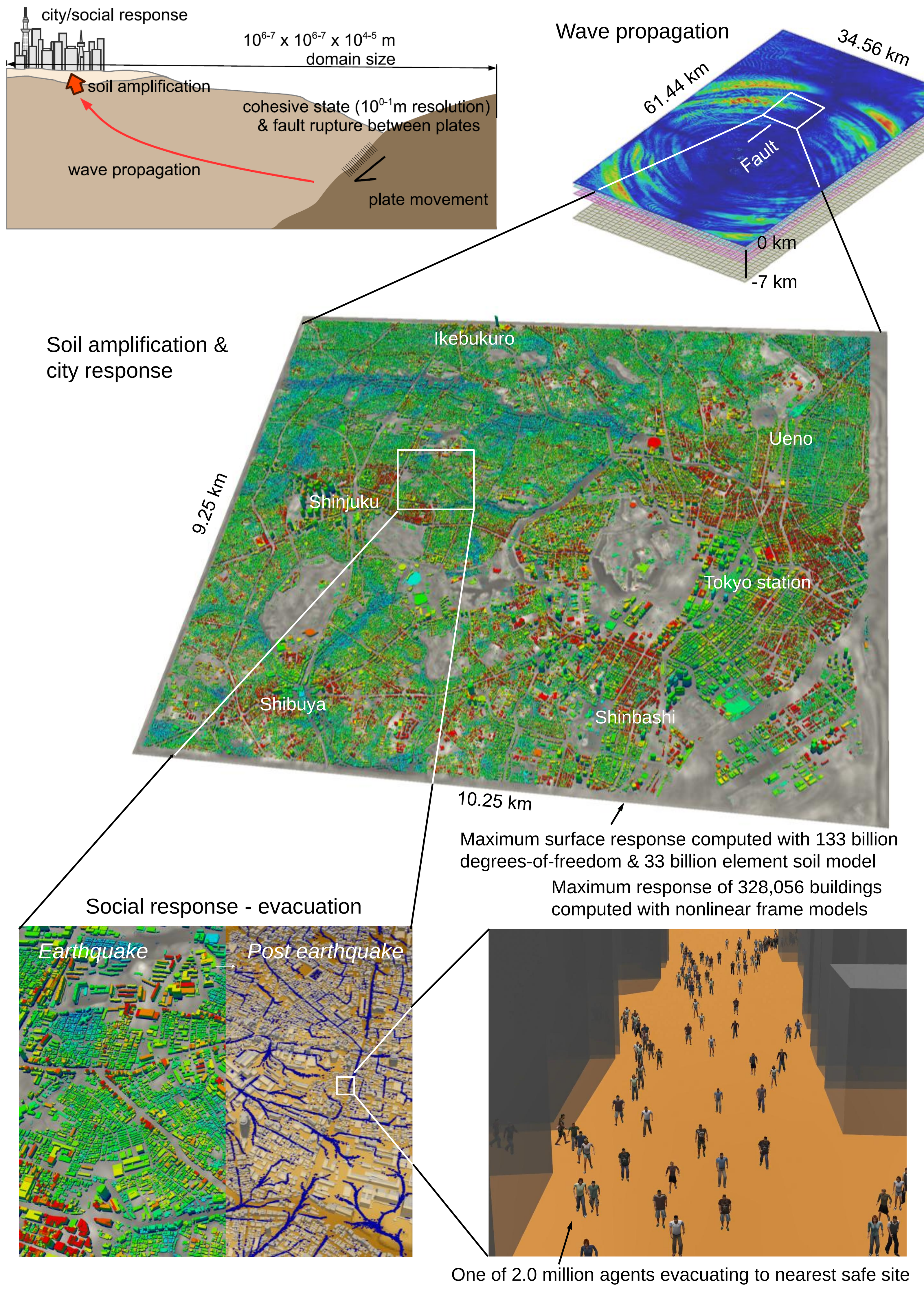
A Fast Implicit Solver with Low Memory Footprint and High Scalability for Comprehensive Earthquake Simulation System

Kohei Fujita[1], Tsuyoshi Ichimura[2], Kentaro Koyama[3], Masashi Horikoshi[4], Hikaru Inoue[3], Larry Meadows[5], Seizo Tanaka[6], Muneo Hori[2,1], Maddegedara Lalith[2] and Takane Hori[7]

