IMAGING AN EVOLVING BLACK HOLE BY LEVERAGING SHARED STRUCTURE

Yvette Y. Lin, Angela F. Gao*, Katherine L. Bouman*

Computing and Mathematical Sciences, California Institute of Technology
*denotes equal contribution

ABSTRACT

High quality black hole videos can provide key evidence of astrophysical processes that single static images cannot provide. However, reconstructing a video of a black hole is a highly ill-posed problem, requiring additional structural constraints to produce a plausible solution. Traditional structural constraints on the spatial or temporal structure are subject to human bias. In our work, we adapt recently developed techniques to solve realistic black hole video reconstruction without direct priors on the spatial or temporal structure, mitigating human bias. In particular, we solve a set of per-frame imaging inverse problems by relying on the shared structure across different underlying frames of the black hole as regularization. We encode this shared structure through a deep generative neural network, requiring that the reconstructed frames all lie within the range of this shared generator. We demonstrate our framework on a set of synthetic measurements of a simulated video of the supermassive black hole M87*, showing that we can substantially outperform both traditional and modern imaging methods and even achieve a level of superresolution in the reconstructed frames.

Index Terms— inverse problems, computational imaging, astronomical imaging, phase retrieval, interferometry

1. INTRODUCTION

Imaging the dynamics of a black hole opens a window into understanding complex black hole properties, such as how they grow and evolve. In 2019, the first images of the supermassive black hole Messier 87* (M87*) was produced by the Event Horizon Telescope (EHT) collaboration. This image demonstrated the possibilities of advancing fundamental physics through black hole images [1]. However, from this single static image, there are important properties of black holes that cannot be observed, such as understanding the jet launching and accretion processes [2, 3]. Characterizing these dynamic properties is a key goal of the next-generation Event Horizon Telescope (ngEHT). To do so requires that the ngEHT create dynamic, rather than static, black hole reconstructions in the form of black hole videos. This involves observing a black hole, namely M87*, at regular intervals over the course of a few months.

The EHT array generates measurements of black holes through *very long baseline interferometry* (VLBI). In this setting, black hole image reconstruction can be characterized by interferometric measurements $y = f(x) + \eta$ where f is the forward model that is dependent on the telescope configuration, x is the true underlying image that we are trying to reconstruct, and η is noise. The difficulty of this problem arises from a non-convex forward model and the inherent physical constraints of the EHT array. Namely, the EHT telescope array is small (i.e., 11 telescope sites in 2023) and the distance between sites is limited by the size of the earth, upper bounding the maximum image resolution. The sparsity of measurements makes this problem highly *ill-posed*. Although the ngEHT does plan to add additional telescopes, the same physical constraints will still apply to any realistic telescope array. Thus, additional structural constraints are necessary for reconstructing black hole images from VLBI measurements.

Currently, black hole image reconstruction methods all rely on defining structural constraints on the image, either through hand-crafted or data-driven priors. A key challenge affecting this choice is that direct images of black holes do not exist, making it hard to identify the optimal choice of constraints. For example, hand-crafted priors such as spatial priors (e.g. total variation [4]) or temporal-consistency priors require selecting hyperparameters, which are subject to human bias. On the other hand, accurate data-driven priors do not exist since we do not have access to direct images of black holes. While we could use data-driven priors of other image distributions (e.g. simulated black hole images), this could bias our reconstructions towards those datasets.

We aim to adapt recently developed techniques to solve black hole video reconstruction without direct priors on the spatial or temporal structure, but instead using priors on shared structure between different images of the same black hole. We adapt the framework in [5, 6] to show how this method can handle the challenging problem of black hole video reconstruction without explicit priors on the spatiotemporal structure. To do so, we solve a set of per-frame imaging inverse problems by inferring the shared structure across the true underlying images. Our method exploits the fact that we expect different images of the same black hole to share common structure, which we encode through a shared deep generative neural network. We demonstrate this method on realistic synthetic measurements of M87* and show that it outperforms other methods.

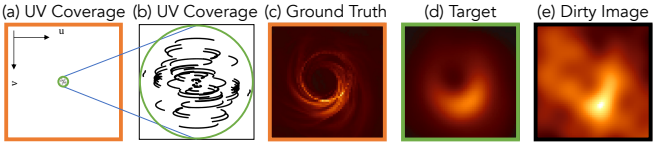


Fig. 1. **Intrinsic resolution of the M87* Event Horizon Telescope (EHT) measurements.** The EHT measures sparse spatial frequencies (i.e. Fourier components) of the image. The 2D Fourier frequency coverage of the array, with positions (u, v), is called the *UV coverage*. To reconstruct the underlying image in (c), the full frequency domain, represented by the orange square in (a), must be measured. Measuring instead only all frequency components within the green circle in (a) and (b) generates the *target* image in (d), where (b) is (a) zoomed in over 10x. The EHT measures a subset of frequencies within the green circle's interior, shown by the black samples in (b). If we use complex visibility measurements (as described in Sec. 2.1) where $y_{\text{vis.}} = Ax + \eta$ and A is a low-rank compressed sensing matrix, then we can naively recover the *dirty image* in (e), computed by $A^H y_{\text{vis.}}$.

