



First AI for Deep Super-resolution Wide-field Imaging in Radio Astronomy: Unveiling Structure in ESO 137-006

A. Dabbech¹ , M. Terris¹, A. Jackson² , M. Ramatsoku³ , O. M. Smirnov^{3,4} , and Y. Wiaux¹ ¹Institute of Sensors, Signals, and Systems, Heriot-Watt University, Edinburgh EH14 4AS, UK; a.dabbech@hw.ac.uk²EPCC, University of Edinburgh, Edinburgh EH8 9BT, UK³Department of Physics and Electronics, Rhodes University, PO Box 94, Makhanda (Grahamstown), 6140, South Africa⁴South African Radio Astronomy Observatory, 2 Fir Street, Black River Park, Observatory, Cape Town 7925, South Africa

Received 2022 August 23; revised 2022 September 28; accepted 2022 October 7; published 2022 October 26

Abstract

We introduce the first AI-based framework for deep, super-resolution, wide-field radio interferometric imaging and demonstrate it on observations of the ESO 137-006 radio galaxy. The algorithmic framework to solve the inverse problem for image reconstruction builds on a recent “plug-and-play” scheme whereby a denoising operator is injected as an image regularizer in an optimization algorithm, which alternates until convergence between denoising steps and gradient-descent data fidelity steps. We investigate handcrafted and learned variants of high-resolution, high dynamic range denoisers. We propose a parallel algorithm implementation relying on automated decompositions of the image into facets and the measurement operator into sparse low-dimensional blocks, enabling scalability to large data and image dimensions. We validate our framework for image formation at a wide field of view containing ESO 137-006 from 19 GB of MeerKAT data at 1053 and 1399 MHz. The recovered maps exhibit significantly more resolution and dynamic range than CLEAN, revealing collimated synchrotron threads close to the galactic core.

Unified Astronomy Thesaurus concepts: [Astronomy image processing \(2306\)](#); [Computational astronomy \(293\)](#); [Convolutional neural networks \(1938\)](#); [Radio galaxies \(1343\)](#); [Aperture synthesis \(53\)](#)

1. Introduction

Image formation in aperture synthesis by radio interferometry (RI) has never been more challenging. On the one hand, the extreme data sampling rates produced by modern radio arrays raise the urgent need for scalable algorithms. On the other hand, the ill-posedness of the underlying inverse problem calls for tailored regularization models to be injected into the image formation process in order to deliver the expected precision and robustness of the reconstruction. Over the last decade, regularization approaches leveraging advanced sparsity-based image models embedded in optimization algorithms have been proposed (e.g., Li et al. 2011; Carrillo et al. 2012; Dabbech et al. 2015; Garsden et al. 2015). In particular, the SARA family of algorithms (Onose et al. 2017; Repetti et al. 2017; Dabbech et al. 2018; Abdulaziz et al. 2019; Dabbech et al. 2021; Thouvenin et al. 2022a, 2022b) have recently delivered a significant increase of resolution and dynamic range (or depth) over CLEAN-based algorithms (e.g., Högbom 1974; Wakker & Schwarz 1988; Cornwell 2008) on the well-known radio galaxy Cygnus A. Owing to the complexity of the regularization models underpinning the imaging accuracy, optimization approaches are significantly more computationally expensive than CLEAN. Their scalability to gigabyte-scale image sizes and data volumes has been enabled by resorting to advanced algorithmic structures enabling a significant degree of parallel processing (e.g., Pesquet & Repetti 2014; Chouzenoux et al. 2016). Nonetheless, scalability to much larger image and data dimensions is required for upcoming instruments, with

the Square Kilometre Array (SKA; Scaife 2020) intended to deliver petabyte-scale images from exabyte-scale data volumes.

Assuming monochromatic, nonpolarized radio emission and a narrow field of view (FoV), the measured complex visibilities are noisy Fourier components of the sky surface brightness, where the sampled (u, v) points are the projections of each antenna pair baseline on the plane perpendicular to the line of sight. Under these assumptions, the visibility vector $\mathbf{y} \in \mathbb{C}^M$ can be modeled as

$$\mathbf{y} = \Phi \bar{\mathbf{x}} + \bar{\mathbf{n}}, \text{ with } \Phi = \mathbf{G}\mathbf{F}\mathbf{Z}, \quad (1)$$

where $\bar{\mathbf{x}} \in \mathbb{R}^N$ is the unknown radio map whose pixel resolution is often set between 1.5 and 2.5 times below the angular resolution of the observations to reduce the limitations of pixel-based image restoration, $\bar{\mathbf{n}} \in \mathbb{C}^M$ is a realization of a complex random Gaussian noise of mean zero and standard deviation $\tau > 0$, and $\Phi \in \mathbb{C}^{M \times N}$ denotes the measurement operator, which encodes the incomplete Fourier sampling. More precisely, $\mathbf{G} \in \mathbb{C}^M \times \mathbb{C}^{N'}$ denotes a sparse degridding matrix whose rows are nonuniform Fourier transform interpolation kernels, $\mathbf{F} \in \mathbb{C}^{N'} \times \mathbb{C}^{N'}$ stands for the discrete Fourier transform, and $\mathbf{Z} \in \mathbb{C}^{N'} \times \mathbb{R}^N$ is a zero-padding operator, allowing for a fine grid in the spatial Fourier domain, which also involves a correction for approximations in the convolution kernels of \mathbf{G} (Fessler & Sutton 2003). Given the remarkable sensitivity of the modern arrays, the RI measurement equation is further complicated by the so-called direction-dependent effects (DDEs). Some of these are unknown and of either atmospheric or instrumental origin and should be calibrated (Smirnov 2011). In contrast, the DDEs originating from the w -component of the antenna pair baselines on the line of sight are known and induce the so-called w -effect



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

(Cornwell & Perley 1992). The DDEs can be encapsulated as additional baseline-specific convolution kernels on each row of \mathbf{G} (Dabbech et al. 2017; Repetti et al. 2017). In this work, we propose a new parallelized and automated framework for wide-field, high-resolution, high dynamic range monochromatic intensity imaging, which we use to revisit observations of the radio galaxy ESO 137-006, the loudest radio galaxy in the Norma cluster.

The remainder of this letter is structured as follows. In Section 2, we provide a summary of the proposed framework, from the underpinning algorithmic structure and the two specific incarnations respectively propelled by sparsity-based and AI-based regularization, to parallelization and automation functionalities critical to scalability. A description of the utilized RI data of ESO 137 from the MeerKAT telescope is given in Section 3. Imaging settings, as well as a description of the utilized computational resources, are provided in Section 4. Imaging results are presented in comparison with a CLEAN-based benchmark method in Section 5, followed by a discussion of the unveiled structure in ESO 137-006. Finally, conclusions are drawn in Section 6.

2. Methods

At the algorithmic level, the proposed framework is underpinned by the versatile forward-backward (FB) convex optimization iterative structure (Bauschke & Combettes 2017). At each iteration, FB simply alternates until convergence between a (forward) gradient-descent step promoting fidelity to data and a (backward) step enforcing a prior image model, critical to the regularization of the inverse problem and the resulting imaging precision (see Appendix A). We investigate two incarnations of a recent plug-and-play (PnP) scheme (Venkatakrisnan et al. 2013; Romano et al. 2017), whereby dedicated denoising operators can be plugged into FB as an image regularizer.

The unconstrained SARA (uSARA) algorithm is a pure optimization variant leveraging a so-called “proximal” denoiser, handcrafted to enforce an advanced sparsity-based image regularization (Carrillo et al. 2012; Repetti & Wiaux 2021; Terris et al. 2022). The sophistication of the underlying prior image model is precisely introduced to deliver the best possible resolution and dynamic range from the data. The resulting denoiser itself is implemented as an iterative algorithm, leading to an overall subiterative FB structure (see Appendix B).

The AI for Regularization in radio interferometric Imaging (AIRI) algorithm (Terris et al. 2022) is an AI-based variant leveraging a learned denoiser in the form of a deep neural network (DNN) trained on a rich database to clean Gaussian random noise from high dynamic range images with a noise level commensurate with the target sensitivity of observation (see Appendix C). By design, AIRI inherits the robustness and interpretability of optimization algorithms and the learning power and speed of DNNs.

