

A distributed wideband imaging algorithm for radio interferometry

P.-A. THOUVENIN^{*}, A. ABDULAZIZ[†], M. JIANG[‡], A. DABBECH[†],
A. REPETTI[†], A. JACKSON[★], J.-P. THIRAN[§], Y. WIAUX[†]

^{*}Université de Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL,

[†]Heriot-Watt University, [‡]Xidian University, [★]EPCC, [§]EPFL

RICOCHET kick-off meeting

MARCH 24, 2022

Table of contents

1 Problem setting

- ▶ Introduction
- ▶ Problem formulation
- ▶ Selecting a prior

2 Faceted HyperSARA

3 Simulation results

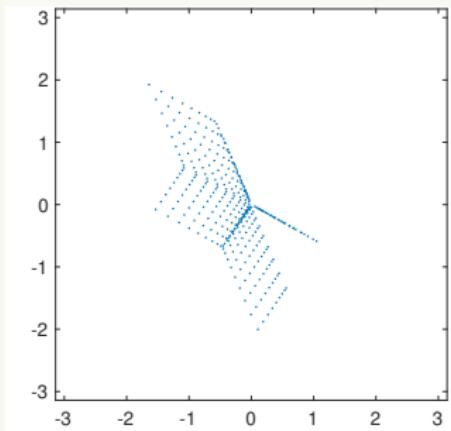
4 Conclusion

Radio-interferometric (RI) imaging: an illustration



VLA radio telescope (27 antennas)

Credit: NRAO



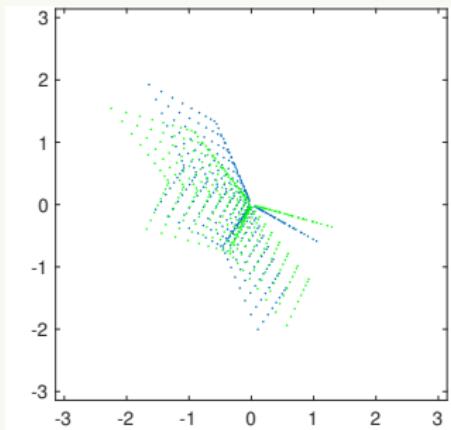
Fourier sampling at
instant $t = 1$

Radio-interferometric (RI) imaging: an illustration



VLA radio telescope (27 antennas)

Credit: NRAO



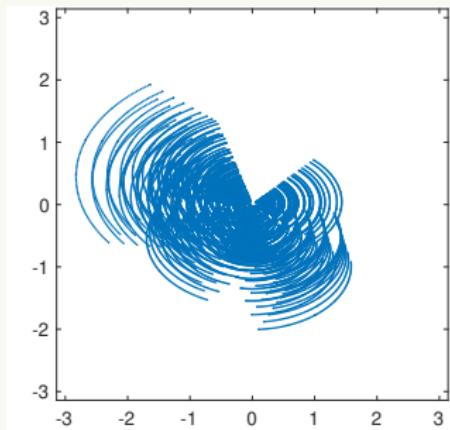
Fourier sampling at
instant $t = 20$

Radio-interferometric (RI) imaging: an illustration



VLA radio telescope (27 antennas)

Credit: NRAO



Fourier sampling for $T = 100$ time instant measurements

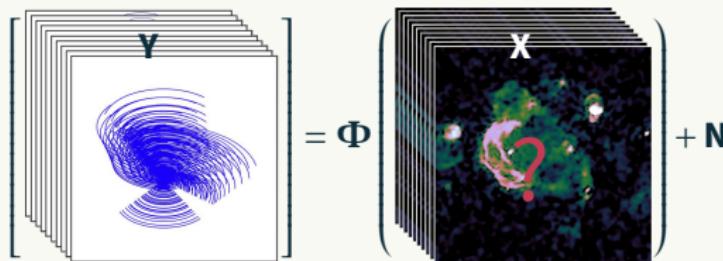
Big data in the SKA era: a few perspectives

- ▶ Modern telescopes (e.g., Square Kilometer Array (SKA): high imaging resolution and sensitivity
 - > gigabyte size image per frequency
 - > 10^4 observation frequencies
 - dynamic range 10^7
- ▶ Imaging challenge in bytes
 - ▶ petabyte size wide-band images
 - ▶ exabyte size data volumes (after correlation)



Wide-band RI imaging

- **Objective:** form wide-band image \mathbf{X} from incomplete data



M	number of measurements per channel
L	number of spectral channels
N	number of pixels
$\mathbf{Y} \in \mathbb{C}^{M \times L}$	wide-band data
$\mathbf{X} \in \mathbb{R}_+^{N \times L}$	wide-band image cube
Φ	measurement operator
$\mathbf{N} \in \mathbb{C}^{M \times L}$	noise

Discrete measurement model

► Measurement equation:

$$\begin{aligned}\mathbf{Y} &= \Phi(\mathbf{X}) + \mathbf{N} \\ \mathbf{y}_l &= \Phi_l \mathbf{x}_l + \mathbf{n}_l, \quad \Phi_l = \Theta_l \mathbf{G}_l \mathbf{F} \mathbf{Z}\end{aligned}\tag{1}$$

$\mathbf{x}_l \in \mathbb{R}_+^N$	image in channel l (column of \mathbf{X})
$\mathbf{y}_l \in \mathbb{C}^M$	data (visibilities) from channel l
$\mathbf{Z} \in \mathbb{R}^{K \times N}$	zero-padding and scaling operator
$\mathbf{F} \in \mathbb{C}^{K \times K}$	Fourier transform
$\mathbf{G}_l \in \mathbb{C}^{M \times K}$	interpolation (Fessler et al. 2003) and calibration kernels (Dabbech et al. 2017)
$\Theta_l \in \mathbb{R}^{M \times M}$	natural weighting (noise whitening)
$\mathbf{n}_l \in \mathbb{C}^M \sim \mathcal{CN}(\mathbf{0}_M, \sigma_l^2 \mathbf{I}_{M \times M})$	noise (realization of a complex Gaussian r.v.)

► Data assumed to be pre-calibrated (\mathbf{G}_l completely known).

Problem formulation

$$\underset{\mathbf{X} \in \mathbb{R}_+^{N \times L}}{\text{minimize}} f(\mathbf{Y}, \Phi(\mathbf{X})) + r(\mathbf{X}). \quad (2)$$

- f data fidelity term
(complex Gaussian noise $\Rightarrow \ell_2$ -norm ball or quadratic term)
- r regularization term
 - ~ sparsity in a transformed domain (Wenger et al. 2014; Ferrari et al. 2015)...
 - ~ low-rankness (source separation model) (Jiang et al. 2017)
 - ~ low-rankness + sparsity (Abdulaziz et al. 2019)
 - ...

- ➊ How to deal with the volume of data (M large)? (split data fitting term f)
- ➋ How to address large image sizes (N large)? (split regularization r)

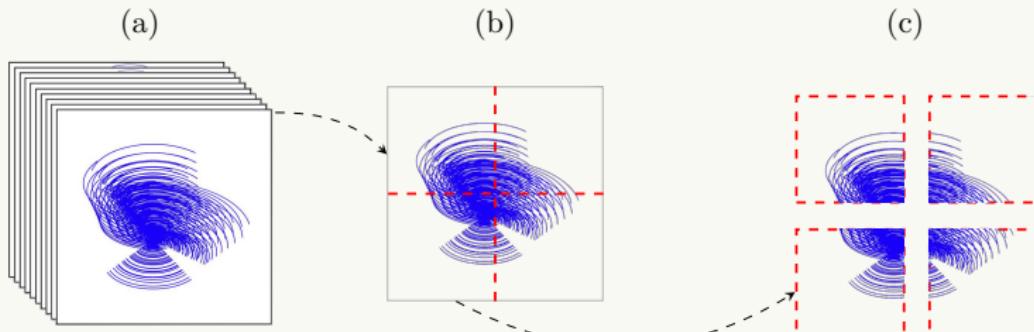
Towards a more scalable procedure

Primary bottleneck: **data size** (Onose et al. 2016)

- ▶ split data into frequency **blocks** (or groups of snapshots)
- ▶ assign data blocks to different **data workers**

$$\rightsquigarrow f(\mathbf{Y}, \Phi(\mathbf{X})) = \sum_{l=1}^L \sum_{b=1}^B \iota_B(\mathbf{y}_{l,b}, \varepsilon_{l,b}) (\Phi_{l,b} \mathbf{x}_l)$$

$\rightsquigarrow \varepsilon_{l,b}$ reflects the noise statistics for the block (b, l) (Onose et al. 2016)



Selecting a prior r : building on the literature

- **SARA** (Carrillo et al. 2012): joint average sparsity

$$r(\mathbf{X}) = \mu \sum_{l=1}^L \sum_{i=1}^I \log(|[\boldsymbol{\Psi}^\dagger \mathbf{x}_l]_i| + \nu) \quad (\text{SARA})$$

$\boldsymbol{\Psi}^\dagger \in \mathbb{R}^{I \times N}$



SARA dictionary (first 8 Daubechies wavelet and Dirac basis)
monochromatic, i.e., no spectral correlations

$\mu, \nu > 0$

regularization parameters

Selecting a prior r : building on the literature

- **HyperSARA** (Abdulaziz et al. 2019): low-rankness and joint average sparsity

$$r(\mathbf{X}) = \bar{\mu} \sum_{j=1}^J \log(|\sigma_j(\mathbf{X})| + \bar{\nu}) + \mu \sum_{i=1}^I \log(\|[\Psi^\dagger \mathbf{X}]_i\|_2 + \nu) \quad (\text{HyperSARA})$$

- $\Psi^\dagger \in \mathbb{R}^{I \times N}$ SARA dictionary (first 8 Daubechies wavelet and Dirac basis)
- $\Delta \Delta$ full image cube \mathbf{X} needed in a single place
- $\mu, \bar{\mu}, \nu, \bar{\nu} > 0$ regularization parameters
- $(\sigma_j(\mathbf{Z}))_{1 \leq j \leq J}$ singular values of the matrix \mathbf{Z}
- $[\mathbf{Z}]_i$ i th row of \mathbf{Z}

Parameter estimation

- **Log priors:** (2) not convex, addressed through reweighting (Candès et al. 2008)
To reweighting algo.

