



# **Introduction of AOBA-S**

## **The world's largest SX-Aurora TSUBASA system operating at Tohoku University**

**June 14, 2024**  
**NUG Society Meeting 35**

**Tohoku University**  
**Hiroyuki Takizawa**  
<[takizawa@tohoku.ac.jp](mailto:takizawa@tohoku.ac.jp)>



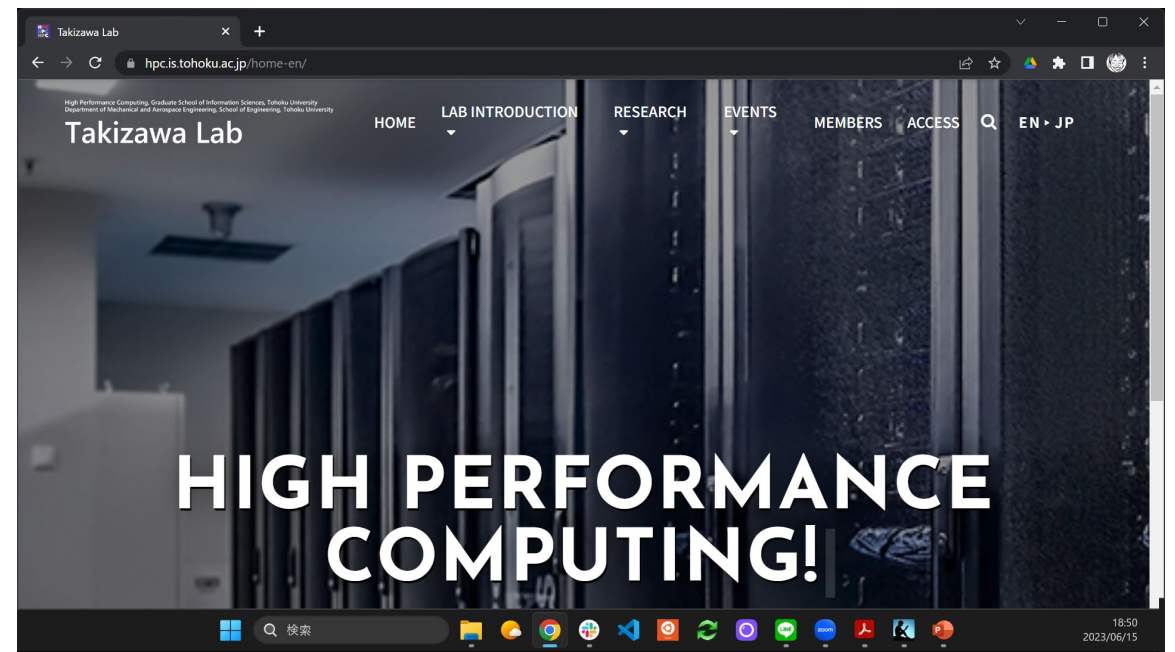
# Self-introduction

- **Hiroyuki TAKIZAWA**

- Professor and Deputy Director of the **Cyberscience Center**, Tohoku University.
- **HPC Lab** at Graduate School of Information Sciences, Tohoku University

<https://www.hpc.is.tohoku.ac.jp>

## Laboratory Members @ Online Drinking Party





# Tohoku University 東北大学

- Tohoku University was established as Japan's third national university in 1907. Located on the ancient site of Aoba Castle in Sendai City, Tohoku University is proud to be ranked among Japan's leading universities.

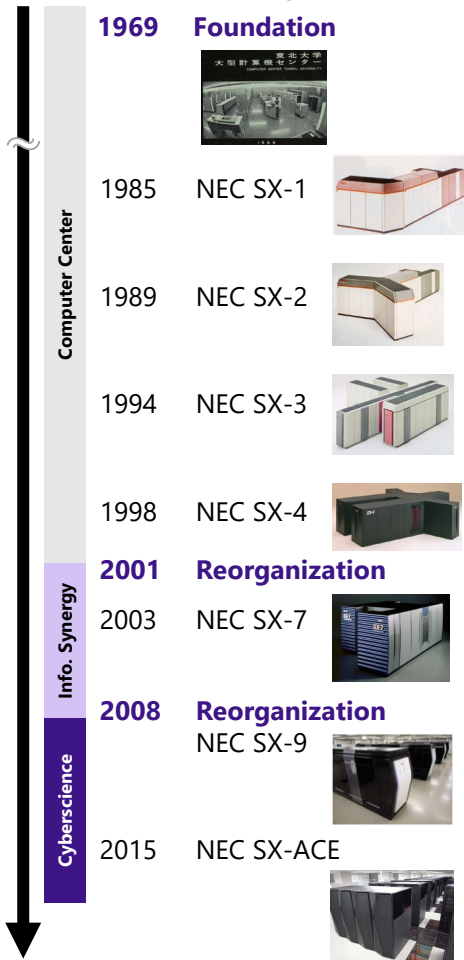
(<https://www.tohoku.ac.jp/en/about/index.html>)

17,685 Students  
3,145 Faculty Staffs



# History and Missions of CSC Tohoku-U

- History



- Missions of **Cyberscience Center**

- **Offering leading-edge computing environments to academic users nationwide in Japan**

- 24/7 service of **Supercomputer AOBA**
- **1,612 users** (as of March 2024)

- User supports

- Benchmarking, analyzing, and tuning users' programs
- Presenting seminars and lectures on supercomputing

- **R&D on supercomputing**

- Joint research projects with users and NEC on HPC
- Designing next-generation high-performance computing systems and their applications for **highly-productive supercomputing**

- **Education**

- Teaching and supervising BS, MS and Ph.D. Students

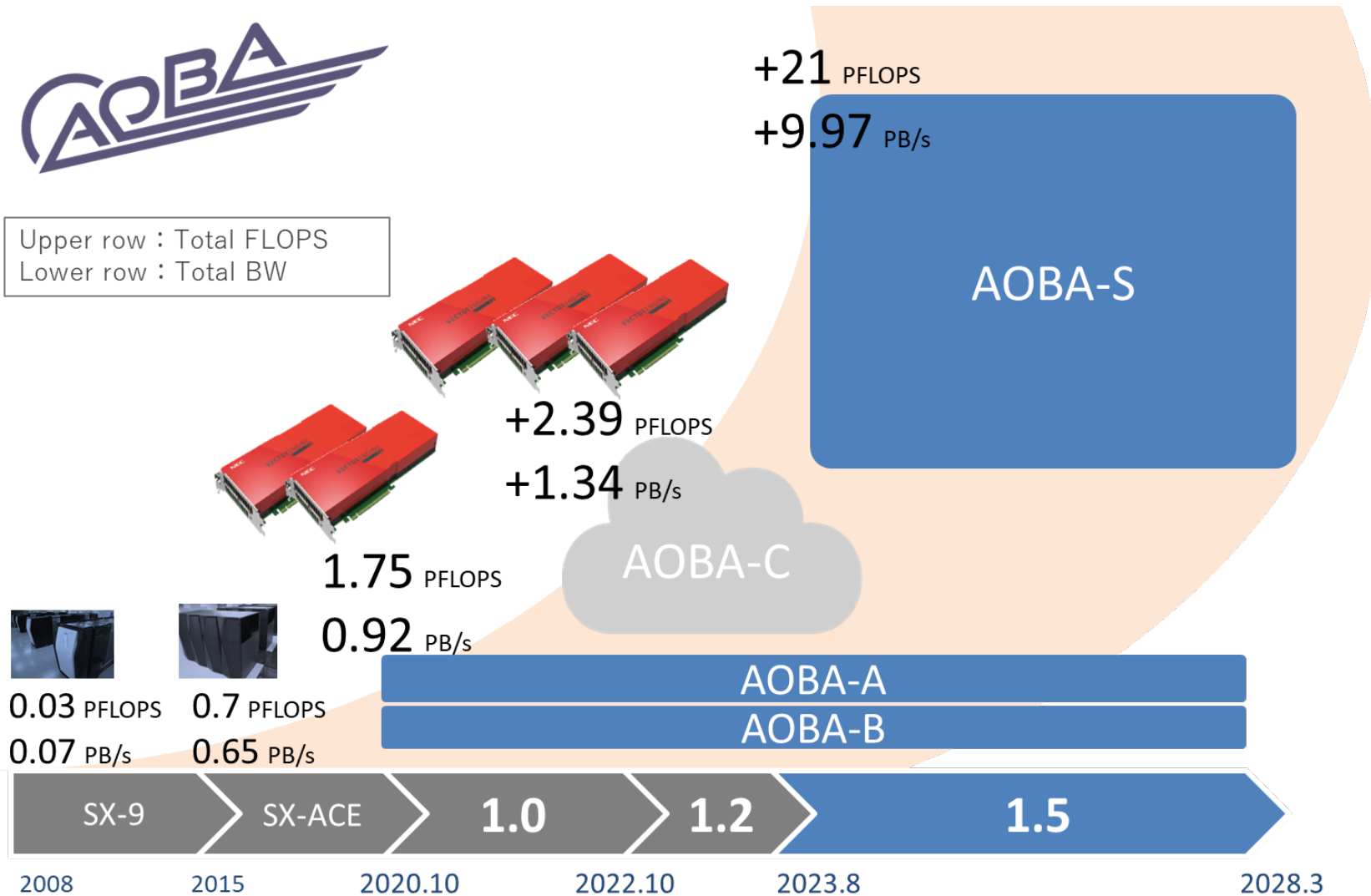


**AOBA (2020)**

# System Updates



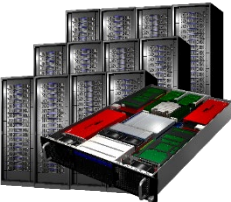
Upper row : Total FLOPS  
Lower row : Total BW




