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

We are *no* longer finding oil...

Push for improved seismic inversion

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

Impediments

Costs of *acquisition* to meet *raised* demands for *full-waveform inversion*...

SLIM

Impediments

Turn-around times to arrive at *final* product...

Impediments

Moore's law is coming to an end...

Impediments

So, we can no longer *compute* ourselves *out* of this...

Impediments

Size of our *discretizations* is *dictated* by a *far* too *pessimistic Nyquist-sampling criterion*...

Wish list

Acquisition & processing costs determined by

- *complexity* of the subsurface
- *controllable* error

Paradigm shift

We are at the cusp of *fundamental* breakthroughs

- *Compressive Sensing* in mathematics
- *Incoherent acquisition* in seismic acquisition & processing practices ...

Aimed at *dimension reduction!*

Johnson–Lindenstrauss lemma

From Wikipedia, the free encyclopedia

In mathematics, the **Johnson–Lindenstrauss lemma** is a result concerning low-distortion embeddings of points from high-dimensional into low-dimensional Euclidean space. The lemma states that a small set of points in a high-dimensional space can be embedded into a space of much lower dimension in such a way that distances between the points are nearly preserved. The map used for the embedding is at least Lipschitz, and can even be taken to be an orthogonal projection.

The lemma has uses in compressed sensing, manifold learning, dimensionality reduction, and graph embedding. Much of the data stored and manipulated on computers, including text and images, can be represented as points in a high-dimensional space. However, the essential algorithms for working with such data tend to become bogged down very quickly as dimension increases. It is therefore desirable to reduce the dimensionality of the data in a way that preserves its relevant structure. The Johnson–Lindenstrauss lemma is a classic result in this vein.

Also the lemma is tight up to a factor $\log(1/\varepsilon)$, i.e. there exists a set of points of size m that needs dimension

$$\Omega\left(\frac{\log(m)}{\varepsilon^2\log(1/\varepsilon)}\right)$$

in order to preserve the distances between all pair of points. See 4.

Lemma

[edit]

Given $0 < \varepsilon < 1$, a set X of m points in $\mathbf{R}^N$, and a number $n > n_0 = O(\ln(m)/\varepsilon^2)$, there is a Lipschitz function $f : \mathbf{R}^N \to \mathbf{R}^n$ such that

$$(1-\varepsilon)\|u-v\|_2 \leq \|f(u)-f(v)\|_2 \leq (1+\varepsilon)\|u-v\|_2$$

for all $u, v \in X$.

One proof of the lemma takes f to be a suitable multiple of the orthogonal projection onto a random subspace of dimension n in $\mathbf{R}^N$, and exploits the phenomenon of concentration of measure.

References

[edit]

- W. Johnson and J. Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. Contemporary Mathematics, 26:189–206, 1984.
- S. Dasgupta and A. Gupta, *An elementary proof of the Johnson–Lindenstrauss lemma*, Technical report 99–006, U. C. Berkeley, March 1999.
- D. Achlioptas, *Database-friendly random projections*, In: Proc. 20-th Annual ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, (2001), pp. 274–281.
- R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, *The Johnson–Lindenstrauss Lemma Meets Compressed Sensing*
- N. Alon, *Problems and results in extremal combinatorics*, I, Discrete Math. 273 (2003), 31–53.

Categories: Lemmas

Combined strategy

Linear *dimension reduction*

- e.g., by ***incoherent*** *randomized simultaneous* acquisition with *source encoding*

Nonlinear *recovery*

- e.g., by curvelet-domain ***sparsity*** *promotion* via *one-norm* minimization

Strategy cont'd

Adapt compressive sensing (CS)

- *randomized* subsampling - turns *aliases/ interference* into *noise*
- *sparsity* promotion - removes subsampling *noise* by exploiting signal *structure*

