Enabling wave-based inversion on GPUs with randomized trace estimation

Mathias Louboutin and Felix J. Herrmann*

June 9 2022

SLIM
Georgia Institute of Technology

ML4Seismic

Georgia Tech College of Computing
School of Computational Science and Engineering

Georgia Tech College of Sciences
School of Earth and Atmospheric Sciences

Georgia Tech College of Engineering
School of Electrical and Computer Engineering

SLIM
ML4Seismic

Motivation

High-memory footprint adjoint-state methods

Computationally expensive checkpointing

Case specific/internal solutions to manage memory

- Fourier (BP patent)
- Compression (No existing GPU porting)
- Serialization/Disk (High IO)
- Boundary methods (reversible only)
- ...

Rajiv Kumar, Marie Graff-Kray, Ivan Vasconcelos, and Felix J. Herrmann, “Target-oriented imaging using extended image volumes—a low-rank factorization approach”, Geophysical Prospecting, vol. 67, pp. 1312-1328, 2019
Mengmeng Yang, Marie Graff, Rajiv Kumar, and Felix J. Herrmann, “Low-rank representation of omnidirectional subsurface extended image volumes”, Geophysics, vol. 86, pp. 1-41, 2021.
Mathias Louboutin, Ali Siahkoohi, Rongrong Wang, and Felix J. Herrmann, “Low-memory stochastic backpropagation with multi-channel randomized trace estimation”. 2021.
Philipp A. Witte, Mathias Louboutin, Fabio Luporini, Gerard J. Gorman, and Felix J. Herrmann, “Compressive least-squares migration with on-the-fly Fourier transforms”, Geophysics, vol. 84, pp. R655-R672, 2019.

Solutions

Take advantage of large-scale randomized linear algebra

Leverage our work on full-subsurface offset Image Volumes

Build on lessons learned form machine learning (convolutional layers)

According to stochastic optimization

- inaccurate gradients can still lead to accurate inversion
- undergirds our compressive imaging & randomized FWI & WRI

Halko, Nathan, Per-Gunnar Martinsson, and Joel A. Tropp. "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions." *SIAM review* 53.2 (2011): 217-288.

SLIM ML4Seismic

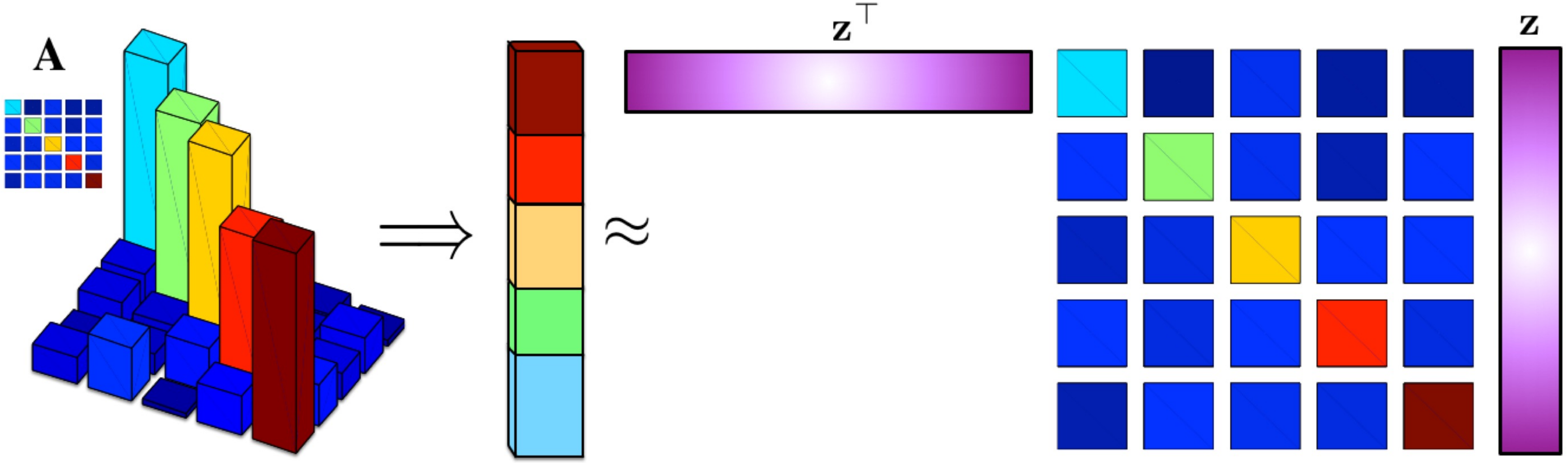
Randomized linear algebra

Randomized SVD:

$$\mathbf{A} \approx \mathbf{U}\mathbf{S}\mathbf{V}^\top \quad \text{with} \quad \begin{cases} [\mathbf{Q}, \sim] & = \mathrm{qr}(\mathbf{A}\mathbf{Z}) \\ [\widetilde{\mathbf{U}}, \mathbf{S}, \mathbf{V}] & = \mathrm{svd}(\mathbf{Q}^\top \mathbf{A}) \\ \mathbf{U} = \mathbf{Q}\widetilde{\mathbf{U}} & \end{cases}$$

- information is reaped during random probing $\mathbf{AZ}$ w/ $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_r]$
- only need access to action of $\mathbf{A}$ (in parallel)
- memory friendly
- unbiased estimator when $\mathbb{E}(\mathbf{z}\mathbf{z}^\top) = \mathbf{I}$ w/ accuracy $\propto r \ll N$, # of sketches w/ random vectors $\mathbf{z}_i$

Random Trace Estimation



$$\mathrm{tr}(\mathbf{A}) \approx \frac{1}{r}\sum_{j=1}^{r} \mathbf{z}_j^\top \mathbf{A}\mathbf{z}_j = \frac{1}{r}\,\mathrm{tr}(\mathbf{Z}^\top \mathbf{A}\mathbf{Z})$$

Hutchinson, Michael F. "A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines." *Communications in Statistics-Simulation and Computation* 18.3 (1989): 1059-1076.
Avron, Haim, and Sivan Toledo. "Randomized algorithms for estimating the trace of an implicit symmetric positive semi-definite matrix." Journal of the ACM (JACM) 58.2 (2011): 1-34.
Meyer, Raphael A., et al. "Hutch++: Optimal Stochastic Trace Estimation." Symposium on Simplicity in Algorithms (SOSA). Society for Industrial and Applied Mathematics, 2021.

SLIM ML4Seismic

Randomized linear algebra

Randomized Trace Estimation:

$$\mathrm{tr}(\mathbf{A}) \approx \frac{1}{r}\sum_{j=1}^{r} \mathbf{z}_j^\top \mathbf{A}\mathbf{z}_j = \frac{1}{r}\mathrm{tr}(\mathbf{Z}^\top \mathbf{A}\mathbf{Z})$$

- only needs matrix-free access to actions of $\mathbf{A}$
- unbiased estimator when $\mathbb{E}(\mathbf{z}\mathbf{z}^\top) = \mathbf{I}$ w/ accuracy $\propto r \ll N$, # of sketches w/ random vectors $\mathbf{z}_j$
- errors studied & understood

Why should we care?

