Publications of the Astronomical Society of Australia (2026), **43**, e026, 15 pages
doi:10.1017/pasa.2026.10167

CAMBRIDGE UNIVERSITY PRESS

Research Article

Conditional image diffusion with interferometric closure invariants: Independent EHT imaging of Centaurus A and 3C 279

Samuel Lai[1], Nithyanandan Thyagarajan[1], O. Ivy Wong[1,2] and Foivos Diakogiannis[3]

[1]Space & Astronomy, Commonwealth Scientific and Industrial Research Organisation, Bentley, Western Australia, 6102, Australia, [2]International Centre for Radio Astronomy Research, The University of Western Australia, Crawley, Western Australia, 6009, Australia and [3]Data 61, Commonwealth Scientific and Industrial Research Organisation, Kensington, Western Australia, 6151, Australia

Abstract

We present independent imaging analyses of Event Horizon Telescope (EHT) observations of the active galactic nuclei in radio galaxy Centaurus A and quasar 3C 279 using Generative Deep learning Image Reconstruction with Closure Terms (GENDIRECT), a recently developed machine-learning framework built on conditional diffusion models that uses interferometric closure invariants as primary observables. For Centaurus A, our reconstruction reveals two prominent emission ridges ($\simeq 80$ μas each) along the jet sheath with a brightness ratio of 1.4 ± 0.1 and an opening angle of 12.3 ± 0.3 deg. For 3C 279, we identify three distinct components in the image, with the southern jet ejecta on sub-parsec scale exhibiting a proper motion of 4.6 ± 1.0 μas over ≈ 5.39 d away from the northern components, corresponding to an apparent superluminal velocity of $\simeq 10 \pm 2$ times light speed. These measurements are consistent with those reported by the EHT Collaboration. The results are significant because we demonstrate that: (1) imaging from interferometric aperture synthesis data, especially in VLBI and most acutely in extremely sparse arrays like the EHT, remains a severely ill-posed and challenging inverse problem, yet closure invariants preserve robust morphological information that can strongly constrain structural features, and (2) more importantly, closure-invariant imaging largely avoids calibration systematics, thus providing a fundamentally independent view of spatial structure with very high angular resolution. The generative nature of GENDIRECT further allows us to sample and characterise clusters of plausible image solutions for each dataset. As a calibration-independent, generative imaging approach, GENDIRECT offers a robust and truly independent blind-imaging tool for current and future VLBI experiments.

Keywords: Methods: data analysis; techniques: image processing; techniques: interferometric

(Received 24 November 2025; revised 5 February 2026; accepted 18 February 2026)

1. Introduction

Very-long baseline radio interferometry is responsible for achieving the highest angular resolution imaging results in the astronomy discipline. By measuring signal correlations, known as visibilities, from pairs of radio receivers sampling the aperture on continental or global scales, one can infer the source intensity distribution by, in the most straightforward case, a two-dimensional inverse Fourier transform of the measured visibilities (Thompson, Moran, & Swenson 2017). However, correlations measured across large spatial distances from a limited array of heterogeneous antennas naturally results in sparse coverage on the aperture plane, which is further complicated by instrumental noise and station-based calibration errors, encompassing amplitude attenuations and phase delays in the signal path both from external influences such as the propagation medium (troposphere, ionosphere, etc.) and from the electronics in the receiver systems.

Accurate image recovery necessitates precise and reliable restoration of the radiation's coherence via signal calibration. Yet, the calibration process itself is often a meticulously fine-tuned iterative process that involves a multi-stage optimisation of both the model image and station-based gains (Pearson & Readhead 1984). Different assumptions adopted during the calibration and deconvolution process can influence the resulting reconstruction, leading to divergent results and a mixture of scientific interpretations. For instance, in the recent observations of the ring-like structure of M87 (Event Horizon Telescope Collaboration et al. 2019b) produced by the Event Horizon Telescope Collaboration (EHTC; Doeleman et al. 2009), the independent analysis of Carilli & Thyagarajan (2022) showed non-negligible variation in the imaged structure produced by a hybrid mapping reconstruction algorithm for different initial models. Likewise, through a markedly different choice of deconvolution window, Miyoshi, Kato, & Makino (2022) demonstrated an alternative reconstruction without a clear ring that is reportedly similarly consistent with the measured visibility data. Multiple other re-analyses of the EHTC data on M87 (e.g. Arras et al. 2022; Broderick et al. 2022; Lockhart & Gralla 2022; Müller 2024; Feng, Bouman, &Freeman 2024), and even independent imaging teams within the EHTC (Event Horizon Telescope Collaboration et al. 2019c), generally produce consistent ring-like morphologies, but finer characteristics (i.e. ring thickness, surface brightness profile, hot spot location, etc.) can vary substantially between different methodologies or prior assumptions, which limits the scientific interpretability of features in the image reconstruction.

Corresponding author: Samuel Lai; Email: samlaihei@gmail.com

Cite this article: Lai S, Thyagarajan N, Wong OI and Diakogiannis F. (2026) Conditional image diffusion with interferometric closure invariants: Independent EHT imaging of Centaurus A and 3C 279. *Publications of the Astronomical Society of Australia* **43**, e026, 1–15. https://doi.org/10.1017/pasa.2026.10167

© The Author(s), 2026. Published by Cambridge University Press on behalf of Astronomical Society of Australia. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

For these reasons, techniques capable of direct imaging from calibration-independent physical observables, free of external fine-tuned hyperparameters, are highly valuable. For decades, it's been known that much of the intricacies of the calibration process can be bypassed by considering specific combinations of interferometric measurements, called 'closure quantities'. Traditionally, closure quantities were constructed in closed triangular loops of stations, which cancels out station-based phase corruption (Jennison 1958; Thyagarajan & Carilli 2022), and quadrilateral loops to cancel out amplitude corruptions (Twiss, Carter, & Little 1960). The closure amplitudes and phases are immune to arbitrarily large station-based multiplicative corruptions, carrying robust information on source properties limited only by additive thermal noise and non-station-based errors (Blackburn et al. 2020; Lockhart & Gralla 2022). An extension of closure quantities to polarimetric measurements, namely 'closure traces', was provided in Broderick & Pesce (2020). In further advancement, a more accurate and unified theory of 'closure invariants' that treats all of them homogeneously is available for co-polar and polarimetric interferometry in Thyagarajan, Nityananda, & Samuel (2022) and Samuel,Nityananda, &Thyagarajan (2022), respectively.

Despite their many desirable characteristics, relying on closure quantities alone instead of traditional visibilities has a few limitations. First, the complete set of closure quantities is always smaller than the full set of visibilities, so some information is inevitably lost, including the absolute amplitude scale and any global phase (overall translation in the image). Second, reconstructing an image directly from closure quantities is an ill-posed inverse problem, because closure quantities relate to the underlying visibilities in a non-linear way. Regularised maximum-likelihood methods (RML; e.g. Ikeda et al. 2016; Akiyama et al. 2017b; Akiyama et al. 2017a; Chael et al. 2018; Blackburn et al. 2020) explored an approach where selected regularisation terms provided the essential additional constraining power to allow images to be produced directly from closure quantities (e.g. Chael et al. 2018). Since then, closure quantities have been used as a constraint in multiple variations of regularised maximum likelihood methods, including techniques based on compressive sensing principles (e.g. Mertens & Lobanov 2015; Müller & Lobanov 2022; Müller 2024) and genetic algorithms for multiobjective optimisation (e.g. Müller, Mus, & Lobanov 2023; Mus, Müller, & Lobanov 2024). Despite this, the resulting image reconstructions are still sensitive to the initialisation of the optimisation problem and hyperparameters introduced by the prior, or selection and relative weighting of regularisation terms (e.g. Chael et al. 2018; Carilli & Thyagarajan 2022), leading to a potentially diverse array of solutions depending on prior assumptions.

Recent advancements in machine learning have demonstrated remarkable performance and fidelity on inverse problems, such as image or video denoising and super-resolution (e.g. Zhang et al. 2017; Rombach et al. 2022; Donike et al. 2025), attracting widespread adoption in astronomy across multiple domains (e.g. Longo, Merényi, & Tiňo 2019; Huertas-Company & Lanusse 2023). In the radio interferometric imaging problem, a large variety of machine learning methods have already been employed, from convolutional neural networks (e.g. Sureau, Lechat, & Starck 2020; Nammour et al. 2022; Schmidt et al. 2022; Chiche et al. 2023; Terris et al. 2023), adversarial networks (e.g. Geyer et al. 2023; Rustige et al. 2023), and normalising flow (e.g. Sun et al. 2022; Feng et al. 2024), to denoising diffusion probabilistic models (e.g. Drozdova et al. 2024; Feng et al. 2024; Lai et al. 2025a). Recently, Lai et al. (2025a) presented a generative diffusion imaging model capable of producing total intensity image reconstructions directly from co-polar closure invariants (Thyagarajan et al. 2022), which is a homogenous and unified alternative to the traditional closure quantities, carrying identical calibration-independent information. Compared to regularised maximum likelihood, the network of Lai et al. (2025a), called GENDIRECT, learns the effective prior and regularisation from its training dataset, containing a variety of natural and non-astronomy images, which eliminates fine-tuned hyperparameters from the imaging sequence and the calibration process altogether, serving as a user-input agnostic imaging technique.

In this work, we present the first application of the Lai et al. (2025a) methodology on real EHT observations. The chosen targets are the active galactic nuclei 3C 279 (Kim et al. 2020) and Centaurus A (Cen A; Janssen et al. 2021), which were observed in 2017 alongside M87. However, despite the public accessibility of the datasets, the reconstructed images from the EHTC have hitherto not been independently verified, to our knowledge, by external groups with alternative imaging techniques. Throughout this work, we present independent reconstructions of both targets and quantitatively compare the results to reference reconstructions from the EHTC, as part of a staged demonstration of the capabilities and reliability of GENDIRECT on real data prior to tackling imaging problems on the event-horizon scale.

The content of this paper is organised as follows: in Sections 2, we briefly describe the Event Horizon Telescope and its observations of blazar 3C 279 and radio galaxy Centaurus A. In Section 3, we describe GENDIRECT, the novel generative deep learning approach to imaging reconstruction with closure invariants. We also describe the validation strategy for trained GENDIRECT models. In Sections 4 and 5, we present the first image reconstructions from GENDIRECT on real EHT observations of 3C 279 and Centaurus A, comparing the results to reference image reconstructions published by the EHTC. We present a summary of this work and conclusions in Section 6. For this study, we adopt a flat ΛCDM cosmology with $H_0 = 67.7\,\mathrm{km\,s^{-1}\,Mpc^{-1}}$ and $(\Omega_m, \Omega_\Lambda) = (0.307, 0.693)$ (Planck Collaboration et al. 2016).

2. The Event Horizon Telescope

The Event Horizon Telescope is a global network of heterogeneous millimeter and sub-millimeter wavelength telescopes separated by long baselines with lengths up to those comparable to the Earth's diameter. The angular scales, $\theta \sim \lambda/D$, probed by the EHT at the nominal operating wavelength of $\lambda \sim 1.3$ mm is $\theta \sim 25\,\mu$as, which is capable of spatially resolving the accretion and jet formation mechanisms near the critical boundaries of several nearby supermassive black holes (e.g. Event Horizon Telescope Collaboration et al. 2019c, 2022).

In 2017, the EHT array was comprised of 8 separate facilities (Event Horizon Telescope Collaboration et al. 2019a): the Atacama Large Millimeter/submillimeter Array (ALMA; Wootten & Thompson 2009; Goddi et al. 2019), the Atacama Pathfinder Experiment telescope (APEX; Güsten et al. 2006), the Large Millimeter Telescope (LMT; Hughes et al. 2010), the Pico Veleta IRAM 30 m telescope (PV; Greve et al. 1995), the Submillimeter Telescope Observatory (SMT; Baars et al. 1999), the James Clerk Maxwell Telescope (JCMT), the Submillimeter Array (SMA; Ho, Moran, & Lo 2004), and the South Pole Telescope (SPT; Carlstrom et al. 2011; Kim et al. 2018). As the Earth rotates,

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

the spatial frequencies sampled by the EHT form tracks in the Fourier plane, resulting in a sparsely sampled Fourier transform of the sky with non-uniform sensitivity.

