First Sagittarius A* Event Horizon Telescope Results. IV. Variability, Morphology, and Black Hole Mass

The Event Horizon Telescope Collaboration
(See the end matter for the full list of authors.)

Received 2022 March 15; revised 2022 April 12; accepted 2022 April 12; published 2022 May 12

Abstract

In this paper we quantify the temporal variability and image morphology of the horizon-scale emission from Sgr A*, as observed by the EHT in 2017 April at a wavelength of 1.3 mm. We find that the Sgr A* data exhibit variability that exceeds what can be explained by the uncertainties in the data or by the effects of interstellar scattering. The magnitude of this variability can be a substantial fraction of the correlated flux density, reaching \(\sim 100\%\) on some baselines. Through an exploration of simple geometric source models, we demonstrate that ring-like morphologies provide better fits to the Sgr A* data than do other morphologies with comparable complexity. We develop two strategies for fitting static geometric ring models to the time-variable Sgr A* data; one strategy fits models to short segments of data over which the source is static and averages these independent fits, while the other fits models to the full data set using a parametric model for the structural variability power spectrum around the average source structure. Both geometric modeling and image-domain feature extraction techniques determine the ring diameter to be 51.8 \(\pm\) 2.3 \(\mu\)as (68\% credible intervals), with the ring thickness constrained to have an FWHM between \(\sim 30\%\) and 50\% of the ring diameter. To bring the diameter measurements to a common physical scale, we calibrate them using synthetic data generated from GRMHD simulations. This calibration constrains the angular size of the gravitational radius to be 4.8\(^{+1.4}_{-0.7}\) \(\mu\)as, which we combine with an independent distance measurement from maser parallaxes to determine the mass of Sgr A* to be \(4.0^{+1.1}_{-0.6} \times 10^6\ M_\odot\).

Unified Astronomy Thesaurus concepts: Black holes (162)

1. Introduction

Sagittarius A* (Sgr A*), the radio source associated with the supermassive black hole (SMBH) at the center of the Milky Way, is thought to subtend the largest angular size of all black holes in the sky. At a distance of \(D \approx 8\) kpc and with a mass of \(M \approx 4 \times 10^6\ M_\odot\) (Do et al. 2019; Gravity Collaboration et al. 2019, 2020), Sgr A* has a Schwarzschild radius of \(\sim 10\) \(\mu\)as. Models of optically thin spherical accretion flows around SMBHs generically predict that they will appear to distant observers as bright rings of emission surrounding a darker central “shadow” (e.g., Bardeen 1973; Luminet 1979; de Vries 2000; Falcke et al. 2000; Broderick & Loeb 2006; Broderick & Narayan 2006; Broderick et al. 2011, 2016; Narayan et al. 2019), and a variety of more general accretion flow simulations have demonstrated that the diameter of this ring is typically \(\sim 5\) times larger than the Schwarzschild radius (e.g., Event Horizon Telescope Collaboration et al. 2019c). The Event Horizon Telescope (EHT) collaboration provided observational verification of this picture, using a global very long baseline interferometry (VLBI) network of radio telescopes observing at a frequency of \(\sim 230\) GHz to resolve the \(\sim 40\) \(\mu\)as ring of emission around the M87* SMBH (Event Horizon Telescope Collaboration et al. 2019a, 2019b, 2019c, 2019d, 2019e, 2019f, 2021a, 2021b, hereafter M87* Papers I–VIII).

The predicted ring diameter for Sgr A* is \(\sim 50\) \(\mu\)as, about 25\% larger than what the EHT observed for M87*. However, because Sgr A* is more than three orders of magnitude less massive than M87*, all dynamical timescales in the system are correspondingly shorter. In particular, the typical gravitational timescale for Sgr A* is \(GM/c^2 \approx 20\) s, implying that the source structure can vary substantially over the several-hour duration of a single EHT observation. Consistent with this expectation, Sgr A* exhibits broadband variability on timescales of minutes to hours (e.g., Genzel et al. 2003; Ghez et al. 2004; Fish et al. 2011; Nilsson et al. 2013; Goddi et al. 2021; Wielgus et al. 2022). The multiwavelength properties of Sgr A* during the 2017 EHT observing campaign are described in Event Horizon Telescope Collaboration et al. (2022b, hereafter Paper II).

The potential for rapid structural variability complicates the analysis of EHT observations of Sgr A*. A standard strategy for ameliorating the sparsity of VLBI data sets is Earth-rotation aperture synthesis, whereby Fourier coverage of the array is accumulated as Earth rotates and baselines change their orientation with respect to the source (Thompson et al. 2017). This strategy is predicated on the source remaining static throughout the observing period, in which case the accumulated data measure a single image structure. However, Sgr A* violates this assumption on timescales as short as minutes. After several hours, the variable components of the image structure in Sgr A* are expected to be uncorrelated (Georgiev et al. 2022; Wielgus et al. 2022). Thus, image reconstructions from the EHT Sgr A* data are focused on reconstructing time-averaged source structures (Event Horizon Telescope Collaboration et al. 2022c, hereafter Paper III).

Despite the necessity of reconstructing an average source structure, the data collected within a single multihour observation epoch are associated with many specific instances of the variable emission from Sgr A*, i.e., they represent an amalgam of observations of instantaneous images. The imaging strategy pursued for the EHT observations of Sgr A* aims to...
mitigate the impact of this changing source structure through the introduction of a “variability noise budget,” which absorbs the structural evolution into inflated uncertainties and thereby permits imaging algorithms to reconstruct a time-averaged image under the usual static source assumption.\footnote{Note that this procedure ignores the subhour correlations that are present in the Sgr A* data; the implications of these correlations are discussed in Section 3.} The image reconstruction procedure is described in detail in Paper III, and the results confirm that the Sgr A* data are consistent with being produced by a ring-like emission structure with a diameter of \( \sim 50 \mu \text{as} \).

For the EHT observations of M87\(^*\), morphological properties of the observed ring (e.g., diameter, thickness, orientation) were quantified using both imaging and geometrical modeling analyses (M87\(^*\) Paper VI), and the measured ring diameter was calibrated using general relativistic magnetohydrodynamic (GRMHD) simulations from M87\(^*\) Paper V to constrain the mass of the SMBH. The current paper applies a conceptually similar strategy to the analysis of the EHT Sgr A\(^*\) data, though significant alterations have been made to meet the new challenges posed by Sgr A\(^*\) and to tailor the analyses appropriately. In this paper, we first characterize the variability seen in the Sgr A\(^*\) data, and we develop a framework for mitigating the impact of variability when imaging or modeling the data. We then make measurements of the ring size and other structural properties using both imaging and geometrical modeling analyses, and we derive and apply a GRMHD-based calibration to bring ring size measurements made using different techniques to a common physical scale.

This paper is organized as follows. Section 2 provides an overview of the Sgr A\(^*\) observations and data processing. In Section 3, we quantify the variability on different spatial scales, and we outline the strategies used to mitigate its impact during imaging and modeling. In Section 4, we discuss salient data properties in the context of a ring-like emission structure, and we describe our procedure for using GRMHD simulations to calibrate different ring size measurement techniques to a common physical scale. Sections 5, 6, and 7 detail our three primary strategies for measuring the ring size and describe their application to the Sgr A\(^*\) data. Our results are presented in Section 8, and we summarize and conclude in Section 9. This paper is the fourth in a series that describes the analysis of the 2017 EHT observations of Sgr A\(^*\). The series is summarized in Event Horizon Telescope Collaboration et al. (2022a, hereafter Paper I). The data processing and calibration are described in Paper II, imaging is carried out in Paper III, physical simulations are described in Event Horizon Telescope Collaboration et al. (2022d, hereafter Paper V), and tests of gravity are presented in Event Horizon Telescope Collaboration et al. (2022e, hereafter Paper VI).

### 2. Observations and Data Products

In this section, we briefly review the interferometric data products used for analyses in this paper (Section 2.1), and we summarize the observations (Section 2.2) and data processing (Section 2.3) that precede these analyses. A more comprehensive description of the Sgr A\(^*\) data collection, correlation, and calibration can be found in Paper II, M87\(^*\) Paper III, and references therein.

#### 2.1. VLBI Data Products

As a radio interferometer, the EHT is natively sensitive to the Fourier transform of the sky-plane emission structure. For a source of emission \( f(x, t) \), the complex visibility \( \mathcal{V}(u, t) \) is given by

\[
\mathcal{V}(u, t) = \mathcal{F}\{I(x, t)\} = \int e^{-2\pi i u \cdot x} I(x, t) d^2x,
\]

where \( t \) is time, \( x = (x, y) \) are angular coordinates on the sky, and \( u = (u, v) \) are projected baseline coordinates in units of the observing wavelength (see, e.g., Thompson et al. 2017).

The ideal visibilities \( \mathcal{V} \) are not directly observable because they are corrupted by both statistical errors and a variety of systematic effects. For the EHT, the dominant systematics are complex station-based gain corruptions. The relationship between an ideal visibility \( \mathcal{V}_i \) and the observed visibility \( V_{ij} \) on a baseline connecting stations \( i \) and \( j \) is given by

\[
V_{ij} = g_i g_j^* \mathcal{V}_{ij} + \sigma_{\mathcal{V}_{ij}} \equiv |V_{ij}| e^{i \phi_{ij}},
\]

where \( \sigma_{\mathcal{V}_{ij}} \) is the statistical (or “thermal”) error on the baseline, \( g_i \) and \( g_j \) are the station gains, and we have defined the visibility amplitude \( |V_{ij}| \) and phase \( \phi_{ij} \). The statistical error is well described as a zero-mean circularly symmetric complex Gaussian random variable with a variance determined (per the radiometer equation) by the station sensitivities, integration time, and frequency bandwidth (Thompson et al. 2017). The station gains vary in time at every site and must in general be either calibrated out or determined alongside the source structure.

The presence of station-based systematics motivates the construction and use of “closure quantities” that are invariant to such corruptions. A closure phase \( \psi_{ijk} \) (Jennison 1958) is the sum of visibility phases around a closed triangle of baselines connecting stations \( i, j, \) and \( k \),

\[
\psi_{ijk} = \phi_{ij} + \phi_{jk} + \phi_{ki}.
\]

Closure phases are invariant to station-based phase corruptions, such that the measured closure phase is equal to the ideal closure phase, up to statistical errors. Similarly, a closure amplitude \( A_{ikl} \) (Twiss et al. 1960) is the ratio of pairs of visibility amplitudes on a closed quadrangle of baselines connecting stations \( i, j, k, \) and \( \ell \),

\[
A_{ikl} = \frac{|V_{ij}| |V_{kl}|}{|V_{ik}| |V_{jl}|}.
\]

Analogous with closure phases, closure amplitudes are invariant to station-based amplitude corruptions. Because closure quantities are constructed from nonlinear combinations of complex visibilities, they have correlated and non-Gaussian error statistics; a detailed discussion is provided in Blackburn et al. (2020).

#### 2.2. EHT Observations of Sgr A\(^*\)

The EHT observed Sgr A\(^*\) on 2017 April 5, 6, 7, 10, and 11 with the phased Atacama Large Millimeter/submillimeter Array (ALMA) and the Atacama Pathfinder Experiment (APEX) on the Llano de Chajnantor in Chile, the Large
Millimeter Telescope Alfonso Serrano (LMT) on Volcán Sierra Negra in Mexico, the James Clerk Maxwell Telescope (JCMT) and phased Submillimeter Array (SMA) on Maunakea in Hawai’i, the IRAM 30 m telescope (PV) on Pico Veleta in Spain, the Submillimeter Telescope (SMT) on Mt. Graham in Arizona, and the South Pole Telescope (SPT) in Antarctica (M87’ Paper II). Only the April 6, 7, and 11 observations included the highly sensitive ALMA station, and the April 11 light curve exhibits strong variability (Wielgus et al. 2022) that is presumably associated with an X-ray flare that occurred shortly before the start of the track (Paper II). In this paper, we thus analyze primarily the April 6 and April 7 data sets. We note that while Paper III focuses on the April 7 data set, with the April 6 data set used for secondary validation, most of the analyses carried out in this paper instead focus on a joint data set that combines the April 6 and April 7 data.

At each site the data were recorded in two 1.875 GHz wide frequency bands, centered around sky frequencies of 227.1 GHz (low band; LO) and 229.1 GHz (high band; HI), and in each of two polarization modes. For all telescopes except ALMA and JCMT, the data were recorded in a dual circular polarization mode: right-hand circular polarization (RCP; R) and left-hand circular polarization (LCP; L). ALMA recorded using linear feeds, and the data were later converted to a circular polarization basis during the DiFX (Deller et al. 2011) correlation (Martí-Vidal et al. 2016; Matthews et al. 2018; Goddi et al. 2019). The JCMT observed only a single hand of circular polarization at a time, with the specific handedness (RCP or LCP) changing from day to day. All other stations observed in a standard dual-polarization mode, which allows the construction of RR, RL, LR, and LL correlation products. The analyses in this paper use only the parallel-hand correlations (i.e., RR and LL), which are averaged to form Stokes $I$ data products. Because JCMT records only a single hand at a time, we instead form “pseudo-$I$” data products for JCMT baselines, using whichever parallel-hand correlation is available as a stand-in for Stokes $I$.\footnote{The “pseudo-$I$” formation is a good approximation for Stokes $I$ when the magnitude of the Stokes $V$ contribution is small. We expect this condition to be met for the 2017 EHT observations of Sgr A$^*$ (Goddi et al. 2021), and the impact of residual Stokes $V$ is captured by the systematic error budget (Paper II).}

2.3. Data Reduction

After correlation, residual phase and bandpass errors are corrected with two independent processing pipelines: EHT-HOPS (Blackburn et al. 2019) producing “HOPS” (Whitney et al. 2004) data and rPICARD (Janssen et al. 2018, 2019) producing “CASA” (McMullin et al. 2007) data. Relative phase gains between RCP and LCP have been corrected based on the assumption of zero circular polarization on baselines between ALMA and other EHT stations. Absolute flux density scales are based on a priori measurements of each station’s sensitivity, resulting in a $\sim 10\%$ typical uncertainty in the amplitude gains (M87’ Paper II). The amplitude gains of the colocated ALMA/APEX and SMA/JCMT stations have been further refined via time-variable network calibration (M87’ Paper III) using a light curve of the compact Sgr A$^*$ flux measured by ALMA and SMA (Wielgus et al. 2022). For the remaining stations, gross amplitude gain errors have been corrected by a transfer of gain solutions from the J1924–2914 and NRAO 5300 calibrator sources as described in Paper II.

Following the completion of the above calibration pipelines, additional preprocessing of the data has been carried out as described in Paper III, including calibration of the LMT and JCMT station gains and normalization of the visibility amplitudes by the total light curve. The characterization of residual calibration effects (e.g., polarization leakage) into a systematic error budget, as well as a more comprehensive description of the overall EHT Sgr A$^*$ data reduction, is provided in Paper II.

3. Variability Extraction and Mitigation

The statistical errors quoted in Paper II and summarized in the preceding section do not account for three additional sources of uncertainty that can otherwise substantially bias any analysis efforts. First, unaccounted-for nonclosing (i.e., baseline-based) systematic errors are present in the data at a level that is on the order of $\sim 1\%$ of the visibility amplitude, which is often larger than the formal statistical errors (for a discussion of their magnitude and potential origins, see Paper II). Second, significant refractive scattering in the interstellar medium produces additional substructure within the image that is not present in the intrinsic emission map (Johnson et al. 2018). Third, there is intraday variability in the source itself. Source variability is theoretically expected to arise on a broad range of timescales, and it is explicitly seen in GRMHD simulations on timescales as short as minutes (Georgiev et al. 2022). Such variability was also observed in the light curve of Sgr A$^*$ during the 2017 EHT campaign on timescales from 1 minute to several hours (Wielgus et al. 2022).

In this section, we summarize the theoretical expectations for and characteristics of the variability based on GRMHD simulations, present an estimate for the degree of structural variability in Sgr A$^*$ directly from the visibility amplitude data, and describe the strategies pursued here and in Paper III to mitigate the impact of the three components of additional error listed above.

3.1. Expectations from Theory

In low-luminosity SMBH systems such as Sgr A$^*$, we expect the emission to originate from the immediate vicinity of the black hole, i.e., on scales comparable to the event horizon size. Here, all characteristic speeds of the hot relativistic gas approach the speed of light. The timescales associated with these processes are therefore set by the gravitational timescale, $GM/c^2$, which is $\sim 20\,$s for Sgr A$^*$. This timescale is $\sim 3$ orders of magnitude shorter than the nightly observations carried out by the EHT, so a single observation contains many realizations of the underlying source variability.

GRMHD simulations can model the dynamical processes in Sgr A$^*$ and, using ray-tracing and radiative transfer, provide a theoretical expectation for the observed emission. Paper V provides a library of GRMHD simulations and associated movies, which have been scaled to the conditions during the EHT 2017 observations (e.g., the average total 230 GHz flux is set to the EHT measurement). We use the variability characteristics of these simulations as our expectation for the Sgr A$^*$ variability seen by the EHT.

GRMHD simulations are universally described by a “red–red” power spectrum, with the largest fluctuations in the emission occurring on the longest timescales and the largest spatial scales (Georgiev et al. 2022). Spatially, the largest scale for variability is limited to the size of the emitting region,
which for an observing frequency of 230 GHz is typically several \( GM/c^2 \) and for the EHT Sgr A* data is constrained to be \( \lesssim 87 \mu \text{as} \) (Paper II). Temporally, the simulations exhibit a red power spectrum that flattens on timescales \( \gtrsim 1000 \text{GM/c}^2 \). Observations of the total flux variability in Sgr A* corroborate this expectation, finding a red-noise spectrum extending to timescales of several hours and flattening on longer timescales (Wielgus et al. 2022).

We can, without loss of generality, express the time-variable image structure \( I \) in terms of some static mean image \( I_{\text{avg}} \) and a zero-mean time-variable component \( \delta I \) that captures all of the variation,

\[
I(x, t) = I_{\text{avg}}(x) + \delta I(x, t). \tag{5}
\]

The linearity of the Fourier transform ensures that an analogous decomposition holds for \( \mathcal{V} \), which is thus simply the sum of an analogous \( \mathcal{V}_0 \) and \( \delta \mathcal{V} \). The variation \( \delta \mathcal{V} \) represents the component of the data we wish to mitigate.

The EHT stations ALMA and SMA are themselves interferometric arrays capable of separating out extended structure (such as the Galactic center “minispiral”, Lo & Claussen 1983; Goddi et al. 2021; Wielgus et al. 2022) from the Sgr A* light curve,

\[
L(t) = \mathcal{V}(0, t) = \iint I(x, t) d^2 x, \tag{6}
\]

on the largest spatial scales, predicted by GRMHD simulations to be the most variable. Using this motivation, the light-curve-normalized image is defined to be

\[
\hat{I}(x, t) \equiv \frac{I(x, t)}{L(t)}, \tag{7}
\]

with \( \hat{I}_{\text{avg}} \) and \( \delta \hat{I} \) similarly defined; here, the “hat” diacritic denotes light-curve normalization. From GRMHD simulations, the expected noise is well approximated by a broken power law,

\[
\sigma_{\text{var}}^2 \equiv \langle \delta \mathcal{V}^2 \rangle = \frac{a^2 |u|/u_0} {1 + (|u|/u_0)^b} e^c, \tag{8}
\]

along any radial direction (Georgiev et al. 2022). This broken power law is described by four parameters: a break at \( u_0 \), an amplitude \( a \) representing the amount of noise at the break location, and long- and short-baseline power-law indices \( b \) and \( c \), respectively. Typically, we expect that \( c \gtrsim 2 \), due to the compact nature of the source.

In Figure 1, red lines show \( \sigma_{\text{var}}^2 \) measured for an example GRMHD simulation about average images that have been constructed on observationally relevant timescales. The variability has been averaged in azimuth and across different black hole spin orientations. As the timescale over which the average image is constructed increases, the location of the break \( u_0 \) decreases and the amount of power at the break increases.\(^{151}\) This behavior can intuitively be understood as the GRMHD simulations changing less for short timescales. For comparison, we show the thermal, systemic, and refractive scattering noise. For timescales longer than \( \sim 10 \) minutes, the variability noise dominates on EHT VLBI baselines.

\(^{151}\) This behavior is generic across a large number of simulations of Sgr A* and explored in detail in Georgiev et al. (2022).

---

**Figure 1.** The amount of excess variance in the visibility amplitudes expected for a GRMHD simulation (red lines, \( \sigma_{\text{var}}^2 \)), after subtraction of the average image and normalization of the flux density by the light curve. Shown also are the typical thermal noise (black dashed line) and a 1% fractional systematic noise (green band) proportional to the mean image visibility amplitudes. The expected degree of refractive scattering is shown by the purple bands, with purple lines evaluated for a Gaussian source at the projected location of EHT data (see Paper III). The variability is shown about a mean image constructed on different observationally relevant timescales. The fractional systematic and variability noises have been averaged over azimuth and over the position angle of the diffractive screen.