# System Configuration of AOBA-1.5

NEC SX-Aurora TSUBASA C401-8 x 504


DDN ES400NVX2



**AOBA-S**  
**21.05 Pflop/s**  
9.97 PB/s



**Storage**  
4.5 PB  
(Lustre)



**Front-end servers**




Internet


InfiniBand NDR 200G

Ethernet 10G


InfiniBand HDR 200G



**Front-end servers**



**Storage**  
2 PB (ScaTeFS)



**AOBA-A**  
**1.48 Pflop/s**  
893 TB/s



**AOBA-B**  
279 Tflop/s  
29 TB/s

DDN SFA7990XE

NEC SX-Aurora TSUBASA B401-8 x 72

NEC LX 406Rz-2 x 68

# AOBA-S Started Operation in August 2023



**System operation experience (esp. user support activity) + research activity**



# HPL and HPCG Benchmarks

- They are **two extremes** for benchmarking HPC systems



## High Performance Linpack (HPL)

solving a dense system of linear equations.

HPL performance is well correlated with **theoretical peak flop/s rate** of the system.

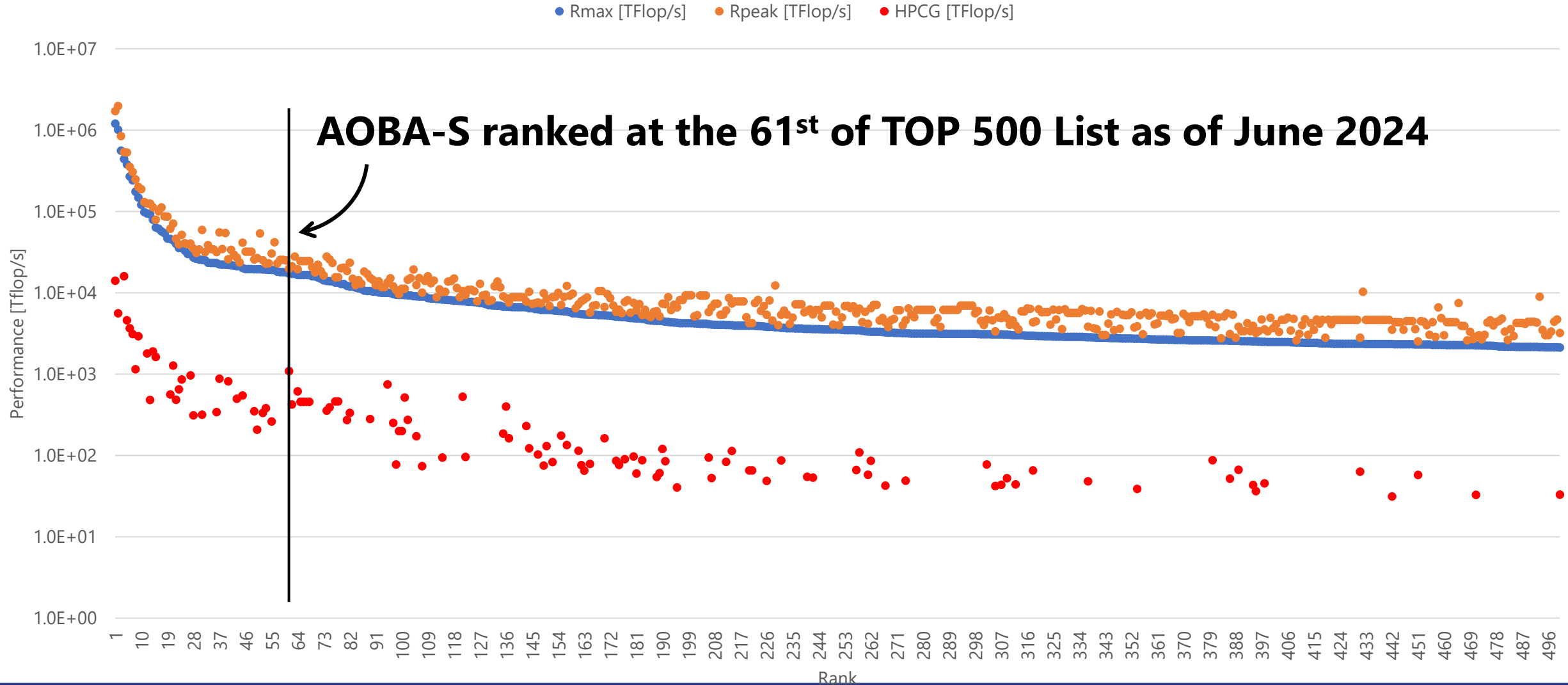
## High Performance Conjugate Gradient (HPCG)

solving a sparse system of linear equations.

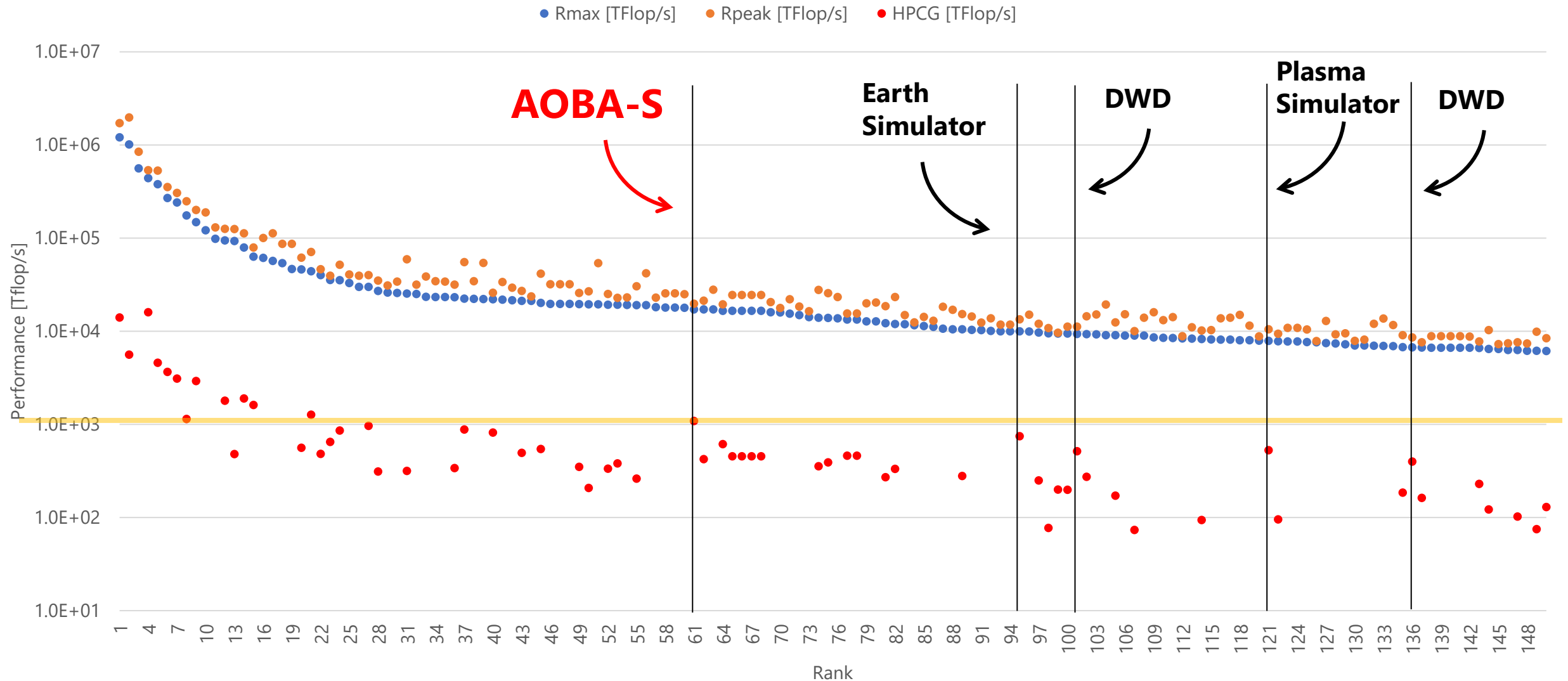
HPCG is intended as a complement to HPL, and the HPCG performance is heavily influenced by **memory bandwidth** of the system.