Despite the challenges, the Event Horizon Telescope Collaboration (EHTC) recently achieved noteworthy scientific results in imaging the black hole shadows of the central supermassive black holes within M87 (Event Horizon Telescope Collaboration et al. 2019c) and Sgr A* (Event Horizon Telescope Collaboration et al. 2022). The images were obtained following a detailed verification process to mitigate both algorithmic and human biases (e.g. Shepherd 2011; Akiyama et al. 2017a; Chael et al. 2018). However, the fine details of the reconstruction depend on different assumptions, initial models, and calibration solutions (e.g. Arras et al. 2022; Broderick et al. 2022; Carilli & Thyagarajan 2022; Lockhart & Gralla 2022; Müller 2024; Feng et al. 2024), which underscores the value of independent imaging techniques in strengthening the reliability of VLBI imaging results.

Outside of the event-horizon scale, the EHT has also been used to observe other active galactic nuclei of interest at 230 GHz, including blazar 3C 279 (Kim et al. 2020) and radio galaxy Centaurus A (Janssen et al. 2021). Notably, although the data is public, the image reconstructions from the EHTC on these targets have, to our knowledge, yet to be independently verified by external groups. In this study, we apply our independent imaging algorithm to both the 3C 279 and Centaurus A EHT datasets and compare the results to the reconstructions from the EHTC. All of the EHT calibrated data products used in this study are from the publicly available data archive on the collaboration's website.[a]

2.1 Blazar 3C 279

3C 279 ($z = 0.5362 \pm 0.0004$; Marziani et al. 1996) is an archetypal blazar, exhibiting rapid time-variability in structure (e.g. Lister et al. 2018) and flux across a wide range of frequencies (e.g. Larionov et al. 2020). It is also one of the first objects with evidence of apparent superluminal motions in the relativistic AGN ejecta revealed by VLBI techniques (e.g. Whitney et al. 1971). Due to its high brightness and the high signal-to-noise ratio (SNR), 3C 279 is an ideal source for robust VLBI fringe detection.

In 2017, observations of 3C 279 were taken with the EHT, interleaved with observations of M87, to independently validate the calibration solution of the M87 image (Event Horizon Telescope Collaboration et al. 2019c). The data were taken over four nights (5, 6, 10, 11) in April at two 2-GHz bands centered at 227.1 (low) and 229.1 GHz (high), where unlike with M87, the SPT was able to participate, albeit at high airmass (Event Horizon Telescope Collaboration et al. 2019c; Kim et al. 2020). Additional details of the observation are described in Kim et al. (2020), hereafter referred to as K20. The frequency-averaged and network-calibrated visibility data produced by the EHT-HOPS pipeline (Blackburn et al. 2019) is publicly available on the collaboration's archive.

In K20, the reference image, reflecting the most robust features detected by several independent imaging methods across teams within the EHTC, consists of two distinct bright emission regions, separated by $\sim 100\,\mu$as (refer to leftmost panel of Figure 2). The northern structure is extended along the NW-SE direction, perpendicular to the orientation of the large-scale jet, while the relative location and elongation of the secondary southern structure is roughly consistent with the jet direction towards the SW (K20). We later refer to the northern structure as the assumed 'core' and the secondary structure as the 'ejecta', following the K20 interpretation. Dramatic inter-day closure phase variations provide strong evidence for rapid variability in the jet structure and surface brightness distribution. Moreover, inter-day difference images show prominent brightness temperature variations in the nuclear region and significant proper motion in the ejected feature. In K20, the imaging problem for 3C 279 was described as more challenging than M87, as a consequence of the source's extended structure and the relative paucity of intermediate baselines in the EHT array (Event Horizon Telescope Collaboration et al. 2019c). Across all of the imaging pipelines, the 3C 279 image was produced within a limited $\sim 100 \times 100\,\mu\text{as}^2$ field-of-view, while a single large-scale Gaussian was used to capture extended emission K20. More recently, observations of 3C 279 by the space VLBI mission, *RadioAstron*, at 22 GHz were published (Fuentes et al. 2023). With its significantly longer (near 10×) space-scale baselines, it achieved angular resolution comparable to that of the EHT and revealed filamentary emission on large ($\sim 500\,\mu$as) scales, possibly threaded by helical magnetic field structure.

2.2 Centaurus A

At a distance of 3.8 ± 0.1 Mpc from Earth (Harris, Rejkuba, & Harris 2010), Centaurus A is the closest radio galaxy, hosting a supermassive black hole which drives its exceptionally bright radio emission. The EHT observed Centaurus A over a six-hour duration track on 10 April 2017 with both low and high bands (Janssen et al. 2021). Details of the data processing and imaging procedure are described in Janssen et al. (2021), hereafter referred to as J21. In brief, the observational data were reduced via two pipelines: rPICARD (Janssen et al. 2019) and EHT-HOPS (Blackburn et al. 2019). For this study, we obtain the EHT-HOPS-reduced data from the EHT collaboration data archive. Imaging in J21 proceeded with a blind imaging challenge undertaken by several individual groups within the EHTC who independently produced twelve images, six of which passed the data fidelity threshold. The six images sourced from a variety of imaging methods converged to the same robust source structure. Subsequently, the EHT-IMAGING script (Chael et al. 2018), used to image M87,[b] was applied to further refine the image.

The final image model presented in J21 appears as a narrow, collimated, and edge-brightened jet with an approaching side that extends towards the NE direction and fainter counter-jet towards the SW. A brightness asymmetry is evident with a flux ratio of $R_{s/n} = 1.6 \pm 0.5$ between the southern and northern ridgelines, most likely caused by relativistic boosting J21. With the high dynamic range of the image reconstruction, J21 empirically located the jet apex and measured the jet collimation profile. Furthermore, by testing modified versions of the image model with and without secondary features using the same data processing and imaging pipeline, J21 confirmed that the counter-jet and extended emission features (out to $\sim 200\,\mu$as from the apex) did not spuriously appear in image reconstructions of simulated observations, lending enhanced confidence in the detection of these fainter features.

[a] https://eventhorizontelescope.org/for-astronomers/data.

[b] https://github.com/eventhorizontelescope/2019-D01-02.

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

3. GenDIReCT imaging

The Generative Deep learning Image Reconstruction with Closure Terms algorithm (GenDIReCT Lai et al. 2025a, hereafter referred to as L25) is a novel imaging pipeline designed using the co-polar closure invariants (Thyagarajan et al. 2022) to tackle the challenge of high-precision calibration and imaging from the sparse aperture coverage characteristic of VLBI data. Building on the earlier work of DIReCT, which showcased the potential of vision transformers in imaging (Lai et al. 2025b), GenDIReCT leverages the conditional image generation (e.g. Rombach et al. 2022) of denoising diffusion probabilistic models (Ho, Jain, & Abbeel 2020) to effectively model the complex distribution of source morphologies conditioned on observed closure invariants. A separate multi-layer convolutional neural network (CNN) produces the final image reconstruction by optimising for the likelihood function containing the selected data fidelity metric. As the latent diffusion model is conditioned with closure invariants, it requires pretraining for each VLBI dataset, while the CNN is optimised for every individual reconstruction. We provide a simplified high-level illustration of the GenDIReCT architecture, imaging sequence, and its outputs in Figure 1, using a model black hole horizon radio image as an example. For a more detailed description of each individual component in the GenDIReCT architecture, their corresponding loss functions, and their training strategy, refer to L25.

Effectively, GenDIReCT replaces the prior and regularisation terms in regularised maximum likelihood-based methods with network-learned priors and regularisations, emulating image characteristics (such as positivity and smooth brightness distribution) from the non-astronomical image training dataset, primarily based on CIFAR-10 natural images (Krizhevsky 2009). Moreover, the generative nature of GenDIReCT enables the imaging pipeline to potentially capture degeneracies and multimodalities in the solution space, by illustrating clusters of candidate solutions for each dataset. However, the integration of the CNN image refinement component does not fit within a Bayesian interpretation; therefore, clusters of GenDIReCT image reconstructions do not directly represent image reconstruction uncertainty.

In the L25 introductory paper, the performance of GenDIReCT was quantitatively evaluated on synthetic datasets derived from the EHT array, achieving excellent performance on data and image metrics for both trained and untrained source morphologies. When tested on the Next Generation Event Horizon Telescope (ngEHT) total intensity analysis challenge, GenDIReCT's performance was competitive with other state-of-the-art image reconstruction algorithms, while not requiring accurate calibration nor hyperparameter fine-tuning. Based on its performance on synthetic datasets, it's expected that GenDIReCT can be an effective blind imaging algorithm on public EHT data, offering an independent constraint on source morphology. However, until this work, GenDIReCT has not been tested on real data from the EHT.

For the first time, we adapt GenDIReCT to real datasets, by modelling the aperture coverage of the calibrated frequency-averaged 3C 279 and Centaurus A data from the EHT-HOPS pipeline (Blackburn et al. 2019), where the visibility data is averaged across ∼5-min scans prior to computing the closure invariants. Models are trained separately for each observing day on identical image datasets containing augmentations of both natural images and simple shapes, as described in L25. Unless otherwise stated, the default field-of-view for 3C 279 is $225 \times 225\,\mu\text{as}^2$ and for Centaurus A a $450 \times 450\,\mu\text{as}^2$ field-of-view was selected. All reconstructed images are produced on a 64×64 pixel grid, which is predetermined by the GenDIReCT architecture. For a discussion on alternative fields-of-view and their effect on the image reconstruction, refer to Appendix 2.1.

Compared to the description of the imaging sequence described in L25, the differences implemented in this study are that we have enlarged the diffusion sample entering the convolutional neural network component by four times, enabling a wider and more diverse sampling of the latent parameter space. To facilitate the larger input, we extended the maximum number of training epochs to allow the convolutional neural network to converge. All other aspects, including the training, optimisation, loss functions, and network architecture remain unchanged from L25.

3.1 Closure invariants

GenDIReCT utilises co-polar interferometric closure invariants (Thyagarajan et al. 2022) to condition the generative latent diffusion model and optimise the CNN image refinement model. Under the Abelian gauge theory formalism, closure invariants are constructed from elementary triangular plaquette variables called 'advariants', which are defined for any pair of array elements (a, b) pinned on a fixed reference vertex indexed at 0,

$$\mathcal{A}'_{0\text{ab}} = \mathcal{V}'_{0a}(\mathcal{V}'^{*}_{ab})^{-1}\mathcal{V}'_{b0} = |\, g_0 \,|^2\, \mathcal{A}_{0\text{ab}}, \tag{1}$$

where $\mathcal{V}_{ab}$ are the corrupted visibilities from the baseline between stations a and b, 0 is the reference station, and $|\, g_0 \,|^2$ is an unknown scaling factor identical on all complex advariants associated with the reference station. The complete and independent set of calibration-independent closure invariants can be constructed by normalising all advariants in each scan by their L_2-norm (other options are possible), resulting in $N_s^2 - 3N_s + 1$ real-valued closure invariants. Throughout this work, we have selected the highest sensitivity station in each scan (often ALMA when available) as the reference station, following the statistical analysis of Blackburn et al. (2020) which proved that constructing triangular closure phases centered around baselines of the most sensitive station station would minimise the closure phase covariances when the array's sensitivity is dominated by a single station, such as in the case of the EHT with ALMA. However, a different selection for the reference vertex station would result in a new set of closure invariants carrying equivalent information. We explore the effect of an alternative selection of reference station on GenDIReCT image reconstructions in Appendix 2.3.

