Dimensionality-reduced estimation of primaries by sparse inversion

Felix J. Herrmann

SLIM
University of British Columbia

Outline

- Motivation
- Theory
- Sparsity Promoting wavefield inversion
- Estimating primaries by sparse inversion
- Dimensionality reduction via Singular Value Decompositions (SVD's)
- Results
- Conclusions
- Future work
- Acknowledgements

Motivation

- Data-driven methods

  Estimation of primaries by sparse inversion (EPSI)

- Curse of dimensionality

  Disproportional growth in computational and storage demands when moving to realistic 3-D field data

Objective

Reduction in computational and storage demands using :

- Dimensionality reduction technique
- Adaptive low-rank approximation

SLIM

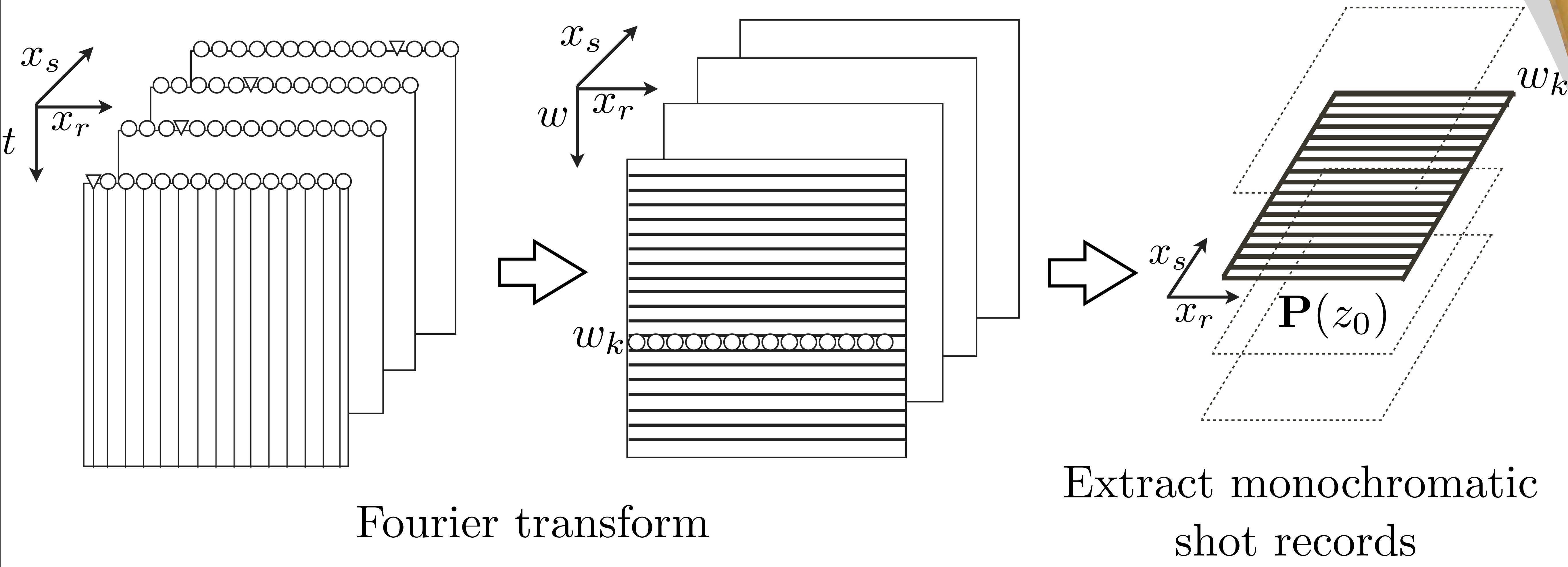
Theory

Success of EPSI Depends on

- Fast sparsifying transform
- Large scale solver - Promotes Sparsity (SPGL1)
- Fast evaluation of monochromatic data matrix and its adjoint (Most Expensive)

E.J van Dedem, 2002

Monochromatic matrix notation



Sparsity-promoting wavefield inversion

Consider the following linear relationship

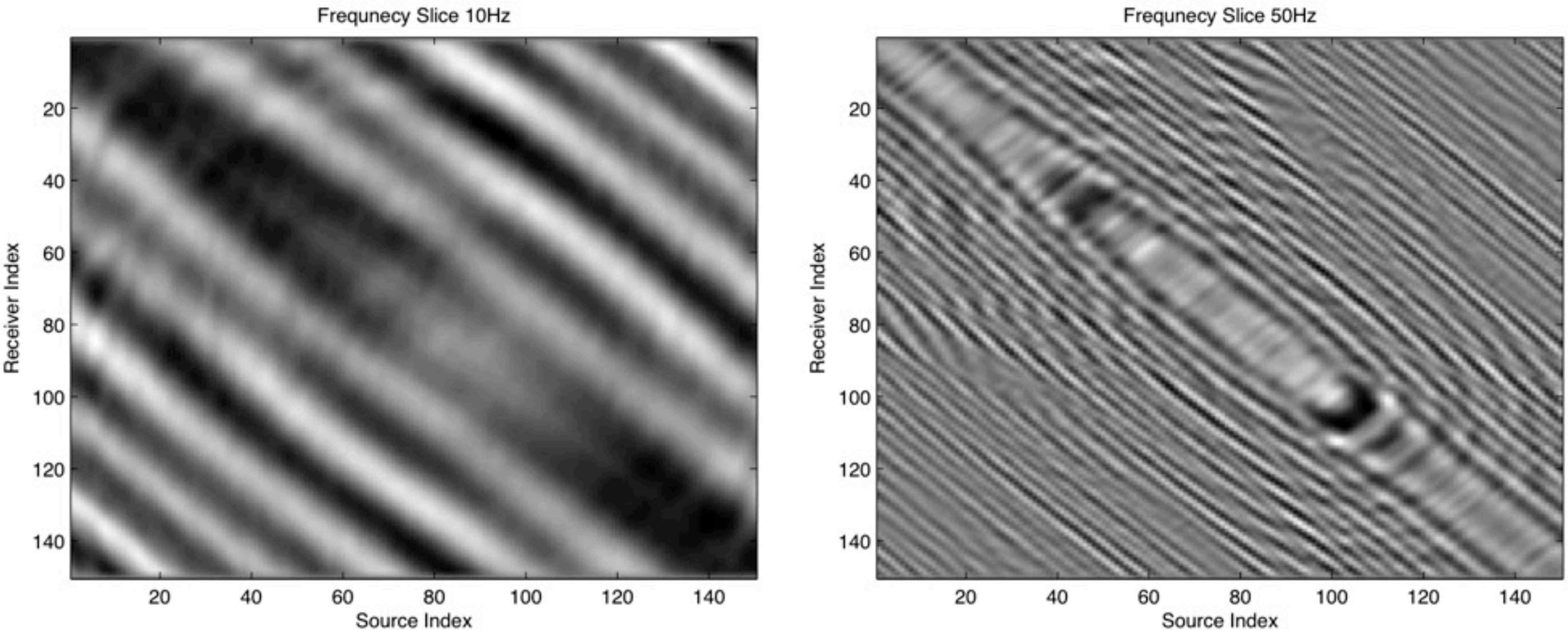
$$\widehat{\mathbf{G}}_i \widehat{\mathbf{U}}_i \approx \widehat{\mathbf{V}}_i,\ i = 1 \cdots n_f$$

$\widehat{\mathbf{U}}_i$, $\widehat{\mathbf{V}}_i$ known discretized monochromatic wavefields
$\widehat{\mathbf{G}}_i$ unknown wavefield
angular frequency $\omega = (i-1)\Delta\omega,\ i = 1 \cdots n_f$
$\Delta\omega$ the sampling rate in the Fourier domain
n_f the number of frequencies

