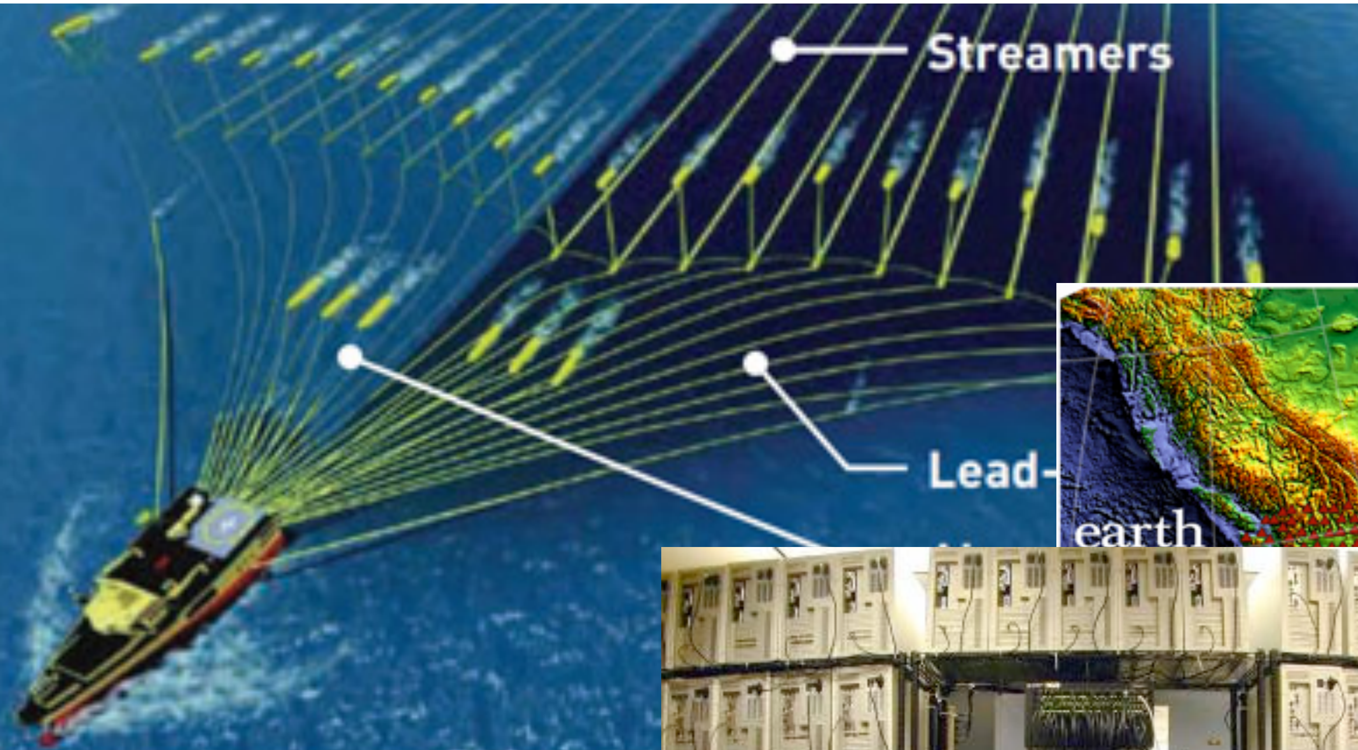
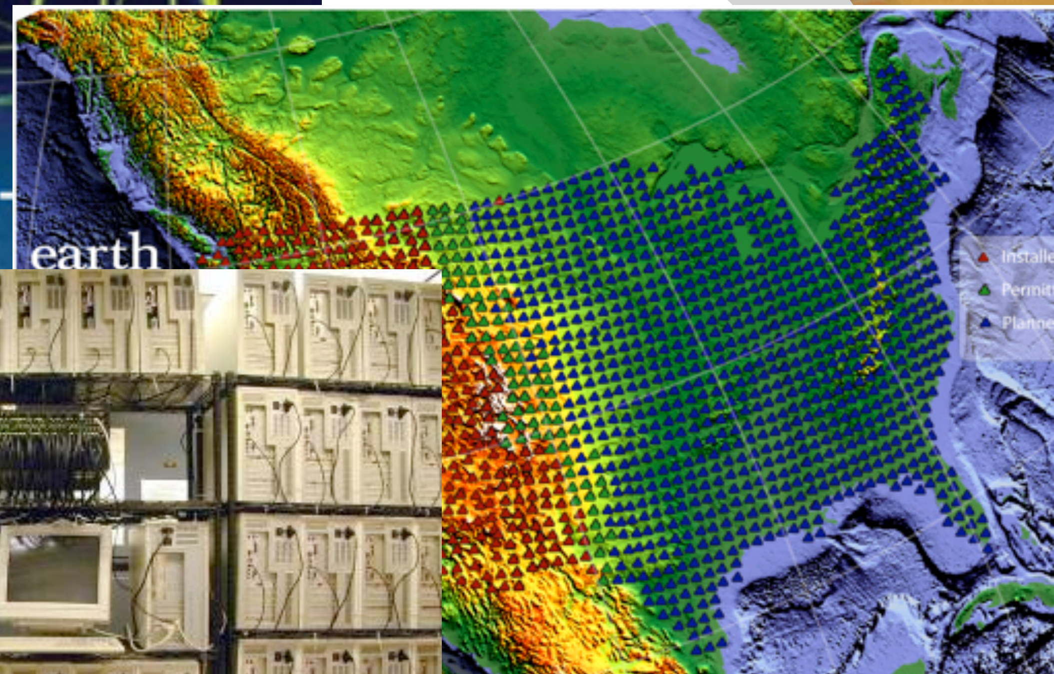
A hybrid stochastic-deterministic method for waveform inversion

T. van Leeuwen, M. Schmidt, M. Friedlander and F.J. Herrmann

SLIM
University of British Columbia

SLIM





SLIM

least-squares fitting of *multi-experiment* data that are linear in the source:

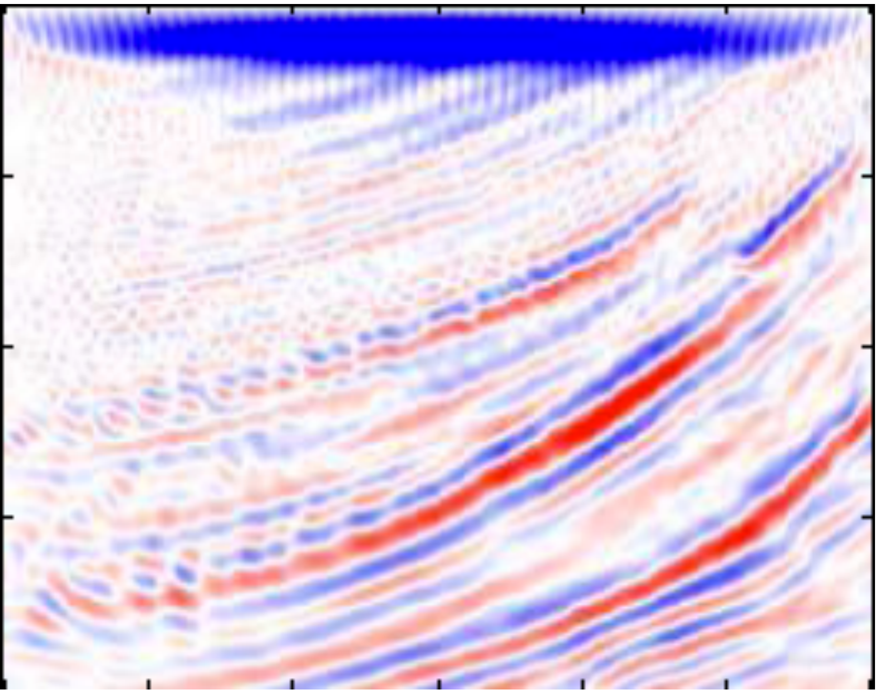
- Costs are proportional to # of sources
- Can be reduced by employing simultaneous sources

- *Computational costs* can be reduced by *synthesizing* simultaneous sources from sequential data

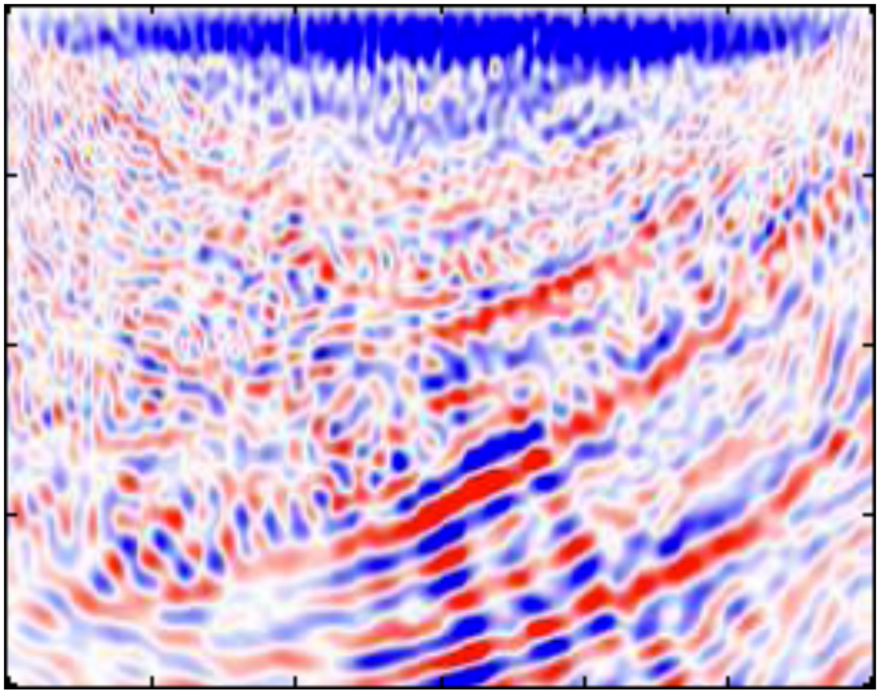
[Beasley `98; Berkhout `08; Romero `00;]
[Ikelle `07; Neelamani `08;Herrmann `09]
[Krebs `09; Haber `10; TvL `10; Ben-Hadj-Ali `11]

What about `cross-talk'?

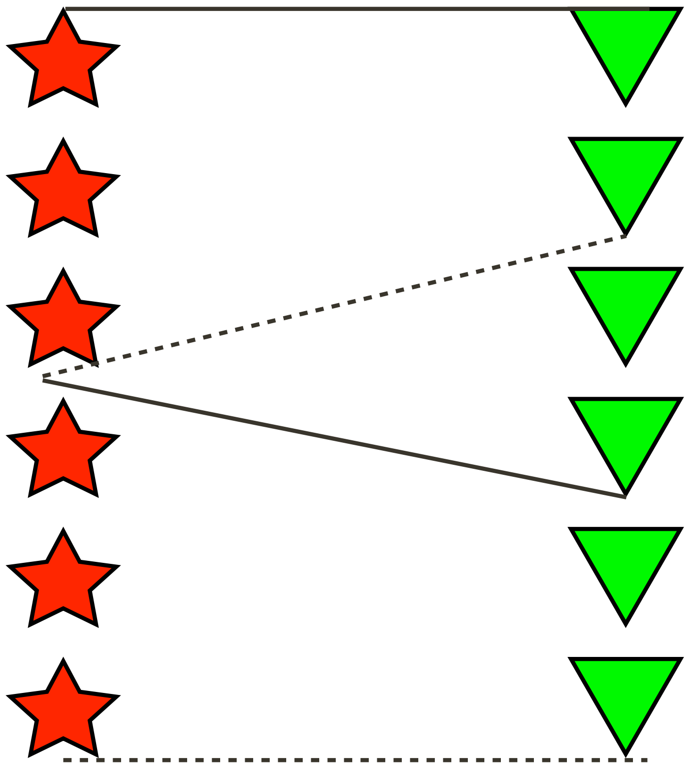
$$\sum_i \mathbf{u}_i \otimes \mathbf{v}_i$$



$$\left(\sum_i \mathbf{u}_i\right) \otimes \left(\sum_i \mathbf{v}_i\right)$$



What about `moving' receiver arrays?



SLIM

Overview

- Optimization problem
- Conventional optimization
- Stochastic optimization
- Hybrid method
- Results
- Conclusions

Optimization problem

Least-squares fitting of *multi-experiment* data

$$\min_{\mathbf{m}} \Phi[\mathbf{m}] = \frac{1}{K}\sum_{i=0}^{K-1} \phi_i[\mathbf{m}]$$

$$\phi_i[\mathbf{m}] = ||\mathbf{d}_i - F[\mathbf{m}]\mathbf{q}_i||_2^2$$

Typically, costs are proportional to K

Source encoding

Replace sequential sources by one simultaneous source: $\widetilde{\mathbf{q}} = \sum_j w_j \mathbf{q}_j$

$$\widetilde{\phi}[\mathbf{m}] = ||\widetilde{\mathbf{d}} - F[\mathbf{m}]\widetilde{\mathbf{q}}||_2^2$$

if $\mathsf{E}\{w_i w_j\} = \delta_{ij}$ we get $\mathsf{E}\left\{\widetilde{\phi}[\mathbf{m}]\right\} = \Phi[\mathbf{m}]$

$$\widetilde{\Phi}[\mathbf{m}] = \frac{1}{K}\sum_{i=0}^{K-1} ||\widetilde{\mathbf{d}}_i - F[\mathbf{m}]\widetilde{\mathbf{q}}_i||_2^2$$

[Beasley `98; Romero `00;]

Source encoding

Choice of random weights:

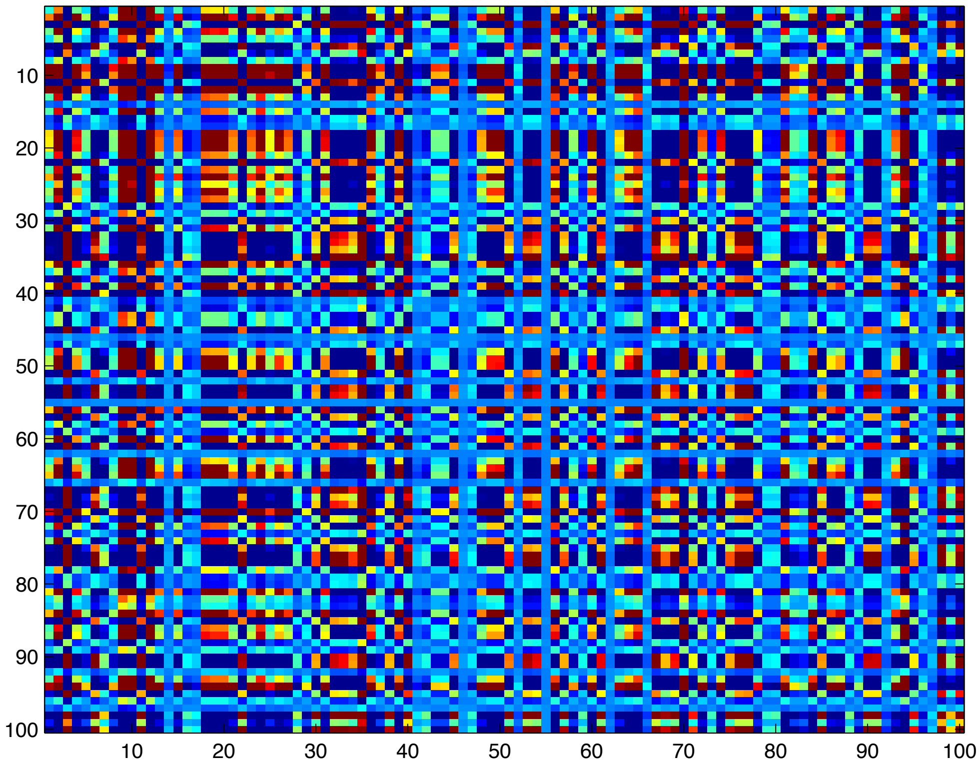
- Gaussian, ± 1 , random phases: *efficient in sampling the whole matrix, but problematic for moving receiver array*

