

Innovative Scientific Computing by Integration of (Simulation+Data+Learning) in Information Technology Center, The University of Tokyo

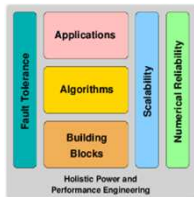


Kengo Nakajima
Information Technology Center
The University of Tokyo

NHR PerfLab Seminar
June 9, 2022

Erlangen, My Favorite Place

- Excellent Colleagues, Unpaved Road
- ESSEX-II (Equipping Sparse Solvers for Exascale) (2015-2018)
 - Leading PI: Gerhard Wellein (FAU)
 - KN was Co-PI of ESSEX-II
 - DFG/SPPEXA (Germany) & JST/CREST (Japan)
- Member of Scientific Advisory Board, NHR (2021-)



- Supercomputing Research Division, Information Technology Center ,The University of Tokyo (SCD/ITC/U.Tokyo)
- JHPCN, HPCI, JCAHPC and NHR
- Supercomputers in SCD/ITC/U.Tokyo
- Integration of (Simulation+Data+Learning)
 - Wisteria/BDEC-01
 - h3-Open-BDEC
- Future Perspective

Information Technology Center The University of Tokyo (ITC/U.Tokyo)



東京大学
THE UNIVERSITY OF TOKYO

- Campus/Nation-wide Services on Infrastructure for Information, related Research & Education
- Established in 1999
 - 4 Divisions
 - Education System, Network, Digital Library/Data Science, **Supercomputing**
- Core Institute of Nation-wide Infrastructure Services/Collaborative Research Projects
 - Joint Usage/Research Center for Interdisciplinary Large-scale Information Infrastructures (JHPCN) (2010-)
 - HPCI (HPC Infrastructure) (2012-)



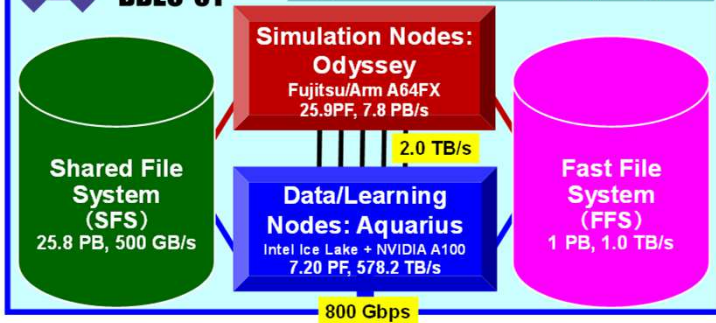
東京大学情報基盤センター
INFORMATION TECHNOLOGY CENTER, THE UNIVERSITY OF TOKYO

Supercomputing Research Division (SCD)

- Services & Operations of Supercomputer Systems, Research, Education
- History
 - Supercomputing Center, the University of Tokyo (1965~1999)
 - Oldest Academic Supercomputer Center in Japan
 - Nation-Wide, Joint-Use Facility: Users are not limited to researchers and students of U.Tokyo
 - Information Technology Center (1999~) (4 divisions)
- 10+ Faculty Members (including part-time members)
 - Architecture, System S/W, Algorithms, Applications, GPU
- 8 Technical Staffs
- 2,600+ Users, 55+% from Outside of U.Tokyo



Platform for Integration of (S+D+L)
Big Data & Extreme Computing



External Resources



External Network



External Resources



Simulation Nodes
(Odyssey)

June 6, 2022



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

Reedbush (HPE, Intel BDW + NVIDIA P100 (Pascal))

- Prototype of “Wisteria/BDEC-01” for Integration of (S+D+L)
- July 2016 – November 2021 (Retired)
- Our First GPU Cluster, 3.36 PF

Oakforest-PACS (OFP) (Fujitsu, Intel Xeon Phi (KNL))

- JCAHPC (U.Tsukuba, U.Tokyo), October 2016 – March 2022
- 25 PF, #39 in 58th TOP 500 (November 2021)

Oakbridge-CX (OBCX) (Fujitsu, Intel Xeon CLX)

- July 2019 – June 2023
- 6.61 PF, #110 in 58th TOP500



Wisteria/BDEC-01 (Fujitsu)

- Simulation Nodes (Odyssey): A64FX (#17)
- Data/Learning Nodes (Aquarius) : Icelake + A100 (#106)
- 33.1 PF, Operation started on May 14, 2021
- 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)



Reedbush



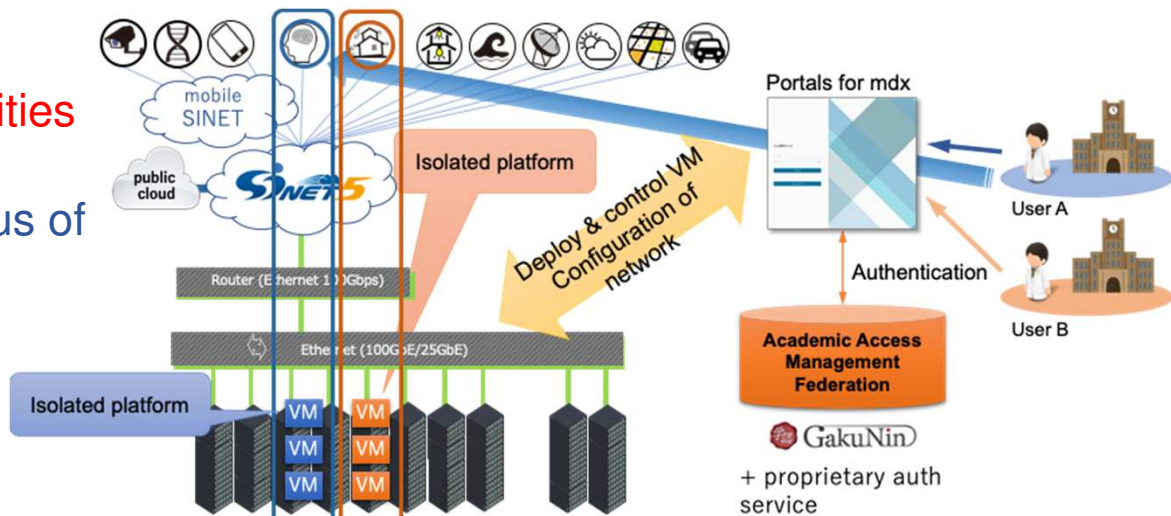
Oakforest-PACS



Oakbridge-CX

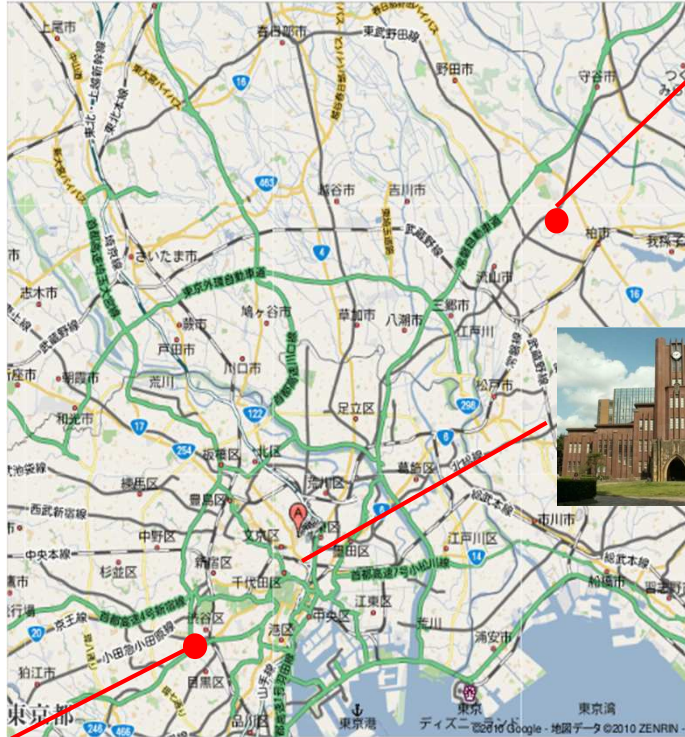
“Data Platform”

- High-performance virtualization environment focusing on leveraging data with security like “Cloud” using same HW as supercomputers.
- National joint usage system jointly operated by 9 universities and 2 research institutes
- Located at Kashiwa II campus of U.Tokyo



June 6, 2022

Location of Information Technology Center



Kashiwa



(Oakforest-PACS)



Oakbridge-CX

Kashiwa II (new)



Hongo



Wisteria/BDEC-01



mdx

Komaba

June 6, 2022

- Supercomputing Research Division, Information Technology Center ,The University of Tokyo (SCD/ITC/U.Tokyo)
- **JHPCN, HPCI, JCAHPC and NHR**
- Supercomputers in SCD/ITC/U.Tokyo
- Integration of (Simulation+Data+Learning)
 - Wisteria/BDEC-01
 - h3-Open-BDEC
- Future Perspective

Users of (Public) Supercomputers in Japan

- Generally, major supercomputers are operated by National Universities in Japan (and Public Research Inst., e.g. RIKEN, AIST, JAMSTEC etc.)
 - Nation-Wide Shared Systems
 - Users are not limited to those from each university
- Main Users
 - Faculty Members/Researchers of Universities/Colleges
 - Students (Graduate, Undergraduate)
 - Researchers in Public Research Institutes (RIKEN, AIST etc.)
 - Basically, limited to “Residents in Japan” by Japanese Trading Law
- After FY. 2008
 - Users from Industry (limited resources (less than total 10%), special qualification)
 - Japanese government strongly recommends Industry-Academia Collaboration
 - Non-Residents (MOU, Research Projects) with several restrictions (e.g. manuals)

In Japan, users usually pay fee for using supercomputers (cost for electricity etc.)

- 1.0+MUSD/year needed for operation of a system with 1MW
- First Come, First Served
 - e.g. Wisteria/BDEC-01 (Odyssey) with A64FX (same as Fugaku)
 - Single node (48 cores) with 2TB Storage: 430 Euro/Year (140JPY= 1 Euro)
 - Users purchase “node-hours (token)”
- Nation-Wide Programs using Supercomputers
 - Proposal-based, Free
 - JHPCN (2010-)
 - Fundamental Research
 - HPCI (2012-)
 - Large-Scale Computing (Simulations)
- Each Supercomputer Center has its own research program
- Recently, 30+% of resources are spent for such programs/projects



HPCI High Performance
Computing Infrastructure

JHPCN (1/4)

<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 (corresponding to NHR in Germany)
 - 7 “Imperial” Universities + Tokyo Tech
 - Core Institute: ITC/U.Tokyo
 - Total 140+PFLOPS (May 2022)
- Promotion of collaborative (fundamental, interdisciplinary) research projects using facilities & human resources in 8 Centers

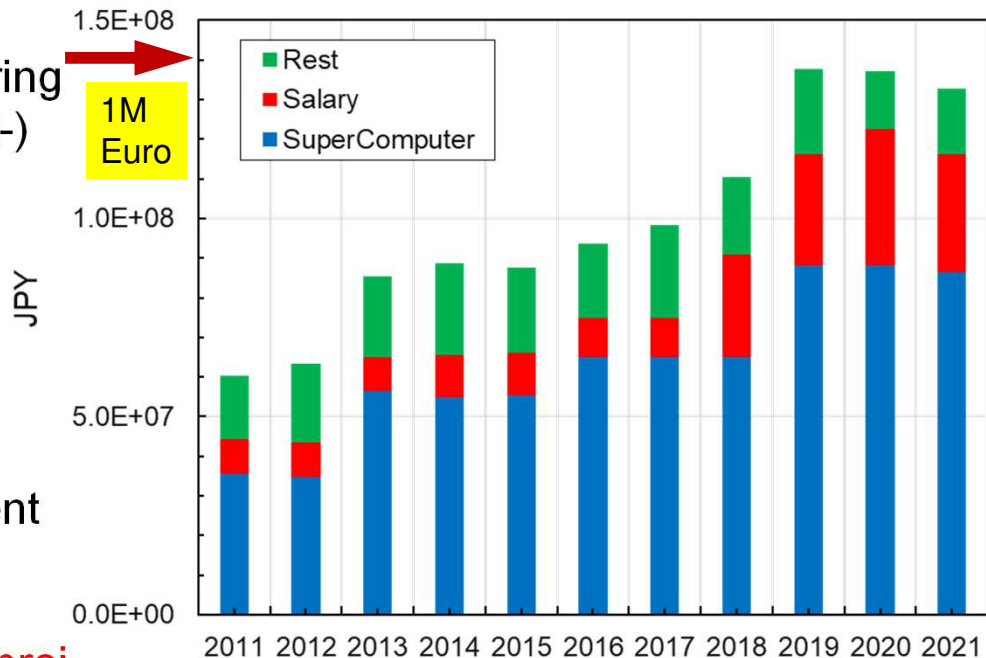


JHPCN (2/4)

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



- Categories for Projects
 - Computational Science & Engineering
 - Data Science/Data Analytics (2022-)
 - General
 - International, Industry
 - 63 Projects in FY.2022
 - 55:CSE, 8:Data
 - Many of the projects are by multiple centers
- Budget
 - Supported by Japanese Government
 - 1/3 of Supercomputer Fee is by 8 centers, Cost for H/W not included
 - Budget is up to 20K Euro for each proj.

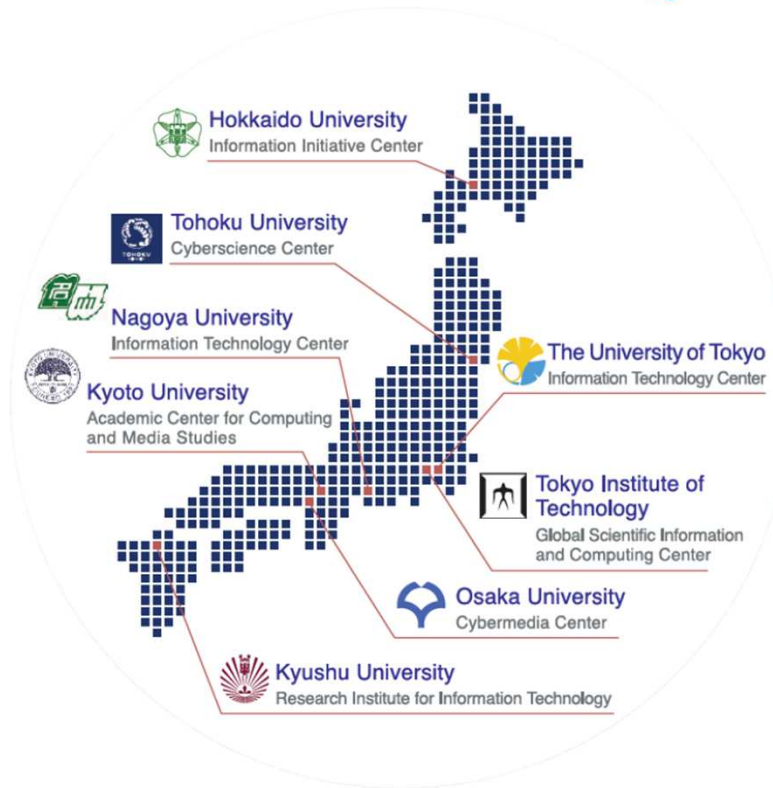


JHPCN (3/4)

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

(KN's personal point of view)

- Before JHPCN (2010)
 - 8 centers were independent, almost NO collaborations
 - NOT necessarily active in research
- After JHPCN
 - Collaborations among centers are much more active, as well as CS-CSE collaboration through projects
 - KN has collaborators in almost all centers
 - Joint proposal for external funding
 - ppOpen-HPC@JST CREST -> ESSEX-II
 - h3-Open-BDEC@JSPS



JHPCN (4/4)

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

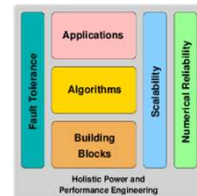


- There are no “assignments” of application area for each center, as NHR
 - e.g.: In U.Tokyo, Weather/Climate, Earthquakes, Engineering, Materials Science are traditional area of usage
- Varieties of HW
 - Cloud: Hokkaido, Osaka
 - Vector: Tohoku, Kyoto
 - GPU: Tokyo Tech
 - Xeon Phi: Hokkaido, Tokyo, Kyoto
 - Integrated/Heterogeneous Systems: Tokyo, Nagoya, Kyushu



ESSEX-II is still continuing in JHPCN Projects

- Innovative Multigrid Methods (FY.2018-)
 - ESSEX-II + ExaStencils
 - International Project
 - Leading PI must be residents, while at least one of Co-PI's must be non-residents
 - Hokkaido, Tokyo, Nagoya, Kyushu
- Numerical Library with High-Performance/Adaptive-Precision/High-Reliability (FY.2018-)
 - ESSEX-II
 - Hokkaido, Tokyo, Tokyo Tech, Nagoya, Kyushu



HPCI (1/2)

HPCI High Performance
Computing Infrastructure

<https://www.hpci-office.jp/folders/english>

- High-Performance Computing Infrastructure

- Structure

- National Flagship System

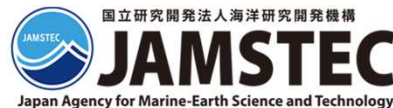
- K, Fugaku (RIKEN R-CCS)

- 2nd Tier

- JHPCN Centers (8 Universities)
- University of Tsukuba
- NII, JAMSTEC, AIST

- Large-Scale Shared Storage (50+PB)

- East Hub: Kashiwa (U.Tokyo)
- West Hub: Kobe (RIKEN R-CCS)



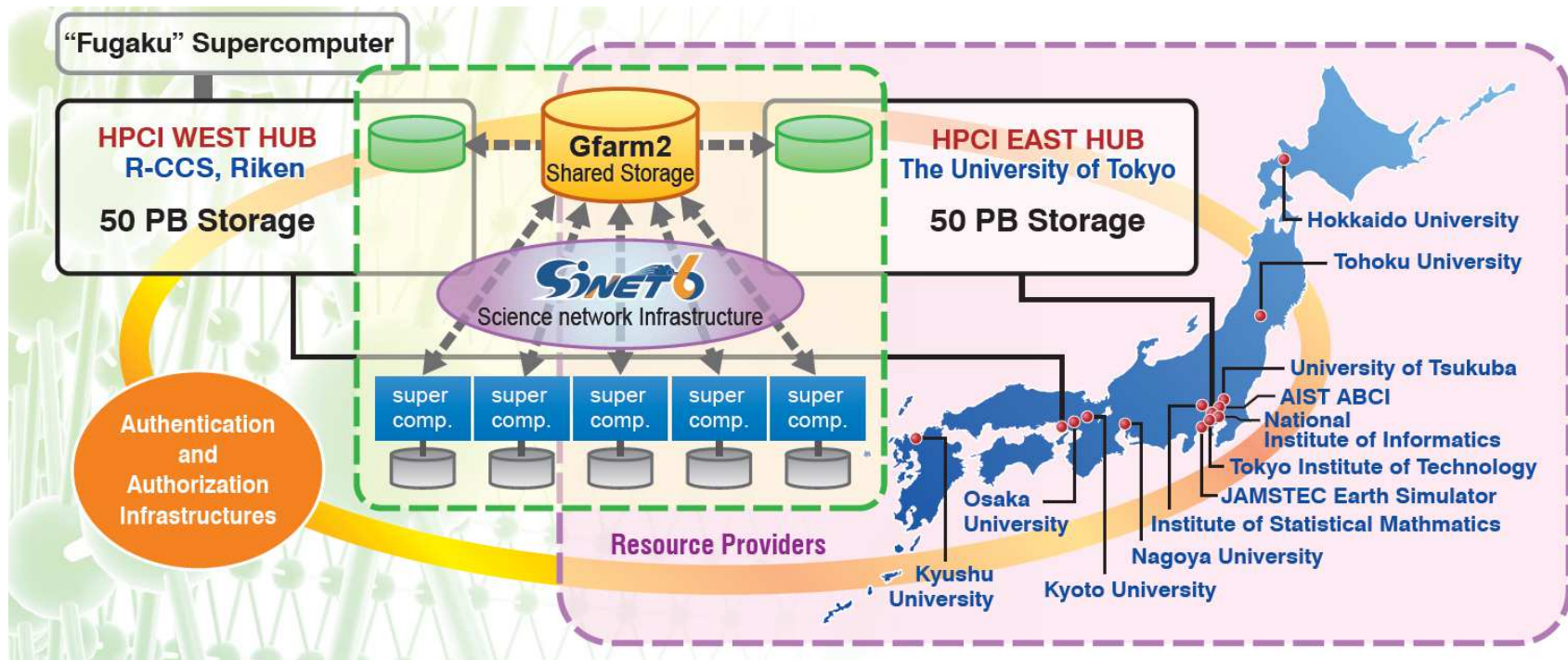
- Focusing on Large-Scale Simulations (not fundamental research)

- Budget for each project is up to 10x of JHPCN project

HPCI (2/2)

<https://www.hpci-office.jp/folders/english>

HPCI High Performance
Computing Infrastructure



2001-2005

2006-2010

2011-2015

2016-2020

2021-2025

2026-2030

Hitachi SR8000
1,024 GF

Hitachi SR11000
J1, J2
5.35 TF, 18.8 TF

Hitachi SR16K/M1
Yayoi
54.9 TF

Hitachi
SR2201
307.2GF

Hitachi
SR8000/MPP
2,073.6 GF


OBCX
(Fujitsu)
6.61 PF

Hitachi HA8000
T2K Today
140 TF

Oakforest-
PACS (Fujitsu)
25.0 PF

OFP-II
100+ PF

Fujitsu FX10
Oakleaf-FX
1.13 PF

 **Wisteria**
BDEC-01 Fujitsu
33.1 PF

BDEC-
02
250+ PF

Reedbush-
U/H/L (SGI-HPE)
3.36 PF

Ipomoea-01 25PB

Ipomoea-
03

Ipomoea-02

Supercomputers @ITC/U.Tokyo

2,600+ Users

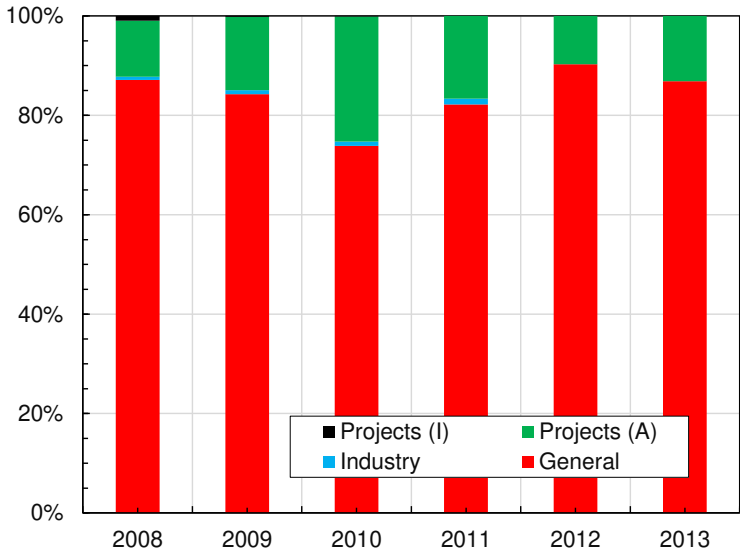
55+% outside of U.Tokyo

Ratio of Assigned Resources (1/3)

