

Hierarchical Interferometric Bayesian Imaging

Paul Tiede^{1,2} , William Moses³ , Valentin Churavy^{4,5} , Michael D. Johnson^{2,1} , Dominic W. Pesce^{2,1} ,
Lindy Blackburn^{2,1} , and Peter Galison^{1,6,7} 

¹ Black Hole Initiative at Harvard University, 20 Garden Street, Cambridge, MA 02138, USA

² Center for Astrophysics|Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA

³ Siebel School of Computing and Data Science, University of Illinois Urbana-Champaign, IL, USA

⁴ Institute of Mathematics, Johannes Gutenberg University Mainz, Staudingerweg 9, 55128 Mainz, Germany

⁵ High-Performance Scientific Computing, Center for Advanced Analytics and Predictive Sciences, University of Augsburg, Universitätsstraße 12a, 86159 Augsburg, Germany

⁶ Department of Physics, Harvard University, Cambridge, MA 02138, USA

⁷ Department of History of Science, Harvard University, Cambridge, MA 02138, USA

Received 2025 October 21; revised 2025 November 30; accepted 2025 December 1; published 2026 January 28

Abstract

Very long baseline interferometry (VLBI) achieves the highest angular resolution in astronomy. VLBI measures corrupted Fourier components, known as visibilities. Reconstructing on-sky images from these visibilities is a challenging inverse problem, particularly for sparse arrays such as the Event Horizon Telescope (EHT) and the Very Long Baseline Array, where incomplete sampling and severe calibration errors introduce significant uncertainty in the image. To help guide convergence and control the uncertainty in image reconstructions, regularization on the space of images is utilized, such as enforcing smoothness or similarity to a fiducial image. Coupled with this regularization is the introduction of a new set of parameters that modulate its strength. We present a hierarchical Bayesian imaging approach (hierarchical interferometric Bayesian Imaging, HIBI) that enables the quantification of uncertainty for all parameters. Incorporating instrumental effects within HIBI is straightforward, allowing for simultaneous imaging and calibration of data. To showcase HIBI’s effectiveness and flexibility, we build a simple imaging model based on Markov random fields and demonstrate how different physical components can be included, e.g., black hole shadow size, and their uncertainties can be inferred. For example, while the original EHT publications were unable to constrain the ring width of M87*, HIBI measures a width of $9.3 \pm 1.3 \mu\text{as}$. We apply HIBI to image and calibrate EHT synthetic data, real EHT observations of M87*, and multifrequency observations of OJ 287. Across these tests, HIBI accurately recovers a wide variety of image structures and quantifies their uncertainties. HIBI is publicly available in the Comrade VLBI software repository.

Unified Astronomy Thesaurus concepts: [Very long baseline interferometry \(1769\)](#); [Black holes \(162\)](#); [Active galactic nuclei \(16\)](#); [Astronomy data analysis \(1858\)](#); [Aperture synthesis \(53\)](#); [Bayesian statistics \(1900\)](#)

1. Introduction

Very long baseline interferometry (VLBI) produces the highest-resolution images among all astronomical techniques, now achieving angular resolutions of $\sim 10 \mu\text{as}$ (e.g., N. S. Kardashev et al. 2013; Event Horizon Telescope Collaboration et al. 2019a). However, VLBI does not directly image the on-sky source. A perfect VLBI interferometer samples interferometric visibilities, which are related to the Fourier transform of the corresponding image (A. R. Thompson et al. 2017). For instance, visibilities $\tilde{I}(u, v)$ corresponding to the total intensity (Stokes I) image $I(x, y)$ are given by

$$\tilde{I}(u, v) = \int I(x, y) e^{2\pi i(ux+vy)} dx dy. \quad (1)$$

Given a pair of sites a, b , this implies that the baseline (a, b) measures a single Fourier component of the image, whose position in the Fourier space, (u, v) , is given by the projected distance in units of wavelength between the sites. We will denote this Fourier component by $\tilde{I}_{ab} = \tilde{I}(u_{ab}, v_{ab})$. The resolution of the interferometer is, therefore, roughly set by

 Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

the maximum distance between two sites,

$$\rho_{\text{Dif}} \approx 1/u_{\text{max}}. \quad (2)$$

VLBI imaging aims to invert the measurement process, removing the impact of instrumental corruption and incomplete Fourier sampling to reconstruct the true on-sky image. As a result, VLBI imaging attempts to rebuild the infinite degree of freedom on-sky intensity map from a finite set of observations.

The primary difficulty in VLBI imaging is that the problem is nonidentifiable. To define identifiability, consider that we have a family of distributions that depends on a set of parameters θ , which we denote p_θ . An example of this is the likelihood distribution $\mathcal{L}(\mathbf{D}|\theta)$, where \mathbf{D} is the hypothetical data we observe. The family of distributions $\theta \rightarrow p_\theta$ is identifiable if, for each set of parameters, the resulting likelihood distribution is unique. That is, the mapping from parameters to the likelihood distribution is one-to-one. If this mapping is not unique, the problem is stated to be nonidentifiable. For VLBI imaging, this degeneracy is typically of a form that there exists a smooth transformation of the parameters that leaves the likelihood invariant. As a result, this kind of nonidentifiability implies that repeating an observation infinitely many times will not result in a single image. Instead, the resulting space of images will converge to a lower-dimensional surface in the space of all

parameters. This kind of identifiability issue is common in VLBI imaging.

Consider an image $I(x, y)$, together with a second function $\delta(x, y)$ that is zero-mean and that has a characteristic scale of variation far below the resolution of the VLBI interferometer (Equation (2)). In this case, $I(x, y) + \delta(x, y)$ will produce an identical likelihood even though we have changed the image potentially everywhere. For posterior inference, this means that naive estimation of the image will result in a non-normalizable distribution. That is, there are infinitely many variations δ that will produce identical fits to the data. To limit the space of images, we must make additional assumptions about the source structure. For example, we could consider a simple parametric model, such as simple geometric shapes (Event Horizon Telescope Collaboration et al. 2019b, hereafter Paper VI), if we have a priori knowledge of what we are analyzing. However, for VLBI, this is often not warranted. Simplified geometric models typically underfit the data, since VLBI images describe complicated environments, e.g., turbulent jets, that are difficult to know a priori. The other, more common way to solve this problem in VLBI imaging is to focus on nonparametric models and then regularize the space of images to make it finite and at least weakly identifiable. In VLBI imaging, this is typically achieved through regularization of the space of images. If we focus on forward modeling of the source, this implies the use of Bayesian priors or similar regularizers within the regularized maximum likelihood (RML) methods that have recently become popular in VLBI imaging (R. Narayan & R. Nityananda 1986; A. A. Chael et al. 2018; K. Kuramochi et al. 2018; B. Zawadzki et al. 2023). The goal of these regularizers is to modify the resulting probability distribution such that the map is now identifiable, although this is often difficult to achieve in practice.

However, pure imaging is not the only source of nonidentifiability; when instrumental effects are considered, additional degeneracies occur. For a realistic interferometer, the observed visibilities are corrupted by both baseline-dependent thermal noise effects (ϵ_{ab}) and station-based complex gains (g_a), giving

$$\hat{V}_{ab} = g_a \tilde{I}_{ab} g_b^* + \epsilon_{ab}, \quad (3)$$

where ϵ_{ab} is drawn from a complex Gaussian with real and imaginary standard deviation, σ_{ab} . In this paper, we will parameterize g with a set of parameters that we will collectively refer to with s . Namely, we will typically decompose g into

$$g_a = \exp(\gamma_a + i\varphi_a), \quad (4)$$

where γ_a and φ_a are real numbers and are typically denoted by log-gain amplitudes and gain phases, respectively.

The VLBI likelihood for a single frequency and time is given by

$$\begin{aligned} \mathcal{L}(\hat{V}_{ab} | \mathbf{r}, \mathbf{s}_a, \mathbf{s}_b) \\ = (2\pi\sigma_{ab}^2)^{-1} \exp\left(-\frac{|\hat{V}_{ab} - g_a(\mathbf{s}_a) \tilde{I}_{ab} g_b^*(\mathbf{s}_b)|^2}{2\sigma_{ab}^2}\right). \end{aligned} \quad (5)$$

Assuming each time and frequency measurement is independent, the total likelihood is given by

$$\mathcal{L}(\hat{\mathbf{V}} | \mathbf{r}, \mathbf{s}) = \prod_{t,a,b} \mathcal{L}(\hat{V}_{t,ab} | \mathbf{r}, \mathbf{s}_{t,a}, \mathbf{s}_{t,b}), \quad (6)$$

where a, b denote the baselines and t the observation times.

For high-frequency VLBI, these gains are typically only known to within 10%–20% and can have order-unity uncertainties for some sites due to technical issues or poor weather. Therefore, these gains are estimated during imaging in a procedure typically denoted iterative self-calibration (P. N. Wilkinson et al. 1977; T. J. Cornwell & P. N. Wilkinson 1981). The need to jointly estimate $I(x, y)$ and g_a induces additional nonidentifiability issues. For example, a shift in the image centroid $I(x) \rightarrow I(x + c)$ can be absorbed in each telescope gain phase, meaning that the VLBI likelihood cannot constrain the absolute image position. Similarly, the change in the total flux of the image degenerates with the scaling of the gain amplitude of each telescope. Moreover, while these degeneracies are in principle easy to eliminate, more complicated degeneracies can occur, depending on the telescope performance and coverage of the array, making them impossible to eliminate a priori.

Given that completely eliminating these degeneracies is extremely difficult, generic VLBI imaging must be able to explore this space of degeneracies and quantify the resulting uncertainty. In that vein, we seek a method that allows us to quantify and even parameterize the space of potential images, and together solve for both the image and instrumental response. To regularize the space of images, we first notice that regularizing the space of images, it is often natural to decompose the set of unknown parameters into different components.

One component is parameters directly related to the image intensities, which we will denote generically as \mathbf{r} . These could be, for instance, the specific intensities for a rasterized image. The second set of parameters, which we denote by ϕ , encode global properties of the image or parameters of the regularization of \mathbf{r} . Such parameters include the image correlation length, image variation, image size, and image shape. Following the language of statistical inference, we call \mathbf{r} the latent variables, and ϕ the parameters or hyperparameters of the model. The space of potential images we wish to explore is then given by the joint prior

$$p(\mathbf{r}, \phi) = p(\mathbf{r} | \phi)p(\phi), \quad (7)$$

where the right-hand side utilizes Bayes' rule to explicitly denote the conditional dependence of the latent variables on the global parameters of the model. This kind of prior is typically called a hierarchical Bayesian prior, and reflects the natural hierarchy of knowledge that is typical in scientific inference.

Within the Bayesian framework, adding the instrumental response is straightforward: include the gains as parameters with priors in the model. Putting the different model components together, the total imaging posterior becomes

$$p(\mathbf{r}, \phi, \mathbf{s} | \mathbf{V}) = \frac{\mathcal{L}(\mathbf{V} | \mathbf{r}, \phi, \mathbf{s})p(\mathbf{r} | \phi)p(\phi)p(\mathbf{s})}{p(\mathbf{V})}. \quad (8)$$

We call this formulation of VLBI imaging hierarchical interferometric Bayesian imaging (HIBI). The benefit of this approach is that depending on the problem at hand, we can

assess a large variety of different problems that arise in VLBI inference. For instance, for standard imaging problems, our main goal is to compute averages or expectations of the image structure. This is typically expressed in terms of the marginal posterior $p(\mathbf{r}|\mathbf{V})$, which averages over both the different parameters and instrumental effects. That is, we are averaging over the impact of different imaging priors and their related instrumental effects. In this setting, ϕ and \mathbf{s} serve as nuisance parameters whose uncertainty we do want to consider but whose values are not of direct interest to the scientific output.

Conversely, suppose that we are modeling the image as some average feature, i.e., the average state of an accretion flow, plus a turbulent process. Within this setting, \mathbf{r} would represent the specific realization of the turbulent field, and ϕ would represent the physical parameters of interest, e.g., black hole mass, ring size, and black hole spin. The distribution of interest in this case would be $p(\phi|\mathbf{V})$ and so \mathbf{r} would be the nuisance parameters that we average over. HIBI provides direct uncertainty quantification in both instances, and in this paper, we will explore both perspectives.

This paper is organized as follows. In Section 2, we present a specific implementation of HIBI in terms of Markov random fields and then relate it to other imaging techniques. In Section 3, we test HIBI on a variety of image structures based on the 2017 Event Horizon Telescope (EHT) observations of M87* (Event Horizon Telescope Collaboration et al. 2019c, hereafter Paper IV) and apply HIBI to the 2017 EHT M87* data reproducing the results from Paper IV. In Section 4, we demonstrate how additional physical information, e.g., the expected appearance of an optically thin accretion flow, can be incorporated into the HIBI framework and reduce the uncertainty of image parameter estimates. In Section 5 we apply HIBI to the active galactic nucleus (AGN) OJ 287 across three frequencies independently, demonstrating HIBI’s super-resolution capabilities. Finally, in Section 6 we summarize our results and discuss future work.

Finally, the HIBI algorithms and a tutorial are available in the Comrade VLBI analysis package (P. Tiede 2022). All tests in this paper were run with Comrade version 0.11.18 and Julia 1.10.

2. Hierarchical Bayesian Imaging

The locations of the pixels are given by

$$\begin{aligned} x_i &= \text{FOV}_x \left[-\frac{1}{2} + \frac{1}{N_x} (i - 1/2) \right] \\ y_j &= \text{FOV}_y \left[-\frac{1}{2} + \frac{1}{N_y} (j - 1/2) \right], \end{aligned} \quad (9)$$

where $\text{FOV}_{x,y}$ is the field of view of the image, and i, j go from 1 to N_x and N_y , respectively. Note that $\Delta(x, y) = \text{FOV}_{x,y}/N_{x,y}$ are the pixels sizes. Associated with each pixel is a specific intensity, F_{ij} , whose collection we denote by \mathbf{F} . To create a continuous representation of the image, we use the formula

$$I_{\mathbf{F}}(x, y) = \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} F_{ij} \kappa \left(\frac{x - x_i}{\Delta x}, \frac{y - y_j}{\Delta y} \right), \quad (10)$$

where $\kappa(x, y)$ is a continuous function often called the *kernel* or *pulse* function.

The Fourier transform of Equation (10) is given by

$$\tilde{I}_{\mathbf{F}}(u, v) = \tilde{\kappa}(u\Delta x, v\Delta y) \Delta x \Delta y \sum_{ij} F_{ij} e^{2\pi i(ux_i + vy_j)}, \quad (11)$$

where $\tilde{\kappa}$ is the Fourier transform of κ . To efficiently compute the visibilities for a set of $u - v$ locations, we use a nonuniform fast Fourier transform implemented within the `NFFT.jl` package (T. Knopp et al. 2023). For our pulse function, we use the third-order B-spline $b_3(x)$, which is given by the square wave pulse,

$$b_0(x) = \begin{cases} 1 & -1/2 < x < 1/2 \\ 0 & \text{otherwise} \end{cases}, \quad (12)$$

convolved with itself three times. Note that the pulse function choice is arbitrary, and our results do not strongly depend on it.

2.1. Prior Properties of the Stochastic Process

Different image priors may give qualitatively different results. Therefore, we aim to construct a prior that encodes the minimal properties we expect our images to obey. For a general image I , two generic properties should be considered:

1. *Intensities are positive.* $I(x, y) \geq 0$ for all (x, y) . For our raster, $F_{ij} \geq 0$ for all i, j .
2. *Spatial Correlation.* Neighboring pixels should have similar fluxes.

Positivity is enforced because each F_{ij} measures an energy flux density. The inclusion of spatial correlation in the model is intended to enforce the idea that astrophysical images are often correlated at some spatial scale. Previous Bayesian imaging methods (A. E. Broderick et al. 2020a; D. W. Pesce 2021) have assumed uncorrelated image priors.⁸ As a result, the number of pixels becomes an important hyperparameter that can influence the structure of the resulting image posterior. If too few pixels are used, the model underfits the data, often leading to overly tight posteriors. At the same time, too many pixels cause overfitting of the data, resulting in an overly broad posterior and weak inference. A. E. Broderick et al. (2020a) used the Bayesian information criterion to find the optimal number of pixels. However, estimating the Bayesian evidence for each set of data is computationally expensive. Furthermore, the Bayesian evidence is also sensitive to the choice of priors and may not always choose the most predictive model (A. Gelman & Y. Yao 2020).