2. VERY LONG BASELINE INTERFEROMETRY

2.1. VLBI measurement process

In VLBI, we measure a single 2D spatial Fourier frequency of the image x for each pair of telescopes a and b at time t. This measurement is called the *complex visibility* $F^t_{a,b}(x)$. This results in $\binom{S_t}{2}$ measurements for S_t observing telescopes at time t. For more details on how these measurements are acquired, see [7]. In Fig. 1, we visualize the measured frequency coverage, measurements of M87* for an EHT array with 11 telescopes, and the intrinsic resolution of the telescope array.

2.2. Data products

In reality, the complex visibility measurements made by a VLBI imaging array, such as the EHT, include different sources of noise. Specifically, the noisy measurements are characterized by $\Gamma^t_{a,b} = e^{i(\phi^t_a - \phi^t_b)} F^t_{a,b}(x) + \eta^t_{a,b}$, where $F^t_{a,b}$ is the ideal Fourier component measured by telescopes a, b at time t, $\eta^t_{a,b}$ is noise arising from Gaussian thermal noise [8], and ϕ_a and ϕ_b are phase errors arising from the inhomogeneity of the atmosphere [9]. These phase errors make the phase from raw visibility measurements unusable at mm and sub-mm wavelengths [9], rendering VLBI imaging as a phase retrieval problem. However, when we consider a set of three telescopes a, b, c, in the triple product $\Gamma^t_{a,b}\Gamma^t_{b,c}\Gamma^t_{c,a}$, we have an identity property $e^{i(\phi_a - \phi_b)} e^{i(\phi_b - \phi_c)} e^{i(\phi_c - \phi_a)} = 1$. Hence, the product $\Gamma_{a,b}\Gamma_{b,c}\Gamma_{c,a}$, called the *bispectrum*, is invariant to atmospheric error [10]. This motivates the usage of the phase of the bispectrum, called the *closure phase*, as a constraint for the image reconstruction problem. Additionally, with calibration, the visibility amplitudes $|\Gamma^t_{a,b}|$ can be well estimated [7], giving us a second set of constraints[1].

[1] Although the amplitudes can be largely calibrated, some error typically remains. As phase errors are much more challenging, we chose to make the simplifying assumption of fully calibrated amplitudes in this work.

Formally, we define our measurements for a single image as

$$
\begin{aligned}
y &:= (y^{\text{amp.}}, y^{\text{clph.}}) \\
y^{\text{amp.}} &:= \left\{|\Gamma^t_{a,b}|\right\}_{(a,b)\in S^t_2,t} = \left\{|F^t_{a,b}(x)| + \eta^{\text{amp.},t}_{a,b}\right\}_{(a,b)\in S^t_2,t} \\
y^{\text{clph.}} &:= \left\{\angle(\Gamma^t_{a,b}\Gamma^t_{b,c}\Gamma^t_{c,a})\right\}_{(a,b,c)\in S^t_3,t} \\
&= \left\{\angle(F^t_{a,b}(x)F^t_{b,c}(x)F^t_{c,a}(x)) + \eta^{\text{clph.},t}_{a,b,c}\right\}_{(a,b,c)\in S^t_3,t}
\end{aligned}
\tag{1}
$$

where a, b, c index telescopes, t is a time stamp from 0 to T, and $S^t_k = \binom{S^t}{k}$ [2]. Following [7, 11] we treat the noise on $y^{\text{amp.}}$ and $y^{\text{clph.}}$ as Gaussian.

3. APPROACH

In this work, we adapt a method proposed in [5, 6] to the task of reconstructing a video of a black hole from noisy VLBI phase-retrieval measurements. Although [5, 6] demonstrated their approach on the simpler task of reconstructing a video from idealized VLBI compressed-sensing measurements, they did not include realistic thermal and atmospheric noise sources on the measurements, which warrants several modifications to the method.

The key assumption of this method is that different underlying images share common low-dimensional structure. This assumption is consistent with our problem setting since we are observing a single evolving target across many nights; while it is changing, many features such as the size of the black hole shadow (i.e., ring diameter) remain consistent. We can use this shared structure as regularization even without knowledge of the true underlying data distribution that generated the underlying images. This common structure can be captured by a shared *Image Generation Model (IGM)* G_θ: a deep generative neural network whose weights θ are inferred directly from N noisy measurements $\{y^{(i)} = f^{(i)}(x) + \eta^{(i)}\}_{i=1}^N$. We solve the reconstruction problem by constraining the reconstructed images $\{\hat{x}^{(i)}\}_{i=1}^N$ to lie within the range of G. Fig. 2 shows an overview of our method.

