Supplementary Material
Deep Probabilistic Imaging: Uncertainty Quantification and Multi-modal Solution Characterization for Computational Imaging

He Sun, Katherine L. Bouman
California Institute of Technology, 1200 E California Blvd, Pasadena, California 91125

Contents

1 DPI with Different Regularizers

In the paper we demonstrated DPI mostly using a Gaussian image prior, but chose to use maximum entropy (MEM) and total squared variation (TSV) regularization for the real M87 EHT data in order to more meaningfully compare to the original reconstruction results presented in [The EHT Collaboration et al.2019]. In this section, we compare the DPI's approximate posterior obtained with different image regularization. The choice of image prior (i.e., regularization) will influence the posterior distribution, and thus also the learned DPI distribution. In this section, we compare how three different regularizers influence the DPI samples: (a) a Gaussian image regularizer that strongly encourages smoothness, (b) a Gaussian image regularizer that weakly encourages smoothness, and (c) regularization that combines total squared variation (TSV) and maximum entropy (MEM) regularizers.

1.1 Gaussian image prior

A Gaussian image regularizer assumes the hidden image satisfies a multivariate Gaussian distribution, $x \sim p(x) = \mathcal{N}(\mu, \Lambda)$, resulting in the image regularization loss

$$\mathcal{R}_{gauss}(x) = \frac{1}{2}(x-\mu)^T\Lambda^{-1}(x-\mu), \tag{1}$$

where $\mu \in \mathbb{R}^{N^2\times 1}$ is the mean image, $\Lambda \in \mathbb{R}^{N^2\times N^2}$ is the pixel covariance matrix, and the image is of size $N \times N$ pixels. We define the covariance matrix of the Gaussian image regularizer based on the dropoff rate of an image's power spectrum, α. In particular, we define the covariance matrix Λ as

$$\begin{aligned}
\Lambda &= c^2 \text{diag}[\mu]^T \Lambda' \text{diag}[\mu] \\
\Lambda' &= W^{\star T} \text{diag}[b] W \\
b[i] &= \begin{cases} (u[i]^2 + v[i]^2)^{-\alpha/2} & u[i]^2 + v[i]^2 > 0 \\ \epsilon & u[i]^2 + v[i]^2 = 0 \end{cases}
\end{aligned} \tag{2}$$

