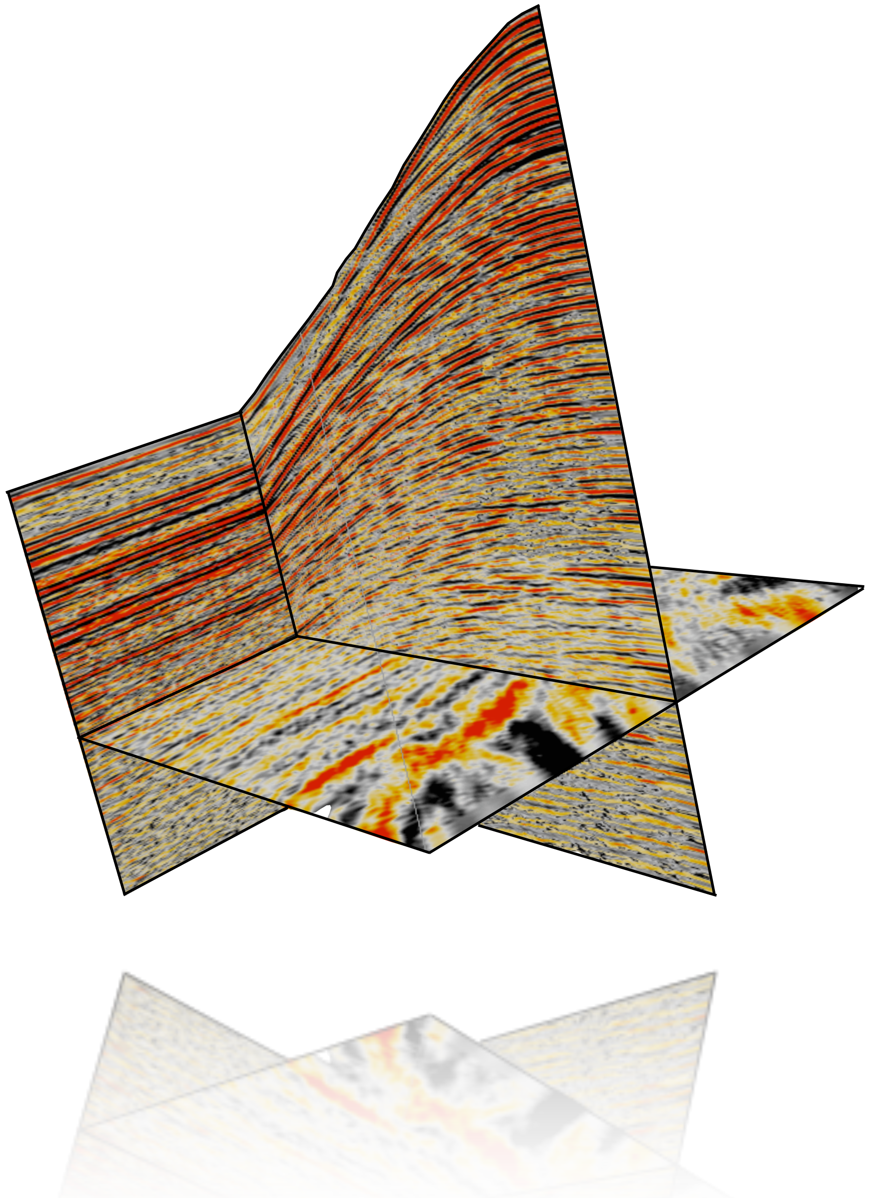
Compressive wavefield simulations

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

Joint work with Yogi Erlangga, and Tim Lin

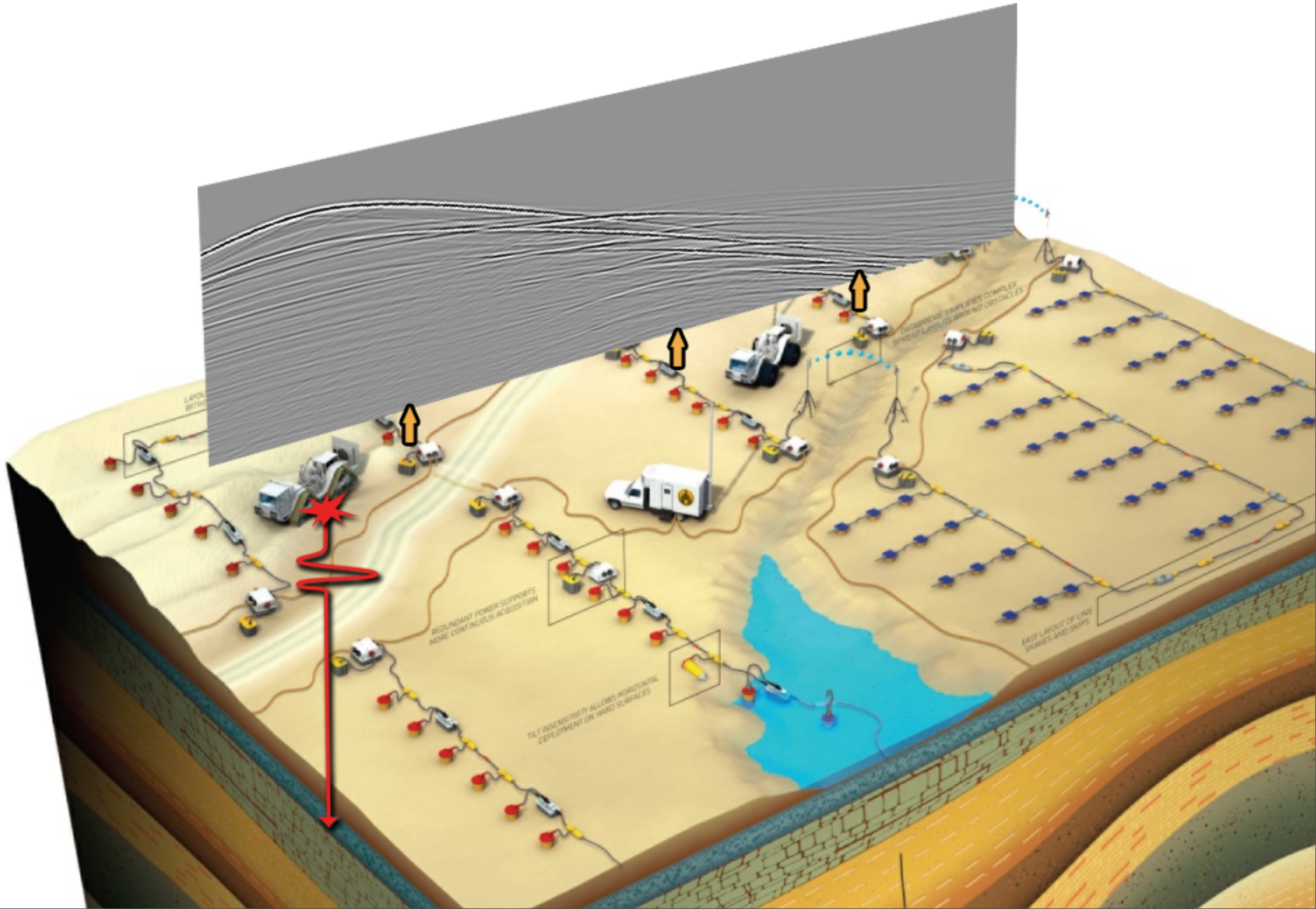
***Seismic Laboratory for Imaging & Modeling**
Department of Earth & Ocean Sciences
The University of British Columbia



