Extreme scale matrix factorizations in Exploration Seismology

Felix J. Herrmann

SLIM

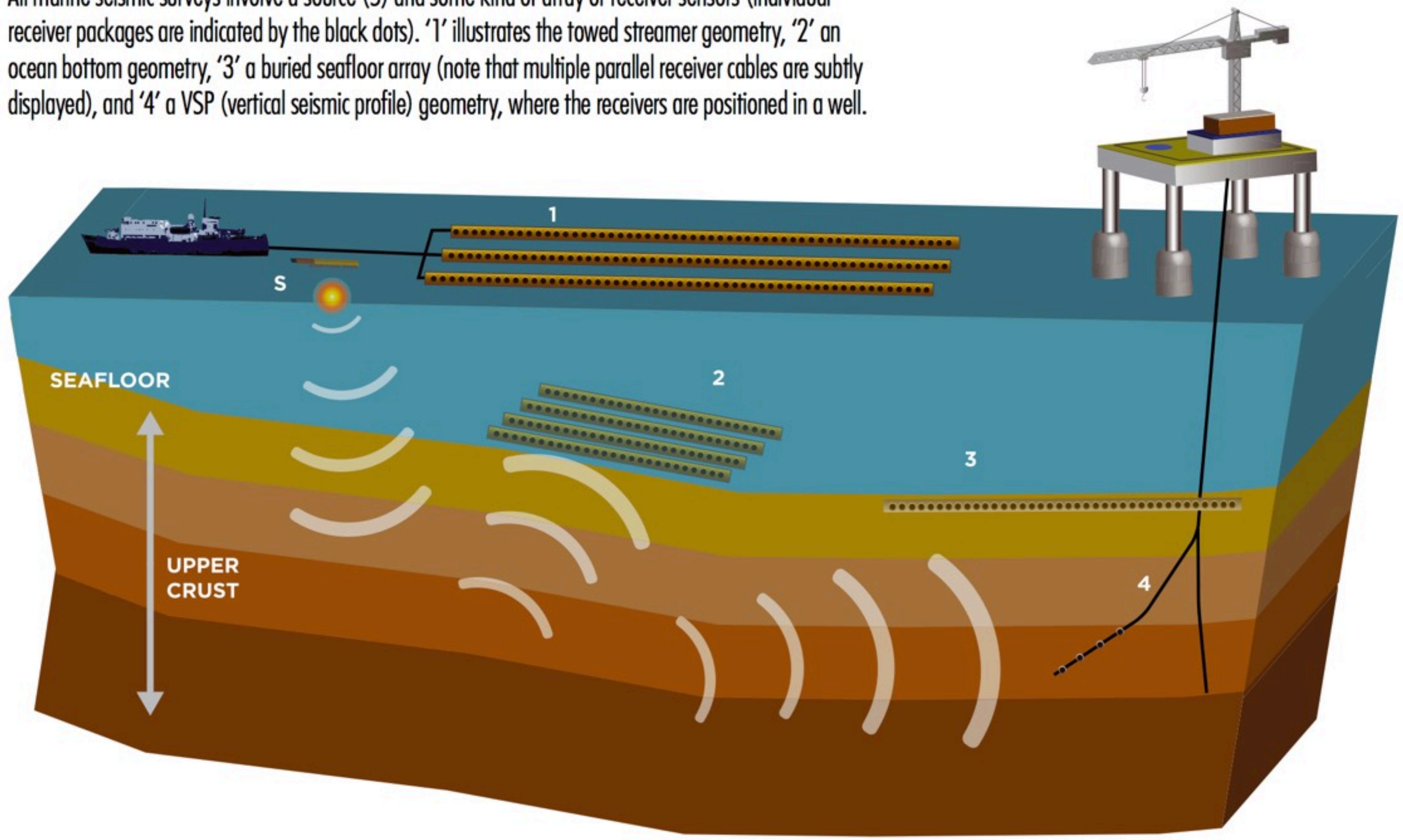
Georgia Institute of Technology

Seismic inversion

Infer 3D images & *velocity* models from *multi-experiment* data:

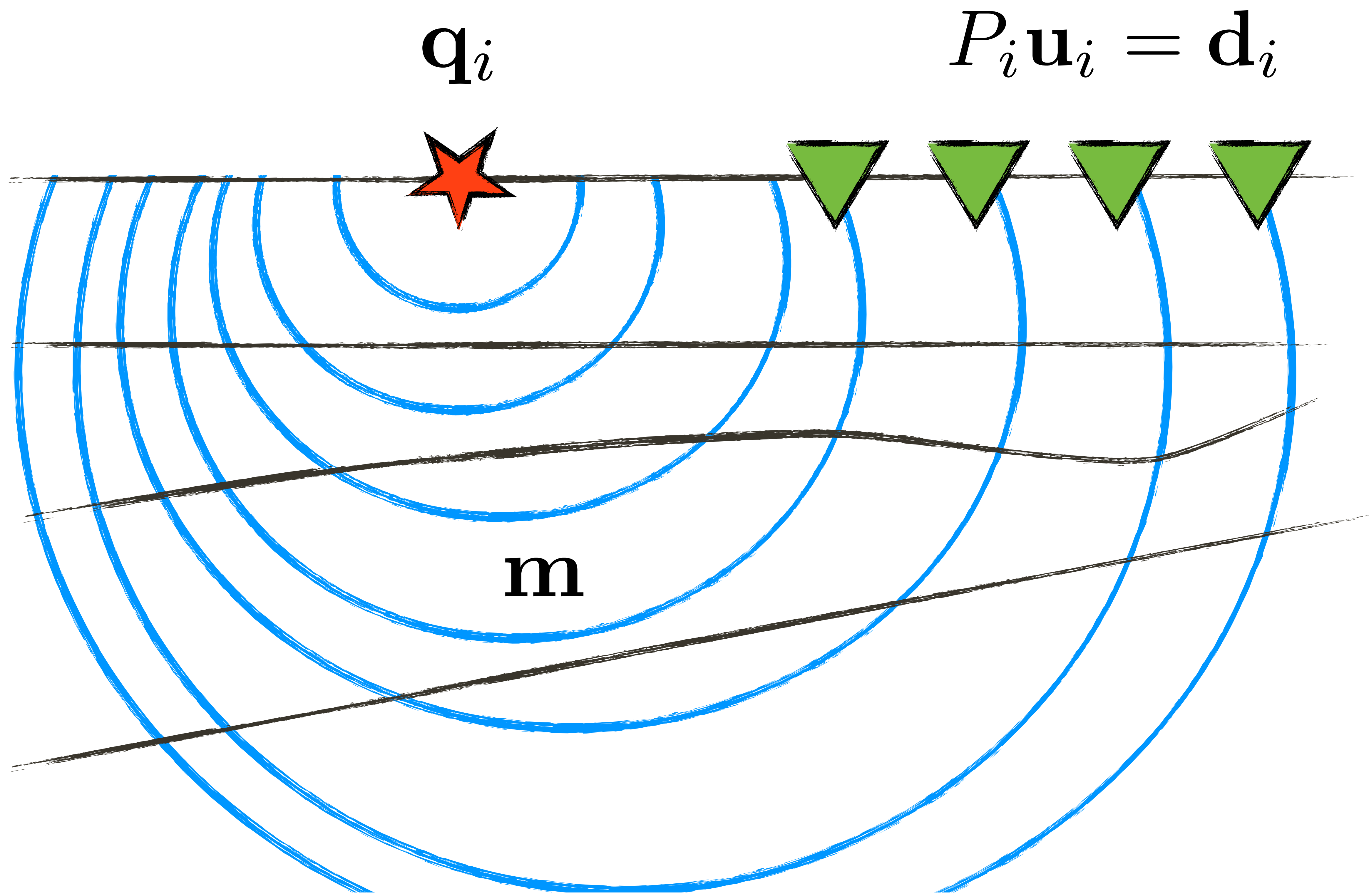
- $\mathcal{O}(10^9)$ unknowns
- $\mathcal{O}(10^{15})$ datapoints
- propagate $\mathcal{O}(10^2)$ wavelengths

All marine seismic surveys involve a source (S) and some kind of array or receiver sensors (individual receiver packages are indicated by the black dots). '1' illustrates the towed streamer geometry, '2' an ocean bottom geometry, '3' a buried seafloor array (note that multiple parallel receiver cables are subtly displayed), and '4' a VSP (vertical seismic profile) geometry, where the receivers are positioned in a well.



Wave-equation based inversion

Retrieve medium parameters $\mathbf{m}$ from partial measurements of the solution of the wave-equation: $A(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$



Wave-equation based inversion

Large-scale parameter estimation problem:

observed data

$$\underset{\mathbf{m}}{\text{minimize}}\, \Phi(\mathbf{m}) = \frac{1}{M}\sum_{i=1}^{M}\phi_i(\mathbf{m}) = \frac{1}{2}\|P_i\mathbf{u}_i - \mathbf{d}_i\|^2$$

- number of field experiments M is large $\mathcal{O}(10^3 - 10^5)$
- $\mathbf{d}_i$ expensive to collect $\mathcal{O}(10^6 - 10^7)$ data points at total survey costs of $30 – 200 M
- ϕ_i expensive to evaluate $\mathcal{O}(10^{14})$ flops per experiment w/ HPC costs of $25 – 500 M
- $\mathbf{m}$ is extremely $\mathcal{O}(10^6 - 10^9)$ large requiring local (= gradient-based) optimization

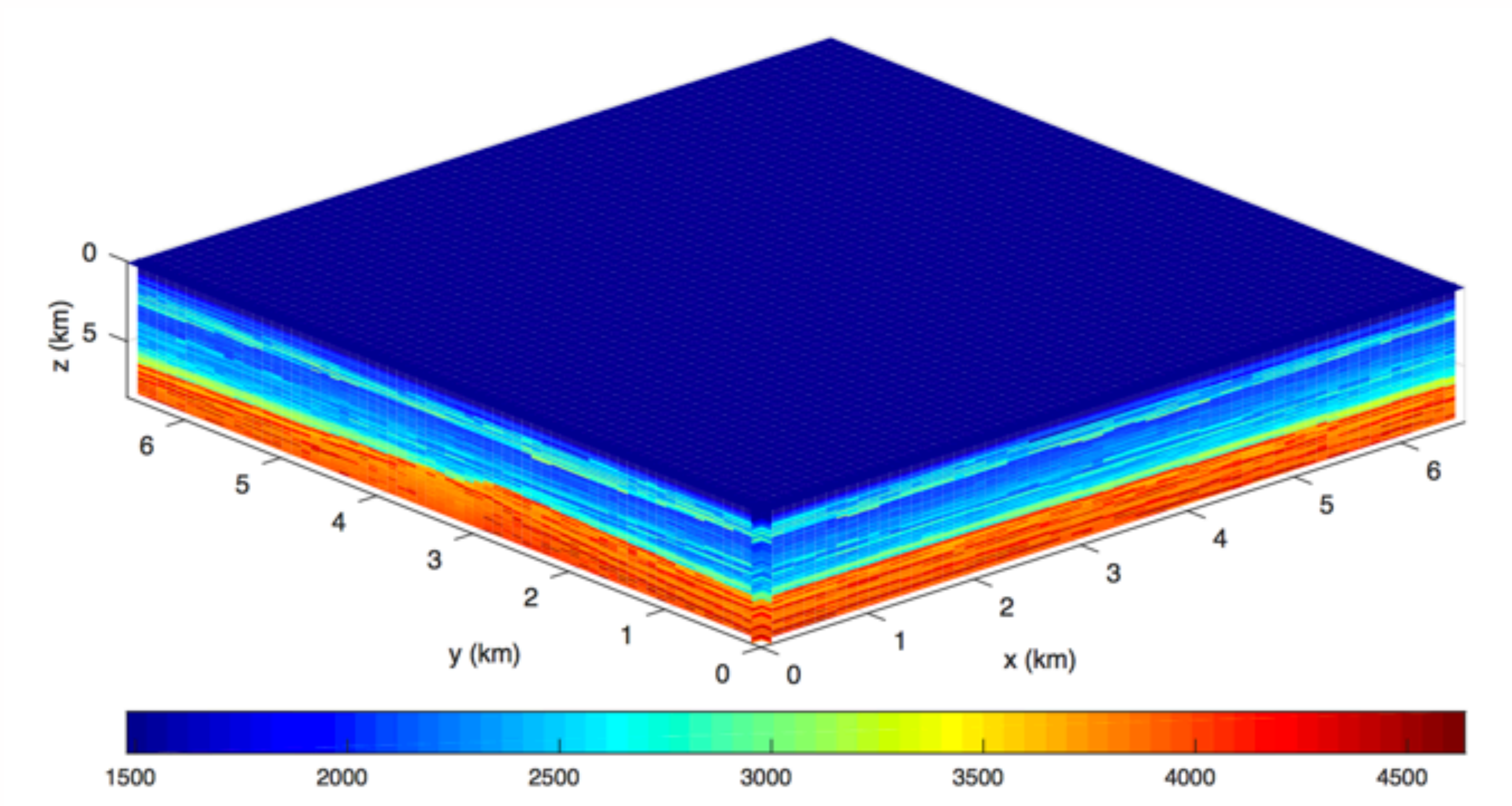
Research questions

"How can we exploit low-rank structure underlying surface seismic data and subsurface image volumes at low to mid frequencies?"

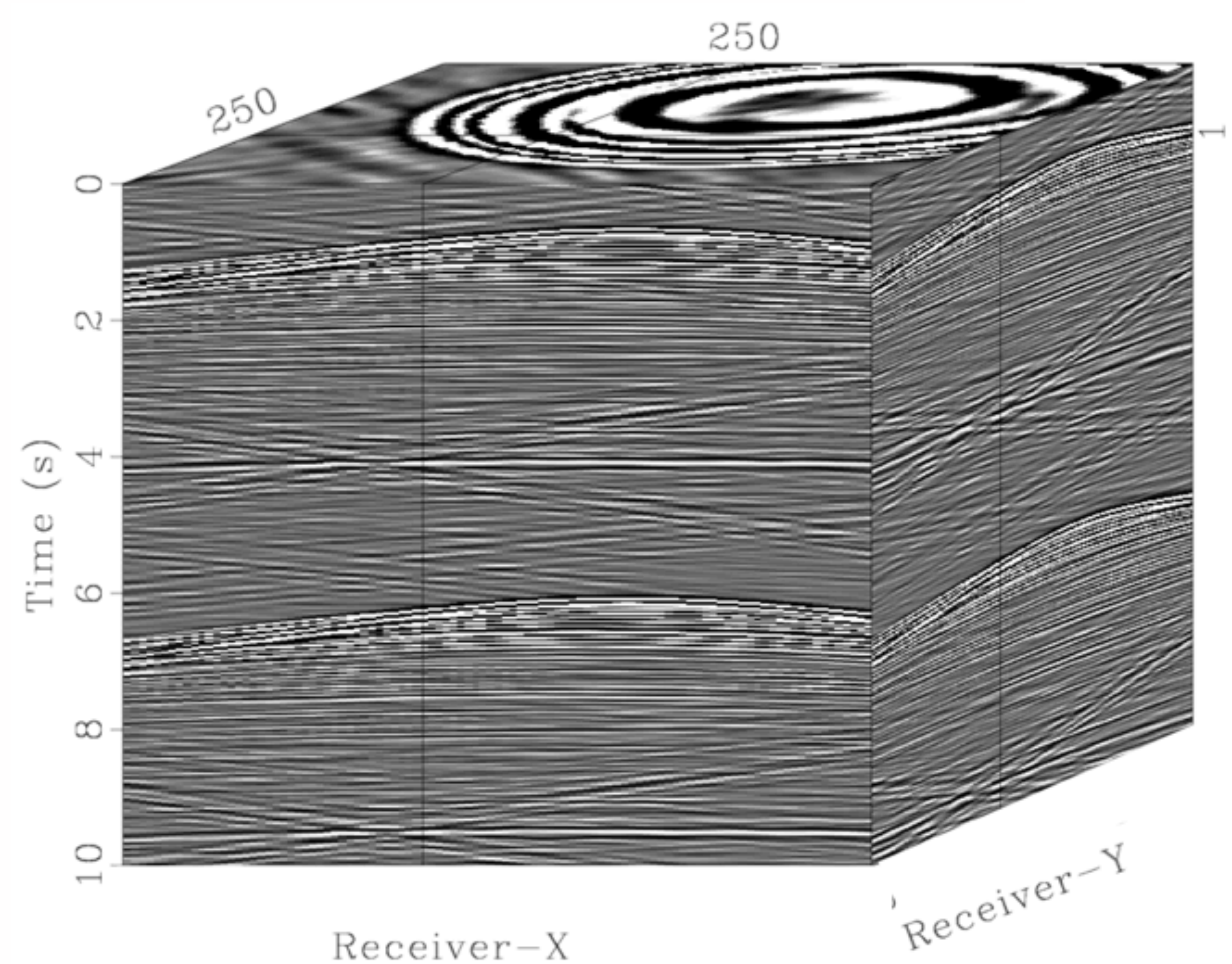
- reduce acquisition time, costs, environmental imprint of via randomized sampling & full azimuth processing
- lower storage & IO cost of wave-equation based inversion via on-the-fly data generation from data represented in factorized form
- form & manipulate massive full-subsurface offset image volumes via randomized probing of the double two-way wave equation

