Evaluation of Tsunami Inundation Simulation using Vector-Scalar Hybrid MPI on SX-Aurora TSUBASA

Akihiro Musa[a),b)], Takashi Soga[a),b)], Takashi Abe[c)], Masayuki Sato[d)], Kazuhiko Komatsu[a)], Shunichi Koshimura[c)], and Hiroaki Kobayashi[d)]

[a)]Cyberscience Center, Tohoku University, [b)]NEC Corporation, [c)]IRIDeS, Tohoku University, [d)]Graduate School of Information Sciences, Tohoku University

Background and Motivation

◻Tsunami disasters may cause many fatalities and extensive damage of infrastructures. To quickly understand their disaster situations, we have developed a real-time tsunami inundation forecast system [1].

- The system consists of three components: Tsunami source modeling, Tsunami inundation simulation, Mapping and delivery.

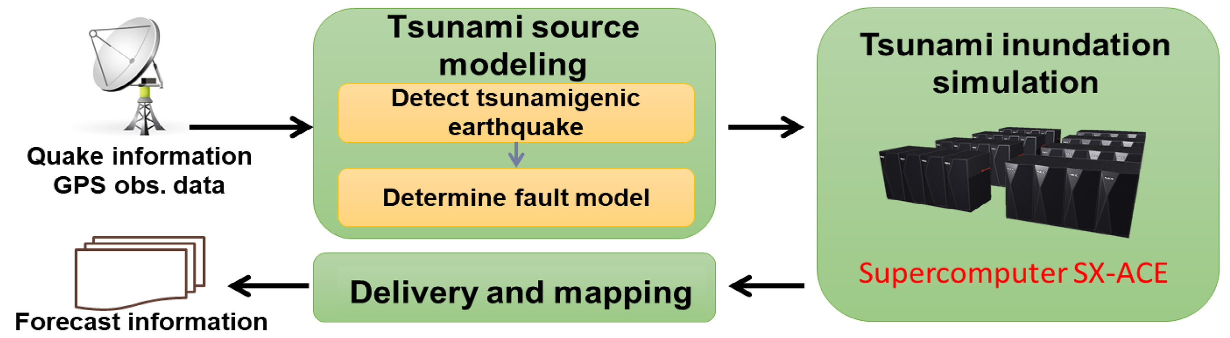
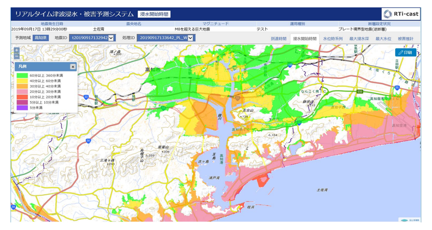
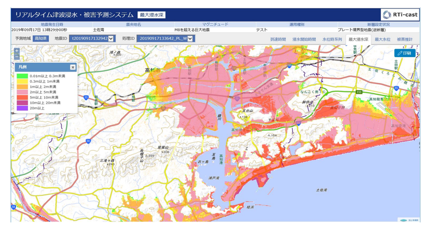


Diagram of the system

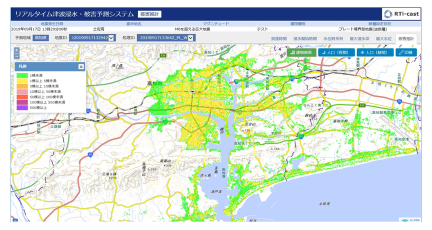
- The Cabinet Office of Japan began using the forecast system [2]. The system can predict damage situations for Japanese coastal regions (costal length: 8,000 km) at the resolution of a 30-meter grid size within 30 minutes. The system contains 550 nodes of NEC supercomputer SX-ACE, and the tsunami inundation simulation has been executed in less than five minutes.



Starting time of inundation



Maximum tsunami inundation depth



Structural damage (houses)

Samples of forecast information

◻As downsizing the system and reducing the processing time are required for local governments and industries, we are developing a vector-scalar hybrid MPI code for the tsunami inundation simulation on SX-Aurora TSUBASA [3], which consists of two Xeon processors and a vector engine (called a workstation type).