Adjoint state gradient

FWI objective

$$\Phi(\mathbf{m}) = \frac{1}{2}||\,\mathbf{P}_r\mathbf{A}^{-1}(\mathbf{m})\mathbf{P}_s^T\mathbf{q} - \mathbf{d}||_2^2$$

with gradient with respect to $\mathbf{m}$

$$\delta\mathbf{m}[\mathbf{x}] = \sum_{t=1}^{n_t} \ddot{\mathbf{u}}[t,\mathbf{x}]\mathbf{v}[t,\mathbf{x}]$$

Randomized trace estimation

Approximate FWI gradient calculation for $r \ll n_t$:

$$\delta \mathbf{m}[\mathbf{x}] = \operatorname{tr}\left(\ddot{\mathbf{u}}[t,\mathbf{x}]\mathbf{v}[t,\mathbf{x}]^\top\right) \approx \frac{1}{r}\operatorname{tr}\left((\mathbf{Z}^\top \ddot{\mathbf{u}}[\mathbf{x}])(\mathbf{v}[\mathbf{x}]^\top \mathbf{Z})\right)$$

- $\ddot{\mathbf{u}}$ second time derivative solution forward wave equation
- $\mathbf{v}$ solution adjoint wave equation
- $\sum \mathbf{x}_i \mathbf{y}_i = \mathbf{x}^\top \mathbf{y} = \operatorname{tr}(\mathbf{x}\mathbf{y}^\top)$
- probing vectors $\mathbf{Z} = \left[\mathbf{z}_1 \cdots \mathbf{z}_r\right]$ with $\mathbb{E}(\mathbf{z}_i^\top \mathbf{z}_i) = 1$

Meyer, R. A., Musco, C., Musco, C., and Woodruff, D. P., 2020, Hutch++: Optimal Stochastic Trace Estimation
Graff-Kray, M., Kumar, R., and Herrmann, F. J., 2017, Low-rank representation of omnidirectional subsurface extended image volumes:

Choice of probing vectors

Range of $\mathbf{u}$ leads to more accurate probing

QR decomposition on the range of $\mathbf{u}$ is too expensive

Use the observed data as a proxy (restriction of $\mathbf{u}$)

$$\left[\mathbf{Q}, \sim\right] = \mathrm{qr}(\mathbf{AZ}) \quad \text{with} \quad \mathbf{A} = \mathbf{D}_{\text{obs}}\mathbf{D}_{\text{obs}}^{\top}$$

Data $\mathbf{D}_{\text{obs}}$ corresponds to

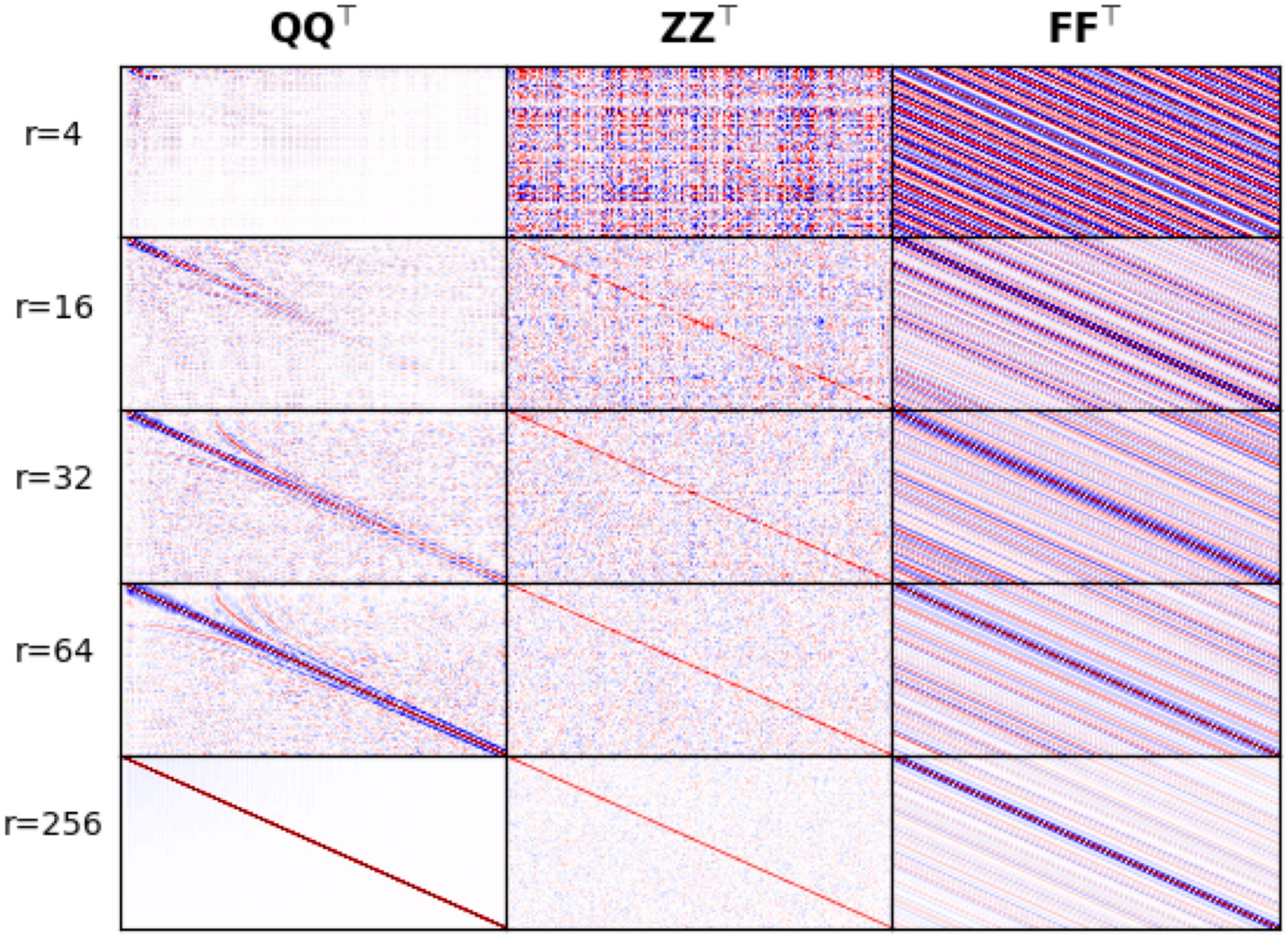
- restriction of the true wavefield to the receivers
- is representative of its range (frequency content, travel time, ...)

