Sub-Nyquist sampling and sparsity: getting more information from fewer samples

Felix J. Herrmann

SLIM

Seismic Laboratory for Imaging and Modeling
the University of British Columbia

Drivers

Our incessant

- demand for *carbonhydrates* while we are *no* longer finding oil...
- desire to understand the Earth's inner workings

Push for improved *seismic inversion* to

- create *more high-resolution* information
- from *noisier* and *incomplete* data

SLIM

Controversial statements

Size of our *discretizations* is *dictated* by

- a *far* too *pessimistic Nyquist-sampling criterion* compounded by the *curse of dimensionality*
- our *insistence* to sample *periodically*

Our desire to work with *all* data

- leads to "over emphasis" on data *collection*
- prohibits *inversion* that requires *multiple* passes through *data*

Wish list

Acquisition & inversion costs determined by *structure of data & complexity* of the *subsurface*

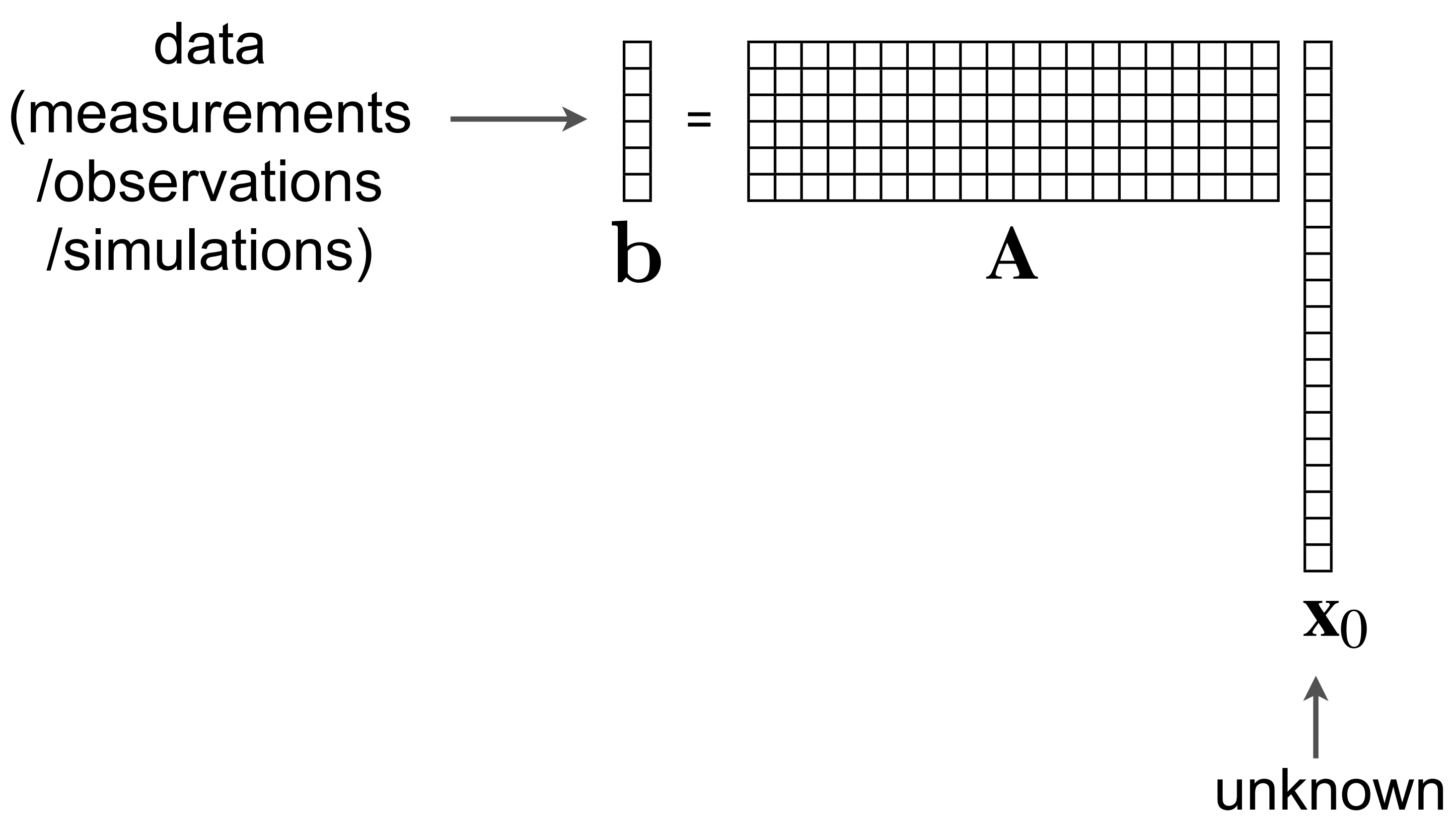
- sampling *criteria* that are *dominated by* transform-domain *sparsity* and *not* by the *size* of the *discretization*

Controllable error that depends on

- degree of *subsampling / dimensionality* reduction
- available computational resources

Problem statement

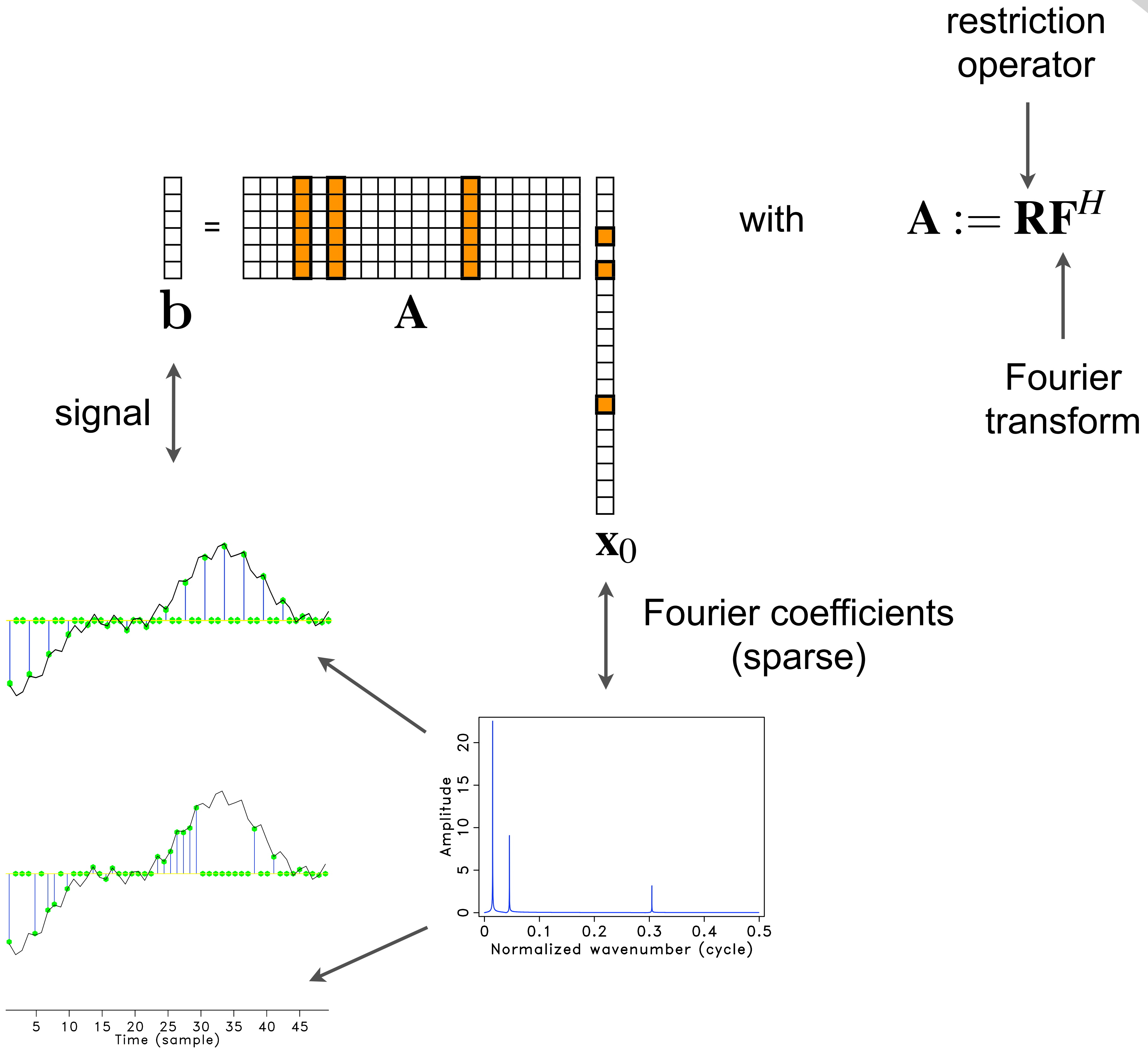
Consider the following (severely) *underdetermined* system of *linear* equations:



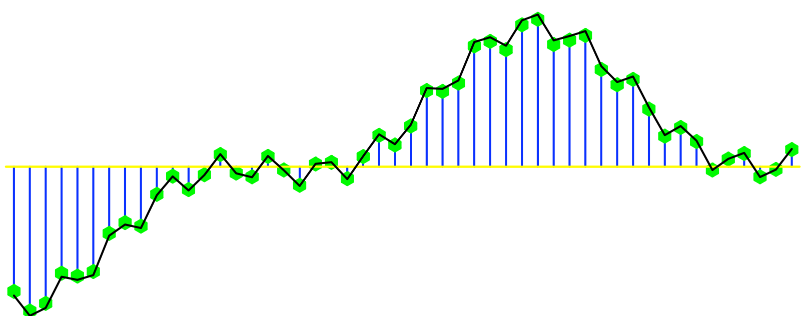
Is it possible to recover $\mathbf{x}_0$ accurately from $\mathbf{b}$?

The new field of *Compressive Sensing* attempts to answer this.

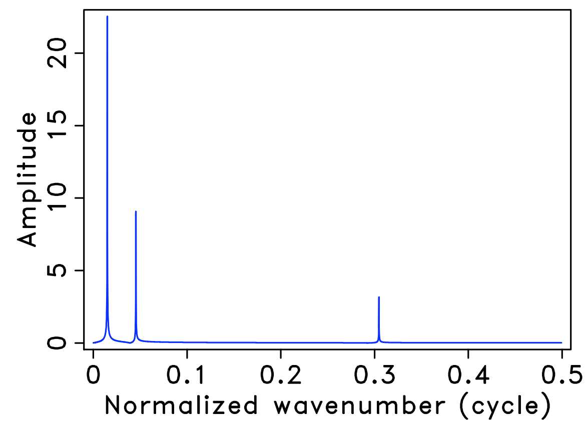
Sparse recovery



Coarse sampling schemes

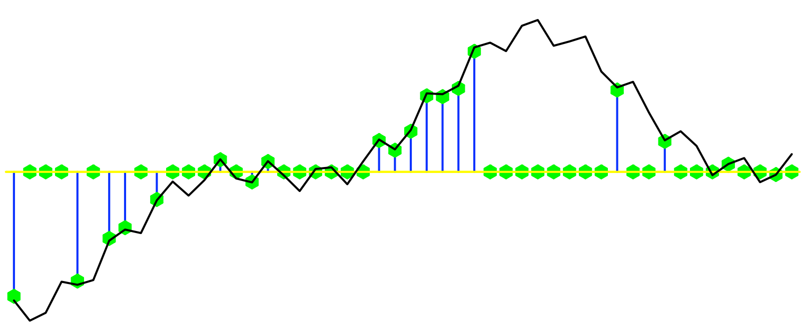


Fourier
transform

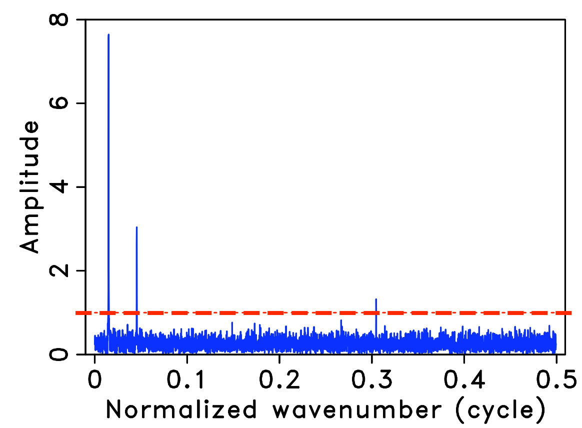


few significant coefficients

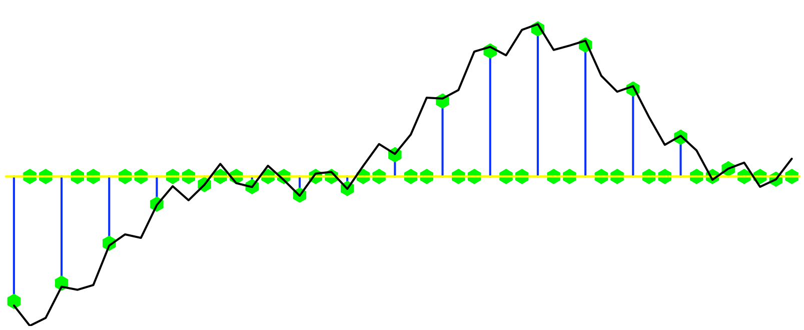
3-fold under-sampling



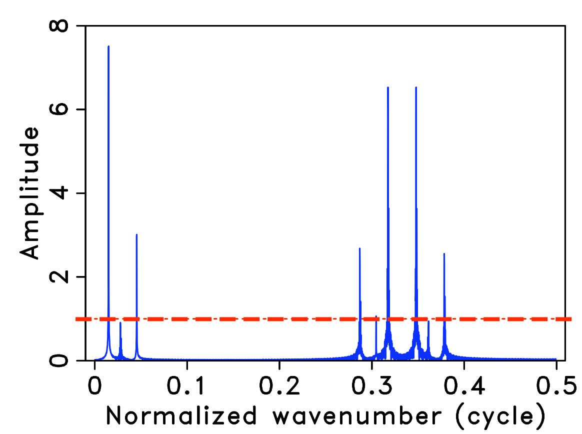
Fourier
transform



✓ significant coefficients detected



Fourier
transform



✗ ambiguity

[Hennenfent & Herrmann, '08]

Sparse one-norm recovery

Signal model

$$\mathbf{b} = \mathbf{A}\mathbf{x}_0 \quad \text{where} \quad \mathbf{b} \in \mathbb{R}^n$$

and $\mathbf{x}_0$ *k* sparse

Sparse one-norm recovery

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \stackrel{\text{def}}{=} \sum_{i=1}^{N} |x[i]| \quad \text{subject to} \quad \mathbf{b} = \mathbf{A}\mathbf{x}$$

with $\mathrm{n} \ll N$

Study recovery as a function of

- the subsampling ratio n/N
- "over sampling" ratio k/n

[Sacchi '98]
[Candès et.al, Donoho, '06]

Case study

Acquisition design according to Compressive Sensing

- ***Periodic*** *subsampling* vs ***randomized jittered*** sampling of ***sequential*** sources
- *Subsampling with randomized jittered* ***sequential*** sources vs *randomized phase-encoded* ***simultaneous*** sources