Using a correlated image prior and fitting for the correlation length, we effectively average over different numbers of effective pixels, finding the optimal configuration in a data-driven manner. As a result, the number of pixels is no longer a critical parameter to consider. Instead, rasterization is a discretization of some underlying continuous model, and as the number of pixels increases, the effects of rasterization decrease. Finally, by specifically introducing a correlation parameter, we can impose a prior that utilizes our prior information that the images should not be dominated by features with a much smaller length scale than what the telescope is sensitive to. For VLBI imaging, it is standard practice to suppress structure on scales below ρ_{Dif} .

⁸ The Dirichlet parameterization used in D. W. Pesce (2021) could include local correlation in the concentration parameter matrix.

To parameterize the raster emission, we formulate F_{ij} as

$$F_{ij} = F_0 \frac{\mu(x_i, y_j) \eta_{ij}}{\sum_{ij} \mu(x_i, y_j) \eta_{ij}}. \quad (13)$$

Here, $\mu(x, y)$ models the a priori structure of the image assumed before imaging the data, which are modulated by the stochastic fluctuations η_{ij} . This model implies that the imaging process is a multiplicative stochastic process, where μ models the geometric mean structure of the process, and η encodes the correlation structure of the image. Finally, we normalize the process to have total flux density F_0 so that the degeneracy between total flux F_0 and a total scaling of gain amplitudes is fixed.

The other main degeneracy is the lack of an image centroid constraint. The critical component of this degeneracy is that it strongly correlates the image structure with the gain phases across all sites. To remove this correlation from the gain phases, we first modify Equation (11) by

$$\tilde{I}_C = e^{-2\pi i(ux_C + vy_C)} \tilde{I}_F, \quad (14)$$

where (x_C, y_C) is the centroid of I_F . This rescaling renders the visibility phases (and, hence, the gain phases) independent of the image centroid. Note that this does not uniquely identify the center of the image with respect to the raster. However, in practice we find that computing this projection dramatically improves sampling efficiency when estimating the image posterior.

To ensure that $\eta_{ij} > 0$, we first augment these variables using a transformation from Euclidean space to strictly positive values. For low dynamic range image reconstructions, such as the 2017 EHT array, we seek a function that is approximately linear for values $r \gg 0$ and then decays to zero $r \ll 0$. A simple function that satisfies these constraints is the softplus function:

$$\eta_{ij,sp} = \text{softplus}(r_{ij}) = \log(1 + e^{r_{ij}}). \quad (15)$$

For $r_{ij} \gg 0$, it is approximately linear, while for $r_{ij} \ll 0$, it is similar to a decaying exponential. This ensures that $\eta_{ij} > 0$ smoothly transitions to zero as $r_{ij} \rightarrow -\infty$, while remaining roughly linear for larger values, i.e., where the image is bright.

For very high-dynamic range images, such as Very Long Baseline Array (VLBA) jet imaging, which we consider in Section 5, placing the image fluctuations on a linear scale is often a poor description of the data, e.g., the jet is often orders of magnitude dimmer than the core. Therefore, for these higher dynamic range images, the exponential function is preferable to map between fluxes F_{ij} and r_{ij} :

$$\eta_{ij,sm} = e^{r_{ij}}. \quad (16)$$

Given the decomposition Equation (13), and our specification of η_{ij} in terms of r_{ij} , the last component of the model is the prior structure for \mathbf{r} , which we now specify.

2.2. Correlated Image Prior

Gaussian random fields (GRFs) provide a flexible framework for parameterizing the space of functions, thereby forming a prior on this space. The Gaussian nature of these fields implies that the distribution of functions is completely characterized by the mean function $m(\mathbf{x})$ and a two-point autocorrelation function $k(\mathbf{x}, \mathbf{x}')$. Therefore, we assume that stochastic fluctuations of our image, r_{ij} , are the discretization

of some continuous GRF. We assume that the process is mean zero: $m(\mathbf{x}) \equiv 0$. Therefore, to specify our GRF, only $k(x_i, x_j')$ must be specified.

Given a discretization, \mathbf{x}_{ij} , the autocorrelation function becomes a matrix Σ_{ij}

$$\Sigma_{ij,(ij)'} = k(\mathbf{x}_{ij}, \mathbf{x}_{(ij)'}), \quad (17)$$

and the transformed image fluctuations are a zero-mean multivariate Gaussian

$$\mathbf{r} \sim \mathcal{N}(0, \Sigma). \quad (18)$$

However, choosing an autocorrelation function k is not straightforward. To ensure that k is a valid GRF, we need to ensure that Σ will always be a positive definite covariance matrix. One way to simplify the construction of k is to assume that the GRF is stationary. That is, the statistical properties at \mathbf{x} and at $\mathbf{x} + \mathbf{c}$ are identical. This implies that $k(\mathbf{x}, \mathbf{x}') = k(\|\mathbf{x} - \mathbf{x}'\|)$ and the Wiener-Khinchin theorem then states that $k(r)$ is the Fourier transform of a positive semidefinite function or measure often called the power spectrum or spectral density. Stationary GRFs⁹ form the basis of the RESOLVE algorithm (P. Arras et al. 2019), and have been successfully applied to previous EHT observations (P. Arras et al. 2022).

For stationary GRFs on a regular Cartesian grid with periodic boundary conditions, the computational complexity to evaluate the GRF is $\mathcal{O}(K \log K)$. However, a general GRF typically scales $\mathcal{O}(K^3)$ for irregular grids, making it infeasible for large rasters. Another complication in estimating the GRFs is that estimating the correct shape of the power spectrum for small data volume, such as for the 2017 EHT arrays, can be quite difficult (G.-A. Fuglstad et al. 2019).

Rather than solving for a general power spectrum, in this paper, we restrict our correlated image priors to Gaussian Markov random fields (GMRFs). The basis of GMRFs is that the inverse covariance matrix, or precision matrix \mathbf{Q} , encodes the condition dependencies between variables. That is, suppose that r_{ij}, r_{jl} are conditionally independent given the rest of the variables $r_{-(ij,jl)}$ ¹⁰, i.e., that

$$p(r_{ij}, r_{kl} | r_{-(ij,kl)}) = p(r_{ij} | r_{-(ij,kl)}) p(r_{kl} | r_{-(ij,kl)}). \quad (19)$$

Then the resultant precision matrix $\mathbf{Q}_{ij,kl} = 0$. For a Markov random field where the pixels will depend only on some small set of neighbors, this implies that the precision matrix will be sparse (see, e.g., H. Rue & L. Held 2005).

In general, GMRFs are given by the formula

$$p(\mathbf{r} | \boldsymbol{\mu}, \mathbf{Q}) = \sqrt{\frac{\det \mathbf{Q}}{2\pi^K}} \exp\left(-\frac{1}{2} \mathbf{r}^T \mathbf{Q} \mathbf{r}\right), \quad (20)$$

where \mathbf{Q} is a sparse precision matrix. GMRF priors have been extensively used in geostatistical and epidemiological (see, e.g., F. Lindgren et al. 2022, for a review) studies, form the basis of the popular R-INLA spatiotemporal statistical modeling package (T. G. Martins et al. 2013), and have already been used in black hole inference by EHT (D. Lee & C. F. Gammie 2021; A. Levis et al. 2021). This paper will consider simple stochastic spatial fluctuations encoded by the precision matrix \mathbf{Q} , which we now describe.

⁹ RESOLVE uses a GRF prior on both the image and the power spectrum.

¹⁰ The negative sign means all pixel values except $-ij$.

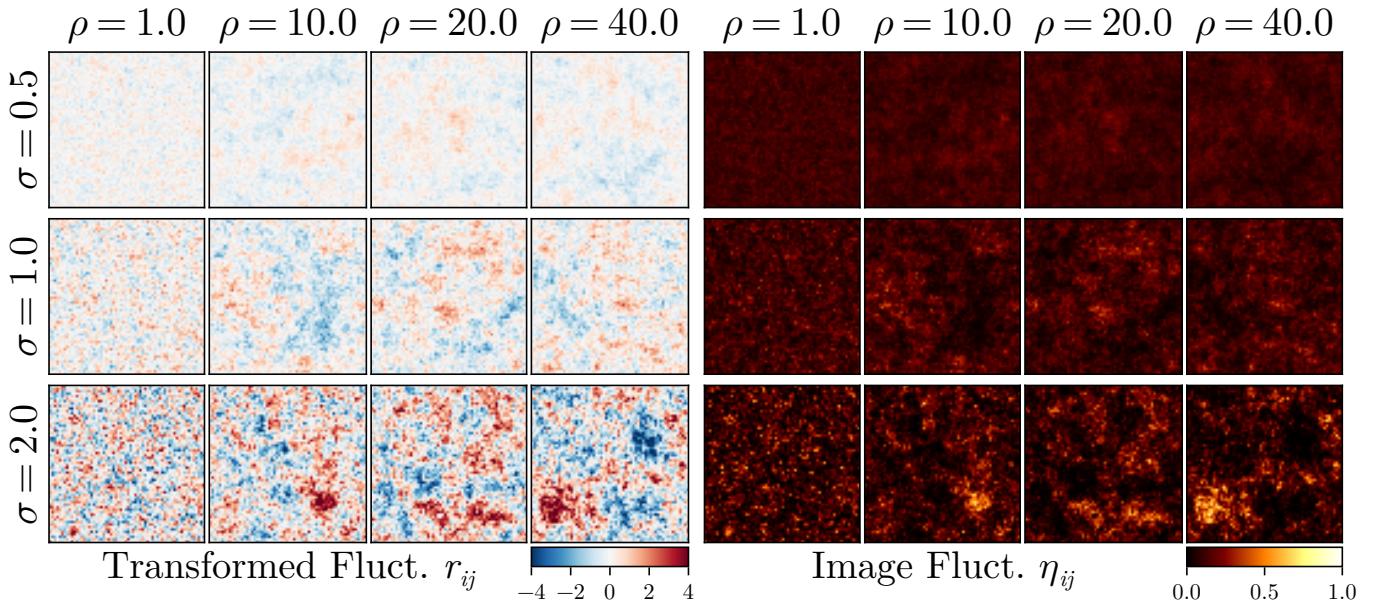


Figure 1. Draws from the GMRF prior in the \mathbf{r} space (left) and in the linear image space η (right) assuming that they are related by Equation (15). The columns of the figures show different correlation lengths for the log-ratio image, while the rows are different values of the image dispersion.

2.2.1. First-order Markov Random Field

For simplicity and to connect with RML imaging, this paper will mainly use a first-order GMRF, whose precision matrix is denoted \mathbf{Q}_1 . To start, we divide our precision matrix into two parts:

$$\mathbf{Q}_1 = \frac{1}{\sigma^2} \left(\frac{1}{\rho^2} \mathbf{1} + \mathbf{G} \right), \quad (21)$$

where $\mathbf{1}$ denotes the identity matrix, and \mathbf{G} sets the correlation structure of the random field. For this work, we followed F. Lindgren et al. (2011) and used the first-order intrinsic GMRF matrix:

$$G_{ij,kl} = \begin{cases} 4 & ij = kl \\ -1 & ij \sim kl \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

where $ij \sim kl$ means that on the image pixel grid, ij is adjacent to kl . With this description, J. Besag (1981) demonstrated that this 2D GMRF Equation (21) converges to a continuous process with spectral density,

$$S(\mathbf{k}) d^2\mathbf{k} \propto \frac{d^2\mathbf{k}}{1 + \rho^2 \|\mathbf{k}\|^2}. \quad (23)$$

This spectral density implies that the first-order GMRF is a stochastic process with a power law with slope -1 and break $\sim \rho$. Therefore, we interpret σ^2 and ρ as hyperparameters that modulate the marginal variance and correlation length of the process.

Random draws from the GMRF distribution are shown in Figure 1 in both \mathbf{r} and the η representation.

To generate higher-order GMRF, F. Lindgren et al. (2011) showed that by convolving this process with itself n times, it approximates the n^{th} -order *Matern* Gaussian process. We briefly overview higher-order GMRFs and Matern processes in Appendix G.¹¹ A benefit of the GMRF approach is that the

eigenvalues of \mathbf{Q} are easily computed in $\mathcal{O}(K)$ time and are given by

$$\det \mathbf{Q}_1 = \sigma^{-2K} \prod_{n=1}^{N_x} \prod_{m=1}^{N_y} (\rho^{-2} + \Lambda_{nm})$$

$$\Lambda_{nm} = 4 + 2 \cos [\pi n / (N_x + 1)] + 2 \cos [\pi m / (N_y + 1)], \quad (24)$$

which is derived in Appendix F. This makes computing the GMRF prior $\mathcal{O}(K)$, cheaper than the Fourier representation of generic stationary random fields ($\mathcal{O}(K \log K)$), albeit at the cost of flexibility.

2.3. Parameter Estimation

Within the Bayesian interpretation of regularization and GMRF priors, there is a natural and self-consistent way to estimate the image and the “regularization” hyperparameters through a hierarchical model. The Bayesian image posterior with the GMRF prior Equation (28) up to constants is given by

$$\log p(\mathbf{r}, \rho, \sigma | \mathbf{V}) = \log \mathcal{L}(\mathbf{V} | \theta) - \frac{1}{2} \mathbf{r}^T \mathbf{Q}_1 \mathbf{r} + \frac{1}{2} \log \det \mathbf{Q}(\rho, \sigma) + \log p(\rho, \sigma) \quad (25)$$

where constants denote terms that do not depend on any of the posterior parameters, and $\log p(\rho, \sigma)$ is the prior on the imaging hyperparameters. Ignoring the hyperparameter prior and constants, we see that Equation (25) is equal to the negative of Equation (26) up to the addition of the log determinant of the precision matrix. Inspecting Equation (24), we see that $\det \mathbf{Q}$ acts as a volume correction that prevents the collapse of the regularizer weights. When the prior is very wide (ρ small and σ large), $\det \mathbf{Q}$ is close to zero. In contrast, as the GMRF prior volume shrinks (ρ large and σ small), we get $\det \mathbf{Q}$ to grow geometrically. Therefore, the Bayesian regularization formulation favors images with large ρ and small σ , namely simpler images closer to our prior image. This

¹¹ See the Julia package `VLBIImagePriors.jl` for its open source implementation.

simplicity is then balanced against the data likelihood terms, resulting in a fit that naturally increases the prior volume to account for data complexity.

2.4. Relating HIBI to Other Imaging Methods

2.4.1. RML Methods

The usage of priors in Bayesian imaging has been qualitatively compared to RML imaging regularizers, which are commonly used in VLBI (R. Narayan & R. Nityananda 1986; K. L. Bouman et al. 2016; A. A. Chael et al. 2016, 2018; K. Kuramochi et al. 2018). This section explores the relationship between the GMRF prior Equation (22) and two commonly used regularizers.

RML imaging is a maximum likelihood method that reconstructs a single image given the data by minimizing the cost function

$$\mathcal{J}(\mathbf{F}; \boldsymbol{\lambda}, \mathbf{D}) = \sum_{d \in |D|} \alpha_d \chi_d^2(\mathbf{F}) + \sum_{r \in |\lambda|} \lambda_r \mathcal{R}_r(\mathbf{F}), \quad (26)$$

where $|\cdot|$ denotes the set of indices for some tuple, (α_i, χ_i^2) is the weight and data chi-square of data product i , and $(\mathcal{R}_j, \lambda_j)$ is the regularizer and its weight, respectively. Relating HIBI to RML, it is natural to associate HIBI’s data likelihood with the chi-square term in Equation (26), its latent variable prior with the regularizer, and its hyperparameters with the regularizer weights.