(local majorant of r at $\mathbf{X}^{(p)}$, $p \in \mathbb{N}$ current iteration index).

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

SARA:

$$r(\mathbf{X}, \mathbf{X}^{(p)}) = \mu \|\Psi^\dagger \mathbf{X}\|_{1,1,\omega(\mathbf{X}^{(p)})}, \quad (4)$$

$$\omega_{i,l}(\mathbf{X}^{(p)}) = \nu \left(|[\Psi^\dagger \mathbf{x}_l^{(p)}]_i| + \nu \right)^{-1}. \quad (5)$$

- Convex “subproblem” (3):
 - ~ primal-dual forward-backward (PDFB) (Condat 2013; Vũ 2013)
 - ~ no costly operator inversions or sub-iterations + splitting
 - ~ address each non-smooth function in parallel through its proximity operator

Parameter estimation

- **Log priors:** (2) not convex, addressed through reweighting (Candès et al. 2008) To reweighting algo.

(local majorant of r at $\mathbf{X}^{(p)}$, $p \in \mathbb{N}$ current iteration index).

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

HyperSARA:

$$r(\mathbf{X}, \mathbf{X}^{(p)}) = \bar{\mu} \|\mathbf{X}\|_{*, \bar{\omega}(\mathbf{X}^{(p)})} + \mu \|\Psi^\top \mathbf{X}\|_{2,1, \omega(\mathbf{X}^{(p)})}, \quad (4)$$

$$\omega_i(\mathbf{X}^{(p)}) = v \left(\|[\Psi^\top \mathbf{X}^{(p)}]_i\|_2 + v \right)^{-1}, \quad (5)$$

$$\bar{\omega}_j(\mathbf{X}^{(p)}) = \bar{v} \left(|\sigma_j(\mathbf{X}^{(p)})| + \bar{v} \right)^{-1}. \quad (6)$$

- Convex “subproblem” (3):
 - ~ primal-dual forward-backward (PDFB) (Condat 2013; Vũ 2013)
 - ~ no costly operator inversions or sub-iterations + splitting
 - ~ address each non-smooth function in parallel through its proximity operator

Image faceting

Secondary bottleneck: image size (focus in this presentation)

- ▶ RI literature: wide-band faceted calibration and imaging DDFacet (Tasse et al. 2018)

- ▶ primarily developed for calibration (piece-wise constant calibration model)
- ▶ tessellation improves imaging efficiency
- ✗ no convergence guarantee...

- ~ Motivation: keep quality of HyperSARA and
- ▶ split image into **3D facets** (spatial + spectral)
 - ▶ assign portions of the image (facets) to different workers (*facet nodes*)

⇒ **Faceted HyperSARA**

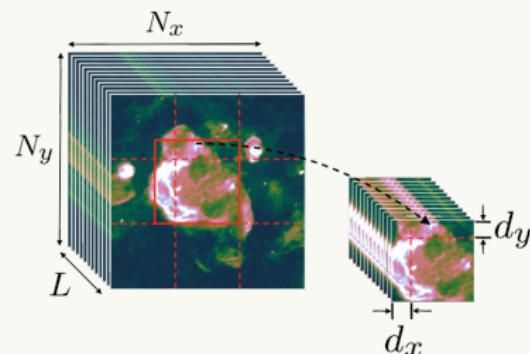


Table of contents

1 Problem setting

2 Faceted HyperSARA

- ▶ Faceted HyperSARA prior
- ▶ Algorithm structure (PDFB)

3 Simulation results

4 Conclusion

Spectral and spatial faceting

(a) Full image cube (b) Spectral sub-cubes (c) Facets & weights

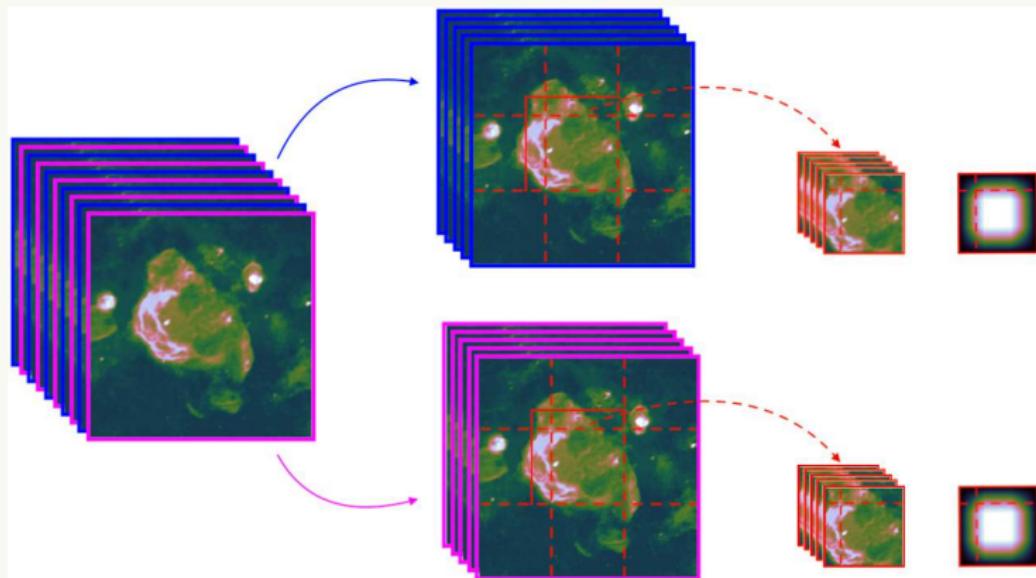


Figure 1: Illustration of the proposed faceting scheme.

Faceted HyperSARA prior

- **HyperSARA** (Abdulaziz et al. 2019): low-rankness and joint average sparsity

$$r(\mathbf{X}) = \bar{\mu} \sum_{j=1}^J \log(|\sigma_j(\mathbf{X})| + \bar{\nu}) + \mu \sum_{i=1}^I \log(\|[\Psi^\dagger \mathbf{X}]_i\|_2 + \nu) \quad (\text{HyperSARA})$$

$\Psi^\dagger \in \mathbb{R}^{I \times N}$



SARA dictionary (first 8 Daubechies wavelets + Dirac)
full cube \mathbf{X} needed in a single place

Faceted HyperSARA prior

- **Faceted HyperSARA:** faceted low-rankness and joint average sparsity

$$r(\mathbf{X}) = \sum_{q=1}^Q \left(\bar{\mu} \sum_{j=1}^{J_q} \log(|\sigma_j(\mathbf{D}_q \tilde{\mathbf{S}}_q \mathbf{X})| + \bar{\nu}) + \mu \sum_{i=1}^{I_q} \log(\|[\Psi_q^\dagger \mathbf{S}_q \mathbf{X}]_i\|_2 + \nu) \right)$$

(faceted HyperSARA)

$$\Psi_q^\dagger \in \mathbb{R}^{I_q \times N_q}$$

✓

$$\tilde{\mathbf{S}}_q \in \mathbb{R}^{\tilde{N}_q \times N}, \mathbf{S}_q \in \mathbb{R}^{N_q \times N}$$

$$\mathbf{D}_q$$

exact faceted implementation of Ψ^\dagger (Prusa 2012)

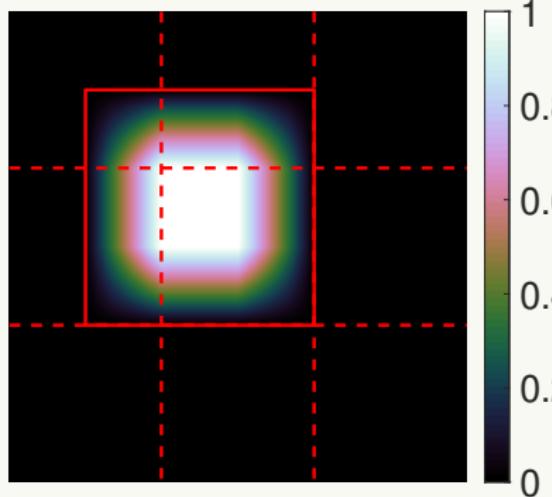
spatial tessellation

content-agnostic facet selection operators

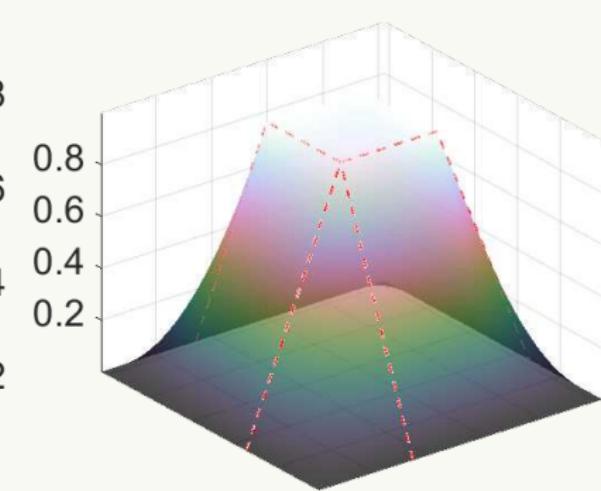
spatial weights (mitigate tessellation artefacts)

- ~ Amount of overlap: free parameter for $\tilde{\mathbf{S}}_q$, fixed for \mathbf{S}_q (Prusa 2012);
- ~ Partially separable expression for the function r ;
- ~ HyperSARA = faceted HyperSARA with $Q = 1$.

