

# Parallel faceted imaging in radio interferometry via proximal splitting (Faceted HyperSARA): when precision meets scalability

Pierre-Antoine Thouvenin,<sup>1,2\*</sup> Abdullah Abdulaziz,<sup>1\*</sup> Ming Jiang,<sup>3\*</sup> Arwa Dabbech,<sup>1\*</sup> Audrey Repetti,<sup>1,4\*</sup> Adrian Jackson,<sup>5</sup> Jean-Philippe Thiran,<sup>3</sup> and Yves Wiaux<sup>1†</sup>

<sup>1</sup>*Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom*

<sup>2</sup>*Université de Lille, CNRS, Centrale Lille, UMR 9189 – CRISTAL – Centre de Recherche en Informatique, Signal et Automatique de Lille, F-59000 Lille, France*

<sup>3</sup>*Signal Processing Laboratory 5 (LTS5), École Polytechnique Fédérale de Lausanne, CH-1015, Lausanne, Switzerland*

<sup>4</sup>*Department of Actuarial Mathematics & Statistics, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom*

<sup>5</sup>*EPCC, University of Edinburgh, Edinburgh EH8 9BT, United Kingdom*

18 March 2020

## ABSTRACT

Upcoming radio interferometers are aiming to image the sky at new levels of resolution and sensitivity, with wide-band image cubes reaching close to the Petabyte scale for SKA. Modern proximal optimization algorithms have shown a potential to significantly outperform CLEAN thanks to their ability to inject complex image models to regularize the inverse problem for image formation from visibility data. They were also shown to be scalable to large data volumes thanks to a splitting functionality enabling the decomposition of data into blocks, for parallel processing of block-specific data-fidelity terms of the objective function. In this work, the splitting functionality is further exploited to decompose the image cube into spatio-spectral facets, and enable parallel processing of facet-specific regularization terms in the objective. The resulting “Faceted HyperSARA” algorithm is implemented in MATLAB (code available on GitHub). Simulation results on synthetic image cubes confirm that faceting can provide a major increase in scalability at no cost in imaging quality. A proof-of-concept reconstruction of a 15 GB image of Cyg A from 7.4 GB of VLA data, utilizing 496 CPU cores on a HPC system for 68 hours, confirms both scalability and a quantum jump in imaging quality from CLEAN. Assuming slow spectral slope of Cyg A, we also demonstrate that Faceted HyperSARA can be combined with a dimensionality reduction technique, enabling utilizing only 31 CPU cores for 142 hours to form the Cyg A image from the same data, while preserving reconstruction quality. Cyg A reconstructed cubes are available online.

**Key words:** techniques: image processing, techniques: interferometric.

## 1 INTRODUCTION

Modern radio interferometers, such as the Karl G. Jansky Very Large Array (VLA) (Perley et al. 2011), the LOw Frequency ARray (LOFAR) (Van Haarlem et al. 2013) and the MeerKAT radio telescope (Jonas et al. 2018) generate extremely large volumes of data, with the aim of producing images of the radio sky at unprecedented resolution and dynamic range over thousands of spectral channels. The upcoming Square Kilometer Array (SKA) (Dewdney et al.

2013) will form wide-band images about 0.56 Petabyte in size (assuming double precision) from even larger visibility data volumes (Scaife 2020). SKA is expected to bring answers to fundamental questions in astronomy<sup>1</sup>, such as improving our understanding of cosmology and dark energy (Rawlings et al. 2004), investigating the origin and evolution of cosmic magnetism (Gaensler et al. 2004) and probing the early universe where the first stars were formed (Carilli et al. 2004). To achieve the expected scientific goals, it is of paramount importance to design efficient imaging algorithms which meet the capabilities of such powerful instru-

\* The first five authors contributed equally to this work.

† E-mail: y.wiaux@hw.ac.uk

<sup>1</sup> <https://www.skatelescope.org/science/>

ments. On the one hand, appropriate algorithms need to inject complex prior image models to regularize the inverse problem for image formation from visibility data, which only provide incomplete Fourier sampling. On the other hand, these algorithms need to be highly parallelizable in order to scale with the sheer amount of data and the large size of the wide-band image cubes to be recovered.

A plethora of radio-interferometric (RI) imaging approaches have been proposed in the literature, which can be classified into three main categories. A first class of methods is the celebrated CLEAN family (*e.g.* Högbom 1974; Schwab & Cotton 1983; Bhatnagar & Cornwell 2004; Cornwell 2008; Rau & Cornwell 2011; Offringa & Smirnov 2017). In particular, Rau & Cornwell (2011) proposed the multi-scale multi-frequency deconvolution algorithm (MS-MFS), leveraging Taylor series and multi-scale CLEAN to promote spectral smoothness of the wide-band image cube. More recently, Offringa & Smirnov (2017) have proposed the Joined Channel CLEAN algorithm (JC-CLEAN), where multi-scale CLEAN components are identified from the integrated residual image (*i.e.* the sum of the residual images over all the channels). Albeit simple and computationally efficient, CLEAN-based algorithms provide a limited imaging quality in high resolution and high sensitivity acquisition regimes. This shortcoming partly results from their greedy nature and their lack of flexibility in injecting complex prior information to regularize the inverse imaging problem. Moreover, these algorithms often require careful tuning of the associated parameters.

The second class of methods relies on Bayesian inference techniques (*e.g.* Sutton & Wandelt 2006; Sutter et al. 2014; Junklewitz et al. 2015, 2016; Arras et al. 2019). For instance, Sutter et al. (2014) proposed a monochromatic Bayesian method based on Markov chain Monte Carlo (MCMC) sampling, considering a Gaussian image prior. Since MCMC sampling methods are computationally very expensive, an efficient variant was proposed in Junklewitz et al. (2016); Arras et al. (2019) to perform approximate Bayesian inference, formulated in the framework of information theory. Importantly, Bayesian methods naturally enable the quantification of uncertainty about the image estimate. However, this type of approaches cannot currently scale to the data regime expected from modern telescopes.

The third class of approaches leverages optimization methods allowing sophisticated prior information to be considered, such as sparsity in an appropriate transform domain, smoothness, etc. (*e.g.* Wiaux et al. 2009; Li, F. et al. 2011; Dabbech et al. 2012; Carrillo et al. 2012; Wenger & Magnor 2014; Garsden et al. 2015; Dabbech et al. 2015; Girard et al. 2015; Ferrari et al. 2015; Abdulaziz et al. 2016; Jiang et al. 2017; Abdulaziz et al. 2019b). From the perspective of optimization theory, the inverse imaging problem is approached by defining an objective function, consisting in a sum of a data-fidelity term and a regularization term promoting a prior image model to compensate for the incompleteness of the visibility data. The sought image is estimated as a minimizer of this objective function, and is computed through iterative algorithms, which benefit from well-established convergence guarantees. For instance, Ferrari et al. (2015) promote spatial sparsity in a redundant wavelet domain and spectral sparsity in a Discrete Cosine Transform. Wenger & Magnor (2014) promote spectra com-

posed of a smooth contribution affected by local sparse deviations. In the last decade, Wiaux and collaborators proposed advanced image models: the *average sparsity* prior in monochromatic imaging (SARA) (Carrillo et al. 2012, 2013, 2014; Onose et al. 2016a,b, 2017; Pratley et al. 2017; Dabbech et al. 2018), the *low-rankness* and *joint average sparsity* priors for wide-band imaging (HyperSARA) (Abdulaziz et al. 2016, 2017, 2019b), and the *polarization constraint* for polarized imaging (Polarized SARA) (Birdi et al. 2018)<sup>2</sup>. These models have been reported to result in significant improvements in the reconstruction quality in comparison with state-of-the-art CLEAN-based imaging methods, at the expense of an increased computation cost.

Note that, from a Bayesian perspective, the objective function can be seen as the negative logarithm of a posterior distribution, with the minimizer corresponding to a Maximum *A Posteriori* (MAP) estimate. Methods for uncertainty quantification by convex optimization have also been tailored recently, which enable assessing the degree of confidence in specific structures appearing in the MAP estimate (Repetti et al. 2018, 2019; Abdulaziz et al. 2019a). In this work we focus solely on image estimation.

Convex optimization offers intrinsically parallel algorithmic structures, such as proximal splitting methods (Combettes & Pesquet 2011; Komodakis & Pesquet 2015). In such algorithmic structures, multi-term objective functions can be minimized, all terms being handled in parallel at each iteration. Each term is involved by application of its so-called proximal operator, which acts as a simple denoising operator (*e.g.* a sparsity regularization term will induce a thresholding operator). The algorithms of the SARA family are all powered by an advanced proximal splitting method known as the primal-dual forward-backward (PDFB) algorithm (Condat 2013; Vũ 2013; Pesquet & Repetti 2015). The splitting functionality of PDFB is utilized in these approaches to enable the decomposition of data into blocks and parallel processing of the block-specific data-fidelity terms of the objective function, which provides scalability to large data volumes. The SARA family however models the image as a single variable, and the computational and storage requirements induced by complex regularization terms can be prohibitive for very large image size, in particular for wide-band imaging.

We address this bottleneck in the present work. We propose to decompose the target image cube into regular, content-agnostic, spatially overlapping spatio-spectral facets, with which are associated facet-specific regularization terms in the objective function. We further exploit the splitting functionality of PDFB to enable parallel processing of the regularization terms and ultimately provide further scalability.

Note that faceting is not a novel paradigm in RI imaging: it has often been considered for calibration purposes in the context of wide-field imaging, assuming piecewise constant direction-dependent effects. For instance, van Haarlem, M. P. et al. (2013) proposed an image tessellation scheme for LOFAR wide-field images, which has been leveraged by Tasse et al. (2018) in the context of wide-

<sup>2</sup> Associated software on the Puri-Psi webpage: <https://basp-group.github.io/Puri-Psi/>.

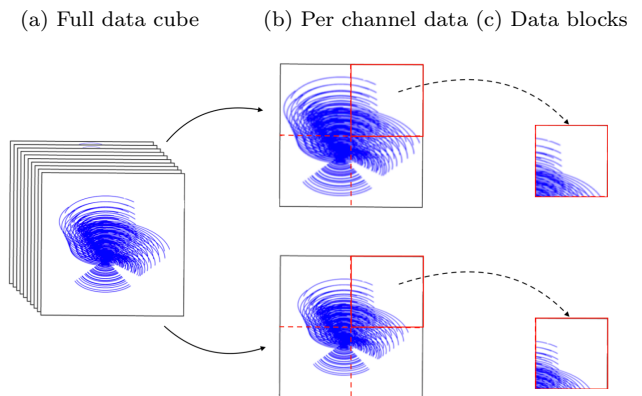
field wide-band calibration and imaging. However, except for (Naghizadeh et al. 2018) and to the best of the authors' knowledge, facet imaging has hitherto been essentially addressed with CLEAN-based algorithms. This class of approaches not only lacks theoretical convergence guarantees, but also does not offer much flexibility to accommodate advanced regularization terms. In contrast with (Naghizadeh et al. 2018), the proposed faceting approach does not need to be tailored to the content of the image, and thus offers more flexibility to design balanced facets exclusively based on computational considerations.

The reconstruction performance of the Faceted HyperSARA is evaluated against HyperSARA and SARA on synthetic data. We further validate the performance and scalability potential of our approach through the reconstruction of a 15 GB image cube of Cyg A from 7.4 GB of VLA observations across 480 channels. Our results confirm the recent discovery of Cyg A2, a second super-massive black hole in Cyg A (Perley et al. 2017). Finally, we combine Faceted HyperSARA with a joint image and data dimensionality reduction technique in order to provide further scalability. In practice, acknowledging the slow spectral slope of Cyg A in the frequency range of interest, we target a 16-fold reduction in spectral resolution. We also apply a data dimensionality reduction technique relying on visibility gridding, offering here a 33-fold reduction in the data volume. We validate the stable performance of Faceted HyperSARA on the reduced-size inverse problem in comparison to the approach without dimensionality reduction.

The remainder of the article is organized as follows. Section 2 introduces the proposed faceted prior model and associated objective function underpinning Faceted HyperSARA. The associated algorithm is described in Section 3, along with the different levels of parallelization exploited in the proposed MATLAB implementation. Performance validation is first conducted on synthetic data in Section 4. We successively evaluate the influence of spectral and spatial faceting for a varying number of facets and spatial overlap, both in terms of reconstruction quality and computing time. Section 5 is focused on the validation of the proposed approach on real VLA observations in terms of precision and scalability. Section 6 illustrates the potential of combining Faceted HyperSARA and dimensionality reduction for further scalability. Conclusions and perspectives are reported in Section 7.

## 2 PRIOR MODEL AND OBJECTIVE FUNCTION

In this section, focusing on optimization-based approaches, we first recall the discrete version of the inverse problem for RI image formation from visibility data. We then formulate the general structure of the objective function for the state-of-the-art SARA and HyperSARA approaches. Finally, we introduce spatio-spectral facets and the associated prior model, leading to the formulation of the objective function for the proposed Faceted HyperSARA approach.



**Figure 1.** Illustration of the data blocking strategy. Starting from the full data cube (a), each of the  $L$  channels represented in (b) is decomposed into  $B$  data blocks (c). The data blocks for each channel are associated with separate data-fidelity terms in the objective function, processed by independent workers.

### 2.1 Wide-band inverse problem

Wide-band RI imaging consists in estimating unknown radio images of the sky over  $L$  frequency channels. Focusing on intensity imaging, and assuming a small field of view on the celestial sphere, each pair of antennae probes, at each observation frequency, a Fourier component of the sky surface brightness. The Fourier mode is given by the projection of the corresponding baseline in the plane perpendicular to the line of sight, in units of the observation wavelength (Thompson et al. 2007). The collection of data (called visibilities) from all baselines accumulated over the whole duration of observation provides an incomplete coverage of the 2D Fourier plane (also called  $uv$ -plane) of the image of interest. The RI measurements can be modeled for each frequency channel index  $l \in \{1, \dots, L\}$  as (Abdulaziz et al. 2016, 2019b)

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

where  $\mathbf{y}_l \in \mathbb{C}^{M_l}$  is the vector of  $M_l$  visibilities acquired in the channel  $l \in \{1, \dots, L\}$ , weighted with the diagonal noise-whitening matrix  $\Theta_l \in \mathbb{R}^{M_l \times M_l}$ .  $\bar{\mathbf{x}}_l \in \mathbb{R}_+^N$  is the underlying image. The vector  $\mathbf{n}_l \in \mathbb{C}^{M_l}$  represents measurement noise, modeled as a realization of a complex white Gaussian noise. The measurement operator  $\Phi_l$  is composed of a zero-padding and scaling operator  $\mathbf{Z} \in \mathbb{R}^{K \times N}$ , the 2D Discrete Fourier Transform represented by the matrix  $\mathbf{F} \in \mathbb{C}^{K \times K}$ , and a non-uniform Fourier transform interpolation matrix  $\mathbf{G}_l \in \mathbb{C}^{M_l \times K}$ . Each row of  $\mathbf{G}_l$  contains a compact support interpolation kernel centered at the corresponding  $uv$ -point (Fessler & Sutton 2003), enabling the computation of the Fourier mode associated with each visibility from surrounding discrete Fourier points. Note that at the sensitivity of interest to the new generation of radio telescopes, direction-dependent effects (DDEs), of either atmospheric or instrumental origin, complicate the RI measurement equation. For each visibility, the sky surface brightness is pre-modulated by the product of a DDE pattern specific to each antenna. The DDEs are often unknown and need to be calibrated jointly with the imaging process (Repetti et al. 2017; Repetti & Wiaux 2017; Thouvenin et al. 2018; Birdi et al. 2019). Focussing here on the imaging problem, *i.e.* assuming DDEs are known,

they can simply be integrated into the forward model (1) by building extended interpolation kernels into each row of  $\mathbf{G}_l$ , resulting from the convolution of the non-uniform Fourier transform kernel with a compact-support representation of the Fourier transform of the involved DDEs. Finally, the matrix  $\Theta_l$  contains on its diagonal the inverse of the noise standard deviation associated with each original measurement. This assumes that the original visibility vector was multiplied by  $\Theta_l$  to produce a measurement vector  $\mathbf{y}_l$  affected by a random independent and identically distributed (i.i.d., or white) Gaussian noise. This noise-whitening operation corresponds to what is known as natural weighting in RI imaging.

When addressing the model (1), a first bottleneck arises from the sheer volume of the data. To address this issue, Onose et al. (2016b, 2017) have proposed a data blocking strategy, which has been exploited in the context of wide-band imaging by Abdulaziz et al. (2019b). The visibility vectors  $\mathbf{y}_l$  are decomposed into  $B$  blocks  $(\mathbf{y}_{l,b})_{1 \leq b \leq B}$ , which can be handled in parallel by advanced imaging algorithms. The data model (1) can thus be formulated for any  $(l, b) \in \{1, \dots, L\} \times \{1, \dots, B\}$  as

$$\mathbf{y}_{l,b} = \Phi_{l,b} \bar{\mathbf{x}}_l + \mathbf{n}_{l,b}, \text{ with } \Phi_{l,b} = \Theta_{l,b} \mathbf{G}_{l,b} \mathbf{F} \mathbf{Z}, \quad (2)$$

where  $\mathbf{y}_{l,b} \in \mathbb{C}^{M_{l,b}}$  is the vector of  $M_{l,b}$  visibilities associated with the  $b$ -th block in the channel  $l$ . Different blocking strategies can be adopted, *e.g.* based on a tessellation of the  $uv$ -space into balanced sets of visibilities (Onose et al. 2016b), or on a decomposition of the data into groups of snapshots (Dabbech et al. 2018) (see Figure 1).

## 2.2 General form of the objective function

Estimating the underlying wide-band sky image  $\bar{\mathbf{X}} = (\bar{\mathbf{x}}_l)_{1 \leq l \leq L}$  from incomplete Fourier measurements is a severely ill-posed inverse problem, which calls for powerful regularization terms to encode a prior image model. In this context, wide-band RI imaging can be formulated as the following constrained optimization problem

$$\underset{\mathbf{X}=(\mathbf{x}_l)_{1 \leq l \leq L} \in \mathbb{R}_+^{N \times L}}{\text{minimize}} \sum_{l=1}^L \sum_{b=1}^B \iota_{\mathcal{B}(\mathbf{y}_{l,b}, \varepsilon_{l,b})}(\Phi_{l,b} \mathbf{x}_l) + r(\mathbf{X}), \quad (3)$$

where the indices  $(l, b) \in \{1, \dots, L\} \times \{1, \dots, B\}$  refer to a data block  $b$  of channel  $l$ ,  $\mathcal{B}(\mathbf{y}_{l,b}, \varepsilon_{l,b}) = \{\mathbf{z} \in \mathbb{C}^{M_{l,b}} \mid \|\mathbf{z} - \mathbf{y}_{l,b}\|_2 \leq \varepsilon_{l,b}\}$  denotes the  $\ell_2$ -ball centred in  $\mathbf{y}_{l,b}$  of radius  $\varepsilon_{l,b} > 0$ , and  $\varepsilon_{l,b}$  reflects the noise statistics. The notation  $\iota_{\mathcal{B}(\mathbf{y}_{l,b}, \varepsilon_{l,b})}$  denotes the indicator function of the  $\ell_2$  ball  $\mathcal{B}(\mathbf{y}_{l,b}, \varepsilon_{l,b})$ . Specifically, let  $C$  be a non-empty, closed, convex subset of  $\mathbb{C}^N$ , then  $\iota_C$  denotes the indicator function of  $C$ , defined by  $\iota_C(\mathbf{z}) = +\infty$  if  $\mathbf{z} \in C$  and 0 otherwise. On the one hand, the indicator functions  $\iota_{\mathcal{B}(\mathbf{y}_{l,b}, \varepsilon_{l,b})}$  act as data-fidelity terms, in that they ensure the consistency of the modeled data with the measurements and reflect the white Gaussian nature of the noise (Carrillo et al. 2012). On the other hand, the function  $r$  encodes a prior model of the unknown image cube. The priors characterizing the state-of-the-art SARA and HyperSARA approaches, as well as the proposed Faceted HyperSARA approach, are discussed in what follows. Finally, note that an additional non-negativity prior is imposed in all approaches of the SARA family focusing on intensity imaging, with the aim to preserve the physical consistency of the

estimated surface brightness. This generalizes to the polarization constraint when solving for all the Stokes parameters

## 2.3 State-of-the-art average sparsity priors

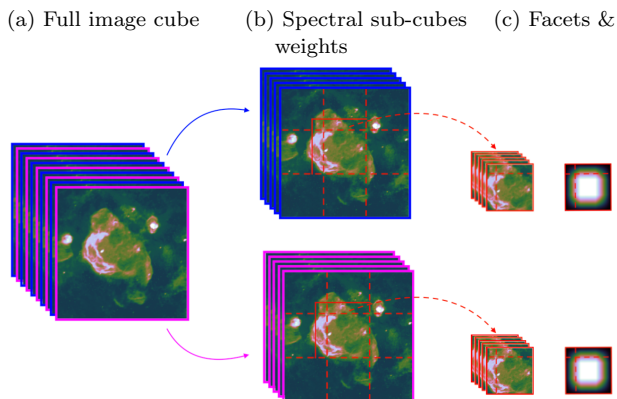
Sparsity-based priors combined with optimization techniques have proved to be very efficient for astronomical imaging, in particular in the context of radio interferometry (Wiaux et al. 2009; Carrillo et al. 2012; Garsden et al. 2015; Onose et al. 2016b). These techniques aim to solve the underlying image recovery problem by enforcing sparsity of the estimated image in an appropriate domain.

