Randomized sampling in *exploration* seismology

Felix J. Herrmann

SLIM

Seismic Laboratory for Imaging and Modeling
the University of British Columbia

Randomized sampling in *exploration* seismology

Felix J. Herrmann

SLIM team & Eldad Haber, and Michael Friedlander

SLIM

Seismic Laboratory for Imaging and Modeling

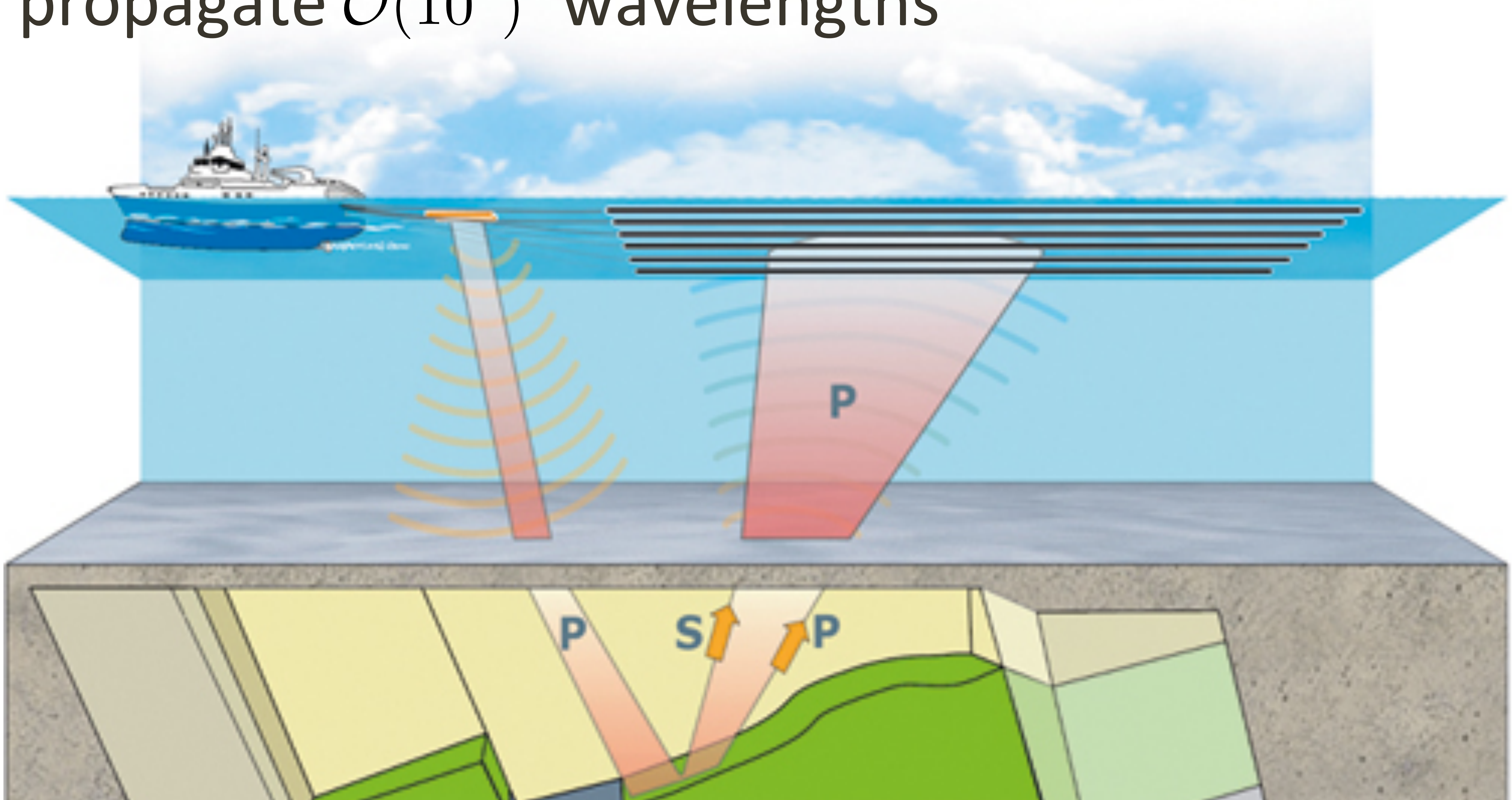
the University of British Columbia

SLIM

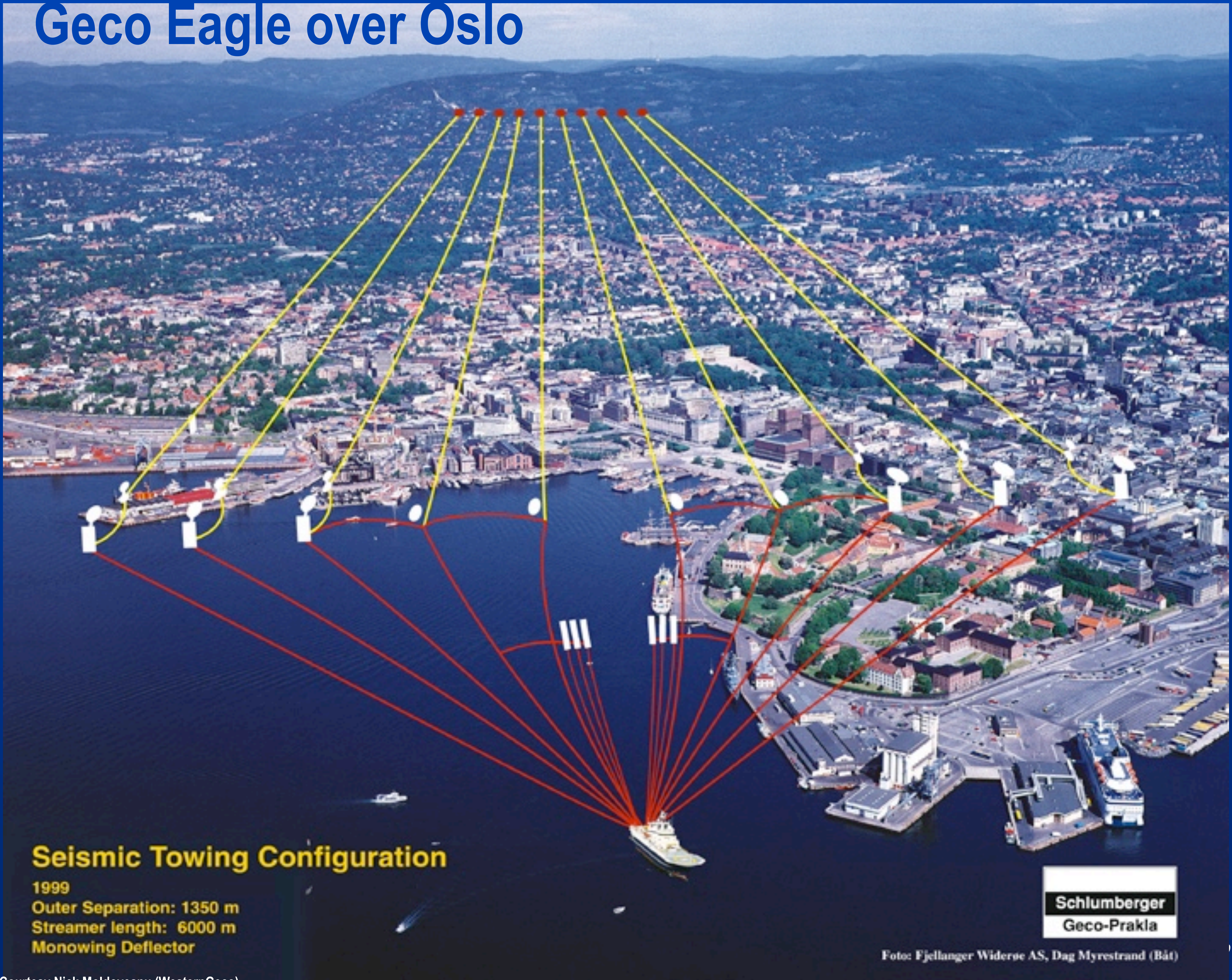
SLIM team

Infer 3D *velocity* model from *multi-experiment* data:

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



Geco Eagle over Oslo



Courtesy Nick Moldoveanu (WesternGeco)

from: T. Keho & P. Kelamis. the Leading Edge

Data *deluge*

"Moore's law" for *channel* count:

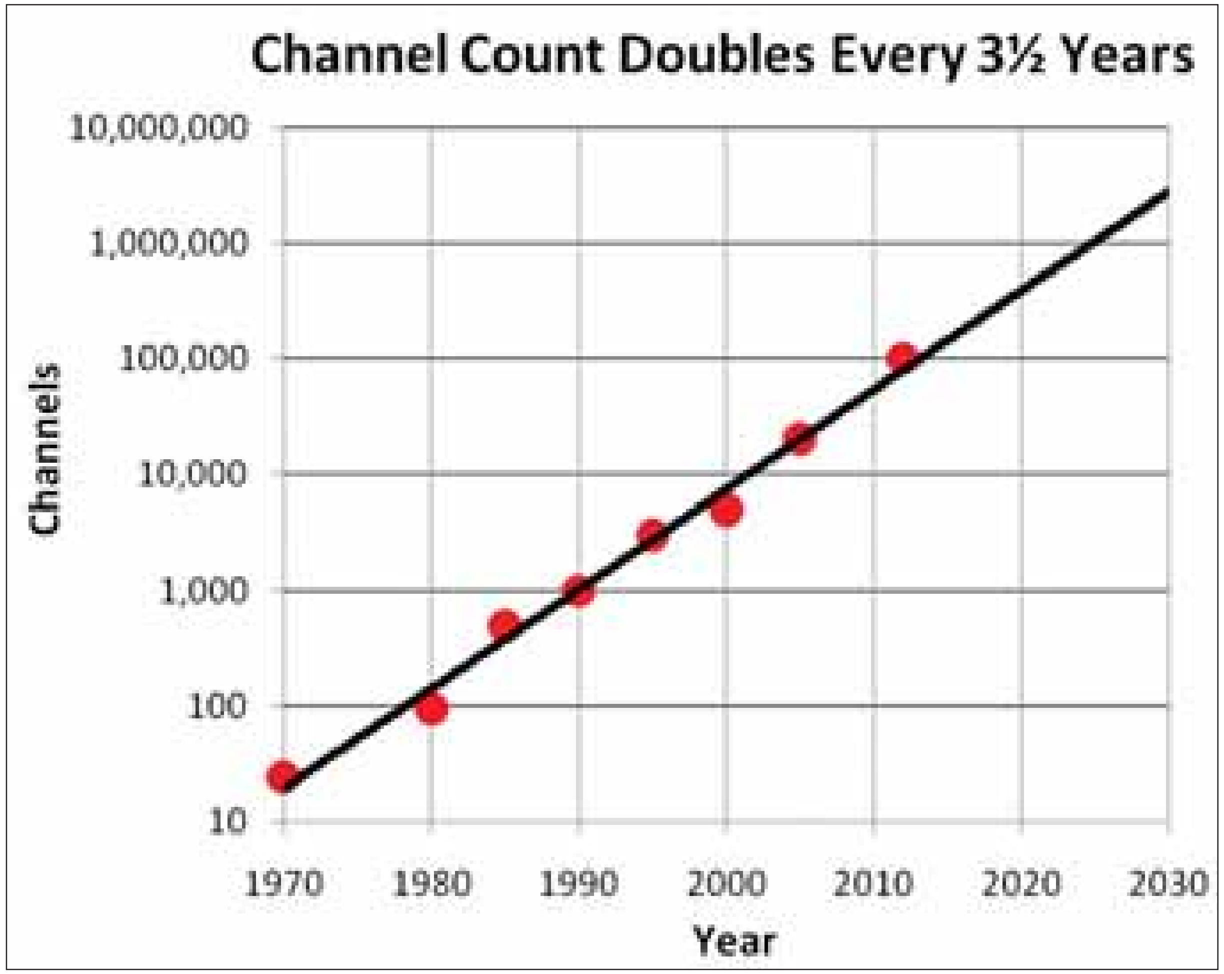


Figure 3. *"Moore's law" for seismic channel count (modified from Monk, 2006).*

Challenges

Main driver: *high* costs of *deep*-sea *drilling* ($250 M a pop) & low 1–in–10 hit *rates*

Push for *wave*-equation based *inversion*/full-waveform *inversion* (FWI):

- *high* costs ($20 M) of *acquisition* – *incomplete* data
- *high* costs (>200k cores) of *computations* – *iterative* algorithms touching *all* data are *prohibitively expensive*
- *nonconvex*, i.e., *local* minima leading to *non*uniqueness

Mathematical structure

Full-waveform inversion:

$$\min_{\mathbf{m}} \phi(\mathbf{m}) = \sum_{i=1}^{K} \phi_i(\mathbf{d}_i, \mathbf{q}_i; \mathbf{m})$$

$$\mathbf{q}_i = \text{"known" monochromatic source}$$

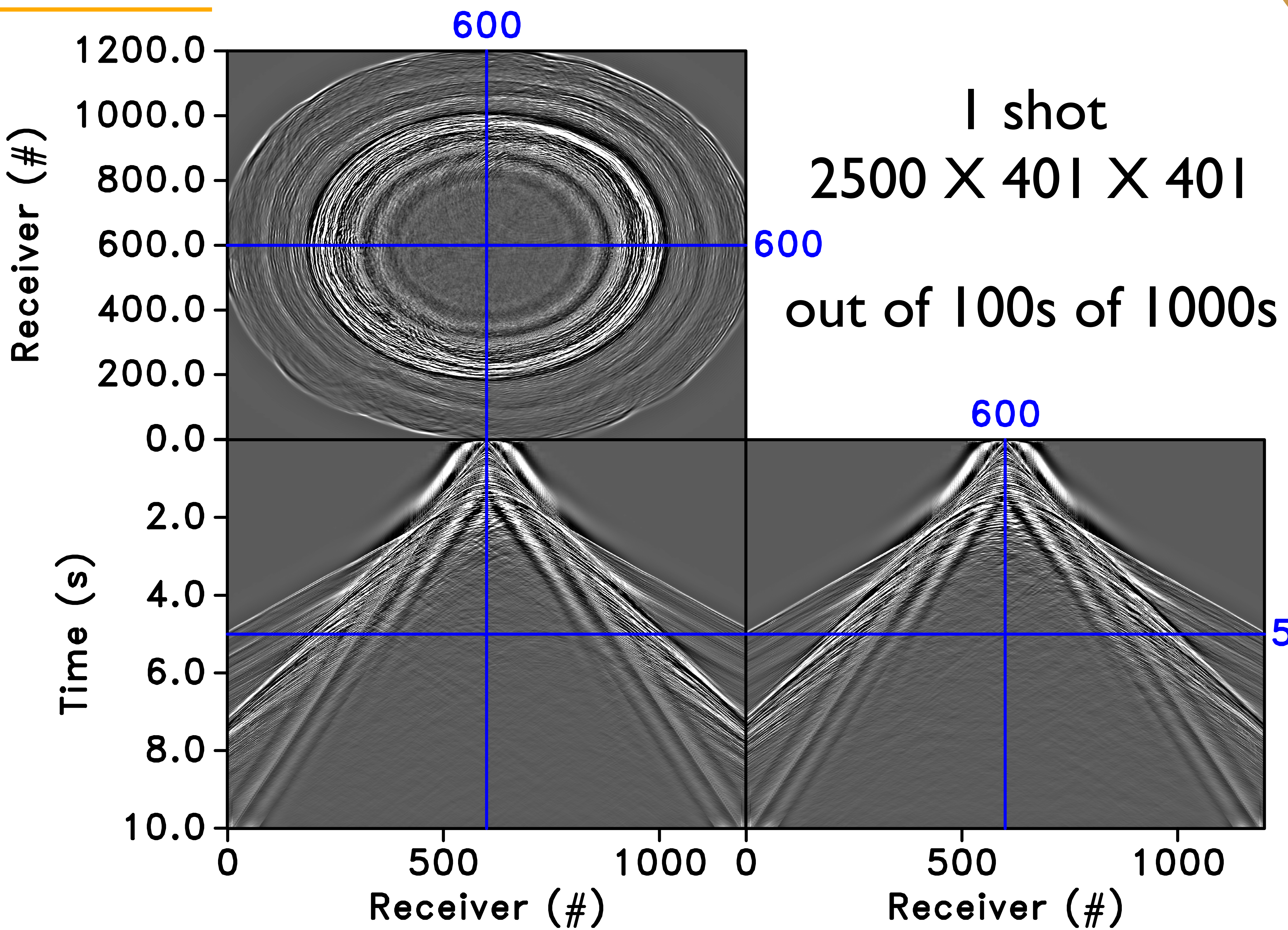
$$\mathbf{d}_i = \text{measured monochromatic shot record}$$

$$\mathbf{m} = \text{unknown medium properties}$$

$$K = n_s \times n_f$$

Parameter estimation / *machine* learning problem w/ PDE *constraints*...

Data deluge

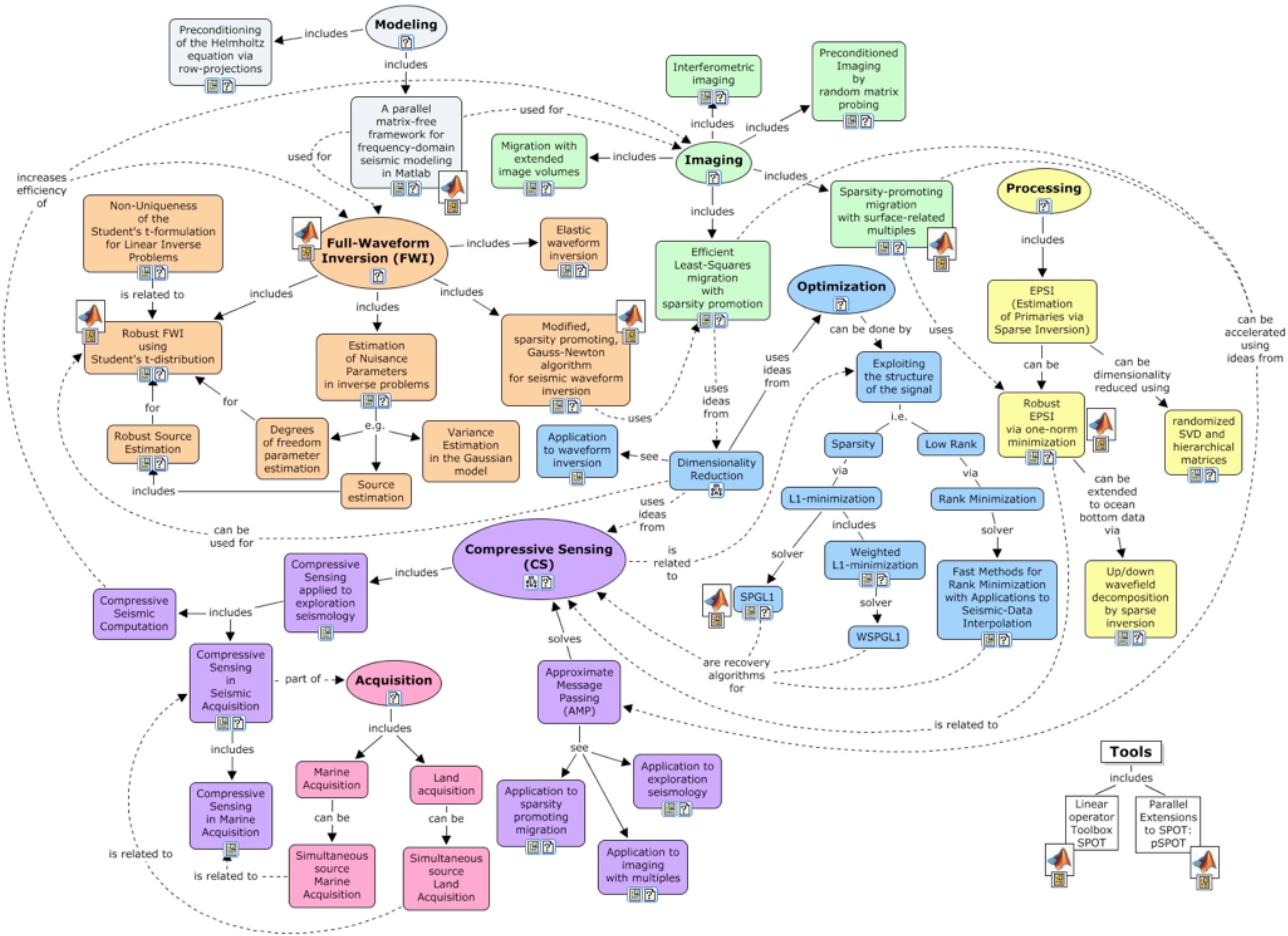


1 shot

2500 X 401 X 401

out of 100s of 1000s

Seismic Laboratory for Imaging and Modeling



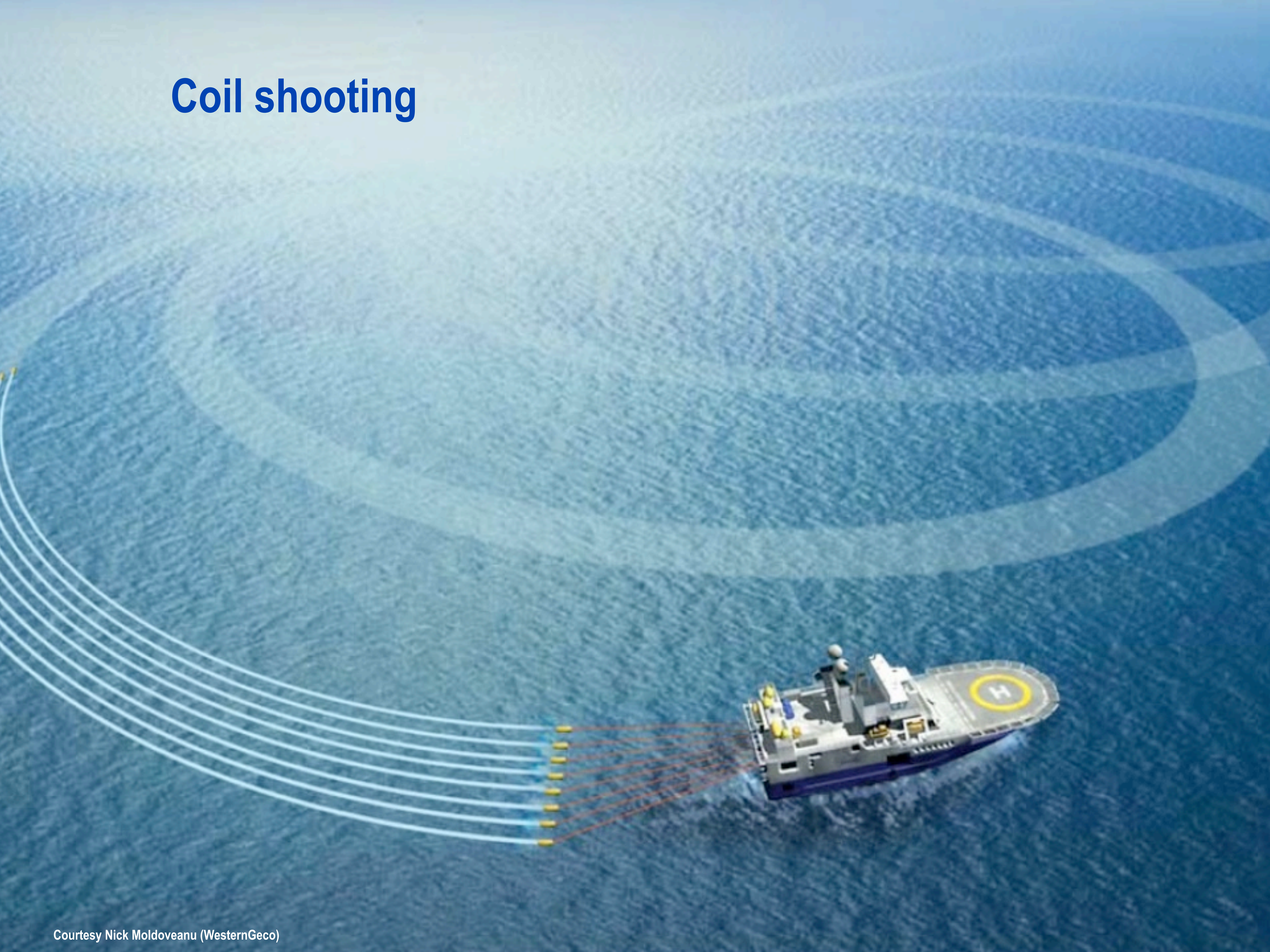
Our contributions

Proposal to *randomize* acquisition

- *random* source/receiver locations
- *jittered* time *dithering* in (simultaneous) source marine acquisition
- recovery via *curvelet*-domain *sparsity* promotion or *low-rank* promotion

Courtesy Nick Moldoveanu (WesternGeco)

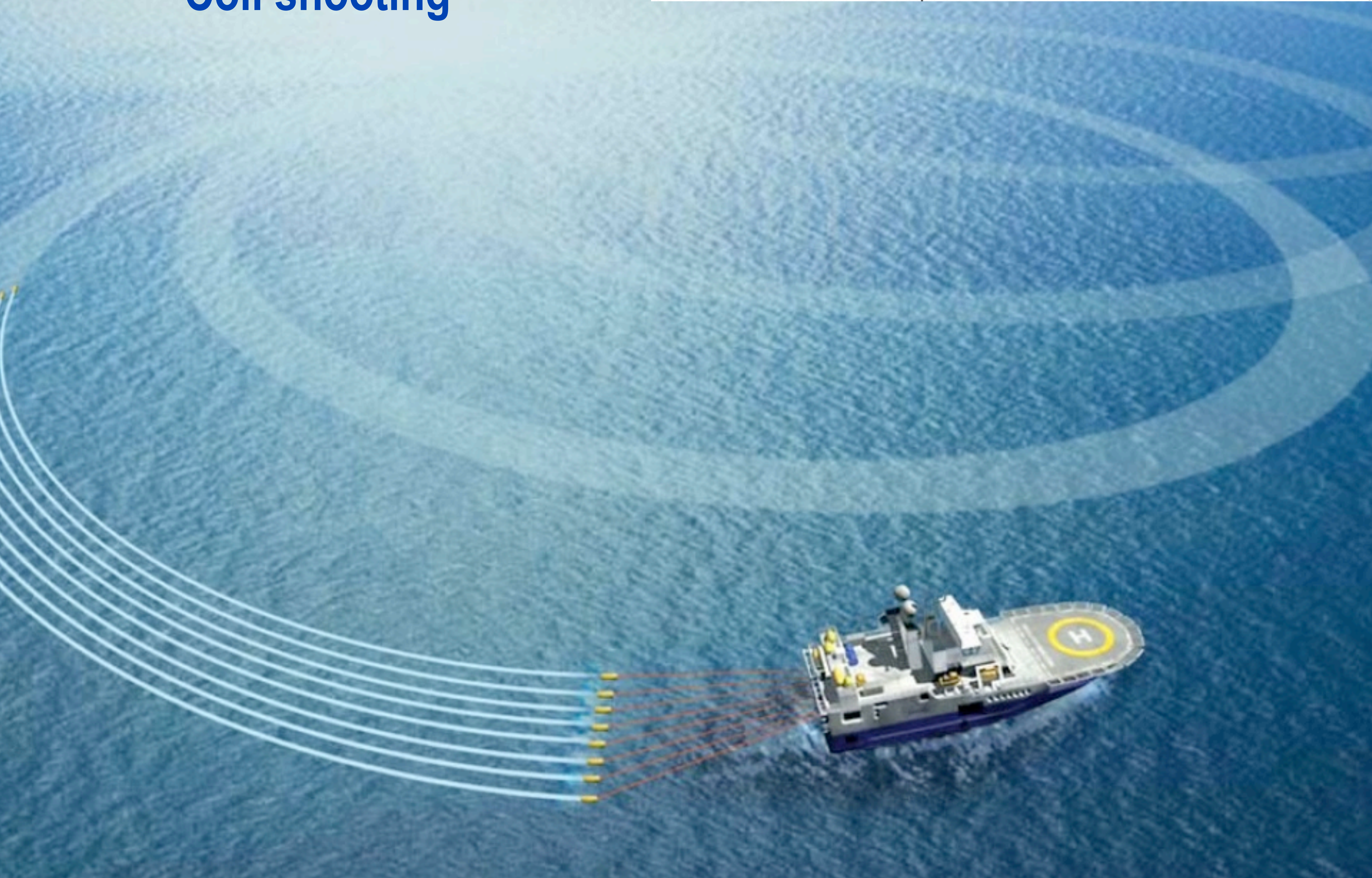
Coil shooting



Courtesy Nick Moldoveanu (WesternGeco)

Coil shooting

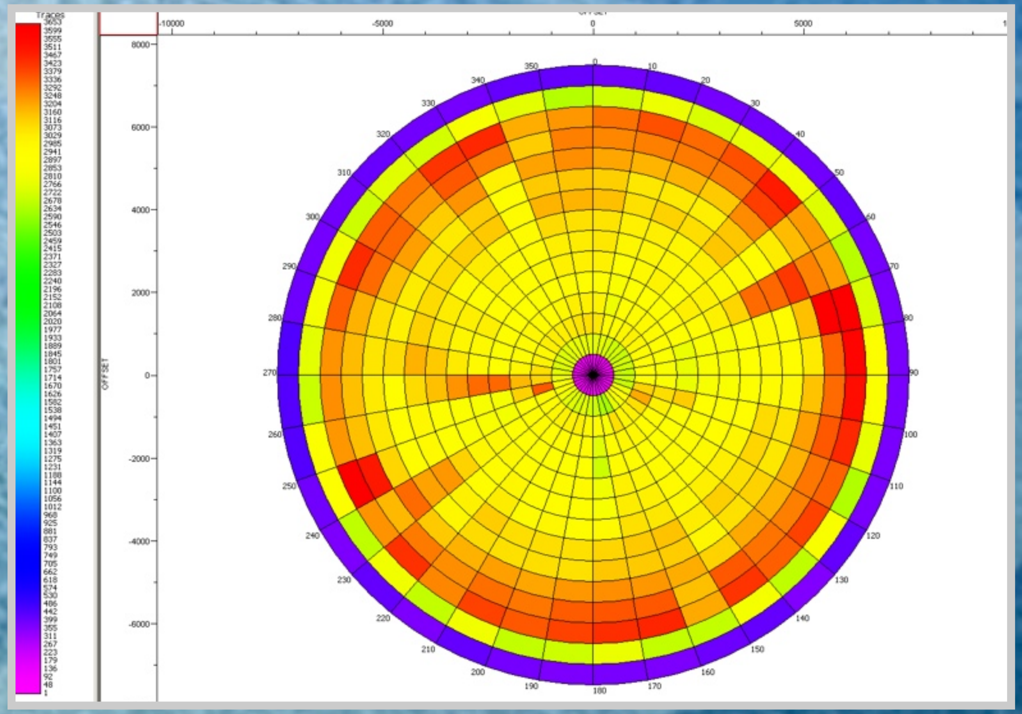
$$W(t, x_s, y_s, x_r, y_r)$$

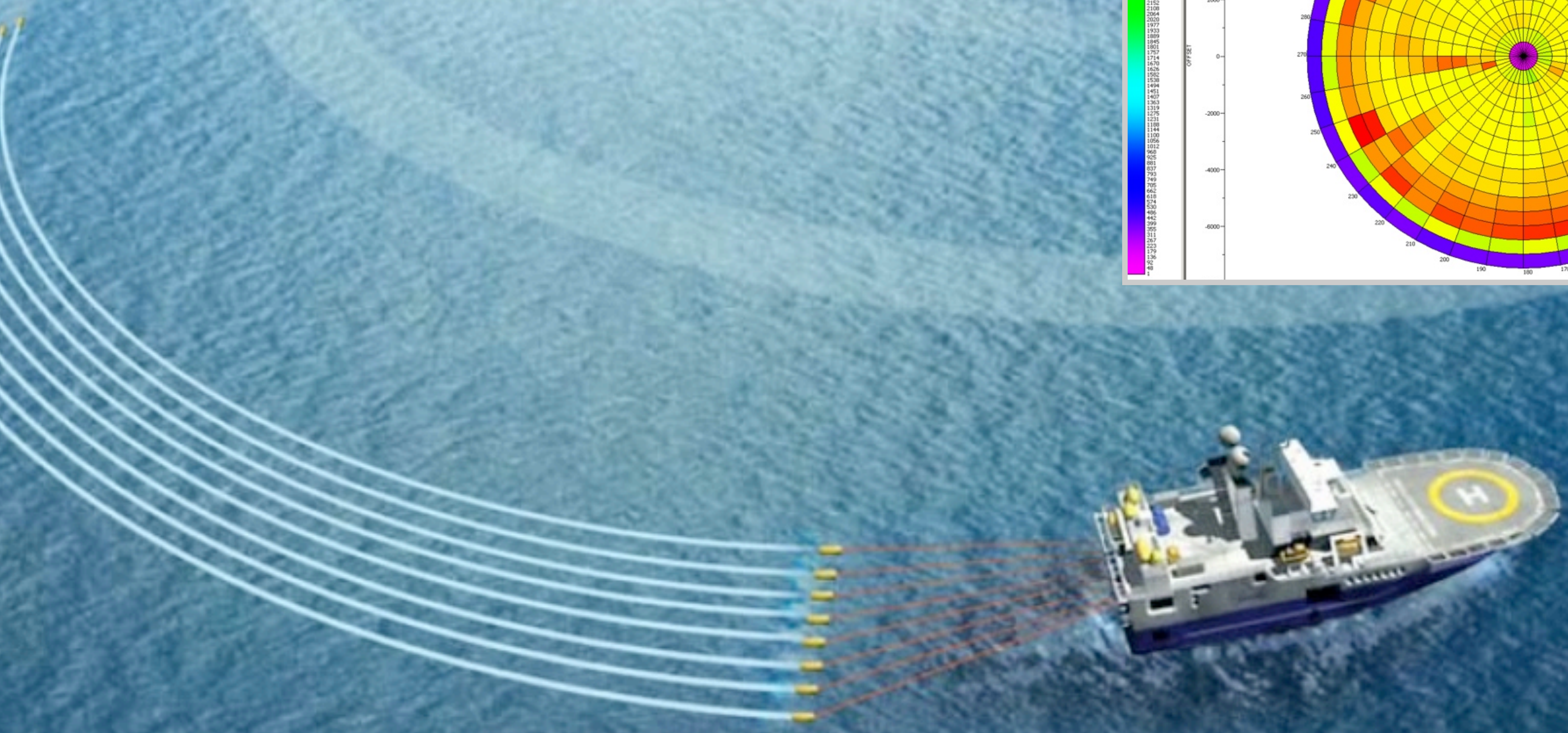


Courtesy Nick Moldoveanu (WesternGeco)

