Integration of Simulation/Data/Learning and Beyond

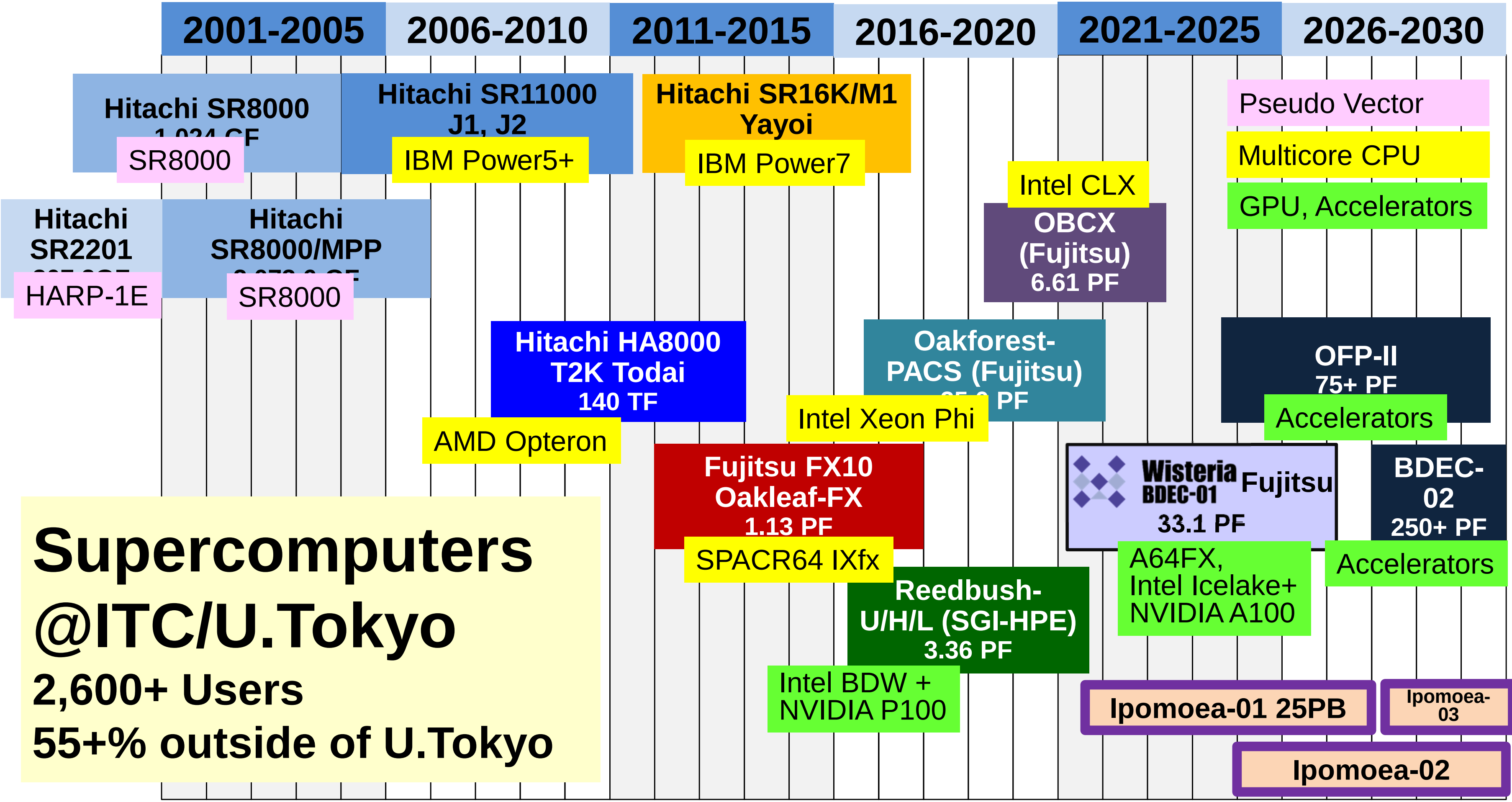
Kengo Nakajima
Information Technology Center
The University of Tokyo
RIKEN R-CCS

16th World Congress on Computational Mechanics & 4th Pan American Congress on Computational Mechanics (WCCM-PANACM Vancouver 2024)
Vancouver, B.C., Canada, July 23, 2024

Acknowledgements

- JSPS Grant-in-Aid for Scientific Research (S) (19H05662)
- New Energy & Industrial Technology Development Organization (NEDO): Cross-ministerial Strategic Innovation Promotion Program (SIP): Big-Data and AI-Enabled Cyberspace Technologies
- Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (JHPCN)
  - jh210022-MDH, jh220029, jh230017, jh230018, jh240029
- Information Technology Center, The University of Tokyo

- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - h3-Open-BDEC
- Applications on Wisteria/BDEC-01 with h3-Open-BDEC


2001-2005
2006-2010
2011-2015
2016-2020
2021-2025
2026-2030
Hitachi SR8000
SR8000
Hitachi SR11000 J1, J2
IBM Power5+
Hitachi SR16K/M1 Yayoi
IBM Power7
Pseudo Vector
Multicore CPU
GPU, Accelerators
Intel CLX
OBCX (Fujitsu) 6.61 PF
Hitachi SR2201
HARP-1E
Hitachi SR8000/MPP
SR8000
Hitachi HA8000 T2K Todai 140 TF
AMD Opteron
Oakforest-PACS (Fujitsu)
PF
Intel Xeon Phi
OFP-II
75+ PF
Accelerators
Fujitsu FX10 Oakleaf-FX 1.13 PF
SPACR64 IXfx
Wisteria BDEC-01
Fujitsu
33.1 PF
A64FX, Intel Icelake+ NVIDIA A100
BDEC-02 250+ PF
Accelerators
Reedbush-U/H/L (SGI-HPE) 3.36 PF
Intel BDW + NVIDIA P100
Ipomoea-01 25PB
Ipomoea-03
Ipomoea-02
Supercomputers @ITC/U.Tokyo
2,600+ Users
55+% outside of U.Tokyo

Integration of (S+D+L) has been our main strategy in recent 10 years

- Various Types of Workloads
  - Computational Science & Engineering: Simulations
  - Big Data Analytics +AI, Machine Learning …
- **Integration of (Simulation+Data+ Learning) (S+D+L) is important towards Society 5.0, Human-Centered Society proposed by Japanese Gov.**
  - **By Integration of Cyber & Physical Space**
- **BDEC (Big Data & Extreme Computing)**
  - **Platform for Integration of (S+D+L)**
  - **Focusing on S (Simulation)**
    - **AI for HPC, (Classical) AI for Science**
  - **Planning started in 2015**

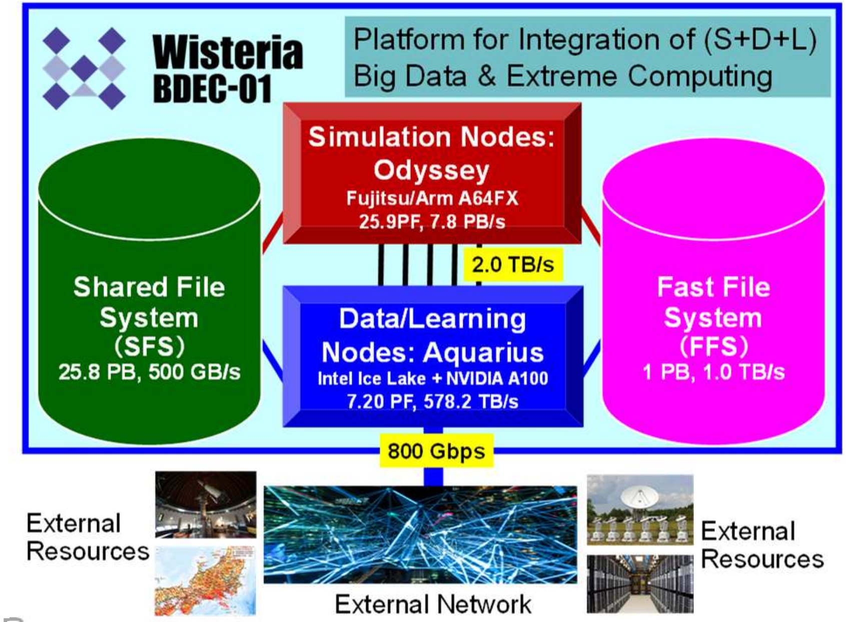
BDEC (Big Data & Extreme Computing)

S + D + L

Wisteria/BDEC-01

The 1st BDEC System
(Big Data & Extreme Computing)
HW Platform for Integration of (S+D+L)

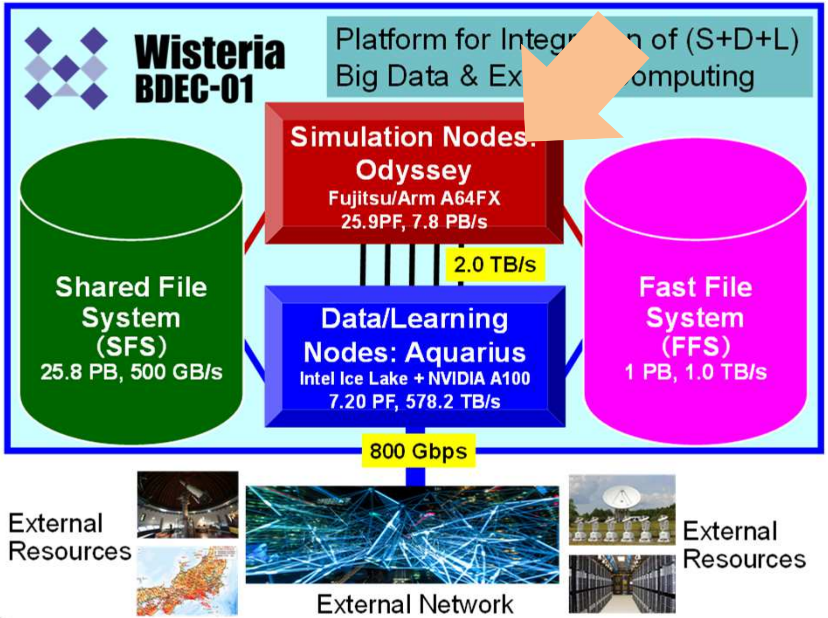
- Operation started on May 14, 2021
- 33.1 PF, 8.38 PB/sec by **Fujitsu**
  - ~4.5 MVA with Cooling, ~360m^2
- 2 Types of Node Groups
  - Hierarchical, Hybrid, Heterogeneous (h3)
  - Simulation Node Group: Odyssey
    - Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF
      - 7,680 nodes (368,640 cores), Tofu-D
      - General Purpose CPU + HBM
      - Commercial Version of “Fugaku”
  - Data/Learning Node Group: Aquarius
    - Data Analytics & AI/Machine Learning
    - Intel Xeon Ice Lake + NVIDIA A100, 7.2PF
      - 45 nodes (90x Ice Lake, 360x A100), IB-HDR
    - DL nodes are connected to external resources directly
- File Systems: SFS (Shared/Large) + FFS (Fast/Small)



Wisteria/BDEC-01

The 1st BDEC System (Big Data & Extreme Computing) HW Platform for Integration of (S+D+L)

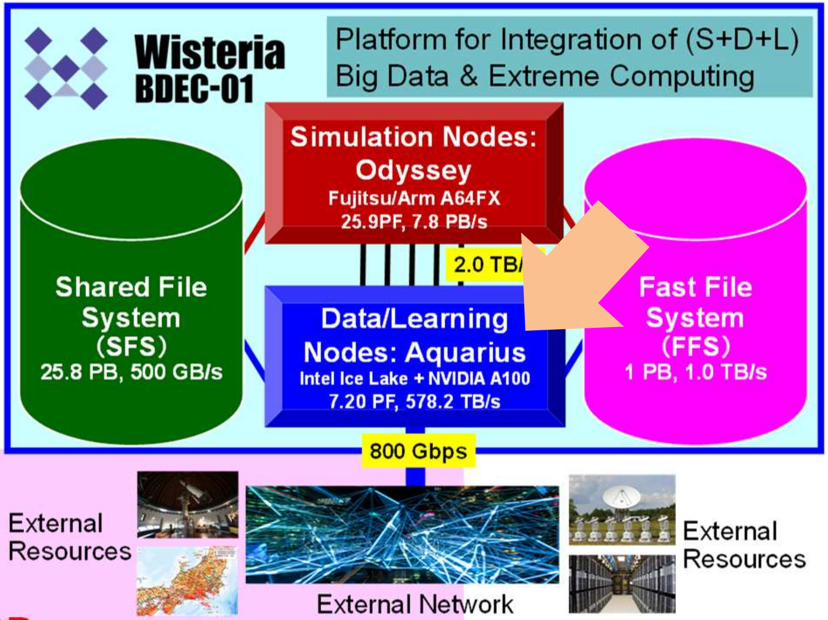
- Operation started on May 14, 2021
- 33.1 PF, 8.38 PB/sec by **Fujitsu**
  - ~4.5 MVA with Cooling, ~360m^2
- **2 Types of Node Groups**
  - Hierarchical, Hybrid, Heterogeneous (h3)
  - **Simulation Node Group: Odyssey**
    - **Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF**
      - **7,680 nodes (368,640 cores), Tofu-D**
      - **General Purpose CPU + HBM**
      - **Commercial Version of "Fugaku"**
  - Data/Learning Node Group: Aquarius
    - Data Analytics & AI/Machine Learning
    - Intel Xeon Ice Lake + NVIDIA A100, 7.2PF
      - 45 nodes (90x Ice Lake, 360x A100), IB-HDR
    - DL nodes are connected to external resources directly
- File Systems: SFS (Shared/Large) + FFS (Fast/Small)

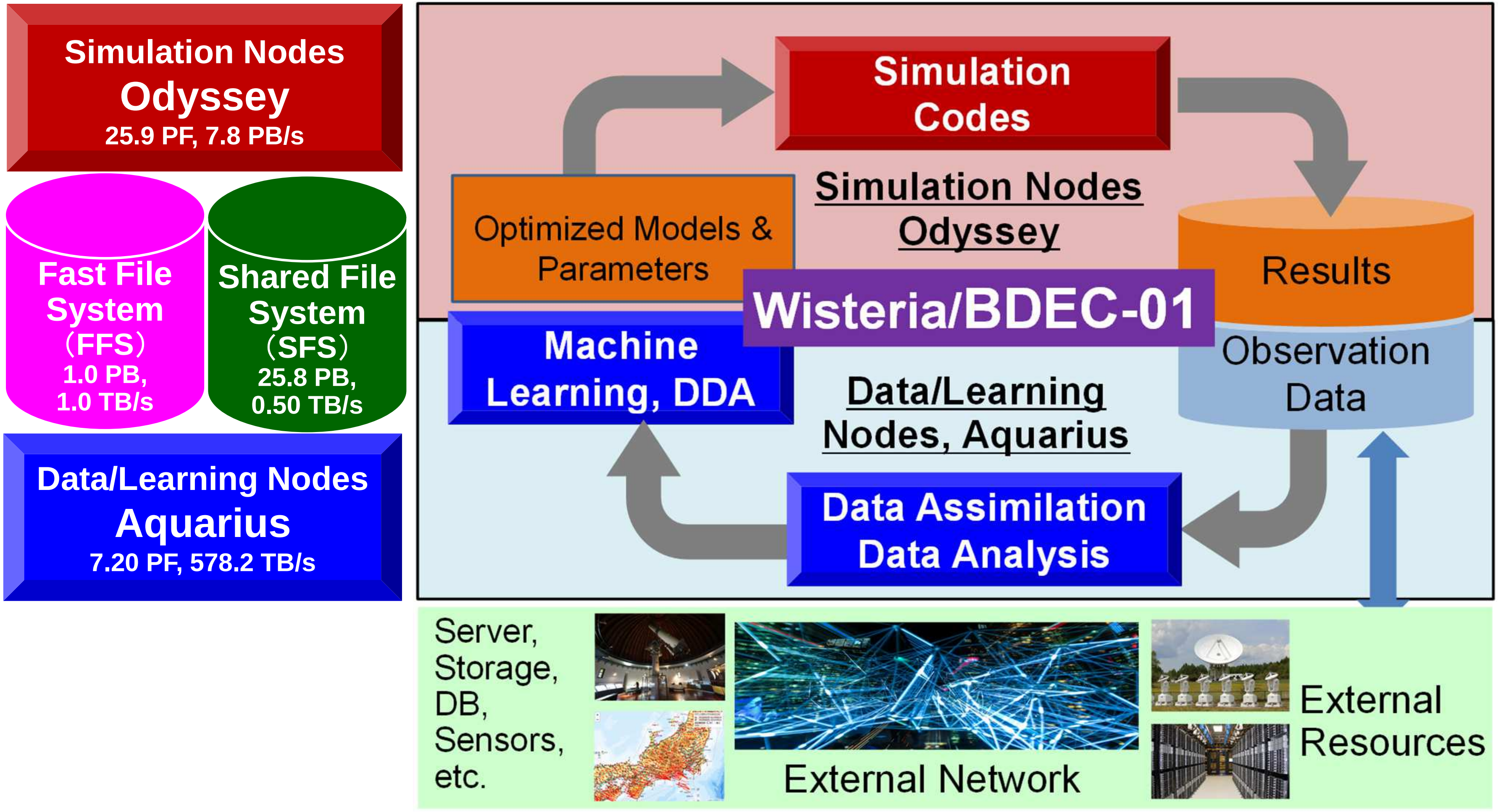


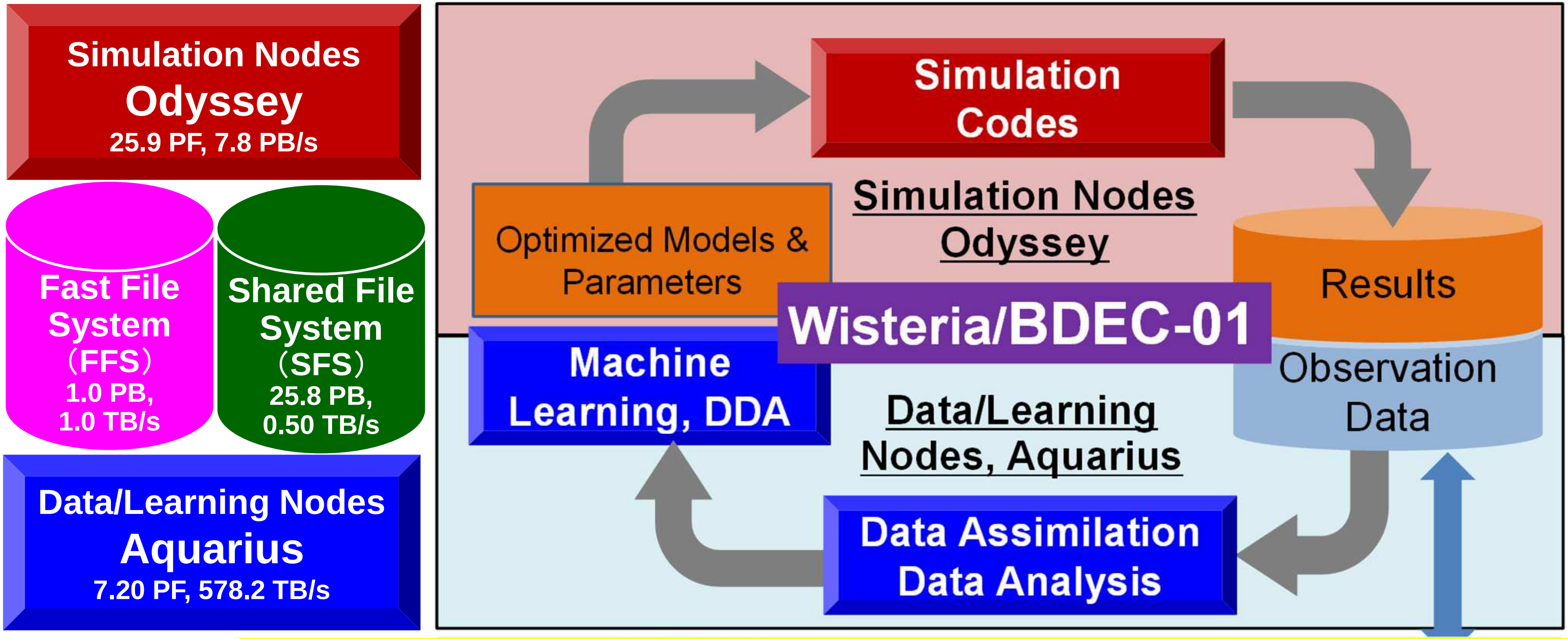
Wisteria/BDEC-01

The 1st BDEC System
(Big Data & Extreme Computing)
HW Platform for Integration of (S+D+L)

- Operation started on May 14, 2021
- 33.1 PF, 8.38 PB/sec by **Fujitsu**
  - ~4.5 MVA with Cooling, ~360m^2
- **2 Types of Node Groups**
  - Hierarchical, Hybrid, Heterogeneous (h3)
  - **Simulation Node Group: Odyssey**
    - **Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF**
      - **7,680 nodes (368,640 cores), Tofu-D**
      - **General Purpose CPU + HBM**
      - **Commercial Version of "Fugaku"**
  - **Data/Learning Node Group: Aquarius**
    - **Data Analytics & AI/Machine Learning**
    - **Intel Xeon Ice Lake + NVIDIA A100, 7.2PF**
      - **45 nodes (90x Ice Lake, 360x A100), IB-HDR**
    - **DL nodes are connected to external resources directly**
- File Systems: SFS (Shared/Large) + FFS (Fast/Small)



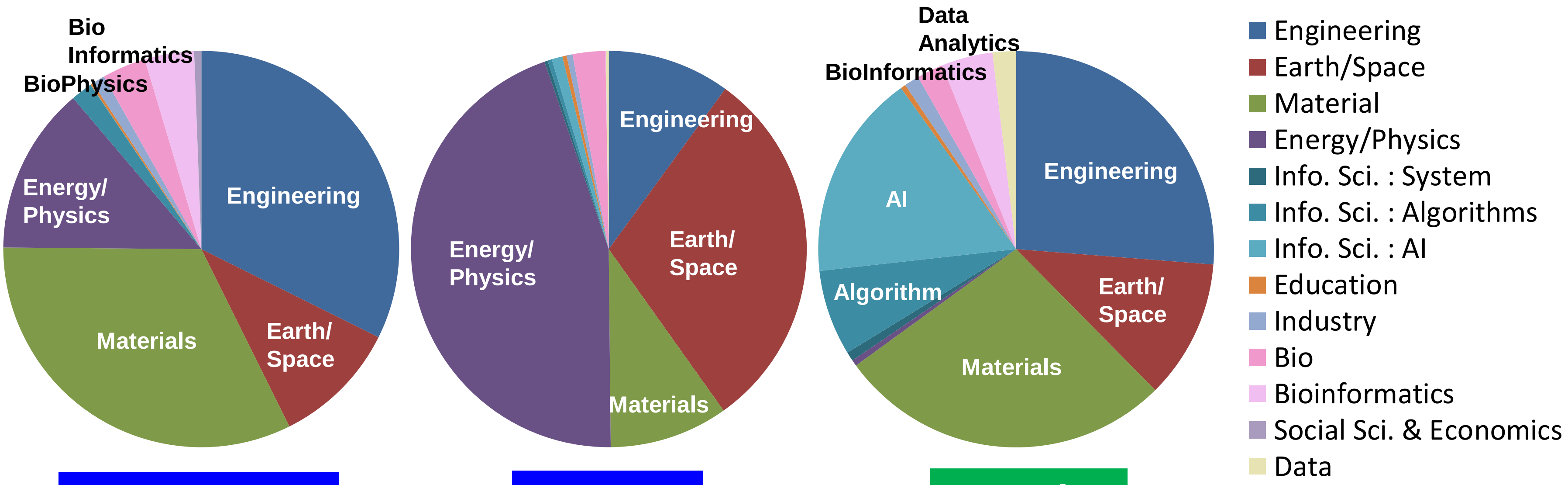

Simulation Nodes
Odyssey
25.9 PF, 7.8 PB/s
Fast File System (FFS)
1.0 PB,
1.0 TB/s
Shared File System (SFS)
25.8 PB,
0.50 TB/s
Data/Learning Nodes
Aquarius
7.20 PF, 578.2 TB/s
Simulation Codes
Optimized Models & Parameters
Simulation Nodes
Odyssey
Results
Wisteria/BDEC-01
Machine Learning, DDA
Observation Data
Data/Learning Nodes, Aquarius
Data Assimilation
Data Analysis
Server,
Storage,
DB,
Sensors,
etc.
External Network
External Resources



Optimization of Models/Parameters for Simulations by Data Analytics & Machine Learning (S+D+L)

Research Area based on Machine Hours (FY.2022)

CPU, GPU

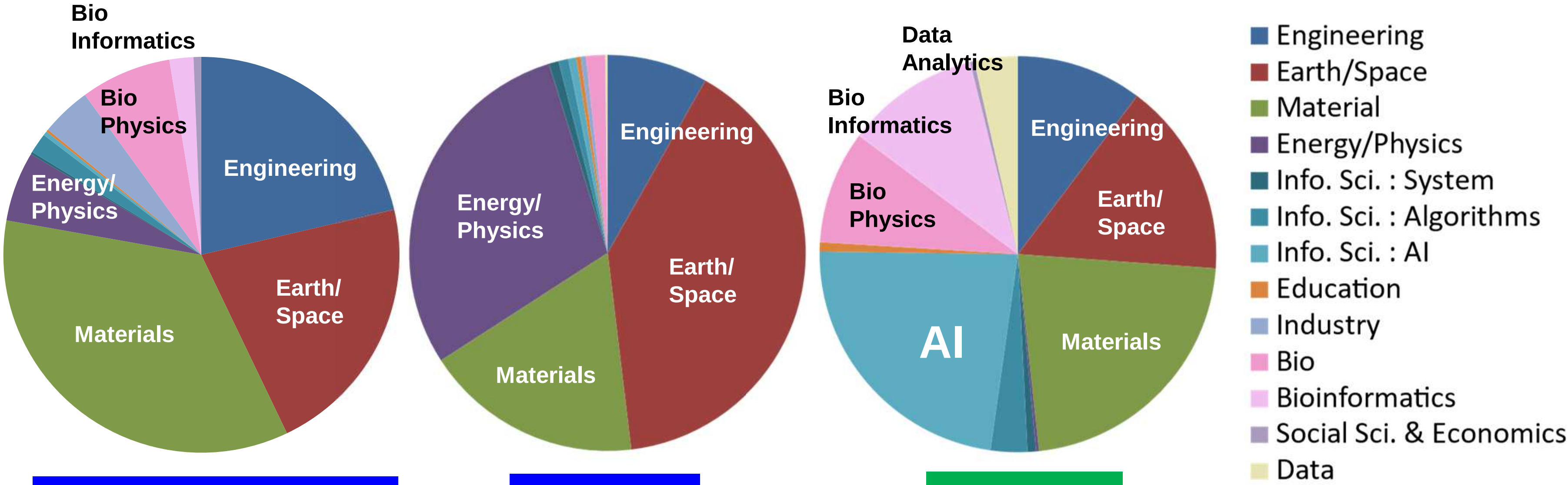


OBCX CascadeLake

Odyssey A64FX

Aquarius A100

Research Area based on Machine Hours (FY.2023) CPU, GPU (April-March)