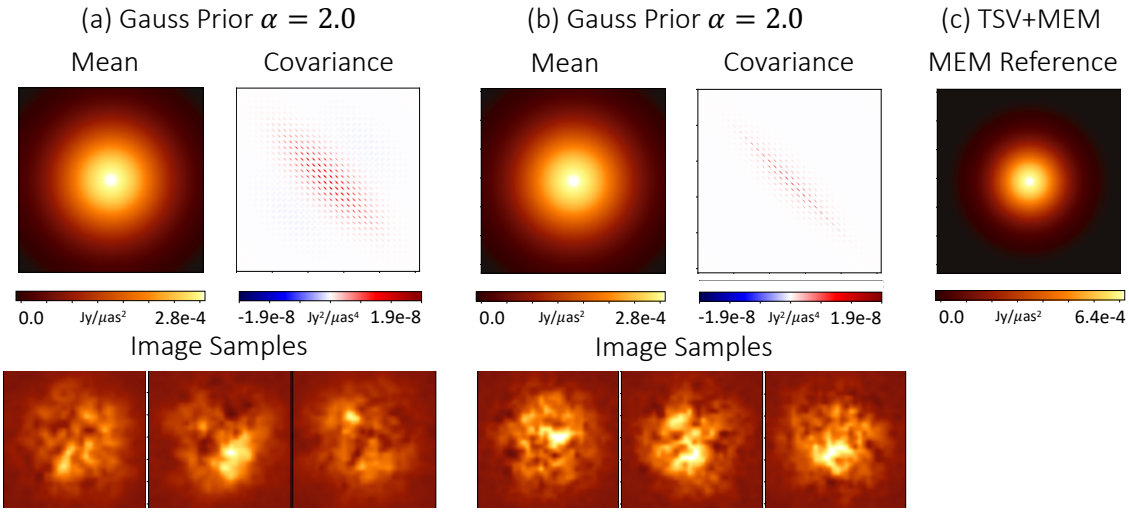

Figure 1: Three image regularizers used in DPI. (a-b) Two Gaussian image priors with different power spectrum dropoff rates ($\alpha = 2.0$ or 1.0). The mean image and covariance matrix of each Gaussian prior distribution and three corresponding randomly sampled images are reported. Larger α results in stronger correlations among neighboring pixels, and thus results in smoother sampled images. (c) A combination of total squared variation (TSV) and MEM regularization. The MEM regularizer uses a circular Gaussian reference image, which encourages the reconstructed image to be bright in the center of the field of view.

where W is a discrete Fourier transform matrix and $c = 1/2$ is a tuning parameter that helps encourage positivity [Bouman et al.2018]. The scalar frequency corresponding to each row i of W is given by $(u[i]^2 + v[i]^2)^{1/2}$. In our experiments, the two Gaussian image priors have the same mean (a circular Gaussian with a full-width-half-max of half the reconstructed image size) but have different covariances, defined by $\alpha = 2.0$ and $\alpha = 1.0$. Figure 1(a) and (b) show the mean and covariance matrices of the two Gaussian priors, and three sample images from each distribution. As expected, the prior with $\alpha = 2.0$ enforces stronger correlations among neighboring pixels resulting in smoother sampled images.

1.2 Maximum entropy and total squared variation

The Maximum Entropy (MEM) regularizer defines a reference image, x_{ref} and forces the reconstructed image to be similar to the reference image by minimizing the relative entropy between the two images,

$$\mathcal{R}_{mem}(x) = \sum_{i=1}^{N^2} x[i] \log \frac{x[i]}{x_{ref}[i]}. \tag{3}$$

Figure 1 (c) shows the reference image used in our MEM regularization, which is a circular Gaussian with a full-width-half-max of roughly 1/3 the reconstructed image field of view. Total squared variation (TSV) is a smoothing regularizer that minimizes the spatial gradient of an image,

$$\mathcal{R}_{tsv}(x) = \sum_{l} \sum_{m} [(\bar{x}[l+1, m] - \bar{x}[l, m])^2 + (\bar{x}[l, m+1] - \bar{x}[l, m])^2], \tag{4}$$

where $\bar{x} \in \mathbb{R}^{N \times N}$ is the two-dimensional image reshaped from the image vector $x \in \mathbb{R}^{N^2 \times 1}$.

1.3 Comparison of DPI distributions using different regularizers

We have demonstrated the DPI approaches using (a) the Gaussian regularizer ($\alpha = 2.0$) in the official paper. Here we report the DPI results using the other two regularizers, (b) the Gaussian regularizer ($\alpha = 1.0$) and (c) the TSV+MEM prior. Fig. 2 and Fig. 3 present the DPI reconstruction of the non-convex interferometric imaging problem with regularizers (b) and (c), similar to Fig. 4 of the official paper using regularizer (a). In all cases, we detect two modes in the