SLIM

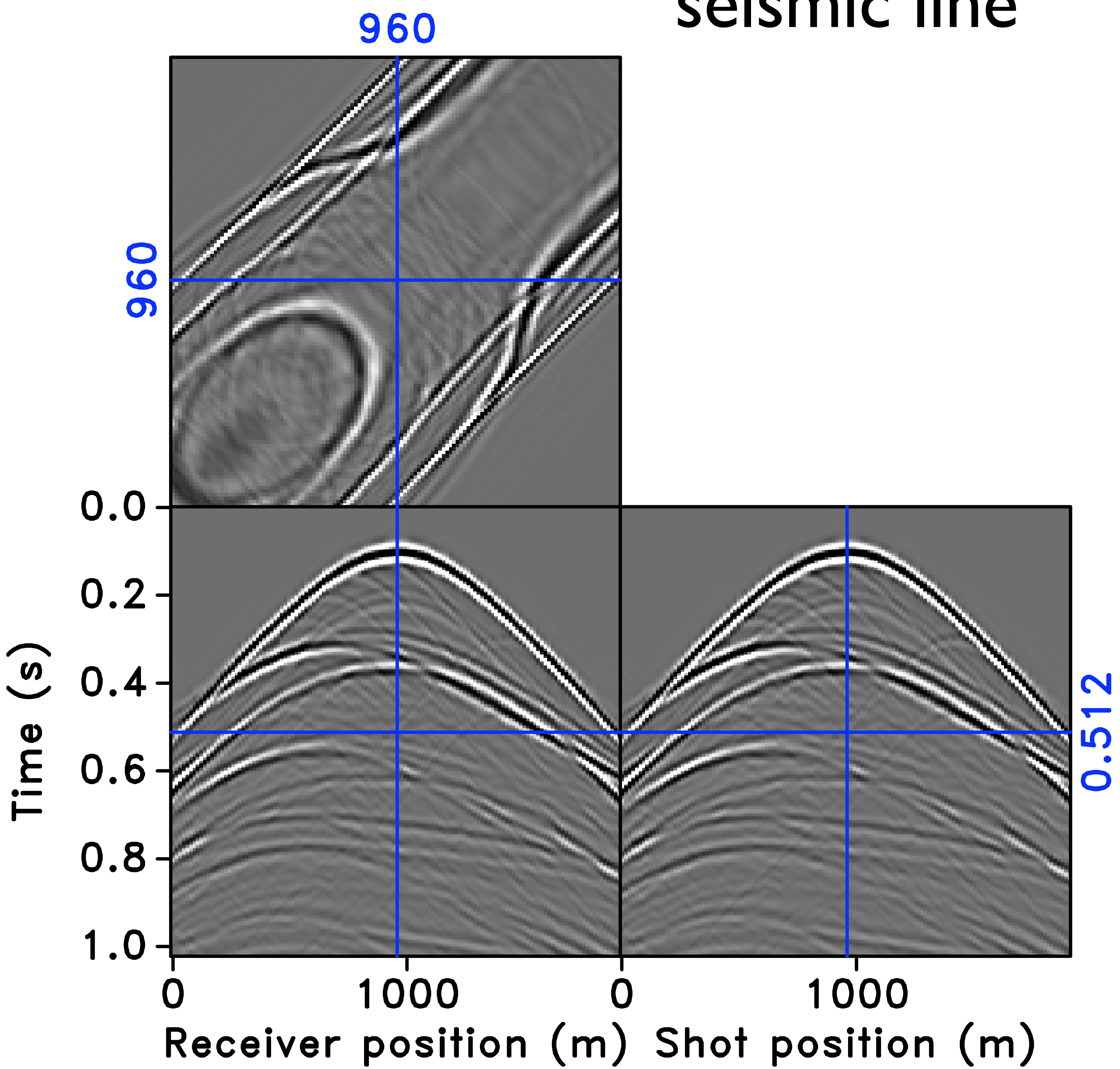
Pathology

shot interpolation

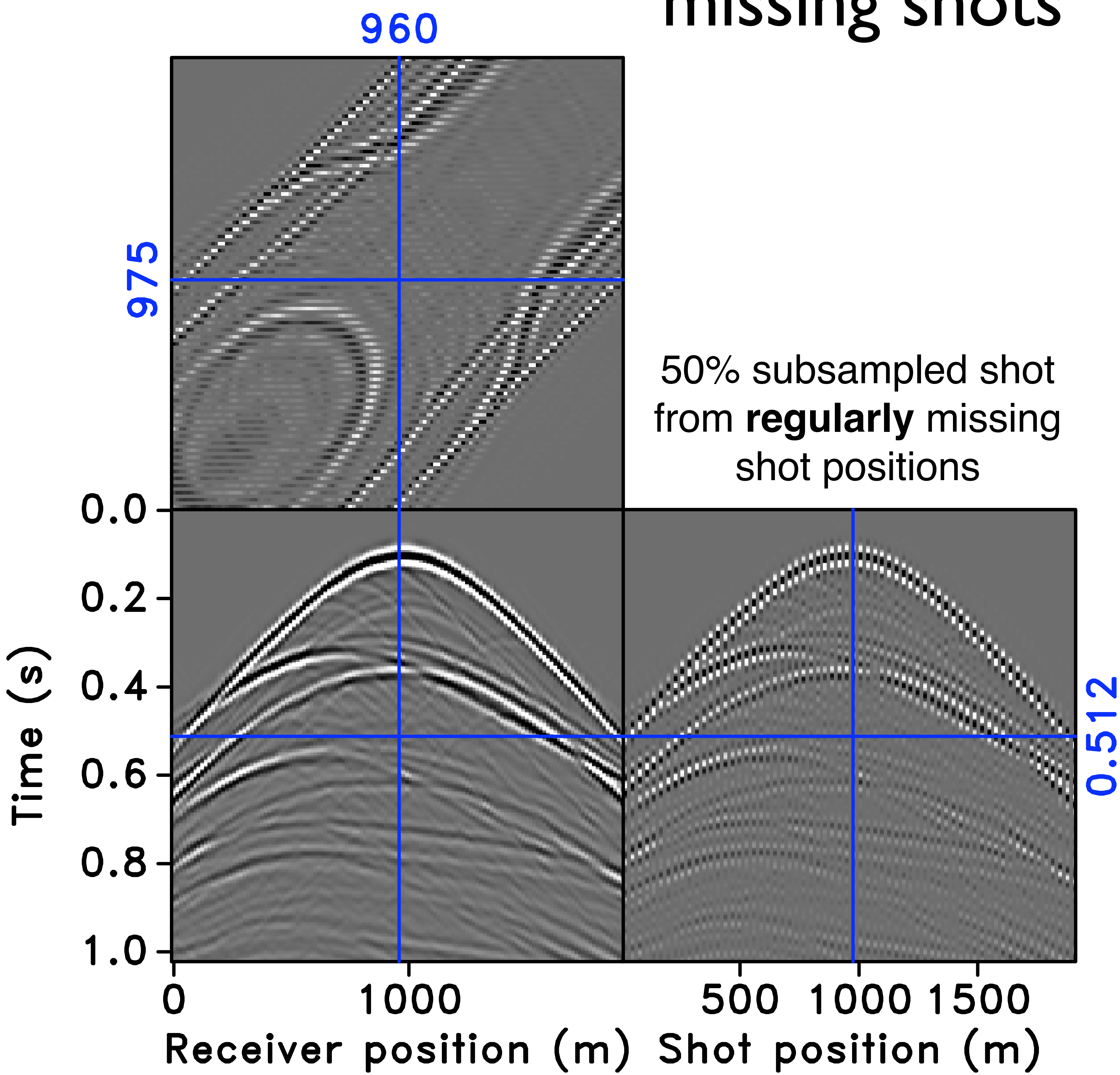
12.5m to 25m

50 % data-size *reduction*

seismic line

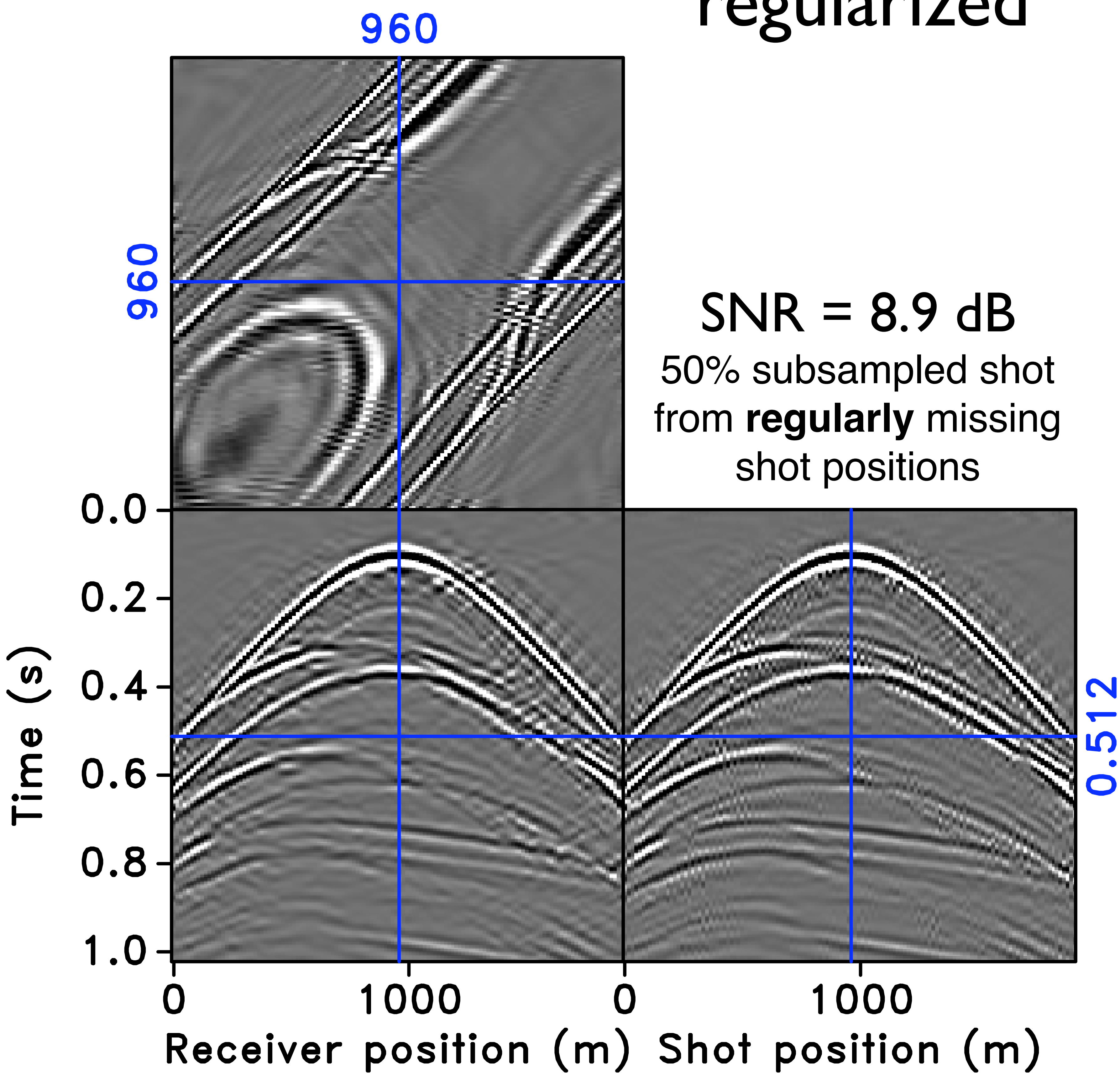


missing shots



50% subsampled shot from **regularly** missing shot positions

regularized



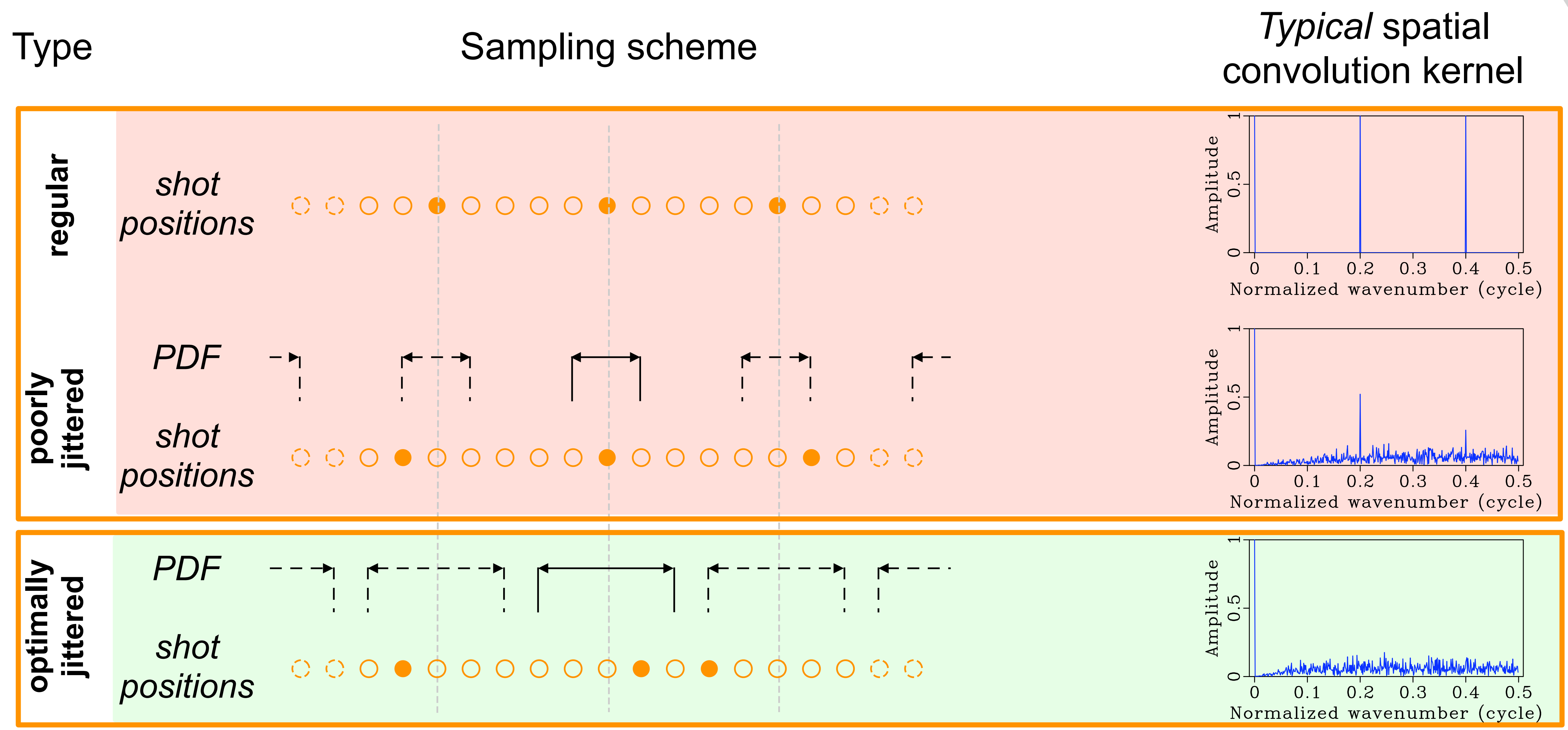
SNR = 8.9 dB

50% subsampled shot from **regularly** missing shot positions

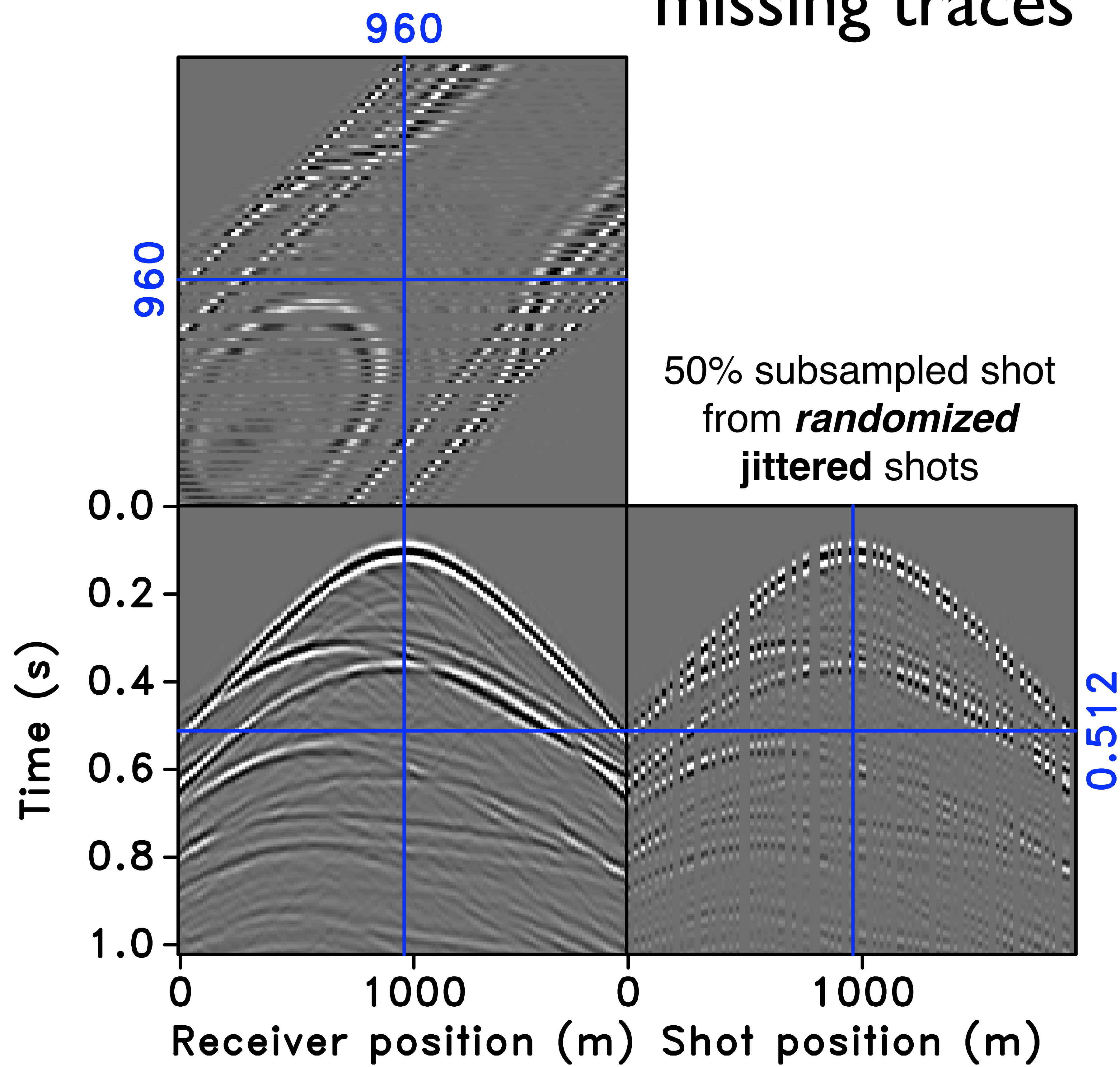
[Hennenfent & FJH, '08]
[Gang et.al., '09]