Closure invariants carry identical calibration-independent information as traditional closure phases and closure amplitudes. However, the Abelian gauge theoretic framework for deriving closure invariants is a unified formalism where measurements are derived entirely from independent triads, the simplest non-trivial loop. This carries some advantages compared to the traditional formalism, which necessitates separate treatment for the closed triangular and quadrilateral station loops. Further discussion of the co-polar closure invariants formalism and its advantages are described in Thyagarajan et al. (2022) with the polarimetric non-Abelian extension in Samuel et al. (2022).

3.2 Model validation

Prior to applying each of the trained GenDIReCT models on their corresponding EHT data, we quantify the performance of each model on a diverse set of synthetic datasets. Each of the validation tests can be categorised into one of two types: data corruption

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

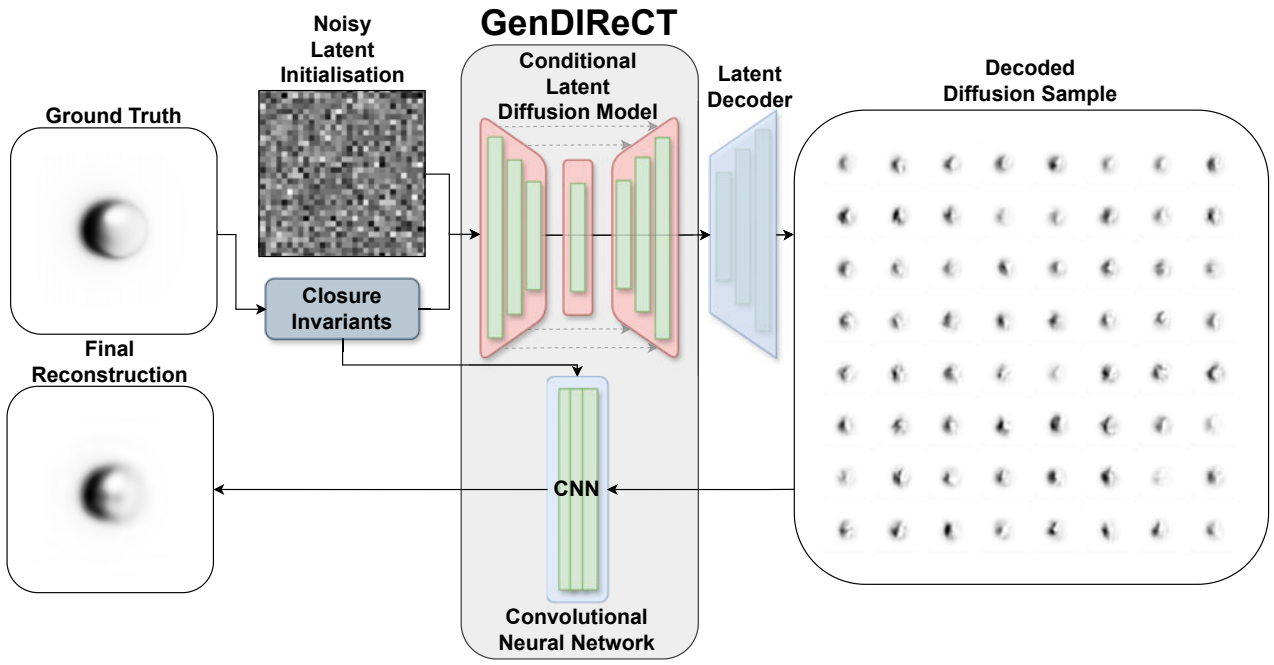


Figure 1. Overview of the GENDIRECT architecture, imaging sequence, and outputs. For a single reconstruction, the GENDIRECT model receives a dataset of closure invariants, which can also be derived from synthetic observation of a ground truth image. The pretrained conditional diffusion component samples images decoded from a distribution in the latent feature space, conditioned on the provided data. Subsequently, the convolutional neural network processes the diffusion sample and optimises for consistency with the input closure invariants, resulting in the final image reconstruction.

and image tests. In the data corruption category, we test the performance of the model under the effects of two additional sources of noise that are not removed by the closure invariants construction: thermal noise and baseline-dependent errors, as well as data corruption coming from an extended source with emission outside the reconstruction field-of-view. In the image test category, we evaluate the performance of the model on basic geometric shapes, selected images from IMAGENET (Deng et al. 2009), the standard MNIST handwritten digits (Deng 2012), and astrophysically relevant images. On each of the challenges, we quantify the performance of the model on the following metrics, which evaluate both the data and image reconstruction fidelity.

3.2.1 Fidelity metrics

In this work, we employ the reduced $\chi^2_{\rm CI}$ metric to evaluate data consistency between different image reconstructions. The reduced $\chi^2_{\rm CI}$ is defined as the sum of error-normalised square residuals, divided by the number of independent data terms,

$$\chi^2_{\rm CI} = \frac{1}{N_{\rm ci}} \sum_{i}^{N_{\rm ci}} + \left[\frac{(\mathcal{C}(A)_i - \mathcal{C}(B)_i)^2}{\sigma_i^2} \right], \tag{2}$$

where the operator $\mathcal{C}$ maps an image to its closure invariants, indexed by i, for a fixed observation arrangement, and $N_{\rm ci}$ is the total number of independent closure invariants. The normalisation, σ_i, is the expected uncertainty of the i-th closure invariant, which is sensitive to the noise model and observation details. Where the ground truth image is unknown, $\mathcal{C}(B)$ is replaced with the observed data, which is compared to $\mathcal{C}(A)$ for the reconstruction, A. Different decisions, such as the reference station selection or data aggregation strategy, can impact the size of the resulting set of closure invariants and their covariances. As such, in all cases throughout this paper, when $\chi^2_{\rm CI}$ is measured with EHT data, we use the public EHT-HOPS-reduced data with no additional post-processing. This ensures that all of the $\chi^2_{\rm CI}$ data metrics reported for each dataset are evaluated on a consistent metric, regardless of the data processing decisions.

While thermal errors are relatively straightforward to propagate from each radio station's system equivalent flux density (SEFD), other sources of systematic errors, such as residual calibration offsets, time-dependent errors from long coherent averaging of visibilities, and limitations of a linear error propagation model, can also affect both visibilities and closure invariants. In some cases and particularly at high SNR, the systematic uncertainty can become dominant over more easily quantifiable thermal noise. In this work, we adopt a relatively conservative homogeneous systematic uncertainty estimate of 5% in addition to the measured thermal noise, which is on the same order as the measured inter-pipeline differences and aligned with the recommended systematic error budget of the 3C 279 and M87 data (Event Horizon Telescope Collaboration et al. 2019b).

In validation tests where the ground truth is known and for relative comparisons between image reconstructions, we introduce the maximum normalised cross-correlation metric, $\rho_{\rm NX}$, to evaluate the image fidelity or correspondence between reconstructions independently from the data fidelity metric. The maximum normalised cross-correlation metric is defined as,

$$\rho_{\rm NX}(A, B) = \frac{1}{M} \max | \mathcal{F}^{-1}\{\mathcal{F}\{\hat{A}\}\mathcal{F}\{\hat{B}\}^*\} | , \tag{3}$$

where A and B are images and the hat operator, $\hat{I} = (I - \bar{I})/\sigma_I$, is an image, I, normalised by its self-evaluated mean and standard deviation. Operations $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent the forward and inverse

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

Fourier transforms, respectively. By taking the maximum value, $\rho_{\rm NX}$ is independent of image translations and changes in the total flux level, which is ideal for evaluating image reconstructions from closure information where these properties are unconstrained. We also employ the $\rho_{\rm NX}$ metric to determine the corresponding image translation for aligning separate image reconstructions by their maximum cross-correlation. We add that when estimating the relative correspondence between reconstructions, both images would be blurred by the fitted uniformly-weighted clean beam prior to computing $\rho_{\rm NX}$. We choose to compare $\rho_{\rm NX}$ after a final convolution with the nominal image resolution due to the heterogeneous origin of EHTC reference reconstructions, where the Centaurus A reference is a single refined image reconstruction, while the 3C 279 reference is a combined reconstruction from three separate imaging algorithms. Generally, in this work, we represent the GENDIRECT output with the median reconstruction, which is an aggregated representation of numerous individual reconstructions, similar to the latter case of 3C 279. By smoothing out high-resolution features, such as those present in the EHTC reference image of Centaurus A, with the nominal image resolution, we can construct a more equitable comparison between image reconstruction methods using the $\rho_{\rm NX}$ correspondence metric.

3.3 Performance on validation tests

In this section, we summarise the performance of two models on both image and data corruption tests: one model is trained with the 11 April 2017 dataset of 3C 279 to produce images on a $225 \times 225\,\mu{\rm as}^2$ field-of-view and the other model is trained with the Centaurus A dataset on a $450 \times 450\,\mu{\rm as}^2$ field-of-view. Note that the 11 April 2017 dataset of 3C 279 has the most complete aperture coverage of the four observation dates. Across all validation tests, we find that both models can consistently achieve $\chi^2_{\rm CI} < 2$ on the data fidelity metric, except in cases of severely corrupted data and significant flux outside the reconstruction field-of-view.

First, we describe the performance of both GENDIRECT models on data corruption tests, which include scaling the thermal noise, the addition of baseline-dependent errors, and an extended source with emission outside the reconstruction field-of-view. Data corruption validation tests are performed with synthetic data on a simulated event-horizon-scale image with a total compact flux density of 1 Jy. The choice of image can shift the nominal performance of the GENDIRECT models; thus, we place no emphasis on the relative performance between the trained models and examine only the trends of a single model with different levels of noise. On thermal noise tests, we scale the thermal noise from 0.01–100× the standard SEFD, and evaluate both image and data fidelity by the $\rho_{\rm NX}$ and $\chi^2_{\rm CI}$. In all cases, the expected σ^2 normalisation of each $\chi^2_{\rm CI}$ measurement is based on the standard SEFD level for a consistent comparison. We find that all models retain consistent performance on $\rho_{\rm NX}$ up to $\sim 3\times$ SEFD; after which, $\rho_{\rm NX}$ indicates that the reconstruction is unreliable. Correspondingly, the models' performance on $\chi^2_{\rm CI}$ worsen gradually as noise level increases, passing $\chi^2_{\rm CI} \sim 1$ between ∼3 and 30× standard SEFDs.

We define the baseline-dependent errors as an additional zero-mean Gaussian corruption applied to each baseline, with its standard deviation defined as some fraction of the visibility amplitude. Like the thermal noise test, we find consistent performance in $\rho_{\rm NX}$ and gradually degraded performance on $\chi^2_{\rm CI}$. The critical threshold for image fidelity is when the baseline noise reaches 10% of the visibility amplitude, while the data fidelity passes $\chi^2_{\rm CI} \sim 1$ at 30%. For the validation test on the extended source with emission beyond the reconstruction window, we first construct a crescent within the field of view and define a large-scale disk with a diameter 2× the size of the reconstruction window width and scale the relative total flux between the two components. In all models, we find that GENDIRECT maintains consistent performance on both image and data metrics as the relative flux of the large-scale disk component grows until the flux ratio becomes unity. Once the extended disk becomes the dominant component, it's not possible for GENDIRECT to reconstruct the disk within its limited field-of-view, which significantly degrades its performance on all metrics. Therefore, as long as the majority of flux is contained within the limited reconstruction window, GENDIRECT can produce an accurate image reconstruction.