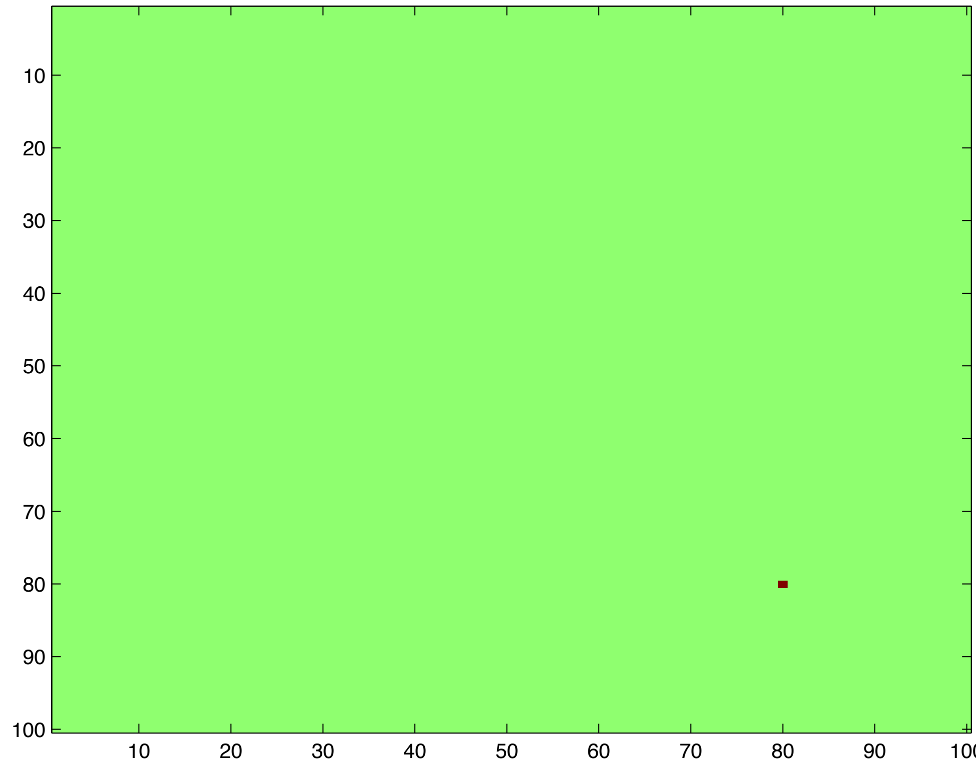
- Random unit vector: *less efficient sampling, but applicable for moving receiver array!*

Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

K=1

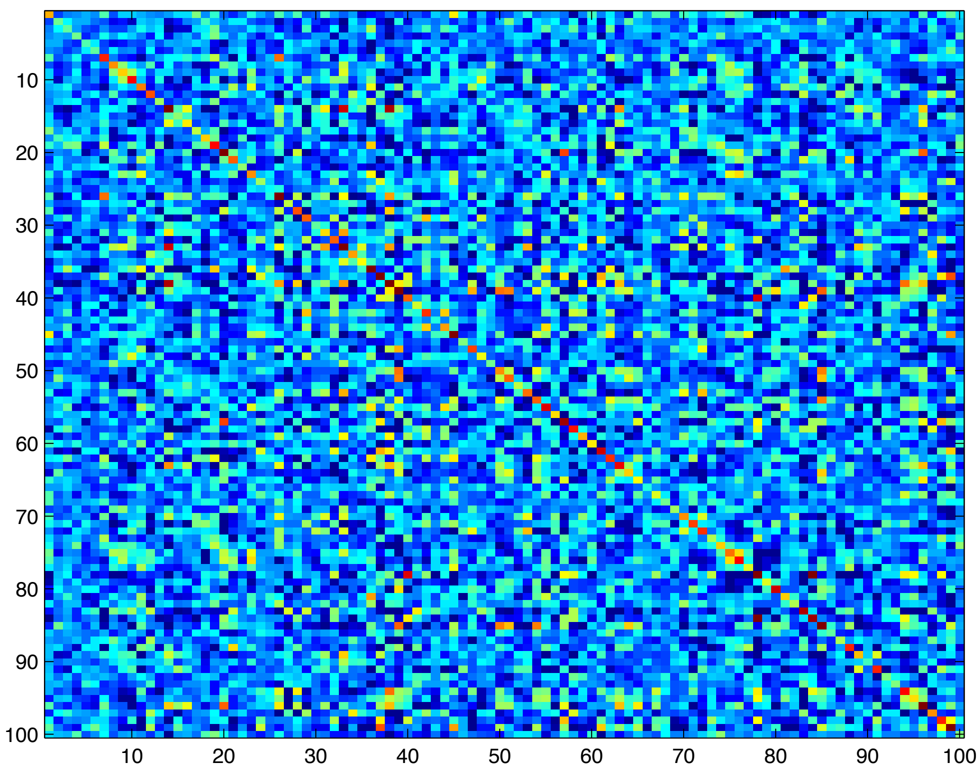


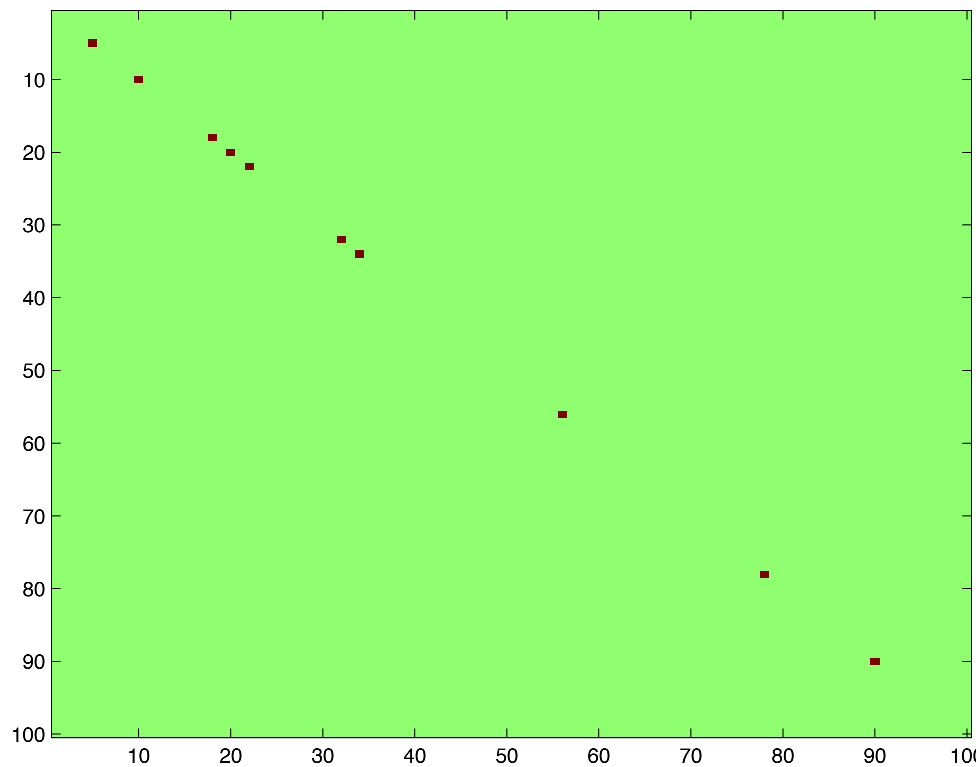


Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

K=10

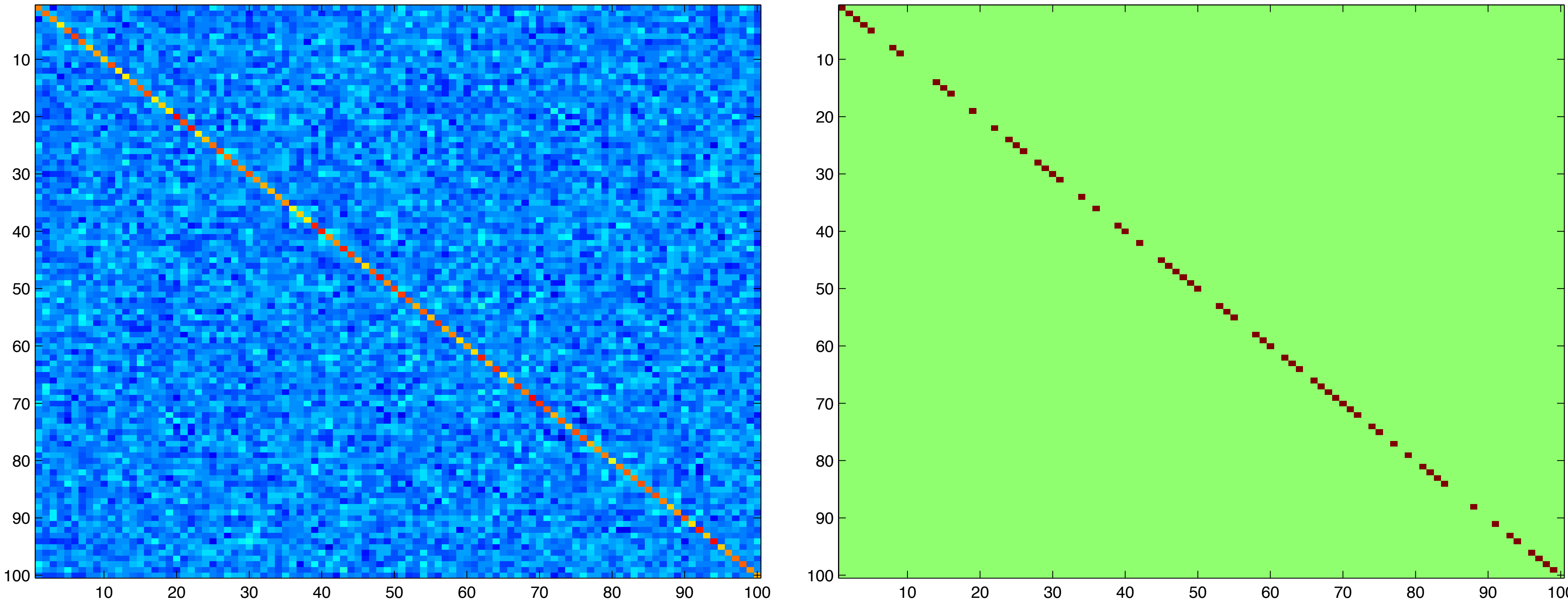




Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

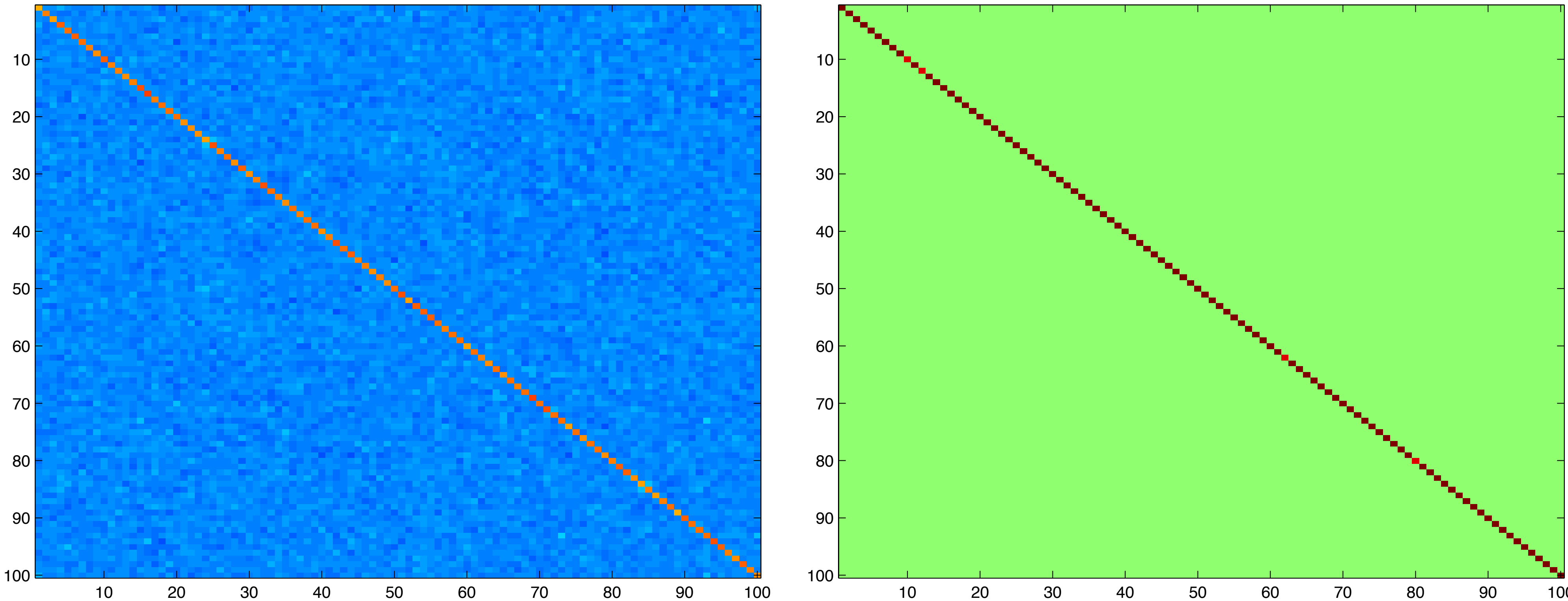
K=100



Source encoding

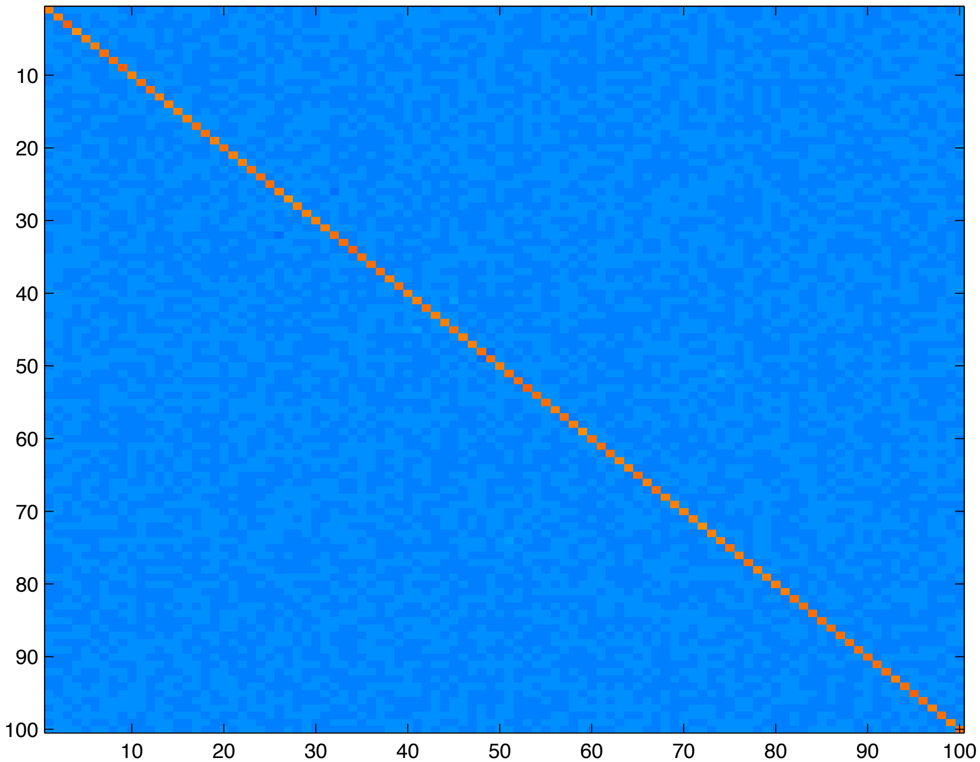
$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I$$

