h3-Open-BDEC: Innovative Software Platform for Scientific Computing in the Exascale Era by Integrations of (Simulation + Data + Learning)

Kengo Nakajima
Information Technology Center
The University of Tokyo

Videos: https://www.cc.u-tokyo.ac.jp/public/sc20.php

Overview

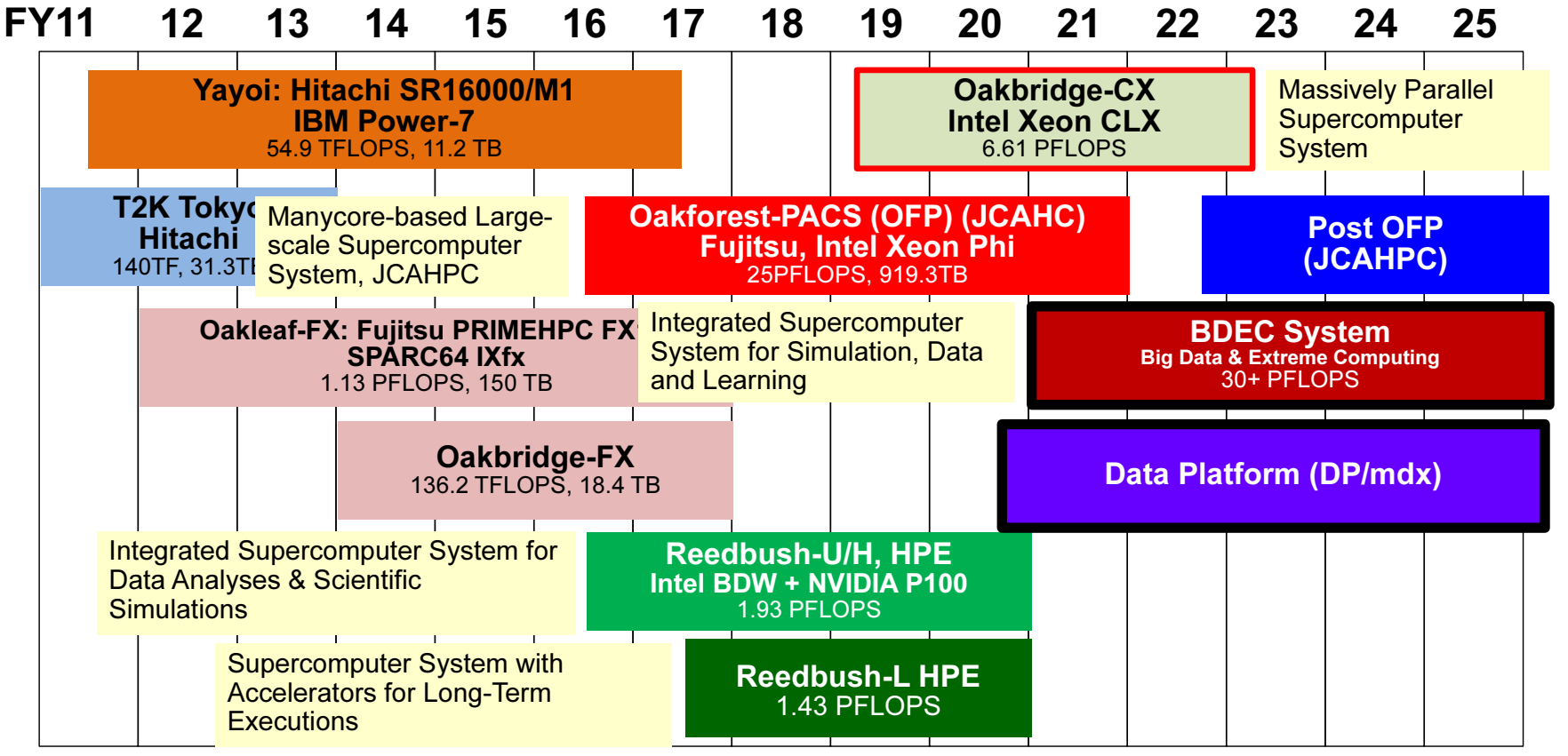
- **In this study, we propose an innovative method of computational science towards the Exascale Era/Society 5.0 by integration of (Simulation + Data + Learning (S+D+L)), where ideas of data science and machine learning are introduced to computational science**

Overview

- In this study, we propose an innovative method of computational science towards the Exascale Era/Society 5.0 by integration of (Simulation + Data + Learning (S+D+L)), where ideas of data science and machine learning are introduced to computational science
- **We are operating 3 supercomputer systems, and now introducing the BDEC (Big Data & Extreme Computing) System with 32+PF as the Platform for Integration of (S+D+L)**

Supercomputers in ITC/U.Tokyo

Information Technology Center, The University of Tokyo

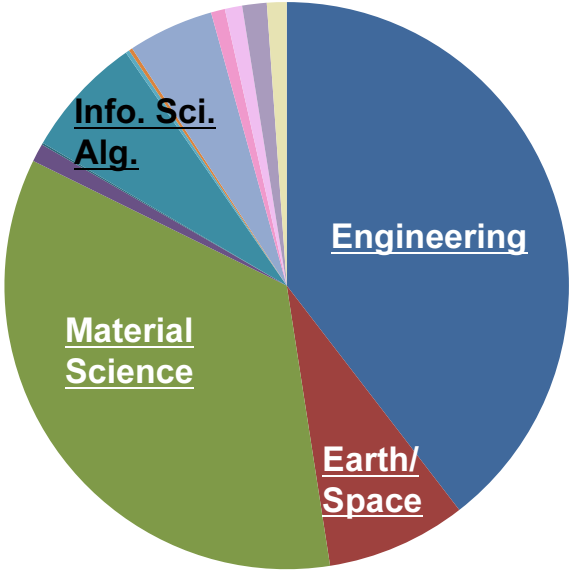


Now operating 3 Systems !!

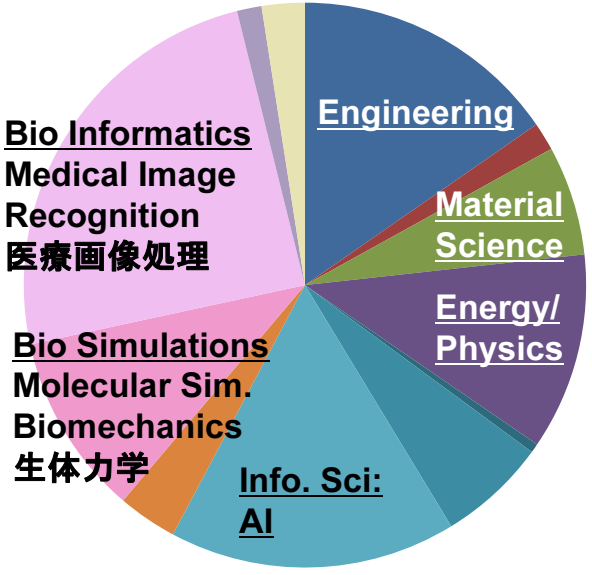
2,600+ users (55+% from outside of U.Tokyo)

- Reedbush (HPE, Intel BDW + NVIDIA P100 (Pascal))
  - Integrated Supercomputer Sys. for Data Analyses & Scientific Simulations
    - Jul.2016-Mar.2021
  - Our first GPU System, DDN IME (Burst Buffer)
  - Reedbush-U: CPU only, 420 nodes, 508 TF (Jul.2016, retired on June 30)
  - Reedbush-H: 120 nodes, 2 GPUs/node: 1.42 PF (Mar.2017)
  - Reedbush-L: 64 nodes, 4 GPUs/node: 1.43 PF (Oct.2017)
- Oakforest-PACS (OFP) (Fujitsu, Intel Xeon Phi (KNL))
  - JCAHPC (U.Tsukuba & U.Tokyo)
  - 25 PF, #18 in 55th TOP 500 (June 2020) (#3 in Japan)
  - Omni-Path Architecture, DDN IME (Burst Buffer)
- Oakbridge-CX (OBCX) (Fujitsu, Intel Xeon Platinum 8280, CLX)
  - Massively Parallel Supercomputer System
  - 6.61 PF, #60 in 55th TOP 500, July 2019-June 2023
  - SSD's are installed to 128 nodes (out of 1,368)

Research Area based on CPU Hours (FY.2019)



Multicore Cluster
Intel BDW Only
(Reedbush-U)