Image validation tests target both general out-of-distribution morphologies, such as MNIST digits and IMAGENET, as well as basic geometric shapes and astrophysically relevant images. All synthetic data for these tests are created with thermal noise from standard SEFD values. As we have previously established that GENDIRECT has minimal difficulty fitting the $\chi^2_{\rm CI}$ data metric for standard SEFDs, we focus primarily on the image fidelity metric for this discussion. We first note that the test on basic geometric shapes, which includes Gaussians, rings, crescents, and doubles, establishes a baseline performance for each model as it tests the trained model's performance on structures similar to those in its training dataset. On this test, we note that all models perform well with image fidelity scores reaching $\rho_{\rm NX} \sim 0.95$. We then test the models on the out-of-distribution sources, such as handwritten digits from the MNIST dataset, natural images from IMAGENET, and a variety of astrophysically relevant images, which include images of a protoplanetary disk, protostellar cluster, protostellar envelope, M51 Hα, g41 continuum, 3C288 radio galaxy, and 30Dor in the near-infrared. All images are collected from the CASA data-bank[c] and resized to the field-of-view of the trained model. While some individual reconstructions reach $\rho_{\rm NX} \sim 0.95$, most reconstructions fall within the range of $\rho_{\rm NX} \sim$ 0.8–0.9. Out of the three datasets, GENDIRECT performs the best on MNIST due to its clean zero-value background. In summary, across all tests, the trained GENDIRECT models are performing within expectations, indicating that the closure invariant datasets are sufficiently constraining for a variety of possible image morphologies.

4. Results

With the GENDIRECT models trained and validated as described in Section 3.3, we present our results on 3C 279 (top) and Centaurus A (bottom) in Figure 2. For each dataset, we run GENDIRECT for 66 GPU-hours on the NVIDIA-H100, which produces several hundred image reconstructions (∼300), although the exact number varies by ∼10% depending on the dataset. The leftmost panels in each row contain the reference reconstructions from the EHTC (Kim et al. 2020; Janssen et al. 2021), followed from left-to-right by the best single reconstruction from the GENDIRECT pipeline as measured by the minimum value of closure invariants reduced $\chi^2_{\rm CI}$, the median of all GENDIRECT reconstructions on the respective dataset aligned by $\rho_{\rm NX}$, the median absolute deviation (MAD), and the ratio image between the

[c] https://casaguides.nrao.edu/index.php/Sim_Inputs.

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

median image and the median absolute deviation. The maximum normalised cross-correlation between the median GENDIRECT reconstruction and the EHTC reference reconstruction is shown on the median panel. We note that although we use the aligned median reconstruction throughout this study to aggregate the generative results of GENDIRECT into a single image, the median is not intended to reflect the best-performing image in metrics $\chi^2_{\rm CI}$ or $\rho_{\rm NX}$. However, it's likely the most suitable image for comparison with the EHTC references, which are likewise informed by multiple pipelines under different imaging conditions, aligned, and averaged (Kim et al. 2020; Janssen et al. 2021). The right-most ratio panel in each row illustrates a 'signal-to-noise' ratio of the reconstruction with darker regions reflecting more consistently reconstructed features.

The top row of Figure 2 corresponds to reconstructions of 3C 279 in a 225 μas × 225 μas field-of-view, specifically reconstructions resulting from the 11 April 2017 dataset, which is the dataset with the best single-day aperture coverage (K20). The median of GENDIRECT reconstructions exhibits a $\rho_{\rm NX} = 0.98$ correspondence with the EHTC reference reconstruction, though there are apparent differences in the significance of the flux bridging the assumed core (northern component) and ejecta (southern component), as well as in the presence of more spurious emission on both the east and west sides. While some flux is marginally detectable in the EHTC reference reconstruction between the bright components, it is significantly fainter than in the GENDIRECT median reconstruction. However, as shown in the median/MAD ratio image, the reconstruction confidence is low in both the emission bridge and the flanking emission compared to the two bright components.

In the bottom row of Figure 2, the median GENDIRECT reconstruction of Centaurus A has a maximum normalised cross-correlation score of $\rho_{\rm NX} = 0.91$ when compared with the EHTC reference model reconstruction. While the highest surface brightness components of the edge-brightened approaching jet were reproduced by the GENDIRECT reconstruction, the counter-jet emission and some of the fainter features extending out to angular distances of ∼ 200 μas from the jet apex are faint or absent. Instead, spurious emission that is not well-constrained by the data can be observed near the top of the reconstruction window in some of the GENDIRECT reconstructions. The normalised cross-correlation score of Centaurus A relative to the EHTC reference reconstruction was estimated after clipping the spurious emission near the top of the GENDIRECT image frame in order to compare the morphologies of the central compact flux distribution. According to the ratio image, the only consistently reconstructed features from GENDIRECT are the two bright ridge-lines along the sheath and near the base of the approaching jet and the overall reconstruction confidence is lower than that of 3C 279.

In Appendix 2, we explore the effects of different data processing and pre-imaging considerations, such as variations in the field-of-view, alternative data aggregation strategies, and a different selection of the closure triad reference station, on the resulting GENDIRECT image reconstruction. We show that the image reconstructions shown in this section are robust against sensible modifications to the reconstruction window and data processing methods. Throughout this section, image reconstructions from both selections of the reference station contribute to the GENDIRECT reconstruction sample, but reconstructions from alternative fields-of-view or with averaging closure invariants are excluded.

4.1 Reconstruction multimodalities

As a consequence of our generative approach, we are able to identify data-consistent, but perceptually diverse multimodalities in the resulting image reconstruction, similar in effect to normalising flow techniques (e.g. Feng et al. 2024) or evolutionary algorithms (e.g. Müller et al. 2023). Clusters of similar reconstructions are identified by the Hamming distance in the discrete cosine transform hash, as described in Appendix 1. Figure 3 illustrates five of the most populated image reconstruction clusters of 3C 279 (top row) and Centaurus A (bottom row), where each cluster is represented by its median image and displayed in descending order (left-to-right) by the share of all reconstructions belonging to the cluster. In identifying clusters of Centaurus A reconstructions, we have applied a Gaussian filtered circular mask to remove some spurious low surface brightness emission typically located near the image periphery, as observed in Figure 2.

While the image clusters of 3C 279 all show at least two extended components with a consistent relative angular separation, the clusters differ in the brightness ratio between the assumed core and ejecta, as well as in the number and flux of any additional components in the intervening space. The dominant cluster with $> 60\%$ share of all reconstructions is the most morphologically consistent with the EHTC reference reconstruction, while several of the alternative solution clusters, such as Clusters 2, 3, and 4, exhibit brighter emission trailing towards the SW direction relative to the core. Often these additional components are accompanied by enhanced faint emission in the western direction of the reconstruction window, from which we can infer that any sparsity regularisation would heavily penalise such solutions.

The dominant cluster for Centaurus A is comprised of a $>50\%$ share of all reconstructions. Alternative solution clusters primarily differ from the dominant cluster in the brightness ratio between the two approaching jet ridges and the low-luminosity extension of the jet stream, although most solutions have the southern jet sheath as the brighter feature. Certain clusters, such as Cluster 1 for 3C 279 as well as in Clusters 2 and 4 for Centaurus A, exhibit phantom repetition of similar structures. These secondary phantom images are not unusual in the VLBI context, as they are symptoms of the poor aperture coverage and phase error (i.e. Chael et al. 2018; Müller et al. 2023).

We emphasise that while the GENDIRECT approach can be used to identify and explore clusters of of candidate reconstructions, the process is not strictly identical to sampling from a posterior distribution in a Bayesian framework. Rather, the diffusion model of GENDIRECT is trained to emulate the expected conditional probability distribution of its training data in the latent domain. Then the convolutional neural network further optimises an aggregated image reconstruction from the diffusion-generated sample according to the error-weighted measurements. While the generative output of this process contains clusters of candidate image reconstructions, the image distribution does not allow for quantification of statistical uncertainty.

5. Discussion

As an imaging algorithm conditioned on only closure information, the results of GENDIRECT are insensitive to different sets of calibration assumptions or residuals in the calibration solution, enabling an independent verification of image morphologies obtained through VLBI methods. Moreover, absent of explicit priors, initialisations, and regularisations, GENDIRECT provides a

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

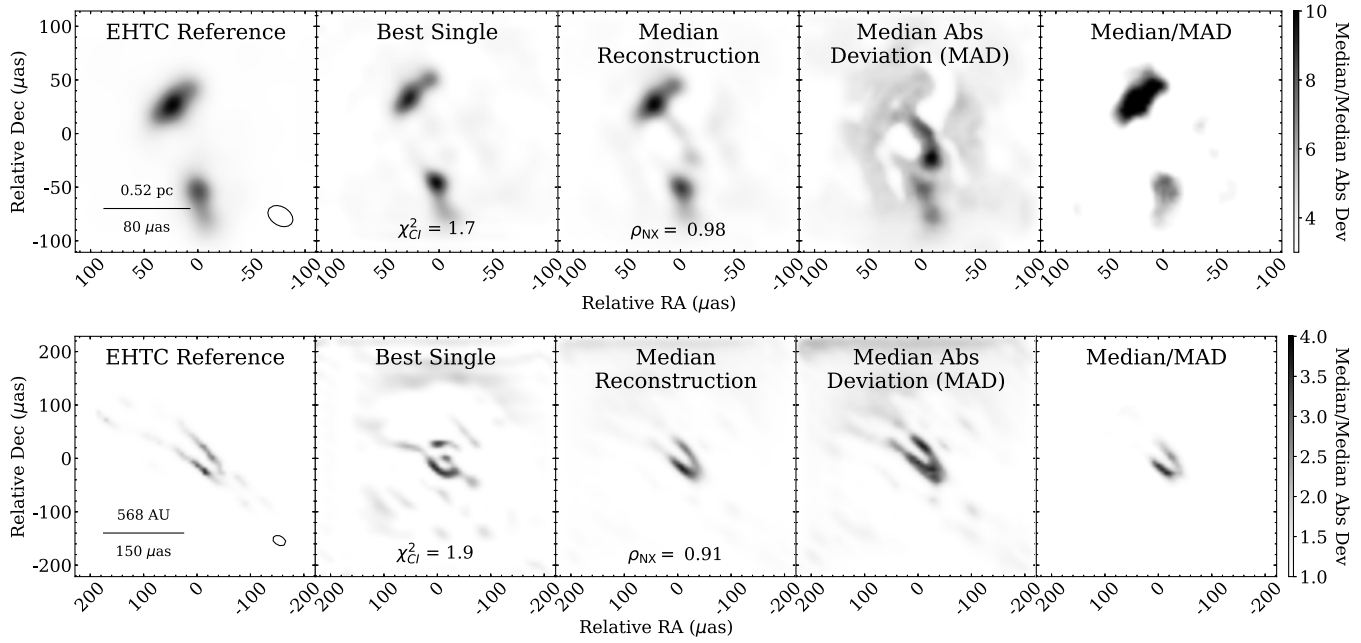


Figure 2. Comparison between the EHTC reference reconstructions of blazar 3C 279 (top) and Centaurus A (bottom) with their corresponding GENDIRECT reconstructions. The dataset used in the reconstruction of 3C 279 is from 11 April 2017. GENDIRECT outputs include the best single reconstruction as measured by the reduced $\chi^2_{\rm CI}$, the median reconstruction, the median absolute deviation, and ratio images. The ratio image displays the median image over the median absolute deviation, illustrating reconstruction confidence. The synthesised beam is visualised in the left panel with an elliptical model. All 3C 279 reconstructions are plotted on a 225 × 225 μas^2 field-of-view, while Centaurus A reconstructions are pictured on a 450 × 450 μas^2 field-of-view. All GENDIRECT reconstructions (and the EHTC Centaurus A reference image) are presented without clean beam convolution, while the EHTC reference for 3C 279 has previously been convolved with a 20 μas circular beam.