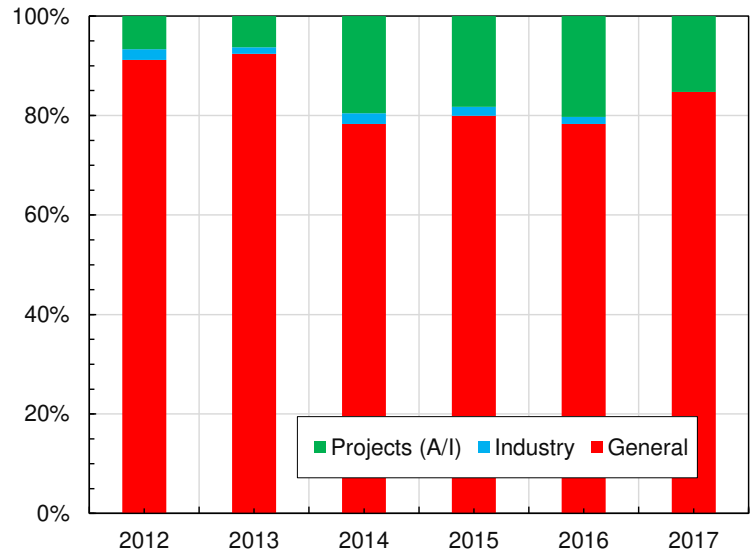
General **Industry (NOT Free)**

Projects: without Fee, A: Academia, I: Industry

T2K Today
2008-2013, AMD Opteron, 140TF



Oakleaf/Oakbridge-FX
2012-2017, Commercial Version of K, 1.2+PF



JCAHPC

<http://jcahpc.jp/eng/index.html>



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO



JCAHPC

- Joint Center for Advanced High Performance Computing, since 2013
 - University of Tsukuba & University of Tokyo
 - Budgets of 2 Centers are combined
 - Promotion on Computational Science, Design/Procurement/Operation of Large-scale Systems
- Oakforest-PACS (OFP), 1st System of JCAHPC
 - 8,208 Intel Xeon Phi, 25PF, Fujitsu
 - Top500 (#6 (Nov.2016), #1 in Japan)
 - Retired in the end of March 2022 (#39 (Nov.2021))
- National Flagship System in FY.2019/2020
 - Between K and Fugaku



Finalist of Hans Meuer Award (Best Paper), ISC-HPC 2018

Chebyshev Filter Diagonalization on Modern Manycore Processors and GPGPUs (Kreutzer, Ernst, Bishop, Fehske, Hager, Nakajima, Wellein)



Chebyshev Filter Diagonalization on Modern Manycore Processors and GPGPUs (Kreutzer, Ernst, Bishop, Fehske, Hager, Nakajima, Wellein)

ISC-HPC 2018 Hans Meuer Award Finalist

Filter Diagonalization

- Computations of All Eigen-Pairs of Large Sparse Matrix
- ChebFD Scheme
 - Apply Chebyshev polynomial filter to search vectors
 - Orthogonalize filtered vectors, Compute Ritz-pairs and restart if necessary

Summary

- ChebFD polynomial filter for Intel KNL and Nvidia GPGPUs
- High Flop-rates for sparse matrix problems & large block vectors
- Subspace blocking allows for:
 - Easy & efficient usage of slow memories via transparent access and
 - Pipelining of communication and computation across subspaces
- Weak scaling: Approx. 0.5 PFLOPD on 2,048 nodes of OFP & PizDaint
- Full scale OFP system: 1.5+ PFLOPS for 100 inner eigenpairs of sparse matrix with 16 billion rows

Polynomial degree

Single kernel

Algorithm 1 Application of the ChebFD polynomial filter to block vectors.

```

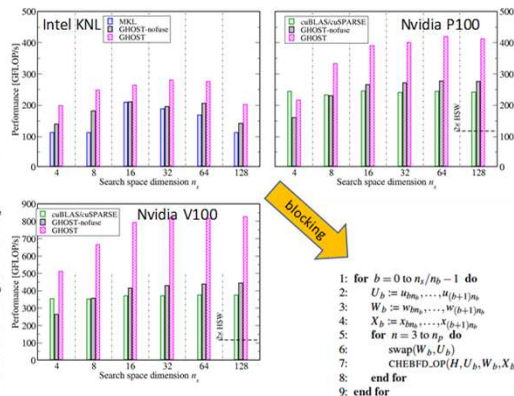
1:  $U := u_1, \dots, u_{n_b}$ 
2:  $W := w_1, \dots, w_{n_b}$ 
3:  $X := x_1, \dots, x_{n_b}$ 
4:  $U \leftarrow (\alpha H + \beta I)X$ 
5:  $W \leftarrow 2(\alpha H + \beta I)U - X$ 
6:  $X \leftarrow g_0 c_0 X + g_1 c_1 U + g_2 c_2 W$ 
7: for  $p = 3$  to  $n_p$  do
8:   swap( $W, U$ )
9:    $W \leftarrow 2(\alpha H + \beta I)U - W$ 
10:   $\eta_p \leftarrow \langle W, U \rangle$ 
11:   $\mu_p \leftarrow \langle U, U \rangle$ 
12:   $X \leftarrow X + g_p c_p W$ 
13: end for
  
```

spmmv

Block vectors of size $n_s \cdot n$

Sparse matrix multiple vector multiplication

Optimistic Comp. Intensity $I(n_b) = 80 + 260/n_b B$

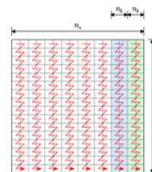


```

1: for  $b = 0$  to  $n_s/n_b - 1$  do
2:    $U_b := w_{n_b \cdot b}, \dots, u_{(b+1)n_b}$ 
3:    $W_b := w_{n_b \cdot b}, \dots, w_{(b+1)n_b}$ 
4:    $X_b := x_{n_b \cdot b}, \dots, x_{(b+1)n_b}$ 
5:   for  $n = 3$  to  $n_p$  do
6:     swap( $W_b, U_b$ )
7:     CHEBFD.OP( $H, U_b, W_b, X_b$ )
8:   end for
9: end for
  
```

Vector subspace blocking

Datalayout

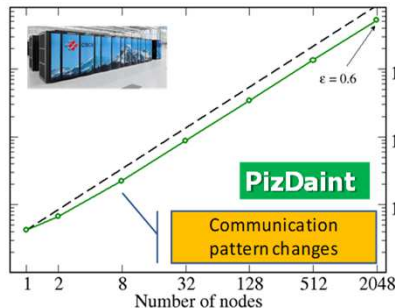
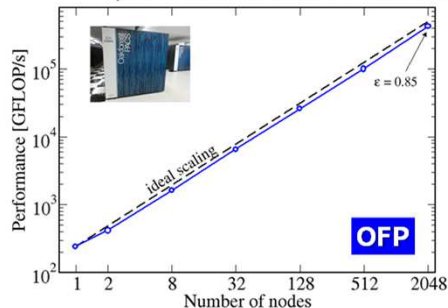


Implementation:

- Intel KNL: SIMD Intrinsics
- Nvidia: CUDA+ opt. reduction

Bottleneck analysis:

- Intel KNL: Core execution
- Nvidia P100: L1 bandwidth
- Nvidia V100: HBM bandwidth (measured: 780 GB/s)



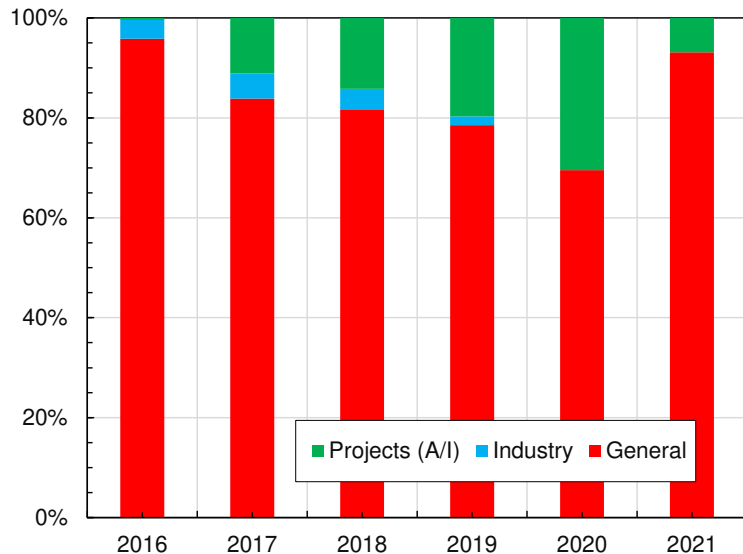
Ratio of Assigned Resources (2/3)

■ General ■ Industry (NOT Free)

■ Projects: without Fee, A: Academia, I: Industry

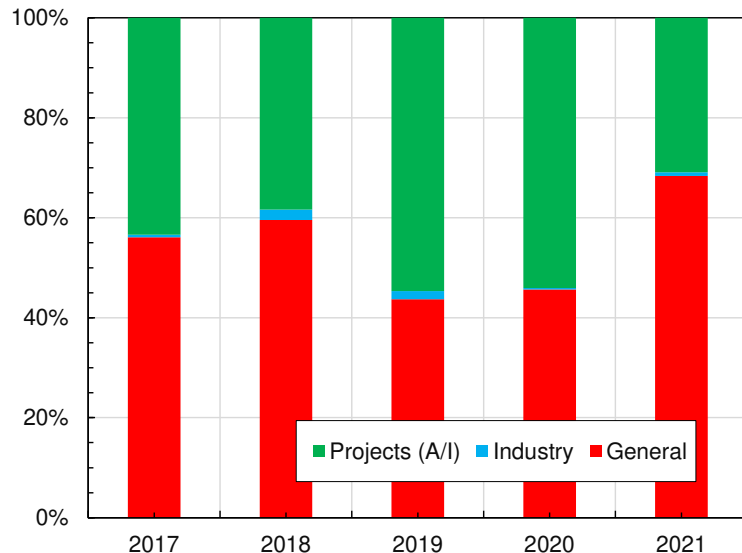
Reedbush U/H/L

2016-2021, Intel BDW+P100, 3.36PF



Oakforest-PACS (OFP)

2016-2021, Intel Xeon Phi, 25 PF



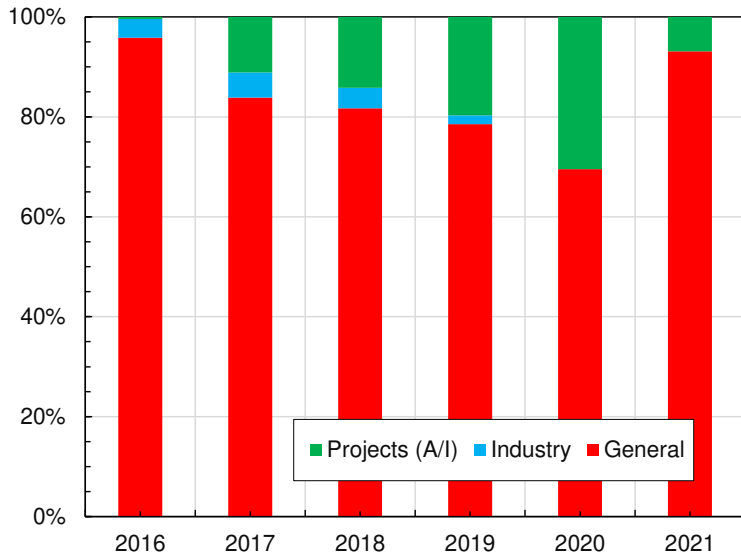
Ratio of Assigned Resources (2/3)

■ General ■ Industry (NOT Free)

■ Projects: without Fee, A: Academia, I: Industry

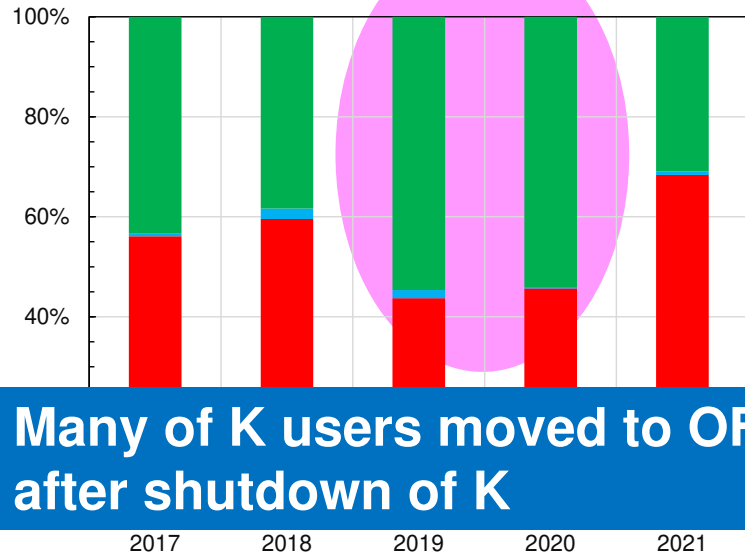
Reedbush U/H/L

2016-2021, Intel BDW+P100, 3.36PF



Oakforest-PACS (OFP)

2016-2021, Intel Xeon Phi, 3.36PF



Many of K users moved to OFP after shutdown of K

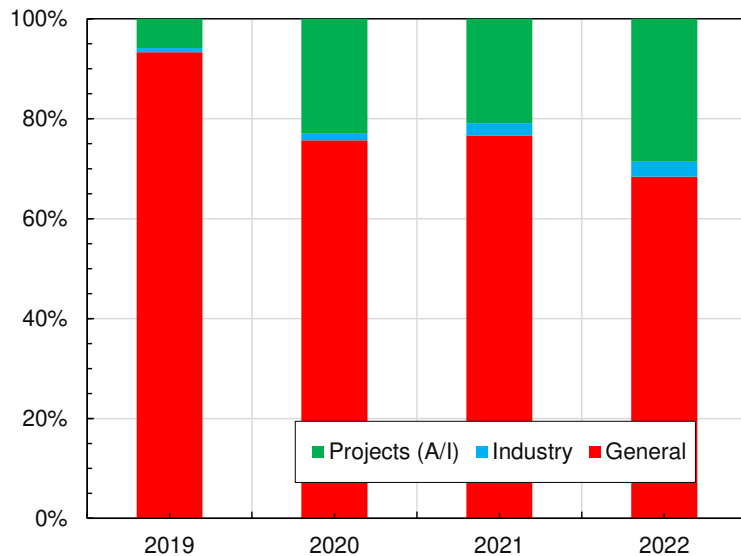
Ratio of Assigned Resources (3/3)

■ General ■ Industry (NOT Free)

■ Projects: without Fee, A: Academia, I: Industry

Oakbridge-CX (OBCX)

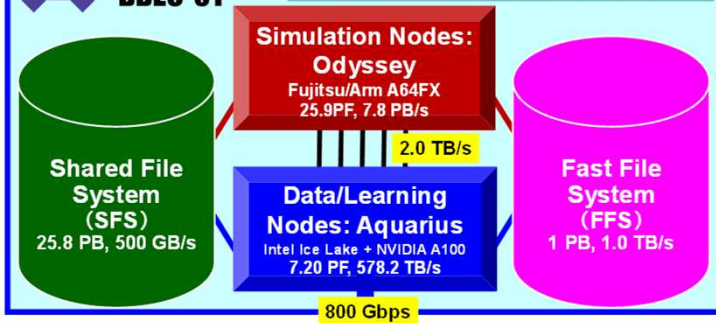
2019-, Intel Xeon Cascadelake, 6.61PF



- Supercomputing Research Division, Information Technology Center ,The University of Tokyo (SCD/ITC/U.Tokyo)
- JHPCN, HPCI, JCAHPC and NHR
- **Supercomputers in SCD/ITC/U.Tokyo**
- Integration of (Simulation+Data+Learning)
 - Wisteria/BDEC-01
 - h3-Open-BDEC
- Future Perspective



Platform for Integration of (S+D+L)
Big Data & Extreme Computing



External Resources



External Network



External Resources



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

Reedbush (HPE, Intel BDW + NVIDIA P100 (Pascal))

- Prototype of “Wisteria/BDEC-01” for Integration of (S+D+L)
- July 2016 – November 2021 (Retired)
- Our First GPU Cluster, 3.36 PF

Oakforest-PACS (OFP) (Fujitsu, Intel Xeon Phi (KNL))

- JCAHPC (U.Tsukuba, U.Tokyo), October 2016 – March 2022
- 25 PF, #39 in 58th TOP 500 (November 2021)

Oakbridge-CX (OBCX) (Fujitsu, Intel Xeon CLX)

- July 2019 – June 2023
- 6.61 PF, #110 in 58th TOP500



Wisteria/BDEC-01 (Fujitsu)

- Simulation Nodes (Odyssey): A64FX (#17)
- Data/Learning Nodes (Aquarius) : Icelake + A100 (#106)
- 33.1 PF, Operation started on May 14, 2021
- 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)



Reedbush



Oakforest-PACS



Oakbridge-CX

2001-2005

2006-2010

2011-2015

2016-2020

2021-2025

2026-2030

Hitachi SR8000
1,024 GF

Hitachi SR11000
J1, J2
5.35 TF, 18.8 TF

Hitachi SR16K/M1
Yayoi
54.9 TF

Hitachi
SR2201
307.2GF

Hitachi
SR8000/MPP
2,073.6 GF


OBCX
(Fujitsu)
6.61 PF

Hitachi HA8000
T2K Today
140 TF

Oakforest-
PACS (Fujitsu)
25.0 PF

OFP-II
100+ PF

Fujitsu FX10
Oakleaf-FX
1.13 PF

 **Wisteria**
BDEC-01 Fujitsu
33.1 PF

BDEC-
02
250+ PF

Reedbush-
U/H/L (SGI-HPE)
3.36 PF

Ipomoea-01 25PB

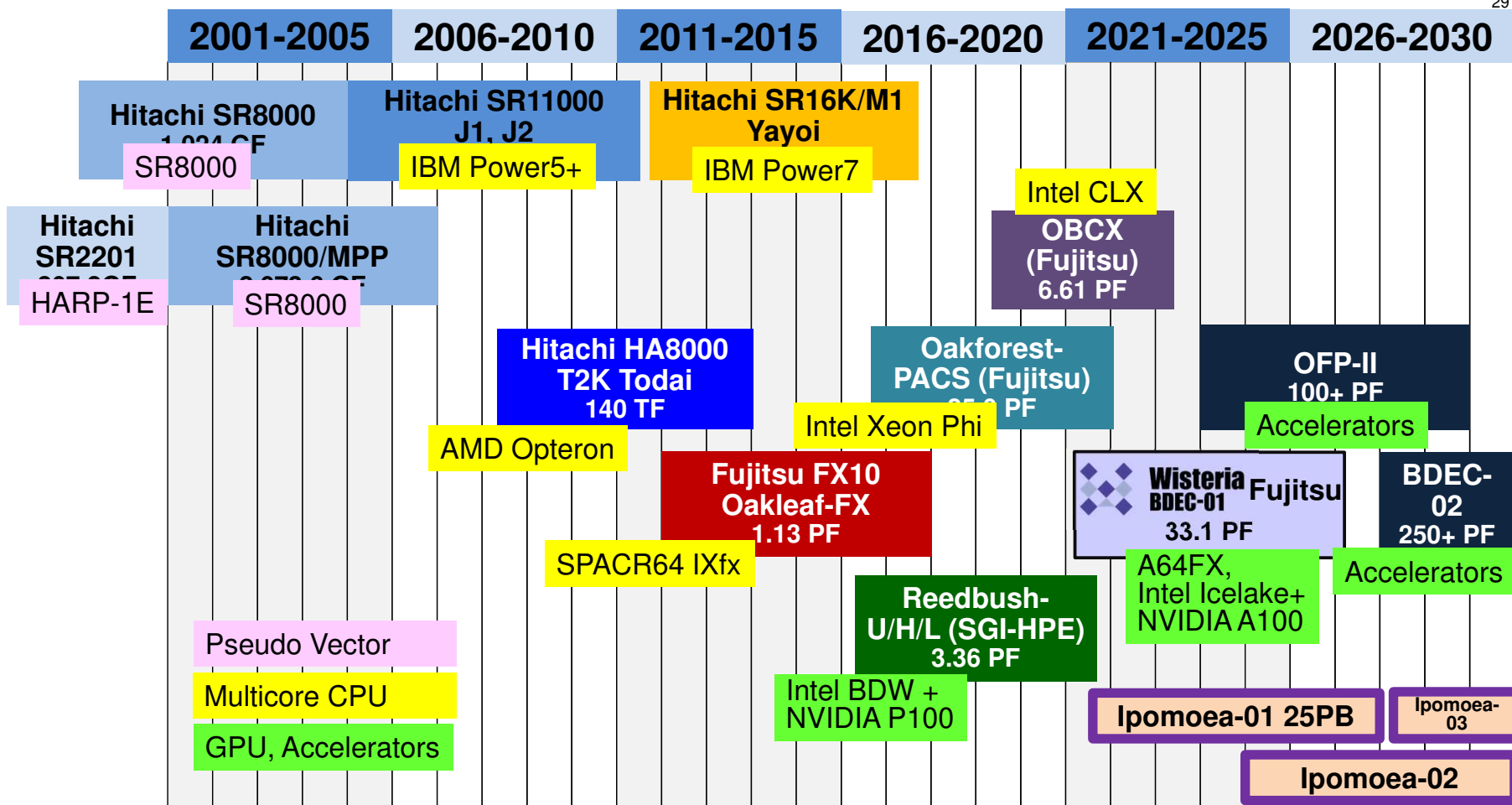
Ipomoea-
03

Ipomoea-02

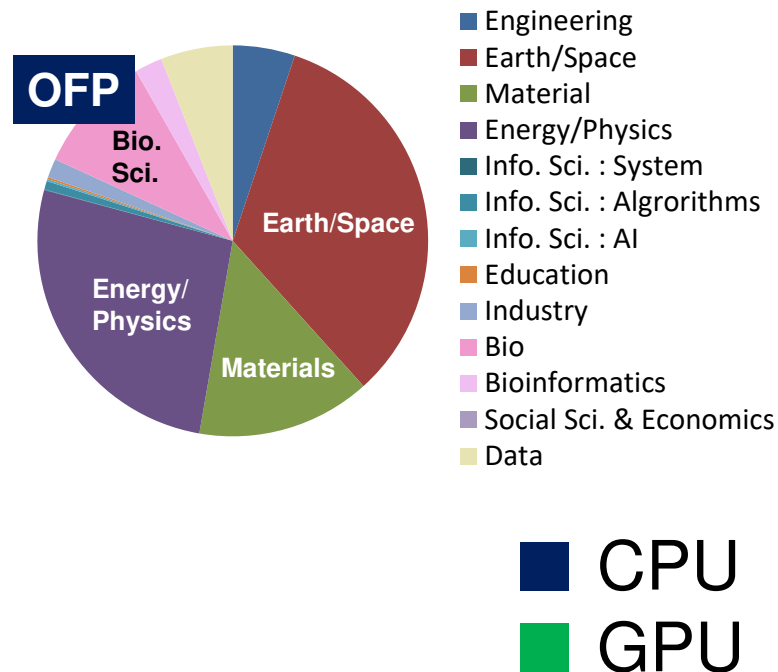
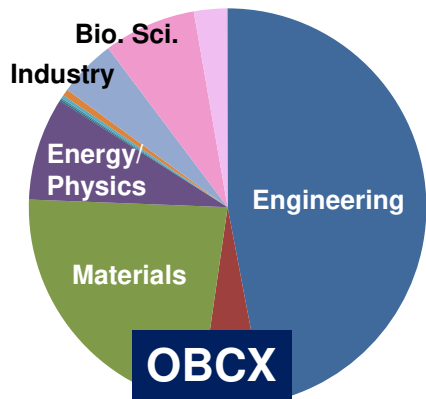
Supercomputers @ITC/U.Tokyo