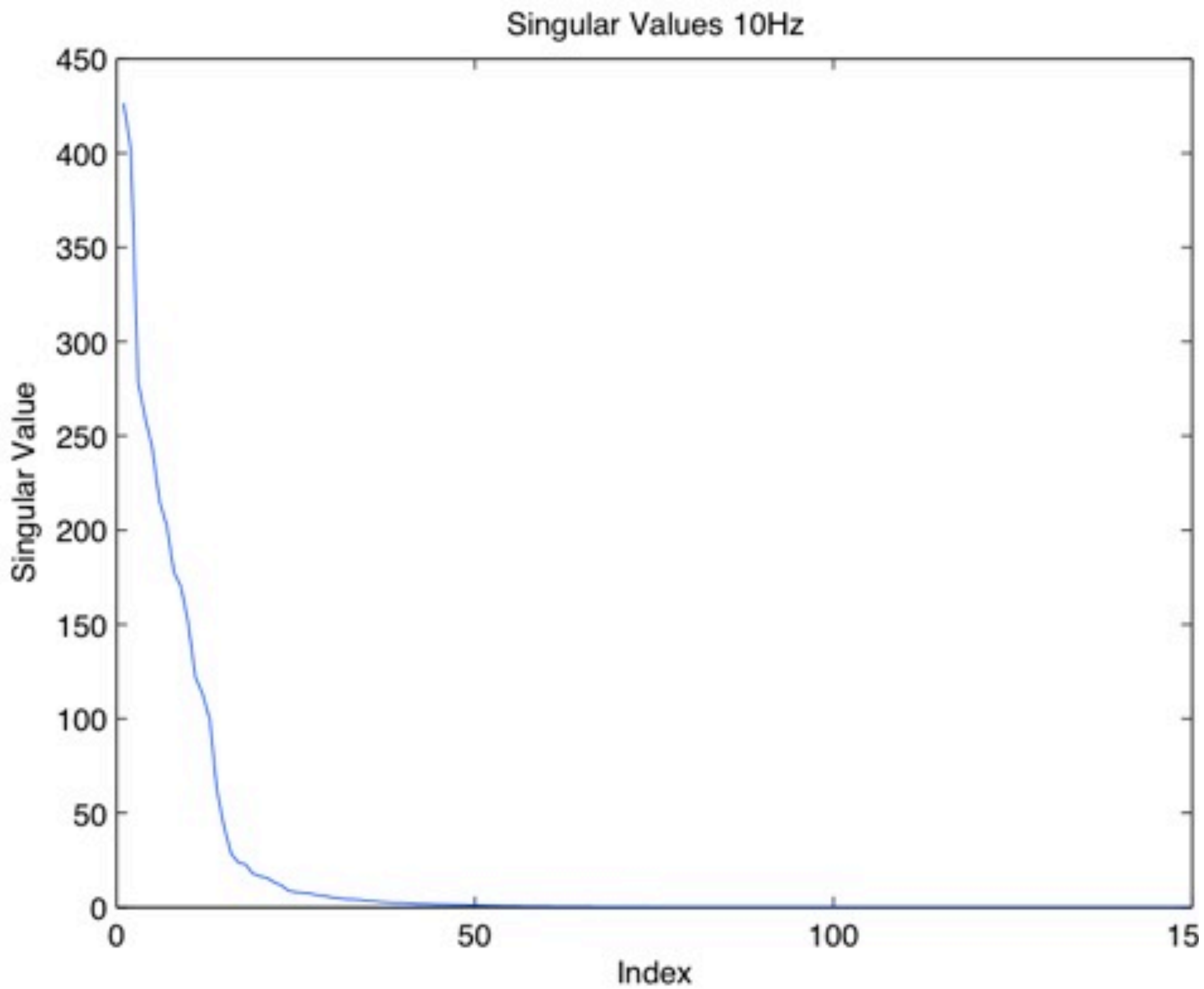
Sparsity-promoting wavefield inversion

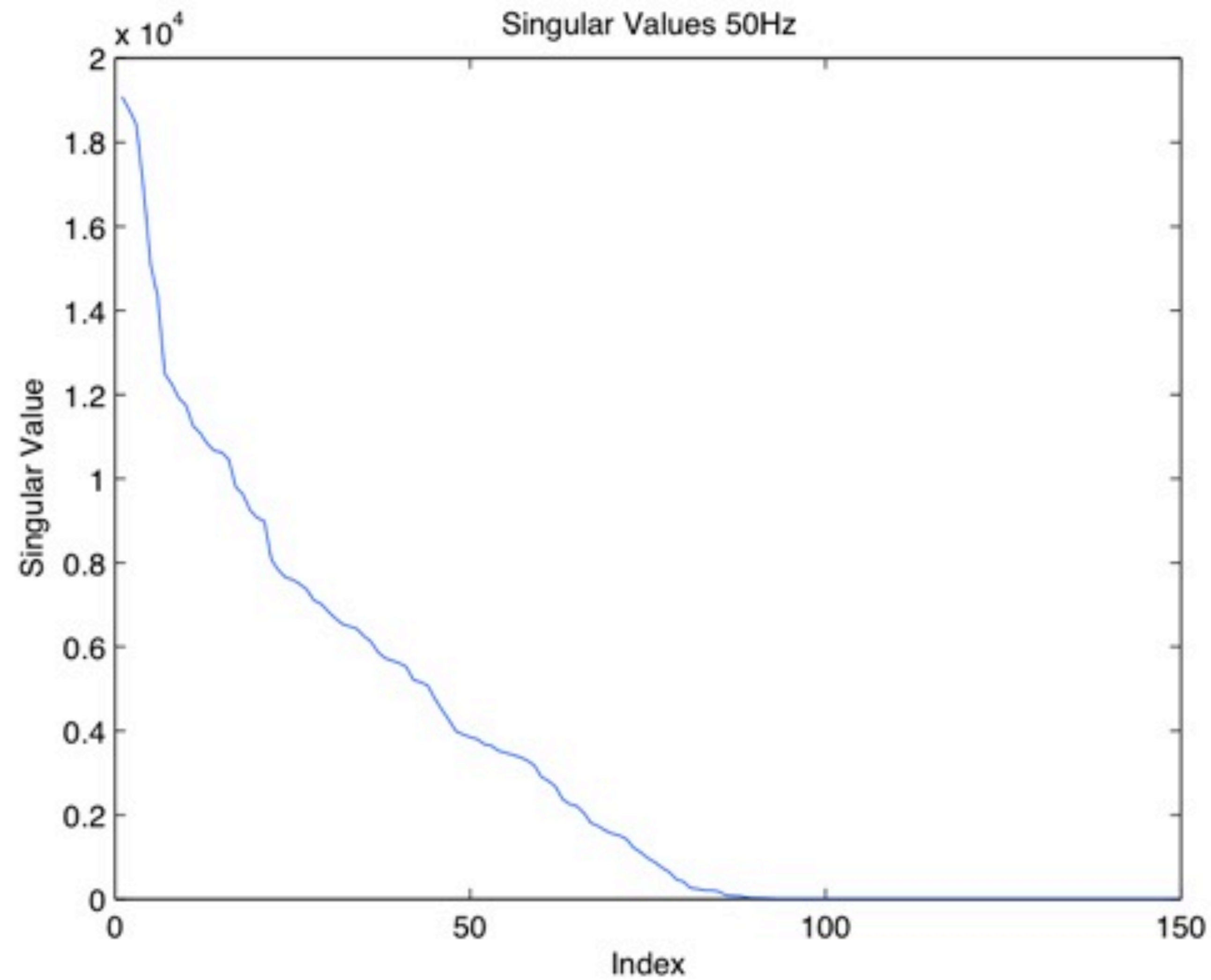
Data Matrices $\widehat{\mathbf{U}}_i, \; i = 1 \cdots n_f$

- Square
- Rank deficient (Finite Aperture)
- Scaled by source wavelet
- ill conditioned and challenging to invert because of instabilities related to small singular values.



Frequnecy Slice 10Hz
Receiver Index
Source Index
20
40
60
80
100
120
140
Frequnecy Slice 50Hz
Receiver Index
Source Index


Singular Values 10Hz
Singular Value
Index
450
400
350
300
250
200
150
100
50
0
0
50
100
15


Singular Values 50Hz
x 10^4
Singular Value
Index
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
50
100
150

Sparsity-promoting wavefield inversion

Countering the instabilities by Imposing an energy penalty on the solution through damped least-squares

$$\widetilde{\widehat{\mathbf{G}}}_i \approx \widehat{\mathbf{V}}_i \widehat{\mathbf{U}}_i^* \left(\widehat{\mathbf{U}}_i \widehat{\mathbf{U}}_i^* + \epsilon_i^2 \mathbf{I} \right)^{-1}, \, i = 1 \cdots n_f$$

ϵ_i a frequency-dependent regularization parameter that controls the data misfit versus the energy penalty on $\widehat{\mathbf{G}}$

SLIM

Sparsity-promoting wavefield inversion

Problems with the previous formulation

- Minimizing the energy leads to loss of high frequencies
- Source function leads to different energy levels at different frequencies (different ϵ for each frequency)
- Minimizing energy does not exploit multi-dimensional structure exhibited by seismic wavefields

Sparsity-promoting wavefield inversion

To address these challenges

- Cast the linear equation onto a form that allows us to solve the unknown wavefield with curvelet-domain sparsity promotion

$$\mathrm{vec}\left(\mathbf{AXB}\right) = \left(\mathbf{B}^T \otimes \mathbf{A}\right) \mathrm{vec}\left(\mathbf{X}\right)$$

$\otimes$ refers to the Kronecker product
vec is a linear operation that stacks the columns of a matrix into a long concatenated vector

Sparsity-promoting wavefield inversion

Our linear equation now becomes :

$$\left(\widehat{\mathbf{U}}_i^* \otimes \mathbf{I}\right) \mathrm{vec}\left(\widehat{\mathbf{G}}_i\right) \approx \mathrm{vec}\left(\widehat{\mathbf{V}}_i\right),\ i = 1 \cdots n_f$$

$\mathbf{I}$ the identity matrix

Sparsity-promoting wavefield inversion

After inclusion of the curvelet synthesis and temporal Fourier transforms ($\mathbf{F}_t = (\mathbf{I} \otimes \mathbf{I} \otimes \mathcal{F}_t)$ with $\mathcal{F}_t$ the temporal Fourier transform)

$$\mathbf{F}_t^* \begin{bmatrix} \left(\widehat{\mathbf{U}}_1^* \otimes \mathbf{I}\right) & & \\ & \ddots & \\ & & \left(\widehat{\mathbf{U}}_{n_f}^T \otimes \mathbf{I}\right) \end{bmatrix} \mathbf{F}_t \begin{bmatrix} \mathrm{vec}\left(\mathbf{G}_1\right) \\ \vdots \\ \mathrm{vec}\left(\mathbf{G}_{n_t}\right) \end{bmatrix} \approx \begin{bmatrix} \mathrm{vec}\left(\widehat{\mathbf{V}}_1\right) \\ \vdots \\ \mathrm{vec}\left(\mathbf{V}_{n_t}\right) \end{bmatrix}$$

$$\mathbf{U}\mathbf{g} \approx \mathbf{v}$$

Sparsity-promoting wavefield inversion

The Previous Equation can be written as

$$\mathbf{A}\mathbf{x} \approx \mathbf{b}$$

with $\mathbf{A} := \mathbf{U}\mathbf{C}^*$, where $\mathbf{x}$ is the discrete curvelet representation of $g(t, x_s, x_r)$, $\mathbf{C}$ the curvelet transform, and $\widehat{\mathbf{v}}$ the discrete representation of $v(t, x_s, x_r)$.

EPSI

In the case of estimation of primaries

$$\mathbf{U} := \mathbf{F}_t^* \,\text{blockdiag}\left[\widehat{\mathbf{Q}}_{1\cdots n_f} - \widehat{\mathbf{P}}_{1\cdots n_f}\right]\mathbf{F}_t,$$

$\widehat{\mathbf{Q}} = \mathbf{I}\,\widehat{\mathbf{q}}(\omega)$ the temporal Fourier transform of the source function (full rank)
$\widehat{\mathbf{P}}$ the Fourier representation of the up-going wavefield (rank deficient)

SLIM

EPSI

To overcome rank deficiency, we regularize the inversion by exploiting sparsity by solving

$$\begin{cases} \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2 \leq \sigma \\ \widetilde{\mathbf{g}} = \mathbf{S}^* \widetilde{\mathbf{x}} \end{cases}$$

σ : noise-dependent tolerance level.

Dimensionality Reduction Via SVD's

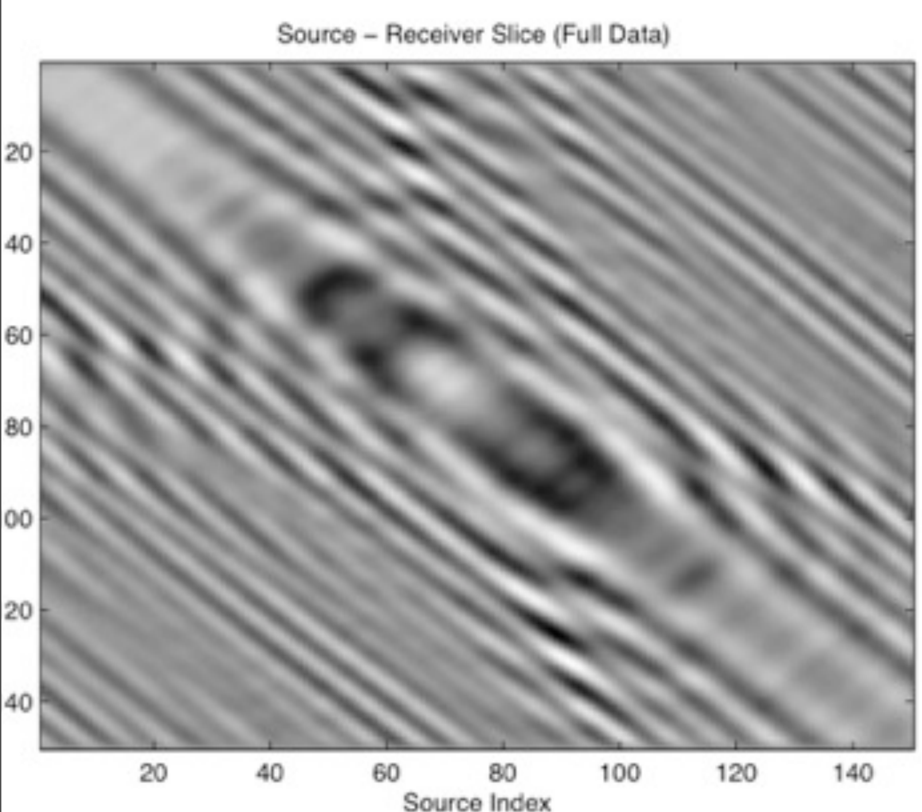
Solving optimization problems require multiple iterations which are challenging because :

- Data matrices are full & extremely large
- Data matrices is incomplete
- Solvers require multiple evaluations of

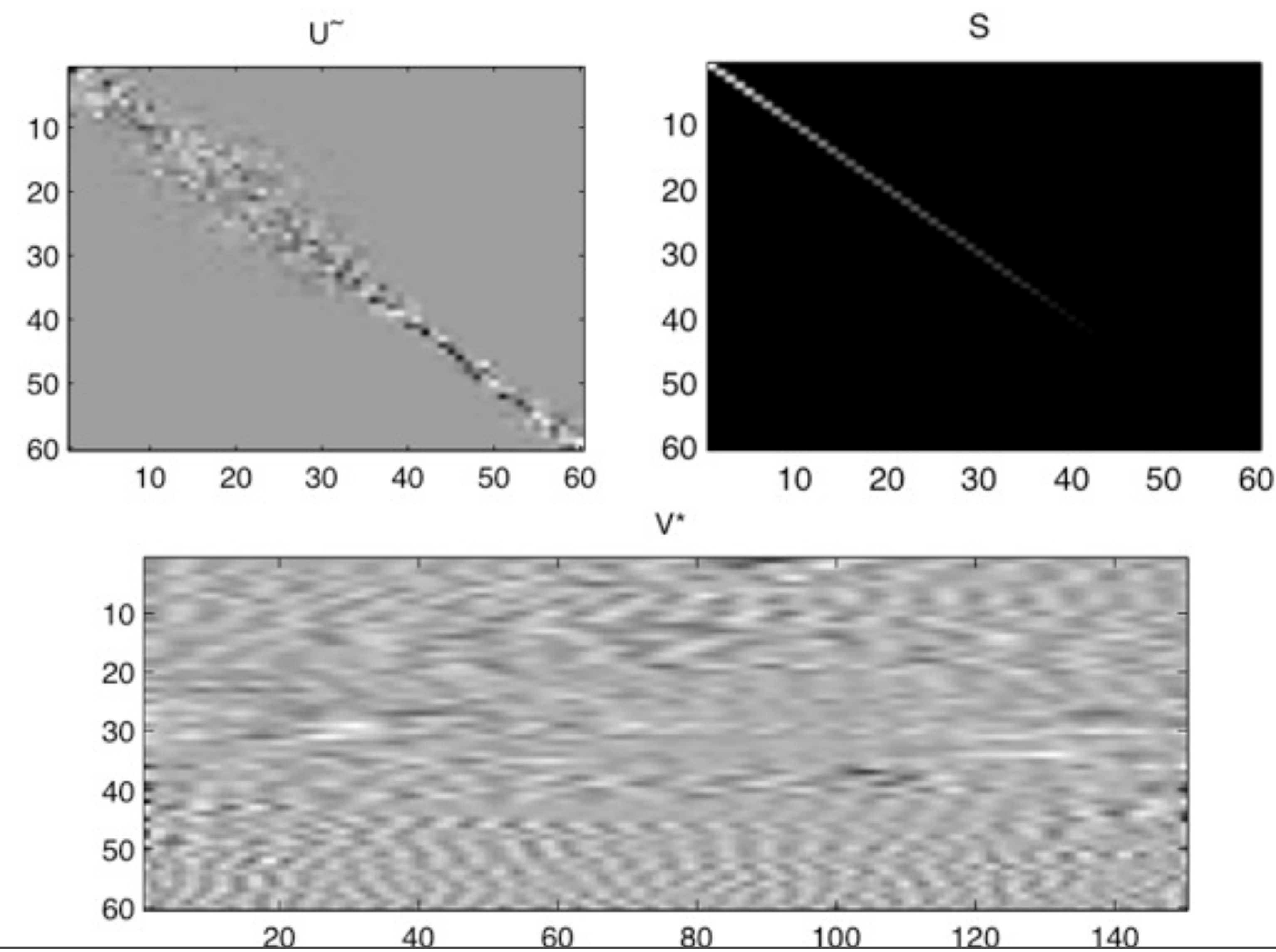
$\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^*\mathbf{A}$

