The Astrophysical Journal, 971:172 (11pp), 2024 August 20 https://doi.org/10.3847/1538-4357/ad566d

Bayesian Constraints on the Ring Ellipticity of M87* 2017 Using THEMIS

Paul Tiede[1,2] and Avery E. Broderick[3,4,5]
[1] Black Hole Initiative at Harvard University, 20 Garden Street, Cambridge, MA 02138, USA; paul.tiede@cfa.harvard.edu
[2] Center for Astrophysics |Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA
[3] Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, ON, N2L 2Y5, Canada
[4] Department of Physics and Astronomy, University of Waterloo, 200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
[5] Waterloo Centre for Astrophysics, University of Waterloo, Waterloo, ON N2L 3G1, Canada
Received 2024 February 1; revised 2024 June 3; accepted 2024 June 7; published 2024 August 16

Abstract

Measuring the properties of black hole images has the potential to constrain deviations from general relativity on horizon scales. Of particular interest is the ellipticity of the ring that is sensitive to the underlying spacetime. In 2019, the Event Horizon Telescope (EHT) produced the first-ever image of a black hole on horizon scales. Here, we reanalyze the M87* EHT 2017 data using Bayesian imaging (BI) techniques, constructing a posterior of the ring shape. We find that BI recovers the true on-sky ring shape more reliably than the original imaging methods used in 2019. As a result, we find that M87*'s ring ellipticity is $0.09^{+0.07}_{-0.06}$ and is consistent with the measured ellipticity from general relativistic magnetohydrodynamic simulations.

Unified Astronomy Thesaurus concepts: Very long baseline interferometry (1769); Astronomy data analysis (1858); Supermassive black holes (1663)

1. Introduction

The Event Horizon Telescope (EHT) imaged the first lensed black hole emission on horizon scales in Event Horizon Telescope Collaboration et al. (2019a, 2019d, hereafter Paper I and Paper IV, respectively). By measuring the size of the image, the EHT provided the first direct measurement of the mass of a supermassive black hole via strong lensing near the horizon (Event Horizon Telescope Collaboration et al. 2019f, hereafter Paper VI). The directly measured EHT mass is consistent with stellar (Gebhardt et al. 2011) and recent gas mass (Liepold et al. 2023; Osorno et al. 2023; Simon et al. 2023) measurements. This consistency provides evidence that the central black hole is consistent with general relativity (GR), as discussed in Paper VI. However, while the size of the measured ring is related to the central mass and, weakly, the spin of the black hole, so is its shape. In particular, the ellipticity of the ring provides a potential probe of the black hole spin and beyond GR effects (Takahashi 2004; Johnson et al. 2020; Medeiros et al. 2020). While GR predicts that the shadow should be highly circular given M87*'s inclination, metrics that violate the no-hair theorem can increase the observed ellipticity in images (Johannsen & Psaltis 2010; Broderick et al. 2014; Johannsen et al. 2016; Gralla & Lupsasca 2020; Gralla et al. 2020). Therefore, measuring a nonzero ellipticity of the shadow of the black hole provides a test of GR on horizon scales.

While the set of image reconstructions presented in Paper IV found nonzero ellipticity, in Tiede et al. (2022a), it was demonstrated that this was due to the imaging method used, the lack of elliptical image structure in the imaging training set, and the EHT coverage. Namely, the set of parameters used to define the distribution of images could not reliably recover the true on-sky ellipticity. It could only constrain the ellipticity of M87*'s image to be less than 3:5. The main source of this uncertainty is that the distribution of images does not provide a statistical measure of uncertainty in image reconstructions. Instead, the distribution is associated with the variation in image reconstructions arising from different choices of the regularized maximum likelihood (RML) hyperparameters, codified in the "top set." Briefly, the top set starts with a coarse grid search over different image hyperparameters, then a set of heuristic cuts based on their performance on some preselected training set, and the image chi-square from fitting the actual data. For complete details of the procedure, see Paper IV. Tiede et al. (2022a) demonstrated that the top-set ellipticity estimates were substantially biased. To account for the biased estimates, a calibration procedure was employed that corrected this bias by imaging 100 general relativistic magnetohydrodynamic (GRMHD) simulation snapshots. The calibration procedure greatly increased the ellipticity uncertainty, preventing a measurement of the ring ellipticity.

The need for this calibration procedure is a direct consequence of the biased ellipticity estimates. Therefore, a new imaging approach or pipeline is needed to substantially reduce the ellipticity estimate bias. One approach to this could be to include various elliptical rings during top-set training. However, choosing the training set is difficult and may introduce new biases in the image reconstructions. Instead, in this paper, we will use Bayesian imaging (BI). BI (Broderick et al. 2020a; Sun & Bouman 2020; Arras et al. 2022; Tiede 2022) provides a statistically motivated procedure to assess uncertainty in reconstructions. As a result of its statistical nature, there is no need for a top set to estimate uncertainty in the image reconstructions. Furthermore, as shown below, BI allows a self-consistent, data-driven approach to hyperparameter optimization and image uncertainty. By combining BI with the image feature extraction tools such as `ReX` and `VIDA` used in Tiede et al. (2022a), we can construct posteriors on image features such as ellipticity, measuring their uncertainty. Therefore, the Bayesian approach may be more immune to the biases found in Tiede et al. (2022a), providing a

direct measurement of image features. To validate the THEMIS and VIDA pipeline and ensure the estimated ellipticity bias is small, we repeat the elliptical ring analysis from Tiede et al. (2022a), demonstrating that the analysis pipeline we develop in this paper provides an accurate estimate of the true on-sky ellipticity. We then apply this same pipeline to the 2017 EHT data and measure the ellipticity of M87* in 2017. The measured ellipticity is then compared to GRMHD simulations to assess whether M87* is consistent with GRMHD simulations.

The paper's layout is as follows: Section 2 describes the BI and feature extraction techniques we use in the paper. Section 3 describes the data products we are fitting and the simulated data generation procedure. Section 4 applies the model to a set of elliptical geometric modeling tests and assesses the amount of bias in the estimated ellipticity. Section 5.2 then applies the same imaging pipeline to the 2017 EHT observations of M87* and compares the results to theoretical expectations from GRMHD simulations.

2. Bayesian Imaging and Feature Extraction

2.1. Image Domain Model

In Broderick et al. (2020a), a new BI technique was developed to quantify image uncertainty statistically. The image model assumes that the on-sky image can be described by a set of rectangular control/raster values c_{ij} positioned at the locations $(x_i, y_i) = (i\mathrm{FOV}_x/(n_x - 1), j\mathrm{FOV}_y/(n_y - 1))$. The on-sky intensity is given by

$$I(x, y) = \sum_{ij} c_{ij}\kappa(x - x_i)\kappa(y - y_j), \tag{1}$$

where κ is an image response or pulse function that converts the discrete set raster to a continuous image. The Fourier transform of the image is

$$\tilde{I}(u, v) = \sum_{ij} e^{2\pi i(ux_i + vy_j)} c_{ij}\tilde{\kappa}(u)\tilde{\kappa}(v), \tag{2}$$

where $\tilde{\kappa}$ is the Fourier transform of the pulse function. The choice of κ is arbitrary. We follow Broderick et al. (2020a) and use a bicubic kernel:

$$\kappa(x) = \begin{cases} 0 & |x| \geqslant 2 \\ b(|x|^3 - 5\,|x|^2 + 8|x| - 4) & 1 \leqslant |x| < 2, \\ (b + 2)|x|^2 - (b + 3)|x|^2 + 1 & |x| < 1 \end{cases} \tag{3}$$

which has Fourier transform

$$\tilde{\kappa}(k) = -\frac{4}{k^3}\sin(k)(2b\cos(k) + (4b + 3)) + \frac{12}{k^4}[b - b\cos(2k) + 2 - 2\cos(k)]. \tag{4}$$

For a derivation of the Fourier transform equations, see Appendix A of Broderick et al. (2020a). The constant b is a free parameter of the interpolation kernel and is typically set to $b = -0.5$.[6] This model has been implemented in THEMIS as `model_image_adaptive_spline_raster`.