Marseille, May 19, 2009

Seismic acquisition

Individual shots



Individual shots

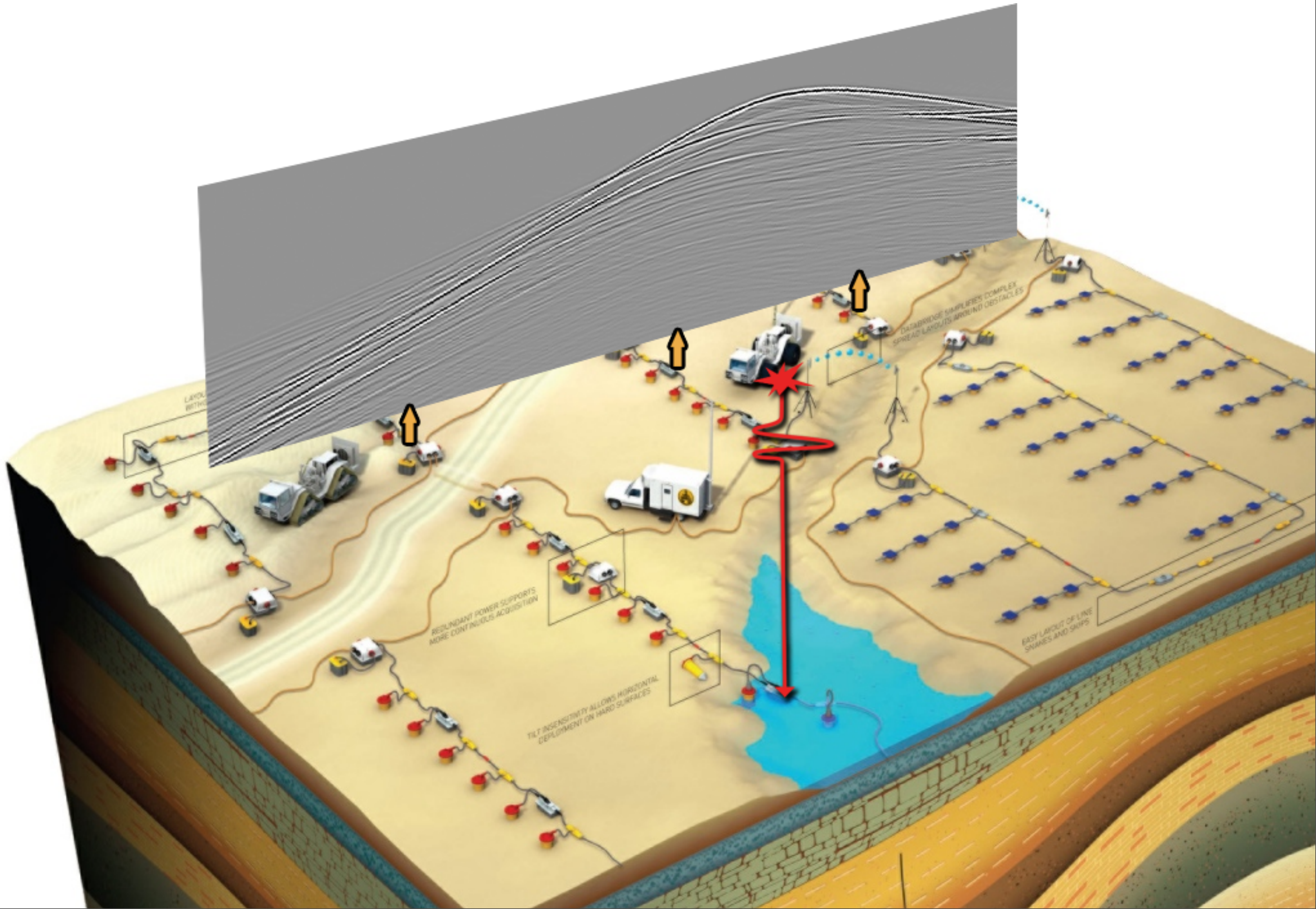
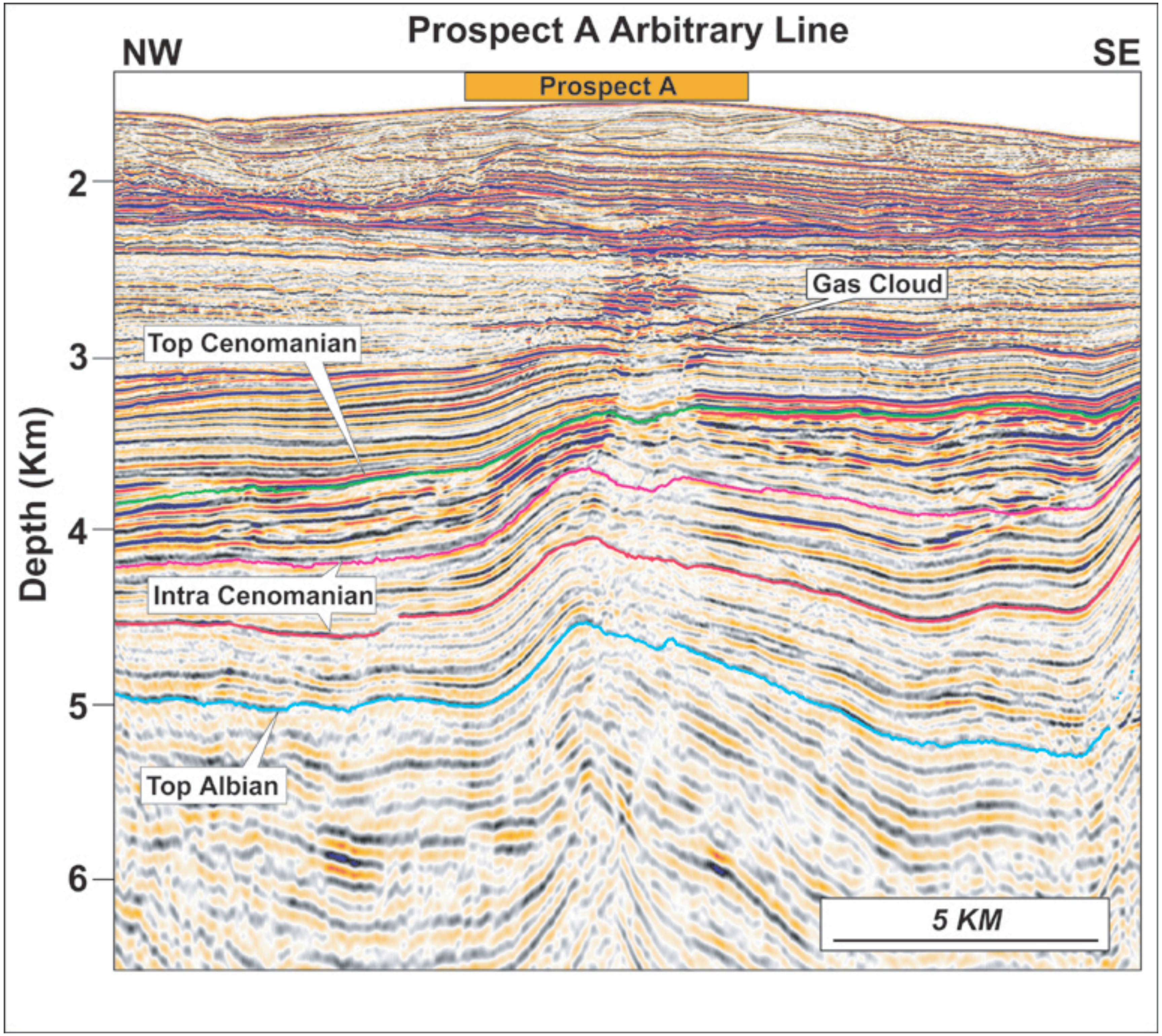


image courtesy vancoenergy (www.vancoenergy.com)

After imaging



Essentials of seismic inversion

Simulation:

sources

residue (data)

$$\mathbf{H}[\mathbf{m}]\mathbf{U} = \mathbf{Q} \quad \text{and} \quad \mathbf{H}^*[\mathbf{m}]\mathbf{V} = \mathbf{R}$$

discretized PDE (Helmholtz)

variables (Earth)

solution (seismic wavefield)

adjoint solution (seismic wavefield)

Essentials of seismic inversion

Simulation:

sources

residue (data)

$$\mathbf{H}[\mathbf{m}]\mathbf{U} = \mathbf{Q} \quad \text{and} \quad \mathbf{H}^{*}[\mathbf{m}]\mathbf{V} = \mathbf{R}$$

discretized PDE (Helmholtz)

variables (Earth)

solution (seismic wavefield)

adjoint solution (seismic wavefield)

- Oscillatory high-dimensional solutions that are ***extremely*** expensive to compute
- Inversion (e.g. via Gauss-Newton) requires multiple solves
- Number of blocks in **H** and number of **rhs** determine simulation & acquisition costs

Essentials of seismic inversion

Simulation:

sources ↓ · **residue (data)** ↓

$$\mathbf{H}[\mathbf{m}]\mathbf{U} = \mathbf{Q} \quad \text{and} \quad \mathbf{H}^*[\mathbf{m}]\mathbf{V} = \mathbf{R}$$

discretized PDE (Helmholtz) → $\mathbf{H}$; **variables (Earth)** → $\mathbf{m}$; **solution (seismic wavefield)** → $\mathbf{U}$; **adjoint solution (seismic wavefield)** → $\mathbf{V}$

