Exploring the Universe with Artificial Intelligence (UniversAI)
Proceedings IAU Symposium No. 397, 2026
I. Liodakis & V. Efthymiou, eds.
doi:10.1017/S1743921325104468

To CLEAN or not to CLEAN: Data Processing in the ngVLA era

Hendrik Mueller

National Radio Astronomy Observatory, 1011 Lopezville Road, 87801 Socorro, New Mexico, USA
email: hmuller@nrao.edu

Abstract. Radio interferometric imaging has long relied on the CLEAN algorithm, valued for its speed, robustness, and integration with calibration pipelines. However, next-generation facilities such as the ngVLA, SKA, and ALMA's Wideband Sensitivity Upgrade will produce data volumes and dynamic ranges that exceed the scalability of traditional methods. CLEAN remains dominant due to its simplicity and accumulated expertise, yet its assumption of modeling the sky as point sources limits its ability to recover extended emission and hampers automation. We review CLEAN's limitations and survey alternatives, including multiscale extensions, compressive sensing, Regularized Maximum Likelihood, Bayesian inference, and AI-driven approaches. Forward-modeling methods enable higher fidelity, flexible priors, and uncertainty quantification, albeit at greater computational cost. Hybrid approaches such as Autocorr-CLEAN, CG-CLEAN, and PolyCLEAN retain CLEAN's workflow while incorporating modern optimization. We argue hybrids are best suited for the near term, while Bayesian and AI-based frameworks represent the long-term future of interferometric imaging.

Keywords. Techniques: interferometric, Techniques: image processing

1. Introduction

Radio interferometry is a cornerstone of modern astrophysics, enabling studies of phenomena from supernova remnants to active galactic nuclei. Signals from multiple antennas are correlated, and by the van Cittert-Zernike theorem these correlations approximate Fourier components of the sky brightness (Thompson et al. 2017). In practice, instrumental effects (e.g., beam shapes, bandpasses, leakage, noise, digitization) and astrophysical corruptions (e.g., RFI, opacity, transients) complicate the measurement. Sparse Fourier sampling further turns image reconstruction into a deconvolution problem with strong sidelobe contamination.

Conventional processing couples calibration and imaging in iterative loops (Fig. 1). CLEAN and its multiscale and multifrequency variants (Högbom 1974; Bhatnagar & Cornwell 2004; Cornwell 2008; Rau & Cornwell 2011; Offringa et al. 2014; Offringa & Smirnov 2017; Müller & Lobanov 2023b; Müller & Bhatnagar 2025) remain the standard for minor-loop deconvolution, while major loops compare models to visibilities through gridding/degridding. High-fidelity imaging requires repeated refinements, making calibration, imaging, and quality control inseparable. For the scope of this summary however, we limit the focus of the discussion mainly to the deconvolution stage. We refer the interested reader to Bhatnagar et al. (2025) for the description of the full algorithm architecture for radio interferometry data processing, and detailed considerations regarding its high performance computing implementation.

© The Author(s), 2026. Published by Cambridge University Press on behalf of The International Astronomical Union.

https://doi.org/10.1017/S1743921325104468 Published online by Cambridge University Press

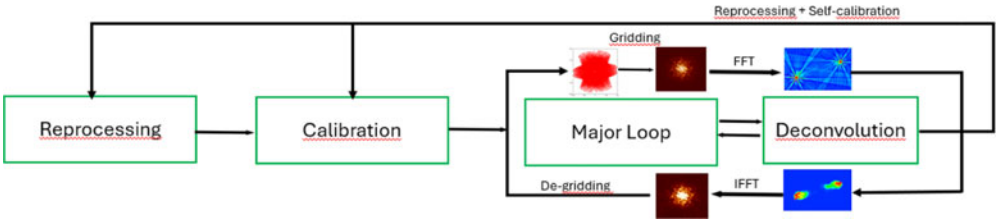


Figure 1. Traditional data processing workflow.

Beyond CLEAN, alternatives such as compressive sensing, Regularized Maximum Likelihood, Bayesian inference, and machine learning are rapidly advancing, and their systematic use in large-scale experiments marks a key frontier.