In Broderick et al. (2020a), the grid itself was specified a priori, and then the preferred field of view (FOV) was found by performing a parameter survey with the FOV and number of raster pixels as free variables. This paper allows the raster size and orientation ξ to vary during sampling, following the procedure from Broderick et al. (2022). As a result, this imaging method only has two discrete hyperparameters, namely, the number of pixels in the two principal directions of the raster.

For our model priors we first use log-uniform priors between e^{35} Jystr^{-1} and e^{50} Jystr^{-1} on all the raster values c_{ij}. The raster FOVs FOV_x and FOV_y were set to be independent with uniform priors over the interval [15 μas, 120 μas] and a uniform prior on the raster orientation. For the gains, we used Gaussian priors on the amplitudes with unit mean and standard deviation 0.1 on all stations except LMT, which used a prior amplitude of 1.0 to model the large pointing offsets during the 2017 observations (Event Horizon Telescope Collaboration et al. 2019c, hereafter Paper III). The gain phase priors were set to a zero-mean Gaussian with a variance of 10^8 radians.

We will use Bayesian model averaging to find the optimal hyperparameters for our model. For a specific raster, the posterior is given by

$$p(\theta|V, N_x, N_y) = \frac{p(V|N_x, N_y, \theta)p(\theta|N_x, N_y)}{p(V|N_x, N_y)}, \tag{5}$$

where $\theta = (c_{ij}, \mathrm{FOV}_{x,y}, \xi)$ represents the nonraster dimension parameters. We can then promote the discrete raster dimension to model parameters using Bayes' rule and introducing a raster dimension prior $p(N_x, N_y)$:

$$p(\theta, N_x, N_y|V) = \frac{p(\theta|V, N_x, N_y)p(V|N_x, N_y)p(N_x, N_y)}{p(V)}. \tag{6}$$

Therefore, marginalizing over the raster dimension gives the image posterior

$$p(\theta|V) \propto \frac{1}{p(V)}\sum_{N_x,N_y} p(\theta|V, N_x, N_y)p(V|N_x, N_y)p(N_x, N_y). \tag{7}$$

Setting the prior $p(N_x, N_y)$ to be uniform, each single raster dimension posterior $p(\theta|V, N_x, N_y)$ is weighted by the Bayesian evidence. Therefore, selecting the optimal raster dimension is equivalent to selecting the model with the highest Bayesian evidence.

In Section 4, we will consider a small survey of different numbers of pixels and use the evidence, or rather an approximation of it, to find the optimal N_x, N_y. This paper uses the Bayesian information criterion (BIC) as an approximation for the evidence. The BIC is defined as

$$\mathrm{BIC} = \chi^2 + k\ln(N), \tag{8}$$

where k is the total number of nongain parameters fit, N is the number of data points, and χ^2 is the minimum chi-square for the nongain marginalized likelihood, i.e., Equation (14).[7] The BIC is a simple approximation of the Bayesian evidence and compares the optimal performance of the model (first term in Equation (8)) to its complexity (second term). The best model is the one that minimizes the BIC or, conversely, approximately maximizes the Bayesian evidence. We will select the model that minimizes the BIC to define our optimal raster model.

[6] This kernel is not positive definite, so the image may have negative intensity. In practice, we find that this effect is minimal.

[7] Note we do not include gains, since they cancel when finding the ΔBIC.

2.2. Image Feature Extraction

The output of BI is a posterior over the image. An additional processing step, feature extraction, is required to move from the image samples to feature posteriors. In this paper, we will follow Tiede et al. (2022a) and use `VIDA` (Tiede et al. 2022b) to extract the image features. `VIDA` works by constructing image templates t_θ that parameterize the features of interest, e.g., image ellipticity. The optimal template given the image reconstruction is found by minimizing an objective function. `VIDA` uses probability divergences as its objective function, using the correspondence between image intensities and probability distributions modulo total flux. To find the optimal template, we use the Julia (Bezanson et al. 2017) package `VIDA.jl`.[8]

For this paper, we follow the procedure in Tiede et al. (2022a) and use the `CosineRing`$\{N, M\}$ template. The `CosineRing`$\{N, M\}$ filter uses an elliptical Gaussian ring, whose azimuthal thickness and brightness profile are described by a cosine expansion of order N and M, respectively. We parameterize the template as follows:

1. $d = 2\sqrt{ab}$ gives the diameter of the elliptical ring, where a and b are the semimajor and semiminor axis lengths of the ellipse.
2. $\tau = 1 - b/a$ is the ellipticity of the ring.
3. ξ_τ, is the position angle of the ellipticity measure east of north.
4. (x_0, y_0) is the ring center.
5. The thickness $\sigma(\phi)$ cosine expansion:

$$\sigma(\phi) = \sigma_0 + \sum_{n=1}^{N} \sigma_n \cos\left[n(\phi - \xi_n^{(\sigma)})\right]. \tag{9}$$

6. The brightness $s(\phi)$ cosine expansion:

$$s(\phi) = 1 - \sum_{m=1}^{M} s_m \cos\left[m(\phi - \xi_n^{(s)})\right]. \tag{10}$$

We have empirically found that the image reconstructions of ring-type images are well described using $N = 1$ and $M = 4$, which will be the template used for the rest of the paper.

The image reconstructions also tend to include a low level of emission distributed throughout the image. We also included an intensity floor template added pointwise to the `CosineRing` template to model the diffuse flux. For our divergence, we use the Bhattacharyya (Bh) divergence:

$$\mathrm{Bh}(t_\theta || I) = \sum_{ij} \sqrt{t_{\theta,ij} \hat{I}_{ij}}\,, \tag{11}$$

where we have assumed that the template and image intensity have unit flux, i.e., they sum to unity. The same strategy described in Tiede et al. (2022b) is used to optimize Equation (11).

3. Data

The EHT is a very-long-baseline interferometer. By the van Cittert–Zernike theorem (Thompson et al. 2017), a perfect interferometer measures the Fourier transform of the on-sky image:

$$\tilde{I}(u, v) = \int I(\alpha, \beta) e^{2\pi i(u\alpha + v\beta)} d\alpha d\beta, \tag{12}$$

where $I(\alpha, \beta)$ is the on-sky image intensity. However, in reality, the measurements are corrupted by telescope and scan specific complex gain terms $g_i = \gamma_i e^{i\theta_i}$ giving

$$\tilde{V}_{ij} = g_i g_j^* \tilde{I}_{ij}, \tag{13}$$

where $\tilde{V}_{ij}$ is known as the complex visibilities. This is the primary data product we will consider in this chapter. The benefit of using complex visibilities is that the observation likelihood of visibilities is just a complex Gaussian (Thompson et al. 2017):

$$\mathcal{L}(V|\theta) = \prod_i (2\pi\sigma_i^2)^{-1} \exp\left(-\frac{|V_i - g_{1,i} g_{2,i}^* \hat{V}_\theta(u_i, v_i)|^2}{2\sigma_i^2}\right), \tag{14}$$

where $\hat{V}$ is the model visibility at u_i and v_i and θ are the model parameters. While closure products are immune to gains, their statistical properties are more complicated. Namely, the likelihoods for closure phases and amplitudes are markedly non-Gaussian at a low signal-to-noise-ratio (SNR; Thompson et al. 2017; Blackburn et al. 2020; Broderick et al. 2020b) and are correlated. As a result, low SNR data are usually flagged, potentially biasing results.

The downside of fitting complex visibilities is that the complex gains must included in the forward model. Unfortunately, modeling gains introduce many additional parameters ($\sim$250 for the M87* observations), potentially dramatically increasing the computational cost of imaging. To combat this, we will use THEMIS' complex gain marginalization scheme. This scheme uses a Laplace approximation to approximate the marginalization at each Markov Chain Monte Carlo (MCMC) step, meaning we do not directly sample the gains. Effectively, this procedure acts as an averaged self-calibration step for each proposed image during sampling.

