jh250027

Innovative Computational Science by Integration of Simulation/Data/Learning on Heterogeneous Supercomputers

Kengo Nakajima（The University of Tokyo, Japan）

This project advances computational science by integrating "Simulation/Data/Learning (S+D+L)" using heterogeneous supercomputers such as "Wisteria/BDEC-01 (U.Tokyo)" and "Miyabi (JCAHPC)", along with "mdx." In FY.2021 and FY.2022, our focus was earthquake simulation with real-time data assimilation. From FY.2023, we expand the "S+D+L" concept to other fields, initiating joint research and software development with international partners (Germany: JSC, FAU, etc., France: CEA, Croatia: RBI). The project's leading PI and colleagues have been working on the innovative software platform "h3-Open-BDEC since FY.2019 to facilitate "S+D+L" integration on heterogeneous systems. Wisteria/BDEC-01 ("Odyssey" for Simulations with A64FX, "Aquarius" for Data/ML/AI with NVIDIA A100) operates as a full-scale platform for "S+D+L" integration with h3-Open-BDEC, gaining global attention. In FY.2023 and FY.2024, we have conducted research in earth science, life science, and library/software/tool for integration of "S+D+L". In FY.2025, we add a new field (quantum sciences), and extend the idea of h3-Open-BDEC for QC (Quantum Computers)-HPC Hybrid Environment.

1. Basic Information

(1) Collaborating JHPCN Centers

Hokkaido University
Tohoku University
The University of Tokyo
Nagoya University
Kyushu University
mdx

(2) Theme Area

Large scale computational science area

(3) Project Members and Their Roles

①: Earth Science, ②Life Science, ③Quantum Science, ④Library/Software/Tool

U.Tokyo: Nakajima (Leading PI)①②③④, Hanawa④, Sumimoto①②③④, Shimokawabe④, Yamazaki①③④, Hu④, Furumura (Co-PI)①, Tsuruoka①, Nagao①, Itoh①, Ichimura①, Fujita①, Suzumura④, Hanai④, Kuno④

Hokkaido U.: Fukaya③

Nagoya U.: Hoshino④, Kawai①④, Ueno④

Kyushu U.: Ohshima④

U.Hyogo: Shiba②④

Osaka U.: Morohoshi④

RIKEN: Hascoet①, Sugita②, Yagi②, Ito②, Imamura③④, Nakao④, Dawson②③④, Tonoyama①

NIES: Yashiro①④

CliMTECH: Arakawa①②③④

CEA (France): Boillod-Cerneux (Co-PI)②, Badri①, Foerster①, Lomet①, Genovese②

JSC (Germany): Di Napoli (Co-PI)③, Caviedes-Voullieme①④, Koh②④, Suarez①②③④, Wu③, Conrads③④

FAU (Germany): Wellein④, Hager④, Afzal④, Ujeniya④, Gruber④

BUW (Germany): Yoda①

TUM (Germany): Chen①

U.Cologne (Germany): Kubicki④

RBI (Croatia): Davidovic③, Mijic③, Badrinarayanan③, Živković③

Fujitsu: Sakaguchi④, Kasai①④, Obinata④

Hitachi: Matsuba④

NVIDIA: Hamamura③, Furuya③

ORNL (USA): Teranishi④, Monil④

UC Riverside (USA): Chung④

2. Purpose and Significance of the Research

This project advances computational science by integrating "Simulation/Data/Learning (S+D+L)" using heterogeneous supercomputers such as "Wisteria/BDEC-01 (U.Tokyo)" and "Miyabi (JCAHPC)", along with "mdx." In FY.2021 and FY.2022, our focus was earthquake simulation with real-time data assimilation. From FY.2023, we expand the "S+D+L" concept to other fields, initiating joint research and software development with international partners (Germany: JSC, FAU, etc., France: CEA, Croatia: RBI). The project's leading PI and colleagues have been working on the innovative software platform "h3-Open-BDEC since FY.2019 to facilitate "S+D+L" integration on heterogeneous systems. Wisteria/BDEC-01 ("Odyssey" for Simulations with A64FX, "Aquarius" for Data/ML/AI with NVIDIA A100) operates as a full-scale platform for "S+D+L" integration with h3-Open-BDEC, gaining global attention. In FY.2023 and FY.2024, we have conducted research in earth science, life science, and library/software/tool for integration of "S+D+L". In FY.2025, we add a new field (quantum sciences), and extend the idea of h3-Open-BDEC for QC (Quantum Computers)-HPC Hybrid Environment.

3. Significance as JHPCN Joint Research Project

"Wisteria/BDEC-01" and "Miyabi" are based on heterogeneous architecture, can cover a wide range of workloads, and are optimum as platforms for integration of "S+D+L", and the collaboration with mdx provides a more flexible usage environment. This project is tightly collaborating with other JHPCN projects, such as jh250012, jh250013, jh250035, jh250046 and jh250055, which use Wisteria/BDEC-01, Miyabi and h3-Open-BDEC. Moreover, other universities in JHPCN have also introduced heterogeneous systems, which provide an opportunity for the deployment and validation of the developed software and applications in this project.