Formally, following [5, 6], we use a proxy for the evidence lower bound (ELBO), termed the ELBOProxy, as our optimization objective. We define 1) a generator of the form $x = G(z)$ where $z \in \mathbb{R}^d$ and $x \in \mathbb{R}^{M\times M}$, 2) N variational distributions for the latent space $\{z^{(i)} \sim q_{\phi^{(i)}}(z^{(i)})\}_{i=1}^N$, and 3) prior distribution $\log p_Z(z|G)$ defined by $z \sim \mathcal{N}(0, I)$. For a single measurement example y,

$$
\begin{aligned}
&\text{ELBOProxy}(G, q_\phi; y) \\
&\quad := \mathbb{E}_{z\sim q_\phi(z)}\Big[\underbrace{\log p(y|G(z))}_{\text{data-fit}} + \underbrace{\log p_Z(z|G)}_{\text{prior}} - \underbrace{\log q_\phi(z)}_{\text{log entropy}}\Big].
\end{aligned}
\tag{2}
$$

[2] It is unnecessary to use all $|S^t_3|$ closure phase measurements. We use the minimum set $\mathcal{S}^t_3$ such that the set of all telescopes $S_t = \cup_{S_i\in\mathcal{S}^t_3} S_i$.

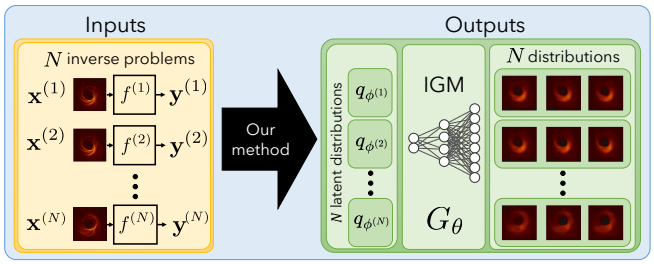


Fig. 2. **Overview of our method.** In this work, we reconstruct a video of the black hole M87* from synthetic sparse, noisy very-long baseline interferometry (VLBI) measurements. This problem is highly ill-posed and non-convex. We propose solving this video reconstruction by learning an Image Generation Model (IGM) directly from noisy measurements of a single black hole evolving over time (described in Sec. 3). Our key insight is that images of different snapshots of the same black hole share common low-dimensional structure. The inputs of our method are N measurement examples $\{y^{(i)}\}_{i=1}^N$ with known forward models $\{f^{(i)}\}_{i=1}^N$. The outputs are a single inferred IGM G_θ leading to N image reconstruction distributions $\{q_{\phi^{(i)}}(x^{(i)})\}_{i=1}^N$, from which we sample to reconstruct the underlying images $\{x^{(i)}\}_{i=1}^N$.

For a collection of measurements $\{y^{(i)} = f^{(i)}(x^{(i)}) + \eta^{(i)}\}_{i=1}^N$ where we assume that the underlying images $\{x^{(i)}\}$ share common structure, we aim to infer the parameters of N latent space distributions $q_{\phi^{(1)}}, \ldots, q_{\phi^{(N)}}$ and a shared generator G_θ by minimizing the loss function,

$$\mathcal{L}_{\text{ELBO}} = -\frac{1}{N}\sum_{i=1}^{N}\left[\text{ELBOProxy}(G_\theta, q_{\phi^{(i)}}; y^{(i)}) + \log p(G_\theta)\right].$$

In the VLBI phase-retrieval problem setting, each measurement y can be described as Eq. 1, which induces $\text{ELBOProxy}(G_\theta, q_{\phi^{(i)}}; \{y^{\text{amp.},(i)}, y^{\text{clph.},(i)}\})$. To combine these into a single objective, we control the relative strength between the visibility amplitude and closure phase data-fits with a hyperparameter α, resulting in the updated data-fit $\log p(y|G(z))$ from Eq. 2:

$$\log p(y|G(z)) = \log p(y^{\text{clph.}}|G_\theta(z)) + \alpha \log p(y^{\text{amp.}}|G_\theta(z))$$

Since phase retrieval problems have intrinsic phase ambiguities, spatial shifts and flips are possible reconstructions. Closure phases remove the flip ambiguity, but the spatial shift ambiguity still remains. Modelling such a multi-modal distribution is challenging, so we introduce a centering loss term to help with the optimization. The center loss is defined by $\mathcal{L}_{\text{center}} := \frac{1}{2N}\sum_{i=1}^{N}\left|(\text{Center}(x^{(i)}) - \text{COM}(x^{(i)}))\right|^2$ where $\text{Center}(x)$[3] and $\text{COM}(x)$[4] are the center point and the center of mass of the image x respectively. We use the hyperparameter β to control the strength of the centering loss, which we anneal from β to 0 as a function of epoch k (ε_k). Thus our final optimization objective is

$$\{\hat{\theta}, \hat{\phi}^{(1)}, \ldots, \hat{\phi}^{(N)}\} = \underset{\theta, \{\phi^{(i)}\}_{i=1}^N}{\text{argmin}} \left\{\mathcal{L}_{\text{ELBO}} + \beta\varepsilon_k\mathcal{L}_{\text{center}}\right\}.$$

Once the parameters have been inferred, $\hat{x}^{(i)}$ is found by sampling $\hat{z}^{(i)} \sim q_{\hat{\phi}^{(i)}}(z^{(i)})$ and computing $\hat{x}^{(i)} = G_\theta(\hat{z}^{(i)})$.