Importantly, the degree of refinement with which the uSARA and AIRI image models are enforced is adjusted to the measurement noise τ , more precisely to the corresponding estimate of the noise level in the image domain, $\tau/\sqrt{2L}$, which results from a normalization by the norm L of the measurement operator. In other words, uSARA and AIRI automatically adapt to the input signal-to-noise ratio or, equivalently, the target dynamic range of reconstruction. Last but not least, we

emphasize that, by construction, PnP denoisers are completely blind to the measurement conditions underpinning the data to be imaged. As a consequence, the learned variants can be trained once and for all at an appropriate dynamic range, significantly alleviating the associated computation cost. They do not suffer from generalizability challenges with respect to measurement conditions either. This stands in stark contrast with the more traditional end-to-end approaches, where a DNN would be trained to reconstruct an image directly from data (Connor et al. 2022; Terris et al. 2022).

At the high-resolution and high dynamic range regime of interest, parallelization and automation functionalities are critical to the scalability of the algorithmic framework. In this context, the image denoisers of uSARA and AIRI are decomposed on small image facets with no loss of precision thanks to their convolutional nature and the compactness of the associated kernels (see Appendix D). Relying on a hybrid approach to efficiently correct for the wide-field w -effect in both image and data spaces (Cornwell et al. 2005; Wiaux et al. 2009; Offringa et al. 2014; Dabbech et al. 2017), the measurement operator is decomposed into sparse and low-dimensional building blocks (see Appendix E). These decompositions are fully automated, enabling a parallel image facet and data block processing, seamlessly adapting to the architecture of the high-performance computing (HPC) system where the reconstruction is run.

3. Data Description

Both uSARA and AIRI are used to revisit MeerKAT L -band observations of a wide FoV containing the radio galaxy ESO 137-006. MeerKAT (Jonas 2016), located in the Karoo desert of South Africa, is a precursor to the SKA. Its 64 antennas with cryogenic receivers are arranged in a close-packed core and baselines of up to 8 km, resulting in superb sensitivity and imaging quality. The array is particularly suited to study faint extended emission and objects with complex morphology, of which ESO 137-006 represents a “flagship” case. Previous analysis of these observations by Ramatsoku et al. (2020) revealed multiple collimated synchrotron threads (CSTs) connecting the lobes of the radio galaxy, whose origin is yet to be unraveled.

The technical details of the observations and the initial calibration (i.e., reference calibration, or 1GC) were reported by Ramatsoku et al. (2020). The 1GC was performed using the CARACAL pipeline (Makhathini 2018; Józsa et al. 2020). The 1GC-calibrated data are averaged down from 4096 to 1024 channels of 0.84 MHz each, spanning the frequency range 856–1712 MHz. We utilize about 7 hr of on-target time and select two subbands relatively free from radio frequency interference, referred to as the “low” band (961–1145 MHz, centered at 1053 MHz) and the “high” band (1295–1503 MHz, centered at 1399 MHz), to form two continuum images. The respective data sizes after flagging are 8.2 GB (~ 532 million data points, double precision) and 10.76 GB (~ 673 million data points). The data were then self-calibrated for phase using a combination of the WSClean imager (Offringa & Smirnov 2017) and the CubiCal calibration suite (Kenyon et al. 2018). Attempts to calibrate for the amplitude and the DDEs (Repetti et al. 2017; Dabbech et al. 2021) did not bring a substantial improvement. Therefore, no further data preprocessing was performed. The resulting WSClean images, obtained with the multiscale variant of CLEAN (Cornwell 2008), are

presented for comparison purposes. Then uSARA and AIRI are used for image reconstruction on these self-calibrated data.

4. Imaging Settings and Computational Resources

The images formed are 4096×4096 pixels in size, spanning an FoV of $1.91 \times 1.91 \text{ deg}^2$ with a cell size of $1''.68$ for super-resolution factors beyond the angular resolution of observation of about 2 and 1.6 at the low and high bands, respectively. The data were weighted using the Briggs weighting scheme (robust parameter zero) to mitigate at best the complicated lobes of the dirty beam, i.e., the point-spread function arising from the Fourier sampling pattern. Specifically to AIRI, a single denoiser with an appropriate dynamic range was trained and used as an AIRI regularizer at both bands. Imaging parameter selection for both uSARA and AIRI is automated (see Appendix F). Finally, the WSClean parameters are set similarly to Ramatsoku et al. (2020).

With regard to computing resources, the MATLAB implementation of uSARA and AIRI and the C++ WSClean imager were run on Cirrus,⁵ a UK Tier-2 HPC system. The uSARA and CLEAN are deployed on CPUs, while AIRI is deployed mainly on CPUs, with AIRI's denoiser utilizing a GPU. More precisely, for uSARA and AIRI, the computation of the measurement operator (see Appendix E for details), decomposed into sparse low-dimensional building blocks, utilized 240 and 280 CPUs at the low and high bands, respectively. For the imaging process itself, forward steps utilized 99 and 180 CPUs at the low and high bands, respectively. The denoiser of uSARA, distributed over 64 image facets, utilized 64 of the CPUs already allocated for the forward steps. The denoiser of AIRI relied on a decomposition of the image into four facets, lowering the memory requirements per facet and enabling each facet to be processed on a single GPU. Given the relatively negligible GPU computation cost, a single GPU was used, with facets denoised sequentially rather than in parallel. Finally, WSClean used 72 CPUs, associated with the considered number of w -stacks.

5. Results and Discussion

The reconstruction results are provided in Figures 1–3, focusing on the ESO 137-006 region of the imaged FoV, and displayed in \log_{10} scale to enable the joint visualization of high-intensity and faint emission. Specific to the CLEAN reconstructions, we display the outcome of the convolution of the associated model image with the so-called restoring beam for a more physical representation of the radio sky. By construction, the uSARA and AIRI images are in units of janskys per pixel. In order to compare intensities in the same units, the CLEAN images are normalized by the flux of the restoring beam. Zooms on selected regions of the imaged FoV are provided on each figure. First, a zoom on the central region of ESO 137-006, including the active galactic nucleus (AGN) at its core, is provided in panels (b) and (e). Second, a zoom on some background compact sources at the high band are shown in panel (c). Third, a zoom on the neighboring radio galaxy ESO 137-007 north of ESO 137-006 at the low band is displayed in panel (f). Images of the full FoV are provided as supplementary material (Dabbech et al. 2022).

Where the residual images contain additional information, CLEAN's restored image, consisting of the sum of the convolved model and the residual image, is considered in our analysis. Both zooms on the selected background compact sources at the high band and the neighboring radio galaxy ESO 137-007 at the low band from CLEAN restored images are included in the respective panels (c') and (f') of Figure 3. No such consideration is necessary for uSARA and AIRI, where the algorithm solution itself, without further processing, is considered to be the final image reconstruction. This advantage was already highlighted for the previous algorithms of the SARA family, which rely on more advanced and physical regularization models than CLEAN (Dabbech et al. 2018; Abdulaziz et al. 2019; Dabbech et al. 2021; Thouvenin et al. 2022b). Finally, the model image of CLEAN is considered to support our analysis, particularly through zooms of the central region of ESO 137-006 at the high and low bands shown in the respective panels (b') and (e') of Figure 3.

Generally speaking, on both bands, one can observe the high level of detail achieved by uSARA and AIRI in comparison with CLEAN, particularly noticeable within the lobes of the radio galaxy. As opposed to the maps produced by CLEAN, whose resolution is, by design, restricted due to the convolution with the restoring beam, the uSARA and AIRI maps show a wealth of filamentary detail within the radio lobes of ESO 137-006. These improvements also come with some pixelation effects, more noticeable in AIRI's reconstructions, in super-resolved structures around the galactic core/jets.

