



**HAL**  
open science

# Multi-step reconstruction of radio-interferometric images

S Wang, S Prunet, S Mignot, A Ferrari

► **To cite this version:**

S Wang, S Prunet, S Mignot, A Ferrari. Multi-step reconstruction of radio-interferometric images. Astronomy and Astrophysics - A&A, 2024, 10.1051/0004-6361/202449921 . hal-04811037

**HAL Id: hal-04811037**

**<https://hal.science/hal-04811037v1>**

Submitted on 29 Nov 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Multi-step reconstruction of radio-interferometric images

S. Wang, S. Prunet, S. Mignot, and A. Ferrari

Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, 06000 Nice, France e-mail: sunrise.wang@oca.eu

October 27, 2024

## ABSTRACT

*Context.* The large aperture arrays for the currently under construction SKA Observatory (SKAO) will allow for observations of the universe in the radio spectrum at unprecedented resolution and sensitivity. However, these telescopes will produce data on the scale of exabytes, introducing a slew of hardware and software design challenges.

*Aims.* This paper proposes a multi-step image reconstruction framework that allows for partitioning of visibility data by baseline length. This enables more flexible data distribution and parallelization, aiding in processing radio-astronomical observations within given constraints. Additionally, as each step of the framework only relies on a subset of the total visibilities, one can perform reconstruction progressively, with the initial step performed on the SKAO Science Data Processors and the second on local clusters.

*Methods.* The multi-step reconstruction is separated into two steps. First a low-resolution image is reconstructed with only short-baseline visibilities, and then using this image together with the long-baseline visibilities, the full-resolution image is reconstructed. The proposed method only operates in the minor cycle, and it can be easily integrated into existing imaging pipelines.

*Results.* We show that our proposed method allows for partitioning of visibilities by baseline without introducing significant additional drawbacks, reconstructing images of similar quality within similar numbers of major cycles compared to a single-step all-baselines approach that uses the same reconstruction method as well as compared to multi-scale CLEAN.

**Key words.** Techniques: interferometric - Techniques: image processing - Methods: numerical - Methods: data analysis

## 1. Introduction

Radio-interferometry allows one to obtain images of the sky in the radio spectrum by using antenna arrays in tandem with aperture synthesis. The upcoming SKA Observatory (SKAO)<sup>1</sup> will be composed of two separate telescopes, one for high frequencies (350 MHz to 15.35 GHz) and one for low frequencies (50–350 MHz), that are currently being built separately in South Africa and Australia. Upon completion, they will provide unprecedented resolution and sensitivity, enabled by the 197 dishes in SKA-Mid in South Africa and the 512 antenna stations in SKA-Low in Australia.

With the large number of antennas comes an equally large amount of data. For SKA-Mid, projections estimate up to 2.375 TB/s = 205.2 PB/day from the dishes to the beamformer and correlator engines and 1.125 TB/s = 97.2 TB/day from these to the imaging super computer, the SKA-Mid Science Data Processor (Swart et al. 2022). For SKA-Low, the estimated data transfer is 0.725 TB/s ≈ 62.5 PB/day from the antennas to the correlator and 0.29 TB/s ≈ 25 PB/day from the correlator to the SKA-Low Science Data Processor (Labate et al. 2022). Such amounts of data naturally lead to hardware and software design challenges. These include transferring data between nodes, which turns out to be very expensive both in terms of computational time and energy, and long-term data storage, which proves impossible given the cost. In addition, memory usage per Science Data Processor node is a potential concern.

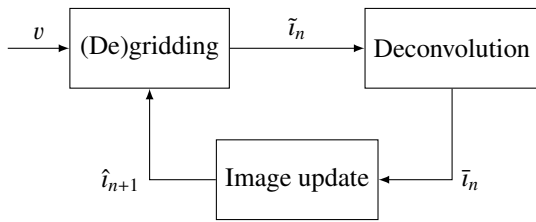
With these complications comes the need to efficiently partition both the data and the workload. This is typically performed along the frequency and time domains, with reconstruction being performed independently for each partition. Approaches

that separate the image into facets for direction-dependent calibration, such as Cornwell & Perley (1992); Van Weeren et al. (2016); Tasse et al. (2018), also enable partitioning by the spatial image domain. In addition to the above, one can also potentially separate the sample data (i.e., visibilities based on the length of their corresponding baselines). However, this is typically not done due to current minor-cycle reconstruction algorithms needing to process all baselines together to achieve full resolution. This limits parallelization flexibility. For example (de)gridding needs to wait for the reconstruction algorithm to finish processing all baselines before restarting, and memory access patterns become complex for large numbers of baselines due to the large number of visibilities needing to be gridded to the same grid.

This paper proposes an image reconstruction framework that allows for the partitioning of visibilities by baseline length. It achieves this by performing reconstruction in two steps, with each processing only a subset of the total visibilities. The first produces a low-resolution image using only the short baseline visibilities. The second produces the final reconstructed image using both the long baseline visibilities as well as the low-resolution image of the first step. This approach is advantageous over previous ones, as there is no need to gather the gridded visibilities each major cycle in order to perform (de)gridding, with the only communication between nodes being an image transfer between the two steps. This also allows for the possibility of progressive reconstruction in cases where the SKA Science Data Processors are not able to fully process the visibilities within the given time constraints and the work has to be offloaded to regional clusters.

We study the proposed framework in the context of a reconstruction method in the family of algorithms that are based on compressive sensing and convex optimization (Wiaux et al.

<sup>1</sup> <https://www.skao.int/>



**Fig. 1.** High-level overview of the radio-interferometric pipeline. A reconstructed image  $\hat{t}_n$  is compared to the measurements  $v$  in the (de)gridding step, which outputs the difference in the spatial domain  $\tilde{t}_n$ . This is passed to the reconstruction algorithm, which generates the next estimate  $\hat{t}_{n+1}$  by deconvolving the residual image  $\tilde{t}_n$  and adding it to  $\hat{t}_n$ .

2009; Carrillo et al. 2014; Garsden et al. 2015) adapted to the traditional major-minor cycle pipeline. We compare our framework to both a single-step approach that processes all baselines simultaneously using the same reconstruction method as well as multi-scale CLEAN (Cornwell 2008). We found it to reconstruct images of comparable quality in similar numbers of major cycles compared to the other two methods.

The remainder of this paper is structured as follows: We provide a brief overview of radio-interferometric imaging as well as a literature review in Sect. 2. We describe our method in Sect. 3, and Sect. 4 describes the datasets used for our experiments. We present and discuss our results in Sect. 5, and we provide conclusions and discuss possible avenues for future work in Sect. 6.

## 2. Radio-interferometric imaging

Radio-interferometers measure the sky using arrays of antennas (i.e., aperture arrays). Baselines produce visibilities, which are the correlated instrumental response of the electrical field for some given time duration and electro-magnetic frequency. These can be defined using the following equation (Smirnov 2011):

$$V(u, v) = C_{uv} \iint D_{uw}(l, m) \frac{I(l, m)}{n} e^{-2\pi i(ul+vm+w(n-1))} dl dm \quad (1)$$

$$n = \sqrt{1 - l^2 - m^2},$$

where  $C$  represents the direction-independent effects, such as antenna gain;  $D$  denotes the direction-dependent effects, such as phase gradients caused by the Earth's ionosphere;  $(u, v, w)$  is the difference between antenna coordinates in the frame, where  $w$  is aligned with the phase center;  $(l, m)$  are the spatial angular coordinates on the celestial sphere, which is also the domain of the integral; and  $I$  is the incident radiance of the sky emission.

If the  $D$  and  $e^{-2\pi i w(n-1)}$  terms are ignored, Equation 1 simplifies into a two-dimensional Fourier transform and allows one to retrieve an image of the sky emission through its inversion. This image contains artifacts caused both by the partial sampling of the Fourier domain due to the geometry of antenna arrays and by the omission of the  $w$  and  $D$  terms. Radio-interferometric imaging aims to correct for these, reconstructing an image that is usable for science.

The general approach used by radio-interferometric algorithms is iterative and is illustrated in Fig. 1. The current image estimate,  $\hat{t}_n$ , is evaluated against the measurements,  $v$ , in the (de)gridding step, which also corrects for the  $w$  and sometimes the  $D$  terms. This step produces the difference between  $v$  and  $\hat{t}_n$  in the image space, termed  $\tilde{t}_n$ . It can be expressed as

$$\tilde{t}_n = F^\dagger G^\dagger (v - GF\hat{t}_n), \quad (2)$$

where  $F$  and  $F^\dagger$  respectively are the fast Fourier transform (FFT) and its inverse,  $G$  is a degridding operator responsible for extrapolating the regular gridded visibilities to their original irregular positions, and  $G^\dagger$  is the gridding operator responsible for interpolating the irregular visibilities to regular gridded positions.

The correction of the  $w$  and  $D$  terms typically occurs before  $F^\dagger$ . There are various approaches to this, such as treating the  $w$  and  $D$  terms as convolution kernels in the Fourier domain and convolving them with the visibilities during  $G^\dagger$  (Cornwell et al. 2008; Bhatnagar et al. 2008; Van Der Tol et al. 2018); discretizing in the  $w$ -plane, which can be seen in methods such as w-stacking (Offringa et al. 2014) and its improvement (Ye et al. 2022); discretizing in the time domain, such as with the snapshots method (Ord et al. 2010); discretizing in the spatial domain with facet-based approaches (Cornwell & Perley 1992; Tasse et al. 2018); or some hybrid of the above, such as between  $w$ -projection and snapshots (Cornwell et al. 2012) and  $w$ -projection and  $w$ -stacking (Pratley et al. 2019a).

The residual image  $\tilde{t}_n$  is then sent to the deconvolution algorithm, which removes the partial sampling artifacts. There are a plethora of methods that aim to achieve this, with some examples being CLEAN and its variants (Högbohm 1974; Cornwell 2008; Rau & Cornwell 2011) that deconvolve  $\tilde{t}_n$  in a greedy non-linear manner, much akin to matching pursuit (Mallat & Z. 1993) and convex optimization methods based on compressive sensing (Wiaux et al. 2009; Carrillo et al. 2014; Dabbech et al. 2015). There has also been work done on progressively reconstructing the final image based on subsets of visibilities, such as the work of Cai et al. (2019), which shares similar general ideas to our approach. It differs in that the partitioning is by time rather than baseline, with their method still requiring all baselines to be present within a subset of visibilities. Another method that has similarities with ours is that of Lauga et al. (2024), where they alternate reconstruction between low and full-resolution iterations. Their work differs from ours in that their framework is primarily aimed at accelerating reconstruction, and it does not allow for easy partitioning of visibilities.

