Efficient large-scale seismic data acquisition and processing using rank minimization

Rajiv Kumar

SLIM
University of British Columbia

Motivation

- Large-scale data acquisition and processing
  - source separation + interpolation
  - interpolate missing data (conventional acquisition)

Motivation

- Exploit low-rank structure of seismic data (2D & 3D)
  - SVD-free rank penalization techniques
  - embarrassingly Parallelized framework
  - super-fast algorithms on full seismic data volumes
  - no need to perform window-based operations
  - achieve very high compression ratios
  - very simple code to adapt in your preferred language

Seismic data acquisition—separation and interpolation via rank-minimization

Rajiv Kumar, Shashin Sharan, Haneet Wason, Felix J. Herrmann



SLIM
University of British Columbia

SLIM

Motivation

How to minimize costs of seismic acquisition?

Solution:

- randomize sampling w/ insights from Compressive Sensing to lower cost

New paradigm:

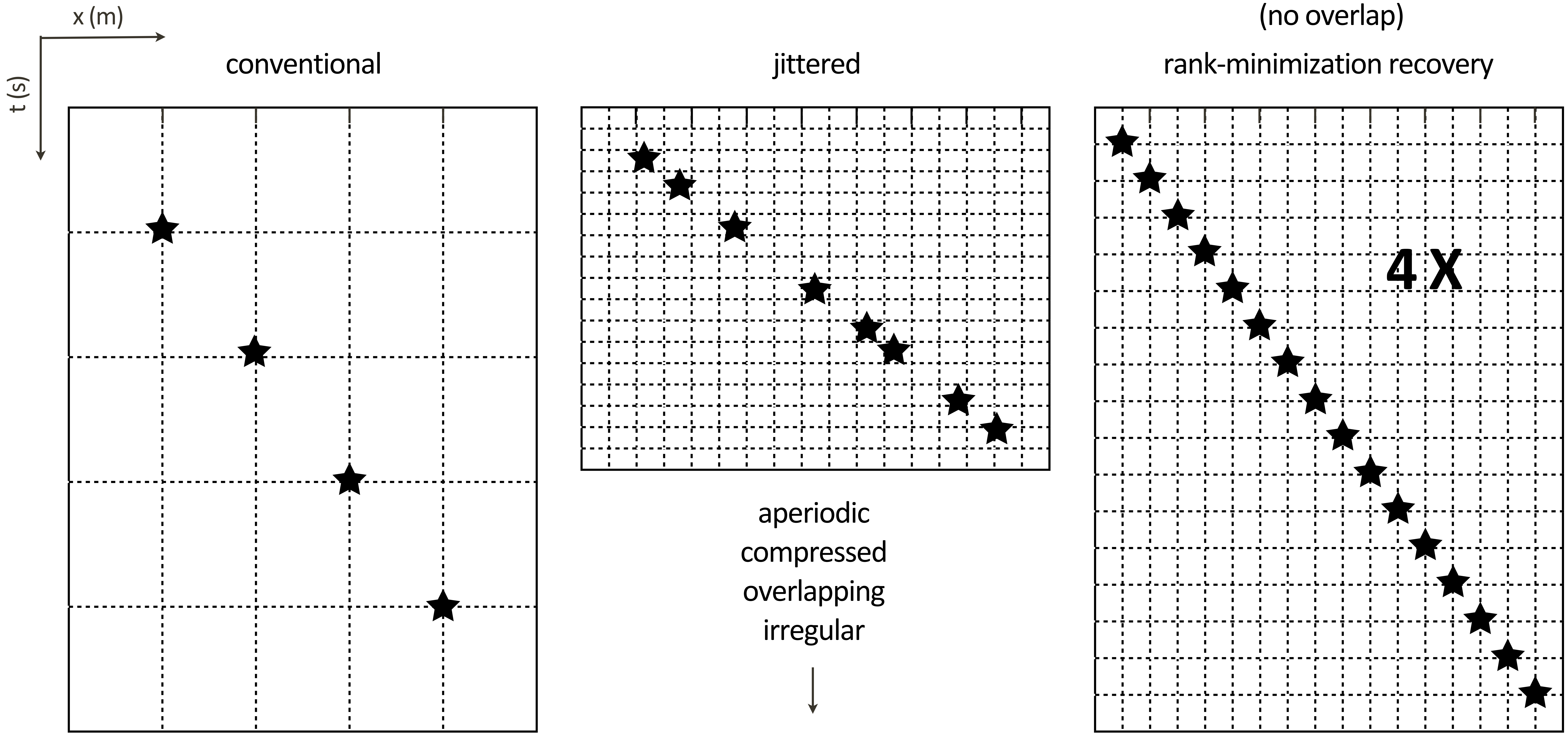
- give up on dense acquisition
- sample coarsely at random
- works as long as we know where we were in the field

Compressive Sensing = increased acquisition productivity

Randomized jitter sampling in marine

(no overlap)

conventional | jittered | rank-minimization recovery



periodic–sparse–no overlap

aperiodic
compressed
overlapping
irregular

separation + regularization
+ interpolation

periodic & dense

source sampling : 25 m (flip-flop) | 25 m (flip-flop) jittered | 6.25 m

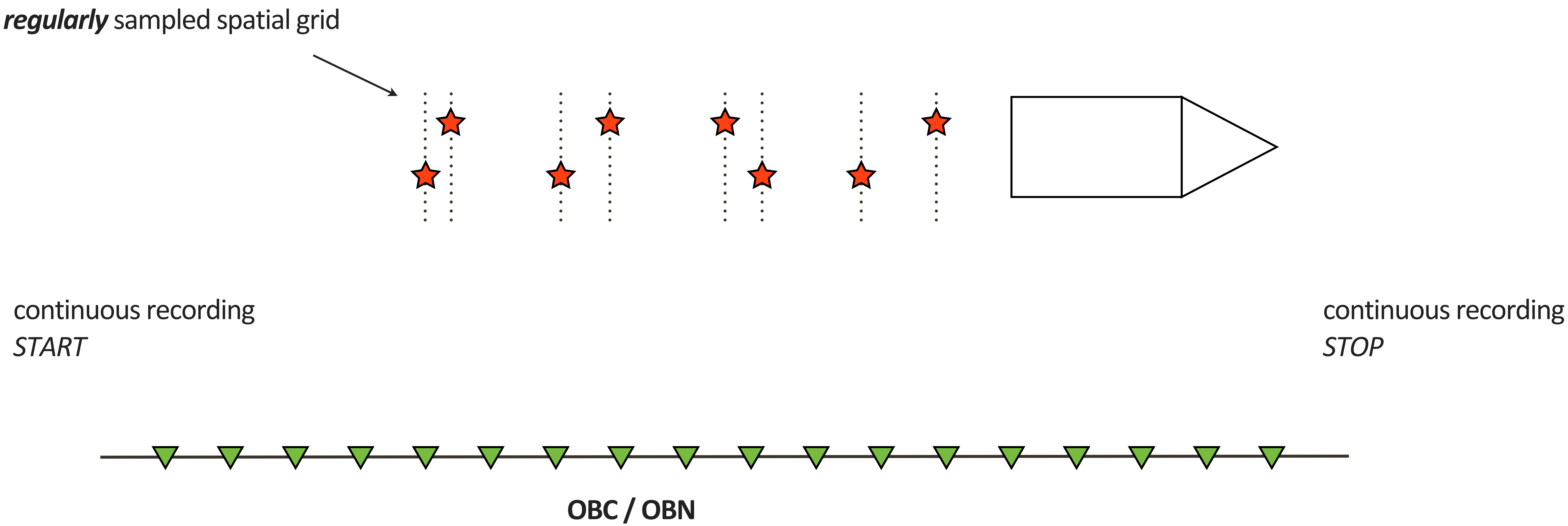
Economical 3D OBN acquisition

Observed grid (m)	Recovered grid (m)	Subsampling %	Economical gain
25	12.5	50	**2X**
25	6.25	75	**4X**
25	3.125	90	**8X - 9X**

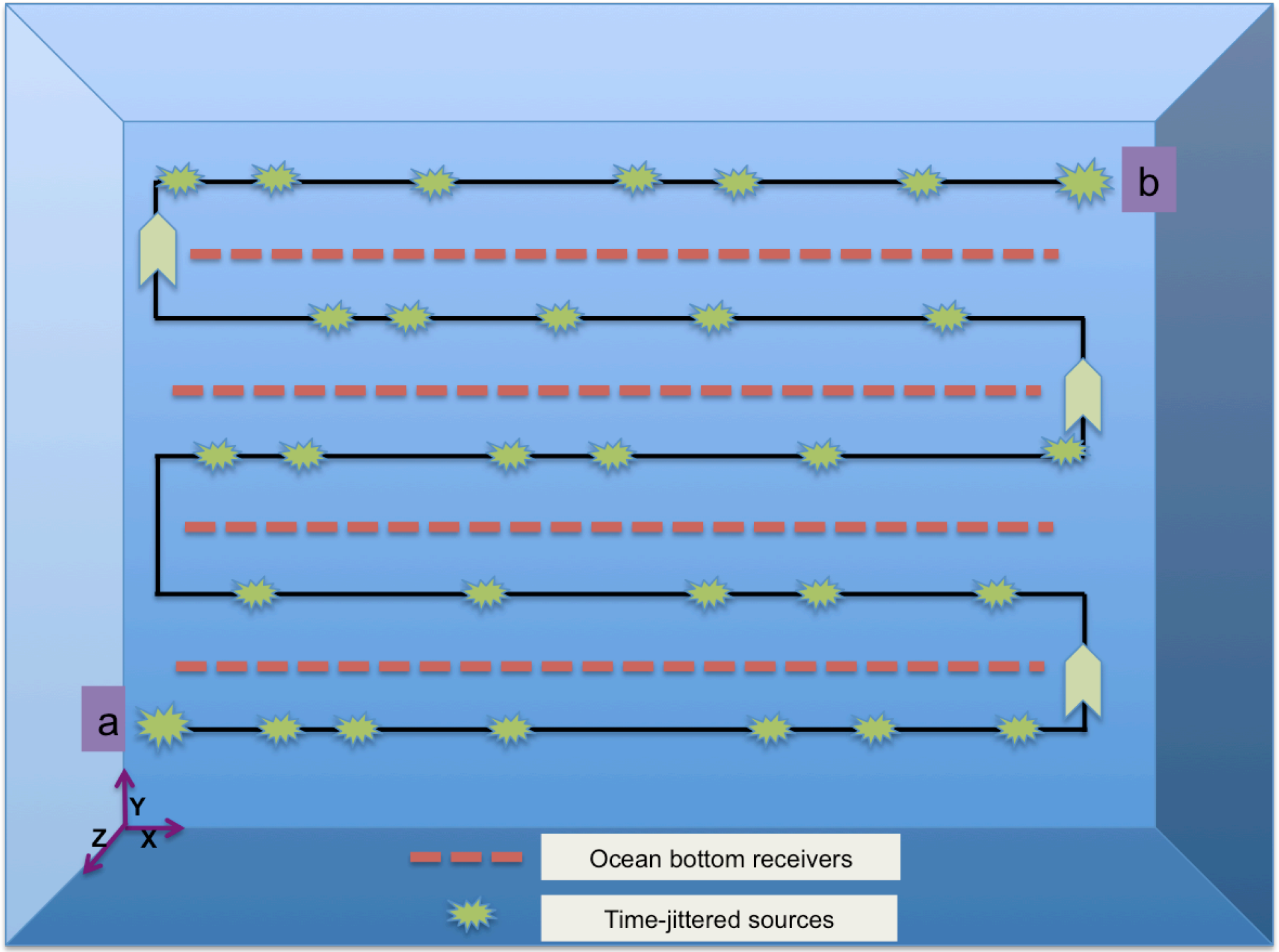
Economical 3D OBN acquisition

Observed grid (m)	Recovered grid (m)	Subsampling %	Economical gain
25	12.5	50	**2X**
25	6.25	75	**4X**
25	3.125	90	**8X - 9X**

Time-jittered acquisition

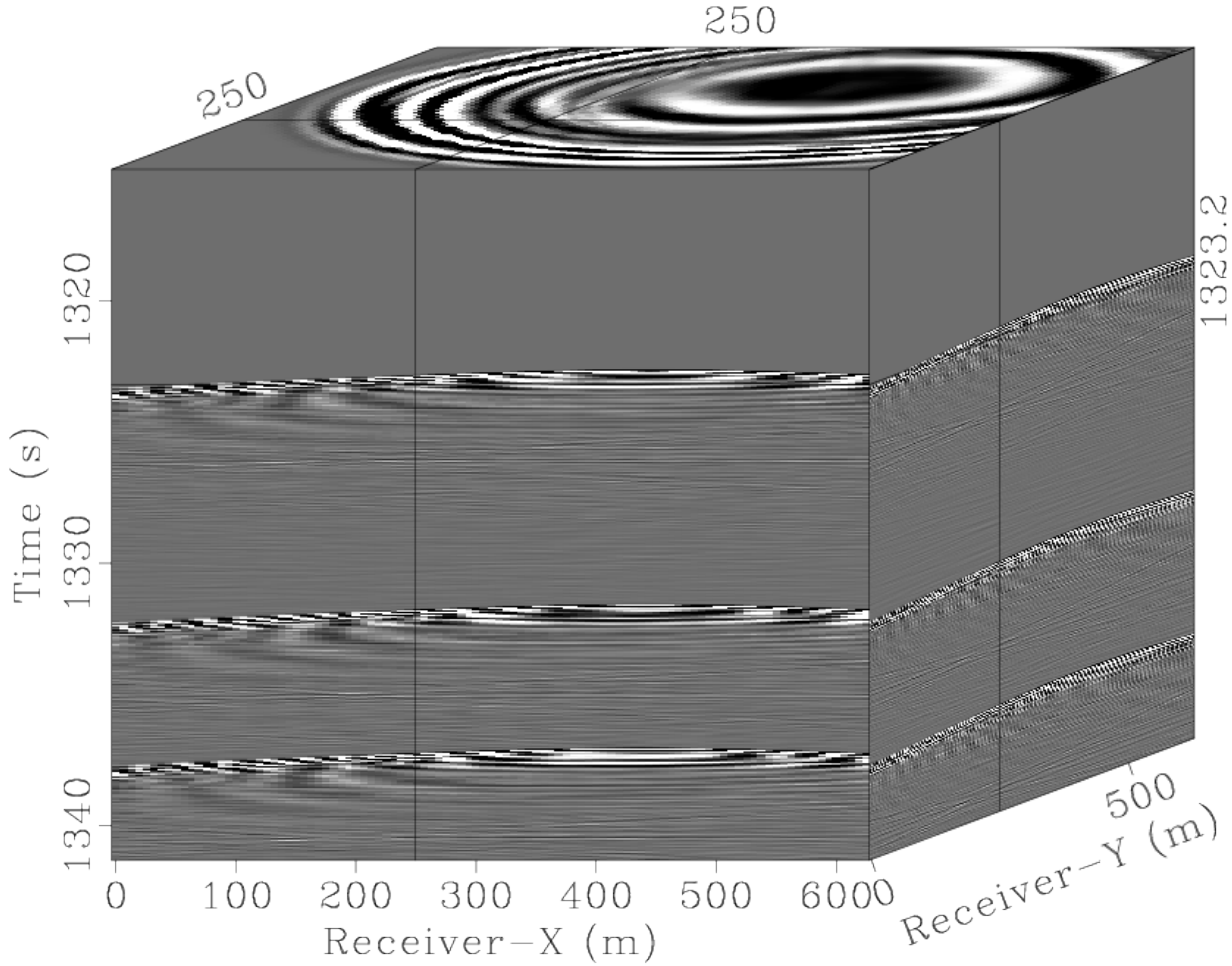


Acquisition setup
speed of source vessel = 5 knots ~ 2.5 m/s



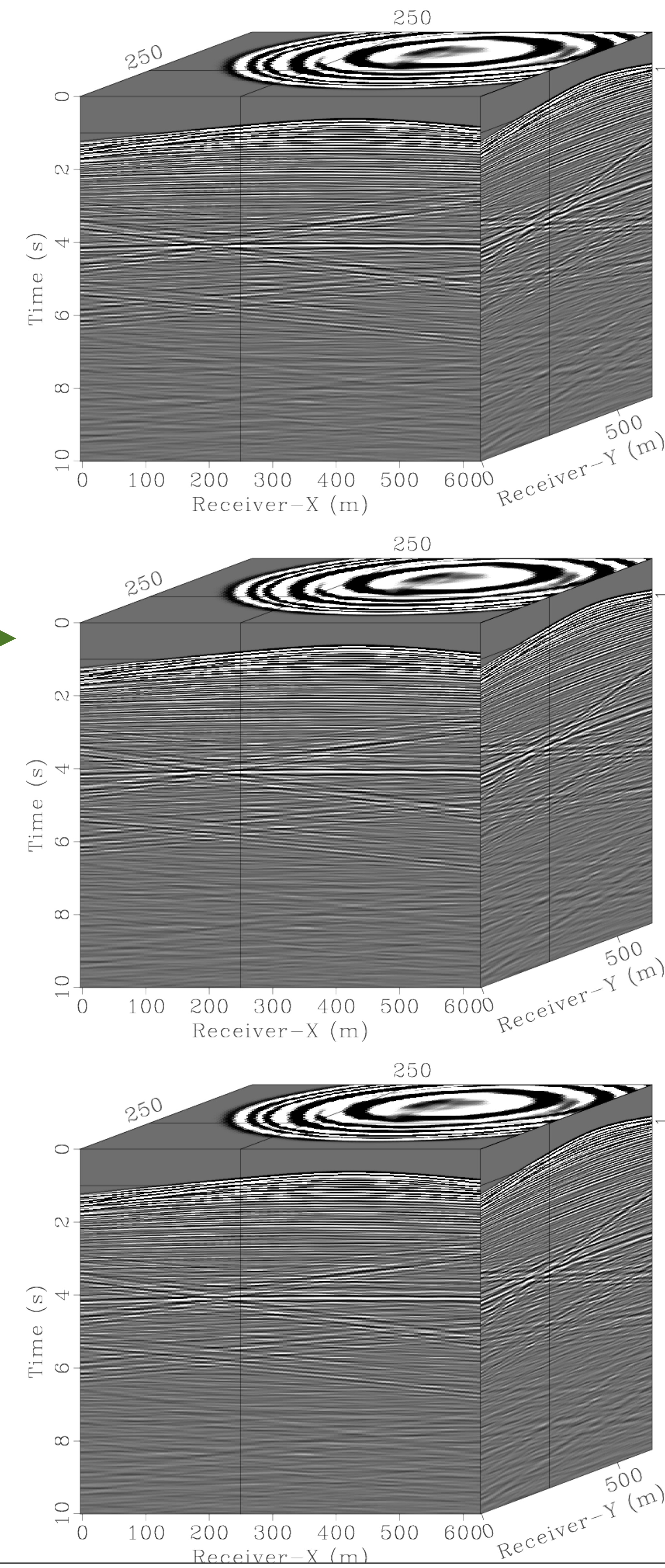
Observed v/s recovered

Observed data @ 25 m flip-flop (overlapping & missing shots)



Recovery

Separation + Interpolation (recovered grid @ 6.25m)



