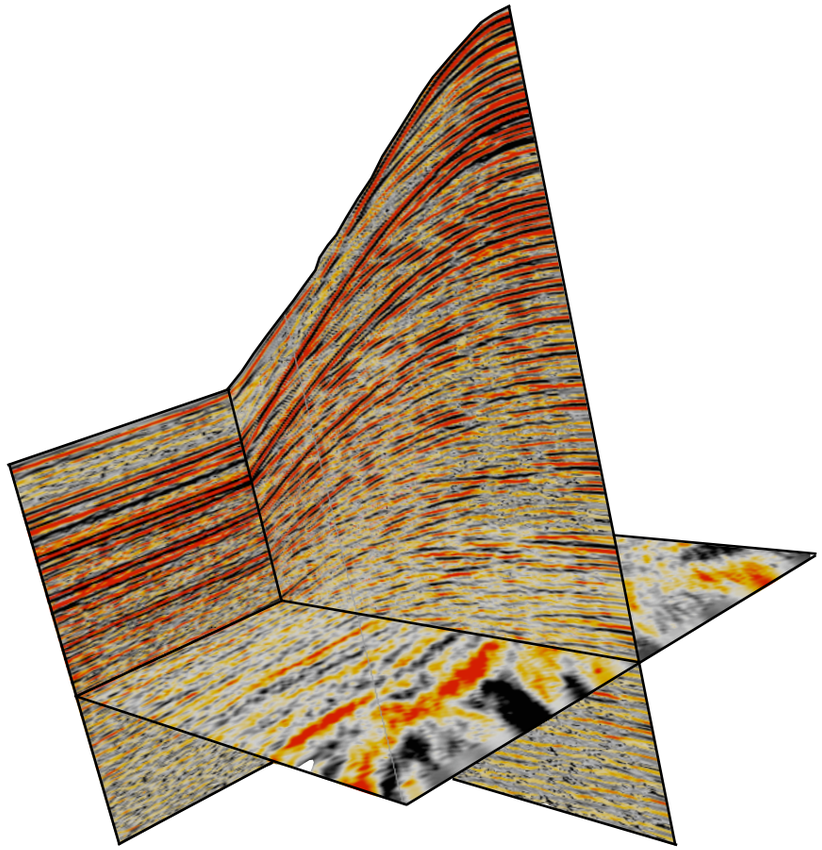
THE UNIVERSITY OF BRITISH COLUMBIA | VANCOUVER

Compressive Sensing Applied to Full-wave Form Inversion

Felix J. Herrmann*
fherrmann@eos.ubc.ca

Joint work with Yogi Erlangga, and Tim Lin

***Seismic Laboratory for Imaging & Modeling**
Department of Earth & Ocean Sciences
The University of British Columbia



EAGE, June 9th, 2009

Seismic imaging & inversion

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

+ BC

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

Major impediments

- "***Curse of dimensionality***"
  - *acquisition* **>>** *processing* & *inversion* **>>** *modeling* **costs** are proportional to the **size** of *data* and *image* space
  - really hits you when having to *model* ***iteratively***

- **Action: size reduction**
  - data volumes
  - Helmholtz system

Curse of dimensionality

- **Seismic data processing, modeling & inversion:**
  - firmly rooted in Nyquist's sampling paradigm for (modeled) wavefields
  - too *pessimistic* for signals with *structure*
  - existence of sparsifying transforms (e.g. curvelets)

- **Solution strategy:**
  - *leverage new* paradigm of *compressive sensing* (CS)
    - identify simultaneous acquisition as CS
    - reduce acquisition, simulation, and inversion costs by ***randomization*** and deliberate ***subsampling***
  - recovery from sample **rates ≈ computational cost** *proportional* to **transform-domain sparsity** of *data* or *model*

- **Remove the "curse of dimensionality" ...**
  - deliberate subsampling followed by nonlinear sparsity-promoting recovery
  - gain when recovery overhead << gain in subsampling

Relation to existing work

- **Simultaneous & continuous acquisition:**
  - *Efficient Seismic Forward Modeling using Simultaneous Random Sources and Sparsity* by N. Neelamani and C. Krohn and J. Krebs and M. Deffenbaugh and J. Romberg, '08
- **Simultaneous simulations & migration:**
  - *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.
- **Imaging:**
  - *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 and Pratt, '04.
- **Full-waveform inversion:**
  - *3D prestack plane-wave, full-waveform inversion* by Vigh and Starr, '08
- **Wavefield extrapolation:**
  - *Compressed wavefield extrapolation* by T. Lin and F.J.H, '07
  - *Compressive wave computations* by L. Demanet (SIAM '08 MS79 & Preprint)

Today's agenda

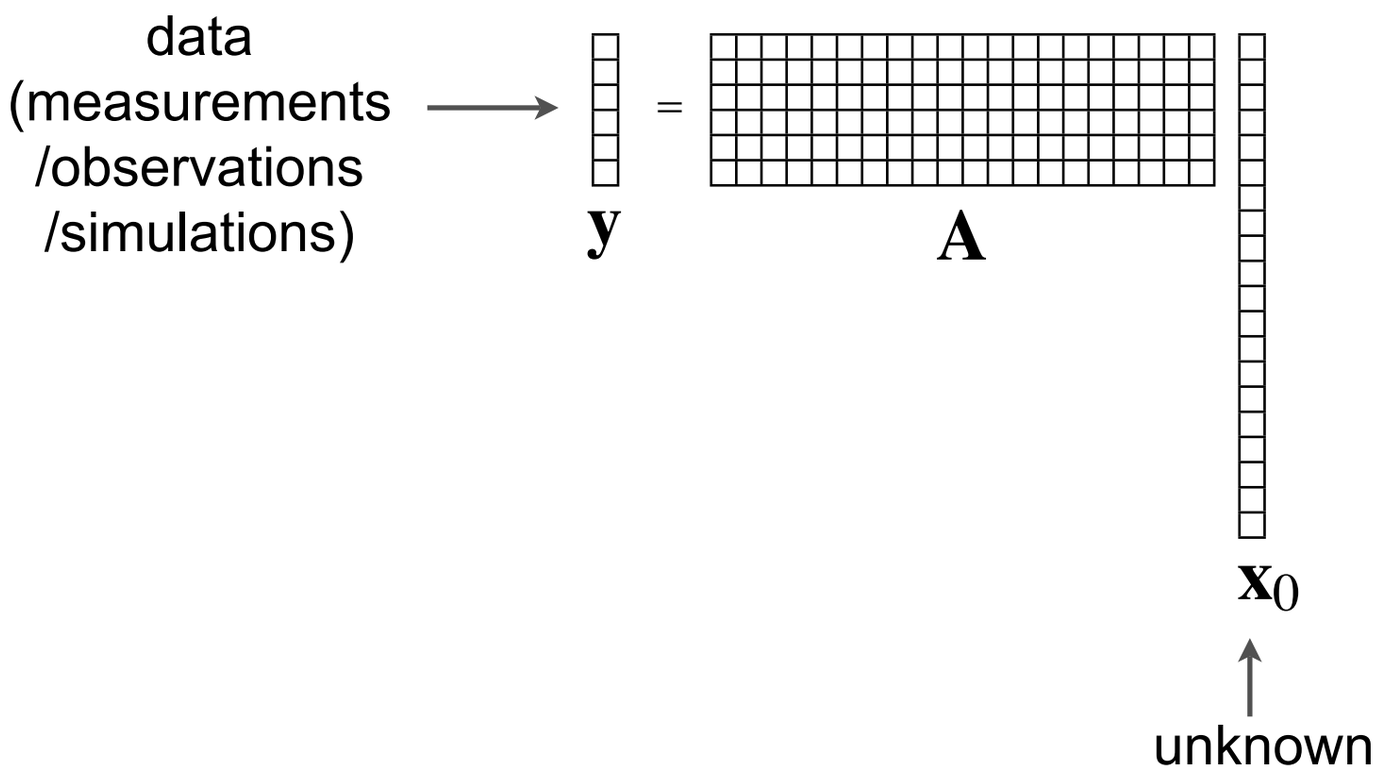
- Brief introduction to *compressive sensing*
  - ***sparsifying*** transforms
  - ***randomized* = *incoherent* downsampling**
  - ***nonlinear*** recovery by ***sparsity*** promotion

- CS applied to ***implicit*** simultaneous full-waveform simulation
  - use ***simultaneous sources*** as the ***sampling*** domain
  - reduction of the number of right-hand sides & frequencies