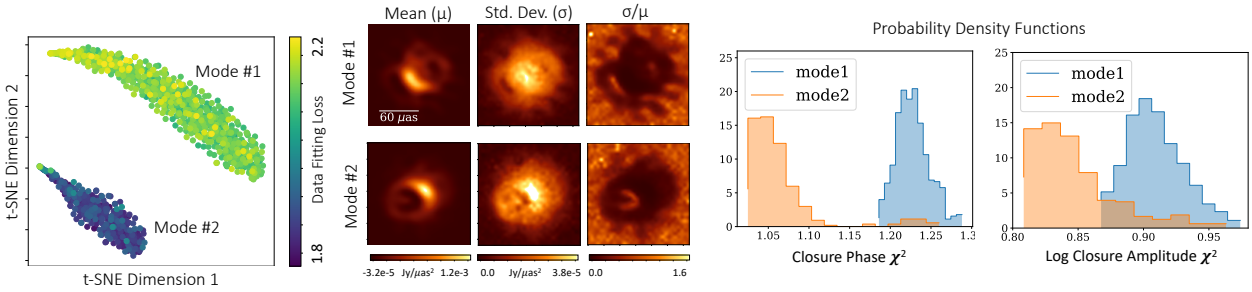

Figure 2: Analysis of 1024 DPI generated reconstructed images trained with closure quantities and Gaussian image regularization ($\alpha = 1.0$). (Left) two-dimensional t-SNE plot of samples with perplexity= 20. The red triangle marks the blurred true image in the embedded space. The samples clearly cluster into two modes. (Middle) The mean, standard deviation, and fractional standard deviation for samples from each mode. (Right) The distributions of data fitting losses (reduced χ^2) of samples from each mode. A χ^2 value of 1 is optimal for high SNR data.

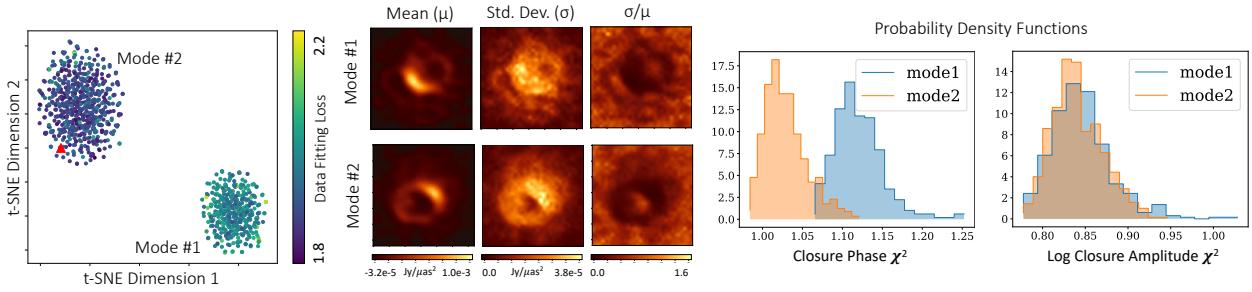

Figure 3: Analysis of 1024 DPI generated reconstructed images trained with closure quantities and TSV+MEM regularization. (Left) two-dimensional t-SNE plot of samples with perplexity= 20. The red triangle marks the blurred true image in the reduced space. (Middle) The mean, standard deviation, and fractional standard deviation for samples from each mode. (Right) The distributions of data fitting losses (reduced χ^2) of samples from each mode. The data fitting losses for the two modes are less separated when using TSV+MEM regularization. Nevertheless, the correct mode has a slightly better data fitting loss on average.

image posterior. However, the distribution of the two modes changes somewhat with the change of regularizer. Since regularizers (b) and (c) allow more small-scale structure to appear in the reconstructions, the resulting images appear to fit the measurements better for Mode #1 compared to when regularizer (a) is used.

Figure 4 shows the DPI reconstruction of two other examples using the TSV+MEM regularizer, similar to Fig. 6 of the official paper. In both the convex and non-convex image reconstructions the blurred true image (red triangle) lies close to the posterior.

2 Variational Inference with Simple Gaussian Models

DPI is a variational inference method that employs a flow-based neural network to learn the hidden image's posterior distribution. Because of the flow network's large modeling capacity, DPI can easily approximate complicated posterior distributions and capture multi-modal solutions. As a comparison, we report results using a Gaussian distribution as the variational model. Models are trained to approximate the posterior for non-convex interferometric image reconstruction.