Dimensionality Reduction Via SVD's

Replace the data matrix P with low-rank approximation



=



Dimensionality Reduction Via SVD's

Low Rank Approximation (Randomized SVD)

- 2 Stages:
    1. Capturing the action of the data matrix P
    2. Forming a SVD on the action of P

SLIM

Dimensionality Reduction Via SVD's

- Stage 1 : Capturing the action of P

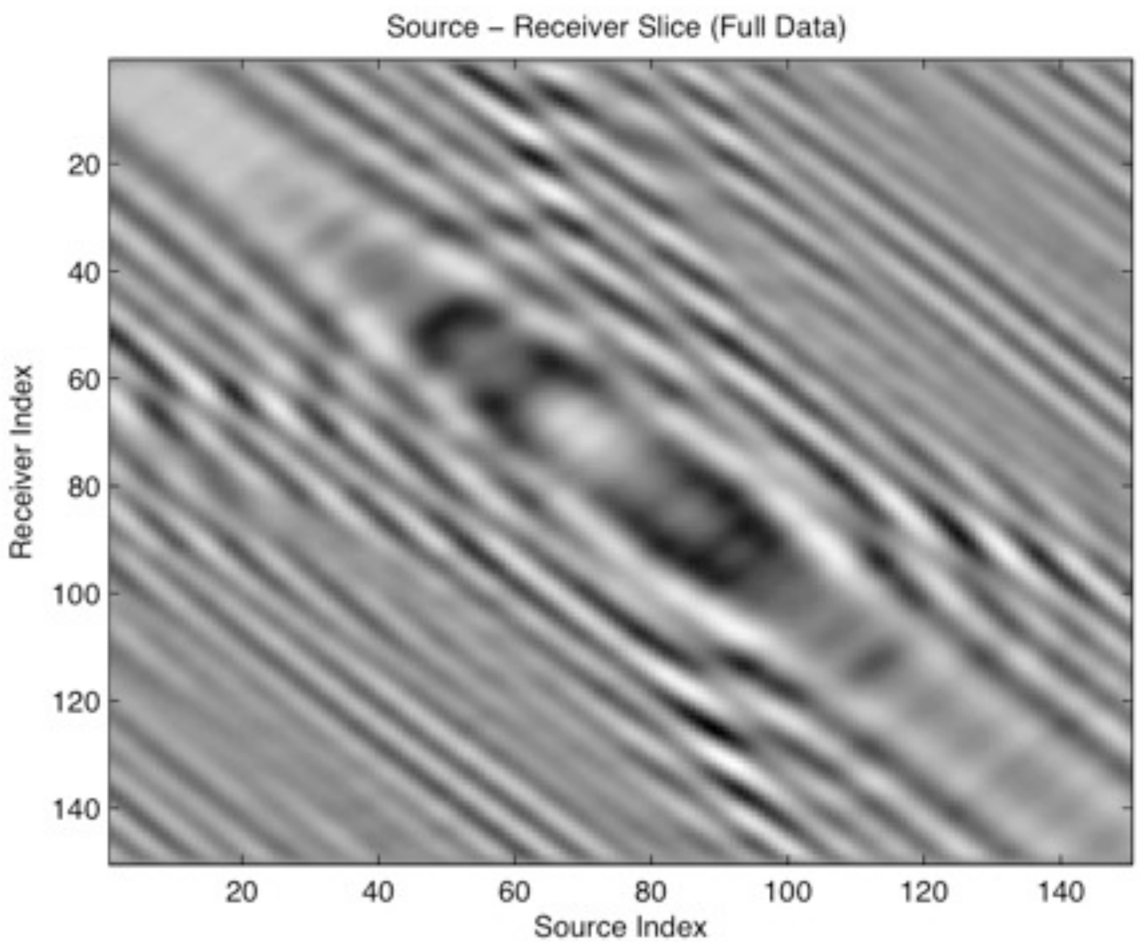
$$\widehat{\mathbf{Y}} = \widehat{\mathbf{P}}\widehat{\mathbf{W}}$$

$\widehat{\mathbf{W}} \in \mathbb{C}^{n_s \times (k+p)}$ a complex-valued Gaussian random matrix with $k+p$ columns and p a small oversampling factor (typically order $5-10$).