◻We evaluate the performance of the vector-scalar hybrid MPI code on the workstation type for examining the possibility of the real-time tsunami inundation forecast system using the Kochi prefecture domain.

Tsunami inundation Simulation

◻The tsunami inundation simulation is based on the TUNAMI code[1].

Governing equations

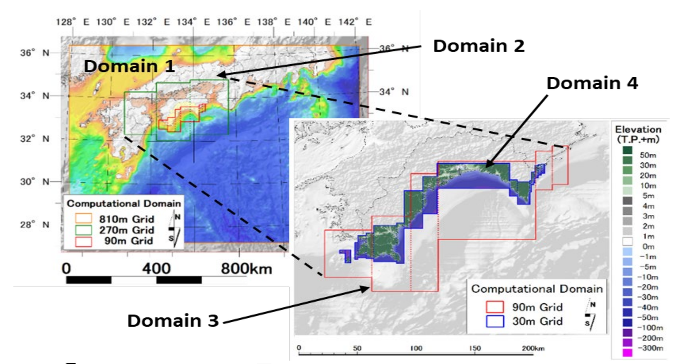
$$\frac{\partial \eta}{\partial t}+\frac{\partial M}{\partial x}+\frac{\partial N}{\partial y}=0 \qquad M=\int_{-h}^{\eta} u dz \qquad N=\int_{-h}^{\eta} v dz$$

$$\frac{\partial M}{\partial t}+\frac{\partial}{\partial x}\left(\frac{M^2}{D}\right)+\frac{\partial}{\partial y}\left(\frac{MN}{D}\right)=-gD\frac{\partial \eta}{\partial x}-\frac{gn^2}{D^{7/3}}M\sqrt{M^2+N^2}$$

$$\frac{\partial N}{\partial t}+\frac{\partial}{\partial x}\left(\frac{MN}{D}\right)+\frac{\partial}{\partial y}\left(\frac{N^2}{D}\right)=-gD\frac{\partial \eta}{\partial y}-\frac{gn^2}{D^{7/3}}N\sqrt{M^2+N^2}$$

- Numerical scheme: Staggered leap-frog difference method
- Polygonally nested grid system [2]: Four nested grid domains

Nested grid Domain	Grid size (meters)	No. of grid points
1	810	2.01×10^6
2	270	1.70×10^6
3	90	1.81×10^6
4	30	4.98×10^6



Nested grid domain in Kochi prefecture, Japan

Supercomputer SX-Aurora TSUBASA

◻SX-Aurora TSUBASA is the latest NEC vector supercomputer.

- It consists of one or more card-type vector engines (VEs) and a vector host (VH).
- VE contains a vector processor that is mainly used for the execution of simulations.
- VH is a standard x86 Linux server that is mainly used to control VEs.

◻NEC MPI/Vector-Scalar Hybrid MPI environment

- The environment supports MPI communications between VE and VH.
- Both VE and VH can be used as a single system image.

◻Conditions for applying the Vector-Scalar Hybrid MPI

- Computing loads should be proportionately allocated to VH and VEs depending on their performances.
- MPI communication between VH and VEs is not a significant overhead.

Specifications of the workstation type

	Model	System				Core	
		No. of CPUs	No . of cores	Perf.	Mem. bandwidth	Perf.	LLC capacity
VH	Xeon 6148	2	40	2.8 Tflop/s	256 GB/s	0.07 Tflop/s	27.5 MB
VE	20B	1	8	2.4 Tflop/s	1.35 TB/s	0.3 Tflop/s	16 MB

	Version	options
Compiler	Intel ifort 19.0.4.243	-O3 –ipo -vec-threshold0 -xCORE-AVX512
	NEC nfort 2.1.28	-O3 -finline-functions -msched-block
MPI lib.	NEC MPI Version 2.6.0	

Implementation of a vector-scalar hybrid MPI

◻The code is parallelized by the one-dimensional domain decomposition method for keeping a longer vector length.