- Oscillatory high-dimensional solutions that are ***extremely*** expensive to compute
- Inversion (e.g. via Gauss-Newton) requires multiple solves
- Number of blocks in **H** and number of **rhs** determine simulation & acquisition costs

Imaging:

image volume ↓ · **multi-D 'cross-correlation'** ↓

$$\boldsymbol{\delta}\mathbf{I}(x_s, x_r, t) = (\mathbf{U} \circ \mathbf{V})$$

$$\delta\mathbf{m}(x_s = x_r, t = 0) = \mathrm{diag}\{\boldsymbol{\delta}\mathbf{I}\}$$

Essentials of seismic inversion

Simulation:

sources; residue (data)

$$\mathbf{H}[\mathbf{m}]\mathbf{U} = \mathbf{Q} \quad \text{and} \quad \mathbf{H}^*[\mathbf{m}]\mathbf{V} = \mathbf{R}$$

discretized PDE (Helmholtz); **variables (Earth)**; **solution (seismic wavefield)**; **adjoint solution (seismic wavefield)**

- Oscillatory high-dimensional solutions that are ***extremely*** expensive to compute
- Inversion (e.g. via Gauss-Newton) requires multiple solves
- Number of blocks in **H** and number of **rhs** determine simulation & acquisition costs

Imaging:

image volume; multi-D 'cross-correlation'

$$\boldsymbol{\delta}\mathbf{I}(x_s, x_r, t) \quad = \quad (\mathbf{U} \circ \mathbf{V})$$

$$\delta\mathbf{m}(x_s = x_r, t = 0) \quad = \quad \mathrm{diag}\{\boldsymbol{\delta}\mathbf{I}\}$$

- Explicit matrix evaluations part of the KKT system
- Only interested in diagonal (focused energy)
- Penalize off diagonals (impose focusing) as part of an extended PDE constrained optimization problem

Impediments & solution strategy

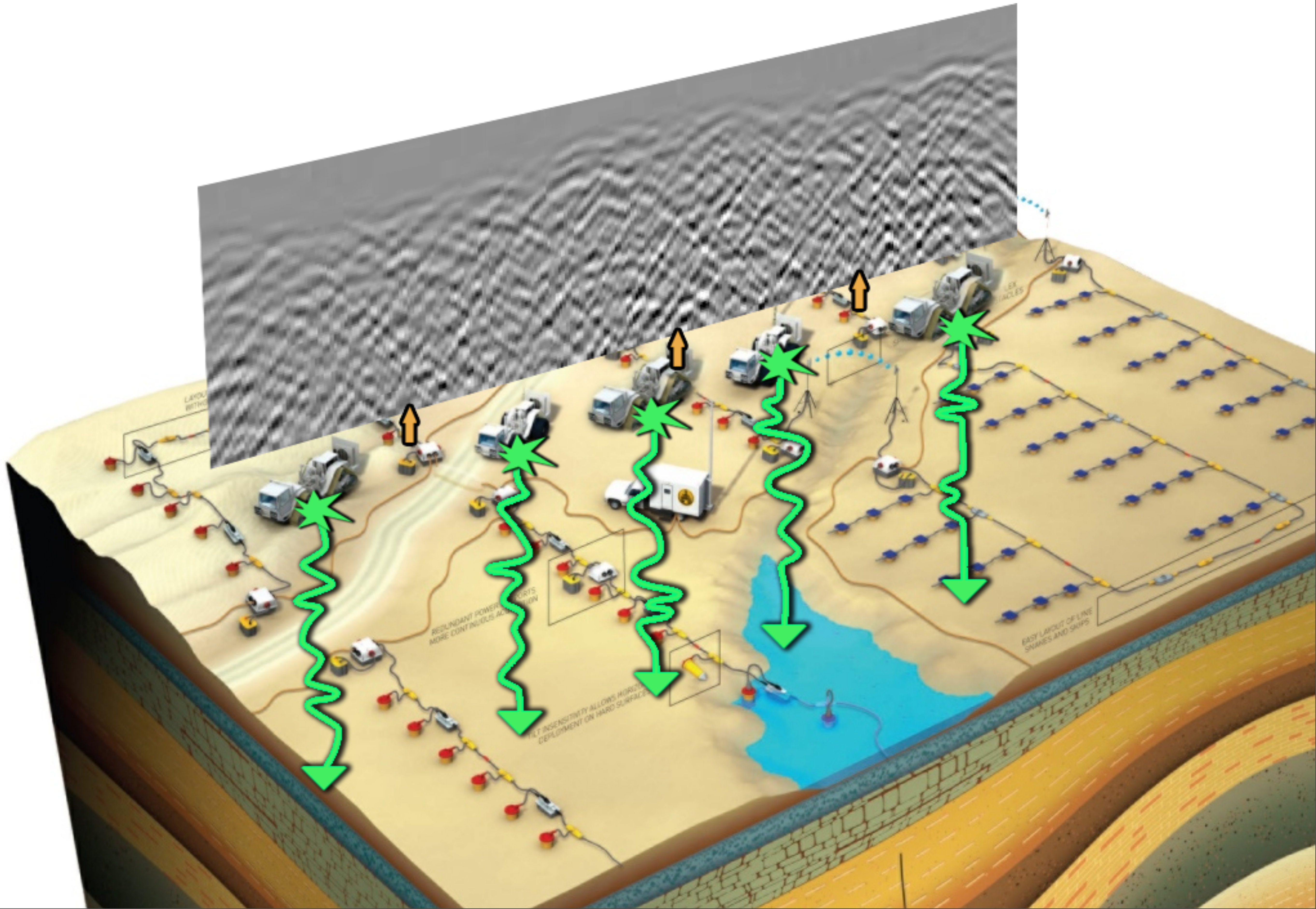
- Acquisition, modeling & inversion **costs** are proportional to the **size** of *data* and *model*
  - curse of dimensionality (d=5 for data)
  - redundancy in data
  - high geometrical structure (wavefronts)
- Computation of image volumes through focusing requires explicit storage of wavefields
  - redundancy (d=7 for image volume, d=3 for model)
  - object of interest is low dimensional and sparse (compressible) in ***phase*** space
  - high geometrical structure (sheet-like singularities)
- **Bottom line: We are drowning in data that may exceed Petabytes**
- **Solution:** use *embedding/compressive* sampling techniques
  - to reduce the size for the Helmholtz system part of forward modeling
  - to reduce the size of the explicit matrix evaluations part of imaging
  - exploit geometrical structure & focusing = joint sparsity promotion with ***curvelets***

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

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

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]