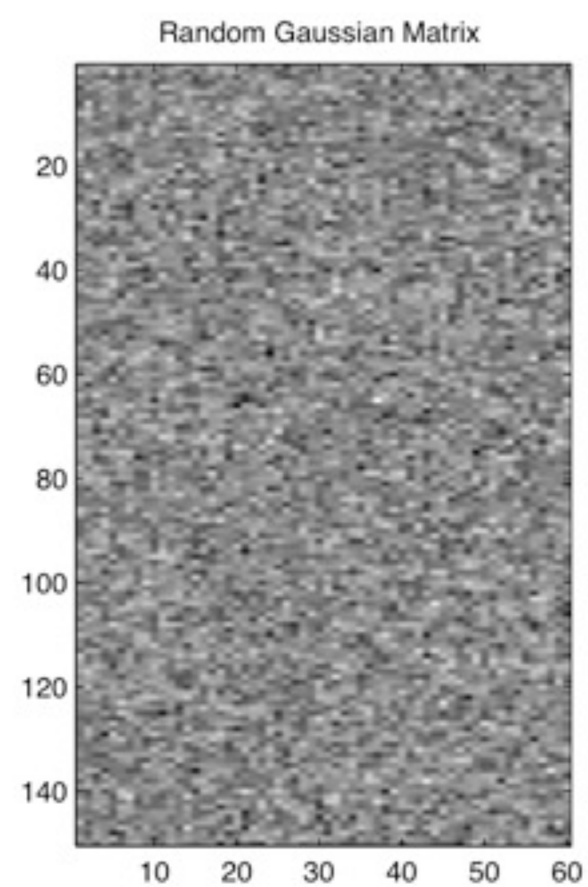
Dimensionality Reduction Via SVD's

- Stage 1 : Capturing the action of P

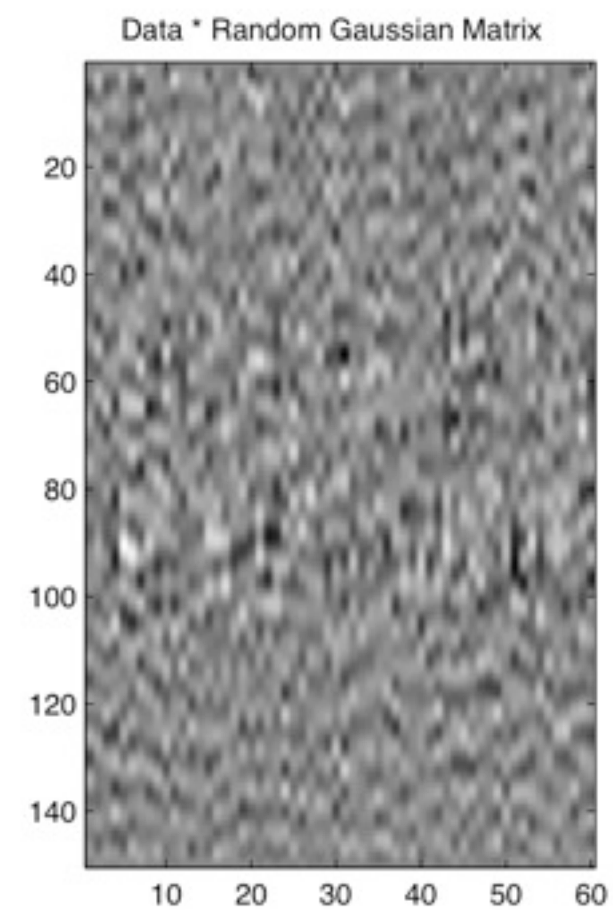
$\mathbf{P}$



$\widehat{\mathbf{W}}$



$\widehat{\mathbf{Y}} = \widehat{\mathbf{P}}\widehat{\mathbf{W}}$

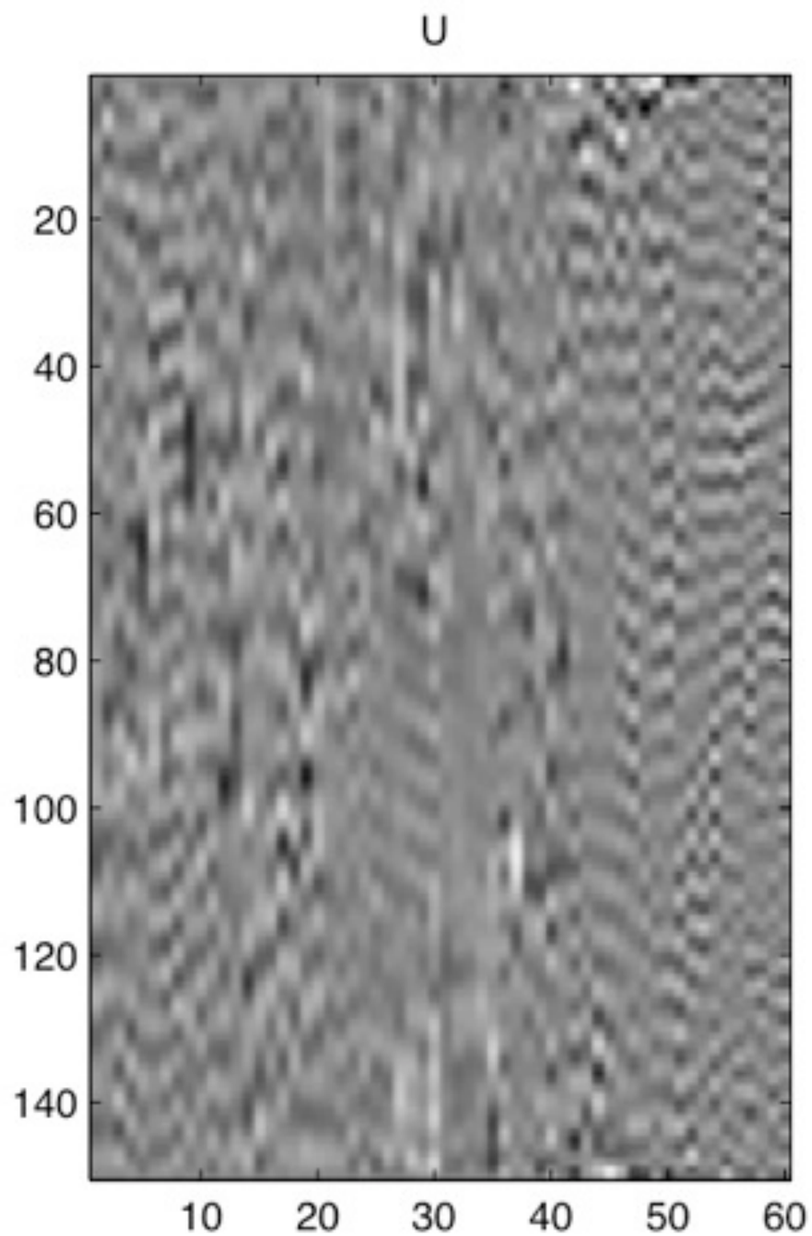


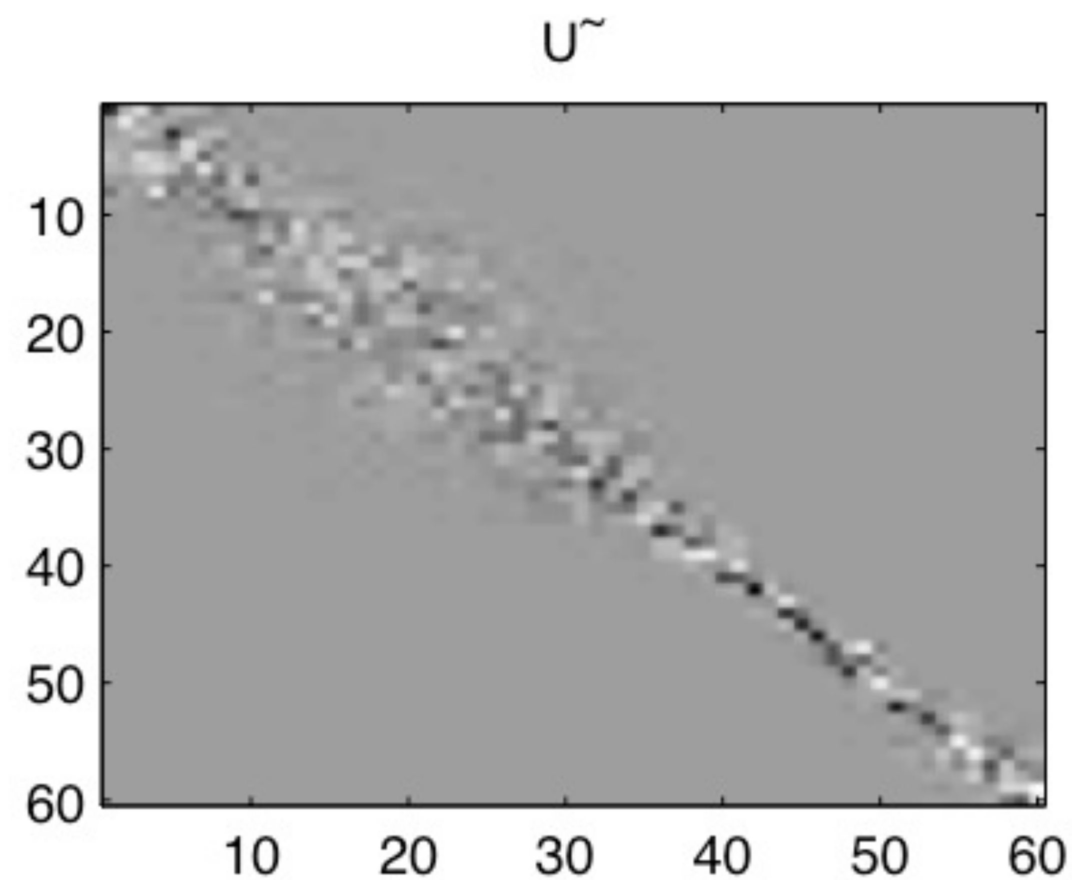
Dimensionality Reduction Via SVD's

- Stage 2 : Compute an approximate SVD of P

1. Form a low-rank factorization $\widehat{\mathbf{P}} \approx \mathbf{QB}$ with $\mathbf{B} = \mathbf{Q}^*\widehat{\mathbf{P}}$ obtained by a QR-factorization of $\widehat{\mathbf{Y}}$.
2. Compute the SVD of the small matrix $\mathbf{B} = \widetilde{\mathbf{U}}\mathbf{S}\mathbf{V}^*$.
3. Compute $\mathbf{U} = \mathbf{Q}\widetilde{\mathbf{U}}$.

SLIM


U
20
40
60
80
100
120
140
10
20
30
40
50
60


U~
10
20
30
40
50
60
10
20
30
40
50
60

Dimensionality Reduction Via SVD's

Low k-rank (with $k \ll \min(n_s, n_r)$) approximation of the action of the data matrix

$$\widehat{\mathbf{P}} \approx \mathbf{USV}^*$$

$$\mathbf{U} \in \mathbb{C}^{n_s \times k},\ \mathbf{S} \in \mathbb{C}^{k \times k},\ \text{and}\ \mathbf{V} \in \mathbb{C}^{n_s \times k}$$

Dimensionality Reduction Via SVD's

Advantages

- Faster $\widehat{\mathbf{P}}\widehat{\mathbf{P}}^*$ and $\widehat{\mathbf{P}}^*\widehat{\mathbf{P}}$ multiplications
- $\widehat{\mathbf{P}}^*\widehat{\mathbf{P}} \approx \mathbf{V}\mathbf{S}^2\mathbf{V}^*$
- Significantly reduced memory imprint

Putting Things Together

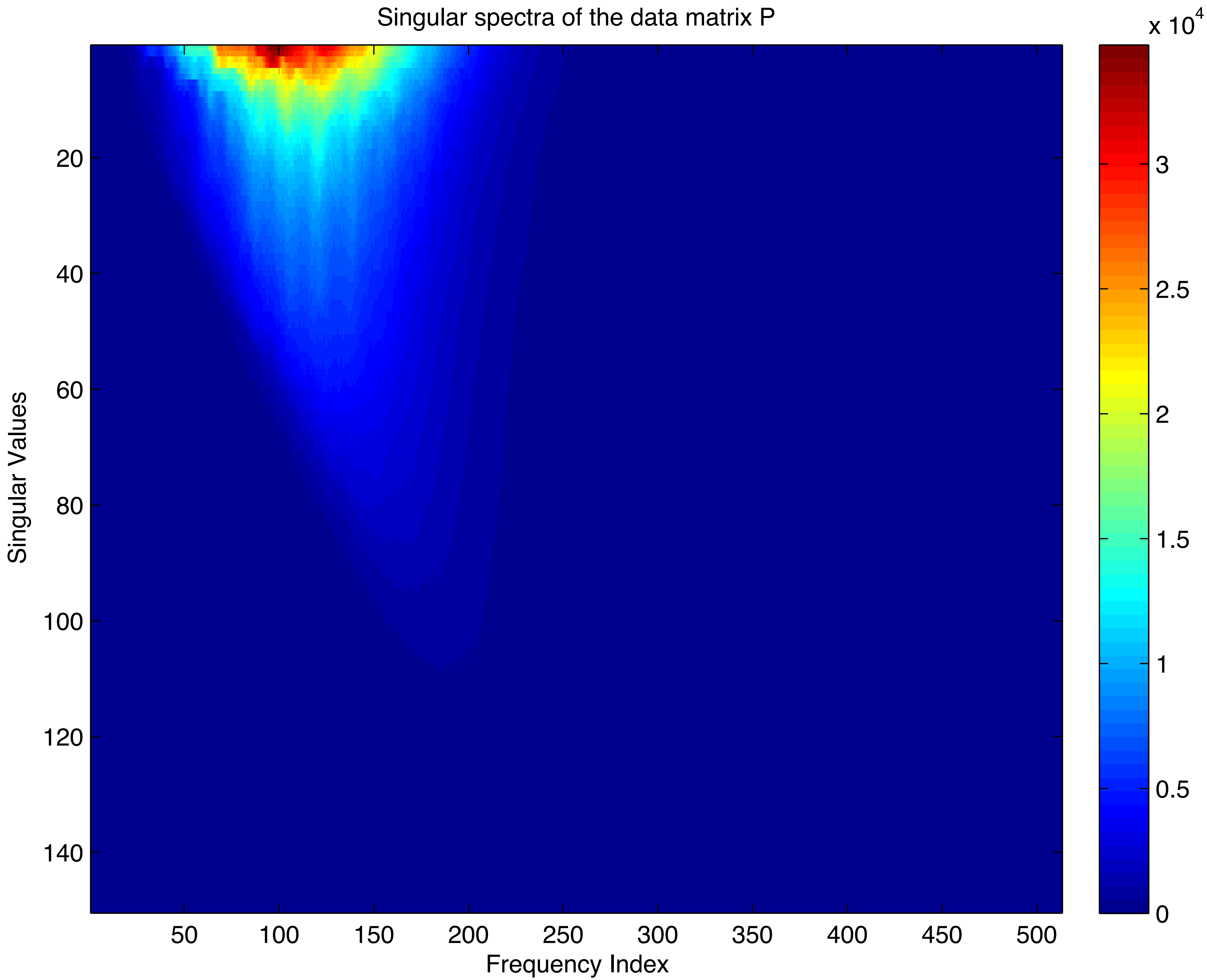
Inorder to combine EPSI and the low rank approximation

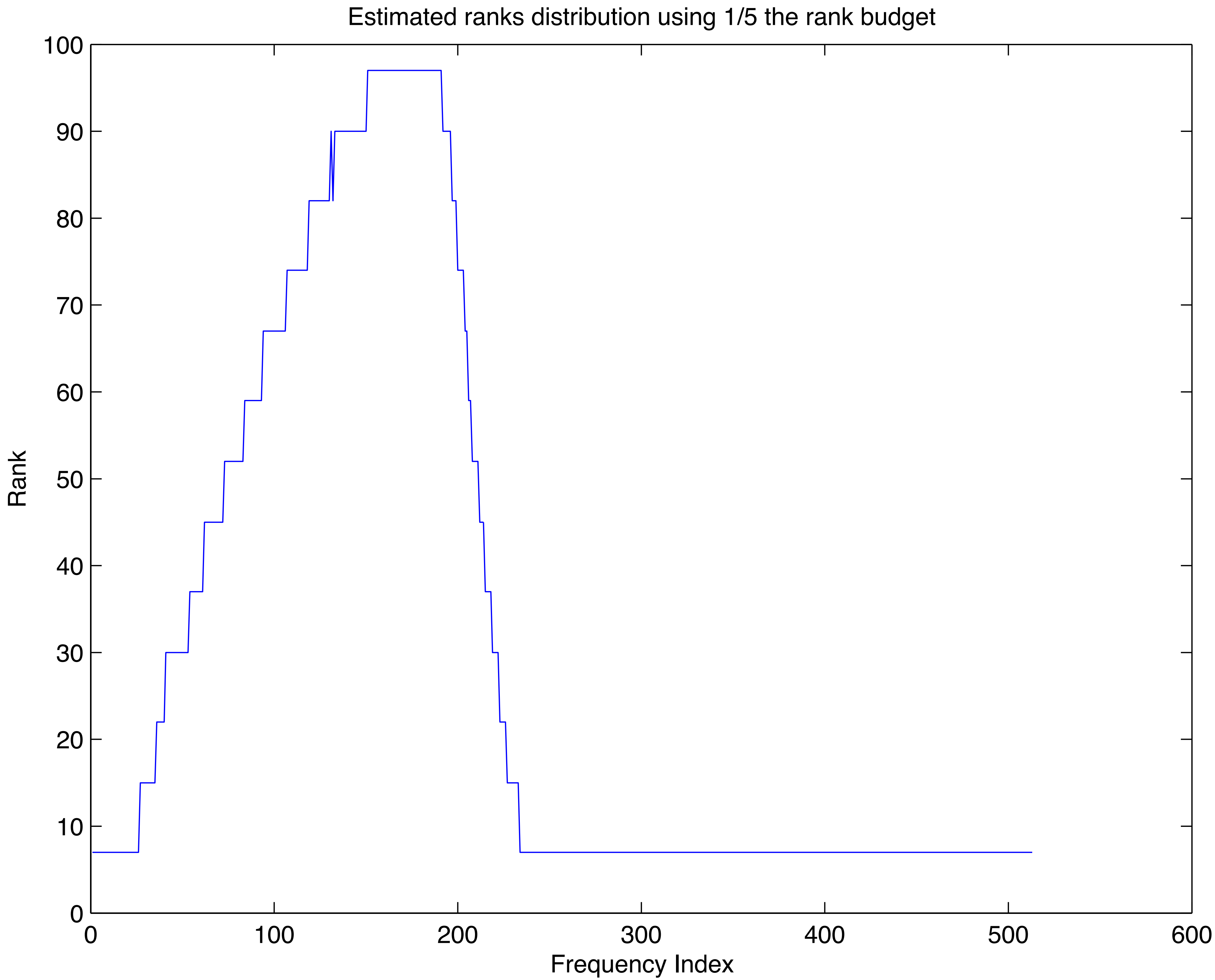
- Balance between
  - Accuracy
  - Memory reduction
  - Matrix multiplication speed up