OBCX CascadeLake
Retired in the end of September 2023

Odyssey A64FX

Aquarius A100

63rd TOP500 List (May, 2024)

R_{max}: Performance of Linpack (TFLOPS)
R_{peak}: Peak Performance (TFLOPS), Power: kW

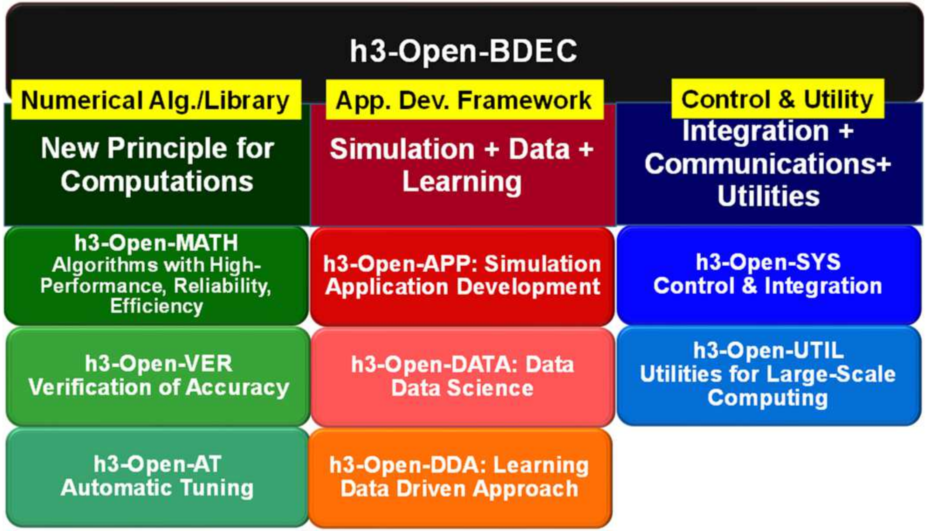
http://www.top500.org/

	Site	Computer/Year Vendor	Cores	R_{max} (PFLOPS)	R_{peak} (PFLOPS)	GFLOPS/ W	Power (kW)
1	**Frontier, 2022, USA** DOE/SC/Oak Ridge National Laboratory	HPE Cray EX235a, AMD Optimized 3rd Gen. EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11	8,699,904	1,194.00 (=1.194 EF)	1,679.82 71.1 %	52.93	22,703
2	**Aurora, 2023, USA** DOE/SC/Argonne National Laboratory	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel	9,264,128	1,012.00	1,980.01 51.1 %	26.15	24,687
3	**Eagle, 2023, USA** Microsoft	Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR	1,123,200	561.20	846.84 66.3 %		
4	**Fugaku, 2020, Japan** R-CCS, RIKEN	Fujitsu PRIMEHPC FX1000, Fujitsu A64FX 48C 2.2GHz, Tofu-D	7,630,848	442.01	537.21 82.3 %	14.78	29,899
5	**LUMI, 2022, Finland** EuroHPC/CSC	HPE Cray EX235a, AMD Optimized 3rd Gen. EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11	2,752,703	379.70	531.51 71.4 %	53.43	7,107
6	**Alps, 2024, Switzerland** Swiss National Supercomputing Centre (CSCS)	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11	1,305,600	270.00	353.75 76.3 %	51.98	7,107
7	**Leonard, 2022, Italy** EuroHPC/Cineca	BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64GB, Quad-rail NVIDIA HDR100	1,824,768	241.20	306.31 78.7 %	32.19	7,494
8	**MareNostrum 5 ACC, 2023, Spain** EuroHPC/BSC	BullSequana XH3000, Xeon Platinum 8460Y+ 40C 2.3GHz, NVIDIA H100 64GB, Infiniband NDR200, EVIDEN	663,040	175.30	249.44 70.3 %	42.15	4,159
9	**Summit, 2018, USA** DOE/SC/Oak Ridge National Laboratory	IBM Power System AC922, IBM POWER9 22C 3.07GHz, NVIDIA Volta GV100, Dual-rail Mellanox EDR InfiniBand	2,414,592	148.60	200.79 74.0 %	14.72	10,096
10	**Eos NVIDIA DGX SuperPOD** NVIDIA Corporation	NVIDIA DGX H100, Xeon Platinum 8480C 56C 3.8GHz, NVIDIA H100, Infiniband NDR400, Nvidia	485,888	121.40	188.65 64.4 %		
11	**Venado, 2024, USA** DOE/NNSA/LANL	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11	481,440	98.51	130.44 75.5 %	59.29	1,662
31	**TSUBAME 4.0, 2024, Japan** Tokyo Institute of Technology	HPE Cray XD665, AMD EPYC 9654 96C 2.4GHz, NVIDIA H100 SXM5 94 GB, Infiniband NDR200	172,800	25.46	59.40 42.9 %	34.78	732
39	**ABCI 2.0, 2021, Japan** AIST	Fujitsu PRIMERGY GX2570 M6, Xeon Platinum 8360Y 36C 2.4GHz, NVIDIA A100 SXM4 40 GB, InfiniBand HDR	504,000	22.21	54.34 40.9 %	13.88	1,600
40	**Wisteria/BDEC-01 (Odyssey), 2021, Japan** ITC, University of Tokyo	Fujitsu PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu interconnect D	368,640	22.12	25.95 85.2 %	15.07	1,468

- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - **h3-Open-BDEC**
- Applications on Wisteria/BDEC-01 with h3-Open-BDEC

h3-Open-BDEC: Innovative Software Platform for Integration of (S+D+L) on the BDEC System, such as Wisteria/BDEC-01

- 5-year project supported by Japanese Government (JSPS) since 2019
  - FY.2023 is the final year
    - Until the end of March 2024
- Leading-PI: Kengo Nakajima (The University of Tokyo)
- Total Budget: 1.41M USD



Members (Co-PI’s) of h3-Open-BDEC Project

Computer Science, Computational Science, Numerical Algorithms, Data Science, Machine Learning

- Kengo Nakajima (ITC/U.Tokyo, RIKEN), Leading-PI
- Takeshi Iwashita (Hokkaido U), Co-PI, Algorithms
- Hisashi Yashiro (NIES), Co-PI, Coupling, Utility
- Hiromichi Nagao (ERI/U.Tokyo), Co-PI, Data Assimilation
- Takashi Shimokawabe (ITC/U.Tokyo), Co-PI, ML/hDDA
- Takeshi Ogita (Waseda U.), Co-PI, Accuracy Verification
- Takahiro Katagiri (Nagoya U), Co-PI, Appropriate Computing
- Hiroya Matsuba (ITC/U.Tokyo, Hitachi), Co-PI, Container

Contributors/Collaborators

- **Information Technology Center, The University of Tokyo**
  - S. Sumimoto, T. Arakawa
  - T. Suzumura, M. Hanai
  - T. Hanawa
- **Earthquake Research Institute, The University of Tokyo**
  - T. Furumura, H. Tsuruoka
  - T. Ichimura, K. Fujita, S. Ito
- **Tokyo Institute of Technology**
  - R. Yokota, R. Sakamoto
- **University of Hyogo**
  - **H. Shiba (Former PD)**
- **Hokkaido University**
  - T. Fukaya
- **Nagoya University**
  - T. Hoshino, **M.Kawai (Former PD)**
- **Kyushu University**
  - S. Oshima, K. Inoue
- **RIKEN R-CCS**
  - M. Nakao, T. Imamura
- **Fujitsu**
  - Y. Sakaguchi, Y. Kasai, D. Obinata
- **My Former Students in U.Tokyo**
  - Y.C. Chen (KIT), R. Yoda (BWU)
  - A.T. Magro (Aitia)

(Part of) International Collaborators

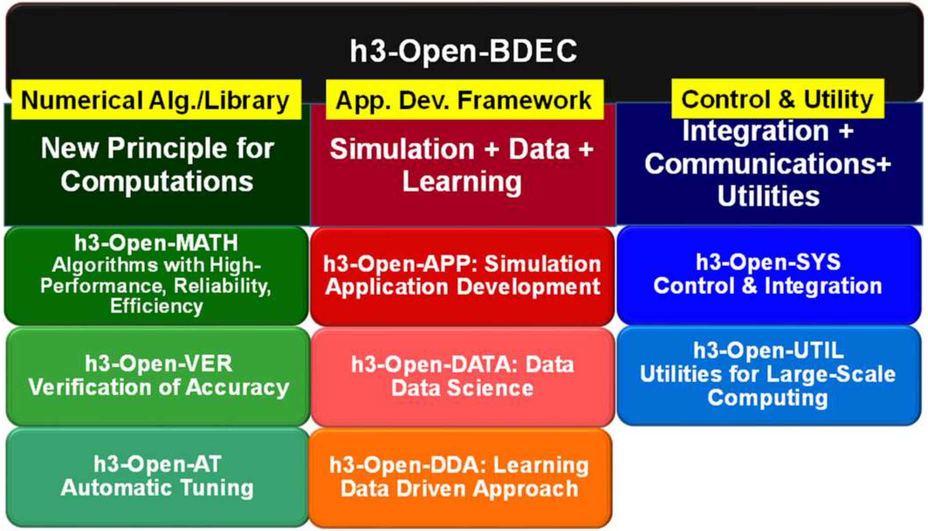
- Osni Marques (Lawrence Berkeley National Laboratory, USA)
- Richard Vuduc (Georgia Institute of Technology, USA)
- Edmond Chow (Georgia Institute of Technology, USA)
- Weichung Wang (National Taiwan University, Taiwan)
- Feng-Nan Hwang (National Central University, Taiwan)
- Gerhard Wellein (FAU Erlangen & Nuremberg, Germany)
- Matthias Bolten (University of Wuppertal, Germany)
- Serge Petiton (University of Liles/CNRS, France)
- Xing Cai (Simula Research Laboratory, Norway)
- Estela Suarez (Jülich Supercomputing Cetner/Univ. Bonn, Germany)
- Edoardo Di Napoli (Jülich Supercomputing Cetner, Germany)
- France Boilllod-Cerneux (CEA, France)

Final Goal stated in the Proposal of h3-Open-BDEC (Nov. 2018)

- We aim to reduce the amount of computations and power consumption **by more than 10 times** while maintaining the same accuracy as conventional methods in multi-level simulations that integrate (S+D+L).
  - Mixed Precision/Adaptive Precision
  - Machine Learning, Hierarchical Data Driven Approach
  - Heterogeneous Computing

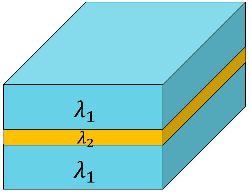
h3-Open-BDEC Innovative Software Platform for Integration of (S+D+L) on the BDEC System, such as Wisteria/BDEC-01

- "Three" Innovations
  - New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
  - Integration of (S+D+L) by Hierarchical Data Driven Approach (*h*DDA)
  - Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01



Adaptive Precision Computing with FP21/FP42

Masatoshi Kawai (kawai@cc.u-tokyo.ac.jp)



Heat Conduction with Heterogeneous Material Property

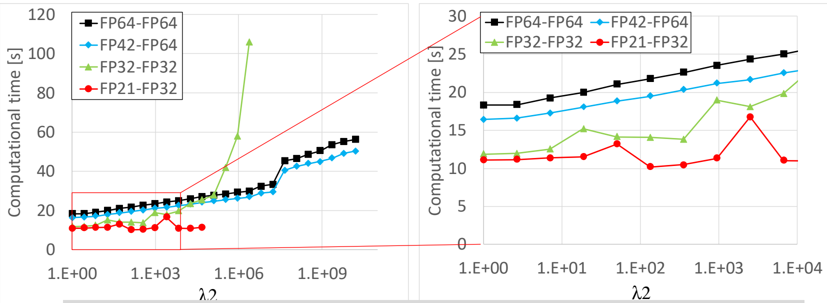
In recent years, the usefulness of low-precision floating-point

Speed-up of ICCG Solver with Mixed Precision on A64FX λ_1/λ_2= 1.00

Preconditioning	Others (SpMV etc.)	Performance
FP64	FP64	1.000
FP21	FP32	1.405
FP32	FP32	1.353
FP32	FP64	1.199
FP42	FP64	1.105

32,

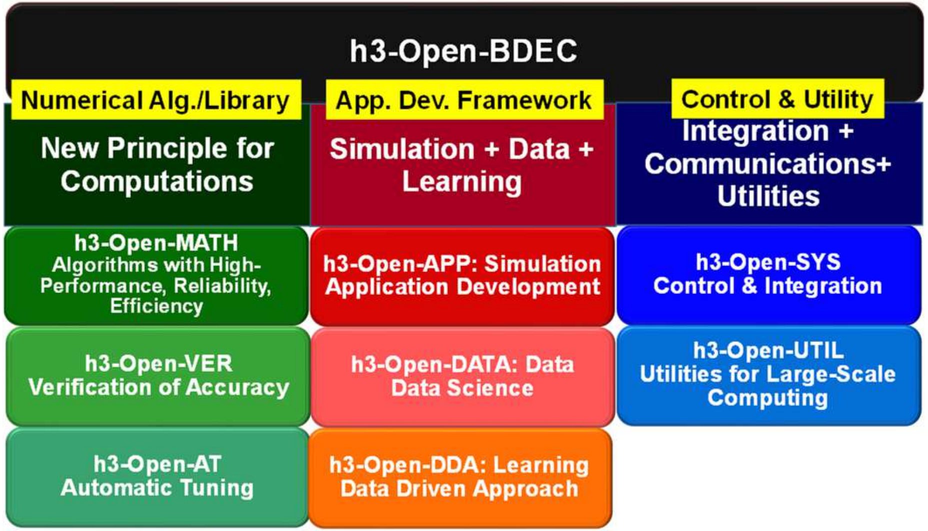
linear equations derived from 3D FVM code for steady-state head conduction with heterogeneous material property ($\lambda_1=10^0$, $\lambda_2=10^0$~10^9). Generally, computation with lower precision (e.g. FP32-FP32, FP21-FP32) becomes unstable, if condition number of the coefficient matrix is larger (λ_2 is larger), FP21-FP32 provides the best performance if λ_2 is up to 10^4. ("FP21-FP32" means "matrices are in FP21, and vectors are in FP32)



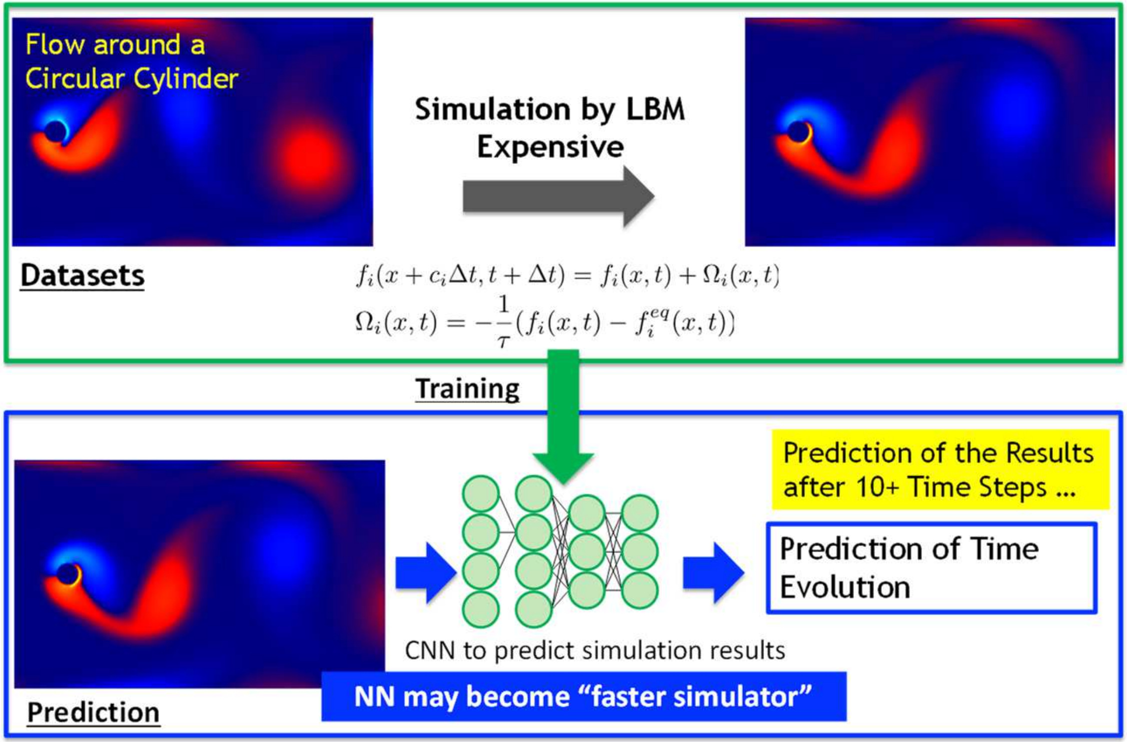
Computation Time for ICCG Solver
Various Types of Precisions on Intel Xeon Cascadelake

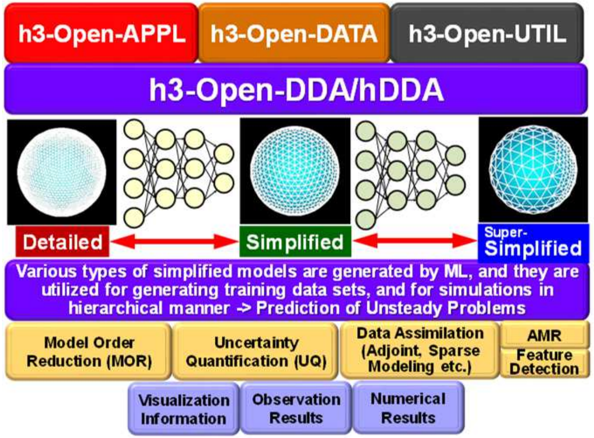
h3-Open-BDEC Innovative Software Platform for Integration of (S+D+L) on the BDEC System, such as Wisteria/BDEC-01

- "Three" Innovations
  - New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
  - Integration of (S+D+L) by Hierarchical Data Driven Approach (*h*DDA)
  - Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01



Acceleration of Transient CFD Simulations using ML/CNN Integration of (S+D+L), AI for HPC/AI for Science



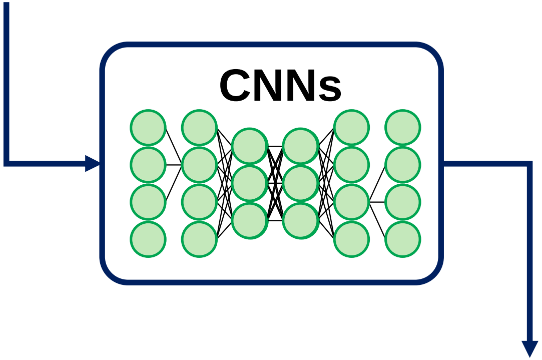


Initial Target: Estimating O(10) time steps ahead in transient CFD simulations

[c/o Takashi Shimokawabe (ITC/U.Tokyo)]

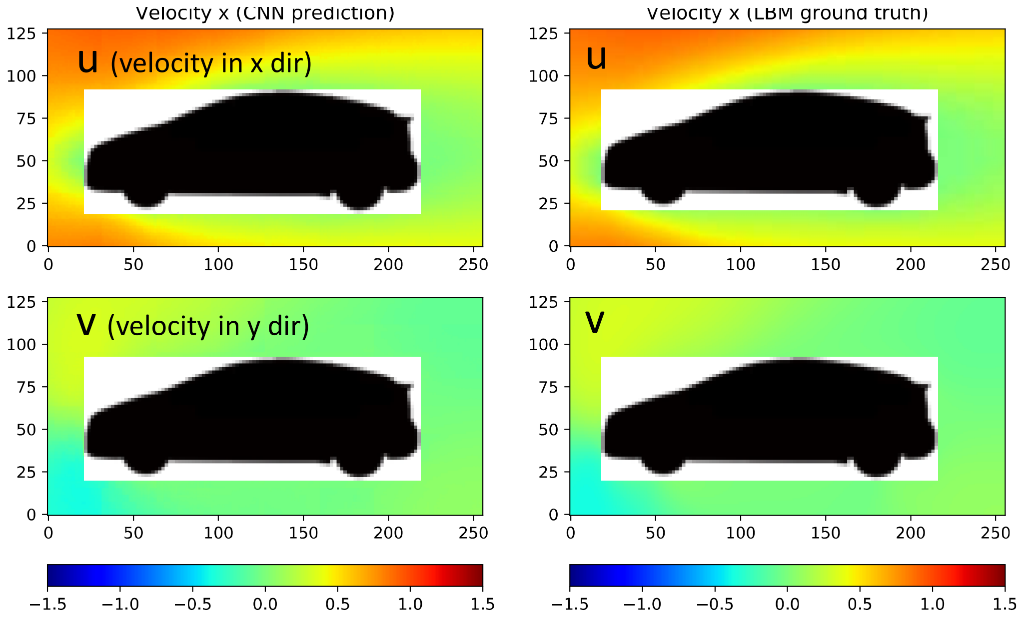
Prediction of steady flows using convolutional neural networks (CNNs)

Input
- Signed distance function (Geometry)
- Boundary conditions of velocity (u, v)



Output
- Velocity (u, v)

CNN Prediction　　LBM Ground truth



Computation time

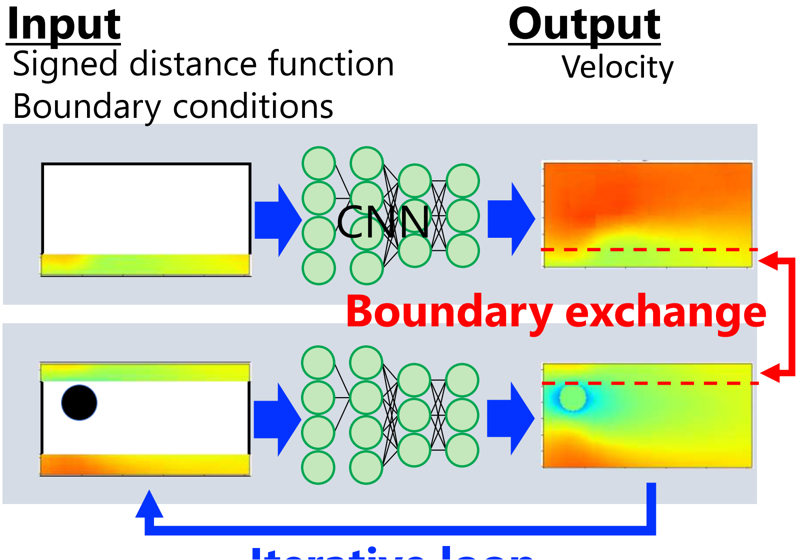
LBM (82,000steps) : 41.1 sec
CNN prediction: 0.6 sec

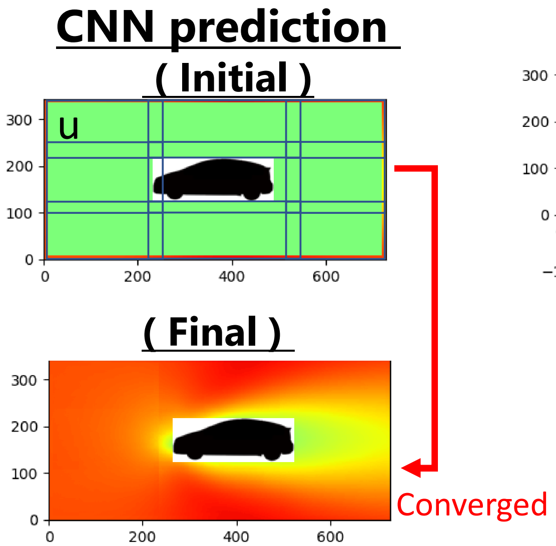
[c/o Takashi Shimokawabe (ITC/U.Tokyo)]

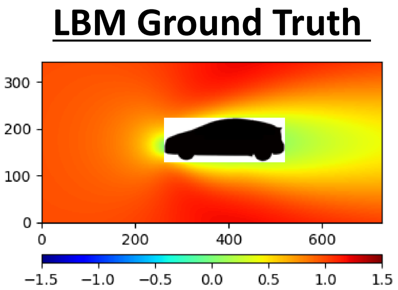
CNN prediction has achieved high accuracy with significant reduction in calculation time.

Prediction by CNN with boundary exchange

- Predicting simulation results on large domain using CNN with boundary exchange.
- The network model trained for a single domain is applied to the decomposed subdomains to predict the simulation results in each subdomain.
- In order to maintain consistency between values in the subdomains, boundary exchange between neighbor subdomains is performed.
- CNN and boundary exchange are performed iteratively until values converge.