Case study I

Acquisition design according to CS

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

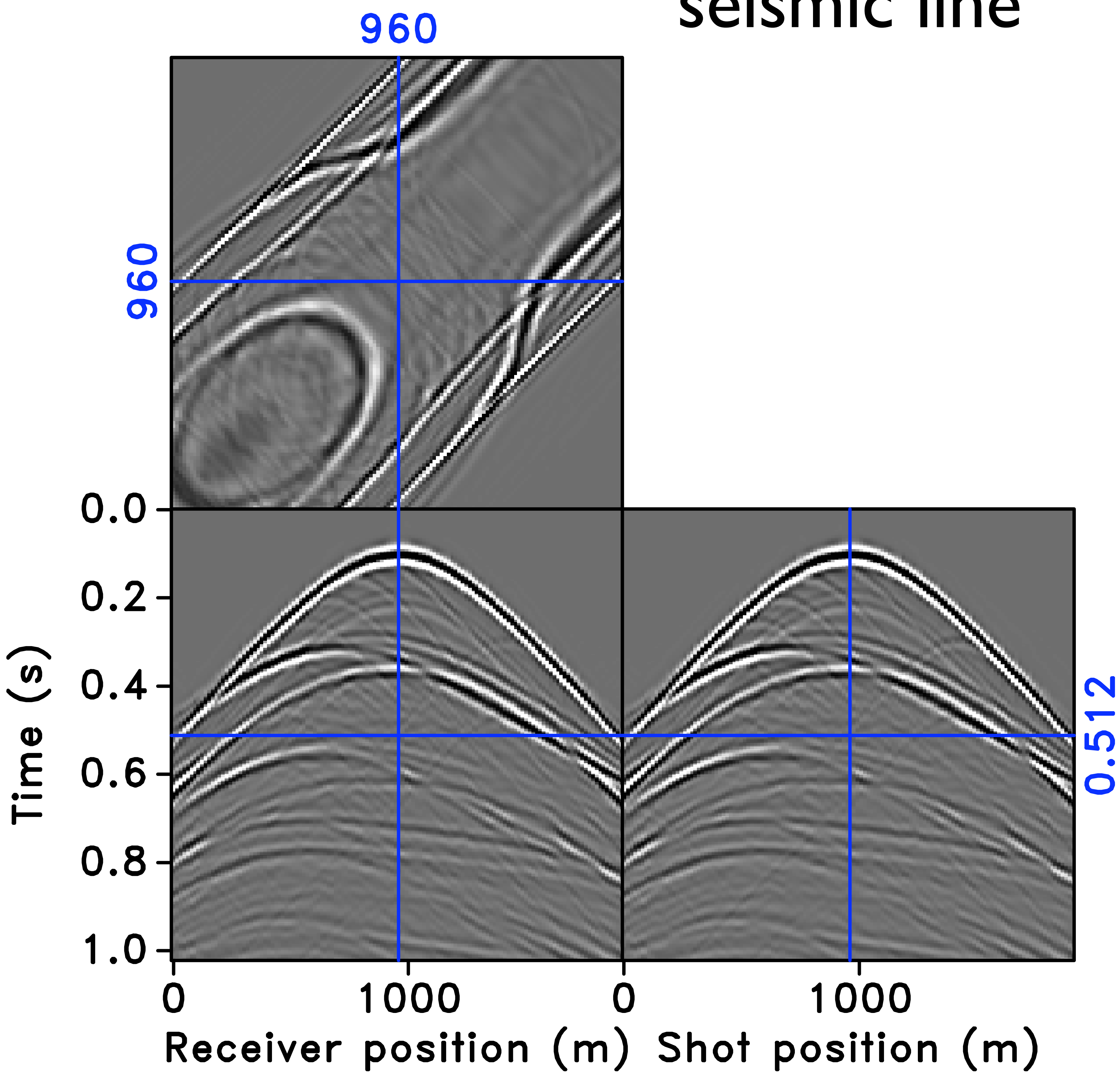
SLIM

pathology

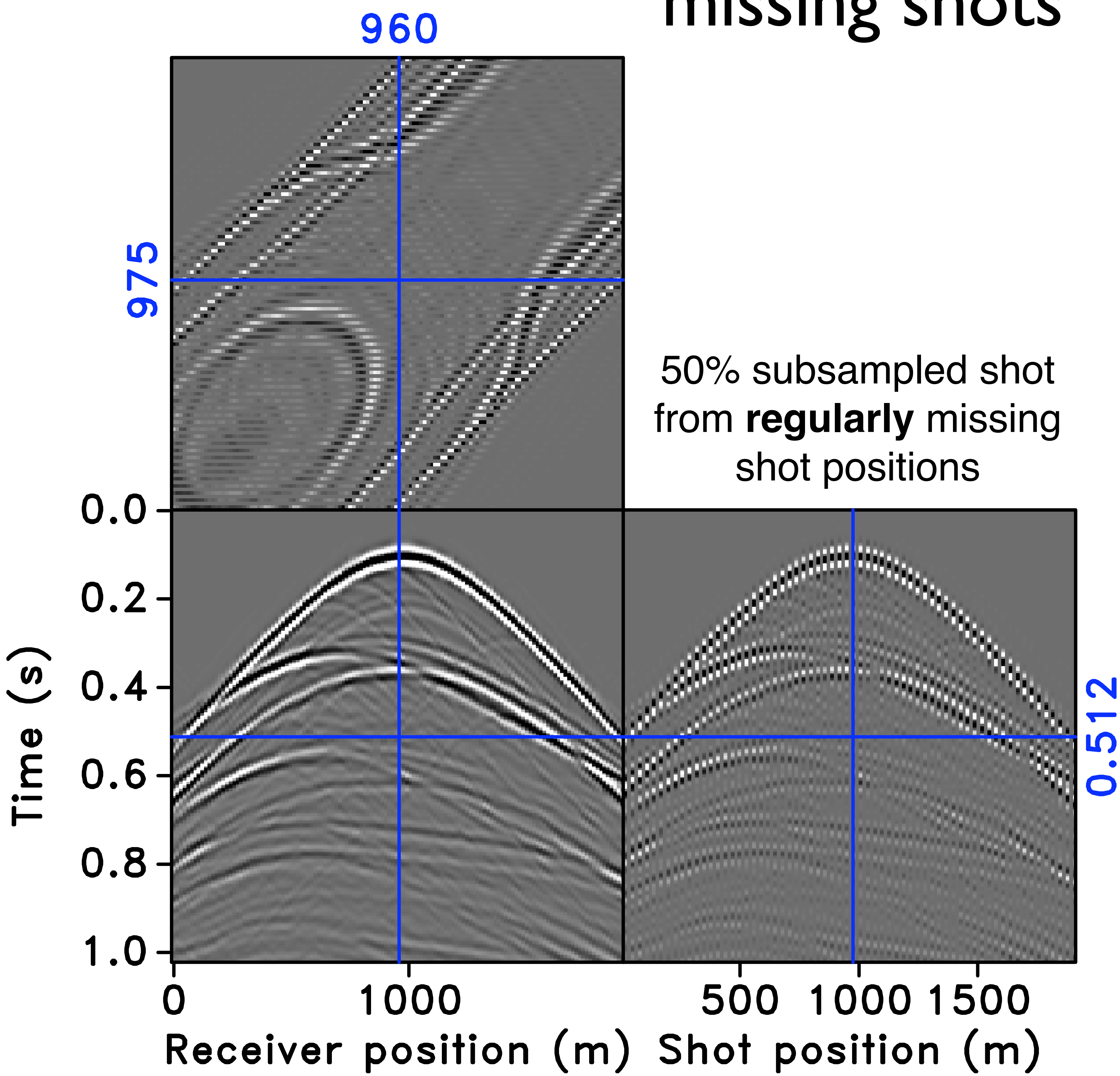
shot interpolation
12.5m to 25m

50 % data-size reduction

seismic line

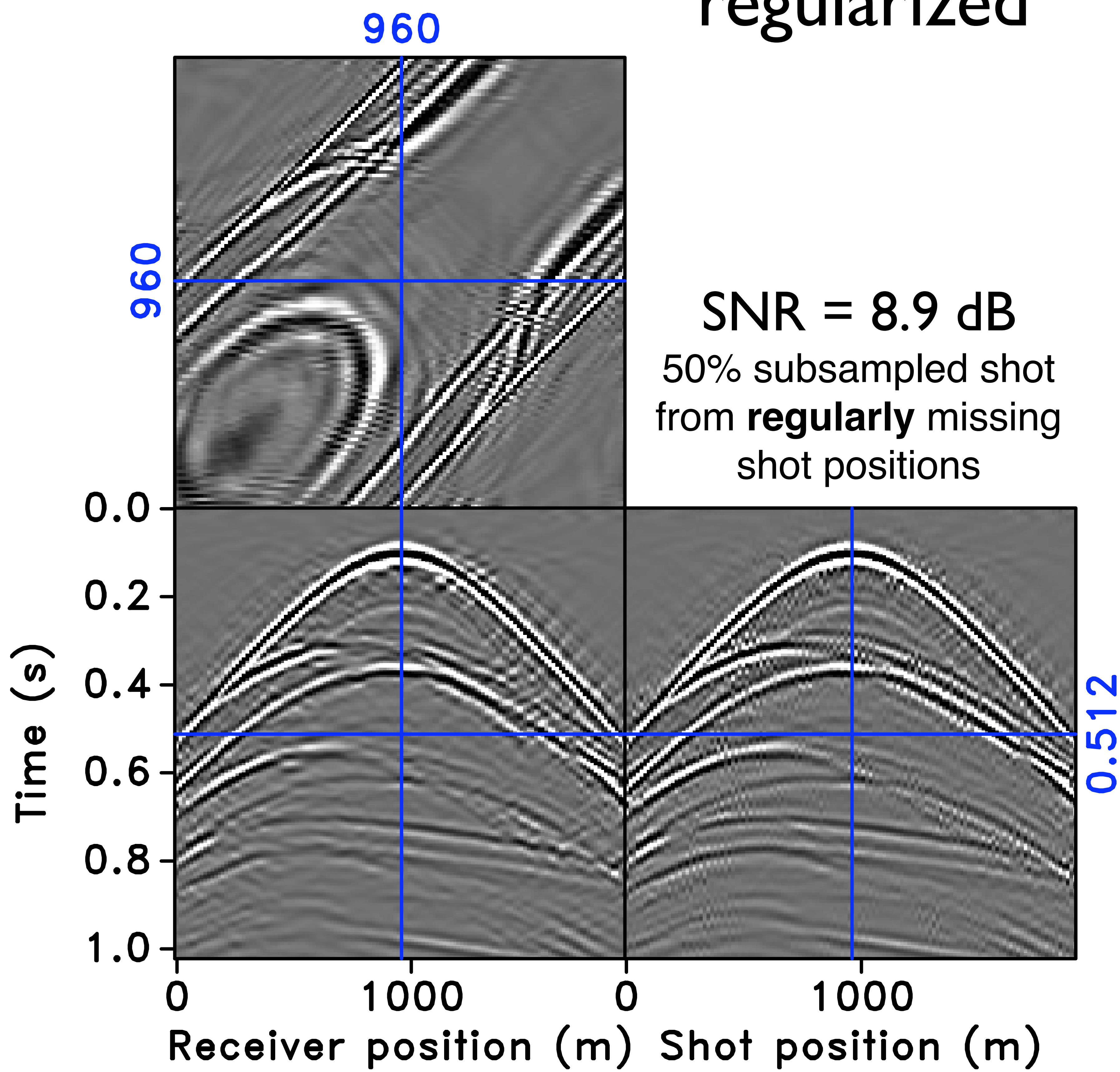


missing shots



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

regularized

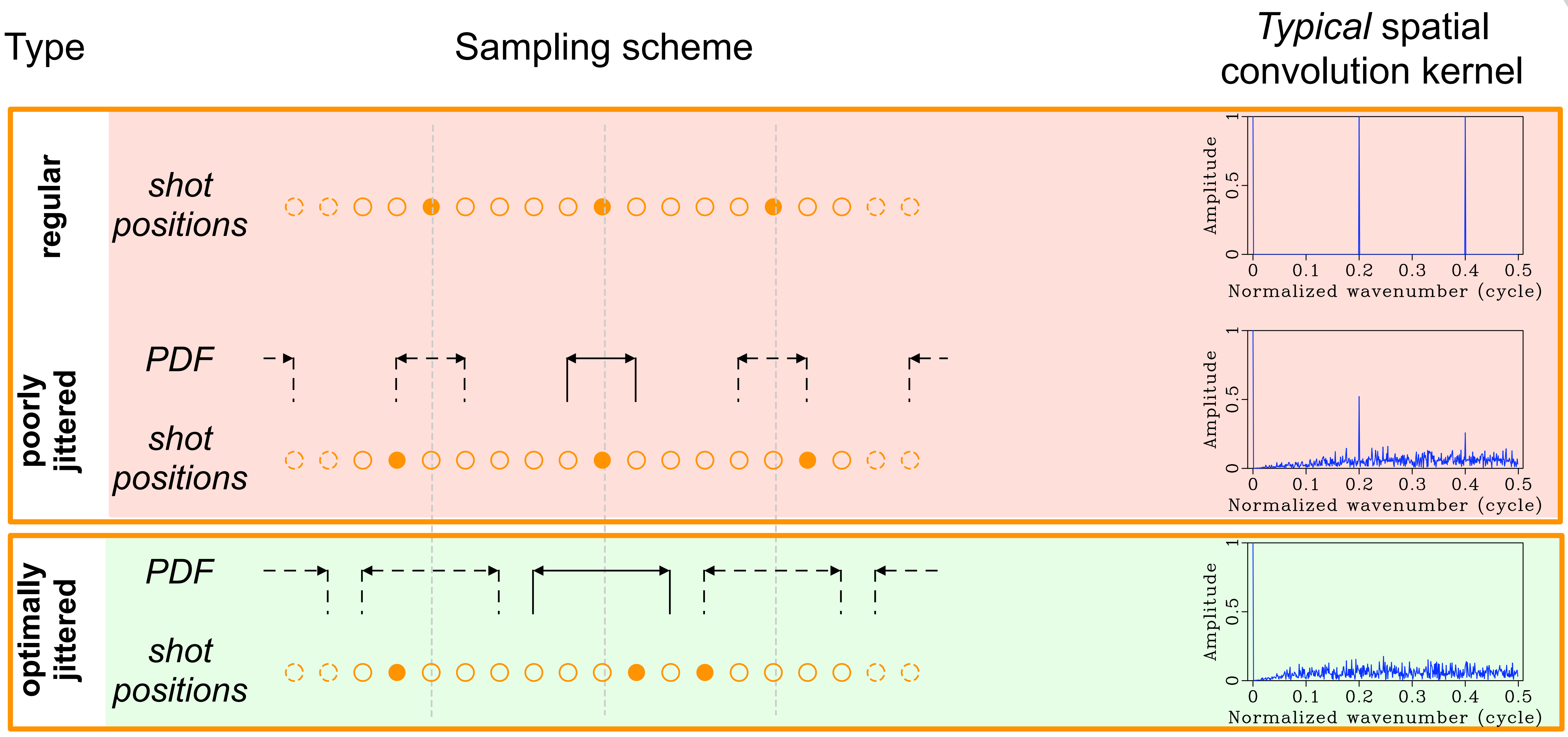


SNR = 8.9 dB

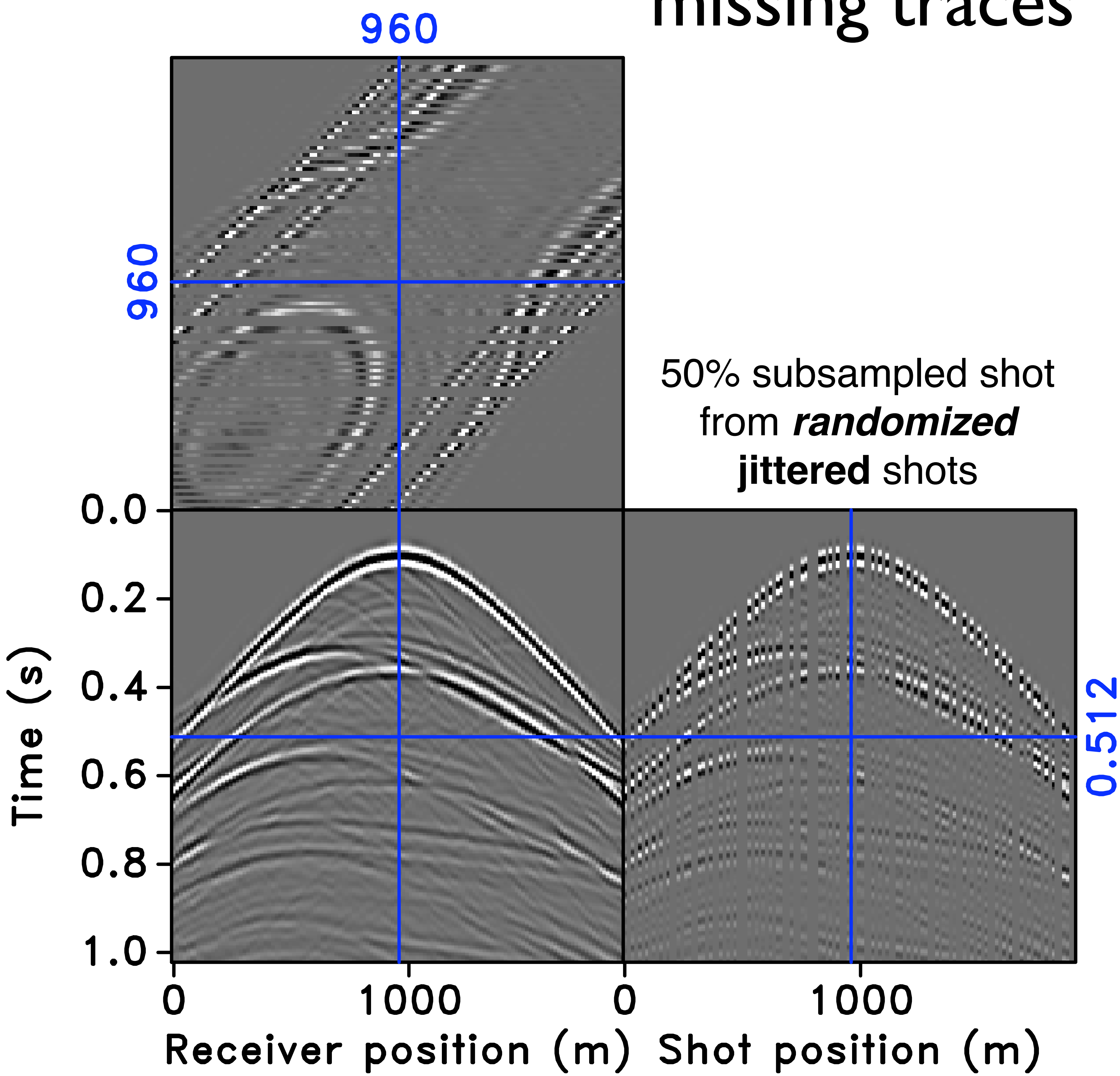
50% subsampled shot from **regularly** missing shot positions

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

Jittered sampling

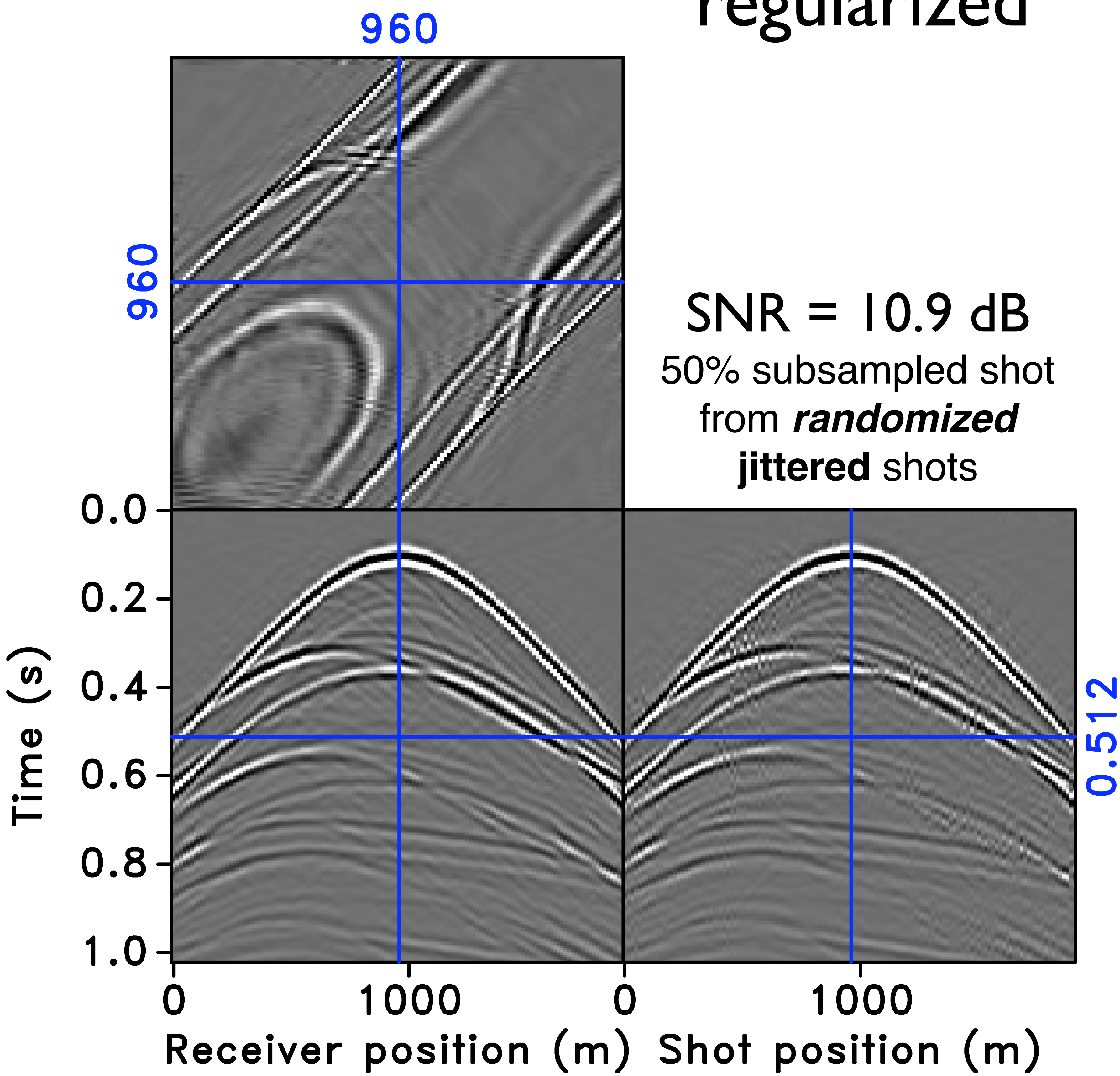


missing traces



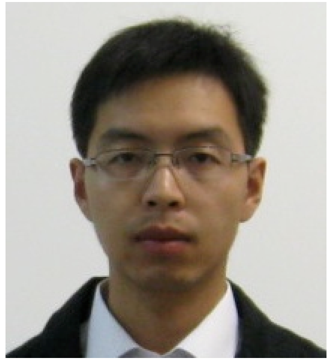
50% subsampled shot from ***randomized*** **jittered** shots

regularized



SNR = 10.9 dB

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



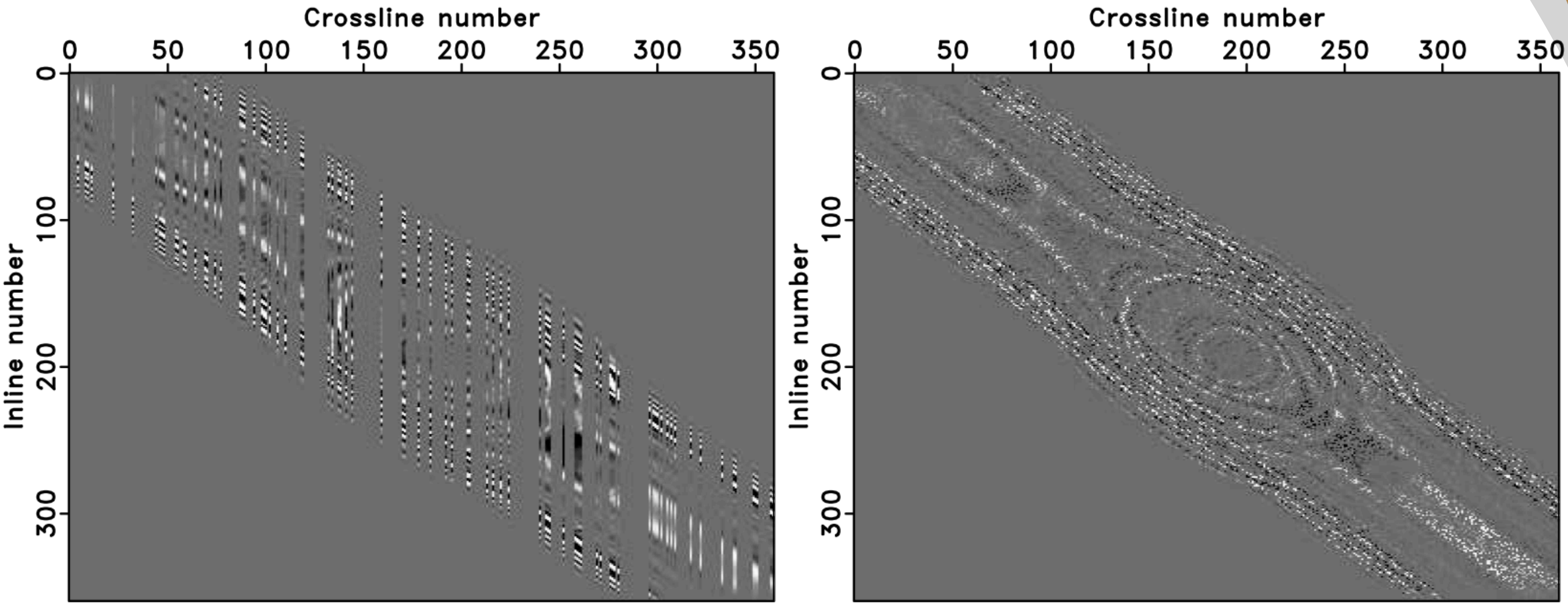
Recent advances

CS applied to acquisition design

"Higher dimensional blue-noise sampling schemes for curvelet-based seismic data recovery"

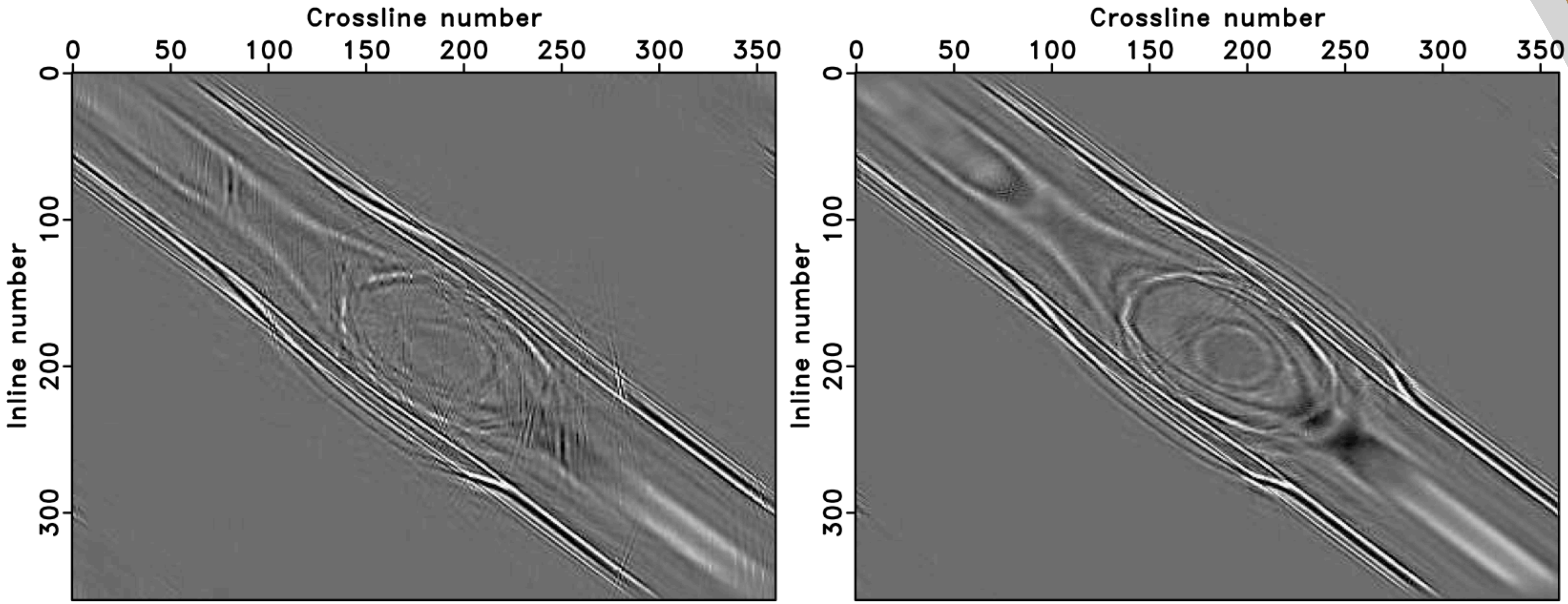
by Gang Tang, Reza Shahidi, Jianwei Ma, and Felix J. Herrmann. SPMUL 2 Multiples II Room: General Assembly C @ 03:10 PM

