Bayesian Imaging for Radio Interferometry with Score-Based Priors

Noé Dia[1,2,4]⬤   **M. J. Yantovski-Barth**[1,2,4]⬤   **Alexandre Adam**[1,2,4]⬤   **Micah Bowles**[5]⬤
Pablo Lemos[1,2,3,4]   **Anna M. M. Scaife**[5,6]   **Yashar Hezaveh**[1,2,3,4,7,8]
Laurence Perreault-Levasseur[1,2,3,4,7,8]

[1]Université de Montréal   [2]Ciela Institute   [3]Flatiron Institute   [4]Mila   [5]University of Manchester
[6]The Alan Turing Institute   [7]Trottier Space Institute   [8]Perimeter Institute
{noe.dia,michael.barth,alexandre.adam,yashar.hezaveh,
laurence.perreault.levasseur}@umontreal.ca
pablo.lemos@mila.quebec
micah.bowles@postgrad.manchester.ac.uk
anna.scaife@manchester.ac.uk

Abstract

The inverse imaging task in radio interferometry is a key limiting factor to retrieving Bayesian uncertainties in radio astronomy in a computationally effective manner. We use a score-based prior derived from optical images of galaxies to recover images of protoplanetary disks from the DSHARP survey. We demonstrate that our method produces plausible posterior samples despite the misspecified galaxy prior. We show that our approach produces results which are competitive with existing radio interferometry imaging algorithms.

1   Introduction

Interferometry, a powerful technique in astronomy, combines signals from an array of radio antennae to achieve resolutions unattainable by any individual telescope. ALMA, a modern radio interferometer, is capable of high angular resolutions with high sensitivity, which enables the study of objects such as protoplanetary disks [e.g., 1] or high-redshift galaxies [e.g., 2]. The analysis of interferometric data, however, is exceedingly challenging due to the incomplete sampling of the Fourier space (uv-space) where data is measured. The most accurate methods fit the data directly in uv-space but due to the large number of the measured Fourier modes (visibilities), these methods require extremely expensive computations. A popular avenue is therefore to convert interferometric data to images of the sky.

The most widely-used imaging methods are derivatives of CLEAN [3–7], a deconvolution algorithm which attempts to iteratively remove from the images the effects of the incomplete uv-space sampling. Unfortunately, all versions of CLEAN suffer from several drawbacks: the algorithm can produce images with unphysical negative flux, results in suboptimal resolution, and does not output statistical uncertainties [8]. These limitations have motivated efforts to develop alternative imaging algorithms. One of these approaches is MPoL [9], which adopts a regularized maximum likelihood approach. Still, this technique does not provide uncertainty quantification and relies on fine-tuning of regularization parameters (an implicit prior). Another approach is `resolve` [10–12], which adopts a Gaussian prior in a Bayesian inference setting. To handle the computational challenges of Bayesian inference, `resolve` employs Metric Gaussian Variational Inference [13].

In this work, we use score-based generative models [14, 15] as principled priors [16] to perform Bayesian inference for radio interferometric imaging. Given a prior (in our case, a trained denoising score network), our formalism is free of hyperparameter tuning and can be adapted on the fly to new datasets without modification. Furthermore, only a fixed number of steps are required to generate

samples from a good approximation to the high-dimensional posterior. We test our pipeline on ALMA observations of 3 DSHARP protoplanetary disks [17] as first application of this method to radio interferometry data. For our prior model, we use high-resolution optical images of galaxies as proxies of images of protoplanetary disks, which would otherwise only be available through simulations [e.g., 18]. We show that, even with a misaligned prior, this framework is capable of competing with state-of-the-art methods

2 Methods

A radio interferometer measures the visibilities of the sky emission. For distant sources, the van Cittert-Zernike theorem [19, 20] states that the visibilities correspond to the Fourier components of the true sky emission. However, since only a subset of the visibilties is observed and since the visibilties are subject to measurement errors (noise), recovering the sky emission is an ill-posed inverse problem. As such, our goal is to recover plausible surface brightness profiles over a pixel grid, $\mathbf{x} \in \mathbb{R}^n$, where n is the number of pixels of our model, using ALMA observations of a protoplanetary disk, $\mathcal{V} \in \mathbb{C}^m$, where m is the number of gridded visibilities (we expand on the gridding process in the next section). Measurements are linearly related to sky emission via the equation

$$\mathcal{V} = S \mathcal{F} P_{\mathrm{beam}} \mathbf{x} + \boldsymbol{\eta} \,, \tag{1}$$

where $\mathcal{F} \in \mathbb{C}^{n \times n}$ is the dense, unitary, 2D Fourier operator and $S \in \{0, 1\}^{m \times n}$ is a mask, otherwise known as the sampling function, which selects visibilities corresponding to measurements. P_{beam} is the primary beam — the response function of the primary antenna — and $\boldsymbol{\eta}$ is additive noise.

In a Bayesian inference setting, solving this problem translates into the task of sampling from the posterior defined by Bayes's theorem as the product of the likelihood, $p(\mathcal{V} \mid \mathbf{x}) = p(\boldsymbol{\eta})$, which encodes the properties of the additive noise distribution, and the prior, $p(\mathbf{x})$:

$$p(\mathbf{x} \mid \mathcal{V}) \propto p(\mathcal{V} \mid \mathbf{x}) p(\mathbf{x}) \,. \tag{2}$$

The prior represents the known information about the uncertain parameters, here the sky emission, $\mathbf{x}$, before considering any measurement. Although it plays an important role in the statistical inversion, Feng et al. [21] show that misaligned score-based priors can recover plausible solutions for the black hole shadow from radio interferometric measurements. This is consistent with a Bayesian inference perspective, in which our knowledge of the world is iteratively updated with new data. When new data is informative, — i.e., when the likelihood is concentrated or narrow — and the prior is permissive — i.e., does not rule out unlikely events — the posterior can be very different from prior expectations.

In this work, we apply a score-based model trained on images of galaxies to posterior inference over real ALMA observations of DSHARP protoplanetary disks [17]. The likelihood of these observations is quite constraining since the number of observed visibilities is large ($m > 10^6$), so it constitutes an ideal test for our hypothesis that a misaligned prior should still allow us to recover plausible solutions. We summarize herein our methodology to characterize the prior and sample from the posterior.

2.1 Sampling from the posterior with score-based priors

To characterize the prior, $p(\mathbf{x})$ — or more specifically its score, $\boldsymbol{\nabla}_{\mathbf{x}} \log p(\mathbf{x})$ — we employ denoising score matching [22–24]. In summary, we train a neural network based on the U-net architecture [25] with a weighted Fisher divergence loss [26] to match a Gaussian perturbation kernel $p(\mathbf{x}_t \mid \mathbf{x}) = \mathcal{N}(\mu(t)\mathbf{x}, \sigma^2(t)\mathbb{1}_{n \times n})$ associated with a Stochastic Differential Equations (SDE), where both the network and the kernel are parameterized by the time index $t \in [0, 1]$ of the SDE [15]. The scalar functions $\mu(t)$ and $\sigma(t)$ are hyper-parameters to specify. In this work, we use the Variance Preserving (VP) SDE with $\beta_{\mathrm{min}} = 0.01$, $\beta_{\mathrm{max}} = 20$ and $\epsilon = 10^{-5}$ [27].