For the data terms, HIBI and RML both utilize similar data products; however, from a Bayesian perspective, RML’s data chi-squares will modify the weights of various terms. Additionally, RML assumes that each data product is independent, which is true if, for instance, closures and visibility are simultaneously included in the data fits, which is common in EHT analyses (Event Horizon Telescope Collaboration et al. 2019c). Additionally, the weights α_i are often different from the VLBI likelihood Equation (6); for example, α is usually given by the inverse of the number of measurements, while the VLBI likelihood omits that factor. Note that while RML only cares about relative weights, Bayesian inference, specifically uncertainty estimates, depends on the absolute scale/weights of all data products. Two commonly used regularizers are the total squared variation (TSV) and L_2 regularizers (K. Kuramochi et al. 2018), which are given by

$$\begin{aligned} \mathcal{R}_{\text{TSV}}(\mathbf{F}) &= \sum_{ij} (F_{i+1,j} - F_{i,j})^2 + (F_{i,j+1} - F_{i,j})^2 \\ \mathcal{R}_{L2}(\mathbf{F}) &= \sum_{ij} F_{ij}^2. \end{aligned} \quad (27)$$

Expanding the negative log density of the GMRF prior (Equation (20)),¹² we have

$$\begin{aligned} & -\log p(\mathbf{r}|\rho, \sigma) \\ &= \frac{1}{2\sigma^2} \sum_{ij} (r_{i+1,j} - r_{i,j})^2 + (r_{i,j+1} - r_{i,j})^2 \\ &+ \frac{1}{2(\rho\sigma)^2} \sum_{ij} r_{ij}^2 \\ & - \frac{1}{2} \log \det \mathbf{Q}_1 + \text{constants.} \end{aligned} \quad (28)$$

¹² We are ignoring boundary terms in this equation.

The first term corresponds to the TSV regularization, while the second is the ℓ_2 regularization. Furthermore, we can relate the regularization hyperparameters to the standard deviation and correlation length

$$\rho = \sqrt{\frac{\lambda_{\text{TSV}}}{\lambda_{\ell_2}}} \quad \sigma = \frac{1}{\sqrt{2\lambda_{\text{TSV}}}}. \quad (29)$$

At face value, our prior example is TSV and L_2 regularization with a nonstandard parameterization of the regularization weights. By letting $\rho \rightarrow \infty$, we recover the standard TSV regularization.

Other classes of regularizers can also be included in this framework. For instance, one could consider a combination of the total variation and ℓ_1 regularizers for the log prior:

$$\begin{aligned} & \log p_{\text{TV}}(\mathbf{r}|\lambda_{\text{TV}}, \lambda_{\ell_1}) \\ &= -\lambda_{\text{TV}} \sum_{ij} \sqrt{(r_{i,j+1} - r_{i,j})^2 + (r_{i+1,j} - r_{i,j})^2} \\ & - \lambda_{\ell_1} \sum_{ij} |r_{ij}| + \log N(\lambda_{\text{TV}}, \lambda_{\ell_1}), \end{aligned} \quad (30)$$

where N is the normalization constant. Similar to TSV and ℓ_2 , we can relate the regularization hyperparameters to a pseudo-variance and correlation length using Equation (29). Unfortunately, a closed-form expression for N is not known unless $\lambda_{\ell_1} = 0$. As we will see in Section 2.3, estimating the hyperparameters in a Bayesian framework requires an expression for the normalization N ; therefore, we do not consider this prior family in this paper.

Before continuing, we note an important consideration: RML imaging tends to regularize the image pixel fluxes directly rather than the transformed quantities. To enforce positivity, the image priors are then effectively truncated multivariate normal distributions. However, this truncation makes the calculation of the normalization term in Equation (25) computationally expensive. Given that this normalization constant is critical to self-consistently estimating the optimal hyperparameters, we do not explore applying the GMRF prior directly to the pixel fluxes.

Finally, we note that there exists a simple extension of this formalism to multiple regularizers. The standard RML approach of summing multiple regularizers is not easily expressible within HIBI. Unfortunately, the prior volume term in Equation (25) is generally not analytic. However, we can extend the formalism described in Section 2.3 to multiple priors through model averaging. Let $p_i(\mathbf{r}|\varphi_i)$ denote each prior considered, with their own set of hyperparameters φ_i . For each p_i , a posterior $p_i(\mathbf{r}, \varphi_i | \mathbf{V})$ can be found with the typical HIBI algorithm. These posteriors can be averaged using a model averaging scheme (see, e.g., Y. Yao et al. 2018 for different averaging approaches). Using this averaging approach, multiple image reconstruction priors can be combined, weighing each posterior according to the specific averaging scheme. In Appendix G, we describe different Markov random field priors that could be considered in such a scheme.

2.4.2. Comparing RML with HIBI

Within the HIBI framework, the estimation of the joint image and hyperparameter posterior $\pi(\mathbf{r}, \rho, \sigma | \mathbf{V})$ is analogous to the image and hyperparameter survey in RML imaging. A

natural question is whether the HIBI formalism could be incorporated into RML imaging by including the missing volume term. The maximum likelihood estimate would then be replaced with the maximum a posteriori (MAP) estimate, but otherwise the rest of the RML algorithm would remain the same. However, as we will show below, the MAP estimate is often a poor approximation of the kind of structure characteristic of the posterior. This is often because the MAP estimate tends to overfit the data and results in various image artifacts. However, another problem exists with the MAP estimates in the HIBI imaging framework: the MAP solution is not invariant to parameterizations. Specifically, the HIBI MAP strongly depends on the choice of parameterization, especially for weakly informative data relative to the model complexity, as is often the case for VLBI imaging.

To demonstrate how the location of the MAP can be changed arbitrarily, we consider the common model known as Neal’s funnel. This distribution has the probability density,

$$p(\mathbf{x}, v) = \prod_{i=1}^N \mathcal{N}(x_i|0, e^v) \mathcal{N}(v|0, b^2), \quad (31)$$

and corresponds to a Gaussian hierarchical model where the variance of the process has a log-normal prior with variance b .¹³ The log probability density of this model (ignoring constant terms) is

$$\log p(\mathbf{x}, v) = -\frac{Nv}{2} - \frac{e^{-v}}{2} \sum_{i=1}^N x_i^2 - \frac{v^2}{2b^2}. \quad (32)$$

The MAP does not exist within this parameterization, but its location tends to $v \rightarrow -\infty, x_i = 0$. Note that this parameterization is similar to the one typically used in RML imaging.

Now consider the transformed parameters \tilde{x}, \tilde{v} , where $x_i = e^{b/2}\tilde{x}_i + \tilde{x}_0$, $v = b\tilde{v} + \tilde{v}_0$, where \tilde{x}_0 and \tilde{v}_0 are the effective origin. This parameterization effectively standardizes the distribution, and the probability density becomes the multiplication of two independent unit normal distributions in \tilde{v} and \tilde{x} . In this case, the MAP is finite and is given in our original parameters by $x_i = v = 0$. By making a relatively simple parameterization change, we shifted the map by an infinite amount.

This dependence makes the parameterization or coordinate frame a critical piece of HIBI imaging, even though the expectations from the posterior are invariant to the choice of parameterization. As a result, we do not recommend that RML methods switch to the hierarchical prior approach, given that the choice of parameterization, especially for sparse data, can significantly impact the MAP’s location and imaging results. Instead, we refer the reader to other methods for hyperparameter optimization, such as the multiobjective optimization approach in H. Müller et al. (2023).

2.5. RESOLVE

The RESOLVE algorithm is an example of HIBI with an additional hierarchical level. Namely, RESOLVE typically uses Equation (13), typically with μ flat and the exponential transfer function Equation (16) (J.-S. Kim et al. 2024). For the stochastic field, RESOLVE uses stationary random fields on a periodic regular grid to utilize the power spectrum

decomposition. However, instead of assuming some parametric form for the power spectrum, they then use a second GRF for the power spectrum itself. The prior decomposition equals

$$p(\mathbf{r}, \mathbf{f}, \phi) = p(\mathbf{r}|\mathbf{f})p(\mathbf{f}|\phi)p(\phi), \quad (33)$$

where \mathbf{f} are the parameters of the discretized power spectrum, and ϕ are the hyperparameters for the stochastic process governing \mathbf{f} . This model is more expressive than the GMRF prior described in Section 2.2, which assumes a specific power spectrum slope.

The tradeoff between RESOLVE’s expressibility and the GMRF approach in this paper comes during inference. Namely, in this paper, we employ Markov Chain Monte Carlo (MCMC) methods to approximate the posterior, whereas RESOLVE typically relies on more approximate parametric methods, such as metric Gaussian variational inference (J. Knollmüller & T. A. Enßlin 2019). A downside of the variational approach is that expectations may be poorly approximated if the variational family, e.g., a Gaussian, does not match the posterior. For very sparse arrays like the EHT or VLBA, the posterior may highly non-Gaussian; however, a more detailed study is warranted to better understand the impact of the variational approximation.

Given the sparsity of the EHT data, this paper uses MCMC methods to approximate posterior expectations. Specifically, given that imaging typically requires 1000–10,000 parameters for EHT data, we employ Hamiltonian Monte Carlo and the NUTS algorithm (M. D. Hoffman & A. Gelman 2014). To demonstrate the effectiveness of this approach, we now consider a suite of synthetic data tests.

3. Imaging With HIBI

3.1. Synthetic Data Tests

To demonstrate the applicability of our approach to different image source morphologies, we consider a suite of synthetic data that model the observations of the 2017 EHT array following Event Horizon Telescope Collaboration et al. (2019c):

1. *Disk*: uniform disk with a diameter of 70 μ as, convolved with a Gaussian with an FWHM of 10 μ as.
2. *Double*: Two circular Gaussian components each with an FWHM of 20 μ as and a separation of 30 μ as R.A. and 12 μ as in decl. One Gaussian has a flux of 0.33 Jy and the other 0.27 Jy.
3. *Ring*: An infinitely thin ring with radius 22 μ as, flux 0.6 Jy. The ring is convolved with a 10 μ as FWHM Gaussian.
4. *Crescent 150*: An infinitely thin crescent model with radius 22 μ as, flux 0.6 Jy, and brightness position angle of 150°, which is blurred with a 10 μ as FWHM Gaussian.
5. *Crescent 180*: An infinitely thin crescent model with radius 22 μ as, flux 0.6 Jy, and brightness position angle of 180°, which is blurred with a 10 μ as FWHM Gaussian.
6. *GRMHD*: A general relativistic magnetohydrodynamic (GRMHD) simulation of M87. We use the same simulation as in Event Horizon Telescope Collaboration et al. (2019c).

¹³ This distribution is equivalent to the VLBI image posterior with the first-order uncorrelated GMRF prior and no observations.

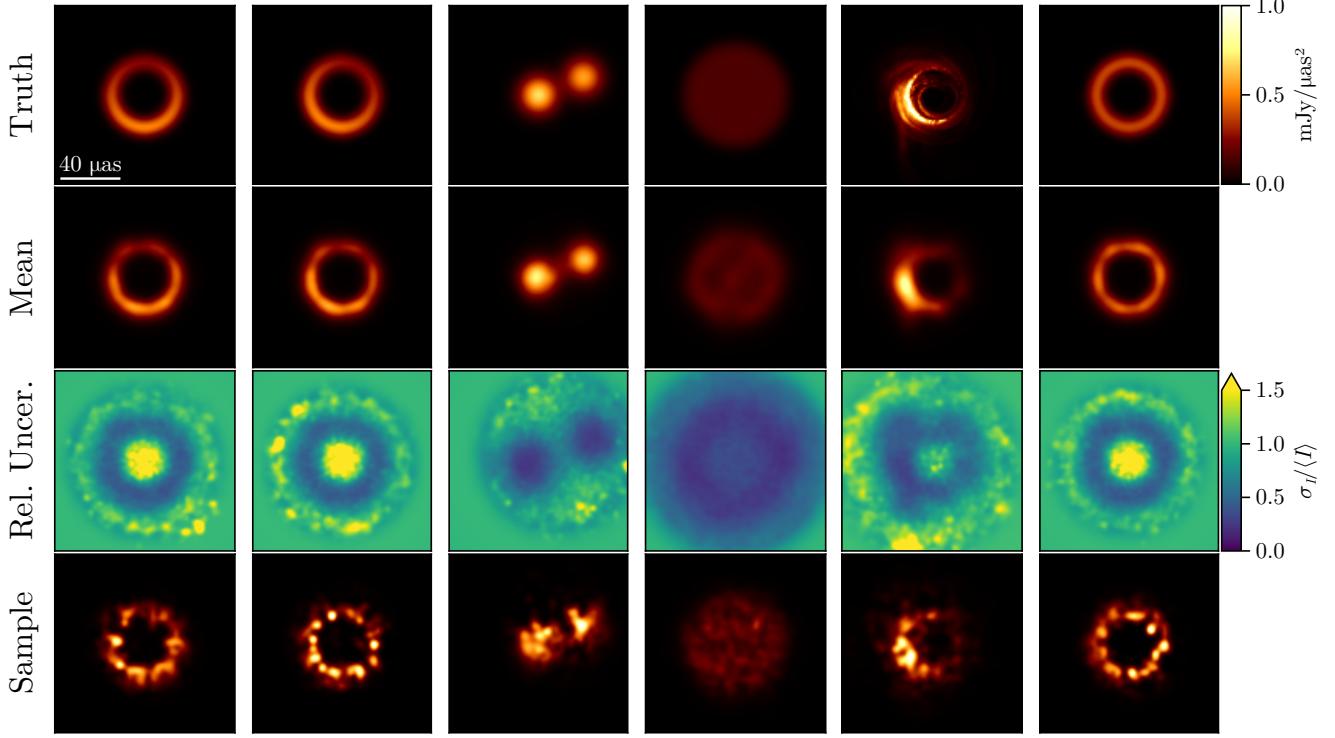


Figure 2. Example reconstructions of various geometric model morphologies using the GMRF prior. The prior is flexible enough to capture various image morphologies and features, including the ring brightness profile.

The ground truth images are shown in the first row of Figure 2. For each image, we then create synthetic data using eht-imaging with the EHT coverage and thermal noise properties of the observation of the LO band on April 6. The synthetic data for each source was also corrupted with random gain phase and gain amplitude errors. In addition, we added a fractional systematic error equal to 2% of the measured visibility amplitudes in quadrature to the thermal noise to all baselines. This additional systematic error models the unresolved errors in the 2017 EHT data, e.g., leakage and coherence loss, that were estimated in Event Horizon Telescope Collaboration et al. (2019d). Note that the same systematic error budget was added to the 2017 EHT data in Section 3.4.

3.2. Model Priors

For imaging, we used the GMRF prior, with a field of view of 200 μ as with a 64×64 raster. A Gaussian profile was used for a priori image structure with the parameterization:

$$\mu_g(\mathbf{x}) = \frac{4 \log(2)}{\pi s_g^2} e^{-4 \log(2) \mathbf{x} \cdot \mathbf{x} / s_g^2}. \quad (34)$$

The size or FWHM of the Gaussian profile, s_g , is included as a free parameter during inference with the prior,

$$s_g \sim \mathcal{TN}(50 \mu\text{as}, 20 \mu\text{as}; a = 20 \mu\text{as}, b = 100 \mu\text{as}), \quad (35)$$

where $\mathcal{TN}(\mu, \sigma; a, b)$ is a truncated normal distribution with mean, μ , standard deviation σ , and lower and upper bounds a and b , respectively. This prior was chosen to match the estimated size of M87* based on the analysis from Paper IV.

For the log-ratio correlation length of the GMRF ρ and dispersion σ , we used the priors

$$\rho \sim \mathcal{IG}[1, -\log(0.01) \lambda_{\text{EHT}} / \Delta x] \quad (36)$$

$$\sigma \sim \mathcal{HN}(0, 0.5^2), \quad (37)$$

where $\mathcal{IG}(\alpha, \beta)$ denotes inverse-gamma distribution,

$$p_{\mathcal{IG}}(x|\alpha, \beta) \propto x^{-(\alpha+1)} e^{-\beta/x} \quad (38)$$

and $\mathcal{HN}(\mu, \sigma)$ is the half-normal distribution. The parameters for ρ were chosen such that 1% of the prior mass for the correlation length of the MRF was below the telescope beam size.