2,600+ Users

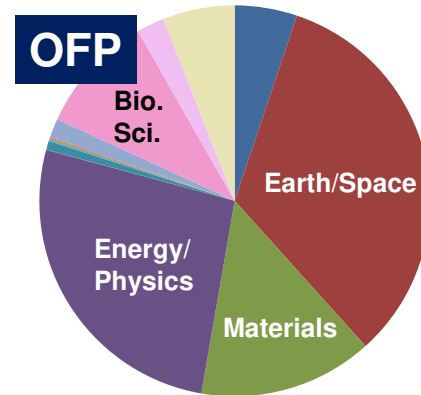
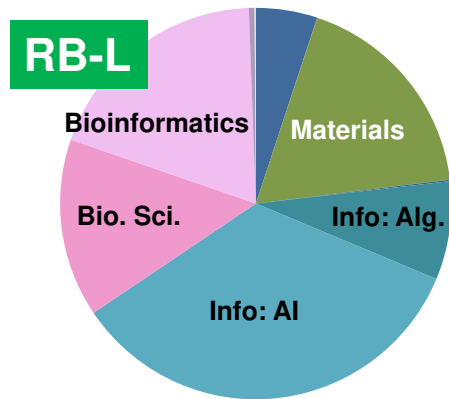
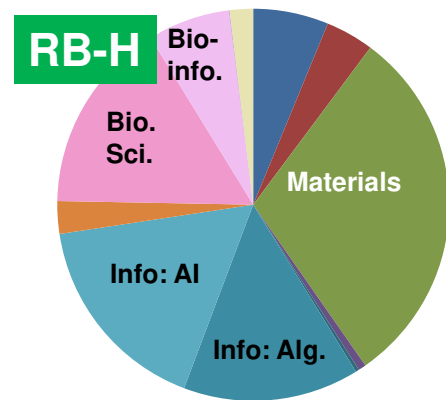
55+% outside of U.Tokyo



Research Area based on CPU Hours (FY.2020)



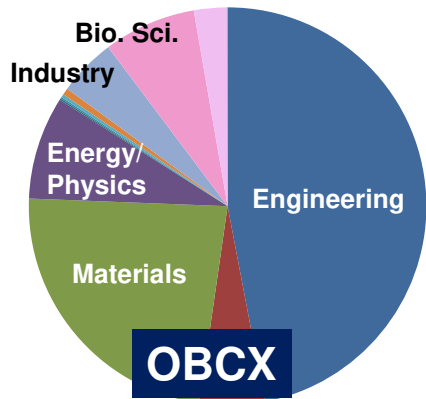
Research Area based on CPU Hours (FY.2020)



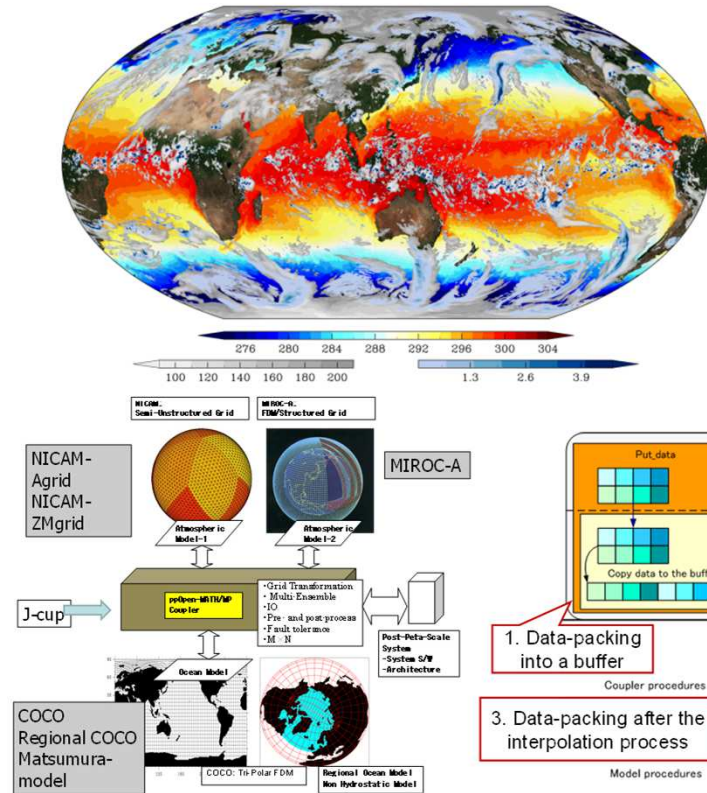
- Engineering
- Earth/Space
- Material
- Energy/Physics
- Info. Sci. : System
- Info. Sci. : Algorithms
- Info. Sci. : AI
- Education
- Industry
- Bio
- Bioinformatics
- Social Sci. & Economics
- Data

■ CPU

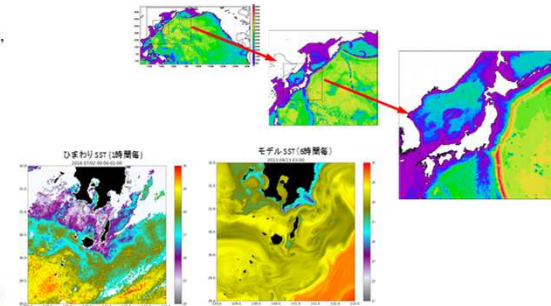
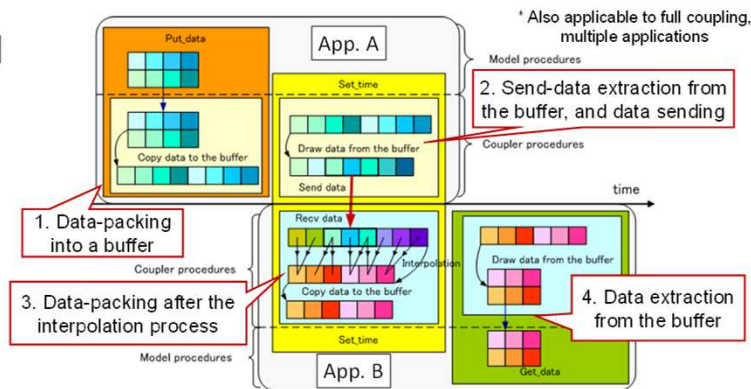
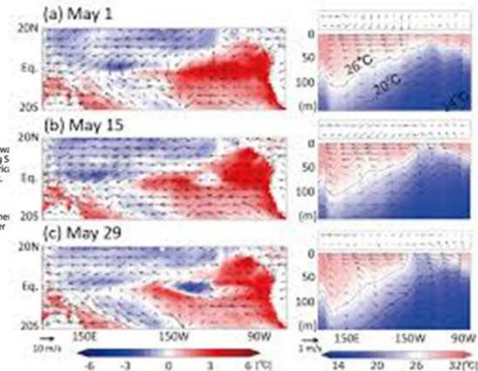
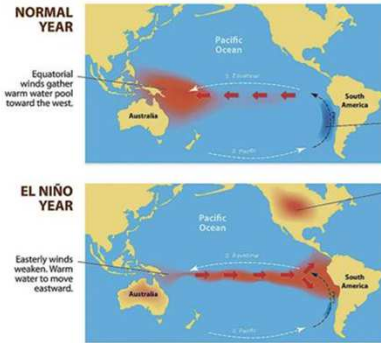
■ GPU



Global Atmosphere-Ocean Coupled Simulations



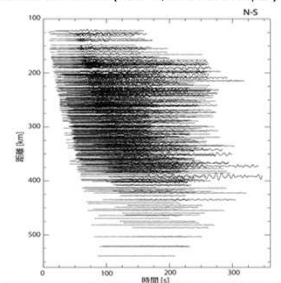
THE EL NIÑO PHENOMENON



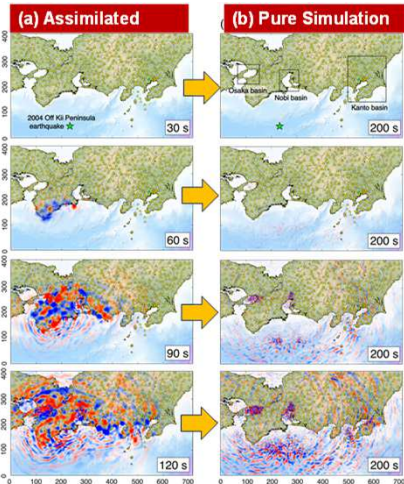
[c/o Prof. M. Sato, Prof. H. Hasumi (AORI/U.Tokyo)]

Solid Earth & Earthquake Simulations

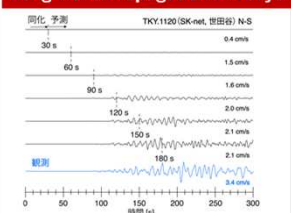
○ Observation (K-NET, KIK-net 446 pts)



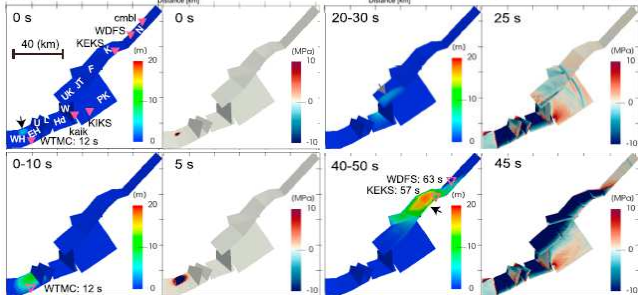
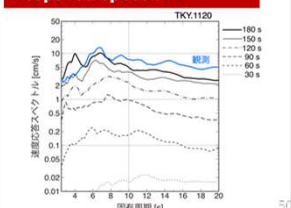
Assimilation at 90 sec. → Pure Simulation



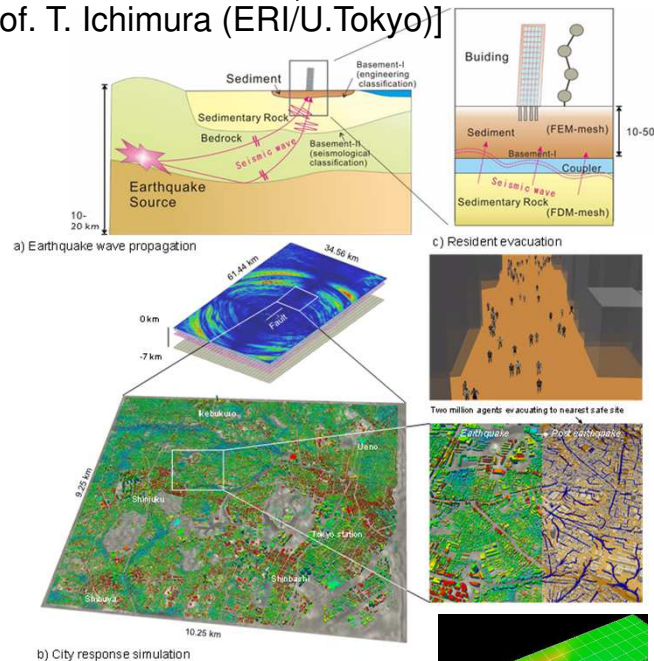
Long Wave Propagation in Tokyo



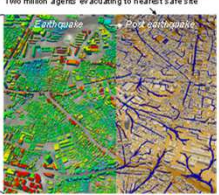
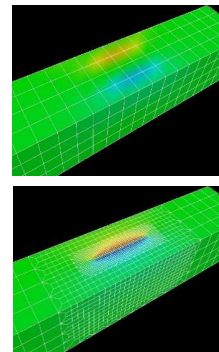
Response Spectrum



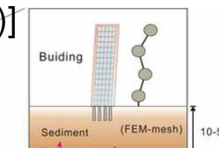
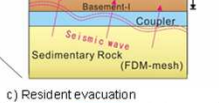
[c/o Prof. T. Furumura,
Prof. T. Ichimura (ERI/U.Tokyo)]



b) City response simulation



Two million agents evacuating to nearest safe site



[c/o Prof. R.
Ando (U.Tokyo)]

Simulation of Geologic CO₂ Storage

[c/o Dr. Hajime Yamamoto
(Taisei Corporation)]

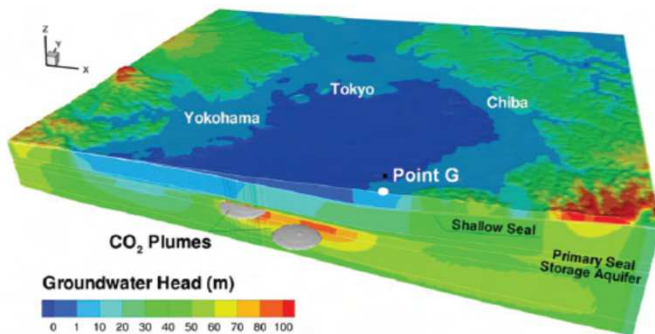
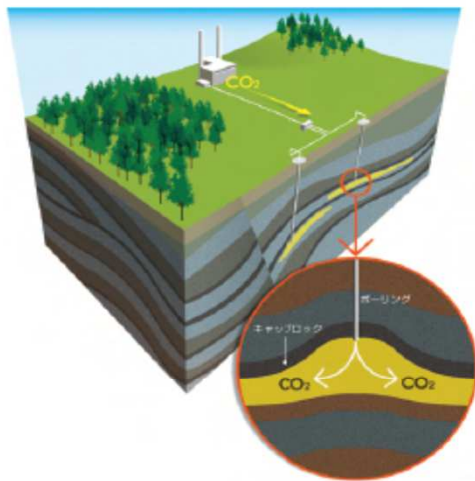
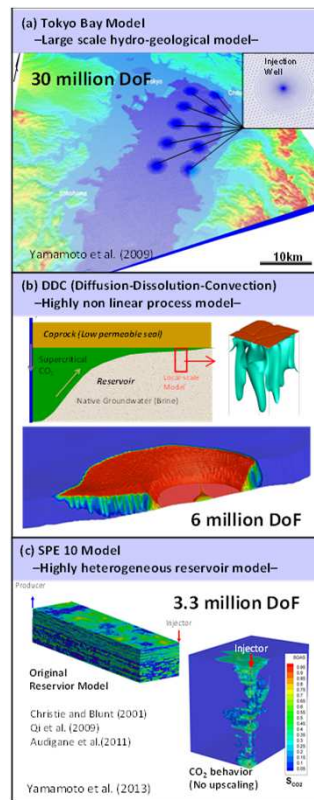
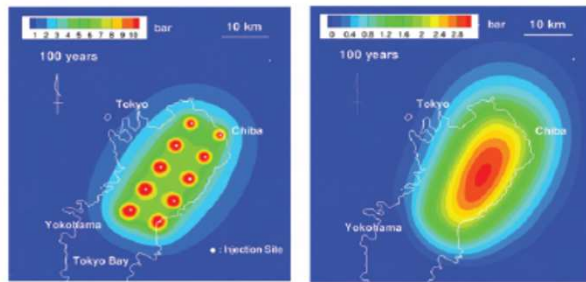


図-4 CO₂ 圧入後の地下水圧（全水頭換算）の分布（100 年後）



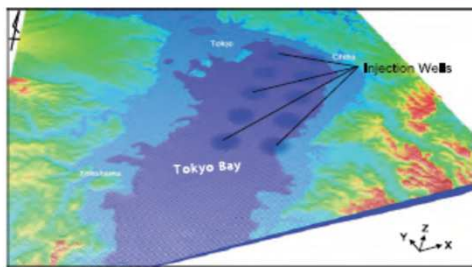
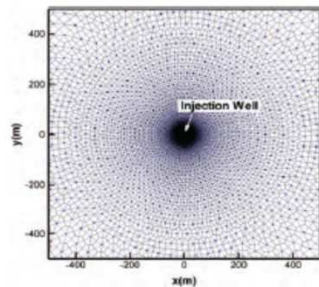
※DOF: degrees of freedom



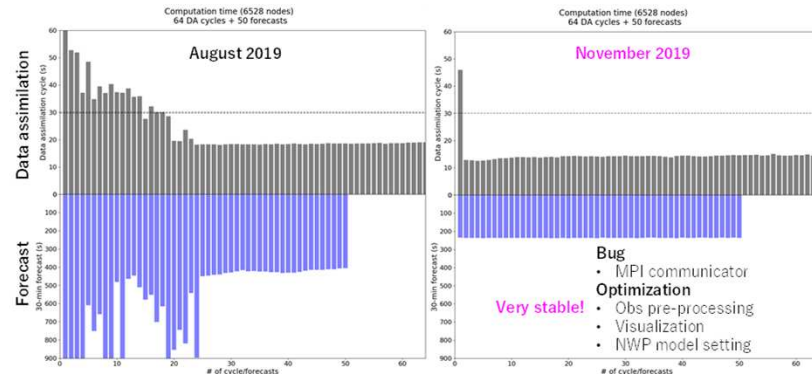
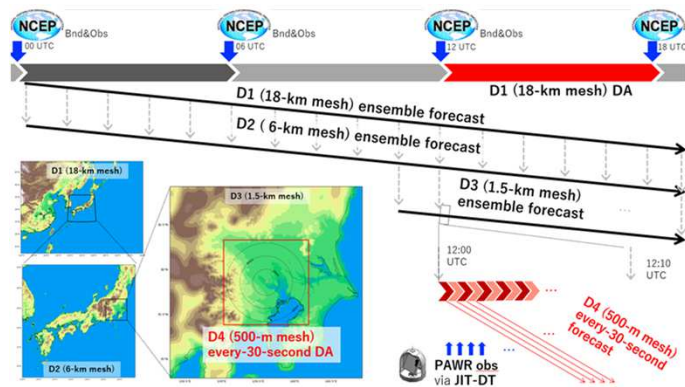
(a) 深部遮蔽層下面

(b) 浅部遮蔽層下面

図-5 圧力上昇量の平面分布（初期状態からの増分、圧入開始から 100 年後）

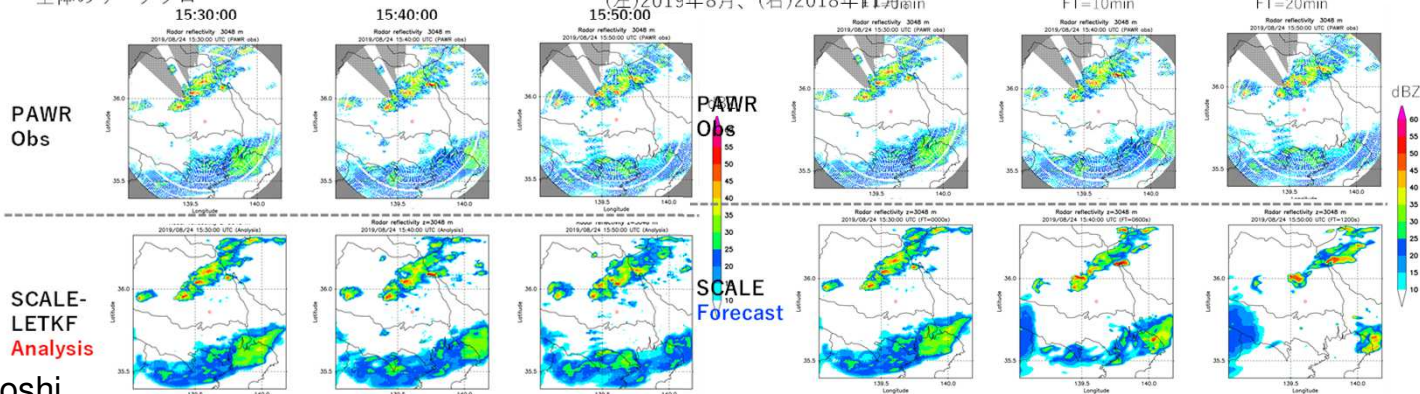


Real-Time Prediction of Severe Rainstorm by OFP



計算性能の向上。上段はデータ同化、下段は30分予報にかかった時間(秒)。
(左)2019年8月、(右)2018年11月

全体のワークフロー



[c/o Dr. Takemasa Miyoshi
(RIKEN R-CCS)]