The SDE used to sample from the posterior is obtained with Anderson's reverse-time formula [28] and setting the posterior as the $t = 0$ boundary condition of the SDE

$$d\mathbf{x} = \big(f(\mathbf{x}, t) - g^2(t) \boldsymbol{\nabla}_{\mathbf{x}} \log p_t(\mathbf{x} \mid \mathbf{y})\big) dt + g(t) d\bar{\mathbf{w}} \,. \tag{3}$$

$\bar{\mathbf{w}}$ is a time-reversed Wiener process and $\mathbf{y} \equiv \mathcal{V}$ is the observation. The drift, $f(\mathbf{x}, t)$, and the homogeneous diffusion coefficient, $g(t)$, are specified by the SDE chosen for the prior [15]. Applying Bayes's theorem, we can decompose the posterior score into a prior score and a likelihood score

$$\boldsymbol{\nabla}_{\mathbf{x}} \log p_t(\mathbf{x} \mid \mathbf{y}) = \boldsymbol{\nabla}_{\mathbf{x}} \log p_t(\mathbf{x}) + \boldsymbol{\nabla}_{\mathbf{x}} \log p_t(\mathbf{y} \mid \mathbf{x}) \tag{4}$$
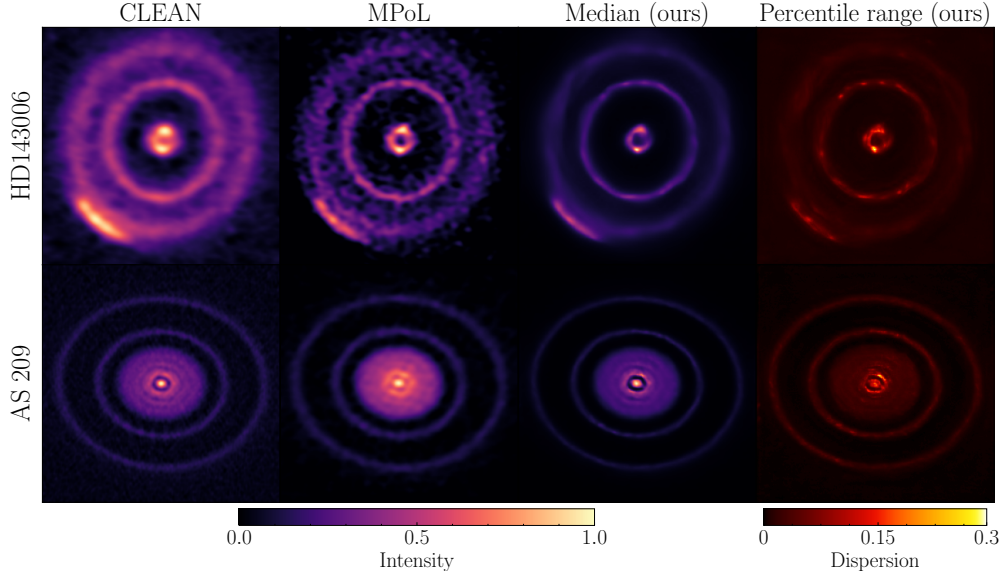
Figure 1: Image reconstructions by CLEAN, MPoL, and our score-based model for two protoplanetary disks. The two columns at right are pixel-wise statistics computed using the posterior obtained from our score-based approach using the SKIRT prior (we also show posterior samples of these two disks in appendix B). From left to right the statistics include the median and the 84%-16% percentile range. The MPoL image for AS 209 (bottom, second from left) is not fully optimized and is shown purely for illustration.

Unfortunately, the likelihood score, $\nabla_{\mathbf{x}} \log p_t(\mathbf{y} \mid \mathbf{x})$, is an intractable quantity [21, 29, 30]. However, since the likelihoods considered in this work are very informative, we can employ the *convolved likelihood approximation* [31, 32], which allows us to write

$$p_t(\mathbf{y} \mid \mathbf{x}_t) \approx \mathcal{N}\left(\mu(t)\mathbf{y} \mid A\mathbf{x}_t, \mu^2(t)\Gamma + \sigma^2(t)\Sigma\right) , \tag{5}$$

where $A \equiv S\mathcal{F}P_{\text{beam}}$, $\Sigma_{jk} = \frac{1}{2}\delta_{jk} + \frac{1}{2}\delta_{j0}\delta_{k0}$, where δ_{jk} denotes the Kronecker delta, and $\Gamma \in \mathbb{R}^{2m \times 2m}$ is the covariance matrix of the noise in the gridded visibilities. Note that the complex distribution in equation (5) is written in terms of a *vectorized* complex random variable, i.e., $\boldsymbol{\eta} \mapsto (\mathrm{Re}(\boldsymbol{\eta}), \mathrm{Im}(\boldsymbol{\eta})) \in \mathbb{R}^{2m}$. We refer to the appendix A for more details.

3 Data and physical model

The **PROBES** dataset is a compendium of high-quality local late-type galaxies [33, 34] that we leverage as a prior for our reconstructions. These galaxies, used in previous studies to train diffusion models [32, 35], have resolved structures of spiral arms, bulges, and disks, which can vaguely resemble the structure of protoplanetary disks. The **SKIRT TNG** [36] dataset is a large public collection of images spanning 0.3-5 microns made by applying dust radiative transfer post-processing [37] to galaxies from the TNG cosmological magneto-hydrodynamical simulations [38]. The SKIRT TNG dataset offers us an opportunity to compare prior distributions based on the same type of objects — galaxies — but with different underlying assumptions about the physics that govern the formation of their structure [39, 40], which in the case of SKIRT TNG is inherited from simulations instead of observations. In this work, we use the z band in both datasets to train the prior score models.

Disk Substructures at High Angular Resolution Project (**DSHARP**) [17] is a recent survey aiming to characterize the substructures of 20 nearby protoplanetary disks by observing their continuum emission around 240 GHz with ALMA. High angular resolution observations of such systems can provide a better understanding of the physical processes occurring in a circumstellar disk which ultimately lead to planetary formation.

We process the calibrated **visibilities** released by the DSHARP survey using the `visread` package created by the MPoL team [41]. For each spectral window, we average the data across the different polarizations and conserve the non-flagged data. We assume a flat spectral index, which permits us
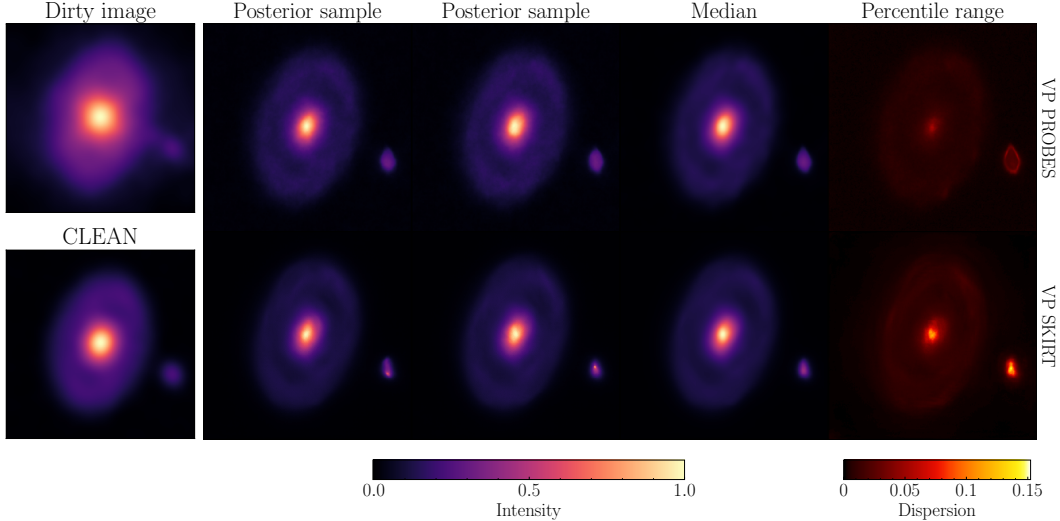
Figure 2: Image reconstruction of HTLup for two different priors. The first column at left shows from top to bottom the dirty image obtained by "naïvely" applying an inverse Fourier transform on the gridded visibilities and the CLEAN reconstruction published by the DSHARP team [43]. The rest of the figure are posterior samples and pixel-wise statistics (median and 84%-16% percentile range) from our approach for a prior trained on two different datasets (PROBES on the first row and SKIRT on the second row).

to combine spectral windows. In order to perform inference over regular grid models, we bin the visibilities using our implementation of a sinc convolutional gridding function, which was chosen for its uniform effect in image space. We estimate the diagonal part of the covariance matrix, $\Gamma \in \mathbb{R}^{2m \times 2m}$, by computing a weighted standard deviation of the real and imaginary parts of the observed visibilities, $\mathcal{V}$, within each bin. If a bin contains a small number of visibilities, we simultaneously expand the bin and the sinc window function until each cell has at least 5 data points.