After deconvolution, the image estimate is updated by adding to it the deconvolved residual  $\tilde{t}_n$ :

$$\hat{t}_{n+1} = \hat{t}_n + \tilde{t}_n. \quad (3)$$

The deconvolution step is iterative, ergo the imaging pipeline has a nested loop structure, which is often referred to as the major-minor loop structure, with the loop shown in Fig. 1 being the major loop and the deconvolution being the minor.

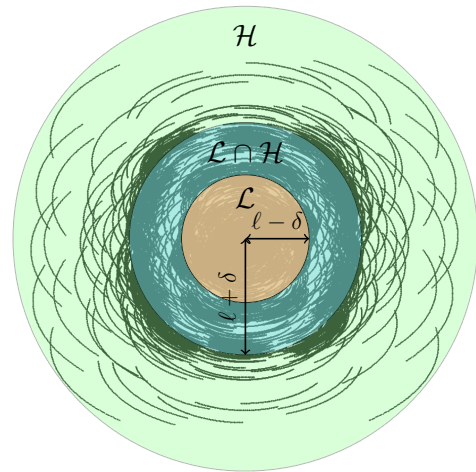
(De)gridding is often the bottleneck of the imaging pipeline due to the sheer number of visibilities needed to be processed, with the work of Tasse et al. (2018) demonstrating that it can reach up to 94% of the total processing time for a fully serial implementation. Due to this fact, there is much work that looks to expedite this stage through parallelization.

This can be done in a coarse- or fine-grained manner. Fine-grained approaches aim to parallelize on the local machine at a fine scale, such as per visibility or grid cell. There has been substantial amounts of work done in this area, both for the central processing unit (CPU), such as in Barnett et al. (2019), as well as the graphics processing unit (GPU), such as in Romein (2012); Merry (2016); Veenboer et al. (2017). Our method does not deal with parallelization on this scale and instead focuses primarily on facilitating distribution and parallelization on coarser scales.

Coarse scale parallelization looks to distribute (de)gridding across multiple nodes within a cluster. A typical method for this is to distribute according to frequency channels and time. More recently, there has also been work that looks to distribute

165 the gridding according to facets (Monnier et al. 2022) as well as sections of the UV plane (Onose et al. 2016; Pratley et al. 2019b). These different distribution strategies can also be combined, such as in the work of Gheller et al. (2023) where visibilities are separated by time and the v-axis.

170 Of the aforementioned coarse-scale methods, the most similar to our work are those by Onose et al. (2016) and Pratley et al. (2019b), both of which distribute and parallelize by regions of the UV grid, or also by baseline length in the case of Pratley et al. (2019b). Although similar in goal, their work differs from ours in that our presented method is not parallel, with our main focus being on investigating partitioning visibilities in the context of radio-interferometric imaging. Furthermore, although our framework is theoretically general, we present it in the context of a major-minor loop reconstruction method, solving for a deconvolution every major loop. This is contrary to both of these works, which aim to directly solve for the measurement equation using primal-dual methods. This has efficiency implications, as while their methods need to perform (de)gridding at every iteration of the optimization algorithm, our approach only needs to do so for each major cycle, which occurs far less often.



**Fig. 2.** Full set of visibilities partitioned according to the domains. The domain  $\mathcal{L}$  (orange) contains the short baseline visibilities, and  $\mathcal{H}$  (green) contains the long baseline visibilities. These are not mutually exclusive but rather have an overlap region indicated in cyan and defined using  $\ell$ , which is the middle of  $\mathcal{L} \cap \mathcal{H}$ , and  $\delta$ , the half-width of  $\mathcal{L} \cap \mathcal{H}$ .

### 3. Proposed algorithm

Our proposed framework performs imaging in two steps, each operating only on a subset of the visibility data separated in the UV plane. These subsets, termed  $V_{\mathcal{L}}$  and  $V_{\mathcal{H}}$ , are partitioned according to the domains  $\mathcal{L}$  and  $\mathcal{H}$ , which cover regions representing short and long baselines, respectively.

Each step performs the entire imaging pipeline and produces an image. The first produces a low-frequency image,  $\hat{i}_{\mathcal{L}}$ , from  $V_{\mathcal{L}}$ , and the second produces the fully reconstructed image,  $\hat{i}$ , using both  $V_{\mathcal{H}}$  and  $\hat{i}_{\mathcal{L}}$ .

The reconstruction method we apply to our framework is based on convex optimization with sparsity regularization, as this has a solid theoretical foundation (Candes et al. 2006; Donoho 2006; Candès et al. 2011) and has been extensively studied in the field of radio-interferometric imaging (Wiaux et al. 2009; Carrillo et al. 2014; Dabbech et al. 2015). Specifically, similar to previous works on reconstruction within the Low-Frequency Array (LOFAR) pipeline (Jiang et al. 2015; Garsden et al. 2015), we solve for a sparse series of atoms  $\alpha$  of a redundant wavelet dictionary  $W$ .

We adapted the method to operate with the traditional major-minor loop framework for efficiency, which we achieved by making the same assumption as algorithms such as CLEAN, namely that  $F^\dagger G^\dagger G F i \approx H i$ , where  $H$  denotes the convolution by the point spread function (PSF). This allows the minor cycles to only solve the deconvolution problem, with the full measurement operator (i.e., degridding and gridding) only needing to be performed every major cycle. As with CLEAN, this assumes that the errors introduced by the approximation will be detected and corrected for in future major-cycle iterations. It is important to note that although detailed for this specific family of reconstruction algorithms, our proposed framework can be adapted to other reconstruction frameworks.

#### 3.1. Partitioning visibilities

220 We partitioned the visibilities into subsets  $V_{\mathcal{L}}$  and  $V_{\mathcal{H}}$  based on whether the visibilities fall under the domain  $\mathcal{L}$  or  $\mathcal{H}$ , respectively. Figure 2 illustrates this, where visibilities in the orange region are part of  $V_{\mathcal{L}}$ , and visibilities in the green belong to  $V_{\mathcal{H}}$ .

It can be seen that  $\mathcal{L} \cap \mathcal{H} \neq \emptyset$ , and this overlap is denoted by the cyan region. This is to alleviate cases where gridding introduces additional spatial frequencies not in the respective domain, which is caused by the visibilities being interpolated onto the grid through convolution using a kernel with a non-zero support size. The visibilities in  $\mathcal{L} \cap \mathcal{H}$  are weighted so that although duplicated, their contribution adds up to one. This was done using filters that are discussed in Sect. 3.3.

The dataset was partitioned using the variables  $\ell$  and  $\delta$ , where  $\delta$  defines the half-width size of  $\mathcal{L} \cap \mathcal{H}$  and  $\ell$  defines the center of  $\mathcal{L} \cap \mathcal{H}$ . These are also shown in Fig. 2. These values are typically defined in units wavelength. However, we opted to use pixel distances in our paper.

#### 3.2. Low-resolution reconstruction

The first step of the framework involves reconstructing a low-resolution image from  $V_{\mathcal{L}}$ . To do this, for every major-cycle iteration,  $n$ , for a total of  $N$  cycles, we solve for the unconstrained problem:

$$\begin{aligned} \alpha_{\mathcal{L}_n} &= \arg \min_{\alpha} \|\tilde{i}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha\|_2^2 + \lambda_{\mathcal{L}_n} \|\alpha\|_1 \\ \tilde{i}_{\mathcal{L}_n} &= W \alpha_{\mathcal{L}_n}, \end{aligned} \quad (4)$$

where  $\tilde{i}_{\mathcal{L}_n}$  is the current residual image (i.e., between  $V_{\mathcal{L}}$  and  $\hat{i}_{\mathcal{L}_{n-1}}$ , computed in the  $n$ th major cycle),  $H_{\mathcal{L}}$  is the operator detailing the convolution by the PSF associated with  $V_{\mathcal{L}}$ ,  $\lambda_{\mathcal{L}_n}$  is the regularization parameter associated with the current major cycle, and  $\tilde{i}_{\mathcal{L}_n}$  is the deconvolved residual of  $\tilde{i}_{\mathcal{L}_n}$ . The final image after  $N$  major cycles is then  $\hat{i}_{\mathcal{L}} = \tilde{i}_{\mathcal{L}_{N+1}} + \sum_{n=1}^N \tilde{i}_{\mathcal{L}_n}$ .

An undesirable side effect when solving for Equation 4 is that  $\hat{i}_{\mathcal{L}}$  can contain frequency information that is not associated with  $\mathcal{L}$  due to  $W$  not being constrained to  $\mathcal{L}$ , which interferes with the reconstruction of the full-resolution image. We handled this in the full-resolution reconstruction step using filtering, which we discuss in Sect. 3.3.

### 3.3. Full-resolution reconstruction

255 The second step of the framework reconstructs the full-resolution image using both  $V_{\mathcal{H}}$  and  $\hat{l}_{\mathcal{L}}$ . We formulated the reconstruction problem using two data fidelity terms one for the high frequencies and one for the low frequencies, in addition to an  $L^1$  regularization term. For every major-cycle iteration,  $n$ ,  
260 over  $N$  total cycles, we solve for

$$\begin{aligned} \alpha_{\mathcal{H}_n} &= \arg \min_{\alpha} \|G_{\mathcal{H}}(\tilde{i}_{\mathcal{H}_n} - H_{\mathcal{H}}W\alpha)\|_2^2 + \|G_{\mathcal{L}}(l_{\mathcal{L}_n} - W\alpha)\|_2^2 \\ &\quad + \lambda_{\mathcal{H}_n} \|\alpha\|_1 \\ l_{\mathcal{L}_n} &= \hat{l}_{\mathcal{L}} - \sum_{j=1}^{n-1} \bar{i}_j, \quad \bar{i}_n = W\alpha_{\mathcal{H}_n}, \end{aligned} \quad (5)$$

where  $\tilde{i}_{\mathcal{H}_n}$  is the current residual image between  $V_{\mathcal{H}}$  and  $\hat{i}_{n-1}$ ,  $H_{\mathcal{H}}$  the operator detailing convolution by the PSF associated to  $V_{\mathcal{H}}$ , and  $\lambda_{\mathcal{H}_n}$  is the regularization parameter for the  $n$ th major cycle of the full-resolution step. Operator  $G_{\mathcal{L}}$  denotes a low-pass linear filter that discards frequencies not in  $\mathcal{L}$ , and  $G_{\mathcal{H}}$  is a high-pass linear filter that discards frequencies not in  $\mathcal{H}$ . Finally,  $\bar{i}_n$  is the deconvolved residual image using both  $\tilde{i}_{\mathcal{H}_n}$  and  $\hat{l}_{\mathcal{L}}$ . The final reconstructed image is computed similarly to the low-resolution step and is  $\hat{i} = \tilde{i}_{\mathcal{H}_{N+1}} + \sum_{n=1}^N \bar{i}_n$ .