We follow the data processing choices used in Paper VI. We will use EHT data that has been coherently averaged over scans. Using scan-averaged data, we effectively assume that the gains are constant over each scan, which is approximately true for M87* Paper III. An additional 1% fractional error to the reported thermal noise was used to model residual calibration systematics based on estimates from Paper III:

$$\sigma_{ij}^2 \to \sigma_{ij}^2 + (0.01 V_{ij})^2. \tag{15}$$

To create simulated data for THEMIS, we use the `eht-imaging` and THEMISPY packages following the same procedures as Paper VI. Namely, we add an overresolved large-scale component for all simulated data sets considered in this paper to bring the zero-baseline flux to 1.2 Jy. Complex station gains are included to model the atmosphere and measurement systematics using an identical procedure to that described in Paper IV. Finally, in 2017, the EHT recorded its data at two frequency bands commonly denoted by LO (227.1 GHz) and HI (229.1 GHz) bands. In Section 4, we only simulate data in the LO band, while in Section 5, we consider both the LO band and HI band data to improve the image reconstructions.

[8] https://github.com/ptiede/VIDA.jl

4. Geometric Tests

In principle, BI should be less susceptible to the ellipticity bias in the `eht-imaging` top set in Tiede et al. (2022a). We will consider the same ellipticity validation from Tiede et al. (2022a) to test the faithfulness of our ellipticity estimates. Namely, we consider an elliptical Gaussian ring with ellipticity $\tau = 0.187$ at various orientations. Given the timescale of the BI scheme used in this paper, we only consider the orientations $\xi_\tau = -45°$, 0°, 45°, 90°, since the remaining cases flip the location of the brightness maximum.

To validate BI, we split the analysis into two steps. First, in Section 4.1, we approximately find the optimal number of pixels given the observed data. The image model described in Section 2 has two discrete hyperparameters that need to be specified, namely, the number of pixels in the two principal directions of the raster grid. We use the Bayesian evidence, or rather an approximation known as the Bayesian information criterion (BIC), which approximates the Bayesian evidence, to score the different numbers of pixels. Given the computational timescale of BI, we only attempt the pixel parameter survey on the elliptical ring with an orientation of 90°. We choose this orientation because the `eht-imaging` top set had the largest biases (Tiede et al. 2022a). Second, in Section 5.1, we use the optimal raster dimension for the $\xi_\tau = 90°$ case and then image the other orientations, i.e., $\xi_\tau = \pm 45°$ and 0°. The reconstructions for the different orientations are then analyzed, and we assess whether the proposed approach is able to recover the true ring shape and orientation.

4.1. Pixel Optimization

We considered 6×4, 5×5, 6×6, and 8×8 rasters in our survey to find the optimal raster dimension. These raster dimensions were inspired by Broderick et al. (2020a), who found that the optimal number of pixels was a 5×5 raster for GRMHD simulations using the uv coverage on April 11. Therefore, using the 5×5 raster as our starting point, we also considered two higher resolutions to test whether increased resolution drastically changes the evidence and image structure. We also considered the 6×4 resolution since this roughly matches the true ellipticity of the ring.

We used the Stan NUTS sampler (Carpenter et al. 2017) with a diagonal mass matrix and 2000 adaptation steps for the local posterior exploration. We used the DEO parallel tempering sampler from Syed et al. (2019) with 60 tempering levels to enable global posterior exploration. The DEO sampler was run for 9–10 adaptation rounds, where we used the Stan adaptation scheme at the beginning of each round. We found an optimal global communication barrier $\Lambda \approx 15$–16, implying an optimal round-trip rate of 0.032. After nine rounds, the ladder had a communication barrier of $E \approx 20$–21, which gives a round-trip rate of 0.023. All chains were run until the split-$\hat{R} < 1.1$ for all image raster c_{ij} parameters.

The mean images from each run are shown in the second row of Figure 1. For the 6×4, 5×5, and 6×6 rasters, the posterior was multimodal, usually displaying 3–4 distinct image clusters.

To move from the posterior of c_{ij} to image features, we randomly selected 2000 images from the last tempering round. These images were then run through `VIDA` using the templates described in Section 2. Each raster dimension's resulting ellipticity and orientation are shown in Figure 1. In the rightmost column, we show the results from Tiede et al. (2022a) using the same `eht-imaging` top set. The true ellipticity and orientation are within the 95% credible regions regardless of the chosen raster dimension. This contrasts the `eht-imaging` top set, which, as detailed in Tiede et al. (2022a), does not contain a single reconstruction consistent with the on-sky image.

The impact of adding more pixels/model freedom can be seen moving from left to right. As the number of image pixels increases, the posterior becomes broader. This is expected since the model has more degrees of freedom to describe the image structure. This pattern could be continued further. Namely, we could keep adding more and more pixels. As we do, the posteriors would become broader until we effectively sample from the prior. However, adding more pixels would eventually shrink the Bayesian evidence, signaling our model is too complicated. Therefore, as described in Section 2, we use the BIC to choose the optimal raster dimension.

The BIC for the different raster dimensions is shown in Table 1. The BIC and reduced chi-square are the smallest for the 5×5 and 6×4 models. These results are consistent with the results from Broderick et al. (2020a). Namely, Broderick et al. (2020a) found that the BIC favored the 5×5 grid for simulated data from GRMHD models given in Event Horizon Telescope Collaboration et al. (2019e). For this paper, we decided to use the 5×5 raster as our optimal model, even though the 6×4 model has a lower BIC. The reason for choosing the 5×5 model is due to its similarity to the results in Broderick et al. (2020a), and it does not presume an anisotropic source structure.

5. M87* Imaging

5.1. Orientation Dependence of Bayesian Imaging

Given that 5×5 optimizes the BIC, we now repeat our analysis on the simulated elliptical ring data with $\xi_\tau = 45°$, 0°, −45°. We used the same sampler settings as above and ran for a similar number of MCMC steps. We again randomly selected 2000 images from the chain of the last tempering round to construct the feature posteriors and used `VIDA` with the template described in Section 2.2. The mean images for THEMIS and `eht-imaging` are shown in Figure 2.

Visually from Figure 2, we show that THEMIS' mean image appears more similar to the ground truth image. Namely, it does not appear to have the same north–south bias as the `eht-imaging` top set. Quantitatively, we find that THEMIS's ellipticity posterior contains the truth in its 95% contours for all the orientations considered. Interestingly, we find that the $\xi_\tau = 90$ orientation THEMIS reported substantially larger uncertainty and has portions with values similar to the `eht-imaging` top set. This implies that the M87* (u, v) coverage poorly contains the image shape in this orientation. Critically, BI incorporates this uncertainty into its ellipticity estimates, and the true value is contained in the bulk of the posterior mass.

The previous section demonstrated that BI can recover the true ellipticity. This result demonstrates that we can directly interpret the measured ellipticity as representing the ellipticity of the on-sky image. Therefore, in this section, we will apply the BI and `VIDA` pipeline to the 2017 EHT M87* data.