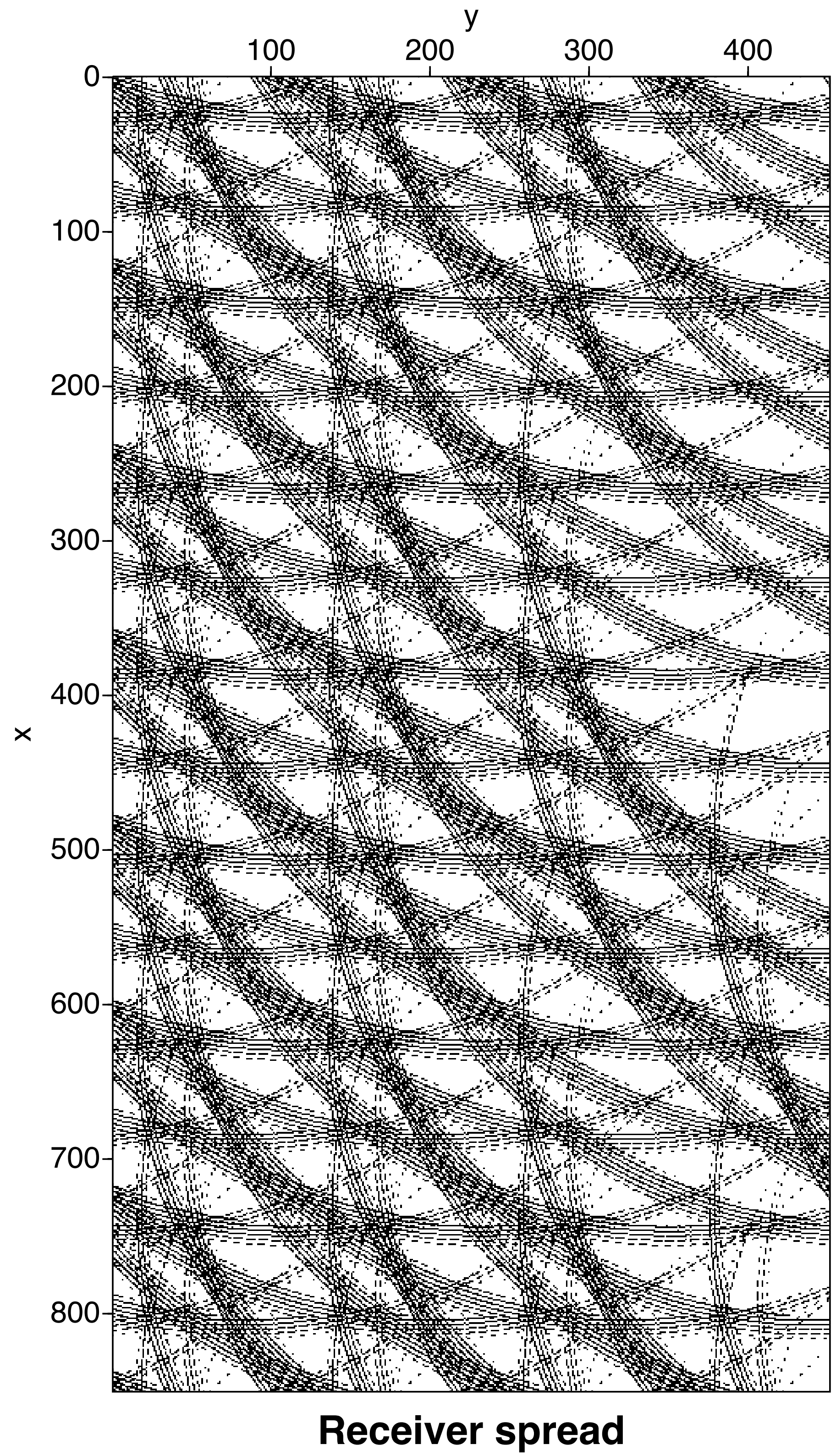
$$W(t, x_s, y_s, x_r, y_r)$$

Coil shooting





Courtesy Nick Moldoveanu (WesternGeco)



34 % of samples

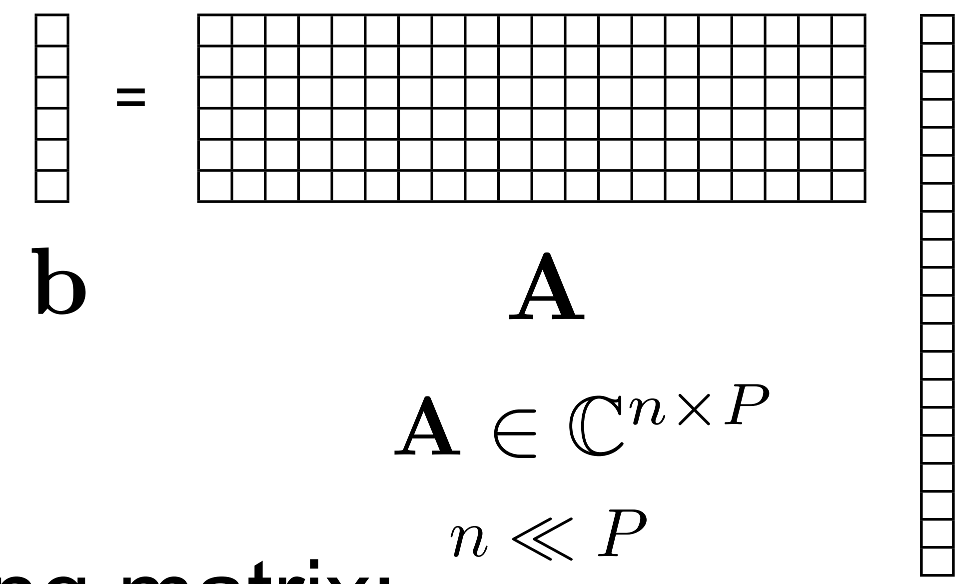
Courtesy Nick Moldoveanu

Problem statement

Solve an *underdetermined* system of *linear* equations:

data (measurements /observations)

$\mathbf{b} \in \mathbb{C}^n$



$\mathbf{b}$ $\mathbf{A}$ $\mathbf{x_0}$

$\mathbf{A} \in \mathbb{C}^{n \times P}$

$n \ll P$

unknown coefficients

$\mathbf{x_0} \in \mathbb{C}^P$

Compressive sensing matrix:

$$\mathbf{A} = \mathbf{RMS}^H$$

Sparse recovery

Sparsity-promoting program:

$$\underbrace{\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1}_{\text{support detection}} \quad \text{subject to} \quad \underbrace{\mathbf{A}\mathbf{x} = \mathbf{b}}_{\text{data-consistent amplitude recovery}}$$

Sparsity-promoting solver: $\mathbf{SPG}\ell_1$ [van den Berg and Friedlander, 2008]

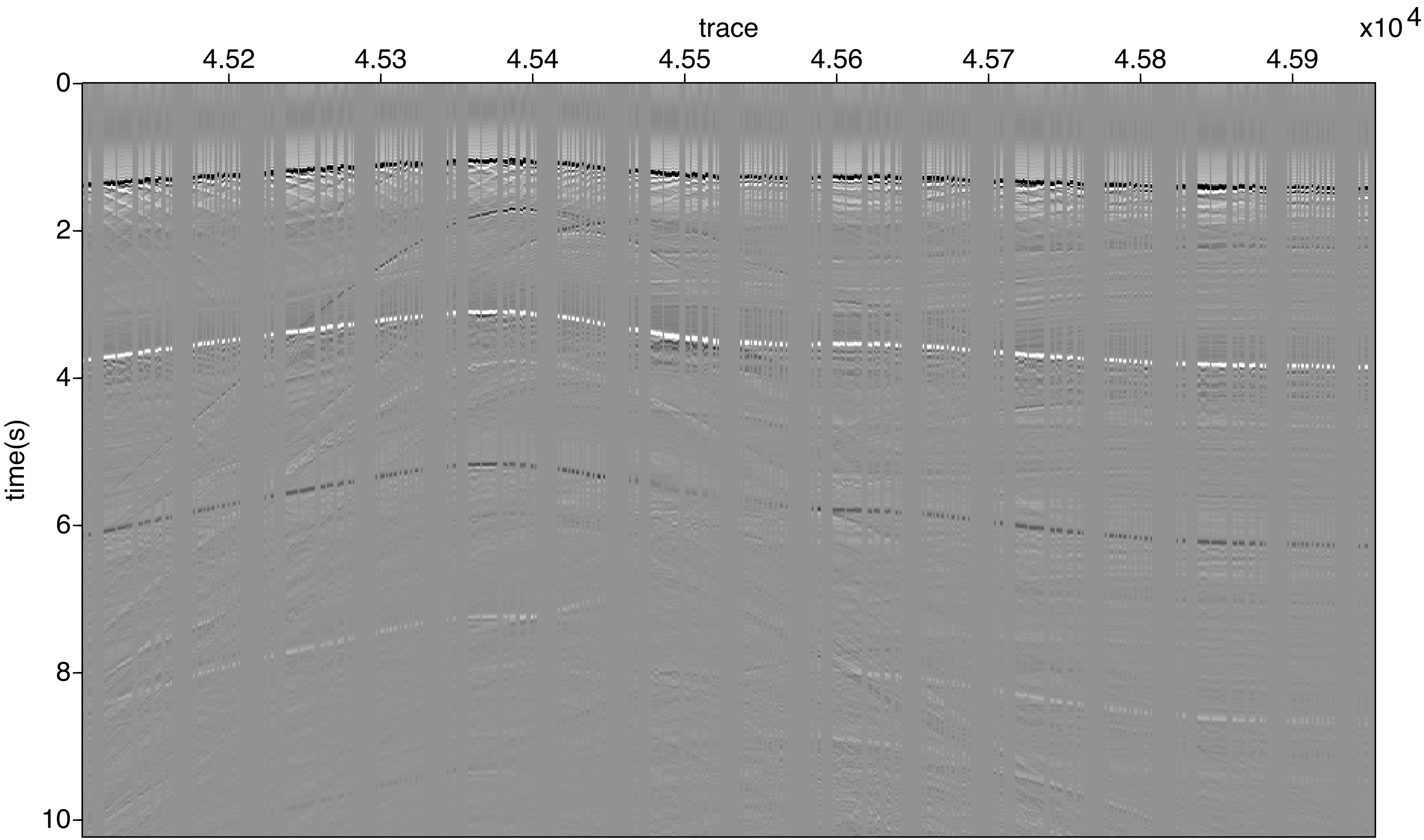
Recover single-source prestack data volume: $\tilde{\mathbf{d}} = \mathbf{S}^{\mathbf{H}}\tilde{\mathbf{x}}$

Randomization favors *sparse* recovery by rendering *interference* into *incoherent* Gaussian noise...That's the hope in practice...

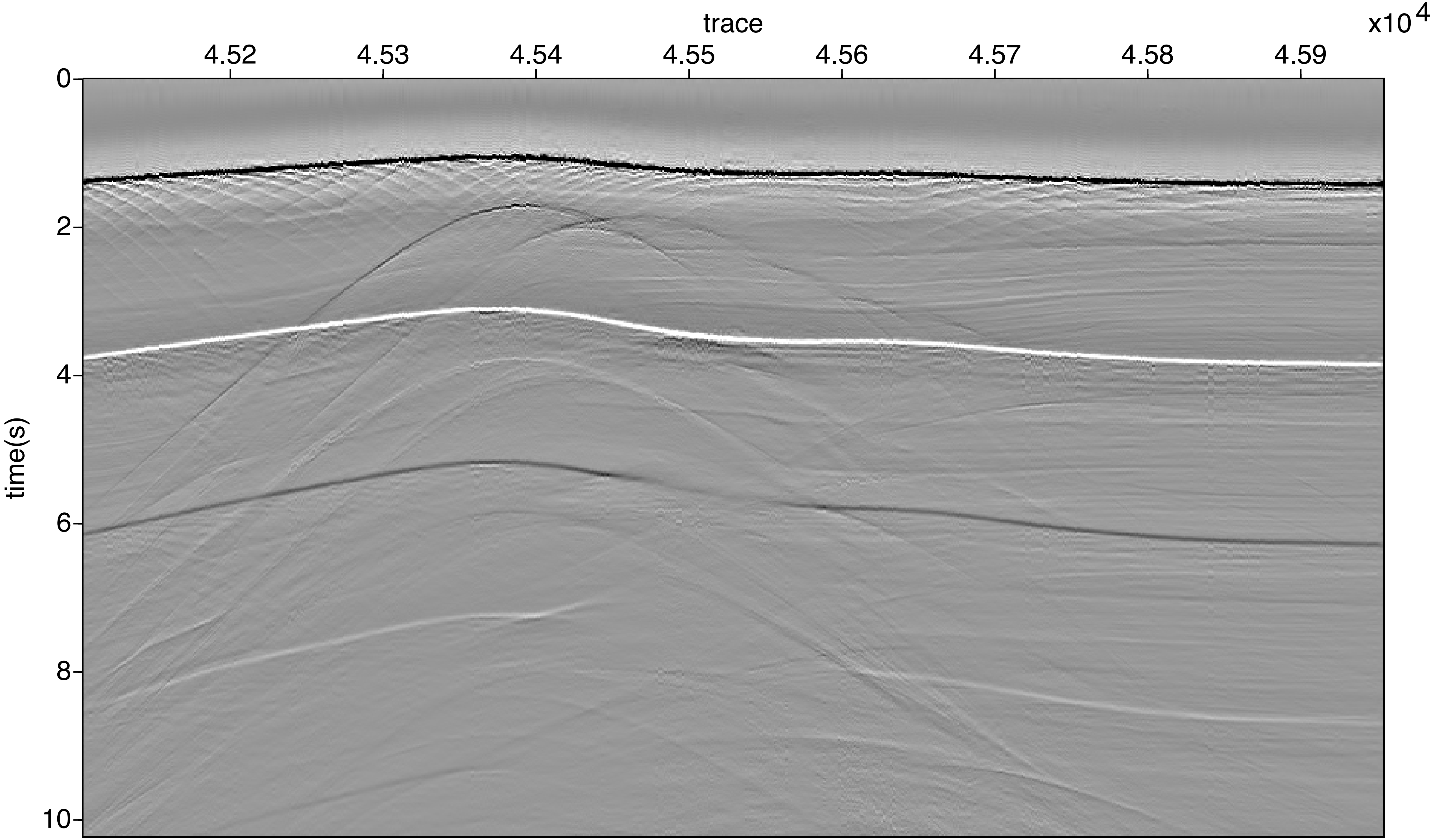
SLIM

Challenge

```
Starting SPGl1 recovery...

 ================================================================================
 SPGL1_SLIM v. 46  (Tue, 14 Jun 2011) based on v.1017
 ================================================================================
 No. rows              : 103672320      No. columns          : 1459253760
 Initial tau           : 0.00e+00      Two-norm of b        : 3.92e+05
 Optimality tol        : 1.00e-04      Target objective     : 0.00e+00
 Basis pursuit tol     : 1.00e-06      Maximum iterations   :      110

  Iter      Objective   Relative Gap  Rel Error      gNorm   stepG   nnzX    nnzG            tau
     0  3.9236638e+05  0.0000000e+00   1.00e+00  6.903e+03    0.0       0       0  2.2303101e+07
     1  3.9219958e+05  1.9364118e+00   1.00e+00  6.677e+03   -0.3       2       0
     2  3.4192692e+05  2.1884194e+00   1.00e+00  5.147e+03    0.0   14452       0
     3  3.2859582e+05  4.1722491e-01   1.00e+00  1.373e+03    0.0   48295       0

   108  1.5609476e+03  1.6347854e+04   1.00e+00  7.335e+00    0.0  356264726       0
   109  1.5850938e+03  9.3198454e+04   1.00e+00  4.283e+01    0.0  346355398       0
   110  1.5641524e+03  6.9308202e+04   1.00e+00  3.104e+01    0.0  345144021       0

 ERROR EXIT -- Too many iterations

 Products with A     :     125        Total time   (secs) :  34838.7
 Products with A'    :     112        Project time (secs) :  2875.2
 Newton iterations   :      26        Mat-vec time (secs) :  25882.1
 Line search its     :      23        Subspace iterations :       0
```



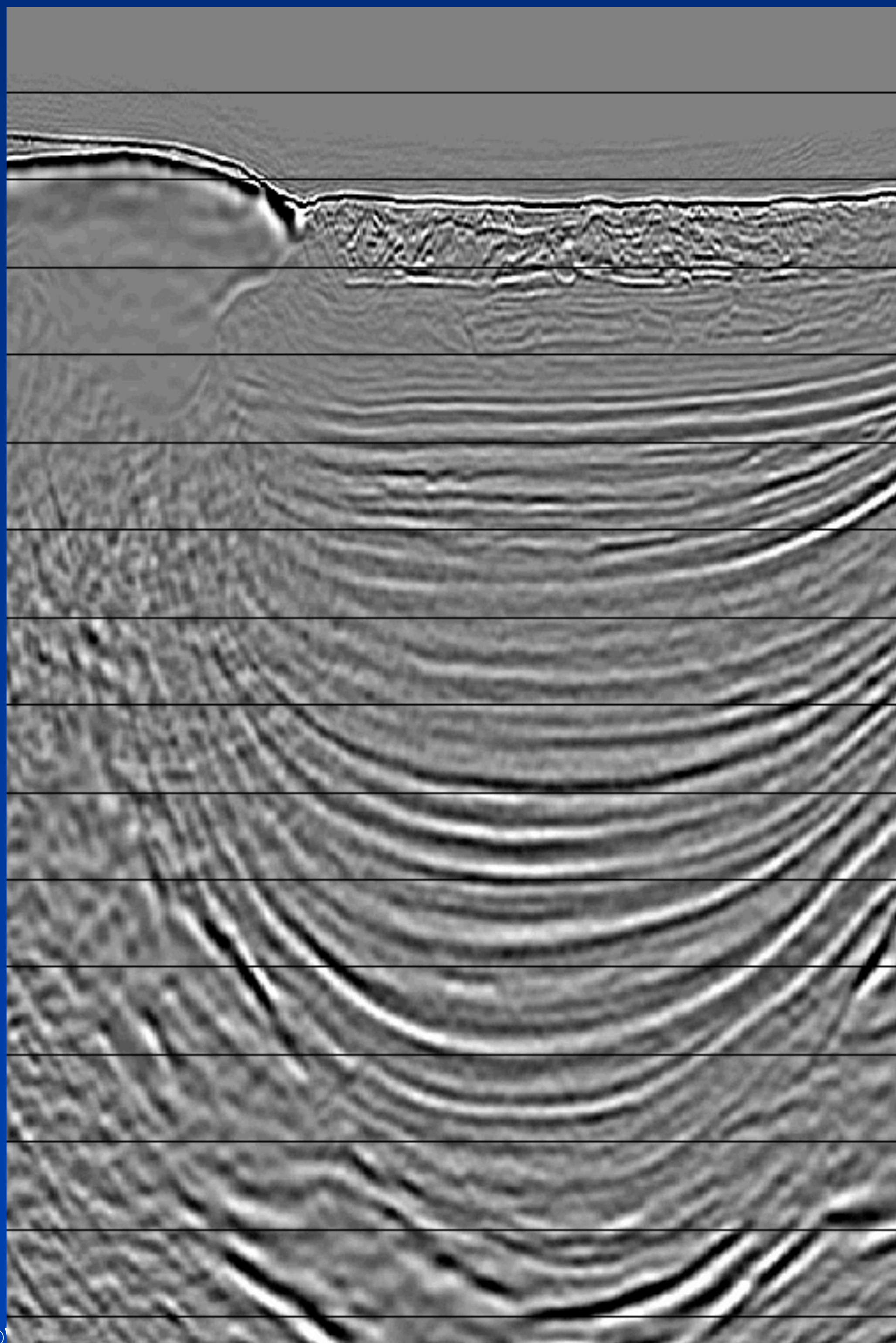
Input data



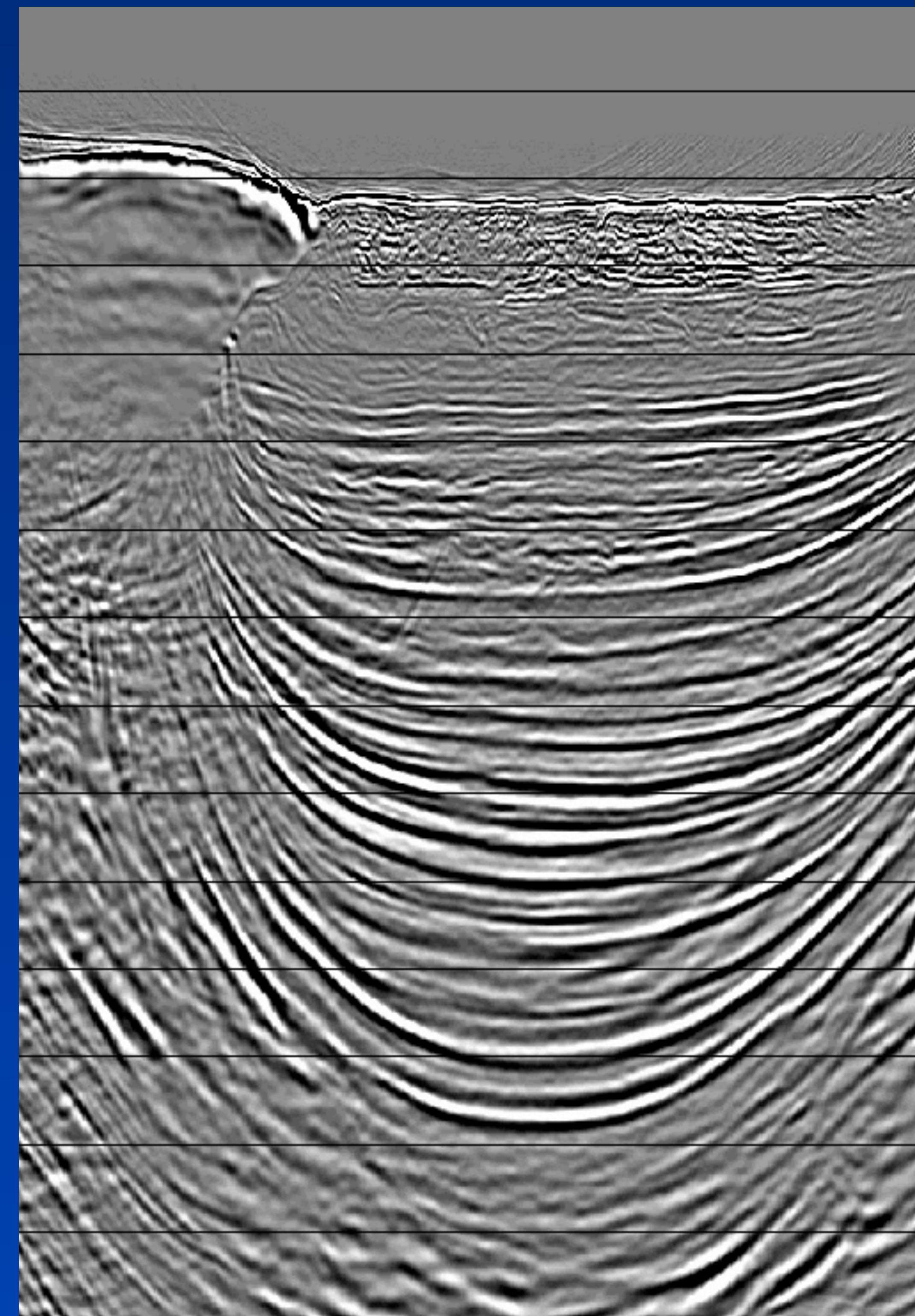
Interpolation with 2D Curvelet

WAZ vs. coil shooting comparison: the same processing sequence was applied on both datasets

WAZ



Coil



Courtesy Nick Moldoveanu (WesternGeco)

SLIM

Challenges

Extension to 3D seismic (5-D data) exposes vulnerabilities

- *redundancy* of *directional* spasifying transforms
- cost of matvecs and # of matvecs for *convex* optimization

Explore a different kind of structure

- "low-rank" SVD-free matrix / tensor factorizations
- rank increasing incoherent sampling

Recent work

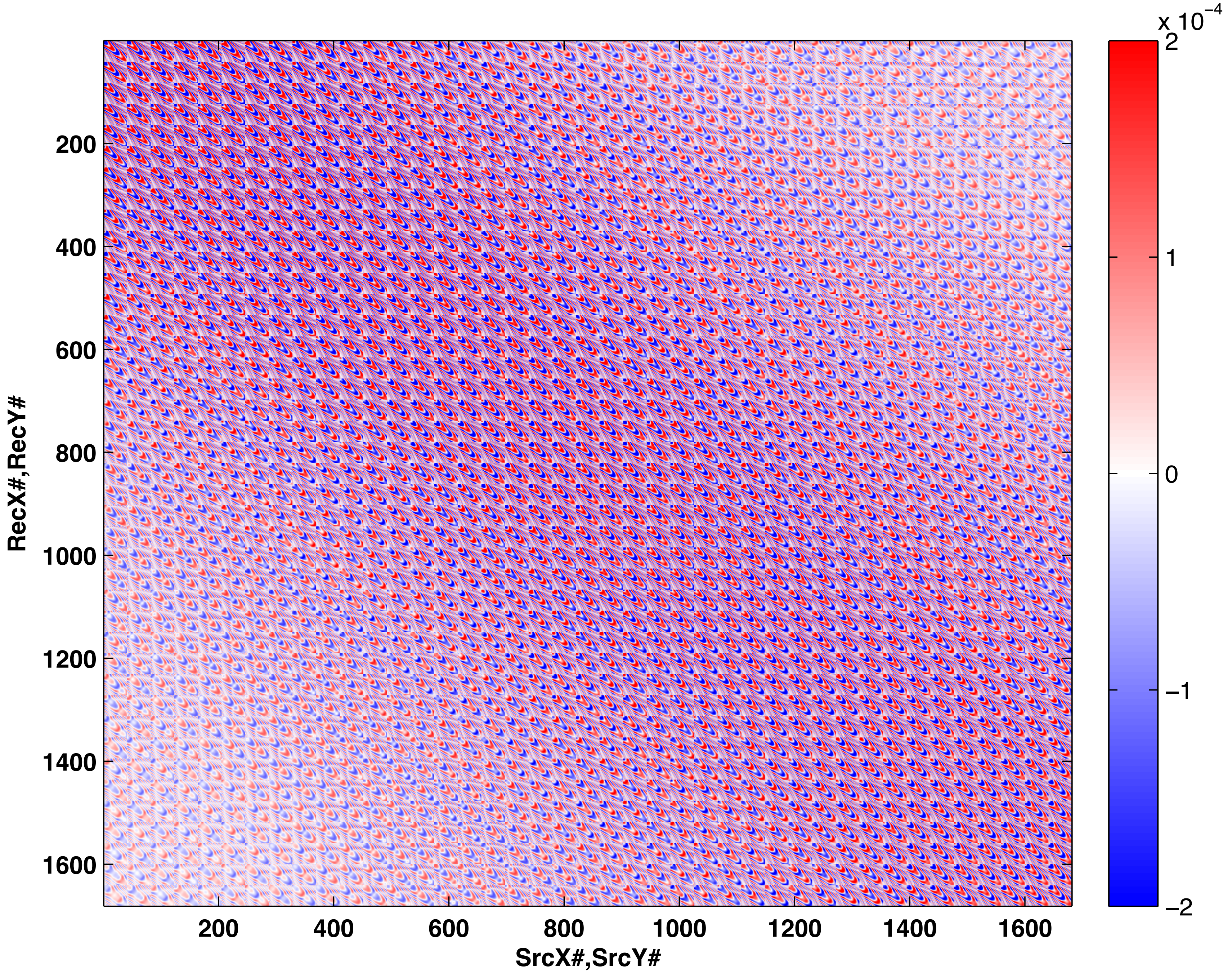
Under certain *perturbations* matrizations/tensorizations

- *low*-frequency frequency slices become *low*-rank
- *randomized* samplings *induce high*-rank

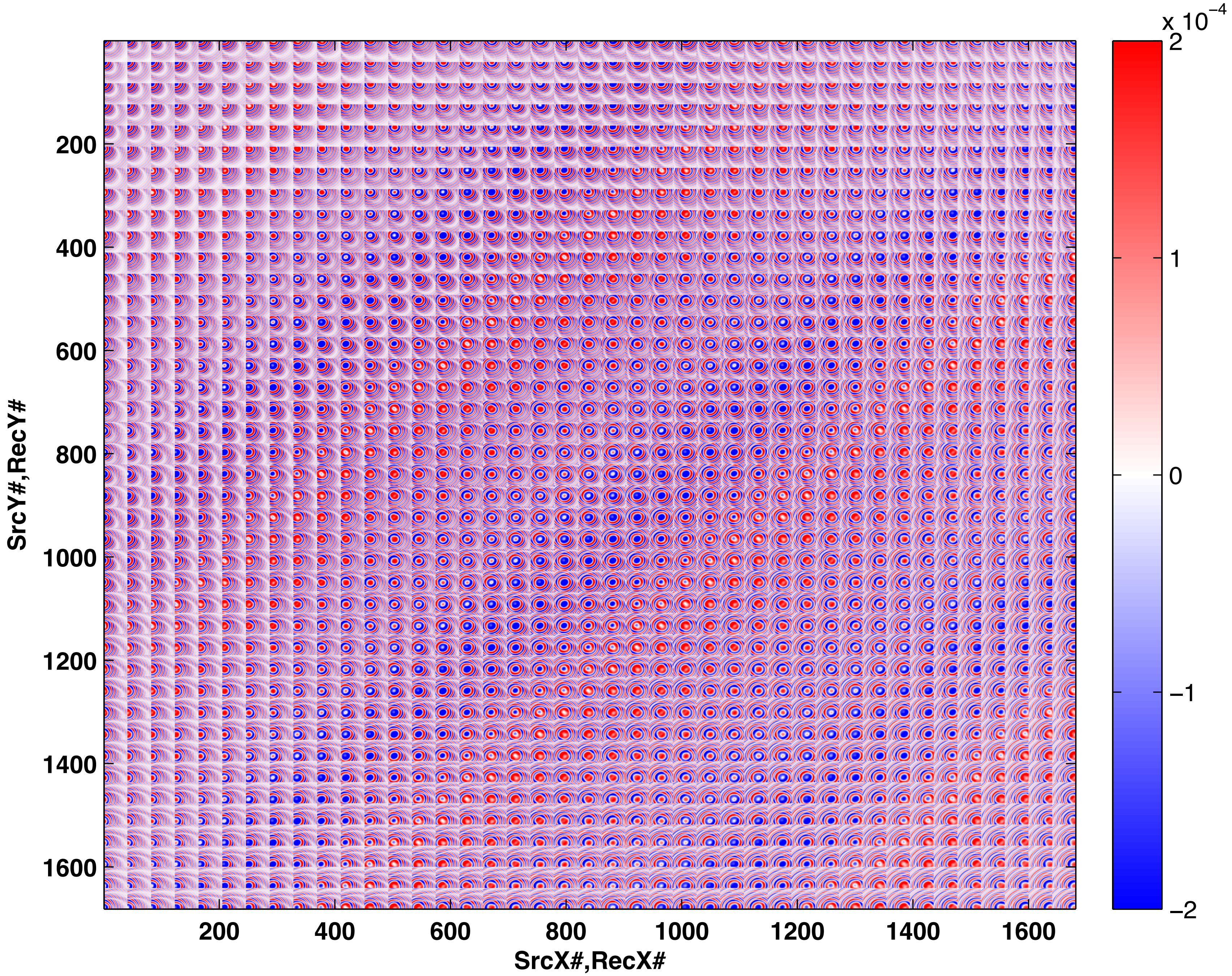
Conducive to *rank*-minimization

- *SVD*-free *nuclear* norm-minimization (w/ Ben Recht)
- SVD-free *hierarchical* Tucker w/ *manifold* optimization

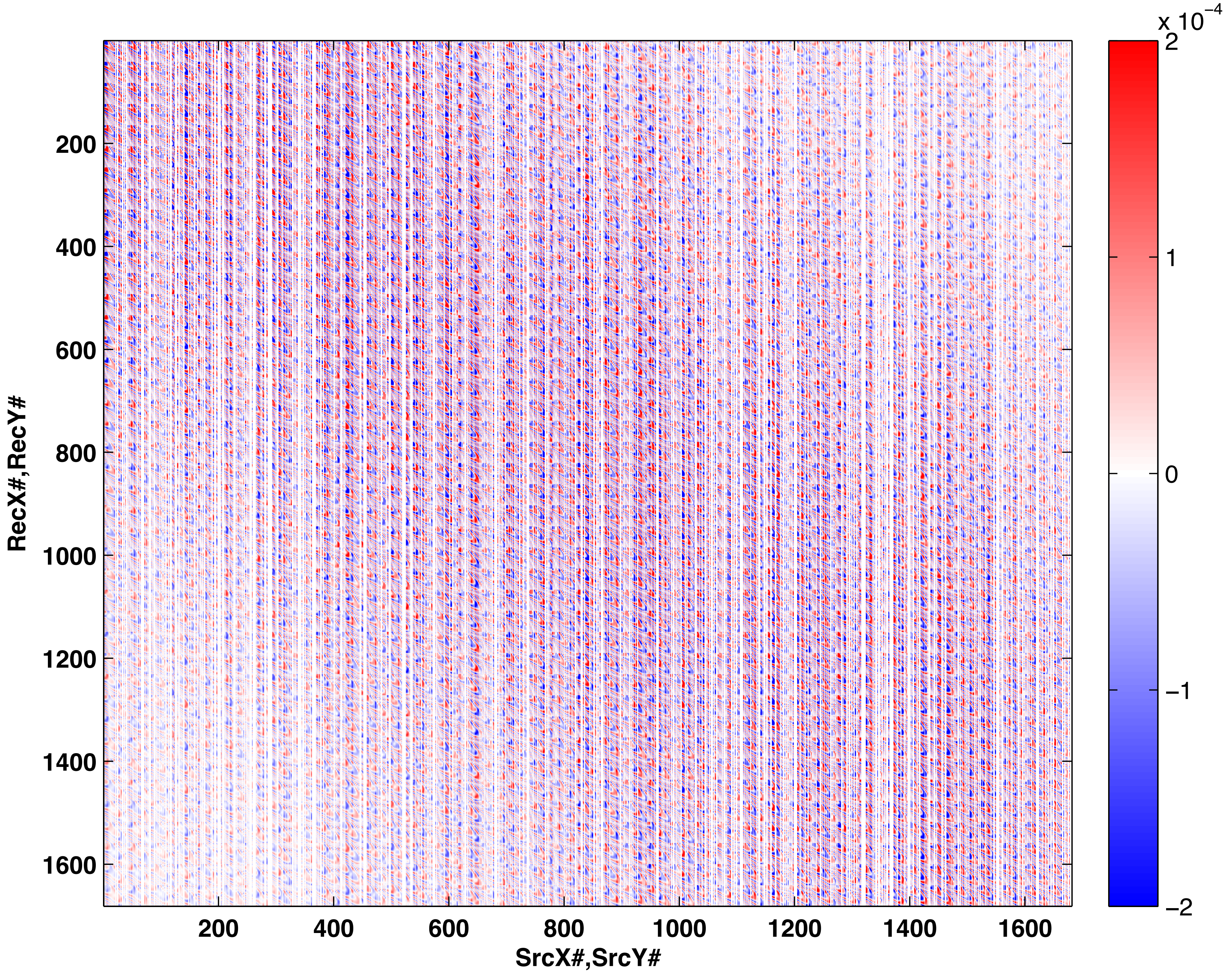
3D Acquisition
[Regular sampled data]