SLIM

Jittered sampling

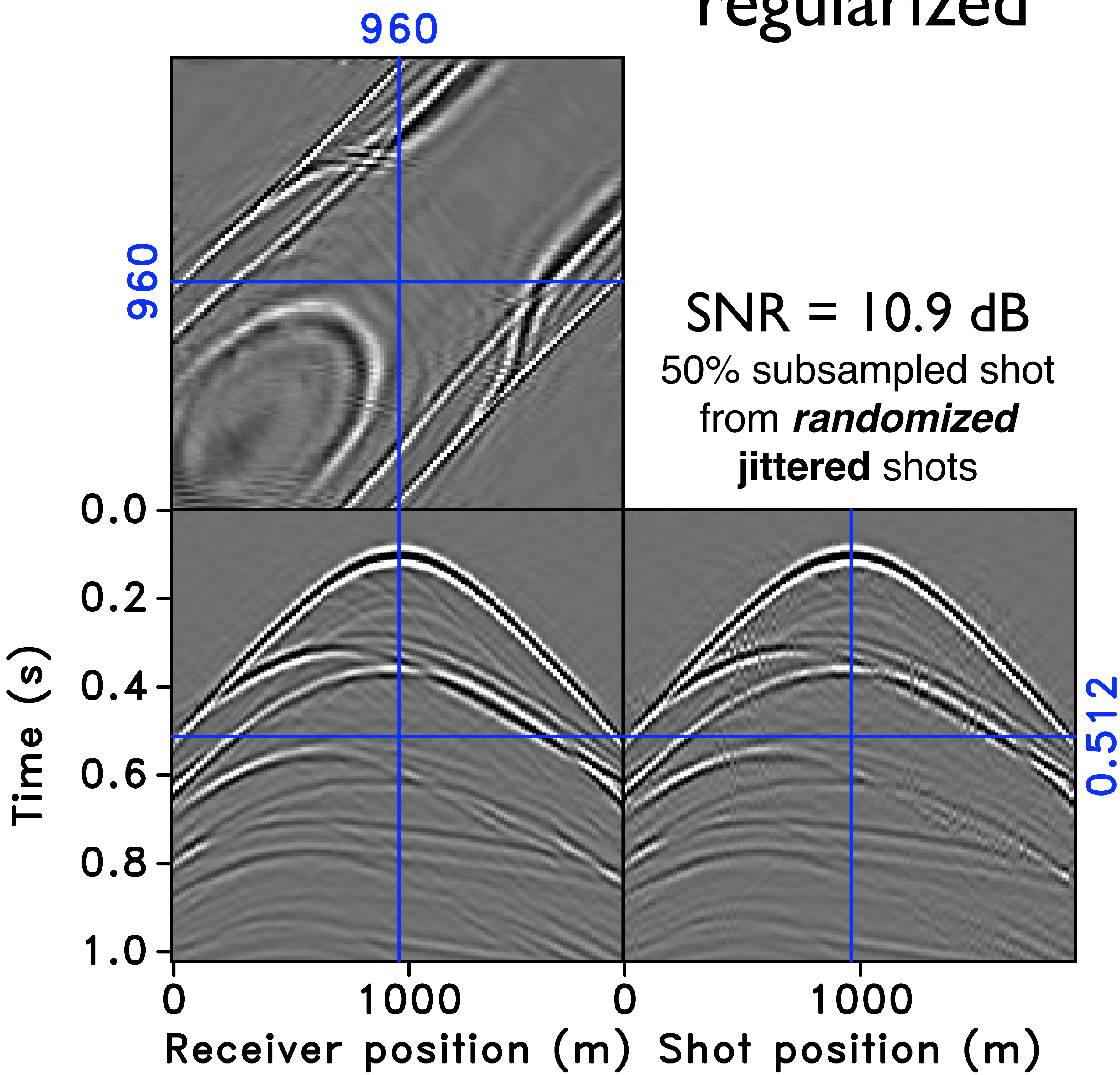


missing traces



50% subsampled shot from ***randomized*** **jittered** shots

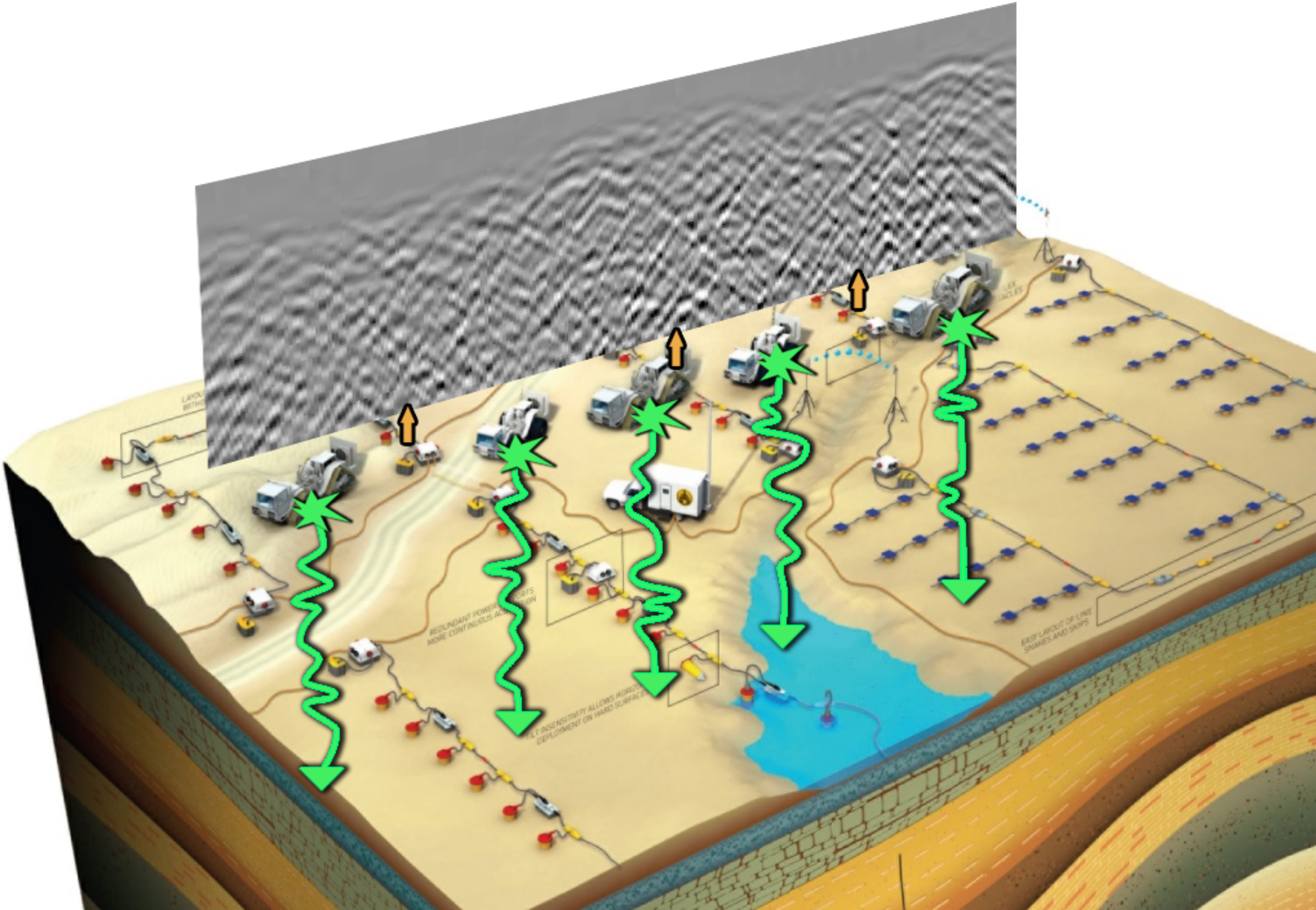
regularized



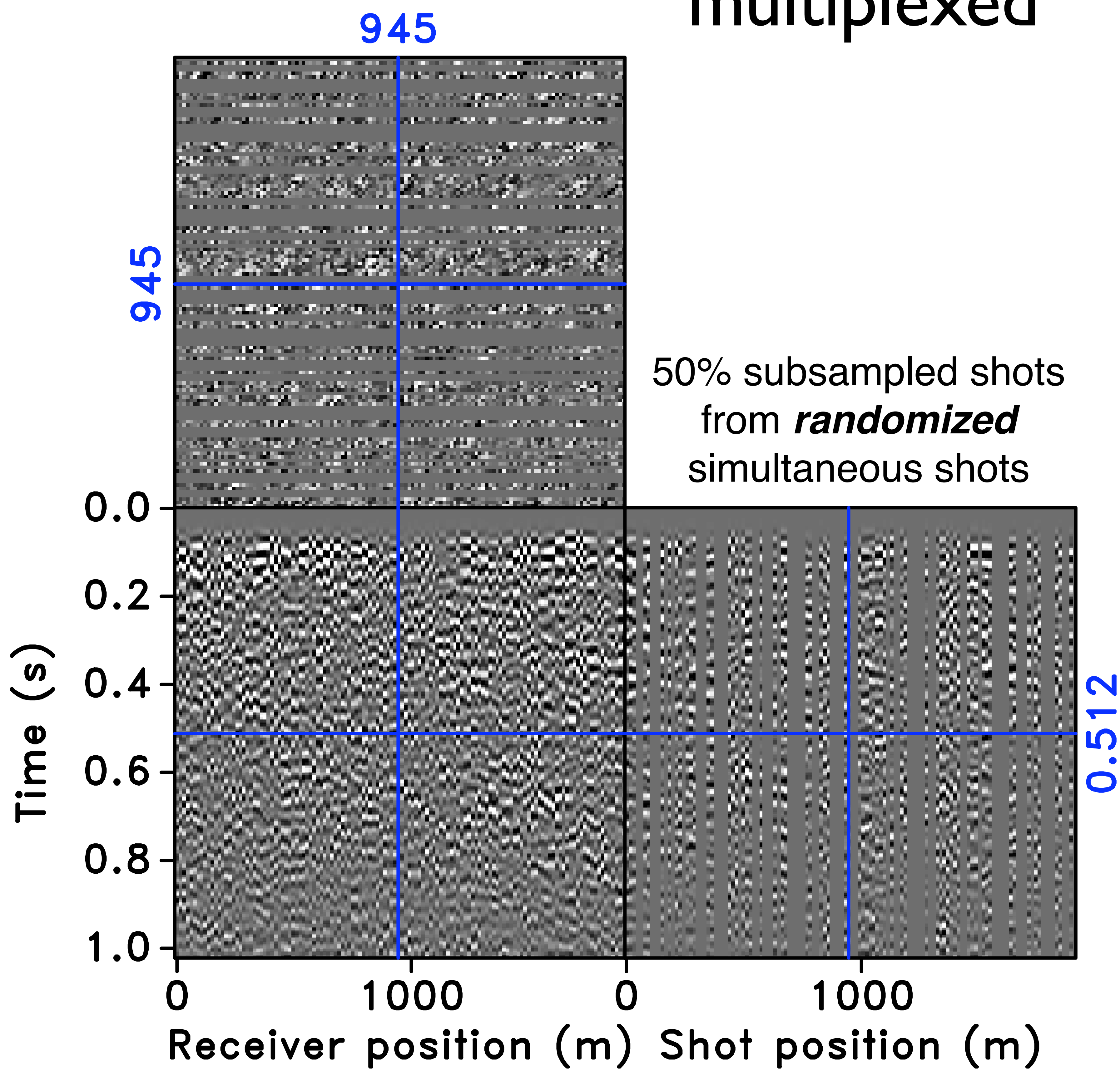
SNR = 10.9 dB

50% subsampled shot from ***randomized*** **jittered** shots

Simultaneous & incoherent sources

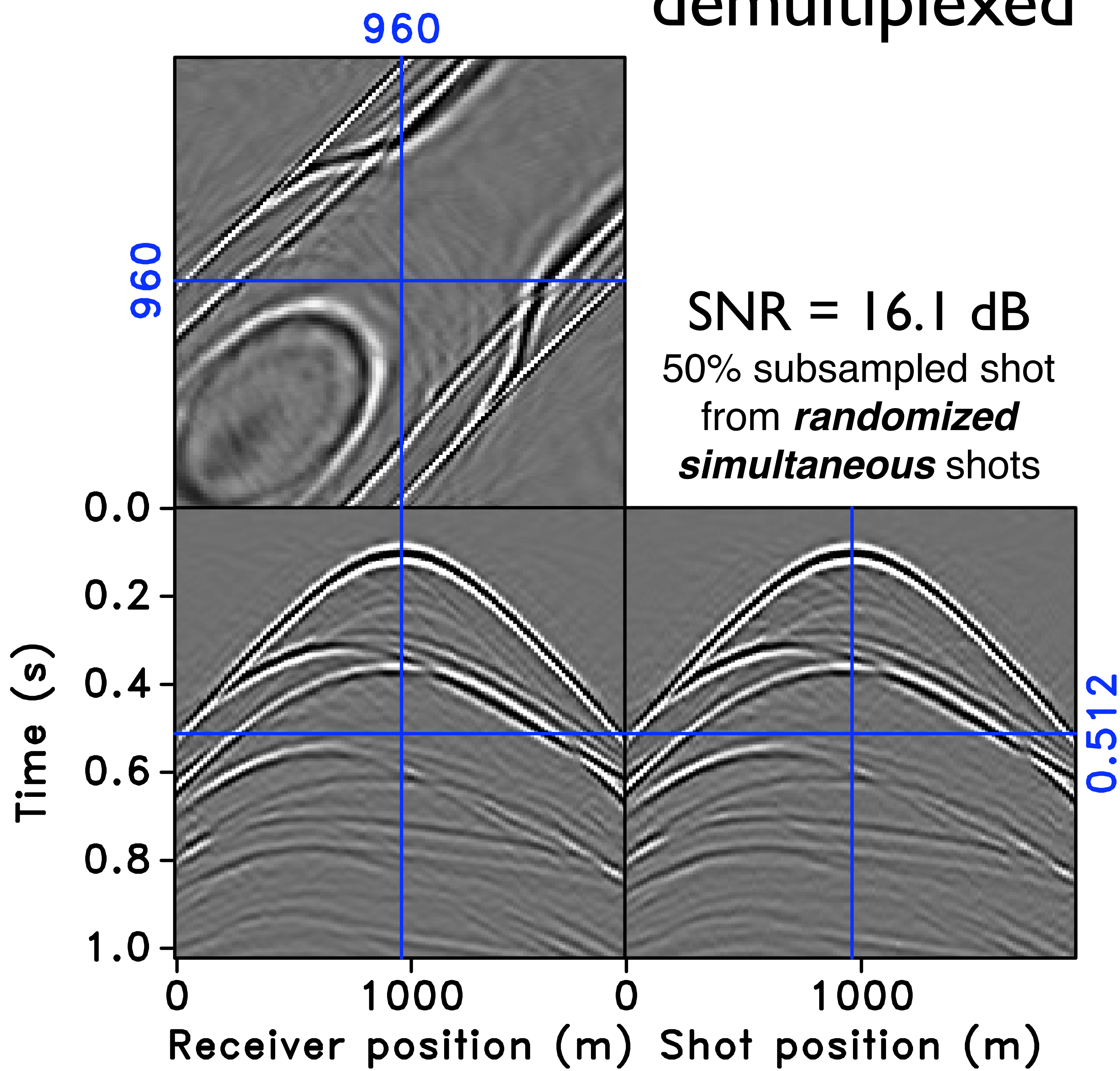


multiplexed



50% subsampled shots from ***randomized*** simultaneous shots

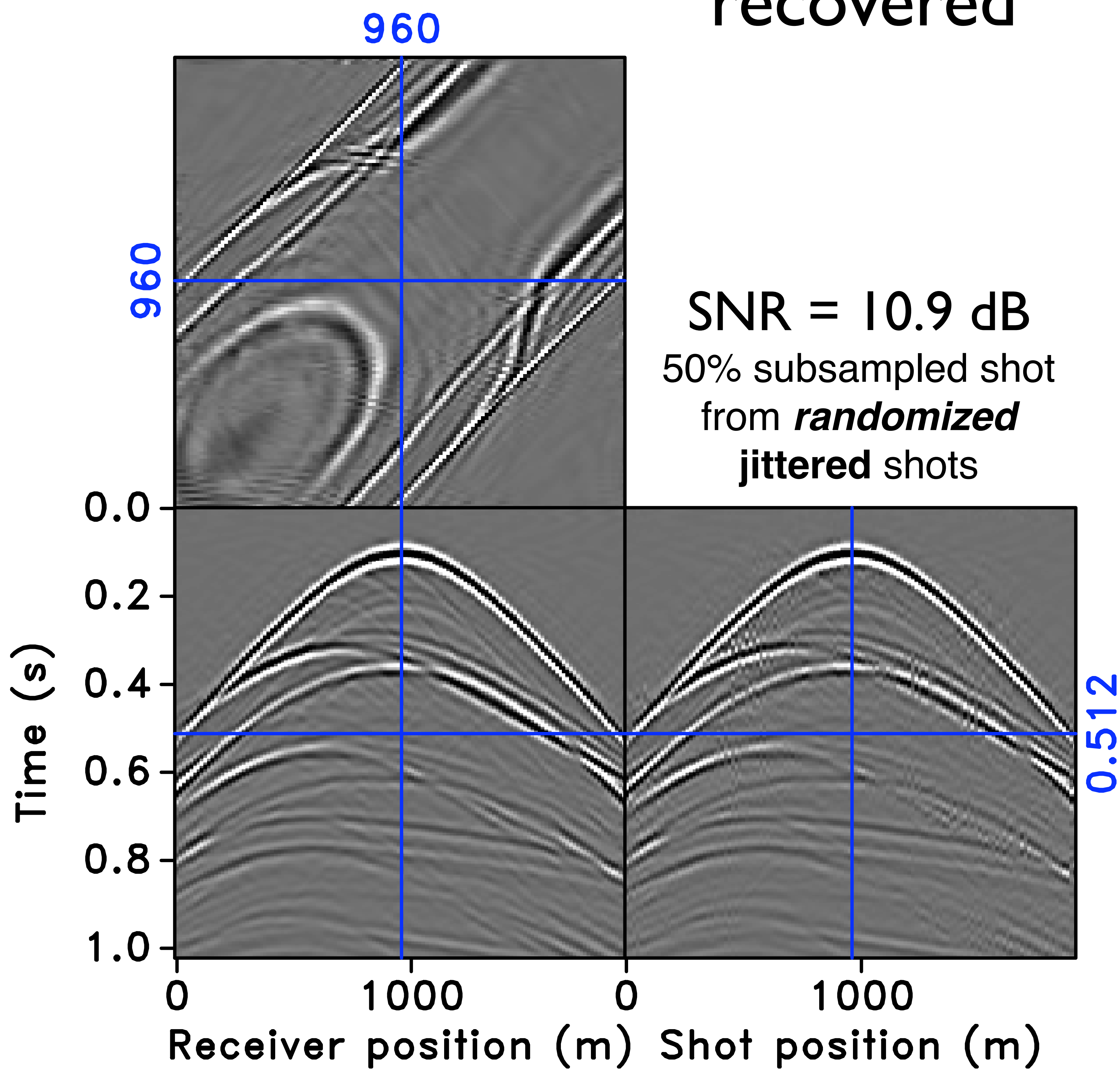
demultiplexed



SNR = 16.1 dB

50% subsampled shot from ***randomized simultaneous*** shots

recovered



SNR = 10.9 dB

50% subsampled shot from ***randomized*** **jittered** shots

The math of Compressive Sensing [Candès et.al, '06]

Recovery is *possible* & *stable* as long as each *subset S* of *k* columns of $\mathbf{A} \in \mathbb{R}^{n \times N}$ with $k \leq N$ the # of nonzeros *approximately* behaves as an orthogonal basis.

In that case, we have

$$(1-\hat{\delta}_k)\|\mathbf{x}_S\|_{\ell_2}^2 \leq \|\mathbf{A}_S\mathbf{x}_S\|_{\ell_2}^2 \leq (1+\hat{\delta}_k)\|\mathbf{x}_S\|_{\ell_2}^2,$$

where S runs over all sets with cardinality $\leq k$

- the smaller the *restricted isometry constant* (RIP) $\hat{\delta}_k$ the more *energy* is captured and the more *stable* the *inversion* of **A**
- determined by the *mutual coherence* of the cols in **A**

Math cont'ed

[Candès et.al, '06]

RIP constant is bounded by

$$\hat{\delta}_k \leq (k-1)\mu$$

where

$$\mu = \max_{1 \leq i \neq j \leq N} |\mathbf{a}_i^H \mathbf{a}_j|$$

Matrices with small $\hat{\delta}_k$ contain subsets of *k* *incoherent* columns.

Gaussian random matrices with *i.i.d.* entries have this property.

One-norm solvers recover $\mathbf{x}_0$ as long it is *k* sparse and

$$k \leq C \cdot \frac{n}{\log_2(N/n)},$$

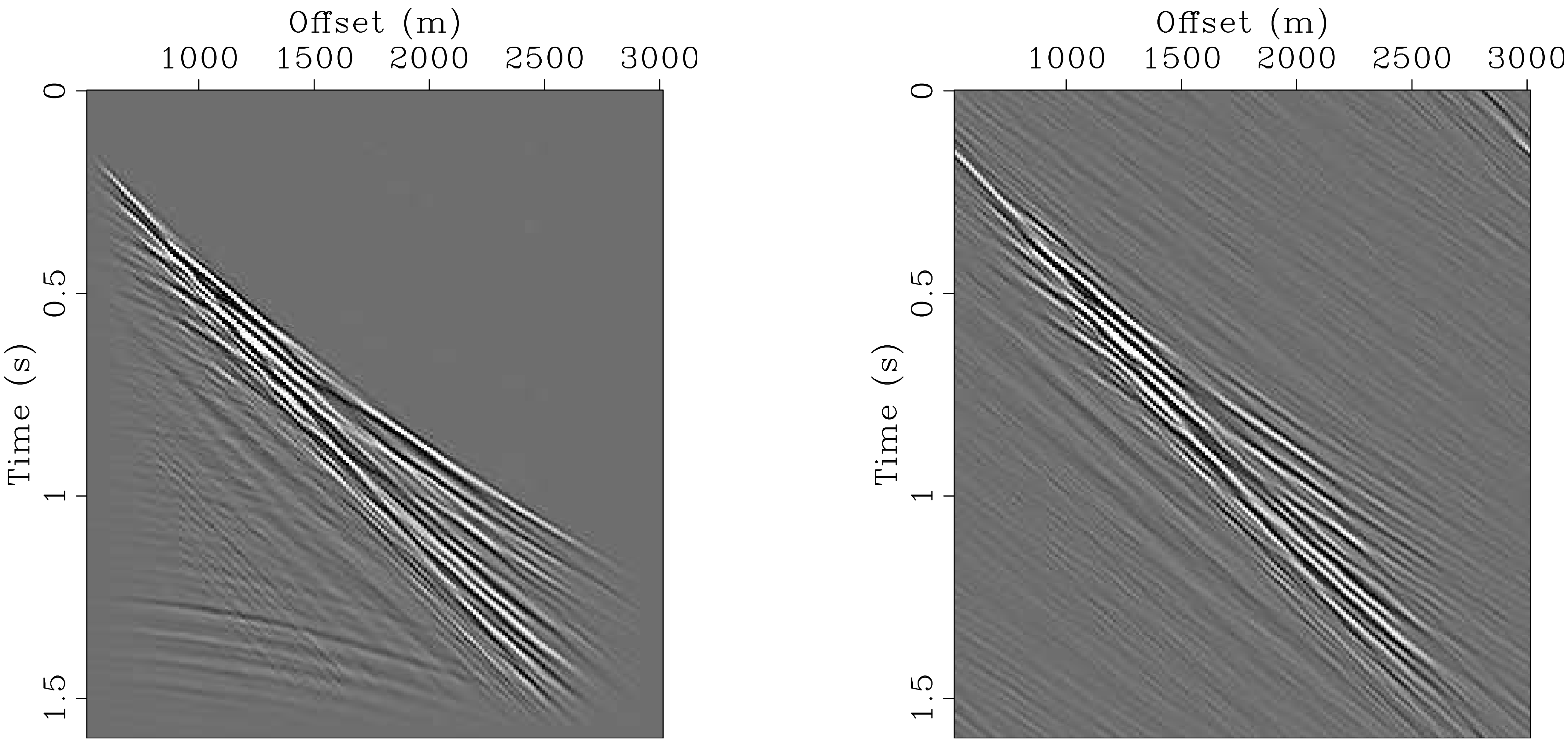
yields an *oversampling ratio* of

$$n/k \approx C \cdot \log_2 N$$

Key elements

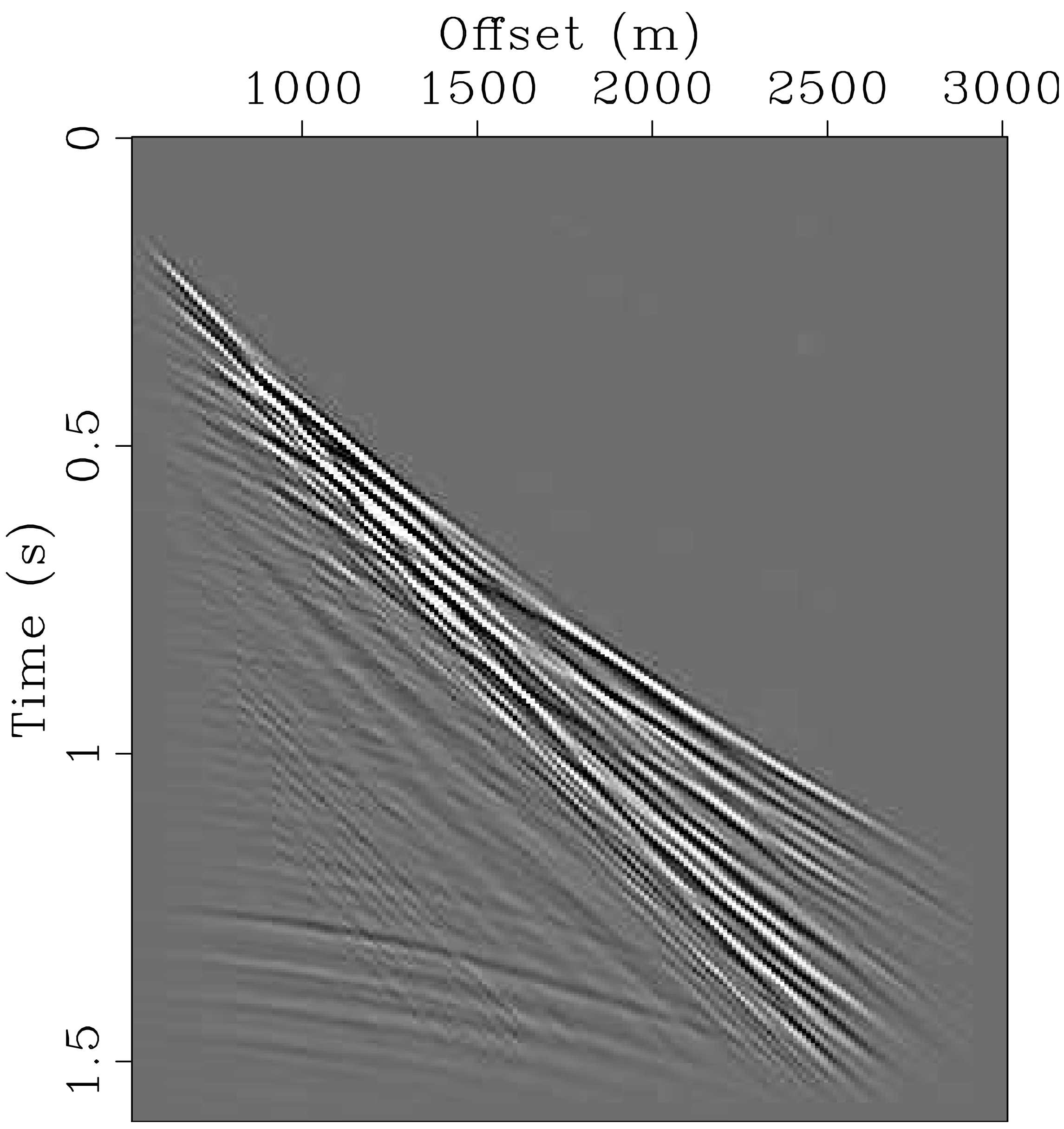
- *sparsifying transform*
  - typically **localized** in the time-space domain to handle the complexity of seismic data
- *advantageous coarse randomized sampling*
  - generates incoherent random undersampling "noise" in the sparsifying domain
- *sparsity-promoting solver*
  - requires few matrix-vector multiplications