Facet weights



(a) Weights for the central facet



(b) 3D view

Figure 2: Facet weights $(\mathbf{d}_q)_{1 \leq q \leq Q}$, for $Q = 9$ (3 facets along each spatial dimension).

Parameter estimation

- **Log priors:** addressed through reweighting (Candès et al. 2008)
(local majorant of r at $\mathbf{X}^{(p)}$, $p \in \mathbb{N}$ current iteration index).

$$\underset{\mathbf{X} \in \mathbb{R}_+^{N \times L}}{\text{minimize}} \sum_{l,b} \iota_B(\mathbf{y}_{l,b}, \varepsilon_{l,b}) (\Phi_{l,b} \mathbf{x}_l) + r(\mathbf{X}, \mathbf{X}^{(p)}),$$

Faceted HyperSARA

$$r(\mathbf{X}, \mathbf{X}^{(p)}) = \sum_q \left(\bar{\mu} \|\mathbf{D}_q \widetilde{\mathbf{S}}_q \mathbf{X}\|_{*, \bar{\omega}_q(\mathbf{X}^{(p)})} + \mu \|\Psi_q^\top \mathbf{S}_q \mathbf{X}\|_{2,1, \omega_q(\mathbf{X}^{(p)})} \right), \quad (7)$$

$$\omega_{q,i}(\mathbf{X}^{(p)}) = v \left(\|[\Psi_q^\top \mathbf{S}_q \mathbf{X}^{(p)}]_i\|_2 + v \right)^{-1}, \quad (8)$$

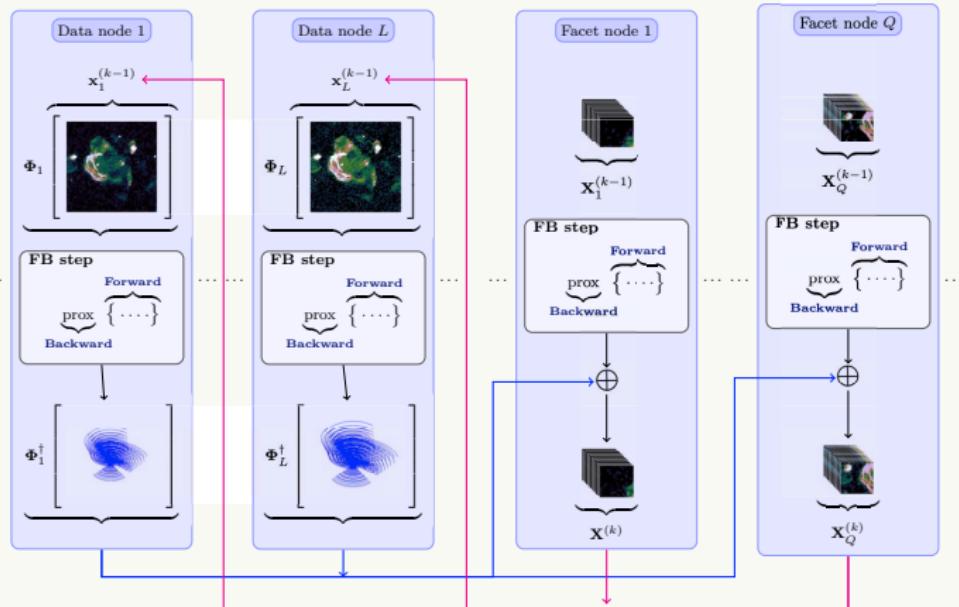
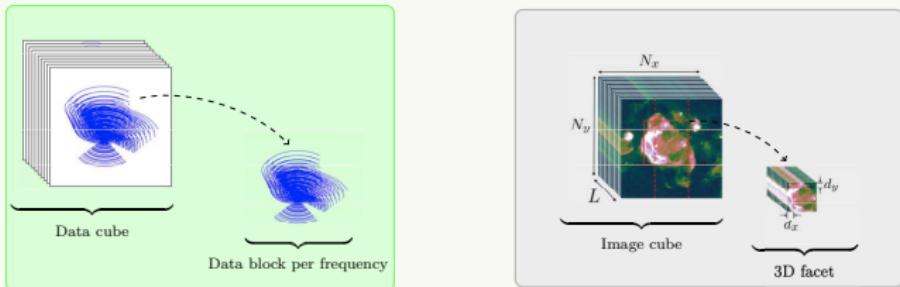
$$\bar{\omega}_{q,j}(\mathbf{X}^{(p)}) = \bar{v} \left(|\sigma_j(\mathbf{D}_q \widetilde{\mathbf{S}}_q \mathbf{X}^{(p)})| + \bar{v} \right)^{-1}. \quad (9)$$

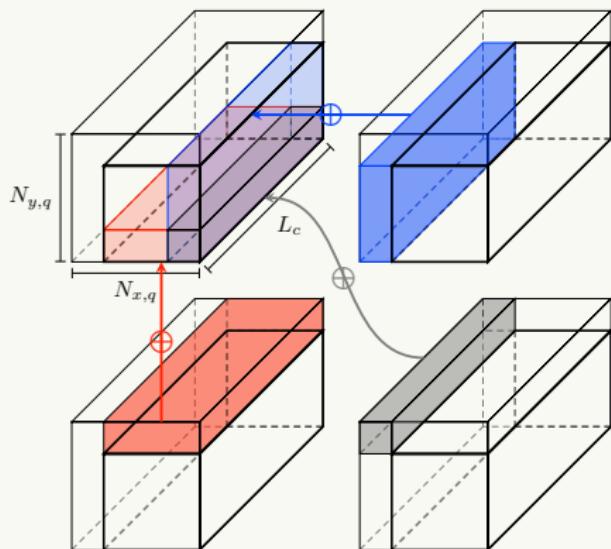
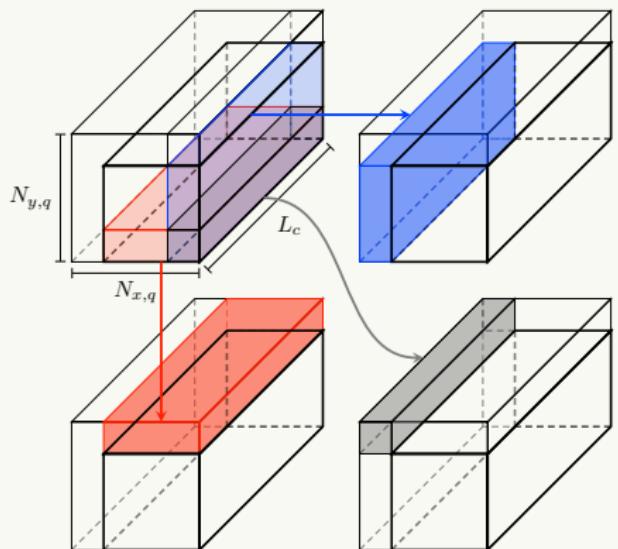
- “subproblem” (3): **primal-dual forward-backward (PDFB)** (Condat 2013)

Parameter estimation (PDFB)

- ▶ Split update of the dual variables between two sets of nodes:
 - ▶ data nodes: contain data & dual variables of full image size (few bands)
 - ▶ facet nodes: contain portions of the image cube (facet size over the full spectrum) + associated dual variables
- ▶ Most of the (dual) variables updated in parallel
- ▶ Exploit parallelization flexibility depending on the size of the problem (N, L, M)

To PDFB code





(a) Update borders (broadcast to neighbours)

(b) Aggregate borders (from neighbour facets)

Figure 3: Communications between the facet nodes, occurring between each single facet and a maximum of three of its neighbours.