In a statistical inference setting, the **forward model** attempts to capture the physics connecting the parameters of interest (in our case, pixelated sky emission) to the observed data.

For our forward model, we represent equation (1) using several simplifying assumptions. We define our images on a regular grid with a constant pixel size and we pad our posterior samples with noise during the diffusion in order to match the resolution of the gridded visibilities in Fourier space. We also model $\mathcal{F}$ with a Fast Fourier Transform (FFT) [42], P_{beam} with a flat response (i.e., $P_{\text{beam}} \equiv \mathbb{1}_{n \times n}$), and the noise, $\boldsymbol{\eta} \in \mathbb{R}^{2m}$, by a Gaussian distribution with diagonal covariance, $p(\boldsymbol{\eta}) = \mathcal{N}(0, \Gamma)$.

4 Results and discussion

We test our approach on three protoplanetary disks, HD143006, AS 209, and HT Lup [17, 44] choosing a pixel size of 4 mas, 9 mas and 1.5 mas respectively. Fig. 1 shows the resulting images generated by CLEAN, MPoL, and this work for HD143006 and AS 209.

The CLEAN images were obtained directly from the DSHARP collaboration [17]. For the MPoL image of HD143006, we adopted the hyperparameters from the MPoL documentation's tutorial, which were optimized for this disk; we then performed a limited hyperparameter search for the disk AS 209. Thus, the performance of MPoL on the disk AS 209 is not optimized and is shown only for illustrative purposes. For generating our posterior samples — and since our method is fully parallel — we use 20 V100 GPUs to get 360 posterior samples (256×256 pixels) in ~ 1 hour wall-time. Sampling is performed with 4000 steps of the discretized Euler-Maruyama SDE solver.

In order to avoid differences in unit conventions between the different algorithms, we normalize each image's maximum pixel value to 1. Our method shows competitive resolution and dynamic range performance compared to MPoL and CLEAN.

Next, for HT Lup, we show in Fig. 2 the effect of two different priors trained under same SDE but on different datasets. Even though the prior has a clear effect on the posterior's sample statistics, we still get consistent imaging results across the two priors since the likelihood is highly informative and the priors are sufficiently permissive.

Finally, we attempt to assess the coverage of our posterior estimator using Tests of Accuracy with Random Points (TARP) [45], a method to assess the accuracy of the posterior encoded by a generative model using samples by the model. To speed up the computation of this coverage test, we use a prior trained on a downsampled version of the SKIRT dataset. Our model shows statistically significant deviation from the ideal case (Figure 3), which is a general indicator of a bias in our uncertainty quantification. We believe that the source of this bias may stem from the convolved likelihood approximation. To explore this further, we plan to conduct additional coverage testing of our inference process while varying the built-in assumptions of our modeling. This exploration may uncover potential strategies to mitigate the bias in our approach.

Figure 3: Results from the TARP coverage test for samples generated by our model. The deviation of the red line (this work) at greater than 3-sigma significance from the diagonal indicates a bias in our uncertainty quantification.

In conclusion, our score-based imaging algorithm's results are shown to be competitive with existing radio interferometry imaging algorithms. Our method, while currently biased, has the potential to provide realistic uncertainty quantification. In future work, we wish to explore the possibility of adding samples of protoplanetary disks from simulations as a way to encode a more accurate prior and quantify the robustness of our reconstructions to prior misspecification.

5 Acknowledgments

Software used: `astropy` [46, 47], `jupyter` [48], `matplotlib` [49] , `numpy` [50], `PyTorch` [51], `tqdm` [52], `CASA` [53], `mpol` [9], `visread` [54]
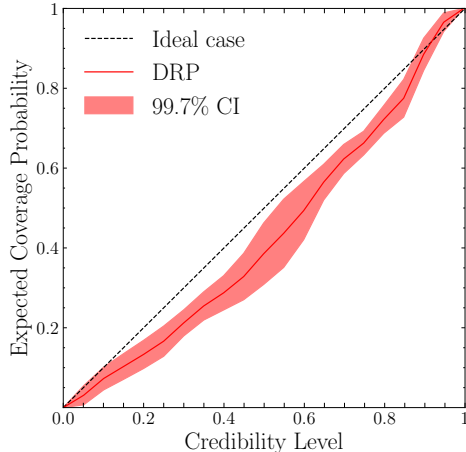
References

[1] ALMA Partnership, C. L. Brogan, L. M. Pérez, T. R. Hunter, W. R. F. Dent, A. S. Hales, R. E. Hills, S. Corder, E. B. Fomalont, C. Vlahakis, Y. Asaki, D. Barkats, A. Hirota, J. A. Hodge, C. M. V. Impellizzeri, R. Kneissl, E. Liuzzo, R. Lucas, N. Marcelino, S. Matsushita, K. Nakanishi, N. Phillips, A. M. S. Richards, I. Toledo, R. Aladro, D. Broguiere, J. R. Cortes, P. C. Cortes, D. Espada, F. Galarza, D. Garcia-Appadoo, L. Guzman-Ramirez, E. M. Humphreys, T. Jung, S. Kameno, R. A. Laing, S. Leon, G. Marconi, A. Mignano, B. Nikolic, L. A. Nyman, M. Radiszcz, A. Remijan, J. A. Rodón, T. Sawada, S. Takahashi, R. P. J. Tilanus, B. Vila Vilaro, L. C. Watson, T. Wiklind, E. Akiyama, E. Chapillon, I. de Gregorio-Monsalvo, J. Di Francesco, F. Gueth, A. Kawamura, C. F. Lee, Q. Nguyen Luong, J. Mangum, V. Pietu, P. Sanhueza, K. Saigo, S. Takakuwa, C. Ubach, T. van Kempen, A. Wootten, A. Castro-Carrizo, H. Francke, J. Gallardo, J. Garcia, S. Gonzalez, T. Hill, T. Kaminski, Y. Kurono, H. Y. Liu, C. Lopez, F. Morales, K. Plarre, G. Schieven, L. Testi, L. Videla, E. Villard, P. Andreani, J. E. Hibbard, and K. Tatematsu. The 2014 ALMA Long Baseline Campaign: First Results from High Angular Resolution Observations toward the HL Tau Region. , 808(1):L3, July 2015. doi: 10.1088/2041-8205/808/1/L3.