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

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

- 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

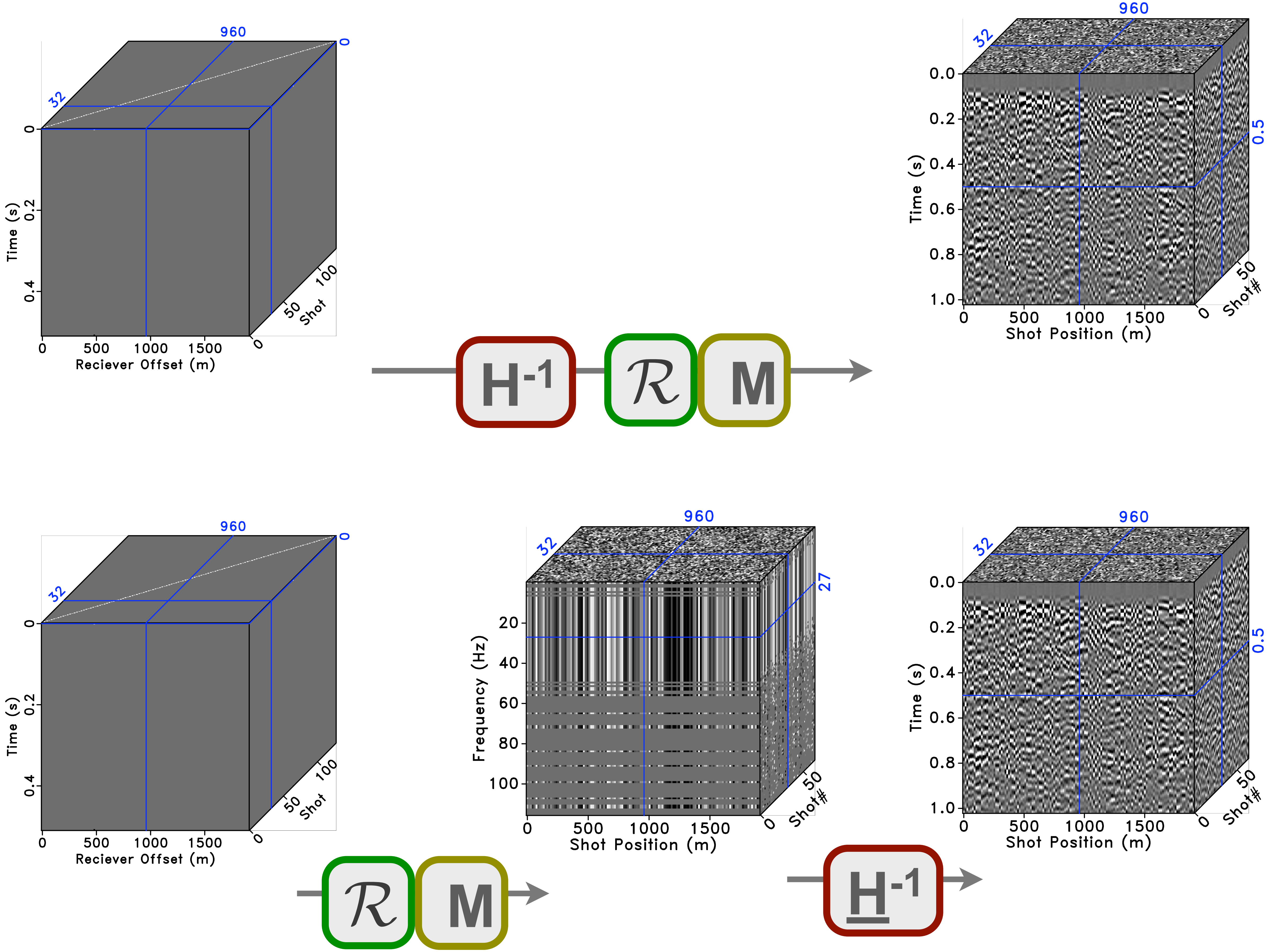
$$\boxed{\mathcal{O}(n^3 \log n)}$$

Source-solution sampling equivalence

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

- 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

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



960
32
0
Time (s)
0.2
0.4
0
500
1000
1500
Reciever Offset (m)
50
100
Shot
H-1
R
M
0.0
0.2
0.4
0.6
0.8
1.0
Time (s)
Shot Position (m)
0.5
Shot#
Frequency (Hz)
20
40
60
80
100
27

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

SLIM Seismic Laboratory for Imaging and Modeling

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)

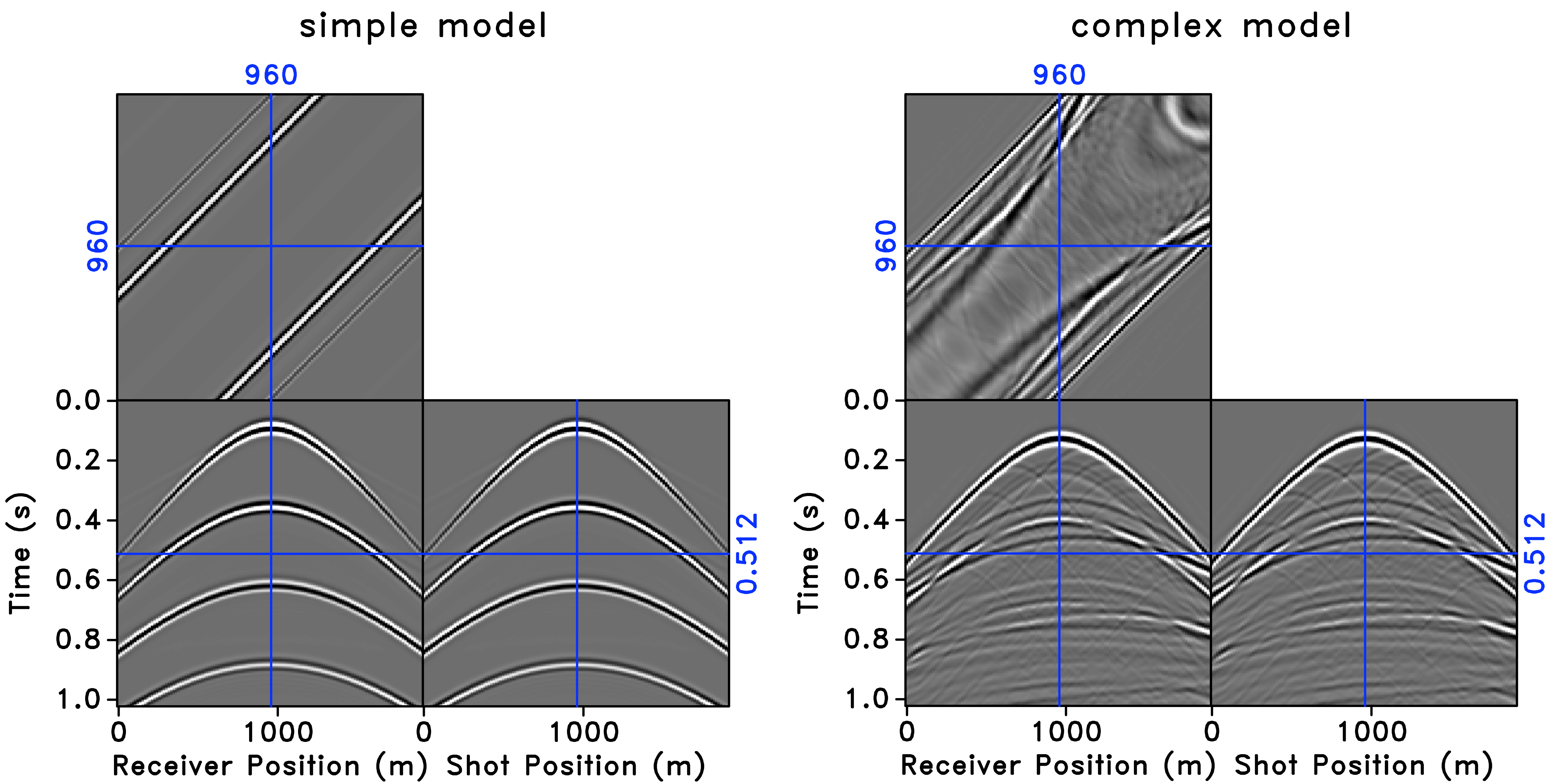
- fast projection in time & shot directions: $\mathcal{O}(n \log n)$
- **Cost** $\mathcal{O}(n^3 \log n)$ instead of $\mathcal{O}(n^4)$