Crosstalk

Stronger diagonal

Less crosstalk

Less coherent noise



Z : Random +-1
F: DFT

SLIM ML4Seismic

Approximate gradient FWI/RTM

Algorithm:

Accumulate over time

0. **for t=2:nt-1** # forward propagation
1. $\mathbf{u}[t+1] = f(\mathbf{u}[t], \mathbf{u}[t-1], \mathbf{m}, \mathbf{q}[t])$
2. $\boxed{\overline{\ddot{\mathbf{u}}}[\mathbf{r}, \mathbf{x}] \mathrel{+}= \mathbf{Q}[\mathbf{r}, t]\ddot{\mathbf{u}}[t, \mathbf{x}] \quad \forall\ \mathbf{r}}$
3. **end for**
4. **for t=nt:-1:1** # back propagation
5. $\mathbf{v}[t-1] = f^\top(\mathbf{v}[t], \mathbf{v}[t+1], \mathbf{m}, \delta\mathbf{d}[t])$
6. $\boxed{\overline{\mathbf{v}}[\mathbf{r}, \mathbf{x}] \mathrel{+}= \mathbf{Q}[\mathbf{r}, t]\mathbf{v}[t, \mathbf{x}] \quad \forall\ \mathbf{r}}$
7. **end for**
8. output: $\frac{1}{r}\,\mathrm{tr}(\overline{\ddot{\mathbf{u}}}\,\overline{\mathbf{v}}^\top)$

$$\ddot{\mathbf{u}}, \mathbf{v} \in \mathbb{R}^{n_t \times N} \implies \overline{\ddot{\mathbf{u}}}, \overline{\mathbf{v}} \in \mathbb{R}^{r \times N},\ r \ll n_t$$

Philipp A. Witte, Mathias Louboutin, Fabio Luporini, Gerard J. Gorman, and Felix J. Herrmann, "Compressive least-squares migration with on-the-fly Fourier transforms", Geophysics, vol. 84, pp. R655-R672, 2019
McMechan, G. A., 1983, Migration by extrapolation of time-dependent boundary values: Geophysical Prospecting, 31, 413–420.

Randomized trace estimation

Ultra-low memory use:

	FWI	DFT	Probing	Optimal checkpointing	Boundary reconstruction
Compute	0	$\mathcal{O}(2r) \times n_t \times N$	$\mathcal{O}(r) \times n_t \times N$	$\mathcal{O}(log(n_t)) \times N \times n_t$	$n_t \times N$
Memory	$N \times n_t$	$2r \times N$	$r \times N$	$\mathcal{O}(10) \times N$	$n_t \times N^{\frac{2}{3}}$

For fixed $r \ll n_t$

- half memory cost of DFT
- half compute cost of DFT
- simple real-valued algorithm

In practice, needs much smaller r compared to DFT.

Op't Root, T. J., C. C. Stolk, and M. V. de Hoop, 2012, Linearized inverse scattering based on seismic reverse time migration: Journal de Mathematiques Pures et Appliquees, 98, 211–238.

Randomized trace estimation

Ultra-cheap imaging conditions:

$$\mathrm{Q}^{\top}\left(\mathbf{D}_x \mathbf{u}[\cdot,\mathbf{x}]\right) = \mathbf{D}_x\left(\mathbf{Q}^{\top}\mathbf{u}[\cdot,\mathbf{x}]\right)$$

Apply space-only imaging condition to time-compressed wavefields:

- $\mathbf{k}$-space filter
- inverse-scattering imaging condition (ISIC)

Imaging condition usually costs extra(s) PDEs (ISIC = 1 PDE)

$$\mathcal{I}(\mathbf{u}[t],\mathbf{v}[t]) = \sum_t \mathbf{m}\ddot{\mathbf{u}}[t]\mathbf{v}[t] + \nabla\mathbf{u}[t]\,.\nabla\mathbf{v}[t]$$

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

SLIM
ML4Seismic

Randomized trace estimation

Subsurface Common-Image Gathers (CIGs):

$$\delta\mathcal{M}[\mathbf{x},\mathbf{h}] \approx \frac{1}{r}\,\mathrm{tr}\left(\bar{\ddot{\mathbf{u}}}[\cdot,\mathbf{x}+\mathbf{h}]\bar{\mathbf{v}}[\cdot,\mathbf{x}-\mathbf{h}]^\top\right)$$

$\mathbf{h}$ subsurface offset

- computed in compressed space
- reduced memory footprint
- less computational cost

FWI example

2D overthrust model

OBN acquisition

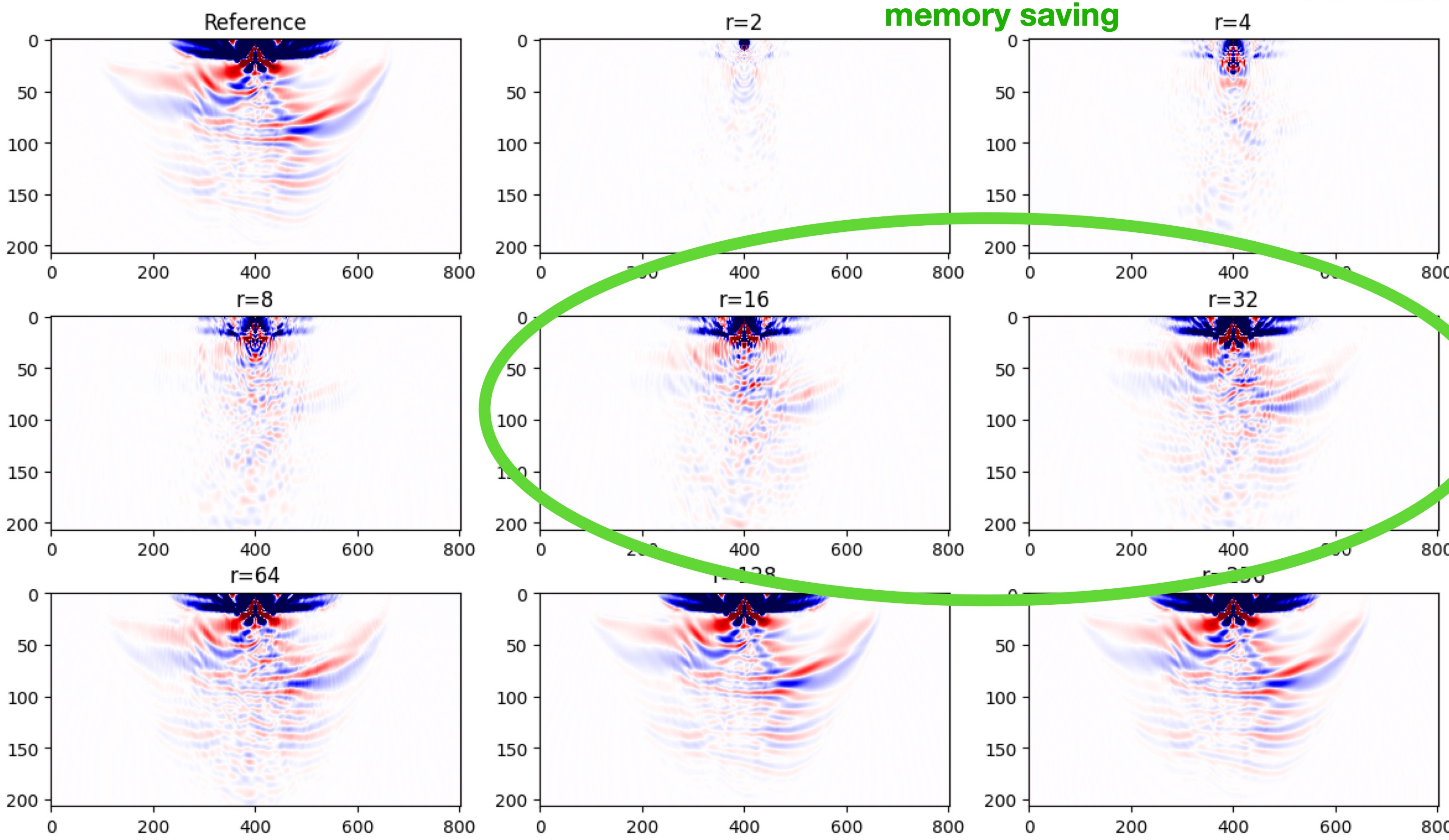
Comparisons:

- standard FWI
- on-the-fly DFT
- randomized trace estimation

Accuracy – gradients

Acceptable accuracy
$50\times - 100\times$
memory saving

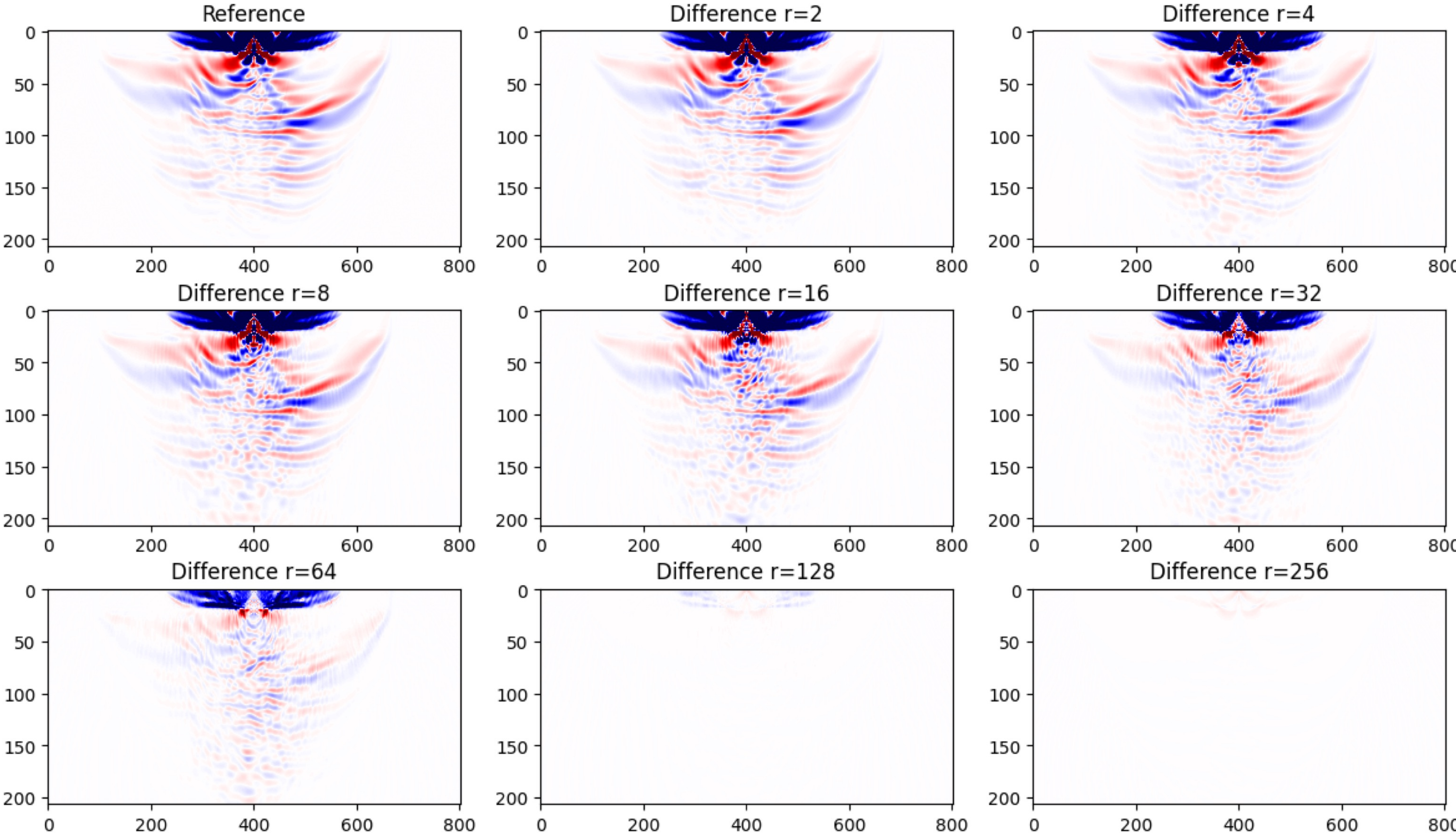
- converges to true gradient as $r \rightarrow n_t$
- less accurate near source