The ability of both uSARA and AIRI to capture complex structure is further showcased when looking at the zoom on ESO 137-007 at the low band (a similar observation is made at the high band), in contrast with CLEAN (panel (f) of each figure). In fact, the filamentary detail of its jet is not recovered in the CLEAN convolved model image. Instead, it is left in the residual image and therefore only seen in the CLEAN restored image (panel (f') of Figure 3). The uSARA and AIRI also deliver further reconstruction depth, recovering lower signal intensities than CLEAN. Additional examination of the bright super-resolved pointlike sources south of the jet showcases the improvement in resolution brought by both AIRI and uSARA. Last but not least, one can notice that AIRI further improves the reconstruction dynamic range over uSARA, as is apparent from the recovered faint compact sources at the high band in panel (c) of each figure (a similar observation is made at the low band). These sources are also visible in the CLEAN restored image (panel (c')), though not directly recovered in the CLEAN model image (panel (c)). This supports the fact that they are real and not hallucination artifacts of the DNN. The AIRI also seems to be less sensitive to calibration errors that take the form of extended ringing structure around ESO 137-006 (panels (a) and (d) of Figures 1 and 2).

Since the low- and high-band images are produced completely independently, flux measurements of unresolved sources provide an important cross-check of the results. Table 1 summarizes the flux measurements of the AGN. We can see that all three methods recover almost the same flux and that the spectrum is relatively flat, as expected from an AGN. Furthermore, spectral index maps of ESO 137-006 obtained from uSARA and AIRI images (Figures 1 and 2, panel (g)) are in broad agreement with the findings of Ramatsoku et al. (2020) based on the CLEAN images of ESO 137-006 at close

⁵ <http://www.cirrus.ac.uk>

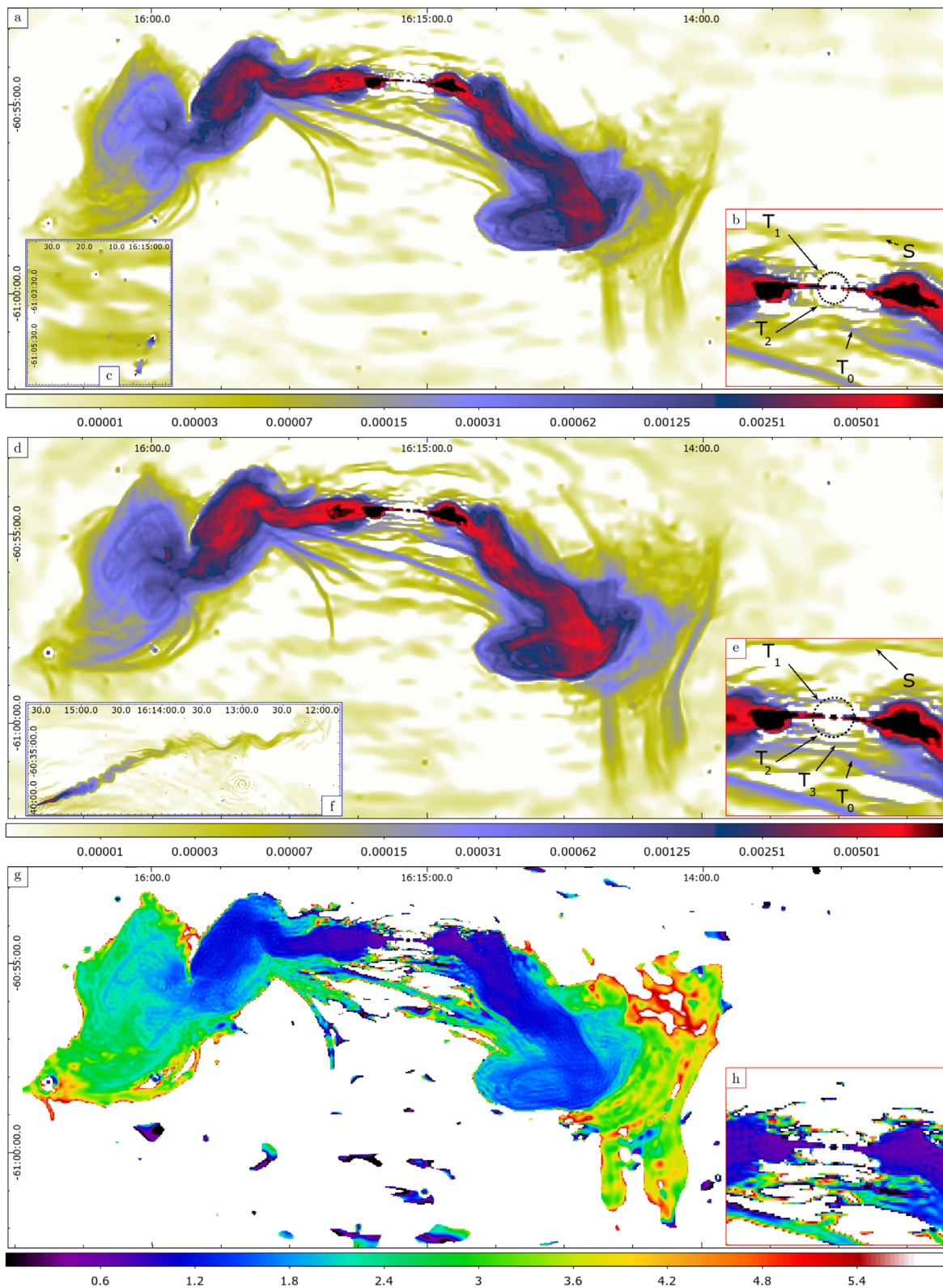


Figure 1. ESO 137-006: uSARA reconstructions (flip pages to visualize differences at a glance with AIRI in Figure 2 and CLEAN in Figure 3). First and second rows: recovered model images (Jy pixel^{-1} , displayed in \log_{10} scale) at the high and low bands (panels (a) and (d)), respectively, overlaid with zooms on the core of ESO 137-006 (panels (b) and (e)), a region with compact sources from the imaged FoV (panel (c)), and a zoom on ESO 137-007, a radio galaxy north of ESO 137-006 (panel (f)). Third row: spectral index map of ESO 137-006 (displayed in linear scale; panel (g)), overlaid with a zoom on its core (panel (h)), same region as in panels (b) and (e). Focusing on the central region (panels (b) and (e)), the first side lobe of the dirty beam is highlighted with a dashed circle. One can see three filaments emerging: T_1 and T_2 , located north and south of the inner core, seen at both bands, and T_3 , located further south, recovered only at low band. A fourth filament, T_0 , detected previously, is also recovered. The filamentary structure S is an example of what mostly likely is a calibration residual artifact, as it moves with the geometry of the dirty beam. The spectral index values of the newly formed filaments (panel (h)) are inconclusive.