### 3.2. Intraday Variability in the Sgr A* Data

The intraday variability expected from theoretical considerations can be observed directly in the Sgr A* data. Figure 2 shows the combined baseline coverage for the EHT’s 2017 Sgr A* campaign, including the observations on April 5, 6, 7, and 10. The upper limit on the source size of 87 \( \mu \)as (see the second-moment analysis in Paper II) implies that the complex visibilities will be correlated in regions of the \((u,v)\)-plane smaller than \( \sim 2 \) G\( \lambda \). In practice, the visibility amplitudes exhibit variations on scales smaller than this and otherwise appear strongly correlated on scales of 1 G\( \lambda \) (see Paper III, Figure 3). Therefore, among the baseline tracks in Figure 2 there are four regions where the \((u,v)\)-coverage is redundant, i.e., multiple baselines pass within 1 G\( \lambda \) of the same \((u,v)\)-position. We separate the redundant baseline combinations into “crossing tracks,” in which two baseline tracks intersect at a single \((u,v)\)-point, and “following tracks,” in which two baselines follow a nearly identical extended track in the \((u,v)\)-plane. Both sets of redundant baselines provide an opportunity to directly probe the degree of intraday variability in the visibilities at specific locations in the \((u,v)\)-plane.

Prior to making comparisons, we apply the data preprocessing steps outlined in Section 2.3 to mitigate unphysical sources of variability. To avoid addressing the unknown atmospheric phase delays, we focus exclusively on visibility amplitudes. Because source structure will produce additional variations in the visibility amplitudes that are hard to visualize in projection and obscure the relative degree of variability, we detrend the visibility amplitudes with a linear model. The crossing and following tracks discussed below are shown in the top and bottom subpanels of Figure 2, respectively.

**Chile–PV versus Chile–SPT:** The first crossing track we consider contains baselines between the Chile stations (ALMA, APEX) and PV and SPT, which both cross near \((u,v) = (4 \text{ G}\lambda, 3.5 \text{ G}\lambda)\) at times separated by 6.2 hr. The concurrent ALMA and APEX baselines are consistent within the reported statistical errors, and thus there is no evidence for unaddressed
baseline-specific dominant systematic errors. The normalized visibility amplitudes for the Chile–PV and Chile–SPT baseline individually vary smoothly with time. Nevertheless, they differ significantly at the crossing point, and this difference is consistent in magnitude with the variation found across days (indicated by the gray band in the relevant panel of Figure 2).

**Chile–SMT versus Chile–SPT:** The second crossing track we consider contains baselines between the Chile stations (ALMA, APEX) and SMT and SPT, which both cross near \((u, v) = (3 \text{ G\lambda}, 4.5 \text{ G\lambda})\) at times separated by 5.2 hr. Again, we find excellent agreement between ALMA and APEX baselines, individually smooth variations on the Chile–SMT and Chile–SPT baselines, and significant differences in the visibility amplitudes between those baselines.

**SMA–SPT versus LMT–SPT:** The first following track we consider contains baselines between the SPT, which is located at the South Pole, and SMA and LMT, which have similar latitudes. Because the baseline tracks are coincident across a large range of locations in the \((u,v)\)-plane, this following track permits many direct comparisons at a baseline length of 8 G\lambda at times separated by 3.4 hr. As with both crossing tracks, significant differences exist between the two sets of baselines, consistent with the range across multiple days.

**SMT–SPT versus PV–SPT:** The final following track we consider again involves the SPT, and now the SMT and PV, which also have similar latitudes. This is the longest set of baselines that we consider, with a length of roughly 8.5 G\lambda and covering similar regions in the \((u,v)\)-plane at times separated by 6.7 hr. Again, significant variations are exhibited, consistent with those across days.

In summary, intraday variability is observed on multiple baselines with lengths ranging from 5 to 8.5 G\lambda and on timescales as short as 3.4 hr. In all cases, this variability is broadly consistent with that observed on interday timescales. Furthermore, the variability behavior is consistent with theoretical expectations from GRMHD simulations and empirical expectations from the Sgr A* light curve, both of which imply that the variable elements of the Sgr A* emission should be uncorrelated beyond a timescale of a few hours (Georgiev et al. 2022; Wielgus et al. 2022). Any average image of Sgr A* reconstructed from data spanning a time range longer than several hours captures the long-timescale asymptotic source structure; the intrinsic image averaged over a single day or multiple days is thus expected to exhibit similar structure.

### 3.3. Model-agnostic Variability Quantification

To quantify the variability observed in the EHT Sgr A* data, we make use of the procedure described in Broderick et al. (2022). This procedure provides an estimate of the excess variability—i.e., the visibility amplitude variance in excess of that caused by known sources, such as average source structure, statistical and systematic uncertainties, and scattering—as a function of baseline length. We apply the same data preparation steps summarized in Section 3.2 and described in Broderick et al. (2022), combining the visibility amplitudes measured on April 5, 6, 7, and 10 in both observing bands. All data points are weighted equally.

The procedure is illustrated in Figure 3. We again make use of the strong correlations induced by the finite source size, and for every location in the \((u,v)\)-plane we consider only those data points falling within a circular region of diameter 1 G\lambda centered at that point (red circular region in the top panel of Figure 3). Within each such region containing at least three data points, we linearly detrend the light-curve-normalized visibility amplitudes with respect to \(u\) and \(v\) to remove variations due to physical structure (bottom panel of Figure 3), and we compute the variance of the residuals. This variance is
then debiased to remove the contributions from the reported statistical errors, as described in Broderick et al. (2022). Finally, the variances from all regions having a common baseline length are averaged to produce an azimuthally averaged set of variances. The uncertainty in the variance estimates is obtained via Monte Carlo sampling of the unknown gains, leakage terms, and statistical errors.

Figure 4 shows the results of applying this procedure to the Sgr A* data, with the normalized visibility amplitude variance measurements given by the black points. For baselines shorter than 2.5 Gλ, the LMT calibration procedure precludes an accurate estimate of the variance, and thus these baselines have been excluded. For baselines between ∼2.5 and 6 Gλ in length, our empirical estimates of the noise exceed the typical contributions from statistical errors and refractive scattering, indicating the presence of an additional source of structural variability. The degree of inferred variability is consistent with that seen in prior millimeter-VLBI data sets, which is discussed further in Appendix A. For baselines longer than 6 Gλ, our measurements are consistent with the degree of variability expected from the statistical uncertainties in the data; we thus do not directly constrain the source variability on these long baselines.

To characterize the variability behavior within the (u,v)-plane, we fit a broken power law of the form in Equation (8) to the normalized variance measurements. As indicated by the filled black circles in Figure 4, significant measurements exist only in the range of baselines with lengths ∼2.5–6 Gλ; on baselines longer than 6 Gλ, we are unable to distinguish the variability from its associated measurement uncertainties. We thus perform the broken power-law fit only to the ∼2.5–6 Gλ range of baselines, where we have significant measurements, and we find no evidence for a break in the power law in this region. As a result, only an upper limit can be placed on $u_0$, and we are not able to constrain the short-baseline power-law index, c. The range of permitted broken power-law fits is illustrated in Figure 4 by the orange shaded region, with several samples from the posterior distribution explicitly plotted as orange lines.

Because the location of the broken power-law break is poorly constrained, the parameters $u_0$ and a (describing the
location of the break and the amplitude of the power law at the break, respectively) are strongly correlated and highly uncertain. However, it is clear from the orange shaded region in Figure 4 that there is only a narrow range of variances permitted over the 2.5–6 Gλ range of baselines over which the data are constraining. We thus choose to characterize the amplitude of the excess variability noise at |u| = 4 Gλ, which we denote as $a_4$. Joint posteriors for $a_4$, the break location $u_0$, and the long-baseline power-law index $b$ are shown in Figure 5. These constraints are used to inform the prior distributions for the full-track geometric modeling described in Section 7; the associated prior ranges on each parameter are indicated by the purple shaded regions in Figure 5.

3.4. Description of Variability Mitigation Approaches

Having established the existence of structural variability and quantified its magnitude in the Sgr A* data, we now turn to strategies for mitigating its impact on downstream analyses. We employ the light-curve-normalized visibility data, which eliminates large-scale variations and correlations by construction. In principle, there are four methods that we might pursue to address the remaining structural variability:

1. Analyze time-averaged data products.
2. Employ explicitly time-variable models.
3. Analyze short time segments of the data and combine the results afterward to characterize the average source structure.
4. Simultaneously reconstruct the average source structure and a statistical characterization of the structural variability.

The first of these options is complicated substantially by the uncertain visibility phases, which limit our ability to coherently average the data on timescales longer than several minutes. The second option can be employed either when a descriptive low-dimensional model for the source structure can be constructed (e.g., Miller-Jones et al. 2019; Kim et al. 2020) or when there is sufficient $(u,v)$-coverage for nonparametric dynamical imaging algorithms to be successful (e.g., Johnson et al. 2017; Bouman et al. 2018; Arras et al. 2022). The latter approach is explored in the dynamical imaging analyses described in Paper III, ultimately demonstrating that the Sgr A* $(u,v)$-coverage is insufficient to permit unambiguous reconstructions of the variable source structure.

We dub the third option “snapshot” modeling, whereby a simple geometric model of the source structure is fit to segments of the data that are short enough in duration ($\lesssim$ 3 minutes) for the impact of structural variability to be subdominant to other sources of visibility uncertainty (e.g., refractive scattering; see Figure 1). Though the data sparsity is exacerbated by restricting the reconstructions to only a single snapshot at a time, the model itself is also correspondingly restricted in its parameterization of the source structure. The results of the fits to each individual snapshot are then combined across the entire data set, effectively averaging over the source variability. Details of our snapshot modeling analyses as applied to Sgr A* are presented in Section 6.

The fourth option we refer to as “full-track” modeling, which aims to simultaneously reconstruct both the average source structure and a set of parameters describing the contribution of the structural variability to the visibility data variances (Broderick et al. 2022). In contrast to the snapshot modeling, full-track modeling considers the entire data set at once and uses a parameterized “variability noise” model to appropriately modify the data uncertainties as part of the fitting procedure. In this way, the full-track modeling retains access to sufficient $(u,v)$-coverage to permit fitting a nonparametric image model to the data (see Paper III), though in Section 7 we also pursue full-track geometric modeling to provide a cross-comparison with the results from the snapshot geometric modeling. Our parameterization of the variability noise follows Equation (8), with the amplitude specified at a baseline length of 4 Gλ as described in Section 3.3. A detailed description of our full-track modeling approach as applied to Sgr A* is presented in Section 7.

Both the snapshot and full-track modeling approaches focus on describing the average source structure and treating the structural variability in a statistical manner. This goal is formally mismatched with what the EHT data measure for a single day, which is instead a collection of complex visibilities that sample different instantaneous realizations of the intrinsic Sgr A* source structure. The nature of this mismatch impacts the full-track analyses significantly.

The variability mitigation scheme employed by full-track modeling presumes that the variability may be modeled as excess uncorrelated fluctuations in the complex visibility data. This assumption is well justified on timescales exceeding a few hours, but significant correlations between visibilities exist on shorter timescales. Within a single day, subhour correlations that are localized in the $(u,v)$-plane can induce significant biases in the source structure reconstructed from the sparse EHT $(u,v)$-coverage of Sgr A*. The noise model is thus fundamentally misspecified for EHT data, with the level of misspecification increasing as shorter-in-time segments of data are analyzed; Appendix B describes pathological behavior that can arise when analyzing EHT Sgr A* data from only a single day. While
prominent artifacts associated with these subhour correlations are present in the April 6 reconstructions shown in Appendices B and C, we note that the underlying origin of these artifacts is no less present on April 7.

The impact of unmodeled correlations on the reconstructed source structure can be ameliorated by combining multiple days, which provides visibility samples associated with independent realizations of the source structure. This additional sampling rapidly brings the statistical properties of the data into better agreement with the assumptions underpinning the full-track analyses; even the combination of just 2 days is often sufficient to mitigate the subhour correlations in analysis experiments that make use of GRMHD simulation data. For this reason, we combine both the April 6 and April 7 data sets individually during the analysis of the Sgr A* data. For comparison, Appendix C presents the results of equivalent analyses applied to the April 6 and April 7 data sets individually.

4. Ring Characterization and Calibration

We have a strong prior expectation—from both prior millimeter-VLBI observations of a different black hole (i.e., the EHT images of M87*; see M87* Paper IV) and theoretical simulations of the accretion flow around Sgr A* itself (see Paper V) —that Sgr A* ought to contain a ring of emission, and we thus aim to determine the characteristics of the ring-like image structure that best describes the Sgr A* data. In this section, we first review the evidence from the Sgr A* data for a ring-like image structure, and we then present a geometric model for fitting parameters of interest and describe our procedure for bringing ring size measurements made using different techniques to a common physical scale.

4.1. Evidence for a Ring

In reconstructing images of Sgr A*, Paper III explores a large space of imaging algorithms and associated assumptions. The resulting “top sets” of images contain primarily “ring-like” image structures, though a small fraction of the images are morphologically ambiguous. These “nonring” images still nominally provide a reasonable fit to the Sgr A* data and so are not ruled out from the Paper III results.

We can quantify the preference for a ring-like image structure by fitting the data with a set of simple geometrical models. Employing the snapshot geometric modeling technique detailed in Section 6, we compare the Bayesian evidence, $\mathcal{Z}$, between these different geometric models. The value of $\mathcal{Z}$ serves as a model comparison metric that naturally balances improvements in fit quality against increases in model complexity, with larger values of $\mathcal{Z}$ indicating preferred models (see, e.g., Trotta 2008).

Figure 6 shows the results of a survey over simple geometric models with varying complexity, captured here by the number of fitting parameters required to specify the model. At all levels of complexity, ring-like models outperform the other tested models. This disparity is most stark for the simplest models but continues to hold as the models increase in complexity.

The remainder of this paper proceeds with analyses that presuppose a ring-like emission structure for Sgr A*.

4.2. Salient Features in the Context of a Ring Model

The overall structure of the Sgr A* visibility amplitudes (see the left panel of Figure 7) exhibits at least three distinct regions:

1. A “short-baseline” region containing baselines shorter than $\sim$2 Gλ. The effects of data calibration and preprocessing—particularly the light-curve normalization and LMT calibration procedures (Paper III)—are evident in the unit total flux density and the Gaussian structure of the visibility amplitudes in this region.

2. An “intermediate-baseline” region containing baselines between $\sim$2 and 6 Gλ. The visibility amplitudes in this region exhibit a general rise and then fall with increasing
baseline length, peaking at a flux density $\sim 20\%$ of the total at a baseline length of $\sim 4 \, G\lambda$.

3. A “long-baseline” region containing baselines with lengths in excess of $\sim 6 \, G\lambda$. The visibility amplitudes in this region generally rise with increasing baseline length from a deep minimum near $\sim 6.5 \, G\lambda$, approximately flattening out at longer baselines to a level that is $\sim 3\%–10\%$ of the total flux density.

The visibility amplitudes exhibit indications of asymmetric source structure, particularly on baselines with lengths of $\sim 3 \, G\lambda$ that fall near the first minimum. Here, the baselines between the SMT and Hawai‘i stations (oriented approximately in the east–west direction) have systematically higher correlated flux densities than the similar-length baselines between the LMT and Chile stations (oriented approximately in the north–south direction). The implication for the source morphology is that we would expect to see more symmetric structure in the north–south than in the east–west direction. Detailed geometric modeling analyses that are able to capture this asymmetry are described in Sections 6 and 7; here, we consider only a simple azimuthally symmetric toy model that captures some salient features of interest.

We attempt to understand the visibility behavior in light of expectations for a ring-like emitting structure. Specifically, we consider a geometric construction whereby an infinitesimally thin circular ring bordering an inner disk of emission is convolved with a Gaussian blurring kernel. The visibility function $V$ produced by such an emission structure is given by:

$$V = F_0 V_{\text{Gauss}} [J_d V_{\text{disk}} + (1 - f_d) V_{\text{ring}}]$$

$$= F_0 \exp \left( -\frac{w^2 \xi^2}{4 \ln(2)} \right) \left[ \frac{2 f_d J_1(\xi) + (1 - f_d) J_0(\xi)}{\xi} \right],$$

(9)

where $F_0$ is the total flux in the image, $f_d$ is the fraction of that flux that is contained in the disk component, $w = W/d$ is a fractional ring width, $W$ is the FWHM of the Gaussian convolving kernel, $d$ is the diameter of the ring and disk components, $\xi \equiv \pi u/d$ is a normalized radial visibility-domain coordinate, and $J_n(\xi)$ is a Bessel function of the first kind of order $n$.

The three regions of Sgr A* data identified above are separated by apparent minima in the visibility amplitudes, and they can be approximately characterized by the baseline locations of those minima and the peak flux density levels achieved at the visibility maxima between them. Figure 7 illustrates how this characterization manifests as constraints on the defining parameters of the geometric toy model. The cyan and purple shaded regions in the left panel indicate the approximate ranges of baseline lengths corresponding to the locations of the first and second visibility minima, respectively. The locations of these minima constrain the diameter of the emitting structure, as shown in the top right panel of Figure 7. To be consistent with both a first minimum falling between $\sim 2.5$ and $3.5 \, G\lambda$ and a second minimum falling between $\sim 6$ and $7 \, G\lambda$, the emitting region must be between $\sim 50$ and $60 \, \mu\text{as}$.
across. The amplitudes of two visibility maxima—one falling between the first and second visibility minima, and the second following the second minimum—constrain a combination of the fractional disk flux $f_d$ and the fractional ring width $w$. The bottom right panel of Figure 7 shows the constraints from the first and second visibility maxima in red and orange, respectively, and from the ratio of the two in green.

Taken together, even these few, simple, and only modestly constrained visibility features result in a rather narrow permitted range of model parameter values for $d$, $w$, and $f_d$; an example of a “best-fit” model from within the permitted range is shown by the gray curve in the left panel of Figure 7. However, we stress that the above constraints only strictly hold within the context of the specific toy model used to derive them. More general and robust constraints on the emission structure require a model that can accommodate more than just the gross features; such models are produced as part of the imaging (Paper III and Section 5) and geometric modeling analyses (see Sections 6 and 7) carried out in this paper series.

4.3. Geometric Ring Model Specification

The ring-like images reconstructed in Paper III are not azimuthally symmetric, but instead show pronounced azimuthal brightness variations that we would like to capture in our geometric modeling analyses. In this section, we specify the “mG-ring” model that we use in Sections 6 and 7 to quantify the morphological properties of the observed Sgr A* emission.

4.3.1. Image-domain Representation of mG-ring Model

Adopting the construction developed by Johnson et al. (2020), we can model an infinitesimally thin circular ring with azimuthal brightness variations using a sum over angular Fourier modes indexed by integer $k$,

$$I_{\text{ring}}(r, \phi) = \frac{F_{\text{ring}}}{\pi d} \delta \left( r - \frac{d}{2} \right) \sum_{k=-m}^{m} \beta_k e^{i\beta_k}.$$

(10)

Here $r$ is the image radial coordinate, $\phi$ is the azimuthal coordinate (east of north), $d$ is the ring diameter, $\{\beta_k\}$ are the set of (dimensionless) complex azimuthal mode coefficients, and $m$ sets the order of the expansion. Because the image is real, $\beta_k = \beta_k^*$; we enforce $\beta_0 \equiv 1$ so that $F_{\text{ring}}$ sets the total flux density of the ring. Given that the images from Paper III show a ring of radius $\sim 25$ $\mu$m and the diffraction-limited EHT resolution is $\sim 20$ $\mu$m, we expect the data to primarily constrain ring modes with $m \lesssim \pi (25/20) \approx 4$. We refer to this asymmetric ring as an “m-ring” of order $m$.

For the purposes of constraining additional image structures, we augment this m-ring in two ways. First, we convolve the m-ring with a circular Gaussian kernel of FWHM $W$,

$$I_{\text{ring}}(r, \phi; W) = I_{\text{ring}}(r, \phi) * \left[ \frac{4 \ln(2)}{\pi W^2} \exp \left( -\frac{4 \ln(2) r^2}{W^2} \right) \right].$$

(11)

Second, we add a circular Gaussian component that is concentric with the ring, which serves to provide a nonzero brightness floor interior to the ring. The Gaussian component has a total flux density of $F_{\text{Gauss}}$ and an FWHM of $W_{\text{Gauss}}$,

$$I_{\text{Gauss}}(r, \phi) = \frac{4 \ln(2) F_{\text{Gauss}}}{\pi W_{\text{Gauss}}^2} \exp \left( -\frac{4 \ln(2) r^2}{W_{\text{Gauss}}^2} \right).$$

(12)

We refer to the resulting composite model $I(r, \phi)$, where

$$I(r, \phi) = I_{\text{ring}}(r, \phi; W) + I_{\text{Gauss}}(r, \phi),$$

(13)

as an “mG-ring.” An example mG-ring is shown in Figure 8.

An mG-ring of order $m$ has $5 + 2m$ model parameters: the flux density in the ring ($F_{\text{ring}}$), the diameter of the ring ($d$), the flux density in the central Gaussian ($F_{\text{Gauss}}$), the FWHM of the central Gaussian ($W_{\text{Gauss}}$), the FWHM of the ring convolving kernel ($W$), and two parameters for each complex Fourier coefficient $\beta_k$ with $1 \leq k \leq m$.

