Centrum Wiskunde & Informatica

Computational and Experimental Challenges of 3D Ultrasound Tomography

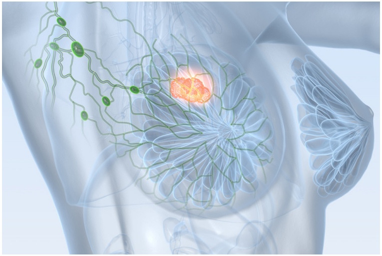
Felix Lucka (he/him/his)

Conference on Mathematics of Wave Phenomena
14 Feb 2022

Motivation: Breast Cancer Imaging

Most common cause of cancer death in women worldwide.

- 25% of all cancer cases in women
- 15% of all cancer deaths in women

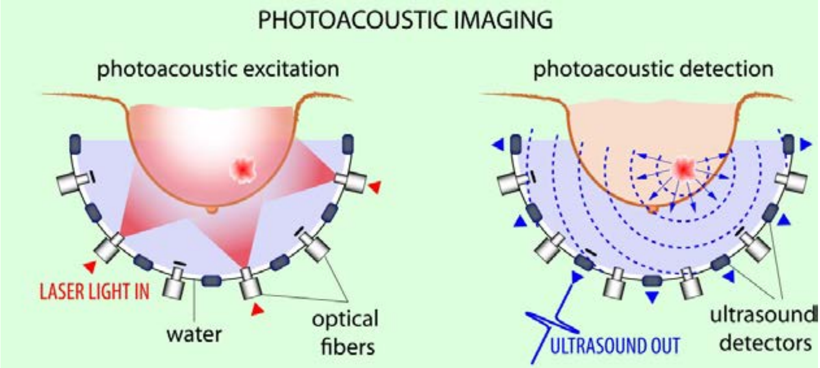


Despite advances in early detection and diagnosis:

Urgent need for novel imaging techniques providing higher specificity, contrast and image resolution than X-ray mammography at lower costs than MRI.

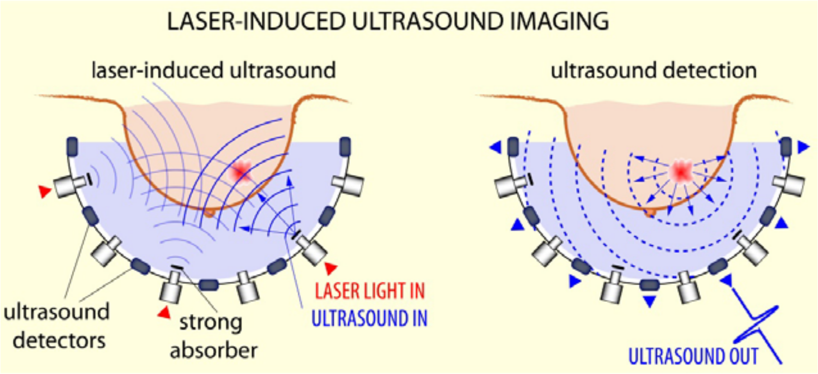
Quantitative Photoacoustic Breast Imaging

- hybrid imaging: "light in, sound out"
- non-ionizing, near-infrared radiation
- quantitative images of optical properties
- novel diagnostic information

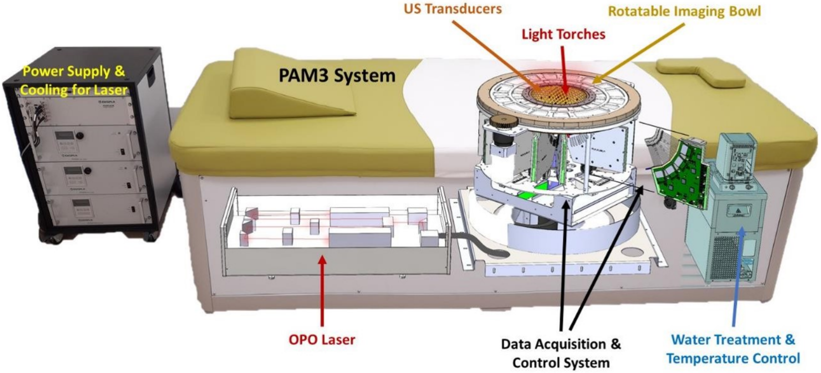


Quantitative Ultrasonic Breast Imaging

- "sound in, sound out"
- different from conventional US but as safe
- quantitative images of acoustic properties
- novel diagnostic information