In this context, the prior of choice is the  $\ell_0$  pseudo-norm, which counts the number of non-zero coefficients of its argument (Donoho 2006). However, minimizing this function, which is neither convex nor smooth, is a NP-hard problem. A common alternative consists in replacing it by its convex envelope, the  $\ell_1$  norm (Donoho & Stark 1989; Donoho & Logan 1992). When combined to other convex terms, *e.g.* the non-negativity constraint and the  $\ell_2$ -ball data-fidelity constraint in (3),  $\ell_1$ -norm priors form a convex objective function, which can be efficiently minimized by powerful iterative algorithms under well-established guarantees on the convergence of the iterates towards a global minimum.

Although the  $\ell_1$  prior has been widely used over the last decades to promote sparsity, it induces an undesirable dependence on the coefficients' magnitude. Indeed, unlike the  $\ell_0$  prior, the  $\ell_1$  norm penalizes more the larger coefficients than the smaller. To address this imbalance, a log-sum prior can be used. In particular, Candès & Boyd (2008); Candès et al. (2009) proposed to use a majorization-minimization framework (Hunter & Lange 2004) to minimize objective functions with a log-sum prior, leading to a reweighted- $\ell_1$  approach consisting in minimizing a sequence of convex objectives with weighted  $\ell_1$  norm priors, acting as convex relaxations of the log-sum prior. In practice, sequentially minimizing convex problems with weighted- $\ell_1$  priors is indeed much simpler than minimizing a non-convex problem with a log-sum prior. From a convergence point of view, multiple works have recently shown that the set of minimizers resulting from a reweighted- $\ell_1$  procedure coincides with the one obtained by minimizing a problem with log-sum prior (Ochs et al. 2015; Geiping & Moeller 2018; Ochs et al. 2019; Repetti & Wiaux 2019). In the following paragraphs, the SARA (Carrillo et al. 2012) and HyperSARA (Abdulaziz et al. 2019b) log-sum priors are presented, considered as benchmarks to assess the proposed spatio-spectral faceted prior.

### 2.3.1 SARA prior

The image prior underpinning the Sparse Averaging Reweighted Analysis (SARA) has proved efficient for astronomical imaging, and in particular for RI imaging (Carrillo et al. 2012; Onose et al. 2016b; Abdulaziz et al. 2019b). It promotes sparsity by minimizing a log-sum prior, considering a highly redundant transformed domain  $\Psi^\dagger \in \mathbb{R}^{N \times I}$  defined as the concatenation of wavelet bases (first eight Daubechies wavelets and the Dirac basis), leading to the notion of average sparsity over the bases of interest. The



**Figure 2.** Illustration of the proposed faceting scheme, using a 2-fold spectral interleaving process and 9-fold spatial tiling process. The full image cube variable (a) is divided into two spectral sub-cubes (b) with interleaved channels (for a 2-fold interleaving, even and odd channels respectively define a sub-cube). Each sub-cube is spatially faceted. A regular tessellation (dashed red lines) is used to define spatio-spectral tiles. The spatio-spectral facets result from the augmentation of each tile to produce an overlap between facets (solid red lines). Panel (c) shows a single facet (left), as well as the spatial weighting scheme (right) with linearly decreasing weights in the overlap region. Note that, though the same tiling process is underpinning the nuclear norm and  $\ell_{21}$  norm regularization terms, the definition of the appropriate overlap region is specific to each of these terms (via the selection operators  $\mathbf{S}_q$  and  $\tilde{\mathbf{S}}_q$  in (9)).

log-sum prior addressed by SARA is of the form

$$r(\mathbf{X}) = \tilde{\mu} \sum_{l=1}^L \sum_{i=1}^I \log (|\Psi^\dagger \mathbf{X}|_{i,l} + \nu), \quad (4)$$

where  $\tilde{\mu} > 0$  and  $\nu > 0$  are regularization parameters, and  $[\Psi^\dagger \mathbf{X}]_{i,l}$  denotes the  $(i, l)$ -th coefficient of  $\Psi^\dagger \mathbf{X}$ .

This prior is fully separable with respect to the spectral channels, similarly to the data-fidelity term in (3). In this setting, the wide-band objective function naturally separates into  $L$  independent, single-channel objective functions underpinning the monochromatic SARA approach defined by Carrillo et al. (2012); Onose et al. (2016b). This approach is therefore highly parallelizable, and will be taken as a reference in terms of computing time.

In practice, each single-channel term in the objective (3) is solved with a reweighting approach leveraging PDFB. The splitting functionality of this advanced proximal algorithm structure is in particular utilized to enable the parallel processing of the block-specific data-fidelity terms for scalability. Note that the parameter  $\tilde{\mu}$  does not affect the minimizers of the objective. Onose et al. (2016b) suggested that setting  $\tilde{\mu}$  in the range  $[10^{-5}, 10^{-3}]$  provides good convergence speed in practice.

### 2.3.2 HyperSARA prior

Unlike the SARA prior, the recently proposed HyperSARA prior (Abdulaziz et al. 2019b) aims to explicitly promote spectral correlations. In particular, it promotes low-rankness of  $\mathbf{X}$  resulting from the correlation of its channels, as well as average spatial sparsity over all frequency channels. Specif-

ically, the log-sum prior of interest is of the form

$$r(\mathbf{X}) = \tilde{\mu} \sum_{j=1}^J \log (|\sigma_j(\mathbf{X})| + \nu) + \mu \sum_{i=1}^I \log (\|\Psi^\dagger \mathbf{X}\|_2 + \nu), \quad (5)$$

where  $(\tilde{\mu}, \mu, \nu) \in ]0, +\infty[^3$  are regularization parameters,  $J \leq \min\{N, L\}$  is the rank of  $\mathbf{X}$ ,  $(\sigma_j(\mathbf{X}))_{1 \leq j \leq J}$  are the singular values of  $\mathbf{X}$ , and  $[\Psi^\dagger \mathbf{X}]_i$  denotes the  $i$ -th row of  $\Psi^\dagger \mathbf{X}$ .

HyperSARA has been shown to produce image cubes with superior quality when compared to SARA and the wide-band CLEAN-based approach JC-CLEAN (Abdulaziz et al. 2019b). The fundamental reason for this is that the number of degrees of freedom to be reconstructed when adding frequency channels increases more slowly than the amount of data due to the correlation between the channels. Also, because the magnitude of the spatial frequency probed by an antenna pair is proportional to the observation frequency, the  $uv$ -coverage at a higher frequency channel is a dilated version of the  $uv$ -coverage at a lower frequency, with the dilation parameter between two channels given by the frequency ratio. Consequently, the data at higher frequency channels provide higher spatial frequency information for the lower frequency channels, thus contributing to better precision of the image reconstruction process, both in terms of resolution and dynamic range. HyperSARA will thus be taken as a reference in terms of imaging quality in Section 4.

In practice, the wide-band objective (3) is also solved with a reweighting procedure relying on PDFB, with the splitting functionality again utilized to enable the parallel processing of the block-specific data-fidelity terms for scalability. The regularization terms at the core of HyperSARA are however not separable, with the full image cube modeled as a single variable. This entails memory and computing requirements scaling with the size of the full image cube. The gist of the present contribution is to address this bottleneck by introducing spatio-spectral facets. Note that Abdulaziz et al. (2019b) have shown that the regularization parameters can be set as  $\tilde{\mu} = 1$ , and  $\mu = \|\mathbf{X}^{\text{dirty}}\|_* / \|\Psi^\dagger \mathbf{X}^{\text{dirty}}\|_{2,1}$ , where  $\mathbf{X}^{\text{dirty}}$  denotes the dirty image cube. Again, in theory, only the ratio of these parameters affects the minimizers of the objective.

## 2.4 Faceting and Faceted HyperSARA prior

The proposed Faceted HyperSARA prior builds on the HyperSARA prior, distributing both the average sparsity and the low-rankness prior over multiple spatio-spectral facets to alleviate the computing and storage requirements inherent to HyperSARA. In particular, we propose to decompose the 3D image cube into  $Q \times C$  spatio-spectral facets, as illustrated in Figure 2 and detailed below.

### 2.4.1 Spectral faceting

The wide-band image cube can first be decomposed into separate image sub-cubes composed of a subset of the frequency channels, with a separate prior for each sub-cube. Since the data-fidelity terms are channel-specific, the overall objective function (3) reduces to the sum of independent objectives for each sub-cube. The smaller-size wide-band imaging sub-problems (smaller data sets, and smaller image volumes)

can thus be solved independently in parallel, offering scalability. Taken to the extreme, this simple spectral faceting can be used to separate all channels and proceed with single-channel reconstructions (leading to SARA), however simultaneously losing completely the advantage of correlations between frames to improve image precision. The key point is to keep an appropriate number of frames per sub-cube in order to optimally take advantage of this correlation. Also, given the data at higher frequency channels provide higher spatial frequency information for the lower frequency channels, it is of critical importance that the whole extent of the frequency band of observation be exploited in each channel reconstruction. In this context, we propose to decompose the cube into channel-interleaved spectral sub-cubes, each of which results from a uniform sub-sampling of the whole frequency band (see Figure 2 (b)). We thus decompose the original inverse problem (1) into  $C$  independent, channel-interleaved sub-problems, each considering  $L_c$  channels from the original data cube, with  $L = L_1 + \dots + L_C$ . For each sub-cube  $c \in \{1, \dots, C\}$ ,  $\mathbf{y}_{c,l,b} \in \mathbb{C}^{M_{c,l,b}}$  denotes the vector of  $M_{c,l,b}$  visibilities associated with the channel  $l \in \{1, \dots, L_c\}$  and data-block  $b \in \{1, \dots, B\}$ , and by  $\Phi_{c,l,b}$  and  $\varepsilon_{c,l,b}$  the associated measurement operator and  $\ell_2$  ball radius, respectively. The initial minimization problem (3) is thus reformulated as

$$\underset{\mathbf{X} \in \mathbb{R}_+^{N \times L}}{\text{minimize}} \sum_{c=1}^C \left( \sum_{l=1}^{L_c} \sum_{b=1}^B \ell_{\mathcal{B}(\mathbf{y}_{c,l,b}, \varepsilon_{c,l,b})}(\Phi_{c,l,b} \mathbf{x}_{c,l}) + r_c(\mathbf{X}_c) \right), \quad (6)$$

where, for every  $c \in \{1, \dots, C\}$ ,  $\mathbf{X}_c = (\mathbf{x}_{c,l})_{1 \leq l \leq L_c} \in \mathbb{R}^{N \times L_c}$  is the  $c$ -th sub-cube of the full image cube  $\mathbf{X}$ , with  $\mathbf{x}_{c,l} \in \mathbb{R}^N$  the  $l$ -th image of the sub-cube  $\mathbf{X}_c$ , and  $r_c: \mathbb{R}^{N \times L_c} \rightarrow ]-\infty, +\infty]$  is a sub-part of the regularization term, only acting on the  $c$ -th sub-cube.

#### 2.4.2 Spatial faceting

Faceting can also be performed in the spatial domain by decomposing the regularization term for each spectral sub-cube into a sum of terms acting only locally in the spatial domain (see Figure 2 (c)). In this context, the resulting facets will need to overlap in order to avoid edge effects, so that the overall objective function (6) will take the form of the sum of inter-dependent facet-specific objectives. This inter-dependence precludes separating the imaging problem into facet problems. However, the splitting functionality of PDFB can be exploited to enable parallel processing of the facet-specific regularization terms and ensure further scalability (see Section 3).

On the one hand, we propose to split the average sparsity dictionary  $\Psi^\dagger$  into  $Q$  smaller wavelet decomposition, leveraging the wavelet splitting technique introduced in Pruša (2012, Chapter 4). Pruša (2012) proposed an exact implementation of the discrete wavelet transform distributed over multiple facets. In this context, the Daubechies wavelet bases are decomposed into a collection of facet-based operators  $\Psi_q^\dagger \in \mathbb{R}^{I_q \times N_q}$  acting only on the  $q$ -th facet of size  $N_q$ , with  $I = I_1 + \dots + I_Q$ . The overlap needed to ensure an exact faceted implementation of the wavelet transforms is composed of a number of pixels between  $15(2^s - 2)$  and  $15(2^s - 1)$

in each spatial direction (Pruša 2012, Section 4.1.4), with  $s$  being the level of decomposition. In practice, the overlap ensures that each facet contains all the information needed to compute the convolutions underlying the discrete wavelet transforms locally.

On the other hand, we consider a faceted low-rank prior enforced by the sum of nuclear norm priors on essentially the same overlapping facets as those introduced for the wavelet decomposition. This provides a more tractable alternative to the global low-rank prior encoded by the nuclear norm of HyperSARA. Unlike the wavelet decomposition, there is no equivalent faceted implementation of the eigen-value decomposition. To mitigate reconstruction artifacts possibly resulting from the faceting of the 3D image cube, for each facet  $q \in \{1, \dots, Q\}$ , of size  $\tilde{N}_q$ , we propose to introduce a diagonal matrix  $\mathbf{D}_q \in ]0, +\infty[^{\tilde{N}_q \times \tilde{N}_q}$  ensuring a smooth transition from the borders of one facet to its neighbours. A natural choice consists in down-weighting the contribution of pixels involved in multiple facets. A tapering window decaying in the overlapping regions is considered, while ensuring that the sum of all the weights associated with each pixel is equal to unity. In this work, we consider weights in the form of a 2D triangular apodization window as considered by Murya et al. (2017) (see Figure 2 (c)). The size of the overlap for this term is taken as an adjustable parameter of the Faceted HyperSARA approach to further promote local correlations. Its influence is investigated in Section 4. In practice, a larger overlap region than the one taken for the faceted wavelet transform is considered, taking advantage of the overlap already imposed by the faceted implementation of the wavelet decomposition and the associated  $\ell_{2,1}$  norm priors.

The spatial faceting procedure therefore results in splitting the original log-sum priors of HyperSARA in (5) into a sum of inter-dependent facet-specific log-sum priors, defining the Faceted HyperSARA prior:

$$r_c(\mathbf{X}_c) = \sum_{q=1}^Q \left( \bar{\mu}_c \sum_{j=1}^{J_{c,q}} \log(|\sigma_j(\mathbf{D}_q \tilde{\mathbf{S}}_q \mathbf{X}_c)| + \nu) + \mu_c \sum_{i=1}^{I_q} \log(\|\Psi_q^\dagger \mathbf{S}_q \mathbf{X}_c\|_i + \nu) \right). \quad (7)$$

In (7),  $(\bar{\mu}_c, \mu_c, \nu) \in ]0, +\infty[^3$  are regularization parameters, and, for every  $q \in \{1, \dots, Q\}$ ,  $J_{c,q} \leq \min(\tilde{N}_q, L_c)$  is the rank of  $\mathbf{D}_q \tilde{\mathbf{S}}_q \mathbf{X}_c$ , and  $\tilde{\mathbf{S}}_q \in \mathbb{R}^{\tilde{N}_q \times N}$  and  $\mathbf{S}_q \in \mathbb{R}^{N_q \times N}$  extract spatially overlapping spatio-spectral facets from the full image cube for the low-rankness prior and the average sparsity prior, respectively. These two operators only differ in the amount of overlapping pixels considered, which is defined as an adjustable parameter for  $\tilde{\mathbf{S}}_q$ , and prescribed by Pruša (2012) for  $\mathbf{S}_q$  (Figure 2). Each facet relies on a spatial decomposition of the image into non-overlapping tiles (see Figure 2 (b), delineated by dashed red lines), each overlapping with its top and left spatial neighbour. In the following, the overlapping regions will be referred to as the borders of a facet, in contrast with its underlying tile (see Figure 2). An edge facet, *i.e.* which does not admit a neighbour in one of the two spatial dimensions, has the same dimension as the underlying tile in the direction where it does not admit a neighbour (*e.g.* corner facets have the same dimension as the under-

lying tile). Note that HyperSARA corresponds to the case  $Q = C = 1$ .

The reweighting approach utilized to minimize the objective (6) with the log-sum priors (7) via convex relaxations powered by PDFB is described in Section 3. Crucially, the splitting functionality of PDFB will be exploited to enable parallel processing of these facet-specific priors, more specifically their convex relaxations. Note that the appropriate values for the parameters  $\bar{\mu}_c$  and  $\mu_c$  will be investigated via simulation and in relation to the corresponding parameter values for HyperSARA in Section 4.

### 3 FACETED HYPERSARA

The parallel algorithmic structure of Faceted HyperSARA is described in this section, leveraging PDFB within a reweighting approach to handle the log-sum priors. Implementation details are also discussed.

#### 3.1 Outer reweighting algorithm

To efficiently address the log-sum prior underpinning the Faceted HyperSARA prior, we resort to a majorize-minimize algorithm similar to the one proposed by (Candès et al. 2009), leading to a reweighting approach described in Algorithm 1.

At each iteration  $p \in \mathbb{N}$  of Algorithm 1, problem (6) is majorized at the local estimate  $\mathbf{X}^{(p)}$  by a convex approximation, and then minimized using a PDFB algorithm described in Algorithm 2 (see Algorithm 1 line 6). For each sub-cube  $c \in \{1, \dots, C\}$ , the convex approximated minimization problem is of the form

$$\begin{aligned} \underset{\mathbf{X}_c \in \mathbb{R}_+^{N \times L}}{\text{minimize}} \quad & \sum_{l=1}^{L_c} \sum_{b=1}^B \iota_{\mathcal{B}(\mathbf{y}_{c,l,b}, \boldsymbol{\varepsilon}_{c,l,b})}(\boldsymbol{\Phi}_{c,l,b} \mathbf{X}_c, l) \\ & + \tilde{r}_c(\mathbf{X}_c, \mathbf{X}_c^{(p)}), \end{aligned} \quad (8)$$

where  $\tilde{r}_c(\cdot, \mathbf{X}_c^{(p)})$  is a convex local majorant function of  $r_c$  at  $\mathbf{X}_c^{(p)}$ , corresponding to the weighted hybrid norm prior

$$\begin{aligned} \tilde{r}_c(\mathbf{X}_c, \mathbf{X}_c^{(p)}) = \sum_{q=1}^Q & \left( \bar{\mu}_c \|\mathbf{D}_q \tilde{\mathbf{S}}_q \mathbf{X}_c\|_{*, \bar{\omega}_q(\mathbf{X}_c^{(p)})} \right. \\ & \left. + \mu_c \|\boldsymbol{\Psi}_q^\dagger \mathbf{S}_q \mathbf{X}_c\|_{2,1, \omega_q(\mathbf{X}_c^{(p)})} \right), \end{aligned} \quad (9)$$

where, for every  $q \in \{1, \dots, Q\}$ , the weights  $\bar{\omega}_q(\mathbf{X}_c^{(p)}) = (\bar{\omega}_{q,j}(\mathbf{X}_c^{(p)}))_{1 \leq j \leq J_{c,q}}$  and  $\omega_q(\mathbf{X}_c^{(p)}) = (\omega_{q,i}(\mathbf{X}_c^{(p)}))_{1 \leq i \leq I_q}$  are given by

$$\bar{\omega}_{q,j}(\mathbf{X}_c^{(p)}) = \left( \sigma_j(\mathbf{D}_q \tilde{\mathbf{S}}_q \mathbf{X}_c^{(p)}) + \nu \right)^{-1}, \quad (10)$$

$$\omega_{q,i}(\mathbf{X}_c^{(p)}) = \left( \|\boldsymbol{\Psi}_q^\dagger \mathbf{S}_q \mathbf{X}_c^{(p)}\|_i + \nu \right)^{-1}. \quad (11)$$

At the beginning of the algorithm, the weights are initialized to one (see Algorithm 1 line 3, where the notation  $\mathbf{1}_{J_c}$  stands for the vector of size  $J_c$  with all coefficients equal to 1, and  $J_c = J_{c,1} + \dots + J_{c,Q}$ ). Note that the weights defined in (10)-(11) are multiplied by the regularization parameter

$\nu$  in Algorithm 1, which is equivalent to re-scaling the sub-problem (8) by  $\nu$ . This does not affect the set of minimizers of the global problem<sup>3</sup>.

A complete description of the PDFB-powered algorithm used to solve the sub-problems (8) is provided in the next section.