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

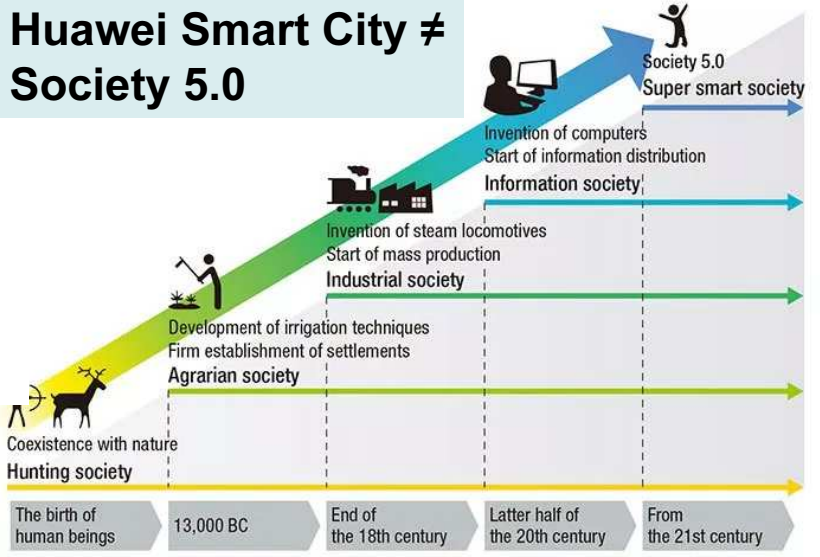
- Engineering
- Earth/Space
- Material
- Energy/Physics
- Info. Sci. : System
- Info. Sci. : Algrorithms
- Info. Sci. : AI
- Education
- Industry
- Bio
- Bioinformatics
- Social Sci. & Economics
- Data

Society 5.0: the Cabinet Office of Japan

- Super Smart & Human-centered Society by Digital Innovation (IoT, Big Data, AI etc.) and by **Integration of Cyber Space & Physical Space (where HPC plays a big role)**

Economic and social innovation by deepening of Society 5.0

Huawei Smart City ≠ Society 5.0



Source: Prepared based on materials from the Japan Business Federation (Keidanren)

5.0: Super Smart

4.0: Information

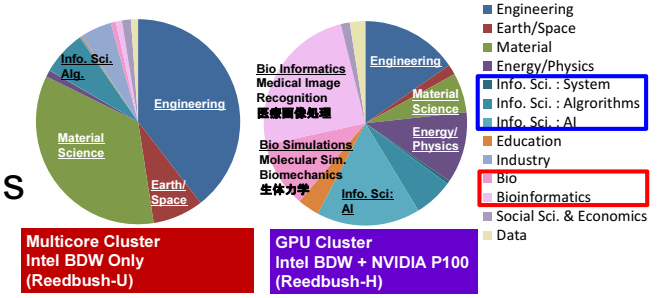
3.0: Industry

2.0: Agrarian

1.0: Hunting

Future of Supercomputing

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



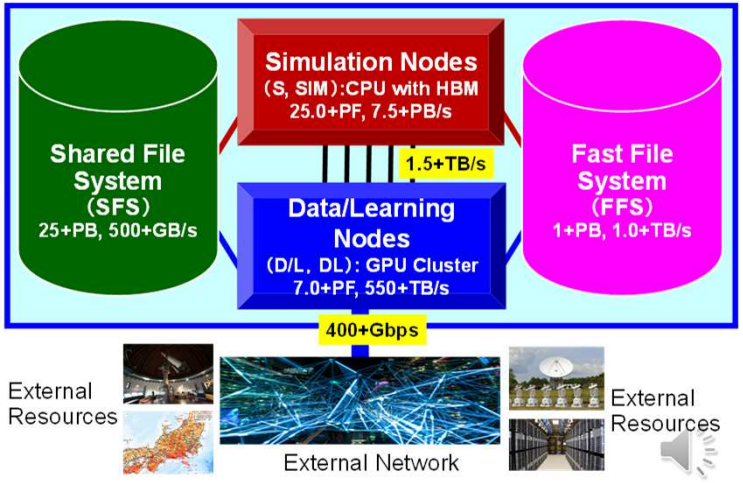
- **Integration/Convergence of (Simulation + Data + Learning) (S+D+L) is important towards Society 5.0: AI for HPC, Sophiscated Simulation**
- **Two Platforms will be introduced in Kashiwa II Campus of the University of Tokyo (Spring 2021)**
  - **BDEC (Big Data & Extreme Computing): Batch**
  - **Data Platform (DP/mdx): Cloud-like, More Flexible/Interactive**