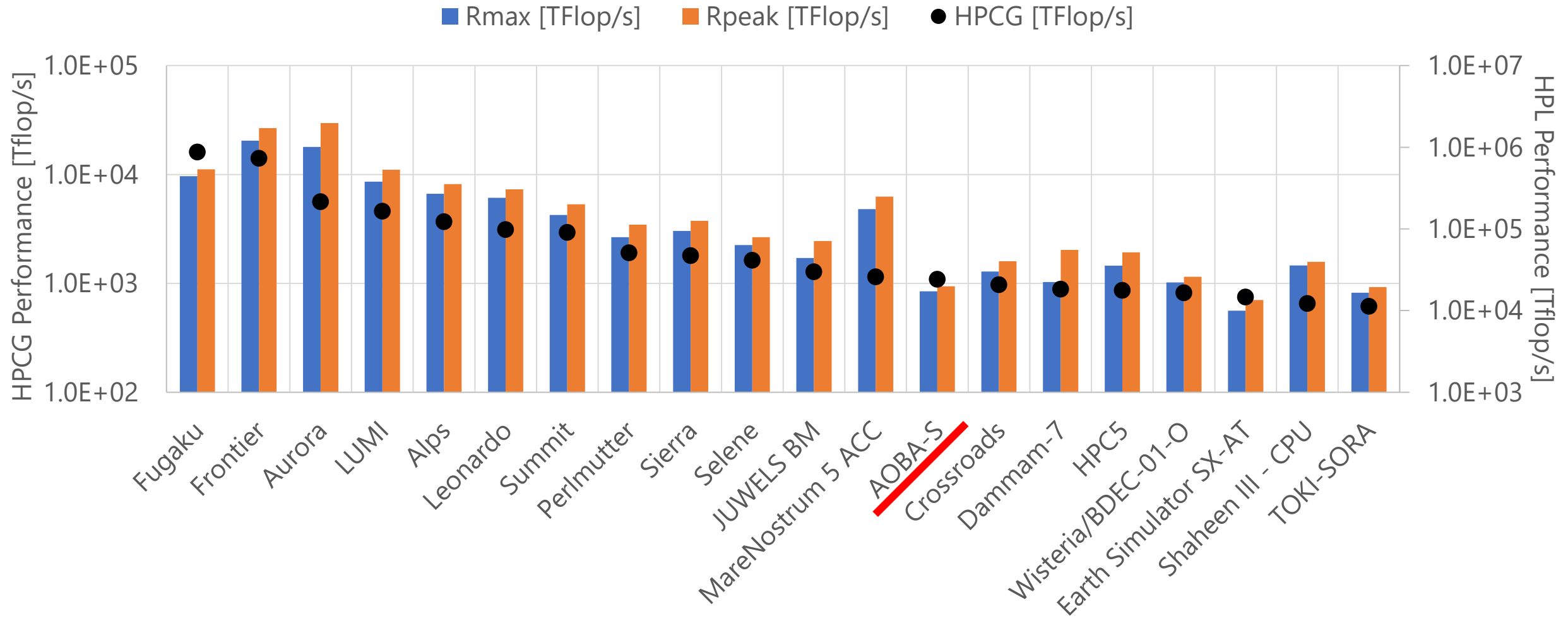
# TOP 500 List



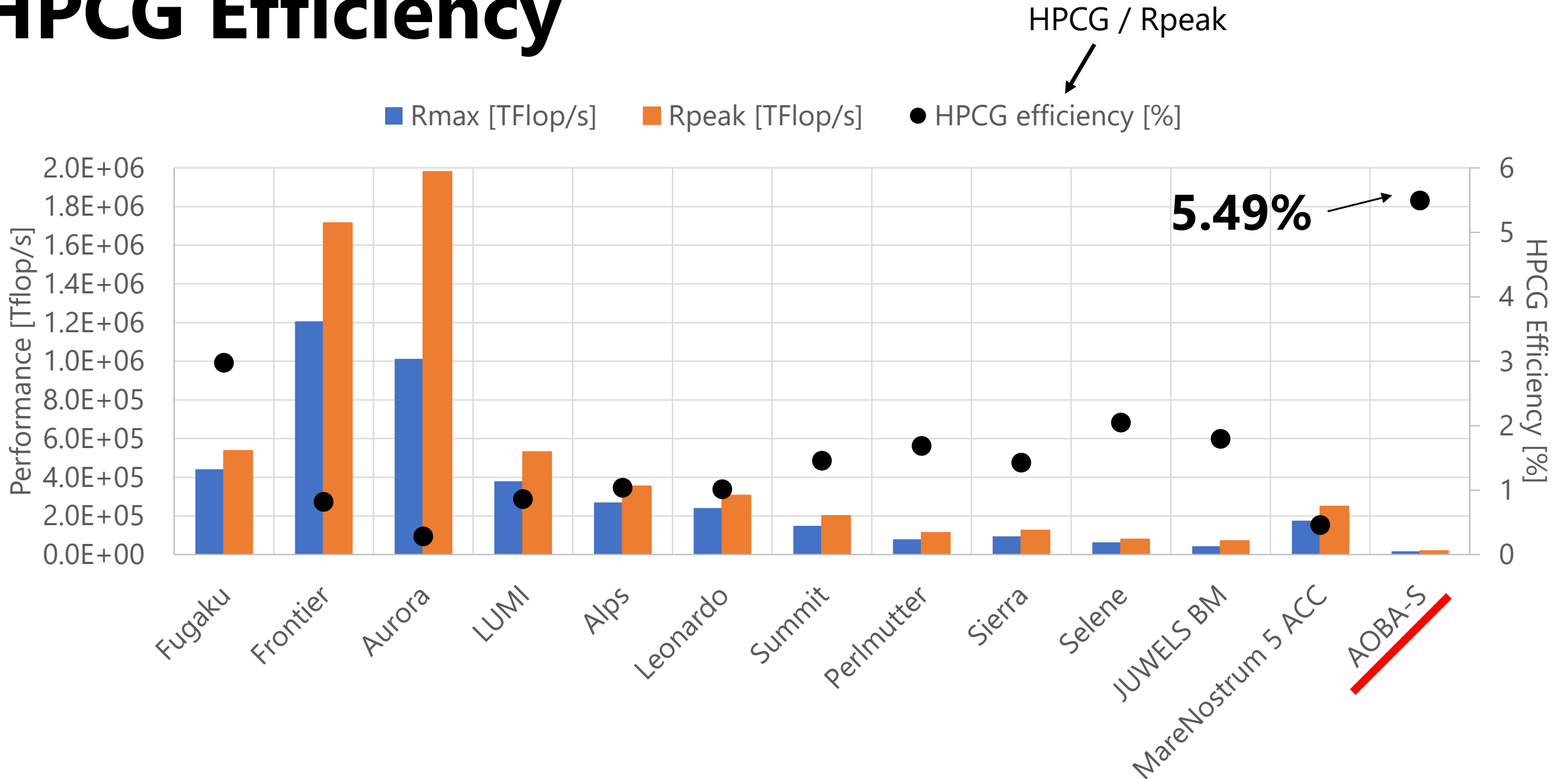
# Closer Look at Top 150 Systems



# Top 20 Systems in HPCG List



# HPCG Efficiency





# The Most Powerful **Vector** Supercomputer!

(As of June 2024)



**Very efficient for memory-intensive workloads**

Takahashi et al: Performance Evaluation of a Next-Generation SX-Aurora TSUBASA Vector Supercomputer, International Conference on High Performance Computing (ISC'23), pp.359-378, 2023.