Methodology

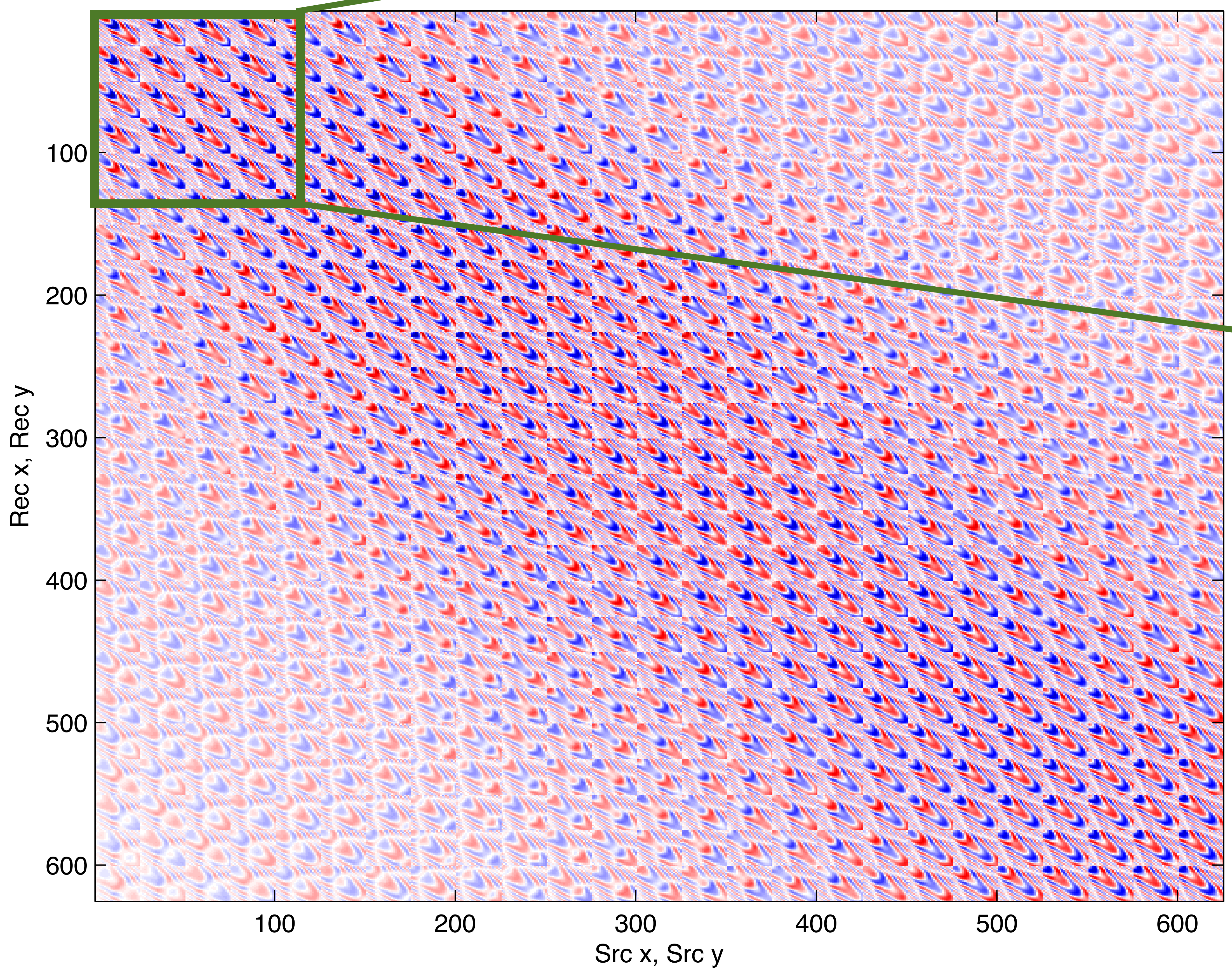
[Candès and Plan, 2009]

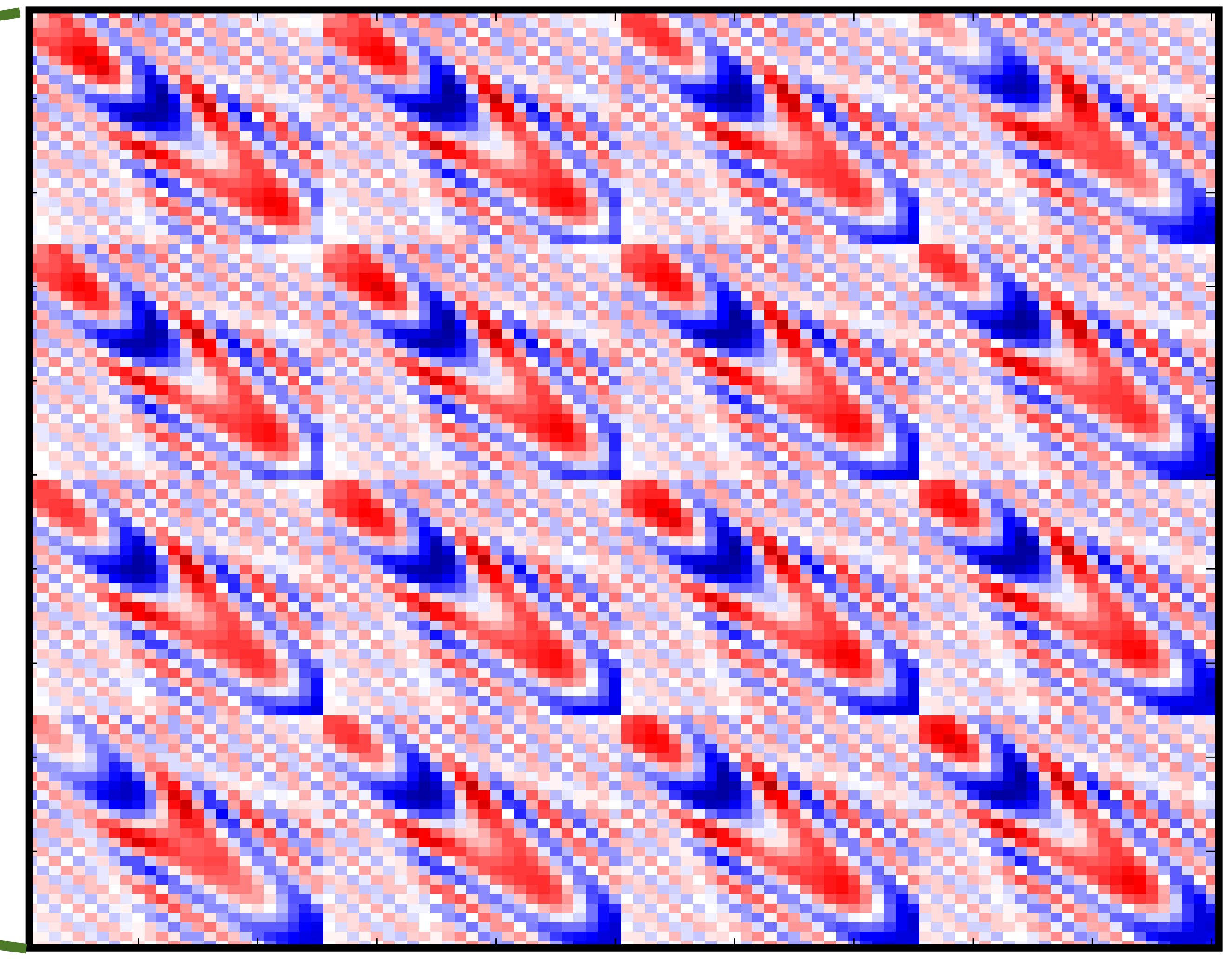
Matrix completion

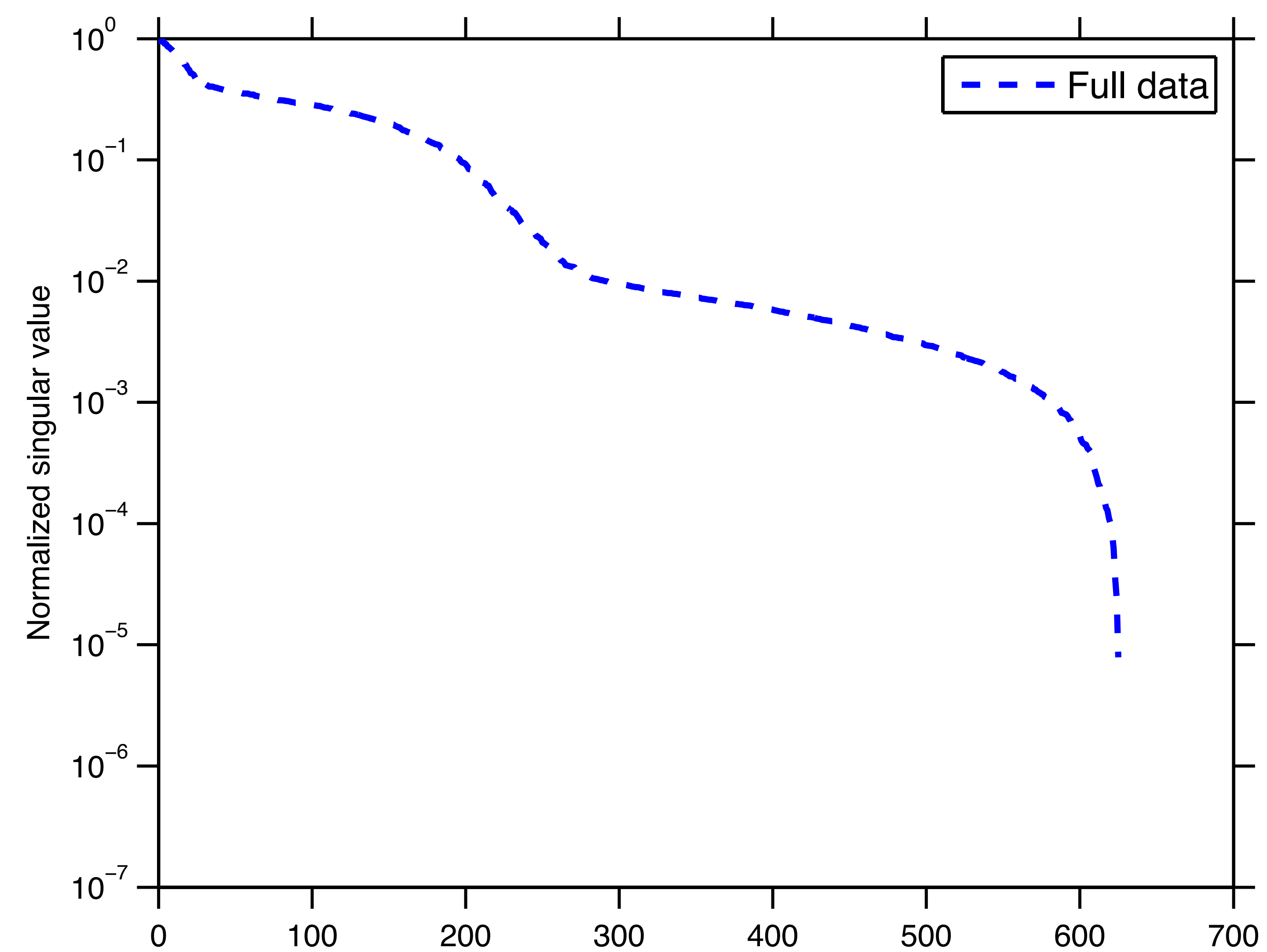
Successful reconstruction scheme

- exploit *structure*
  - *low-rank* / *fast decay* of singular values
- sampling
  - randomness *increases* rank in "transform domain"
- optimization
  - via *rank-minimization (nuclear norm-minimization)*

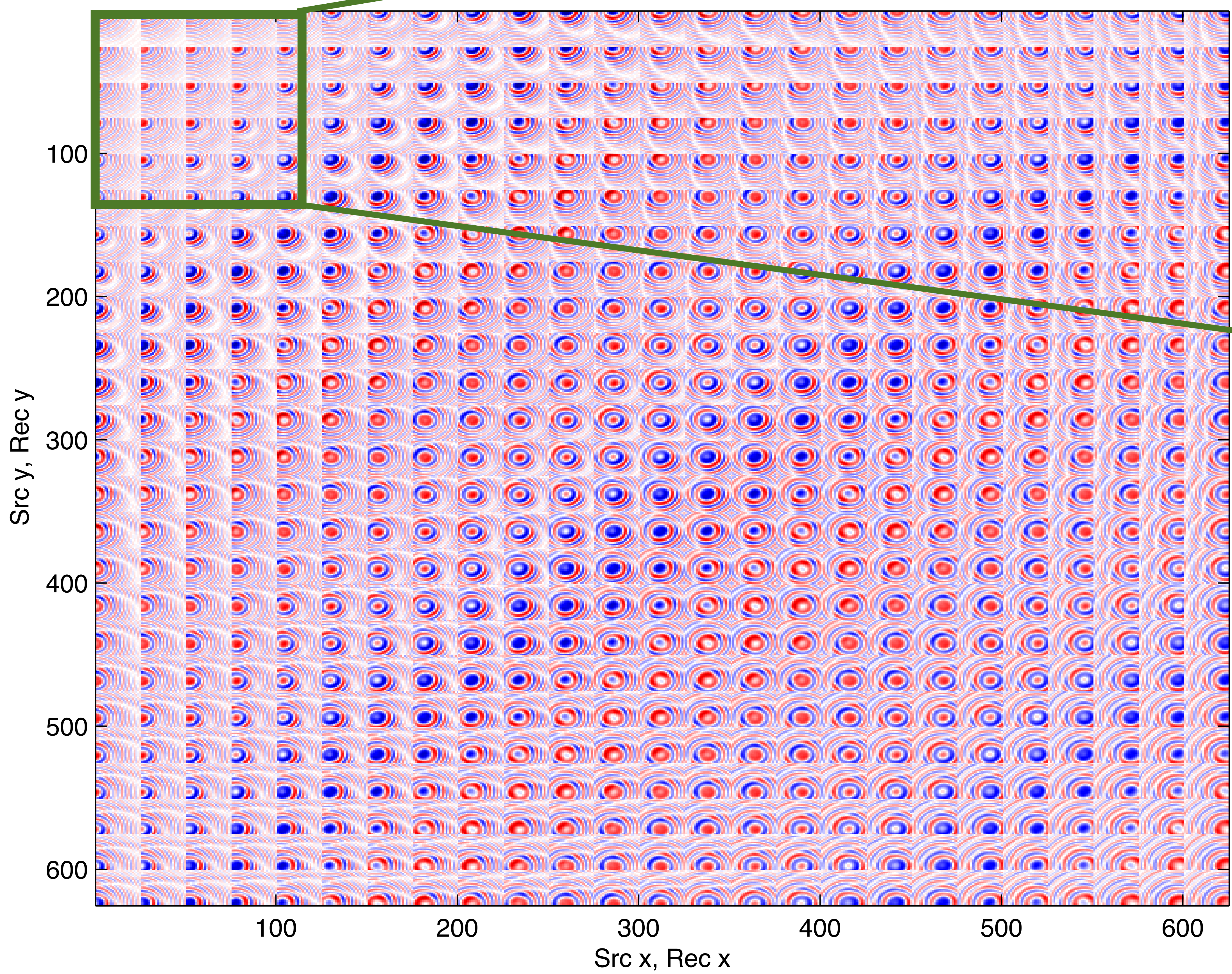
Low-rank structure
conventional 5D data, monochromatic slice, Sx-Sy matricization

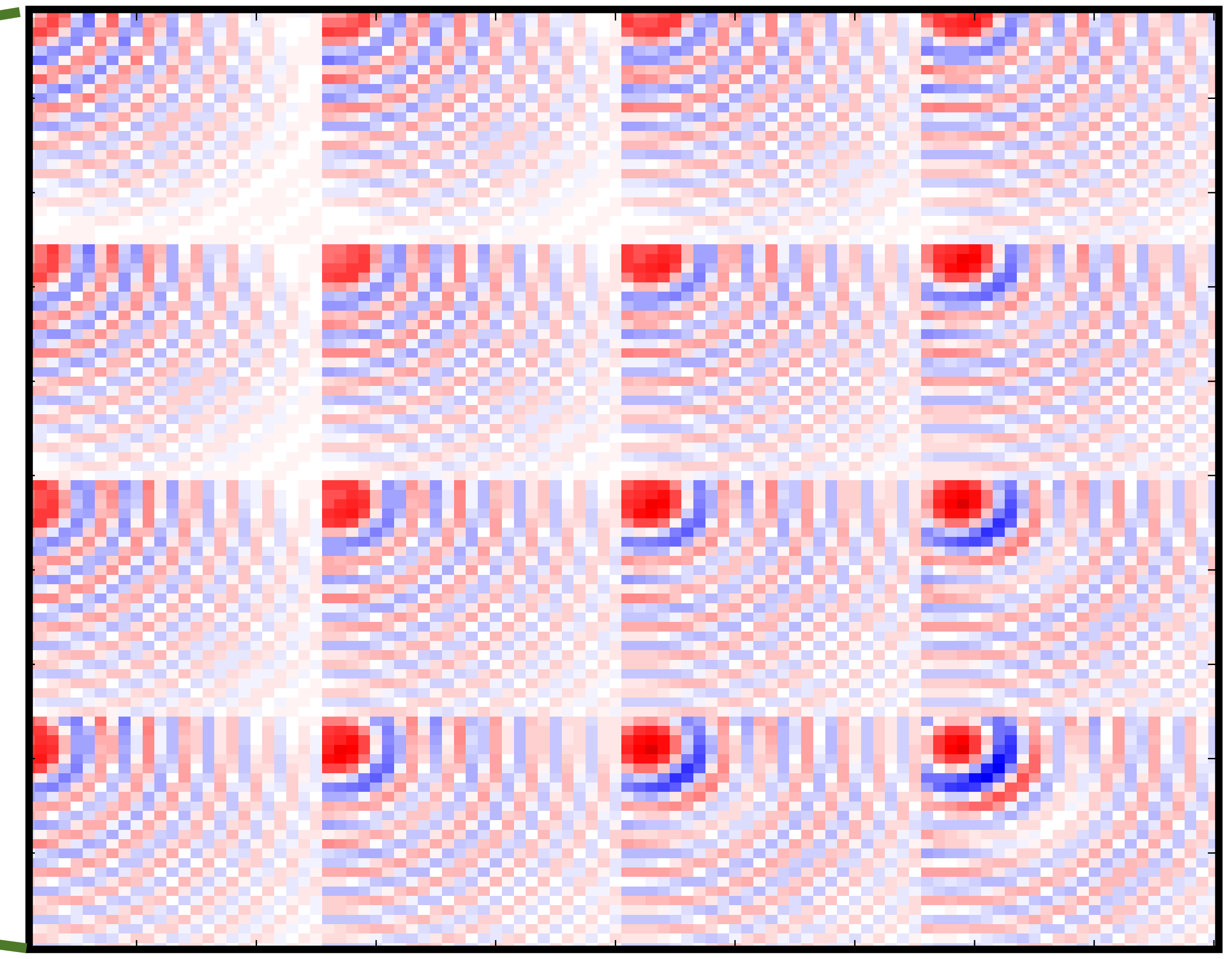


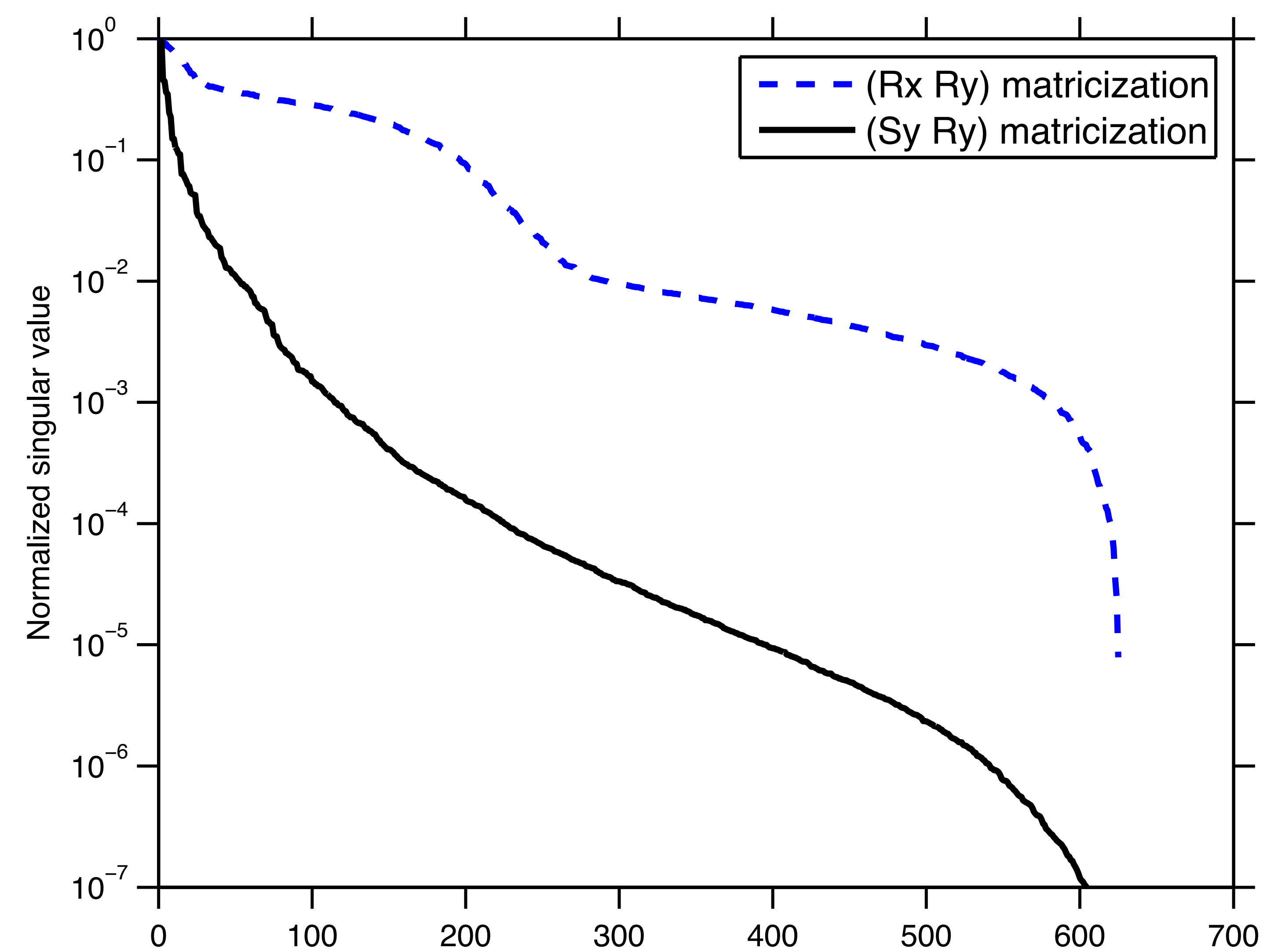




Low-rank structure
conventional 5D data, monochromatic slice, Sx-Rx matricization