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

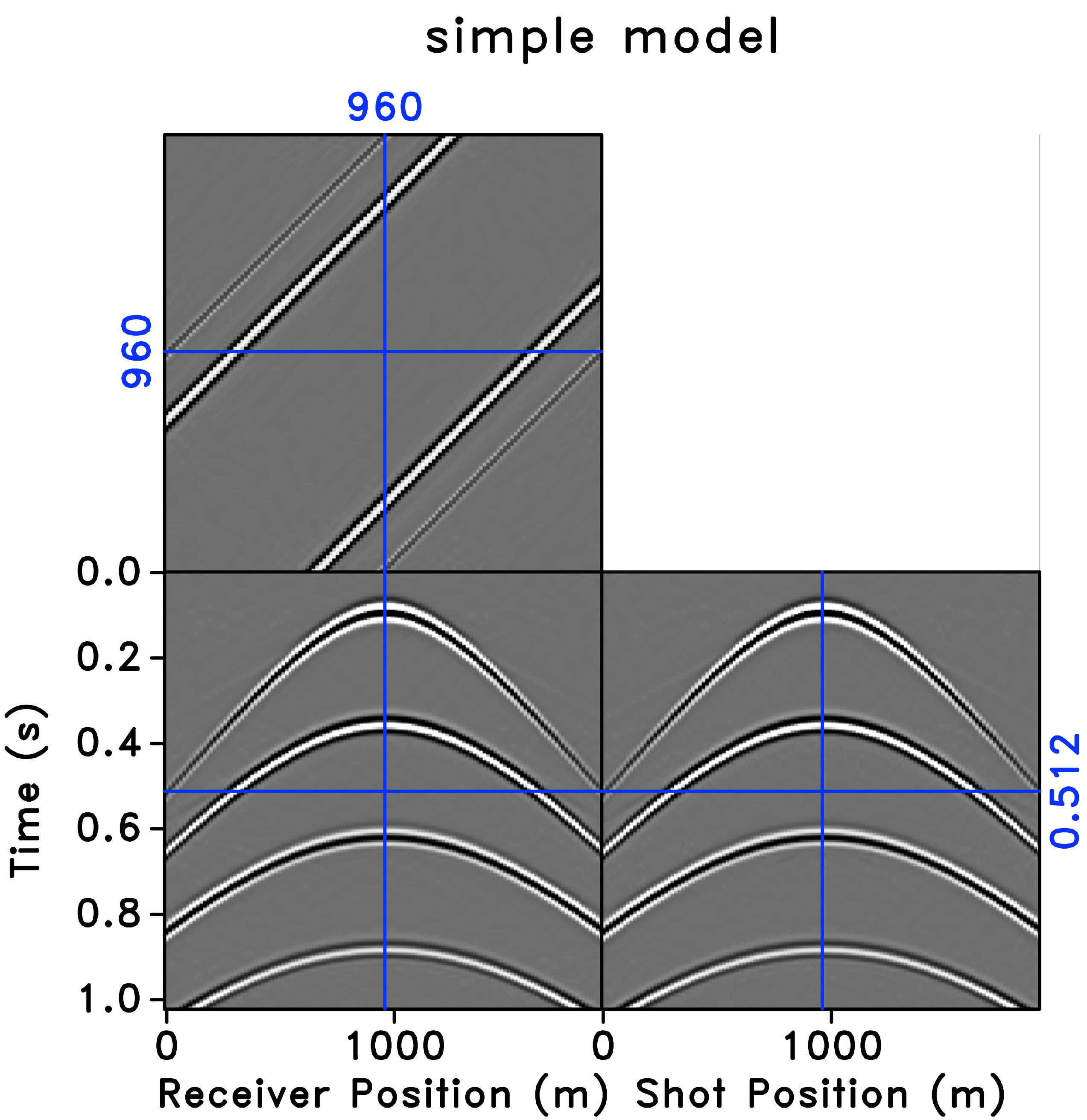
SLIM
Seismic Laboratory for Imaging and Modeling

Green’s functions

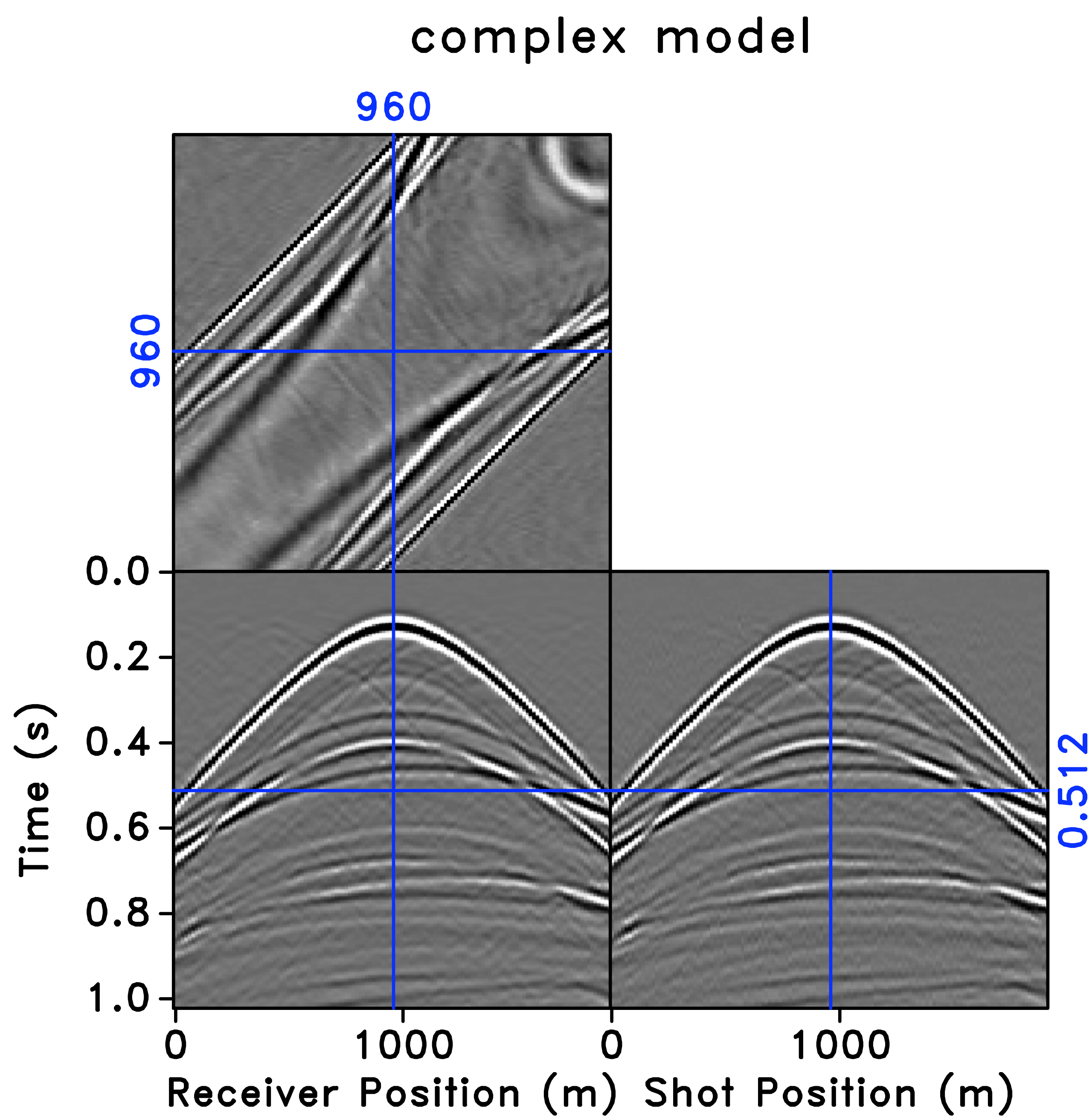


SLIM
Seismic Laboratory for Imaging and Modeling

Recovered data



28.1dB



18.2dB

300 SPGL1 iteration

SLIM
Seismic Laboratory for Imaging and Modeling

The next step

- **Fast matrix computations**
  - Improved Approximation Algorithms for Large Matrices via Random Projections by Tamás Sarlós, ’08

$$\mathbf{AB} \approx \mathbf{A} \left(\mathbf{RM}\right)^{*} \left(\mathbf{RM}\right) \mathbf{B}$$

  - Faster Least Squares Approximation by Petros Drineas et. al., ’08

$$\arg\min_{\mathbf{X}} \|\mathbf{B} - \mathbf{AX}\|_2 \approx \arg\min_{\mathbf{X}} \|\mathbf{RM}\left(\mathbf{B} - \mathbf{AX}\right)\|_2$$

- **Joint sparsity-promotion with mixed (1,2) norm minimization**
  - Joint-sparse recovery from multiple measurements by E. van den Berg and M. Friedlander, ‘09

$$\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{1,2} \quad \text{subject to} \quad \|\mathbf{AX} - \mathbf{B}\|_{2,2} \leq \sigma,$$

- Based on Johnson-Lindenstrauss type embeddings.

Wavefield focusing

Define linear mid-point/offset coordinate transformation

$$\boldsymbol{\delta}\mathbf{I}'(m,h,t) = \mathbf{T}^{\Delta h}_{(x_s,x_r)\mapsto(m,h)}\boldsymbol{\delta}\mathbf{I}(x_s,x_r,t),$$

with $m = \frac{1}{2}(x_s + x_r) \quad \text{and} \quad h = \frac{1}{2}(x_s - x_r)$

Penalize ***defocusing*** via minimizing [Symes, '09]

$$\|\mathsf{P}_h\mathbf{I}'(\cdot,\, h)\|_2 \text{ with } \mathsf{P}_h\cdot = \mathbf{h}\cdot$$

an *annihilator* that increasingly *penalizes* non-zero offsets.