[To PDFB code](#)

Table of contents

- 1 Problem setting
- 2 Faceted HyperSARA
- 3 Simulation results
- 4 Conclusion

Experiments on synthetic data

Simulation settings:

- ▶ synthetic wide-band image of Cygnus A
- ▶ $L = 20$ spectral channels, $N = 1024 \times 2048$ pixels
- ▶ $M \approx 0.5N$ measurements per channel, SNR = 60 dB
- ▶ $B = 1$ data block
- ▶ Comparison: SARA (Carrillo et al. 2012), HyperSARA (HS) (Abdulaziz et al. 2019) and Faceted HyperSARA (FHS)

Assessment criteria:

- ▶ average SNR (aSNR, in dB) (\pm std of the SNR over the channels)
- ▶ average SNR of \log_{10} of the image (better characterize dynamic range)
- ▶ timing: runtime per PDFB iteration (cpu_{pi}), active CPU time per iteration (cpu)
- ▶ total runtime (run), total active CPU time (run_{pi})

Varying number of facets

	aSNR (dB)	aSNR _{log} (dB)	CPU cores	PDFB iterations	run _{pi} (s)	run (h)	cpu _{pi} (s)	cpu (h)
SARA	35.05 (± 0.59)	13.72 (± 0.16)	240	3275	3.28 (± 0.38)	3.38	7.13 (± 0.95)	129.77
HS	39.47 (± 2.15)	16.36 (± 1.27)	22	9236	25.36 (± 0.85)	65.06	84.49 (± 2.79)	216.76
FHS ($Q = 4$)	39.79 (± 2.34)	16.47 (± 1.39)	24	10989	26.50 (± 1.88)	80.90	184.41 (± 9.22)	562.90
FHS ($Q = 9$)	40.00 (± 2.40)	16.82 (± 1.42)	29	11009	15.38 (± 1.38)	47.04	226.52 (± 11.00)	692.71
FHS ($Q = 16$)	40.08 (± 2.40)	16.85 (± 1.43)	36	10945	11.62 (± 0.50)	35.32	286.06 (± 10.80)	869.71

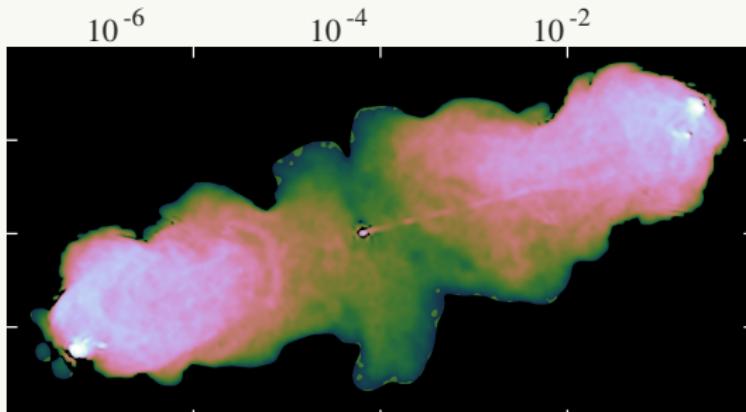
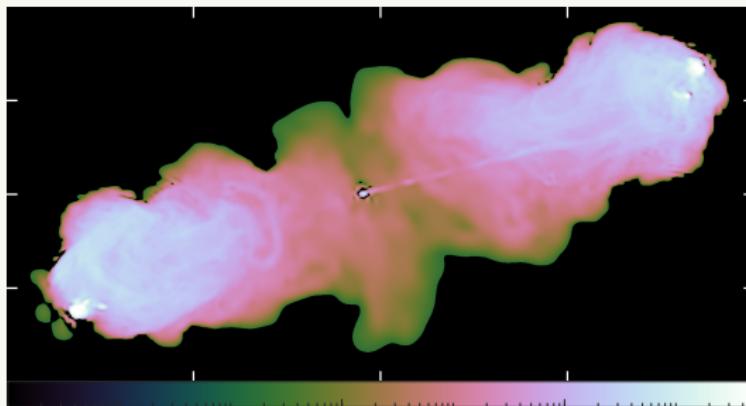
Table 1: Varying number of facets Q . Faceted HyperSARA (FHS, overlap of 50%), HyperSARA (HS) and SARA.

Varying overlap between facets

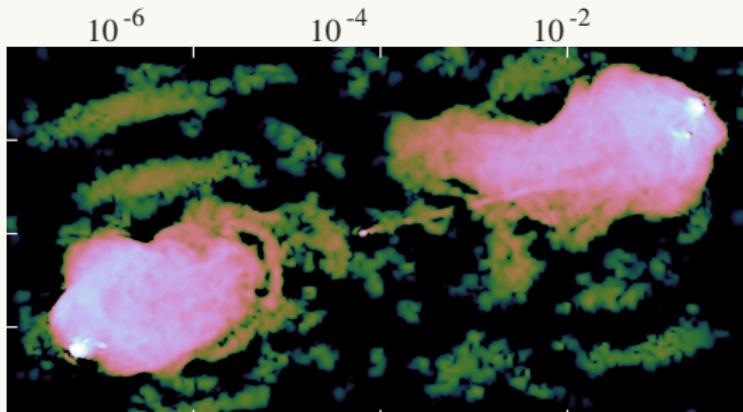
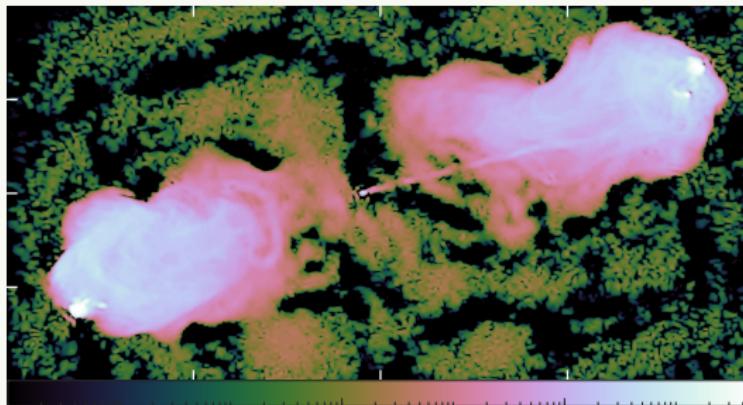
	aSNR (dB)	aSNR _{log} (dB)	CPU cores	PDFB iterations	run _{p1} (s)	run (h)	cpu _{p1} (s)	cpu (h)
SARA	35.05 (± 0.59)	13.72 (± 0.16)	240	3275	3.28 (± 0.38)	3.38	7.13 (± 0.95)	129.77
HS	39.47 (± 2.15)	16.36 (± 1.27)	22	9236	25.36 (± 0.85)	65.06	84.49 (± 2.79)	216.76
FHS (0% overlap)	40.03 (± 2.41)	16.79 (± 1.42)	36	10961	11.55 (± 0.70)	35.18	284.17 (± 13.40)	865.22
FHS (10% overlap)	40.08 (± 2.40)	16.85 (± 1.43)	36	10945	11.62 (± 0.50)	35.32	286.06 (± 10.80)	869.71
FHS (25% overlap)	40.22 (± 2.41)	16.96 (± 1.47)	36	10918	11.96 (± 0.53)	36.26	290.71 (± 13.90)	881.66
FHS (40% overlap)	40.24 (± 2.42)	17.05 (± 1.51)	36	10934	12.67 (± 0.55)	38.47	298.32 (± 14.30)	906.08
FHS (50% overlap)	40.08 (± 2.53)	16.95 (± 1.52)	36	10962	13.69 (± 0.65)	41.68	311.14 (± 16.00)	947.41

Table 2: Varying size of the overlap region (faceted low-rank prior). Faceted HyperSARA (FHS) with $Q = 16$, HyperSARA (HS) and SARA.

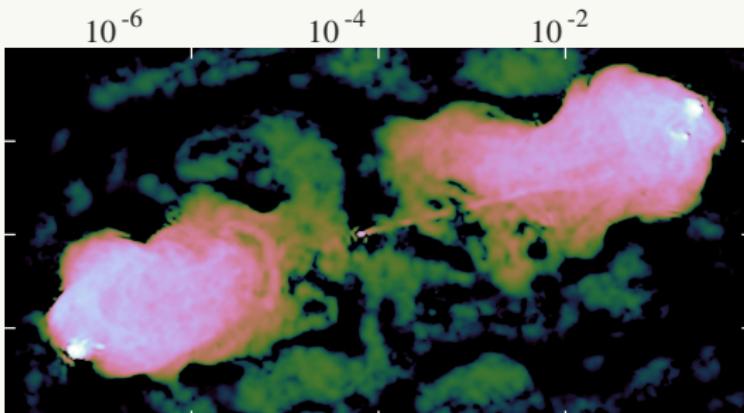
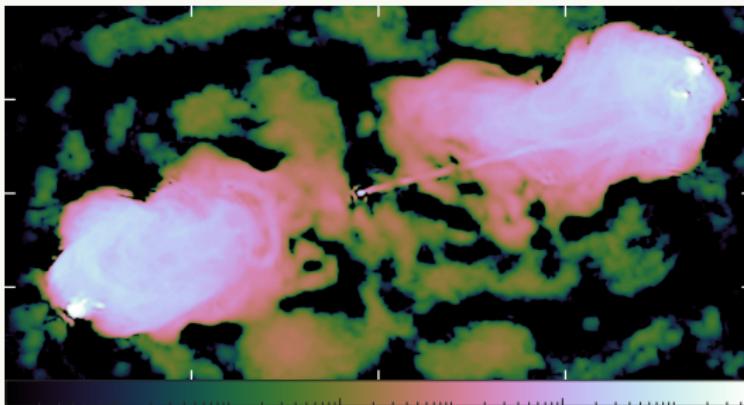
Experiments on synthetic data (ground truth)