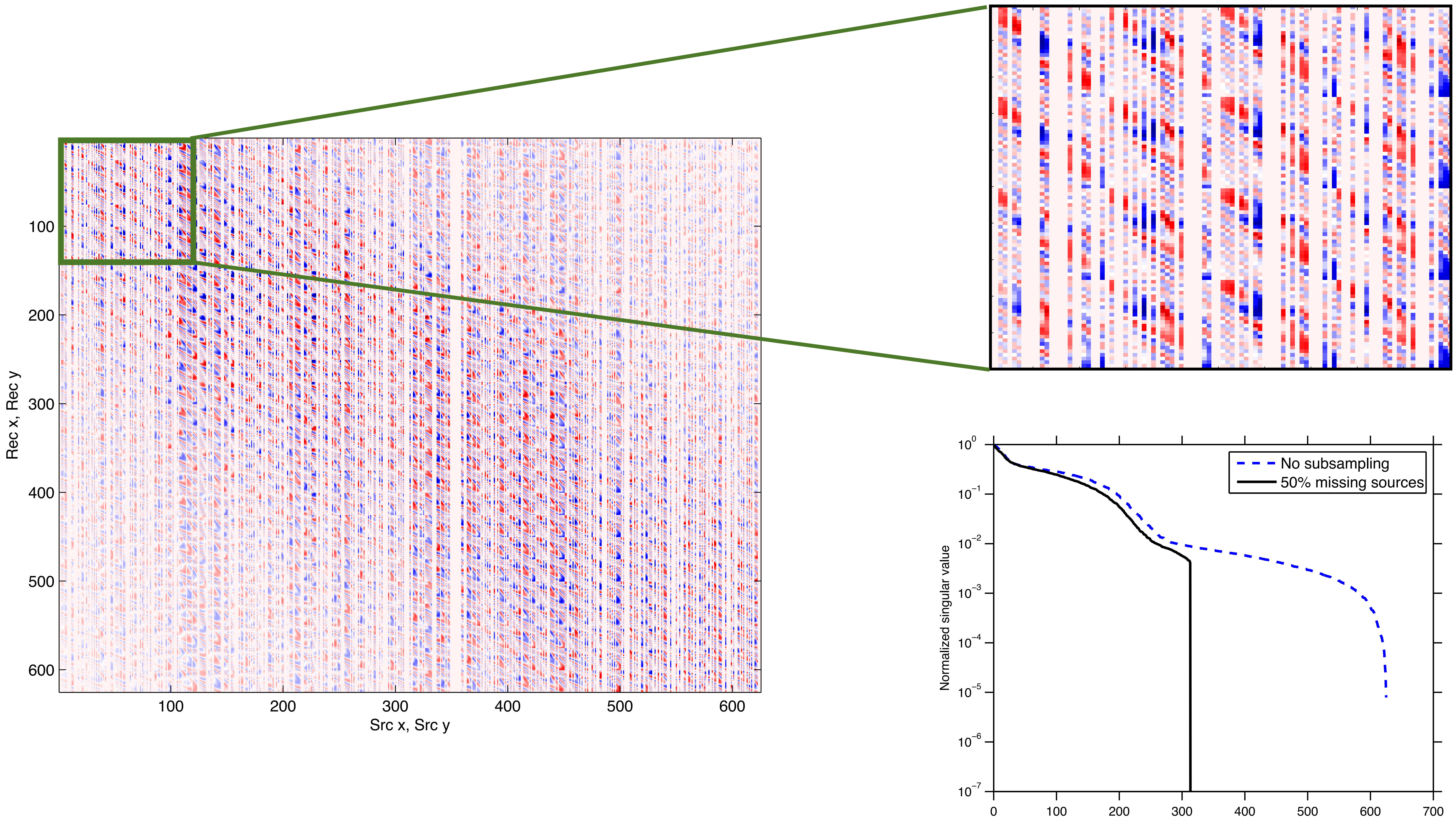
[Candès and Plan, 2009]

Matrix completion

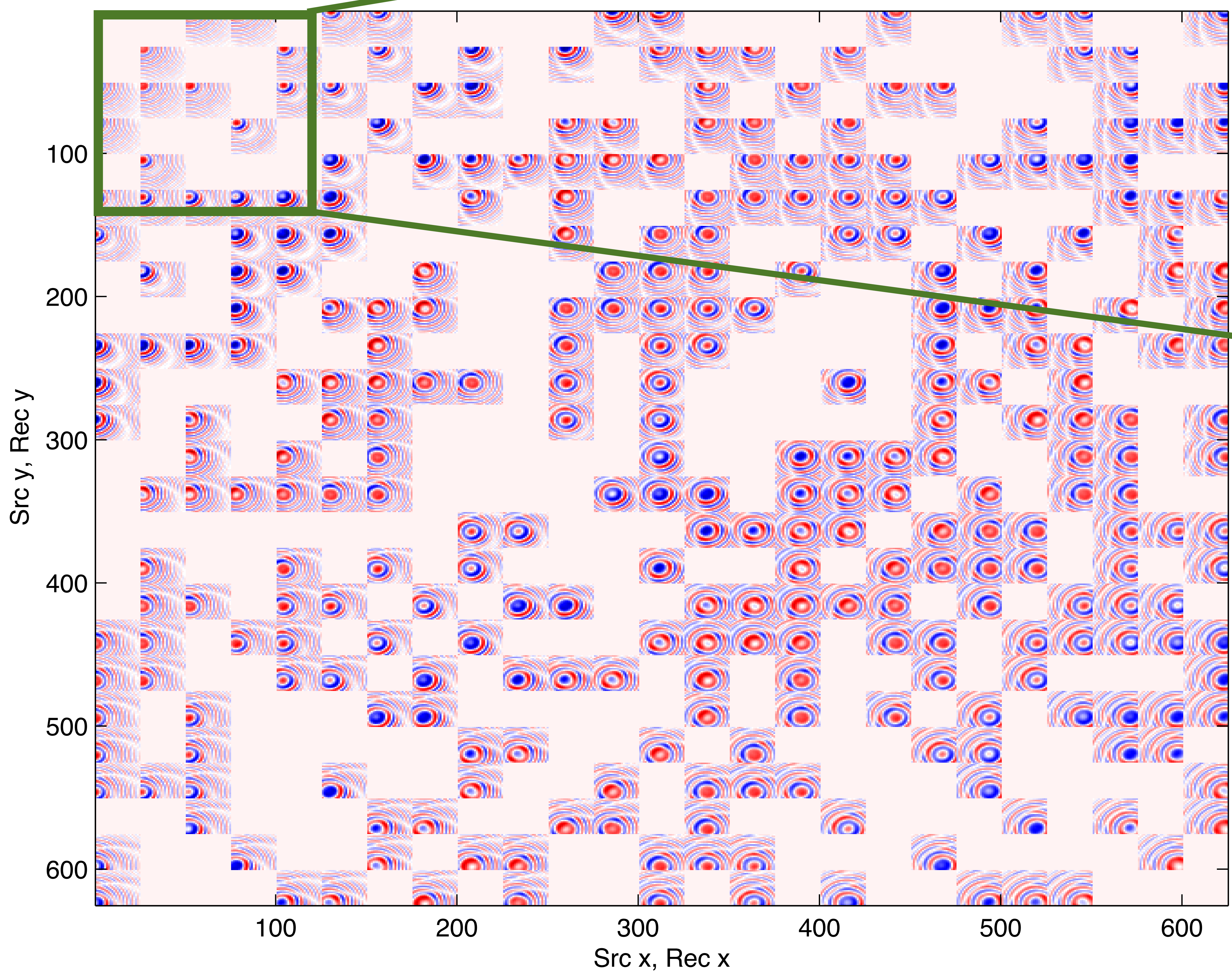
Successful reconstruction scheme

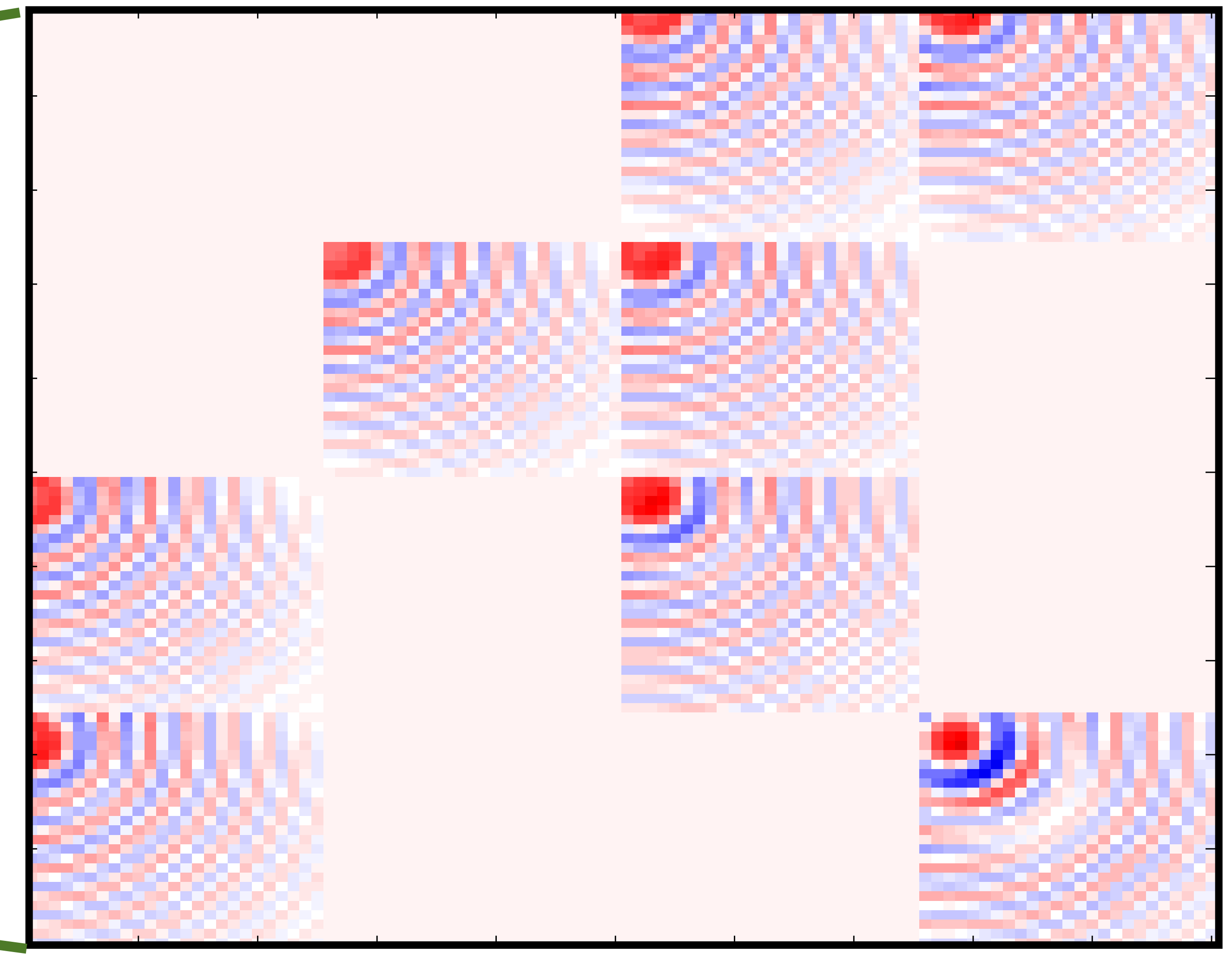
- exploit *structure*
  - *low-rank* / *fast decay* of singular values
- sampling
  - randomness *increases* rank in "transform domain"
- optimization
  - via *rank-minimization (nuclear norm-minimization)*

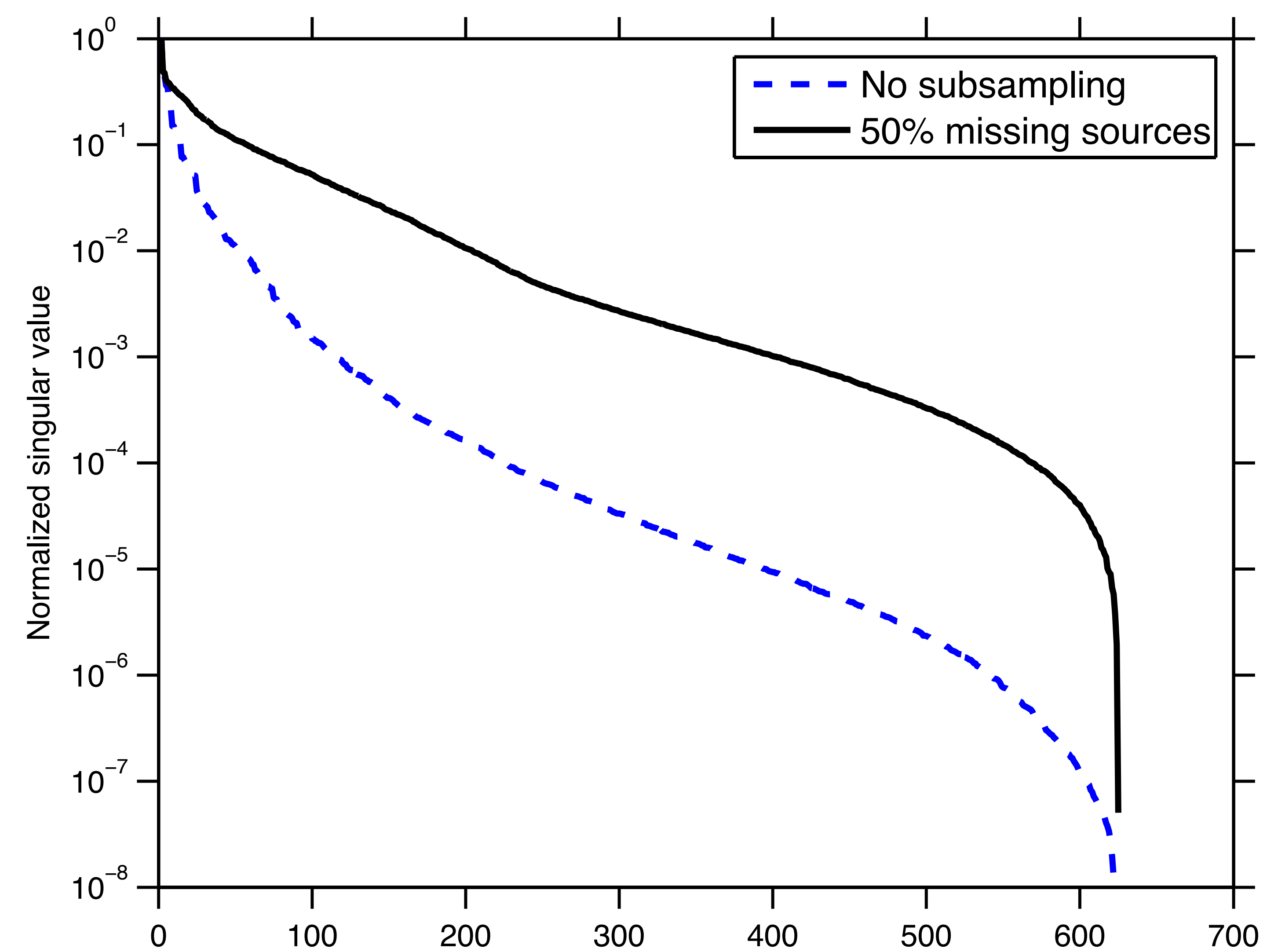
Low-rank structure
time-jittered data, monochromatic slice, Sx-Sy matricization



Low-rank structure
time-jittered data, monochromatic slice, Sx-Rx matricization







[Candès and Plan, 2009]

SLIM

Matrix completion

Successful reconstruction scheme

- exploit *structure*
  - *low-rank / fast decay* of singular values
- sampling
  - randomness *increases* rank in "transform domain"
- optimization
  - via *rank-minimization (nuclear norm-minimization)*

Rank minimization

expensive
(search over all possible values of rank)

$$\min_{\mathbf{X}} \quad \underbrace{\mathrm{rank}(\mathbf{X})}_{} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

number of singular values of $\mathbf{X}$

Rank minimization

expensive
(search over all possible values of rank)

$$\min_{\mathbf{X}} \quad \underbrace{\mathrm{rank}(\mathbf{X})}_{} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

number of singular values of $\mathbf{X}$

Nuclear-norm minimization

convex relaxation of rank-minimization

[Recht et. al., 2010]

$$\min_{\mathbf{X}} \quad \underbrace{||\mathbf{X}||_*}_{} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

sum of singular values of $\mathbf{X}$

Matrix-Completion framework

- Restriction operator is constant across frequencies

- Perform matrix-completion across frequencies in parallel

5D Jittered marine acquisition

- Restriction operator is non-separable
  - combination of time-shifting and shot-jittered operator
- Can't perform matrix-completion over independent frequencies
  - reformulate nuclear-norm minimization over temporal-frequency domain

SLIM

Rank-minimization problem

- Let $\mathbf{X} \in \mathbb{C}^{\mathbf{n_f} \times \mathbf{n_{rx}} \times \mathbf{n_{sx}} \times \mathbf{n_{ry}} \times \mathbf{n_{sy}}}$ be the conventional 5D seismic data volume represented as a tensor.