- Extensions
  - integration into formulation for full-waveform inversion

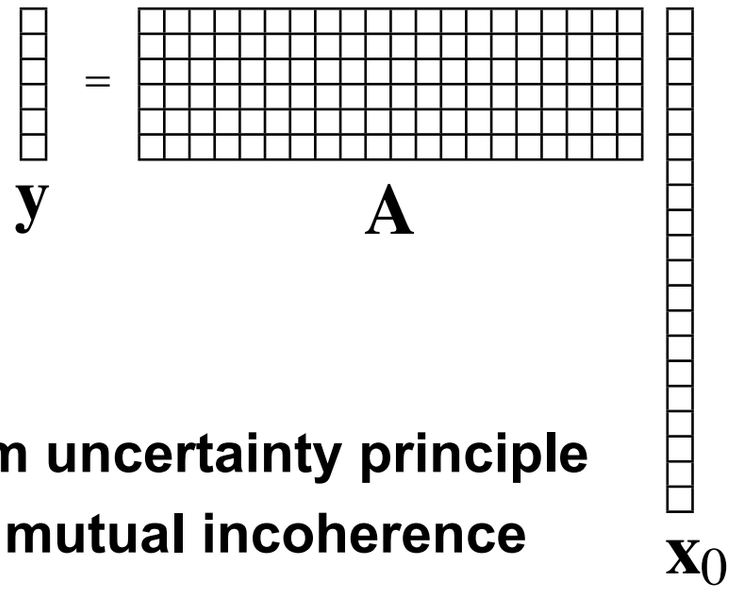
Problem statement

Consider the following (severely) underdetermined system of linear equations



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

Perfect recovery



- conditions
  - **A** obeys the **uniform uncertainty principle**
  - ***randomized* A <=> mutual incoherence**
  - $\mathbf{x}_0$ is **sufficiently sparse**

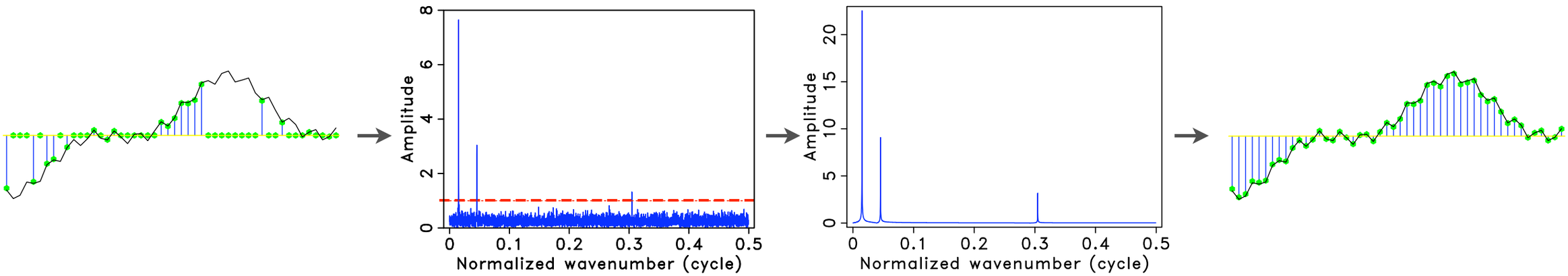
- ***nonlinear*** recovery procedure:

$$\underbrace{\min_{\mathbf{x}} \|\mathbf{x}\|_1}_{\text{sparsity}} \quad \text{s.t.} \quad \underbrace{\mathbf{A}\mathbf{x} = \mathbf{y}}_{\text{perfect reconstruction}}$$

- performance
  - *S*-sparse vectors recovered from roughly on the order of *S* measurements (to within constant and *log* factors)

[Candès et al.'06]
[Donoho'06]

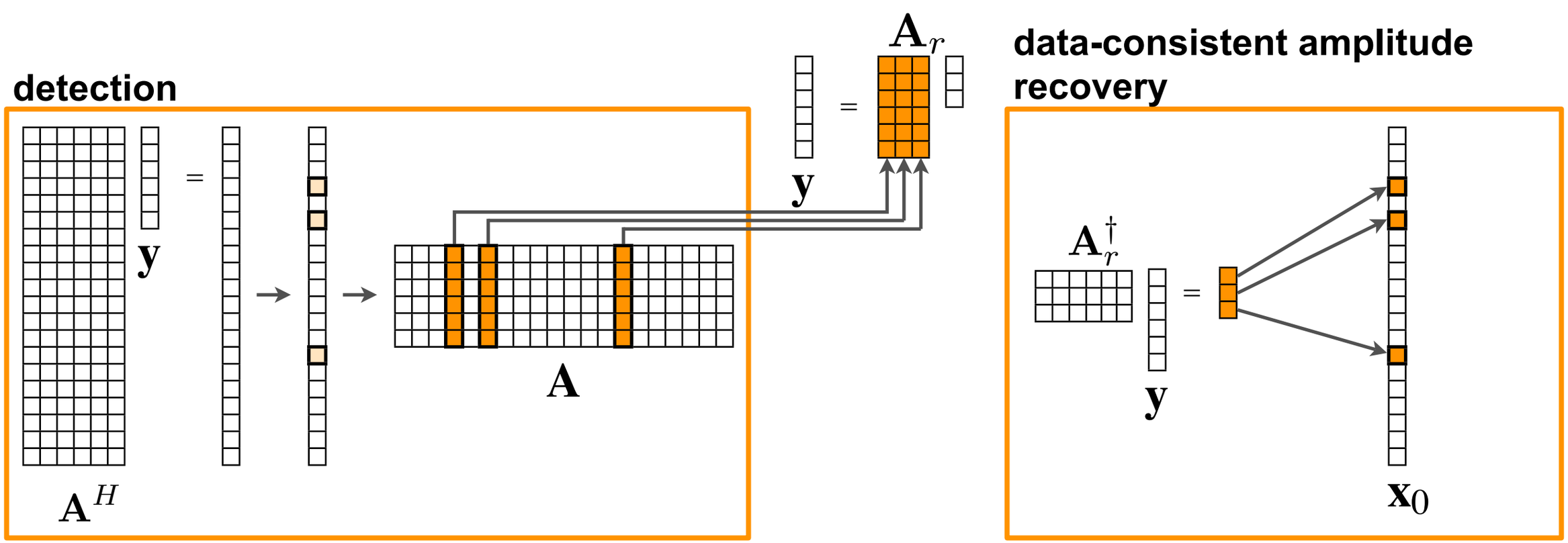
NAIVE sparsity-promoting recovery



inverse Fourier transform

detection + data-consistent amplitude recovery

Fourier transform



Extensions

- Use CS principles to select *physically* appropriate
  - *measurement* basis **M = *random* phase encoder**
  - ***randomized*** *restriction* matrix **R = downsampler**

  } **"blending"**
  - sparsifying transform **S** (e.g. curvelets)
  - driven by signal type, physics, and type of acquisition (e.g. fMRI vs seismic)
- Sparse signal representation:

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0$$

with

$$\mathbf{A} = \mathbf{R}\mathbf{M}\mathbf{S}^H$$

restriction matrix **measurement matrix** **sparsity matrix**

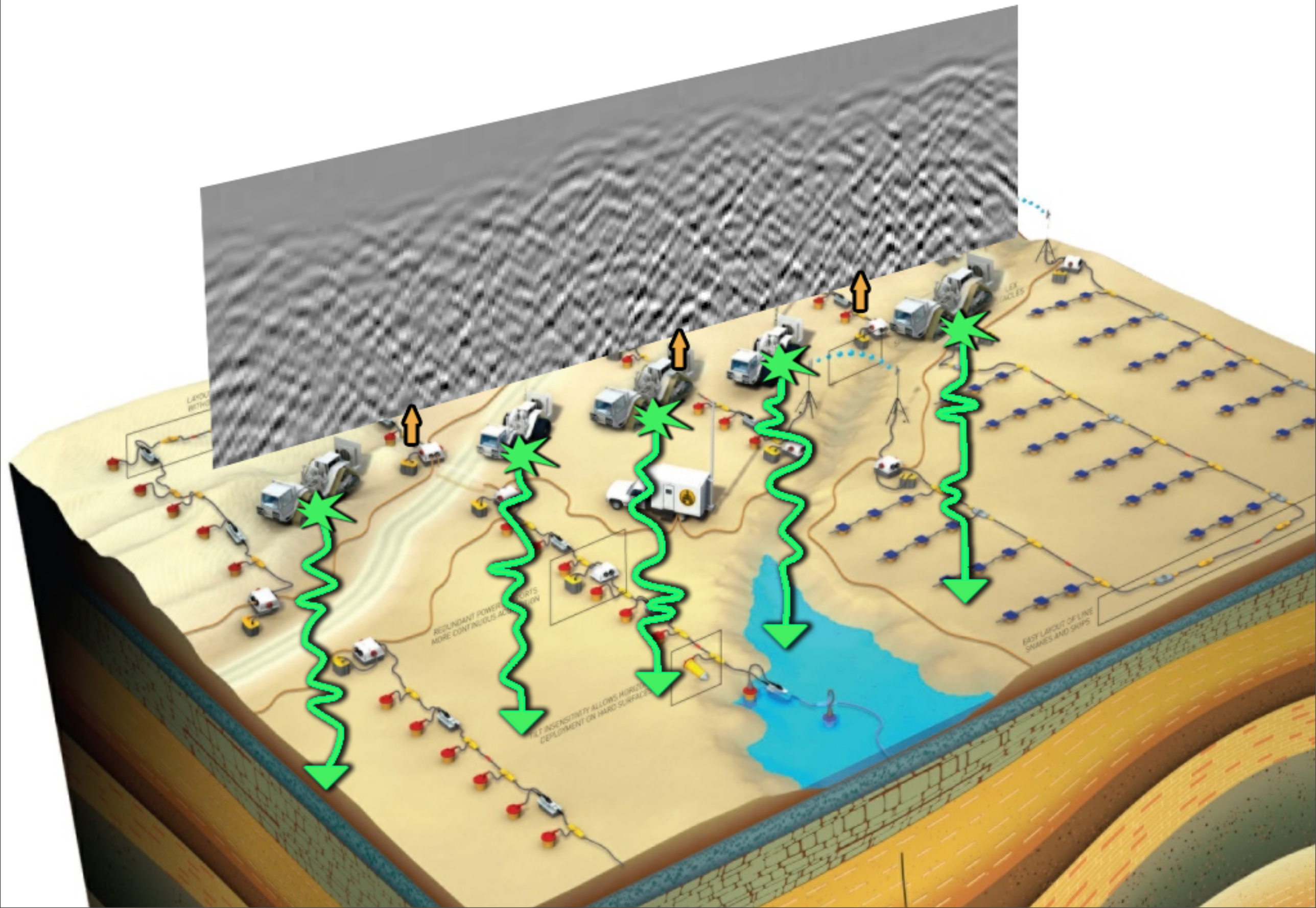
Selection is aimed at turning *aliases/coherent subsampling artifacts* into harmless ***noise*** ...

Wavefield computations

$$\overbrace{\begin{bmatrix} \mathcal{H}_{\omega_1} & 0 & & \\ 0 & \mathcal{H}_{\omega_2} & \ddots & \\ & \ddots & \ddots & 0 \\ & & 0 & \mathcal{H}_{\omega_{n_f}} \end{bmatrix}}^{\mathbf{H}} \overbrace{\begin{bmatrix} \mathbf{U}_{\omega_1} \\ \mathbf{U}_{\omega_1} \\ \vdots \\ \mathbf{U}_{n_f} \end{bmatrix}}^{\mathbf{U}} = \overbrace{\begin{bmatrix} \mathbf{B}_{\omega_1} \\ \mathbf{B}_{\omega_2} \\ \vdots \\ \mathbf{B}_{n_f} \end{bmatrix}}^{\mathbf{B}} \quad \Rightarrow \quad \mathbf{HU} = \mathbf{B}$$

- Matrix-free preconditioned indirect solver based on multilevel Krylov with deflation [Erlanga, Nabben, '08, Erlanga and F.J.H, '08]
- Solution gives multidimensional wavefield $\mathbf{u}(x_s, x_r, t)$
- Block-diagonal structure **H** and multiple rhs are amenable to CS as long as CS sampling matrix ***commutes*** with **H**
- Corresponds to simultaneous acquisition
  - replaces *impulsive* individual sources by *simultaneous* randomized sources
  - reduces number *simultaneous* sources (rhs) & *angular* frequencies (blocks)

Simultaneous & continuous sources



Sparse recovery

$$\mathbf{P_1}: \quad \begin{cases} \mathbf{y} & = \mathbf{RMd} \\ \tilde{\mathbf{x}} & = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{y} \\ \mathbf{A} & = \mathbf{RMS}^* \\ \tilde{\mathbf{d}} & = \mathbf{S}^*\tilde{\mathbf{x}} \end{cases}$$

Challenges:

- large to extreme large system size (number of unknowns is 2^{25} for a really small problem)
- find proper subsampling matrix that is physically realizable and numerically fast
- find proper sparsifying transforms that balances ***sparsity*** with **mutual coherence**

Solver:

- bring in as many entries per iteration as possible
- projected gradient with root finding method ($\mathrm{SPG}\ell_1$, Friedlander & van den Berg, '07-'08)
- few matrix-vector multiplies
- use matrix-free implementations where possible

CS sampling matrix

Subsample along source and frequency coordinates

Use ***fast*** transform-based sampling algorithms such as ***scrambled Fourier*** [Romberg, '08] or ***Hadamard*** ensembles [Gan et. al., '08]

$$\mathbf{RM} = \overbrace{\begin{bmatrix} \mathbf{R}_1^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{R}_1^{\Omega} \\ \vdots \\ \mathbf{R}_{n_{s'}}^{\Sigma} \otimes \mathbf{I} \otimes \mathbf{R}_{n_{s'}}^{\Omega} \end{bmatrix}}^{\text{sub sampler}} \overbrace{\left(\mathbf{F}_2^* \operatorname{diag}\left(e^{\hat{i}\boldsymbol{\theta}}\right) \otimes \mathbf{I}\right) \mathbf{F}_3}^{\text{random phase encoder}},$$

$$\theta_w = \mathrm{Uniform}([0, 2\pi])$$

- Different random restriction for each $n_s' \ll n_s$ simultaneous experiments
- Restriction reduces system size
- Different from implementations of sampling matrices based on Kronecker-products
- Numerical complexity CS sampling

$$\boxed{\mathcal{O}(n^3 \log n)}$$

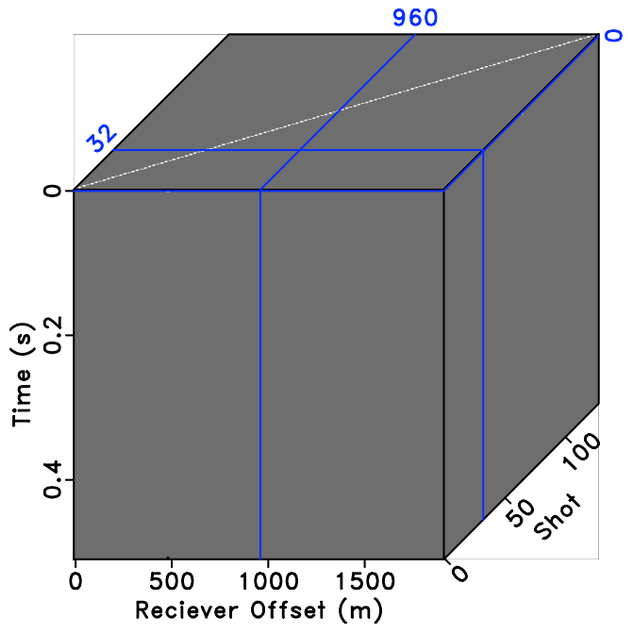
Source-solution sampling equivalence

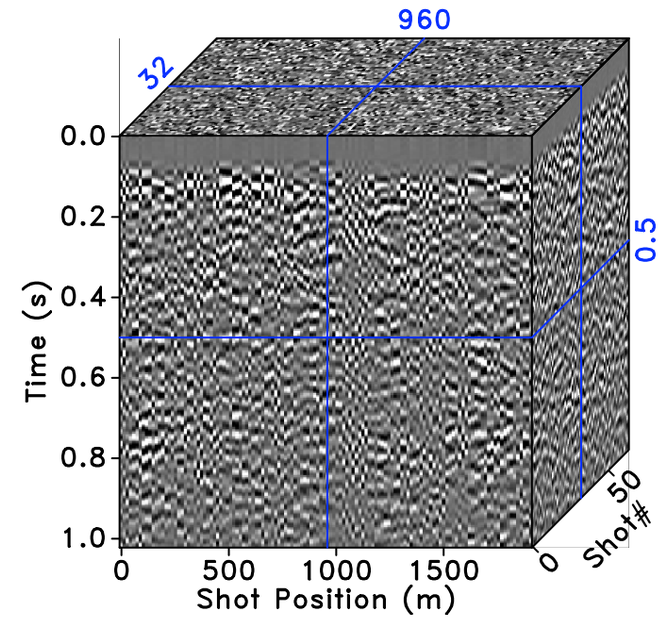
$$\begin{cases} \mathbf{B} = \mathbf{D}^* \quad \underbrace{\mathbf{s}}_{\text{single shots}} \\ \mathbf{HU} = \mathbf{B} \\ \mathbf{y} = \mathbf{RMDU} \end{cases} \iff \begin{cases} \underline{\mathbf{B}} = \underline{\mathbf{D}}^* \quad \underbrace{\mathbf{RMs}}_{\text{simul. shots}} \\ \underline{\mathbf{HU}} = \underline{\mathbf{B}} \\ \underline{\mathbf{y}} = \underline{\mathbf{DU}} \end{cases}$$