For our instrument model priors, we used a log-normal prior on the gain amplitudes with a log-mean of zero and a standard deviation of 0.2 for all baselines except LMT, which assumed a standard deviation of 1 to model the poor performance of LMT in 2017 observations (Event Horizon Telescope Collaboration et al. 2019d). For the gain phases, we used a von Mises distribution,

$$p_{\text{VM}}(\varphi|\mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} e^{\kappa \cos(\varphi - \mu)}, \quad (39)$$

with mean zero and concentration parameter $\kappa = \pi^{-2}$ for all non-Atacama Large Millimeter/submillimeter Array (ALMA) gain phases. This prior is essentially flat on its support $[0, 2\pi]$ and is temporally uncorrelated. ALMA is used as the reference station of the array, meaning we set its gain phases to 0 for all scans.

3.3. Synthetic Data Results

To reconstruct the image, we used a two-step procedure. The first step is similar to RML, where we found the MAP

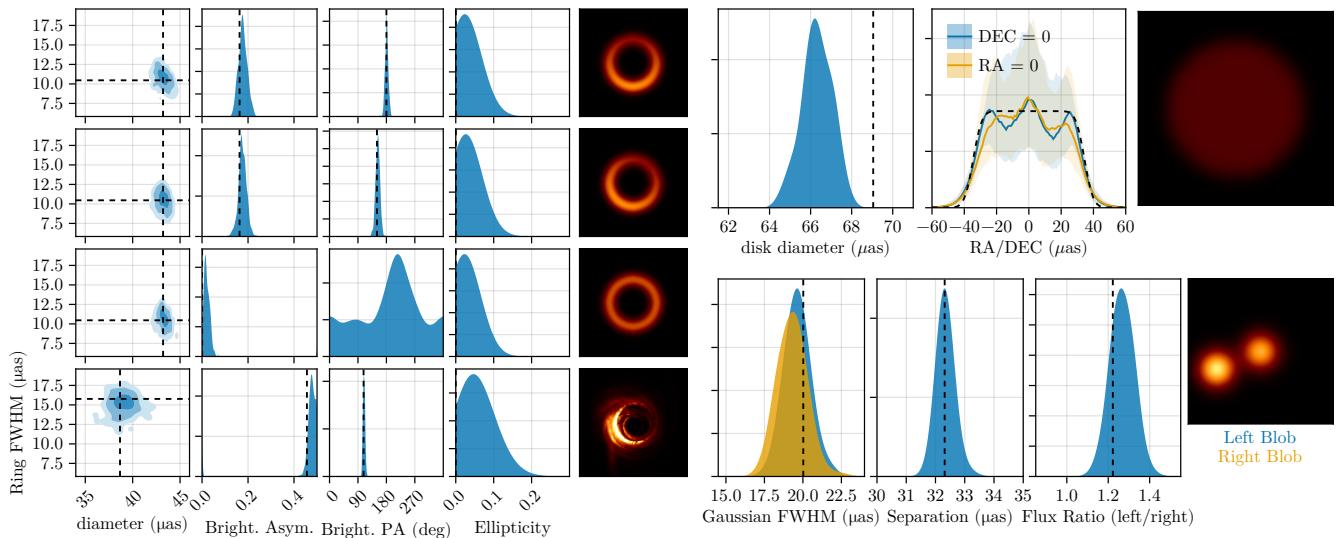


Figure 3. The parameter estimation results from the synthetic data tests in Figure 2. Overall, HIBI can measure all of the parameters of the ground truth image except the disk, where the diameter definition gives a slightly too small result. Analyzing the profile of the disk reconstruction, shown between the disk diameter marginal and on-sky truth, we see that at smaller distances from the origin ($<20\ \mu\text{as}$), the disk is slightly too dim, while at larger distances ($>40\ \mu\text{as}$) it is too bright.

estimate using the `Optimization.jl` library and the Adam optimizer from `Optimisers.jl`. The goal of the first step was to reduce the burn-in time when sampling the posterior, which is the second step. To sample the posterior, we used the `AdvancedHMC.jl` NUTS sampler with a diagonal mass matrix adaptively tuned over 8000 MCMC steps. To compute the posterior gradient with respect to the parameters, we used the automatic differentiation software `Enzyme` and its Julia extension `Enzyme.jl` (W. Moses & V. Churavy 2020; W. S. Moses et al. 2021, 2022). After tuning, the sampler ran for an additional 7000 MCMC steps. The runtime of the sampler depended on the dataset, but it generally took 1–2 hr for Stokes I imaging on an AMD Ryzen 7950X CPU using a single core.

The image reconstructions for each synthetic data test are shown in Figure 2. The second row shows the mean reconstruction, and the third row shows the relative uncertainty of the posterior reconstructions. Qualitatively, we see that HIBI recovers all of the different image morphologies considered.

Note that the disk displays a ring-like inner depression in the mean image. This is only visible in the mean image, and reflects the fact that the central region of the disk is quite uncertain (see Figure 3). In conjunction with the a priori image being a Gaussian that prefers to concentrate flux near the center, the mean image tends to be brighter near the middle of the disk. A benefit of HIBI’s Bayesian approach is that we can assess the significance of this feature and other quantitative features. From the third row of Figure 3, we see that the relative uncertainty in the central region of the disk is larger than near the edge. Samples from the disk do not consistently show an increased flux near the center, which means that this enhanced brightness near the disk center is not statistically significant.

To quantitatively assess image reconstruction quality, we apply `VIDA` (P. Tiede et al. 2022a), a template matching scheme, to extract estimates of key image features. `VIDA` requires a cost function (usually a probability divergence) and a template family to match to the observed image. For the crescent, ring, and GRMHD models, we followed P. Tiede

et al. (2022b) and used an elliptical ring template with a fourth-order cosine expansion for the azimuthal brightness profile. For a definition of the template, see P. Tiede et al. (2022b) for a complete description. The parameters we report in this paper are:

1. $d = \sqrt{ab}$ the geometric average of the ellipse semi-major/minor axis of the elliptical ring template
2. w the FWHM of the Gaussian ring profile
3. A one-half the first-order amplitude in the cosine expansion, which we refer to as the *brightness asymmetry*
4. ξ the first-order phase in the cosine expansion, which we refer to as the *ring position angle (PA)*.

For the double synthetic data, we used a two-component circular Gaussian blob template, and modeled the two Gaussian blob relative separations and their absolute size. For the disk synthetic data, we used a top-hat disk with a Gaussian taper for the disk synthetic data template (see P. Tiede et al. 2022a, for a description) and only report the diameter of the disk.

Given a template family, the optimal template parameters are found by minimizing the Bhattacharyya divergence,

$$\text{Bh}(t|r) = -\log \sum_{n,m} \sqrt{r_{nm} t_{nm}}. \quad (40)$$

To optimize the function, we use the *Evolutionary Centers Algorithm* from the Julia package `Metaheuristics.jl` (J.-A. M. de Dios & E. Mezura-Montes 2022) within the `Optimization.jl` package.

The parameter estimation results for all models are summarized in Figure 3. For each synthetic data test, the true values are within the bulk of the posterior except for the disk diameter. Analyzing the reason for the bias, we plot the true disk profile compared to the recovered profile in Figure 3. The origin of the bias is that the profile of the disk edge is not accurately reconstructed, especially in the vertical direction, where it is biased to be small near the center and larger near the edge. This is due to the choice of a Gaussian profile μ , which a priori concentrates the emission near the center, as

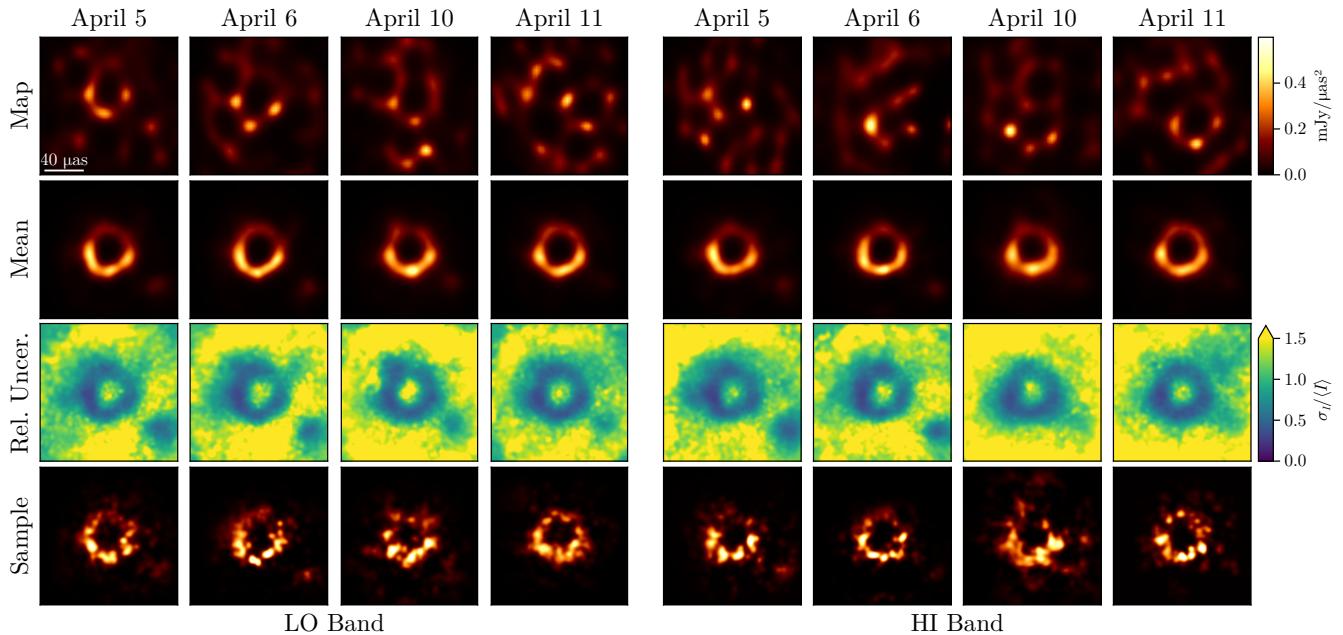


Figure 4. HIBI image reconstructions of M87* for each observing day and frequency band. The top row shows the MAP image, which is often littered with artifacts, such as ghost rings. The second and third rows display the mean image and its relative uncertainty, respectively. Finally, the bottom row shows a random sample from the posterior, blurred to match the resolution of the mean image using the NXCORR metric.

seen in the profile. In Appendix D, we demonstrate that a different μ can remove this bias and recover the true diameter of the disk. Regardless, this bias is minor and generally indicates that HIBI can recover a wide variety of image structures, even when the a priori image structure is not an accurate description. This robustness is critical for arrays such as the EHT, since it provides novel insights into the core region of a variety of AGNs (J.-Y. Kim et al. 2020; M. Janssen et al. 2021; S. Issaoun et al. 2022; S. Jorstad et al. 2023; J. Röder et al. 2025) whose structure is highly uncertain a priori.

3.4. Application to 2017 EHT M87 Data

In the previous section, we demonstrated that the GMRF prior and hyperparameter selection algorithms we developed are capable of recovering a wide array of image features from synthetic EHT data. This section applies an identical procedure to the 2017 EHT data.¹⁴ We explore the stability of the image and the properties of the ring.

For the reconstructions of M87*, we followed a similar procedure to Section 3. The instrument priors are identical to those for the synthetic data, except on April 11, where a large initial gain in JCMT was found. Therefore, for the data on April 11, we used a log-normal prior with zero-mean and a standard deviation of 0.5 for JCMT.

The imaging results for the 4 days and the two frequency bands are shown in Figure 4. The top row of Figure 4 demonstrates the dangers of using an MAP estimate compared to posterior expectations (e.g., the second row of Figure 4). For the imaging results for M87*, the MAP estimate had a reduced chi-square of $\sim 0.4 - 0.5$, significantly smaller than the reduced chi-square in the posterior bulk (1.0 ± 0.1). This result implies that the MAP estimate overfits the data and produces

significant artifacts in the image. Samples from the posterior row in the lower row of Figure 4 demonstrate that the MAP is not characteristic of the posterior bulk and should not be used to assess the features of the image when using HIBI.

The second row shows that the mean posterior image is qualitatively similar throughout the four days, as found in Paper IV. We examined the relative uncertainty map and found that the ring is robustly recovered using HIBI. However, beyond the ring, there is evidence for an extended emission knot in the southwestern region of the image. This extended emission was also found in P. Arras et al. (2022), A. E. Broderick et al. (2022), and C. L. Carilli & N. Thyagarajan (2022).

4. Parameter Estimation with HIBI

Typically, VLBI imaging attempts to be largely agnostic about the on-sky source structure. This approach is ideal when we do not want to make strong structural assumptions about the source, as was the case for the first images of M87*'s shadow. However, this approach often ignores our physical understanding of the source and introduces additional sources of uncertainty in our conclusions. Relatedly, being entirely agnostic to the underlying physics often means that physical quantities of interest, e.g., black hole mass and spin, must be extracted, introducing new sources of uncertainty and methodology.

In this section, we take a different perspective and view the image itself as a set of nuisance parameters, while the parameters ϕ represent the physical quantities of interest. By using this formulation, we will demonstrate how HIBI can incorporate our knowledge that optically thin accretion flows around supermassive black holes generically produce ring-like morphologies (e.g., Event Horizon Telescope Collaboration et al. 2019e; Event Horizon Telescope Collaboration et al. 2022), and estimate physical parameters of interest directly, e.g., the ring size and its profile.

We seek to create a simple model that describes the generic parameters of the emission profile around a supermassive

¹⁴ We used the public EHT data on the CyVerse Data Commons (The Event Horizon Telescope Collaboration 2019).

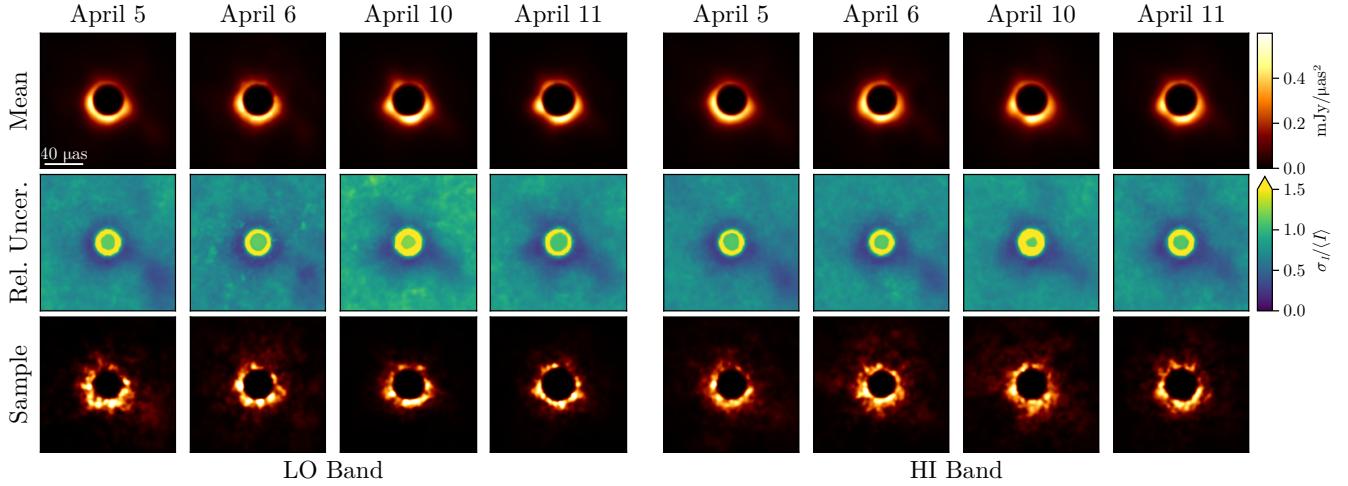


Figure 5. Reconstructions of M87* using the ring profile prior and first-order GMRF for multiplicative noise.

black hole. From simulations of accretion flows, these emission profiles often appear to have a profile where the inner profile differs from the outer profile (Event Horizon Telescope Collaboration et al. 2019e; W. Lockhart & S. E. Gralla 2022a). A simple model that captures this behavior is the double power-law radial profile

$$\mu_r(\mathbf{r}|r_0, \alpha_{\text{in}}, \alpha_{\text{out}}) = N \frac{(r/r_0)^{\alpha_{\text{in}}}}{1 + (r/r_0)^{\alpha_{\text{in}} + \alpha_{\text{out}} + 1}}, \quad (41)$$

Here, $\alpha_{\text{in,out}}$ are the inner and outer power-law indices, r_0 denotes the rough location of the transition of the power law from the inner to outer scale, and N is a normalization constant to ensure the mean image raster has unit flux. The radius of the peak intensity in Equation (41) is given by

$$r_{\text{peak}} = r_0 \left(\frac{\alpha_{\text{in}}}{\alpha_{\text{out}} + 1} \right)^{1/(\alpha_{\text{in}} + \alpha_{\text{out}} + 1)}. \quad (42)$$

To incorporate this model into the HIBI framework, we replace the mean profile μ with Equation (41) and include $\alpha_{\text{in,out}}, r_0$ as parameters to be estimated as part of the HIBI posterior. Compared to Equation (10), we kept the other parameters and priors identical, except we increased the number of pixels in the raster to 95×95 to better resolve the mean image structure when α was large. Specifically, we maintain the same priors for \mathbf{r} and its related parameters ρ and σ . The only remaining piece is to specify the priors for the ring parameters. For the characteristic radius r_0 , we used a uniform prior $\mathcal{U}(5 \mu\text{as}, 40 \mu\text{as})$; for the inner power law, we used the uniform prior $\mathcal{U}(0, 10)$, and for the outer power law, we used the prior $\mathcal{U}(1, 10)$. Setting $\alpha_{\text{in}} = 0$ yields profiles without central depression, i.e., disk-like profiles.