# User Support Activity at Tohoku University

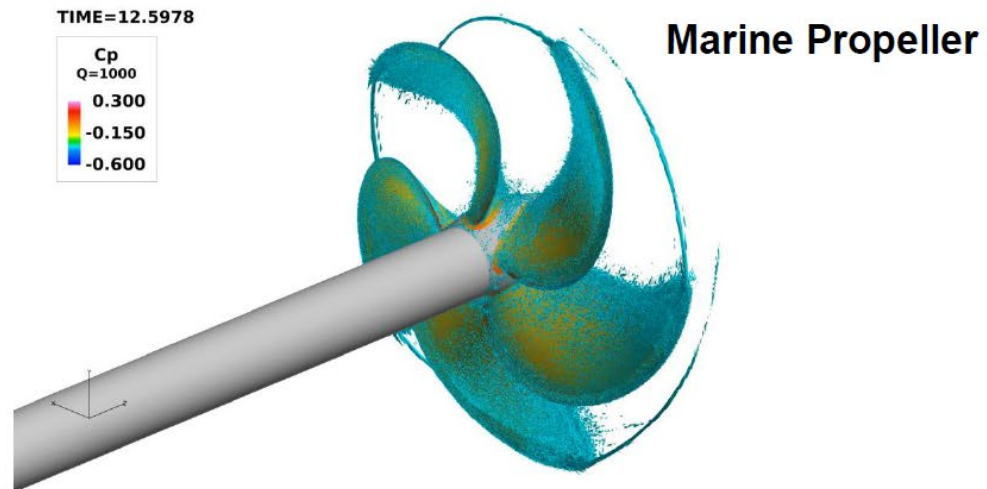
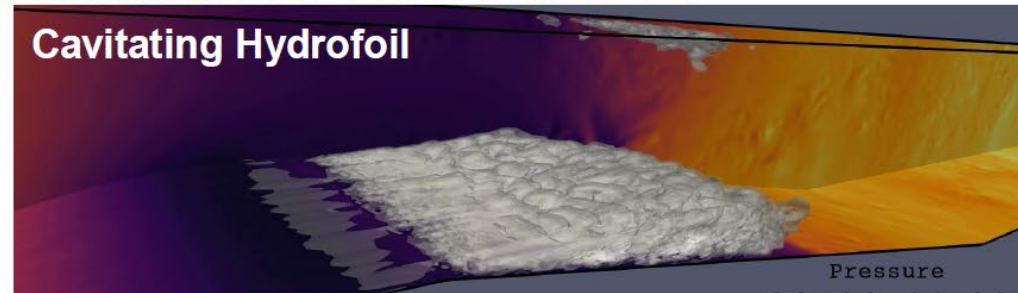
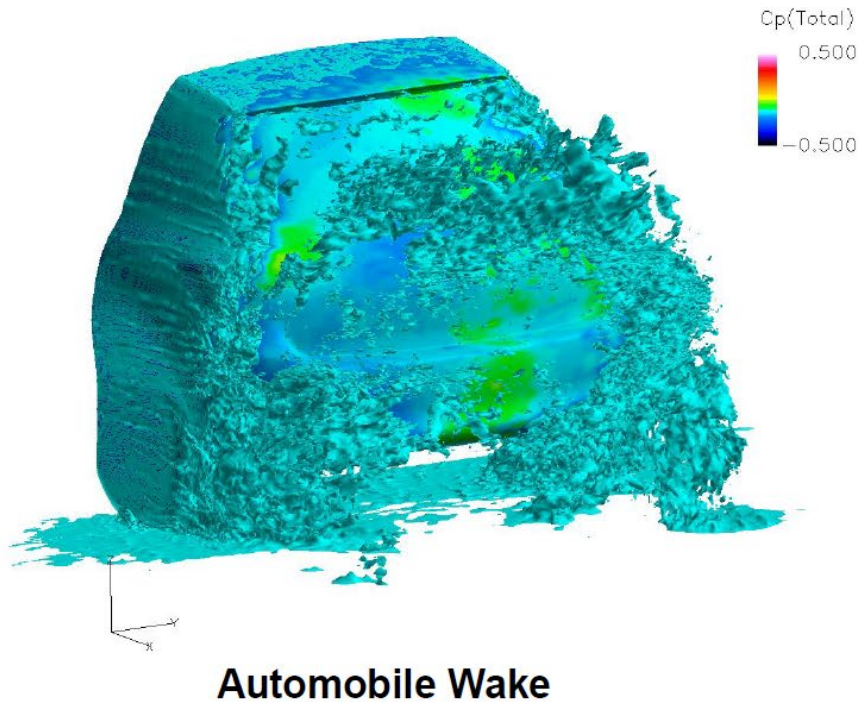
NUG Society Meeting 35





# FrontFlow/blue (FFB) Flow Solver

- FEM-based incompressible/compressible Flow Solver
- Developed for Industrial Applications of WR-LES
- Features Automated Mesh Refinement and Overset Method

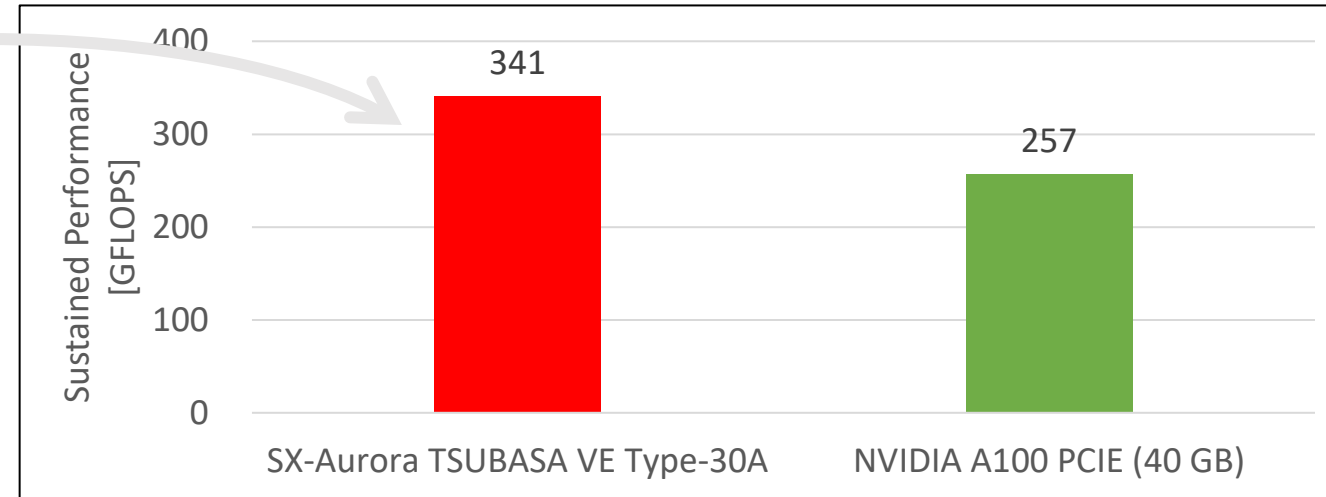


# Code Optimization for SX-Aurora TSUBASA

- **Exploiting the memory bandwidth of SX-Aurora TSUBASA VE3 for real-world applications**
  - Promoting compiler's vectorization
  - Use of Basic Sparse Matrix Operation library (NLC SBLAS)
  - Optimization of reordering parameters

The speedup ratio is reasonable by considering the difference in sustained memory bandwidths.

→ **Code is well optimized for SX-Aurora TSUBASA**  
(thanks to **Toshihiro Kato** at NEC)



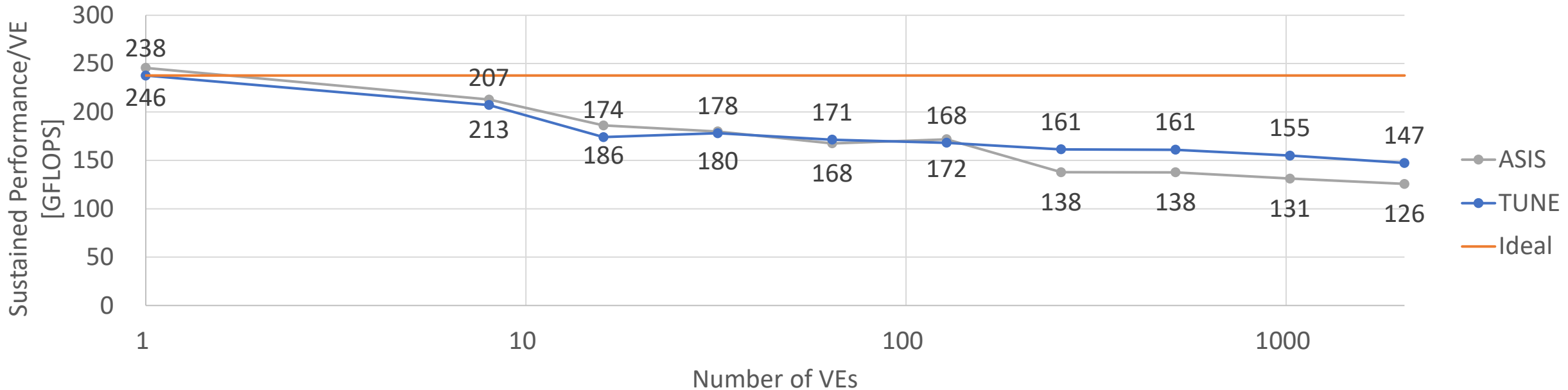
**NEC Numeric Library Collection SBLAS**

[https://sxauroratsubasa.sakura.ne.jp/documents/sdk/SDK\\_NLC/UsersGuide/sblas/f/en/index.html](https://sxauroratsubasa.sakura.ne.jp/documents/sdk/SDK_NLC/UsersGuide/sblas/f/en/index.html)