Multi-D jittering



25 % samples

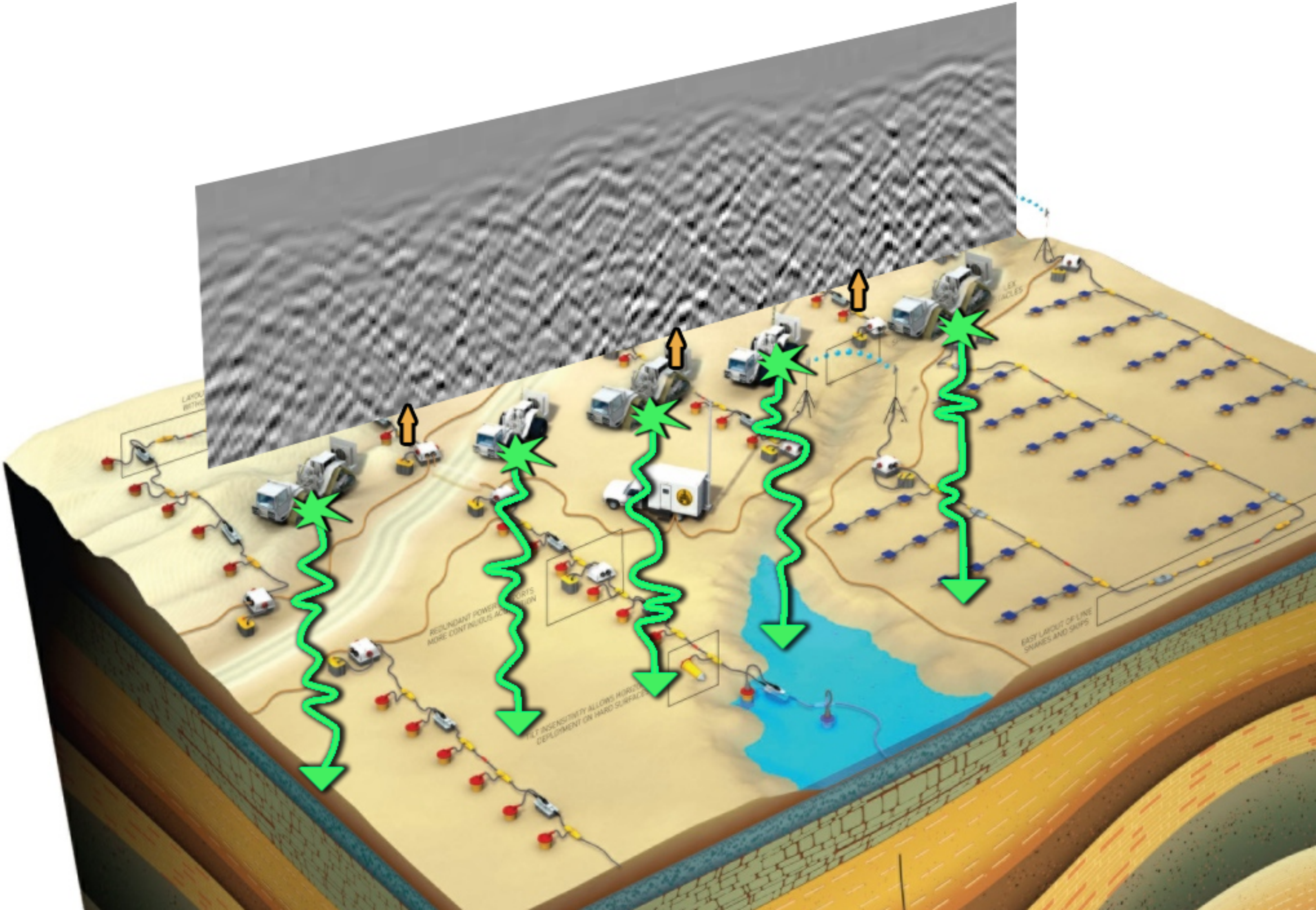
Multi-D jittering



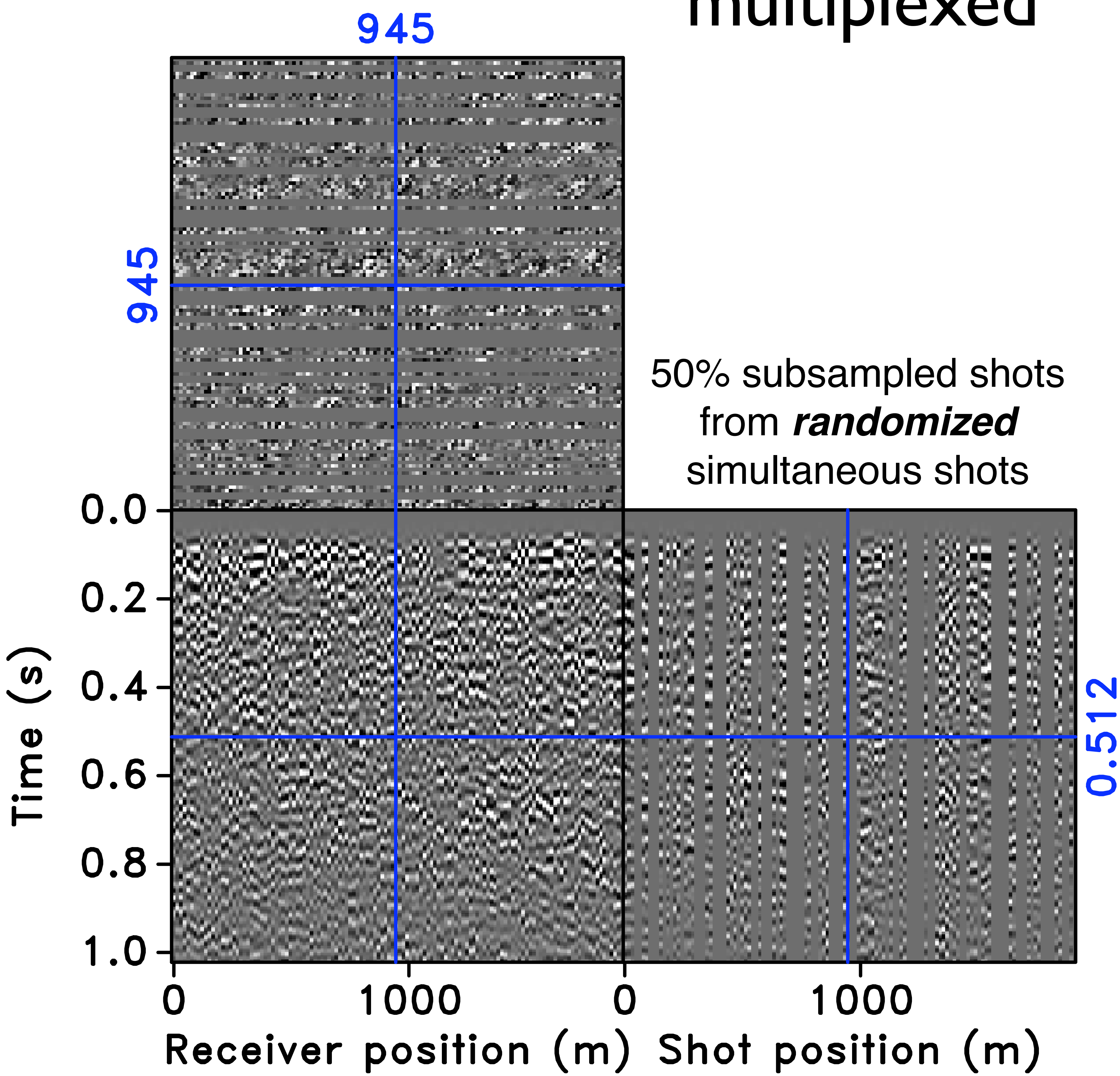
SNR=6.77 dB

SNR=9.75 dB

Simultaneous & incoherent sources

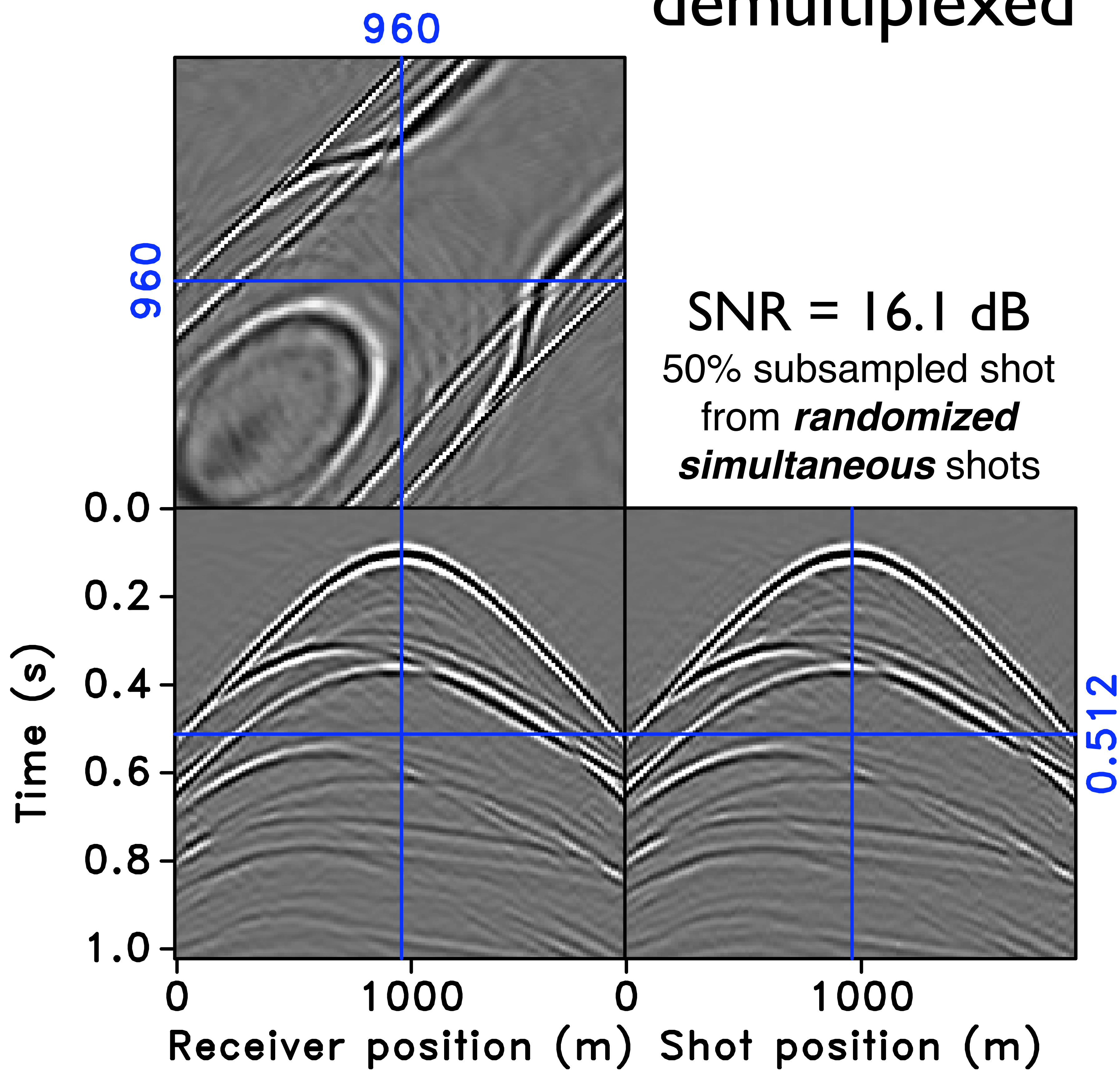


multiplexed



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

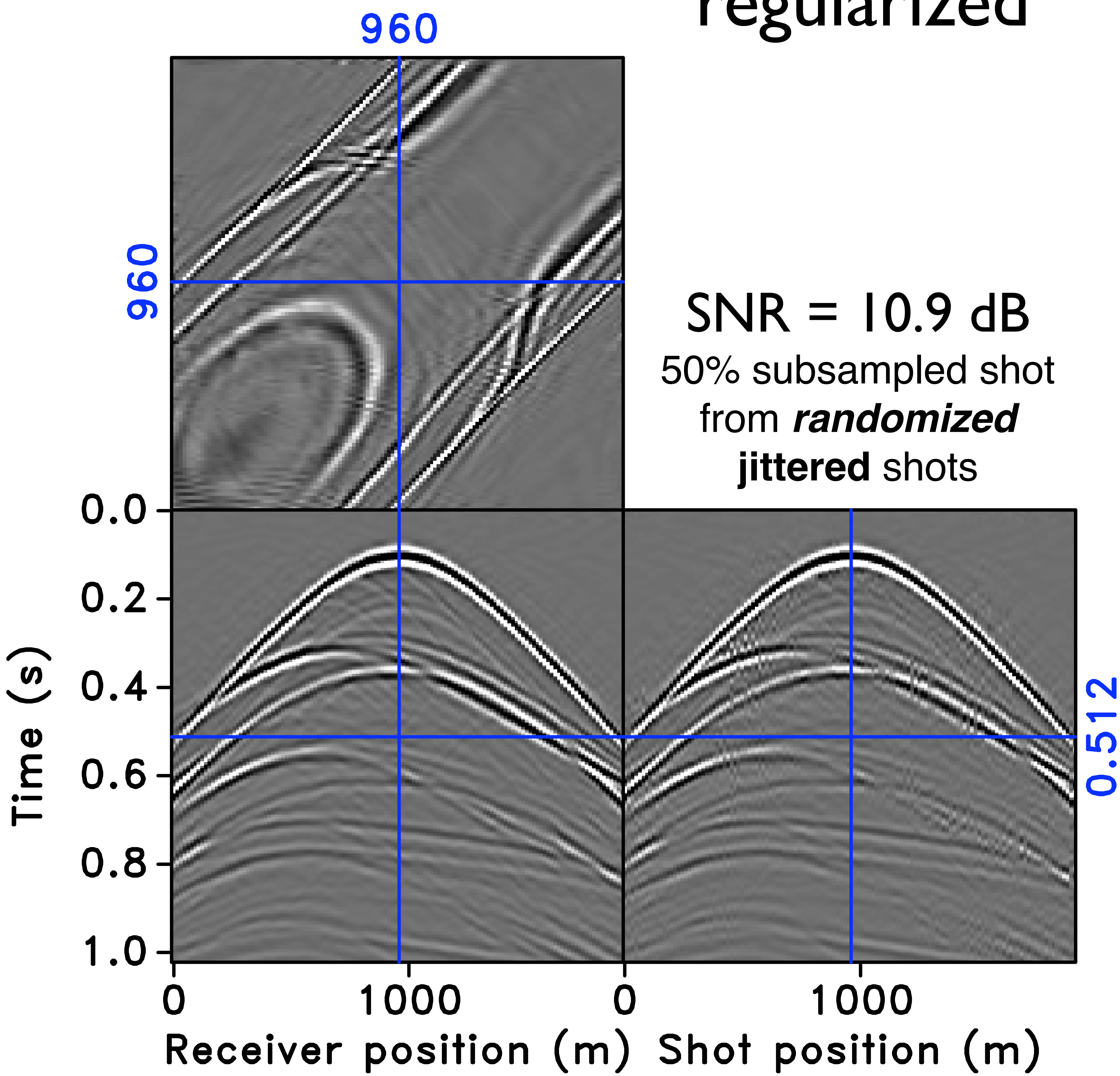
demultiplexed



SNR = 16.1 dB

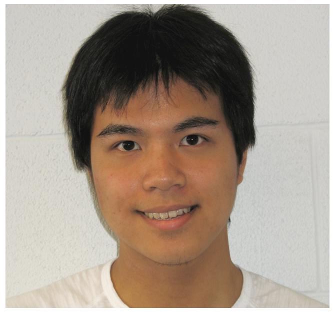
50% subsampled shot from ***randomized simultaneous*** shots

