Parallelizing radio-interferometric image reconstruction by baselines

Sunrise Wang, Simon Prunet, Shan Mignot, André Ferrari

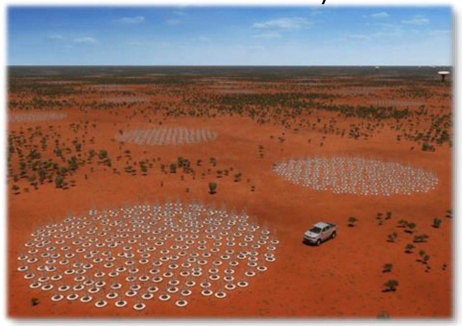
Outline

- Introduction
- Partitioning image reconstruction by baseline
- Parallelizing image reconstruction by baseline

Introduction

Antenna arrays

SKA-Low



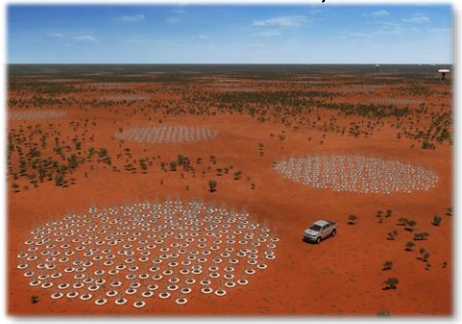
Radio dishes

SKA-Mid

Introduction

SKA-Low

Antenna arrays



~0.7 TB/s

Correlator & Beamformer

~0.3 TB/s

Imaging supercomputer: SDP-Low

SKA-Mid

Radio dishes

~2.4 TB/s

Correlator

~1.1 TB/s

Imaging supercomputer: SDP-Mid

image sources: [1, 2, 3, 4, 5]
data source: [6, 7]

[1] Farnes, Jamie, et al. "Science pipelines for the square kilometre array." Galaxies 6.4 (2018): 120.
[2] https://www.kit.edu/kit/english/pi_2021_059_kit-supercomputer-one-of-the-15-fastest-in-europe.php
[3] https://www.technologyreview.com/2023/09/21/1079909/whats-next-for-the-worlds-fastest-supercomputers/
[4] https://www.eso.org/public/france/images/eso1253c
[5] https://www.atnf.csiro.au/projects/askap/news_commissioning_24022014.html
[6] Labate, Maria G., et al. "Highlights of the square kilometre array low frequency (SKA-LOW) telescope." Journal of Astronomical Telescopes, Instruments, and Systems 8.1 (2022): 011024-011024.
[7] Swart, Gerhard P., Peter E. Dewdney, and Andrea Cremonini. "Highlights of the SKA1-Mid telescope architecture." Journal of Astronomical Telescopes, Instruments, and Systems 8.1 (2022): 011021-011021.

Scaling the computation

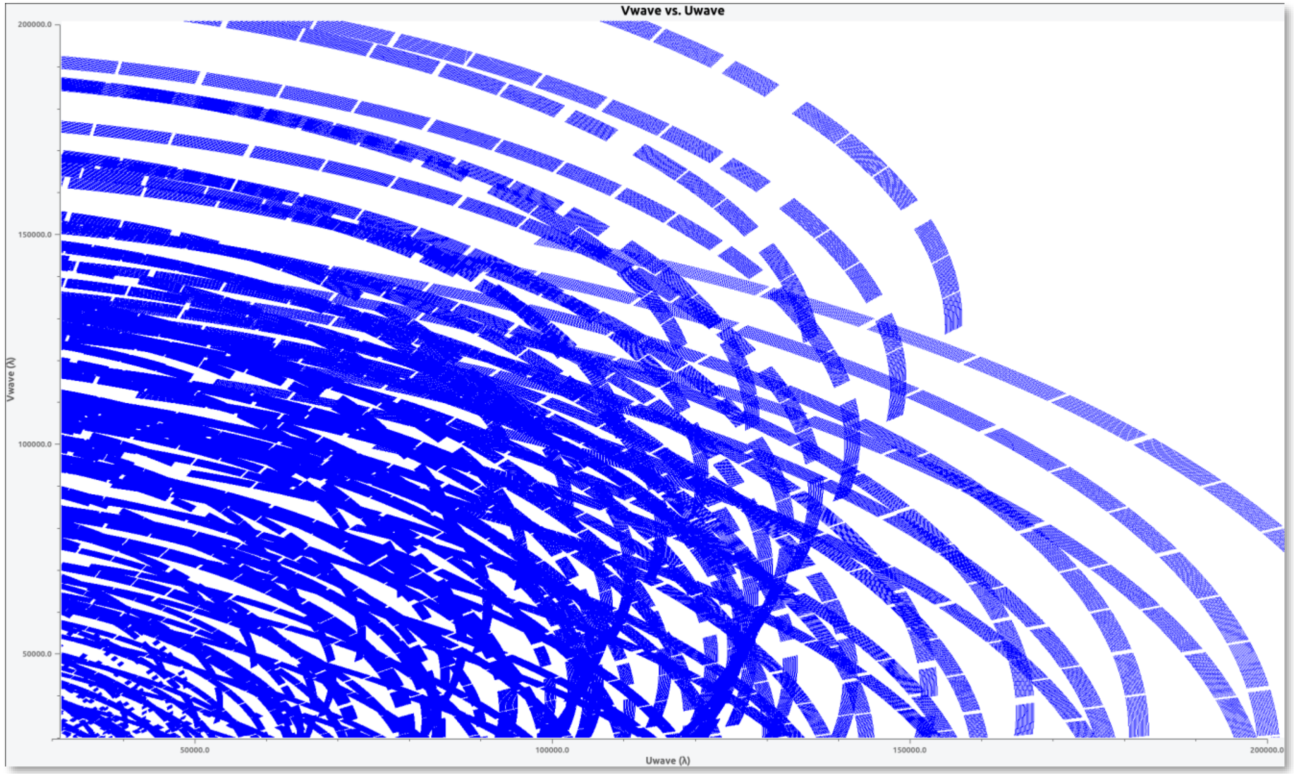
- Need to scale the computation, particularly de/gridding
  - Shown to take up to 94% of the time for a serial implementation[1]
- Partition visibilities, process separately

[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

Scaling the computation

- Need to scale the computation, particularly de/gridding
    - Shown to take up to 94% of the time for a serial implementation[1]
- Partition visibilities, process separately

Partition & Parallelization strategies:



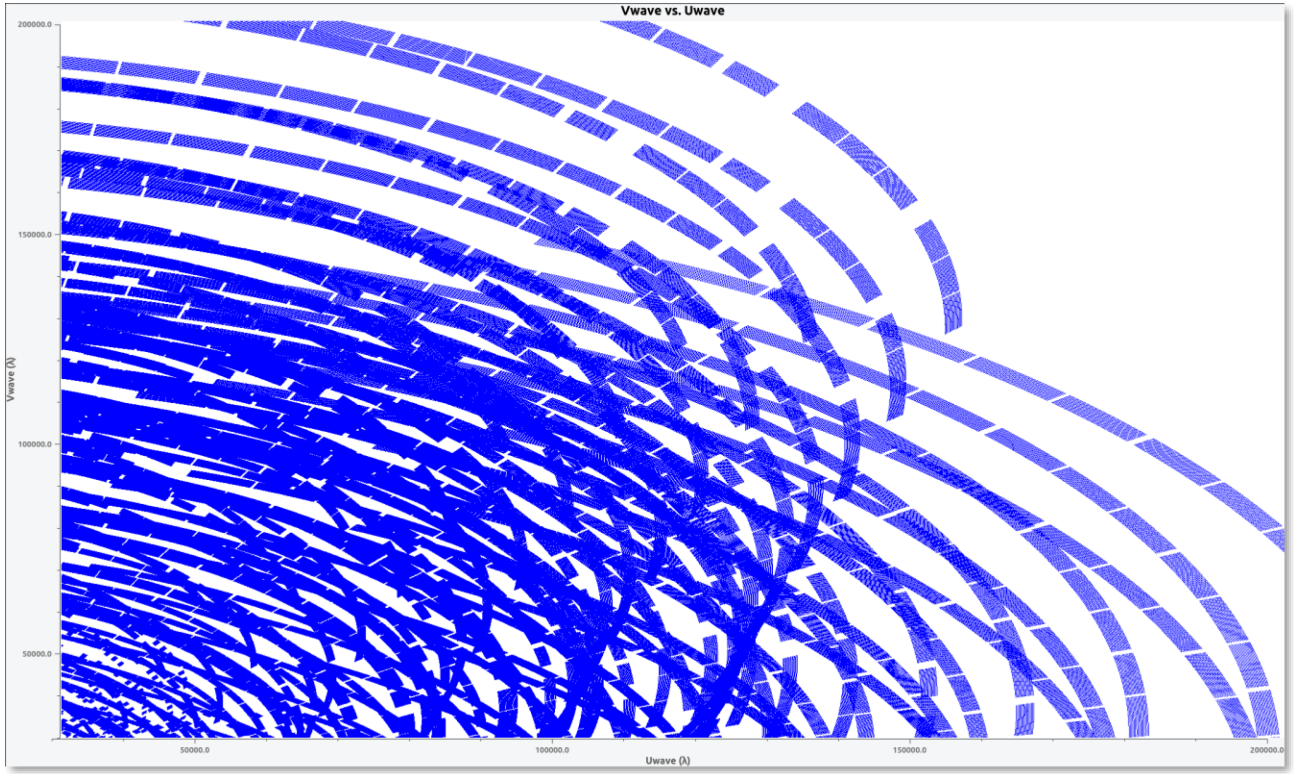
[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

Scaling the computation

- Need to scale the computation, particularly de/gridding
  - Shown to take up to 94% of the time for a serial implementation[1]
- Partition visibilities, process separately

Partition & Parallelization strategies:

- By frequency



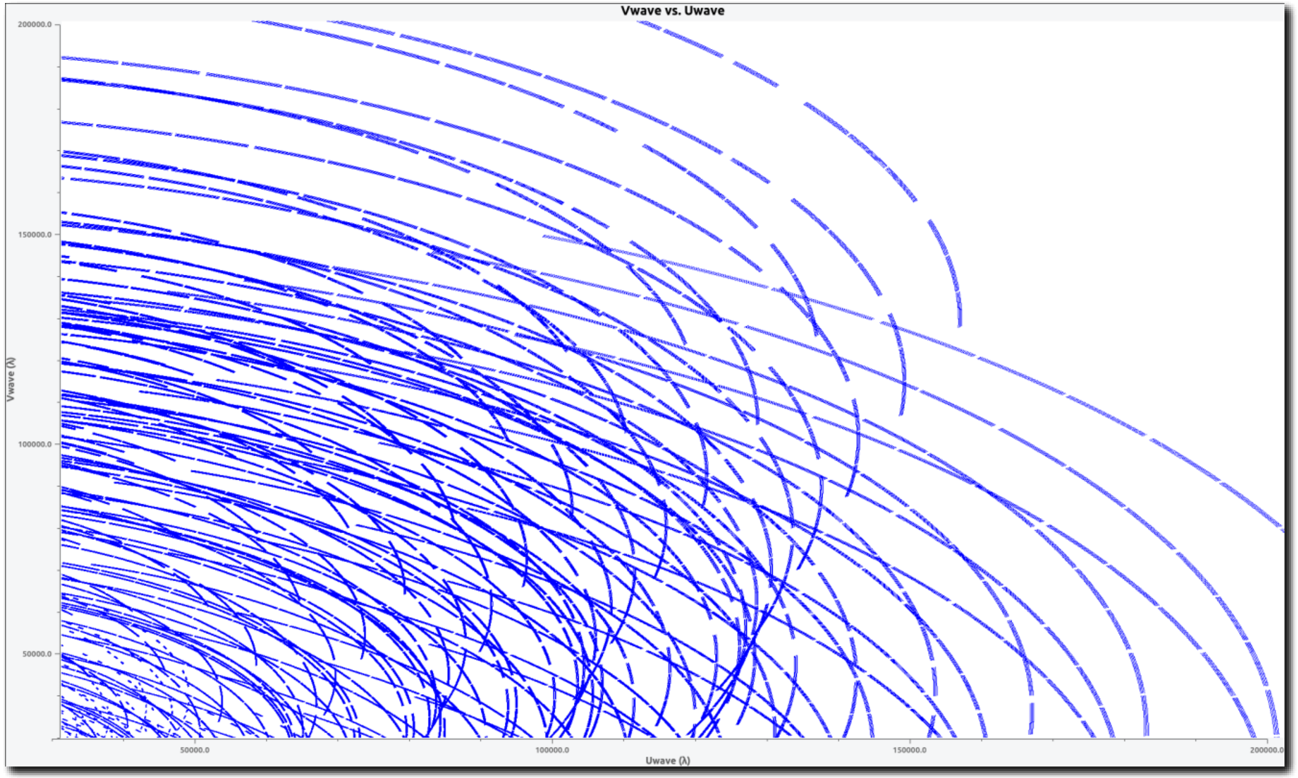
[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

Scaling the computation

- Need to scale the computation, particularly de/gridding
  - Shown to take up to 94% of the time for a serial implementation[1]
- Partition visibilities, process separately

Partition & Parallelization strategies:

- By frequency



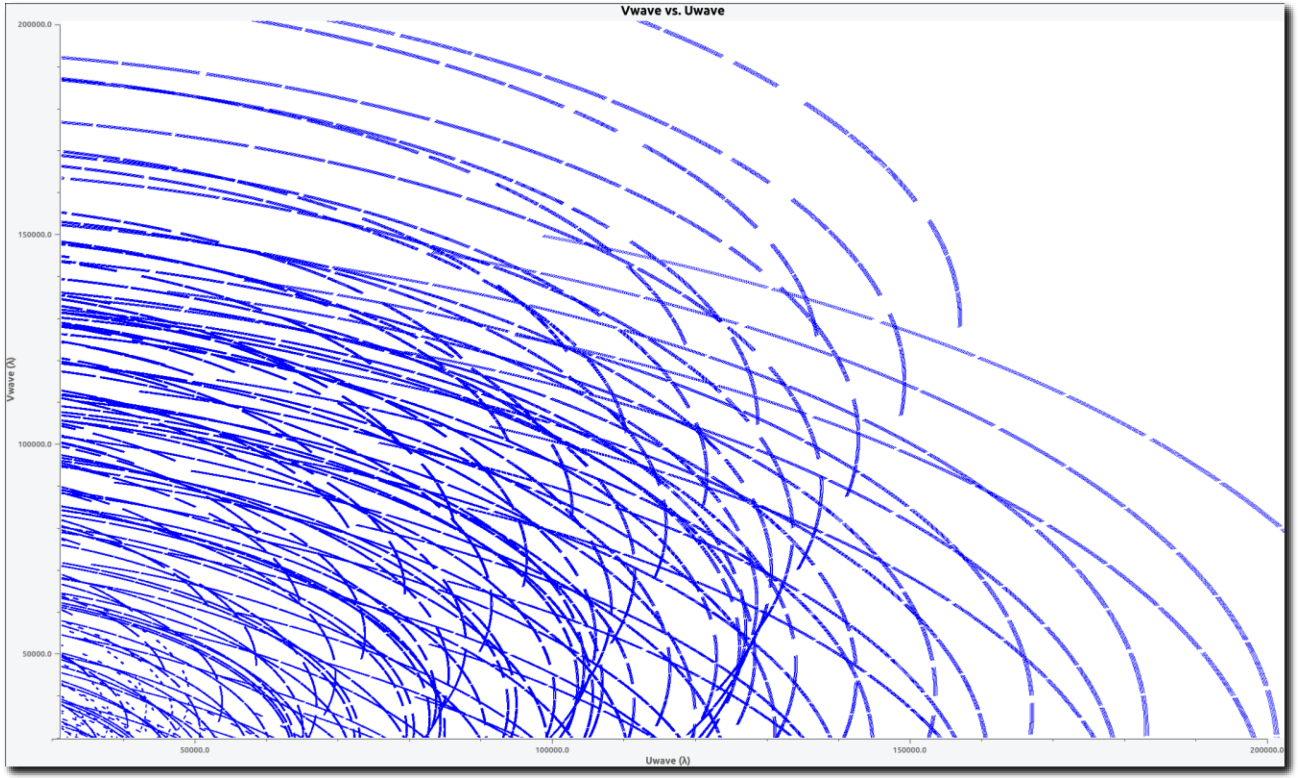
[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

Scaling the computation

- Need to scale the computation, particularly de/gridding
    - Shown to take up to 94% of the time for a serial implementation[1]
- Partition visibilities, process separately

Partition & Parallelization strategies:

- By frequency
- By time



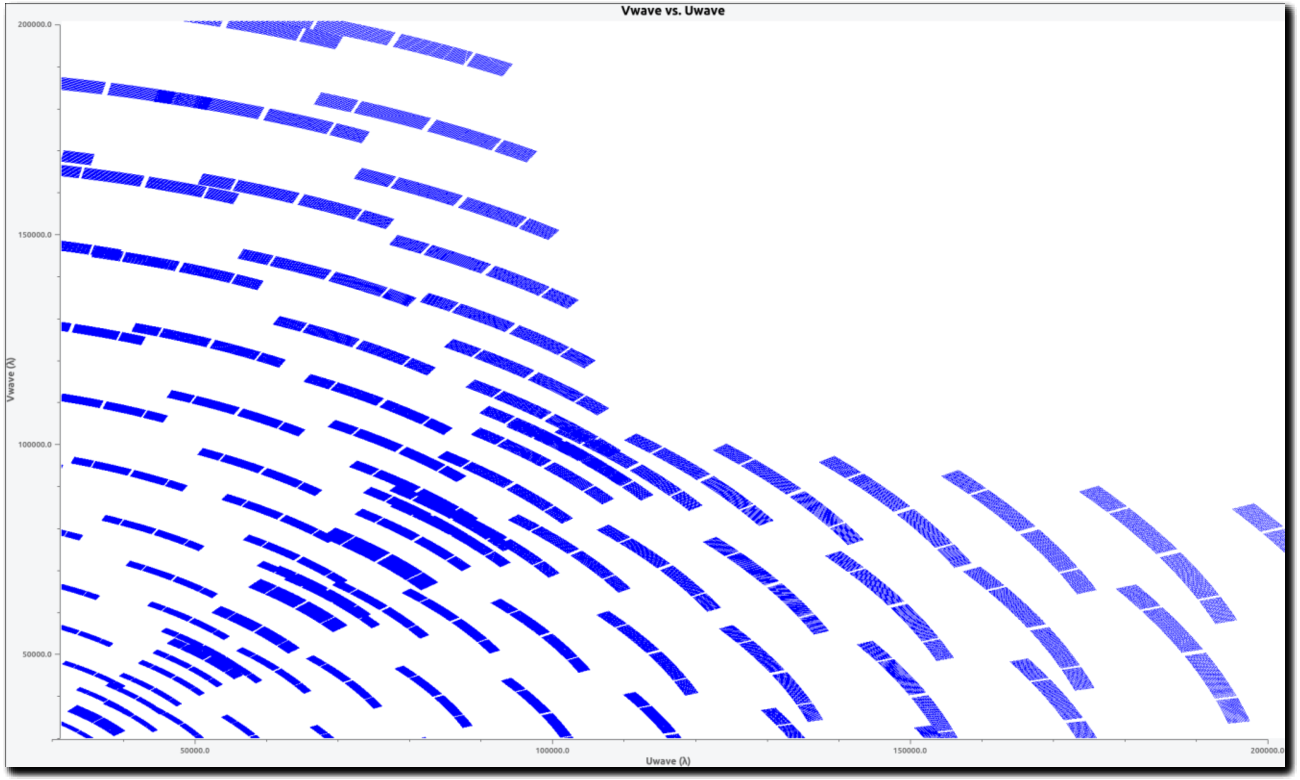
[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

Scaling the computation

- Need to scale the computation, particularly de/gridding
  - Shown to take up to 94% of the time for a serial implementation[1]
- Partition visibilities, process separately

Partition & Parallelization strategies:

- By frequency
- By time

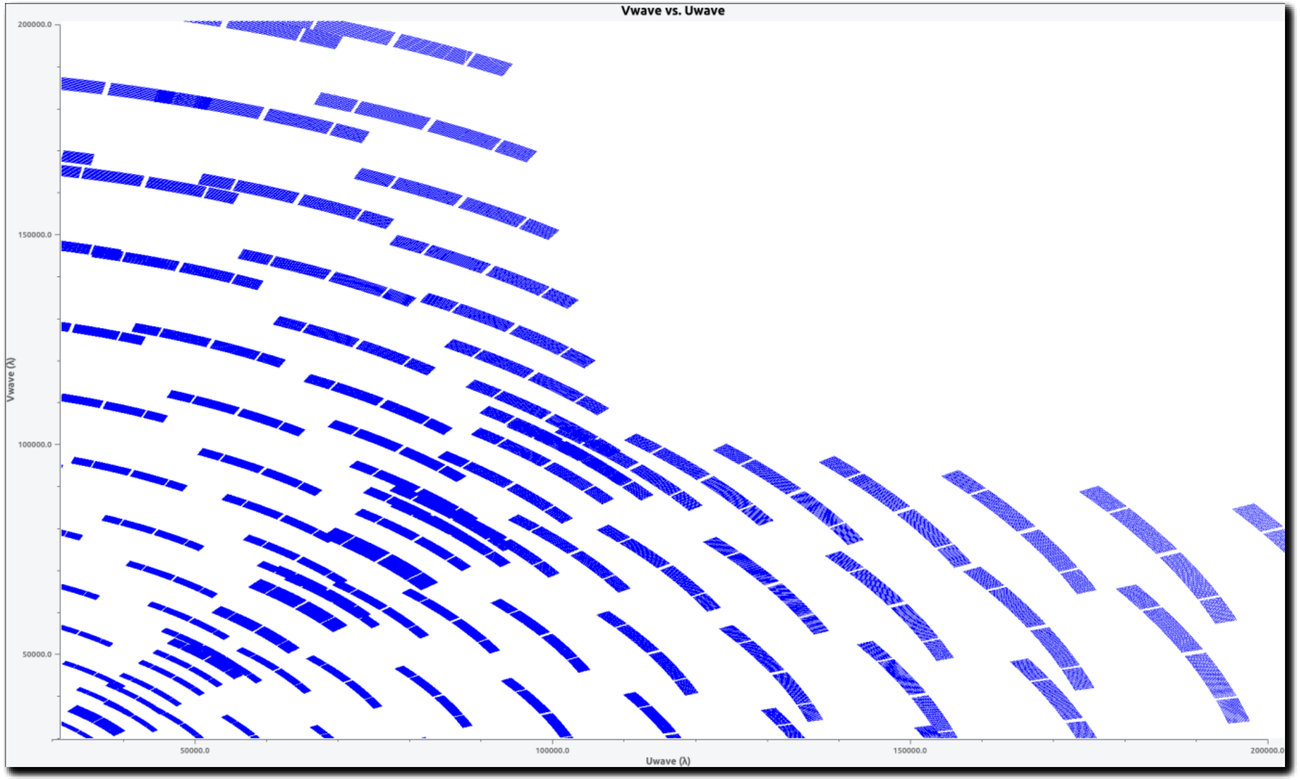


[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

Scaling the computation

- Need to scale the computation, particularly de/gridding
  - Shown to take up to 94% of the time for a serial implementation[1]
- Partition visibilities, process separately

Partition & Parallelization strategies:
- By frequency
- By time
- What about by baseline length?