270 In Equation 5, the first data fidelity term aims to reconstruct the high-frequency components of the residual image from  $V_{\mathcal{H}}$  (i.e.,  $\tilde{i}_{\mathcal{H}_n}$  for the  $n$ th major loop). The second data fidelity term guarantees that the low-frequency components of the reconstructed image match the reconstruction obtained using  $V_{\mathcal{L}}$  in the first low-resolution reconstruction step. For the  $n$ th major loop, the low-pass filtered reconstructed image must match the residual between  $\hat{l}_{\mathcal{L}}$  and the low-frequency components of the reconstructions at the previous major loop iterations:  $G_{\mathcal{L}} \sum_{j=1}^{n-1} \bar{i}_j$ . As mentioned in the previous section,  $\hat{l}_{\mathcal{L}}$  may contain spatial frequencies not in  $\mathcal{L}$  due to  $W$  not being limited to  $\mathcal{L}$ , which may bias the reconstruction of  $\hat{i}$  toward these frequencies. These components are removed by applying  $G_{\mathcal{L}}$  also to  $\hat{l}_{\mathcal{L}}$ , as in Equation 5.

285 The gains of the two filters have to be properly normalized, in particular in  $\mathcal{L} \cap \mathcal{H}$ , to account for visibilities being duplicated. We assumed that the two filters  $G_{\mathcal{L}}$  and  $G_{\mathcal{H}}$  have circularly symmetric frequency responses denoted as  $g_{\mathcal{L}}(r)$  and  $g_{\mathcal{H}}(r)$ . We considered that  $l_{\mathcal{L}_n}$  contains some reconstruction noise with variance  $\eta^2$  and that  $\tilde{i}_{\mathcal{H}_n}$  has noise with variance  $\sigma^2$ . This suggests defining the two filters as

$$r > \ell + \delta : |g_{\mathcal{H}}(r)|^2 = 1/\sigma^2, \quad g_{\mathcal{L}}(r) = 0 \quad (6)$$

$$r < \ell - \delta : g_{\mathcal{H}}(r) = 0, \quad |g_{\mathcal{L}}(r)|^2 = 1/\eta^2 \quad (7)$$

$$\ell - \delta < r < \ell + \delta : \sigma^2 |g_{\mathcal{H}}(r)|^2 + \eta^2 |g_{\mathcal{L}}(r)|^2 = 1. \quad (8)$$

These constraints ensure normalization of the variance noise across the three frequency domains  $\mathcal{L} - \mathcal{H}$ ,  $\mathcal{H} - \mathcal{L}$ , and  $\mathcal{H} \cap \mathcal{L}$ .

We propose defining the frequency response of the filters for  $\ell - \delta < r < \ell + \delta$  as

$$g_{\mathcal{L}}(r) = \alpha(r) \left( 1 - \sin\left(\frac{\pi}{2\delta}(r - \ell)\right) \right) \quad (9)$$

$$g_{\mathcal{H}}(r) = \alpha(r) \left( 1 + \sin\left(\frac{\pi}{2\delta}(r - \ell)\right) \right), \quad (10)$$

295 where  $\alpha(r)$  is such that  $\eta^2 g_{\mathcal{L}}(r)^2 + \sigma^2 g_{\mathcal{H}}(r)^2 = 1$ . An example is given in Fig. 3.

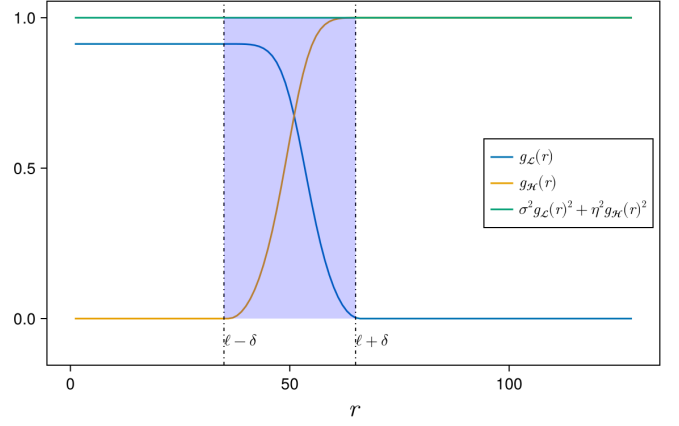


Fig. 3. Radial frequency response of  $G_{\mathcal{H}}$  and  $G_{\mathcal{L}}$  for  $\sigma^2 = 1$  and  $\eta^2 = 1.2$ .

### 3.4. Optimization algorithm

We opted to use the fast iterative shrinkage-thresholding algorithm (FISTA) (Beck & Teboulle 2009) for our optimization algorithm. FISTA is a fast converging algorithm aimed at optimizing convex problems that comprise both a smooth term and a term that has an easy to solve proximal operator. Equations 4 and 5 both fall under this umbrella.

---

**Algorithm 1:** FISTA for  $\alpha \rightarrow f(\alpha) + \gamma \|\alpha\|_1$

---

- 1: Initialize  $\beta_p, \alpha \leftarrow \beta_p$
  - 2: **for**  $k = 1 \dots N - 1$  **do**
  - 3:    $\beta \leftarrow \tau_{\gamma}(\alpha - \mu \nabla f(\alpha))$
  - 4:    $\alpha \leftarrow \beta + \frac{k-1}{k+2}(\beta - \beta_p)$
  - 5:    $\beta_p \leftarrow \beta$
  - 6: **end for**
  - 7: **return**  $\alpha$
- 

FISTA operates iteratively and involves computing the gradient of the smooth portion and the proximal operator of the non-smooth portion, which for the  $L^1$ -norm is the soft-thresholding operator. It then computes the candidate solution for the next iteration using a gradient step size,  $\mu$ ; a soft-thresholding step size; and a momentum term to allow for faster convergence. We describe FISTA in Algorithm 1.

FISTA requires the gradients of the objective function for each iteration. These are, for the  $n$ th major cycle,

$$\nabla f(\alpha)_{\mathcal{L}_n} = 2W^\dagger H_{\mathcal{L}}^\dagger (H_{\mathcal{L}}W\alpha - \tilde{i}_{\mathcal{L}_n}) \quad (11)$$

$$\begin{aligned} \nabla f(\alpha)_{\mathcal{H}_n} &= 2W^\dagger \left( H_{\mathcal{H}}^\dagger G_{\mathcal{H}}^\dagger G_{\mathcal{H}}(H_{\mathcal{H}}W\alpha - \tilde{i}_{\mathcal{H}_n}) \right. \\ &\quad \left. + G_{\mathcal{L}}^\dagger G_{\mathcal{L}}(W\alpha - l_{\mathcal{L}_n}) \right) \end{aligned} \quad (12)$$

for Equations 4 and 5 respectively. The gradient step size is set as  $\mu = \frac{1}{\vartheta}$ , where  $\vartheta$  is the Lipschitz constant of  $\nabla f(\alpha)$ , defined for our problem as

$$\vartheta_{\mathcal{L}} = 2\lambda_{\max} \left( W^\dagger H_{\mathcal{L}}^\dagger H_{\mathcal{L}} W \right) \quad (13)$$

$$\vartheta_{\mathcal{H}} = 2\lambda_{\max} \left( W^\dagger (H_{\mathcal{H}}^\dagger G_{\mathcal{H}}^\dagger G_{\mathcal{H}} H_{\mathcal{H}} + G_{\mathcal{L}}^\dagger G_{\mathcal{L}}) W \right) \quad (14)$$

for Equations 4 and 5 respectively, where  $\lambda_{\max}$  is the largest eigenvalue.

In the case where  $W$  consists of the concatenation of  $M$  orthogonal decompositions, as in (Onose et al. 2016), and the convolutions are circular, we obtain the following:

$$\vartheta_{\mathcal{L}} = 2M \max\{|\hat{H}_{\mathcal{L}}|^2\} \quad (15)$$

$$\vartheta_{\mathcal{H}} = 2M \max\{|\hat{G}_{\mathcal{H}} \odot \hat{H}_{\mathcal{H}}|^2 + |\hat{G}_{\mathcal{L}}|^2\}, \quad (16)$$

where  $\hat{A}$  denotes the discrete Fourier transform of the PSF associated to  $A$ . In other cases, such as in Starck et al. (2007),  $\lambda_{\max}$  can be obtained using an algorithm such as power iteration.

Finally, there is the question of how to approximate  $\sigma^2$  and  $\eta^2$  and how to select the regularization parameters  $\lambda_{\mathcal{L}}$  and  $\lambda_{\mathcal{H}}$ . The former is discussed in Sect. 4.3, whereas the latter is presented in Sect. A.2.

### 3.5. Implementation

We implemented our method in Julia (Bezanson et al. 2017) and integrated it into the Radio Astronomy Simulation, Calibration and Imaging Library (RASCIL) (Cornwell et al. 2020) framework under the clean algorithm name `mstep`. Our implementation and results as well as instructions on how to use our code are available in our repository,<sup>2</sup> with code specific to this paper available as a release.<sup>3</sup>

## 4. Simulated datasets

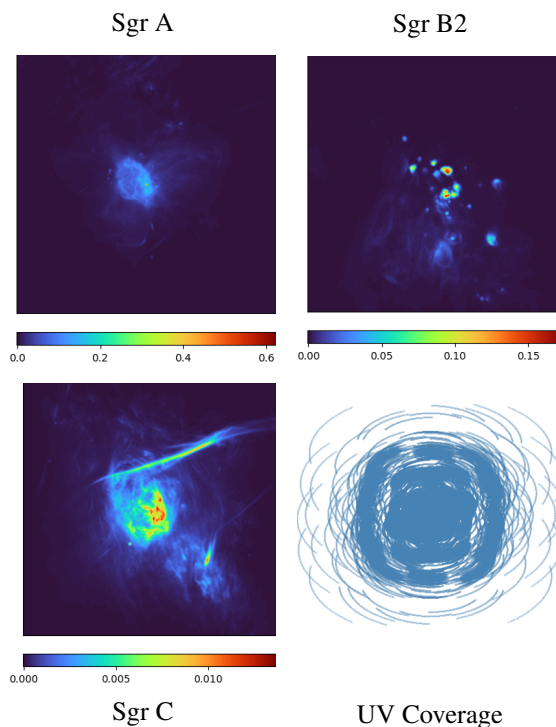
### 4.1. Full datasets

We used three different simulated datasets for our experiments. They were obtained from  $512 \times 512$  pixel tapered cutouts of the 1.28 GHz MeerKAT mosaic (Heywood et al. 2022) and consist of the regions surrounding Sgr A, B2, and C with pixel resolutions of  $1.1''$ . These can be seen along with their UV coverage in Fig. 4.