regularized



SNR = 10.9 dB

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

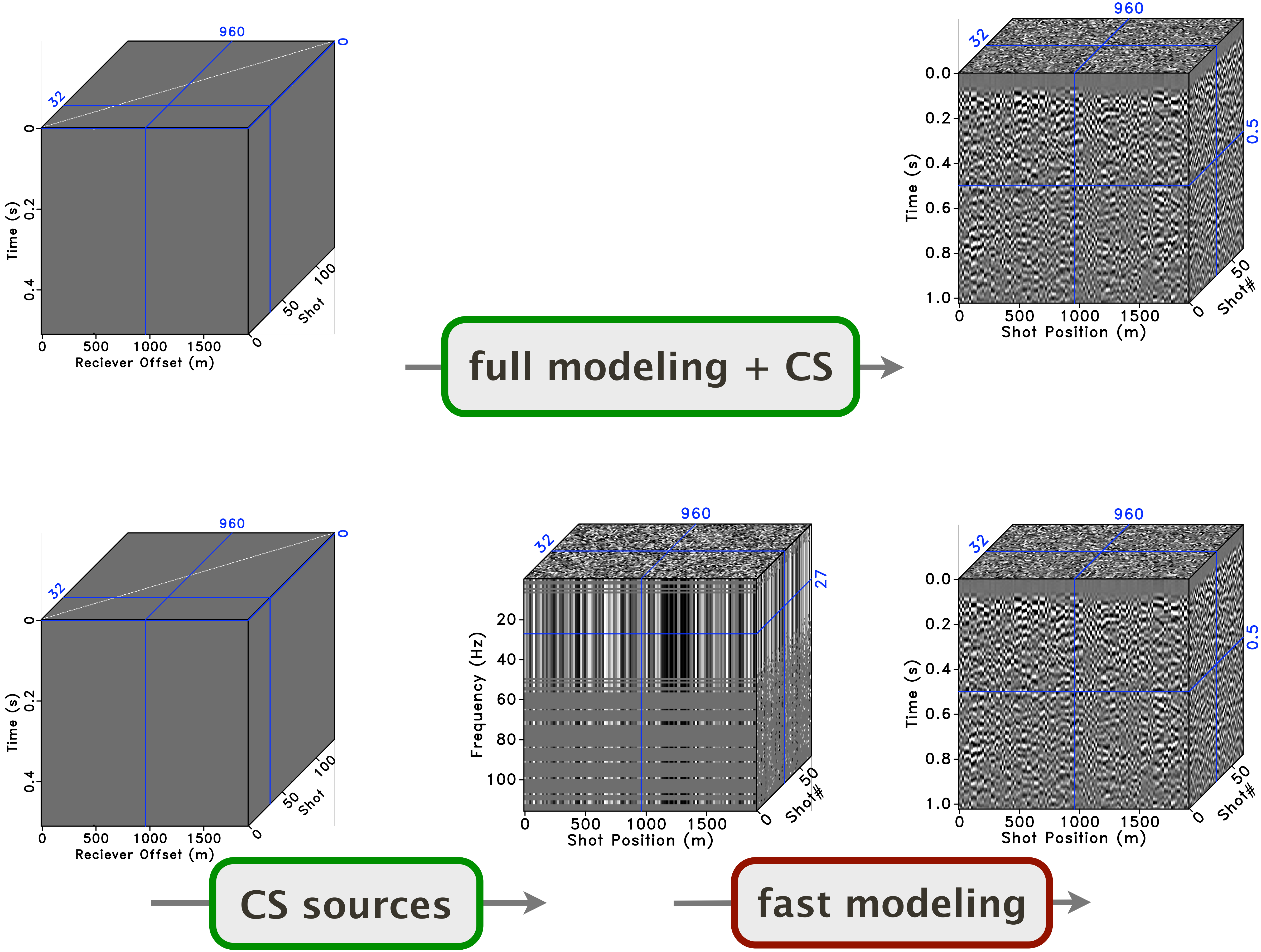


Recent advances

CS applied to forward modeling

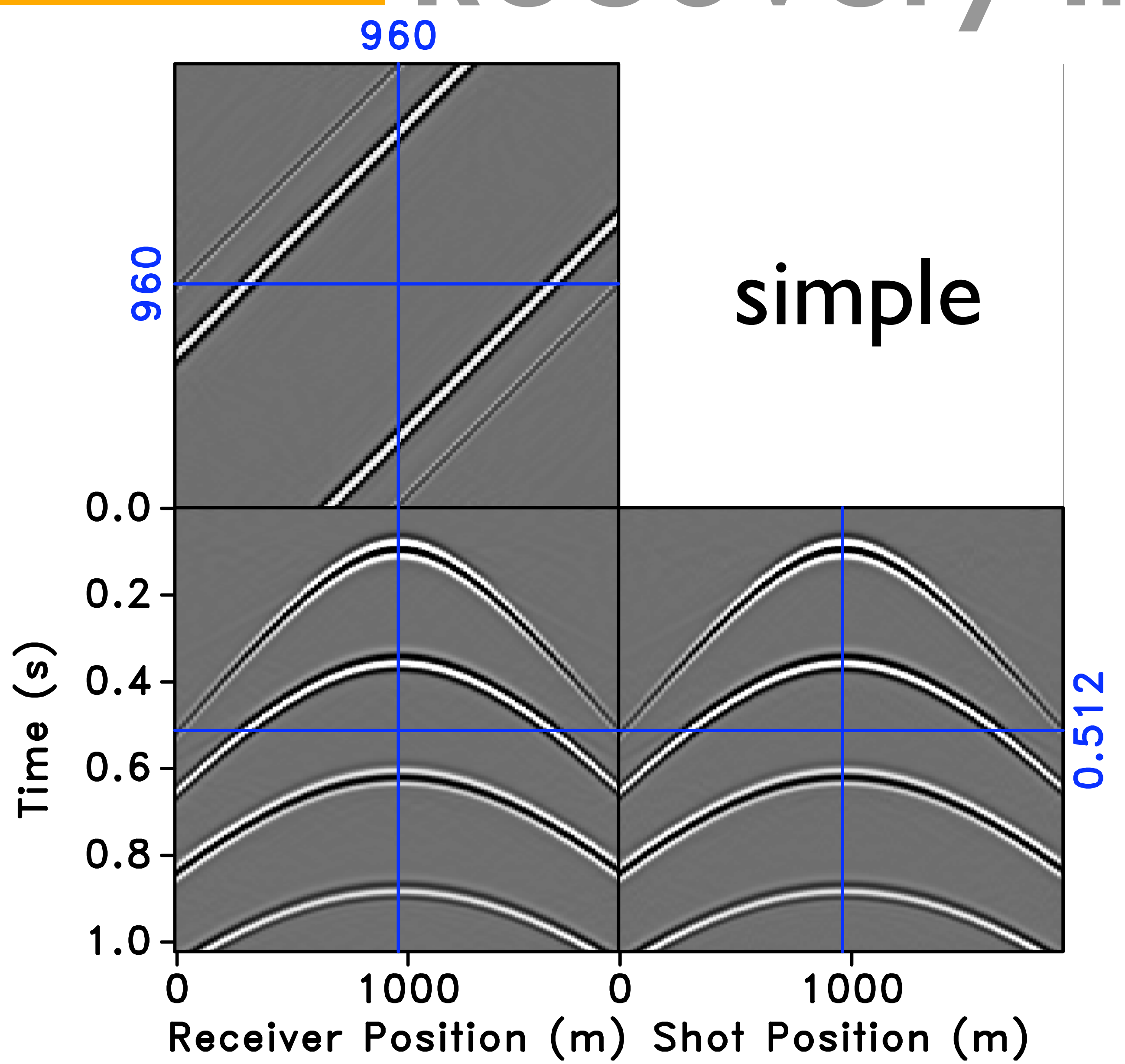
"Compresive simultaneous full-waveform simulation"

by Felix J. Herrmann, Tim T.Y. Lin*, Yogi A. Erlangga. SM 1 Algorithms and Methods Room 360 A @ 11:25 AM

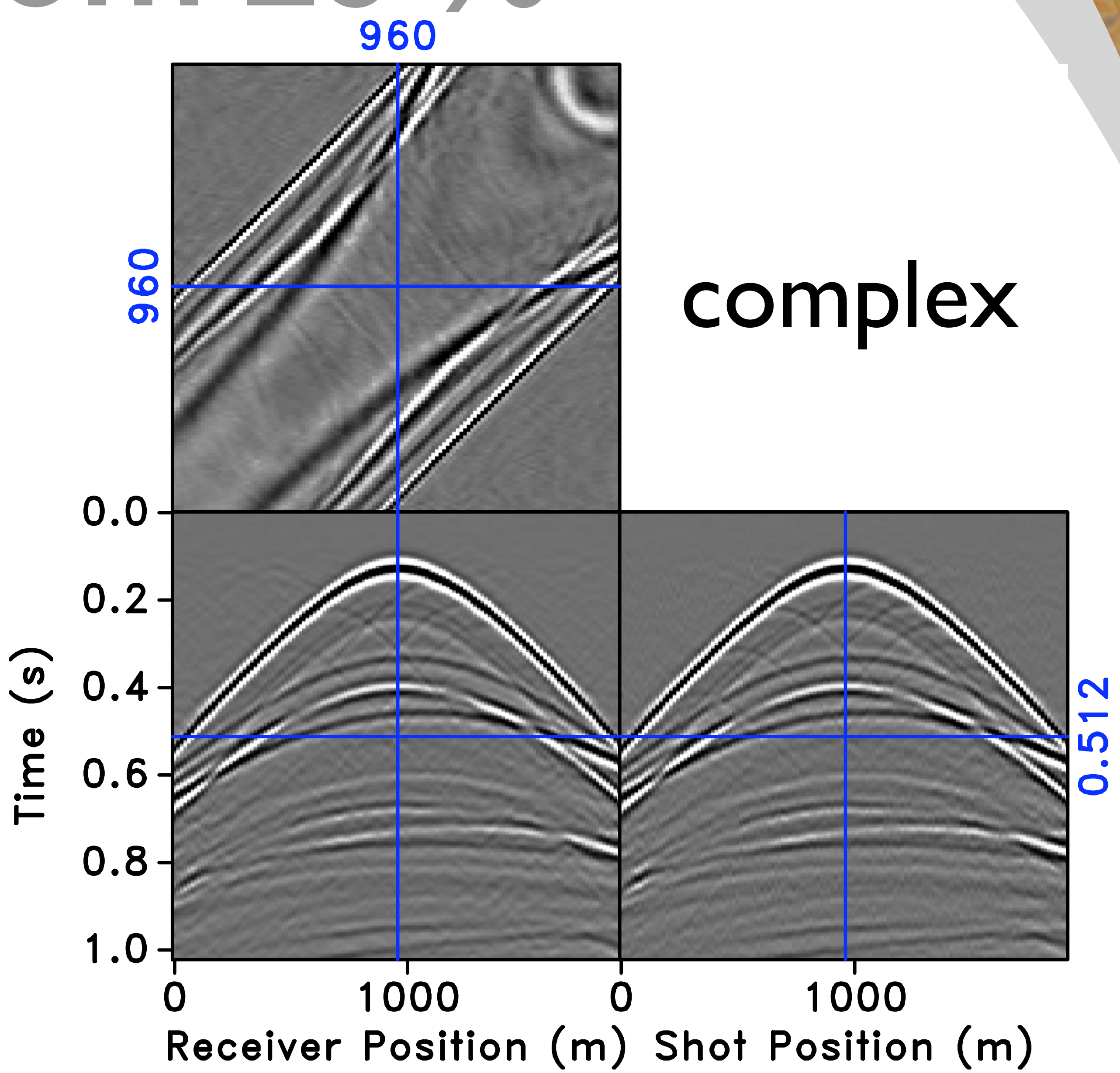


Time (s)
Reciever Offset (m)
Shot
full modeling + CS
Shot Position (m)
Shot#
CS sources
Frequency (Hz)
fast modeling

Recovery from 25 %



28.1dB



18.2dB

SLIM

Strategy

Adapt Compressive Sensing (CS)

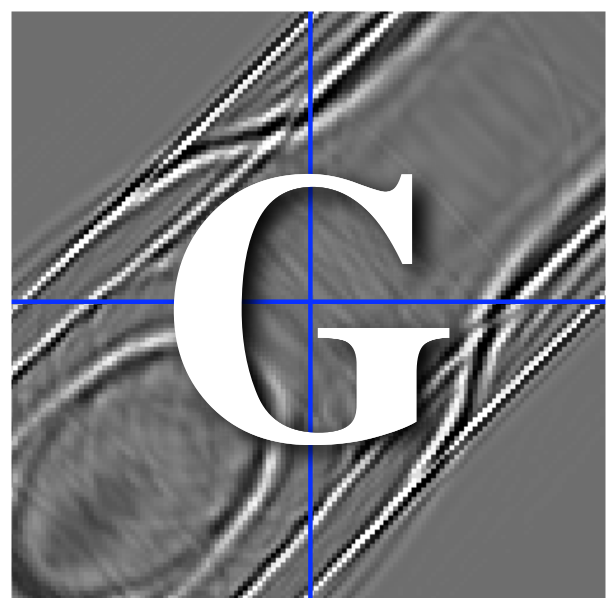
- *randomized* subsampling - turns *aliases/ interferences* into *noise*
- *sparsity* promotion - removes subsampling *noise* by exploiting signal *structure*

Ideal coverage

shot position

sources

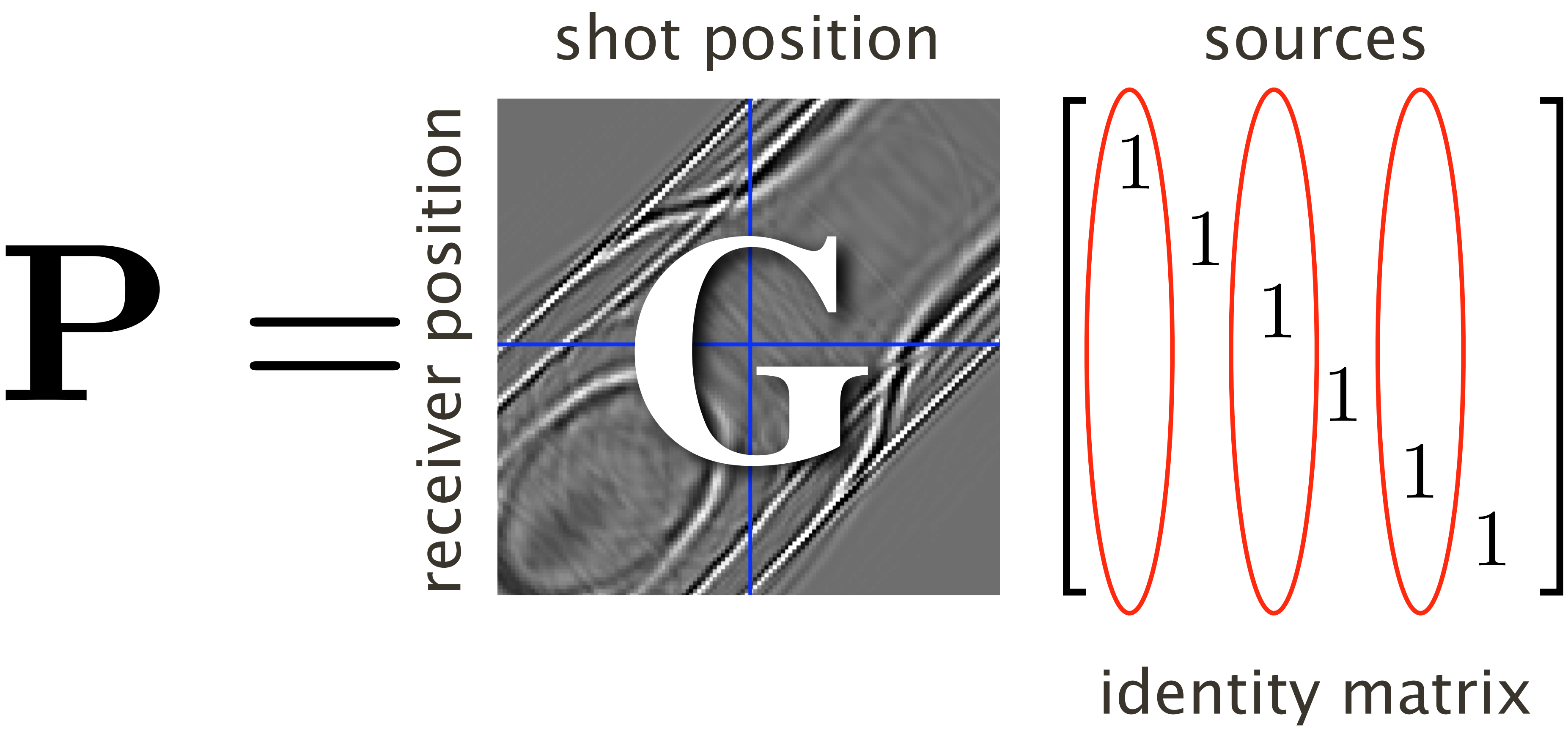
receiver position



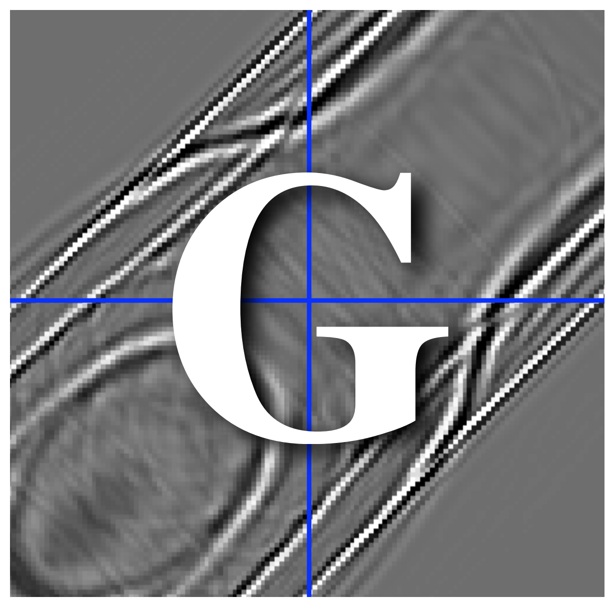
$$\mathbf{P} = \mathbf{G}\begin{bmatrix}1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1\end{bmatrix}$$

identity matrix

Ideal coverage



Actual coverage



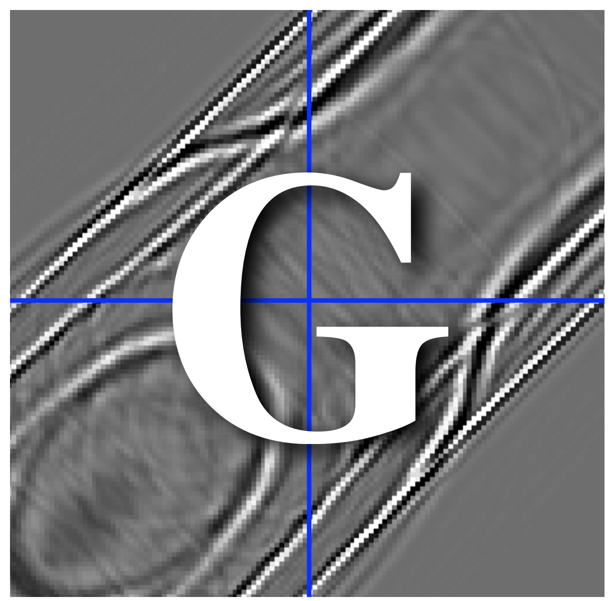
subset sources

$$\mathbf{P} = \mathbf{G} \begin{bmatrix} 1 & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

sampling matrix

Periodic 50 % subsampling

Sampling



subsampling matrix

$$\mathbf{P} = \mathbf{G}\left[\mathbf{R}^*\right]$$

CS matrix

$$\mathbf{A} = \left(\mathbf{R} \otimes \mathbf{I}\right)$$

linear compressive-sampling matrix

bad, bad examples

$$\mathbf{A} = \left(\begin{bmatrix} 1\;0 & & \\ & 1\;0 & \\ & & 1\;0 \end{bmatrix} \otimes \mathbf{I} \right)$$

(2x shot subsampling)

bad, bad examples

$$\mathbf{A} = \left(\begin{bmatrix} 1\ 1 & & \\ & 1\ 1 & \\ & & 1\ 1 \end{bmatrix} \otimes \mathbf{I} \right)$$

(every-other source simultaneous)