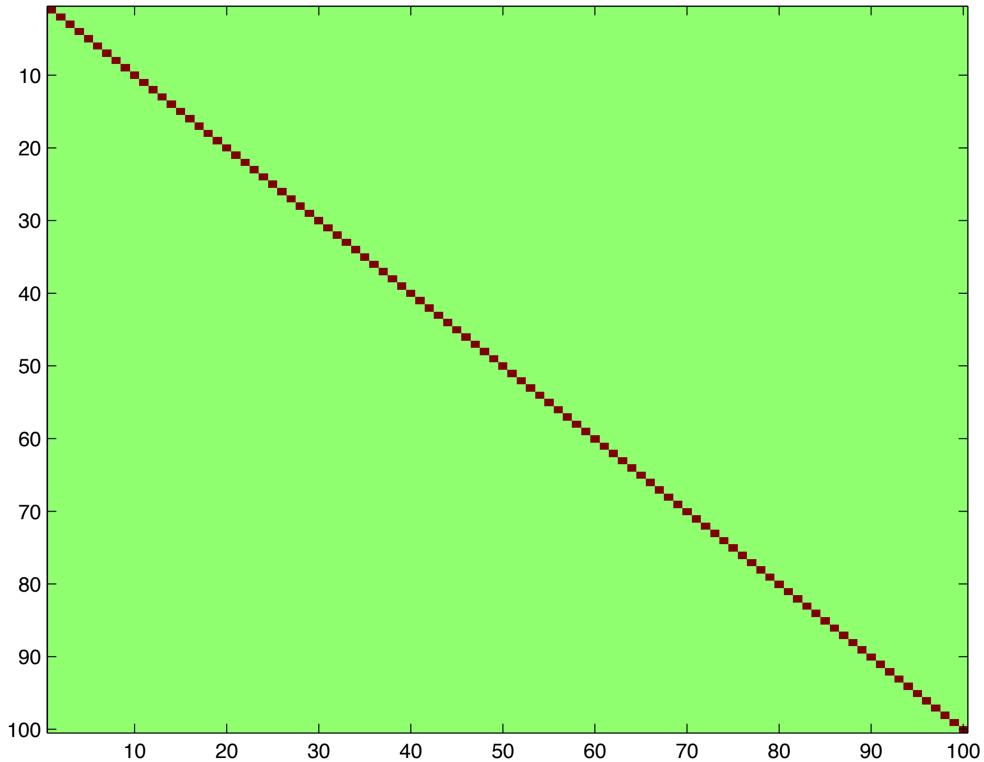
K=1000



Source encoding

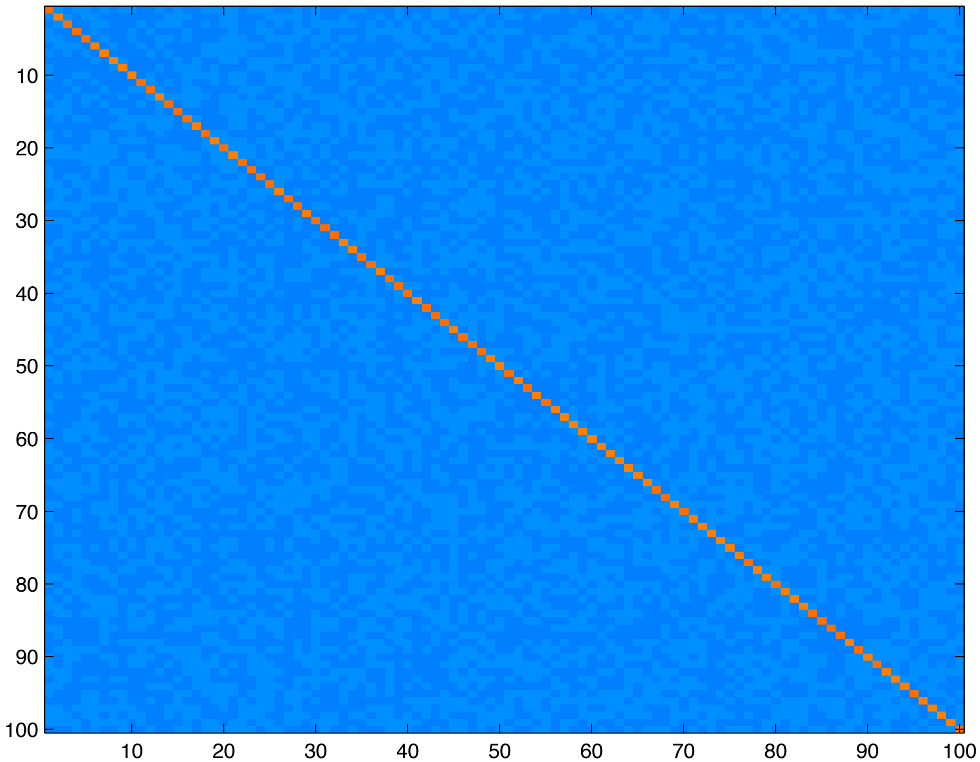
$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I \qquad K{=}10000$$

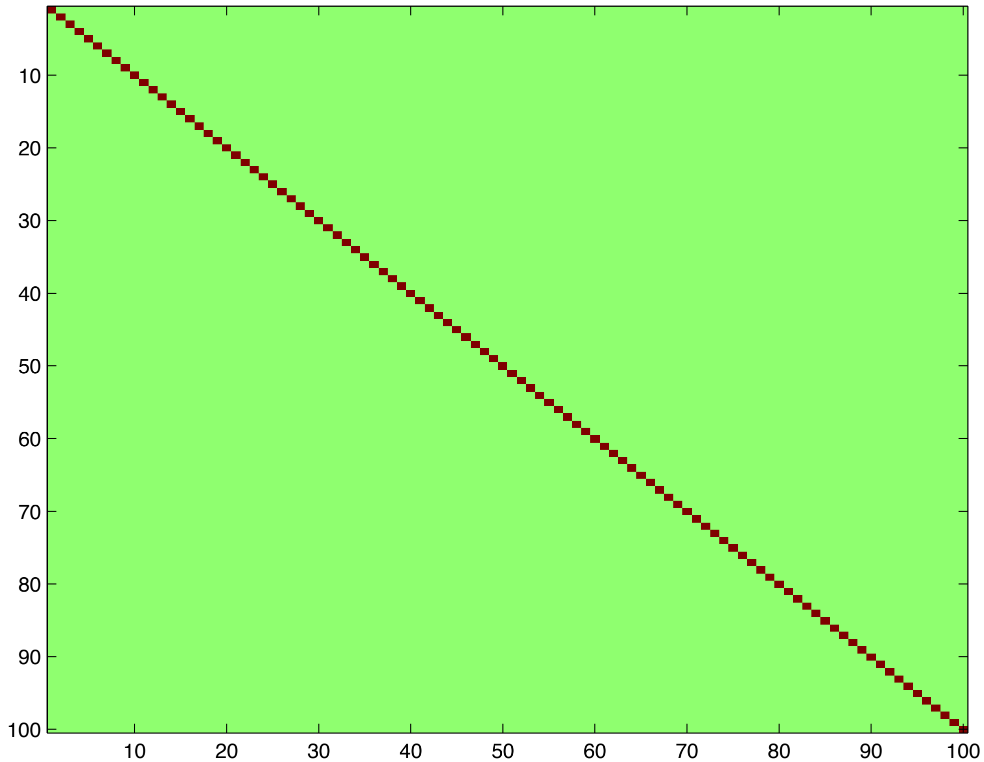




Source encoding

$$\frac{1}{K}\sum_{i=0}^{K-1}\mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \approx I \qquad K=100000$$





Deterministic optimization

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \gamma_k \mathbf{s}_k$$
$$\mathbf{s}_k = -H_k^{-1} \left(\frac{1}{K} \sum_{i=0}^{K-1} \nabla \phi_i[\mathbf{m}_k] \right)$$

- cost per iteration: $\mathcal{O}(K)$
- convergence rate:

$$|\Phi[\mathbf{m}_*] - \Phi[\mathbf{m}_k]| = \mathcal{O}(c^k), \quad 0 < c \leq 1$$

Stochastic optimization

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \gamma_k \mathbf{s}_k$$

$$\mathbf{s}_k = -\nabla \phi_i[\mathbf{m}_k], \quad i \sim U[0, K-1]$$

- assumption: $\mathsf{E}\left\{\mathbf{s}_k\right\} = -\nabla\Phi[\mathbf{m}_k]$
- cost per iteration: $\mathcal{O}(1)$
- convergence rate:

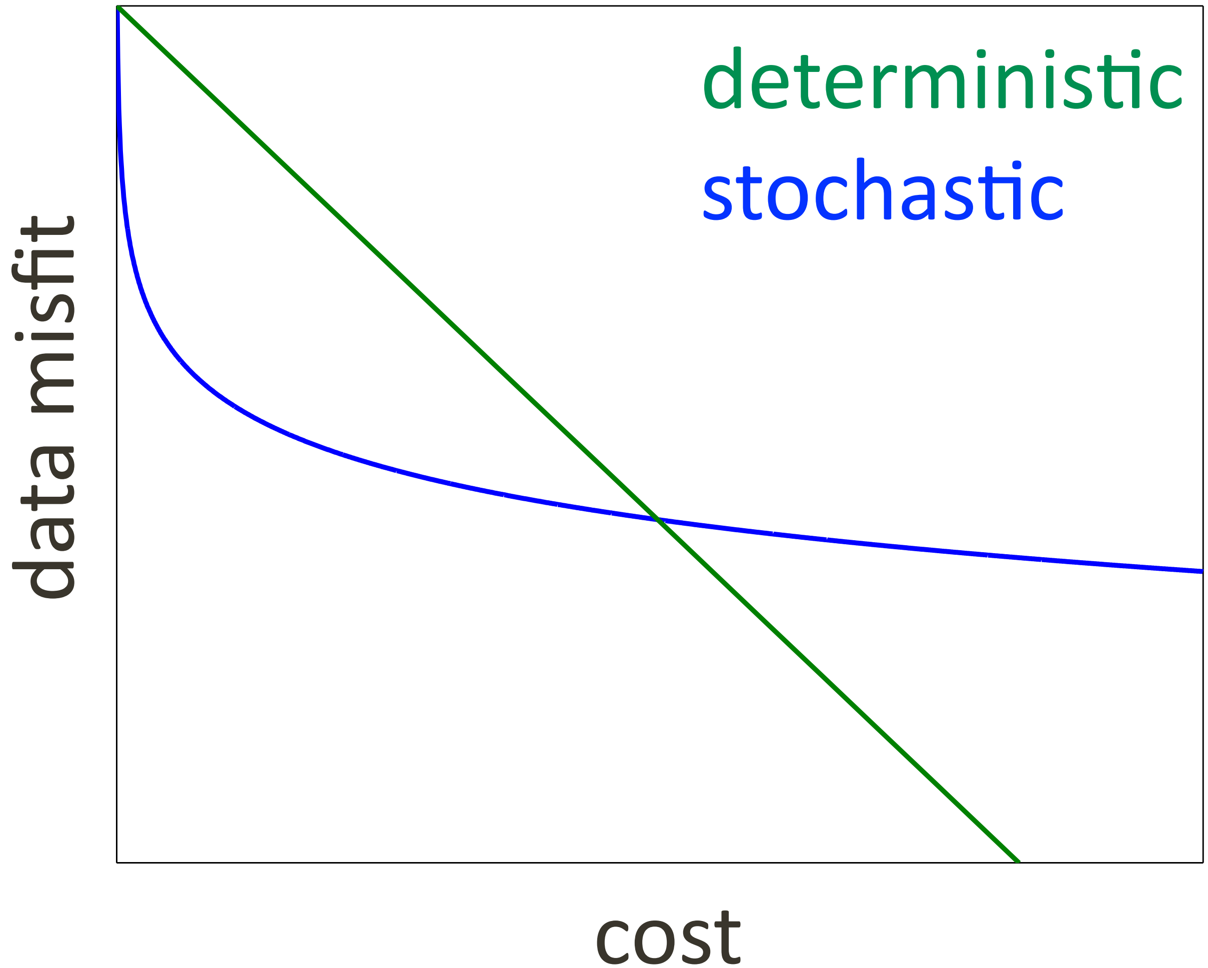
$$|\Phi[\mathbf{m}_*] - \Phi[\mathbf{m}_k]| = \mathcal{O}(1/k)$$

[Robbins et al. '50; Bertsekas et. al '96]

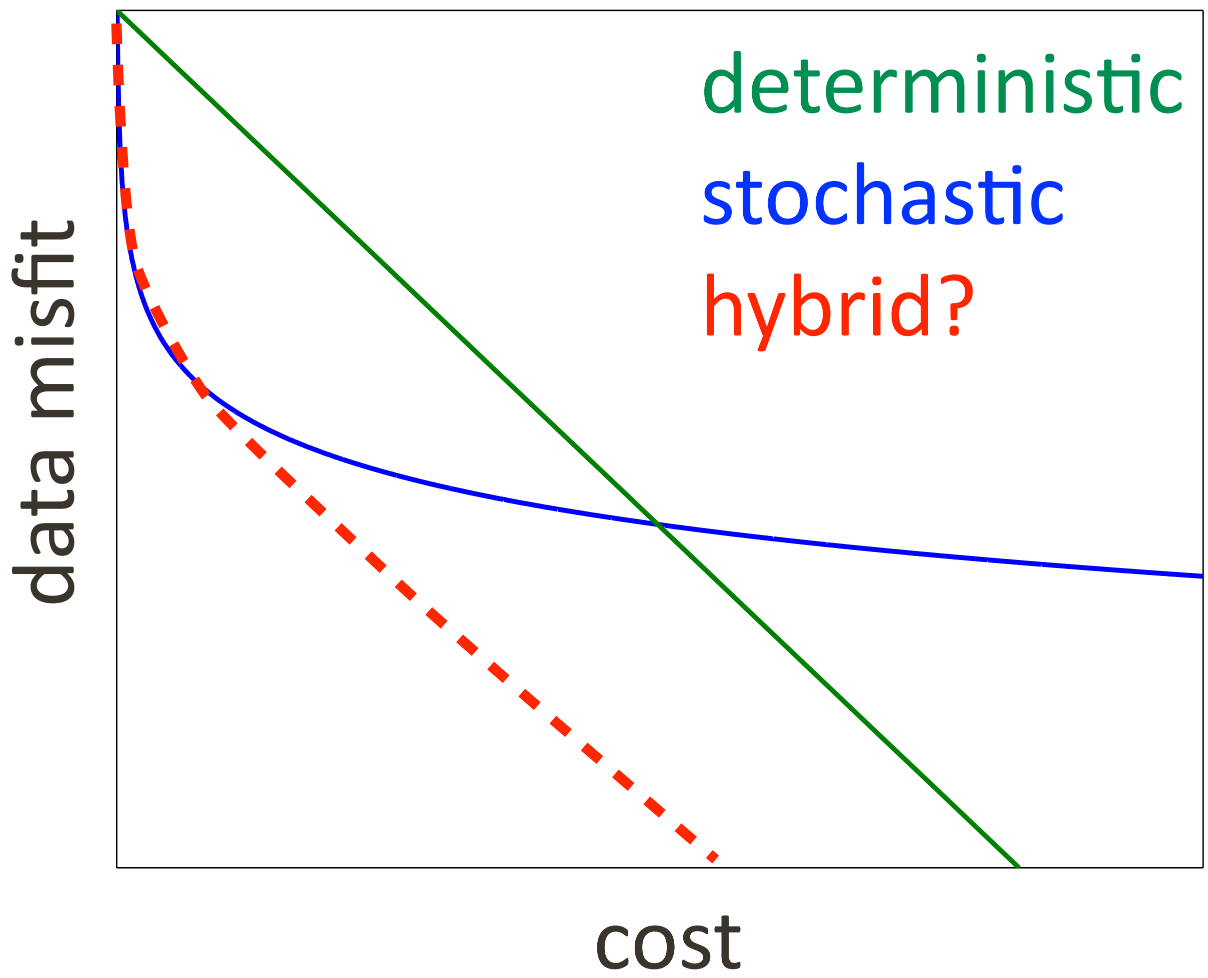
Stochastic optimization

- cheap iterations
- sample data *with* replacement
- slow convergence (relies on law of large numbers)

Stochastic vs. Deterministic



Stochastic vs. Deterministic



Hybrid method

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \gamma_k \mathbf{s}_k$$

$$\mathbf{s}_k = -H_k^{-1} \left(\frac{1}{|B_k|} \sum_{i \in B_k} \nabla \phi_i[\mathbf{m}_k] \right)$$

- assumption: $|| |B_k|^{-1} \sum_{i \in B_k} \nabla \phi_i[\mathbf{m}_k] - \nabla \Phi[\mathbf{m}_k] ||_2^2 < C_k$
- cost per iteration: $\mathcal{O}(|B_k|)$
- convergence rate

$$|\Phi[\mathbf{m}_*] - \Phi[\mathbf{m}_k]| = \mathcal{O}(\max\{c^k, C^k\}), \quad 0 < c \leq 1$$

[Friedlander et. al '11]