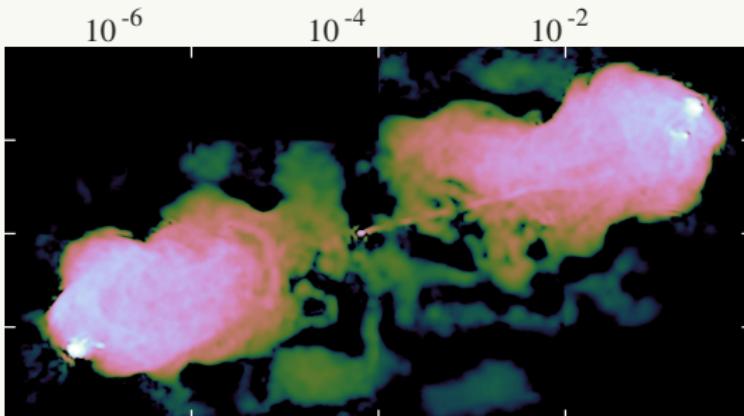
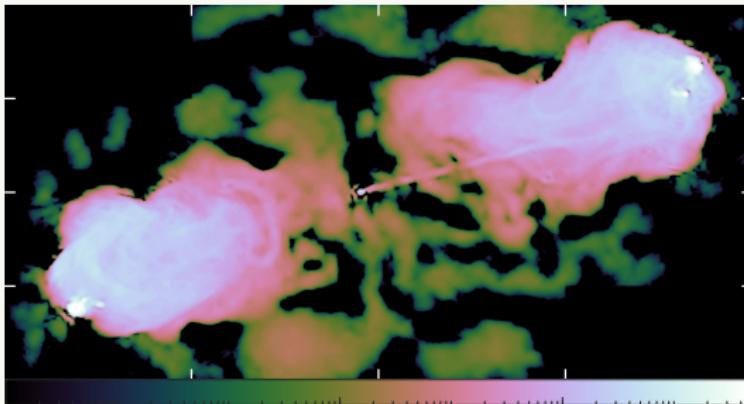
Experiments on synthetic data (SARA)



Experiments on synthetic data (HS)



Experiments on synthetic data (FHS, 0% overlap)



Experiments on synthetic data (FHS, 10% overlap)

Real data

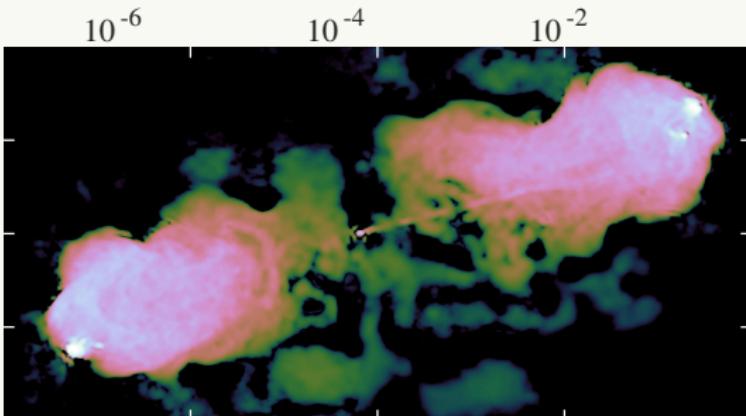
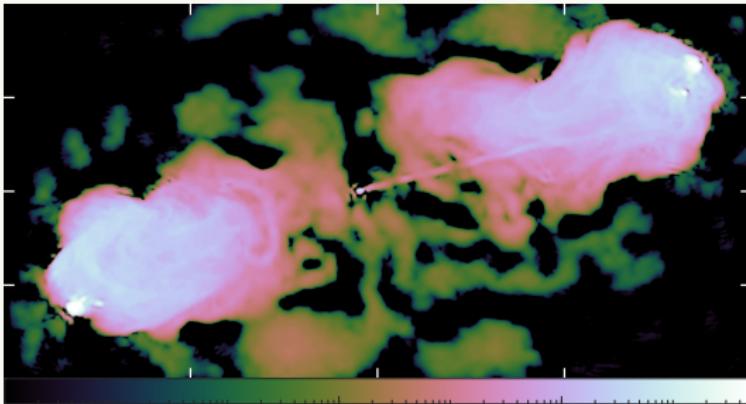


Table of contents

- 1 Problem setting**
- 2 Faceted HyperSARA**
- 3 Simulation results**
- 4 Conclusion**

Conclusions and perspectives

Conclusions: faceted prior for wide-band imaging

- ✓ quality comparable to HyperSARA...
- ✓ for a much lower computing time (increased distribution flexibility)
- ✓ spectral faceting (Thouvenin et al. 2020)

Conclusions and perspectives

Conclusions: faceted prior for wide-band imaging

- ✓ quality comparable to HyperSARA...
- ✓ for a much lower computing time (increased distribution flexibility)
- ✓ spectral faceting (Thouvenin et al. 2020)

Perspectives:

- limited to Stokes I for now
 - ↝ reconstruct other Stokes parameters (prior ?)
- investigate faceted approximation to the Fourier transform
 - ↝ reduce communications, facilitate load balancing
- faceted prior for joint calibration and imaging?

Conclusions and perspectives

Conclusions: faceted prior for wide-band imaging

- ✓ quality comparable to HyperSARA...
- ✓ for a much lower computing time (increased distribution flexibility)
- ✓ spectral faceting (Thouvenin et al. 2020)

Perspectives:

- limited to Stokes I for now
 - ↝ reconstruct other Stokes parameters (prior ?)
- investigate faceted approximation to the Fourier transform
 - ↝ reduce communications, facilitate load balancing
- faceted prior for joint calibration and imaging?

Thank you for your attention.

A distributed wideband imaging algorithm for radio interferometry

P.-A. THOUVENIN^{*}, A. ABDULAZIZ[†], M. JIANG[‡], A. DABBECH[†],
A. REPETTI[†], A. JACKSON[★], J.-P. THIRAN[§], Y. WIAUX[†]

^{*}Université de Lille, CNRS, Centrale Lille, UMR 9189 CRISTAL,

[†]Heriot-Watt University, [‡]Xidian University, [★]EPCC, [§]EPFL

RICOCHET kick-off meeting

MARCH 24, 2022

Backup slides

- ▶ References
- ▶ Reweighting algorithm
- ▶ PDFB algorithm
- ▶ Spectral faceting
- ▶ Real data

References I

- Abdulaziz, A., A. Dabbech, and Y. Wiaux (2019). "Wideband Super-resolution Imaging in Radio Interferometry via Low Rankness and Joint Average Sparsity Models (HyperSARA)". In: *Monthly Notices of the Royal Astronomical Society* 489.1, pp. 1230–1248.
- Candès, Emmanuel J., Michael B. Wakin, and Stephen P. Boyd (2008). "Enhancing sparsity by reweighted ℓ_1 minimization". In: *J. Fourier Anal. Appl.* 4.5–6, pp. 877–905.
- Carrillo, R. E., J. D. McEwen, and Y. Wiaux (2012). "Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging". In: *Monthly Notices of the Royal Astronomical Society* 426.2, pp. 1223–1234.
- Condat, Laurent (2013). "A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms". In: *J. Optim. Theory Appl.* 158, pp. 460–479.
- Dabbech, Arwa et al. (July 2017). "The w-effect in interferometric imaging: from a fast sparse measurement operator to superresolution". In: *Monthly Notices of the Royal Astronomical Society* 471.4, pp. 4300–4313.
- Ferrari, A. et al. (2015). "Multi-frequency image reconstruction for radio interferometry. A regularized inverse problem approach". In: *arXiv preprint arXiv:1504.06847*.
- Fessler, Jeffrey A. and Bradley P. Sutton (Jan. 2003). "Nonuniform fast Fourier transforms using min-max interpolation". In: *IEEE Trans. Signal Process.* 51.2, pp. 560–574.

References II

- Jiang, Ming, Jérôme Bobin, and Jean-Luc Starck (2017). "Joint Multichannel Deconvolution and Blind Source Separation". In: *SIAM J. Imaging Sciences* 10.4, pp. 1997–2021.
- Onose, Alexandru et al. (2016). In: *Scalable splitting algorithms for big-data interferometric imaging in the SKA era* 462.4, pp. 4314–4335.
- Prusa, Zdenek (2012). "Segmentwise discrete wavelet transform". PhD thesis. Brno university of technology.
- Tasse, C. et al. (Apr. 2018). "Faceting for direction-dependent spectral deconvolution". In: 611.1, A87.
- Thouvenin, Pierre-Antoine et al. (2020). "Parallel faceted imaging in radio interferometry via proximal splitting (Faceted HyperSARA): when precision meets scalability". In: *arXiv preprint*. Under revision. arXiv: 2003.07358 [astro-ph.IM].
- Vũ, Bang Cõng (Apr. 2013). "A splitting algorithm for dual monotone inclusions involving cocoercive operators". In: *Advances in Computational Mathematics* 38.3, pp. 667–681.
- Wenger, S. and M. Magnor (2014). *A sparse reconstruction algorithm for multi-frequency radio images*. Tech. rep. Computer Graphics Lab, TU Braunschweig.