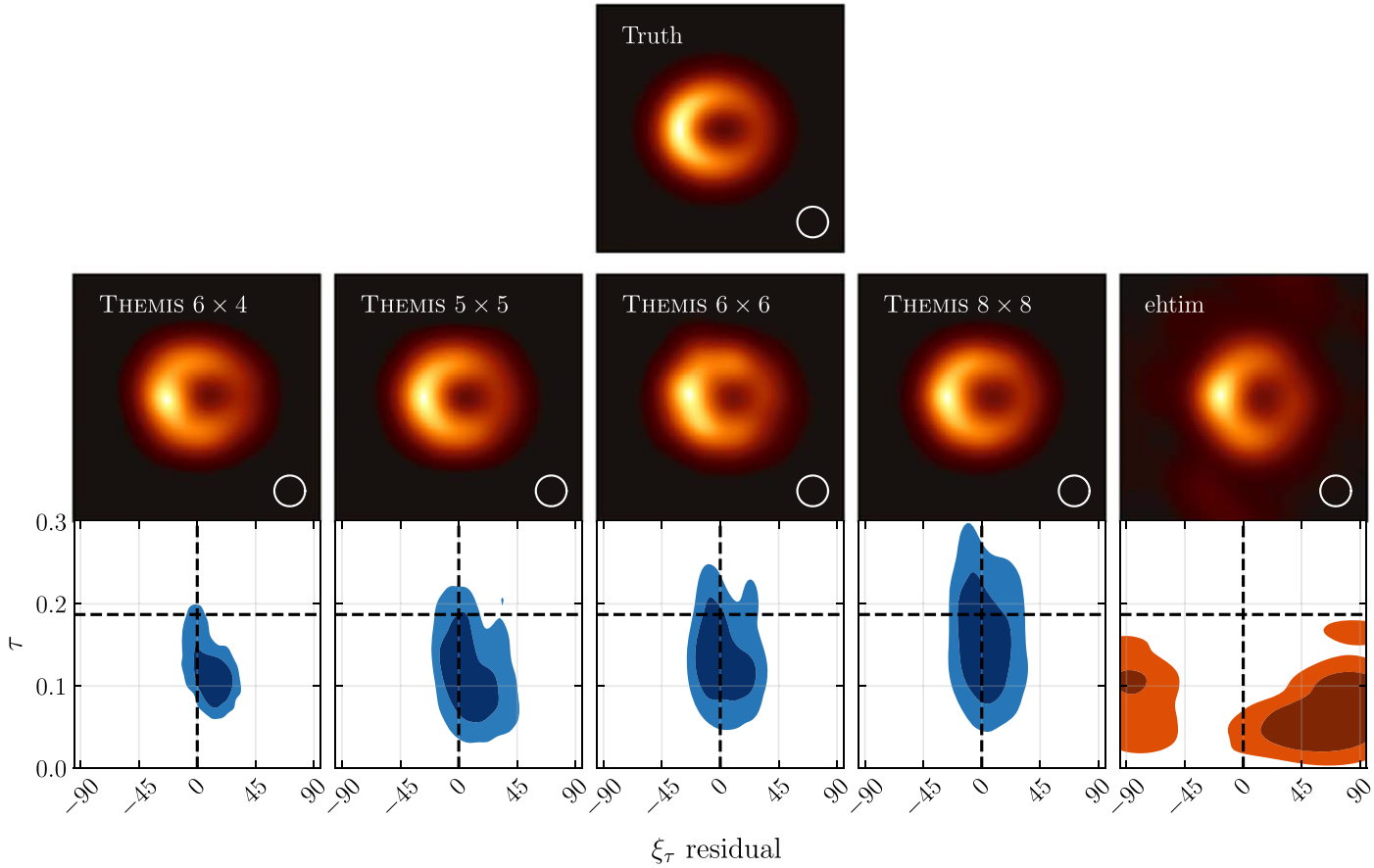

Figure 1. Comparison of the mean image reconstructions (middle row) and ellipticity posterior (bottom row) compared to the sky-truth image (top row). The probability contours show the 68% and 95% probability regions. All images have been blurred with a circular Gaussian with a 15 μas FWHM, but feature extraction is done on the unblurred images. The first row columns are the results using the different rasters described in Section 4.1. The last columns show the results from the `eht-imaging` top set from Paper IV. Overall, we find the ellipticity is within the 95% probability regions for all raster dimensions considered. This is the opposite of the `eht-imaging` top set, which failed to produce a single image representing the true ellipticity and orientation.

Table 1
Elliptical Ring Raster Survey Results

	8 × 8	6 × 6	5 × 5	6 × 4
ΔBIC	240.6	68.9	5.2	0.0
$\chi^2_{\rm red}$	1.49	1.28	1.23	1.22

5.2. Imaging Pipeline for M87*

We consider two data sets to analyze the 2017 M87* EHT data. First, we analyze the LO band-only data to compare the results with those in Tiede et al. (2022a). Second, we combine HI and LO bands and fit the total HI+LO band data. We expect that the combined data will increase the constraining power since we are effectively doubling the amount of data. Additionally, splitting the data into two sets effectively performs a jackknife test as it allows us to assess which features are robust. Namely, if our model adequately represents the on-sky image, we expect consistent results between the two data sets.

For both data sets, we will use the same image model. Following the parameter exploration in Section 4.1, we consider a 5 × 5 raster with an adaptive FOV and orientation. Note that for the combined HI+LO analysis, we do not assume that the gains are the same across bands but do assume the image is identical.

For the LO band data, we use similar sampler settings for the elliptical ring simulated data. Namely, 60 tempering levels and 2000 adaptation steps at the beginning of each tempering round. We found that the posterior was multimodal on April 5, 10, and 11, while on April 6, we only found a single mode.

We apply k-means clustering on the last 50% of the chain from the last tempering round to separate the image modes. Each cluster's mean and the relative number of samples in each cluster are shown in the left panels of Figure 3 for the April 11 LO band data. The three image models have a qualitatively similar structure. Namely, each mode is dominated by a ring-like feature whose brightness peak is in the south. The nonring emission, however, differs across the three modes. The dominant mode has emission in the northeastern part of the image, and the two subdominant modes either lack extended emission or place it in the southwestern part.

For the combined HI and LO data, we used 320 tempering levels. Each chain was run for >50,000 MCMC steps and achieved a split-$\hat{R} \leqslant 1.03$. Interestingly, the additional data from fitting the combined bands eliminated the multimodality on April 5 and 11. This is shown in Figure 3 for April 11. The LO band-only reconstructions have three distinct image modes; however, the other two lower probability modes disappear once we include the HI band data. On April 10, we found the same three modes for the HI+LO band as the LO band-only fits. To create the feature posteriors, we created 2000 images from the last tempering round's chain and then applied `VIDA` using the same template from Section 5.1.