SLIM

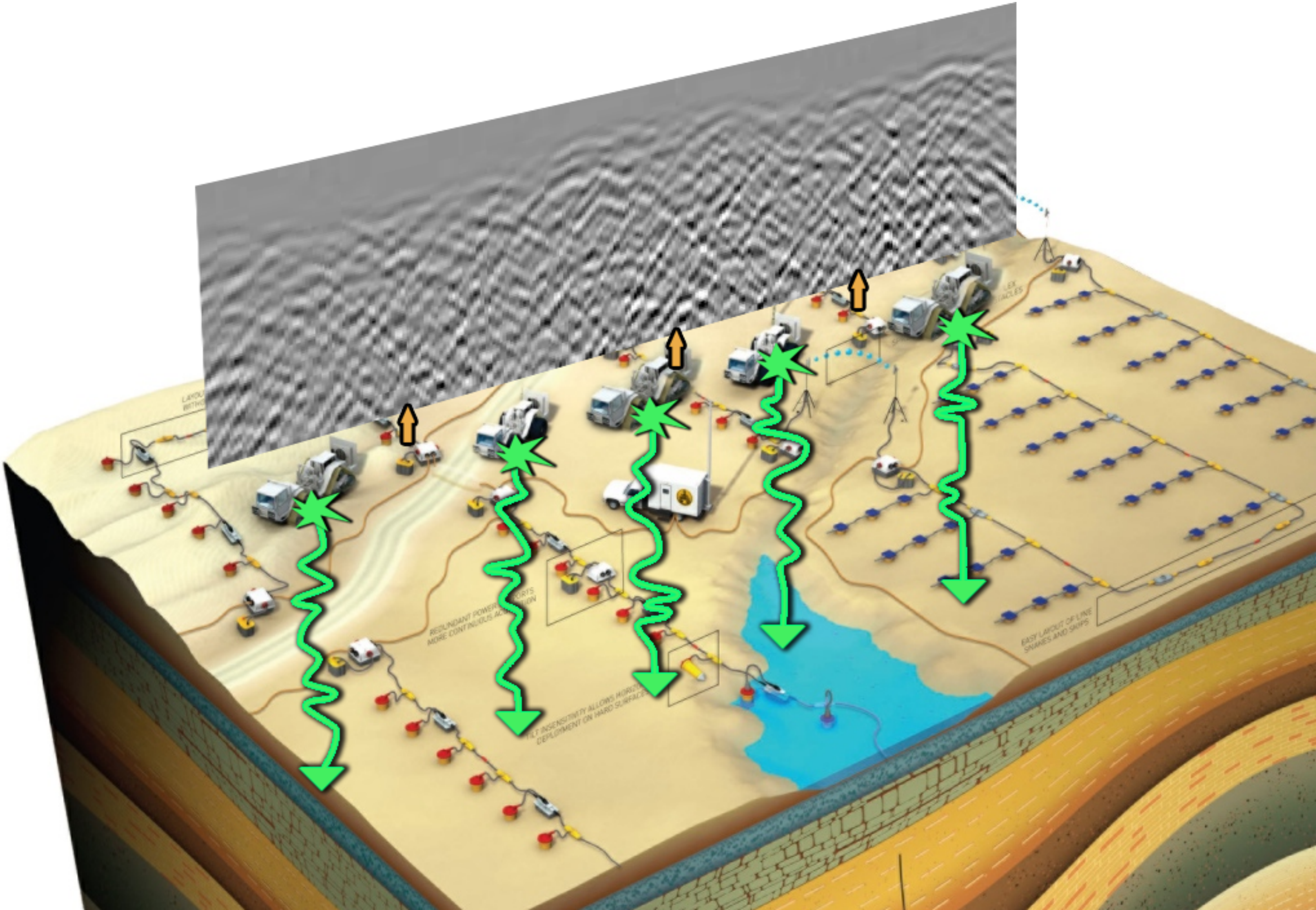
good example

$$\mathbf{A} = \left(\mathbf{R}\begin{bmatrix}\textbf{Gaussian}\\ \textbf{noise}\end{bmatrix} \otimes \mathbf{I}\right)$$

incoherent simultaneous sources

(Subsampled simultaneous-source experiments)

Simultaneous & incoherent sources



SLIM

Reality check

A new look at simultaneous sources by Beasley et. al., '98.

Reality check

Changing the mindset in seismic data acquisition by Berkhout '08.

now, P ← G ✓

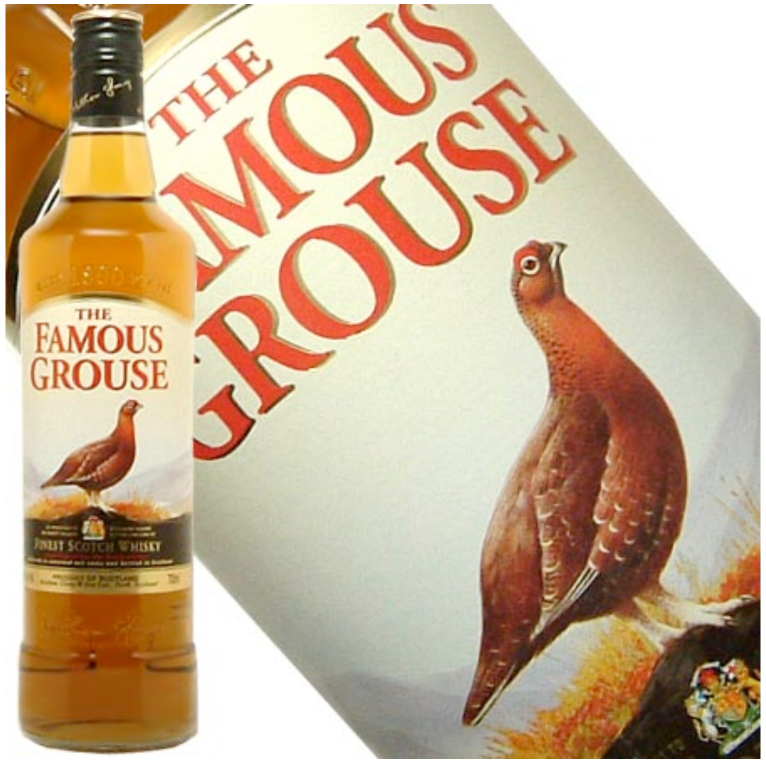
P → G ?

SLIM

Or

✓ →

? ←



$$$$$$$$$$ $

Blending versus *unblending* ...

Strategy cont'd

Adapt Compressive Sensing (CS)

- *randomized* subsampling - turns *aliases/ interference* into *noise*

➡ *sparsity* promotion - removes subsampling *noise* by exploiting signal *structure*

Least squares

We know that

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|$$

with $\mathbf{b} = \mathrm{vec}(\mathbf{P})$ and $\mathbf{A} = (\mathbf{R} \otimes \mathbf{I})$

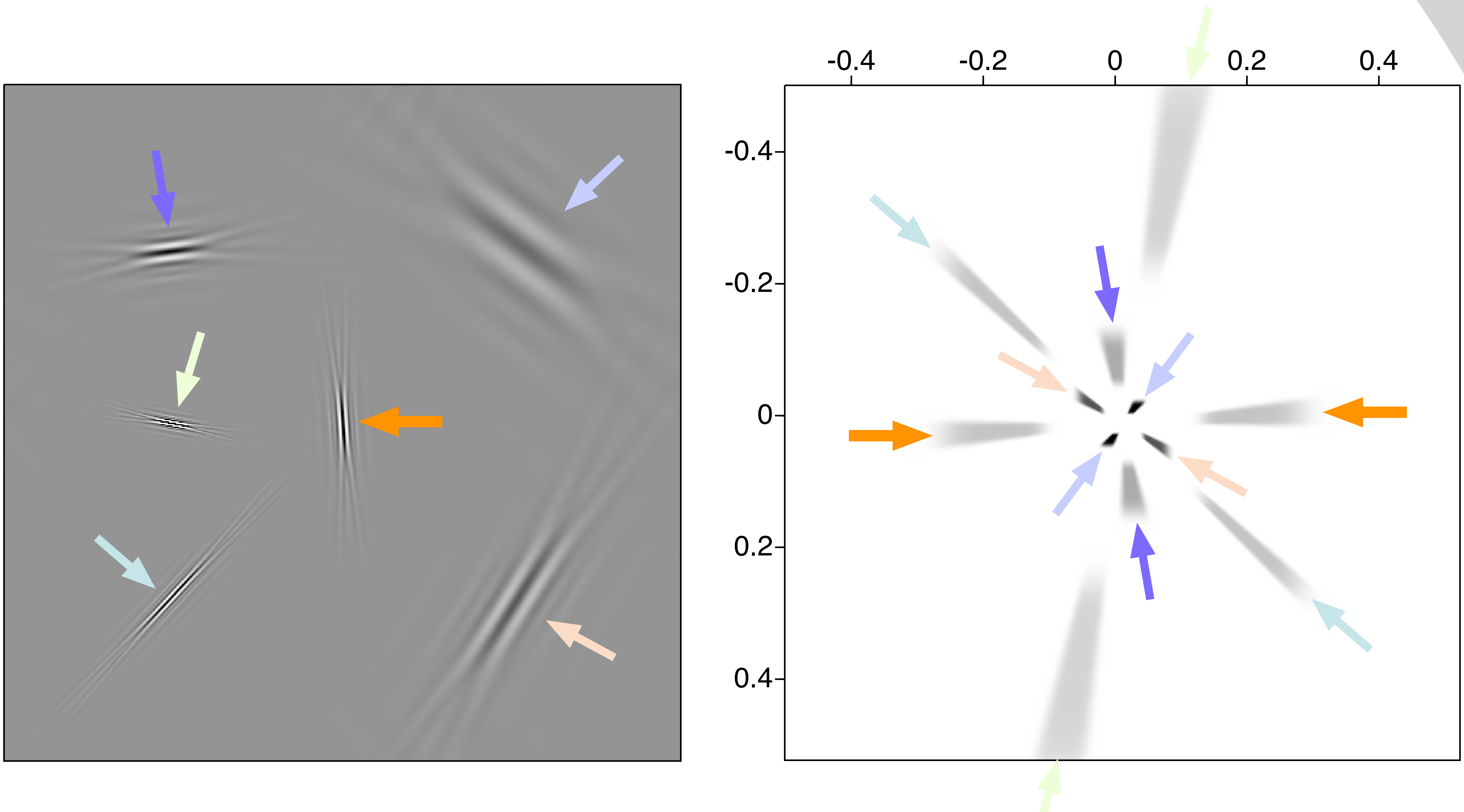
does not promote sparsity...

but wait...

I know geophysics!
G has some sort of
structure

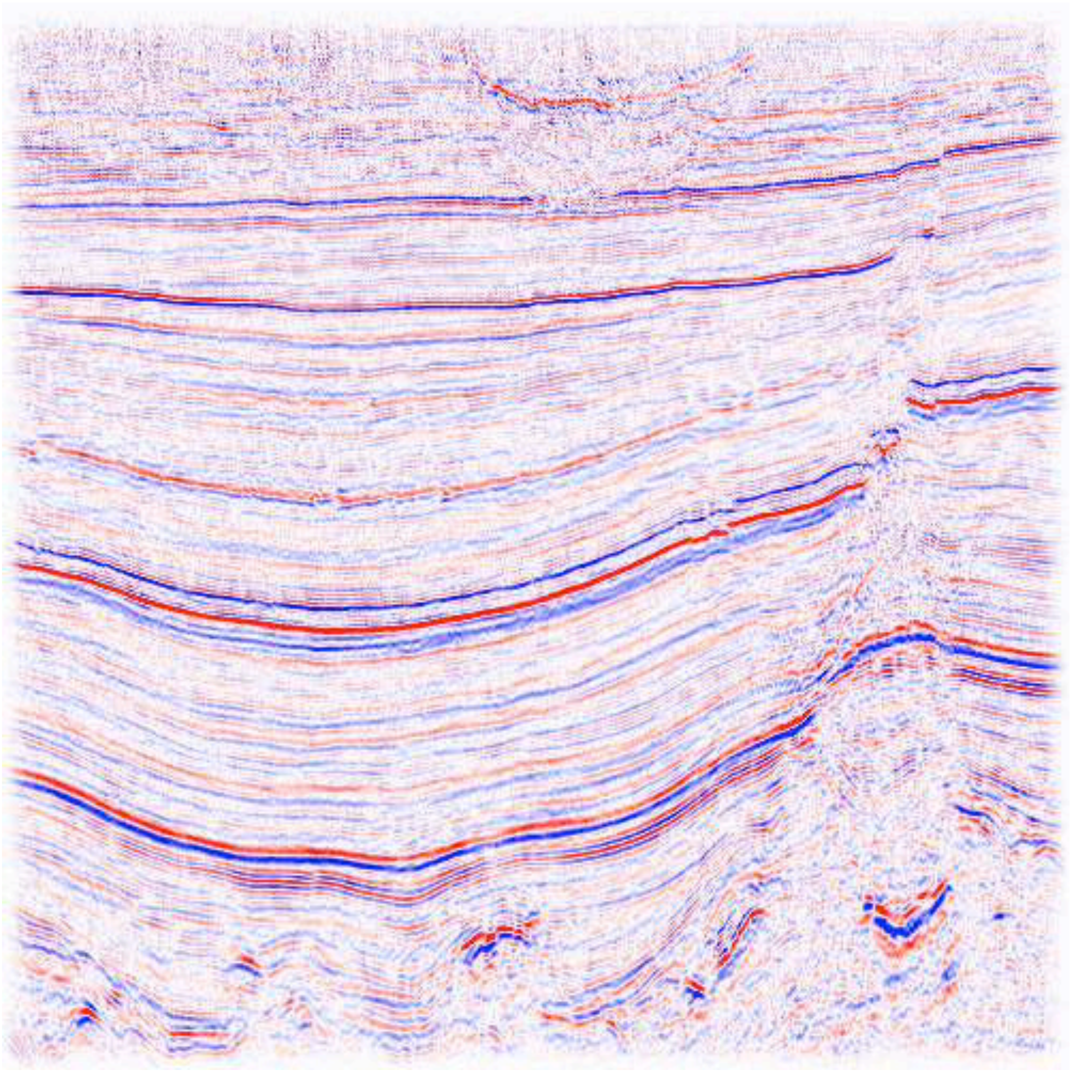
[Demanet et. al., '06]
[Hennenfent & FJH, '06]