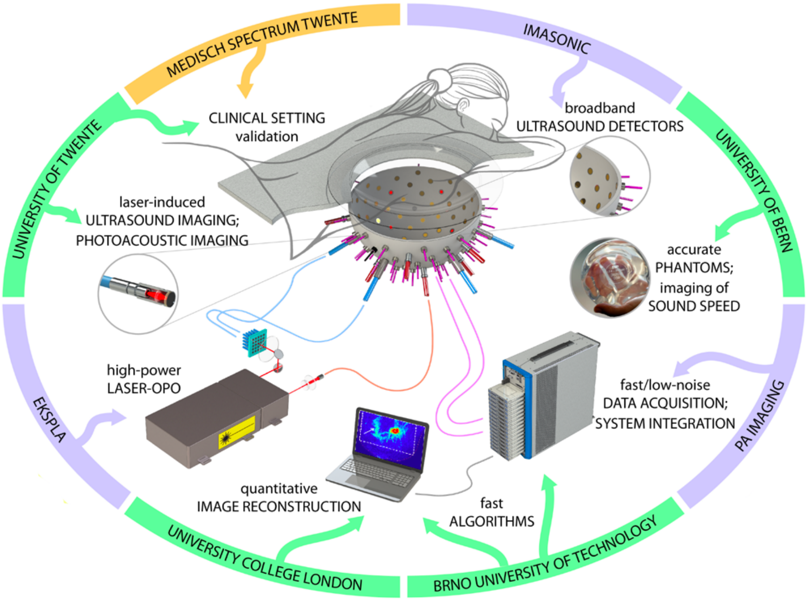
Photoacoustic and Ultrasonic Mammography Scanner



Aim: novel diagnostic information from high resolution maps of optical and acoustic properties

- 512 US transducers on rotatable half-sphere
- 40 optical fibers for photoacoustic excitation

Partners in H2020 Project



UST Reconstruction Approaches

$$(c(x)^{-2}\partial_t^2 - \Delta)p_i(x,t) = s_i(x,t), \qquad f_i = M_i p_i, \qquad i = 1, \dots, n_{src}$$

Travel time tomography (TTT): geometrical optics approximation.

- ✓ robust & computationally efficient
- ! valid for high frequencies (attenuation!), low res, lots of data

Reverse time migration (RTM): forward wavefield correlated in time with backward wavefield (adjoint wave equation) via imaging condition.

- ✓ 2 wave simulations, better quality than TTT.
- ! approximation, needs initial guess, quantitative errors

Full waveform inversion (FWI): fit full model to all data.

- ✓ high res from little data, transducer modelling, constraints
- ! many wave simulations, complex numerical optimization
- low TRL but already used in 2D systems

time domain vs frequency domain methods

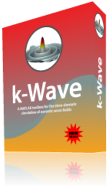
Acoustic Wave Propagation: Numerical Solution

- **Direct methods**, such as finite-difference, pseudospectral, finite/spectral element, discontinous Galerkin.
- **Integral equation methods**, e.g. boundary element
- **Asymptotic methods**, e.g., geometrical optics, Gaussian beams

Acoustic Wave Propagation: Numerical Solution

- Direct methods, such as finite-difference, **pseudospectral**, finite/spectral element, discontinous Galerkin.
- Integral wave equation methods, e.g. boundary element.
- Asymptotic methods, e.g., geometrical optics, Gaussian beams.

k-Wave: *k*-space pseudospectral method solving the underlying system of first order conservation laws.

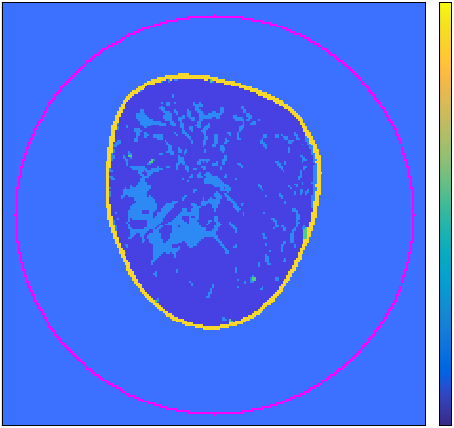


- Compute spatial derivatives in Fourier space: 3D FFTs.
- Parallel/GPU computing leads to massive speed-ups.
- Modify finite temporal differences by *k*-space operator and use staggered grids for accuracy and robustness.
- Perfectly matched layer to simulate free-space propagation.

B. Treeby and B. Cox, 2010. k-Wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields, *Journal of Biomedical Optics.*

FWI Illustration in 2D

SOS ground truth c^{true}

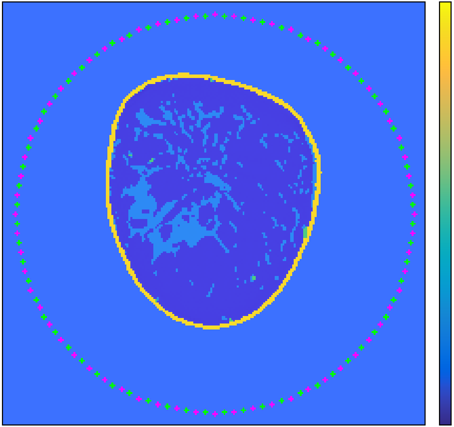


color range 1450 - 1670 m/s

- 1mm resolution
- 222^2 voxel
- 836 voxels on surface (pink)
- TTT would need 836^2 source-receiver combos for high res result

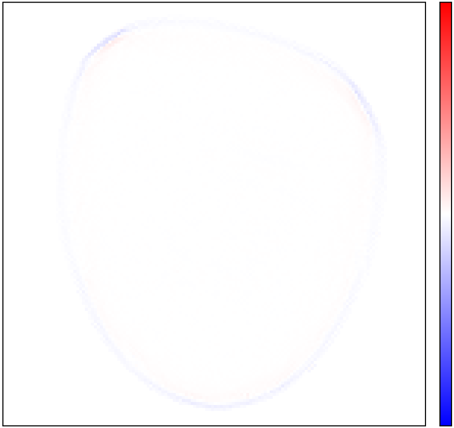
FWI Illustration in 2D: 64 Sensors, 64 Receivers