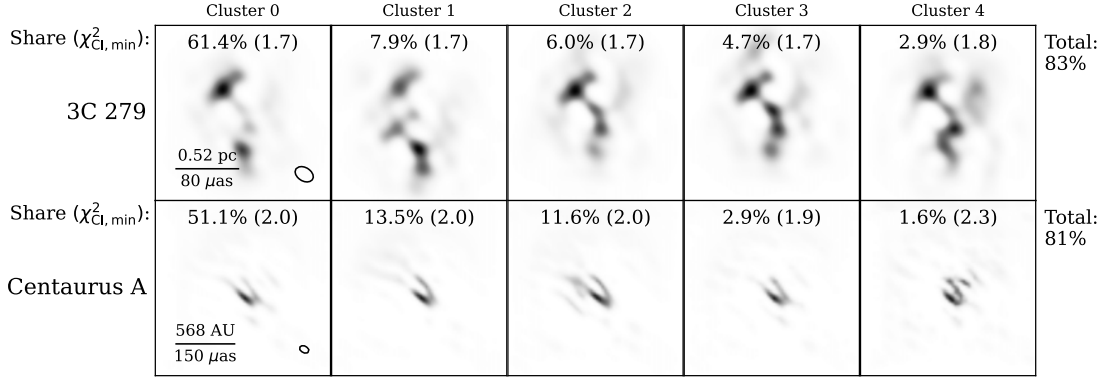


Figure 3. Reconstruction clusters for the EHT observations of 3C 279 (top row) and Centaurus A (bottom row). Each cluster is represented by the median image and we show the percentage share alongside the minimum $\chi^2_{\rm CI}$ (in parenthesis) of all reconstructions identified with that cluster. The first image of each row shows the linear scale and fitted clean beam.

relatively simple and reproducible imaging tool for VLBI datasets, which is also robust against several different pre-imaging considerations as shown in Appendix 2. With only minimal and sensible modifications to the size of the diffusion sample and maximum number of epochs for convergence, we have applied the methodology and imaging sequence of L25 on real EHT data for the first time in this work.

In the previous section, we presented the image reconstructions of the GENDIRECT model trained on the 11 April 2017 dataset of 3C 279 and 10 April 2017 dataset of Centaurus A. We illustrated the multimodalities in the resulting image reconstructions by clustering candidate solutions and presented an aggregated median reconstruction, alongside the median absolute deviation, and ratio images. We measured the best single reconstruction data fidelity $\chi^2_{\rm CI}$ and maximum cross-correlation, $\rho_{\rm NX}$ with the EHT reference images, finding consistent results with the published images. In this Section, we present a more detailed morphological comparison between the GENDIRECT reconstructions and the EHTC images. Additionally, we study the other EHT datasets of 3C 279 taken on different days in April 2017, and evaluate the inter-day variability through independent GENDIRECT image reconstructions.

5.1 Morphological comparison

We have shown in Figure 2 that the GENDIRECT median reconstructions of both 3C 279 and Centaurus A are rated at $\rho_{\rm NX} > 0.9$

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

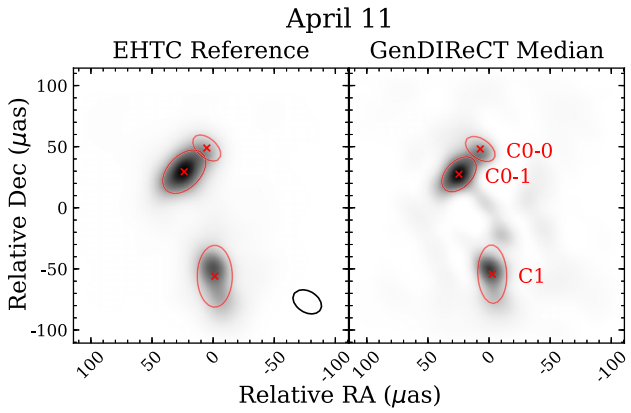


Figure 4. Illustration of the three elliptical Gaussian component fits of both the EHTC reference (left) and median GENDIRECT (right) reconstruction on the 11 April 2017 dataset of 3C 279. The centroid of each component is marked with an 'x' and a single contour is drawn at the half-amplitude.

on the maximum normalised cross-correlation metric when compared to the EHTC reference reconstructions. In this section, we quantitatively compare measurable morphological characteristics between the EHTC and GENDIRECT reconstructions by applying identical parameterised fitting algorithms to both images.

We jointly fit the reconstructions of 3C 279 with three elliptical Gaussian components: two for the extended core and a single component for the ejecta, applying the Levenberg-Marquardt algorithm for parameter estimation. We adopt the nomenclature used in K20 to label each of the components. Note that while K20 fits 6 Gaussian components in the visibility model-fitting analysis, we combine the C1-0, C1-1, and C1-2 components into a single C1 component, and do not attempt to fit C0-2, which is faint and likely undetectable in the GENDIRECT reconstruction. Figure 4 illustrates the three fitted components on both the K20 reference and median GENDIRECT reconstructions, with the centroid of each component marked by an 'x' and contour drawn to illustrate the Gaussian half-amplitude. We measure the intensity ratio between C0-1 and C1 ($I(\mathrm{C0}-1)/I(\mathrm{C1})$), separation ($|\,\mathbf{C0}-\mathbf{C1}\,|$), and angle ($\theta_{\mathbf{C1-C0}}$) between the flux weighted C0 and C1, as well as the angular extension of the C0-1 component ($\mathrm{PA}_{\mathbf{C0\text{-}1}}$). The measured quantities are tabulated and compared in Table 1. Note that $\theta_{\mathbf{C1-C0}}$ and $\mathrm{PA}_{\mathbf{C0\text{-}1}}$ are reported in degrees north of west. We find that all measured quantities to be statistically similar within $\sim 1\sigma$ uncertainty intervals, except for the angular separation between C1 and the flux-weighted centroid of C0. However, the angular discrepancy is ~ 0.7 μas, which is insignificant compared to the clean beam semi-major axis (~ 25 μas) and the pixel scale of the GENDIRECT reconstruction (~ 3.5 μas/pixel). If one considers adding in quadrature a systematic position error of 5% the clean beam geometric mean half-width at half-maximum, the angular separation discrepancy is no longer statistically significant. Altogether, these results effectively demonstrate that the robustly detected core and ejecta components of 3C 279 are morphologically similar within expected uncertainties between the EHTC reference and median GENDIRECT reconstruction, which is supported by the image correspondence metric, $\rho_{\mathrm{NX}} = 0.98$.

For Centaurus A, we jointly fit each of the edge-brightened ridges with separate linear profiles and consider the intensity ratio between the north-south ridgelines as well as the opening angle, which we define as the angle formed between the two fitted lines. We show both the J21 reference and GENDIRECT image

Table 1. Tabulated quantities comparing fitted morphological parameters for both 3C 279 and Centaurus A.

	Parameter	EHTC reference	GENDIRECT median
3C 279	$I(\mathrm{C0}-1)/I(\mathrm{C1})$	1.52 ± 0.01	1.53 ± 0.02
(K20)	$\lvert\,\mathbf{C0}-\mathbf{C1}\,\rvert$ (μas)	90.9 ± 0.2	90.2 ± 0.2
	$\theta_{\mathbf{C1-C0}}$ (deg*)	-75.7 ± 0.1	-75.6 ± 0.1
	$\mathrm{PA}_{\mathbf{C0\text{-}1}}$ (deg*)	44.5 ± 0.5	44.2 ± 0.7
CenA	$R_{\mathrm{s/n}}$	1.5 ± 0.2	1.4 ± 0.1
(J21)	$\theta_{\mathrm{s/n}}$ (deg)	12.2 ± 0.1	12.3 ± 0.3

* Measured north of west.

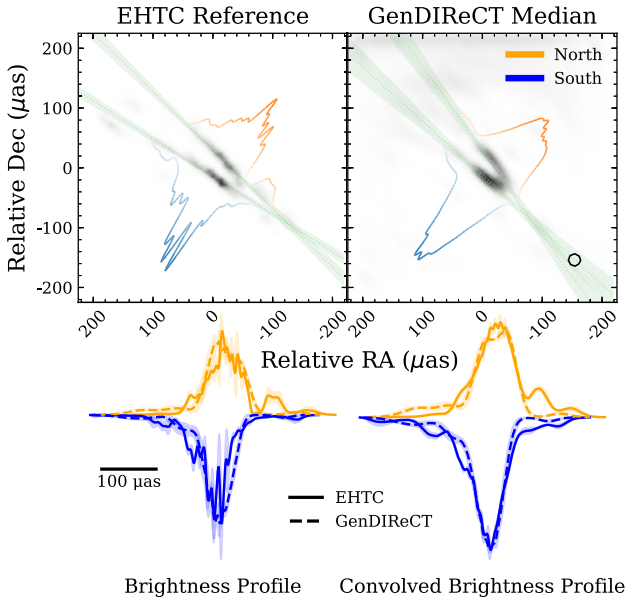


Figure 5. Illustration of the EHTC reference (top-left) and median GENDIRECT (top-right) reconstructions fitted with straight lines for each of the northern and southern edge-brightened jet ridgelines. The surface brightness along the fitted ridgeline is represented by the distribution plotted orthogonally to each line. In the bottom row, we compare the northern and southern surface brightness profiles between the EHTC (solid line) and GENDIRECT (dotted line) reconstructions. The bottom-right panel compares the profiles for both reconstructions after convolving with a common 20 μas circular beam, which is illustrated on the GENDIRECT median image.

reconstructions fitted with the double linear model in Figure 5, alongside a comparison of the brightness profiles normalised by the peak flux. The linear fits are performed on both reconstructions blurred by a common 20 μas circular beam. From the fitted model, we measure the brightness asymmetry ($R_{\mathrm{s/n}}$) between the two jet ridges and the opening angle ($\theta_{\mathrm{s/n}}$), tabulating the results in Table 1. We find no statistically significant discrepancies between the two fitted models on either quantity; however, we note that while the opening angle is near-identical, the two fitted line models appear to be rotated by $\sim 10 \pm 4$ deg with respect to each other. This is likely caused by the extended lower-luminosity features located further away (~ 200 μas) from the jet apex, which are present in the EHTC reference image, but significantly fainter or altogether absent in the GENDIRECT reconstruction. The extended emission drives the fitted model slope shallower in the EHTC reference reconstruction compared to the median GENDIRECT reconstruction. We also find the Centaurus A jet to exhibit more high-frequency structure in the Janssen_2021_CenA

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

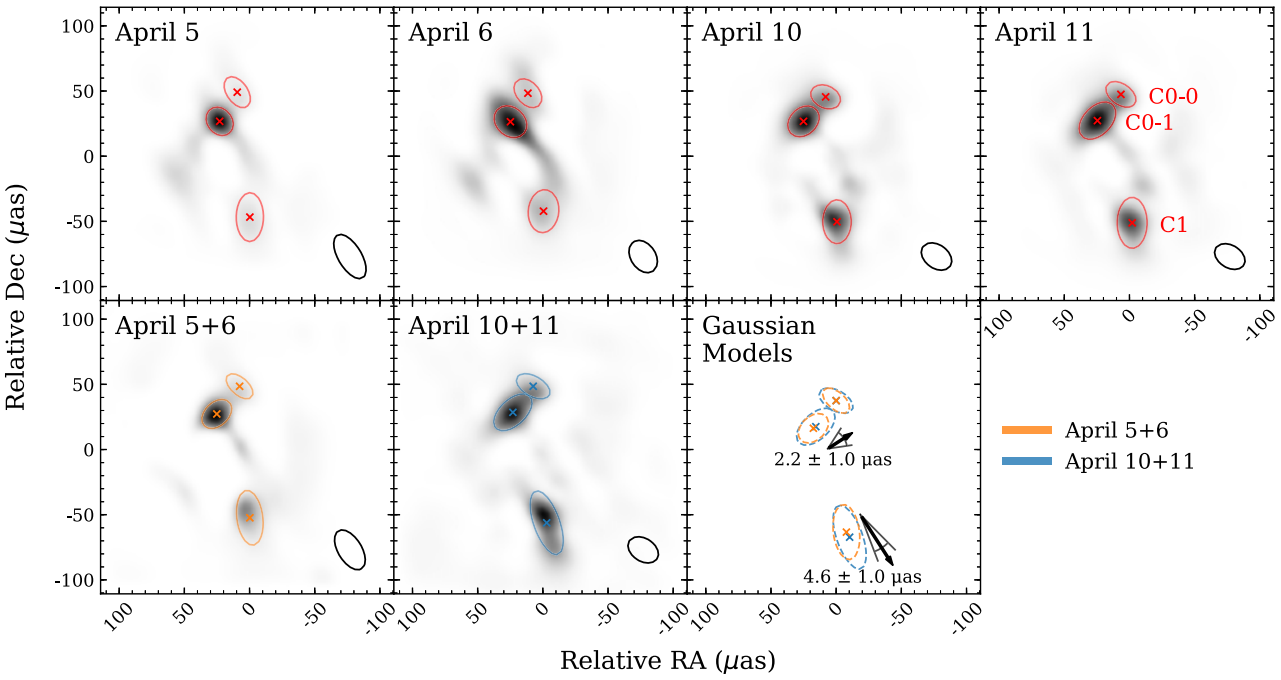


Figure 6. (Top row) Median GENDIRECT reconstructions of 3C 279 on each observation date. (Bottom row) Median GENDIRECT reconstructions using the combined dataset of the first and latter two observations dates. Each image is fitted with three Gaussian components, where the centroid is marked and a single contour is plotted at the half-maximum flux value. The components are labeled similarly to the designations from Kim et al. (2020). The final bottom-right panel plots the centroid and contours of the fitted Gaussian components in the combined dataset with C0-0 as the kinematic reference. Arrows point in the direction of proper motion and wedges illustrate the 1σ uncertainty in direction.

jet model, but differences in the brightness profiles in the proximity of the brightest components become insignificant once both images are convolved with a common 20 μas beam.