Note that while in this section we are mainly concerned with the parameters $r_0, \alpha_{\text{in,out}}$, we will also produce a new set of images of M87*. Given that M87* has been established as a ring, we expect the uncertainty of the image to decrease, since a large portion of the image will be described by the simple profile μ_r . One potential concern with this informative model for μ is that it can bias the results if there is no ring in the image. To test this, we repeated the synthetic data tests from Section 3. Generally, we found better results with the ring prior than the previous Gaussian prior image. We suspect that the reason for this is that the ring mean image better describes the general

image structure for four out of five of the synthetic datasets. The profile Equation (41) is a near-perfect match for the symmetric ring and disk models, and the GMRF models the brightness asymmetry in the crescent and GRMHD models. The only nontrivial result is the double Gaussian blob model. In this case, the ring radius matches the blob separation, and the GMRF then modulates the rest of the brightness to match the true on-sky image. For more details on the results, see Appendix D.

Moving to the estimation of M87*'s ring parameters, we repeat the analysis of Section 3.4 using this ring mean image prior. To fit the M87* data, we use the same sampling procedure as Section 3.3. The image reconstructions are shown in Figure 5. Qualitatively, the ring morphology appears similar to Figure 4, with a circular ring close to $\sim 40 \mu\text{as}$ and brighter at the bottom. To extract the ring parameters, we again use VIDA with the same fourth-order ring template. The parameter estimates relative to the more agnostic prior are shown in Figure 6. We find consistent values with the Gaussian profile. However, the uncertainty of the diameter and width of the ring are reduced by around 50%, and the brightness asymmetry and position angle are reduced by 10%–20%. This reduced uncertainty is expected, as the observed ring more closely matches the assumed μ structure, thereby reducing the variance of the GMRF and lowering the overall uncertainty in the image reconstructions.

Comparing both the ring and Gaussian profile HIBI imaging results to the original EHT results (Paper IV), we find consistent results, although the uncertainties reported tend to differ. For both the Gaussian and ring profiles, the measured uncertainties for the asymmetry and position angle reported by HIBI tend to be larger than those originally reported by the EHT. This is likely because the uncertainty reported by the EHT did not account for the uncertainty in image reconstruction given a set of hyperparameters. Instead, the EHT's hyperparameter surveys focused on estimating changes in the image structure due to different hyperparameter choices. HIBI estimates both the image and hyperparameter uncertainties jointly. For the ring width, we find that both the Gaussian and ring profiles give consistent measurements, unlike the EHT results, which only reported an upper bound of $20 \mu\text{as}$.

To test whether HIBI's ring width estimate is reliable, we imaged synthetic data of a symmetric delta ring blurred with a Gaussian with FWHM of $5 \mu\text{as}, 10 \mu\text{as}, 15 \mu\text{as}$, and $20 \mu\text{as}$ using both the Gaussian and ring profiles. The detailed results are shown in Appendix C. We found that the Gaussian profile

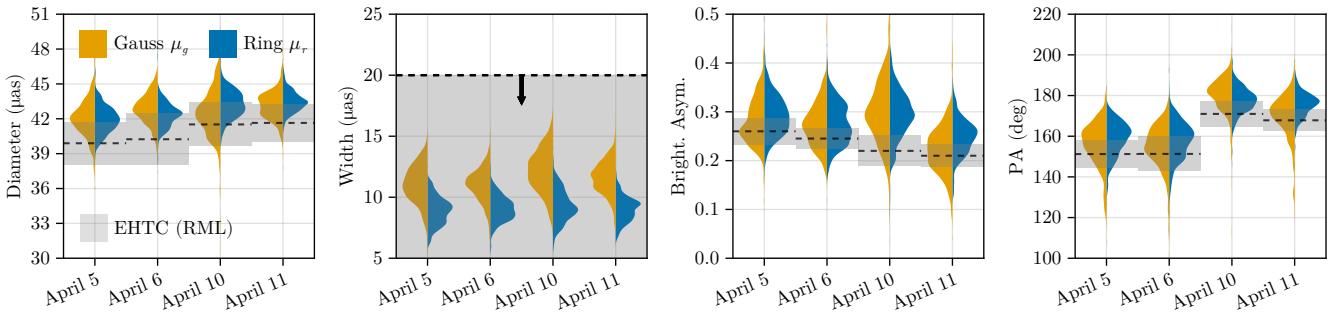


Figure 6. Marginal distribution of the extracted ring parameters from the M87* LO band data from the Gaussian (orange) and ring (blue) a priori image model. Overall, the reconstructed parameters are similar across the two models, but the ring prior produces tighter measurements of all of the ring quantities. The EHT values are from Paper IV Table 7, and combine the two RML imaging pipelines. The black dotted lines show the mean and gray band the standard deviation, except for the ring width, where the dotted line and gray region show the upper limit reported by the EHT (Paper I). In general, we find that HIBI is consistent with the reported EHT values.

recovered the correct width for the 10, 15, and 20 μ as cases. However, for the 5 μ as synthetic data, the Gaussian profile reconstruction was biased high, providing a measurement of $8.0^{+0.7}_{-0.8}$ μ as, rather than the actual value of 5.7 μ as measured by VIDA. However, the ring profile μ_r accurately measured the ring width for all cases, likely due to the ring profile being a better match to the true image. Given that the synthetic data tests demonstrate the reliability of the measurement of the width of the ring profile, we believe that it is an accurate measurement of the width of M87* in 2017. Combining the measurements of M87* across all 4 days and two frequency bands, the total width estimate is 9.3 ± 1.3 μ as.

Turning to the estimation of the parameters of the ring profile, Table 1 reports the parameter estimates for the peak radius, characteristic radius, and the inner and outer power-law slope. Interestingly, we find that the inner slope is quite large, typically ranging from 10–40, while the outer power-law slope is closer to between 1.5 and 2.5 for all 4 days. This suggests that the ring emission tends to peak closer to the inner edge of the shadow and declines more gradually.

5. Application to VLBA Data

Although the algorithms developed in this paper focus on black hole imaging with the EHT, they apply to any VLBI array. To demonstrate HIBI's capabilities, we now consider VLBA AGN data. Specifically, we analyzed the AGN OJ 287 observations on November 1 at 43 and 86 GHz and on November 2 at 15 GHz. For the 43 and 86 GHz data, we utilize public data from the Boston University BEAM-ME project (S. Jorstad & A. Marscher 2016; S. G. Jorstad et al. 2017; Z. R. Weaver et al. 2022). For the 15 GHz data, we use public data from the MOJAVE AGN monitoring program (M. L. Lister et al. 2018).

For the 15 GHz data, a 144×112 raster with an FOV of 7.8×6 mas was chosen. For the 43 GHz data, a 192×96 raster with a 4×2 mas raster was used. For the 86 GHz data, a 96×96 raster with an FOV of 1×1 mas was used. The raster size was chosen for efficiency when computing the NFFT. For the total flux of each image, we use the total flux reported from each experiment: $F_{230} = 7.17$ Jy, $F_{86} = 6.50$ Jy, and $F_{86\text{GHz}} = 4.58$ Jy. Similarly to the above, a first-order GMRF prior is assumed for each dataset, with the same priors on the hyperparameters as mentioned in Section 3.4. For the a priori image structure of the GMRF, we used

$$\mu_{\text{OJ}}(\mathbf{x}|f_b) = (1 - f_b)\mu_{\text{core}}(\mathbf{x}) + f_b/K. \quad (43)$$

Table 1
Ring Profile Estimates LO Band (68% CI)

Param.	April 5	April 6	April 10	April 11
$2r_{\text{peak}} (\mu\text{as})$	$34.4^{+1.7}_{-1.8}$	$34.6^{+1.5}_{-1.6}$	$36.0^{+2.0}_{-1.8}$	$36.4^{+1.7}_{-1.5}$
$2r_0 (\mu\text{as})$	$32.6^{+1.4}_{-1.6}$	$32.8^{+1.2}_{-1.5}$	$34.2^{+1.8}_{-1.7}$	$34.6^{+1.4}_{-1.3}$
α_{in}	$58.8^{+27.6}_{-29.0}$	$62.8^{+23.9}_{-31.8}$	$55.4^{+30.5}_{-31.5}$	$54.7^{+31.5}_{-30.9}$
α_{out}	$1.7^{+0.4}_{-0.3}$	$1.6^{+0.4}_{-0.3}$	$2.3^{+0.8}_{-0.6}$	$2.5^{+0.6}_{-0.5}$
VIDA est.				
$d (\mu\text{as})$	$41.9^{+1.1}_{-1.1}$	$42.4^{+1.0}_{-1.2}$	$43.2^{+1.3}_{-1.1}$	$43.4^{+1.0}_{-0.9}$
$w (\mu\text{as})$	$8.9^{+1.1}_{-1.2}$	$9.0^{+1.1}_{-0.9}$	$9.1^{+1.6}_{-1.2}$	$9.1^{+0.8}_{-1.1}$
A	$0.29^{+0.04}_{-0.04}$	$0.27^{+0.05}_{-0.04}$	$0.29^{+0.05}_{-0.05}$	$0.25^{+0.03}_{-0.04}$
$\xi (\text{deg})$	$160.2^{+6.6}_{-9.9}$	$159.0^{+7.6}_{-10.5}$	$177.9^{+6.1}_{-5.5}$	$176.3^{+4.7}_{-6.6}$

Here $\mu_{\text{core}}(\mathbf{x})$ is a Gaussian profile with FWHM 400 μ as, f_b is the fraction of the flux in the background component, and K is the total number of image pixels. This profile corresponds to a core image with a potential constant background whose fractional flux, f_b , is also a free parameter fit during sampling. The point of the Gaussian core component is to encourage OJ 287's core to reside near the phase center of the image, which the data does not constrain because we refit for instrumental gain phases as mentioned below. Finally, we used the exponential map Equation (16) to map from r_{ij} to δ_{ij} due to the potentially high-dynamic range image structure.

For simplicity, we used the publicly available calibrated data. However, we refit the gain solutions to correct for any residual calibration errors. Given that we used well-calibrated data, we averaged each dataset over a “scan” using the `scan_average` function from `Pyehtim.jl` and added 0.2% systematic error. Similarly to the M87* analysis, we fit the gain phases for each station using a von Mises prior with the concentration parameter π^{-2} . We used a normal distribution for the log-gain amplitudes with a mean of zero and a standard deviation of 0.25 for all stations. The fitting and sampling strategy is identical to Section 3.3.

Figure 7 shows the resulting fits. The first column shows the native resolution image reconstruction results at 15, 43, and 86 GHz from top to bottom. For all frequencies, a qualitatively similar core structure consisting of 2–3 pointlike objects was recovered using HIBI. The third column shows the image reconstructions blurred to match the CLEAN images (fourth column), demonstrating mostly consistent results, although the

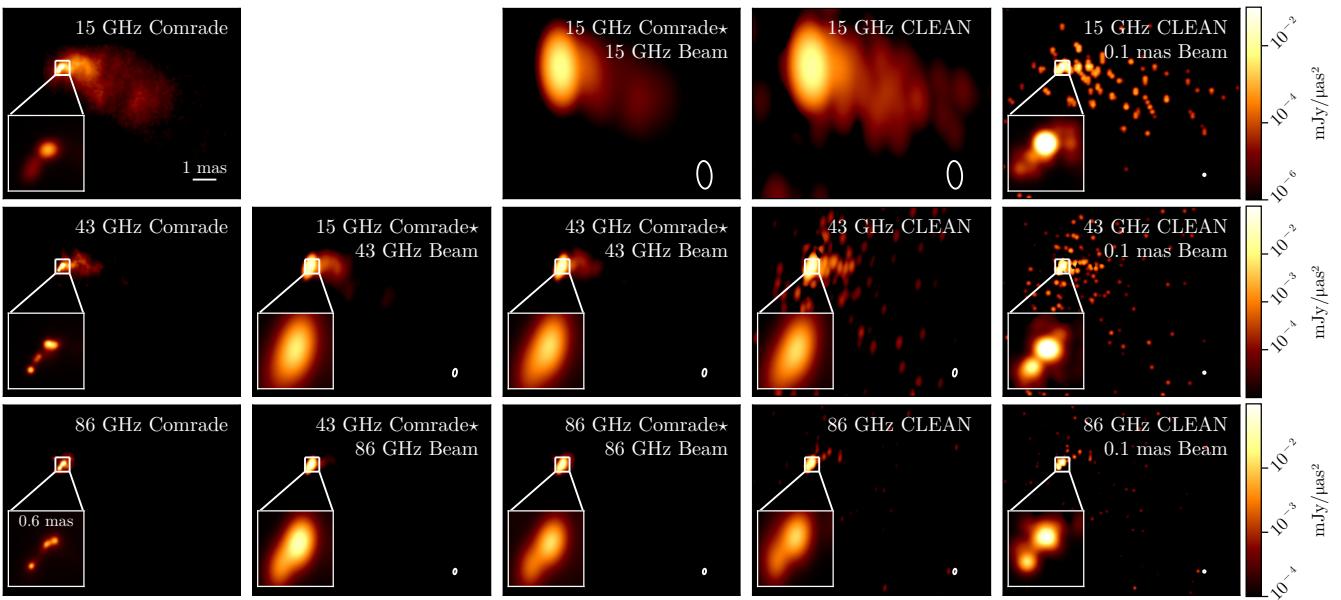


Figure 7. The mean Comrade reconstructions of OJ 287 on 2016 November 1 and 2 at 15 GHz, 43 GHz, and 86 GHz compared to CLEAN images. The first (leftmost) column shows the Comrade reconstructions results at its native resolution. The second column shows the Comrade reconstructions blurred to match the beam at the next higher frequency. For example, the second row is the 15 GHz reconstruction blurred to match the resolution of the 43 GHz CLEAN reconstructions. The third column shows the Comrade reconstructions blurred to match the resolution of the native CLEAN reconstructions shown in the fourth column. Overall, we find that Comrade’s super resolution of the core is supported by the image reconstructions at higher frequencies.

HIBI images from Comrade images are less noisy. The second column of the figure shows the 15(43) GHz HIBI reconstruction, blurred to match the 43(86) GHz CLEAN resolution in the second(third) rows. For all three frequencies, we see that the HIBI reconstruction achieves significantly improved resolution compared to the CLEAN reconstruction blurred with the nominal beam. The improved resolution observed at 15 and 43 GHz is similar to the observed structure at higher frequencies and supports the conclusion that the improved resolution is faithful to the true source structure. Note, however, that the improved resolution is significantly higher near the core region of the image, where we have the largest effective signal-to-noise ratio (SNR) of the image. For the more extended emission observed at 15 GHz, the resolution is more similar to CLEAN reconstructions. The result demonstrates how effective resolution depends on the SNR of various image components (A. P. Lobanov 2005).