SOS reconstruction c^{rec}



color range 1450 - 1670 m/s

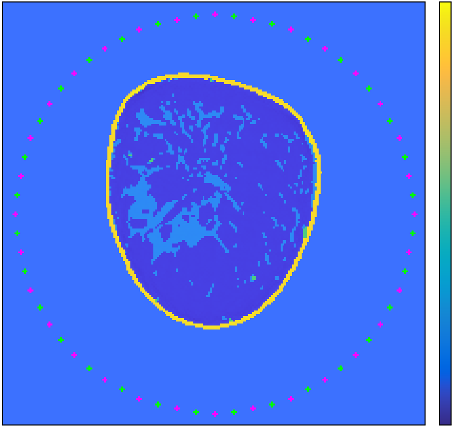
reconstruction error $c^{true} - c^{rec}$



color range -50 - 50 m/s

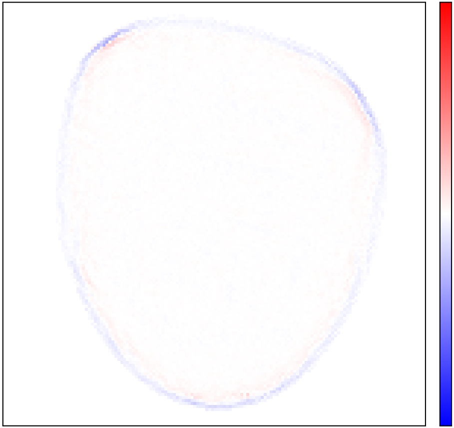
FWI Illustration in 2D: 32 Sensors, 32 Receivers

SOS reconstruction c^{rec}



color range 1450 - 1670 m/s

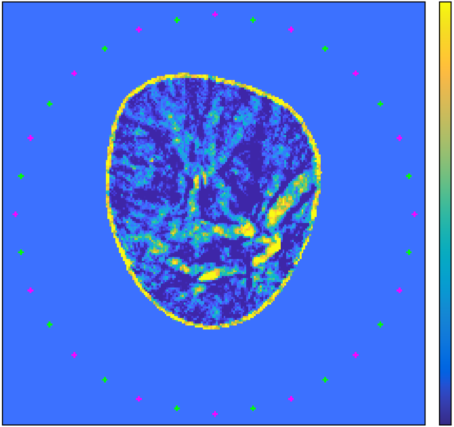
reconstruction error $c^{true} - c^{rec}$



color range -50 - 50 m/s

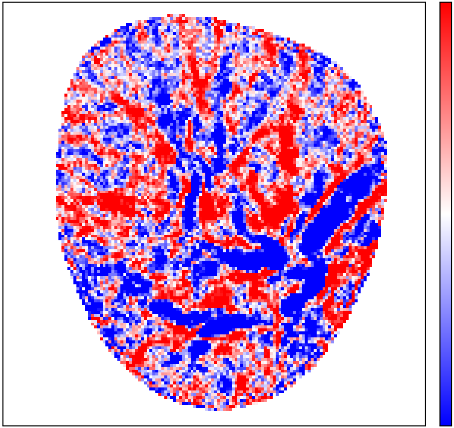
FWI Illustration in 2D: 16 Sensors, 16 Receivers

SOS reconstruction c^{rec}



color range 1450 - 1670 m/s

reconstruction error $c^{true} - c^{rec}$



color range -50 - 50 m/s

Time Domain Full Waveform Inversion

$$A(c)p_i := (c^{-2}\partial_t^2 - \Delta)p_i = s_i, \qquad f_i = M_i\, p_i, \quad i = 1, \dots, n_{src}$$

$$\min_{c \in \mathcal{C}} \sum_i^{n_{src}} \mathcal{D}\left(f_i(c), f_i^\delta\right) \quad s.t. \quad f_i(c) = M_i A^{-1}(c) s_i$$

gradient for **first-order optimization** via **adjoint state method**:

$$\nabla_c \mathcal{D}\left(f(c), f^\delta\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t) \quad ,$$

where $(c^{-2}\partial_t^2 - \Delta)q^* = s^*$, $s^*(x,t)$ is time-reversed data discrepancy

$\rightarrow$ **two wave simulations for one gradient**

Starting point in 2D:

Pérez-Liva, Herraiz, Udías, Miller, Cox, Treeby 2017. Time domain reconstruction of sound speed and attenuation in ultrasound computed tomography using full wave inversion, *JASA*.

