TEXAS
The University of Texas at Austin

OPTIMUS
BEYOND FORWARD SIMULATION

ODEN INSTITUTE FOR COMPUTATIONAL ENGINEERING & SCIENCES

Can We Predict Tsunamis in Real Time?

Omar Ghattas
Oden Institute for Computational Engineering & Sciences
Dept of Mechanical Engineering
Dept of Earth & Planetary Sciences (by courtesy)
The University of Texas at Austin

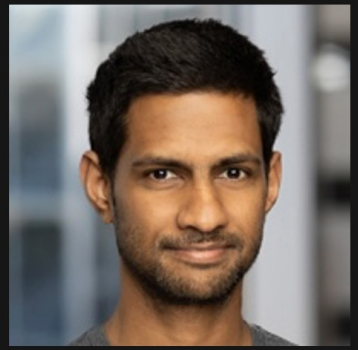
Stefan Henneking

Sreeram Venkat

2025 Gordon Bell Prize team

Stefan Henneking (UTA)

Sreeram Venkat (UTA)



Milinda Fernando (UTA)

Omar Ghattas (UTA)

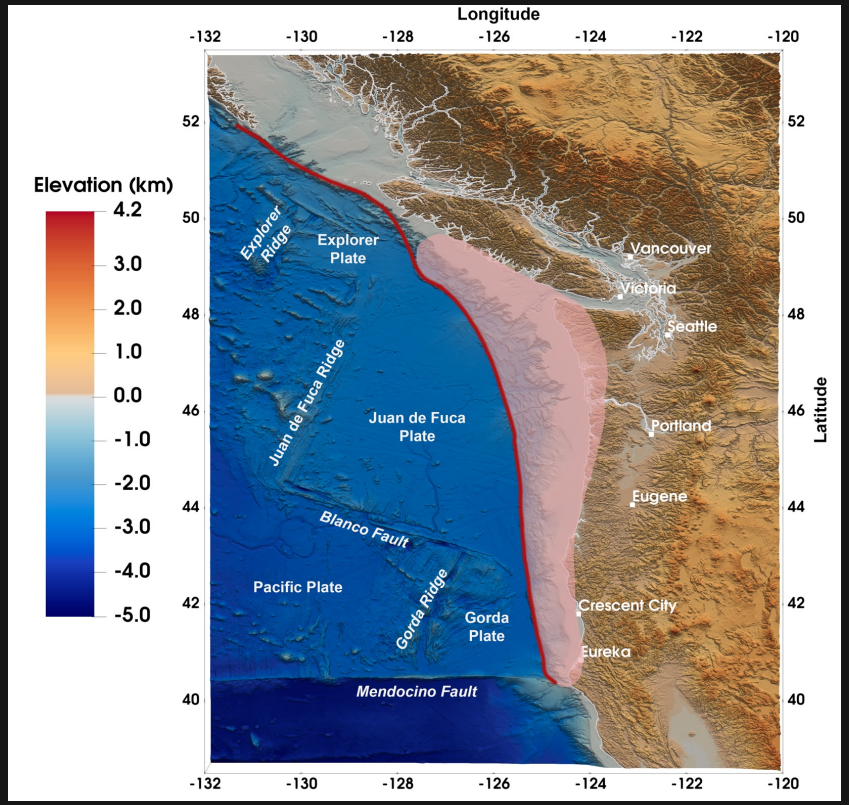
Alice Gabriel (UCSD)

Veselin Dobrev (LLNL)

John Camier (LLNL)

Tzanio Kolev (LLNL)

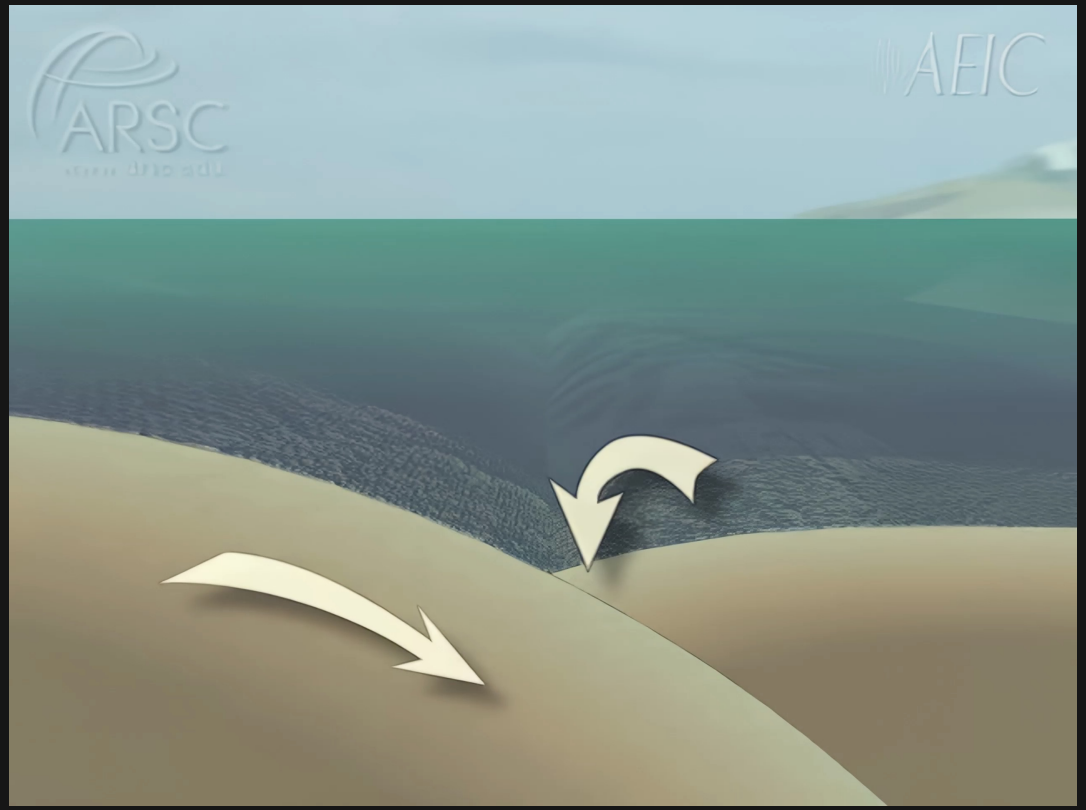
Cascadia subduction zone



- Last major earthquake and orphan tsunami: January 26, 1700
- Paleoseismic evidence indicates a 250–500 years recurrence interval

Image credit: Rob DeGraff/Flickr, 2008

Subduction earthquake tsunami generation



Video credit: Alaska Earthquake Center

Cascadia subduction zone tsunami (USGS model)

Image credit:
Washington State Dept. of Natural Resources

Proposed sensor network

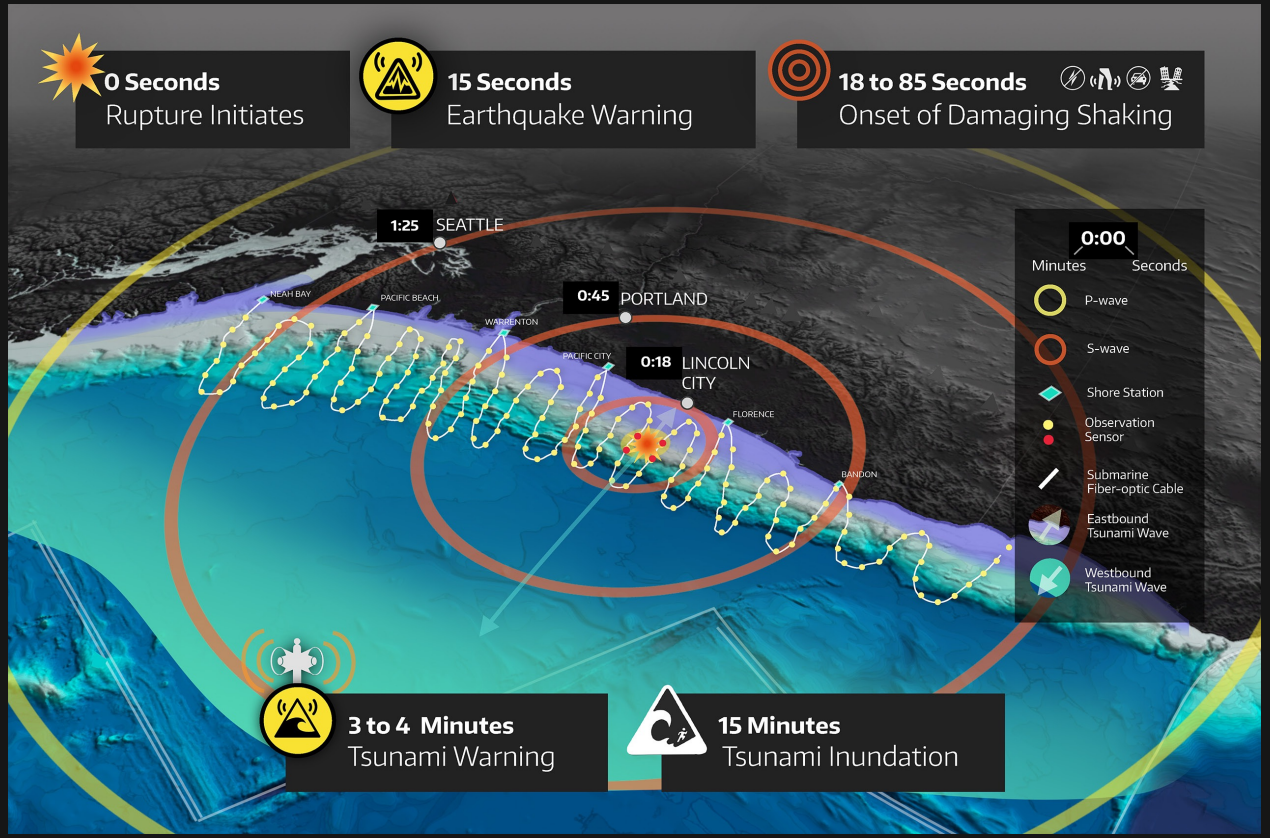


Image credit: D. Schmidt et al., 2019

Digital twin for tsunami forecasting

Current forecasting models rely on shallow water equations for tsunami propagation

- Efficient in the far-field; does not make use of near-field pressure transients from hydroacoustic waves

Our digital twin approach

- Use near-field pressure observations to **infer** the seafloor motion and **forward predict** the tsunami propagation
- Employ high-fidelity, **full-physics** model (coupled 3D acoustic–gravity wave)
- Quantify **uncertainty** via Bayesian inference
- Solve inverse problem in **real time** (order of seconds)

Acoustic–gravity wave propagation forward model

- Acoustic wave equation in mixed velocity-pressure form

$$\rho \frac{\partial \vec{u}}{\partial t} + \nabla p = 0, \qquad \frac{1}{K} \frac{\partial p}{\partial t} + \nabla \cdot \vec{u} = 0, \qquad \text{in } \Omega \times (0, \mathrm{T})$$

- Boundary conditions
  - Sea surface (coupling with surface gravity wave)

$$p = \rho g \eta, \qquad \partial \eta / \partial t = \vec{u} \cdot \vec{n}, \qquad \text{on } \Gamma_{\text{surface}} \times (0, \mathrm{T})$$

  - Seafloor velocity (boundary source)