Backup (Reweighting) I

Parameters: $T > 0, 0 < \underline{\xi}_{\text{rw}} < 1$

$t \leftarrow 0, \xi \leftarrow +\infty$

// Solve spectral sub-problem c

while ($t < T$) and ($\xi > \underline{\xi}_{\text{rw}}$) **do**

for $q = 1$ **to** Q **do**

 // Update weights (low-rankness prior)

$$\bar{\theta}_{c,q}^{(t)} = \bar{v}_{c,q} \left(|s_j(\mathbf{D}_q \widetilde{\mathbf{S}}_q \mathbf{X}_c^{(t)})| + \bar{v}_{c,q} \right)^{-1}, \quad \bar{v}_{c,q} > 0;$$

 // Update weights (joint-sparsity prior)

$$\theta_{c,q}^{(t)} = v_c \left(\|[\Psi_q^\top \mathbf{S}_q \mathbf{X}_c^{(t)}]_i\|_2 + v_c \right)^{-1}, \quad v_c > 0;$$

 // Run PDFB algorithm

$$(\mathbf{X}_c^{(t+1)}, \mathbf{P}_c^{(t+1)}, \mathbf{W}_c^{(t+1)}, \mathbf{v}_c^{(t+1)}) = \mathbf{PDFB}(\mathbf{X}_c^{(t)}, \mathbf{P}_c^{(t)}, \mathbf{W}_c^{(t)}, \mathbf{v}_c^{(t)}, \theta_c^{(t)}, \bar{\theta}_c^{(t)});$$

$$\xi = \|\mathbf{X}_c^{(t+1)} - \mathbf{X}_c^{(t)}\|_F / \|\mathbf{X}_c^{(t)}\|_F;$$

$$t \leftarrow t + 1;$$

Backup (PDFB) I

Data: $(\mathbf{y}_{l,b})_{l,b}, l \in \{1, \dots, L\}, b \in \{1, \dots, B\}$

Input: $\mathbf{X}^{(0)}, \mathbf{P}^{(0)} = (\mathbf{P}_q^{(0)})_q, \mathbf{W}^{(0)} = (\mathbf{W}_q^{(0)})_q, \mathbf{v}^{(0)} = (\mathbf{v}_{l,b}^{(0)})_{l,b}, \theta = (\theta_q)_{1 \leq q \leq Q}, \bar{\theta} = (\bar{\theta}_q)_{1 \leq q \leq Q}$

Parameters: $(\mathbf{D}_q)_q, (\mathbf{U}_{l,b})_{l,b}, \varepsilon = (\varepsilon_{l,b})_{l,b}, \mu, \bar{\mu}, \tau, \zeta, \eta, \kappa$

$k \leftarrow 0; \xi = +\infty$

$\check{\mathbf{X}}^{(0)} = \mathbf{X}^{(0)}$

$\mathbf{r}^{(0)} = (r_{l,b}^{(0)})_{l,b} \in \mathbb{R}^{LB}, \text{ with } r_{l,b}^{(0)} = \|\mathbf{y}_{l,b} - \Phi_{l,b} \mathbf{x}_l^{(0)}\|_2$

while $\left[(k < R_{\max}) \text{ and } (\xi > 10^{-5} \text{ or } \|\mathbf{r}^{(k)}\|_2 > 1.01 \|\varepsilon\|_2) \right]$ **do**

// Broadcast auxiliary variables

for $q = 1$ **to** Q **do**

$\widetilde{\mathbf{X}}_q^{(k)} = \widetilde{\mathbf{S}}_q \check{\mathbf{X}}^{(k)}; \check{\mathbf{X}}_q^{(k)} = \mathbf{S}_q \check{\mathbf{X}}^{(k)}$

Backup (PDFB) II

```

for  $l = 1$  to  $L$  do
   $\hat{\mathbf{x}}_l^{(k)} = \mathbf{FZ}\check{\mathbf{x}}_l^{(k)}$ ; // Fourier transforms
    for  $b = 1$  to  $B$  do
       $\hat{\mathbf{x}}_{l,b}^{(k)} = \mathbf{M}_{l,b}\hat{\mathbf{x}}_l^{(k)}$ ; // send to data cores
  // Update low-rankness variables [facet cores]
for  $q = 1$  to  $Q$  do
   $\mathbf{P}_q^{(k+1)} = (\mathbf{I}_{\widetilde{N}_q \times L} - \text{prox}_{\zeta^{-1}\bar{\mu}\|\cdot\|_{*,\bar{\theta}_q}})(\mathbf{P}_q^{(k)} + \mathbf{D}_q\widetilde{\mathbf{X}}_q^{(k)})$ 
   $\widetilde{\mathbf{P}}_q^{(k+1)} = \mathbf{D}_q^\dagger \mathbf{P}_q^{(k+1)}$ 
// Update sparsity variables [facet cores]
for  $q = 1$  to  $Q$  do
   $\mathbf{W}_q^{(k+1)} = (\mathbf{I}_{I_q \times L} - \text{prox}_{\kappa^{-1}\mu\|\cdot\|_{2,1,\theta_q}})(\mathbf{W}_q^{(k)} + \boldsymbol{\Psi}_q^\dagger \check{\mathbf{x}}_q^{(k)})$ 
   $\widetilde{\mathbf{W}}_q^{(k+1)} = \boldsymbol{\Psi}_q \mathbf{W}_q^{(k+1)}$ 

```

Backup (PDFB) III

// Update data fidelity variables [data cores]

for $(l, b) = (1, 1)$ **to** (L, B) **do**

$$\left| \begin{array}{l} \mathbf{v}_{l,b}^{(k+1)} = \mathbf{U}_{l,b} \left(\mathbf{I}_{M_{l,b}} - \text{prox}_{\ell_B(\mathbf{y}_{l,b}, \varepsilon_{l,b})} \right) \left(\mathbf{U}_{l,b}^{-1} \mathbf{v}_{l,b}^{(k)} + \boldsymbol{\Theta}_{l,b} \mathbf{G}_{l,b} \hat{\mathbf{x}}_{l,b}^{(k)} \right) \\ \widetilde{\mathbf{v}}_{l,b}^{(k+1)} = \mathbf{G}_{l,b}^\dagger \boldsymbol{\Theta}_{l,b}^\dagger \mathbf{v}_{l,b}^{(k+1)} \\ r_{l,b}^{(k+1)} = \|\mathbf{y}_{l,b} - \boldsymbol{\Theta}_{l,b} \mathbf{G}_{l,b} \hat{\mathbf{x}}_{l,b}^{(k)}\|_2 \end{array} \right.$$

// Inter node communications

for $l = 1$ **to** L **do**

$$\left| \mathbf{a}_l^{(k)} = \sum_{q=1}^Q \left(\zeta \widetilde{\mathbf{S}}_q^\dagger \widetilde{\mathbf{p}}_{q,l}^{(k+1)} + \kappa \mathbf{S}_q^\dagger \widetilde{\mathbf{w}}_{q,l}^{(k+1)} \right) + \eta \mathbf{Z}^\dagger \mathbf{F}^\dagger \sum_b \mathbf{M}_{l,b}^\dagger \widetilde{\mathbf{v}}_{l,b}^{(k+1)} \right.$$

// Update image tiles [on facet cores, in parallel]

$$\mathbf{X}^{(k+1)} = \text{prox}_{\ell_{\mathbb{R}^{N \times L}_+}} \left(\mathbf{X}^{(k)} - \tau \mathbf{A}^{(k)} \right);$$

$$\check{\mathbf{X}}^{(k+1)} = 2\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)};$$

$$\xi = \|\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)}\|_F / \|\mathbf{X}^{(k)}\|_F \quad k \leftarrow k + 1$$

$$\mathbf{A}^{(k)} = \left(\mathbf{a}_l^{(k)} \right)_{1 \leq l \leq L}$$

// communicate facet borders

Spectral faceting

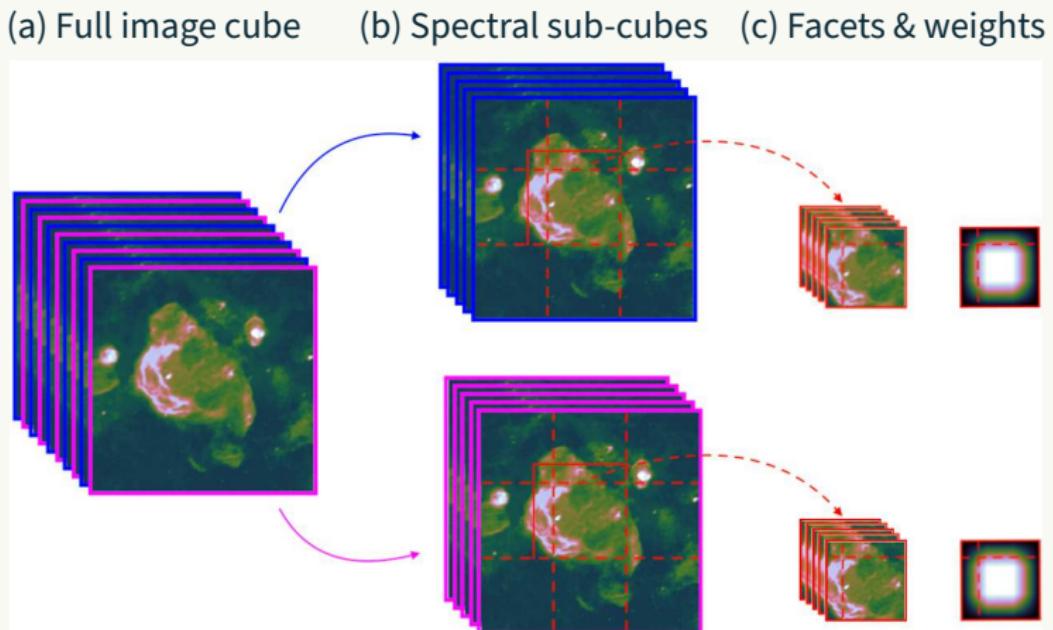


Figure 4: Illustration of the proposed faceting scheme.