4. Outline of Research Achievements until FY.2024

In FY.2021 and FY.2022, we focused on earthquake simulation with real-time data assimilation. Since FY.2023, we expanded S+D+L to other fields, initiating joint research with international partners. Using the software platform h3-Open-BDEC, Wisteria/BDEC-01 has gained global attention. Recent research includes earth science, life science and system software for integration of S+D+L.

5. Details of FY.2025 Research Achievements

(a) Earth Science

(a-1) Weather/Climate [5,9,20,33,34]

We proposed an innovative computational method "ensemble coupling" (https://dl.acm.org/doi/10.1145/3712031.3712035), that combines low-resolution ensemble calculations with a high-resolution single one in global weather/climate simulations. While initial plan for FY.2025 was to conduct more detailed meteorological analysis using NICAM with 6-days, the proper data-set has not been prepared after preliminary evaluation. Instead, we evaluated a coupled atmospheric–AI workflow on modular supercomputers, integrating the NICAM global nonhydrostatic model with an AI-based cloud microphysics surrogate called NICE (NICAM Cloud Emulator). Using the h3-Open-SYS/WaitIO (WaitIO) communication library and the h3-Open-UTIL/MP (UTIL/MP) coupler in h3-Open-BDEC, the system enables asynchronous CPU–

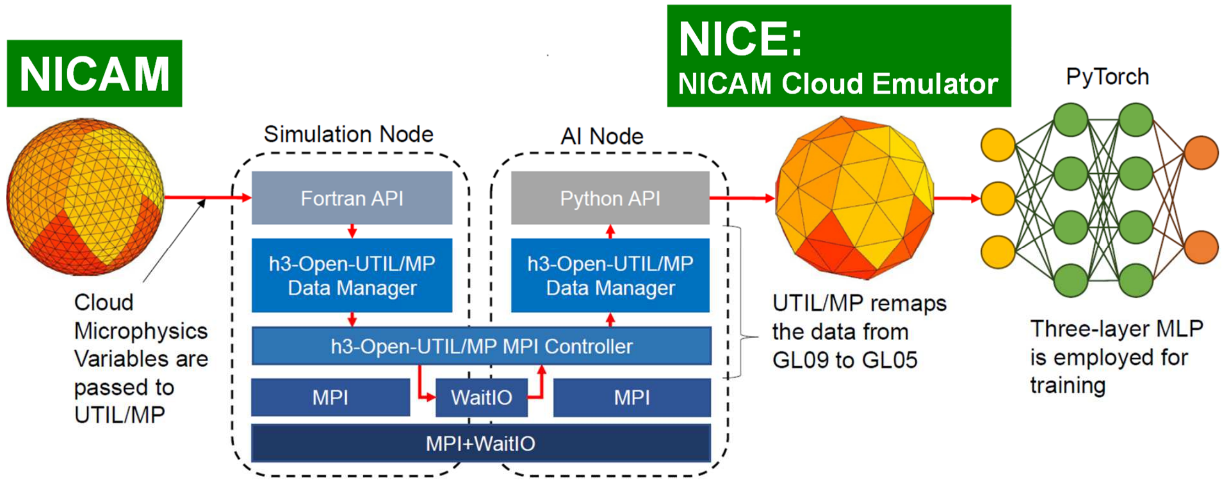

Fig.1 Data Flow of NICAM-AI (NICE) Coupling [9]

GPU data exchange across heterogeneous nodes. The AI model, a three-layer MLP trained with PyTorch, successfully reproduces key atmospheric variables, achieving correlation coefficients above 0.92 and demonstrating strong agreement in horizontal field predictions. In the previous work by us, agreement was not so good, as shown in upper graphs of Fig.2. In the present work [9], we added 5 more parameters for estimation of three parameters. Performance tests on the Wisteria/BDEC-01 and Miyabi systems show that execution efficiency depends heavily on resource balance: faster NICAM runs can cause AI idle time, while slower NICAM shifts the bottleneck. Training performance is approximately twice as fast on H100 GPUs on Miyabi compared to A100 GPUs on Aquarius and is limited mainly by memory and data-transfer bandwidth. The results demonstrate the feasibility of practical NICAM–AI coupling and highlight the importance of optimizing resource allocation and training parameters.