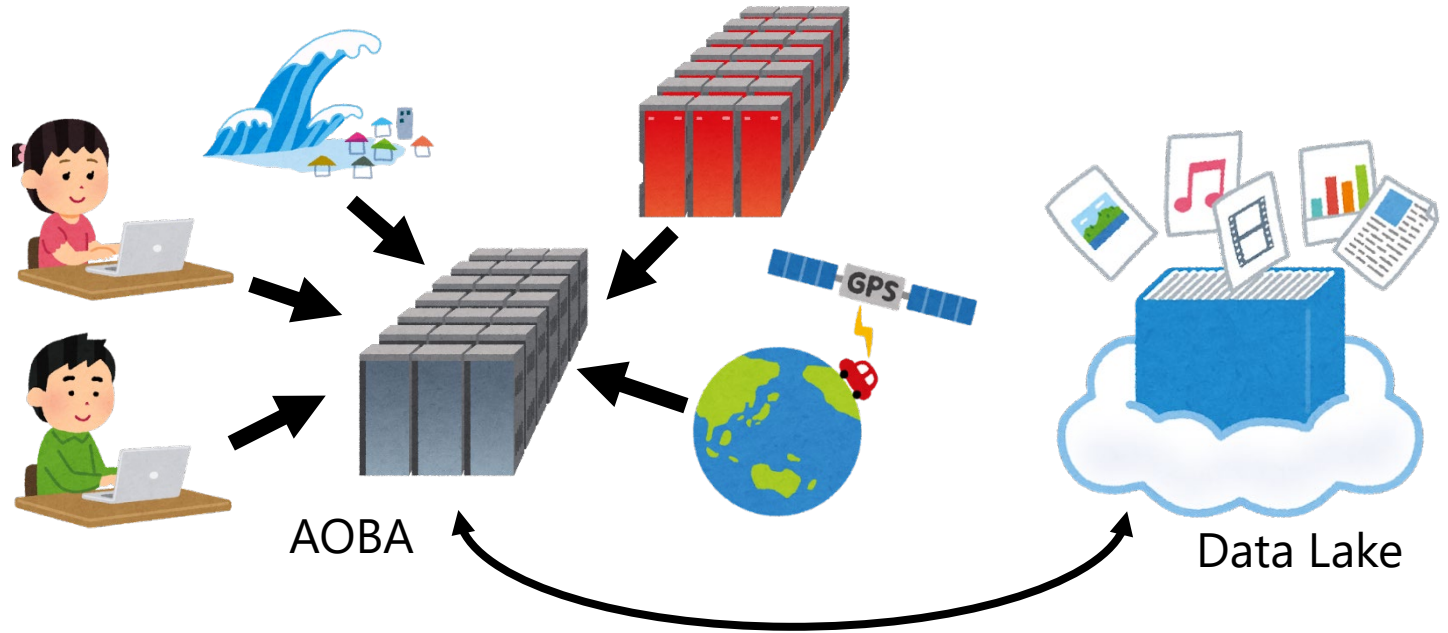
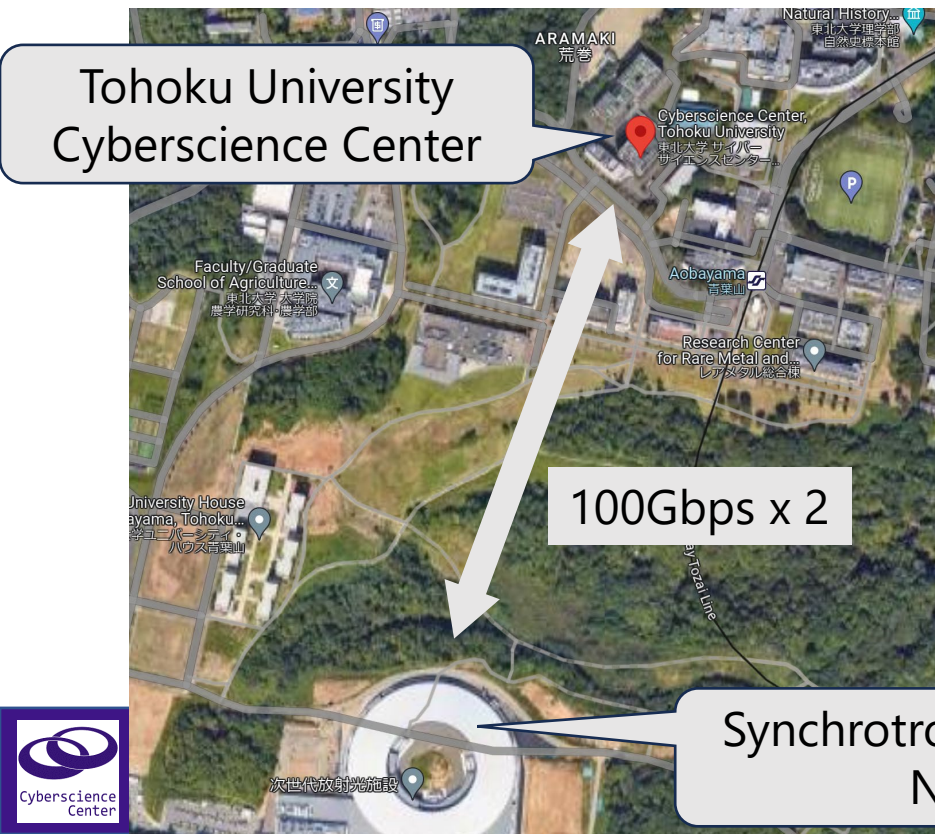
# Improving Weak Scalability

- **Scalability at increasing the problem size with the system size**
  - Ideally the performance per VE should be constant but actually not.
- **Optimizations**
  - Hybrid parallelization + MPI comm. optimization + Bank conflict avoiding



# AOBA for Big Data Analysis

- The most important metric at AOBA system design was memory bandwidth
  - The highest priority is given to efficient execution of numerical simulations
- **AOBA is now expected to work as an infrastructure for data analysis.**



# AOBA as a Cloud Storage for NT users

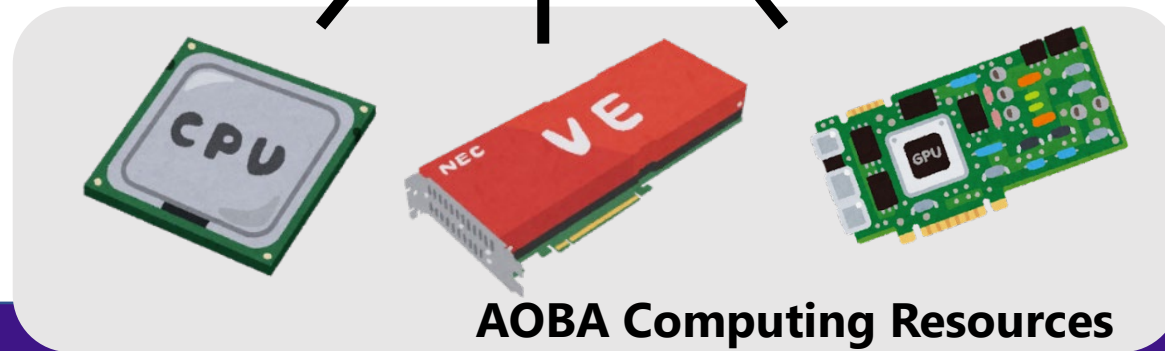
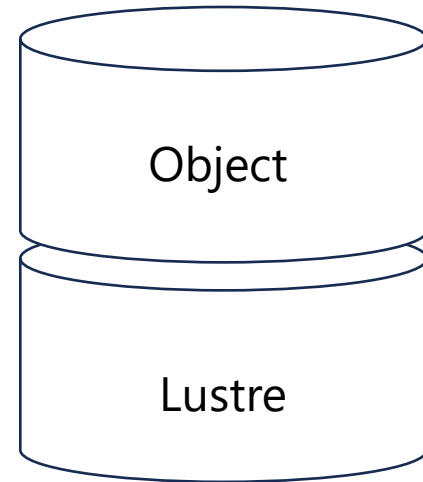
AOBA-S's storage is split into two.

Data sharing with other cloud storages and authorized users on the Internet.

NanoTerasu users can quickly and easily access data and resources with their web browsers



OPEN OnDemand



AOBA Computing Resources

NUG Society Meeting 35

# AOBA in 1<sup>st</sup> NanoTerasu Paper

- Top 1 at "Most read" ranking!

Open all abstracts

---

**OPEN ACCESS**

**Towards sub-10 nm spatial resolution by tender X-ray ptychographic coherent diffraction imaging**

Nozomu Ishiguro *et al* 2024 *Appl. Phys. Express* 17 052006

▼ Open abstract    View article    PDF

Applied Physics Express

**APPLIED**    SUPPORTS OPEN ACCESS

Applied Physics Express (APEX) is an open access letters journal dedicated solely to rapid dissemination of up-to-date reports on new findings in applied physics. The journal is characterized by high scientific quality and prompt publication.

Receive monthly Spotlight research from APEX

Article    Track my article

Sign up for new issue notifications

---

**13 days**    **2.3**    **5.6**  
Median submission to first decision after peer review    Impact factor    Citescore

[Full list of journal metrics](#)

**Most read**

- Latest articles
- Review articles
- Accepted manuscripts
- Trending
- Open Access
- Spotlights

