# Parallel faceted imaging in radio interferometry via proximal splitting (Faceted HyperSARA)

P.-A. Thouvenin[*,†], A. Abdulaziz[†], M. Jiang[‡,†], A. Dabbech[†], A. Repetti[†], A. Jackson[⋆], J.-P. Thiran[§], Y. Wiaux[†]

[*]Centrale Lille (CRIStAL), [†]Heriot-Watt University, [‡]Xidian University, [⋆]EPCC, [§]EPFL

URSI GASS 2021

September 4, 2021

# Table of contents

# Big data in the SKA era: a few perspectives

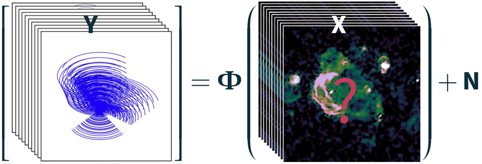


- Modern telescopes (*e.g.*, Square Kilometer Array (SKA): high imaging resolution and sensitivity
  - $>$ gigabyte size image per frequency
  - $> 10^4$ observation frequencies
  - dynamic range $10^7$
- Imaging challenge in bytes
  - petabyte size wide-band images
  - exabyte size data volumes (after 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 (*visibilities*) |
| $\mathbf{X} \in \mathbb{R}_+^{N \times L}$ | wide-band image cube |
| $\boldsymbol{\Phi}$ | measurement operator |
| $\mathbf{N} \in \mathbb{C}^{M \times L}$ | noise |

# Discrete measurement model



- **Measurement equation**:

$$\mathbf{Y} = \boldsymbol{\Phi}(\mathbf{X}) + \mathbf{N}$$
$$\mathbf{y}_l = \boldsymbol{\Phi}_l \mathbf{x}_l + \mathbf{n}_l, \quad \boldsymbol{\Phi}_l = \boldsymbol{\Theta}_l \mathbf{G}_l \mathbf{F} \mathbf{Z} \tag{1}$$

| | |
|---|---|
| $\mathbf{x}_l \in \mathbb{R}_+^N$ | 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) |
| $\boldsymbol{\Theta}_l \in \mathbb{R}^{M \times M}$ | 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



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

$f$ data fidelity term

(complex Gaussian noise $\Rightarrow$ $\ell_2$-norm ball constraint or quadratic term)

$$f\Big(\mathbf{Y},\boldsymbol{\Phi}(\mathbf{X})\Big)=\sum_{l=1}^{L}\iota_{\mathcal{B}(\mathbf{y}_l,\varepsilon_l)}\Big(\boldsymbol{\Phi}_l\mathbf{x}_l\Big)$$

$r$ regularization term

$\rightsquigarrow$ sparsity in a transformed domain (Wenger et al. 2014; Ferrari et al. 2015).
$\rightsquigarrow$ low-rankness (source separation model) (Jiang et al. 2017)
$\rightsquigarrow$ low-rankness + sparsity in a transformed domain (Abdulaziz et al. 2019)
(shown to yield a good quality wideband image in terms of both sensitivity and dynamic range)
...

1. How to deal with the volume of data ($M$ large)? (split $f$)
2. How to address large image sizes ($N$ large)? (split $r$)

# Towards a more scalable procedure

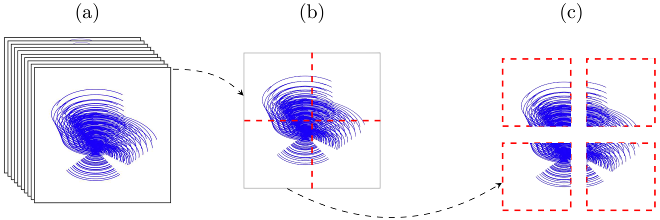


**Primary bottleneck**: data size (Onose et al. 2016)

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

$$f\Big(\mathbf{Y}, \boldsymbol{\Phi}(\mathbf{X})\Big) = \sum_{l=1}^{L} \sum_{b=1}^{B} \iota_{\mathcal{B}(\mathbf{y}_{l,b}, \varepsilon_{l,b})}\Big(\boldsymbol{\Phi}_{l,b}\mathbf{x}_l\Big)$$

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



## Selecting a prior $r$: building on the literature



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

- Low-rankness: sum of log functions acting on the singular values of $\mathbf{X}$
- Average joint-sparsity: sum of log functions acting on $\|[\boldsymbol{\Psi}^\dagger \mathbf{X}]_i\|_2$ ($\ell_2$ norm $i$th of $\boldsymbol{\Psi}^\dagger \mathbf{X}$)

| | |
|---|---|
| ⚠ ⚠ | full image cube $\mathbf{X}$ needed in a single place (SVD of $\mathbf{X}$) |
| $\boldsymbol{\Psi}^\dagger \in \mathbb{R}^{I \times N}$ | SARA dictionary (first 8 Daubechies wavelet and Dirac basis) |
| $[\mathbf{Z}]_i$ | $i$th row of $\mathbf{Z}$ |

# Parameter estimation



- **Log priors**: (2) not convex: use reweighting (Candès et al. 2008) (local majorant of $r$ at $\mathbf{X}^{(t)}$, $t \in \mathbb{N}$ current iteration index).

$$\underset{\mathbf{X} \in \mathbb{R}_+^{N \times L}}{\text{minimize}} \sum_{l,b} \iota_{\mathcal{B}(\mathbf{y}_{l,b}, \varepsilon_{l,b})}\big(\boldsymbol{\Phi}_{l,b}\mathbf{x}_l\big) + \bar{r}(\mathbf{X}, \mathbf{X}^{(t)}). \tag{3}$$

- Convex subproblem (4):

⇝ primal-dual forward-backward (PDFB) (Condat 2013; Vũ 2013)

⇝ no costly operator inversions or sub-iterations + splitting

⇝ handle non-smooth functions in parallel (through proximity operator)

# Image faceting

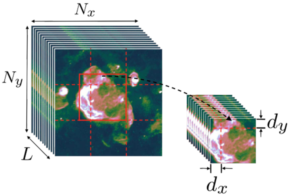


**Secondary bottleneck**: image size (focus in this presentation)

- 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

$\rightsquigarrow$ Motivation:
  - benefit from the same convergence guarantees as HyperSARA
  - keep reconstruction quality of HyperSARA
  - split image into 3D facets
  - assign portions of the image (facets) to different workers (*facet* cores)