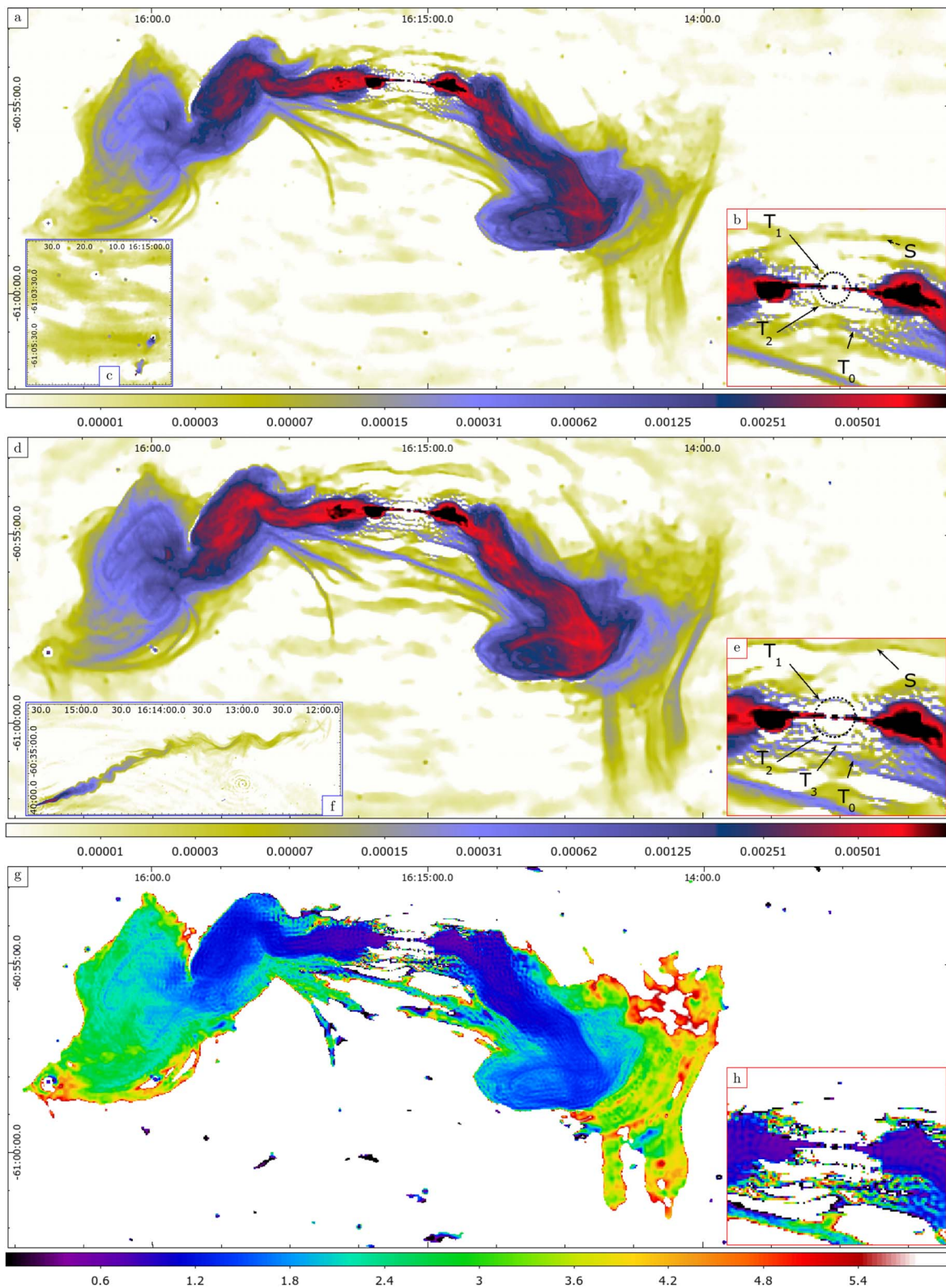


Figure 2. ESO 137-006: AIRI reconstructions (flip pages to visualize differences at a glance with uSARA in Figure 1 and CLEAN in Figure 3). First and second rows: recovered model images (Jy pixel^{-1} , displayed in \log_{10} scale) at the high and low bands (panels (a) and (d)), respectively, overlaid with zooms on the core of ESO 137-006 (panels (b) and (e)), a region with compact sources from the imaged FoV (panel (c)), and a zoom on ESO 137-007, a radio galaxy north of ESO 137-006 (panel (f)). Third row: spectral index map of ESO 137-006 (displayed in linear scale; panel (g)), overlaid with a zoom on its core (panel (h)), same region as in panels (b) and (e). Focusing on the central region (panels (b) and (e)), the first side lobe of the dirty beam is highlighted with a dashed circle. Similarly to uSARA reconstructions, one can see three filaments emerging: T_1 and T_2 , located north and south of the inner core, seen at both bands, and T_3 , located further south, recovered only at low band. A fourth filament, T_0 , detected previously, is also recovered. The filamentary structure S is an example of what mostly likely is a calibration residual artifact, as it moves with the geometry of the dirty beam. The spectral index values of the newly formed filaments (panel (h)) are inconclusive.

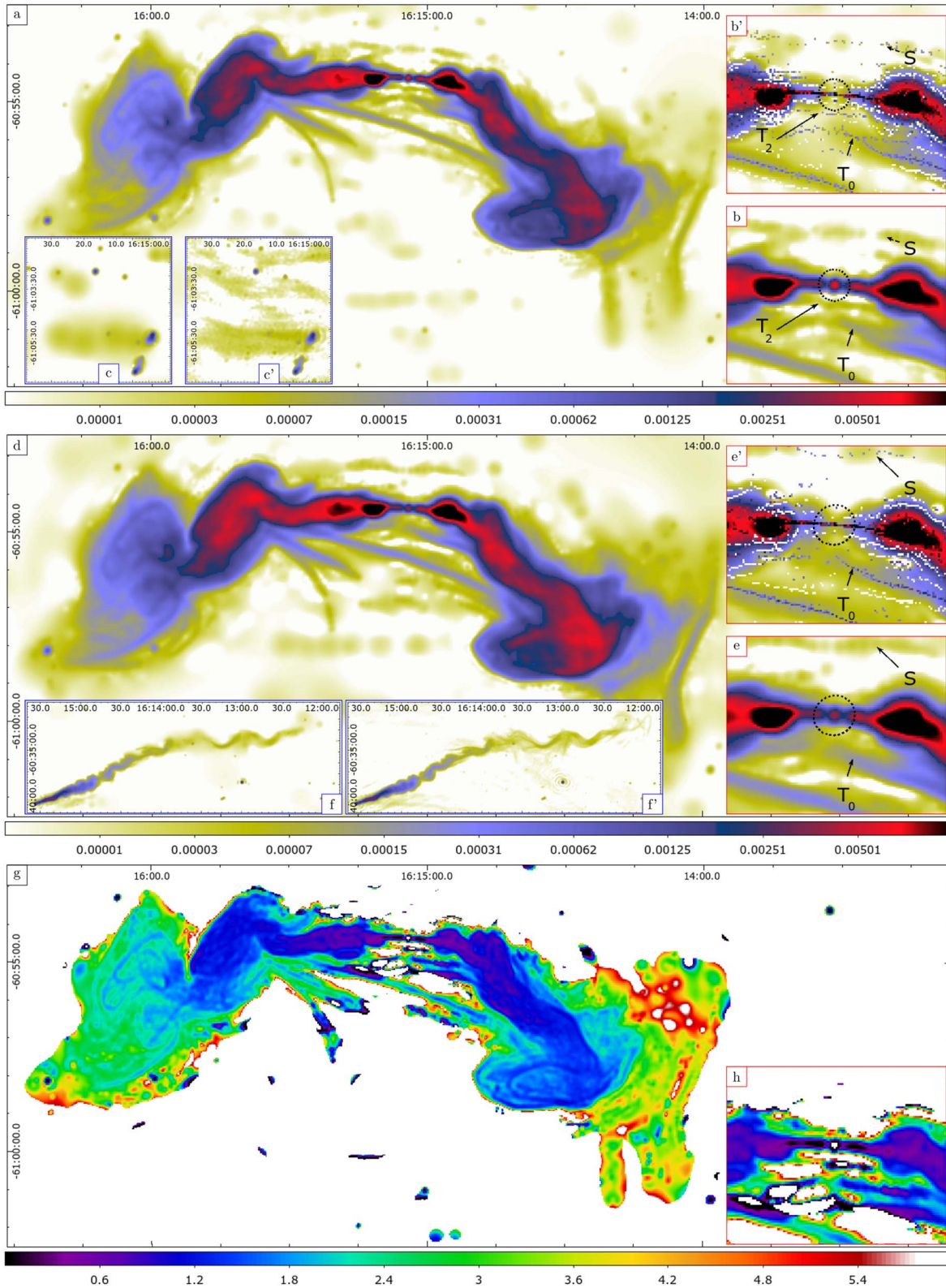


Figure 3. ESO 137-006: CLEAN reconstructions (flip pages to visualize differences at a glance with uSARA in Figure 1 and AIRI in Figure 2). First and second rows: recovered convolved model images (Jy pixel^{-1} , displayed in \log_{10} scale) at the high and low bands (panels (a) and (d)), respectively, overlaid with zooms on the core (panels (b) and (e)), a region with compact sources from the imaged FoV (panels (c) and (c')), and a zoom on ESO 137-007, a radio galaxy north of ESO 137-006 (panels (f) and (f')). Third row: estimated spectral index map of ESO 137-006 (displayed in linear scale; panel (g)), overlaid with a zoom on its core (panel (h)), same region as in panels (b), (b'), and (e). First, regarding the central region, the first side lobe of the dirty beam is highlighted with a dashed circle. Unlike the uSARA and AIRI reconstructions, only one new filament, T_2 , has clearly emerged south of the core at the high band. The previously detected filament, T_0 , is recovered at both bands. The filamentary structure S is an example of what mostly likely is a calibration residual artifact, as it moves with the geometry of the dirty beam. Inspection of the zooms on the core from CLEAN model images (panels (c') and (e')) confirms these findings. Second, looking at the zooms on the compact sources, one can see that some faint sources are not recovered in the CLEAN model image (panel (c)). Instead, they are left in the residual and so only visible in the CLEAN restored image (panel (c')). Finally, a close look at ESO 137-007 in the convolved model image (panel (f)) indicates that much of the filamentary structure is not captured and is only visible on the CLEAN restored image (panel (f')).