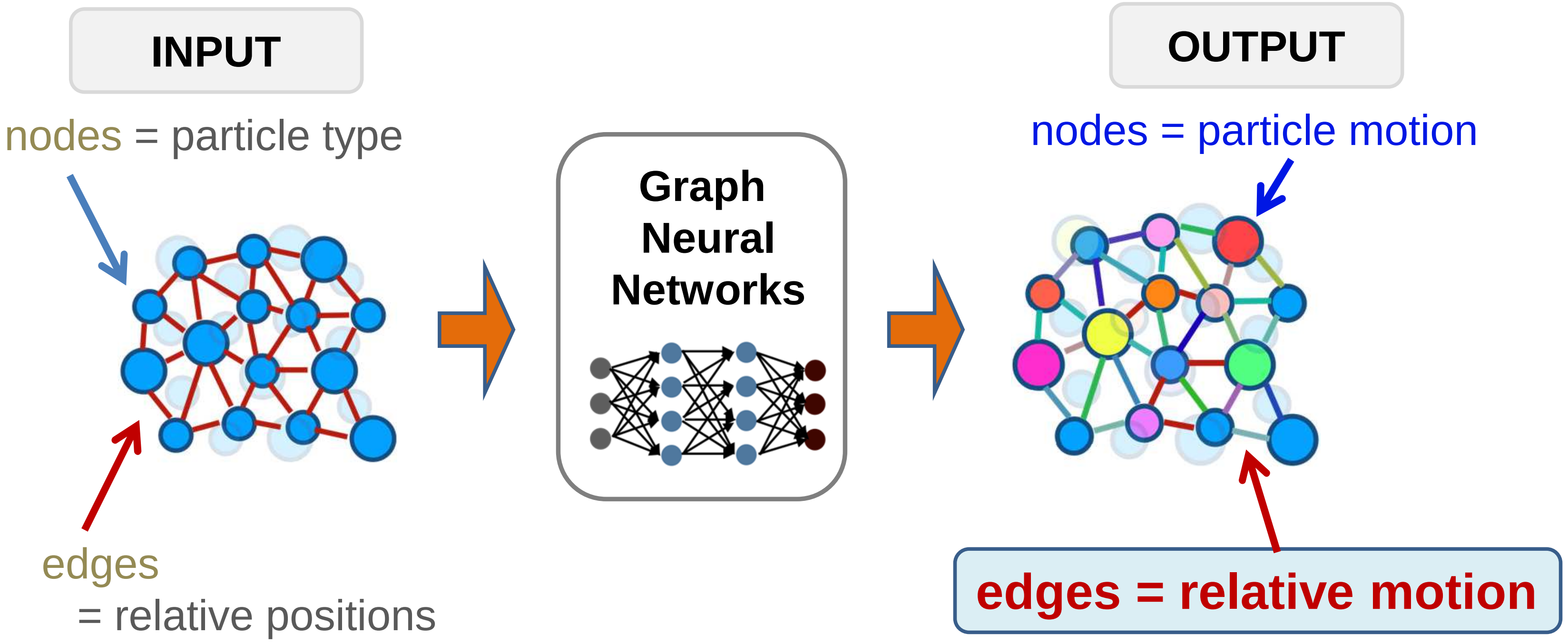




Domain size：748 x 364
(9 decomposed subdomains)
Mean error ：3.89％
Comp. time ：3.82 s

Machine learning slow molecular dynamics

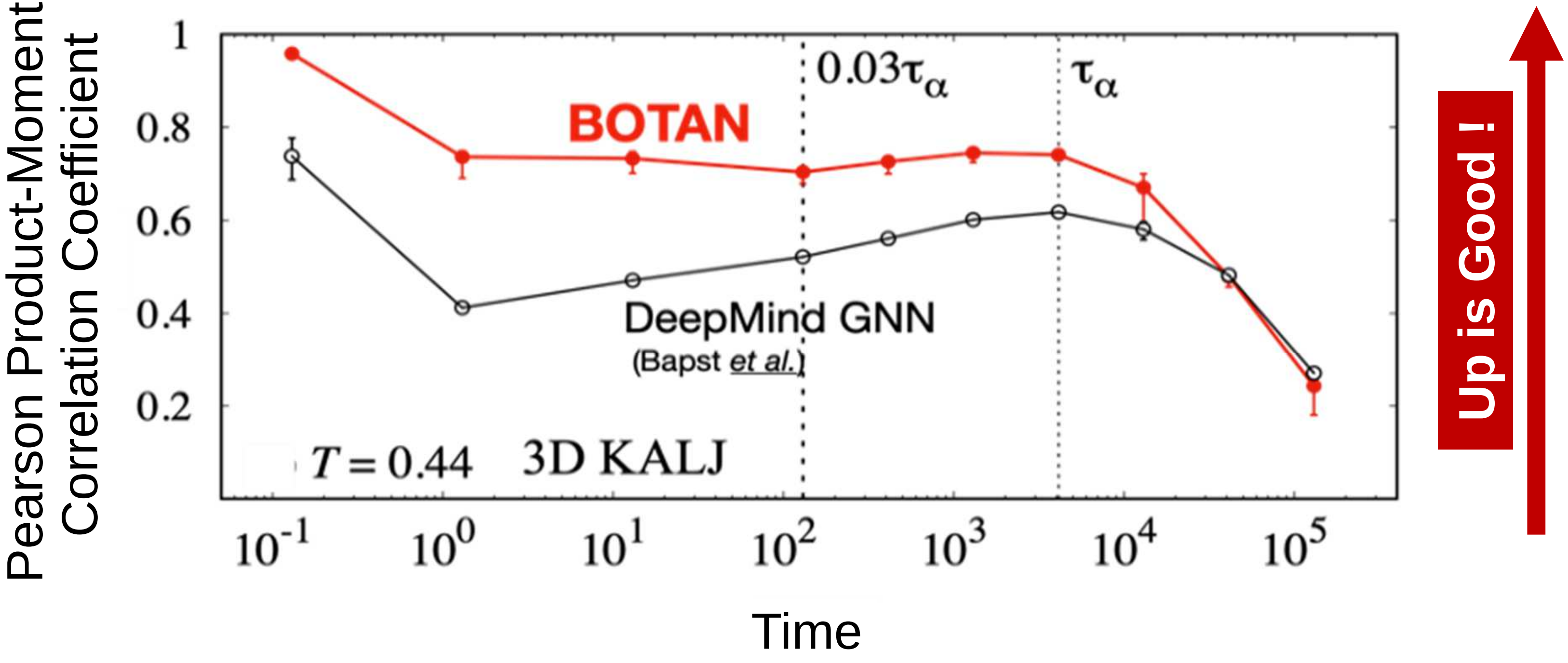
Our proposal ― **BOnd Targeting Network (BOTAN)**



H. Shiba, M. Hanai, T. Suzumura, and T. Shimokawabe, arXiv:2206.14024 (2022)

Machine learning slow molecular dynamics

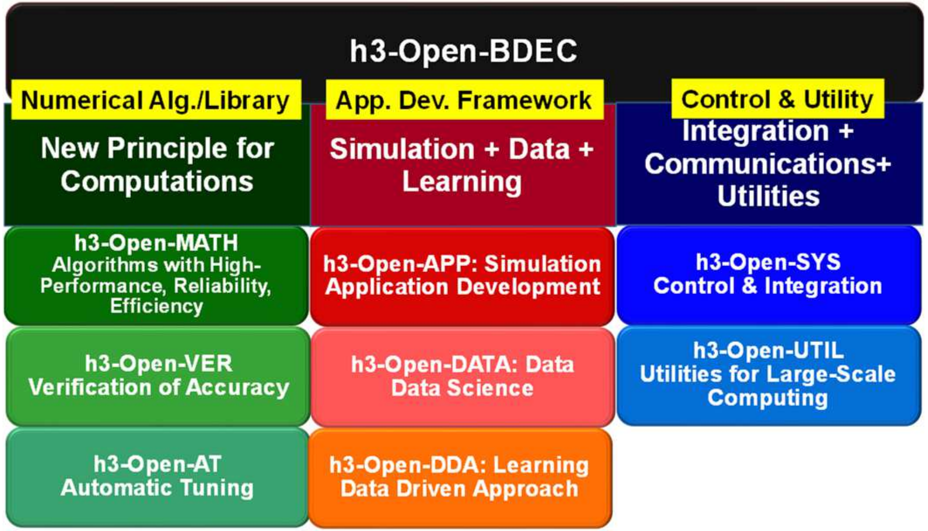
Our proposal － **BOnd Targeting Network (BOTAN)**



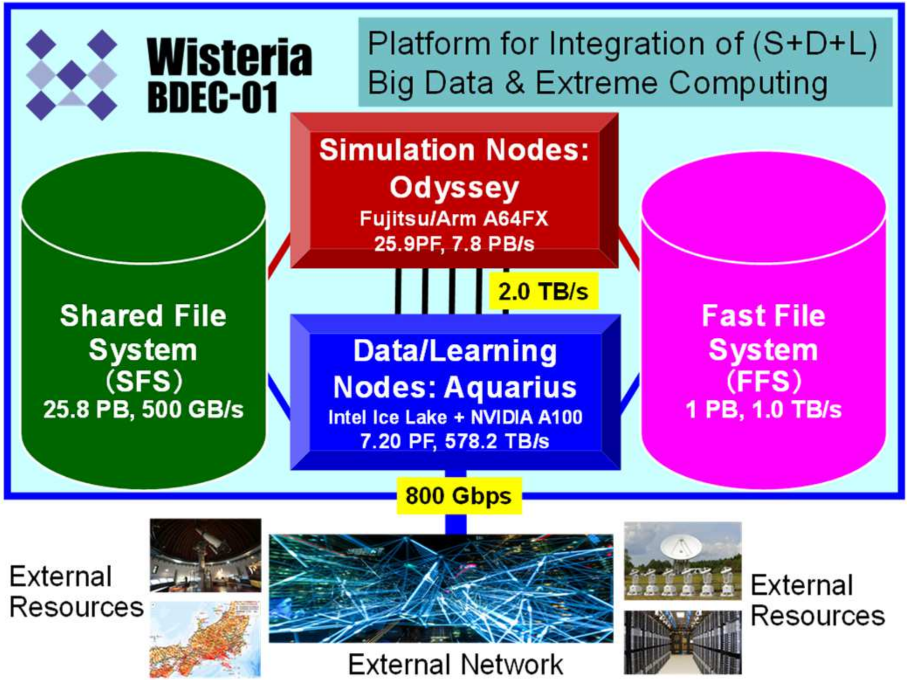
H. Shiba, M. Hanai, T. Suzumura, and T. Shimokawabe, arXiv:2206.14024 (2022)

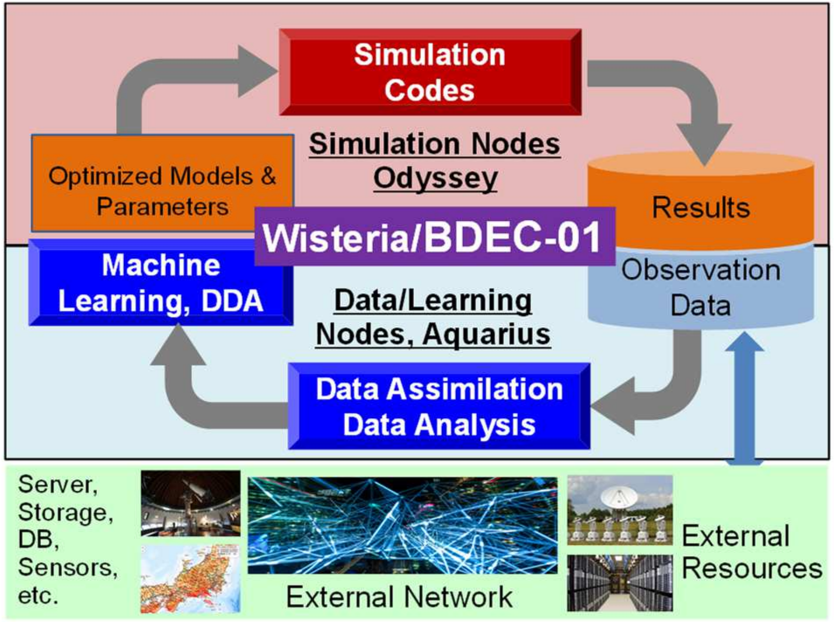
h3-Open-BDEC Innovative Software Platform for Integration of (S+D+L) on the BDEC System, such as Wisteria/BDEC-01

- “Three” Innovations
  - New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
  - Integration of (S+D+L) by Hierarchical Data Driven Approach (*h*DDA)
  - Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01



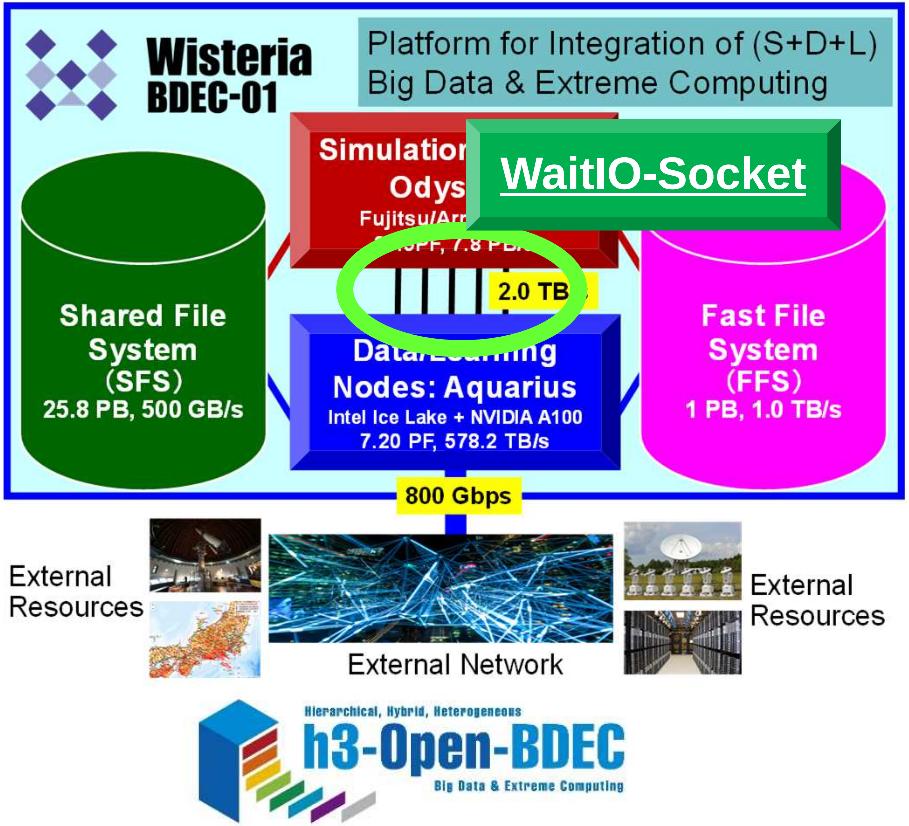
Wisteria/BDEC-01: The First “Really Heterogenous” System in the World





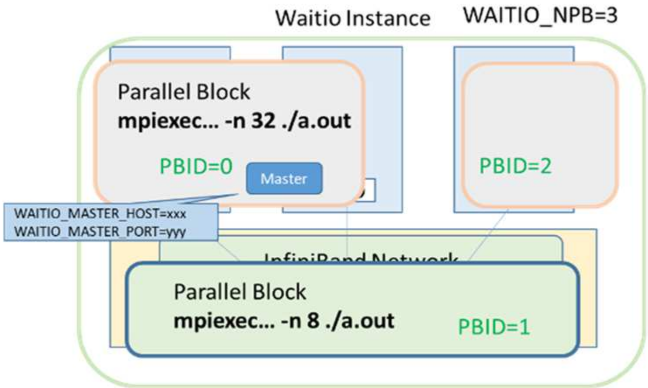
h3-Open-SYS/WaiIO-Socket

- Wisteria/BDEC-01
  - Aquarius (GPU: NVIDIA A100)
  - Odyssey (CPU: A64FX)
- Combining Odyssey-Aquarius
  - Single MPI Job over O-A is impossible
- **Connection between Odyssey-Aquarius**
  - **IB-EDR with 2TB/sec.**
  - **Fast File System**
  - **h3-Open-SYS/WaitIO-Socket**
    - **Library for Inter-Process Communication through IB-EDR with MPI-like interface**



API of h3-Open-SYS/WaitIO-Socket PB (Parallel Block): Each Application

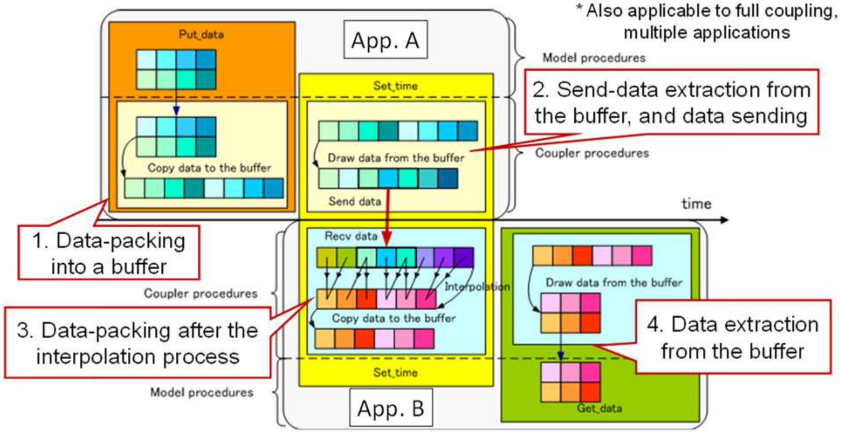
WaitIO API	Description
waitio_isend	Non-Blocking Send
waitio_irecv	Non-Blocking Receive
waitio_wait	Termination of waitio_isend/irecv
waitio_init	Initialization of WaitIO
waitio_get_nprocs	Process # for each PB (Parallel Block)
waitio_create_group waitio_create_group_wranks	Creating communication groups among PB's
waitio_group_rank	Rank ID in the Group
waitio_group_size	Size of Each Group
waitio_pb_size	Size of the Entire PB
waitio_pb_rank	Rank ID of the Entire PB



[Sumimoto et al. 2021]

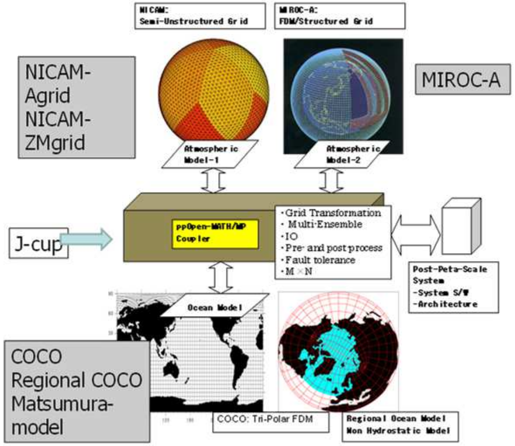
Multiphysics Coupler

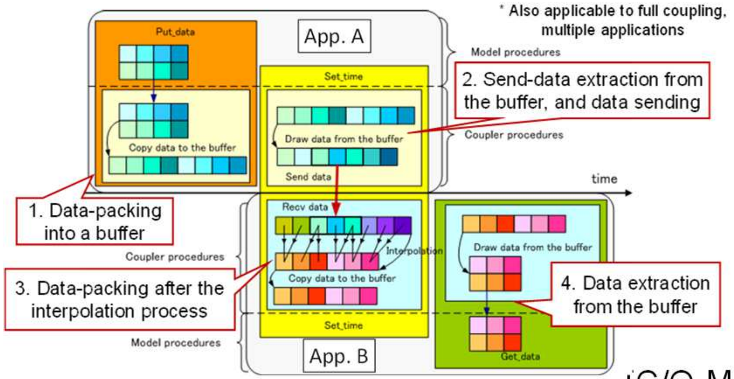
- Traditional Coupler: ppOpen-MATH/MP
- Weak-Coupling of Multiple (usually two) Applications
  - Each application does a single computation
  - Ocean-Atmosphere
  - Fluid-Structure

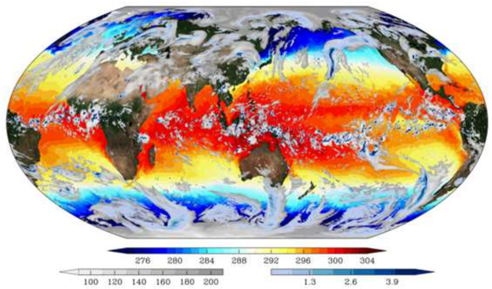


Atmosphere-Ocean Coupling by ppOpen-MATH/MP (Previous Project)

- High-resolution global atmosphere-ocean coupled simulation by NICAM (Atmosphere) and COCO (Ocean) through ppOpen-MATH/MP on the K computer is achieved.
  - ppOpen-MATH/MP is a coupling software for the models employing various discretization method.





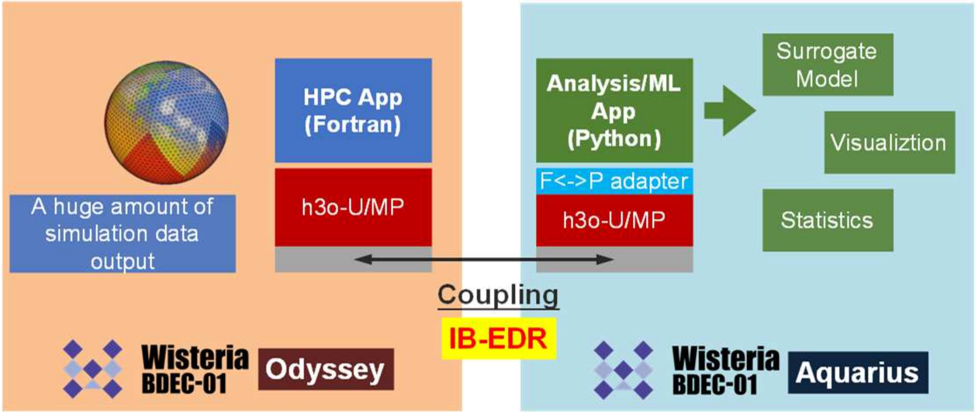


[C/O M. Satoh (AORI/UTokyo)@SC16]

h3-Open-UTIL/MP

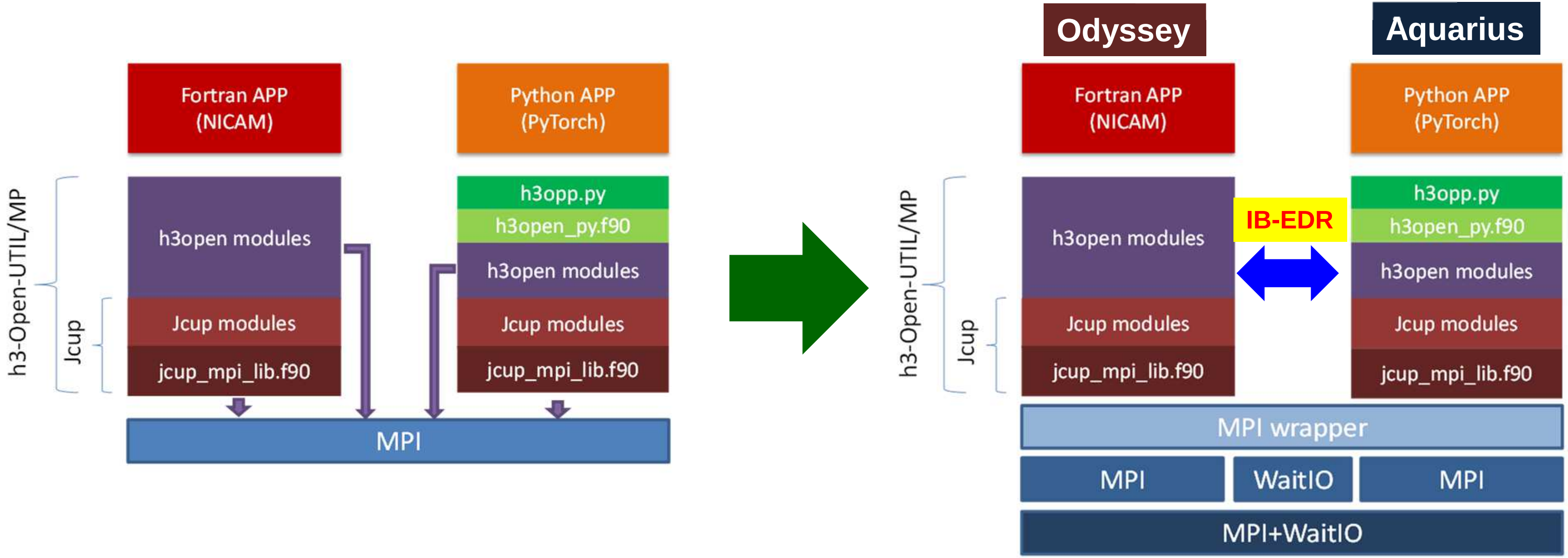
Multilevel Coupler/Data Assimilation Integration of (S+D+L)

- Extended Version of Multy-Physics Coupler
- Data Assimilation (Multiple Computations: Ensemble)
  - Assimilation of Computations with Different Resolutions
  - Data Assimilation by Coupled Codes
    - e.g. Atmosphere-Ocean
- Coupling of Simulations on Odyssey and AI on Aquarius



h3-Open-UTIL/MP +
h3-Open-SYS/WaitIO-Socket

Available in June 2022



May 2021: MPI Only

June 2022: Coupler＋WaitIO

- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - h3-Open-BDEC
- **Applications on Wisteria/BDEC-01 with h3-Open-BDEC**
  - **Earthquake Simulations**
  - (Global Cloud Simulation+AI) Coupling
  - Ensemble Coupling
  - International Collaboration through JHPCN

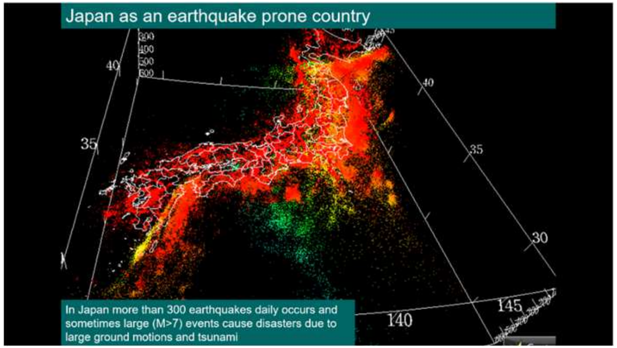
Early Forecast of Long-Period Ground Motions via Data Assimilation of Observation and Simulations [Furumura et al. 2018]

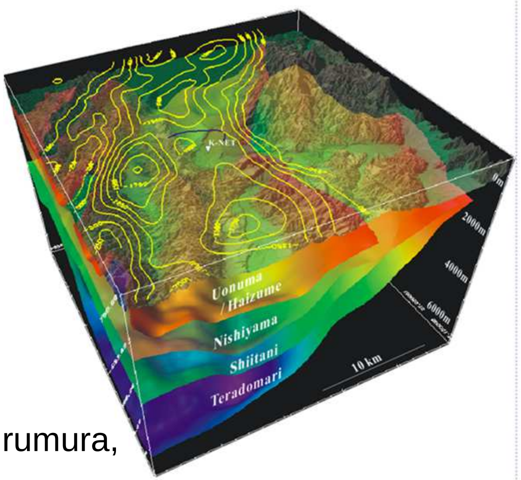
https://doi.org/10.1029/2018GL081163

- New method for the early forecast of long-period (> 3–10 s) ground motions generated by large earthquakes based on the data assimilation of observed ground motions and FDM simulations of seismic wave propagation in a 3-D heterogeneous structure (**Seism3D/OpenSWPC-DAF（Data-Assimilation-Based Forecast）**).
- **This approach uses the dense nationwide network in Japan and supercomputers to perform forecasts using the assimilated wavefields at speeds much faster than the actual wave propagation speed.**
- **An early alert can be issued prior to the occurrence of strong motions due to large and distant earthquakes.**
- **This research inspired me to develop a system like Wisteria/BDEC-01, where (Simulation, Data, Learning) are integrated on a single system.**

Earthquake simulation is always with uncertainty

- Subsurface/Underground Structure
  - Heterogenous, Random, Stochastic
  - Fluctuations
- Traditional Simulations
  - Forward Simulations
- **Integration of Simulation/Observation is essential**
- **New Types of Methods for Simulations combined with Data Assimilation/Real-Time Observation is under development**
  - **Forecast by Simulations, Correction by Data Assimilation**