SLIM

Putting Things Together

- Spectral/Operator Norm $\|\cdot\|_S$
  - Maximum Eigen - Value
- Numerical Rank depends on frequency
- Amplitude spectrum of seismic wavelet varies with frequency
  - Adaptive rank selection


Singular spectra of the data matrix P
Singular Values
20
40
60
80
100
120
140
Frequency Index
50
100
150
200
250
300
350
400
450
500
x 10^4
3
2.5
2
1.5
1
0.5
0


Estimated ranks distribution using 1/5 the rank budget
Rank
Frequency Index
0
10
20
30
40
50
60
70
80
90
100
0
100
200
300
400
500
600

PERFORMANCE

Up front cost of Low-Rank Approximation

- Classical methods

$$\mathcal{O}(n_r \times n_s \times K)$$

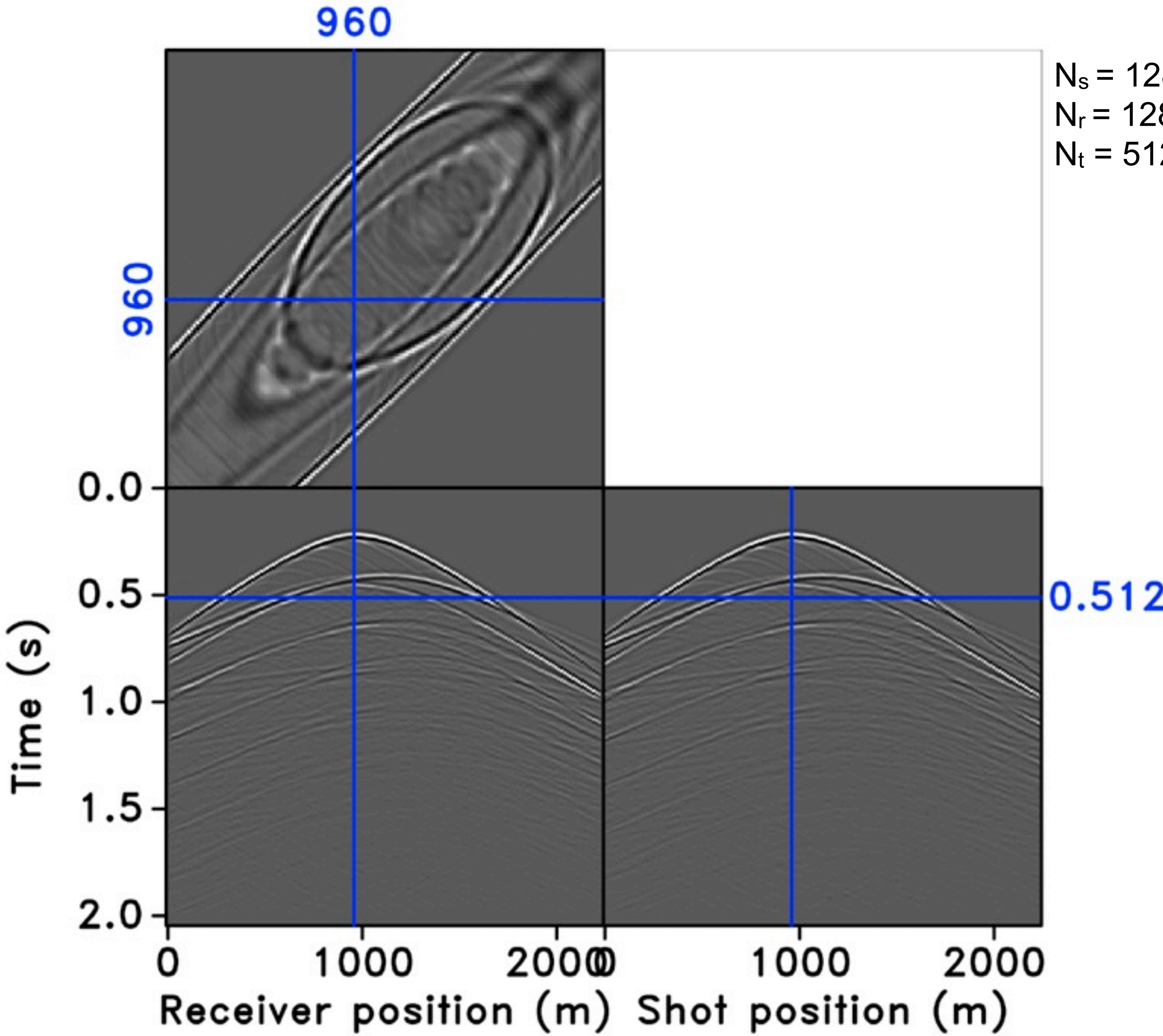
- Randomized methods

$$\mathcal{O}(n_r \times n_s \times \log K)$$

PERFORMANCE

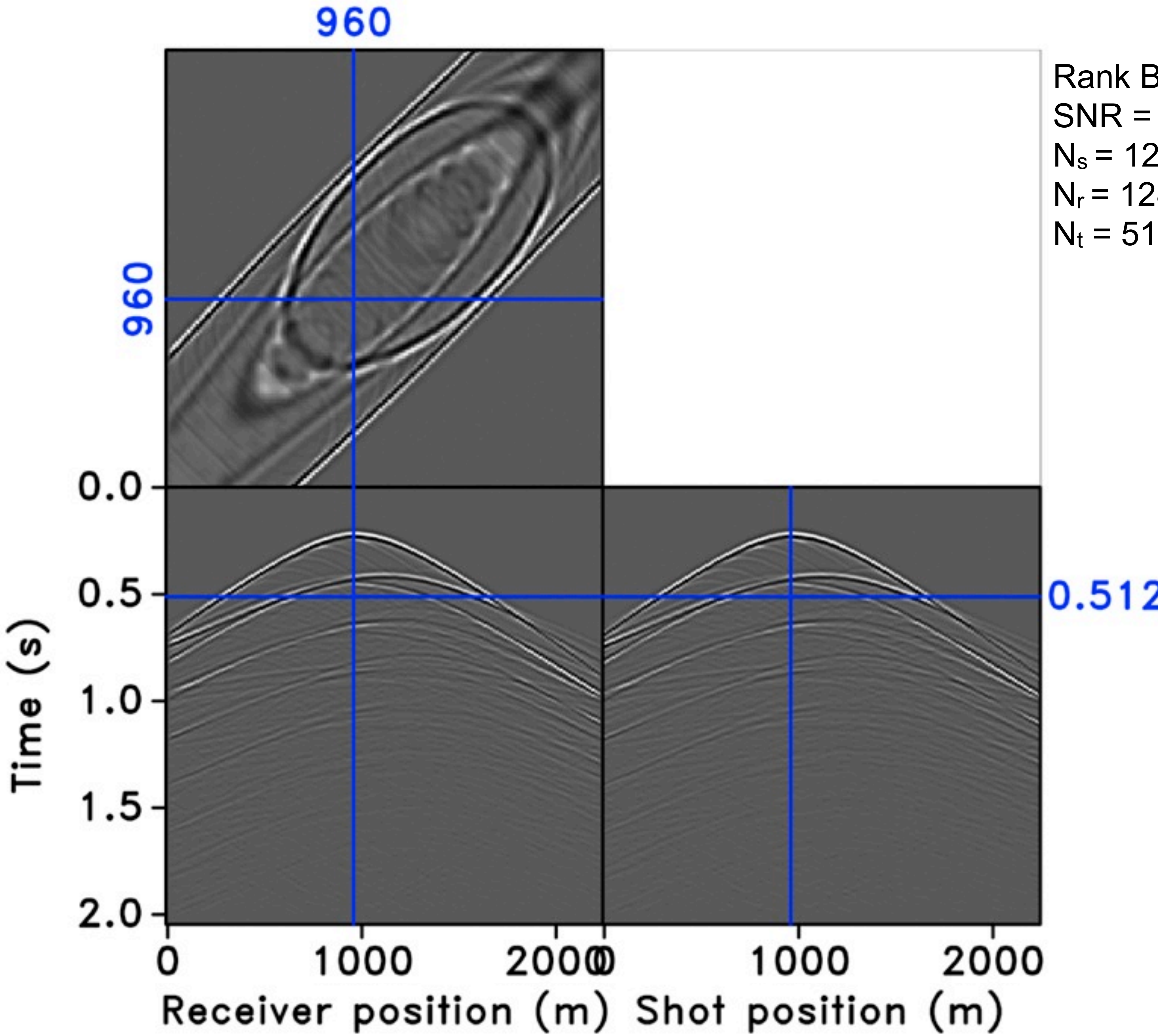
Subsample ratio δ	1/2	1/5	1/8	1/12
	recovery error (dB) / spectral norms ($\times 10^3$)			
	88 (44)	**20** (121)	**16** (144)	**13** (152)
Speed up ($\times$)	2	5	8	12

Synthetic Data



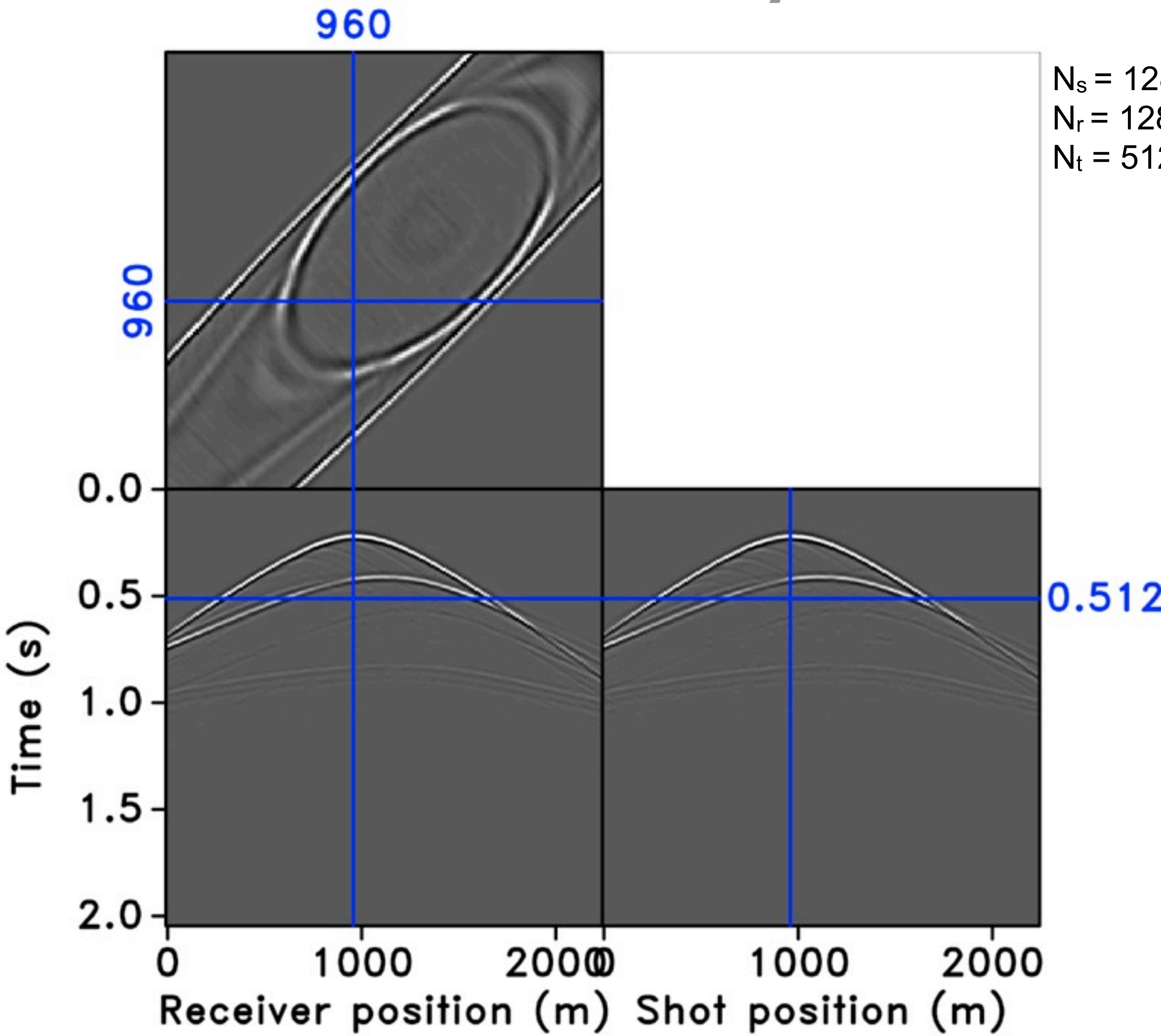
$N_s = 128$
$N_r = 128$
$N_t = 512$

Synthetic Data (Approximated)



