rev.* **174**, 1559 (1968).
- [48] K. P. Rauch and R. D. Blandford, *Astrophys. J.* **421**, 46 (1994).
- [49] S. E. Vázquez and E. P. Esteban, *Nuovo Cimento B* **119**, 489 (2004).
- [50] L.-X. Li, E. R. Zimmerman, R. Narayan, and J. E. McClintock, *Astrophys. J. Suppl. Ser.* **157**, 335 (2005).
- [51] L.-X. Li, R. Narayan, and J. E. McClintock, *Astrophys. J.* **691**, 847 (2009).
- [52] A. Lupsasca, D. Kapec, Y. Shi, D. E. A. Gates, and A. Strominger, *Proc. R. Soc. A* **476**, 20190618 (2020).
- [53] E. Himwich, M. D. Johnson, A. Lupsasca, and A. Strominger, *Phys. Rev. D* **101**, 084020 (2020).
- [54] R. Narayan *et al.* (Event Horizon Telescope Collaboration), *Astrophys. J.* **912**, 35 (2021).
- [55] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford University Press, New York, 1983).
- [56] J. M. Bardeen, W. H. Press, and S. A. Teukolsky, *Astrophys. J.* **178**, 347 (1972).
- [57] E. Hackmann, Geodesic equations in black hole space-times with cosmological constant, Ph.D. thesis, Bremen University, 2010.
- [58] G. Compère, Y. Liu, and J. Long, *Phys. Rev. D* **105**, 024075 (2022).
- [59] S. Hadar, D. Kapec, A. Lupsasca, and A. Strominger, *Classical Quantum Gravity* **39**, 215001 (2022).
- [60] K. Beckwith and C. Done, *Mon. Not. R. Astron. Soc.* **359**, 1217 (2005).
- [61] O. James, E. v. Tunzelmann, P. Franklin, and K. S. Thorne, *Classical Quantum Gravity* **32**, 065001 (2015).
- [62] S. E. Gralla, A. Lupsasca, and A. Strominger, *Mon. Not. R. Astron. Soc.* **475**, 3829 (2018).
- [63] D. E. A. Gates, S. Hadar, and A. Lupsasca, *Phys. Rev. D* **103**, 044050 (2021).
- [64] V. Bozza, *Phys. Rev. D* **78**, 063014 (2008).
- [65] O. Y. Tsupko, *Phys. Rev. D* **106**, 064033 (2022).
- [66] H.-Y. Pu, K. Akiyama, and K. Asada, *Astrophys. J.* **831**, 4 (2016).
- [67] C. T. Cunningham, *Astrophys. J.* **202**, 788 (1975).
- [68] D. H. Roberts, J. F. C. Wardle, and L. F. Brown, *Astrophys. J.* **427**, 718 (1994).
- [69] C. T. Cunningham and J. M. Bardeen, *Astrophys. J.* **183**, 237 (1973).
- [70] A. E. Broderick and A. Loeb, *Mon. Not. R. Astron. Soc.* **363**, 353 (2005).
- [71] A. E. Broderick and A. Loeb, *Mon. Not. R. Astron. Soc.* **367**, 905 (2006).
- [72] N. Hamaus, T. Paumard, T. Müller, S. Gillessen, F. Eisenhauer, S. Trippe, and R. Genzel, *Astrophys. J.* **692**, 902 (2009).
- [73] M. Zamaninasab, A. Eckart, G. Witzel, M. Dovciak, V. Karas, R. Schödel, R. Gießbübel, M. Bremer, M. García-Marín, D. Kunneriath, K. Mužić, S. Nishiyama, N. Sabha, C. Straubmeier, and A. Zensus, *Astron. Astrophys.* **510**, A3 (2010).
- [74] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, *Astrophys. J.* **304**, 15 (1986).
- [75] P. Guttorp and T. Gneiting, *Biometrika* **93**, 989 (2006).
- [76] F. Lindgren, H. Rue, and J. Lindström, *J. R. Stat. Soc. Ser. B* **73**, 423 (2011).
- [77] G.-A. Fuglstad, F. Lindgren, D. Simpson, and H. Rue, *Statistica Sinica* **25**, 115 (2015).
- [78] X. Guan, C. F. Gammie, J. B. Simon, and B. M. Johnson, *Astrophys. J.* **694**, 1010 (2009).
- [79] S. C. Noble and J. H. Krolik, *Astrophys. J.* **703**, 964 (2009).
- [80] K. Chatterjee, Z. Younsi, M. Liska, A. Tchekhovskoy, S. B. Markoff, D. Yoon, D. van Eijnatten, C. Hesp, A. Ingram, and M. B. M. van der Klis, *Mon. Not. R. Astron. Soc.* **499**, 362 (2020).
- [81] L. I. Gurvits *et al.*, *Acta Astronaut.* **196**, 314 (2022).
- [82] P. Kurczynski, M. D. Johnson, S. S. Doeleman *et al.*, *Proc. SPIE* **12180**, 121800M (2022).
- [83] R. F. Penna, A. Kulkarni, and R. Narayan, *Astron. Astrophys.* **559**, A116 (2013).
- [84] I. S. Gradshteyn, I. M. Ryzhik, A. Jeffrey, and D. Zwillinger, *Table of Integrals, Series, and Products* (Academic Press, New York, 2007).