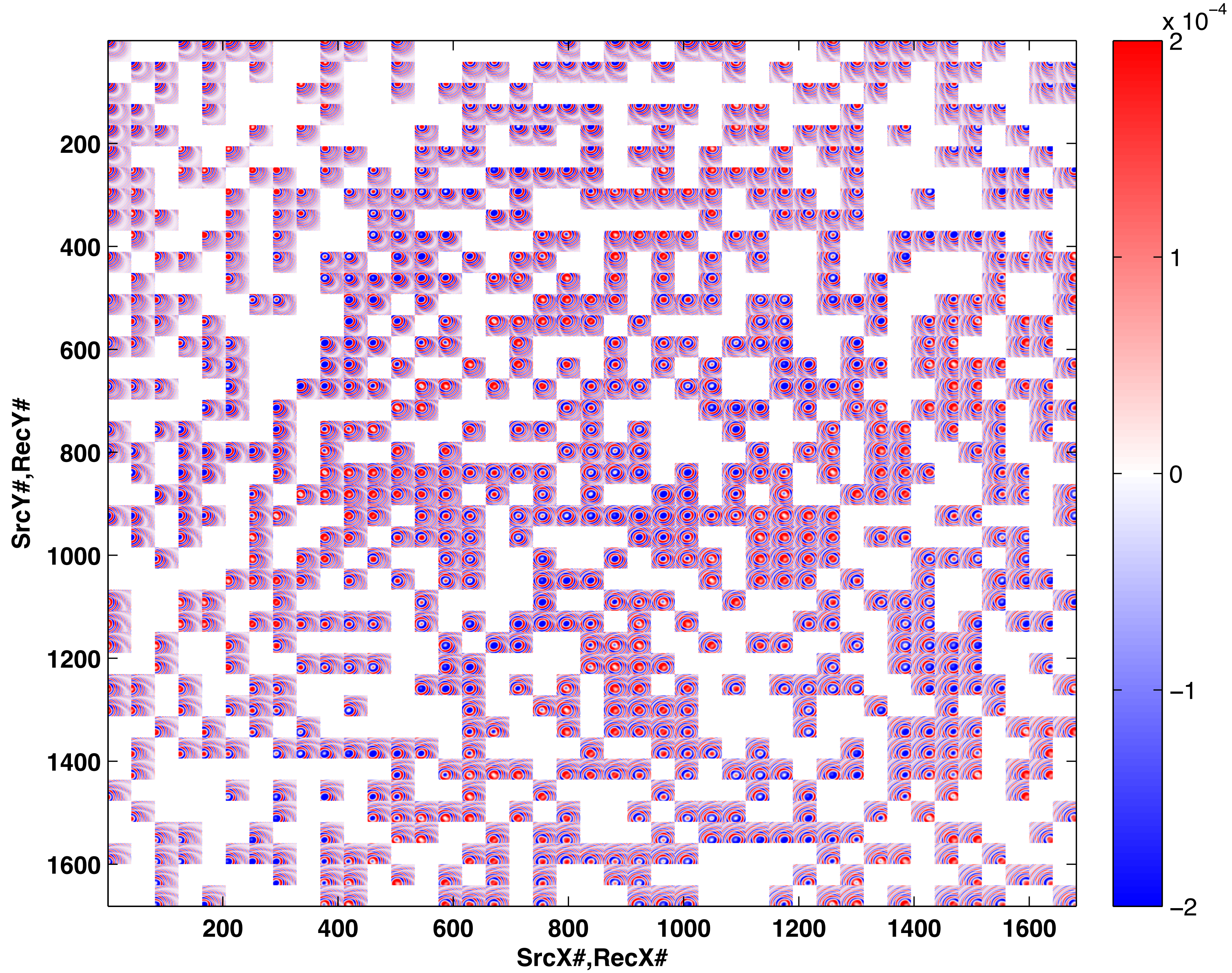
3D Acquisition
[Regular sampled data - “Transform” domain]



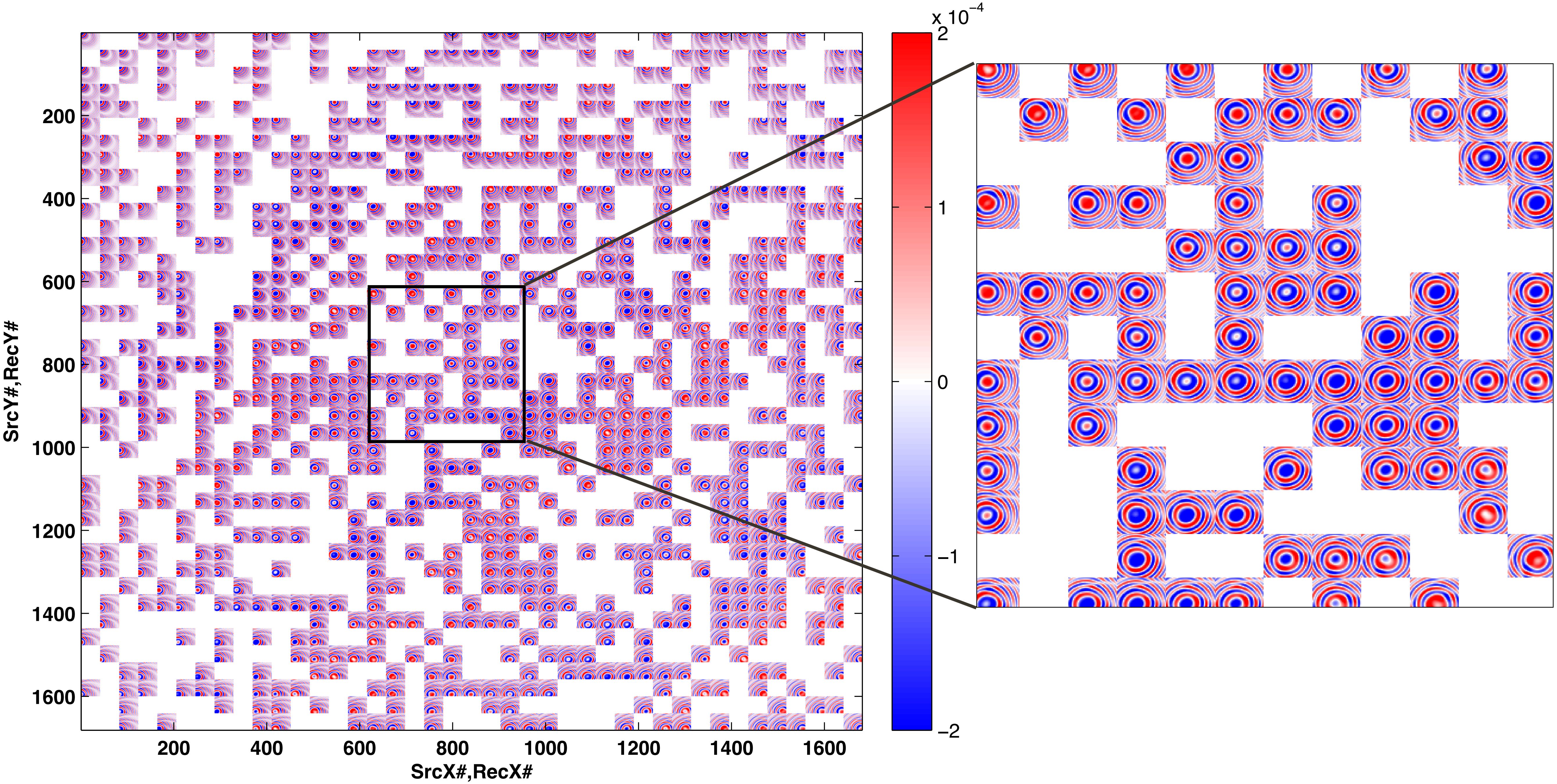
3D Acquisition
[Irregular sampled data]



3D Acquisition
[Irregular sampled data - “Transform” domain]

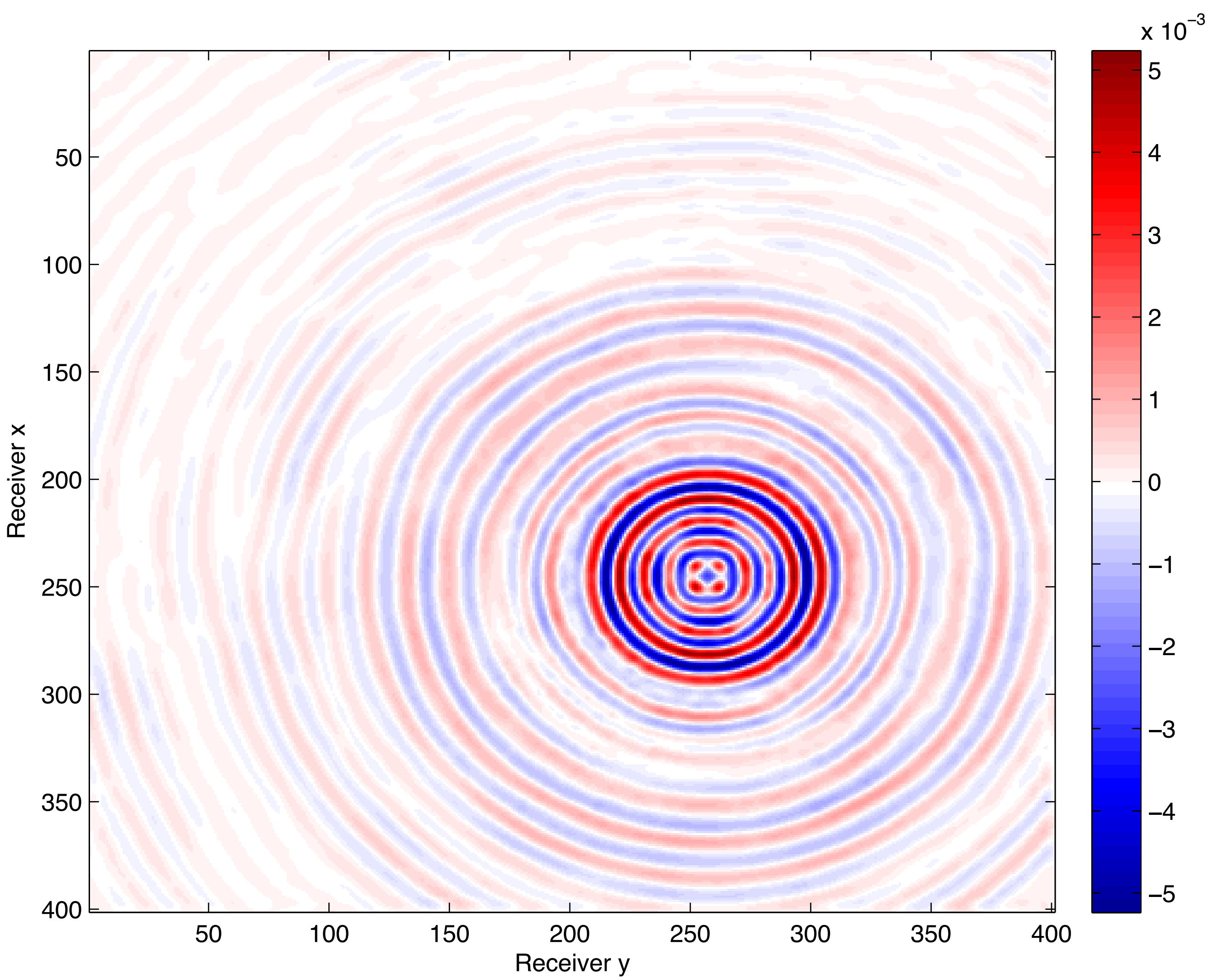


3D Acquisition
[Irregular sampled data - “Transform” domain]



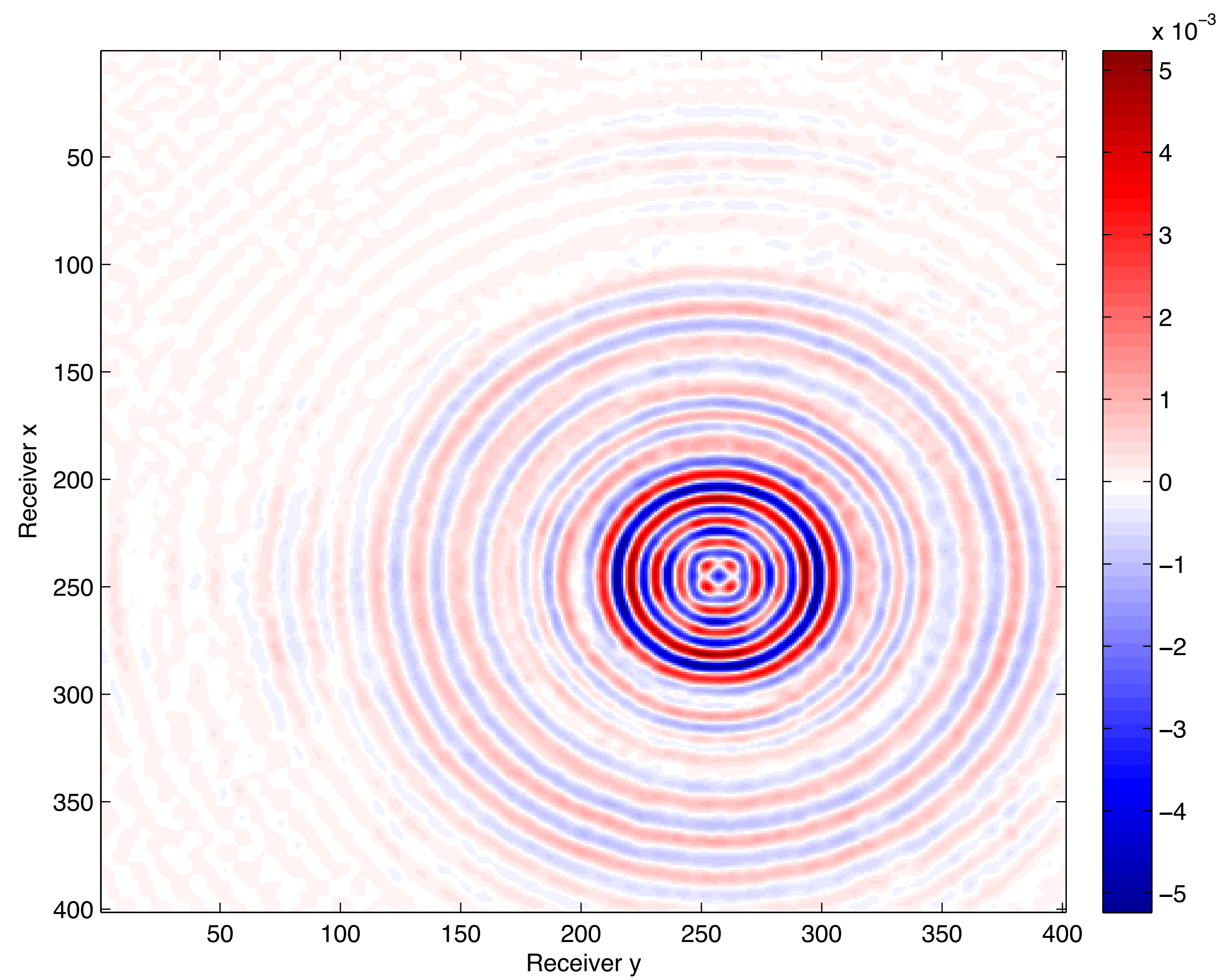
Hierarchical Tucker Interpolant

Reconstruction from 200 shots -> 6400 shots



Known data

(Src x, Src y) = (63,66)

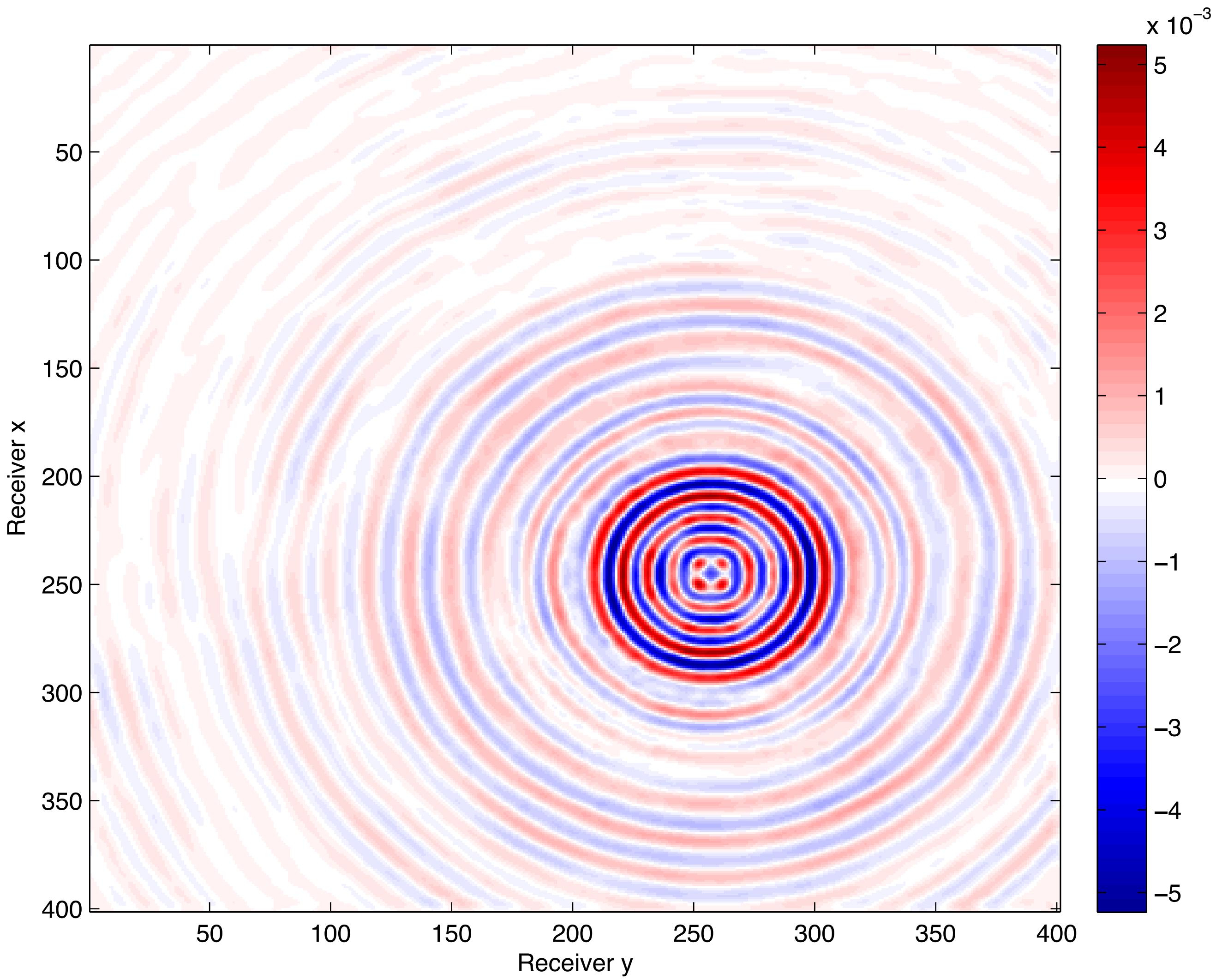


Interpolated data -

SNR 13.2 dB

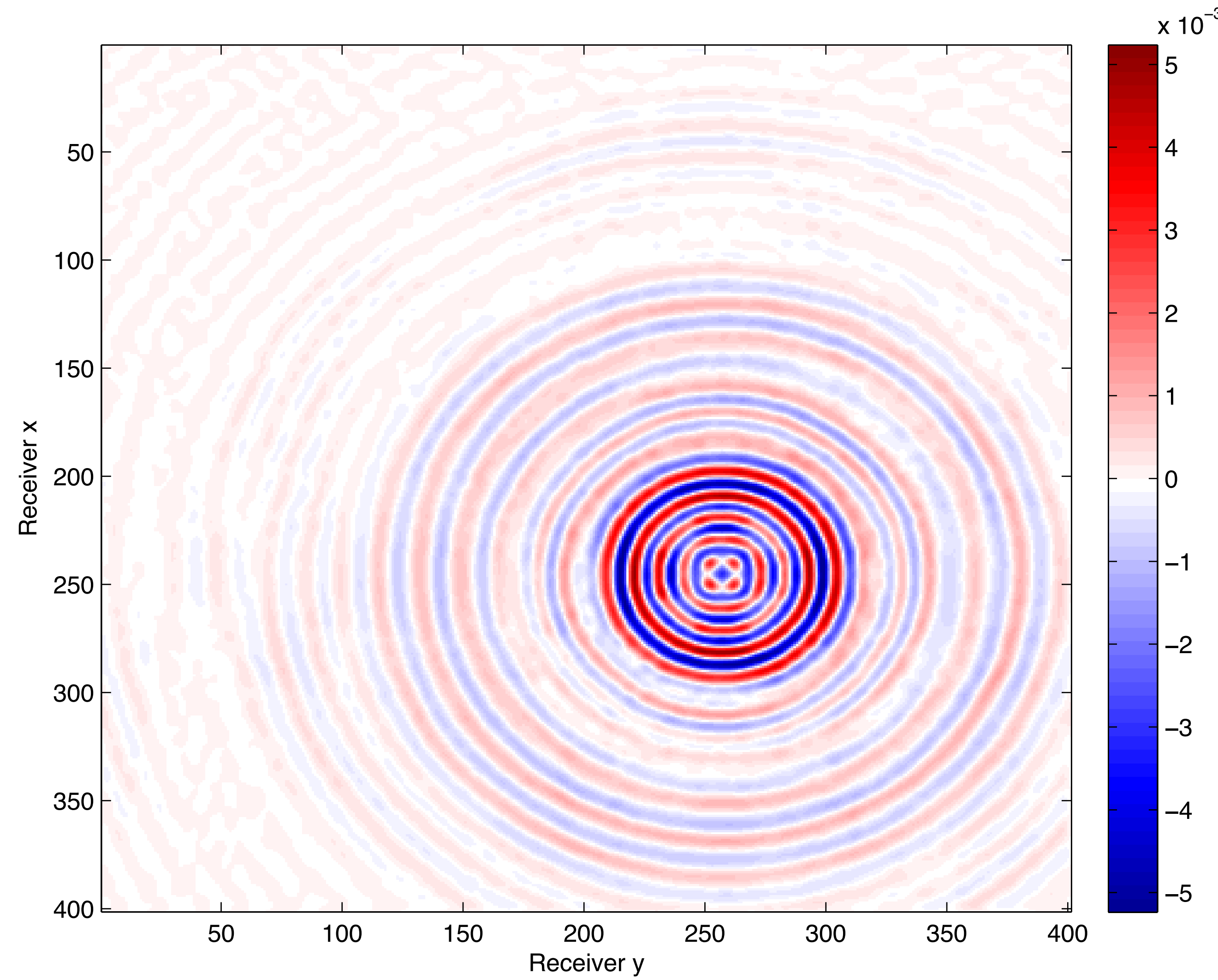
HTuck Interpolant - Regularized

Reconstruction from 200 shots -> 6400 shots



Known data

(Src x, Src y) = (63,66)



Interpolated data -

SNR 15 dB

SLIM

Observations

Acquisition costs *reduced* by *randomization* sampling

- e.g. via *multiple* randomly *dithered* sources

Cost *reduction* at *cost* of *large*-scale *sparsity-promoting* program

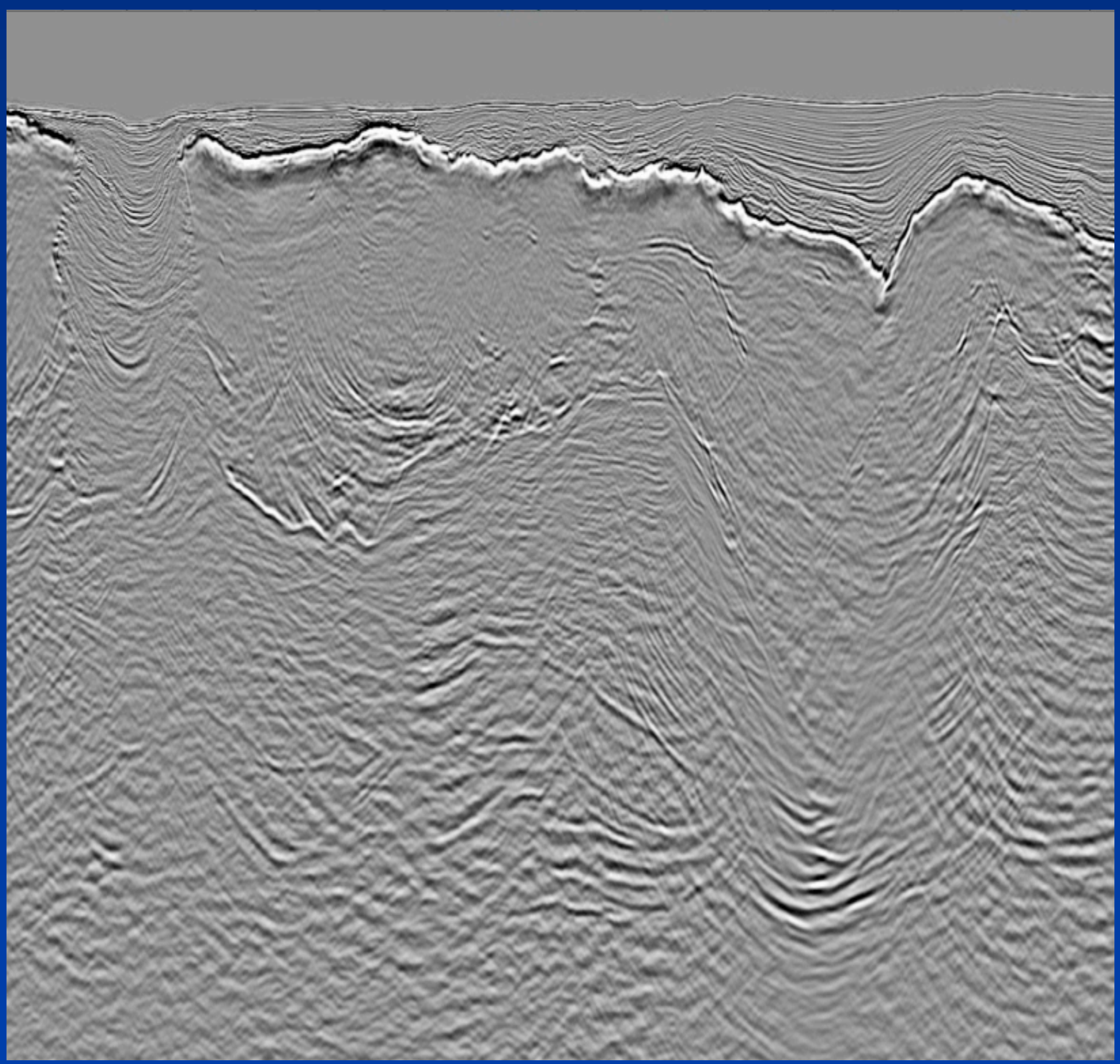
- *dominated* by *sparsifying* transform, which is $\mathcal{O}(n \log n)$
- or by factorization $\mathcal{O}(dN^{d+1})$

We ***win*** because *processing* costs << *acquisition* costs

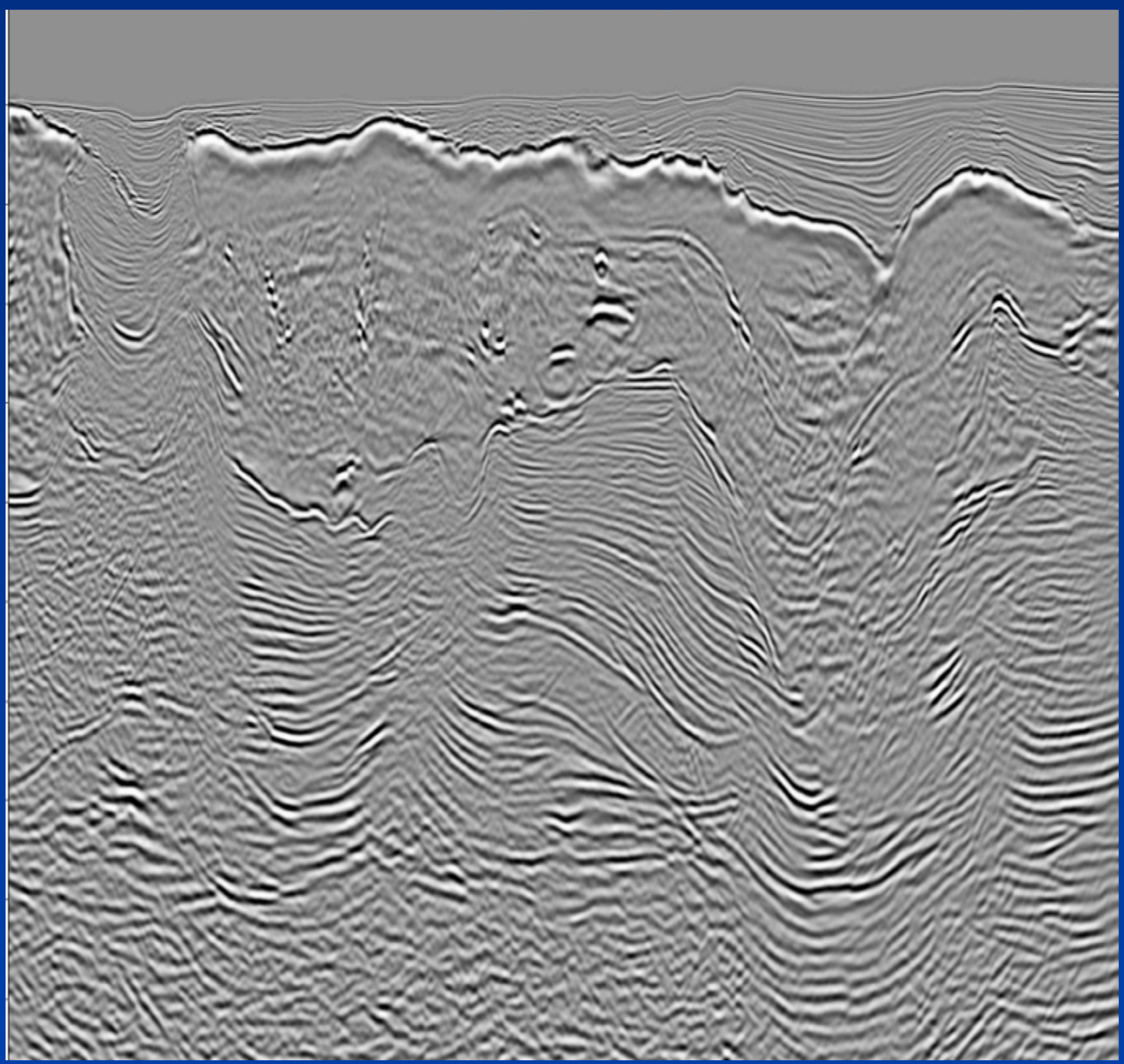
- in 3D *redundancy* & processing *turn-around* times become main *issues*

Subsalt imaging improvements from 2005 to 2010: GSMP, FWI, RTM

2005 technologies
NAZ/SRME/WEM



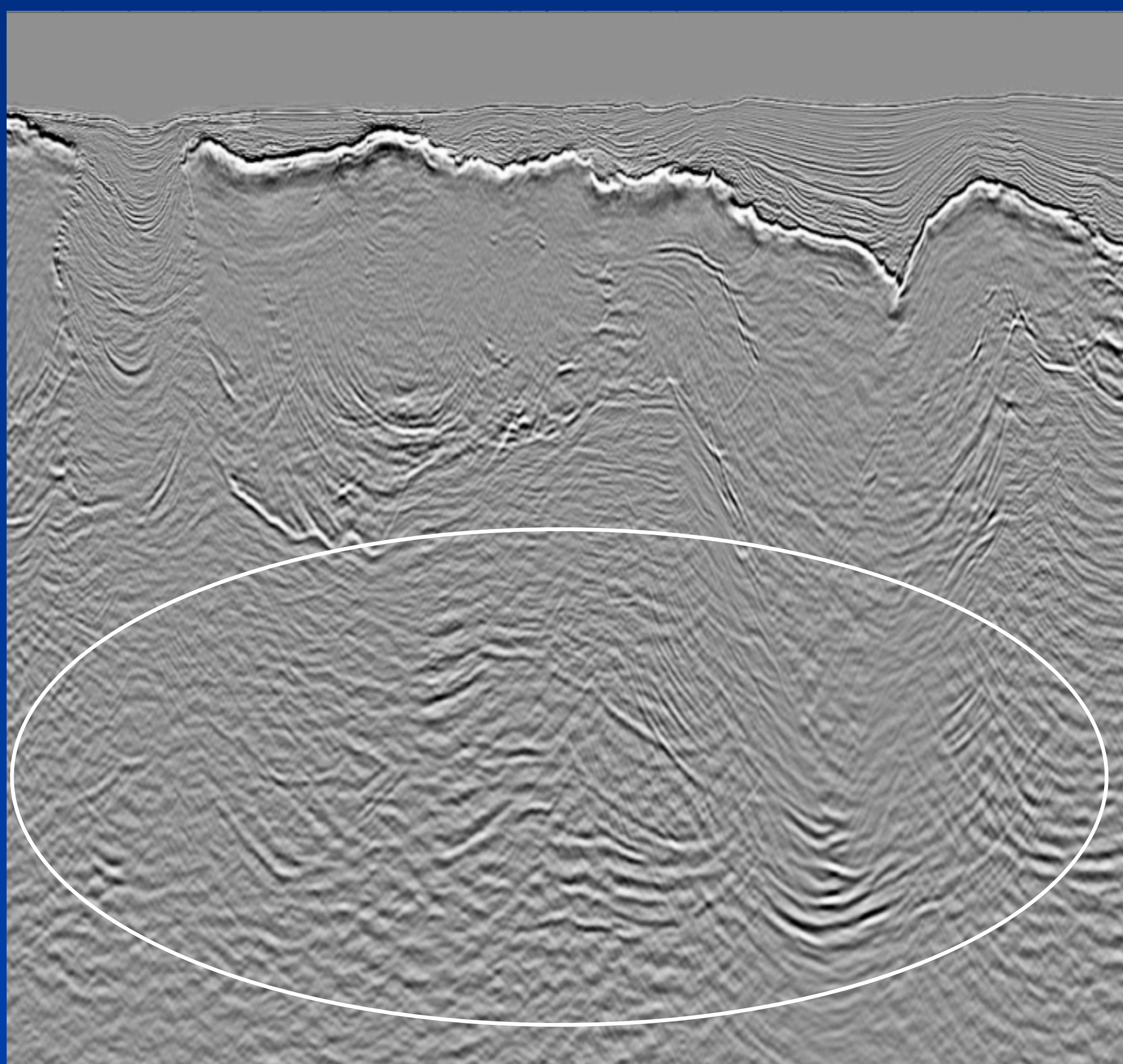
2010 technologies
WAZ/GSMP/FWI/RTM



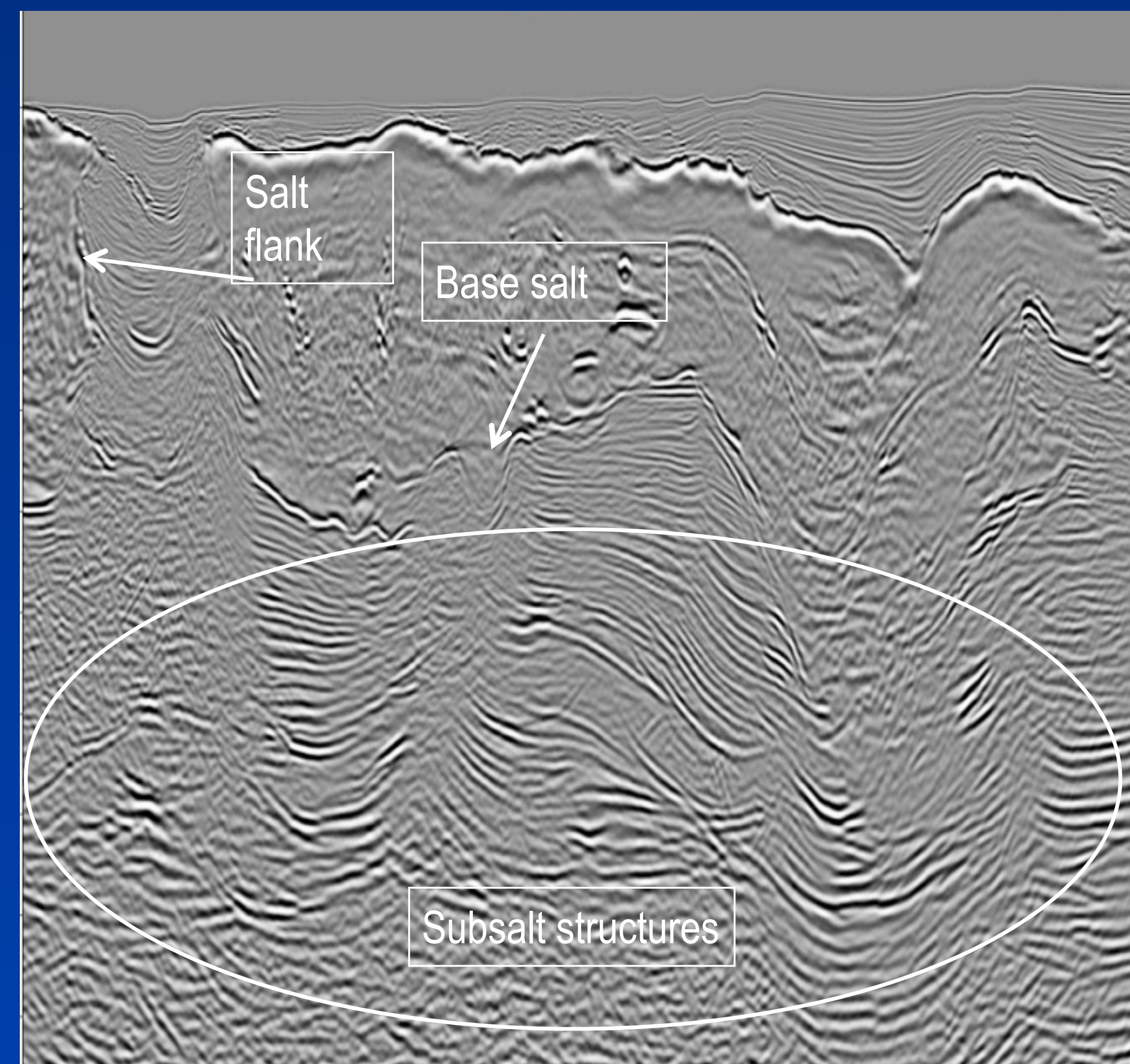
Courtesy Nick Moldoveanu (WesternGeco)