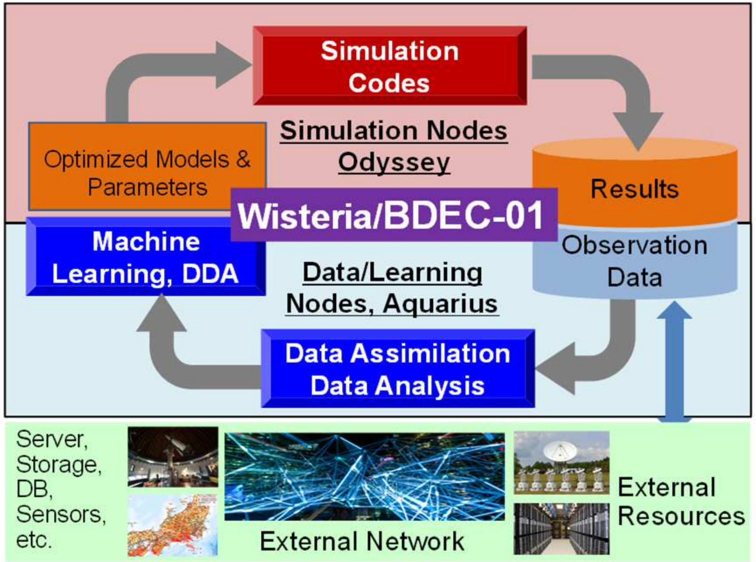


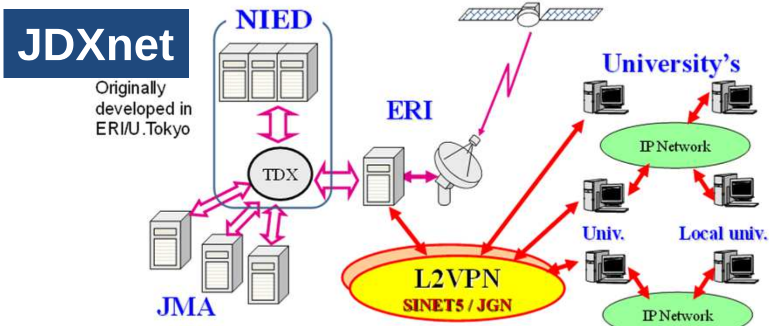


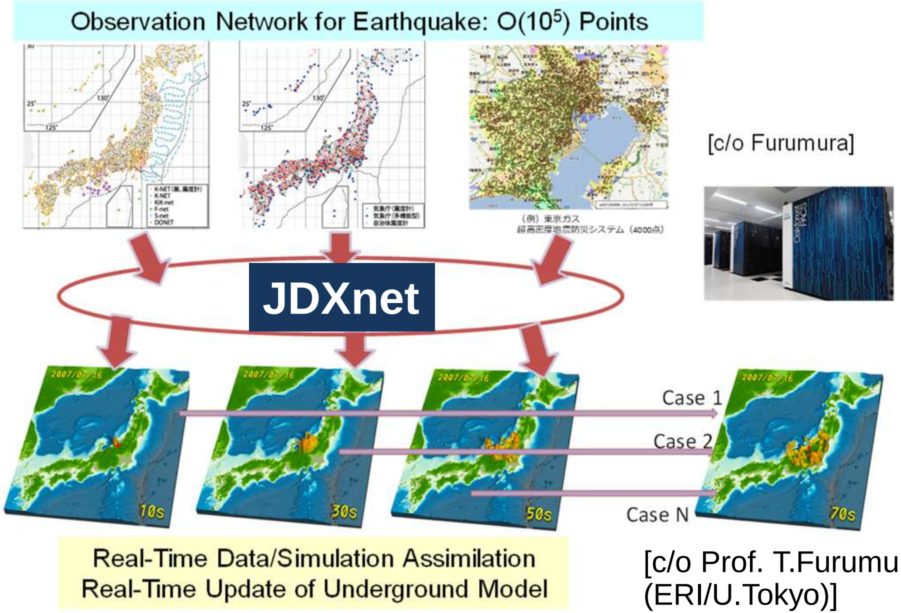
[c/o Prof. T. Furumura, ERI/U.Tokyo]

3D Earthquake Simulation with Real-Time Data Observation/Assimilation

Simulation of Strong Motion (Wave Propagation) by 3D FDM





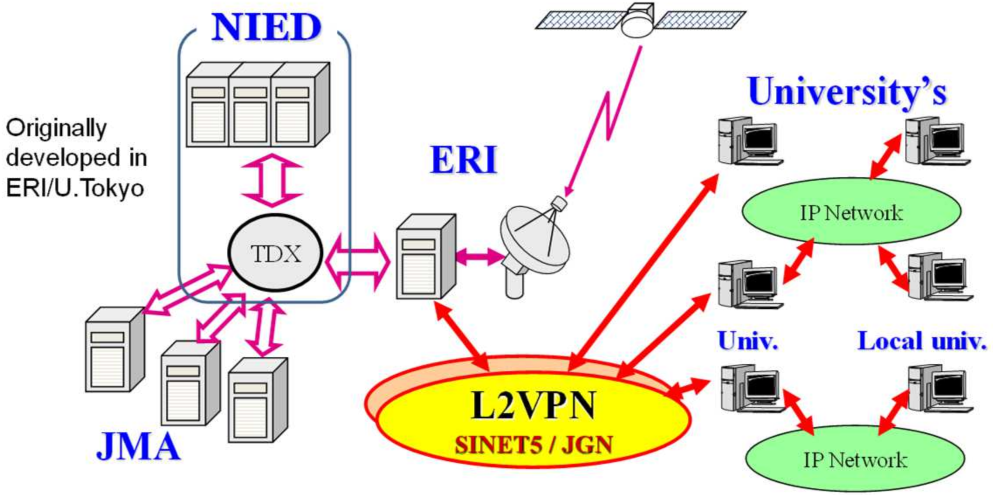


[c/o Prof. T.Furumura (ERI/U.Tokyo)]

Real-Time Sharing of Seismic Observation is possible in Japan by JDXnet with SINET

Japan Data eXchange network

- Seismic Observation Data (100Hz/3-dir's/O(10^3) observation points) by JDXnet is available through SINET in Real Time
  - O(10^2) GB/day: available at Website of NIED
  - O(10^5) pts in future including stations operated by industry



[c/o Prof. H.Tsuruoka (ERI/U.Tokyo)]

Real-Time Assimilation of “Observation+Computation” in Seismic Wave Propagation [c/o Oba & Furumura]

- **Data Assimilation of Wave Propagation by “<u>Optimal Interpolation Technique</u>”**

Assim. Comp. Obs. Residual Comp.

$$\boldsymbol{x}_n^a = \boldsymbol{x}_n^f + \boldsymbol{W}(\boldsymbol{y}_n - \boldsymbol{H}\boldsymbol{x}_n^f)$$

Comp. Assim.

$$\boldsymbol{x}_{n+1}^f = \boldsymbol{F}\boldsymbol{x}_n^a$$

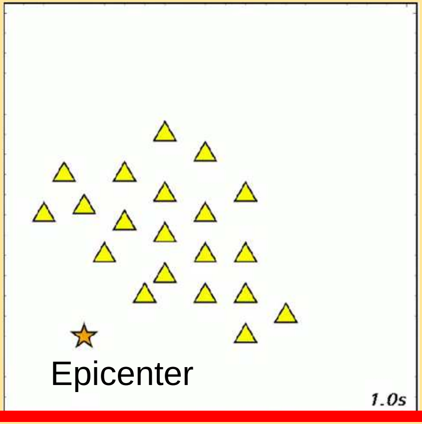
n: Time Step

$\boldsymbol{W}$: **Weighting Matrix**

$\boldsymbol{F}$: Wave Propagation simulation

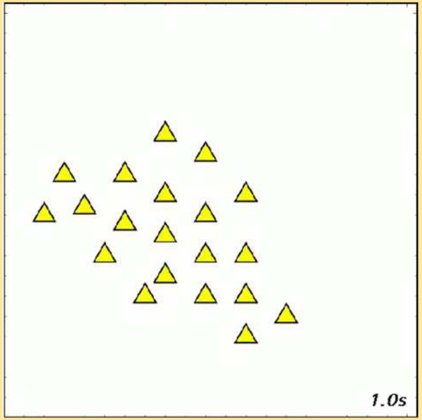
（A) Pure Simulation

△：Obs. Pts.



（B） Assimilation+Sim.

(No info for Epicenter needed)

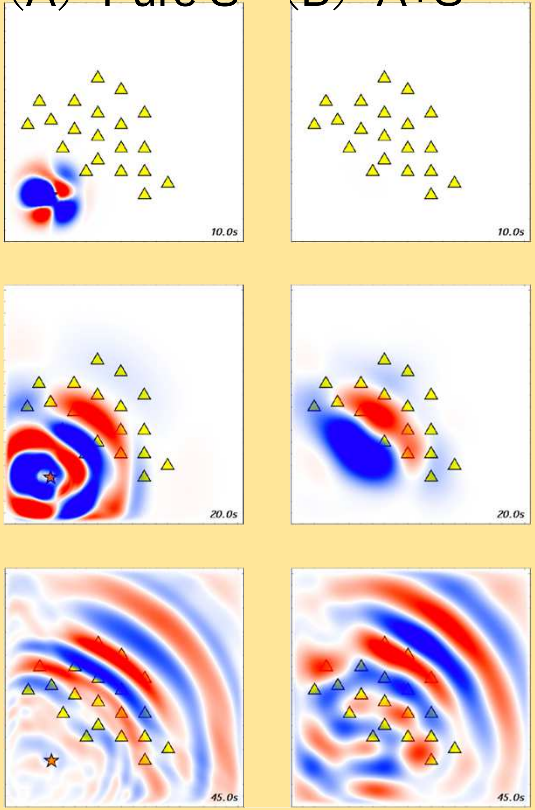


（A） Pure S （B） A+S

10 sec.

20 sec.

30 sec.



Real-Time Assimilation of "Observation+Computation" in Seismic Wave Propagation [c/o Oba & Furumura]

- **Data Assimilation of Wave Propagation by "<u>Optimal Interpolation Technique</u>"**

Assim. Comp. Obs. Residual Comp.

$$x_n^a = x_n^f + W(y_n - Hx_n^f)$$

Comp. Assim.

$$x_{n+1}^f = Fx_n^a$$

n: Time Step

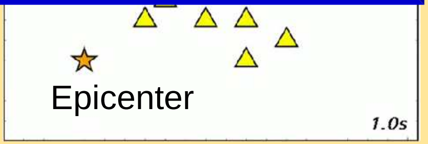
W: **Weighting Matrix**

F: Wave Propagation simulation

(A) Pure Simulation

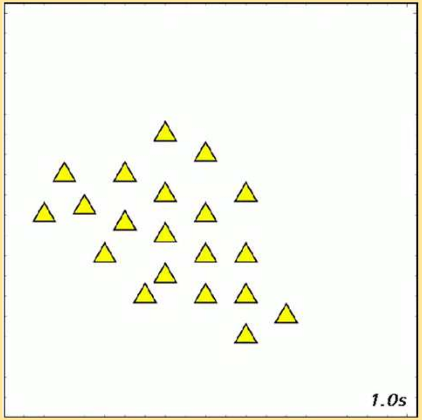
△: Obs. P s.

Initial Conditions are created by Interpolation of Observed Results



(B) Assimilation+Sim.

(No info for Epicenter needed)

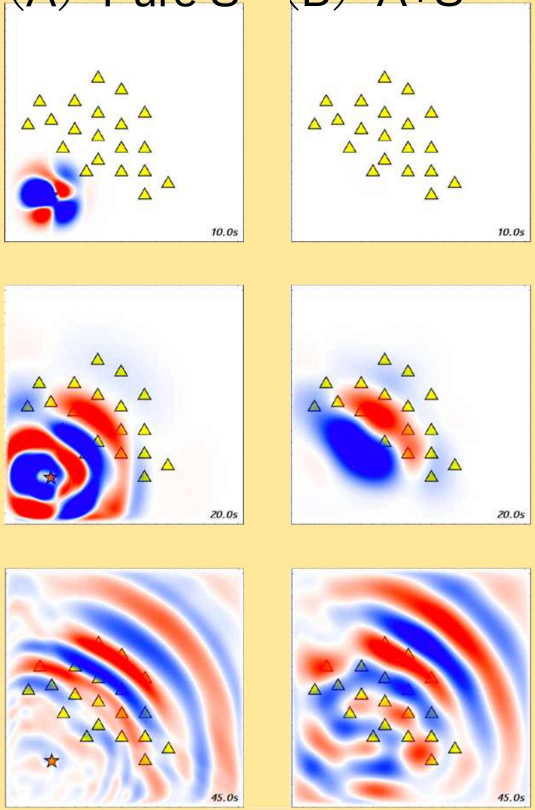


(A) Pure S (B) A+S

10 sec.

20 sec.

30 sec.



Starting from (A+S: Assim+Sim.) to (Pure S: Pure Simulation)

Residual

Assim. Comp. Obs. Comp.

$$x_n^a = x_n^f + W(y_n - Hx_n^f)$$

Comp. Assim.

$$x_{n+1}^f = Fx_n^a$$

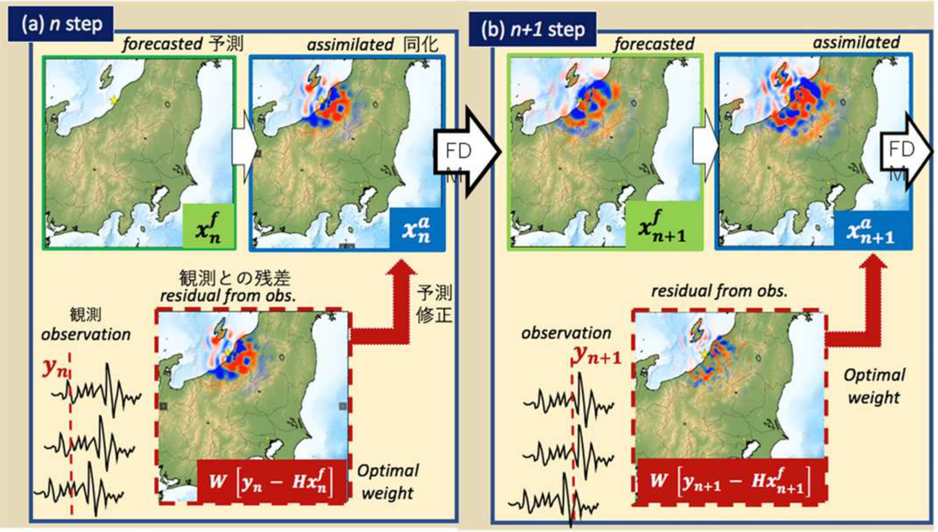
n: Time Step

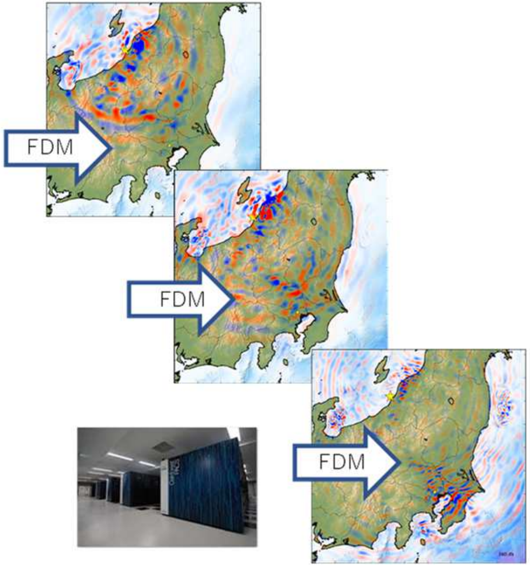
W: **Weighting Matrix**

F: Wave Propagation simulation

(A+S) Assimilation+Simulation

(Pure S) Pure Simulation/Forecast

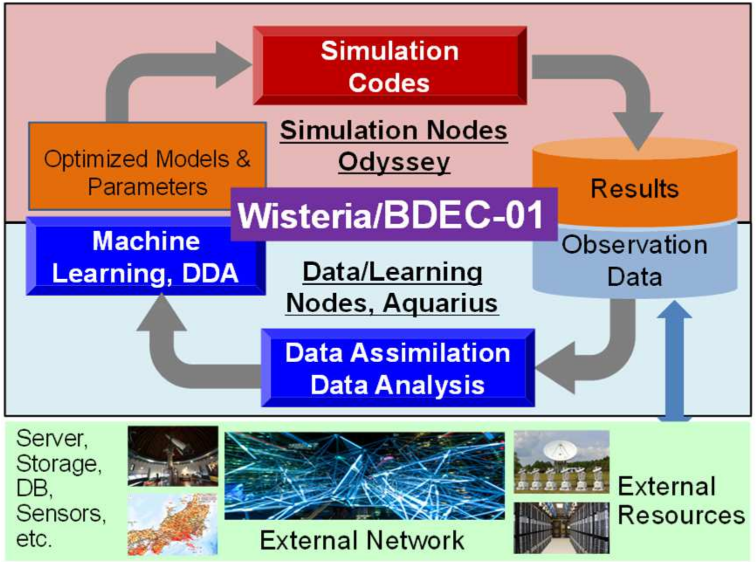


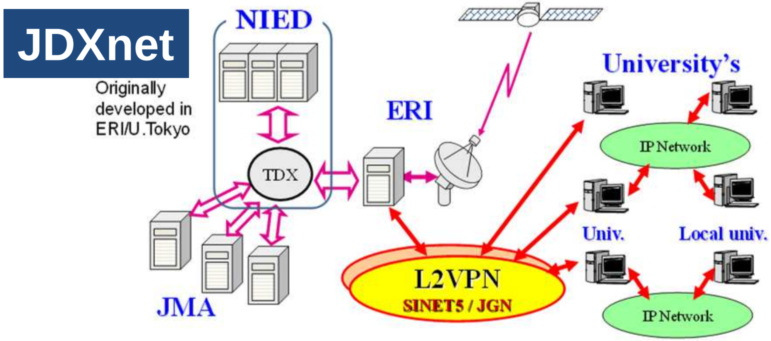


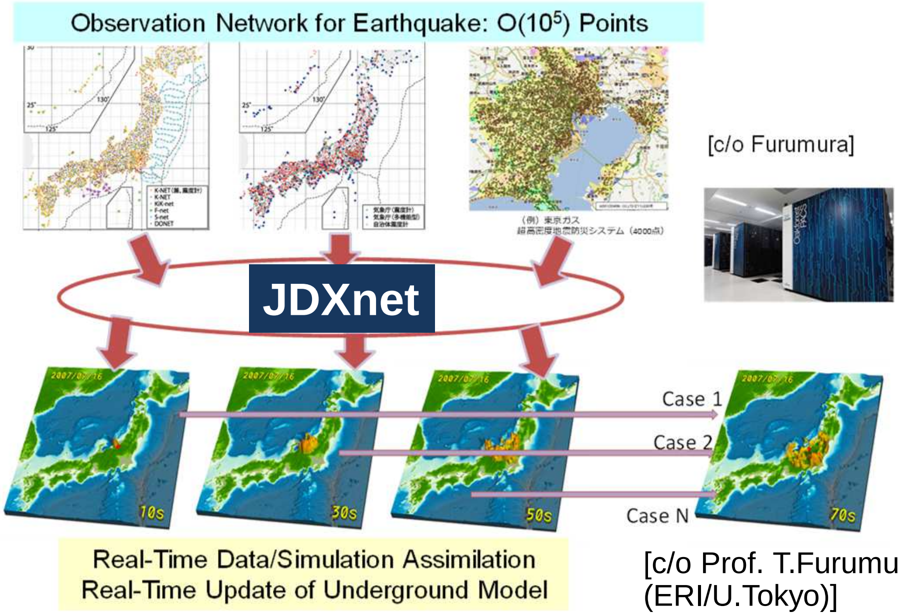
[c/o Prof. T. Furumura, ERI/U.Tokyo]

3D Earthquake Simulation with Real-Time Data Observation/Assimilation

Simulation of Strong Motion (Wave Propagation) by 3D FDM

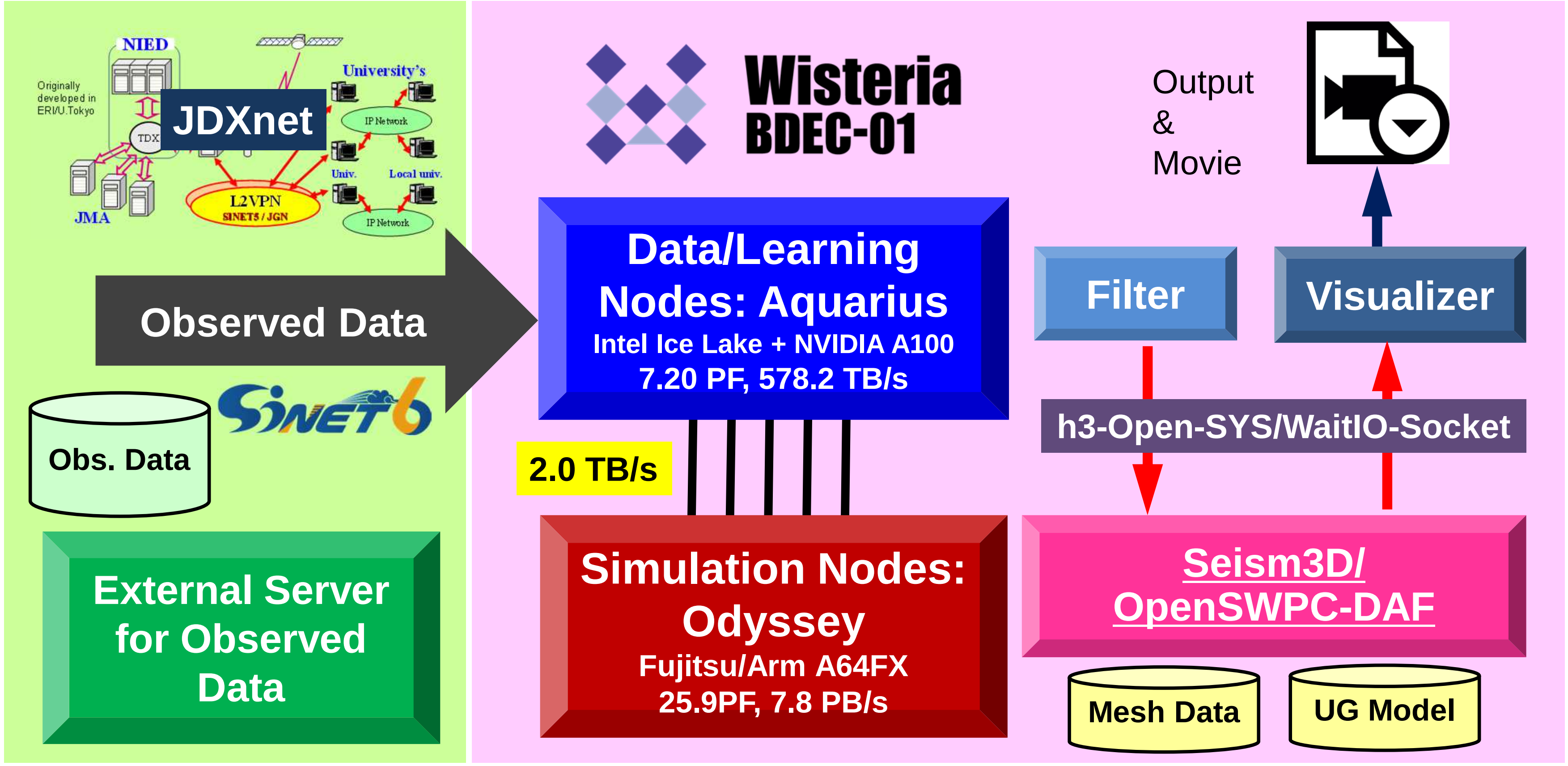






[c/o Prof. T.Furumura (ERI/U.Tokyo)]

3D Earthquake Simulation with Real-Time Data Observation/Assimilation on Wisteria/BDEC-01



Communications by WaitIO-Socket

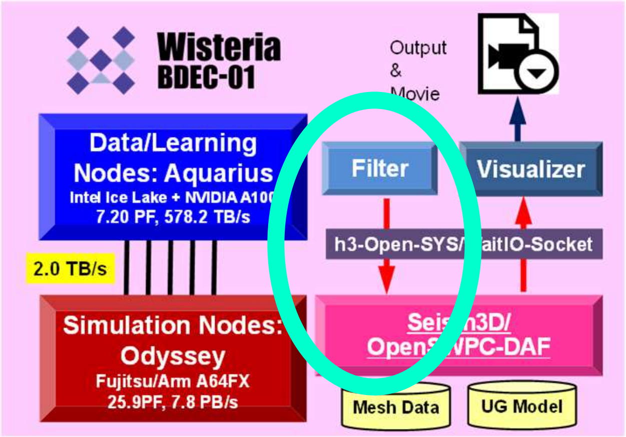
[Kasai et al. 2021]

Aquarius: SEND

```
program dmy_filter
<省略: 型宣言等>
call mpi_init (ierr)
call mpi_comm_size (MPI_COMM_WORLD, nprocs, ierr)
call mpi_comm_rank (MPI_COMM_WORLD, myrank,  ierr)
call WAITIO_CREATE_UNIVERSE (WAITIO_COMM_UNIVERSE, ierr)

if (myrank==0) then
  open(100,file='./obsfile_list.txt', form='formatted', status='old', iostat=ierr)
  do i=1,300
   <省略: obsデータ読み込み処理>
    print *,"Send obs data ...... "
    call WAITIO_MPI_ISEND (NTMAX1_o, 1,           WAITIO_MPI_INTEGER,  2,1, WAITIO_COMM_UNIVERSE,req(1,1), ierr)
    call WAITIO_MPI_ISEND (DT_o,     1,           WAITIO_MPI_FLOAT,    2,2, WAITIO_COMM_UNIVERSE,req(1,2), ierr)
    call WAITIO_MPI_ISEND (NST_o,    1,           WAITIO_MPI_INTEGER,  2,3, WAITIO_COMM_UNIVERSE,req(1,3), ierr)
    call WAITIO_MPI_ISEND (AT_o,     1,           WAITIO_MPI_FLOAT,    2,4, WAITIO_COMM_UNIVERSE,req(1,4), ierr)
    call WAITIO_MPI_ISEND (T0_o,     1,           WAITIO_MPI_FLOAT,    2,5, WAITIO_COMM_UNIVERSE,req(1,5), ierr)
    call WAITIO_MPI_ISEND (ISO_X_o,  NSMAX,       WAITIO_MPI_INTEGER,  2,6, WAITIO_COMM_UNIVERSE,req(1,6), ierr)
    call WAITIO_MPI_ISEND (ISO_Y_o,  NSMAX,       WAITIO_MPI_INTEGER,  2,7, WAITIO_COMM_UNIVERSE,req(1,7), ierr)
    call WAITIO_MPI_ISEND (ISO_Z_o,  NSMAX,       WAITIO_MPI_INTEGER,  2,8, WAITIO_COMM_UNIVERSE,req(1,8), ierr)
    call WAITIO_MPI_ISEND (ISTX_o,   NST,         WAITIO_MPI_INTEGER,  2,9, WAITIO_COMM_UNIVERSE,req(1,9), ierr)
    call WAITIO_MPI_ISEND (ISTY_o,   NST,         WAITIO_MPI_INTEGER,  2,10,WAITIO_COMM_UNIVERSE,req(1,10),ierr)
    call WAITIO_MPI_ISEND (ISTZ_o,   NST,         WAITIO_MPI_INTEGER,  2,11,WAITIO_COMM_UNIVERSE,req(1,11),ierr)
    call WAITIO_MPI_ISEND (STC_o,    6*NST,       WAITIO_MPI_CHAR,     2,12,WAITIO_COMM_UNIVERSE,req(1,12),ierr)
    call WAITIO_MPI_ISEND (VxAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,    2,13,WAITIO_COMM_UNIVERSE,req(1,13),ierr)
    call WAITIO_MPI_ISEND (VyAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,    2,14,WAITIO_COMM_UNIVERSE,req(1,14),ierr)
    call WAITIO_MPI_ISEND (VzAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,    2,15,WAITIO_COMM_UNIVERSE,req(1,15),ierr)
    call WAITIO_MPI_WAITALL (15,req, status, ierr)
    call sleep(1)
  enddo
  close (100)
endif
call WAITIO_FINALIZE (ierr)
call mpi_finalize (ierr)
end
```

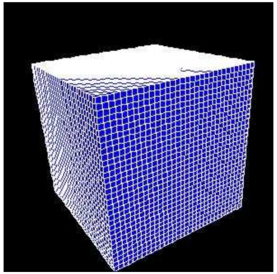
Odyssey: RECV

```
call WAITIO_MPI_IRECV (NTMAX1_o, 1,           WAITIO_MPI_INTEGER, 0,1, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (DT_o,     1,           WAITIO_MPI_FLOAT,   0,2, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (NST_o,    1,           WAITIO_MPI_INTEGER, 0,3, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (AT_o,     1,           WAITIO_MPI_FLOAT,   0,4, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (T0_o,     1,           WAITIO_MPI_FLOAT,   0,5, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISO_X_o,  NSMAX,       WAITIO_MPI_INTEGER, 0,6, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISO_Y_o,  NSMAX,       WAITIO_MPI_INTEGER, 0,7, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISO_Z_o,  NSMAX,       WAITIO_MPI_INTEGER, 0,8, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISTX_o,   NST,         WAITIO_MPI_INTEGER, 0,9, WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISTY_o,   NST,         WAITIO_MPI_INTEGER, 0,10,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (ISTZ_o,   NST,         WAITIO_MPI_INTEGER, 0,11,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (STC_o,    6*NST,       WAITIO_MPI_CHAR,    0,12,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (VxAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,   0,13,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (VyAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,   0,14,WAITIO_COMM_UNIVERSE,...)
call WAITIO_MPI_IRECV (VzAll_obs,NST*NOBS_LEN,WAITIO_MPI_FLOAT,   0,15,WAITIO_COMM_UNIVERSE,...)
```



Example: Off Niigata 2007 Mw6.6 Earthquake

- Observed Data: Stored in External Server
- Aquarius receives observed data, and apply filtering
- “Data Assimilation + Simulation (A+S)”, and “Forecast by Simulation (Pure S)” are separated codes, while same number of computing nodes were used on Odyssey
- Movies were created after simulations (O(10) sec.)