Next-generation facilities, including the ngVLA, SKA, and ALMA's Wideband Sensitivity Upgrade (WSU), will produce unprecedented data volumes with higher dynamic range, broader bandwidths, and finer resolution. These advances place heavy demands on imaging pipelines for speed, accuracy, and automation. For the WSU, computing requirements reach $\sim$100 Tflop/s, while the ngVLA will require $\sim$50 Petaflop/s, with gridding as the dominant cost (Bhatnagar et al. 2021; Kepley et al. 2024). Even modest algorithmic improvements at this stage can yield significant savings. In the post-Moore era, scalability depends on massive parallelization, especially via GPUs. Quantum computing is not expected to solve the data size problem in radio interferometry in the near future (Brunet et al. 2024).

Future-ready pipelines require scalable algorithms that accelerate major loops through optimized GPU implementations and/or reduce their number through faster convergence. While higher sensitivity may necessitate advanced methods for super-resolution and extreme dynamic range, the majority of science cases should remain well served by CLEAN. The challenge is therefore to balance novel, high-fidelity methods with robust, efficient algorithms that scale to upcoming arrays, addressing both immediate upgrades and longer-term developments.

2. Beyond Traditional CLEAN

The CLEAN algorithm (Högbom 1974) remains the workhorse of interferometric deconvolution. Its peak-finding and beam-subtraction procedure is fast, robust, and tightly integrated with calibration workflows. With heuristics such as automasking (Kepley et al. 2020) and adaptive stopping (Homan et al. 2024), CLEAN is pragmatic, extensible, and, especially in multiscale variants, still competitive to newer approaches including compressive sensing, Bayesian inference, and AI (Offringa & Smirnov 2017; Bester et al. 2024).

Nonetheless, increasing sensitivity and data complexity expose its limits. CLEAN's assumption that the sky is a collection of point sources fails for extended or diffuse emission, while the method's reliance on user input (masks, gains, stopping rules) restricts automation. It also enforces a rigid separation between model and image and offers no uncertainty quantification. Alternatives based on sparse modeling, Bayesian inference, and machine learning provide more flexible priors, potentially higher fidelity reconstructions, and principled error estimates.

Imaging algorithms are often divided into inverse modeling and forward modeling approaches (Event Horizon Telescope Collaboration 2019). Inverse modeling refines CLEAN through improved basis functions and heuristics. MS-CLEAN and multifrequency variants (Cornwell 2008; Rau & Cornwell 2011; Offringa et al. 2014;

https://doi.org/10.1017/S1743921325104468 Published online by Cambridge University Press

Offringa & Smirnov 2017) use Gaussian components of varying scale sizes, Asp-CLEAN (Bhatnagar & Cornwell 2004; Hsieh & Bhatnagar 2021; Hsieh et al. 2022a,b) optimizes scale sizes adaptively, and wavelet-based methods such as DoB-CLEAN (Müller & Lobanov 2023b) introduce elliptical difference-of-Gaussian bases. These approaches significantly improve fidelity while still remaining competitive with more advanced forward modeling paradigms (Offringa & Smirnov 2017; Hsieh & Bhatnagar 2021; Hsieh et al. 2022b; Bester et al. 2024).

Forward modeling, in contrast, formulates imaging as a direct optimization problem. Maximum entropy methods evolved into Regularized Maximum Likelihood, balancing data fidelity and multiple regularization terms (Cornwell & Evans 1985; Briggs 1995), and have been extensively developed in the Event Horizon Telescope context (Chael et al. 2016; Akiyama et al. 2017; Chael et al. 2018). More recent directions include multiobjective evolution to explore multimodality in degenerate solutions (Müller et al. 2023; Mus et al. 2024b,a), compressive sensing with sparsity-promoting priors (Starck et al. 2002, 2005; Wiaux et al. 2009; Carrillo et al. 2012, 2014; Dabbech et al. 2018; Müller & Lobanov 2022), and Bayesian inference frameworks such as `resolve` (Junklewitz et al. 2016; Knollmüller & Enßlin 2019; Frank et al. 2021; Arras et al. 2021a; Roth et al. 2024) or sampling-based methods in the EHT (Broderick et al. 2020; Tiede 2022). AI-based techniques are also emerging rapidly (Sun et al. 2022; Connor et al. 2022; Terris et al. 2023; Aghabiglou et al. 2024; Dabbech et al. 2024).

Forward-modeling methods often achieve superior fidelity for complex structures, potentially offer an uncertainty quantification and can directly address nonlinear corruptions, e.g. closure-only imaging and polarization (Chael et al. 2018; Müller et al. 2023; Müller 2024), joint calibration and imaging (Arras et al. 2019; Kim et al. 2024), direction-dependent effects (Dabbech et al. 2021; Arras et al. 2021b; Roth et al. 2023), or time-dynamic sources (Bouman et al. 2017; Arras et al. 2022; Müller & Lobanov 2023a; Mus & Martí-Vidal 2024; Mus et al. 2024a). Their main drawback typically is computational cost, which remains a barrier to routine use.