$\Rightarrow$ Faceted HyperSARA



# Table of contents

# Spectral and spatial faceting



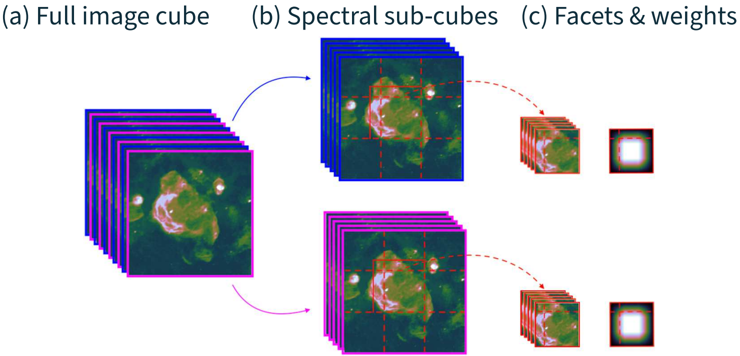



**Figure:** Illustration of the proposed faceting scheme.

- Spectral faceting: define interleaved groups of channels $\rightsquigarrow$ independent problems.
- Spatial faceting: tessellate the prior along the spatial dimension.

# Faceted HyperSARA prior



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

⚠⚠ full image cube $\mathbf{X}$ needed in a single place

$\boldsymbol{\Psi}^\dagger \in \mathbb{R}^{I \times N}$ SARA dictionary (first 8 Daubechies wavelets + Dirac)

# Faceted HyperSARA prior



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

| | |
|---|---|
| ✓ | spatial tessellation |
| $\boldsymbol{\Psi}_q^\dagger \in \mathbb{R}^{I_q \times N_q}$ | *exact* faceted implementation of $\boldsymbol{\Psi}^\dagger$ (Prusa 2012) |
| $\widetilde{\mathbf{S}}_q \in \mathbb{R}^{\widetilde{N}_q \times N}$, $\mathbf{S}_q \in \mathbb{R}^{N_q \times N}$ | content-agnostic facet selection operators |
| $\mathbf{D}_q$ | spatial weights (mitigate tessellation artefacts) |

⇝ Amount of overlap: free parameter for $\widetilde{\mathbf{S}}_q$, fixed for $\mathbf{S}_q$ (Prusa 2012);

⇝ Partially separable expression for the function $r$;

⇝ HyperSARA = faceted HyperSARA with $Q = 1$ facets.

**Parameter estimation**: same approach as for HyperSARA

- reweighting approach (to address log priors)
- convex sub-problems solved with PDFB

# Parameter estimation (PDFB)

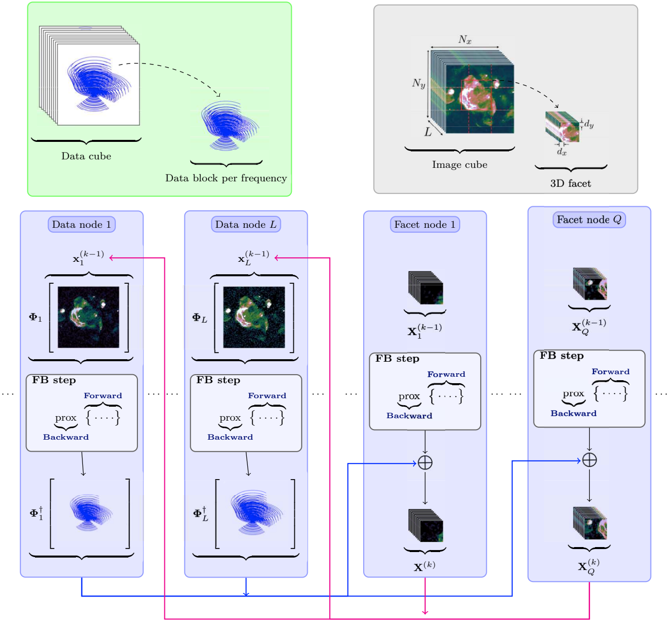


Split update of the auxiliary variables between two sets of cores:

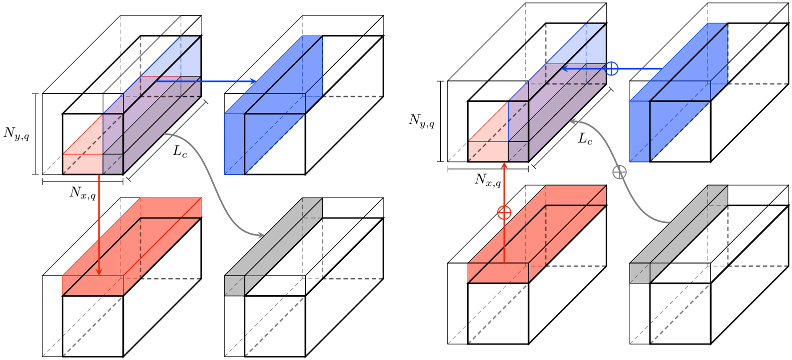
- data cores: contain data & auxiliary variables of full image size (few channels)
- facet cores: contain portions of the image cube (facet size over the full spectrum) + associated auxiliary variables

Most of the (dual) variables updated in parallel

Parallelization flexibility: adjust to the size of the problem ($N$, $L$, $M$)



Data cube
Data block per frequency
$N_x$
$N_y$
$L$
$d_y$
$d_x$
Image cube
3D facet
Data node 1
Data node $L$
Facet node 1
Facet node $Q$
$\mathbf{x}_1^{(k-1)}$
$\mathbf{x}_L^{(k-1)}$
$\mathbf{\Phi}_1$
$\mathbf{\Phi}_L$
$\mathbf{X}_1^{(k-1)}$
$\mathbf{X}_Q^{(k-1)}$
FB step
Forward
prox $\{\cdots\}$
Backward
$\mathbf{\Phi}_1^\dagger$
$\mathbf{\Phi}_L^\dagger$
$\mathbf{X}^{(k)}$
$\mathbf{X}_Q^{(k)}$




**(a)** Update borders (broadcast to neighbours) **(b)** Aggregate borders (from neighbour facets)

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

PDFB algo.

