# Compressive-wavefield simulations

Felix J. Herrmann, Yogi Erlangga, and Tim. T. Y. Lin

Department of Earth and Ocean Sciences, the University of British Columbia, Canada
fherrmann,yerlangga,tlin@eos.ubc.ca

## Abstract:

Full-waveform inversion's high demand on computational resources forms, along with the non-uniqueness problem, the major impediment withstanding its widespread use on industrial-size datasets. Turning modeling and inversion into a compressive sensing problem—where simulated data are recovered from a relatively small number of independent simultaneous sources—can effectively mitigate this high-cost impediment. The key is in showing that we can design a sub-sampling operator that commutes with the time-harmonic Helmholtz system. As in compressive sensing, this leads to a reduction in simulation cost. Moreover, this reduction is commensurate with the transform-domain sparsity of the solution, implying that computational costs are no longer determined by the size of the discretization but by transform-domain sparsity of the solution of the CS problem which forms our data. The combination of this sub-sampling strategy with our recent work on implicit solvers for the Helmholtz equation provides a viable alternative to full-waveform inversion schemes based on explicit finite-difference methods.

## 1. Introduction

With the recent resurgence of full-waveform inversion—i.e., adjoint-state methods applied to solve PDE-constrained optimization problems—the computational cost of solving forward modeling has become one of the major impediments withstanding successful application of this technology to industry-size data volumes. To overcome this impediment, we argue that further improvements will depend on a problem formulation with a computational complexity that is no longer strictly determined by the *size* of the discretization but by transform-domain *sparsity* of its solution. In this new paradigm, we bring computational costs in par with our ability to compress solutions to certain PDEs. This premise is related to two recent developments. First, there is the new field of compressive sensing [CS in short throughout the paper, 5, 6]—where the argument is made, and rigorously proven—that compressible signals can be recovered from severely sub-Nyquist sampling by solving a sparsity promoting program. Second, there is in the seismic community the recent resurgence of simultaneous-source acquisition [1, 13, 3, 17, 12], and continuing efforts to reduce the cost of seismic modeling, imaging, and inversion through phase encoding of simultaneous sources [15, 20, 13, 12], removal of subsets of angular frequencies [21, 16, 14, 12] or plane waves [22]. All these approaches correspond to instances of CS and by using its principles we are able to remove the associated sub-sampling interferences through a combination of exploiting transform-domain sparsity, properties of certain sub-sampling schemes, and the existence of sparsity promoting solvers.

## 2. Theory

### 2.1 Compressive full-waveform inversion

Full-waveform inversion entails solving PDE-constrained optimization problems of the following type:

$$\min_{\mathbf{U},\,\mathbf{m}} \frac{1}{2}\|\mathbf{R}\mathbf{M}\big(\mathbf{d}-\mathbf{D}\mathbf{U}\big)\|_2^2 \quad \text{s.t.} \quad \mathbf{H}[\mathbf{m}]\mathbf{U}=\mathbf{B}, \qquad (1)$$

where $\mathbf{d}$ is the observed data volume, $\mathbf{U}$ the solution of the multi-source (in its columns)-frequency Helmholtz equation, $\mathbf{D}$ the detection operator that extracts the simulated data from time-harmonic solutions at the receiver locations, $\mathbf{H}$ a matrix with the discretized multi-frequency Helmholtz equation, and $\mathbf{B}$ a matrix with the frequency-transformed source distributions in its columns. In the above optimization problem (from which, after casting Eq. 1 in its unconstrained form, most quasi-Newton type full-waveform inversion schemes derive), solutions for the unknown velocity model, $\mathbf{m}$, and for the wave equation, $\mathbf{U}$, that minimize the energy mismatch are pursued. Solutions to Eq.1 require multiple solves of the implicit Helmholtz equation, which may prove computationally prohibitive even after preconditioning (yielding a complexity for this solver of $\mathcal{O}(n^4)$ in 2-D [8, 7]). We address this problem by using CS [19, 12] to reduce the size of the seismic data volume through $\mathbf{y}=\mathbf{R}\mathbf{M}\mathbf{d}$ where