5.2 3C 279 inter-day variability

During the 2017 EHT observing campaign, 3C 279 was observed on four separate days between April 5 and 11. We chose to image 3C 279 with the April 11 data in Section 4, as it is the day with the best single-day aperture coverage. However, J21 successfully imaged 3C 279 using all of the datasets in the 2017 observing campaign and robustly identified six compact bright features through a Gaussian model-fitting analysis (THEMIS; Broderick et al. 2020). With the Gaussian component parameters, K20 observed prominent and rapid inter-day brightness variations, as well as relative proper motion between components of $\simeq 1$ μas day^{-1}, corresponding to apparent superluminal velocities of $\beta_{\rm app} = \frac{v_{\rm app}}{c} \sim (13-15) \pm 2$.

In this section, we independently image the April 5, 6, and 10 datasets with GENDIRECT, presenting the median reconstructions in Figure 6 alongside the previously presented results on the April 11 dataset. Due to the more limited dynamic range of the closure-based reconstructions (e.g. Chael et al. 2018), we identify three distinct emission components (C0-0, C0-1, and C1) as in Section 5.1, each labeled accordingly with reference to K20. Unlike the model-fitting analysis performed in K20 which fit Gaussian models to scan-averaged visibilities, we attempt a similar independent analysis in the image domain. We fit an elliptical Gaussian to each of the three identified components in the median image, marking its centroid and contour at the half-maximum value in Figure 6. On April 5, the C0-0 component is only be marginally detected. In this case, the Gaussian parameters are fit to the image with loose constraints from the detections on the other days.

As shown in Fig. 1 of K20, the aperture coverage on April 5 and 6 are notably more sparse than on April 10 and 11, particularly due to the absence of baselines involving the SPT, limiting the sensitivity in the North-South direction. Both images exhibit spurious emission features that are not clearly present in reconstructions from April 10 and 11 datasets. Meanwhile, both images from the later two days show a high degree of consistency. Following K20 in adopting C0-0 as the VLBI core and kinematic reference, we compare the motion of the centroids from the fitted C0-1 and C1-0 Gaussian components. When considering each day individually, the uncertainties in the fit parameters are too high to resolve proper motion on the order of 1 μas day^{-1} in a 64×64 pixel grid covering 225×225 μas^2. Therefore, to increase the signal's significance, we train a GENDIRECT model on the concatenated datasets of April 5 and 6, as well as April 10 and 11, producing a median image reconstruction from each new combined dataset, and comparing the averaged differential motion across the approximately five-day separation.

The bottom row of Figure 6 presents the median GENDIRECT reconstructions on the combined datasets with the three-component fits. In the third panel on the bottom row, the fitted Gaussian components on the two separate epochs are overplotted with C0-0 set as the kinematic reference. We include a fiducial position uncertainty of 5% the clean beam's geometric mean half-width at half-maximum as a systematic error added in quadrature to the measurement parameter uncertainties. For the April 5+6 and April 10+11 datasets, the additional systematic uncertainty on the measured centroids amounts to ~ 0.62 μas and ~ 0.53 μas, respectively. Relative to C0-0, we detect a 2.2 ± 1.0 μas proper motion in the C0-1 component toward the NW direction

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

(33 ± 24 deg) between the combined datasets. The proper motion measured for the C1 component is 4.6 ± 1.0 μas in the SW direction (-57 ± 12 deg), which is aligned with the large-scale jet direction (e.g. Fuentes et al. 2023). Generally, the magnitude and direction of the proper motion estimates for C1 are consistent with those reported in K20 from the visibility model-fitting analysis, but the motion for our measured C0-1 sub-component is more consistent with that of K20's C0-2, though we don't attempt to model C0-2. The nearly perpendicular motions of K20's C0-1 and C0-2 complicate a direct comparison with our single C0-1 model. On the other hand, all of the C1 sub-components (C1-0, C1-1, and C1-2) are moving in a consistent direction in the K20 analysis. Therefore, we calculate that the motion of the flux-weighted C1 centroid in K20's analysis is $\Delta\mathrm{RA} = -1.2 \pm 0.4$ μas, $\Delta\mathrm{Dec} = -4.2 \pm 0.9$ μas, with a total angular displacement of 4.4 ± 0.9 μas. Altogether, these values are mutually consistent within 1σ with GENDIRECT's independent analysis, which measured $\Delta\mathrm{RA} = -2.5 \pm 1.0$ μas, $\Delta\mathrm{Dec} = -3.9 \pm 1.0$ μas, and total displacement of 4.6 ± 1.0 μas, despite the fact that our analysis was performed on the image plane. We measure that the difference in the mean observation epoch between the combined two-day datasets is 5.39 d, with which we estimate the apparent proper motion of the C1 component to be $\beta_{\mathrm{app}} = 10 \pm 2$.

6. Conclusion

GENDIRECT is a novel deep learning imaging algorithm designed to solve the ill-posed inverse problem of directly reconstructing images from a set of interferometric closure invariant measurables in the highly challenging VLBI context (Lai 2025; Lai et al. 2025a). In previous work, the co-polar closure invariants formalism presented in Thyagarajan et al. (2022) and applied under the conditions of sparse aperture coverage was shown to be effective carriers of calibration-independent information on true source structure.

In this work, we present, for the first time, the application of the GENDIRECT imaging technique on real data from the Event Horizon Telescope on active galactic nuclei. The two targets in this work are the archetypal blazar 3C 279 and the radio galaxy Centaurus A, observed by the EHT in 2017. Below, we summarise the main results of this study:

- GENDIRECT was trained as described in L25 utilising the aperture coverage from datasets of 3C 279 and Centaurus A. Once fully trained, validated, and applied on the real data, we find a high degree of consistency between the median GENDIRECT reconstruction and the appropriate reference reconstruction from the EHTC, as measured by the maximum normalised cross-correlation $\rho_{\mathrm{NX}} > 0.9$, where a score of unity denotes exact image correspondence. The data fidelity score of the best single reconstruction from GENDIRECT is $\chi^2_{\mathrm{CI}} \lesssim 2$.
- We identify clusters of candidate image reconstructions from GENDIRECT by clustering image solutions based upon perceptual hashing algorithms. In both cases of 3C 279 and Centaurus A, a significant portion of all reconstructions fall within a single cluster, while less populated clusters exhibit morphological diversity which are symptoms of the sparse aperture coverage VLBI image reconstruction, such as the phantom repetition of similar structures.
- By fitting the image reconstructions with parameterised image models, we construct quantitative measures of morphological characteristics which we compare between the EHTC reference images and the median GENDIRECT reconstruction. For 3C 279, we fitted three asymmetric Gaussian components to compare the intensity ratio between the core and ejecta ($I(\mathrm{C0}-1)/I(\mathrm{C1})$), the separation ($|\,\mathbf{C0}-\mathbf{C1}\,|$), and angle ($\theta_{\mathbf{C1-C0}}$) formed between the two components, as well as the position angle of the bright extended component ($\mathrm{PA}_{\mathbf{C0\text{-}1}}$). We find all measured features to be statistically similar between the EHTC ($I(\mathrm{C0}-1)/I(\mathrm{C1}) = 1.52 \pm 0.01$, $|\,\mathbf{C0}-\mathbf{C1}\,| = 90.9 \pm 0.2$ μas, $\theta_{\mathbf{C1-C0}} = -75.7 \pm 0.1$ deg, $\mathrm{PA}_{\mathbf{C0\text{-}1}} = 44.5 \pm 0.5$ deg) and GENDIRECT reconstructions ($I(\mathrm{C0}-1)/I(\mathrm{C1}) = 1.53 \pm 0.02$, $|\,\mathbf{C0}-\mathbf{C1}\,| = 90.2 \pm 0.2$ μas, $\theta_{\mathbf{C1-C0}} = -75.6 \pm 0.1$ deg, $\mathrm{PA}_{\mathbf{C0\text{-}1}} = 44.2 \pm 0.7$ deg), especially if taking into consideration a systematic position error at a fraction (5%) of the clean beam dimensions.
- For Centaurus A, we fitted linear components to the two ridgelines and measured the intensity ratio ($R_{\mathrm{s/n}}$) as well as the jet opening angle ($\theta_{\mathrm{s/n}}$), defined by the angle formed between the two fitted lines. Both quantities were near-identical between the EHTC ($R_{\mathrm{s/n}} = 1.5 \pm 0.2$, $\theta_{\mathrm{s/n}} = 12.2 \pm 0.1$ deg) and median GENDIRECT reconstructions ($R_{\mathrm{s/n}} = 1.4 \pm 0.1$, $\theta_{\mathrm{s/n}} = 12.3 \pm 0.3$ deg).
- We examine the relative motion of nuclear and ejecta components in 3C 279 across four days of EHT observations using GENDIRECT reconstructions on the combined dataset of the first two and latter two dates, quantifying the relative proper motion on sub-parsec scales. Despite performing our analysis on the image reconstructions rather than model-fitted visibilities, we find consistent estimates of the ejecta proper motion in both magnitude and direction, with a total angular displacement of 4.6 ± 1.0 μas over $\simeq 5.39$ d, indicating apparent superluminal motion of $\beta_{\mathrm{app}} = 10 \pm 2$. This can be compared to the 4.4 ± 0.9 μas displacement in the flux-weighted centroid of the three ejecta subcomponents in Kim et al. (2020).

The results of GENDIRECT on the datasets of 3C 279 and Centaurus A show remarkable consistency with the EHTC image reconstructions. As an imaging algorithm conditioned on purely closure information, GENDIRECT offers a calibration-independent constraint on source morphology, which ultimately enhances the reliability of image reconstruction on sparsely sampled datasets. Moreover, we showcase the utility of novel machine learning algorithms in solving challenging inverse problems characteristic of radio interferometric imaging. This study presents, for the first time, independent image reconstructions from EHT datasets on 3C 279 and Centaurus A as a demonstration of the GENDIRECT imaging method on real EHT datasets, paving the way for GENDIRECT to be applied on the event-horizon-scale imaging problem of M87 and Sgr A∗. With potential to be expanded for dynamic and polarimetric imaging, the techniques underlying GENDIRECT have the potential to be more widely applied on imaging inverse problems throughout radio interferometry.

Acknowledgements. We acknowledge Jae-young Kim and Michael Janssen for supplying the Event Horizon Telescope image reconstructions of 3C 279

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

and of Centaurus A, respectively, which were for the comparative analysis. Inputs from Phil Edwards, Ron Ekers, and Mark Cheung are gratefully acknowledged. We acknowledge use of GitHub Copilot.