- Seism3D/OpenSWPC-DAF
  - 3D FDM + Optimal Interpolation Technique for Data Assimilation
  - Each Mesh: 240m×240m×240m
  - 1,920×1,920×240 meshes（8.85×10^8）
  - 460.8 km×460.8 km×57.6 km

$$v_p^n = v_p^{n-1} + \frac{1}{\rho}\left(\frac{\partial\sigma_{xp}^{n-1/2}}{\partial x} + \frac{\partial\sigma_{yp}^{n-1/2}}{\partial y} + \frac{\partial\sigma_{zp}^{n-1/2}}{\partial z}\right)\Delta t \quad (p = x, y, z)$$

Assim. Comp. Obs. Comp. Residual

$$\boldsymbol{x}_n^a = \boldsymbol{x}_n^f + \boldsymbol{W}(\boldsymbol{y}_n - \boldsymbol{H}\boldsymbol{x}_n^f)$$

Comp. Assim.

$$\boldsymbol{x}_{n+1}^f = \boldsymbol{F}\boldsymbol{x}_n^a$$

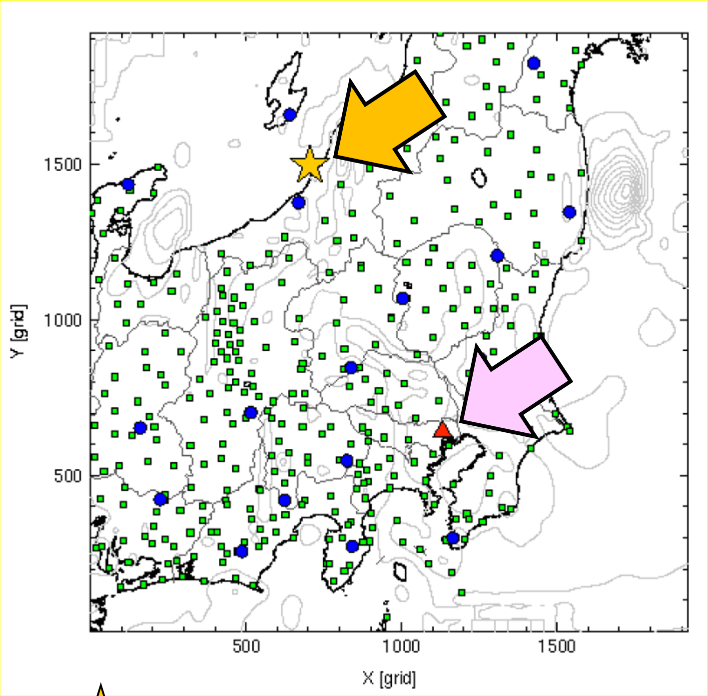
n: Time Step

$\boldsymbol{W}$: **Weighting Matrix**

$\boldsymbol{F}$: Wave Propagation simulation

Off Niigata 2007 Mw6.6 Earthquake

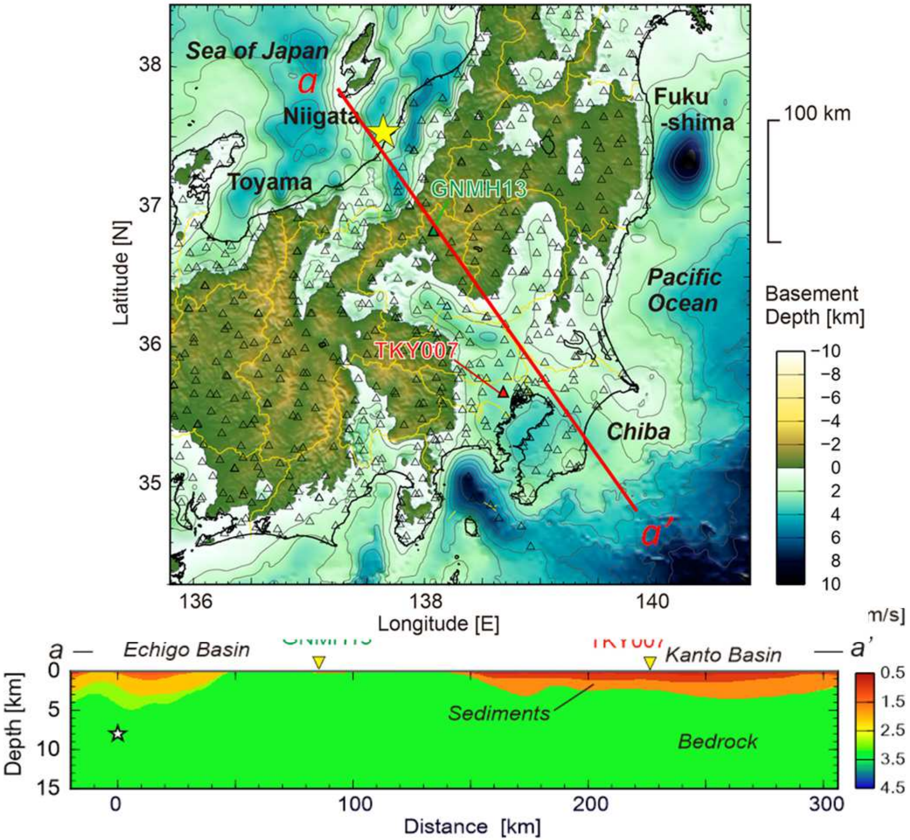
[c/o Prof. T. Furumura, ERI/U.Tokyo]



☆ Epicenter

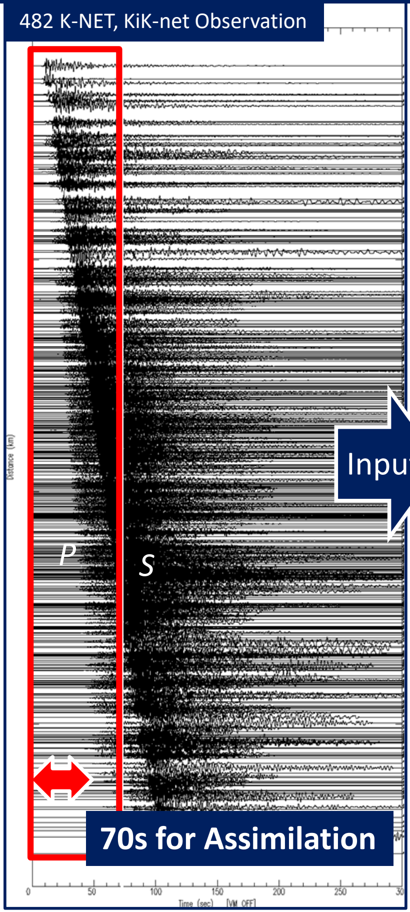
■ Hi-net (Short Period) 349 pts

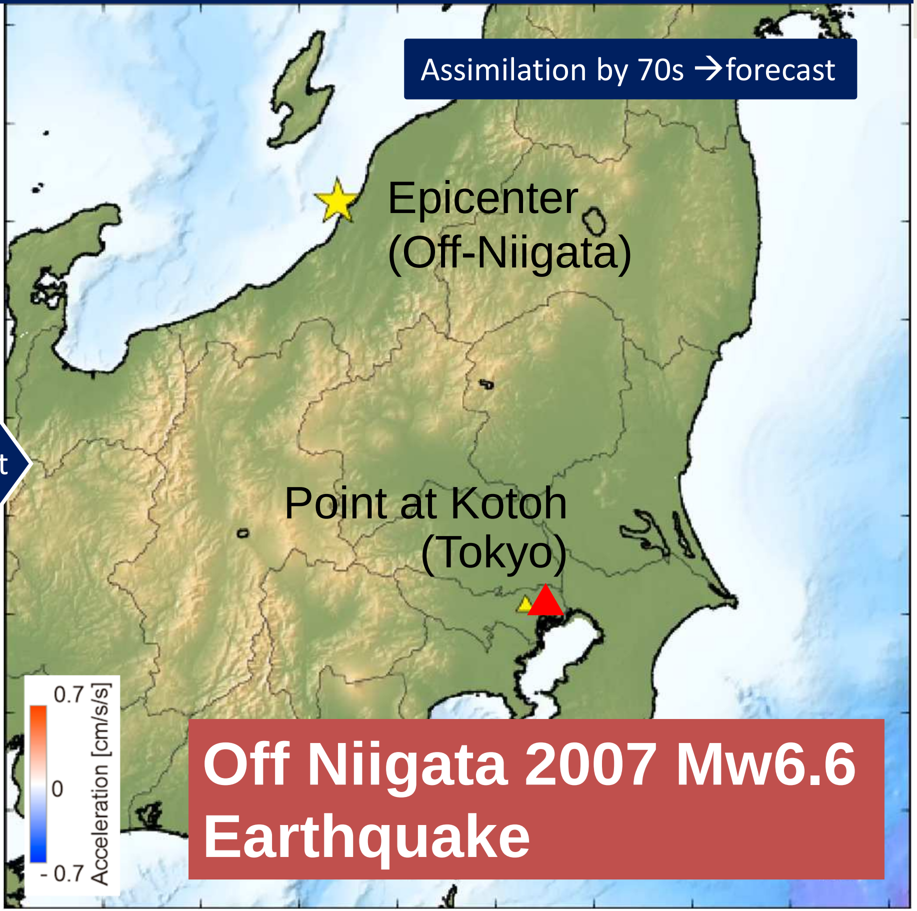
● F-net (Broadband) 18 pts



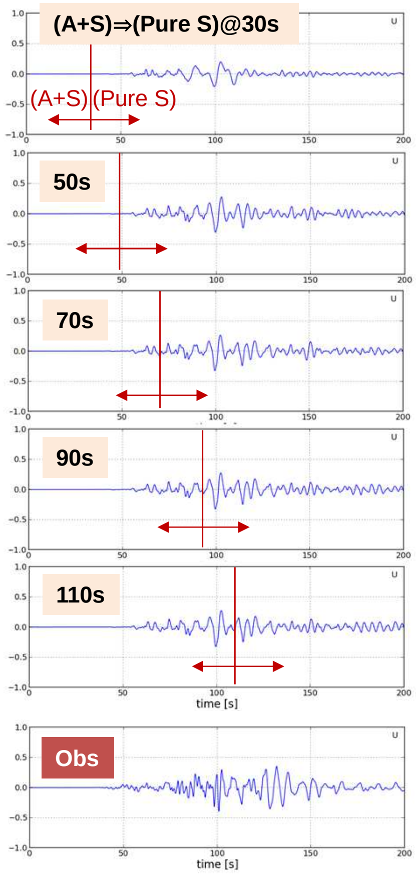
Data Assimilation + Pure Simulation/Forecast

Results at Kotoh ▲ (N.KOTH)
N 35° 37.0’
E 139° 46.9’



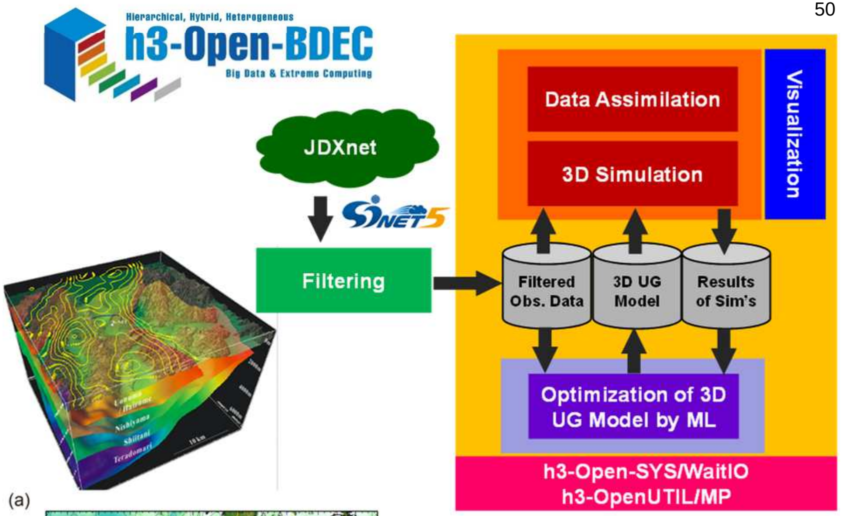


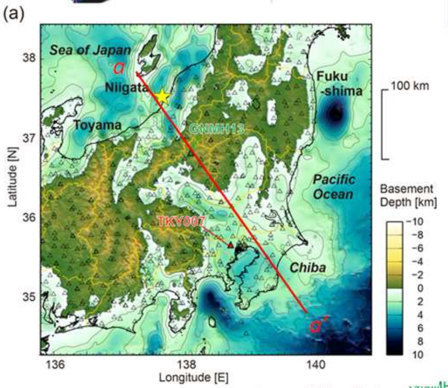
Off Niigata 2007 Mw6.6 Earthquake

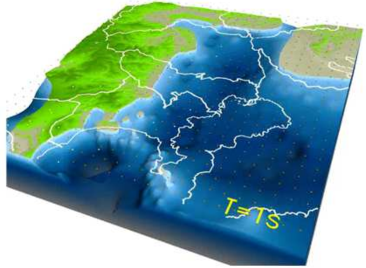


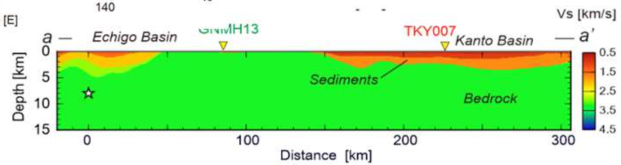
Future Directions towards Integration of (S+D+L)

- Accurate Prediction of Seismic Wave Propagation with Real-Time Data Observation/Assimilation
  - Emergency Info. for Safer Evacuation
  - 10x faster than real phenomena with $O(10^3)$ nodes of supercomputers
- 3D Underground Model
  - Heterogeneous, Observation is difficult
  - Inversion analyses of seismic waves are important for prediction of structure of underground model
  - ML may be utilized for acceleration of this prediction based on analyses of small earthquakes in normal time (e.q. Mw < 3.0)
  - More sophisticated DA method (e.g. 4DVar)









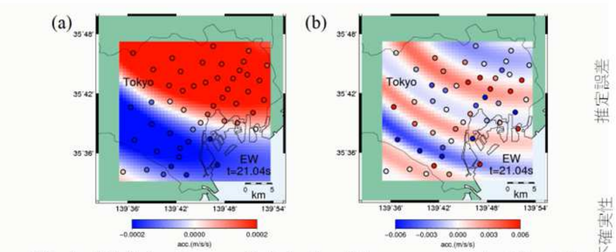
Actually, construction of 3D Underground Model by this Model for Long-Period Seismic Wave Propagation is not realistic

- Local models with smaller meshes should be used

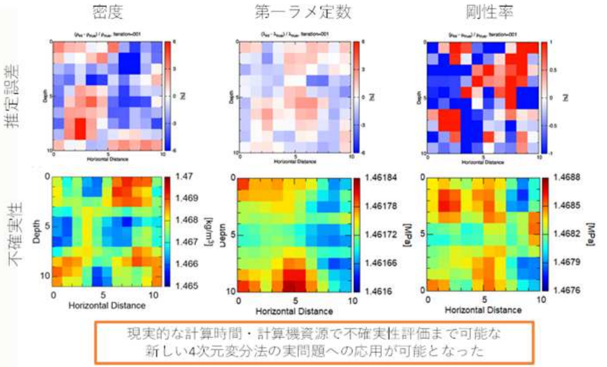
Replica Exchange Monte Carlo
Nagao et al.

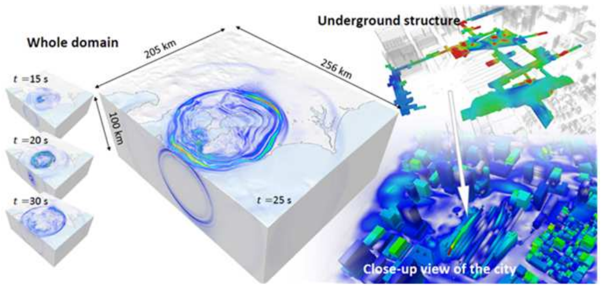
2nd Order Adjoint
Nagao et al.

Large-Scale ML
Ichimura, Fujita
SC22 GB Finalists



Movie S2. Seismic wavefield in the Tokyo area for the Mw 5.5 earthquake of 16 September 2014 in the northern Kanto area, in the frequency band (a) 0.10–0.20 Hz and (b) 0.10–1.0 Hz, computed with the optimum model parameters, compared to the observations (circles).

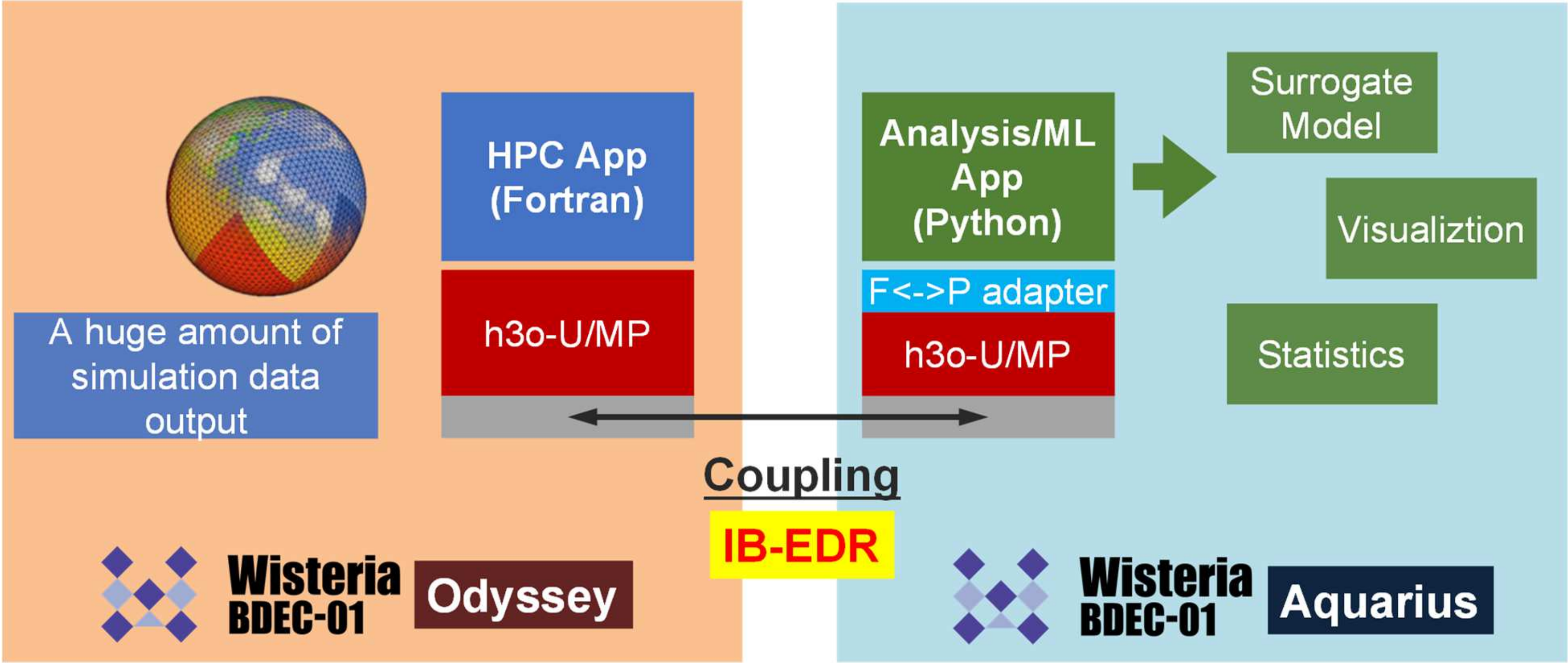




- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - h3-Open-BDEC
- **Applications on Wisteria/BDEC-01 with h3-Open-BDEC**
  - Earthquake Simulations
  - **(Global Cloud Simulation+AI) Coupling**
  - Ensemble Coupling
  - International Collaboration through JHPCN

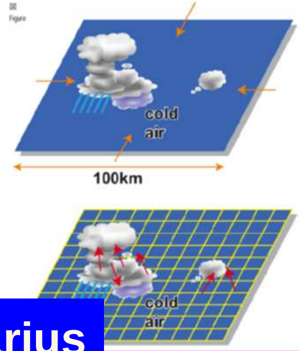
h3-Open-UTIL/MP (h3o-U/MP) Extended Multiphysics Coupler

Hierarchical, Hybrid, Heterogeneous
h3-Open-BDEC
Big Data & Extreme Computing

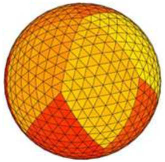


Atmosphere-ML Coupling
[Yashiro (NIES), Arakawa (ClimTech/U.Tokyo)]

- Motivation of this experiment
  - Tow types of Atmospheric models: Cloud resolving VS Cloud parameterizing
  - Could resolving model is difficult to use for climate simulation
  - Parameterized model has many assumptions
  - Replacing low-resolution cloud processes calculation with ML!