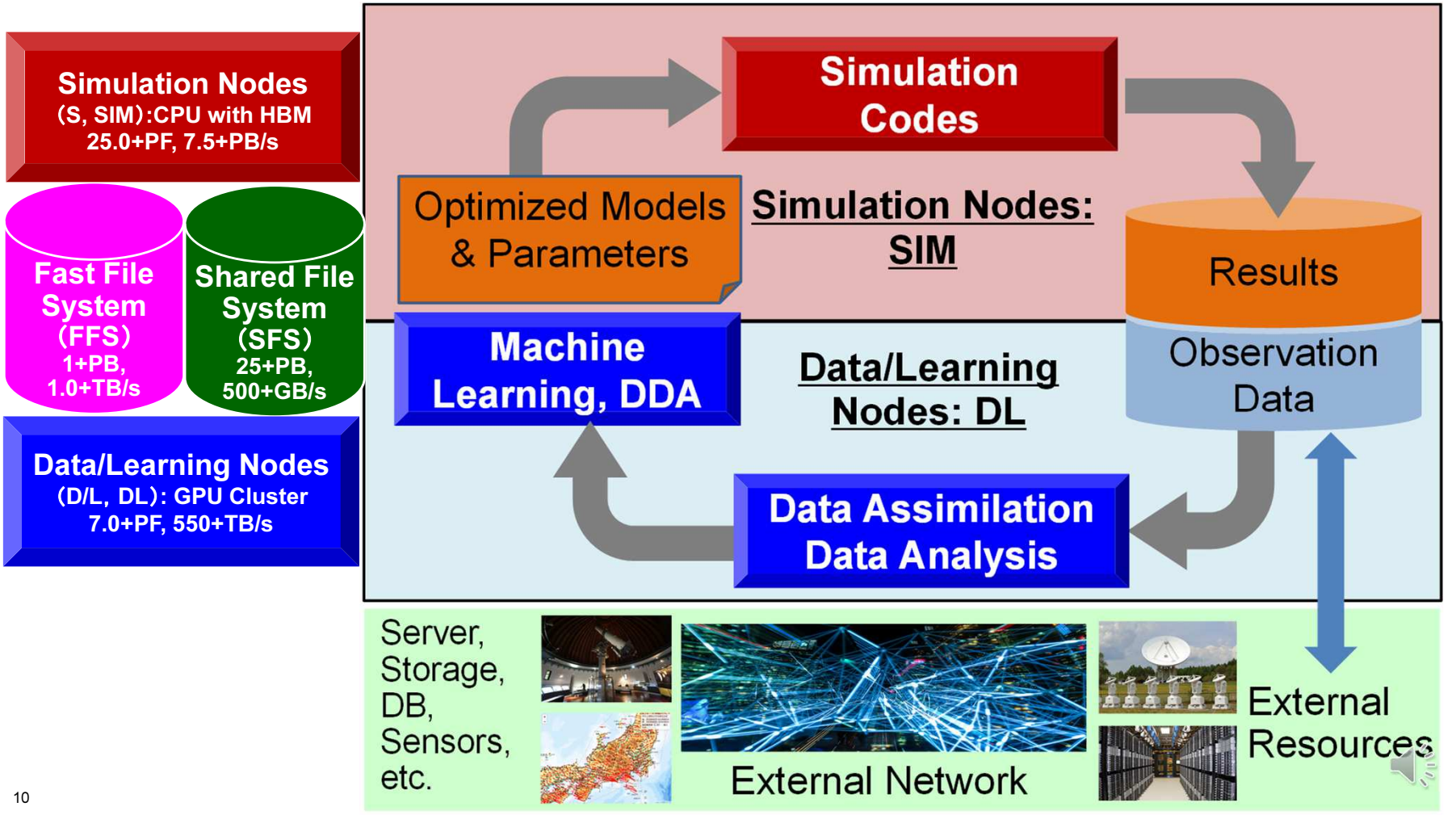
BDEC: S + D + L

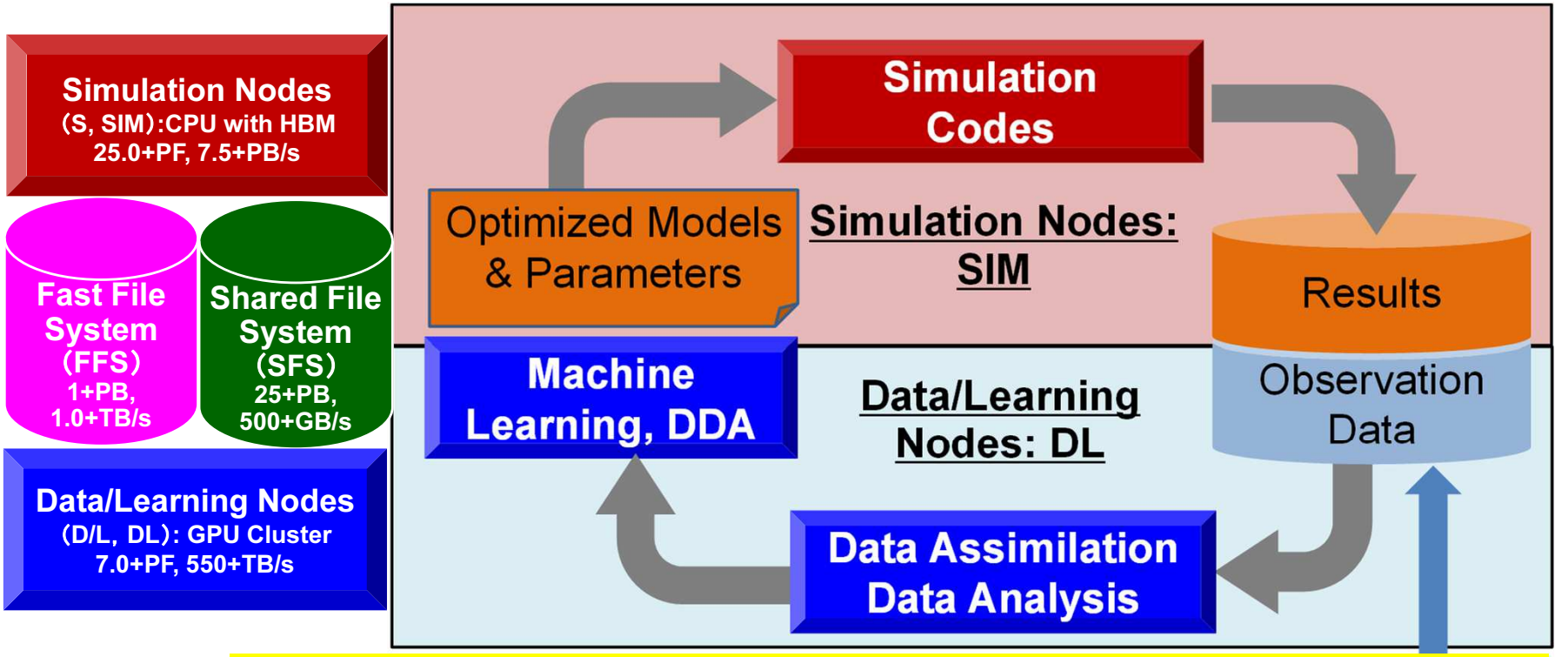
mdx: s + D + L

BDEC System (Big Data & Extreme Computing) Platform for (S+D+L)

- The BDEC System (Big Data & Extreme Computing) is a platform for the **integration of (S+D+L) (32+PF , 8.05+PB/sec**)
  - Operation starts in Spring 2021 at the Info. Technology Ctr., the Tokyo University
- **Hierarchical, Hybrid, Heterogeneous (h3)**
- **Simulation Nodes for CSE (SIM)**
  - **Manycore CPU with HBM, 25+PF**
- **Data/Learning Nodes for Data Analytics & AI/ML Workloads (DL)**
  - **GPU Cluster, 7+PF**
  - Some of the DL nodes are connected to external resources directly through an external network




Simulation Nodes
(S, SIM):CPU with HBM
25.0+PF, 7.5+PB/s
Fast File System (FFS)
1+PB, 1.0+TB/s
Shared File System (SFS)
25+PB, 500+GB/s
Data/Learning Nodes
(D/L, DL): GPU Cluster
7.0+PF, 550+TB/s
Simulation Codes
Optimized Models & Parameters
Simulation Nodes: SIM
Results
Machine Learning, DDA
Data/Learning Nodes: DL
Observation Data
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)

Overview

- In this study, we propose an innovative method of computational science towards the Exascale Era/Society 5.0 by integration of (Simulation + Data + Learning (S+D+L)), where ideas of data science and machine learning are introduced to computational science
- We are operating 3 supercomputer systems, and now introducing the BDEC (Big Data & Extreme Computing) System with 32+PF as the Platform for Integration of (S+D+L)
- **h3-Open-BDEC: Innovative Software Platform for Integration of (S+D+L) on the BDEC System**
  - **5-year project supported by Japanese Government through JSPS Grant-in-Aid for Scientific Research (S) since 2019**
  - **Leading-PI: Kengo Nakajima (The University of Tokyo)**
  - **Total Budget: 152.7M JPY= 1.41M USD**

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

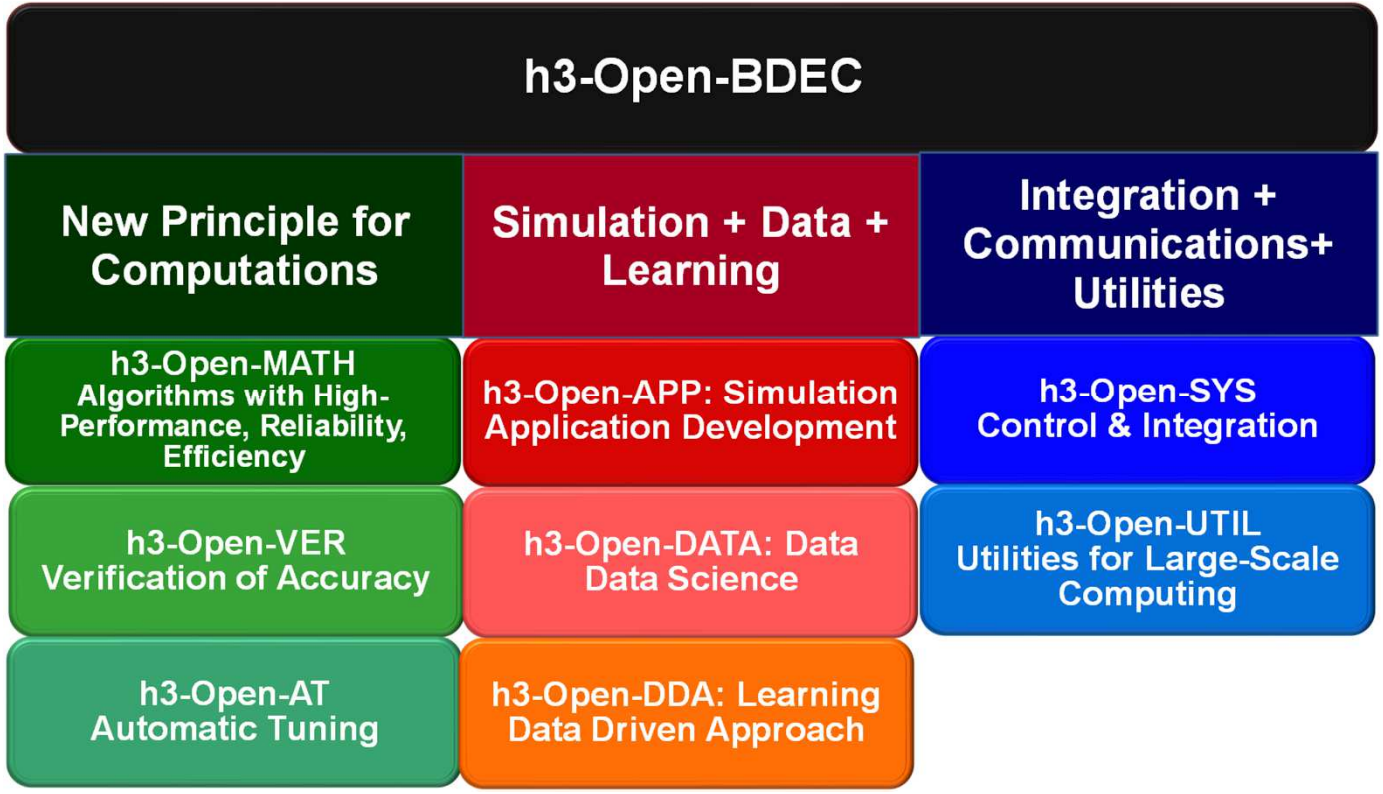
Computer Science, Computational Science, Numerical Algorithms, Data Science, Machine Learning

- 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), Co-PI, Container

東京大学 THE UNIVERSITY OF TOKYO
北海道大学 HOKKAIDO UNIVERSITY
国立研究開発法人 国立環境研究所 National Institute for Environmental Studies
Tokyo Woman's Christian University 東京女子大学
名古屋大学 NAGOYA UNIVERSITY
RIKEN R-CCS