Motivation – seismic surface data

Large 5D volumes of seismic data

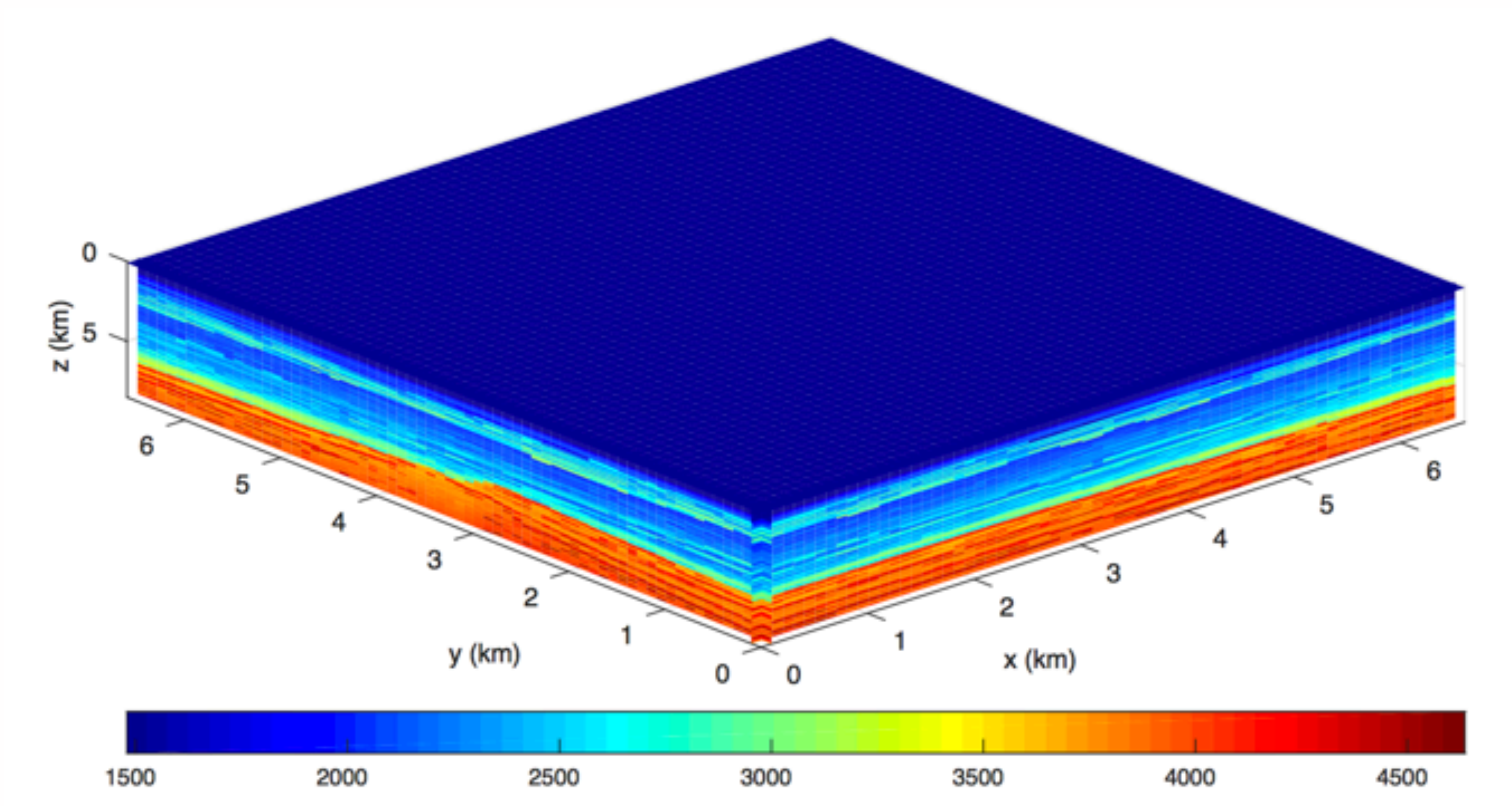


100’s of thousands of shots

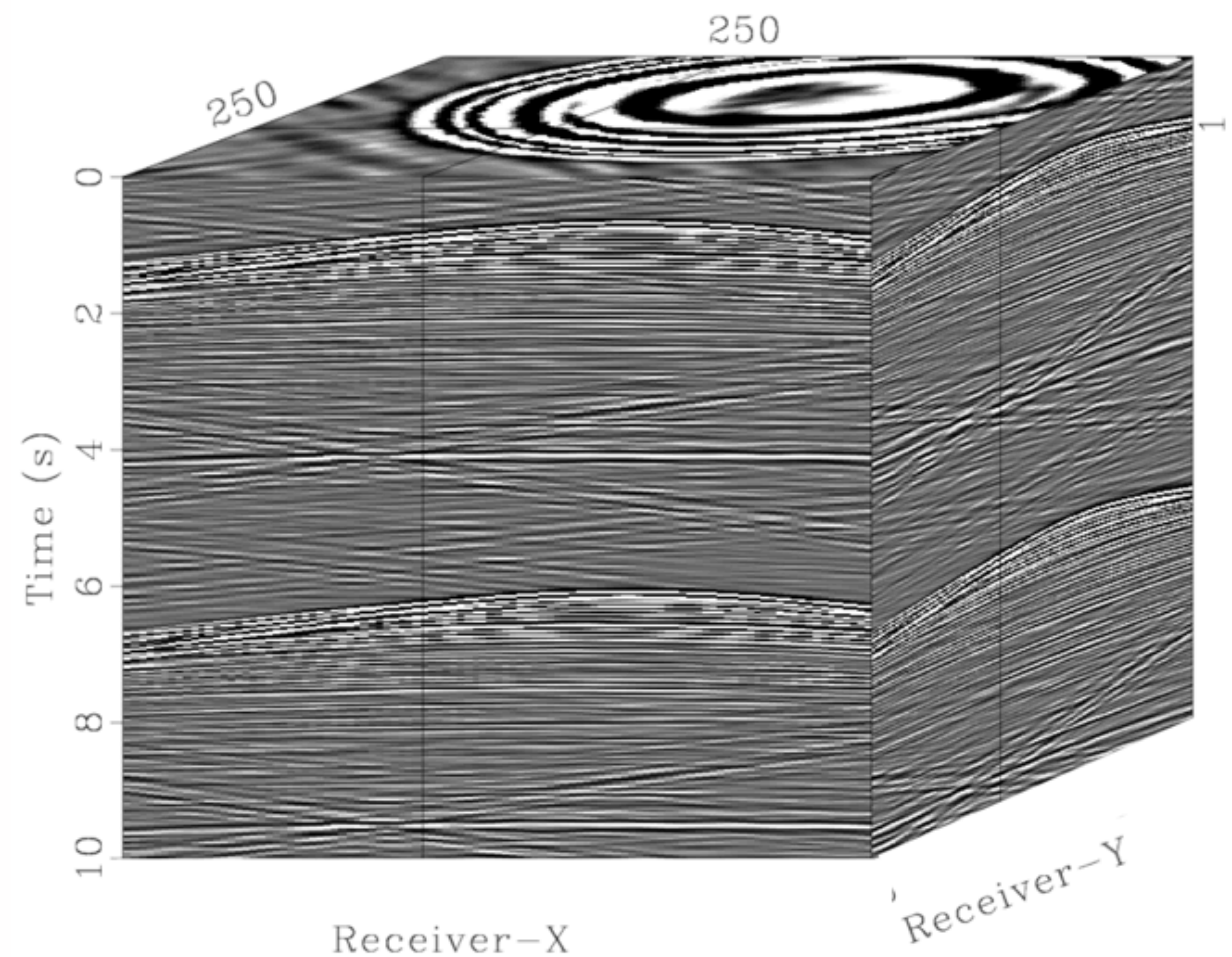


Motivation – seismic surface data

Large 5D volumes of seismic data



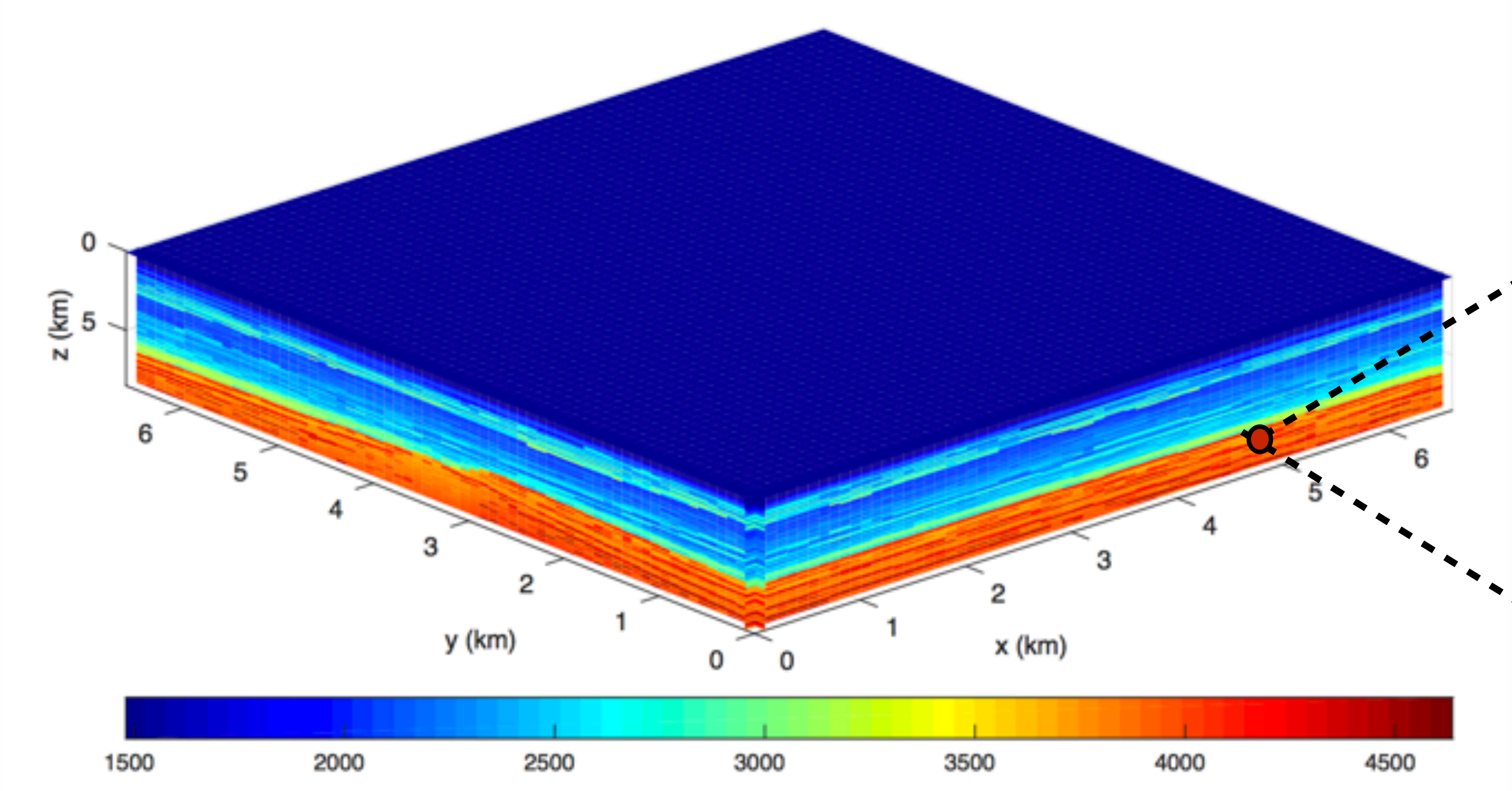
100's of thousands of shots



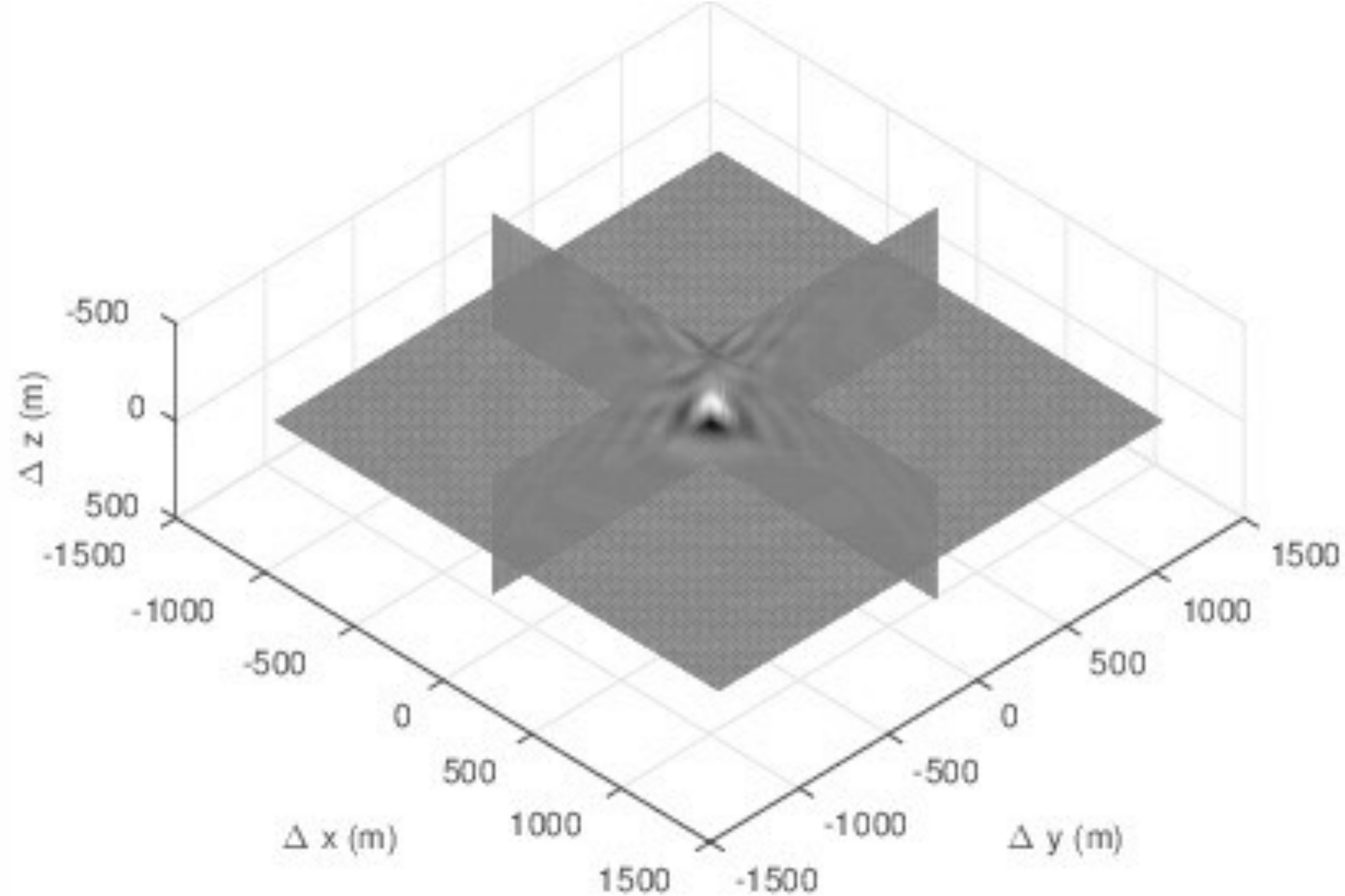
Will soon reach petabytes.

Motivation – image volumes

Extremely large 6D image volumes



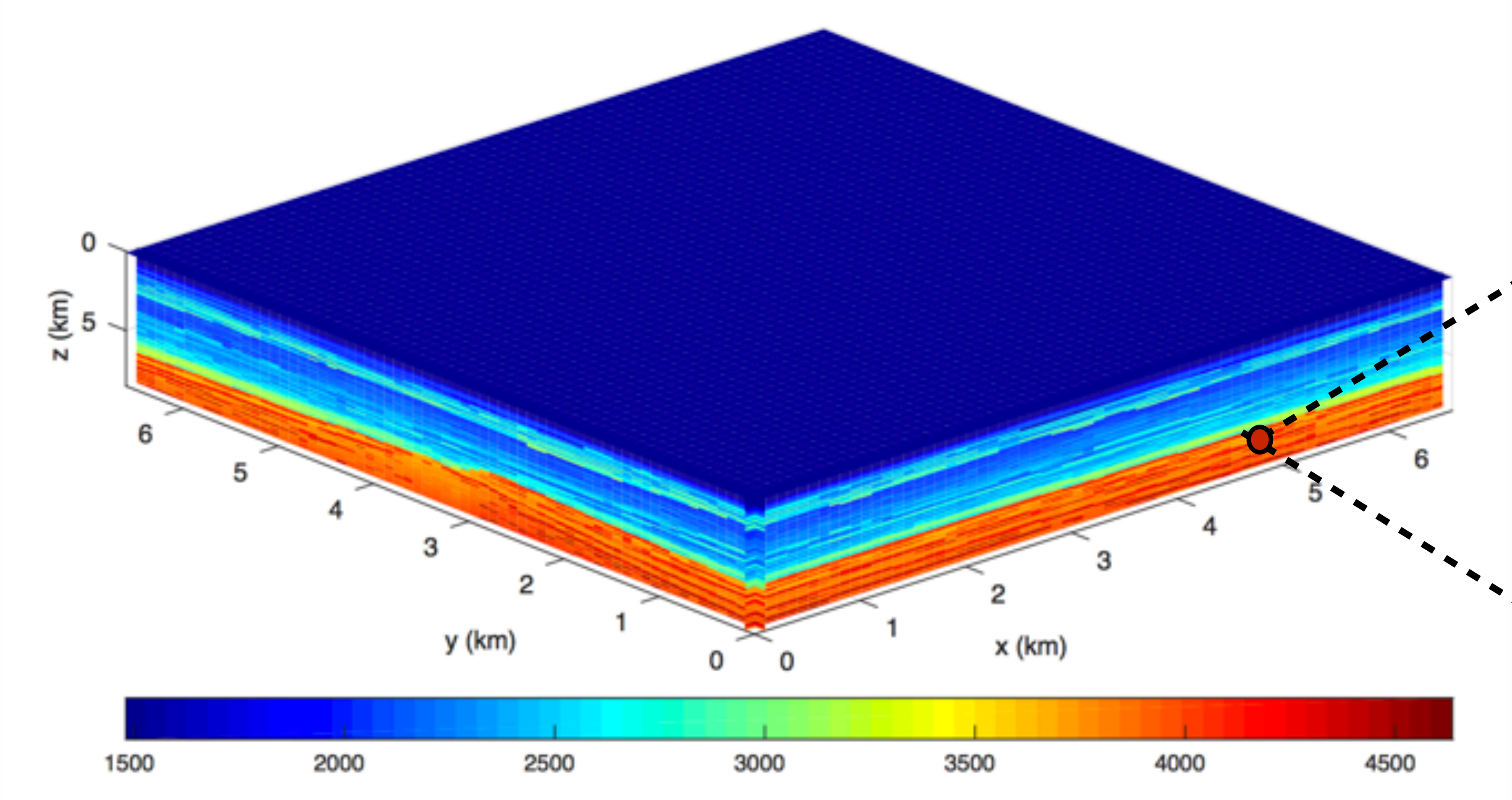
quadratic in image size

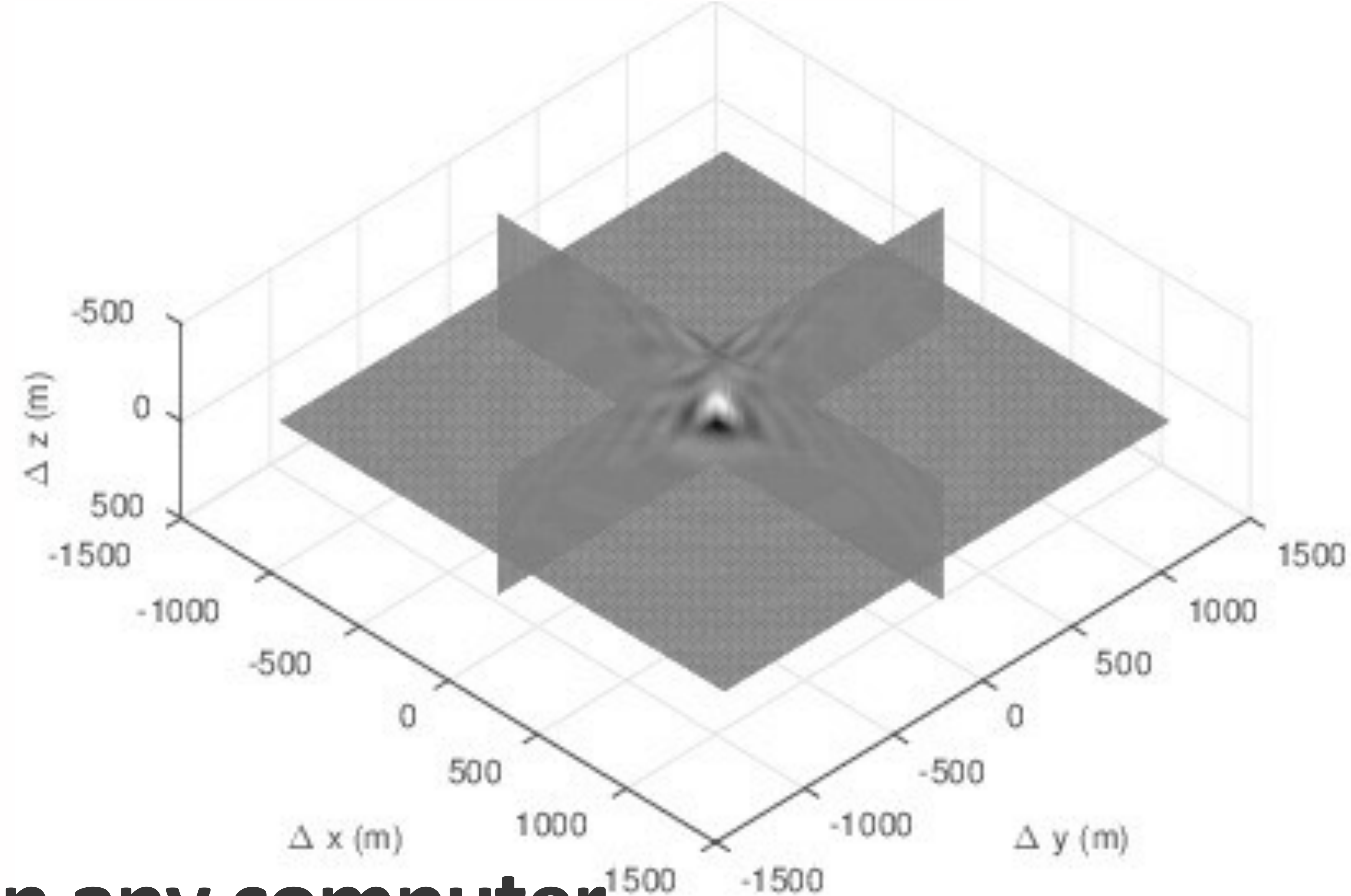


Motivation – image volumes

Extremely large 6D image volumes

quadratic in image size





Can not be formed explicitly on any computer...

Full-azimuth seismic data processing with coil acquisition

Rajiv Kumar, Nick Moldoveanu, Keegan Lensink, and Felix J. Herrmann

SLIM
Georgia Institute of Technology

3D seismic

Challenge seismic data collected in 5 dimensions

- 1 for time
- 2 for the receivers
- 2 for the sources

Compressive Sensing works well for vectorial transform-domain sparsity

- curvelets & other non-separable transforms are too slow & memory intensive
- prohibits scale up to 5D

Can we exploit a different kind of structure ...

Quick recap—matrix completion

Aleksandr Y. Aravkin, Rajiv Kumar, Hassan Mansour, Ben Recht, and Felix J. Herrmann, "**Fast methods for denoising matrix completion formulations, with applications to robust seismic data interpolation**", *SIAM Journal on Scientific Computing*, vol. 36, p. S237-S266, 2014

Rajiv Kumar, Haneet Wason, and Felix J. Herrmann, "**Source separation for simultaneous towed-streamer marine acquisition –- a compressed sensing approach**", *Geophysics*, vol. 80, p. WD73-WD88, 2015.

Rajiv Kumar, Curt Da Silva, Okan Akalin, Aleksandr Y. Aravkin, Hassan Mansour, Ben Recht, and Felix J. Herrmann, "**Efficient matrix completion for seismic data reconstruction**", *Geophysics*, vol. 80, p. V97-V114, 2015.

Curt Da Silva and Felix J. Herrmann, "**Optimization on the Hierarchical Tucker manifold - applications to tensor completion**", *Linear Algebra and its Applications*, vol. 481, p. 131-173, 2015.

[Candes and Plan 2010, Oropeza and Sacchi 2011]

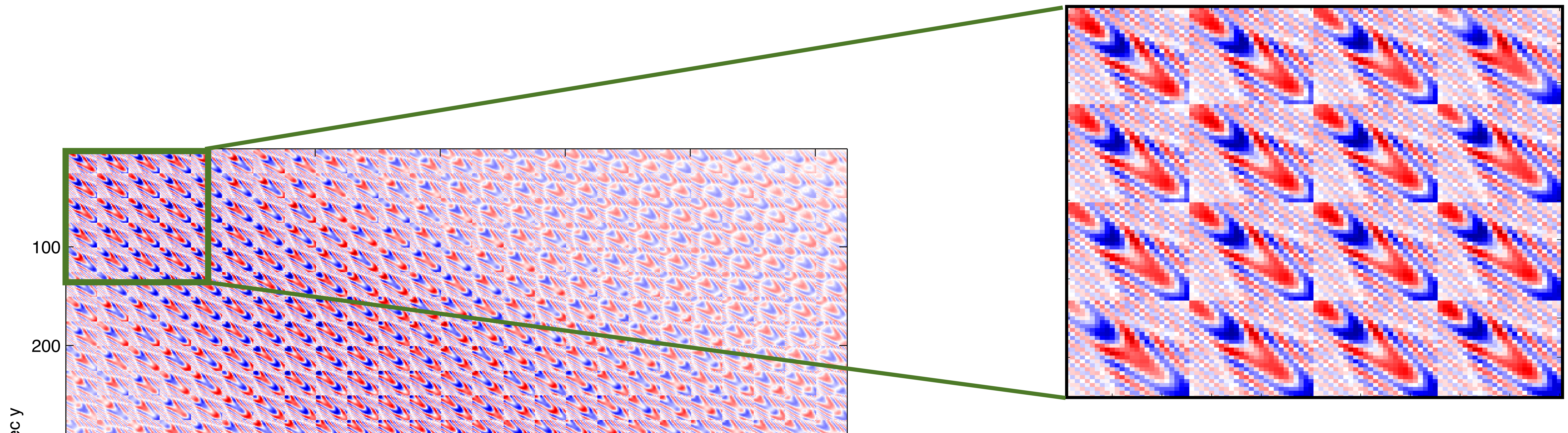
Matrix completion

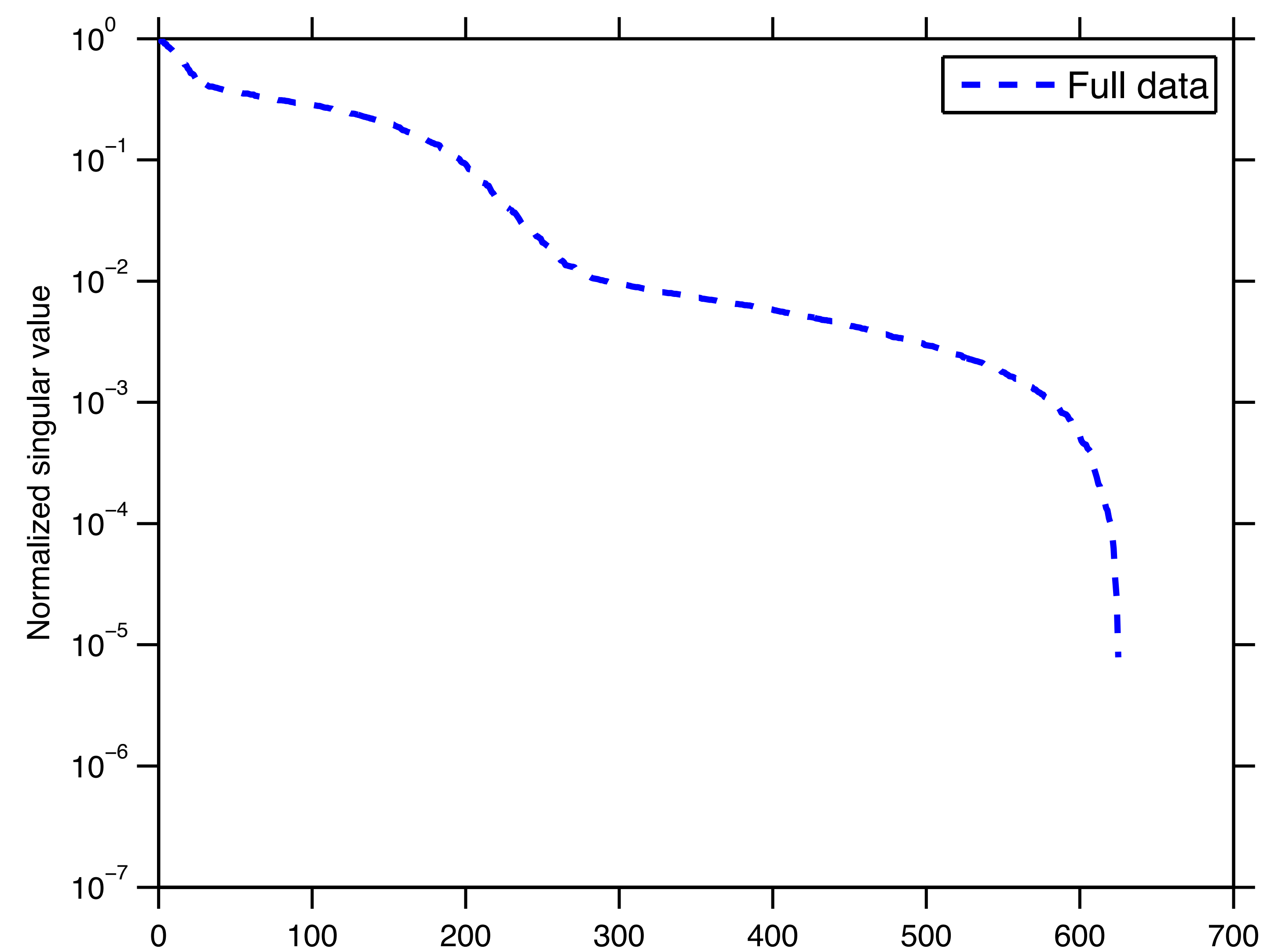
- signal structure
  - *low rank/fast decay* of singular values

- sampling scheme
  - missing data *increase* rank in "transform domain"