4. EXPERIMENTAL RESULTS

4.1. Synthetic data generation

We show our results on a collection of realistic synthetic measurements for M87* generated using the `eht-imaging` library [12]. We use 60 sets of synthetic measurements, which are computed using 60 frames of a simulated black hole video from [13, 14] with a realistic flux of 1 Jansky. We use the telescope array EHT2017+, which consists of the 8 telescopes used for the EHT in 2017, with 3 additional augmenting telescopes that have been added or are in the process of being added to the EHT [5]. We generate visibility measurements $\Gamma^t_{a,b}$ with realistic Gaussian thermal noise. The visibility amplitudes $y^{\text{amp.}}$ and closure phases $y^{\text{clph.}}$ are then computed according to Eq. 1.

4.2. Black hole video reconstruction of M87*

We show results of our reconstruction method on selected frames in Fig. 3. The target image is computed by blurring the ground-truth image to the intrinsic resolution of the EHT telescope array (∼25 μas), as shown in Fig. 1.d. The time $\times$ angle plots, which visualize the temporal trajectory of the ring, are created by plotting the intensity counter-clockwise for each of the 60 frames. Since we reconstruct not just a single image but an image distribution, we show images of the empirical mean and standard deviation. Our reconstructions are visually similar to the target and accurately reconstruct the primary features while reconstructing some high-frequency features. Additionally, the time $\times$ angle plots of our reconstructions are similar to that of the target, indicating that our reconstruction accurately captures the dynamics of the black hole even without any temporal regularization. We find that our reconstructions best match the ground truth image at a 10 μas resolution, substantially smaller than the intrinsic 25 μas resolution of the telescope, implying that our approach also achieves a level of superresolution.

Baseline Comparisons We further quantify the quality of the reconstruction by comparing our reconstruction to baseline methods in Fig. 4. We used the official EHT published code in `eht-imaging` [12] to produce the regularized maximum likelihood (RML) baselines with the following regularizers: maximum entropy (MEM-RML) [17], total variation (TV-RML) [18], and total squared variation (TSV-RML) [19].

[3] $\text{Center}(x) := \left(\frac{1}{2}(M+1), \frac{1}{2}(M+1)\right)$

[4] $\text{COM}(x) := \frac{1}{\sum_{m,n=1}^{M} x_{mn}} \sum_{m,n=1}^{M} x_{mn}(m,n)$

[5] 2017 array plus OVRO, Kitt Peak, IRAM NOEMA, and Greenland Telescopes

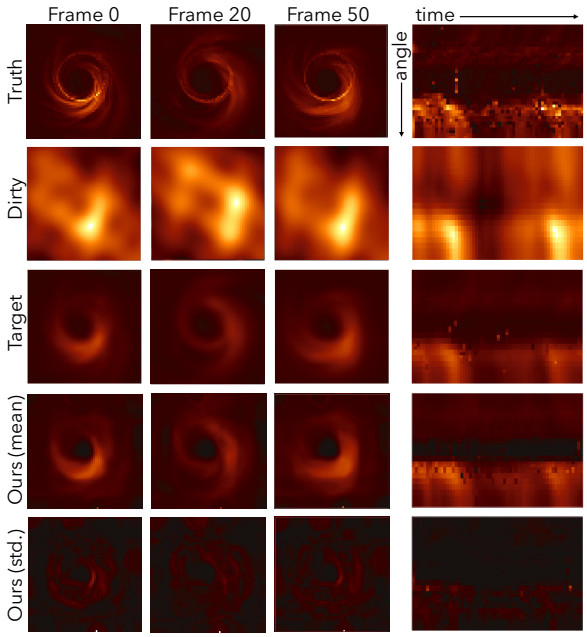
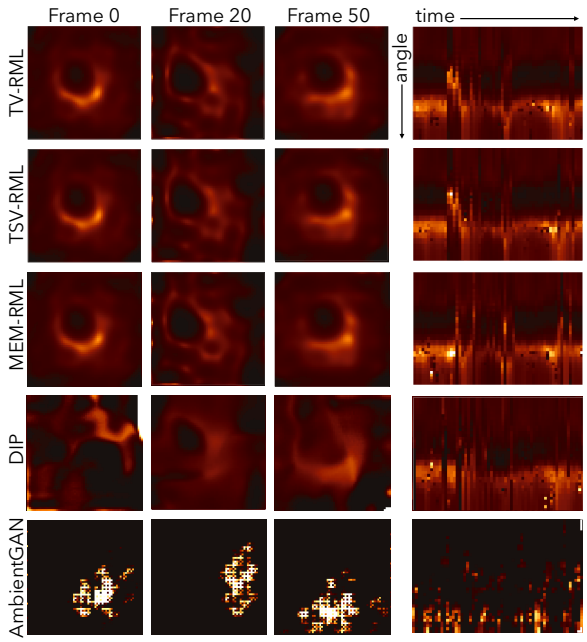


Fig. 3. **Our reconstructions of selected frames from a video of M87*.** We show the ground truth, dirty image, target image (see Fig. 1), and empirical mean and standard deviation of our reconstructed image distribution for selected frames from the 60-frame M87* video. Additionally, we show the unwrapped space vs. time image, which is taken counter-clockwise along the ring; the trajectory of the bright spot in our reconstruction matches the true image. Our method reconstructs the primary features of the true underlying images while also reconstructing some high frequency features.