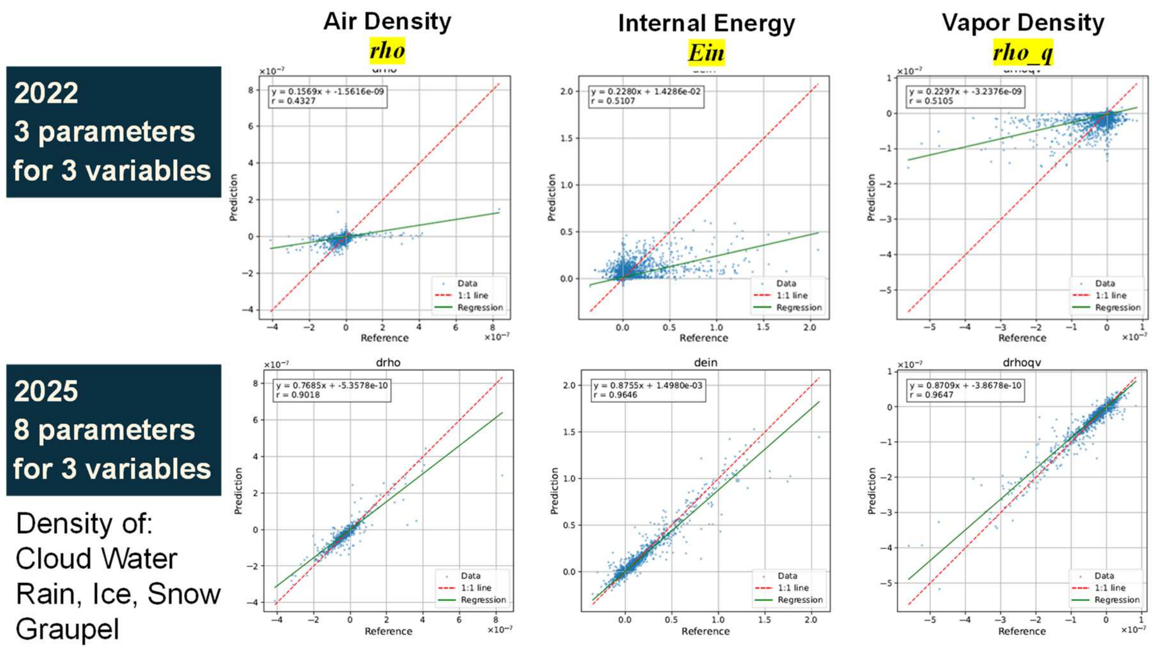

Fig.2 Correlation of Simulation vs. Prediction by AI (Upper: Previous work in HPC Asia 2022, Lower: Present Work) [9]

(a-2) Causality-Drive Earthquake Simulation [5,6,10,14,16,36,50]

We proposed a new causal inference approach designed to improve seismic data assimilation for large-scale earthquake simulations based on the finite element method (FEM). Traditional assimilation techniques rely on correlation-based coupling, which fails to capture the inherently directional nature of seismic wave propagation. To address this limitation, the authors introduce PCMat-HSIC, a causal discovery algorithm that identifies directional cause–effect relations among sensor time series using the Hilbert–Schmidt Independence Criterion (HSIC). Exact HSIC requires Gram matrix construction, leading to time complexity between $O(n^2)$ and $O(n^3)$ and memory complexity $O(n^2)$, which restricts its applicability to long time series. By contrast, the Random Fourier Feature (RFF)-based formulation reduces computational complexity to $O(nD)$, with $D \ll n$, enabling the analysis of substantially longer signals and higher-dimensional systems. The algorithm is further accelerated through GPU implementation using PyTorch, achieving speedups of up to 100× on an NVIDIA A100 on Wisteria/BDEC-01(Aquarius). Experiments show that PCMat-HSIC with RFF maintains high F1-scores while significantly reducing runtime compared to existing methods such as PCMCI-Parcorr and PCMCI-CMIknn. Future work in FY.2026 includes integrating RFF into the GPU pipeline and comparing with other assimilation strategies.

In FY.2025, we also started to enhance PSD code developed by CEA with data assimilation through optimal interpolation technique in Seism3D/OpenSWPC-DAF [5]. This enables PSD to perform 3D simulations with real-time data assimilation.

(a-3) 3D Underground Model [5,7,43]

We examined a CNN-based data assimilation approach to directly estimate subsurface structure parameters from seismic waveforms generated by the Hisada-code for a 1D model. Although this approach is conceptually reasonable due to the nonlinear waveform–parameter relationship, our previous attempts have not been successful. Direct CNN regression showed unstable performance because subsurface parameters have highly nonuniform identifiability from waveforms, especially under noisy conditions. Parameter-space errors alone were insufficient to evaluate waveform consistency, and even simplified settings failed