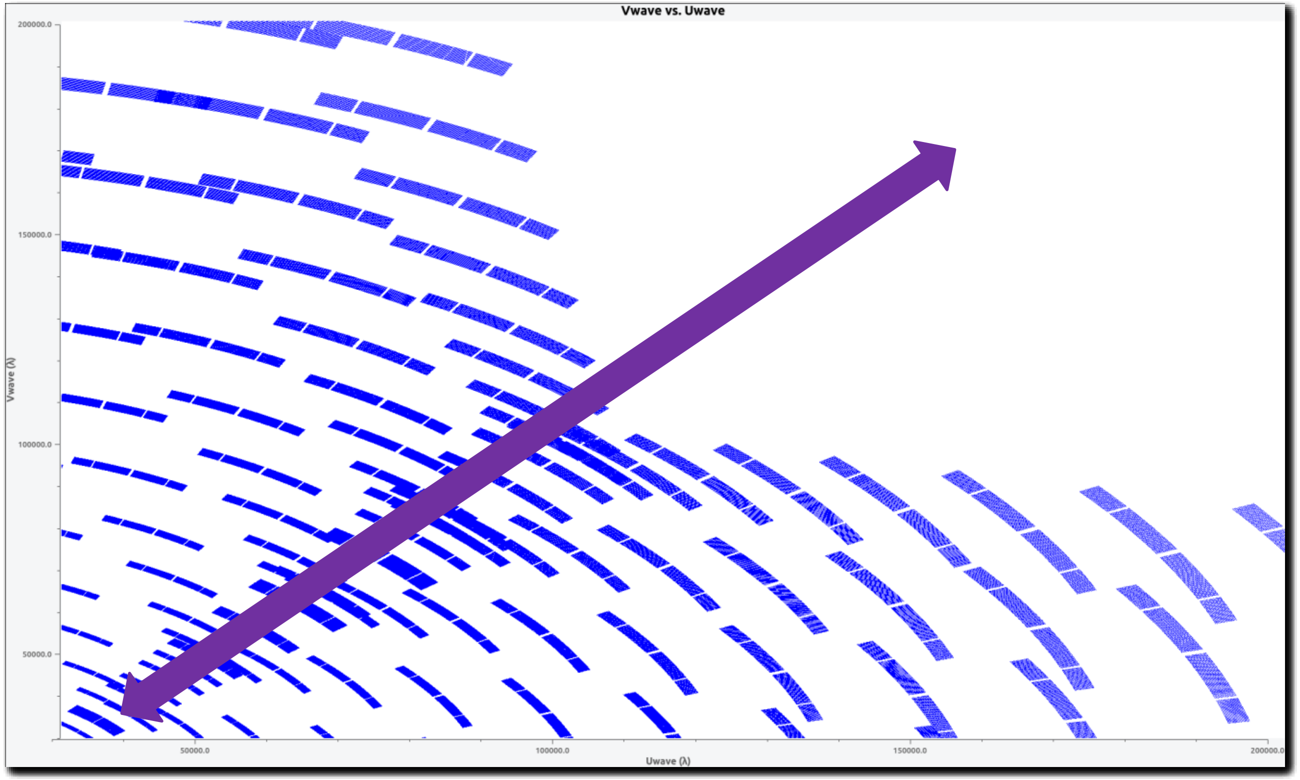
[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

Scaling the computation

- ➢ Need to scale the computation, particularly de/gridding
    - Shown to take up to 94% of the time for a serial implementation[1]
- ➢ Partition visibilities, process separately

Partition & Parallelization strategies:

- ➢ By frequency
- ➢ By time
- ➢ What about by baseline length?

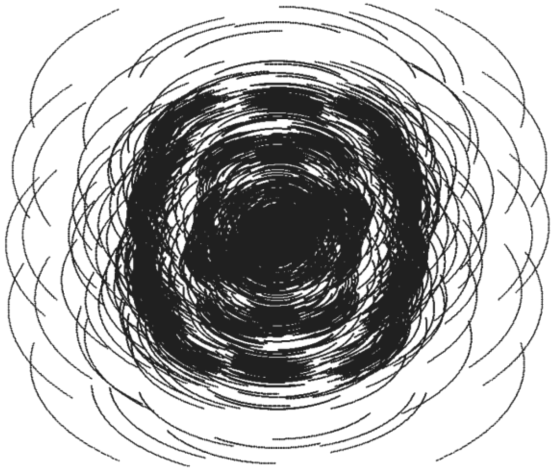


[1]Tasse, C., et al. "Faceting for direction-dependent spectral deconvolution." *Astronomy & Astrophysics* 611 (2018): A87.

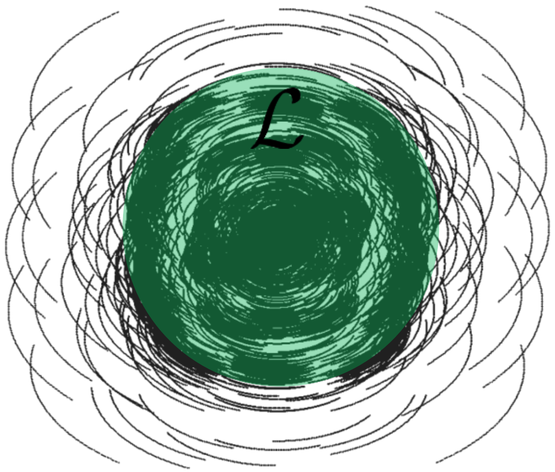
Partitioning image reconstruction by baseline

Multi-step image reconstruction: General Framework

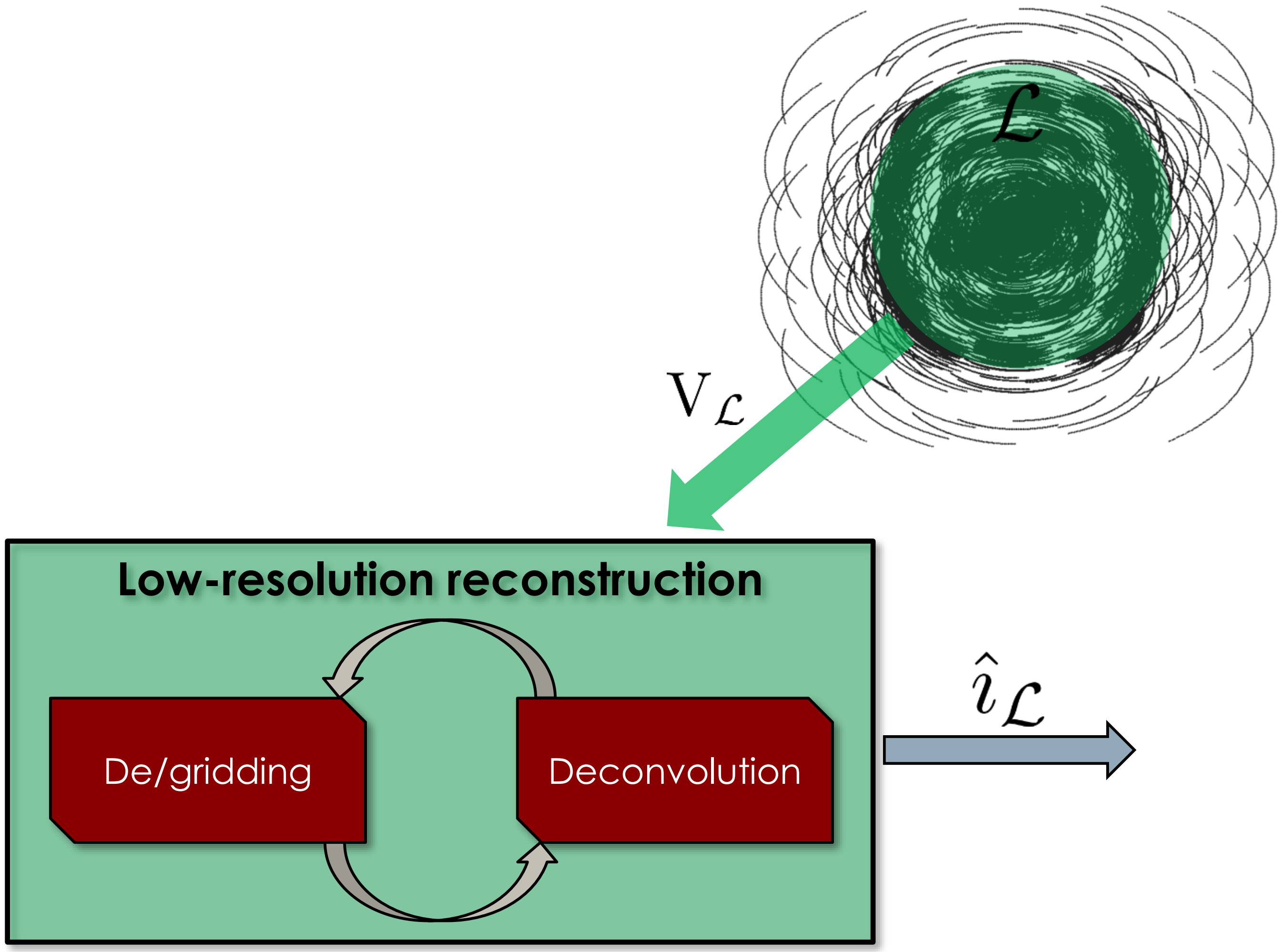
Multi-step image reconstruction: General Framework



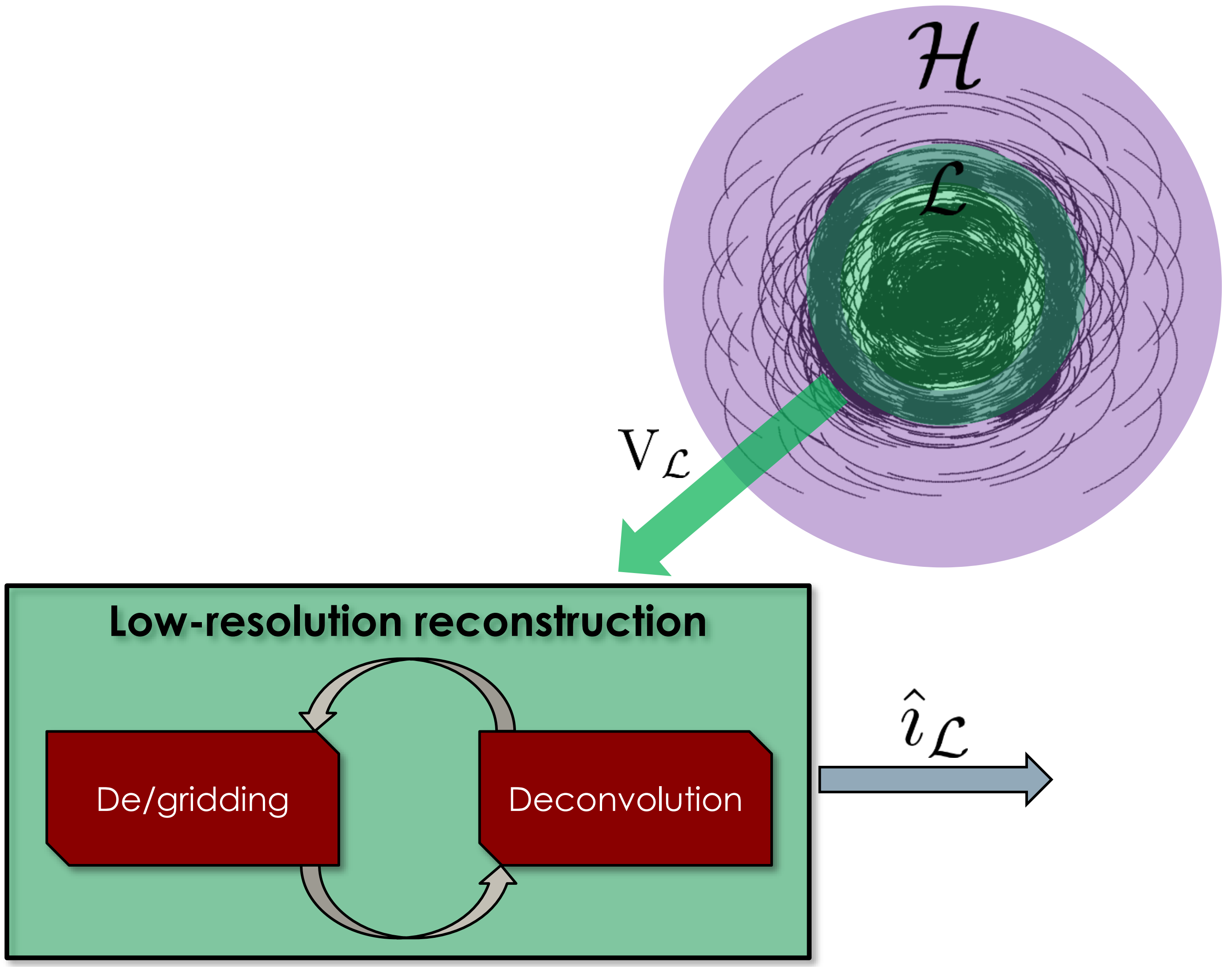
Multi-step image reconstruction: General Framework



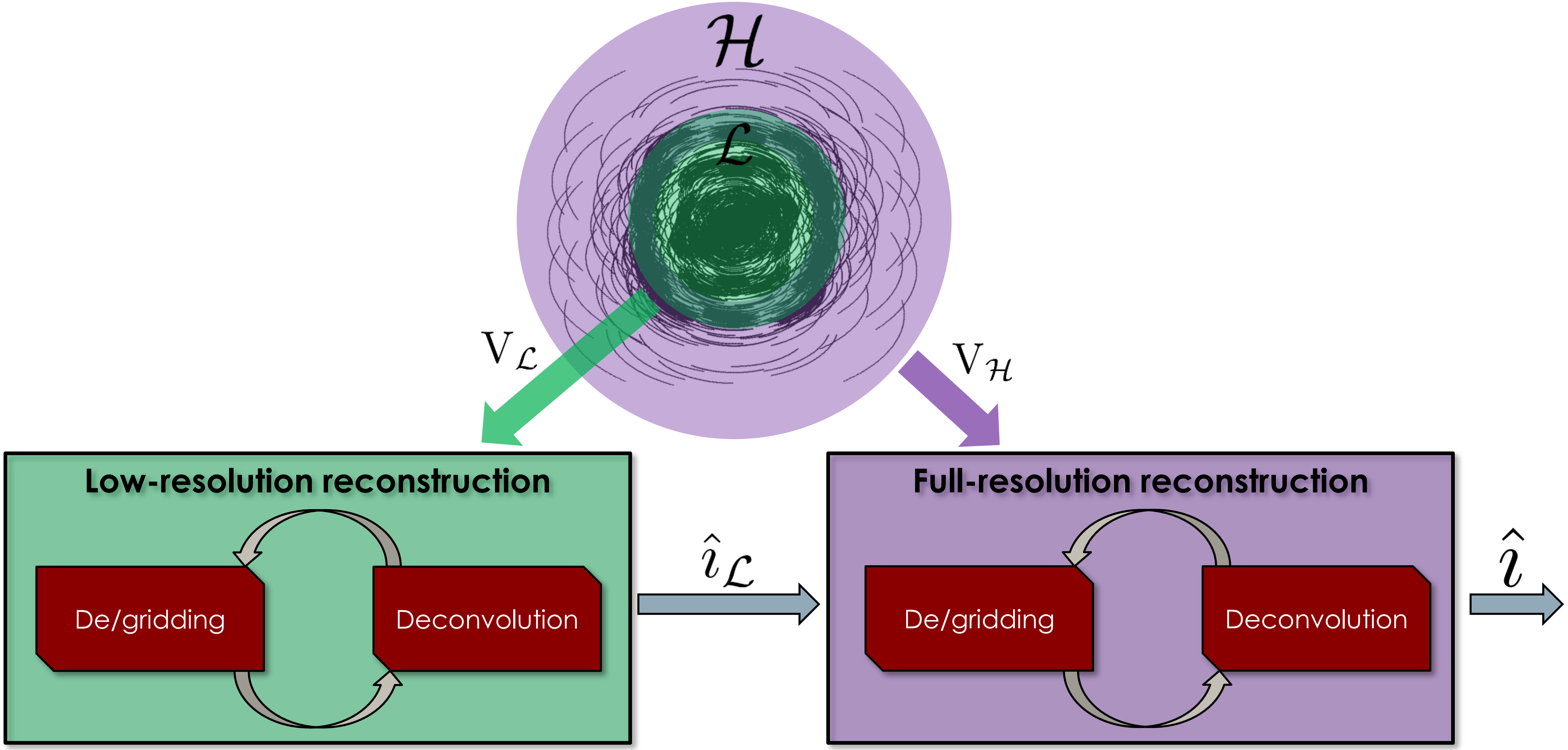
Multi-step image reconstruction: General Framework



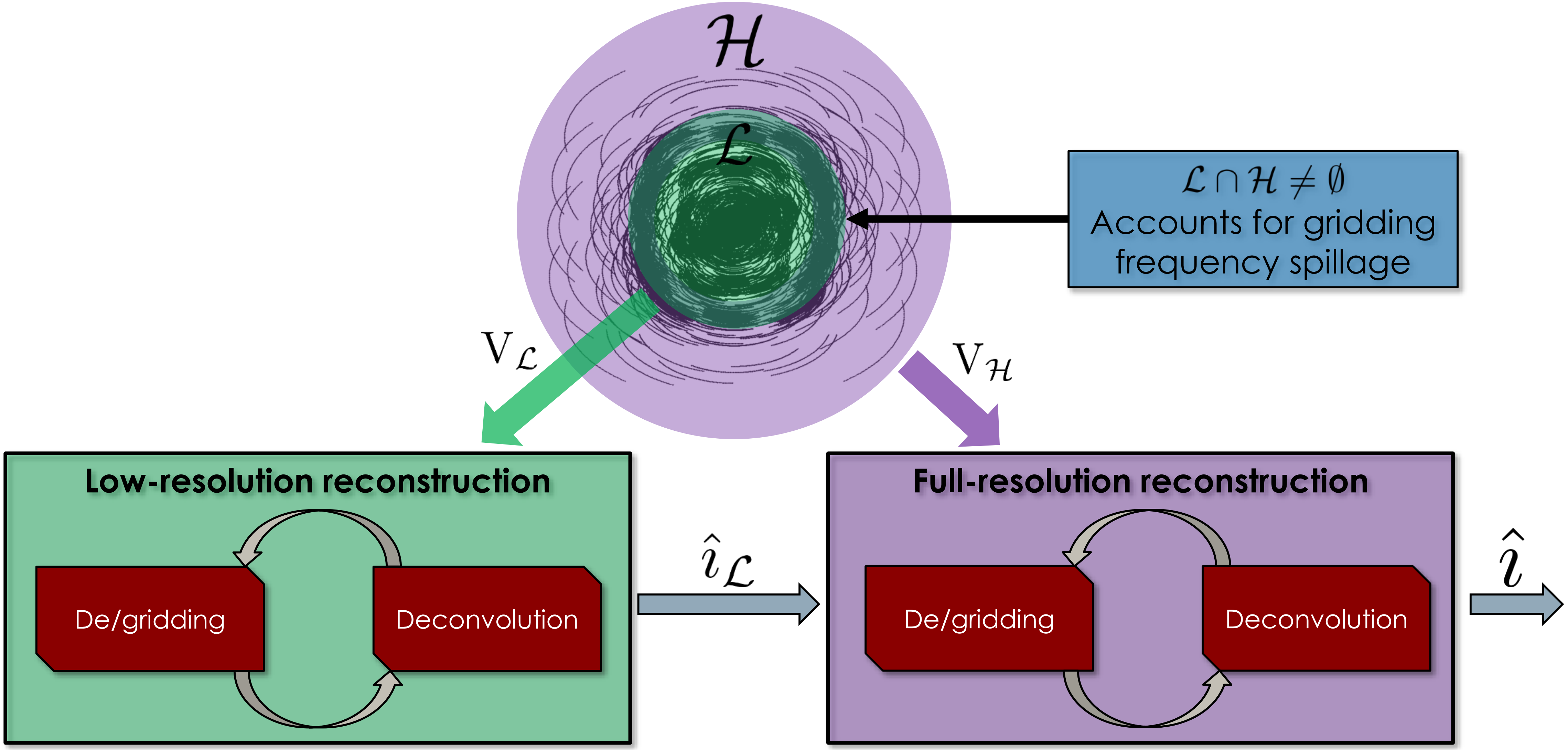
Multi-step image reconstruction: General Framework



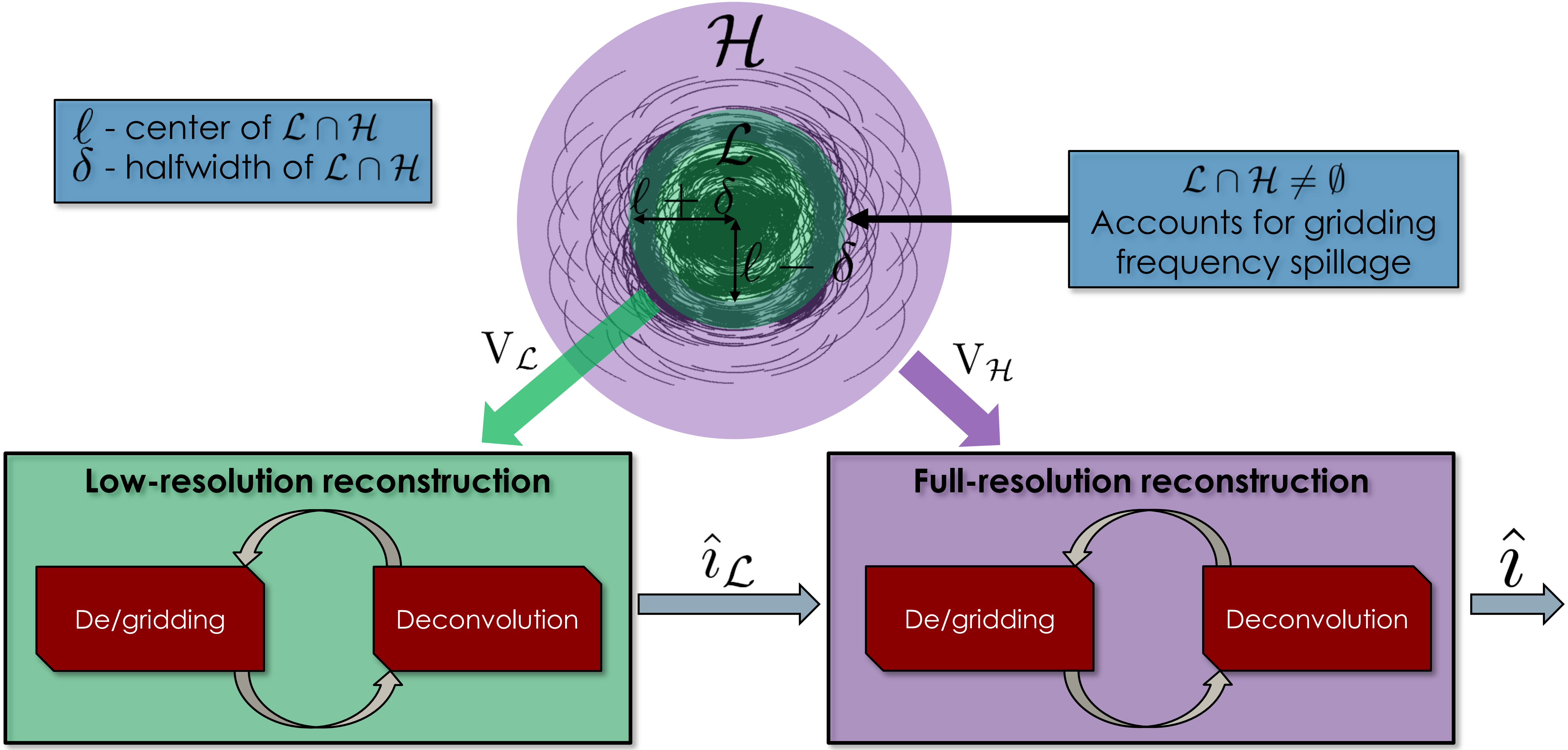
Multi-step image reconstruction: General Framework



Multi-step image reconstruction: General Framework



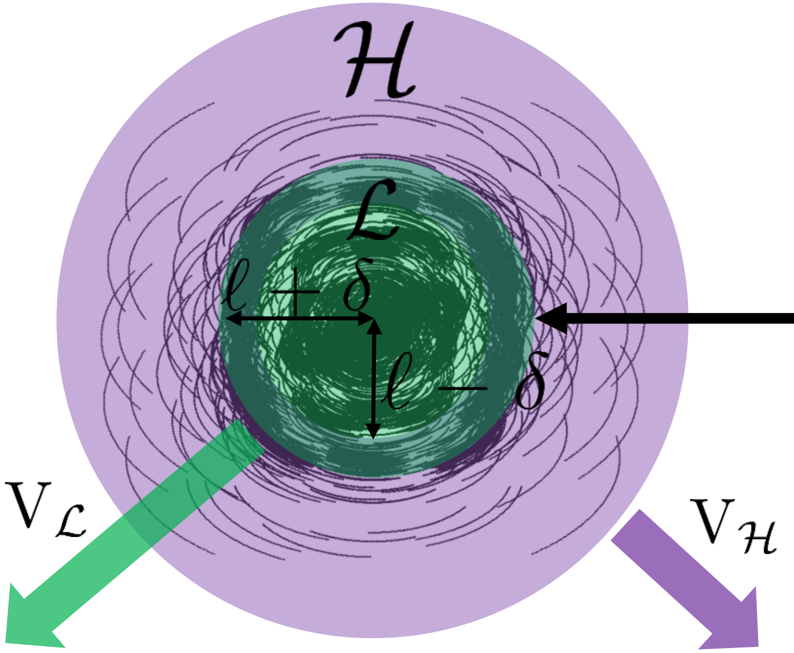
Multi-step image reconstruction: General Framework