Odyssey



Phys

High Resolution Atmospheric Model (Convection-Resolving Mode)

75%

ML App (Python)

h3-Open-UTIL/MP (Coupler) + h3-Open-SYS/WaitIO-Socket

Training with high-resolution NICAM data

Replacing Physical Process in Low-Resolution NICAM with Machine Learning

Aquarius

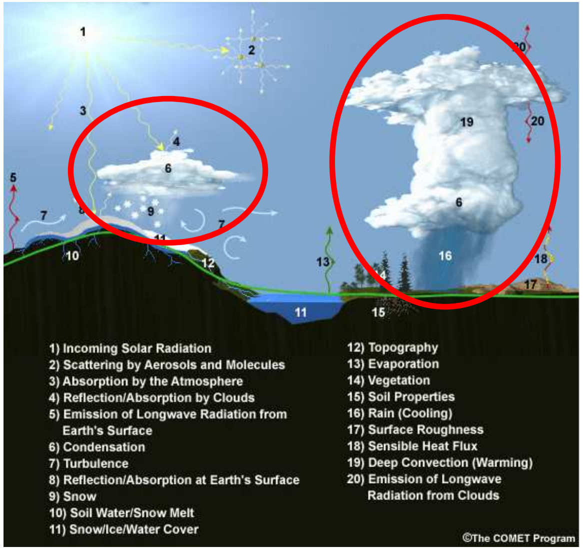
ow Resolution Atmospheric Model onvection-Parameterization Mode)

Replacing this part with AI

25% → **~0%**

Diagram of applying ML to an atmospheric model

Atmosphere-ML Coupling

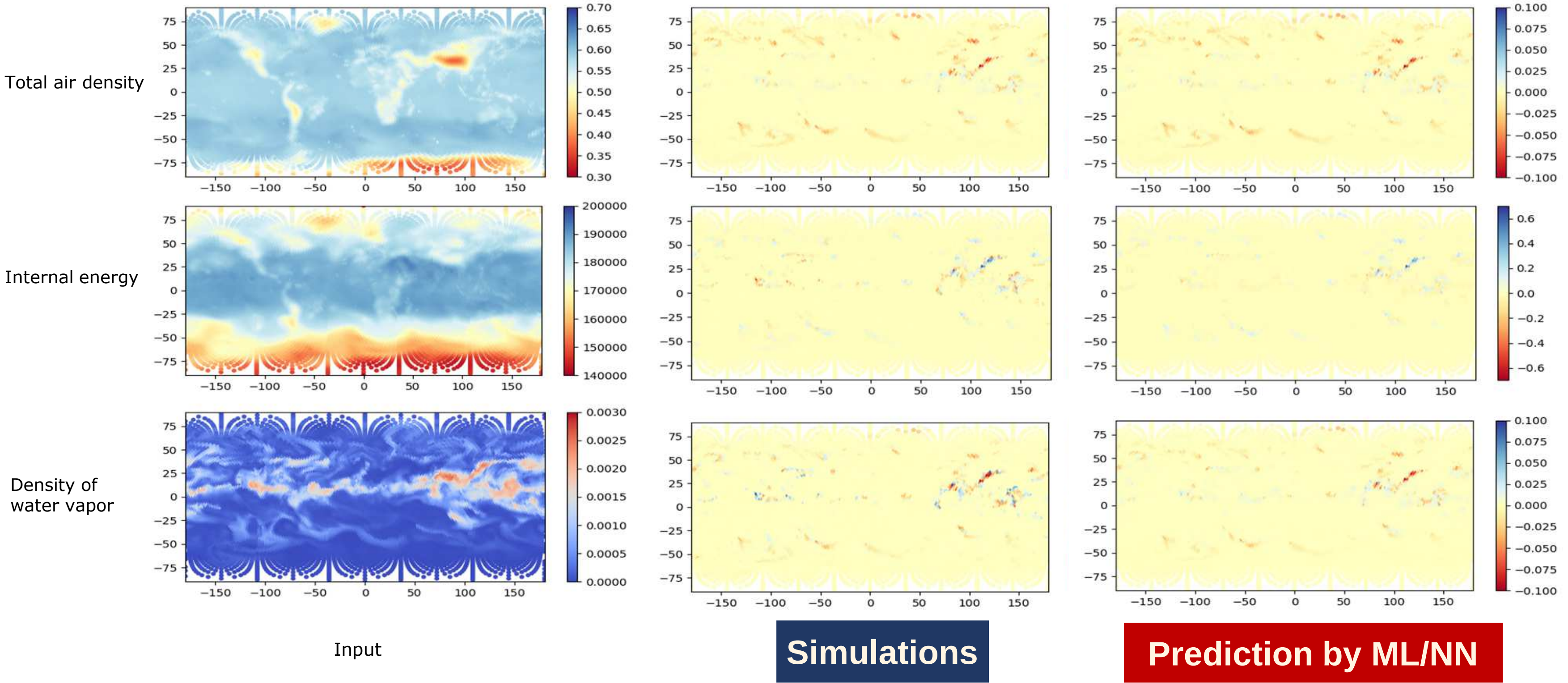


- Model component emulation (surrogation)
  - The emulation target in this study is cloud microphysical processes (phase changes, collision, coagulation, and precipitation)
  - Atmospheric pressure, temperature, and vertical distribution of water will change between before and after computing the cloud microphysical processes
- Atmospheric model and ML Library
  - NICAM (global non-hydrostatic model with icosahedral grid) + Pytorch (three layers MLP)
- Methodology
  - ML is trained to reproduce output variable from input variables of cloud microphysical subroutine
- Training data
  - Input : total air density (rho), internal energy (ein), density of water vapor (rho_q)
  - Output : tendencies of input variables computed within the cloud physics subroutine

$$\frac{\Delta rho}{\Delta T} \quad \frac{\Delta ein}{\Delta T} \quad \frac{\Delta rho\_q}{\Delta T}$$

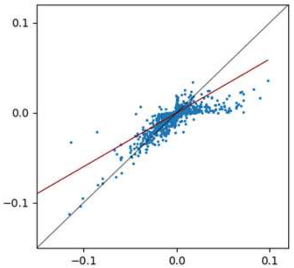
Test calculation

- Compute output variables from input variables and PyTorch
  - The rough distribution of all variables is well reproduced
  - The reproduction of extreme values is no good



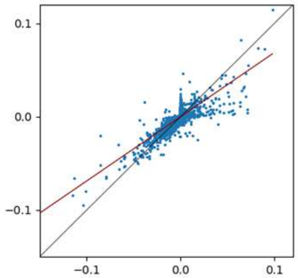
Reproducibility Improvement

- for more accurate reproducibility
  - Variable selection is important
    - NICAM subroutine mp_driver has INPUT:23, OUTPUT: 27, INOUT: 11
    - Reproducibility was improved by increasing the number of input variables to five.



d_rho calculated
from thee input variables
(rho, ein, rhoq)

	slope	intercept	coef.
d_rho	0.598	-0.0001	0.807
d_ein	0.555	-0.0004	0.798
d_rhoq	0.532	0.0000	0.781

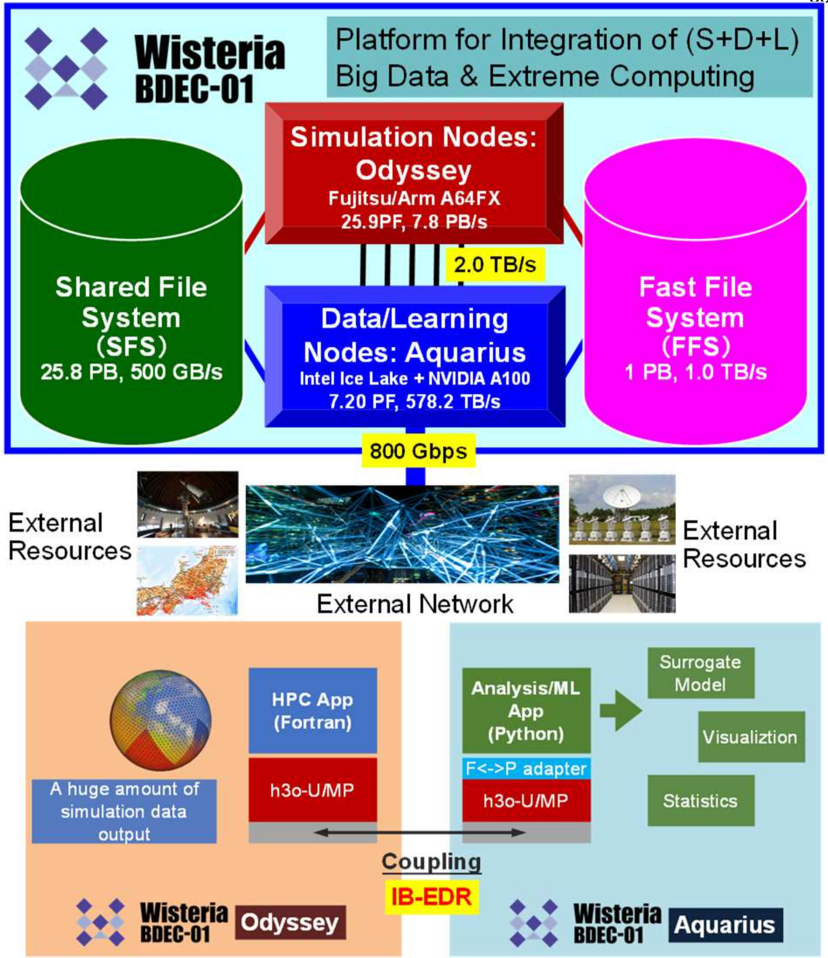


d_rho calculated
from five input variables
(three + vertical wind and precipitation)

	slope	intersept	coef.
d_rho	0.688	-0.0000	0.857
d_ein	0.710	0.0011	0.858
d_rhoq	0.692	0.0003	0.843

How to run the workloads

- Total Number of Nodes
  - Odyssey: 7,680 nodes: not so crowded
  - Aquarius: 45 nodes, 360 GPUs, very crowded
- One node of Aquarius is reserved for this type of workload on the integration of (S+D+L)
- 2 separate jobs (Odyssey, Aquarius) should be submitted
- If both jobs “grab” resources, execution starts.
- More flexible (& complicated) policy needed



Examples of Scripts [Sumimoto, Arakawa]

Odyssey for Simulation

```
#!/bin/bash
#PJM -N "test_waitio"
#PJM -L rscgrp=coupler-lec-o
#PJM -L node=10:noncont
#PJM --mpi proc=80
#PJM -L elapse=00:10:00
#PJM -g gt00
#PJM -j
#PJM -e err

module load fj
module load fjmpi
module load waitio

export WAITIO_MASTER_HOST=`hostname`
export WAITIO_MASTER_PORT=7100
export WAITIO_PBID=0
export WAITIO_NPB=2

hostname
waitio-serv-a64fx -d -m $WAITIO_MASTER_HOST

#mpiexec -oferr-proc errnicam -np 160 ./nicam
mpiexec -np 80 ./nicam
```

Aquarius for AI

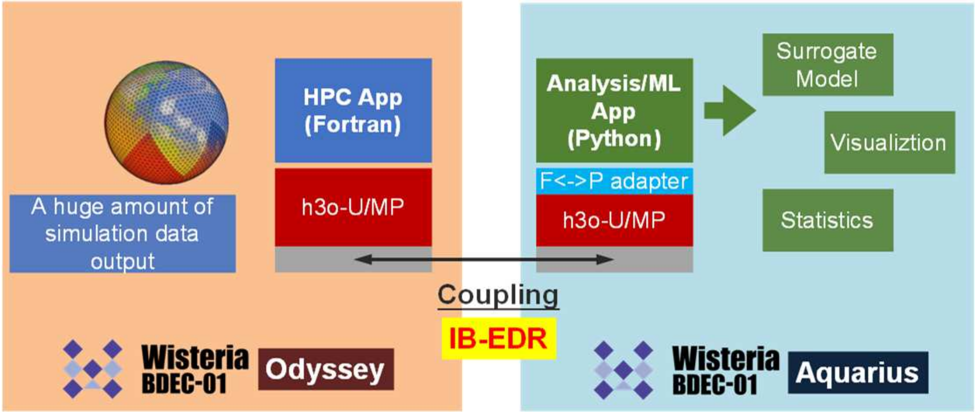
```
#!/bin/bash
#PJM -N "test_waitio"
#PJM -L rscgrp=coupler-lec-a
#PJM -L node=1
#PJM --mpi proc=10
#PJM -L elapse=00:10:00
#PJM -g gt00
#PJM -j
#PJM -e err

module unload aquarius
module unload gcc ompi
module load intel
module load impi
module load waitio

export WAITIO_MASTER_HOST=`waitio-serv -c`
export WAITIO_MASTER_PORT=7100
export WAITIO_PBID=1
export WAITIO_NPB=2

module unload intel
module unload impi
module load gcc ompi

mpiexec -n 10 ./ada
```

- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - h3-Open-BDEC
- **Applications on Wisteria/BDEC-01 with h3-Open-BDEC**
  - Earthquake Simulations
  - (Global Cloud Simulation+AI) Coupling
  - **Ensemble Coupling [Yashiro, Arakwa]**
  - International Collaboration through JHPCN

h3-Open-UTIL/MP

Multilevel Coupler/Data Assimilation
Integration of (S+D+L)

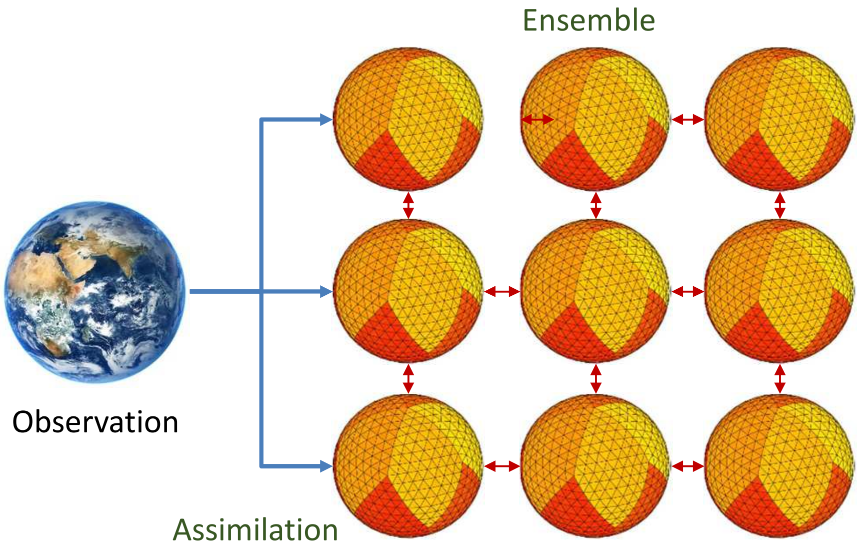
- Extended Version of Multy-Physics Coupler
- Data Assimilation (Multiple Computations: Ensemble)
  - Assimilation of Computations with Different Resolutions
  - Data Assimilation by Coupled Codes
    - e.g. Atmosphere-Ocean
- Coupling of Simulations on Odyssey and AI on Aquarius



Ensemble-based Data Assimilation (1/2)

- **<u>Ensemble-based data assimilation</u>** (in global atmospheric simulation for climate/weather prediction) combines data assimilation and ensemble simulations for accurate predictions, demanding significant computational resources for high-resolution simulations.

- 160-nodes of Odyssey are needed for running Global Atmospheric Simulation by NICAM with 14km meshes

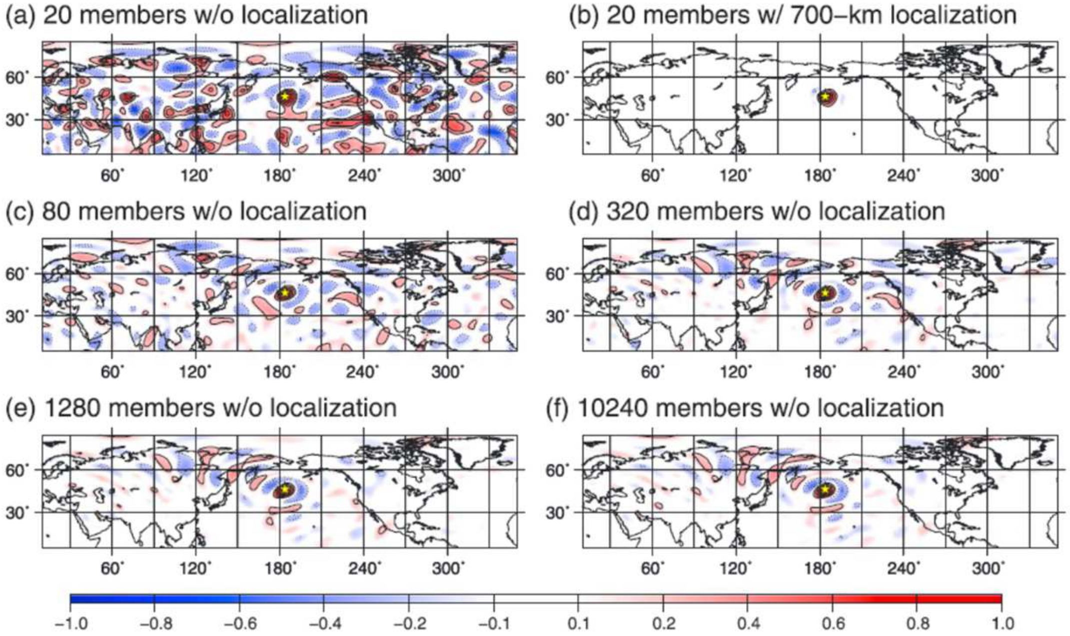


14km ensemble DA (GL09)

160 nodes x 64 members

Ensemble-based Data Assimilation (2/2)

- Usually, we do $O(10^2)$ ensembles for mid-range forecasts, while we can obtain very accurate prediction if we can do $O(10^3)$ ensembles [Miyoshi et al. 2014]

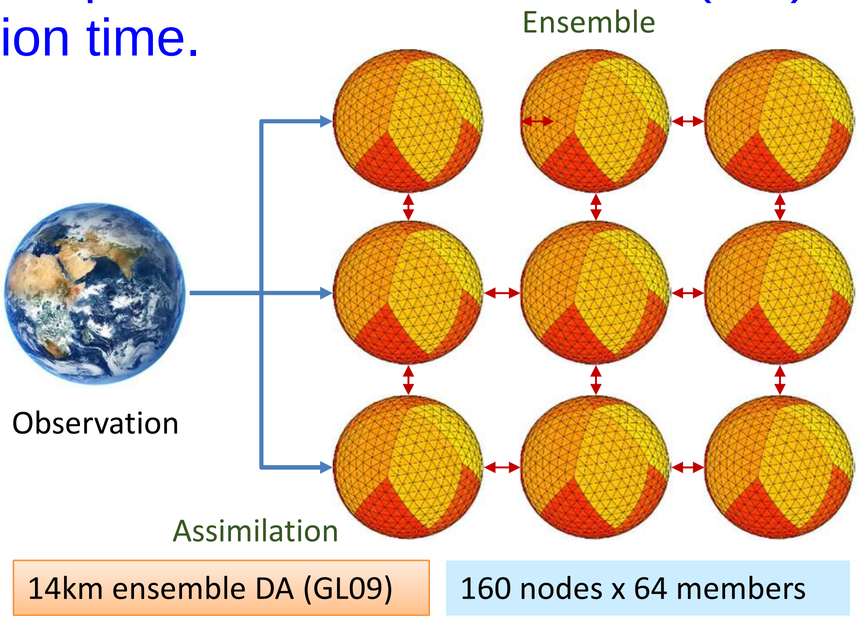


[Miyoshi et al. 2014]

Ensemble-based Data Assimilation (2/2)

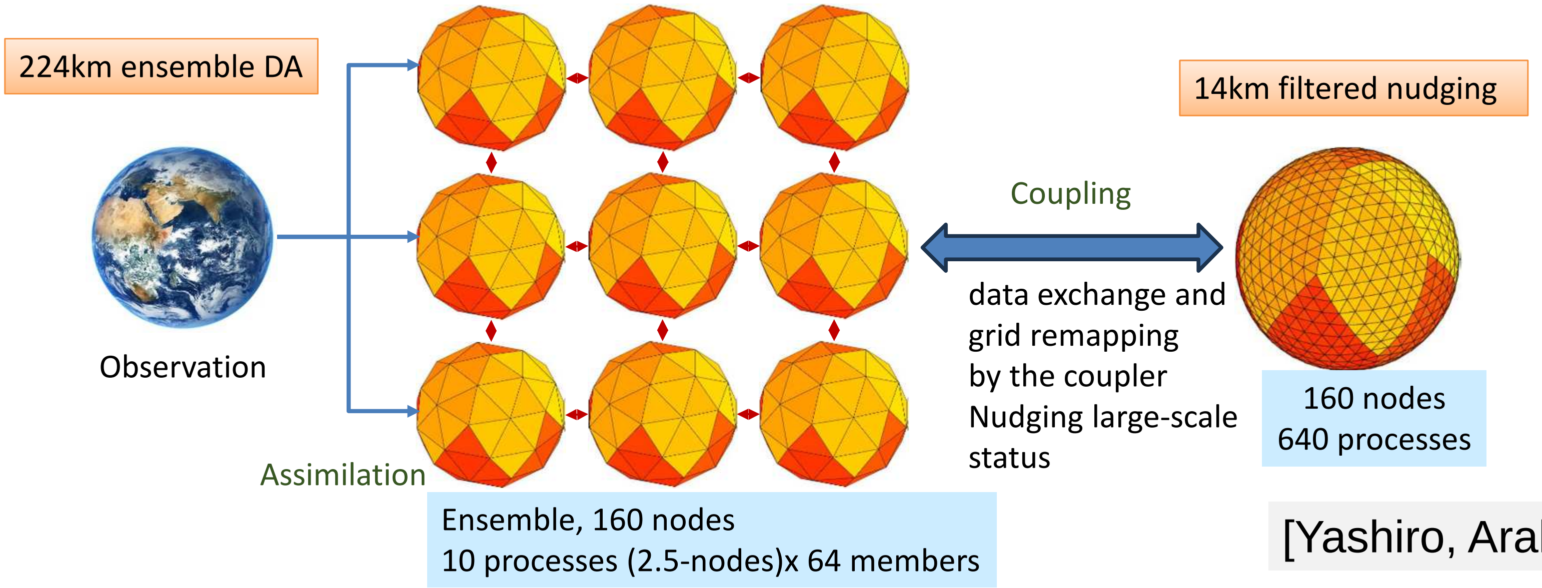
- Usually, we do O(10^2) ensembles for mid-range forecasts, while we can obtain very accurate prediction if we can do O(10^3) ensembles [Miyoshi et al. 2014]
- Currently, we do not have enough computational resources for O(10^3) ensembles in reasonable computation time.

- If we do 64 ensembles for 9-Hour Ensemble-based Data Assimilation, we need 2,240 Node-Hours (NH), using 160-nodes for each ensemble
  - ✓ 787.5 sec for each ensemble



Ensemble Coupling (1/3) Ensemble+Coupling

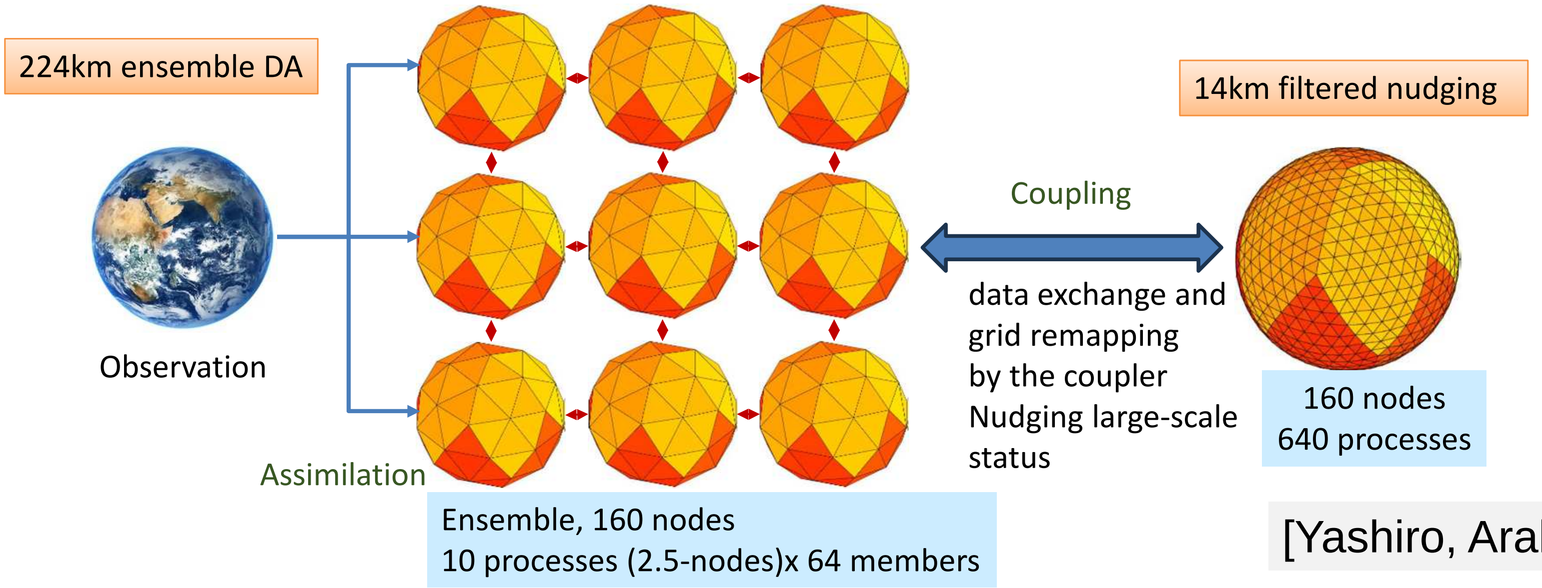
- **Coupling low-resolution ensemble data assimilation with a high-resolution simulation ⇒ reducing resource requirements.**
- In FY.2023, preliminary evaluations of ensemble coupling were conducted for 9-hour simulations by NICAM on 320 nodes of Odyssey.
  – 160-nodes for low resolution (224km), 160-nodes for high-resolution (14km)



[Yashiro, Arakawa]

Ensemble Coupling (2/3) Ensemble+Coupling

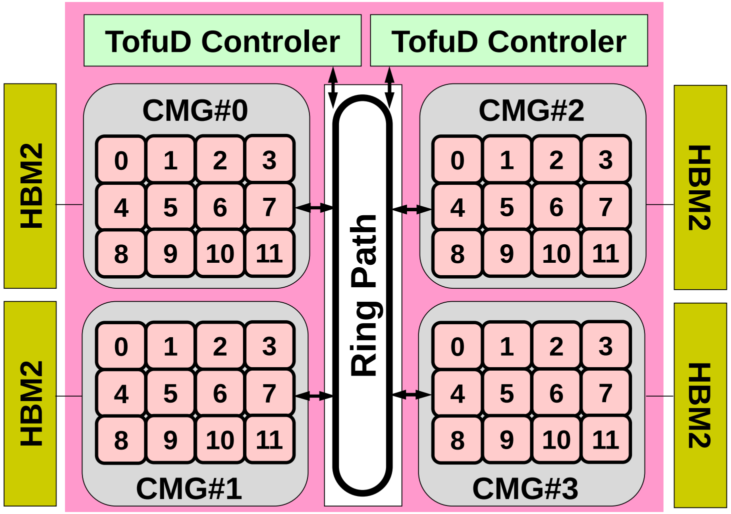
- 64 ensembles on a 224km low-resolution mesh, coupled with a 14km high-resolution mesh model
  - ✓ <u>FP32 (single-precision) was applied, while original code was by FP64.</u>
- 160-nodes for low resolution (224km): 2.5-nodes x 64-members



[Yashiro, Arakawa]

Ensemble Coupling (2/3) Ensemble+Coupling

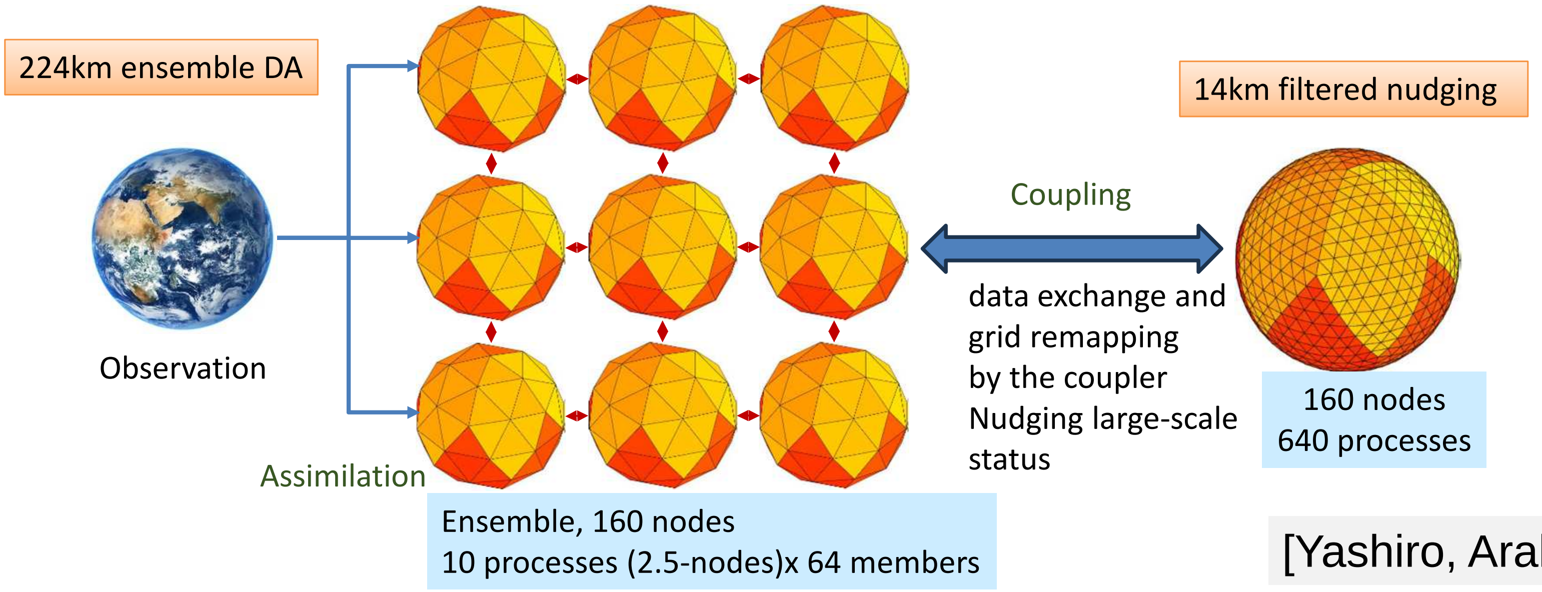
- 64 ensembles on a 224km low-resolution mesh, coupled with a 14km high-resolution mesh model
  - ✓ FP32 (single-precision) was applied, while original code was by FP64.
- 160-nodes for low resolution (224km): 2.5-nodes x 64-members
  - ✓ 10 MPI Processes for Each Case: 4 for each Compute Node of A64FX, 2.5 nodes