Despite these advances, CLEAN persists as the de facto standard, valued for its speed, competitiveness of its modern multiscalar variants, and simply sufficiency for many science cases. Novel paradigms have shown improved performance in targeted tests, but large-scale validation across observatory pipelines is lacking, and deployment will require substantial additional development and commissioning.

Hybrid approaches offer strong near-term potential for pipelines supporting upcoming facilities such as SKA and ALMA's WSU. They preserve the speed, robustness, and validation of CLEAN while replacing its suboptimal regularization, providing an evolutionary improvement without requiring new infrastructure. Notable examples include Autocorr-CLEAN (Müller & Bhatnagar 2025), CG-CLEAN (Müller et al. 2025), and PolyCLEAN (Jarret et al. 2025). Even modern AI and Bayesian methods, such as R2D2 (Aghabiglou et al. 2024) and fast-resolve (Roth et al. 2024), adopt major/minor loop structures for efficiency.

PolyCLEAN integrates CLEAN with an explicit LASSO formulation, while CG-CLEAN interprets the major loop as a gradient step and the minor loop as a Hessian inversion, casting the process into a conjugate gradient scheme. This accelerates convergence and drives residuals toward a noise-like distribution more efficiently than traditional CLEAN (Müller et al. 2025). We show the residuals as a function of number of major loops exemplary in Fig. 2.

Other hybrids focus on enhancing multiscale representations. Asp-CLEAN (Bhatnagar & Cornwell 2004) and DoB-CLEAN (Müller & Lobanov 2023b) improve accuracy by introducing flexibility in scale size or ellipticity, extending the gains of MS-CLEAN (Cornwell 2008). Autocorr-CLEAN instead clusters components via residual

https://doi.org/10.1017/S1743921325104468 Published online by Cambridge University Press

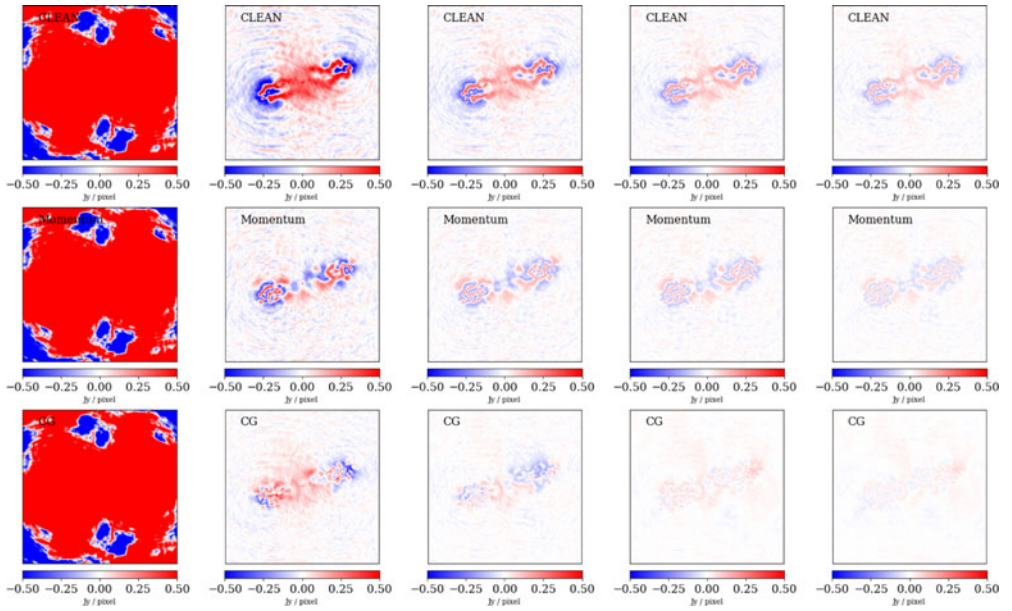


Figure 2. Residual as a function of major loop iteration for CLEAN and two variants that couple CLEAN with optimization techniques (CG-CLEAN and Momentum-CLEAN) for Cygnus A. Image adapted from Müller et al. (2025).

autocorrelation, reducing iteration counts and improving convergence by up to an order of magnitude while remaining CASA-compatible. We show the residual of Autocorr-CLEAN and Asp-CLEAN as a function of number of minor loop iterations in Fig. 3, demonstrating the significant improvements of latter two over traditional CLEAN. Further medium-term developments may include subgridding techniques, which selectively only grid visibilities based on component scale to achieve additional computational savings.