Subsalt imaging improvements from 2005 to 2010: GSMP, FWI, RTM

2005 technologies
NAZ/SRME/WEM

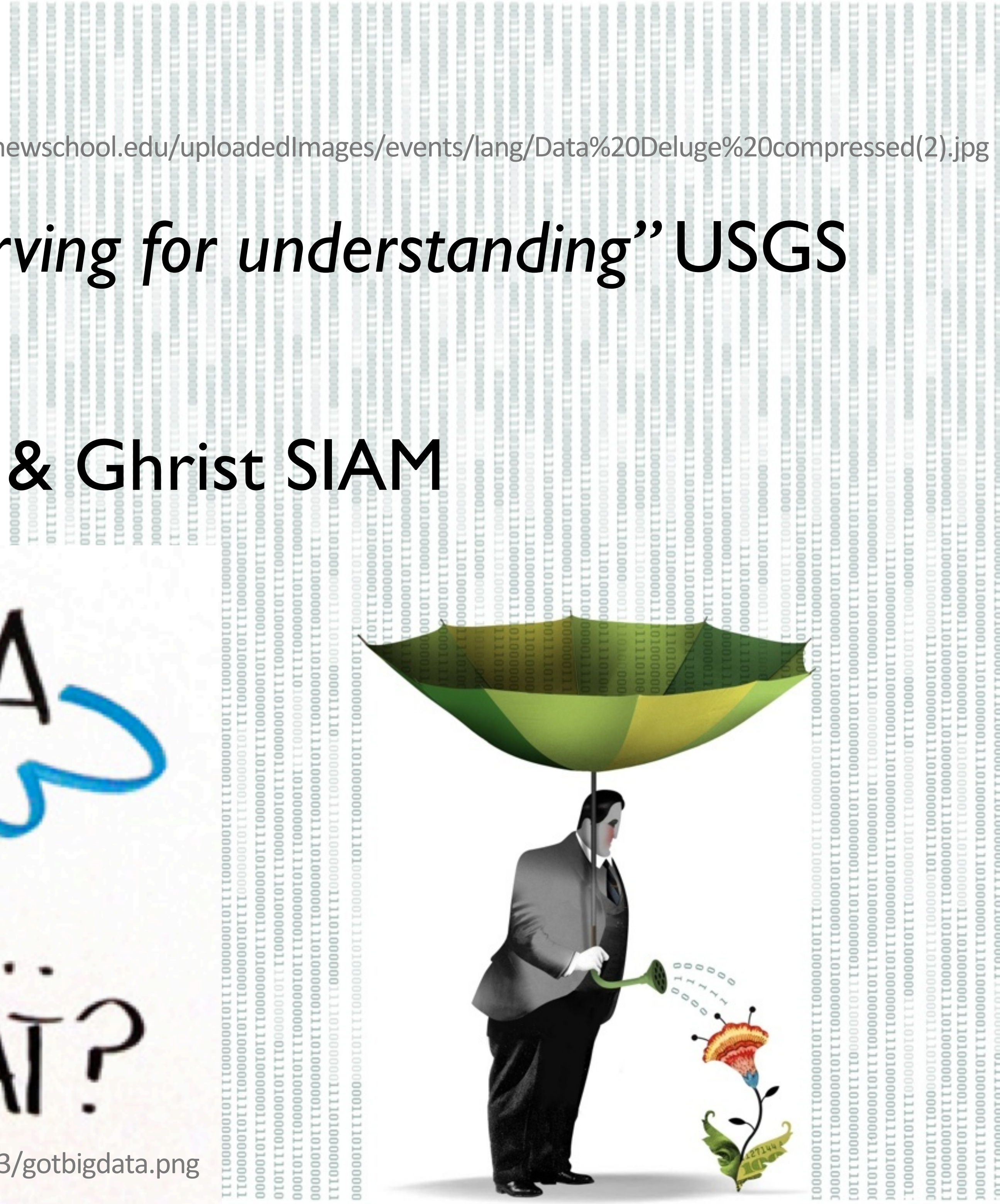


2010 technologies
WAZ/GSMP/FWI/RTM



Courtesy Nick Moldoveanu (WesternGeco)

Big data



http://www.newschool.edu/uploadedImages/events/lang/Data%20Deluge%20compressed(2).jpg

"We are drowning in data but starving for understanding" USGS director Marcia McNutt

"Got data now what" Carlsson & Ghrist SIAM

http://bigdatablog.emc.com/wp-content/uploads/2012/03/gotbigdata.png

Wave-equation based *inversion*

Industry has difficulty *replenishing* produced *resources*

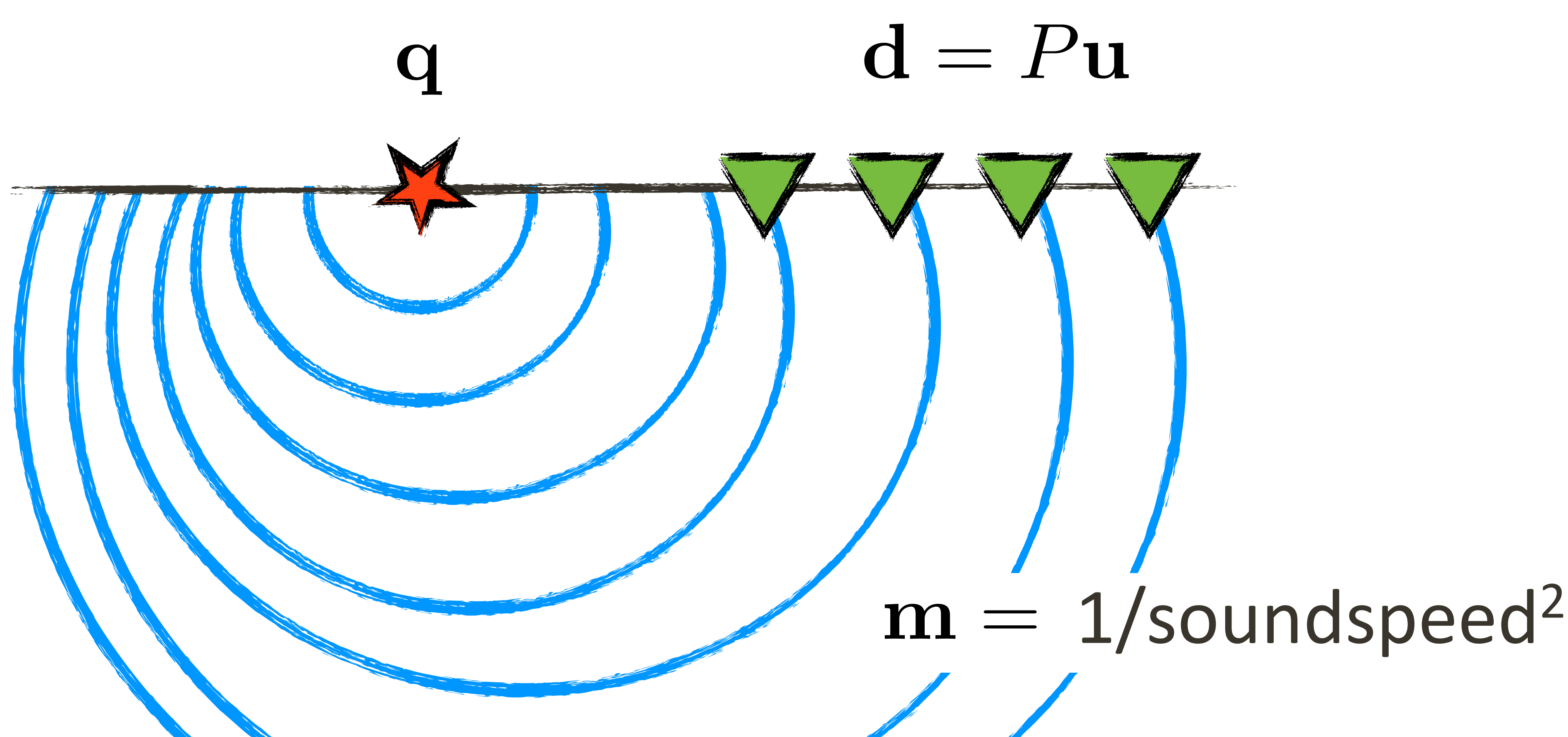
- basically are *no* longer finding oil & gas

Big drive for *transformative wave*-equation based technology

- *PDE* constrained optimization or full-waveform inversion (FWI)

Full-waveform inversion

We model the data in the *acoustic* approximation $(\omega^2 \mathbf{m} + \nabla^2)\mathbf{u} = \mathbf{q}$



Formulation

non-linear least-squares problem:

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) = \sum_{i=1}^{M} ||\mathbf{d}_i - P_i \mathbf{u}_i||_2^2$$

gradient:

$$\frac{\partial \Phi}{\partial m_k} = \sum_{i=1}^{M} \mathbf{u}_i^H \left(\frac{\partial H(\mathbf{m})}{\partial m_k} \right)^H \mathbf{v}_i$$

where:

$$H(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

$$H(\mathbf{m})^H \mathbf{v}_i = P_i^T (\mathbf{d}_i - P_i \mathbf{u}_i)$$

Formulation

non-linear least-squares problem:

$$\min_{\mathbf{m}} \Phi(\mathbf{m}) = \sum_{i=1}^{M} ||\mathbf{d}_i - P_i \mathbf{u}_i||_2^2$$

gradient:

$$\frac{\partial \Phi}{\partial m_k} = \sum_{i=1}^{M} \mathbf{u}_i^H \left(\frac{\partial H(\mathbf{m})}{\partial m_k} \right)^H \mathbf{v}_i$$

where:

$$H(\mathbf{m})\mathbf{u}_i = \mathbf{q}_i$$

$$H(\mathbf{m})^H \mathbf{v}_i = P_i^T (\mathbf{d}_i - P_i \mathbf{u}_i)$$

Inversion of very large sparse linear systems

Batched optimization

$$\min_{\mathbf{m}} \Phi[\mathbf{m}] = \frac{1}{K}\sum_{i=1}^{K} \phi_i[\mathbf{m}]$$

Quasi-Newton approach

$$\mathbf{s}_k = -B_k \nabla \Phi[\mathbf{m}_k]$$

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \lambda_k \mathbf{s}_k$$

But: evaluation of *full* misfit and gradient is very expensive.

Batched optimization

The gradient is the *average*

$$\nabla \Phi = \frac{1}{K} \sum_{i=1}^{K} \nabla \phi_i$$

which we can approximate by

$$\nabla \Phi \approx \nabla \widetilde{\Phi} = \frac{1}{|\mathcal{I}|} \sum_{i \in \mathcal{I}} \nabla \phi_i$$

Optimization

Grow the sample by adding elements

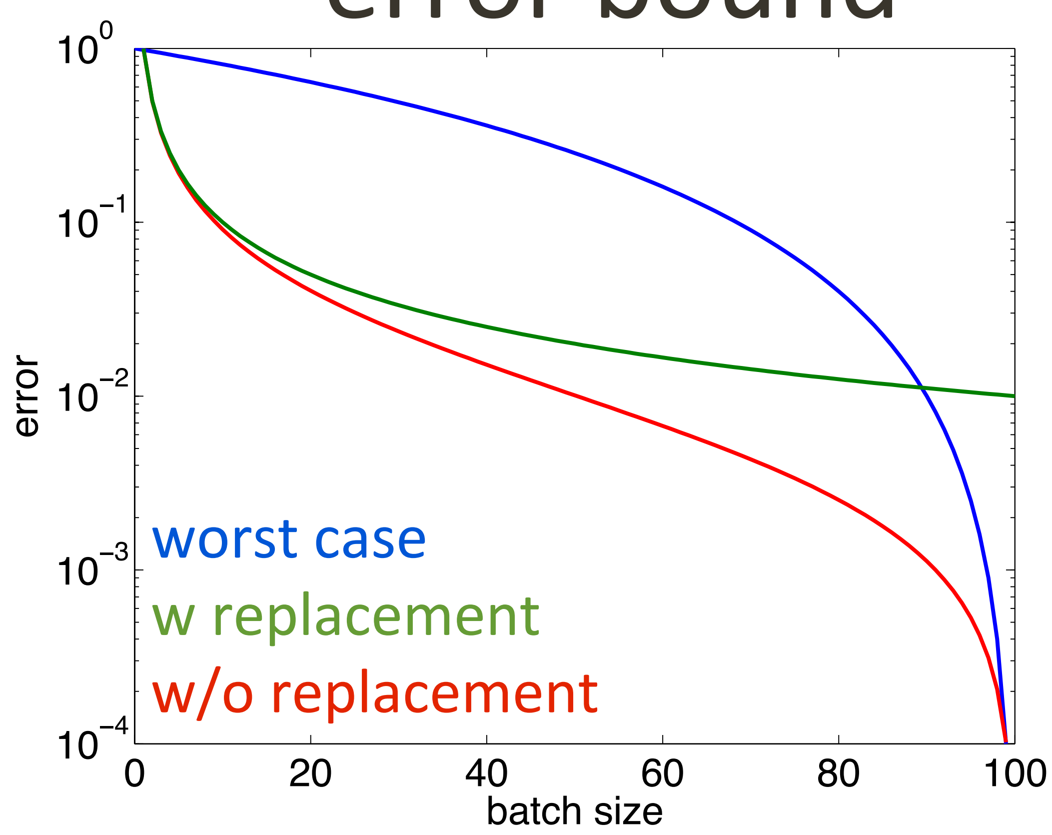
- in a pre-scribed order
- chosen at random *without* replacement
- chosen at random *with* replacement

SLIM

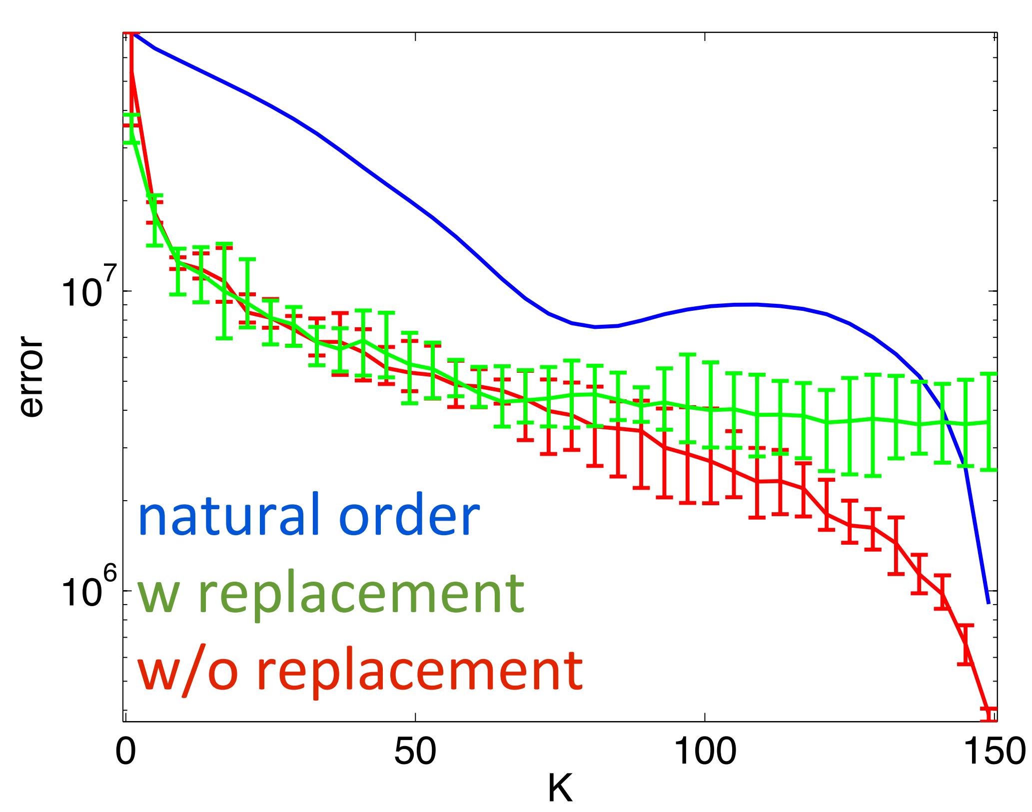
Optimization

Error in the gradient

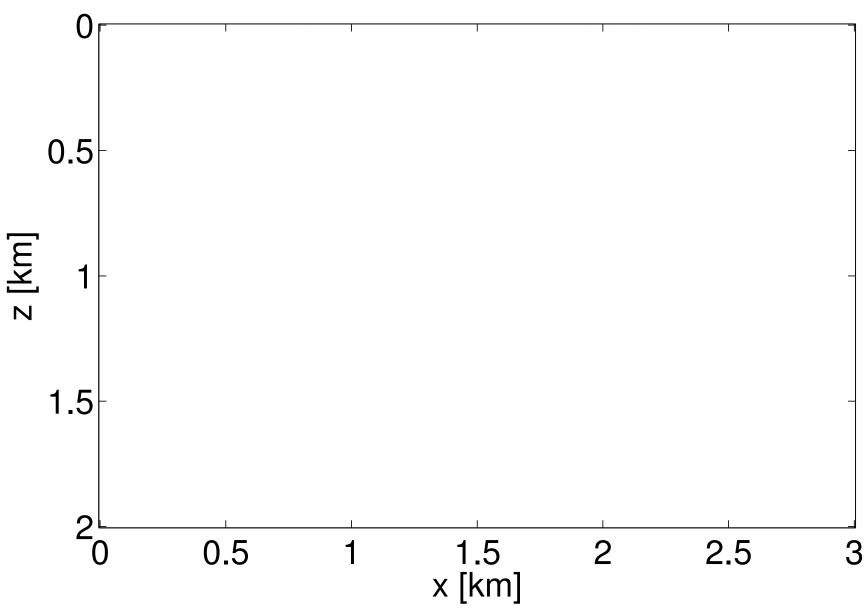
error bound



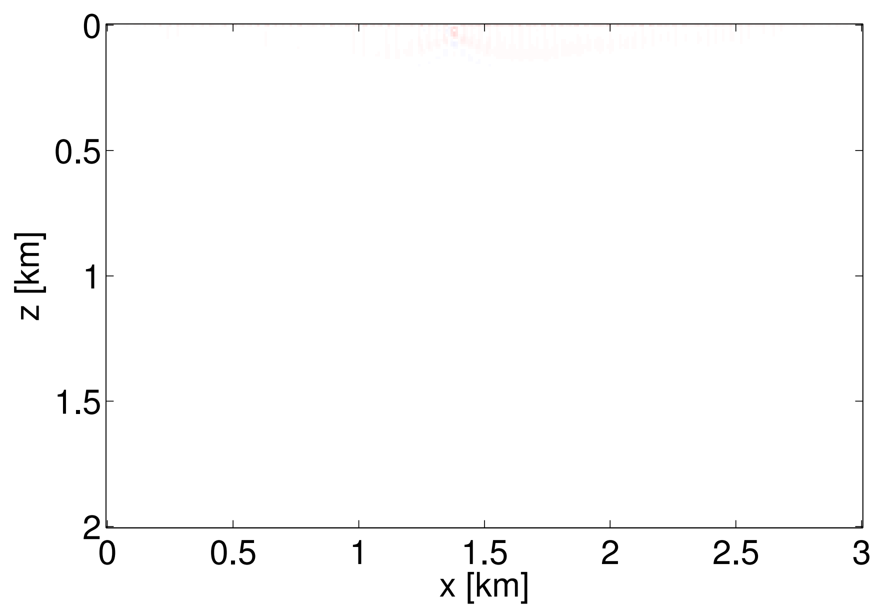
numerical result



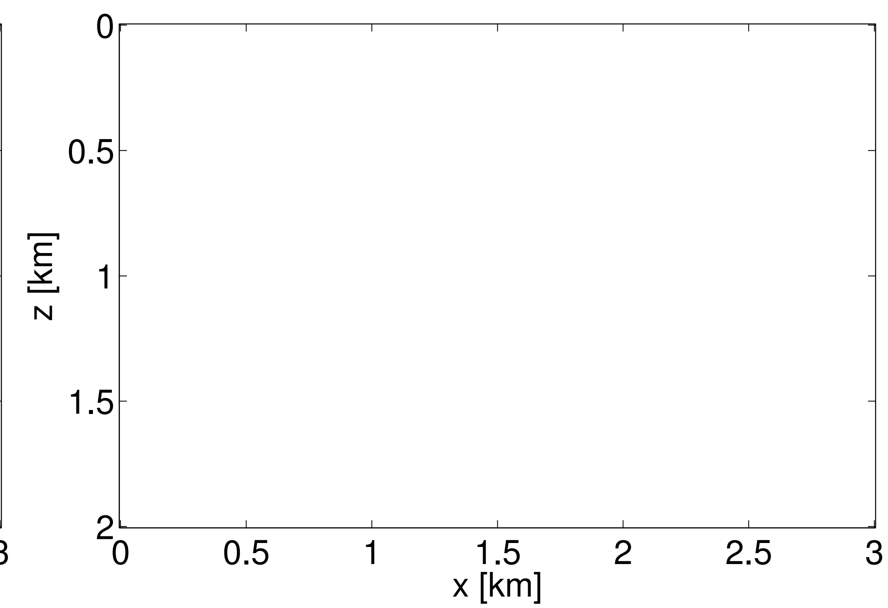
full gradient



incremental gradient

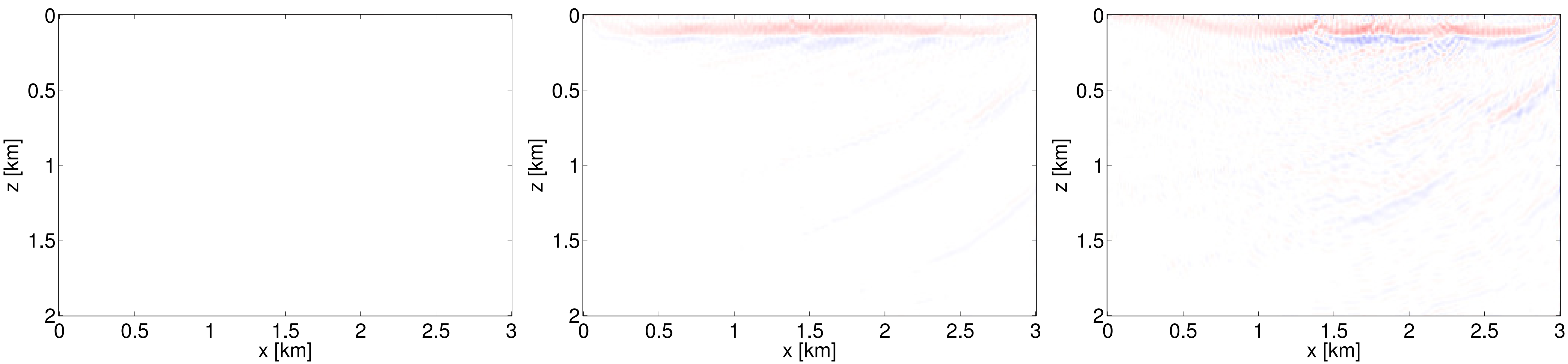


gradient sampling



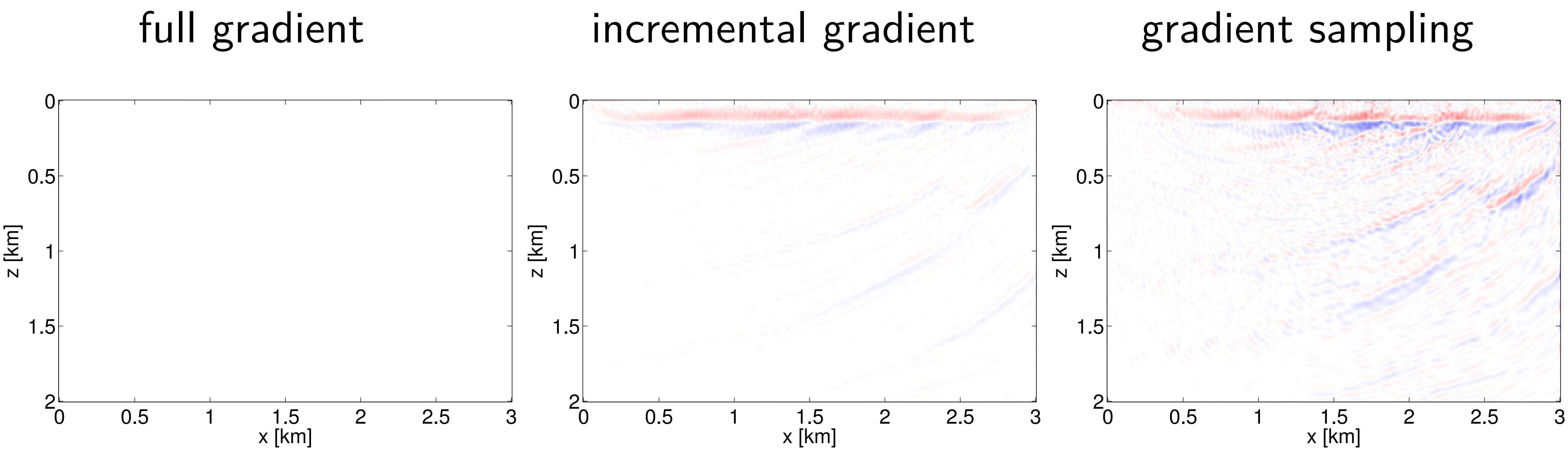
0.01 of 39 passes

full gradient | incremental gradient | gradient sampling

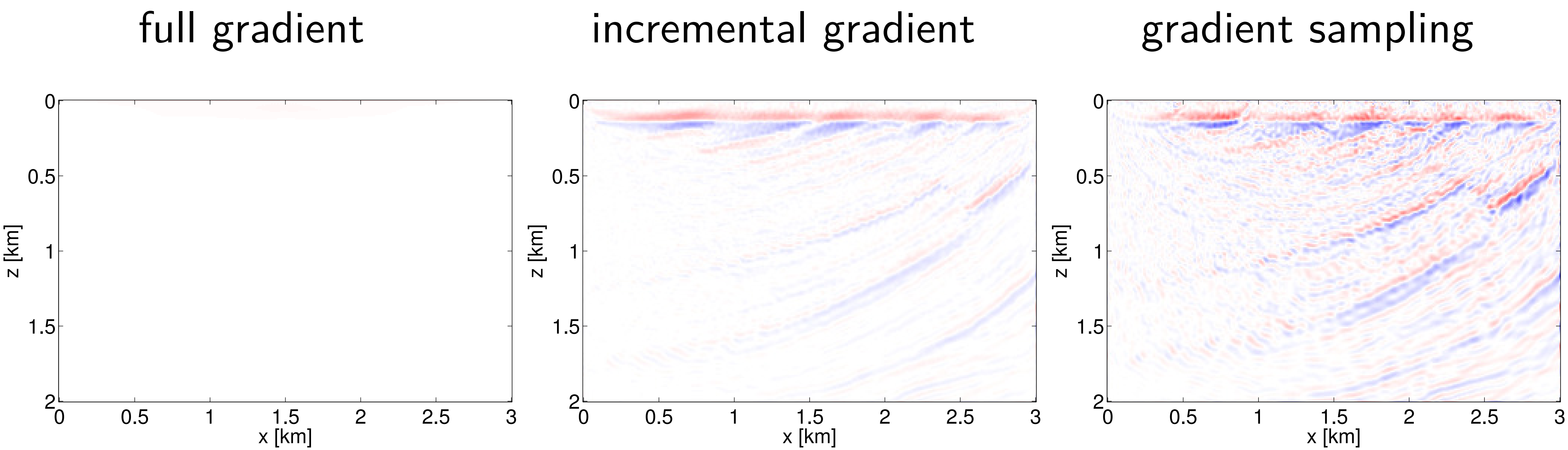


0.4 of 39 passes

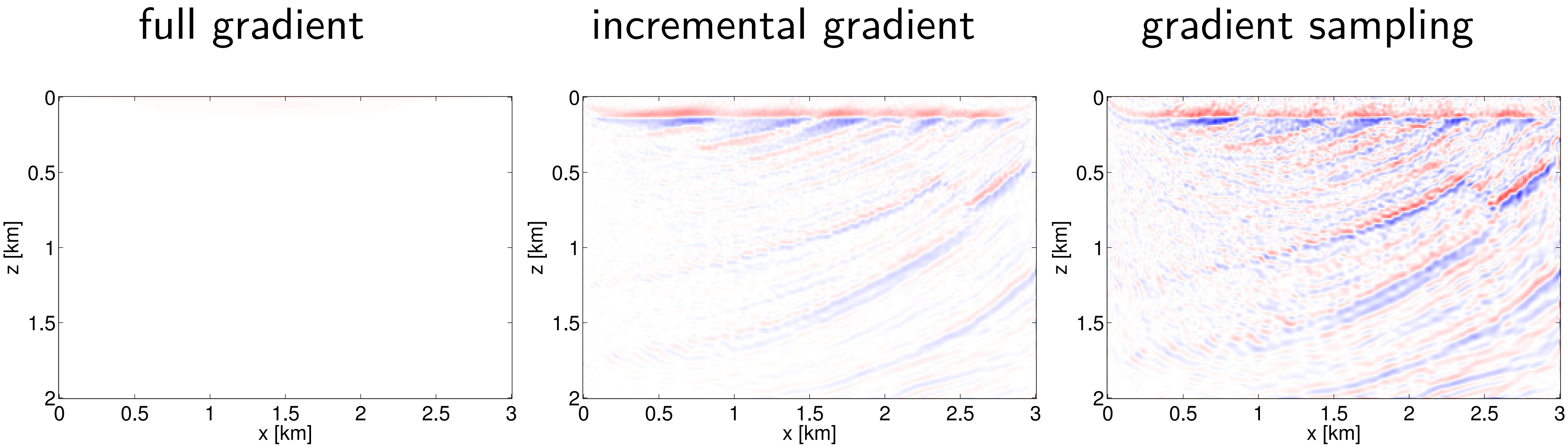
full gradient — incremental gradient — gradient sampling