# Table of contents

# Validation on synthetic data



**Simulation settings**:

- synthetic wide-band image of Cyg A: power law spectral model, ground truth image S band (2 GHz) (Dabbech et al. 2021)
- $L = 20$ spectral channels in frequency range $[2.052, 3.572]$ GHz
- $N = 1024 \times 2048$ pixels
- $M \approx 7.62 \times 10^5$ measurements per channel, iSNR = 40 dB for each channel
- $B = 1$ data block per channel
- Comparison: SARA (Carrillo et al. 2012), HyperSARA (HS) (Abdulaziz et al. 2019) and Faceted HyperSARA (FHS).

**Assessment criteria**:

- average (over the channels) reconstruction SNR (aSNR, in dB)
- runtime per PDFB iteration ($\texttt{run}_{\texttt{pi}}$), active CPU time per iteration ($\texttt{cpu}_{\texttt{pi}}$)
- total runtime ($\texttt{run}$), total active CPU time ($\texttt{cpu}$)

# Varying number of facets



| | aSNR (dB) | CPU cores | PDFB iter. | $\mathtt{run}_{\mathtt{pi}}$ (s) | run (h) | $\mathtt{cpu}_{\mathtt{pi}}$ (s) | cpu (h) |
|---|---|---|---|---|---|---|---|
| SARA | **35.05** ($\pm$0.59) | 240 | 3275 | **3.28** ($\pm$0.38) | 3.38 | **7.13** ($\pm$0.95) | 129.77 |
| HS | **39.47** ($\pm$2.15) | 22 | 9236 | **25.36** ($\pm$0.85) | 65.06 | **84.49** ($\pm$2.79) | 216.76 |
| FHS ($Q = 4$) | **39.79** ($\pm$2.34) | 24 | 10989 | **26.50** ($\pm$1.88) | 80.90 | **184.41** ($\pm$9.22) | 562.90 |
| FHS ($Q = 9$) | **40.00** ($\pm$2.40) | 29 | 11009 | **15.38** ($\pm$1.38) | 47.04 | **226.52**($\pm$11.00) | 692.71 |
| FHS ($Q = 16$) | **40.08** ($\pm$2.40) | 36 | 10945 | **11.62** ($\pm$0.50) | 35.32 | **286.06** ($\pm$10.80) | 869.71 |

**Table:** Varying number of facets $Q$. SARA, HyperSARA (HS) and Faceted HyperSARA (FHS, overlap of 10%).

- SARA: 12 cores per channel (3 for the data-fidelity terms, 9 for the average sparsity)
- HS: 22 cores (20 for data-fidelity terms, primal variable and average joint-sparsity terms, 2 for the low-rank prior)
- FHS: 20 cores for the data-fidelity terms + 1 core per facet (primal variable, low-rank and joint average priors)

⚠ The implementation of HS is not equivalent to the implementation of FHS with $Q = 1$ (too slow in this case, $\mathtt{run}_{\mathtt{pi}} \approx 50$ s).

# Varying overlap between facets

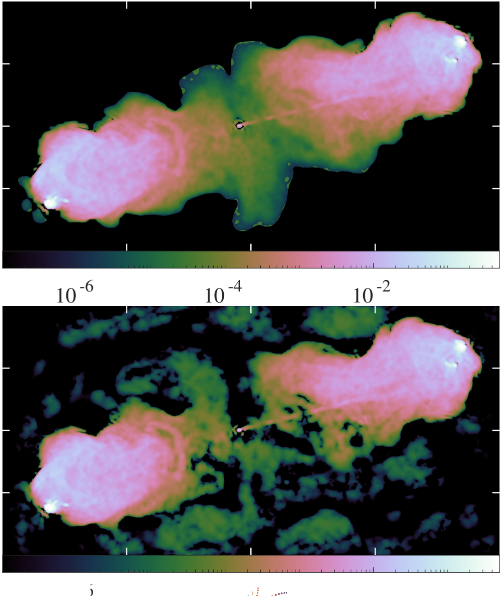


| | aSNR (dB) | CPU cores | PDFB iter. | $\text{run}_{\text{pi}}$ (s) | run (h) | $\text{cpu}_{\text{pi}}$ (s) | cpu (h) |
|---|---|---|---|---|---|---|---|
| SARA | **35.05** (**±0.59**) | 240 | 3275 | **3.28** (**±0.38**) | 3.38 | **7.13** (**±0.95**) | 129.77 |
| HS | **39.47** (**±2.15**) | 22 | 9236 | **25.36** (**±0.85**) | 65.06 | **84.49** (**±2.79**) | 216.76 |
| FHS (0% overlap) | **40.03** (**±2.41**) | 36 | 10961 | **11.55** (**±0.70**) | 35.18 | **284.17** (**±13.40**) | 865.22 |
| FHS (10% overlap) | **40.08** (**±2.40**) | 36 | 10945 | **11.62** (**±0.50**) | 35.32 | **286.06** (**±10.80**) | 869.71 |
| FHS (25% overlap) | **40.22** (**±2.41**) | 36 | 10918 | **11.96** (**±0.53**) | 36.26 | **290.71** (**±13.90**) | 881.66 |
| FHS (40% overlap) | **40.24** (**±2.42**) | 36 | 10934 | **12.67** (**±0.55**) | 38.47 | **298.32** (**±14.30**) | 906.08 |
| FHS (50% overlap) | **40.08** (**±2.53**) | 36 | 10962 | **13.69** (**±0.65**) | 41.68 | **311.14** (**±16.00**) | 947.41 |

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

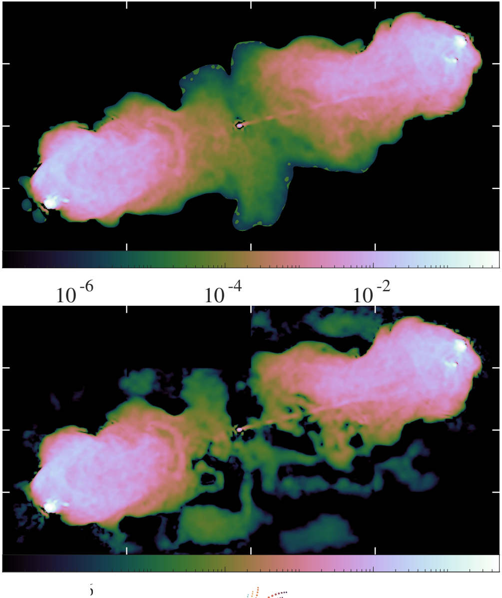
# Model image (truth / HS, ch. 20)