Spectral facetting

	aSNR (dB)	aSNR _{log} (dB)	CPU cores	PDFB iterations	run _{pi} (s)	run (h)	cpu _{pi} (s)	cpu (h)
SARA	19.76 (± 3.19)	19.85 (± 1.80)	1200	2205	0.55 (± 0.05)	0.41	0.87 (± 0.05)	53.01
HS	22.27 (± 2.56)	23.57 (± 0.13)	16	3800	11.30 (± 1.01)	12.01	64.71 (± 2.42)	68.75
FHS ($C = 2$)	21.77 (± 2.51)	23.37 (± 1.19)	32	2400	5.68 (± 0.45)	3.80	32.25 (± 1.72)	43.18
FHS ($C = 5$)	21.85 (± 2.72)	24.73 (± 0.61)	80	2380	2.67 (± 0.44)	2.01	13.78 (± 1.17)	45.74
FHS ($C = 10$)	22.04 (± 2.85)	25.03 ($\pm 3.97e - 01$)	160	2540	1.53 (± 0.29)	1.36	7.04 (± 0.91)	49.58

Table 3: Spectral facetting: FHS with a varying number of spectral sub-problems C and $Q = 1$, compared to HyperSARA (= FHS with $Q = C = 1$) and SARA (=FHS with $Q = 1$ and $C = L$).

[Back to presentation](#)[Backup slide](#)

Real data (I)

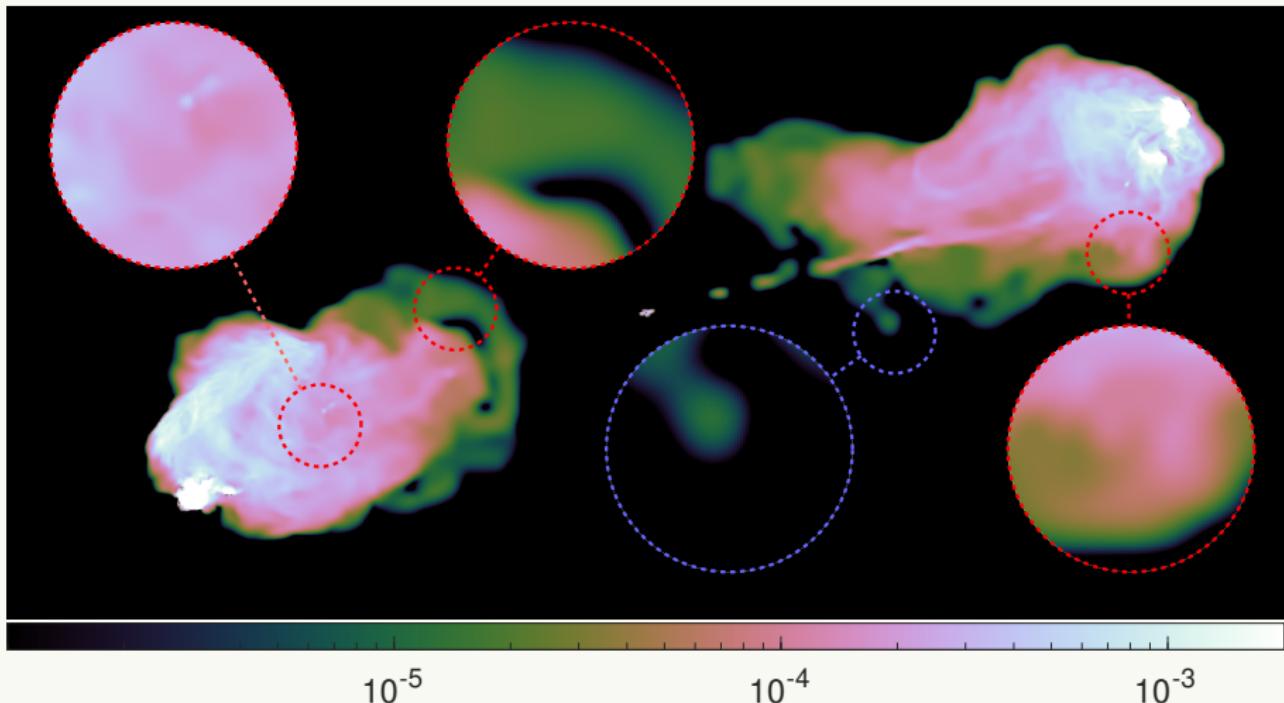


Figure 5: Cyg A (SARA), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_{241} = 5.899$ GHz.
Images in Jy/pixel, angular resolution 0.06'' (2.38x spatial bandwidth).

[Back to presentation](#)[Backup slide](#)

Real data (II)

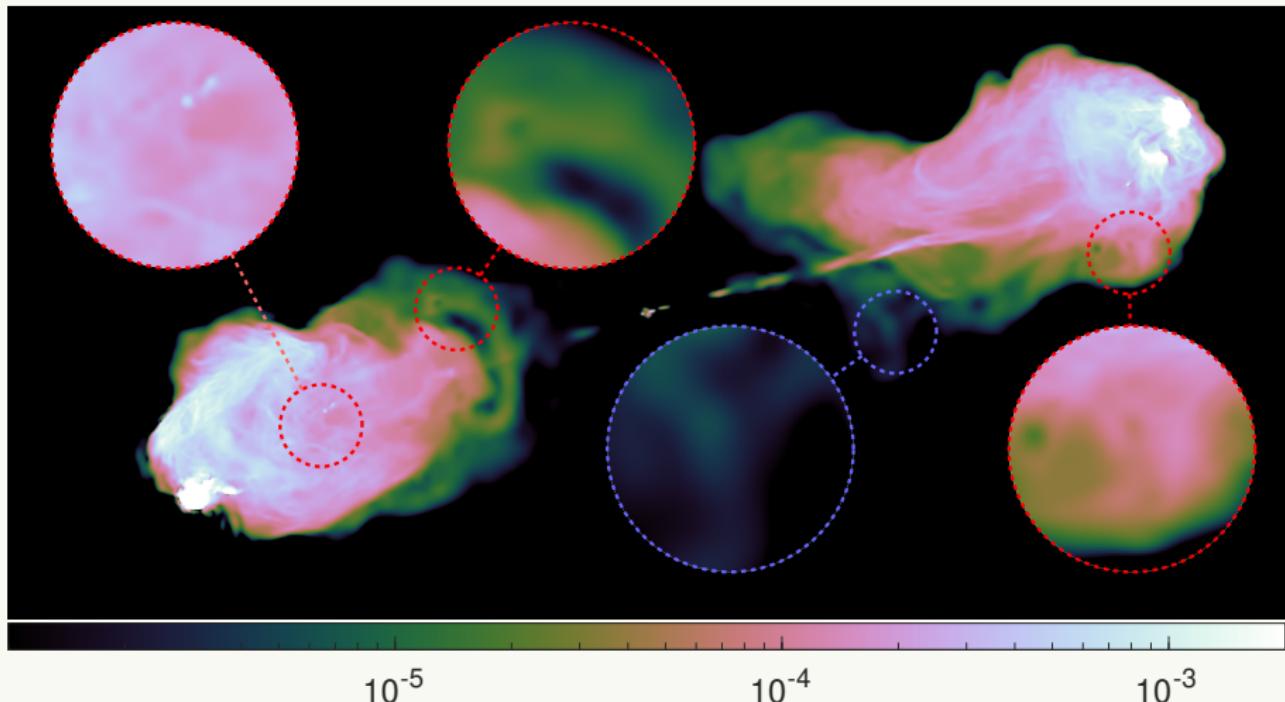


Figure 6: Cyg A (FHS, $Q = 60$ and $C = 16$), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_{241} = 5.899$ GHz. Images in Jy/pixel, angular resolution $0.06''$ (2.38x spatial bandwidth).