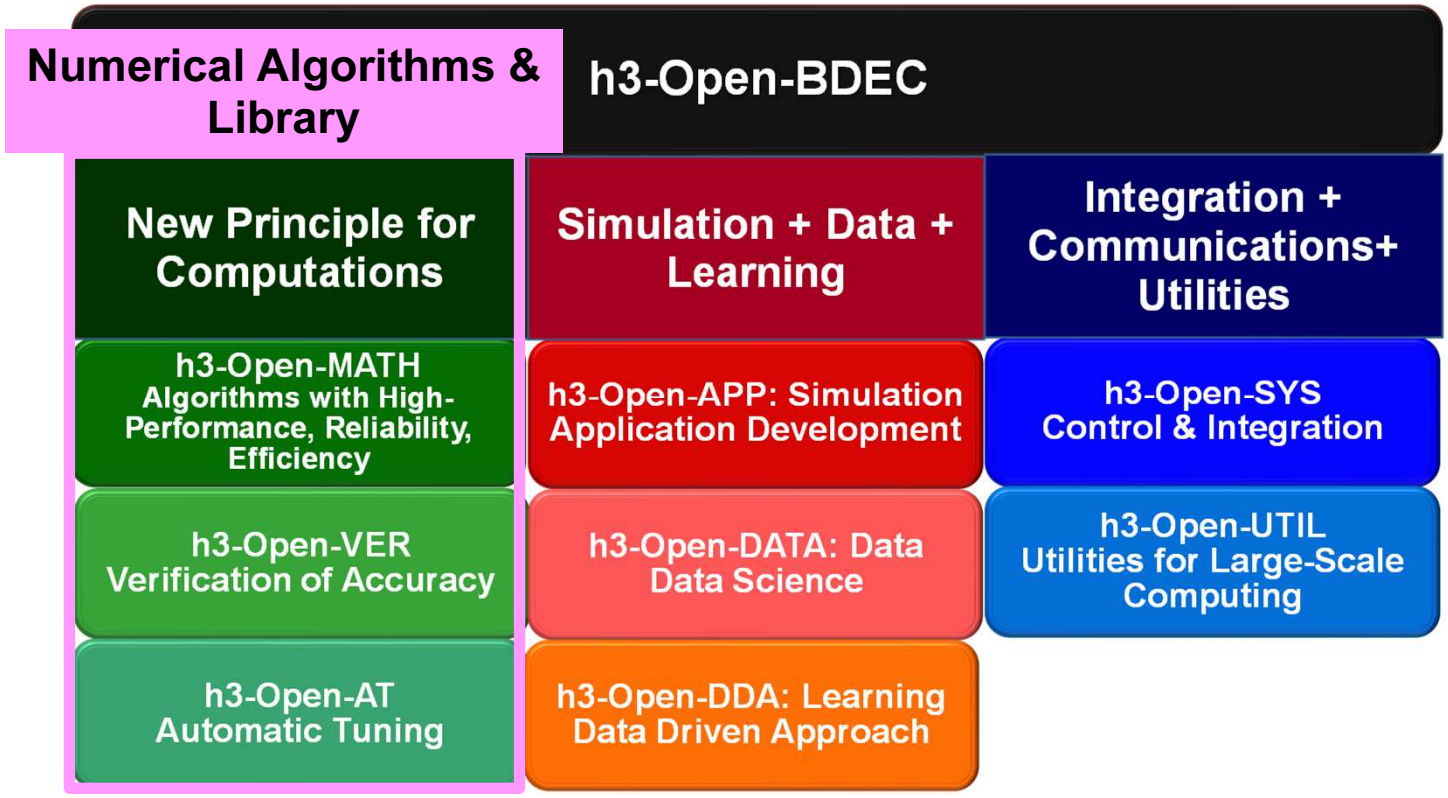
h3-Open-BDEC

Innovative Software Platform for Integration of (S+D+L) on BDEC



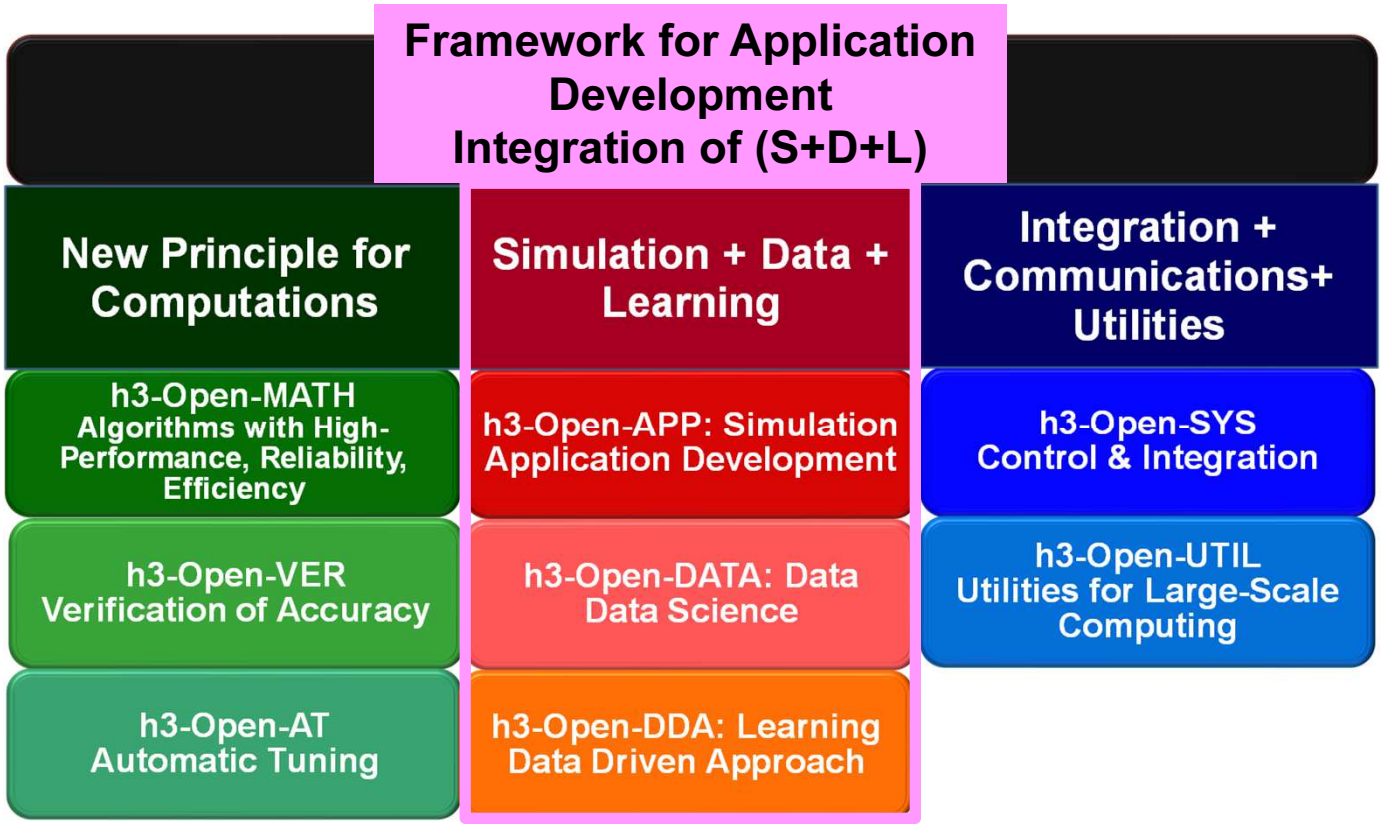
h3-Open-BDEC

Innovative Software Platform for Integration of (S+D+L) on BDEC



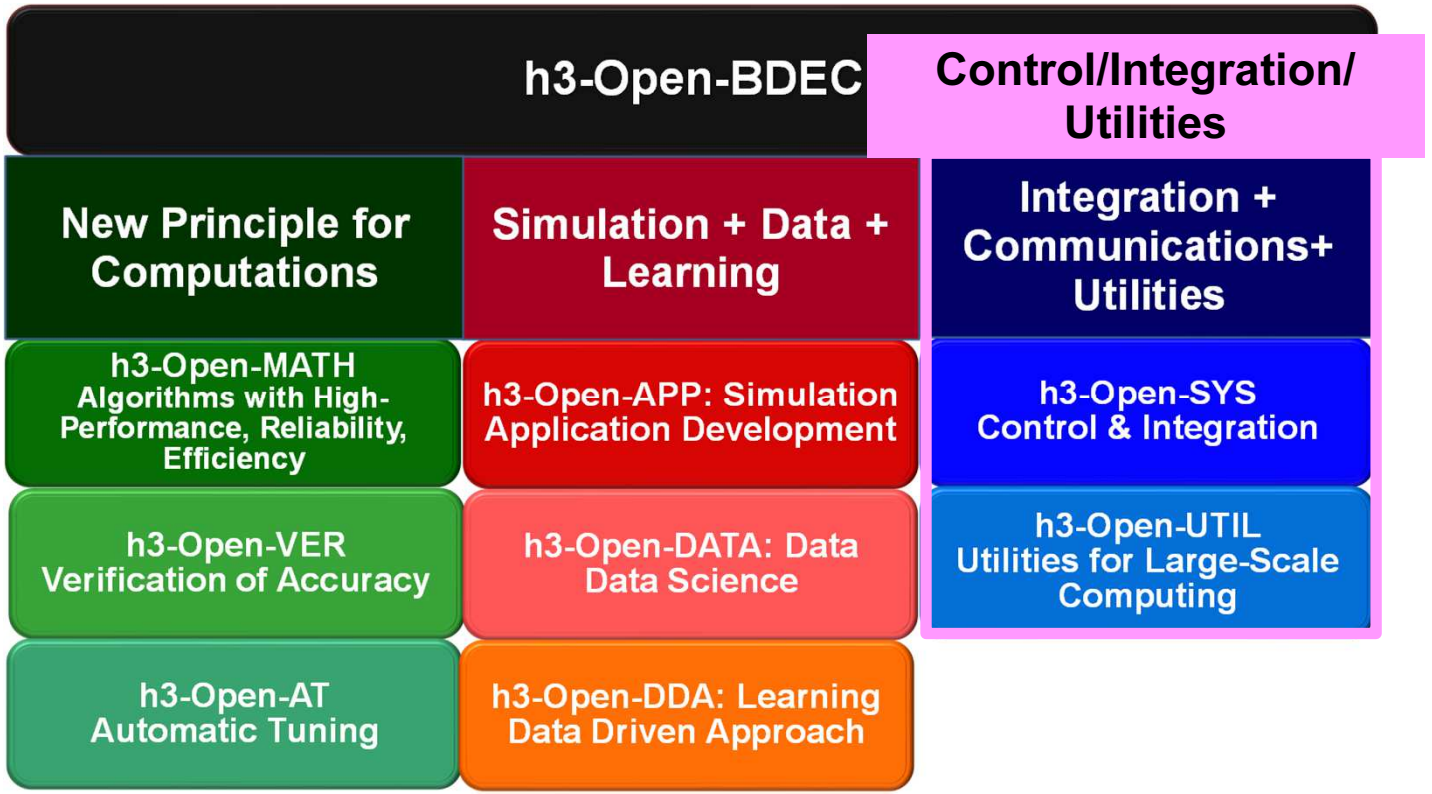
h3-Open-BDEC

Innovative Software Platform for Integration of (S+D+L) on BDEC



h3-Open-BDEC

Innovative Software Platform for Integration of (S+D+L) on BDEC



h3-Open-SYS/EXEC

Control & Integration by Heterogeneous Container Technology: Flexible Interface with 2-layers