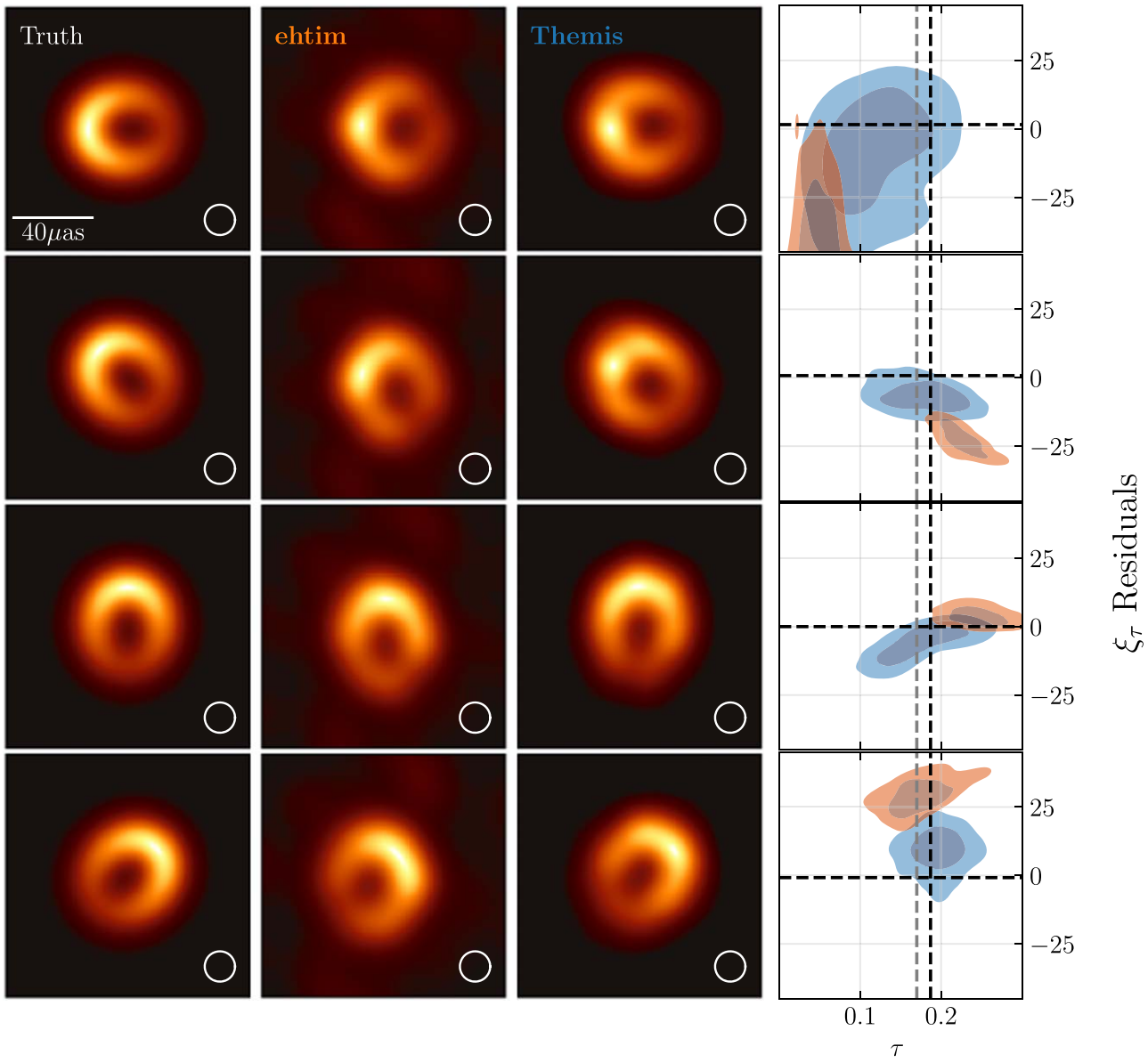

Figure 2. Comparison between `eht-imaging` and THEMIS' BI reconstructions of the elliptical ring test from Tiede et al. (2022a). The left column shows the on-sky-truth image. The second column shows the mean image from the `eht-imaging` top set. The third column shows the mean image from the THEMIS posterior using a 5×5 raster. All images have been blurred by a Gaussian with a 15 μas FWHM. The rightmost column shows the joint distribution for the recovered ellipticity τ and its position angle ξ_τ for `eht-imaging` (orange) and THEMIS (blue). The contours display the 68% and 95% probability regions. The black dashed line shows the on-sky-truth values, and the gray dashed line shows the truth values after being blurred by a Gaussian with 15 μas FWHM.

5.3. Ellipticity Results

The resulting ellipticity and orientation joint posteriors for THEMIS are shown in Figure 4. Overall, we find that the combined HI+LO band ellipticity posterior is contained in the LO band posterior. Therefore, our single-band and multiband fits are consistent. Given that the combined band fits have less uncertainty in the remaining discussion, we will focus on the HI+LO band fits exclusively.

In Figure 5, we compare the `eht-imaging` and THEMIS results across each observation day. The top two rows show the mean THEMIS and `eht-imaging` images blurred by a Gaussian with 15 μas FWHM. The bottom row shows the joint ellipticity and ellipticity position angle posterior for THEMIS compared to the join distribution estimated from the `eht-imaging` top set. The THEMIS and `eht-imaging` measured ellipticity are discrepant at over 2σ on April 5, 6, and 11. On April 10, the ellipticity and position angles between `eht-imaging` and THEMIS do overlap. However, April 10 has substantially less data than the other days. As a result, the relative error bars are much larger, likely aiding in the consistency.

Focusing on the THEMIS results, we see that on April 5 and 6, the resulting ellipticity distributions are consistent at a 2σ level, while April 10 and 11 are consistent. Quantitatively, we find that the measured ellipticity from THEMIS on April 5 is $\tau = 0.09^{+0.02}_{-0.03}$, April 6 is $\tau = 0.14^{+0.03}_{-0.02}$, April 10 is $\tau = 0.08^{+0.14}_{-0.05}$, and April 11 is $\tau = 0.05^{+0.02}_{-0.02}$, where the main value is the median and the error bars are the 95% credible intervals around the median. To create a conservative estimate of the average ellipticity of M87*, we combine ellipticity samples for each day weighted inversely by the standard deviation of the posterior. Computing the median and 95% credible interval on the combined samples, we find that the ring ellipticity of M87* is $0.09^{+0.07}_{-0.06}$.

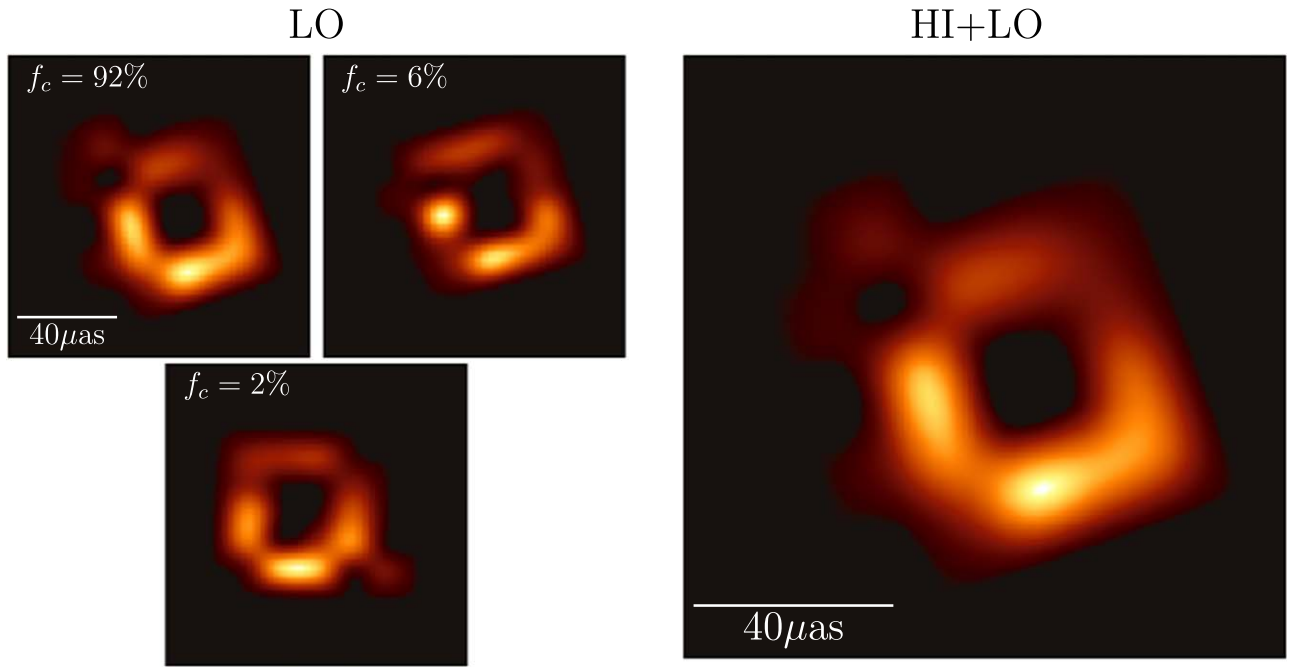

Figure 3. Example image reconstructions of M87* LO band on April 11 (left three plots) and HI+LO band (right). For the LO band, we show three images corresponding to the three image clusters found in the MCMC chains. In the top-left corner of the LO band plots, we show the percentage of images in each cluster taken from the last 50% of the MCMC chain. For the HI+LO band, we find that the extra data removes the two lower percentage modes.