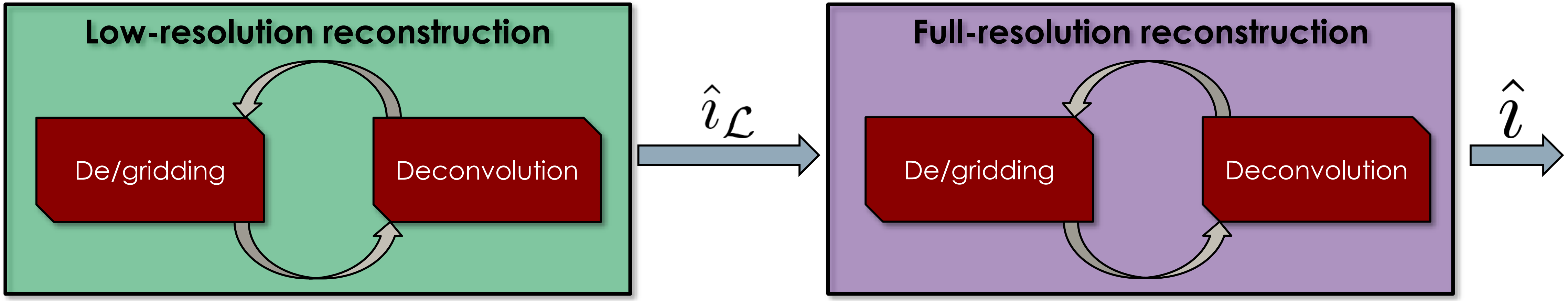
Multi-step image reconstruction: General Framework

ℓ - center of $\mathcal{L} \cap \mathcal{H}$
δ - halfwidth of $\mathcal{L} \cap \mathcal{H}$

Can theoretically work for arbitrary numbers of partitions



$\mathcal{L} \cap \mathcal{H} \neq \emptyset$
Accounts for gridding frequency spillage



Multi-step image reconstruction: Deconvolution framework

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} ||\tilde{\imath}_n - HW\alpha||_2^2 + \lambda_n ||\alpha||_1$$

$$\bar{\imath}_n = W\alpha_n$$

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: Deconvolution framework

n_{th} major cycle residual

$\tilde{i}_n = F^\dagger G(v - G^\dagger F \sum_{n=1}^{N} \bar{i}_{\mathcal{L}_n})$

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} ||\tilde{i}_n - HW\alpha||_2^2 + \lambda_n ||\alpha||_1$$

$$\bar{i}_n = W\alpha_n$$

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: Deconvolution framework

n_{th} major cycle residual

$$\tilde{i}_n = F^\dagger G(v - G^\dagger F \textstyle\sum_{n=1}^{N} \bar{i}_{\mathcal{L}_n})$$

Convolution by PSF operator

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} ||\tilde{i}_n - HW\alpha||_2^2 + \lambda_n ||\alpha||_1$$

$$\bar{i}_n = W\alpha_n$$

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: Deconvolution framework

n_{th} major cycle residual
$\tilde{\imath}_n = F^\dagger G(v - G^\dagger F \sum_{n=1}^{N} \bar{\imath}_{\mathcal{L}_n})$

Convolution by PSF operator

Regularization parameter

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} ||\tilde{\imath}_n - HW\alpha||_2^2 + \lambda_n ||\alpha||_1$$

$$\bar{\imath}_n = W\alpha_n$$

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: Deconvolution framework

n_{th} major cycle residual
$\tilde{i}_n = F^{\dagger}G(v - G^{\dagger}F\sum_{n=1}^{N}\bar{i}_{\mathcal{L}_n})$

Convolution by PSF operator

Regularization parameter

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} ||\tilde{i}_n - HW\alpha||_2^2 + \lambda_n ||\alpha||_1$$

$$\bar{i}_n = W\alpha_n$$

Wavelet transform operator (Daubechies 1-8)

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: Deconvolution framework

n_{th} major cycle residual
$\tilde{\imath}_n = F^{\dagger}G(v - G^{\dagger}F\sum_{n=1}^{N}\bar{\imath}_{\mathcal{L}_n})$

Convolution by PSF operator

Regularization parameter

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} ||\tilde{\imath}_n - HW\alpha||_2^2 + \lambda_n||\alpha||_1$$

$$\bar{\imath}_n = W\alpha_n$$

Wavelet transform operator (Daubechies 1-8)

Deconvolved image

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: Deconvolution framework

n_{th} major cycle residual
$\tilde{\imath}_n = F^\dagger G(v - G^\dagger F \sum_{n=1}^{N} \bar{\imath}_{\mathcal{L}_n})$

Convolution by PSF operator

Regularization parameter

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} \underbrace{||\tilde{\imath}_n - HW\alpha||_2^2}_{} + \lambda_n ||\alpha||_1$$

$$\bar{\imath}_n = W\alpha_n$$

Wavelet transform operator (Daubechies 1-8)

Used over $||v - G^\dagger F W\alpha||_2^2$ for efficiency
Assumes $Hi \approx F^\dagger G G^\dagger F i$
Errors corrected in major cycle

Deconvolved image

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: Deconvolution framework

n_{th} major cycle residual
$\tilde{\imath}_n = F^\dagger G(v - G^\dagger F \sum_{n=1}^{N} \bar{\imath}_{\mathcal{L}_n})$

Convolution by PSF operator

Regularization parameter

Deconvolution framework for every major cycle n, similar to [1, 2]

$$\alpha_n = \arg\min_{\alpha} \|\tilde{\imath}_n - HW\alpha\|_2^2 + \lambda_n \|\alpha\|_1$$

$$\bar{\imath}_n = W\alpha_n$$

Wavelet transform operator (Daubechies 1-8)

Deconvolved image

Used over $\|v - G^\dagger F W\alpha\|_2^2$ for efficiency
Assumes $Hi \approx F^\dagger G G^\dagger F i$
Errors corrected in major cycle

Final reconstructed image: $\hat{\imath} = \tilde{\imath}_{N+1} + \sum_{n=1}^{N} \bar{\imath}_n$

[1] Wiaux, Yves, et al. "Compressed sensing imaging techniques for radio interferometry." Monthly Notices of the Royal Astronomical Society 395.3 (2009): 1733-1742.
[2] Garsden, Hugh, et al. "LOFAR sparse image reconstruction." Astronomy & astrophysics 575 (2015): A90.

Multi-step image reconstruction: The low and full-resolution steps

Low-resolution deconvolution

$$\alpha_{\mathcal{L}_n} = \arg\min_{\alpha} \|\tilde{\imath}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha\|_2^2 + \lambda_{\mathcal{L}_n} \|\alpha\|_1$$

$$\bar{\imath}_{\mathcal{L}_n} = W \alpha_{\mathcal{L}_n}$$

Multi-step image reconstruction: The low and full-resolution steps

Low-resolution deconvolution

$$\alpha_{\mathcal{L}_n} = \arg\min_{\alpha} \|\tilde{\imath}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha\|_2^2 + \lambda_{\mathcal{L}_n} \|\alpha\|_1$$
$$\bar{\imath}_{\mathcal{L}_n} = W \alpha_{\mathcal{L}_n}$$

Full resolution deconvolution

$$\alpha_{\mathcal{H}_n} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{\imath}_{\mathcal{H}_n} - H_{\mathcal{H}} W \alpha)\|_2^2 + \|G_{\mathcal{L}}(l_{\mathcal{L}_n} - W\alpha)\|_2^2 + \lambda_{\mathcal{H}_n} \|\alpha\|_1$$
$$\bar{\imath}_n = W \alpha_{\mathcal{H}_n}$$

Multi-step image reconstruction: The low and full-resolution steps

Low-resolution deconvolution

$$\alpha_{\mathcal{L}_n} = \arg\min_{\alpha} \|\tilde{\imath}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha\|_2^2 + \lambda_{\mathcal{L}_n} \|\alpha\|_1$$

$$\bar{\imath}_{\mathcal{L}_n} = W \alpha_{\mathcal{L}_n}$$

First data fidelity term contains only high-resolution information

Full resolution deconvolution

$$\alpha_{\mathcal{H}_n} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{\imath}_{\mathcal{H}_n} - H_{\mathcal{H}} W \alpha)\|_2^2 + \|G_{\mathcal{L}}(l_{\mathcal{L}_n} - W \alpha)\|_2^2 + \lambda_{\mathcal{H}_n} \|\alpha\|_1$$

$$\bar{\imath}_n = W \alpha_{\mathcal{H}_n}$$

Multi-step image reconstruction: The low and full-resolution steps

Low-resolution deconvolution

$$\alpha_{\mathcal{L}_n} = \arg\min_{\alpha} \|\tilde{\imath}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha\|_2^2 + \lambda_{\mathcal{L}_n} \|\alpha\|_1$$

$$\bar{\imath}_{\mathcal{L}_n} = W \alpha_{\mathcal{L}_n}$$

First data fidelity term contains only high-resolution information

Full resolution deconvolution

$$\alpha_{\mathcal{H}_n} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{\imath}_{\mathcal{H}_n} - H_{\mathcal{H}} W \alpha)\|_2^2 + \|G_{\mathcal{L}}(l_{\mathcal{L}_n} - W \alpha)\|_2^2 + \lambda_{\mathcal{H}_n} \|\alpha\|_1$$

$$\bar{\imath}_n = W \alpha_{\mathcal{H}_n}$$

Second data fidelity term

$l_{\mathcal{L}_n} = \hat{\imath}_{\mathcal{L}} - \sum_{j=1}^{n-1} \bar{\imath}_j$

Multi-step image reconstruction: The low and full-resolution steps

Low-resolution deconvolution

$$\alpha_{\mathcal{L}_n} = \arg\min_{\alpha} \|\tilde{\imath}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha\|_2^2 + \lambda_{\mathcal{L}_n} \|\alpha\|_1$$

$$\bar{\imath}_{\mathcal{L}_n} = W \alpha_{\mathcal{L}_n}$$

First data fidelity term contains only high-resolution information

Full resolution deconvolution

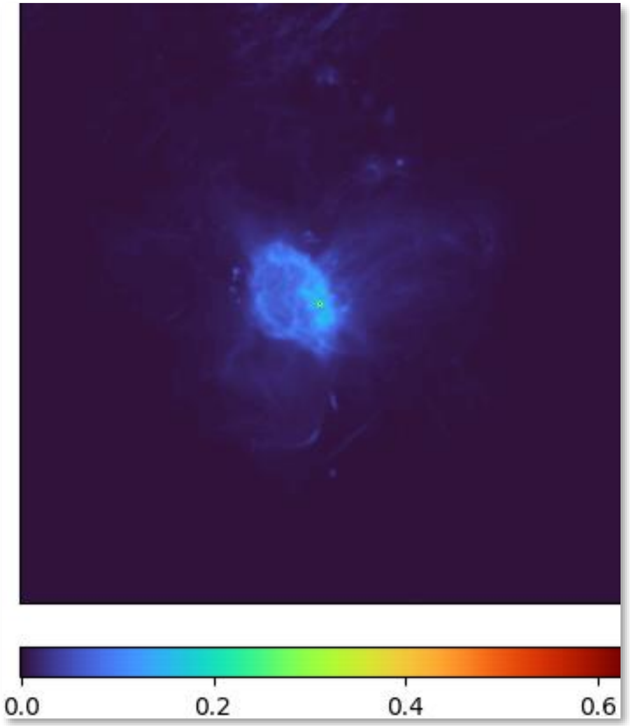
$$\alpha_{\mathcal{H}_n} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{\imath}_{\mathcal{H}_n} - H_{\mathcal{H}} W \alpha)\|_2^2 + \|G_{\mathcal{L}}(l_{\mathcal{L}_n} - W\alpha)\|_2^2 + \lambda_{\mathcal{H}_n} \|\alpha\|_1$$

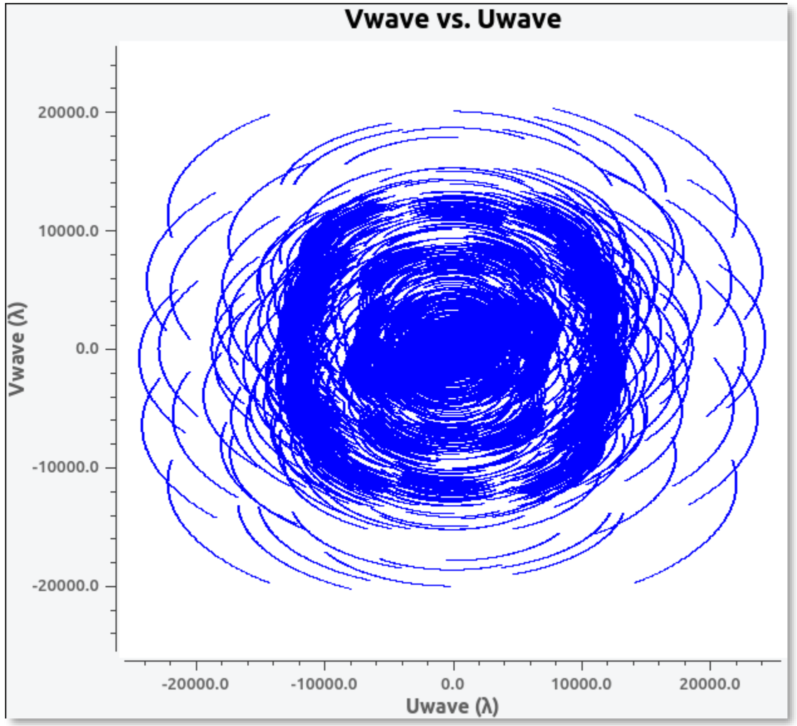
$$\bar{\imath}_n = W \alpha_{\mathcal{H}_n}$$

High and low resolution filters
Normalizes data fidelity terms
Ensures visibilities in $\mathcal{L} \cap \mathcal{H}$ sum to 1

Second data fidelity term
$l_{\mathcal{L}_n} = \hat{\imath}_{\mathcal{L}} - \sum_{j=1}^{n-1} \bar{\imath}_j$

Results – Measurement set



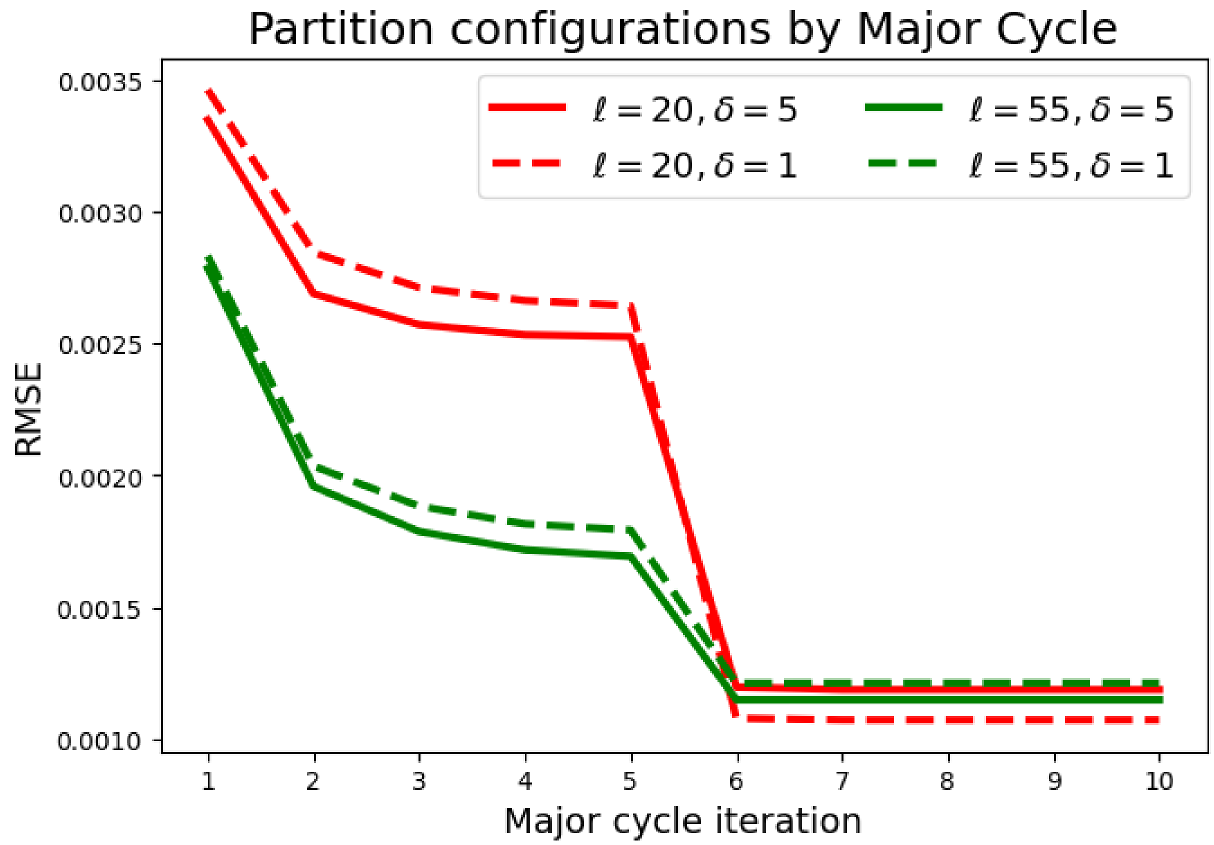


- Initial images tapered and cutout from 1.28GHz mosaic produced in [1]
- Visibilities generated with Meerkat configuration
- Exposure time of 4h, samples every 120s for to generate visibility positions
- Degrid to get visibility values
- Visibility noise artificially added

[1] Heywood, I., et al. "The 1.28 GHz MeerKAT galactic center mosaic." The Astrophysical Journal 925.2 (2022): 165.
[2] Sebokolodi, M. Lerato L., et al. "A Wideband Polarization Study of Cygnus A with the Jansky Very Large Array. I. The Observations and Data." The Astrophysical Journal 903.1 (2020): 36.

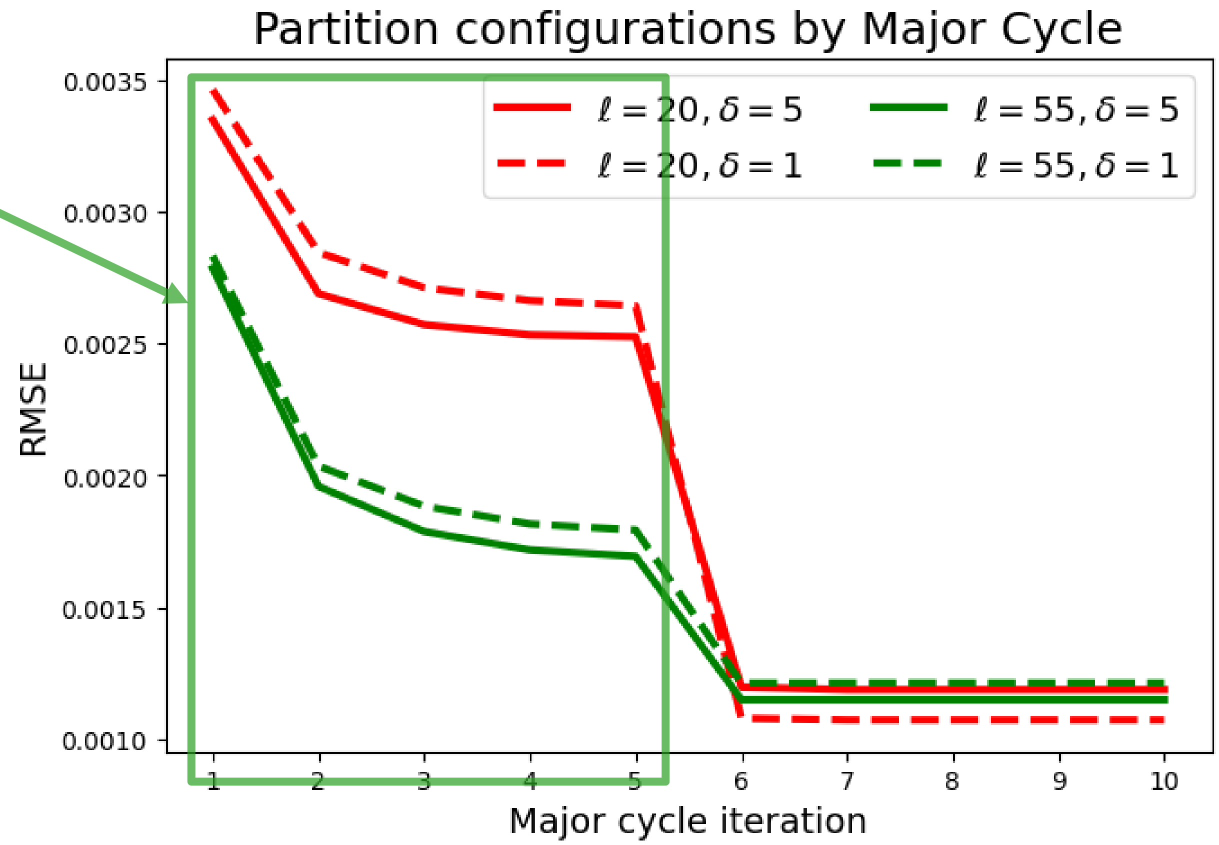
Results: Partition configuration affect on reconstruction accuracy and speed