Interface: End-Pt. of Application Development, Start-Pt. of Library Development (Hadoop-like)

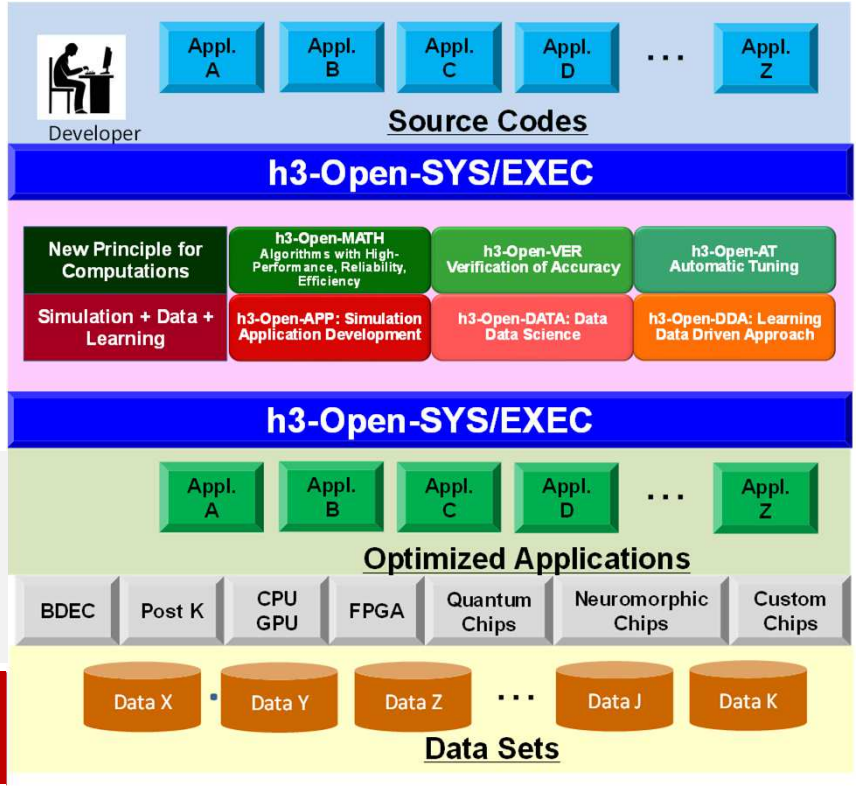
Appl's - h3-Open-BDEC

General/Common Interface for Multiple Applications:
Very Critical/Important, experience in ppOpen-HPC

Harware - Software

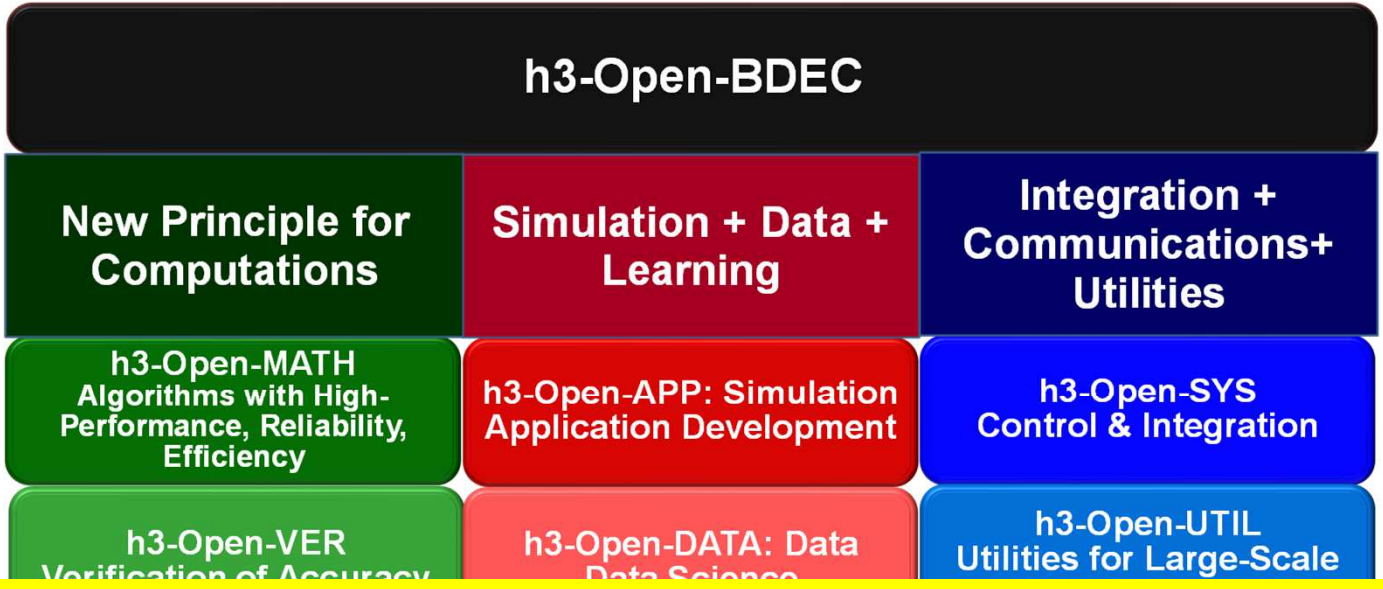
Common Programming Language/Framework generating Optimum Binaries on Various/Heterogeneous Environments: collaboration with h3-Open-AT etc.

Suitable for Heterogenous Workloads/HW Environments in the Exascale/Post Moore Era



h3-Open-BDEC

Innovative Software Platform for Integration of (S+D+L) on BDEC

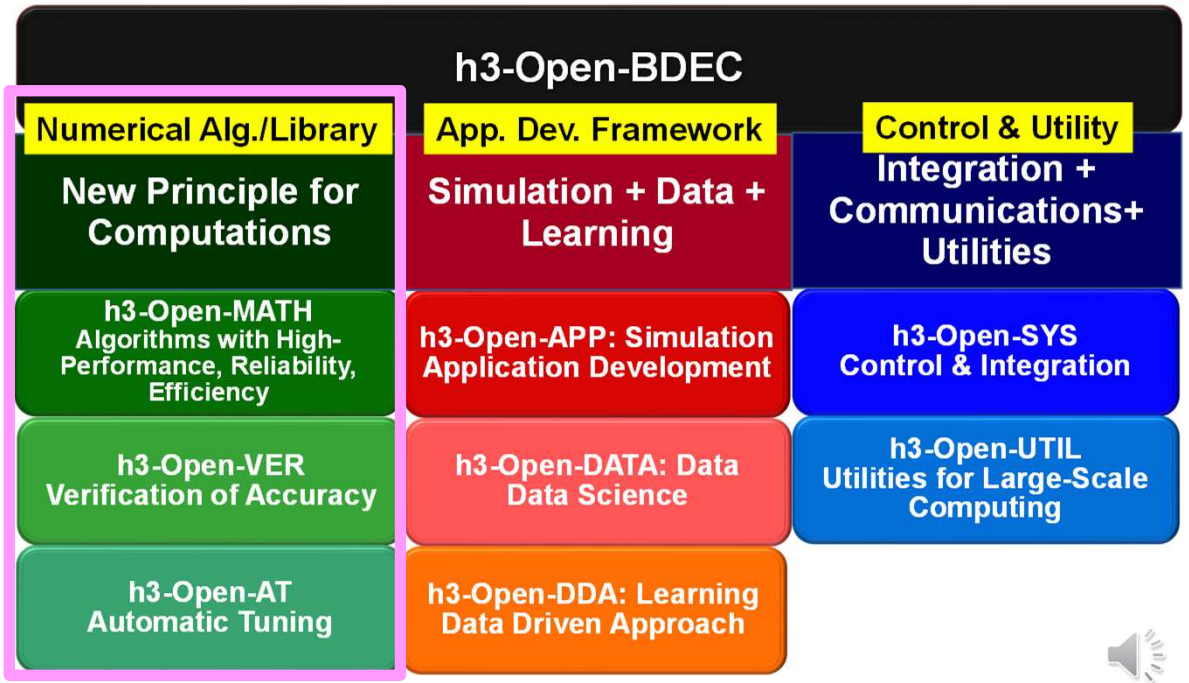


Extracting the maximum performance of the supercomputers with minimum energy consumption

h3-Open-BDEC: Two Significant Innovations

① Methods for Numerical Analysis with High-Performance/High-Reliability/Power-Saving based on the New Principle of Computing by