Table 1
Flux Measurements of the AGN in ESO 137-006 at Both Bands and Its Spectral Index Values Recovered by uSARA, AIRI, and CLEAN

	uSARA	AIRI	CLEAN	Frequency (MHz)
Flux (mJy)	143	143	146	1053
	124	125	123	1399
Spectral index	0.50	0.47	0.60	

Note. Flux measurements of CLEAN are computed from the convolved model images over a region centered at the AGN and of the size of the main lobe of the associated dirty beams. The source being super-resolved in uSARA and AIRI images, its flux measurements are computed over its active pixels.

frequencies, where the lobes exhibit steep spectra, with a spectral index close to 5 at the tails.

The central region zooms (panels (b) and (e) of each figure) highlight both the super-resolution potential of uSARA and AIRI and the difficulty of interpreting features in RI images. At the low band, both AIRI and uSARA recover what appear to be additional sets of filaments, or CSTs: T_1 and T_2 , respectively north and south of the core/jet structure; T_3 , further south; and T_0 , a filament already detected by Ramatsoku et al. (2020). The pixel intensity values of these filaments are within the range [0.1, 0.4] mJy, not only well above the image domain noise level of about 0.0014 mJy but also at least five times higher than the level of the imaging artifacts, induced by the lack of amplitude calibration. The filaments are roughly parallel to the linear core/jet structure. This is a cause for both excitement and wariness. On the one hand, one explanation for the origin of CSTs is shearing of relativistic electrons off the jets, which then follow the ambient magnetic field with possibly stretched field lines (see also Condon et al. 2021). From that point of view, additional inner filaments parallel to the jets make physical sense. On the other hand, the core/jet structure is one of the brightest features in the image, and one must always be wary of secondary image features that appear to trace bright features too closely, since calibration and deconvolution artifacts could easily take this form (being modulated by the side lobes of the dirty beam). One must therefore carefully compare reconstructions made with different methods and at different frequencies, since imaging artifacts tend to scale spectrally following the geometry of the dirty beam.

The innermost filaments T_1 and T_2 appear in both the low- and high-band images made by AIRI and uSARA. Filament T_2 is also confirmed by the CLEAN high-band image and is not inconsistent with the CLEAN low-band image, where there is emission blending with the core/jets. Filament T_1 is not inconsistent with the CLEAN images either, where there is also emission, however, not resolved at all. This is also backed up by the associated model images (Figures 3(b') and (e')), where large-scale components are recovered around the same region. The first side lobe of the dirty beam (indicated by dashed circles in panels (b) and (e) of each figure) is, in any case, only slightly smaller than the separation between the jets and T_1 and T_2 , explaining why the filaments are not well resolved by CLEAN. Finally, the core/jet structure recovered by AIRI and uSARA has a clear discontinuity between the core and the jets, while the innermost filaments show no such thing. For CLEAN, although the core and jets are fully connected on the model images (Figures 3(b') and (e')), the connection seems

Table 2
Computational Cost of uSARA, AIRI, and CLEAN in CPU Hours

	uSARA (CPU hr)	AIRI (CPU hr, GPU hr)	CLEAN (CPU hr)	Frequency (MHz)
Computation of $\Phi^\dagger\Phi^a$	427	427, —	...	1053
Imaging	1120	480, 5	132	
Computation of $\Phi^\dagger\Phi^a$	652	652, —	...	1399
Imaging	2377	1028, 6	236	

Notes. Specific to AIRI, GPU hours used by the DNN denoiser are also reported.

^a Computation cost of the underlying sparse matrices.

weaker after convolution with the restoring beam (panels (b) and (e) of the same figure). The position of the filaments shifts slightly (by about a pixel) between the low- and high-band images but in the opposite direction than would be expected from a dirty beam-modulated artifact. This is possibly due to pixelation effects in the reconstructions. The recovered spectral indices are inconclusive. On balance, we must conclude that the innermost filaments are likely to be physical CSTs and not imaging artifacts.

Although reconstructed by both uSARA and AIRI, the nature of T_3 is less conclusive, as it only appears at the low band. It may be that, at the high band, the reconstruction confuses it with the more complex structure of T_0 , just to the south of it. We note that T_0 is reconstructed by all algorithms, with an impressively clear and finely resolved east-west connection at the low band by uSARA and AIRI, while the structure is very blurred and interrupted in the CLEAN image.

Finally, examples of what are almost certainly artifacts are the fainter extended structures labeled S. They are roughly parallel to the radio galaxy structure and scale inwardly with the dirty beam in the high-band images. Since they are reconstructed by all three methods, they are likely to be residual amplitude calibration errors.

Computational cost. Table 2 summarizes the computational cost of the imaging algorithms. Specific to uSARA and AIRI, the computation cost associated with the decomposition of the measurement operator is reported alongside the cost to run the imaging algorithm. As expected, with AIRI leveraging a fast denoiser on GPU and uSARA relying on a subiterative denoiser on CPU, the former brings a significant reduction of the imaging cost over the latter, about 2.3 times fewer CPU hours at both bands, with a negligible amount of GPU hours. The AIRI was only four times more expensive than WSClean in the imaging process and seven when including the computation cost of the measurement operator. As AIRI denoisers are trained completely independently of the data to be imaged, the training cost associated with the single denoiser used for both bands is not considered part of the computational cost.

6. Conclusions

We have introduced the first AI-based framework for deep, super-resolution, wide-field RI imaging based on a plug-and-play scheme whereby a dedicated denoising operator is injected as an image regularizer in an optimization algorithm. We have

demonstrated two image reconstruction algorithms, uSARA and AIRI, respectively propelled by powerful handcrafted and learned denoisers, aiming at delivering a high level of imaging precision. Both algorithms are highly parallelized for scalability via automated image faceting and decomposition of the RI measurement operator into sparse low-dimensional building blocks. An in-depth study of practical scalability to the extreme data and image dimensions expected in the SKA context, in particular for wideband imaging, is warranted. The uSARA and AIRI were used to revisit MeerKAT *L*-band observations of a wide FoV containing ESO 137-006 from 19 GB of visibility data. Our results confirm the ability of uSARA and AIRI to access a new regime of imaging resolution and dynamic range with respect to CLEAN. In particular, we have studied the wealth of filamentary structure revealed within ESO 137-006’s radio lobes, some of which are likely CSTs. Our results also demonstrate further improvement brought by AIRI over uSARA in both dynamic range and speed, underpinned by the hybrid approach at the interface of optimization and AI.

The research of A.D., A.J., and Y.W. was supported by the UK Research and Innovation under the EPSRC grant EP/T028270/1 and STFC grant ST/W000970/1. The work of O.S. was supported by the South African Research Chairs Initiative of the National Research Foundation (NRF), an agency of the Department of Science and Innovation (DST). The research of M.R. is supported by the New Scientific Frontiers grant of the South African Radio Astronomy Observatory (SARAO). The research used Cirrus, a UK National Tier-2 HPC Service at EPCC funded by the University of Edinburgh and EPSRC (EP/P020267/1). The MeerKAT telescope is operated by SARAO, which is a facility of DST/NRF.