Results: Partition configuration affect on reconstruction accuracy and speed

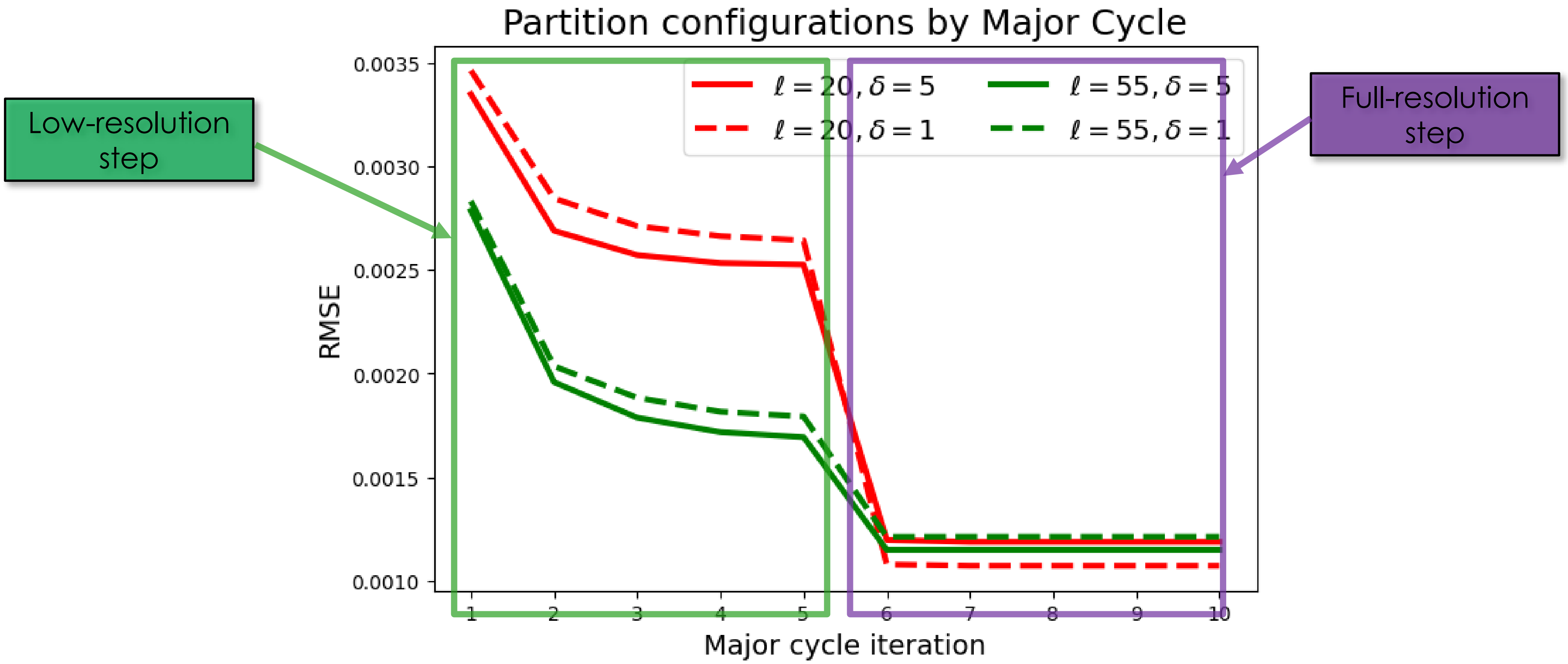


Results: Partition configuration affect on reconstruction accuracy and speed

Low-resolution step

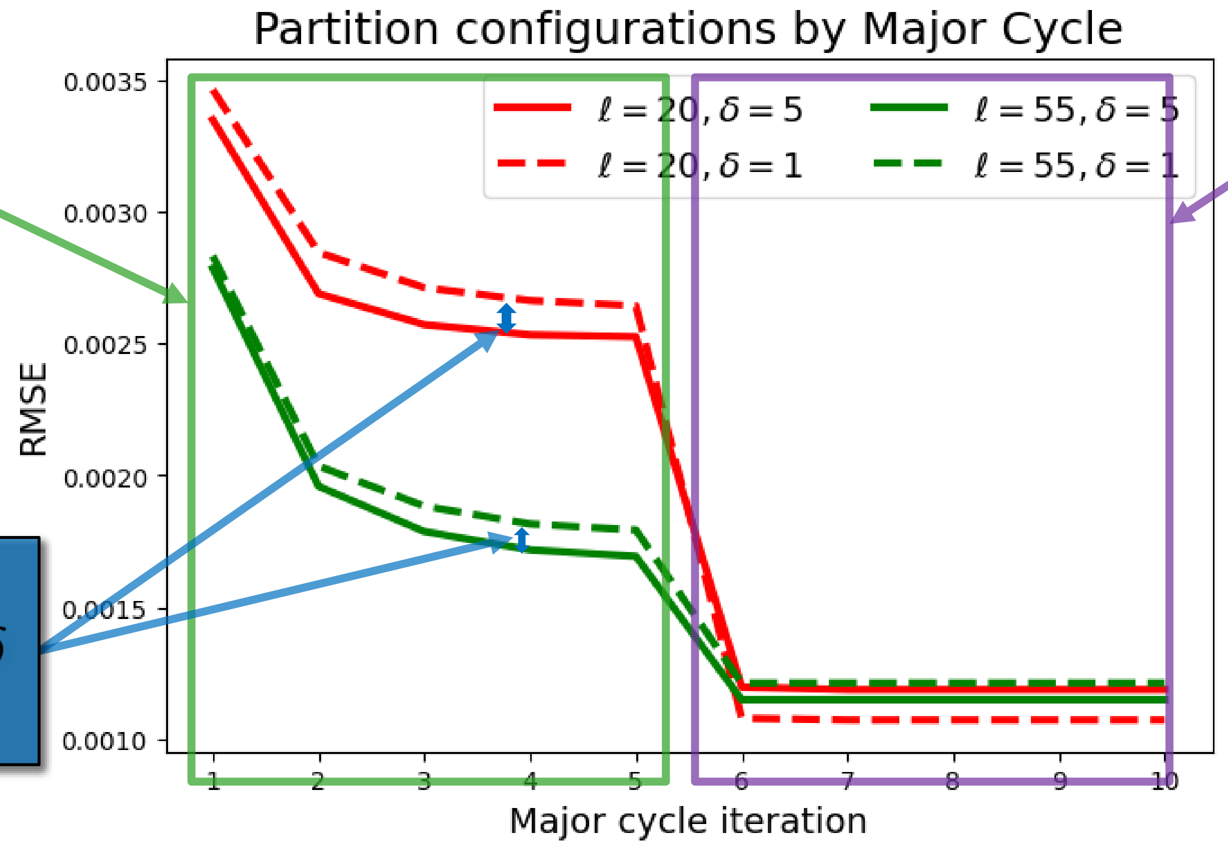


Results: Partition configuration affect on reconstruction accuracy and speed



Results: Partition configuration affect on reconstruction accuracy and speed

Low-resolution step

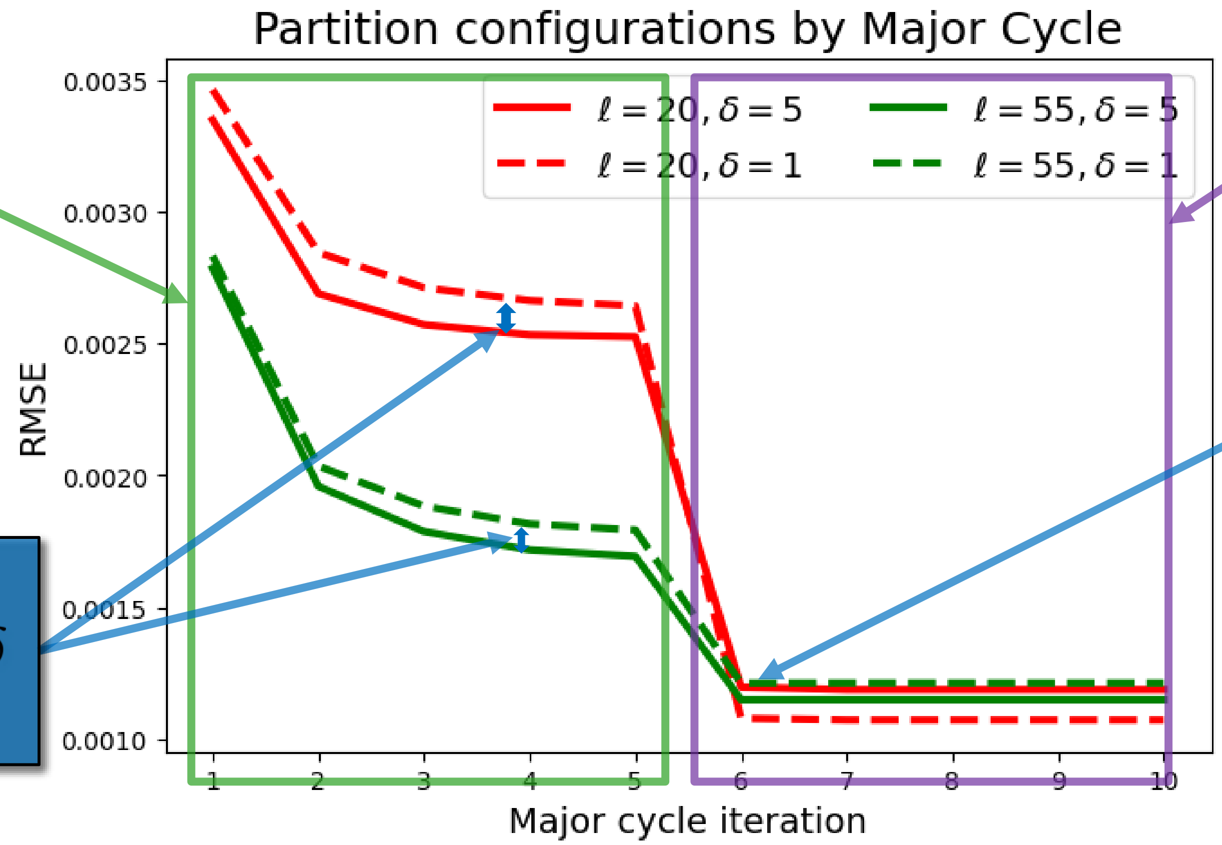


Full-resolution step

Caused by $V_{\mathcal{L}}$ being larger when δ is larger

Results: Partition configuration affect on reconstruction accuracy and speed

Low-resolution step

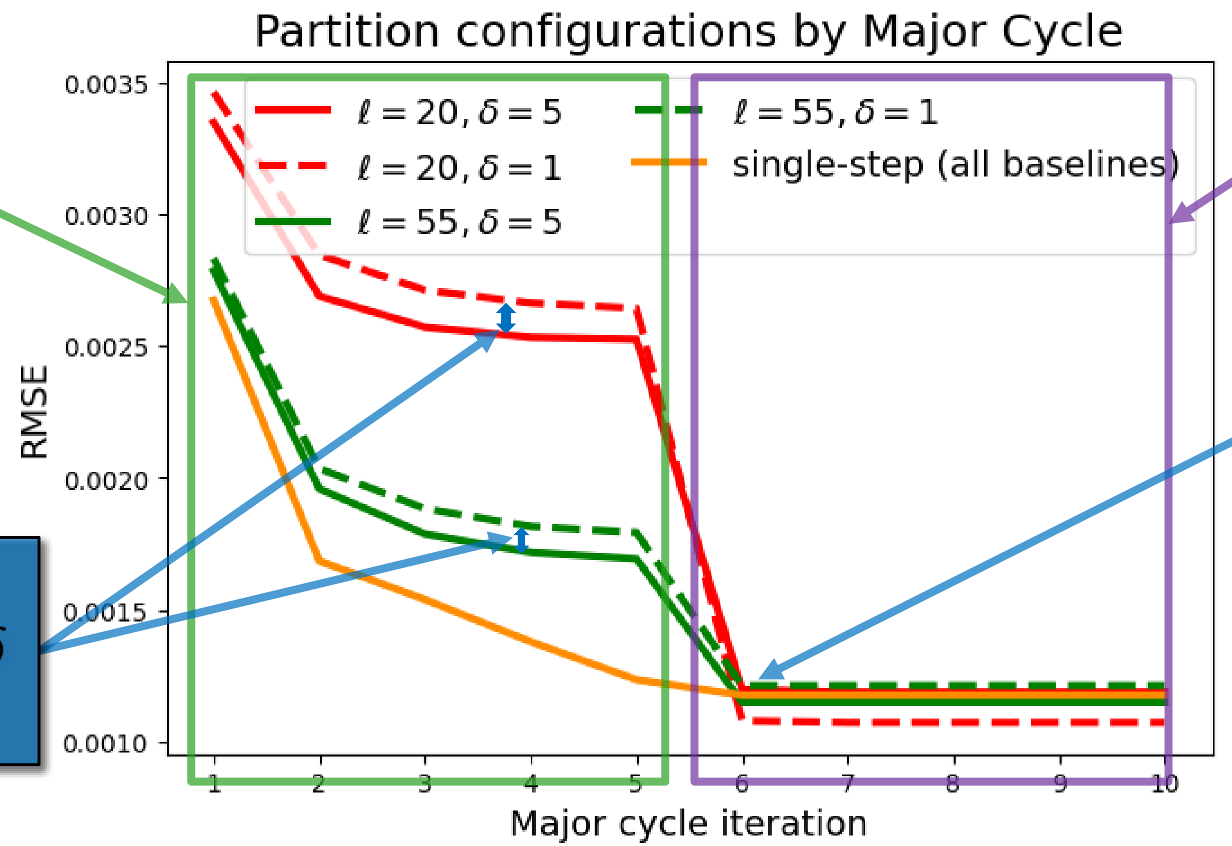


Full-resolution step

- Eventual reconstruction quality similar
- No clear trend as to which configuration is better
- No noticeable issues from smaller delta, at least when using rmse...

Caused by $V_{\mathcal{L}}$ being larger when δ is larger

Results: Partition configuration affect on reconstruction accuracy and speed



Low-resolution step

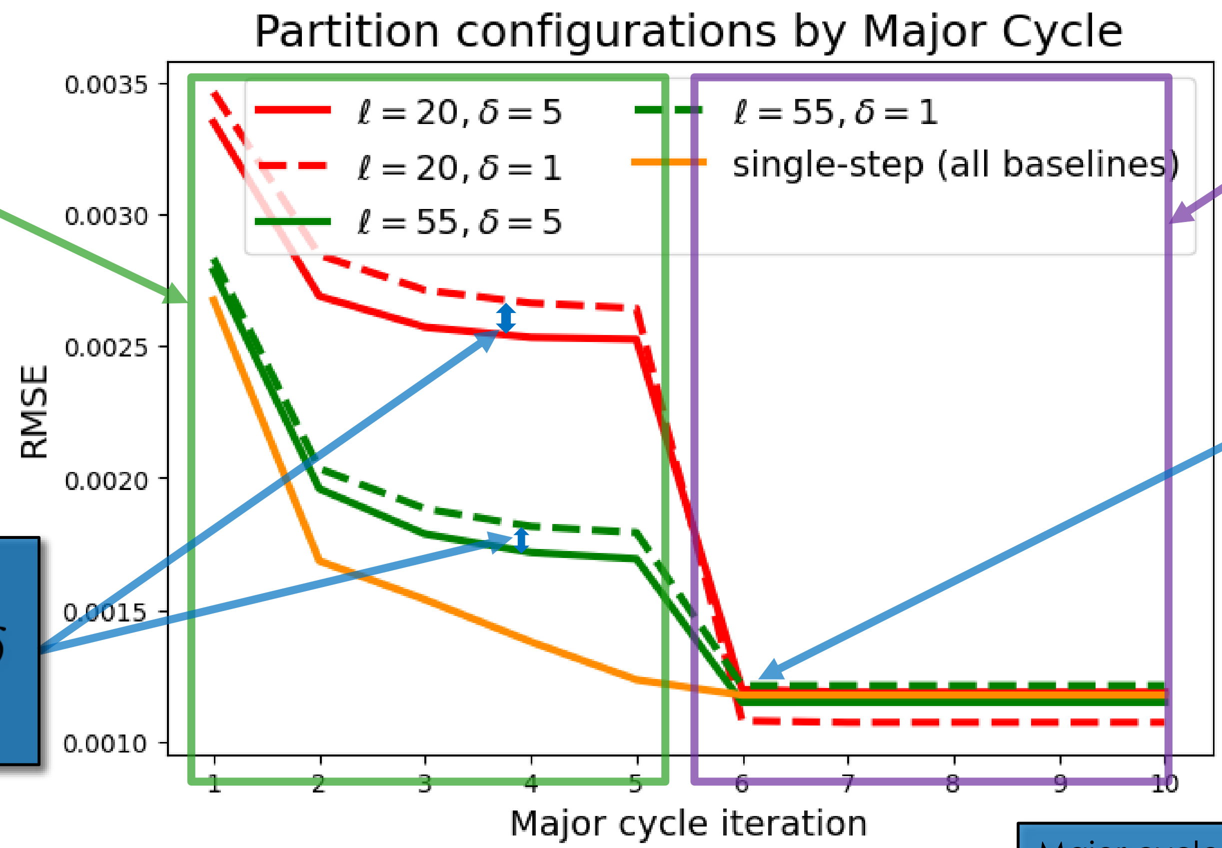
Full-resolution step

Caused by $V_{\mathcal{L}}$ being larger when δ is larger

- Eventual reconstruction quality similar
- No clear trend as to which configuration is better
- No noticeable issues from smaller delta, at least when using rmse…

Results: Partition configuration affect on reconstruction accuracy and speed

Low-resolution step



Full-resolution step

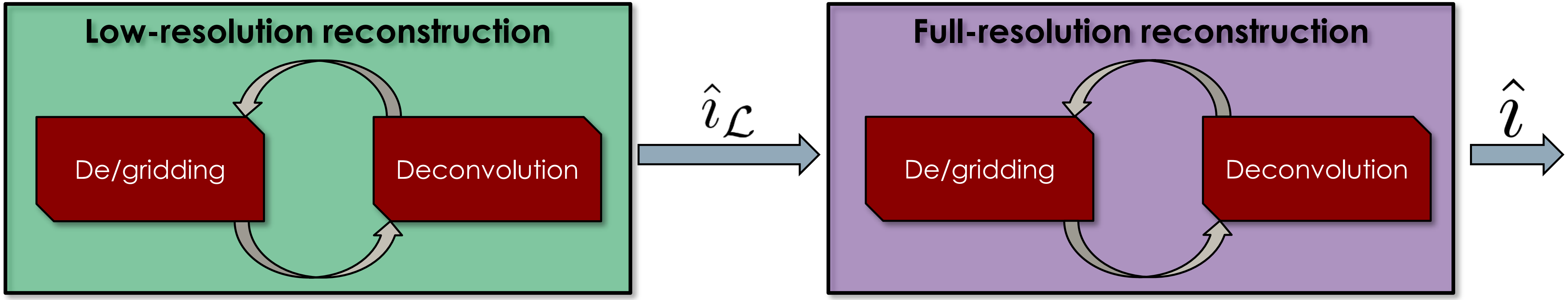
Caused by $V_{\mathcal{L}}$ being larger when δ is larger

- Eventual reconstruction quality similar
- No clear trend as to which configuration is better
- No noticeable issues from smaller delta, at least when using rmse...

Major cycles not the same cost between single and multi-step approaches!

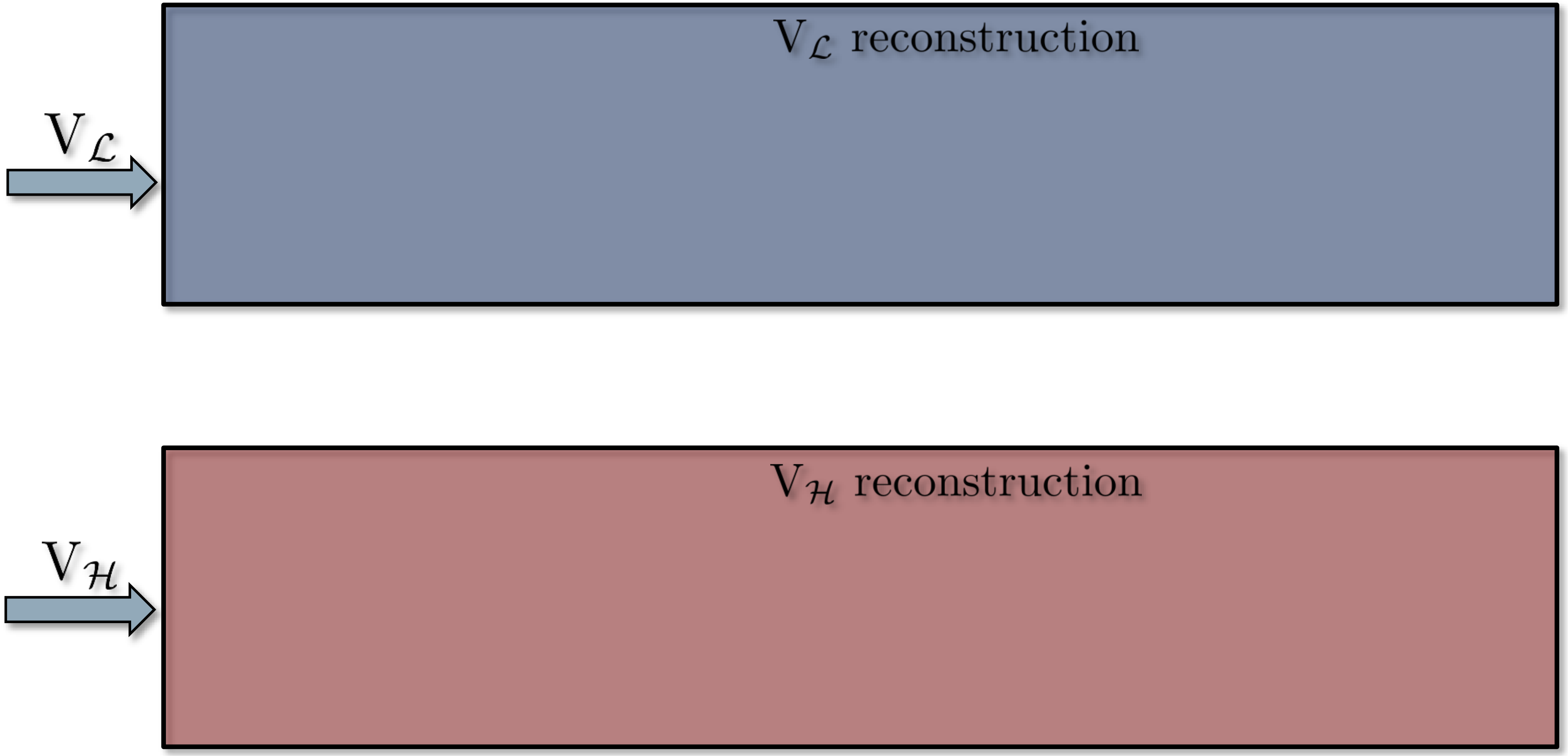
Parallelizing image reconstruction by baseline

Parallelizing the Multi-step Image Reconstruction

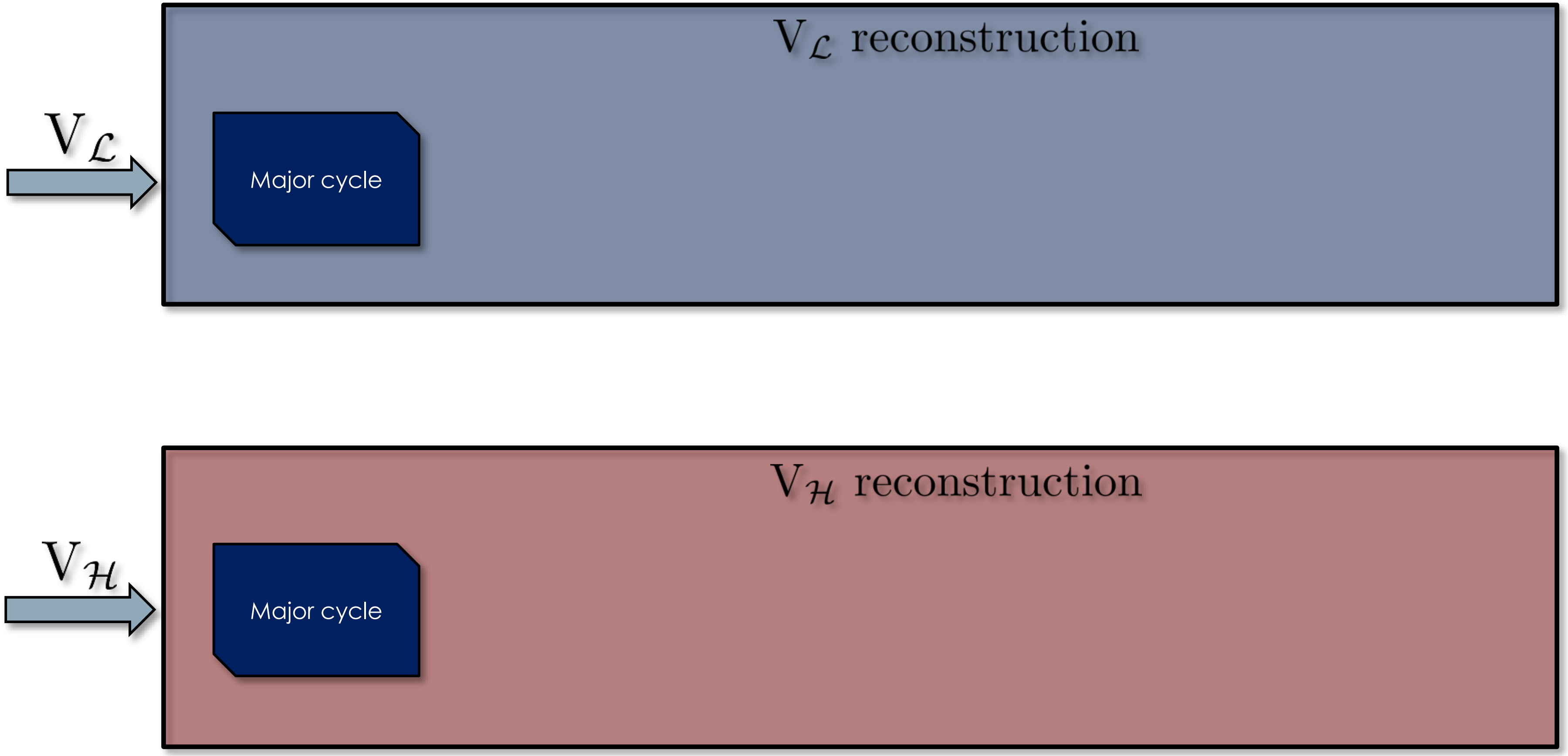


- Described approach allows for better visibility distribution, but is still serial
- Framework can be modified for parallelization