2019年8月24日の事例についてのテスト結果。(上)レーダー観測と(下)SCALE-LETKFによる解析で得られたレーダー反射強度(dBZ)を示す。

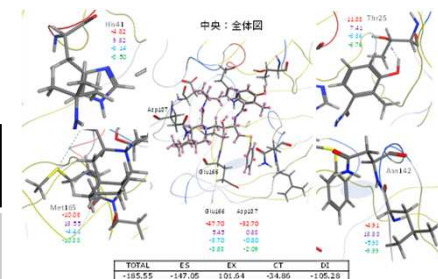
2019年8月24日の事例についてのテスト結果。(上)レーダー観測と(下)SCALE-LETKFによる予報で得られたレーダー反射強度(dBZ)を示す。

HPCI Urgent Call for Fighting against COVID-19 in Japan (FY.2020)

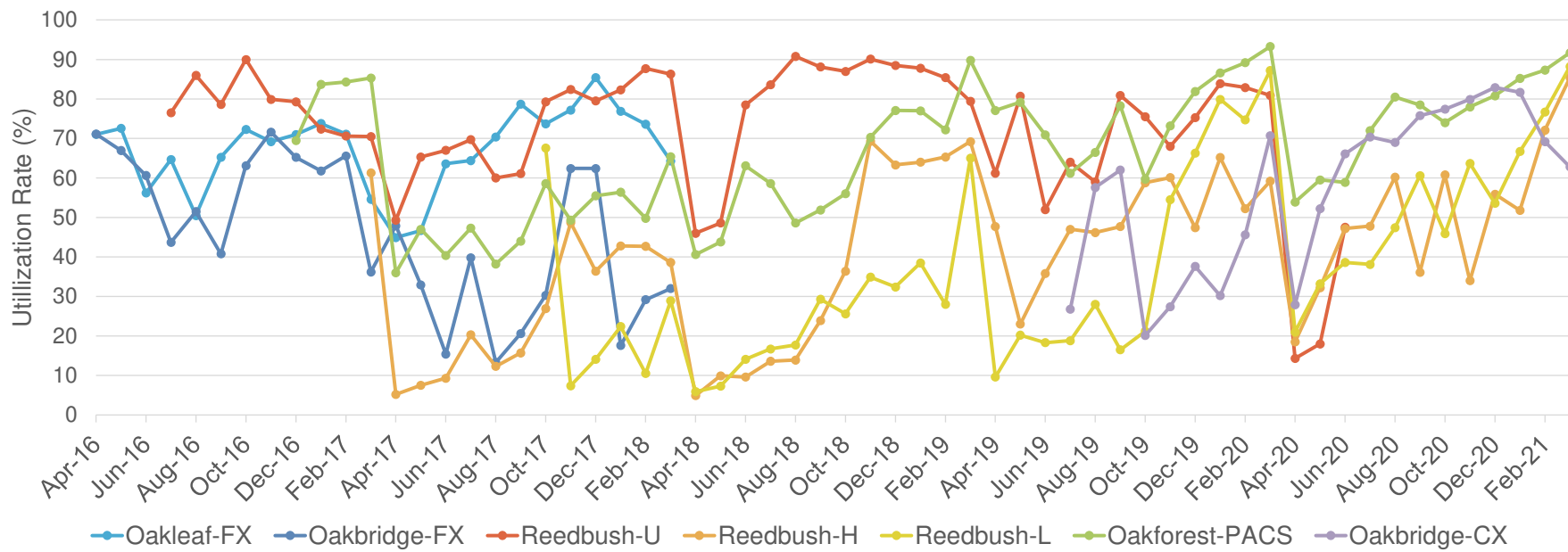
by 8 SC Centers of Natl. Univ., AIST etc.

6 of 14 accepted projects use U.Tokyo's Systems

Project Name	PI	System
Fragment molecular orbital calculations on the main protease of COVID-19	Yuji Mochizuki (Rikkyo U.)	OFP
Study on the evaluation of arrhythmogenic risk of COVID-19 candidate drugs	Toshiaki Hisada (UT Heart)	
Prediction of dynamical structure of Spike protein of SARS-COV19	Yuji Sugita (RIKEN)	
Computer-assisted search for inhibitory agents for SARS-CoV-2	Tyuji Hoshino (Chiba U.)	OBCX
Prediction and Countermeasure for virus droplet Infection under Indoor Environment: Case studies for massively-parallel simulation on Fugaku	Makoto Tsubokura (Kobe U.)	
Spreading of polydisperse droplets in a turbulent puff of saturated exhaled air	Marco Edoardo Rosti (OIST)	



Usage Ratio for Supercomputer Systems



Ipomoea: Common Shared Storage

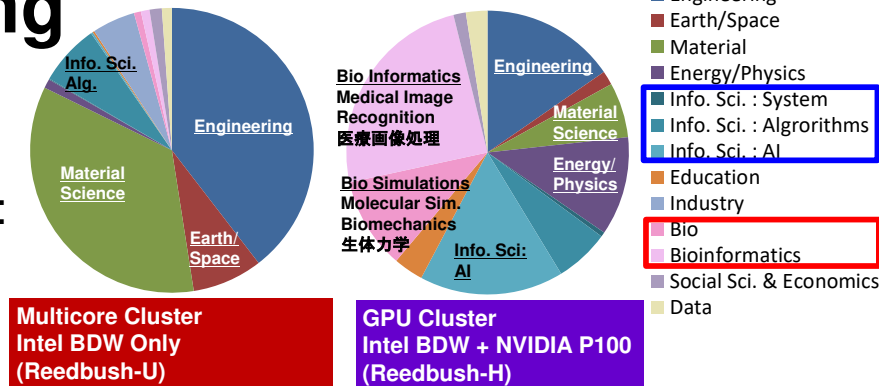
- Each Supercomputer has its own File System: Independent
 - Number and size of files are increasing -> Inconvenient
 - Shutdown of OFP with 25+PB Storage in March 2022
 - We decided to introduce a “Common Shared Storage”
- Available from Various Systems in ITC/U.Tokyo and JCAHPC
 - OFP, OBCX, mdx, Wisteria/BEDC-01
 - Future Systems
- Operation Period: 5-6 Years, 25+PB
 - 2 File Systems, One of these is replaced every 3-year
- Ipomoea-01
 - Operation started in January 2022, Public use from June 2022
 - Supporting smooth transition from OFP to OFP-II (Successor)
- Ipomoea-02 (April 2025)



- Supercomputing Research Division, Information Technology Center ,The University of Tokyo (SCD/ITC/U.Tokyo)
- JHPCN, HPCI, JCAHPC and NHR
- Supercomputers in SCD/ITC/U.Tokyo
- **Integration of (Simulation+Data+Learning)**
 - **Wisteria/BDEC-01**
 - **h3-Open-BDEC**
- Future Perspective

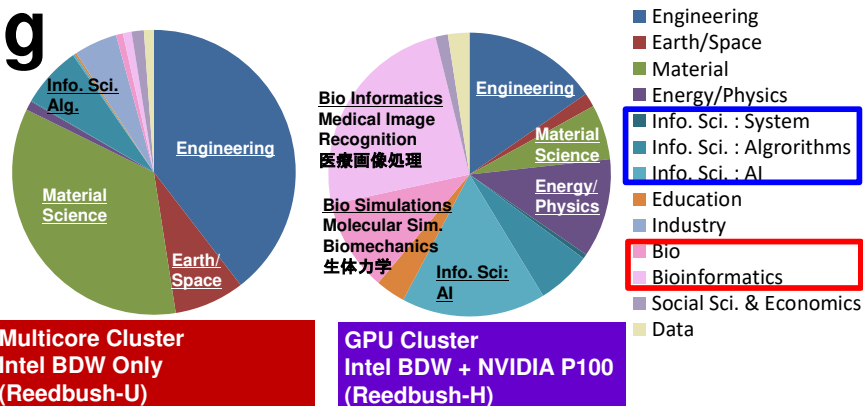
Future of Supercomputing

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



Future of Supercomputing

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

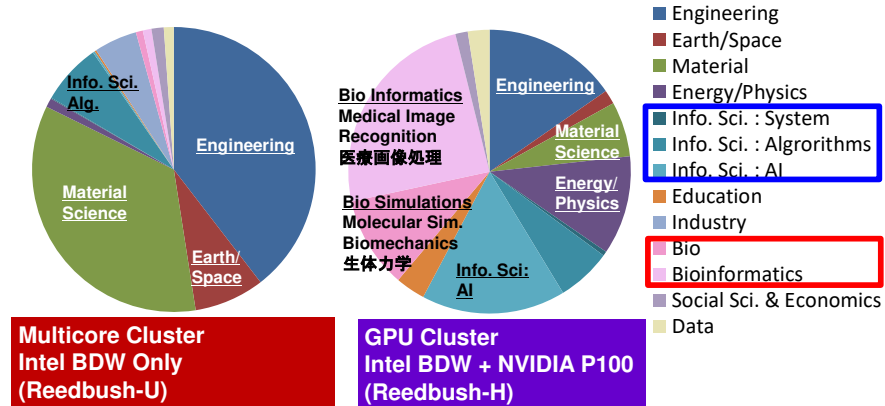


- Integration/Convergence of (Simulation + Data + Learning) (S+D+L) is important towards Society 5.0
 - Super Smart & Human-centered Society by Digital Innovation (IoT, Big Data, AI etc.) and by Integration of Cyber Space & Physical Space



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**
- **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



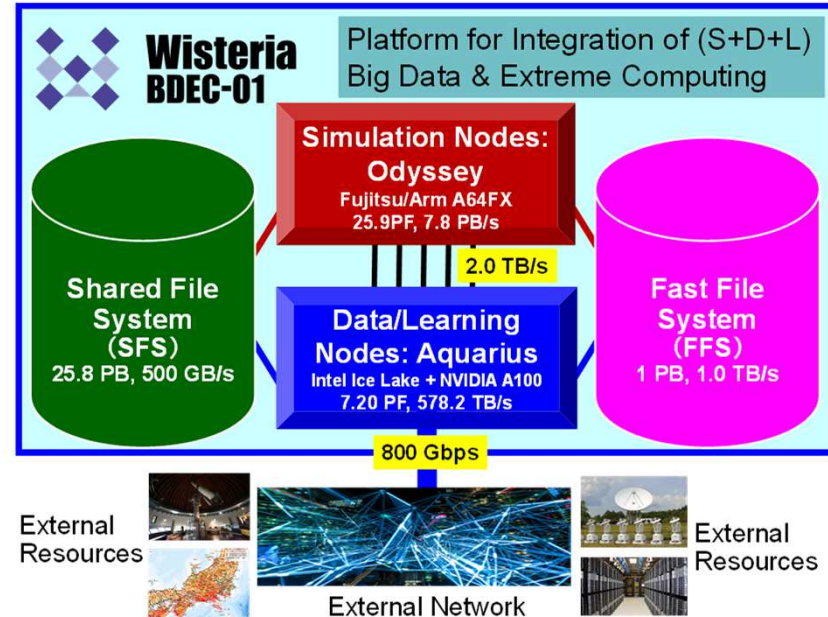
BDEC (Big Data & Extreme Computing)

S + D + L

Wisteria/BDEC-01

- 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)

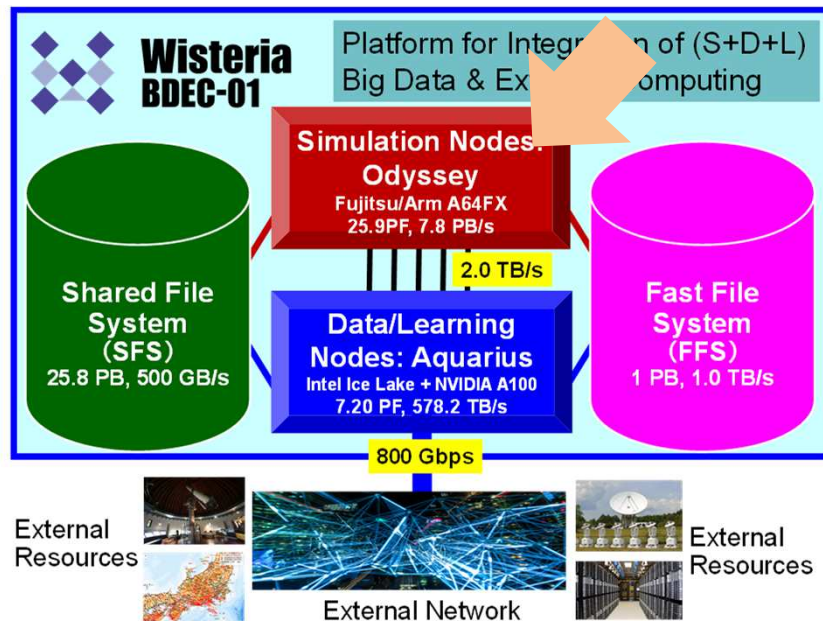
The 1st BDEC System (Big Data & Extreme Computing) Platform for Integration of (S+D+L)



Wisteria/BDEC-01

- 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)

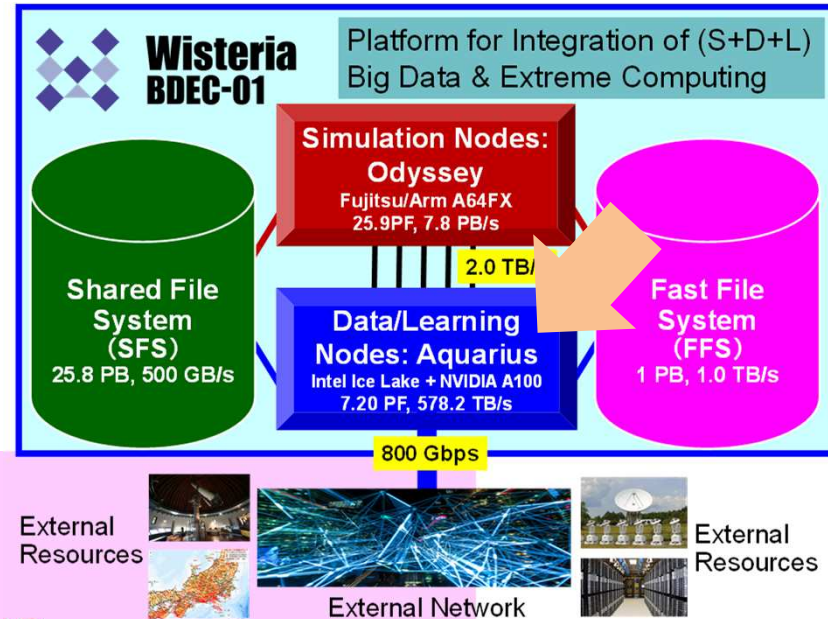
The 1st BDEC System (Big Data & Extreme Computing) Platform for Integration of (S+D+L)



Wisteria/BDEC-01

- 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)

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



Rankings@ISC 2022

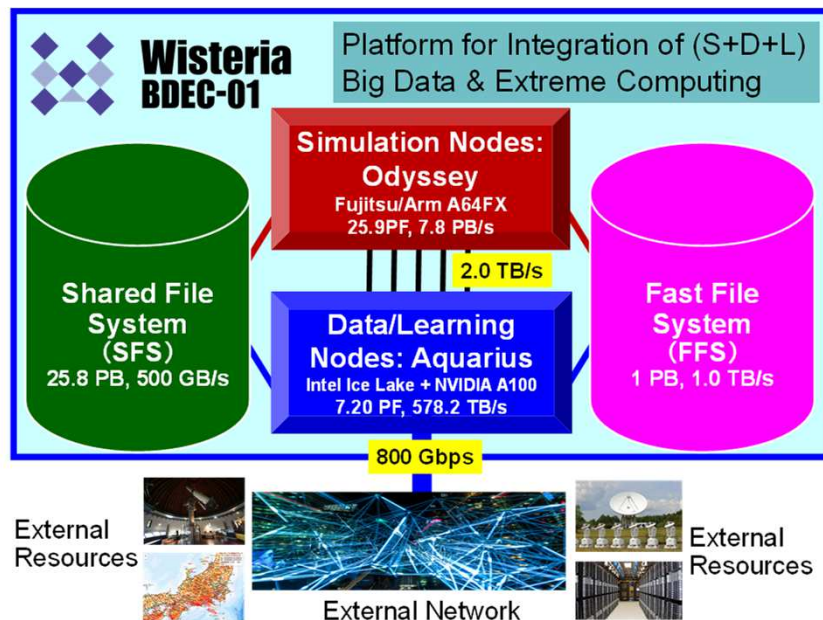
June 2022



ISC HIGH
PERFORMANCE
2021 DIGITAL

JUNE 24 - JULY 2, 2021
ISC-HPC.COM

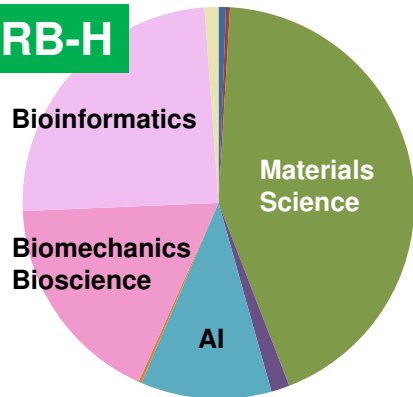
	Odyssey	Aquarius
TOP 500	20	115
Green 500	34	21
HPCG	10	62
Graph 500 BFS	3	-
HPL-AI	10	-



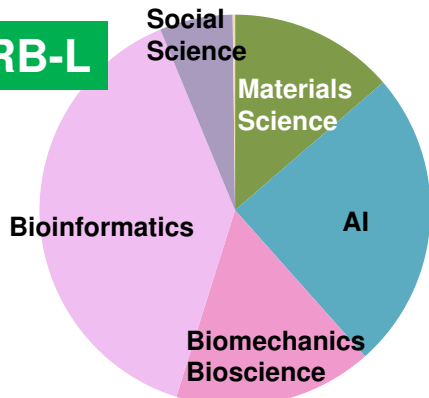
Research Area based on CPU-H's (FY.2021)

Odyssey, Aquarius: After Aug., RB-H, RB-L: Nov.E

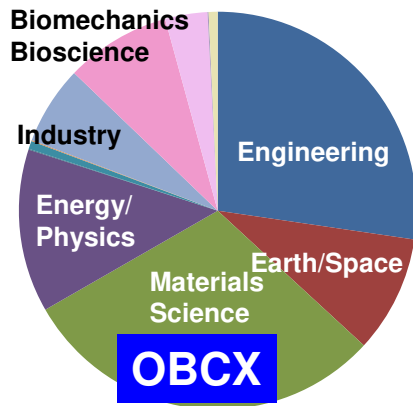
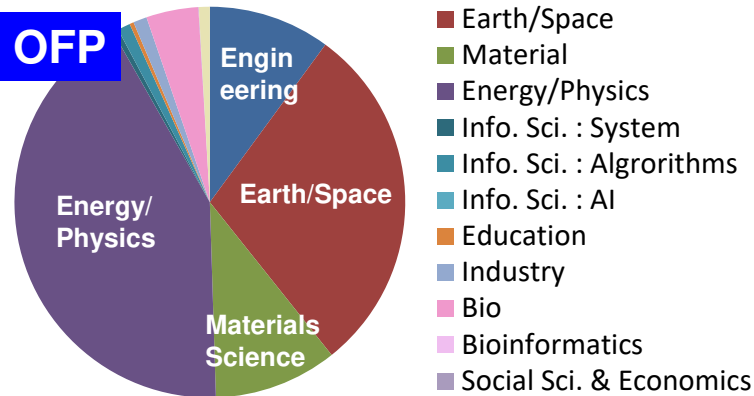
RB-H