- Given a set of measurements $\mathbf{b}$, aim is to solve

$$\min_{\mathbf{X}_f} \sum_f \|\mathbf{X}_f\|_* \quad \text{s.t.} \quad ||\mathcal{A}(\mathbf{X}_f) - \mathbf{b}||_2^2 \leq \sigma$$

where

$$\|\mathbf{X}_f\|_* = \sum_{i=1}^{m} \lambda_i = \|\lambda\|_1$$

Sampling-measurement operator

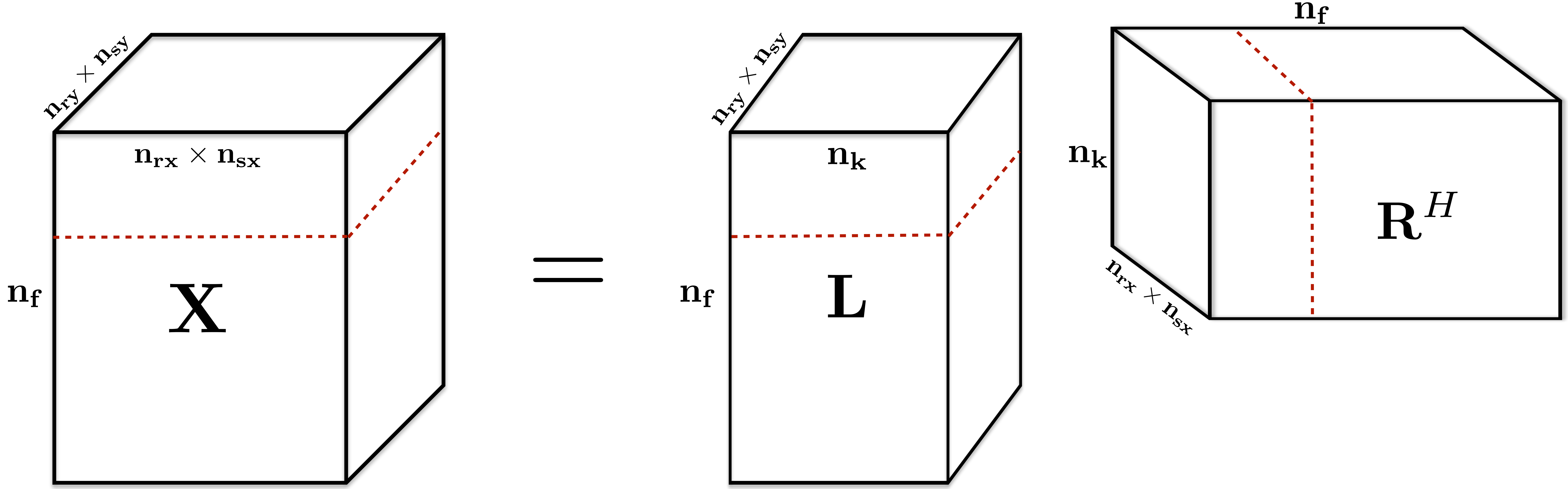
- $\mathcal{A}$ is the transform-sampling operator defined as

$$\mathcal{A}(.) = \mathbf{M}\mathbf{F}^H\mathcal{S}^H(.)$$

$\mathbf{M}$	time-jittered operator
$\mathbf{F}^H$	inverse Fourier transform along frequency axis
$\mathcal{S}^H$	rank-revealing transform domain

Factorized formulation

$$\mathbf{X} = \mathbf{L}\mathbf{R}^H$$



$\mathbf{X} \in \mathbb{C}^{\mathbf{n_f \times n_{rx} \times n_{sx} \times n_{ry} \times n_{sy}}}$

$\mathbf{L} \in \mathbb{C}^{\mathbf{n_f \times n_{rx} \times n_{sx} \times n_k}}$

$\mathbf{R} \in \mathbb{C}^{\mathbf{n_f \times n_{ry} \times n_{sy} \times n_k}}$

Factorized formulation

- Costly SVD's

- Nuclear norm satisfies

$$\sum_{j}^{n_f} \|\mathbf{D_j^{(i)}}\|_* \leq \sum_{\mathbf{j}}^{\mathbf{n_f}} \frac{\mathbf{1}}{\mathbf{2}} \|\mathbf{L_j^{(i)}} \mathbf{R_j^{(i)}}\|_{\mathbf{F}}^{\mathbf{2}}$$ [Rennie and Srebro 2005]

where $\|\cdot\|_F^2$ is sum of squares of all entries

- Choose rank $\mathbf{k}$ *explicitly* & avoid costly *SVD's*

How to choose the rank parameter?

Typical abridged result from low-rank matrix recovery theory:

If $\mathcal{A} : \mathbb{C}^{n \times m} \mapsto \mathbb{C}^k$ is a random linear operator (e.g., Ω chosen randomly, subgaussian), then we can recover a rank-r matrix via nuclear norm minimization if

$$k \geq Cr \max(n, m) \log(\max(n, m))$$ [Candes and tao 2009]

with high probability.

SLIM

How to choose the rank parameter?

$$k \geq Cr\max(n,m)\log(\max(n,m))$$

In our case: $k = .25 \cdot nm$, where 0.25 is subsampling ratio,

$$n = m = 4141$$

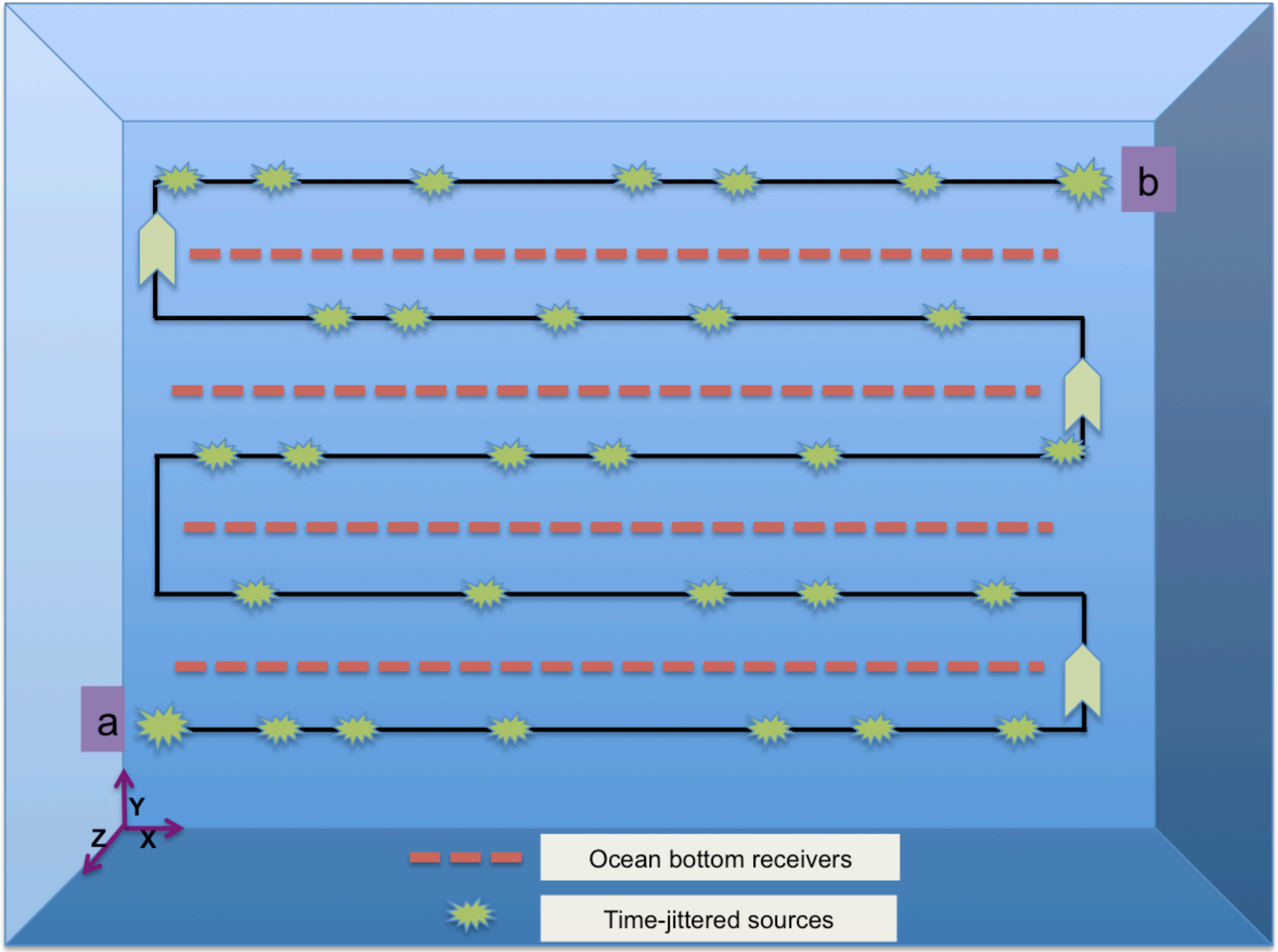
(with $C = 1$ and rounding) $\Longrightarrow r \leq 100$

Choose upper bound as rank.

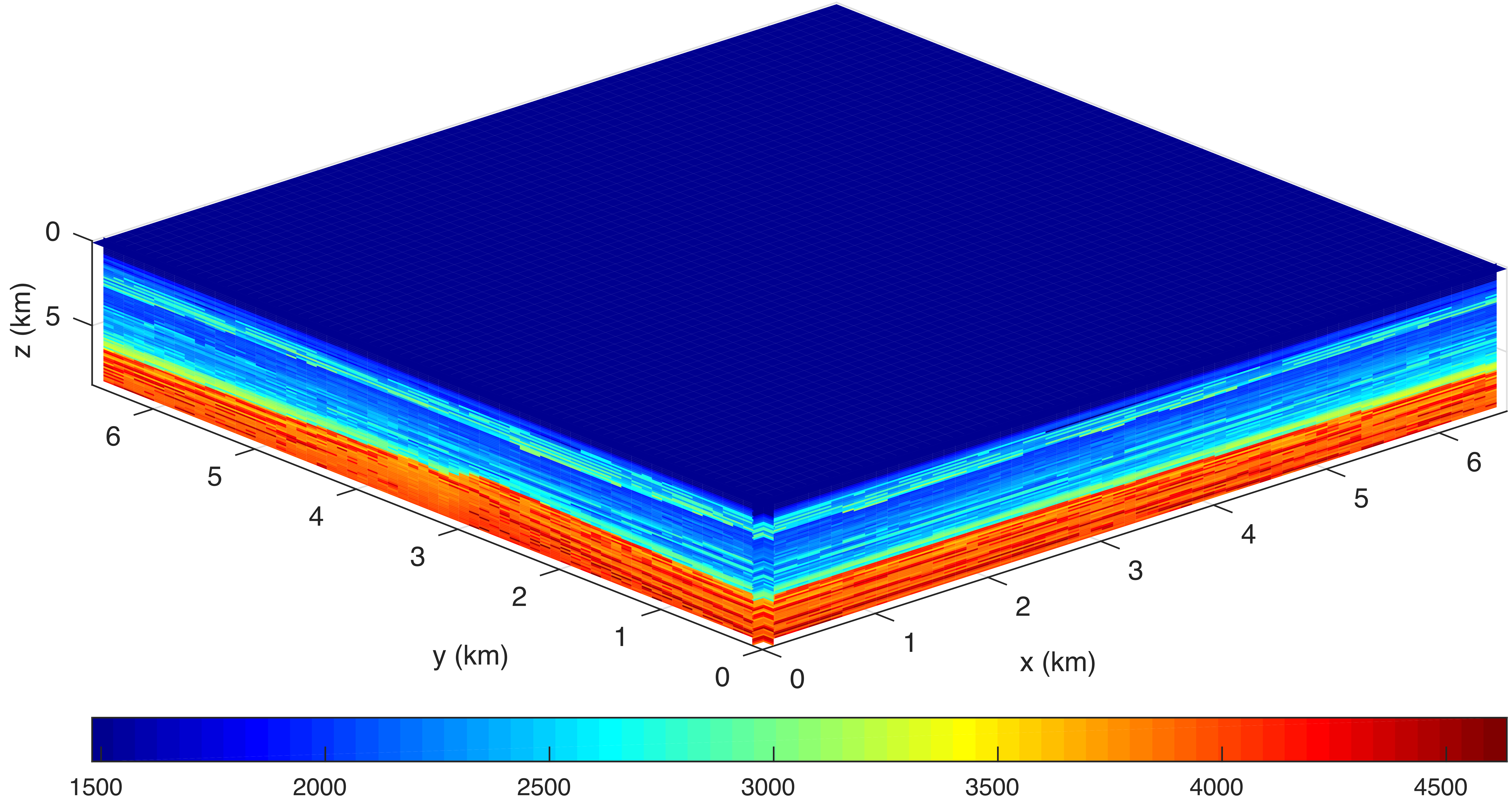
Experimental results

Acquisition setup

SLIM



3D BG Compass model



SLIM

Acquisition information

- 10s temporal length

- 25 m flip-flop shooting
  - source-sampling ranges from 25 m to 175 m
  - effective 50 m source sampling for each airgun array
  - acquired 400 sources

- 10201 receivers

- Ricker wavelet with central frequency of 20 Hz

- size of the recovered 5D seismic data volume is 0.3 TB

Optimization information

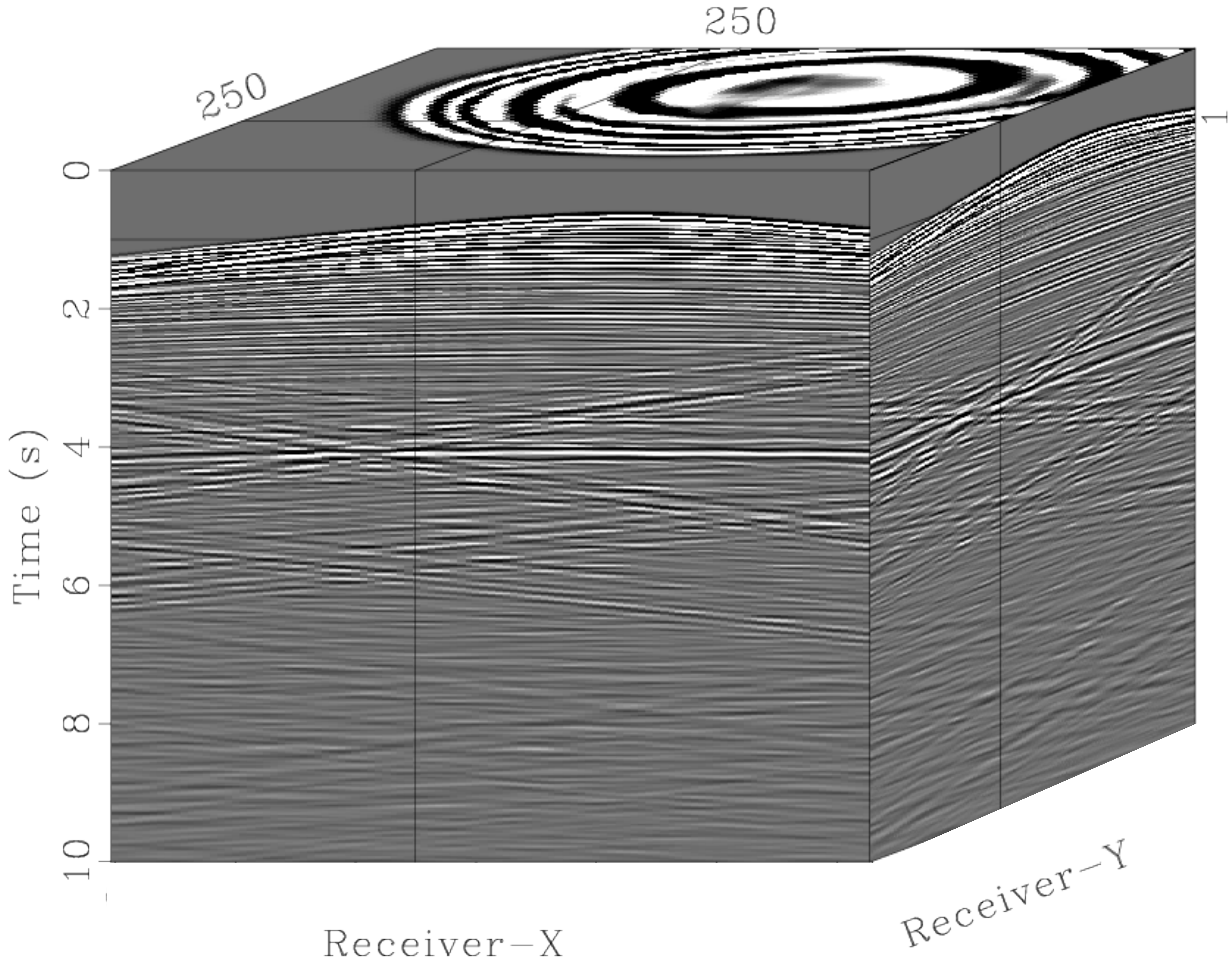
- Parallelized factorization framework over sources and receivers
- 200 iterations, computational time 42 hours
- fixed 100 rank values across frequencies
- Separation + interpolation @ 6.25 m grid
  - recovered 1600 sources

Computational Environment

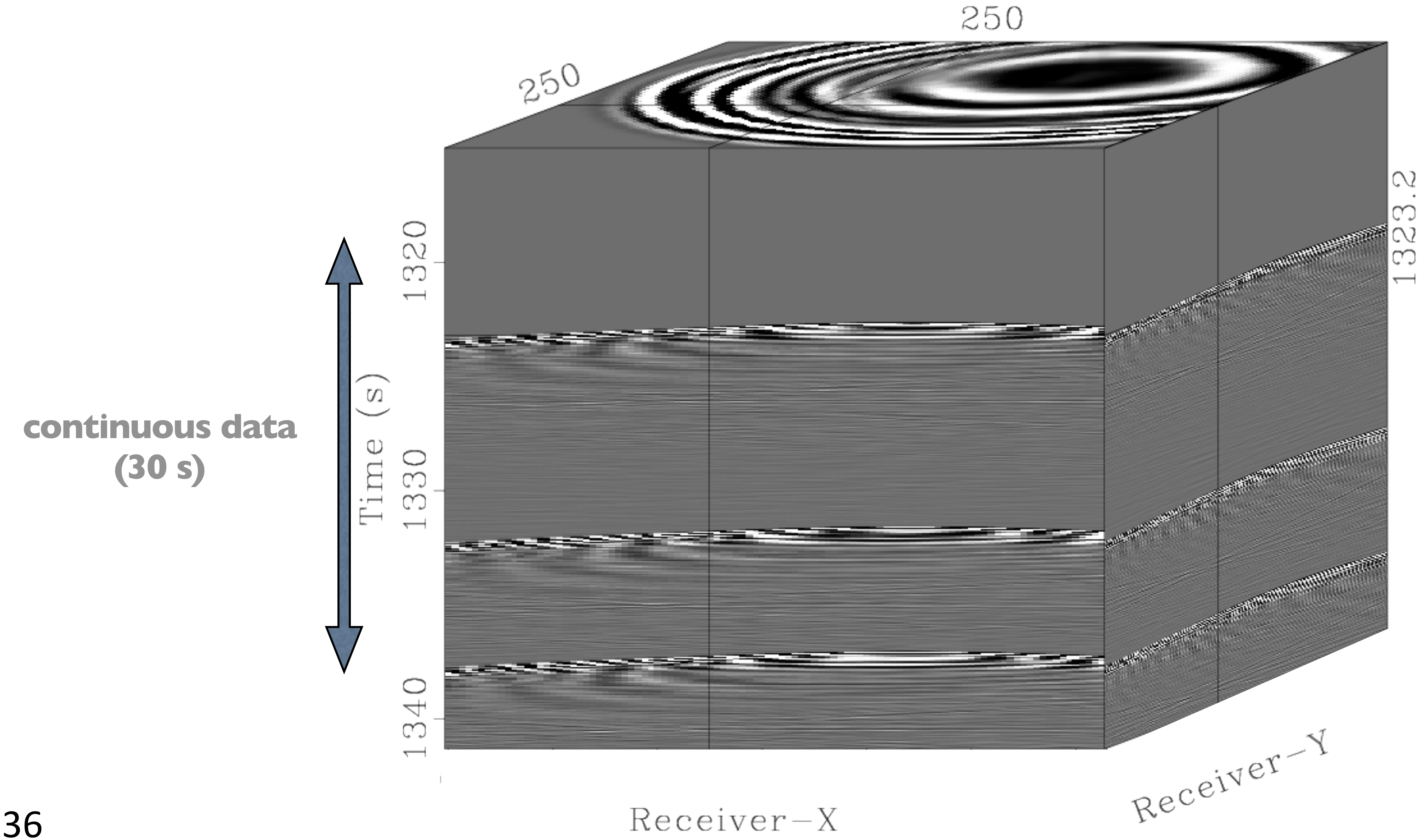
SENAI Yemoja cluster

- 30 nodes, 128 GB RAM each, 20-core processors
- 300 Parallel Matlab workers (10 per node), multithread - full core utilization