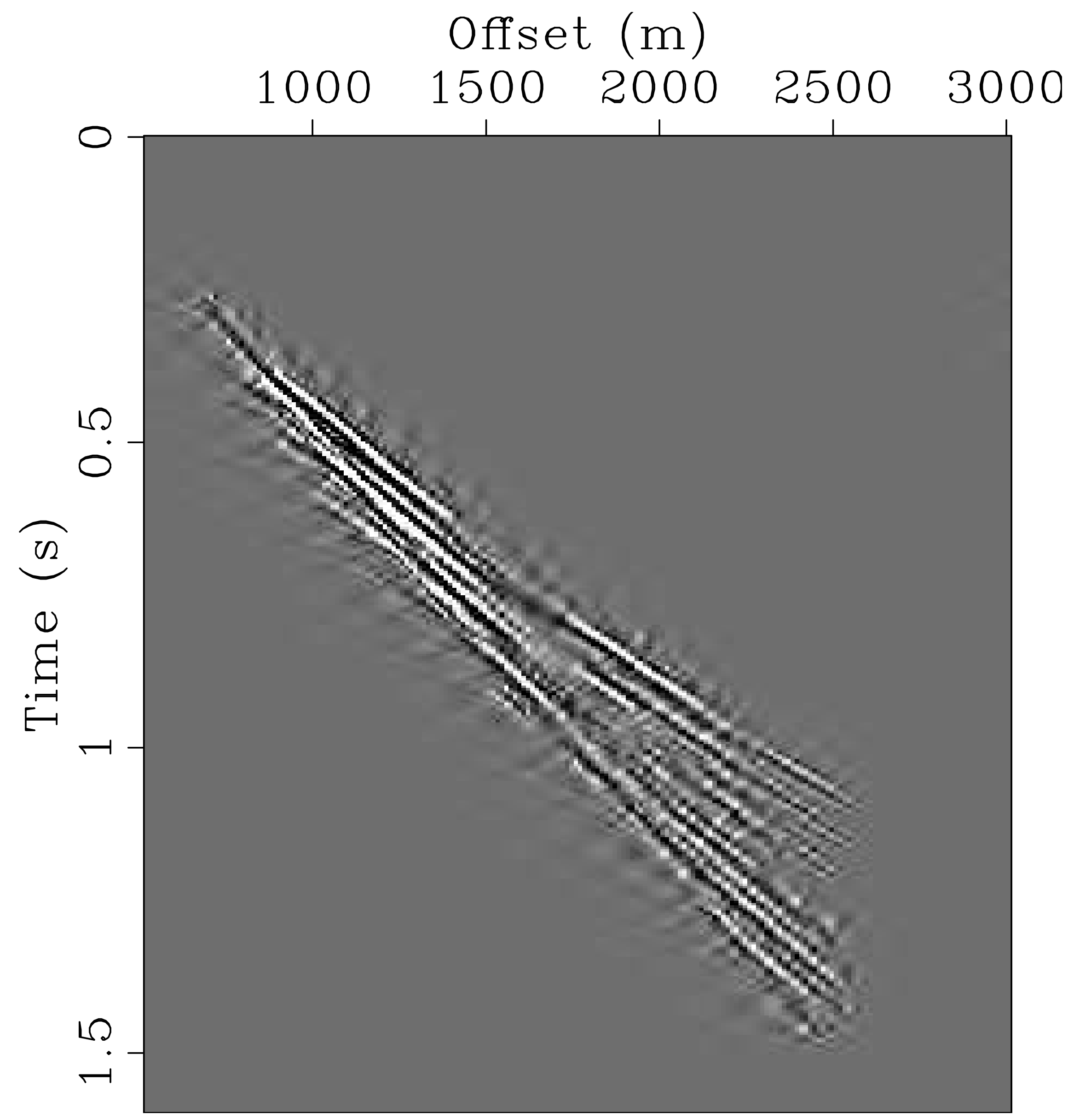
Fourier reconstruction



1 % of coefficients

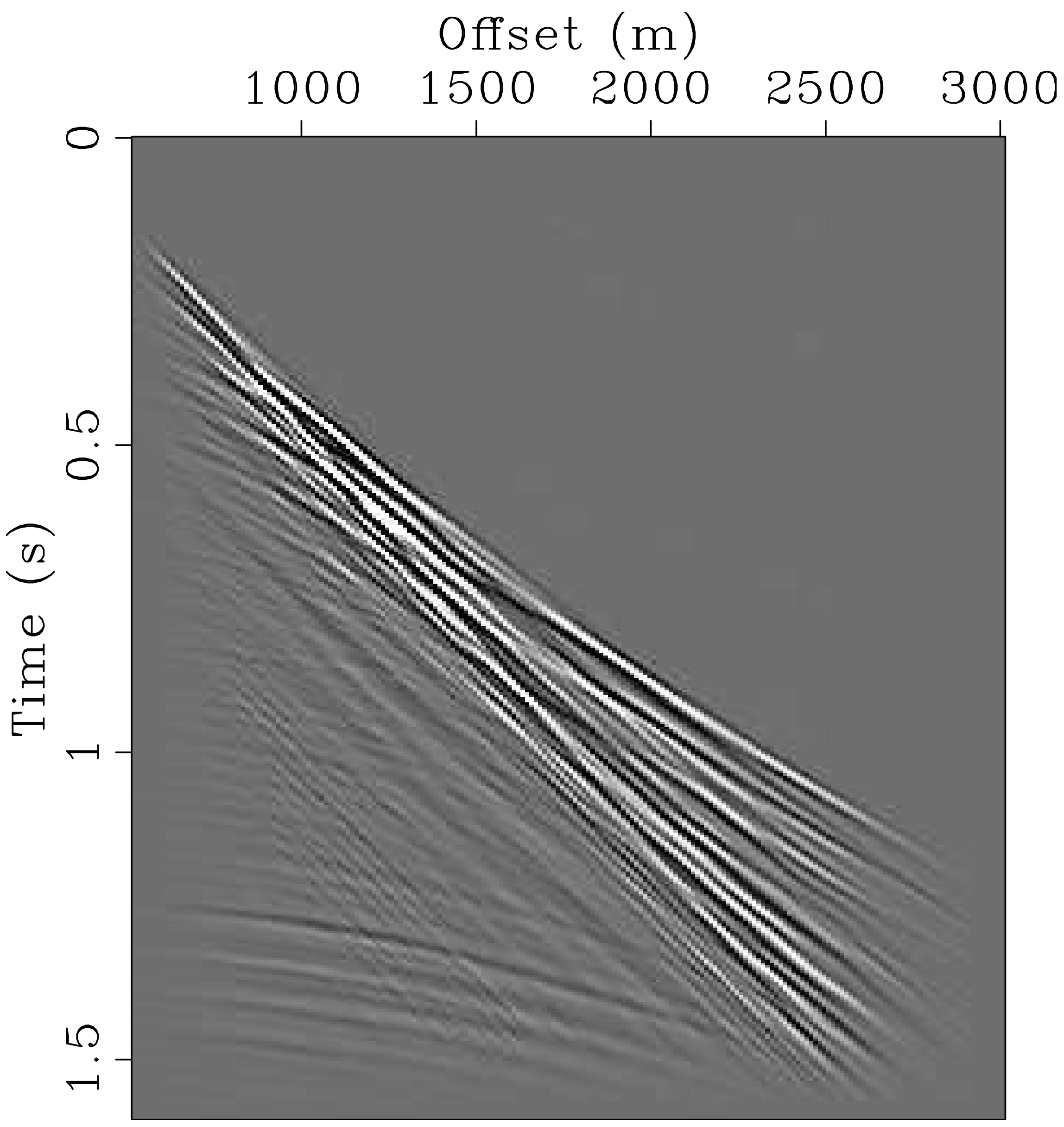
Wavelet reconstruction

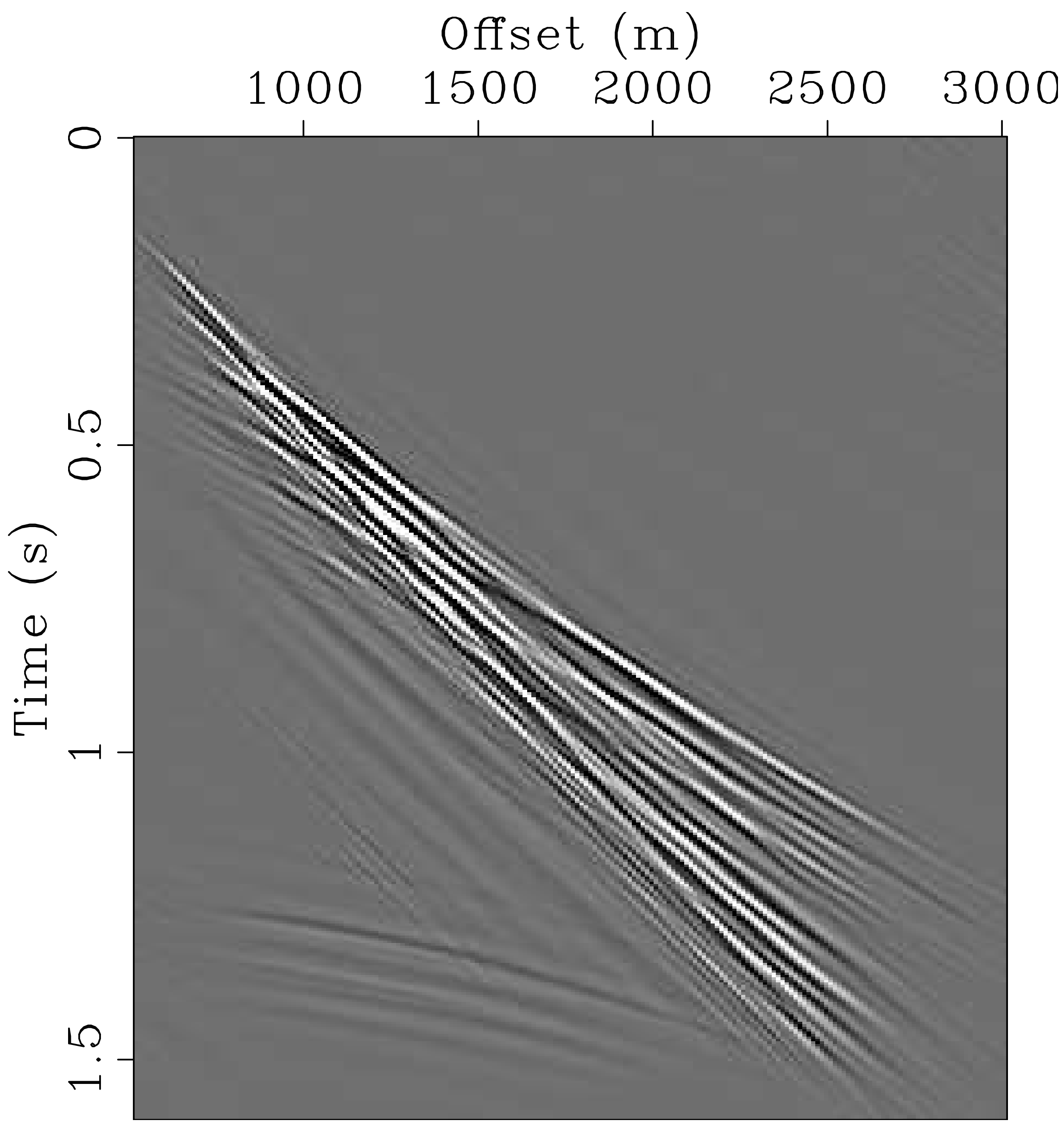




1 % of coefficients

Curvelet reconstruction

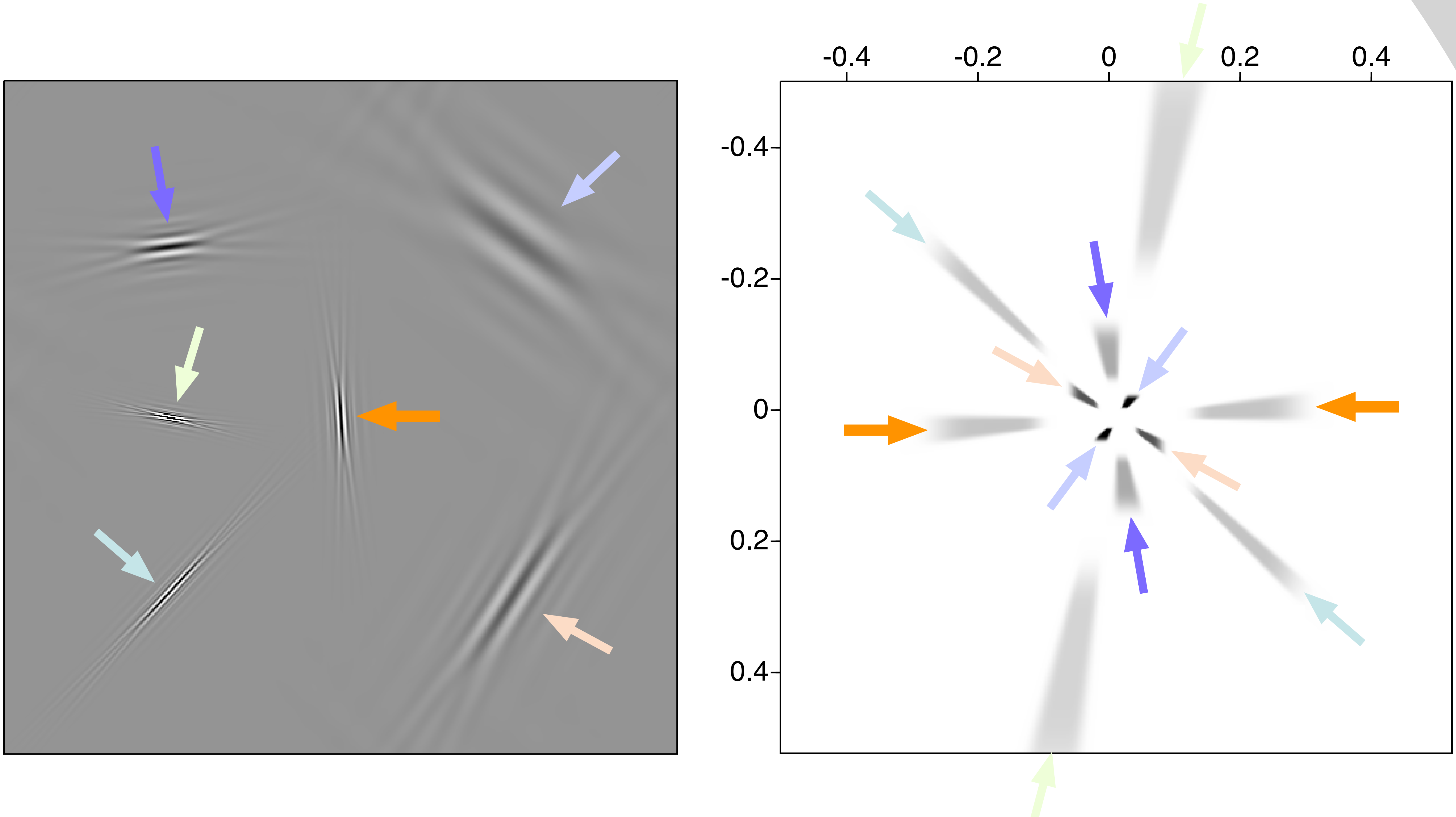




1 % of coefficients

[Demanet et. al., ‘06]

Curvelets



Extension

Extend CS framework:

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

restriction matrix | measurement matrix | sparsity matrix

Expected to perform well when

$$\mu = \max_{1 \le i \ne j \le N} |\left(\mathbf{RMs}^i\right)^H \mathbf{RMs}^j|$$

Generalizes to *redundant* transforms for cases where

- max of RIP constants for **M**, **S** are small [Rauhut et.al, '06]
- or $\mathrm{SS}^H\mathrm{x}$ remains sparse for **x** sparse [Candès et.al, '10]

Open research topic...

SLIM

Empirical performance analysis

Selection of the appropriate sparsifying transform

- nonlinear approximation error

$$\mathrm{SNR}(\rho) = -20\log\frac{\|\mathbf{f}-\mathbf{f}_\rho\|}{\|\mathbf{f}\|} \quad \text{with} \quad \rho = k/P$$

- recovery error

$$\mathrm{SNR}(\delta) = -20\log\frac{\|\mathbf{f}-\tilde{\mathbf{f}}_\delta\|}{\|\mathbf{f}\|} \quad \text{with} \quad \delta = n/N$$

- oversampling ratio

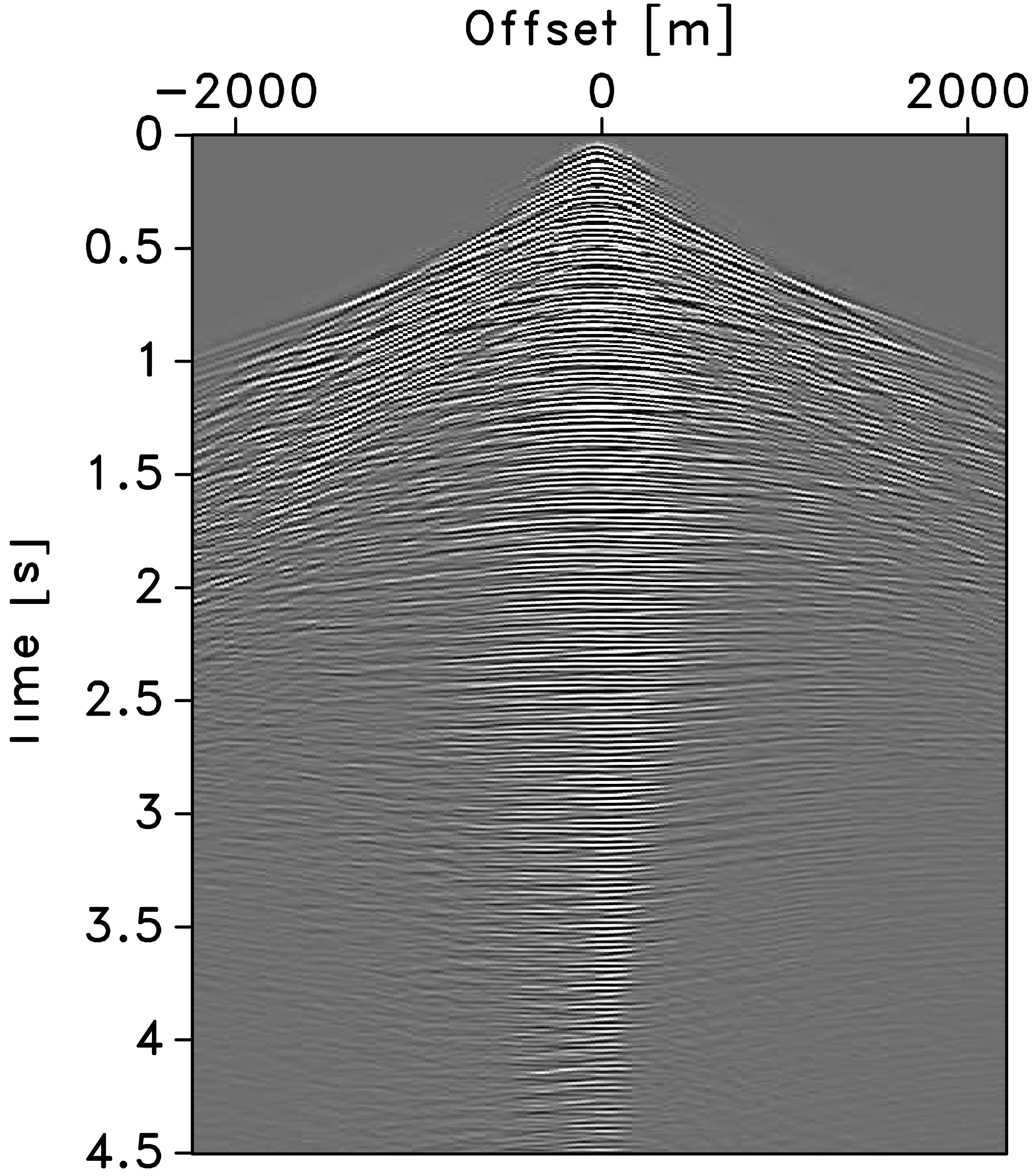
$$\delta/\rho \quad \text{with} \quad \rho = \inf\{\tilde{\rho} : \quad \overline{\mathrm{SNR}}(\delta) \leq \mathrm{SNR}(\tilde{\rho})\}$$

[FJH, '10]

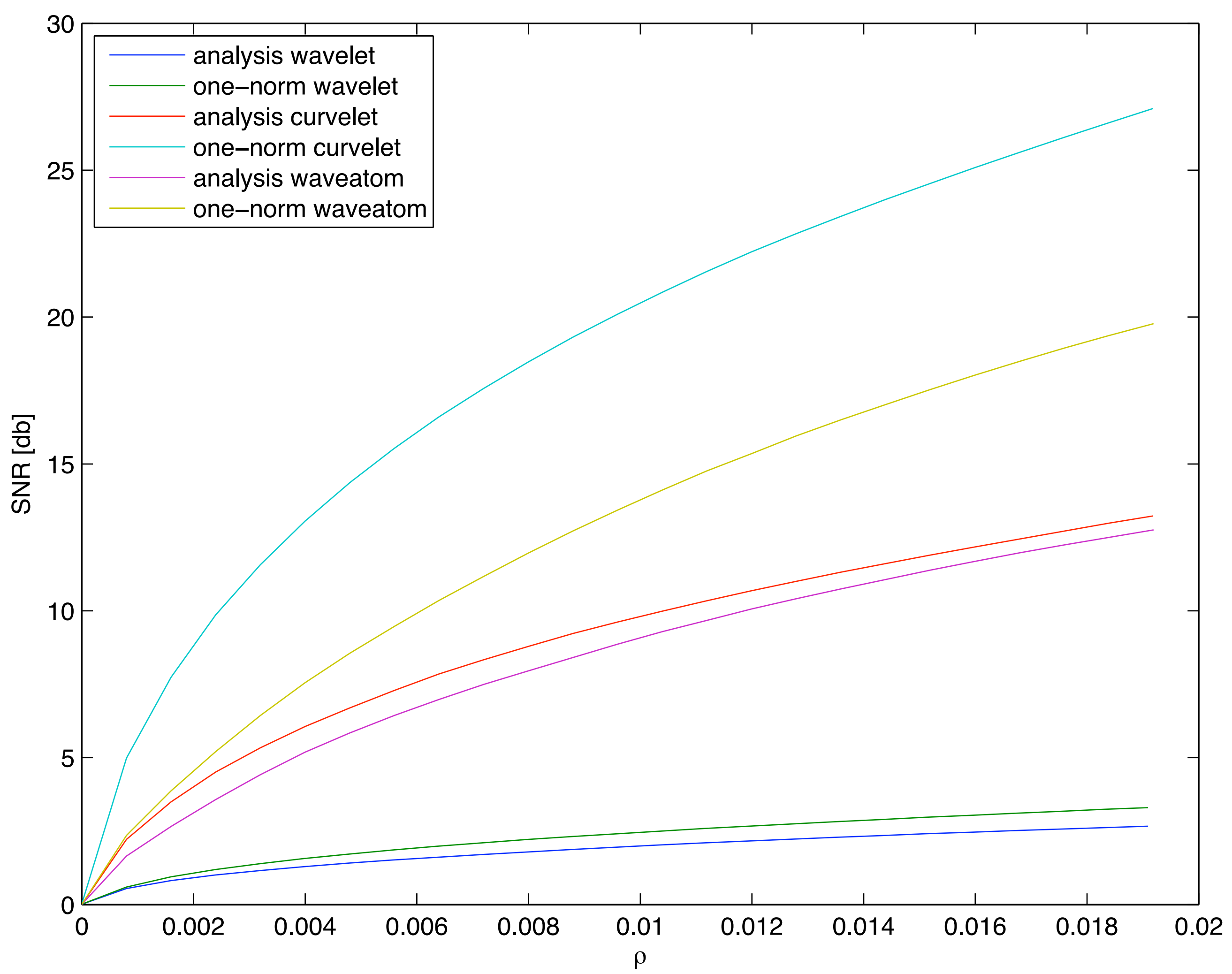
Nonlinear approximation error

SLIM

common receiver gather



recovery error



Key elements

☑ *sparsifying transform*

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☐ *advantageous coarse sampling*

- generates incoherent random undersampling "noise" in the sparsifying domain

☐ *sparsity-promoting solver*

- requires few matrix-vector multiplications

Key elements

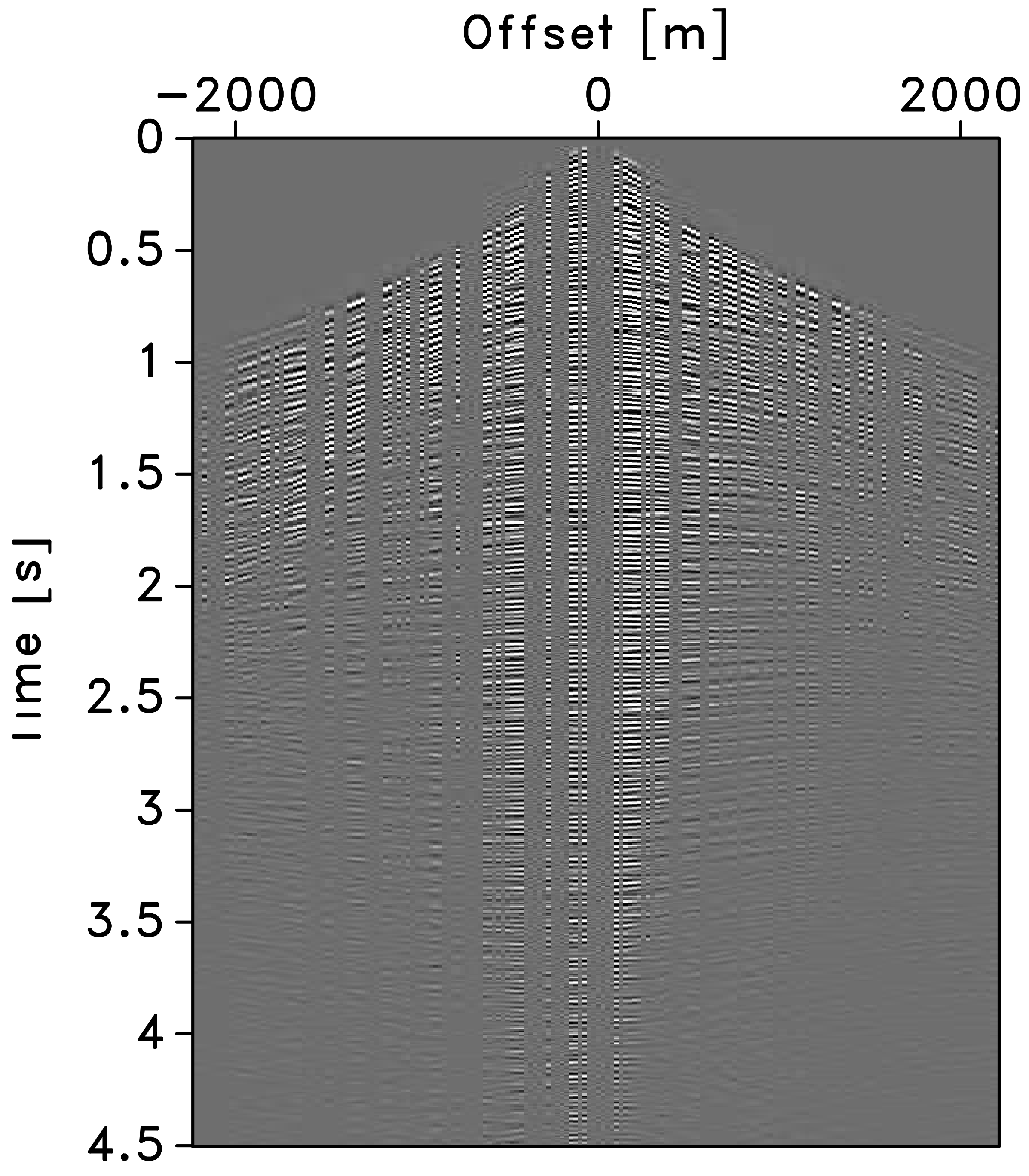
- [x] *sparsifying transform*
  - typically **localized** in the time-space domain to handle the complexity of seismic data
  - **curvelets**