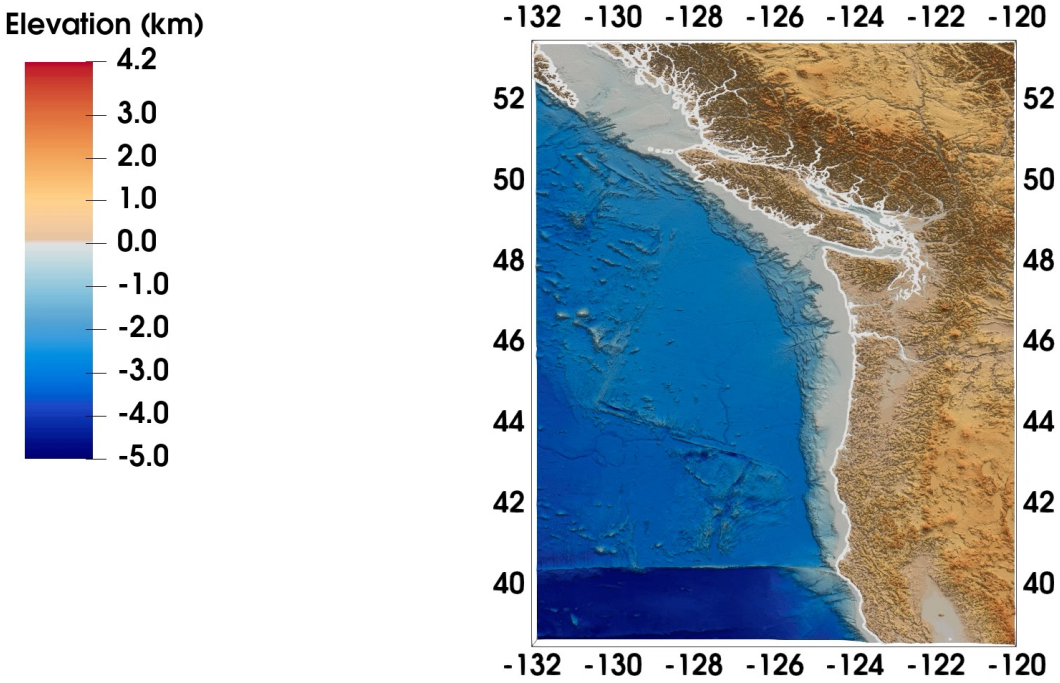
$$\vec{u} \cdot \vec{n} = m \text{ on } \Gamma_{\text{bottom}} \times (0, \mathrm{T})$$

  - Lateral absorbing boundary (outgoing waves)

$$\vec{u} \cdot \vec{n} = p / \rho c, \qquad \text{on } \Gamma_{\text{absorb}} \times (0, \mathrm{T})$$

- Homogeneous initial conditions
- Implemented with MFEM using high-order finite elements and RK4 time-stepping

Bathymetry-adapted meshing



The inverse problem

Given:

- Observations d^{obs} of acoustic pressure p at sensor locations
- Linear parameter-to-observable map $F: m(\vec{x}, t) \mapsto d(\vec{x}, t)$ governed by forward acoustic–gravity wave propagation model

Infer:

- Seafloor normal velocity parameter field m

This inverse problem is ill-posed

- Observations are sparse and noisy
- Parameter field is infinite-dimensional in space-time
- Many different parameter fields may be consistent with the data & model

Bayesian framework for inverse problems

- The uncertain parameter is treated as a random field, and the solution of the inverse problem becomes a probability density, the posterior distribution
- The posterior π_{post} characterizes the probability that any parameter field is consistent with the data, the model, and any prior knowledge

$$\pi_{\text{post}}(m|d^{\text{obs}}) \propto \pi_{\text{like}}(d^{\text{obs}}|m)\,\pi_{\text{prior}}(m)$$

- With Gaussian prior and noise:

$$\pi_{\text{post}}(m|d^{\text{obs}}) \propto \exp\left(-\frac{1}{2}\left\|Fm - d^{\text{obs}}\right\|^2_{\Gamma^{-1}_{\text{noise}}} - \frac{1}{2}\left\|m - m_{\text{prior}}\right\|^2_{\Gamma^{-1}_{\text{prior}}}\right)$$

Solving the Bayesian inverse problem

Gaussian posterior

$$\pi_{\text{post}}(m|d^{\text{obs}}) \propto \mathcal{N}\left(m_{\text{map}}, \Gamma_{\text{post}}\right)$$

where

$$\Gamma_{\text{post}} = H^{-1} = \left(F^{*}\Gamma_{\text{noise}}^{-1}F + \Gamma_{\text{prior}}^{-1}\right)^{-1}$$

$$Hm_{\text{map}} = F^{*}\Gamma_{\text{noise}}^{-1}d^{\text{obs}} + \Gamma_{\text{prior}}^{-1}m_{\text{prior}}$$

To compute the mean m_{map}, solve the linear system $Hm_{\text{map}} = f(d^{\text{obs}})$

To compute parameter uncertainties, compute covariance Γ_{post}

Conventional methods

- Need to solve large-scale linear system $\mathrm{Hm}_{\mathrm{map}} = \mathrm{f}(\mathrm{d}^{\mathrm{obs}})$ in real time!
- Cascadia digital twin: $\dim(\mathrm{m}) \sim 10^9$ parameters, $\dim(\mathrm{d}^{\mathrm{obs}}) \sim 2.5 \cdot 10^5$ data

Conventional methods are intractable!

- Constructing the Hessian $\mathrm{H} = \Gamma_{\mathrm{post}}^{-1}$ requires 1 adjoint PDE solve per column of F: $\sim 2.5 \cdot 10^5$ adjoint PDE solutions → ~25 years on 512 A100 GPUs
- Storing the Hessian in dp requires ~4 exabytes

Traditional fix of surrogates (e.g., ROMs, NNs, ...) is not feasible for high-frequency wave propagation

- Very large parameter dimension
- Kolmogorov *N*-width barrier

Instead, exploit time-shift invariance of p2o map F

Autonomous dynamical system

- Evolution of the physics model does not depend explicitly on the independent variable (e.g. time)
- Here: the mapping $m(\vec{x}, t+\tau) \mapsto d(\vec{x}, t+\tau)$ is the same as $m(\vec{x}, t) \mapsto d(\vec{x}, t)$
- The discrete p2o map F is shift-invariant with respect to time-stepping

Block Toeplitz structure, FFT-based solver

F is block-Toeplitz ($N_t \times N_t$ blocks: $\mathrm{F}_{ij} \in \mathbb{R}^{N_d N_m}$, where $N_d \ll N_m$)

$$\begin{bmatrix} \mathrm{d}_1 \\ \mathrm{d}_2 \\ \mathrm{d}_3 \\ \vdots \\ \mathrm{d}_{N_t} \end{bmatrix} = \begin{bmatrix} \mathrm{F}_{11} & 0 & 0 & \cdots & 0 \\ \mathrm{F}_{21} & \mathrm{F}_{11} & 0 & \cdots & 0 \\ \mathrm{F}_{31} & \mathrm{F}_{21} & \mathrm{F}_{11} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \mathrm{F}_{N_t,1} & \mathrm{F}_{N_t-1,1} & \cdots & \mathrm{F}_{21} & \mathrm{F}_{11} \end{bmatrix} \begin{bmatrix} \mathrm{m}_1 \\ \mathrm{m}_2 \\ \mathrm{m}_3 \\ \vdots \\ \mathrm{m}_{N_t} \end{bmatrix}, \quad \begin{aligned} \mathrm{d}_i &= d(\vec{x}, t_i) \in \mathbb{R}^{N_d} \\ \mathrm{m}_i &= m(\vec{x}, t_i) \in \mathbb{R}^{N_m} \end{aligned}$$

- Precomputation of F by only N_d (number of sensors) adjoint PDE solutions
- Compact storage of F: $\mathcal{O}(N_m N_d N_t)$
- PDE-free application of F and F^* in Fourier space: $\mathcal{O}(N_m N_d N_t \log N_t)$
  Block Toeplitz matrix is block-diagonalized by FFT
- Algorithms extend to parameter-to-QoI (quantity of interest) map F_q

Offline–online decomposition

Phase 1 (offline): Construct F and F_q maps from adjoint wave props (MFEM)

Phase 2 (offline): Compute compact representation of the posterior covariance

Apply Sherman–Morrison–Woodbury formula to reformulate Hessian in data space

$$\Gamma_{\text{post}} = \left(F^*\Gamma_{\text{noise}}^{-1}F + \Gamma_{\text{prior}}^{-1}\right)^{-1} = \Gamma_{\text{prior}}(\mathrm{I} - F^*(\underbrace{\Gamma_{\text{noise}} + F\Gamma_{\text{prior}}F^*}_{K})^{-1}F\Gamma_{\text{prior}})$$

Exploit that $G^* \coloneqq \Gamma_{\text{prior}}F^*$ is block Toeplitz,
where Γ_{prior} is a Matern prior, equivalent to a block inverse screened Poisson operator

Phase 3 (offline): Compute uncertainties in QoI

$$q \sim \mathcal{N}\left(F_q m_{\text{map}}, F_q\Gamma_{\text{post}}F_q^*\right), \qquad F_q\Gamma_{\text{post}}F_{\text{q}}^* = F_q(I - G^*K^{-1}F)G_q^*$$

Phase 4 (online): Compute parameter and QoI means in real time

Cascadia margin-wide rupture inversion

Inverse problem configuration:

- Spatial resolution: 300 m
- Simulation time: T = 420 s
- Temporal dim.: N_t = 420 (1 Hz frequency)
- Spatial parameter dim.: $N_m \approx$ 2.42M
- Number of sensors: N_d = 600
- Tsunami forecast (QoI) locations: N_q = 21
- Total parameter dim.: $N_m N_t \approx$ 1.01B
- Total data dim.: $N_d N_t$ = 252,000
- Total QoI dim.: $N_q N_t$ = 8,820
- Total state dim.: $\approx$ 3.74B spatial DOF; 336,000 RK4 timesteps

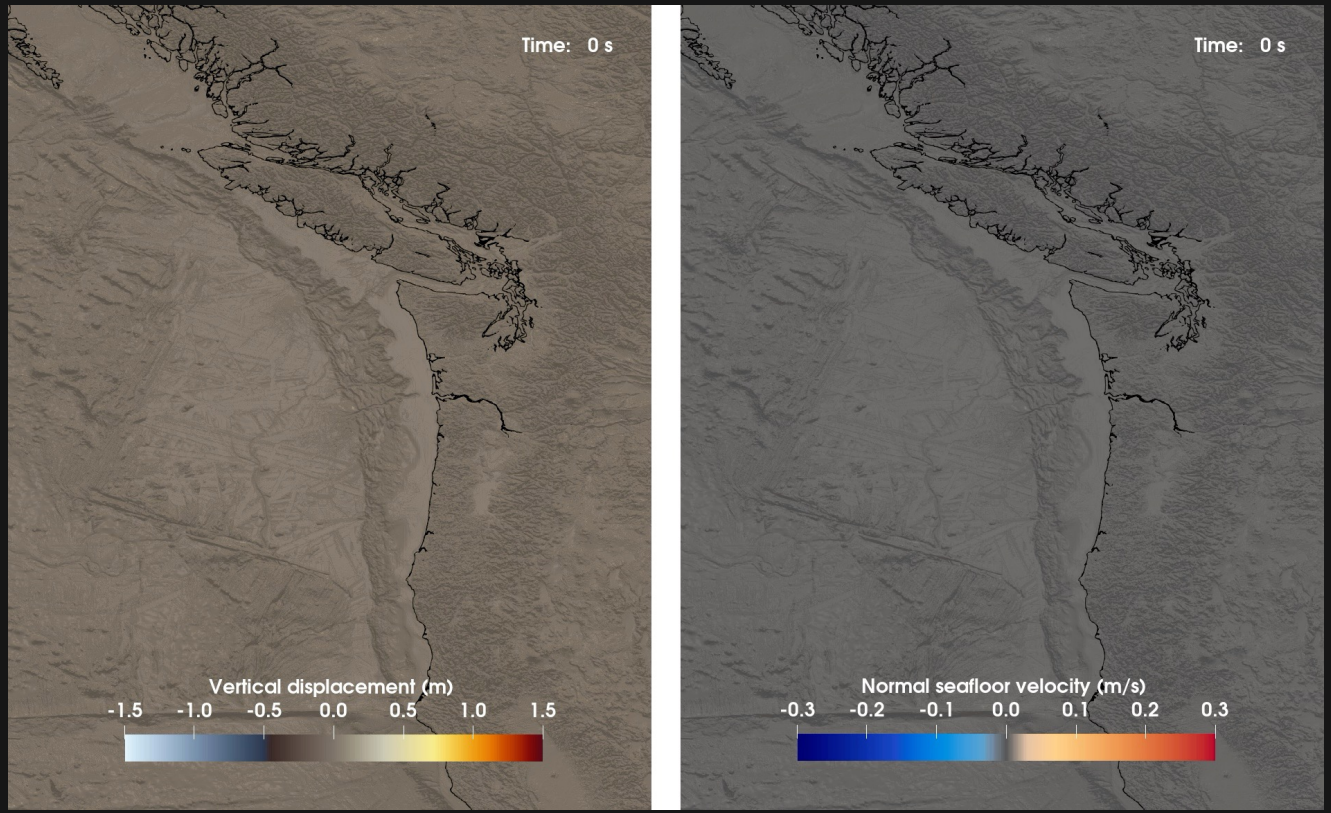
Cascadia margin-wide rupture scenario

Left:

Vertical seafloor uplift

Right:

Seafloor normal velocity



Glehman et al., 2025

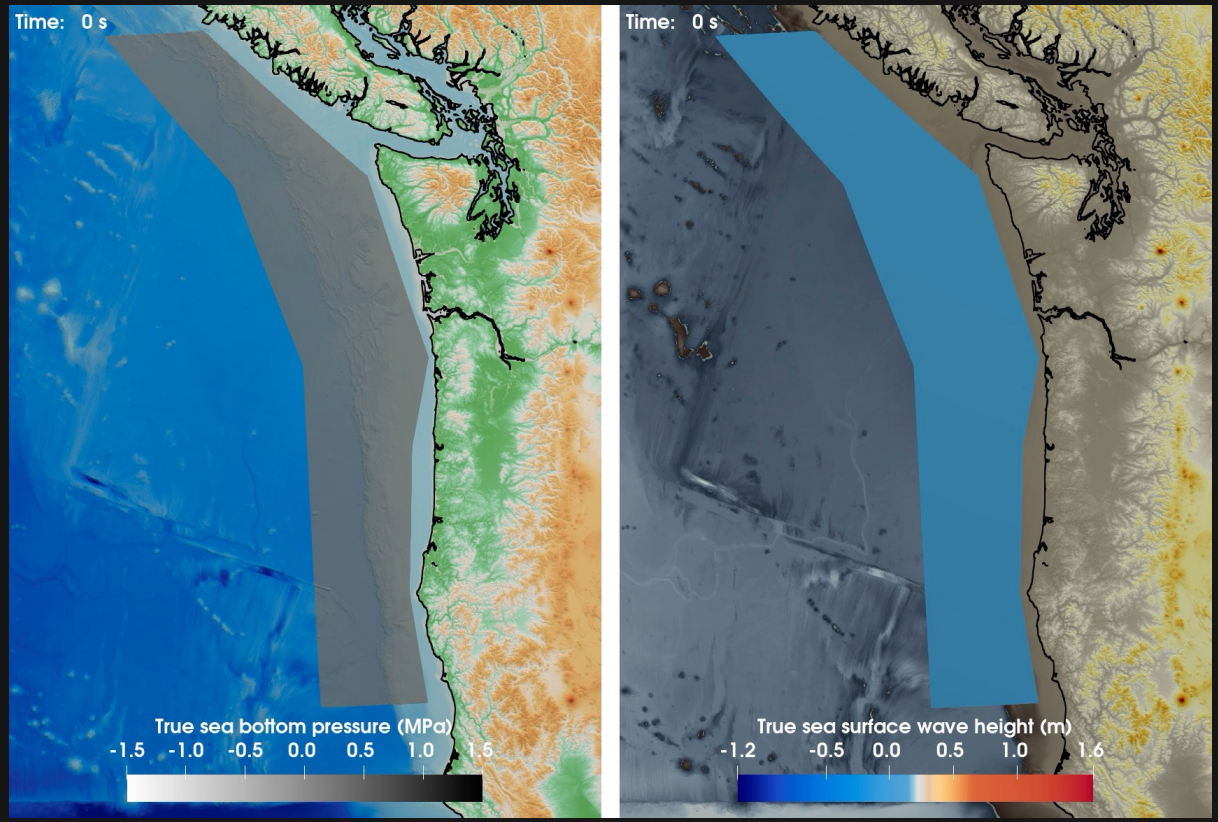
Acoustic–gravity simulation

Left:

Seafloor pressure change

Right:

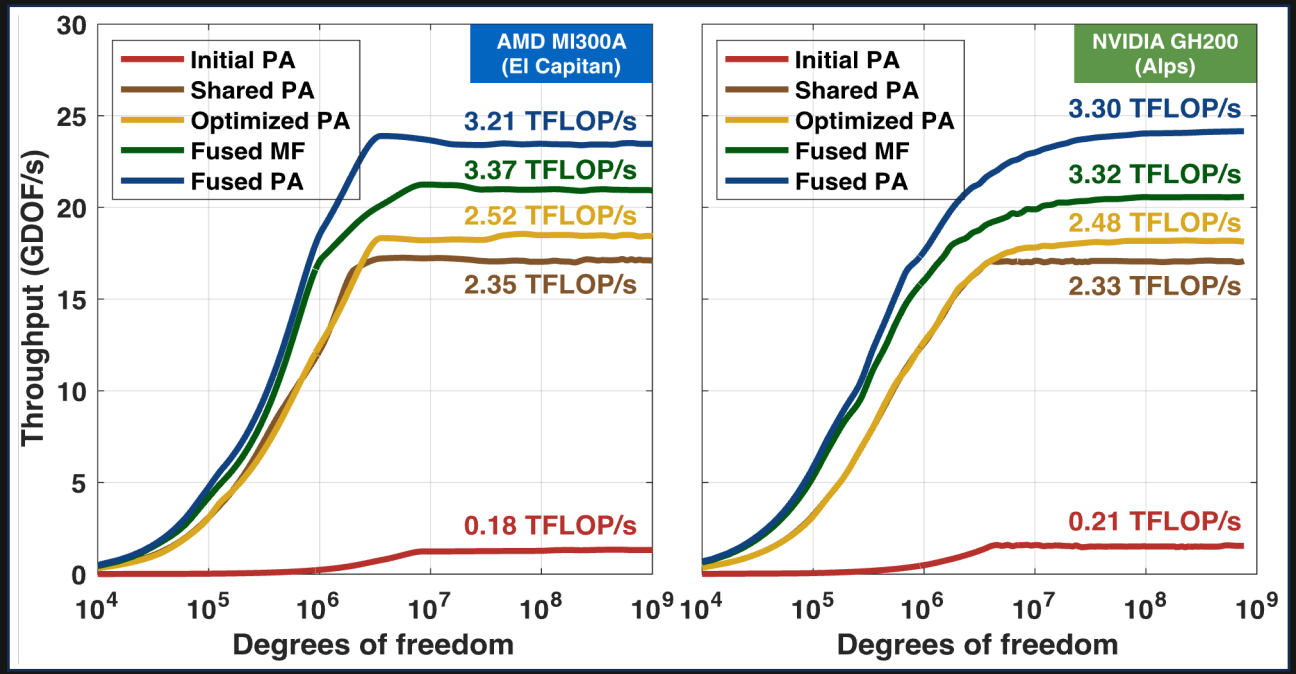
Sea surface wave height



GPU Kernel optimizations

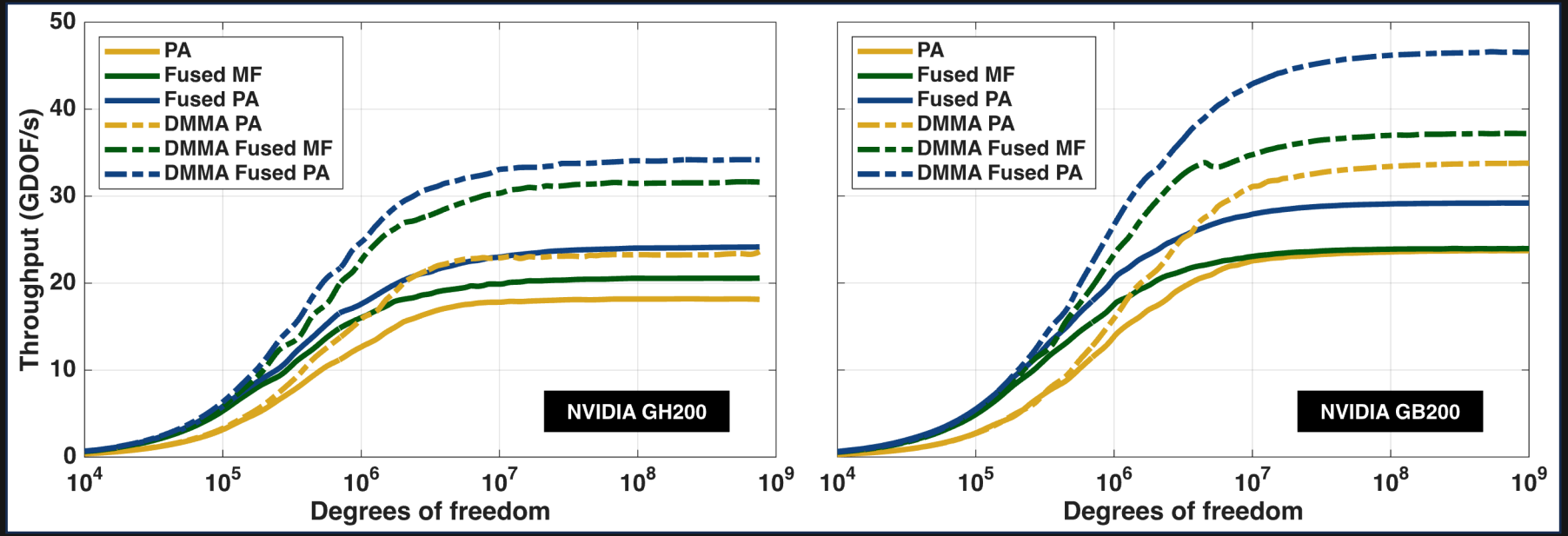
GPU-enabled finite element kernels in MFEM:

- **Partial assembly (PA)**
  Stores data at quadrature points
- **Shared PA**
  Exploits GPU shared-memory
- **Optimized PA**
  Uses explicit launch bounds
- **Fused PA**
  Fuses operators into single kernel
- **Fused matrix-free (MF)**
  Computes entirely matrix-free