---

#### Algorithm 1: Outer reweighting algorithm.

---

**Input:**  $\mathbf{X}^{(0)} = (\mathbf{X}_c^{(0)})_c$ ,  $\mathbf{P}^{(0)} = (\mathbf{P}_c^{(0)})_c$ ,  $\mathbf{W}^{(0)} = (\mathbf{W}_c^{(0)})_c$ ,  
 $\mathbf{v}^{(0)} = (\mathbf{v}_c^{(0)})_c$

- 1  $p \leftarrow 0$ ;
- // Initialization of the weights
- 2 **for**  $c = 1$  **to**  $C$  **do**
- 3    $\boldsymbol{\theta}_c^{(0)} = (\boldsymbol{\theta}_{c,q}^{(0)})_{1 \leq q \leq Q} = \mathbf{1}_I$ ;  $\bar{\boldsymbol{\theta}}_c^{(0)} = (\bar{\boldsymbol{\theta}}_{c,q}^{(0)})_{1 \leq q \leq Q} = \mathbf{1}_{J_c}$ ;
- 4 **while** *stopping criterion not satisfied* **do**
- 5   // Solve spectral sub-problems in parallel
- 6   **for**  $c = 1$  **to**  $C$  **do**
- 7     // Run Algorithm 2
- 8      $(\mathbf{X}_c^{(p+1)}, \mathbf{P}_c^{(p+1)}, \mathbf{W}_c^{(p+1)}, \mathbf{v}_c^{(p+1)}) =$   
      **Algorithm2** $(\mathbf{X}_c^{(p)}, \mathbf{P}_c^{(p)}, \mathbf{W}_c^{(p)}, \mathbf{v}_c^{(p)}, \boldsymbol{\theta}_c^{(p)}, \bar{\boldsymbol{\theta}}_c^{(p)})$ ;
- 9     **for**  $q = 1$  **to**  $Q$  **do**
- 10      // Update weights: low-rankness prior
- 11       $\bar{\boldsymbol{\theta}}_{c,q}^{(p+1)} = \nu \bar{\omega}_q(\mathbf{X}_c^{(p+1)})$ ; // using (10)
- 12      // Update weights: joint-sparsity prior
- 13       $\boldsymbol{\theta}_{c,q}^{(p+1)} = \nu \omega_q(\mathbf{X}_c^{(p+1)})$ ; // using (11)
- 14      $p \leftarrow p + 1$ ;

**Result:**  $\mathbf{X}^{(p)}$ ,  $\mathbf{P}^{(p)}$ ,  $\mathbf{W}^{(p)}$ ,  $\mathbf{v}^{(p)}$

---

#### 3.2 Inner convex optimization algorithm

A primal-dual algorithmic structure as PDFB works by jointly solving the problem (8), referred to as the primal problem, and its dual formulation in the sense of the Fenchel-Rockafellar duality theory (Bauschke & Combettes 2017). The splitting functionality will enable all block-specific data-fidelity terms and facet-specific regularization terms to be updated in parallel via their proximal operator. In this work, we resort to a preconditioned variant of PDFB, which uses proximal operators with respect to non-Euclidean metrics in order to reduce the number of iterations necessary to converge. Let  $\mathbf{U} \in \mathbb{R}^{n \times n}$  be a symmetric, positive definite matrix. The proximal operator of a proper, convex, lower semi-continuous function  $f: \mathbb{R}^n \rightarrow ]-\infty, +\infty]$  at  $\mathbf{z} \in \mathbb{R}^n$  with respect to the metric induced by  $\mathbf{U}$  is defined by (Moreau 1965; Hiriart-Urruty & Lemaréchal 1993)

$$\text{prox}_f^{\mathbf{U}}(\mathbf{z}) = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} \left\{ f(\mathbf{x}) + \frac{1}{2}(\mathbf{x} - \mathbf{z})^\dagger \mathbf{U}(\mathbf{x} - \mathbf{z}) \right\}. \quad (12)$$

The more compact notation  $\text{prox}_f$  is used when  $\mathbf{U} = \mathbf{I}_n$ , where  $\mathbf{I}_n$  is the identity matrix in  $\mathbb{R}^{n \times n}$ . In addition, when the function  $f$  corresponds to an indicator function of a

<sup>3</sup> Previous works from Carrillo et al. (2012); Onose et al. (2016b); Abdulaziz et al. (2019b) suggest that the regularization parameter  $\nu$  in (10)-(11) should decrease from one iteration  $p$  to another by a factor 80% to improve the convergence rate and the stability of the algorithm. This procedure is also adopted in this article.

**Algorithm 2:** Inner convex optimisation algorithm for each spectral sub-problem (8), powered by PDFB.

---

**Data:**  $(\mathbf{y}_{c,l,b})_{l,b}$ ,  $l \in \{1, \dots, L_c\}$ ,  $b \in \{1, \dots, B\}$   
**Input:**  $\mathbf{X}_c^{(0)}$ ,  $\mathbf{P}_c^{(0)} = (\mathbf{P}_{c,q}^{(0)})_q$ ,  $\mathbf{W}_c^{(0)} = (\mathbf{W}_{c,q}^{(0)})_q$ ,  
 $\mathbf{v}_c^{(0)} = (\mathbf{v}_{c,l,b}^{(0)})_{c,l,b}$ ,  $\boldsymbol{\theta}_c = (\boldsymbol{\theta}_{c,q})_{1 \leq q \leq Q}$ ,  
 $\bar{\boldsymbol{\theta}}_c = (\bar{\boldsymbol{\theta}}_{c,q})_{1 \leq q \leq Q}$   
**Parameters:**  $(\mathbf{D}_{c,q})_q$ ,  $(\mathbf{U}_{c,l,b})_{l,b}$ ,  $(\boldsymbol{\varepsilon}_{c,l,b})_{l,b}$ ,  $\mu_c$ ,  $\bar{\mu}_c$ ,  $\tau$ ,  
 $\zeta$ ,  $\eta$ ,  $\kappa$

- 1  $k \leftarrow 0$ ;  $\xi \leftarrow +\infty$ ;  $\check{\mathbf{X}}_c^{(0)} = \mathbf{X}_c^{(0)}$ ;
- 2 **while**  $\xi > 10^{-5}$  **do**
  - 3 **// Broadcast auxiliary variables**
  - 4 **for**  $q = 1$  **to**  $Q$  **do**  
 $\check{\mathbf{X}}_{c,q}^{(k)} = \tilde{\mathbf{S}}_q \check{\mathbf{X}}_c^{(k)}$ ;  $\check{\mathbf{X}}_c^{(k)} = \mathbf{S}_q \check{\mathbf{X}}_c^{(k)}$ ;
  - 5 **for**  $l = 1$  **to**  $L_c$  **do**
  - 6  $\hat{\mathbf{x}}_{c,l}^{(k)} = \mathbf{FZ}\hat{\mathbf{x}}_{c,l}^{(k)}$ ; **// Fourier transforms**
  - 7 **for**  $b = 1$  **to**  $B$  **do**
  - 8  $\hat{\mathbf{x}}_{c,l,b}^{(k)} = \mathbf{M}_{c,l,b} \hat{\mathbf{x}}_{c,l}^{(k)}$ ; **// send to data cores**
  - 9 **// Update low-rankness variables [facet cores]**
  - 10 **for**  $q = 1$  **to**  $Q$  **do**  
 $\mathbf{P}_{c,q}^{(k+1)} = (\mathbf{I}_{I_q} - \text{prox}_{\zeta^{-1} \bar{\mu}_c \|\cdot\|_{\mathbf{S}, \bar{\boldsymbol{\theta}}_{c,q}}}) (\mathbf{P}_{c,q}^{(k)} + \mathbf{D}_{c,q} \check{\mathbf{X}}_c^{(k)})$ ;
  - 11  $\bar{\mathbf{P}}_{c,q}^{(k+1)} = \mathbf{D}_q^\dagger \mathbf{P}_{c,q}^{(k+1)}$ ;
  - 12 **// Update sparsity variables [facet cores]**
  - 13 **for**  $q = 1$  **to**  $Q$  **do**  
 $\mathbf{W}_{c,q}^{(k+1)} =$   
 $(\mathbf{I}_{I_q} - \text{prox}_{\kappa^{-1} \mu_c \|\cdot\|_{2,1, \boldsymbol{\theta}_{c,q}}}) (\mathbf{W}_{c,q}^{(k)} + \boldsymbol{\Psi}_q^\dagger \check{\mathbf{X}}_c^{(k)})$ ;
  - 14  $\bar{\mathbf{W}}_{c,q}^{(k+1)} = \boldsymbol{\Psi}_q \mathbf{W}_{c,q}^{(k+1)}$ ;
  - 15 **// Update data fidelity variables [data cores]**
  - 16 **for**  $(l, b) = (1, 1)$  **to**  $(L_c, B)$  **do**  
 $\mathbf{v}_{c,l,b}^{(k+1)} =$   
 $\mathbf{U}_{c,l,b} (\mathbf{M}_{c,l,b} - \text{prox}_{\mathcal{B}(\mathbf{y}_{c,l,b}, \boldsymbol{\varepsilon}_{c,l,b})}) (\mathbf{U}_{c,l,b}^{-1} \mathbf{v}_{c,l,b}^{(k)} +$   
 $\boldsymbol{\Theta}_{c,l,b} \mathbf{G}_{c,l,b} \hat{\mathbf{x}}_{c,l,b}^{(k)})$ ;
  - 17  $\bar{\mathbf{v}}_{c,l,b}^{(k+1)} = \mathbf{G}_{c,l,b}^\dagger \boldsymbol{\Theta}_{c,l,b}^\dagger \mathbf{v}_{c,l,b}^{(k+1)}$ ;
  - 18 **// Inter node communications**
  - 19 **for**  $l = 1$  **to**  $L_c$  **do**  
 $\mathbf{a}_{c,l}^{(k)} = \sum_{q=1}^Q (\zeta \bar{\mathbf{S}}_q^\dagger \bar{\mathbf{v}}_{c,q,l}^{(k+1)} + \kappa \bar{\mathbf{S}}_q^\dagger \bar{\mathbf{w}}_{c,q,l}^{(k+1)}) +$   
 $\eta \mathbf{Z}^\dagger \mathbf{F}^\dagger \sum_b \mathbf{M}_{c,l,b}^\dagger \bar{\mathbf{v}}_{c,l,b}^{(k+1)}$ ;
  - 20 **// Update image tiles [on facet cores, in parallel]**
  - 21  $\mathbf{X}_c^{(k+1)} = \text{prox}_{\mathbb{R}_+^{N \times L_c}} (\mathbf{X}_c^{(k)} - \tau \mathbf{A}_c^{(k)})$ ;
  - 22  $\check{\mathbf{X}}_c^{(k)} = 2\mathbf{X}_c^{(k+1)} - \mathbf{X}_c^{(k)}$ ; **// communicate facet borders**
  - 23  $\xi = \|\mathbf{X}_c^{(k+1)} - \mathbf{X}_c^{(k)}\|_{\mathbf{F}} / \|\mathbf{X}_c^{(k)}\|_{\mathbf{F}}$ ;
  - 24  $k \leftarrow k + 1$ ;

**Result:**  $\mathbf{X}_c^{(k)}$ ,  $\mathbf{P}_c^{(k)}$ ,  $\mathbf{W}_c^{(k)}$ ,  $\mathbf{v}_c^{(k)}$

---

closed, non-empty, convex set, then its proximal operator reduces to the projection operator onto this set.

A graphical illustration of the PDFB-powered algorithm to solve problem (8) is given in Figure 3. A formal description is reported in Algorithm 2. First, the faceted low-rankness prior is handled in lines 9-11 by computing in parallel the proximal operator of the per facet weighted nuclear

**Table 1.** proximal operators involved in Algorithm 2.

proximal operator for $\alpha > 0$	Details
$\text{prox}_{\alpha \ \cdot\ _{\mathbf{S}}}(\mathbf{Z}) = \mathbf{U} \text{Diag}(\text{prox}_{\alpha \ \cdot\ _{\mathbf{S}}}(\boldsymbol{\sigma})) \mathbf{V}^\dagger$	$\mathbf{Z} = \mathbf{U} \boldsymbol{\Sigma} \mathbf{V}^\dagger \in \mathbb{R}^{N \times L}$ , $\boldsymbol{\Sigma} = \text{Diag}(\boldsymbol{\sigma})$ singular value decomposition of $\mathbf{Z}$
$\text{prox}_{\alpha \ \cdot\ _{2,1}}(\mathbf{Z}) = \left( \max\{\ \mathbf{z}_n^\dagger\ _2 - \alpha, 0\} \frac{\mathbf{z}_n^\dagger}{\ \mathbf{z}_n^\dagger\ _2} \right)_{1 \leq n \leq N}$	$\mathbf{Z} = [\mathbf{z}_1, \dots, \mathbf{z}_N]^\dagger \in \mathbb{R}^{N \times L}$
$\text{prox}_{\mathcal{B}(\mathbf{y}_{c,l,b}, \boldsymbol{\varepsilon}_{c,l,b})}(\mathbf{z}) = \boldsymbol{\varepsilon}_{c,l,b} \frac{\mathbf{z} - \mathbf{y}_{c,l,b}}{\ \mathbf{z} - \mathbf{y}_{c,l,b}\ _2} + \mathbf{y}_{c,l,b}$	$\mathbf{z} \in \mathbb{C}^{M_{c,l,b}}$
$\text{prox}_{\mathbb{R}_+^{N \times L}}(\mathbf{Z}) = \max\{\mathbf{0}, \Re(\mathbf{Z})\}$	$\mathbf{Z} \in \mathbb{C}^{N \times L}$ , $\Re(\mathbf{Z})$ denotes the real part of $\mathbf{Z}$

norms (see Table 1). Second, the average sparsity prior is addressed in lines 12-14 by computing the proximal operator of the per facet weighted  $\ell_{2,1}$  norm in parallel (see Table 1). Third, the data-fidelity terms are handled in parallel in lines 15-17 by computing, for every data block  $(c, l, b)$ , the projection onto the  $\ell_2$  balls  $\mathcal{B}(\mathbf{y}_{c,l,b}, \boldsymbol{\varepsilon}_{c,l,b})$  with respect to the metric induced by the diagonal matrices  $\mathbf{U}_{c,l,b}$ , chosen using the preconditioning strategy proposed by Onose et al. (2017); Abdulaziz et al. (2019b). More precisely, their diagonal coefficients are the inverse of the sampling density in the vicinity of the probed Fourier modes. The projections onto the  $\ell_2$  balls for the metric induced by  $\mathbf{U}_{c,l,b}$  do not admit an analytic expression, and thus need to be approximated numerically through sub-iterations. In this work, we resort to FISTA (Beck & Teboulle 2009), which iteratively approximates this projection by computing Euclidean projections onto the  $\ell_2$  balls  $\mathcal{B}(\mathbf{y}_{c,l,b}, \boldsymbol{\varepsilon}_{c,l,b})$  (see Table 1). Finally, the non-negativity constraint is handled in line 21 by computing the Euclidean projection onto the non-negative orthant (see Table 1).

Algorithm 2 is guaranteed to converge to a global solution to problem (8), for a given sub-cube  $c \in \{1, \dots, C\}$ , provided that the preconditioning matrices  $(\mathbf{U}_{c,l,b})_{c,l,b}$  and the parameters  $(\tau, \zeta, \eta, \kappa)$  satisfy technical conditions described in (Pesquet & Repetti 2015, Lemma 4.3). In particular, these conditions are satisfied for our choice of parameters:  $\tau = 1/3$ ,  $\zeta = 1$ ,  $\eta = 1/\|\mathbf{U}_c^{1/2} \boldsymbol{\Phi}_c\|_{\mathbf{S}}^2$ ,  $\kappa = 1/\|\boldsymbol{\Psi}^\dagger\|_{\mathbf{S}}^2$ , where  $\|\cdot\|_{\mathbf{S}}$  denotes the spectral norm of a linear operator.

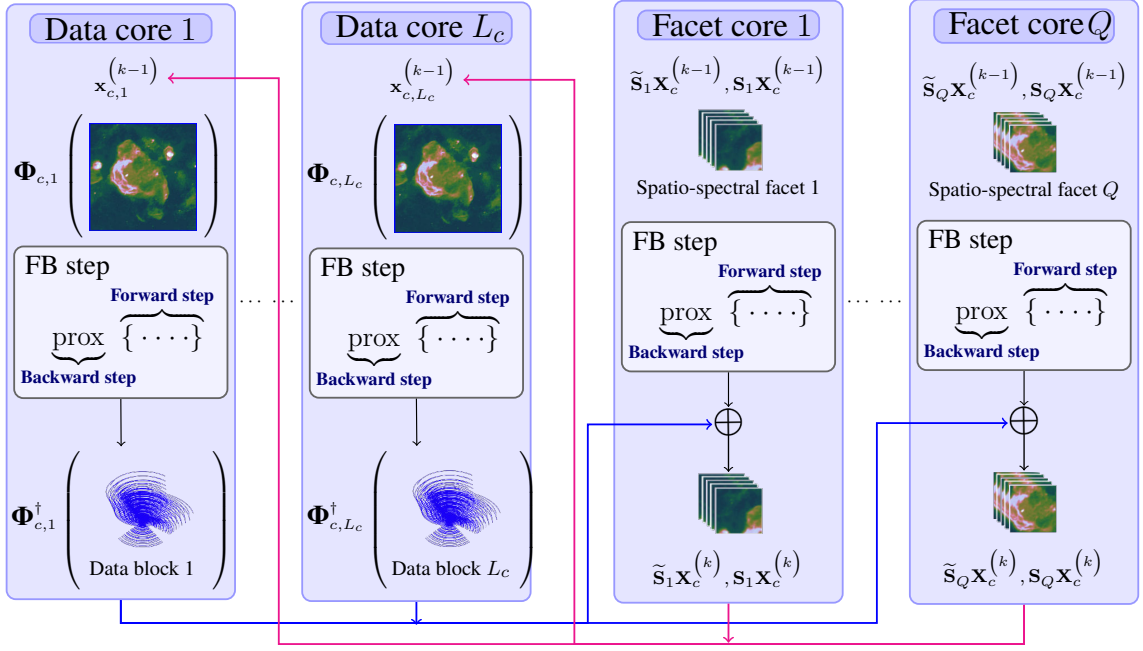
Note that PDFB can accommodate randomization in the update of the variables, *e.g.* by randomly selecting a subset of the data and facet dual variables to be updated at each iteration. This procedure can significantly alleviate the memory load per node (Pesquet & Repetti 2015) at the expense of an increased number of iterations for the algorithm to converge. This feature, which has been specifically investigated for wide-band imaging (Abdulaziz et al. 2017) and facet-based monochromatic imaging (Naghizadeh et al. 2018), is not leveraged in the implementation of Algorithm 2 used for the experiments reported in Sections 4, 5 and 6.

### 3.3 Implementation

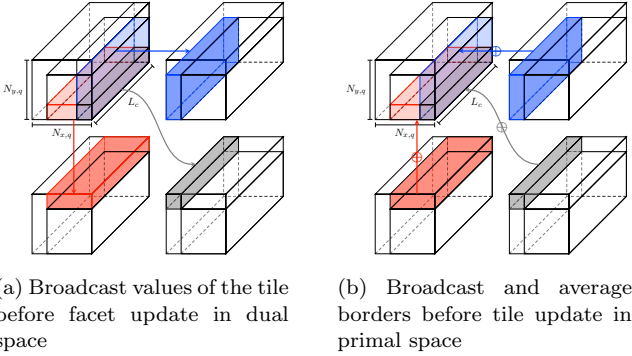
To solve a spectral sub-problem  $c \in \{1, \dots, C\}$ , different parallelization strategies can be adopted, depending on the computing resources available and the size of the problem to be addressed. We propose to divide the variables to be estimated into the two following groups of computing cores.

- *Data cores:* Each core involved in this group is responsible for the update of several dual variables  $\mathbf{v}_{c,l,b} \in \mathbb{C}^{M_{c,l,b}}$  associated with the data-fidelity terms (see Algorithm 2 line 16). These cores produce auxiliary variables  $\bar{\mathbf{v}}_{c,l,b} \in \mathbb{R}^N$





**Figure 3.** Illustration of the two groups of cores described in Section 3, with the main steps involved in Algorithm 2 applied to each independent sub-problem  $c \in \{1, \dots, C\}$ , using  $Q$  facets (along the spatial dimension) and  $B = 1$  data block per channel. Data cores handle variables of the size of data blocks (Algorithm 2 lines 15–17), whereas facet cores handle variables of the size of a spatio-spectral facet (Algorithm 2 lines 9–14), respectively. Communications between the two groups are represented by colored arrows. Communications between facet cores, induced by the overlap between the spatio-spectral facets, are illustrated in Figure 4.



**Figure 4.** Illustration of the communication steps involving a facet core (represented by the top-left rectangle in each sub-figure) and a maximum of three of its neighbours. The tile underpinning each facet, located in its bottom-right corner, is delineated in thick black lines. At each iteration, the following two steps are performed sequentially. (a) Facet borders need to be completed before each facet is updated independently in the dual space (Algorithm 2 lines 9–14): values of the tile of each facet (top left) are broadcast to cores handling the neighbouring facets in order to update their borders (Algorithm 2 line 4). (b) Parts of the facet tiles overlapping with borders of nearby facets need to be updated before each tile is updated independently in the primal space (Algorithm 2 line 20): values of the parts of the borders overlapping with the tile of each facet are broadcast by the cores handling neighbouring facets, and averaged.

of single channel image size, each assumed to be held in the memory of a single core (line 17). Note that the Fourier transform computed for each channel  $l$  in line 6 is performed once per iteration on the data core  $(l, 1)$ . Each data core

$(l, b)$ , with  $b \in \{2, \dots, B\}$ , receives only a few coefficients of the Fourier transform of  $\mathbf{x}_l$  from the data core  $(l, 1)$ , selected by the operator  $\mathbf{M}_{c,l,b}$  (line 8);

- *Facet cores:* Each worker involved in this group, composed of  $Q$  cores, is responsible for the update of an image tile (*i.e.* a portion of the primal variable) and the dual variables  $\mathbf{P}_{c,q}$  and  $\mathbf{W}_{c,q}$  associated with the low-rankness and the joint average sparsity prior respectively (Algorithm 2, lines 10 and 13). Note that the image cube is stored across different facet cores, which are responsible for updating their image tiles (line 21). Since the facets underlying the proposed prior overlap, communications involving a maximum of 4 contiguous facet cores are needed to build the facet borders prior to updating the facets independently in the dual space (Algorithm 2 lines 9–14). Values of the tile of each facet are broadcast to cores handling neighbouring facets in order to update their borders (Algorithm 2 line 4, see Figure 4(a)). In a second step, parts of the facet tiles overlapping with borders of nearby facets need to be updated before each tile is updated independently in the primal space (Algorithm 2 line 20). More precisely, values of the parts of the borders overlapping with the tile of each facet are broadcast by the workers handling neighbouring facets, and averaged (see Figure 4(b)).