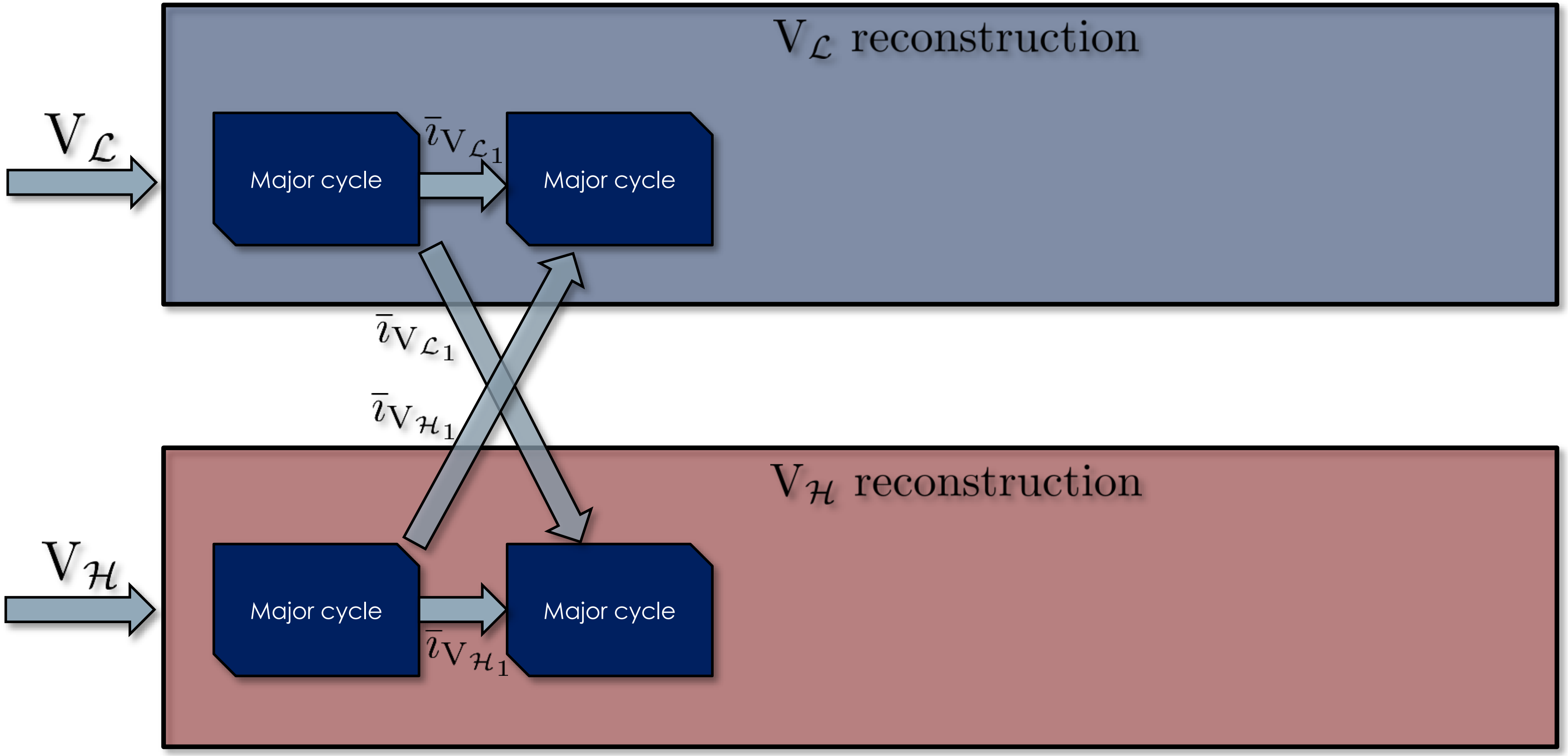
Interleaved parallelization strategy



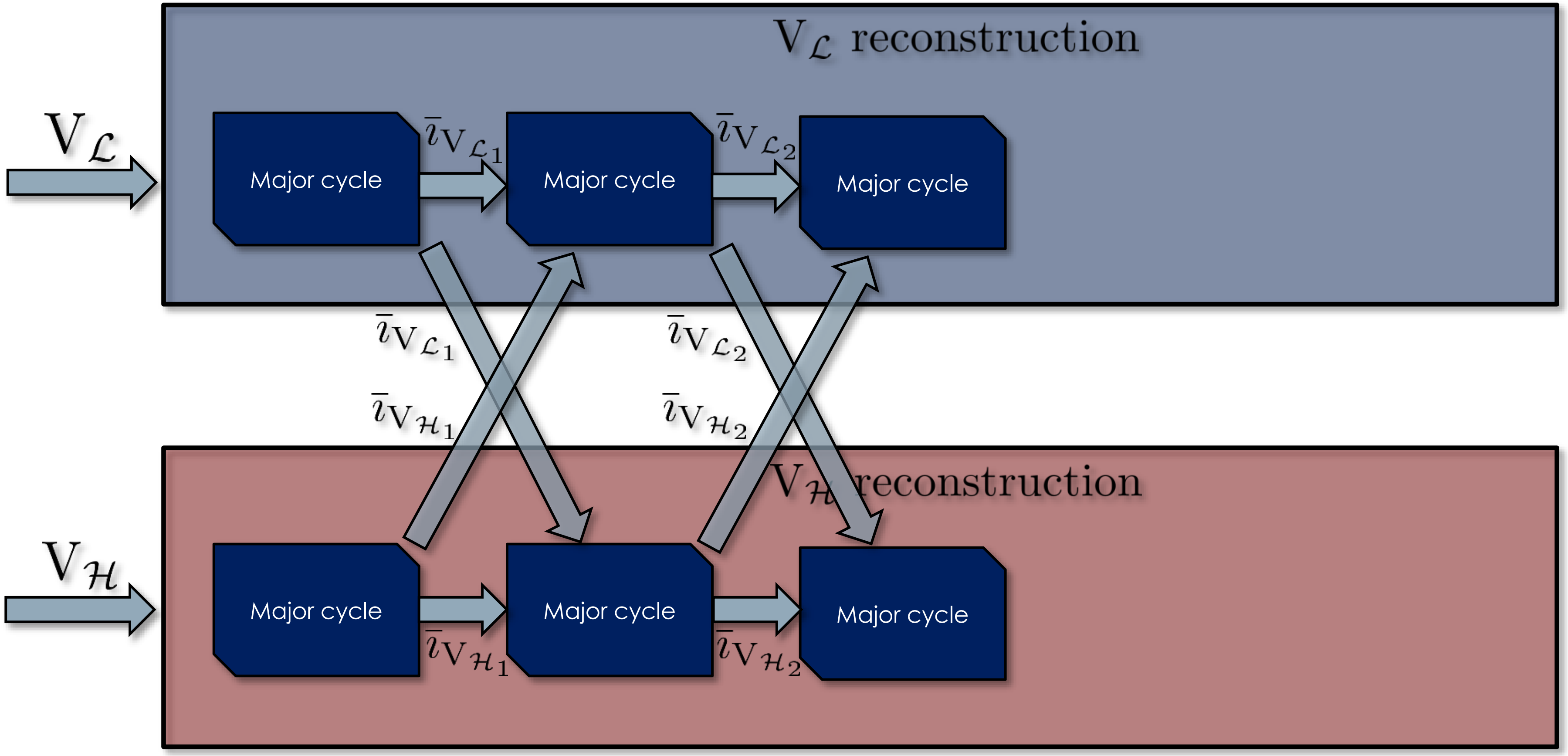
Interleaved parallelization strategy



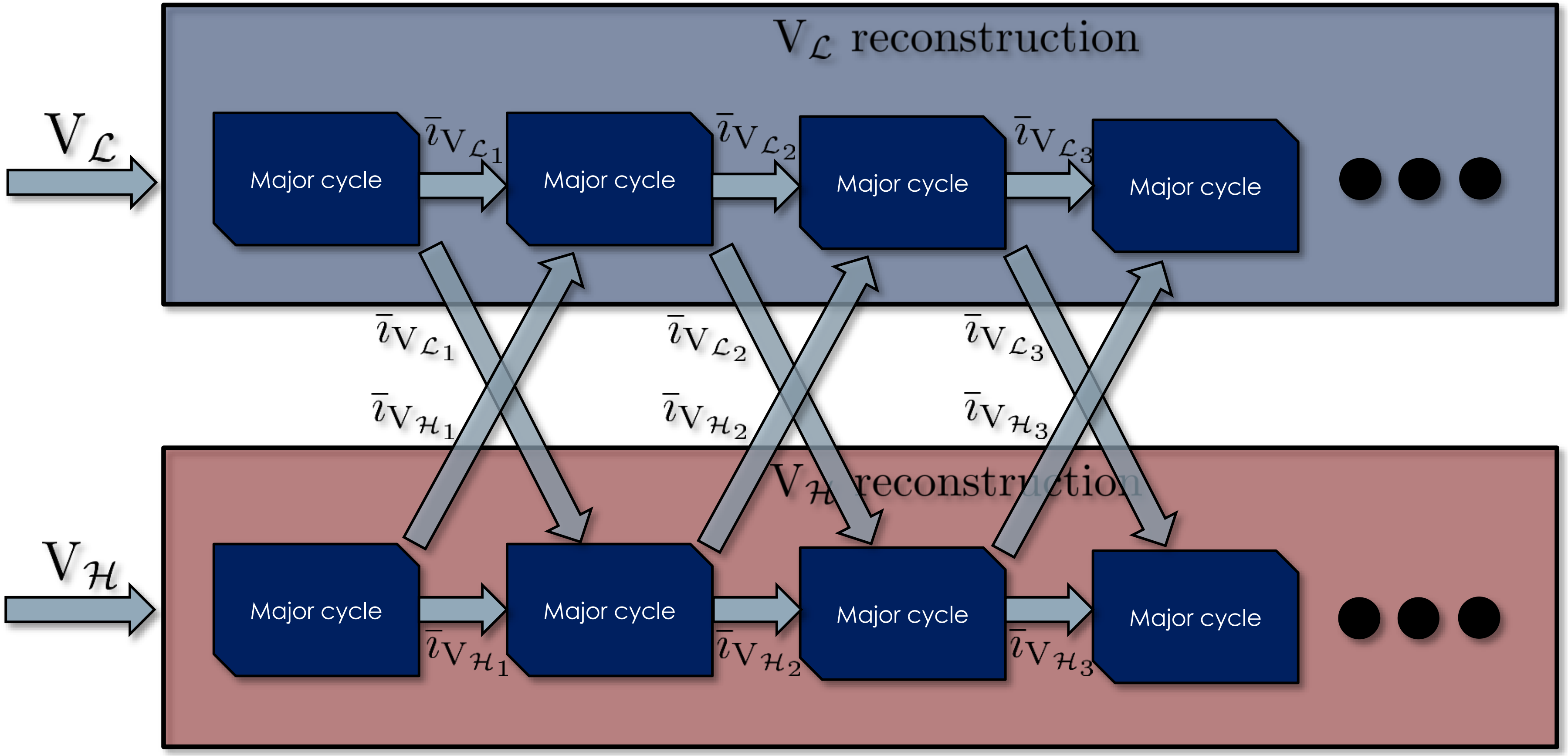
Interleaved parallelization strategy



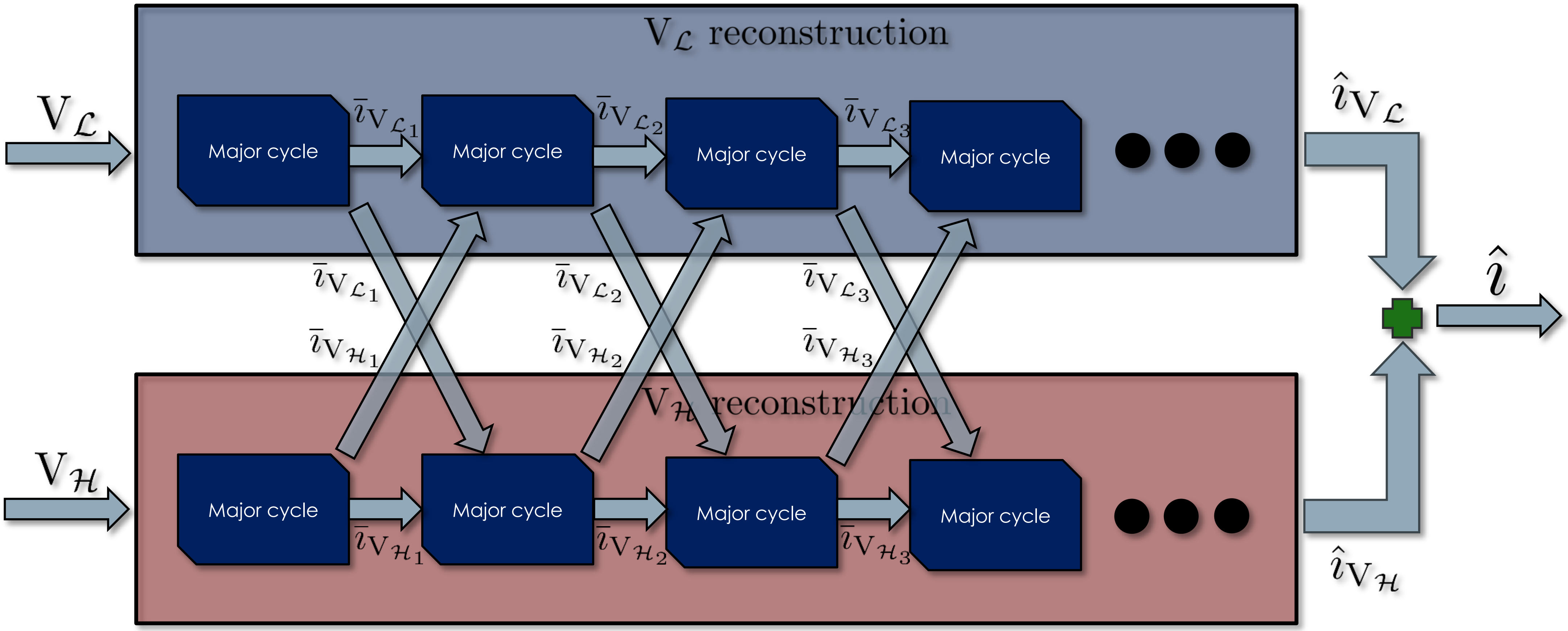
Interleaved parallelization strategy



Interleaved parallelization strategy



Interleaved parallelization strategy



Interleaved parallelization strategy

$\mathrm{V}_{\mathcal{L}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{L}_n}} = \arg\min_{\alpha} \|G_{\mathcal{L}}(\tilde{\imath}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha)\|_2^2 + \|G_{\mathcal{H}}(h_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{L}_n}} \|\alpha\|_1$$

$$\bar{\imath}_{\mathrm{V}_{\mathcal{L}_n}} = W \alpha_{\mathrm{V}_{\mathcal{L}_n}}, h_n = \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{H}_j}} - \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{L}_n}} = \hat{\imath}_{\mathrm{V}_{\mathcal{H}_{n-1}}} - \hat{\imath}_{\mathrm{V}_{\mathcal{L}_{n-1}}}$$

$\mathrm{V}_{\mathcal{H}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{H}_n}} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{\imath}_{\mathcal{H}_n} - H_{\mathcal{H}} W \alpha)\|_2^2 + \|G_{\mathcal{L}}(l_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{H}_n}} \|\alpha\|_1$$

$$\bar{\imath}_{\mathrm{V}_{\mathcal{H}_n}} = W \alpha_{\mathrm{V}_{\mathcal{H}_n}}, l_n = \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{L}_j}} - \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{H}_j}} = \hat{\imath}_{\mathrm{V}_{\mathcal{L}_{n-1}}} - \hat{\imath}_{\mathrm{V}_{\mathcal{H}_{n-1}}}$$

Interleaved parallelization strategy

$\mathrm{V}_{\mathcal{L}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{L}_n}} = \arg\min_{\alpha} \|G_{\mathcal{L}}(\tilde{i}_{\mathcal{L}_n} - H_{\mathcal{L}} W\alpha)\|_2^2 + \|G_{\mathcal{H}}(h_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{L}_n}}\|\alpha\|_1$$

$$\bar{i}_{\mathrm{V}_{\mathcal{L}_n}} = W\alpha_{\mathrm{V}_{\mathcal{L}_n}}, h_n = \sum_{j=1}^{n-1}\bar{i}_{\mathrm{V}_{\mathcal{H}_j}} - \sum_{j=1}^{n-1}\bar{i}_{\mathrm{V}_{\mathcal{L}_n}} = \hat{i}_{\mathrm{V}_{\mathcal{H}_{n-1}}} - \hat{i}_{\mathrm{V}_{\mathcal{L}_{n-1}}}$$

Visibility data-fidelity term

$\mathrm{V}_{\mathcal{H}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{H}_n}} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{i}_{\mathcal{H}_n} - H_{\mathcal{H}} W\alpha)\|_2^2 + \|G_{\mathcal{L}}(l_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{H}_n}}\|\alpha\|_1$$

$$\bar{i}_{\mathrm{V}_{\mathcal{H}_n}} = W\alpha_{\mathrm{V}_{\mathcal{H}_n}}, l_n = \sum_{j=1}^{n-1}\bar{i}_{\mathrm{V}_{\mathcal{L}_j}} - \sum_{j=1}^{n-1}\bar{i}_{\mathrm{V}_{\mathcal{H}_j}} = \hat{i}_{\mathrm{V}_{\mathcal{L}_{n-1}}} - \hat{i}_{\mathrm{V}_{\mathcal{H}_{n-1}}}$$

Interleaved parallelization strategy

$\mathrm{V}_{\mathcal{L}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{L}_n}} = \arg\min_{\alpha} \|G_{\mathcal{L}}(\tilde{\imath}_{\mathcal{L}_n} - H_{\mathcal{L}} W \alpha)\|_2^2 + \|G_{\mathcal{H}}(h_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{L}_n}} \|\alpha\|_1$$

$$\bar{\imath}_{\mathrm{V}_{\mathcal{L}_n}} = W \alpha_{\mathrm{V}_{\mathcal{L}_n}}, h_n = \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{H}_j}} - \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{L}_n}} = \hat{\imath}_{\mathrm{V}_{\mathcal{H}_{n-1}}} - \hat{\imath}_{\mathrm{V}_{\mathcal{L}_{n-1}}}$$

Visibility data-fidelity term

Additional data-fidelity term, only from 2nd major cycle onwards

$\mathrm{V}_{\mathcal{H}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{H}_n}} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{\imath}_{\mathcal{H}_n} - H_{\mathcal{H}} W \alpha)\|_2^2 + \|G_{\mathcal{L}}(l_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{H}_n}} \|\alpha\|_1$$

$$\bar{\imath}_{\mathrm{V}_{\mathcal{H}_n}} = W \alpha_{\mathrm{V}_{\mathcal{H}_n}}, l_n = \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{L}_j}} - \sum_{j=1}^{n-1} \bar{\imath}_{\mathrm{V}_{\mathcal{H}_j}} = \hat{\imath}_{\mathrm{V}_{\mathcal{L}_{n-1}}} - \hat{\imath}_{\mathrm{V}_{\mathcal{H}_{n-1}}}$$

Interleaved parallelization strategy

$\mathrm{V}_{\mathcal{L}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{L}_n}} = \arg\min_{\alpha} \|G_{\mathcal{L}}(\tilde{i}_{\mathcal{L}_n} - H_{\mathcal{L}} W\alpha)\|_2^2 + \|G_{\mathcal{H}}(h_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{L}_n}}\|\alpha\|_1$$

$$\bar{i}_{\mathrm{V}_{\mathcal{L}_n}} = W\alpha_{\mathrm{V}_{\mathcal{L}_n}}, h_n = \sum_{j=1}^{n-1} \bar{i}_{\mathrm{V}_{\mathcal{H}_j}} - \sum_{j=1}^{n-1} \bar{i}_{\mathrm{V}_{\mathcal{L}_n}} = \hat{i}_{\mathrm{V}_{\mathcal{H}_{n-1}}} - \hat{i}_{\mathrm{V}_{\mathcal{L}_{n-1}}}$$

Visibility data-fidelity term

Additional data-fidelity term, only from 2nd major cycle onwards

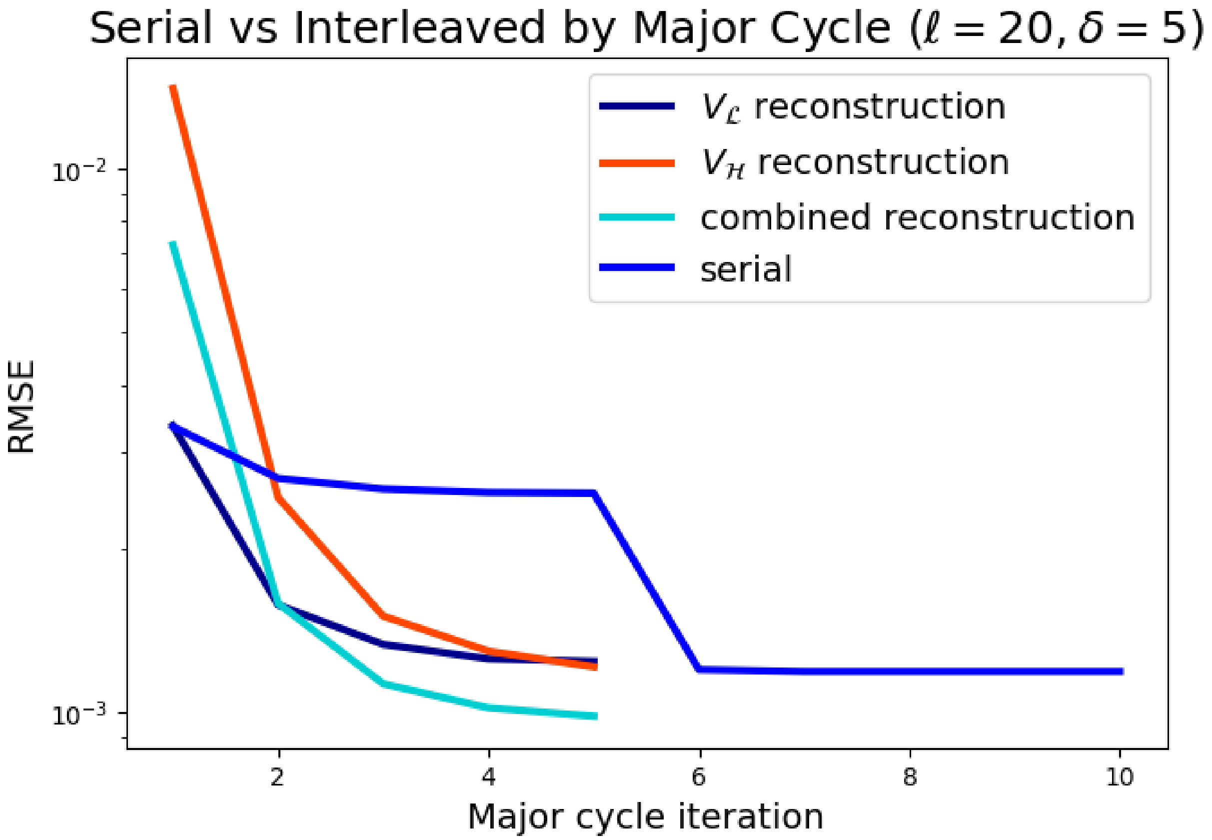
$\mathrm{V}_{\mathcal{H}}$ deconvolution

$$\alpha_{\mathrm{V}_{\mathcal{H}_n}} = \arg\min_{\alpha} \|G_{\mathcal{H}}(\tilde{i}_{\mathcal{H}_n} - H_{\mathcal{H}} W\alpha)\|_2^2 + \|G_{\mathcal{L}}(l_n - W\alpha)\|_2^2 + \lambda_{\mathrm{V}_{\mathcal{H}_n}}\|\alpha\|_1$$

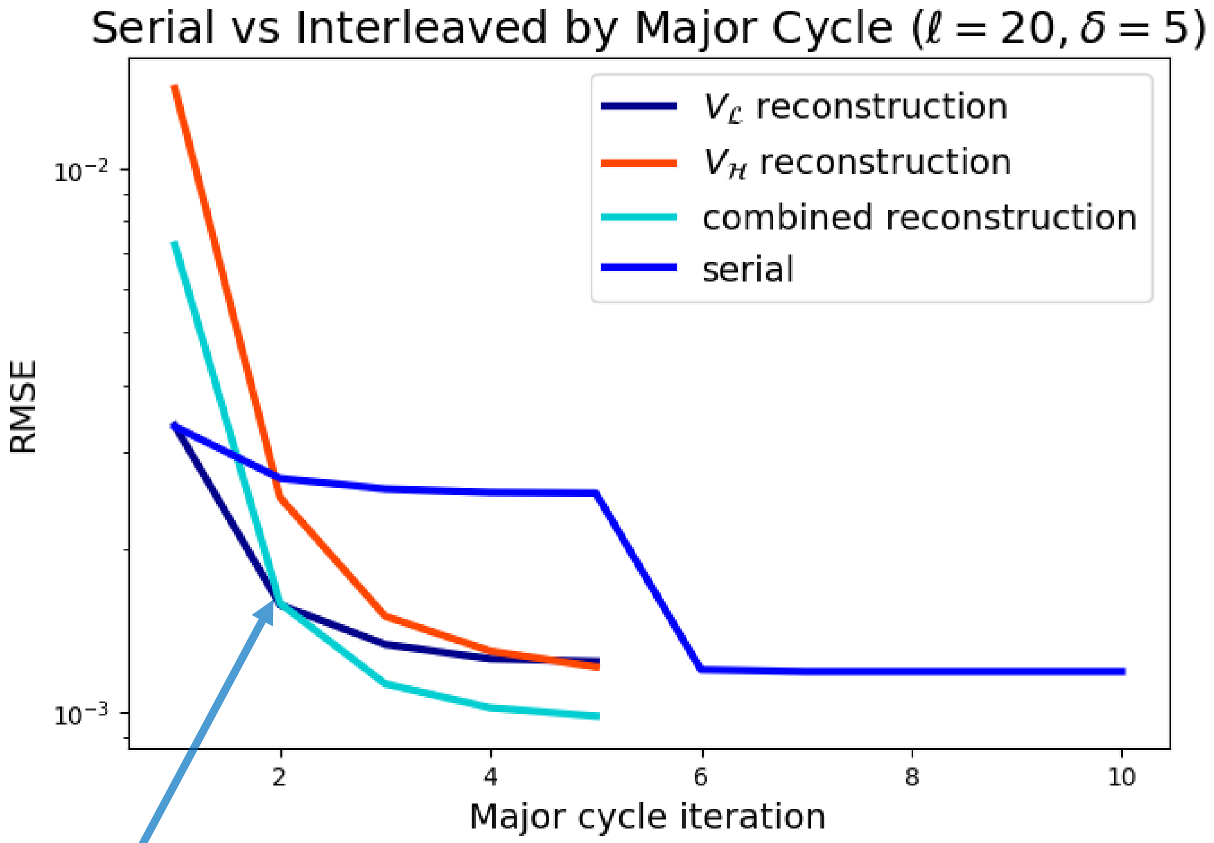
$$\bar{i}_{\mathrm{V}_{\mathcal{H}_n}} = W\alpha_{\mathrm{V}_{\mathcal{H}_n}}, l_n = \sum_{j=1}^{n-1} \bar{i}_{\mathrm{V}_{\mathcal{L}_j}} - \sum_{j=1}^{n-1} \bar{i}_{\mathrm{V}_{\mathcal{H}_j}} = \hat{i}_{\mathrm{V}_{\mathcal{L}_{n-1}}} - \hat{i}_{\mathrm{V}_{\mathcal{H}_{n-1}}}$$