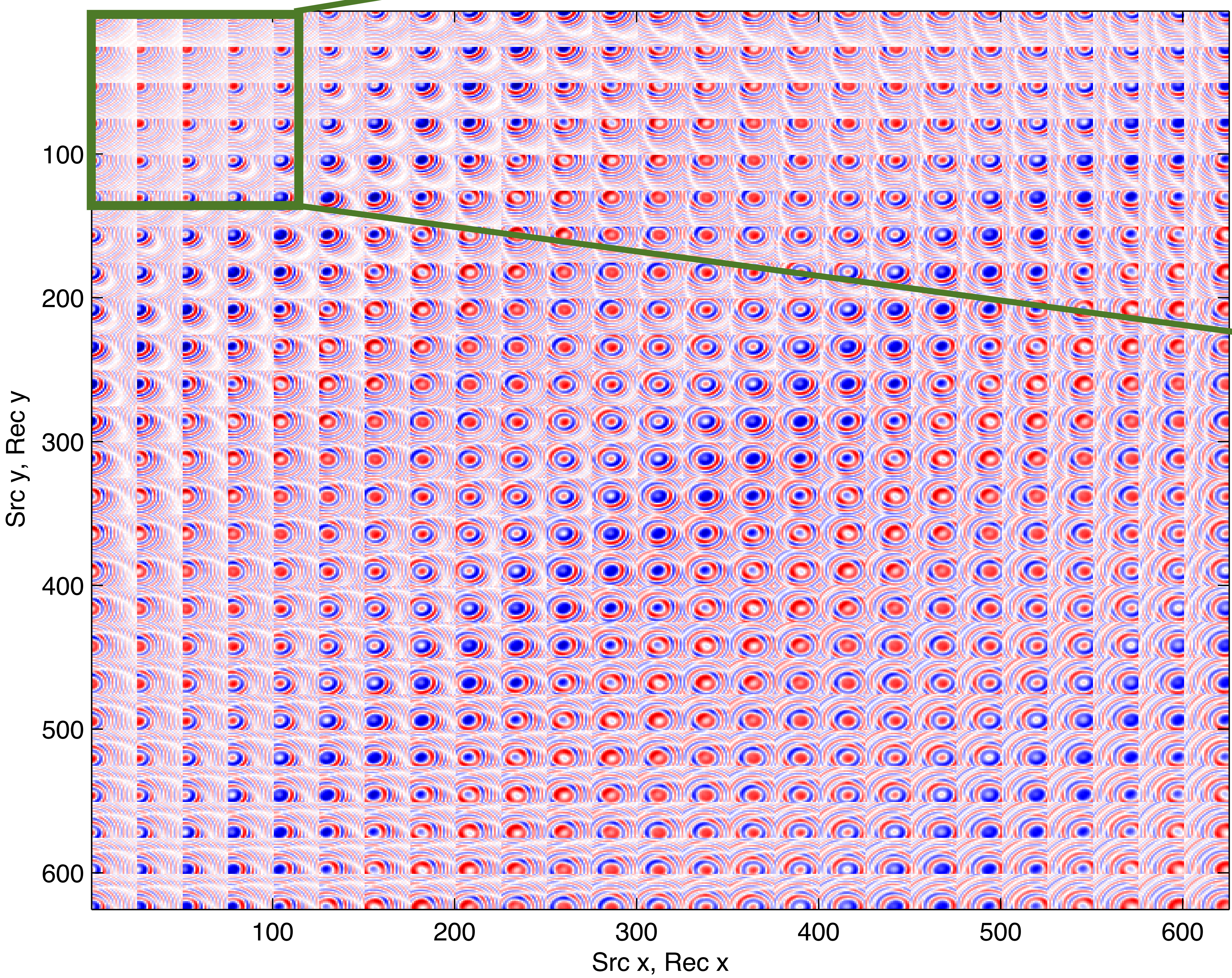
- recovery using *rank penalization* scheme

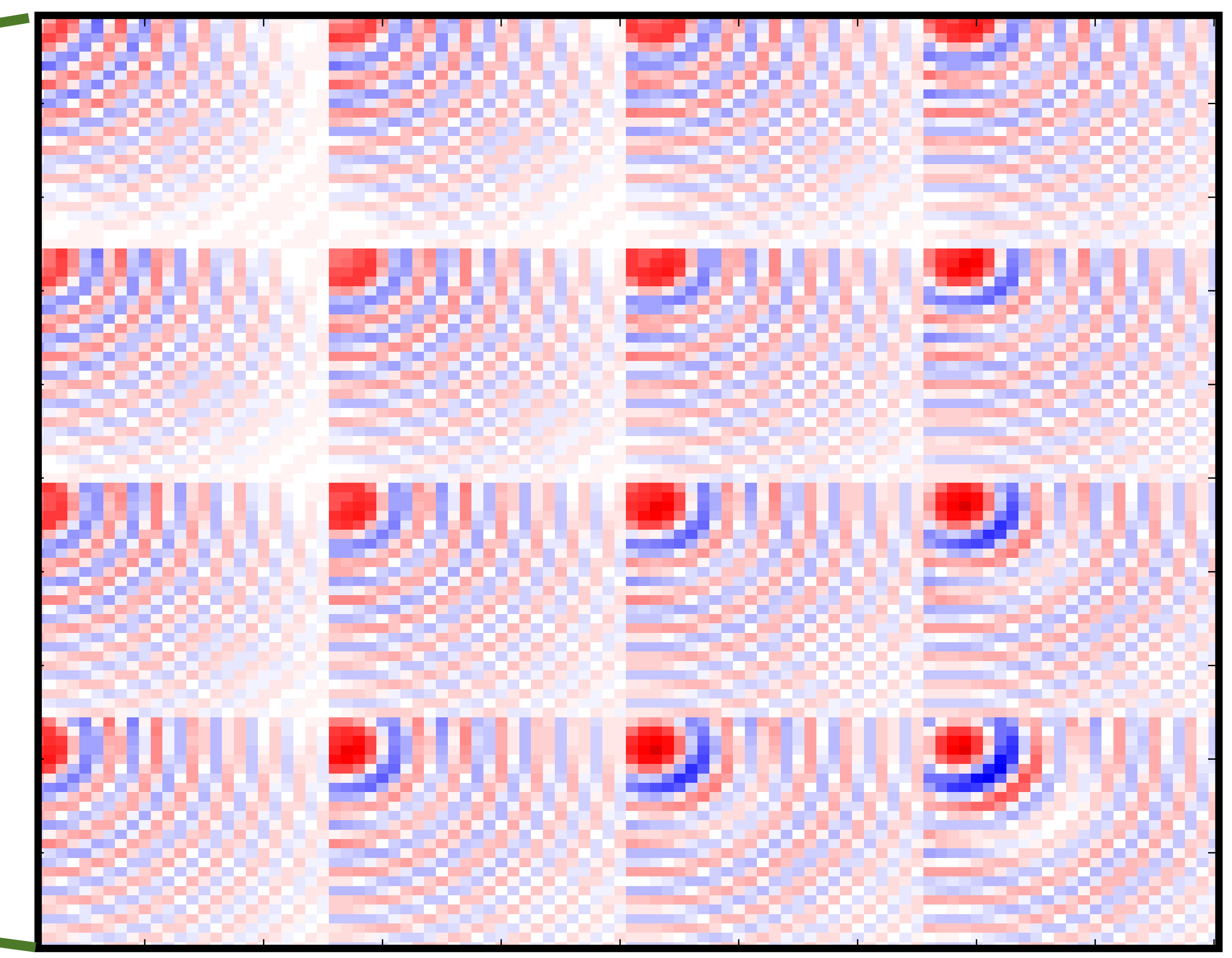
Low-rank structure
conventional 5D data, 5 Hz monochromatic slice, Sx-Sy matricization

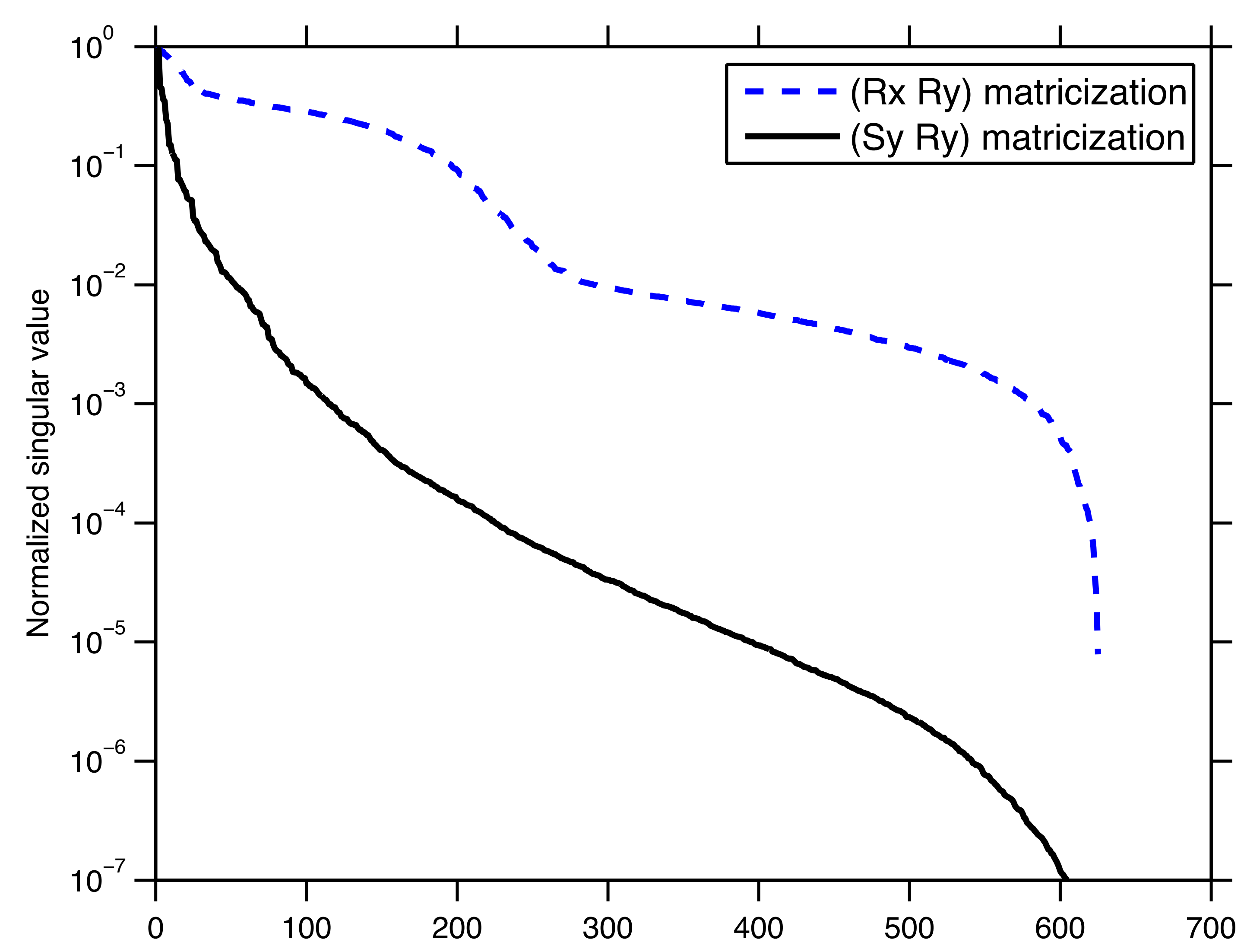




Low-rank structure
conventional 5D data, 5 Hz monochromatic slice, Sx-Rx matricization



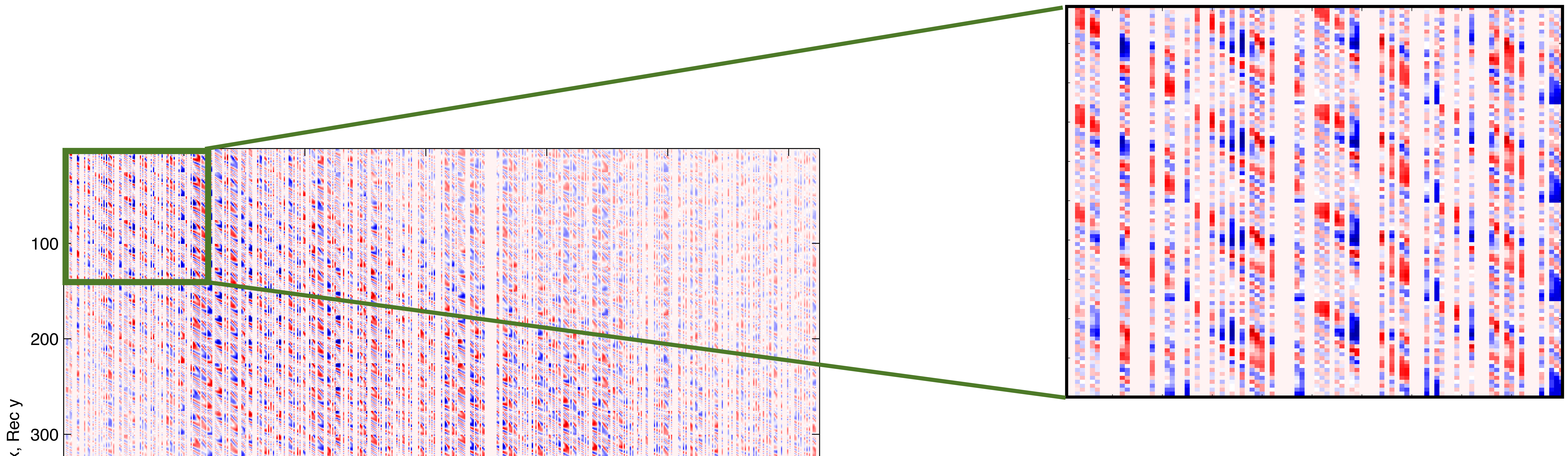


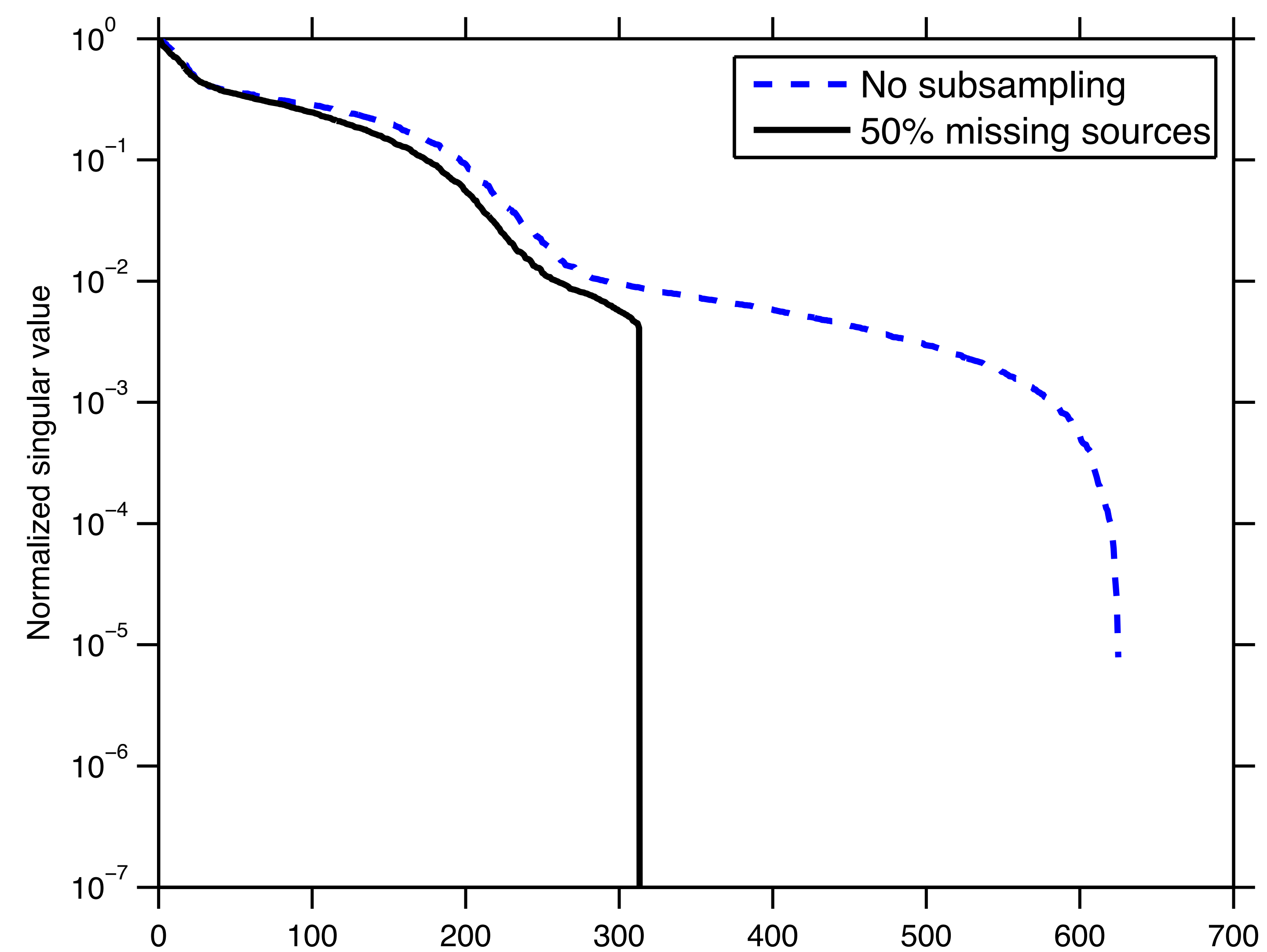


Matrix completion

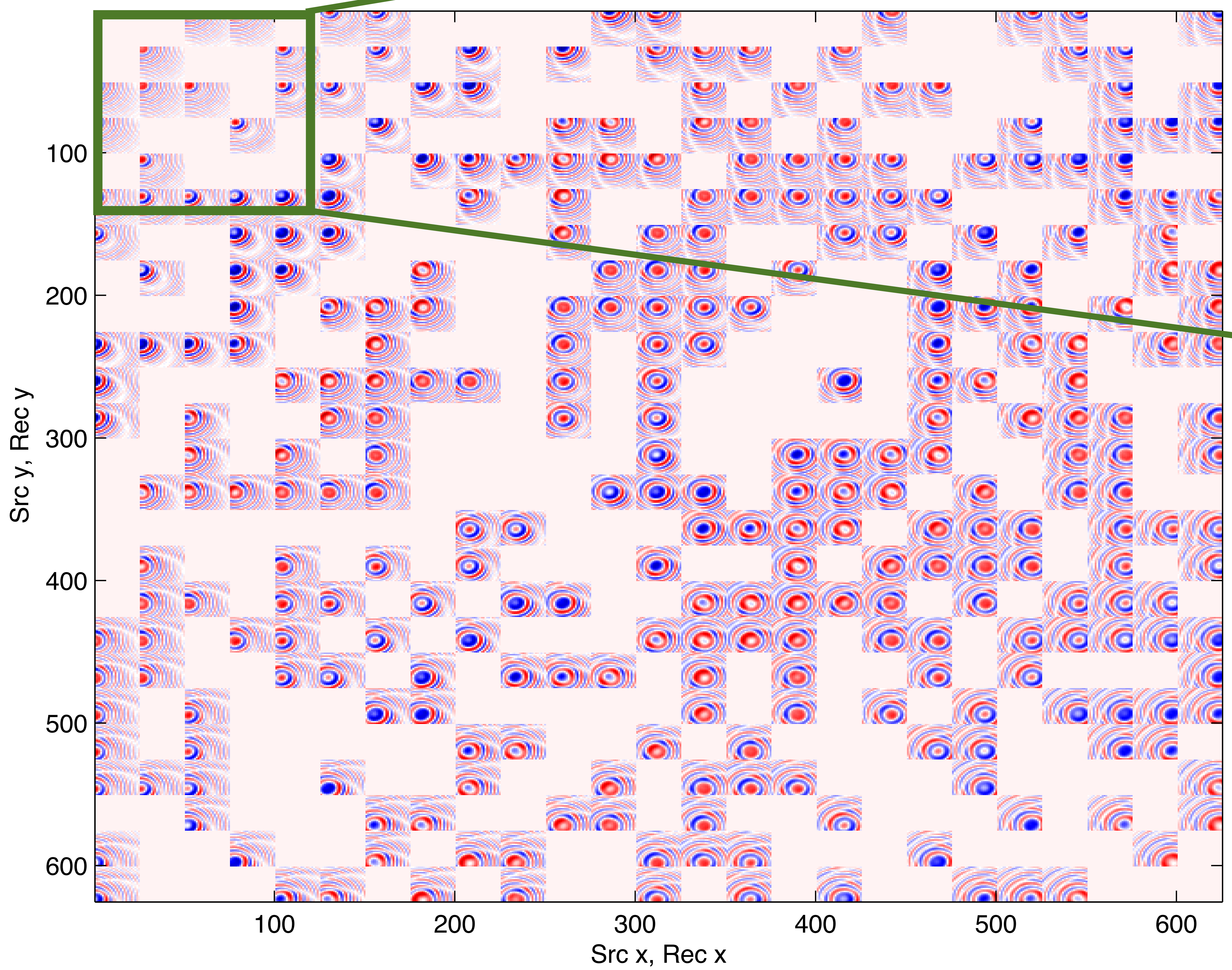
- signal structure
  - *low rank/fast decay* of singular values
- sampling scheme
  - missing data *increase* rank in "transform domain"
- recovery using *rank penalization* scheme

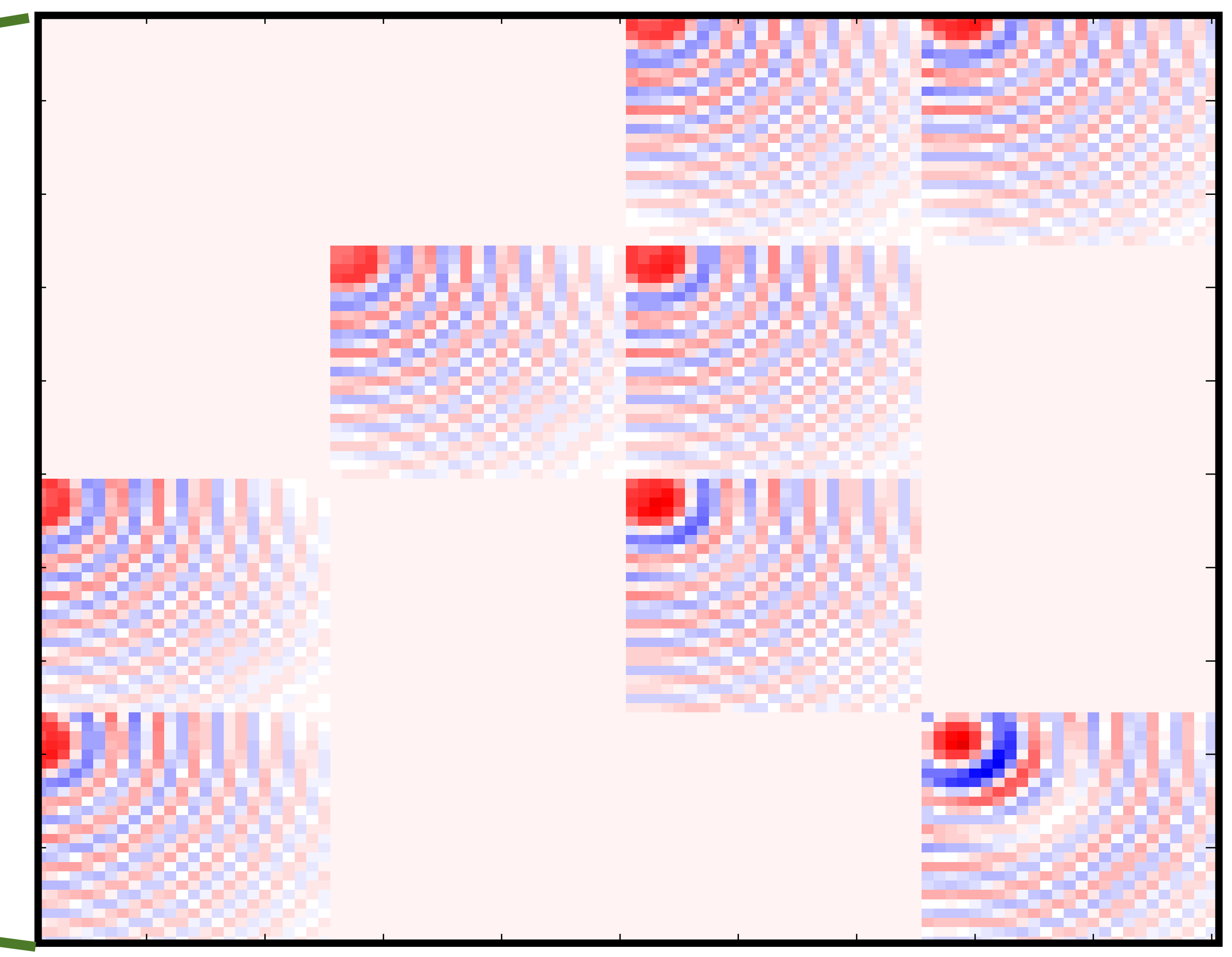
Low-rank structure
jittered data, 5 Hz monochromatic slice, Sx-Sy matricization