6. Summary

This paper presents a new perspective on VLBI imaging, casting it as a hierarchical Bayesian inference problem. HIBI separates imaging into two distinct types of parameters: the latent parameters, which are the image itself, and the higher-level or structural parameters, sometimes referred to as hyperparameters. Within this hierarchical framework, HIBI can reconstruct both types of parameters and estimate their joint uncertainty.

To demonstrate HIBI, we implemented a scalable version using Markov random fields and demonstrated how physical information can be incorporated, e.g., ring-like morphology, into the model. The HIBI algorithm itself is available in the Comrade VLBI imaging software, and examples can be found in documentation.¹⁵

To test the applicability of HIBI to various image structures, we repeated the synthetic data tests from Paper IV. In all cases,

after marginalizing over the gain and hyperparameter parameters, we found that HIBI reconstructs the correct image morphology and quantitative features for EHT 2017 coverage. We then applied HIBI to the 2017 EHT data, reproducing the original EHT results, and measuring M87*’s diameter, width, brightness asymmetry, and position angle.

While the HIBI posterior can be used to construct images marginalized over different structural and instrumental assumptions, we can also marginalize over image fluctuations to estimate more fundamental parameters. In Section 4 we demonstrated how we can use HIBI to directly measure properties of the ring, such as its radius and profile shape, after marginalizing over different image fluctuations. We anticipate that this approach to feature extraction will be highly beneficial for science cases where the rough morphology or physics is known a priori. For example, the demographics of marginally resolved black hole shadow candidates from future missions like the ngEHT (S. S. Doeleman et al. 2023; M. D. Johnson et al. 2023) and the Black Hole Explorer (M. D. Johnson et al. 2024), could utilize the ring modeling in this paper to measure the radius of the shadow of other black holes. This approach to imaging could also be expanded to include more physically interesting models, including semianalytic accretion models (e.g., A. E. Broderick et al. 2016; D. C. M. Palumbo et al. 2022; D. Chang et al. 2024) where the imaging component would focus on modeling the stochastic turbulence in plasmas around black holes.

In Section 5 we applied HIBI to non-EHT data, imaging the well-known blazar OJ 287 and demonstrating the reliability of HIBI’s super-resolution by comparing the reconstructions across frequency for near-simultaneous observations. We found that HIBI dramatically improved the image reconstruction resolution in the bright core region, and the core morphology was confirmed upon examining higher-frequency data.

¹⁵ <https://ptiede.github.io/Comrade.jl/stable/>

Acknowledgments

The authors thank Sara Issaoun, Iniyan Natarajan, Aleksander (Sasha) Plavin, and Andrew Chael for informative discussions during paper writing. This publication was supported by the AWS Impact Computing Project at the Harvard Data Science Initiative, award No. A61166. We acknowledge financial support from the National Science Foundation (AST-2307887). This publication is funded in part by the Gordon and Betty Moore Foundation, grant GBMF12987. This work was supported by the Black Hole Initiative, which is funded by grants from the John Templeton Foundation (grant No. 62286) and the Gordon and Betty Moore Foundation (grant GBMF-8273)—although the opinions expressed in this work are those of the author(s) and do not necessarily reflect the views of these Foundations. This material is based upon work supported in part by the National Science Foundation under grant No. 2346519. This work was also supported in part by the Alan Turing Institute, and the U.S. Department of Energy. We gratefully acknowledge the support of the NSF-Simons AI-Institute for the Sky (SkAI) via grants NSF AST-2421845 and Simons Foundation MPS-AI-00010513. The work was supported in part by the Deutsche Forschungsgemeinschaft (DFG-528753982) and the European Union's Horizon Europe research and innovation funding programme under grant agreement number 101135064. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them. This research has made use of data from the MOJAVE database that is maintained by the MOJAVE team (M. L. Lister et al. 2018). This study makes use of VLBA data from the VLBA-BU Blazar Monitoring Program (BEAM-ME and VLBA-BU-BLAZAR; <http://www.bu.edu/blazars/BEAM-ME.html>), funded by NASA through the Fermi Guest Investigator Program. The VLBA is an instrument of the National Radio Astronomy Observatory. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated by Associated Universities, Inc.

Facilities: EHT (ALMA, APEX, IRAM, JCMT, LMT, SMA, SMT, SPT), VLBA.

Software: Comrade (P. Tiede 2022), AdvancedHMC.jl (K. Xu et al. 2020), eht-imaging (A. A. Chael et al. 2016, 2018), Enzyme (W. Moses & V. Churavy 2020; W. S. Moses et al. 2021, 2022), Julia (J. Bezanson et al. 2018), NFFT.jl (T. Knopp et al. 2023), Makie.jl (S. Danisch & J. Krumbiegel 2021), Optimization.jl (V. K. Dixit & C. Rackauckas 2023), Optimisers.jl (<https://github.com/FluxML/Optimisers.jl>).

Appendix A Imaging Parameters

Table 2 shows the parameters used for the image reconstructions for the various sets of data considered in this paper.

Appendix B Impact of Correlation in the Image Prior

To explore the impact of including correlation in the image prior, this section will re-image the April 11 lo-band data using two commonly used priors in Bayesian imaging from A. E. Broderick et al. (2020a) and D. W. Pesce (2021). The A. E. Broderick et al. (2020a) inspired prior assumes a log-uniform prior for the pixel fluxes \mathbf{F} with lower bound A and upper bound B . This prior tends to enforce a high degree of sparsity in the data and does not have any hyperparameters other than the number of pixels in the raster and the field of view of the image. The D. W. Pesce (2021) prior assumes that \mathbf{F} is drawn from a symmetric Dirichlet distribution with probability density

$$p_{\mathcal{D}}(\mathbf{F}|\alpha) = \frac{\Gamma(cK)}{\Gamma(c)^K} \prod_{I=1}^K F_I^{c-1}, \quad (B1)$$

where c is the concentration parameter and enforces a global notion of smoothness or sparsity. For $c = 1$, this distribution corresponds to the uniform distribution on the simplex. For $c < 1$, the prior tends to promote sparsity in the image, while for $c > 1$, it tends to prefer homogeneous images. Given that it is unclear what to choose for c a priori, we included it as a hyperparameter during imaging. Similarly to the GMRF prior, we used a hyperprior that tends to prefer simple images, i.e., images with similar flux everywhere, i.e., $c > 1$. To enforce the preference for a smooth image, we use an inverse-gamma prior with $\alpha = 6$, $\beta = 9.25$. This distribution has its mode at $c \approx 1.3$ and has 10% of its probability mass $c < 1.0$.

As the log-uniform and Dirichlet priors did not include correlation parameters, we ideally should search over different FOVs and raster dimensions to find the optimal number of pixels according to some cost function, such as the Bayesian evidence or Bayesian leave-one-out cross-validation. However, in this paper, we instead fixed the values of the field of view of to $80\mu\text{as}$ since it was large enough to contain all of the flux seen in Figure 4. To select the pixel size, we followed A. E. Broderick et al. (2020a) and used a pixel size of $\sim 12\mu\text{as}$, which corresponds to a 7×7 raster. This is far below the resolution of the GMRF prior in the main text, so we also considered a raster identical to the GMRF run to demonstrate the impact of enforcing zero correlation in the image. To sample the image posterior, we employed the same procedure from Section 3.3.

The image reconstructions for the log-uniform (left), Dirichlet (middle), and GMRF (right) priors are shown in Figure 8. Overall, the qualitative image structure looked similar except for the uncorrelated priors that use a large FOV and a large number of pixels. Both lower-resolution and log-uniform priors produce similar rings in appearance, and qualitatively, as can be seen in Table 3. However, when the

Table 2
Imaging Parameters

Data	$\mu(x, y)$	$\eta(r)$	GMRF Order	Raster Size	FOV
EHT M87*/synth. (Gauss.)	μ_g (34)	η_{sp} (15)	1	64×64	$(200\mu\text{as}, 200\mu\text{as})$
EHT M87*/synth. (ring)	μ_r (41)	η_{sp} (15)	1	95×95	$(200\mu\text{as}, 200\mu\text{as})$
VLBA OJ 287 (15 GHz)	μ_{OJ} (43)	η_{sm} (16)	1	143×111	$(7.8\text{ mas}, 6\text{ mas})$
VLBA OJ 287 (43 GHz)	μ_{OJ} (43)	η_{sm} (16)	1	191×95	$(4\text{ mas}, 2\text{ mas})$
VLBA OJ 287 (86 GHz)	μ_{OJ} (43)	η_{sm} (16)	1	95×95	$(1\text{ mas}, 1\text{ mas})$

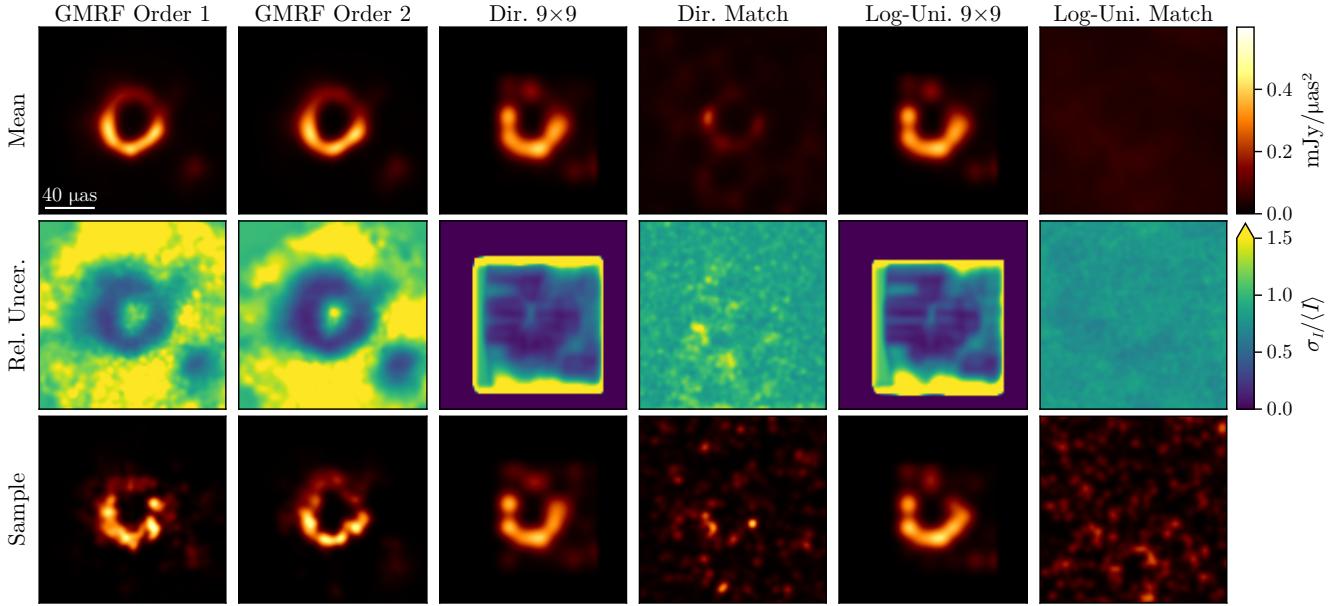


Figure 8. Comparison between three alternative Bayesian imaging priors.

Table 3
April 6 Ring Parameters

Prior	d (μas)	w (μas)	A	ξ (deg)
GMRF Order 1 (μ_g)	$41.0^{+2.3}_{-3.5}$	$11.1^{+2.0}_{-2.5}$	$0.27^{+0.09}_{-0.06}$	$155.2^{+14.6}_{-19.5}$
GMRF Order 2 (μ_g)	$41.4^{+2.2}_{-3.1}$	$11.5^{+1.9}_{-1.6}$	$0.26^{+0.07}_{-0.05}$	$158.0^{+11.6}_{-20.9}$
Dirichlet (9 × 9)	$40.3^{+1.4}_{-1.2}$	$14.8^{+1.0}_{-0.8}$	$0.29^{+0.04}_{-0.04}$	$177.3^{+3.7}_{-4.4}$
Dirichlet (64 × 64)	$36.7^{+9.8}_{-9.5}$	$7.6^{+5.8}_{-2.4}$	$0.33^{+0.17}_{-0.16}$	$128.4^{+39.2}_{-44.7}$
Log-Uniform (9 × 9)	$40.5^{+1.2}_{-1.1}$	$14.2^{+0.9}_{-0.5}$	$0.31^{+0.04}_{-0.05}$	$176.9^{+4.4}_{-3.5}$
Log-Uniform (64 × 64)	$53.4^{+6.6}_{-15.2}$	$8.2^{+27.1}_{-3.4}$	$0.33^{+0.17}_{-0.26}$	$157.0^{+146.6}_{-94.9}$

FOV and number of pixels were increased to match the GMRF values, both the Dirichlet and log-uniform prior struggled to identify a unique image structure. For the Dirichlet prior, we see that the mean image displays a ring-like feature, albeit at a low significance. For the log-uniform prior, no significant ring-like feature can be identified in the image reconstruction, and samples from the posterior look like random noise. This random noise is due to the extreme overfitting. Since no spatial correlation is enforced, the number of degrees of freedom (64 × 64) in the image is much greater than the size of the data (~200 visibilities). As a result, the data are not very informative relative to the model, and the posterior starts to approach the prior distribution. If we view this model within the HIBI framework, the number of pixels could be included as a discrete hyperparameter in the model. A. E. Broderick et al. (2020a) approximately demonstrated that in such a scheme, high-resolution rasters are heavily disfavored compared to lower-resolution ones.

Appendix C EHT and GMRF Prior Sensitivity to On-sky Ring Thickness

The thickness of M87*'s ring was reported by the EHT as an upper limit of 20 μas across imaging and modeling methods (Event Horizon Telescope Collaboration et al. 2019b, 2019c). Later work (W. Lockhart & S. E. Gralla 2022a, 2022b)

analyzed the 2017 EHT data with a variety of geometric models and found that the fractional width of the ring model was ≤ 0.25 , which, for a ~ 40 μas ring, gives a width of ~ 10 μas. However, the authors noted that the fractional width bound could increase to 0.3–0.4 or 12–16 μas when using a different approximate closure likelihood.

Using Comrade and the GMRF prior with a Gaussian profile, we measured the thickness of the ring to be 11.7 ± 0.7 μas averaged over the 4 days and two frequency bands. Meanwhile, for the ring profile, we found a similar width of 11.0 ± 0.5 μas. Both of the measurements are consistent with Event Horizon Telescope Collaboration et al. (2019c), Event Horizon Telescope Collaboration et al. (2019b), and W. Lockhart & S. E. Gralla (2022a). However, due to the finite resolution of the EHT, the width measurements may be biased. Namely, the EHT is not sensitive to changes in the ring width below a certain limit. To test whether HIBI's ring thickness measurement is reliable for the scales measured by EHT, we imaged rings with several different thicknesses. Namely, we used the same azimuthally symmetric delta ring from Section 3, but blurred using different Gaussian kernels with FWHM of 5, 10, 15, and 20 μas. We then imaged the said synthetic data using both the Gaussian and ring profiles.

The results are shown in Figure 9. We found that HIBI recovers the true diameter and width of the rings for the 10, 15, and 20 μas rings for both μ_g and μ_r . However, for the 5 μas ring, in Figure 9, the Gaussian profile reconstructions have widths that are slightly too large compared to the ground truth measurement 5.1–10 μas. However, the ring profile μ_r recovers the true width of the ring. To ensure that μ_g was not biased by the rasterization resolution, we reran the 5 μas synthetic data with a higher-resolution raster and found equivalent results. Therefore, this bias is likely due to a combination of the EHT's limited resolution and the fact that the Gaussian profile does not accurately match the true value.