- Two full-resolution images instead of one low and one full
- Roughly symmetric reconstruction costs
- Combined image should be of higher quality
- Quadratic communication with number of partitions
- At least 4 major-cycles needed for an image

Results: Reconstruction quality of interleaved method

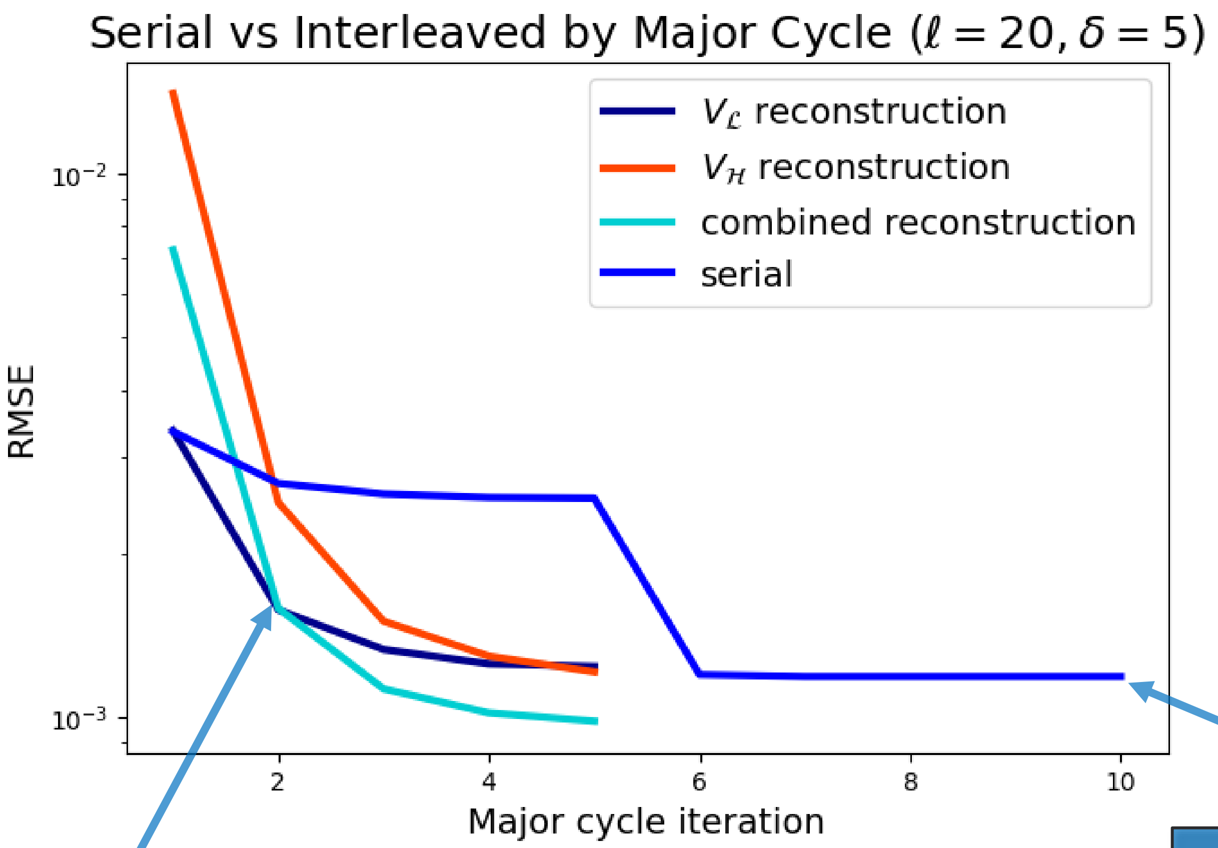


Results: Reconstruction quality of interleaved method



Combined reconstruction equal to or outperforms others after 2nd major cycle

Results: Reconstruction quality of interleaved method



Combined reconstruction equal to or outperforms others after 2nd major cycle

Serial reconstruction worse than parallel after roughly same amount of computation

Conclusions and Future Work

➢ To conclude:

- ➢ Method to partition visibilities by baseline length when reconstructing images
  - ❖ Shown to have similar cost and quality to processing all baselines together
- ➢ Parallelization strategy for given method
  - ❖ Better reconstruction quality than serial for same cost

Future work:

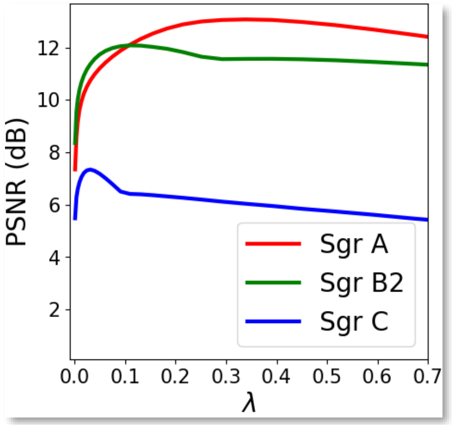
- ➢ More partitions
- ➢ Implementation and benchmarking on large cluster
- ➢ Different metrics for reconstruction quality
- ➢ Investigate the framework with other deconvolution frameworks

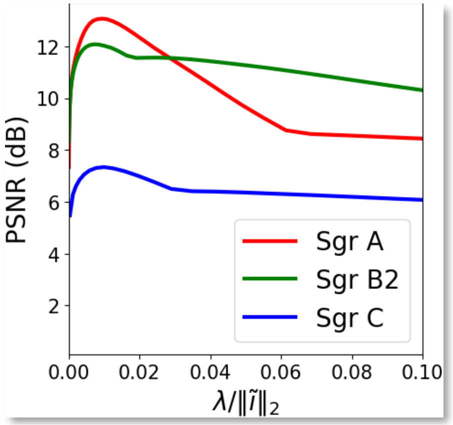
Thank you!
Questions?

Appendices

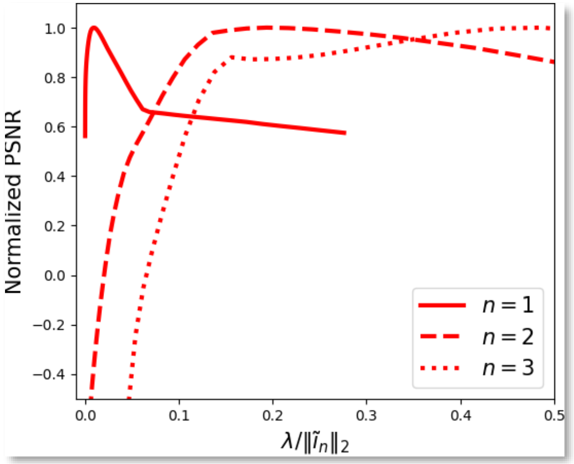
Selection of λ

Across different datasets for first major-cycle:





Across first three major-cycles for Sgr A dataset:



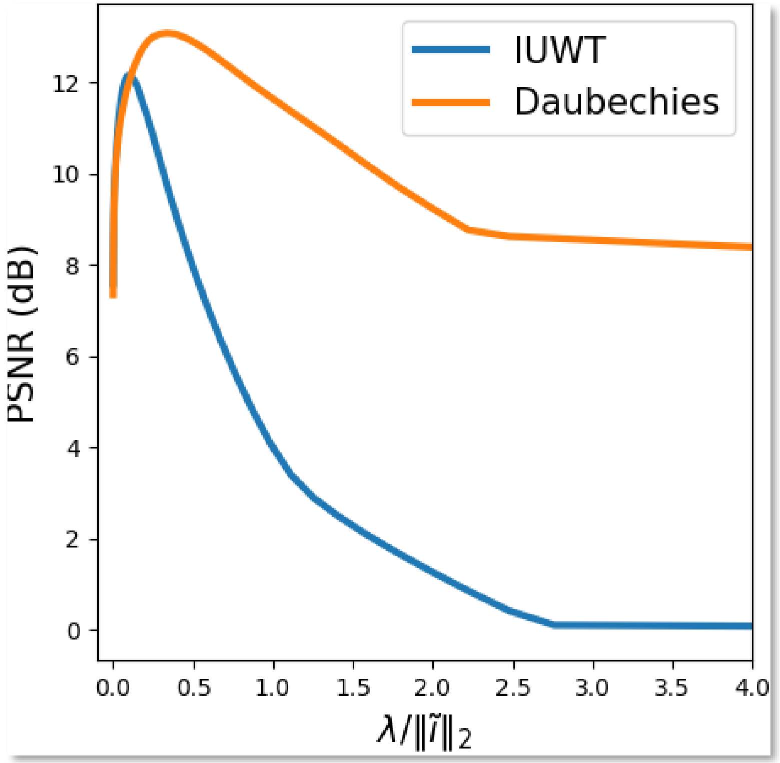
Results suggest that lambda should be normalized by the norm of the image, and be increased as the major-cycles progress to maximize RMSE/PSNR. We use:

$$\lambda_{\mathcal{L}_n} = 0.05\|\tilde{i}_{\mathcal{L}_n}\|_2 \times 2^n$$

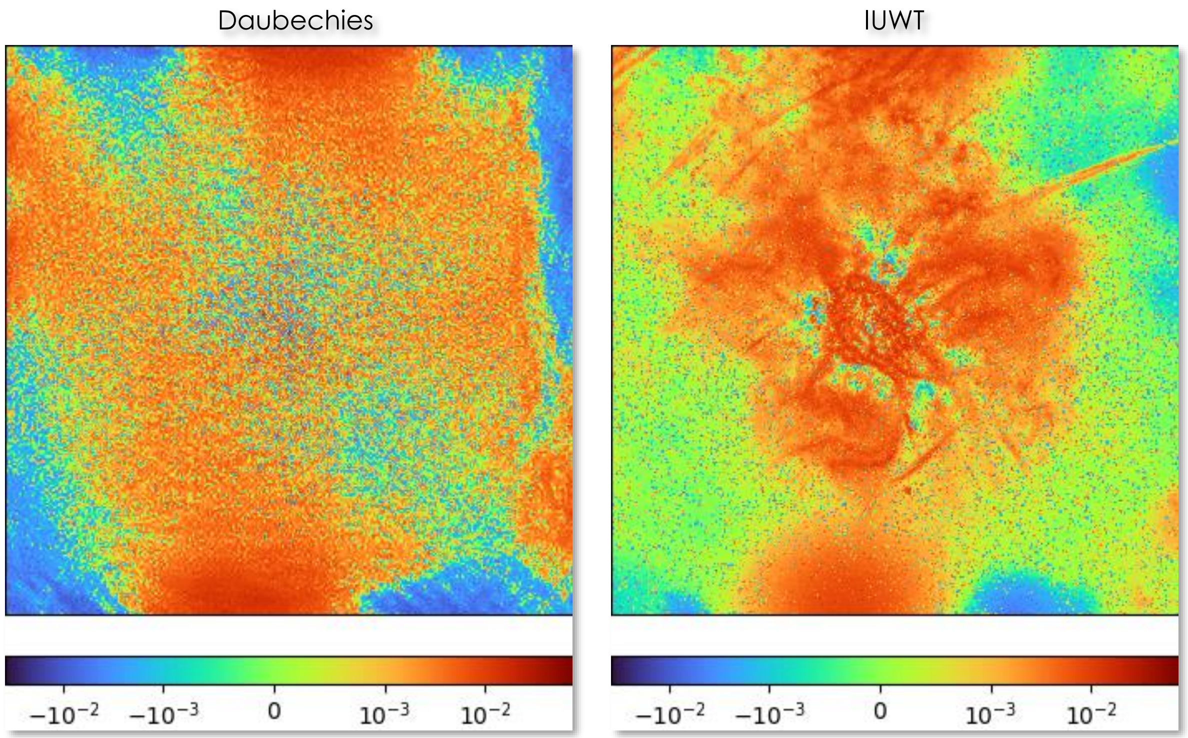
$$\lambda_{\mathcal{H}_n} = 0.05\|\tilde{i}_{\mathcal{H}_n}\|_2 \times 2^n$$

$$\lambda_n = 0.01\|\tilde{i}_n\|_2 \times 2^n$$

IUWT vs Daubechies



First major-cycle residuals for Sgr A

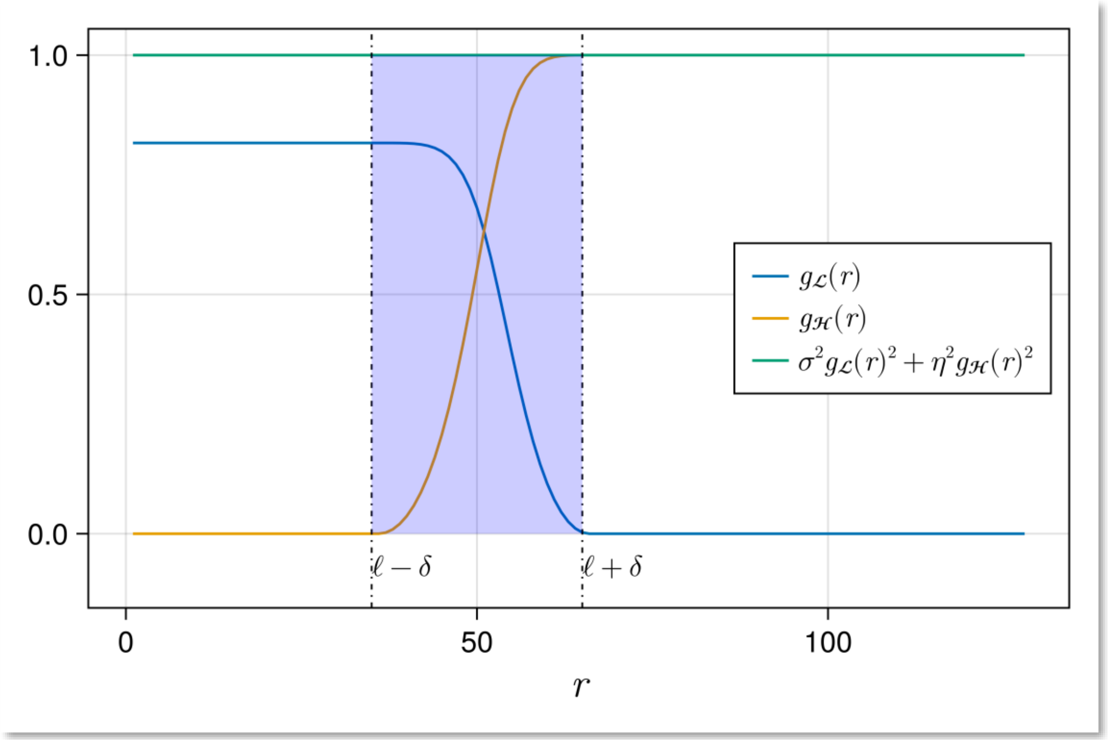


IUWT seems worse at reconstructing large-scale extended emissions, possibly due to its isotropic nature.

Filters

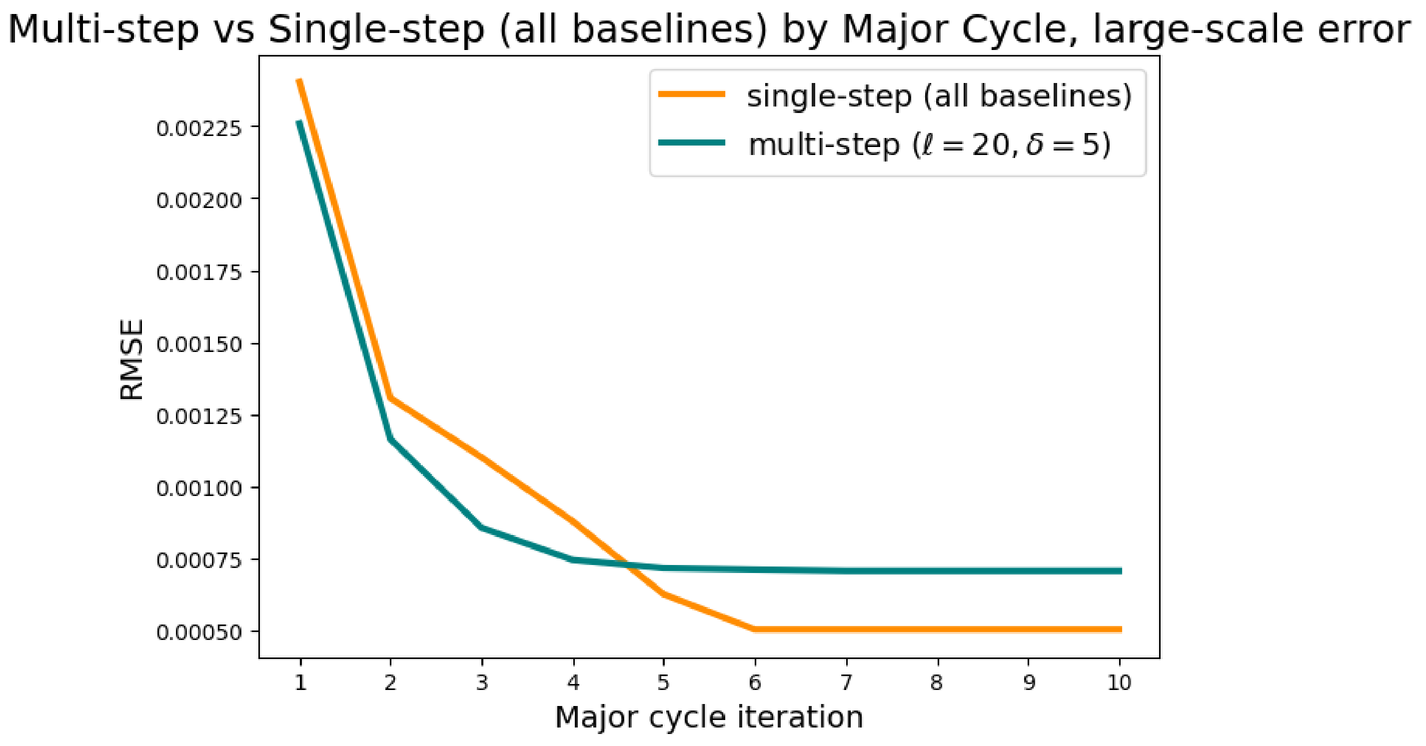
$$r > \ell + \delta : \quad |g_{\mathcal{H}}(r)|^2 = 1/\sigma^2,\ g_{\mathcal{L}}(u) = 0$$
$$r < \ell - \delta : \quad g_{\mathcal{H}}(r) = 0,\ |g_{\mathcal{L}}(r)|^2 = 1/\eta^2$$
$$\ell - \delta < r < \ell + \delta : \quad \sigma^2 |g_{\mathcal{H}}(r)|^2 + \eta^2 |g_{\mathcal{L}}(r)|^2 = 1$$

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

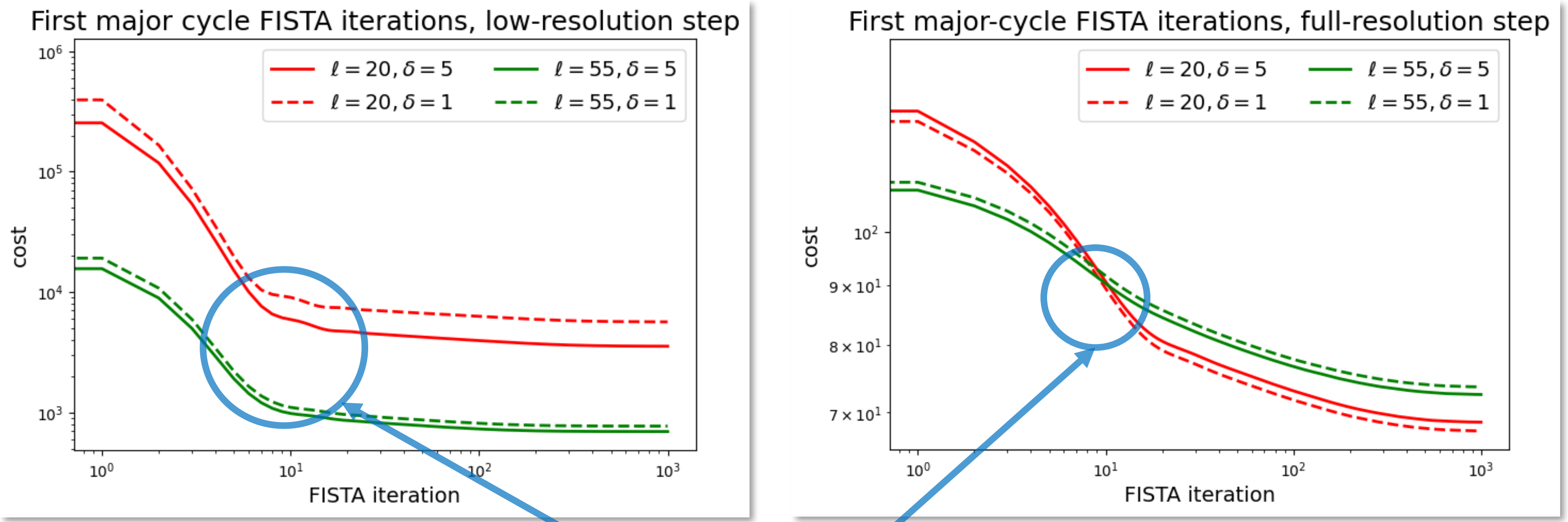


Reconstruction of large-scale features

Full-resolution step does not seem to change the large-scales

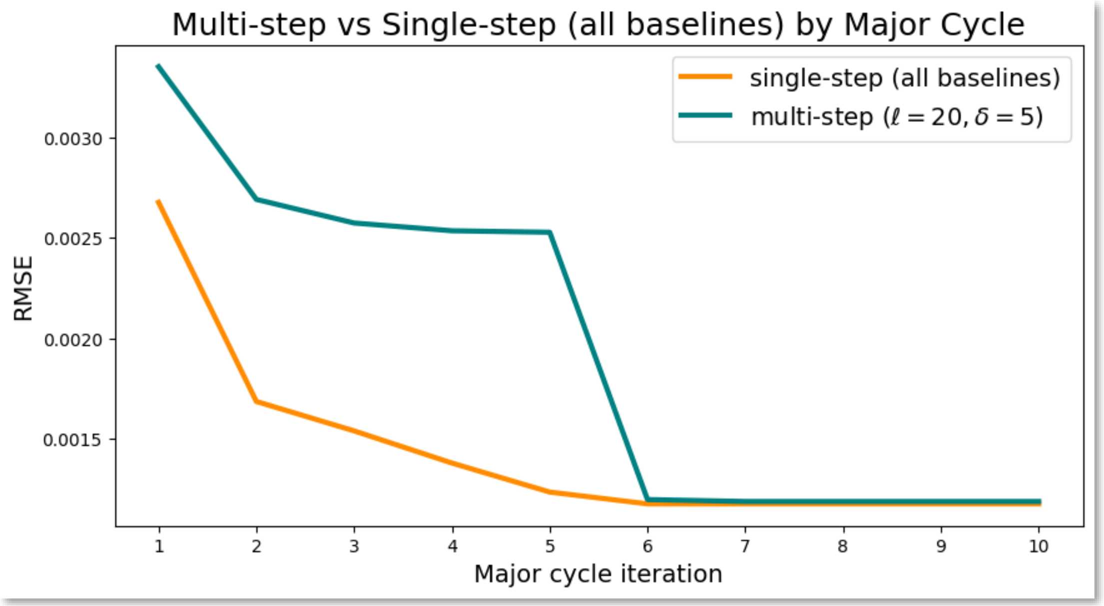


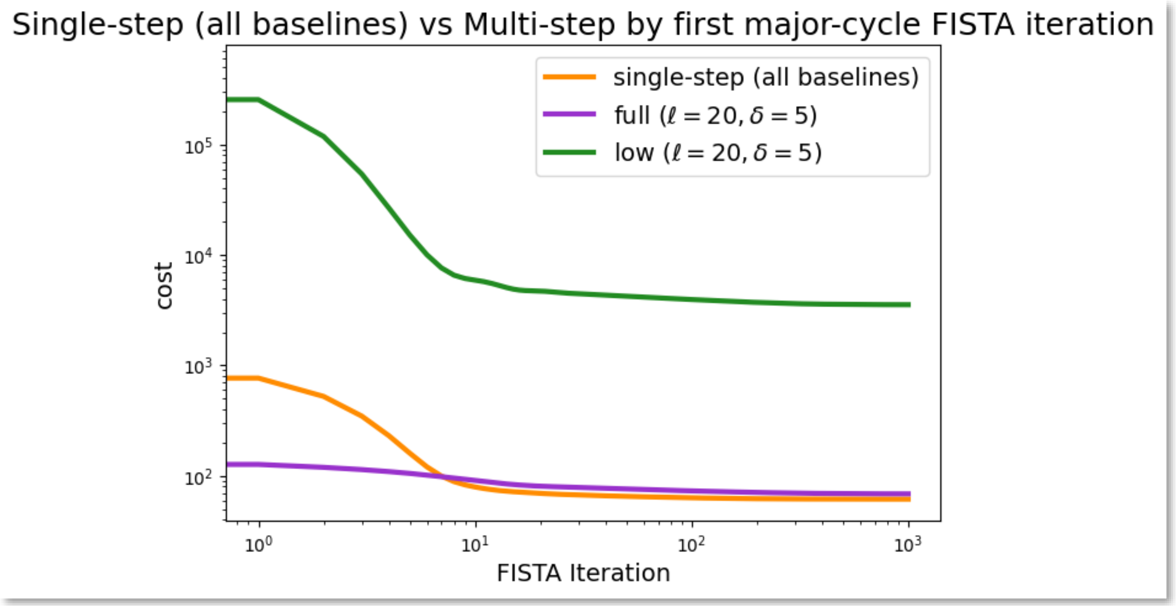
Results: Partition configuration affect on reconstruction accuracy and speed, FISTA