Conventional data
common-shot gather, @6.25 m source sampling

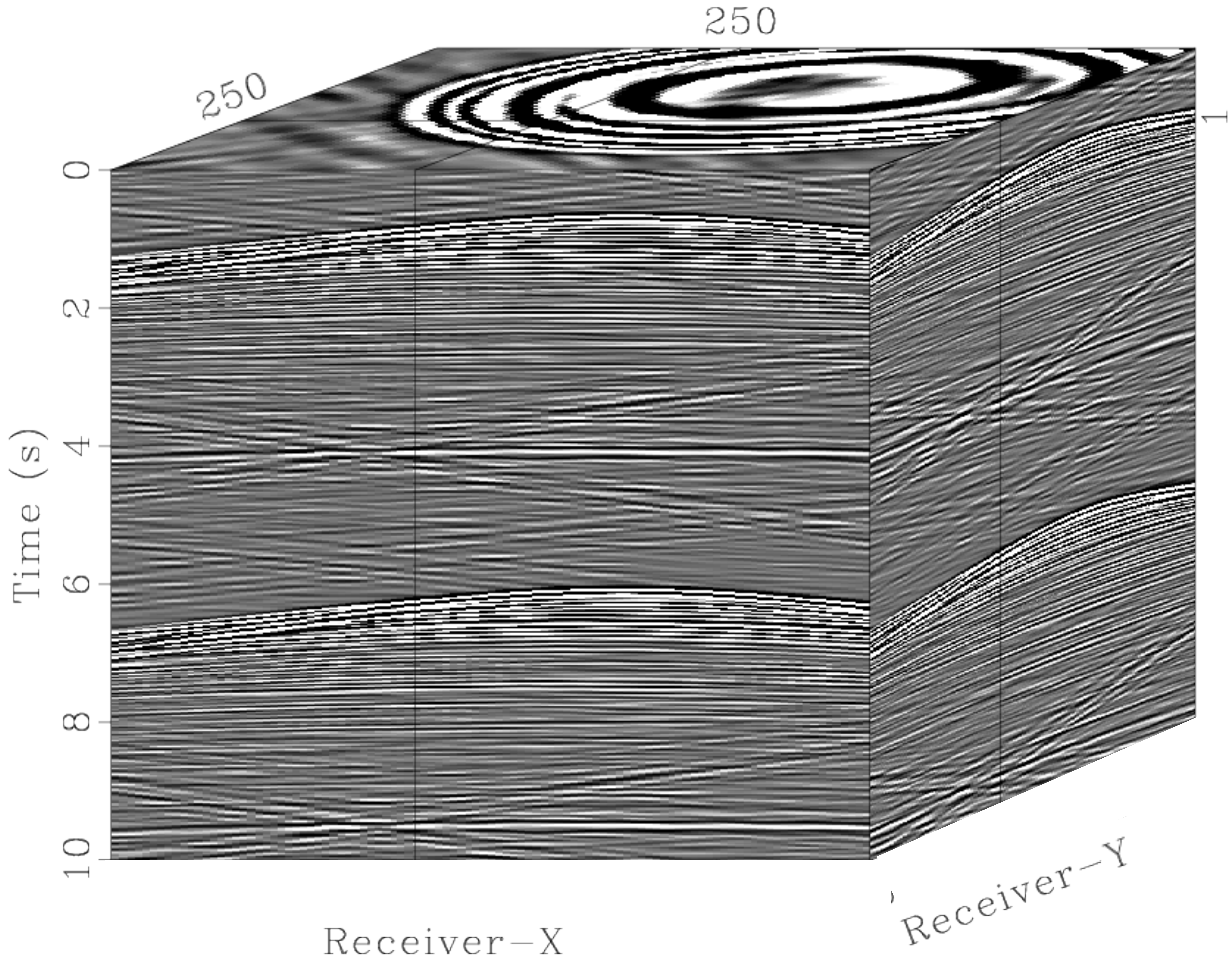


Time-jittered continuous record
@ 25m flip-flop shooting, blended & missing shots



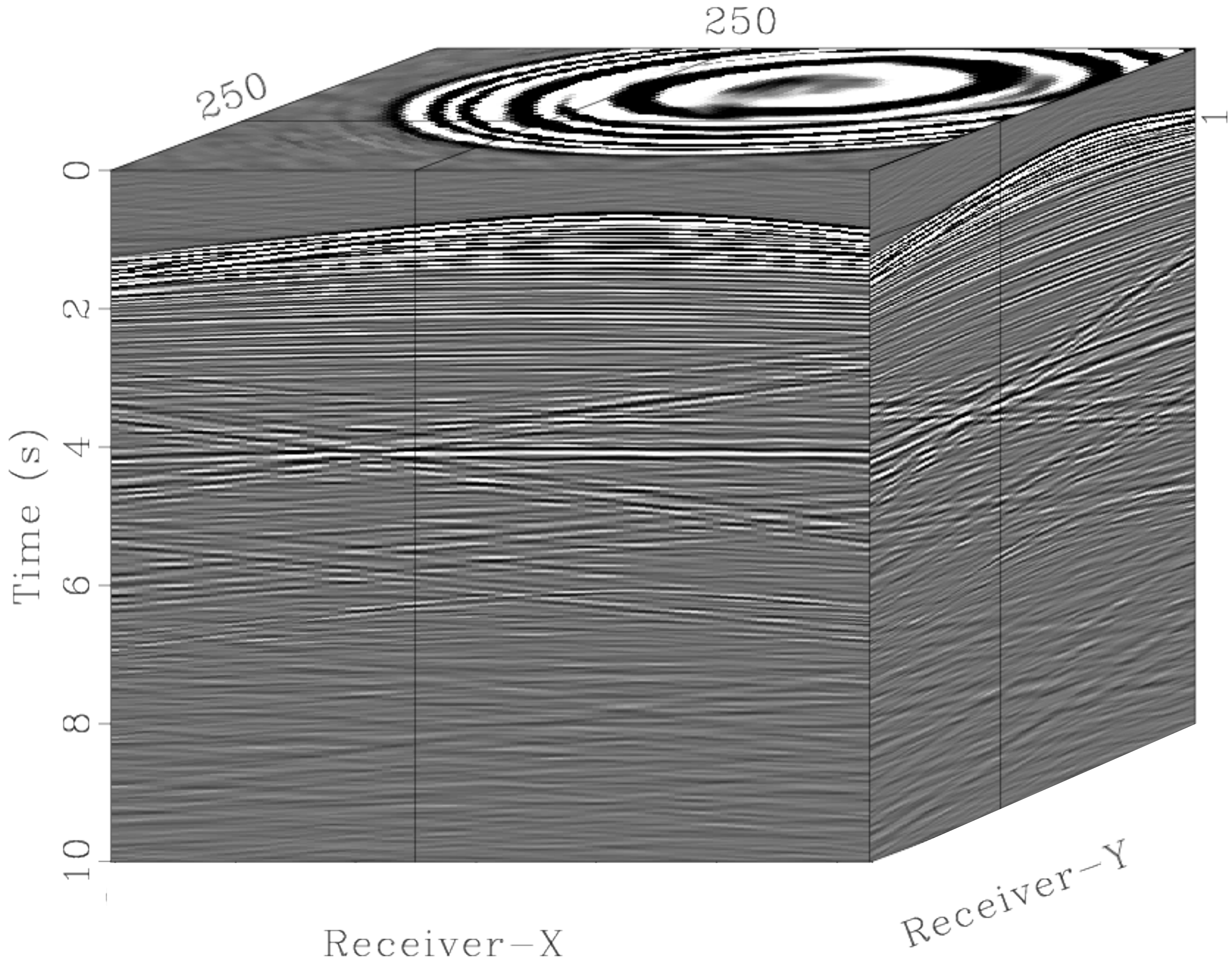
Adjoint of sampling-operator

common-shot gather

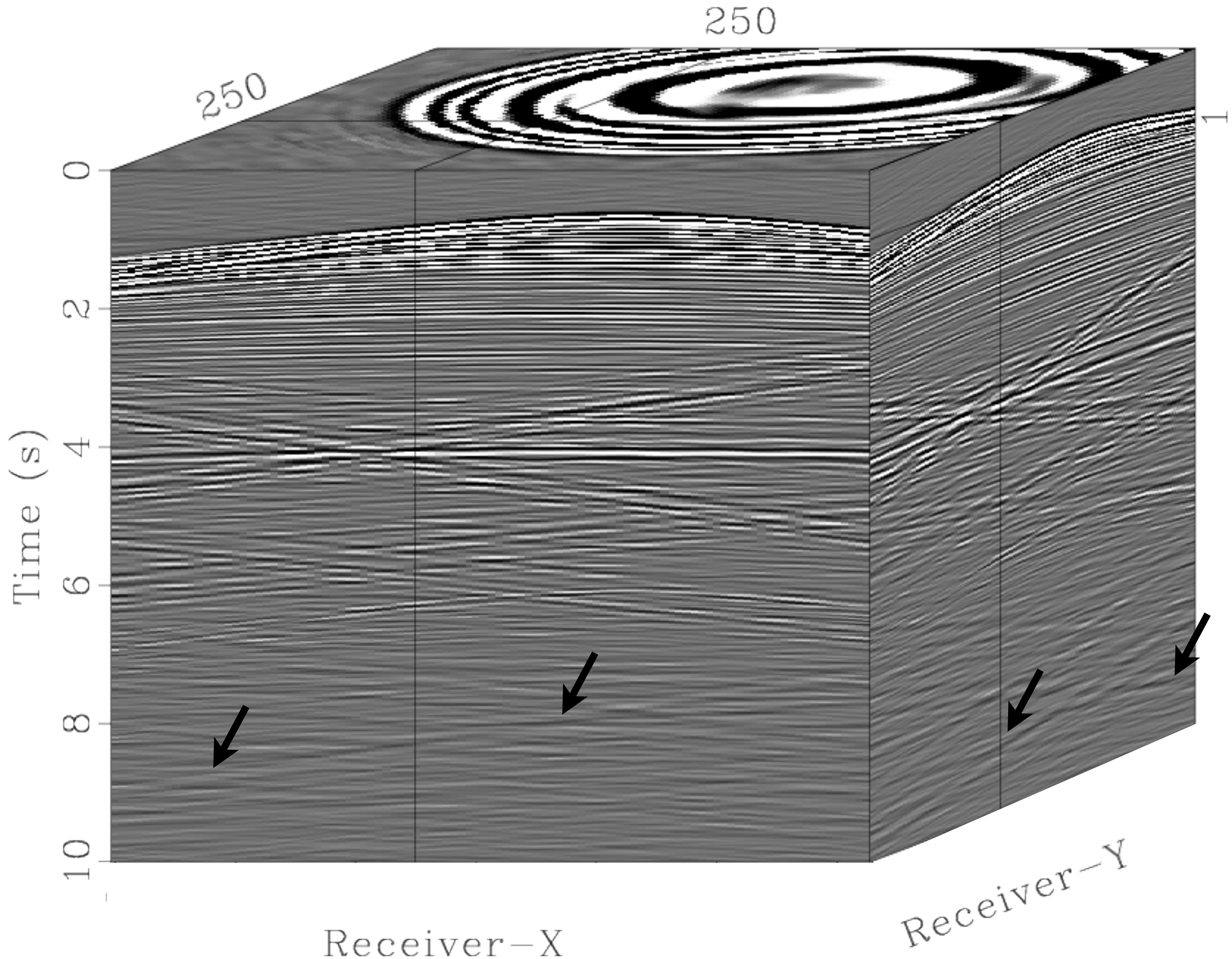


After Source-Separation

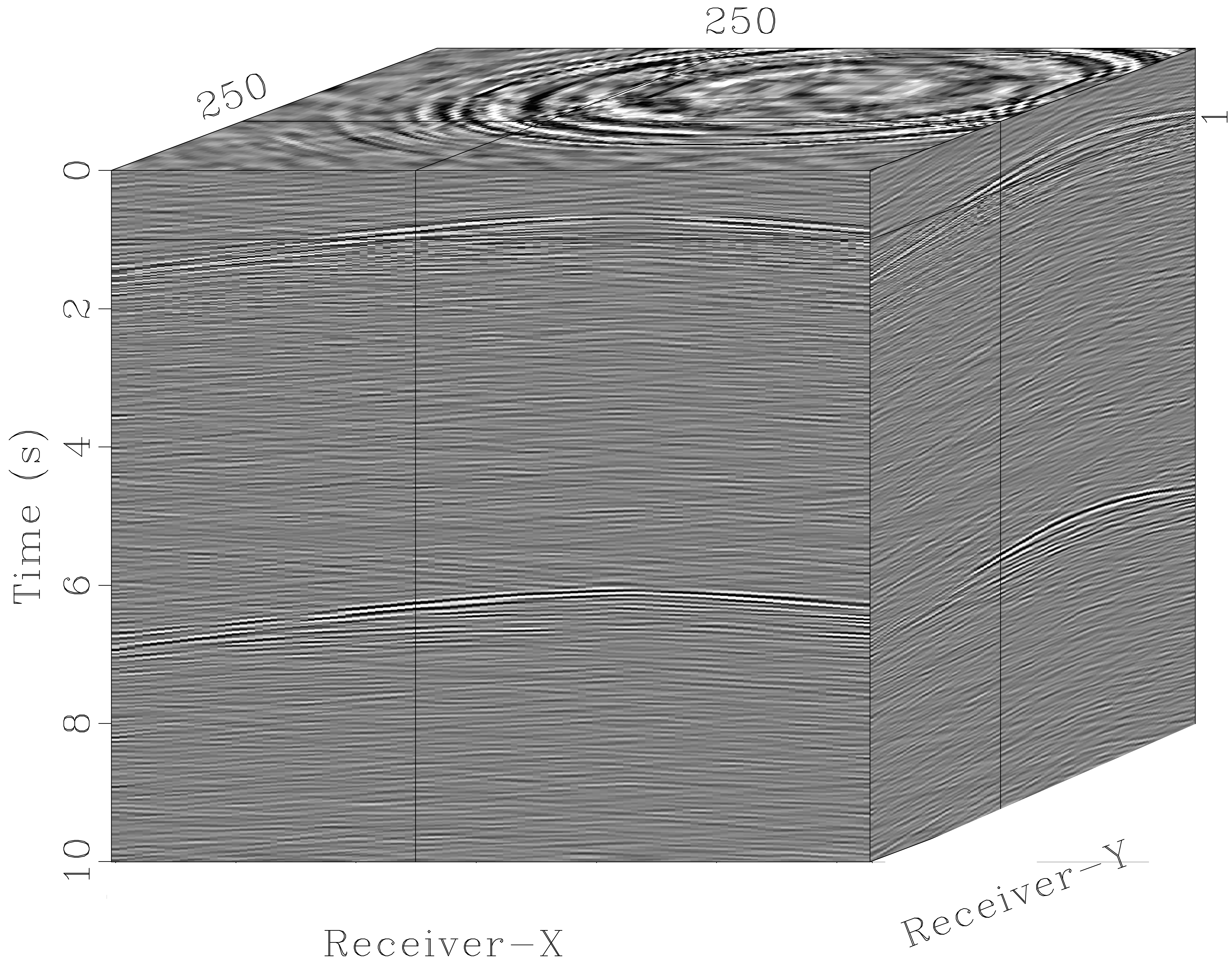
common-shot gather, 21dB signal-to-noise ratio