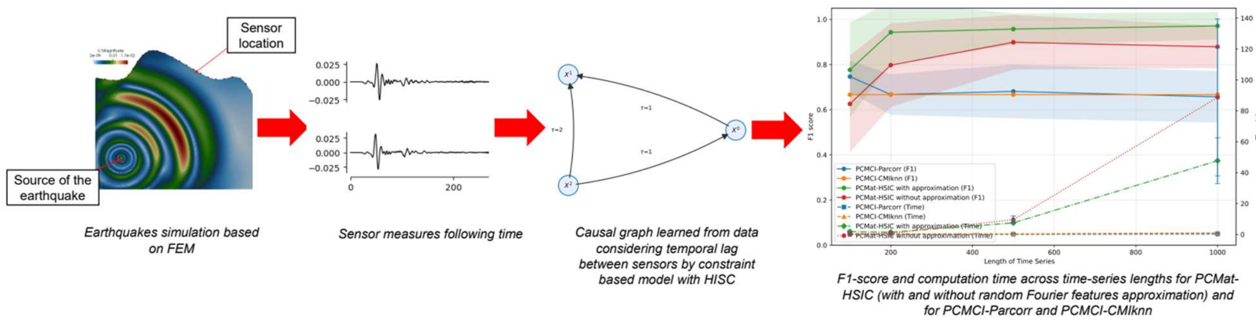

Fig.3 Causal Graph Learning and Scaling with Time-Series Length [10]
Earthquake simulation based on the finite element method generates time series at sensor locations. A causal graph is learned from data by considering temporal lag between sensors using a constraint-based model with the Hilbert–Schmidt Independence Criterion. The right panel reports F1-score and computation time across time-series lengths for PCMat-HSIC (with and without random Fourier features approximation) and for PCMCI-Parcorr and PCMCI-CMIknn..

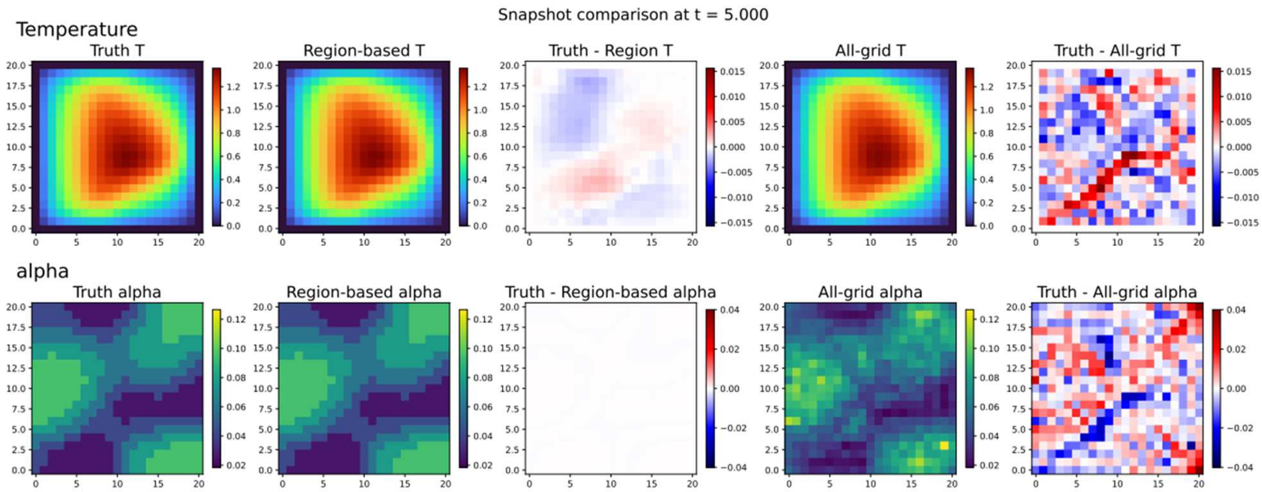

Fig.4 Simulation results for temperature prediction and α estimation using the region-based and all-grid methods

to provide reliable results. These findings indicate that CNN-based direct estimation is inadequate as a final assimilation method and should be considered only for limited or auxiliary purposes.

We also investigated the use of data assimilation to estimate subsurface structures through a simplified heat transfer model. A two-dimensional heat conduction simulation was conducted on a 20×20 grid with heterogeneous thermal diffusivity, serving as a toy model for seismic wave propagation. Temperature evolution over time was treated as observational data, and parameter estimation was performed using the Extended Kalman Filter. Two estimation approaches were compared: a region-based method, in which the spatial structure was assumed known and diffusivity was estimated per region, and an all-grid method, in which diffusivity was estimated independently at each grid point without prior structural information. The region-based estimation accurately reproduced both temperature evolution and thermal diffusivity values, while the all-grid estimation successfully reconstructed the spatial distribution of diffusivity (Fig.4). These results demonstrate that temperature diffusion observations contain sufficient information to constrain material properties. The study suggests that similar data assimilation techniques may be applied to estimate subsurface structures from observed seismic waveforms in future work.