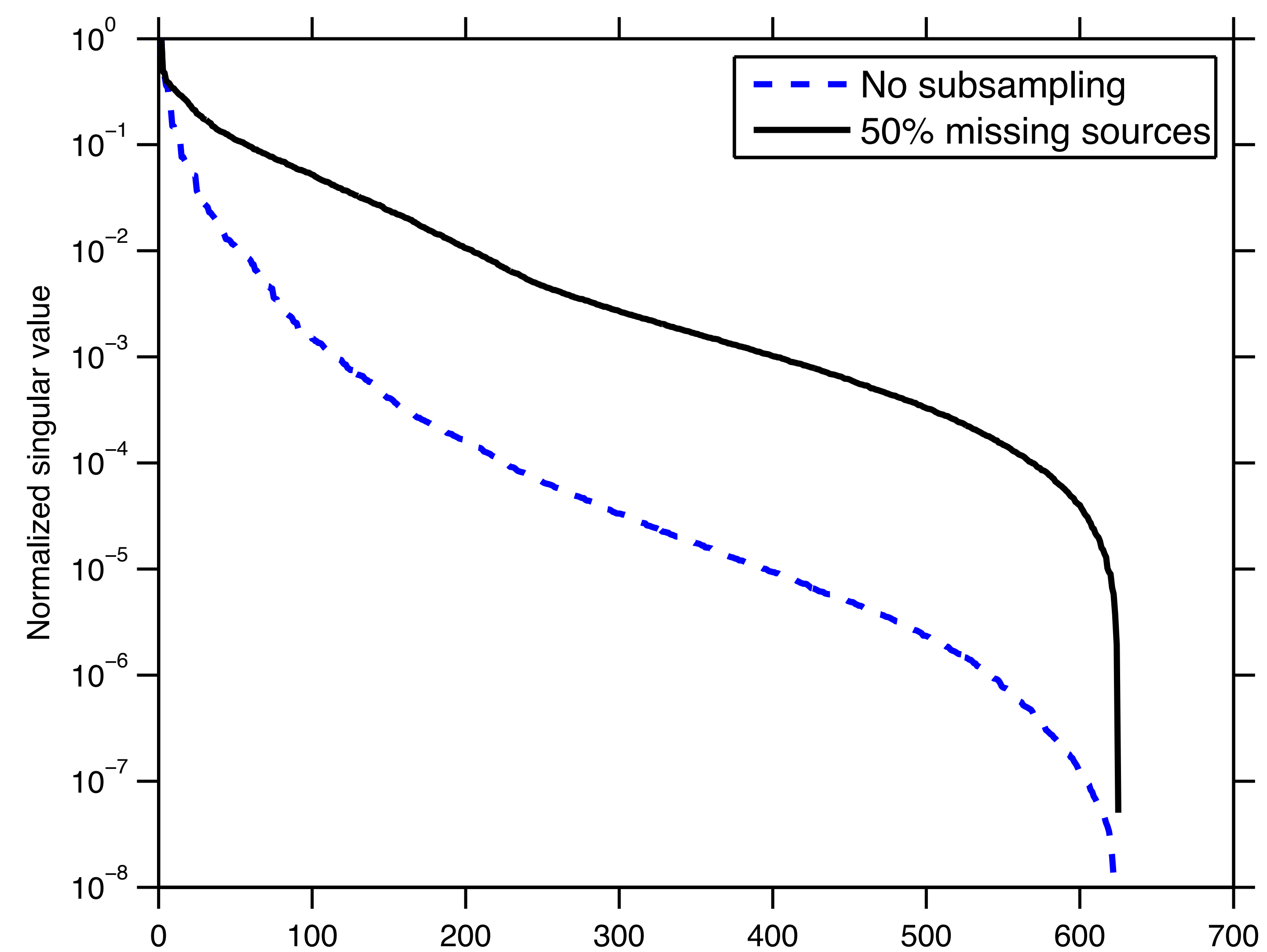




Low-rank structure
jittered data, 5 Hz monochromatic slice, Sx-Rx matricization







Matrix completion

- signal structure
  - *low rank/fast decay* of singular values
- sampling scheme
  - missing data *increase* rank in "transform domain"
- recovery using *rank penalization* scheme

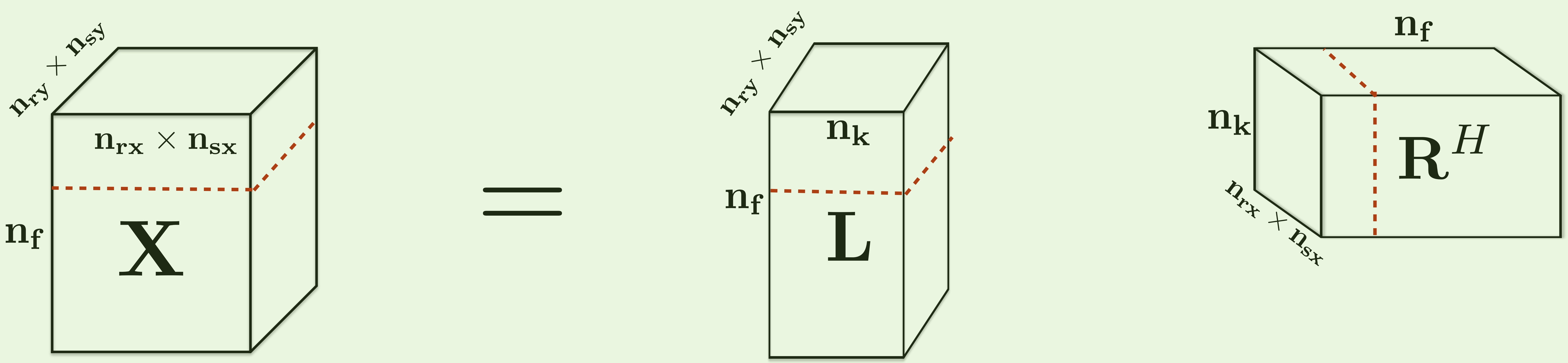
Nuclear-norm minimization *convex relaxation of rank-minimization*

[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]

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



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

[Rennie and Srebro 2005]

Factorized formulation

- Upper-bound on nuclear norm is defined as

$$\|\mathbf{L}\mathbf{R}^H\|_* \leq \frac{1}{2}\left\|\begin{bmatrix}\mathbf{L}\\ \mathbf{R}\end{bmatrix}\right\|_F^2$$

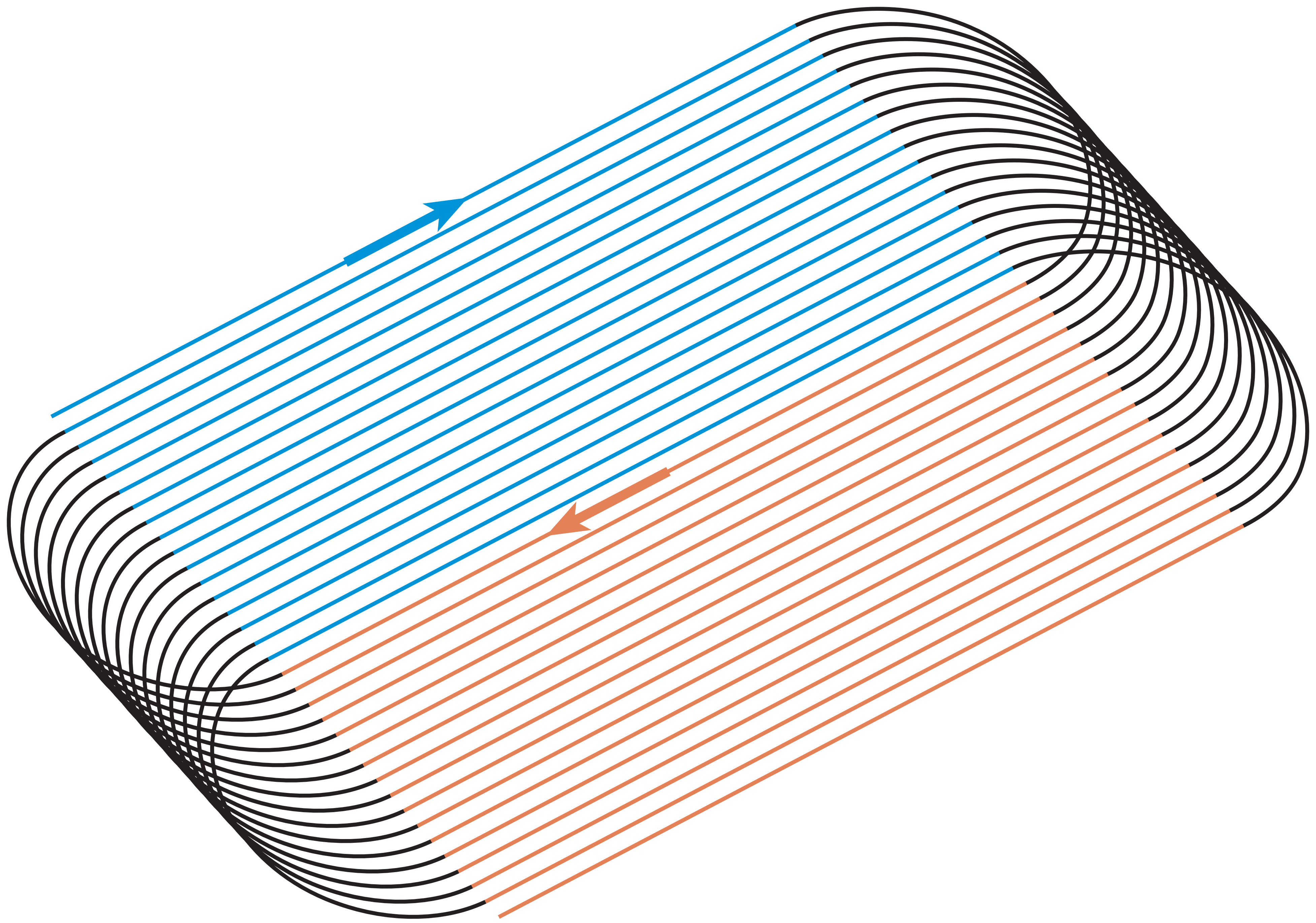
where $\|\cdot\|_F^2$ is sum of squares of all entries

- choose k *explicitly* & avoid costly *SVD's*

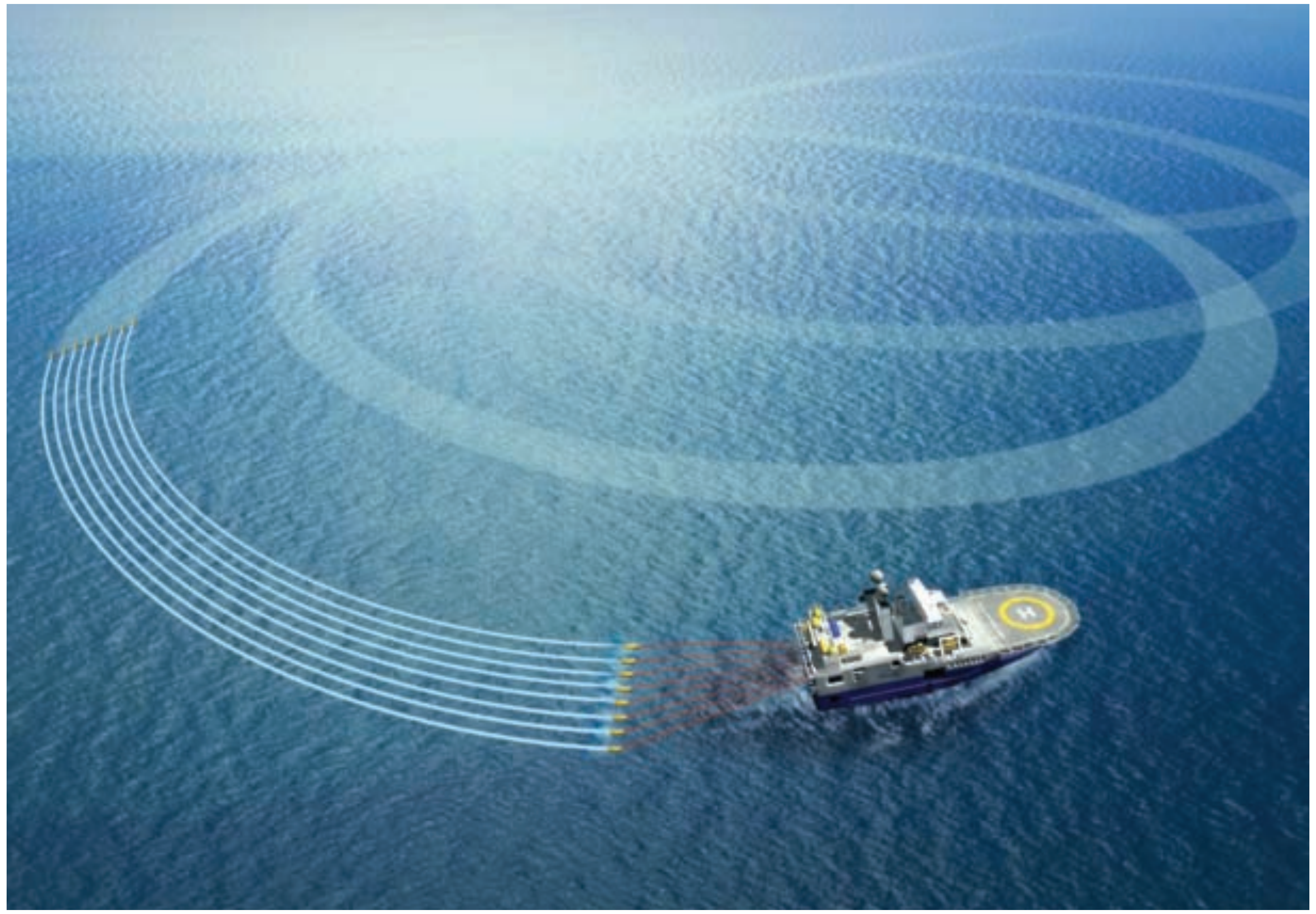
Survey information — coil acquisition

Old vs new

Conventional acquisition

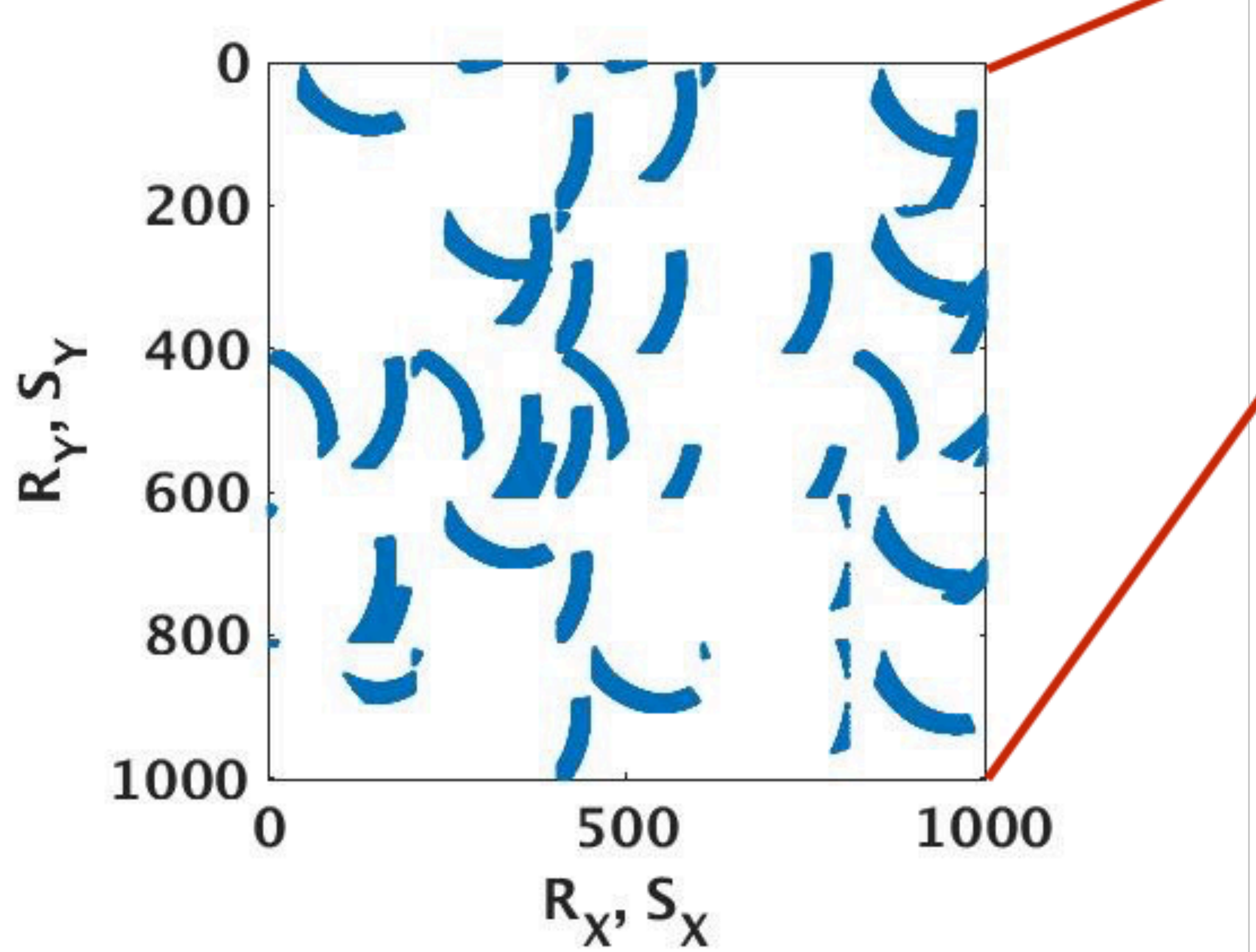


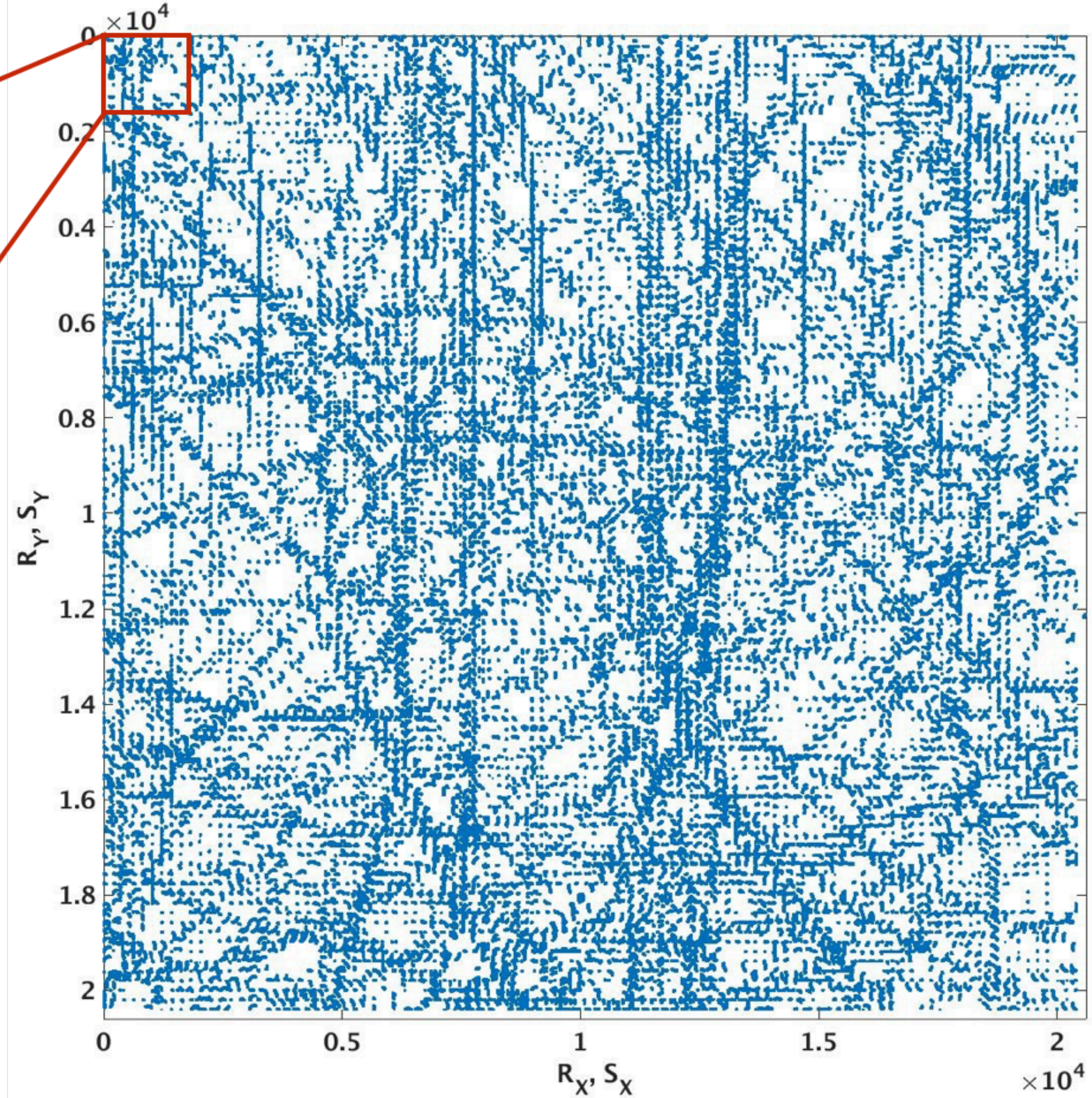
random coil acquisition



from https://www.slb.com/~/media/Files/resources/oilfield_review/ors08/aut08/shooting_seismic_surveys_in_circles.pdf

Acquisition mask – non-canonical matrix (10 x10 km)





Observed sampling = 16%
Effective sampling = 4%
Spectral Gap = 0.4

Acquisition information

3D overthrust model, 5km x 12km x 12km

10404 sources @ 100m

40804 receivers @ 50m

Time length : 3 seconds @ 0.004s

Interpolation from 1-50 Hz

Acquisition information

3D overthrust model, 5km x 12km x 12km

10404 sources @ 100m

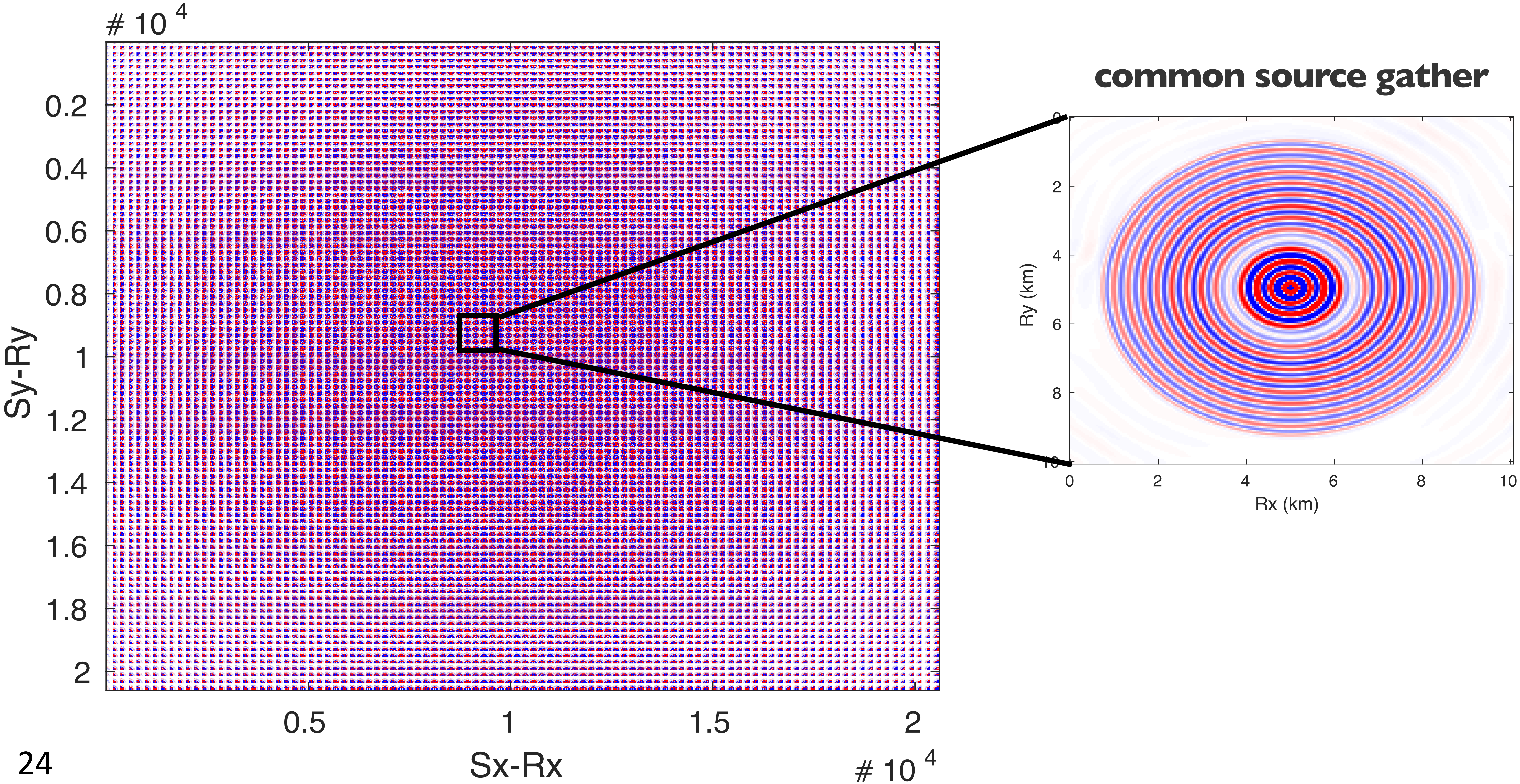
Unknown 20k X 20k matrix for each frequency!