After Source-Separation
preserved late-arrivals energy

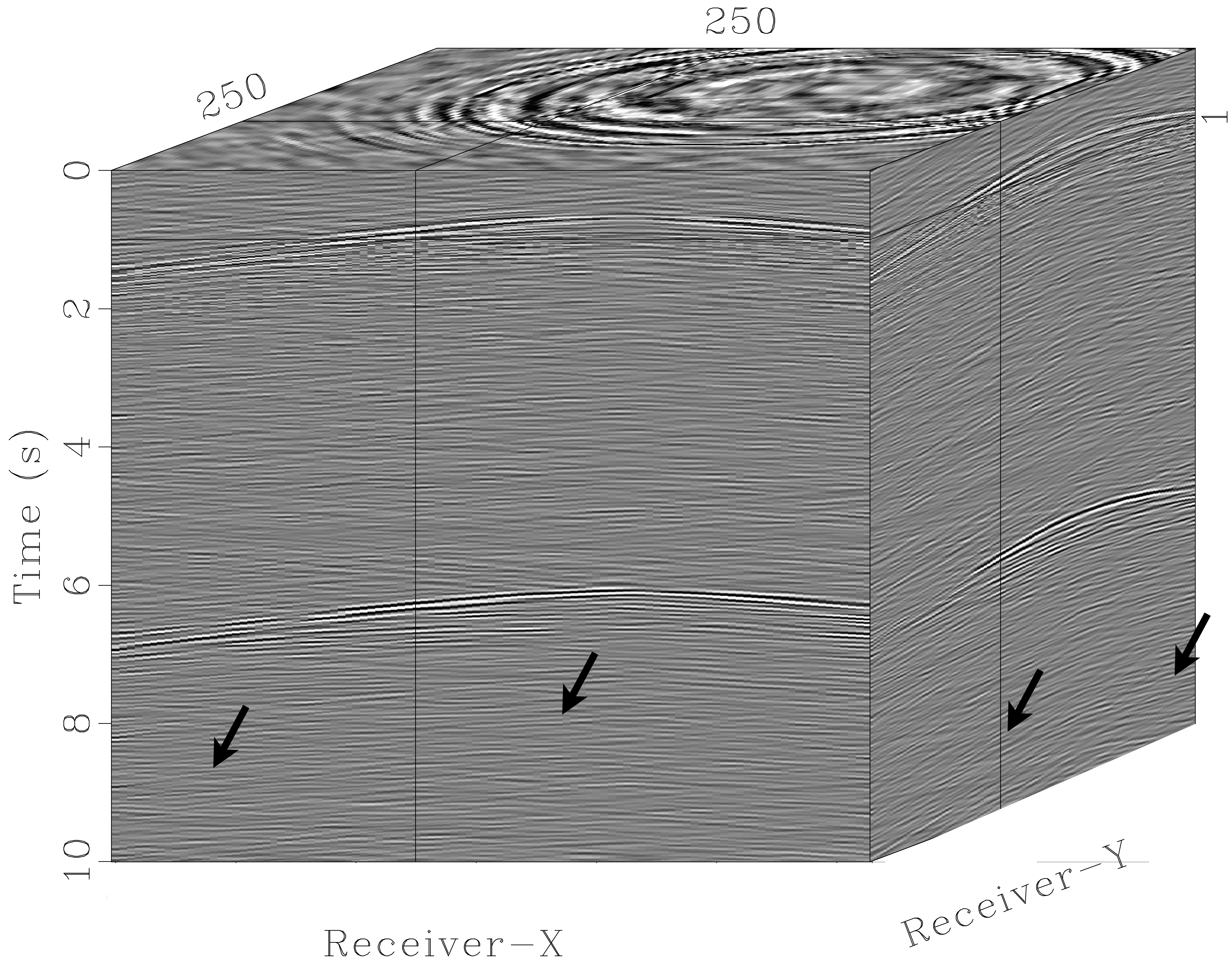


Residual



10x magnified

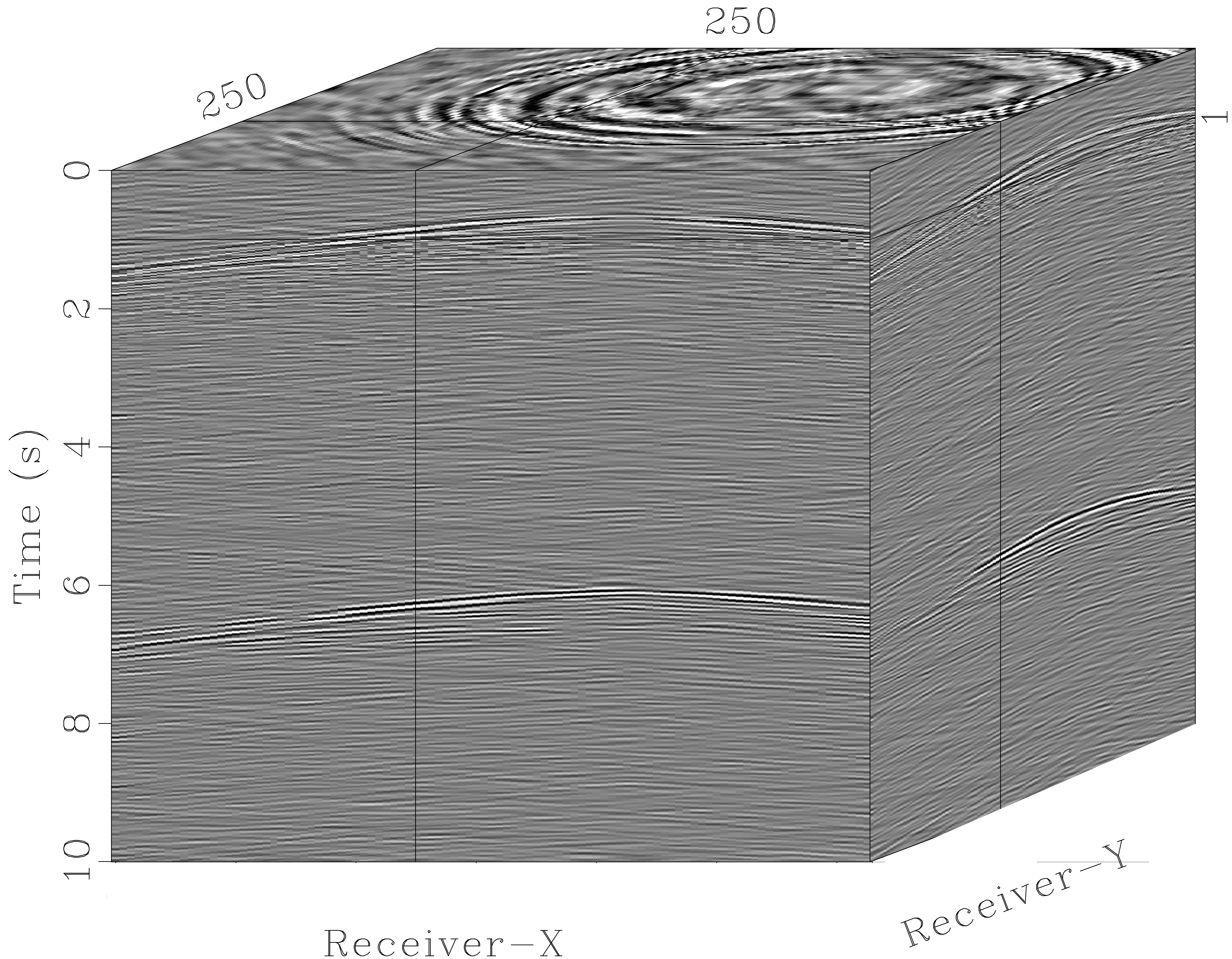
Residual



10x magnified

reconstructed late-arrivals

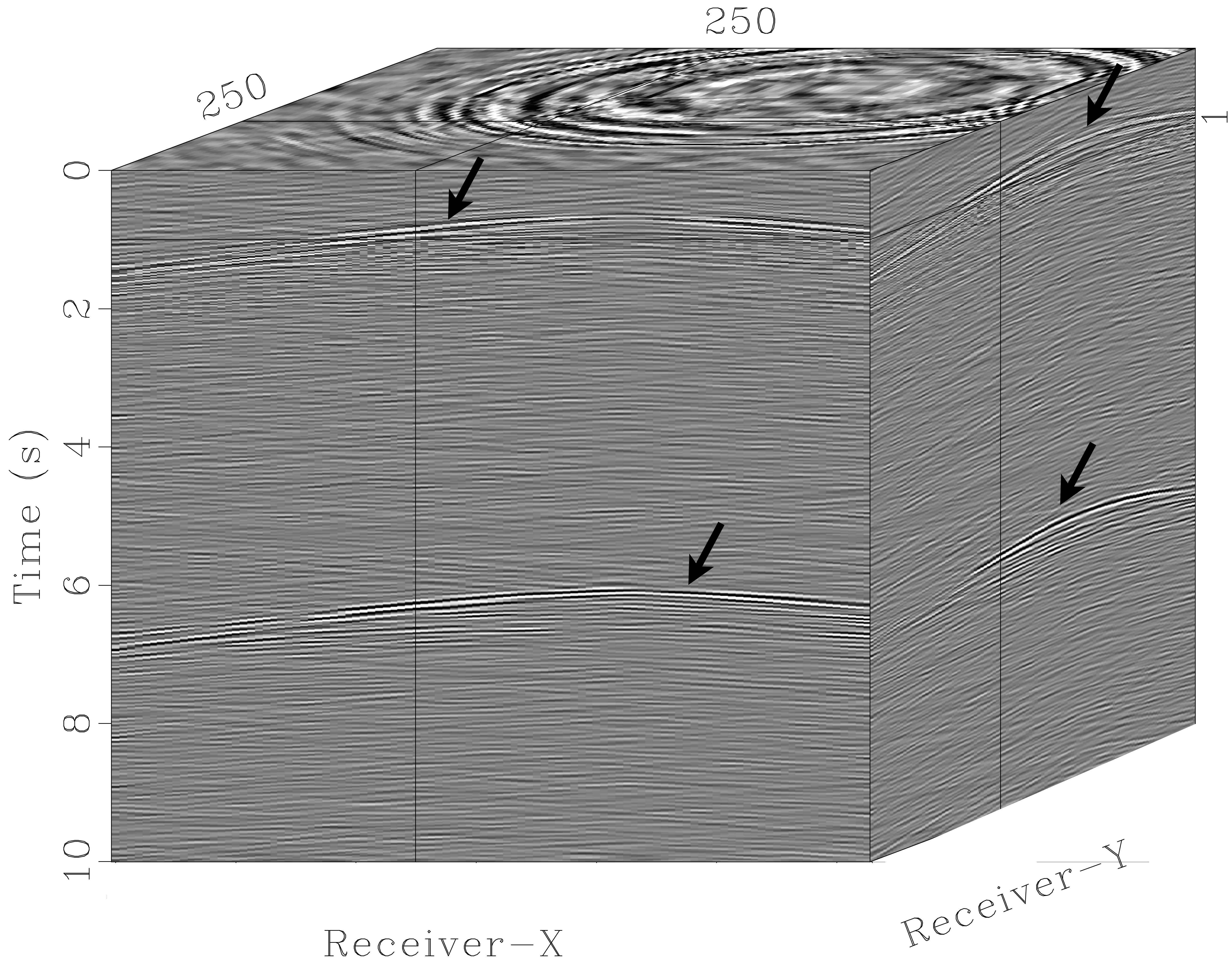
Residual



10x magnified

Residual

coherent energy can be reconstructed using 2nd pass over data



10x magnified

Take-away message

- **4X** up-sampling (@ 6.25m) & saving in acquisition time
- size of final recovered data volume is **0.3 TB**
  - no need to save fully sampled seismic data volume
- save $\mathbf{L}$ and $\mathbf{R}$ factors
  - compression rate is **98%**
  - size of final compressed 5D seismic volume is **~7 GB**

Seismic data processing—interpolation via rank-minimization

Rajiv Kumar, Oscar Lopez and Felix J. Herrmann

SLIM
University of British Columbia

Thursday, October 27, 2016

Motivation

- infill the missing/coarser source-receiver acquisition grid
- minimize the acquisition related artifacts during waveform-inversion

[Candès and Plan, 2009]

SLIM

Matrix completion

Successful reconstruction scheme

- exploit *structure*
  - *low-rank* / *fast decay* of singular values
- sampling
  - randomness *increases* rank in "transform domain"
- optimization
  - via *rank-minimization (nuclear norm-minimization)*

Nuclear-norm minimization

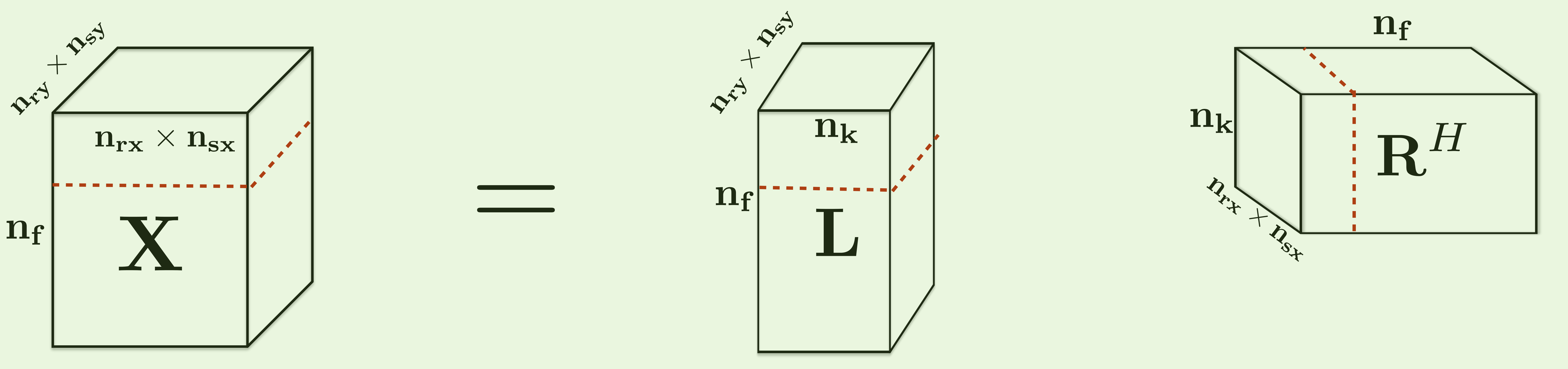
convex relaxation of rank-minimization

[Recht et. al., 2010]

$$\min_{\mathbf{X}} \quad \underbrace{||\mathbf{X}||_*}_{\text{sum of singular values of } \mathbf{X}} \quad \text{s.t.} \quad \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2 \leq \epsilon$$

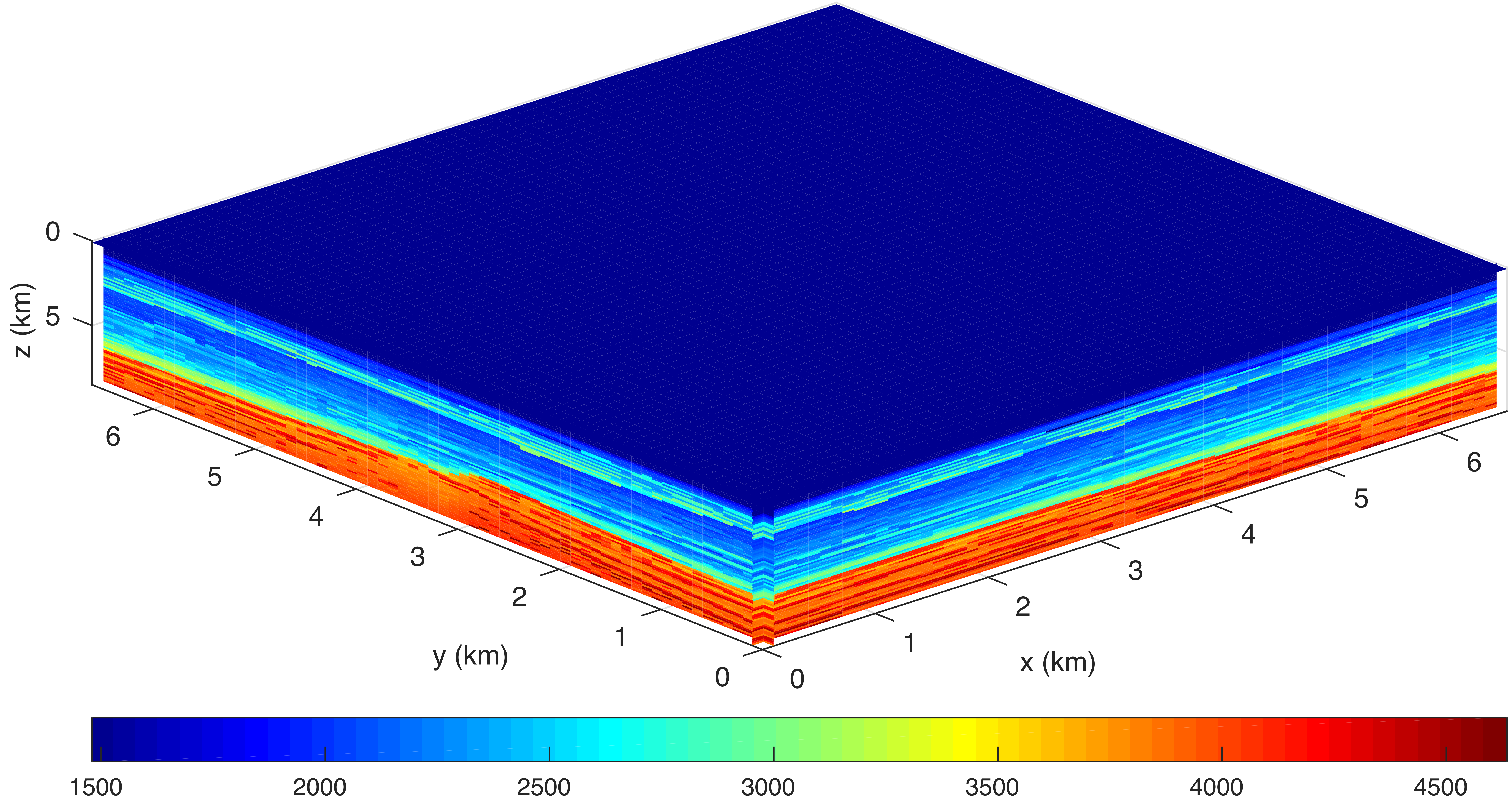
[Rennie and Srebro 2005, Lee et. al. 2010, Recht and Re 2011]

$$\mathbf{X} = \mathbf{L}\mathbf{R}^H$$

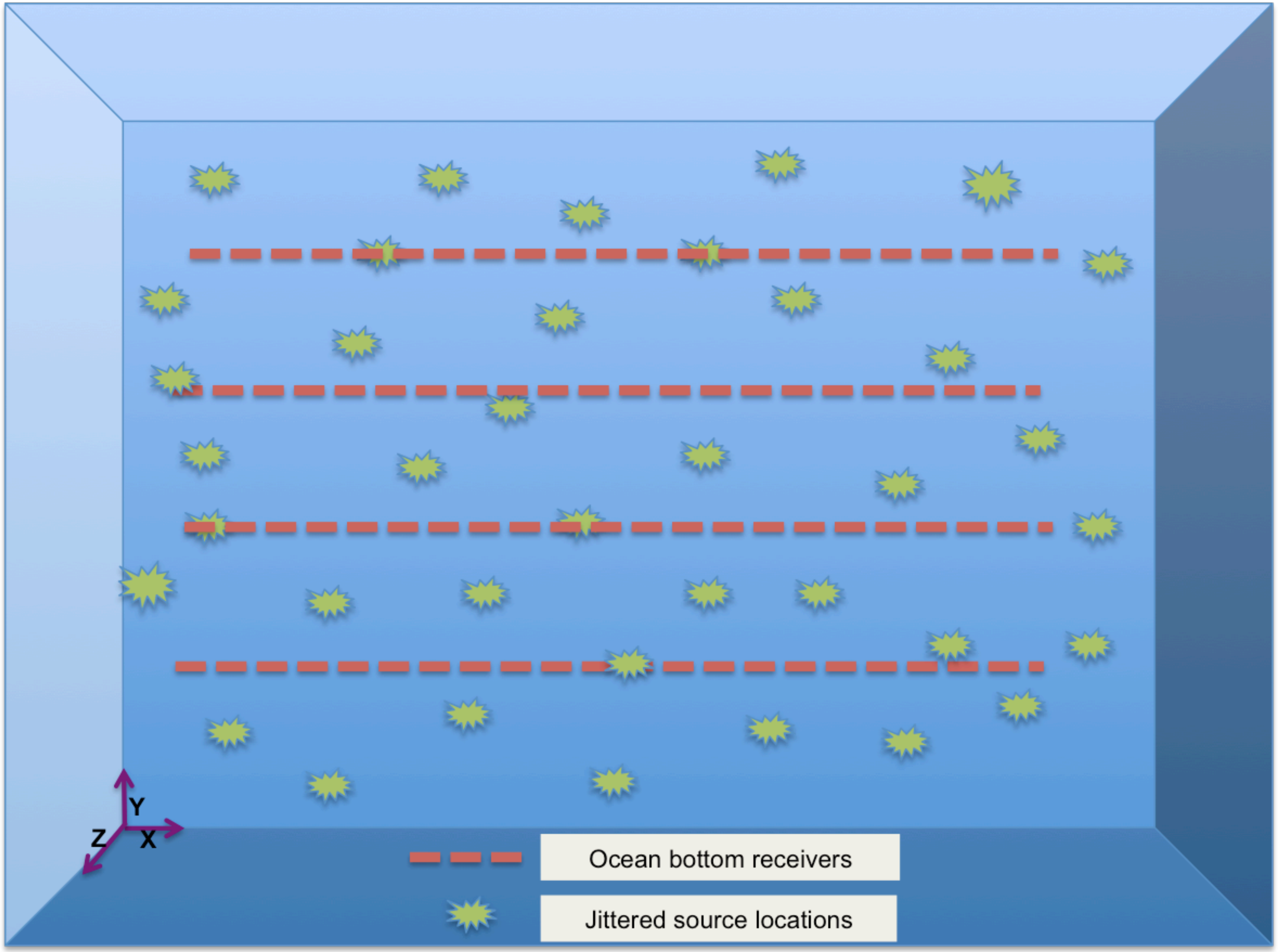


$\mathbf{X} \in \mathbb{C}^{\mathbf{n_f} \times \mathbf{n_{rx}} \times \mathbf{n_{sx}} \times \mathbf{n_{ry}} \times \mathbf{n_{sy}}}$ $\quad$ $\mathbf{L} \in \mathbb{C}^{\mathbf{n_f} \times \mathbf{n_{rx}} \times \mathbf{n_{sx}} \times \mathbf{n_k}}$ $\quad$ $\mathbf{R} \in \mathbb{C}^{\mathbf{n_f} \times \mathbf{n_{ry}} \times \mathbf{n_{sy}} \times \mathbf{n_k}}$