---

Open all abstracts

**OPEN ACCESS**

**Towards sub-10 nm spatial resolution by tender X-ray ptychographic coherent diffraction imaging**

Nozomu Ishiguro *et al* 2024 *Appl. Phys. Express* 17 052006

▼ Open abstract    View article    PDF

---

**Defect engineering in SiC technology for high-voltage power devices**

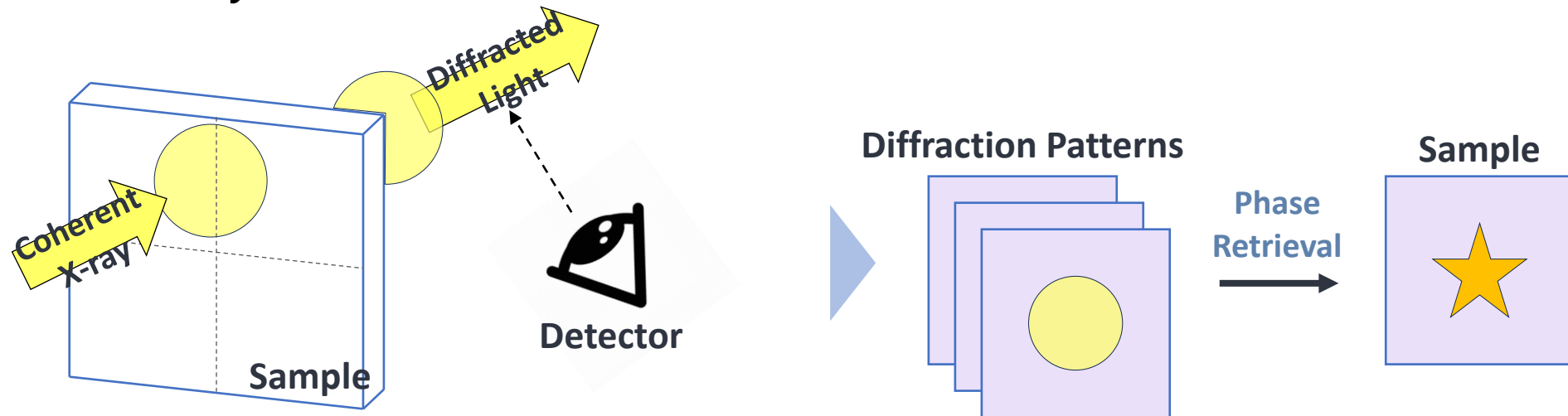
Tsunenobu Kimoto and Heiji Watanabe 2020 *Appl. Phys. Express* 13 120101

▼ Open abstract    View article    PDF



# X-ray Ptychographic Coherent Diffraction Imaging (PCDI)

- PCDI scans a sample using X-ray beam to obtain a diffraction pattern at each scan position.
- The sample image is then reconstructed from the diffraction patterns using phase retrieval algorithms such as the Extended Ptychographical Iterative Engine (ePIE).
- Phase retrieval algorithms are computationally intensive and generally bottlenecked by Fast Fourier Transform (FFT).



# Implementation of ePIE on the Vector Engine

- **Basic design**

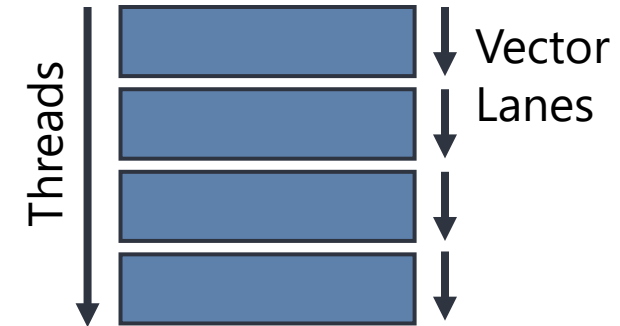
- Core computation is implemented in C and invoked from Python using NLCpy's (NumPy-like library for VE) JIT compilation feature.
- Anything else (such as I/O and memory management) is implemented in Python and NLCpy for productivity.

- **2D FFT optimized for small arrays**

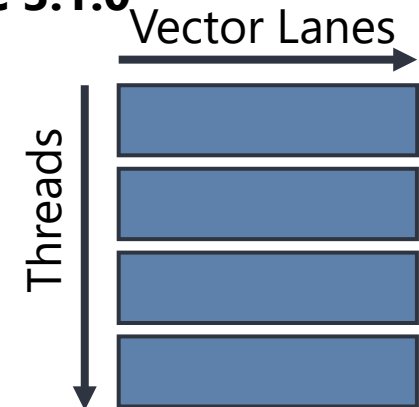
(thanks to **Arihiro Yoshida** at NEC)

- $NLC \leq 3.0.0$  parallelized and vectorized the same axis, leading to insufficient average vector length for the array sizes we are targeting.
- To improve the FFT performance for small arrays, NEC engineers added a new code path [1] in NLC 3.1.0 that parallelizes and vectorizes different axes.

**NLC 3.0.0**



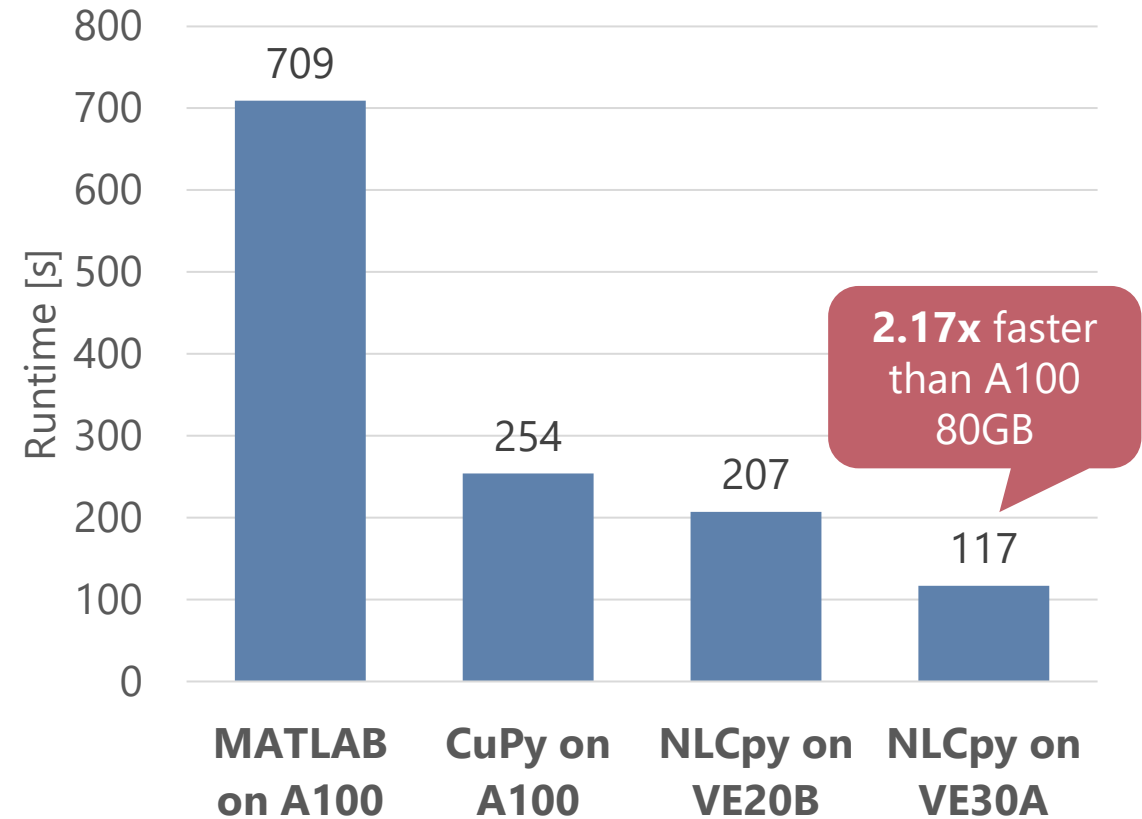
**NLC 3.1.0**



[1] P. Vizcaino et al., "Acceleration with long vector architectures: Implementation and evaluation of the FFT kernel on NEC SX-Aurora and RISC-V vector extension," CCPE, vol. 35, no. 20, 2024.

# Performance evaluation

- Ran ePIE on an intermediate-size dataset using three ePIE implementations:
  - Original MATLAB code
  - CuPy (NumPy-like library for GPU) port
  - NLCpy (NumPy-like library for VE) port [2]
- NLCpy on VE30A is fastest and achieves over 2x speedup over CuPy on A100 80GB PCIe.
- Scientific outcomes have been published in the very first academic paper from NanoTerasu [3].