- [x] *advantageous coarse sampling*
  - generates incoherent random undersampling "noise" in the sparsifying domain
  - does not create large gaps for measurement in the physical domain
  - does not create coherent interferences in simultaneous acquisition

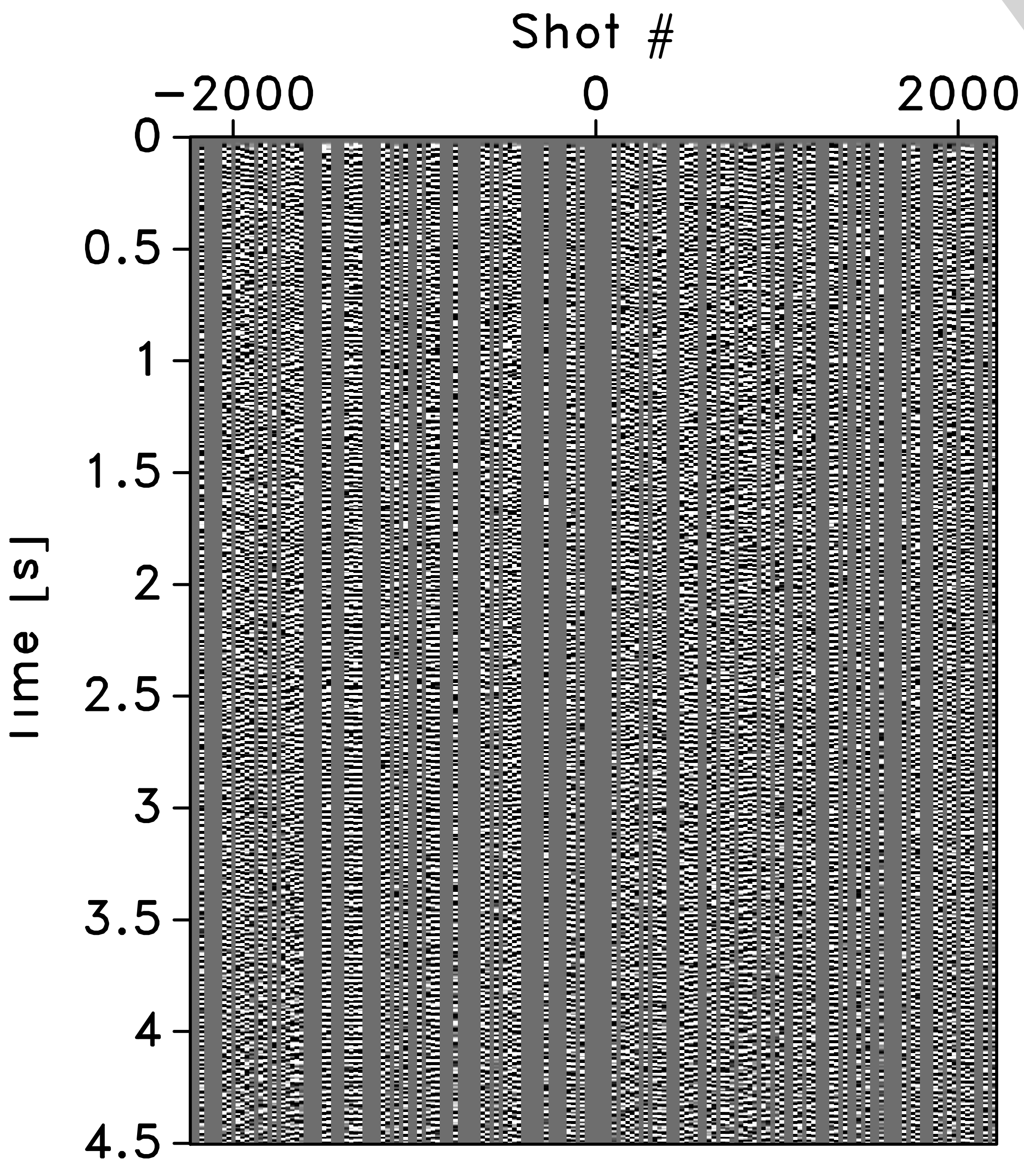
- [ ] *sparsity-promoting solver*
  - requires few matrix-vector multiplications

Data

missing shots

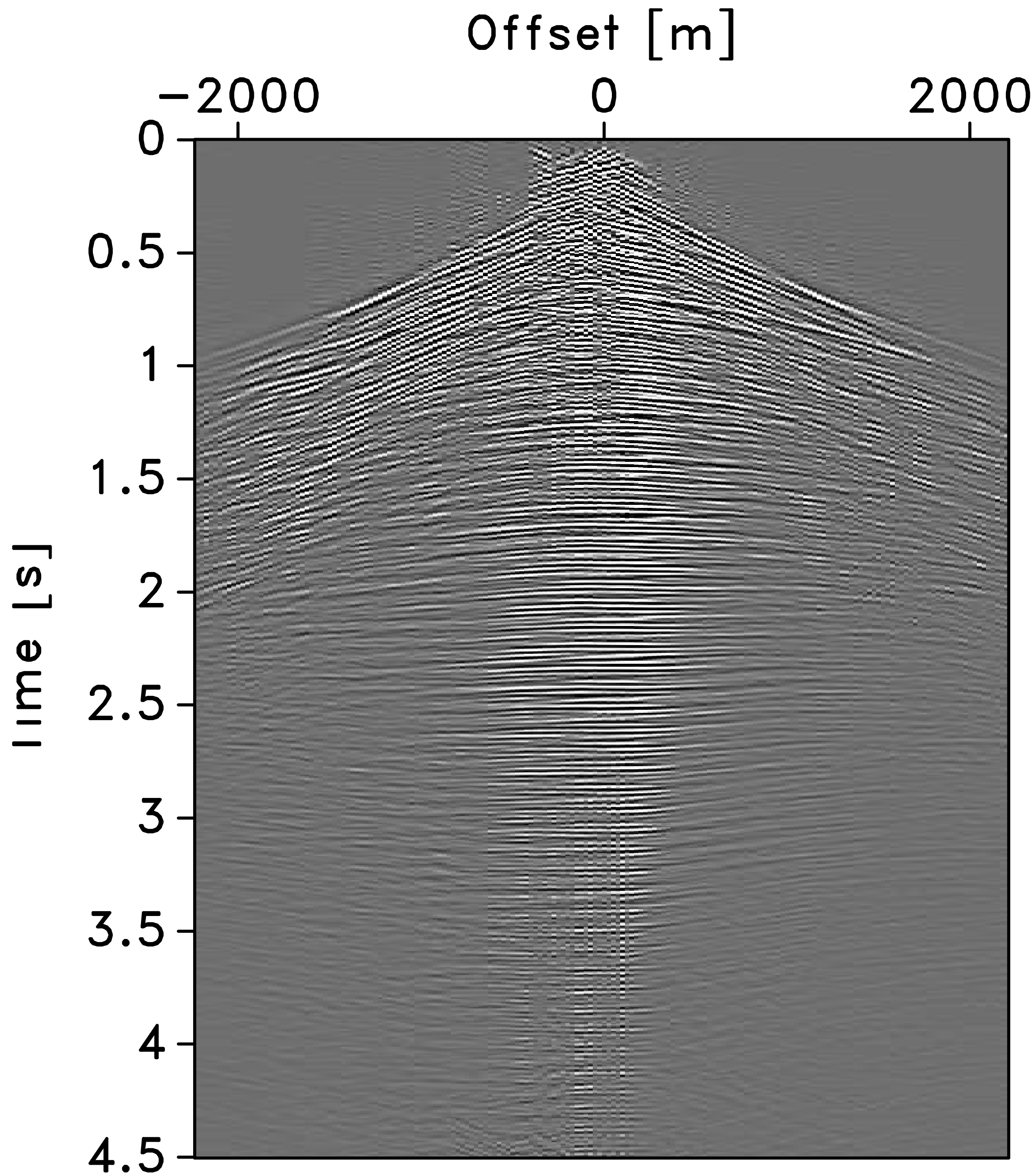


sim. shots

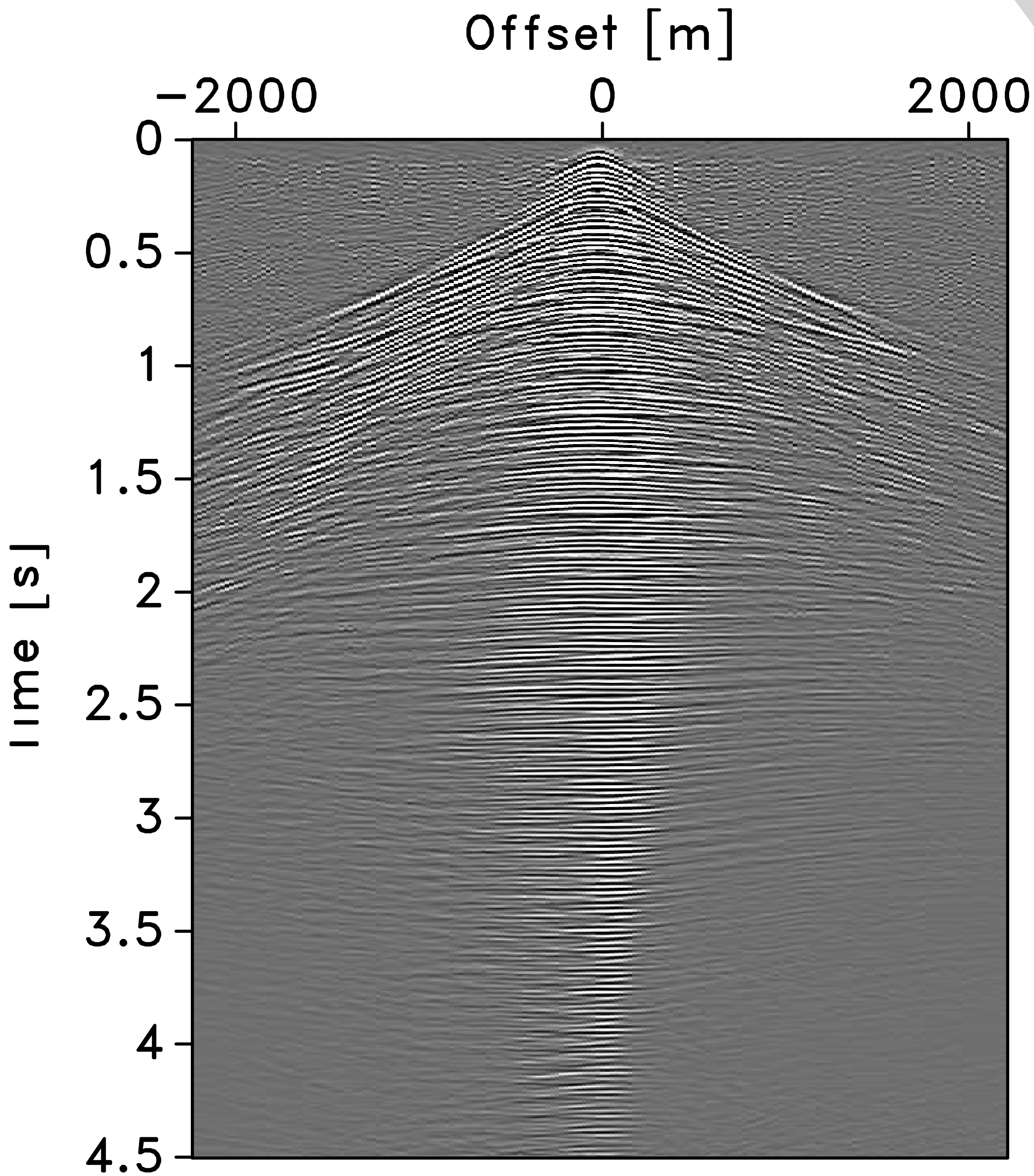


Sparse recovery

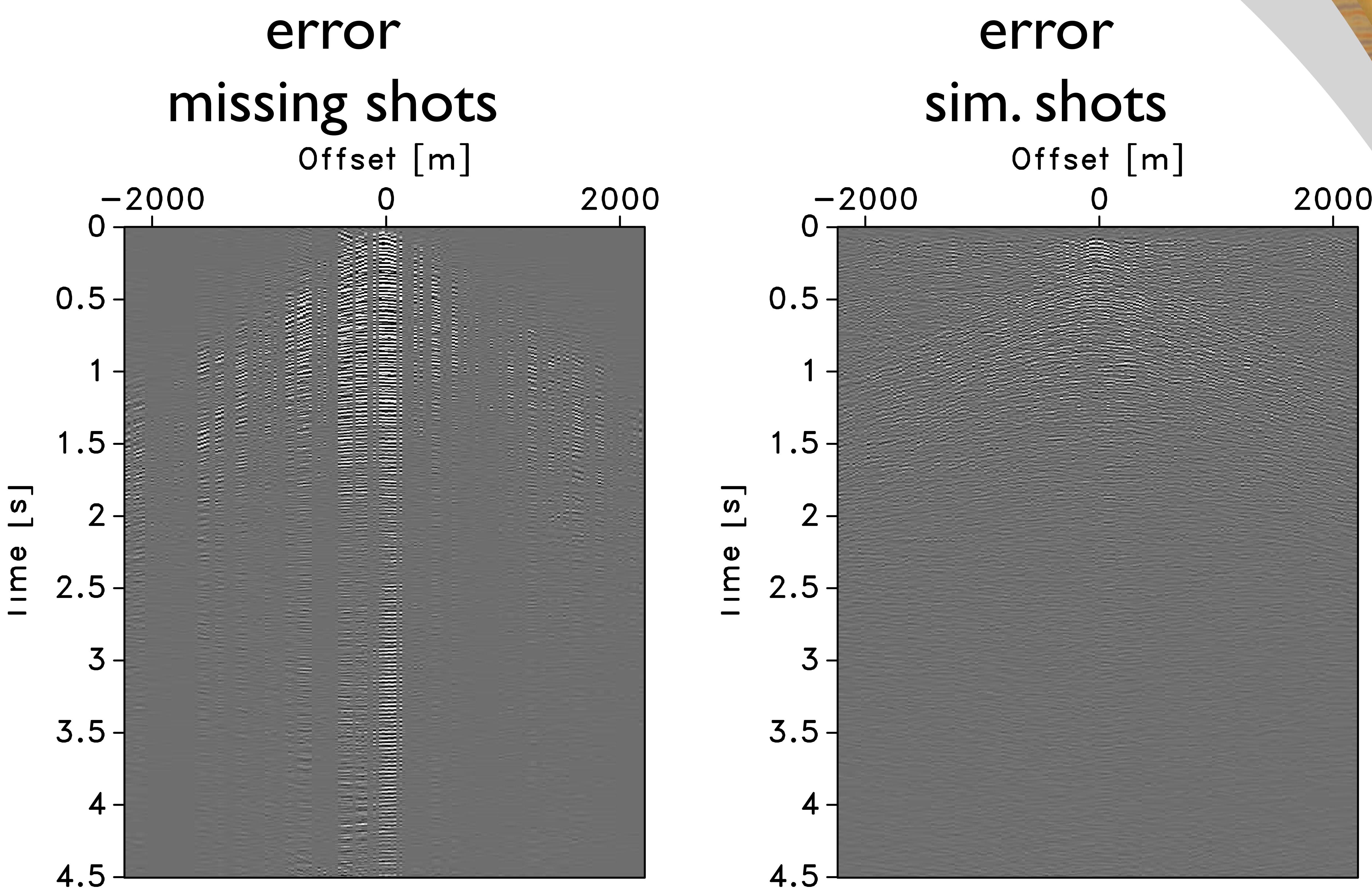
recovery
missing shots



recovery
sim. shots

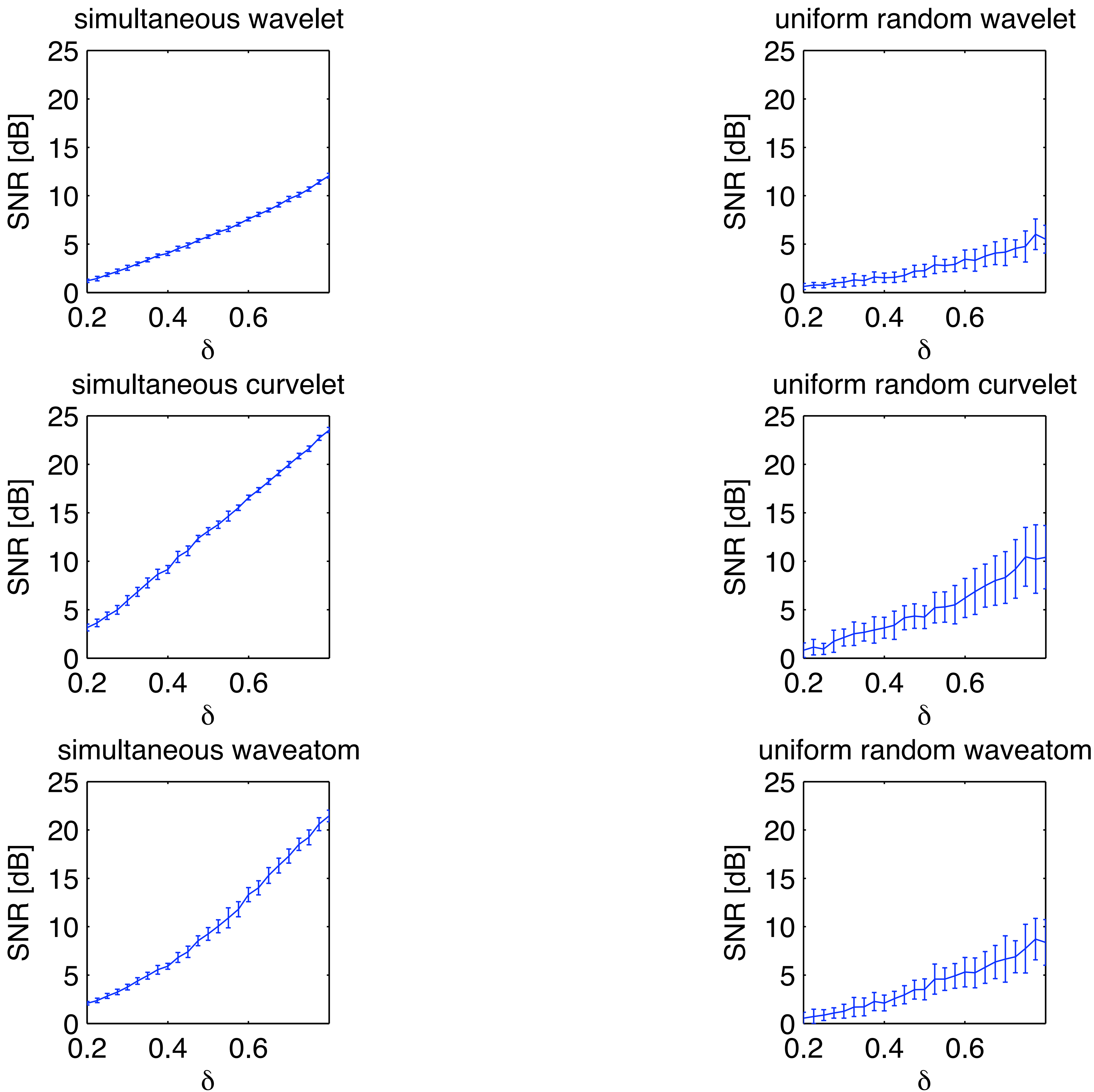


Sparse recovery error

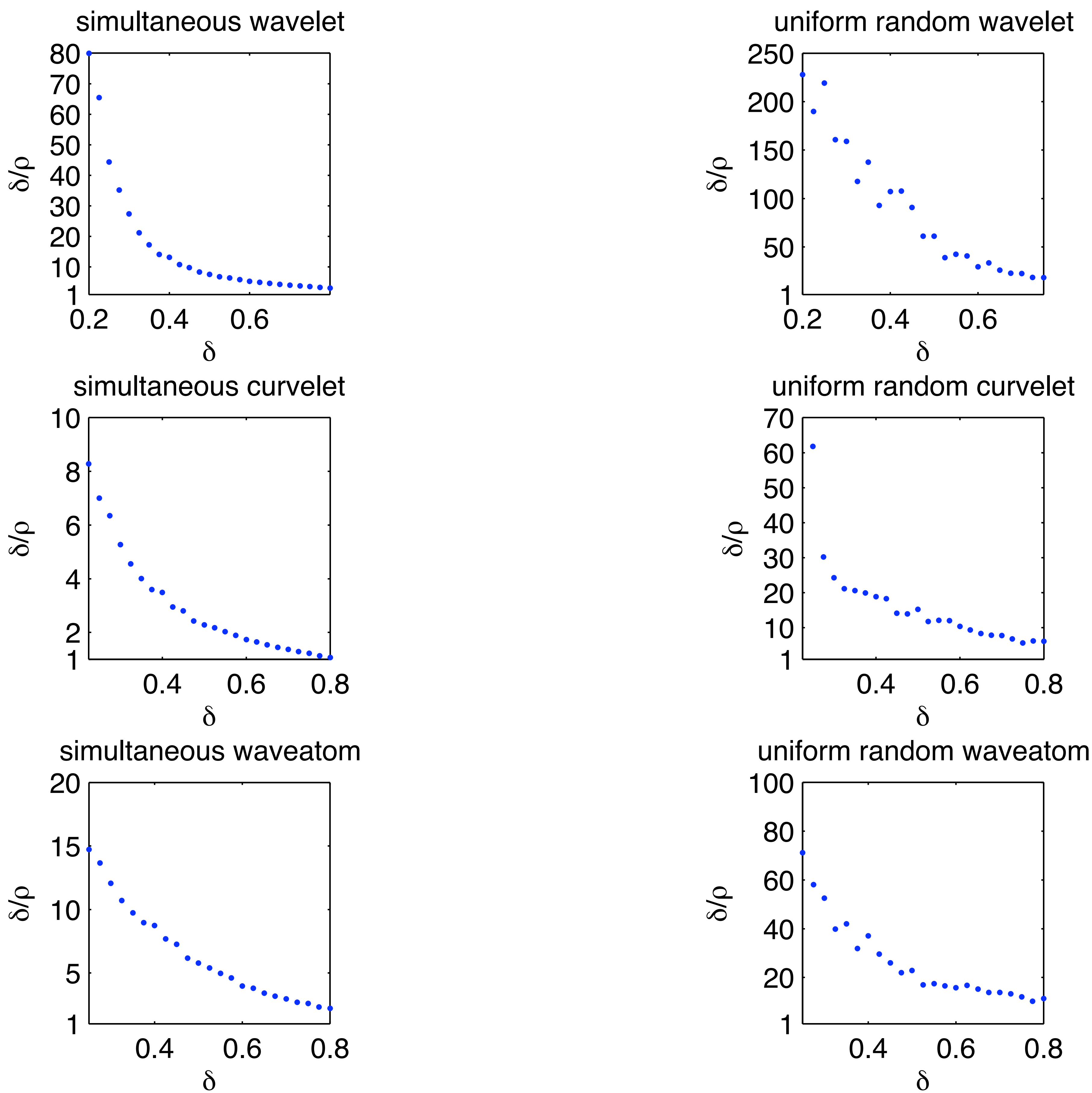


SLIM

Multiple experiments



Oversampling ratios



Key elements

SLIM

☑ *sparsifying transform*

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☑ *advantageous coarse sampling (mixing)*