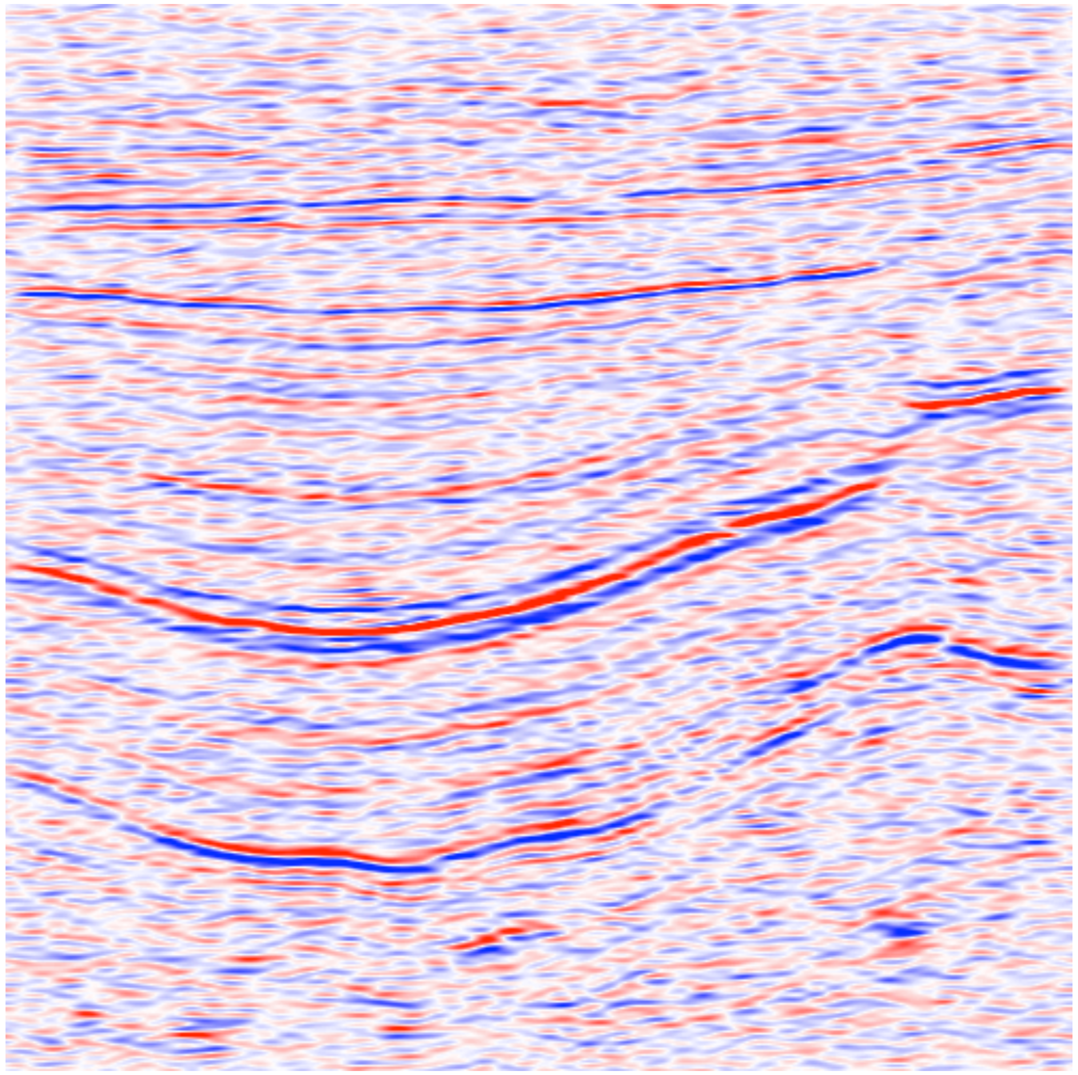
Curvelets



Fourier

SNR 2.1 dB

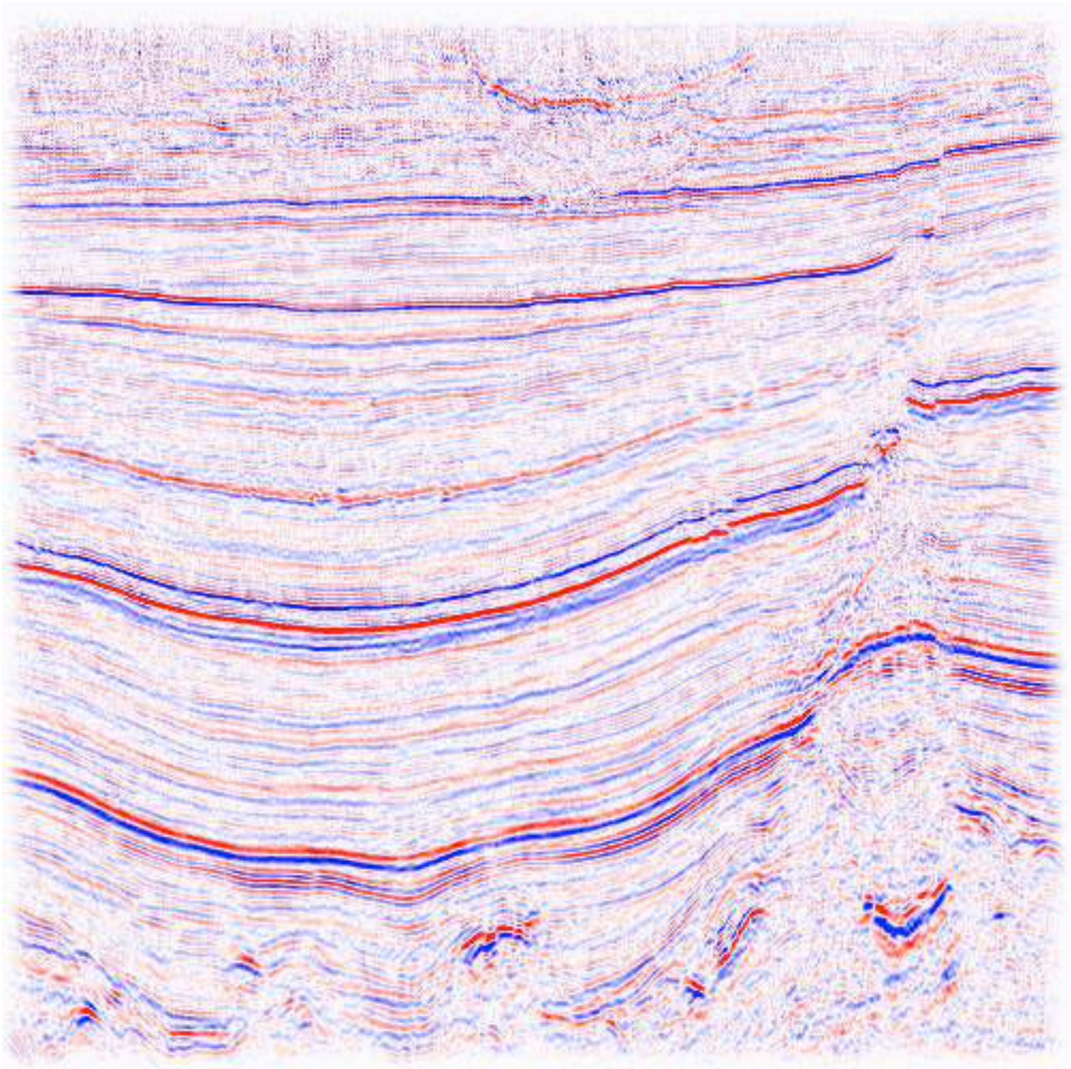


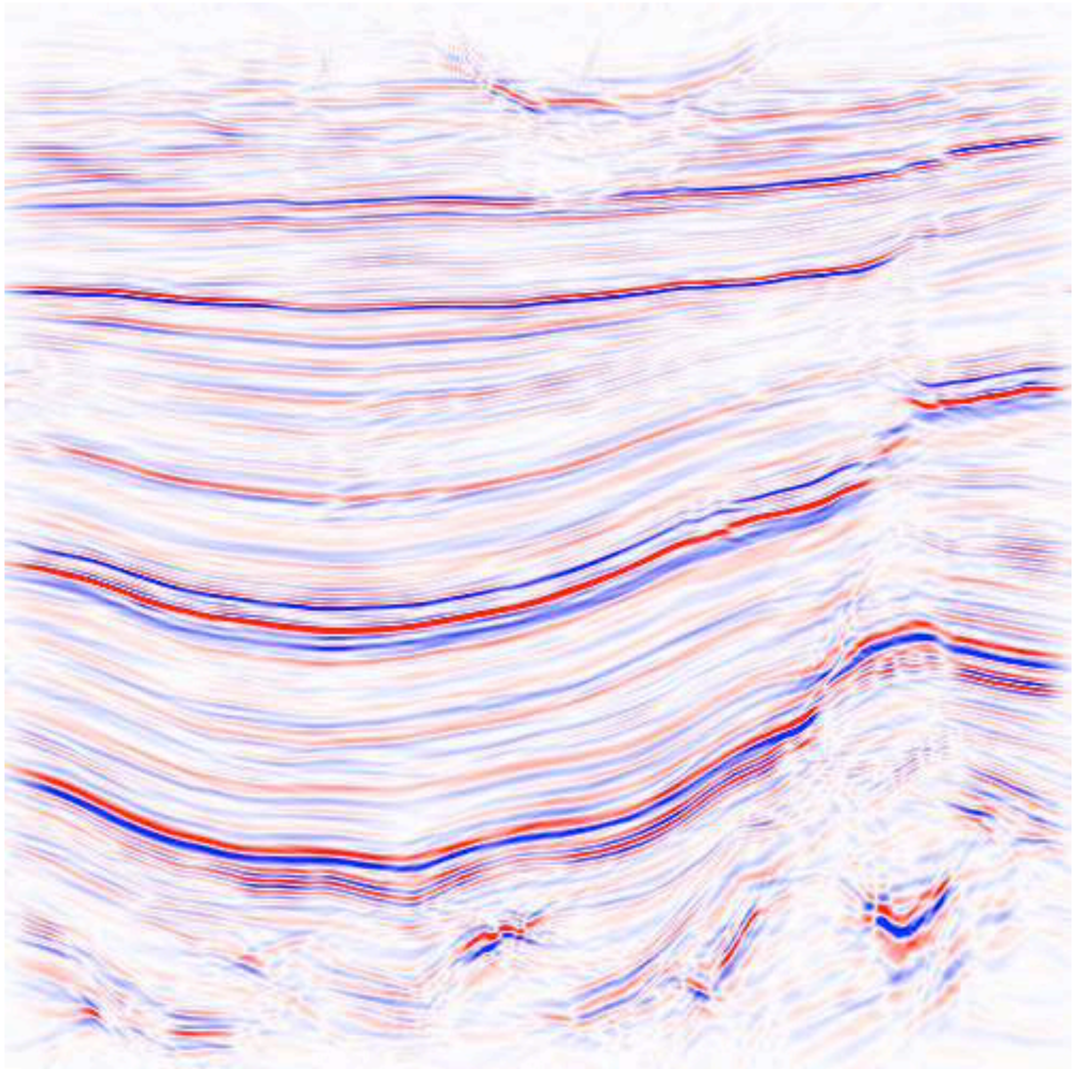


1 % of coefficients

Curvelets

SNR 6.0 dB





1 % of coefficients

Sparsifying transform

So, we know... a compressive representation **S**

$$\mathbf{g} = \mathbf{S}^{*}\mathbf{x}_0$$

(x_0 is compressible or sparse)

CS matrix

$$\mathbf{A} = (\mathbf{R} \otimes \mathbf{I})\mathbf{S}^*$$

linear compressive-sampling matrix
with sparsifying transform

Promote sparsity

$$\min_{\mathbf{x}} \quad \mathrm{nnz}(\mathbf{x}) \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

Is unfortunately NP hard ...

SLIM

talk to ~~strangers~~ mathematicians

Candes

Tao

Donoho

Romberg

talk to ~~strangers~~ mathematicians

"Look at A!"

(Compressive Sensing)

Recovery conditions

$$(1-\delta_k)\|\mathbf{x}_T\|_{\ell 2} \leq \|\mathbf{A}_T\mathbf{x}\|_{\ell 2} \leq (1+\delta_k)\|\mathbf{x}_T\|_{\ell 2}$$

(Restricted Isometry Property)

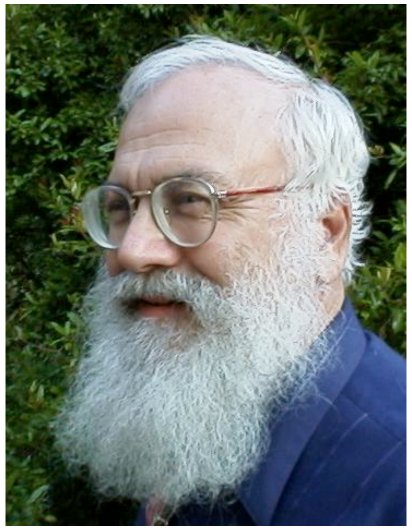
- Related to Johnson-Lindenstrauss Lemma
- CS establishes links between
  - subsampling *rate* & *sparsity*
  - recovery *error* & subsampling *rate*
- *Equivalence one*- and *zero*-norm *minimization*

Convexification

We know that

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_{\ell_1}$$
$$\text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}$$

is a very good
convex relaxation



Reality check

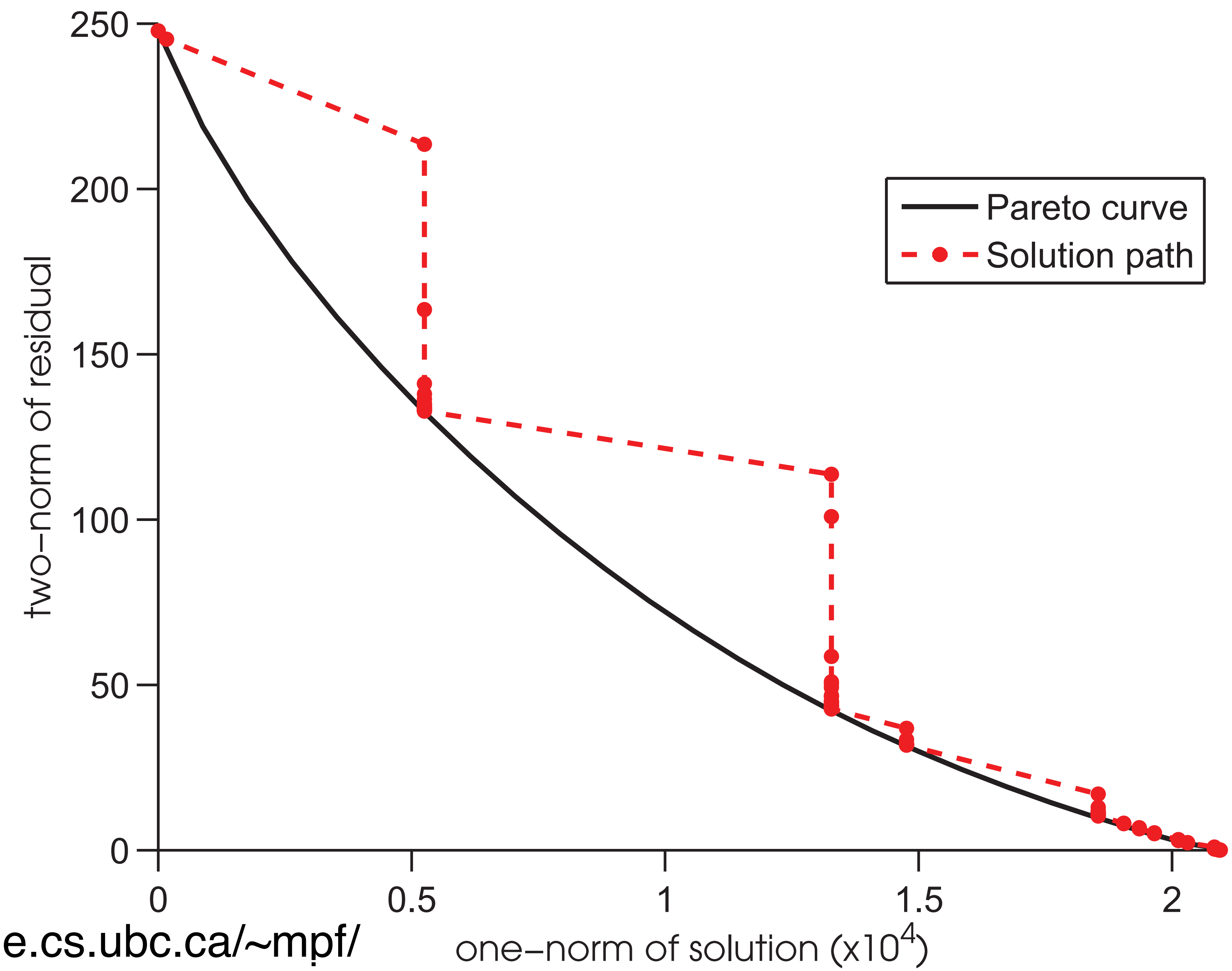
"When a traveler reaches a fork in the road, the l_1-norm tells him to take either one way or the other, but the l_2 -norm instructs him to head off into the bushes."

John F. Claerbout and Francis Muir, 1973

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

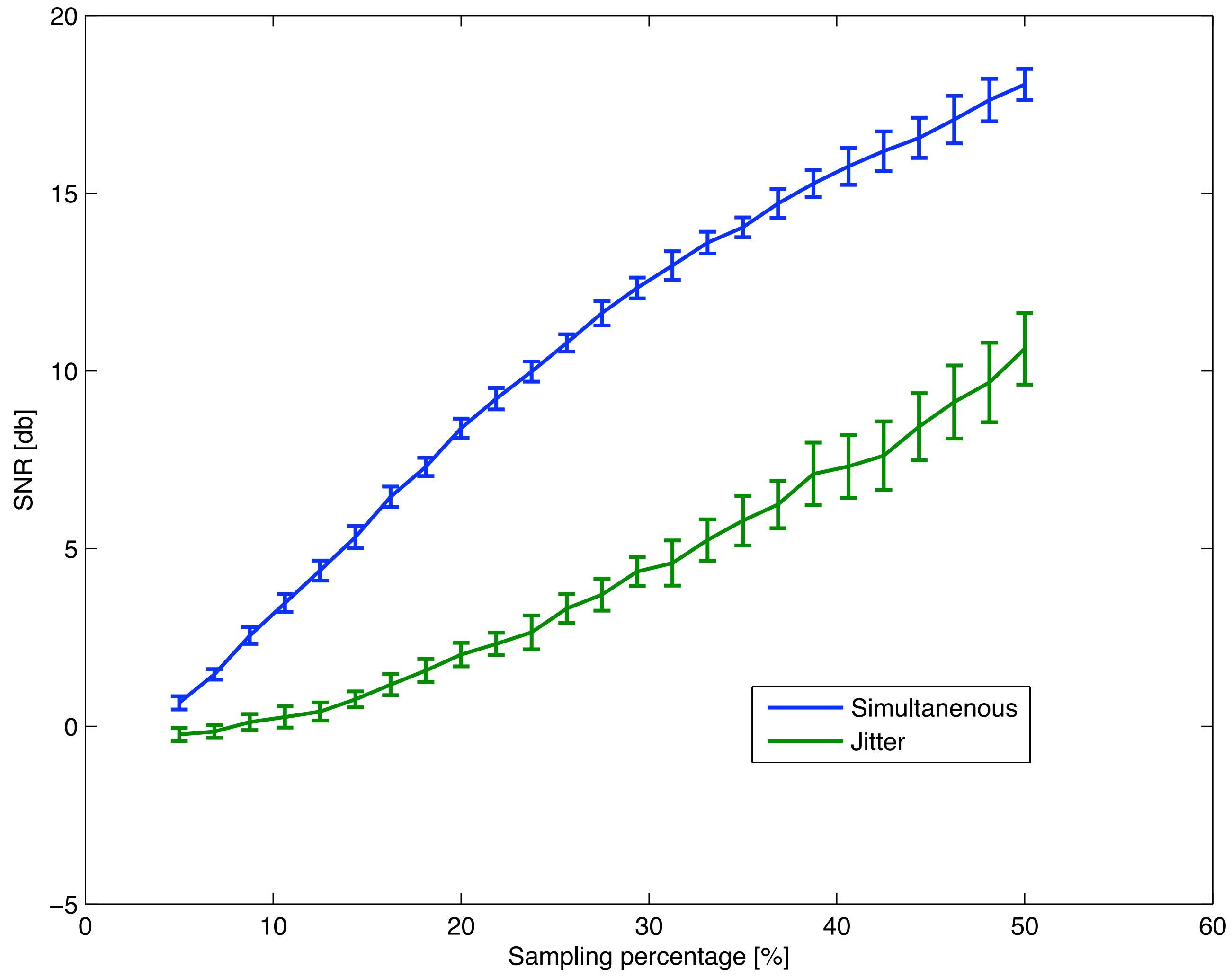
SLIM

One-norm solver



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

Controlled experiments



Bottom line

CS acquisition & recovery costs are *proportional* to

- transform-domain sparsity: the sparser the cheaper acquisition/the faster the turnaround
- recovery error: the larger the permissible error the cheaper the acquisition/the faster the turnaround ...