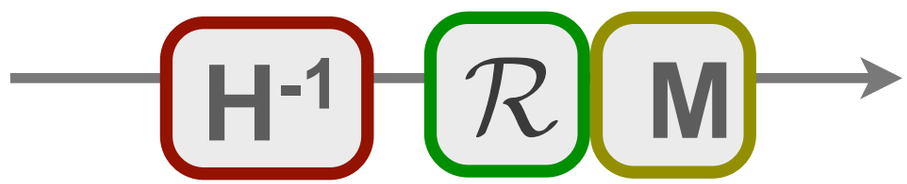
- Show equivalence between
  - CS sampling the ***full*** solution for separate single-source experiments
  - Solution of ***reduced*** system after CS sampling the collective single-shot source wavefield **s**

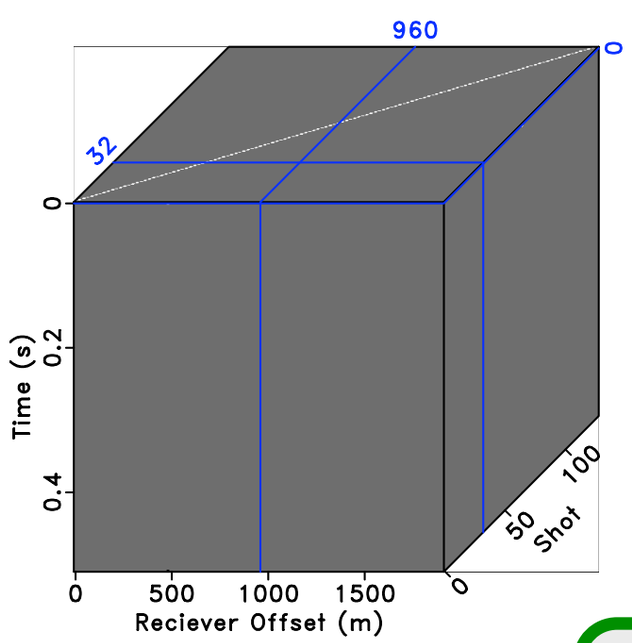
- Have to show that

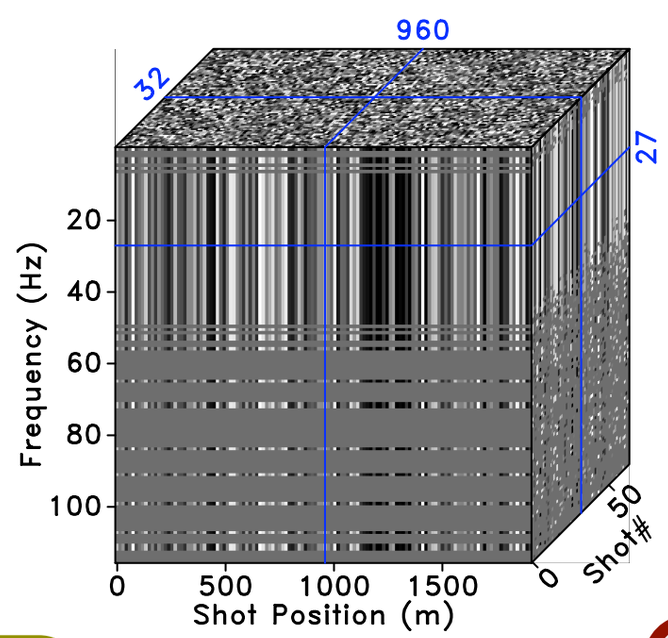
$$\boxed{\mathbf{y} = \underline{\mathbf{y}}}$$

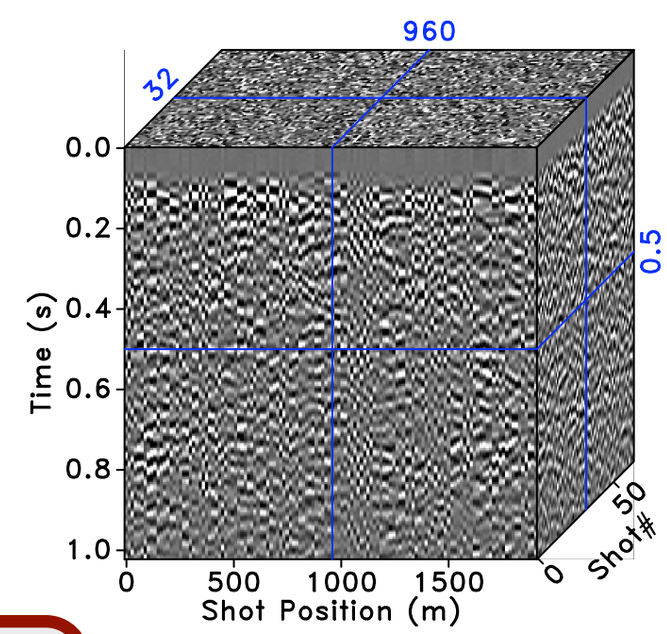

960
32
0
Time (s)
0
0.2
0.4
0
500
1000
1500
Reciever Offset (m)
0
50
100
Shot


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


H-1
ℛ
M


960
32
0
Time (s)
0
0.2
0.4
0
500
1000
1500
Reciever Offset (m)
0
50
100
Shot


960
32
27
Frequency (Hz)
20
40
60
80
100
0
500
1000
1500
Shot Position (m)
0
50
Shot#


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


ℛ
M


H-1

Sparsifying transform

- Use fast discrete 2-D Curvelet transform based on wrapping [Demanet '06] along shot and receiver coordinates
  - compresses highly geometrical features of monochromatic wavefields
  - incoherent with compressive-sampling matrix that acts along the source coordinate

- Use fast discrete wavelet transform along the time coordinate
  - compresses front-like features arriving along the time direction
  - reasonable incoherent with sampling of angular frequencies

- Combine both transforms through a **Kronecker** product

$$\mathbf{S} = \mathbf{C}_{2d} \otimes \mathbf{W}$$

- Numerical complexity *sparsifying* transform

$$\boxed{\mathcal{O}(n^3 \log n)}$$

Complexity analysis

Assume discretization size in each dimension is *n*, and

$$n_s = n_t = n_f = \mathcal{O}(n)$$

Time-domain finite differences:

- $\mathcal{O}(n^4)$ in 2-D
- large constants

Multilevel-Krylov preconditioned [Erlangga, Nabben, FJH, '08]

- $\mathcal{O}(n^4) = n_f n_s n_{it} \mathcal{O}(n^2)$ with $n_{it} = \mathcal{O}(1)$
- small constants

Complexity analysis cont'd

Cost sparsity promoting optimization dominated by matrix-vector products

- Sparsity transform is $\mathcal{O}(n^3 \log n)$
- Gaussian projection $\mathcal{O}(n^3)$ per frequency
- **Cost** $\mathcal{O}(n^4)$, which does not lead to asymptotic improvement

Use fast transforms (e.g. Random Convolutions by Romberg '08)

- fast projection in time & shot directions: $\mathcal{O}(n \log n)$
- **Cost** $\mathcal{O}(n^3 \log n)$ instead of $\mathcal{O}(n^4)$