4.3.2. Visibility-domain Representation of mG-ring Model

To aid in efficient parameter space exploration, the mG-ring model is intentionally constructed using components and transformations that permit analytic Fourier transformations. The Fourier transform of the m-ring image (Equation (10)) is given by

$$V_{\text{ring}}(|u|, \phi_u; W) = \exp \left( -\frac{\pi^2 W^2 |u|^2}{4 \ln(2)} \right) V_{\text{ring}}(|u|, \phi_u).$$

(14)

The Fourier transform of the Gaussian image (Equation (12)) is given by

$$V_{\text{Gauss}}(|u|, \phi_u) = F_{\text{Gauss}} \exp \left( -\frac{\pi^2 W_{\text{Gauss}}^2 |u|^2}{4 \ln(2)} \right).$$

(15)

By the linearity of the Fourier transform, the visibility-domain representation of the mG-ring model is then simply the sum of these two components,

$$V(|u|, \phi_u) = V_{\text{ring}}(|u|, \phi_u; W) + V_{\text{Gauss}}(|u|, \phi_u).$$

(16)

When interpreting model-fitting results in subsequent sections, we are interested in a number of derivative quantities. We will typically work with the fractional thickness of the ring, $w$, defined to be

$$w \equiv \frac{W}{d}.$$  

(18)

Similarly, we are typically interested in fractional representations of flux densities. We define

$$F_0 \equiv F_{\text{ring}} + F_{\text{Gauss}},$$

$$f_{\text{ring}} \equiv \frac{F_{\text{ring}}}{F_0}$$

(19)

to be the total flux density, and then
and
\[ f_{\text{Gauss}} \equiv \frac{F_{\text{Gauss}}}{F_0} \]  (20)

are the fraction of the total flux density that is contained in the ring and in the Gaussian components, respectively. Note that \( F_0 \) is typically close to or fixed to unity as a consequence of normalizing the data by the light curve. We also define a fractional central flux as
\[ f_c \equiv \frac{F_{\text{Gauss}}(r < d/2)}{F_{\text{ring}} + F_{\text{Gauss}}(r < d/2)}, \]  (21)

where \( F_{\text{Gauss}}(r < d/2) \) is the integrated flux density of the central Gaussian component interior to the ring radius, given by
\[ F_{\text{Gauss}}(r < d/2) = F_{\text{Gauss}} \left[ 1 - \exp \left( -\frac{d^2 \ln(2)}{W_{\text{Gauss}}^2} \right) \right]. \]  (22)

Following M87* Paper IV, the m-ring position angle \( \eta \) and degree of azimuthal asymmetry \( A \) are both determined by the coefficient of the \( m = 1 \) mode,
\[ \eta \equiv \arg \left( \int_0^{2\pi} I(\phi)e^{i\phi} d\phi \right) = -\arg (\beta_1), \]  (23)
\[ A \equiv \frac{\int_0^{2\pi} I(\phi)e^{i\phi} d\phi}{\int_0^{2\pi} I(\phi)e^{i\phi} d\phi} = |\beta_1|. \]  (24)

A number of these derivative quantities are illustrated in the example mG-ring shown in Figure 8.

4.4. Calibrating Ring Size Measurements to a Common Physical Scale

The parameters returned by the geometric modeling and feature extraction analyses used in this paper to describe the Sgr A* emission structure do not correspond directly to physical quantities. Instead, the relationship between measured and physical quantities must be calibrated using data for which we know the correct underlying physical system’s defining parameters. For ring size measurements, the associated physical quantity of interest is related to the angular size of the gravitational radius,
\[ \theta_g = \frac{GM}{c^2 D} \]  (25)

which sets the absolute scale of the system.

Under the assumption that the emission near the black hole originates from some “typical” radius, a measurement of the angular diameter \( d \) of the emitting region will be related to \( \theta_g \) by a scaling factor \( \alpha \),
\[ d = \alpha \theta_g. \]  (26)

If the observations were directly sensitive to the critical curve bounding the black hole shadow, then \( \alpha \) could be determined analytically and would take on a value ranging from \(~9.6\) to \(10.4\) depending on the black hole spin and inclination (Bardeen 1973; Takahashi 2004). For more realistic emission structures and measurement strategies, the value of \( \alpha \) cannot be determined from first principles and must instead be calibrated.

Our \( \alpha \) calibration strategy generally follows the procedure developed in M87* Paper VI. Using the library of GRMHD simulations described in Paper V, we generate a suite of 100 synthetic data sets that emulate the cadence and sensitivity of
the 2017 EHT observations and that contain a realistic character and magnitude of data corruption; Appendix D describes the generation of these synthetic data sets. In the analyses described in Sections 5, 6, and 7, 90 of these 100 synthetic data sets are used to derive the α calibration for each analysis pathway, while the remaining 10 data sets are used to validate the calibration.

After carrying out ring size measurements on each of the data sets in the suite, we determine α (for each specific combination of data set and measurement technique) by dividing the measured ring diameter by the known value of θg (per Equation (26)). For a given measurement technique, the distribution of α values that results from applying this procedure to the entire suite of synthetic data sets then provides a measure of α and its theoretical uncertainty. The α value associated with each measurement technique can then be used to translate Sgr A* ring size measurements into their corresponding θg constraints. We note that this calibration strategy assumes that the images contained in the GRMHD library provide a reliable representation of the emission structure in the vicinity of Sgr A*: a separate calibration strategy that relaxes this GRMHD assumption is presented in Paper VI.

Appendix I describes elements of the calibration and validation strategy that are specific to each of the analysis pathways detailed in Sections 5, 6, and 7.

5. Image-domain Feature Extraction

The imaging carried out in Paper III permits very flexible emission structures to be reconstructed from the Sgr A* data, but the majority of these images exhibit a ring-like morphology whose properties we seek to characterize. In this section, we describe our image-domain feature extraction (IDFE) procedure, which uses a topological classification scheme to identify the presence of a ring-like structure in an image and quantifies the parameters that best describe this ring using two different algorithms. We apply this IDFE procedure to the Sgr A* image reconstructions from Paper III.

5.1. Imaging Methods and Products

The imaging analyses carried out in Paper III use four different algorithms classified into three categories: one sampling-based posterior exploration algorithm (THEMIS; Broderick et al. 2020a, 2020b), one CLEAN-based deconvolution algorithm (DIFMAP; Shepherd 1997), and two “regularized maximum likelihood” (RML) algorithms (eht-imaging, Chael et al. 2016, 2018; SMILI, Akiyama et al. 2017a, 2017b). All methods produce image reconstructions using band-combined data (i.e., both low band and high band), and the latter three are run on two versions of the Sgr A* data: a “descattered” version that attempts to deconvolve the effects of the diffractive scattering kernel from the data, and a “scattered” version that applies no such deconvolution. The posterior exploration imaging method THEMIS instead applies the effects of diffractive scattering as part of its internal forward model, rather than deconvolving the data; the analogous “scattered” and “descattered” versions of the THEMIS images thus correspond simply to those for which the scattering kernel has been applied or not, respectively. The posterior exploration imaging jointly reconstructs the combined April 6 and April 7 data sets (see Appendix B), while the CLEAN and RML imaging reconstructs each day individually, focusing primarily on the April 7 data and using the April 6 data for cross-validation. Example fits and residuals for each of the imaging pipelines are shown in Figure 9, and χ² statistics for each image are provided in Appendix E; detailed descriptions of the data preprocessing and imaging procedures for each imaging algorithm are provided in Paper III.

For the CLEAN and RML imaging methods, there are a number of tunable hyperparameters associated with each algorithm whose values are determined through extensive “parameter surveys” carried out on synthetic data sets. During a parameter survey, images of each synthetic data set are reconstructed using a broad range of possible values for each hyperparameter. Settings that produce high-fidelity image reconstructions across all synthetic data sets are collected into a “top set” of hyperparameters, and these settings are then applied for imaging the Sgr A* data. The resulting top sets of Sgr A* images capture emission structures that are consistent with the data, and we use these top-set images for the feature extraction analyses in this paper.

The THEMIS imaging algorithm explores a posterior distribution over the image structure, and there are no hyperparameters that require synthetic data surveys to determine. Rather than producing a top set of images, THEMIS instead produces a sample of images drawn from the posterior determined from the Sgr A* data. We use these posterior image samples for the feature extraction analyses in this paper.

5.2. Image-domain Feature Extraction Methods

Given the top-set and posterior images from Paper III, we carry out IDFE analyses using two separate tools: REx and VIDA. An independent cross-validation of both IDFE tools has been carried out in P. Tiede et al. (2022, in preparation). In this section, we provide a brief overview of each method and specify the details relevant for the analyses presented in this paper.

5.2.1. REx

The Ring Extractor (REx) is an IDFE tool for quantifying the morphological properties of ring-like images. It is available as part of the eht-imaging software library and is described in detail in Chael (2019). REx was the main tool used in M87+ Paper IV to extract ring properties from the M87+ images, and detailed definitions of the various REx parameters are provided in that paper.

For the majority of the REx-derived ring parameters, we retain the same definitions as used in M87+ Paper IV. REx first defines a ring center (x₀, y₀), which is determined to be the point in the image from which radial intensity profiles have a minimum dispersion in their peak intensity radii. The ring radius, r₀, is then taken to be the average of these peak intensity radii over all angles, and the ring thickness w is taken to be the angular average of the FWHM about the peak measured along each radial intensity profile. To avoid biases associated with a nonzero floor to the image brightness.

---

152 Note that the preprocessing and data products used during imaging are not the same across imaging pipelines; DIFMAP and THEMIS fit to complex visibilities, while eht-imaging and SMILI fit iteratively to different combinations of data products that include visibility amplitudes, closure phases, and log closure amplitudes (see Paper III, for details). For clarity in Figure 9, we simply show residual complex visibilities for each imaging pipeline using a representative image from that pipeline’s top set or posterior.
outside of the ring, we subtract out the quantity

\[ I_{\text{floor}} = \frac{1}{2\pi} \int_0^{2\pi} I(r_{\text{max}} = 60 \, \mu\text{as}, \phi) d\phi \] (27)

when computing the FWHM, i.e., we compute the average FWHM of \( I(r, \phi) - I_{\text{floor}} \). For all other ring parameters, the definitions remain the same as those used in M87* Paper IV.

\text{REx} defines the ring position angle \( \eta \) and asymmetry \( A \) as the argument and amplitude, respectively, of the first circular mode,

\[ \beta_1 = \left( \frac{\int_0^{2\pi} I(\phi) \cos(\phi) d\phi}{\int_0^{2\pi} I(\phi) d\phi} \right) \left( \frac{\int_0^{2\pi} I(\phi) \sin(\phi) d\phi}{\int_0^{2\pi} I(\phi) d\phi} \right) \] (28)

where the angled brackets denote a radial average between \( r_0 - w/2 \) and \( r_0 + w/2 \).\footnote{We note that this definition for the position angle \( \eta \) does not necessarily return the azimuthal location of the intensity peak; rather, it tracks the circular mean of the azimuthal intensity profile.} These definitions are analogous to...
those used to define the corresponding position angle and asymmetry of the mG-ring model (Equations (23) and (24), respectively). The fractional central brightness \( f_c \) is defined to be the ratio of the mean brightness within 5 \( \mu \)as of the center to the azimuthally averaged brightness along the ring (i.e., along \( r = r_0 \)).

As in Paper III, we replace the negative pixels in THEMIS images with zero values before performing REx analyses.

5.2.2. VIDA

Variational Image-Domain Analysis (VIDA; P. Tiede et al. 2022, in preparation) is an IDFE tool for quantifying the parameters describing a specifiable image morphology; it is written in Julia (Bezanson et al. 2017) and contained in the package VIDA.jl.\footnote{https://github.com/ptiede/VIDA.jl} VIDA employs a template-matching approach for image analysis, using parameterized templates to approximate an image and adjusting the parameters of the templates until a specified cost function is minimized. Within VIDA, the cost function takes the form of a probability divergence, which provides a distance metric between the image and template; the template parameters that minimize this divergence are taken to provide the best description of the image. The VIDA optimization strategy and additional details are provided in P. Tiede et al. (2022, in preparation).

For the IDFE analyses in this paper, we use VIDA’s SymCosineRingFloor template and the least-squares divergence (for details, see Section 8 of Paper III). This template describes an image structure that is similar to the mG-ring model (Section 4.3), and it is characterized by a ring center \((x_0, y_0)\), a ring diameter \(d = 2r_0\), an FHWM fractional ring thickness \(w\), and a cosine expansion describing the azimuthal brightness distribution \(S(\phi)\),

\[
S(\phi) = 1 - 2 \sum_{k=1}^{m} A_k \cos[k(\phi - \eta_k)].
\]  

To maintain consistency with the geometric modeling analyses (see Sections 6 and 7), we use \(m = 4\). We also restrict the value of the \(A_1\) parameter to be \(<0.5\) to avoid negative flux in the template. As with the mG-ring model, the orientation \(\eta\) is equal to the first-order phase \(\eta_1\), and the asymmetry \(A\) is equal to the first-order coefficient \(A_1\).

To permit the presence of a central brightness floor, the SymCosineRingFloor template contains an additional component in the form of a circular disk whose center point is fixed to coincide with that of the ring. The disk radius is fixed to be \(r_0\). A Gaussian falloff is stitched to the outer edge of the disk, such that for radii larger than \(r_0\) the intensity profile becomes a Gaussian with mean \(r_0\) and an FHWM that matches the ring thickness. The flux of this disk component is a free parameter in the template. We then retain the same definition of the fractional central brightness \(f_c\) as used by REx.

5.3. Identifying Rings via Topological Classification

The output of the IDFE analysis is a set of distributions for the ring parameters from each imaging method; Figure 10 shows an example set of results from applying both IDFE software packages to the descattered Sgr A* posterior and top-set images. However, both REx and VIDA implicitly assume that the images fed into them contain a ring-like emission structure. If the input image does not contain a ring, then the output measurements may not be meaningful. For each input image, we thus wish to determine both whether the image contains a ring-like structure and how sensitive the IDFE

---

Figure 10: Morphological parameter distributions from IDFE analyses, applied to the descattered Sgr A* top-set and posterior images corresponding to the combined LO+HI band data from the HOPS calibration pipeline. The distributions shown correspond to combined April 6+7 results for posterior imaging and April 7 data for top-set imaging. No metronization-based filtering has been applied.
results are to the specific manner in which “a ring-like structure” is defined.

To determine whether the images we are analyzing with REx and VIDA contain ring-like structures, we use metronization\(^{155}\), a software that preprocesses the images into a form suitable for topological analysis and extracts topologically relevant features with the help of the open-source computational topology code Dionysus 2\(^{156}\) (Morozov 2017). A detailed description of metronization can be found in Christian et al. (2022).

The metronization preprocessing procedure consists of the following steps:

1. First, the image undergoes a “robust” thresholding step, in which the pixels are sorted by brightness in a cumulative sequence, and all pixels below a certain threshold in this sequence have their values set to zero and the rest are set to a value of one.
2. Next, in a process called “skeletonization,” the Boolean image produced in the first step is reduced to its topological skeleton that preserves the topological characteristics of the original shape. This step thins large contiguous areas of flagged pixels and enlarges the “holes.”
3. The topological skeleton is rebinned and downsampled. Holes smaller than the rebinning resolution are preserved by the skeletonization in the previous step.
4. The downsampled image undergoes skeletonization once more.

The resulting output is a low-resolution image that preserves the topologically relevant information from the original image, speeding up the application of computationally expensive topological algorithms that follow. A technique known as persistent homology is then used to convert this low-resolution image into a topological space that preserves features that are topologically invariant. It computes a quantity known as the first Betti number that provides a metric for measuring the number of holes present in the image.

The metronization software contains a number of tunable parameters that determine how closely the emission structure in the input image must resemble that of a topological ring, and for how many cumulative threshold levels it must persist, for it to be classified as a ring. We identify three modes for these parameters—a “permissive” mode, a “moderate” mode, and a “strict” mode—and explore the impact on the REx and VIDA measured parameter distributions when the input top sets and posterior images are restricted only to those that are classified as containing rings. We compare these results with those of a fourth, default setting in which the top sets and posterior images are not filtered by the classification prescribed by metronization.

We note that metronization differs from the ring identification methods presented in Paper III in that it searches for the presence of a topological ring in the input image. Figure 11 compares the mean ring and nonring descattered images for each imaging pipeline as classified by metronization in the “permissive” mode and the clustering analysis from Paper III. Both methods classify all the posterior imaging samples as rings, while the top-set imaging samples contain both ring and nonring images. We find that the mean ring and nonring images for each imaging pipeline are broadly consistent between the two classification methods.

The definition of what constitutes a ring is subjective, and there will always be images that are ambiguous to the human eye. Different automated methods will classify these images differently. Hence, it is important to verify that the ring parameters measured by REx and VIDA are robust against the specifics of the ring identification scheme used. Figure 12 shows the resulting diameter distributions from ring fitting to the descattered Sgr A* images from all imaging pipelines, split out by metronization setting. As we move from the most to the least permissive classification scheme, the tails in the distributions are diminished while the primary peaks are sharpened, but the mean and general shape of the distribution remain largely unchanged. This trend indicates that while metronization penalizes images with emission structures deviating from a topological ring, the distributions of the REx and VIDA measurements are robust against the choice of metronization mode employed.

6. Snapshot Geometric Modeling

Because the Sgr A* data are observed to be time-variable (see Section 3), a static model cannot reproduce the observed data. As described in Section 3.4, one method for mitigating the effects of this variability on the reconstructed source structure...
is through the use of an inflated variability noise budget, as pursued during the imaging (Paper III: Section 5) and full-track geometric modeling (Section 7) analyses. In this section we instead pursue “snapshot” geometric modeling, whereby we fit a geometric model—the mG-ring model described in Section 4.3—for which the parameters are allowed to vary as a piecewise constant function of time. To this end, we divide the Sgr A* data into many independent and short “snapshots” over which the source is assumed to be static. In this section, we detail our formalism for fitting the mG-ring model to snapshots of data and for combining the fits from across snapshots into a global posterior distribution.

### 6.1. Data Preparation

Prior to fitting the mG-ring model to real or synthetic data, we process the data using the pre-imaging pipeline described in Paper III. This preprocessing procedure entails light-curve normalization and an inflation of the error budget to account for residual calibration uncertainties and the effects of refractive scattering in the interstellar medium toward Sgr A*. Specifically, the total error budget \( \sigma_{sb} \) for a visibility measured on the baseline \( b \) during snapshot \( s \) is given by

\[
\sigma_{sb}^2 = \sigma_{fn, sb}^2 + f^2 |V|_b^2 + \sigma_{ref, sb}^2
\]  

(30)

Here the first term corresponds to the baseline-specific thermal noise (see Equation (2)), the second term is a component that is multiplicative in the visibility amplitude and is intended to capture residual (nongain) calibration errors (e.g., residual polarization leakage), and the third term is the J18modell1 refractive scattering noise from Paper III. For the snapshot modeling, we fix \( f = 0.02 \) per the analyses carried out in Paper II. The pre-imaging pipeline also mitigates the impact of diffractive scattering by “deblurring” the data using the Johnson et al. (2018) model.

Following the application of the pre-imaging pipeline, we split the data into 120 s segments, or “snapshots,” and coherently average the visibilities in each snapshot over the 120 s window. Finally, we flag snapshots that contain fewer than four unique stations, so as to retain snapshots during which closure amplitudes can be formed.

### 6.2. Snapshot Fitting Procedure

The first step of our snapshot modeling procedure is to determine the posterior distribution for the mG-ring model parameters on each snapshot of data. The observation is divided up into \( N_s \) independent snapshots, which we label using a snapshot index \( s \). Within each snapshot we fit the mG-ring model described in Section 4.3, whose parameter vector we denote as \( \theta_s \). For a single snapshot, the posterior is given by Bayes’s theorem,

\[
P(\theta_s|D_s) \propto \frac{L_s(D_s|\theta_s)\pi_s(\theta_s)}{Z_s(D_s)},
\]

(31)

where \( D_s \) denotes the data available on snapshot \( s \), \( L_s \) is the likelihood, \( \pi_s \) is the prior distribution, and \( Z_s \) is the Bayesian evidence.

In our snapshot modeling analyses we make use of three different classes of interferometric data products: visibility amplitudes \( |V| \), log closure amplitudes \( \ln A \), and closure phases \( \psi \). Each analysis uses only a single amplitude data product (either visibility amplitudes or log closure amplitudes) along with the closure phases. For analyses that use visibility amplitudes and closure phases, the likelihood is given by

\[
L_s = L_{|V|, s}L_{\psi, s},
\]

(32)

while for those that use log closure amplitudes, we instead have

\[
L_s = L_{A, s}L_{\psi, s}.
\]

(33)

Here \( L_{|V|, s} \), \( L_{\psi, s} \), and \( L_{A, s} \) are components of the likelihood on snapshot \( s \) associated with the visibility amplitudes, closure phases, and log closure amplitudes, respectively. We assume Gaussian likelihood functions for the amplitude data components and a von Mises likelihood function for the closure phases; the detailed expressions for each likelihood function are provided in Appendix F.

### 6.3. Averaging the Snapshot Results

The output of a snapshot fitting analysis is a set of posterior samples for the model parameters from each individual snapshot; Figure 13 shows an example set of posterior distributions for the mG-ring diameter parameter on each snapshot in the April 6 and April 7 data sets. To arrive at a single posterior on these parameters that combines the information from all snapshots across both days, we use a Bayesian hierarchical model similar to the one used in Baronchelli et al. (2020). This approach treats the model fit to each snapshot as a realization from some average model or “hypermmodel.”
is the data vector across all snapshots. denotes the density for a truncated normal distribution, where \( q \) is the prior distribution for the hypermodel parameters (the “hyperprior”), \( \Theta = (\theta_1, \theta_2, \ldots, \theta_N) \) is the parameter vector across all snapshots, and \( \mathcal{D} = (D_1, D_2, \ldots, D_N) \) is the data vector across all snapshots.