- generates incoherent random undersampling "noise" in the sparsifying domain
- does not create large gaps for measurement in the physical domain
- does not create coherent interferences in simultaneous acquisition

☐ *sparsity-promoting solver*

- requires few matrix-vector multiplications

Reality check

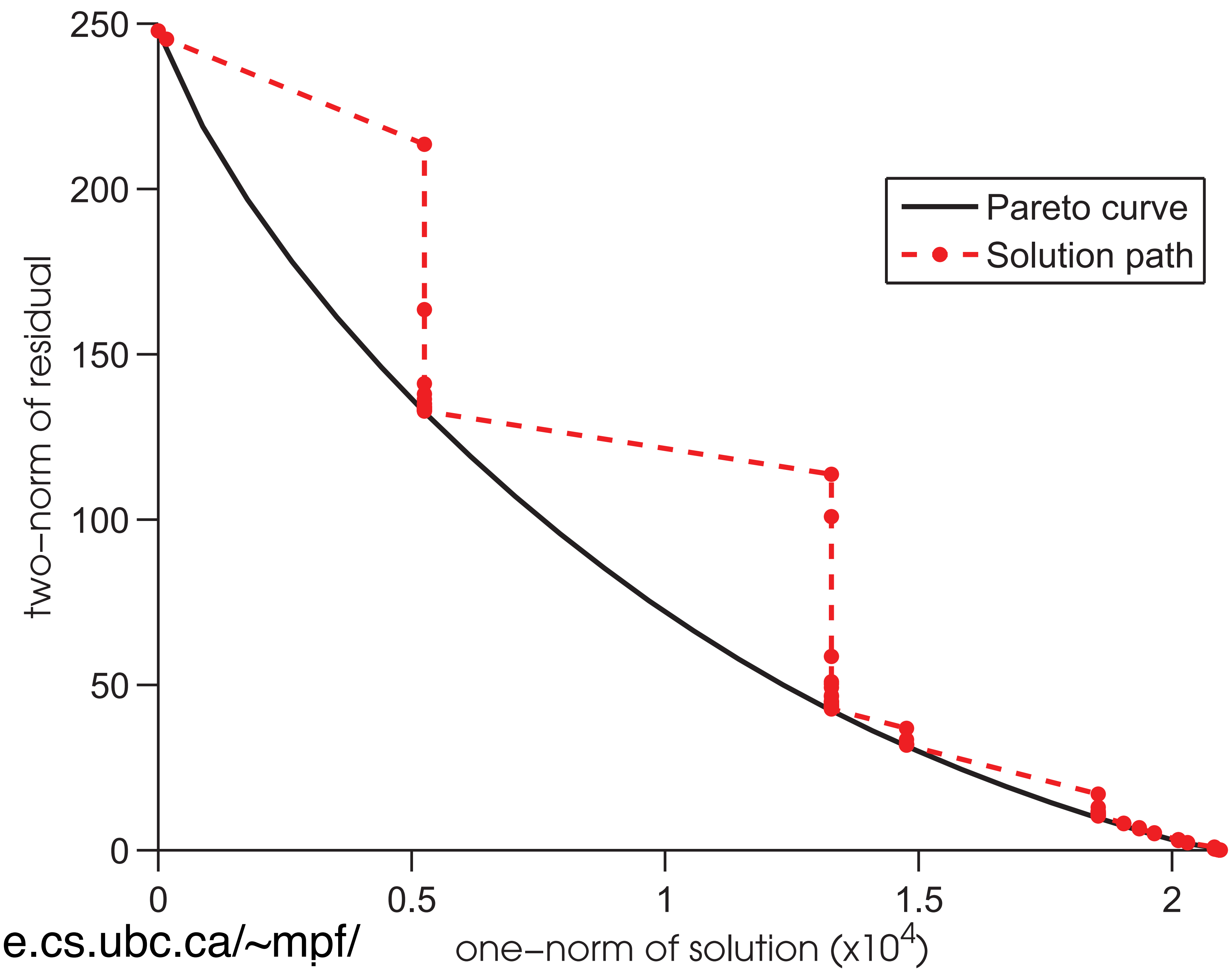
"When a traveler reaches a fork in the road, the l_1-norm tells him to take either one way or the other, but the l_2 -norm instructs him to head off into the bushes."

John F. Claerbout and Francis Muir, 1973

[van den Berg & Friedlander, '08]
[Hennenfent, FJH, et. al, '08]

SLIM

One-norm solver



from http://people.cs.ubc.ca/~mpf/

Key elements

SLIM

☑ *sparsifying transform*

- typically **localized** in the time-space domain to handle the complexity of seismic data
- **curvelets**

☑ *advantageous coarse sampling (mixing)*

- generates incoherent random undersampling "noise" in the sparsifying domain
- does not create large gaps for measurement in the physical domain
- does not create coherent interferences in simultaneous acquisition

☑ *sparsity-promoting solver*

- requires few matrix-vector multiplications

Observations

Controllable error for reconstruction from *randomized* subsamplings

Curvelets and *simultaneous* acquisition perform the best

Oversampling compared to *conventional compression* is small

Combination of *sampling* & *encoding* into a *single* ***linear*** step has profound implications

- *acquisition* costs ***no*** longer determined by *resolution* & *size*
- *but* by *transform-domain sparsity* & *recovery error*

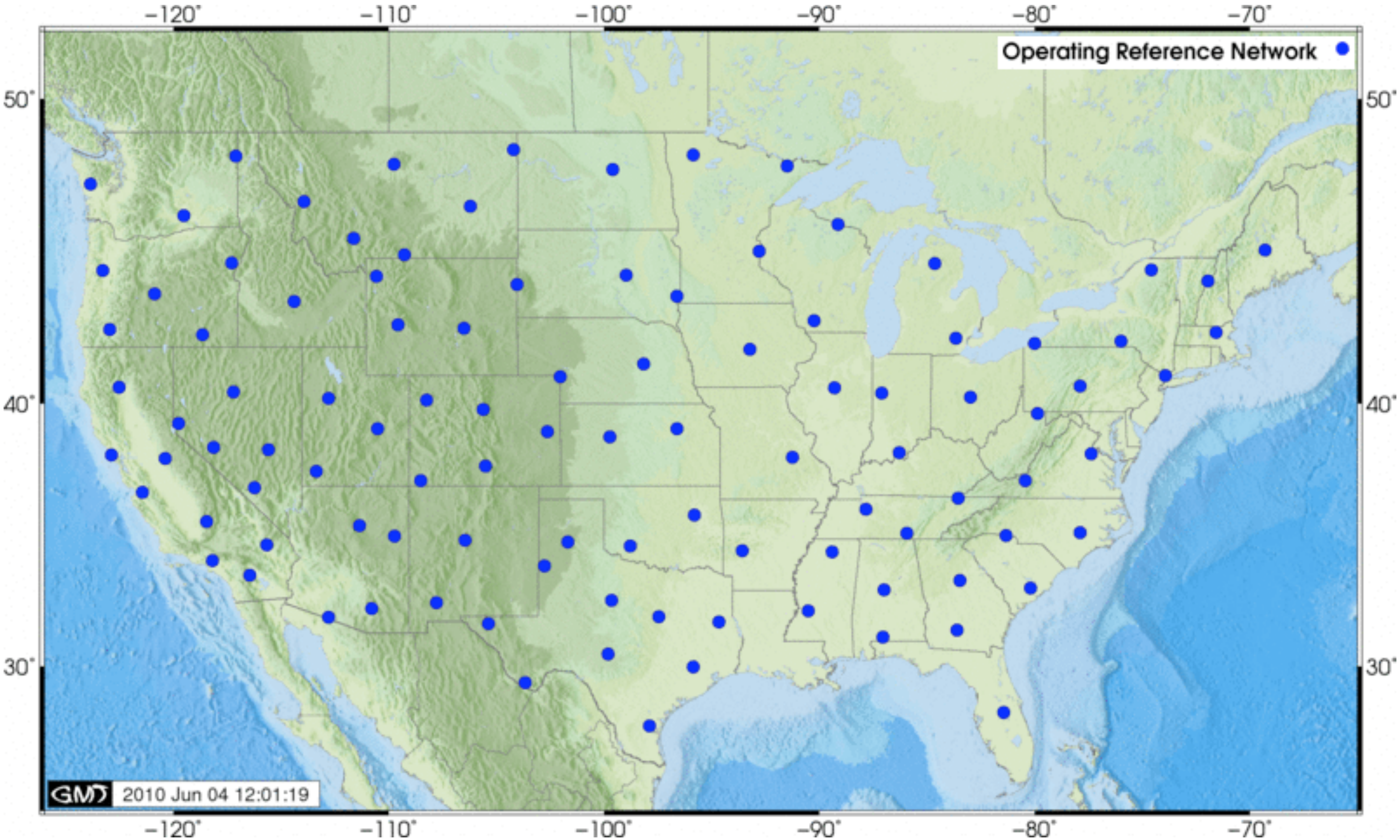
Implications

Periodic sampling is *detrimental* to *sparse* recovery

“Random nature” of receiver functions is highly favorable

- know the source-time function
- deal with surface-related multiples & surface waves

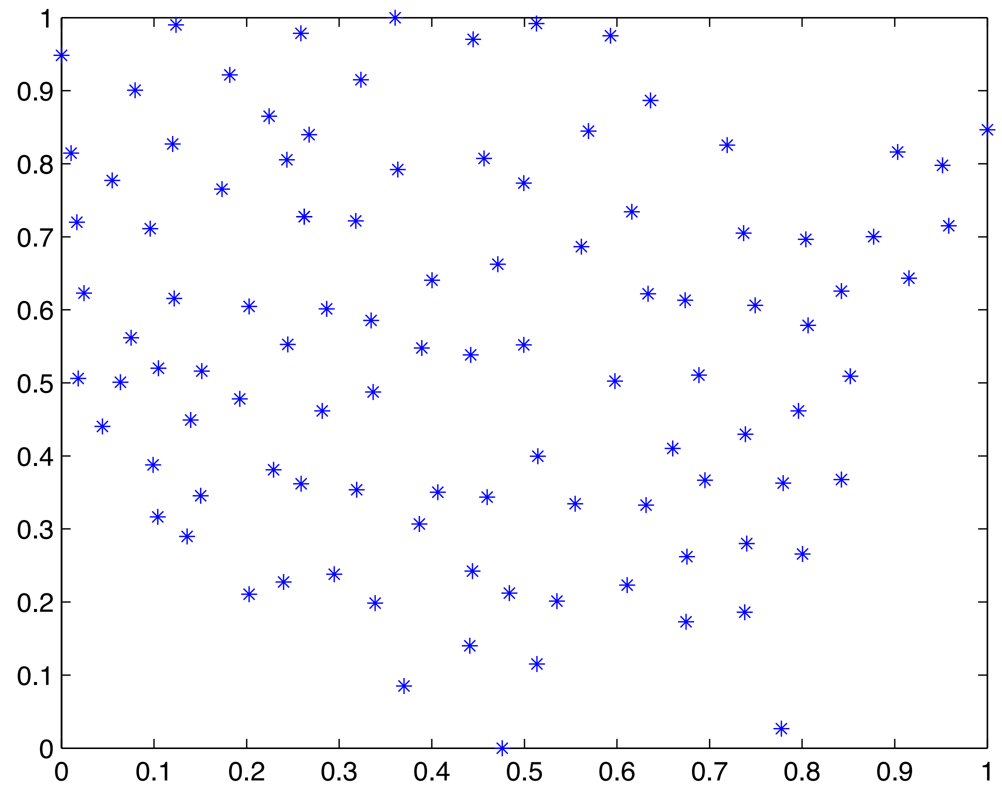
US array



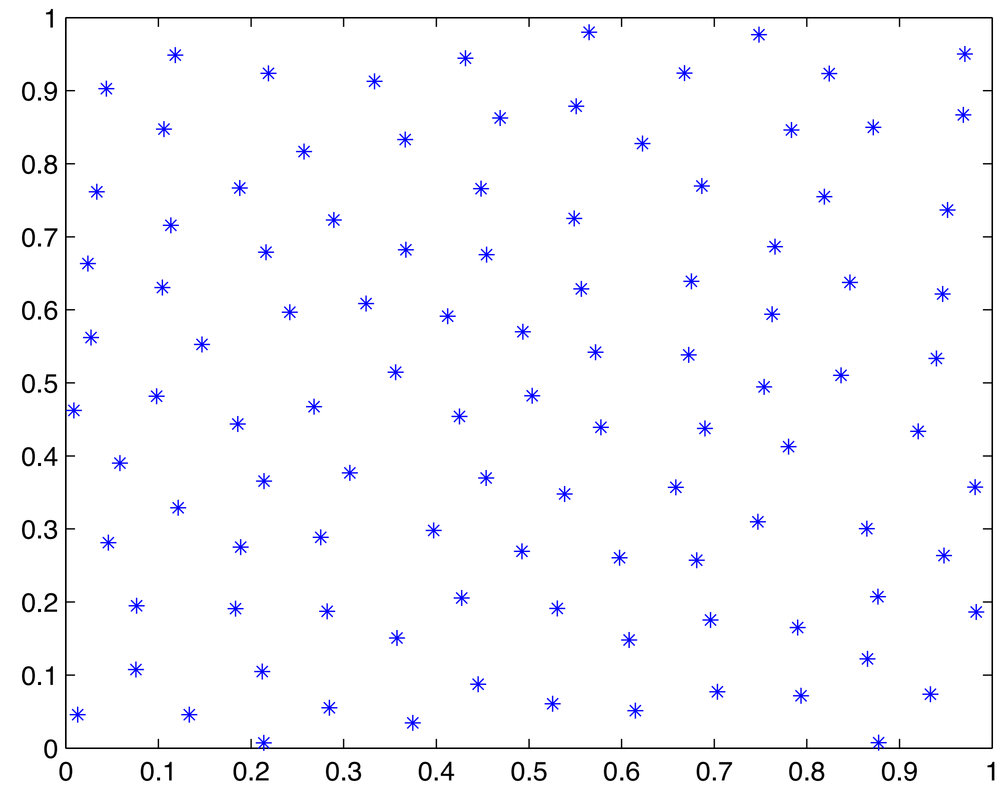
−120˚
−110˚
−100˚
−90˚
−80˚
−70˚
50˚
40˚
30˚
Operating Reference Network
GMT
2010 Jun 04 12:01:19