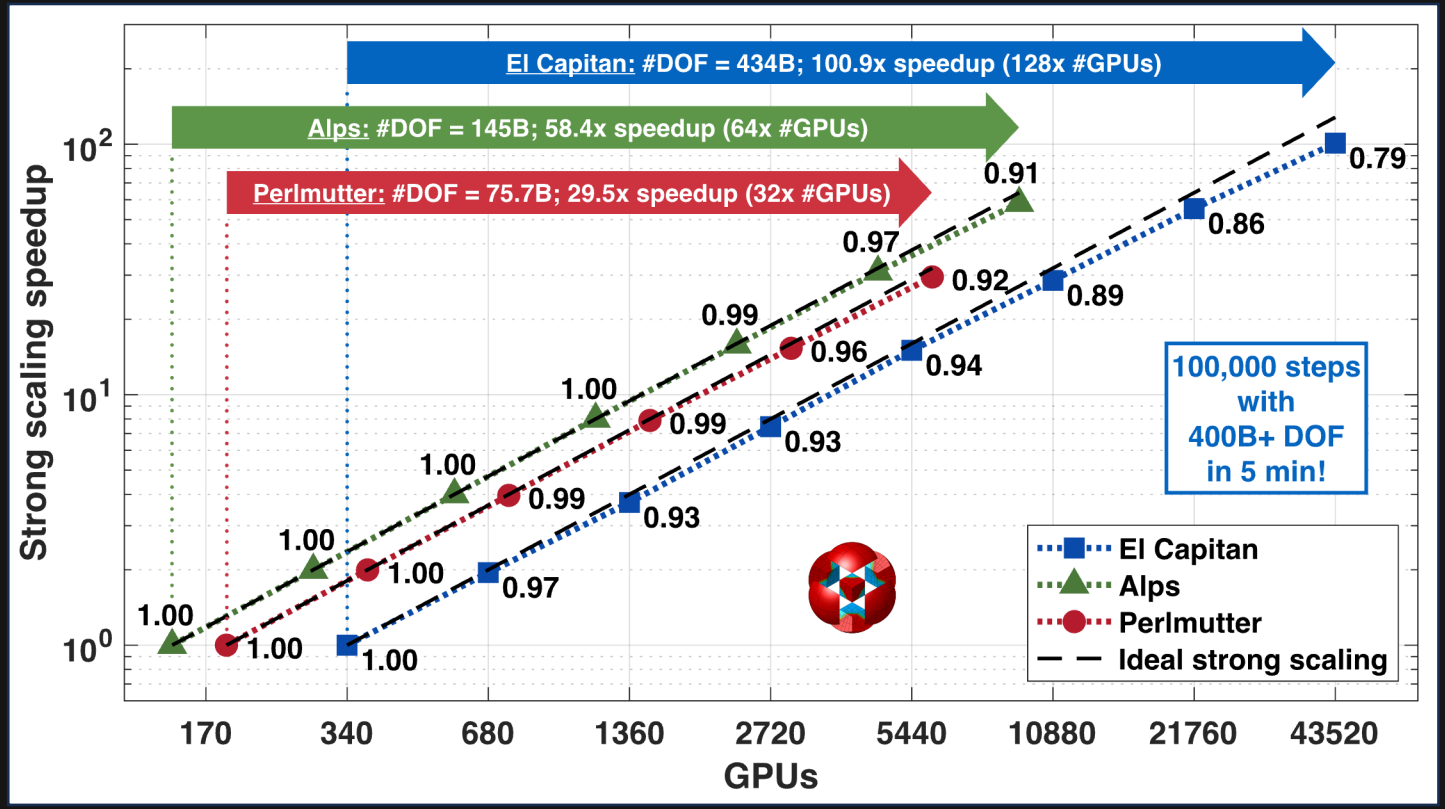
Kernel optimizations with FP64 Tensor Cores

Tensor-core-enabled DMMA [double precision matrix-multiply-accumulate] finite elements with MFEM on Grace–Hopper and Grace–Blackwell Superchips

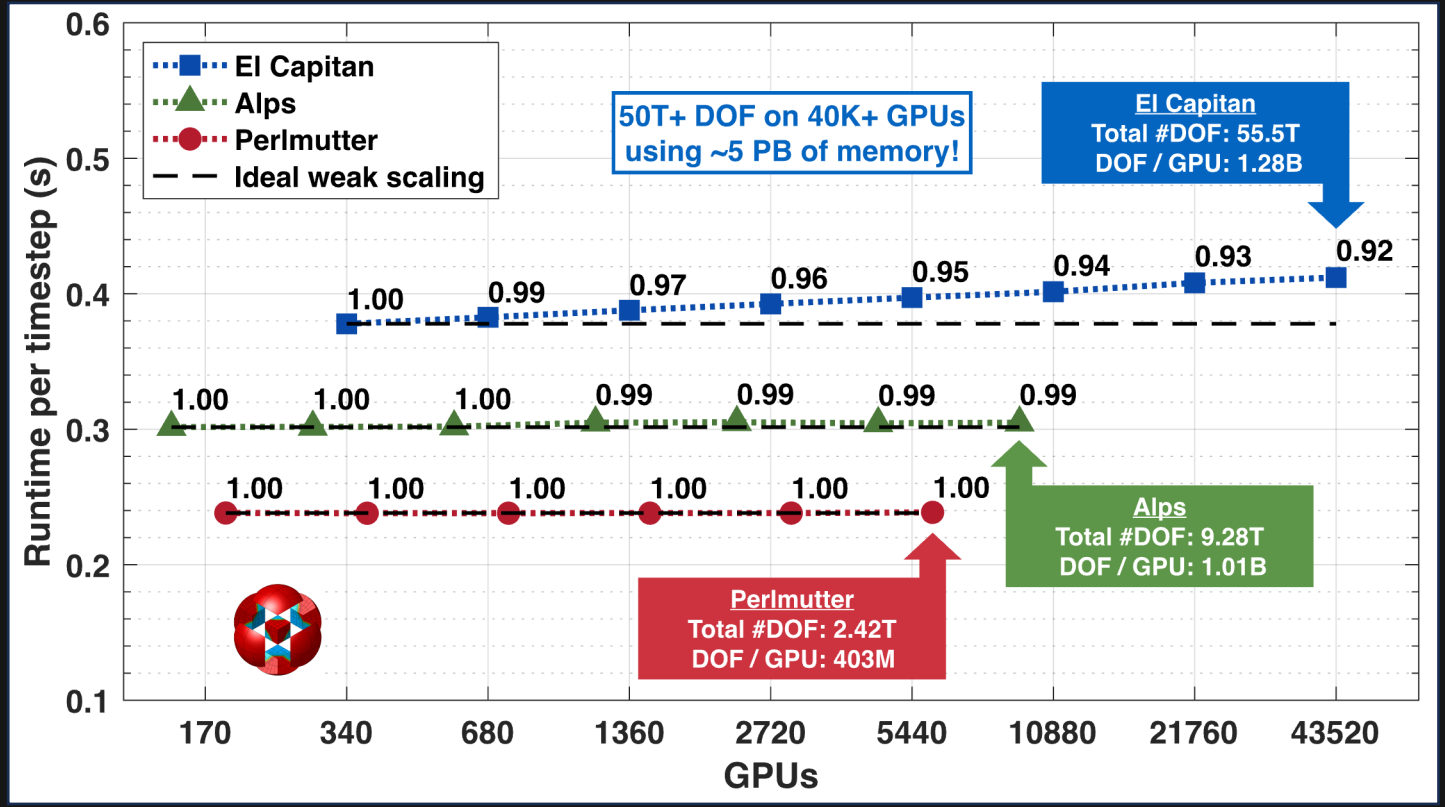


In collaboration with Jiqun Tu (NVIDIA Corporation) & Ian Karlin (Queen's University)

Strong scaling of finite element solver



Weak scaling of finite element solver



Compute times for each phase

Phase	Task	#GPUs	Compute time
1	form $F : m \mapsto d$	512	600 x 52 m ~ 520 hours
	form $F_q : m \mapsto q$	512	21 x 52 m ~ 18 hours
2	form $G^* = \Gamma_{\text{prior}} F^*$	16	600 x 4.5 s ~ 45 minutes
	form $G_q^* = \Gamma_{\text{prior}} F_q^*$	16	21 x 4.5 s ~ 1.5 minutes
	form $K = \Gamma_{\text{noise}} + F G^*$	512	252,000 x 24 ms ~ 100 minutes
	factorize K	25	22 seconds
3	compute $\Gamma_{\text{post}(q)}$	512	8,820 x 150 ms ~ 25 minutes
	compute $Q : d \mapsto q$	512	8,820 x 150 ms ~ 25 minutes
4	**infer parameters m_{map}**	512	**< 0.2 seconds**
	predict QoI q_{map}	1	**< 1 millisecond**

Time-to-solution for the online computation is < 0.2 seconds!

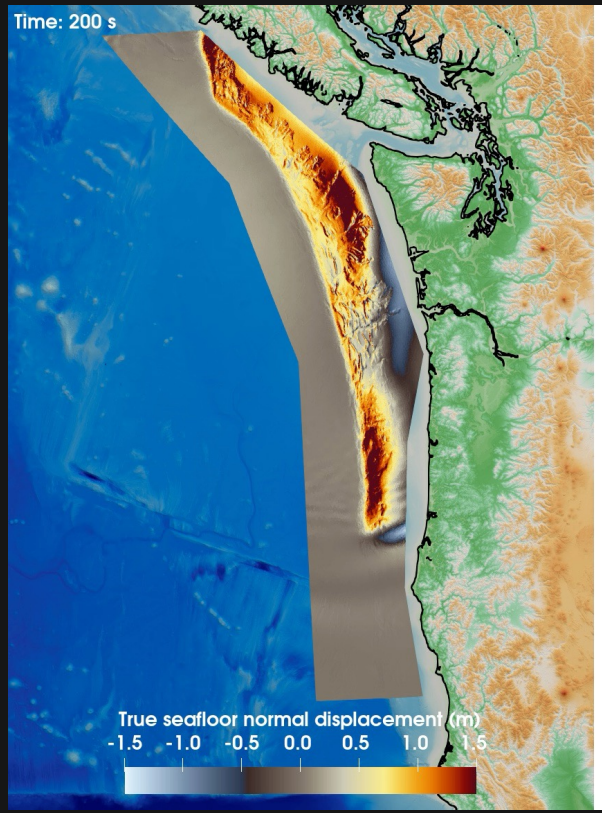
True vs. inferred spatiotemporal seafloor motion

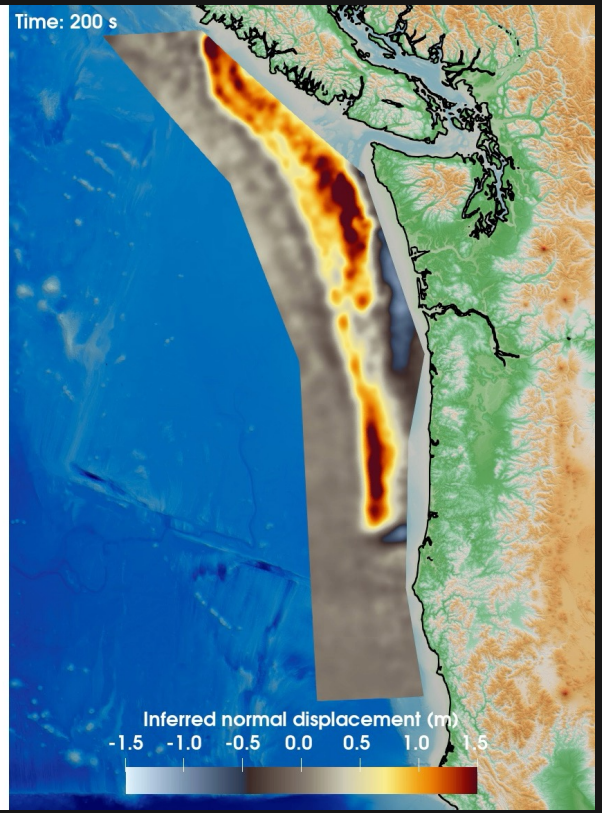
Left:

"True" seafloor normal uplift

Middle:

Mean of inferred seafloor normal uplift

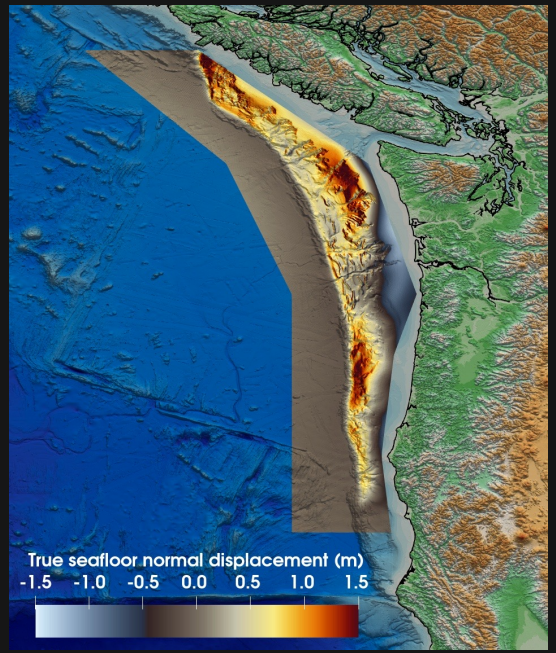




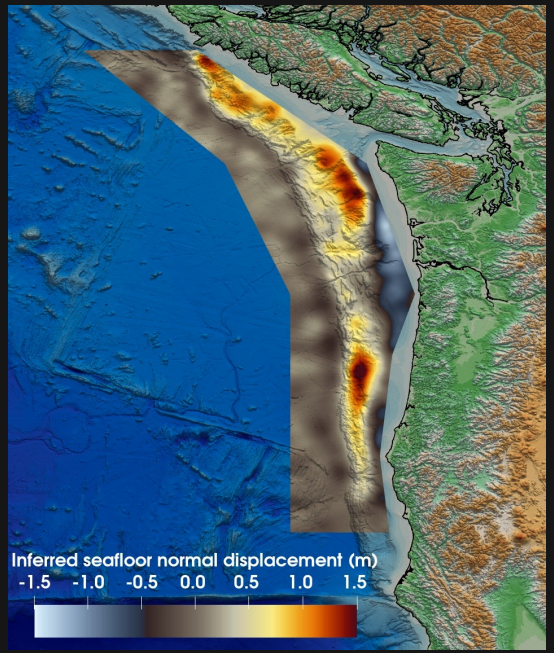
600 hypothesized sensor locations

Inference of seafloor motion

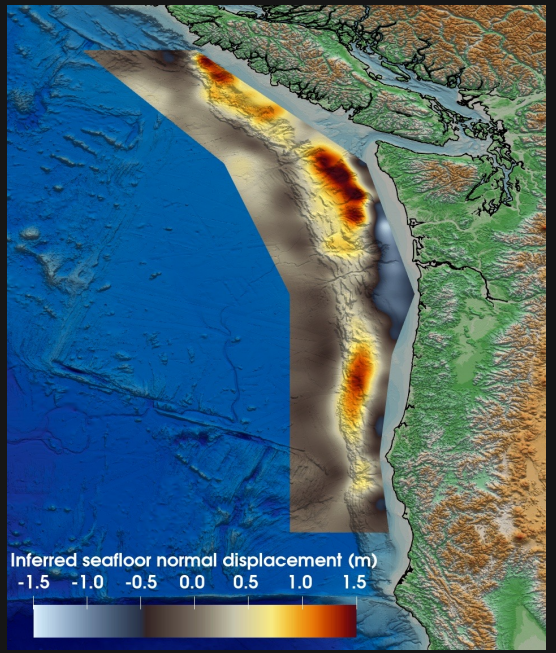
Ground truth



600 sensors



175 sensors

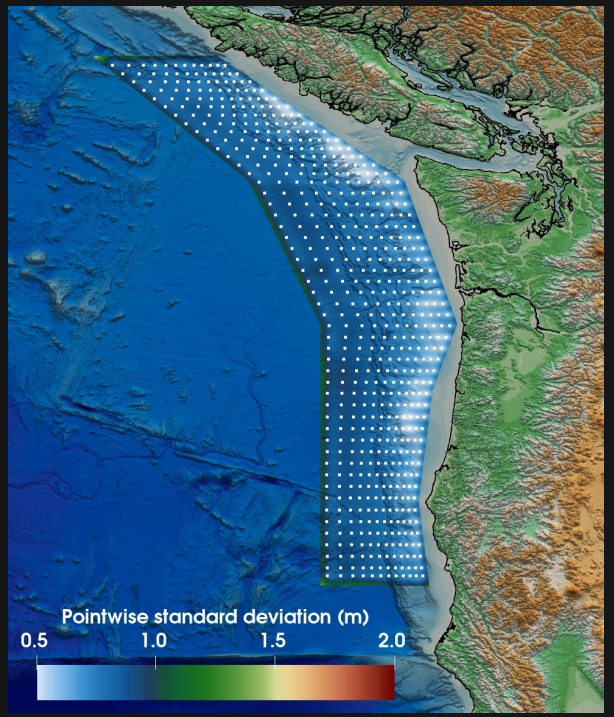


Relative errors of inferred seafloor displacement: 29.6% (600 sensors); 36.1% (175 sensors)

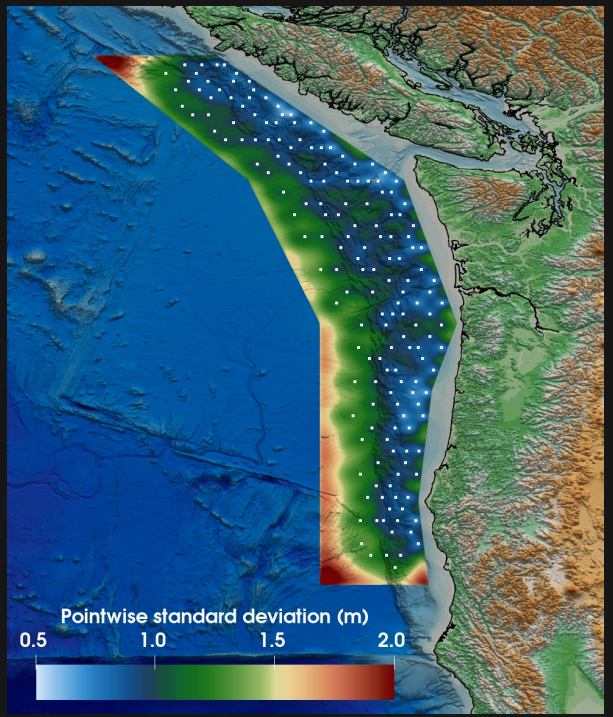
Uncertainties of inverse solution

Uncertainties depicted as pointwise standard deviations of the total seafloor normal uplift

600 sensors

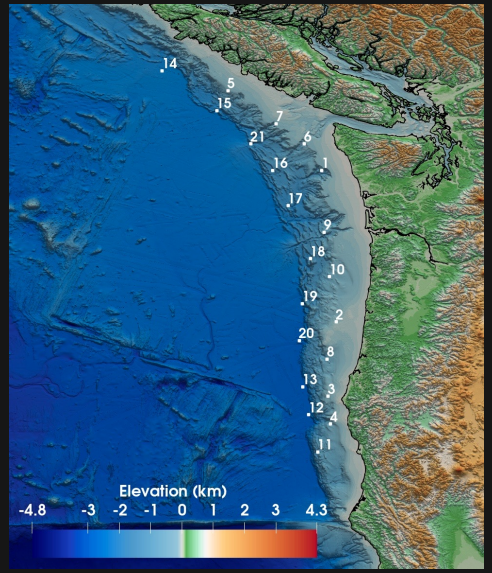


175 sensors

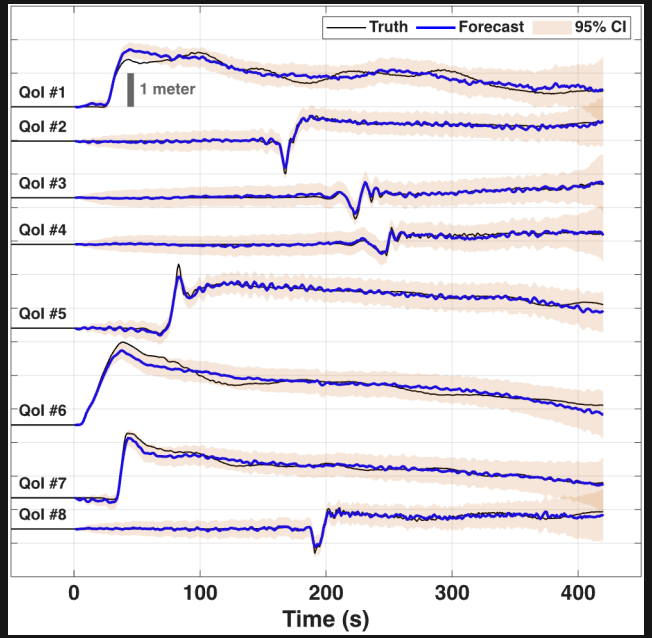


Real-time tsunami wave height forecasts

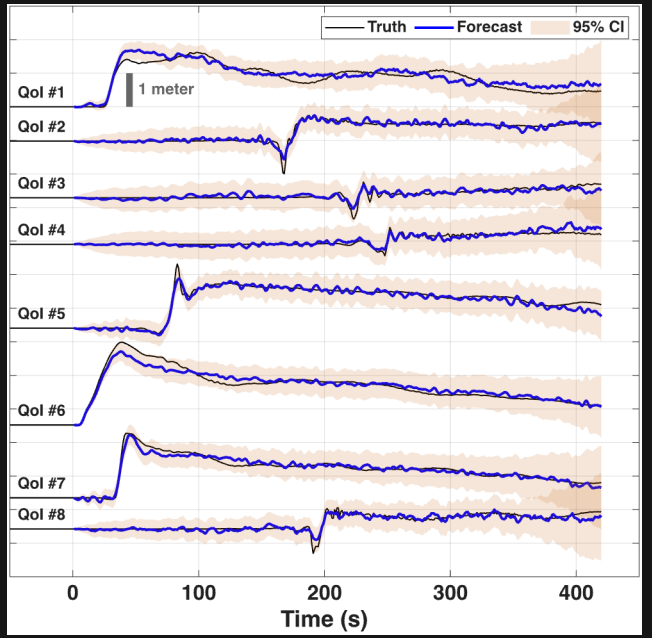
Forecast locations



600 sensors



175 sensors



Relative errors of tsunami wave height forecasts: 18.6% (600 sensors); 22.1% (175 sensors)

Sensor placement for Cascadia

- Cascadia is currently lacking sufficient offshore instrumentation

- Deploying and operating sensor equipment is expensive, so we must answer the following questions:
  - *How many sensors are required for accurate forecasts?*
  - *What is the optimal sensor placement?*

→ Bayesian optimal experimental design (Bayesian OED)

Optimal experimental design (OED)

Given:

- Set of candidate sensor locations (e.g., $N_d = 600$)
- Fixed budget of B sensors