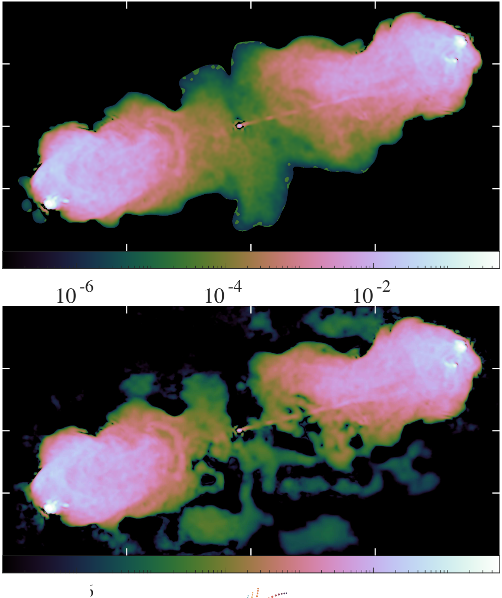
# Model image (truth / FHS, no overlap, ch. 20)





# Model image (truth / FHS, 10% overlap, ch. 20)





# Real data experiment



Imaging problem: 15 GB image cube of Cyg A from 7.4 GB of JVLA data

- Data acquired in 2015-2016 within 2–18 GHz (courtesy of R. Perley)
- Observations phase center: $\mathrm{RA} = 19\mathrm{h}\,59\mathrm{mn}\,28.356\mathrm{s}$ ($J$2000), $\mathrm{DEC} = +40°44'2.07''$
- 4 acquisitions instances: JVLA configurations A and C
  frequency ranges (GHz): $[\nu_1, \nu_{256}] = [3.979, 6.019]$,
  $[\nu_{257}, \nu_{480}] = [5.979, 8.019]$
- Channel-width $\delta\nu = 8$ MHz, total bandwidth of 4.04 GHz;
- Field-of-view (FoV): $\Omega_0 = 2.56' \times 1.536'$, pixel size $\delta x = 0.06''$
  $\rightsquigarrow N = 1536 \times 2560$
- $B = 2$ data blocks per channel (one per configuration)
- $Q = 3 \times 5$ facets, $C = 16$ subcubes (30 channels each)
- Pre-processing: monochromatic joint calibration and imaging ((Dabbech et al. 2021) used to initialize SARA and FHS (DDE + image))

# Real data (I)



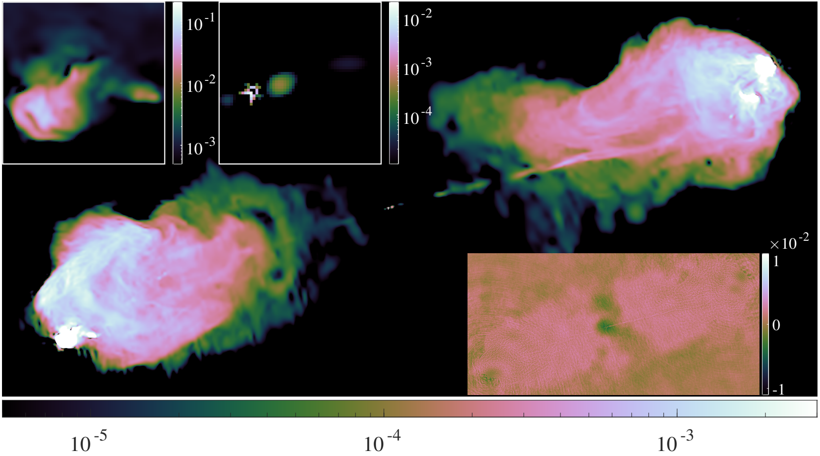



**Figure:** Cyg A (SARA), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_1 = 3.979$ GHz. Images in Jy/pixel, angular resolution 0.06″ (3.53× spatial bandwidth).

# Real data (II)



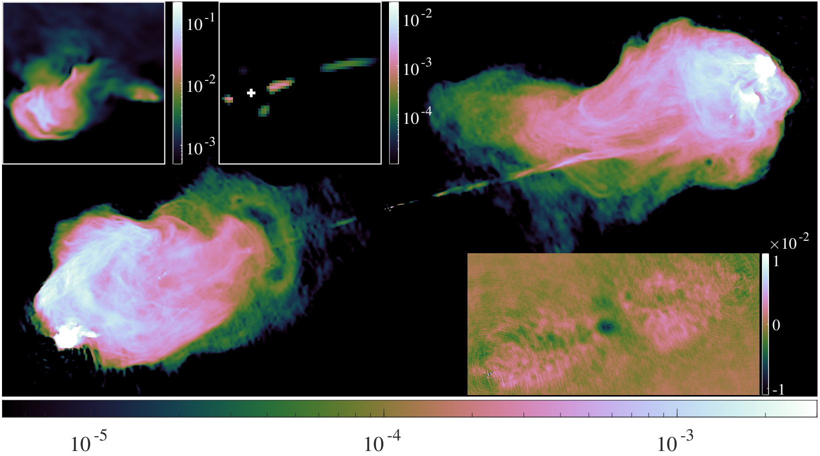



**Figure:** Cyg A (FHS), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_1 = 3.979$ GHz. Images in Jy/pixel, angular resolution 0.06″ (3.53× spatial bandwidth).

# Real data (III)



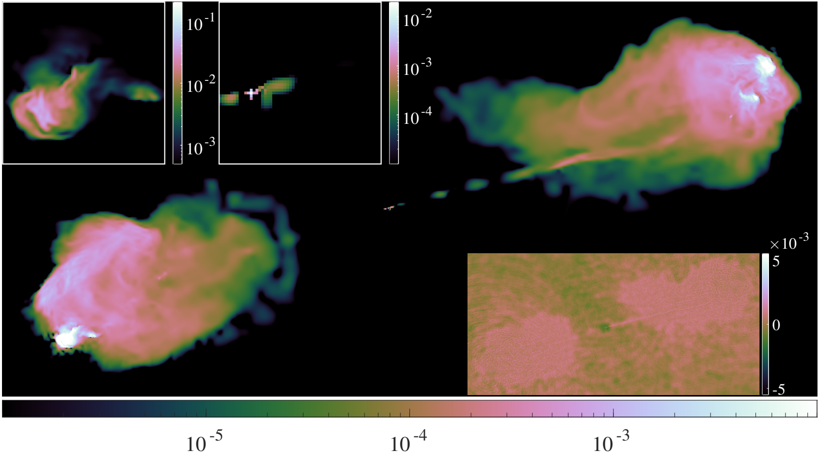



**Figure:** Cyg A (SARA), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_{480} = 8.019$ GHz. Images in Jy/pixel, angular resolution 0.06$''$ (1.75$\times$ spatial bandwidth).