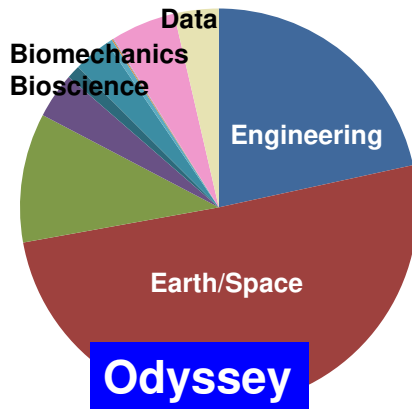
RB-L



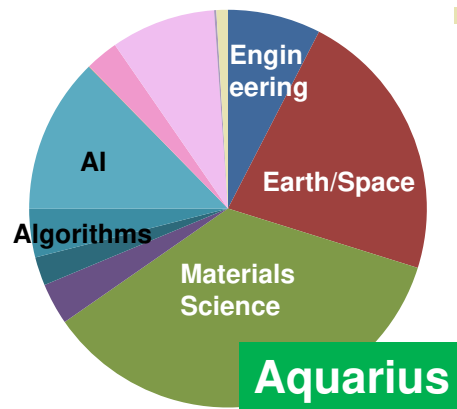
OFP



Odyssey



Aquarius



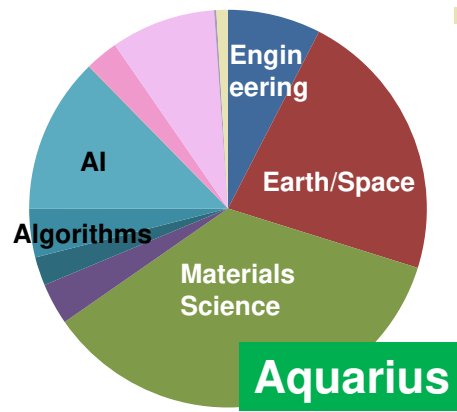
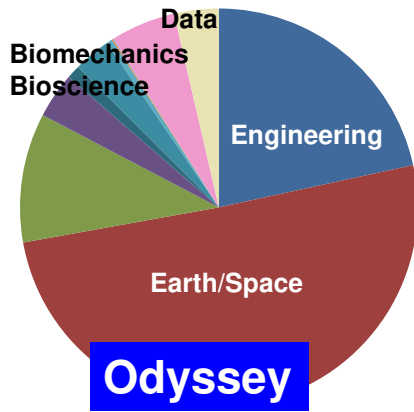
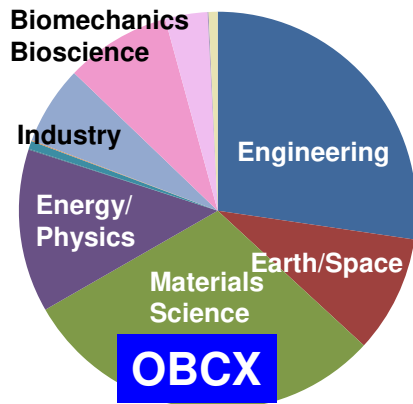
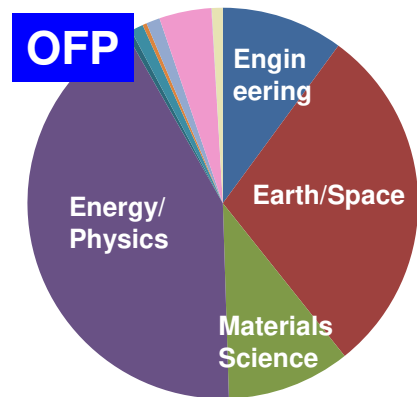
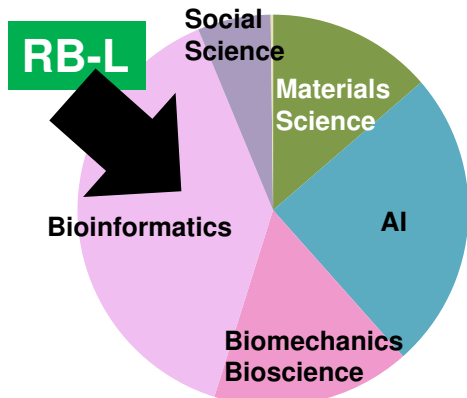
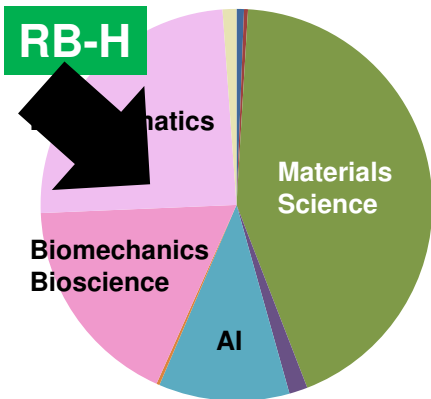
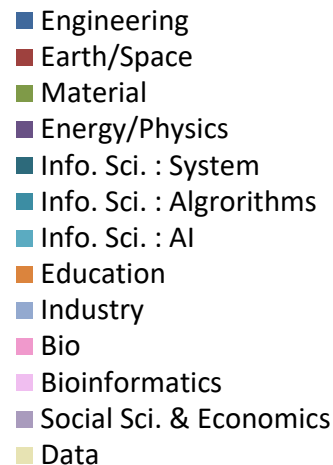
- Engineering
- Earth/Space
- Material
- Energy/Physics
- Info. Sci. : System
- Info. Sci. : Algorithms
- Info. Sci. : AI
- Education
- Industry
- Bio
- Bioinformatics
- Social Sci. & Economics
- Data

CPU

GPU

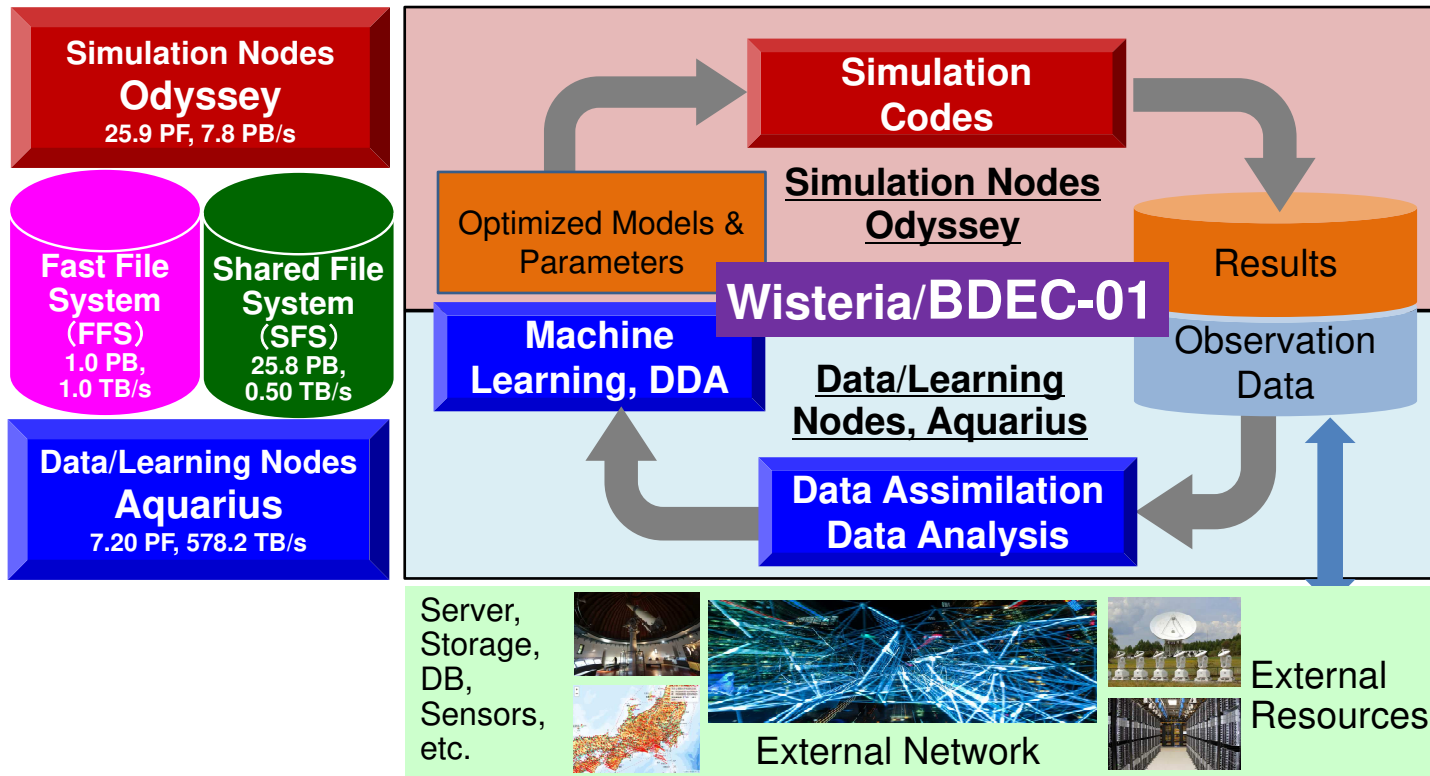
Research Area based on CPU-H's (FY.2021)

Odyssey, Aquarius: After Aug., RB-H, RB-L: Nov.E



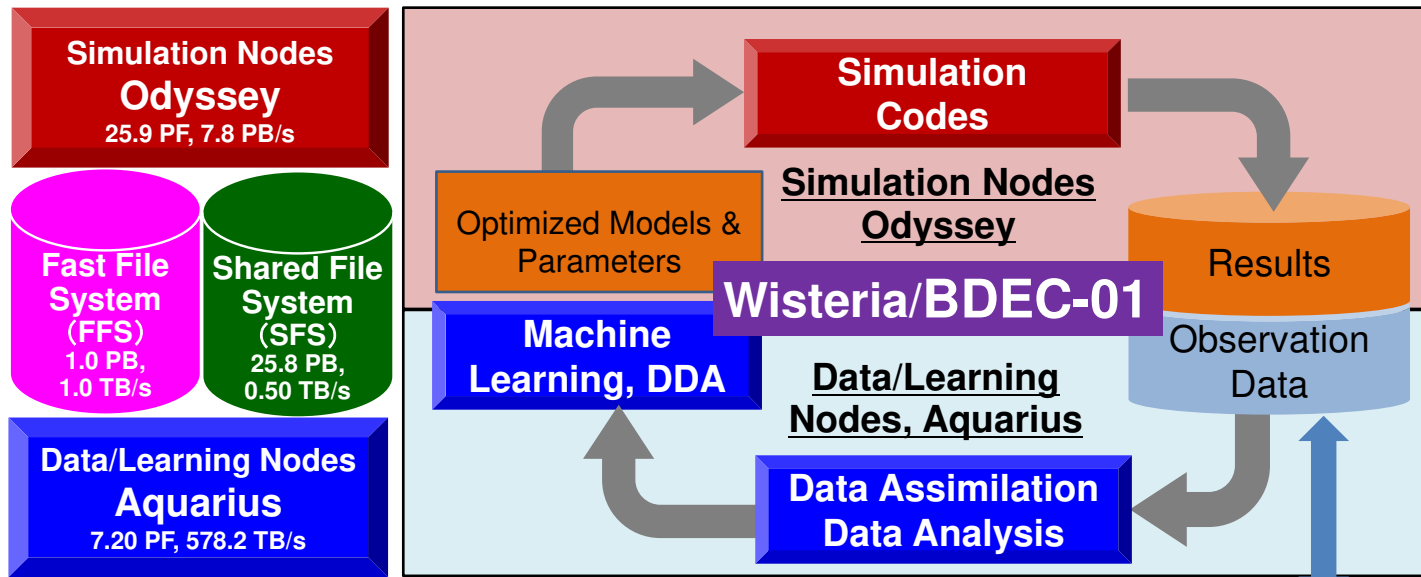
Wisteria/BDEC-01

Platform for Integration of (Simulation+Data+Learning) (S+D+L)



Wisteria/BDEC-01

Platform for Integration of (Simulation+Data+Learning) (S+D+L)

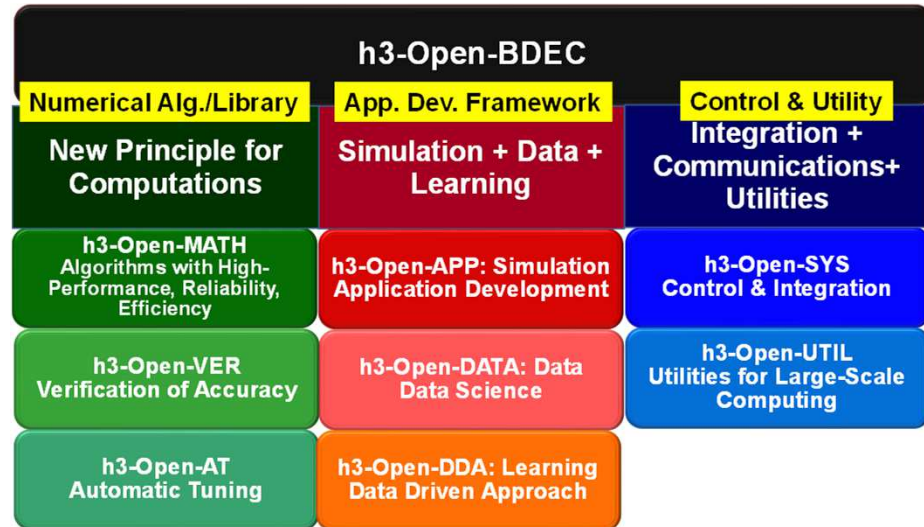


Optimization of Models/Parameters for Simulations by Data Analytics & Machine Learning (S+D+L)

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), Co-PI, Container



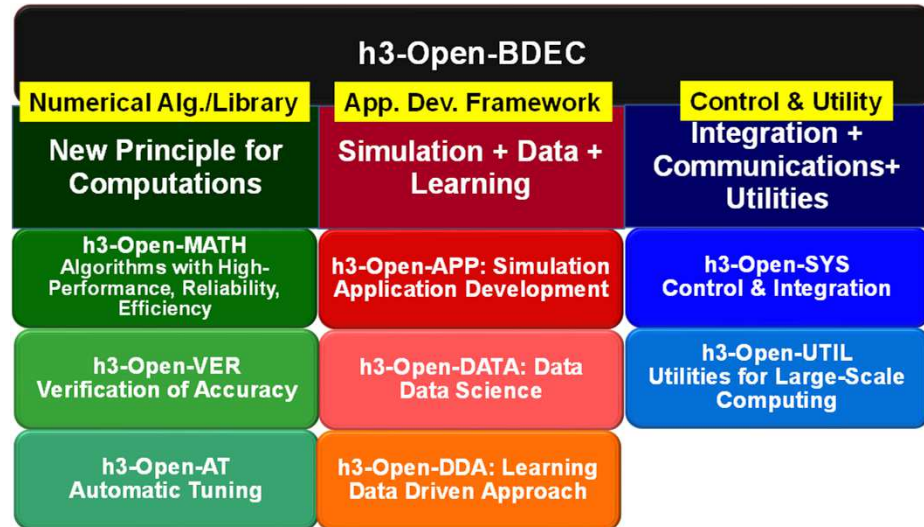
International Collaborators

- Osni Marques (LBNL, USA)
- Edmond Chow (Georgia Tech, USA)
- Richard Vuduc (Georgia Tech, USA)
- Gerhard Wellein (FAU Erlangen, Germany)
- Weichung Wang (National Taiwan University, Taiwan)
- Feng-Nan Hwang (National Central University, Taiwan)

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

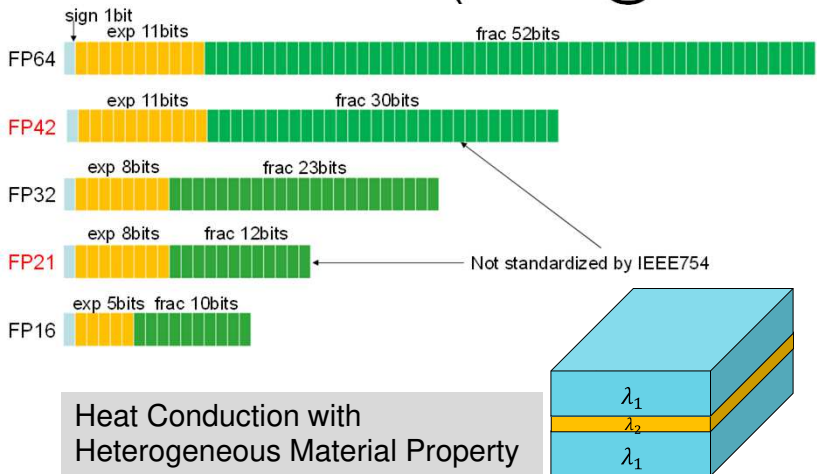


- “Three” Innovations
 - New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
 - Hierarchical Data Driven Approach (*hDDA*) based on Machine Learning
 - Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01

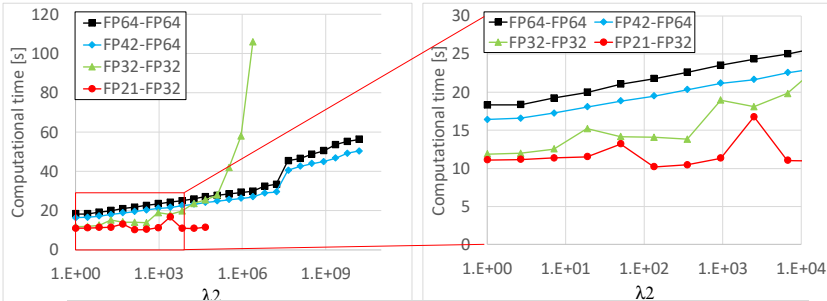


Adaptive Precision Computing with FP42/FP21

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



Heat Conduction with Heterogeneous Material Property



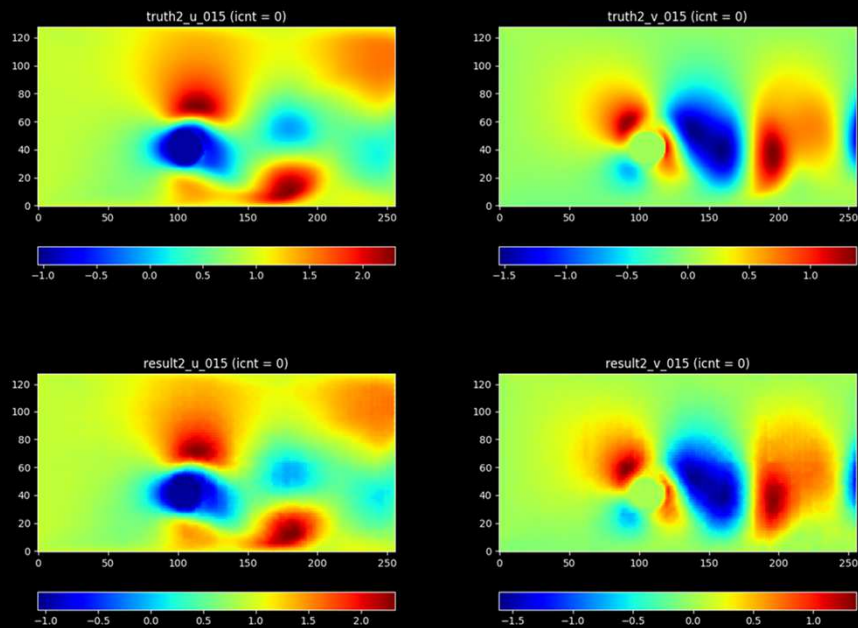
Computation Time for ICCG Solver
Various Types of Precisions on Intel Xeon Cascadelake

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 ($\lambda_1=10^0$, $\lambda_2=10^0 \sim 10^9$). 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”).

Prediction of CFD Simulation by Deep Learning

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



Computational fluid dynamics (CFD) is widely used in science and engineering. However, since CFD simulations require 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.

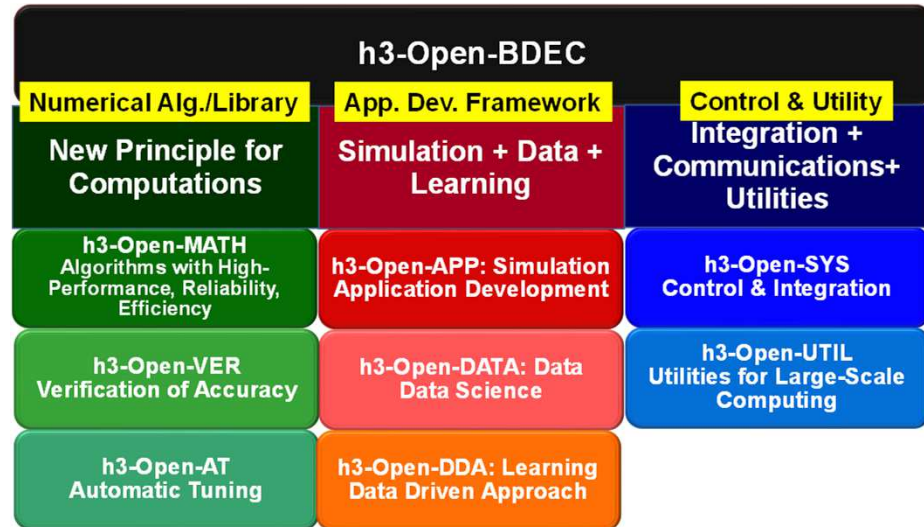
Comparison of the flow velocity results obtained by the conventional simulation (upper) and the prediction of these results by deep learning (lower)

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

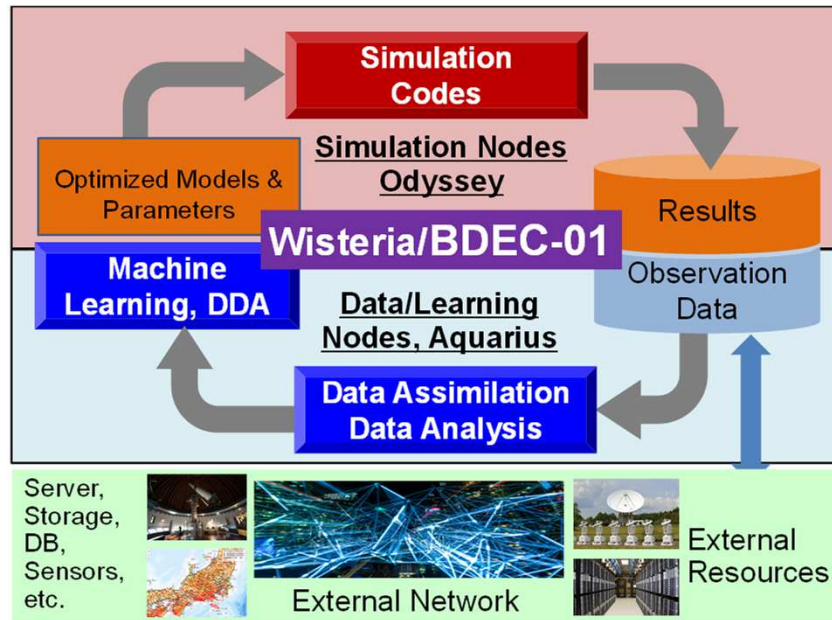
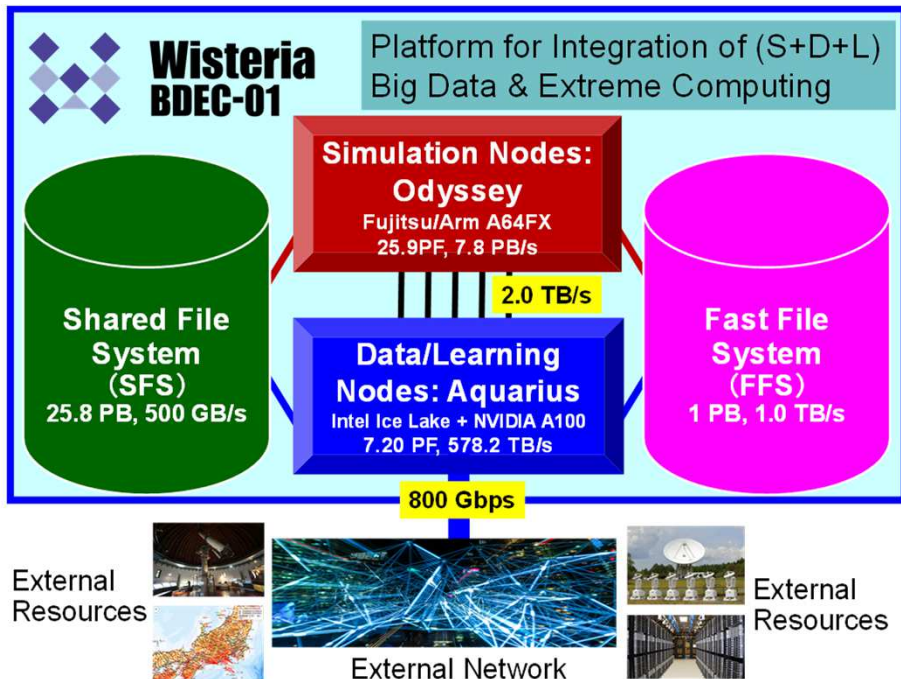