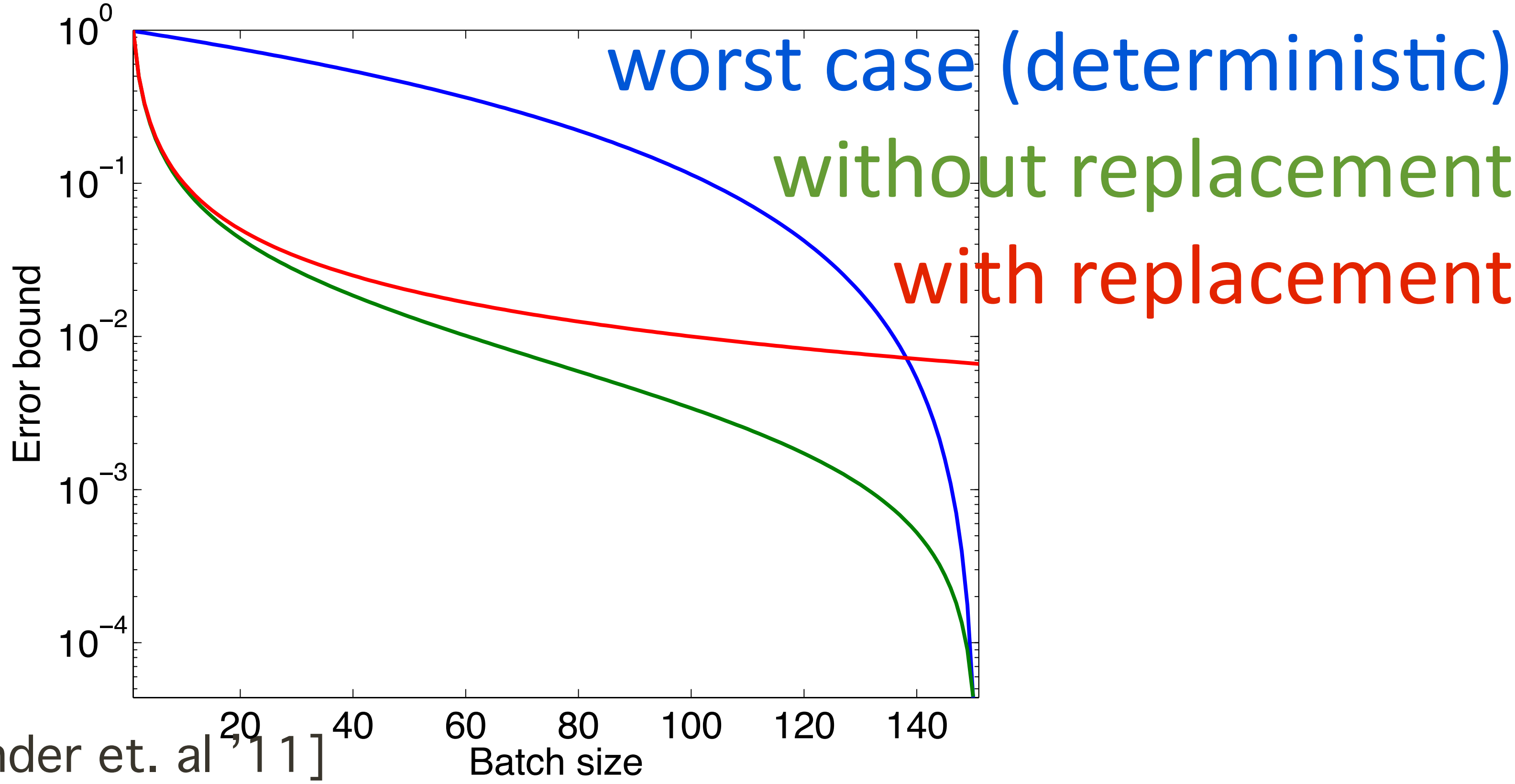
Batching strategy

We can grow the batch according to the natural order, or by random selection *without* replacement.

The SO strategy relies on drawing *with* replacement.

Batching strategy

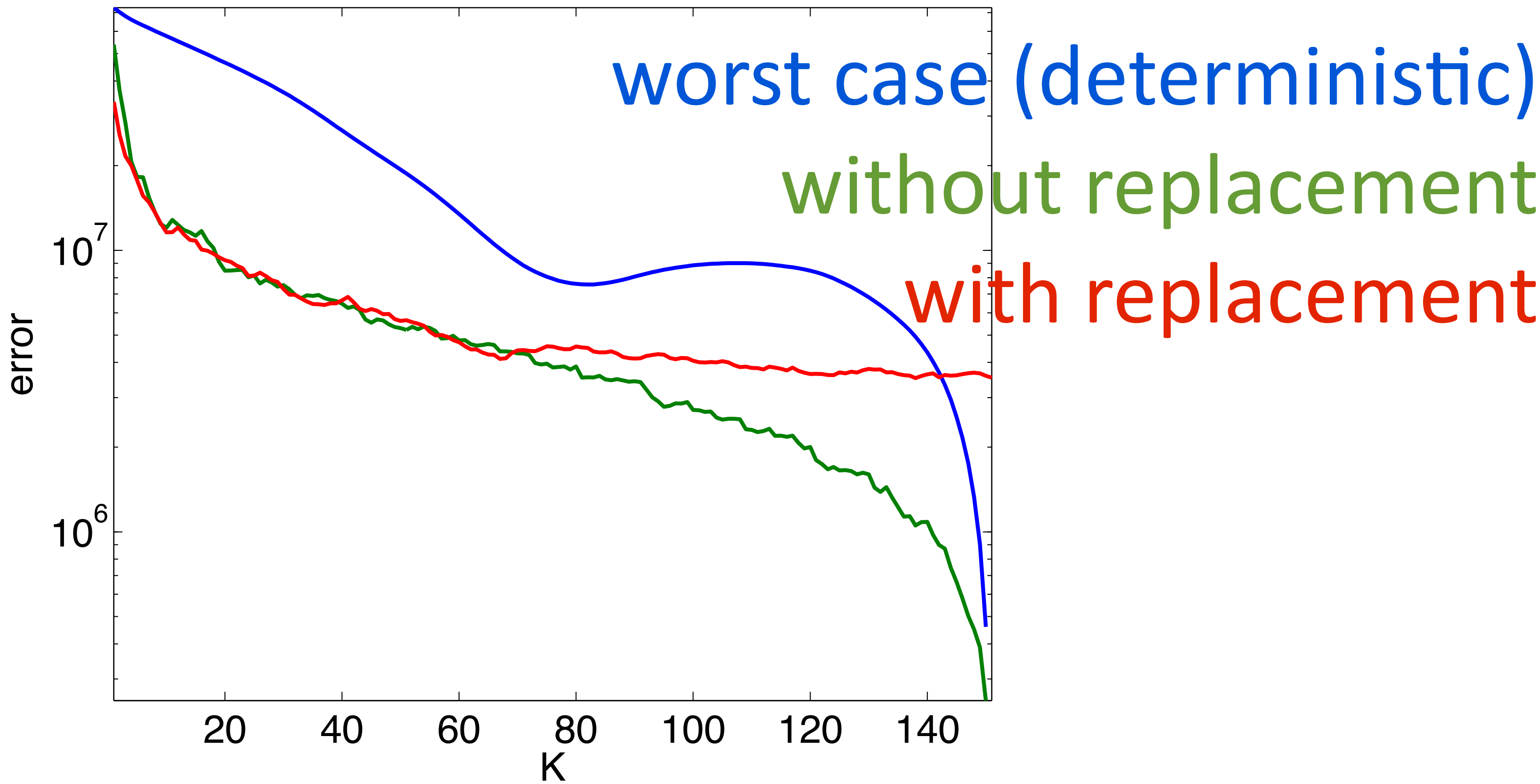
Batching strategy controls theoretical decay of the error.



[Friedlander et. al '11]

Batching strategy

Actual decay of the error in the gradient



Hybrid method

while not converged **do**

$\mathbf{g}_k = \sum_{i \in B_k} \nabla \phi_i[\mathbf{m}_k]$ gradient

$\mathbf{s}_k = -H_k^{-1} \mathbf{g}_k$ search dir.

find γ s.t. $\sum_{i \in B_k} \phi_i[\mathbf{m}_k + \gamma \mathbf{s}_k] < \sum_{i \in B_k} \phi_i[\mathbf{m}_k]$ linesearch

$\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{s}_k$ update model

$B_{k+1} = B_k \bigcup \{i\}$ increase batchsize

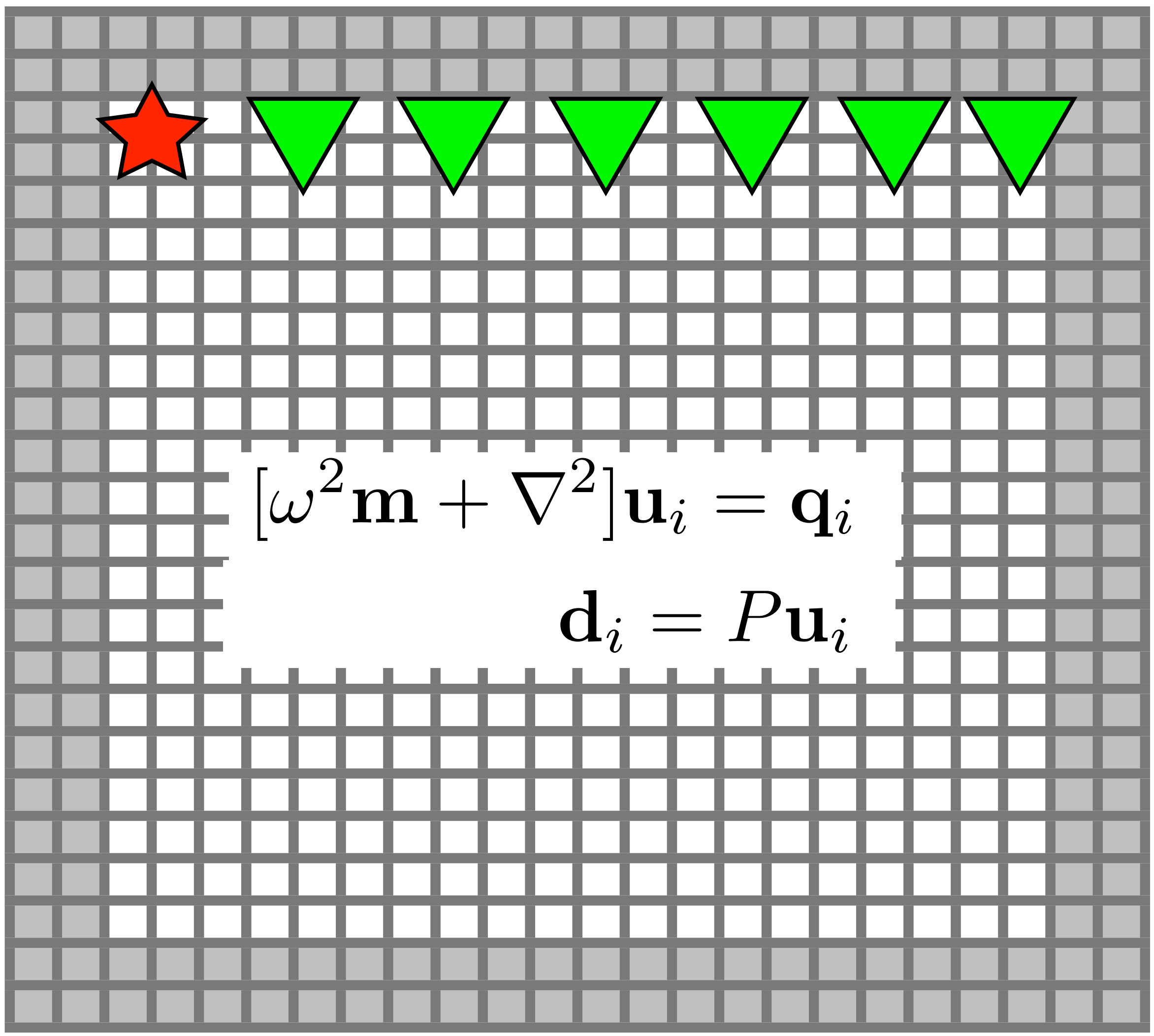
end while

[Friedlander et. al '11]

Hybrid method

- Will work with *any* source encoding strategy (none, randomized, plane wave, eigenvectors, …)
- Draw samples *without* replacement
- Random batching strategy is important to efficiently probe the data

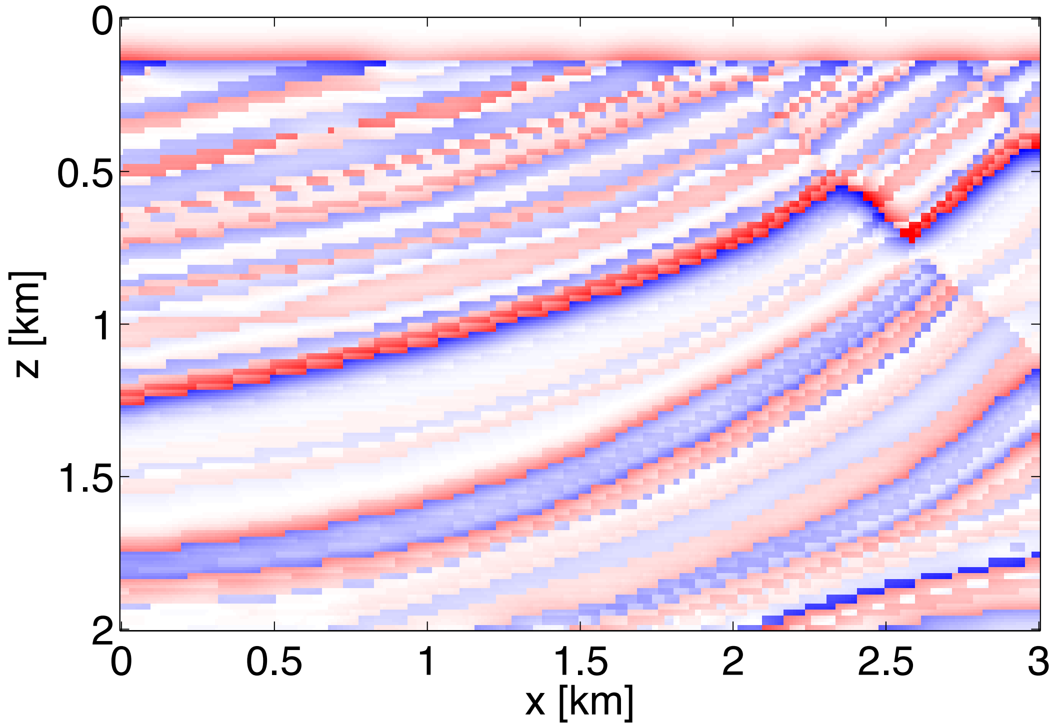
Results



$$[\omega^2\mathbf{m} + \nabla^2]\mathbf{u}_i = \mathbf{q}_i$$

$$\mathbf{d}_i = P\mathbf{u}_i$$

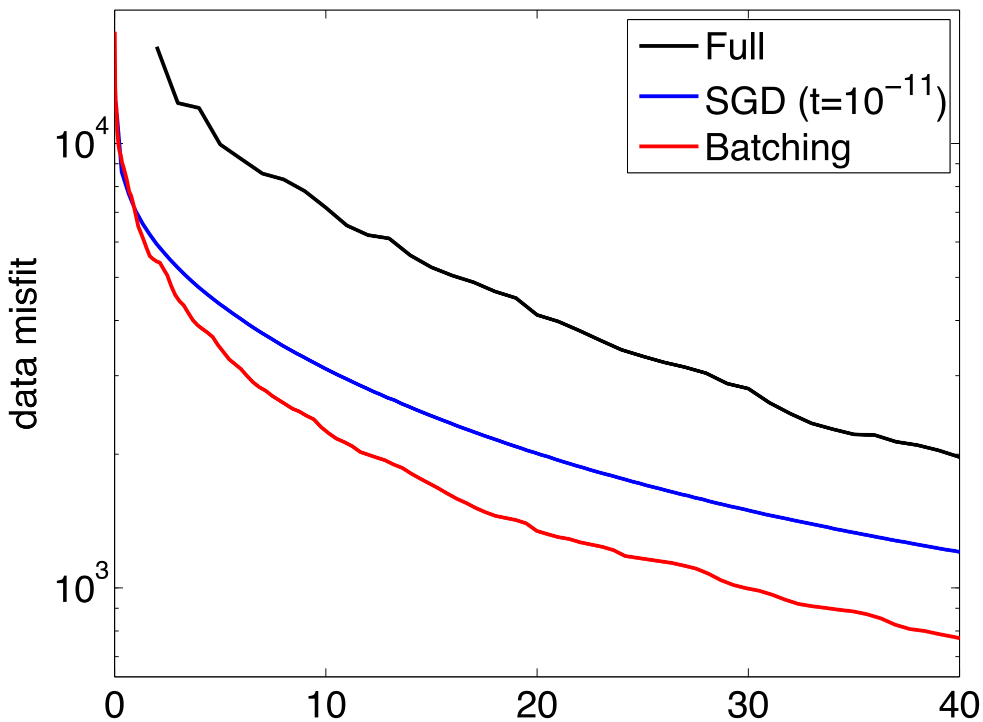
`Nonlinear migration'