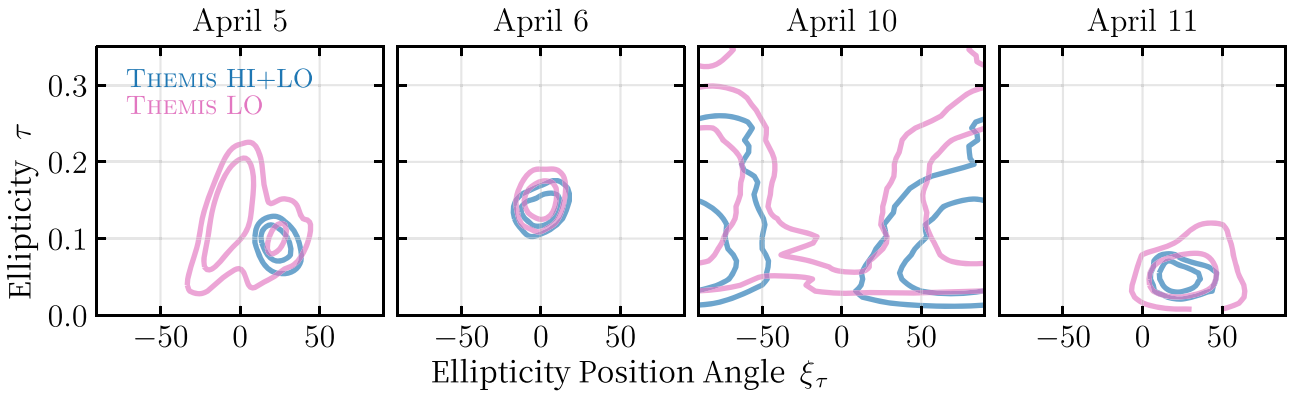

Figure 4. Joint ellipticity value and orientation from M87* on April 5, 6, 10, and 11. The contours show the 68% and 95% probability regions from THEMIS using the combined HI+LO band data (blue) and LO band only (pink). The LO band and HI+LO ellipticity results are consistent between bands.

In Tiede et al. (2022a), the authors reported an upper limit on the ring ellipticity of $\tau \lesssim 0.3$, assuming the image of M87* is approximately described by a GRMHD simulation. The origin of the upper limit was due to requiring GRMHD calibration to account for biased ring ellipticity measurements from the `eht-imaging` imaging pipeline in Paper IV. The imaging pipeline in this paper does not require the same calibration pipeline. As a result of the ellipticity measurement being entirely data driven, we are able to improve on the results compared to Tiede et al. (2022a).

5.4. Physical Interpretation of Ellipticity

A comparison to expectations from theoretical simulations is required to provide a physical interpretation of the measured ellipticity. The source of ellipticity in the reconstructed image can come from spacetime effects such as the photon ring (Johannsen et al. 2016; Medeiros et al. 2018), inner shadow (Chael et al. 2021), and the accretion flow (Tiede et al. 2022a). We will use the scaled set of GRMHD simulations from Tiede et al. (2022a) to make this comparison. The scaled set of GRMHD simulations consists of 100 random GRMHD snapshots rescaled to the best-fit mass from Paper VI, whose image size is further stretched by a factor of 0.8, 0.9, 1.0, 1.1, and 1.2 isotropically to approximate the uncertainty in the mass measurement. The result is 500 GRMHD simulations.

To compare the simulations to the measured M87* asymmetry, one approach would be to first apply the same synthetic data and imaging pipeline from Section 5.2 for each of the 500 GRMHD simulations and construct the empirical GRMHD ellipticity distribution. The computation timescale of the BI pipeline used in this paper (∼5–6 days for a single source using hundreds of CPUs) makes this approach prohibitively expensive. To approximate the true GRMHD ellipticity distribution from the BI pipeline, we will instead blur the GRMHD simulations by 15 μas and run `VIDA` on the blurred images to approximate the ellipticity we expect to see from the image reconstructions. If the M87* inferred ellipticity distribution lives within the values from the blurred GRMHD,

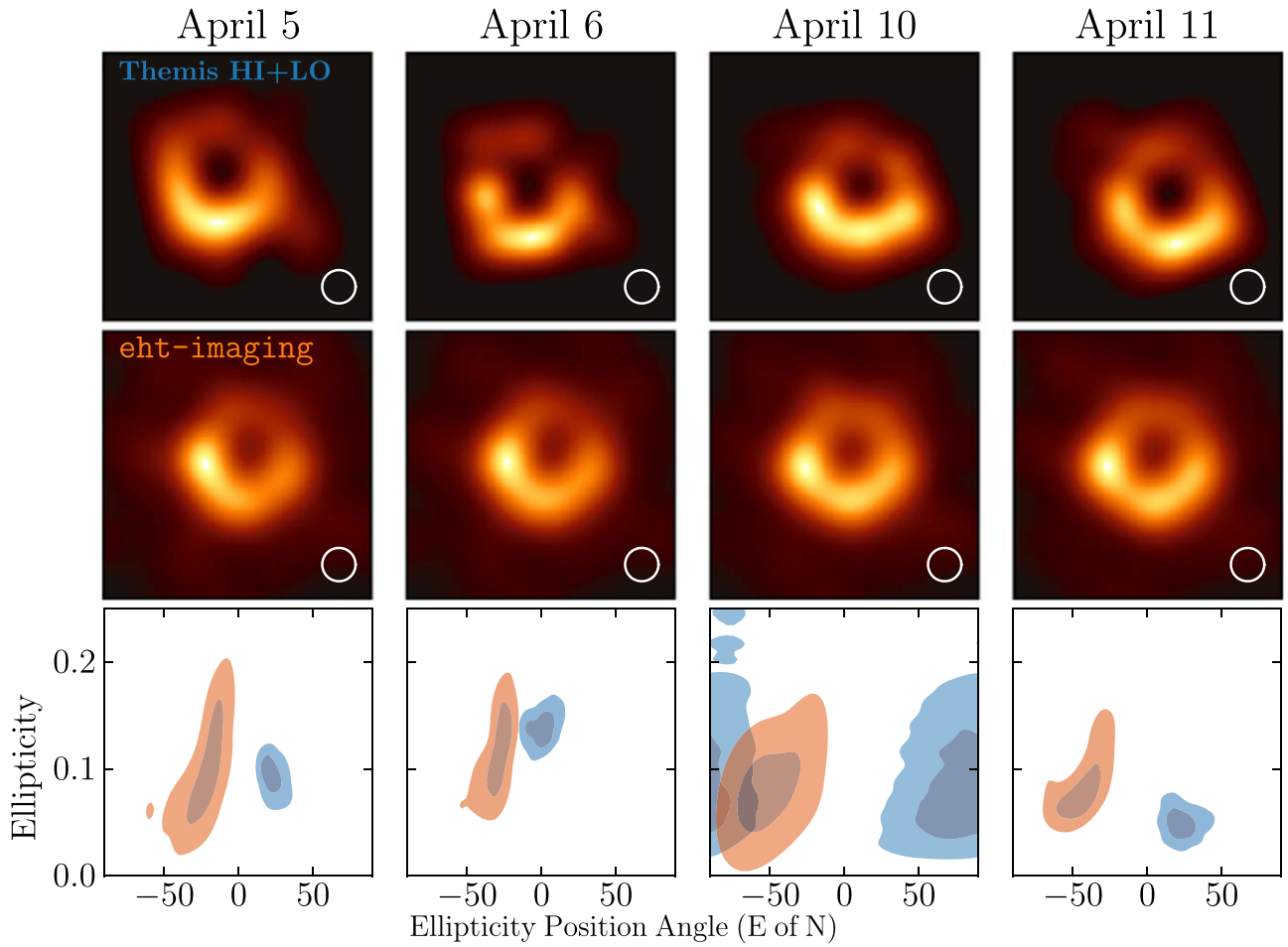

Figure 5. Mean images from the THEMIS BI posterior (top) and `eht-imaging` top set (middle) both blurred by a 15 μas FWHM Gaussian. The bottom row shows the joint distribution of the image ellipticity vs. position angle for the THEMIS results (blue) and `eht-imaging` results (orange). The contours shown are 68% and 95% credible regions.