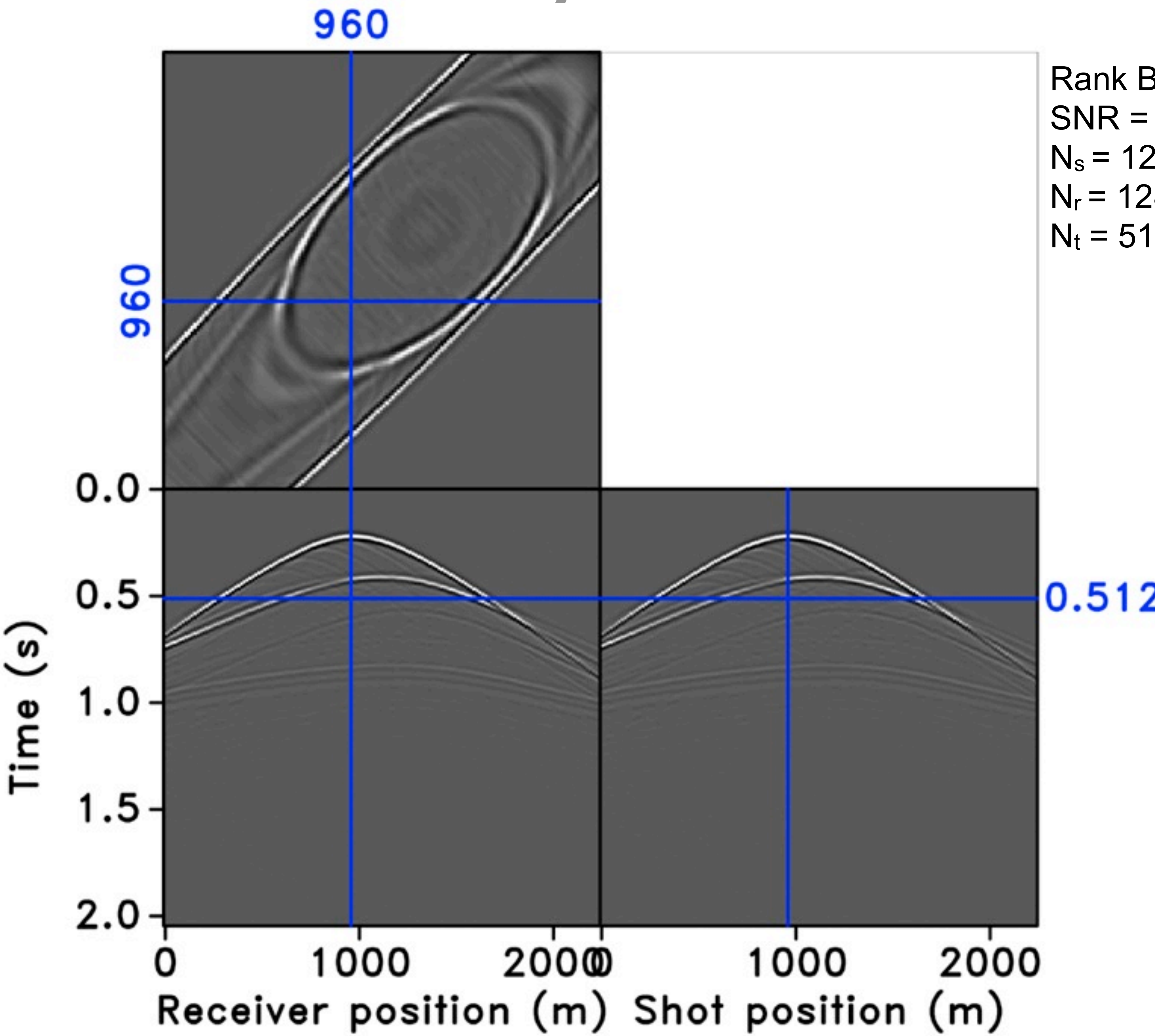
Rank Budget =1/5
SNR = 17 dB
N_s = 128
N_r = 128
N_t = 512

EPSI :Primary



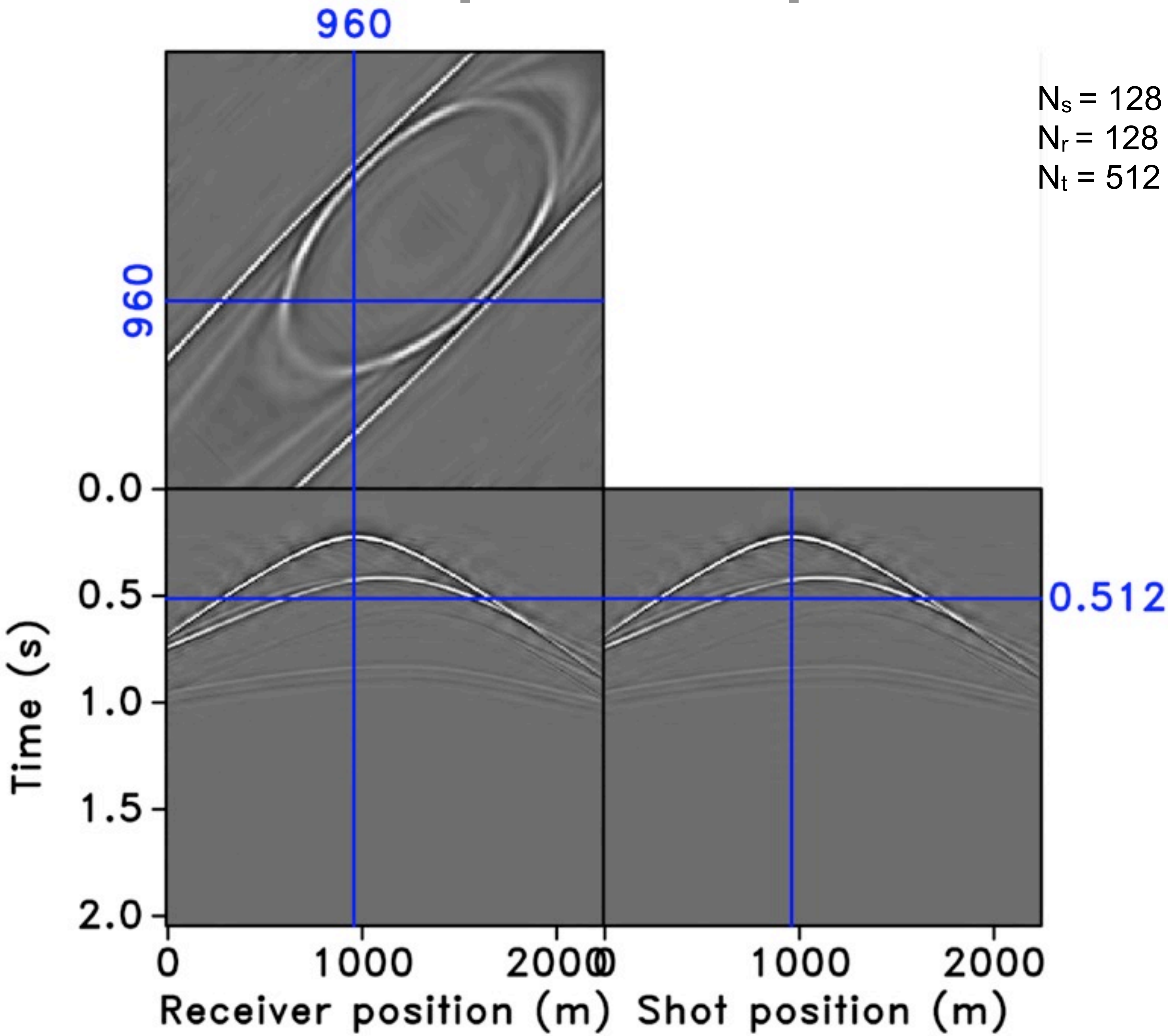
$N_s = 128$
$N_r = 128$
$N_t = 512$

EPSI : Primary (Approximated)



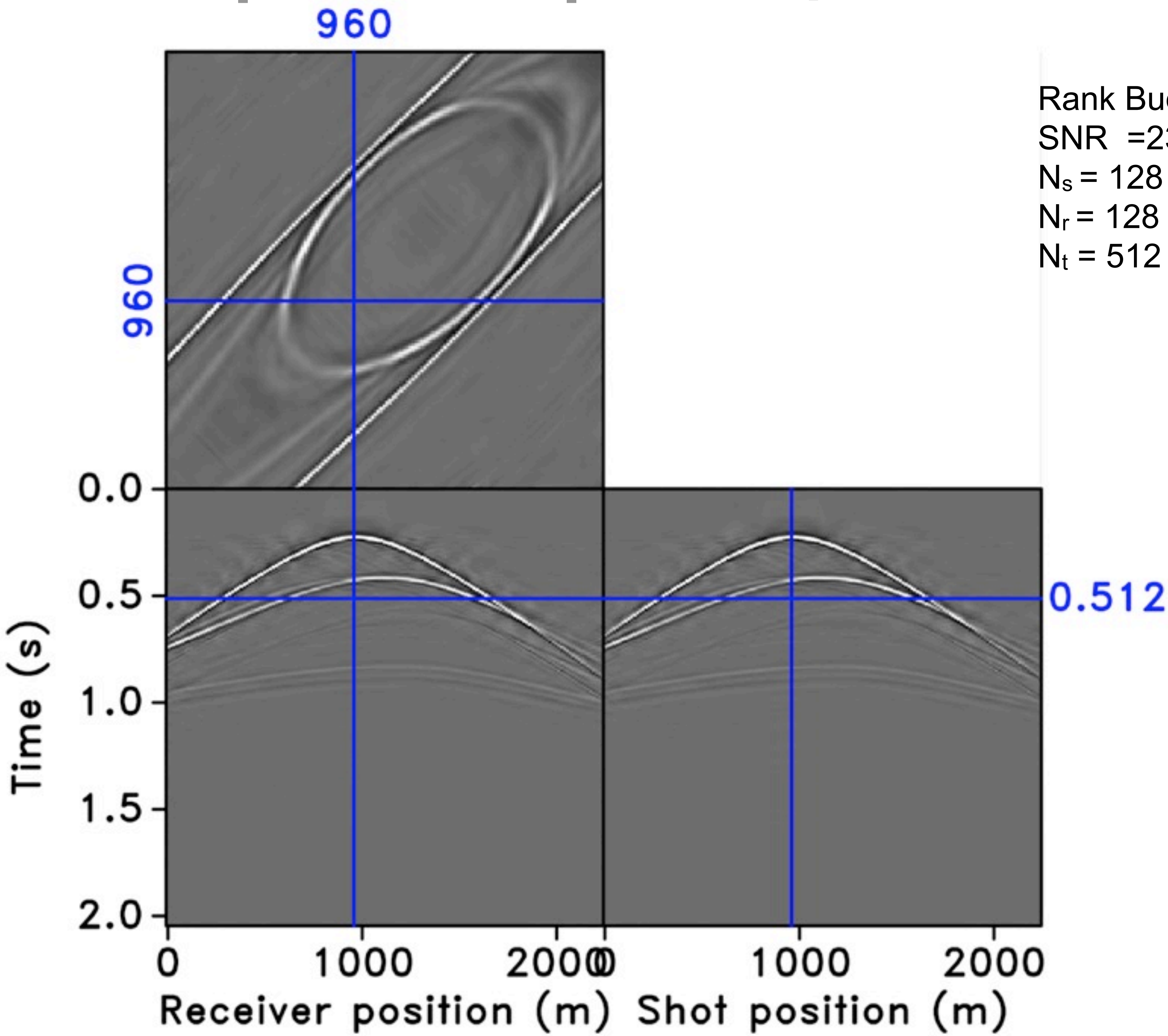
Rank Budget =1/5
SNR = 28 dB
N_s = 128
N_r = 128
N_t = 512