151 sources/receivers, frequencies [5-25] Hz
15 Hz Ricker

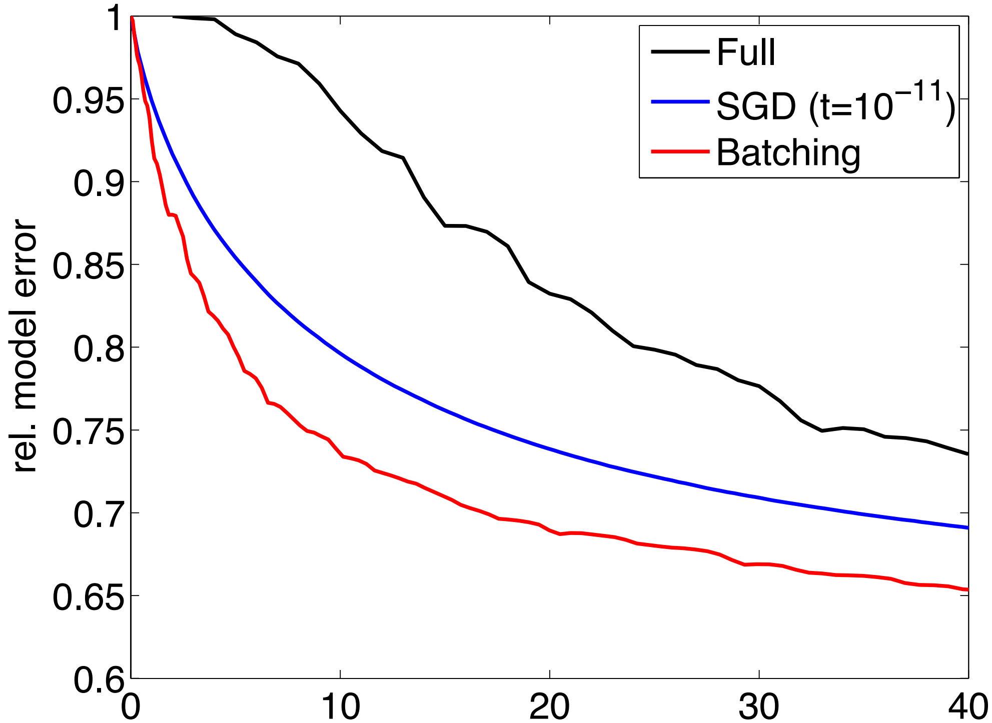
Non-linear migration

data misfit



# of full evaluations

model error

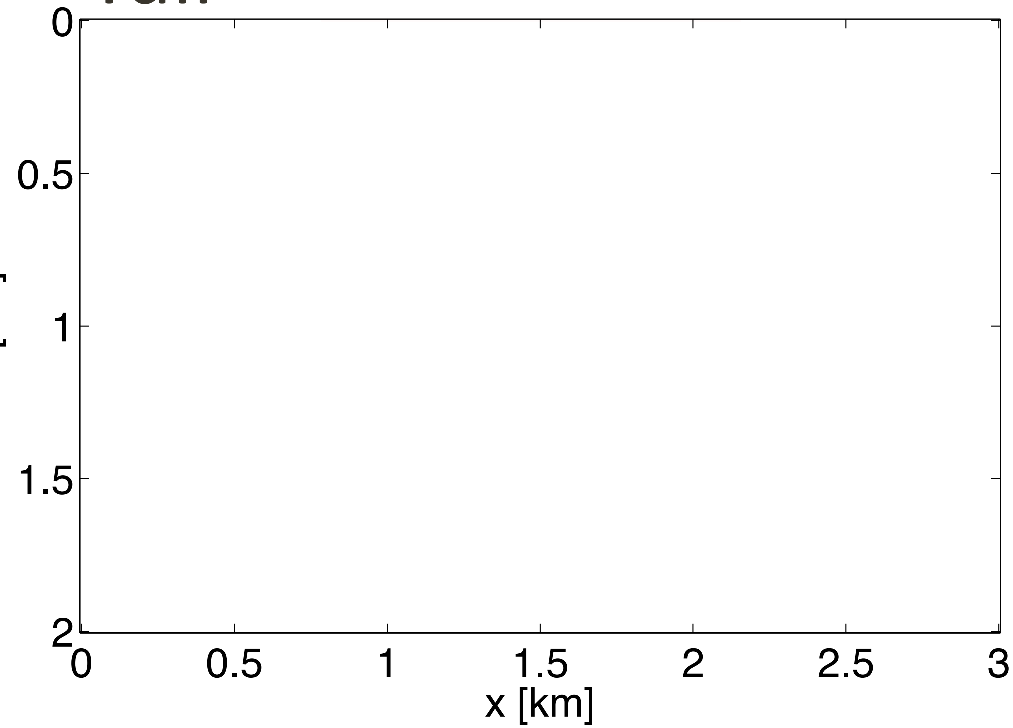


# of full evaluations

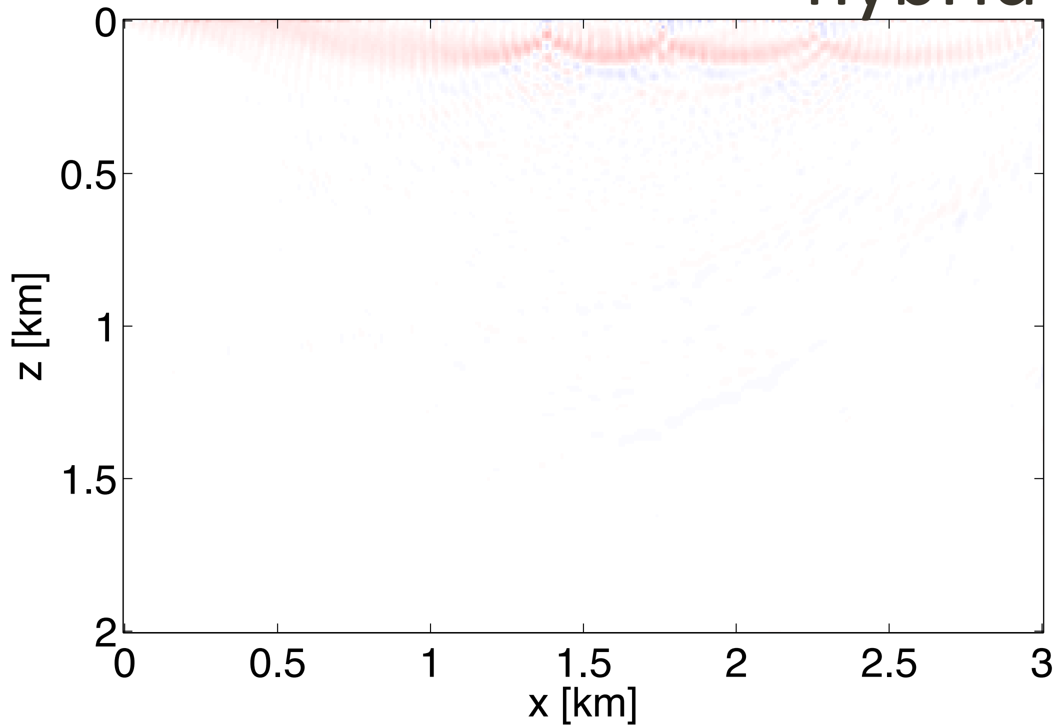
Non-linear migration

0.18 full evaluations

full



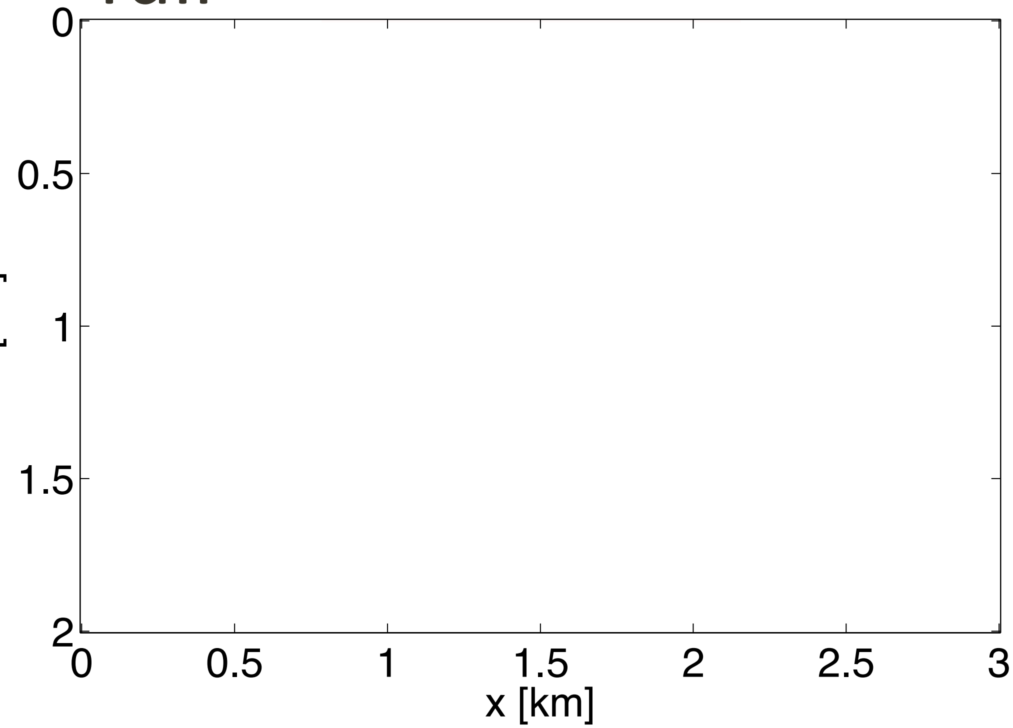
hybrid



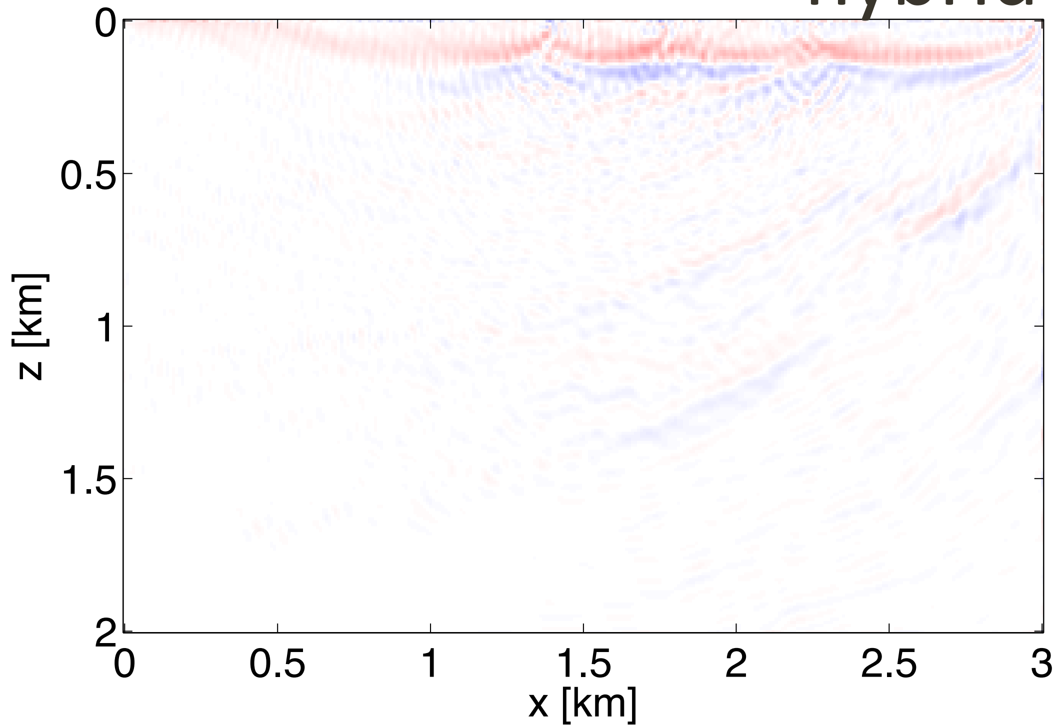
Non-linear migration

0.43 full evaluations

full



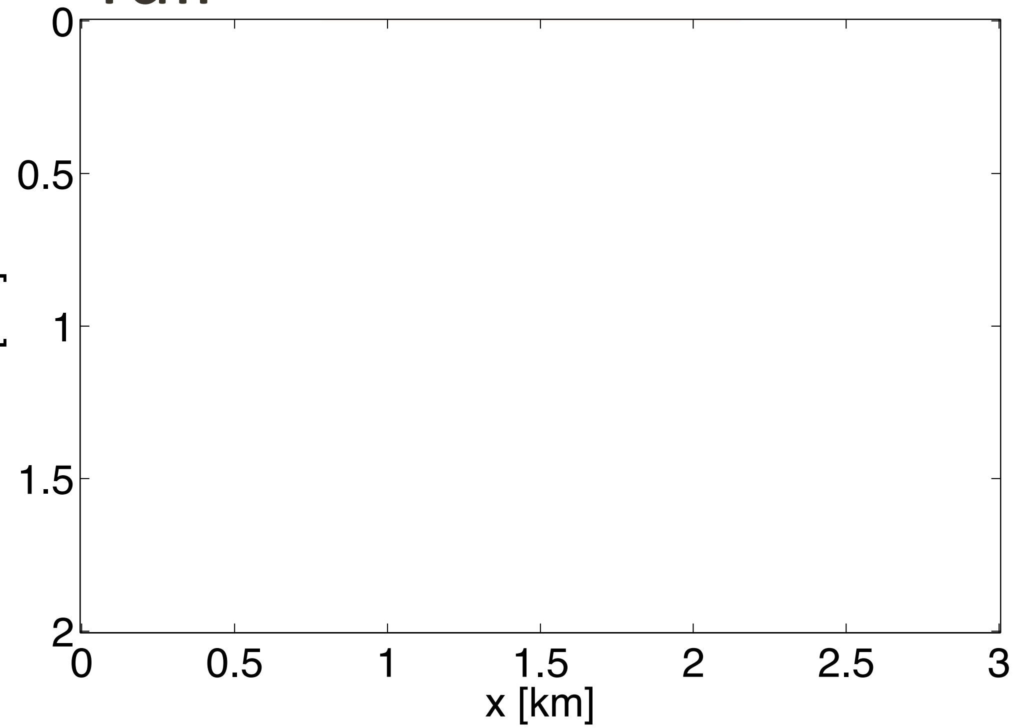
hybrid



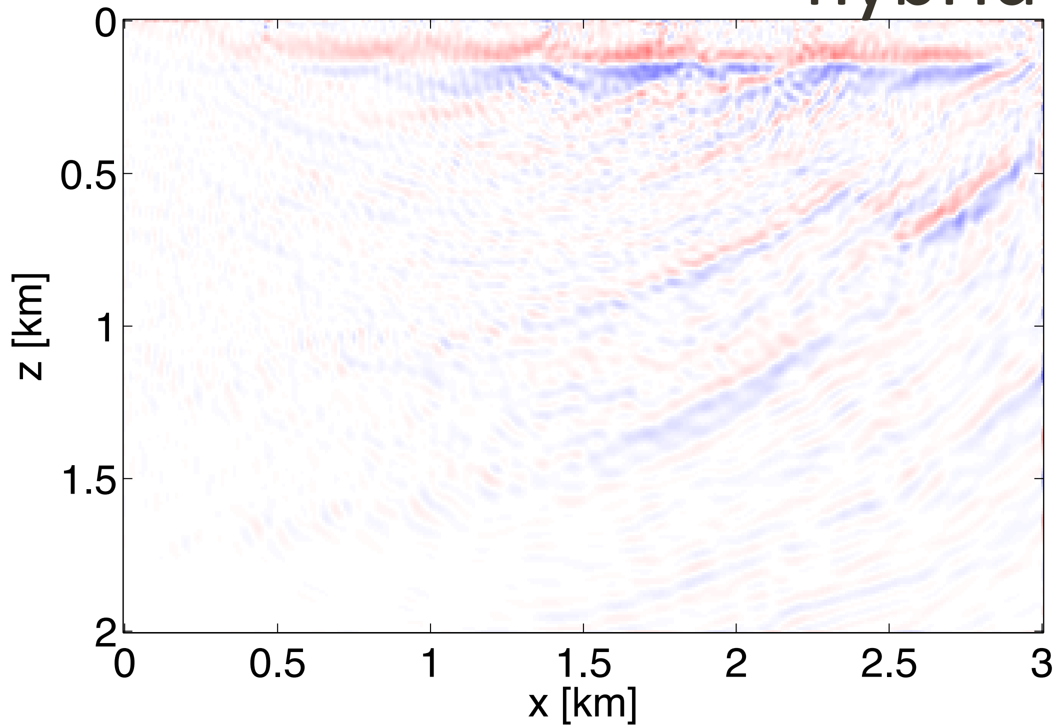
Non-linear migration

0.89 full evaluations

full

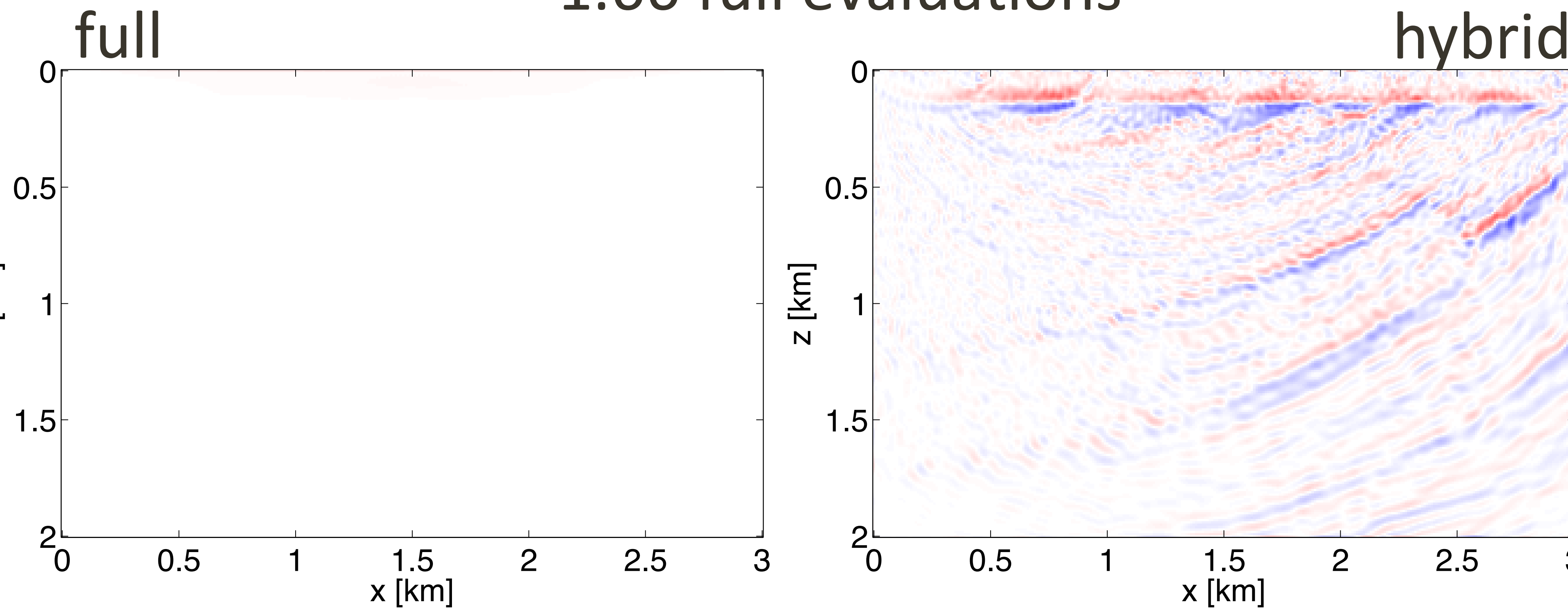


hybrid



Non-linear migration