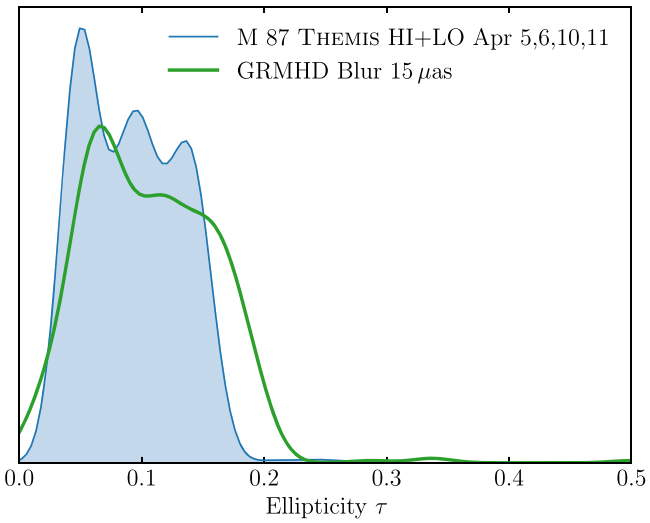

Figure 6. Comparison of recovered ellipticity of M87* using BI (blue), and the expected ellipticity from GRMHD simulations blurred with a 15 μas Gaussian kernel (green line). Both the GRMHD and BI asymmetries were found using `VIDA`. We find that the recovered M87* ellipticity is within the expected GRMHD ellipticity distribution and thus is consistent with theoretical expectations.

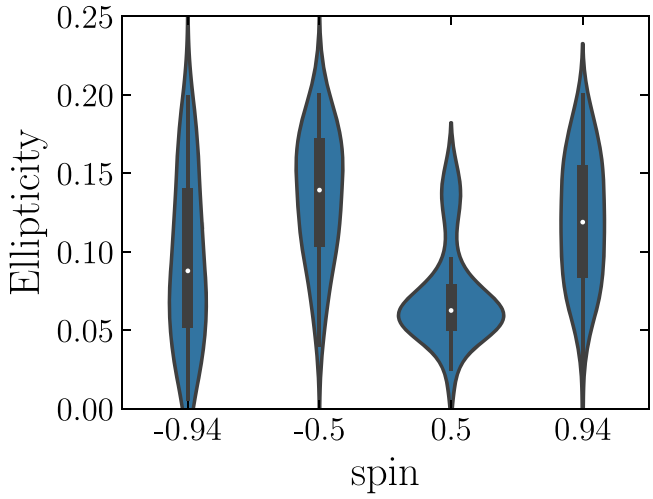

Figure 7. Recovered ellipticity τ for each image in the GRMHD set as a function of the black hole spin. The thick/thin lines show the 68% and 95% quantile ranges for the ellipticity posteriors, and the white dot is the median.

we can conclude that M87*'s shape is consistent with expectations from the GRMHD simulations. Figure 6 shows the inferred distribution of ring ellipticity from the GRMHD simulations (green) versus the stacked posterior[9] across M87* on April 5, 6, and 11. We can see that the ellipticity observed

[9] We combine the samples from the three posteriors as a conservative estimate of the ellipticity uncertainty.

from M87* is consistent with expectations from the GRMHD simulations.

Of particular interest is whether the recovered ellipticity is related to the underlying spacetime or, rather, is more measuring ellipticity due to the accreting plasma. To assess the ellipticity's origin, Figure 7 shows each snapshot image's measured ellipticity as a function of spin. Comparing the measured ellipticity to the GRMHD simulations as a function of spin, we see that the effect of spin is minor on the overall value of ellipticity. Namely, the measured ring shape is driven not by the underlying spacetime but rather by the properties of the emitting plasma. The lack of spin dependence results from the finite resolution of the EHT array, where any ellipticity due to the black hole shadow or photon ring (expected to be 1% for M87*; Farah et al. 2020) is washed out by the background accretion flow. Therefore, we conclude that the shape of M87*'s shadow is consistent with black hole spin.

6. Summary

This paper presents the first-ever Bayesian estimates of the ring ellipticity of M87*'s shadow. We found that M87*'s averaged ellipticity is $0.09^{+0.07}_{-0.06}$. Compared to the `eht-imaging` results from Paper IV and Tiede et al. (2022a), the Bayesian approach has provided a weak measurement of the actual ellipticity of M87* rather than an upper limit.

Additionally, the ellipticity measurement in this paper is less susceptible to incomplete Fourier coverage and instrumental effects than the `eht-imaging` top set approach in Paper IV. To demonstrate this improvement, we considered four simulated data tests based on realistic simulated data to validate our ellipticity measurement. The BI approach recovered the correct ellipticity for all simulated data sets, while the RML approach failed to recover the truth in every case. Furthermore, by using the BIC to score our different image reconstructions, we are able to select the optimal hyperparameters given our data, providing a data-driven, hyperparameter-optimized imaging algorithm.

Comparing the measured ellipticity of M87* with GRMHD, we assessed whether M87* is consistent with theoretical expectations. We found that M87* is consistent with GRMHD simulations, and the measured ellipticity is dominated by the accretion flow and not the spin of the black hole. To disentangle the ellipticity due to gravitational effects from the accretion flow will require new observations. Future EHT observations will be able to provide more stringent measurements of M87* as more telescopes are included in the array, such as GLT (added in 2018), Kitt Peak, and NOEMA (added in 2021; Event Horizon Telescope Collaboration et al. 2019b). These new sites will dramatically lower the ellipticity uncertainty and provide more stringent comparisons with theoretical simulations. However, conclusively identifying the gravitationally induced ellipticity in M87*'s shadow will likely require a space-based radio dish or very-long-baseline interferometry (VLBI) arrays, such as the recently proposed black hole Explorer (Kurczynski et al. 2022) or THEZA project (Gurvits et al. 2021). Space-based VLBI arrays will provide the necessary resolution to resolve the properties of the $n = 1$ photon ring, providing a more stringent measurement of the properties of spacetime around supermassive black holes.

Acknowledgments

P.T. receives support from the Natural Science and Engineering Research Council through the Alexander Graham Bell CGS-D scholarship. This work was supported by the Black Hole Initiative, which is funded by grants from the John Templeton Foundation and the Gordon and Betty Moore Foundation (although the opinions expressed in this work are those of the author(s) and do not necessarily reflect the views of these Foundations). P.T. was also supported by the Gordon and Betty Moore Foundation grants GBMF-3561, GBMF-5278, GBMF-10423. P.T. was also supported by National Science Foundation grants AST 19-35980 and AST 20-34306. The Perimeter Institute for Theoretical Physics partially supported this work. Funding for research at the institute is provided by the Department of Innovation, Science and Economic Development Canada and the Ministry of Economic Development, Job Creation and Trade of Ontario, both of which are branches of the Government of Canada. A.E.B. expresses gratitude to the Delaney family for their generous financial backing through the Delaney Family John A. Wheeler Chair at Perimeter Institute. Additionally, A.E.B. receives further financial support for this research through a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

Software: BlackBoxOptim.jl, `eht-imaging`, Julia (Bezanson et al. 2017), matplotlib 3.3 (Hunter 2007), Pandas (The pandas development team 2020; McKinney 2010), Python 3.8.3 (Van Rossum & Drake 2009), SciPy (Virtanen et al. 2020), ThemisPy,[10] VIDA.jl (Tiede et al. 2022b) P. Tiede et al. (2024, in preparation), THEMIS Broderick et al. (2020a).

Appendix
M87 2017 Feature Posteriors

Figure 8 shows the marginal feature posteriors for the THEMIS results and the top set distribution from the 2017 `eht-imaging` results. The THEMIS results for the diameter are stable across days and frequency bands, although the measured diameter is slightly larger than the `eht-imaging` results for each day. The brightness asymmetry for THEMIS is consistent on April 5, 6, and 10, while April 11 may be marginally more azimuthally symmetric. On all days, THEMIS' measured brightness asymmetry is larger than the measured value from `eht-imaging`. Finally, for the brightness position angle (PA), THEMIS does not find significant evolution, remaining consistent with an angle of 170° for all days. On the other hand, `eht-imaging` sees modest evolution, with the PA increasing from April 5/6 to April 10/11.