[Yashiro, Arakawa]

Ensemble Coupling (3/3) Ensemble+Coupling

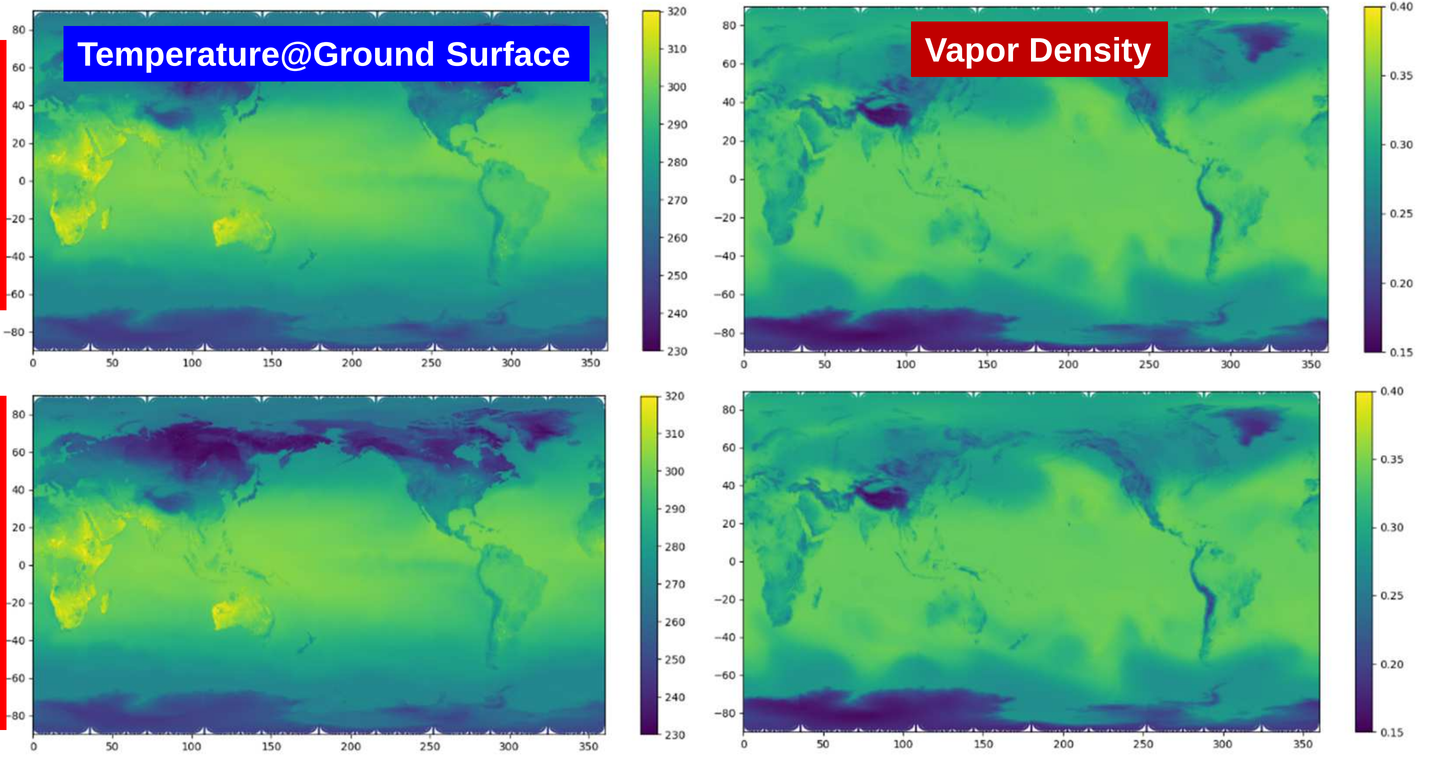
- This resulted in a performance improvement of over 100 times (with FP64⇒FP32): 2,240 NH⇒19.3 NH
- Accurate prediction by O(10^3) ensembles is possible using reasonable computational resources, and in reasonable computation time.



[Yashiro, Arakawa]

Data Assimilation
14km
2,240 NH

Ensemble-Coupling
224km+
14km
19.3 NH



- Preliminary Results for 9-Hour Integration
- Detailed verification of reproducibility requires integration over at least ONE MONTH and meteorological analyses.
- Accurate prediction by O(10^3) ensembles is possible

[Yashiro, Arakawa]

Final Goal stated in the Proposal of h3-Open-BDEC (Nov. 2018)

- We aim to reduce the amount of computations and power consumption **by more than 10 times** while maintaining the same accuracy as conventional methods in multi-level simulations that integrate (S+D+L).
  - Mixed Precision/Adaptive Precision
  - Machine Learning, Hierarchical Data Driven Approach
  - Heterogeneous Computing

- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - h3-Open-BDEC
- **Applications on Wisteria/BDEC-01 with h3-Open-BDEC**
  - Earthquake Simulations
  - (Global Cloud Simulation+AI) Coupling
  - Ensemble Coupling [Yashiro, Arakwa]
  - **International Collaboration through JHPCN**

JHPCN

https://jhpcn-kyoten.itc.u-tokyo.ac.jp/en/

- Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (2010-)
- Alliance of SC Centers of 8 National Universities in Japan
  - 7 “Imperial” Universities + Tokyo Tech
  - Core Institute: ITC/U.Tokyo
  - Total 185+PFLOPS (April 2024)
- MoU with NHR/Germany since July 11, 2024
- Promotion of collaborative (fundamental, interdisciplinary) research projects using facilities & human resources in 8 Centers
  - Proposal-based, Resources of Supercomputers are awarded for accepted proposals

FY.2023-2025, JHPCN Project

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

- ✓ Jülich Supercomputing Centre（JSC）
- ✓ Rudjer Boskovic Institute, Centre for Informatics and Computing, Croatia
- ✓ Friedrich-Alexander-Universität Erlangen-Nürnberg（FAU）
- ✓ French Atomic Energy Commission (CEA)
- ✓ Bergische Universität Wuppertal (BUW)
- ✓ Karlsruher Institut für Technologie (KIT)

(FY.2023:35, FY.2024:49)

History & Plans

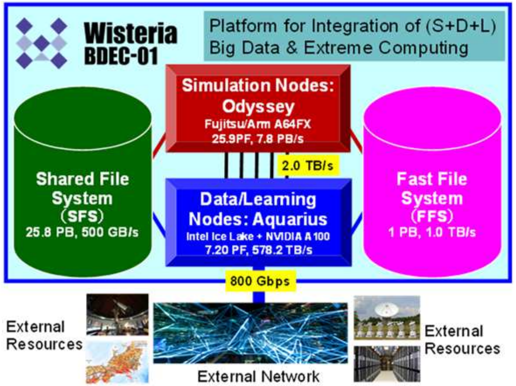
https://jhpcn-kyoten.itc.u-tokyo.ac.jp/en/

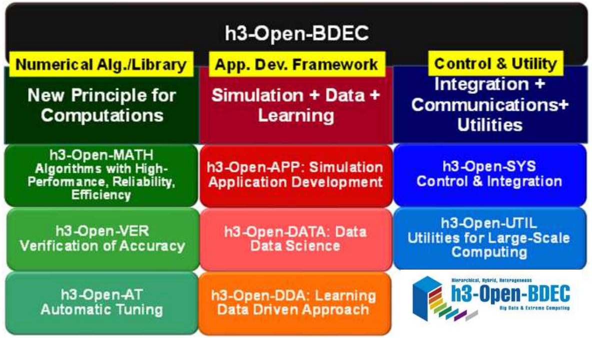
- Innovative Computational Science by Integration of Simulation/Data/Learning on Heterogeneous Supercomputers
  - FY.2021 & 2022: Focused on Earthquake Simulations
    - Univ. Tokyo (ITC, ERI), Nagoya U., Kyushu U., NIES, Fujitsu
  - FY.2023-2025 (plan): Other applications and International Collaborations, Popularization of SW usage (e.g. WaitIO, Coupler)
    - Jülich Supercomputing Centre（JSC）：Modular Supercomputing
    - Rudjer Boskovic Institute, Centre for Informatics and Computing, Croatia
    - Friedrich-Alexander-Universität Erlangen-Nürnberg（FAU）
    - French Atomic Energy Commission (CEA)
- Target Systems in Japan
  - Wistreia/BDEC-01，Flow@Nagoya U., mdx

Collaborations related to Heterogeneous Computing (U.Tokyo-JSC)

Wisteria/BDEC-01
h3-Open-BDEC
U.Tokyo
K. Nakajima et al.

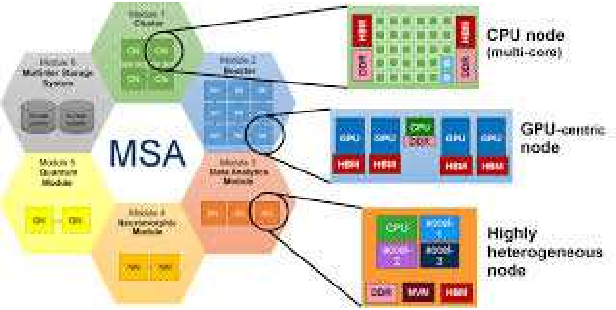
東京大学
THE UNIVERSITY OF TOKYO

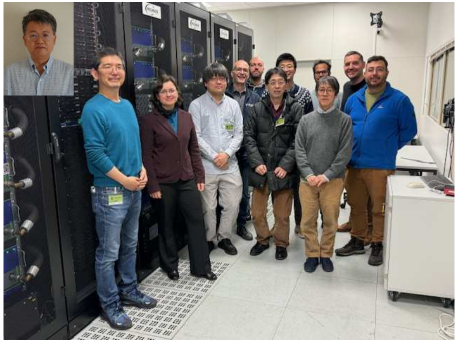




Modular Supercomputing Architecture (MSA), JSC
E. Suarez et al.

JÜLICH
FORSCHUNGSZENTRUM

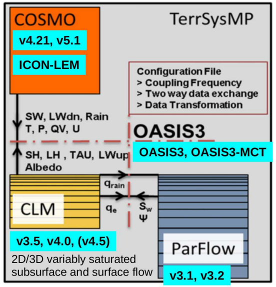


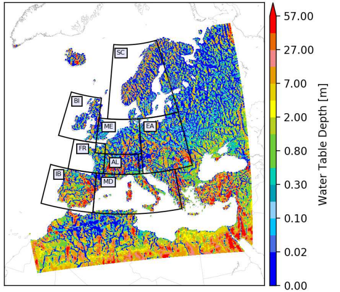


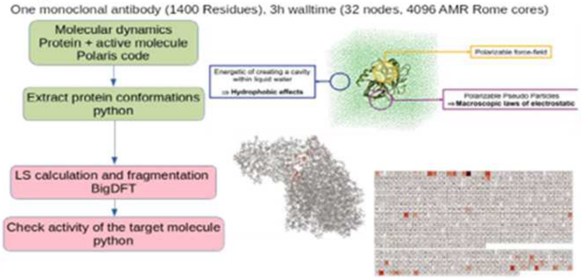
JHPCN WS Mar.13-15, JSC

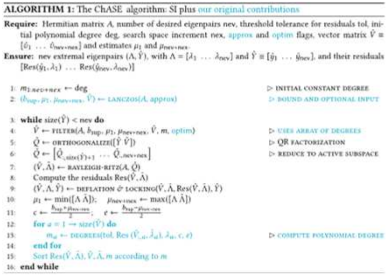
Target Applications

- JSC
  - Terrestrial Systems Modeling Platform (TSMP)
    - Coupling: Groundwater Flow & Atmosphere
    - https://www.terrsysmp.org/
  - Chebyshev Accelerated Subspace Eigensolver (ChASE)
    - Quantum Chemistry, Heterogeneous Environment
    - https://github.com/ChASE-library
  - Brain Aneurysm Simulations
    - Multiscale, Multiphysics
    - CFD Codes (m-AIA) at JSC
      - https://www.hpccoe.eu/2021/06/04/m-aia/
- CEA
  - Selection of inhibitors of the SARS-CoV-2 Main Protease
    - BigDFT + Polaris/GENESIS
  - Earthquake Simulation (PSD), ML with Causality



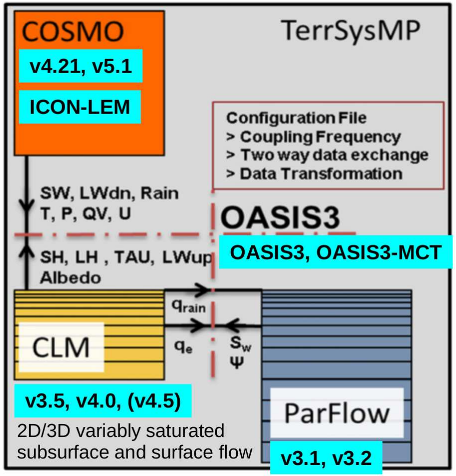


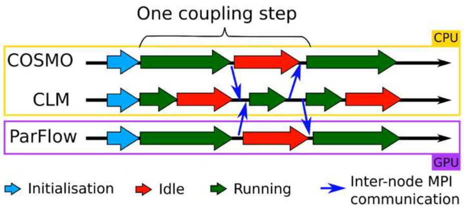




Terrestrial Systems Modeling Platform (TSMP) (JSC, U.Tokyo)

- TSMP is a scale-consistent, highly modular, massively parallel, fully integrated soil-vegetation-atmosphere modeling system by JSC.
- **Our target is coupling COSMO/ICON (Atmosphere)-ParFlow (Surface/Subsurface Flow)-CLM(Land Surface Model).**
  - **The coupling of 3 models has been already done using OASIS3 library on CPU-GPU heterogeneous environment.**
- **In this project, we replace OASIS3 with h3-Open-BDEC, and coupled simulations will be possible on really heterogeneous systems, such as Wisteria/BDEC-01.**
  - In FY.2023, we mainly ported codes to Odyssey and made preliminary evaluations.
  - In FY.2024, we fucus on replacing OASIS3 with h3-Open-BDEC, devel preliminary version of the coupled codes, and conduct preliminary evaluations on Wisteria/BDEC-01.



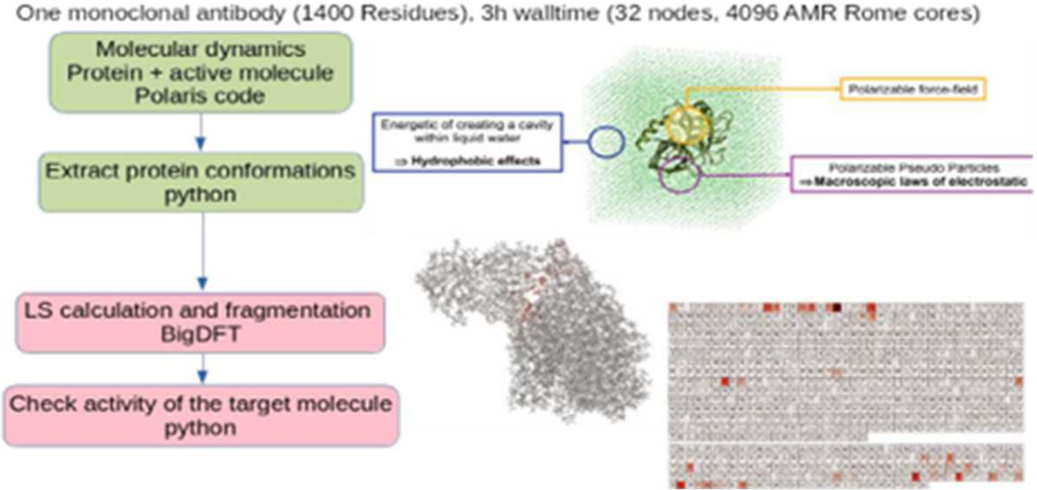


Big-DFT with GENESIS for SARS-CoV-2 Main Protease (CEA, RIKEN, U.Tokyo) (1/2)

- Developing medicines for viruses like SARS-CoV-2 faces challenges, including drug resistance (SARS-CoV-2: Virus, COVID-19: Infection)
  - Understanding and predicting drug resistance involves modeling structural changes from point mutations, utilizing long trajectories from classical molecular dynamics (MD/MM).
  - Mechanistic insight into mutation effects can benefit from quantum mechanical (QM) modeling.

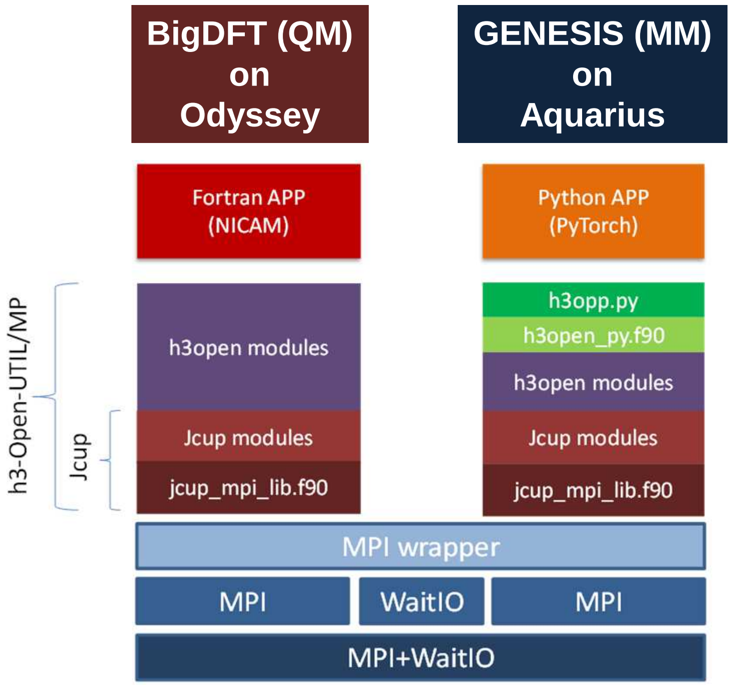
GENESIS (RIKEN)
Polaris (CEA)
(MM/MD)

BigDFT (QM)



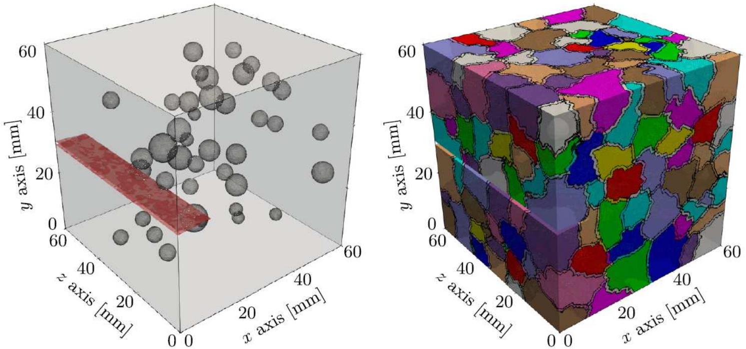
Big-DFT with GENESIS for SARS-CoV-2 Main Protease (CEA, RIKEN, U.Tokyo) (2/2)

- **In this project, we will exploit the heterogeneous architecture of Wisteria/BDEC-01 to build a coupled QM-MM workflow.**
  - **The MM workflow will run the "GENESIS" (RIKEN) on Aquarius to exploit its GPU nodes and provide samples from a trajectory that are sent to the QM-MM workflow running "BigDFT" on Odyssey.**
  - **BigDFT was already optimized for A64FX architecture under CEA-RIKEN collaboration.**
- In FY.2024, we will construct preliminary version of QM-MM workflow using h3-Open-BDEC on Wisteria/BDEC-01, and make evaluations.



PSD (1/3)
Parallel Seismic Dynamics

- Since FY.2021 (or before), we have continued to advance FDM-based Seism3D/OpenSWPC-DAF by h3-Open-BDEC on Wisteria/BDEC-01.
  - Using observation data from JDXnet, we have achieved a 3D seismic wave simulation by combining it with real-time data assimilation.
- PSD developed by CEA is a massively parallel 3D seismic wave propagation simulation code based on FEM and implicit time-marching
  - Seism3D: FDM, structured, explicit time-marching
  - PSD: FEM, unstructured (tetrahedron), implicit time-marching



PSD (2/3)
Parallel Seismic Dynamics

- In FY.2024, we will enhance PSD with data assimilation through optimal interpolation technique.
- This enables PSD to perform 3D simulations with real-time data assimilation.
  – We validate results by comparing PSD simulations with Seism3D/OpenSWPC-DAF.

PSD (3/3) Parallel Seismic Dynamics

- Furthermore, we explore machine learning-based earthquake propagation prediction using **causality** from combined simulation results.
  - Unlike other machine and deep learning methods that are based on correlations between events, causality is a method that learns the cause-effect relationships between variables that provides more realistic links between them and the necessary information to intervene on the phenomena.
  - In this context, our goal is to enhance either the time and computational efficiency of earthquake simulations or gain a deeper understanding of the data.
  - We plan to achieve this by utilizing earthquake data obtained from either FEM simulations or real-world sensors, employing causality learning methods.
- **ML with Causality easily detects sensor errors and such observations are excluded for DA. Moreover, it can automatically select optimum set of sensors for optimum interpolation.**

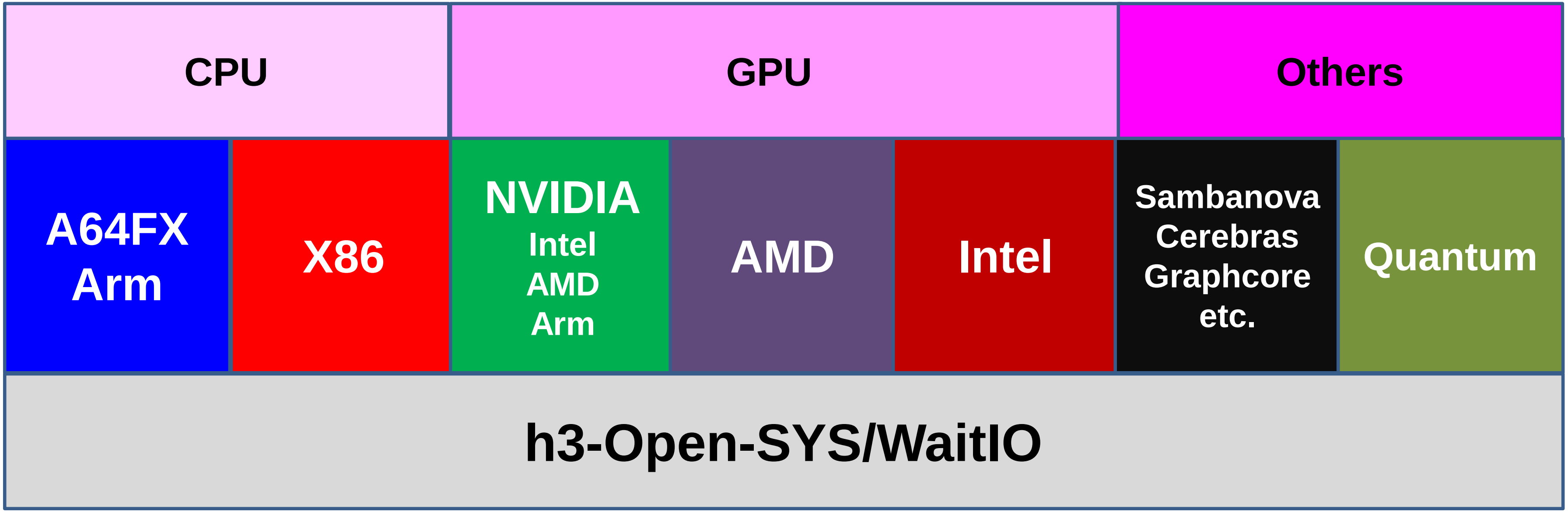
- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - h3-Open-BDEC
- Applications on Wisteria/BDEC-01 with h3-Open-BDEC
- **Integration of (Simulation/Data/Learning) and Beyond**
- Summary

Integration of Simulation/Data/Learning and Beyond

Kengo Nakajima
Information Technology Center
The University of Tokyo
RIKEN R-CCS

16th World Congress on Computational Mechanics & 4th Pan American Congress on Computational Mechanics (WCCM-PANACM Vancouver 2024)
Vancouver, B.C., Canada, July 23, 2024

Anything is possible with WaitIO
WaitIO over Internet/cloud is possible



"JHPC-Quantum" for QC-HPC Hybrid Platform:

Research & Development of Quantum/HPC Hybrid Platform for Exploring the Computable Domain (FY.2023-2028)