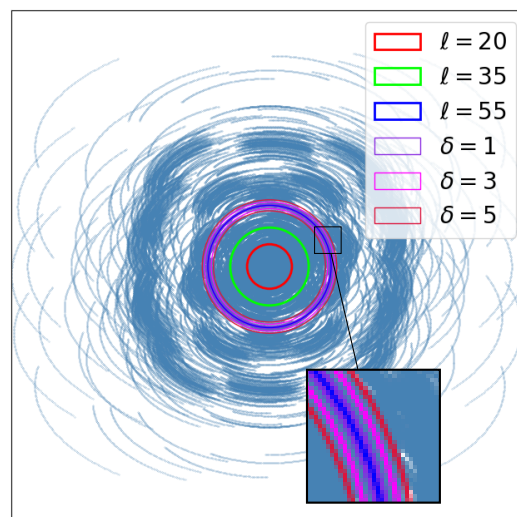
We used RASCIL to generate the visibilities. We first generated the visibility positions using a telescope configuration detailing the 64 MeerKAT dishes, assuming that the observation lasts 4 hours (-2 to 2 hour angles) and with visibilities sampled every 120 s, resulting in 249600 unique positions. We then performed an FFT on our ground-truth images to obtain their respective gridded visibilities, which we degridged to the generated irregular positions with uniform weighting using the improved w-stacking gridded (Ye et al. 2022). Finally, we modeled the noise received by antennas by perturbing the visibilities with noise sampled from  $\mathcal{N}(0, \sigma/50)$ , where  $\sigma$  is the standard deviation of the visibilities. A more thorough study of how noise affects our reconstruction method is outside the scope of this paper.

### 4.2. Partitioning configurations

We partitioned each initial dataset with three different centers of separation,  $\ell$ , and each had three different half-width sizes,  $\delta$ , resulting in nine different partitioning configurations for each. Figure 5 illustrates both  $\ell$  and  $\delta$  in the context of our datasets, whereas Table 1 details the number of visibilities in  $V_{\mathcal{L}}$  and  $V_{\mathcal{H}}$  for each configuration. It also details the number of duplicated visibilities, which lie in the region  $\mathcal{L} \cap \mathcal{H}$ . These numbers are



**Fig. 4.** Ground truth images of our three datasets obtained by cutting out and tapering sections of the 1.28 GHz Meerkat galactic center mosaic (Heywood et al. 2022). The images are  $512 \times 512$  pixels in size with a resolution of  $1.1''$  per pixel. The UV coverage for our simulated datasets is shown at the bottom right. The same coverage was used for all three datasets.



**Fig. 5.** Visualization of partition configuration parameters  $\delta$  and  $\ell$  used for our datasets. We only show  $\delta$  for  $\ell = 55$  in this figure to avoid clutter. The combinations of the three values of  $\delta$  and three values of  $\ell$  result in nine different partitioning configurations per dataset.

identical for all three of our datasets, as they have the same observational parameters. We show an example of how the partition configuration affects the dirty images for the Sgr A dataset in Fig. 6.

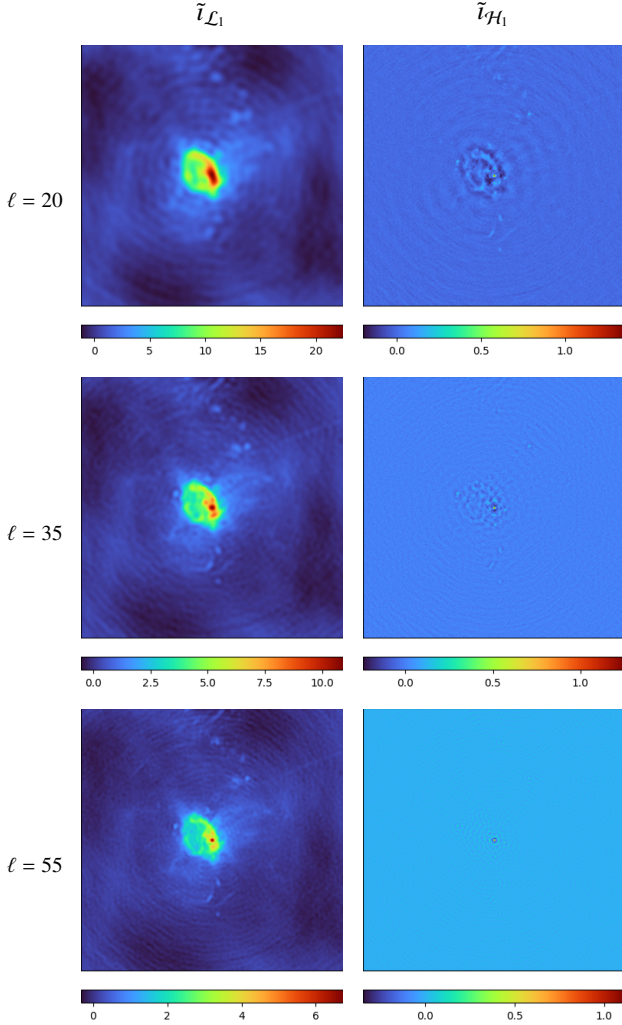
<sup>2</sup> <https://github.com/simon-prunet/radio-imaging/>

<sup>3</sup> <https://github.com/simon-prunet/radio-imaging/releases/tag/msteprelease>

**Table 1.** Visibilities per partition.

		$V_{\mathcal{L}}$	$V_{\mathcal{H}}$	$V_{\mathcal{L} \cap \mathcal{H}}$
$\ell = 20$	$\delta = 1$	106548	150641	7589
	$\delta = 3$	113791	160061	24252
	$\delta = 5$	118941	169477	38818
$\ell = 35$	$\delta = 1$	141653	111758	3811
	$\delta = 3$	146107	115792	12299
	$\delta = 5$	149271	120204	19875
$\ell = 55$	$\delta = 1$	167772	84034	2206
	$\delta = 3$	169771	85758	5929
	$\delta = 5$	171508	87509	9417

**Notes.** Numbers apply to all three datasets, as they have the same observational parameters.



**Fig. 6.** Dirty images  $\tilde{t}_{\mathcal{L}_1}$  and  $\tilde{t}_{\mathcal{H}_1}$  for different values of  $\ell$  for the Sgr A dataset. We set  $\delta = 5$  for these images.

### 4.3. Visibility and reconstruction variance

Our proposed method requires knowing the variances  $\sigma^2$  and  $\eta^2$  of  $\tilde{t}_{\mathcal{H}}$  and  $\hat{t}_{\mathcal{L}}$  respectively. Ensemble statistics are required to estimate them, and thus these statistics need to be approximated. To derive an appropriate strategy, we first computed estimations using ensemble populations of 50 for both  $V_{\mathcal{L}}$  and  $V_{\mathcal{H}}$ , each with independently sampled and identical noise properties to the original.

For  $\sigma^2$ , we used the average per-pixel variances of the different realizations of  $\tilde{t}_{\mathcal{H}_1}$  computed from the different realizations of  $V_{\mathcal{H}}$  using RASCIL. For  $\eta^2$ , we reconstructed  $\hat{t}_{\mathcal{L}}$  over three major cycles for each realization of  $V_{\mathcal{L}}$ , and then we set  $\eta^2$  to be equal to the average of the per-pixel variance among these reconstructions. For the reconstructions, we used the regularization parameter  $\lambda_{\mathcal{L}_n} = 0.05 \|\hat{t}_{\mathcal{L}_n}\|_2 \times 2^n$  for the  $n$ th major cycle as well as the concatenation of the first eight Daubechies wavelets for our dictionary. These choices are discussed in Sect. A.

**Table 2.** Estimated  $\sigma^2$  and  $\eta^2$  and their approximations

		$\sigma^2$	$\hat{\sigma}^2$	$\eta^2$	$\hat{\eta}^2$
Sgr A	$\ell = 20$	2.8e-04	2.2e-04	2e-08	2.2e-07
	$\ell = 35$	3e-04	2.3e-04	1e-07	2.3e-07
	$\ell = 55$	3.6e-04	2.5e-04	2.6e-07	2.5e-07
Sgr B2	$\ell = 20$	2e-05	2.2e-05	2e-08	2.2e-07
	$\ell = 35$	2.2e-05	6e-05	9.5e-09	6e-08
	$\ell = 55$	2.5e-05	3.8e-05	2e-08	3.8e-08
Sgr C	$\ell = 20$	4e-06	2.8e-06	3e-10	2.8e-09
	$\ell = 35$	4.2e-06	2.7e-06	1.6e-09	2.7e-09
	$\ell = 55$	5e-06	2.8e-06	3.2e-09	2.8e-09

We approximated  $\sigma^2$  with  $\hat{\sigma}^2$  by first estimating the variance of pixels in  $\tilde{t}_{\mathcal{H}_1}$  within a  $5 \times 5$  sliding window and then averaging the values. Approximating  $\eta^2$  is more complicated, as it is dependent on the details of the reconstruction algorithm in the low-resolution step. For example, an exceedingly large value of  $\lambda_{\mathcal{L}}$  would result in  $\eta^2 = 0$ , as all realizations of  $\hat{t}_{\mathcal{L}}$  will be zero. Rather than deriving a strategy for this, we applied a constant factor of  $\hat{\eta}^2 = \hat{\sigma}^2 10^{-3}$  for our experiments, as we observed this to be roughly the relationship between the two variables for many cases.

Table 2 shows  $\eta^2$  and  $\sigma^2$  for the different values of  $\ell$  as well as our estimated values  $\hat{\sigma}^2$  and  $\hat{\eta}^2$ . We set  $\delta = 5$  and did not vary it, as it does not significantly change the variance. It can be seen that  $\hat{\sigma}^2$  is relatively close to  $\sigma^2$ , diverging by at most 2.75 $\times$ , and in most cases it is much closer to the estimated value. Although our approximations for  $\eta^2$  are also close in some cases, there are others where they are off by more than a factor of ten, with the worst being for the dataset Sgr B2 with  $\ell = 35$ , where  $\hat{\eta}^2$  is almost 100 times larger. In practice, we found that these variations did not greatly change the quality of the final reconstruction. However, investigation into better approximation strategies may yield improved convergence speeds.

## 5. Simulation results

We evaluated our proposed multi-step method by first studying how the partitioning configuration affects the final image reconstruction, both in terms of quality and speed. We then compared our method against a method that reconstructs with all baselines in  $\mathcal{L} \cup \mathcal{H}$  in a single step, which afforded us information on how partitioning visibilities by baseline compares to processing all the visibilities simultaneously without any partitioning. We also performed a comparison to images reconstructed using RASCIL's multi-scale CLEAN implementation in order to evaluate how our method compares to current popular methods used in production.