Accuracy – gradients

SLIM ML4Seismic

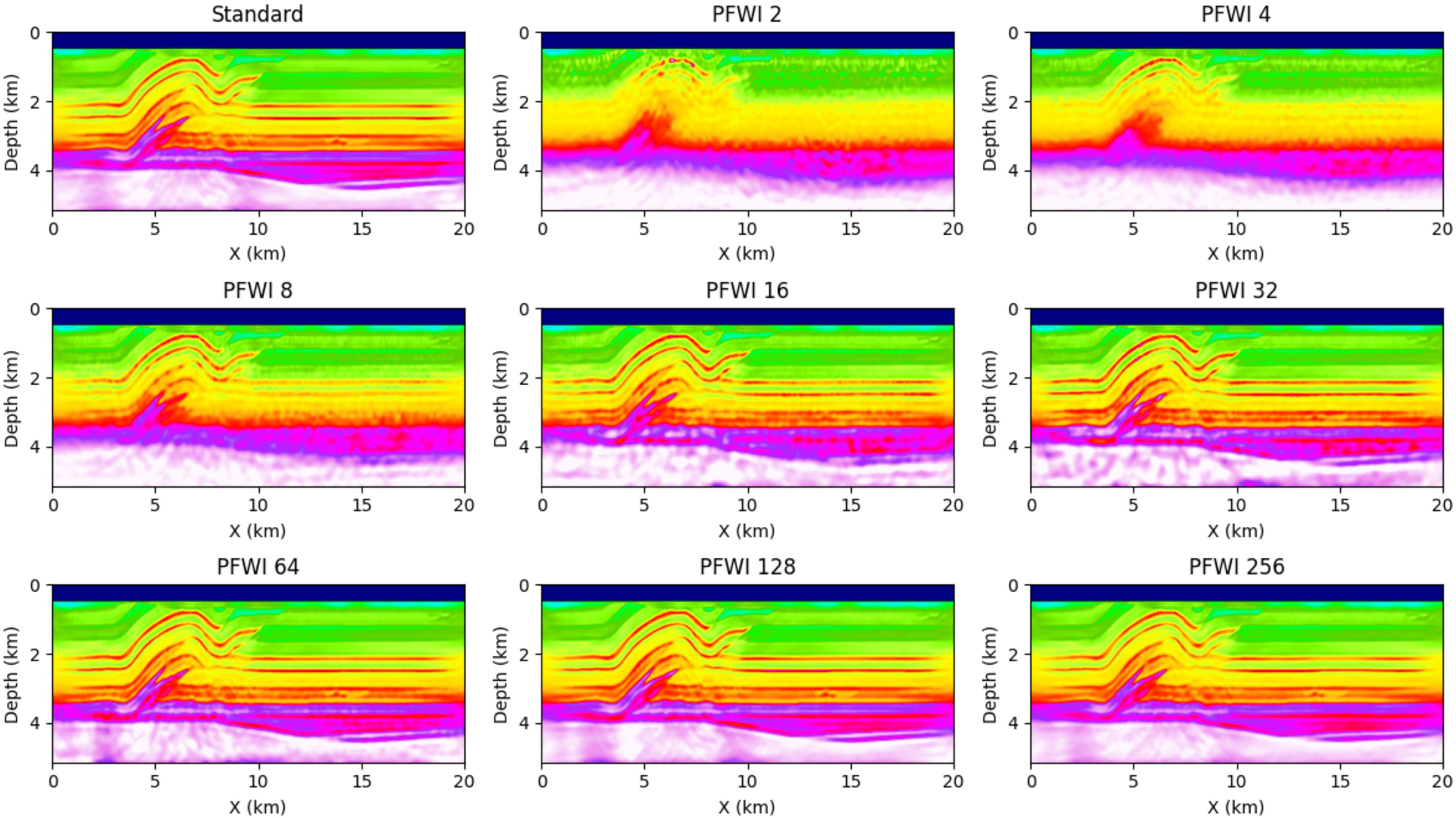
- Exact for large r
- Noisy error



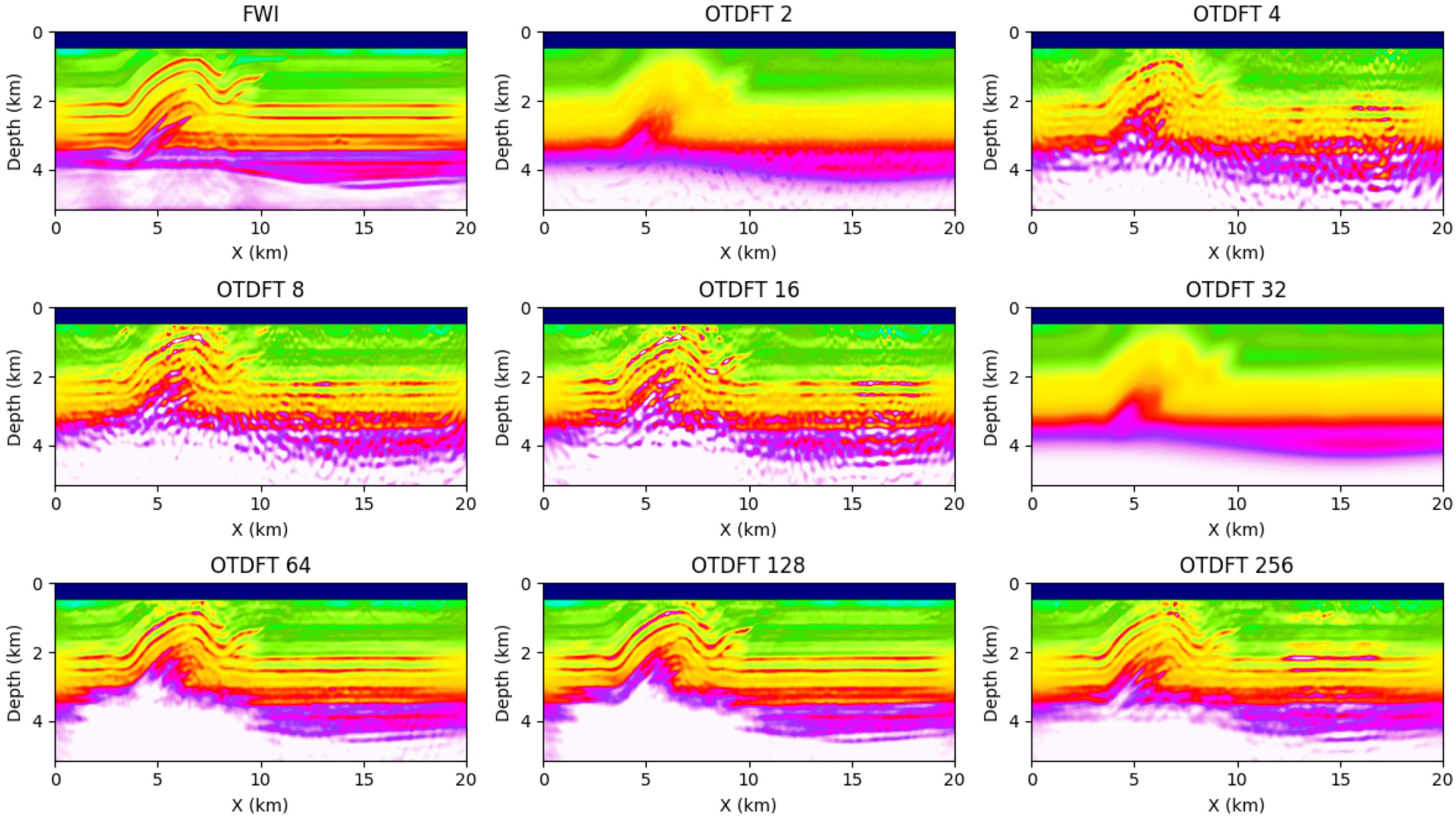
Accurate but becomes expensive

