東京大学
THE UNIVERSITY OF TOKYO

RIKEN R-CCS

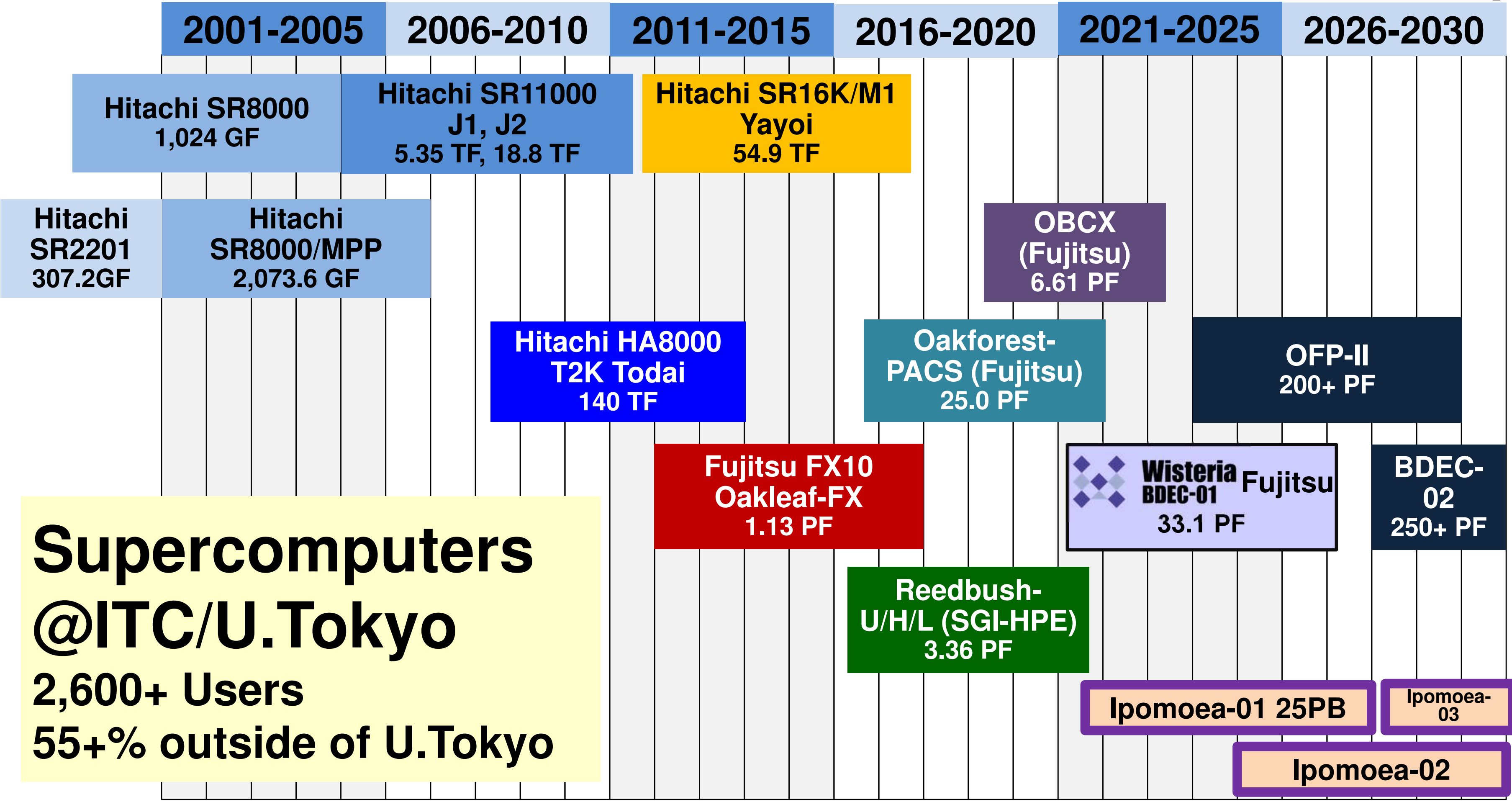
# Integration of (Simulation/Data/Learning) using h3-Open-BEDC on Wisteria/BDEC-01


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


Wisteria BDEC-01

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

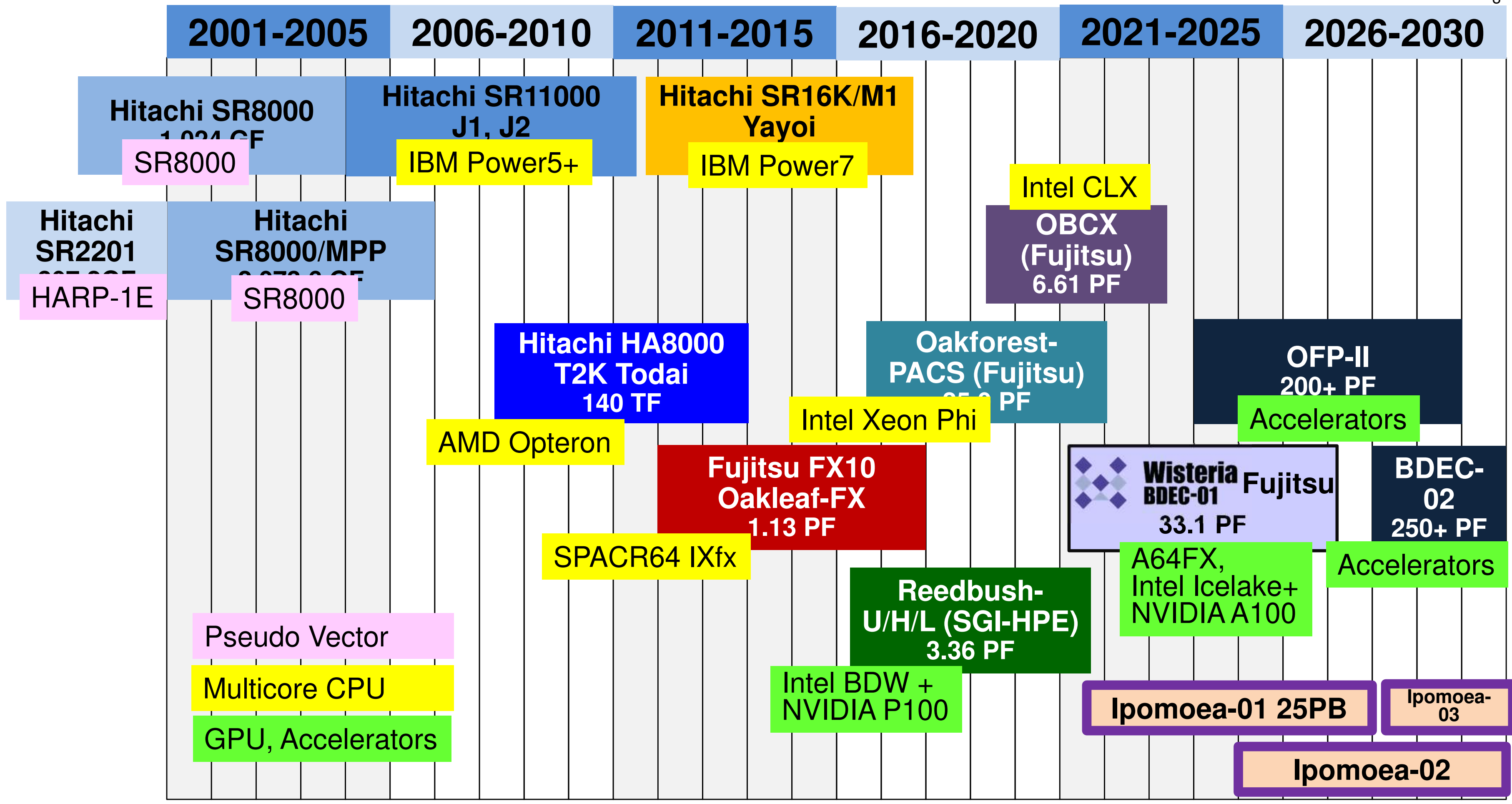

**ADAC12 Workshop, February 10-11, 2023, Kobe, Japan**
**Accelerated Data Analytics and Computing Institute**

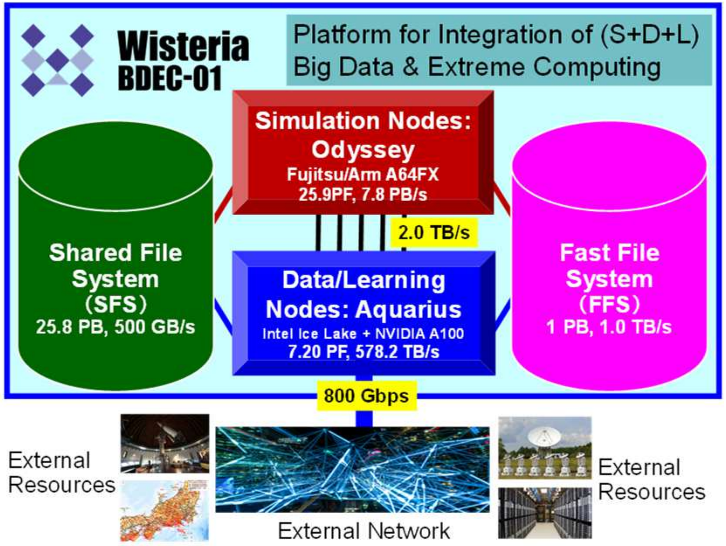



# Supercomputers @ITC/U.Tokyo

## 2,600+ Users

## 55+% outside of U.Tokyo


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
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)
Intel Xeon Phi
OFP-II
200+ 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
Pseudo Vector
Multicore CPU
GPU, Accelerators
Ipomoea-01 25PB
Ipomoea-
03
Ipomoea-02