For our algorithm parameters, we set  $\lambda_n$ ,  $\lambda_{\mathcal{L}_n}$ , and  $\lambda_{\mathcal{H}_n}$  for the  $n$ th major cycle to  $0.015 \|\tilde{t}_n\|_2 \times 2^n$ ,  $0.05 \|\tilde{t}_{\mathcal{L}_n}\|_2 \times 2^n$ , and  $0.05 \|\tilde{t}_{\mathcal{H}_n}\|_2 \times 2^n$ , respectively, and for  $W$ , we used a concatenation of the first eight Daubechies wavelets. We discuss our motivations for these in Sect. A. We used a static value of 100 FISTA

iterations as the stopping condition for both the single-step full-resolution method as well as for both steps in our multi-step method.

For multi-scale CLEAN, we used a maximum of 2000 minor iterations per major cycle, with a clean threshold of 0.001, a clean fractional threshold of 0.01, and CLEAN scales of [0, 1, 2, 4, 6, 10, 30]. Although these parameters are conservative, we found that they are necessary in reconstructing many of the smaller scale features, with larger CLEAN scales or higher thresholds leading to significantly poorer quality images.

Finally, we evaluated our reconstructions using peak signal-to-noise ratio (peak S/N), which we calculated as  $20\log\left(\frac{\|i\|_2}{\|\hat{i} - \tilde{i}\|_2}\right)$  to obtain the decibel (dB) value, where  $i$  is the ground truth and  $\hat{i}$  is the reconstruction without the final added residual. We opted to use this rather than the maximum signal, as it evaluates the reconstruction over the entire image while maintaining a greater importance for brighter sources.

### 5.1. Partition configuration effect on reconstruction accuracy

**Table 3.** Final reconstruction quality across partition configurations.

		Sgr A	Sgr B2	Sgr C
$\ell = 20$	$\delta = 1$	22.6 (0.99)	20.9 (0.92)	19 (0.97)
	$\delta = 3$	<b>22.9 (1)</b>	21.2 (0.93)	19.3 (0.98)
	$\delta = 5$	21.9 (0.96)	21.2 (0.93)	18.5 (0.94)
$\ell = 35$	$\delta = 1$	22.5 (0.98)	21.6 (0.95)	<b>19.6 (1)</b>
	$\delta = 3$	22.1 (0.97)	21.6 (0.95)	<b>19.6 (1)</b>
	$\delta = 5$	21.6 (0.94)	21.5 (0.94)	19.2 (0.98)
$\ell = 55$	$\delta = 1$	21.2 (0.93)	21 (0.92)	18.6 (0.95)
	$\delta = 3$	21.6 (0.94)	22.3 (0.98)	18.8 (0.96)
	$\delta = 5$	21.9 (0.96)	<b>22.8 (1)</b>	19.2 (0.98)

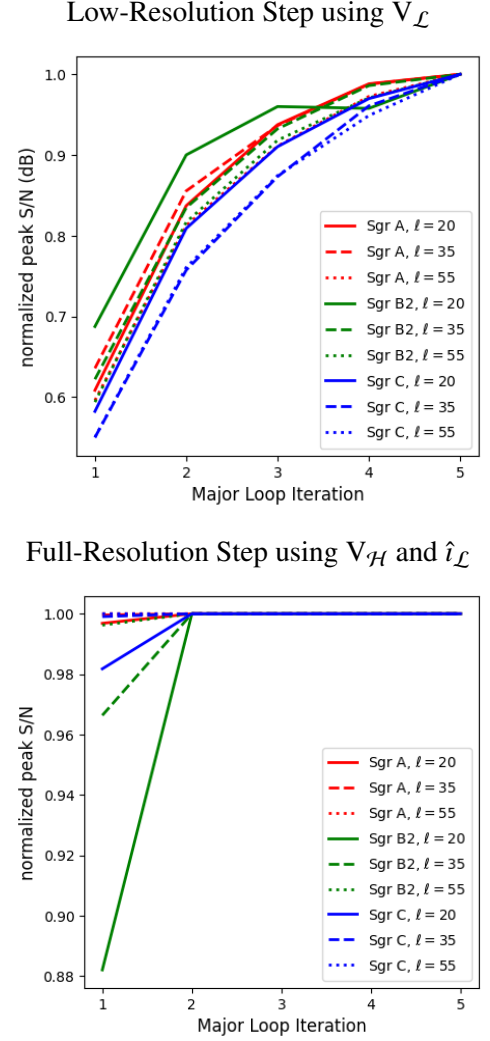
**Notes.** Values given are in peak S/N (dB), and normalized values are given in parenthesis for easier comparison. Reconstructions were obtained using five major-cycle iterations for each step.

Table 3 compares the peak S/Ns in dB of the final reconstructed images for each partition configuration. These reconstructions were obtained by running the multi-step reconstruction algorithm for five low-resolution and five full-resolution major cycles.

We found that there are only slight differences between the peak S/Ns when varying  $\ell$ , without any obvious trend. This led us to believe that the difference in final image quality is caused mainly by the choice and approximation of  $\lambda$ ,  $\sigma^2$ , and  $\eta^2$ . This is advantageous, as it implies that one can adjust the partition sizes to suit one's needs without drastically impacting the quality of the final reconstructed image. We also found that the choice of  $\delta$  does not greatly impact the quality of the final image. This could be because it only affects a few spatial frequencies, and thus the effect on the final image quality is dominated by other factors, such as parameter selection. Practically, it means that one can decrease  $\delta$  without greatly impacting the final reconstructed image, reducing the number of duplicate visibilities. However, care needs to be taken, as reducing  $\delta$  too much may result in frequency spillage, biasing the reconstruction.

### 5.2. Reconstruction speed

We evaluated two aspects of reconstruction speed: the number of major cycles required for convergence, and the number of



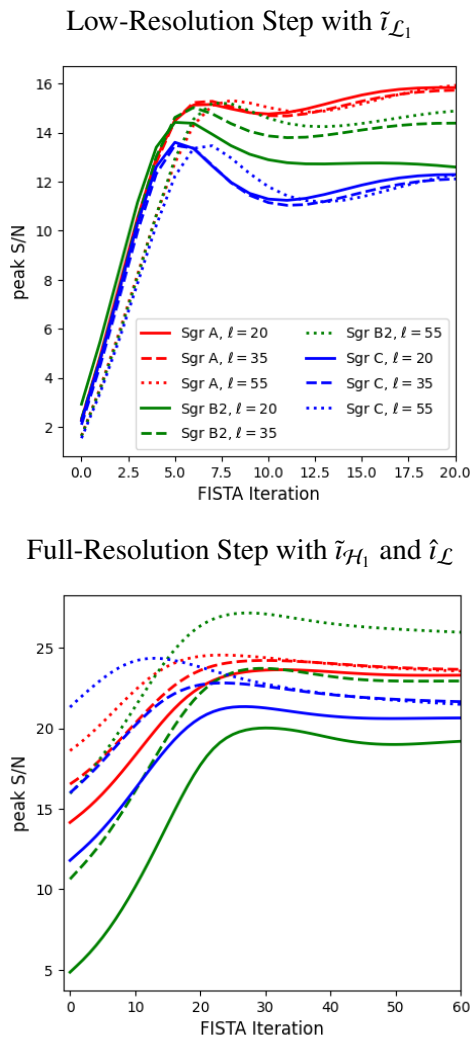
**Fig. 7.** Progression of peak S/N of reconstructed images for various partitioning configurations across the major cycles for both the low- and full-resolution reconstruction steps. peak S/Ns have been normalized by the maximum value of each configuration to allow for easier comparison. An exponential scale is used for the bottom image to better illustrate the differences between the different configurations. We fixed  $\delta = 5$  here, as we found it does not greatly impact the results.

required FISTA iterations. We evaluated the latter solely in the context of the first major cycle, as this is where the majority of information is reconstructed.

Figure 7 shows the progression of the peak S/Ns of the reconstructed images across major cycles. We normalized these by the maximum obtained peak S/N for the configuration across the major cycles for easier comparison. We found that for the low-resolution step, a lower value of  $\ell$  often results in a minor increase in convergence speed. This is expected, as a lower  $\ell$  corresponds to a simpler image to reconstruct. However, this is counteracted in the second step, where the corresponding configurations have more information and are more challenging to reconstruct. This can also be seen for the Sgr B2 and Sgr C datasets, which have more information in  $\tilde{i}_{H_1}$  and thus have a worse initial reconstruction quality than the Sgr A dataset.

A surprising result is that the full-resolution step converges after at most only two major cycles. There are two main reasons for this. The first is due to us selecting a larger  $\lambda$  to maximize peak S/N, which leads to deconvolved images in the later ma-





**Fig. 8.** Peak signal-to-noise ratio of the candidate solution of the specified FISTA iteration for the first major cycle of the low- and full-resolution steps, respectively. The full-resolution images were constructed with  $\hat{\mathcal{L}}$  set as a low-pass filtered version of the ground-truth image. We observed slightly faster reconstruction speeds for low values of  $\ell$  in the low-reconstruction step, but these configurations are significantly slower in the full-resolution step. It should be noted that although lower values of  $\ell$  improve faster for the low-resolution step, they do not necessarily converge to a better result, as illustrated in the dataset Sgr B2.

major cycles being zero, as the regularization term dominates the data fidelity term. We experimented with regularizing less aggressively, but we found that the deconvolved images, despite still reducing the residual, overfits the noise, detracting from the overall peak S/N. We supply some results in Sect. A.2 to illustrate this outcome. The second reason is that the peak S/N is dominated by large-scale features that have already been reconstructed in the low-resolution step.

Figure 8 shows the peak S/N of the candidate solutions across FISTA iterations for both reconstruction steps and their respective first major-cycle iterations. We set  $\hat{\mathcal{L}}$  to be a low-pass filtered version of the ground truth for the full-resolution step for these reconstructions. We observed a slight change in the low-resolution step reconstruction speed when varying  $\ell$ , but we observed a much more significant change in the full-resolution step. This is expected, as the low-resolution recon-

struction is simpler than the full one and uses less FISTA iterations, so changes are less pronounced. We note that for the low-resolution step, although lower values of  $\ell$  typically result in faster initial improvement, they do not guarantee that the final image will be of higher quality, such as with the Sgr B2 dataset. However, there are additional influencing factors in these cases, such as differences in ideal parameters. Thus, whether the final difference in image quality is due to  $\ell$  specifically is inconclusive. It should be noted that despite this, the different configurations all converge to roughly similar peak S/N values, as shown in Sect. 5.1.