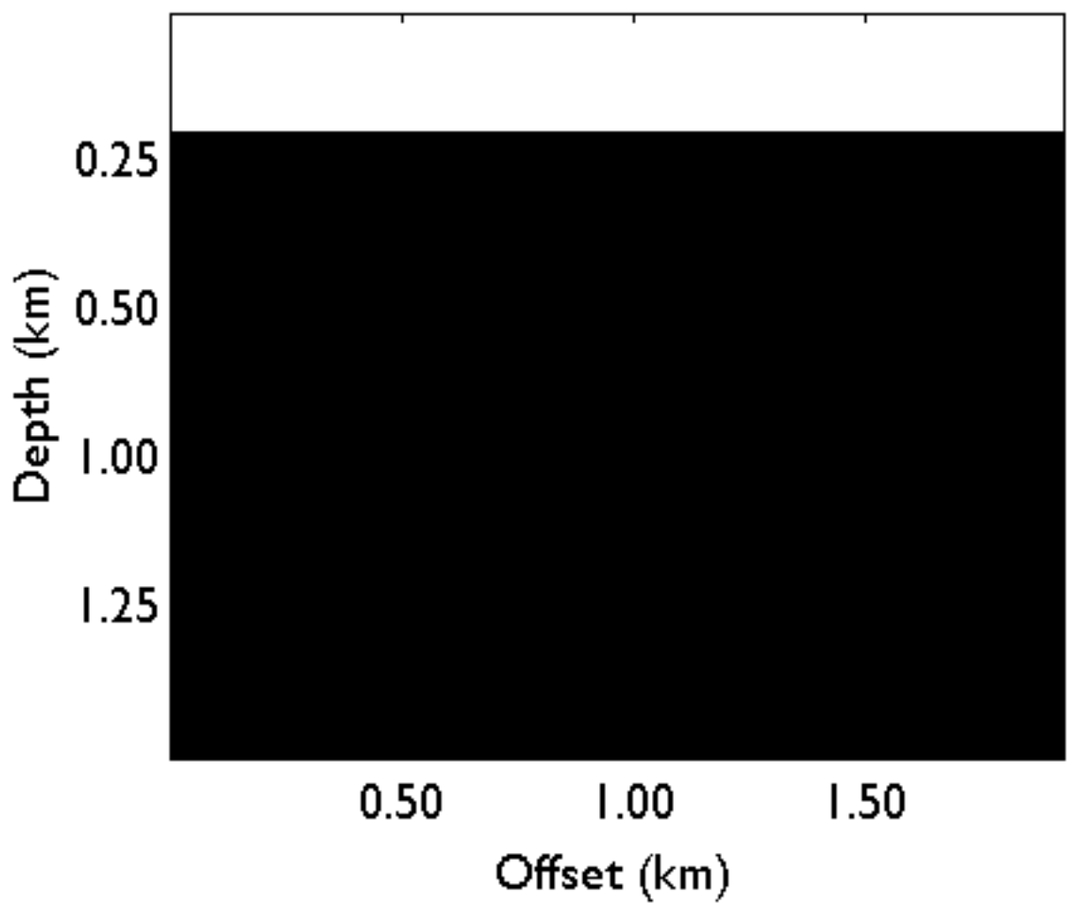
Bottom line: Computational cost for the ℓ_1-solver is less ($\mathcal{O}(n^3 \log n)$ vs. $\mathcal{O}(n^4)$) than the cost of solving Helmholtz

- smaller memory imprint
- cost reduction dependent on complexity = transform-domain sparsity of the solution

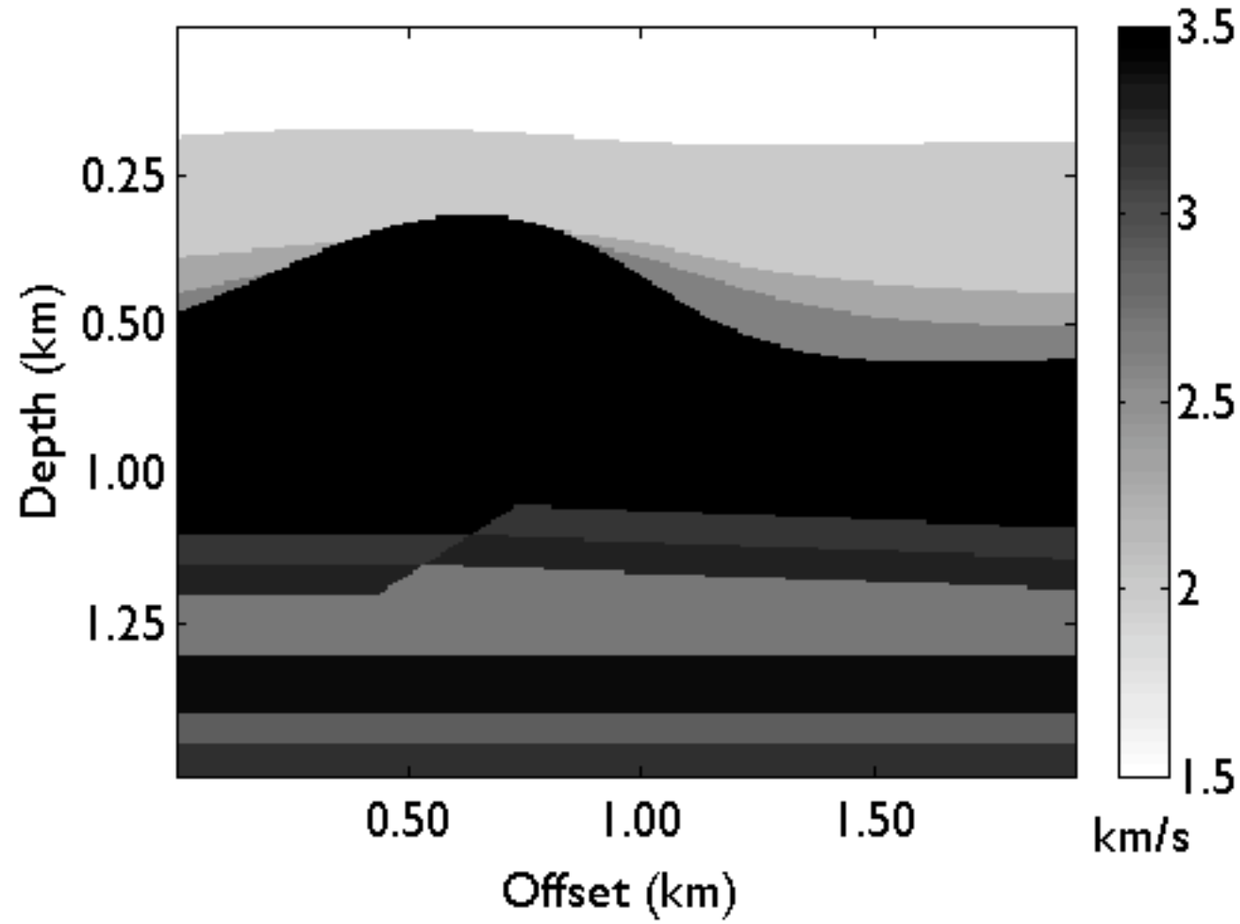
SLIM
Seismic Laboratory for Imaging and Modeling