Simulation Nodes (Odyssey)


Data/Learning Nodes (Aquarius)

東京大学 THE UNIVERSITY OF TOKYO

東京大学情報基盤センター INFORMATION TECHNOLOGY CENTER, THE UNIVERSITY OF TOKYO

## Oakbridge-CX (OBCX) (Fujitsu)

Oakbridge-CX

- Intel Xeon Cascade Lake
- July 2019-September 2023
- 6.61 PF, #129 in 69th TOP500

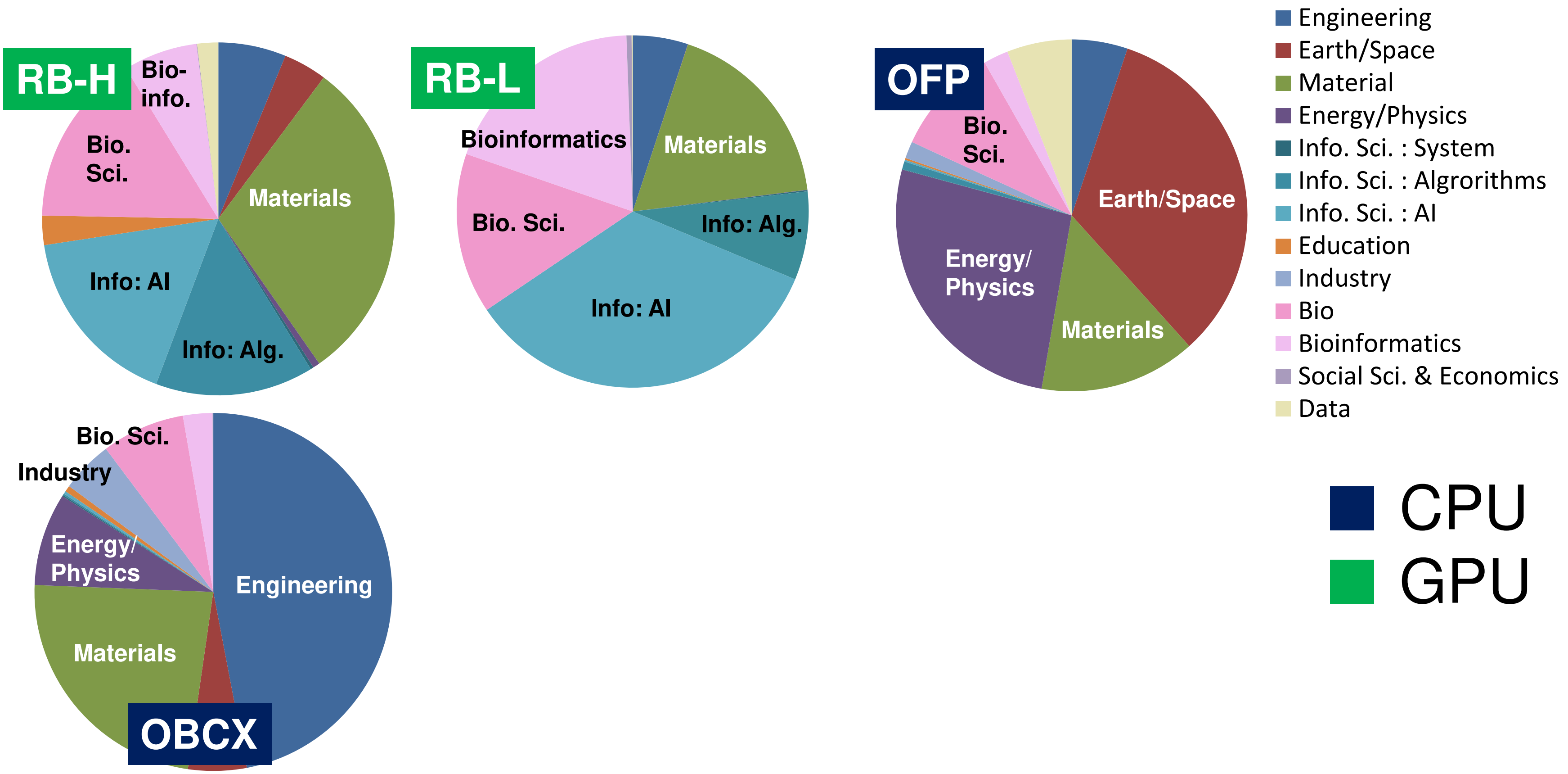
## Wisteria/BDEC-01（Fujitsu）

- **Simulation Nodes (Oddysey): A64FX (#23)**
- **Data/Learning Nodes (Aquarius) : Icelake + A100 (#125)**
- 33.1 PF, May 2021-April 2027
- Platform for Integration of "Simulation+Data+Learning (S+D+L)"
- Innovative Software Platform "h3-Open-BDEC" supported by Japanese Government (JSPS Grant-in-Aid for Scientific Res. (S) FY.2019-2023)

Wisteria BDEC-01

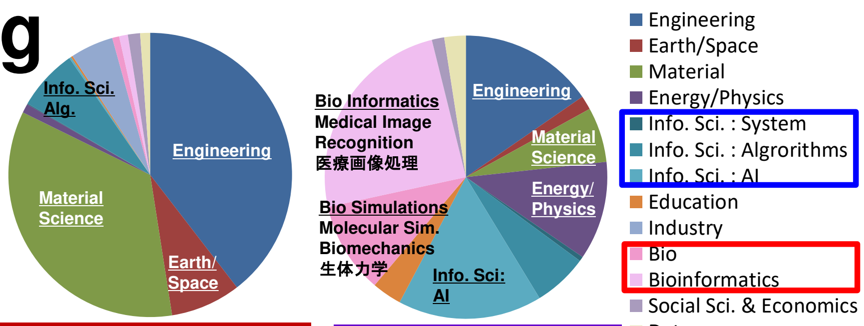
Hierarchical, Hybrid, Heterogeneous h3-Open-BDEC Big Data & Extreme Computing

# Research Area based on CPU Hours (FY.2020)



# Future of Supercomputing

- Various Types of Workloads
  - Computational Science & Engineering: Simulations
  - Big Data Analytics
  - AI, Machine Learning …

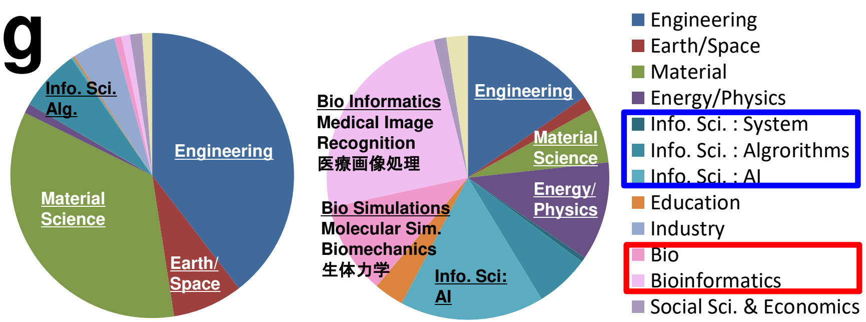



**Multicore Cluster**
**Intel BDW Only**
**(Reedbush-U)**

**GPU Cluster**
**Intel BDW + NVIDIA P100**
**(Reedbush-H)**

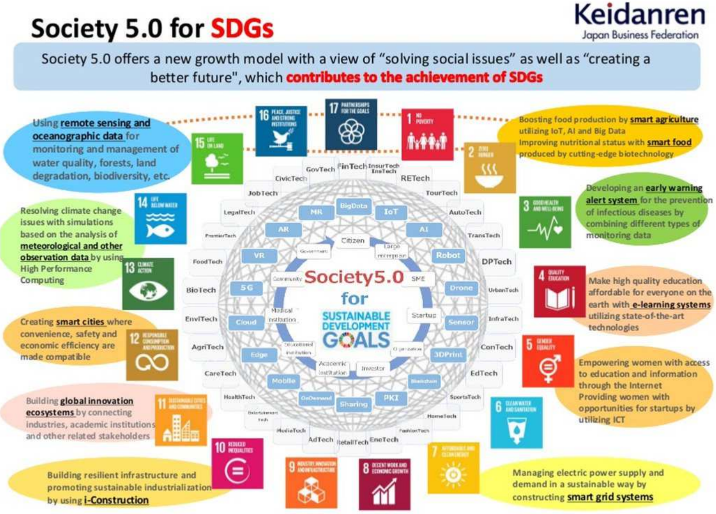
# Future of Supercomputing

- Various Types of Workloads
  - Computational Science & Engineering: Simulations
  - Big Data Analytics
  - AI, Machine Learning …
- Integration/Convergence of (Simulation + Data + Learning) (S+D+L) is important towards Society 5.0 proposed by Japanese Government
  - Super Smart & Human-centered Society by Digital Innovation (IoT, Big Data, AI etc.) and by <u>Integration of Cyber Space & Physical Space</u>