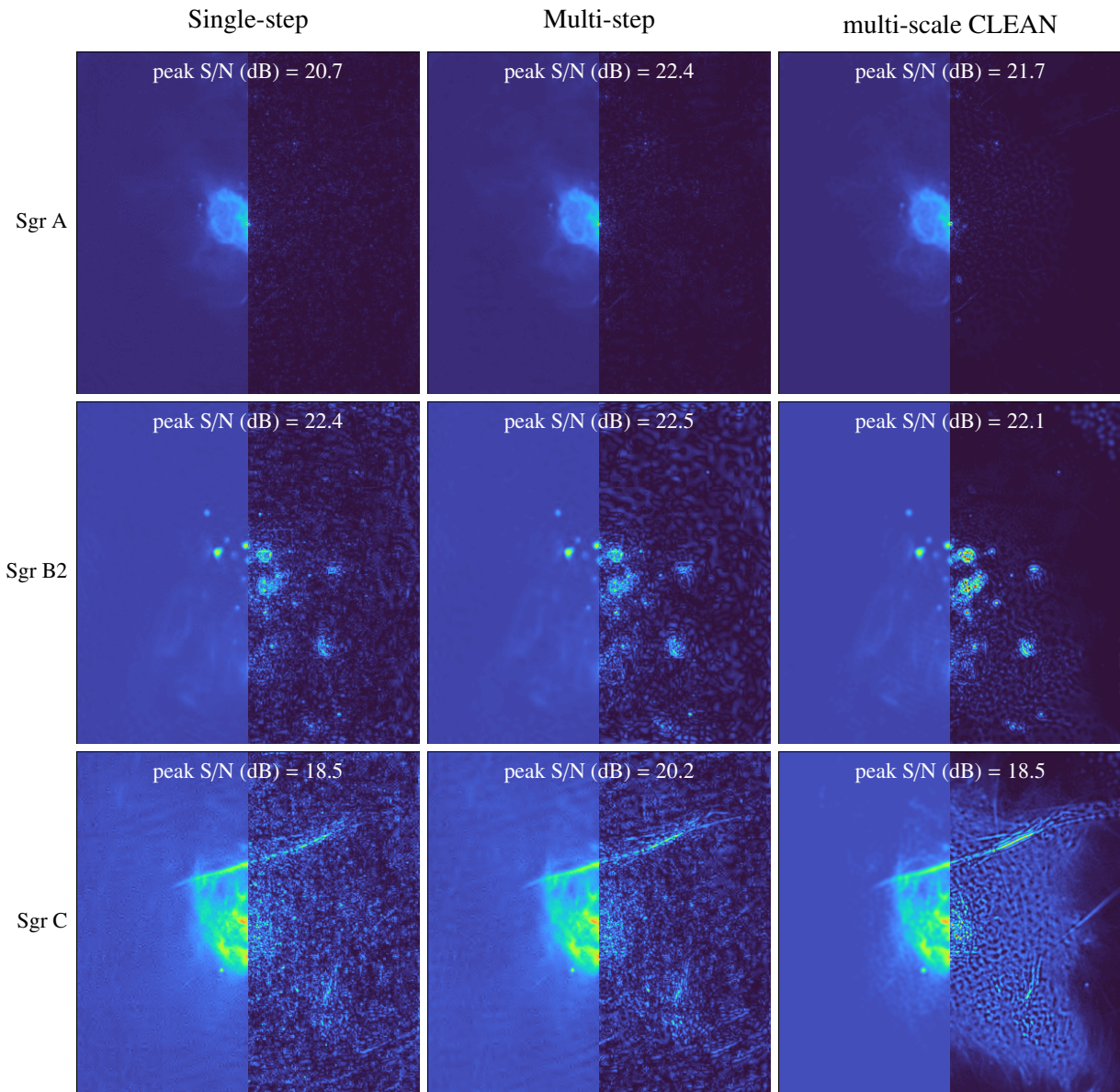
Taking into account the reconstruction speeds for both major-cycle and FISTA iterations, the slight change in FISTA iterations is counteracted by the larger number of required major cycles for the low-resolution step, and vice-versa for the full-resolution step. Thus, we can conclude that the choice of  $\ell$  and  $\delta$  should not greatly change the reconstruction speed, as the gains and losses roughly cancel out.

### 5.3. Comparison to reconstruction without partition of visibilities and to multi-scale CLEAN

We compare our multi-step method to both a single-step method that reconstructs using all visibilities, and to multi-scale CLEAN. Figure 9 shows the final reconstructed images and their absolute errors with their respective peak S/Ns in dB for a single-step reconstruction using all the visibilities  $\mathcal{V}_{\mathcal{L}} \cup \mathcal{V}_{\mathcal{H}}$ , multi-scale CLEAN, and the proposed multi-step reconstructions using first  $\mathcal{V}_{\mathcal{L}}$  and then  $\mathcal{V}_{\mathcal{H}}$ . We more aggressively increased  $\lambda_{\mathcal{L}}$  for the multi-step reconstructions compared to what was stated in Sect. 5, as we found this to remove wave-like artifacts that were occurring on the edges of the images. Specifically, we used  $\lambda_{\mathcal{L}_n} = 0.05 \|\tilde{\mathcal{L}}_n\|_2 \times 4^n$ . Although the three methods achieved comparable peak S/Ns, we also found there to be a visual difference in terms of their error images. In particular, we found multi-scale CLEAN to perform better at reconstructing the larger-scale features but worse at reconstructing the smaller scales, which the other two methods reconstruct to a similar level of accuracy.

Figure 10 shows the progression of all three methods across major-cycle iterations, with ten major-cycles for both the single-step method and multi-scale CLEAN and five low-resolution (major cycles 1-5) and five full-resolution (major cycles 6-10) major cycles for our method. We found that the three methods obtained reconstructions with comparable peak S/Ns after six to seven major cycles, with the single-step method and multi-scale CLEAN sometimes improving after this point, whereas our multi-step method does not. The main reason for this is that the other two methods have double the number of major cycles to improve the larger scale features, which dominate the peak S/N. In practice, as our method only improves for one to two major cycles in the full-resolution step, we can allocate more major cycles to the low-resolution step in order to achieve better peak S/Ns at the end of ten major cycles. Another important aspect to note is that the major cycles for our multi-step method are less computationally expensive, as less visibilities need to be (de)gridded. Thus, reconstructions for the same number of major cycles should be faster to compute when compared to the other tested methods. However, thorough benchmarking is required to determine the actual amount of speedup.

We also found the total number of FISTA iterations required by the single-step method to be comparable to the multi-step one. Figure 11 shows the reconstruction speed across FISTA iterations for the single-step method during the first major cycle. We observed that the single-step method requires two to three



**Fig. 9.** Reconstructed images (left) and absolute error images (right) for our multi-step method ( $\ell = 35, \delta = 5$ ) using five low-resolution major cycles and five full-resolution major cycles, a single-step all-baselines reconstruction method using ten major cycles, and multi-scale CLEAN using ten major cycles. The corresponding images share scales. Overall, the different methods reconstruct images of comparable peak S/Ns, with multi-scale CLEAN performing better with the larger scales, and the other two methods performing better with the smaller.

565 times less FISTA iterations compared to the full-resolution step  
 in the multi-step method, and it requires roughly 1.5 times more  
 than the low-resolution step. This, coupled with there being more  
 major-cycle iterations needed for the low-resolution step than the  
 full, means that the total number of FISTA iterations between  
 570 the single and multi-step approaches roughly equalize, leading  
 to comparable costs for each major cycle.

## 6. Conclusion and future work

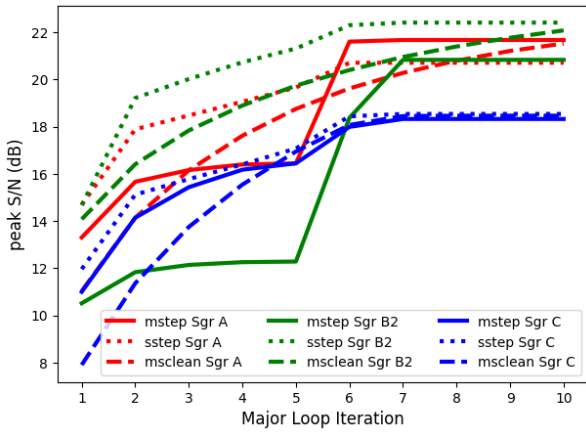
### 6.1. Conclusions

575 This paper proposes a radio-interferometric imaging framework  
 that allows for partitioning of visibilities by baseline. This has  
 several advantages over previous approaches that require all  
 baselines to be treated simultaneously in that (1) it alleviates  
 memory concerns, as not all baselines need to be gridded simul-

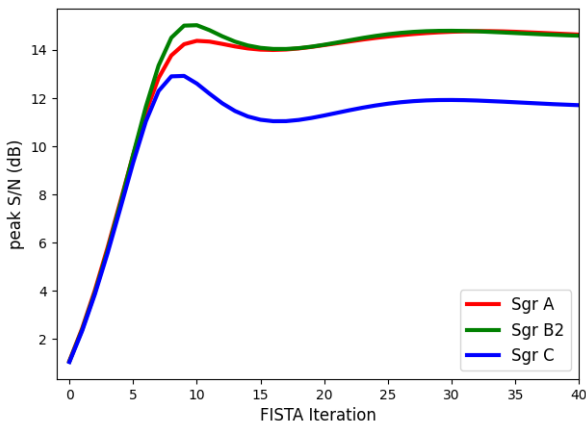
taneously; (2) it allows for more flexible distribution of visibility  
 data in a cluster; and (3) it enables progressive reconstruction.

We have presented our method in the context of sparsity 580  
 regularized convex optimization problems with over-redundant  
 wavelet dictionaries and compared it to a single-step approach in  
 the same framework that processes all baselines simultaneously  
 without partitioning as well as to RASCIL's implementation of  
 multi-scale CLEAN. In addition to a better data distribution, we 585  
 find our method reconstructs images of similar quality in roughly  
 the same total number of major cycles, assuming that the low-  
 and full-resolution major cycles for our method are counted sep-  
 arately. This is promising for the overall computational costs, as  
 each major-cycle iteration for our method should be cheaper to 590  
 compute given that we only process a subset of the visibilities.

The main drawback to our proposed method is that it creates  
 a secondary data product, the low-resolution image  $\hat{I}_{\mathcal{L}}$  used in the  
 full-resolution reconstruction step. This could be a significant



**Fig. 10.** Progression of image reconstruction quality across five low-resolution and five full-resolution major cycles for our proposed multi-step method and ten major cycles for a single-step method that reconstructs using all visibilities and multi-scale CLEAN. For our multi-step method, major cycles 1-5 are the low-resolution step, and major cycles 6-10 are the full-resolution step. We found the methods to have comparable peak S/Ns after 6-7 major cycles, with the single-step method and multi-scale CLEAN sometimes improving after this point. It should be noted that the major cycles for the multi-step method are computationally cheaper than the others, as there are less visibilities to (de)grid.