(a-4) TSMP [5]

The initial goal for FY.2025 was to port and validate TSMP (Terrestrial System Modeling Platform, https://www.terrsysmp.org/), developed by JSC, on Wisteria/BDEC-01 using h3-Open-BDEC. However, we decided instead to port TSMP2, the new version of TSMP designed for the JUPITER system at JSC (https://www.fz-juelich.de/en/jsc/jupiter), to Miyabi. Although the porting of TSMP2 for JUPITER, which consists of ICON (atmosphere), ParFlow (groundwater), and eCLM (land surface flow), has not yet been completed, in FY.2026 we will conduct preliminary verification on Miyabi using TSMP2's original coupler, OASIS3. First, we will validate coupled execution of eCLM (CPU) and ParFlow (GPU) on Miyabi-G, followed by three-way coupling with ICON on Miyabi-C. The implementation of h3-Open-BDEC is scheduled for FY.2027.

(b) Life Science

(b-1) Brain Aneurysm Simulations [5]

In FY.2025, we conducted performance evaluation and optimization of the m-AIA computational fluid dynamics (CFD) code (https://m-aia-aia-m-aia-60defb05e8a289b7b7c07554d8d66f6c0373baacfa2d23b.pages.rwth-aachen.de/) and machine learning workloads on modern heterogeneous HPC systems. For CFD, a lattice Boltzmann (LB) solver was optimized on ARM based processors by tuning SIMD vectorization using ARM NEON and SVE intrinsics. Benchmarks on a circular pipe flow simulation with approximately 150 million grid cells show that runtime is strongly dependent on vector width and OpenMP parallelism, with up to 16 threads per task. Strong scaling tests were conducted on problems up to 347 million cells, while weak scaling experiments increased workload from 100 million cells on one node to 3.2 billion cells on 32 nodes, demonstrating good scalability on Miyabi-G with GH200 and JUWELS Booster at JSC system with A100 (Fig.5). In addition, distributed machine learning training using PyTorch was benchmarked on the JUWELS Booster with ResNet 50. Quantitative results show that PyTorch Distributed Data Parallel (DDP) achieves higher parallel efficiency and data throughput than Horovod, making it the preferred approach for large scale ML workloads on multi GPU clusters.

(b-2) Big-DFT with GENESIS for SARS-CoV-2 Main Protease [5,51]

Predicting how to address mutations that affect the binding of a ligand to a target enzyme remains an important challenge in drug discovery. In this work, we propose to use the heterogeneous computational capabilities of the

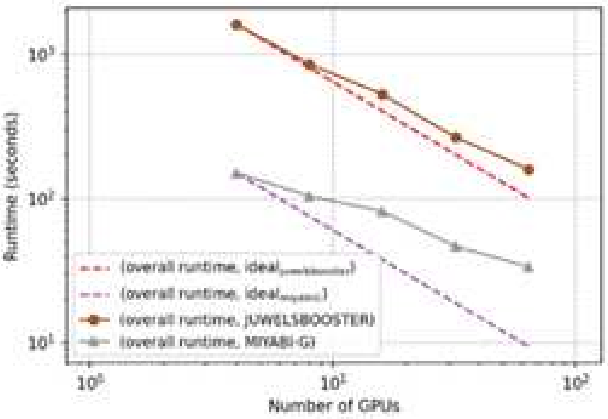

Fig.5 Runtime obtained by a strong scaling test on the code m-AIA

Wisteria/BDEC-0 computer to perform simulations coupling multiple theoretical methods to shed light on mutations that affect the binding of the Darnavir (DRV) medicine to the main protease (PR) of HIV-1. In particular, we utilize the Aquarius GPU partition to perform classical molecular dynamics simulations; as samples are generated through the classical trajectory, Density Functional Theory calculations are performed in-operando on the Odyssey CPU partition to gain insight into binding interactions. The coupled simulation produces results with the statistical fidelity available to classical methods and the level of detail made possible through quantum mechanical simulations with minimal impact on the overall time to solution. In Fig. 6, we show an example of the data generated through this coupled workflow. The distribution of interactions over snapshots — defined in terms of the quantum mechanical bond order quasi-observable — between DRV, PR, the solution (WAT), individual amino acids, and DRV fragments (APC, ANL, TOL) are calculated and sorted by median interaction. This analysis reveals the key interaction signatures at multiple granularities. We compare three variants of the HIV-1 protease: 6OPS (Wild