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

with $\mathbf{F}_{2,3}$ the 2,3-D Fourier transforms, and $\boldsymbol{\theta}=\mathrm{Uniform}([0,2\pi])$ a random phase rotation. The matrices $\mathbf{R}^{\Omega}$ and $\mathbf{R}^{\Sigma}$ represent CS-subsampling matrices (see Figure 1) acting along the rows (frequency coordinate) and columns (source coordinate) of the data volume, respectively. As shown by [12] application of this CS-sampling

matrix, $\mathbf{RM}$, to the data is equivalent to applying it to the source wavefields directly, which turns single-impulsive sources into a smaller set ($n_s' \ll n_s$ with $n_s$ the number of separated single-impulsive source) of time-harmonic simultaneous sources that are randomly phase encoded and that have for each source a different set of angular frequencies missing—i.e., there are $n_f' \ll n_f$ (with $n_f$ the number of frequencies of fully sampled data) frequencies non-zero (see Figure 1). This implies that the subsampling operator commutes with the Helmholtz system and this allows us to recast Eq. 1 into the following reduced form (consisting of fewer frequencies and fewer right-hand sides):

$$\min_{\underline{\mathbf{U}},\,\mathbf{m}} \frac{1}{2}\|\mathbf{y} - \underline{\mathbf{D}}\underline{\mathbf{U}}\|_2^2 \quad \text{s.t.} \quad \underline{\mathbf{H}}[\mathbf{m}]\underline{\mathbf{U}} = \underline{\mathbf{B}}, \qquad (2)$$

where the underlined quantities are related to the reduced Helmholtz system.

### 2.2 The time-harmonic Helmholtz system

Since their inception, iterative implicit matrix-free solutions to the time-harmonic Helmholtz equation have been plagued by lack of numerical convergence for decreasing mesh sizes and increasing angular frequencies [18]. The inclusion of deflation—a way to handle small eigenvalues that lead to slow convergence [8, 7]—can successfully remove this impediment, bringing 2 and 3-D solvers for the time-harmonic Helmholtz into reach. For a given source (right-hand side $\mathbf{b}$) and angular frequency $\omega$ ($:= 2\pi f$, with $f$ the temporal frequency in Hz), the frequency-domain wavefield $\mathbf{u}$ is computed with a Krylov method that involves the following system of equations:

$$\boldsymbol{\mathcal{H}}[\omega]\boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Q}}\hat{\mathbf{u}} = \mathbf{b}, \quad \mathbf{u} = \boldsymbol{\mathcal{M}}^{-1}\boldsymbol{\mathcal{Q}}\hat{\mathbf{u}},$$

where $\boldsymbol{\mathcal{H}}[\omega]$, $\boldsymbol{\mathcal{M}}$, and $\boldsymbol{\mathcal{Q}}$ respectively represent the discretized monochromatic Helmholtz equation, the preconditioner, and the projection matrices. As shown by [9, 10], convergence is guaranteed by defining the preconditioning matrix $\boldsymbol{\mathcal{M}}$ in terms of the discretized shifted or damped Helmholtz operator $\mathcal{M} := -\nabla\cdot\nabla - \frac{\omega^2}{c(x)^2}(1-\beta\hat{i}), \quad \hat{i} = \sqrt{-1}$, with $\beta > 0$. With this preconditioning, the eigenvalues of $\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{M}}^{-1}$ are clustered into a circle in the complex plane. By the action of the projector matrix $\boldsymbol{\mathcal{Q}}$, these eigenvalues move towards unity on the real axis. These two operations effectively lower the condition number, which explains the superior performance of this solver.

### 2.3 Source-solution CS-sampling equivalence

Aside from the required number of frequencies, the computational cost of full-wavefield simulation is determined by the number of sources—i.e., the number of right-hand sides. In the current simulation paradigm, the number of sources coincides with the number of single-impulsive source simulations. As prescribed by CS, this number can be reduced by designing a survey that consists of a relatively *small number* of simultaneous experiments with simultaneous sources that contain *subsets* of angular frequencies. Mathematically, we can accomplish this by applying a CS-sampling matrix, $\mathbf{RM}$, to the individual-impulsive sources collected in the vector $\mathbf{s}$. Now, if we can show that the solution from this set of "compressed" sources $\underline{\mathbf{s}} = \mathbf{RMs}$, is identical to the compressively sampled solution yielded from modeling the *complete*, we are in the position to speed up our computations. This speed up is the result of a decreased number of experiments and angular frequencies that are present in the simultaneous source vector. For this to work, the solution $\underline{\mathbf{y}}$ must be equivalent to the solution $\mathbf{y}$ obtained by compressively sampling the full solution. More specifically, we need to demonstrate that the solutions for the full and compressed systems are equivalent—i.e., $\mathbf{y} = \underline{\mathbf{y}}$ in