# Real data (IV)



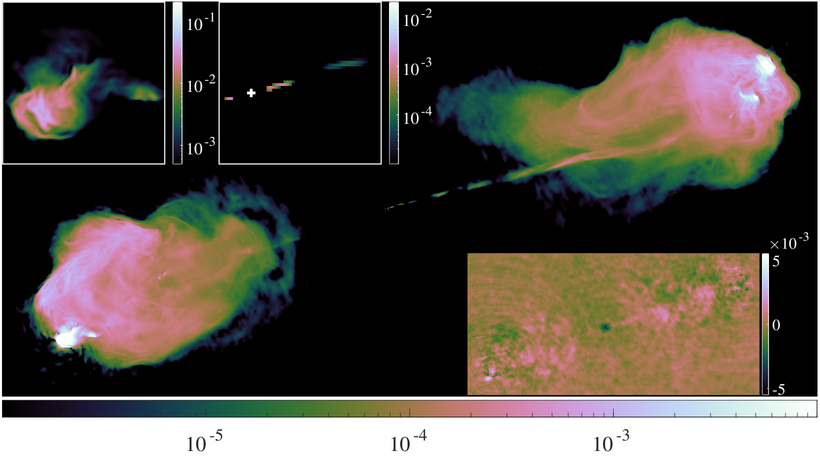



**Figure:** Cyg A (FHS), spectral resolution 8 MHz, 7.4 GB data, channel $\nu_{480} = 8.019$ GHz. Images in Jy/pixel, angular resolution 0.06″ (1.75× spatial bandwidth).

# Table of contents

# Conclusions and perspectives



**Conclusions**: faceted prior for wide-band imaging

- ✓ quality comparable to HyperSARA
- ✓ lower computing time (increased distribution flexibility)
- ✓ spectral faceting, possible combination with dim. reduction (Thouvenin et al. 2020) (not addressed today)

# Conclusions and perspectives



**Conclusions**: faceted prior for wide-band imaging

- ✓ quality comparable to HyperSARA
- ✓ lower computing time (increased distribution flexibility)
- ✓ spectral faceting, possible combination with dim. reduction (Thouvenin et al. 2020) (not addressed today)

**Perspectives**:

- → investigate faceted approximation to the Fourier transform
  - ⇝ reduce communications, facilitate load balancing
- → faceted prior for joint calibration and imaging?
  - ⇝ PDFB not applicable in this context.

# Conclusions and perspectives



**Conclusions**: faceted prior for wide-band imaging

- ✓ quality comparable to HyperSARA
- ✓ lower computing time (increased distribution flexibility)
- ✓ spectral faceting, possible combination with dim. reduction (Thouvenin et al. 2020) (not addressed today)

**Perspectives**:

- → investigate faceted approximation to the Fourier transform
  - ⇝ reduce communications, facilitate load balancing
- → faceted prior for joint calibration and imaging?
  - ⇝ PDFB not applicable in this context.

Thank you for your attention.

# Parallel faceted imaging in radio interferometry via proximal splitting (Faceted HyperSARA)

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

[*]Centrale Lille (CRIStAL), [†]Heriot-Watt University, [‡]Xidian University, [⋆]EPCC, [§]EPFL

URSI GASS 2021

SEPTEMBER 4, 2021

# Backup slides



- References
- Facet weights
- Priors: detailed expressions
- Reweighting algorithm (outer loop)
- PDFB algorithm (inner loop)
- More spatial faceting results
- Results: spectral faceting

# References I



Abdulaziz, A., A. Dabbech, and Y. Wiaux (2019). “Wideband Super-resolution Imaging in Radio Interferometry via Low Rankness and average joint-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 $\ell_1$ 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.

# References II



Dabbech, Arwa et al. (Jan. 2021). “Cygnus A jointly calibrated and imaged via non-convex optimisation from JVLA data”. In: *arXiv e-prints*. arXiv: `2102.00065 [astro-ph.IM]`.

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.

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.

# References III

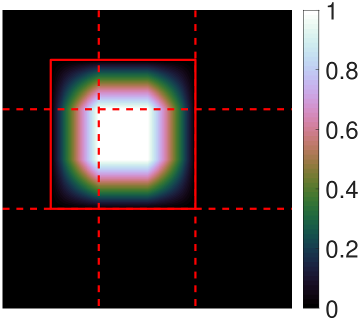

Thouvenin, Pierre-Antoine et al. (2020). “Parallel faceted imaging in radio interferometry via proximal splitting (Faceted HyperSARA): when precision meets scalability”. In: *arXiv preprint*. Under revision. arXiv: `2003.07358 [astro-ph.IM]`.

Vũ, Bang Công (Apr. 2013). “A splitting algorithm for dual monotone inclusions involving cocoercive operators”. In: *Advances in Computational Mathematics* 38.3, pp. 667–681.

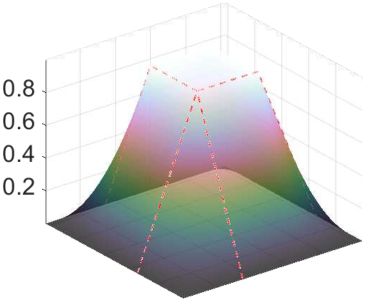
Wenger, S. and M. Magnor (2014). *A sparse reconstruction algorithm for multi-frequency radio images*. Tech. rep. Computer Graphics Lab, TU Braunschweig.