- The outermost loop is parallelized and the innermost loop is vectorized.
- Each nested grid domain is independently divided into subdomains, and each subdomain is assigned to an MPI process.

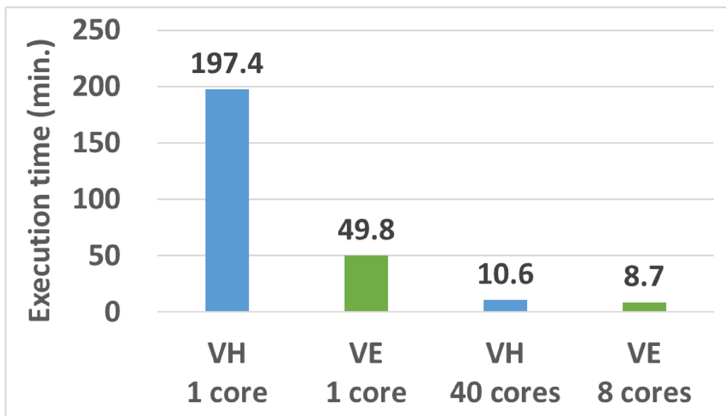
```
subroutine nlmnt2
  ...
  do j=jsta,jend      ! Parallelized
    do i=ista,iend    ! Vectorized
      call xmmt       ! Inlined
    end do
  end do
  ...
end subroutine nlmnt2
```

Sample of program structure

◻We perform two preliminary evaluations of the tsunami inundation simulation on SX-Aurora TSUBASA for the vector-scalar hybrid MPI.

(1) The computational performances

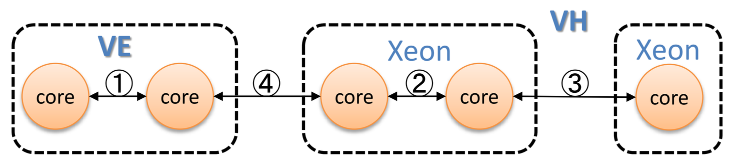
- To determine the number of grid points per core in the hybrid MPI parallelization
- The single core of VE calculates the code in one-fourth the execution time of VH.
- The execution time of the 40 cores on VH is 1.2 times longer than that of the 8 cores on VE. The case of eight cores cannot achieve high parallel performance due to an imbalance of computational load.



Execution times on VH and VE

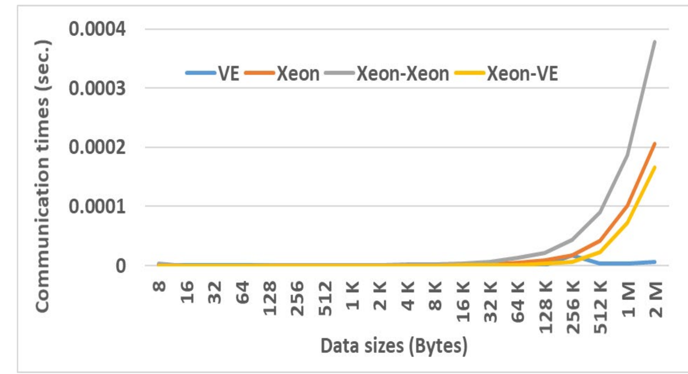
(2) The MPI communication

- To clarify the impact of MPI communication on the vector-scalar hybrid MPI code. The performance of four cases is evaluated by using the Pingpong benchmark. The data are transferred ① within one VE, ② within one Xeon processor, ③ between two Xeon processors, and ④ between one Xeon processor and one VE.



Four evaluation cases

- The cases of "Xeon", "Xeon-Xeon", and "Xeon-VE" quickly increase the times with 256 KB or more.
- The case of "VE" is nearly constant, 1.0×10^{-6} seconds.
- The maximum communication data size in the simulation is about 64 KB, which indicates that the MPI communication times has a little impact on the performance of the vector-scalar hybrid code.