Sample points

US Array

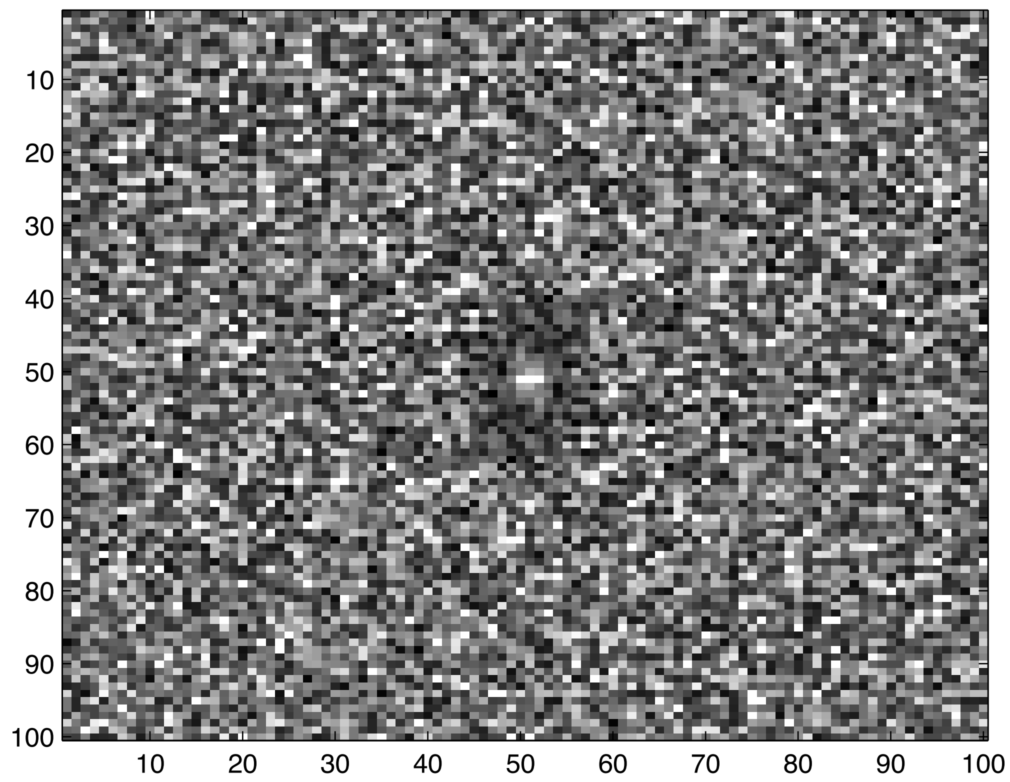


Poisson disk

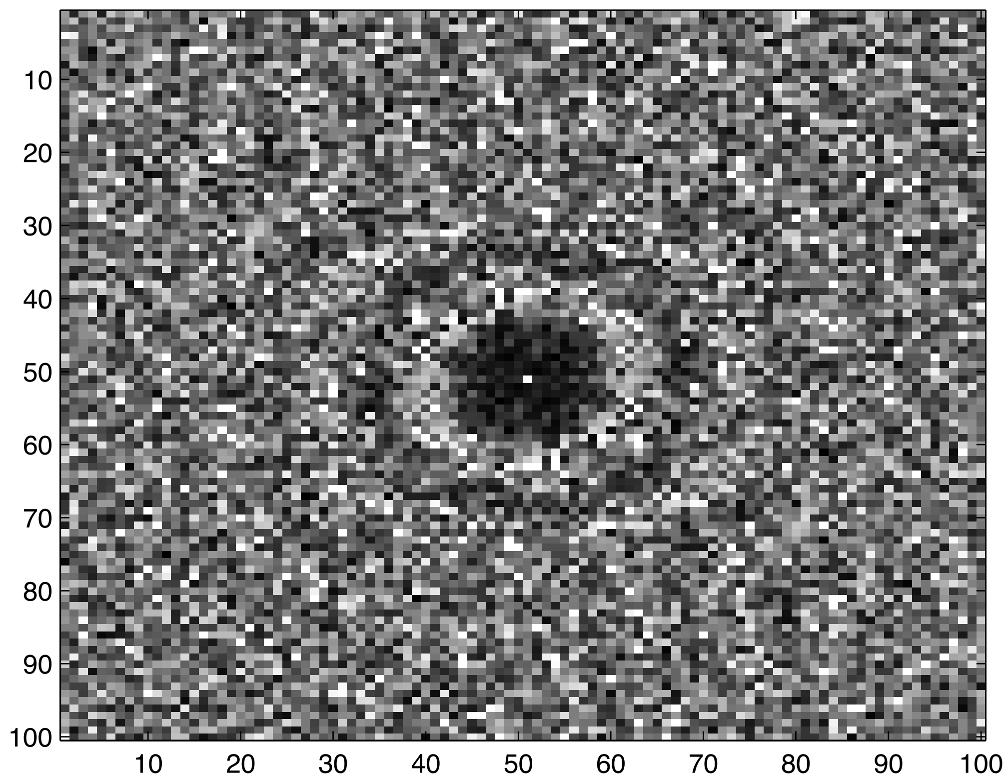


Spectra

US Array

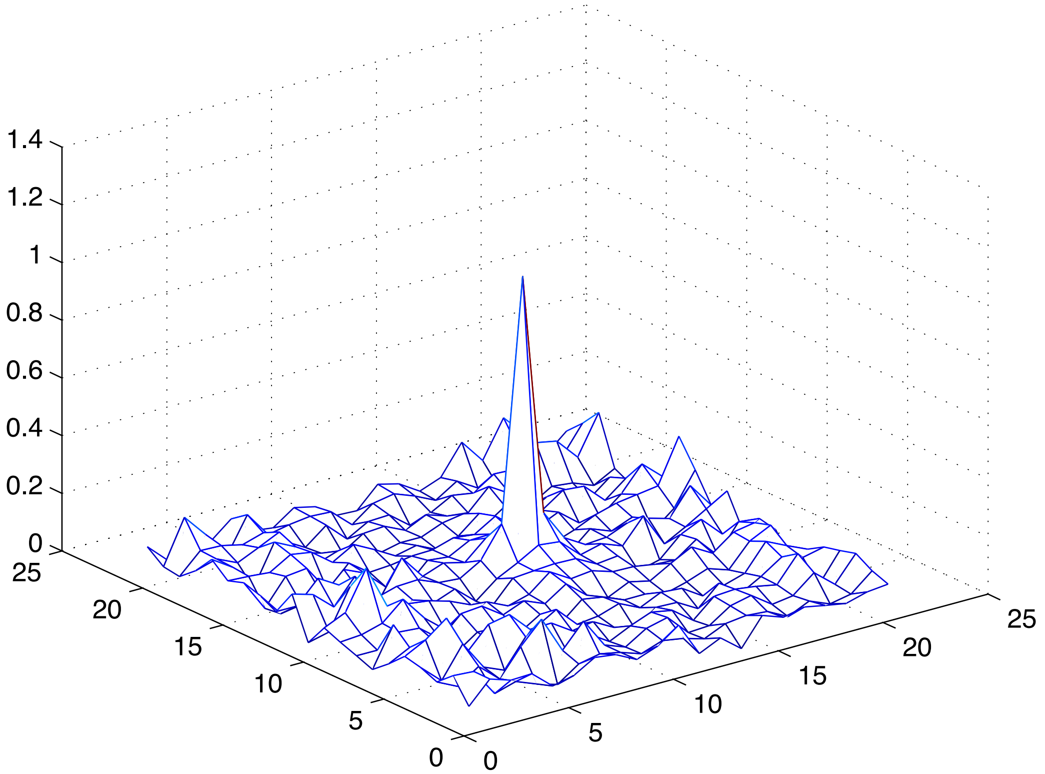


Poisson disk

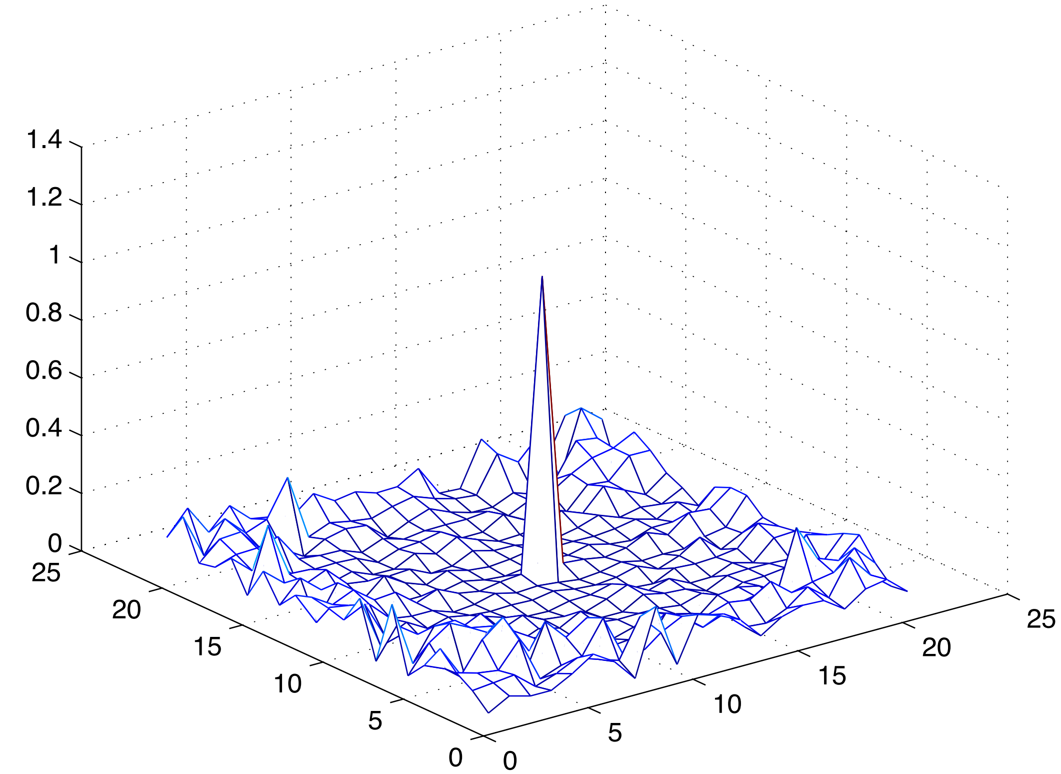


Spectra

US Array



Poisson disk



Extensions

Include more "physics" in the formulation

- via *discretization* of *integral* equations of the *second* kind
- *prediction* of *surface-related* multiples

Incorporate *dimensionality reductions* in *full-waveform* inversion

- via creation of *supershots*
- *stochastic gradients* as part of *stochastic optimization*

EPSI L1 formulation

Use L1-norm relaxation for the sparsity objective

$$\underset{\mathbf{Q},\,\mathbf{X_o}}{\textbf{minimize}} \quad ||\mathbf{X_o}||_1 \quad \textbf{s.t.} \quad ||\mathbf{P}^- - \mathbf{X_o}(\mathbf{Q} + \mathbf{R}\mathbf{P}^-)||_2^2 \leq \sigma$$

$\mathbf{P}^-$ total up-going wavefield

$\mathbf{Q}$ down-going source signature

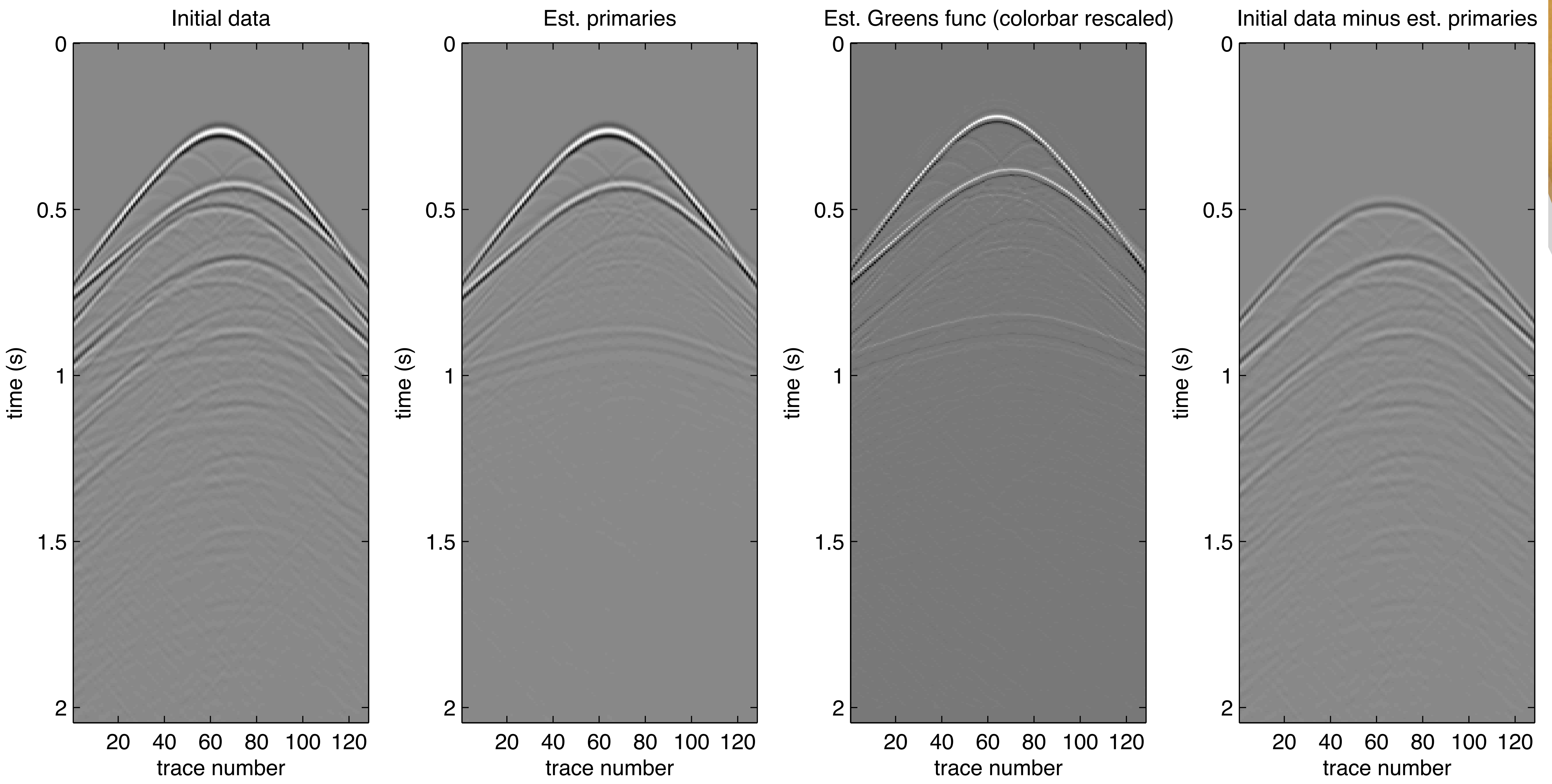
$\mathbf{R}$ reflectivity of free surface (assume -1)

$\mathbf{X_o}$ primary impulse response

(all single-frequency data volume, implicit ω)

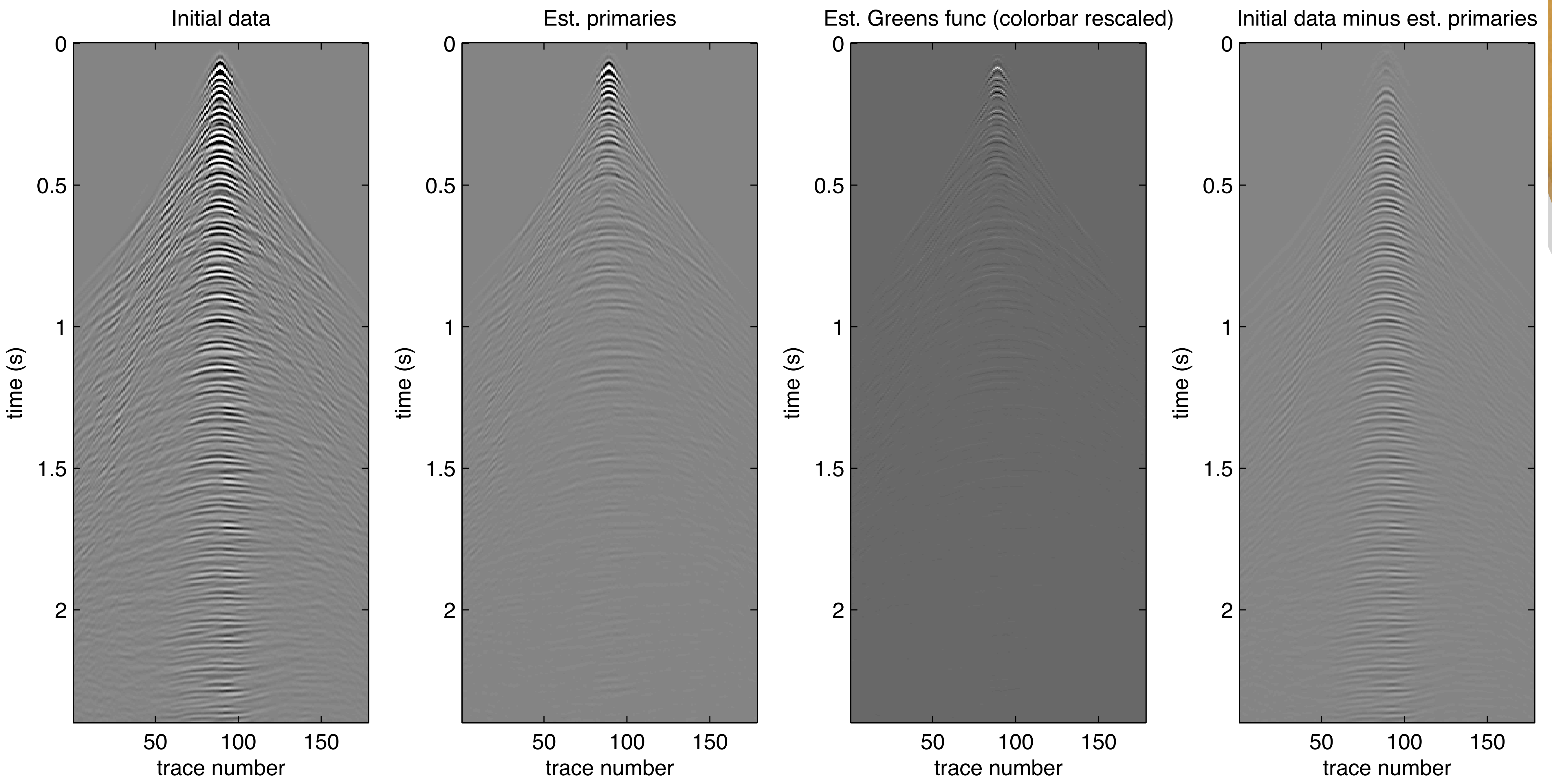
[Groenestijn et. al. '09]
[Lin and Herrmann, '09]

EPSI L1: synthetic

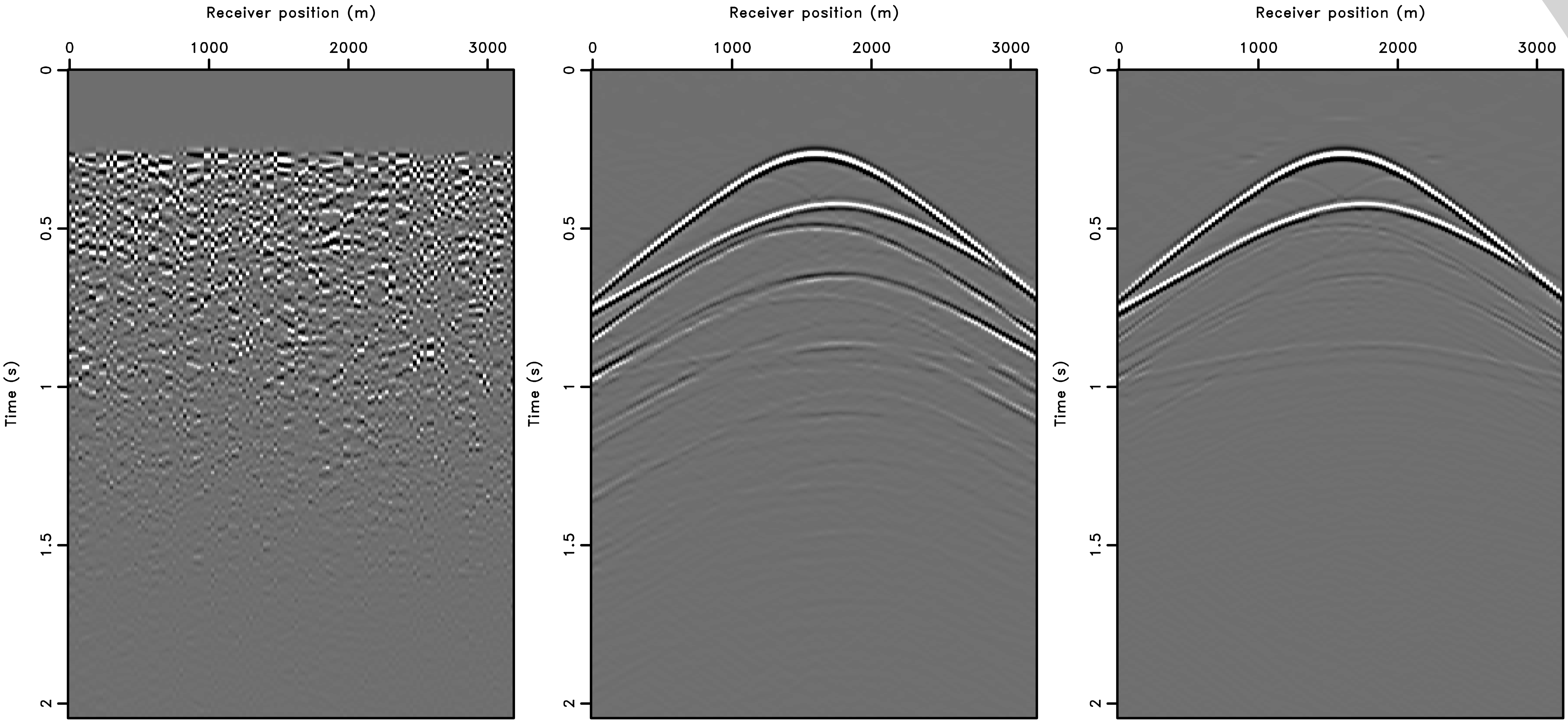


~70 mat-vec, 6 source matching

EPSI L1: real data



multisourcing



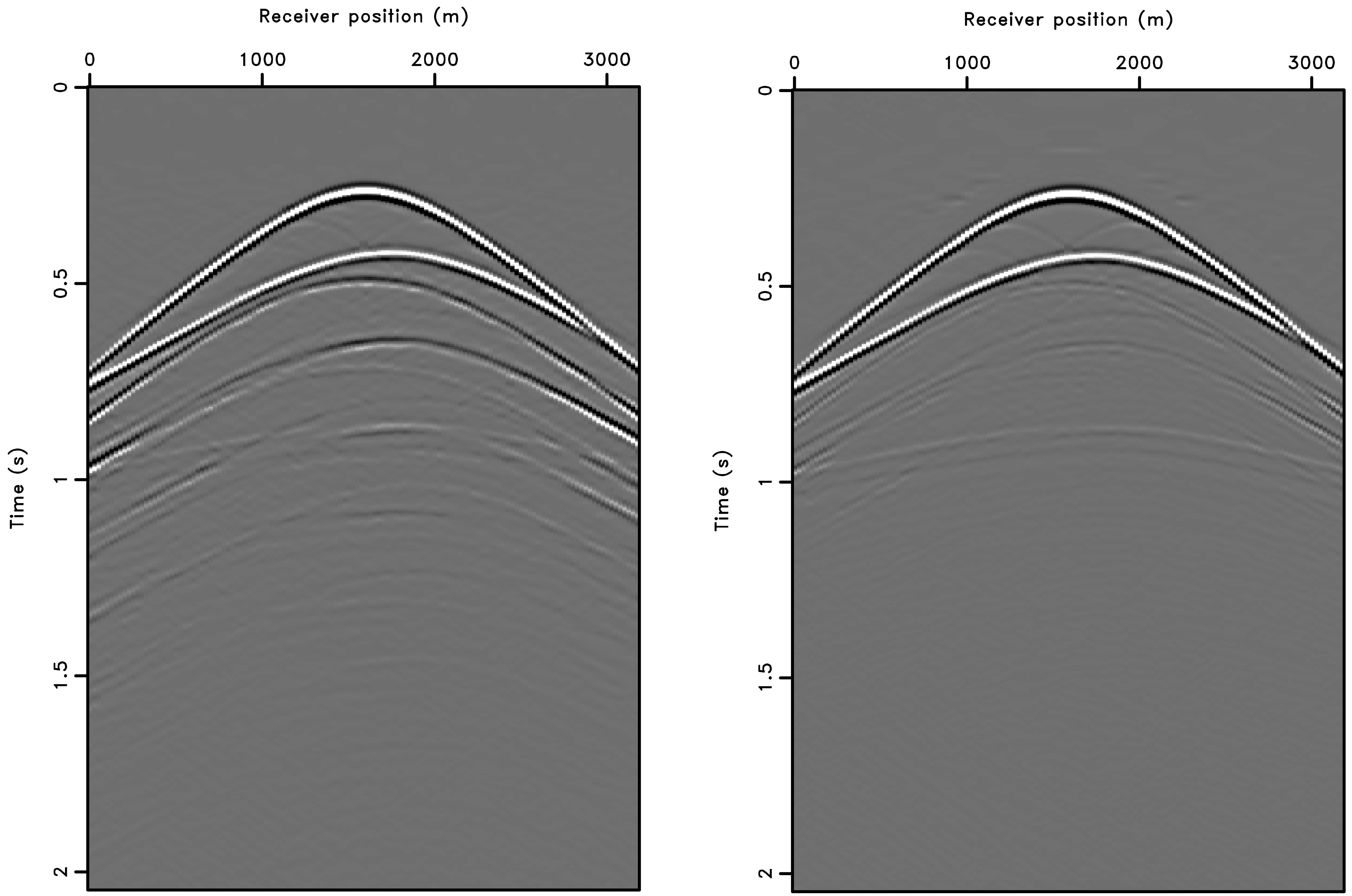
Single simultaneous shot

separate

separate + primary inversion

~100 projected gradient, 5 source matching

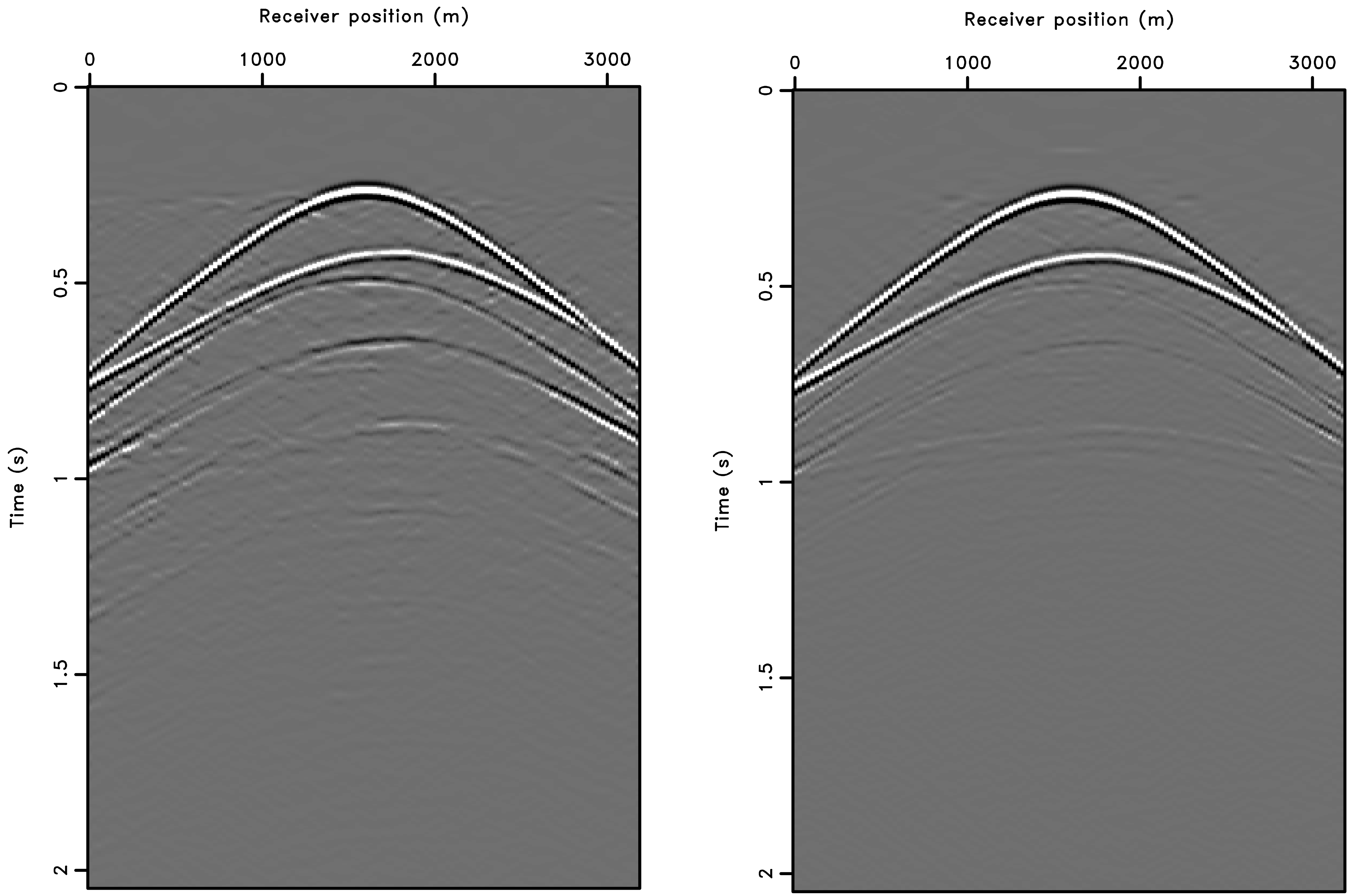
50% measurement



separate — **separate + primary inversion**

SLIM

20% measurement



separate **separate + primary inversion**

Full-waveform inversion

Multiexperiment PDE-constrained optimization problem:

$$\min_{\mathbf{U}\in\mathcal{U},\,\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{P}-\mathbf{D}\mathbf{U}\|_2^2 \quad \text{subject to} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{Q}$$

$$\begin{aligned}
\mathbf{P} &= \text{Total multi-source and multi-frequency data volume}\\
\mathbf{D} &= \text{Detection operator}\\
\mathbf{U} &= \text{Solution of the Helmholtz equation}\\
\mathbf{H} &= \text{Discretized multi-frequency Helmholtz system}\\
\mathbf{Q} &= \text{Unknown seismic sources}\\
\mathbf{m} &= \text{Unknown model, e.g. } c^{-2}(x)
\end{aligned}$$