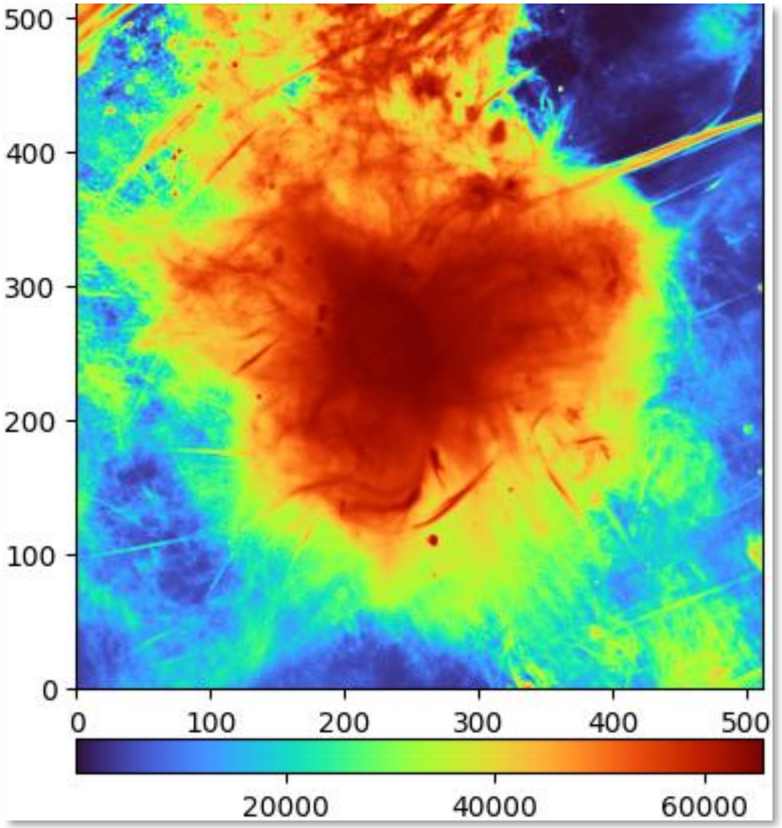
Convergence roughly the same, gradient of cost function remains very similar between partition configurations and starts tapering off around 10-25 iterations.

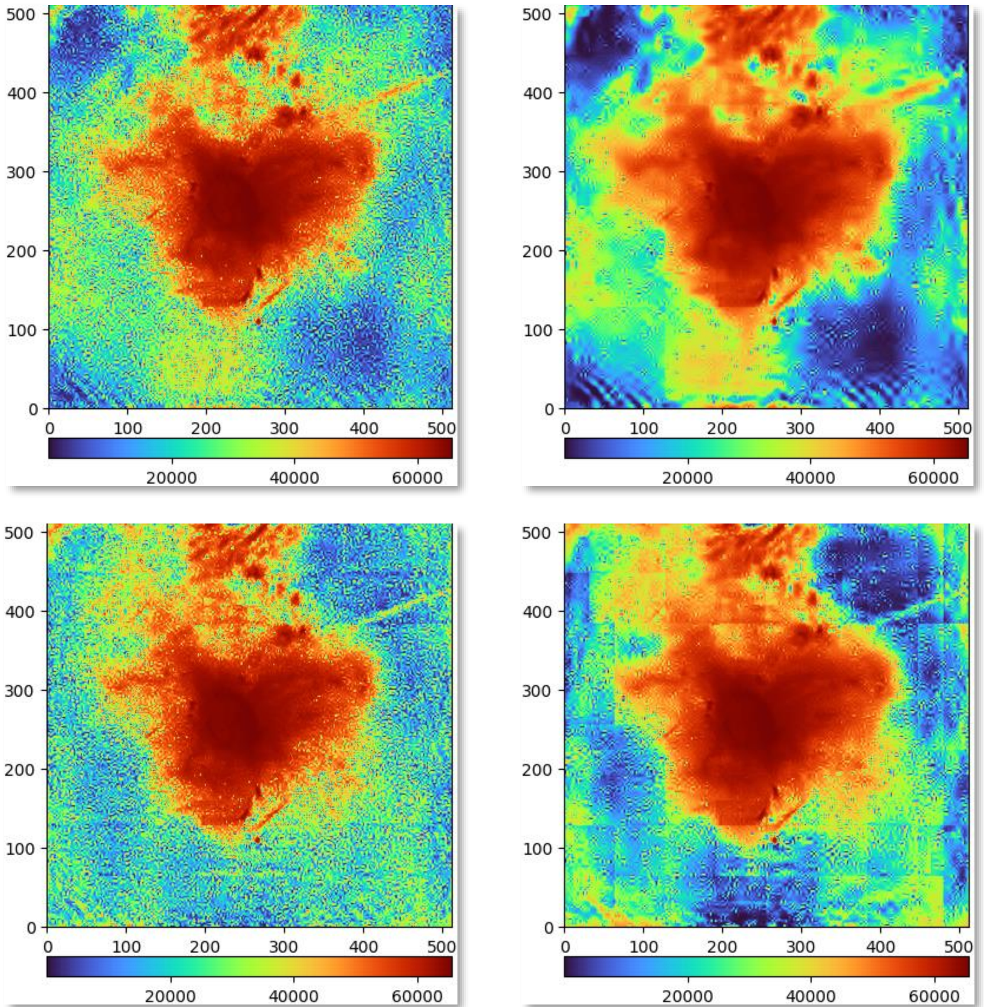
Results: Comparison to all-baselines reconstruction





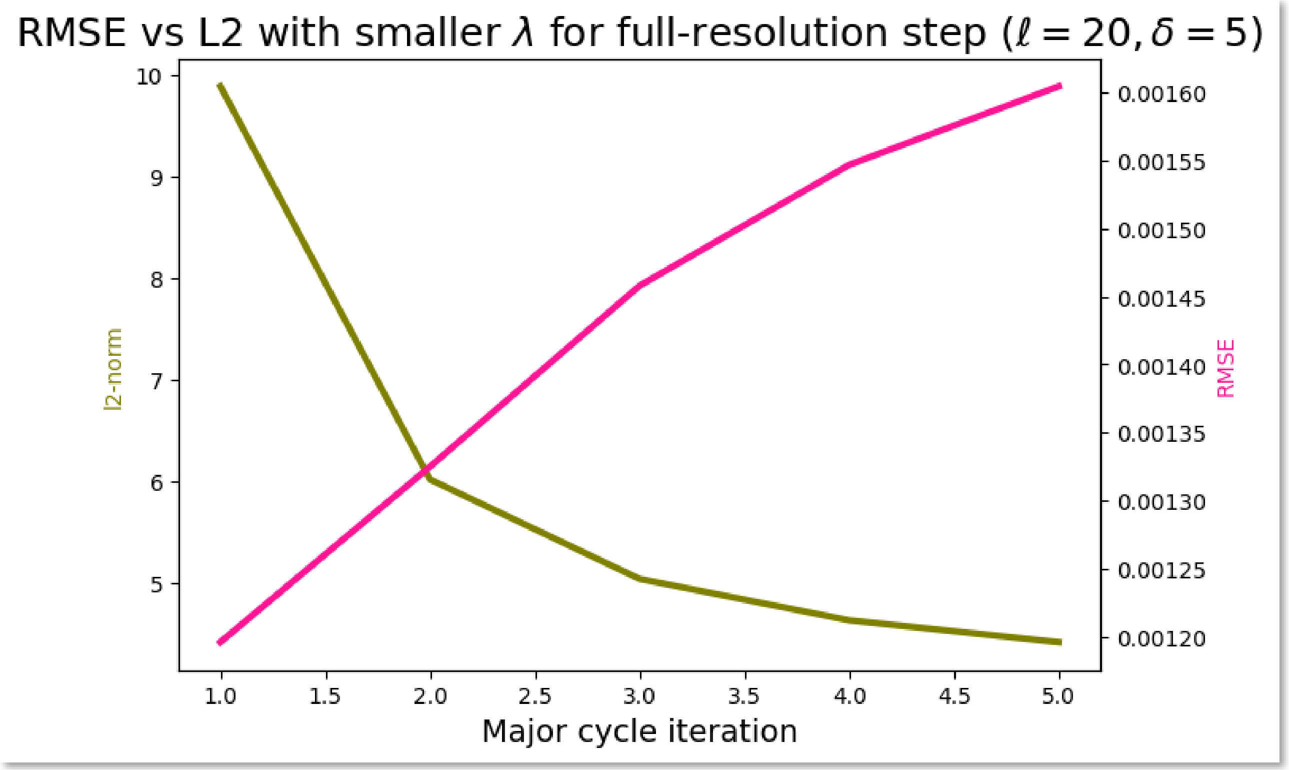
Example image reconstructions (Histogram equalized)



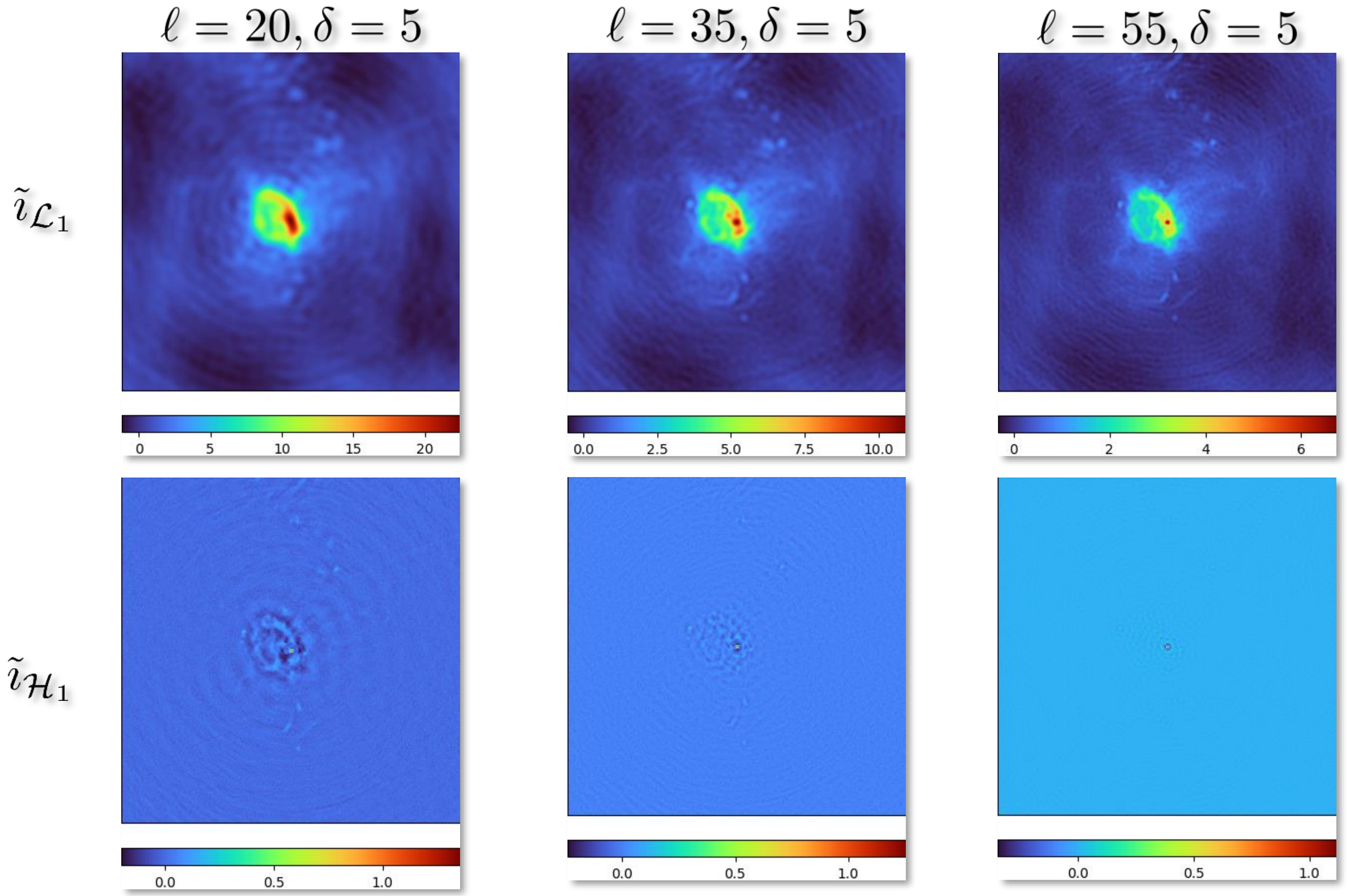


Using a less aggressive lambda for the full-resolution step

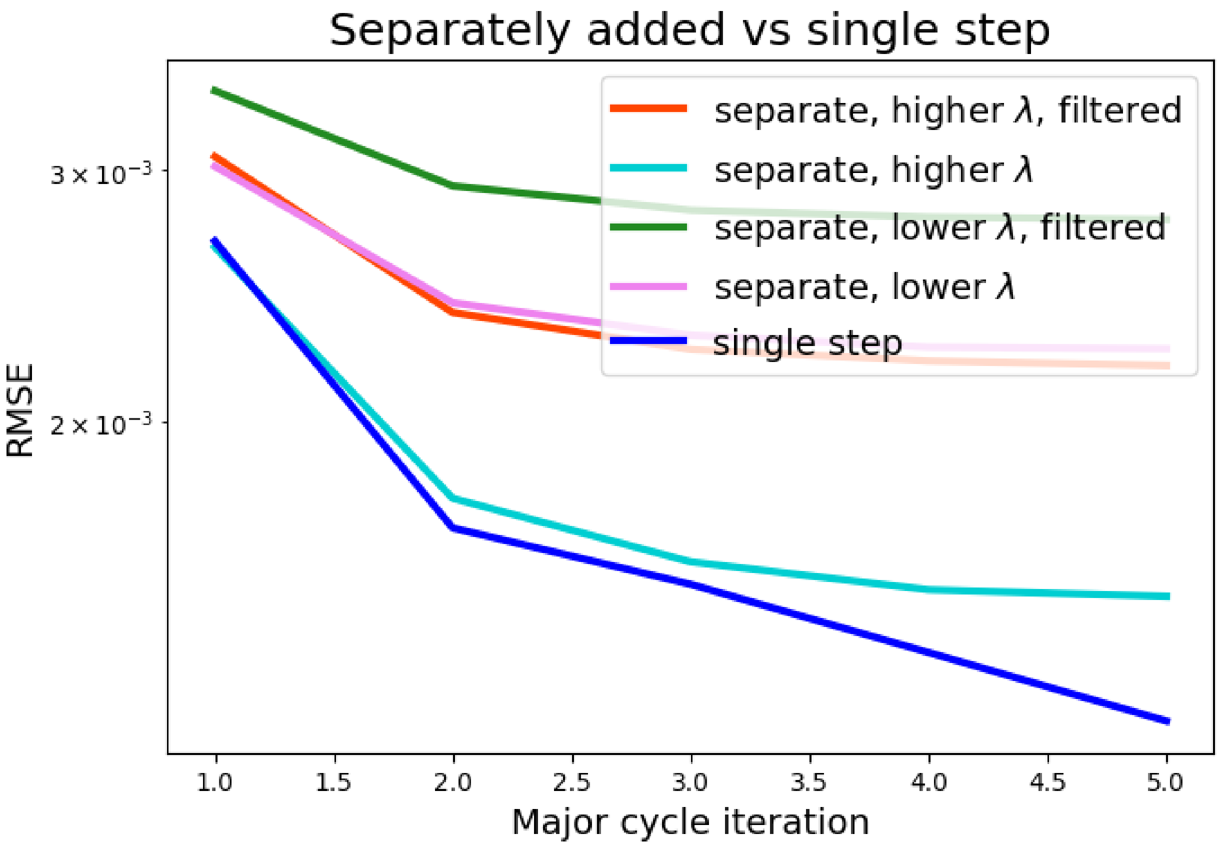
A less aggressive lambda can result in gradual reduction of residual. This can overfit to noise, ergo the worsening RMSE.