[2] N Laporte, R A Meyer, R S Ellis, B E Robertson, J Chisholm, and G W Roberts-Borsani. Probing cosmic dawn: Ages and star formation histories of candidate z 9 galaxies. *Monthly Notices of the Royal Astronomical Society*, 505(3):3336–3346, 06 2021. ISSN 0035-8711. doi: 10.1093/mnras/stab1239. URL https://doi.org/10.1093/mnras/stab1239.

[3] JA Högbom. Aperture synthesis with a non-regular distribution of interferometer baselines. *Astronomy and Astrophysics Supplement, Vol. 15, p. 417*, 15:417, 1974.

[4] Tim J Cornwell. Multiscale clean deconvolution of radio synthesis images. *IEEE Journal of selected topics in signal processing*, 2(5):793–801, 2008.

[5] Urvashi Rau and Tim J Cornwell. A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry. *Astronomy & Astrophysics*, 532:A71, 2011.

[6] AR Offringa and O Smirnov. An optimized algorithm for multiscale wideband deconvolution of radio astronomical images. *Monthly Notices of the Royal Astronomical Society*, 471(1): 301–316, 2017.

[7] T. J. Cornwell. Hogbom's CLEAN algorithm. Impact on astronomy and beyond. Commentary on: Högbom J. A., 1974, A&AS, 15, 417. , 500(1):65–66, June 2009. doi: 10.1051/0004-6361/200912148.

[8] Philipp Arras, Hertzog L. Bester, Richard A. Perley, Reimar Leike, Oleg Smirnov, Rüdiger Westermann, and Torsten A. Enßlin. Comparison of classical and Bayesian imaging in radio interferometry. Cygnus A with CLEAN and resolve. , 646:A84, February 2021. doi: 10.1051/0004-6361/202039258.

[9] Ian Czekala, Brianna Zawadzki, Ryan Loomis, Hannah Grzybowski, Robert Frazier, and Tyler Quinn. Mpol-dev/mpol: v0.1.1 release, June 2021. URL https://doi.org/10.5281/zenodo.4939048.

[10] H. Junklewitz, M. R. Bell, M. Selig, and T. A. Enßlin. RESOLVE: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy. , 586:A76, February 2016. doi: 10.1051/0004-6361/201323094.

[11] Philipp Arras, Philipp Frank, Reimar Leike, Rüdiger Westermann, and Torsten A. Enßlin. Unified radio interferometric calibration and imaging with joint uncertainty quantification. , 627:A134, July 2019. doi: 10.1051/0004-6361/201935555.

[12] Philipp Arras, Jakob Knollrnüller, Henrik Junklewitz, and Torsten A. Enßlin. Radio imaging with information field theory. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 2683–2687, 2018. doi: 10.23919/EUSIPCO.2018.8553533.

[13] Jakob Knollmüller and Torsten A. Enßlin. Metric gaussian variational inference. 2020.

[14] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. 33:6840–6851, 2020. URL https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

[15] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

[16] Alexandros Graikos, Nikolay Malkin, Nebojsa Jojic, and Dimitris Samaras. Diffusion models as plug-and-play priors. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=yhlMZ3iR7Pu.

[17] Sean M. Andrews, Jane Huang, Laura M. Pérez, Andrea Isella, Cornelis P. Dullemond, Nicolás T. Kurtovic, Viviana V. Guzmán, John M. Carpenter, David J. Wilner, Shangjia Zhang, Zhaohuan Zhu, Tilman Birnstiel, Xue-Ning Bai, Myriam Benisty, A. Meredith Hughes, Karin I. Öberg, and Luca Ricci. The Disk Substructures at High Angular Resolution Project (DSHARP). I. Motivation, Sample, Calibration, and Overview. , 869(2):L41, December 2018. doi: 10.3847/2041-8213/aaf741.

[18] Pablo Benítez-Llambay and Frédéric S. Masset. Fargo3d: A new gpu-oriented mhd code. *The Astrophysical Journal Supplement Series*, 223(1):11, 2016. URL http://stacks.iop.org/0067-0049/223/i=1/a=11.

[19] P.H van Cittert. Die wahrscheinliche schwingungsverteilung in einer von einer lichtquelle direkt oder mittels einer linse beleuchteten ebene. *Physica*, 1(1):201–210, 1934. ISSN 0031-8914. doi: https://doi.org/10.1016/S0031-8914(34)90026-4. URL https://www.sciencedirect.com/science/article/pii/S0031891434900264.

[20] F. Zernike. The concept of degree of coherence and its application to optical problems. *Physica*, 5(8):785–795, 1938. ISSN 0031-8914. doi: https://doi.org/10.1016/S0031-8914(38)80203-2. URL https://www.sciencedirect.com/science/article/pii/S0031891438802032.

[21] Berthy T. Feng, Jamie Smith, Michael Rubinstein, Huiwen Chang, Katherine L. Bouman, and William T. Freeman. Score-Based Diffusion Models as Principled Priors for Inverse Imaging. *arXiv e-prints*, art. arXiv:2304.11751, April 2023. doi: 10.48550/arXiv.2304.11751.

[22] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005. URL http://jmlr.org/papers/v6/hyvarinen05a.html.

[23] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Comput.*, 23(7):1661–1674, 2011. doi: 10.1162/NECO_a_00142. URL https://doi.org/10.1162/NECO_a_00142.

[24] Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating distribution. *J. Mach. Learn. Res.*, 15(1):3563–3593, jan 2014. ISSN 1532-4435.

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv e-prints*, art. arXiv:1505.04597, May 2015. doi: 10.48550/arXiv.1505.04597.

[26] Yang Song, Conor Durkan, Iain Murray, and Stefano Ermon. Maximum Likelihood Training of Score-Based Diffusion Models. *arXiv e-prints*, art. arXiv:2101.09258, January 2021. doi: 10.48550/arXiv.2101.09258.

[27] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the Design Space of Diffusion-Based Generative Models. *arXiv e-prints*, art. arXiv:2206.00364, June 2022. doi: 10.48550/arXiv.2206.00364.

[28] Brian D.O. Anderson. Reverse-time diffusion equation models. *Stochastic Processes and their Applications*, 12(3):313–326, 1982. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(82)90051-5. URL https://www.sciencedirect.com/science/article/pii/0304414982900515.

[29] Hyungjin Chung, Jeongsol Kim, Michael T. Mccann, Marc L. Klasky, and Jong Chul Ye. Diffusion Posterior Sampling for General Noisy Inverse Problems. *arXiv e-prints*, art. arXiv:2209.14687, September 2022. doi: 10.48550/arXiv.2209.14687.

[30] Berthy T. Feng and Katherine L. Bouman. Efficient Bayesian Computational Imaging with a Surrogate Score-Based Prior. *arXiv e-prints*, art. arXiv:2309.01949, September 2023. doi: 10.48550/arXiv.2309.01949.

[31] Benjamin Remy, Francois Lanusse, Niall Jeffrey, Jia Liu, Jean-Luc Starck, Ken Osato, and Tim Schrabback. Probabilistic Mass Mapping with Neural Score Estimation. *arXiv e-prints*, art. arXiv:2201.05561, January 2022.

[32] Alexandre Adam, Adam Coogan, Nikolay Malkin, Ronan Legin, Laurence Perreault-Levasseur, Yashar D. Hezaveh, and Yoshua Bengio. Posterior samples of source galaxies in strong gravitational lenses with score-based priors. *ArXiv*, abs/2211.03812, 2022. URL https://api.semanticscholar.org/CorpusID:253397812.

[33] Connor Stone and Stéphane Courteau. The Intrinsic Scatter of the Radial Acceleration Relation. *The Astrophysical Journal*, 882(1):6, September 2019. doi: 10.3847/1538-4357/ab3126.