1.66 full evaluations

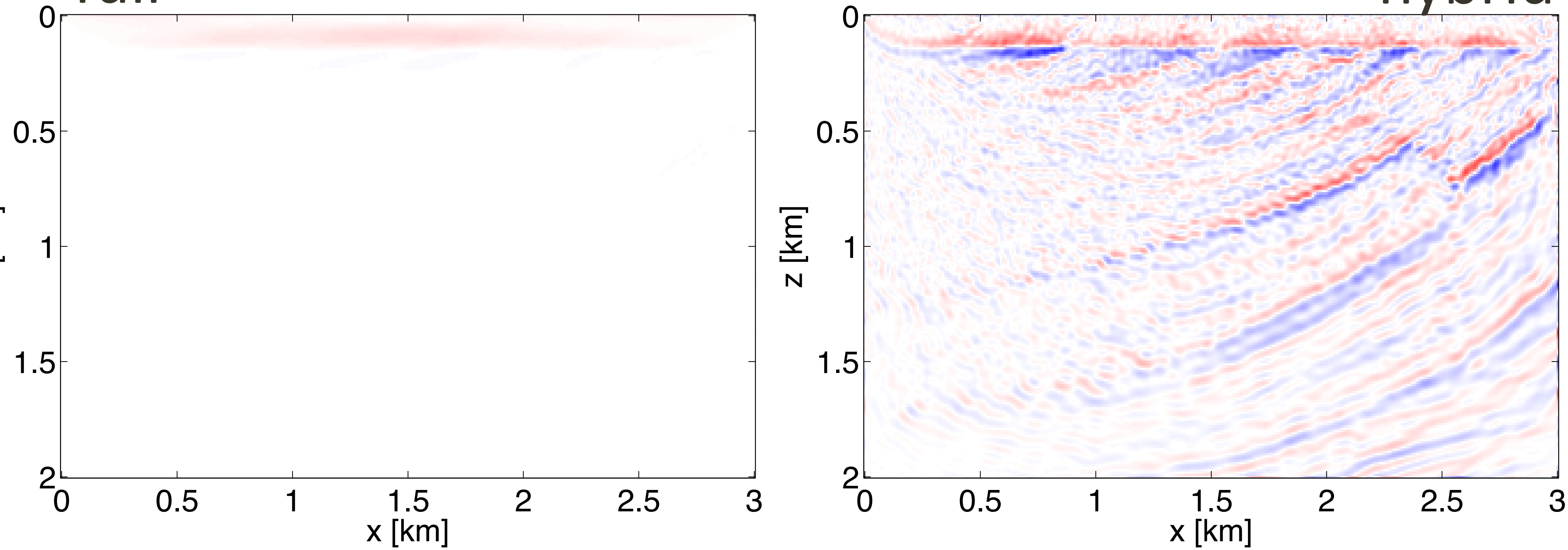


Non-linear migration

2.87 full evaluations

full

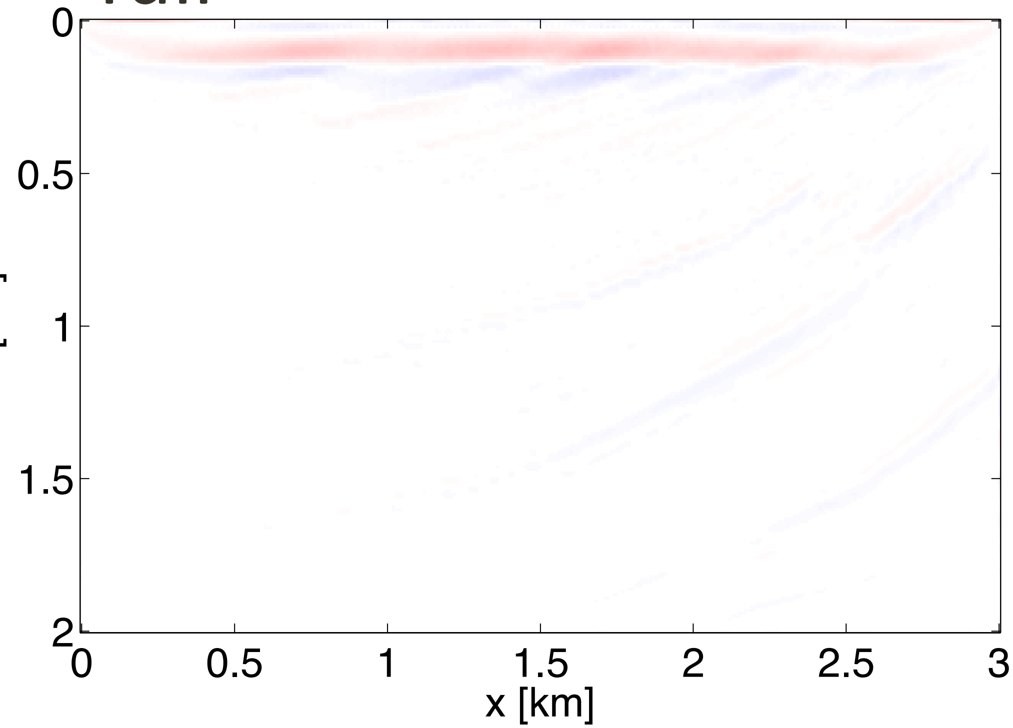
hybrid



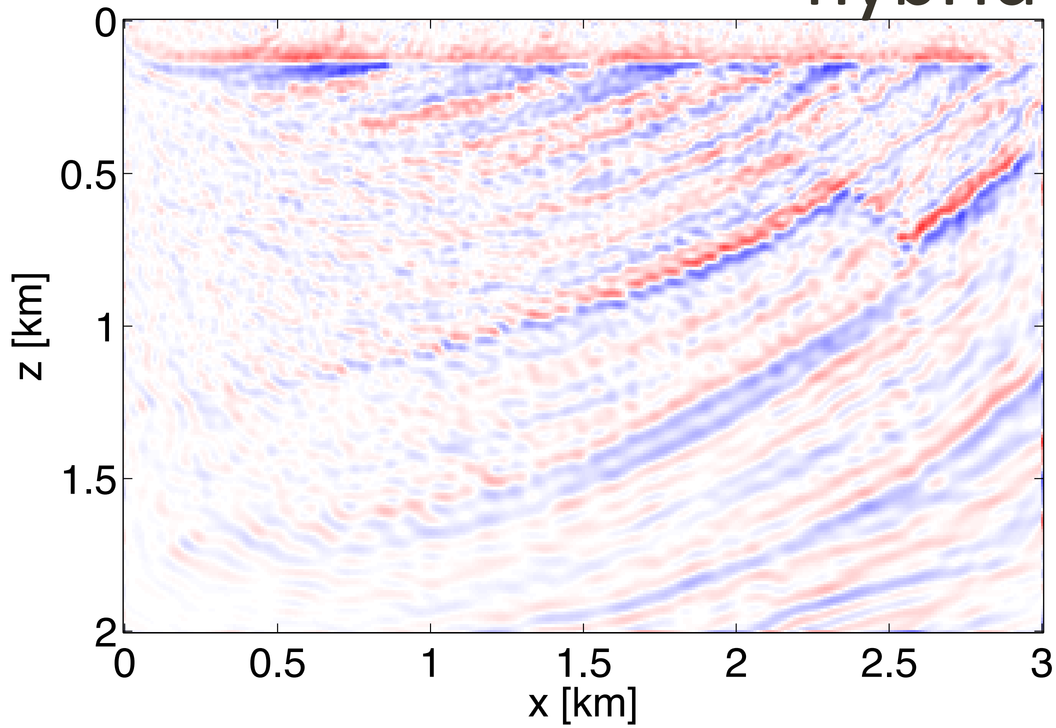
Non-linear migration

4.65 full evaluations

full

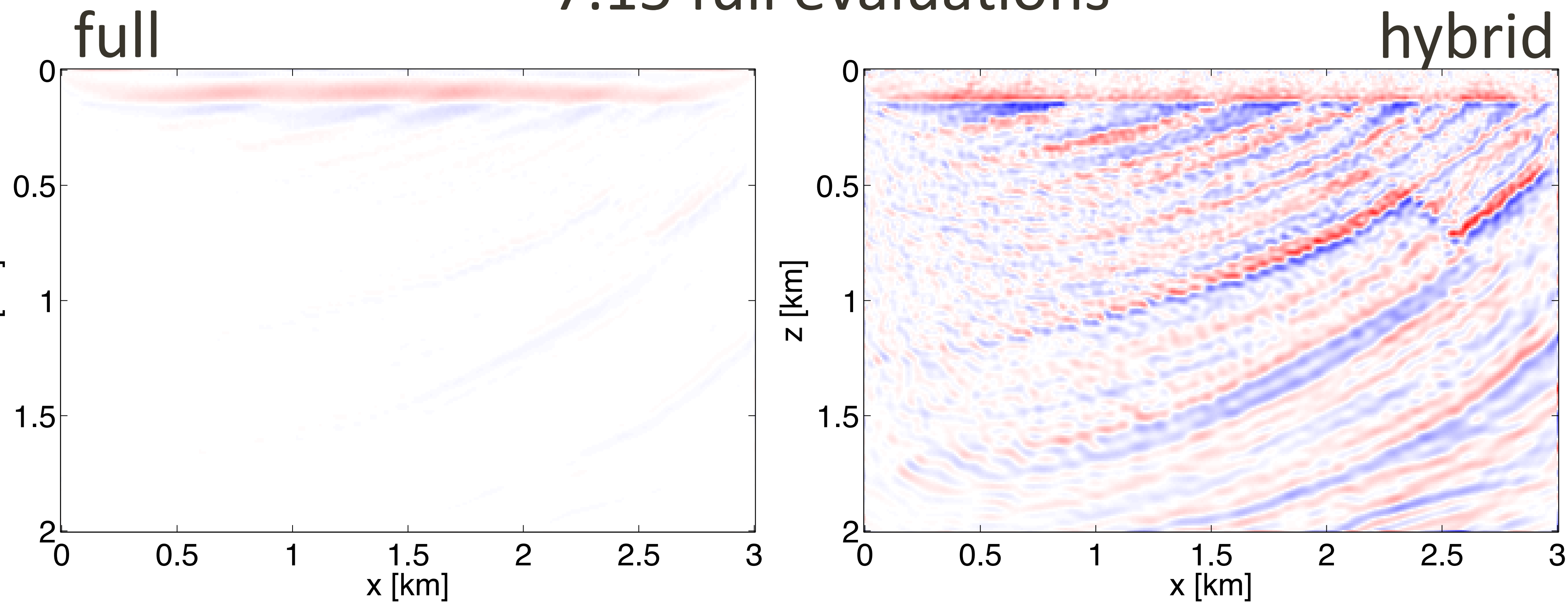


hybrid



Non-linear migration

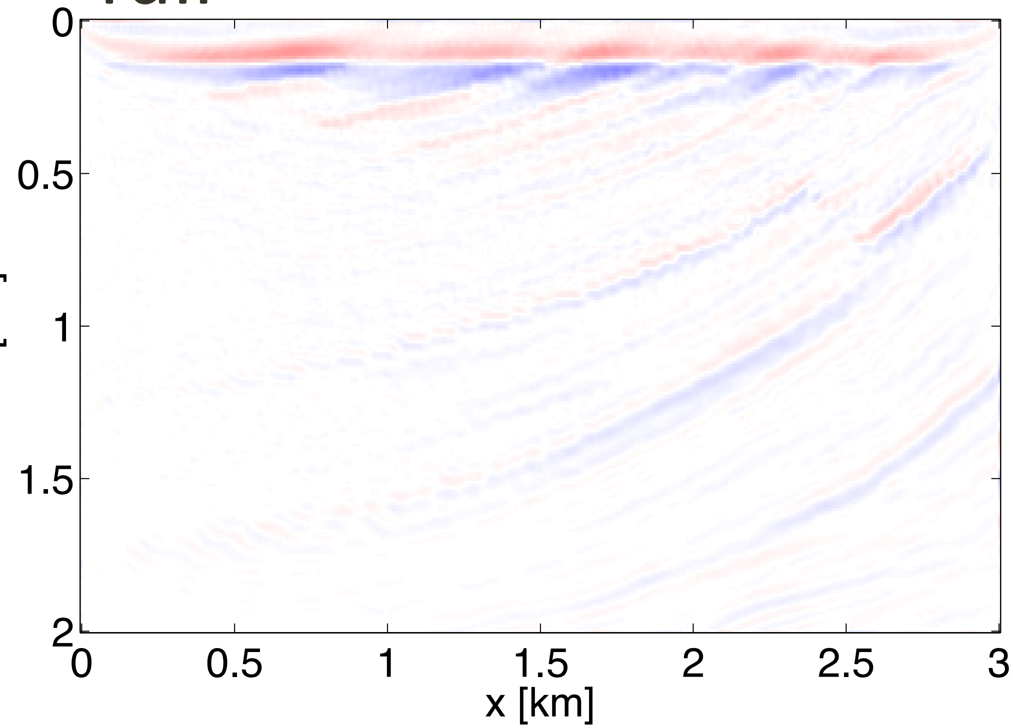
7.15 full evaluations



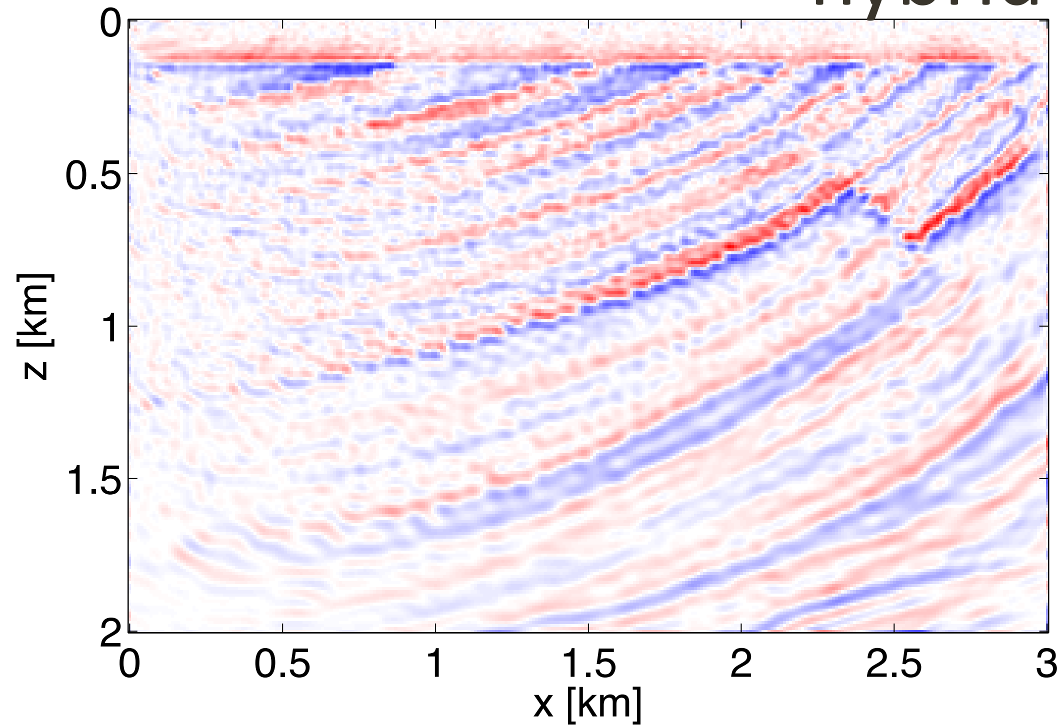
Non-linear migration

10.87 full evaluations

full

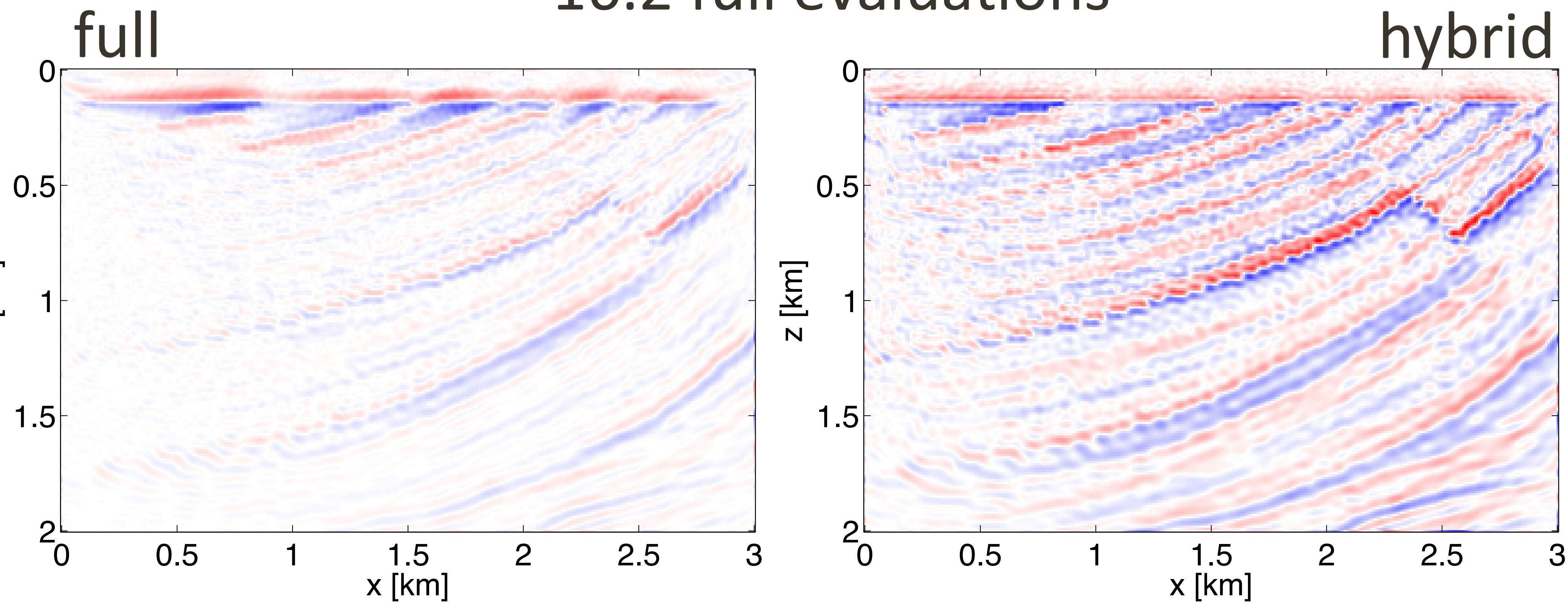


hybrid



Non-linear migration

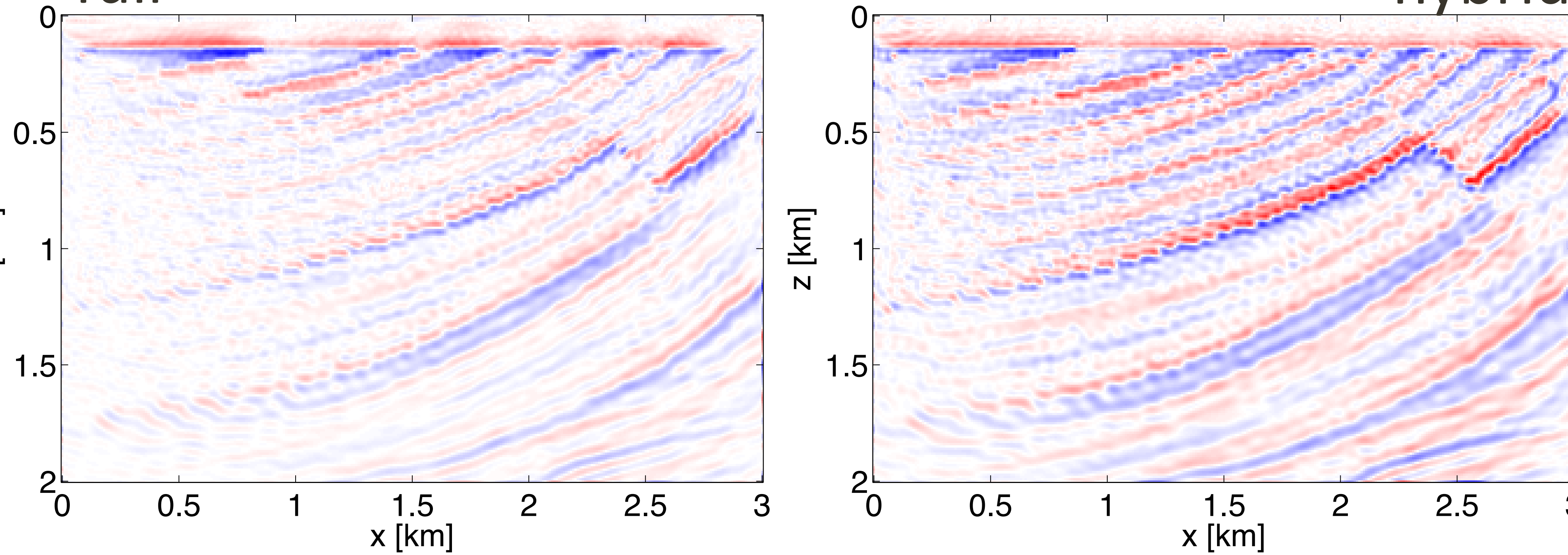
16.2 full evaluations



Non-linear migration

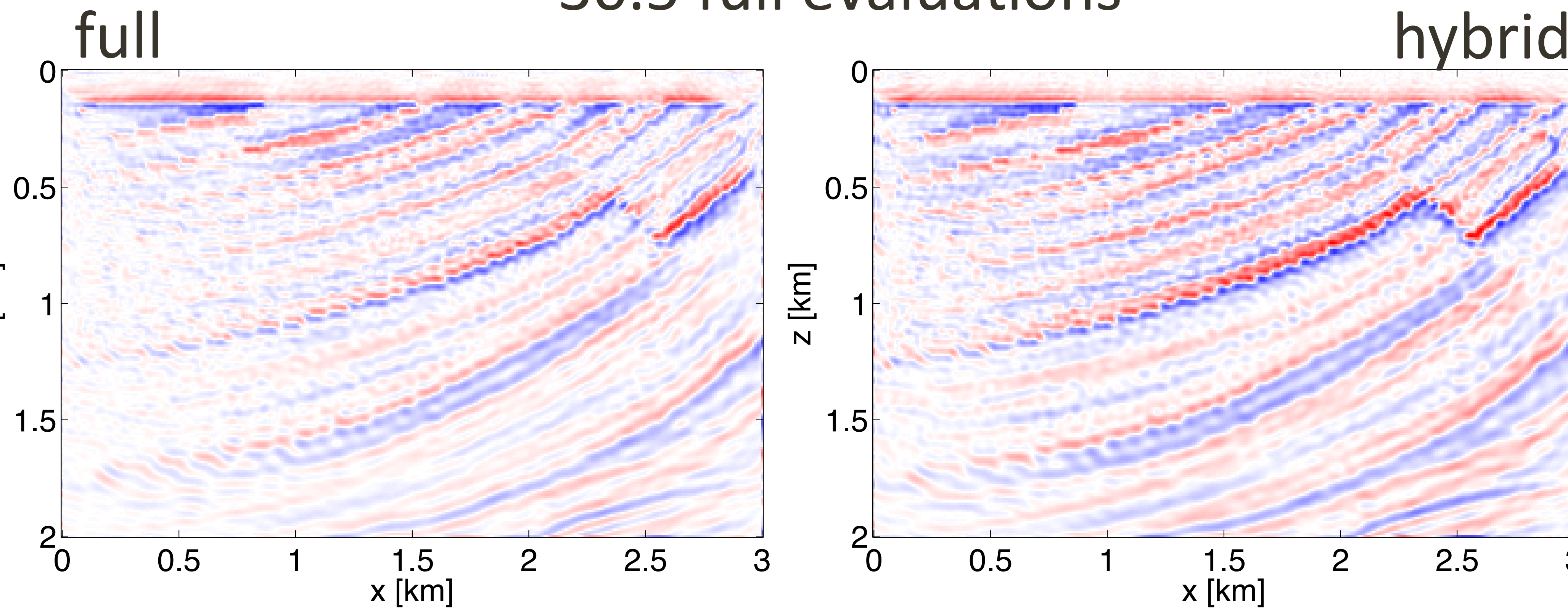