1. Advanced Institute for Computational Science, RIKEN; 2. Earthquake Research Institute & Department of Civil Engineering, The University of Tokyo; 3. Frontier Computing Center, Fujitsu Limited; 4. Software and Solutions Group, Intel K.K.; 5. Data Center Group, Intel Corporation; 6. Faculty of Engineering, Information and System, University of Tsukuba; 7. R&D Center for Earthquake and Tsunami, Japan Agency for Marine-Earth Science and Technology

Comprehensive Earthquake Simulation System

- We are developing a comprehensive earthquake simulation system designed to model all phases of an earthquake disaster
- Large-scale forward modeling of wave propagation and soil amplification enabled by a fast, unstructured finite-element simulation (e.g., Ichimura et al. SC14 [Gordon Bell Finalist], SC15 [Gordon Bell Finalist])
- The next step is **physics-based earthquake forecasting** via assimilation of observational data and multiple crust-deformation analyses

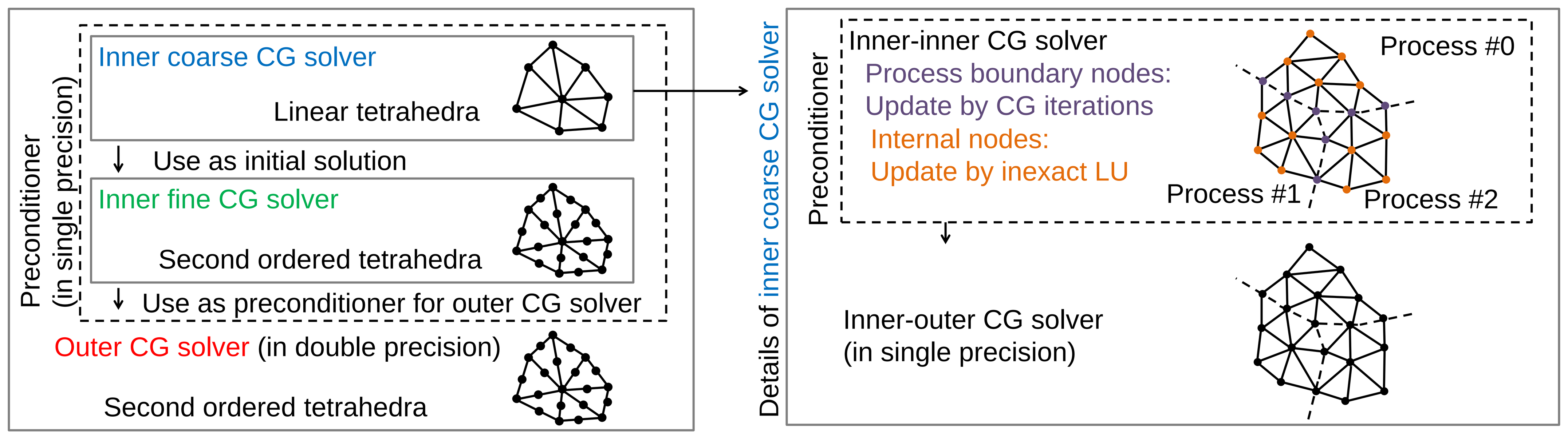


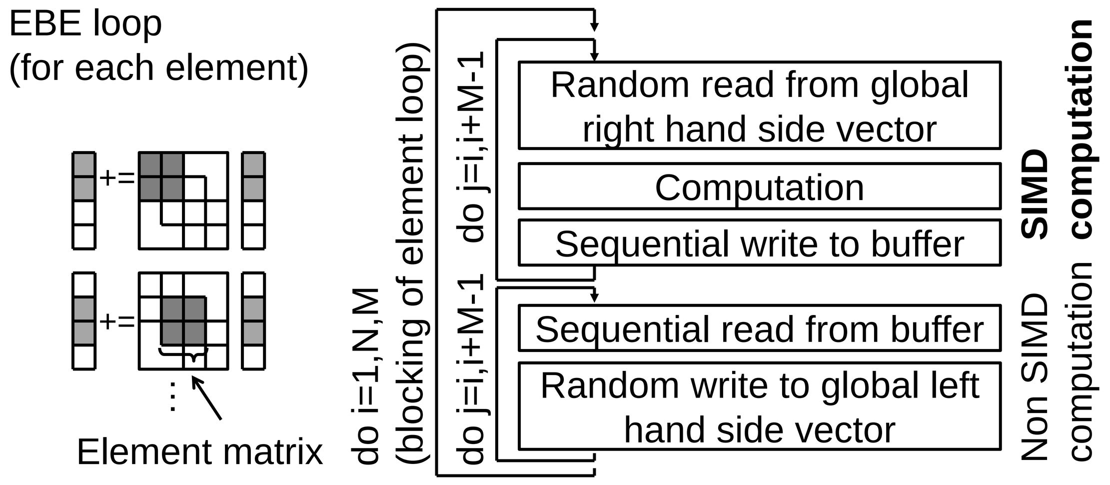
Target problem: T-DOF crust-deformation analysis for physics-based earthquake forecasting