3. Conclusions and Future AI Outlook

The question of whether *to CLEAN or not to CLEAN* is ultimately one of practicality. CLEAN is not obsolete, but its limitations become clearer as we push toward higher sensitivity, resolution, and data volume. Incremental improvements and hybrid extensions, coupled with modern computational techniques, provide a bridge until more advanced paradigms are ready for large-scale deployment.

In the near term, hybrid frameworks, such as Autocorr-CLEAN or convex optimization-based variants, offer faster convergence and higher dynamic range while retaining CLEAN's reliability and simplicity. Looking further ahead, Bayesian and AI-driven methods are poised to redefine interferometric imaging, offering flexible priors, principled uncertainty quantification, and improved fidelity.

AI is expected to play a key role throughout this transition. Beyond plug-and-play reconstructions that exploit learned low-dimensional priors, AI can already enhance traditional workflows, for example, by learning policies to trigger calibration and imaging tasks (Kirk et al. 2025) or by transforming Hogbom CLEAN components into multiscale representations (Fiedler et al. 2025). We demonstrate this approach in Fig. 4, where a network trained on single-scale and MS-CLEAN representations transforms single-scale images into multiscale ones. This overcomes the main limitation of the former and achieves the accuracy of the latter with minimal computational overhead and bias from the training set (Fiedler et al. 2025).

https://doi.org/10.1017/S1743921325104468 Published online by Cambridge University Press

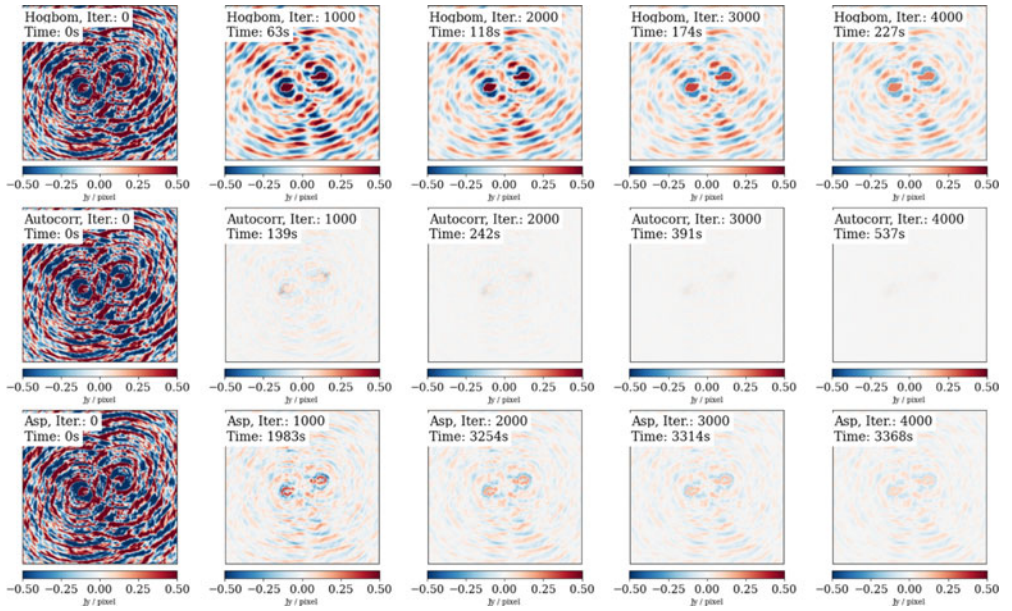


Figure 3. Residual as a function of minor loop iteration and time for CLEAN (upper row), Autocorr-CLEAN (middle row) and Asp-CLEAN (bottom row) for Cygnus A. Image adapted from Müller & Bhatnagar (2025).