22 full evaluations

full hybrid



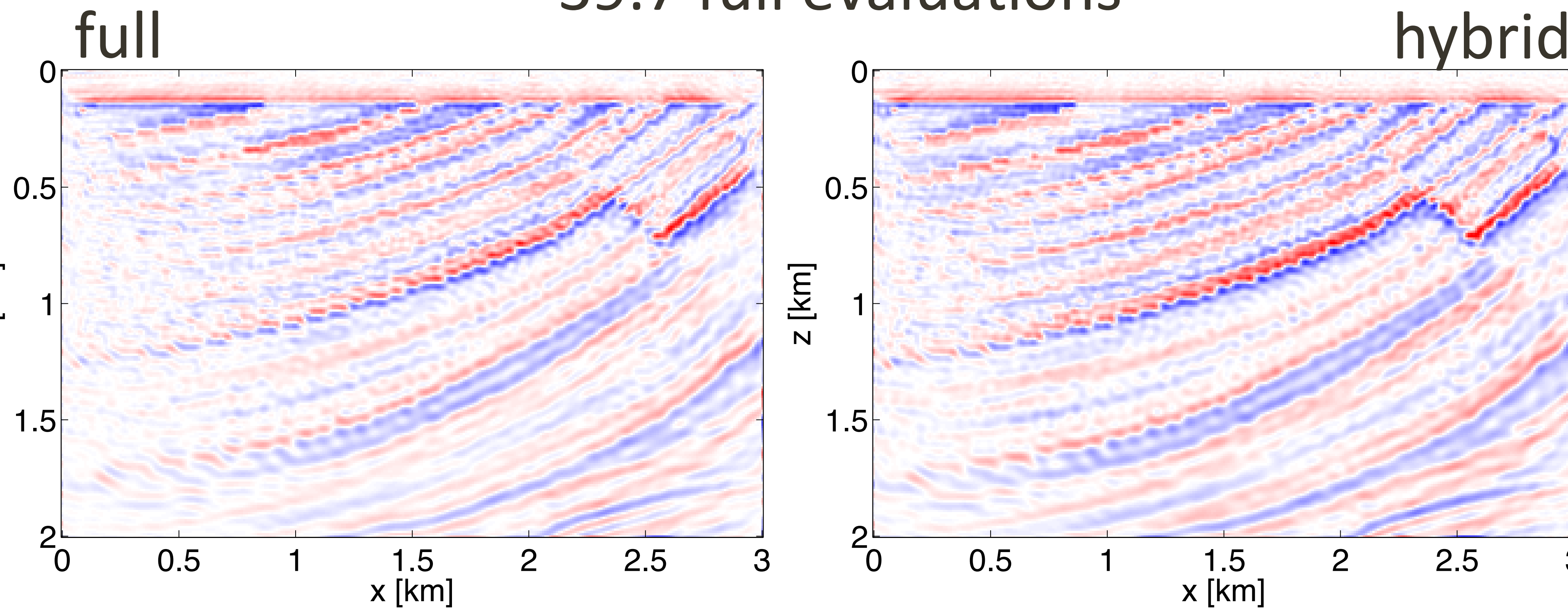
Non-linear migration

30.5 full evaluations



Non-linear migration

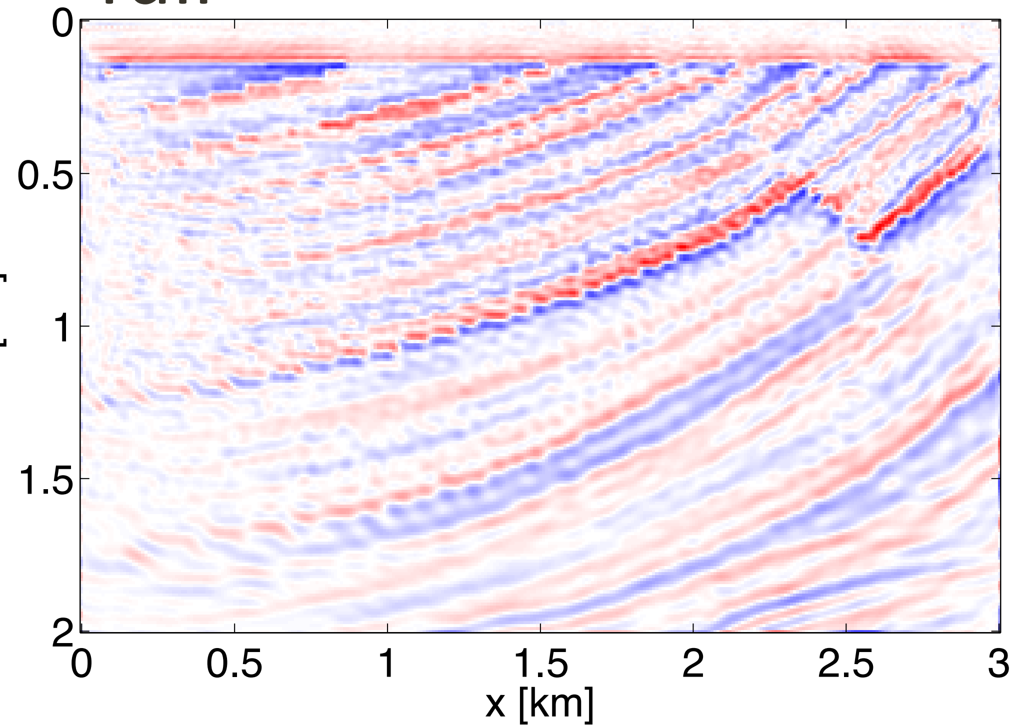
39.7 full evaluations



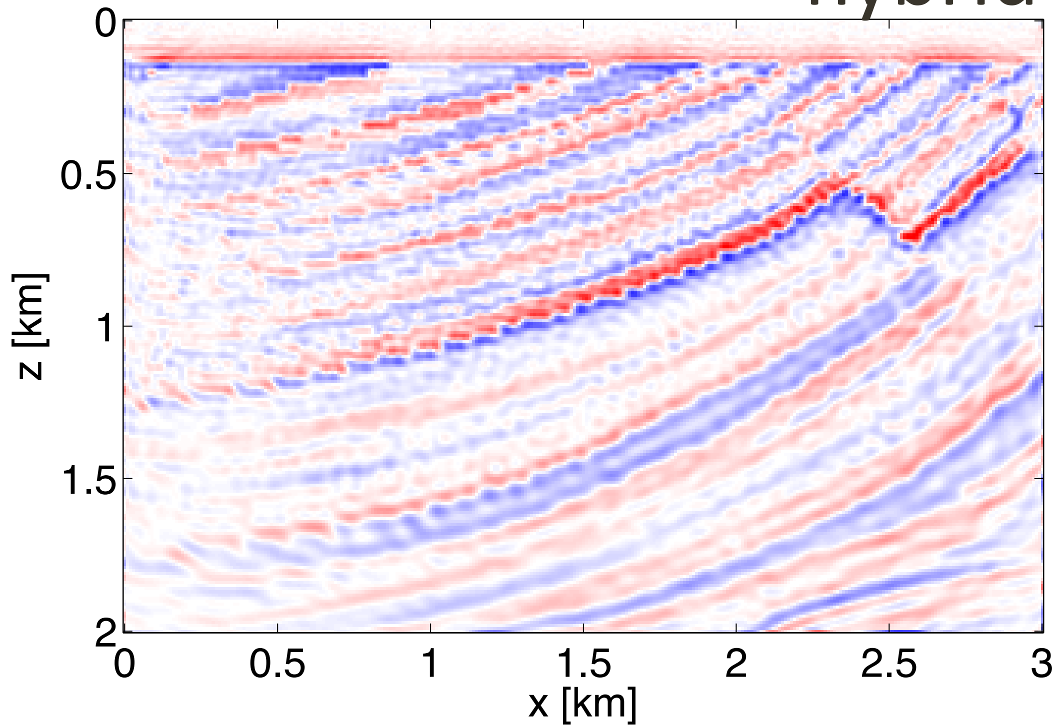
Non-linear migration

50.24 full evaluations

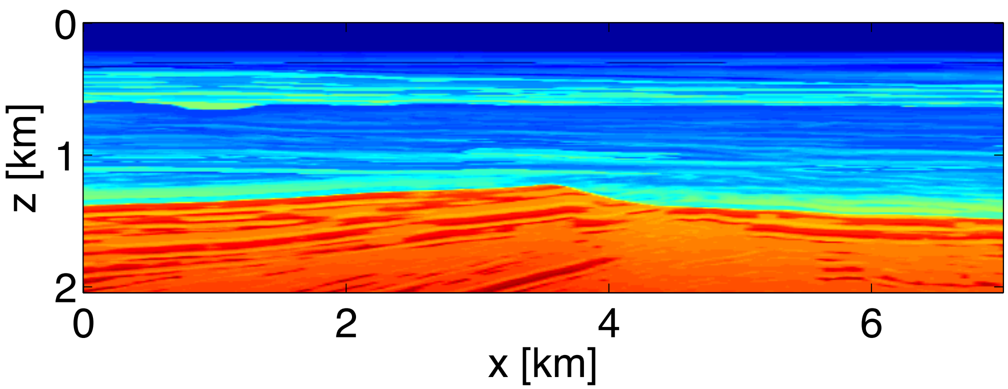
full

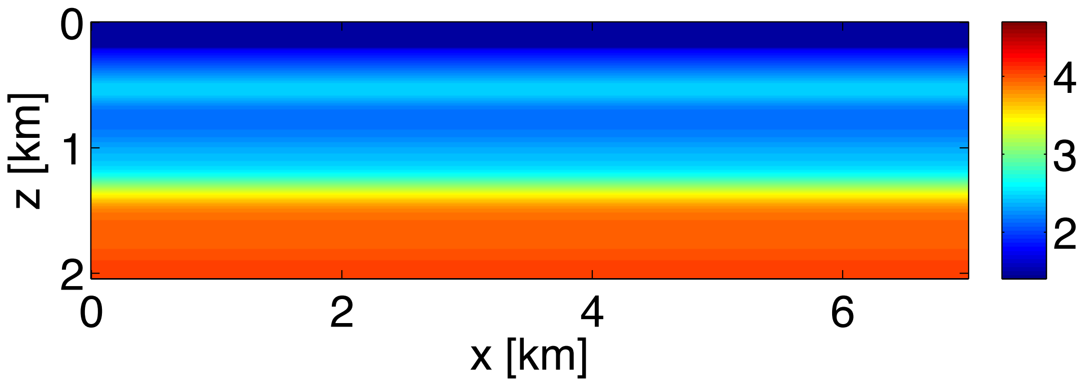


hybrid



Full waveform inversion



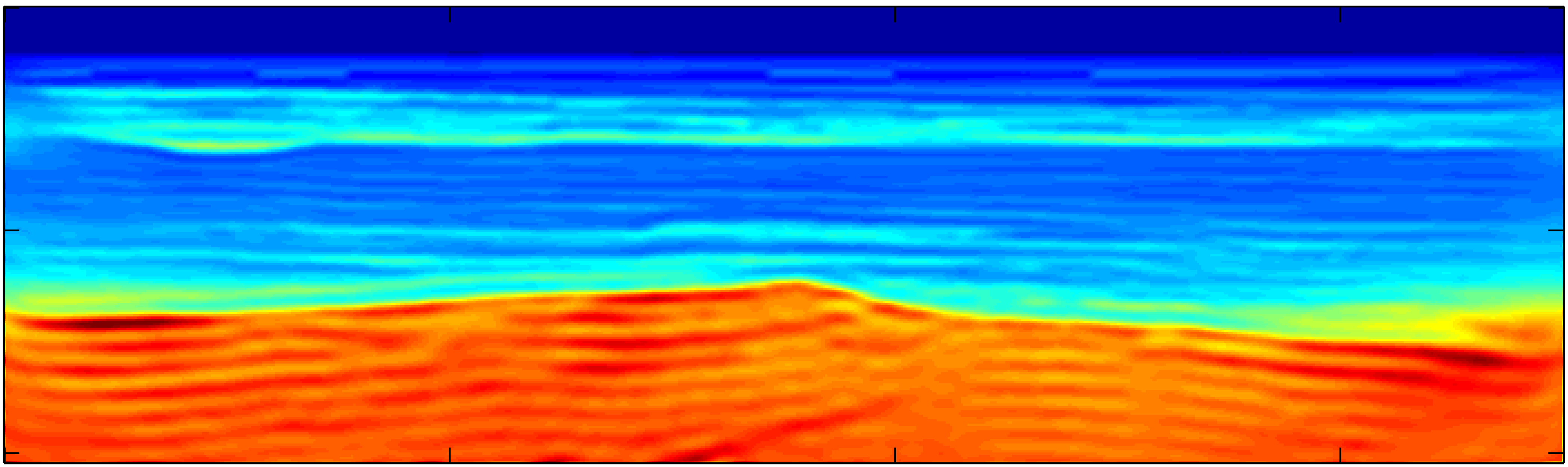


data for
141 sources, 281
receivers, 15 Hz Ricker

multiscale frequency
domain inversion:
[2.5-20] Hz in 16 bands

[Bunks `95; Pratt `98]

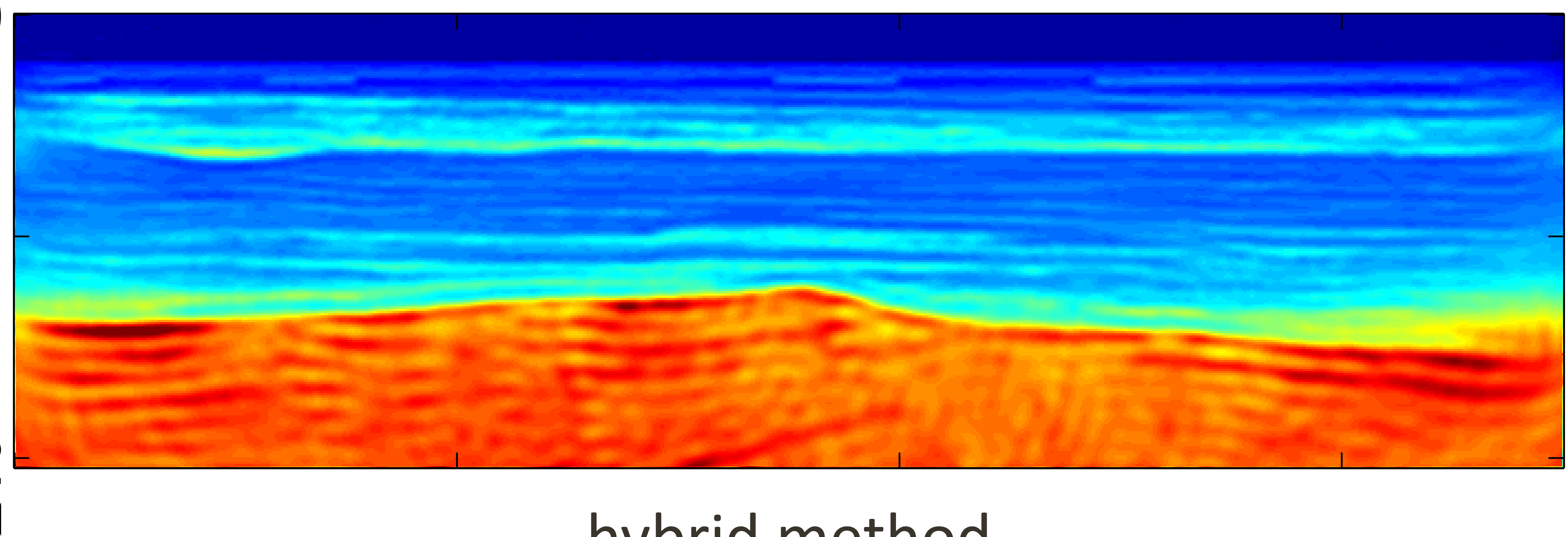
FWI



traditional L-BFGS

~10 full evaluations per frequency band