# Facet weights






**(a)** Weights for the central facet




**(b)** 3D view

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

# Backup (priors)



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

$$r(\mathbf{X}) = \mu \sum_{i=1}^{I} \upsilon \log\left(\frac{\|[\boldsymbol{\Psi}^\dagger \mathbf{X}]_i\|_2}{\upsilon} + 1\right) + \overline{\mu} \sum_{j=1}^{J} \overline{\upsilon} \log\left(\frac{|\sigma_j(\mathbf{X})|}{\overline{\upsilon}} + 1\right) \quad \text{(HyperSARA)}$$

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

# Parameter estimation



- **Log priors**: (2) not convex: use reweighting (Candès et al. 2008) (local majorant of $r$ at $\mathbf{X}^{(t)}$, $t \in \mathbb{N}$ current iteration index).

$$\underset{\mathbf{X}\in\mathbb{R}_+^{N\times L}}{\text{minimize}} \sum_{l,b} \iota_{\mathcal{B}(\mathbf{y}_{l,b},\varepsilon_{l,b})}\left(\mathbf{\Phi}_{l,b}\mathbf{x}_l\right) + r(\mathbf{X},\mathbf{X}^{(t)}). \tag{4}$$

**HyperSARA**:

$$r(\mathbf{X},\mathbf{X}^{(t)}) = \mu \|\mathbf{\Psi}^\dagger \mathbf{X}\|_{2,1,\omega(\mathbf{X}^{(t)})} + \overline{\mu}\|\mathbf{X}\|_{*,\overline{\omega}(\mathbf{X}^{(t)})}, \tag{5}$$

$$\omega_i(\mathbf{X}^{(t)}) = \upsilon\left(\|[\mathbf{\Psi}^\dagger\mathbf{X}^{(t)}]_i\|_2 + \upsilon\right)^{-1}, \tag{6}$$

$$\overline{\omega}_j(\mathbf{X}^{(t)}) = \overline{\upsilon}\left(\left|\sigma_j(\mathbf{X}^{(t)})\right| + \overline{\upsilon}\right)^{-1}. \tag{7}$$

- Convex subproblem (4):

$\rightsquigarrow$ primal-dual forward-backward (PDFB) (Condat 2013; Vũ 2013)

$\rightsquigarrow$ no costly operator inversions or sub-iterations + splitting

$\rightsquigarrow$ handle non-smooth functions in parallel (through proximity operator)

# Faceted HyperSARA prior



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

$$r(\mathbf{X}) = \mu \sum_{i=1}^{I} \upsilon \log\left(\frac{\|[\boldsymbol{\Psi}^\dagger \mathbf{X}]_i\|_2}{\upsilon} + 1\right) + \overline{\mu} \sum_{j=1}^{J} \overline{\upsilon} \log\left(\frac{|\sigma_j(\mathbf{X})|}{\overline{\upsilon}} + 1\right) \quad \text{(HyperSARA)}$$

⚠⚠ full image cube $\mathbf{X}$ needed in a single place

$\boldsymbol{\Psi}^\dagger \in \mathbb{R}^{I \times N}$ SARA dictionary (first 8 Daubechies wavelets + Dirac)

# Faceted HyperSARA prior



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

$$r(\mathbf{X}) = \sum_{q=1}^{Q} \left( \mu \sum_{i=1}^{I_q} \upsilon \log\left( \frac{\| [\boldsymbol{\Psi}_q^\dagger \mathbf{S}_q \mathbf{X}]_i \|_2}{\upsilon} + 1 \right) + \overline{\mu}_q \sum_{j=1}^{J_q} \overline{\upsilon}_q \log\left( \frac{|\sigma_j(\mathbf{D}_q \widetilde{\mathbf{S}}_q \mathbf{X})|}{\overline{\upsilon}_q} + 1 \right) \right)$$

(faceted HyperSARA)

| | |
|---|---|
| ✓ | spatial tessellation |
| $\boldsymbol{\Psi}_q^\dagger \in \mathbb{R}^{I_q \times N_q}$ | *exact* faceted implementation of $\boldsymbol{\Psi}^\dagger$ (Prusa 2012) |
| $\widetilde{\mathbf{S}}_q \in \mathbb{R}^{\widetilde{N}_q \times N}$, $\mathbf{S}_q \in \mathbb{R}^{N_q \times N}$ | content-agnostic facet selection operators |
| $\mathbf{D}_q$ | spatial weights (mitigate tessellation artefacts) |

⇝ Amount of overlap: free parameter for $\widetilde{\mathbf{S}}_q$, fixed for $\mathbf{S}_q$ (Prusa 2012);

⇝ Partially separable expression for the function $r$;

⇝ HyperSARA = faceted HyperSARA with $Q = 1$ facets.

# Backup (reweighting scheme)



**Data:** $(\mathbf{y}_{l,b})_{l,b}, c \in \{1, \ldots, C\}, l \in \{1, \ldots, L_c\}, b \in \{1, \ldots, B\}$
**Input:** $\mathbf{X}_c^{(0)}, \mathbf{P}_c^{(0)}, \mathbf{W}_c^{(0)}, \mathbf{v}_c^{(0)}$
**Parameters:** $T > 0$, $0 < \underline{\xi}_{\mathrm{rw}} < 1$

$t \leftarrow 0$, $\xi \leftarrow +\infty$
**while** $(t < T)$ *and* $(\xi > \underline{\xi}_{\mathrm{rw}})$ **do**
  **for** $q = 1$ **to** $Q$ **do**
    `// Update weights (low-rankness prior)`
    $\overline{\theta}_{c,q}^{(t)} = \overline{\omega}_{c,q}(\mathbf{X}_c^{(t)})$;
    `// Update weights (joint-sparsity prior)`
    $\theta_{c,q}^{(t)} = \omega_{c,q}(\mathbf{X}_c^{(t)})$;
  `// Run inner PDFB algorithm`
  $(\mathbf{X}_c^{(t+1)}, \mathbf{P}_c^{(t+1)}, \mathbf{W}_c^{(t+1)}, \mathbf{v}_c^{(t+1)}) = \mathbf{PDFB}\Big(\mathbf{X}_c^{(t)}, \mathbf{P}_c^{(t)}, \mathbf{W}_c^{(t)}, \mathbf{v}_c^{(t)}, \theta_c^{(t)}, \overline{\theta}_c^{(t)}\Big)$;
  $\xi = \|\mathbf{X}_c^{(t+1)} - \mathbf{X}_c^{(t)}\|_{\mathrm{F}} / \|\mathbf{X}_c^{(t)}\|_{\mathrm{F}}$;
  $t \leftarrow t + 1$;