$$\begin{cases} \mathbf{H} = \mathbf{D}^* \underbrace{\mathbf{s}}_{\text{impulsive sources}} \\ \mathbf{HU} = \mathbf{B} \\ \mathbf{y} = \mathbf{RMDU} := \mathbf{RMd} \end{cases} \qquad \begin{cases} \underline{\mathbf{B}} = \underline{\mathbf{D}}^*\underline{\mathbf{s}} = \underline{\mathbf{D}}^* \underbrace{\mathbf{RMs}}_{\text{sim. sources}} \\ \underline{\mathbf{HU}} = \underline{\mathbf{B}} \\ \underline{\mathbf{y}} = \underline{\mathbf{DU}}. \end{cases}$$

Here, $\mathbf{H} = \mathrm{diag}(\boldsymbol{\mathcal{H}}[\omega_i])$ is the block-diagonal discretized Helmholtz equation for each $\omega_i := 2\pi i\cdot\Delta f$, $i = 1\cdots n_f$, with $n_f$ the number of frequencies and $\Delta f$ its sample interval. The adjoint (denoted by $^*$) of the detection matrix $\mathbf{D}$ injects the individual sources into the multiple right-hand sides, $\mathbf{B} = [\mathbf{b}_1\ \mathbf{b}_2\ \cdots\ \mathbf{b}_{n_s}]$, with $n_s$ the number of shots. This detection matrix extracts data at the receiver positions. Its adjoint inserts data at the co-located source positions. Each column of $\mathbf{U}$ contains the wavefields for all frequencies induced by the shots located in the columns of $\mathbf{B}$. Consequently, the full simulation requires the inversion of the block-diagonal system (for all shots), followed by a detection—i.e., we have $\mathbf{d} = \mathbf{DH}^{-1}\mathbf{B}$, with $\mathbf{H}^{-1} = \mathrm{diag}(\boldsymbol{\mathcal{H}}^{-1}[\omega_i])$, $i = 1\cdots n_s$. After CS sampling, this volume is reduced to $\mathbf{y} = \mathbf{RMd}$ by applying the flat rectangular CS-sampling matrix $\mathbf{RM}$ (defined explicitly in the next section) to the full simulation. Applying $\mathbf{RM}$ directly to the sources $\mathbf{s}$ leads to a compressed system $\underline{\mathbf{H}}$, which after inversion gives $\underline{\mathbf{y}}$. To illustrate why $\mathbf{y}$ is equivalent to $\underline{\mathbf{y}}$, consider a compressive sampling of the solution over frequency by the subsampling matrix $\mathbf{R}^\Omega$ (for clarity, we removed the orthonormal measurement matrix). This restriction matrix removes arbitrary rows from the right-hand side. By virtue of the block-diagonal structure of our system, we have $\mathbf{R}^\Omega\mathbf{H}^{-1} = \underline{\mathbf{H}}^{-1}\mathbf{R}^\Omega$ with $\underline{\mathbf{H}}^{-1} = \mathrm{diag}(\boldsymbol{\mathcal{H}}^{-1}[\omega_i]), i \subset \{1\cdots n_f\}$, yielding $\mathbf{R}^\Omega\mathbf{U} = \underline{\mathbf{H}}^{-1}\underline{\mathbf{B}} = \underline{\mathbf{U}}$, where $\underline{\mathbf{B}} := \mathbf{R}^\Omega\mathbf{B}$. This means that frequency subsampling the right-hand side, followed by solving the system for the corresponding frequencies, is the same as solving the full system, followed by frequency subsampling. A similar argument holds when subsampling the shots (removing arbitrary columns of $\mathbf{B}$). Now, we have the reduced system $\mathbf{R}^\Omega\mathbf{U}(\mathbf{R}^\Sigma)^* = \underline{\mathbf{H}}^{-1}\underline{\mathbf{B}} = \underline{\mathbf{U}}$, with $\underline{\mathbf{B}} := \mathbf{R}^\Omega\mathbf{B}(\mathbf{R}^\Sigma)^*$. Using Kronecker products, these relations can be written succinctly as $(\mathbf{R}^\Sigma\otimes\mathbf{R}^\Omega)\mathrm{vec}\,(\mathbf{U}) = \mathrm{vec}\,(\underline{\mathbf{U}})$ and $(\mathbf{R}^\Sigma\otimes\mathbf{R}^\Omega)\mathrm{vec}\,(\mathbf{B}) = \mathrm{vec}\,(\underline{\mathbf{B}})$ with $\mathrm{vec}\,(\cdot)$ being a linear operator that maps a matrix into a lexicographically-sorted array.