	Ours	MEM-RML	TV-RML	TSV-RML	DIP	AmbientGAN
PSNR (↑)	**30.870**	26.112	26.119	26.079	24.986	13.573
NXCorr (↑)	**0.980**	0.955	0.956	0.953	0.916	0.519

Fig. 4. **Baseline comparisons.** We show results from the following baseline methods: regularized maximum likelihood with maximum entropy (MEM-RML), total variation (TV-RML), and total-squared-variation (TSV-RML), Deep Image Prior (DIP) [15], and AmbientGAN [16]. We show the average peak signal to noise ratio (PSNR) and normalized cross correlation (NXCorr) compared to the target images (see Fig. 3). Our method exhibits fewer artifacts and has the highest average PSNR and NXCorr, although PSNR is not an ideal metric due to the shift ambiguities in phase retrieval.

We also include the following baselines: 1) Deep Image Prior (DIP) [15], which uses a deep implicit prior that we modify to have a centering loss and 2) AmbientGAN [16], which learns the underlying data distribution through a generative model directly from noisy measurements.

We show the peak signal-to-noise ratio (PSNR) and normalized cross-correlation (NXCorr) for each method compared to the target image. Our method exhibits the highest PSNR and NXCorr, is visually more accurate in reconstructing the image features, and exhibits less artifacts than the other reconstruction methods. Unlike the RML baselines, our method is able to reconstruct the dynamics of the spiral structure in Frame 20. Moreover, our results have temporal consistency, substantially outperforming DIP even when using the same centering loss throughout DIP's inference.

For our forward model dependent hyperparameters, we find that the choice of α has a substantial impact on the data-fit while the reconstructions are less sensitive to β since it is annealed quickly during optimization.

5. CONCLUSION

In this work, we showcase how one can reconstruct images of black holes by inferring an IGM directly from noisy VLBI measurements, without any explicit spatial or temporal priors that would introduce human bias. By leveraging the assumed common structure between different images of the same black hole, we can infer an IGM capable of simultaneously solving the N inverse problems from an N-frame video of a black hole, reconstructing a full movie of a black hole. We demonstrate our method on realistic synthetic interferometric data modelled after the black hole M87*, showing that we can accurately recover the black hole's features and dynamics without any explicit spatial or temporal priors. Our work showcases that we are able to solve the challenging ill-posed and non-convex black hole image reconstruction problem in an unsupervised manner while mitigating human bias. In the future, paired with data collected over the span of months with the ngEHT, our approach could help shed light on potentially surprising phenomenon in M87*'s evolving structure.

6. ACKNOWLEDGEMENTS

This work was sponsored by NSF Award 2048237 and 1935980, an Amazon AI4Science Partnership Discovery Grant, and the Caltech/JPL President's and Director's Research and Development Fund (PDRDF). This research was carried out at the Jet Propulsion Laboratory and Caltech under a contract with the National Aeronautics and Space Administration and funded through the PDRDF. We would like to acknowledge Ben Prather, Abhishek Joshi, Vedant Dhruv, Chi-kwan Chan, and Charles Gammie for providing black hole simulations used in this work.

7. REFERENCES

[1] The Event Horizon Telescope Collaboration, "First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole," *The Astrophysical Journal Letters*, vol. 875, no. 1, pp. L1, Apr 2019.

[2] Michael D. Johnson et al., "Key science goals for the next-generation Event Horizon Telescope," *Galaxies*, vol. 11, no. 3, pp. 61, Apr 2023.

[3] Lindy Blackburn et al., "Studying black holes on horizon scales with VLBI ground arrays," 2019.

[4] Leonid I. Rudin, Stanley Osher, and Emad Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–268, 1992.

[5] Angela F. Gao, Oscar Leong, He Sun, and Katie L. Bouman, "Image reconstruction without explicit priors," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2023.

[6] Oscar Leong, Angela F Gao, He Sun, and Katherine L Bouman, "Ill-posed image reconstruction without an image prior," *arXiv preprint arXiv:2304.05589*, 2023.

[7] A.R. Thompson, J.M. Moran, and G.W. Swenson, Jr., *Interferometry and Synthesis in Radio Astronomy, 3rd Edition*, Springer Cham, 2017.

[8] *Synthesis Imaging in Radio Astronomy II*, vol. 180 of *Astronomical Society of the Pacific Conference Series*, January 1999.

[9] John D. Monnier and Ronald J. Allen, *Radio and Optical Interferometry: Basic Observing Techniques and Data Analysis*, pp. 325–373, Springer Netherlands, Dordrecht, 2013.

[10] R.C. Jennison, "A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent," *MNRAS*, vol. 118, pp. 276, January 1958.