- Dramatically improves the reliability of damage estimation by forward modeling, as shown above: most important core technology for comprehensive earthquake simulation system
- Problem size $10^{2\text{-}3}$ times larger than the state-of-the-art crust-deformation simulation: need of fast and scalable implicit solver with low memory footprint

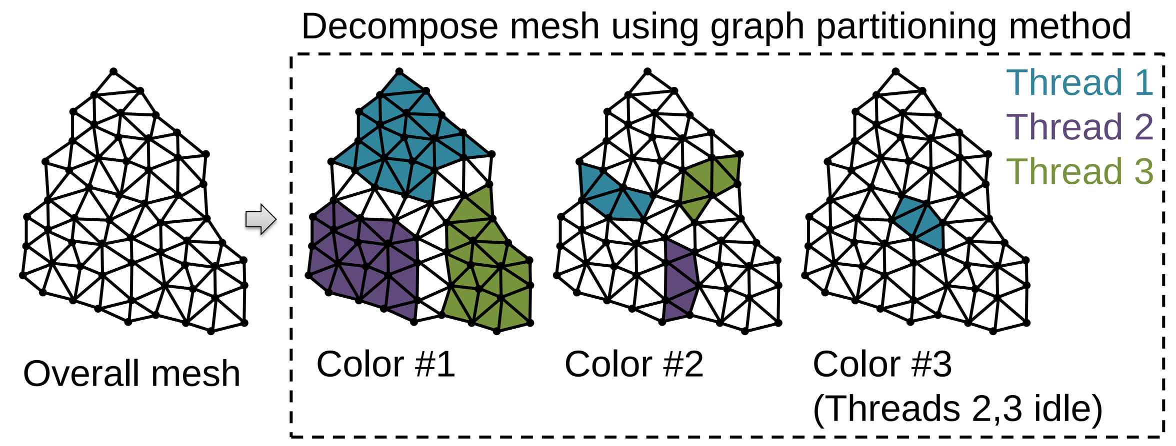
Key Ideas and Innovations

- **High convergence with a low-memory footprint:** communication avoiding inexact LU preconditioning used for coarse grid of a mixed-precision multi-grid preconditioner in an inexact conjugate gradient (CG) method
- **Fast computation with low-memory footprint:** a SIMD buffering and multi-core coloring method developed for efficient matrix-free matrix-vector multiplication [the element-by-element (EBE) method]
- **Highly scalable:** the on-cache EBE method circumvents load-imbalance from difference in memory access patterns in the unstructured computation





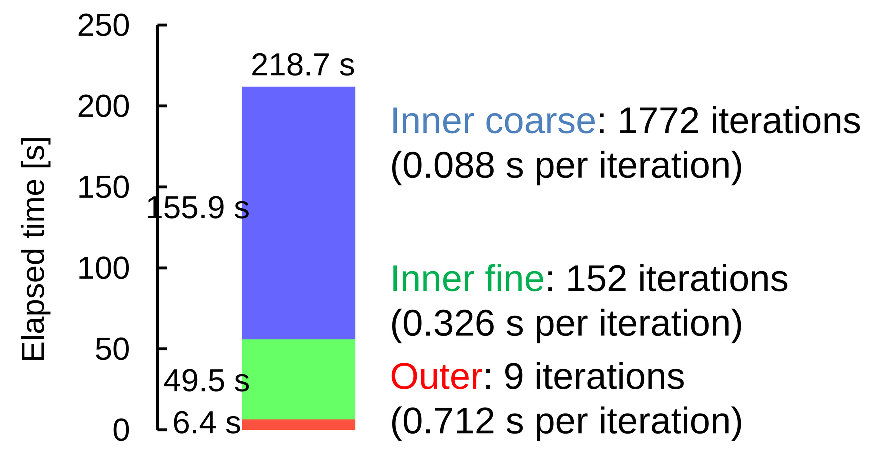
Buffering for the SIMD computation of an EBE kernel