To find the marginal average parameter posterior, we integrate this expression over all the snapshot parameters,

\[
P(\bar{\Theta}|\mathcal{D}) = \frac{\pi(\bar{\Theta})}{Z(\mathcal{D})} \int \prod_s L_s(D_s|\theta_s) \pi_s(\theta_s|\bar{\Theta}) d\Theta.
\]  

In general, this integral is analytically intractable. However, a bit of manipulation permits us to use the posterior samples from the individual snapshot fits to make headway. Because the snapshots are independent, we can swap the order of the integral and product in Equation (35) and use Bayes’s theorem to substitute in for the snapshot likelihood (Equation (31)), giving

\[
P(\bar{\Theta}|\mathcal{D}) = \pi(\bar{\Theta}) \prod_s \left( \int P(\theta_s|D_s) \frac{\pi_s(\theta_s|\bar{\Theta})}{\pi(\theta_s)} d\theta_s \right).
\]  

Note that the evidence term from the prefactor denominator in Equation (35) has now been subsumed into the posterior term \( P(\theta_s|D_s) \) inside of the integral. To evaluate Equation (36), we make use of the fact that the snapshot posterior samples \( \theta_s^{(i)} \) permit us to approximate the integral by a sum,

\[
P(\bar{\Theta}|\mathcal{D}) \approx \pi(\bar{\Theta}) \prod_s \sum_i \frac{\pi_s(\theta_s^{(i)}|\bar{\Theta})}{\pi(\theta_s^{(i)})}.
\]  

We can use this expression to sample from the posterior distribution over just the hypermodel parameters \( \bar{\Theta} \), having fully marginalized over the parameters from each individual snapshot.

We note that the averaging procedure described here is simply a generalization of standard inverse-variance weighting. If we consider a delta-function hypermodel that contains only a single parameter (i.e., the to-be-determined mean value) for each snapshot model parameter, then in the limit where the individual snapshot posteriors \( P(\theta_s|D_s) \) are Gaussian and the priors on \( \bar{\Theta} \) and \( \theta_s \) are uninformative, the posterior maximum for \( P(\bar{\Theta}|\mathcal{D}) \) is equal to the mean of the snapshot posterior means weighted by their inverse posterior variances. However, because the model we employ (described in the following section) does not conform to the necessary conditions (see, e.g., the non-Gaussian snapshot posteriors shown in Figure 13), we proceed with the more general averaging procedure.

### 6.3.2. Hypermodel Specification

We now need to specify the hypermodel \( \pi(\theta_s|\bar{\Theta}) \) that determines the distribution from which the individual snapshot models are drawn; for simplicity, we choose a hypermodel that is approximately Gaussian. Let \( \bar{\Theta} = (\mu, \sigma) \), where \( \mu \) is a vector of the mean parameter values and \( \sigma \) is a vector containing their standard deviations across scans. We assign most hypermodel parameters to be distributed according to a truncated normal distribution,

\[
\pi_{\text{NS}}(\theta_s^{(i)}|\mu, \sigma) = T(\theta_s^{(i)}|\mu, \sigma; a_i, b_i). \tag{38}
\]

Here \( T(\theta|\mu, \sigma; a, b) \) denotes the density for a truncated normal distribution with mean \( \mu \) and standard deviation \( \sigma \), and whose lower and upper bounds are given by \( a \) and \( b \), respectively; we index the separate parameters by \( i \). This truncation is necessary to ensure that the support of the hypermodel parameters matches that of the individual snapshot model parameters. However, for angular parameters—i.e., those with values that are periodic in [0,
Table 1  

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Comrade</th>
<th>DPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_0$</td>
<td>$U(0.8, 1.2)$</td>
<td>$U(1)$</td>
</tr>
<tr>
<td>$d$ (mas)</td>
<td>$U(25, 85)$</td>
<td>...</td>
</tr>
<tr>
<td>$d'$ (mas)</td>
<td>...</td>
<td>$U(25, 85)$</td>
</tr>
<tr>
<td>$W$ (mas)</td>
<td>$U(1, 40)$</td>
<td>...</td>
</tr>
<tr>
<td>$W'$ (mas)</td>
<td>...</td>
<td>$U(1, 40)$</td>
</tr>
<tr>
<td>$\beta_{\text{rad}}$</td>
<td>$U(0, 0.5)$</td>
<td>$U(0, 0.5)$</td>
</tr>
<tr>
<td>$\phi_\text{ang}$ (deg)</td>
<td>$U(-180, 180)$</td>
<td>$U(-180, 180)$</td>
</tr>
<tr>
<td>$J_{\text{max}}$</td>
<td>$U(0, 1)$</td>
<td>$U(0, 1)$</td>
</tr>
<tr>
<td>$W_{\text{max}}$ (mas)</td>
<td>$U(40, 200)$</td>
<td>$U(40, 200)$</td>
</tr>
</tbody>
</table>

Note. Prior distributions for Comrade and DPI snapshot geometric modeling analyses. $U(a, b)$ denotes a uniform prior on the interval $[a, b]$, and $\delta(a)$ denotes a delta-function (i.e., fixed-value) prior, with the parameter value fixed at $a$. For the definitions of the parameters see Section 4.3.

We instead use a von Mises distribution,

$$
\pi_{\text{vM}}(\theta_{ij} | \mu_j, \sigma_j) = \frac{1}{2\pi I_0(\sigma_j^{-2})} \exp \left[ \frac{\cos(\theta_{ij} - \mu_j)}{\sigma_j^2} \right],
$$

(39)

The subscripted “N” in Equation (38) indicates that the corresponding parameters use a truncated normal prior, while the subscripted “vM” in Equation (39) similarly indicates that the corresponding parameters use a von Mises prior. The total hypermodel is then given by

$$
\pi(\theta | \mu, \sigma) = \prod_{i} \pi_{\text{vM}}(\theta_{ij} | \mu_j, \sigma_j) \prod_{f} \pi_{\text{tN}}(\theta_{fj} | \mu_j, \sigma_j),
$$

(40)

where the first product runs over nonangular parameters and the second runs over angular parameters.

We set the hyperpriors for $\mu$ to be equal to the corresponding snapshot priors, which are specified in Table 1. For $\sigma$ we instead use a half-normal hyperprior,

$$
\pi(\sigma) = \prod_i T(\sigma | 0, L_i/4; 0, L_i),
$$

(41)

where $L_i = (b_i - a_i)$ is the breadth of support for parameter $\theta_{i,j}$. Appendix G describes the level of consistency between these selected hyperpriors and the priors for the individual snapshot model parameters.

6.4. Software Implementations

We use three different software packages to carry out snapshot geometric modeling on the Sgr A$^*$ data and a fourth software to perform the hypermodel sampling. In this section, we specify the relevant implementation specifics for these different tools. Cross-validation tests are detailed in Appendix H.

6.4.1. Comrade

Our primary snapshot fitting software is the modeling framework Comrade (P. Tiede 2022, in preparation), which is written in the dynamic programming language Julia (Bezanson et al. 2017). Comrade does not natively include functionality for constructing a joint probability describing both observations and model. Instead, it interfaces with existing probabilistic programming languages present in Julia. For the analyses presented in this paper, we use the probabilistic programming package Soss57 to construct the joint probability. This interface is specified in the package ComradeSoss.jl.158 To sample from the posterior, we use the nested sampling package dynesty, which also produces estimates of the Bayesian evidence (Speagle 2020).

Given a model specification, Comrade can fit a variety of interferometric data products, including visibility amplitudes, closure phases, and log closure amplitudes. Unless otherwise specified, for the snapshot modeling analyses performed in this paper, we use Comrade to fit to visibility amplitudes and closure phases to the mG-ring model; the snapshot likelihood is thus given by Equation (32). Prior to fitting, all time stamps that contain fewer than four baselines are flagged.

When fitting to visibility amplitudes, we include the station gain amplitudes as model parameters alongside the geometric parameters that describe the mG-ring model. For the gain amplitudes we use a lognormal prior with a log-mean of zero (i.e., corresponding to unit gain amplitude) and a log standard deviation of 0.1 on all stations except for LMT, for which we use a log standard deviation of 0.2 to accommodate its larger variations (M87” Paper III; Paper II).

6.4.2. eht-imaging

We also utilize the geometric model-fitting tools developed within the eht-imaging Python library (Chael et al. 2016, 2018). This library enables visibility-domain fitting to arbitrary combinations of simple analytic models, including the mG-ring model, and it can do so using a variety of interferometric data products, including visibility amplitudes, closure phases, and log closure amplitudes. eht-imaging is also able to interface with a variety of external packages to perform parameter optimization or posterior exploration.

For the snapshot modeling analyses performed in this paper, we match the operation of eht-imaging with that of Comrade. Unless otherwise specified, we use eht-imaging to fit to visibility amplitudes and closure phases, so that the snapshot likelihood is given by Equation (32), and we use dynesty (Speagle 2020) for posterior exploration and evidence estimation. We also specify the same priors for the station gain parameters as used in the Comrade fits.

Given that both eht-imaging and Comrade use identical model specifications, priors, and samplers, we expect all results produced by these softwares to be identical up to sampling precision. We thus use only Comrade fits for all Sgr A$^*$ snapshot geometric analyses in this paper.

6.4.3. DPI

The third software we use for snapshot geometric modeling is the Python code Deep Probabilistic Imaging/Inference (DPI/α-DPI; Sun & Bouman 2021; Sun et al. 2022). DPI approximates the posterior over all model parameters by fitting a normalizing flow neural network (Rezende & Mohamed 2015) to the data using a Rényi $\alpha$-divergence variational inference technique (Li & Turner 2016). DPI is an optimization-based posterior estimation framework, and it uses the auto-differentiation package PyTorch (Paszke et al. 2017) to optimize the neural network weights. The posterior

57. https://github.com/cscherrer/Soss.jl
158. https://github.com/pptiede/ComradeSoss.jl
estimation accuracy is further improved post-optimization through importance reweighting of the samples generated by the normalizing flow neural network.

DPI supports fitting to multiple data products, including visibility amplitudes and closure quantities, but it does not currently support the inclusion of station gain amplitudes as model parameters. We thus use DPI to fit to closure phases and log closure amplitudes; the snapshot likelihood is given by Equation (33). Prior to fitting, all time stamps that are unable to form at least one closure phase and at least one closure amplitude are flagged.

DPI differs from both Comrade and eht-imaging in that it defines geometric models in the image domain rather than in the visibility domain, and it uses a nonuniform fast Fourier transform (NFFT) to compute the necessary data products. For the analyses carried out in this paper, we discretize the model as an image containing $32 \times 32$ pixels spanning a 160 $\mu$as field of view.

Because the pixel size is finite, DPI cannot support a model containing infinitesimally thin rings such as that in Equation (10); furthermore, convolutions in the image domain are computationally expensive. The DPI fits in this paper thus employ a modified version of the mG-ring model specification,

$$I'_\text{ring}(r, \phi) = F_{\text{ring}} \exp \left[ \frac{4 \ln 2}{W^2} \left( r - d^2/2 \right)^2 \sum_{n=1}^{m} \beta_n e^{i n \phi} \right],$$

$$I'(r, \phi) = I'_{\text{ring}}(r, \phi) + I_{\text{Gauss}}(r, \phi),$$

$$F_0 = F_{\text{ring}} + F_{\text{Gauss}} = 1,$$

where we note that $d'$ and $W'$ are conceptually distinct from $d$ and $W$. The quantities $d$ and $W$ in the mG-ring model from Section 4.3 determine the diameter of the infinitesimally thin ring and the FWHM of its convolving kernel, respectively. In contrast, the quantities $d'$ and $W'$ determine the intensity peak and FWHM, respectively, of a radial Gaussian function. These two specifications congregate only in the limit of large $d$ and small $W$. In addition, the total flux of the DPI model implementation is fixed to be 1 Jy because DPI fits only to closure quantities and closure amplitudes are not sensitive to the absolute flux scale.

6.4.4. Sampling the Hypermodel Posterior

To sample from the hypermodel posterior $P(\theta|D)$ (Equation (37)), we use the adaptive Metropolis sampler from Vihola (2010) via its implementation in the Julia package RobustAdaptiveMetropolisSampler.jl.159 The sampler is initialized by first running an adaptive genetic algorithm from the Julia package BlackBoxOptim.jl,160 which provides a starting point near the maximum posterior density. We run the sampler for a minimum of 2 million Markov Chain Monte Carlo (MCMC) steps or until we have effective sample sizes of 500 for all parameters.

6.5. Model Selection

The mG-ring model described in Section 4.3 is not actually a single model but rather a class of models, delineated by the

order $m$ (Equation (10)). To determine the $m$-order that is preferred by the Sgr A* data, we carry out a series of snapshot mG-ring fits to the Sgr A* using different values of $m$ and compare Bayesian evidence estimates. Given a set of log-evidences $\ln Z_i$ computed for every snapshot $i$ in a single observation, the total evidence for the entire observation is simply given by their sum,

$$\ln Z = \sum_i \ln Z_i.$$ (43)

The Comrade snapshot fitting analyses directly estimate the Bayesian evidence on every snapshot, and so the total evidence across an entire observation can be computed directly using Equation (43). The results of a Comrade $m$-order survey covering $m = \{1, 2, 3, 4, 5\}$ are shown in Figure 14. We find that the $m = 4$ order is preferred in both bands and across both calibration pipelines. We thus use the $m = 4$ mG-ring as our fiducial model for all Comrade Sgr A* analyses in this paper.

Unlike Comrade, DPI does not directly estimate the Bayesian evidence during each fit. Instead, we use the evidence lower bound (ELBO) to determine the $m$-order preference. The ELBO is a combination of the true evidence modified by a relative entropy term that encodes the performance of the variational approximation,

$$\text{ELBO}(m) = \log p(D_i|m) - D_{KL}[q(\theta_*)||p(\theta_i|D_i, m)],$$ (44)

where $D_{KL}[A||B]$ is the Kullback–Leibler divergence of $A$ from $B$, and $q(\theta_*)$ is the optimized DPI normalizing flow distribution. The relative entropy term is zero when the DPI distribution $q(\theta_*)$ and the true posterior $p(\theta_i|D_i)$ are identical, so the ELBO provides a rough estimate of the log-evidence.

Figure 14. Relative Bayesian evidence and best-fit mG-ring model diameter vs. $m$-order, from the snapshot-averaging geometric modeling from the Comrade pipeline applied to the Sgr A* data. Because the absolute values of $\ln Z$ can be substantially different for each data set (i.e., each combination of observing day, frequency band, and calibration pipeline), and because only the relative values carry information about model specification preferences, we reference all $\ln Z$ values to the maximum value attained at any $m$ for that data set.
The results of a DPI m-order survey covering \( m = \{1, 2, 3, 4\} \) indicate that either \( m = 1 \) or \( m = 2 \) is preferred, depending on the day and band. We choose to err on the side of increased model flexibility and use the \( m = 2 \) mG-ring as our fiducial model for all DPI Sgr A* analyses in this paper.

Figure 15 shows representative mG-ring fits to the Sgr A* HOPS low-band data for both the Comrade and DPI pipelines. In all cases, we find that the normalized residuals are distributed around a value of zero with a subunity variance, and there is no evidence of systematic structure. The \( \chi^2 \) statistics for each of these fits are provided in Appendix E.

### 7. Full-track Geometric Modeling

The snapshot modeling analysis presented in the previous section addresses the variability of the Sgr A* data by explicitly permitting the source structure to vary in time. As described in Section 3.4, an alternative approach to fitting variable data is to statistically capture the impact of variability, treating it as an additional source of uncertainty modifying data that otherwise describe a static (or average) source structure. We pursue such an approach here in the form of “full-track” geometric modeling, whereby we fit the mG-ring model (see Section 4.3) to an entire data set at once and account for the variability by simultaneously fitting a parameterized noise.
model. In this section we detail our formalism for fitting the mG-ring geometric model alongside a model that captures the noise budget inflation associated with source variability.

7.1. Data Preparation

The data preparation for the full-track geometric modeling analyses is similar to that used for snapshot geometric modeling analyses (see Section 6.1). The data are first processed through the pre-imaging pipeline described in Paper III, which applies light-curve normalization and performs some a priori gain calibration. However, unlike in Paper III and Section 6.1, we do not modify the data uncertainties at all beyond their thermal noise values; neither a systematic error term nor a refractive scattering term is added to the error budget. Additionally, no “deblurring” is applied to the data; instead, the blurring is applied directly to the model as described in the next section.

Following the application of the pre-imaging pipeline, we coherently average the visibilities from each baseline on a per-scan basis. A scan length (~10 minutes) is approximately the amount of time over which we expect structural variability to be subdominant to other sources of uncertainty (see Section 3.1, in particular Figure 1). Furthermore, the station gains are expected to be constant in time across a single scan but not from one scan to the next (M87* Paper III; Paper II), meaning that a scan length is also the longest coherent integration time that the a priori calibration can support.

While the full-track modeling is necessarily focused on reconstructing a time-averaged image structure, the underlying data remain a collection of complex visibilities that sample different instantaneous realizations of the intrinsic Sgr A* source structure. As a consequence, the Sgr A* data exhibit subhour correlations that over a single day are localized in the (u,v)-plane. As previously noted in Section 3.4 and detailed in Appendix B, these unmodeled correlations can result in significant biases in the reconstructed properties of Sgr A*.

However, by fitting to multiple days of Sgr A* data, and thus combining multiple samplings of the variable source structure at each location in the (u,v)-plane, we better match the statistical properties of the data to those assumed by the full-track analysis. An additional benefit of combining days is that the multiday analyses more clearly emphasize the static signatures of gravitational lensing from the spurious astrophysical variability. For these reasons, all full-track analyses presented in Section 8 make use of the combined April 6 and April 7 Sgr A* data. For comparison we provide single-day analysis results in Appendix C.

7.2. Model Specification and Implementation

The goal of the full-track geometric modeling procedure is to determine the posterior distribution for the parameters of the static mG-ring model and parameterized noise model that best describe an entire Sgr A* data set. Our specification for the mG-ring model is described in Section 4.3, and we retain the same notation and terminology in this section. Additionally, we incorporate the blurring effects of scattering in the same manner described in Section 4 of Paper III, through multiplication of the mG-ring visibilities by the Fourier transform of the scattering kernel. We note, however, that all images shown in the figures in this paper correspond to the underlying (i.e., nonscattered) image.

7.2.1. Parameterized Noise Model

Our parameterized noise model for a complex visibility $V_i$ measured on a baseline $u_i$ is given by

$$\sigma_i^2 = \sigma_{th,i}^2 + f^2 |V_i|^2 + \sigma_{ref}^2 + \sigma_{var}^2 (|u_i|).$$

(45)

Here the first term is the thermal noise in the measurement (see Equation (2)), the second term is a component that is multiplicative in the visibility amplitude $|V_i|$ and is intended to capture residual (nongain) calibration errors (e.g., residual polarization leakage), the third term is a component that is additive and is intended to account for refractive scattering noise, and the fourth term is a component that is a function of the baseline length $|u_i|$ and is intended to capture the effects of source variability. With the exception of the variability term, Equation (45) is similar to the noise budget used in the snapshot modeling (see Equation (30)); the only difference is that now $f$ and $\sigma_{ref}$ enter into the model as free parameters.

The variability noise $\sigma_{var}$ is described in Section 3 (see Equation (8)) and consists of a broken power law in $|u|$ specified by four parameters: an overall amplitude $a_4$ specified at a baseline length of 4 GA, a falling long-baseline power-law index $b$, a rising short-baseline power-law index $c$, and a baseline length $u_0$ at which the power-law breaks. Informative prior bounds for each of these parameters are determined from the model-agnostic variability quantification described in Section 3.3, and these bounds are listed in Table 2.

7.2.2. THEMIS Implementation

We have implemented the combined mG-ring plus noise parameterization as a model within the sampling-based parameter estimation framework THEMIS developed for the EHT (Broderick et al. 2020a, 2020b). Given a model specification and a data set, THEMIS works within a Bayesian

<table>
<thead>
<tr>
<th>Table 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full-track Modeling mG-ring Priors</strong></td>
</tr>
<tr>
<td>Parameter</td>
</tr>
<tr>
<td>$f_{\text{ring}}$</td>
</tr>
<tr>
<td>$d$ (µas)</td>
</tr>
<tr>
<td>$W$ (µas)</td>
</tr>
<tr>
<td>$</td>
</tr>
<tr>
<td>$\arg(\beta_0)$ (deg)</td>
</tr>
<tr>
<td>$f_{\text{class}}$</td>
</tr>
<tr>
<td>$W_{\text{class}}$ (µas)</td>
</tr>
<tr>
<td>$\sigma_{\text{ref}}$</td>
</tr>
<tr>
<td>$f$</td>
</tr>
<tr>
<td>$a_4$</td>
</tr>
<tr>
<td>$b$</td>
</tr>
<tr>
<td>$c$</td>
</tr>
<tr>
<td>$u_0$ (Gλ)</td>
</tr>
</tbody>
</table>

**Note.** Prior distributions for THEMIS full-track geometric modeling analyses. The top section lists priors for the mG-ring model parameters, and the bottom section lists priors for the parameterized noise model, $U(a, b)$ denotes a uniform prior on the interval $[a, b]$, $U_\ref(a, b)$ denotes a log-uniform prior on the interval $[a, b]$, and $N_{\text{G}}(\mu, \sigma^2)$ denotes a lognormal prior with mean $\mu$ and variance $\sigma^2$. Priors for the variability noise parameters $a_4$, $b$, $c$, and $u_0$ are informed by the model-agnostic variability quantification analysis described in Section 3.3.