Real data (III)

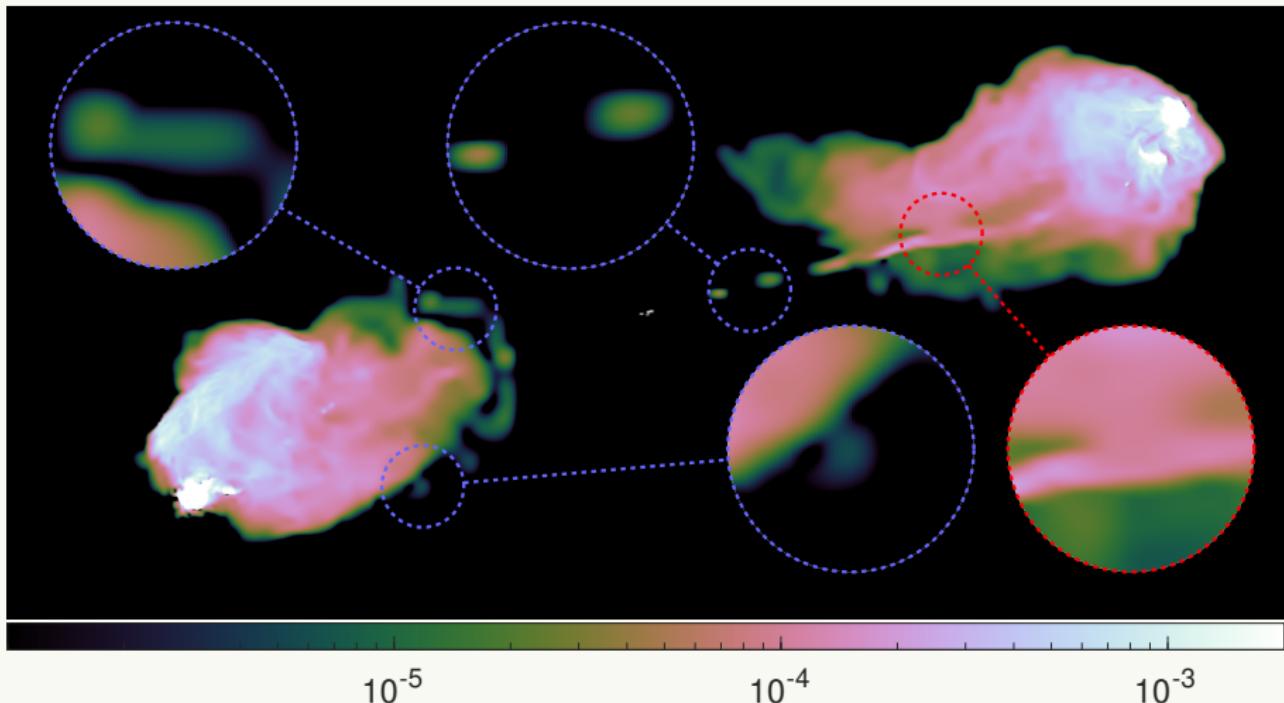


Figure 7: Cyg A (SARA), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_{480} = 8.019$ GHz.
Images in Jy/pixel, angular resolution $0.06''$ (1.75x spatial bandwidth).

[Back to presentation](#)[Backup slide](#)

Real data (IV)

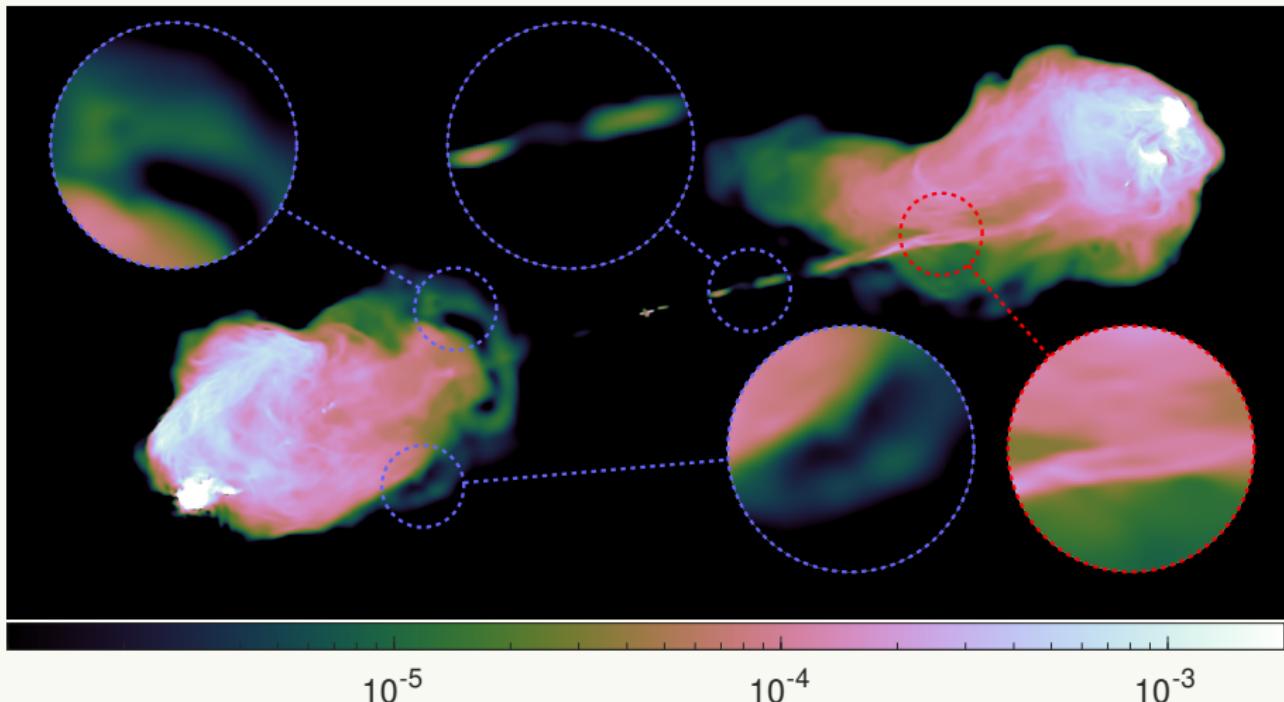


Figure 8: Cyg A (FHS, $Q = 60$ and $C = 16$), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_{480} = 8.019$ GHz. Images in Jy/pixel, angular resolution $0.06''$ (1.75 \times spatial bandwidth).

[Back to presentation](#)[Backup slide](#)

Real data (V)

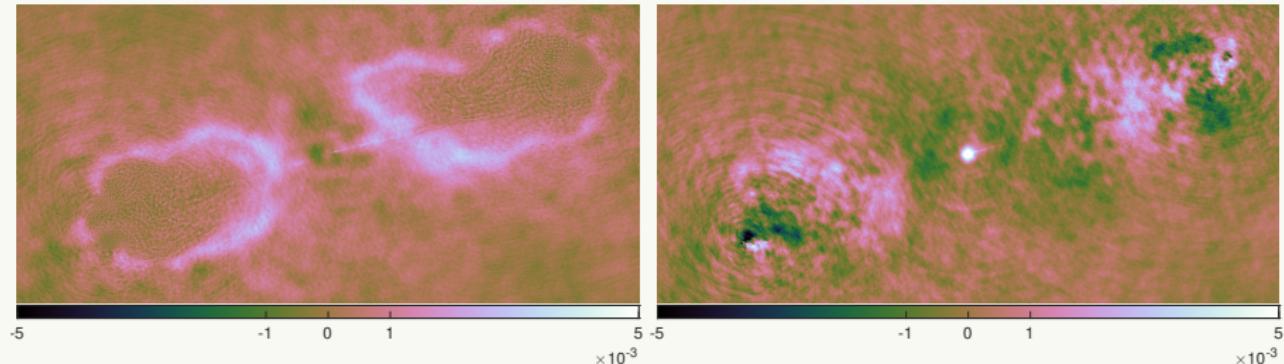


Figure 9: Cyg A, residual images in linear scale, $\nu_{241} = 5.899$ GHz. Left to right: SARA, FHS.

Real data (VI)

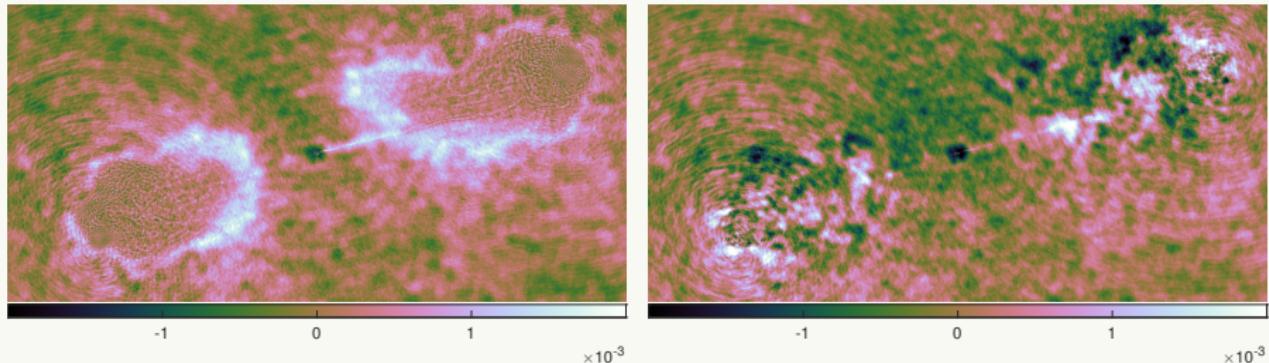


Figure 10: Cyg A, residual images in linear scale, $\nu_{480} = 8.019$ GHz. Left to right: SARA, FHS.