Figure 5(left) shows the mean and standard deviation of a learned approximate posterior using an independent Gaussian distribution (with diagonal covariance matrix) as the variational model. We show the result of three runs obtained with different random initializations. The independent pixel distribution identifies only one mode in each run.

Figure 5(right) shows the result of fitting a full multivariate Gaussian distribution, $\mathcal{N}(\mu, C)$, to the posterior distribution. The mean and covariance of the variational approximation, along with random samples from the learned Gaussian distribution, are shown. Approximating the posterior with Gaussian parameters m and C is equivalent to

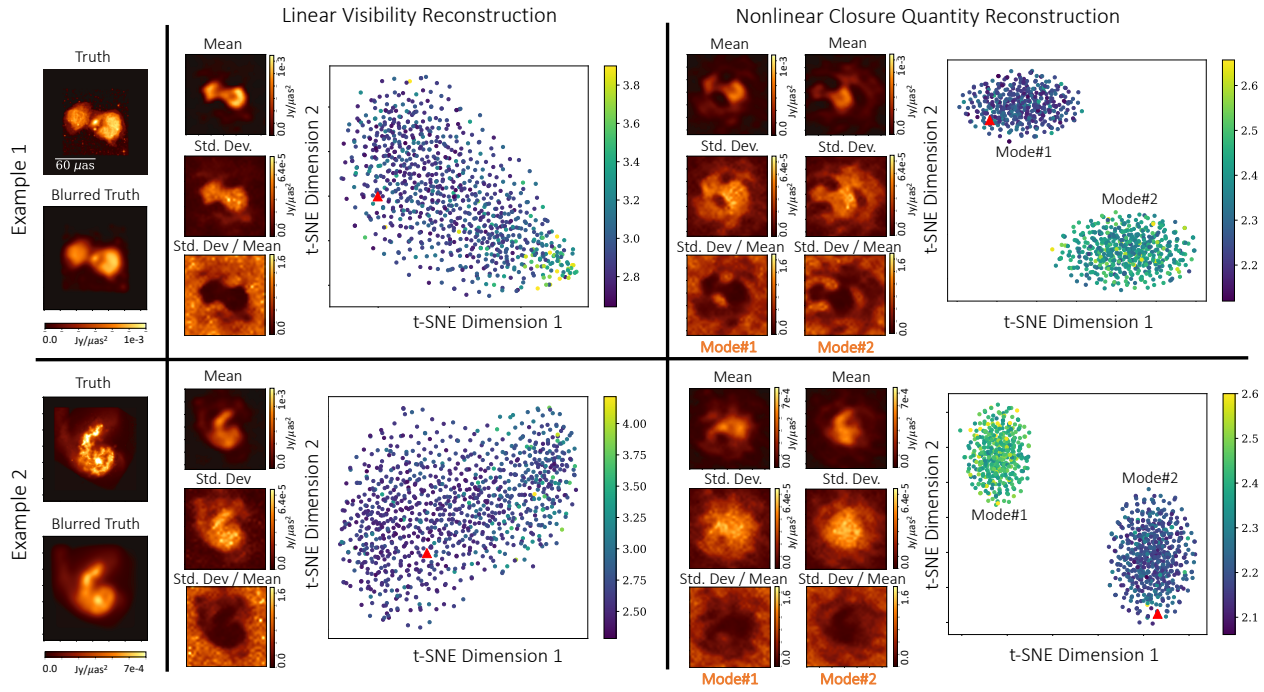

Figure 4: Examples of DPI uncertainty quantification with TSV+MEM regularization. For each image, we report the mean, standard deviation, and two-dimensional t-SNE plot of DPI posterior samples in both convex and non-convex interferometric imaging. Since the TSV+MEM does not define an explicit prior distribution, the image posterior distribution is not analytically solvable even in convex reconstruction. However, DPI still is able to learn an approximate posterior distributions.

learning a generative model with a single dense layer, i.e.,

$$x = \mu + Wz, \text{ where } C = WW^T, \tag{5}$$

Since the multivariate Gaussian model introduces second order correlations among pixels it should be able to somewhat capture multiple image modes. However, although all samples from the learned distribution show ring like structures, these images do not fit the measurement data well. We believe this is due primarily to difficulty in optimizing to a global minimum when solving for a full covariance matrix.