40804 receivers @ 50m

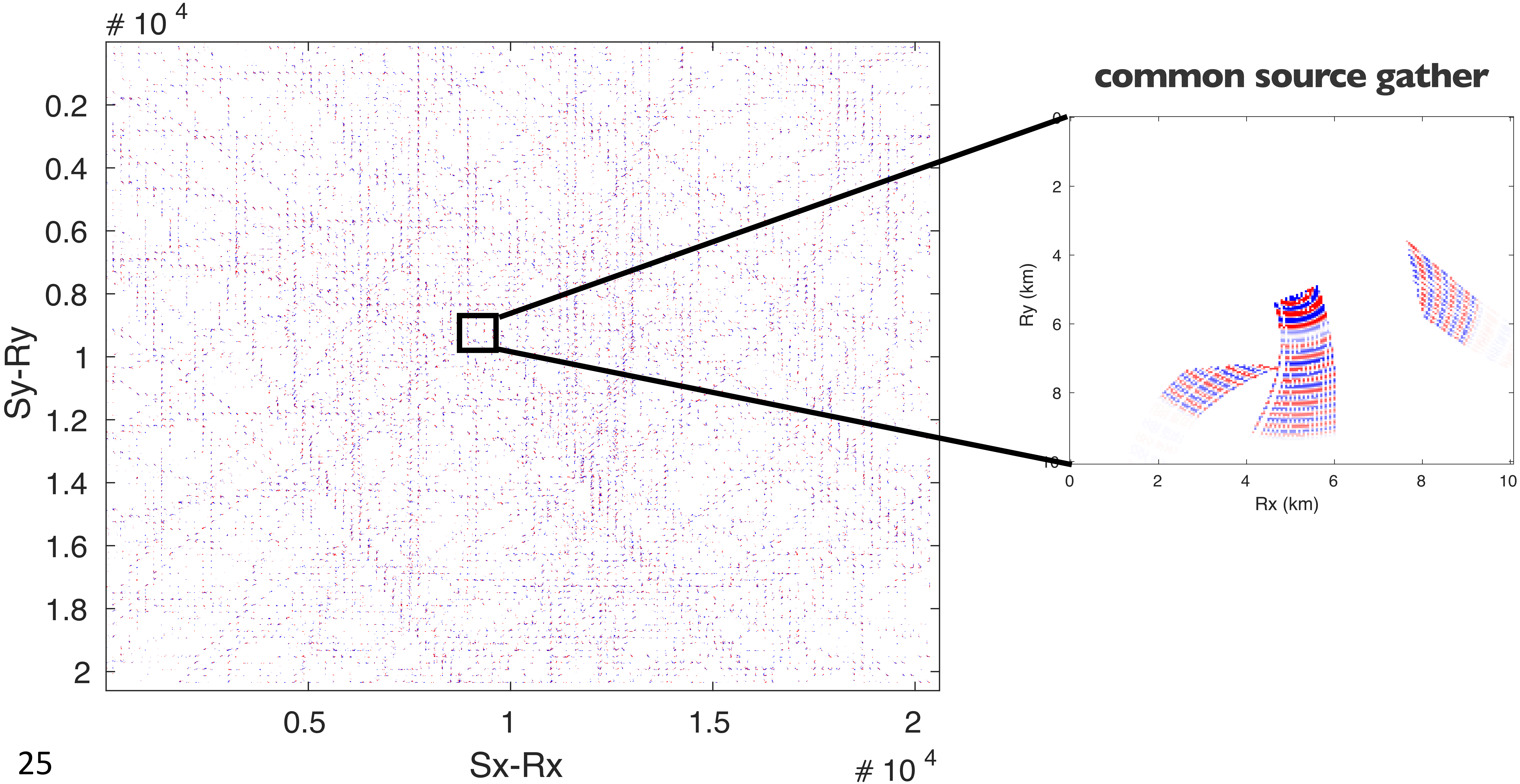
Time length : 3 seconds @ 0.004s

Interpolation from 1-50 Hz

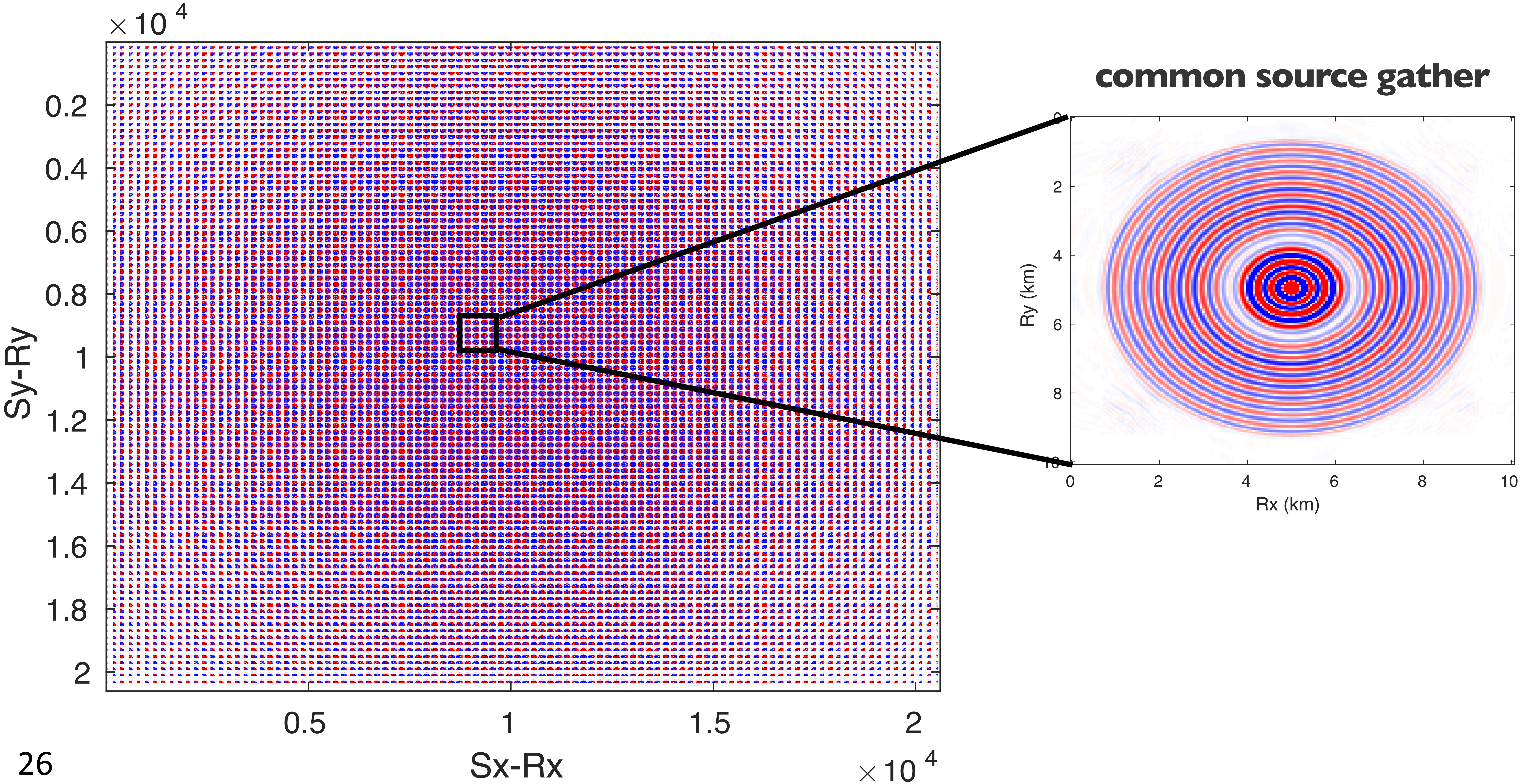
Frequency slice @ 7Hz
ground truth



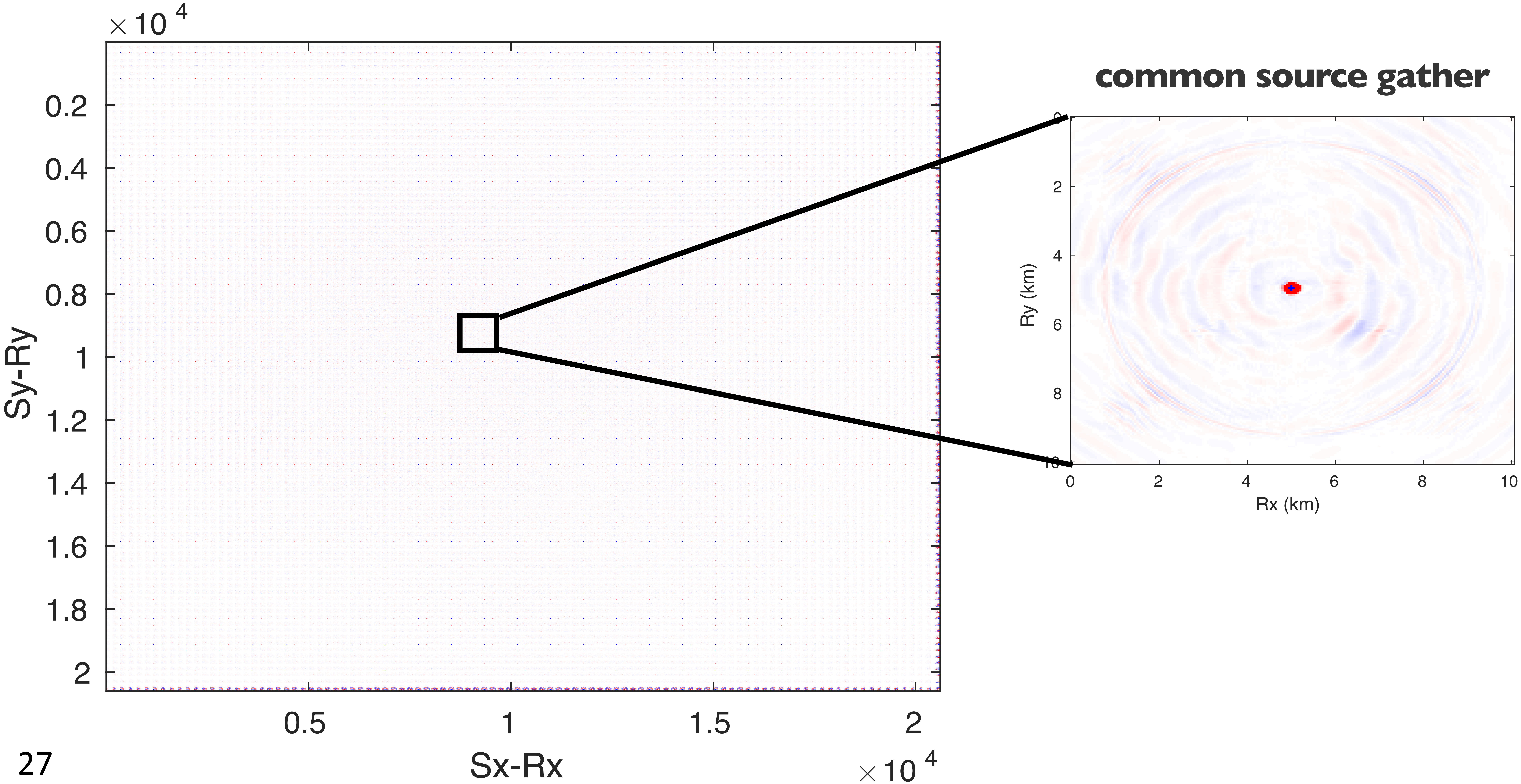
Frequency slice @ 7Hz
observed



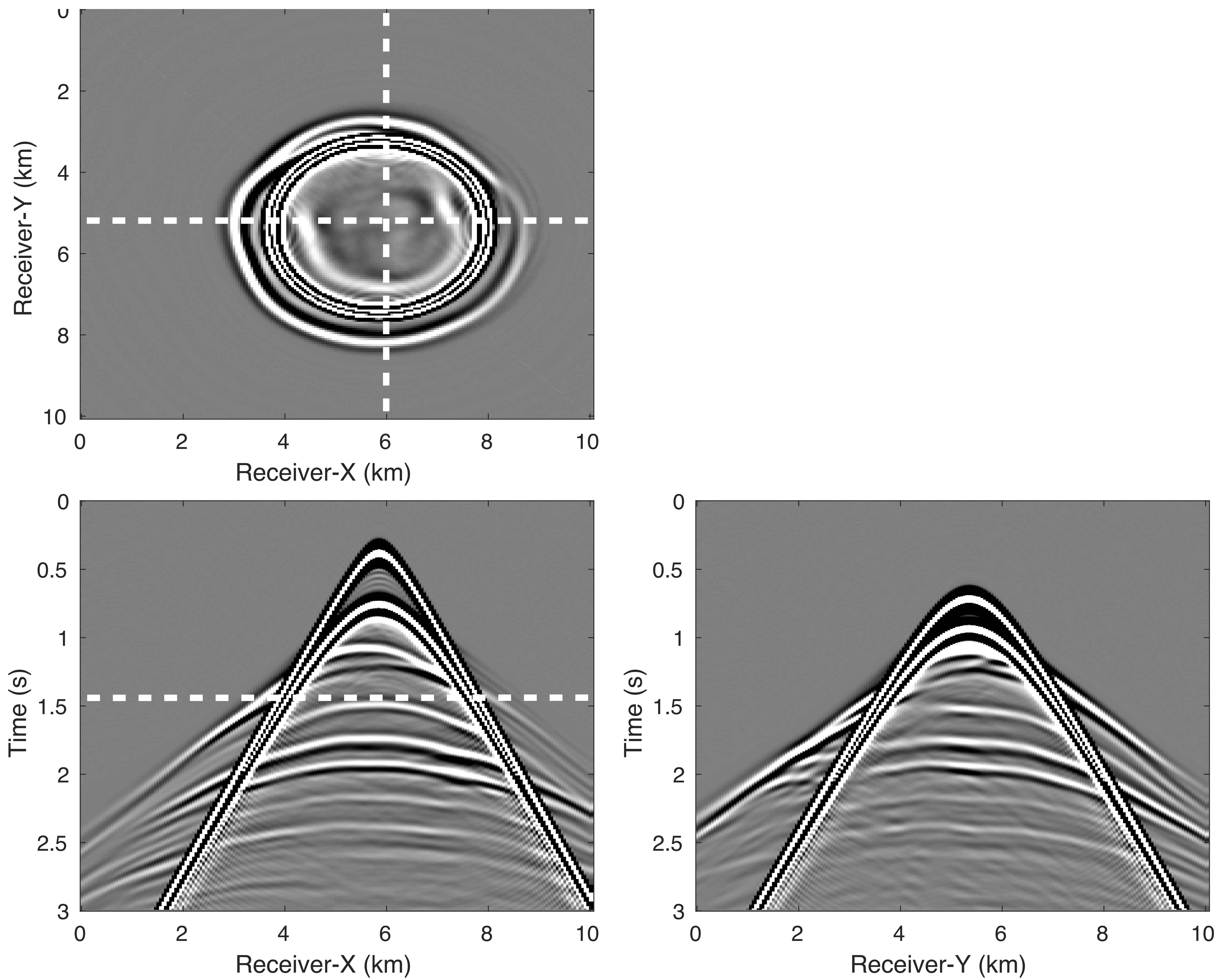
Frequency slice @ 7Hz
interpolated



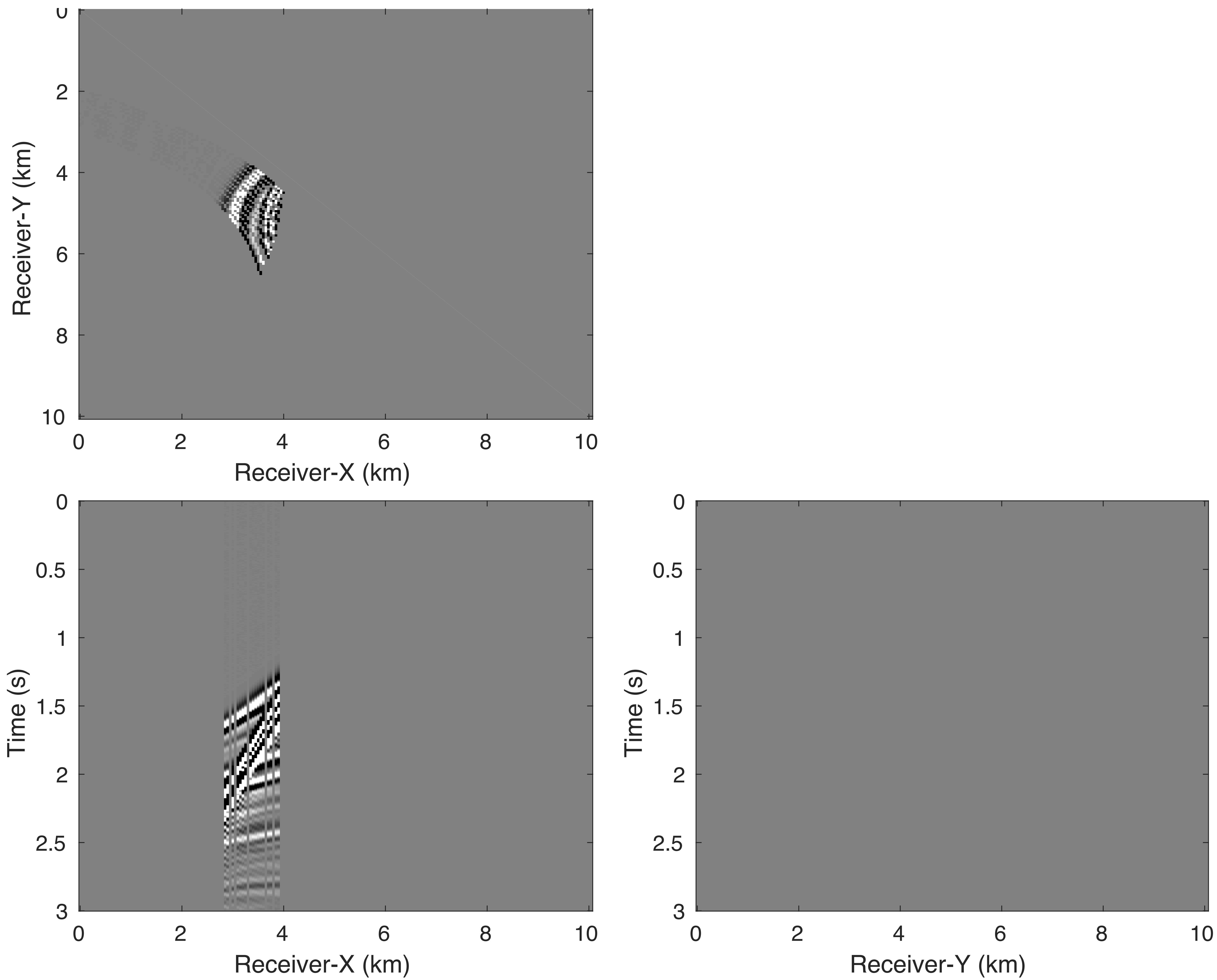
Frequency slice @ 7Hz
residual



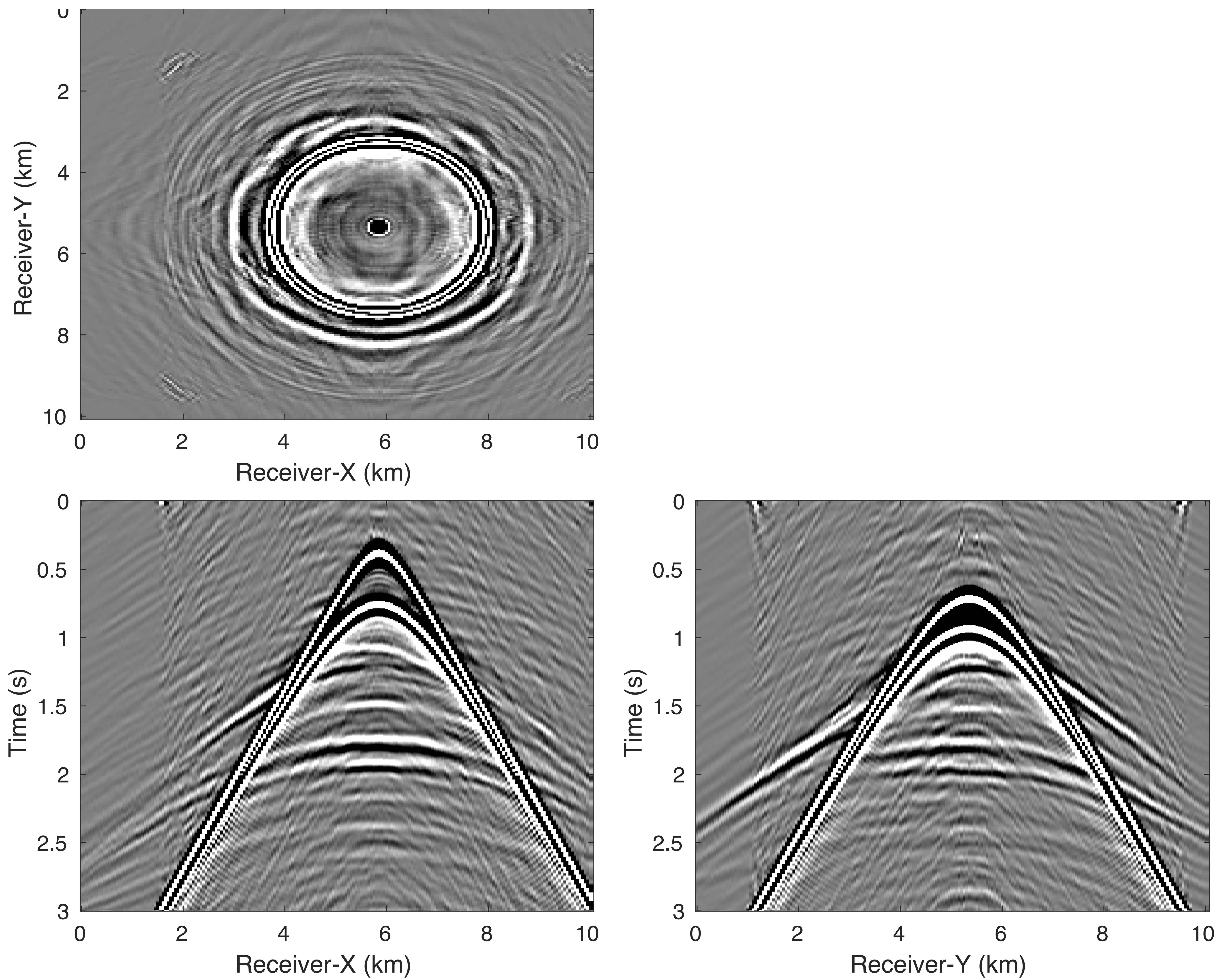
Common source gather
ground truth



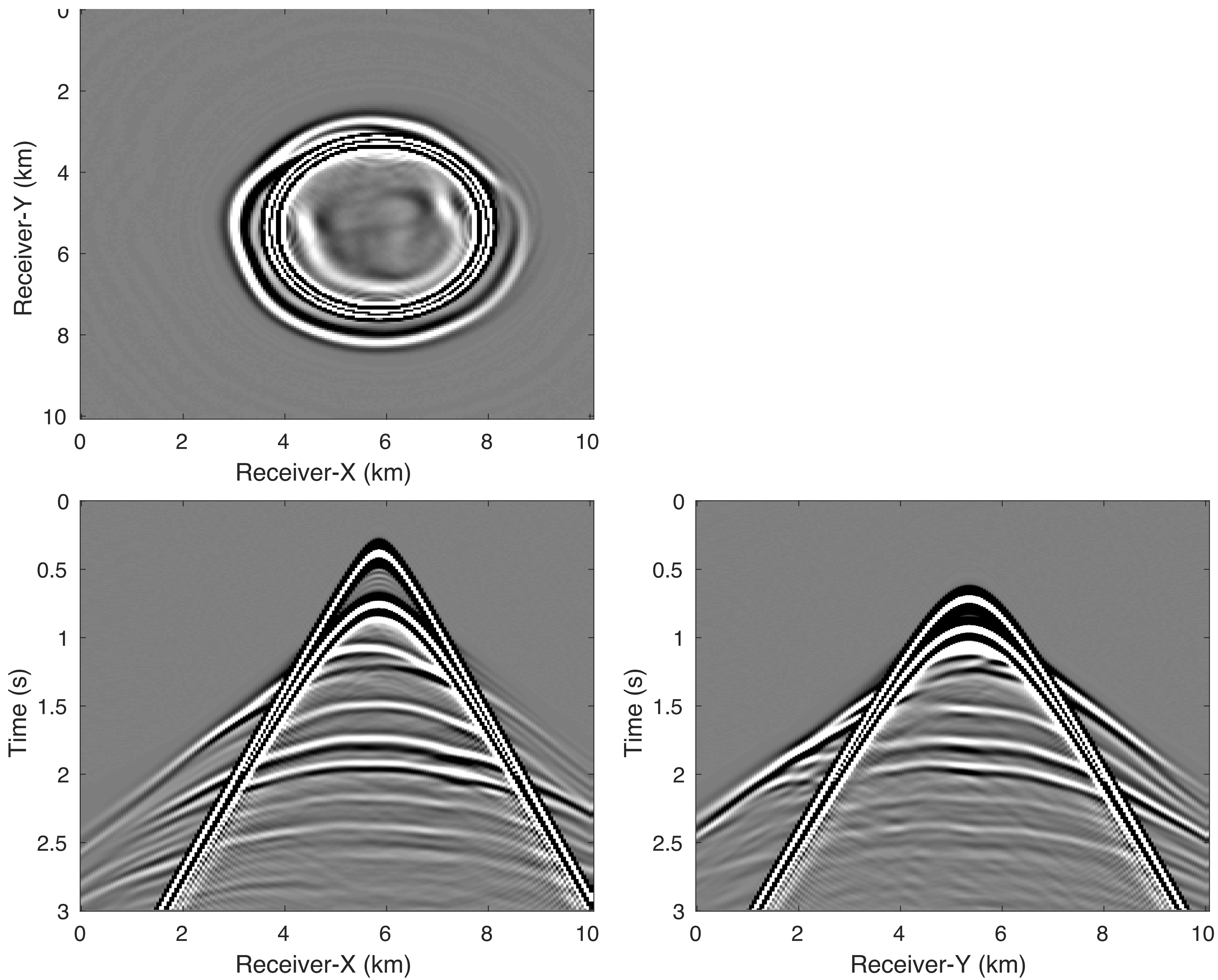
Common source gather
subsampled



Common source gather
interpolated



Common source gather
ground truth



Computational & memory advantages

Size of fully sampled interpolated volume : 2.5 TB

Save only low-rank factors

- compression rate: **99.5%**
- size of final compressed 5D seismic volume : **15GB**