<table>
<thead>
<tr>
<th>Section 3.3.</th>
<th>Parameter Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>THEMIS full-track geometric modeling analyses. The top section lists priors for the mG-ring model parameters, and the bottom section lists priors for the parameterized noise model, $U(a, b)$ denotes a uniform prior on the interval $[a, b]$, $U_\ref(a, b)$ denotes a log-uniform prior on the interval $[a, b]$, and $N_{\text{G}}(\mu, \sigma^2)$ denotes a lognormal prior with mean $\mu$ and variance $\sigma^2$. Priors for the variability noise parameters $a_4$, $b$, $c$, and $u_0$ are informed by the model-agnostic variability quantification analysis described in Section 3.3.</td>
<td></td>
</tr>
</tbody>
</table>
formalism to produce a set of samples from the posterior distribution of the model parameters. THEMIS uses a MCMC sampling scheme to explore the posterior space, employing a parallel tempering scheme (Syed et al. 2022) to ensure traversal over the entire prior volume and the Hamiltonian Monte Carlo sampling kernel from the Stan package (Carpenter et al. 2017) to efficiently sample within each tempering level. A detailed description of the THEMIS sampling framework can be found in Tiede (2021).

The full-track geometric modeling analyses carried out in this paper fit to complex visibility data. Given a vector of geometric model parameters \( \mathbf{p} \) and a vector of noise model parameters \( \mathbf{n} \), the THEMIS likelihood function for complex visibilities is Gaussian,

\[
\ln \mathcal{L} = -\frac{1}{2} \sum_i \left( \frac{V_i - \hat{V}(u_i; \mathbf{p})}{\sigma_i^2} + \ln{2\pi\sigma_i^2} \right).
\]

Here \( V_i \) is a measured visibility, \( u_i \) is its baseline vector, \( \hat{V} \) is the corresponding modeled visibility, and the sum is taken over all data points \( i \). The noise \( \sigma_i \) in each visibility is specified as in Equation (45), with the noise model parameters \( \mathbf{n} = [f, \sigma_{\text{ref}}, \alpha, \beta, \gamma, \delta] \). THEMIS internally solves for and marginalizes over the full set of complex gain parameters (i.e., one complex gain per station per time stamp) at every sampling step using a Laplace approximation (see Broderick et al. 2020a). It also applies the Johnson et al. (2018) diffractive scattering kernel directly to the model prior to computing visibilities. A validation test of the THEMIS mG-ring plus noise model implementation is described along with other tests in a dedicated paper on the noise modeling approach (Broderick et al. 2022).

We assess MCMC convergence through both visual inspection of the traces and a number of quantitative chain statistics, including the integrated autocorrelation time, split-R, and parameter rank distributions (Vehtari et al. 2021). The number of tempering levels is selected to ensure efficient communication between the highest and lowest levels (per Syed et al. 2022), which typically requires about 20 levels. We run the sampler for between \( 5 \times 10^4 \) and \( 10^5 \) steps per tempering level.

To compute the Bayesian evidence, THEMIS uses thermodynamic integration (e.g., Lartillot & Philippe 2006), which computes the log-evidence through

\[
\ln \mathcal{Z} = \int_0^1 \mathrm{d} \beta \langle \ln \mathcal{L} \rangle_{\beta},
\]

where \( \langle \ln \mathcal{L} \rangle_{\beta} \) is the expectation of the log-likelihood taken over the distribution

\[
P^{(\beta)}(\mathbf{p}, \mathbf{n}) = \left( \frac{P(\mathbf{p}, \mathbf{n})}{\pi_{\text{ref}}(\mathbf{p}, \mathbf{n})} \right)^{\beta} \pi_{\text{ref}}(\mathbf{p}, \mathbf{n}).
\]

Note that THEMIS does not take \( \pi_{\text{ref}} \) to be the prior distribution. Instead, THEMIS uses a uniform distribution whose support matches the support of the priors given in Table 2. To compute Equation (47), we compute the average log-likelihood for each tempering level and then use trapezoidal integration to numerically compute the integral.

Prior for all mG-ring model and noise model parameters are listed in Table 2. We impose mean-zero lognormal priors for all station gain amplitudes, with a log standard deviation of 0.01 for all network calibrated stations (ALMA, APEX, JCMT, SMA), 0.2 for the LMT, and 0.1 for the remaining stations (PV, SMT, SPT); these gain priors are motivated by the expected performance of each station after the post-processing described in Section 7.1 (see also Paper II). All station gain phase priors are uniform on the unit circle.

### 7.3. Model Selection

As with the snapshot geometric analyses (see Section 6.5), the mG-ring model used for the full-track analyses is really a class of models that increases in complexity with \( m \). The results of a THEMIS \( m \)-order survey covering \( m = \{2, 3, 4, 5, 6\} \) are shown in Figure 16. In contrast to the snapshot geometric modeling (see Section 6.5), we find that the full-track analysis is able to support more complex model specifications, exhibiting a strong preference for \( m > 4 \) over \( m = 4 \). However, we find that increasing the \( m \)-order does not significantly impact the values of the primary morphological parameters of choice (see the similar diameter measurements and uncertainties for \( m \geq 4 \) in Figure 16). Thus, to maintain consistency among model specifications and to facilitate comparison with the snapshot geometric modeling analyses, we proceed with the \( m = 4 \) mG-ring as our fiducial model for all full-track Sgr A* analyses in this paper.

A representative \( m = 4 \) mG-ring fit to the Sgr A* HOPS low-band data is shown in Figure 17. We find that the normalized residuals are distributed around a value of zero with near-unity variance and that there is no evidence of systematic structure. The \( \chi^2 \) statistics for this fit are discussed in Appendix E.
8. Results

In this section we aggregate and present the results from the analyses described in Sections 3–7 as applied to the 2017 EHT Sgr A* data.

8.1. Structural Variability Measurements

The model-agnostic variability quantification analysis carried out in Section 3 demonstrates that the Sgr A* data exhibit variability—quantified here in terms of a normalized visibility amplitude variance—that is significantly in excess of that expected from thermal noise, station gains, and refractive scattering. As illustrated in Figure 4, the measured variability can be broken down into three regions with qualitatively distinct behavior:

1. On short baselines with lengths \(|u| \lesssim 2.5 \, \mu \text{as} \), corresponding to spatial scales \(\gtrsim 100 \, \mu \text{as} \), limitations in our calibration and the subsequent choices made in our preprocessing procedure preclude meaningful constraints on the variability. The light-curve normalization procedure removes all variability on intrasite baselines and suppresses variability on short intersite baselines that are highly correlated with the light curve; the variability of the light curve itself is thoroughly characterized in Wielgus (2022). The source size constraint used to perform gain calibration of the LMT-SMT baseline (see Paper II; Paper III) imposes a further, more artificial suppression of the variability on this baseline. We thus do not obtain any variability measurements for baselines shorter than 2.5 G\(\lambda\).

2. On intermediate baselines with lengths between 2.5 G\(\lambda \lesssim |u| \lesssim 6 \, \mu \text{as} \), corresponding to spatial scales between \(\sim 30\) and \(\sim 100 \, \mu \text{as} \), we measure significant variability that exhibits an approximately power-law decline with increasing baseline length. The power-law index is between \(~2\) and \(3\), and the magnitude of the variability ranges from a peak rms of \(~5\)% of the total flux density \((\sim 120 \, \text{mJy})\) near 2.5 G\(\lambda\) down to \(~1\)% of the total flux density \((\sim 25 \, \text{mJy})\) near 6 G\(\lambda\).

3. On long baselines with lengths \(|u| \gtrsim 6 \, \mu \text{as} \), corresponding to spatial scales \(\lesssim 30 \, \mu \text{as} \), the measured variability is comparable in magnitude to that expected from statistical errors and refractive scattering. These measurements thus do not contain statistically significant detections of structural variability.

These measurements describe the level of excess variance that the data exhibit about an underlying average source model. The parameters describing a broken power-law noise model fit to these measurements are thus used as a variability noise budget during image reconstruction (Paper III) and to define priors on the corresponding parameters in the full-track modeling analyses (Section 7).

Determining the intrinsic (i.e., infinite-time) source variability from these measurements requires an additional debiasing step to remove the impact of correlations between data points that are closely spaced in time. The analysis carried out in Section 3 involves binning the visibility data in the \((u,v)\)-plane for the purpose of computing variances. However, many data points within a single bin are from measurements taken close in time, which can introduce correlations that bias the computed variance. A procedure for removing this bias is detailed in Broderick et al. (2022), whereby the factor relating the measured variability to the intrinsic variability at every baseline length is calibrated using synthetic measurements of GRMHD simulations. In practice, Broderick et al. (2022) derive the debiasing function using the same 90 GRMHD simulations we use in this paper for \(\theta_g\) calibration (see Section 4.4 and Appendix D), and the resulting debiasing factor is close to unity everywhere except between \(~6\) and 7.5 G\(\lambda\) (see Figure 5 in Broderick et al. 2022). Applying this debiasing function to the variability measurements from the Sgr A* data yields the results shown in Figure 18 and reported in Table 3.

Due to the near-unit debiasing factor, the variability measurements shown in Figure 18 are similar to those from Figure 4. Quantitative constraints on the parameters of the noise model that are well constrained are presented in Figure 19 and Table 3. Where strong constraints on the excess noise exist (i.e., on baseline lengths between \(~2\) and 6 G\(\lambda\)), it continues to
be well described by a single power law with index $b = 2.4 \pm 0.8 \pm 0.1$ and amplitude (evaluated at a baseline length of $4 \, G\lambda$) of $a_4 = 1.90 \pm 0.2 \pm 0.3\%$, where in each value the $1\sigma$ and $2\sigma$ ranges are indicated. The measured variability magnitude is between $\sim 2$ and 10 times higher than that expected from refractive scattering alone. The lack of an observable break places an upper limit on its location of $u_0 < 1.3 \, G\lambda$ at $1\sigma$ and $u_0 < 3.1 \, G\lambda$ at $2\sigma$.

The excess variability is broadly consistent with that due to structural fluctuations anticipated by the GRMHD simulations discussed in Paper V and Georgiev et al. (2022). The magnitude of the excess variability lies within the range of that predicted by GRMHD simulations for $|u| > 2 \, G\lambda$, though it does appear to marginally favor less variable models. The long-baseline power-law index is consistent with all GRMHD simulations. A detailed discussion of the implications for GRMHD models is contained within Paper V.

### 8.2. Image Morphology Measurements

Both the snapshot (Section 6) and full-track (Section 7) geometric modeling analyses produce reconstructions of the Sgr A* emission structure, and the posterior distributions determined for the parameters describing these geometric model reconstructions provide a quantification of the morphological properties directly from the EHT interferometric data. Similarly, the IDFE analyses carried out in Section 5 quantify the morphological properties of the top-set and posterior images reconstructed in Paper III. Figure 20 compares the geometric models and image reconstructions determined for Sgr A*; for the geometric modeling analyses we show posterior mean images (i.e., the mean of many images sampled from the posterior distribution), while for the image reconstructions we show averages over the top sets (for eht-imaging, SMILI, and DIFMAP) or posterior means (for THEMIS). We see that both the snapshot and full-track geometric modeling analyses each recover a grossly similar overall structure across frequency bands and that this structure is also similar between the snapshot and full-track analyses. The image reconstructions permit much more flexibility in the permitted image structure, and so we see correspondingly more variation both within the imaging and geometric modeling. The primary point of consistency between the ring-like structures recovered from imaging and the rings fit via geometric modeling seems to be their sizes.

We use the geometric modeling and IDFE analyses to quantify a number of morphological parameters of interest, which are shown in Figure 21 and listed in Table 4.

#### 8.2.1. Ring Size

The parameter of most interest for gravitational studies (e.g., M87' Paper VI; Paper VI) is the ring size, which we quantify using its diameter. For the mG-ring modeling results from both snapshot and full-track analyses, we report a debiased diameter $\hat{d}$, given by

$$\hat{d} = d - \frac{1}{4 \ln(2)} \frac{W^2}{d},$$

(49)

where $d$ and $W$ are the mG-ring ring diameter and thickness, respectively (see Section 4.3). This debiasing corrects for the lowest-order impact of the Gaussian blurring kernel on the radial location of the peak intensity—which is shifted inward with respect to the radius of the pre-convolved ring (see Appendix G of M87' Paper IV)—and thus aids a more direct comparison of the geometric modeling diameter values with those obtained from IDFE.
The diameter measurements from the geometric modeling and IDFE analyses are compared in the top row of Figure 21 across frequency bands and calibration pipelines. We find that the diameter is the most well constrained of the geometric parameters we attempt to quantify, with both a typical measurement uncertainty and a scatter between measurement types that is substantially smaller than the magnitude of the value itself. An average of the geometric modeling results across both frequency bands and calibration pipelines yields a debiased diameter of $51.9 \pm 2.0 \mu\text{as}$, with a corresponding symmetrized uncertainty of $2.0 \mu\text{as}$. The quoted error is the 68% (i.e., approximately $1\sigma$) probability and corresponds to the samples from each measurement weighted equally. The IDFE measurements are broadly consistent with the results from geometric modeling, yielding $\approx 51.8 \pm 2.6 \mu\text{as}$.

We note that even after debiasing the diameter measurements from different analysis pathways remain interpretationally distinct quantities. To ensure mutual consistency between different measurement methods, in Section 8.3 we calibrate the diameter measurements to a common physical scale using the GRMHD synthetic data sets generated for this purpose (see Section 4.4).

### 8.2.2. Ring Thickness

The thickness of the ring is of interest for its ability to constrain the location and size of the emitting region near the black hole (e.g., Lockhart & Gralla 2022). The ring thickness measurements from the geometric modeling and IDFE analyses are compared in the second row of Figure 21 across frequency bands and calibration pipelines. We find that the geometric modeling methods recover similar ring thicknesses, with the snapshot analyses obtaining $\approx 16$–$22 \mu\text{as}$ and the full-track analyses obtaining $\approx 19$–$23 \mu\text{as}$. The IDFE analyses obtain consistently thicker rings, with $\approx 30 \mu\text{as}$ being a more typical value and a somewhat larger scatter (from $\approx 25$ to $35 \mu\text{as}$) seen both across and within pipelines. The increased ring thickness recovered from the IDFE analyses likely arises in part from image smoothing introduced by some reconstruction algorithms (e.g., the CLEAN algorithm used by DIFMAP).

The thickness of the M87$^*$ ring was a parameter that the geometric modeling analyses carried out in M87$^*$ Paper VI had difficulty constraining, and it showed substantial variation across days and between modeling methods; only a relatively weak upper limit for the fractional thickness of $\mathcal{W} = 0.5$ was obtained. In contrast, we find for Sgr A$^*$ that the thickness parameter is relatively well constrained by geometric modeling approaches. The fractional thickness is $\mathcal{W} = 0.35 \pm 0.05$, where the uncertainties quoted are symmetrized $1\sigma$. IDFE analyses obtain systematically larger fractional thicknesses, finding $\mathcal{W} = 0.53 \pm 0.1$.

Unlike with the diameter measurements, we do not debias the ring thicknesses obtained from different analysis pathways or attempt to calibrate them to a common scale. The ring thicknesses from geometric modeling and IDFE thus represent two interpretationally distinct quantities, and we do not
produce an analysis-agnostic measurement of the ring thickness.

8.2.3. Position Angle and Asymmetry

The magnitude and orientation of any asymmetry in the azimuthal brightness distribution around the ring is of interest because it can be related to the spin and inclination of the black hole (e.g., M87* Paper V; Paper V). As described in Section 4.3, the mG-ring position angle ($\eta$) and degree of azimuthal asymmetry ($A$) are both determined by the coefficient of the $m = 1$ mode as specified in Equations (23) and (24), respectively. These definitions match closely the corresponding IDFE quantities defined in Equation (5.2). The asymmetry and position angle measurements we obtain for Sgr A* are shown in the third and fourth rows of Figure 21, respectively.

Unlike for M87*, where the image morphology exhibits a clearly defined asymmetry axis whose magnitude and orientation can be consistently quantified using either geometric modeling or IDFE analyses (M87* Paper IV; M87* Paper VI), the image structure for Sgr A* is less amenable to such a description. The asymmetry magnitude measurements show a large scatter across methods, spanning $\sim0.15–0.3$ for the geometric modeling methods and $\sim0.04–0.20$ for the IDFE methods. The IDFE methods recover systematically smaller median levels of asymmetry than the geometric modeling methods, but the uncertainties are large; several of the measurement methods have statistical uncertainties that cover nearly the entire 0–0.5 range of the prior distribution for $A$.

The position angle measurements show similarly little consistency between different analysis methods, spanning essentially the full ($–180^\circ$, $180^\circ$) range when compared across all data sets and measurement techniques. The geometric modeling analyses find position angles that are loosely confined to a region between $–100^\circ$ and $0^\circ$, while the IDFE analyses show a large ($>100^\circ$) scatter between methods and a similar magnitude of uncertainty for individual measurements.

8.2.4. Brightness Depression

The depth of the brightness depression interior to the ring is a key signature of the presence of a black hole. Additionally, it
Table 4
Parameters Describing Sgr A* Image Morphology

<table>
<thead>
<tr>
<th>Analysis class</th>
<th>Software</th>
<th>Day</th>
<th>Band</th>
<th>$\tilde{d}$ (µas)</th>
<th>$W$ (µas)</th>
<th>$A$ (deg)</th>
<th>$f_c$</th>
<th>$\tilde{d}$ (µas)</th>
<th>$W$ (µas)</th>
<th>$A$ (deg)</th>
<th>$f_c$</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIFMAP+ REx</td>
<td>April 7</td>
<td>HI+LO</td>
<td>50.6±1.1</td>
<td>31.9±2.2</td>
<td>0.09±0.07</td>
<td>48.4±57.1</td>
<td>0.43±0.12</td>
<td>0.07±0.01</td>
<td>50.1±1.2</td>
<td>31.8±3.3</td>
<td>0.05±0.09</td>
</tr>
<tr>
<td>DIFMAP+ VIDA</td>
<td>April 7</td>
<td>HI+LO</td>
<td>51.2±2.2</td>
<td>33.4±3.1</td>
<td>0.10±0.11</td>
<td>49.8±61.1</td>
<td>0.40±0.09</td>
<td>0.07±0.03</td>
<td>51.3±1.4</td>
<td>33.1±2.1</td>
<td>0.07±0.18</td>
</tr>
<tr>
<td>eht-imaging+ REx</td>
<td>April 7</td>
<td>HI+LO</td>
<td>50.3±1.9</td>
<td>26.7±3.1</td>
<td>0.13±0.06</td>
<td>-33.9±56.3</td>
<td>0.33±0.12</td>
<td>0.03±0.16</td>
<td>53.6±1.8</td>
<td>27.2±4.6</td>
<td>0.11±0.05</td>
</tr>
<tr>
<td>eht-imaging+ VIDA</td>
<td>April 7</td>
<td>HI+LO</td>
<td>52.7±2.6</td>
<td>28.0±3.1</td>
<td>0.19±0.11</td>
<td>-40.5±65.2</td>
<td>0.30±0.22</td>
<td>0.11±0.23</td>
<td>55.3±3.6</td>
<td>27.4±5.2</td>
<td>0.11±0.05</td>
</tr>
<tr>
<td>SMILI+ REx</td>
<td>April 7</td>
<td>HI+LO</td>
<td>52.0±5.5</td>
<td>25.1±1.1</td>
<td>0.16±0.04</td>
<td>130.9±70.3</td>
<td>0.29±0.09</td>
<td>0.04±0.07</td>
<td>51.4±4.0</td>
<td>26.7±1.2</td>
<td>0.12±0.04</td>
</tr>
<tr>
<td>SMILI+ VIDA</td>
<td>April 7</td>
<td>HI+LO</td>
<td>49.3±1.7</td>
<td>28.5±1.7</td>
<td>0.15±0.04</td>
<td>124.4±77.0</td>
<td>0.26±0.07</td>
<td>0.00±0.11</td>
<td>50.6±1.1</td>
<td>27.2±0.8</td>
<td>0.07±0.01</td>
</tr>
<tr>
<td>THEMIS + REx</td>
<td>April 6-7</td>
<td>HI+LO</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>THEMIS + VIDA</td>
<td>April 6-7</td>
<td>HI+LO</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Comrade</td>
<td>April 6-7</td>
<td>HI</td>
<td>53.9±0.6</td>
<td>16.4±0.3</td>
<td>0.23±0.02</td>
<td>3.8±5.3</td>
<td>0.23±0.01</td>
<td>0.01±0.01</td>
<td>52.3±0.7</td>
<td>17.7±0.2</td>
<td>0.23±0.03</td>
</tr>
<tr>
<td>DPI</td>
<td>April 6-7</td>
<td>LO</td>
<td>51.4±0.5</td>
<td>17.0±0.2</td>
<td>0.22±0.03</td>
<td>-11.4±9.0</td>
<td>0.20±0.01</td>
<td>0.03±0.03</td>
<td>53.3±0.8</td>
<td>17.7±0.2</td>
<td>0.22±0.02</td>
</tr>
<tr>
<td>Full-track</td>
<td>THEMIS</td>
<td>April 6-7</td>
<td>HI</td>
<td>52.1±0.4</td>
<td>19.7±0.4</td>
<td>0.16±0.02</td>
<td>-4.0±11.8</td>
<td>0.09±0.02</td>
<td>0.02±0.03</td>
<td>52.6±0.5</td>
<td>21.5±0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>April 6-7</td>
<td>LO</td>
<td>52.7±0.5</td>
<td>20.3±0.4</td>
<td>0.15±0.02</td>
<td>-74.0±10.5</td>
<td>0.16±0.03</td>
<td>0.03±0.03</td>
<td>51.4±0.5</td>
<td>22.5±0.5</td>
</tr>
</tbody>
</table>

Note. Median values and 68% credible intervals for the morphological quantities of interest, measured from the EHT Sgr A* data. Because medians and quantiles are not well defined for angular variables, for the position angle $\eta$ we instead quote the circular mean and standard deviation.
can be used to constrain the presence of an emitting or reflecting surface, as a potential alternative to a horizon (e.g., Broderick et al. 2015; Paper VI). For the mG-ring model, the fractional central flux \( f_c \) is given by Equation (21). For the IDFE analyses, we retain the definitions from Equation (5.2) for \( f_c \).