# Backup (PDFB) I



**Data:** $(\mathbf{y}_{c,l,b})_{l,b}, l \in \{1,\ldots,L_c\}, b \in \{1,\ldots,B\}$

**Input:** $\mathbf{X}_c^{(0)}, \mathbf{P}_c^{(0)} = \left(\mathbf{P}_{c,q}^{(0)}\right)_q, \mathbf{W}_c^{(0)} = \left(\mathbf{W}_{c,q}^{(0)}\right)_q, \mathbf{v}_c^{(0)} = \left(\mathbf{v}_{c,l,b}^{(0)}\right)_{c,l,b},$
$\theta_c = \left(\theta_{c,q}\right)_{1\leq q\leq Q}, \overline{\theta}_c = \left(\overline{\theta}_{c,q}\right)_{1\leq q\leq Q}$

**Parameters:** $(\mathbf{D}_q)_q, (\mathbf{U}_{c,l,b})_{l,b}, \varepsilon = (\varepsilon_{c,l,b})_{l,b}, \mu_c, (\overline{\mu}_{c,q})_q, \tau, \zeta, (\eta_{c,l})_{1\leq l\leq L}, \kappa,$
$0 < P_{\min} < P_{\max}, 0 < \underline{\xi}_{\mathtt{pdfb}} < 1$

$p \leftarrow 0; \xi = +\infty;$
$\check{\mathbf{X}}_c^{(0)} = \mathbf{X}_c^{(0)}, \hat{\mathbf{X}}_c^{(0)} = (\hat{\mathbf{x}}_{c,l}^{(0)})_{1\leq l\leq L_c} = \mathbf{FZX}_c^{(0)};$
$\mathbf{r}_c^{(0)} = (r_{c,l,b}^{(0)})_{l,b} \in \mathbb{R}^{L_c B}$, with $r_{c,l,b}^{(0)} = \|\mathbf{y}_{c,l,b} - \boldsymbol{\Phi}_{c,l,b}\mathbf{x}_{c,l}^{(0)}\|_2;$
**while** $(p < P_{\min})$ *or* $\left[(p < P_{\max}) \text{ and } (\xi > \underline{\xi}_{\mathtt{pdfb}} \text{ or } \|\mathbf{r}_c^{(p)}\|_2 > 1.01\|\varepsilon\|_2)\right]$ **do**

```
// Update low-rankness and sparsity variables
```

split $(\widetilde{\mathbf{X}}_{c,q}^{(p)})_{1\leq q\leq Q} = (\widetilde{\mathbf{S}}_q\check{\mathbf{X}}_c^{(p)})_{1\leq q\leq Q};$
split $(\check{\mathbf{X}}_{c,q}^{(p)})_{1\leq q\leq Q} = (\mathbf{S}_q\check{\mathbf{X}}_c^{(p)})_{1\leq q\leq Q};$

# Backup (PDFB) II



```
// [Parallel on facet cores]
```

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

$$\mathbf{P}_{c,q}^{(p+1)} = \left(\mathbf{I}_{\widetilde{N}_q \times \widetilde{N}_q} - \operatorname{prox}_{\zeta^{-1}\overline{\mu}_c \|\cdot\|_{*,\overline{\theta}_{c,q}}}\right)\left(\mathbf{P}_{c,q}^{(p)} + \mathbf{D}_q \widetilde{\mathbf{X}}_{c,q}^{(p)}\right);$$

$$\widetilde{\mathbf{P}}_{c,q}^{(p+1)} = \mathbf{D}_q^{\dagger} \mathbf{P}_c^{(p+1)};$$

$$\mathbf{W}_{c,q}^{(p+1)} = \left(\mathbf{I}_{I_q \times I_q} - \operatorname{prox}_{\kappa^{-1}\mu_c \|\cdot\|_{2,1,\theta_{c,q}}}\right)\left(\mathbf{W}_{c,q}^{(p)} + \boldsymbol{\Psi}_q^{\dagger} \check{\mathbf{X}}_{c,q}^{(p)}\right);$$

$$\widetilde{\mathbf{W}}_{c,q}^{(p+1)} = \boldsymbol{\Psi}_q \mathbf{W}_{c,q}^{(p+1)};$$

```
// Update data fidelity variables [data cores]
```

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

$$\hat{\mathbf{x}}_{c,l}^{(p+1)} = \mathbf{FZx}_{c,l}^{(p)}$$

$$\text{split } (\hat{\mathbf{x}}_{c,l,b}^{(p+1)})_{1 \leq b \leq B} = (\mathbf{M}_{c,l,b} \hat{\mathbf{x}}_{c,l}^{(p+1)})_{1 \leq b \leq B};$$

**for** $b = 1$ **to** $B$ **do**

$$\mathbf{v}_{c,l,b}^{(p+1)} =$$

$$\mathbf{U}_{c,l,b}\left(\mathbf{I}_{M_{c,l,b}} - \operatorname{prox}^{\mathbf{U}_{c,l,b}}_{\iota_{\mathcal{B}(\mathbf{y}_{c,l,b}, \varepsilon_{c,l,b})}}\right)\left(\mathbf{U}_{c,l,b}^{-1} \mathbf{v}_{c,l,b}^{(p)} + \mathbf{G}_{c,l,b}\left(2\hat{\mathbf{x}}_{c,l,b}^{(p+1)} - \hat{\mathbf{x}}_{c,l,b}^{(p)}\right)\right);$$

$$\widetilde{\mathbf{v}}_{c,l,b}^{(p+1)} = \mathbf{G}_{c,l,b}^{\dagger} \mathbf{v}_{c,l,b}^{(p+1)};$$

$$r_{c,l,b}^{(p+1)} = \|\mathbf{y}_{c,l,b} - \mathbf{G}_{c,l,b} \hat{\mathbf{x}}_{c,l,b}^{(p+1)}\|_2;$$

# Backup (PDFB) III



```
// Inter node communications
```

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

$$\mathbf{a}_{c,l}^{(p)} = \sum_{q=1}^{Q} \Big( \zeta \widetilde{\mathbf{S}}_q^\dagger \widetilde{\mathbf{p}}_{c,q,l}^{(p+1)} + \kappa \mathbf{S}_q^\dagger \widetilde{\mathbf{w}}_{c,q,l}^{(p+1)} \Big) + \eta_{c,l} \mathbf{Z}^\dagger \mathbf{F}^\dagger \sum_b \mathbf{M}_{c,l,b}^\dagger \widetilde{\mathbf{v}}_{c,l,b}^{(p+1)};$$