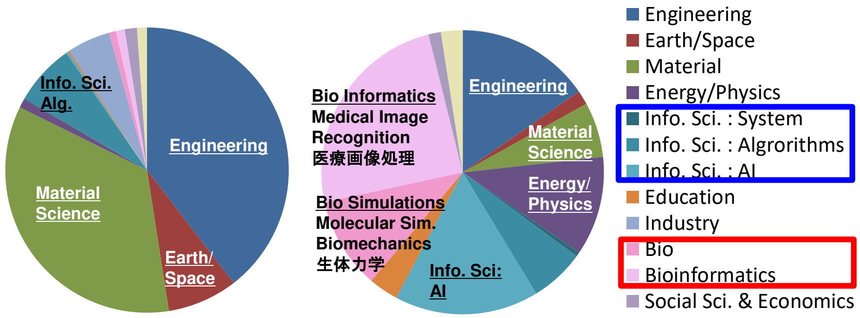

Multicore Cluster
Intel BDW Only
(Reedbush-U)

GPU Cluster
Intel BDW + NVIDIA P100
(Reedbush-H)



# Future of Supercomputing

- Various Types of Workloads
  - Computational Science & Engineering: Simulations
  - Big Data Analytics
  - AI, Machine Learning …
- **Integration/Convergence of (Simulation + Data + Learning) (S+D+L) is important towards Society 5.0**




**Multicore Cluster Intel BDW Only (Reedbush-U)**

**GPU Cluster Intel BDW + NVIDIA P100 (Reedbush-H)**

- **BDEC (Big Data & Extreme Computing)**
  - **Platform for Integration of (S+D+L)**
  - **Focusing on S (Simulation)**
    - **AI for HPC, AI for Science, Digital Twins**
  - **Planning started in 2015**

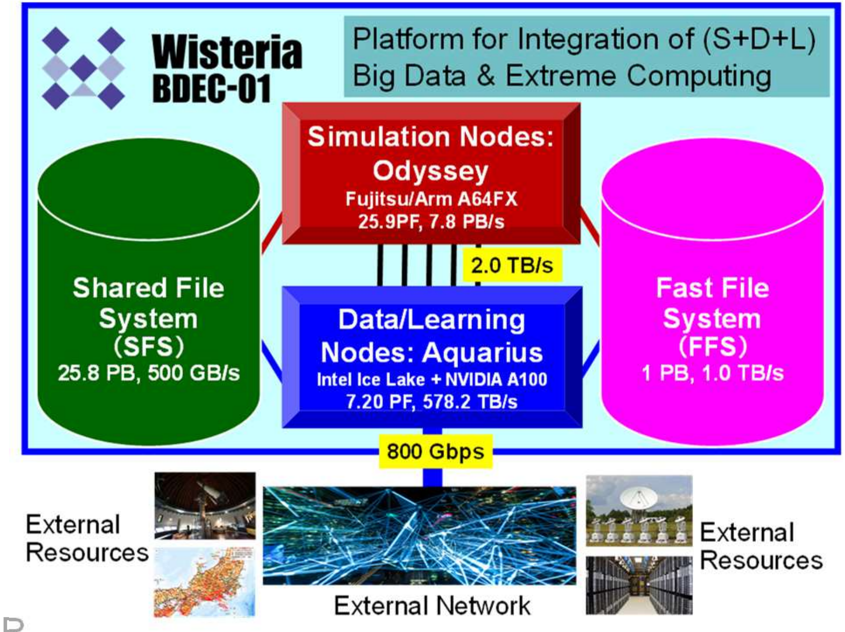
**BDEC (Big Data & Extreme Computing)**

**S + D + L**

# Wisteria/BDEC-01

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

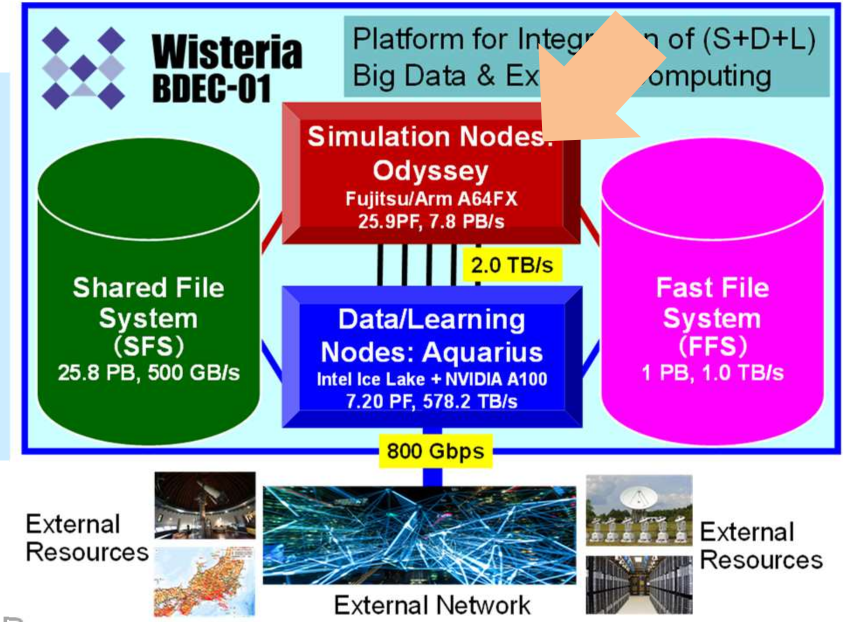
- Operation starts on May 14, 2021
- 33.1 PF, 8.38 PB/sec by **Fujitsu**
  - ~4.5 MVA with Cooling, ~360m²
- 2 Types of Node Groups
  - Hierarchical, Hybrid, Heterogeneous (h3)
  - Simulation Nodes: Odyssey
    - Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF
      - 7,680 nodes (368,640 cores), Tofu-D
      - General Purpose CPU + HBM
      - Commercial Version of "Fugaku"
  - Data/Learning Nodes: Aquarius
    - Data Analytics & AI/Machine Learning
    - Intel Xeon Ice Lake + NVIDIA A100, 7.2PF
      - 45 nodes (90x Ice Lake, 360x A100), IB-HDR
    - Some of the 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)**
**Platform for Integration of (S+D+L)**

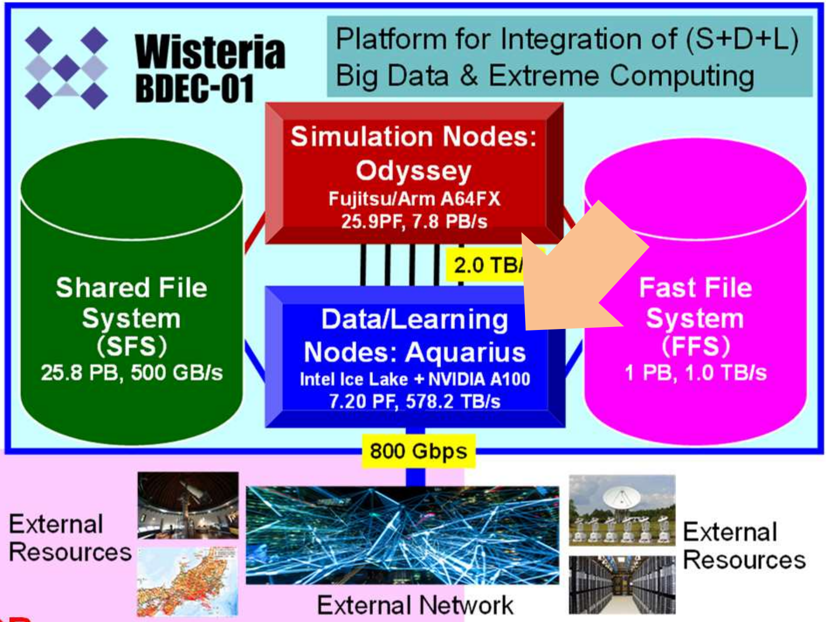
- Operation starts on May 14, 2021
- 33.1 PF, 8.38 PB/sec by **Fujitsu**
  – ~4.5 MVA with Cooling, ~360m²
- **2 Types of Node Groups**
  – Hierarchical, Hybrid, Heterogeneous (h3)
  – **Simulation Nodes: Odyssey**
    - **Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF**
      – **7,680 nodes (368,640 cores), Tofu-D**
      – **General Purpose CPU + HBM**
      – **Commercial Version of “Fugaku”**
  – Data/Learning Nodes: Aquarius
    - Data Analytics & AI/Machine Learning
    - Intel Xeon Ice Lake + NVIDIA A100, 7.2PF
      – 45 nodes (90x Ice Lake, 360x A100), IB-HDR
    - Some of the 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)**
**Platform for Integration of (S+D+L)**