[34] Connor Stone, Stéphane Courteau, and Nikhil Arora. The Intrinsic Scatter of Galaxy Scaling Relations. *The Astrophysical Journal*, 912(1):41, May 2021. doi: 10.3847/1538-4357/abebe4.

[35] Michael J Smith, James E Geach, Ryan A Jackson, Nikhil Arora, Connor Stone, and Stéphane Courteau. Realistic galaxy image simulation via score-based generative models. *Monthly Notices of the Royal Astronomical Society*, 511(2):1808–1818, 01 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac130. URL https://doi.org/10.1093/mnras/stac130.

[36] Connor Bottrell, Hassen M. Yesuf, Gergö Popping, Kiyoaki Christopher Omori, Shenli Tang, Xuheng Ding, Annalisa Pillepich, Dylan Nelson, Lukas Eisert, Hua Gao, Andy D. Goulding, Boris S. Kalita, Wentao Luo, Jenny E. Greene, Jingjing Shi, and John D. Silverman. IllustrisTNG in the HSC-SSP: image data release and the major role of mini mergers as drivers of asymmetry and star formation. *arXiv e-prints*, art. arXiv:2308.14793, August 2023. doi: 10.48550/arXiv.2308.14793.

[37] P. Camps and M. Baes. Skirt 9: Redesigning an advanced dust radiative transfer code to allow kinematics, line transfer and polarization by aligned dust grains. *Astronomy and Computing*, 31:100381, 2020. ISSN 2213-1337. doi: https://doi.org/10.1016/j.ascom.2020.100381. URL https://www.sciencedirect.com/science/article/pii/S2213133720300354.

[38] Dylan Nelson, Volker Springel, Annalisa Pillepich, Vicente Rodriguez-Gomez, Paul Torrey, Shy Genel, Mark Vogelsberger, Ruediger Pakmor, Federico Marinacci, Rainer Weinberger, Luke Kelley, Mark Lovell, Benedikt Diemer, and Lars Hernquist. The IllustrisTNG simulations: public data release. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi: 10.1186/s40668-019-0028-x.

[39] Rainer Weinberger, Volker Springel, Lars Hernquist, Annalisa Pillepich, Federico Marinacci, Rüdiger Pakmor, Dylan Nelson, Shy Genel, Mark Vogelsberger, Jill Naiman, and Paul Torrey. Simulating galaxy formation with black hole driven thermal and kinetic feedback. , 465(3):3291–3308, Mar 2017. doi: 10.1093/mnras/stw2944.

[40] A. Pillepich, V. Springel, D. Nelson, S. Genel, J. Naiman, R. Pakmor, L. Hernquist, P. Torrey, M. Vogelsberger, R. Weinberger, and F. Marinacci. Simulating galaxy formation with the IllustrisTNG model. , 473:4077–4106, January 2018. doi: 10.1093/mnras/stx2656.

[41] Brianna Zawadzki, Ian Czekala, Ryan A. Loomis, Tyler Quinn, Hannah Grzybowski, Robert C. Frazier, Jeff Jennings, Kadri M. Nizam, and Yina Jian. Regularized maximum likelihood image synthesis and validation for ALMA continuum observations of protoplanetary disks. *Publications of the Astronomical Society of the Pacific*, 135(1048):064503, jun 2023. doi: 10.1088/1538-3873/acdf84. URL https://doi.org/10.1088%2F1538-3873%2Facdf84.

[42] James Cooley and John Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[43] Nicolás T. Kurtovic, Laura M. Pérez, Myriam Benisty, Zhaohuan Zhu, Shangjia Zhang, Jane Huang, Sean M. Andrews, Cornelis P. Dullemond, Andrea Isella, Xue-Ning Bai, John M. Carpenter, Viviana V. Guzmán, Luca Ricci, and David J. Wilner. The Disk Substructures at High Angular Resolution Project (DSHARP). IV. Characterizing Substructures and Interactions in Disks around Multiple Star Systems. , 869(2):L44, December 2018. doi: 10.3847/2041-8213/aaf746.

[44] Shangjia Zhang, Zhaohuan Zhu, Jane Huang, Viviana V. Guzmá n, Sean M. Andrews, Tilman Birnstiel, Cornelis P. Dullemond, John M. Carpenter, Andrea Isella, Laura M. Pérez, Myriam Benisty, David J. Wilner, Clément Baruteau, Xue-Ning Bai, and Luca Ricci. The disk

substructures at high angular resolution project (DSHARP). VII. the planet–disk interactions interpretation. *The Astrophysical Journal*, 869(2):L47, dec 2018. doi: 10.3847/2041-8213/aaf744. URL https://doi.org/10.3847%2F2041-8213%2Faaf744.

[45] Pablo Lemos, Adam Coogan, Yashar Hezaveh, and Laurence Perreault-Levasseur. Sampling-based accuracy testing of posterior estimators for general inference, 2023.

[46] Astropy Collaboration, T. P. Robitaille, E. J. Tollerud, P. Greenfield, M. Droettboom, E. Bray, T. Aldcroft, M. Davis, A. Ginsburg, A. M. Price-Whelan, W. E. Kerzendorf, A. Conley, N. Crighton, K. Barbary, D. Muna, H. Ferguson, F. Grollier, M. M. Parikh, P. H. Nair, H. M. Unther, C. Deil, J. Woillez, S. Conseil, R. Kramer, J. E. H. Turner, L. Singer, R. Fox, B. A. Weaver, V. Zabalza, Z. I. Edwards, K. Azalee Bostroem, D. J. Burke, A. R. Casey, S. M. Crawford, N. Dencheva, J. Ely, T. Jenness, K. Labrie, P. L. Lim, F. Pierfederici, A. Pontzen, A. Ptak, B. Refsdal, M. Servillat, and O. Streicher. Astropy: A community Python package for astronomy. *Astronomy and Astrophysics*, 558:A33, October 2013. doi: 10.1051/0004-6361/201322068.

[47] Astropy Collaboration, A. M. Price-Whelan, B. M. Sipőcz, H. M. Günther, P. L. Lim, S. M. Crawford, S. Conseil, D. L. Shupe, M. W. Craig, N. Dencheva, A. Ginsburg, J. T. Vand erPlas, L. D. Bradley, D. Pérez-Suárez, M. de Val-Borro, T. L. Aldcroft, K. L. Cruz, T. P. Robitaille, E. J. Tollerud, C. Ardelean, T. Babej, Y. P. Bach, M. Bachetti, A. V. Bakanov, S. P. Bamford, G. Barentsen, P. Barmby, A. Baumbach, K. L. Berry, F. Biscani, M. Boquien, K. A. Bostroem, L. G. Bouma, G. B. Brammer, E. M. Bray, H. Breytenbach, H. Buddelmeijer, D. J. Burke, G. Calderone, J. L. Cano Rodríguez, M. Cara, J. V. M. Cardoso, S. Cheedella, Y. Copin, L. Corrales, D. Crichton, D. D'Avella, C. Deil, É. Depagne, J. P. Dietrich, A. Donath, M. Droettboom, N. Earl, T. Erben, S. Fabbro, L. A. Ferreira, T. Finethy, R. T. Fox, L. H. Garrison, S. L. J. Gibbons, D. A. Goldstein, R. Gommers, J. P. Greco, P. Greenfield, A. M. Groener, F. Grollier, A. Hagen, P. Hirst, D. Homeier, A. J. Horton, G. Hosseinzadeh, L. Hu, J. S. Hunkeler, Ž. Ivezić, A. Jain, T. Jenness, G. Kanarek, S. Kendrew, N. S. Kern, W. E. Kerzendorf, A. Khvalko, J. King, D. Kirkby, A. M. Kulkarni, A. Kumar, A. Lee, D. Lenz, S. P. Littlefair, Z. Ma, D. M. Macleod, M. Mastropietro, C. McCully, S. Montagnac, B. M. Morris, M. Mueller, S. J. Mumford, D. Muna, N. A. Murphy, S. Nelson, G. H. Nguyen, J. P. Ninan, M. Nöthe, S. Ogaz, S. Oh, J. K. Parejko, N. Parley, S. Pascual, R. Patil, A. A. Patil, A. L. Plunkett, J. X. Prochaska, T. Rastogi, V. Reddy Janga, J. Sabater, P. Sakurikar, M. Seifert, L. E. Sherbert, H. Sherwood-Taylor, A. Y. Shih, J. Sick, M. T. Silbiger, S. Singanamalla, L. P. Singer, P. H. Sladen, K. A. Sooley, S. Sornarajah, O. Streicher, P. Teuben, S. W. Thomas, G. R. Tremblay, J. E. H. Turner, V. Terrón, M. H. van Kerkwijk, A. de la Vega, L. L. Watkins, B. A. Weaver, J. B. Whitmore, J. Woillez, V. Zabalza, and Astropy Contributors. The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package. *Astronomical Journal*, 156(3):123, September 2018. doi: 10.3847/1538-3881/aabc4f.