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

$$\mathbf{X}_c^{(p+1)} = \operatorname{prox}_{\iota_{\mathbb{R}_+^{N \times L_c}}} \Big( \mathbf{X}_c^{(p)} - \tau \mathbf{A}_c^{(p)} \Big); \qquad // \ \mathbf{A}_c^{(p)} = \big( \mathbf{a}_{c,l}^{(p)} \big)_{1 \leq l \leq L}$$

$$\check{\mathbf{X}}_c^{(p+1)} = 2\mathbf{X}_c^{(p+1)} - \mathbf{X}_c^{(p)}; \qquad // \text{ communicate facet borders}$$

$$\xi = \|\mathbf{X}_c^{(p+1)} - \mathbf{X}_c^{(p)}\|_\mathrm{F} / \|\mathbf{X}_c^{(p)}\|_\mathrm{F};$$

$$p \leftarrow p + 1;$$

# Varying number of facets



| | aSNR (dB) | CPU cores | PDFB iter. | $\mathtt{run}_{\mathtt{pi}}$ (s) | run (h) | $\mathtt{cpu}_{\mathtt{pi}}$ (s) | cpu (h) |
|---|---|---|---|---|---|---|---|
| SARA | **35.05** ($\pm$0.59) | 240 | 3275 | **3.28** ($\pm$0.38) | 3.38 | **7.13** ($\pm$0.95) | 129.77 |
| HS | **39.47** ($\pm$2.15) | 22 | 9236 | **25.36** ($\pm$0.85) | 65.06 | **84.49** ($\pm$2.79) | 216.76 |
| FHS ($Q = 4$) | **39.79** ($\pm$2.34) | 24 | 10989 | **26.50** ($\pm$1.88) | 80.90 | **184.41** ($\pm$9.22) | 562.90 |
| FHS ($Q = 9$) | **40.00** ($\pm$2.40) | 29 | 11009 | **15.38** ($\pm$1.38) | 47.04 | **226.52**($\pm$11.00) | 692.71 |
| FHS ($Q = 16$) | **40.08** ($\pm$2.40) | 36 | 10945 | **11.62** ($\pm$0.50) | 35.32 | **286.06** ($\pm$10.80) | 869.71 |

**Table:** Varying number of facets $Q$. SARA, HyperSARA (HS) and Faceted HyperSARA (FHS, overlap of 10%).

- SARA: 12 cores per channel (3 for the data-fidelity terms, 9 for the average sparsity)
- HS: 22 cores (20 for data-fidelity terms, primal variable and average joint-sparsity terms, 2 for the low-rank prior)
- FHS: 20 cores for the data-fidelity terms + 1 core per facet (primal variable, low-rank and joint average priors)

⚠ The implementation of HS is not equivalent to the implementation of FHS with $Q = 1$ (too slow in this case, $\mathtt{run}_{\mathtt{pi}} \approx 50$ s).

# Varying overlap between facets



| | aSNR (dB) | CPU cores | PDFB iter. | $\mathtt{run_{pi}}$ (s) | run (h) | $\mathtt{cpu_{pi}}$ (s) | cpu (h) |
|---|---|---|---|---|---|---|---|
| SARA | **35.05 (**$\pm$0.59**)** | 240 | 3275 | **3.28 (**$\pm$0.38**)** | 3.38 | **7.13 (**$\pm$0.95**)** | 129.77 |
| HS | **39.47 (**$\pm$2.15**)** | 22 | 9236 | **25.36 (**$\pm$0.85**)** | 65.06 | **84.49 (**$\pm$2.79**)** | 216.76 |
| FHS (0% overlap) | **40.03 (**$\pm$2.41**)** | 36 | 10961 | **11.55 (**$\pm$0.70**)** | 35.18 | **284.17 (**$\pm$13.40**)** | 865.22 |
| FHS (10% overlap) | **40.08 (**$\pm$2.40**)** | 36 | 10945 | **11.62 (**$\pm$0.50**)** | 35.32 | **286.06 (**$\pm$10.80**)** | 869.71 |
| FHS (25% overlap) | **40.22 (**$\pm$2.41**)** | 36 | 10918 | **11.96 (**$\pm$0.53**)** | 36.26 | **290.71 (**$\pm$13.90**)** | 881.66 |
| FHS (40% overlap) | **40.24 (**$\pm$2.42**)** | 36 | 10934 | **12.67 (**$\pm$0.55**)** | 38.47 | **298.32 (**$\pm$14.30**)** | 906.08 |
| FHS (50% overlap) | **40.08 (**$\pm$2.53**)** | 36 | 10962 | **13.69 (**$\pm$0.65**)** | 41.68 | **311.14 (**$\pm$16.00**)** | 947.41 |

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

# Spectral faceting



| | aSNR (dB) | CPU cores | PDFB iter. | $\texttt{run}_{\texttt{pi}}$ (s) | run (h) | $\texttt{cpu}_{\texttt{pi}}$ (s) | cpu (h) |
|---|---|---|---|---|---|---|---|
| SARA | **19.76 (**$\pm$3.19**)** | 1200 | 2205 | **0.55 (**$\pm$0.046**)** | 0.41 | **0.87 (**$\pm$0.056**)** | 53.01 |
| HS | **22.27 (**$\pm$2.56**)** | 16 | 3800 | **11.30 (**$\pm$1.01**)** | 12.01 | **64.71 (**$\pm$2.42**)** | 68.75 |
| FHS ($C = 2$) | **21.77 (**$\pm$2.51**)** | 32 | 2400 | **5.68 (**$\pm$0.45**)** | 3.80 | **32.25 (**$\pm$1.72**)** | 43.18 |
| FHS ($C = 5$) | **21.85 (**$\pm$2.72**)** | 80 | 2380 | **2.67 (**$\pm$0.44**)** | 2.01 | **13.78 (**$\pm$1.17**)** | 45.74 |
| FHS ($C = 10$) | **22.04 (**$\pm$2.85**)** | 160 | 2540 | **1.53 (**$\pm$0.29**)** | 1.36 | **7.04 (**$\pm$0.91**)** | 49.58 |

**Table:** Spectral faceting: 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 $C = L$).