3D Time Domain FWI for Breast UST

$$\min_{c \in \mathcal{C}} \sum_{i}^{n_{src}} \mathcal{D}\left(M_i A^{-1}(c) s_i, f_i^{\delta}\right)$$

$$\nabla_c \mathcal{D}\left(f(c), f^{\delta}\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)$$

Challenges and solutions for 3D:

- **!** 2 x n_{src} wave simulations per gradient
- **!** computationally & stochastically efficient gradient estimator
- **!** memory requirements of gradient computation
- **!** slow convergence and local minima
- **!** computational resources

3D Time Domain FWI for Breast UST

$$\min_{c \in \mathcal{C}} \sum_{i}^{n_{src}} \mathcal{D}\left(M_i A^{-1}(c) s_i, f_i^\delta\right)$$

$$\nabla_c \mathcal{D}\left(f(c), f^\delta\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)$$

Challenges and solutions for 3D:

- ! 2 x n_{src} wave simulations per gradient
  $\longrightarrow$ **stochastic quasi-newton optimization (SL-BFGS)**
- ! computationally & stochastically efficient gradient estimator
- ! memory requirements of gradient computation
- ! slow convergence and local minima
- ! computational resources

3D Time Domain FWI for Breast UST

$$\min_{c \in \mathcal{C}} \sum_{i}^{n_{src}} \mathcal{D}\left(M_i A^{-1}(c) s_i, f_i^{\delta}\right)$$

$$\nabla_c \mathcal{D}\left(f(c), f^{\delta}\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)$$

Challenges and solutions for 3D:

! 2 x n_{src} wave simulations per gradient
$\longrightarrow$ **stochastic quasi-newton optimization (SL-BFGS)**

! computationally & stochastically efficient gradient estimator
$\longrightarrow$ **source encoding for time-invariant systems**

! memory requirements of gradient computation

! slow convergence and local minima

! computational resources

3D Time Domain FWI for Breast UST

$$\min_{c \in \mathcal{C}} \sum_{i}^{n_{src}} \mathcal{D}\left(M_i A^{-1}(c) s_i, f_i^\delta\right)$$

$$\nabla_c \mathcal{D}\left(f(c), f^\delta\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)$$

Challenges and solutions for 3D:

! 2 x n_{src} wave simulations per gradient
$\longrightarrow$ **stochastic quasi-newton optimization (SL-BFGS)**

! computationally & stochastically efficient gradient estimator
$\longrightarrow$ **source encoding for time-invariant systems**

! memory requirements of gradient computation
$\longrightarrow$ **time-reversal based gradient computation**

! slow convergence and local minima

! computational resources

3D Time Domain FWI for Breast UST

$$\min_{c \in \mathcal{C}} \sum_{i}^{n_{src}} \mathcal{D}\left(M_i A^{-1}(c) s_i, f_i^\delta\right)$$

$$\nabla_c \mathcal{D}\left(f(c), f^\delta\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)$$

Challenges and solutions for 3D:

- ! 2 x n_{src} wave simulations per gradient
  $\longrightarrow$ **stochastic quasi-newton optimization (SL-BFGS)**
- ! computationally & stochastically efficient gradient estimator
  $\longrightarrow$ **source encoding for time-invariant systems**
- ! memory requirements of gradient computation
  $\longrightarrow$ **time-reversal based gradient computation**
- ! slow convergence and local minima
  $\longrightarrow$ **coarse-to-fine multigrid schemes**
- ! computational resources

3D Time Domain FWI for Breast UST

$$\min_{c \in \mathcal{C}} \sum_{i}^{n_{src}} \mathcal{D}\left(M_i A^{-1}(c) s_i, f_i^\delta\right)$$

$$\nabla_c \mathcal{D}\left(f(c), f^\delta\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)$$

Challenges and solutions for 3D:

! 2 x n_{src} wave simulations per gradient

$\longrightarrow$ **stochastic quasi-newton optimization (SL-BFGS)**

! computationally & stochastically efficient gradient estimator

$\longrightarrow$ **source encoding for time-invariant systems**

! memory requirements of gradient computation

$\longrightarrow$ **time-reversal based gradient computation**

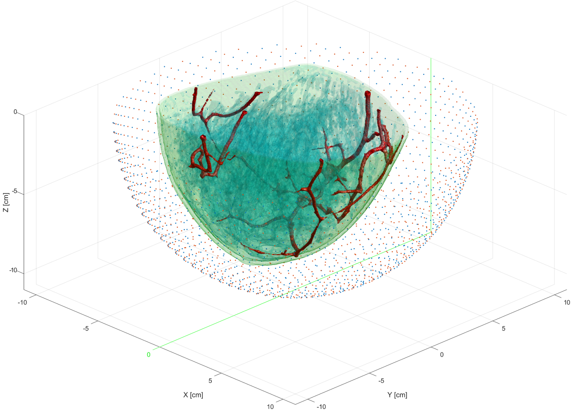
! slow convergence and local minima

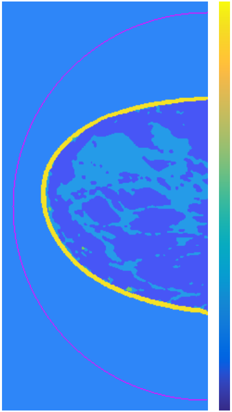
$\longrightarrow$ **coarse-to-fine multigrid schemes**