Visualization of partition configurations

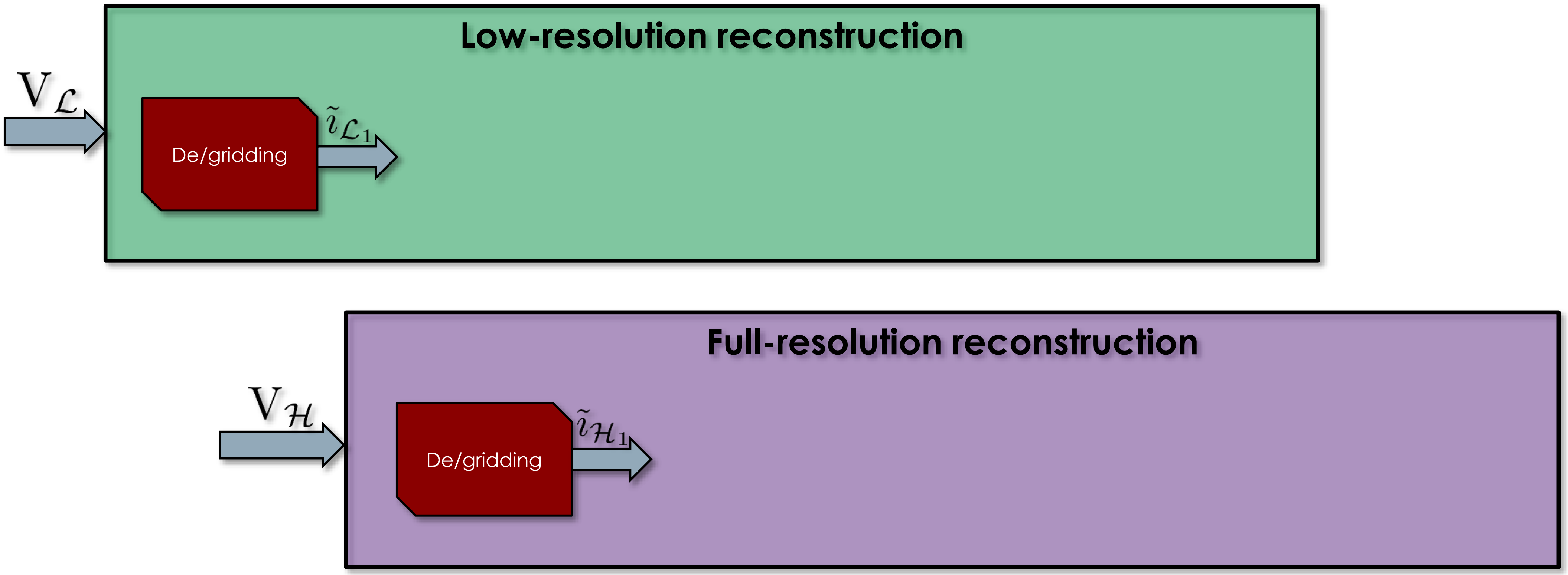


Just adding separately deconvolved images

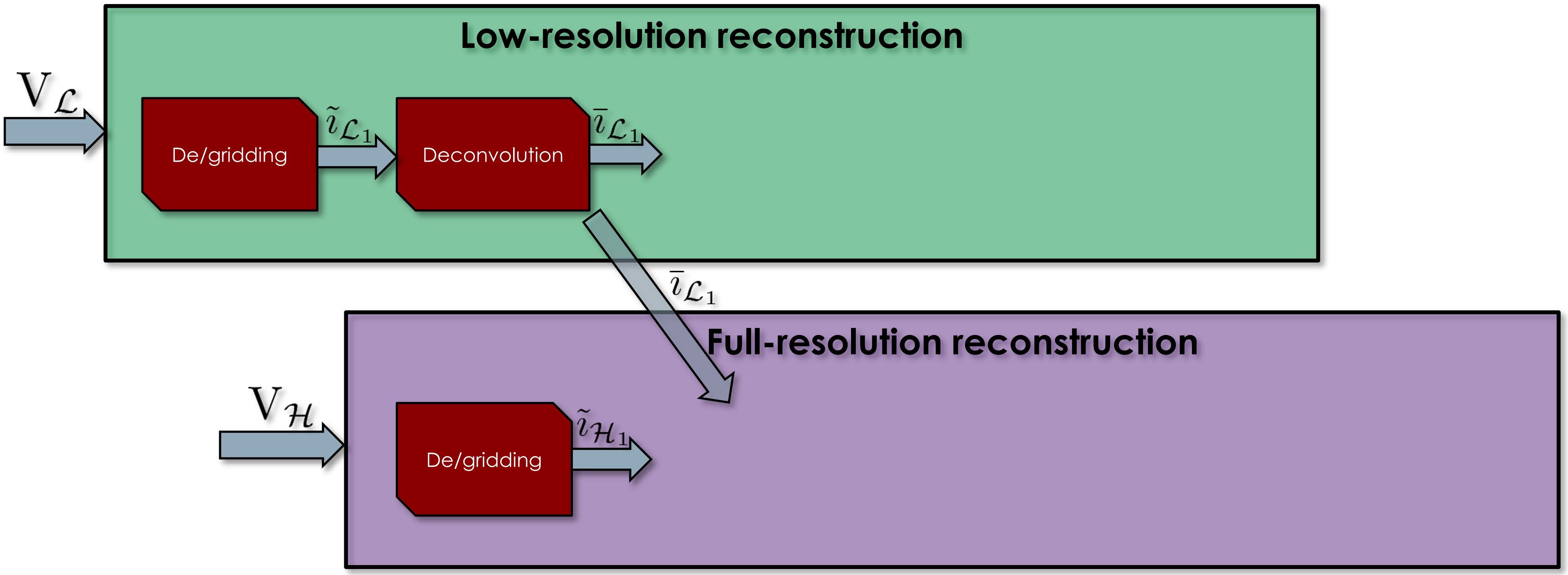


Pipelined parallelization strategy

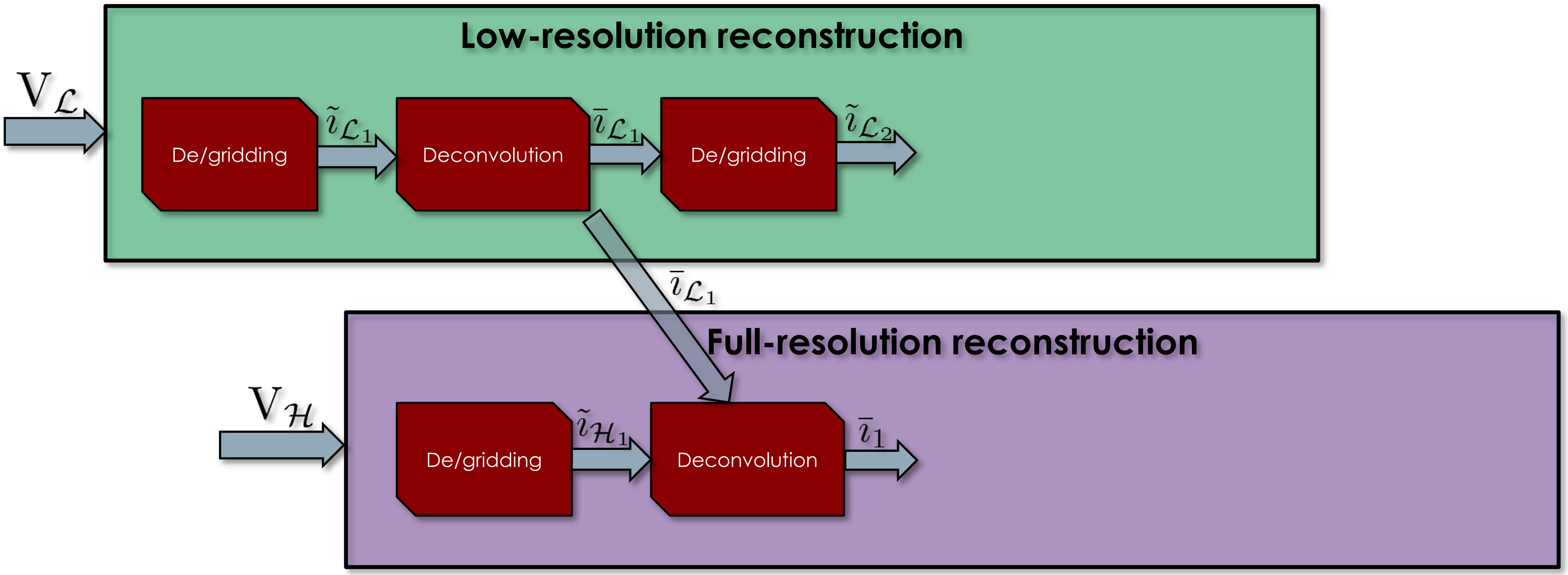
Pipelined parallelization strategy



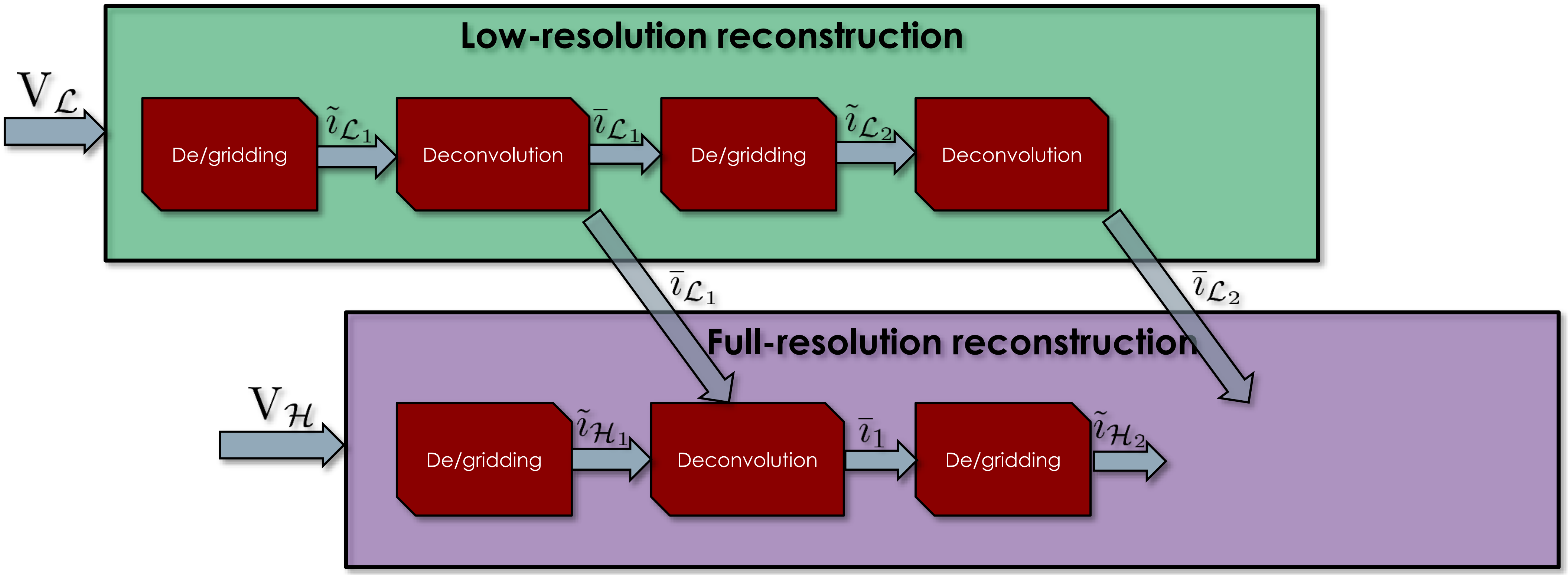
Pipelined parallelization strategy



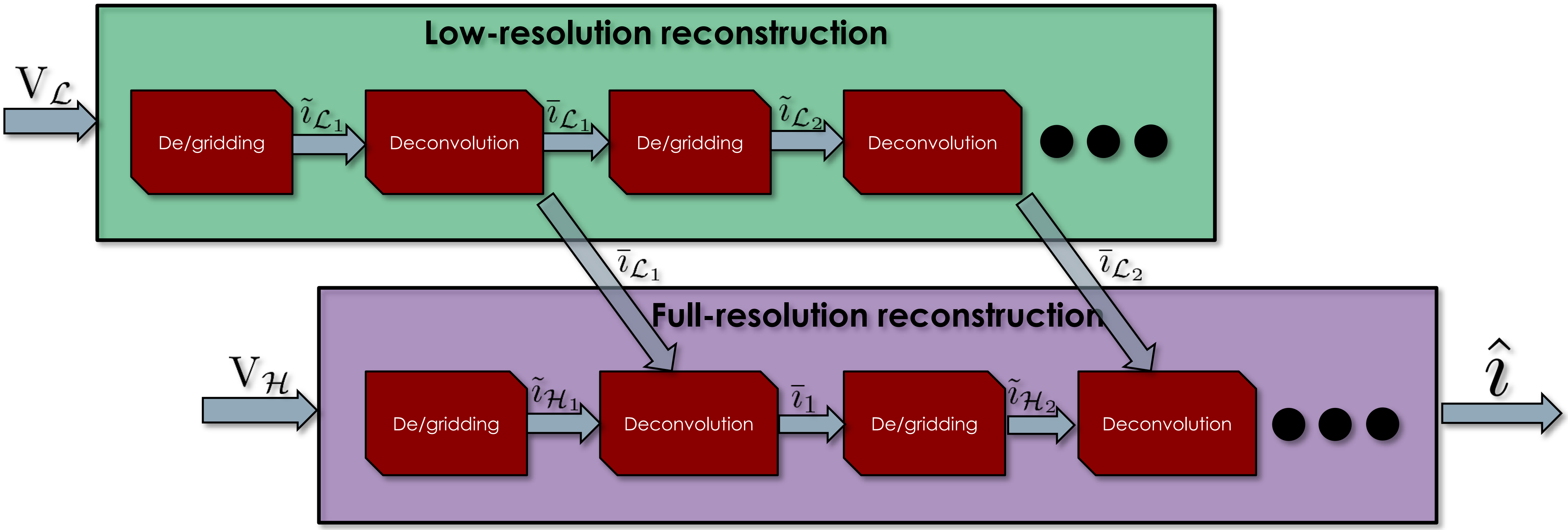
Pipelined parallelization strategy



Pipelined parallelization strategy



Pipelined parallelization strategy



Pipelined parallelization strategy

- Low-resolution step remains the same
- For the full-resolution step:

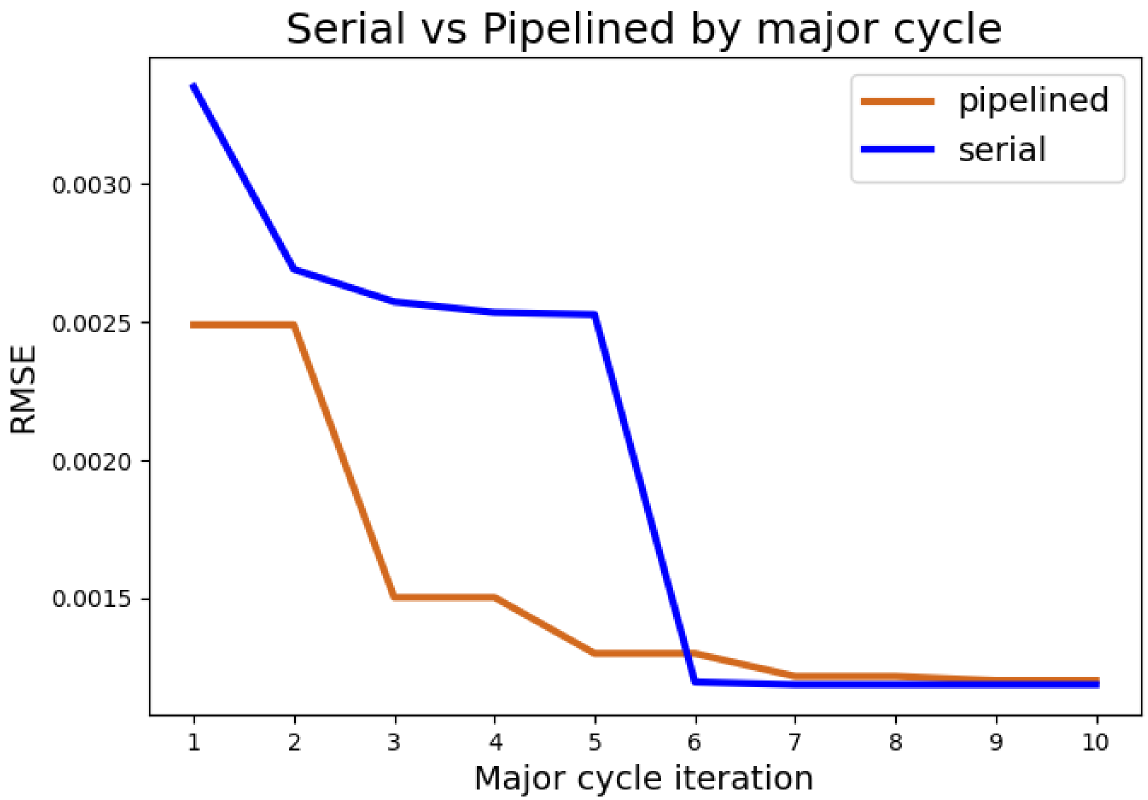
$$l_n = \sum_{j=1}^{N} \bar{\imath}_{\mathcal{L}_j} - \sum_{j=1}^{n-1} \bar{\imath}_j = \hat{\imath}_{\mathcal{L}} - \hat{\imath}_{n-1}$$

changes to:

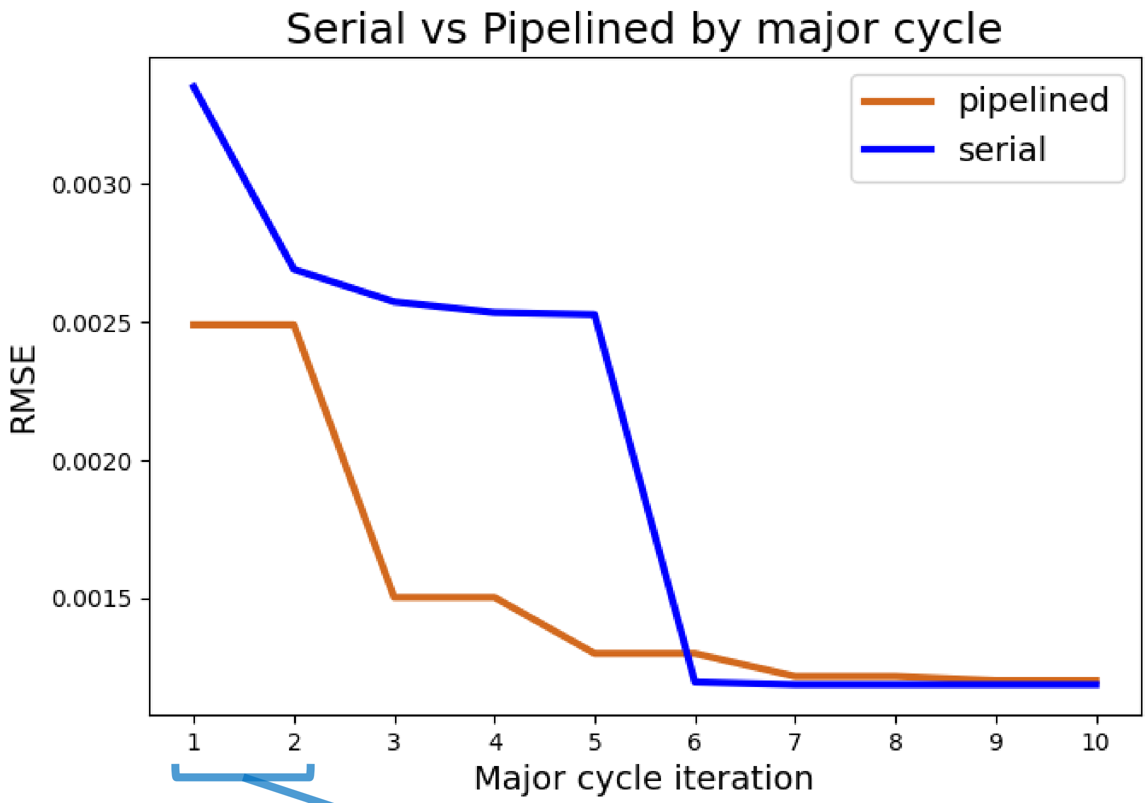
$$l_n = \sum_{j=1}^{n} \bar{\imath}_{\mathcal{L}_j} - \sum_{j=1}^{n-1} \bar{\imath}_j = \hat{\imath}_{\mathcal{L}_{n-1}} - \hat{\imath}_{n-1} + \bar{\imath}_{\mathcal{L}_n}$$

- Can result in waiting if computation costs of each step not similar
  - Asynchronous strategy can alleviate this somewhat
- Quality upper bound of serial (but most likely slightly worse)
- One image transmitted per major cycle

Results: Reconstruction quality of pipelined parallel method

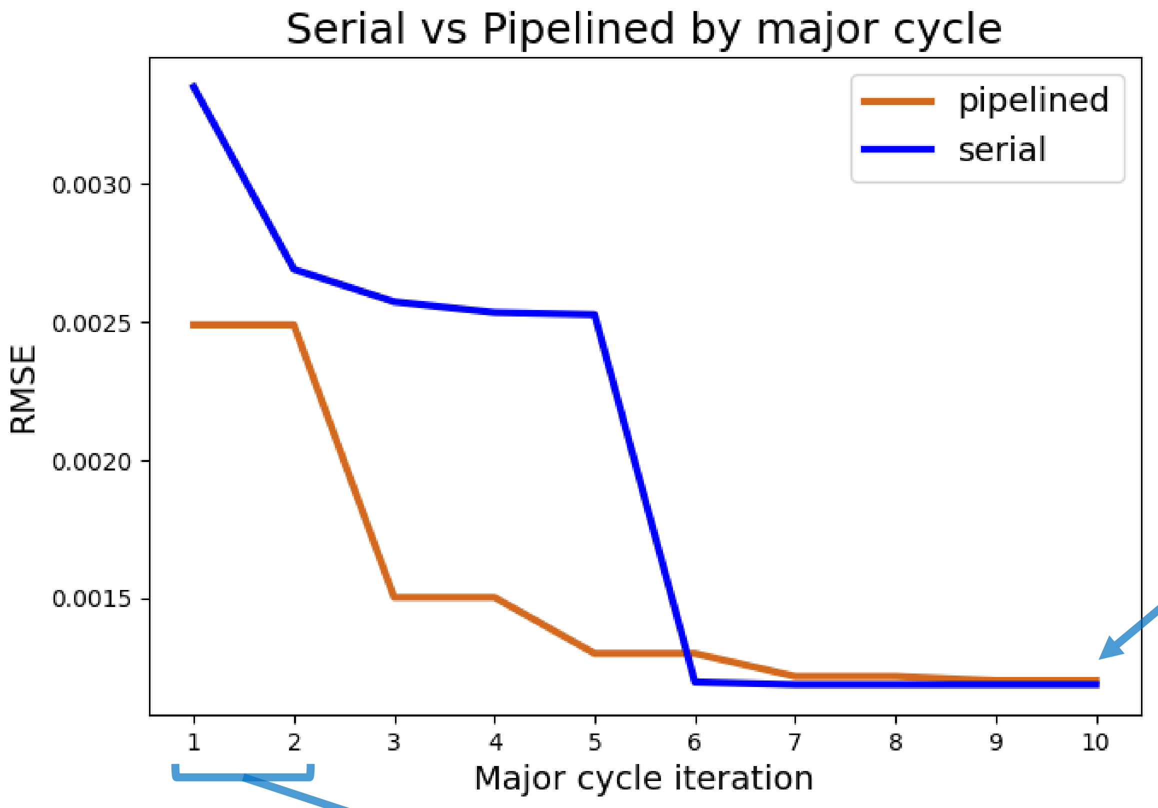


Results: Reconstruction quality of pipelined parallel method



For pipelined, only showing full-resolution reconstruction after 1 low and 1 full-resolution major-cycle

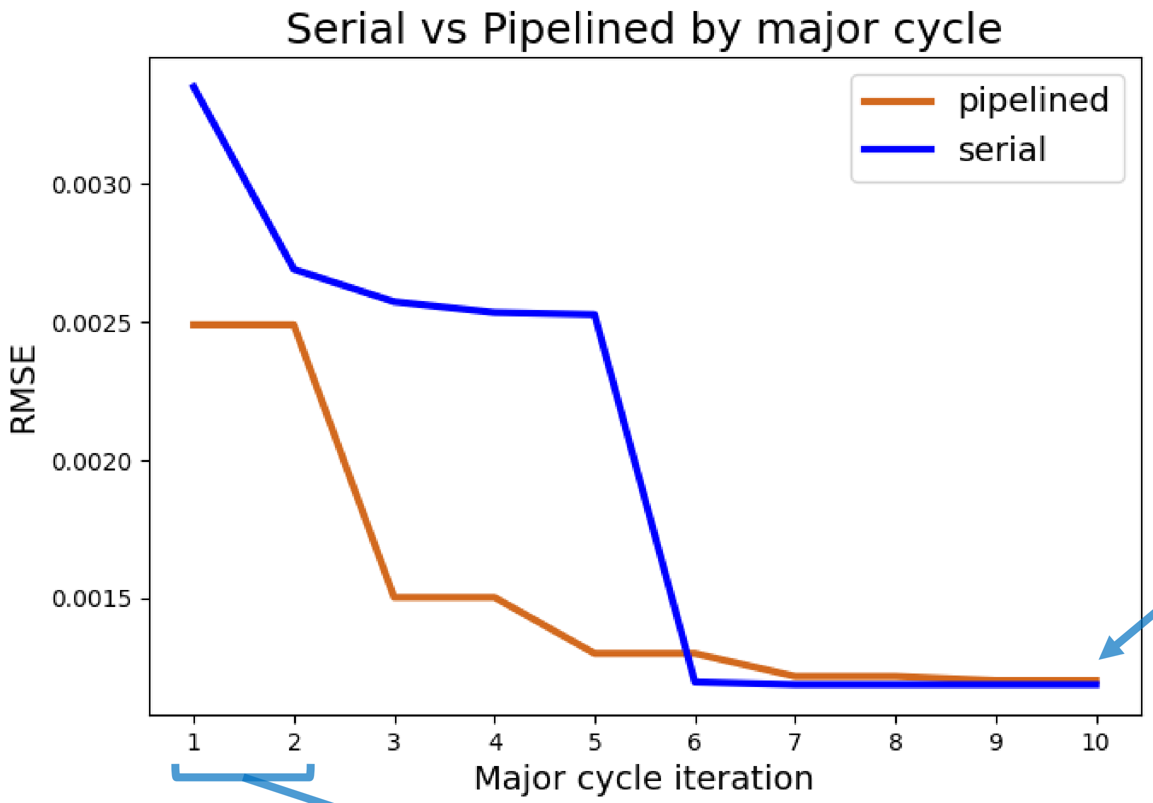
Results: Reconstruction quality of pipelined parallel method



As expected, converges to similar result but slightly worse

For pipelined, only showing full-resolution reconstruction after 1 low and 1 full-resolution major-cycle

Results: Reconstruction quality of pipelined parallel method



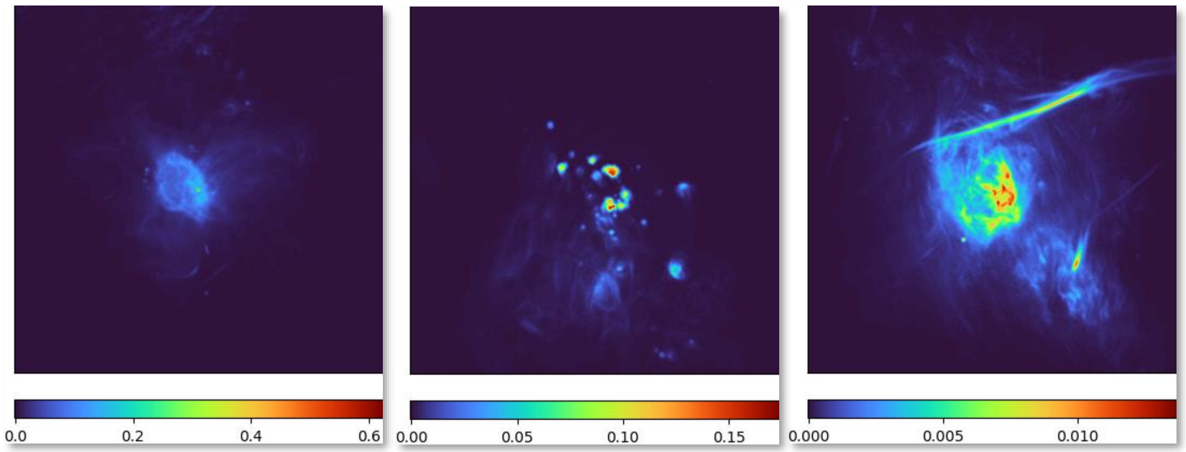
As expected, converges to similar result but slightly worse

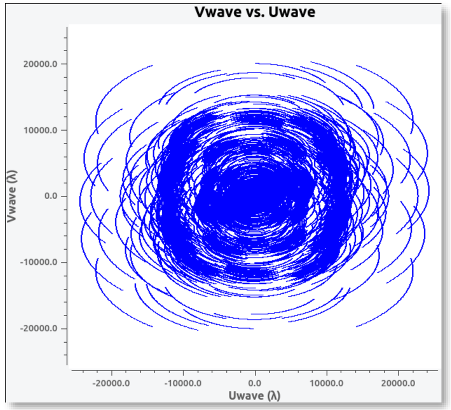
For pipelined, only showing full-resolution reconstruction after 1 low and 1 full-resolution major-cycle

Trading off quality for parallelism

Results – Measurement sets

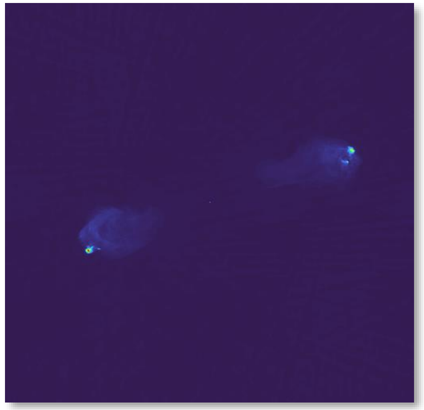
Simulated

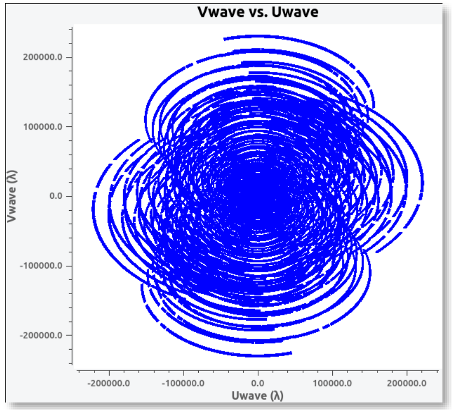




- ➢ Initial images tapered and cutout from 1.28GHz mosaic produced in [1]
- ➢ Visibilities generated with Meerkat configuration
- ➢ Exposure time of 4h, samples every 120s for to generate visibility positions
- ➢ Degrid to get visibility values
- ➢ Visibility noise artificially added

Real





- ➢ Measurement set fully described in [2]
- ➢ VLA, all 4 configurations, S-band
- ➢ Rather than the full dataset, we only use visibilities from 1988.5 MHz – 2020.5 MHz

[1] Heywood, I., et al. "The 1.28 GHz MeerKAT galactic center mosaic." The Astrophysical Journal 925.2 (2022): 165.
[2] Sebokolodi, M. Lerato L., et al. "A Wideband Polarization Study of Cygnus A with the Jansky Very Large Array. I. The Observations and Data." The Astrophysical Journal 903.1 (2020): 36.