• “Three” Innovations

- New Principles for Numerical Analysis by Adaptive Precision, Automatic Tuning & Accuracy Verification
- Hierarchical Data Driven Approach (*hDDA*) based on Machine Learning
- Software & Utilities for Heterogenous Environment, such as Wisteria/BDEC-01

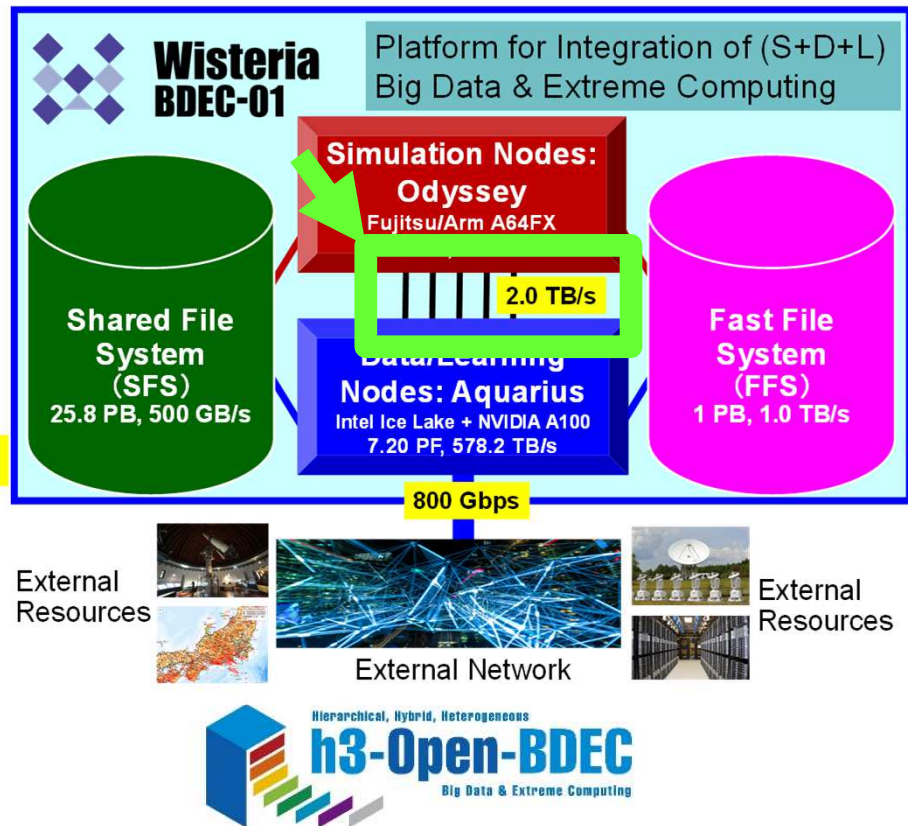


Wisteria/BDEC-01: The First “Really Heterogenous” System in the World



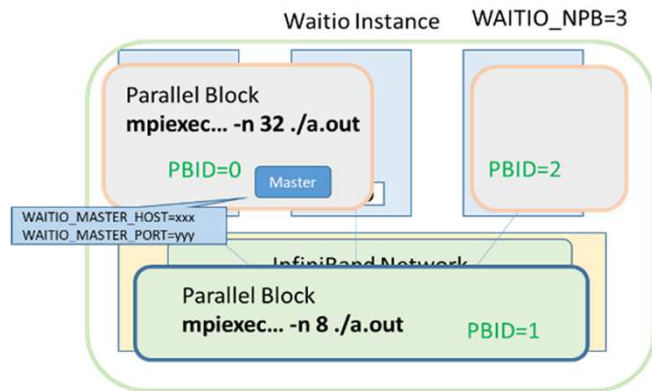
AI for HPC/Science on Wisteria/BDEC-01

- Wisteria/BDEC-01
 - Aquarius (GPU: NVIDIA A100)
 - Filtering, ML, Visualization
 - Odyssey (CPU: A64FX)
 - Data Assimilation, Simulation
- Combining Odyssey-Aquarius
 - Single MPI Job over O-A is impossible
 - Actually, O-A are connected through IB-EDR with 2TB/sec.
 - h3-Open-SYS/WaitIO-Socket
 - Library for Inter-Process Communication through IB-EDR with MPI-like interface
 - h3-Open-UTIL/MP
 - Multiphysics Coupler



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

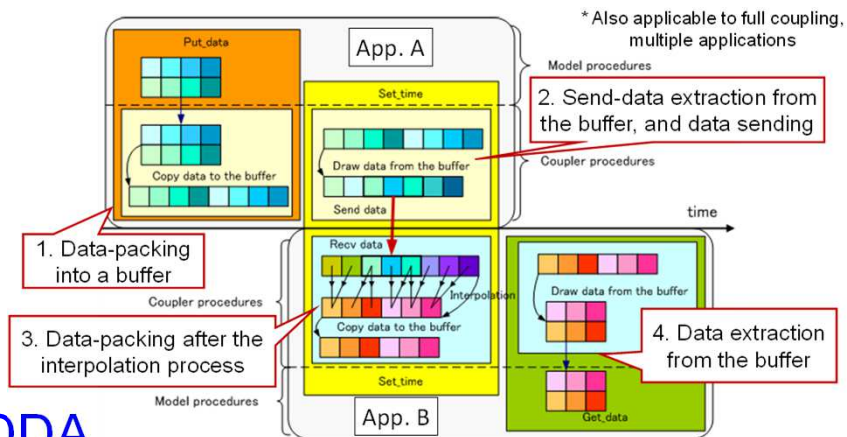


[Sumimoto et al. 2021]

h3-Open-UTIL/MP

Multilevel Coupler/Data Assimilation

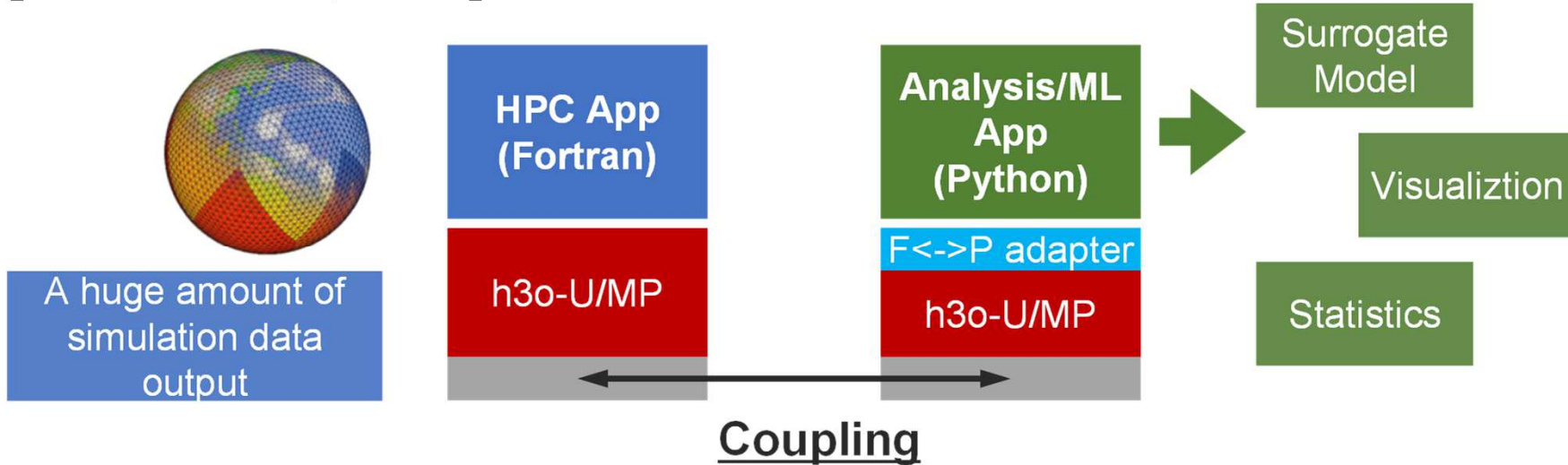
- Current Coupler: ppOpen-MATH/MP
 - Weak-Coupling of Multiple (usually two) Applications
 - Each application does a single computation
- h3-Open-UTIL/MP
 - Data Assimilation (Multiple Computations: Ensemble)
 - Assimilation of Computations with Different Resolutions
 - h3-Open-DATA, h3-Open-APP
 - Data Assimilation by Coupled Codes
 - e.g. Atmosphere-Ocean
- Data Assimilation: h3-Open-DATA
 - Karman Filter, Particle Karman Filter
 - LETKF
 - Adjoint Method
- Generation of Simplified Models in hDDA



h3-Open-UTIL/MP (h3o-U/MP)

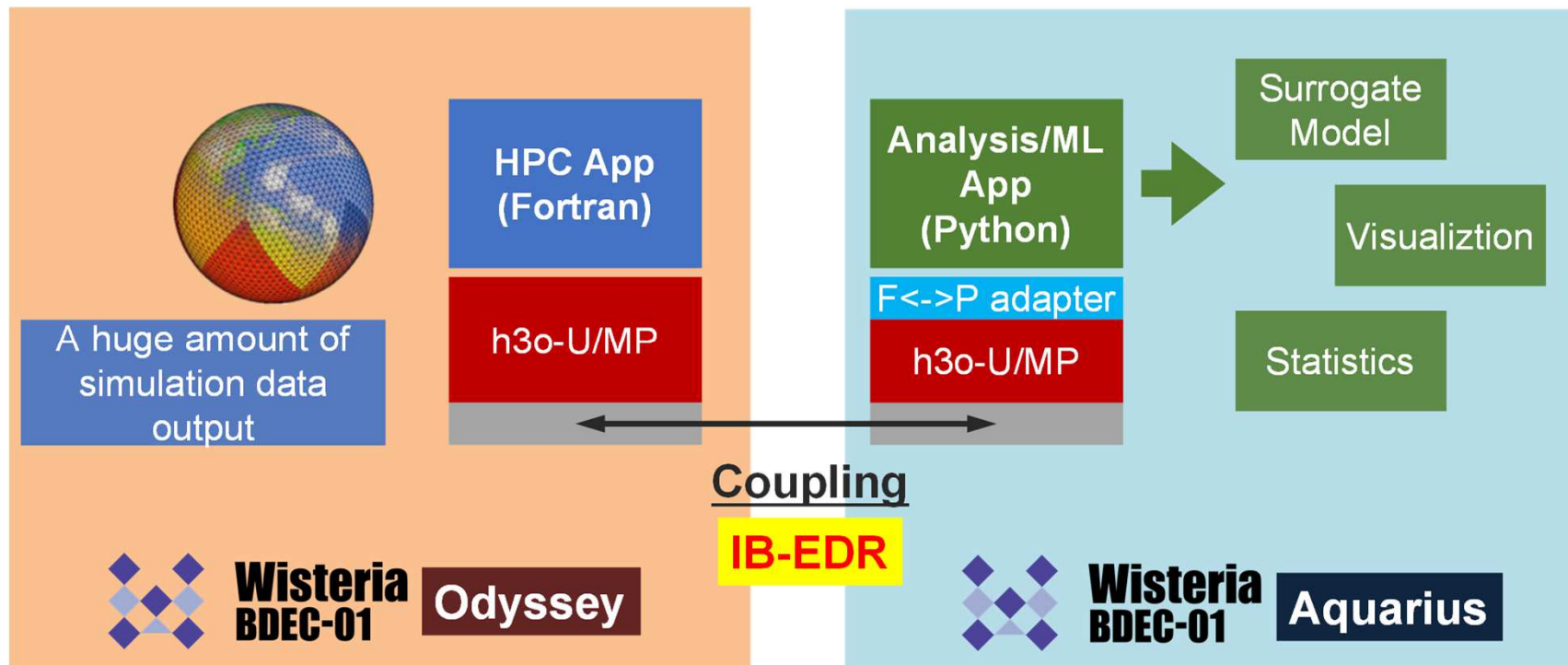
(HPC+AI) Coupling

[Dr. H. Yashiro, NIES]

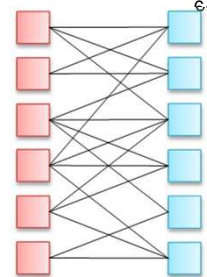
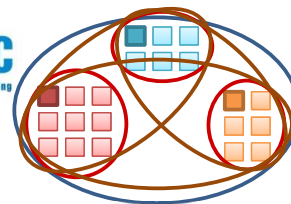


- Providing on-the-fly input/output/training data to the Analysis/ML tools
 - Easy to apply to existing HPC applications
 - Easy access to existing Python-based tools for AI/ML

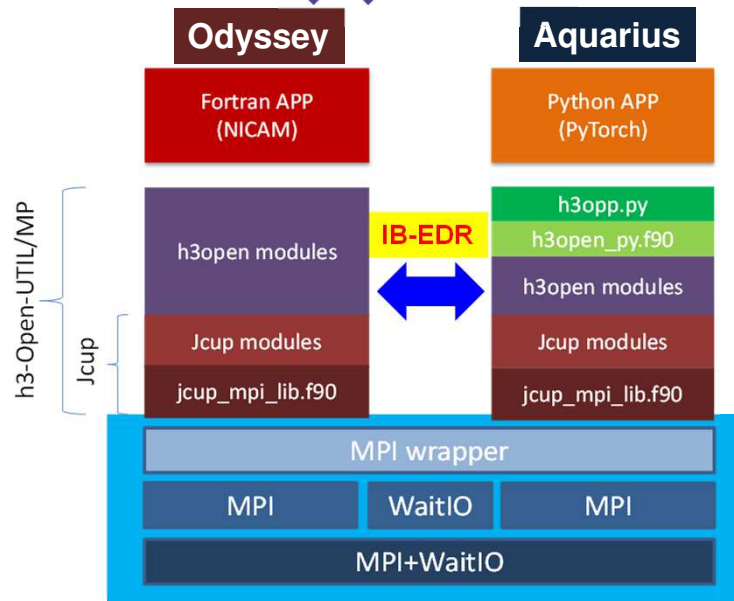
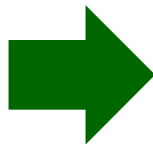
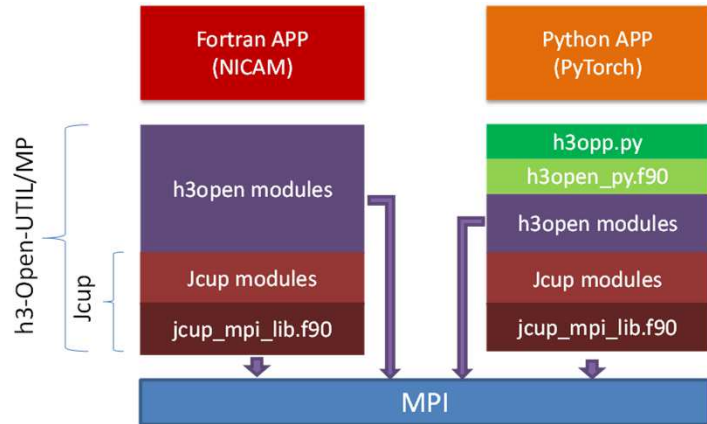
h3-Open-UTIL/MP (h3o-U/MP) + h3-Open-SYS/WaitIO-Socket



h3-Open-UTIL/MP + h3-Open-SYS/WaitIO-Socket available from June 2022

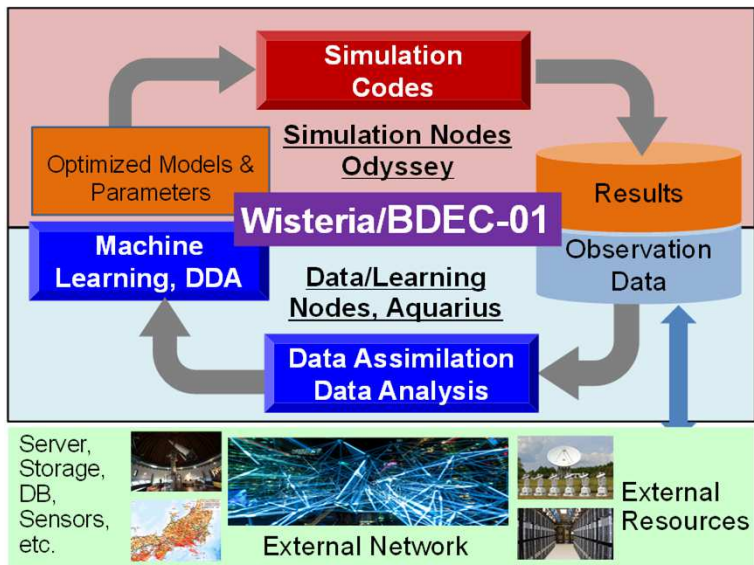


**Wisteria
BDEC-01**



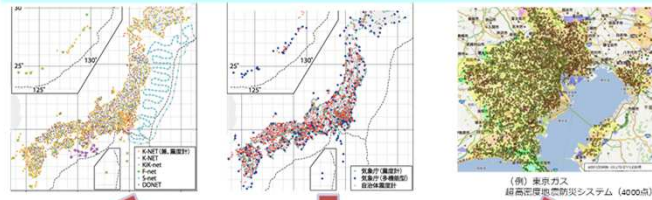
May 2021: Single MPI Environment

June 2022: Coupler + WaitIO

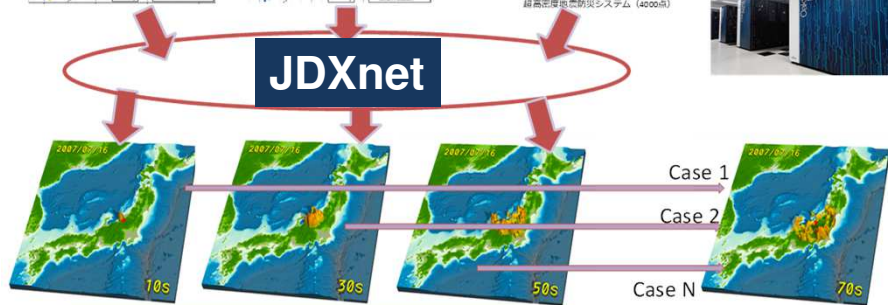
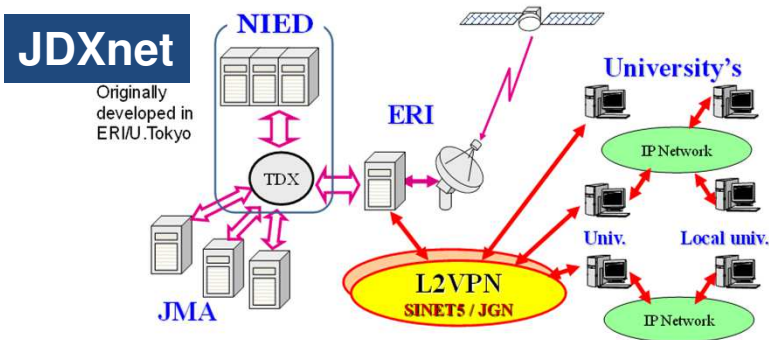


3D Earthquake Simulation with Real-Time Data Observation/Assimilation Simulation of Strong Motion (Wave Propagation) by 3D FDM

Observation Network for Earthquake: $O(10^5)$ Points



[c/o Furumura]



Real-Time Data/Simulation Assimilation
Real-Time Update of Underground Model

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

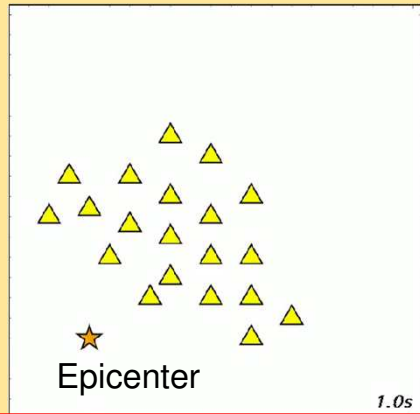
Real-Time Assimilation of “Observation+Computation” in Seismic Wave Propagation [c/o Oba & Furumura]

• Data Assimilation of Wave Propagation by “Optimal Interpolation Technique”

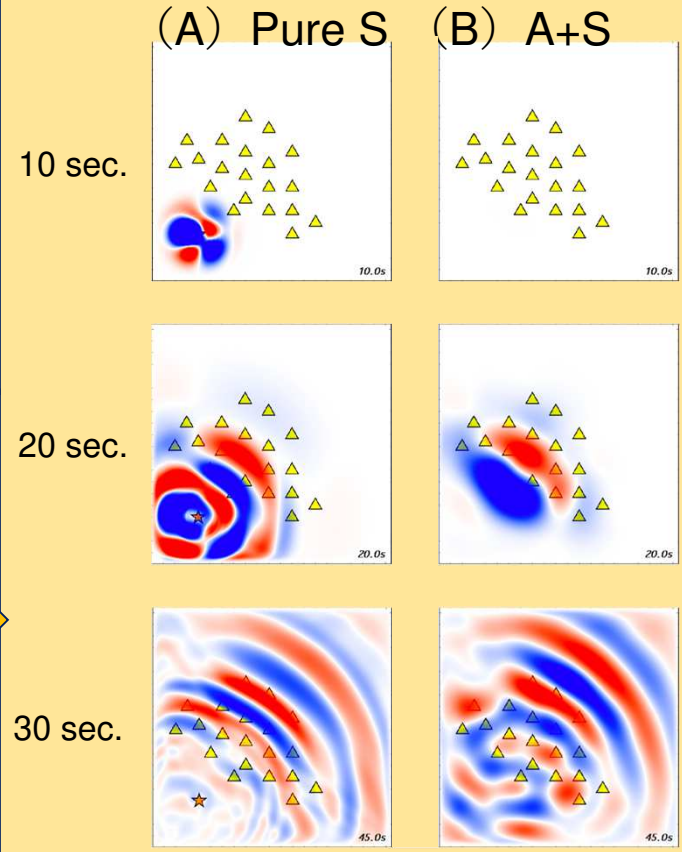
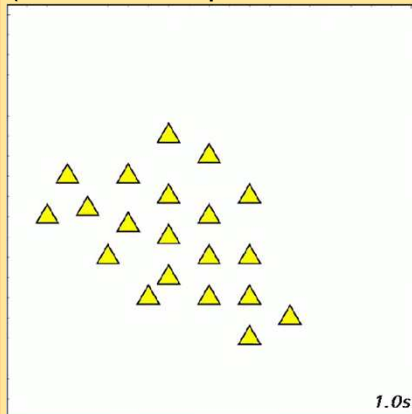
$$\begin{array}{c}
 \text{Assim.} \quad \text{Comp.} \quad \text{Residual} \quad \text{Comp.} \\
 x_n^a = x_n^f + W(y_n - Hx_n^f) \\
 \text{Comp.} \quad \text{Assim.} \\
 x_{n+1}^f = Fx_n^a
 \end{array}
 \quad
 \begin{array}{l}
 n: \text{Time Step} \\
 W: \text{Weighting Matrix} \\
 F: \text{Wave Propagation simulation}
 \end{array}$$

(A) Pure Simulation

▲ : Obs. Pts.