Non-canonical vs. canonical – 396 x 396 x 50 x 50 volume (~5.8 GB)

	Frequency (Hz)	Parameter Size	SNR	Compression Ratio
Non-canonical	3	71MB	42.8	98.8%
canonical	3	501MB	42.9	91.6%
Non-canonical	6	421MB	43.0	92.9%
canonical	6	1194MB	43.1	79.9%

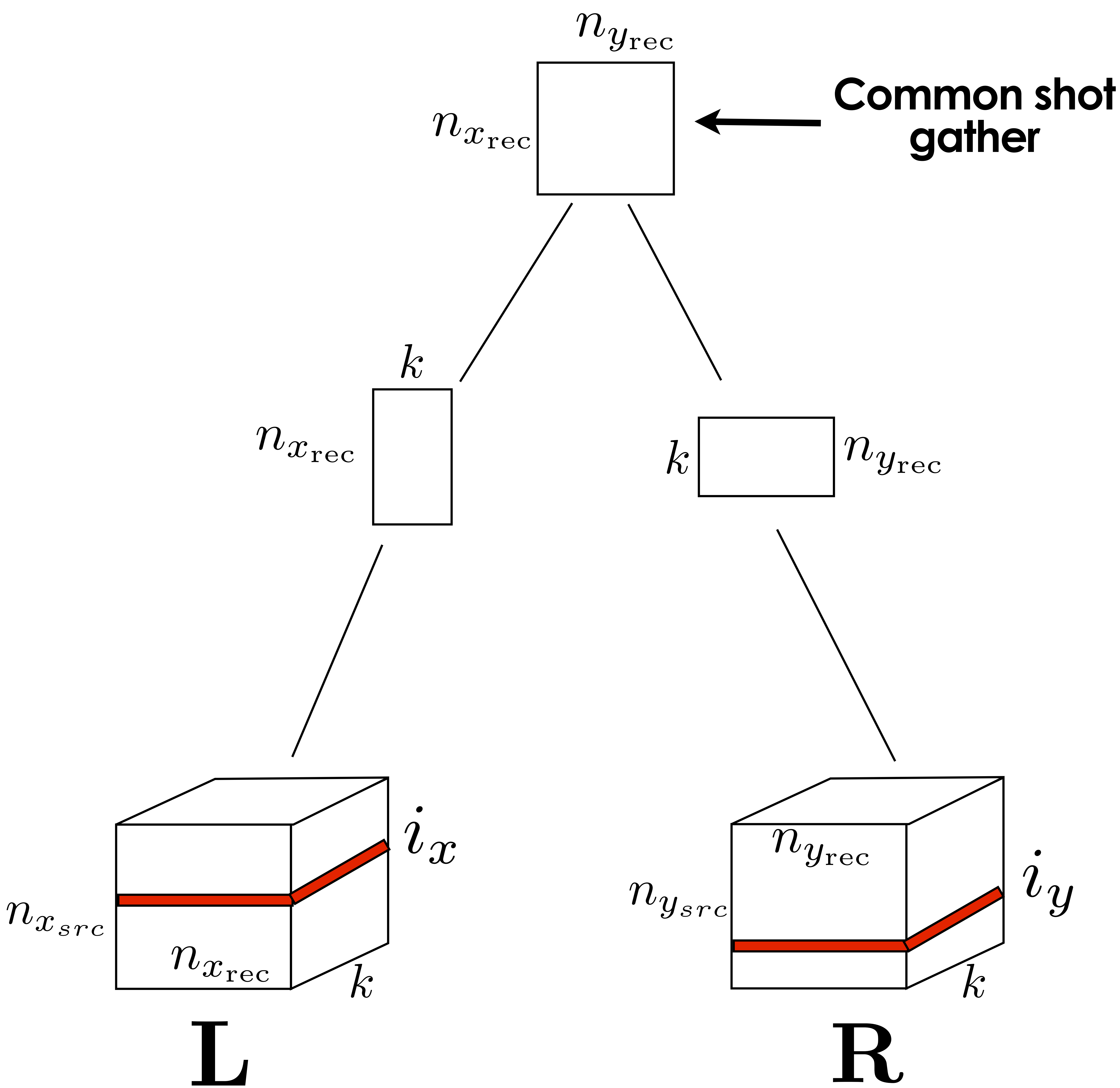
Non-canonical vs. Nyquist – 396 x 396 x 50 x 50 volume (~5.8 GB)

	Frequency (Hz)	Compression Ratio
Non-canonical	3	98.8%
Nyquist $\theta = 45^o, V = 1500\ m/s$	3	89%
Non-canonical	6	92.9%
Nyquist $\theta = 45^o, V = 1500\ m/s$	6	0 %

Nyquist Criteria : $\Delta x \leq \dfrac{V}{4f\sin(\theta)}$

On-the-fly extraction

- i_x, i_y: Common source index number



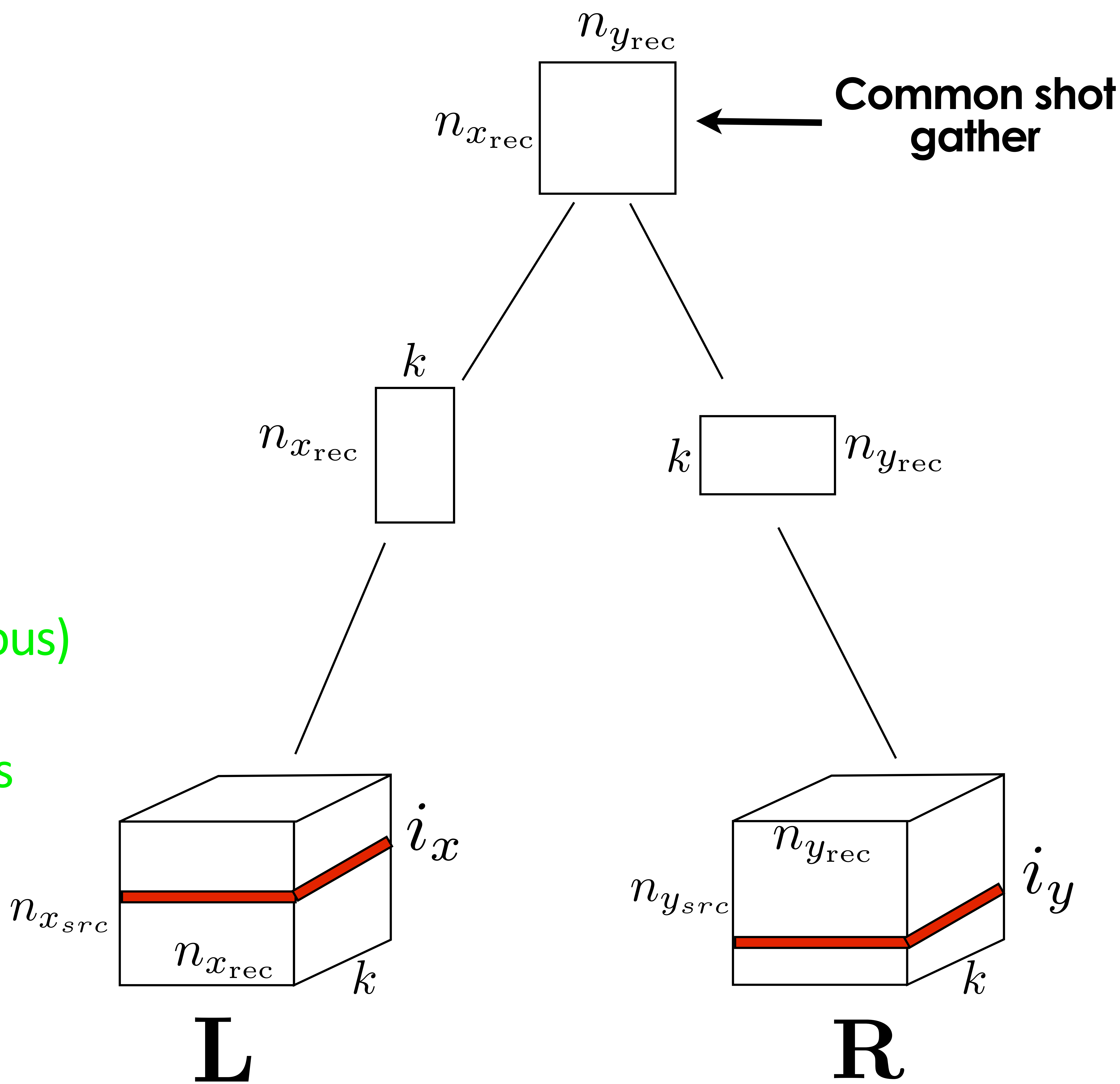
On-the-fly extraction

SLIM

i_x, i_y : Common source index number

Able to extract (simultaneous)

- common source gathers
- common receiver gathers



Observations

Seismic surface data is highly redundant

- exhibits low-rank structure in proper permutation
- low-rank structure can only be observed w/o working in small windows

Parallel scalable algorithms are available that work on real data

- source experiments can be generated on the fly

Instance of true multi-azimuth processing

Observations

Seismic surface data is highly redundant

- exhibits low-rank structure in proper permutation
- low-rank structure can only be observed w/o working in small windows

Parallel scalable algorithms are available that work on real data

- source experiments can be generated on the fly

Instance of true multi-azimuth processing

Compression is remarkable despite inherent oversampling...

SLIM

Observations

Seismic surface data is highly redundant

- exhibits low-rank structure in proper permutation
- low-rank structure can only be observed w/o working in small windows

Parallel scalable algorithms are available that work on real data

- source experiments can be generated on the fly

Instance of true multi-azimuth processing

SLIM

Observations

Seismic surface data is highly redundant

- exhibits low-rank structure in proper permutation
- low-rank structure can only be observed w/o working in small windows

Parallel scalable algorithms are available that work on real data

- source experiments can be generated on the fly

Instance of true multi-azimuth processing

Attained compression will be a game changer in how we handle data during inversion.

Low-rank representation of omnidirectional subsurface extended image volumes

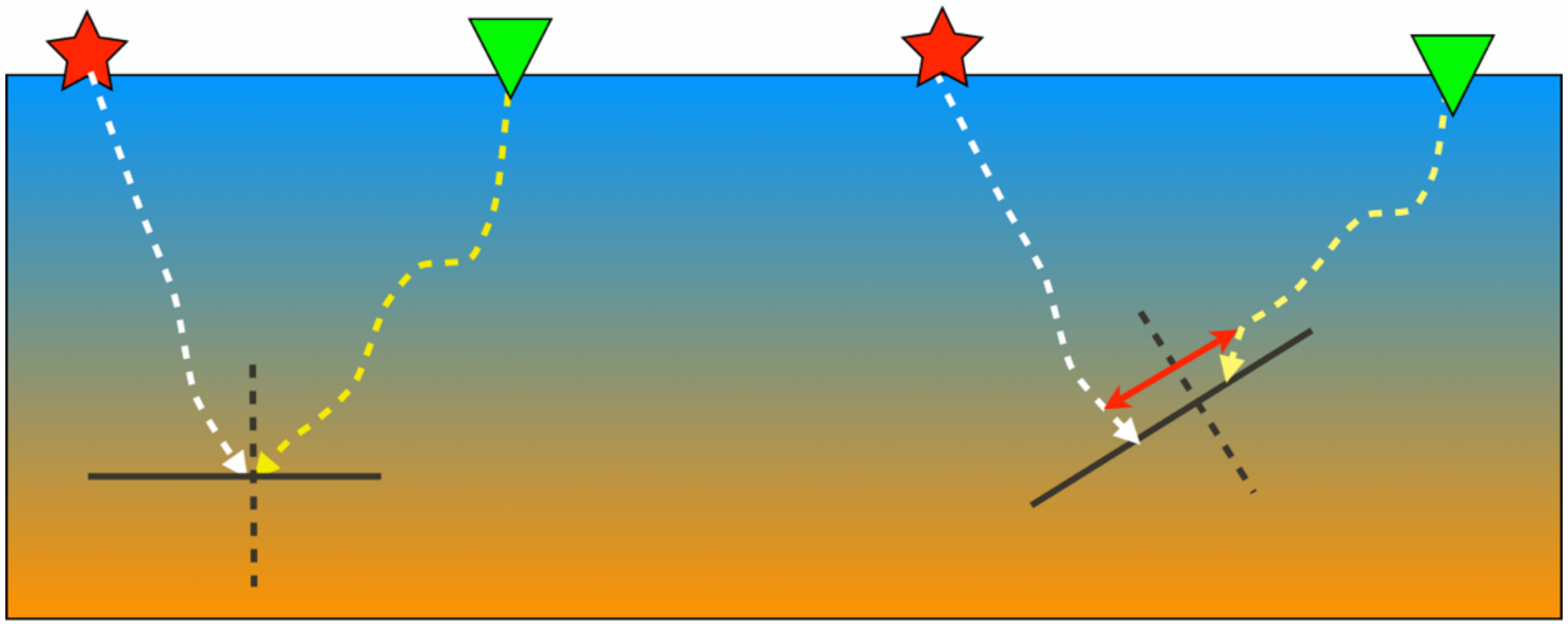
Marie Graff-Kray, Rajiv Kumar and Felix J. Herrmann



SLIM
University of British Columbia

Seismic imaging

- Forward propagate source wavefields
- Back propagate receiver wavefields
- Cross-correlate wavefields at subsurface locations

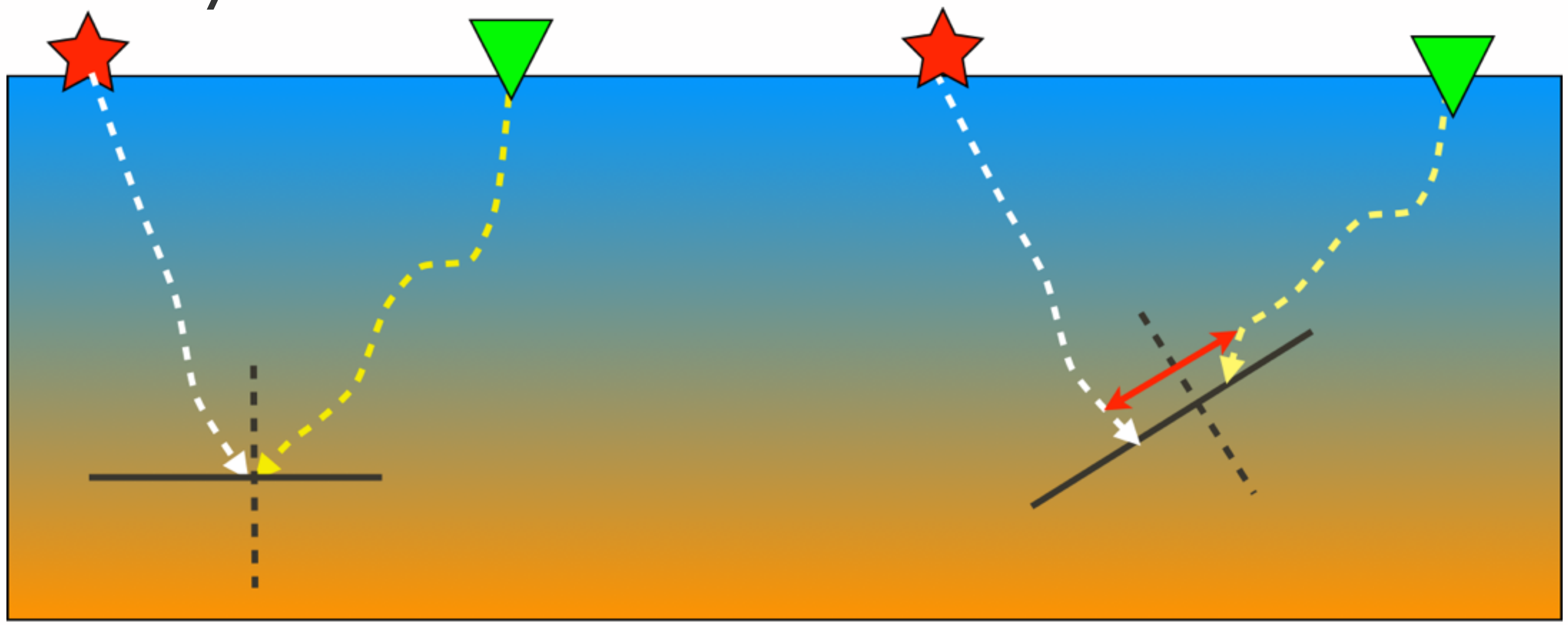


Zero offset (migration)

All offsets (Extended image volume)

Seismic imaging w/ extensions

- Conventional imaging extracts zero-offset section only
- Extension/lifting corresponds to new experiment w/ sources/receivers anywhere in subsurface
- Near isometry

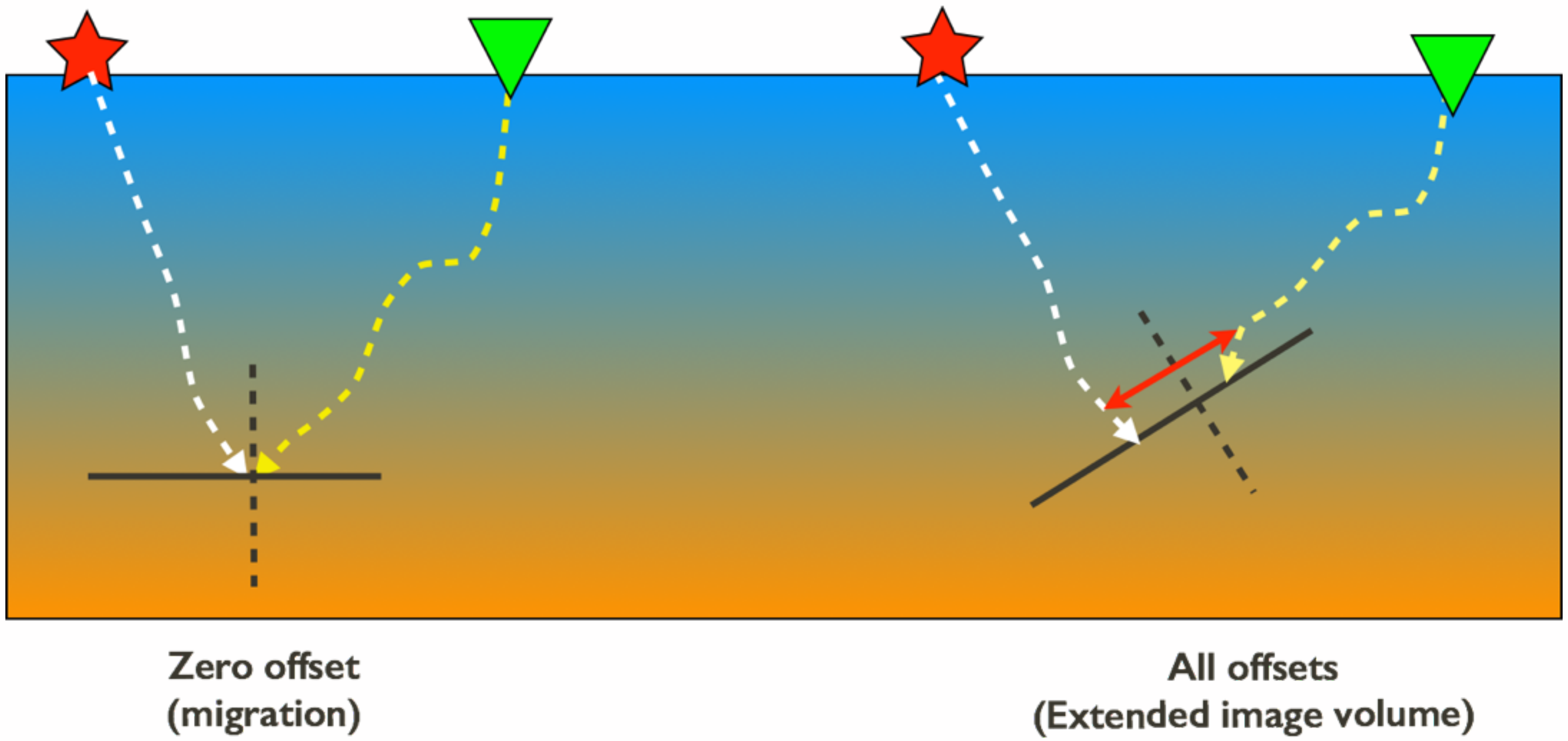


Zero offset (migration)

All offsets (Extended image volume)

Seismic imaging w/ extensions

- Parametrized by subsurface horizontal offset or angles
- Computed & stored for small subsets of offsets/angles
- Do not explore underlying low-rank structure