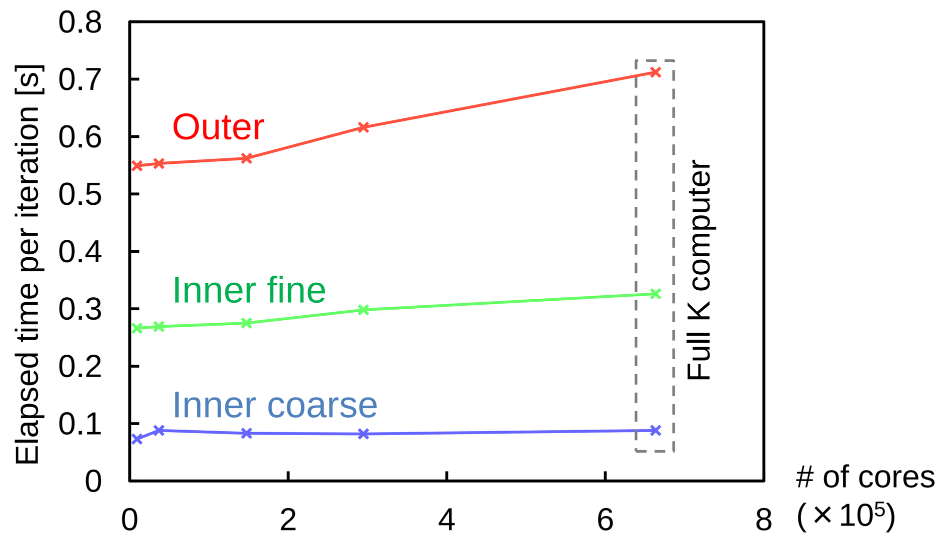
Efficient coloring for multi-core computation of an EBE kernel: Good load balance, high-locality, and fewer colors

Performance Measurement Results

Performance on the K computer



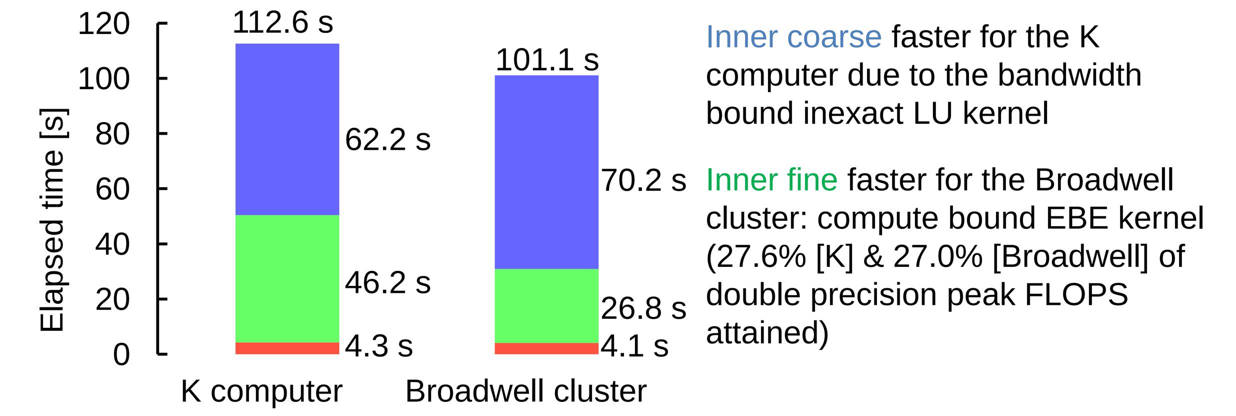
High DOF/memory attained: 2.36 T-DOF problem solved using the full K computer with 1.3 PB memory
High performance for unstructured FEM: 1.21 PFLOPS (11.4%) attained for the whole solver