Velocity models

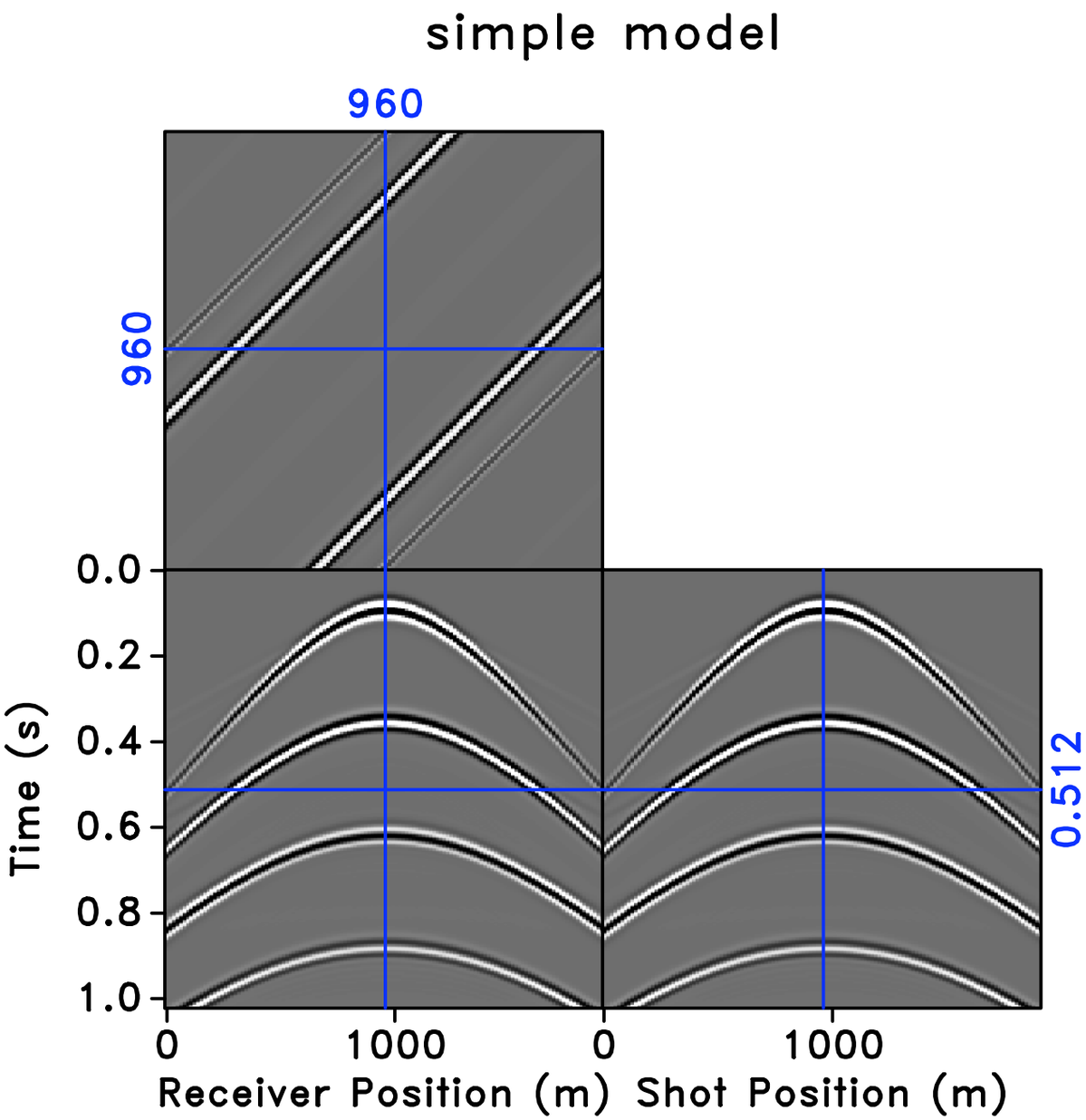
simple model

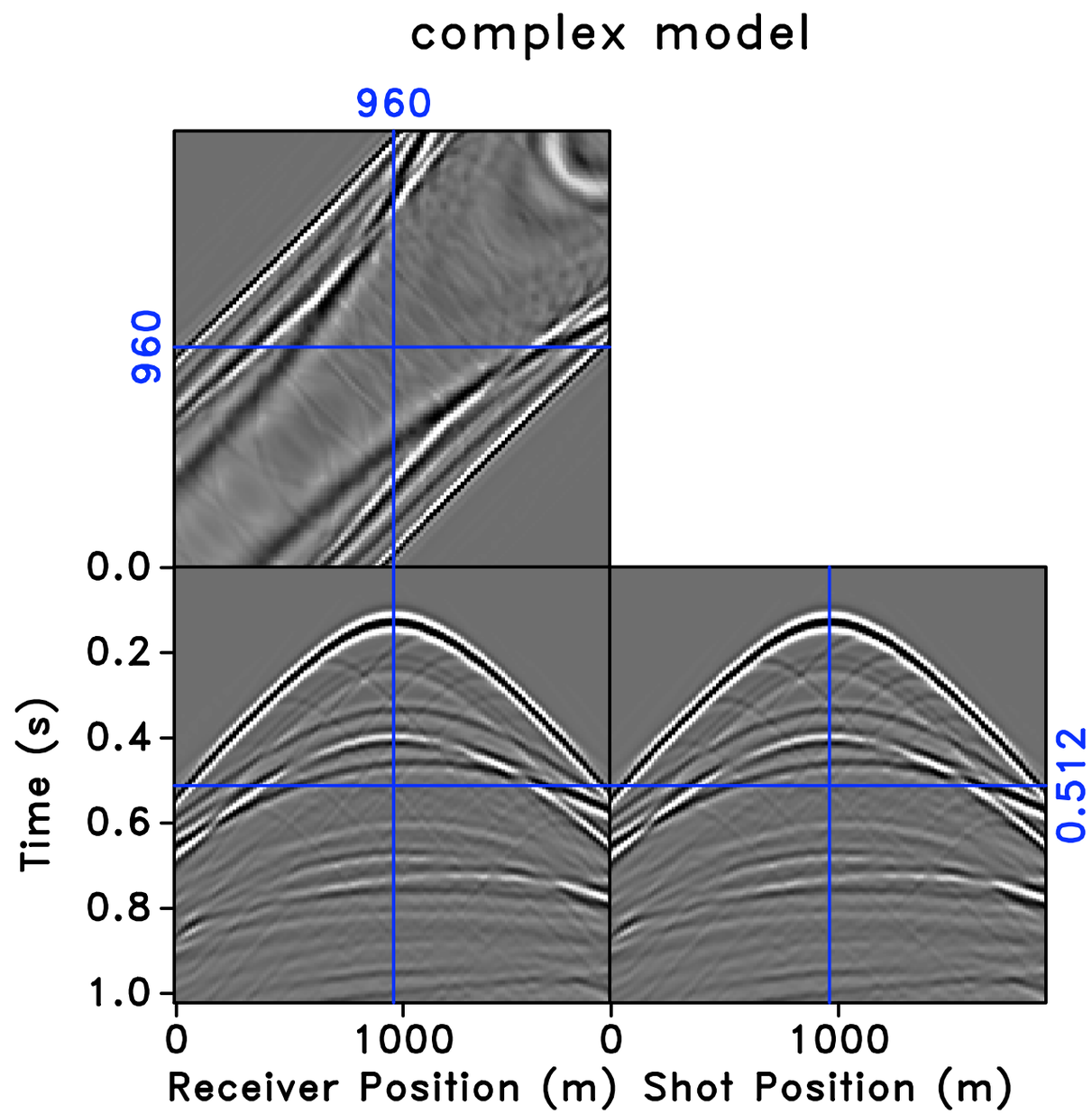


complex model



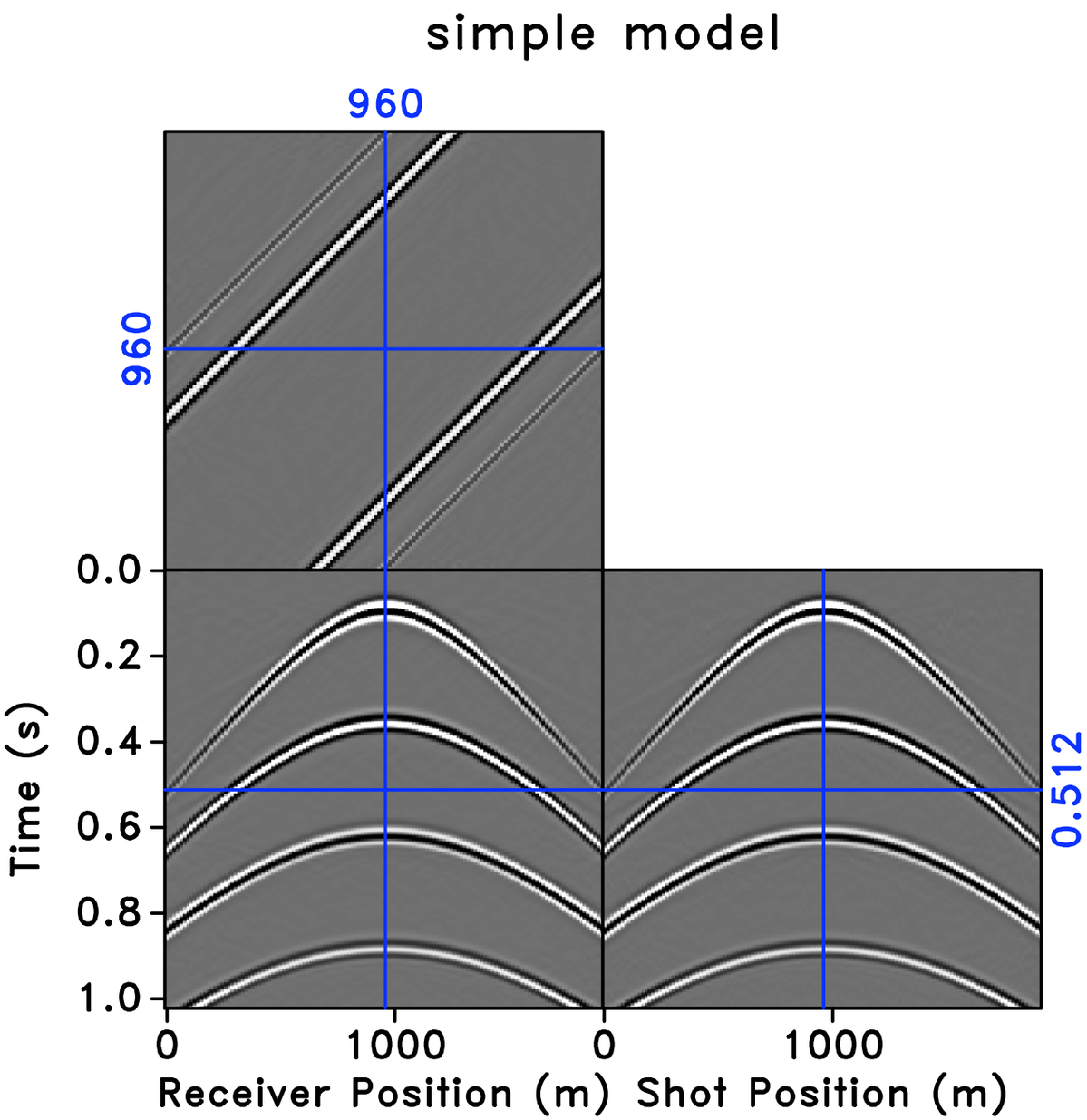
Green's functions



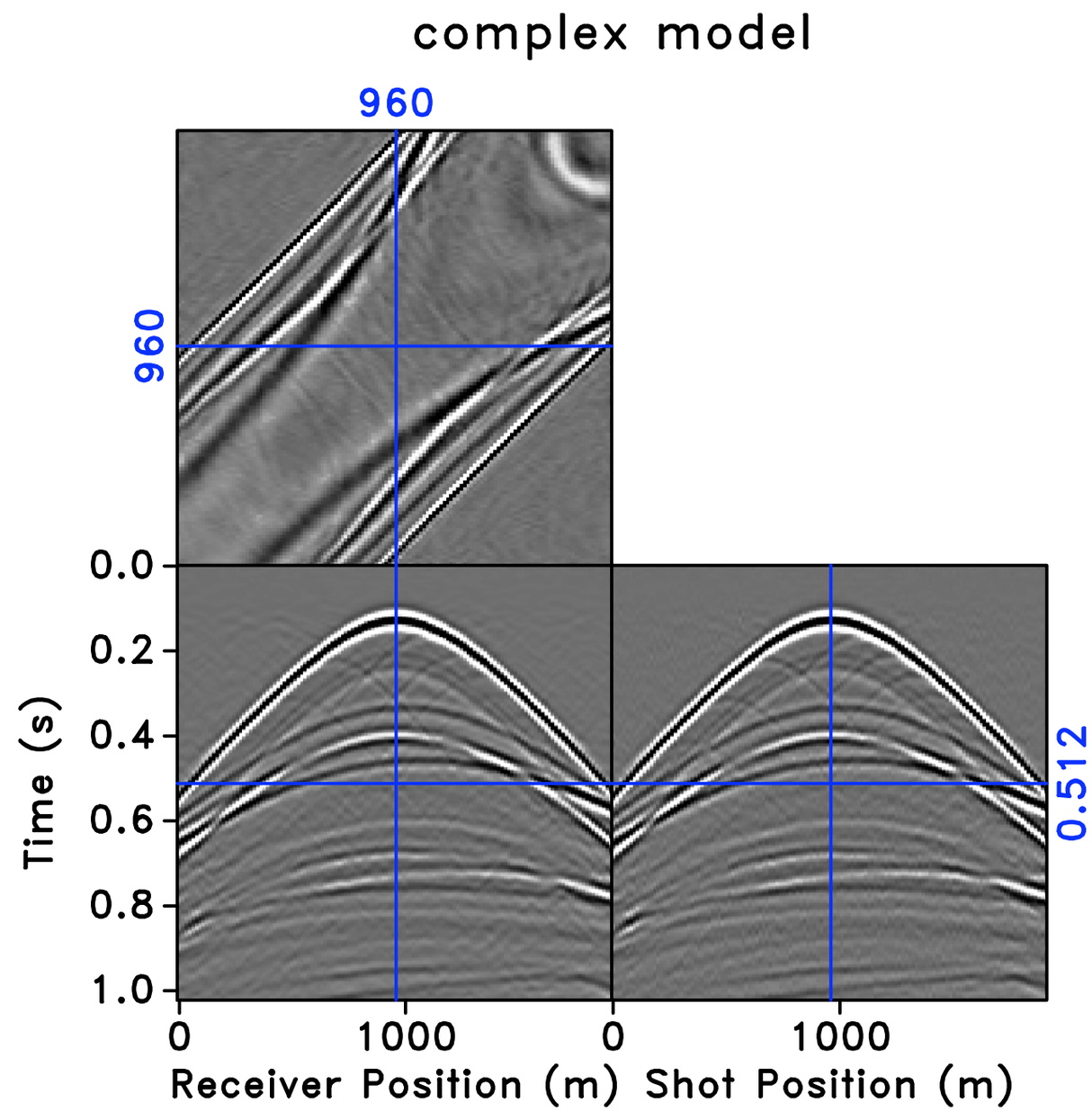


SLIM
Seismic Laboratory for Imaging and Modeling

Recovered data



28.1dB

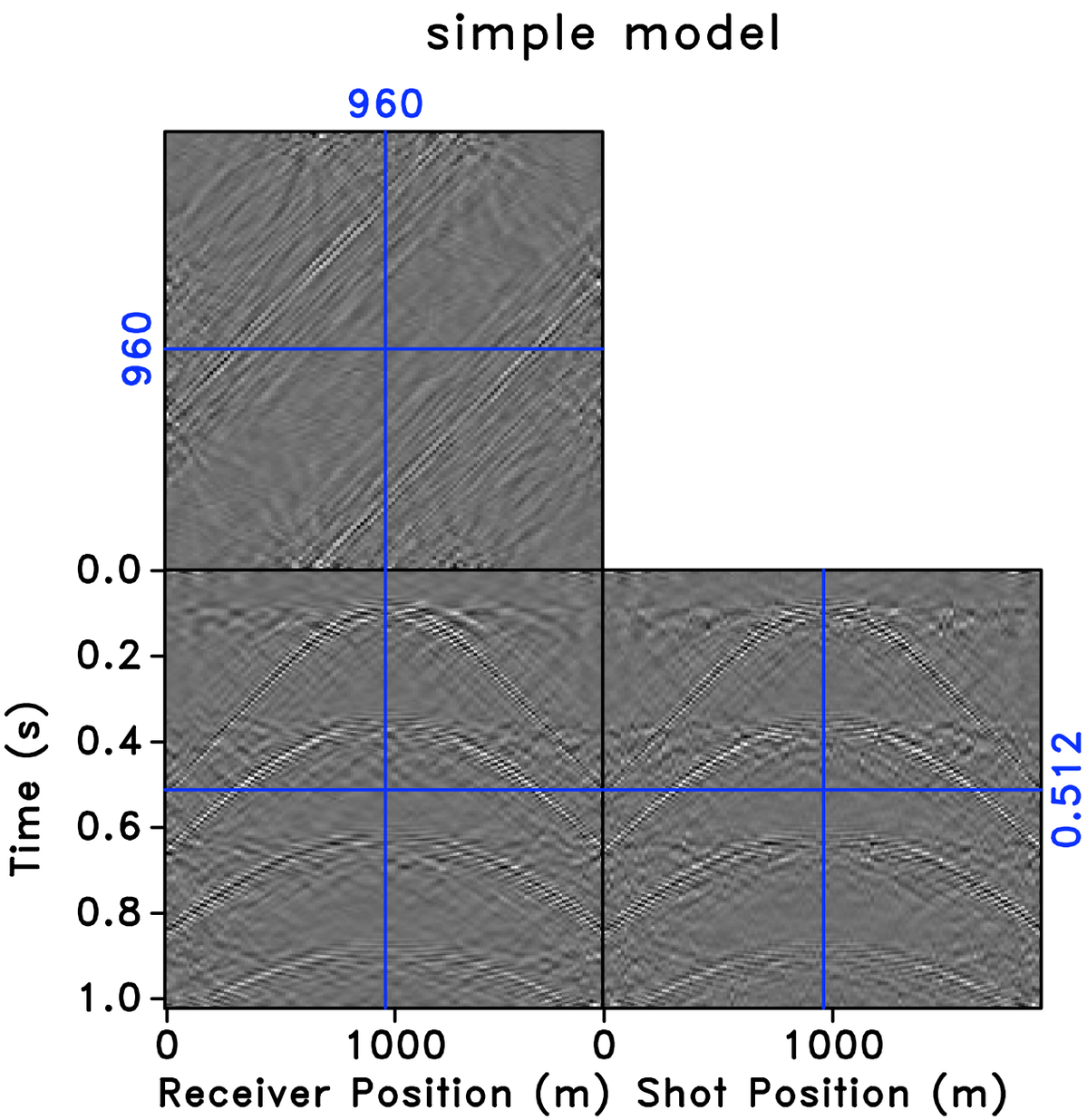


18.2dB

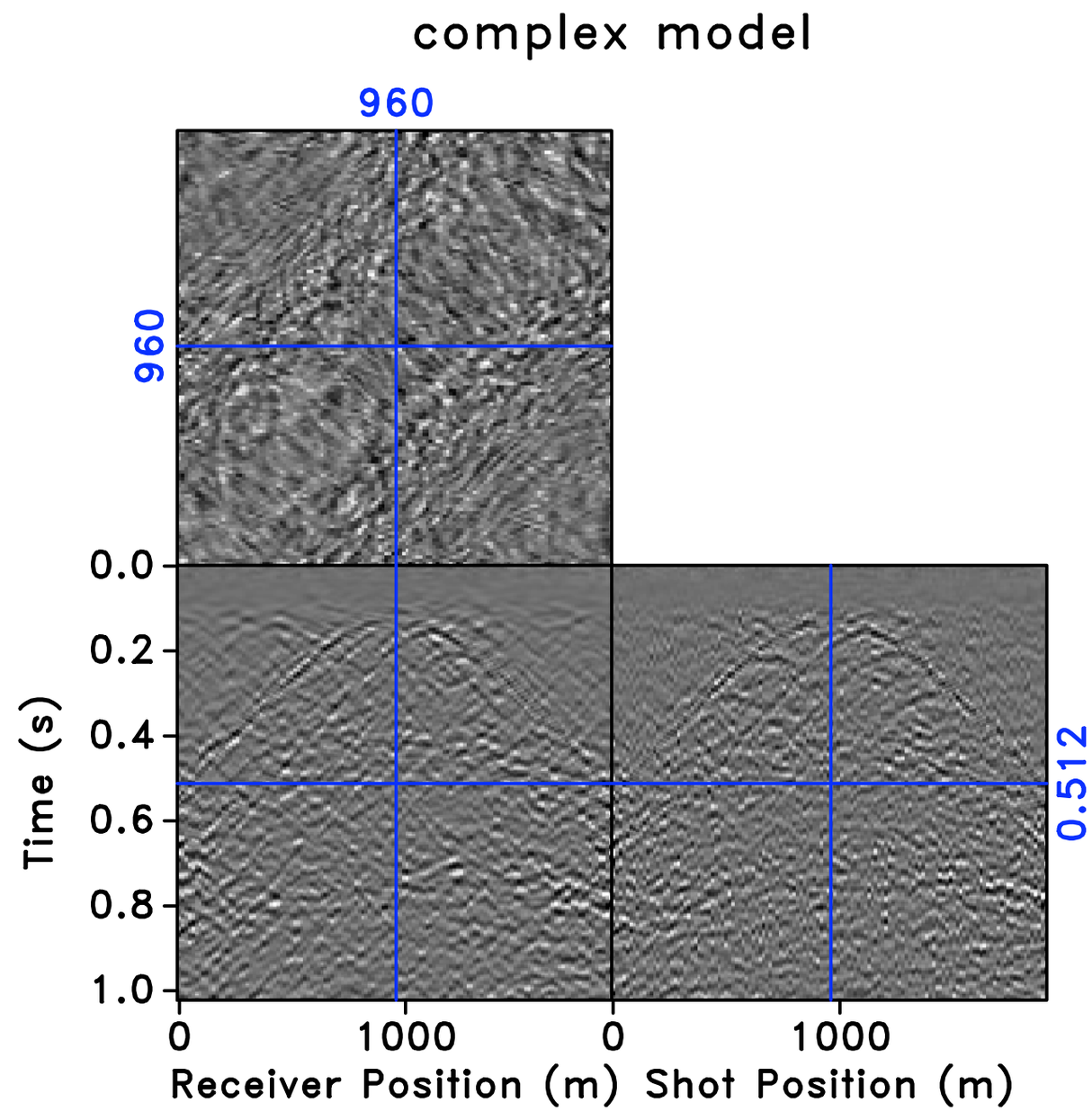
300 SPGL1 iteration

SLIM
Seismic Laboratory for Imaging and Modeling

Difference



28.1dB



18.2dB

300 SPGL1 iteration

SLIM
Seismic Laboratory for Imaging and Modeling

Sample ratio SNR (dB)

problem size 2^{22}

Total computed data fraction

# Frequencies / # Shots	0.25	0.15	0.07
2	14.3	12.1	8.6
1	18.2	14.5	10.2
0.5	22.2	16.5	10.7

$$\mathrm{SNR} = -20 \log \frac{\|\mathbf{d} - \tilde{\mathbf{d}}\|_2}{\|\mathbf{d}\|_2}$$

SLIM
Seismic Laboratory for Imaging and Modeling

Integration into adjoint-state method

- Reduce system size and add sparsity promoting ***prior***