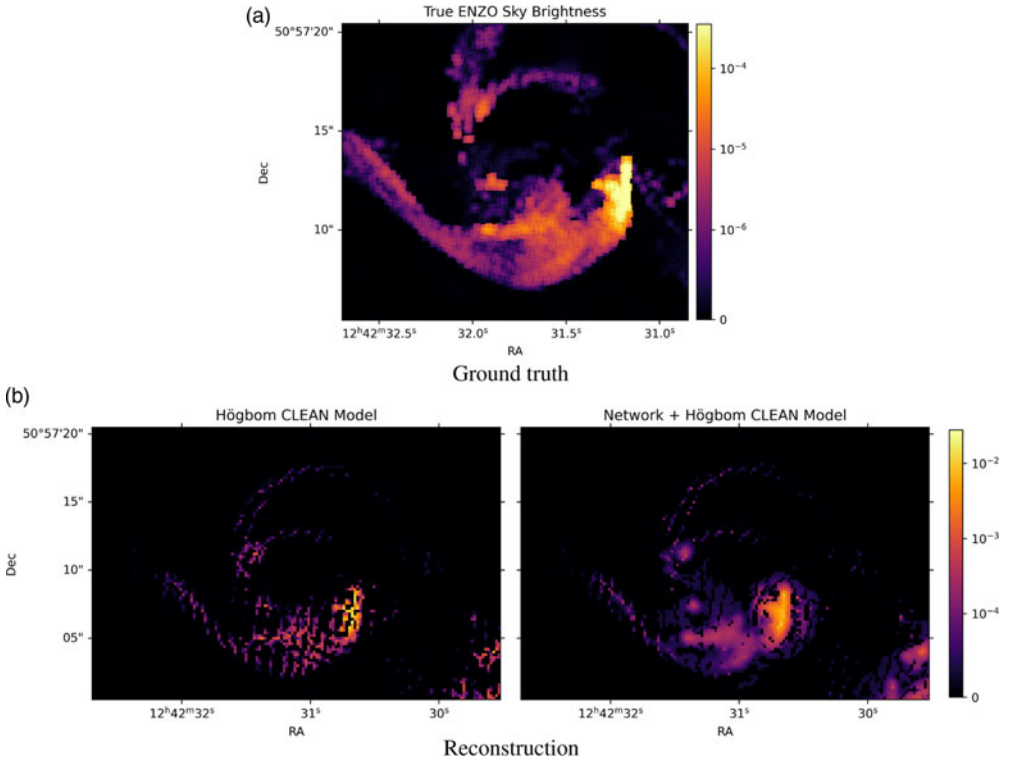


Figure 4. Reconstruction with traditional CLEAN and with a network transforming CLEAN components into a multiscale representation. Images are adapted from Fiedler et al. (2025).

https://doi.org/10.1017/S1743921325104468 Published online by Cambridge University Press

The future of radio interferometry will be shaped by our ability to balance accuracy, scalability, and practicality, ensuring imaging methods evolve in step with the demands of next-generation telescopes.