Data Availability

The utilized data are observations with the MeerKAT telescope (Project ID SCI-20190418-SM-01). The images underlying this article are available from the research portal of Heriot-Watt University with the digital object identifier [10.17861/b3a4ea6b-805c-4629-8c5c-fbb1f06ab53a](https://doi.org/10.17861/b3a4ea6b-805c-4629-8c5c-fbb1f06ab53a).

Appendix A The FB Algorithmic Structure

In this work, the inverse RI imaging problem is approached as an optimization problem. In the context of optimization theory, an “objective function” is defined, typically as the sum of a data fidelity term f and a regularization term r injecting a prior image model to compensate for data incompleteness. The image estimate is defined as the minimizer of this objective and reached via provably convergent algorithms (Bauschke & Combettes 2017). The obtained solution can also be understood in a Bayesian framework as a maximum a posteriori estimate with respect to a posterior distribution, the negative logarithm of which is the objective. More specifically to our setting, we aim at solving

$$\underset{\mathbf{x} \in \mathbb{R}^N}{\text{minimize}} f(\mathbf{x}; \mathbf{y}) + \lambda r(\mathbf{x}). \quad (\text{A1})$$

In this objective, f is a convex Lipschitz differentiable function of a variable $\mathbf{x} \in \mathbb{R}^N$ representing the image variable, also a function of the data vector $\mathbf{y} \in \mathbb{C}^M$, whose role is to enforce fidelity to data. The function r is a convex and possibly

nondifferentiable function of \mathbf{x} , encoding the prior image model. The regularization parameter $\lambda > 0$ acts as a trade-off between the two terms. Problems of the form (A1) can be solved via the iterative FB algorithm, alternating between a forward step in the negative direction of the gradient of f and a backward step involving a simple denoising operator, known as the proximal operator of the regularization function r . The proximal operator is itself defined as the solution of a (simpler) minimization problem, $\text{prox}_{\lambda r}(\mathbf{z}) = \text{argmin}_{\mathbf{u} \in \mathbb{R}^N} \lambda r(\mathbf{u}) + \|\mathbf{z} - \mathbf{u}\|_2^2/2$, for any $\mathbf{z} \in \mathbb{R}^N$ and $\lambda > 0$. The proximal operator of simple functions r often benefits from a closed-form solution (e.g., the proximal operator of the ℓ_1 norm is a simple component-wise soft-thresholding operator). However, proximal denoisers of sophisticated regularizations must usually be computed iteratively as solutions of the minimization task by which they are defined.

The FB iterative structure reads

$$(\forall k \in \mathbb{N}) \quad \mathbf{x}^{k+1} = \text{prox}_{\gamma \lambda r}(\mathbf{x}^k - \gamma \nabla f(\mathbf{x}^k)), \quad (\text{A2})$$

where the step size $\gamma > 0$ is strictly upper-bounded by $2/L$ to ensure convergence, with L being the Lipschitz constant of ∇f .

Interestingly, the recently emerged PnP scheme has established that proximal optimization algorithms such as FB enable not only the use of proximal operators of handcrafted regularization functions but also the injection of learned DNN denoisers defining the regularization term implicitly (Venkatakrisnan et al. 2013; Romano et al. 2017). We note that, in order to preserve algorithm convergence and the interpretability of its solution, the PnP denoiser must typically satisfy a “firm nonexpansiveness” constraint, ensuring that it contracts distances (Pesquet et al. 2021; Hurault et al. 2022).

Our RI imaging framework relies on a data fidelity term that reflects the Gaussian nature of the noise and is given by $f(\mathbf{x}; \mathbf{y}) = 1/2 \|\Phi \mathbf{x} - \mathbf{y}\|_2^2$, where $\|\cdot\|_2$ denotes the standard ℓ_2 norm. Its gradient reads $\nabla f(\mathbf{x}) = \Phi^\dagger \Phi \mathbf{x} - \Phi^\dagger \mathbf{y}$, where $(\cdot)^\dagger$ denotes the adjoint of its argument. The Lipschitz constant of ∇f is given by $L = \|\text{Re}\{\Phi^\dagger \Phi\}\|_S$, where $\|\cdot\|_S$ denotes the spectral norm. As for the image regularization, we leverage the versatility of the PnP framework and investigate both advanced handcrafted and learned prior image models, respectively propelling the uSARA and AIRI imaging algorithms.

Appendix B uSARA’s Handcrafted Denoiser

The image model of uSARA, originally proposed in Carrillo et al. (2012), promotes the nonnegativity of the intensity image and its sparsity in an overcomplete dictionary $\Psi \in \mathbb{R}^{N \times B}$, which consists of a normalized concatenation of orthogonal wavelet bases. The sparsity model is encoded via a nondifferentiable log-sum regularization function r , generalizing the ℓ_1 norm. The resulting multiterm regularization thus reads (Thouvenin et al. 2022a; Repetti & Wiaux 2021)

$$\lambda r(\mathbf{x}) = \lambda \sum_{n=1}^B \rho \log(1 + \rho^{-1} |(\Psi^\dagger \mathbf{x})_n|) + \iota_{\mathbb{R}_+^N}(\mathbf{x}), \quad (\text{B1})$$

where $(\cdot)_n$ denotes the n^{th} coefficient of its argument vector, and $\iota_{\mathbb{R}_+^N}$ denotes the indicator function of the real positive orthant, imposing the nonnegativity constraint $\iota_{\mathbb{R}_+^N}(\mathbf{x}) = +\infty$ if $\mathbf{x} \notin \mathbb{R}_+^N$ and zero otherwise. The log-sum regularization being nonconvex, the minimization task is approached via a

reweighting procedure, where a series of convex surrogate minimization tasks, composed of weighted- ℓ_1 regularization with nonnegativity constraints, are solved using FB (Repetti & Wiaux 2020, 2021; Terris et al. 2022). The denoising proximal operator of the resulting multiterm regularization does not have a closed-form solution and is solved iteratively. In other words, uSARA relies on FB with subiterative regularization denoisers.

We note that the parameter λ controls a soft-thresholding operation acting on the wavelet coefficients. As proposed by Terris et al. (2022), the exact thresholding parameter $\gamma\lambda$ is set equal to the measurement noise transferred to the image domain $\tau/\sqrt{2L}$,⁶

$$\gamma\lambda = \tau/\sqrt{2L}, \quad (\text{B2})$$

with the step size γ typically set to $\gamma = 1.98/L$. The parameter $\rho > 0$ represents a floor level on the wavelet coefficients and is naturally set to the noise level $\rho = \gamma\lambda$ (Thouvenin et al. 2022a).

Appendix C AIRI's Learned Denoiser

Following Terris et al. (2022), we trained a convolutional DNN with a simple DnCNN architecture (Zhang et al. 2017) on a rich, synthetic database \mathcal{U} of normalized images with an adaptive dynamic range. The training loss is a classical ℓ_1 loss enhanced with a firm nonexpansiveness constraint on the denoiser,

$$\begin{aligned} & \underset{\mathbb{D}}{\text{minimize}} \mathbb{E}_{\mathbf{u} \sim \mathcal{U}, \mathbf{n} \sim \mathcal{N}(0,1)} [\|\mathbb{D}(\mathbf{u} + \sigma\mathbf{n}) - \mathbf{u}\|_1] \\ & \text{such that } (\forall \mathbf{u} \in \mathbb{R}^N) \|\nabla_{\mathbf{u}}(2\mathbb{D} - \mathbb{I})\|_s \leq 1, \end{aligned} \quad (\text{C1})$$

where $\mathbf{u} \in \mathbb{R}^N$ are samples of the training database \mathcal{U} , $\mathbf{n} \sim \mathcal{N}(0, 1)$ is the additive Gaussian random noise, $\sigma > 0$ is the training noise level, \mathbb{E} denotes the expectation taken over \mathbf{u} and \mathbf{n} , \mathbb{I} denotes the identity operator, and $\|\cdot\|_1$ denotes the ℓ_1 norm.

We emphasize that training under the firm nonexpansiveness constraint is a highly challenging task. As proposed in Pesquet et al. (2021), in practice, we relax the constraint and introduce a variant of the regularization in Equation (C1), which penalizes softly nonfirmly nonexpansive networks. We further note that, while Terris et al. (2022) demonstrated in simulation that this leads to a robust way to ensure convergence of the resulting PnP algorithms, we have witnessed that, when used for real data and at large image sizes and dynamic ranges, such as those of interest here, some denoisers lead to algorithm instability, requiring further training.