The inversion of $\underline{\mathbf{HU}} = \underline{\mathbf{B}}$ is easier because it involves only a subset of angular frequencies and simultaneous shots—i.e., $\{\underline{\mathbf{U}}, \underline{\mathbf{B}}\}$ contain only $n_s'$ columns with $n_f'$ frequency components each. Finally, the matrix $\underline{\mathbf{D}}$ extracts the compressed data from the solution.

### 2.4 The sparsifying transform

Aside from proper CS sampling the recovery from simultaneous simulations depends on a sparsifying transform that compresses seismic data, is fast, and reasonably incoherent with the CS sampling matrix. We accomplish this by defining the sparsity transform as the Kronecker product between the 2-D discrete curvelet transform [4] along the source-receiver coordinates, and the discrete wavelet transform along the time coordinate—i.e., $\mathbf{S} := \mathbf{C} \otimes \mathbf{W}$ with $\mathbf{C}$, $\mathbf{W}$ the curvelet- and wavelet-transform matrices, respectively.

### 2.5 Recovery by sparsity promotion

We reconstruct the seismic wavefield by solving the following nonlinear optimization problem

$$\widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{y}, \qquad (3)$$

with $\widetilde{\mathbf{d}} = \mathbf{S}^*\widetilde{\mathbf{x}}$ the reconstruction, $\mathbf{A} := \mathbf{R}\mathbf{M}\mathbf{S}^*$ the CS matrix, and $\mathbf{y}\,(=\underline{\mathbf{y}})$ the compressively simulated data (cf. Equation 2-right). Equation 4 is solved by SPG$\ell_1$ [2], a projected-gradient algorithm with root finding.

## 3. Computational complexity analysis

According to [18], the cost of the iterative Helmholtz solver equals $n_f n_s n_{it} \mathcal{O}(n^d)$, typically with $n_{it} = \mathcal{O}(n)$ the number of iterations. For $d = 2$ and assuming $n_s = n_f = \mathcal{O}(n)$, this cost becomes $\mathcal{O}(n^5)$. Under the same assumption, the cost of a time-domain solver is $\mathcal{O}(n^4)$. The iterative Helmholtz solver can only become competitive if $n_{it} = \mathcal{O}(1)$, yielding an $\mathcal{O}(n^4)$ computational complexity. [8, 7] achieve this by the method explained earlier. Despite this improvement, this figure is still overly pessimistic for simulations that permit sparse representations. As long as the simulation cost exceeds the $\ell_1$-recovery cost (cf. Equation 4), CS will improve on this result. This reduction depends on the cost of $\mathbf{A}$, which is dominated by the CS-matrix. For naive choices, such as Gaussian projections, these sampling matrices cost $\mathcal{O}(n^3)$ for each frequency, which offers no gain. However, with our choice of fast $\mathcal{O}(n \log n)$ projections with random convolutions [19], we are able to reduce this cost to $\mathcal{O}(n^2 \log n)$. Note that these costs are of the same order as those of calculating the sparsifying transforms. Now, the leading order cost of the $\ell_1$ recovery is reduced to $\mathcal{O}(n^3 \log n)$, which is significantly less than the cost of solving the full Helmholtz system, especially for large problems ($n \to \infty$) and for extensions to $d = 3$.