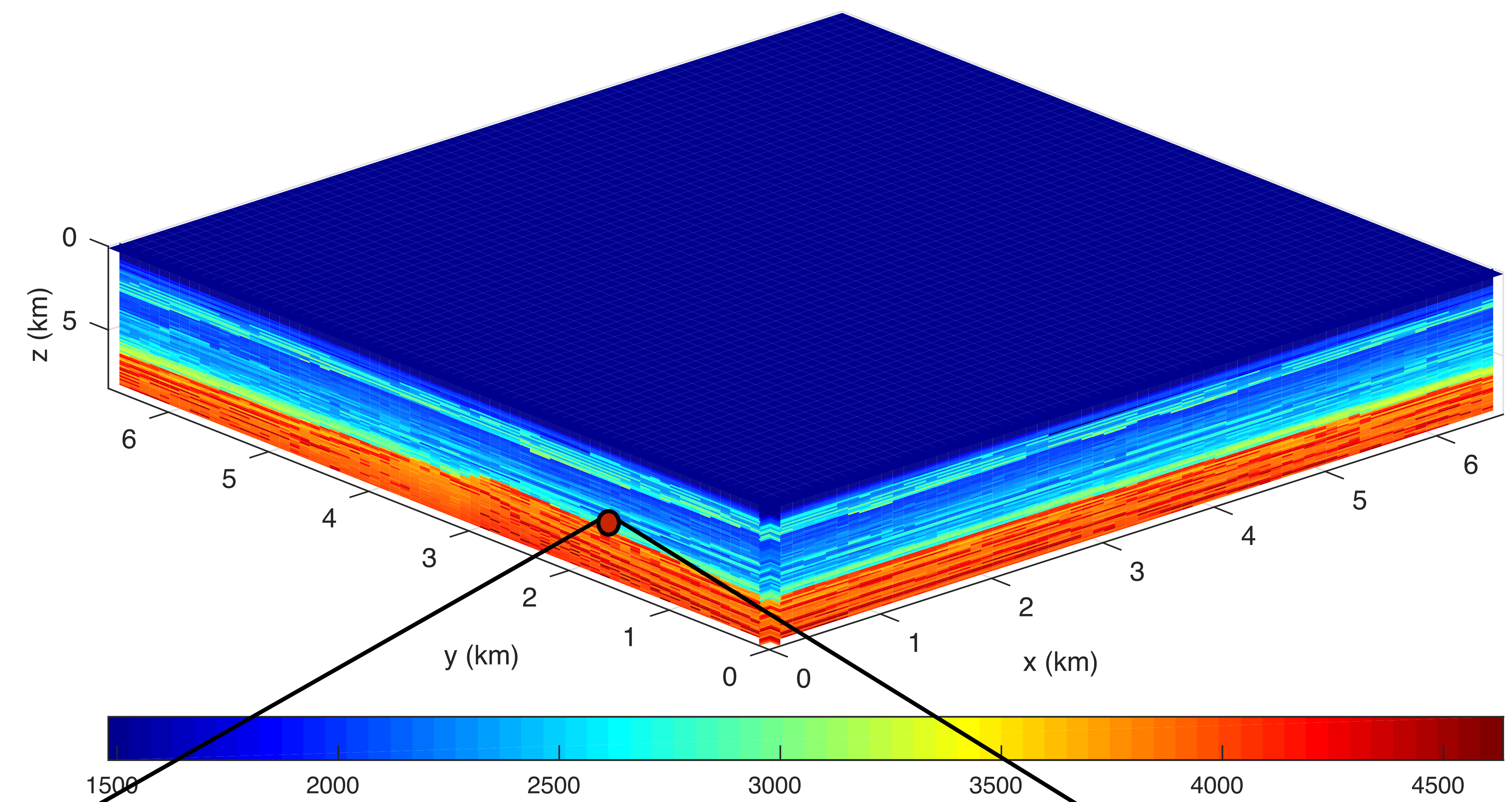


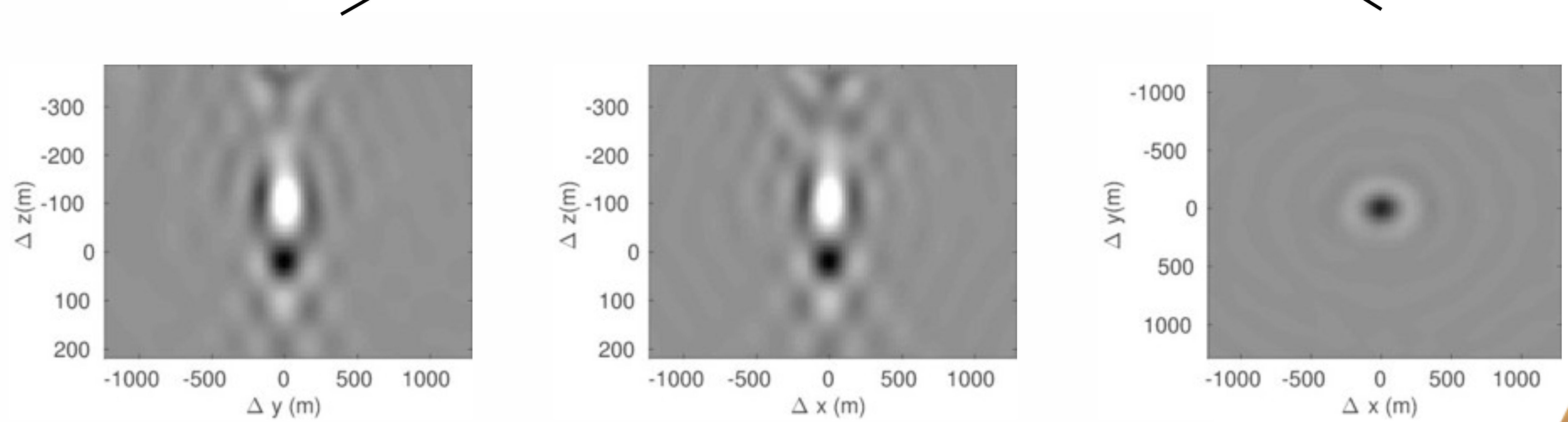
Extended images: challenges

- use *all* subsurface offsets (6D volume for 3D model)
- 2-way wave-equation

but…. we can never hope to compute or store such an image volume!

Can we work with these volumes *implicitly*?





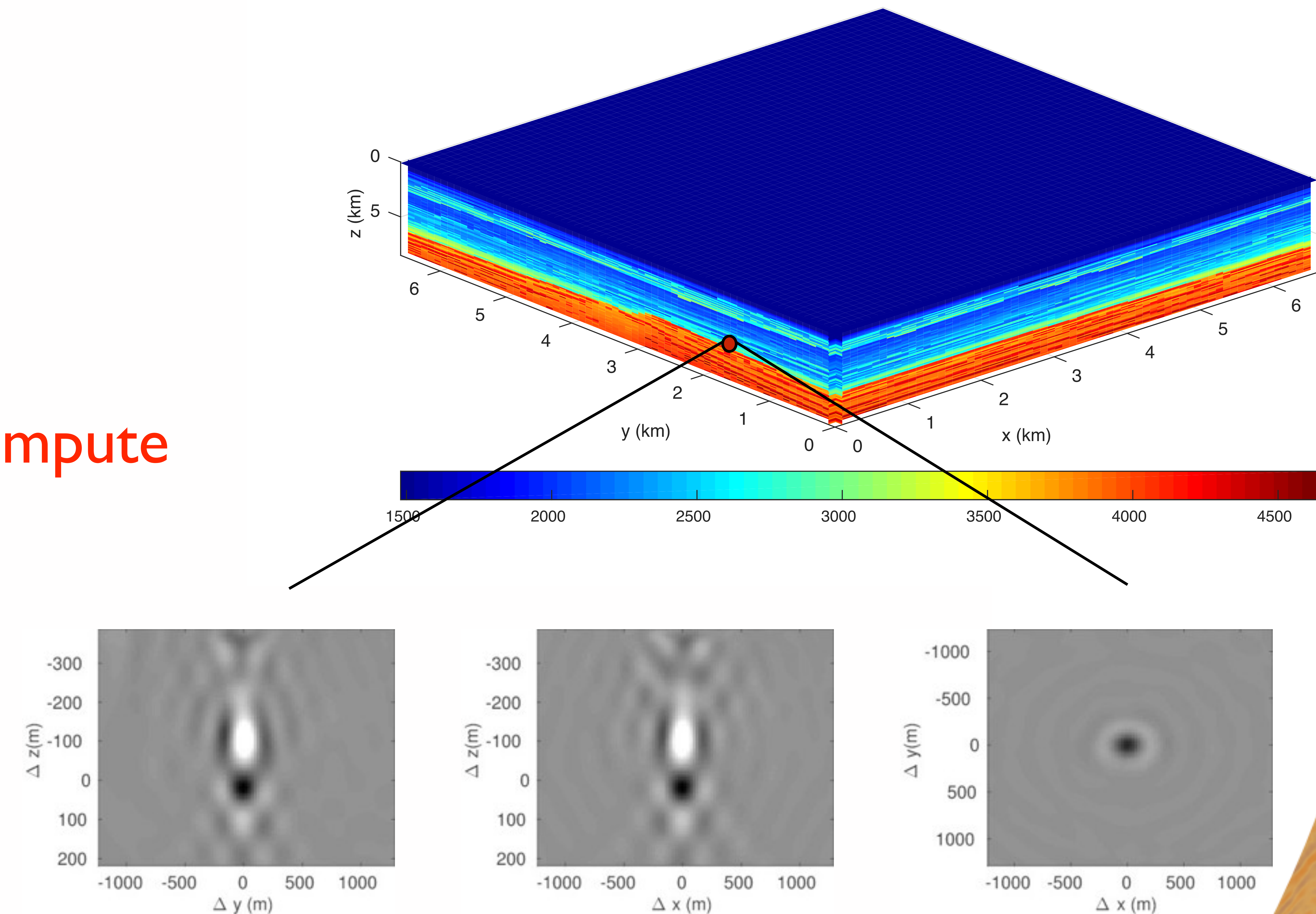
Extended images: challenges

- use *all* subsurface offsets (6D volume for 3D model)
- 2-way wave-equation

but…. we can never hope to compute or store such an image volume!

quadratic in image size

Can we work with these volumes *implicitly*?



SLIM

When "the dream" comes true

Computation of full-subsurface offset volumes is prohibitively expensive in 3D (storage & computation time)

Past

Can **not** form full *E* **but** *action* on (random) vectors allows us to get information from *all* or *subsets* of *subsurface points*

When “the dream” comes true

Computation of full-subsurface offset volumes is prohibitively expensive in 3D (storage & computation time)

Past

Can **not** form full *E* **but** *action* on (random) vectors allows us to get information from *all* or *subsets* of *subsurface points*

Present

Can ~~**not**~~ form full *E* **using** *action* on (random) vectors allows us to get information from *all* or *subsets* of *subsurface points*

Efficient ways to extract information from highly compressed image volumes

Extended images via probing

Extended images

Given two-way wave equations, source & receiver wavefields are defined as

$$H(\mathbf{m})U = P_s^T Q$$
$$H(\mathbf{m})^* V = P_r^T D$$

where

$H(\mathbf{m})$: discretization of the Helmholtz operator

Q : source

D : data matrix

P_s, P_r : samples the wavefield at the source and receiver positions

$\mathbf{m}$: slowness

Extended images

Organize wavefields in monochromatic data *matrices* where each *column* represents a *common* shot gather

Express image volume *tensor* for *single* frequency as a *matrix*

$$E = VU^*$$

SLIM

Extended images – in the past

Too expensive to compute *(storage & computational time)*

Instead, *probe* volume with *tall* matrix $W = [\mathbf{w}_1, \ldots, \mathbf{w}_\ell]$

$$\widetilde{E} = EW = H^{-*} P_r^\top D Q^* P_s H^{-*} W$$

where $\mathbf{w}_i = [0, \ldots, 0, 1, 0, \ldots, 0]$ represents *single* scattering points

Tristan van Leeuwen, Rajiv Kumar, and Felix J. Herrmann, "Enabling affordable omnidirectional subsurface extended image volumes via probing",
Geophysical Prospecting, 2016

SLIM

Extended images – at present

Too expensive to compute *(storage & computational time)*

Instead, *probe* volume with *tall* matrix $W = [\mathbf{w}_1, \dots, \mathbf{w}_\ell]$

$$\widetilde{E} = EW = H^{-*} P_r^\top D Q^* P_s H^{-*} W$$

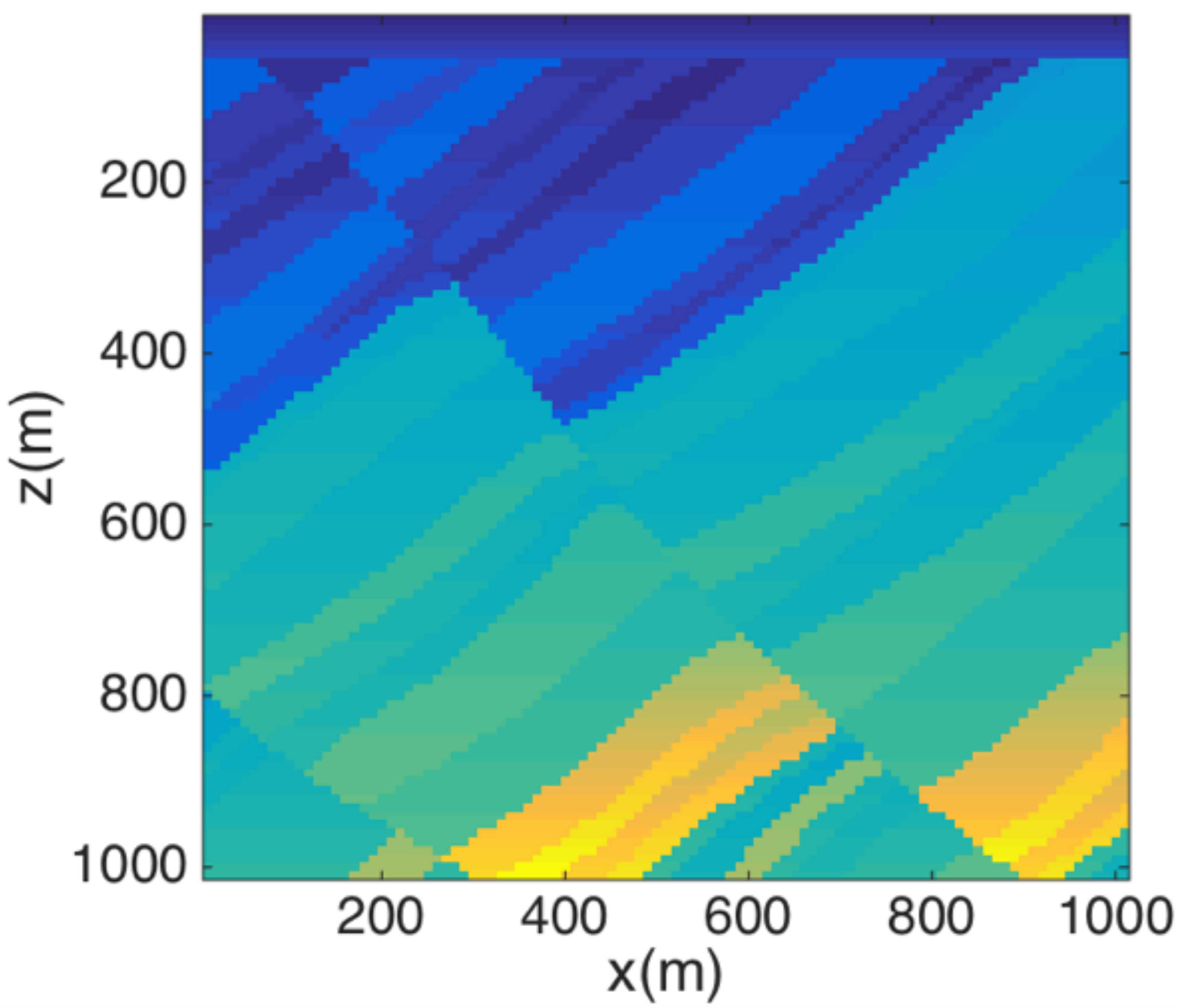
where $\mathbf{w}_i = [0, \dots, 0, 1, 0, \dots, 0]$ represents *single* scattering points

Other choice for W ? And how many vectors are needed ? for example:

- random (Gaussian or Rademacher) vectors
- singular vectors from (randomized) SVD

Low-rank representation (5 Hz)

SVD on monochromatic extended image volumes



Model (101x101)

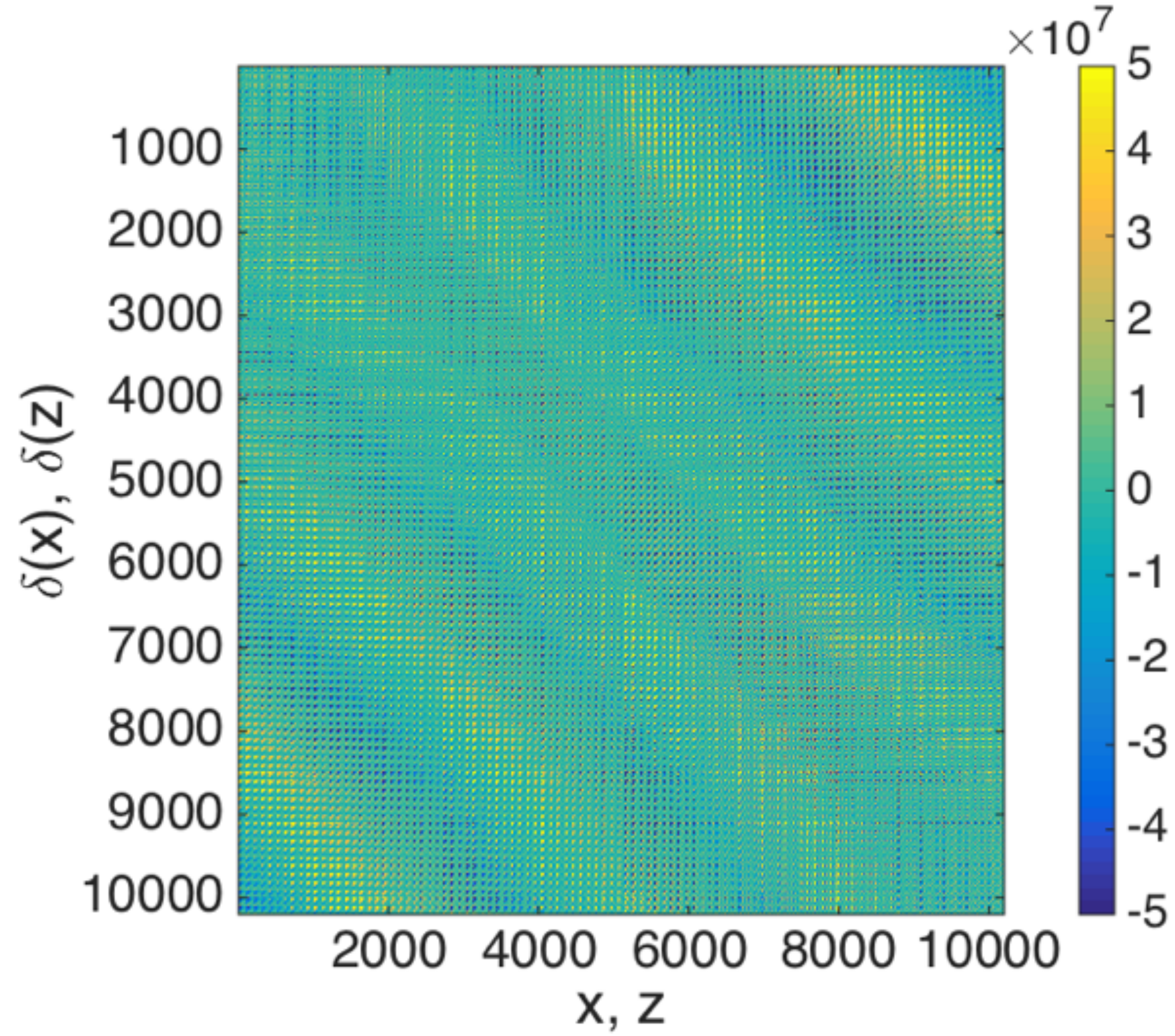
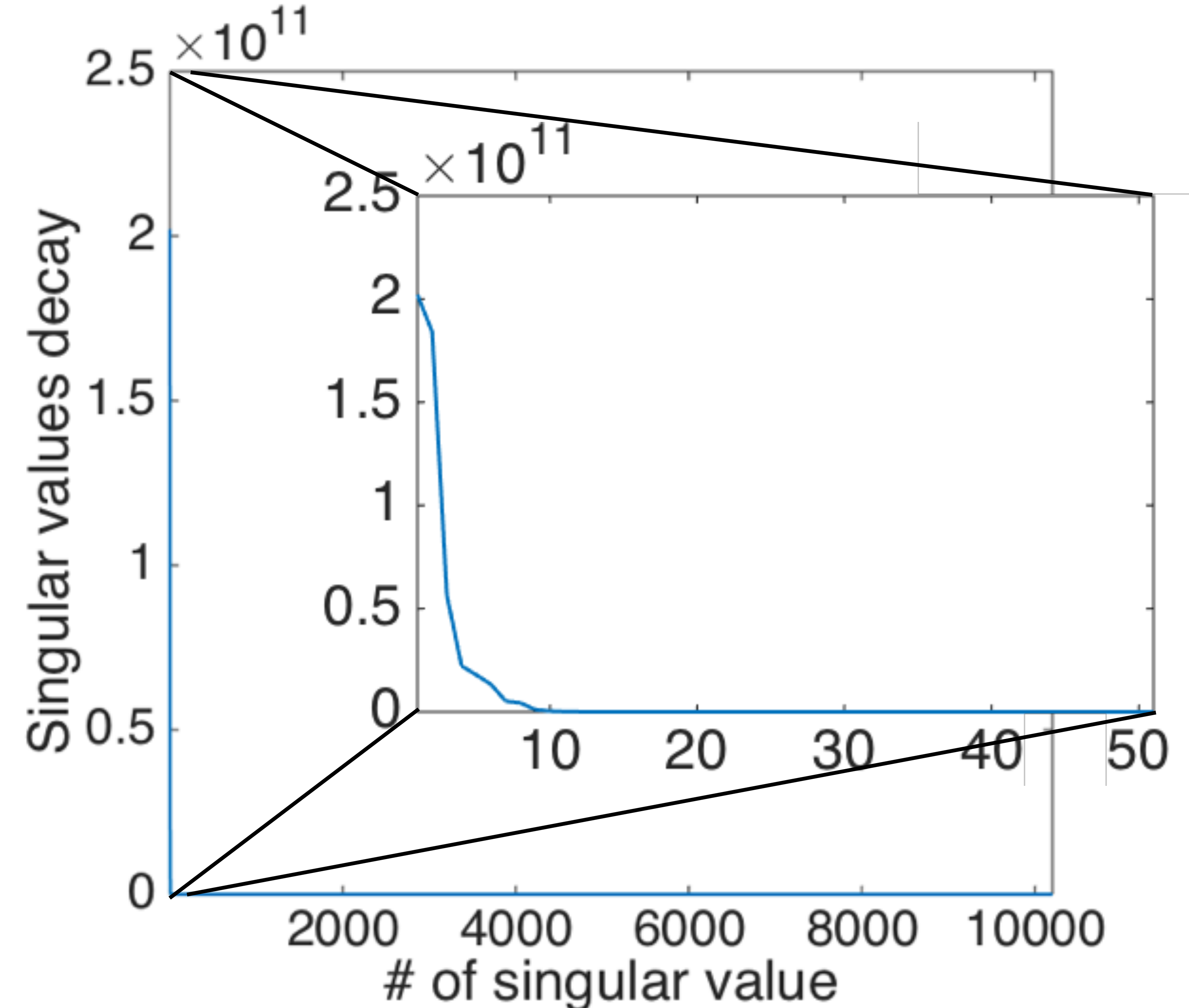


Image Volume (IV)



Singular Values of IV

Saturday, November 11, 17

Rank of the extended image volume

From the formula

$$\widetilde{E} = EW = H^{-*} P_r^{\top} D Q^* P_s H^{-*} W$$

the rank of E is given by the rank of the data matrix D

So, we take r probing vector $W = [w_1, \ldots, w_r]$

— random +1/-1 with probability 0.5

— Gaussian random with 0 mean & variance 1

— our contribution: orthogonal basis of the range of E

Representation of the extended image

From the formula

$$\widetilde{E} = EW = H^{-*} P_r^\top D Q^* P_s H^{-*} W$$

where $W = [w_1, \ldots, w_r]$ are Gaussian random vectors

Our representation consists of building an **orthogonal** basis Q of the range of E

such that Q is the r first columns of Q-matrix of the QR-factorization of $\tilde{E} = EW$

Notation: $[Q, EQ]$

[1] Halko et. al, Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions, 2010

[2] Bekas et. al, An Estimator for the Diagonal of a Matrix, 2007

Representation of the extended image

From

$$\boxed{[Q, EQ]}$$

we want to **extract information** about E (diagonal, columns, off-diagonals...)