[11] Katherine L. Bouman, Michael D. Johnson, Daniel Zoran, Vincent L. Fish, Sheperd S. Doeleman, and William T. Freeman, "Computational imaging for VLBI image reconstruction," 2016.

[12] Andrew A. Chael, Michael D. Johnson, Katherine L. Bouman, Lindy L. Blackburn, Kazunori Akiyama, and Ramesh Narayan, "Interferometric Imaging Directly with Closure Phases and Closure Amplitudes," *The Astrophysical Journal*, vol. 857, no. 1, pp. 23, April 2018.

[13] The Event Horizon Telescope Collaboration, "First m87 Event Horizon Telescope results. V. Physical origin of the asymmetric ring," *The Astrophysical Journal Letters*, vol. 875, no. 1, pp. L5, 2019.

[14] George N. Wong, Ben S. Prather, Vedant Dhruv, Benjamin R. Ryan, Monika Mościbrodzka, Chi-kwan Chan, Abhishek V. Joshi, Ricardo Yarza, Angelo Ricarte, Hotaka Shiokawa, Joshua C. Dolence, Scott C. Noble, Jonathan C. McKinney, and Charles F. Gammie, "PATOKA: Simulating Electromagnetic Observables of Black Hole Accretion," , vol. 259, no. 2, pp. 64, Apr. 2022.

[15] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.

[16] Ashish Bora, Eric Price, and Alexandros G Dimakis, "AmbientGAN: Generative models from lossy measurements," in *International conference on learning representations*, 2018.

[17] J. Skilling and R. K. Bryan, "Maximum entropy image reconstruction: general algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 211, no. 1, pp. 111–124, 11 1984.

[18] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. 2, no. 3, pp. 296–310, 1993.

[19] Kazuki Kuramochi, Kazunori Akiyama, Shiro Ikeda, Fumie Tazaki, Vincent L Fish, Hung-Yi Pu, Keiichi Asada, and Mareki Honma, "Superresolution interferometric imaging with sparse modeling using total squared variation: application to imaging the black hole shadow," *The Astrophysical Journal*, vol. 858, no. 1, pp. 56, 2018.

[20] Reinhard Heckel and Paul Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," *International Conference on Learning Representations (ICLR)*, 2019.

[21] Mohammad Zalbagi Darestani and Reinhard Heckel, "Accelerated MRI with un-trained neural networks," *IEEE Transactions of Computational Imaging*, vol. 7, pp. 724–733, 2021.

[22] Giannis Daras, Kulin Shah, Yuval Dagan, Aravind Gollakota, Alexandros G Dimakis, and Adam Klivans, "Ambient diffusion: Learning clean distributions from corrupted data," *arXiv preprint arXiv:2305.19256*, 2023.

[23] Katherine L. Bouman, Michael D. Johnson, Adrian V. Dalca, Andrew A. Chael, Freek Roelofs, Sheperd S. Doeleman, and William T. Freeman, "Reconstructing video of time-varying sources from radio interferometric measurements," *IEEE Transactions on Computational Imaging*, vol. 4, no. 4, pp. 512–527, 2018.

Supplementary Materials

1. ADDITIONAL DETAILS OF OUR APPROACH

1.1. Centering loss details

$\text{Center}(x)$ is the center of the $M \times M$-pixel image x defined by

$$\text{Center}(x) = \left(\frac{1}{2}(M+1), \frac{1}{2}(M+1)\right) \tag{3}$$

The center of mass $\text{COM}(x)$ of the image x is defined as

$$\text{COM}(x) = \frac{1}{\text{Flux}(x)} \sum_{m=1}^{M} \sum_{n=1}^{M} x_{mn}(m,n) \tag{4}$$

where $\text{Flux}(x) = \sum_{m=1}^{M} \sum_{n=1}^{M} x_{mn}$ is the total flux (i.e. intensity) of the image x.

The function ε_k that anneals the centering loss in

$$\{\hat{\theta}, \hat{\phi}^{(1)}, \ldots, \hat{\phi}^{(N)}\} = \underset{\theta, \{\phi^{(i)}\}_{i=1}^{N}}{\text{argmin}} \; \{\mathcal{L}_{\text{ELBO}} + \beta \varepsilon_k \mathcal{L}_{\text{center}}\}. \tag{5}$$

is defined as

$$\varepsilon_k = \begin{cases} 1 & \text{if } k \leq 10000 \\ -\frac{k}{10000} + 2 & \text{if } 10000 \leq k < 20000 \\ 0 & \text{if } k \geq 20000 \end{cases} \tag{6}$$

where k is the epoch.

1.2. Array used for the M87 results

NAME	X	Y	Z
PDB	4523998	468045	4460310
PV	5088968	-301682	3825016
SMT	-1828796	-5054407	3427865
SMA	-5464523	-2493147	2150612
LMT	-768714	-5988542	2063276
ALMA	2225061	-5440057	-2481681
APEX	2225040	-5441198	-2479303
JCMT	-5464585	-2493001	2150654
OVRO	-2397431	-4482019	3843524
KP	-1995679	-5037318	3357328
GLT	1500692	-1191735	6066409

Note that SPT is not present in our array as it does not observe M87*.