3D BG Compass model



Acquisition setup



Acquisition information

- 4s temporal length
- source-sampling ranges from 25 m to 175 m
  - effective 25 m source sampling
  - acquired 320 sources, 80% missing scenario
- 10201 receivers
- Ricker wavelet with central frequency of 20 Hz

Optimization information

- Parallelization over frequencies
- 400 iterations (SPG-LR framework)
- fixed 100 rank value across frequencies
- Interpolation @ 25 m grid
  - recovered 1600 sources

Computational Environment

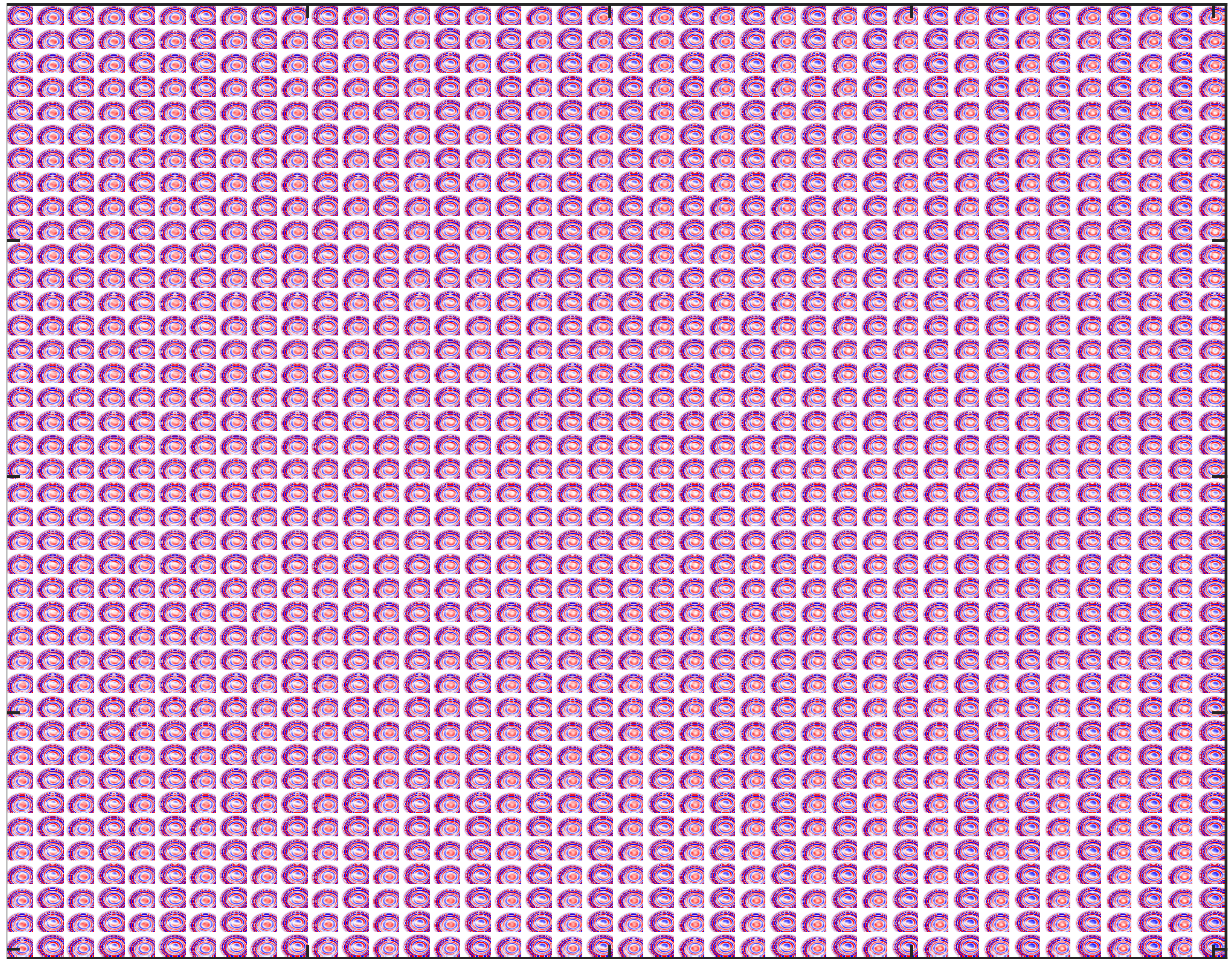
SENAI Yemoja cluster

- 10 nodes, 128 GB RAM each, 20-core processors

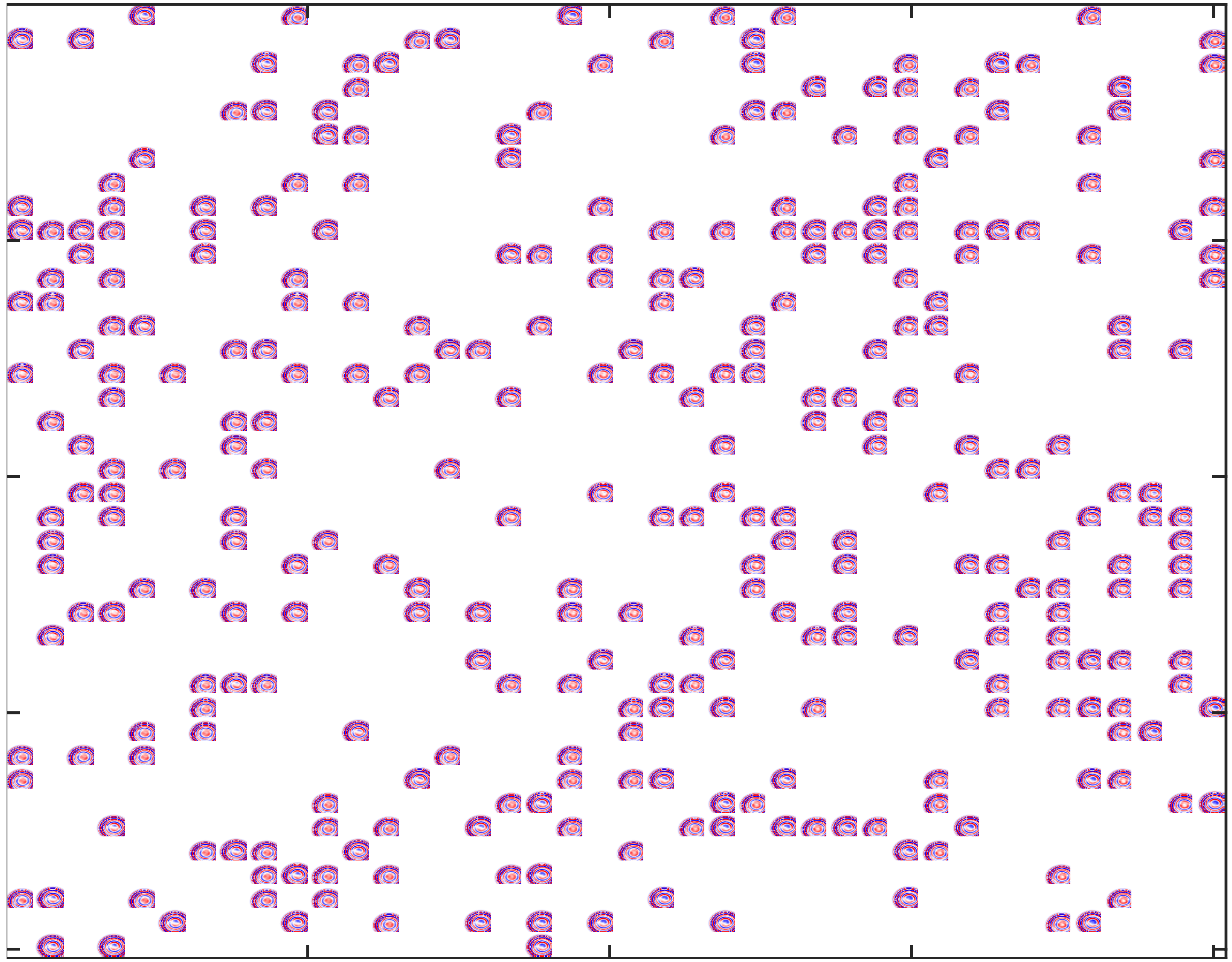
- 100 Parallel Matlab workers (10 per node), multithread - full core utilization

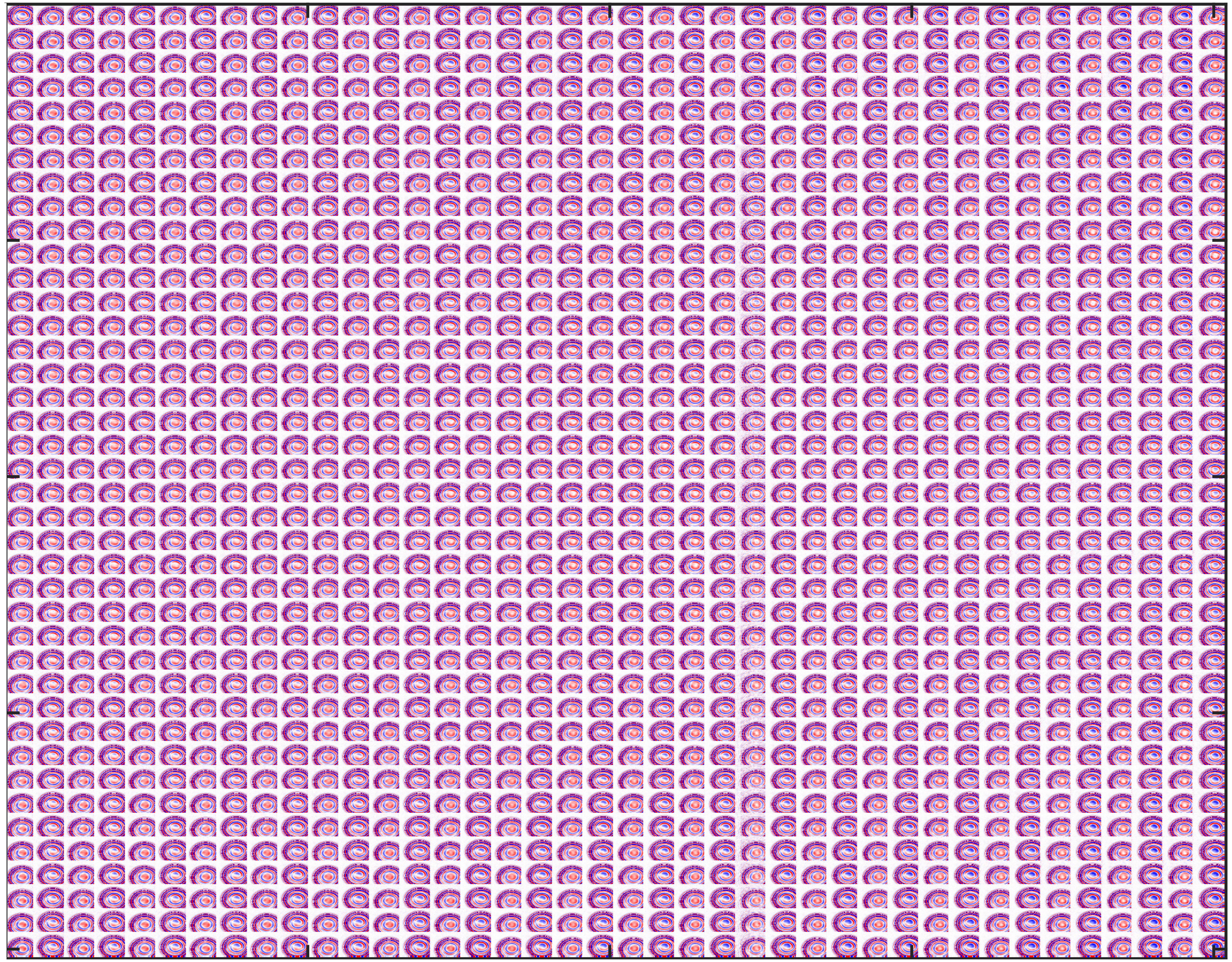
monochromatic slice
(4000 x 4000)

ground truth



5-times subsampling

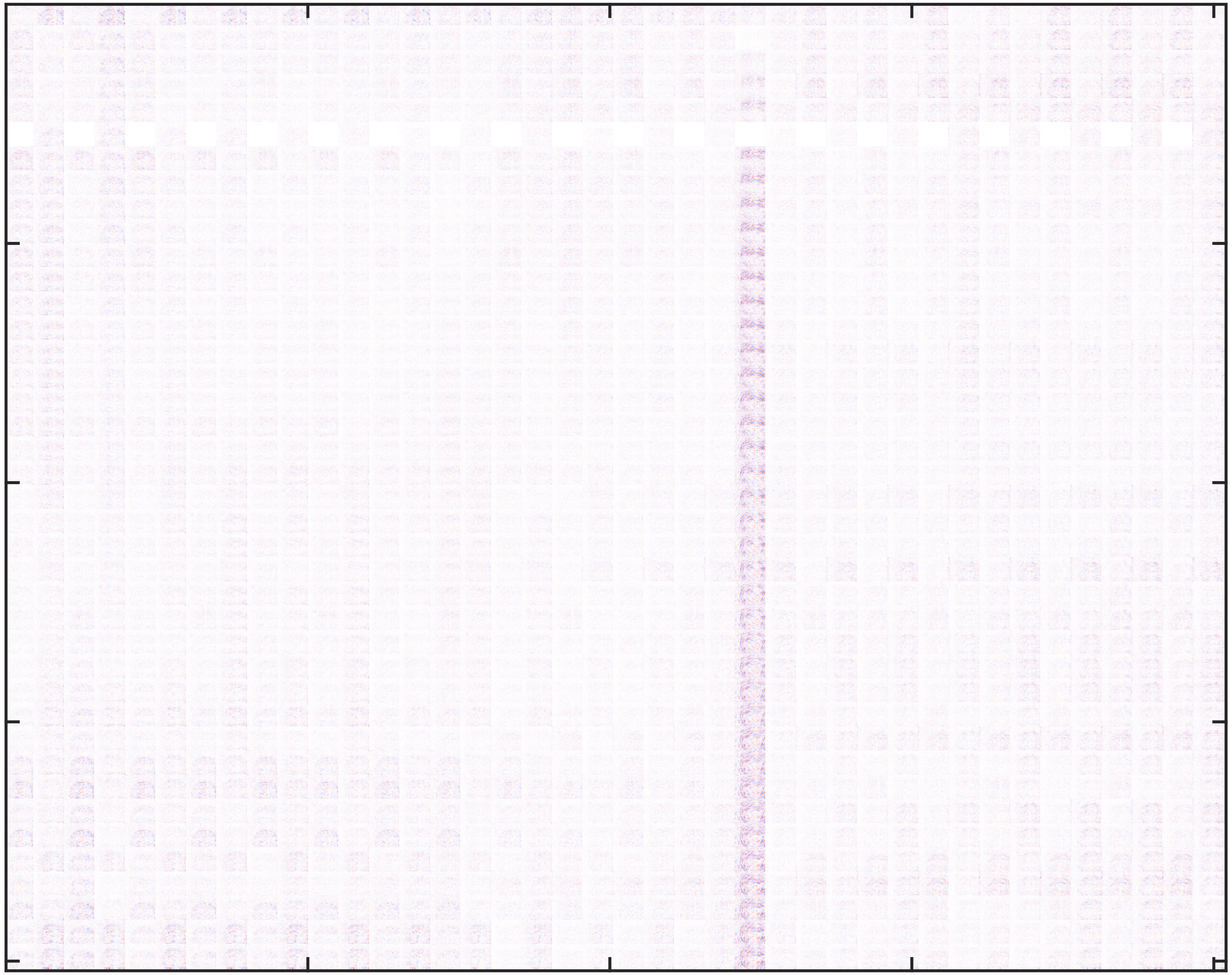




after interpolation

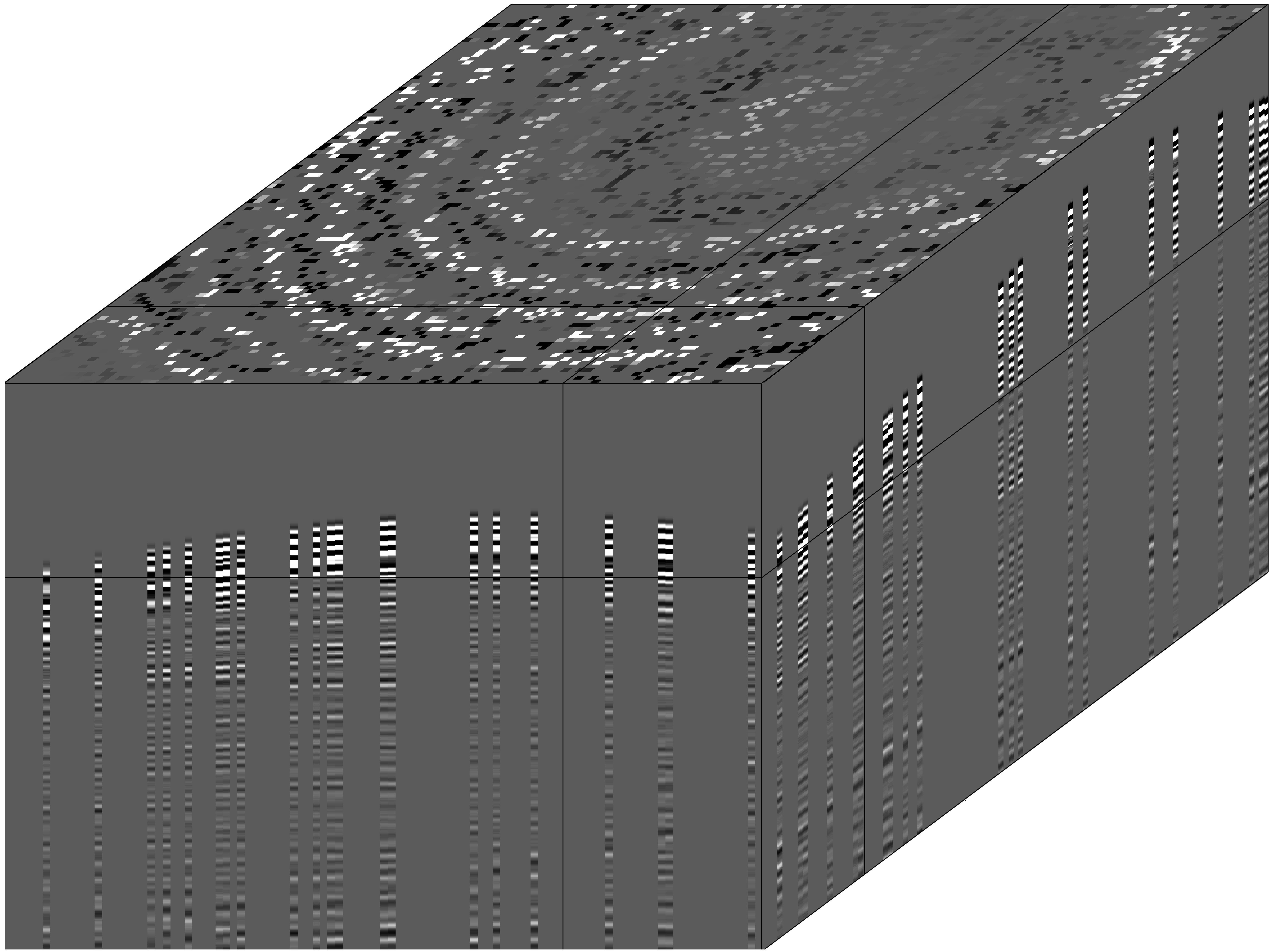
10 nodes, 10 worker each

run time : 7 minute per monochromatic slice difference



Common-receiver gather (2501 x 101 x 101)