FWI w/ randomized Trace estimation

SLIM
ML4Seismic



FWI w/ on-the-fly DFT

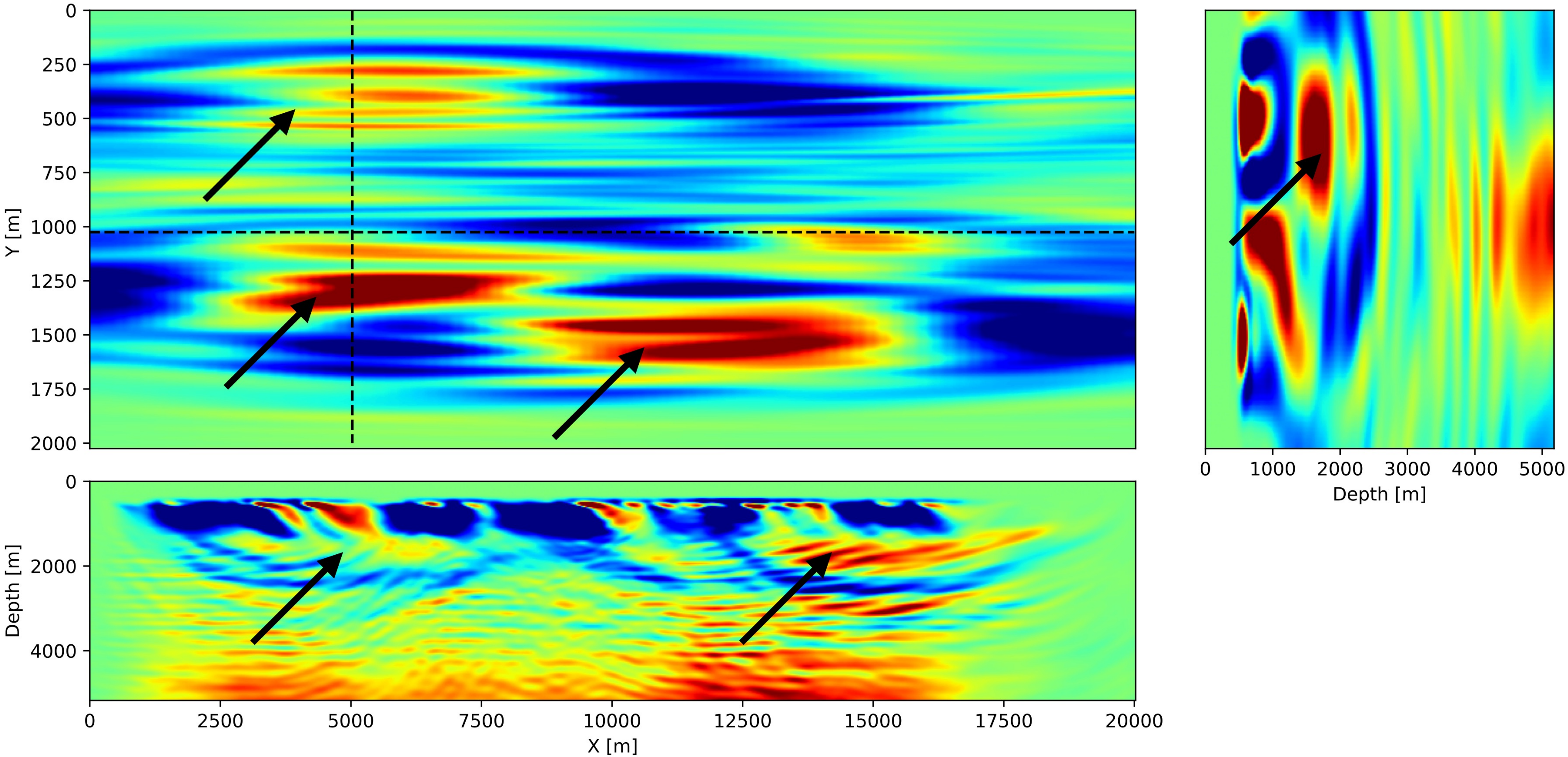


3D, first gradient

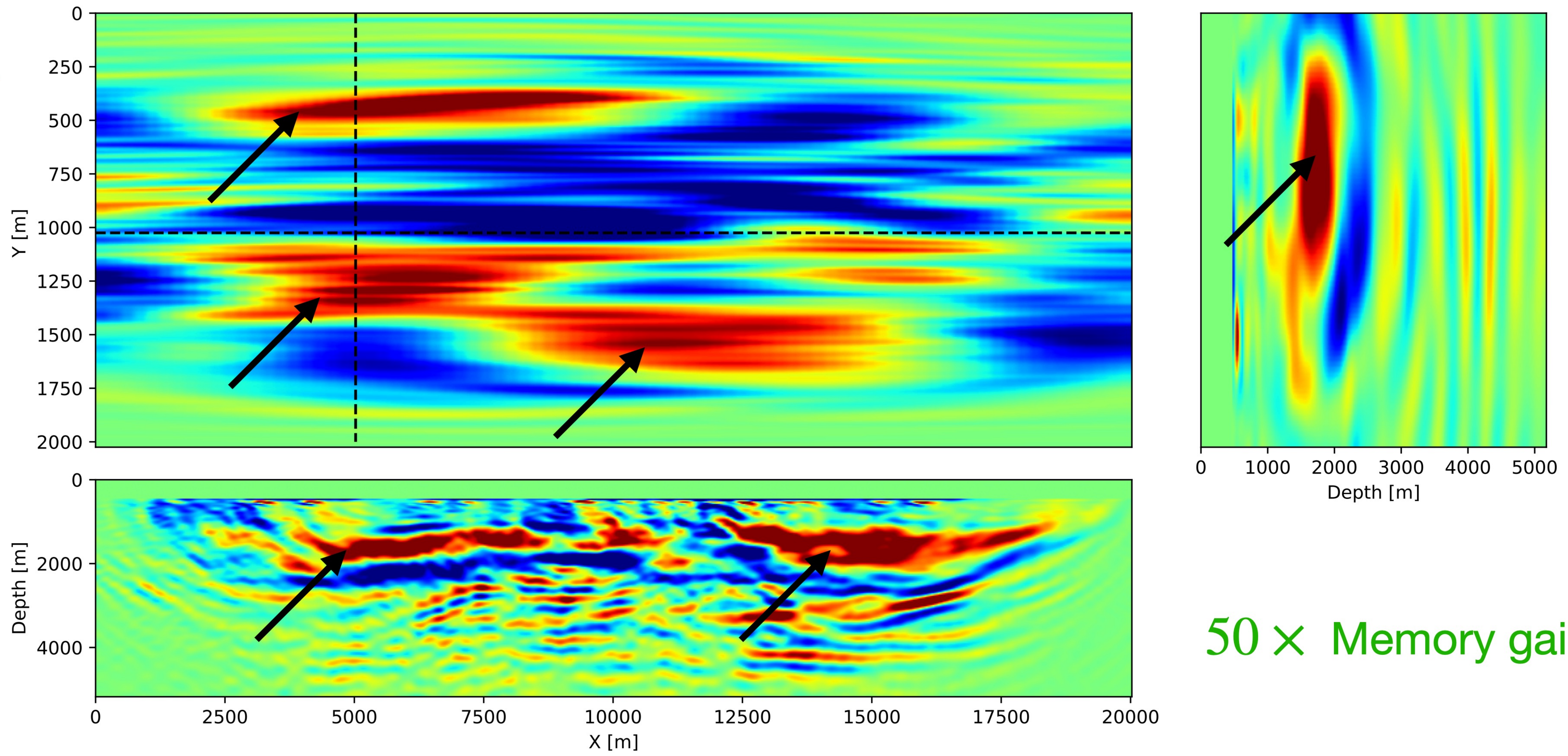
Overthrust 3D:

- Marine acquisition
- 12.5Hz Ricker wavelet bandpass filtered at 3-15Hz
- 32 probing vectors $\implies 40\times$ memory reduction
- Probing on GPU (NVidia M60, \$45/hr)
- True gradient on CPU (Intel Skylake, \$65/hour)

True gradient



Estimated gradient with 32 vectors



$50\times$ Memory gain

Imaging

- Sparse OBN
- TTI imaging
- Subsurface common image gather

Makes RTM conducive to acceleration w/ GPUs

All on Azure NC6 (NVidia M60 w/ 8Gb memory)

RTM

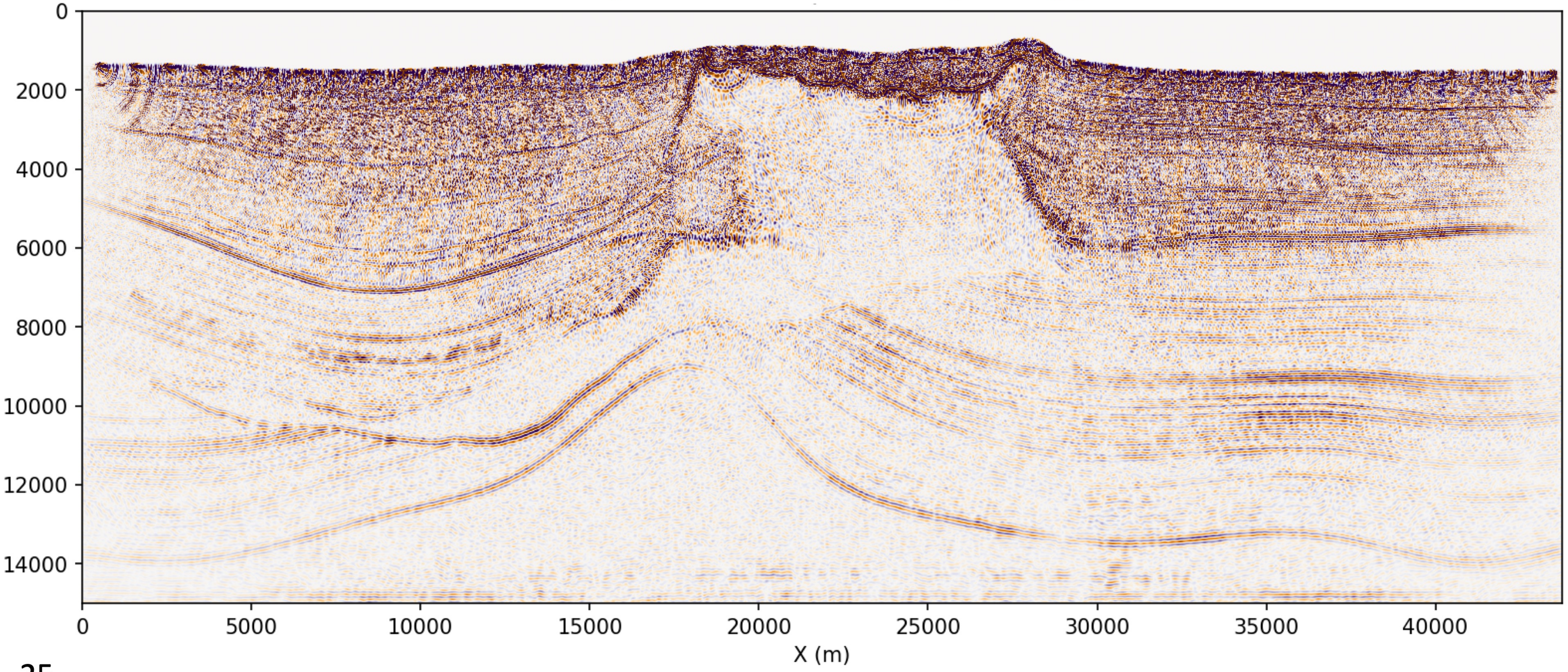
SEAM 2D:

- 44 OBN 1km apart
- 3521 sources 12.5m apart
- 14.5Hz Ricker wavelet
- 64 probing vectors ($160\times$ memory savings, 84Gb vs .5Gb)

Makes RTM conducive to acceleration w/ GPUs

Noisy but accurate

RTM ($r = 64 \rightarrow 160\times$ memory savings)



TTI RTM

BP 2D TTI:

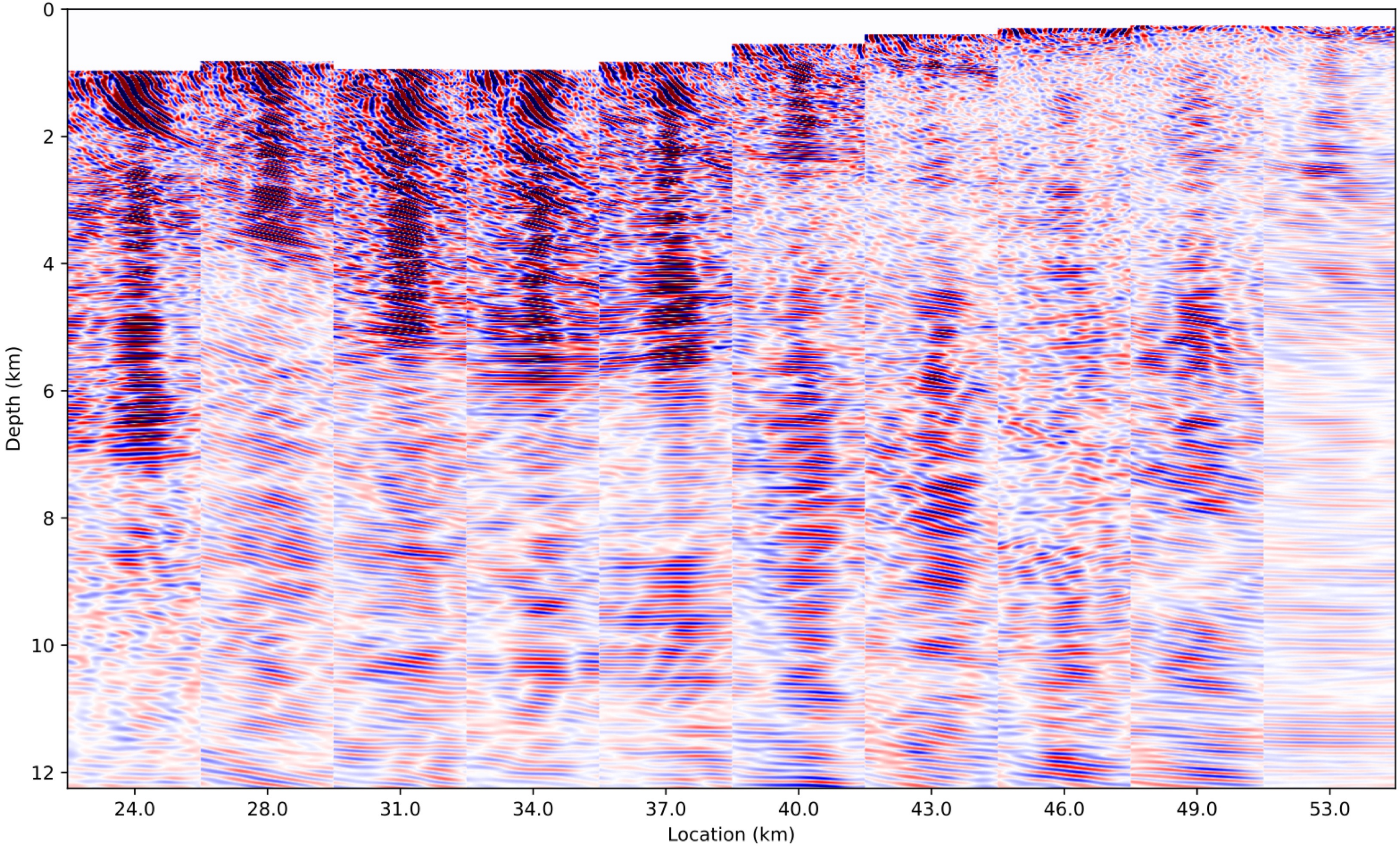
- 1600 Marine sources
- 8km offset
- 19.5Hz Ricker wavelet
- 64 probing vectors ($160 \times$ memory savings, 84Gb vs .5Gb)