The \( f_c \) measurements for Sgr A* are shown in the bottom row of Figure 21. We find a large spread in values across analysis methods, ranging from \(-0.1\) to \(0.25\) for the geometric modeling analyses and from \(-0.0\) to \(0.5\) for the IDFE analyses. Compared against the constraints from geometric modeling of the M87* ring structure, which consistently found \( f_c \lesssim 0.1 \) (M87* Paper VI), the results obtained here for Sgr A* allow for the possibility of substantially more emission interior to the ring.

8.3. Gravitational Radius and Mass

The ring size measurements presented in Section 8.2.1 have been made using a variety of different analysis techniques, with different inherent assumptions and biases. To bring these otherwise disparate measurement techniques to a common scale, we follow a strategy similar to that developed in M87* (Paper VI and calibrate the diameter measurements using simulations from the Paper V GRMHD library. As described in Section 4.4, our calibration suite consists of synthetic data sets constructed from 90 GRMHD simulations spanning a range of accretion flow and black hole parameters, and for which an absolute reference size scale (i.e., the angular size of the gravitational radius \( \theta_g \)) is known. We apply the same data processing and ring diameter measurement strategies as used for the Sgr A* data to each of these synthetic data sets, and we use the resulting distribution of diameter measurements to derive the value and uncertainty in the scaling factor (\( \alpha \)) between \( \theta_g \) and \( d \) for every method (Equation (26)). We note that a conceptually similar calibration is carried out in the companion Paper VI, in which the GRMHD assumption is relaxed and a more diverse set of spacetimes and accretion flow models is used to calibrate ring size measurements.

8.3.1. Calibrated Scaling Factors

When applied to the calibration suite data sets, each diameter measurement technique produces a discrete distribution of \( \alpha \) scaling factors. We use a kernel density estimator (KDE) from the scikit-learn package (Pedregosa et al. 2011) to produce a nonparametric estimate of the continuous distribution corresponding to these discrete samples, and we use this KDE to construct the \( \theta_g \) distribution for Sgr A* described below (Section 8.3.2).

Table 5 lists the derived \( \alpha \) value and its uncertainty—as computed from the KDE distribution—for each of the ring size measurement methods used in this paper. The uncertainty in \( \alpha \) contains two main components: a statistical uncertainty \( \sigma_{\alpha}^{\text{stat}} \) associated with the fidelity of the ring measurement from each data set, and a theoretical uncertainty \( \sigma_{\alpha}^{\text{theory}} \) associated with the intrinsic scatter in \( \alpha \) as measured across different GRMHD calibration data sets. The total uncertainty \( \sigma_{\alpha}^{\text{tot}} \) is a combination of both the statistical and theoretical uncertainties, meaning that in practice we do not have access to \( \sigma_{\alpha}^{\text{theory}} \) in isolation, and thus we do not report it in Table 5. Nevertheless, the relative values of \( \sigma_{\alpha}^{\text{stat}} \) and \( \sigma_{\alpha}^{\text{theory}} \) indicate that the theoretical uncertainty is typically the dominant component.

The calibrated \( \alpha \) values span a range of \(-10\)–\(12\), depending on the specific measurement technique. Statistical uncertainties in the \( \alpha \) values determined using the geometric modeling techniques are a few percent, while for the IDFE techniques the statistical uncertainty is typically larger and reaches as high as \(-20\%\) in the worst cases. Folding in both the statistical and the theoretical components, the total uncertainties are more comparable across methods, though they still exhibit a large range spanning \(-15\%\)–\(-35\%). Overall, the calibrated \( \alpha \) values show similar magnitudes to what M87* Paper VI derived from fits to M87*, but the calibration uncertainty in the case of Sgr A* is substantially larger. This increased uncertainty reflects the increased flexibility that has been built into the ring size measurement techniques to capture structural variability in the source, as well as the increased morphological diversity of the GRMHD calibration suite that is necessary to accommodate the a priori unknown inclination of Sgr A*.

8.3.2. Sgr A* Angular Gravitational Radius

We apply the calibrated \( \alpha \) values to the measured Sgr A* ring diameters for each measurement technique, which produces a distribution of \( \theta_g \) values that captures the uncertainties in both the ring measurements and the GRMHD
The resulting $\theta_\phi$ distributions are shown in Figure 22 and listed in Table 6. An average across all methods and datasets yields a joint constraint of $\theta_\phi = 4.8^{+0.6}_{-0.7}$ mas, where the uncertainties are quoted at the 68% (i.e., 1σ) probability level and the systematic uncertainty is taken to be the standard deviation across all measurement methods. This value is consistent with the considerably more precise constraints obtained from measurements of stellar orbits (Do et al. 2019; Gravity Collaboration et al. 2019, 2020), and the gravitational implications of this consistency are explored in Paper VI.

8.3.3. Sgr A* Mass

The angular size of the gravitational radius is proportional to the ratio of the mass and distance to Sgr A*, per Equation (25). Our constraints on $\theta_\phi$ can thus be mapped directly to constraints on the black hole mass $M$ by incorporating an independent distance measurement to Sgr A*. Reid et al. (2019) report a distance of $D = 8.15 \pm 0.15$ (at 1σ probability) to the Galactic center, measured using trigonometric VLBI parallaxes of a large number (∼200) of masers. Using this distance measurement along with our measurement of $\theta_\phi$ from Section 8.3.2 yields a constraint on the mass of Sgr A* of $M = 4.0^{+1.1}_{-0.6} \times 10^6 M_\odot$, where we again quote 1σ uncertainties and the systematic component is taken to be the weighted standard deviation across methods. This measurement is once again consistent with the more precise constraints obtained from stellar orbits (Do et al. 2019; Gravity Collaboration et al. 2019, 2020), and the uncertainty in the mass remains dominated by our uncertainty in $\theta_\phi$.

9. Summary and Conclusions

In this paper we quantify the temporal variability and morphological properties of the horizon-scale emission from Sgr A*, using data taken by the EHT in 2017 April. Our primary morphological quantity of interest is the diameter of the observed ring of emission, which we quantify using multiple independent analysis pathways. We then use the ring diameter to place constraints on the angular size of the gravitational radius ($\theta_\phi$) and on the mass ($M$) of Sgr A*. The analyses presented here have been carried out using data taken on April 6 and April 7, across two frequency bands and two data calibration pipelines.

Motivated by theoretical expectations that the dynamical timescales in Sgr A* should be much shorter than the duration of EHT observing tracks, we employ a new method developed in Broderick et al. (2022) for quantifying the time variability observed in the visibilities in a manner that is agnostic to the specifics of the average underlying source structure. We find that the visibility amplitudes exhibit a light-curve-normalized variance that is in excess of that expected from thermal noise, station gains, or refractive scattering effects, and we attribute this excess variance to intrinsic structural changes in the source. The detected variability is most statistically significant on baselines with lengths between 2.5 and 6 Gλ, where it exhibits an approximately power-law decline with increasing baseline length, with a power-law index of $\sim$-2–3. The magnitude of this unimodal distribution representing a diverse family of probability density functions. We use the GAD parameterization from Freimer et al. (1988), and we use the GLDEX package in R (Su 2007a, 2007b) to carry out the fitting.


EvErt HoriEen TeleEscoPc EoPrlalCollepafEo etal.
variability on baselines near 3 GA in length exceeds 0.1 Jy, which is roughly equal to the value of the correlated flux density on these same baselines.

Through an exploration of potential simple geometric source structures, we demonstrate that the EHT Sgr A* data statistically prefer ring-like morphologies over other morphologies with comparable complexity. We develop and deploy two new methods for fitting the time-variable Sgr A* data using static geometric ring models with azimuthally modulated brightness structures. In the first method, called "snapshot" geometric modeling, we first fit the models to short segments of data over which the source variability is subdominant to other sources of uncertainty. The fits from individual segments are then combined via a hierarchical model averaging scheme to determine a value of 51.8 ± 0.5 μas. The thickness of the ring is well measured by geometric modeling methods, both with one another and with the IDFE techniques jointly determining a value of 51.8 ± 0.5 μas. We thank an anonymous referee for insightful and constructive comments that helped improve the quality of this paper.

Table 6

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Software</th>
<th>Day</th>
<th>HOPS LO</th>
<th>HOPS HI</th>
<th>CASA LO</th>
<th>CASA HI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Snapshot</td>
<td>Comrade</td>
<td>April 6+7</td>
<td>4.6±0.4</td>
<td>4.5±0.7</td>
<td>4.4±0.4</td>
<td>4.6±0.7</td>
</tr>
<tr>
<td>Full-track</td>
<td>DPI</td>
<td>April 6+7</td>
<td>4.5±1.7</td>
<td>...</td>
<td>4.9±1.8</td>
<td>...</td>
</tr>
<tr>
<td>Imaging</td>
<td>DIFMAP+ REx</td>
<td>April 7</td>
<td>4.8±1.2</td>
<td>4.8±1.5</td>
<td>4.8±1.5</td>
<td>4.8±1.5</td>
</tr>
<tr>
<td></td>
<td>DIFMAP+ VIDA</td>
<td>April 7</td>
<td>4.8±1.3</td>
<td>4.8±1.5</td>
<td>4.8±1.5</td>
<td>4.8±1.5</td>
</tr>
<tr>
<td></td>
<td>eht-imaging+ REx</td>
<td>April 7</td>
<td>4.9±1.2</td>
<td>4.9±0.5</td>
<td>4.9±0.5</td>
<td>4.9±0.5</td>
</tr>
<tr>
<td></td>
<td>eht-imaging + VIDA</td>
<td>April 7</td>
<td>5.0±0.8</td>
<td>5.0±0.8</td>
<td>5.0±0.8</td>
<td>5.0±0.8</td>
</tr>
<tr>
<td></td>
<td>SMILI + REx</td>
<td>April 7</td>
<td>5.0±1.8</td>
<td>5.2±0.8</td>
<td>5.2±0.8</td>
<td>5.2±0.8</td>
</tr>
<tr>
<td></td>
<td>SMILI + VIDA</td>
<td>April 7</td>
<td>5.1±0.9</td>
<td>5.2±1.1</td>
<td>5.2±1.1</td>
<td>5.2±1.1</td>
</tr>
<tr>
<td></td>
<td>THEMIS + REx</td>
<td>April 6+7</td>
<td>5.1±1.1</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>THEMIS + VIDA</td>
<td>April 6+7</td>
<td>5.0±1.3</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Note. The angular gravitational radius, θ_g, measured across analysis schemes and data sets; all values are quoted as medians and 68% credible intervals in units of μas. Entries that straddle both low-band and high-band columns have been computed from fits to band-combined data sets.

Using a suite of synthetic data sets generated from the Paper V GRMHD simulation library, we calibrate the diameter measurements from both geometric modeling and IDFE methods to a common physical scale. The resulting constraint on the angular size of the Sgr A* gravitational radius, combined across all methods and data sets, is θ_g = 4.8±1.5 μas. This large uncertainty arises from both the model flexibility necessary to capture structural variability in the source and the broad morphological diversity of the GRMHD calibration suite that reflects the a priori unknown inclination of Sgr A*. Combining our θ_g constraint with an independent distance measurement from Reid et al. (2019), we determine the mass of Sgr A* to be $M = 4.0 \pm 0.5 \times 10^6 M_\odot$. Though the uncertainties are large compared to those derived using other techniques (e.g., stellar orbit modeling), our measurement represents the first time that the mass of Sgr A* has been constrained by observations of light bending near the horizon.
78-409; the South African Research Chairs Initiative, through the South African Radio Astronomy Observatory (SARAO, grant ID 77948), which is a facility of the National Research Foundation (NRF), an agency of the Department of Science and Innovation (DSI) of South Africa; the Toray Science Foundation; Swedish Research Council (VR); the US Department of Energy (USDOE) through the Los Alamos National Laboratory (operated by Triad National Security, LLC, for the National Nuclear Security Administration of the USDOE (Contract 89233218CNA000001); and the YCAA Prize Postdoctoral Fellowship.

We thank the staff at the participating observatories, correlation centers, and institutions for their enthusiastic support. This paper makes use of the following ALMA data: ADS/JAO.ALMA#2016.1.01154.V. ALMA is a partnership of the European Southern Observatory (ESO; Europe, representing its member states), NSF, and National Institutes of Natural Sciences of Japan, together with National Research Council (Canada), Ministry of Science and Technology (MOST; Taiwan), Academia Sinica Institute of Astronomy and Astrophysics (ASIAA; Taiwan), and Korea Astronomy and Space Science Institute (KASI; Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, Associated Universities, Inc. (AUI)/NRAO, and the National Astronomical Observatory of Japan (NAOJ). The NRAO is a facility of the NSF operated under cooperative agreement by AUI. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. We also thank the Center for Computational Astrophysics, National Astronomical Observatory of Japan. The computing cluster of Shanghai VLBI correlator supported by the Special Fund for Astronomy from the Ministry of Finance in China is acknowledged.

APEX is a collaboration between the Max-Planck-Institut für Radioastronomie (Germany), ESO, and the Onsala Space Observatory (Sweden). The SMA is a joint project between the SAO and ASIAA and is funded by the Smithsonian Institution and the Academia Sinica. The JCMT is operated by the East Asian Observatory on behalf of the NAOJ, ASIAA, and KASI, as well as the Ministry of Finance of China, Chinese Academy of Sciences, and the National Key Research and Development Program (No. 2017YFA0402700) of China and Natural Science Foundation of China grant 11873028. Additional funding support for the JCMT is provided by the Science and Technologies Facility Council (UK) and participating universities in the UK and Canada. The LMT is a project operated by the Instituto Nacional de Astrofísica, Óptica, y Electrónica (Mexico) and the University of Massachusetts at Amherst (USA). The IRAM 30-m telescope on Pico Veleta, Spain is operated by IRAM and supported by CNRS (Centre National de la Recherche Scientifique, France), MPG (Max-Planck-Gesellschaft, Germany) and IGN (Instituto Geográfico Nacional, Spain). The SMT is operated by the Arizona Radio Observatory, a part of the Steward Observatory of the University of Arizona, with financial support of operations from the State of Arizona and financial support for instrumentation development from the NSF. Support for SPT participation in the EHT is provided by the National Science Foundation through award OPP-1852617 to the University of Chicago. Partial support is also provided by the Kavli Institute of Cosmological Physics at the University of Chicago. The SPT hydrogen maser was provided on loan from the GLT, courtesy of ASIAA.

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), supported by NSF grant ACI-1548562, and CyVerse, supported by NSF grants DBI-0735191, DBI-1265383, and DBI-1743442. XSEDE Stampede2 resource at TACC was allocated through TG-AST170024 and TG-AST080026N. XSEDE JetStream resource at P1T and TACC was allocated through AST170028. This research is part of the Frontera computing project at the Texas Advanced Computing Center through the Frontera Large-Scale Community Partnerships allocation AST20023. Frontera is made possible by National Science Foundation award OAC-1818253. This research was carried out using resources provided by the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy Office of Science. Additional work used ABACUS2.0, which is part of the eScience center at Southern Denmark University. Simulations were also performed on the SuperMUC cluster at the LRZ in Garching, on the LOEWE cluster in CSC in Frankfurt, on the HazelHen cluster at the HLRS in Stuttgart, and on the Pi2.0 and Siyuan Mark-I at Shanghai Jiao Tong University. The computer resources of the Finnish IT Center for Science (CSC) and the Finnish Computing Competence Infrastructure (FCCI) project are acknowledged. This research was enabled in part by support provided by Compute Ontario (http://computeontario.ca), Calcul Quebec (http://www.calculquebec.ca) and Compute Canada (http://www.computecanada.ca).

The EHTC has received generous donations of FPGA chips from Xilinx Inc., under the Xilinx University Program. The EHTC has benefited from expertise and technologies shared under open-source license by the Collaboration for Astronomy Signal Processing and Electronics Research (CASPER). The EHT project is grateful to T4Science and Microsemi for their assistance with Hydrogen Masers. This research has made use of NASA’s Astrophysics Data System. We gratefully acknowledge the support provided by the extended staff of the ALMA, both from the inception of the ALMA Phasing Project through the observational campaigns of 2015 and 2017. We would like to thank A. Deller and W. Brisken for EHT-specific support with the use of DiFX. We thank Martin Shepherd for the addition of extra features in the Difmap software that were used for the CLEAN imaging results presented in this paper. We acknowledge the significance of Maunakea, where the SMA and JCMT EHT stations are located, has for the indigenous Hawaiian people. IMV acknowledges the use of LLuis Vives HPC resources of the University of Valencia.

Facility: EHT.

Software: Comrade, DIFMAP, DFI, eht-imaging, metronization, REx, SMILI, Stan, THEMIS, VIDA.

Appendix A

Comparison of Sgr A* Variability in 2013 and 2017

Additional prior epochs of millimeter-VLBI observations of Sgr A* provide a means to explore a wider range of baseline lengths and assess the consistency of the variability across many years. Johnson et al. (2015) reported 1.3 mm VLBI observations of Sgr A* from 2013 March 21, 22, 23, 26, and 27 with an array that included JCMT, SMA, SMT, and the
Combined Array for Research in Millimeter-wave Astronomy (CARMA, which has since been decommissioned). By virtue of the small number of participating stations, these observations were much more limited in their \((u,v)\)-coverage than the 2017 EHT observations. Nevertheless, the 2013 observations provide a second multiday data set for which normalized variance estimates may be produced and compared to those reported in this paper.

Johnson et al. (2015) reported only the visibility amplitudes, which we average in time on a per-scan basis. This averaging presumes that there is no intrinsic phase evolution in the visibilities over the approximately 10-minute scan lengths; this approximation is well justified for the source sizes inferred from the 2017 EHT observations. The visibility data are normalized by the intrasite SMA-JCMT, SMA-SMA, and CARMA-CARMA baselines. Note that this normalization eliminates the need to perform a phase calibration like that applied to JCMT and LMT for the 2017 EHT data.

Figure 23 shows the reconstructed normalized variances from the 2013 observations in comparison to those from the 2017 EHT observations. Baselines between 2.5 and 4 G\(\lambda\) provide variability estimates that are broadly consistent between the two sets of observations. This agreement suggests that the degree of structural variability exhibited by Sgr A* during the 2017 EHT campaign is not anomalous.

Similarly, because the station gains were well characterized for all stations during the 2013 observations, it is not necessary to make assumptions about the source size on short baselines. Therefore, the 2013 observations provide estimates of the normalized variance on baselines shorter than 2 G\(\lambda\). The large statistical errors of these measurements preclude strong constraints on the variability below 1 G\(\lambda\), but there are nevertheless hints of a turnover in the variability power between 1 and 2 G\(\lambda\).

Appendix B
Origin and Mitigation of Biases in Single-day Analyses

As discussed in Section 3, there is compelling evidence that Sgr A* exhibits structural variability on timescales ranging from minutes to a full observation night. The degree of variability is estimated in Sections 3.3 and 8.1 and found to be the dominant contribution to the difference between the observed visibilities and those associated with a mean image for baselines with lengths between \(\sim 3\) and 6 G\(\lambda\). Underlying the noise modeling mitigation method (see Section 3.4) is the assumption that the added “variability noise” modifies the data in a stochastic manner, i.e., coherent deviations do not persist throughout large patches of the \((u,v)\)-plane. However, Earth-rotation aperture synthesis naturally results in correlated variability between visibilities that are nearby in the \((u,v)\)-plane, because visibilities that have small \((u,v)\) separations tend to also have small temporal separations. Furthermore, the sparsity of the EHT array prevents most locations in the \((u,v)\)-plane from being sampled more than once in a single observation (though see Section 3.2 for several exceptions), meaning that multiple observing days must be combined to access more than a single instantiation of the source variability.