! computational resources

$\longrightarrow$ **runs on single GPU, can utilize multiple GPUs**

3D FWI: Setup

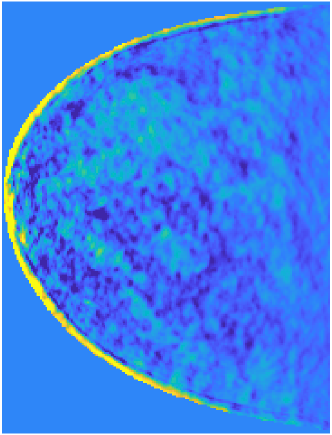




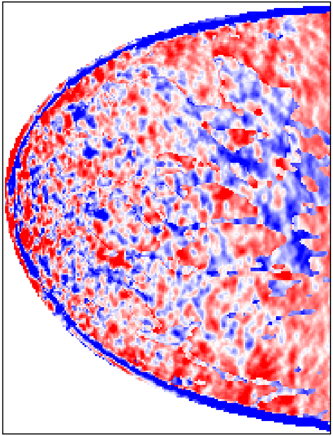
- color range 1435-1665 m/s
- 3D breast phantom at 0.5mm resolution, 1024 sources and receivers
- $442 \times 442 \times 222$ voxel, 3912 time steps

Yang Lou et al. Generation of anatomically realistic numerical phantoms for photoacoustic and ultrasonic breast imaging, *JBO, 2017.*

Starting point in 24h on desktop with single GPU



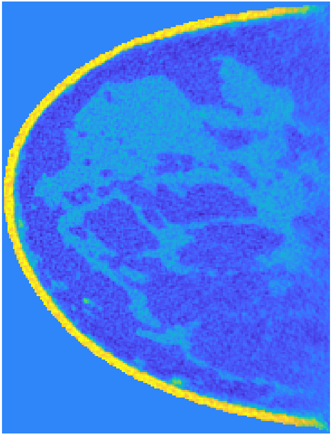
color range 1435 to 1665 m/s



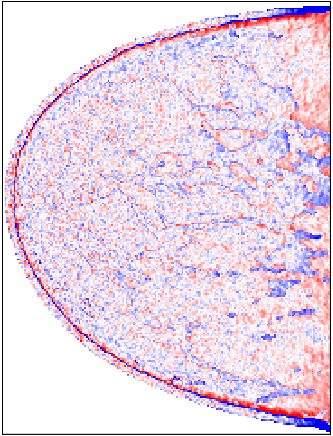
color range -50 to +50 m/s

- single grid
- SGD
- normal single source gradient estimator

3D FWI in 24h on desktop with single GPU



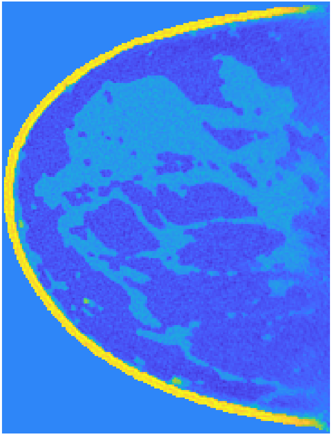
color range 1435 to 1665 m/s



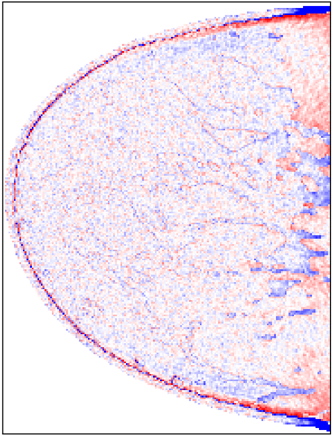
color range -50 to +50 m/s

- multi-grid with 3 level, coarsening factor 2
- SL-BFGS, slowness transform, prog. iter averaging
- time-reversal based source encoding gradient estimator

3D FWI in 24h on cluster with 4 GPU



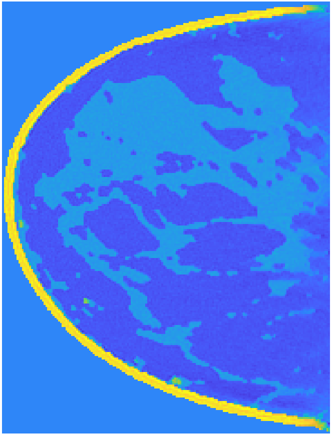
color range 1435 to 1665 m/s



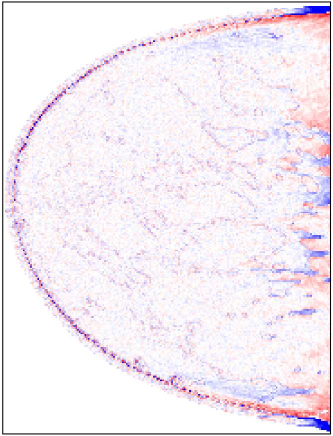
color range -50 to +50 m/s

- multi-grid with 3 level, coarsening factor 2
- SL-BFGS, slowness transform, prog. iter averaging
- time-reversal based source encoding gradient estimator

3D FWI in 24h on cluster with 16 GPU



color range 1435 to 1665 m/s



color range -50 to +50 m/s

- multi-grid with 3 level, coarsening factor 2
- SL-BFGS, slowness transform, prog. iter averaging
- time-reversal based source encoding gradient estimator

FWI for Experimental Data: Where Are We?

✓ data from phantom objects, volunteers & patients

✓ ray-based SOS reconstructions

✓ photoacoustic reconstructions $\rightarrow$ data pre-processing, scanner & transducer modelling, wave simulations

✓ modeling of US protocol, data read-in & pre-processing

! model calibration

! FWI of phantom objects, quantitative evaluation

! FWI of volunteer data

! FWI of phantom objects, clinical evaluation

Summary & Outlook

Summary:

- proof-of-concept studies of TD-FWI for high resolution 3D UST
- combination of computational tricks: multi-grid, SLBFGS, source encoding, preconditioning, time-reversal based gradient estimation
- reasonable results within 24h
- extensions towards experimental data

Outlook:

- **reconstruction of experimental data!**
- multi-GPU CUDA code
- refined source/receiver modeling
- extension to acoustic attenuation, density, etc.