[48] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, and Jupyter development team. Jupyter notebooks ? a publishing format for reproducible computational workflows. In Fernando Loizides and Birgit Scmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90. IOS Press, 2016. URL https://eprints.soton.ac.uk/403913/.

[49] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9 (3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

[50] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL https://doi.org/10.1038/s41586-020-2649-2.

[51] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-

Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

[52] Casper O. da Costa-Luis. 'tqdm': A fast, extensible progress meter for python and cli. *Journal of Open Source Software*, 4(37):1277, 2019. URL https://doi.org/10.21105/joss.01277.

[53] The CASA Team, Ben Bean, Sanjay Bhatnagar, Sandra Castro, Jennifer Donovan Meyer, Bjorn Emonts, Enrique Garcia, Robert Garwood, Kumar Golap, Justo Gonzalez Villalba, Pamela Harris, Yohei Hayashi, Josh Hoskins, Mingyu Hsieh, Preshanth Jagannathan, Wataru Kawasaki, Aard Keimpema, Mark Kettenis, Jorge Lopez, Joshua Marvil, Joseph Masters, Andrew McNichols, David Mehringer, Renaud Miel, George Moellenbrock, Federico Montesino, Takeshi Nakazato, Juergen Ott, Dirk Petry, Martin Pokorny, Ryan Raba, Urvashi Rau, Darrell Schiebel, Neal Schweighart, Srikrishna Sekhar, Kazuhiko Shimada, Des Small, Jan-Willem Steeb, Kanako Sugimoto, Ville Suoranta, Takahiro Tsutsumi, Ilse M. van Bemmel, Marjolein Verkouter, Akeem Wells, Wei Xiong, Arpad Szomoru, Morgan Griffith, Brian Glendenning, and Jeff Kern. Casa, the common astronomy software applications for radio astronomy. *Publications of the Astronomical Society of the Pacific*, 134(1041):114501, nov 2022. doi: 10.1088/1538-3873/ac9642. URL https://dx.doi.org/10.1088/1538-3873/ac9642.

[54] Ian Czekala, Ryan Loomis, Sean Andrews, Jane Huang, and Katherine Rosenfeld. Mpoldev/visread, January 2021. URL https://doi.org/10.5281/zenodo.4432501.

A Convolved likelihood approximation

A.1 Background

In this appendix, we perform the derivation for the core approximation used in this work to make posterior sampling tractable, namely the convolved likelihood $p_t(\mathbf{y} \mid \mathbf{x})$ in equation (5).

In the context of continuous-time diffusion models introduced by [15], we work with the stochastic process $\mathbf{X}_t(\omega) : \Omega \times [0,1] \to \mathbb{R}^n$, defined on the measurable space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \{\mathcal{F}_t\}, \mathbf{W})$ where $\mathcal{B}(\mathbb{R}^n)$ is the Borel σ-algebra associated with $\mathbb{R}^n$, $\{\mathcal{F}_t\}$ is a set of filtrations and $\mathbf{W}$ is the Wiener measure. We will describe the stochastic process associated with the SDE of the random variable of interest $\mathbf{X}_t$ using the perturbation kernel of the SDE, $p_t(\mathbf{x}_t \mid \mathbf{x}_0)$, where $t \in [0,1]$ is the time index of the SDE. More specifically, we consider the Gaussian perturbation kernel that correspond to the Variance-Preserving (VP) SDE [14, 15]

$$p_t(\mathbf{x}_t \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t \mid \mu(t)\mathbf{x}_0, \sigma^2(t)\mathbb{1}_{n \times n}). \tag{6}$$

In this work, $\mathbf{x}_0$ is an image of the protoplanetary disk we wish to infer from the interferometric data. Since the kernel is Gaussian, we can express $\mathbf{x}_t$, a noisy image of the protoplanetary disk, directly in term of $\mathbf{x}_0$ and pure noise,

$$\mathbf{x}_t = \mu(t)\mathbf{x}_0 + \sigma(t)\mathbf{z}, \tag{7}$$

where $\mathbf{z} \sim \mathcal{N}(0, \mathbb{1}_{n \times n})$.

Since our goal is to sample from the posterior, we specify the $t = 0$ boundary condition of the SDE as the posterior distribution $p(\mathbf{x}_0 \mid \mathbf{y})$. We can construct the marginal of the posterior at any time t by applying Bayes' theorem, taking the logarithm and taking the gradient with respect to $\mathbf{x}_t$. However, the quantity $\nabla_{\mathbf{x}_t} \log p(\mathbf{y} \mid \mathbf{x}_t)$, the second term on the RHS of equation (4), is intractable to compute since it involves an expectation over $p(\mathbf{x}_0 \mid \mathbf{x}_t)$

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int d\mathbf{x}_0 \, p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_0 \mid \mathbf{x}_t). \tag{8}$$

To simplify this expression, we reverse the conditional $p(\mathbf{x}_0 \mid \mathbf{x}_t)$ using Bayes' theorem to get the known perturbation kernel of the SDE

$$p_t(\mathbf{y} \mid \mathbf{x}_t) = \int d\mathbf{x}_0 \, p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_t \mid \mathbf{x}_0) \frac{p(\mathbf{x}_0)}{p(\mathbf{x}_t)}. \tag{9}$$

The convolved likelihood approximation involves in treating the ratio $p(\mathbf{x}_0)/p(\mathbf{x}_t)$ as a constant or setting it to 1. The argument in favor of this simplification is that at low temperature ($t \sim 0$), the prior and the convolved prior are roughly equal, $p(\mathbf{x}_t) \approx p(\mathbf{x}_0)$. Hence, our approximation converges to the true likelihood at low temperature. At high temperature ($t \sim 1$), this approximation is much worse unless $p(\mathbf{x}_0)$ can be treated as a constant over the region where $p(\mathbf{y} \mid \mathbf{x}_0)$ contributes to the integral. In other words, the convolved likelihood approximation is more accurate when the likelihood is informative (narrow).