High scalability attained up to the full K computer: 88.8% size-up efficiency for memory bandwidth utilization

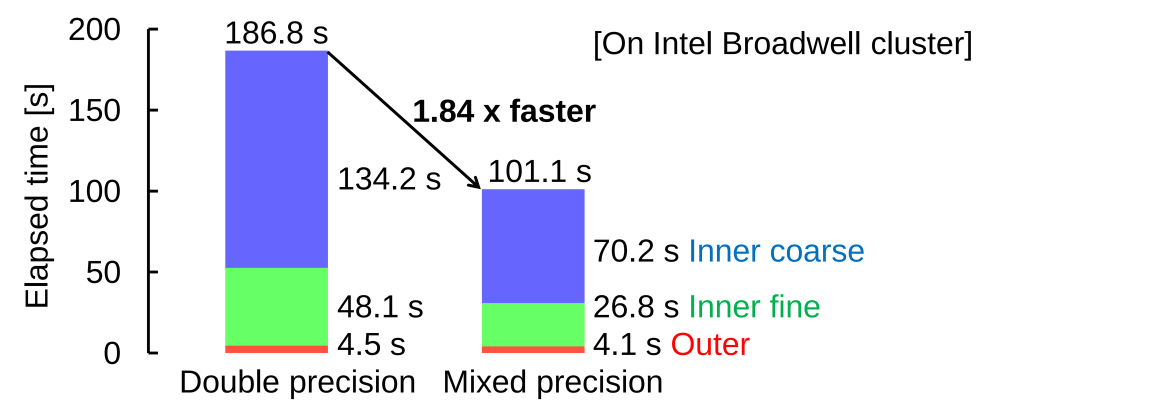
Performance portability

	Hardware peak double precision FLOPS	Hardware peak memory bandwidth
36 nodes of K computer with single eight-core SPARC VIIIfx	4.60 TFLOPS	2.30 TB/s
9 nodes of Intel Broadwell cluster with dual 18-core Intel Xeon E5-2697 v4	11.9 TFLOPS	1.38 TB/s



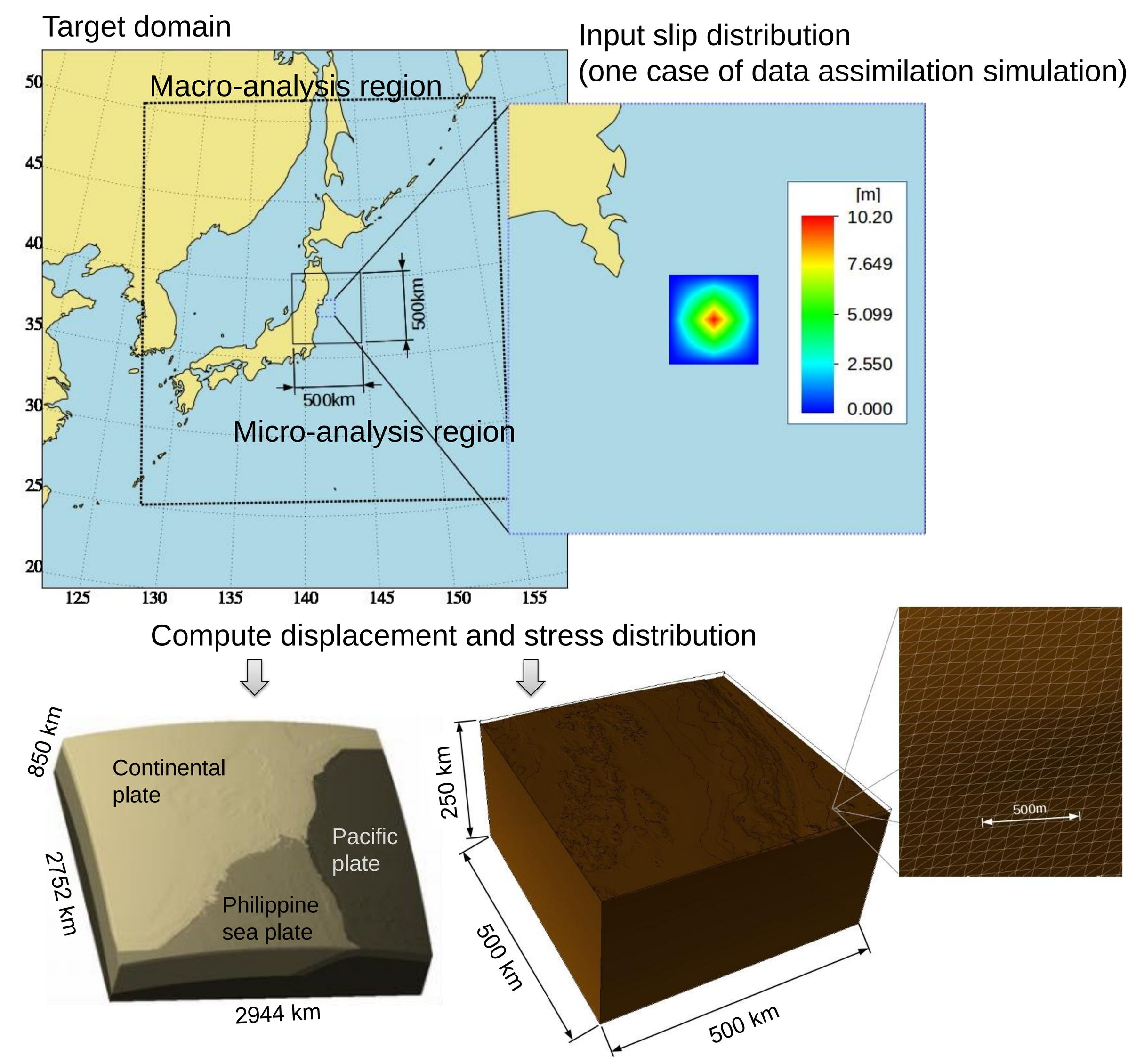
Performance improvement reflecting the increase in the number of cores and SIMD length. Further speedup on many-core machines expected by changing inexact LU kernel to EBE kernels