- ✓ Adaptive Precision
- ✓ Accuracy Verification
- ✓ Automatic Tuning



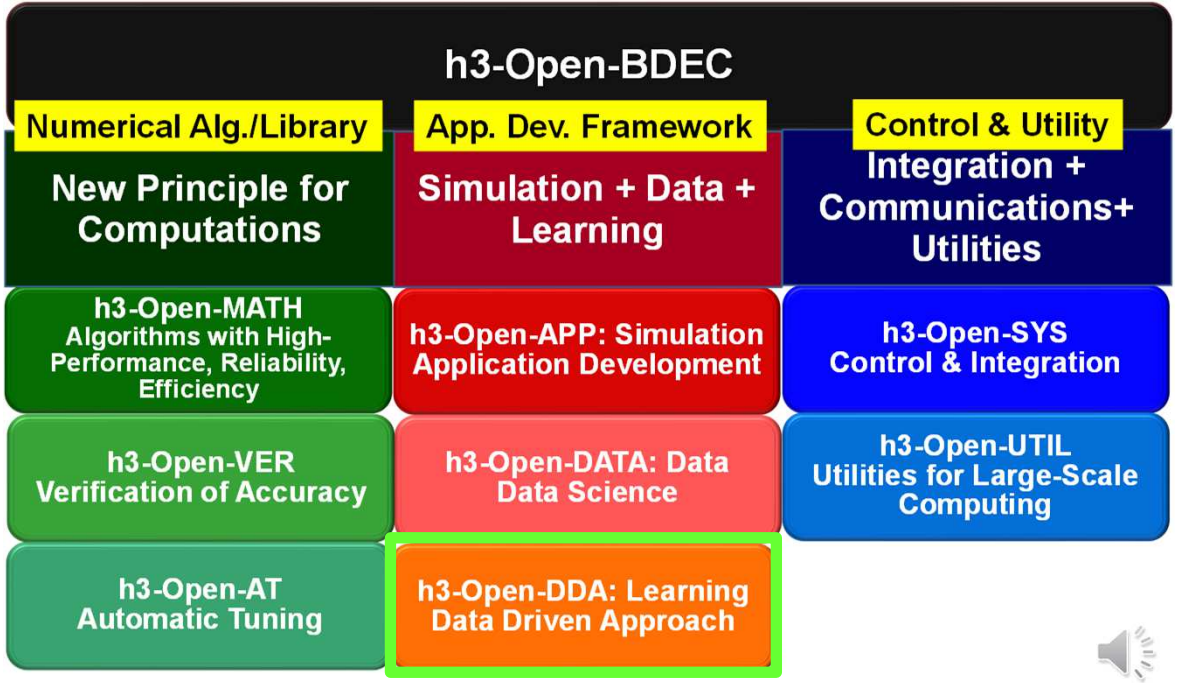
h3-Open-BDEC: Two Significant Innovations

① Methods for Numerical Analysis with High-Performance/High-Reliability/Power-Saving based on the New Principle of Computing by

- ✓ Adaptive Precision
- ✓ Accuracy Verification
- ✓ Automatic Tuning

② Hierarchical Data Driven Approach (*h*DDA) based on machine learning

- ✓ Integration of (S+D+L) **AI for HPC**

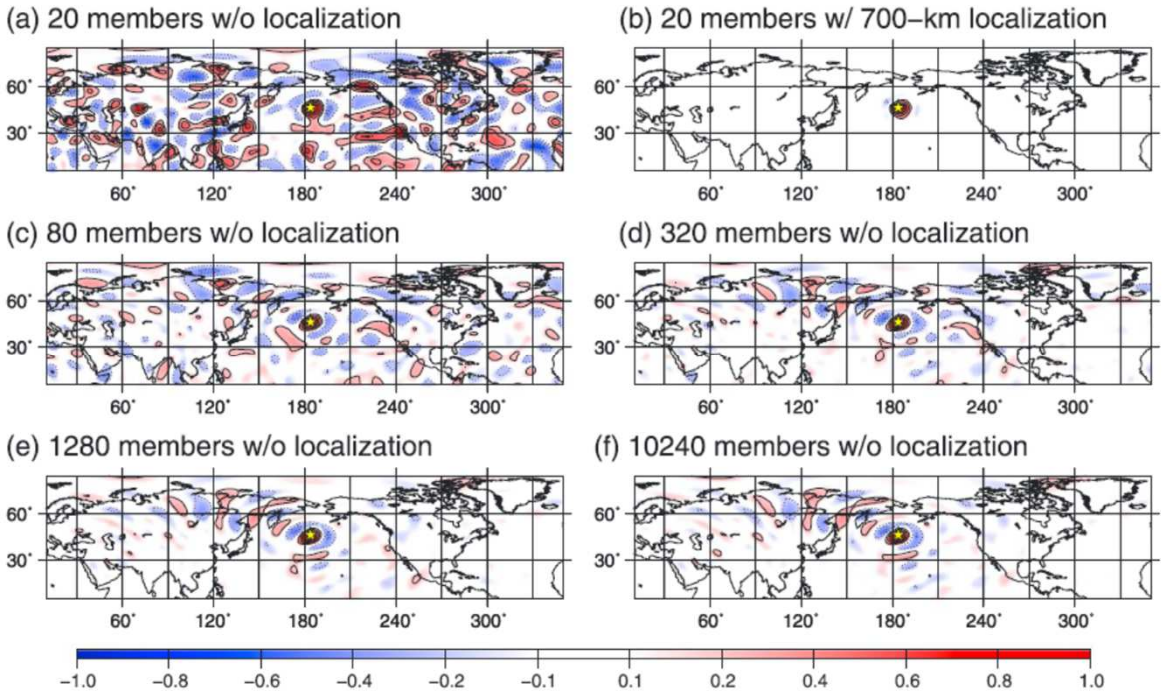


Real-World Scientific Simulations

- Non-Linear: Huge Number of Parameter Studies needed
  - ✓ Reduction of cases is very crucial

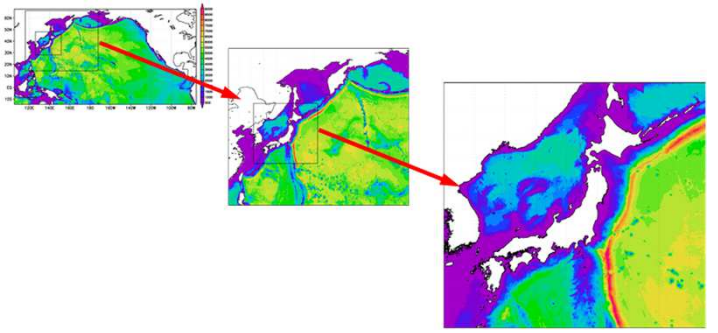
[Miyoshi et al. 2014]

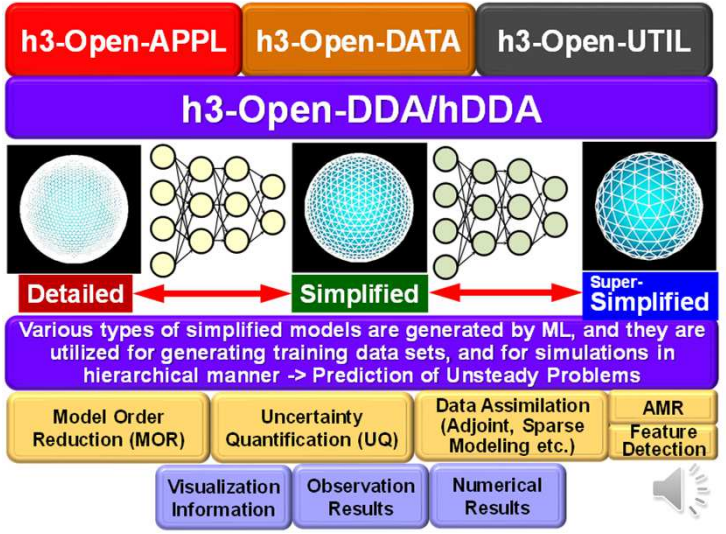
- Data Assimilation
  - ✓ Mid-Range Weather Prediction: 50-100 Ensemble Cases, 1,000 needed for accurate solution.
  - ✓ 50-100 (or fewer) may be enough for accurate solution, if opt. parameters are selected (e.g. by ML),