## 4. Example

To illustrate CS-recovery quality, we conduct a series of experiments for two velocity models, namely the complex model used in [11], and a simple single-layer model. These models generate seismic lines that differ in complexity. During these experiments, we vary the subsampling ratio and the frequency-to-shot subsampling ratio. All simulations are carried out with a fully parallel Helmholtz solver for a spread with 128 collocated shots and receivers sampled at a 30 m interval. The time sample interval is $0.004\,\mathrm{s}$ and the source function is a Ricker wavelet with a central frequency of 10 Hz. By solving Equation 4, we recover the full simulation for the two datasets. Comparison between the full and compressive simulations in Figure 2 shows remarkable high-fidelity results even for increasing subsampling ratios. As expected, the SNR for the simple model is better because of the reduced complexity, whereas the numbers in Table 1 for the complex model confirm increasing recovery errors for increasing subsampling ratios. Moreover, the bandwidth limitation of seismic data explains improved recovery with decreasing frequency-to-shot ratio for a fixed subsampling ratio. Because the speedup of the solution is roughly proportional to the subsampling ratio, we can conclude that speedups of four to six times are possible at the expense of a minor drop in SNR.

| Subsample ratio | 0.25 | 0.15 | 0.07 |
|---|---|---|---|
| $n'_f/n'_s$ | recovery error (dB) | | |
| 2 | 14.3 | 12.1 | 8.6 |
| 1 | 18.2 | 14.5 | 10.2 |
| 0.5 | 22.2 | 16.5 | 10.7 |
| Speed up (%) | 400 | 670 | 1420 |

Table 1: Signal-to-noise ratios based on the complex model, SNR $= -20\log_{10}(\frac{\|\mathbf{d}-\widetilde{\mathbf{d}}\|_2}{\|\mathbf{d}\|_2})$ for reconstructions with the curvelet-wavelet sparsity transform for different subsample and frequency-to-shot ratios..

## 5. Discussion and conclusions

Compressive sampling—where data with structure (read transform-domain sparsity) can be reconstructed from degrees of subsamping commensurate their sparsity—is considered a paradigm shift. We show that this new paradigm can also be applied to reduce the computational complexity of solving PDEs (or, even better, that of PDE-constrained optimization problems). When we have control over the sources—such as on land—CS extends to acquisition itself. This is exciting because CS decouples simulation- and acquisition-related costs from the discretization size. Instead, these costs depend on sparsity. Because the model space is even sparser, we expect further improvements when we apply CS principles to inversion itself.

## References


[1] Craig J. Beasley. A new look at marine simultaneous sources. *The Leading Edge*, 27(7):914–917, 2008.

[2] E. van den Berg and M. P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. Technical Report 2, Department of Computer Science, University of British Columbia, January 2008.

[3] A. J. Berkhout. Changing the mindset in seismic data acquisition. *The Leading Edge*, 27(7):924–938, 2008.

[4] E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Modeling and Simulation*, 5:861–899, 2006.

[5] E.J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[6] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[7] Y. A. Erlangga and F. J. Herrmann. An iterative multilevel method for computing wavefields in frequency-domain seismic inversion. In *SEG Technical Program Expanded Abstracts*, volume 27, pages 1957–1960. SEG, November 2008.

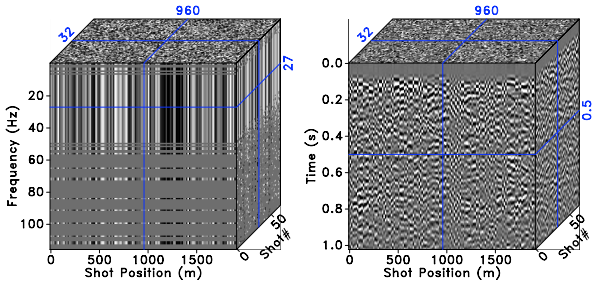


Figure 1: Compressive sampling with simultaneous sources. **(a)** Amplitude spectrum for the source signatures emitted by each source as part of the simultaneous-source experiments. These signatures appear noisy in the shot-receiver coordinates because of the phase encoding (cf. Equation 1). Observe that the frequency restrictions are different for each simultaneous source experiment. **(b)** CS-data after applying the inverse Fourier transform. Notice the noisy character of the simultaneous-shot interferences..


[8] Y A Erlangga and R Nabben. On multilevel projection Krylov method for the preconditioned Helmholtz system. 2007. Submitted for publication.

[9] Y A Erlangga, C Vuik, and C W Oosterlee. On a class of preconditioners for solving the Helmholtz equation. *Applied Numerical Mathematics*, 50:409–425, 2004.

[10] Y A Erlangga, C Vuik, and C W Oosterlee. Comparison of multigrid and incomplete LU shifted-Laplace preconditioners for the inhomogeneous Helmholtz equation. *Applied Numerical Mathematics*, 56:648–666, 2006.