$$\min_{\underline{\mathbf{U}}\in\underline{\boldsymbol{\mathcal{U}}},\,\mathbf{x}\in\boldsymbol{\mathcal{X}}} \frac{1}{2}\|\mathbf{y}-\underline{\mathbf{DU}}\|_2^2 \quad \text{subject to} \quad \underline{\mathbf{H}}[\mathbf{S}^H\mathbf{x}]\underline{\mathbf{U}} = \underline{\mathbf{Q}} \quad \wedge \quad \|\mathbf{x}\|_1 \leq \tau$$

- Recast into *unconstrained* optimization problem:

$$\min_{\mathbf{x}\in\boldsymbol{\mathcal{X}}} \frac{1}{2}\|\mathbf{y}-\underline{\boldsymbol{\mathcal{F}}}[\mathbf{x}]\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau$$

with

$$\underline{\boldsymbol{\mathcal{F}}}[\mathbf{x}] = \underline{\mathbf{DH}}^{-1}[\mathbf{S}^H\mathbf{x}]\underline{\mathbf{Q}}$$

- Requires extension of projected gradient ℓ_1-solver (SPGℓ_1) to nonlinear forward map ...

Observations

- **CS** provides a **new** *linear sampling paradigm* based on ***randomization***
  - ***reduces*** *data* volumes and hence *acquisition*, *processing* & *inversion* **costs**
  - **linearity** allows for compressive processing & inversion

- **CS** leads to
  - "acquisition" of *smaller* data volumes that carry the ***same*** *information*
  - ***improved*** *inferences* from data using the *same* resources

- **Bottom line: acquisition & processing & inversion costs** are **no** longer determined by the ***size*** of the ***discretization*** but by ***transform-domain sparsity*** of the **solution ...**

Acknowledgments

- E. van den Berg and M. P. Friedlander for *SPGL1* (www.cs.ubc.ca/labs/scl/spgl1) & *Sparco* (www.cs.ubc.ca/labs/scl/sparco)
- Sergey Fomel and Yang Liu for Madagascar (rsf.sf.net)
- E. Candes and the Curvelab team

This work was carried out as part of the Collaborative Research & Development (CRD) grant DNOISE (334810-05) funded by the Natural Science and Engineering Research Council (NSERC) and matching contributions from BG, BP, Chevron, ExxonMobil and Shell. FJH would also like to thank the Technische University for their hospitality.

slim.eos.ubc.ca

and... Thank you!