A.2 Convolution with a complex Gaussian likelihood

We now evaluate the convolution between the likelihood and the perturbation kernel

$$p_t(\mathbf{y} \mid \mathbf{x}_t) \approx \int d\mathbf{x}_0 \, p(\mathbf{y} \mid \mathbf{x}_0) p(\mathbf{x}_t \mid \mathbf{x}_0). \tag{10}$$

The likelihood is a Gaussian distribution with covariance matrix $\Gamma \in \mathbb{R}^{2m \times 2m}$, measured empirically as described in section 3,

$$p(\mathbf{y} \mid \mathbf{x}_0) = \mathcal{N}(\mathbf{y} \mid A\mathbf{x}_0, \Gamma). \tag{11}$$

In what follows, we assume that the physical model, A, is not a singular matrix. We define $\tilde{A} \equiv S\mathcal{F}P_{\text{beam}} \in \mathbb{C}^{m \times n}$ to be the physical model. For the construction in this appendix to work, we redefine this complex matrix into its real-valued equivalent matrix

$$A \equiv \begin{pmatrix} \operatorname{Re}(\tilde{A}) & -\operatorname{Im}(\tilde{A}) \\ \operatorname{Im}(\tilde{A}) & \operatorname{Re}(\tilde{A}) \end{pmatrix} \in \mathbb{R}^{2m \times 2n} \tag{12}$$

In this construction, every complex random variable, e.g., $\tilde{\mathbf{z}} \in \mathbb{C}^n$, is reformulated as a vectorized complex random variable

$$\mathbf{z} = (\text{Re}(\tilde{\mathbf{z}}), \text{Im}(\tilde{\mathbf{z}})) \in \mathbb{R}^{2n} \tag{13}$$

Naturally, we also reformulate real random variables; i.e., $\mathbf{z} \in \mathbb{R}^n$ is mapped to $(\mathbf{z}, 0) \in \mathbb{R}^{2n}$. As such, we can use the same probability rules that apply to real random variables to the complex random variables in the derivation below.

To evaluate the convolution in equation (10), we change the variables of integration for both the likelihood and the perturbation kernel. We first recall the data generating process, equation (1), $\mathbf{y} = A\mathbf{x}_0 + \boldsymbol{\eta}$, where $\mathbf{y} \equiv \mathcal{V}$ are the observed or simulated visibilities. We now multiply equation (1) by $\mu(t)$, then use the reparameterization in equation (7) to get

$$\boldsymbol{\eta}_t \equiv \mu(t)\mathbf{y} - A\mathbf{x}_t = \mu(t)\boldsymbol{\eta} - A\sigma(t)\mathbf{z} . \tag{14}$$

We now extract the form of the perturbation kernel from equation (14). Recall that $\mathbf{z} \sim \mathcal{N}(0, \mathbb{1}_{n \times n})$ is a normally distributed random variable (see appendix A.1). We now cast this variable in the space $\mathbb{R}^{2n}$. In an abuse of notation, we write

$$p(\mathbf{z}) = \mathcal{N}(0, \mathbb{1}_{n \times n} \oplus \mathbb{0}_{n \times n}) , \tag{15}$$

where $\mathbb{0}_{n \times n}$ is a matrix filled with zeros and

$$\mathbb{1}_{n \times n} \oplus \mathbb{0}_{n \times n} \equiv \begin{pmatrix} \mathbb{1}_{n \times n} & \mathbb{0}_{n \times n} \\ \mathbb{0}_{n \times n} & \mathbb{0}_{n \times n} \end{pmatrix} . \tag{16}$$

Despite the fact that we introduce a singular covariance matrix to account for the Dirac delta distribution of $\text{Im}(\mathbf{z}) = 0$, the end product of our derivation will not necessarily be singular. Using this, we can now obtain the perturbation kernel of the random variable $\boldsymbol{\eta}$

$$p(\boldsymbol{\eta}_t \mid \boldsymbol{\eta}) = \mathcal{N}(\boldsymbol{\eta}_t \mid \mu(t)\boldsymbol{\eta}, \sigma^2(t)\Sigma) \tag{17}$$

where

$$\Sigma \equiv A(\mathbb{1}_{n \times n} \oplus \mathbb{0}_{n \times n})A^T = \begin{pmatrix} \text{Re}(\tilde{A})\text{Re}(\tilde{A})^T & \text{Re}(\tilde{A})\text{Im}(\tilde{A})^T \\ \text{Im}(\tilde{A})\text{Re}(\tilde{A})^T & \text{Im}(\tilde{A})\text{Im}(\tilde{A})^T \end{pmatrix} \tag{18}$$

We then rewrite equation (10) in terms of $\boldsymbol{\eta}$ and $\boldsymbol{\eta}_t$. Using equation (14), we can relate the kernel $p(\boldsymbol{\eta}_t \mid \boldsymbol{\eta})$ to $p(\mathbf{x}_t \mid \mathbf{x}_0)$ by making use of the change of variable formula for probability densities, we obtain

$$p(\boldsymbol{\eta}_t \mid \boldsymbol{\eta}) = \frac{1}{|\det A|} p(\mathbf{x}_t \mid \mathbf{x}_0) . \tag{19}$$

This change of variable can only occur if A is not singular. By changing the variable of integration in equation (10) to $\boldsymbol{\eta}$, we introduce another Jacobian determinant, $d\boldsymbol{\eta} = |\det A| d\mathbf{x}_0$. Finally, we make use of the equality $p(\mathbf{y} \mid \mathbf{x}_0) = p(\boldsymbol{\eta})$ to write

$$p_t(\mathbf{y} \mid \mathbf{x}_t) \approx |\det A|^2 \int d\boldsymbol{\eta} \, p(\boldsymbol{\eta})p(\boldsymbol{\eta}_t \mid \boldsymbol{\eta}) . \tag{20}$$

With the approximations made in section (3), we have $|\det A| = 1$. We can thus ignore this determinant factor in what follows. We note that the factor $\mu(t)$ scales the mean and covariance of the likelihood

$$p(\mu(t)\boldsymbol{\eta}) = \mathcal{N}(0, \mu^2(t)\Gamma) . \tag{21}$$

We obtain the distribution of $\boldsymbol{\eta}_t$ by evaluating analytically the convolution implied by the sum of random variable on the RHS of equation (14)

$$p(\boldsymbol{\eta}_t) = \mathcal{N}(\boldsymbol{\eta}_t \mid 0, \mu^2(t)\Gamma + \sigma^2(t)\Sigma) , \tag{22}$$

which we can rewrite as

$$p(\mathbf{y} \mid \mathbf{x}_t) \approx \mathcal{N}(\mu(t)\mathbf{y} \mid A\mathbf{x}_t, \mu^2(t)\Gamma + \sigma^2(t)\Sigma) . \tag{23}$$

In the appendix that follows, we discuss in more detail the structure of the covariance matrix Σ.

A.3 The convolved likelihood for imaging in Fourier space

We now discuss in more detail the structure of the covariance Σ and show how we arrive to the expression of equation (5) with $\Sigma_{jk} = \frac{1}{2}\delta_{jk} + \frac{1}{2}\delta_{j0}\delta_{k0}$. By choosing a flat beam, $P_{\text{beam}} \equiv \mathbb{1}_{n \times n}$, the structure of the covariance is mostly determined by the Fourier operator, $\mathcal{F}$. The effect of the sampling function, S, is only to select or remove rows from the covariance matrix. As such, we invoke this function only at the end of this section. To make things simpler, we only consider the 1D Fourier operator in in order to give an intuition of the behavior for the 2D case. This intuition will extend naturally to the 2D case.