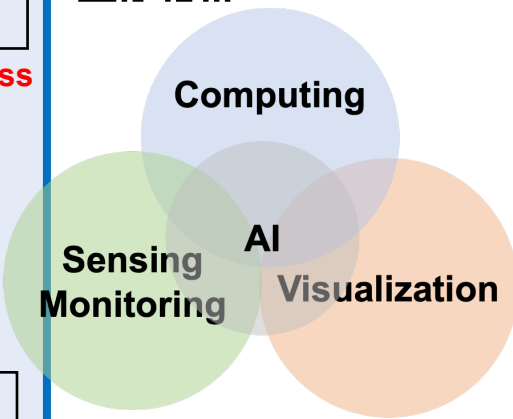
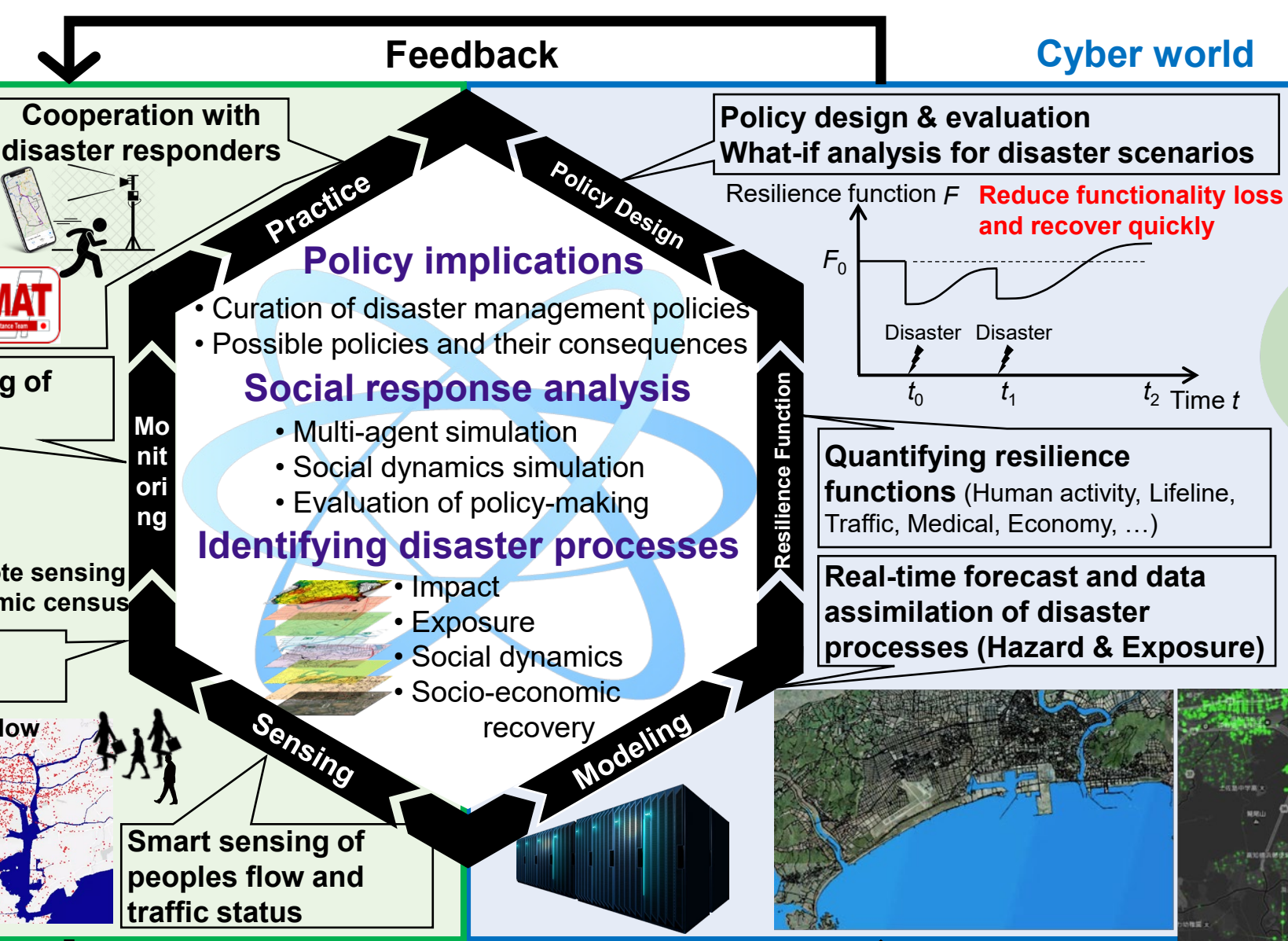
[2] <https://github.com/keichi/aoba-ptycho>

[3] N. Ishiguro et al., "Towards Sub-10 nm Spatial Resolution by Tender X-ray Ptychographic Coherent Diffraction Imaging," APEX, vol. 17, no. 5, 2024.

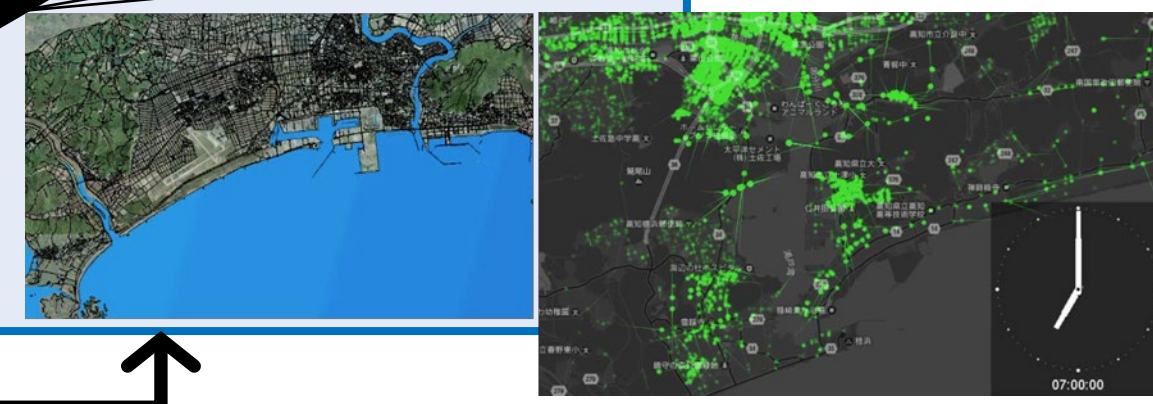


# Recent Research Activity





TsunamiCast by RTi-cast, Inc.



Physical world

Copy (Twin)

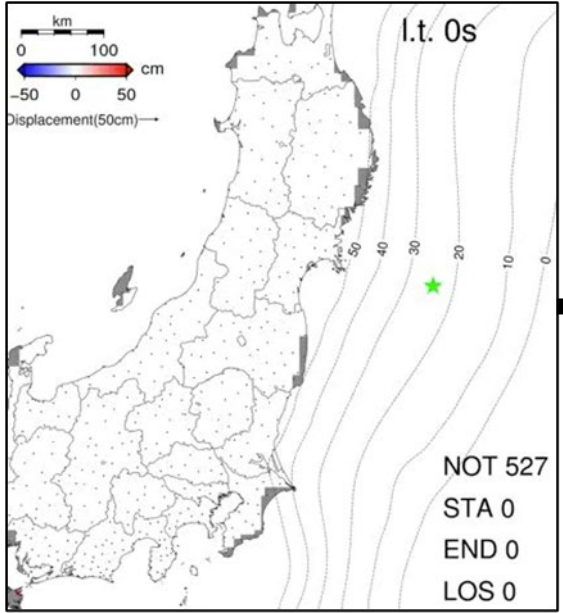
Cyber world



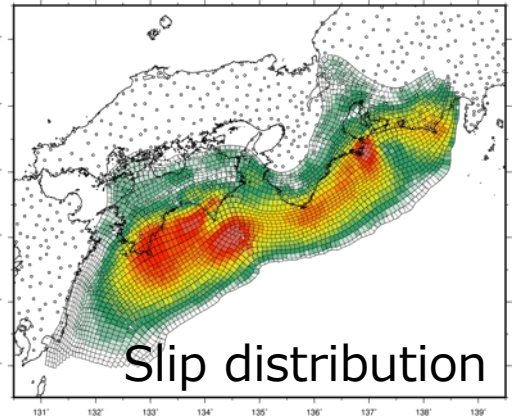
# Full-automatic real-time estimation of tsunami damage

Courtesy of Prof. Koshimura

>7min (present)  
>3min (target)



Real-time estimation of fault model from GEONET data (REGARD; GSI)



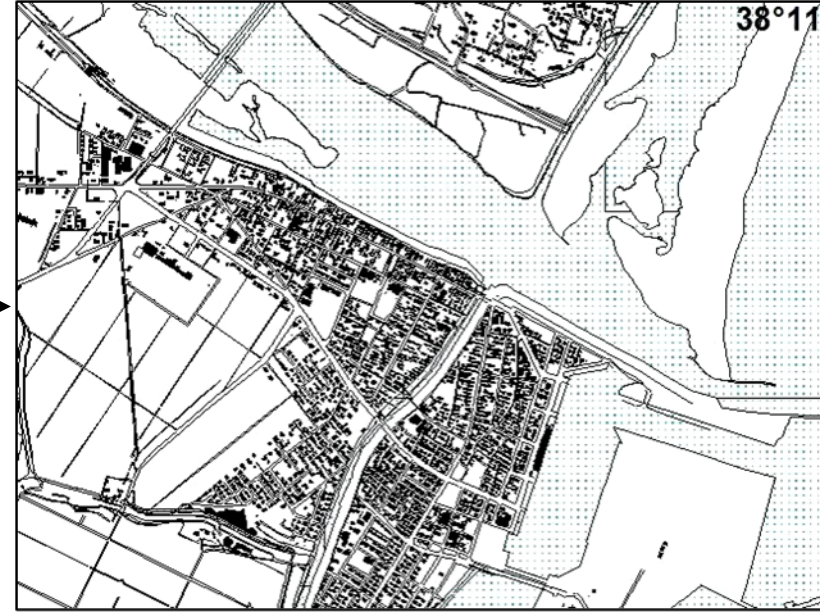
>10min (present)  
>1min (target)



Real-time simulation on Supercomputer AOBAs at Tohoku University with considering current costal facilities and tide height at the occurrence.