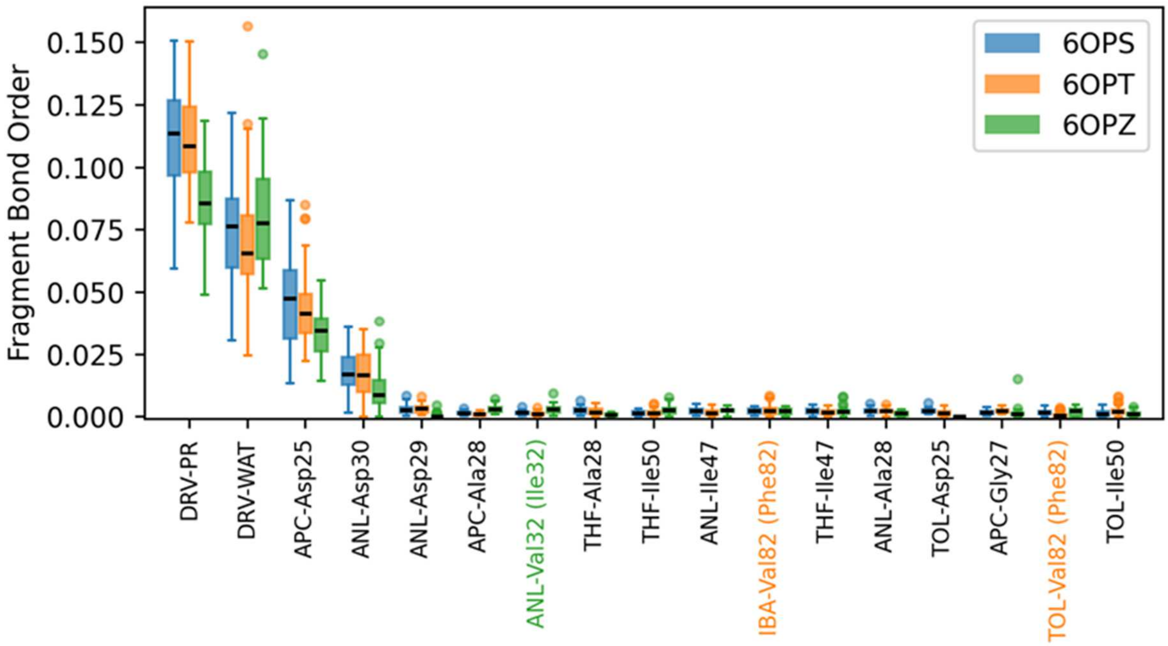

Fig.6 FBO interactions between amino acid and DRV fragments sorted by median strength. Residues mutated in 6OPT and 6OPZ are written in orange and green, respectively [51]

Type), 6OPT (2 Mutations), and 6OPZ (11 Mutations) to see the signature of affinity loss. In the 11 mutation variant, a clear signal of weakening is revealed that is distributed across multiple residues.

(c) Quantum Sciences

(c-1) ChASE [3,5]

In FY.2025, we extended the Chebyshev Accelerated Subspace iteration Eigensolver (ChASE) to solve the full Bethe–Salpeter Equation (BSE) in its pseudo Hermitian form, targeting interior eigenvalues corresponding to the smallest positive excitations. Since this problem departs from standard Hermitian and extremal eigenvalue settings, several algorithmic components are redesigned. Interior eigenvalues are addressed by applying Chebyshev filtering to the squared operator H^2, effectively transforming them into extremal eigenvalues, while spectral bounds are estimated using a pseudo Hermitian Lanczos method. The solver further exploits the positive–negative symmetry of the BSE to reduce computational cost and replaces the Rayleigh–Ritz step with an oblique projection to preserve quadratic convergence.

A GPU enabled implementation is developed and integrated into the Yambo code. Strong scaling experiments on NVIDIA H100 GPUs for a Hamiltonian of size 100,000 show good scalability when computing 1–3% of the eigenpairs, demonstrating both robust convergence and efficient multi GPU performance.

(c-2) QC-HPC Hybrid Computing [5,8,46]

In [8], we prosed a prototype implementation and benchmark of Quantum Selected Configuration Interaction (QSCI) on a heterogeneous hybrid platform integrating quantum simulation and high performance computing (HPC). In this framework, GPU nodes emulate quantum processors using CUDA-Q by NVIDIA to prepare and sample reference states via the Variational Quantum

Eigensolver (VQE), while CPU nodes perform classical optimization and large scale subspace diagonalization using the EigenExa eigensolver.

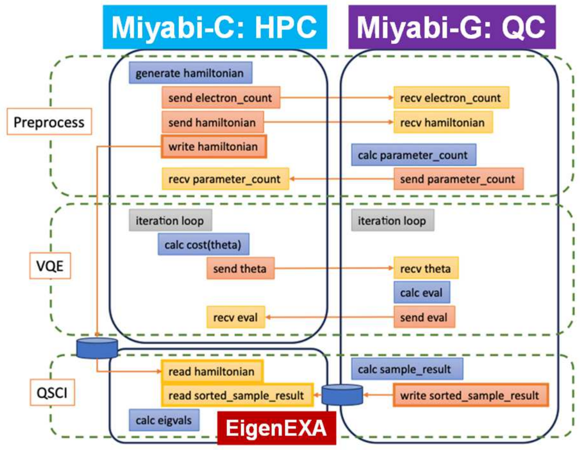

Fig.7 Hybrid CPU–GPU collaborative workflow across the Preprocess, VQE, and QSCI stages [8]