A MATLAB implementation of Algorithms 1 and 2 is available on the [Puri-Psi](#) webpage. Both HyperSARA and Faceted HyperSARA rely on MPI-like MATLAB parallelization features based on the `spmd` MATLAB function, based on composite MATLAB variables to handle parameters distributed across several cores (*e.g.* for the wide-band image cube). In practice, 1 process running either on 1 CPU core (physical core) or one hypertext (logical core) specifically

ensures communication synchronization between the data and facet cores. In the following, this process will be referred to as the master process, hosted on a CPU core referred to as the master CPU core.

## 4 VALIDATION ON SYNTHETIC DATA

In this section, the impact of spatial faceting is first assessed in terms of both reconstruction quality and computing time for a single spectral sub-problem, using a varying number of facets and a varying size of the overlapping regions. The impact of spectral faceting on the reconstruction performance of Faceted HyperSARA is then quantified for a single underlying facet along the spatial dimension ( $Q = 1$ ). Results are compared with those of both SARA and HyperSARA.

### 4.1 Simulation setting

#### 4.1.1 Images and data

Following the procedure described by [Abdulaziz et al. \(2019b\)](#), a wide-band model image composed of  $L$  spectral channels is simulated from an image of the W28 supernova remnant of size  $N$ , considering  $B = 1$  data block per channel. The measurement operator relies on a realistic VLA  $uv$ -coverage, generated within the frequency range  $[\nu_1, \nu_L] = [1, 2]$  GHz with uniformly sampled channels and a total observation time of 6 hours. Note that the  $uv$ -coverage associated with each channel  $l$  corresponds to the reference  $uv$ -coverage at the frequency  $\nu_1$  scaled by the factor  $\nu_l/\nu_1$ . The data are corrupted by an additive, zero-mean complex white Gaussian noise of variance  $\sigma^2$ . An input signal-to-noise ratio (iSNR) of 60 dB is considered, which is defined as

$$\text{iSNR} = 10 \log_{10} \left( \frac{\sum_l \|\Phi_l \bar{x}_l\|_2^2 / M_l}{L \sigma^2} \right).$$

Note that, given the larger computational cost of HyperSARA, the size of the data is chosen so that it can be run in a reasonable amount of time for the different simulation scenarios described below.

#### 4.1.2 Spatial faceting

The performance of Faceted HyperSARA is first evaluated with  $C = 1$  (number of facets along the spectral dimension) for different parameters of the spatial faceting. Data generated from a  $N = 1024 \times 1024$  image composed of  $L = 20$  channels are considered, with  $M_l = 0.5N$  measurements per channel. The assessment is conducted with (i) varying  $Q$  (number of facets along the spatial dimensions) and a fixed overlap; (ii) a fixed number of facets and a varying spatial overlap for the nuclear norm regularization. Additional details can be found in the following lines. Regarding the choice of the regularization parameters, we set  $\bar{\mu} = 10^{-3}$  for SARA as explained in Section 2.3.1. As prescribed in Section 2.3.2, the regularization parameters of HyperSARA are set as  $\bar{\mu} = 1$ , and  $\mu = \|\mathbf{X}^{\text{dirty}}\|_* / \|\Psi^\dagger \mathbf{X}^{\text{dirty}}\|_{2,1} = 10^{-3}$ . For Faceted HyperSARA, we have observed that setting  $\bar{\mu}_c = 1$  and  $\mu_c = 10^{-2} \|\mathbf{X}_c^{\text{dirty}}\|_* / \|\Psi_c^\dagger \mathbf{X}_c^{\text{dirty}}\|_{2,1} = 10^{-5}$  leads

to a good trade-off to recover high resolution, high dynamic range model cubes.

- *Varying overlap*: Reconstruction performance and computing time are evaluated with  $C = 1$  and  $Q = 16$  (4 facets along each spatial dimension) and a varying size of the overlapping region for the faceted nuclear norm (0%, 6%, 20%, 33% and 50% of the spatial size of the facet, corresponding to 0, 16, 64, 128 and 256 pixels respectively) in each of the two spatial dimensions. Note that the overlap for the  $\ell_{2,1}$  prior is a fixed parameter ([Pruša 2012](#)). The comparison is conducted between SARA, HyperSARA, and Faceted HyperSARA.

- *Varying number of facets*: The reconstruction performance and computing time of Faceted HyperSARA are reported for experiments with  $Q \in \{4, 9, 16\}$  (corresponding to 2, 3 and 4 facets along each spatial dimension) with a fixed overlap corresponding to 50% of the spatial size of a facet. The regularization parameters are set to the same values as those considered in the experiment with a varying overlap.

#### 4.1.3 Spectral faceting

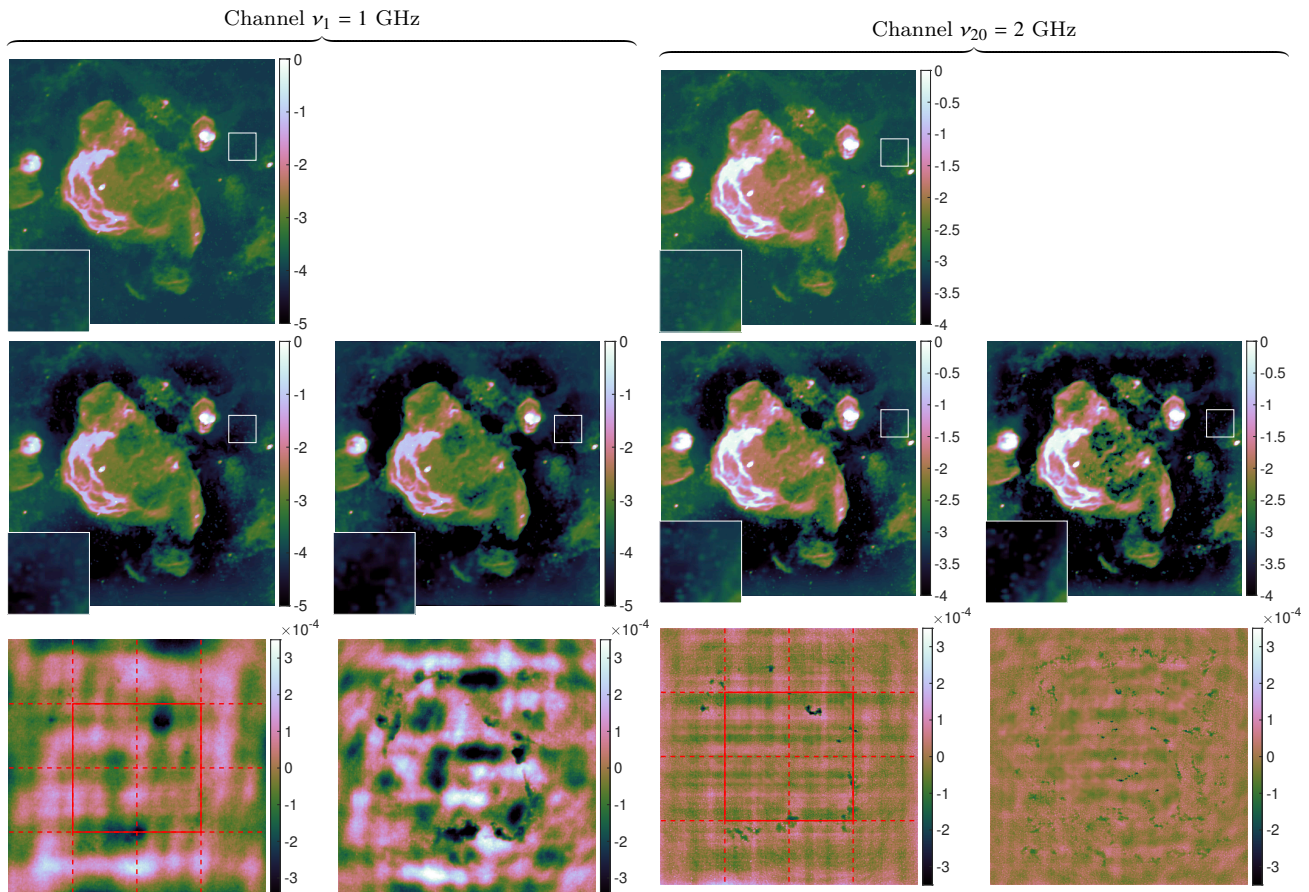
The influence of spectral faceting is evaluated in terms of computing time and reconstruction quality from data generated with a ground truth image composed of  $N = 256 \times 256$  pixels in  $L = 100$  channels, with  $M_l = N$  measurements per channel. The overall reconstruction performance of SARA, HyperSARA and Faceted HyperSARA with a single facet along the spatial dimension ( $Q = 1$ ) is compared. For faceted HyperSARA, a channel-interleaving process with a varying number of facets along the spectral dimension  $C$  is considered (see Section 2.4 and Figure 2 (b)). The simulation scenario involves facets composed of a varying number of channels  $L_c$  ( $L_c \approx 6, 10, 14, 20, 33$  and 50 channels for each sub-problem  $c \in \{1, \dots, C\}$ ) obtained by down-sampling the data cube along the frequency dimension. For the choice of the regularization parameters, we set  $\bar{\mu} = 10^{-2}$  for SARA. Our simulations indeed show that increasing the value beyond the range suggested in [Onose et al. \(2016b\)](#) provides better convergence speed. As prescribed in Section 2.3.2, the regularization parameters of HyperSARA are set as  $\bar{\mu} = 1$ , and  $\mu = \|\mathbf{X}^{\text{dirty}}\|_* / \|\Psi^\dagger \mathbf{X}^{\text{dirty}}\|_{2,1} = 10^{-2}$ . Similarly for Faceted HyperSARA, we set  $\bar{\mu}_c = 1$  and  $\mu_c = \|\mathbf{X}_c^{\text{dirty}}\|_* / \|\Psi_c^\dagger \mathbf{X}_c^{\text{dirty}}\|_{2,1} = 10^{-2}$ .

## 4.2 Hardware

All the methods compared in this section have been run on multiple compute nodes of Cirrus, one of the UK's Tier2 HPC services<sup>4</sup>. Cirrus is an SGI ICE XA system composed of 280 compute nodes, each with two 2.1 GHz, 18-core, Intel Xeon E5-2695 (Broadwell) series processors. The compute nodes have 256 GB of memory shared between the two processors. The system has a single Infiniband FDR network connecting nodes with a bandwidth of 54.5 GB/s.

The different methods have been applied in the following setting. For all the experiments, SARA uses 12 CPU

<sup>4</sup> <https://epsrsc.ukri.org/research/facilities/hpc/tier2/>



**Figure 5.** Spatial faceting analysis for synthetic data: reconstructed images (in Jy/pixel) reported in  $\log_{10}$  scale for channels  $\nu_1 = 1$  GHz (first two columns) and  $\nu_{20} = 2$  GHz (last two columns) for Faceted HyperSARA with  $Q = 16$  and  $C = 1$  (columns 1 and 3), and HyperSARA (*i.e.* Faceted HyperSARA with  $Q = C = 1$ , in column 2 and 4). From top to bottom are reported the ground truth image, the reconstructed and residual images. The overlap for the faceted nuclear norm regularization corresponds to 50% of the spatial size of a facet. The non-overlapping tiles underlying the definition of the facets are delineated on the residual images in red dotted lines, with the central facet displayed in continuous lines.

	Time (h)	aSNR (dB)	aSNR <sub>log</sub> (dB)	CPU cores
SARA	5.89	32.78 ( $\pm 2.76$ )	<b>-1.74</b> ( $\pm 0.83$ )	240
HyperSARA	133.1	38.63 ( $\pm 0.23$ )	<b>-0.39</b> ( $\pm 0.95$ )	22
Faceted nuclear norm overlap (0%)	26.26	37.03 ( $\pm 2.90 \cdot 10^{-3}$ )	<b>5.09</b> ( $\pm 1.09$ )	36
Faceted nuclear norm overlap (6%)	18.01	37.01 ( $\pm 1.00 \cdot 10^{-3}$ )	<b>4.09</b> ( $\pm 0.99$ )	36
Faceted nuclear norm overlap (20%)	18.11	36.86 ( $\pm 0.90 \cdot 10^{-3}$ )	<b>4.51</b> ( $\pm 1.07$ )	36
Faceted nuclear norm overlap (33%)	17.94	36.98 ( $\pm 1.60 \cdot 10^{-3}$ )	<b>6.00</b> ( $\pm 1.05$ )	36
Faceted nuclear norm overlap (50%)	20.75	37.08 ( $\pm 1.60 \cdot 10^{-3}$ )	<b>7.88</b> ( $\pm 0.91$ )	36

**Table 2.** Spatial faceting experiment: varying size of the overlap region for the faceted nuclear norm regularization. Reconstruction performance of Faceted HyperSARA with  $Q = 16$  and  $C = 1$ , compared to HyperSARA (*i.e.* Faceted HyperSARA with  $Q = C = 1$ ) and SARA. The results are reported in terms of reconstruction time, aSNR and aSNR<sub>log</sub> (both in dB with the associated standard deviation), and total number of CPU cores used to reconstruct the full image. The evolution of the aSNR<sub>log</sub>, of specific interest for this experiment, is highlighted in bold face.

cores to reconstruct each single channel, based on the parallelization strategy proposed by Onose et al. (2016b): 1 master CPU core, 2 CPU cores for the data-fidelity terms and 9 CPU cores to handle the average sparsity terms (associated with the nine bases of the SARA dictionary). HyperSARA and Faceted HyperSARA have been applied in the following configuration, given the different number of visibilities considered in the two simulation scenarios.

- *Spatial faceting*: HyperSARA addresses the full problem (3)-(5) with 22 CPU cores: 1 master CPU core, 20 CPU cores for the data fidelity terms (1 CPU core per data term), and 1 CPU core for the regularization term. To address each sub-problem in (6), Faceted HyperSARA uses 20 CPU cores for the data fidelity terms (1 CPU core per data channel), and 1 CPU core for each of the  $Q$  facets. The master process runs on one of the hyperthreads of the node (logical core) to ensure communication synchronizations.

	Time (h)	aSNR (dB)	aSNR <sub>log</sub> (dB)	CPU cores
SARA	<b>6.23</b>	32.78 ( $\pm 2.76$ )	-1.74 ( $\pm 0.83$ )	240
HyperSARA	<b>133.08</b>	38.63 ( $\pm 0.23$ )	-0.39 ( $\pm 0.95$ )	22
Faceted HyperSARA ( $Q = 4$ )	<b>42.04</b>	36.58 ( $\pm 1.80 \cdot 10^{-3}$ )	10.19 ( $\pm 0.88$ )	24
Faceted HyperSARA ( $Q = 9$ )	<b>21.60</b>	37.00 ( $\pm 1.70 \cdot 10^{-3}$ )	5.88 ( $\pm 1.00$ )	29
Faceted HyperSARA ( $Q = 16$ )	<b>17.94</b>	37.08 ( $\pm 1.60 \cdot 10^{-3}$ )	7.88 ( $\pm 1.05$ )	36

**Table 3.** Spatial faceting experiment: varying number of facets along the spatial dimension  $Q$ . Reconstruction performance of Faceted HyperSARA ( $C = 1$ , overlap of 50%), compared to HyperSARA (*i.e.* Faceted HyperSARA with  $Q = C = 1$ ) and SARA. The results are reported in terms of reconstruction time, aSNR and aSNR<sub>log</sub> (both in dB with the associated standard deviation), and total number of CPU cores used to reconstruct the full image. The evolution of the computing time, of specific interest for this experiment, is highlighted in bold face.

- *Spectral faceting:* HyperSARA addresses the full problem (3)-(5) with 7 CPU cores: 1 master CPU core, 5 CPU cores for the data fidelity terms (20 data-fidelity terms handled by each core), and 1 CPU core for the regularization term. To address each sub-problem in (6), Faceted HyperSARA uses 1 master CPU core, 5 CPU cores for the data fidelity terms, and 1 CPU core per facet ( $Q$  facets in total).

Note that for each experiment, the number of cores assigned to each group of cores in Faceted HyperSARA (*i.e.* data and facet cores) has been chosen to ensure a reasonable balance between the different computing tasks.

### 4.3 Evaluation metrics

Performance is evaluated in terms of global computing time (elapsed real time) and reconstruction SNR, defined for each channel  $l \in \{1, \dots, L\}$  as

$$\text{SNR}_l(\mathbf{x}_l) = 20 \log_{10} \left( \frac{\|\bar{\mathbf{x}}_l\|_2}{\|\bar{\mathbf{x}}_l - \mathbf{x}\|_2} \right).$$

Results are reported in terms of the average SNR (aSNR)

$$\text{aSNR}(\mathbf{X}) = \frac{1}{L} \sum_{l=1}^L \text{SNR}_l(\mathbf{x}_l).$$

Since the above criterion shows limitations to reflect the dynamic range and thus appreciate improvements in the quality of faint emissions, the following criterion is computed over images in log<sub>10</sub> scale

$$\text{SNR}_{\log, l}(\mathbf{x}_l) = 20 \log_{10} \left( \frac{\|\log_{10}(\bar{\mathbf{x}}_l + \epsilon \mathbf{1}_N)\|_2}{\|\log_{10}(\bar{\mathbf{x}}_l + \epsilon \mathbf{1}_N) - \log_{10}(\mathbf{x} + \epsilon \mathbf{1}_N)\|_2} \right),$$

where the log<sub>10</sub> function is applied term-wise, and  $\epsilon$  is an arbitrarily small parameter to avoid numerical issues ( $\epsilon$  is set to machine precision). Results are similarly reported in terms of the average log-SNR, defined as  $\text{aSNR}_{\log}(\mathbf{X}) = \frac{1}{L} \sum_{l=1}^L \text{SNR}_{\log, l}(\mathbf{x}_l)$ .

## 4.4 Results and discussion

### 4.4.1 Spatial faceting

- *Varying spatial overlap:* The results reported in Table 2 show that spatial faceting gives a good reconstruction of high intensity pixels (reflected by an aSNR close to HyperSARA). Even if the performance of the proposed approach does not vary much in terms of aSNR as the overlap for the faceted nuclear norm increases, the aSNR<sub>log</sub> improves significantly.

This reflects the ability of the proposed prior to enhance the estimation of faint emissions and finer details by promoting local correlations. This observation is further confirmed by the reconstructed images, reported in Jy/pixel in Figure 5 for the channels  $\nu_1 = 1$  GHz and  $\nu_L = 2$  GHz, showing that Faceted HyperSARA reconstructs images with a higher dynamic range (see the zoomed region delineated in white in Figure 5). The associated residual images (last row of Figure 5) are comparable to or better than HyperSARA. Note that the regular patterns observed on the residual images do not result from the faceting, as they are not aligned with the facet borders and appear for both approaches. From a computational point of view, Table 2 shows that increasing the overlap size results in a moderate increase in the computing time. Overall, an overlap of 50% gives the best reconstruction SNR for a reasonable computing time, and will thus be considered as a default faceting setting for the real data experiments reported in Sections 5 and 6.

- *Varying number of facets  $Q$  along the spatial dimension:* The reconstruction performance and computing time reported in Table 3 show that Faceted HyperSARA gives an almost constant reconstruction performance as the number of facets increases, for an overall computing time getting closer to the SARA approach. The dynamic range of the reconstructed images is notably higher for the Faceted approach, as indicated by the aSNR<sub>log</sub> values reported in Table 3. These results confirm the potential of the proposed approach to scale to large image sizes by increasing the number of facets along the spatial dimensions, while ensuring a stable reconstruction level as the number of facets increases. In particular, the setting  $Q = 16$  is reported to ensure a satisfactory reconstruction performance for a significantly reduced computing time.

In both experiments, Faceted HyperSARA has a much lower SNR standard deviation than HyperSARA and SARA (see Tables 2 and 3), *i.e.* ensures a more stable recovery quality across channels. This results from the stronger spatio-spectral correlations induced by the proposed faceted regularization, in comparison with both the HyperSARA and SARA priors.

### 4.4.2 Spectral faceting

The results reported in Table 4 show that Faceted HyperSARA using channel-interleaved facets retains most of the overall reconstruction performance of HyperSARA, ensuring a reconstruction quality significantly better than SARA. As expected, the reconstruction quality of faint emissions,

reflected by the  $\text{aSNR}_{\log}$  values, gradually decreases as fewer channels are involved in each facet (*i.e.* as  $C$  increases). This observation is qualitatively confirmed by the images reported in Figure 6 (in Jy/pixel) for facets composed of 10 channels each (see the zoomed regions in Figure 6). The slight loss of dynamic range is likely due to the reduction in the amount of data per spectral sub-cube. Spectral faceting remains however computationally attractive, in that it preserves the overall imaging quality of HyperSARA up to an already significant amount of interleaving (see discussion in Section 2.4.1), while allowing lower-dimension wide-band imaging sub-problems to be considered (see discussion in Section 2.4). This strategy offers an increased scalability potential to Faceted HyperSARA over HyperSARA, which may reveal of significant interest in extreme dimension.