3 DPI Reconstruction Results using Real VLBA data

In Fig. 7 we demonstrate DPI on the reconstruction of two astronomical sources using real interferometric data collected by the VLBA [Jorstad et al.2005]. By constraining only the closure quantities, DPI solves for a distribution of images that fit the measurements under multivariate Gaussian regularization. Although ground truth images are unavailable, we compare our reconstructions to those generated by the BU Blazar Group, reconstructed using CLEAN with an interactive calibration procedure described in [Högbom1974].

Figures 6 and 7 shows the DPI reconstruction results using Architecture A (allowing negative image pixels) and Architecture B (all positive image pixels), respectively. In both figures, the top row shows the DPI reconstruction of the blazar 3C111 and the bottom row corresponds to reconstructions of the blazar 3C84. For the 3C111 dataset, the frequency space is sufficiently covered to only result in small image perturbations in the posterior (resulting in one reconstruction mean). In contrast, the dataset for 3C84 has sparser coverage, resulting in higher uncertainty and multiple modes in the posterior (resulting in two reconstruction mode means). The results using Architecture A appear cleaner and more similar to CLEAN images. This is likely due to negative pixels helping to absorb artifacts caused by an overly constrained field of view.

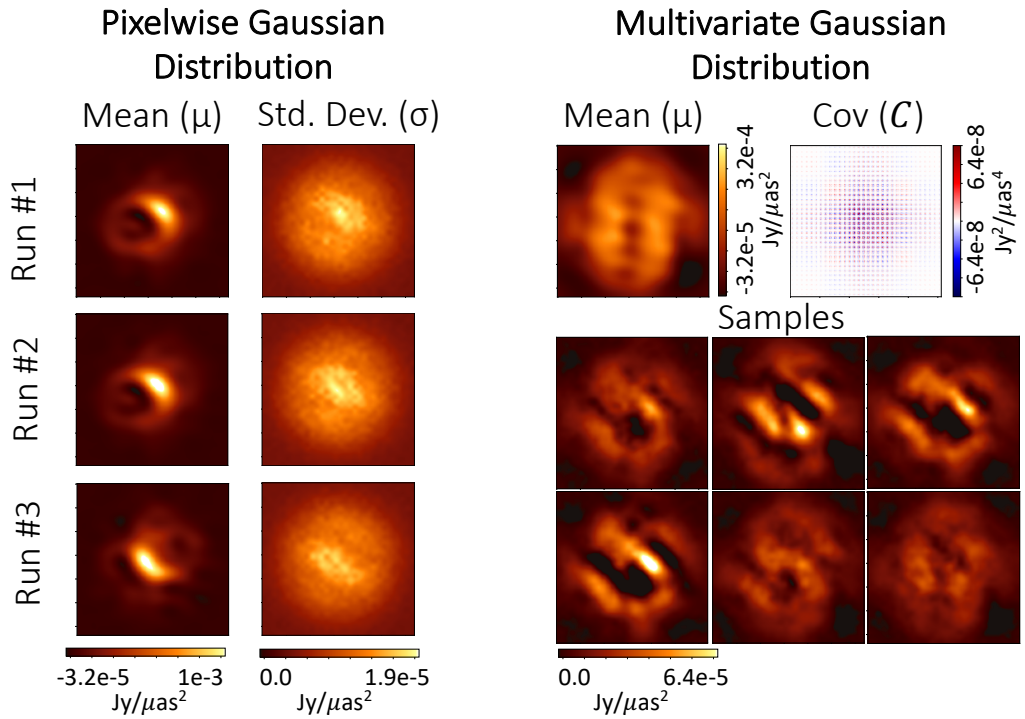

Figure 5: Learned image posterior using variational inference with an (left) independent pixel Gaussian model, or (right) multivariate Gaussian model.