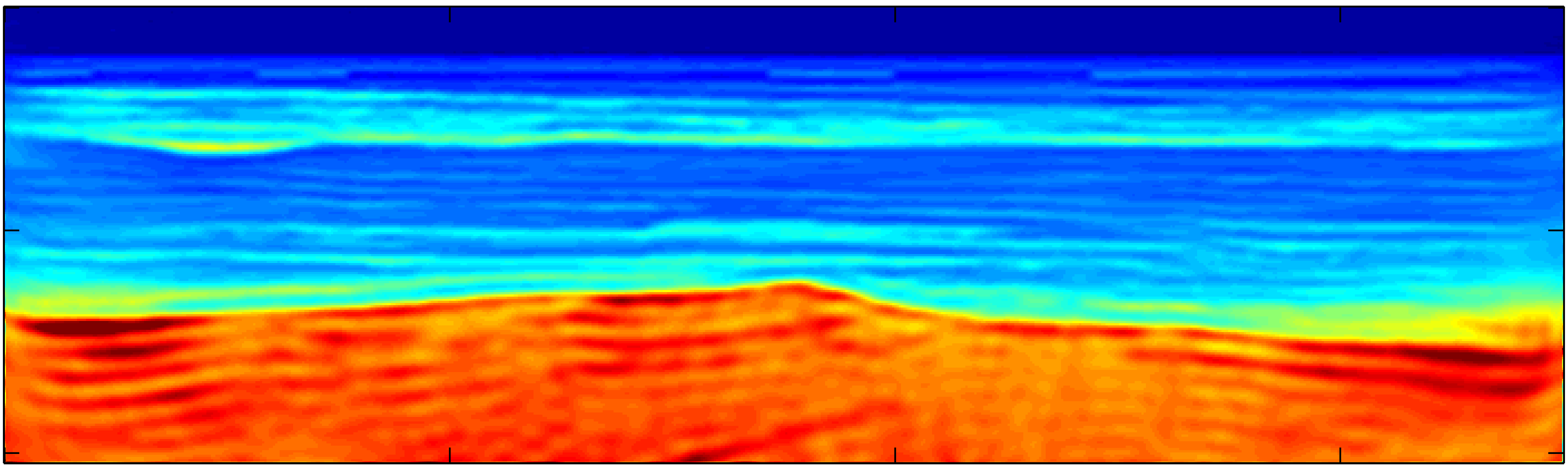
FWI



hybrid method

~2 full evaluations per frequency band

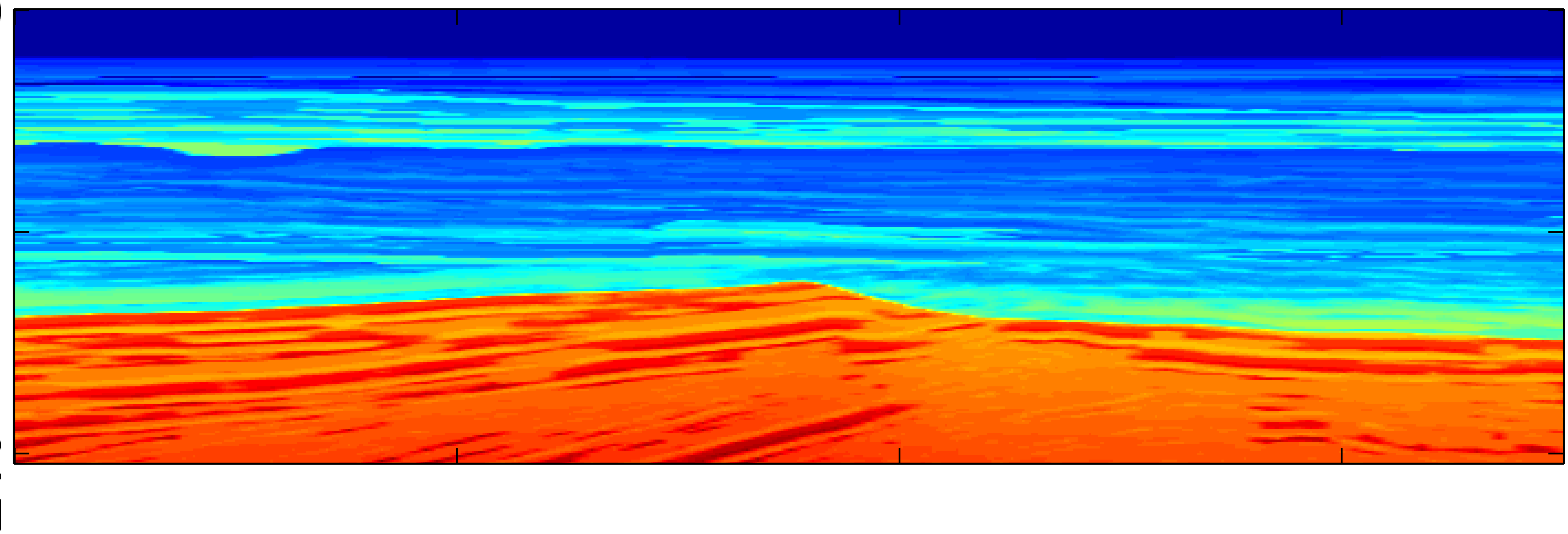
FWI



hybrid method

~5 full evaluations per frequency band

FWI



Conclusions

- Hybrid method gives both speed-up of stochastic method and convergence rate of deterministic method
- Not restricted to randomized source encoding: *can be applied to marine data*!
- Applicable to *any* optimization problem of the form

$$\min_{\mathbf{m}} \Phi[\mathbf{m}] = \frac{1}{K}\sum_{i=0}^{K-1} \phi_i[\mathbf{m}]$$

Acknowledgements

Eldad Haber, Aleksandr Aravkin

SINBAD

NSERC CRSNG

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.