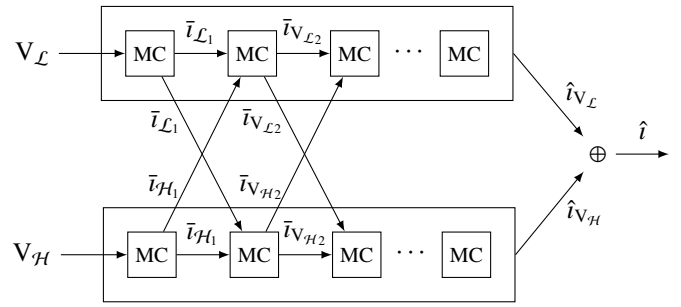


**Fig. 11.** Progression of peak S/N across FISTA iterations for the first major-cycle iteration of the single-step all-baselines reconstruction. Compared to the multi-step method, the convergence is faster than the full-resolution step but slower than the low-resolution step.

595 stumbling block, especially in the context of the SKA, where storage costs are a concern. However, this can be alleviated by reconstructing in Equation 4 a down-sampled image according to the partition configuration. This smaller-sized image can then be incorporated into the second data fidelity term of 5 after an appropriate modification of  $G_{\mathcal{L}}$  and the addition of a decimation operator to  $W\alpha$ .

## 6.2. Future work

605 There are several avenues to extending our work. One natural extension is studying how our framework performs with more than two steps, which will afford us information on the limit of partitioning by baseline. Furthermore, we are also looking to optimize our implementation in order to obtain quantitative results with regard to computational costs.



**Fig. 12.** Example workflow of how parallelism will work with our proposed method. We look to reconstruct full-resolution images in parallel, achieved by sharing the deconvolved residuals across different nodes after each major cycle, denoted as MC in the figure. A low and high-resolution image are produced after the first major cycle, denoted as  $\bar{v}_{L1}$  and  $\bar{v}_{H1}$  respectively. Afterwards, full-resolution images  $\bar{v}_{Ln}$  and  $\bar{v}_{Hn}$  are produced. The final reconstructed images of their respective nodes,  $\hat{v}_{L}$  and  $\hat{v}_{H}$ , are then combined to produce the final reconstructed image  $\hat{i}$ .

In addition, we are looking to incorporate parallelism into our framework. This can be done by reconstructing two full-resolution images simultaneously (rather than a low-resolution image first and then a full-resolution one), which can be achieved by sharing the deconvolved residuals between nodes after every major cycle. Thus, low- and high-resolution images are reconstructed after the first major cycle and full-resolution images from then on. The full-resolution reconstructions from each node can then be combined to produce the final reconstructed image. An example workflow of this can be seen in Fig. 12.

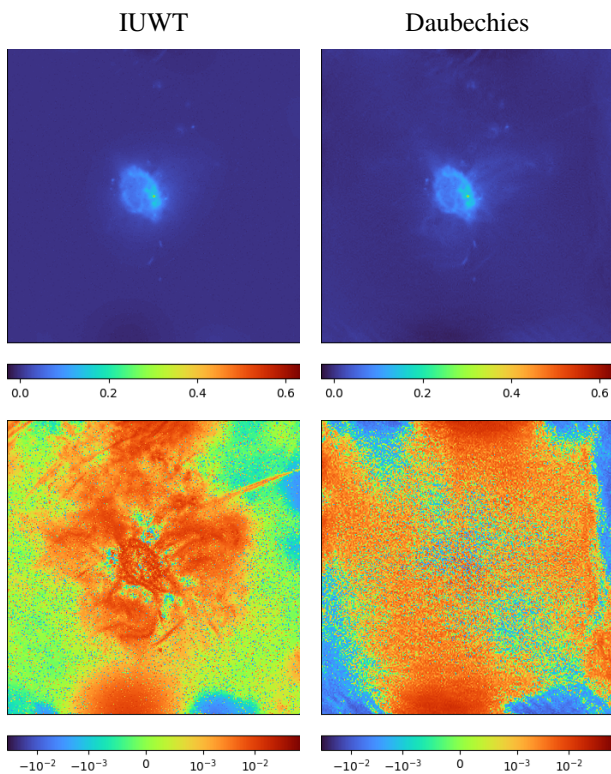
610  
615  
620  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000

*Acknowledgements.* This work was supported by DARK-ERA (ANR-20-CE46-0001-01)

## References

- Barnett, A. H., Magland, J., & Af Klinteberg, L. 2019, SISC, 41, C479  
 Beck, A. & Teboulle, M. 2009, SIIMS, 2, 183  
 Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. 2017, SIAM Rev, 59, 65  
 Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, A & A, 487, 419  
 Cai, X., Pratley, L., & McEwen, J. D. 2019, MNRAS, 485, 4559  
 Candes, E. J., Romberg, J., & Tao, T. 2006, IEEE Trans. Inf. Theory, 52, 489  
 Candès, E. J., Eldar, Y. C., Needell, D., & Randall, P. 2011, ACHA, 31, 59  
 Carrillo, R. E., McEwen, J. D., & Wiaux, Y. 2014, MNRAS, 439, 3591  
 Cornwell, T. J. 2008, IEEE J-STSP, 2, 793  
 Cornwell, T. J., Golap, K., & Bhatnagar, S. 2008, IEEE J-STSP, 2, 647  
 Cornwell, T. J. & Perley, R. A. 1992, A & A, 261, 353  
 Cornwell, T. J., Voronkov, M. A., & Humphreys, B. 2012, in Image Reconstruction from Incomplete Data VII, Vol. 8500, SPIE, 186  
 Cornwell, T. J., Wortmann, P., Nikolic, B., Wang, F., & Stolyarov, V. 2020, Radio Astronomy Simulation, Calibration and Imaging Library  
 Dabbech, A., Ferrari, C., Mary, D., et al. 2015, A & A, 576, A7  
 Daubechies, I. 1992, Ten lectures on wavelets (SIAM)  
 Donoho, D. L. 2006, IEEE Trans. Inf. Theory, 52, 1289  
 Garsden, H., Girard, J. N., Starck, J.-L., et al. 2015, A & A, 575, A90  
 Gheller, C., Taffoni, G., & Goz, D. 2023, RASTI, 2, 91  
 Heywood, I., Rammala, I., Camilo, F., et al. 2022, ApJ, 925, 165  
 Högbom, J. A. 1974, A & A Supplement, 15, 417  
 Jiang, M., Girard, J. N., Starck, J.-L., Corbel, S., & Tasse, C. 2015, in EUSIPCO (IEEE), 1646  
 Labate, M. G., Waterson, M., Alachkar, B., et al. 2022, JATIS, 8, 011024  
 Lauga, G., Repetti, A., Riccietti, E., et al. 2024, arXiv preprint arXiv:2403.13385  
 Mallat, S. G. & Z., Z. 1993, IEEE Trans. Signal Process, 41, 3397  
 Merry, B. 2016, A & C, 16, 140

- 655 Monnier, N., Guibert, D., Tasse, C., et al. 2022, in Proceedings of the 2022 IEEE  
SIPS
- Offringa, A. R., McKinley, B., Hurley-Walker, N., et al. 2014, MNRAS, 444, 606
- Onose, A., Carrillo, R. E., Repetti, A., et al. 2016, MNRAS, 462, 4314
- Ord, S. M., Mitchell, D. A., Wayth, R. B., et al. 2010, PASP, 122, 1353
- 660 Pratley, L., Johnston-Hollitt, M., & McEwen, J. D. 2019a, ApJ, 874, 174
- Pratley, L., McEwen, J. D., d’Avezac, M., et al. 2019b, arXiv preprint  
arXiv:1903.04502
- Rau, U. & Cornwell, T. J. 2011, A & A, 532, A71
- Romein, J. W. 2012, in Proceedings of the 26th ACM ICS, 321
- 665 Smirnov, O. M. 2011, A & A, 527, A106
- Starck, J.-L., Fadili, J., & Murtagh, F. 2007, IEEE Trans. Image Process., 16, 297
- Swart, G. P., Dewdney, P., & Cremonini, A. 2022, JATIS, 8, 011021
- Tasse, C., Hugo, B., Mirmont, M., et al. 2018, A & A, 611, A87
- Van Der Tol, S., Veenboer, B., & Offringa, A. R. 2018, A & A, 616, A27
- 670 Van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016, ApJ Supplement  
Series, 223, 2
- Veenboer, B., Petschow, M., & Romein, J. W. 2017, in IEEE IPDPS, 545
- Wiaux, Y., Jacques, L., Puy, G., Scaife, A. M. M., & Vanderghenst, P. 2009,  
MNRAS, 395, 1733
- 675 Ye, H., Gull, S. F., Tan, S. M., & Nikolic, B. 2022, MNRAS, 510, 4110



**Fig. A.1.** Reconstructed images from the Sgr A full visibilities dataset using both a concatenated dictionary of the first eight Daubechies wavelets (right) and IUWT (left) wavelets for a single major iteration using a single-step method that reconstructs with  $V_{\mathcal{L}} \cup V_{\mathcal{H}}$ . Provided below are the error images against the ground truth, which show that the Daubechies dictionary is able to better reconstruct the large-scale gas structures. We use a  $\log_{10}$  scale to better illustrate the structural differences for the error images.

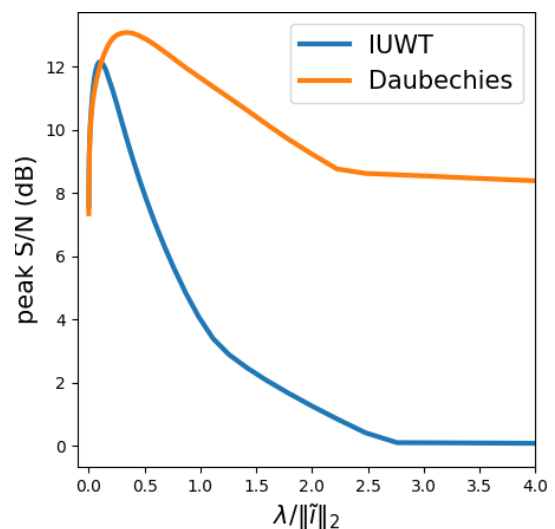
## Appendix A: Preliminary simulations for parameter selection

Our proposed method requires knowing the regularization parameters  $\lambda_{\mathcal{L}}$  and  $\lambda_{\mathcal{H}}$ , as well as the wavelet dictionary  $W$ . This appendix discusses our preliminary experiments to determine these.