0.8 of 39 passes

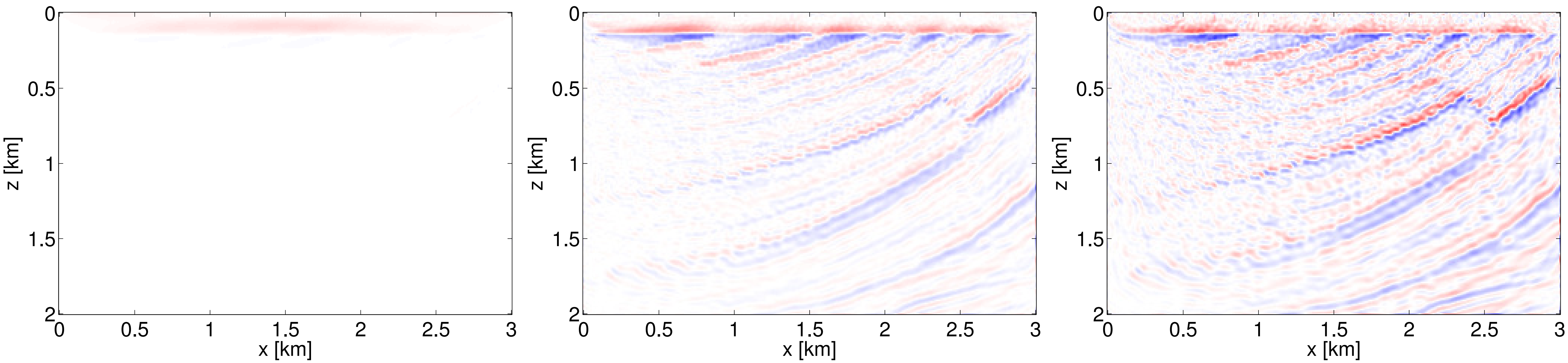


2 of 39 passes



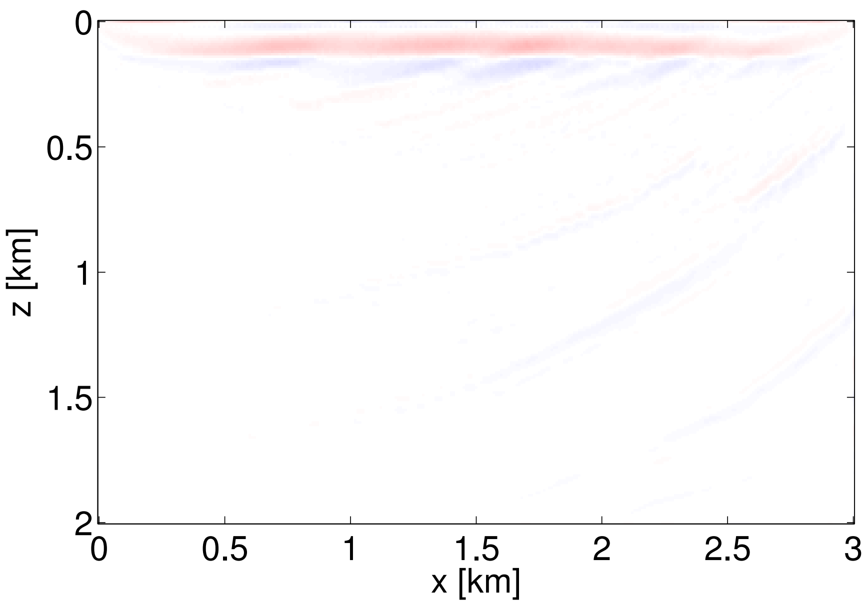
2.6 of 39 passes

full gradient | incremental gradient | gradient sampling

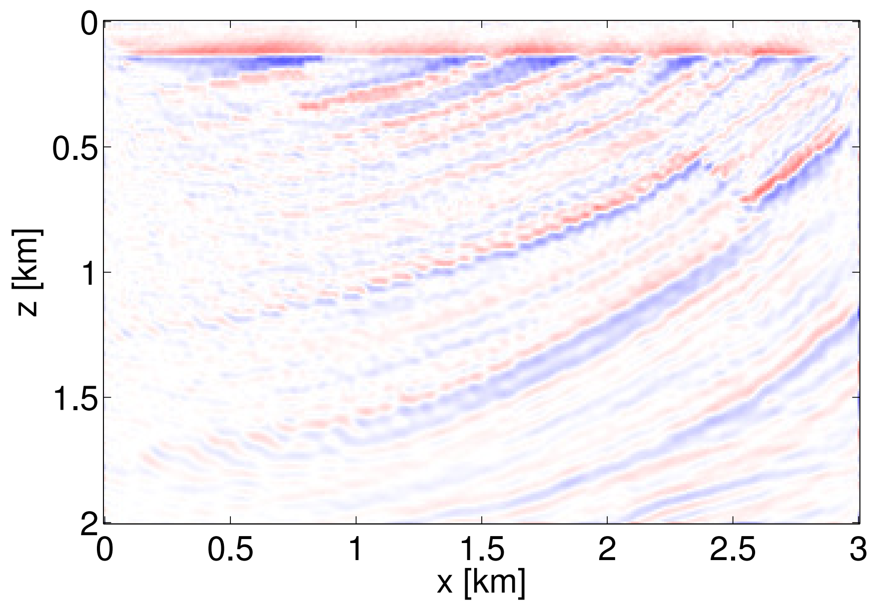


4 of 39 passes

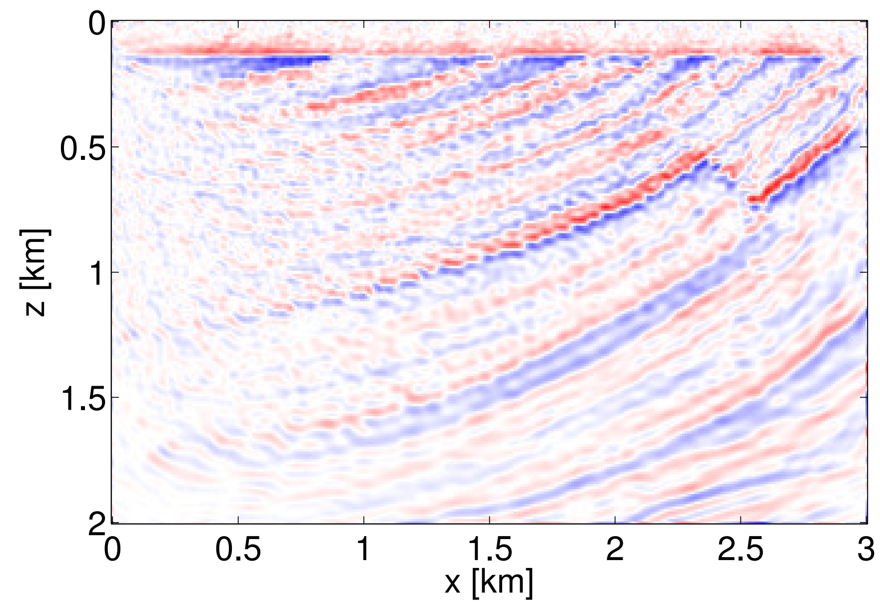
full gradient



incremental gradient

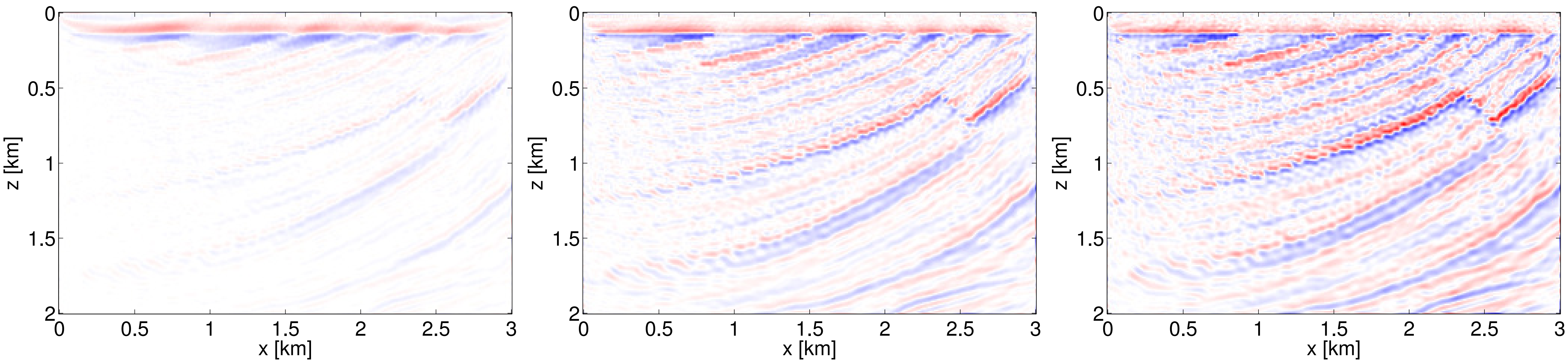


gradient sampling



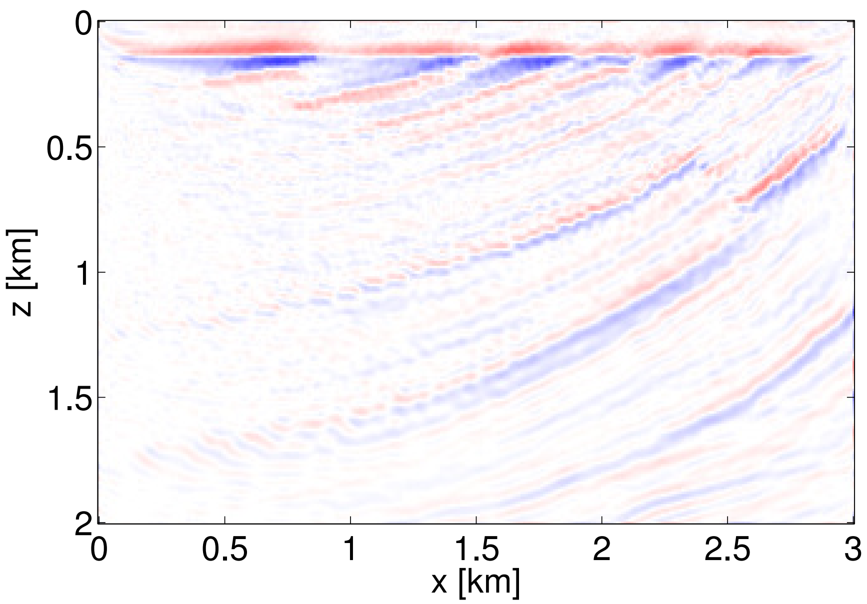
7 of 39 passes

full gradient | incremental gradient | gradient sampling

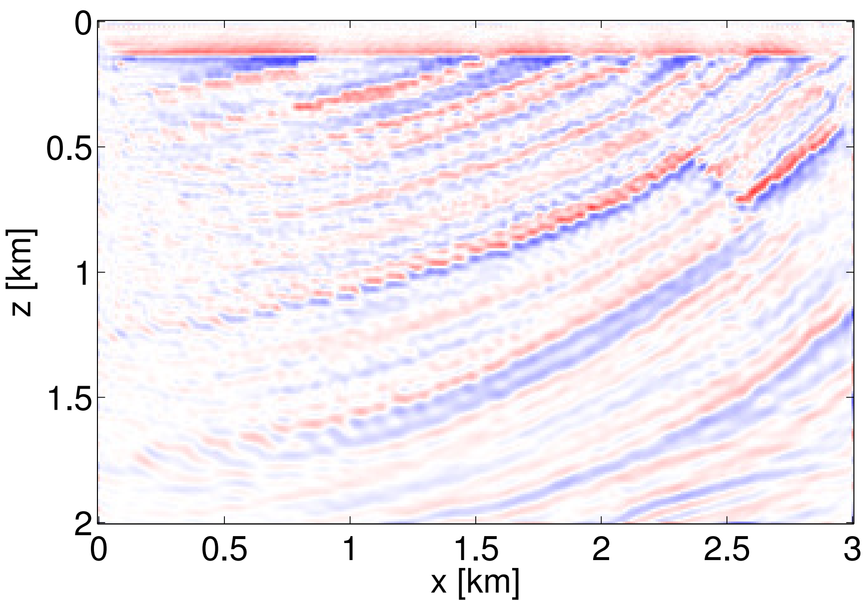


10 of 39 passes

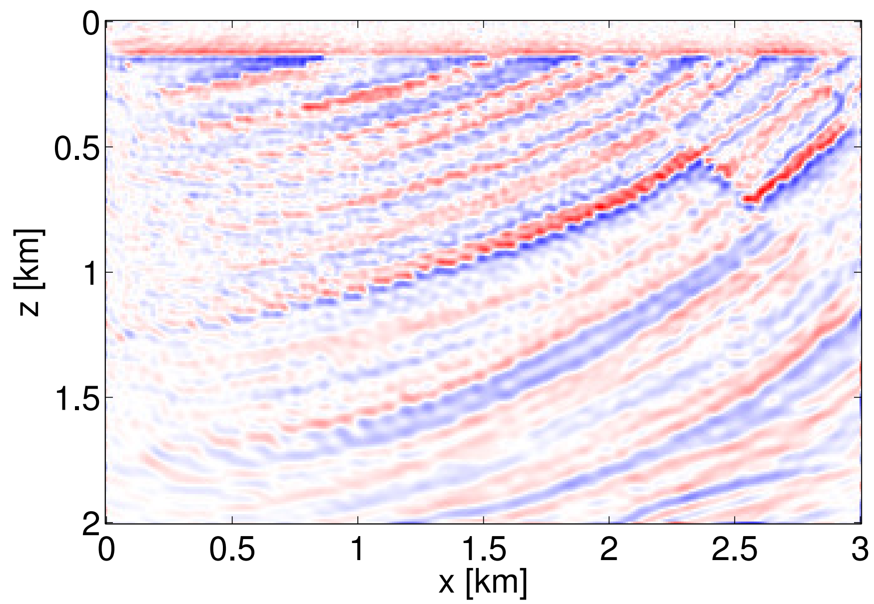
full gradient



incremental gradient

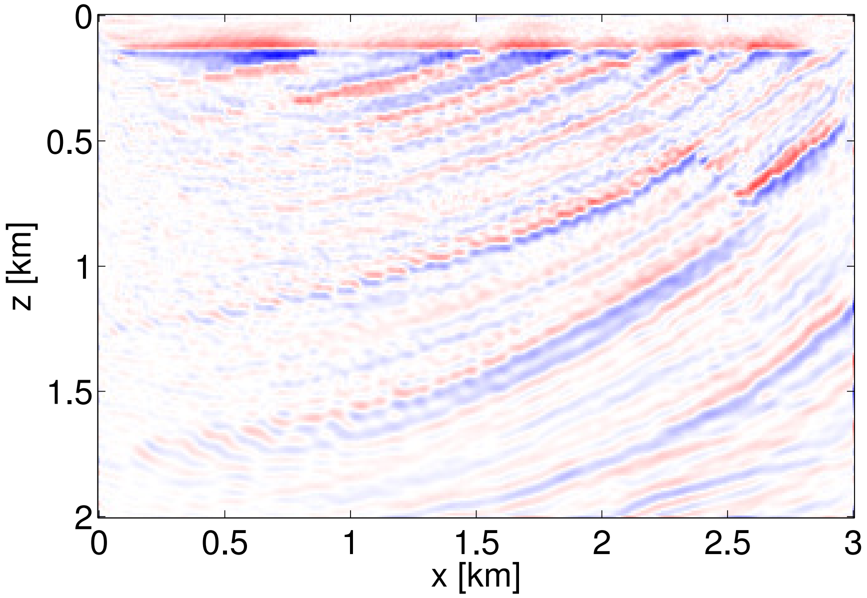


gradient sampling

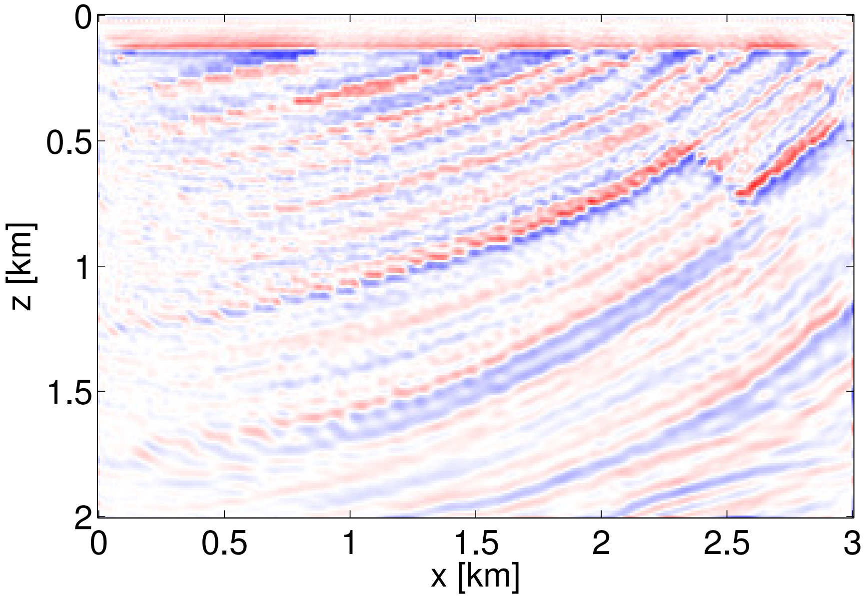


16 of 39 passes

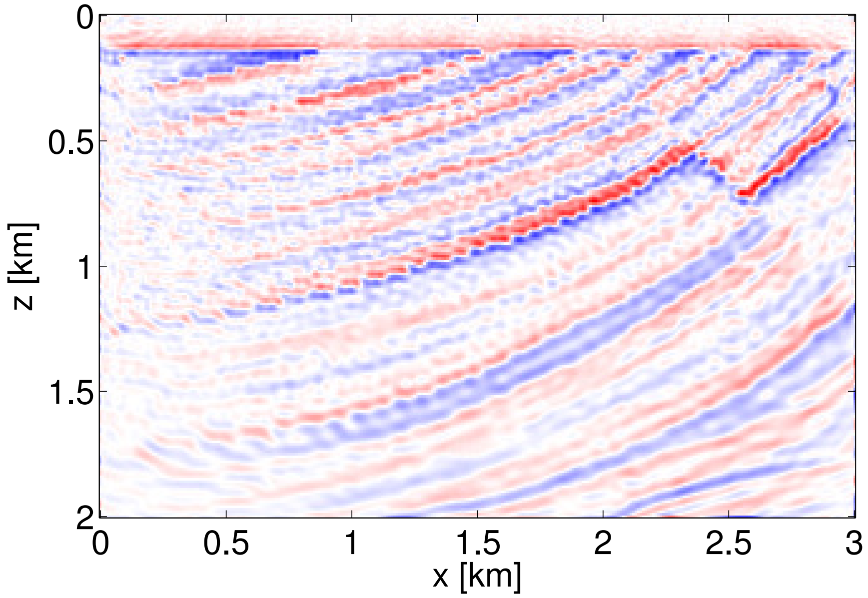
full gradient



incremental gradient

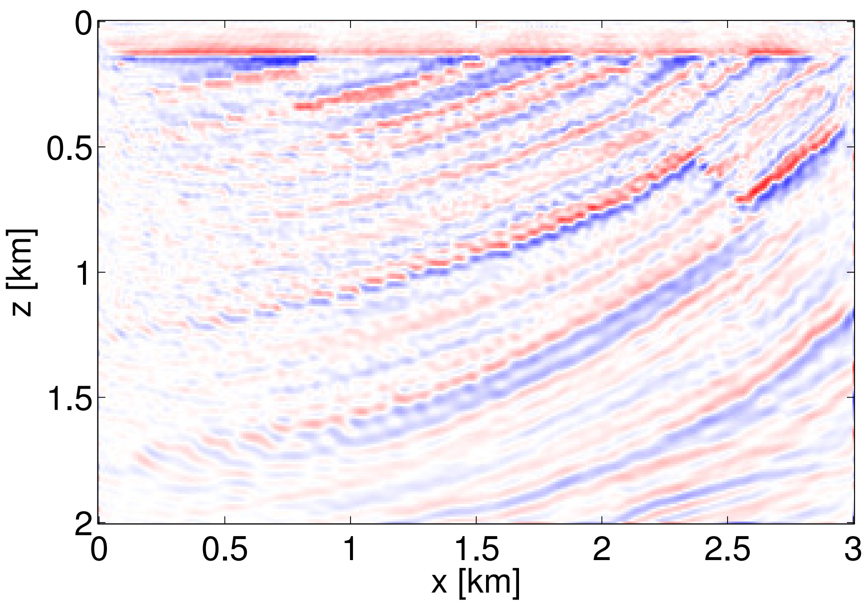


gradient sampling

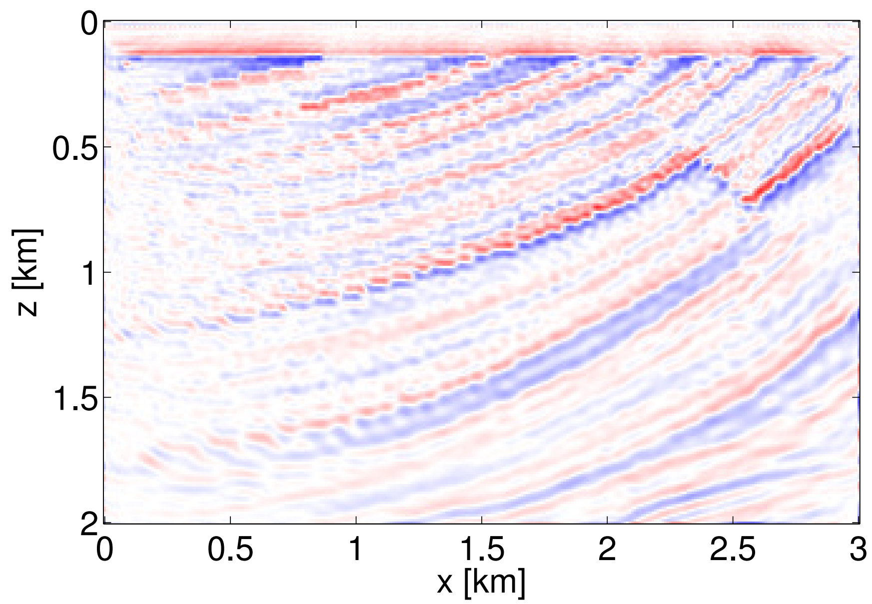


22 of 39 passes

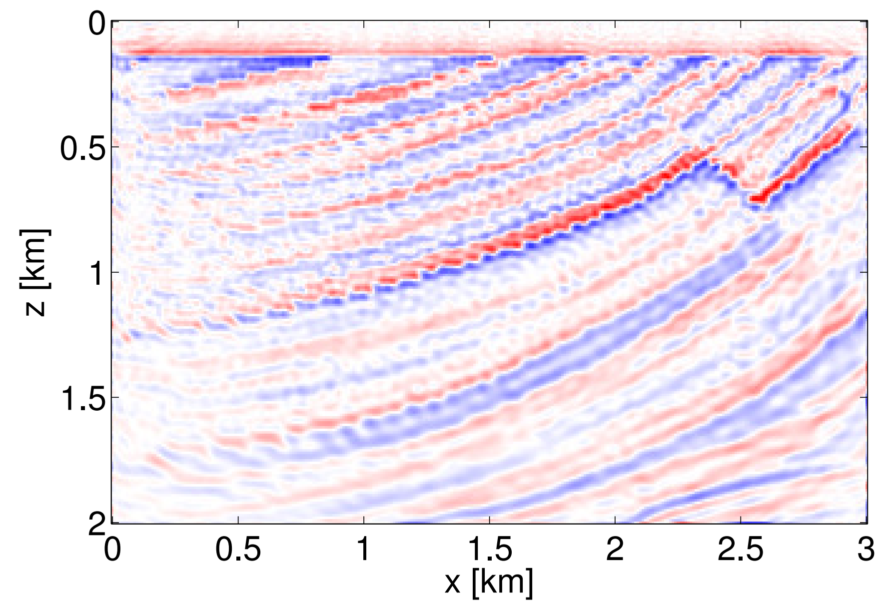
full gradient



incremental gradient

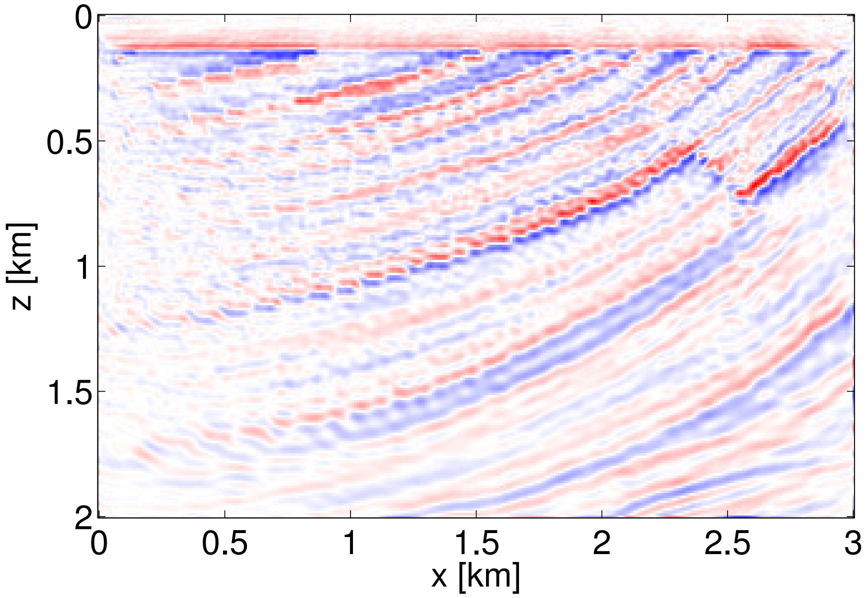


gradient sampling

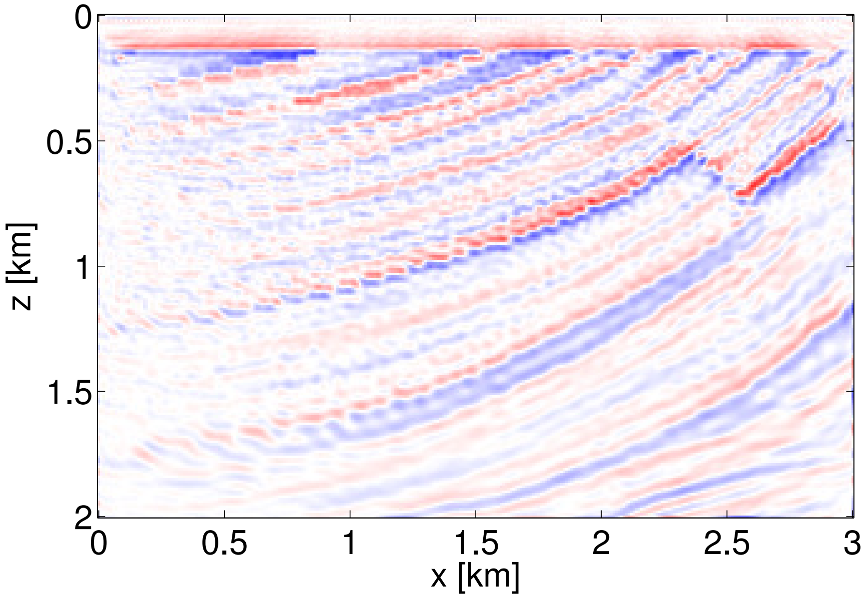


30 of 39 passes

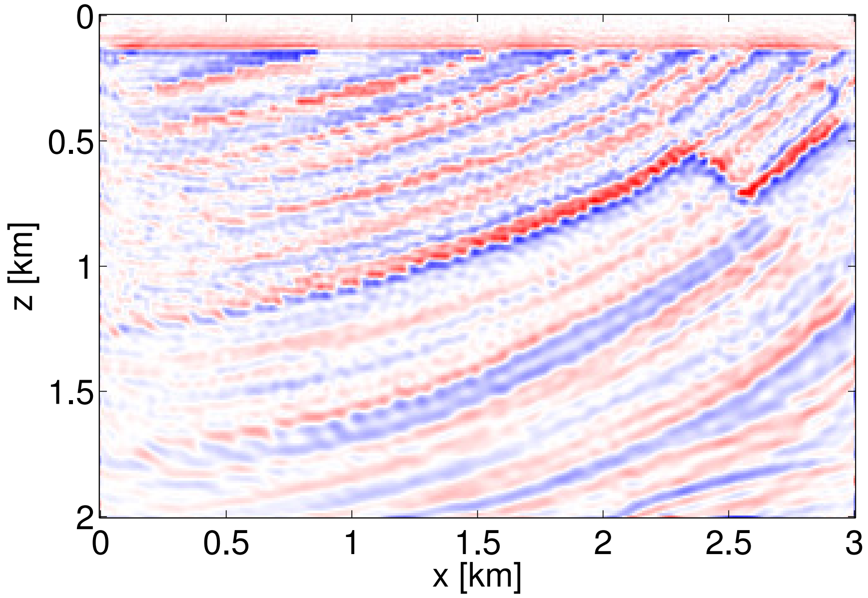
full gradient



incremental gradient



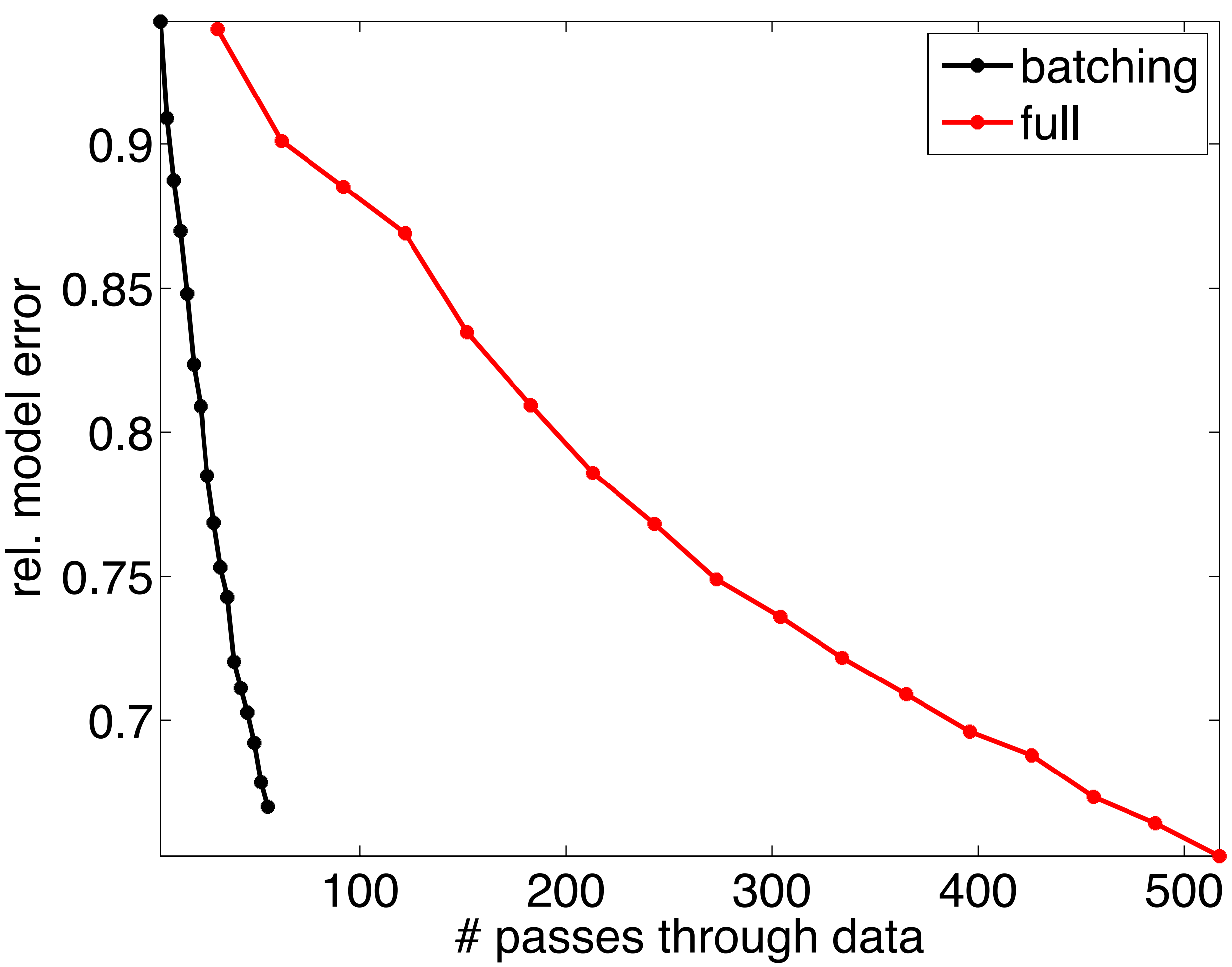
gradient sampling

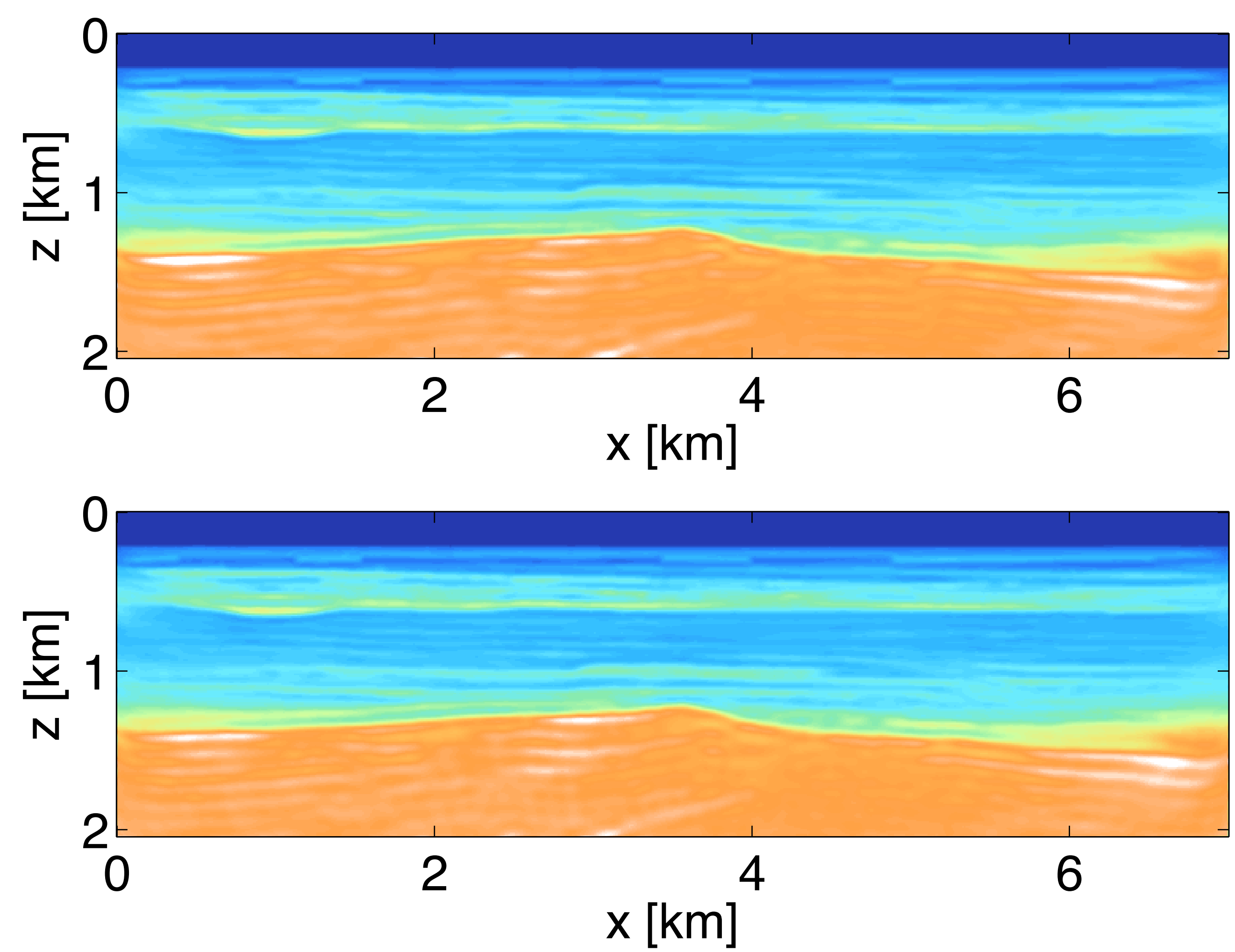


39 of 39 passes

Optimization

10 x speedup





[van Leeuwen et al ’11]

Observations

Our *batching* strategy controls *sampling* and/or *simulation* errors

- by *growing* the *batch* size in *accordance* w/ *convergence* rate
- best of both worlds: *stochastic* versus *deterministic*
- removes noise *sensitivity* of *stochastic* gradients

Can we exploit *sparse* structure of gradient *updates*

- *Dimensionality* reduction w/ Compressive Sensing
- *Acceleration* w/ Approximate Message Passing

Convex composite structure [Burke & Ferris, '95.]