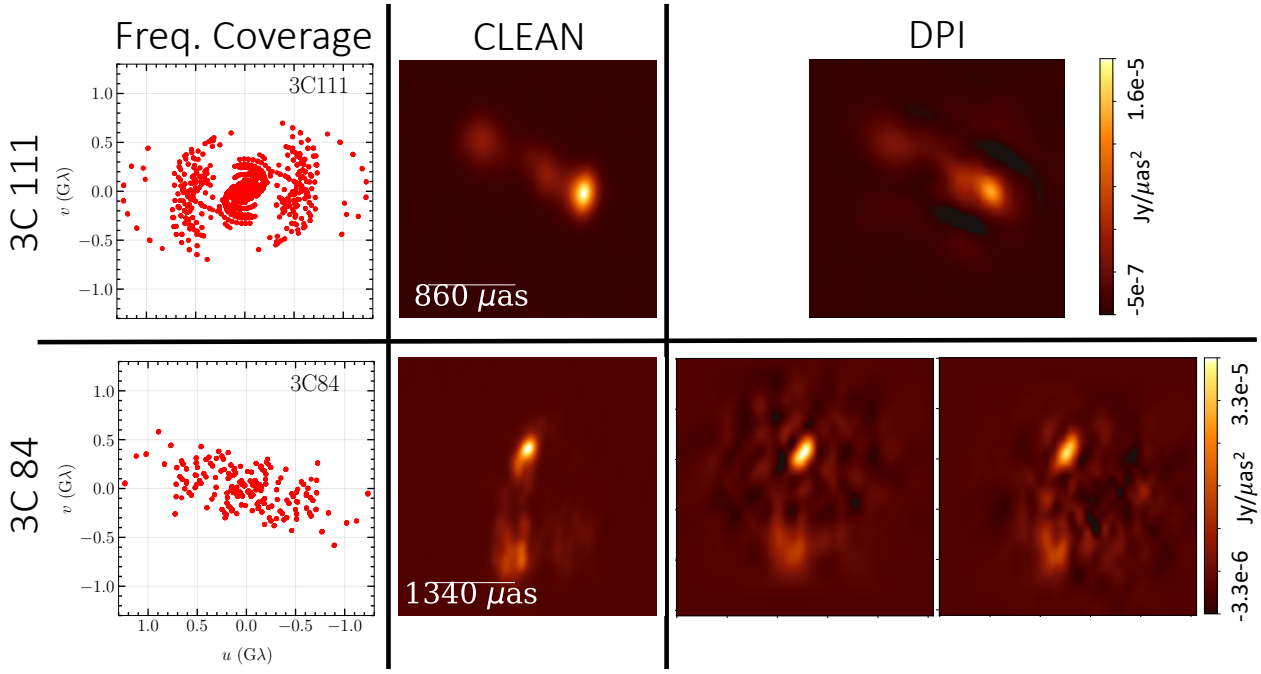

Figure 6: VLBA DPI reconstruction using Architecture A (no non-negativity constraint). We empirically identify one mode for the reconstruction of target 3C111 and two modes for target 3C84. For computational reasons, we have chosen to use a restrictive field of view that may result in artifacts. However, allowing negative pixels helps to partially absorb these errors.

References

[Bouman et al.2018] Bouman, K. L.; Johnson, M. D.; Dalca, A. V.; Chael, A. A.; Roelofs, F.; Doeleman, S. S.; and Freeman, W. T. 2018. Reconstructing video of time-varying sources from radio interferometric measurements.