Thank you for your attention!

L, Pérez-Liva, Treeby, Cox, 2021. High Resolution 3D Ultrasonic Breast Imaging by Time-Domain Full Waveform Inversion, *Inverse Problems 38(2).*

Challenges of High-Resolution FWI in 3D

$$\min_{c \in \mathcal{C}} \sum_{i}^{n_{src}} \mathcal{D}\left(f_i(c), f_i^\delta\right) \quad s.t. \quad f_i(c) = M_i A^{-1}(c) s_i$$

$$\nabla_c \mathcal{D}\left(f(c), f^\delta\right) = 2 \int_0^T \frac{1}{c(x)^3} \left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)$$

PAMMOTH scanner example:

- 0.5mm res: comp grid $560 \times 560 \times 300$ voxel $= 94$M, ROI $= 7$M
- 1024 transducers, 4000 time samples (multiple sources);

Gradient computation:

- 1 wave sim: $\sim$30 min.
- ! **2 wave sim per source**, $n_{src} = 1024 \rightarrow 20$ days per gradient.
  stochastic gradient methods $\rightarrow$ 60 min per gradient
- ! **storage of forward field** in ROI: $\sim$ 200GB.
  time-reversal based gradient computation $\rightarrow 5 - 25$GB.

Stochastic Gradient Optimization

$$\mathcal{J} := n_{src}^{-1} \sum_{i}^{n_{src}} \mathcal{D}_i(c) := n_{src}^{-1} \sum_{i}^{n_{src}} \mathcal{D}\left(M_i A^{-1}(c) s_i, f_i^{\delta}\right)$$

approx $\nabla \mathcal{J}$ by $|\mathcal{S}|^{-1} \sum_{j \in \mathcal{S}} \nabla \mathcal{D}_j(c)$, $\mathcal{S} \subset \{1, \ldots, n_{src}\}$ predetermined.
$\rightarrow$ **incremental gradient, ordered sub-set methods**

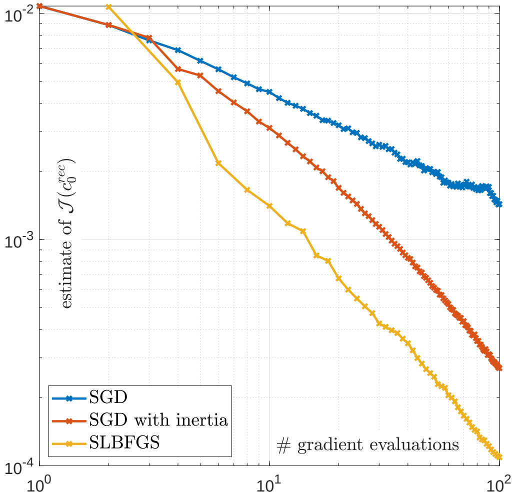
Instance of **finite sum minimization** similar to **training in machine learning**. Use **stochastic gradient descent (SGD)**:

- momentum, gradient/iterate averaging (SAV, SAGA), variance reduction (SVRG), choice of step size, mini-batch size
- include non-smooth regularizers (SPDHG, SADMM)
- quasi-Newton-type methods, e.g., **stochastic L-BFGS**

Bottou, Curtis, Nocedal. Optimization Methods for Large-Scale Machine Learning, *arXiv:1606.04838.*

Fabien-Ouellet, Gloaguen, Giroux, 2017. A stochastic L-BFGS approach for full-waveform inversion, *SEG.*

Stochastic Gradient Optimization



Gradient Estimates: Sub-Sampling vs Source Encoding

Computationally & stochastically efficient gradient estimator?

Source Encoding for linear PDE constraints:

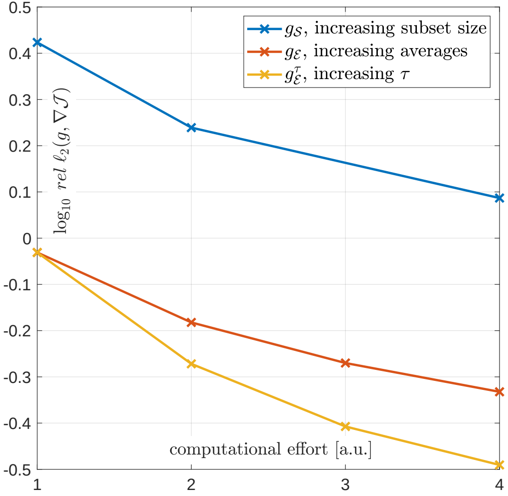
$$\text{Let} \quad \hat{s} := \sum_{i}^{n_{srt}} w_i s_i, \quad \hat{f}^{\delta} := \sum_{i}^{n_{srt}} w_i f_i^{\delta}, \quad \text{with} \quad \mathbb{E}\left[w\right] = 0,\ \mathbb{C}\text{ov}[w] = I,$$

$$\text{then} \quad \mathbb{E}\left[\nabla \left\| MA^{-1}(c)\hat{s} - \hat{f}^{\delta} \right\|_2^2 \right] = \nabla \sum_{i}^{n_{src}} \left\| MA^{-1}(c)s_i - f_i^{\delta} \right\|_2^2$$