## 5 VALIDATION ON REAL DATA

In this section, we illustrate both the precision and scalability potential of Faceted HyperSARA through the reconstruction of a 15 GB image cube of Cyg A from 7.4 GB of VLA data. The algorithm is mapped on 496 CPU cores on a high performance computing system, achieving a TerafLOPS proof of concept. The performance of the proposed approach is evaluated in comparison with the monochromatic imaging approach SARA (Onose et al. 2017) and the CLEAN-based wide-band imaging algorithm JC-CLEAN in the software WSCLEAN (Offringa & Smirnov 2017). Note that HyperSARA (Abdulaziz et al. 2019b) is not considered in this study due to its prohibitive cost.

### 5.1 Dataset description and imaging settings

The data analyzed in this section are part of wide-band VLA observations of the celebrated radio galaxy Cyg A, acquired over two years (2015-2016) within the frequency range 2–18 GHz. We consider 480 channels in C band spanning the frequency range  $[\nu_1, \nu_{480}] = [3.979, 8.019]$  GHz, with a frequency step  $\delta\nu = 8$  MHz and a total bandwidth of 4.04 GHz. Observations phase center is at the position RA = 19h 59mn 28.356s ( $J2000$ ) and DEC =  $+40^\circ 44' 2.07''$ . The dataset was acquired at four instances which correspond to the frequency ranges  $[\nu_1, \nu_{256}] = [3.979, 6.019]$  GHz and  $[\nu_{257}, \nu_{480}] = [5.979, 8.019]$  GHz, and VLA configurations A and C. The wideband data consists of 30 spectral windows (each spanning 128 MHz), composed of 16 channels each, with approximately  $10^6$  complex visibilities per channel (about  $8 \times 10^5$  and  $2 \times 10^5$  measurements for configurations A and C, respectively), stored as double precision complex numbers.

In order to improve the accuracy of the modeled measurement operator, a pre-processing step has been conducted. It consists in a joint DDE calibration and imaging, applied to each channel separately. The approach, originally proposed by Repetti et al. (2017), consists in the alternate estimation of the unknown DDEs and the image of interest, with a spatio-temporal smoothness DDE prior and an average sparsity image prior (Repetti et al. 2017; Repetti & Wiaux 2017; Thouvenin et al. 2018). The underpinning algorithmic structure offers convergence guarantees to a critical point of the global non-convex optimization problem for

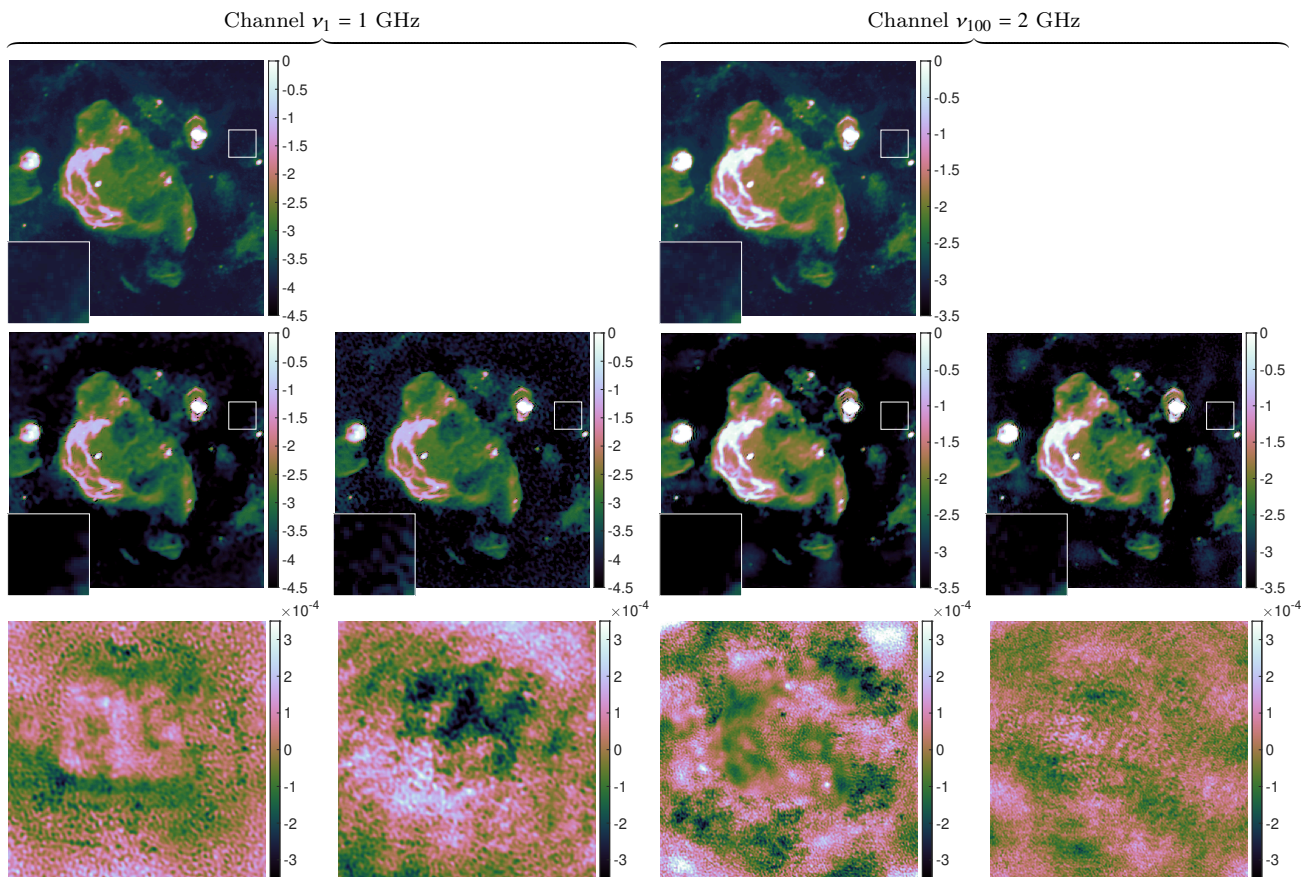
joint calibration and imaging and the approach was suggested to open the door to a significant improvement over the state-of-the-art (Repetti et al. 2017). Note that, one would ultimately want to resort to such a joint calibration and imaging approach to reconstruct the final wide-band image cube (Dabbech et al. 2019), rather than applying it as a pre-processing step on each channel separately. However, the underpinning algorithmic structure does not enable a faceting approach as the one proposed here, thereby severely limiting its scalability. We thus restrict ourselves to using this approach separately on each channel for scalability, and essentially to estimate DDEs. These are easily integrated into the forward model (1), as explained in Section 2. The estimated model visibilities are also exploited to determine estimates of the noise statistics, thus defining the  $\ell_2$  constraints in the minimization problem (3). Note that both SARA and Faceted HyperSARA take advantage of this pre-processing step, in contrast with JC-CLEAN, as the antenna-based DDEs estimates cannot be incorporated into WSCLEAN.

From the wide-band data acquired at  $L = 480$  channels, we consider imaging the field-of-view (FoV)  $\Omega_0 = 2.56' \times 1.536'$  at the spatial resolution given by the pixel size  $\delta x = 0.06''$  (in both directions), corresponding to the spatial dimension  $N = 2560 \times 1536$ . The imaged cube is of size  $2560 \times 1536 \times 480$ . Note that the selected pixel size is such that the spatial bandwidth of the recovered signal is up to 1.75 times the nominal resolution at the highest channel  $\nu_L = 8.019$  GHz, and 3.53 times the nominal resolution at the lowest channel  $\nu_1 = 3.979$  GHz. For both SARA and Faceted HyperSARA,  $B = 2$  data blocks are considered per channel. The data blocks are associated with VLA configurations A and C, presenting different noise statistics. More specifically to Faceted HyperSARA,  $C = 16$  channel-interleaved sub-problems are considered, with  $L_c = 30$ , for all  $c \in \{1, \dots, C\}$ , and  $Q = 5 \times 3$  facets along the spatial dimension, resulting in a total of  $Q \times C = 240$  spatio-spectral facets.

Concerning the initialization, both SARA and Faceted HyperSARA are initialized with the wide-band model cube  $\mathbf{X}^{(0)} = (\mathbf{X}_c^{(0)})_{1 \leq c \leq C}$ , obtained by the monochromatic joint calibration and imaging pre-processing step. From now on, for all  $(c, l) \in \{1, \dots, C\} \times \{1, \dots, L_c\}$ , the gridding matrices  $\mathbf{G}_{c,l}$  include the estimated DDEs, and  $\Phi_{c,l}$  refers to the resulting measurement operator. The  $\ell_2$  bounds defining the data-fidelity constraints in (3) are approximated as follows. For each data block indexed by  $(c, l, b) \in \{1, \dots, C\} \times \{1, \dots, L_c\} \times \{1, \dots, B\}$ , we set  $\varepsilon_{c,l,b} = \|\mathbf{y}_{c,l,b} - \Phi_{c,l,b} \mathbf{x}_{c,l}^{(0)}\|_2$ . More specifically to Faceted HyperSARA, the weights, which are defined in (10) and (11) and involved in the reweighting scheme (Algorithm 1), are initialized from the image cube  $\mathbf{X}^{(0)}$ . The regularization parameters  $\bar{\mu}_c$  and  $\mu_c$  are set as  $\bar{\mu}_c = 1/\|\mathbf{X}_c^{(0)}\|_* = 10^{-2}$  and  $\mu_c = 10^{-2}/\|\Psi^\dagger \mathbf{X}_c^{(0)}\|_{2,1} = 5 \times 10^{-6}$ . The rescaling by  $1/\|\mathbf{X}_c^{(0)}\|_*$  with respect to the values proposed in Section 4 theoretically does not affect the minimizers of the objective, but was empirically shown to optimize convergence. Finally, the SARA regularization parameter  $\bar{\mu}$  is fixed to  $\bar{\mu} = 5 \times 10^{-6}$ .

	Time (h)	aSNR	aSNR <sub>log</sub>	CPU cores
	0.19	<b>25.04</b> ( $\pm 4.06$ )	-6.28 ( $\pm 0.60$ )	120
HyperSARA	14.83	<b>31.74</b> ( $\pm 1.31$ )	-1.24 ( $\pm 0.57$ )	7
Faceted HyperSARA ( $C = 16, L_c \approx 6$ )	1.31	<b>31.05</b> ( $\pm 0.98$ )	-3.54 ( $\pm 1.37$ )	112
Faceted HyperSARA ( $C = 10, L_c \approx 10$ )	1.87	<b>31.48</b> ( $\pm 0.82$ )	-3.26 ( $\pm 1.43$ )	70
Faceted HyperSARA ( $C = 7, L_c \approx 14$ )	2.36	<b>31.68</b> ( $\pm 0.90$ )	-2.90 ( $\pm 1.38$ )	49
Faceted HyperSARA ( $C = 5, L_c \approx 20$ )	3.31	<b>31.84</b> ( $\pm 0.92$ )	-2.33 ( $\pm 0.91$ )	35
Faceted HyperSARA ( $C = 3, L_c \approx 33$ )	5.10	<b>32.00</b> ( $\pm 1.04$ )	-2.33 ( $\pm 1.07$ )	21
Faceted HyperSARA ( $C = 2, L_c \approx 50$ )	7.56	<b>31.97</b> ( $\pm 1.08$ )	-1.63 ( $\pm 0.64$ )	14

**Table 4.** Spectral faceting experiment: reconstruction performance of Faceted HyperSARA with a varying number of spectral sub-problems  $C$  and  $Q = 1$ , compared to HyperSARA (*i.e.* Faceted HyperSARA with  $Q = C = 1$ ) and SARA. The results are reported in terms of reconstruction time, aSNR and aSNR<sub>log</sub> (both in dB with the associated standard deviation) and total number of CPU cores. The reconstruction performance of Faceted HyperSARA, specifically investigated in this experiment, is highlighted in bold face.



**Figure 6.** Spectral faceting analysis for synthetic data: reconstructed images (in Jy/pixel) reported in  $\log_{10}$  scale for channels  $\nu_1 = 1$  GHz (first two columns) and  $\nu_{100} = 2$  GHz (last two columns) with Faceted HyperSARA for  $C = 10$  and  $Q = 1$  (columns 1 and 3) and HyperSARA (*i.e.* Faceted HyperSARA with  $Q = C = 1$ , in columns 2 and 4). Each sub-cube is composed of 10 out of the  $L = 100$  channels. From top to bottom: ground truth image, estimated model images and residual images.

## 5.2 Hardware

All the methods investigated in this section have been run on multiple nodes of Cirrus, having 36 cores and 256 GB of memory each (see Section 4 for further details). JC-CLEAN was run with 36 CPU cores assigned to each sub-problem. SARA was applied based on the parallelization procedure proposed by Onose et al. (2016b). More precisely, each channel is reconstructed using 12 CPU cores: 1 master CPU core, 2 CPU cores for the data-fidelity terms (one core per data-fidelity term) and 9 CPU cores to handle the average sparsity terms (associated with the nine bases of the SARA

dictionary). Finally, each sub-problem (6) (composed of 30 channels) is solved with Faceted HyperSARA using 1 master CPU core, 15 CPU cores to process the  $2 \times 30$  data-fidelity terms (4 data-fidelity blocks handled per core), and 15 CPU cores to handle the 15 spatio-spectral facets. We recall that SARA and Faceted HyperSARA are implemented using MATLAB whilst JC-CLEAN is implemented in C++.

### 5.3 Evaluation metrics

Imaging precision of the proposed approach is evaluated via the visual inspection of the reconstructed images, in comparison with the benchmark methods SARA and JC-CLEAN. For Faceted HyperSARA and SARA, we consider the examination of the estimated model cube  $\mathbf{X}$  and the naturally-weighted residual image cube  $\mathbf{R}$  whose columns, indexed by  $l \in \{1, \dots, L\}$ , are given by  $\mathbf{r}_l = \eta_l \Phi_l^\dagger (\mathbf{y}_l - \Phi_l \mathbf{x}_l)$ , where  $\mathbf{y}_l$  are the naturally-weighted RI measurements. Let  $\delta \in \mathbb{R}^N$  denote an image with value 1 at the phase center and zero elsewhere. The normalization factor  $\eta_l$  is such that the channel-associated point spread function (PSF), given by  $\eta_l \Phi_l^\dagger \Phi_l \delta$ , has a peak value equal to 1. In contrast with SARA and Faceted HyperSARA, for which natural weighting is adopted, optimal results for JC-CLEAN are obtained with Briggs weighting (Briggs 1995). JC-CLEAN examined images are the restored image cube  $\mathbf{T} = (\mathbf{t}_l)_{1 \leq l \leq L}$  and Briggs-weighted residual cube  $\tilde{\mathbf{R}} = (\tilde{\mathbf{r}}_l)_{1 \leq l \leq L}$ . We recall that the columns of the restored image cube, indexed by  $l \in \{1, \dots, L\}$ , are defined as  $\mathbf{t}_l = \mathbf{x}_l * \mathbf{c}_l + \tilde{\mathbf{r}}_l$ , where  $\mathbf{x}_l$  is the estimated model image (consisting of the CLEAN components) and  $\mathbf{c}_l$  is the CLEAN beam (typically a Gaussian fitted to the primary lobe of the associated PSF). As a quantitative metric of fidelity to data, the average standard deviation (aSTD) is reported for the three residual image cubes, and is defined for an image cube  $\mathbf{Z} \in \mathbb{R}^{N \times L}$  as  $\text{aSTD}(\mathbf{Z}) = \frac{1}{L} \sum_{l=1}^L \text{STD}_l(\mathbf{z}_l)$ , where  $\text{STD}_l(\mathbf{z}_l)$  is the standard deviation of the image  $\mathbf{z}_l \in \mathbb{R}^N$  at a given channel  $\nu_l$ . Finally, scalability of the different approaches is assessed through their computing time (elapsed time), resources (number of CPU cores) and overall computing cost.

## 5.4 Results and discussion

### 5.4.1 Imaging quality

To assess the reconstruction quality, we first examine the estimated images of channels  $\nu_1 = 3.979$  GHz and  $\nu_{480} = 8.019$  GHz, displayed in Figures 7 and 8, respectively. Note that these channels correspond to the respective channel indexes 1 and 30 of sub-problems 1 and 16. Average images of the estimated cubes, computed as the mean along the spectral dimension, are also examined and provided in Figure 9. From top to bottom are displayed in  $\log_{10}$  scale the model images of Faceted HyperSARA and SARA, and the restored images of JC-CLEAN. These images are overlaid with zooms on selected key regions of the radio galaxy. These are (i) the west hotspot (top left, left panel) over the angular area  $\Omega_1 = 0.08' \times 0.08'$ , centered at the position given by RA = 19h 59mn 33.006s ( $J2000$ ) and DEC = +40° 43' 40.889" and (ii) the inner core of Cyg A (top left, right panel) over the angular area  $\Omega_2 = 0.03' \times 0.03'$ , centered at the position RA = 19h 59mn 28.345s ( $J2000$ ) and DEC = +40° 44' 2.015". Note that the scale ranges of the displayed zooms are adapted to ensure a clear visualization of the contrast within the different structures of Cyg A.

In general, a visual inspection of the reconstructed images displayed in Figures 7 and 8 indicates superior imaging quality of Faceted HyperSARA model images, compared to the model images of SARA and the restored images of JC-CLEAN, with SARA imaging quality outperforming JC-

CLEAN. On the one hand, the higher resolution of Faceted HyperSARA is reflected by a better reconstruction of the hotspots and the inner core of Cyg A, in particular at the low-frequency channels (see Figure 7, first row, top left zooms). On the other hand, its higher dynamic range is reflected by the enhanced estimation of faint emissions in Cyg A, in particular, structures whose surface brightness is within the range [0.01,0.1] mJy (see the arc around the right end of the west jet in Figure 8, first row). We further observe that the proposed spatial tessellation does not introduce artifacts in the estimated images over the large dynamic range of interest. For SARA, given that no spectral correlation is promoted, the reconstruction quality of the different channels is restricted to their inherent resolution and sensitivity. This explains the lower reconstruction quality of SARA in comparison with Faceted HyperSARA. JC-CLEAN restored images exhibit a comparatively poorer reconstruction quality, which is likely to be attributed to both the much simpler regularization enforced, and the greedy nature of the algorithm. They are also limited to the instrument's resolution (through convolutions with the channel-associated synthesized CLEAN beams). The associated dynamic range is also limited by the prominent artifacts resulting from the lack of DDE calibration. The inspection of the average images displayed in Figure 9 confirms the ability of the proposed approach to recover fine details of Cyg A in comparison with SARA. Notice the faint emissions within the surface brightness range [0.01,0.1] mJy arising in the average JC-CLEAN restored image as opposed to the channel-specific restored images in Figures 7–8. This validates the different structures recovered by both Faceted HyperSARA and SARA across all channels, thanks to their accurate measurement operators.

Naturally-weighted residual images obtained with Faceted HyperSARA and SARA, and Briggs-weighted residual images obtained with JC-CLEAN are reported in Figures 7–9, displayed on bottom right panels overlaying the full recovered images, in linear scale. Their respective aSTD values are  $5.46 \times 10^{-4}$ ,  $4.53 \times 10^{-4}$  and  $5.2 \times 10^{-4}$ , indicating a comparable fidelity to data. Yet, a visual inspection of the residual images, in particular for the average residual images (Figure 9, bottom right panels), indicates that details of Cyg A jets are not fully recovered by SARA, as opposed to Faceted HyperSARA. Given that both approaches satisfy the same data constraints, this demonstrates the efficiency of the Faceted HyperSARA prior to capture the details of the galaxy. Although no DDE solutions are incorporated into the forward-modelling of JC-CLEAN, its residuals are homogeneous due to the absence of the non-negativity constraint. In fact, negative components are absorbed in its model images to compensate for spurious positive components.

The examination of the full image cubes provided by the three methods, available online (Thouvenin et al. 2020), confirms our analysis of the imaging quality offered by Faceted HyperSARA, SARA, and JC-CLEAN. The general shape of the spectra of the different emissions in Cyg A are consistent continuum emission following a typical power-law decay. Even though a spectral index map can be directly inferred from the reconstructed images, such an analysis lies beyond the scope of the present work. We also note a prominent spectral discontinuity at channel  $\nu_{257} = 5.979$  GHz, particularly noticeable for emissions with surface brightness above 1 mJy (*e.g.* see the inner core of Cyg A).