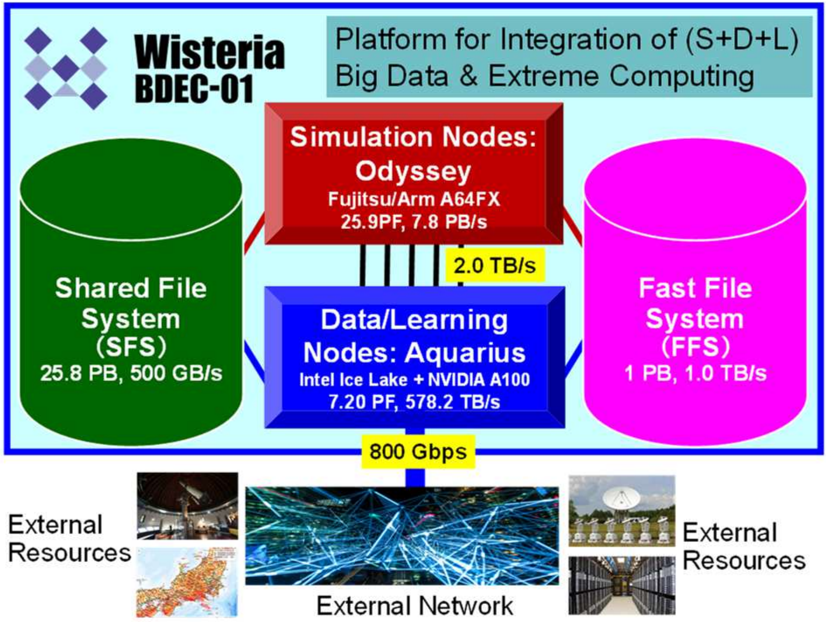
- Operation starts on May 14, 2021
- 33.1 PF, 8.38 PB/sec by **Fujitsu**
  - ~4.5 MVA with Cooling, ~360m²
- **2 Types of Node Groups**
  - Hierarchical, Hybrid, Heterogeneous (h3)
  - **Simulation Nodes: Odyssey**
    - **Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF**
      - **7,680 nodes (368,640 cores), Tofu-D**
      - **General Purpose CPU + HBM**
      - **Commercial Version of "Fugaku"**
  - **Data/Learning Nodes: Aquarius**
    - **Data Analytics & AI/Machine Learning**
    - **Intel Xeon Ice Lake + NVIDIA A100, 7.2PF**
      - **45 nodes (90x Ice Lake, 360x A100), IB-HDR**
    - **Some of the DL nodes are connected to external resources directly**
- File Systems: SFS (Shared/Large) + FFS (Fast/Small)