1.3. Implementation details

In our experiments, following [5, 6] we use a multivariate Gaussian distribution $\mathcal{N}(\mu^{(i)}, \mathcal{L}^{(i)}\mathcal{L}^{(i)^T} + \varepsilon I)$ to parameterize each posterior distribution $q^{(i)}$. We use a $\varepsilon = 10^{-3}$ to help with the stability of the optimization. For our IGM, we use a Deep Decoder [20] with 6 layers, 150 channels, a latent input size of 40, and a dropout of 10^{-4}.

1.4. Synthetic data generation details

We generate visibility measurements $\Gamma^{t}_{a,b}$ with the Python `eht-imaging` library [12], using a bandwidth $\Delta\nu$ of 4 GHz and an integration time t_{int} of 100 seconds. We add realistic Gaussian thermal noise according to standard deviation

$$\sigma_{a,b} = \frac{1}{0.88}\sqrt{\frac{\text{SEFD}_a \times \text{SEFD}_b}{2 \times \Delta\nu \times t_{\text{int}}}} \tag{7}$$

where $\text{SEFD}_a, \text{SEFD}_b$ are the System Equivalent Flux Densities (SEFD) of telescopes a, b respectively [8].

2. BASELINES

2.1. Description

We compare our method with several relevant baselines. A classic approach to solving inverse problems is regularized maximum likelihood (RML), which solves $\hat{x} = \text{argmin}_x\{\mathcal{L}(y, f(x)) + \lambda\mathcal{R}(x)\}$ where $\mathcal{L}$ is the data-fit, $\mathcal{R}$ is a regularizer, and λ is a hyperparameter that controls the relative impact of $\mathcal{L}$ and $\mathcal{R}$. Common regularizers that we compare to are total variation (TV) [18], total squared variation (TSV) [19], and maximum entropy (MEM) [17]. Using explicit regularizers often requires hyperparameter tuning, which requires knowledge of the unknown image distribution. Additionally, these priors are often weak, and unable to represent the true underlying image distribution [**?**]. In contrast to explicit regularizers, there are deep learning based approaches that leverage the structure of neural networks as implicit priors [15, 20, 21]. One such method that we compare to is Deep Image Prior (DIP) [15]. While these work well in some settings, they similarly require some degree of hyperparameter tuning (i.e. early stopping), which again requires knowledge of the unknown underlying data distribution.

In contrast to these single image reconstruction methods, AmbientGAN [16] and AmbientDiffusion [22] leverage the shared structure of the underlying images by learning a generative model directly from noisy measurements, obtaining a generator G that can generate realistic samples from the underlying data distribution. While they have shown strong performance with natural images, there are some challenges to apply AmbientGAN to a few-shot setting with scientific images [5].

A particularly relevant baseline is StarWarps [23], which aims to reconstruct a video of a dynamic black hole by incorporating a temporal prior between measurements as a regularizer. This approach is well-suited to quickly evolving sources such as the black hole Sagittarius A* (SgrA*), but requires tuning the hyperparameter on the temporal regularization.

2.2. Implementation details of baseline methods

We present implementation details for the results for the baselines shown in Fig. 4.

The results for the RML methods TV-RML, TSV-RML, and MEM-RML were computed following the example in `eht-imaging` by initializing a circular Gaussian, optimizing, then 2 cycles of blurring the result and then optimizing. For each of the RML baselines, the default hyperparameters were used, with each data term given weight 100, the regularizer term given weight 1, and additional flux and centering of mass constraints (not annealed) each given weight 500.

The result for DIP was computed with the amplitude term given weight 1, the closure phase term given weight 10, and a centering of mass constraint (not annealed) given weight 10. The result was optimized for 3000 epochs. The results for AmbientGAN were computed with the amplitude term given weight 1, the closure phase term given weight 10, learning rate 5×10^{-6}. The result was optimized for around 10,000 epochs.

3. ADDITIONAL RESULTS

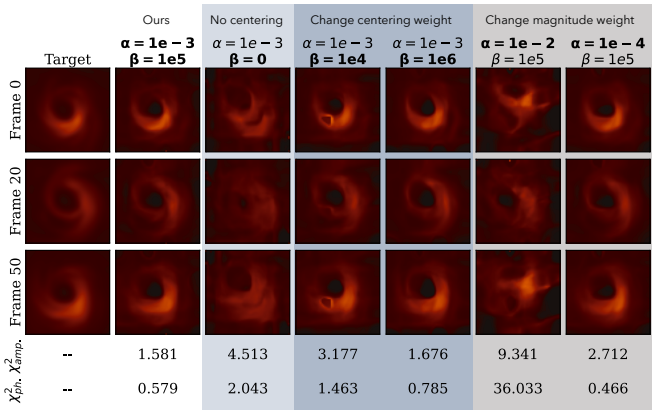


Fig. 5. **Effect of hyperparameters in the data-fit (α) and centering weight (β).** The hyperparameter α controls the relative weight between the amplitude and closure phase terms in Eq. 3, and hyperparameter β controls the weight of the centering loss term in Eq. 5. We show the parameters used for our reconstructions (col. 2) alongside reconstructions with no centering loss $\beta = 0$ (col. 3), decreased and increased centering loss β (col. 4 & 5), and decreased and increased amplitude weight α (col. 6 & 7). For all results, we show the mean $\chi^2_{\text{amp.}}$ (see Eq. 8) and $\chi^2_{\text{clph.}}$ (see Eq. 9), which represent the data-fit where lower is a better data-fit.