EPSI : Impulse Response



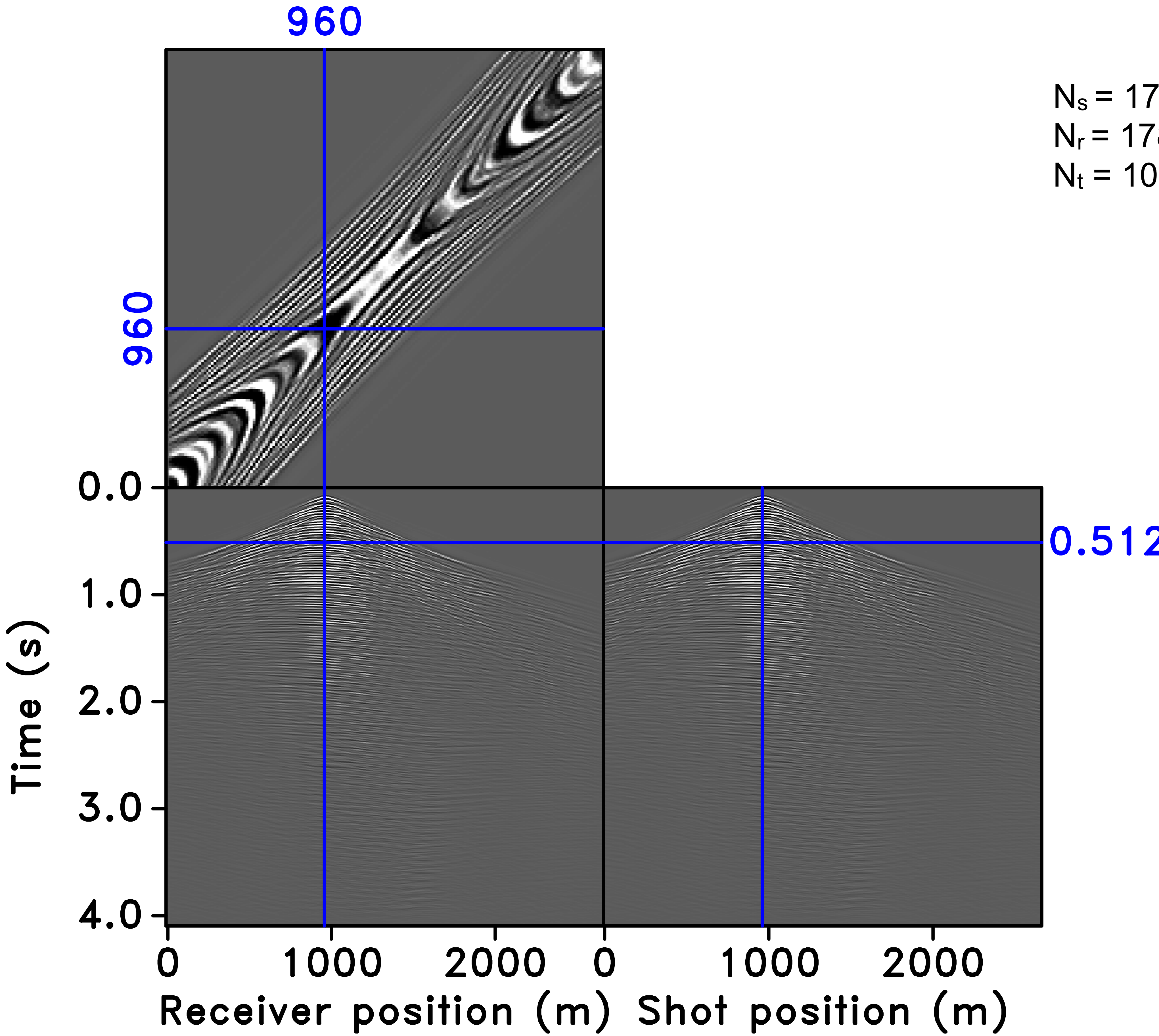
$N_s = 128$
$N_r = 128$
$N_t = 512$

EPSI : Impulse Response (Approximated)



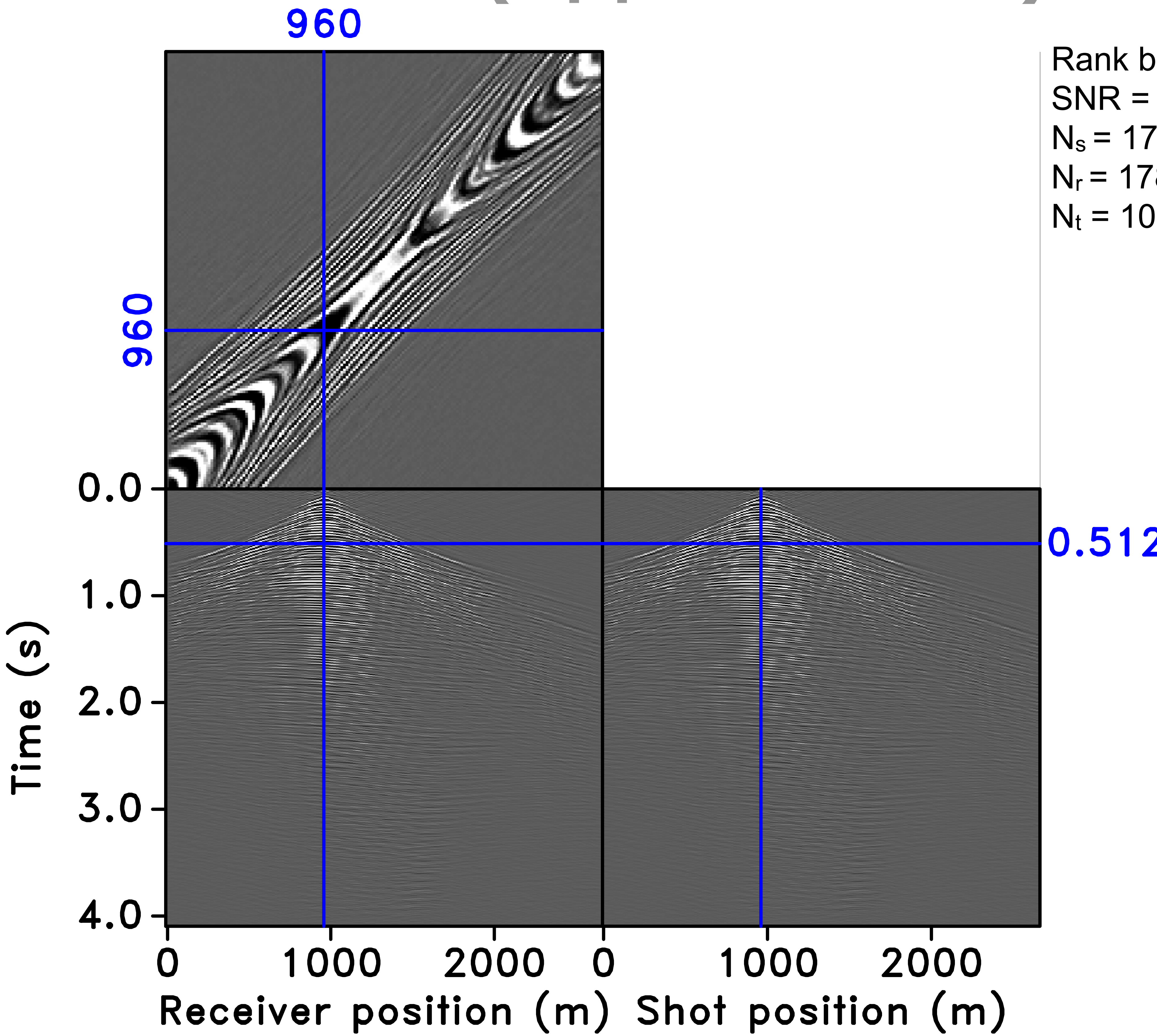
Rank Budget =1/5
SNR =23 dB
N_s = 128
N_r = 128
N_t = 512

Real Data



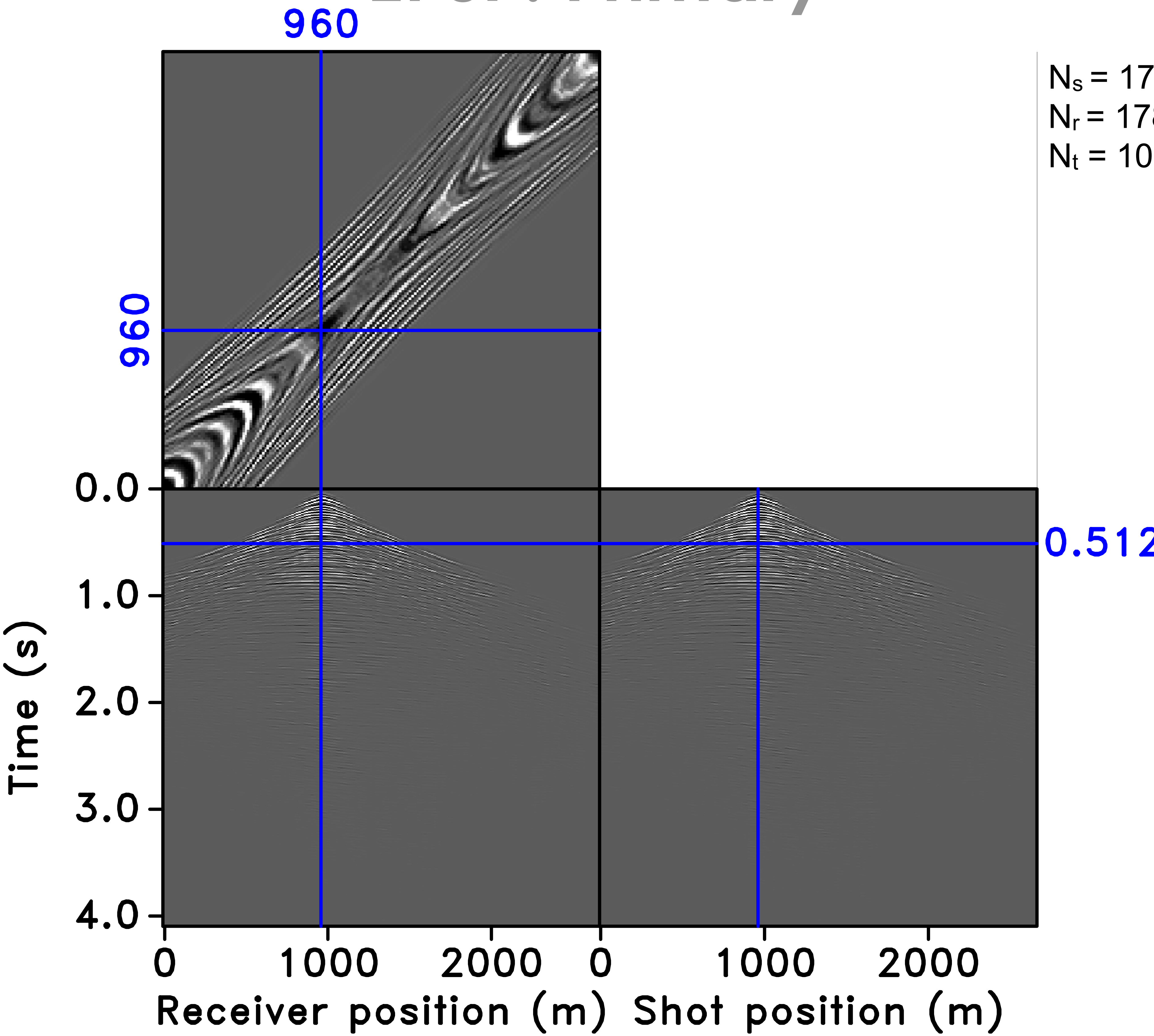
N_s = 178
N_r = 178
N_t = 1024

Real Data (Approximated)