For M87* we found that both the ring and Gaussian profile provided identical width measurements. Additionally, M87*'s measured width value more closely matches the 10 μas ring

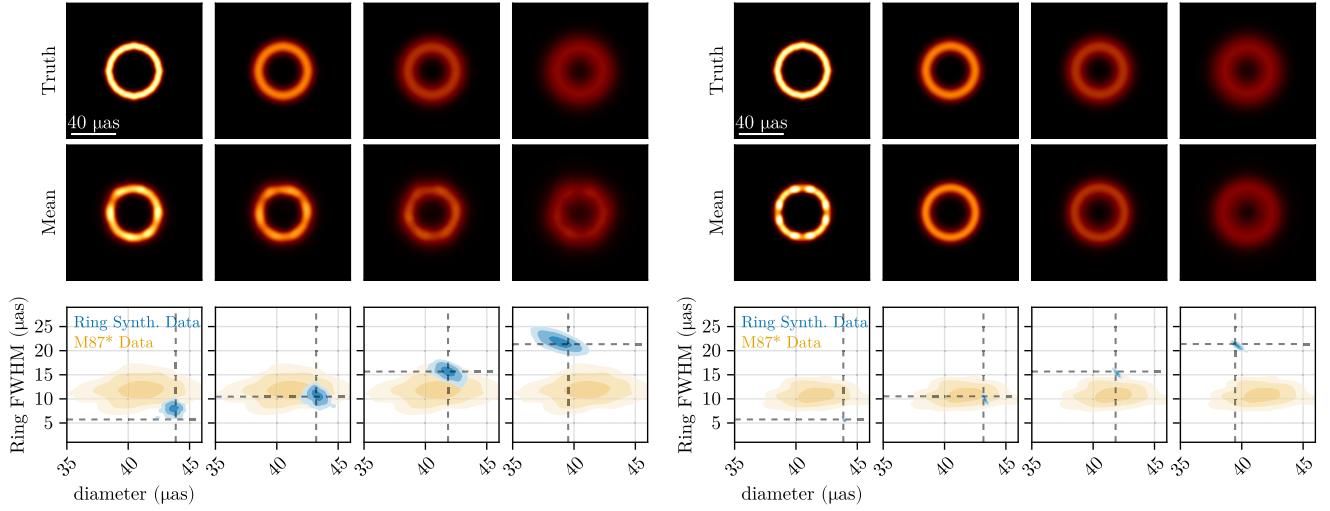


Figure 9. Testing the sensitivity of the EHT array to different ring thicknesses for the Gaussian (left group) and ring (right group) mean prior. For each group, the columns show rings with intrinsic widths 5 μas , 10 μas , 15 μas , and 20 μas from left to right. The first row is the ground truth, the middle row is the mean reconstruction using the first-order GMRF, and the bottom row shows the posterior density of the ring width versus diameter, with contours denoting the 50%, 90%, and 95% credible intervals. The measured M87* width and diameter averaged over the observing days and frequency are shown by the orange region.

data, which was recovered for both profiles considered. Therefore, we believe that HIBI's estimate of the ring width of M87* is reliable.

Appendix D Robustness of the a priori Ring Profile to Different Structures

In this appendix, we demonstrate that the EHT data is informative enough to produce non-ring images even when a ring model is used for $\mu(x, y)$. This is complementary to the synthetic data tests in Paper IV, which demonstrate that the EHT can reconstruct various image structures with minimal assumptions about the source morphology.

For the data, we used the same synthetic data described in Section 3. However, for the a priori profile $\mu(x, y)$, we used the ring profile Section 4. This profile consists of a ring, although

the ring radius and the inner and outer power-law slopes are included as parameters in the model. Furthermore, the inner power-law prior has support around zero, meaning that a flat disk is in the space of prior images.

The imaging results are shown in Figure 10. For the ring synthetic data, we observed that the ring prior significantly reduces the uncertainty in the image reconstructions, because the profile μ is a near-perfect model of the image. The disk reconstruction also improved considerably, as it is included in the family of $\mu(x, y)$ profiles. Encouragingly, even when a ring-like structure was assumed, the double synthetic data were constructed correctly. This result implies that the EHT data are informative enough to create non-ring images even when a ring prior is assumed when using HIBI. Therefore, we conclude that the ring profile does not rigidly enforce ring-like images when analyzing the 2017 EHT data.

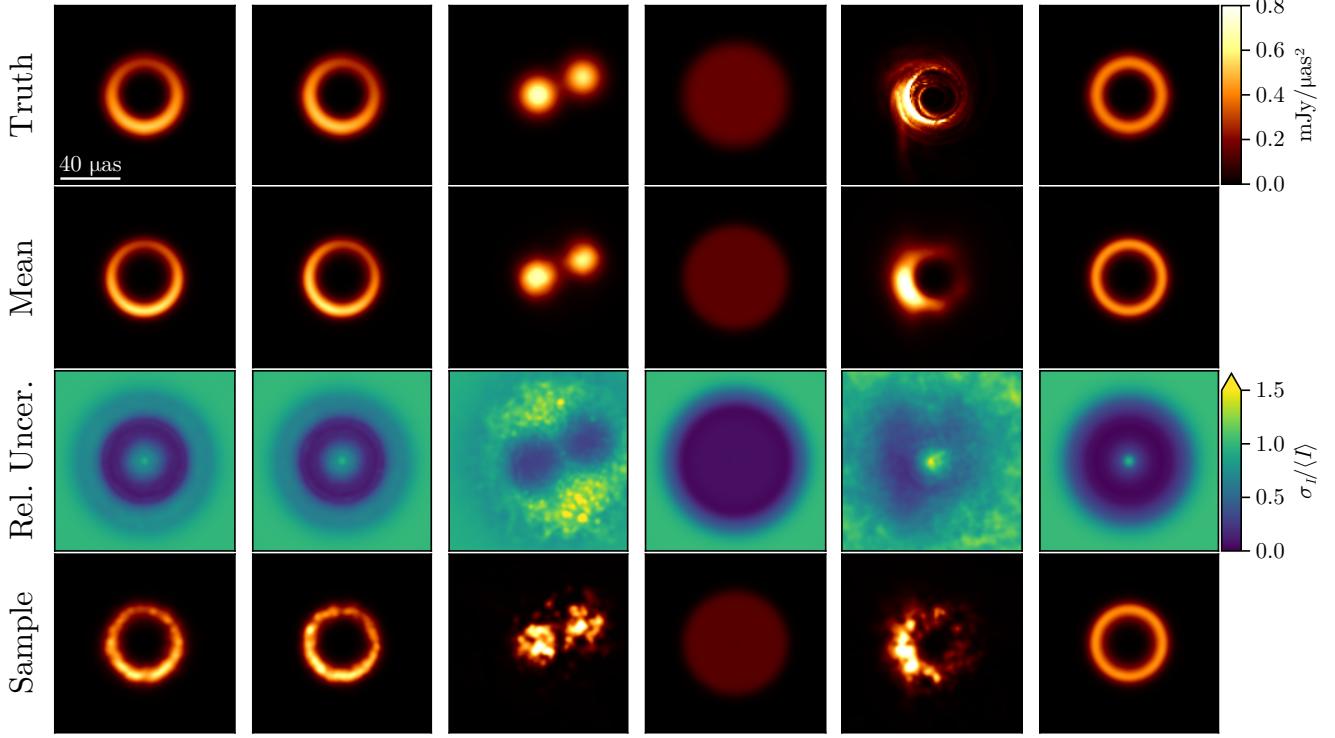


Figure 10. Reconstructions of synthetic data from Section 3, but using a ring prior on the mean image of the MRF.

Appendix E Robustness of M87* Image to Different Gain Assumptions

E.1. Closure Likelihood

Fitting complex visibilities directly means that we require an instrument model to describe both atmospheric and instrumental effects. Since VLBI imaging is typically only interested in the image posterior $p(\theta|V)$, another mathematically equivalent approach is to first marginalize over the instrument model parameters s :

$$p(\theta|V) = \frac{p(\theta)}{p(V)} \int p(V|\theta, s)p(s)ds. \quad (\text{E1})$$

L. Blackburn et al. (2020) demonstrated that the marginal image posterior is equivalent to fitting closure products in the high-SNR limit when improper flat priors are assumed for the instrumental gains. This assumption can be viewed as a maximally conservative view of our a priori knowledge of the instrument. Therefore, to explore the impact of our instrument model in Section 3.4, this section considers image reconstructions using log-closure amplitudes and closure phases.

Closure phases are defined as the argument of the bispectrum of three measured visibilities:

$$\psi_{ijk} = \arg V_{ij}V_{jk}V_{ki}, \quad (\text{E2})$$

and provide the gain-invariant phase information of the set of measured visibilities. Amplitude information can be recovered by analyzing the closure amplitudes:

$$A_{ijkl} = \frac{|V_{ij}||V_{kl}|}{|V_{ik}||V_{jl}|}. \quad (\text{E3})$$

Instead, we consider the log-closure amplitudes $\mathcal{A} = \log A$. The reason for this is that log-closure amplitudes are related to log amplitudes by a linear transformation, and their errors are

better approximated by Gaussian noise than the direct closure amplitudes (A. R. Thompson et al. 2017; A. E. Broderick et al. 2020b).

For high-SNR measurements ($\gtrsim 3$), the closure likelihood is approximately a correlated Gaussian (A. R. Thompson et al. 2017; A. E. Broderick et al. 2020b; L. Blackburn et al. 2020). For the closure phases, we used the approximate likelihood

$$\mathcal{L}(\psi|\theta) \propto \exp\left(-\frac{1}{2}\Delta\mathbf{c}^\dagger \Sigma_\psi \Delta\mathbf{c}\right), \quad (\text{E4})$$

where $\Delta\mathbf{c}_I = e^{i\psi_I} - e^{i\psi_{\theta,I}}$, and $\psi_{\theta,I}$ is the image model, I th closure phase. Note that the normalization of this likelihood is not easily calculable; however, it amounts to a scaling constant, so it does not affect posterior estimation.

Similarly, for the log-closure amplitudes, we make a similar high-SNR Gaussian approximation for the likelihood:

$$\mathcal{L}(\mathcal{A}|\theta) \propto \exp\left(-\frac{1}{2}\Delta\mathcal{A}^\dagger \Sigma_\mathcal{A} \Delta\mathcal{A}\right). \quad (\text{E5})$$

Here, $\Delta\mathcal{A}_I = \mathcal{A}_I - \mathcal{A}_{\theta,I}$, and $\mathcal{A}_{\theta,I}$ is the image model, I th log-closure amplitude.

To construct the set of closure phases and log-closure amplitudes, we follow L. Blackburn et al. (2020). Let L_A and L_ψ denote the linear maps from log amplitudes/phases to log-closure amplitudes/phases. We will assume that L is constructed such that both are full-rank matrices. Then we have that

$$\Sigma_{\mathcal{A},\psi} = L_{\mathcal{A},\psi} \frac{\sigma^2}{|V|} L_{\mathcal{A},\psi}^T. \quad (\text{E6})$$

The total closure likelihood is therefore given by

$$p(\theta|V) \approx p(\theta|\psi, \mathcal{A}) \propto \mathcal{L}(\psi|\theta) \mathcal{L}(\mathcal{A}|\theta) p(\theta). \quad (\text{E7})$$

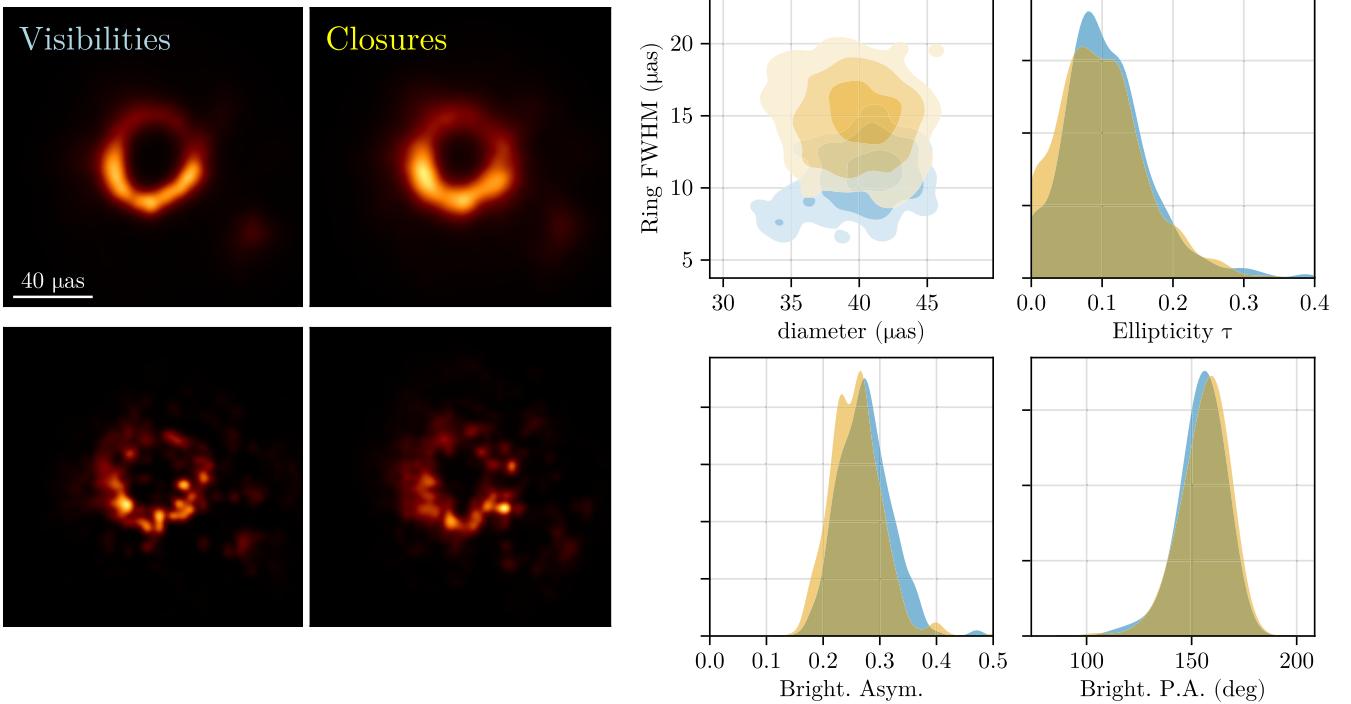


Figure 11. Comparison between complex visibility (blue) and closure-only (orange) reconstructions using the April 6 LO band data. The top row of the images shows the mean image, while the bottom are random draws from their respective posterior. The right grid of images show the parameter estimates computed using the VIDA template matching approach described in Section 3

To compare the impact of our gain priors on the images, we compare the results in Section 3.4 with closure-only fits. To simplify the comparison, we only show the results for the April 11 LO band. However, we found similar results on other days. For closure-only reconstructions, we flag any closures whose $\text{SNR} < 3$ to remove closures that violate our Gaussian likelihood assumption. Otherwise, we use identical fitting and sampling procedures.

The imaging results are shown in Figure 11. We measured similar diameters, ellipticities, brightness asymmetries (A), and position angles (ξ) for both complex visibilities and closures. The major difference in the reconstruction is the width or resolution of the ring. That is, the complex visibility fits appear sharper, and the inferred ring width posterior is shifted slightly downward. The difference is likely explained by the loss of information when moving to closures, and the flagging of low-SNR points, which tend to remove longer baselines and baselines near M87*'s amplitude null (Paper VI). Regardless, this result implies that the measurements of M87*'s properties in the main text are not significantly biased by the assumed instrument model.

Appendix F Eigenvalues of the First-order GMRF

In this section, we prove Equation (24). The precision matrix Q_1 for the unit variance first-order GMRF is

$$Q_1 = \frac{1}{\rho^2} \mathbb{1} + \mathbf{G}. \quad (\text{F1})$$

To compute its eigenvalues, we first note that \mathbf{G} is given by the Kronecker sum

$$\mathbf{G} = \mathbf{D}_x \otimes \mathbb{1} + \mathbb{1} \otimes \mathbf{D}_y, \quad (\text{F2})$$

where \mathbf{D} is the tridiagonal matrix,

$$\mathbf{D}_{ij} = \begin{cases} 2 & i = j, \\ -1 & i = j \pm 1, \\ 0 & \text{otherwise,} \end{cases} \quad (\text{F3})$$

the subscripts denote the number of pixels in the x - and y -directions, and \otimes denotes the tensor product.