Design principles: the road ahead

- ***randomize*** **-** break subsampling *interferences*
- ***sparsify*** **-** exploit *structure* by *transform-domain sparsity promotion*

➡ ***focus*** **-** leverage *physics of waves*

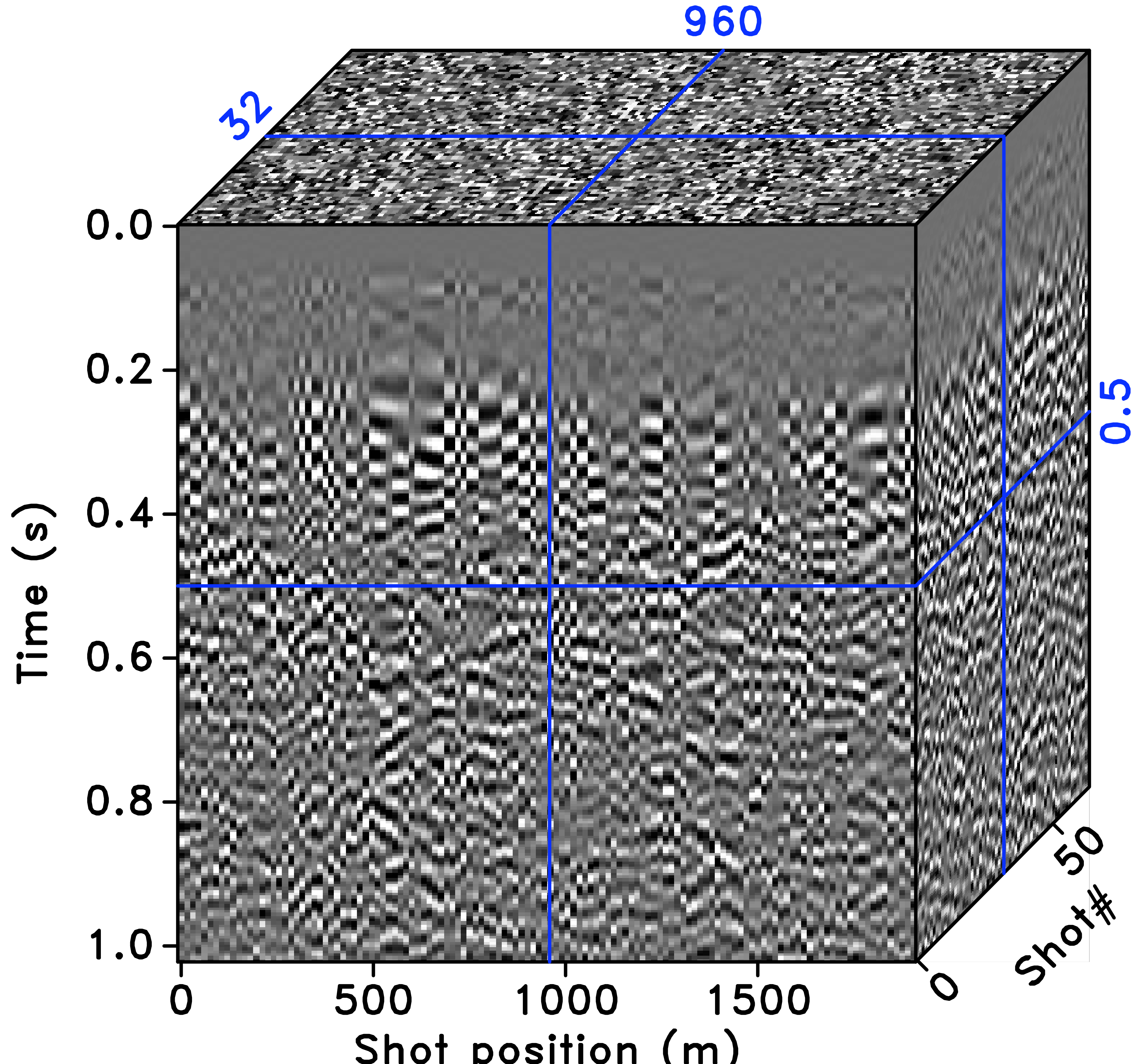
Case study II

Processing according to CS

- CS recovery from *simultaneous* data, followed by *primary* estimation

vs.

- *Primary* estimation *directly* from *simultaneous* data

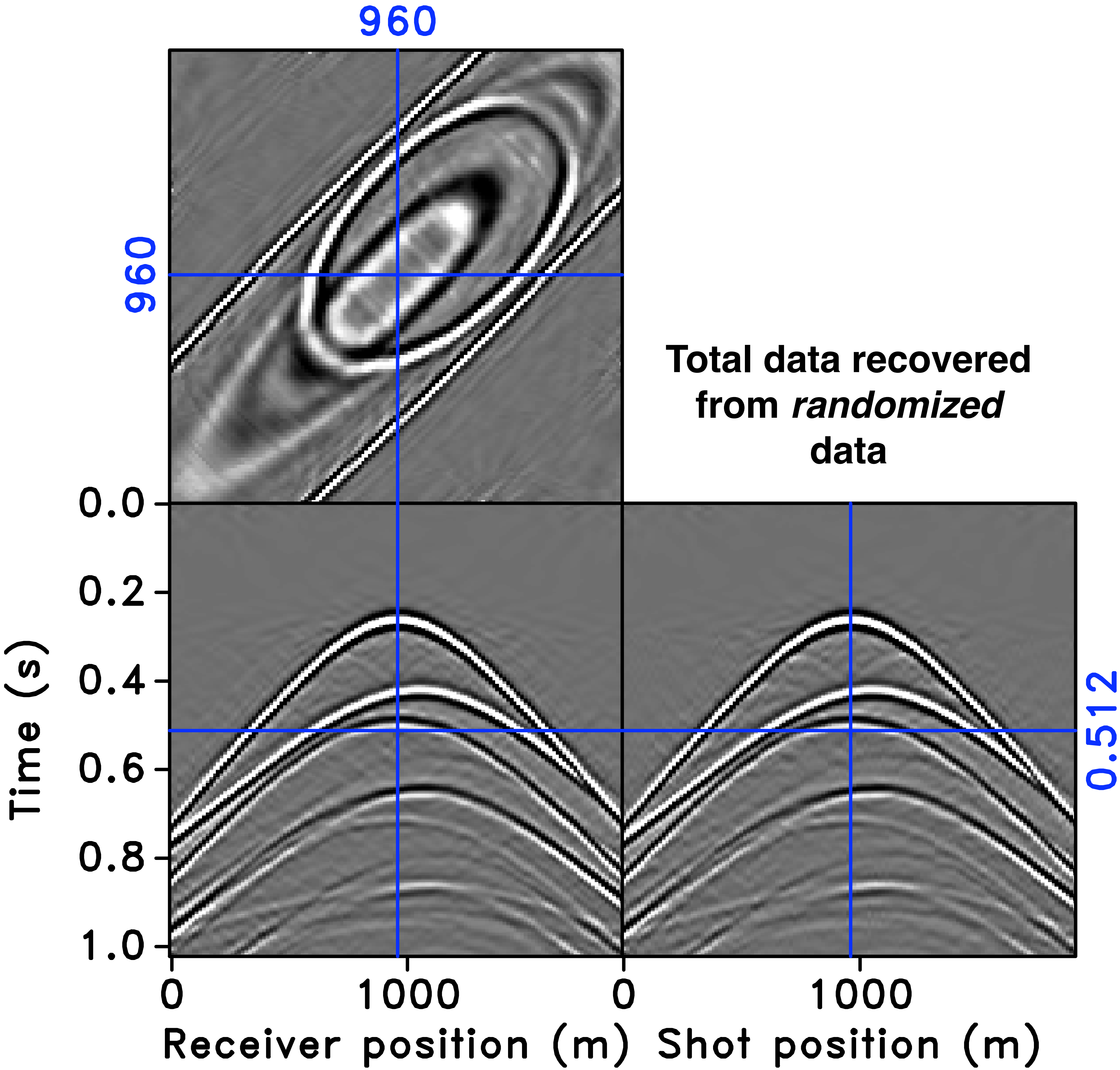

960
32
0.5
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
0
500
1000
1500
Shot position (m)
0
50
Shot#

CS

Use to demultiplex

$$\mathbf{A} = \left(\mathbf{R}\begin{bmatrix}\textbf{Gaussian}\\\textbf{noise}\end{bmatrix} \otimes \mathbf{I}\right)\mathbf{S}^*$$

(Randomized simultaneous sources)



Total data recovered from *randomized* data

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

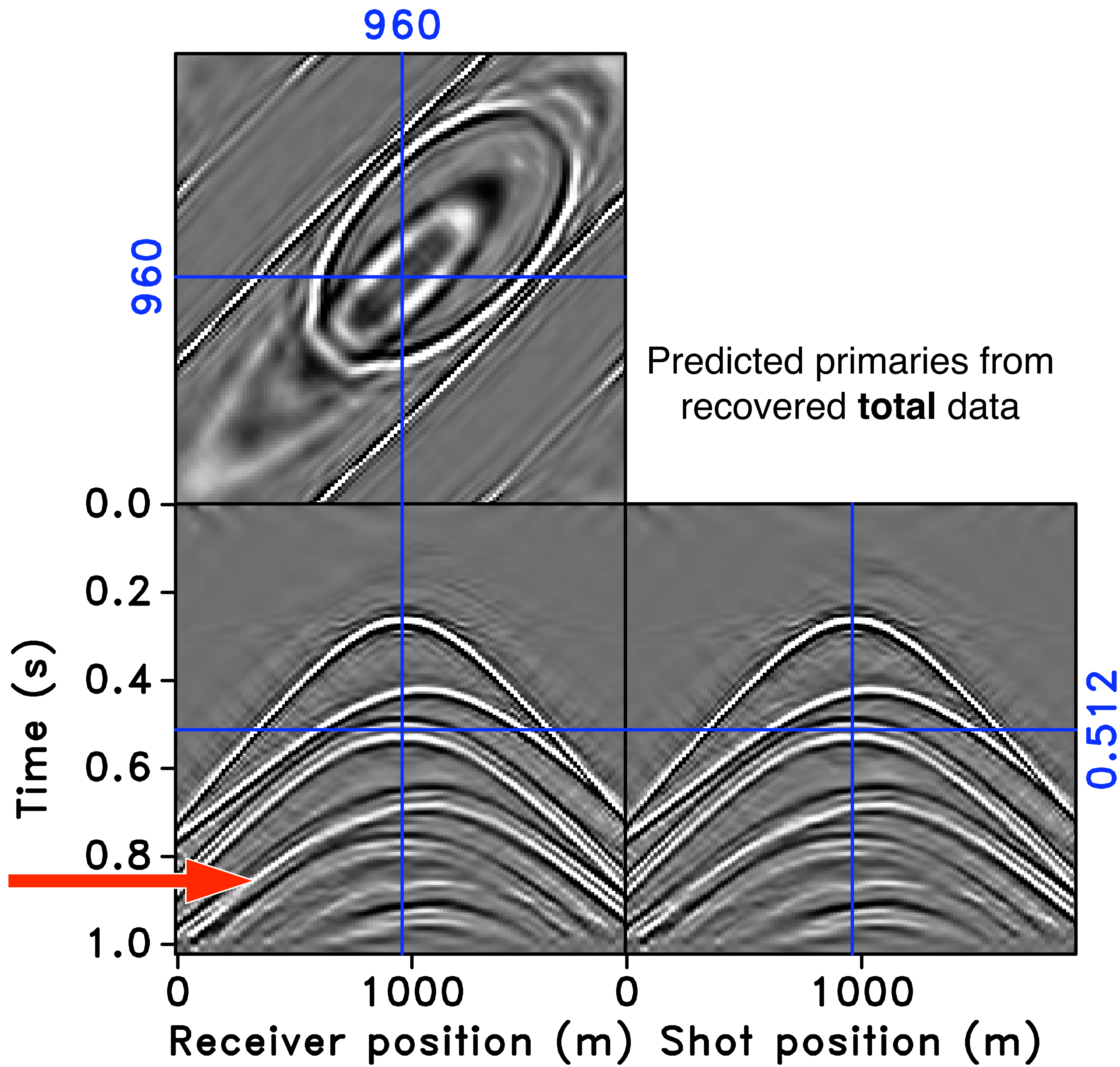
SLIM

Physical principle

Modeling the surface:

$$\overbrace{\mathbf{P}}^{\text{upgoing wavefield}} \approx \underbrace{\mathbf{G}}_{\text{surface-free impulse response}} \overbrace{[\mathbf{Q}-\mathbf{P}]}^{\text{downgoing wavefield}}$$

Inversion "focusses" multiples onto primaries ...


960
960
Predicted primaries from
recovered **total** data
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
0.512
0
1000
0
1000
Receiver position (m)
Shot position (m)

[Lin & FJH, '09]

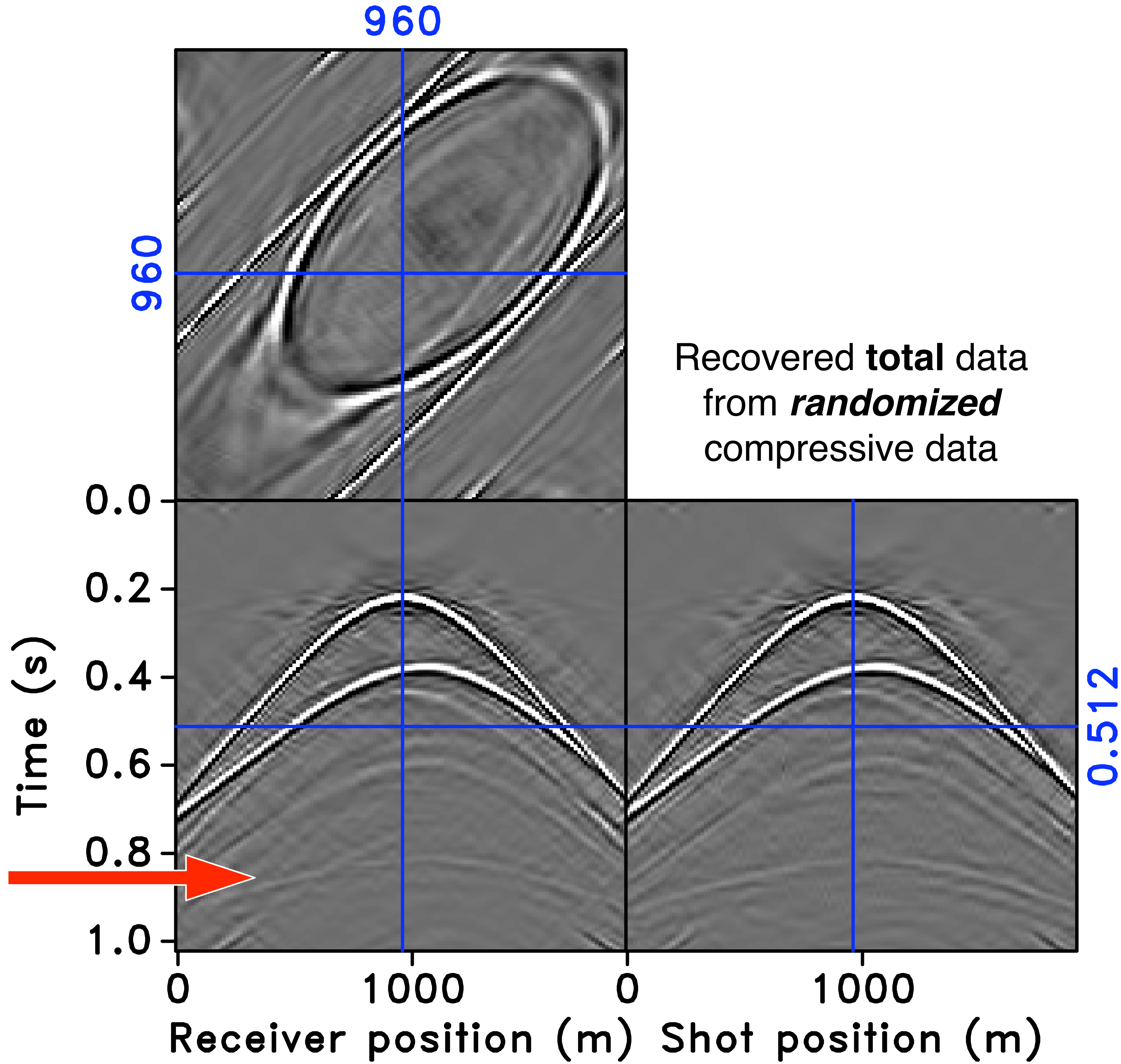
Extension CS

Use to demultiplex & predict

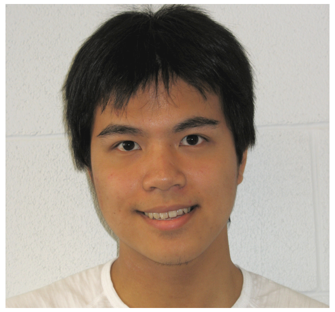
randomized physics

$$\mathbf{A} = \left[\mathbf{R}\right]\left[\text{Gaussian noise}\right]\overbrace{\left[\mathbf{M}\right]}\left[\mathbf{S}^{\dagger}\right]$$