Two possible ways to do it:

1. using the *randomized SVD algorithm* [1]

   (actually only steps 4 and 5, see next slide)

2. using the *randomized (off) diagonal extraction formula* [2]

   (or any other diagonal of E thanks to a permutation matrix P)

SLIM

1. Randomized SVD algorithm

Original algorithm from [1]:

1. $Y = EW$ probe full extended image volume with virtual sources
2. $[Q, R] = \mathrm{qr}(Y)$ QR factorization
3. $Z = Q^* E$ probe again with new virtual sources
4. $[U, S, V] = \mathrm{svd}(Z)$ SVD factorization (first few singular values)
5. $U \leftarrow QU$ update left singular vectors

For us, steps 1 to 3 are given by $[Q, EQ]$ by probing only from the right

if doing so, step 5 becomes an update of right singular vectors: $V \leftarrow QV$

Finally

$$E \simeq USV^*$$

SLIM

2. Randomized diagonal extraction

Original formula from [2]:

$$\mathrm{diag}(E) \approx \left(\sum_{i=1}^{\ell} w_i \odot (E w_i)\right) \oslash \left(\sum_{i=1}^{\ell} w_i \odot w_i\right)$$

for $W = [\mathbf{w}_1, \ldots, \mathbf{w}_\ell]$, +1/-1 with probability 0.5 random vectors and $\ell \gg N$ (too expensive)

With an orthogonal basis Q:

$$\boxed{\mathrm{diag}(E) = \sum_{i=1}^{r} q_i \odot (E q_i)}$$

Our contribution: take only r vectors spanning an orthogonal basis of the range of E (exact if r is the rank of E)

2. Randomized part extraction

For the diagonal:

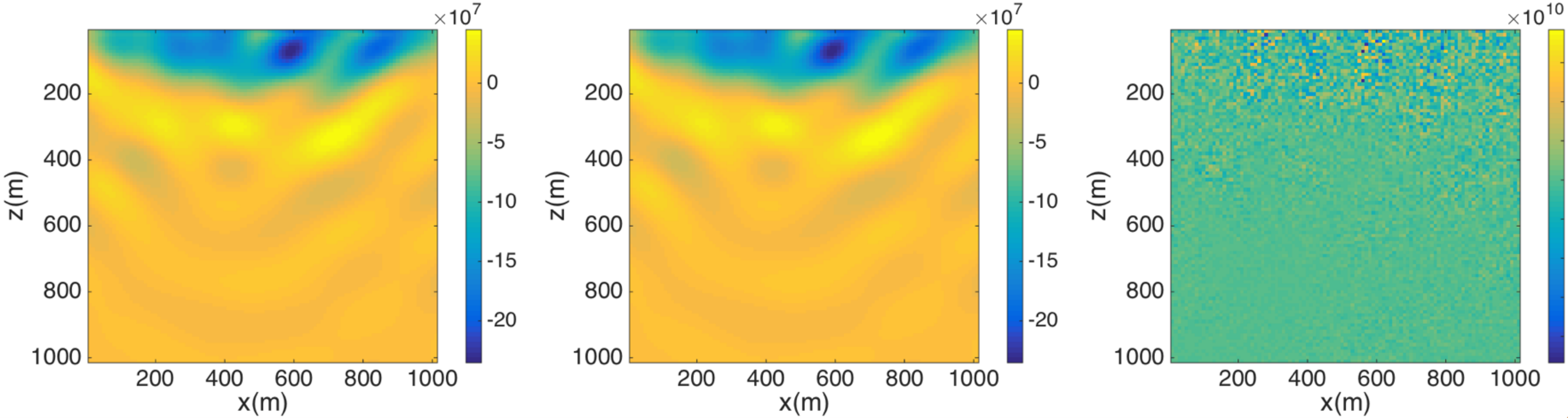
$$\boxed{\mathrm{diag}(E) = \sum_{i=1}^{r} q_i \odot (E q_i)}$$

For another diagonal, let P be a permutation matrix

$$\mathrm{offdiag}(E) = \sum_{i=1}^{r} (P q_i) \odot (E q_i)$$

Orthogonal basis vs random basis

Diagonal extraction of the EIV for different representation (5 Hz, r = 15)



full EIV

E

with orthogonal basis

[Q,EQ]

with Gaussian basis

[W,EW]

Invariance formula for EIVs

Tristan van Leeuwen and Felix J. Herrmann, "**Wave-equation extended images: computation and velocity continuation**", in *EAGE Annual Conference Proceedings*, 2012.

SLIM

Invariance formulation for EIVs...

For monochromatic data and sources

$$E = H[m]^{-*} \underbrace{P_r^\top D Q^* P_s}_{invariant} H[m]^{-*}$$

then for two models m_1 and m_2

$$H[m_1]^* E_1 H[m_1]^* = H[m_2]^* E_2 H[m_2]^*$$

Invariance formulation for EIVs...

For monochromatic data and sources

$$E = H[m]^{-*} \underbrace{P_r^\top D Q^* P_s}_{invariant} H[m]^{-*}$$

then for two models m_1 and m_2

$$H[m_1]^* E_1 H[m_1]^* = H[m_2]^* E_2 H[m_2]^*$$

we deduce E_2 from E_1

$$E_2 = H[m_2]^{-*} H[m_1]^* E_1 H[m_1]^* H[m_2]^{-*}$$

Only $2r$ PDEs solves!

...from Low-Rank representation

From $[Q_1, E_1 Q_1]$, we get a low-rank formulation for E_1

$$E_1 = L_1 R_1^*$$

with L_1 and R_1 two $N \times r$ matrices given by

$$L_1 = U_1 \sqrt{S_1}$$

$$R_1 = V_1 \sqrt{S_1}$$

$[U_1, S_1, V_1]$ from randomized SVD

New extended image

Now we deduce

$$L_2 = H[m_2]^{-*} H[m_1]^* L_1$$

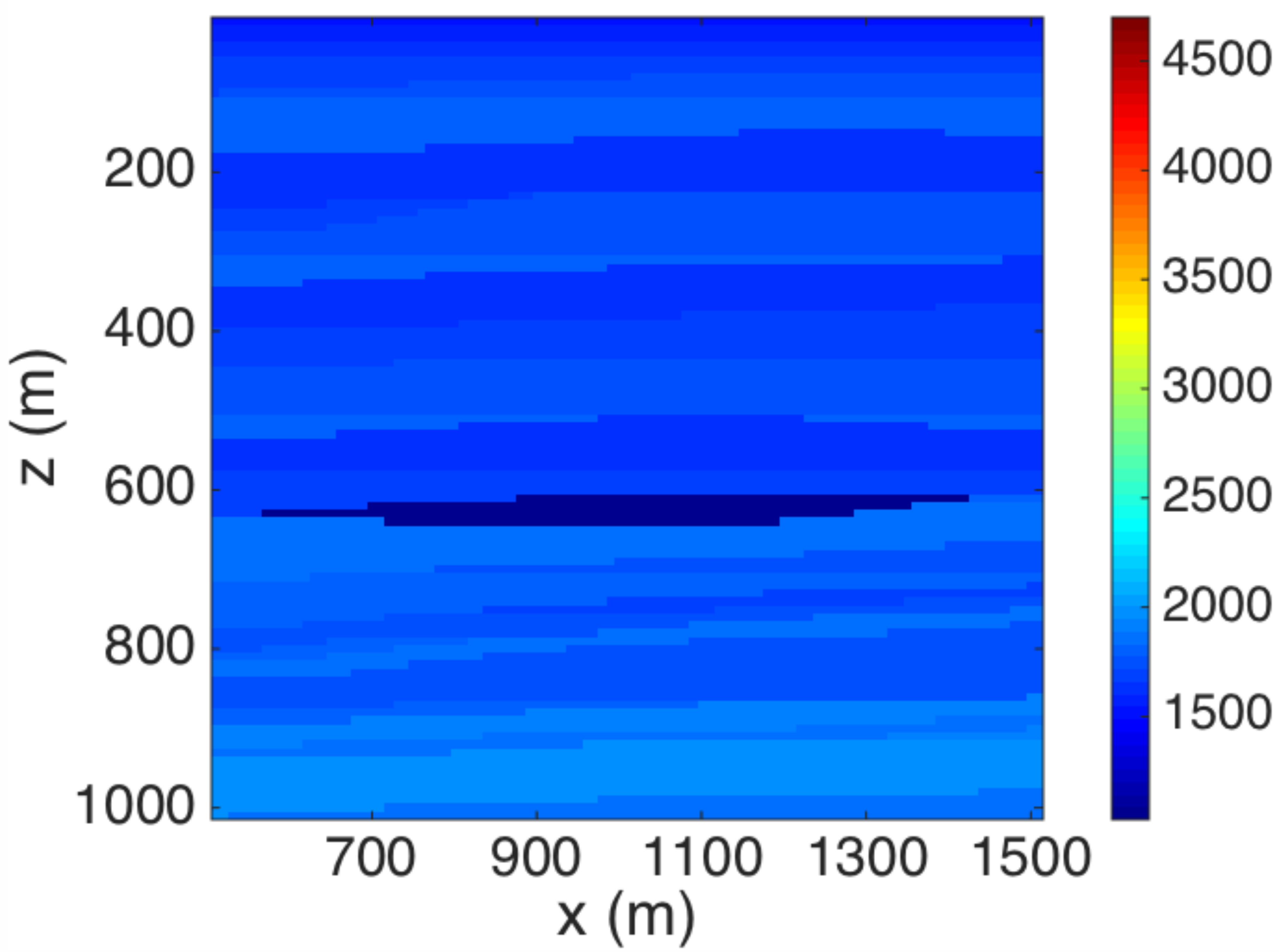
$$R_2 = H[m_2]^{-1} H[m_1] \; R_1$$

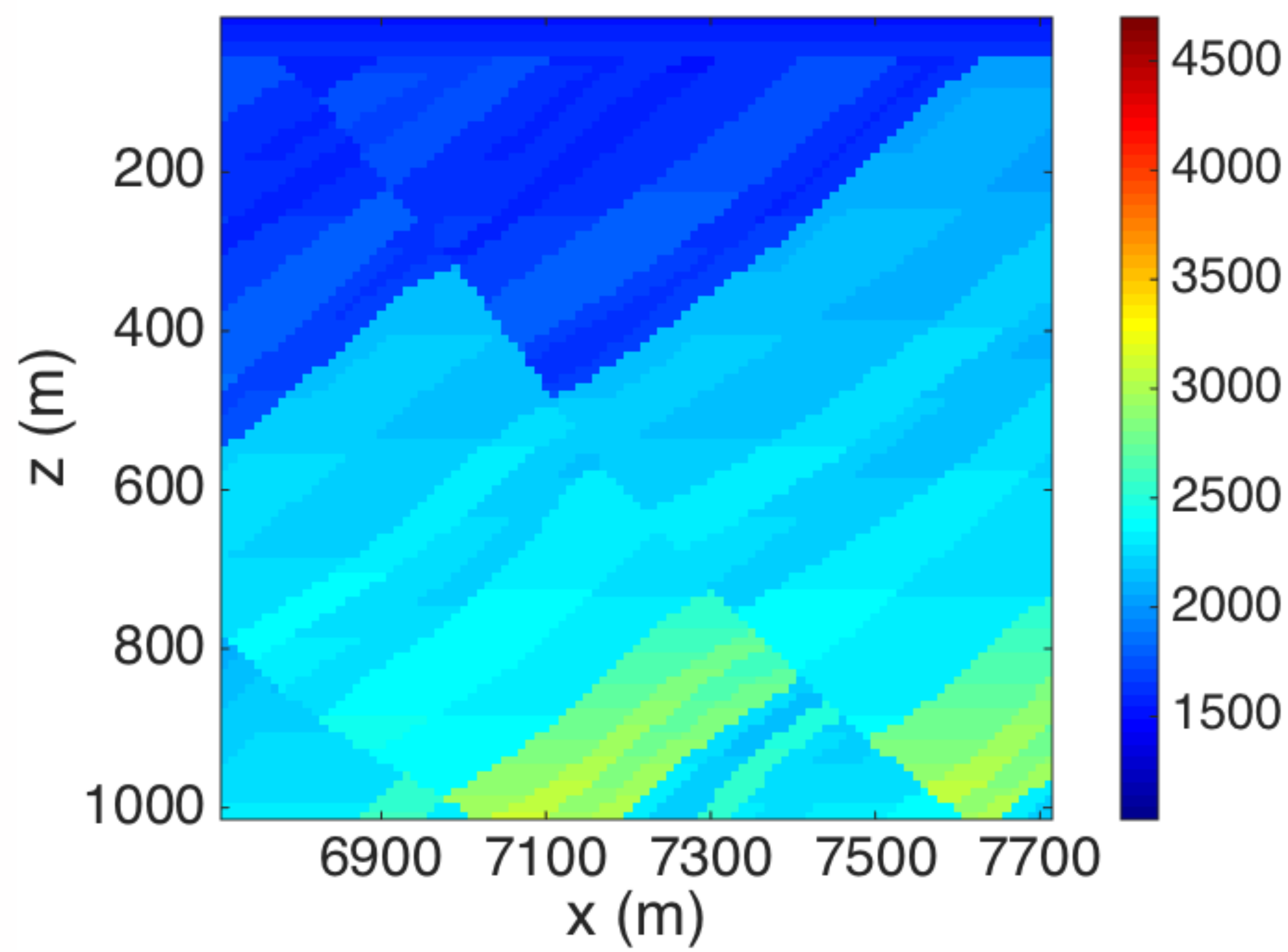
to compute

$$E_2 = L_2 R_2^*$$

with *only* $2r$ *extra PDEs solves!*

Invariance formula for EIVs (example 1)





background model 1

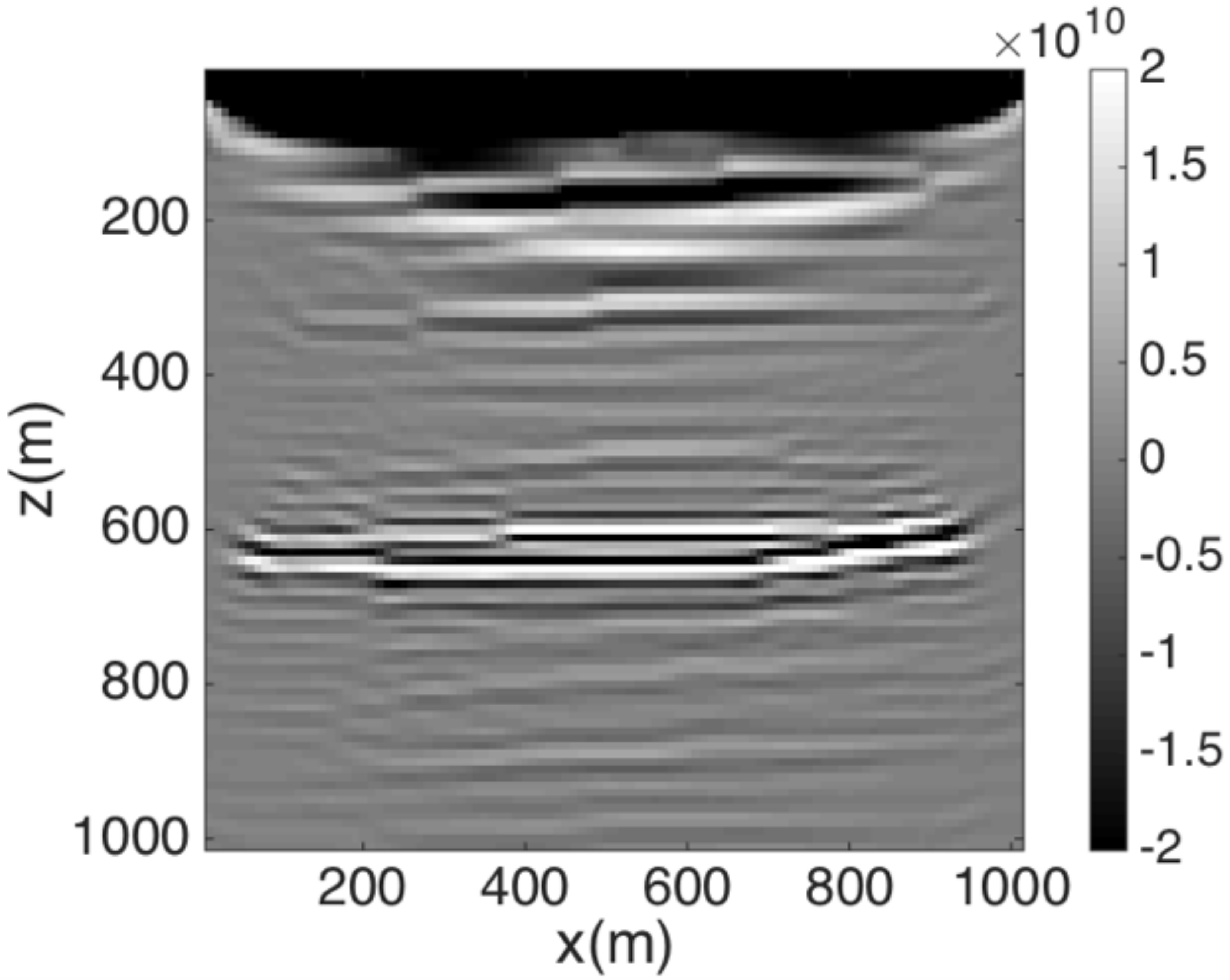
(correct)

background model 2

(incorrect)

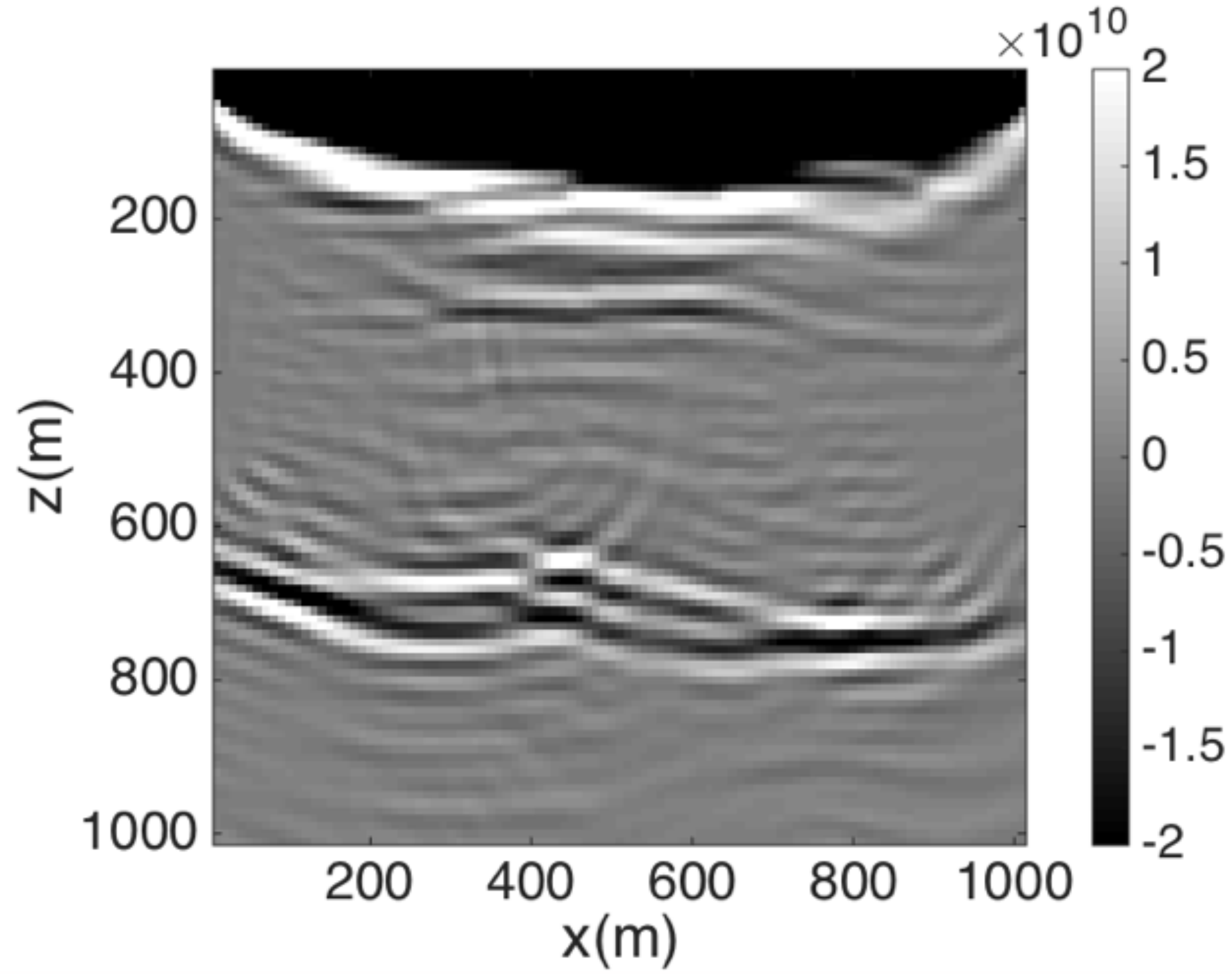
Invariance formula for EIVs (example 1)

Diagonal extraction of the low-rank EIV (5-30 Hz, step 0.5Hz, r = 15-45)



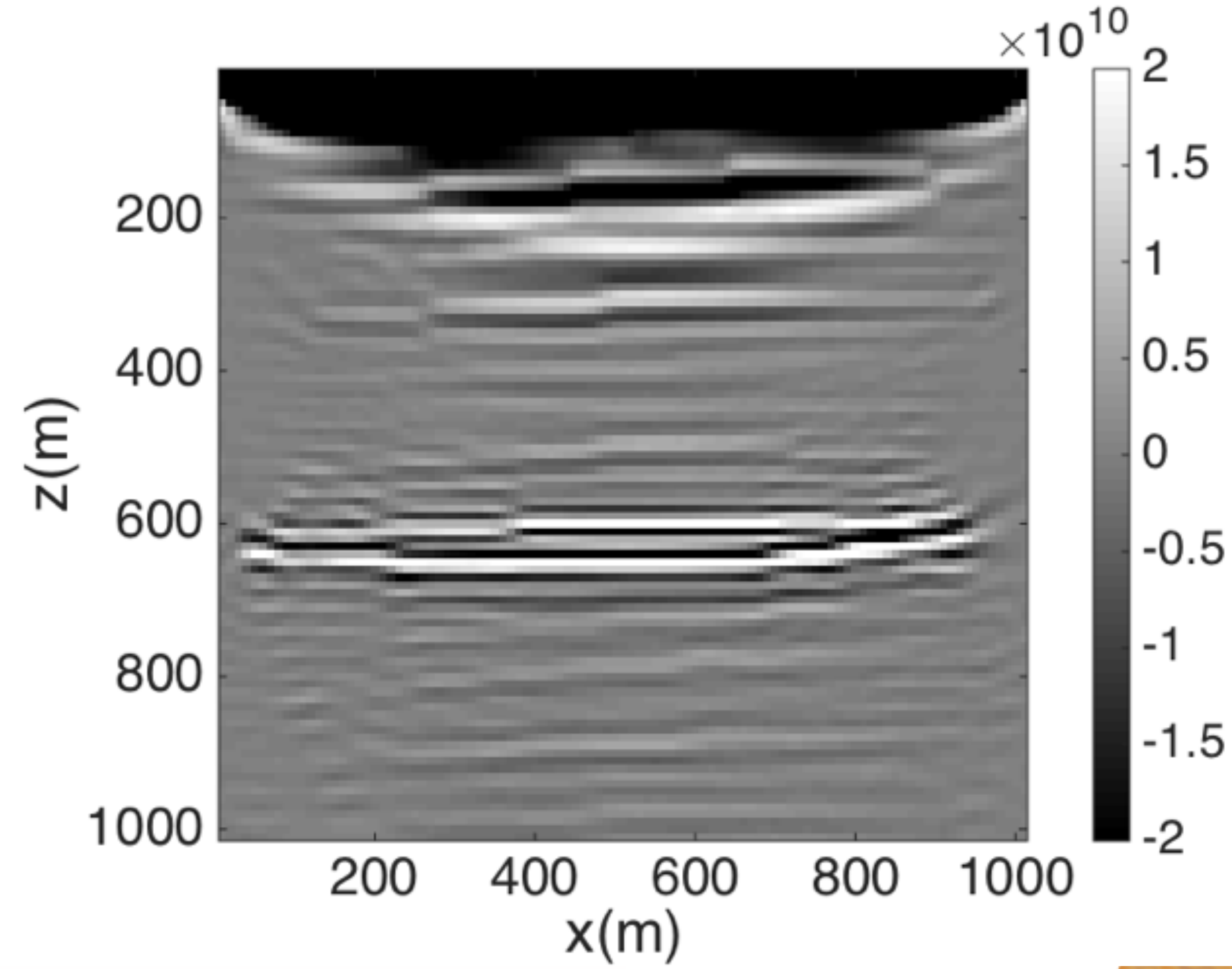
direct reconstruction

model 1



direct reconstruction

model 2



using invariance formula

from model 2 to get model 1

from wrong to correct!!!

Complexity analysis

Full subsurface offset extended images:

	# of PDE solves	size of EIV
conventional E	2Ns	N x N
mat-vec $\tilde{E} = EW$	2Nx	N x Nx
low-rank L, R	4r	2N x r

Ns = # sources

N = # grid points

Nx = # probing points

r = # estimated rank

Complexity analysis

Full subsurface offset extended images:

	# of PDE solves	size of EIV
conventional E	2Ns	N x N
mat-vec $\tilde{E} = EW$	2Nx	N x Nx
low-rank L, R	4r	2N x r

Ns = # sources

N = # grid points

Nx = # probing points

r = # estimated rank

we win when Nx << Ns
but usually Nx ~ N
(Dirac probing vectors)

SLIM

Complexity analysis

Full subsurface offset extended images:

	# of PDE solves	size of EIV
conventional E	2Ns	N x N
mat-vec $\tilde{E} = EW$	2Nx	N x Nx
low-rank L, R	4r	2N x r

Ns = # sources

N = # grid points

Nx = # probing points

r = # estimated rank

we win when r << Ns
okay from low-rank approx.
of data matrix!

Observations & Conclusions

Full-offset image volumes can be formed via probing

Form orthonormal basis that spans its range

— low-rank approximation via randomized SVD

— extract (off)diagonals from image volumes

Natural "parametrization" from linear algebra

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