(B) Assimilation+Sim. (No info for Epicenter needed)



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

67

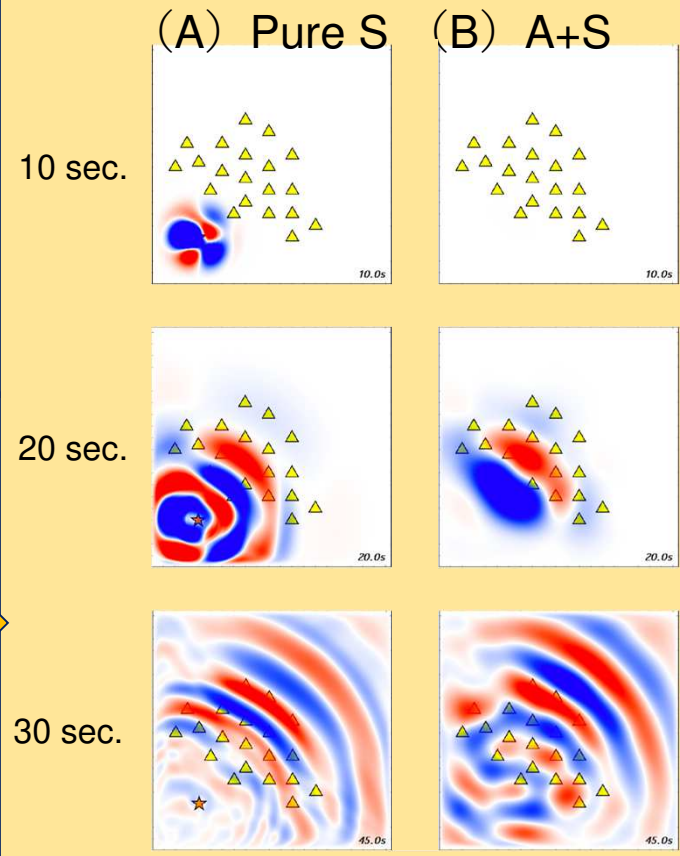
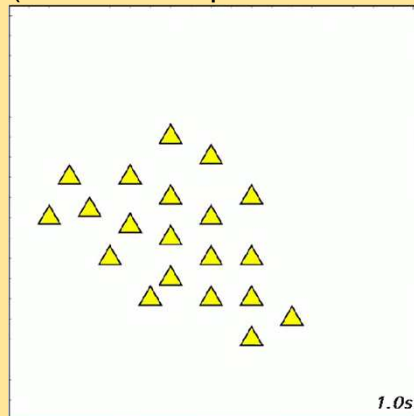
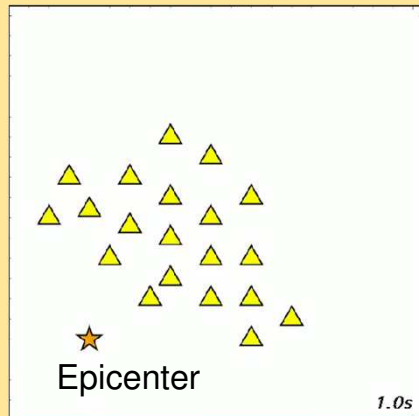
• Data Assimilation of Wave Propagation by “Optimal Interpolation Technique”

$$\begin{array}{c}
 \text{Assim.} \quad \text{Comp.} \quad \text{Residual} \quad \text{Comp.} \\
 x_n^a = x_n^f + W(y_n - Hx_n^f) \\
 \text{Comp.} \quad \text{Assim.} \\
 x_{n+1}^f = Fx_n^a
 \end{array}
 \quad
 \begin{array}{l}
 n: \text{Time Step} \\
 W: \text{Weighting Matrix} \\
 F: \text{Wave Propagation simulation}
 \end{array}$$

(A) Pure Simulation

(B) Assimilation+Sim.

▲ : Obs. Pts. (No info for Epicenter needed)



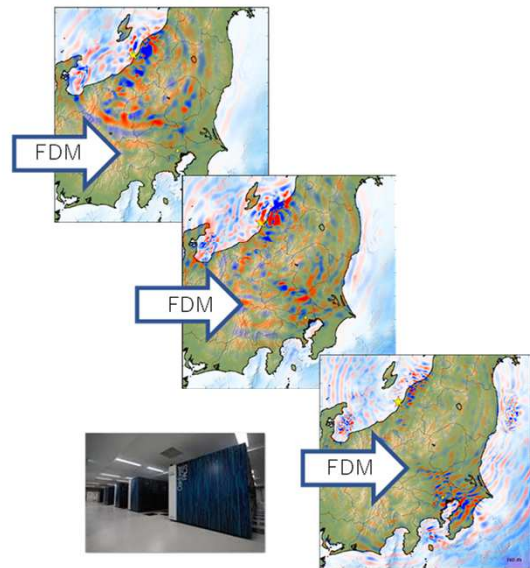
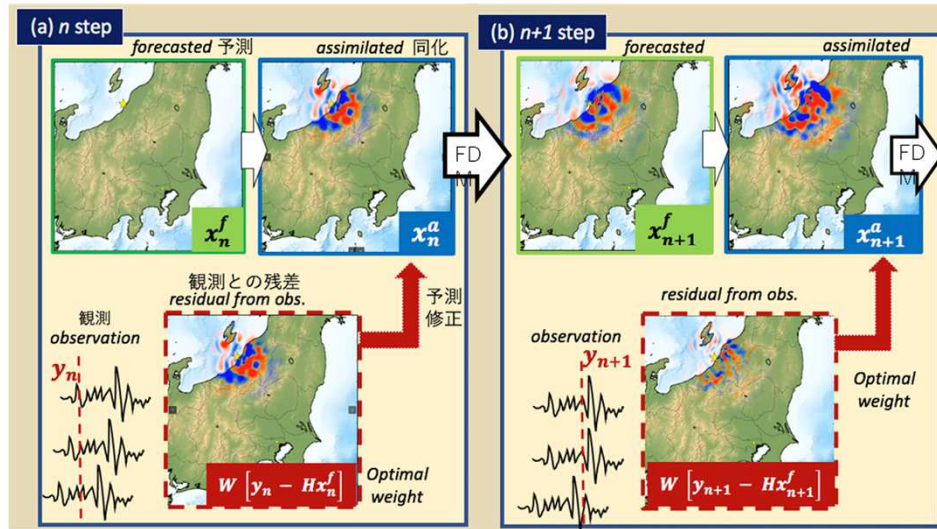
67

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

$$\begin{aligned}
 & \text{Assim. Comp.} \quad x_n^a = x_n^f + W(y_n - Hx_n^f) \\
 & \text{Comp. Assim.} \quad x_{n+1}^f = Fx_n^a \\
 & \text{Residual Comp.} \quad y_n - Hx_n^f \\
 & \text{Obs.} \quad y_n \\
 & \text{F: Wave Propagation simulation} \\
 & n: \text{Time Step} \\
 & W: \text{Weighting Matrix}
 \end{aligned}$$

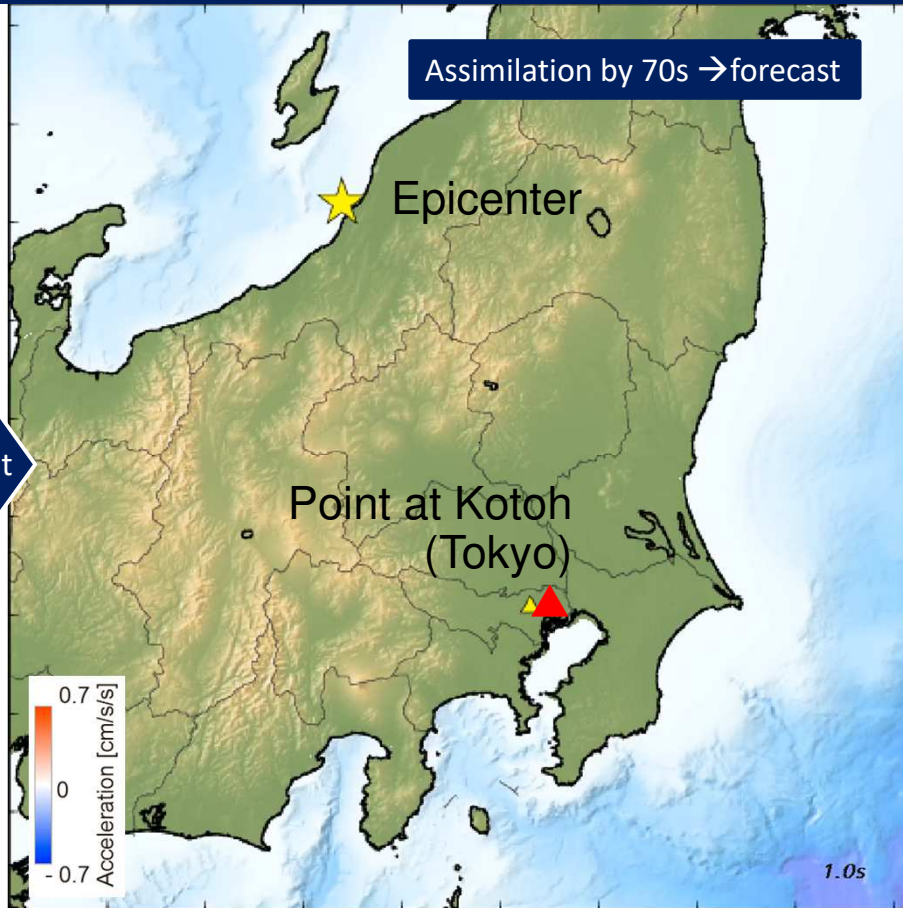
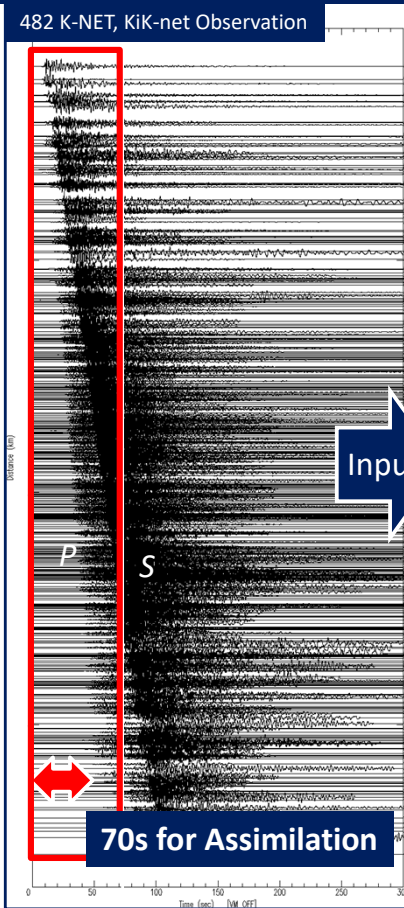
(A+S) Assimilation+Simulation

(Pure S) Pure Simulation/Forecast

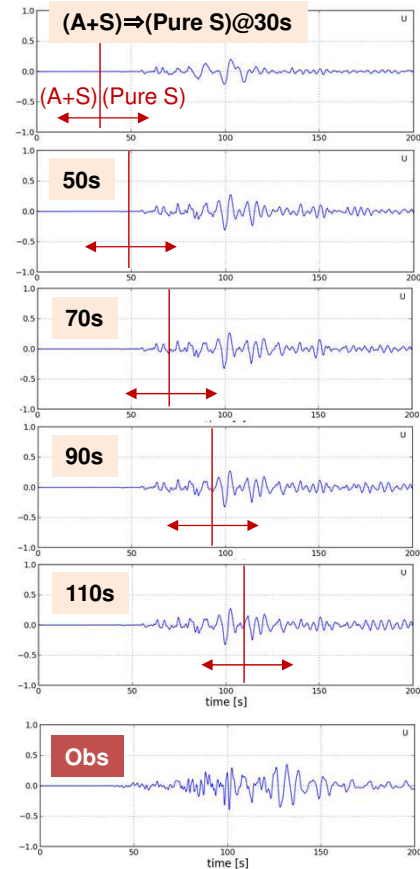


Data Assimilation + Pure Simulation/Forecast

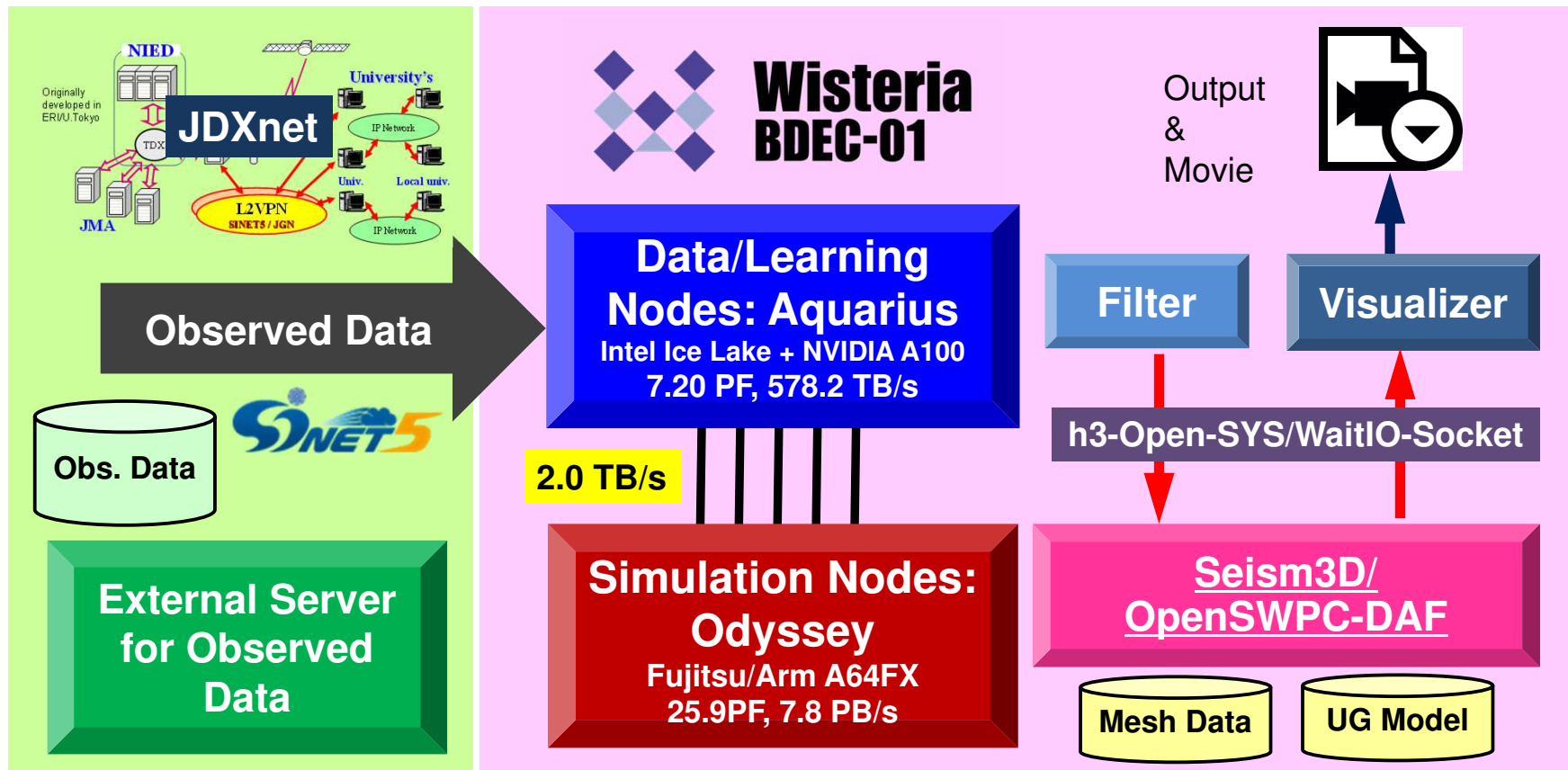
482 K-NET, KiK-net Observation



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

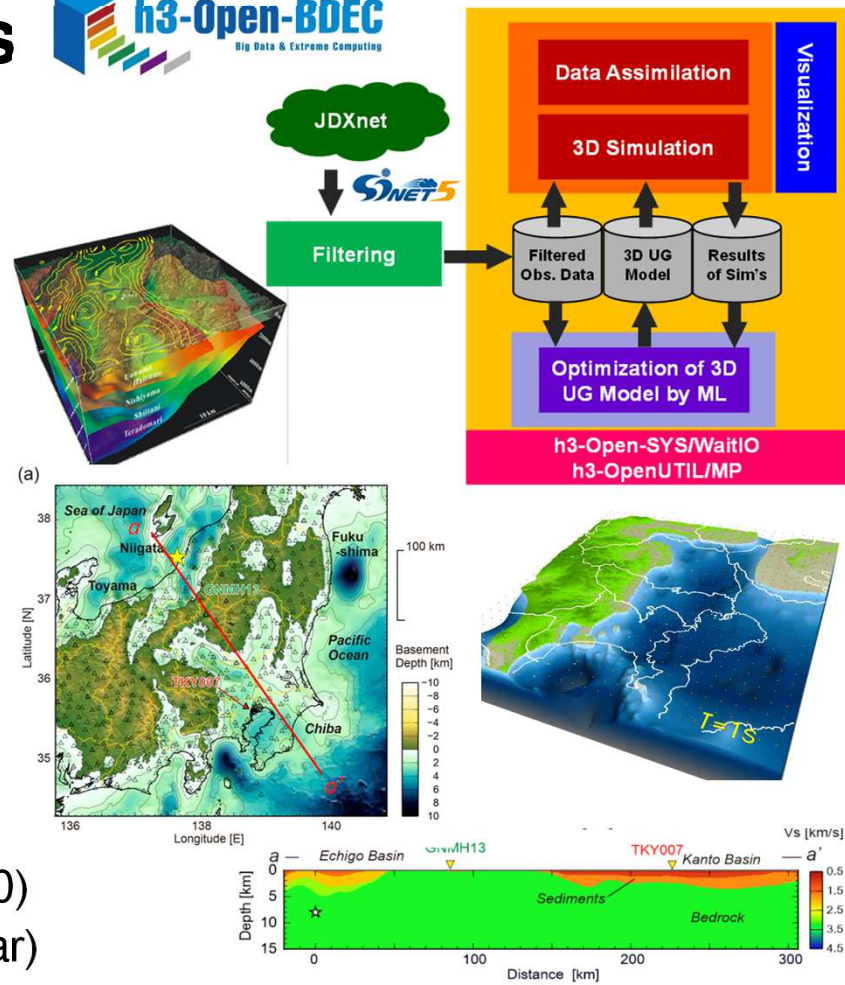


System on Wisteria/BDEC-01 using WaitIO

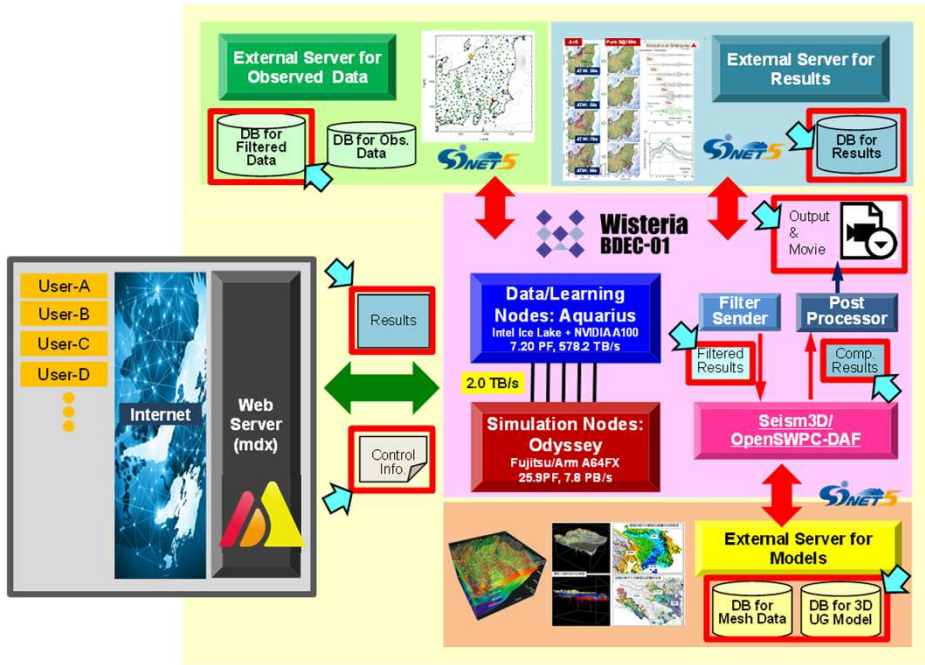


Future Directions towards Integration of (S+D+L)

- Accurate Prediction of Seismic Wave Propagation with Real-Time Data Observation/Assimilation
 - Emergency Info. for Safer Evacuation
- 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.g. $M_w < 3.0$)
 - More sophisticated DA method (e.g. 4DVar)

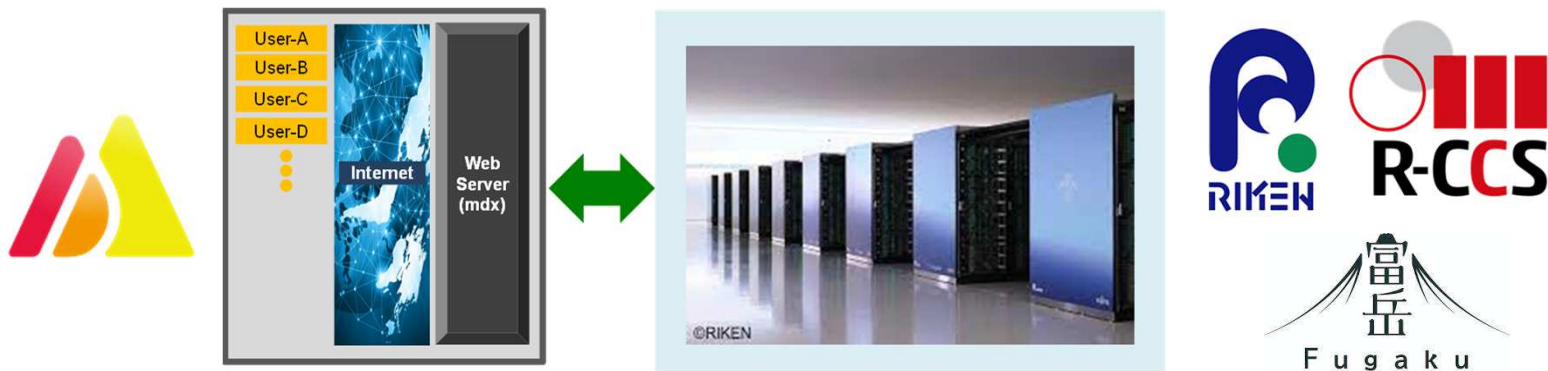


Web-based Simulation System for Outreach Activities



- Web-based simulation system for enlightenment of disaster prevention/mitigation using “3D Earthquake Simulation with Real-Time Data Observation/Assimilation”
- Users (e.g. students of high-school/junior-high-school/elementary school) access the web-server on the mdx, and manipulate simulations on the Wisteria/BDEC-01.
- The framework can be utilized in various types of applications.