MPI communication times

◻Implementation method

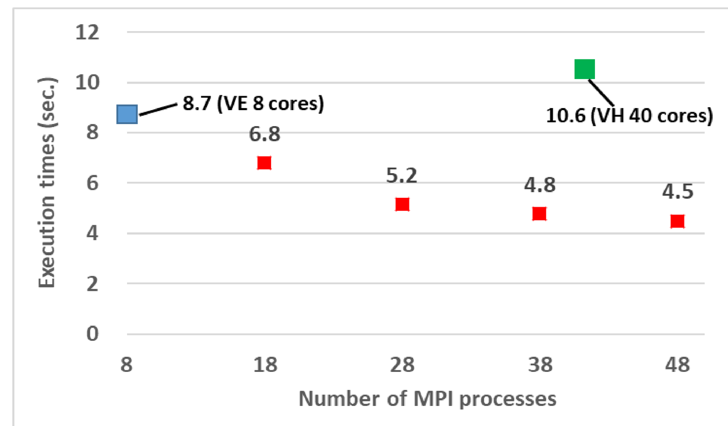
- To keep load among VE and VH, four time more task are assigned to a core of VE.
- To minimize the MPI communication time, neighboring domains are assigned into the same processors as much as possible.

Evaluation on the vector-scalar hybrid code

◻The vector-scalar hybrid code is evaluated on the workstation type from 18 MPI processes (8 cores from VE and 10 cores from VH) to 48 MPI processes (8 cores from VE and 40 cores from VH).

◻The vector-scalar hybrid code achieves the shortest time of 4.5 min. with a hybrid of 8 VE cores and 40 VH cores.

- The vector-scalar hybrid code achieves roughly two times faster than VE or VH only implementation.
- The performance improvement is saturated from 28 cores due to the bottleneck of memory bandwidth on VH.



Execution times on Hybrid code

Conclusions and Future Work

◻We have developed the vector-scalar hybrid MPI coed for the tsunami inundation simulation, and achieved the execution time of less than 4.5 minutes in the Kochi prefecture domain. The workstation type with the vector-scalar hybrid code can be utilized as the real-time tsunami inundation forecast system for a prefecture's disaster information system. The *tsunami source modeling* and *mapping and delivery* processes can also run on VH. We are currently developing the forecast system on the workstation type of SX-Aurora TSUBASA.

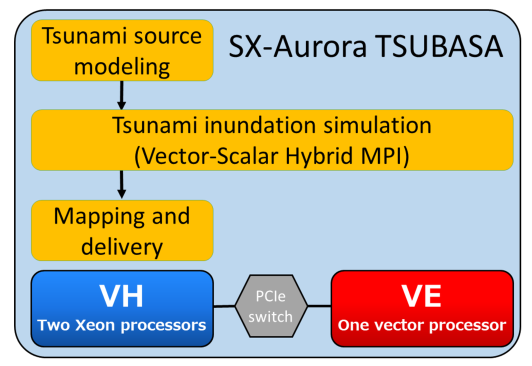


Diagram of the new system

Acknowledgements

This study was partially supported by JSPS KAKENHI Grant Numbers 18K11322 and Ministry of Education, Culture, Sports, Science and Technology of Japan (Next Generation High-Performance Computing Infrastructures and Applications R&D Program).

Reference

[1] A. Musa, O. Watanabe, H. Matsuoka, H. Hokari, T. Inoue, Y. Murashima, Y. Ohta, R. Hino, S. Koshimura, H. Kobayashi, "Real-time tsunami inundation forecast system for tsunami disaster prevention and mitigation," Journal of Supercomputing, vol. 74, issue 7, pp 3093--3113, 2018.

[2] T. Inoue, T. Abe, S. Koshimura, A. Musa, Y. Murashima, H. Kobayashi, ``Development and Validation of a Tsunami Numerical Model with the Polygonally Nested Grid System and its MPI-Parallelization for Real-Time Tsunami Inundation Forecast on a Regional Scale," Journal of Disaster Research, 14(3):416-434, 2019.

[3] SX-Aurora TSUBASA. https://www.nec.com/en/global/solutions/hpc/