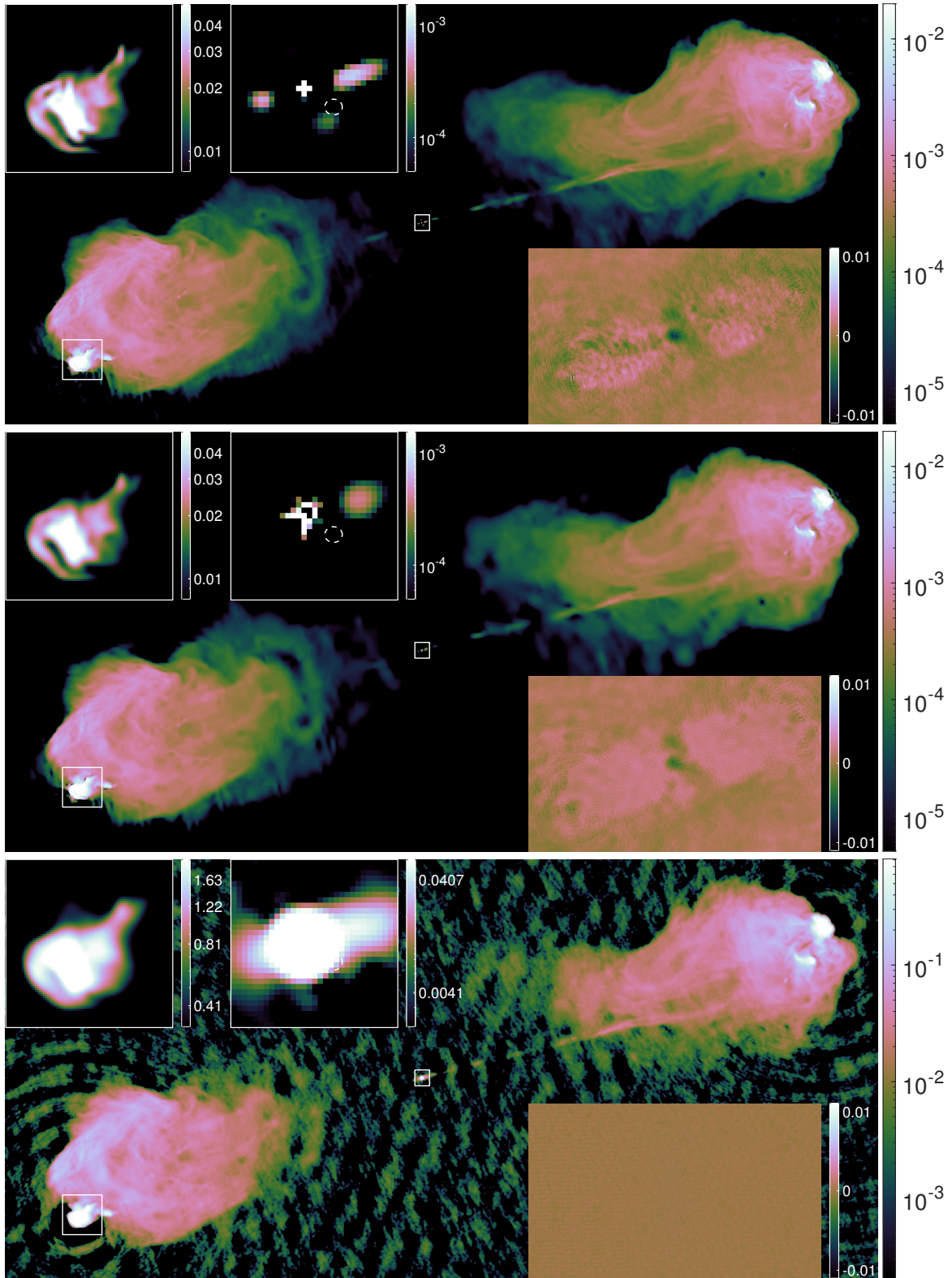
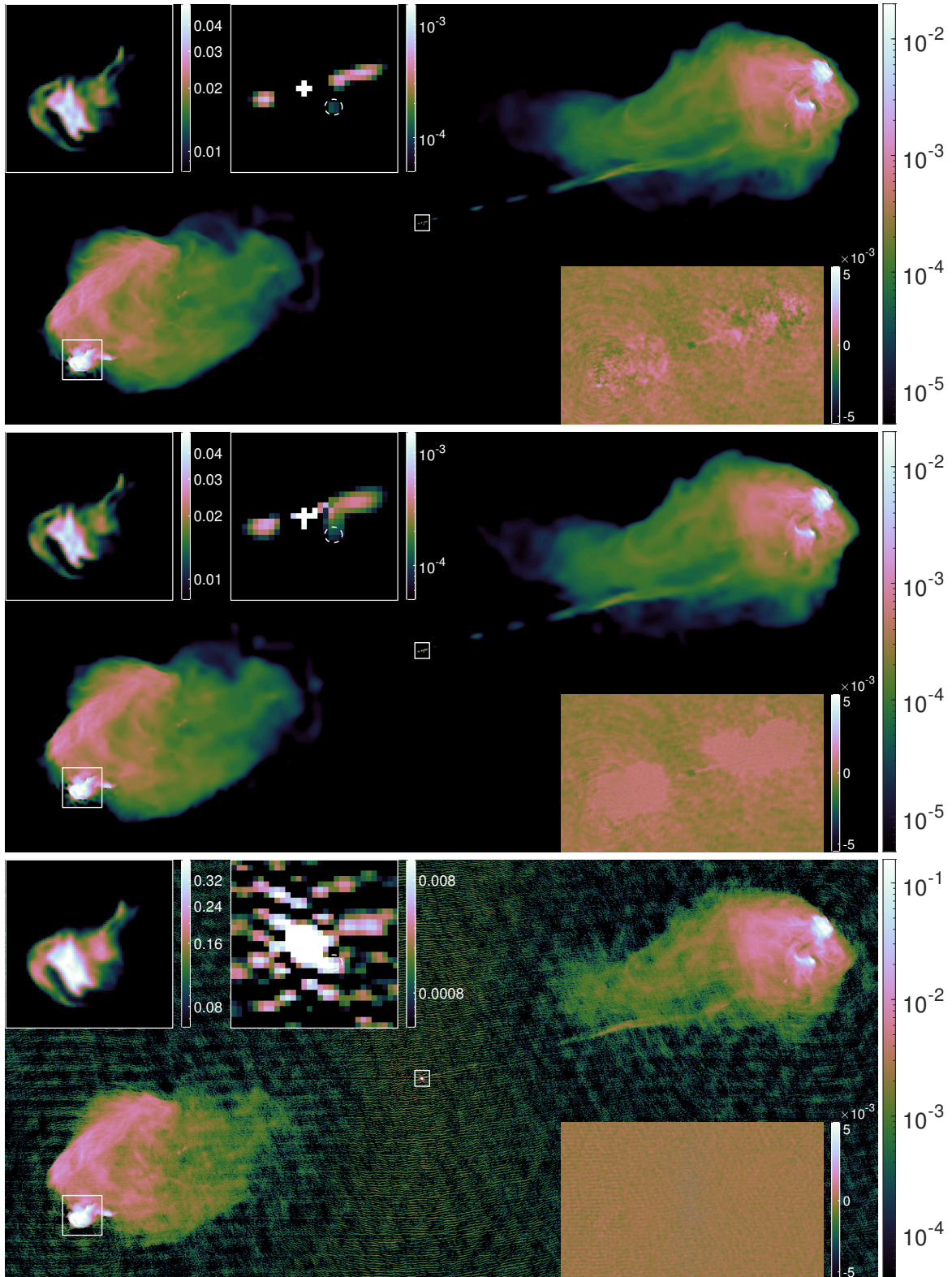


Figure 7. (Compressed images inserted due to size constraints. These exhibit very light compression artifacts such as loss of color contrast and blocking artifacts visible on full zoom. See [researchportal.hw.ac.uk](https://researchportal.hw.ac.uk) ID 25219759 for the full article with uncompressed images.) Cyg A imaged at the spectral resolution 8 MHz from 7.4 GB of data. Imaging results of channel  $\nu_1 = 3.979$  GHz. Estimated images at the angular resolution  $0.06''$  (3.53 times the associated sensed spatial bandwidth). From top to bottom: the respective estimated model images of Faceted HyperSARA ( $Q = 15$ ,  $C = 16$ ) and SARA, both in units of Jy/pixel, and restored image of JC-CLEAN in units of Jy/beam. The associated synthesized beam is of size  $0.37'' \times 0.35''$  and its flux is 42.18 Jy. The full FoV images ( $\log_{10}$  scale) are overlaid with the residual images (bottom right, linear scale) and zooms on selected regions in Cyg A (top left,  $\log_{10}$  scale). These correspond to the west hotspot (left) and the inner core of Cyg A (right). The zoomed regions are displayed with different value ranges for contrast visualization purposes and highlighted with white boxes in the full images. Cyg A-2 location is highlighted with a white dashed circle. Negative pixel values of JC-CLEAN restored image and associated zooms are set to 0 for visualization purposes. Full image cubes are available online (Thouvenin et al. 2020).





**Figure 8.** (Compressed images inserted due to size constraints. These exhibit very light compression artifacts such as loss of color contrast and blocking artifacts visible on full zoom. See [researchportal.hw.ac.uk](https://researchportal.hw.ac.uk) ID 25219759 for the full article with uncompressed images.) Cyg A imaged at the spectral resolution 8 MHz from 7.4 GB of data. Reconstruction results of channel  $\nu_{480} = 8.019$  GHz. Estimated images at the angular resolution  $0.06''$  (1.75 times the associated sensed spatial bandwidth). From top to bottom: the respective estimated model images of Faceted HyperSARA ( $Q = 15$ ,  $C = 16$ ) and SARA, both in units of Jy/pixel, and restored image of JC-CLEAN in units of Jy/beam. The associated synthesized beam is of size  $0.17'' \times 0.15''$  and its flux is 8.32 Jy. The full FoV images ( $\log_{10}$  scale) are overlaid with the residual images (bottom right, linear scale) and zooms on selected regions in Cyg A (top left,  $\log_{10}$  scale). These correspond to the west hotspot (left) and the inner core of Cyg A (right). The zoomed regions are displayed with different value ranges for contrast visualization purposes and highlighted with white boxes in the full images. Cyg A-2 location is highlighted with a white dashed circle. Negative pixel values of JC-CLEAN restored image and associated zooms are set to 0 for visualization purposes. Full image cubes are available online (Thouvenin et al. 2020).

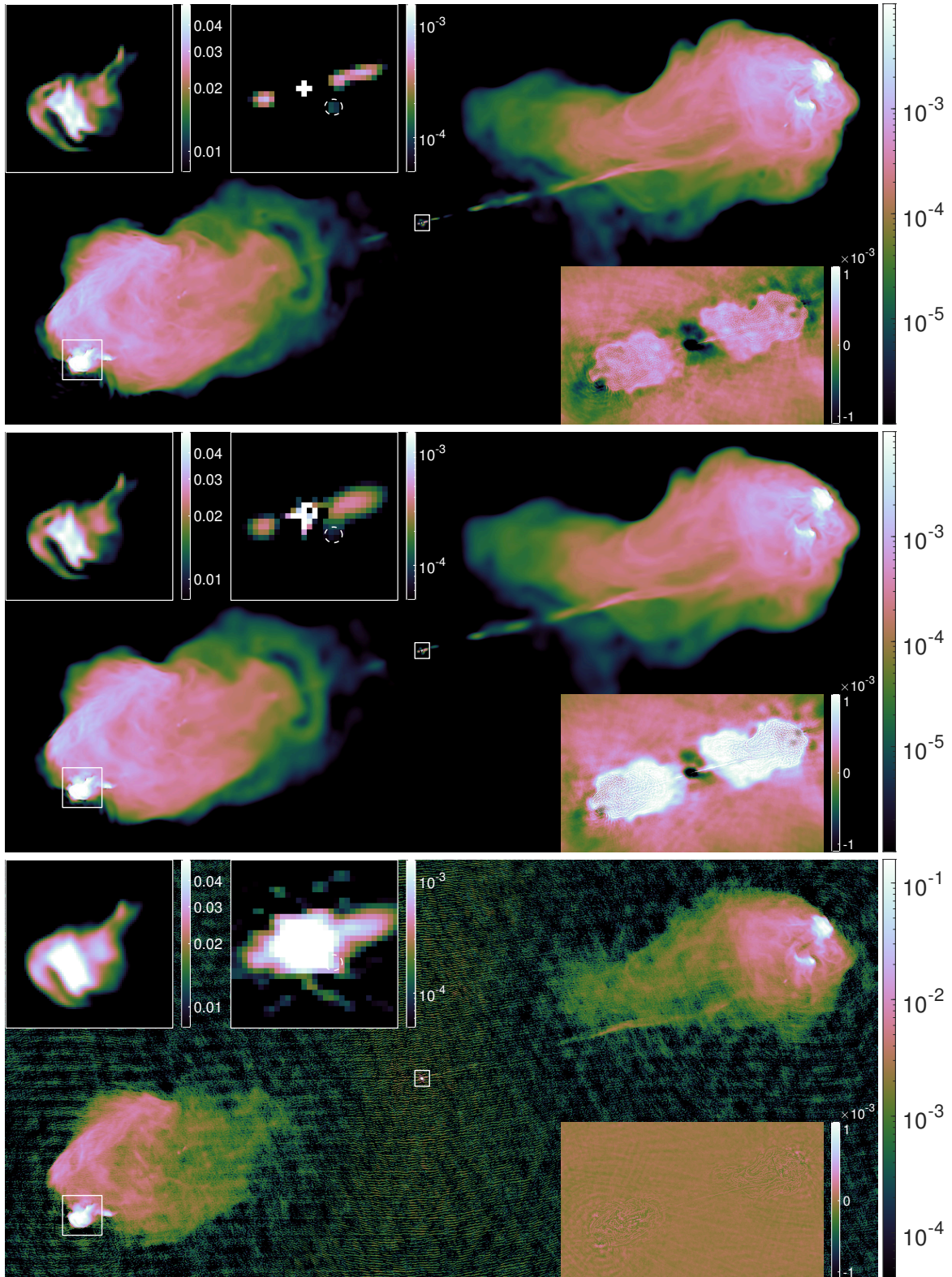


Figure 9. (Compressed images inserted due to size constraints. These exhibit very light compression artifacts such as loss of color contrast and blocking artifacts visible on full zoom. See [researchportal.hw.ac.uk](https://researchportal.hw.ac.uk) ID 25219759 for the full article with uncompressed images.) Cyg A imaged at the spectral resolution 8 MHz from 7.4 GB of data. Average estimated images, computed as the mean along the spectral dimension. From top to bottom: the respective estimated average model images of Faceted HyperSARA ( $Q = 15$ ,  $C = 16$ ) and SARA, and the average restored image of JC-CLEAN (obtained as the mean of the restored images normalized by the flux of their associated synthesized beam). The full FoV images ( $\log_{10}$  scale) are overlaid with the residual images (bottom right, linear scale) and zooms on selected regions in Cyg A (top left,  $\log_{10}$  scale). These correspond to the west hotspot (left) and the inner core of Cyg A (right). The zoomed regions are displayed with different value ranges for contrast visualization purposes and highlighted with white boxes in the full images. Cyg A-2 location is highlighted with a white dashed circle. Negative pixel values of JC-CLEAN restored image and associated zooms are set to 0 for visualization purposes.

	Time (h)	CPU cores	CPU time (h)
Faceted HyperSARA	68	496	<b>33728</b>
SARA	12.5	5760	<b>72000</b>
JC-CLEAN	22	576	<b>12672</b>

**Table 5.** Computing cost of Cyg A imaging at the spectral resolution 8 MHz from 7.4 GB of data. Results are reported for Faceted HyperSARA, SARA, and JC-CLEAN in terms of reconstruction time, number of CPU cores and overall CPU time (highlighted in bold face).

Since the dataset has been acquired separately in the frequency ranges  $[\nu_1, \nu_{256}] = [3.979, 6.019]$  GHz and  $[\nu_{257}, \nu_L] = [5.979, 8.019]$  GHz, the spectral discontinuity is most likely due to the different calibration errors and noise statistics of the two frequency ranges. We further notice small-scale spectral discontinuities, demarcating the spectral windows, which again are most likely to be attributed to calibration errors.

Recently, Perley et al. (2017) have reported the presence of a bright object in the inner core of the galaxy with a flux value reaching 0.5 mJy at the frequency 8.422 GHz. The object, dubbed Cyg A-2, has been identified with a white dashed circle in Figures 7–9 (top left, right panel), centered at the position given by RA = 19h 59mn 28.322s ( $J2000$ ) and DEC = +40° 44′ 1.89″ with a radius of size 0.1″. The discovery was further confirmed in Dabbech et al. (2018) by imaging two monochromatic VLA data sets at C band (6.678 GHz) and X band (8.422 GHz) with SARA. Interestingly, the inspection of Faceted HyperSARA estimated image cube shows that Cyg A-2 is discernible at frequencies lower than ever. More precisely, the source is resolved in all channels within the range [5.979, 8.019] GHz with an average flux value of 0.5164 ( $\pm 0.1394$ ) mJy. SARA, however, succeeds in detecting it within the range [7.131, 8.019] GHz with an average flux value of 0.5157 ( $\pm 0.3957$ ) mJy. Given the important calibration errors present in the associated restored images, JC-CLEAN is not able to resolve Cyg A-2.

#### 5.4.2 Computing cost

The computing time and resources required by the different methods, corresponding to the hardware resource allocation described in Section 5.2, are reported in Table 5. In this setting, JC-CLEAN required 12672 CPU hours, whereas SARA leveraged 72000 CPU hours. Finally, Faceted HyperSARA required 33728 CPU hours. These numbers indicate an overall higher efficiency of the parallelization procedure adopted for Faceted HyperSARA when compared to the one taken for SARA.

Interestingly, the computing cost of Faceted HyperSARA, implemented in MATLAB, is not far from the JC-CLEAN C++ implementation. This gap can be reduced, if not closed, with a C++ implementation of Faceted HyperSARA. It is worth noting that the reported numbers are sensitive to the number of CPU cores allocated for each method.

## 6 COMBINATION WITH DIMENSIONALITY REDUCTION

In the previous section, spectral faceting has been leveraged to divide the imaging problem into smaller, independent sub-problems to form an image of Cyg A at the full spectral resolution of the data. The reconstructions confirm a relatively flat spectrum, suggesting that the image could be formed at a much coarser spectral resolution without violating the Nyquist limit. This motivates a reduction of the image dimension. In this context, the amount of data per effective channel increases significantly. Following the recent work of Kartik et al. (2017), the data of each effective channel can be embedded into a lower dimensional sub-space to reduce the data volume. Such a joint image and data dimensionality reduction (DR) represents a significant potential for reducing both memory requirements and computing cost to form RI image cubes. This section illustrates how applying Faceted HyperSARA to a lower-dimension inverse imaging problem resulting from joint image and data DR can enable handling large wide-band visibility volumes, thereby boosting the scalability of the image formation process.

### 6.1 Image DR

In light of the reconstruction results obtained at the observations spectral resolution  $\delta\nu = 8$  MHz and presented in Section 5.4, we consider herein a coarse spectral sampling that is given by the spectral resolution  $\widehat{\delta\nu} = 128$  MHz by binning every 16 consecutive channels – constituting a spectral window – into one effective channel. By doing so, the total number of effective channels is  $\widehat{L} = 30$ . Furthermore, the number of data points per effective channel rises to  $\widehat{M}_l = 1.6 \times 10^7$  approximately, larger than the associated effective channel image size  $N = 2560 \times 1536$  (assuming the imaged FoV and spatial resolution considered in Section 5.1). In this setting, further reduction of the data volume per effective channel can be of high interest to lower memory requirements and computing cost for image reconstruction while preserving the overall reconstruction quality.

### 6.2 Data DR

Various data DR techniques have been studied by Kartik et al. (2017) in the context of single-channel imaging. In particular, data DR considered herein effectively relies on visibility gridding. In what follows, data DR is described in the context of Faceted HyperSARA.

Considering  $C$  independent interleaved sub-problems, let  $(c, l, b) \in \{1, \dots, C\} \times \{1, \dots, L_c\} \times \{1, \dots, B\}$  denote the index of a data block consisting of  $\widehat{M}_{c,l,b}$  data points. The new forward model in the DR setting reads as

$$\mathbf{y}'_{c,l,b} = \Phi'_{c,l,b} \mathbf{x}_l + \mathbf{n}'_{c,l,b}, \quad (13)$$

with the reduced data and noise vectors  $\mathbf{y}'_{c,l,b} \in \mathbb{C}^{\widehat{M}'_{c,l,b}}$  and  $\mathbf{n}'_{c,l,b} \in \mathbb{C}^{\widehat{M}'_{c,l,b}}$ , respectively, and the measurement operator  $\Phi'_{c,l,b} \in \mathbb{C}^{\widehat{M}'_{c,l,b} \times N}$ , where the reduced data dimension  $\widehat{M}'_{c,l,b}$  is much smaller than both the original data dimension  $\widehat{M}_{c,l,b}$  and the image dimension  $N$ . DR is achieved by applying a linear embedding operator  $\Gamma_{c,l,b} \in \mathbb{C}^{\widehat{M}'_{c,l,b} \times \widehat{M}_{c,l,b}}$  to the

data, leading to  $\mathbf{y}'_{c,l,b} = \mathbf{\Gamma}_{c,l,b}\mathbf{y}_{c,l,b}$ ,  $\mathbf{n}'_{c,l,b} = \mathbf{\Gamma}_{c,l,b}\mathbf{n}_{c,l,b}$ , and  $\mathbf{\Phi}'_{c,l,b} = \mathbf{\Gamma}_{c,l,b}\mathbf{\Phi}_{c,l,b}$ . The specific DR operator considered here reads as

$$\mathbf{\Gamma}_{c,l,b} = \mathbf{\Xi}_{c,l,b}\mathbf{\Lambda}_{c,l,b}\mathbf{G}_{c,l,b}^\dagger\mathbf{\Theta}_{c,l,b}^\dagger, \quad (14)$$

where  $\mathbf{\Lambda}_{c,l,b} \in \{0, 1\}^{\widehat{M}'_{c,l,b} \times K}$  is a selection operator, and the diagonal matrix  $\mathbf{\Xi}_{c,l,b} \in \mathbb{R}^{\widehat{M}'_{c,l,b} \times \widehat{M}'_{c,l,b}}$  is a noise-whitening operator. The measurement operator  $\mathbf{\Phi}'_{c,l,b}$  can be reformulated as

$$\mathbf{\Phi}'_{c,l,b} = \mathbf{\Xi}_{c,l,b}\mathbf{\Lambda}_{c,l,b}\mathbf{H}_{c,l,b}\mathbf{F}\mathbf{Z}, \quad (15)$$

where the holographic matrix  $\mathbf{H}_{c,l,b} \in \mathbb{C}^{K \times K}$  is given by

$$\mathbf{H}_{c,l,b} = \mathbf{G}_{c,l,b}^\dagger\mathbf{\Theta}_{c,l,b}^\dagger\mathbf{\Theta}_{c,l,b}\mathbf{G}_{c,l,b}.$$

Having introduced the holographic matrix  $\mathbf{H}_{c,l,b}$ , further details of the different matrices involved in the embedding operator  $\mathbf{\Gamma}_{c,l,b}$  can now be provided. On the one hand, the selection matrix  $\mathbf{\Lambda}_{c,l,b}$  acts on the holographic matrix  $\mathbf{H}_{c,l,b}$ , such that the retained rows are associated with the elements of  $\text{Diag}(\mathbf{H}_{c,l,b})$  with the largest amplitude values and contributing to a percentage  $P$  of  $\|\text{Diag}(\mathbf{H}_{c,l,b})\|_1$ . Note that the larger the number of visibilities clustered in a reduced amount of Fourier cells, the more significant the reduction of dimension, *i.e.* the smaller the ratio between each data block size  $\widehat{M}'_{c,l,b}$  and  $N$ . On the other hand, the diagonal of the matrix  $\mathbf{\Xi}_{c,l,b}$  is given by the vector  $(\text{Diag}(\mathbf{\Lambda}_{c,l,b}\mathbf{H}_{c,l,b}))^{-\frac{1}{2}}$ . This form of weighting ensures that the noise affecting the embedded data vector  $\mathbf{y}'_{c,l,b}$  is, identically distributed. Yet, it does not remove the noise correlation introduced by the gridding operation itself, *i.e.* the noise is not formally white.