[10] https://github.com/aeb/ThemisPy

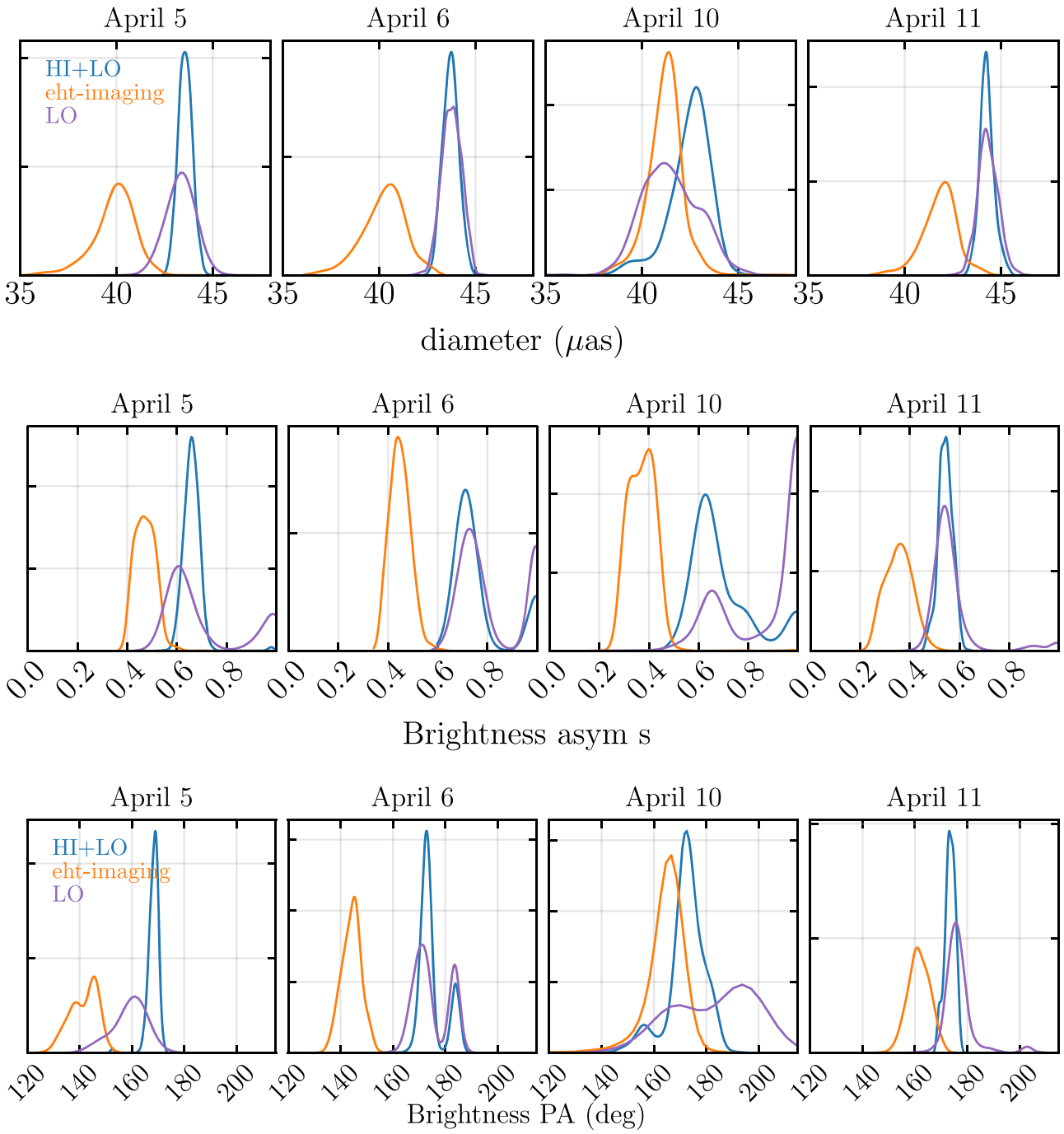

Figure 8. Marginal feature distributions for the diameter (top row), brightness asymmetry (middle row), and brightness PA (bottom row). The THEMIS distribution for the LO band only (purple) and HI+LO band (blue) fits are the marginal posteriors estimated using `VIDA`. The `eht-imaging` distributions are estimated by applying `VIDA` to the top set described in Paper IV and do not form a posterior.

ORCID iDs

Paul Tiede https://orcid.org/0000-0003-3826-5648
Avery E. Broderick https://orcid.org/0000-0002-3351-760X

References

Arras, P., Frank, P., Haim, P., et al. 2022, NatAs, 6, 259
Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. 2017, SIAMR, 59, 65
Blackburn, L., Pesce, D. W., Johnson, M. D., et al. 2020, ApJ, 894, 31
Broderick, A. E., Gold, R., Karami, M., et al. 2020b, ApJ, 897, 139
Broderick, A. E., Johannsen, T., Loeb, A., & Psaltis, D. 2014, ApJ, 784, 7
Broderick, A. E., Pesce, D. W., Gold, R., et al. 2022, ApJ, 935, 61
Broderick, A. E., Pesce, D. W., Tiede, P., Pu, H.-Y., & Gold, R. 2020a, ApJ, 898, 9
Carpenter, B., Gelman, A., Hoffman, M. D., et al. 2017, Journal of Statistical Software, 76, 1
Chael, A., Johnson, M. D., & Lupsasca, A. 2021, ApJ, 918, 6
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019a, ApJL, 875, L1
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019b, ApJL, 875, L2
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019c, ApJL, 875, L3
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019d, ApJL, 875, L4
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019e, ApJL, 875, L5

Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019f, ApJL, 875, L6
Farah, J. R., Pesce, D. W., Johnson, M. D., & Blackburn, L. 2020, ApJ, 900, 77
Gebhardt, K., Adams, J., Richstone, D., et al. 2011, ApJ, 729, 119
Gralla, S. E., & Lupsasca, A. 2020, PhRvD, 102, 124003
Gralla, S. E., Lupsasca, A., & Marrone, D. P. 2020, PhRvD, 102, 124004
Gurvits, L. I., Paragi, Z., Casasola, V., et al. 2021, ExA, 51, 559
Hunter, J. D. 2007, CSE, 9, 90
Johannsen, T., & Psaltis, D. 2010, ApJ, 718, 446
Johannsen, T., Wang, C., Broderick, A. E., et al. 2016, PhRvL, 117, 091101
Johnson, M. D., Lupsasca, A., Strominger, A., et al. 2020, SciA, 6, eaaz1310
Kurczynski, P., Johnson, M. D., Doeleman, S. S., et al. 2022, Proc. SPIE, 12180, 121800M
Liepold, E. R., Ma, C.-P., & Walsh, J. L. 2023, ApJL, 945, L35
McKinney, W. 2010, in Proc. 9th Python in Science Conf. 445, ed. S. van der Walt & J. Millman, 51
Medeiros, L., Chan, C.-k., Özel, F., et al. 2018, ApJ, 856, 163
Medeiros, L., Psaltis, D., & Özel, F. 2020, ApJ, 896, 7
Osorno, J., Nagar, N., Richtler, T., et al. 2023, A&A, 679, A37
Simon, D. A., Cappellari, M., & Hartke, J. 2023, MNRAS, 527, 2341
Sun, H., & Bouman, K. L. 2021, in Proc. of the AAAI Conf. on Artificial Intelligence 35, 2628
Syed, S., Bouchard-Côté, A., Deligiannidis, G., & Doucet, A. 2021, Journal of the Royal Statistical Society Series B: Statistical Methodology, 84, 321
Takahashi, R. 2004, ApJ, 611, 996
The pandas development team, 2020 pandas-dev/pandas: Pandas v2.2.2, Zenodo
Thompson, A. R., Moran, J. M., & Swenson, G. W., Jr. 2017, Interferometry and Synthesis in Radio Astronomy (3rd ed.; Berlin: Springer International Publishing)
Tiede, P. 2022, JOSS, 7, 4457
Tiede, P., Broderick, A. E., & Palumbo, D. C. M. 2022b, ApJ, 925, 122
Tiede, P., Broderick, A. E., Palumbo, D. C. M., & Chael, A. 2022a, ApJ, 940, 182
Van Rossum, G., & Drake, F. L. 2009, Python 3 Reference Manual (Scotts Valley, CA: CreateSpace)
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, NatMe, 17, 261