The performance of the learned image regularization is highly dependent on the training noise level σ , the impact of which mirrors that of the regularization parameter λ in uSARA's denoiser. Terris et al. (2022) proposed a heuristic according to which σ should be set equal to the measurement noise transferred to the image domain $\tau/\sqrt{2L}$. However, the training database is normalized, with peak image values upper-bounded by 1. To avoid any generalizability issues, the trained denoisers should therefore be used on similarly normalized images, which was the case for the test images in the simulation

framework of Terris et al. (2022). In general, this constraint can be accommodated by rescaling the inverse problem (Equation (1)), effectively dividing it by an upper bound on the peak intensity of the sought image, $\alpha \geq \max_j \{x_j\}$, which can be inferred from the peak of the dirty image. The rescaled inverse problem,

$$\mathbf{y}/\alpha = \Phi(\bar{\mathbf{x}}/\alpha) + \bar{\mathbf{n}}/\alpha, \quad (\text{C2})$$

now targets the recovery of $\bar{\mathbf{x}}/\alpha$, with a peak value upper-bounded by 1. As a result, the heuristic generalizes to setting σ equal to the inverse input image domain peak signal-to-noise ratio, which can be understood as the target reconstruction dynamic range rather than an absolute noise level:

$$\sigma = \tau/\alpha\sqrt{2L}. \quad (\text{C3})$$

Interestingly, if a predefined denoiser trained at some high dynamic range is available, any RI data set with a signal-to-noise ratio a priori pointing to a lower dynamic range denoiser can be further rescaled (with a larger α) to match the existing denoiser, according to Equation (C3). Here we adopt this single denoiser approach, where different data are matched to the denoiser, rather than the contrary. Naturally, PnP solutions are multiplied by α after reconstruction.

Appendix D Denoiser Faceting

Specific to the algorithm scalability requirement arising from the large image dimensions of interest, we propose an automated parallelization of the studied denoisers enabled by image faceting. First, uSARA's proximal denoiser takes advantage of a faceted implementation of the sparsity dictionary Ψ enabled by its convolutional nature and the compact support of the wavelet kernels (Pruša 2012). The number of facets is set to optimize the parallelization of the processing across the available CPUs under communication constraints. Second, AIRI's learned denoiser is decomposed and applied independently across the facets of the image. This procedure is enabled by the convolutional nature of the DNNs, which rely on kernels with compact support, in turn yielding a small receptive field (Luo et al. 2016). The number of facets is set to optimize the parallel processing across the available GPUs for scalability under memory constraints.

Appendix E Parallel Wide-field Measurement Operator

First, on large FoVs, such as the one of interest here, the w -component of the baselines induces a nonnegligible baseline-dependent chirp-like phase modulation on the radio sky (Cornwell et al. 2005; Wiaux et al. 2009). This w -effect can be formulated in closed form and needs to be accounted for in the model of the measurement operator. Its modeling as a simple phase modulation in the image domain for each baseline is, however, impractical when used in combination with the fast Fourier transform (FFT) underpinning the fast implementation of $\mathbf{F} \in \mathbb{C}^{N' \times N'}$ in Equation (1), which computes all of the discrete coefficients of the Fourier plane at once, rather than a selected (u, v) point. For accurate and computationally efficient modeling, we consider a hybrid approach combining the w -stacking (Offringa et al. 2014) and w -projection approaches (Cornwell et al. 2005), whereby the measurements are grouped

⁶ When considering a data weighting scheme other than natural weighting, such as uniform or Briggs weighting, a multiplicative correction factor is applied to L for a more accurate noise estimate in the image domain (A. Wilber et al. 2022, in preparation).

into P w -stacks composed of M_p data points each, with $1 \leq p \leq P$, resulting from binning the visibilities in the w -dimension. The w -modulation of each visibility is decomposed into two components: (i) a large phase modulation associated with the central w -value of the w -stack to which it belongs, incorporated into the measurement operator through phase modulation in the image domain, and (ii) an offset phase modulation injected through convolution with a small w -kernel (Dabbech et al. 2017) in the Fourier plane. The resulting measurement operator Φ is decomposed into a series of sparse operators as

$$\Phi = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_P \end{bmatrix}, \quad (\text{E1})$$

where the operator $\Phi_p = \mathbf{G}_p \mathbf{F} \mathbf{Z}_p \in \mathbb{C}^{M_p \times N}$ is the measurement operator associated with the p th w -stack. More specifically, $\mathbf{Z}_p \in \mathbb{C}^{N' \times N}$ denotes the zero-padding operator that encompasses the w -modulation of the associated w -stack, in addition to the correction for the convolution with the approximate nonuniform Fourier transform interpolation kernels, and $\mathbf{G}_p \in \mathbb{C}^{M \times N'}$ is the sparse degridding matrix encoding row-based convolutions between these kernels and the small w -kernels implementing the phase modulation of the w -offsets in the Fourier plane, combining w -stacking and w -projection results in a memory-efficient and accurate measurement operator. We also emphasize that any DDE calibration solutions modeled as Fourier kernels can be easily injected into the measurement operator model via further row-based convolution (Dabbech et al. 2021).

Second, the operator of interest in the gradient step of the FB iterative structure (Equation (A2)) is not $\Phi \in \mathbb{C}^{M \times N}$ but rather $\Phi^\dagger \Phi \in \mathbb{C}^{N \times N}$, which now reads

$$\Phi^\dagger \Phi = \sum_{p=1}^P \Phi_p^\dagger \Phi_p, \quad (\text{E2})$$

with $\Phi_p^\dagger \Phi_p = \mathbf{Z}_p^\dagger \mathbf{F}^\dagger \mathbf{H}_p \mathbf{F} \mathbf{Z}_p \in \mathbb{C}^{N \times N}$, and where the holographic matrices $\mathbf{H}_p = \mathbf{G}_p^\dagger \mathbf{G}_p \in \mathbb{C}^{N' \times N'}$ encode both the degridding and gridding steps. The scalability of $\Phi^\dagger \Phi$ to large data acquisition regimes is promoted by enabling three key features.

First, a dimensionality reduction feature to reduce the memory requirements of $\Phi^\dagger \Phi$ is supported. The functionality consists in gridding the data and encoding the degridding and gridding steps as a single operation with the holographic matrices \mathbf{H}_p directly implemented as sparse operators. By doing so, the operator $\Phi^\dagger \Phi$ becomes effectively blind to the data dimension M .

Second, a planning strategy to automate the choice of the number of the w -stacks and the decision to enable the dimensionality reduction from a subset of the data is devised. In the first instance, estimates of the computational complexity of the application of $\Phi^\dagger \Phi$ (derived from the number of FFTs and the sparsity of the degridding matrices) and of the memory required to host the degridding matrices are obtained for a wide range of values of the w -stacks number. The retained value is the one presenting the best trade-off between the computational cost and the memory requirements under constraints set by the computing architecture on which the imaging algorithm is

deployed (number of compute nodes, number of CPUs per node, available memory per CPU, etc). Data dimensionality reduction via visibility gridding is enabled when the memory requirements exceed the available resources.




Third, a fully automated parallelization of $\Phi^\dagger \Phi$ is achieved through memory-based partitioning of its underlying degridding/holographic matrices and data vectors. The sparse matrices are computed as part of the initialization of the imaging algorithm. A data-clustering step is first performed in parallel for each w -stack to further distribute its degridding/holographic matrix into blocks. The clusters are made of visibilities belonging to the same radial slice of the Fourier plane, minimizing the amount of underpinning discrete Fourier coefficients and subsequent communication requirements. The angular opening of each radial slice is determined by the memory needed to compute and host the resulting blocks. From the identified number of clusters, the CPUs dedicated to the forward step are allocated. The blocks of the degridding/holographic matrices are then computed only once and hosted directly on the compute nodes to be applied in parallel at each FB iteration.