Unconstrained problem

For each ***separate*** source **q** solve the **unconstrained problem:**

$$\min_{\mathbf{m}\in\mathcal{M}} \frac{1}{2}\|\mathbf{p}-\mathcal{F}[\mathbf{m},\mathbf{q}]\|_2^2$$

with

$$\mathcal{F}[\mathbf{m},\mathbf{q}] = \mathbf{D}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{q}$$

and **q** a single source function

Multiexperiment

Gradient updates:

$$\mathbf{m}^{k+1} := \mathbf{m}^k - \eta_k \nabla J(\mathbf{m}^k, \mathbf{Q})$$

with

$$J(\mathbf{m}^k, \mathbf{Q}) := \|\mathbf{P} - \boldsymbol{\mathcal{F}}[\mathbf{m}^k, \mathbf{Q}]\|_{2,2}^2$$

and

$$\boldsymbol{\mathcal{F}}[\mathbf{m}, \mathbf{Q}] = \mathbf{D}\mathbf{H}^{-1}[\mathbf{m}]\mathbf{Q}$$

Dimensionality reduction

$$
\begin{bmatrix}
\mathcal{H}_{\omega_1} & 0 & & \\
0 & \mathcal{H}_{\omega_2} & \ddots & \\
 & \ddots & \ddots & 0 \\
 & & 0 & \mathcal{H}_{\omega_{n_f}}
\end{bmatrix}
\begin{bmatrix}
\overbrace{[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_{n_s}]_{\omega_1}}^{\mathbf{U}_{\omega_1}} \\
\vdots \\
\vdots \\
\underbrace{[\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_{n_s}]_{\omega_{n_f}}}_{\mathbf{U}_{n_f}}
\end{bmatrix}
=
\begin{bmatrix}
\overbrace{[\mathbf{q}_1\ \mathbf{q}_2\ \cdots\ \mathbf{q}_{n_s}]_{\omega_1}}^{\mathbf{Q}_{\omega_1}} \\
\vdots \\
\vdots \\
\underbrace{[\mathbf{q}_1\ \mathbf{q}_2\ \cdots\ \mathbf{q}_{n_s}]_{\omega_{n_f}}}_{\mathbf{Q}_{n_f}}
\end{bmatrix}
$$

Explicit solves 3D models (>1000^3) extremely challenging

Use *preconditioned indirect* Krylov solvers and reduce # right-hand sides and blockdiagonals

$$
\mathcal{F}[\mathbf{m},\mathbf{Q}] \mapsto \mathcal{F}[\mathbf{m},\underline{\mathbf{Q}}] \quad \text{with} \quad \underline{\mathbf{Q}} = \mathbf{RMQ}
$$

Removes the main disadvantage of *indirect* methods.

[Erlanga, Nabben, ’08, Erlanga and F.J.H, ’08, FJH et. al., ’09-’10]

SLIM

Stochastic optimization

Stochastic "batch" gradient decent:

$$\mathbf{m}^{k+1} := \mathbf{m}^{k} - \eta_k \nabla \sum_{i=1}^{n} J(\mathbf{m}^{k}, \underline{\mathbf{q}}^{i}) \quad \text{with} \quad \underline{\mathbf{q}}^{i} := (\mathbf{RM})_i \, \mathbf{Q}$$

- for $n \rightarrow \infty$, the updates become *deterministic*
- *prohibitively* expensive

[Bersekas, '96]

Stochastic optimization

Stochastic "online" gradient descent:

$$\mathbf{m}^{k+1} := \mathbf{m}^{k} - \eta_k \nabla J(\mathbf{m}^{k}, \underline{\mathbf{Q}}^{k}) \quad \text{with} \quad \underline{\mathbf{Q}}^{k} := (\mathbf{RM})_k \, \mathbf{Q}$$

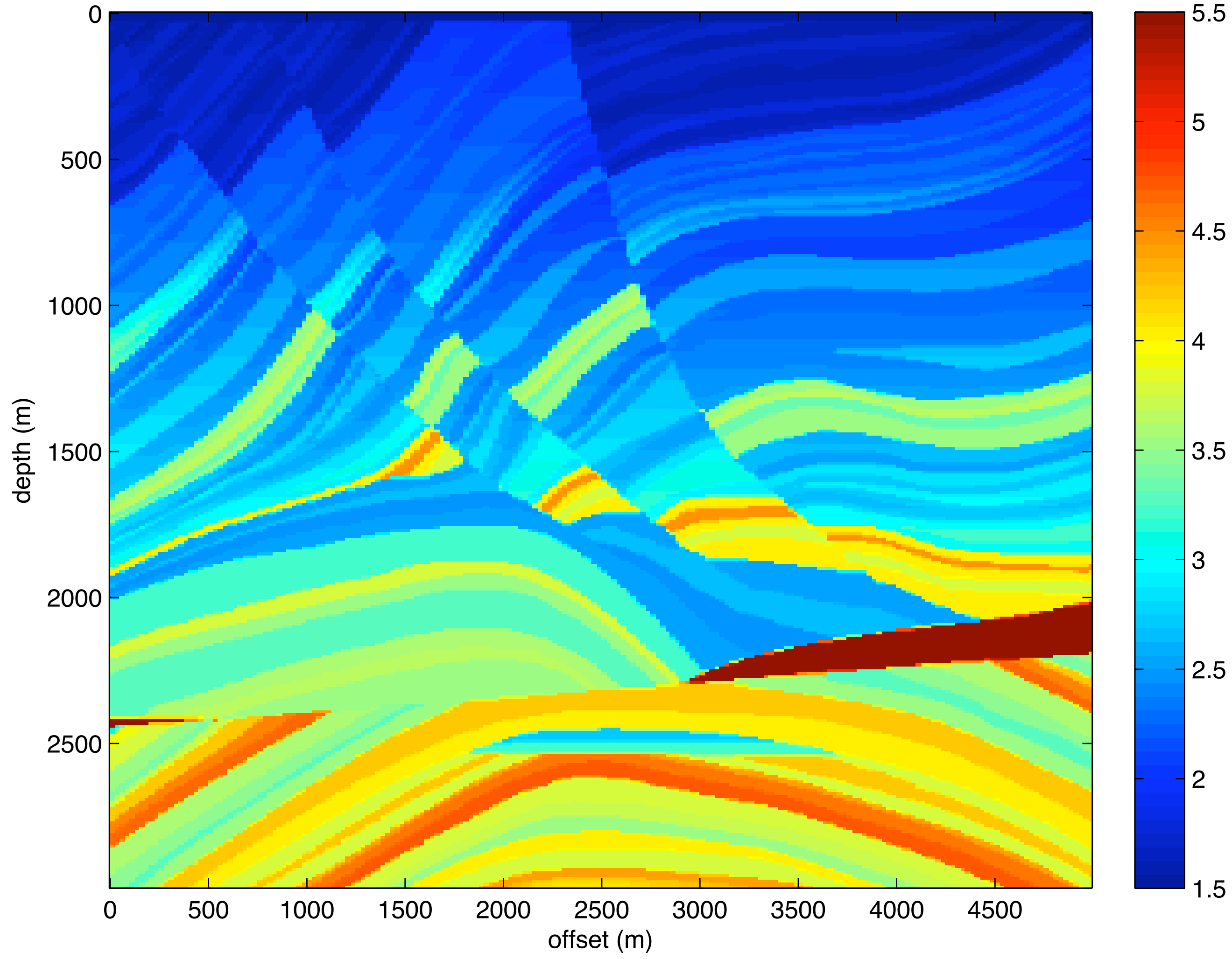
- uses *different* random **RM** for each *iteration*
- involves a *single* $\underline{\mathbf{Q}}^k$ for *each* gradient update
- costs depend on # RHS and freq. blocks
- cheap but introduces "noise"... Sounds familiar?

[Krebs et.al, '09]

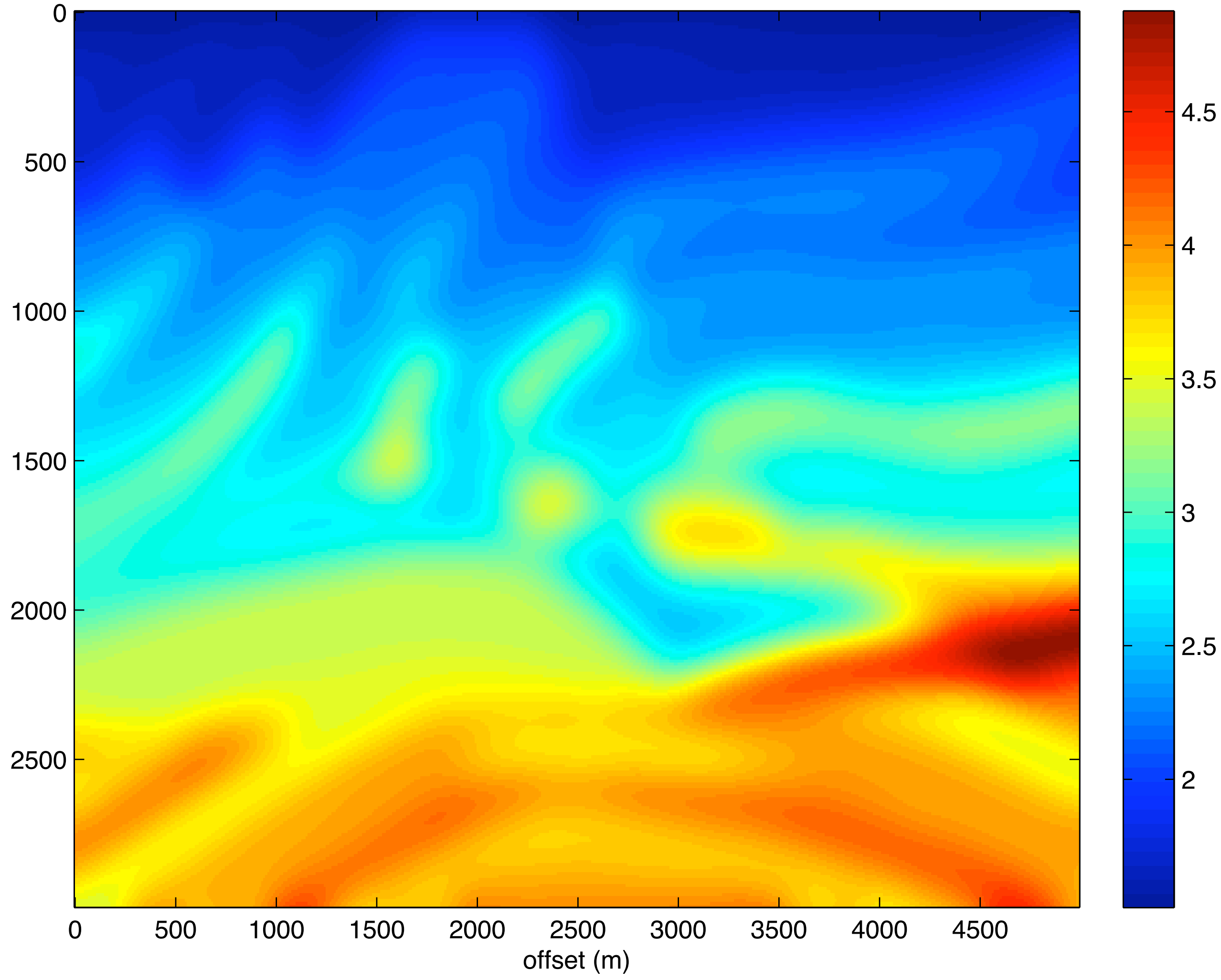
[Bersekas, '96]

Marmoussi model

original model

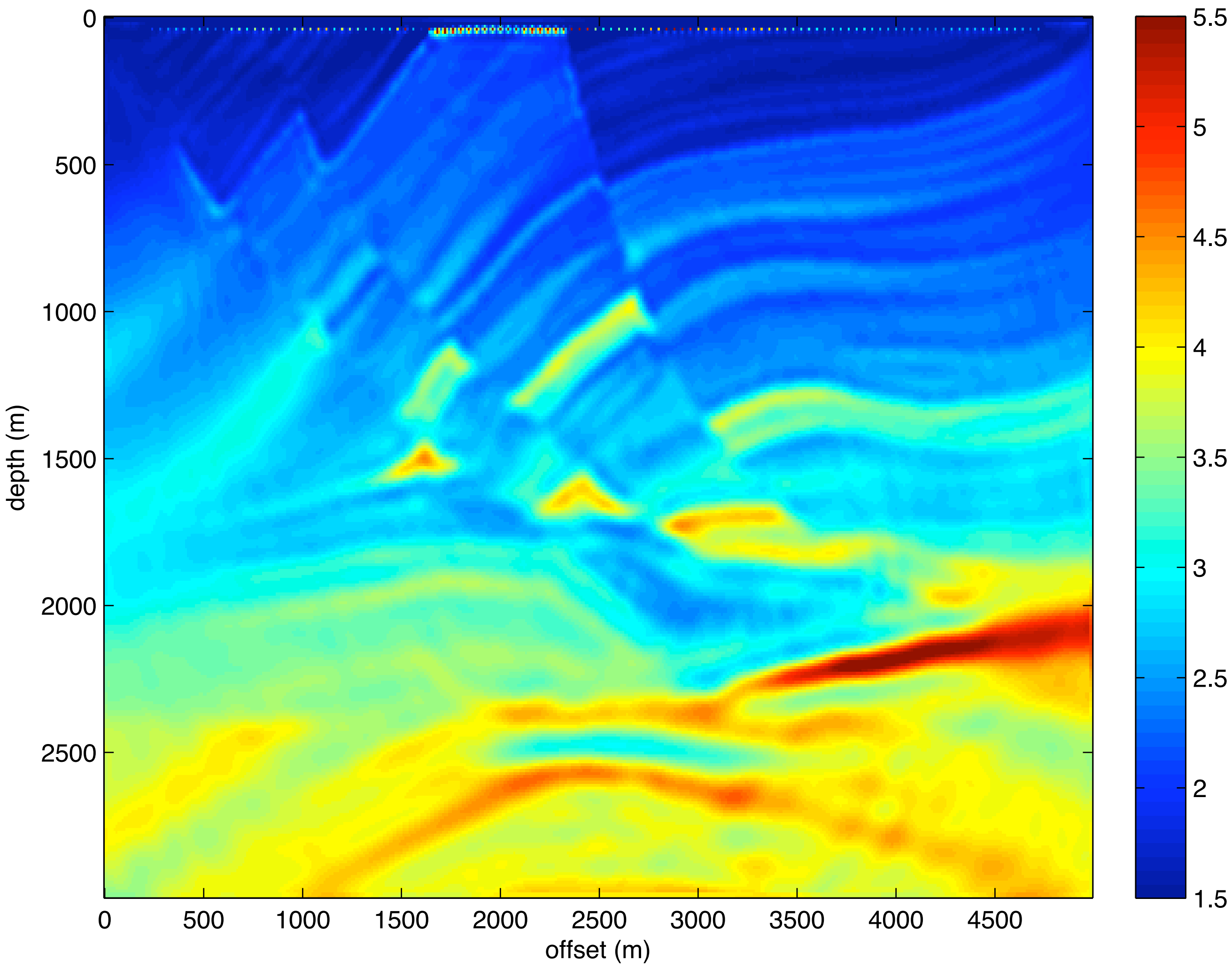


initial model

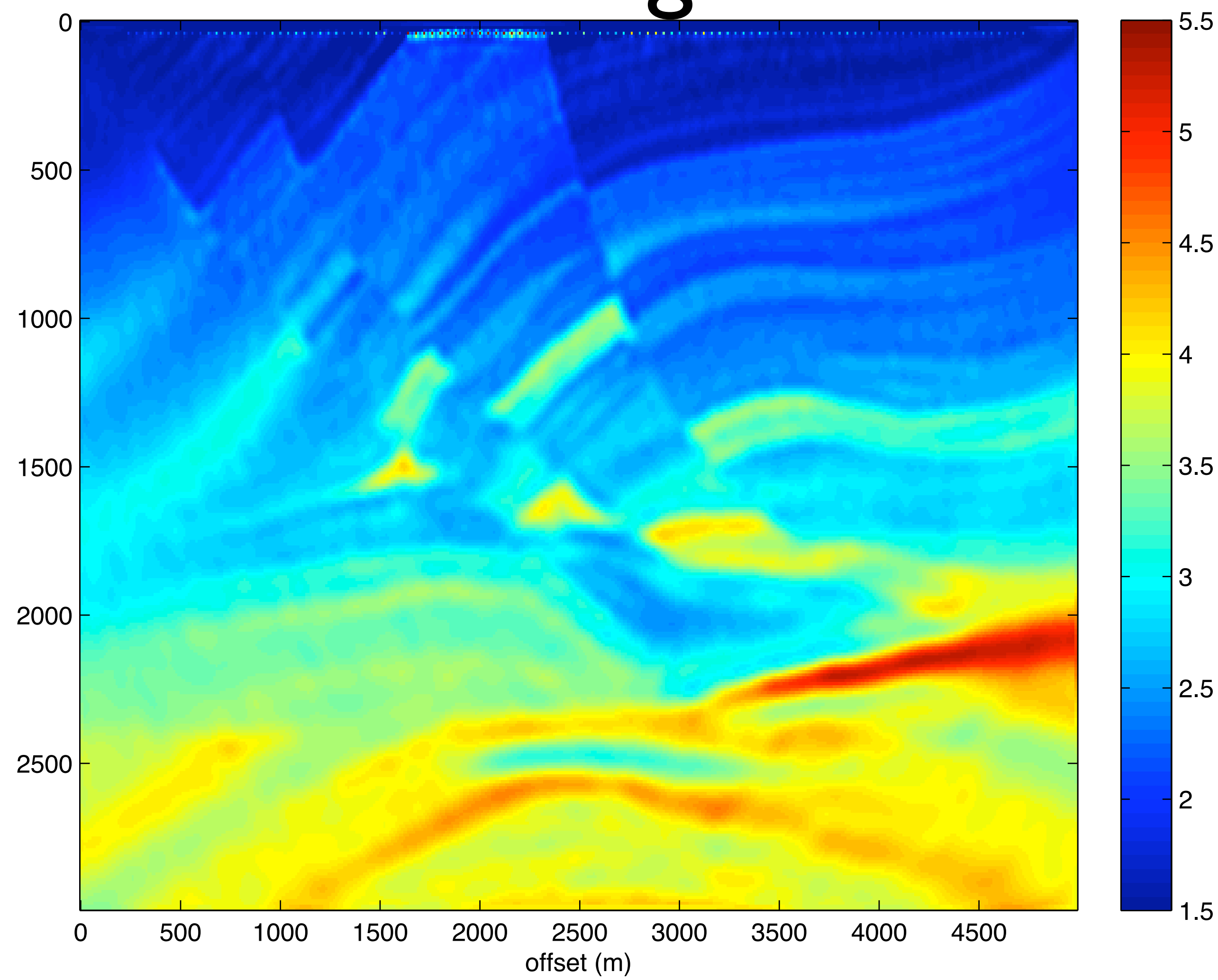


Full-waveform inversion

recovered model
l-BFGS



recovered
stochastic gradient



SLIM

Speed up

Full scenario:

- 113 sequential shots with 50 frequencies
- 18 iterations of l-BFGS (90=5*18 Helmholtz solves)

Reduced scenario:

- 16 randomized simultaneous shots with 4 frequencies
- 40 iterations of SA (2.27=16*4/(113*50)*40*5 solves)

Speed up of 40 X or > week vs 8 h on 32 CPUs

Conclusions

Dimensionality reduction will *revolutionize* our field

- *reduction* of *acquisition* costs
- *less reliance* on *full sampling*
- *decrease* in *processing* time
- *high-resolution* inversions that are otherwise *infeasible* with fully-sample (Nyquist-based) methods

It is only going to work ...

... if *all* components are in place

- Applied & Computational Harmonic analysis / Compressive Sensing
- Convex & PDE-constrained optimization
- Numerical Linear Algebra
- Stochastic optimization & machine learning

This combination will lead to the breakthroughs we need...

Acknowledgments

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE (334810-05).

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

Thank you

slim.eos.ubc.ca

Further reading

Compressive sensing

- *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information by Candes, 06.*
- *Compressed Sensing* by D. Donoho, '06

Simultaneous acquisition

- *A new look at simultaneous sources* by Beasley et. al., '98.
- *Changing the mindset in seismic data acquisition* by Berkhout '08.

Transform-based seismic data regularization

- *Interpolation and extrapolation using a high-resolution discrete Fourier transform* by Sacchi et. al, '98
- *Non-parametric seismic data recovery with curvelet frames* by FJH and Hennenfent.,'07
- *Simply denoise: wavefield reconstruction via jittered undersampling* by Hennenfent and FJH, '08

Estimation of surface-free Green's functions:

- *Estimating primaries by sparse inversion and application to near-offset data reconstruction* by Groenestijn, '09
- *Unified compressive sensing framework for simultaneous acquisition with primary estimation* by T. Lin & FJH, '09

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.
- *Fast full-wavefield seismic inversion using encoded sources* by Krebs et. al., '09
- *Randomized dimensionality reduction for full-waveform inversion* by FJH & X. Li, '10