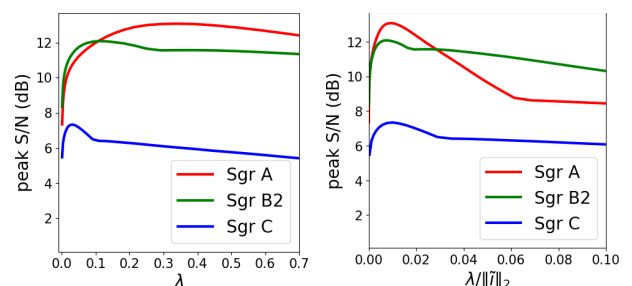
### A.1. Choice of wavelet dictionary

The choice of the wavelet dictionary can greatly affect the quality of the reconstruction. We experimented with using both a concatenation of the first 8 Daubechies (Daubechies 1992) wavelets and Isotropic Undecimated Wavelet Transform (IUWT) (Starck et al. 2007). Figure A.1 shows a comparison between reconstructed images of both dictionaries when using a single-step  $L^1$  regularized reconstruction of the Sgr A database when simultaneously processing all visibilities  $V_{\mathcal{L} \cup \mathcal{H}}$ .

We found that both dictionaries exhibited artifacts, with Daubechies exhibiting various tiling effects and discontinuities at larger values of  $\lambda$ , and IUWT exhibiting false sources. However, we found that the Daubechies dictionary allowed for both better reconstruction of large-scale structures, and was more robust to changes in  $\lambda$ , as shown in Fig. A.2, supporting our use for it in our experiments. A possible reason for why IUWT performs worse for our test cases is due to its isotropic nature, which performs poorly with the large-scale anisotropic features prevalent in our datasets, such as the gas clouds.



**Fig. A.2.** peak S/N (dB) of different  $\lambda$  values for the first major cycle when performing a single-step reconstruction using  $V_{\mathcal{L}} \cup V_{\mathcal{H}}$  for Sgr A using a concatenation of the first eight Daubechies wavelets and IUWT wavelets. We found Daubechies wavelets better reconstruct the large-scale features, as evidenced by the higher peak S/N values, and to also be more robust to changes of  $\lambda$ , having a larger range of values that work well.



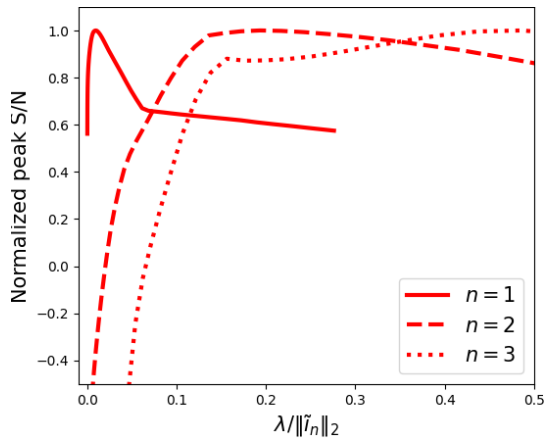
**Fig. A.3.** Reconstruction quality of images for our three datasets for different values of  $\lambda$ . The left image shows peak S/N in dB versus  $\lambda$  for the first major cycle for our three initial datasets, illustrating the different ideal values across different objects. This becomes less problematic when normalizing by the  $L^2$ -norm of the dirty image  $\|\tilde{t}_1\|_2$ . As shown on the right,  $\frac{\lambda}{\|\tilde{t}_1\|_2}$  is much better behaved, with the region of well-performing values all roughly lining up. The reconstruction algorithm used here is the single-step sparsity regularized method that reconstructs using all visibilities without partitioning.

### A.2. The regularization parameter $\lambda$

The selection of the ideal regularization parameters  $\lambda_{\mathcal{L}_n}$  and  $\lambda_{\mathcal{H}_n}$  depends on a slew of variables, such as the strength and nature of the object being imaged, the amount of noise present in the measurements, and the wavelet dictionary used. We perform preliminary experiments to determine these – deriving a general strategy is outside the scope of this work.

All our preliminary experiments are done with the single-step  $L^1$  regularized method that reconstructs with  $V_{\mathcal{L}} \cup V_{\mathcal{H}}$ , and evaluate  $\lambda$ . We generalize the findings here to  $\lambda_{\mathcal{L}}$  and  $\lambda_{\mathcal{H}}$  for our multi-step method, as the nature of reconstructed images are similar, albeit at different resolutions. Finally, as discussed in the previous section, we used a concatenation of the first eight Daubechies wavelets as our dictionary.

Figure A.3 shows the results for our three datasets for the first major cycle. The left image plots  $\lambda$  as the x-axis, whereas the right plots the normalized value  $\frac{\lambda}{\|\tilde{t}_1\|_2}$ . We found that normal-



**Fig. A.4.** Normalized peak S/N against  $\frac{\lambda}{\|\tilde{r}_n\|_2}$  for the first three major cycles when reconstructing the Sgr A database using the single-step method that reconstructs using  $V_{\mathcal{L}} \cup V_{\mathcal{H}}$ . The ideal values for  $\frac{\lambda}{\|\tilde{r}_n\|_2}$  increase as we progress through the major cycles.

ization allows for the well performing regions to line-up, implying that we can pass a constant value for our simulations without the need for dataset specific parameter tuning. We found  $\frac{\lambda}{\|\tilde{r}_1\|_2} = 0.015$  to work well for the single-step reconstructions, and  $\frac{\lambda_{\mathcal{L}}}{\|\tilde{r}_{\mathcal{L}_1}\|_2} = \frac{\lambda_{\mathcal{H}}}{\|\tilde{r}_{\mathcal{H}_1}\|_2} = 0.05$  to work well for our multi-step method.

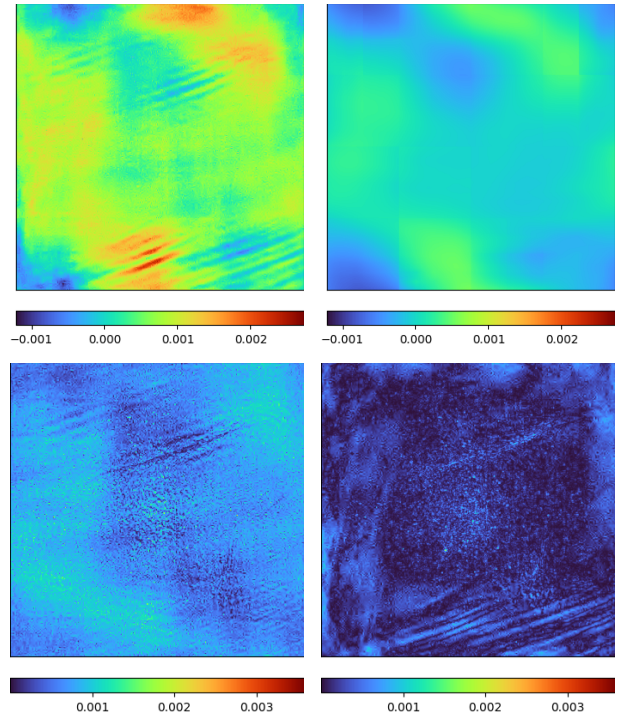
We hypothesize that the larger ideal values for the low-resolution step can be explained by there being less frequencies, ergo needing less wavelet coefficients. On the other hand, the larger values for the full-resolution step could be explained by the low-resolution data fidelity term biasing the reconstruction to have less noise and fine-scale features, which also increases the sparsity of the wavelet coefficients.

We also evaluate how  $\lambda$  changes across major-cycle iterations, which is necessary as the nature of residual images  $\tilde{r}_n$  changes. Figure A.4 illustrates the behavior of  $\frac{\lambda}{\|\tilde{r}_n\|_2}$  across three different major cycles for the Sgr A dataset. We normalize by the maximum peak S/N obtained for each respective major cycle for easier comparison. The ground truth images we use to compute the peak S/N are defined as  $I - \sum_N \hat{r}_n$ , where  $I$  is the initial ground truth, and  $\hat{r}_n$  is the best reconstructed image among the different  $\frac{\lambda}{\|\tilde{r}_n\|_2}$  values for the  $n$ th major cycle.

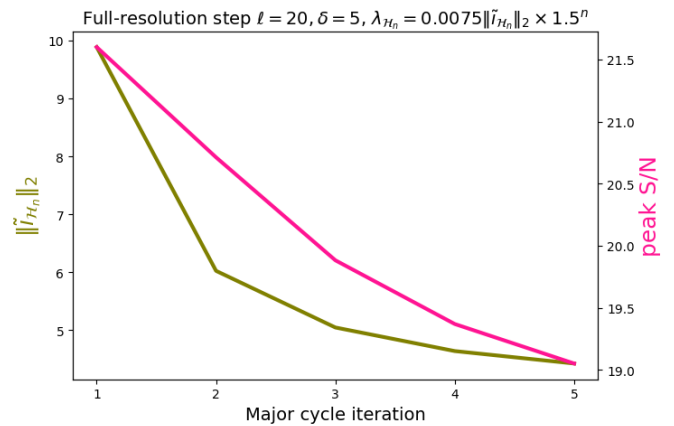
It can be seen that the ideal values for  $\frac{\lambda}{\|\tilde{r}_n\|_2}$  generally increase as we progress through the major-cycle iterations. This is expected as the residual images are increasingly dominated by noise, necessitating larger values of  $\frac{\lambda}{\|\tilde{r}_n\|_2}$  to suppress it. However, care needs to be taken not to increase it too aggressively, as this will lead to all fine scale features being ignored, as shown in Fig. A.5. We found that multiplying  $\frac{\lambda}{\|\tilde{r}_n\|_2}$  by  $2^n$  for the  $n$ th major cycle provides a good trade-off between the two.

Finally, we would like to note that for the full-resolution reconstruction, a less aggressive regularization parameter  $\lambda$  does allow for further reduction of the residual in latter major cycles. However, we found this to overfit to noise, reducing the peak S/N of the overall reconstruction. Figure A.6 illustrates this for the Sgr A  $\ell = 20, \delta = 5$  dataset.

$\lambda = 0.018$ , peak S/N (dB)=-6.74  $\lambda = 0.516$ , peak S/N (dB)=2.73



**Fig. A.5.** Reconstructed images as well as their absolute errors of the Sgr C dataset for the second major cycle residual image  $\tilde{r}_2$  using the single-step method, with a lower (left) and higher (right)  $\lambda$ . The larger  $\lambda$  has a better peak S/N, as it reduces the error on a large scale, but it is not able to reconstruct any of the finer-scale details.



**Fig. A.6.**  $L^2$ -norm of the residual and the peak S/N of the reconstructed image for the full-resolution step across major cycles. A less aggressive regularization parameter was used here, which results in further reduction of the residual in the later major cycles. However, this also detracts from the peak S/N, as the reconstructed image is overfits the noise.