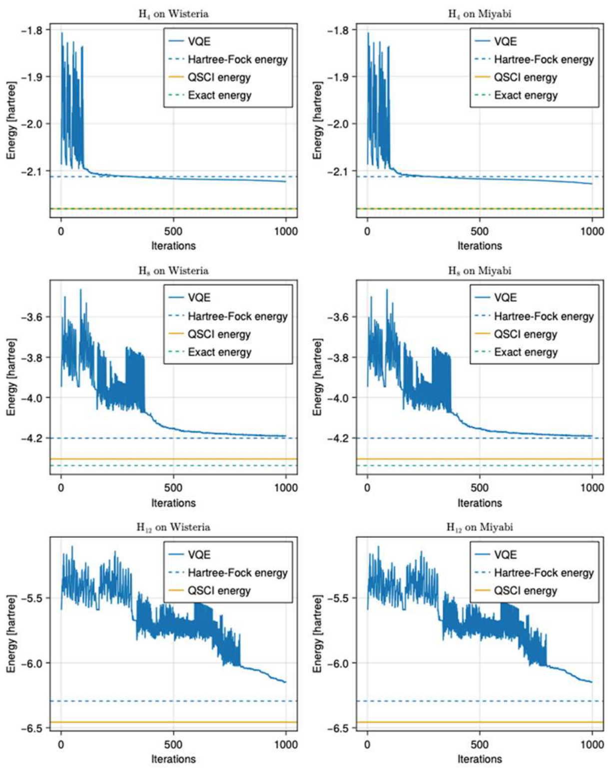

Fig.8 VQE and QSCI results for H4, H8, and H12 on Wisteria/BDEC-01 and Miyabi [8]

The system is built on modular supercomputers, Wisteria/BDEC 01 and Miyabi, and coordinated through the h3-Open-BDEC software stack with the WaitIO communication library. Benchmark calculations on linear hydrogen chains (H_4, H_8, and H_{12}) demonstrate that QSCI can achieve ground state energies close to exact values while handling subspaces containing up to tens of thousands of determinants. Performance analysis shows that computational cost is dominated by quantum state simulation, whereas communication and diagonalization overheads remain manageable (Fig.8). The results validate the feasibility of scalable hybrid HPC–Quantum workflows and highlight their potential for future large scale quantum chemistry applications.

Furthermore, we investigated a quantum–HPC hybrid computing framework implemented on heterogeneous supercomputers, focusing on improving the efficiency and accuracy of Quantum Selected Configuration Interaction (QSCI) through parallelization of the Variational Quantum Eigensolver (VQE) in [46]. GPU nodes are used as quantum computer simulators via CUDA-Q, while CPU nodes perform classical optimization and large scale diagonalization using EigenExa, with inter node communication enabled by the h3-Open-BDEC framework. The authors demonstrate that running multiple VQE instances in parallel with different initial parameters significantly improves QSCI accuracy. For an H_8 benchmark system, increasing VQE parallelism from 1 to 20 raises the number of discovered configurations from 961 to 2,182, enlarging the subspace dimension by more than 2.3×. As a result, the QSCI ground state energy systematically approaches the exact value, even though the total number of samples remains fixed. These results show that parallel VQE enhances solution quality by diversifying sampled reference states rather than by increasing sampling cost, highlighting an effective strategy for scalable quantum–HPC hybrid workflows.

(d) Library/Software/Tool for Integration of "S+D+L"

(d-1) h3-Open-BDEC [4,5,42,47]

h3-Open-BDEC was originally developed for integration of heterogeneous node groups in systems, such as Wisteria/BDEC-01. In FY.2025, enhancements were made to h3-Open-SYS/WaitIO (WaitIO, heterogeneous communication library) and h3-Open-UTIL/MP (multi-physics coupler), including WaitIO-Router for cross-supercomputer communication, WaitIO/MPI-Hybrid for high-performance workloads, and WaitIO4Julia for Julia integration. Additionally, We started to integrate WaitIO with the IRIS runtime (https://iris-programming.github.io/) by ORNL on Miyabi-G/Miyabi-C for dynamic task scheduling across heterogeneous architectures.