Remark: conventional imaging principle

$$\delta\mathbf{m} = \boldsymbol{\delta}\mathbf{I}'(\cdot, h=0, t=0)$$

Compressive wavefield inversion with focussing

Compressively sample augmented system

$$\begin{aligned} \mathbf{RM}\left(\mathbf{U}^* \circ \mathbf{S}^*\mathbf{X}\right) &\approx \mathbf{RMV}^T \\ \mathsf{P}_h\mathbf{X} &\approx \mathbf{0} \end{aligned} \quad \text{or} \quad \mathbf{AX} \approx \mathbf{B}$$

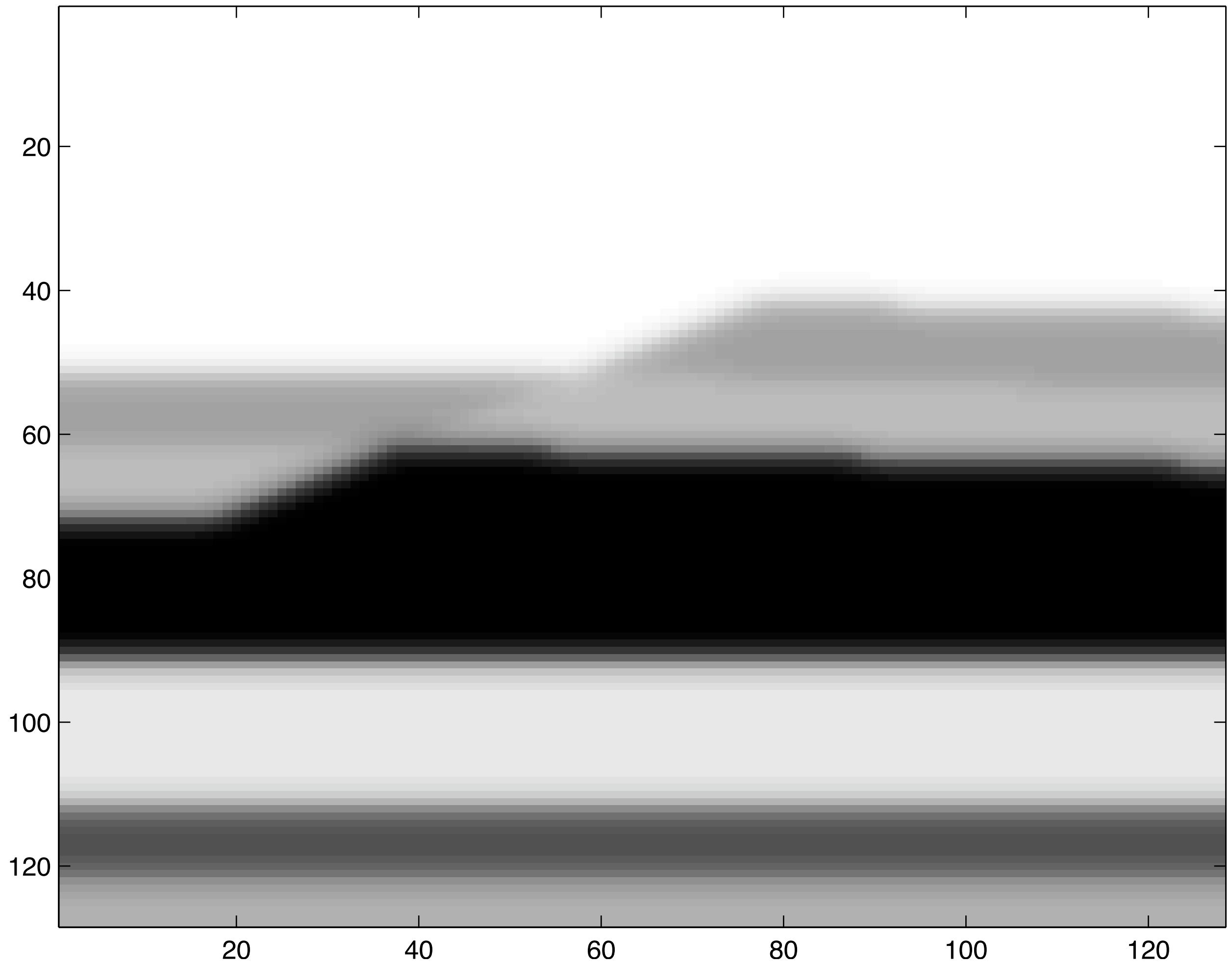
Recover focused solution by mixed (1,2)-norm minimization

$$\tilde{\mathbf{X}} = \arg\min_{\mathbf{X}} \|\mathbf{X}\|_{1,2} \quad \text{subject to} \quad \|\mathbf{AX} - \mathbf{B}\|_{2,2} \leq \sigma,$$

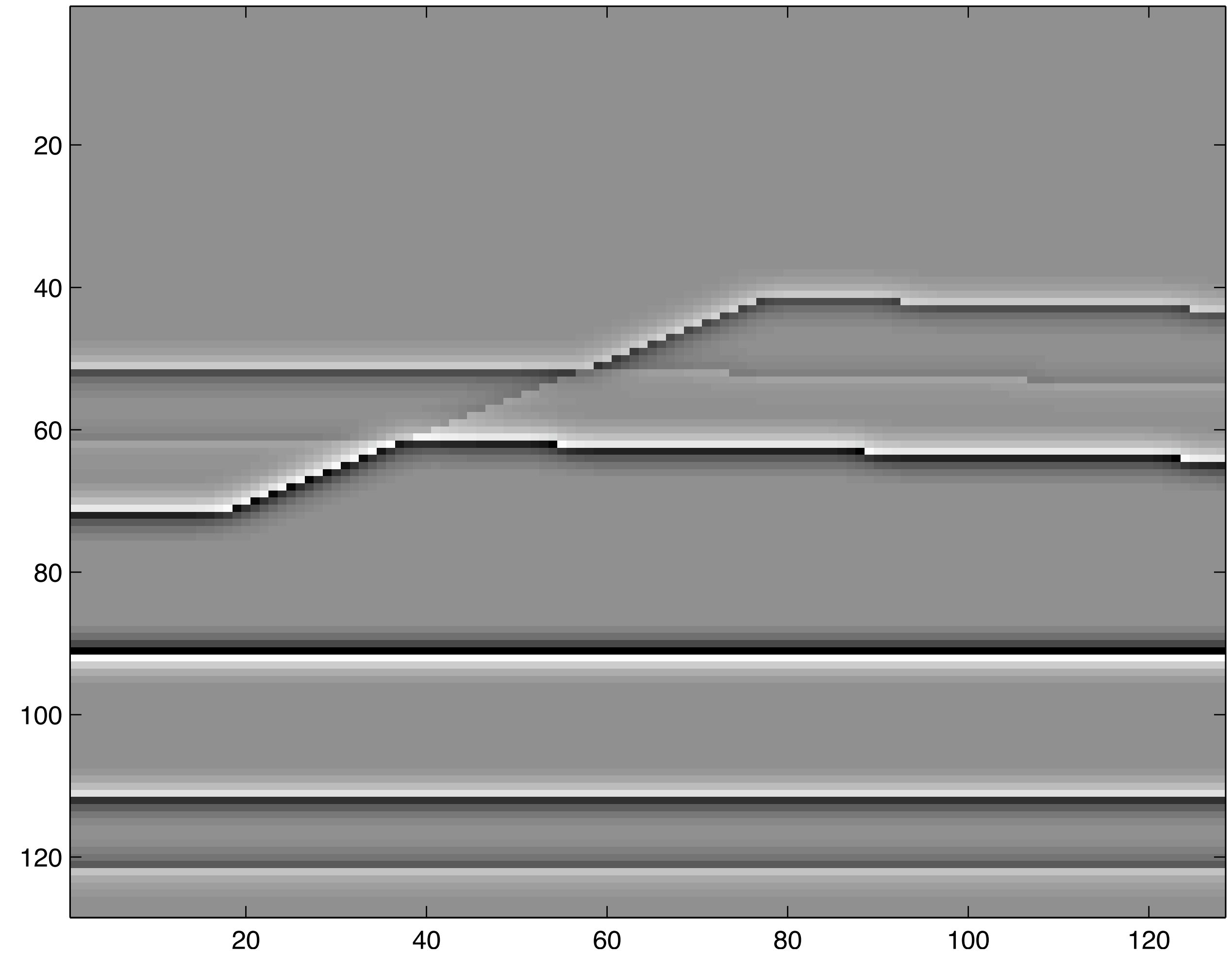
with $$\|\mathbf{X}\|_{1,2} := \sum_{i \in \text{rows}(\mathbf{X})} \|\text{row}_i(\mathbf{X})^*\|_2$$

and $$\|\mathbf{X}\|_{2,2} := \left(\sum_{i \in \text{rows}(\mathbf{X})} \|\text{row}_i(\mathbf{X})^*\|_2^2 \right)^{\frac{1}{2}}.$$