FWI:

$$\min_{\mathbf{m}} \quad \phi(\mathbf{m}) := \underbrace{\frac{1}{2}\| \overbrace{\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]}^{\text{smooth}} \|_F^2}_{\text{convex}}$$

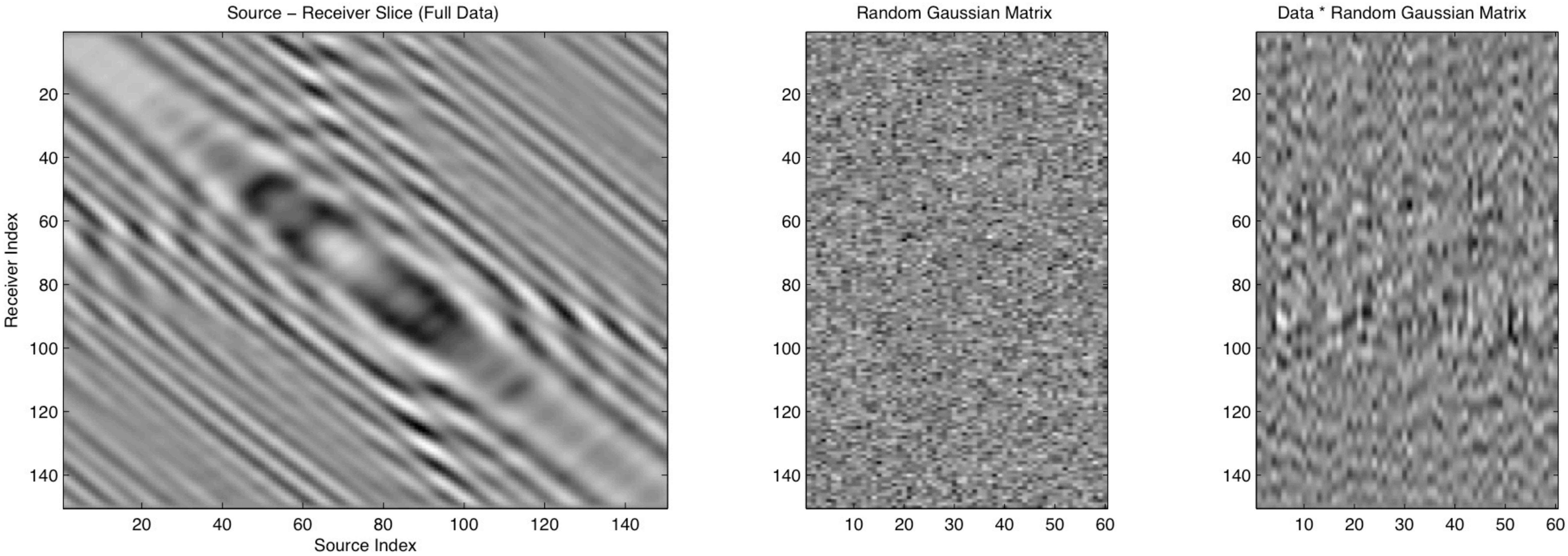
- exploit *convexity* by linearizing *within*

$$\min_{\boldsymbol{\delta}\mathbf{m}} \quad \phi(\boldsymbol{\delta}\mathbf{m}) := \frac{1}{2}\|\mathbf{D} - \boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}] - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m};\mathbf{Q}]\boldsymbol{\delta}\mathbf{m}\|_F^2$$

- control the norm of the updates to *guarantee* convergence

Randomized source aggregates

$\mathbf{D}$ $\mathbf{W}$ $\underline{\mathbf{D}} = \mathbf{D}\mathbf{W}$



Convex optimization

[p=2 or p=1]

Linearized inversion with *randomized supershots:*

$$\delta\widetilde{\mathbf{m}} = \mathbf{S}^* \mathop{\arg\min}_{\delta\mathbf{x}} \|\delta\mathbf{x}\|_{\ell_p} \quad \text{subject to} \quad \| \underbrace{\delta\underline{\mathbf{d}}}_{\mathbf{b}} - \underbrace{\overbrace{\nabla\mathcal{F}[\mathbf{m}_0;\underline{\mathbf{Q}}]}^{\text{lin. modelling}}\mathbf{S}^*}_{\mathbf{A}} \delta\mathbf{x}\|_2 \leq \sigma$$

$$\begin{aligned} \delta\mathbf{x} &= \text{Sparse curvelet-coefficient vector} \\ \mathbf{S}^* &= \text{Curvelet synthesis} \\ \underline{\mathbf{Q}} &= \text{Simultaneous sources} \\ \delta\underline{\mathbf{d}} &= \text{Super shots} \end{aligned}$$

Fast Gauss-Newton step

[*via* stochastic optimization]

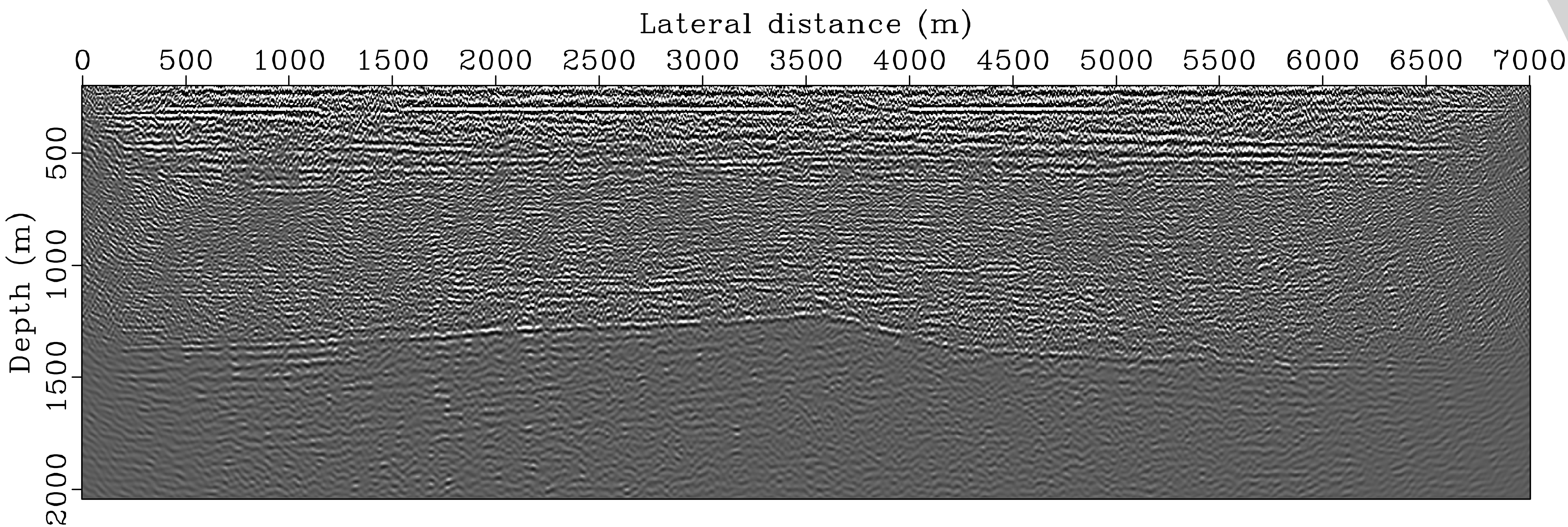
*E*xploit multi-experiment *redundancy* of seismic data volumes by *re*randomized sampling

- *regularly* draw *independent* subsets of shot *aggregates*
- cancels *crosstalk/interference* by *rerandomization*

Heuristic of *current* phase-encoding/*dimensionality* reduction for imaging/FWI

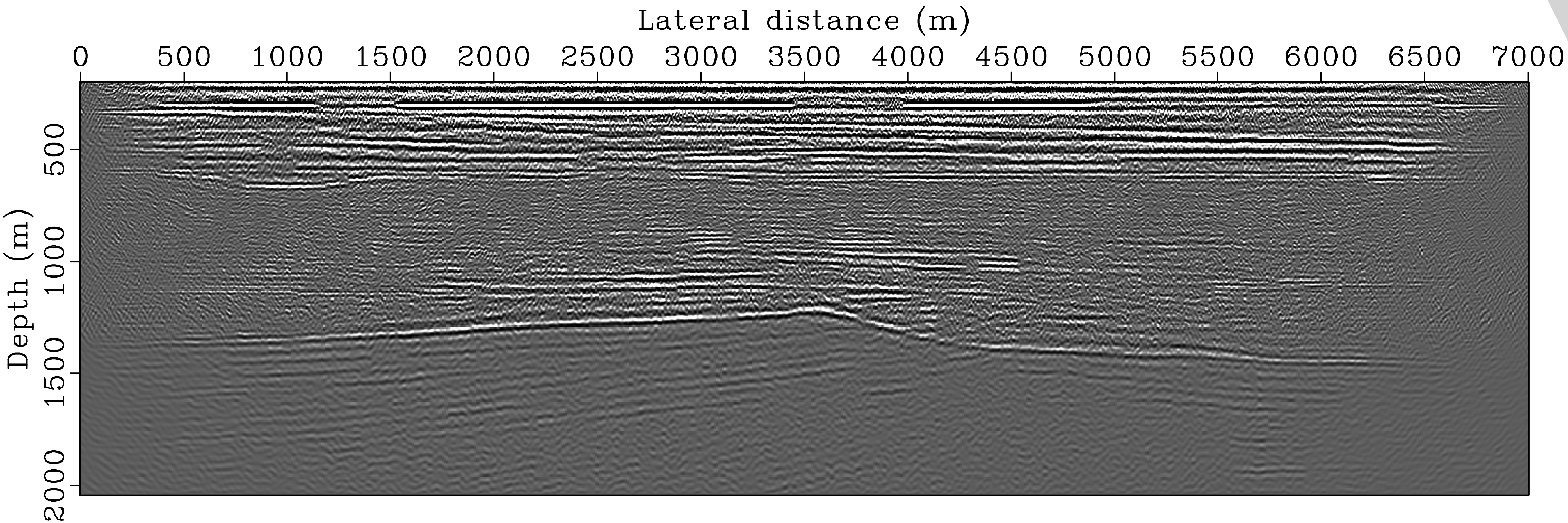
Fast Gauss-Newton step

[ℓ_2 w/o rerandomization 3 *super* shots]



Fast Gauss-Newton step

[ℓ_2 *w/ rerandomization* 3 *super* shots]



Fast Gauss-Newton step
[*via* compressive sensing]

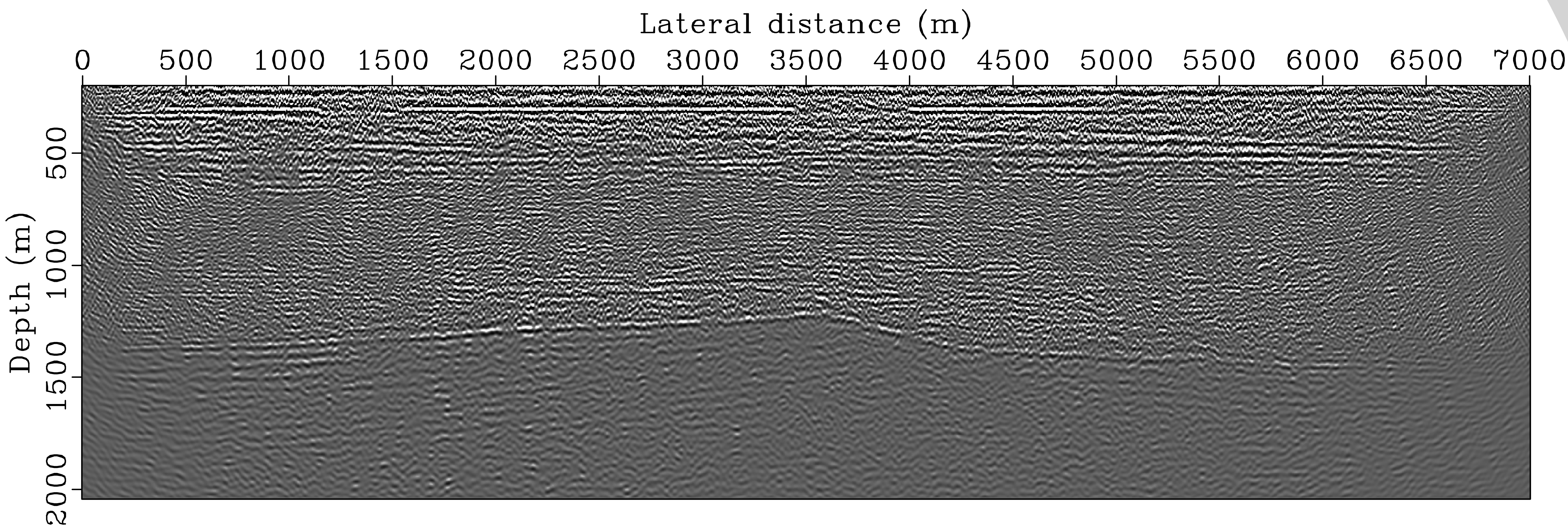
Randomized sampling turns *coherent* source crosstalk/ interferences into ***non-sparse*** *incoherent* noise

Exploits transform-domain *structure* exhibited *within* GN *updates*

- leverage *curvelet*-domain ***sparsity*** promotion
- *map* “noisy” crosstalk/interferences to *coherent* reflectors

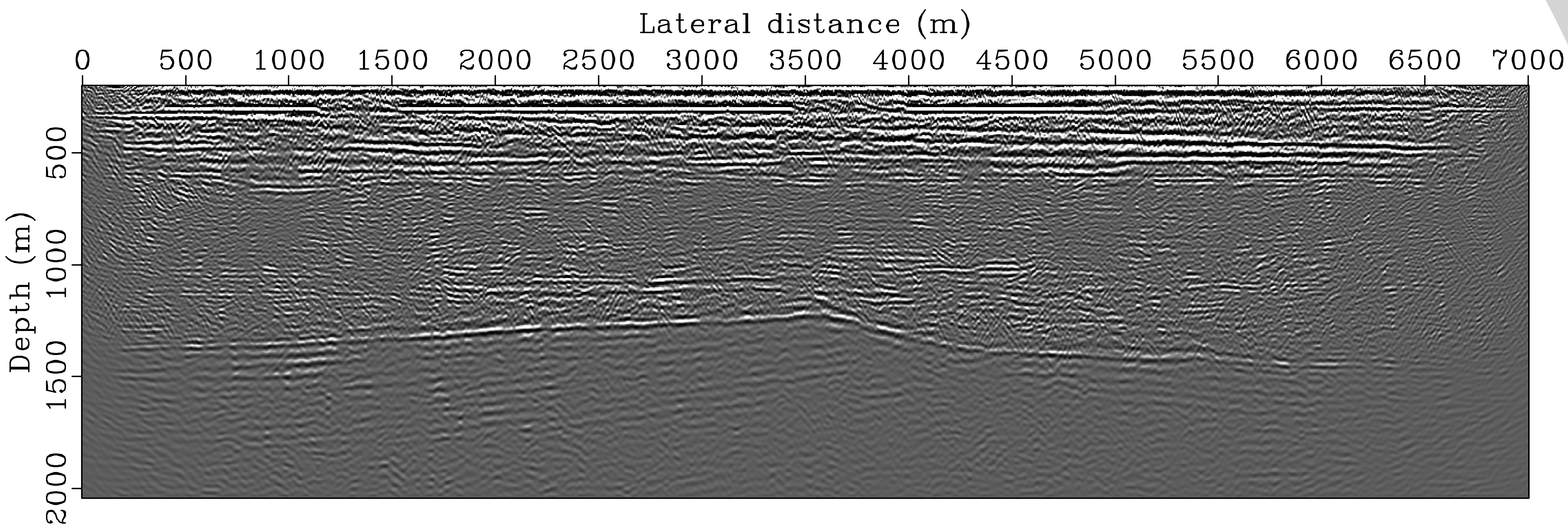
Fast Gauss-Newton step

[ℓ_2 3 *super* shots]



Fast Gauss-Newton step

[ℓ_1 3 *super* shots]



Observations
[w/ reasonable PDE solve budget]

*R*erandomization and *curvelet*-domain *sparsity* promotion:

- *partly* eliminate "*noisy*" *crosstalk*
- *fail* to *remove* "small" *incoherent* crosstalk

Can we somehow combine these two methods?

- *continuation* method for large-scale *convex* optimization
- use *insights* from *approximate* message passing

Supercooling

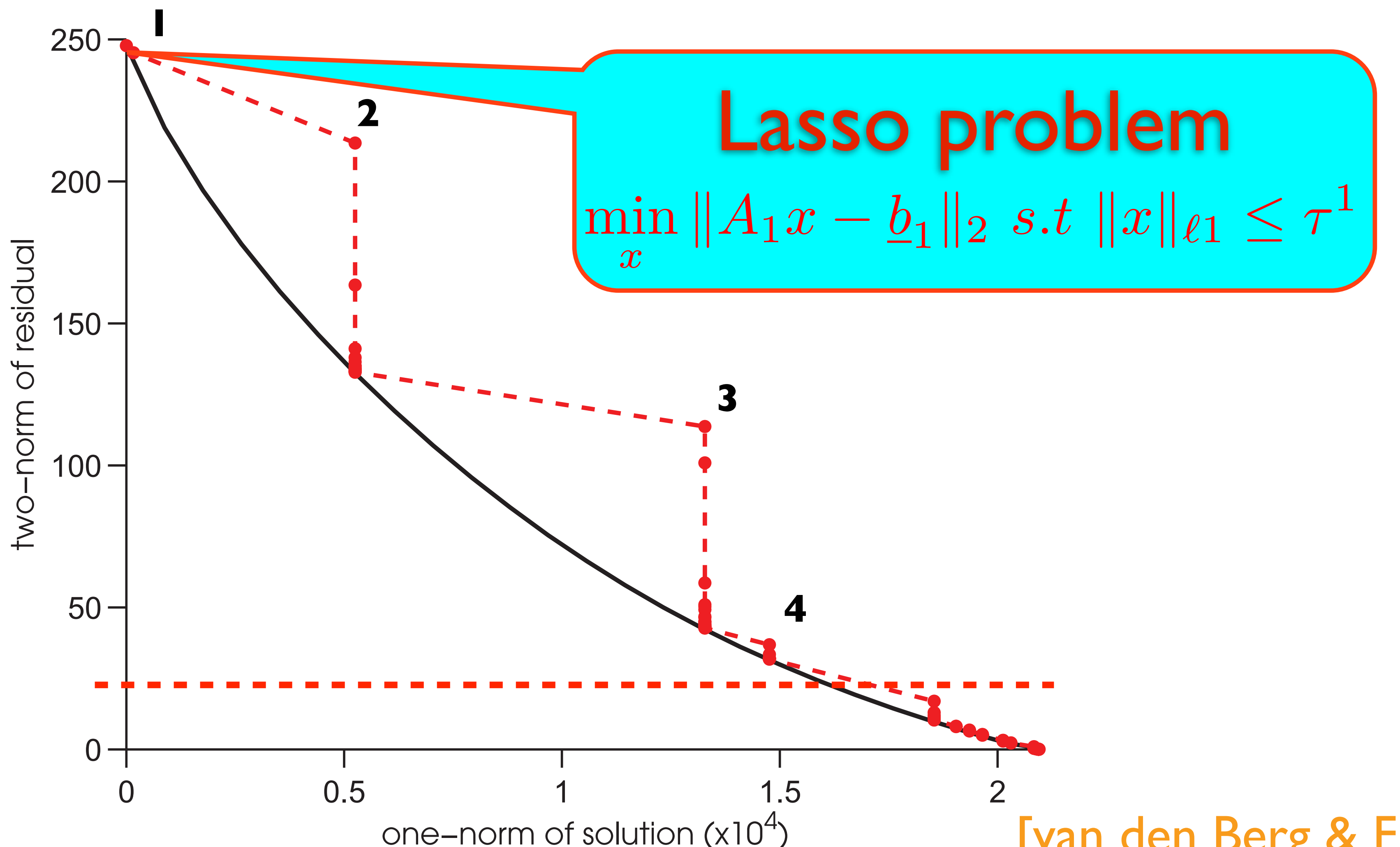
Break *correlations* between the model *iterate* and matrix **A** by *rerandomization*

- draw new *independent* $\{\mathbf{b}_t, \mathbf{A}_t\}$ after each subproblem is solved
- brings in "*extra*" information *without* growing the *system*
- ***minimal*** *extra* computational & memory cost

Progress one-norm solvers *no* longer *stalled*...

Supercooled

spectral-projected gradients



[van den Berg & Friedlander, '08]

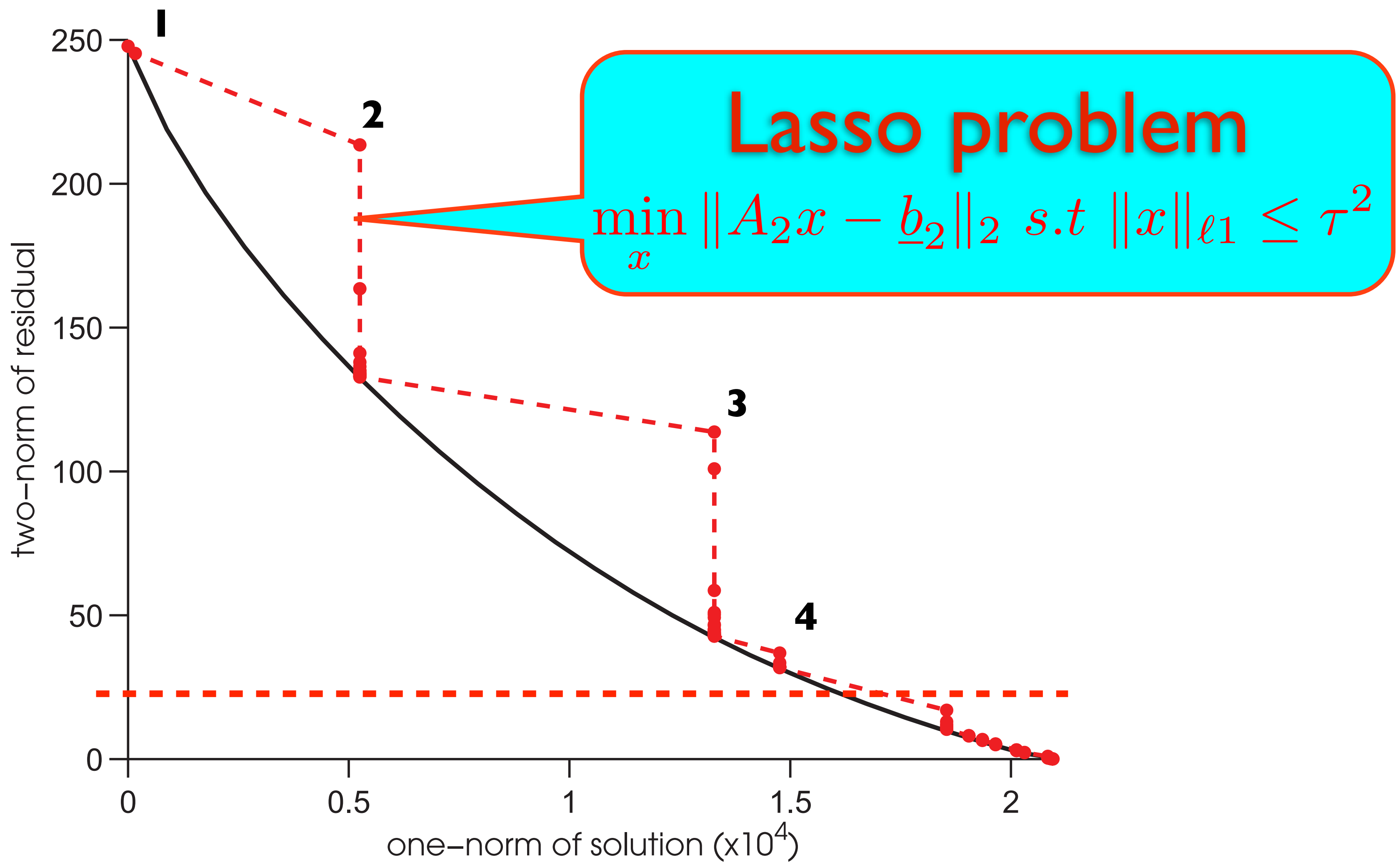
[Hennefent et. al., '08]

[Lin & FJH, '09-]

SLIM

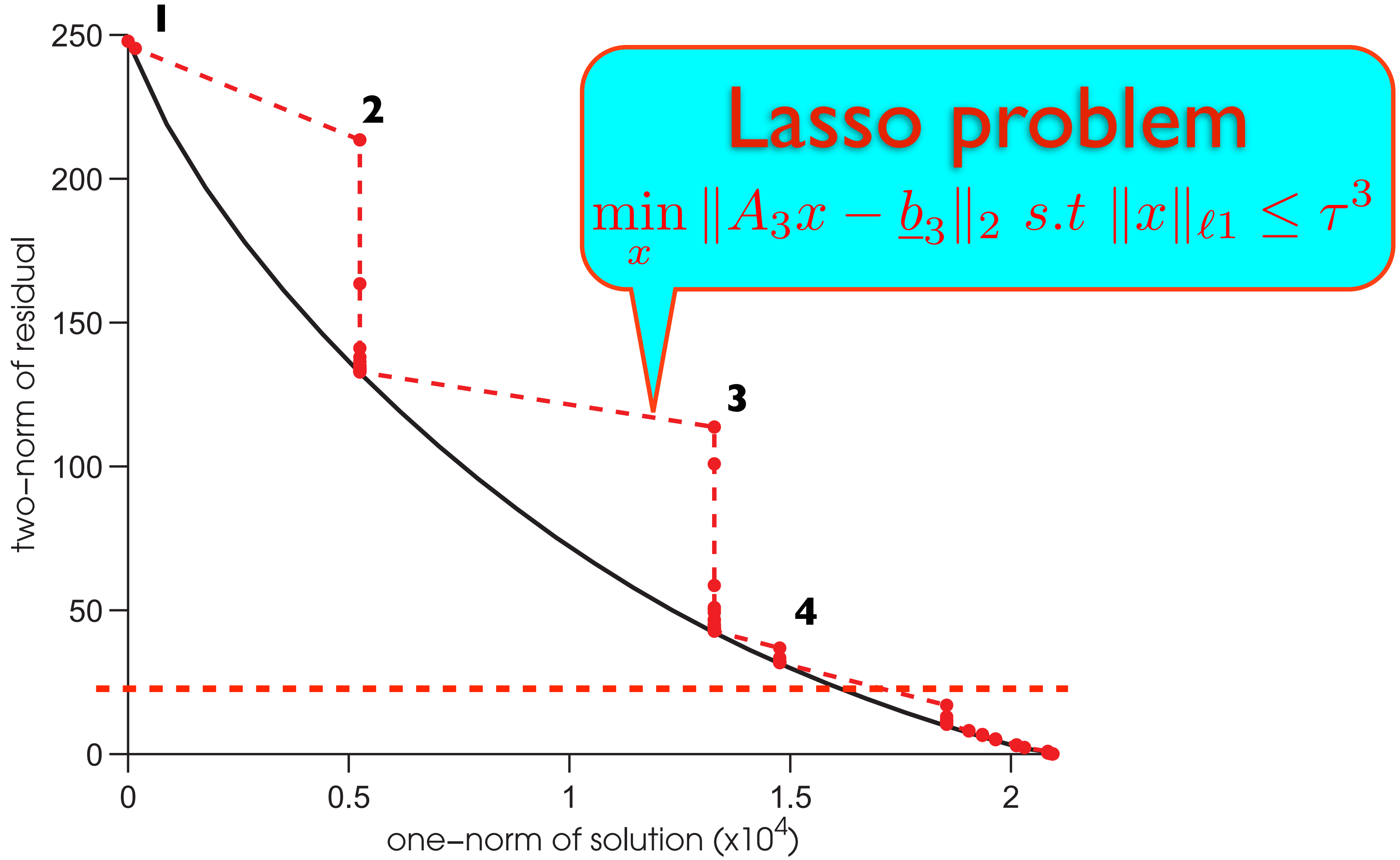
Supercooled

spectral-projected gradients



Supercooled

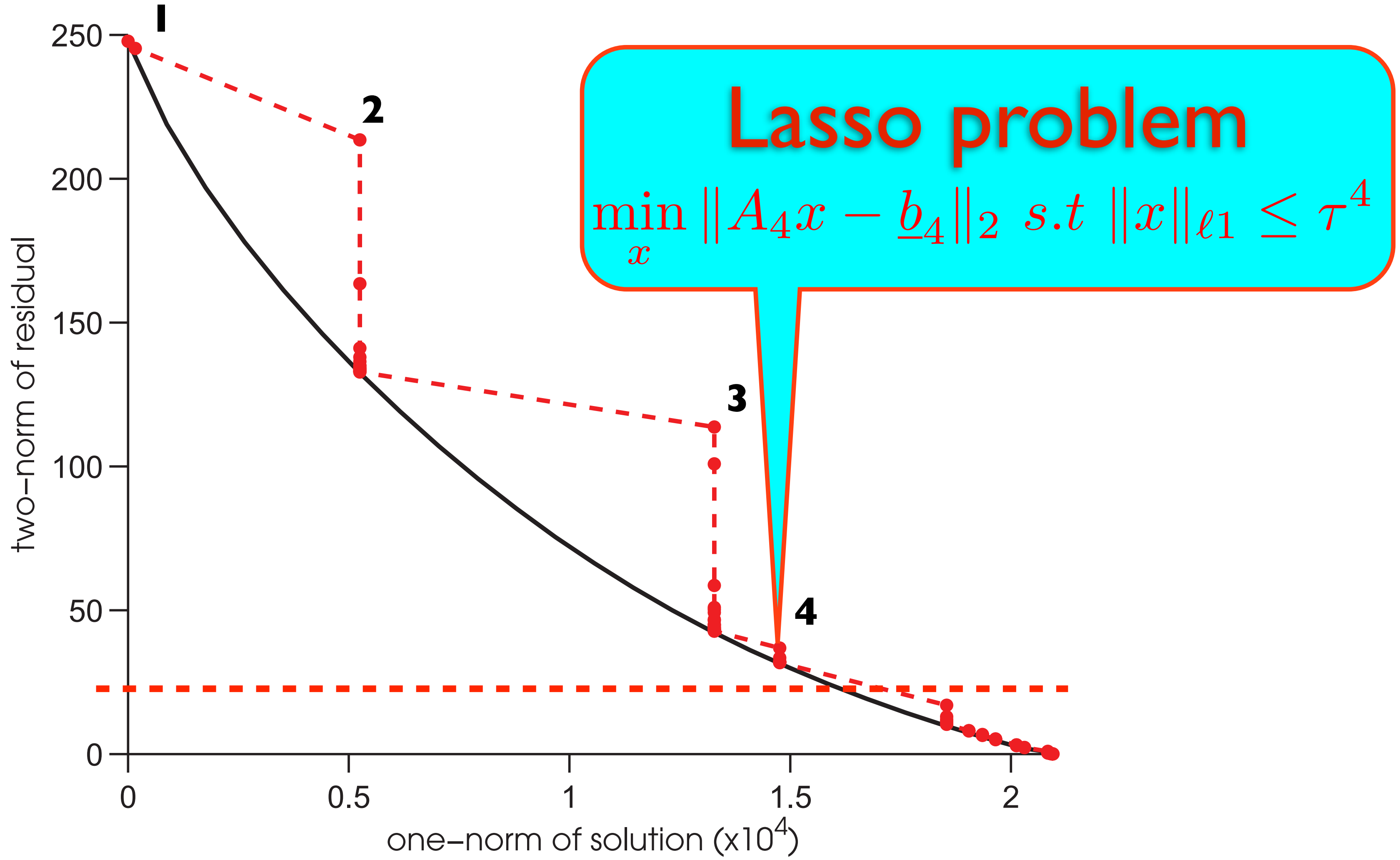
spectral-projected gradients



SLIM

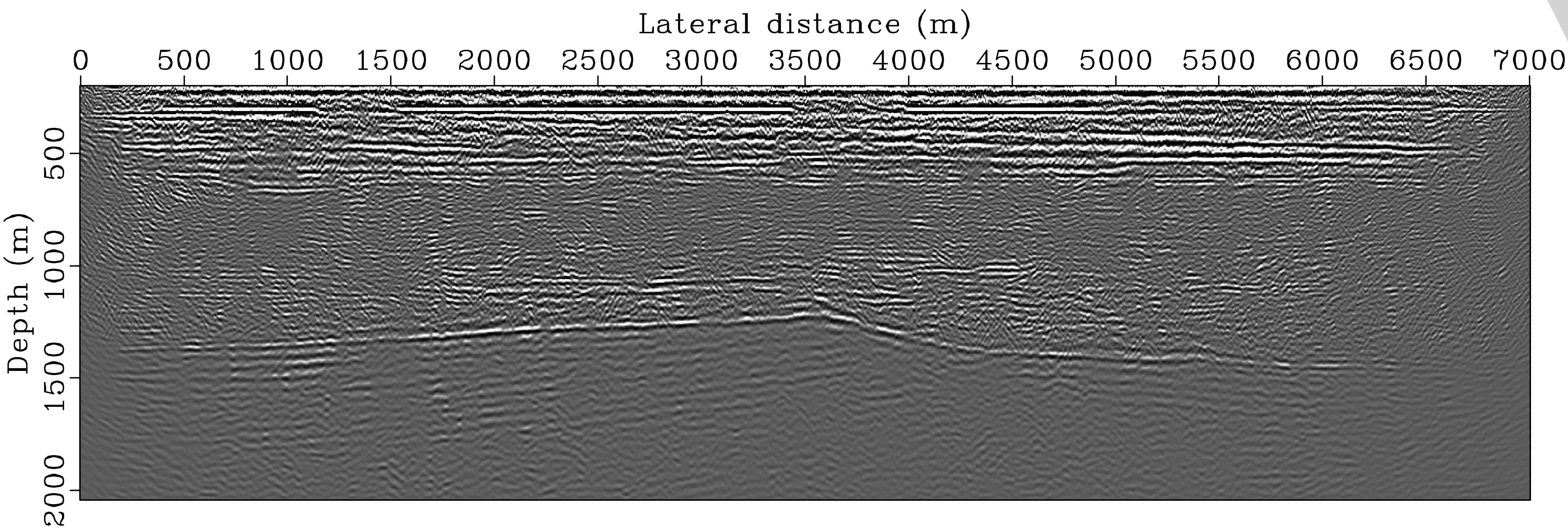
Supercooled

spectral-projected gradients



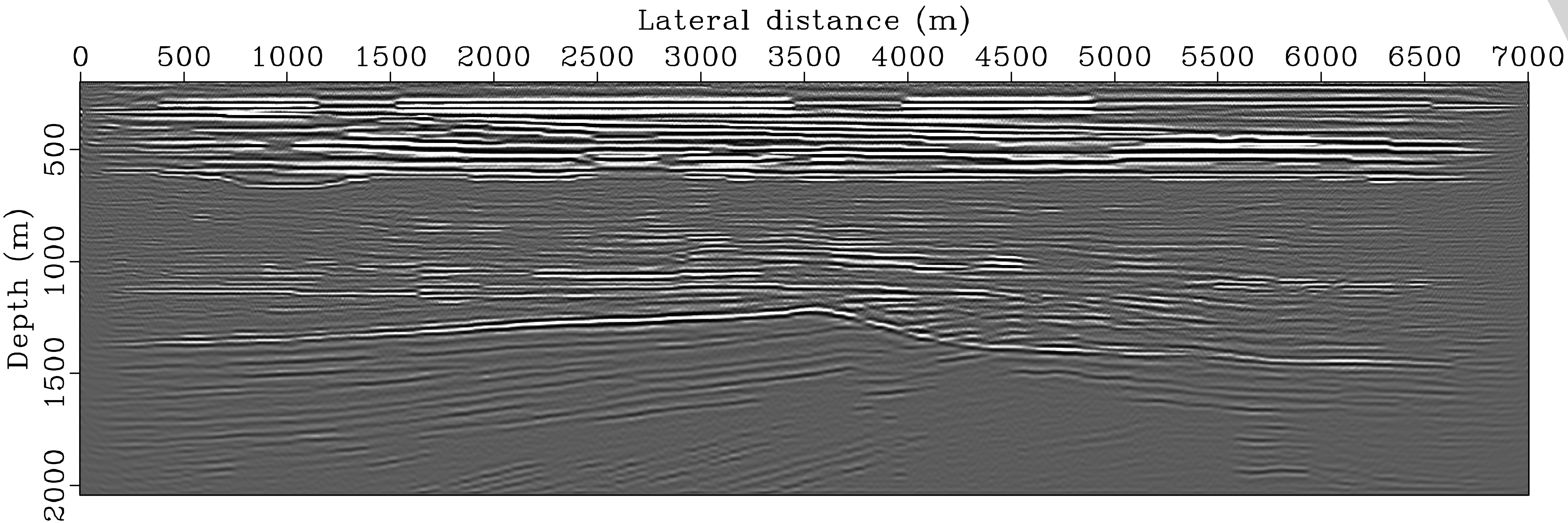
Fast Gauss-Newton step

[ℓ_1 w/o rerandomization 3 *super* shots]



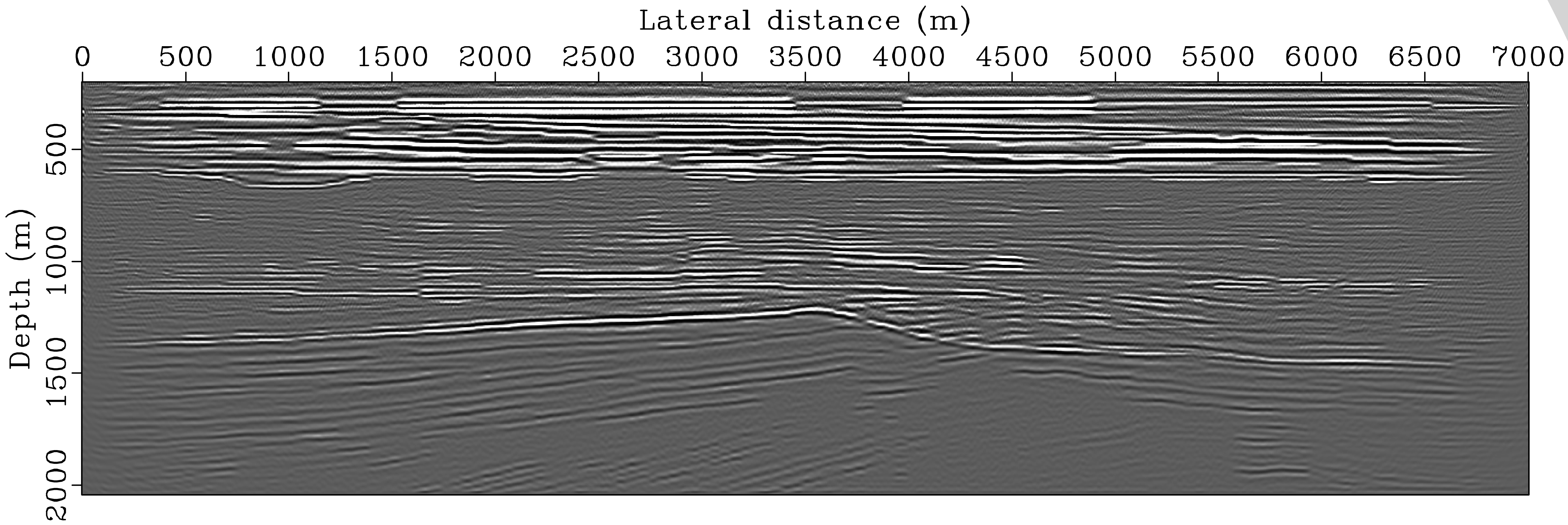
Fast Gauss-Newton step

[ℓ_1 w/ rerandomization 3 *super* shots]