Hierarchical Data Driven Approach: *h*DDA

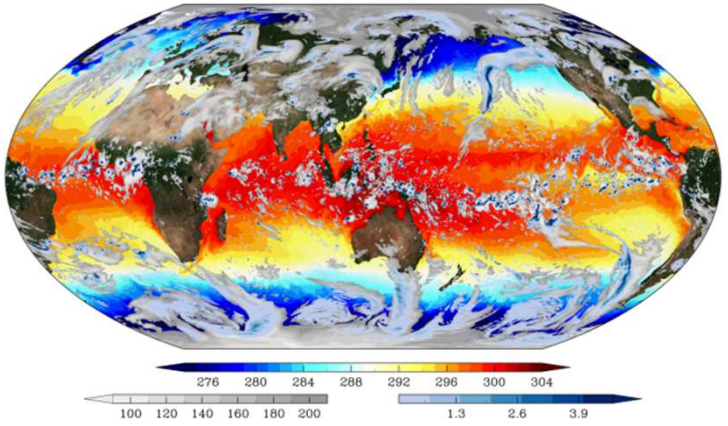
- Data Driven Approach (DDA)
  - Technique of AI/ML is introduced for predicting the results of simulations with different parameters.
  - DDA generally requires O(10^3-10^4) runs for generation of training data.
- ***h*DDA (Hierarchical DDA)**
  - Simplified models with coarser meshes (but preserving original features of physics) for efficient training are constructed automatically by Machine Learning using:
    - Feature Detection, AMR
    - MOR (Model Order Reduction)
    - UQ (Uncertainty Quantification)
    - Sparse Modeling

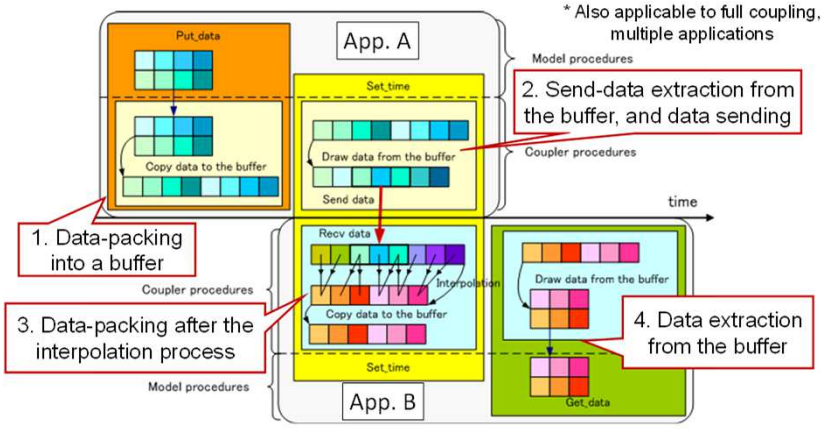




h3-Open-BDEC: Possible Applications (S+D+L)

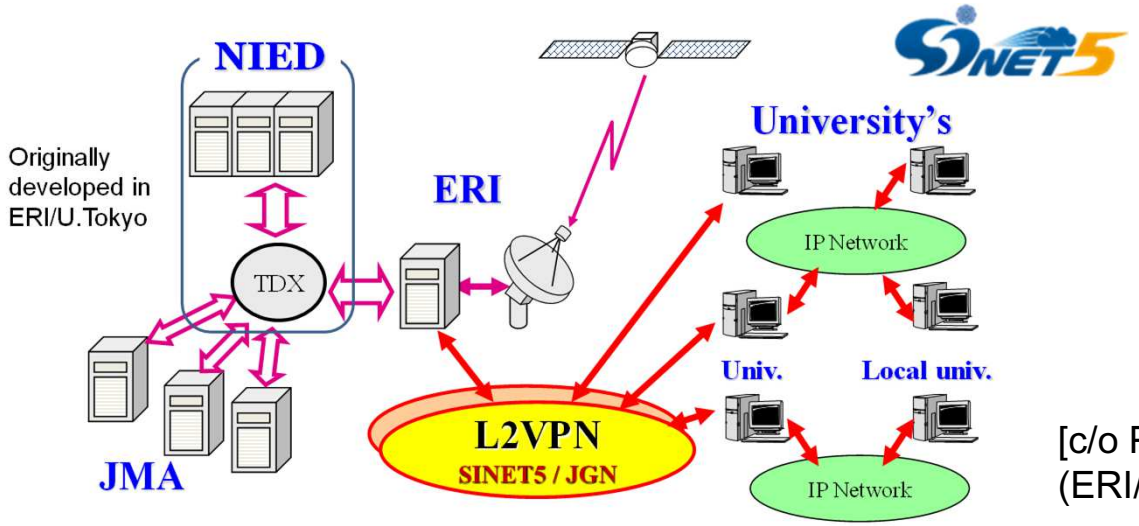
- Simulations/Data Assimilation
  - Very Typical Example of (S+D+L)
- Atmosphere-Ocean Simulations with Data Assimilation
  - AORI/U.Tokyo
  - RIKEN R-CCS
- **Earthquake Simulations with Data Assimilation**
  - **ERI/U. Tokyo**
- Real-Time Disaster Simulations
  - Flood, Tsunami
- (S+D+L) for Existing Simulation Codes (OSS)
  - OpenFOAM





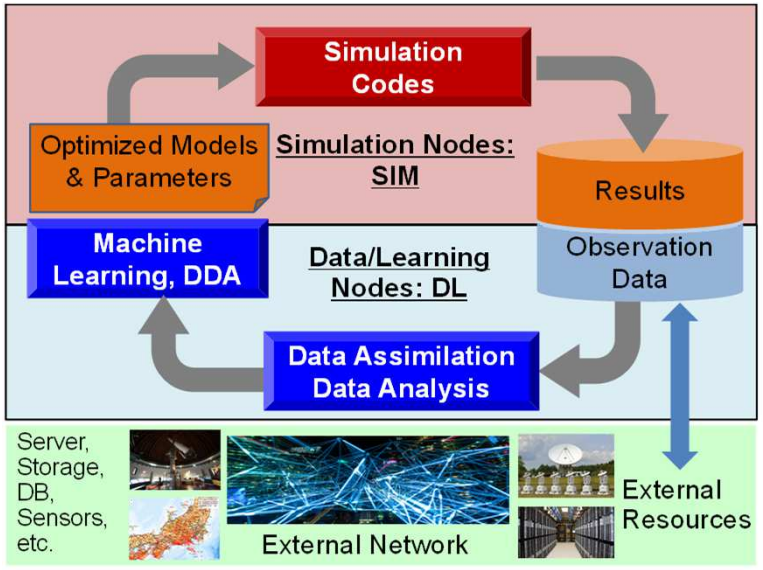
Real-Time Sharing of Seismic Observation is possible in Japan by JDXnet with SINET

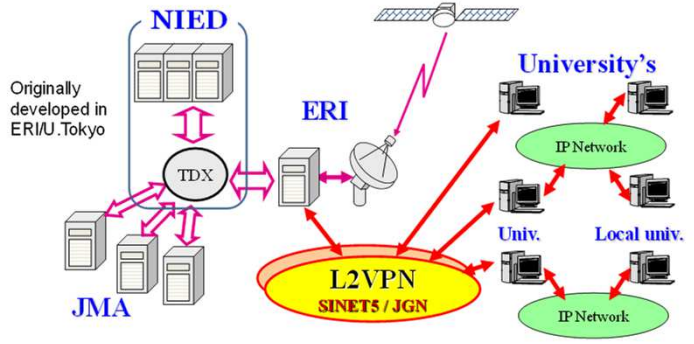
- Seismic Observation Data (100Hz/3-dir's/O(10^3) pts) by JDXnet is available through SINET in Real Time
  - Peta Server in ERI/U.Tokyo: O(10^2) GB/day⇒DL-N of BDEC
  - O(10^5) pts in future including stations operated by industry

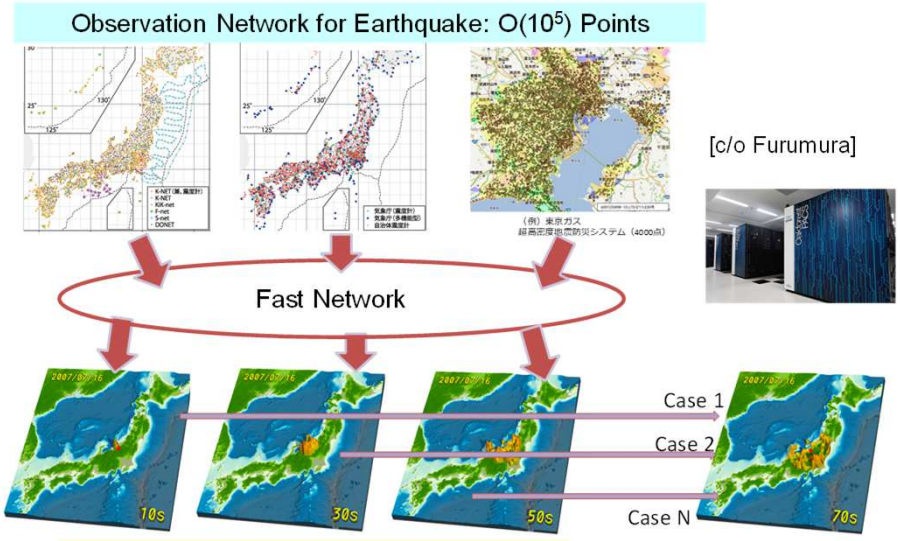


[c/o Prof. H.Tsuruoka (ERI/U.Tokyo)]