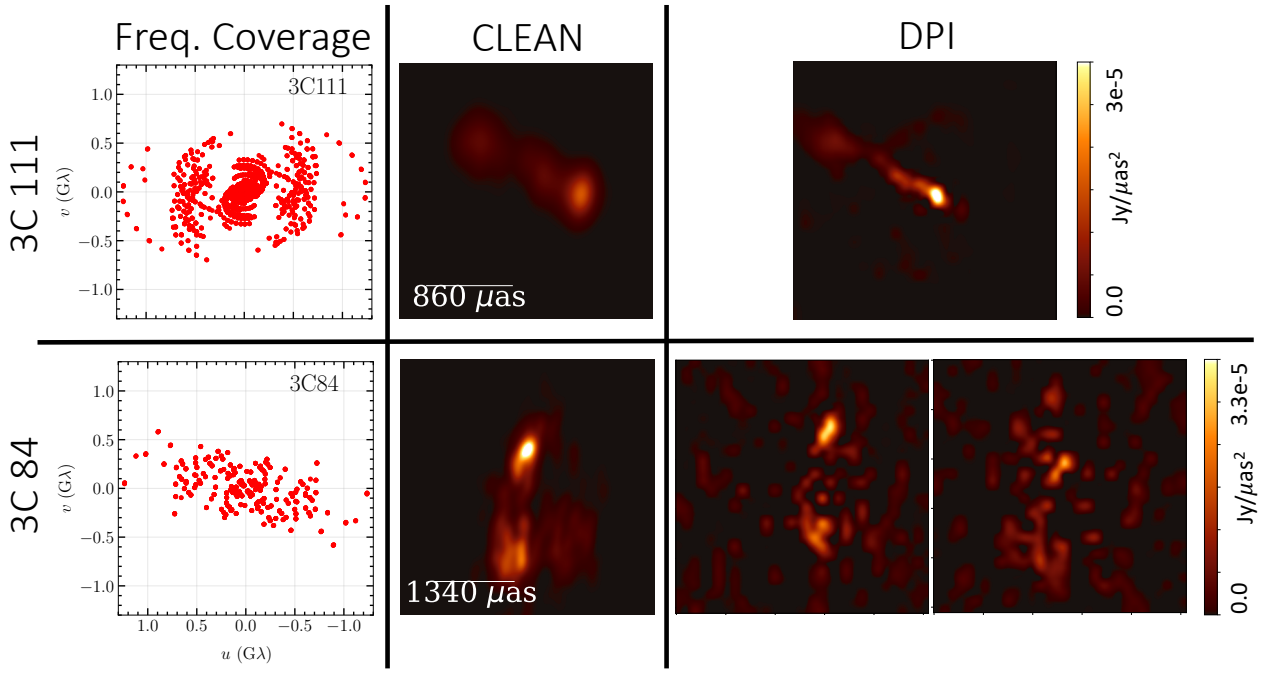

Figure 7: VLBA DPI reconstruction using Architecture B (non-negativity constraint). Strong artifacts that appear in the DPI images are likely due to an overly constrained field of view.

IEEE Transactions on Computational Imaging 4(4):512–527.

[Högbom1974] Högbom, J. 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement Series* 15:417.

[Jorstad et al.2005] Jorstad, S. G.; Marscher, A. P.; Lister, M. L.; Stirling, A. M.; Cawthorne, T. V.; Gear, W. K.; Gómez, J. L.; Stevens, J. A.; Smith, P. S.; Forster, J. R.; et al. 2005. Polarimetric observations of 15 active galactic nuclei at high frequencies: jet kinematics from bimonthly monitoring with the very long baseline array. *The Astronomical Journal* 130(4):1418.

[The EHT Collaboration et al.2019] The EHT Collaboration, Akiyama, K.; Alberdi, A.; Alef, W.; Asada, K.; Azulay, R.; Baczko, A.-K.; Ball, D.; Baloković, M.; Barrett, J.; Bintley, D.; et al. 2019. First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *The Astrophysical Journal Letters* 875(1):L4.