Given the form of G , its eigenvalues are just the sum of the eigenvalues of D_x and D_y , i.e., the eigenvalues of G are given by

$$\lambda_{nm}^G = \lambda_n + \lambda_m \quad 1 \leq n \leq N_x, 1 \leq m \leq N_y. \quad (\text{F4})$$

Therefore, to find the spectrum of G , we need to find the eigenvalues for D . First, note that the eigenvalues of the $N \times N$ matrix

$$\begin{pmatrix} 0 & -1 & & & \\ -1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & & \ddots & \end{pmatrix}, \quad (\text{F5})$$

are given by

$$2 \cos\left(\pi \frac{n}{N+1}\right), \quad 1 \leq n \leq N. \quad (\text{F6})$$

This equation can be derived by noting that the characteristic polynomial of Equation (F5) defines the recurrence relation,

$$T_n(\lambda) = \lambda T_{n-1}(\lambda) - T_{n-2}(\lambda), \quad (\text{F7})$$

where $T_1 = \lambda$ and we set $T_0 = 1$. Let $x = \lambda/2$, then Equation (F7) becomes

$$\begin{aligned} T_0(x) &= 1 \\ T_1(x) &= 2x \\ T_n(x) &= 2xT_{n-1}(x) - T_{n-2}(x), \end{aligned} \quad (\text{F8})$$

which are the recurrence relations for Chebyshev polynomials of the second kind, whose roots give Equation (F6).

Given Equation (F6), the eigenvalues for D are given by

$$\lambda_n = 2 + 2 \cos\left(\pi \frac{n}{N+1}\right), \quad (\text{F9})$$

and thus eigenvalues for Q_1 are

$$\lambda_{nm}^{Q_1} = \rho^{-2} + 4 + 2 \cos\left(\pi \frac{n}{N_x+1}\right) + 2 \cos\left(\pi \frac{m}{N_y+1}\right), \quad (\text{F10})$$

whose product gives Equation (24).

Note that using a similar computation, one can show that the eigenvectors of D , are given by

$$(\mathbf{v}_n)_j \propto \sin\left(\pi \frac{n j}{N+1}\right), \quad (\text{F11})$$

for the j th component of the n th eigenvector. This implies that the type 1 discrete sine transform diagonalizes the matrix D .

Appendix G Extending the GMRF Priors

While the GMRF prior used in the main text is flexible, it assumes that the noise structure of the image is conditionally dependent only on the nearest neighbors and that the noise structure is Gaussian. This appendix demonstrates how you can relax both constraints.

G.1. Non-Gaussian Markov Random Fields

Recall that the general form of a GMRF is given by Equation (20). We can easily extend this to non-GRFs, such as the exponential distribution,

$$p_{\text{exp}}(\mathbf{r}|\boldsymbol{\mu}, \mathbf{Q}) = \sqrt{\frac{\det \mathbf{Q}}{2\pi^K}} \exp(-K\sqrt{\mathbf{r}^T \mathbf{Q} \mathbf{r}}) \quad (\text{G1})$$

and T-distribution,

$$\begin{aligned} p_{\text{T}}(\mathbf{r}|\boldsymbol{\mu}, \mathbf{Q}, s) &= \frac{\Gamma(\nu/2 + K/2)}{\Gamma(\nu/2)} \sqrt{\frac{\det \mathbf{Q}}{(\pi\nu)^K}} \left[1 + \frac{1}{\nu} \mathbf{r}^T \mathbf{Q} \mathbf{r}\right]^{-(\nu+K)/2}, \end{aligned} \quad (\text{G2})$$

where $\nu \geq 1$ denotes the degrees of freedom of the process.

The net effect of the exponential and T-distribution is that the tails of their density function are heavier than those of the GMRF. These heavier tails may be more applicable for sources that have some outliers, such as pointlike features, in the image.

G.2. Higher-order Markov Random Fields

To generalize any Markov Random fields, including the exponential and T-distribution MRF, we need to alter the structure of the precision matrix \mathbf{Q} . Following F. Lindgren et al. (2011), we can easily extend the first-order process by changing the precision matrix to

$$\mathbf{Q}_n = \tau(\mathbf{I} + \kappa \mathbf{G})^n, \quad (\text{G3})$$

where n is a positive integer. In the spectral domain and assuming Gaussian noise and large raster, this process has the power spectrum (F. Lindgren et al. 2011)

$$S(\mathbf{k}) d^2\mathbf{k} \propto \frac{1}{(\kappa^2 + k^2)^n}, \quad (\text{G4})$$

and is a discrete approximation of the integer-order 2D *Matern* Gaussian process, which has autocorrelation function:

$$C_\alpha(r) = \frac{\sigma^2}{2^{\alpha-1}\Gamma(\alpha)} \left(\sqrt{8\alpha} \frac{r}{\rho} \right)^\alpha K_\alpha \left(\sqrt{8\alpha} \frac{r}{\rho} \right). \quad (\text{G5})$$

Here, $\alpha = n - 1$ and σ and ρ are related to τ and κ by

$$\begin{aligned} \kappa &= \sqrt{8\alpha}/\rho \\ \sigma &= \sqrt{\frac{1}{\tau\kappa^\alpha} \frac{\Gamma(\alpha)}{\Gamma(\alpha+1)4\pi}}. \end{aligned} \quad (\text{G6})$$

Increasing the order of the GMRF increases the power-law slope and, thus, the smoothness of the image reconstruction. Note that the first-order GMRF is not an example of the Matern process but remains a valid random field on a grid.

The higher-order random field can be easily included in the exponential and T-distribution random fields by setting $\mathbf{Q} = \mathbf{Q}_n$. The impact of changing the base distribution and order of the Markov random field is shown in Figure 12 for the 2017 M87 observations on April 6, LO band. In general, Figure 12 demonstrates that the mean reconstructions are nearly identical across the different priors. The fractional uncertainty maps are also very similar across the different priors and orders. However, the second-order MRF reconstructions tend to be smoother, which is expected due to its steeper power spectrum compared to the first-order process.

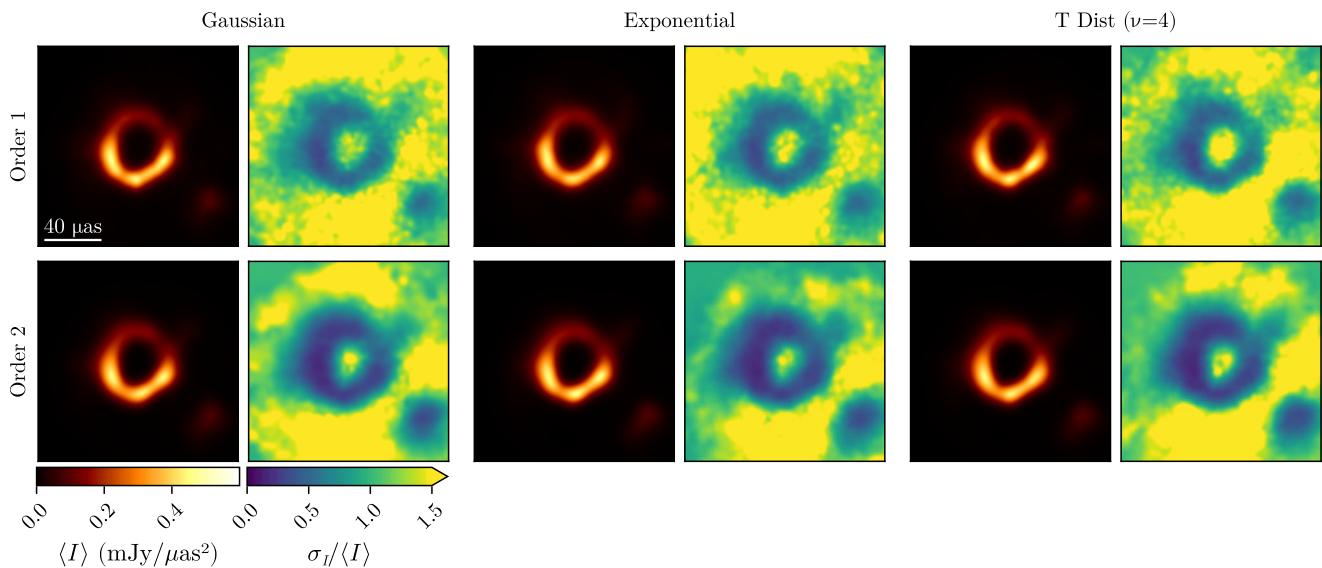


Figure 12. Imaging results on April 6 LO band using different MRF priors. The left group shows the results for a Gaussian base distribution, the middle shows an exponential distribution, and the right shows a T-distribution with degrees of freedom $\nu = 4$. The top row shows the results for the first-order MRF, and the bottom row shows the second-order MRF.

ORCID iDs

Paul Tiede <https://orcid.org/0000-0003-3826-5648>
 William Moses <https://orcid.org/0000-0003-2627-0642>
 Valentin Churavy <https://orcid.org/0000-0002-9033-165X>
 Michael D. Johnson <https://orcid.org/0000-0002-4120-3029>
 Dominic W. Pesce <https://orcid.org/0000-0002-5278-9221>
 Lindy Blackburn <https://orcid.org/0000-0002-9030-642X>
 Peter Galison <https://orcid.org/0000-0002-6429-3872>

References

Arras, P., Frank, P., Haim, P., et al. 2022, *NatAs*, **6**, 259
 Arras, P., Frank, P., Leike, R., Westermann, R., & Enßlin, T. A. 2019, *A&A*, **627**, A134
 Besag, J. 1981, *Journal of the Royal Statistical Society: Series B (Methodological)*, **43**, 302
 Bezancon, J., Chen, J., Chung, B., et al. 2018, *Proc. ACM Program. Lang.*, **2**, 120
 Blackburn, L., Pesce, D. W., Johnson, M. D., et al. 2020, *ApJ*, **894**, 31
 Bouman, K. L., Johnson, M. D., Zoran, D., et al. 2016, in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (IEEE)
 Broderick, A. E., Fish, V. L., Johnson, M. D., et al. 2016, *ApJ*, **820**, 137
 Broderick, A. E., Gold, R., Karami, M., et al. 2020b, *ApJ*, **897**, 139
 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*, **988**, 9
 Carilli, C. L., & Thyagarajan, N. 2022, *ApJ*, **924**, 125
 Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018, *ApJ*, **857**, 23
 Chael, A. A., Johnson, M. D., Narayan, R., et al. 2016, *ApJ*, **829**, 11
 Chang, D., Tiede, P., Johnson, M., & Palumbo, D. 2024, *IAUGA*, **32**, 1132
 Cornwell, T. J., & Wilkinson, P. N. 1981, *MNRAS*, **196**, 1067
 Danisch, S., & Krumbeigel, J. 2021, *JOSS*, **6**, 3349
 de Dios, J.-A. M., & Mezura-Montes, E. 2022, *JOSS*, **7**, 4723
 Dixit, V. K., & Rackauckas, C. 2023, Optimization.jl: A Unified Optimization Package, v3.12.1, Zenodo, doi: [10.5281/zenodo.7738525](https://doi.org/10.5281/zenodo.7738525)
 Doeleman, S. S., Barrett, J., Blackburn, L., et al. 2023, *Galax*, **11**, 107
 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**, L6
 Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019c, *ApJL*, **875**, L4
 Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019d, *ApJL*, **875**, L3
 Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019e, *ApJL*, **875**, L5
 Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2022, *ApJL*, **930**, L15
 Fuglstad, G.-A., Simpson, D., Lindgren, F., & Rue, H. 2019, *JASA*, **114**, 445
 Gelman, A., & Yao, Y. 2020, *JPhG*, **48**, 014002
 Hoffman, M. D., & Gelman, A. 2014, *JMLR*, **15**, 1593, <https://doi.org/10.5555/2627435.2638586>
 Issaoun, S., Wielgus, M., Jorstad, S., et al. 2022, *ApJ*, **934**, 145
 Janssen, M., Falcke, H., Kadler, M., et al. 2021, *NatAs*, **5**, 1017
 Johnson, M. D., Akiyama, K., Baturin, R., et al. 2024, *SPIE*, **13092**, 130922D
 Johnson, M. D., Akiyama, K., Blackburn, L., et al. 2023, *Galax*, **11**, 61
 Jorstad, S., & Marscher, A. 2016, *Galax*, **4**, 47
 Jorstad, S., Wielgus, M., Lico, R., et al. 2023, *ApJ*, **943**, 170
 Jorstad, S. G., Marscher, A. P., Morozova, D. A., et al. 2017, *ApJ*, **846**, 98
 Kardashev, N. S., Khartov, V. V., Abramov, V. V., et al. 2013, *ARep*, **57**, 153
 Kim, J.-S., Nikonorov, A. S., Roth, J., et al. 2024, *A&A*, **690**, A129
 Kim, J.-Y., Krichbaum, T. P., Broderick, A. E., et al. 2020, *A&A*, **640**, A69
 Knollmüller, J., & Enßlin, T. A. 2019, arXiv:1901.11033
 Knopp, T., Boberg, M., & Grosser, M. 2023, *SJSC*, **45**, C179
 Kuramochi, K., Akiyama, K., Ikeda, S., et al. 2018, *ApJ*, **858**, 56
 Lee, D., & Gammie, C. F. 2021, *ApJ*, **906**, 39
 Levis, A., Lee, D., Tropp, J. A., Gammie, C. F., & Bouman, K. L. 2021, in Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV), 2340
 Lindgren, F., Bolin, D., & Rue, H. 2022, *SpaSt*, **50**, 100599
 Lindgren, F., Rue, H., & Lindström, J. 2011, *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **73**, 423
 Lister, M. L., Aller, M. F., Aller, H. D., et al. 2018, *ApJS*, **234**, 12
 Lobanov, A. P. 2005, arXiv:0503225
 Lockhart, W., & Gralla, S. E. 2022a, *MNRAS*, **517**, 2462
 Lockhart, W., & Gralla, S. E. 2022b, *MNRAS*, **509**, 3643
 Martins, T. G., Simpson, D., Lindgren, F., & Rue, H. 2013, *Computational Statistics & Data Analysis*, **67**, 68
 Moses, W., & Churavy, V. 2020, in Advances in Neural Information Processing Systems, 33, ed. H. Larochelle (Curran Associates, Inc), 12472, <https://proceedings.neurips.cc/paper/2020/file/9332c513ef44b682e9347822c2e457ac-Paper.pdf>
 Moses, W. S., Churavy, V., Paehler, L., et al. 2021, in Proc. Int. Conf. High Performance Computing, Networking, Storage and Analysis, SC '21 (ACM)
 Moses, W. S., Narayanan, S. H. K., Paehler, L., et al. 2022, in Proc. Int. Conf. High Performance Computing, Networking, Storage and Analysis, SC '22 (IEEE)
 Müller, H., Mus, A., & Lobanov, A. 2023, *A&A*, **675**, A60
 Narayan, R., & Nityananda, R. 1986, *ARA&A*, **24**, 127
 Palumbo, D. C. M., Gelles, Z., Tiede, P., et al. 2022, *ApJ*, **939**, 107
 Pesce, D. W. 2021, *AJAJ*, **161**, 178
 Röder, J., Wielgus, M., Lobanov, A. P., et al. 2025, *A&A*, **695**, A233
 Rue, H., & Held, L. 2005, Gaussian Markov Random Fields: Theory and Applications (Chapman & Hall/CRC)

The Event Horizon Telescope Collaboration 2019, First M87 EHT Results: Calibrated Data, CyVerse Data Commons, doi:[10.25739/G85N-F134](https://doi.org/10.25739/G85N-F134)

Thompson, A. R., Moran, J. M., Swenson, G. W., & George, W. J. 2017, Interferometry and Synthesis in Radio Astronomy (3rd ed.; Springer)

Tiede, P. 2022, *JOSS*, **7**, 4457

Tiede, P., Broderick, A. E., & Palumbo, D. C. M. 2022a, *ApJ*, **925**, 122

Tiede, P., Broderick, A. E., Palumbo, D. C. M., & Chael, A. 2022b, *ApJ*, **940**, 182

Weaver, Z. R., Jorstad, S. G., Marscher, A. P., et al. 2022, *ApJS*, **260**, 12

Wilkinson, P. N., Readhead, A. C. S., Purcell, G. H., & Anderson, B. 1977, *Natur*, **269**, 764

Xu, K., Ge, H., Tebbutt, W., et al. 2020, in Symp. Advances in Approximate Bayesian Inference (PMLR), 1

Yao, Y., Vehtari, A., Simpson, D., & Gelman, A. 2018, *Bayesian Analysis*, **13**, 917

Zawadzki, B., Czekala, I., Loomis, R. A., et al. 2023, *PASP*, **135**, 064503