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





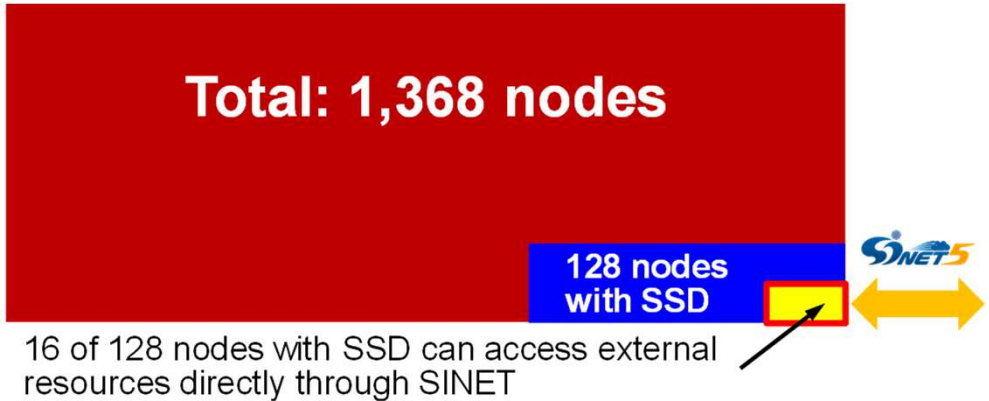


[c/o Furumura]

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

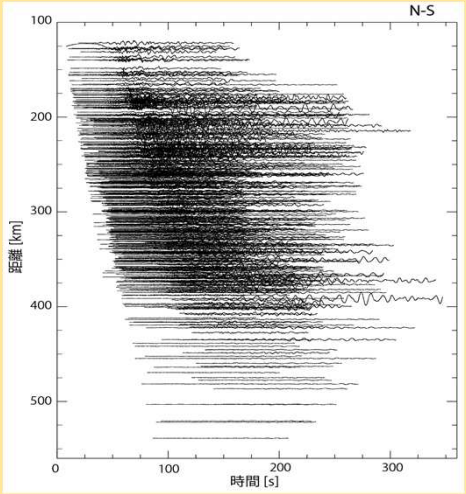
Preliminary Works on Oakbridge-CX (OBCX)

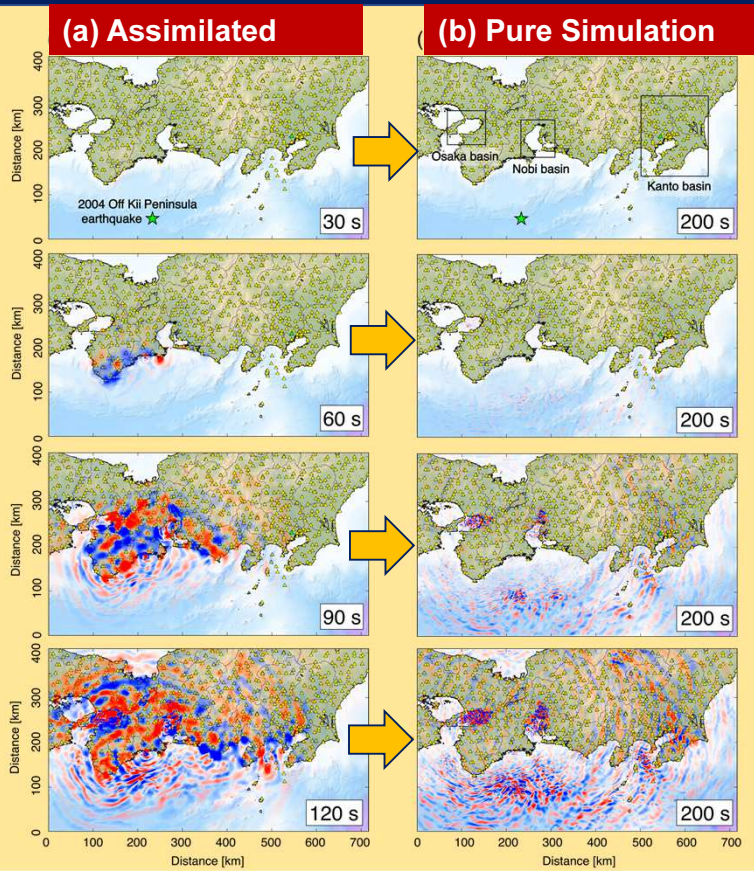
- Intel Xeon Platinum 8280 (Cascade Lake, CLX), Fujitsu
  - 1,368 nodes, 6.61 PF peak, 385.1 TB/sec, 4.2+ PF for HPL **#60 in 55th Top500**
  - Fast Cache: SSD's for 128 nodes: Intel SSD, BeeGFS: 200+TB Fast FS
    - 1.6 TB/node, 3.20/1.32 GB/s/node for R/W
    - Staging, Check-Pointing, Data Intensive Application
    - **16 of these nodes can directly access external resources (server, storage, sensor network etc.) through SINET**
  - Network: Intel Omni-Path, 100 Gbps, Full Bi-Section
  - Storage: DDN EXAScaler (Lustre) 12.4 PB, 193.9 GB/sec
  - Power Consumption: 950.5 kVA
- Prototype of BDEC



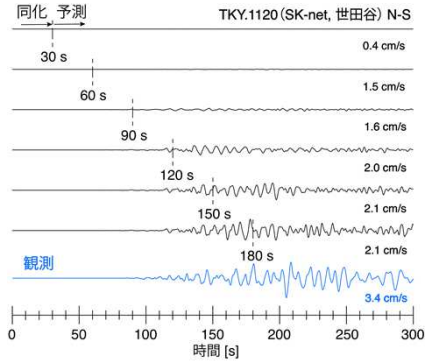
Example of Real-Time Assimilation of (Obs.+Comp.): 2004 Kii Peninsula Earthquake (Mw 7.4) [c/o Oba & Furumura]

○ **Observation** (K-NET, KiK-net 446 pts)

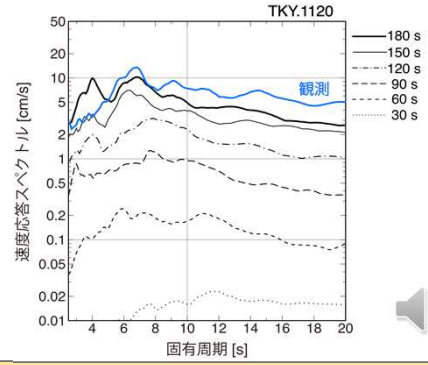




Long Wave Propagation in Tokyo

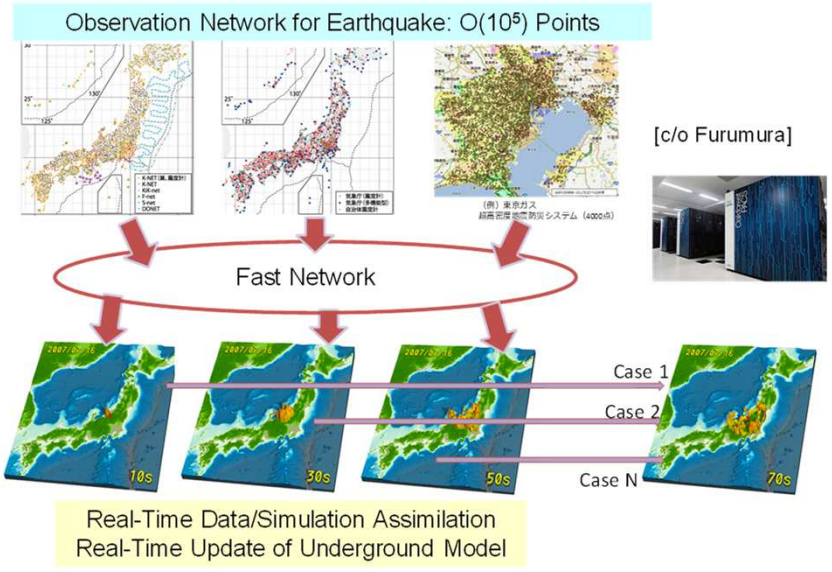


Response Spectrum



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

- Accurate Prediction of Seismic Wave Propagation with Real-Time Data Observation/Assimilation
  - Emergency Information for Safer Evacuation
- 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



- FY.2019
  - Prototype on OBCX
- FY.2020
  - More realistic System on OBCX

h3-Open-BDEC: Summary

http://nkl.cc.u-tokyo.ac.jp/h3-Open-BDEC/

- By Integration of (S+D+L) using **h3-Open-BDEC (Adaptive Precision + *h*DDA)**, total energy consumption (=total computation time) for simulations will be 10% of that by the conventional methods for simulations with parameter studies
- h3-Open-BDEC is the 1st innovative software platform for integration of (S+D+L) on Exascale systems, where computational scientists can achieve such integration without supports by other experts in data analytics and AI/ML.
- Source codes and documents (in English) are open to public for various kinds of computational environments.

Video materials are available …

https://www.cc.u-tokyo.ac.jp/public/sc20.php

- Overview of h3-Open-BDEC (this PDF)
  - h3-Open-BDEC: Innovative Software Platform for Scientific Computing in the Exascale Era by Integrations of (Simulation + Data + Learning)

- More Detailed Topics
  - Accuracy Verification of Parallel Preconditioned Iterative Solvers with Mixed Precision Computing
  - Parallel Multigrid on Manycore Clusters