>10min (present)  
>1min (target)



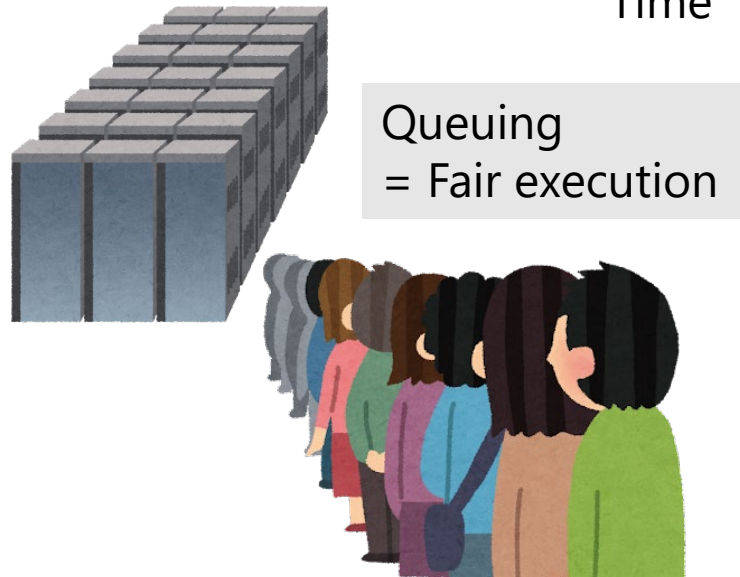
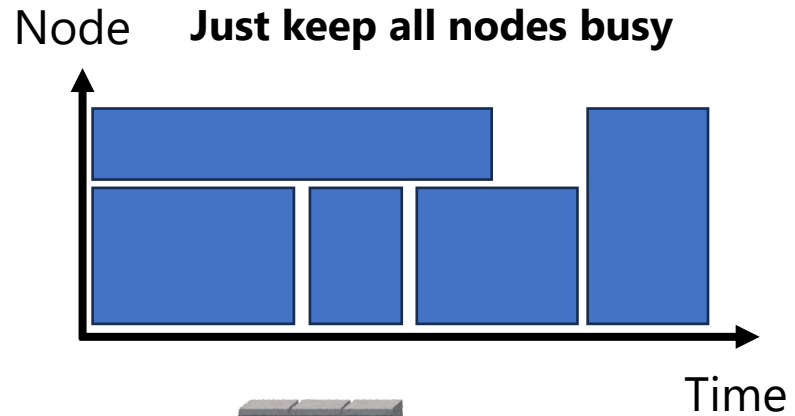
Quantitative estimation of damages



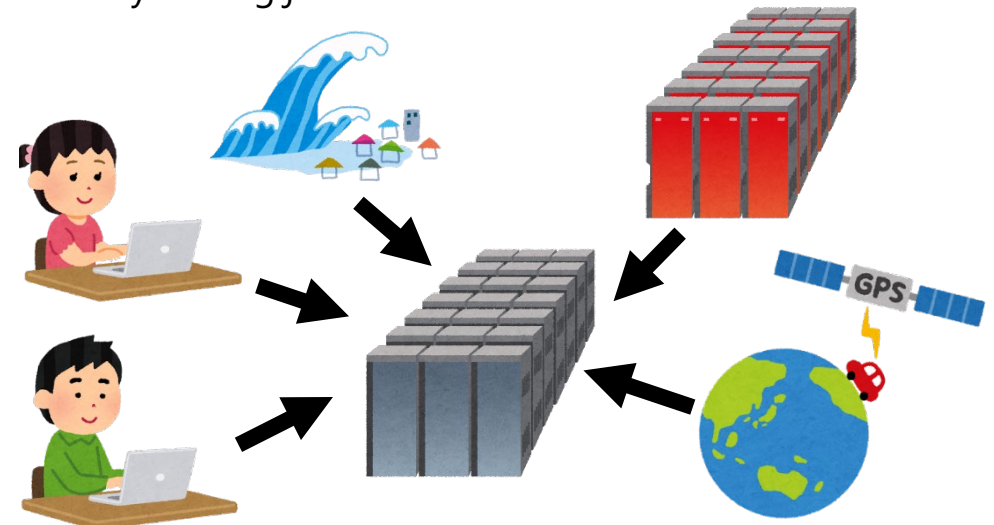
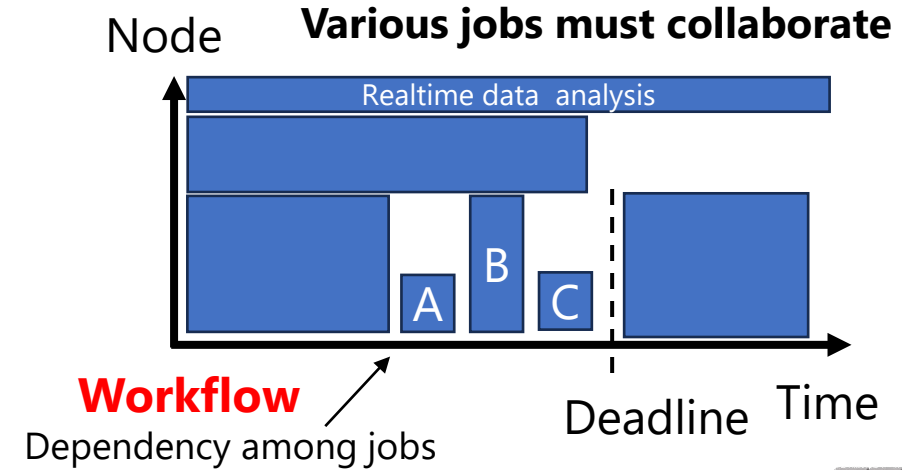
Successfully operated upon the earthquake in March 2023

# Workflow + Batch Job Scheduling

For academic use



External data sources





# Working very actively in this research field!

- Workshop on Job Scheduling Strategies for Parallel Processing (JSSPP'24)

## Session 1 [8:00 AM - 10:00 AM]

### Workshop Opening

- Dalibor Klusacek and Vaclav Chlumsky : *Real-life HPC Workload Trace Featuring Refined Job Runtime Estimates*
- Monish Soundar Raj, Thomas MacDougall, Di Zhang and Dong Dai : *An Empirical Study of Machine Learning-based Synthetic Job Trace Generation Methods*
- Hang Cui, Keichi Takahashi, Yoichi Shimomura and Hirovuki Takizawa : *Clustering Based Job Runtime Prediction for Backfilling Using Classification*
- Vanamala Venkataswamy : *Launchpad: Learning to Schedule Using Offline and Online RL Methods*

## Coffee Break [10:00 AM - 10:30 AM]

## Keynote Lecture [10:30 AM - 11:30 AM]

Walfredo Cirne (Google): *Managing Private Clouds*

Abstract: This talk defines the Private Cloud Management program and presents Flex, Google's solution for it. It covers how Flex makes Google's easier to operate, as well as the key techniques used to make it more efficient. It also discusses how we preserve governance by charging internal users the resources they prompted Google to buy, in spite of aggressive resource sharing and on-demand allocation.

Biography: Walfredo Cirne has worked on the many aspects of parallel scheduling and cluster management for the past 25 years. He is currently with the Technical Infrastructure Group at Google, where he leads Flex, Google's solution for resource management of its internal Cloud. Previously, he was faculty at the Universidade Federal de Campina Grande, where he led the OurGrid project. Walfredo holds a PhD in Computer Science from the University of California San Diego, and Bachelors and Masters from the Universidade Federal de Campina Grande.

## Lunch break [11:30 AM - 1:00 PM]

## Session 2 [1:00 PM - 3:00 PM]

- Arup Kumar Sarker, Aymen Al-Saadi, Niranda Perera, Mills Staylor, Gregor Von Laszewski, Matteo Turilli, Ozgur Ozan Kilic, Mikhail Titov, Andre Merzky, Shantenu Jha and Geoffrey Fox : *Radical-Cylon: A Heterogeneous Data Pipeline for Scientific Computing*
- Mohammad Samadi, Tiago Carvalho, Luis Miguel Pinho and Sara Royuela : *Evaluation of Heuristic Task-to-Thread Mapping Using Static and Dynamic Approaches*
- Roy Nissim, Oded Schwartz and Reut Shabo : *Challenges in parallel matrix chain multiplication*
- Daiki Nakai, Keichi Takahashi, Yoichi Shimomura and Hirovuki Takizawa : *A node selection method for on-demand job execution with considering deadline constraints*

## Coffee Break [3:00 PM - 3:30 PM]

## Session 3 [3:30 PM - 4:30 PM]

- Sho Ishii, Keichi Takahashi, Yoichi Shimomura and Hirovuki Takizawa : *Maximizing Energy Budget Utilization Based on Dynamic Power Cap Control*
- Luc Angelelli, Danilo Carastan-Santos and Pierre-Francois Dutot : *Run your HPC jobs in Eco-Mode: revealing the potential of user-assisted power capping in supercomputing systems*

## Workshop Closing

**3 out of 10 technical papers are from our research group!**

# Summary

- **AOBA-1.5 (2023.8-2028.3)**

- **10 months since AOBA-S started operation**

- AOBA-S can achieve excellent performance on memory-intensive workloads
  - Ranked at the 13<sup>th</sup> place in the HPCG benchmark due to the high efficiency of approx. 5.5%.

- **User Support Activities**

- FrontFlow/blue (FFB) Flow Solver
- X-ray Ptychographic Coherent Diffraction Imaging (ePIE)

- **Research Activities**

- Resource Management for Urgent Computing
  - We are working actively in this research field.