The gist of data DR resides in the fact that, the final embedded data size is  $\widehat{M}'_{c,l,b} \ll N$ , and the measurement operators  $\mathbf{\Phi}'_{c,l,b}$  become blind to the original data dimension  $\widehat{M}_{c,l,b}$  thanks to the interplay of the precomputed (reduced and weighted) holographic matrices  $\mathbf{\Xi}_{c,l,b}\mathbf{\Lambda}_{c,l,b}\mathbf{H}_{c,l,b}$  of size  $\widehat{M}'_{c,l,b} \times K$ . Given the original sparse nature of  $\mathbf{G}_{c,l,b}$ , the holographic matrices can be efficiently encoded as sparse matrices.

The updated forward model (13) can be substituted to the original model and solved efficiently using Algorithms 1 and 2, replacing the initial measurement operators  $\mathbf{\Phi}_{c,l,b}$ , data vectors  $\mathbf{y}_{c,l,b}$  and the  $\ell_2$  bounds  $\varepsilon_{c,l,b}$  by the reduced measurement operators  $\mathbf{\Phi}'_{c,l,b}$ , the embedded data vectors  $\mathbf{y}'_{c,l,b}$  and updated  $\ell_2$  bounds  $\varepsilon'_{c,l,b}$ , respectively. Note that, despite the correlated nature of the noise affecting the reduced data induced by visibility gridding, [Kartik et al. \(2017\)](#) report that imposing data-fidelity through  $\ell_2$  constraints, which emanate from the assumption of white Gaussian noise, does not affect the reconstruction quality.

### 6.3 DR and imaging settings

Similarly to Section 5.1, the imaged FoV is  $\Omega_0 = 2.56' \times 1.536'$ , with spatial resolution given by the pixel size  $\delta x = 0.06''$  (in both directions) resulting in the spatial dimension  $N = 2560 \times 1536$ . In the context of image DR, the considered spectral resolution of the imaged cube is  $\widehat{\delta\nu} = 128$  MHz, resulting in  $\widehat{L} = 30$  effective channels. The image cube is of

size  $2560 \times 1536 \times 30$ , that is a 16-fold reduction from 15 GB down to 0.94 GB. On a further note, the spatial bandwidth of the recovered signal is up to 1.76 times the nominal resolution at the highest effective channel  $\widehat{\nu}_30 = 7.959$  GHz, and 3.48 times the nominal resolution at the lowest effective channel  $\widehat{\nu}_1 = 4.039$  GHz. For both SARA and Faceted HyperSARA, we consider  $B = 2$  data blocks per effective channel (associated with VLA configurations). More specifically to Faceted HyperSARA, given the important reduction in the spectral dimension, only one spectral problem is considered, (*i.e.*  $C = 1$ , and  $L_C = \widehat{L} = 30$ ). Therefore, the index  $c$  will be omitted in the following paragraphs from data-block vectors and their associated operators. JC-CLEAN is systematically applied to the binned data, described in Section 6.1. In the context of data DR, we found that the threshold percentage  $P = 85\%$  yields a reasonable compromise between the image reconstruction quality and the size of the resulting imaging problem. In fact, the data, initially composed of  $4.65 \times 10^8$  points, are reduced down to  $1.4 \times 10^7$  points. That is a 33-fold reduction in data volume from 7.4 GB down to 0.22 GB. This process also yields a significant decrease in memory requirements, reduced from 2.5 TB required to store the initial gridding matrices  $(\mathbf{G}_{l,b})_{(1,1) \leq (l,b) \leq (\widehat{L}, B)}$ , down to 220 GB required to store the reduced holographic matrices  $(\mathbf{\Lambda}_{l,b}\mathbf{H}_{l,b})_{(1,1) \leq (l,b) \leq (\widehat{L}, B)}$ , that is a 11-fold reduction in memory requirements. Note that data DR is applied to the measurement operators encapsulating the estimated DDEs, obtained with the pre-processing step described in Section 5.1.

Concerning imaging initialization, both Faceted HyperSARA and SARA exploit the image cube estimated via the monochromatic joint calibration and imaging pre-processing step as follows. For each effective channel  $l \in \{1, \dots, \widehat{L}\}$ , the associated initial image  $\widehat{\mathbf{x}}_l^{(0)}$  is taken as the average image over the associated spectral window. In Faceted HyperSARA,  $Q = 5 \times 3$  facets along the spatial dimension are considered, leading to a total of  $Q \times C = 15$  spatio-spectral facets. Data constraints are defined such that, for each data block indexed by  $(l, b) \in \{1, \dots, \widehat{L}\} \times \{1, \dots, B\}$ , the associated  $\ell_2$  bound is estimated as  $\varepsilon'_{l,b} = \|\mathbf{y}'_{l,b} - \mathbf{\Phi}'_{l,b}\widehat{\mathbf{x}}_l^{(0)}\|_2$ . For optimal convergence, the regularization parameters  $\bar{\mu}_1$  and  $\mu_1$  are set as  $\bar{\mu}_1 = 0.15/\|\mathbf{X}_c^{(0)}\|_* = 10^{-2}$  and  $\mu_1 = 4 \times 10^{-3}/\|\mathbf{\Psi}^\dagger\mathbf{X}_c^{(0)}\|_{2,1} = 5 \times 10^{-6}$ , similarly to Section 5.1. Finally, the SARA regularization parameter  $\bar{\mu}$  is fixed to  $\bar{\mu} = 5 \times 10^{-6}$ .

### 6.4 Hardware

Despite the drastic reduction enabled by DR, memory requirements of Faceted HyperSARA exceed the memory available on a single node of the Cirrus system used in Section 5. Therefore, we conducted DR experiments on a different HPC system, whose nodes have 2 Intel Casade Lake 8260M processors, each with 24 cores running at 2.4 GHz. The nodes, equipped with 192 GB main memory and 3 TB of Intel DCPMM, were used in Memory Mode: the 192 GB main memory were used as a cache, and the 3 TB persistent memory were exploited by the algorithm. Note that Faceted HyperSARA required around 300 GB of memory for this experiment.

The different methods were compared in the following setting. JC-CLEAN used 36 cores to form the 30 effective channels from the full dataset. As for SARA, each effective

channel was reconstructed using 12 CPU cores: 1 master CPU core, 2 CPU cores for the data-fidelity terms (one core per data block) and 9 CPU cores for the average sparsity (associated to the 9 bases of the SARA dictionary). Finally, Faceted HyperSARA was deployed on 31 CPU cores, including 1 master CPU core and two groups of 15 CPU cores handling 15 spatial facets and 15 data-fidelity terms (one core for two effective channels, *i.e.* 4 data blocks), respectively.

## 6.5 Evaluation metrics

The performance of Faceted HyperSARA is studied in comparison with SARA and JC-CLEAN. For imaging quality assessment, we examine the estimated model image cubes  $\widehat{\mathbf{X}}$  obtained by Faceted HyperSARA and SARA and the restored image cube  $\widehat{\mathbf{T}}$  obtained by JC-CLEAN. Data-fidelity is assessed through the visual inspection of the associated residual image cubes, described in what follows. For both SARA and Faceted HyperSARA, and for a given channel  $l \in \{1, \dots, \widehat{L}\}$ , the associated residual image is  $\mathbf{r}'_l = \eta'_l (\Phi'_l)^\dagger (\mathbf{y}' - \Phi'_l \widehat{\mathbf{x}}_l)$ . Let  $\delta \in \mathbb{R}^N$  denote an image with value 1 at the phase center and zero elsewhere. The normalization factor  $\eta'_l$  is such that the channel-associated PSF, given by  $\eta'_l = \Phi'^{\dagger}_l \Phi'_l \delta$ , has a peak value equal to 1. As for JC-CLEAN, similarly to Section 5.3, Briggs weighting is adopted for optimal results. Therefore, associated Briggs-weighted residual image cube is examined. Data-fidelity is also assessed in terms of the aSTD metric defined in Section 5.3. Finally, the scalability potential of the different methods in the context of DR is provided.

## 6.6 Results and discussion

### 6.6.1 Imaging quality

The estimated images associated with the low-frequency channel  $\widehat{\nu}_1 = 4.039$  GHz and high-frequency channel  $\widehat{\nu}_{30} = 7.959$  GHz are respectively reported in Figures 10 and 11. The average of the estimated images are also reported in Figure 12. From top to bottom are displayed the images obtained by Faceted HyperSARA, SARA and JC-CLEAN. Each image is overlaid with zooms on the west hotspot (top left, left panel), and the inner core of Cyg A (top left, right panel), defined over the respective angular areas  $\Omega_1$  and  $\Omega_2$ , detailed in Section 5.4.

When comparing SARA and Faceted HyperSARA, one can observe that the latter exhibits better reconstruction quality in terms of resolution and dynamic range, owing to its sophisticated wide-band underlying prior. Notice the resolved details of the jets and the west hotspot displayed in Figure 10. In agreement with the observations made in Section 5, JC-CLEAN reconstruction quality is comparatively poorer, and of lower resolution, due to the simpler regularization approach. The associated dynamic range is further limited by strong artifacts resulting from the lack of DDE calibration. Importantly, the images obtained by both Faceted HyperSARA and SARA preserve well the overall reconstruction quality reported in Section 5.4 in the absence of DR. To some extent, one can even notice a better resolution at specific regions, such as filaments and jets in both Faceted HyperSARA and SARA in the DR context.

	Time (h)	CPU cores	CPU time (h)
Faceted HyperSARA	142	31	<b>4402</b>
SARA	10	360	<b>3600</b>
JC-CLEAN	36	36	<b>1296</b>

**Table 6.** Computing cost of Cyg A imaging in the context of DR at the spectral resolution 128 MHz. Results are reported for Faceted HyperSARA with DR, SARA with DR and JC-CLEAN in terms of reconstruction time, total number of CPU cores and overall CPU time (highlighted in bold face).

The naturally-weighted residual images at the scrutinized channels and the average residual images obtained by Faceted HyperSARA, SARA and JC-CLEAN are displayed in the bottom right panels overlaying the associated estimated images in Figures 10–12. The respective aSTD values are  $4.12 \times 10^{-4}$ ,  $2.20 \times 10^{-4}$  and  $5.88 \times 10^{-4}$ . These values are in agreement with the ones reported in Section 5.4. Despite lower aSTD value obtained by SARA in comparison with Faceted HyperSARA, one can clearly see from the average residual images displayed in Figure 12, details left on the jets locations in the average residual image of SARA as opposed to Faceted HyperSARA. The residuals of both methods also exhibit less structure than the corresponding method in Section 5.4. JC-CLEAN average Briggs-weighted residual image also exhibits significant details left on the location of the jets.

The general examination of the full reconstructed image cubes, available online (Thouvenin et al. 2020), confirms our analysis of the imaging quality offered by Faceted HyperSARA, SARA, and JC-CLEAN. In the present DR setting with a single effective channel per spectral window, the cubes exhibit smooth spectra (*i.e.* free from the small spectral discontinuities reported in Section 5.4) of the different emissions in Cyg A, confirming a typical power-law decay behaviour. On a further note, a spectral discontinuity at the effective channel  $\widehat{\nu}_{17} = 6.039$  GHz, demarcating the two effective frequency ranges  $[\widehat{\nu}_1, \widehat{\nu}_{16}] = [4.039, 5.959]$  GHz and  $[\widehat{\nu}_{17}, \widehat{\nu}_{30}] = [6.039, 7.959]$  GHz is noticed. We recall that this discontinuity is most likely induced by the calibration errors affecting the observations associated with these two frequency ranges.

With regards to Cyg A-2, both Faceted HyperSARA and SARA succeed in detecting it on some channels, though with lower flux values than the ones reported in Section 5.4. In fact, Faceted HyperSARA detects the source within the frequency range [7.063, 7.831] GHz with an average flux value of 0.1411 ( $\pm 0.07709$ ) mJy, whereas SARA detects it within the frequency range [7.575, 7.959] GHz with an average flux value of 0.6485 ( $\pm 0.6064$ ) mJy. JC-CLEAN fails to detect the source, due to significant calibration errors.

### 6.6.2 Computing cost

The computing cost required by the different methods with hardware resource allocation described in Section 6.4 is reported in Table 6. Faceted HyperSARA requires approximately 4402 CPU hours to reconstruct the whole image cube, while SARA requires 3600 CPU hours and JC-CLEAN requires 1296 CPU hours. In agreement with Section 5, note that the MATLAB implementation of SARA and Faceted

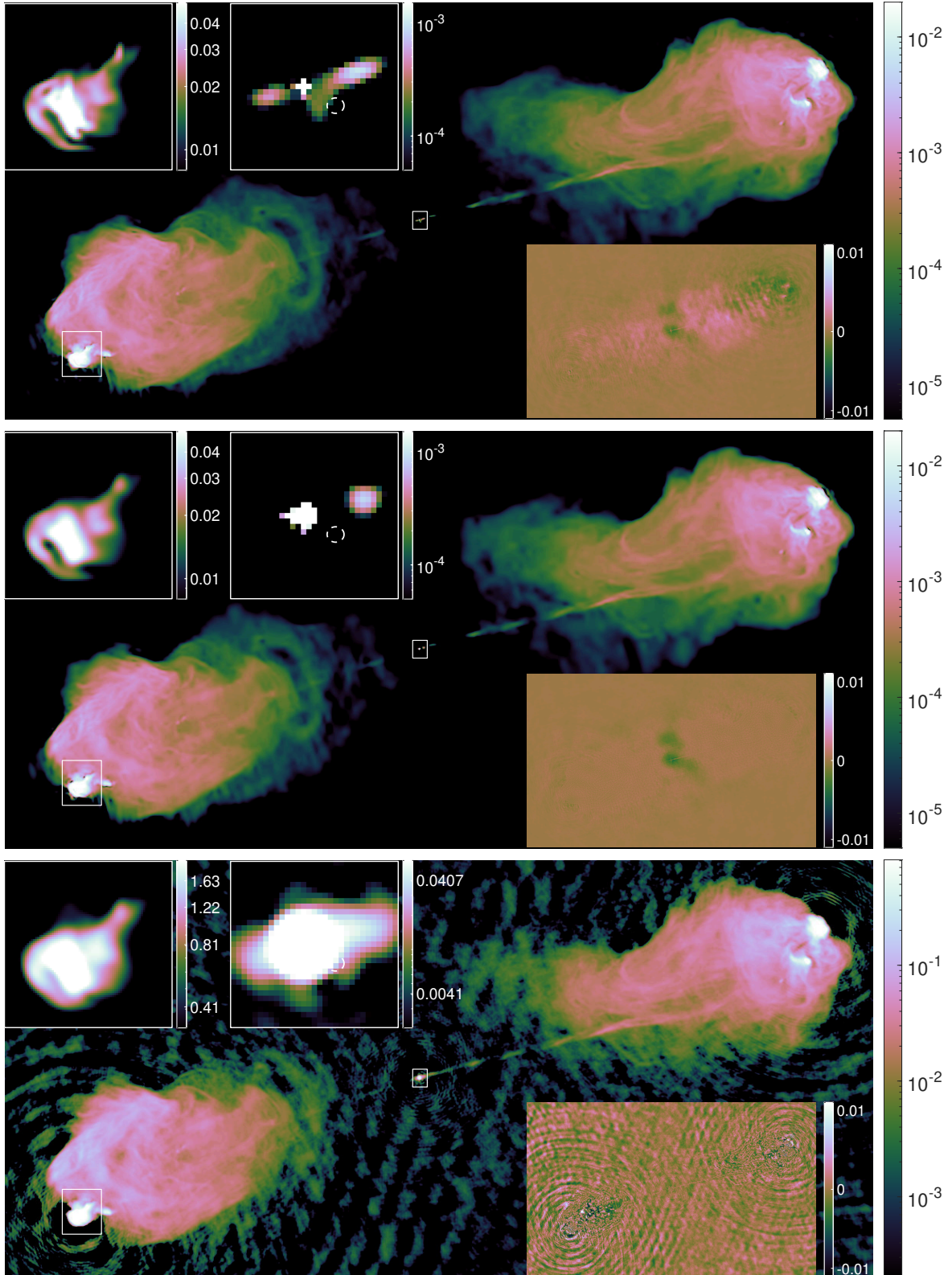


Figure 10. (Compressed images inserted due to size constraints. These exhibit very light compression artifacts such as loss of color contrast and blocking artifacts visible on full zoom. See [researchportal.hw.ac.uk](https://researchportal.hw.ac.uk) ID 25219759 for the full article with uncompressed images.) Cyg A imaged in the context of DR at the spectral resolution 128 MHz. Reconstruction results of channel  $\hat{\nu}_1 = 4.039$  GHz. Estimated images at the angular resolution  $0.06''$  (3.48 times the associated sensed spatial bandwidth). From top to bottom: the respective estimated model image of Faceted HyperSARA ( $Q = 15$ ,  $C = 1$ ) and SARA, both in units of Jy/pixel, and restored image of JC-CLEAN in units of Jy/beam. The associated synthesized beam is of size  $0.36'' \times 0.35''$  and its flux is 40.7 Jy. The full FoV images ( $\log_{10}$  scale) are overlaid with the residual images (bottom right, linear scale) and zooms on selected regions in Cyg A (top left,  $\log_{10}$  scale). These correspond to the west hotspot (left) and the inner core of Cyg A (right). The zoomed regions are displayed with different value ranges for contrast visualization purposes and highlighted with white boxes in the full images. Cyg A-2 location is highlighted with a white dashed circle. Negative pixel values of JC-CLEAN restored image and associated zooms are set to 0 for visualization purposes. Full image cubes are available online (Thouvenin et al. 2020).