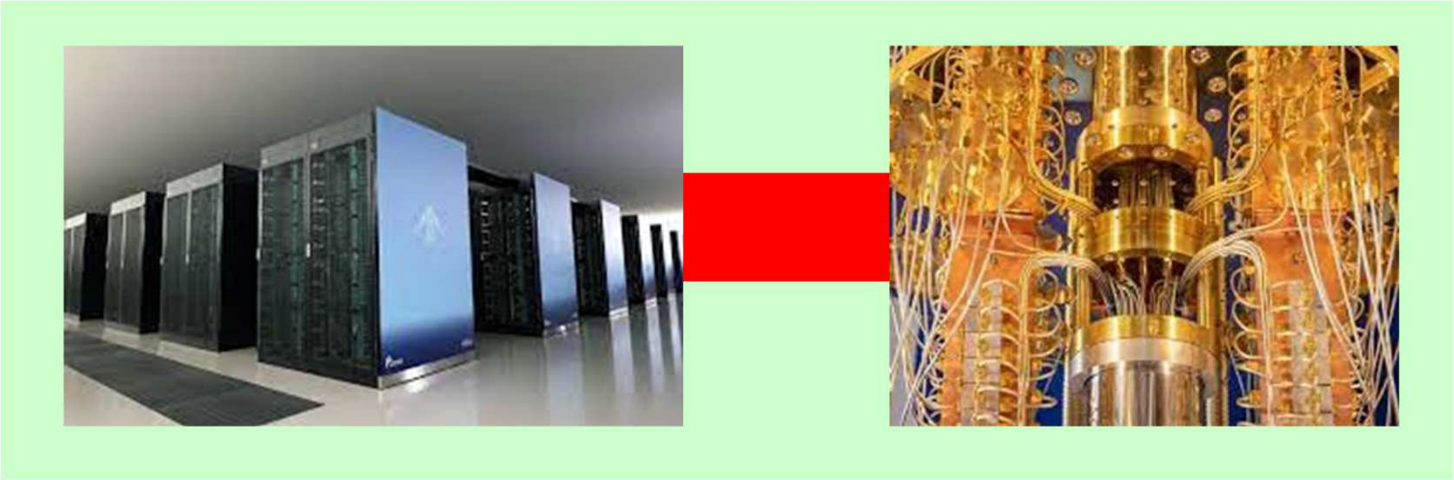
- **RIKEN R-CCS, SoftBank**
  - **Leading PI: Prof. Mitsuhisa Sato (RIKEN R-CCS)**
  - **Cooperating Organizations: U.Tokyo, Osaka U.**
- Supported by New Energy & Industrial Technology Development Organization (NEDO): Post-5G Project
  - This project has a strong focus on industrial applications.
  - FY.2023-2028 (5 Years)
- **Two Real Quantum Computers will be installed**
  - **IBM's Superconducting QC at RIKEN-Kobe (100+Qubit)**
  - **Quantinuum's Ion-Trap QC at RIKEN-Wako (20+Qubit)**
- **Target Applications**
  - **Quantum Physics, Error Mitigation, Quantum ML**

新エネルギー・産業技術総合開発機構
New Energy and Industrial Technology Development Organization

Post-5G Project
ポスト5G情報通信システム
基盤強化研究開発事業

System SW for QC-HPC Hybrid Environment (1/2)

- **Quantum Computer as Accelerator of Supercomputers**
  - **QC-HPC Hybrid**

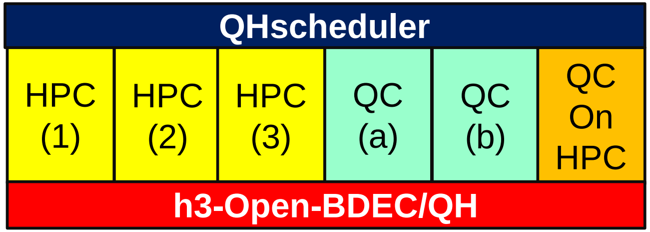


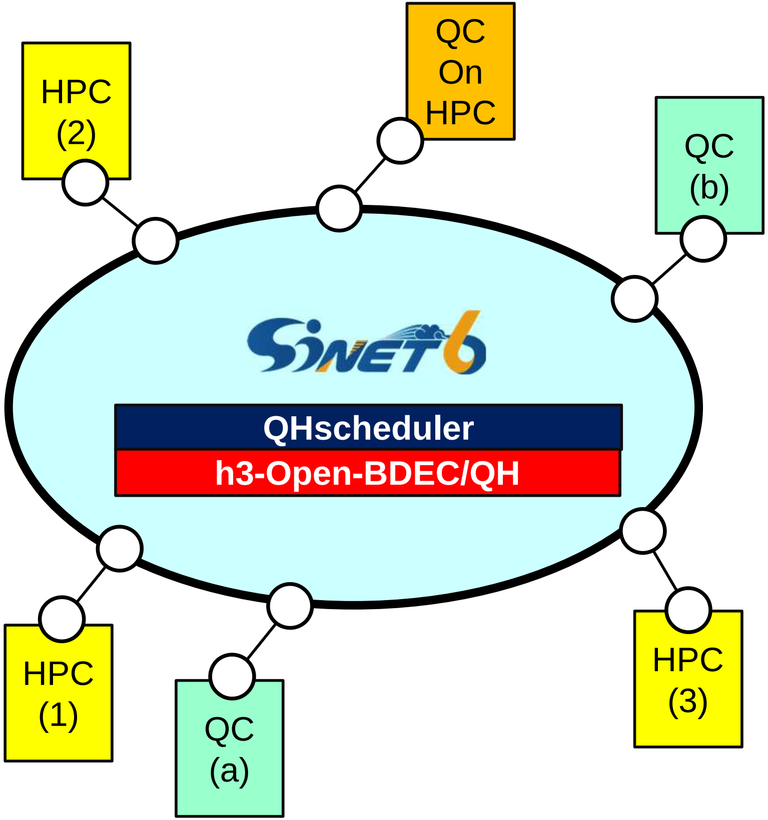
- **Role of U.Tokyo**
  - **R&D on System SW for QC-HPC Hybrid Environment**
  - **Extension of h3-Open-BDEC**

JHPC-quantum

System SW for QC-HPC Hybrid Environment (2/2)

- System SW for Efficient & Smooth Op. of QC-HPC Hybrid Environment
  - <u>QHscheduler</u>: A job scheduler that can simultaneously use multiple computer resources distributed in remote locations
  - **<u>h3-Open-BDEC/QH</u>: Coupling to efficiently implement and integrate communication and data transfer between QC-HPC on-line and in real time: Extension of WaitIO, Coupler**



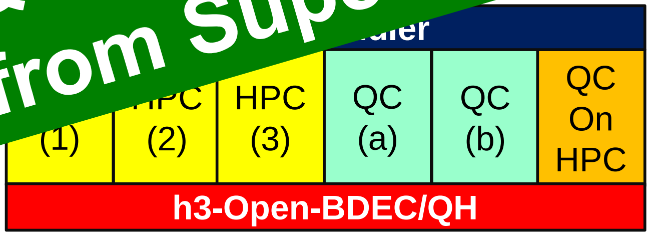


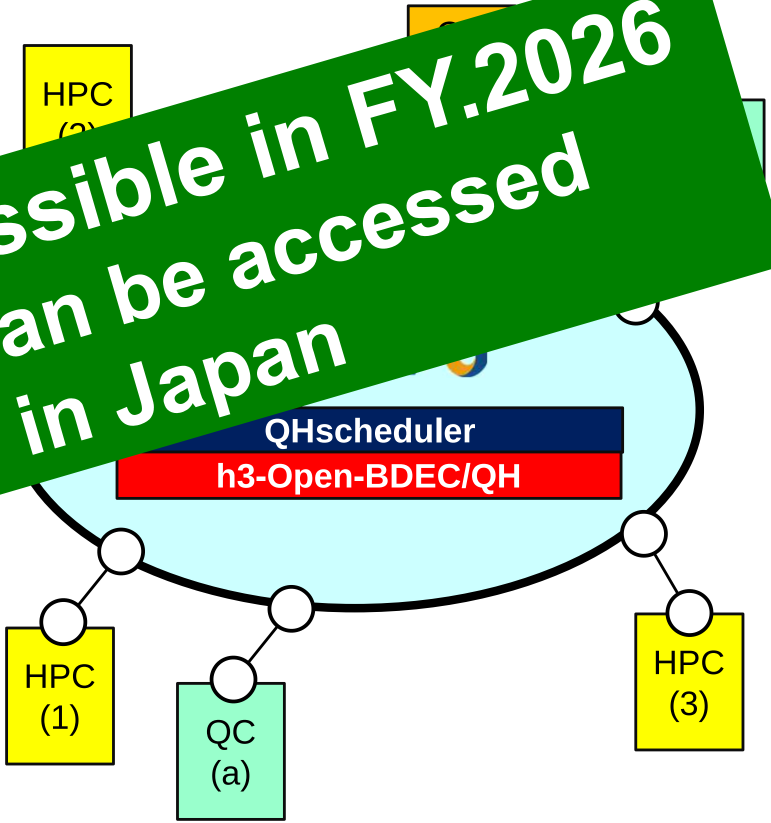
System SW for QC-HPC Hybrid Environment (2/2)

- System SW for Efficient & Smooth Op. of QC-HPC Hybrid Environment
  - QHscheduler: A job scheduler that can simultaneously use multiple computa… resources distributed in re…
  - **h3-Open-BDEC/… effici…**

…pler

QC-HPC Hybrid is possible in FY.2026
Quantum Computers can be accessed from Supercomputers in Japan



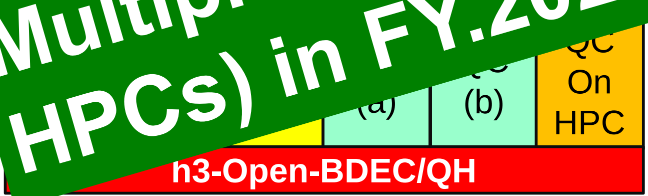


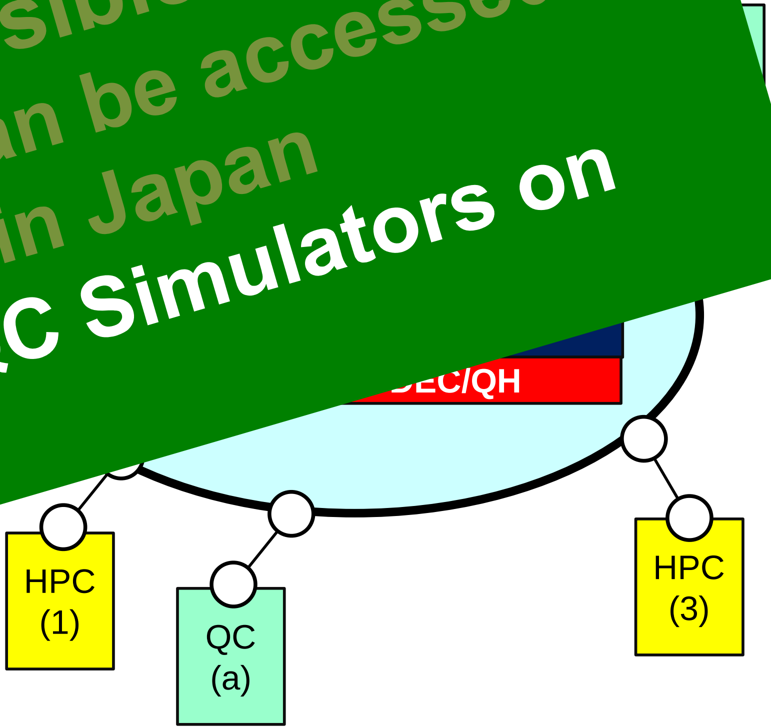
System SW for QC-HPC Hybrid Environm (2/2)

- System SW for Efficient & Smooth Op. of QC-HPC Hybrid Environment
  - QHscheduler: A job schedu
    simultaneously us
    resourc

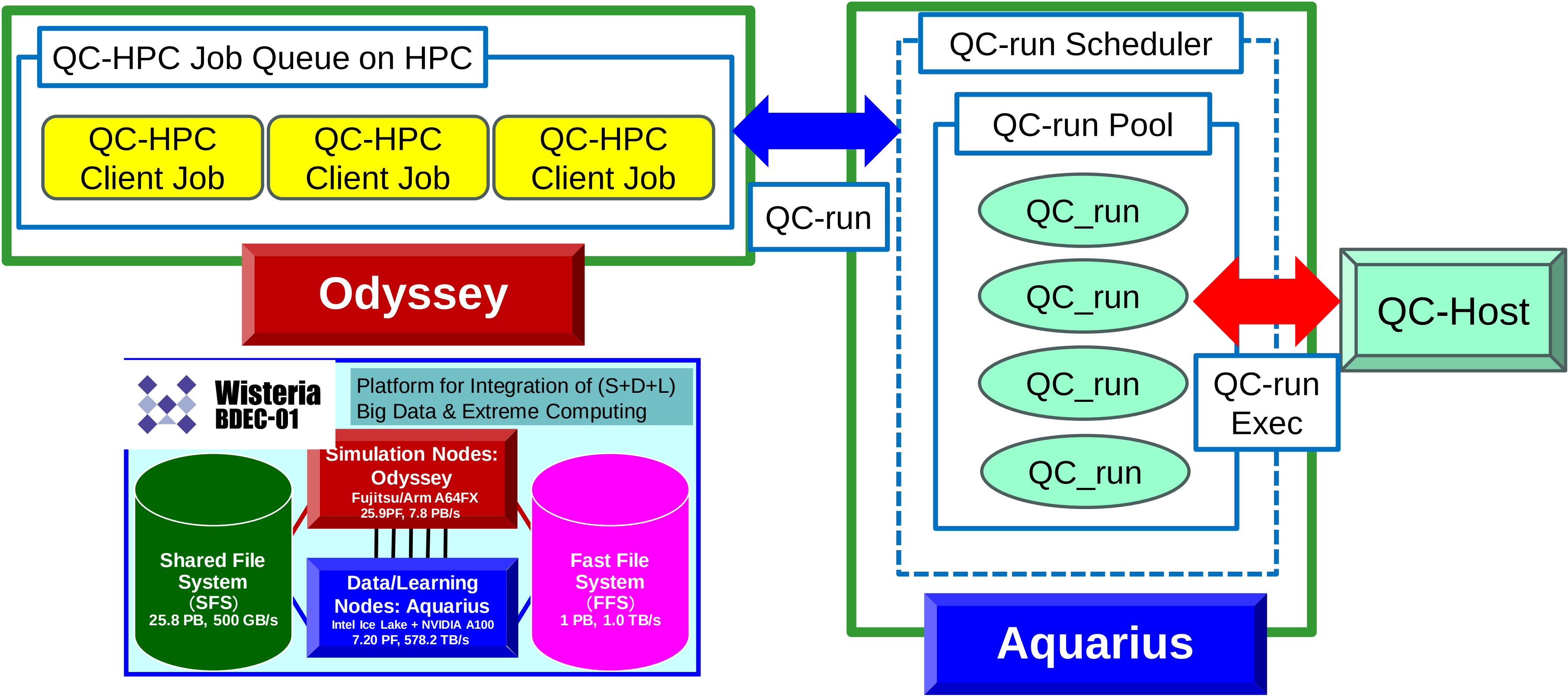
QC-HPC Hybrid is possible in FY.2026
Quantum Computers can be accessed from Supercomputers in Japan

Multiple HPCs+QC+(QC Simulators on HPCs) in FY.2028





How to run QC-HPC Hybrid Workloads Prof. M.Sato (RIKEN)



- Integration of (Simulation/Data/Learning)
  - Wisteria/BDEC-01
  - h3-Open-BDEC
- Applications on Wisteria/BDEC-01 with h3-Open-BDEC
- Integration of (Simulation/Data/Learning) and Beyond
- **Summary**

Summary

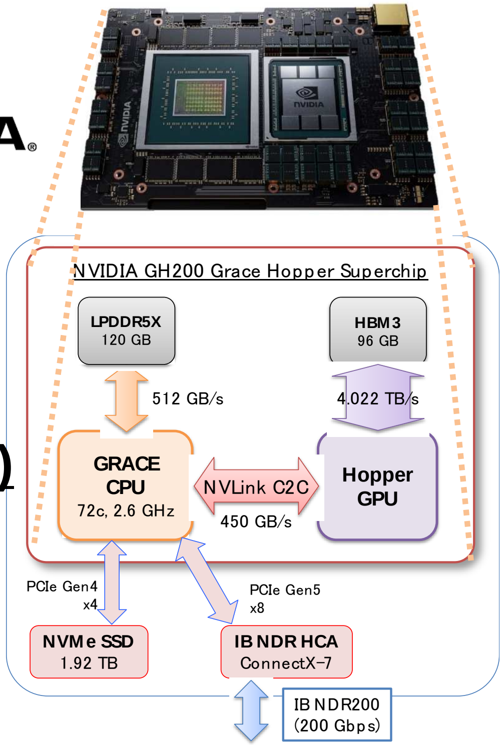
- Integration of (Simulation/Data/Learning) at ITC/U.Tokyo
- Wisteria/BDEC-01
- h3-Open-BDEC
- Applications
- Challenges towards Quantum Computing

- Collaborations are welcome
  - nakajima@cc.u-tokyo.ac.jp

OFP-II: Miyabi (1/2)

Operation starts in January 2025

- **<u>Acc-Group: CPU+GPU: NVIDIA GH200</u>**
  - Node: NVIDIA GH200 Grace-Hopper Superchip
    - Grace: 72c, 3.456 TF, 120 GB, 512 GB/sec (LPDDR5X)
    - H100: 66.9 TF DP-Tensor Core, 96 GB, 4,022 GB/sec (HBM3)
      - Cache Coherent between CPU-GPU
    - NVMe SSD for each GPU: 1.9TB, 8.0GB/sec, GPUDirect Storage
  - **Total (Aggregated Performance: CPU+GPU)**
    - **1,120 nodes, 78.8 PF, 5.07 PB/sec, IB-NDR 200**
- **<u>CPU-Group: CPU Only: Intel Xeon Max 9480 (SPR)</u>**
  - Node: Intel Xeon Max 9480 (1.9 GHz, 56c) x 2
    - 6.8 TF, 128 GiB, 3,200 GB/sec (<u>HBM2e only</u>)
  - **Total**
    - **190 nodes, 1.3 PF, IB-NDR 200**
    - **372 TB/sec for STREAM Triad (Peak: 608 TB/sec)**



OFP-II: Miyabi (2/2)

Operation starts in January 2025

- **File System: DDN EXA Scalar, Lustre FS**
  - 11.3 PB (NVMe SSD) 1.0TB/sec, "Ipomoea-01" with 26 PB is also available
- **All nodes are connected with Full Bisection Bandwidth**
  - **(400Gbps/8)×(32×20＋16×1) = 32.8 TB/sec**
- **Operation starts in January 2025, h3-Open-SYS/WaitoIO will be adopted for communication between Acc-Group and CPU-Group**

IB-NDR（400Gbps）

IB-NDR200（200）

IB-HDR（200）

Acc-Group
NVIDIA GH200 1,120
78.2 PF, 5.07 PB/sec

CPU-Group
Intel Xeon Max (HBM2e) 2 x 190
1.3 PF, 608 TB/sec

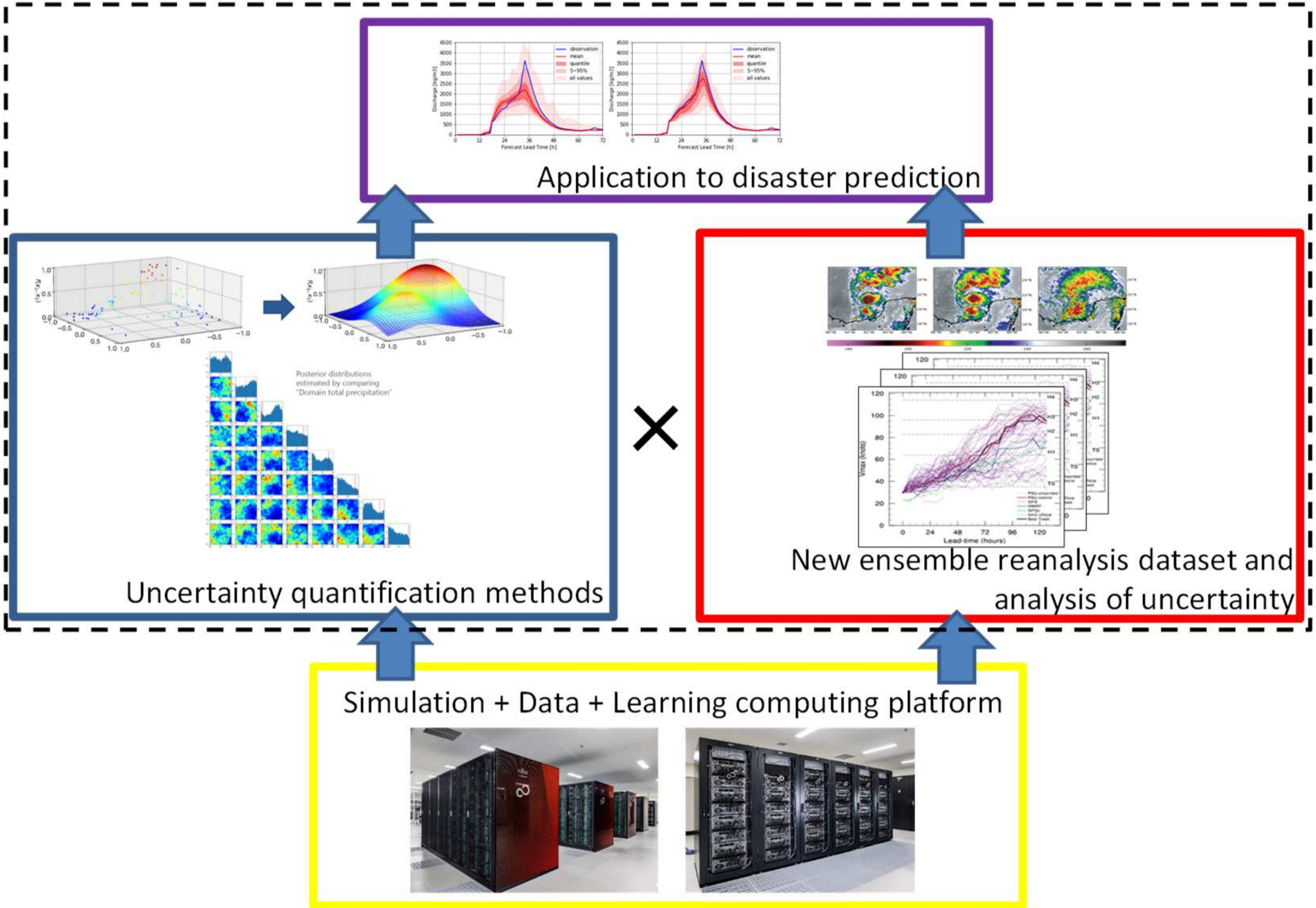
File System
DDN EXA Scaler
11.3 PB, 1.0TB/sec

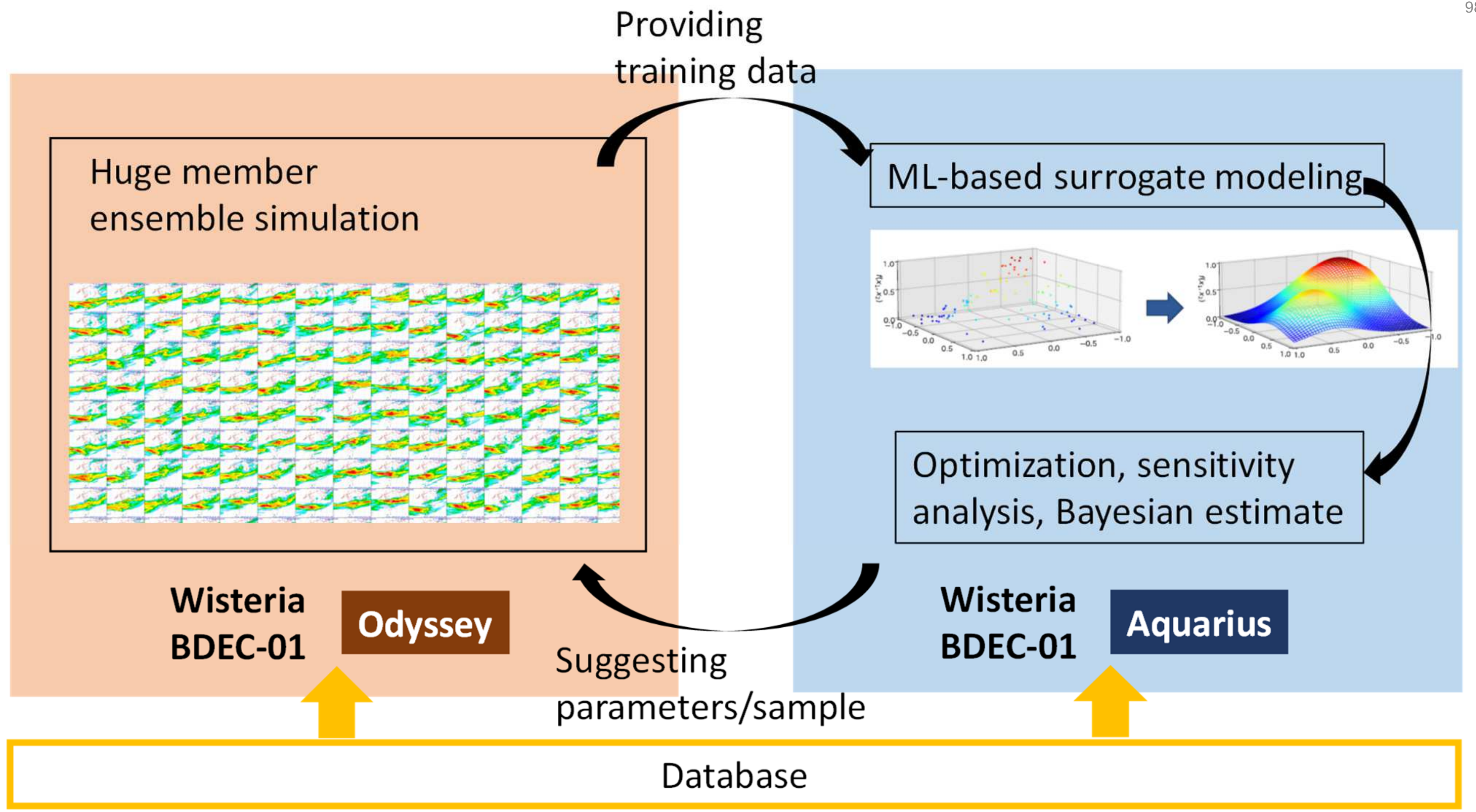
Ipomoea-01
Common Shared Storage
26 PB

Uncertainty Quantification of Extreme Weather Prediction

Y. Sawada (U.Tokyo)

- Predicting extreme weather phenomena that lead to flooding and inundation remains highly uncertain.
- While existing research has focused on analyzing uncertainties related to initial conditions and boundary values in large-scale weather simulations, a comprehensive understanding of all sources of uncertainty within these simulations is crucial.
- In this study, our goal is to construct a software framework for efficiently estimating all inherent uncertainties in large-scale weather simulations using Bayesian methods.
- Additionally, we aim to create and publicly share large-scale weather data with added uncertainty information, enabling an investigation into the origins of uncertainty in predicting extreme weather events.
- By maximizing the performance of Wisteria/BDEC-01, we address this challenging task.


Application to disaster prediction
Uncertainty quantification methods
Posterior distributions
estimated by comparing
"Domain total precipitation"
×
New ensemble reanalysis dataset and
analysis of uncertainty
Simulation + Data + Learning computing platform


Providing training data
Huge member ensemble simulation
ML-based surrogate modeling
Optimization, sensitivity analysis, Bayesian estimate
Wisteria BDEC-01
Odyssey
Wisteria BDEC-01
Aquarius
Suggesting parameters/sample
Database

FY.2024 Proposal

- (Leading-PI) Kengo Nakajima (ITC/U.Tokyo)
- (Co-PI) Takashi Furumura (ERI/U.Tokyo)
- (Co-PI) France Boillod-Cerneux (CEA)
- (Co-PI) Edoardo Di Napoli (JSC)