Fast Gauss-Newton step

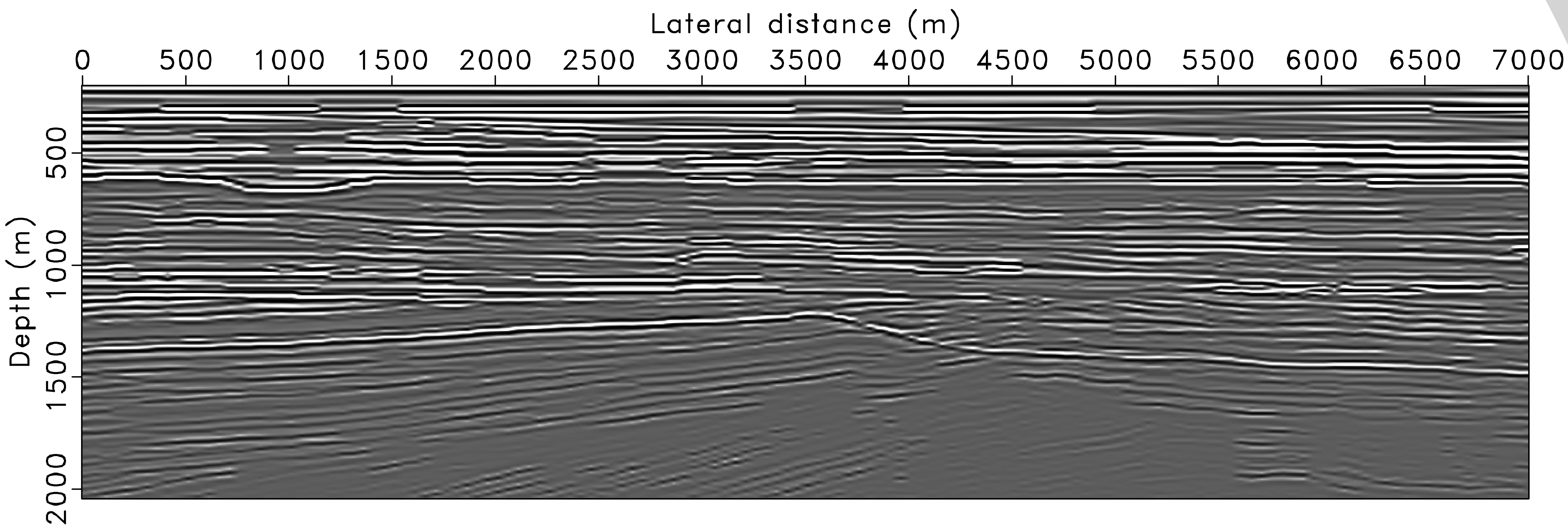
[ℓ_1 w/ rerandomization 3 *super* shots]



cost of 1/2 gradient update w/ *all* data

Fast Gauss-Newton step

[*true* update]



Approximate message passing

Add a *term* to *iterative* soft *thresholding*, i.e.,

$$\mathbf{x}^{t+1} = \eta_t\left(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t\right)$$

$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n}\mathbf{r}^{t-1}$$

← "message term"

Holds for

- *normalized* Gaussian matrices $\mathbf{A}_{ij} \in n^{-1/2}N(0,1)$
- large-scale limit and for specific thresholding *strategy*

[Montanari, ’12]

SLIM

Approximate message passing

Statistically equivalent to

$$\mathbf{x}^{t+1} = \eta_t \left(\mathbf{A}_t^* \mathbf{r}^t + \mathbf{x}^t \right)$$
$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t \mathbf{x}^t$$

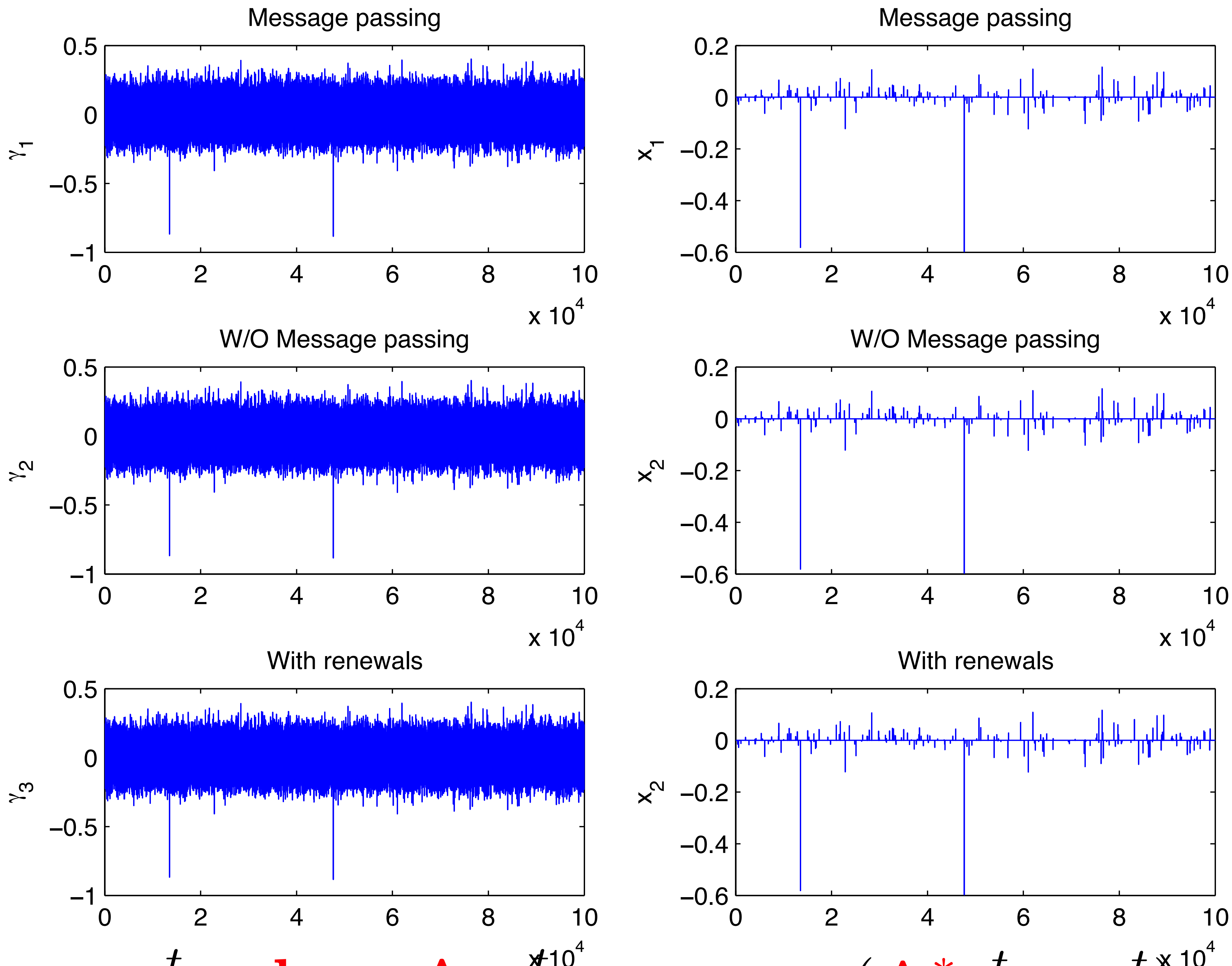
by drawing *new independent* pairs $\{\mathbf{b}_t, \mathbf{A}_t\}$ for *each* iteration

Changes the story *completely*

- breaks *correlation* buildup
- *faster* convergence

Iteration t=1

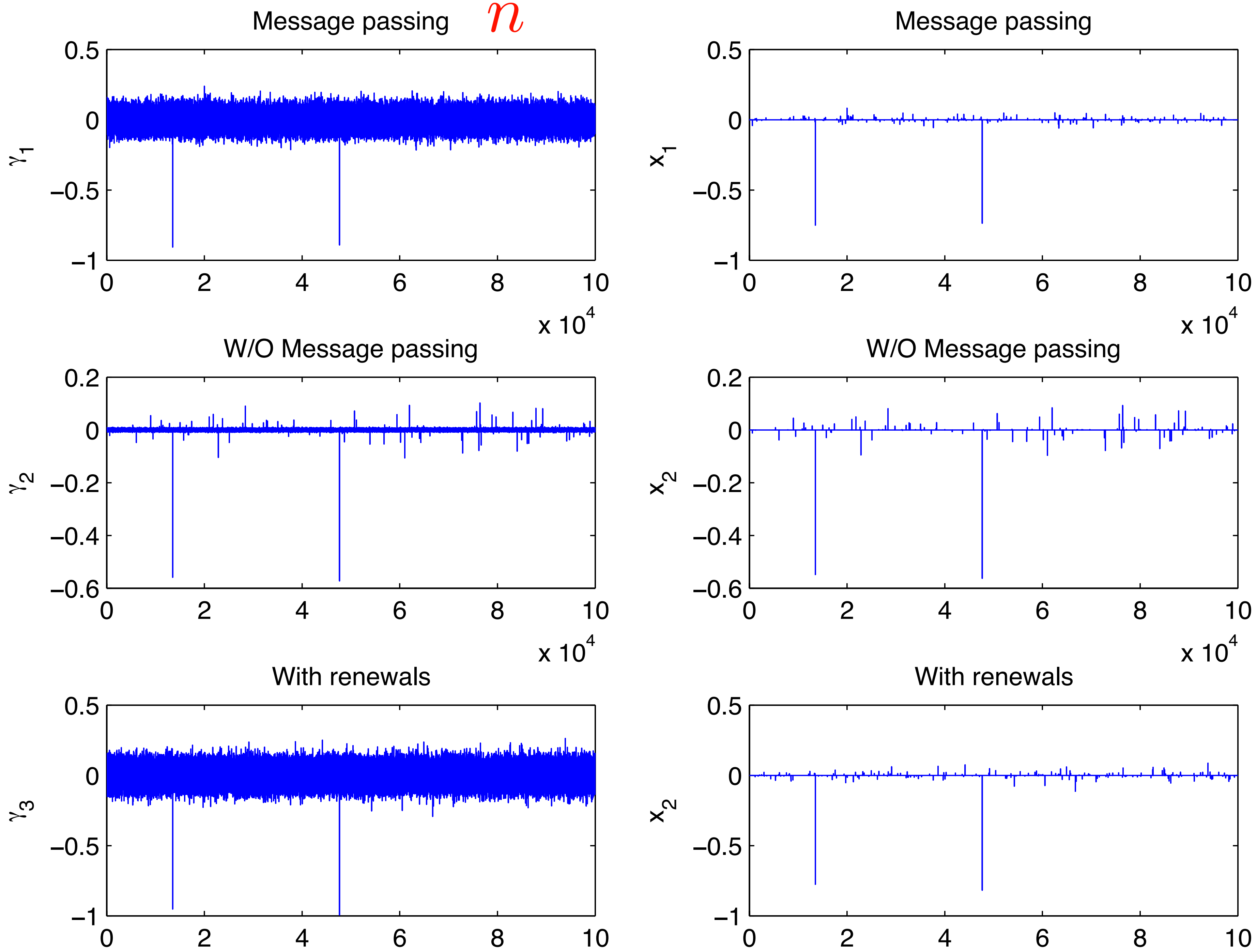
$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

Iteration t=2

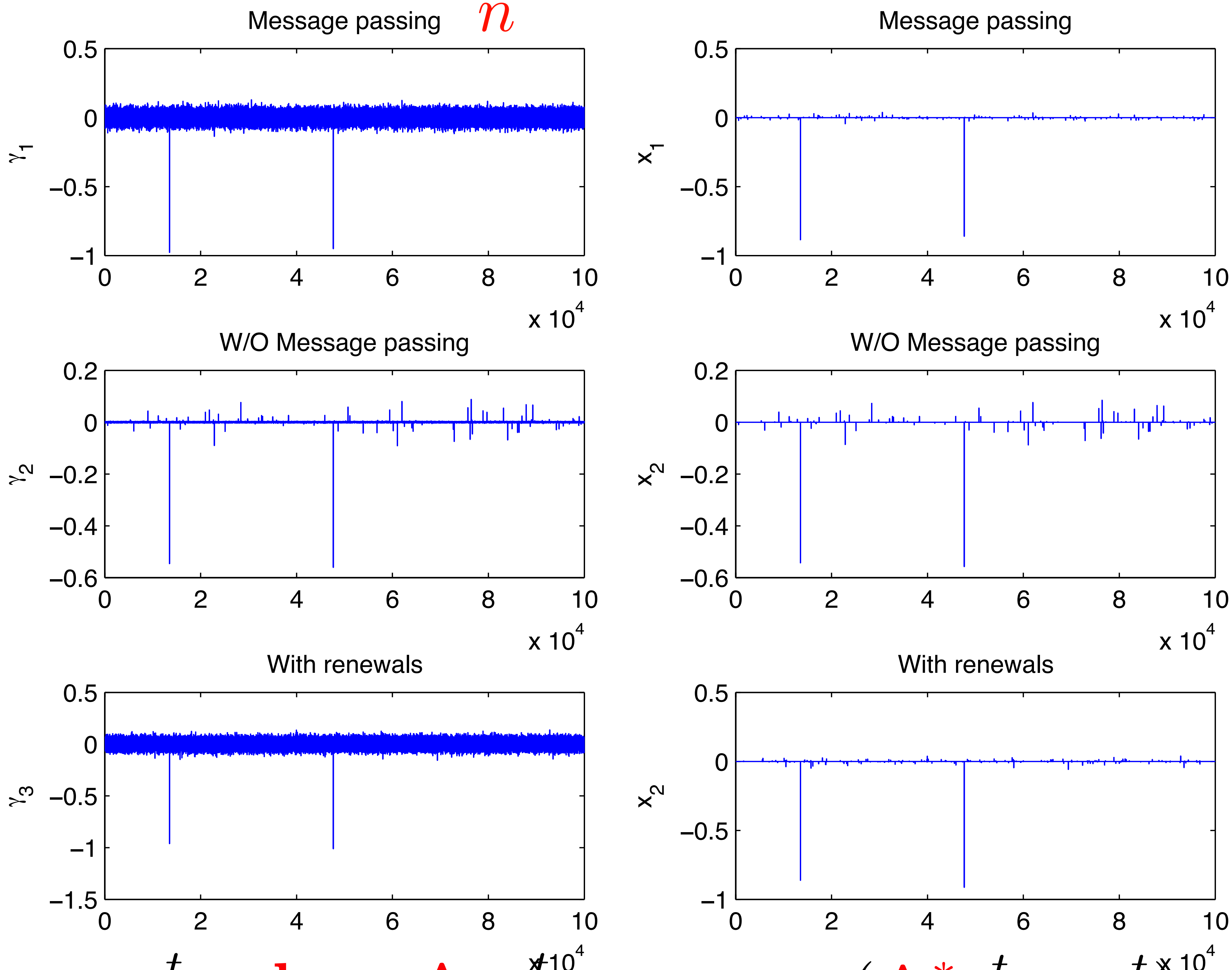
$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

Iteration t=3

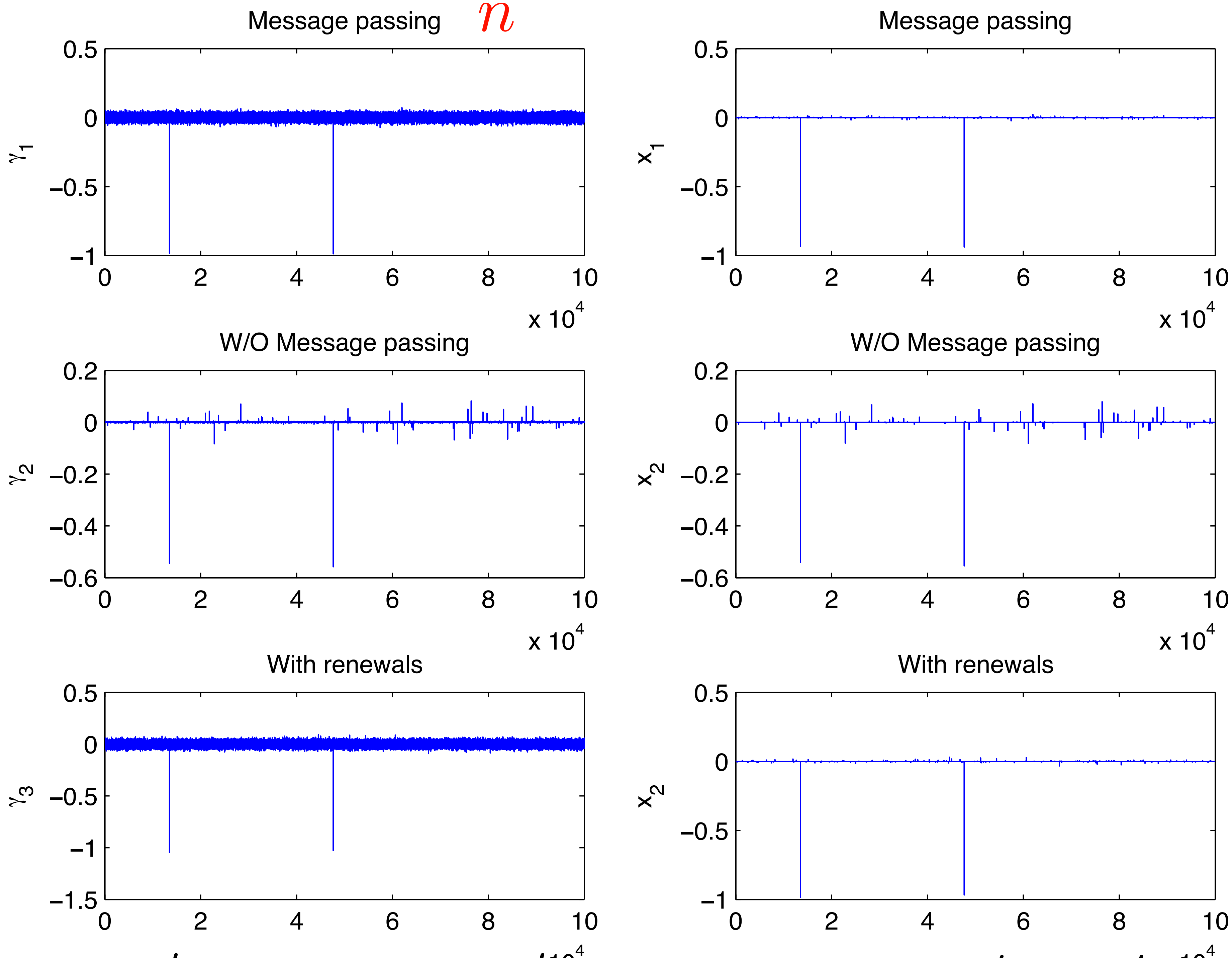
$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

Iteration t=4

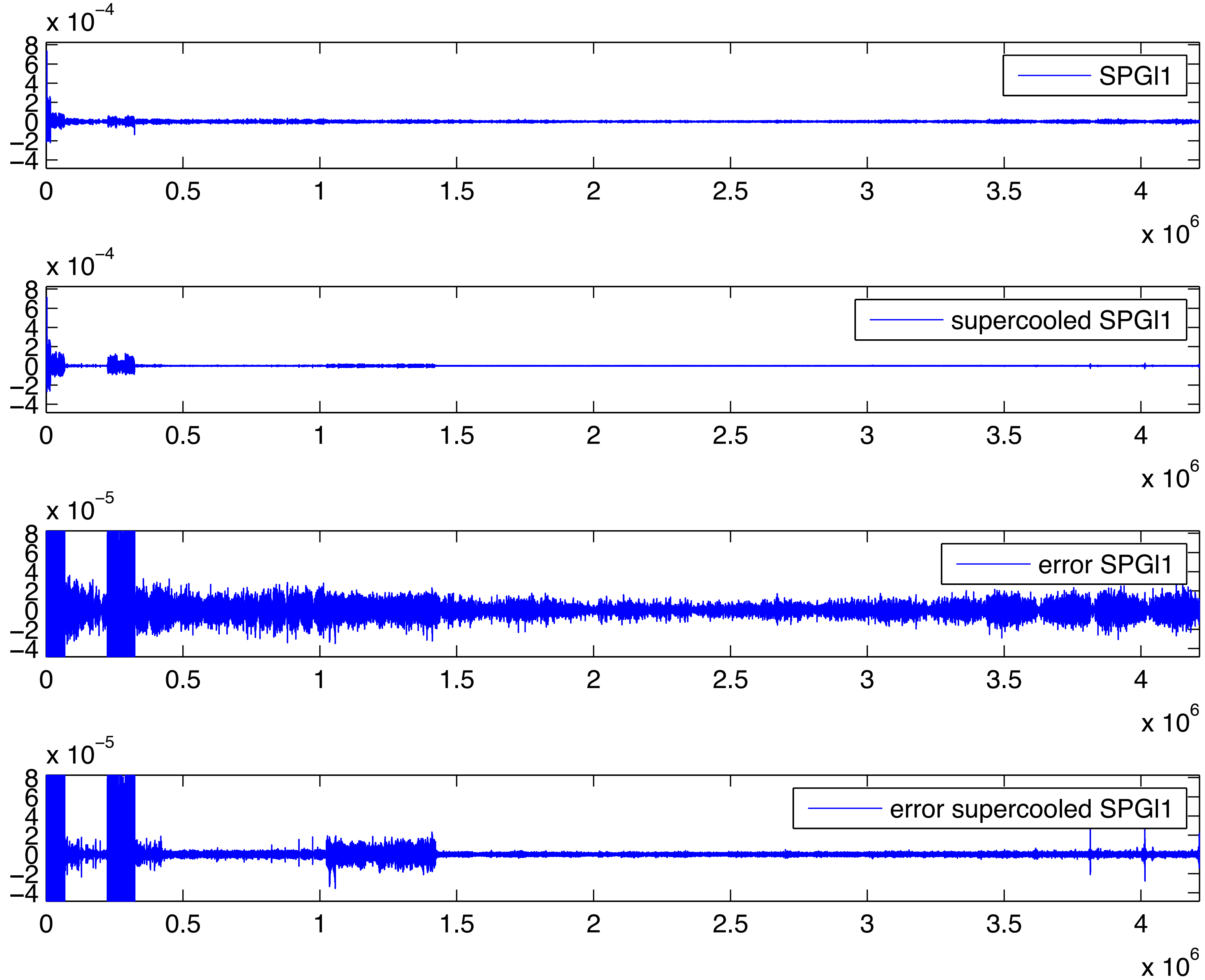
$$\mathbf{r}^t = \mathbf{b} - \mathbf{A}\mathbf{x}^t + \frac{\|\mathbf{x}^{t+1}\|_0}{n}\mathbf{r}^{t-1} \qquad \eta_t(\mathbf{A}^*\mathbf{r}^t + \mathbf{x}^t)$$



$$\mathbf{r}^t = \mathbf{b}_t - \mathbf{A}_t\mathbf{x}^t \qquad \eta_t(\mathbf{A}_t^*\mathbf{r}^t + \mathbf{x}^t)$$

Fast Gauss-Newton step

[estimated coefficients]



10 X

10 X

Blind case study

Synthetic data for *unknown* earth model was generated by a team from Chevron, ExxonMobil, and Schlumberger

Several contractor companies worked w/ large teams for weeks/months to get results w/ a lot of "hand holding"

We did not do too bad but do not really now...

Algorithm
modified Gauss-Newton

Result: Output estimate for the model $\mathbf{m}$

$k \longleftarrow 0$; $\mathbf{m}^k \longleftarrow \mathbf{m}_0$

while not converged **do**

$\{\underline{\mathbf{D}}^k, \underline{\mathbf{Q}}^k\} \longleftarrow \{\mathbf{D}\mathbf{W}^k, \mathbf{Q}\mathbf{W}^k\}$ with $\mathbf{W}^k \subset [\mathbf{e}_1, \cdots, \mathbf{e}_{n_s}]$

$\underline{\delta\mathbf{D}}^k \longleftarrow \underline{\mathbf{D}}^k - \boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]$ $\tau^k \longleftarrow \|\underline{\delta\mathbf{D}}^k\|_F \,/\, \|\mathbf{C}_2 \nabla\boldsymbol{\mathcal{F}}^*[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\underline{\delta\mathbf{D}}^k\|_\infty$

$\delta\mathbf{x} \leftarrow \arg\min_{||\mathbf{x}||_1 \leq \tau_k} \|\underline{\delta\mathbf{D}}^k - \nabla\boldsymbol{\mathcal{F}}[\mathbf{m}^k; \underline{\mathbf{Q}}^k]\mathbf{C}_2^H \mathbf{x}\|_F^2$

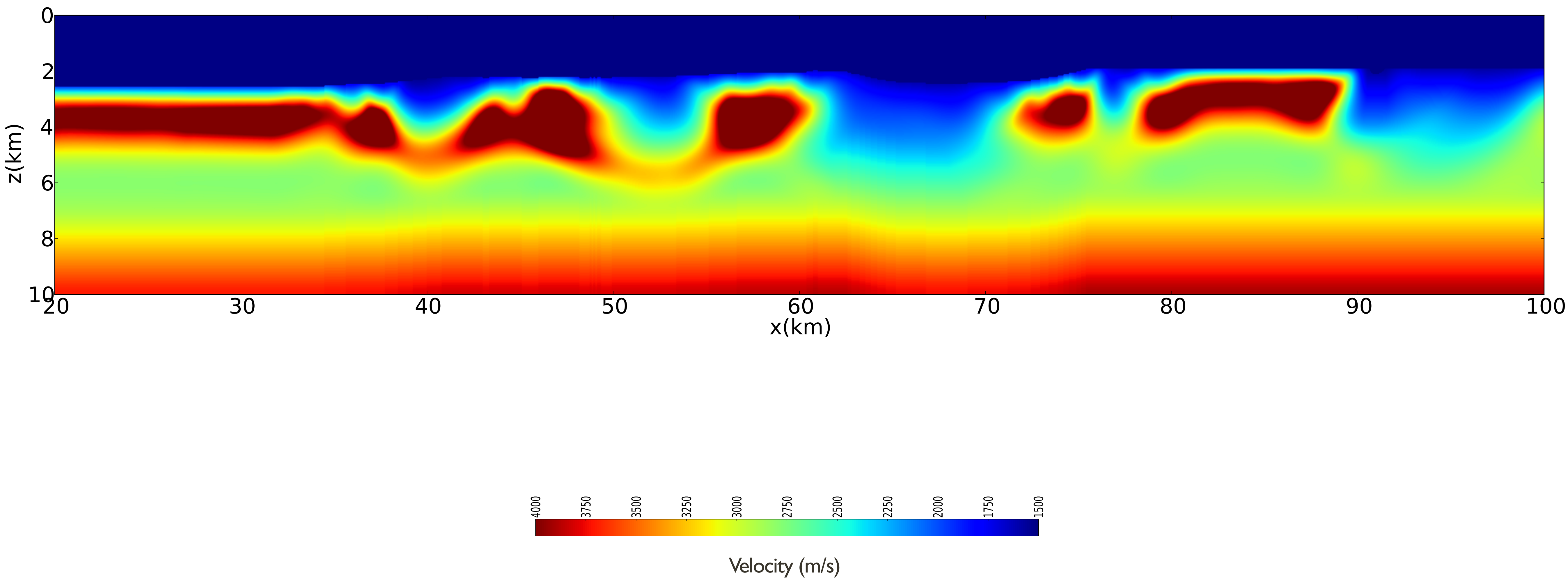
$\mathbf{m}^{k+1} \longleftarrow \mathbf{m}^k + \gamma^k \mathbf{C}_2^H \delta\mathbf{x}$

$k \longleftarrow k + 1$;

end

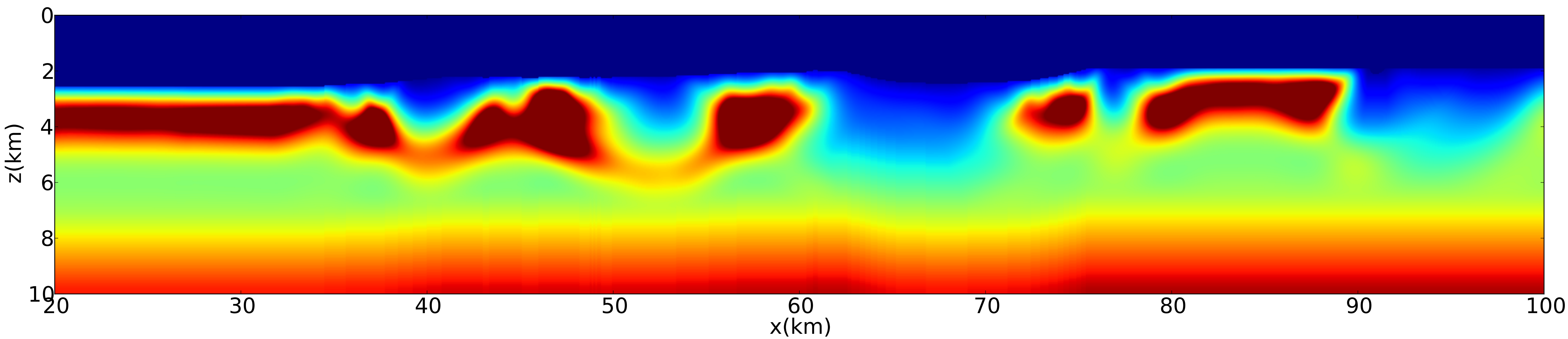
Algorithm 1: modified Gauss Newton with sparsity promotion

Input Model
[ray-based tomography + NMO]

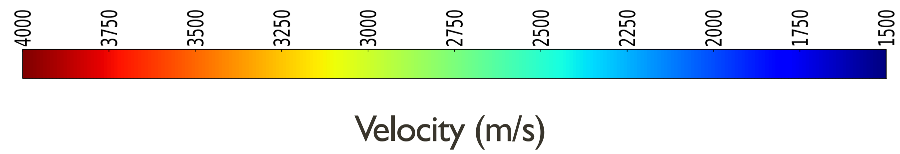


SLIM

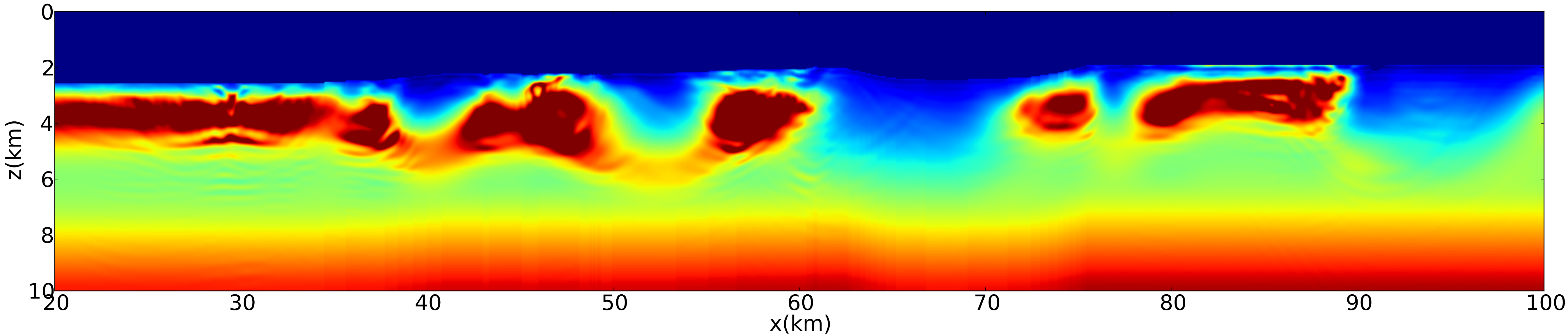
Input Model
[ray-based tomography + NMO]

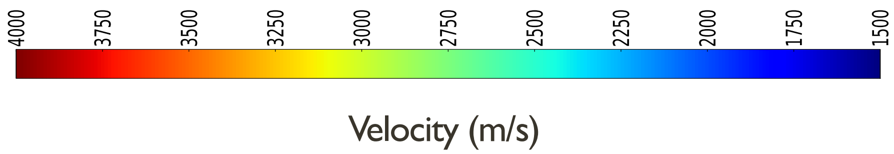


after hand picking of first breaks in 600k traces

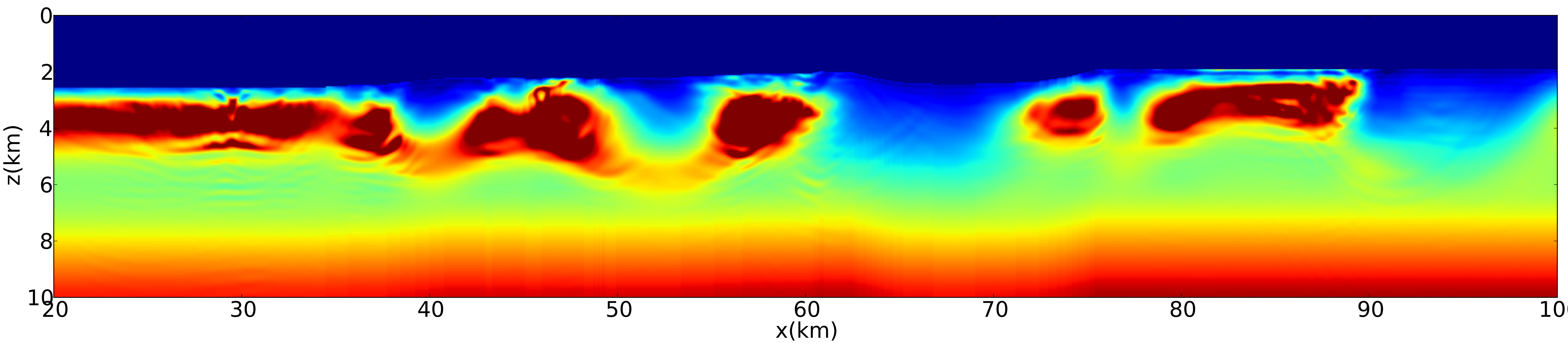


Final result
[Quasi-Newton]

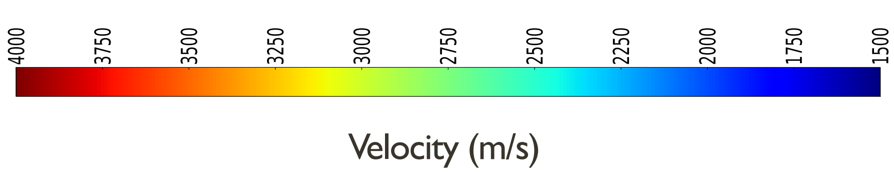




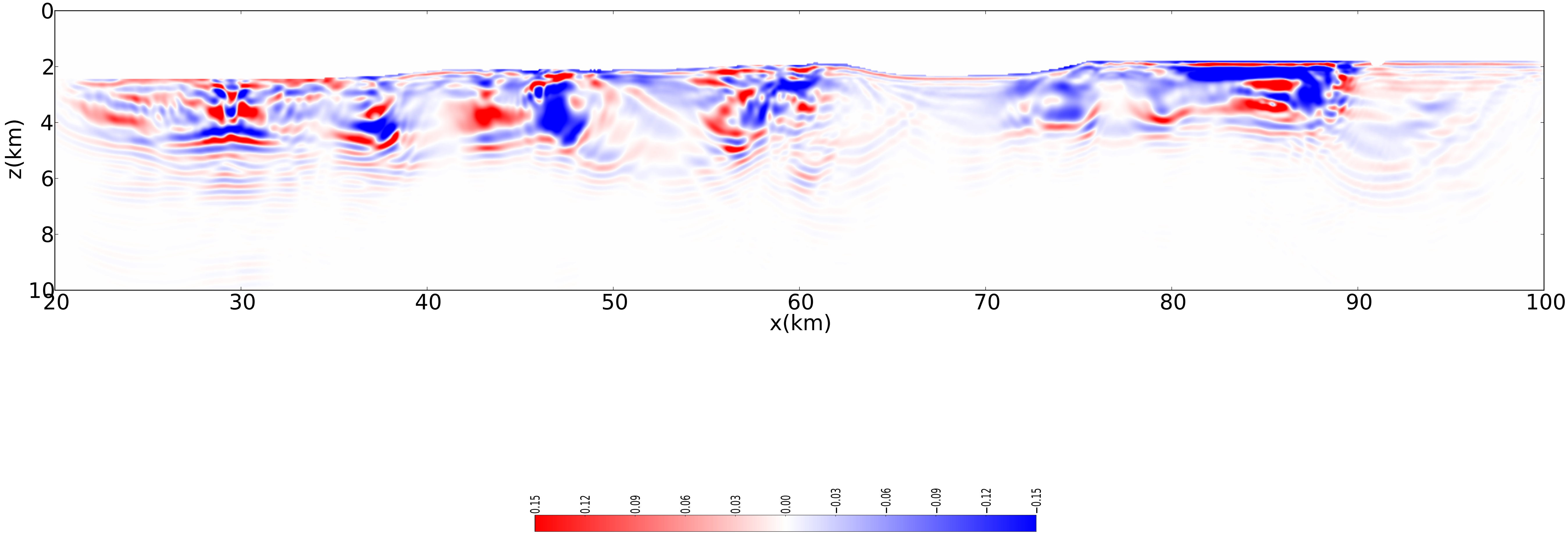
Final result
[Quasi-Newton]