Web-based System for CSE Simulations in the Program of Human Res. Dev. in RIKEN R-CCS

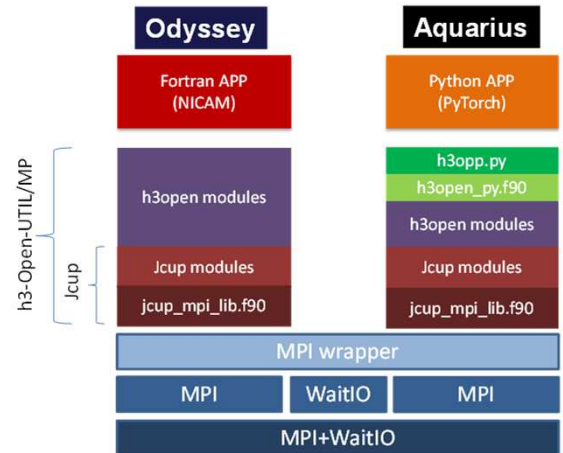
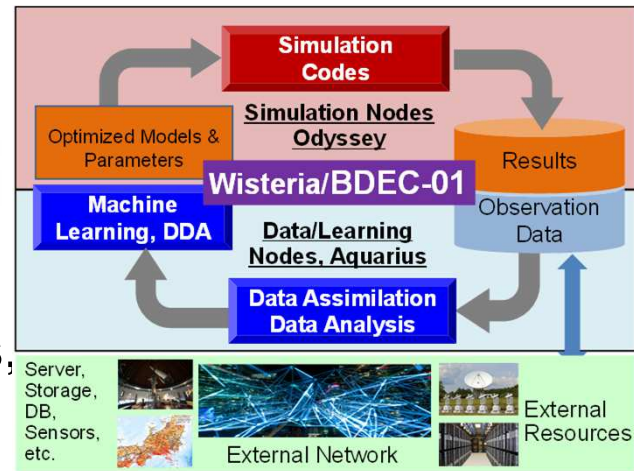


- Users can easily have experiences in supercomputing by the Fugaku
 - ✓ Resources of Fugaku is utilized like cloud, Simulations are “manipulated” by the web-server on mdx
- In FY.2022 prototype system on “computational solid mechanics” for university-level students will be developed
 - ✓ to be extended to younger generations, to other areas

MATLAB is available on Aquarius & OBCX after March 2022



- We are expecting excellent capabilities of MATLAB for machine learning and data analytics, which can be called from Fortran/C/C++ programs
- MATLAB only works on Aquarius, but if you apply h3-Open-BDEC (WaitIO, Coupler), you can combine the MATLAB workloads (called from Fortran/C/C++ programs) with large-scale simulations on Odyssey for optimization of parameters etc.
- (h3-Open-BDEC + MATLAB) could be an excellent tool for integration of (S+D+L)

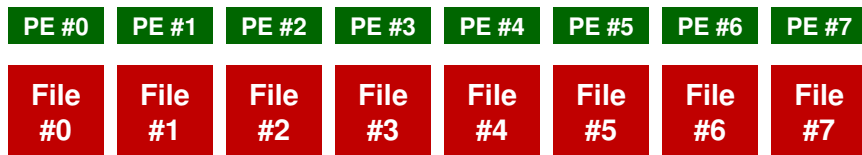
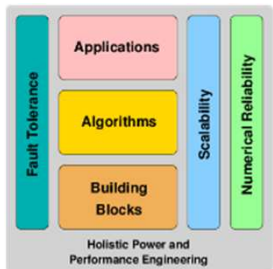


- Supercomputing Research Division, Information Technology Center ,The University of Tokyo (SCD/ITC/U.Tokyo)
- JHPCN, HPCI, JCAHPC and NHR
- Supercomputers in SCD/ITC/U.Tokyo
- Integration of (Simulation+Data+Learning)
 - Wisteria/BDEC-01
 - h3-Open-BDEC
- **Future Perspective**

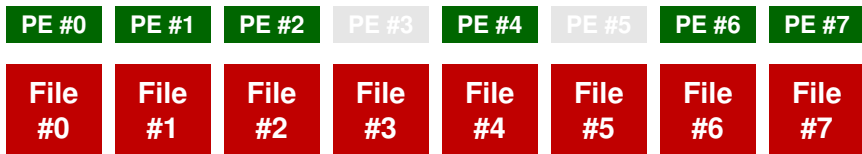
pFEM-CRAFT

[Fukasawa, Shahzad, KN, Wellein, SIAM PP18]

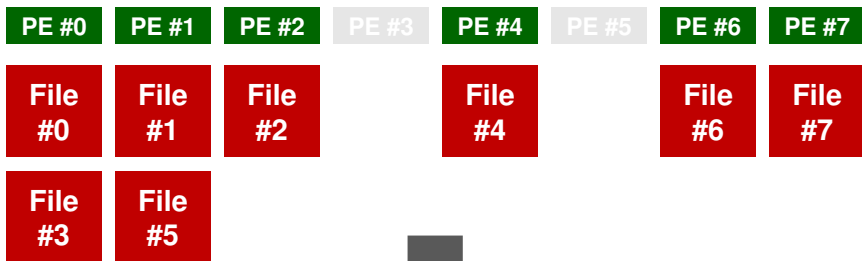
Parallel FEM Applications with Fault Resilience without Spare Nodes using CRAFT



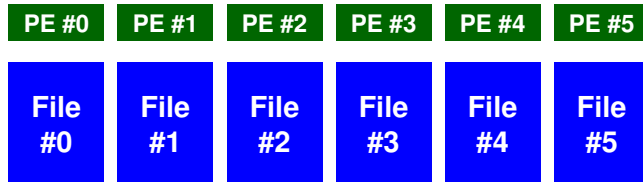
Starting with 8 Proc's



#3 & #5 failed



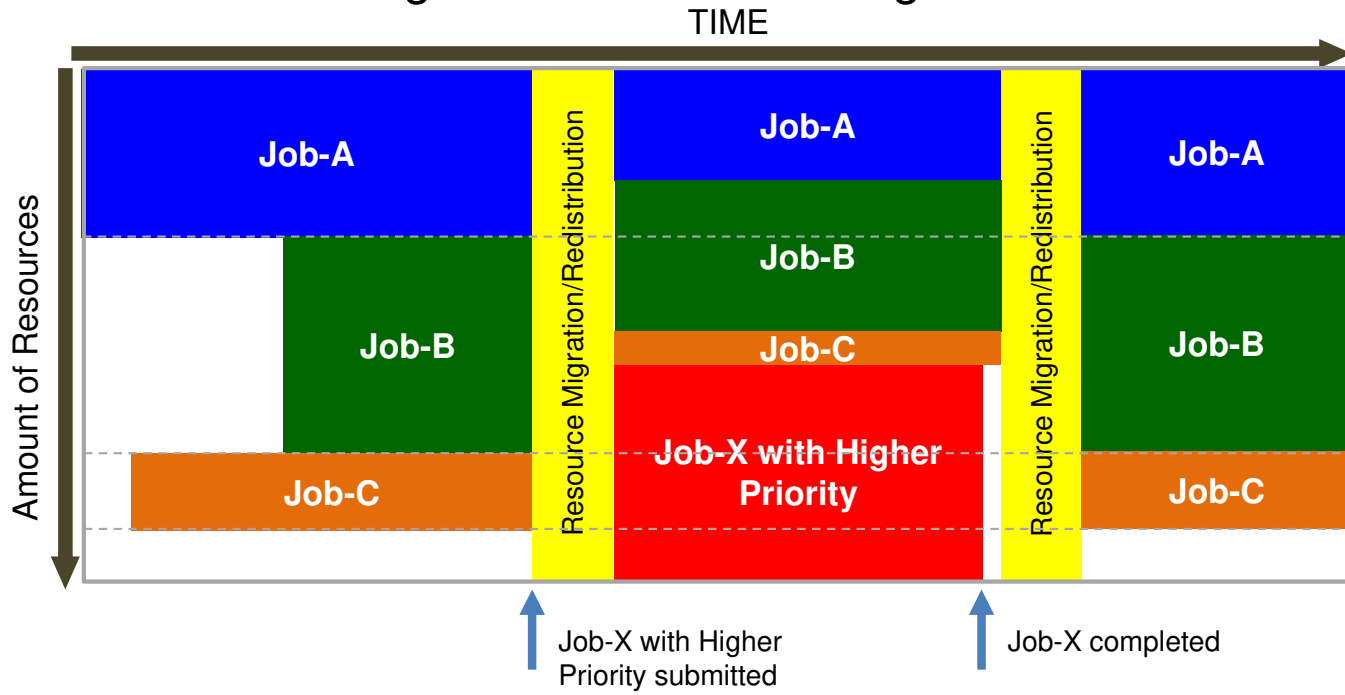
Files of #3 & #5 are read by #0 & #1 (Merging)



Repartitioning & Data Migration

CRAFT for Resource Management/Provisioning of Supercomputers under Urgent Situations

Each Application is equipped with pFEM-CRAFT-like capability/interface for flexible load balancing & resource/data migration



Future Perspective (1/2)

- U.Tokyo is shifting to GPUs/Accelerators in next 10 years
 - Maximum performance under constraint of power consumption
- Wisteria-Mercury (April-June 2023)
 - GPU Cluster, for supporting “Aquarius”
 - Prototype of OFP-II (128+ GPU's, 32+ nodes)
- OFP-II (April 2024)
 - Successor of OFP (JCAHPC, U.Tsukuba & U.Tokyo)
 - GPU Cluster (same GPUs as those of Mercury), 100+PF

2001-2005

2006-2010

2011-2015

2016-2020

2021-2025

2026-2030

Hitachi SR8000
1,024 GF

Hitachi SR11000
J1, J2
5.35 TF, 18.8 TF

Hitachi SR16K/M1
Yayoi
54.9 TF

Hitachi
SR2201
307.2GF

Hitachi
SR8000/MPP
2,073.6 GF

Hitachi HA8000
T2K Today
140 TF

OBCX
(Fujitsu)
6.61 PF

Oakforest-
PACS (Fujitsu)
25.0 PF

OFP-II
100+ PF

Fujitsu FX10
Oakleaf-FX
1.13 PF

Wisteria
BDEC-01 Fujitsu
33.1 PF

BDEC-
02
250+ PF

Reedbush-
U/H/L (SGI-HPE)
3.36 PF

Mercury

Ipomoea-01 25PB

Ipomoea-
03

Ipomoea-02

Supercomputers @ITC/U.Tokyo

2,600+ Users

55+% outside of U.Tokyo

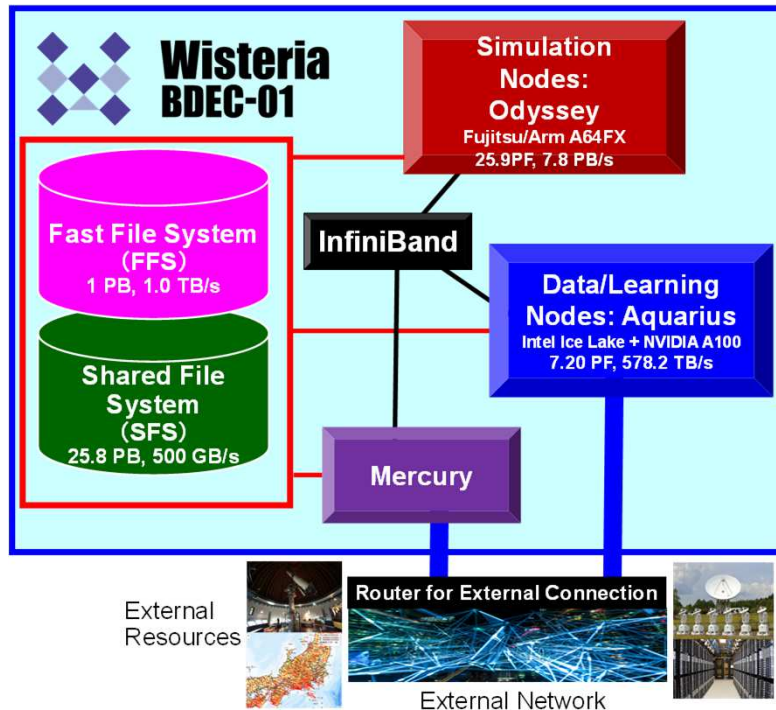
Future Perspective (1/2)

- U.Tokyo is shifting to GPUs/Accelerators in next 10 years
 - Maximum performance under constraint of power consumption
- Wisteria-Mercury (April-June 2023)
 - GPU Cluster, for supporting “Aquarius”
 - Prototype of OFP-II (128+ GPU's, 32+ nodes)
- OFP-II (April 2024)
 - Successor of OFP (JCAHPC, U.Tsukuba & U.Tokyo)
 - GPU Cluster (same GPUs as those of Mercury), 100+PF
- It is difficult for 3,000 users of OFP to perfectly switch to GPU cluster
 - We need certain amount of time and human resources
 - U.Tokyo considered this for Wisteria/BDEC-01, but finally gave up (Fall 2019)
- Current Situation
 - Various GPU's & Programming Environment
 - Good for NEW users of GPU's (e.g. Standard Language, OpenMP Offloadnig)

Wisteria-Mercury

Winged Messenger

- General Purpose CPU + Accelerators
 - Target Performance: 6+ PF, 128+GPU's
 - Prototype of OFP-II
- Supporting “Aquarius”: Primary Mission
 - Also used for HPC Workload
 - Direct Link to External Network
 - Working with “Odyssey” directly
- Some of infrastructures of Wisteria/BDEC-01 are shared
 - Login Nodes, HW for Management
 - SFS, FFS, Job Management System etc.
- Operation starts in April-June 2023



OFP-II (1/2)



JCAHPC



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO

- OFP-II: Successor of Oakforest-PACS (OFP) with Intel Xeon Phi
 - Operation starts in April 2024 (1 year after Mercury)
- Target Peak Performance: 100+PF (or 200 PF?)
 - Integration of (Simulation+Data+Learning)
 - Accelerators (e.g. GPU) are introduced because of power restriction
 - Group-A (CPU+GPU), Group-B (Only CPU): CPUs in Group-A and Group-B could be different

OFP-II (2/2)



- We should start porting ASAP, hopefully BEFORE Fall 2022
- **Additional Mission for Mercury**
 - **Prototype of OFP-II: Testbed for Porting Applications to GPU**
- 7 Benchmarks in CSE for decision of GPU on OFP-II & Mercury
 - OpenMP+MPI or CUDA/OpenACC
 - 3 Types of Benchmarks, 2-levels for each category
 - Performance Estimation on H/W for Mercury (Apr.2023) and OFP-II (Apr.2024)

A	Performance Evaluation of Codes for CPU on the Host CPU (Optional)	A-1	As-Is
		A-2	Optimized
B	Porting Codes for CPU to GPU	B-1	OpenACC, OpenMP, Standard Language etc.
		B-2	To be tuned for maximum performance o the target GPU
C	Optimization of Codes written in OpenACC/CUDA	C-1	As-Is
		C-2	To be tuned for maximum performance o the target GPU

Seven Benchmarks



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO

Name of the Code	Description	Lang.	Parallelization	GPU	Category
P3D	3-D Poisson's Equation by Finite Volume Method	C	OpenMP		A & B
GeoFEM/ICCG	Finite Element Method	Fortran	OpenMP, MPI		
H-Matrix	Hierarchical-Matrix calculation	Fortran	OpenMP, MPI		
QCD	Quantum-Chromo Dynamics simulation	Fortran	OpenMP, MPI	CUDA	C
N-Body	N-Body simulation using FDPS	C++	OpenMP, MPI	CUDA	
GROMACS	Molecular Dynamics simulation	C++	OpenMP, MPI	CUDA, HIP, SYCL	
SALMON	Ab-initio quantum-mechanical simulator for optics and nanoscience	Fortran	OpenMP, MPI	(OpenACC)	B

Purpose & Strategy of Each Category

- Category-B: (OpenMP+MPI) Codes
 - B-1: Minimum changes to the original codes (e.g. OpenMP, OpenACC, Standard Language etc.)
 - B-2: Fully Optimized by CUDA, OpenACC etc.
 - Evaluation: Performance of B-2, Ratio of (B-1)/(B-2) (Highly evaluated if performance of B-1 and B-2 would be close)
- Category-C: (OpenACC, CUDA) Codes
 - C-1: As Is
 - C-2: Fully Optimized
 - Evaluation: Performance of C-2
- Both of Performance and Portability are evaluated
 - Portability is especially important in B-1

Status of Evaluation

- Report due May 20, 2022
- Online presentations/interviews just before ISC
- GPU for Mercury system will be decided soon
 - The result will be shown in the “draft specification” to be released in mid-late June
- This time, we select GPU for Mercury (128+GPU's, U.Tokyo), but it (or its successor) will be automatically adopted to OFP-II (100+PF, JCAHPC)
 - Very Critical Decision
 - Thus, we can start porting this Fall !!

Summary of Schedule

- June 2022 (Mid-Late)
 - GPU for Mercury and OFP-II decided
- Late Summer-Fall 2022
 - Porting of applications starts on target GPU
 - Procurement of OFP-II starts (we start discussions about that with vendors at SC22)
- April-June 2023
 - Operation of Mercury starts
 - Performance evaluation and further optimization of ported codes on Mercury
- April 2024
 - Operation of OFP-II starts

2001-2005

2006-2010

2011-2015

2016-2020

2021-2025

2026-2030

Hitachi SR8000
1,024 GF

Hitachi SR11000
J1, J2
5.35 TF, 18.8 TF

Hitachi SR16K/M1
Yayoi
54.9 TF

Hitachi
SR2201
307.2GF

Hitachi
SR8000/MPP
2,073.6 GF

OBCX
(Fujitsu)
6.61 PF

Hitachi HA8000
T2K Today
140 TF

Oakforest-
PACS (Fujitsu)
25.0 PF

OFP-II
100+ PF

Fujitsu FX10
Oakleaf-FX
1.13 PF

Wisteria
BDEC-01 Fujitsu
33.1 PF

BDEC-
02
250+ PF

Supercomputers @ITC/U.Tokyo

2,600+ Users

55+% outside of U.Tokyo

Reedbush-
U/H/L (SGI-HPE)
3.36 PF

Mercury

Ipomoea-01 25PB

Ipomoea-
03

Ipomoea-02

Future Perspective (2/2)

- BDEC-02 (Fall 2027-Spring 2028)
 - Platform for “Digital Twin”, “S+D+L”
 - We are thinking about introducing DPU, IPL Quantum-Inspired Devices etc. for supporting workloads for (D+L)
 - Some of them can be also used for HPC simulations with certain constraints
 - ✓ Graphcore, SambaNova, Cerebras
 - ✓ PIUMA@Intel
 - ✓ Digital Annealer@Fujitsu
 - We have been using Fujitsu’s Digital Annealer since 2019: Combinatorial Optimization
 - Programming Environment & Communication Library for Integration of HPC and Such Devices are needed: Examples of prototype in DOE Lab’s