[11] F. J. Herrmann, U. Boeniger, and D. J. Verschuur. Non-linear primary-multiple separation with directional curvelet frames. *Geophysical Journal International*, 170:781–799, 2007.

[12] Felix J. Herrmann, Yogi A. Erlangga, and Tim T.Y. Lin. Compressive simultaneous full-waveform simulation. TR-2008-09. to appear in geophysics. 2009.

[13] C.E. Krohn and R. Neelamani. Simultaneous sourcing without compromise. In *Rome 2008, 70th EAGE Conference & Exhibition*, page B008, 2008.

[14] T.T.Y. Lin, E. Lebed, Y. A. Erlangga, and F. J. Herrmann. Interpolating solutions of the helmholtz equation with compressed sensing. In *SEG Technical Program Expanded Abstracts*, volume 27, pages 2122–2126. SEG, November 2008.

[15] S. A. Morton and C. C. Ober. Faster shot-record depth migrations using phase encoding. In *SEG Technical Program Expanded Abstracts*, volume 17, pages 1131–1134. SEG, 1998.

[16] W. Mulder and R. Plessix. How to choose a subset of frequencies in frequency-domain finite-difference migration. 158:801–812, 2004.

[17] N. Neelamani, C. Krohn, J. Krebs, M. Deffenbaugh, and J. Romberg. Efficient seismic forward modeling using simultaneous random sources and sparsity. In *SEG International Exposition and 78th Annual Meeting*, pages 2107–2110, 2008.

[18] C. D. Riyanti, Y. A. Erlangga, R.-E. Plessix, W. A. Mulder, C. Vuik, and C. Oosterlee. A new iterative solver for the time-harmonic wave equation. *Geophysics*, 71(5):E57–E63, 2006.

[19] J. Romberg. Compressive sensing by random convolution. *submitted*, 2008. Preprint available at `http://users.ece.gatech.edu/~justin/Publications_files/RandomConvolutio%n.pdf`.

[20] L. A. Romero, D. C. Ghiglia, C. C. Ober, and S. A. Morton. Phase encoding of shot records in prestack migration. *Geophysics*, 65(2):426–436, 2000.

[21] Laurent Sirgue and R. Gerhard Pratt. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248, 2004.

[22] Denes Vigh and E. William Starr. 3d prestack plane-wave, full-waveform inversion. *Geophysics*, 73(5):VE135–VE144, 2008.


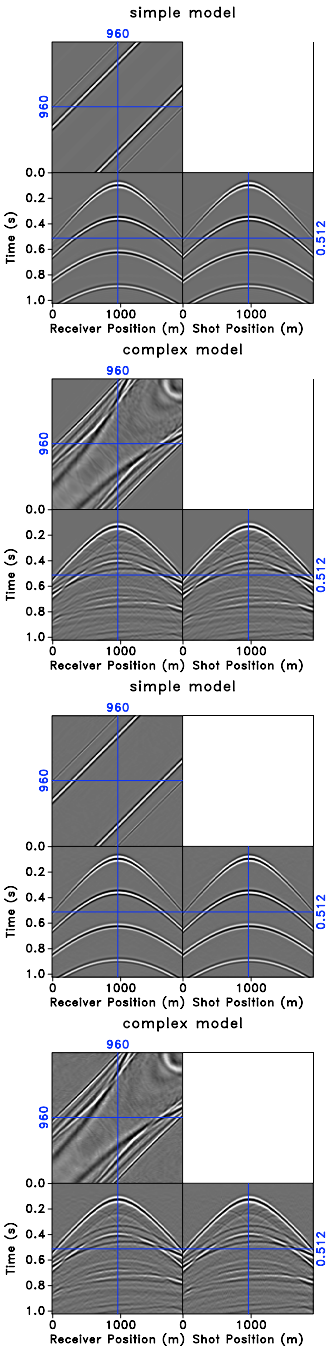


Figure 2: Comparison between conventional and compressive simulations in for simple and complex velocity models. **(a)** Crossing-planes view of the seismic line for the simple model. **(b)** The same for the complex model. **(c)**. Recovered simulation (with a SNR of 28.1 dB) for the simple model from 25 % of the samples with the $\ell_1$-solver running to convergence. **(d)** The same but for the complex model now with a SNR of 18.2 dB..