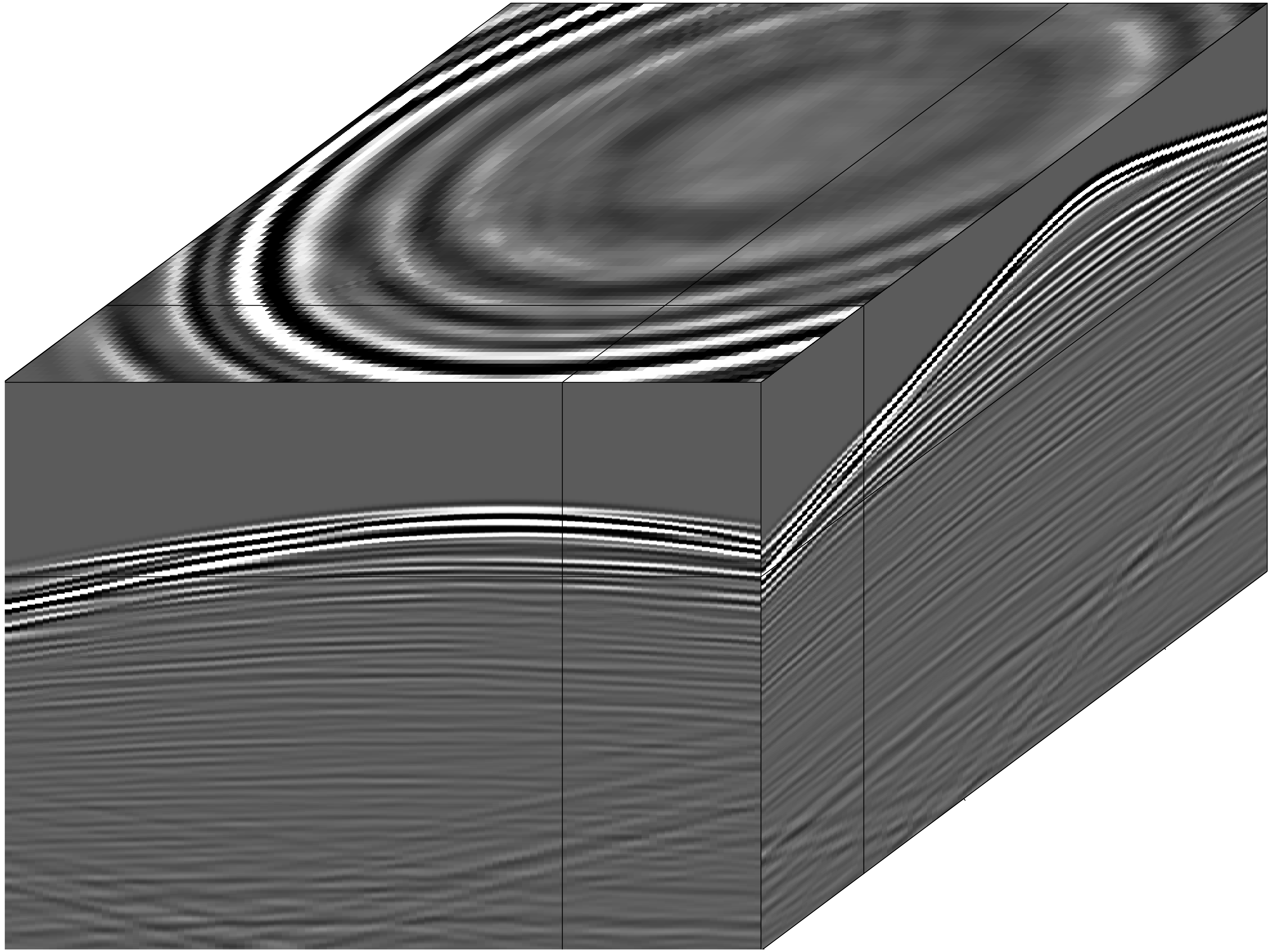
before interpolation



SLIM

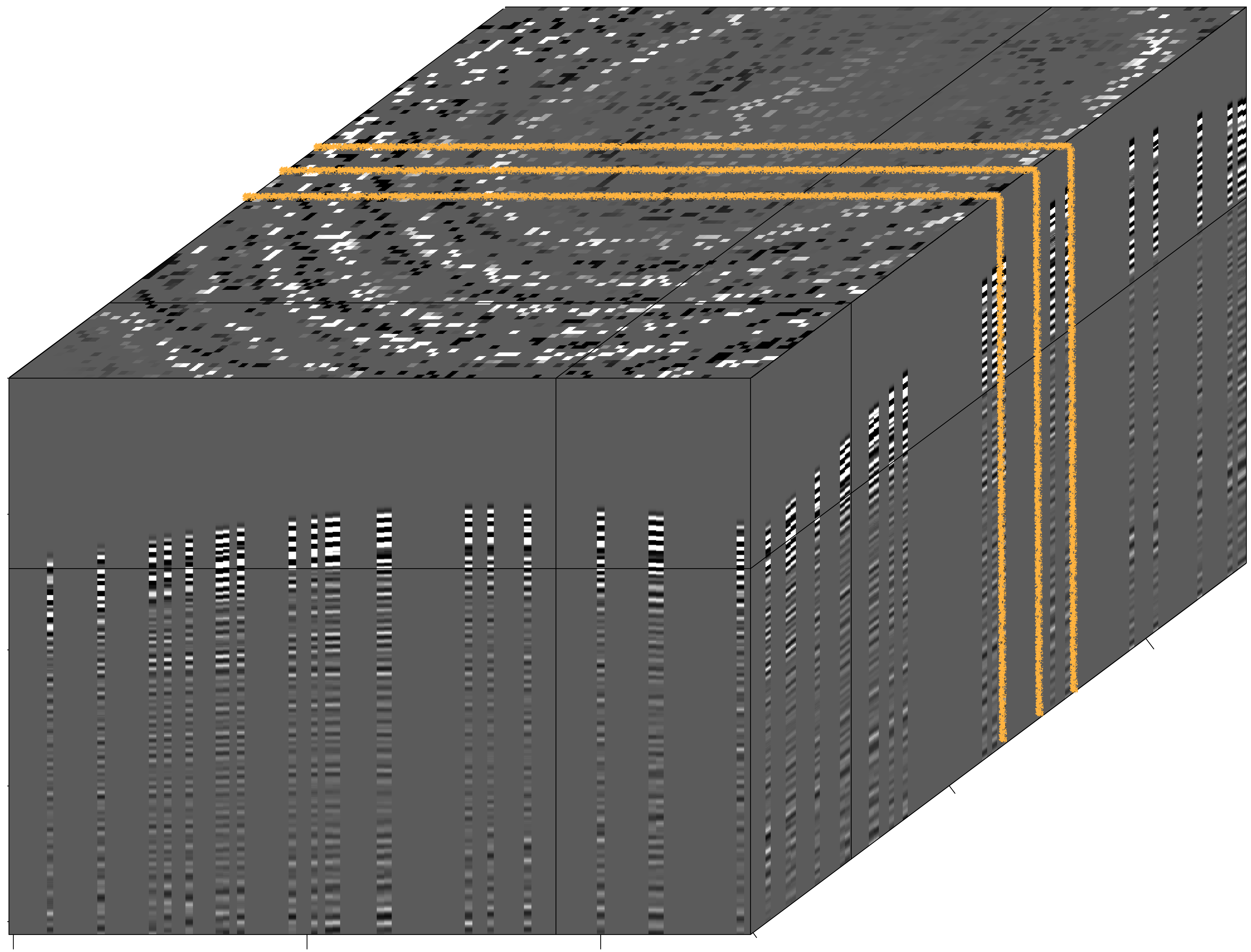
Common-receiver gather (2501 x 101 x 101)

after interpolation



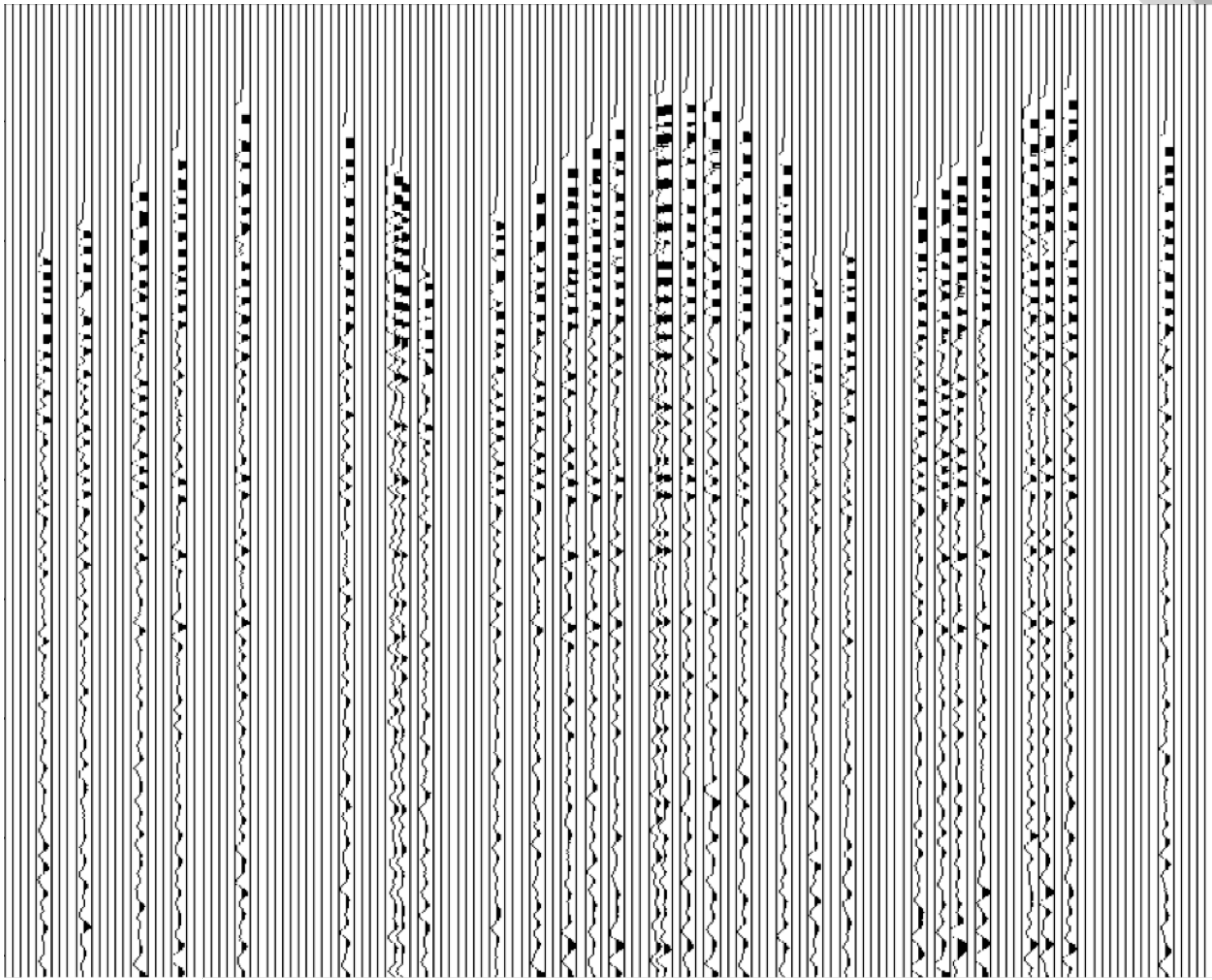
Common-receiver gather (2501 x 101 x 101)

5-times subsampling

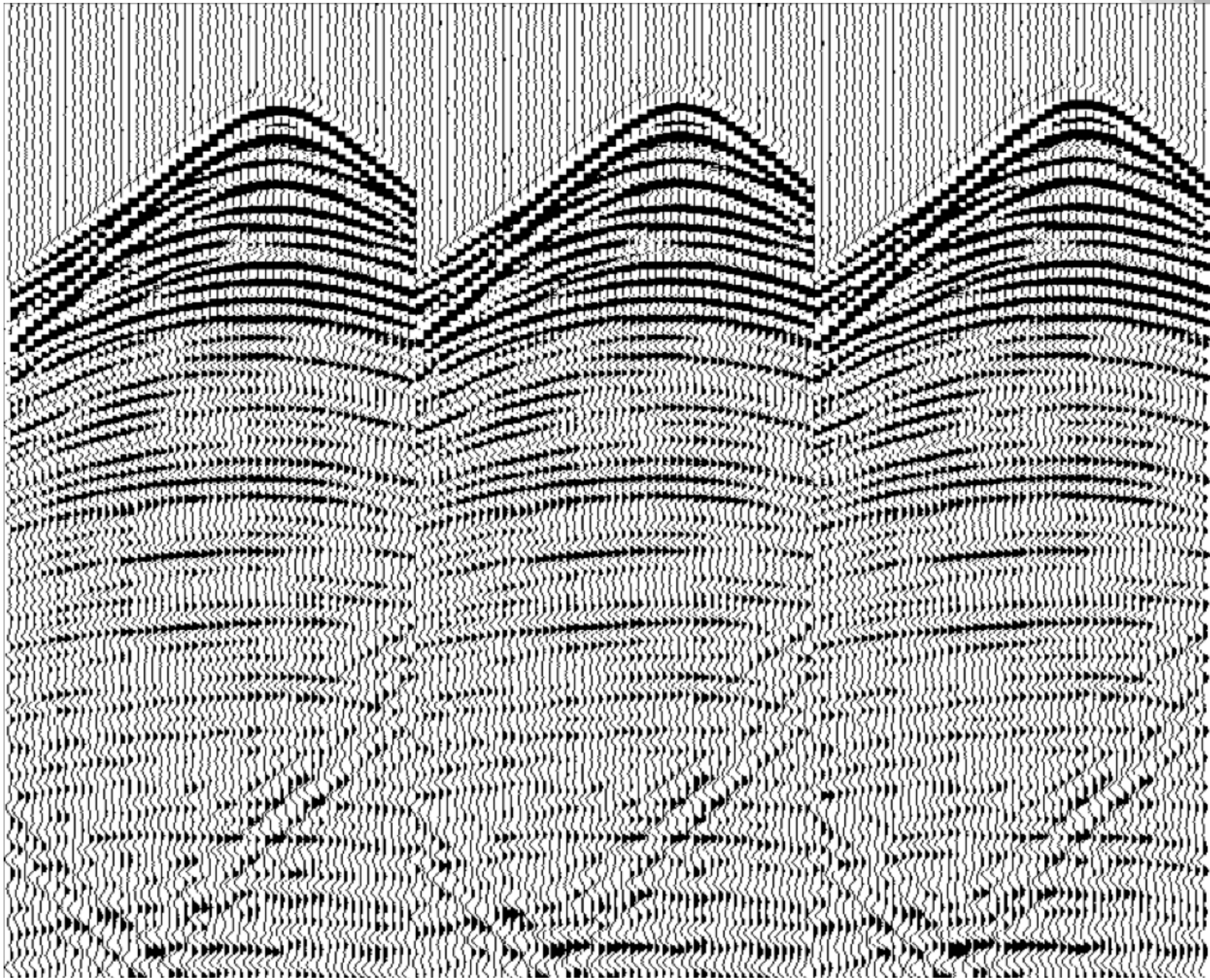


Common-receiver gather (2501 x 101 x 3)

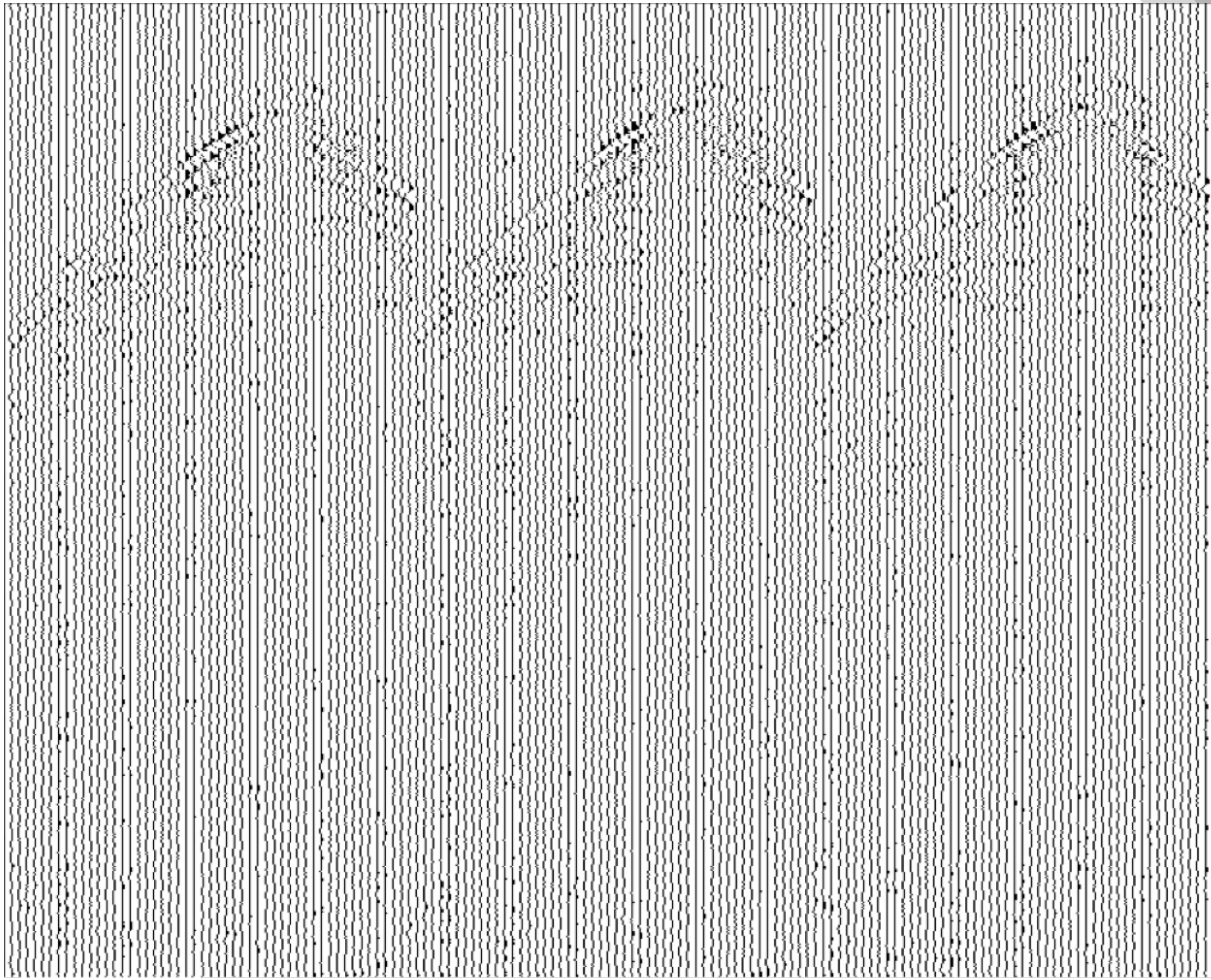
5-times subsampling



After Interpolation



Difference



Take-away message

- size of final recovered data volume is **0.15 TB**
  - no need to save fully sampled seismic data volume
- save $\mathbf{L}$ and $\mathbf{R}$ factors
  - compression rate is **98%**
  - size of final compressed 5D seismic volume is **~3 GB**

Conclusions

- Low-cost 3D OBN acquisition and processing techniques
- expandable to time-lapse static/dynamic acquisition
- Factorization based rank-minimization framework can handle large-scale seismic data
- Achieve very high compression rate for separated and interpolated volumes

Future work

- Adapt the rank-minimization framework in Julia
  - embarrassingly parallel alternate-minimization framework
- Cloud-based separation and interpolation framework
- Testing with realistic noise scenarios

Acknowledgements

This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.

Acknowledgements

The authors wish to acknowledge the SENAI CIMATEC Supercomputing Center for Industrial Innovation, with support from BG Brasil, Shell, and the Brazilian Authority for Oil, Gas and Biofuels (ANP), for the provision and operation of computational facilities and the commitment to invest in Research & Development.

Thank you for your attention