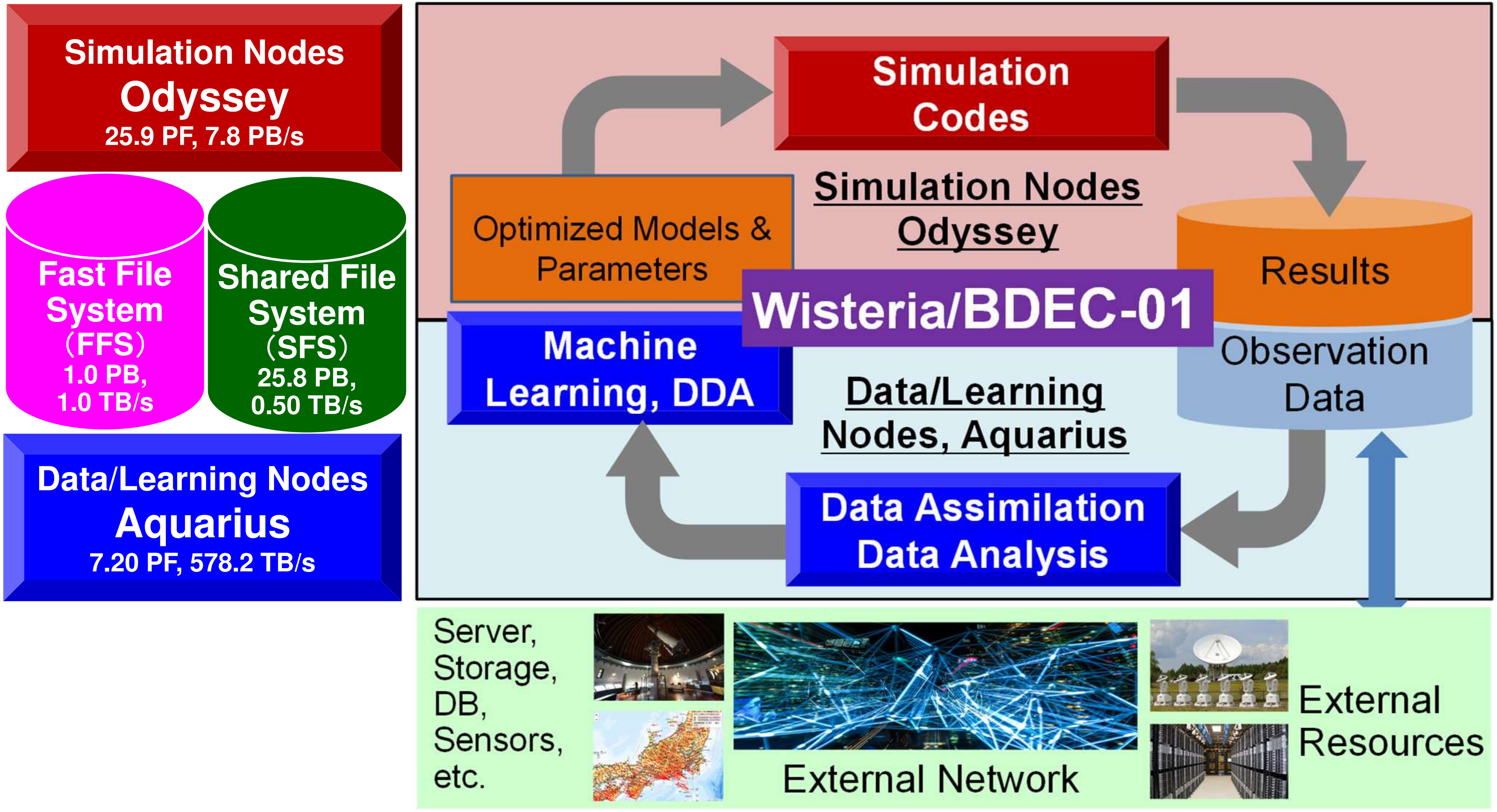


# Rankings@SC22 November 2022

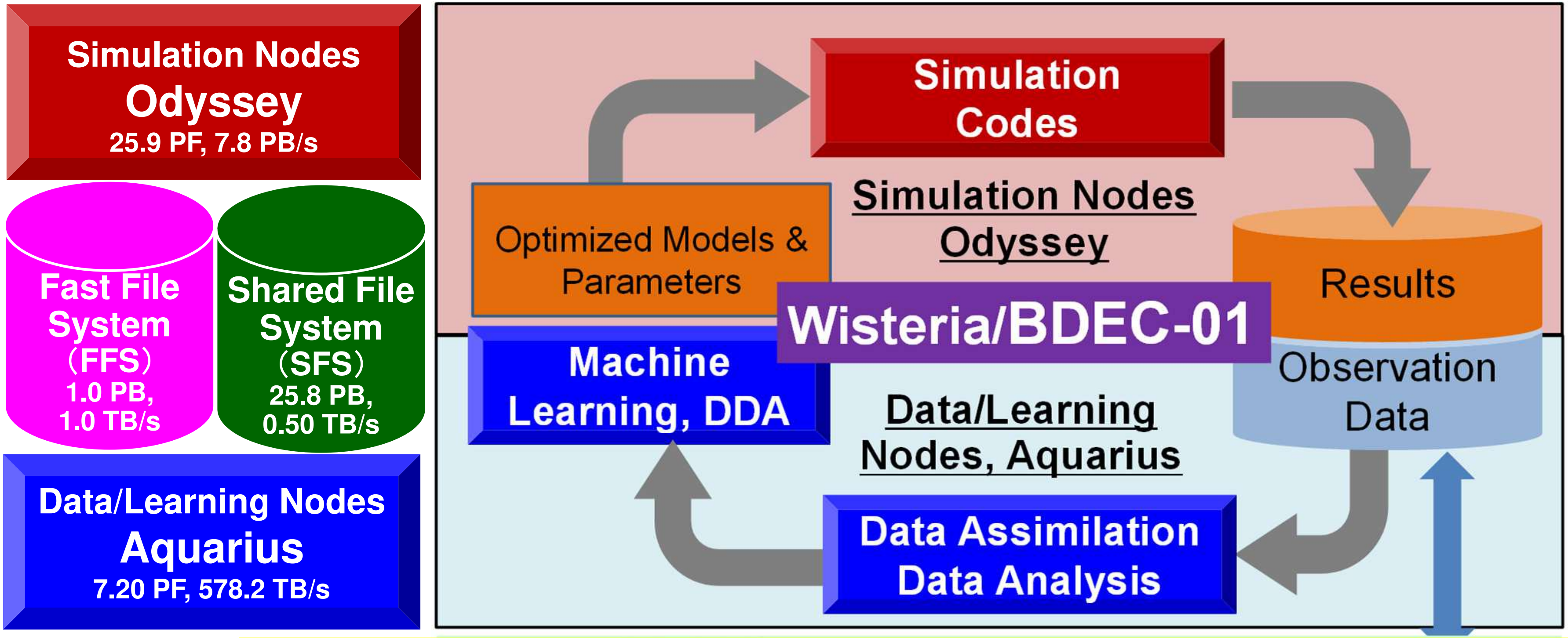
| | Odyssey | Aquarius |
|---|---|---|
| TOP 500 | 23 | 125 |
| Green 500 | 45 | 28 |
| HPCG | 12 | 68 |
| Graph 500 BFS | 4 | - |
| HPL-MxP (HPL-AI) | 10* | - |

*) ISC 2022 (June 2022)





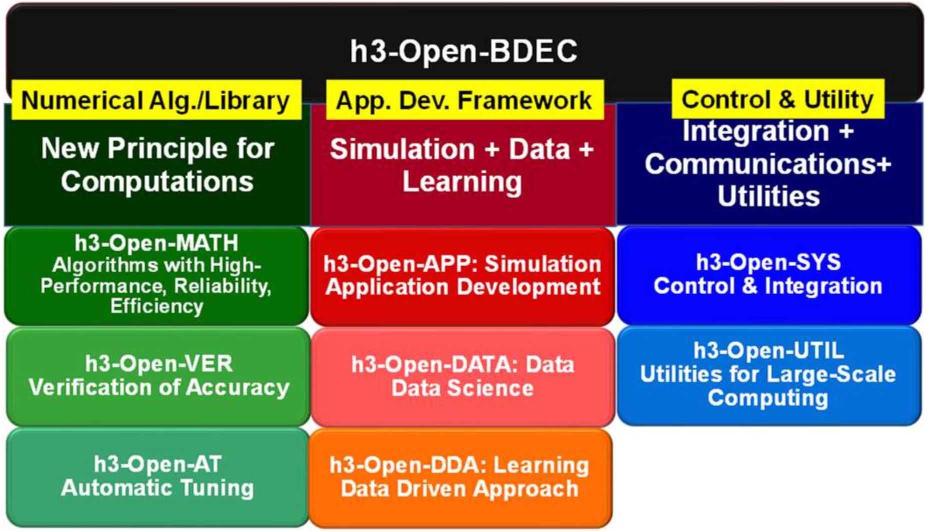
Simulation Nodes
Odyssey
25.9 PF, 7.8 PB/s
Fast File System (FFS)
1.0 PB, 1.0 TB/s
Shared File System (SFS)
25.8 PB, 0.50 TB/s
Data/Learning Nodes
Aquarius
7.20 PF, 578.2 TB/s
Simulation Codes
Optimized Models & Parameters
Simulation Nodes Odyssey
Results
Wisteria/BDEC-01
Machine Learning, DDA
Observation 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)**

# 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
- Leading-PI: Kengo Nakajima (The University of Tokyo)
- Total Budget: 1.41M USD



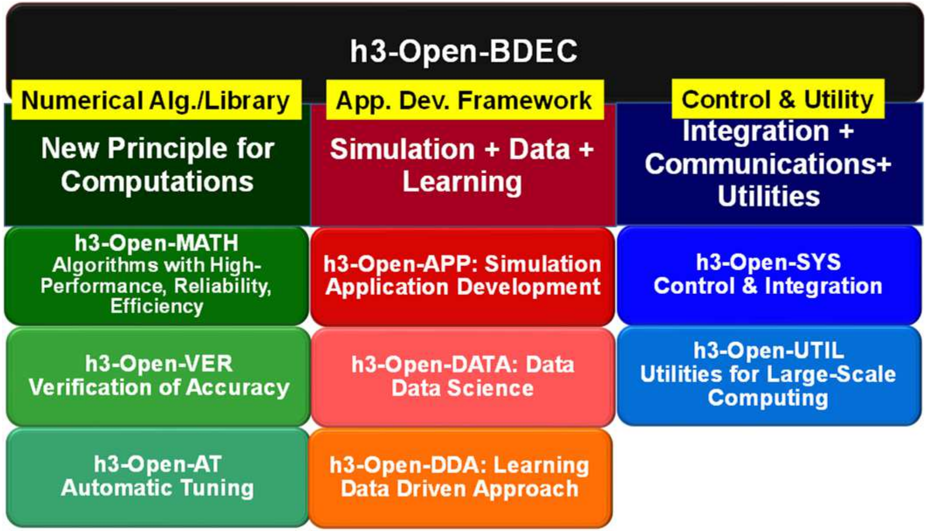
# Members (Co-PI's) of h3-Open-BDEC Project

## 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 (TWCU), Co-PI, Accuracy Verification
- Takahiro Katagiri (Nagoya U), Co-PI, Appropriate Computing
- Hiroya Matsuba (ITC/U.Tokyo, Hitachi), Co-PI, Container

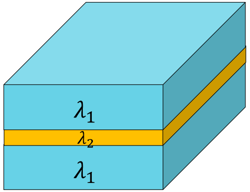
# h3-Open-BDEC Innovative Software Platform for Integration of (S+D+L) on the BDEC System, such as Wisteria/BDEC-01

- "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

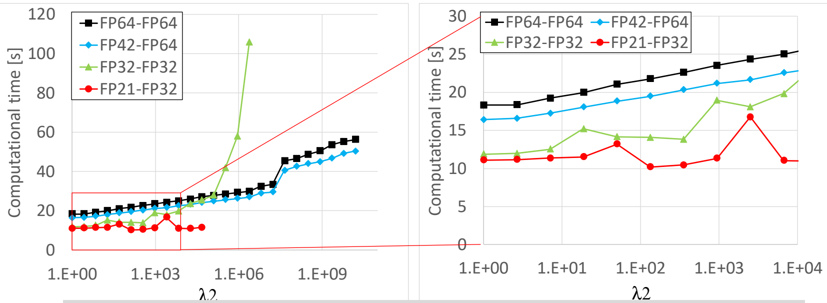


# Adaptive Precision Computing with FP21/FP42

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


Heat Conduction with
Heterogeneous Material Property

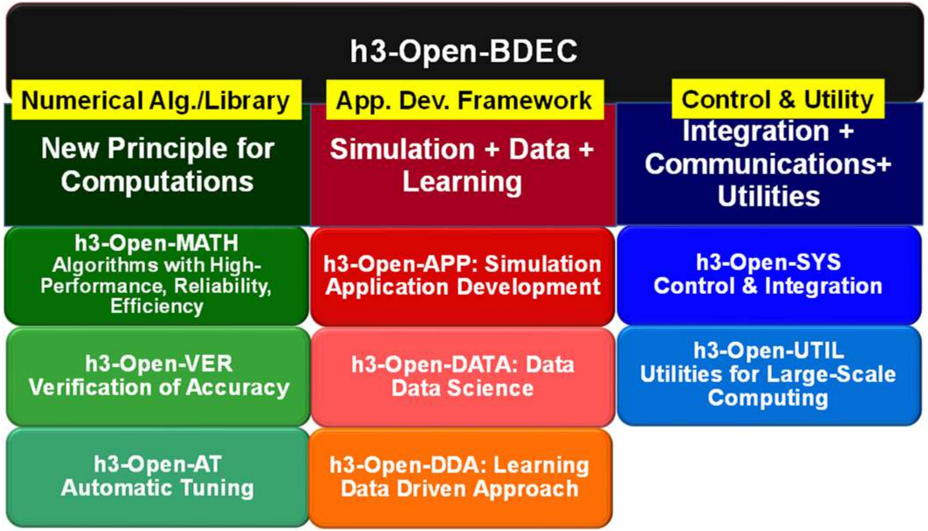



Computation Time for ICCG Solver
Various Types of Precisions on Intel Xeon Cascadelake