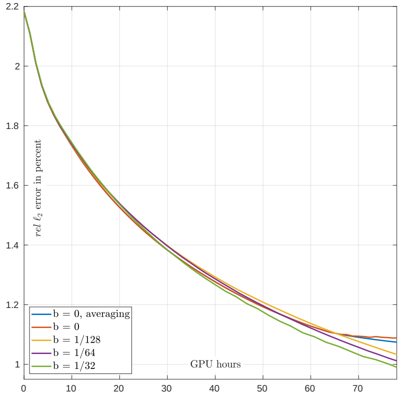
- related to covariance trace estimators
- Rademacher distribution ($w_i = \pm 1$ with equal prob)
- add time-shifting for time-invariant PDEs $\rightarrow$ variance control
- can be turned into scanning strategy

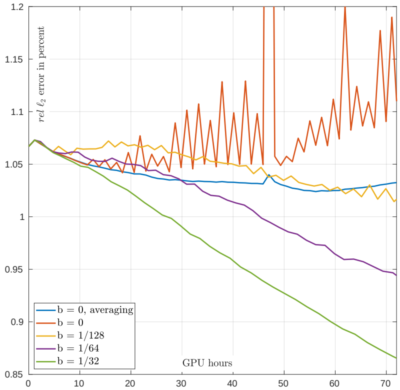
Haber, Chung, Herrmann, 2012. An effective method for parameter estimation with PDE constraints with multiple right-hand sides, *SIAM J. Optim.*

Stochastic Gradient Estimates



Delayed Source Encoding





Time-Reversal Gradient Computations

Avoid storage of forward fields!

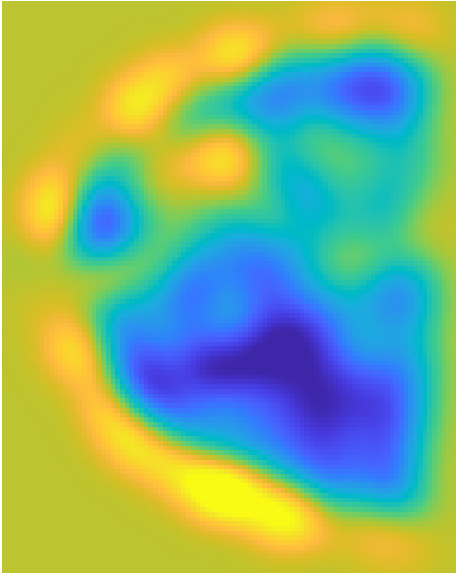
$$
\begin{aligned}
(c(x)^{-2}\partial_t^2 - \Delta)p(x,t) &= s(x,t), \qquad \textit{in } \mathbb{R}^d \times [0,T] \\
\nabla_c \mathcal{D} &= 2\int_0^T \frac{1}{c(x)^3}\left(\frac{\partial^2 p(x,t)}{\partial t^2}\right) q^*(x,t)
\end{aligned}
$$

Idea: ROI Ω, $\mathrm{supp}(s) \in \Omega^c \times [0,T]$. As $p(x,0) = p(x,T) = \partial_t p(x,0) = \partial_t p(x,T) = 0$ in Ω, $p(x,t)$ can be reconstructed from $p(x,t)$ on $\partial\Omega \times [0,T]$ by **time-reversal (TR)**.

- store fwd fields on ROI boundary during forward wave simulation
- interleave backward (adjoint) simulation with TR of boundary data
- 3 instead of 2 wave simulations (unless 2 GPUs used).
- code up efficiently
- multi-layer boundary increases accuarcy for pseudospectral method

Multigrid Schemes

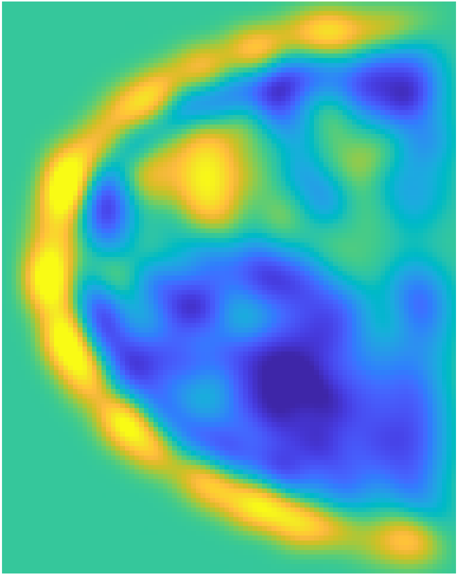
- easy due to regular grids in space and time
- coarsening by 2: (in principle) **speed up of 16**
- most basic multi-grid usage for now: initialization



level 6: upsampled from 5.66mm.