(**M** models free surface & source function)


960
960
Recovered **total** data
from ***randomized***
compressive data
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
0.512
0
1000
0
1000
Receiver position (m)
Shot position (m)

SLIM



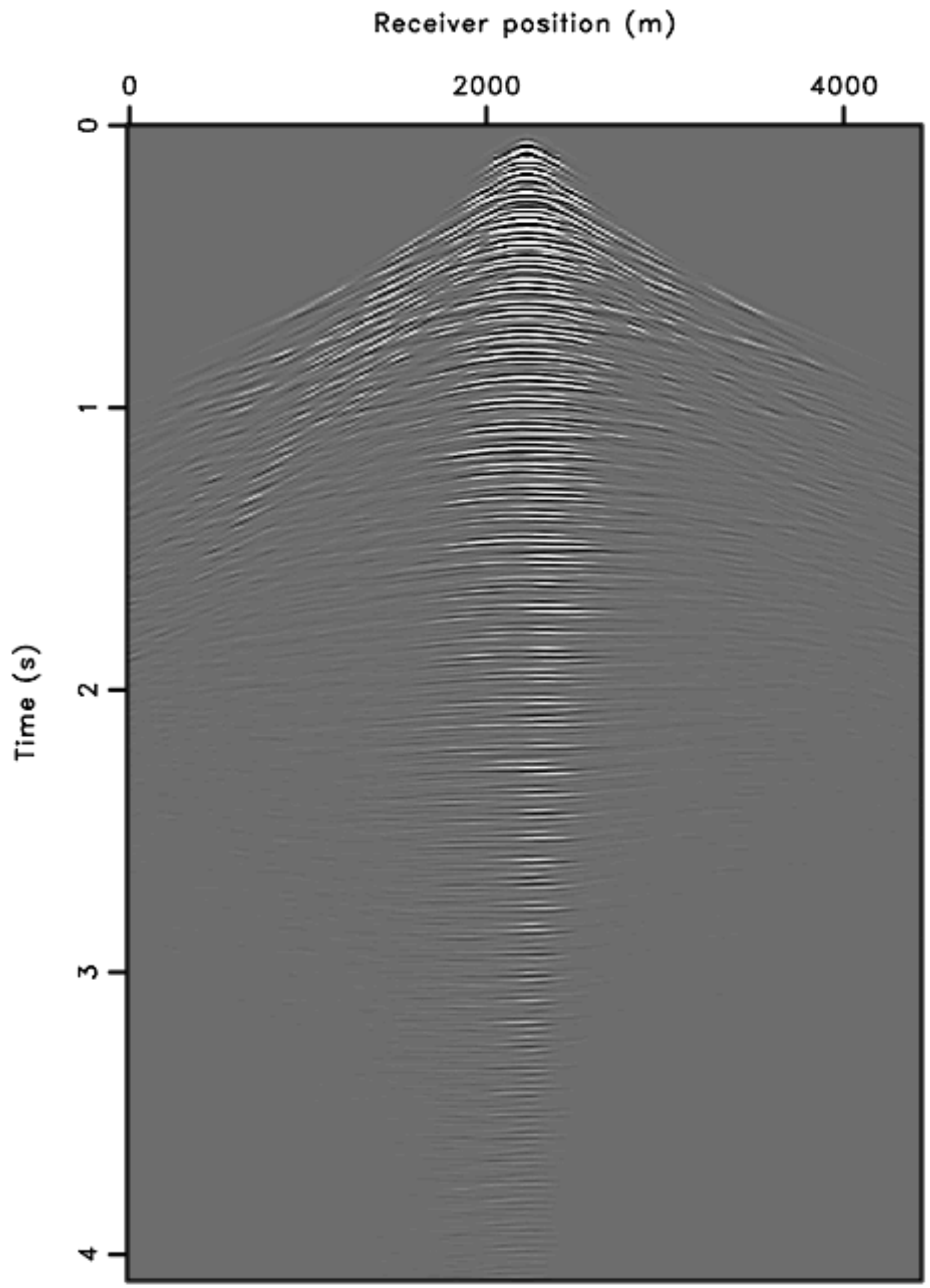
Recent advances

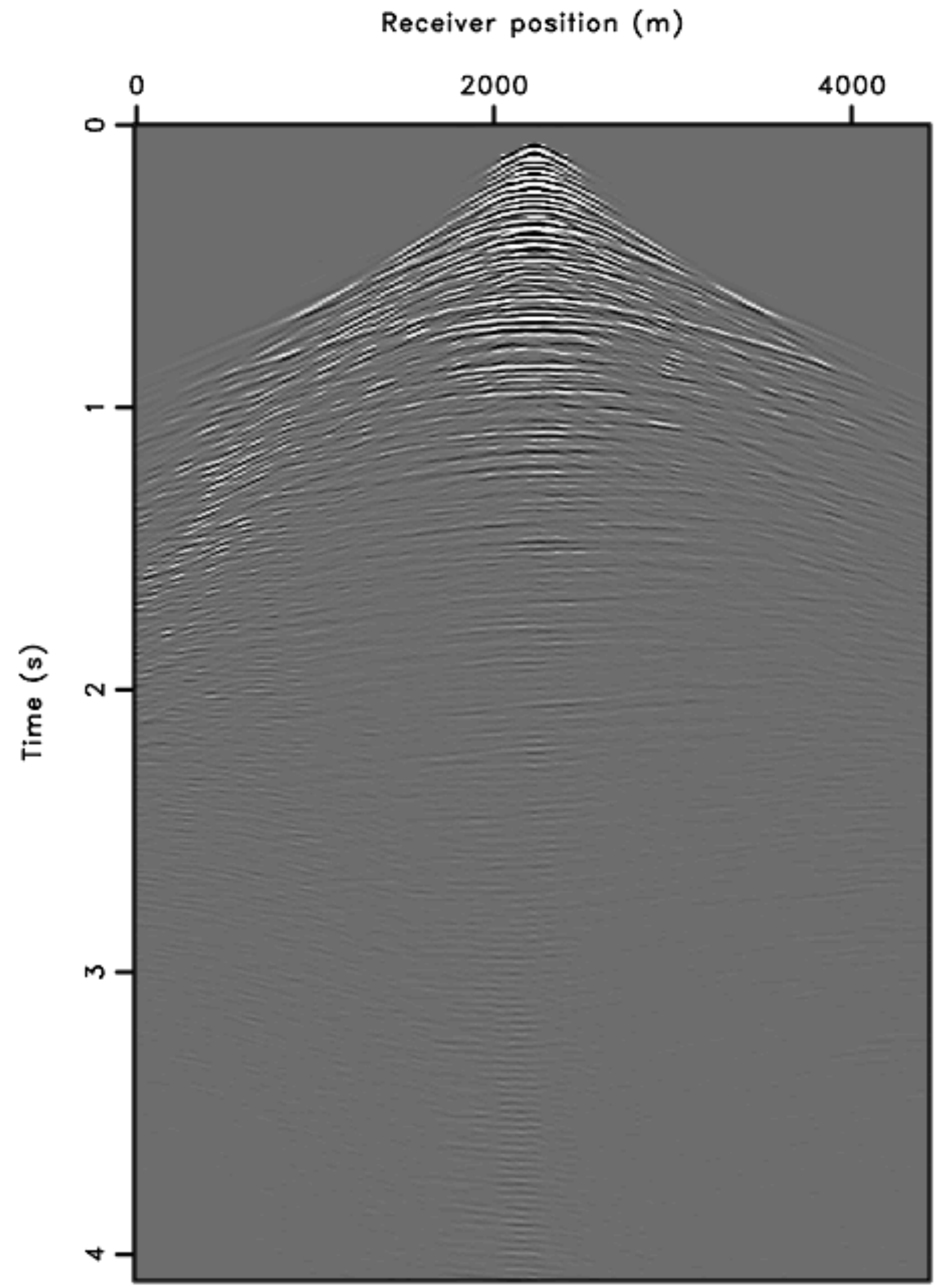
CS applied to wavefield inversion

"Unified compressive sensing framework for simultaneous acquisition with primary estimation"

by Tim Lin and Felix J. Herrmann. Session: SPMUL 2 Multiples II. Room: General Assembly C @ 03:10 PM

Primary estimation





Bottom line

CS explains *improved* recovery

CS leads to *reduction* of data volumes & computational costs

Incorporating *physics* really pays ...

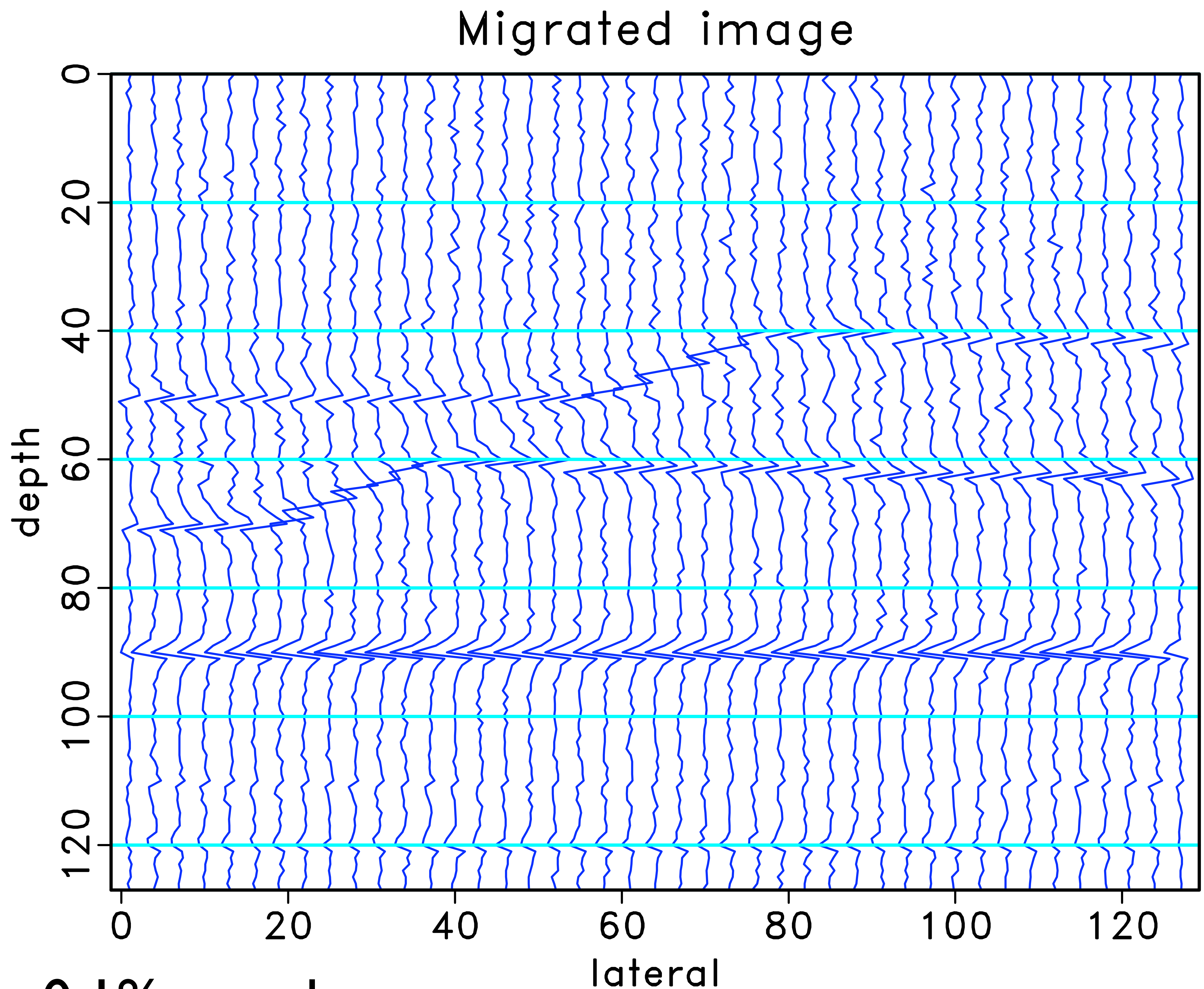
SLIM

Recent advances

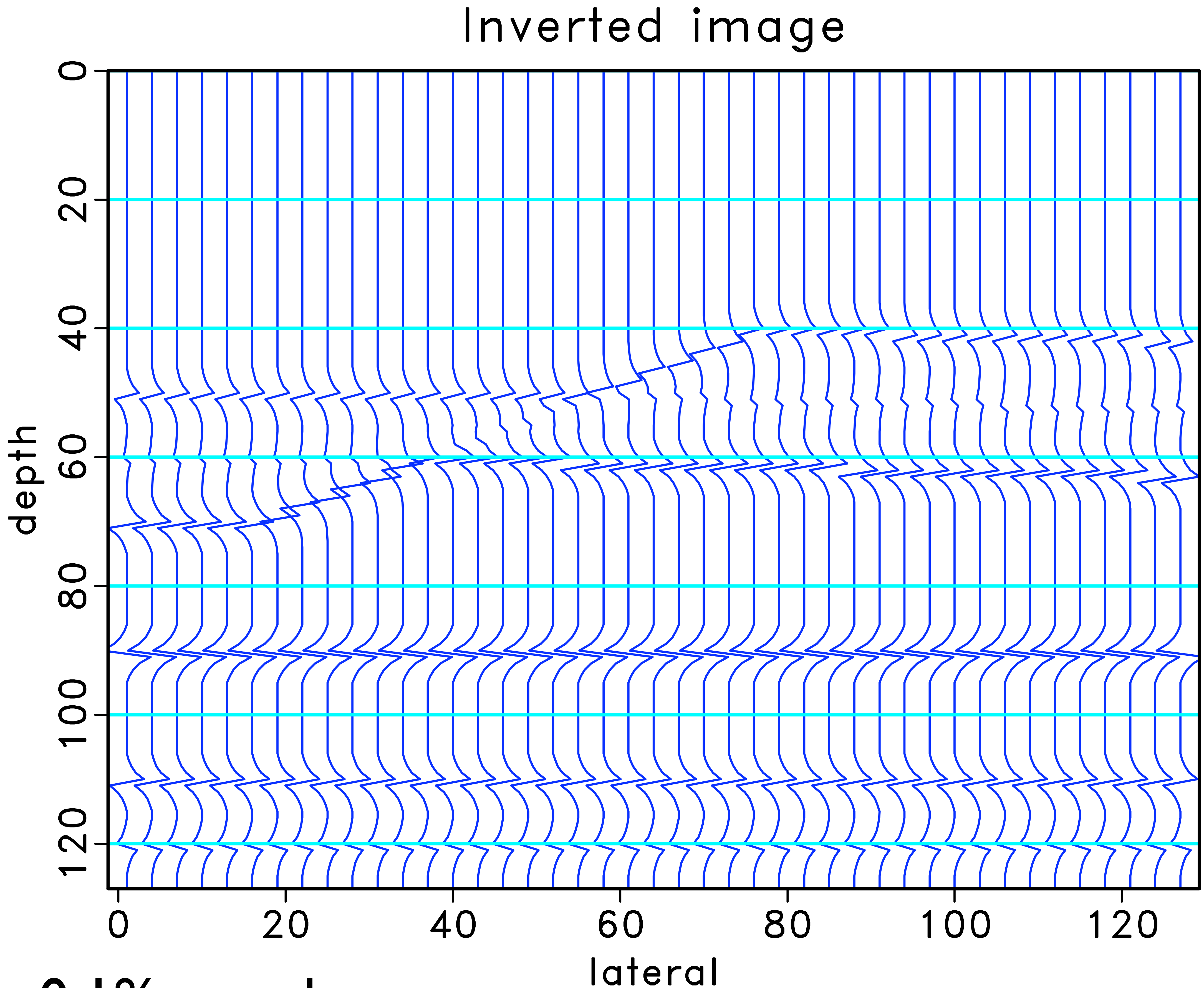
Model-space CS applied to imaging with extensions

"Compressive imaging by wavefield inversion with group sparsity"

by Felix J. Herrmann. SI 3 Methods
Room: 351 F @ 02:45 PM.



From 0.1% samples



From 0.1% samples

Recent advances

CS applied to full-waveform inversion

"Seismic waveform inversion with Gauss-Newton-Krylov method"

by Yogi Erlangga and Felix J Herrmann. SI 3 Methods. Room: 351 F @ 04:25 PM.

Conclusions & outlook

Dimensionality reduction will *revolutionize* our field

- *reduction* of *acquisition* costs
- *decrease* in *processing* time
- *high-resolution* inversions that are otherwise *infeasible* with Nyquist-based methods

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) of Felix J. Herrmann.

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

Thanks to the DELPHI consortium for their hospitality and Eric Verschuur for many fruitful discussions.

Relation to existing work

Simultaneous & continuous 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. ald, '98
- *Reconstruction of band-limited signals, irregularly sampled along one spatial direction* by Duijndam et. al., '99
- *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

Wavefield extrapolation:

- *Compressed wavefield extrapolation* by T. Lin and F.J.H, '07
- *Compressive wave computations* by L. Demanet and G. Peyré ,'08

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.
- *How to choose a subset of frequencies in frequency-domain finite-difference migration* by Mulder & Plessix, '04.
- *Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies* by Sirque & Pratt, '04.
- *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.

SLIM

Thank you for your attention!

more information

slim.eos.ubc.ca