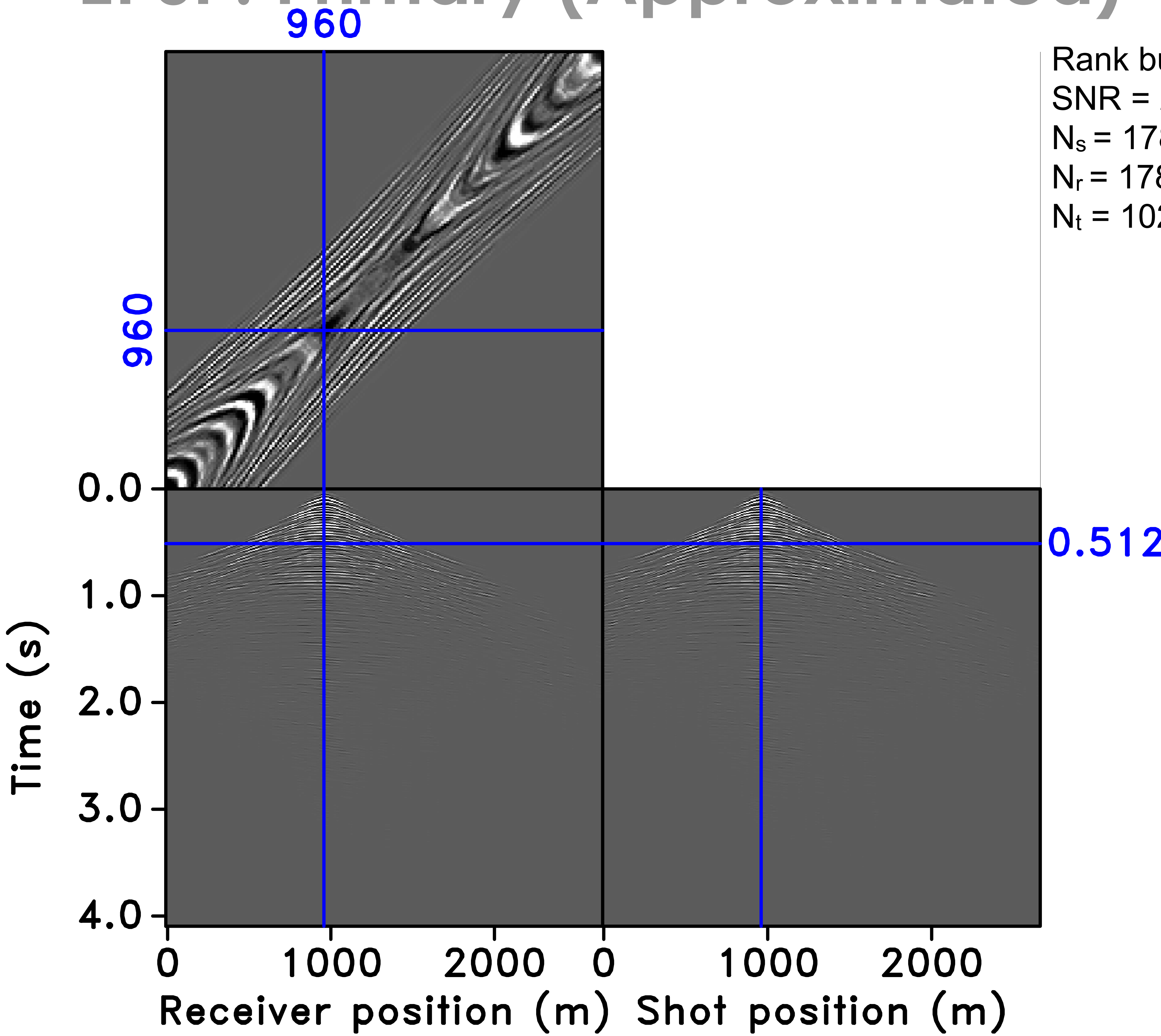
Rank budget =1/4
SNR = 13 dB
N_s = 178
N_r = 178
N_t = 1024

EPSI : Primary



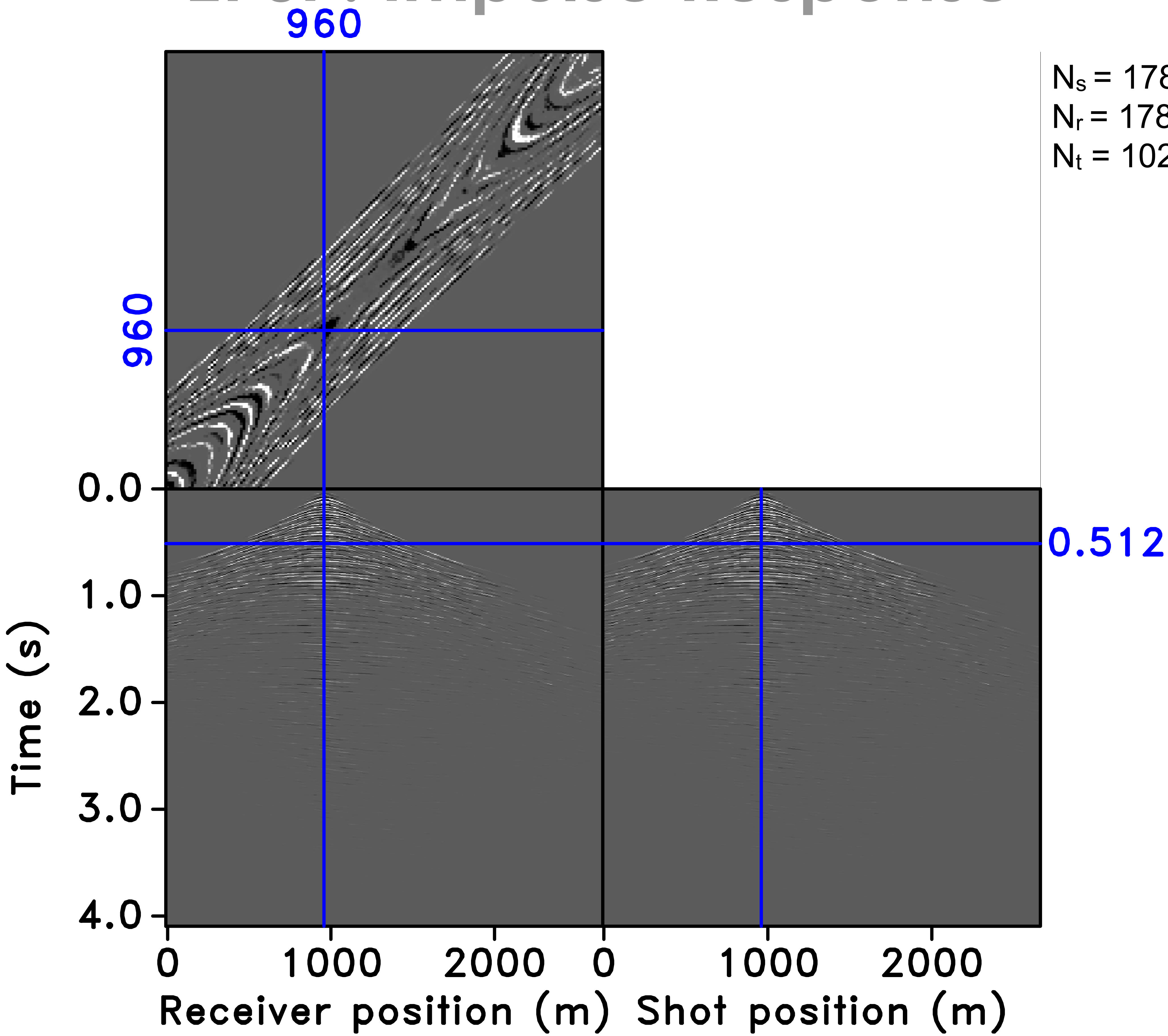
$N_s = 178$
$N_r = 178$
$N_t = 1024$

EPSI : Primary (Approximated)



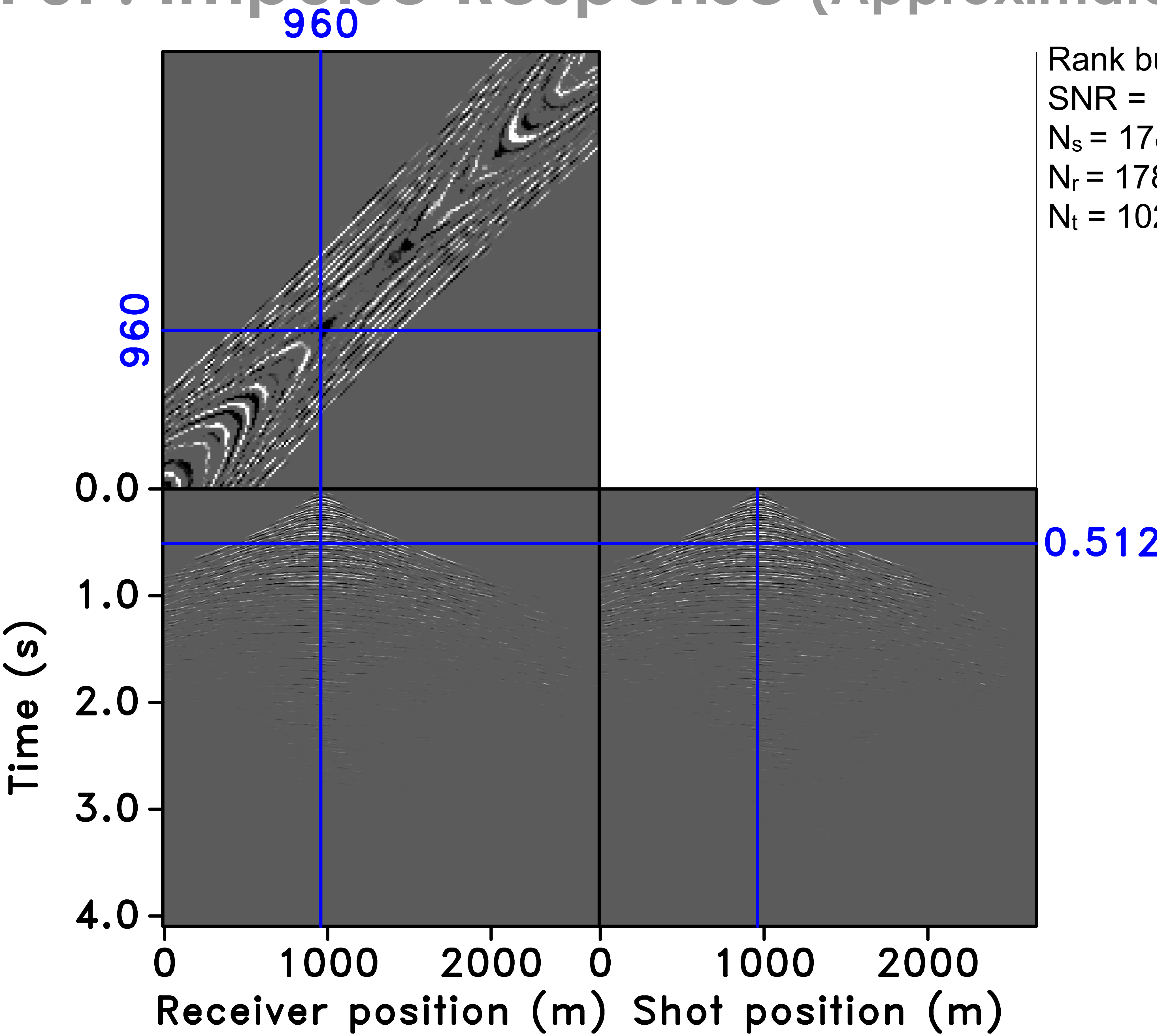
Rank budget =1/4
SNR = 21 dB
N_s = 178
N_r = 178
N_t = 1024

EPSI : Impulse Response



$N_s = 178$
$N_r = 178$
$N_t = 1024$

EPSI : Impulse Response (Approximated)



Rank budget =1/4
SNR = 13 dB
N_s = 178
N_r = 178
N_t = 1024

SLIM

Conclusion

- Data driven methods - e.g. EPSI - suffers from the 'curse of dimensionality'
- We utilize insights from random matrix theory to approximate action of the data matrix
- Reductions in multiplication and storage costs
- Up-Front cost is cheap
- Can be implemented in parallel
- Instance of compressive Sensing

Future Work

- Application of low rank approximation on 3D data
- Parallel implementation of the randomized approximation techniques
- Extending EPSI to work with 3D data

Acknowledgements

- 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 II (375142-08).
- This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BP, Chevron, ConocoPhillips, Petrobras, Total SA, and WesternGeco.
  E. J. Candès, L. Demanet, D. L. Donoho, and L. Ying for CurveLab (www.curvelet.org)
- E. van der Berg and M. Friedlander for SPGl1 (www.cs.ubc.ca/labs/scl/spgl1/)

SLIM

Thanks

slim.eos.ubc.ca