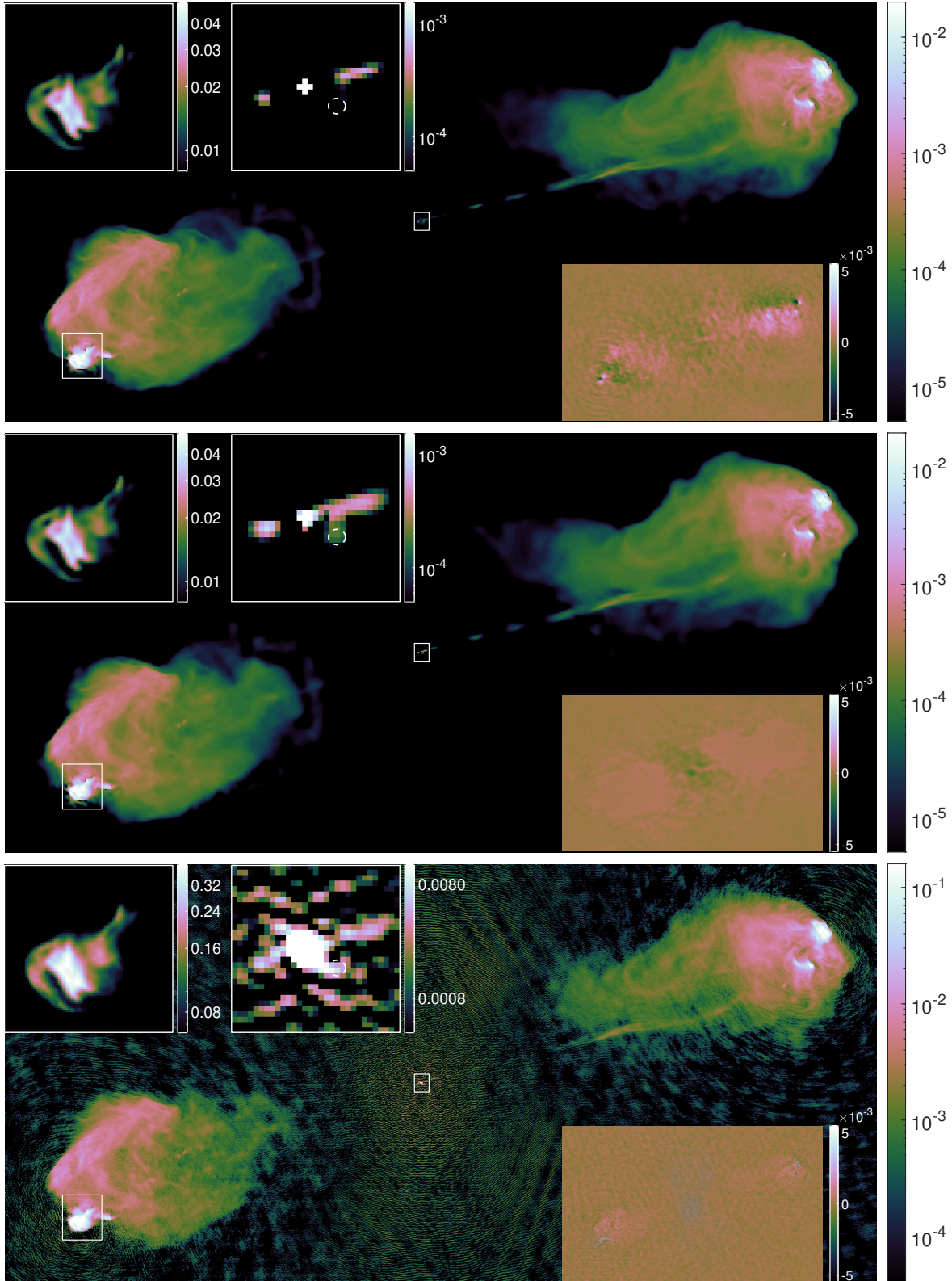


Figure 11. (Compressed images inserted due to size constraints. These exhibit very light compression artifacts such as loss of color contrast and blocking artifacts visible on full zoom. See [researchportal.hw.ac.uk](https://researchportal.hw.ac.uk) ID 25219759 for the full article with uncompressed images.) Cyg A imaged in the context of DR at the spectral resolution 128 MHz. Reconstruction results of channel  $\hat{\nu}_{30} = 7.959$  GHz. Estimated images at the angular resolution  $0.06''$  (1.76 times the observations spatial bandwidth). From top to bottom: the respective estimated model images of Faceted HyperSARA ( $Q = 15$ ,  $C = 1$ ) and SARA, both in units of Jy/pixel, and restored image of JC-CLEAN in units of Jy/beam. The associated synthesized beam is of size  $0.16'' \times 0.15''$  and its flux is 7.97 Jy. The full FoV images ( $\log_{10}$  scale) are overlaid with the residual images (bottom right, linear scale) and zooms on selected regions in Cyg A (top left,  $\log_{10}$  scale). These correspond to the west hotspot (left) and the inner core of Cyg A (right). The zoomed regions are displayed with different value ranges for contrast visualization purposes and highlighted with white boxes in the full images. Cyg A-2 location is highlighted with a white dashed circle. Negative pixel values of JC-CLEAN restored image and associated zooms are set to 0 for visualization purposes. Full image cubes are available online (Thouvenin et al. 2020).

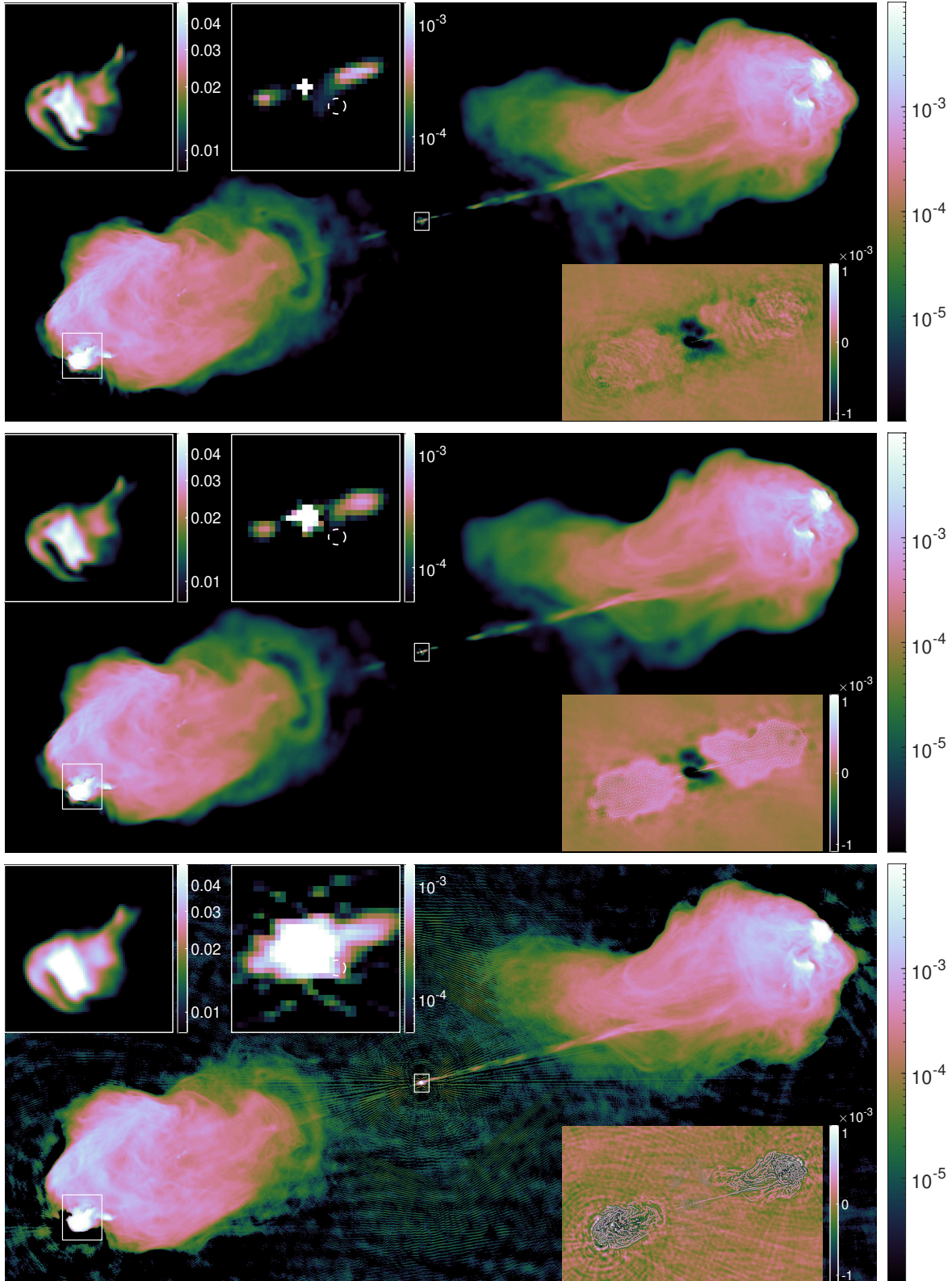


Figure 12. (Compressed images inserted due to size constraints. These exhibit very light compression artifacts such as loss of color contrast and blocking artifacts visible on full zoom. See [researchportal.hw.ac.uk](https://researchportal.hw.ac.uk) ID 25219759 for the full article with uncompressed images.) Cyg A imaged in the context of DR at the spectral resolution 128 MHz. Average estimated images, computed as the mean along the spectral dimension. From top to bottom: the respective estimated average model images of Faceted HyperSARA ( $Q = 15$ ,  $C = 1$ ) and SARA, and average restored image of JC-CLEAN (obtained as the mean of the restored images normalised by the flux of their associated synthesized beam). The full FoV images ( $\log_{10}$  scale) are overlaid with the residual images (bottom right, linear scale) and zooms on selected regions in Cyg A (top left,  $\log_{10}$  scale). These correspond to the west hotspot (left) and the inner core of Cyg A (right). The zoomed regions are displayed with different value ranges for contrast visualization purposes and highlighted with white boxes in the full images. Cyg A-2 location is highlighted with a white dashed circle. Negative pixel values of JC-CLEAN restored image and associated zooms are set to 0 for visualization purposes.



HyperSARA does not require significantly more CPU time than the optimized C++ implementation of JC-CLEAN.

More interestingly, in comparison with the computing cost of the experiments conducted in Section 5 and reported in Table 5, joint image and data DR allows a significant reduction in the computing cost, that is about 8-fold and 20-fold reduction for Faceted HyperSARA and SARA, respectively. We re-emphasize that the reported numbers are sensitive to the number of CPU cores allocated for each method.

## 7 CONCLUSION AND FUTURE WORK

We have introduced the Faceted HyperSARA method, which leverages a spatio-spectral facet prior model for wide-band radio-interferometric imaging. The underlying regularization encodes a sophisticated facet-specific prior model to ensure precision of the image reconstruction, allowing the bottleneck induced by the size of the image cube to be efficiently addressed via parallelization. Experiments conducted on synthetic data confirm that the proposed approach can provide a major increase in scalability in comparison with the original HyperSARA algorithm (Abdulaziz et al. 2019b), at no cost in imaging quality and showing potential to improve the reconstruction of faint emissions.

Leveraging the power of a large scale High Performance Computing system, our MATLAB implementation (available on GitHub <https://basp-group.github.io/Puri-Psi/>) has been further validated on the reconstruction of a 15 GB image cube of Cyg A from 7.4 GB of VLA data. The associated results are a practical proof of concept of the scalability of Faceted HyperSARA, which is also shown to provide a significant improvement in the imaging quality with respect to JC-CLEAN. Since a comparison with HyperSARA would have been impractical, we show that Faceted HyperSARA also supersedes the early monochromatic SARA approach in imaging precision. Interestingly, our results confirm the recent discovery of a super-massive second black hole in the inner core of Cyg A at lower frequencies than both JC-CLEAN and SARA (the black hole is detected and resolved at C band, starting from 5.979 GHz).

We have finally revisited the image formation process of Cyg A and proposed to reduce the dimension of the inverse imaging problem, both in terms of image and data dimensions, prior to applying Faceted HyperSARA. The combined approach was shown to be very effective in preserving the imaging quality obtained in the absence of dimensionality reduction, while offering significant reductions in both memory requirements and computing cost, ultimately translating into an enhanced scalability.

Having addressed the computational bottlenecks raised by both the volume of the data and the size of the image, future work should contemplate the definition and implementation of a faceted Fourier transform to improve the data and image locality in the proposed algorithm. Another perspective consists in developing a production C++ version of Faceted HyperSARA, building from the existing C++ version of HyperSARA (see the [Puri-Psi](#) webpage), with the aim of achieving maximum performance and scalability of a software implementation. Ultimately, beyond image estimation, the faceted approach should also be integrated into

the recently proposed methods for uncertainty quantification by convex optimization (*e.g.* Repetti et al. 2018, 2019; Abdulaziz et al. 2019a).

Last but not least, our work further illustrates the potential of advanced algorithms to enhance imaging quality beyond instrument resolution, opening the door to cost saving considerations for forthcoming arrays.

## ACKNOWLEDGEMENTS

The authors warmly thank R.A. Perley (NRAO, USA) for providing Cyg A observations with the VLA. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. This work was supported by EPSRC, grants EP/M011089/1, EP/M008843/1, EP/M019306/1, EP/R026173/1, the Swiss-South Africa Joint Research Program (IZLSZ2\_170863/1), and used the Cirrus UK National Tier-2 HPC Service at EPCC (<http://www.cirrus.ac.uk>) funded by the University of Edinburgh and EPSRC (EP/P020267/1). The dimensionality reduction experiments were run on hardware funded by the European Union's Horizon 2020 Research and Innovation program under Grant Agreement no. 671951.

## REFERENCES

- Abdulaziz A., Dabbech A., Onose A., Wiaux Y., 2016, in Proc. European Signal Process. Conf. (EUSIPCO). Budapest, Hungary, pp 388–392
- Abdulaziz A., Onose A., Dabbech A., Wiaux Y., 2017, in International Biomedical and Astronomical Signal Processing Frontiers Workshop. p. 6
- Abdulaziz A., Repetti A., Wiaux Y., 2019a, in Signal Processing with Adaptive Sparse Structured Representations (SPARS) workshop.
- Abdulaziz A., Dabbech A., Wiaux Y., 2019b, Monthly Notices of the Royal Astronomical Society, 489, 1230
- Arras P., Frank P., Leike R., Westermann R., Enßlin T., 2019, arXiv preprint arXiv:1903.11169
- Bauschke H. H., Combettes P. L., 2017, Convex analysis and monotone operator theory in Hilbert spaces. Springer, New York, NY, doi:<https://doi.org/10.1007/978-3-319-48311-5>
- Beck A., Teboulle M., 2009, SIAM Journal on Imaging Sciences, 2, 183
- Bhatnagar S., Cornwell T., 2004, Astronomy & Astrophysics, 426, 747
- Birdi J., Repetti A., Wiaux Y., 2018, Monthly Notices of the Royal Astronomical Society, 478, 4442
- Birdi J., Repetti A., Wiaux Y., 2019, Monthly Notices of the Royal Astronomical Society
- Briggs D. S., 1995, PhD thesis, The New Mexico Institute of Mining and Technology, Socorro
- Candès E. J., Boyd M. B. W. S. P., 2008, J. Fourier Anal. Appl., 4, 877
- Candès E. J., Li X., Ma Y., Wright J., 2009, journal of ACM, 58, 1
- Carilli C., Furlanetto S., Briggs F., Jarvis M., Rawlings S., Falcke H., 2004, New Astronomy Reviews, 48, 1029
- Carrillo R. E., McEwen J. D., Wiaux Y., 2012, Monthly Notices of the Royal Astronomical Society, 426, 1223
- Carrillo R. E., McEwen J. D., Van De Ville D., Thiran J., Wiaux Y., 2013, IEEE Signal Processing Letters, 20, 591

- Carrillo R. E., McEwen J. D., Wiaux Y., 2014, *Monthly Notices of the Royal Astronomical Society*, 439, 3591
- Combettes P. L., Pesquet J.-C., 2011, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, 49, 185
- Condat L., 2013, *J. Optim. Theory Appl.*, 158, 460
- Cornwell T. J., 2008, *IEEE Journal of Selected Topics in Signal Processing*, 2, 793
- Dabbech A., Mary D., Ferrari C., 2012, in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. pp 3665–3668
- Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, *Astronomy and Astrophysics*, 576, 16
- Dabbech A., Onose A., Abdulaziz A., Perley R. A., Smirnov O. M., Wiaux Y., 2018, *Monthly Notices of the Royal Astronomical Society*, 476, 2853
- Dabbech A., Repetti A., Wiaux Y., 2019. <http://www.basfrontiers.org>
- Dewdney P., Turner W., Millenaar R., McCool R., Lazio J., Cornwell T., 2013, Document number SKA-TEL-SKO-DD-001 Revision, 1
- Donoho D. L., 2006, *IEEE Trans. Inf. Theory*, 52, 1289
- Donoho D. L., Logan B. F., 1992, *SIAM Journal on Applied Mathematics*, 52, 577
- Donoho D. L., Stark P. B., 1989, *SIAM Journal on Applied Mathematics*, 49, 906
- Ferrari A., Deguignet J., Ferrari C., Mary D., Schutz A., Smirnov O., 2015, arXiv preprint arXiv:1504.06847
- Fessler J. A., Sutton B. P., 2003, *IEEE Trans. Signal Process.*, 51, 560
- Gaensler B. M., Beck R., Feretti L., 2004, *New Astronomy Reviews*, 48, 1003
- Garsden H., et al., 2015, *Astronomy & astrophysics*, 575, A90
- Geiping J., Moeller M., 2018, *SIAM J. Imaging Sci.*, 11, 2494
- Girard J., Garsden H., Starck J., Corbel S., Woiselle A., Tasse C., McKean J., Bobin J., 2015, *Journal of Instrumentation*, 10, C08013
- Hiriart-Urruty J.-B., Lemaréchal C., 1993, *Convex Analysis and Minimization Algorithms II*. Springer-Verlag Berlin Heidelberg, doi:10.1007/978-3-662-06409-2
- Högbom J., 1974, *Astronomy and Astrophysics Supplement Series*, 15, 417
- Hunter D. R., Lange K., 2004, *The American Statistician*, 58, 30
- Jiang M., Bobin J., Starck J.-L., 2017, *SIAM J. Imaging Sci.*, 10, 1997
- Jonas J., et al., 2018, in *MeerKAT Science: On the Pathway to the SKA*. p. 001
- Junklewitz H., Bell M. R., Enßlin T., 2015, *Astronomy & Astrophysics*, 581, A59
- Junklewitz H., Bell M., Selig M., Enßlin T., 2016, *Astronomy & Astrophysics*, 586, A76
- Kartik S. V., Carrillo R. E., Thiran J.-P., Wiaux Y., 2017, *Monthly Notices of the Royal Astronomical Society*, 480, 2382
- Komodakis N., Pesquet J.-C., 2015, *IEEE Signal Process. Mag.*, 32, 31
- Li, F. Cornwell, T. J. de Hoog, F. 2011, *Astronomy & Astrophysics*, 528, A31
- Moreau J.-J., 1965, *Bull. Soc. Math. France*, 93, 273
- Murya R., Ferrari A., Flamary R., Richard C., 2017, arXiv preprint
- Naghizadeh S., Repetti A., van der Veen A.-J., Wiaux Y., 2018, in *Proc. European Signal Process. Conf. (EUSIPCO)*. Roma, Italy
- Ochs P., Dosovitskiy A., Brox T., Pock T., 2015, *SIAM J. Imaging Sci.*, 8, 331
- Ochs P., Fadili J., Brox T., 2019, *SIAM J. Optim. Theory. Appl.*, 181, 244
- Offringa A., Smirnov O., 2017, *Monthly Notices of the Royal Astronomical Society*, 471, 301
- Onose A., Carrillo R. E., McEwen J. D., Wiaux Y., 2016a, in *Proc. European Signal Process. Conf. (EUSIPCO)*. Budapest, Hungary, pp 1448–1452
- Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016b, *Monthly Notices of the Royal Astronomical Society*, 462, 4314
- Onose A., Dabbech A., Wiaux Y., 2017, *Monthly Notices of the Royal Astronomical Society*, 469, 938
- Perley R., Chandler C., Butler B., Wrobel J., 2011, *The Astrophysical Journal Letters*, 739, L1
- Perley D. A., Perley R. A., Dhawan V., Carilli C. L., 2017, *ApJ*, 841, 117
- Pesquet J.-C., Repetti A., 2015, *Journal of nonlinear and convex analysis*, 16, 2453
- Pratley L., McEwen J. D., d’Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2017, *Monthly Notices of the Royal Astronomical Society*, 473, 1038
- Pruša Z., 2012, PhD thesis, Brno university of technology
- Rau U., Cornwell T. J., 2011, *Astronomy & Astrophysics*, 532, A71
- Rawlings S., Abdalla F., Bridle S., Blake C., Baugh C., Greenhill L., van der Hulst J., 2004, *New Astronomy Reviews*, 48, 1013
- Repetti A., Wiaux Y., 2017, in *Proceedings of the conference on Wavelets and Sparsity XVII*, part of the SPIE Optical Engineering + Applications. San Diego, California, United States
- Repetti A., Wiaux Y., 2019, arXiv preprint
- Repetti A., Birdi J., Dabbech A., Wiaux Y., 2017, *Monthly Notices of the Royal Astronomical Society*, 470, 3981
- Repetti A., Pereyra M., Wiaux Y., 2018, in *2018 26th European Signal Processing Conference (EUSIPCO)*. pp 2668–2672
- Repetti A., Pereyra M., Wiaux Y., 2019, *SIAM J. Imaging Sci.*, 12, 87
- Scaife A. M. M., 2020, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378, 20190060
- Schwab F. R., Cotton W. D., 1983, *The Astronomical Journal*, 88, 688
- Sutter P. M., et al., 2014, *Monthly Notices of the Royal Astronomical Society*, 438, 768
- Sutton E. C., Wandelt B. D., 2006, *The Astrophysical Journal Supplement Series*, 162, 401
- Tasse C., et al., 2018, *Astronomy & Astrophysics*, 611, A87
- Thompson A. R., Moran J. M., Jr. G. W. S., 2007, *Interferometry and Synthesis in Radio Astronomy*, Second Edition. Wiley-VCH Verlag, doi:10.1002/9783527617845
- Thouvenin P.-A., Repetti A., Dabbech A., Wiaux Y., 2018, in *Proc. IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM)*. Sheffield, UK, pp 475–479
- Thouvenin P.-A., Abdulaziz A., Jiang M., Dabbech A., Repetti A., Jackson A., Thiran J.-P., Wiaux Y., 2020, Cygnus A image cubes at C band (4-8 GHz) obtained with Faceted HyperSARA, doi:10.17861/e94f61c7-4950-4339-99a3-e6c3a820efdf
- Van Haarlem M., et al., 2013, *Astronomy & Astrophysics*, 556, A2
- Vũ B. C., 2013, *Advances in Computational Mathematics*, 38, 667
- Wenger S., Magnor M., 2014, Technical report, A sparse reconstruction algorithm for multi-frequency radio images. Computer Graphics Lab, TU Braunschweig
- Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vandergheynst P., 2009, *Monthly Notices of the Royal Astronomical Society*, 395, 1733
- van Haarlem, M. P. et al., 2013, *A&A*, 556, A2

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.