(d-2) Web-based Framework for QC-HPC Hybrid Computing using mdx [5,48]

We introduced application-level telemetry for quantum-centric supercomputing (QCSC) workflows and formalized it through a workflow metrics pyramid, emphasizing observability beyond traditional system-centric metrics. The framework extends monitoring to domain-level insights that support algorithm design and scientific interpretation. Building on this concept, the paper proposes an observability architecture that decouples computation from monitoring, enabling persistent telemetry collection, structured storage, and post-hoc analysis without disrupting primary workload execution. The approach is validated through a reference implementation on the Miyabi supercomputer and IBM Quantum systems using mdx platform, demonstrating its effectiveness in analyzing solver behavior and performance within a closed-loop sample-based quantum diagonalization (SQD) workflow. Persistent storage of telemetry data allows new metrics to be introduced without re-running experiments, supporting iterative research. While the current implementation focuses on mid- to high-level telemetry (L2–L4), the framework is extensible to lower-level metrics and other hybrid quantum–classical algorithms, offering a path toward infrastructure-aware algorithm development and performance optimization.

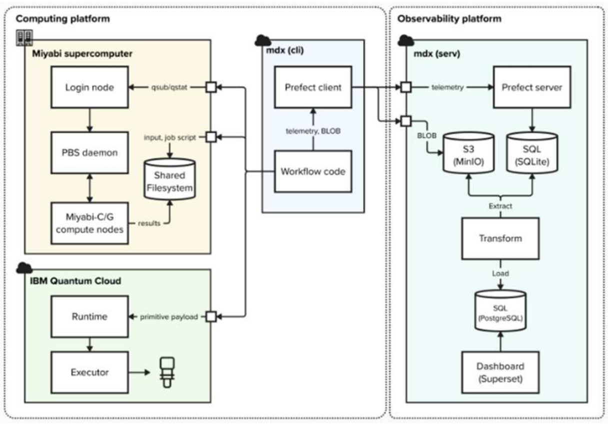

Fig.9 Component diagram of QCSC observability architecture implemented on the Miyabi supercomputer and mdx platform [48]

6. Self-review of Current Progress and Future Prospects

This is an international collaboration with 10 topics by more than 60 researchers from 18 organizations of 4 countries. In FY.2023 and FY.2024, we have conducted research in earth science, life science, and library/software/tool for integration of "S+D+L". In FY.2025, we added a new field (quantum sciences), and extend the idea of h3-Open-BDEC for QC-HPC Hybrid Environment. Since FY.2021, we have been continuously holding monthly online meetings with domestic researchers. We also held online meetings with overseas groups every 3-4 months, and held one or two in-person meetings during international conferences in FY.2025. The most important principle in collaborative research is continuous communication. By faithfully adhering to this principle, we have been able to achieve significant results in FY.2025. Generally, our activities have been highly recognized internationally, and we had 18 invited talks in FY.2025 [11-16, 21-32]. **Minister of MEXT Award for Science and Technology has been awarded to core members of this project [55].**

Since November 2023, h3-Open-BDEC has been utilized as software for constructing QC-HPC Hybrid environment in JHPC-quantum project (https://jhpc-quantum.org/). Works in (c-2) and (d-2) in the previous section are part of the JHPC-quantum project.-

We are also passionately promoting the idea of integration of "S+D+L". We regularly conduct tutorials on h3-Open-BDEC and on QC-HPC Hybrid using h3-Open-BDEC [52,55]. In FY.2025, we were invited to the 8th INFIERI Summer School in Pisa, Italy, September 1-13, 2025 (Intelligent Signal Processing for FrontIEr Research and Industry, https://indico.cern.ch/event/1441933/) and provided tutorials on h3-Open-BDEC and QC-HPC Hybrid Computing [53].

In FY.2025, research and development activities and dissemination of results were carried out actively across all areas as described above, and almost all of the planned objectives were achieved. The level of achievement for each category is as follows:

- Earth Science: 80%
- Life Science: 100%
- Quantum Science: 100%
- Library / Software / Tool: 100%

The achievement level for Earth Science is evaluated somewhat conservatively due to the inability to carry out the full-scale Ensemble Coupling simulations and delays in porting TSMP to the JUPITER system, which prevented substantial progress. On the other hand, a long-standing issue since 2022—the improvement of the NICAM-AI model—was successfully addressed, and results were presented in SCA/HPCAsia 2026 [9]. This outcome was not originally planned. **Taking these circumstances into account, the overall achievement level for FY.2025 is assessed at 92%.** Objectives that could not be achieved in FY.2025 have been postponed to FY.2026. However, Ensemble Coupling will be conducted under a separate project in FY.2026, and therefore it is not planned to be carried out within the continuation project of this research.

In 2025, RIKEN announced its plan for Fugaku's successor (FugakuNEXT), which will start its operation after 2029 and will adopt Fujitsu's Monaka-X with NVIDIA GPUs. The NVIDIA GH200 installed in Miyabi-G adopts NVLink-C2C, enabling efficient utilization of CPU and GPU, and similar technology is expected to be adopted in FugakuNEXT. In FY.2026, with FugakuNEXT in mind, we will make further promotion of Integration of (S+D+L) while developing applications in earth sciences, life sciences, and quantum sciences that efficiently coordinate CPU-to-GPU interactions on Miyabi-G.