The impact of structural variability can be seen most prominently on the Chile-LMT baselines, which exhibit coherent deviations on \(\sim 1\) hr timescales that are evident in the visibility amplitudes presented in Figure 24. These deviations are most pronounced near 1 hr GMST \((\sim 12\) UTC), where they are evident on both April 6 and April 7. The residual gain uncertainties of \(\sim 10\%-20\%\) (Paper II) are insufficient to explain the dramatic drop near 4 G\(\lambda\) around 1 hr GMST on April 6. At the same GMST on April 7, the visibility amplitudes fluctuate upward by a similar amount. On the remaining days the amplitudes at this GMST lie between the April 6 and April 7 values, indicating that the variations are associated with a process that is uncorrelated on interday timescales.

Similar coherent deviations are also observed in the synthetic data sets produced from GRMHD simulations for the purposes of calibrating ring size measurements to a common physical scale (see Section 4.4 and Appendix D). Figure 25 shows a similar set of visibility amplitudes on the Chile-LMT baselines for one of these synthetic data sets (data set 092 in Table 7). While the date and time of the spurious feature differ from those seen in the real Sgr A* data—arising in the simulated data set near 3 hr GMST \((\sim 14\) UTC)—a dramatic deviation is present and persists for \(\sim 0.5\) hr. In both the Sgr A* data and GRMHD simulations, these observed deviations are \(\gtrsim 3\sigma\) outliers (per the variance expected from the variability quantification scheme detailed in Section 3.3), they are nearly exclusively confined to the Chile-LMT baseline, and they are rare. However, due to their coherent nature—i.e., many data points on a given day are similarly displaced—these fluctuations violate the assumption of statistical independence made by the noise modeling mitigation scheme on a single day.

The origin of the visibility amplitude deviations for the GRMHD simulations can be identified with coherent variable structures moving about the ring. As shown in Figure 26, instantaneous images from the GRMHD simulations can differ qualitatively from the average image, with the former sometimes dominated by small, bright patches of emission. When aligned in the NW–SE direction and separated by the \(\sim 50\) \(\mu\)as ring diameter, these bright emission regions significantly impact the visibilities on the \(|u| \approx 4\) G\(\lambda\) Chile-LMT baselines. Days that do not exhibit large variations correspond to periods less impacted by such patchy emission structures.

The coherent deviations seen on the Chile-LMT baselines can manifest as pathological behaviors in the single-day Sgr A*
analyses carried out in this paper and in Paper III, particularly when applied to the sparser April 6 data set (see left column of Figure 27). The specific manner in which the visibility fluctuations impact the reconstructed source structure depends on the details of the analysis scheme and the freedom each underlying image model has to accommodate a subset of visibilities that are discrepant with the time-averaged structure.

Images reconstructed from the April 6 Sgr A* data exhibit parallel NE–SW streaks, typical of baseline artifacts associated with miscalibration of a single baseline, and a natural consequence of visibility amplitudes on the Chile-LMT baseline that are discrepant with the time-averaged image. The top left panel of Figure 27 shows such artifacts in a THEMIS image reconstruction; similar image artifacts are observed in the RML (eht-imaging and SMILI) and, to a lesser extent, the CLEAN (DIFMAP) reconstructions presented in Paper III.

By virtue of their specification, the mG-ring source models are not capable of introducing streak-like features into the image structure. When applied to Sgr A*, full-track mG-ring fits to the April 6 data instead exhibit smaller ring sizes than those applied to the April 7 data (see the first and third panels in the second row of Figure 27), with the April 6 fits preferring rings with a ∼40 μas diameter and the April 7 fits preferring rings with a ∼55 μas diameter. This discrepancy in ring size may be associated with a shift in the location of the visibility minimum from ∼3 to ∼4 Gλ between April 7 and April 6, respectively. However, we note that the behavior of the full-track mG-ring model when applied to the April 6 data is a sensitive function of the m-order; changing the m-order can cause the model to prefer a ∼55 μas diameter. Snapshot mG-ring fits exhibit qualitatively similar behavior to the full-track fits, as shown in the third row of Figure 27. We again find that the April 6 data prefer a somewhat smaller ring diameter than the April 7 fits, though for snapshot fits the posterior distributions for the diameter parameter are consistent between the 2 days.

Despite their disparate forms, the various artifacts observed in April 6 reconstructions are effectively ameliorated by flagging the Chile-LMT baselines prior to carrying out the
### Table 7: GRMHD Synthetic Data Set Parameters

<table>
<thead>
<tr>
<th>Data Set Index</th>
<th>Accretion State</th>
<th>Spin</th>
<th>Inclination</th>
<th>$R_{\text{high}}$</th>
<th>Position Angle</th>
<th>Input $\theta_g$</th>
</tr>
</thead>
<tbody>
<tr>
<td>000</td>
<td>MAD –0.94</td>
<td>10</td>
<td>10</td>
<td>155</td>
<td>4.477</td>
<td></td>
</tr>
<tr>
<td>001</td>
<td>MAD –0.94</td>
<td>10</td>
<td>40</td>
<td>–114</td>
<td>4.140</td>
<td></td>
</tr>
<tr>
<td>002</td>
<td>MAD –0.94</td>
<td>10</td>
<td>160</td>
<td>24</td>
<td>5.317</td>
<td></td>
</tr>
<tr>
<td>003</td>
<td>MAD –0.94</td>
<td>50</td>
<td>10</td>
<td>168</td>
<td>5.493</td>
<td></td>
</tr>
<tr>
<td>004</td>
<td>MAD –0.94</td>
<td>50</td>
<td>40</td>
<td>90</td>
<td>5.493</td>
<td></td>
</tr>
<tr>
<td>005</td>
<td>MAD –0.94</td>
<td>50</td>
<td>160</td>
<td>89</td>
<td>6.423</td>
<td></td>
</tr>
<tr>
<td>006</td>
<td>MAD –0.94</td>
<td>90</td>
<td>10</td>
<td>–177</td>
<td>3.843</td>
<td></td>
</tr>
<tr>
<td>007</td>
<td>MAD –0.94</td>
<td>90</td>
<td>40</td>
<td>–73</td>
<td>5.503</td>
<td></td>
</tr>
<tr>
<td>008</td>
<td>MAD –0.94</td>
<td>90</td>
<td>160</td>
<td>112</td>
<td>6.152</td>
<td></td>
</tr>
<tr>
<td>009</td>
<td>MAD –0.5</td>
<td>10</td>
<td>10</td>
<td>168</td>
<td>5.704</td>
<td></td>
</tr>
<tr>
<td>010</td>
<td>MAD –0.5</td>
<td>10</td>
<td>40</td>
<td>51</td>
<td>5.684</td>
<td></td>
</tr>
<tr>
<td>011</td>
<td>MAD –0.5</td>
<td>10</td>
<td>160</td>
<td>–12</td>
<td>4.502</td>
<td></td>
</tr>
<tr>
<td>012</td>
<td>MAD –0.5</td>
<td>50</td>
<td>10</td>
<td>–22</td>
<td>5.724</td>
<td></td>
</tr>
<tr>
<td>013</td>
<td>MAD –0.5</td>
<td>50</td>
<td>40</td>
<td>178</td>
<td>5.563</td>
<td></td>
</tr>
<tr>
<td>014</td>
<td>MAD –0.5</td>
<td>50</td>
<td>160</td>
<td>105</td>
<td>4.039</td>
<td></td>
</tr>
<tr>
<td>015</td>
<td>MAD –0.5</td>
<td>90</td>
<td>10</td>
<td>–171</td>
<td>5.935</td>
<td></td>
</tr>
<tr>
<td>016</td>
<td>MAD –0.5</td>
<td>90</td>
<td>40</td>
<td>146</td>
<td>3.596</td>
<td></td>
</tr>
<tr>
<td>017</td>
<td>MAD –0.5</td>
<td>90</td>
<td>160</td>
<td>–3</td>
<td>5.110</td>
<td></td>
</tr>
<tr>
<td>018</td>
<td>MAD 0.0</td>
<td>10</td>
<td>10</td>
<td>35</td>
<td>3.677</td>
<td></td>
</tr>
<tr>
<td>019</td>
<td>MAD 0.0</td>
<td>10</td>
<td>40</td>
<td>143</td>
<td>5.719</td>
<td></td>
</tr>
<tr>
<td>020</td>
<td>MAD 0.0</td>
<td>10</td>
<td>160</td>
<td>115</td>
<td>5.030</td>
<td></td>
</tr>
<tr>
<td>021</td>
<td>MAD 0.0</td>
<td>50</td>
<td>10</td>
<td>112</td>
<td>3.813</td>
<td></td>
</tr>
<tr>
<td>022</td>
<td>MAD 0.0</td>
<td>50</td>
<td>40</td>
<td>–101</td>
<td>4.301</td>
<td></td>
</tr>
<tr>
<td>023</td>
<td>MAD 0.0</td>
<td>50</td>
<td>160</td>
<td>–12</td>
<td>4.909</td>
<td></td>
</tr>
<tr>
<td>024</td>
<td>MAD 0.0</td>
<td>90</td>
<td>10</td>
<td>76</td>
<td>4.059</td>
<td></td>
</tr>
<tr>
<td>025</td>
<td>MAD 0.0</td>
<td>90</td>
<td>40</td>
<td>11</td>
<td>4.029</td>
<td></td>
</tr>
<tr>
<td>026</td>
<td>MAD 0.0</td>
<td>90</td>
<td>160</td>
<td>97</td>
<td>6.323</td>
<td></td>
</tr>
<tr>
<td>027</td>
<td>MAD 0.5</td>
<td>10</td>
<td>10</td>
<td>40</td>
<td>3.974</td>
<td></td>
</tr>
<tr>
<td>028</td>
<td>MAD 0.5</td>
<td>10</td>
<td>40</td>
<td>–4</td>
<td>4.658</td>
<td></td>
</tr>
<tr>
<td>029</td>
<td>MAD 0.5</td>
<td>10</td>
<td>160</td>
<td>126</td>
<td>6.272</td>
<td></td>
</tr>
<tr>
<td>030</td>
<td>MAD 0.5</td>
<td>50</td>
<td>10</td>
<td>–42</td>
<td>4.472</td>
<td></td>
</tr>
<tr>
<td>031</td>
<td>MAD 0.5</td>
<td>50</td>
<td>40</td>
<td>25</td>
<td>4.089</td>
<td></td>
</tr>
<tr>
<td>032</td>
<td>MAD 0.5</td>
<td>50</td>
<td>160</td>
<td>–135</td>
<td>5.598</td>
<td></td>
</tr>
<tr>
<td>033</td>
<td>MAD 0.5</td>
<td>90</td>
<td>10</td>
<td>108</td>
<td>5.251</td>
<td></td>
</tr>
<tr>
<td>034</td>
<td>MAD 0.5</td>
<td>90</td>
<td>40</td>
<td>171</td>
<td>5.432</td>
<td></td>
</tr>
<tr>
<td>035</td>
<td>MAD 0.5</td>
<td>90</td>
<td>160</td>
<td>140</td>
<td>5.015</td>
<td></td>
</tr>
<tr>
<td>036</td>
<td>MAD 0.94</td>
<td>10</td>
<td>10</td>
<td>–54</td>
<td>5.679</td>
<td></td>
</tr>
<tr>
<td>037</td>
<td>MAD 0.94</td>
<td>10</td>
<td>40</td>
<td>1</td>
<td>4.200</td>
<td></td>
</tr>
<tr>
<td>038</td>
<td>MAD 0.94</td>
<td>10</td>
<td>160</td>
<td>–92</td>
<td>5.915</td>
<td></td>
</tr>
<tr>
<td>039</td>
<td>MAD 0.94</td>
<td>50</td>
<td>10</td>
<td>–2</td>
<td>6.282</td>
<td></td>
</tr>
<tr>
<td>040</td>
<td>MAD 0.94</td>
<td>50</td>
<td>40</td>
<td>161</td>
<td>5.131</td>
<td></td>
</tr>
<tr>
<td>041</td>
<td>MAD 0.94</td>
<td>50</td>
<td>160</td>
<td>–89</td>
<td>5.699</td>
<td></td>
</tr>
<tr>
<td>042</td>
<td>MAD 0.94</td>
<td>90</td>
<td>10</td>
<td>–48</td>
<td>5.025</td>
<td></td>
</tr>
<tr>
<td>043</td>
<td>MAD 0.94</td>
<td>90</td>
<td>40</td>
<td>–99</td>
<td>4.587</td>
<td></td>
</tr>
<tr>
<td>044</td>
<td>MAD 0.94</td>
<td>90</td>
<td>160</td>
<td>54</td>
<td>4.467</td>
<td></td>
</tr>
<tr>
<td>045</td>
<td>SANE –0.94</td>
<td>10</td>
<td>10</td>
<td>97</td>
<td>5.880</td>
<td></td>
</tr>
<tr>
<td>046</td>
<td>SANE –0.94</td>
<td>10</td>
<td>40</td>
<td>127</td>
<td>6.388</td>
<td></td>
</tr>
<tr>
<td>047</td>
<td>SANE –0.94</td>
<td>10</td>
<td>160</td>
<td>–142</td>
<td>6.267</td>
<td></td>
</tr>
<tr>
<td>048</td>
<td>SANE –0.94</td>
<td>50</td>
<td>10</td>
<td>–59</td>
<td>6.016</td>
<td></td>
</tr>
<tr>
<td>049</td>
<td>SANE –0.94</td>
<td>50</td>
<td>40</td>
<td>144</td>
<td>3.652</td>
<td></td>
</tr>
<tr>
<td>050</td>
<td>SANE –0.94</td>
<td>50</td>
<td>160</td>
<td>–110</td>
<td>4.411</td>
<td></td>
</tr>
<tr>
<td>051</td>
<td>SANE –0.94</td>
<td>90</td>
<td>10</td>
<td>46</td>
<td>3.783</td>
<td></td>
</tr>
<tr>
<td>052</td>
<td>SANE –0.94</td>
<td>90</td>
<td>40</td>
<td>–129</td>
<td>5.075</td>
<td></td>
</tr>
<tr>
<td>053</td>
<td>SANE –0.94</td>
<td>90</td>
<td>160</td>
<td>68</td>
<td>6.106</td>
<td></td>
</tr>
<tr>
<td>054</td>
<td>SANE –0.5</td>
<td>10</td>
<td>10</td>
<td>53</td>
<td>5.276</td>
<td></td>
</tr>
<tr>
<td>055</td>
<td>SANE –0.5</td>
<td>10</td>
<td>40</td>
<td>76</td>
<td>4.281</td>
<td></td>
</tr>
<tr>
<td>056</td>
<td>SANE –0.5</td>
<td>10</td>
<td>160</td>
<td>144</td>
<td>4.854</td>
<td></td>
</tr>
<tr>
<td>057</td>
<td>SANE –0.5</td>
<td>50</td>
<td>10</td>
<td>–173</td>
<td>6.418</td>
<td></td>
</tr>
<tr>
<td>058</td>
<td>SANE –0.5</td>
<td>50</td>
<td>40</td>
<td>55</td>
<td>5.070</td>
<td></td>
</tr>
<tr>
<td>059</td>
<td>SANE –0.5</td>
<td>50</td>
<td>160</td>
<td>66</td>
<td>5.000</td>
<td></td>
</tr>
<tr>
<td>060</td>
<td>SANE –0.5</td>
<td>90</td>
<td>10</td>
<td>154</td>
<td>5.080</td>
<td></td>
</tr>
<tr>
<td>061</td>
<td>SANE –0.5</td>
<td>90</td>
<td>40</td>
<td>–154</td>
<td>5.236</td>
<td></td>
</tr>
</tbody>
</table>
proceed instead with the noise modeling scheme described in Section 3.4, which is itself intended to mitigate the effects of intrinsic variability on the source reconstructions.

As described above, the apparent failure of the noise modeling method to produce consistent results on some individual single-day analyses can be attributed to the coherent (rather than stochastic) nature of the variability sampled at any single location in the $(\mu, \nu)$-plane, which violates the key assumption underlying the noise modeling approach that each data point represent an independent sample of the source variability. This assumption would be more faithfully adhered to by a data set containing a larger number of independent variability realizations; within the context of the EHT Sgr A* observations, additional variability realizations are most naturally acquired by combining data sets across observing days. Multiday analyses are also more consistent in spirit with the model-agnostic variability estimates in Sections 3.3 and 8.1. Combining the independent realizations of the structural variability from multiple days improves the estimate of the mean visibilities, as evident in Figure 24, in which the multiday mean of the visibility amplitudes is both smoother than and intermediate between those of April 6 and 7 individually.

The fourth column in Figure 27 shows reconstructions made using the combined April 6 and 7 data sets. The improved behavior of the visibility means is reflected in better consistency across methods and a reduction in image artifacts.

Figure 26. Average (left) and snapshot (right) images from the GRMHD simulation used to generate the validation synthetic data set shown in Figure 25. Three frames from the portion of the simulation used to generate the April 10 (top, outlined in red) and April 7 (bottom, outlined in blue) synthetic data sets are shown at the GMST times specified (corresponding to the times indicated by the vertical dotted green lines in Figure 25). All of the images share a common brightness color scale; the absolute brightness scale is arbitrary because each image has been normalized to have unit total flux density, and a modest amount of saturation has been permitted in the brightest regions to enhance the visibility of low-brightness features.
Appendix C
Single-day Fits

In this section we present the results from analyses carried out on the April 6 and April 7 data sets individually. Figures 28 and 29 show single-day images from each of the analysis pipelines (analogous to those shown in Figure 20), and Figure 30 shows the corresponding measurements of morphological properties (analogous to those shown in Figure 21).
Figure 28. Same as Figure 20, but showing results obtained from using only the April 6 data.

Figure 29. Same as Figure 20, but showing results obtained from using only the April 7 data.
Appendix D

GRMHD Synthetic Data Set Generation

To calibrate measurements of the angular gravitational radius $\theta_g$ (see Section 4.4), we rely on the library of GRMHD simulations and associated GRRT synthetic movies produced and described in Paper V. In this appendix we provide an overview of our model selection and data generation procedures, which are conceptually similar to the calibration analysis carried out in M87*. Paper VI.

We select 90 simulations from the GRMHD library to be used for $\theta_g$ calibration and another 10 simulations to be used to validate this calibration. The 90 calibration data sets uniformly grid a range of GRMHD parameters: every combination of the two MAD and SANE accretion states, five black hole spin values of $[-0.94, -0.5, 0, 0.5, 0.94]$, three inclinations of $[10^\circ, 50^\circ, 90^\circ]$, and three $R_{\text{high}}$ values of $[10, 40, 160]$ are represented. The 10 validation data sets are split evenly between MAD and SANE, but the black hole spins are randomly selected from $[-0.94, -0.5, 0, 0.5, 0.94]$, the inclinations are randomly selected from $[30^\circ, 70^\circ, 110^\circ, 130^\circ, 150^\circ]$, and the $R_{\text{high}}$ values are randomly selected from $[10, 40, 160]$. The resulting model images contain a large variety of possible image morphologies, along with self-consistent dynamics as governed by the equations of GRMHD and GRRT. The GRMHD parameters corresponding to each selected calibration and validation model are listed in Table 7, and some example average images are shown in Figure 31.

After selecting the GRMHD models and prior to generating synthetic data, we first modify their orientations and angular sizes from their default simulation values. We rotate each simulated movie by a position angle that is a uniformly chosen integer in the range $[-180^\circ, 180^\circ]$. Each of the simulations from Paper V was produced assuming a mass of $M = 4.14 \times 10^6 M_\odot$ and a distance of 8.127 kpc (Do et al. 2019; Gravity Collaboration et al. 2019), corresponding to an angular gravitational radius of $\theta_g = 5.03 \mu$as. To avoid biasing our calibration in favor of any one value of $\theta_g$, we modify the overall spatial scale of each simulated movie by a random factor that is uniformly drawn from the range $[0.7, 1.3]$. The input position angles and gravitational radii for each movie are listed in Table 7.

Once the GRMHD models are selected and their movies rotated and scaled, we generate synthetic data sets in the same manner as for the synthetic data sets used in Paper III. We use the eht-imaging software to first apply artificial scattering to the source structure per the scattering model from Johnson et al. 2022.
et al. (2018) and then sample the Fourier transform of each movie at a cadence and at \((u,v)\) locations identical to those of the EHT observations of Sgr A*.

The resulting visibilities are then corrupted with thermal noise and station-based gain and leakage effects at a level that is consistent with the Sgr A* data (Paper II). Eight synthetic data sets are generated for each GRMHD model, corresponding to the \((u,v)\)-coverage on four observing nights—2017 April 5, 6, 7, and 10—and two frequency bands (see Section 2).

### Appendix E

**Representative \(\chi^2\) Values for Each Analysis Method**

In this section we provide some example representative \(\chi^2\) values and associated quantities for each of the analysis methods used in this paper. Specifically, we report values corresponding to the example fits shown in Figures 9, 15, and 17.