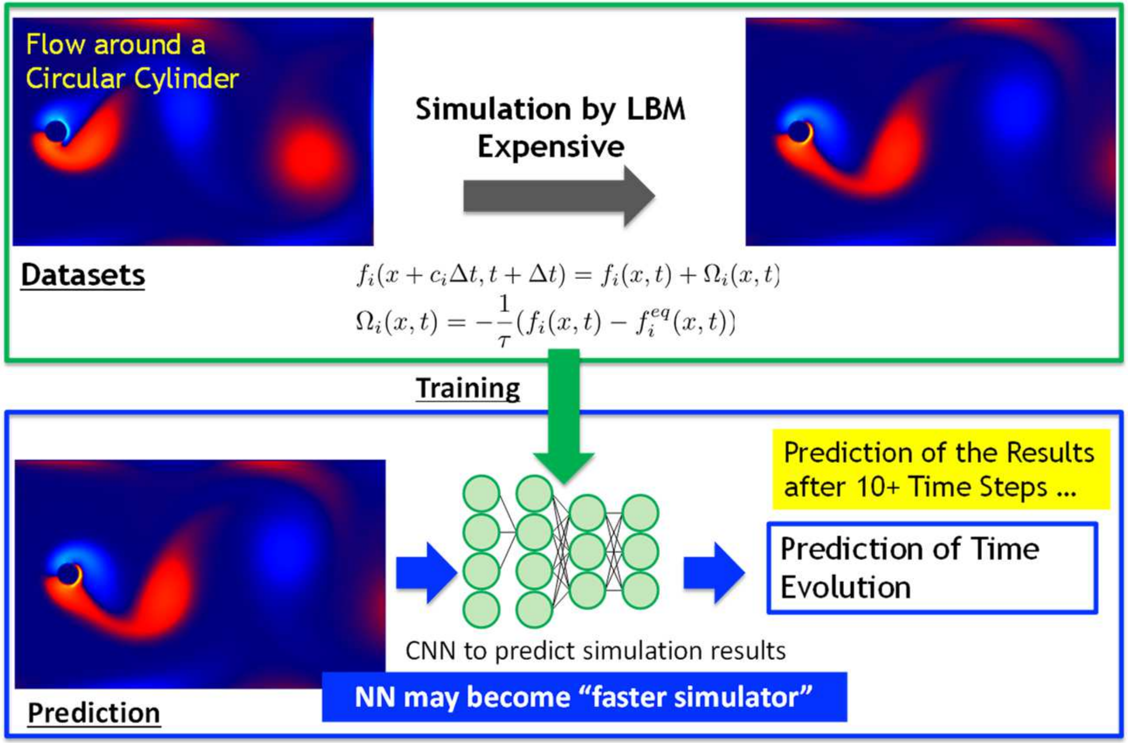
In recent years, the usefulness of low-precision floating-point representation has been studied in various fields such as machine learning. Low accuracy can be expected to have effects such as shortening calculation time and reducing power consumption. For example, in an application with a memory bandwidth bottleneck, the effect of reducing the calculation time by reducing the amount of memory transfer is significant. However, in fields such as iterative methods, it is common to use FP64 because the calculation accuracy strongly affects the convergence, and there are few application examples of low-precision arithmetic. This study investigates the applicability of low-precision representation to the Krylov subspace and stationary iterative methods. In this research, we focus on the FP32, FP16, and FP42, FP21, which are not standardized by IEEE754. Developed method has been evaluated for ICCG solver, which solves linear equations derived from 3D FVM code for steady-state head conduction with heterogeneous material property ($\lambda_1=10^0$, $\lambda_2=10^0\sim10^9$). Generally, computation with lower precision (e.g. FP32-FP32, FP21-FP32) becomes unstable, if condition number of the coefficient matrix is larger ($\lambda_2$ is larger), FP21-FP32 provides the best performance if $\lambda_2$ is up to $10^4$. (“FP21-FP32” means “matrices are in FP21, and vectors are in FP32)

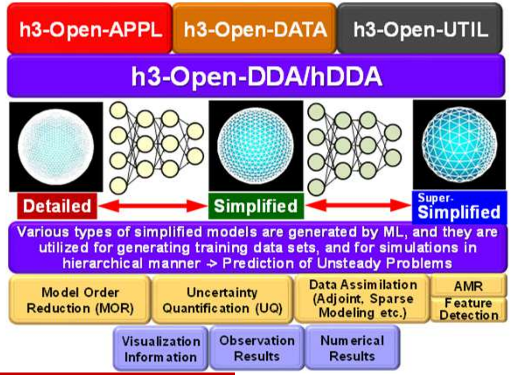
# h3-Open-BDEC Innovative Software Platform for Integration of (S+D+L) on the BDEC System, such as Wisteria/BDEC-01

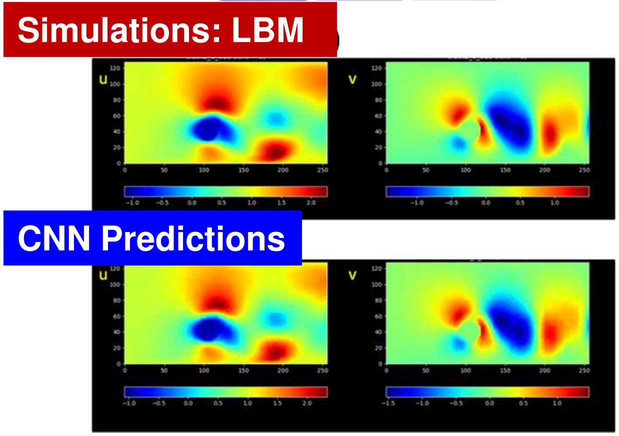
- "Three" Innovations
  - New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
  - Integration of (S+D+L) by Hierarchical Data Driven Approach (*h*DDA)
  - Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01



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



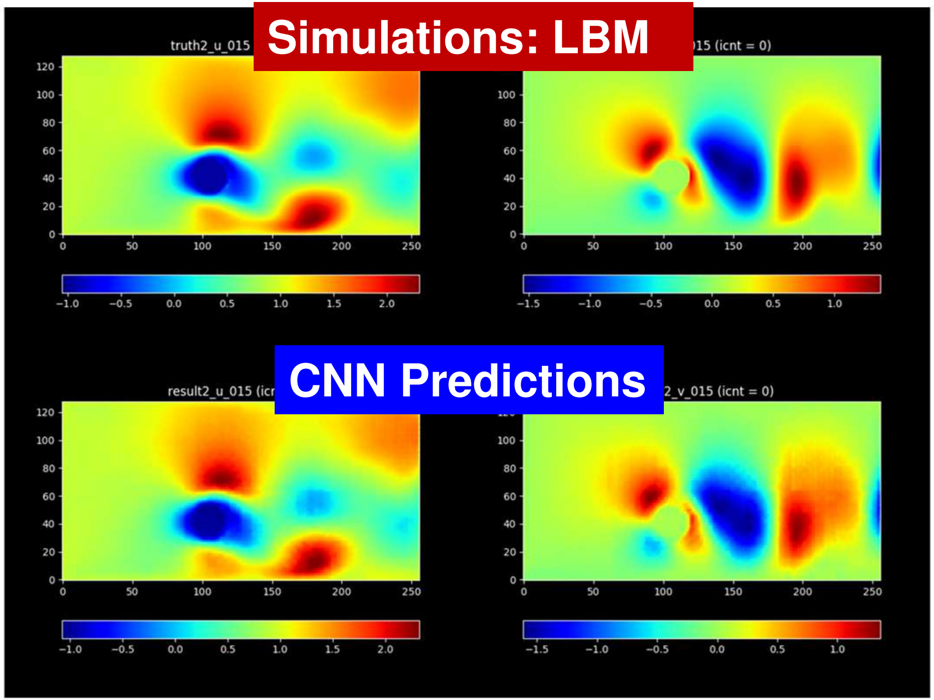




[c/o Takashi Shimokawabe (ITC/U.Tokyo)]