Example

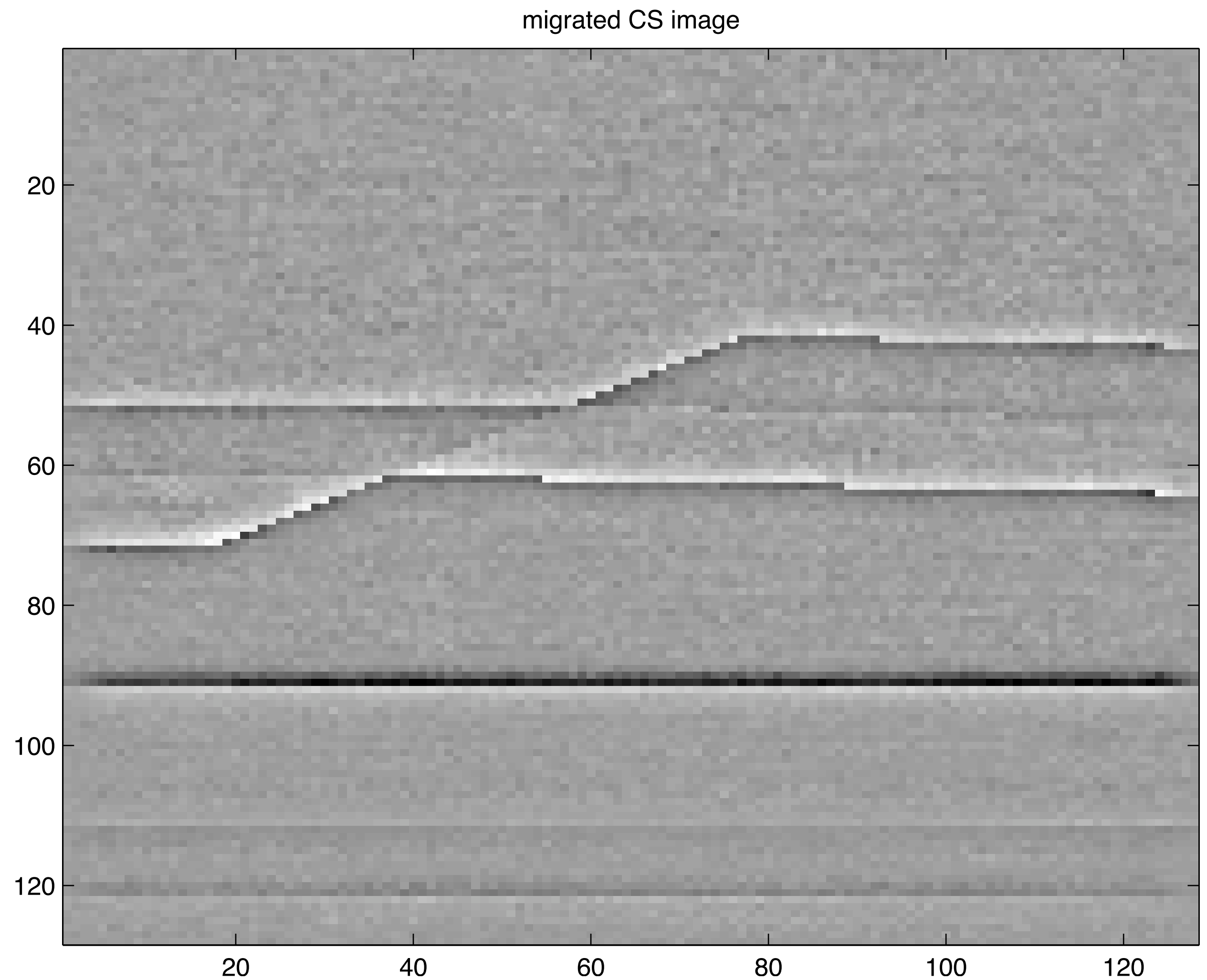


background velocity model

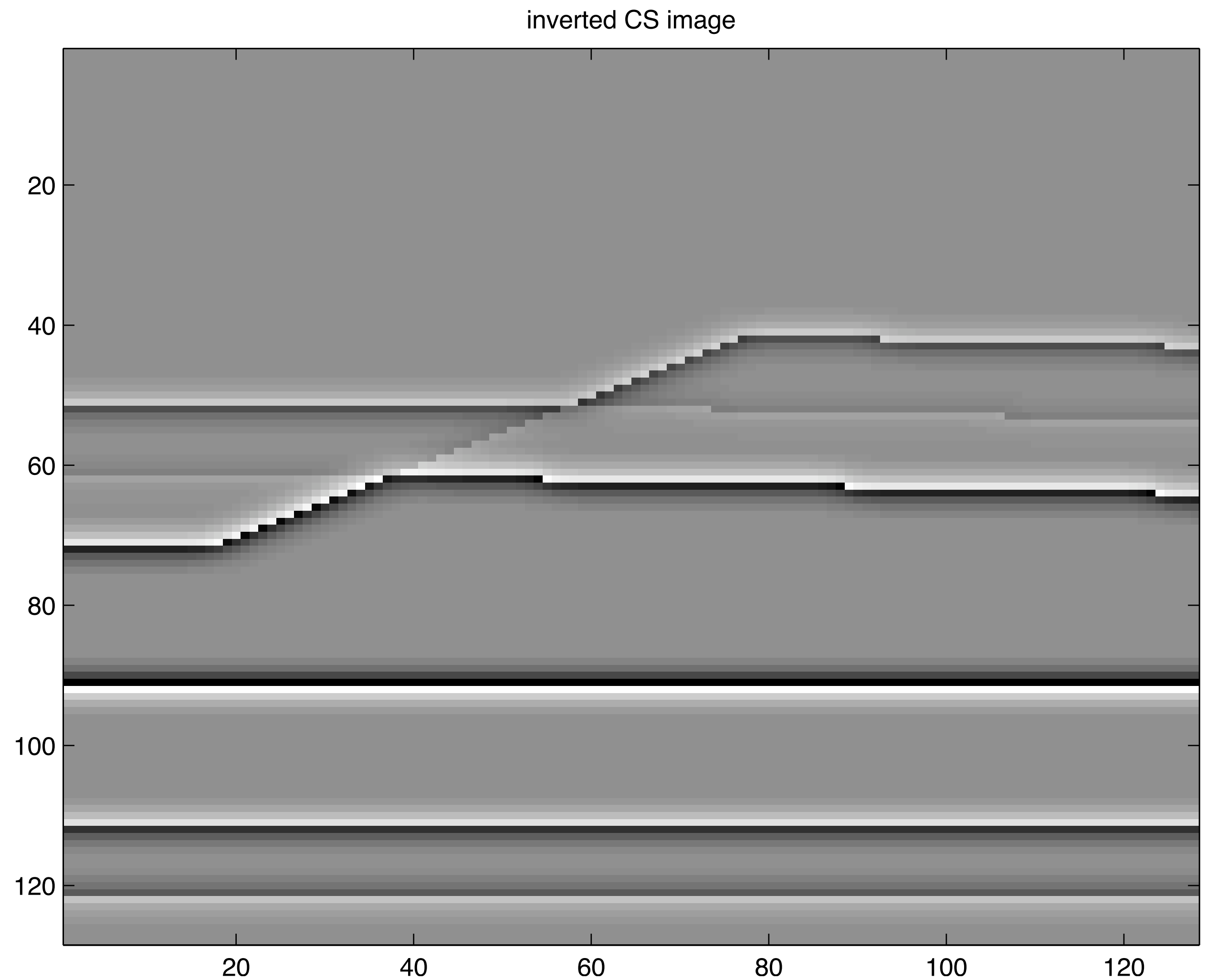


perturbation

Example



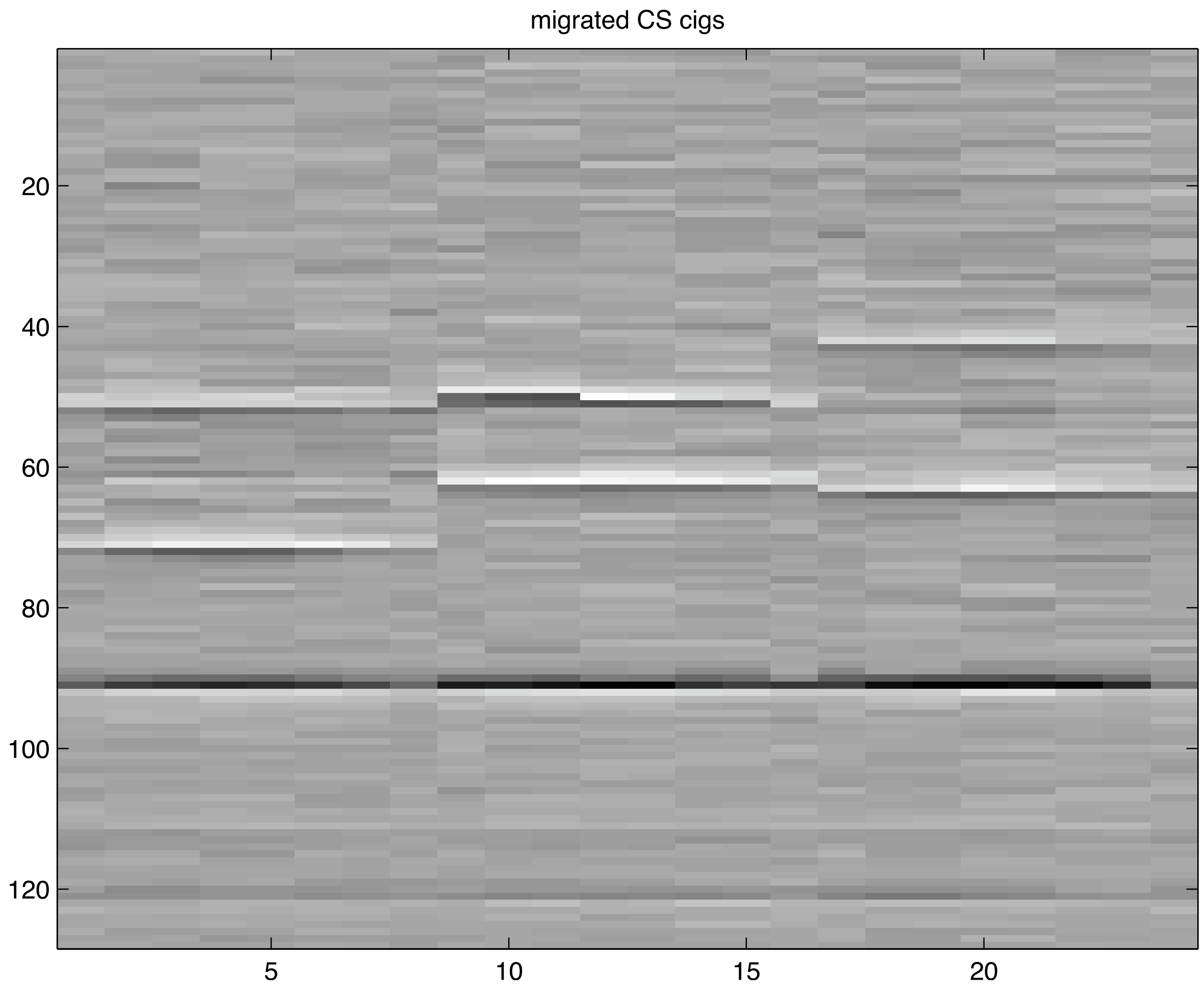
matched filter



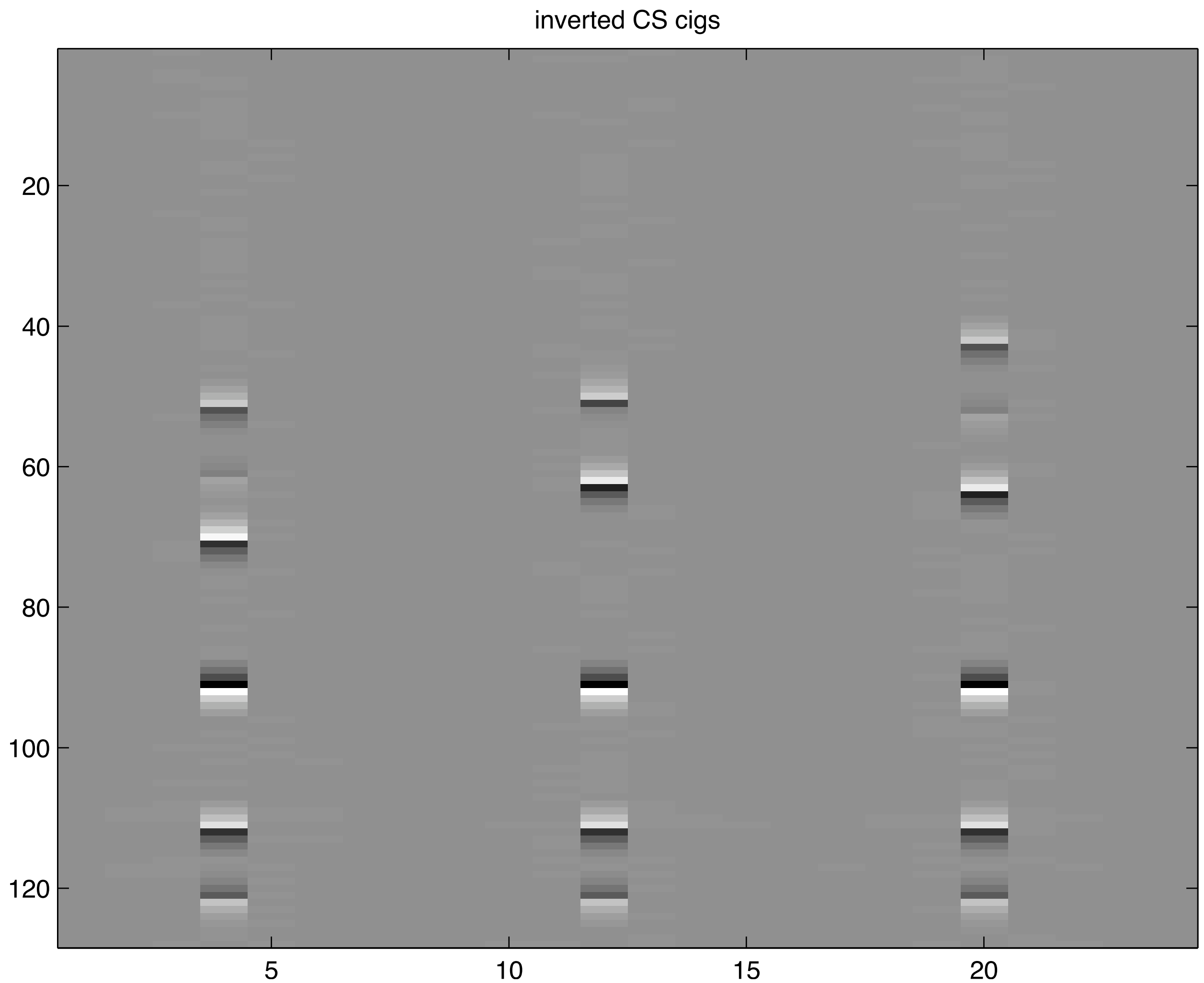
wavefield inversion

Recovery from 64-fold subsampling ...

Example



correlation based



wavefield inversion

Common-image gathers are focussed.

Conclusions & outlook

- **CS** provides a **new** *linear sampling paradigm*
  - ***degree*** of *subsampling commensurate* with transform-domain ***sparsity***
  - subsampling of seismic data volumes
    - missing source-receiver locations
    - simultaneous acquisition
  - *subsampling* of solutions to PDEs

- **CS** leads to
  - “acquisition” of *smaller* data volumes that carry the ***same*** *information* or
  - to ***improved*** *inferences* from data using the *same* resources
  - feasibility of nonlocal extensions of Helmholtz solvers that are otherwise impossible

- Bottom line: **acquisition & numerical modeling costs** are **no** longer determined by the ***size*** of the ***discretization*** but by the ***transform-domain compressibility*** 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 SINBAD project with financial support from 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.

slim.eos.ubc.ca

and... Thank you!