Effectiveness of the mixed-precision algorithm



1.84 times speedup attained by use of a mixed precision algorithm: inner fine and inner coarse loops accelerated by use of single precision

Application to a Practical Problem



Macro-analysis region (0.01 T-DOF problem with 500 m resolution in [1])
Micro-analysis region (2.05 T-DOF problem with 45 m resolution)

- Also developed a mathematical method to rigorously split a whole-island scale Peta-DOF problem into a macro-region problem in coarse resolution (0.01T-DOF) and micro-region problem in fine resolution (2.05 T-DOF)
- **An extremely large practical problem was solved**: a micro-analysis model with 2.05 T-DOF and 0.513 Tera elements (205 times larger than the current state-of-the-art) computed using the full K computer in 3199 s. High-resolution analysis enables computation of change of stress at plate boundaries
- **High performance and fast time-to-solution attained for the practical problem**: 1.07 PFLOPS (10.1% of peak) achieved for the solver, and 30 times faster than the memory-efficient CG solver developed in SC14 [2] (3×3 block Jacobi preconditioning and EBE method)

[1] Ichimura et al. Geophysical Journal International, 2016.
[2] Ichimura et al. SC14 Gordon Bell Prize Finalist, 2014.

Summary and Future Prospective

- Fast, scalable and portable solver with low-memory footprint developed for a comprehensive earthquake simulation system
- Short time-to-solution attained by putting a high load on the compute units, cache, and memory; the effective use of all of these units is required for high-performance
- This system running on the next generation of supercomputers is expected to enable $10^{4\text{-}5}$ times crust-deformation simulation and will be key to developing a dramatic reduction in seismic loss levels
- Also useful for non-seismic applications where appropriate modeling of geometry is essential

Acknowledgements

Results are obtained using the K computer at the RIKEN Advanced Institute for Computational Science (Proposal numbers: hp160221, hp150175, hp150127, hp160160, 160157). We acknowledge support from theJapan Society for the Promotion of Science (15K18110, 26249066, 25220908) and the FOCUS Establishing Supercomputing Center of Excellence. We thank the Operations and Computer Technologies Division of RIKEN Advanced Institute for Computational Science and the help-desk of High Performance Computing Infrastructure for support concerning the use of the K computer system. We used the crust data of Japan from the J-SHIS of National Research Institute for Earth Science and Disaster Prevention.