# Prediction of CFD Simulation by ML/CNN

Takashi Shimokawabe (shimokawabe@cc.u-tokyo.ac.jp)

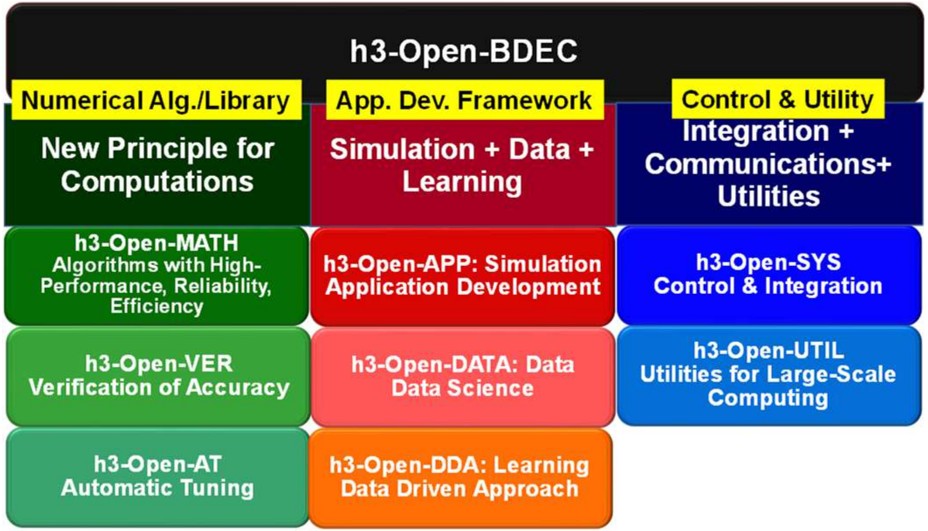



Comparison of the flow velocity results obtained by the conventional simulation (upper) and the prediction of these results by deep learning (lower)

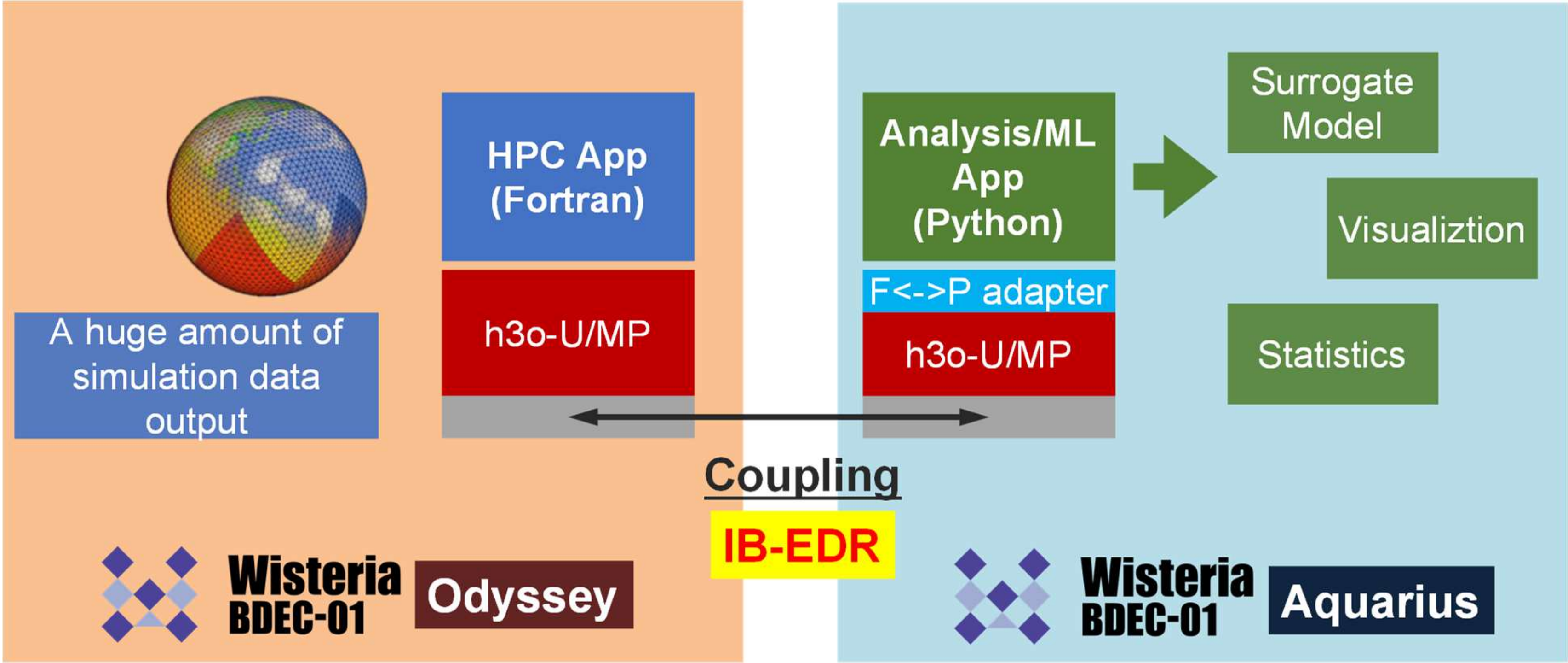
Computational fluid dynamics (CFD) is widely used in science and engineering. However, since CFD simulations requires a large number of grid points and particles for these calculations, these kinds of simulations demand a large amount of computational resources such as supercomputers. Recently, deep learning has attracted attention as a surrogate method for obtaining calculation results by CFD simulation approximately at high speed. We are working on a project to develop a parallelization method to make it possible to apply the surrogate method based on the deep learning to large scale geometry. Unlike the model parallel computing, the method we are currently developing predicts large-scale steady flow simulation results by dividing the input geometry into multiple parts and applying a single small neural network to each part in parallel. This method is developed based on considering the characteristics of CFD simulation and the consistency of the boundary condition of each divided subdomain. By using the physical values on the adjacent subdomains as boundary conditions, applying deep learning to each subdomain can predict simulation results consistently in the entire computational domain. It is possible to predict the simulation results in about 36.9 seconds by the developed method, compared to about 286.4 seconds by the conventional numerical method. In addition to this, we are also attempting to develop a method for fast prediction of time evolution calculations using deep learning.

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

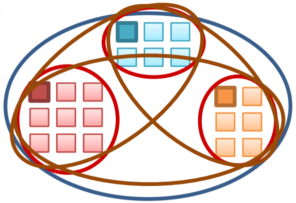
- “Three” Innovations
  - New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
  - Integration of (S+D+L) by Hierarchical Data Driven Approach (*h*DDA)
  - Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01