Software packages used in this study include Numpy (van der Walt, Colbert, & Varoquaux 2011), Scipy (Virtanen et al. 2020), PyTorch (Paszke et al. 2019), eht-imaging (Chael et al. 2018), GenDIReCT (Lai 2025), ClosureInvariants (Thyagarajan & Lai 2025), and Matplotlib (Hunter 2007).

Data availability statement. The data underlying this article will be shared on reasonable request to the corresponding author. Code for GenDIReCT is publicly accessible on GitHub.[d]

Author contributions. Conceptualisation: S.L; N.T. Methodology: S.L; F.D; N.T; I.W. Data curation: S.L. Data visualisation: S.L. Writing original draft: S.L.; N.T. All authors approved the final submitted draft.

Funding statement. None.

Competing interests. None.

Ethical standards. The research meets all ethical guidelines, including adherence to the legal requirements of the study country.

References

Akiyama, K., et al. 2017a, ApJ, 838, 1
Akiyama, K., et al. 2017b, AJ, 153, 159
Arras, P., et al. 2022, NatAs, 6, 259
Baars, J. W. M., Martin, R. N., Mangum, J. G., McMullin, J. P., & Peters, W. L. 1999, PASP, 111, 627
Blackburn, L., et al. 2020, ApJ, 894, 31
Blackburn, L., et al. 2019, ApJ, 882, 23
Broderick, A. E., & Pesce, D. W. 2020, ApJ, 904, 126
Broderick, A. E., Pesce, D. W., Tiede, P., Pu, H.-Y., & Gold, R. 2020, ApJ, 898, 9
Broderick, A. E., et al. 2022, ApJ, 935, 61
Carilli, C. L., & Thyagarajan, N. 2022, ApJ, 924, 125
Carlstrom, J. E., et al. 2011, PASP, 123, 568
Chael, A. A., et al. 2018, ApJ, 857, 23
Chiche, B. N., Girard, J. N., Frontera-Pons, J., Woiselle, A., & Starck, J.-L. 2023, A&A, 675, A116
Deng, J., et al. 2009, in 2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), 248
Deng, L. 2012, IEEE SPM, 29, 141
Doeleman, S., et al. 2009, in astro2010: The Astronomy and Astrophysics Decadal Survey, Vol. 2010, 68
Donike, S., Aybar, C., Gómez-Chova, L., & Kalaitzis, F. 2025, IEEE JSTAEORS, 18, 6940
Drozdova, M., et al. 2024, A&A, 683, A105
Event Horizon Telescope Collaboration, et al. 2019a, ApJ, 875, L2
Event Horizon Telescope Collaboration, et al. 2019b, ApJ, 875, L3
Event Horizon Telescope Collaboration, et al. 2019c, ApJ, 875, L4
Event Horizon Telescope Collaboration, et al. 2022, ApJ, 930, L14
Feng, B. T., Bouman, K. L., & Freeman, W. T. 2024, ApJ, 975, 201
Fuentes, A., et al. 2023, NatAs, 7, 1359
Geyer, F., et al. 2023, A&A, 677, A167
Goddi, C., et al. 2019, PASP, 131, 075003
Greve, A., et al. 1995, A&A, 299, L33
Güsten, R., et al. 2006, A&A, 454, L13
Hamming, R. W. 1950, BSTJ, 29, 147
Harris, G. L. H., Rejkuba, M., & Harris, W. E. 2010, PASA, 27, 457
Ho, J., Jain, A., & Abbeel, P. 2020, in Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS'20 (Red Hook, NY, USA: Curran Associates Inc.)
Ho, P. T. P., Moran, J. M., & Lo, K. Y. 2004, ApJ, 616, L1
Huertas-Company, M., & Lanusse, F. 2023, PASA, 40, e001
Hughes, D. H., et al. 2010, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7733, Ground-based and Airborne Telescopes III, ed. L. M. Stepp, R. Gilmozzi, & H. J.Hall, 773312
Hunter, J. D. 2007, CSE, 9, 90
Ikeda, S., Tazaki, F., Akiyama, K., Hada, K., & Honma, M. 2016, PASJ, 68, 45
Janssen, M., et al. 2019, A&A, 626, A75
Janssen, M., et al. 2021, NatAs, 5, 1017
Jennison, R. C. 1958, MNRAS, 118, 276
Kim, J., et al. 2018, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 10708, Millimeter, Submillimeter, and Far-Infrared Detectors and Instrumentation for Astronomy IX, ed. J. Zmuidzinas, & J.-R. Gao, 107082S
Kim, J.-Y., et al. 2020, A&A, 640, A69
Krizhevsky, A. 2009, Learning multiple layers of features from tiny images, Tech. rep., University of Toronto, Toronto, Ontario, https://www.cs.toronto.edu/kriz/learning-features-2009-TR.pdf
Lai, S. 2025, samlaihei/GenDIReCT: GenDIReCT_v1.0.0, doi:10.5281/zenodo.15460681
Lai, S., Thyagarajan, N., Wong, O. I., & Diakogiannis, F. 2025a, PASA, 42, e148
Lai, S., Thyagarajan, N., Wong, O. I., Diakogiannis, F., & Hoefs, L. 2025b, MNRAS, 536, 446
Larionov, V. M., et al. 2020, MNRAS, 492, 3829
Lister, M. L., et al. 2018, ApJS, 234, 12
Lockhart, W., & Gralla, S. E. 2022, MNRAS, 509, 3643
Longo, G., Merényi, E., & Tiňo, P. 2019, PASP, 131, 100101
Marziani, P., Sulentic, J. W., Dultzin-Hacyan, D., Calvani, M., & Moles, M. 1996, ApJS, 104, 37
Mertens, F., & Lobanov, A. 2015, A&A, 574, A67
Miyoshi, M., Kato, Y., & Makino, J. 2022, ApJ, 933, 36
Müller, H. 2024, A&A, 689, A299
Müller, H., & Lobanov, A. P. 2022, A&A, 666, A137
Müller, H., Mus, A., & Lobanov, A. 2023, A&A, 675, A60
Mus, A., Müller, H., & Lobanov, A. 2024, A&A, 688, A100
Nammour, F., et al. 2022, A&A, 663, A69
Paszke, A., et al. 2019, PyTorch: An Imperative Style, High-Performance Deep Learning Library (Red Hook, NY, USA: Curran Associates Inc.)
Pearson, T. J., & Readhead, A. C. S. 1984, ARA&A, 22, 97
Planck Collaboration, et al. 2016, A&A, 594, A13
Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. 2022, in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10674
Rustige, L., et al. 2023, RASTI, 2, 264
Samuel, J., Nityananda, R., & Thyagarajan, N. 2022, PhRvL, 128, 091101
Schmidt, K., et al. 2022, A&A, 664, A134
Shepherd, M. 2011, Difmap: Synthesis Imaging of Visibility Data, Astrophysics Source Code Library, record ascl:1103.001
Sun, H., et al. 2022, ApJ, 932, 99
Sureau, F., Lechat, A., & Starck, J. L. 2020, A&A, 641, A67
Terris, M., Dabbech, A., Tang, C., & Wiaux, Y. 2023, MNRAS, 518, 604
Thompson, A. R., Moran, J. M., & Swenson, George W., J. 2017, Interferometry and Synthesis in Radio Astronomy, Astronomy & Astrophysics Library (3rd edn.), doi:10.1007/978-3-319-44431-4
Thyagarajan, N., & Carilli, C. L. 2022, PASA, 39, e014
Thyagarajan, N., & Lai, S. 2025, nithyanandan/ClosureInvariants-1.0.0, doi:10.5281/zenodo.15524509
Thyagarajan, N., Nityananda, R., & Samuel, J. 2022, PhRvD, 105, 043019
Twiss, R. Q., Carter, A. W. L., & Little, A. G. 1960, Obs, 80, 153
van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, CSE, 13, 22
Virtanen, P., et al. 2020, NM, 17, 261
Whitney, A. R., et al. 1971, Sci, 173, 225
Wootten, A., & Thompson, A. R. 2009, IEEE Proc., 97, 1463
Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. 2017, IEEE TIP, 26, 3142

[d]https://github.com/samlaihei/GenDIReCT.

Appendix A. Perceptual hashing

Perceptual hashing is a form of fingerprinting algorithm that is often applied on various types of multimedia, including images. Unlike cryptographic hashing techniques, perceptual hash algorithms are designed to minimise variation and preserve the resulting hash when the inputs are similar.

To quantify image similarity across a generated sample for clustering purposes, we calculate the discrete cosine transform (DCT) hash for each reconstructed image. As illustrated in Figure A1, the DCT hash is a transformation which begins by computing the discrete cosine transform,

$$X_k = \sum_{n=0}^{N-1} 2x_n \cos\left(\frac{\pi}{2N}(2n+1)k\right), \tag{A1}$$

for each row and column. We then select the $s \times s$ box of pixels corresponding to the lowest frequency features, which are the least sensitive to noisy or small-scale variations. Finally, we convert the DCT image into binary using the median of the transformed data as the cutoff value. Optionally, we can identify the hash of any image by its hexadecimal representation as shown in the bottom right panel of Figure A1.

Similarity is measured by the Hamming distance (Hamming 1950), defined as the number of bits for which the corresponding bit in the comparison sequence is flipped. In our implementation, we select $s = 5$ and apply a Hamming distance threshold of 6 for identifying members of a cluster unless stated otherwise.

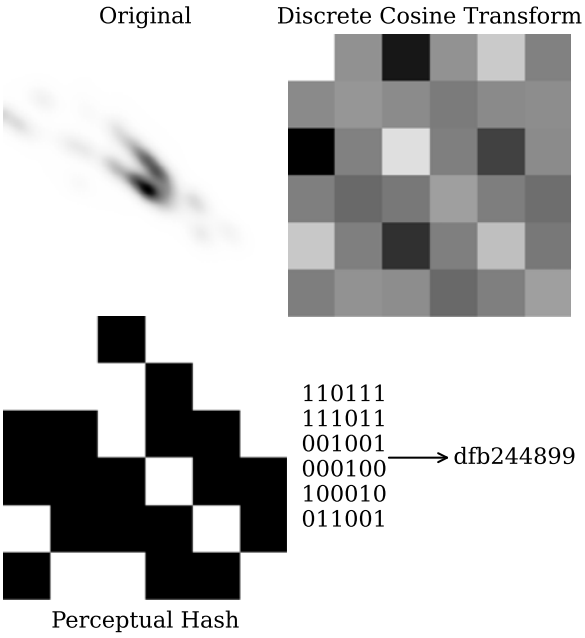


Figure A1. (Top left) Model image of Centaurus A's jet (Janssen et al. 2021). (Top right) Processed image after applying the discrete cosine transform on both axes and selecting the 6 × 6 box of the lowest frequency features. (Bottom left) The image is converted to binary by using the median value of the data as the threshold. White and black pixels correspond to 1 and 0, respectively. (Bottom right) Hexadecimal representation of the binary hash.

Appendix B. Alternative reconstructions

Although GenDIReCT imaging does not depend on fine-tuned hyperparameters, such as selected priors or independently weighted regularisation terms, its result is not immune from pre-imaging or data processing considerations. In this appendix, we explore the outcome of different selections of the image reconstruction window and data processing strategies.

Appendix B.1. Field-of-view effects

Due to the paucity of short and intermediate ($\lesssim 1$ Gλ) baselines in the array, EHT observations exhibit limited sensitivity to extended structure on scales of $\gtrsim 100$ μas (e.g. Kim et al. 2020). As such, the EHT field-of-view is relatively compact. Consequently, we have carefully selected the image reconstruction field-of-view for 3C 279 and Centaurus A based on the EHTC reference images to capture as much of the contributing emission sources while adhering to the spatial scales that fall within the sensitivity range of the EHT. However, as we have described regarding the validation tests in Section 3.3, in cases where significant emission lies outside the reconstruction field-of-view, limiting the image reconstruction window can produce biased and unfaithful reconstructions. Therefore, in this section, we train GenDIReCT for the 3C 279 and Centaurus A datasets with modified reconstruction windows to ensure consistency in the image reconstruction output.