Multigrid Schemes

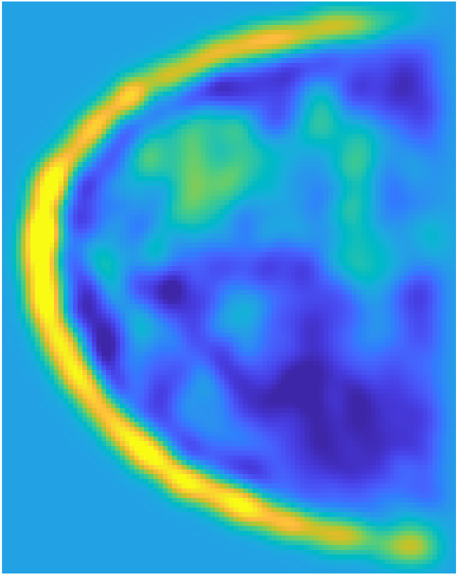
- easy due to regular grids in space and time
- coarsening by 2: (in principle) **speed up of 16**
- most basic multi-grid usage for now: initialization



level 5: upsampled from 4mm.

Multigrid Schemes

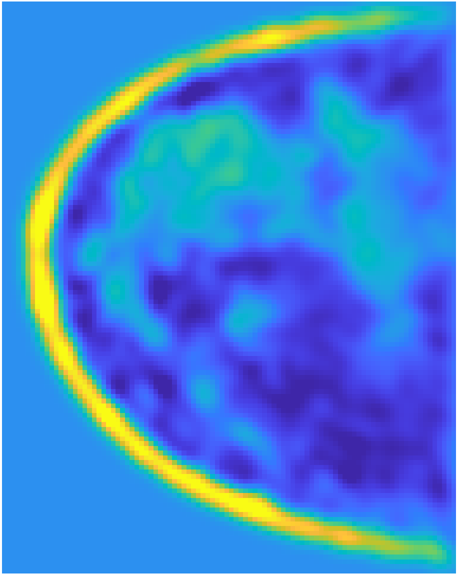
- easy due to regular grids in space and time
- coarsening by 2: (in principle) **speed up of 16**
- most basic multi-grid usage for now: initialization



level 4: upsampled from 2.83mm.

Multigrid Schemes

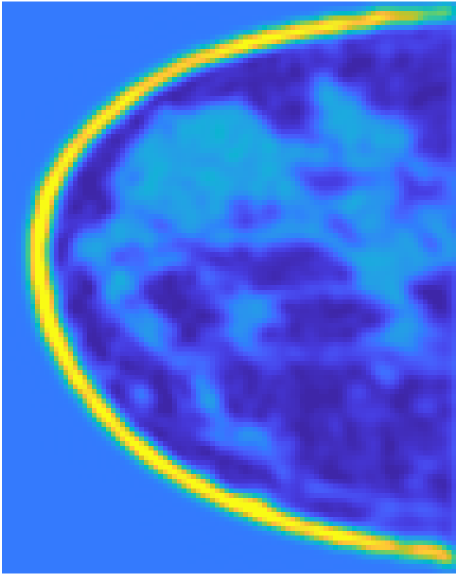
- easy due to regular grids in space and time
- coarsening by 2: (in principle) **speed up of 16**
- most basic multi-grid usage for now: initialization



level 3: upsampled from 2mm.

Multigrid Schemes

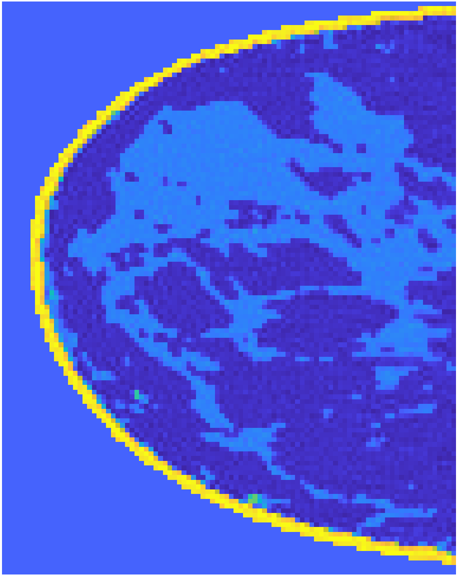
- easy due to regular grids in space and time
- coarsening by 2: (in principle) **speed up of 16**
- most basic multi-grid usage for now: initialization



level 2: upsampled from 1.41mm.

Multigrid Schemes

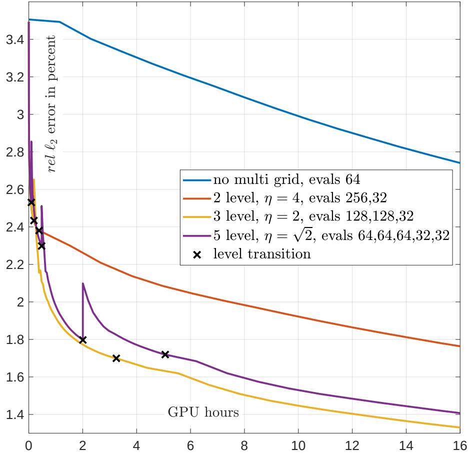
- easy due to regular grids in space and time
- coarsening by 2: (in principle) **speed up of 16**
- most basic multi-grid usage for now: initialization



level 1: resolution 1mm

Multigrid Schemes

- easy due to regular grids in space and time
- coarsening by 2: **speed up of 16** (in principle)
- most basic multi-grid usage for now: initialization



Utilizing Multiple GPUs

- average independent gradient estimates to reduce variance
- not be the best way to use multiple GPUs