With these simplifications in place, we replace the physical model by the unitary Fourier operator, with elements

$$\tilde{A}_{k\ell} = \frac{1}{\sqrt{n}}\omega^{k\ell}, \qquad k, \ell \in \{0, \ldots, n-1\}. \tag{24}$$

$\omega = e^{-2\pi i/n}$ is the n$^{\text{th}}$-root of unity and $i \equiv \sqrt{-1}$ is the imaginary number. We start by evaluating elements of the first bloc of the covariance Σ (see equation (18)). Using Euler's formula and taking the real part of ω, we get

$$[\text{Re}(\tilde{A})\,\text{Re}(\tilde{A})^T]_{jk} = \frac{1}{n}\sum_{\ell=0}^{n-1}\cos\left(\frac{2\pi j\ell}{n}\right)\cos\left(\frac{2\pi k\ell}{n}\right). \tag{25}$$

By using the trigonometric identity $\cos(\alpha)\cos(\beta) = \frac{1}{2}(\cos(\alpha-\beta) + \cos(\alpha+\beta))$ and the general result for the sum of a geometric series, we obtain

$$[\text{Re}(\tilde{A})\,\text{Re}(\tilde{A})^T]_{jk} = \frac{1}{2n}\sum_{\ell=0}^{n-1}\cos\left(\frac{2\pi\ell(j-k)}{n}\right) + \frac{1}{2n}\sum_{\ell=0}^{n-1}\cos\left(\frac{2\pi\ell(j+k)}{n}\right)$$

$$= \begin{cases} 1, & j = k = 0 \mod n/2 \\ \frac{1}{2}, & j = k \neq 0 \mod n/2 \\ \frac{1}{2}, & j + k = n \\ 0, & \text{otherwise} \end{cases} \tag{26}$$

We get a formula for the other blocs of the covariance matrix using similar arguments

$$[\text{Im}(\tilde{A})\,\text{Im}(\tilde{A})^T]_{jk} = \begin{cases} \frac{1}{2}, & j = k \neq 0 \mod n/2 \\ -\frac{1}{2}, & j + k = n \\ 0, & \text{otherwise} \end{cases}, \tag{27}$$

$$[\text{Re}(\tilde{A})\,\text{Im}(\tilde{A})^T]_{jk} = 0, \tag{28}$$

$$[\text{Im}(\tilde{A})\,\text{Re}(\tilde{A})^T]_{jk} = 0. \tag{29}$$

In summary, we obtain a block-diagonal structure for the covariance. The diagonal components of the real and imaginary blocks are equal to $1/2$, except for their $j = k = 0$ components (sometimes called the DC component) and their component at the Nyquist frequency, $j = k = n/2$. In Figure 4, we show the covariance matrix associated to the 2D Fourier operator with

Figure 4: The covariance matrix Σ corresponding to the 2D Fourier operator for an input in $\mathbb{R}^{(4\times4)}$.

$n = 4 \times 4$, assuming a row-major ordering of the dimensions for an image. This covariance has a similar structure to the 1D case. As before, the diagonal components equal $1/2$ with special cases for the DC and Nyquist components (3 per block in total). Since our images are real-valued, the covariance is unity for the real DC component and null for the imaginary DC component. The null entries of the covariance can be removed with the sampling function since real vectors do not contribute to these Fourier components. As it turns out, the sampling function considered in this work also removes the Nyquist frequencies. This is to say that the size of the pixels considered in section 4 over-samples the signal in the data. As such, these cases can safely be ignored in our final expression.

To speed up computation, we simplify further the covariance matrix by ignoring off-diagonal terms in the covariance. This practical solution has little effects on the results in this work. Thus, our final expression for the covariance is $\Sigma_{jk} = \frac{1}{2}\delta_{jk} + \frac{1}{2}\delta_{j0}\delta_{k0}$, where δ_{jk} denotes the Kronecker delta.
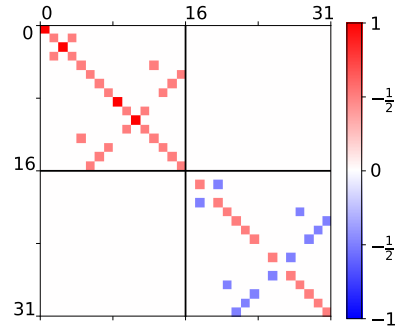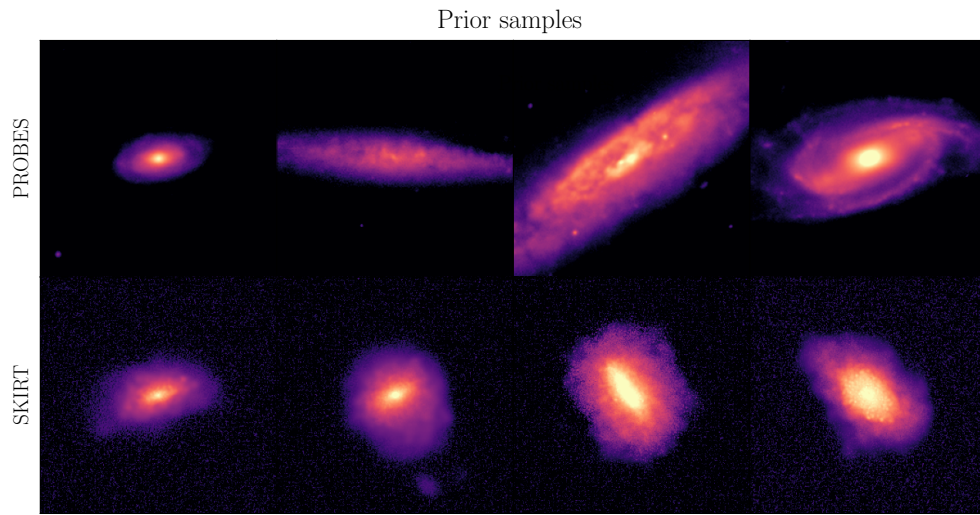
B Additional Figures

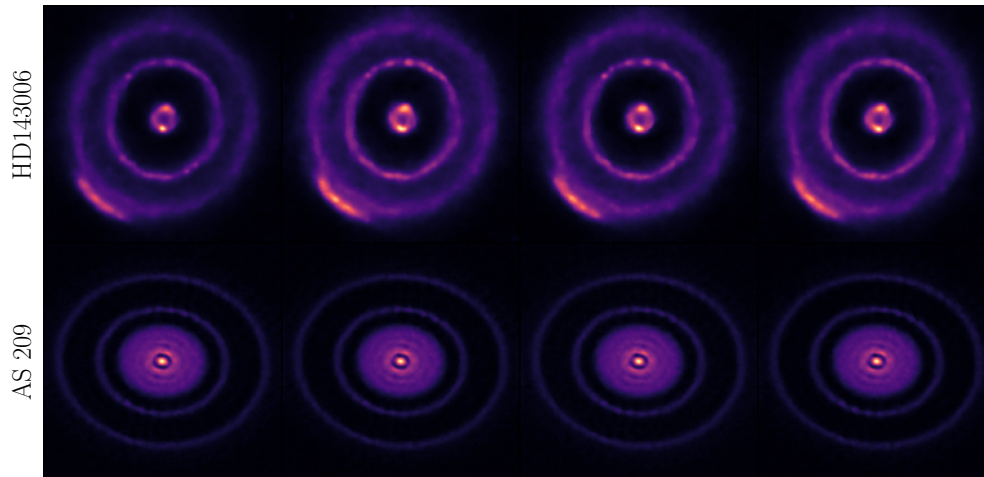Figure 5: Prior samples from the two score-based models used as prior in this work.

Figure 6: Posterior samples of the disks HD143006 (first row) and AS209 (second row) presented in Figure 1.