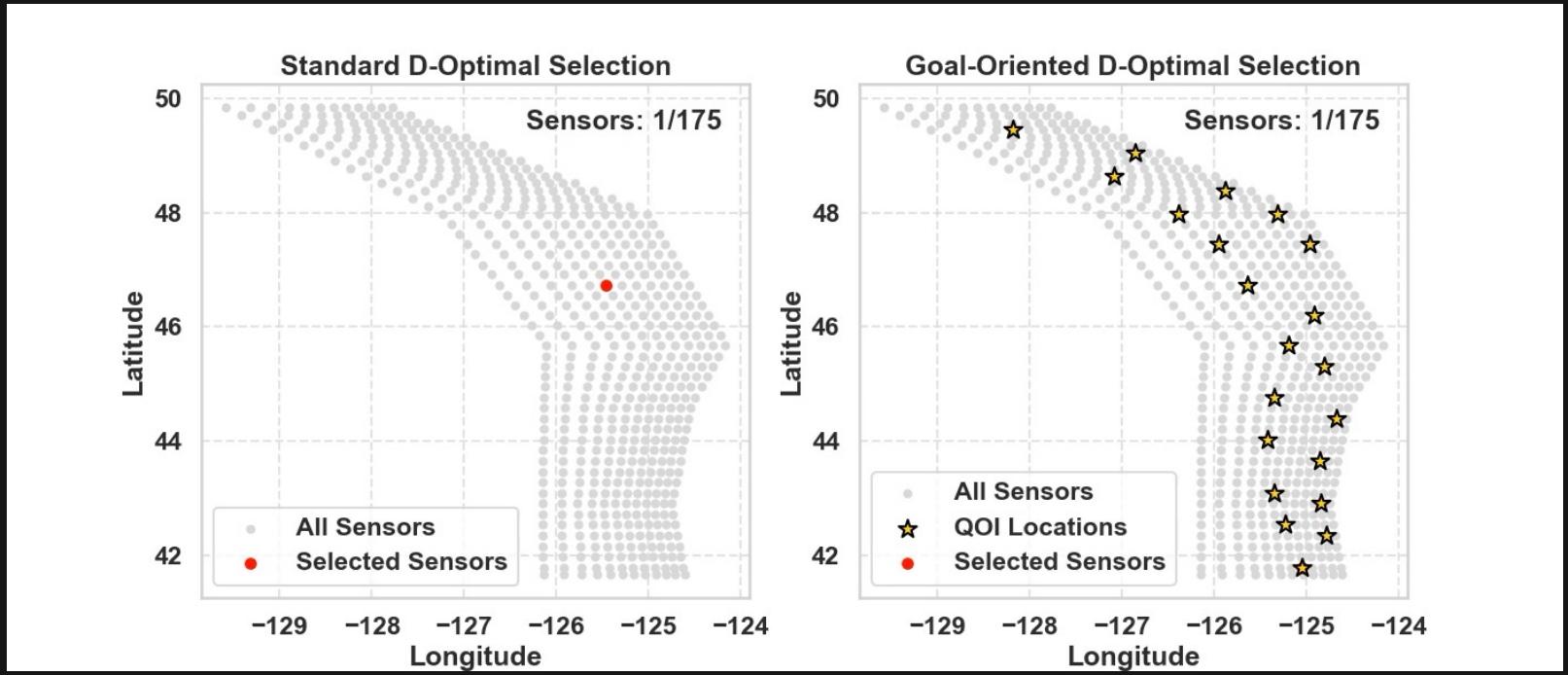
Task:

- Select optimal sensor subset S of size $|S| \leq B$ that provides maximum information about
  - uncertain parameters m (seafloor motion) → *standard OED*
  - quantities of interest q (tsunami forecasts) → *goal-oriented OED*

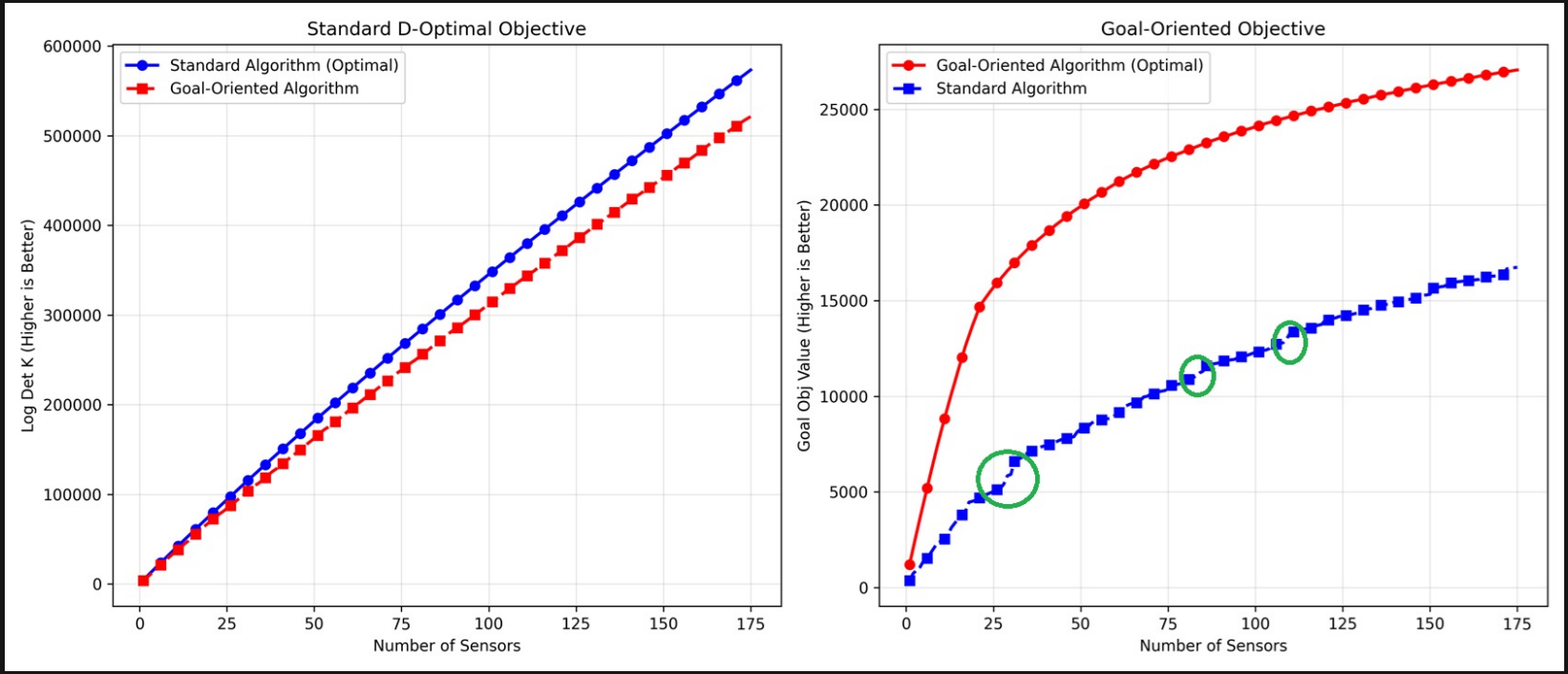
Bayesian approach:

- Maximize the expected information gain from the prior to the posterior
- For linear inverse problems, this is equivalent to the D-optimal design criterion (minimizes log-determinant of the posterior covariance)
- Use greedy algorithm to solve combinatorial optimal sensor placement problem
  - Submodular objective function implies that greedy is within 63% of optimal

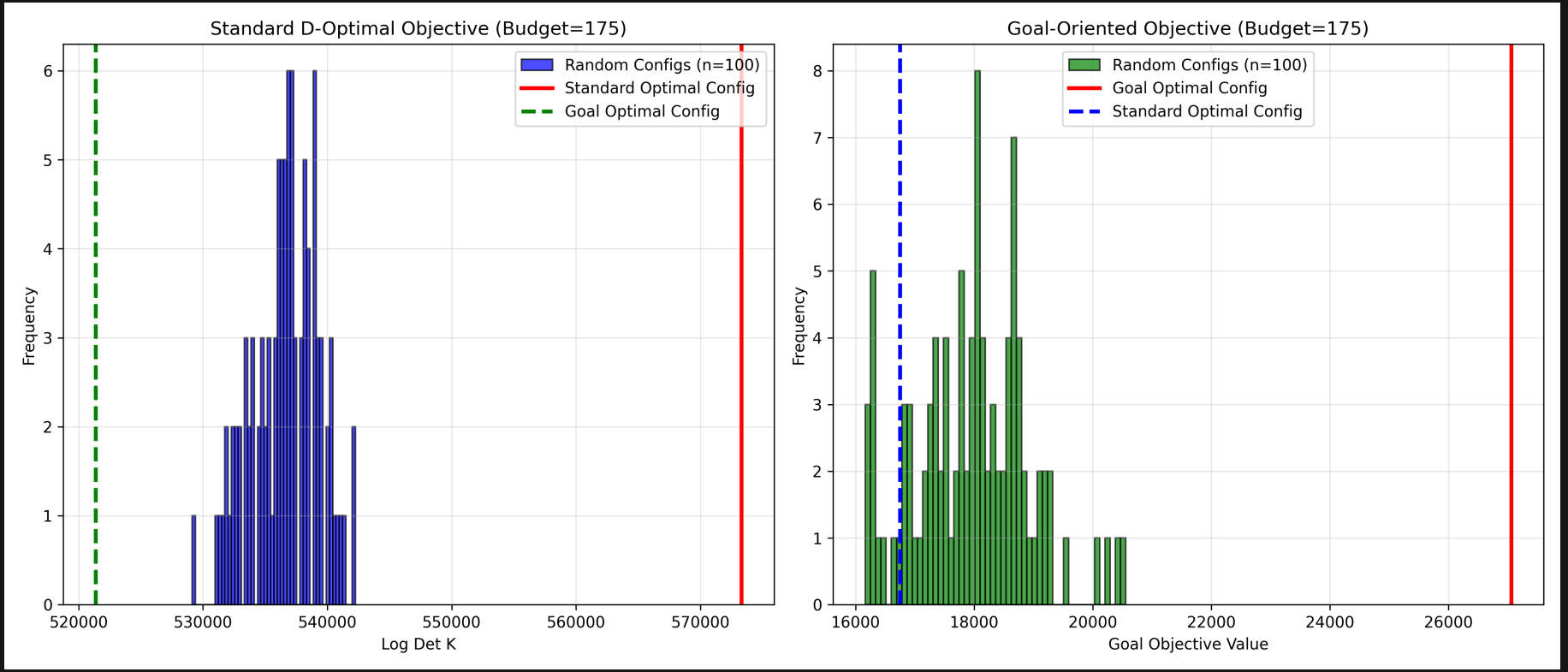
Greedy algorithm-based optimal sensor locations: Standard vs. goal-oriented EIG, 175 out of 600 sensors