For any fitted data quantity \(q\) with modeled counterpart \(\hat{q}\) and associated measurement uncertainty \(\sigma\), we determine the \(\chi^2\) as

\[
\chi^2 = \sum_i \left( \frac{q_i - \hat{q}_i}{\sigma_i} \right)^2,
\]

where the sum is understood to be taken over all \(N_{\text{data}}\) fitted data points. We also define a reduced-\(\chi^2\) value,

\[
\chi^2_{\text{red}} = \frac{\chi^2}{N_{\text{dof}}},
\]

where \(N_{\text{dof}}\) is the number of degrees of freedom remaining in the data after accounting for the free parameters in the model.

However, we note that despite their familiarity, the interpretation of either of these \(\chi^2\) statistics is complicated by several aspects of the analyses presented in this paper.

The first complication is that the number of degrees of freedom is generally unknown, rendering \(\chi^2_{\text{red}}\) difficult to define in practice. The use of informative priors and the presence of correlations among model parameters mean that \(N_{\text{dof}}\) cannot be determined as simply the difference between \(N_{\text{data}}\) and the number of free parameters in the model. For example, in the RML imaging methods (eht-imaging and SMILI), the number of effective image parameters is implicitly reduced—relative to the number that would be assumed by simply counting the total number of image pixels—by a large factor by the imposition of regularization terms (e.g., smoothness, sparseness) in the objective function (Paper III). Additionally, for all methods that fit to complex visibilities or visibility amplitudes, station gains are reconstructed as part of the fitting process; for those methods that simultaneously reconstruct high- and low-band data, the gains for the two bands are necessarily strongly correlated. Furthermore, strong priors are imposed by network calibration (Paper II; M87* Paper III), further reducing the number of effective model parameters and growing the effective number of degrees of freedom.

The second complication is that, with the exception of the snapshot modeling presented in Section 6 (Comrade and DPI), all models make use of an added uncertainty budget to account for source variability. For the THEMIS imaging and full-track modeling analyses, the parameters describing this excess variability noise are simultaneously fit alongside those describing the image structure. In both cases, the impact is to

Event Horizon Telescope Collaboration et al.

Table 8

Representative \( \chi^2 \) Values and Ancillary Model Information

<table>
<thead>
<tr>
<th>Method</th>
<th>Day</th>
<th>Band</th>
<th>( \chi^2 )</th>
<th>( N_{\text{data}} )</th>
<th>( N_{\text{gains}} )</th>
<th>( N_{\text{params}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>eht-imaging imaging</td>
<td>Apr 7</td>
<td>HI+LO</td>
<td>4174.5</td>
<td>12082</td>
<td>6358</td>
<td>6400</td>
</tr>
<tr>
<td>SMILI imaging</td>
<td>Apr 7</td>
<td>HI+LO</td>
<td>2192.9</td>
<td>6806</td>
<td>3562</td>
<td>5625</td>
</tr>
<tr>
<td>DIFMAP imaging</td>
<td>Apr 7</td>
<td>HI+LO</td>
<td>4194.5</td>
<td>12082</td>
<td>6358</td>
<td>2262</td>
</tr>
<tr>
<td>THEMIS imaging</td>
<td>Apr 6+7</td>
<td>HI+LO</td>
<td>1444.5</td>
<td>2990</td>
<td>1323</td>
<td>58</td>
</tr>
<tr>
<td>THEMIS full-track modeling</td>
<td>Apr 6+7</td>
<td>LO</td>
<td>1237.0</td>
<td>1562</td>
<td>666</td>
<td>19</td>
</tr>
<tr>
<td>Comrade snapshot modeling</td>
<td>Apr 6+7</td>
<td>LO</td>
<td>1987.2</td>
<td>4977</td>
<td>1345</td>
<td>3328</td>
</tr>
<tr>
<td>DPI snapshot modeling</td>
<td>Apr 6+7</td>
<td>LO</td>
<td>1752.7</td>
<td>3176</td>
<td></td>
<td>1664</td>
</tr>
</tbody>
</table>

Note. The \( \chi^2 \) values, number of data points \( (N_{\text{data}}) \), number of station gain parameters \( (N_{\text{gains}}) \), and number of parameters describing the image structure \( (N_{\text{params}}) \) for representative example fits from each of the analysis methods used in this paper. For the eht-imaging and SMILI analyses, \( N_{\text{params}} \) is equal to the number of pixels in the image Paper III; for the DIFMAP analyses, \( N_{\text{params}} \) is determined by the number of CLEAN components, with each component contributing three parameters. For the THEMIS imaging and full-track modeling analyses, \( N_{\text{params}} \) is equal to the number of model parameters used to specify the image structure (see Paper III and Section 7, respectively). For the Comrade and DPI snapshot modeling analyses, each of the \( N_{\text{params}} \) snapshots contributes 13 and 8 parameters, respectively, to the total \( N_{\text{params}} \) count (see Section 6). Note that because the DPI analysis fits only to closure quantities, it does not contain any station gain parameters.

* Note that the listed numbers of gains and model parameters are solely those necessary to forward-compute the model data values, i.e., these values characterize a property of the model specification. Importantly, due to model nonlinearity and strong correlations between parameters, these numbers are not generally suitable for determining \( N_{\text{params}} \) (see Appendix E).

Appendix F

Snapshot Modeling Likelihood Functions

In this appendix we provide specific expressions for the likelihood functions used during the snapshot geometric modeling analyses described in Section 6. We assume the high signal-to-noise ratio limit for all data products, which is not strictly satisfied for the relatively short integration times (120 s) employed in the snapshot modeling analyses, but which has the benefit of reducing all likelihood functions to Gaussians.

F.1. Visibility Amplitude Likelihood

For each snapshot and baseline, the visibility amplitudes are distributed according to a Rice distribution, which in the high signal-to-noise ratio limit becomes Gaussian (e.g., Wardle & Kronberg 1974; Broderick et al. 2020a). We can thus write the visibility amplitude likelihood function as

\[
\mathcal{L}_{|V|,ab} = \frac{1}{\sqrt{2\pi \sigma_{ab}^2}} \exp \left[ -\frac{(|V_{ab}| - |g_i||g_j||V_{ij}|)^2}{2\sigma_{ab}^2} \right],
\]

where \( b \) is a baseline index that runs over all station pairs \( \{i, j\} \) in snapshot \( s \). Here \( |V_{ij}| \) is the model visibility amplitude and \( |g_i| \) and \( |g_j| \) are the individual station gain amplitudes (see Equation (2)). We use this Gaussian approximation to the Rice distribution for all of the snapshot geometric modeling analyses. For a snapshot \( s \), the joint visibility amplitude likelihood across all baselines is then given by

\[
\mathcal{L}_{|V|,s} = \prod_b \mathcal{L}_{|V|,ab},
\]

where the product is taken over all baselines \( b \).

F.2. Closure Phase Likelihood

The visibility phases in EHT data sets are heavily corrupted by atmospheric fluctuations (M87’ Paper II; M87’ Paper III), so all of our snapshot geometric modeling analyses work instead with closure phases \( \psi \) (see Equation (3)). In the high signal-to-noise ratio limit, the variance in the closure phase on the triangle containing stations \( i, j, \) and \( k \) is given by

\[
\sigma_{\psi,ijk}^2 = \sigma_{\text{ln}|V|,ij}^2 + \sigma_{\text{ln}|V|,ik}^2 + \sigma_{\text{ln}|V|,ji}^2,
\]

where

\[
\sigma_{\text{ln}|V|,ij} = \frac{\sigma_{ij}}{|V_{ij}|}
\]

is the uncertainty in the log visibility amplitude in the same limit and \( \sigma_{ij} \) is the uncertainty in \( V_{ij} \). Hereafter we replace the triangle indices \( ijk \) with a single multi-index \( t \) for clarity. We approximate the closure phase likelihood for a single triangle \( t \) and snapshot \( s \) by a von Mises distribution,

\[
\mathcal{L}_{\psi,t} = \frac{1}{2\pi I_0(\sigma_{\hat{\psi},t}^2)} \exp \left[ \frac{\cos(\psi_t - \hat{\psi}_t)}{\sigma_{\psi,t}^2} \right],
\]

where \( \hat{\psi}_t \) denotes a measured closure phase and \( I_0(\cdot) \) is a modified Bessel function of the first kind of order 0. In the high signal-to-noise ratio limit, the von Mises becomes a Gaussian distribution with mean \( \hat{\psi}_t \) and standard deviation \( \sigma_{\psi,t} \). Note that

 drive the \( \chi^2_{\text{red}} \) value toward unity, rendering the resulting \( \chi^2 \) value not particularly meaningful as a metric of fit quality. Nevertheless, in Table 8 we present the \( \chi^2 \) values and relevant properties of the data sets and models used for representative examples from each analysis pathway. In lieu of a well-defined number of degrees of freedom, we consider two limits. An optimistic estimate is given by the procedure adopted in Paper III, in which \( N_{\text{dof}} = N_{\text{data}} \) appropriate when the total number of effective model parameters is much less than the number of data points. In this limit, \( \chi^2 \) ranges from 0.35 to 0.79 for the fits listed in Table 8. A more pessimistic accounting is given by \( N_{\text{dof}} = N_{\text{data}} - N_{\text{gains}} - N_{\text{params}} \). This quantity is negative for some analyses, a consequence of the strong correlations that limit the effective number of model parameters in practice (e.g., for RML imaging and for certain individual snapshot models). Among those analysis methods that do not exhibit this pathology in the \( N_{\text{dof}} \), the pessimistic \( \chi^2_{\text{red}} \) estimates range from 0.9 to 1.4.
up to a normalization $\ln L_\psi$ is equal to the closure phase $\chi^2$ defined in Paper III.

Because a full set of closure phases is highly redundant (i.e., the set is not linearly independent; see Blackburn et al. 2020), we use eht-imaging (Chael et al. 2016, 2018) to construct a minimal nonredundant set within every snapshot. The minimal set of closure phases is constructed by selecting all triangles that contain the most sensitive station in the array (typically ALMA), which ensures that the resulting closure phases are minimally covariant. We ignore the remaining covariances and approximate the joint closure phase likelihood for a snapshot $s$ as

$$L_{\psi,s} = \prod_t L_{\psi,t},$$

where the product is taken over all triangles $t$ in the minimal set.

### F.3. Log Closure Amplitude Likelihood

Some of our snapshot modeling fits use log closure amplitudes in place of visibility amplitudes. In the high signal-to-noise ratio limit, the variance in the log closure amplitudes in place of visibility amplitudes. In the high minimal set.

Hereafter we replace the quadrangle indices $ijk\ell$ with a single multi-index $q$ for clarity. For a single log closure amplitude on quadrangle $q$, we thus have the Gaussian likelihood

$$L_{\ln A,q} = \frac{1}{\sqrt{2\pi\sigma^2_{\ln A,q}}} \exp \left[ -\frac{(\ln A_{sq} - \ln \hat{A}_{sq})^2}{2\sigma^2_{\ln A,q}} \right].$$

in the same limit. We use eht-imaging to construct a minimal nonredundant set of log closure amplitudes within every snapshot, and for the joint likelihood over all of the linearly independent quadrangles within that snapshot we again ignore covariance and treat the data products as being statistically independent. The joint likelihood is then given by

$$L_{A,s} = \prod_q L_{A,q},$$

where the product is taken over all quadrangles $q$ in the minimal set.

### Appendix G

**Snapshot Modeling Prior Distribution Consistency**

In specifying the hypermodel for combining geometric modeling results from individual snapshots (see Section 6.3), we have selected a set of snapshot priors $\pi(\theta_s)$, an average model prior $\pi(\hat{\theta})$, and a hypermodel $\pi(\theta|\hat{\theta})$. A self-consistent set of hypermodel and snapshot priors should satisfy the relation

$$\pi(\theta_s) = \int \pi(\theta_s|\mu, \sigma)\pi(\mu)\pi(\sigma)d\mu d\sigma.$$  

However, because we have carried out the hierarchical modeling from Section 6 in two stages—i.e., first fitting the mG-ring model to individual snapshots, then combining the results from all snapshots—the priors on the model parameters during each stage are selected for local convenience, and the relation in Equation (G1) does not strictly hold.

As a result, the final posteriors for each parameter behave as though the individual snapshot fits were carried out using effective priors that differ from the priors we actually imposed (see Table 1). Figure 32 illustrates how the original priors differ from the effective priors induced by the snapshot-combining procedure. The net result is a modest ($\leq 20\%$ maximum probability density deviation) bias for parameter values to fall toward the center of the original (flat) prior range and away from the edges of that range. Because the widths of the individual snapshot posteriors are typically much narrower than the full prior range (see, e.g., Figure 13), we expect the impact of this centralizing bias on our reported values to be negligible.

### Appendix H

**Snapshot Geometric Modeling Validation Tests**

We cross-validate the results of each snapshot modeling software using a single snapshot of the low-band Sgr A* data from April 7 processed with the HOPS pipeline. The specific snapshot selected begins at 12.65 hr UTC, where the $(u,v)$-coverage of the observation is maximized (Farah et al. 2022). Given the differences in model specification and fitted data products between Comrade/eht-imaging and DPI, we run two separate cross-validation tests:

1. The first test compares the results between Comrade and eht-imaging when fitting an $m = 3$ mG-ring model to visibility amplitudes (including gain amplitudes as model
Figure 33. Comparison of the 2D joint posterior distributions obtained from fitting an mG-ring model to a 120 s snapshot starting at 12.65 hr UT in the Sgr A* April 7 HOPS low-band data set. The left plot compares the results from Comrade (blue) and eht-imaging (orange), fitting to visibility amplitudes (including gains) and closure phases with an $m = 3$ mG-ring. The right triangle plot shows the results from Comrade (blue), eht-imaging (orange), and DPI (green), fitting to closure amplitudes and closure phases for an $m = 2$ mG-ring. In both cases we only show the results for the diameter, width, and fractional Gaussian component flux parameters. Since DPI fits the diameter of the blurred m-ring $d^*$ (Equation (42)), the DPI diameter was debiased so that it corresponds to the infinitesimally thin m-ring diameter that is fit by eht-imaging and Comrade (see also M87 Paper VI). The contours show $1\sigma$, $2\sigma$, and $3\sigma$ levels of the posterior distributions.

parameters) and closure phases. We do not include DPI in this comparison because it cannot currently fit for station gain parameters.

2. The second test compares the results between Comrade, DPI, and eht-imaging when fitting an $m = 2$ mG-ring model to log closure amplitudes and closure phases.

Figure 33 shows the diameter, thickness, and fractional central flux posteriors obtained from performing the tests described above. The posteriors show generally good agreement across codes.

Appendix I
Analysis Specifics and Validation of Calibration Strategy

Our calibration strategy for determining the scaling factor $\alpha$ that relates measured ring diameters to intrinsic angular gravitational radii $\theta_\text{g}$ is described in Section 4.4. In this appendix, we summarize the elements of this strategy that are specific to the different analysis pathways described in Sections 5, 6, and 7. We also validate the calibration procedure by applying the calibrated $\alpha$ values to ring diameter measurements from 10 synthetic GRMHD data sets. For these data sets we know the underlying ground-truth $\theta_\text{g}$ values, and so we can use them to verify whether our measurement and calibration strategy is working as intended.

I.1. IDFE Specifics

To perform an IDFE-based $\theta_\text{g}$ calibration, top-set and posterior images are produced for each of the 90 synthetic GRMHD-based data sets described in Section 4.4 (see also Appendix D). Each of these images is run through both REx and VIDA in the same manner as described in Section 5 for the Sgr A* data.

I.2. Snapshot Geometric Modeling Specifics

We carry out snapshot geometric modeling of the GRMHD calibration and validation data sets in the same manner as described in Section 6 for the Sgr A* data. For all synthetic data sets we use the same data preparation and snapshot timescale as for the fits to the Sgr A* data. We also retain the same model specification, fitting an $m = 4$ mG-ring for all Comrade analyses and an $m = 2$ mG-ring for all DPI analyses. The DPI analyses are carried out on the low-band data sets only, while the Comrade analyses are carried out on both low- and high-band data sets.

I.3. Full-track Geometric Modeling Specifics

We carry out full-track geometric modeling of the GRMHD calibration and validation data sets in the same manner as described in Section 7 for the Sgr A* data. For all synthetic data sets we use the same data preparation as for the fits to the Sgr A* data (see Section 7.1). In particular, we derive appropriately individualized priors on the noise model parameters by performing model-agnostic variability quantification (per Section 3.3) on multiday instantiations (corresponding to April 5, 6, 7, and 10; see Appendix D) of each of the synthetic data sets. We also retain the same geometric model specification, fitting an $m = 4$ mG-ring for all analyses.
Figure 34. Distributions of recovered $\theta_g$ values relative to the known input value, from the GRMHD validation exercise for each analysis pathway. For geometric modeling results the left- and right-hand distributions show the results from fitting to LO and HI bands, respectively. For the IDFE results, the left- and right-hand distributions show the results from using REx and VIDA, respectively. No metronization-based culling has been applied to the IDFE results.
I.4. Validation

For each of the analysis pathways, we validate the $\theta_g$ calibration using an additional 10 synthetic GRMHD-based data sets. Figure 34 shows the results of carrying out ring diameter measurements and subsequent $\theta_g$ conversions on these 10 validation suite data sets, for each of the IDFE, snapshot, and full-track analyses. All analysis pathways are able to successfully recover the correct value of $\theta_g$ to within their determined level of calibration uncertainty.

I.5. Origin and Nature of Calibration Outliers

The distributions of calibrated $\alpha$ values from each ring diameter measurement technique show heavy tails toward small $\alpha$, which manifest as heavy tails to large $\theta_g$ in Figure 34 (see also Figure 22). This behavior appears to be generic across all classes of geometric modeling and IDFE analyses used in this paper, and it implies that some fraction of the calibration data sets are reconstructed to have systematically smaller rings than would be predicted from the known values of $\theta_g$ in each of the input ground-truth simulations.

The left half of Figure 35 shows average images from the input GRMHD calibration suite simulations corresponding to the five smallest median $\alpha$ values (computed across the applicable posterior or top set). Above each collection of images, the number of calibration data sets that find $\alpha < 7.5$ is listed, in comparison to the total number of calibration experiments (HI/LO band, REX/VIDA ring radius measurements). Right: five GRMHD simulations randomly chosen from within the peak of the distribution of $\alpha$ values. Above each collection of images, the number of calibration data sets within one standard deviation of the mean $\alpha$ across the calibration set is listed. In each panel, the corresponding simulation index in Table 7 and the median $\alpha$ across the relevant posterior or top set are given in the upper left and right, respectively.

Figure 35. Time-averaged images for GRMHD simulations that produce anomalously small (left) and typical (right) $\alpha$ calibrations for each geometric modeling and IDFE method used to estimate the mass of Sgr A*. Left: shown are the GRMHD simulations within the calibration set that result in the five smallest median $\alpha$ values (computed across the applicable posterior or top set). Above each collection of images, the number of calibration data sets that find $\alpha < 7.5$ is listed, in comparison to the total number of calibration experiments (HI/LO band, REX/VIDA ring radius measurements). Right: five GRMHD simulations randomly chosen from within the peak of the distribution of $\alpha$ values. Above each collection of images, the number of calibration data sets within one standard deviation of the mean $\alpha$ across the calibration set is listed. In each panel, the corresponding simulation index in Table 7 and the median $\alpha$ across the relevant posterior or top set are given in the upper left and right, respectively.
prominent diffuse emission extending well outside the shadow region of interest. Such structures cannot necessarily be well measured by, e.g., the mG-ring model or IDE techniques aimed at extracting signatures of a ring-like emission morphology, and attempts to apply these techniques to such data sets can yield results that are difficult to interpret. However, we note that not all of the small-α simulations exhibit such morphological difficulties; some of the poor reconstructions are obtained from simulations with readily apparent ring-like structures, indicating that other difficulties (e.g., strong variability) may be playing a more important role in these cases.

The right half of Figure 35 shows average images from the input GRMHD calibration suite simulations corresponding to five “typical” α values recovered by each analysis pathway: each of these images has been randomly selected from the set whose reconstructed α falls within one standard deviation of the mean. In contrast to the small-α simulations, these images more commonly exhibit ring-like morphologies of the sort that we would expect to be amenable to mG-ring modeling or ring extraction techniques. Furthermore, for most analysis methods a large fraction (~75%) of the data sets are contained within one standard deviation of the mean; the fact that this fraction is larger than the α~68% that we would expect for a Gaussian distribution is another manifestation of the heavy tails in the α distributions, and it indicates that the majority of reconstructions are narrowly peaked around the mean (see also Figure 34). However, even among the well-reconstructed images we still find a small number of less obvious ring structures, including some that are dominated by compact emission regions like many of the small-α simulations. Again, the presence of such simulations indicates that the ground-truth emission morphology is not the sole driver of whether or not the underlying ring structure can be successfully reconstructed.

The ring measurement analysis techniques developed in this paper are designed to be appropriate for application to the EHT Sgr A* data. When applying these techniques to a suite of GRMHD simulations containing very diverse image morphologies, we find that a fraction of the reconstructed rings have unreliable diameter measurements. These poor reconstructions contribute to the uncertainty in our α calibration, where they manifest as heavy tails in our calibrated α distribution. The corresponding large uncertainty in α is a consequence of the fact that many of the images in the calibration suite do not resemble Sgr A*, and thus analysis techniques designed for the latter do not necessarily function well when applied to the former. A calibration suite that was more directly tailored to match the properties of the EHT Sgr A* observations may result in smaller α calibration uncertainties and a correspondingly tighter constraint on θ∗.