Appendix F Automated Parameter Selection

The estimated image noise levels at the low and high bands are, respectively, $\tau/\sqrt{2L} \simeq 0.0014$ and 0.0017 mJy. The peak intensity values as estimated from the normalized⁷ dirty images are 0.69 Jy at the low band and 0.37 Jy at the high band. For uSARA, $\gamma\lambda$ and ρ are set equal to the estimated noise levels, per Equation (B2). For AIRI, the estimated noise and peak values suggest target dynamic ranges of 5×10^5 and 2.2×10^5 at the low and high bands, respectively. Owing to the chosen normalization of the dirty image for peak estimation, the dirty peak value consistently overestimates the true peak value, so that the real target dynamic ranges are below these values.⁸ In this context, we have used a single denoiser trained for target dynamic range 4×10^5 , rescaling the inverse problems by the appropriate α at each band independently as in Equation (C2). In other words, after rescaling, and as in Equation (C3), the inverse problem at each band is affected by a noise of standard deviation $\tau/\alpha\sqrt{2L}$ equal to the training noise level of the chosen denoiser, i.e., $\sigma = 2.5 \times 10^{-6}$ Jy.

The uSARA and AIRI denoisers were respectively applied on 8×8 and 2×2 facet decompositions of the images. The low- and high-band data were decomposed into 12 and 14 w -stacks in Equation (E1), with $\Phi^\dagger \Phi$ encoded via the holographic matrices \mathbf{H}_p . At the low and high bands, respectively, this enabled lower memory requirements from 470 and 645 GB needed to host the degridding matrices \mathbf{G}_p down to 81 and 159 GB. Finally, in WSClean, multiscale CLEAN utilized 72 w -stacks for both bands.

ORCID iDs

A. Dabbech  <https://orcid.org/0000-0002-7903-3619>
A. Jackson  <https://orcid.org/0000-0003-0073-682X>
M. Ramatsoku  <https://orcid.org/0000-0003-0231-3249>

⁷ The dirty images are normalized by $\beta = \max_i(\Phi^\dagger \Phi \delta)_i$, where δ is an image with a value of 1 at the phase center and zero otherwise. By doing so, the dirty beam $\Phi^\dagger \Phi \delta / \beta$ has a peak value equal to 1.

⁸ In fact, the peak values reconstructed by uSARA and AIRI are around 0.05 Jy for both bands, suggesting an initial overestimation with the normalized dirty image by more than a factor of 10. Note that this uncertainty in the initial estimate is not a problem, as only an upper bound is required.

O. M. Smirnov  <https://orcid.org/0000-0003-1680-7936>
 Y. Wiaux  <https://orcid.org/0000-0002-1658-0121>

References

- Abdulaziz, A., Dabbech, A., & Wiaux, Y. 2019, *MNRAS*, **489**, 1230
- Bauschke, H. H., & Combettes, P. L. 2017, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Vol. 468 (Cham: Springer)
- Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, *MNRAS*, **426**, 1223
- Chouzenoux, E., Pesquet, J.-C., & Repetti, A. 2016, *J. Glob. Optim.*, **66**, 457
- Condon, J. J., Cotton, W. D., White, S. V., et al. 2021, *ApJ*, **917**, 18
- Connor, L., Bouman, K. L., Ravi, V., & Hallinan, G. 2022, *MNRAS*, **514**, 2614
- Cornwell, T. J. 2008, *ISTSP*, **2**, 793
- Cornwell, T. J., Golap, K., & Bhatnagar, S. 2005, in *ASP Conf. Ser.* 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 86
- Cornwell, T. J., & Perley, R. A. 1992, *A&A*, **261**, 353
- Dabbech, A., Ferrari, C., Mary, D., et al. 2015, *A&A*, **576**, A7
- Dabbech, A., Onose, A., Abdulaziz, A., et al. 2018, *MNRAS*, **476**, 2853
- Dabbech, A., Repetti, A., Perley, R. A., Smirnov, O. M., & Wiaux, Y. 2021, *MNRAS*, **506**, 4855
- Dabbech, A., Terris, M., Jackson, A., et al. 2022, *Unveiling structure in ESO 137-006: AIRI and uSARA reconstructions*, doi:[10.17861/b3a4ea6b-805c-4629-8c5c-fbb1f06ab53a](https://doi.org/10.17861/b3a4ea6b-805c-4629-8c5c-fbb1f06ab53a)
- Dabbech, A., Wolz, L., Pratley, L., McEwen, J. D., & Wiaux, Y. 2017, *MNRAS*, **471**, 4300
- Fessler, J. A., & Sutton, B. P. 2003, *ITSP*, **51**, 560
- Garsden, H., Girard, J. N., Starck, J. L., et al. 2015, *A&A*, **575**, A90
- Högbom, J. A. 1974, *A&AS*, **15**, 417
- Hurault, S., Leclaire, A., & Papadakis, N. 2022, arXiv:[2201.13256](https://arxiv.org/abs/2201.13256)
- Jonas, J. & MeerKAT Team 2016, in *MeerKAT Science: On the Pathway to the SKA* (Trieste: SISSA), 1
- Józsa, G. I. G., White, S. V., Thorat, K., et al. 2020, in *ASP Conf. Ser.* 527, *Astronomical Data Analysis Software and Systems XXIX*, ed. R. Pizzo et al. (San Francisco, CA: ASP), 635
- Kenyon, J. S., Smirnov, O. M., Grobler, T. L., & Perkins, S. J. 2018, *MNRAS*, **478**, 2399
- Li, F., Cornwell, T. J., & de Hoog, F. 2011, *A&A*, **528**, A31
- Luo, W., Li, Y., Urtasun, R., & Zemel, R. 2016, *NIPS*, **29**, 4905
- Makhathini, S. 2018, PhD thesis, Rhodes Univ.
- Offringa, A. R., & Smirnov, O. 2017, *MNRAS*, **471**, 301
- Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, *MNRAS*, **444**, 606
- Onose, A., Dabbech, A., & Wiaux, Y. 2017, *MNRAS*, **469**, 938
- Pesquet, J.-C., & Repetti, A. 2014, arXiv:[1406.6404](https://arxiv.org/abs/1406.6404)
- Pesquet, J.-C., Repetti, A., Terris, M., & Wiaux, Y. 2021, *SIAM J. Imaging Sci.*, **14**, 1206
- Pruša, Z. 2012, PhD thesis, Brno Univ. of Technology
- Ramatsoku, M., Murgia, M., Vacca, V., et al. 2020, *A&A*, **636**, L1
- Repetti, A., Birdi, J., Dabbech, A., & Wiaux, Y. 2017, *MNRAS*, **470**, 3981
- Repetti, A., & Wiaux, Y. 2020, in *IEEE ICASSP* (Piscataway, NJ: IEEE), 1434
- Repetti, A., & Wiaux, Y. 2021, *SIAM J. Optim.*, **31**, 1215
- Romano, Y., Elad, M., & Milanfar, P. 2017, *SIAM J. Imaging Sci.*, **10**, 1804
- Scaife, A. 2020, *RSPTA*, **378**, 20190060
- Smirnov, O. M. 2011, *A&A*, **527**, A107
- Terris, M., Dabbech, A., Tang, C., & Wiaux, Y. 2022, *MNRAS*, in press
- Thouvenin, P.-A., Abdulaziz, A., Dabbech, A., Repetti, A., & Wiaux, Y. 2022a, *MNRAS*, in press (arXiv:[2003.07358](https://arxiv.org/abs/2003.07358))
- Thouvenin, P.-A., Dabbech, A., Jiang, M., et al. 2022b, arXiv:[2209.07604](https://arxiv.org/abs/2209.07604)
- Venkatakrisnan, S. V., Bouman, C. A., & Wohlberg, B. 2013, in *IEEE GlobalSIP* (Piscataway, NJ: IEEE), 945
- Wakker, B. P., & Schwarz, U. J. 1988, *A&A*, **200**, 312
- Wiaux, Y., Puy, G., Boursier, Y., & Vanderghynst, P. 2009, *MNRAS*, **400**, 1029
- Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. 2017, *ITIP*, **26**, 3142