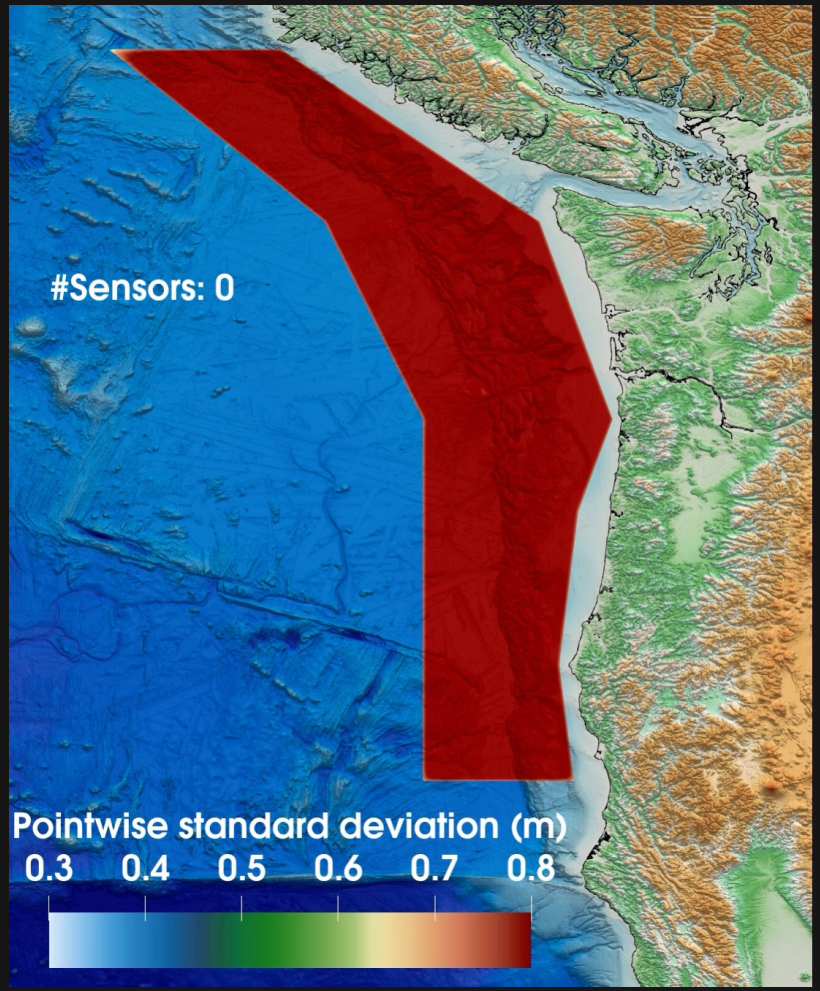
EIG objective function: Standard vs goal-oriented



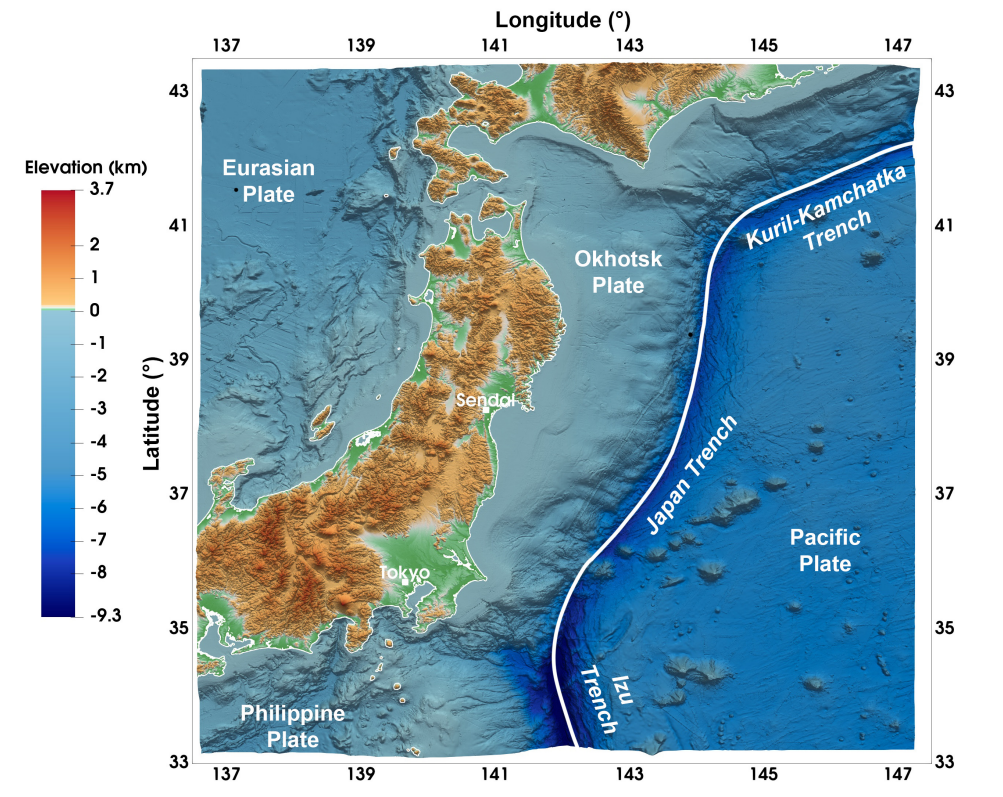
Comparison of greedy solution vs random designs



Animation of posterior pointwise standard deviation with OED-driven sequential addition of sensors



Digital twin of the Japan Trench



Last major rupture:

2011 magnitude 9.0 Tōhoku, Japan, megathrust earthquake

- > 20,000 deaths
- > $300B damages
- Fukushima nuclear reactor meltdown

In collaboration with Jeremy Wong & Alice-Agnes Gabriel (UC San Diego)

Japan's cabled ocean bottom sensor networks

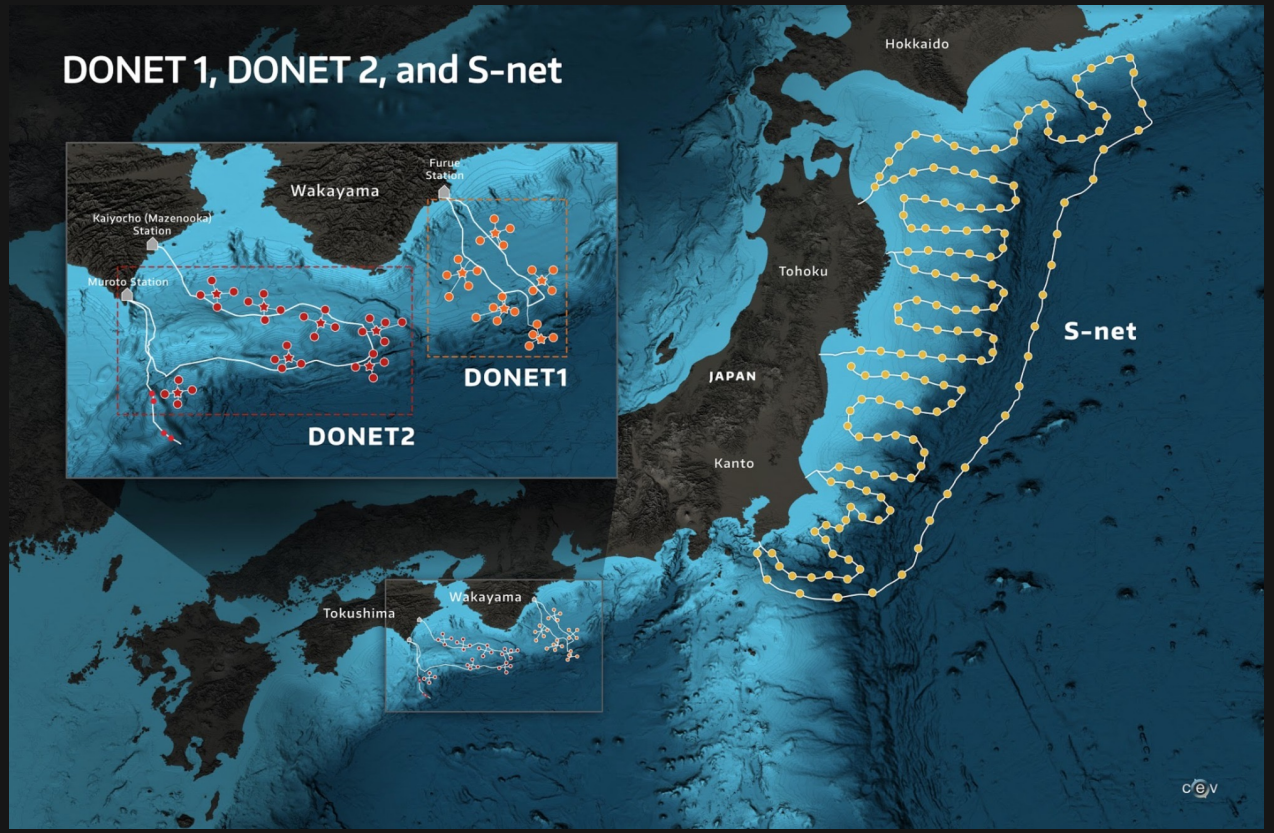


Image Credit:
CEV, U. Washington;
Schmidt et al., 2019

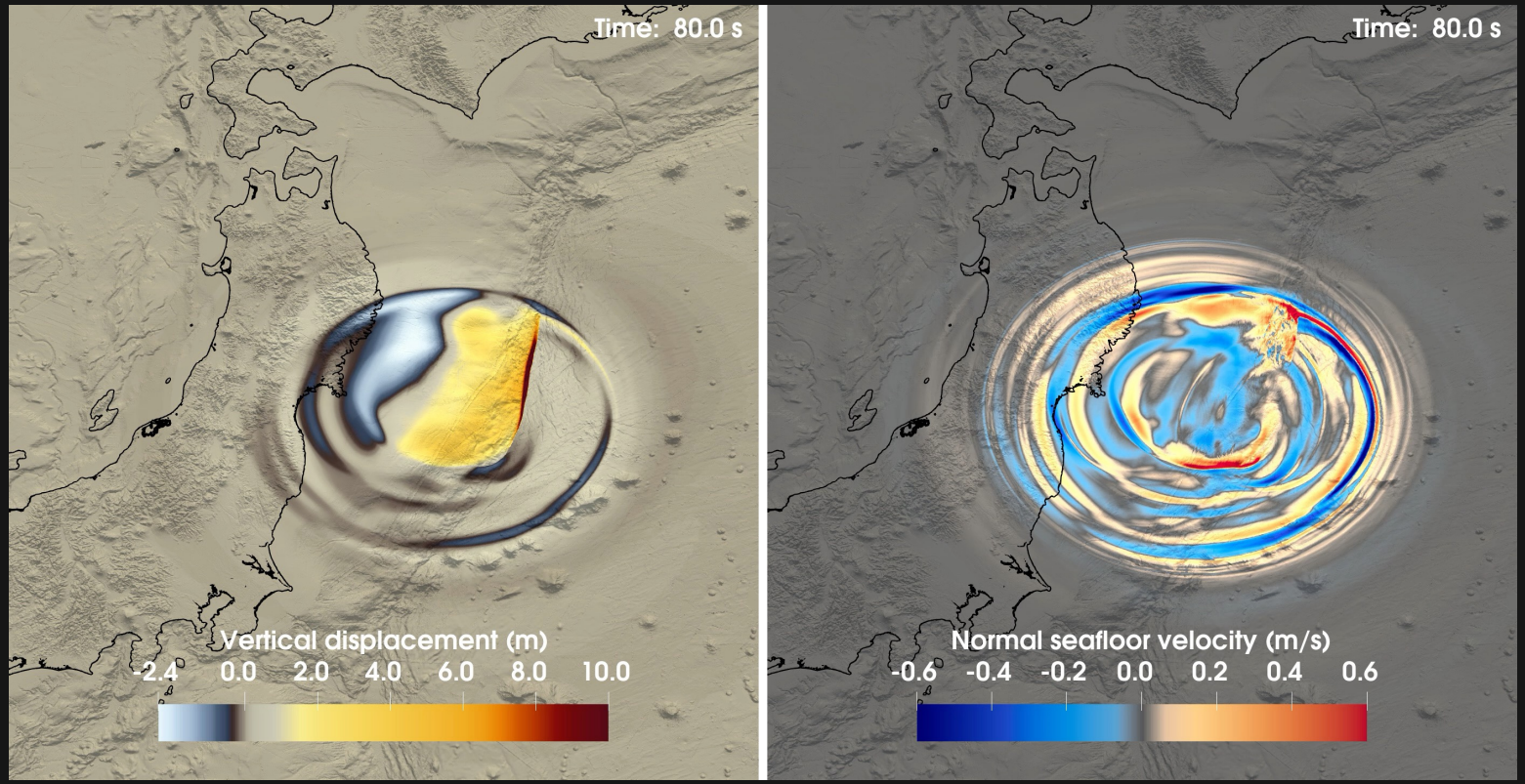
Tōhoku earthquake: Dynamic rupture simulation