3.1. Ablation study

In Fig. 5, we show the effects of the choice of hyperparameters α, which controls the relative weight between the amplitude and closure phase terms (see Eq. 3), and β, which controls the weight of the centering loss term (see Eq. 5). For each hyperparameter setting, we include the quantitative metrics $\chi^2_{\text{amp.}}$ and $\chi^2_{\text{clph.}}$, which express the data-fit for the amplitude and closure phase respectively, where a lower number indicates a better fit. Formally, let $y^{\text{amp.},(i)} = \left(y_j^{\text{amp.},(i)}\right)_{j=1}^{M_i}$ and $y^{\text{clph.},(i)} = \left(y_j^{\text{clph.},(i)}\right)_{j=1}^{N_i}$. The quantities $\chi^2_{\text{amp.}}$ and $\chi^2_{\text{clph.}}$ are computed according to

$$\chi^2_{\text{amp.}} = \frac{1}{M_i}\sum_{j=1}^{M_i}\left(\frac{y_j^{\text{amp.},(i)} - \hat{y}_j^{\text{amp.},(i)}}{\sigma_j^{\text{amp.},(i)}}\right)^2 \tag{8}$$

$$\chi^2_{\text{clph.}} = \frac{2}{N_i}\sum_{j=1}^{N_i}\left(\frac{1-\cos\left(y_j^{\text{clph.},(i)} - \hat{y}_j^{\text{clph.},(i)}\right)}{\left(\sigma_j^{\text{clph.},(i)}\right)^2}\right) \tag{9}$$

where $\hat{y}^{(i)} = (\hat{y}^{\text{amp.},(i)}, \hat{y}^{\text{clph.},(i)})$ are the observations made on the reconstructed image $\hat{x}$ (see Eq. 1).

We note that removing the centering loss $\beta = 0$ entirely (col. 3) results in a posterior with multiple shifted modes due to the shift ambiguities in phase retrieval. This causes the mean image to appear as multiple overlaid rings. Increasing the centering loss β (col. 5 & 5) does not have a pronounced effect on the reconstruction, since the $\mathcal{L}_{\text{center}}$ term is quickly annealed to 0 in the loss function (see Eq. 5). However, if the centering weight is not sufficiently strong enough (col. 4), sometimes this affects the reconstruction quality. We also note the effect of decreasing or increasing α (cols. 6 & 7), which respectively decreases or increases the weight on the amplitude term relative to the closure phase term (see Eq. 3). Decreasing α worsens the fit of the amplitude $\chi^2_{\text{amp.}}$, but improves the fit of the closure phase $\chi^2_{\text{clph.}}$ (col. 6). Increasing α to a sufficiently high value (col. 7) makes it difficult to recover the full image structure as the closure phase information, which determines much of the image structure, has too low a weight. This substantially worsens the fit of closure phase $\chi^2_{\text{clph.}}$ but also does not improve the fit of the amplitude $\chi^2_{\text{amp.}}$ by much.

	Blur strength in μas							
	0	5	10	15	20	25	30	35
NCC	0.932	0.973	0.980	0.969	0.950	0.928	0.906	0.884

Table 1. **Intrinsic resolution of our result.** To find the intrinsic resolution of our reconstruction, we find the optimal amount of blur σ that maximizes the normalized cross correlation (NXCorr) between our reconstruction $\hat{x}$ and the blurred true underlying image $\text{Blur}(x, \sigma)$. Formally, we are solving $\arg\max_\sigma \sum_{i=1}^{N} \text{NXCorr}(\text{Blur}(x^{(i)}, \sigma), \hat{x}^{(i)})$ through a grid search over values of σ, which have units μas. We take the empirical mean of our posterior distribution as $\hat{x}$.

3.2. Intrinsic resolution

To find the intrinsic resolution of our reconstruction, we find the optimal amount of blur σ that maximizes the normalized cross correlation (NXCorr) between our reconstruction $\hat{x}$ and the blurred true underlying image $\text{Blur}(x, \sigma)$ in Table 1. Formally, we are solving

$$\arg\max_{\sigma} \sum_{i=1}^{N} \text{NXCorr}(\text{Blur}(x^{(i)}, \sigma), \hat{x}^{(i)}) \tag{10}$$

through a grid search over values of σ, which have units μas. We take the empirical mean of our posterior distribution as $\hat{x}$. We find that our reconstructions maximize the normalized cross correlation at a blur of $\sim$10 μas, achieving a level of superresolution.