References

Aghabiglou, A., Chu, C. S., Dabbech, A., & Wiaux, Y. 2024, ApJS, 273, 3
Akiyama, K., Kuramochi, K., Ikeda, S., et al. 2017, ApJ, 838, 1
Arras, P., Bester, H. L., Perley, R. A., et al. 2021a, A&A, 646, A84
Arras, P., Frank, P., Haim, P., et al. 2022, Nature Astronomy, 6, 259
Arras, P., Frank, P., Leike, R., Westermann, R., & Enßlin, T. A. 2019, A&A, 627, A134
Arras, P., Reinecke, M., Westermann, R., & Enßlin, T. A. 2021b, A&A, 646, A58
Bester, L., Perkins, S., Kenyon, J., & Smirnov, O. 2024, in 32nd General Assembly International Union (IAUGA 2024), 2280
Bhatnagar, S. & Cornwell, T. J. 2004, A&A, 426, 747
Bhatnagar, S., Hiriart, R., & Pokorny, M. 2021, Size-of-Computing Estimates for ngVLA Synthesis Imaging, NRAO-NGVLA Computing Memo 4
Bhatnagar, S., Rau, U., Hsieh, M., Kern, J., & Xue, R. 2025, arXiv e-prints, arXiv:2508.17187
Bouman, K. L., Johnson, M. D., Dalca, A. V., et al. 2017, arXiv e-prints, arXiv:1711.01357
Briggs, D. S. 1995, PhD thesis, New Mexico Institute of Mining and Technology
Broderick, A. E., Pesce, D. W., Tiede, P., Pu, H.-Y., & Gold, R. 2020, ApJ, 898, 9
Brunet, T., Tolley, E., Corda, S., et al. 2024, Astronomy and Computing, 47, 100796
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2012, mnras, 426, 1223
Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2014, MNRAS, 439, 3591
Chael, A. A., Johnson, M. D., Bouman, K. L., et al. 2018, ApJ, 857, 23
Chael, A. A., Johnson, M. D., Narayan, R., et al. 2016, ApJ, 829, 11
Connor, L., Bouman, K. L., Ravi, V., & Hallinan, G. 2022, MNRAS, 514, 2614
Cornwell, T. J. 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 793
Cornwell, T. J. & Evans, K. F. 1985, A&A, 143, 77
Dabbech, A., Aghabiglou, A., Chu, C. S., & Wiaux, Y. 2024, ApJL, 966, L34
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
Event Horizon Telescope Collaboration. 2019, ApJl, 875, L4
Fiedler, K., Müller, H., & Dong, D. 2025, Accelerating Radio Interferometric Imaging: A Neural Network Approximation of Multiscale CLEAN, in prep.
Frank, P., Leike, R., & Enßlin, T. A. 2021, Entropy, 23, 853
Högbom, J. A. 1974, A&As, 15, 417
Homan, D. C., Roth, J. S., & Pushkarev, A. B. 2024, AJ, 167, 11
Hsieh, G., Bhatnagar, S.and Hiriart, R., & Pokorny, M. 2022a, Recommended Developments Necessary for Applying (W)Asp Deconvolution Algorithms to ngVLA, NRAO-ARDG Memo Series
Hsieh, G. & Bhatnagar, S. 2021, Efficient Adaptive-Scale CLEAN Deconvolution in CASA for Radio Interferometric Images, NRAO-ARDG Memo Series
Hsieh, G., Rau, U., & Bhatnagar. 2022b, An Adaptive-Scale Multi-Frequency Deconvolution of Interferometric Images, NRAO-ARDG Memo Series
Jarret, A., Kashani, S., Rué-Queralt, J., et al. 2025, A&A, 693, A225
Junklewitz, H., Bell, M. R., Selig, M., & Enßlin, T. A. 2016, A&A, 586, A76
Kepley, A. et al. 2024, Estimates of ALMA WSU Data Properties, ALMA Memo 626 626, ALMA Observatory, based on ALMA Cycle 7 & 8 data
Kepley, A. A., Tsutsumi, T., Brogan, C. L., et al. 2020, PASP, 132, 024505
Kim, J.-S., Nikonov, A. S., Roth, J., et al. 2024, A&A, 690, A129
Kirk, B. M., Rau, U., & Ramyaa, R. 2025, AJ, 169, 43
Knollmüller, J. & Enßlin, T. A. 2019, arXiv e-prints, arXiv:1901.11033
Müller, H. 2024, A&A, 689, A299
Müller, H. & Bhatnagar, S. 2025, A&A, 698, A176

https://doi.org/10.1017/S1743921325104468 Published online by Cambridge University Press

Müller, H., Hsieh, G., & Bhatnagar, S. 2025, Accelerating CLEAN with convex optimization, submitted to A&A
Müller, H. & Lobanov, A. P. 2022, A&A, 666, A137
Müller, H. & Lobanov, A. P. 2023a, A&A, 673, A151
Müller, H. & Lobanov, A. P. 2023b, A&A, 672, A26
Müller, H., Mus, A., & Lobanov, A. 2023, A&A, 675, A60
Mus, A. & Martí-Vidal, I. 2024, MNRAS, 528, 5537
Mus, A., Müller, H., & Lobanov, A. 2024a, A&A, 688, A100
Mus, A., Müller, H., Martí-Vidal, I., & Lobanov, A. 2024b, A&A, 684, A55
Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606
Offringa, A. R. & Smirnov, O. 2017, MNRAS, 471, 301
Rau, U. & Cornwell, T. J. 2011, A&A, 532, A71
Roth, J., Arras, P., Reinecke, M., et al. 2023, A&A, 678, A177
Roth, J., Frank, P., Bester, H. L., et al. 2024, A&A, 690, A387
Starck, J.-L., Candes, E. J., & Donoho, D. L. 2002, IEEE Transactions on Image Processing, 11, 670
Starck, J. L., Elad, M., & Donoho, D. L. 2005, IEEE Transactions on Image Processing, 14, 1570
Sun, H., Bouman, K. L., Tiede, P., et al. 2022, ApJ, 932, 99
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, 3rd Edition
Tiede, P. 2022, The Journal of Open Source Software, 7, 4457
Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vandergheynst, P. 2009, mnras, 395, 1733

https://doi.org/10.1017/S1743921325104468 Published online by Cambridge University Press