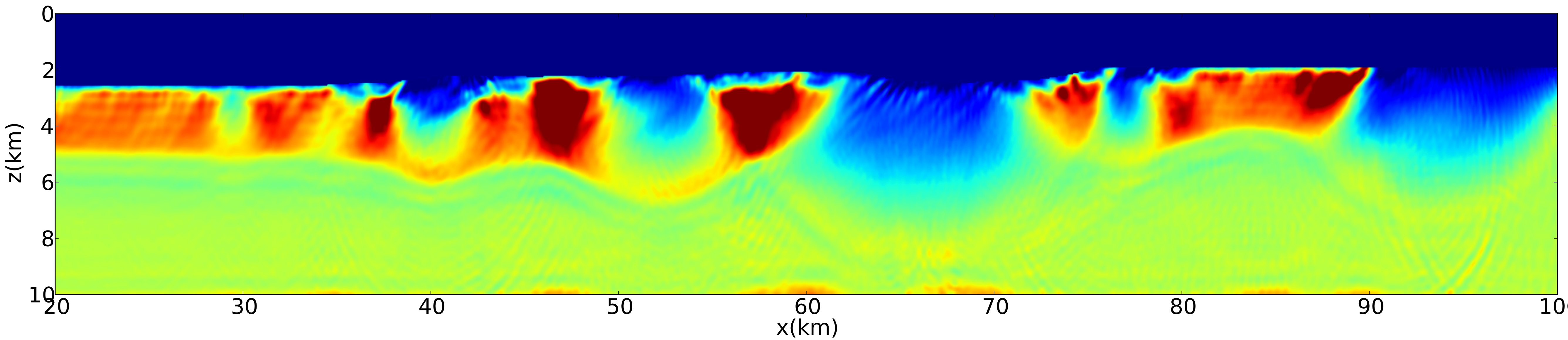
got *stuck* in a *local* minimum

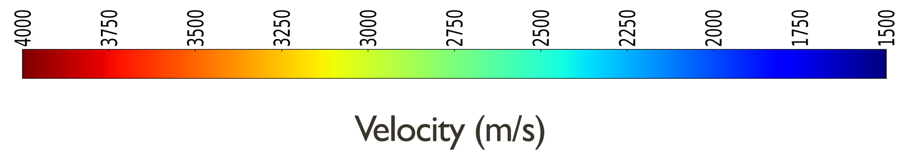


Relative update $\Delta(V)/V$

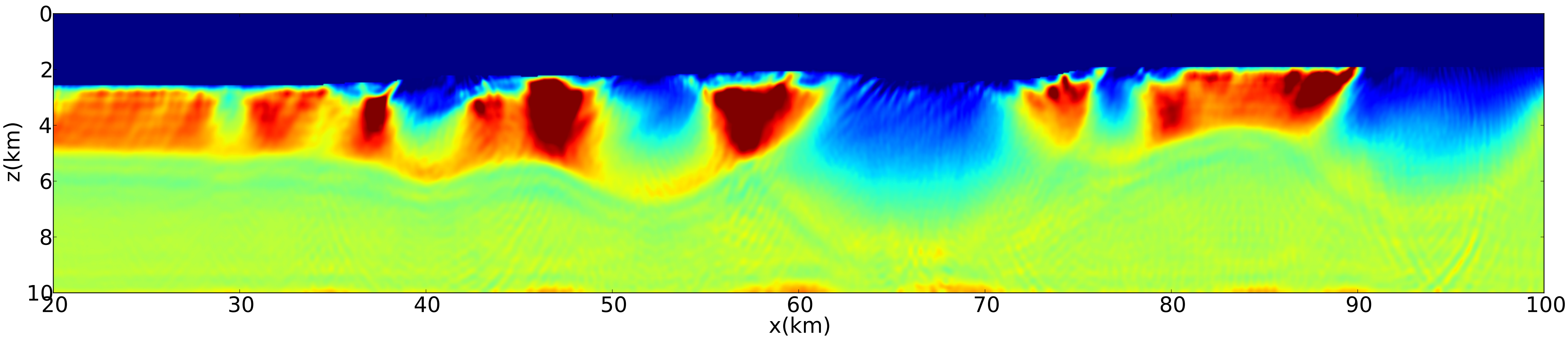


Final result
[w/ *modified* Gauss-Newton]

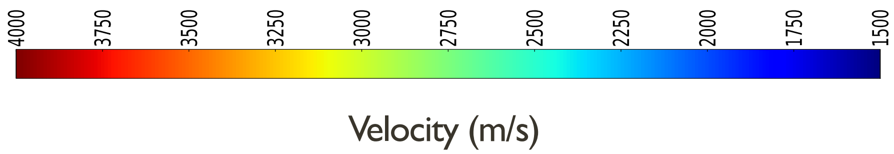




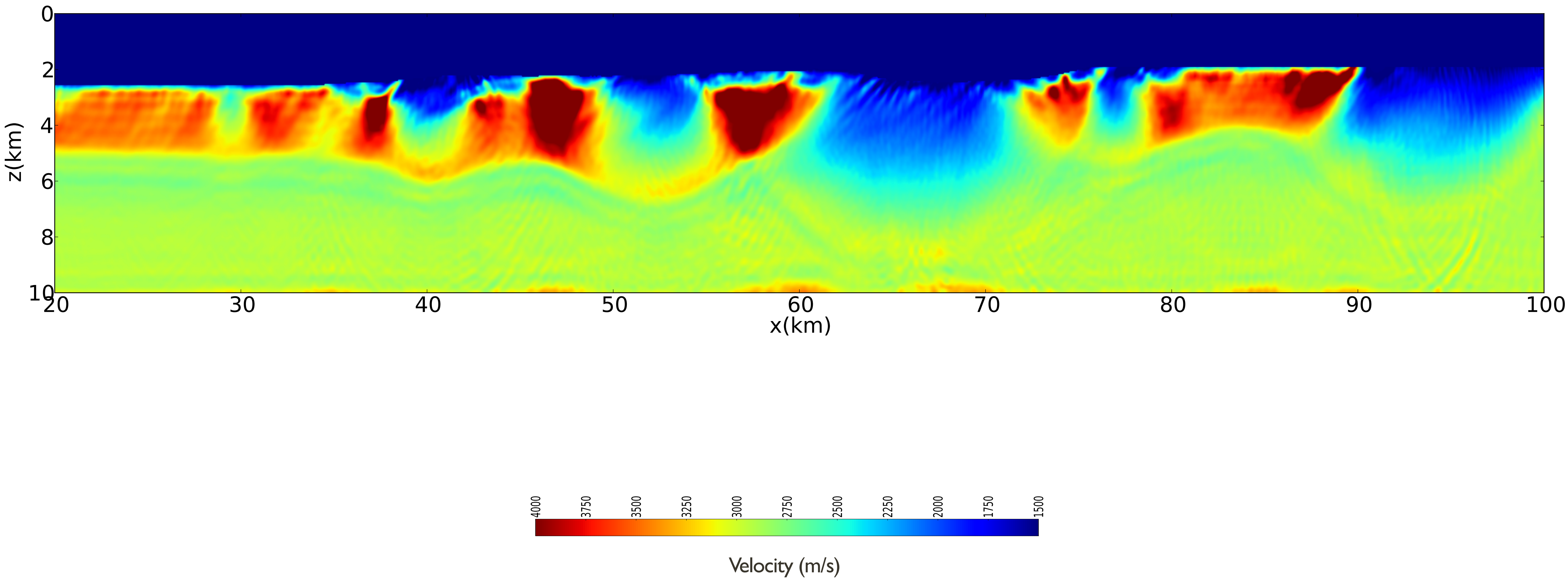
Final result
[w/ *modified* Gauss-Newton]



230400 PDE solves

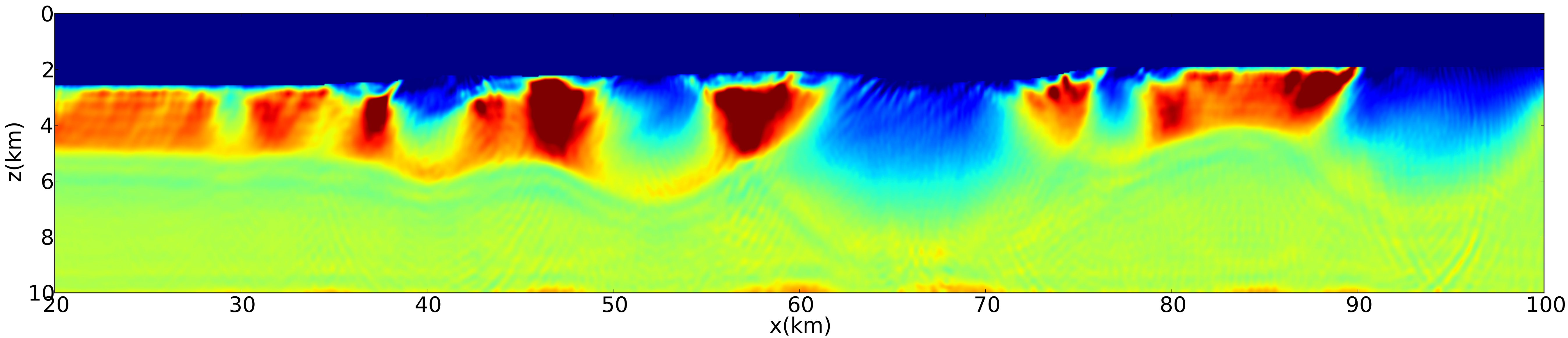


Final result
[w/ *modified* Gauss-Newton]

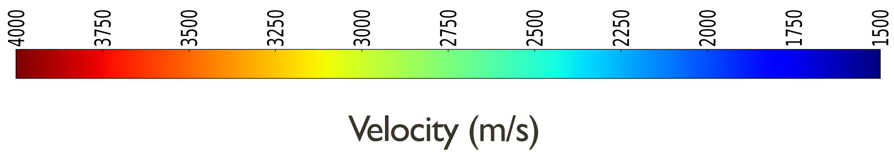


SLIM

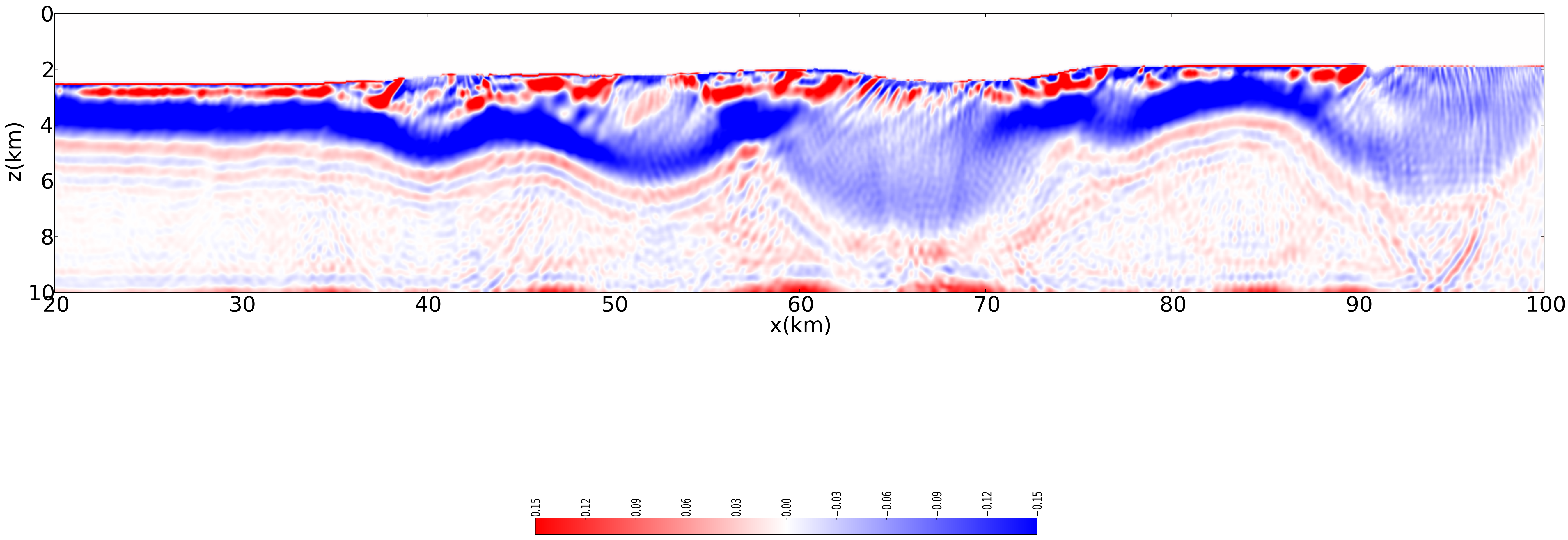
Final result
[w/ *modified* Gauss-Newton]



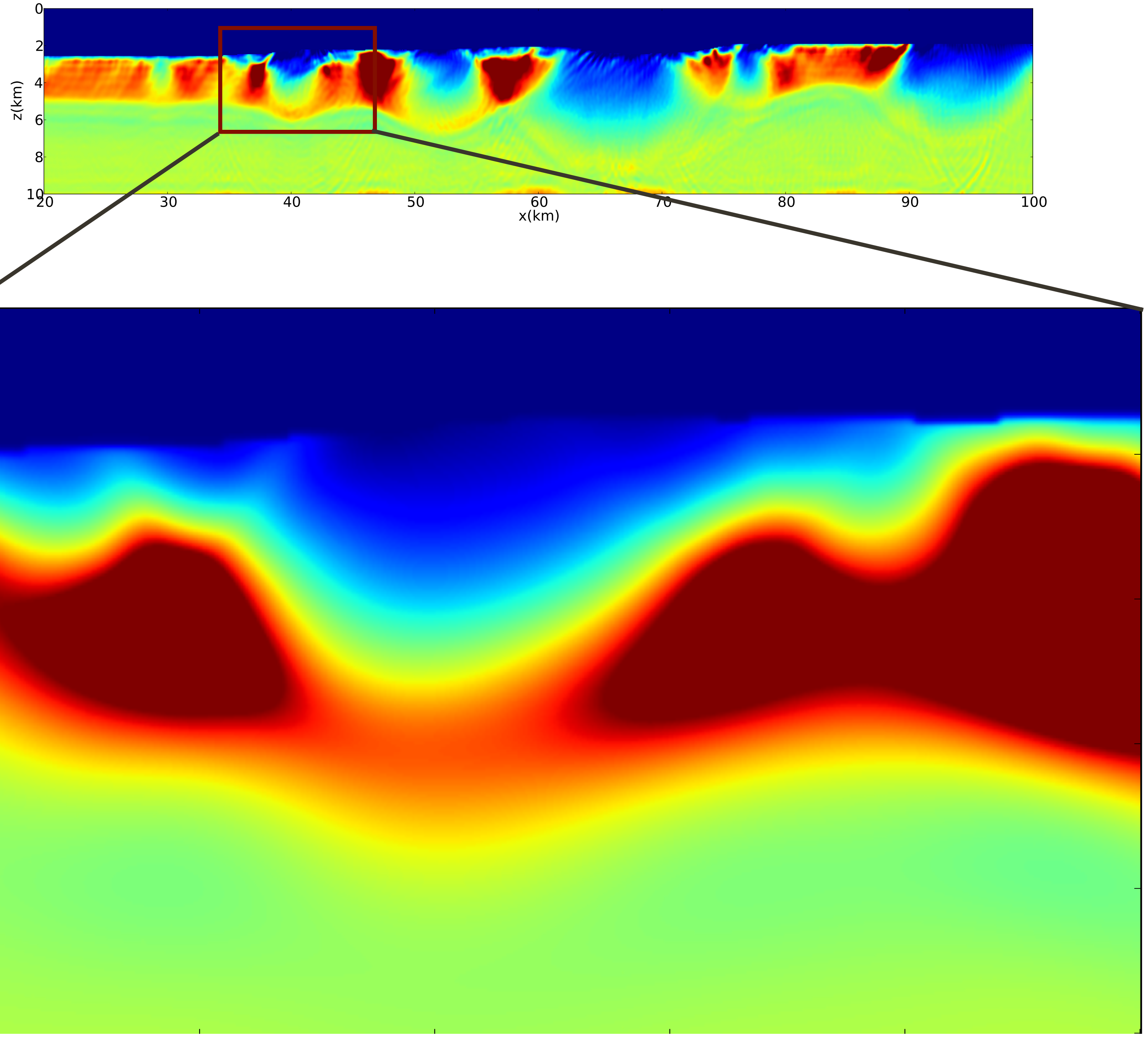
7682400 X 2815407



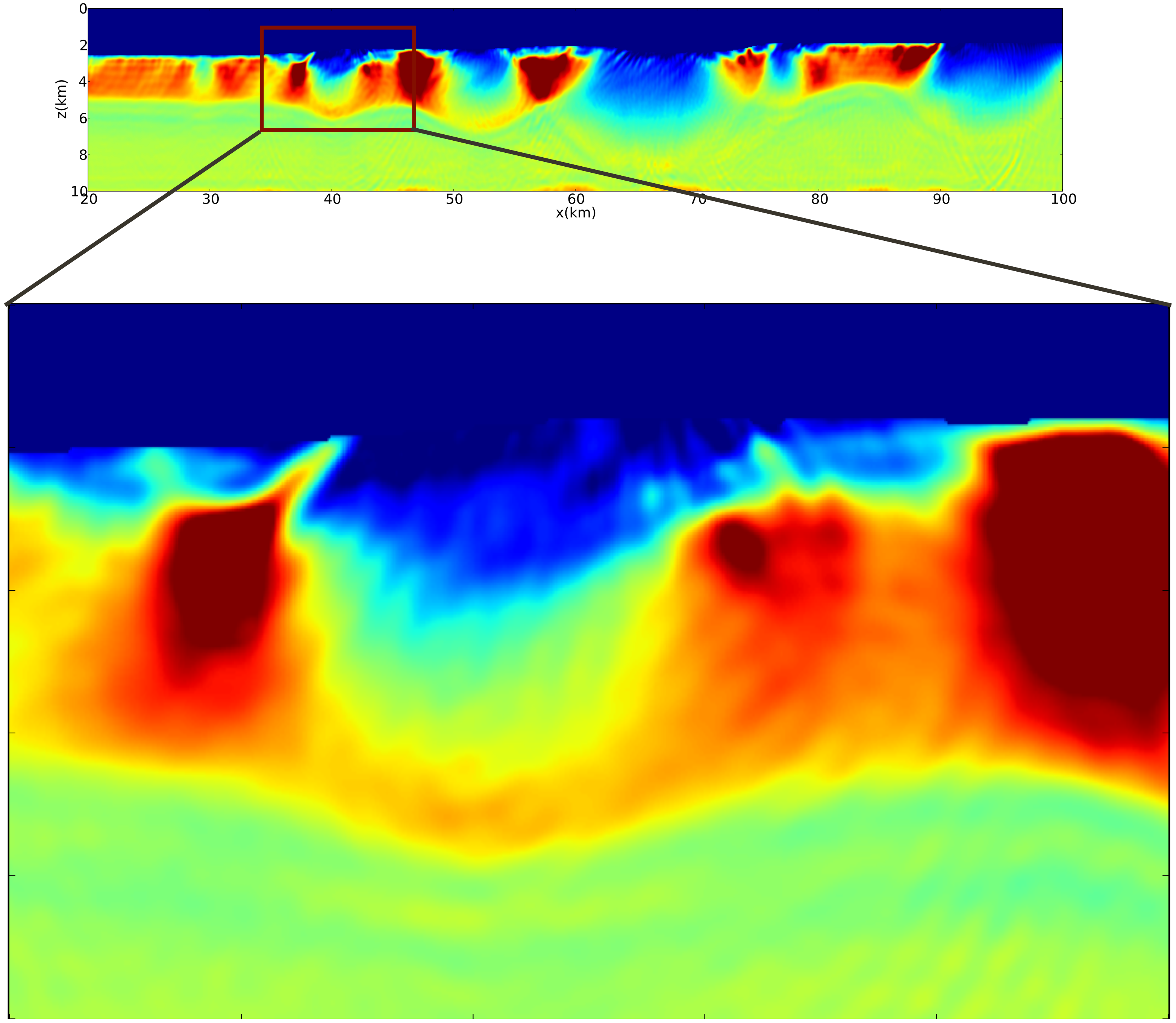
Relative update $\Delta(V)/V$



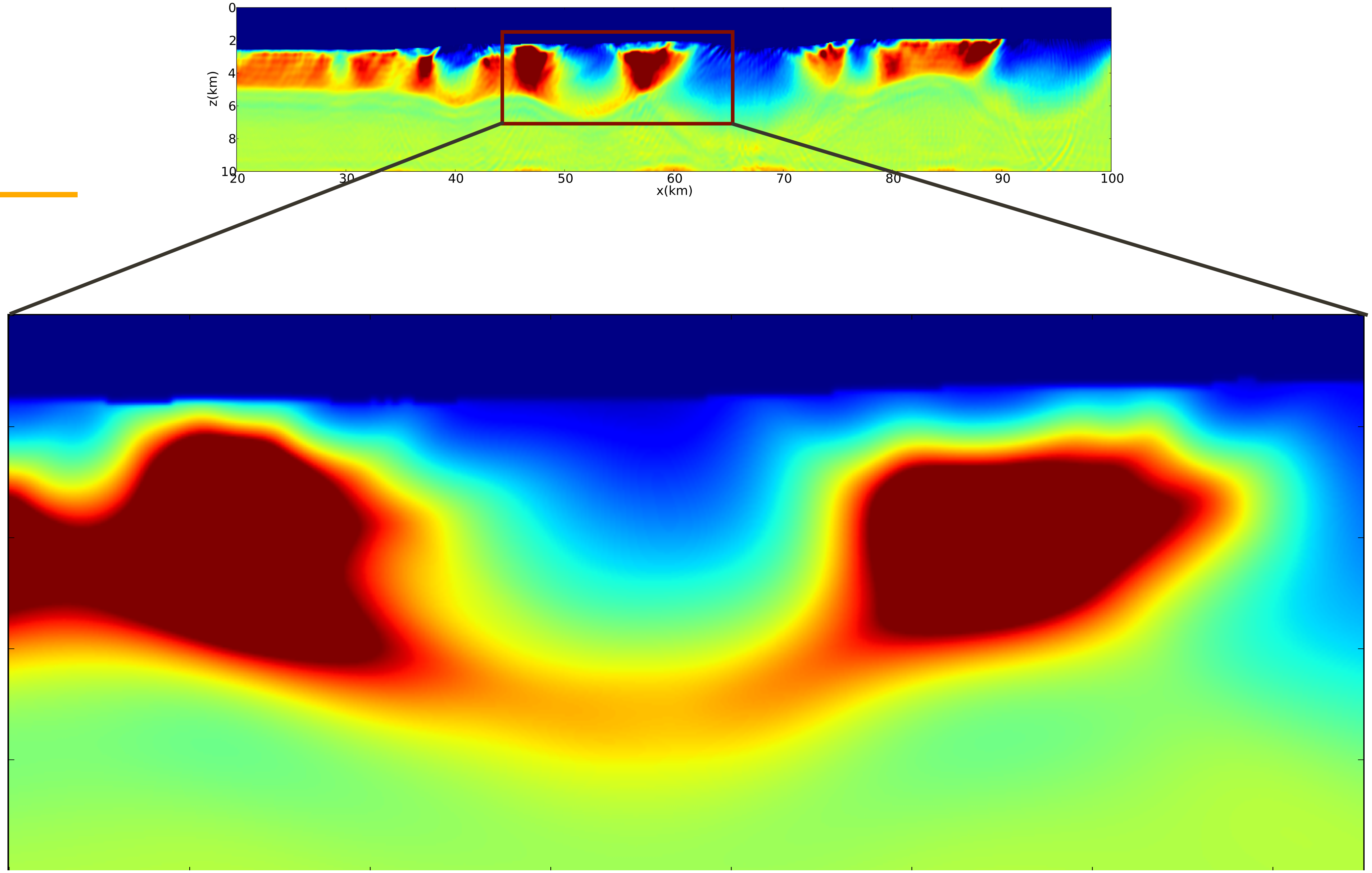
Final result
[w/ denoising]



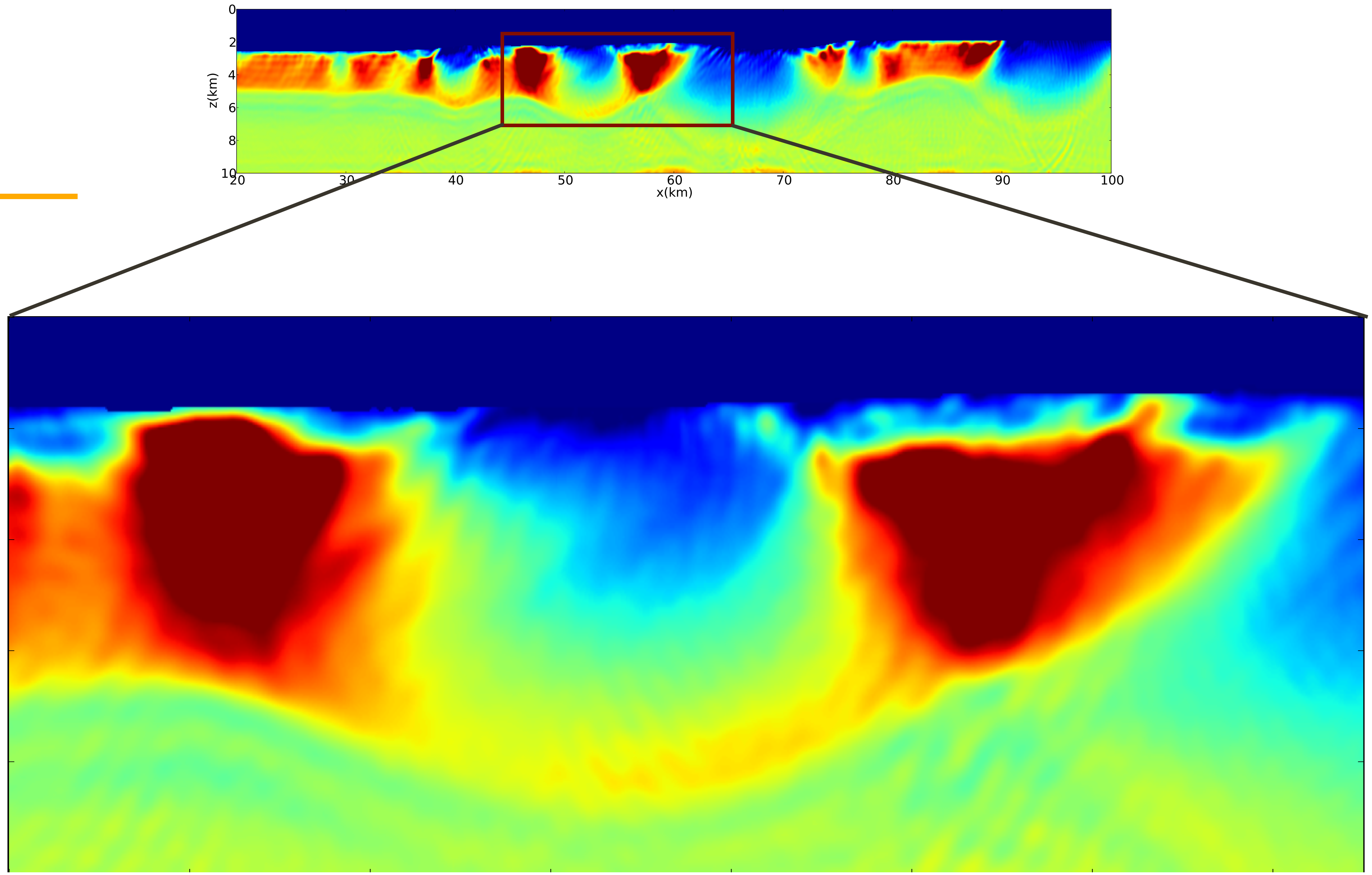
Final result
[w/ denoising]



Final result
[w/ denoising]



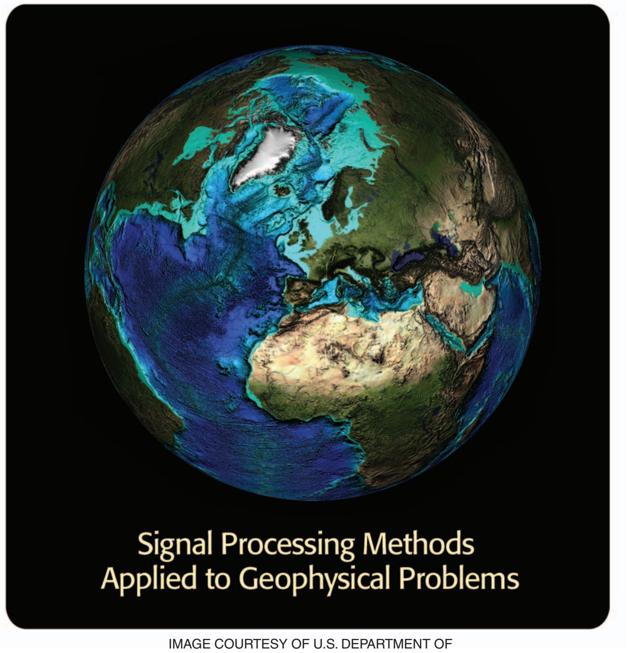
Final result
[w/ denoising]



Felix J. Herrmann, Michael P. Friedlander, and Özgür Yılmaz

Fighting the Curse of Dimensionality

Compressive sensing in exploration seismology



Signal Processing Methods Applied to Geophysical Problems

IMAGE COURTESY OF U.S. DEPARTMENT OF COMMERCE/NOAA/NESDIS/NATIONAL GEOPHYSICAL DATA CENTER

Many seismic exploration techniques rely on the collection of massive data volumes that are mined for information during processing. This approach has been extremely successful, but current efforts toward higher-resolution images in increasingly complicated regions of Earth continue to reveal fundamental shortcomings in our typical workflows. The "curse of dimensionality" is the main roadblock and is exemplified by Nyquist's sampling criterion, which disproportionately strains current acquisition and processing systems as the size and desired resolution of our survey areas continues to increase.

We offer an alternative sampling strategy that leverages recent insights from compressive sensing (CS) towards seismic acquisition and processing for data that are traditionally considered to be undersampled. The main outcome of this approach is a new technology where acquisition and processing related costs are no longer determined by overly stringent sampling criteria.

Compressive sensing is a novel nonlinear sampling paradigm, effective for acquiring signals that have a sparse representation in some transform domain. We review basic facts about this new sampling paradigm that revolutionized various areas of signal processing and illustrate how it can be successfully exploited in various problems in seismic exploration to effectively fight the curse of dimensionality.

THE CURSE OF DIMENSIONALITY IN SEISMIC EXPLORATION

Modern-day seismic-data processing, imaging, and inversion increasingly rely on computationally and data-intensive techniques to meet society's continued demand for hydrocarbons. This approach is problematic because it leads to exponentially increasing costs as the size of the area of interest increases. Motivated by recent findings from CS and earlier work in seismic data regularization [1] and phase encoding [2], we confront the challenge of the "curse of dimensionality" with a randomized dimensionality-reduction approach that decreases the cost of acquisition and subsequent processing significantly. Before we discuss possible solutions to the curse of dimensionality in exploration seismology, we first discuss how sampling is typically conducted in exploration seismology.

CLASSICAL APPROACHES

During seismic data acquisition, data volumes are collected that represent dicretizations of analog finite-energy wave fields in up to five dimensions including time. So, we are concerned with the

Digital Object Identifier 10.1109/MSP.2012.2185859
Date of publication: 9 April 2012

1053-5888/12/$31.00©2012IEEE

Further reading

Simultaneous, continuous, and random acquisition:

- *A new look at simultaneous sources* by Beasley et. al., '98.
- *Changing the mindset in seismic data acquisition* by Berkhout '08.
- Random Sampling: A New Strategy for Marine Acquisition, Moldoveanu, '10

Simultaneous simulations, imaging, and full-wave inversion:

- *Faster shot-record depth migrations using phase encoding* by Morton & Ober, '98.
- *Phase encoding of shot records in prestack migration* by Romero et. al., '00.
- *Efficient Seismic Forward Modeling using Simultaneous Random Sources and Sparsity* by N. Neelamani et. al., '08.
- Compressive simultaneous full-waveform simulation by FJH et. al., '09.
- Randomized dimensionality reduction for full-waveform inversion by FJH & X. Li, '10
- Fast full-wavefield seismic inversion using encoded sources by Krebs et. al., '09
- *An effective method for parameter estimation with PDE constraints with multiple right hand sides. by* Eldad Haber, Matthias Chung, and Felix J. Herrmann. '10
- Seismic waveform inversion by stochastic optimization. Tristan van Leeuwen, Aleksandr Aravkin and FJH, 2010.
- Efficient least-squares imaging with sparsity promotion and compressive sensing by FJH & Li, '12
- Fast randomized full-waveform inversion with compressive sensing by Xiang Li et. al., '12
- Accelerated large-scale inversion with message passing by FJH, '12

Further reading

Compressive sensing in seismic acquisition

- Non-parametric seismic data recovery with curvelet frames FJH & Hennenfent ’08
- Simply denoise: wavefield reconstruction via jittered undersampling” Hennenfent & FJH ’08
- Non-uniform optimal sampling for seismic survey design Mosher et. al. ’12
- Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis by Oropeza, V., and M. Sacchi, ‘11,
- Parallel Stochastic Gradient Algorithms for Large-Scale Matrix Completion by Recht, B., and C. Ré, ’11
- Randomized marine acquisition with compressive sampling matrices, Mansour et. al., ’12
- Fast Methods for Rank Minimization with Applications to Seismic-Data Interpolation, R. Kumar et. al., ’12
- Only dither: efficient simultaneous marine acquisition by Wason et. al., ’12

Compressive sensing, sparse solvers, and weighting

- Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information by Candes, 06.
- Compressed Sensing by D. Donoho, ’06
- Probing the Pareto frontier for basis pursuit solutions by E. van den Berg and M. Friedlander, ’08
- Recovering compressively sampled signals using partial support information by Friedlander et. al., ’12
- Beyond $\ell 1$ norm minimization for sparse signal recovery by Mansour, ’12

Further reading

Message passing

- Message passing algorithms for compressed sensing by David Donoho et. al., 2009
- Graphical Models Concepts in Compressed Sensing by Andrea Montanari, '2012

Acknowledgments

We would like to thank Charles Jones from BG and Dimitri Bevc for providing for providing synthetic models. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Collaborative Research and Development Grant DNOISE II (375142-08).

This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, BGP, Chevron, ConocoPhillips, Petrobras, PGS, Total SA, and WesternGeco.

NSERC
CRSNG

Thank you

www.slim.eos.ubc.ca