Left:

Vertical seafloor uplift

Right:

Seafloor normal velocity



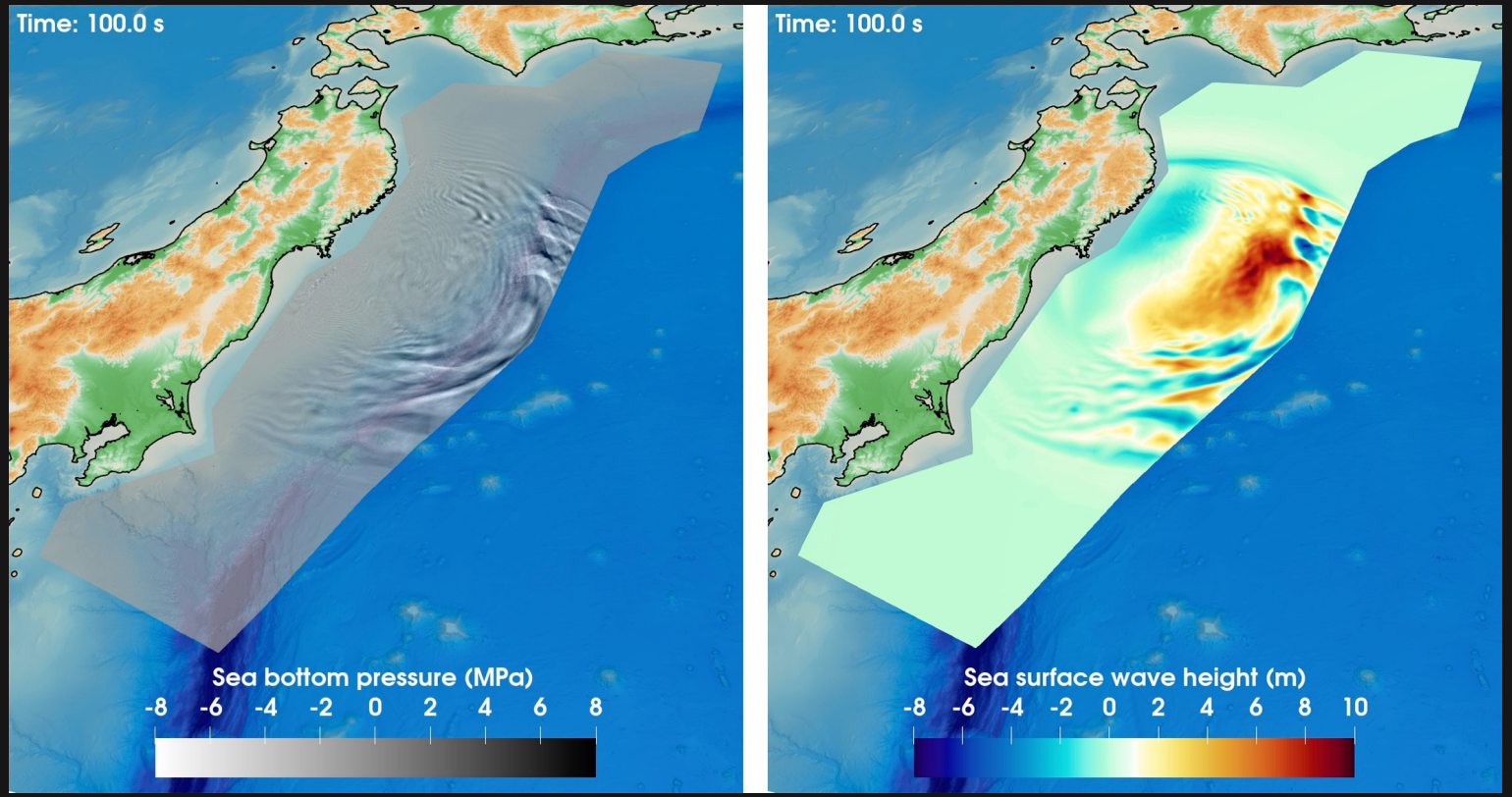
Forward model (acoustic–gravity simulation)

Left:

Seafloor pressure change

Right:

Sea surface wave height



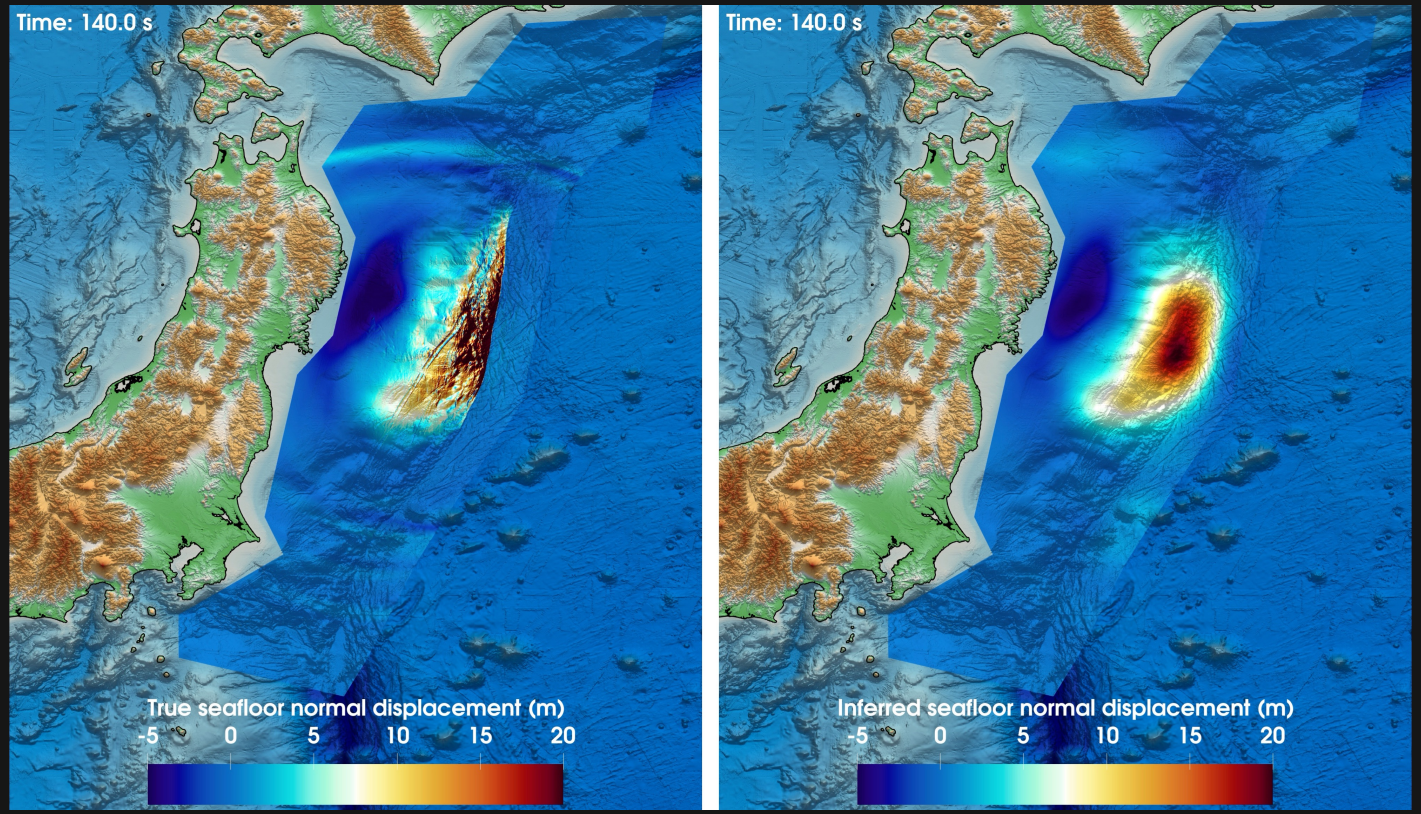
Tōhoku earthquake: Inverse solution

Left:

"True" seafloor normal uplift

Right:

Mean of inferred seafloor normal uplift



Summary

- Bayesian inverse problem to infer tsunamigenic seafloor motion directly from near-field pressure data using an acoustic–gravity model in the deep ocean
- Large-scale real-time inversion algorithms for linear autonomous dynamical systems, exploiting block Toeplitz structure of p2o map F, enabling fast construction and compact storage of F and fast FFT-based F and F* matvecs
- MAP point computed exactly in <0.2 s on 512 A100s, for inverse problem with 1B parameters and 3.7B spatial DOFs
- Wave heights at critical locations and their uncertainties computed in a fraction of a second by exploiting structure of push forward of posterior to QoI
- Optimal sensor selection for inverse problems governed by autonomous dynamical systems via standard and goal-oriented D-optimal design criteria.

Ongoing and future work

- Data-driven prior: Incorporation of database of hyothesized ruptures on the Cascadia fault into a more informative prior
  - Will result in a non-Gaussian prior; need to retain fast algorithms
- Inference of fault slip from not only seafloor acoustic sensors, but also land-based seismic waveform data
  - Permits forecasting of seismicity and production of “shake maps”
  - Seismic waves strike populated regions in less than a minute, so this is primarily to inform first responders
- Extensions to other problems with similar structure, e.g.
  - Threat detection (airborne, underwater, subsurface)
  - Nuclear testing detection
  - Contaminant transport

Acknowledgments

- This research was supported by DOD MURI grants FA9550-21-1-0084 and FA9550-24-1-0327 and DOE ASCR grant DE-SC0023171
- This research used resources from the Swiss National Supercomputing Centre (CSCS), the Texas Advanced Computing Center (TACC), the National Energy Research Scientific Computing Center (NERSC, ALCC-ERCAP0030671), and Livermore Computing (LC).

Further details in:

- S. Venkat, M. Fernando, S. Henneking, O. Ghattas. Fast and scalable FFT-based GPU-accelerated algorithms for block-triangular Toeplitz matrices with application to linear inverse problems governed by autonomous dynamical systems, SIAM Journal on Scientific Computing 47(5) , B1201–B1226, 2025.
- S. Henneking, S. Venkat, V. Dobrev, J. Camier, T. Kolev, M. Fernando, A.-A. Gabriel, O. Ghattas. Real-time Bayesian inference at extreme scale: A digital twin for tsunami early warning applied to the Cascadia subduction zone, Proceedings of The International Conference for High Performance Computing, Networking, Storage and Analysis (SC '25), ACM, November 16–21, 2025, St Louis, MO, USA.
- S. Henneking, S. Venkat, O. Ghattas. Goal-oriented real-time Bayesian inference for linear autonomous dynamical systems with application to digital twins for tsunami early warning, Journal of Computational Physics, 114682, 2026.
- J. Tu, I. Karlin, J. Camier, V. Dobrev, T. Kolev, S. Henneking, O. Ghattas. Accelerating high-order finite element simulations at extreme scale with FP64 tensor cores. Proceedings of ISC, to appear, 2026.
- S. Henneking, F. Kutschera, S. Venkat, A.-A. Gabriel, O. Ghattas. Real-time probabilistic tsunami forecasting in Cascadia from sparse offshore pressure observations. arXiv:2603.14966, 2026.
- S. Venkat, S. Henneking, O. Ghattas. Sensor placement for tsunami early warning via large-scale Bayesian optimal experimental design, arXiv:2604.08812, 2026.