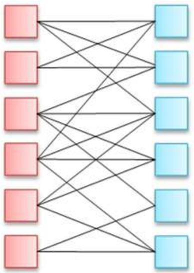


# h3-Open-UTIL/MP (h3o-U/MP) + h3-Open-SYS/WaitIO-Socket

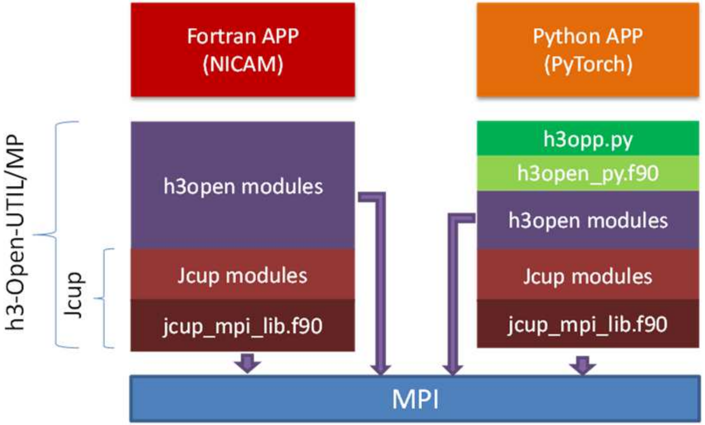


# h3-Open-UTIL/MP + h3-Open-SYS/WaitIO-Socket

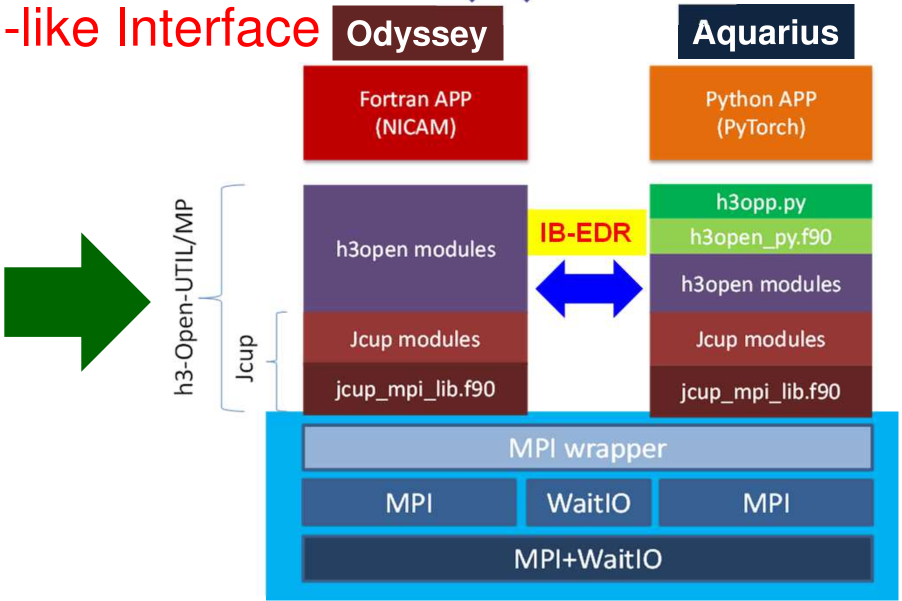




- Single MPI Job (May 2021)
- Direct Communication between Odyssey-Aquarius through IB-EDR by h3-Open-SYS/WaitIO, which provides MPI-like Interface

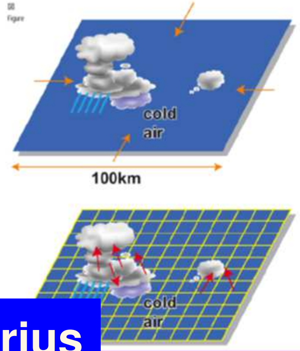



May 2021 (Initial): Single MPI Job



# 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!



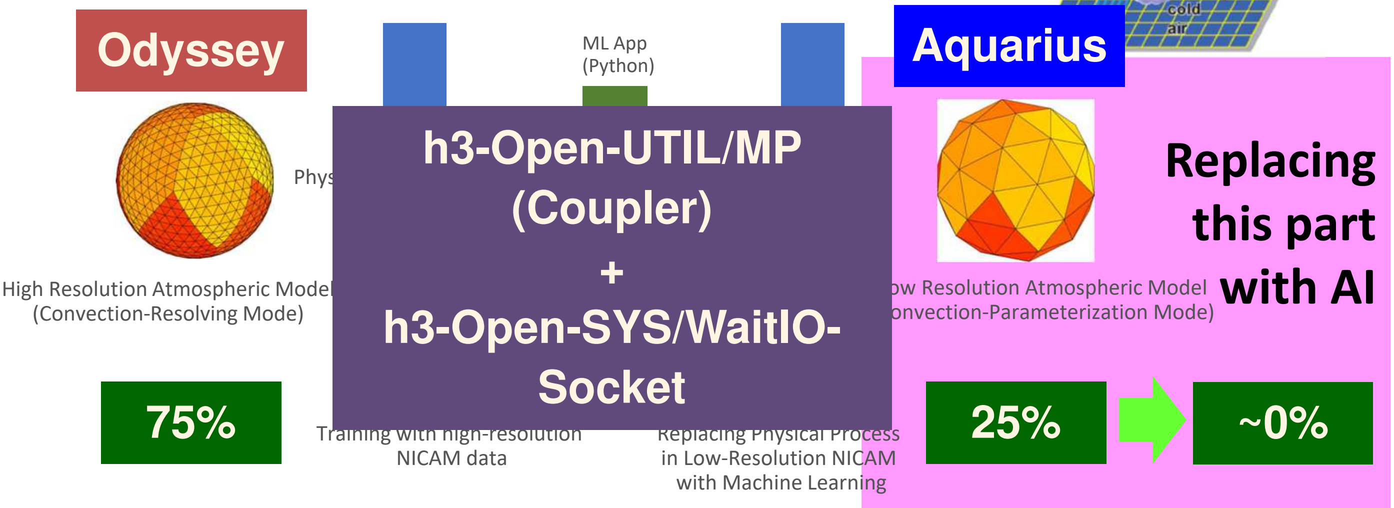
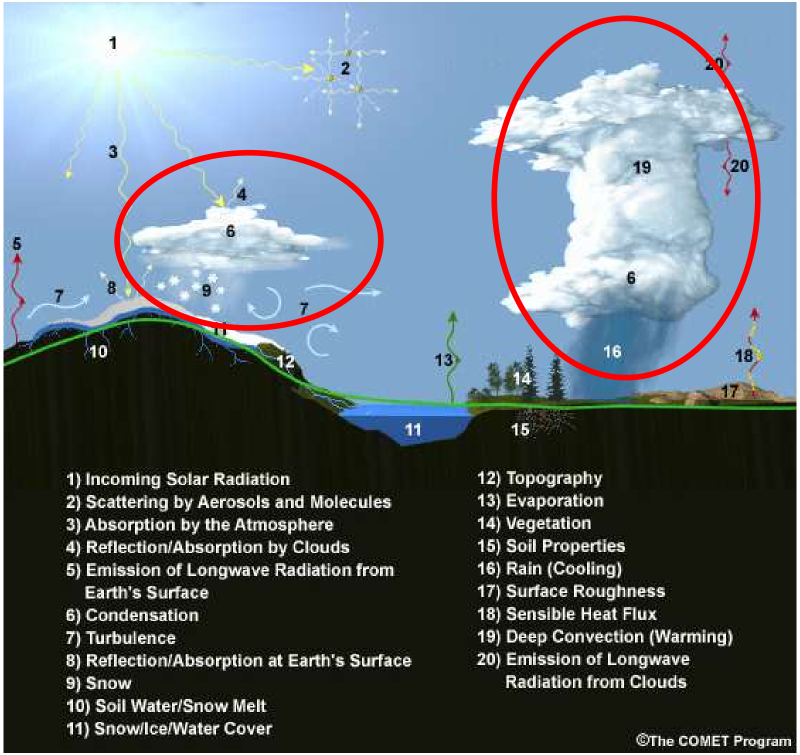




Diagram of applying ML to an atmospheric model

# Atmosphere-ML Coupling

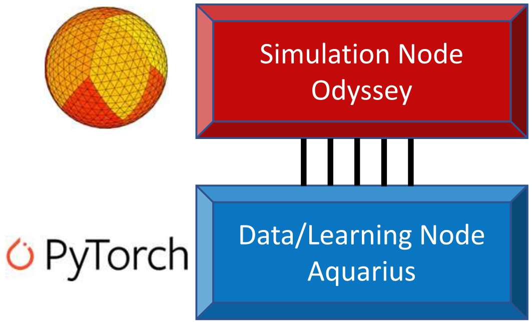


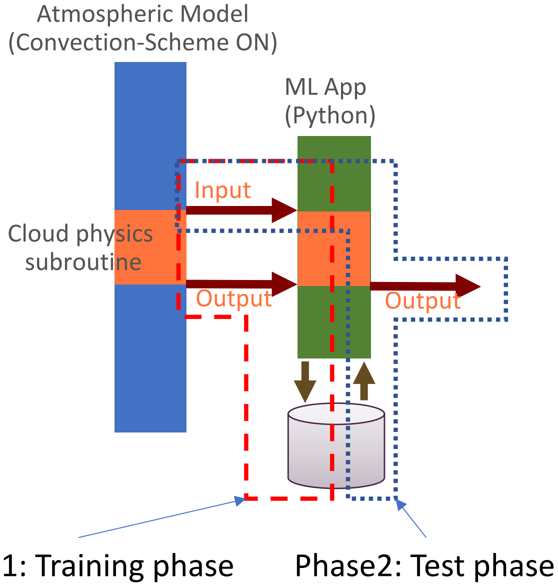

- Model component emulation (surrogation)
  - The emulation target in this study is cloud microphysical processes (phase changes, collision, coagulation, and precipitation)
  - Atmospheric pressure, temperature, and vertical distribution of water will change between before and after computing the cloud microphysical processes
  - The data-driven cloud model predicts atmospheric state changes per unit of time

# Experimental Design

- Atmospheric model on Odyssey
  - NICAM : global non-hydrostatic model with an icosahedral grid
  - Resolution : horizontal : 10240, vertical : 78
- ML on Aquarius
  - Framework : PyTorch
  - Method : Three-Layer MLP
  - Resolution : horizontal : 10240, vertical : 78
- Experimental design
  - Phase1: PyTorch is trained to reproduce output variables from input variables of cloud physics subroutine.
  - Phase2:Reproduce the output variables from Input variables and training results
- 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