In Figure A2, we present the median reconstructions from GenDIReCT on 3C 279 with the 11 April 2017 dataset and Centaurus A, each with three different reconstruction windows. For 3C 279, we have chosen 128×128, 225×225, and 300×300 μas^2. On Centaurus A, the original 450×450 μas^2 field-of-view is already large relative to the EHT array's sensitivity. Therefore, we selected smaller alternatives of 225×225 and 300×300 μas^2 to visualise the edge-brightened jet sheath in closer detail. We present the relative $\rho_{\rm NX}$ image correspondence score between the median GenDIReCT reconstruction in each of the alternative fields-of-view and the default median image used in this study (225 μas for 3C 279 and 450 μas for Centaurus A), which is also highlighted with the gold border.

As described in Section 4, the image reconstruction artifacts northwards of the edge-brightened Centaurus A jet are clipped to focus the $\rho_{\rm NX}$ comparison on the central feature. For any fields-of-view smaller than the reference, we illustrate the edges of the reconstruction windows for visual clarity as all reconstructions are presented in the reference field-of-view. Additionally, we compute the best $\chi^2_{\rm CI}$ of the individual GenDIReCT reconstructions in each field-of-view, showing that the widths of 225 μas for 3C 279 and 450 μas for Centaurus A result in the optimum $\chi^2_{\rm CI}$ compared to distinct field-of-view alternatives. All of the panels showcasing alternative reconstruction windows in Figure A2 are closely consistent ($\rho_{\rm NX} > 0.80$) with the reference image reconstruction without significant deviations in the overall reconstructed image morphology, lending enhanced credibility to the imaging results obtained in this study.

Appendix B.2. Alternative data aggregation strategy

As described in Section 3, the imaging results in this section are produced by averaging visibility information across ∼ 5-min scans to compress the dataset into a more manageable volume and improve SNR. However, while station-based multiplicative corruptions are removed by the closure invariants construction, visibilities can be further corrupted by time-dependent errors across the duration of the scan, the effect of which would propagate into the closure invariants. To mitigate the time-dependent effects, it's possible to compute the closure invariants in the 10s-binned

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

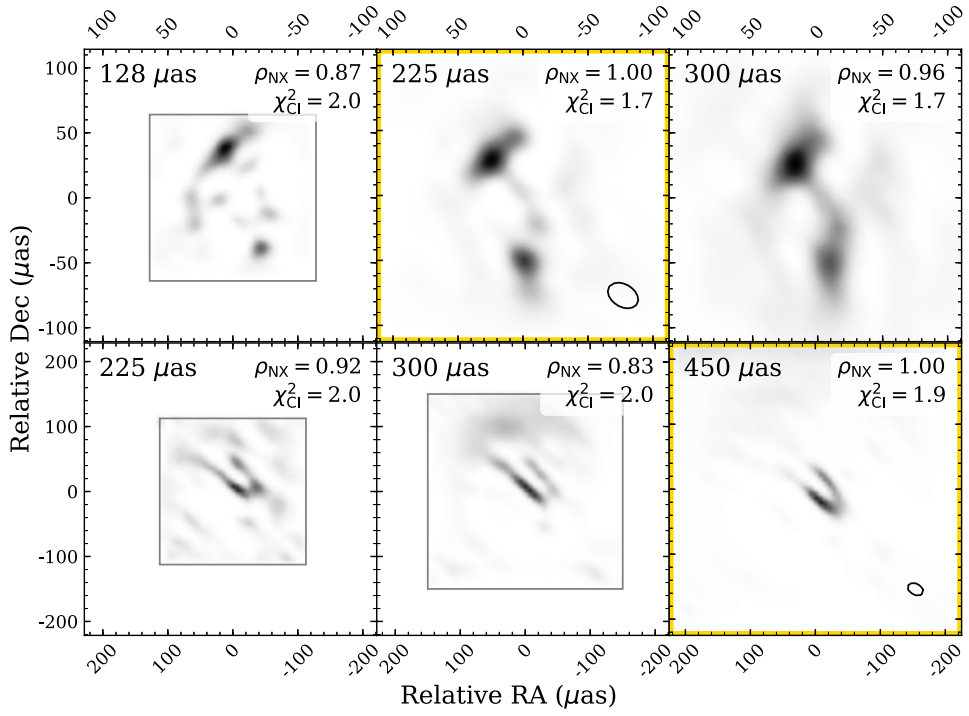


Figure A2. Median GenDIReCT reconstructions of 3C 279 (top) and Centaurus A (bottom) at varying square fields-of-view from widths of (128, 225, 300 μas) for 3C 279 and (225, 300, 450 μas) for Centaurus A. The fitted elliptical clean beam is plotted on the reconstruction that is used as reference for computing the relative ρ_{NX} and the reference image is also highlighted with a gold border. The best χ^2_{CI} data fidelity score on individual GenDIReCT reconstructions is presented in each panel. For any reconstruction windows smaller than the reference field-of-view, we plot the edges of the reconstruction window.

cadence of the public dataset. In L25, it was demonstrated empirically that pre-computing and averaging the closure invariants at a higher temporal resolution is generally a more robust method of data aggregation than averaging the visibilities on the same timescale prior to computing closure invariants in the presence of noise. However, some exceptions were identified, such as when the total flux density is low, then under standard SEFDs, there is little difference between both strategies. Additionally, when the total flux density is on the order of ~ 1 Jy and the visibility phase is well-calibrated, then averaging visibilities is superior at preserving the true closure invariant quantities than pre-computing closure invariants at a higher temporal resolution prior to averaging. Indeed, if the visibilities are well-calibrated, averaging the visibilities produces a more robust result for any source with ~ 1 Jy or brighter. On the other hand, if the calibration is uncertain, averaging the closure invariants is superior and its relative advantage grows further with source brightness.

For Figure B1, we train new GenDIReCT models with the closure invariant averaging strategy and present the resulting median image reconstructions of 3C 279 on the 11 April 2017 dataset and Centaurus A, alongside the best single-reconstruction χ^2_{CI} data metric and ρ_{NX} of the median closure invariant averaging reconstruction relative to the corresponding visibility averaging dataset image presented in Section 4. We find that for 3C 279, the resulting median reconstruction from the closure invariant averaged dataset is $\rho_{NX} = 0.96$ relative to the visibility averaged dataset with statistically similar performance on χ^2_{CI}. However, the quality of the image reconstruction of Centaurus A is degraded. Although the general tuning fork morphology is visibly identifiable, the median reconstruction shows parallel features similar to structures in the

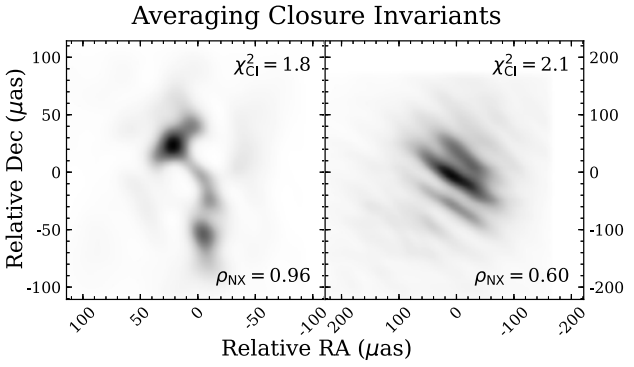


Figure B1. Median GenDIReCT reconstructions of 3C 279 (left) and Centaurus A (right) from the closure invariant averaged dataset. The best single-reconstruction χ^2_{CI} is shown on the top-right of each panel and the ρ_{NX} compared to the median visibility-averaged GenDIReCT reconstruction is displayed in the bottom-right corner of each panel.

dirty image, which decreases the image correspondence with the visibility averaged reconstruction to $\rho_{NX} = 0.6$. The total compact flux density of Centaurus A is ~ 2 Jy (Janssen et al. 2021) and after averaging the closure invariants, the median closure invariant SNR is ~ 2, which is likely too low for the GenDIReCT reconstruction to be considered reliable, even on tests with synthetic data (L25). For comparison, the median closure invariant SNR from the visibility-averaged dataset is ~ 5. Thus, we find that due to its low compact flux density, the phase-calibrated dataset of Centaurus A falls within the range where visibility averaging results in a superior reconstruction. Comparably, we find minimal difference in

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press

reconstruction quality of 3C 279 due to its high SNR. We further anticipate that with a compact flux density of ~ 1 Jy, the calibrated datasets of M87 are expected to display similar characteristics as Centaurus A in this respect.

Appendix B.3. Alternative closure triad reference stations

One can identify all independent triangles for an interferometer by fixing a base vertex or reference station and selecting all triangles containing the vertex (Thompson et al. 2017; Thyagarajan et al. 2022). A different selection of the reference station produces a different set of closure invariants carrying equivalent information. In this section, we investigate the choice of the reference station on the resulting image reconstruction by alternatively selecting the second most sensitive station in each scan instead of the most sensitive station. With this selection, the overall SNR of the closure invariants decreases due to the less sensitive base station and increased covariances.

We illustrate the median GenDIReCT reconstruction for both 3C 279 (top) and Centaurus A (bottom) in Figure B2 alongside their respective $\chi^2_{\rm CI}$ data fidelity scores and relative $\rho_{\rm NX}$. The left column reconstructions utilise the most sensitive station as the reference in each scan, while the right column reconstructions utilise the second most sensitive station. For both targets, we find a high degree of image consistency with either reference station selection, with $\rho_{\rm NX} = 0.95$ correspondence between the median GenDIReCT reconstructions. As for the data fidelity, we note that with reference station 1, the reconstruction performance on $\chi^2_{\rm CI}$ is degraded for Centaurus A, but not for 3C 279. The reason is likely that when the less sensitive station is used consistently in construction of closure triads, the overall SNR of the resulting closure invariants also decreases, which affects the observations of Centaurus A, observed with a median visibility SNR ~ 5, disproportionately compared to observations of 3C 279, where the median visibility is significantly greater (SNR ~ 23).

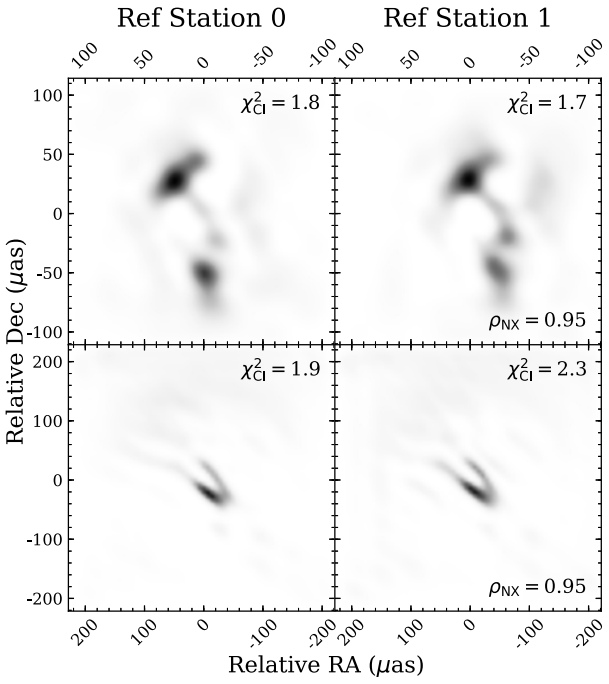


Figure B2. Comparison illustration of median GenDIReCT reconstructions of 3C 279 (top) and Centaurus A (bottom) with different choices of reference station in each scan (0— most sensitive station, 1— second-most sensitive station). Both reconstructions of each target are structurally consistent to $\rho_{\rm NX} = 0.95$.

https://doi.org/10.1017/pasa.2026.10167 Published online by Cambridge University Press