Makes TTI-RTM conducive to acceleration w/ GPUs

Noise stacks out /w dense source sampling

Subsurface CIG

SLIM ML4Seismic



Conclusions

Leverage Randomized Linear Algebra

Low-memory foot print and low algorithmic complexity

Controllable error

Allows for accelerators (GPUs)

Drop-in extension for existing open source framework JUDI/Devito

Discussion

Proof of concept

- incurs somewhat of computational hit (negating acceleration of GPUs)
- we stay 100% on GPU
- as w/ optimal checkpointing memory gains offset by computational overhead
- implementation can be improved so overhead is minimal

In theory, we outperform DFT methods

- similar structure
- real valued
- require fewer basis vectors

Open source software

TimeProbeSeismic.jl:

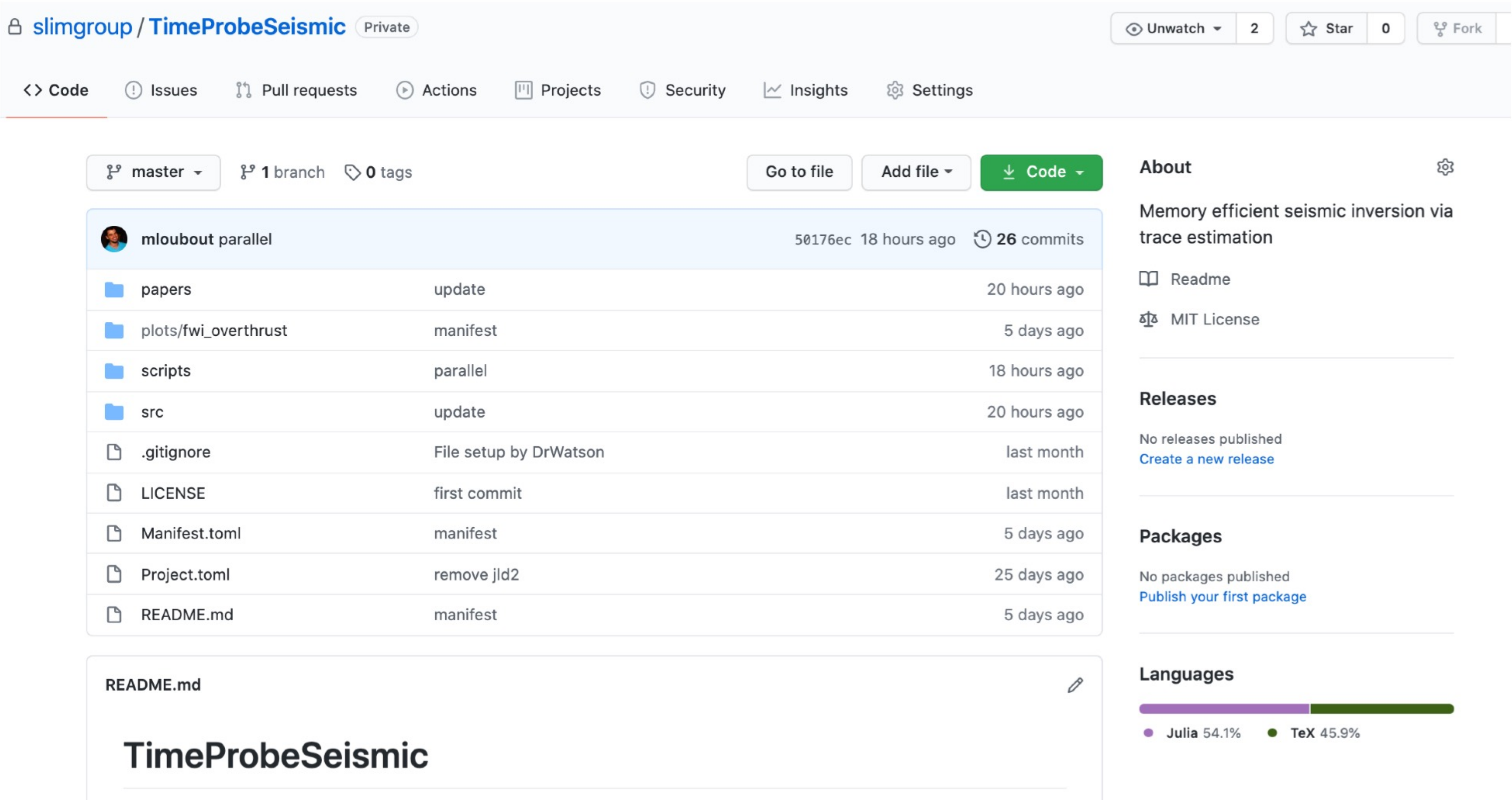
- Open-source MIT license
- Built on top of JUDI.jl
- Leverages Devito
- https://github.com/slimgroup/TimeProbeSeismic.jl

ImageGather.jl:

- Open-source MIT license
- Built on top of JUDI.jl
- Leverages Devito
- https://github.com/slimgroup/ImageGather.jl

```julia
# Standard JUDI
function objective_function(x)
  model0.m .= x
   f, g = fwi_objective(model0, q[idx], d_obs[idx]; options=opt)
end
options = spg_options(verbose = 3, maxIter = fevals, memory = 3,
iniStep = 1f0)
g_const = 0
sol = spg(x->objective_function(x), vec(m0), ProjBound, options)

# Probing extension
function objective_function(x, ps)
  model0.m .= x
   f, g = fwi_objective(model0, q[idx], d_obs[idx], ps; options=opt)
end
ps = 32
global g_const = 0
sol = spg(x->objective_function(x, ps), vec(m0), ProjBound, options)
```



This research was carried out with the support of Georgia Research Alliance and partners of the ML4Seismic consortium.