



“Brief Introduction” to “An Introduction to GPU Programming Models”

Kengo Nakajima

**Lead, Research & Development Division of JCAHPC
Information Technology Center, The University of Tokyo**

October 21, 2022

- Joint Center for Advanced High Performance Computing, since 2013
 - University of Tsukuba & University of Tokyo
 - Budgets of 2 Supercomputing 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 (KNL), 25PF, Fujitsu
 - Top500 (#6 (Nov.2016), #1 in Japan)
 - National Flagship System “in fact” (Oct.2019-Mar.2021) after shutdown of the K computer
 - Retired on March 31, 2022 (#39 (Nov.2021))
- We are starting procurement for OFP-II, successor of OFP, whose operation starts in April 2024



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
200+ 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

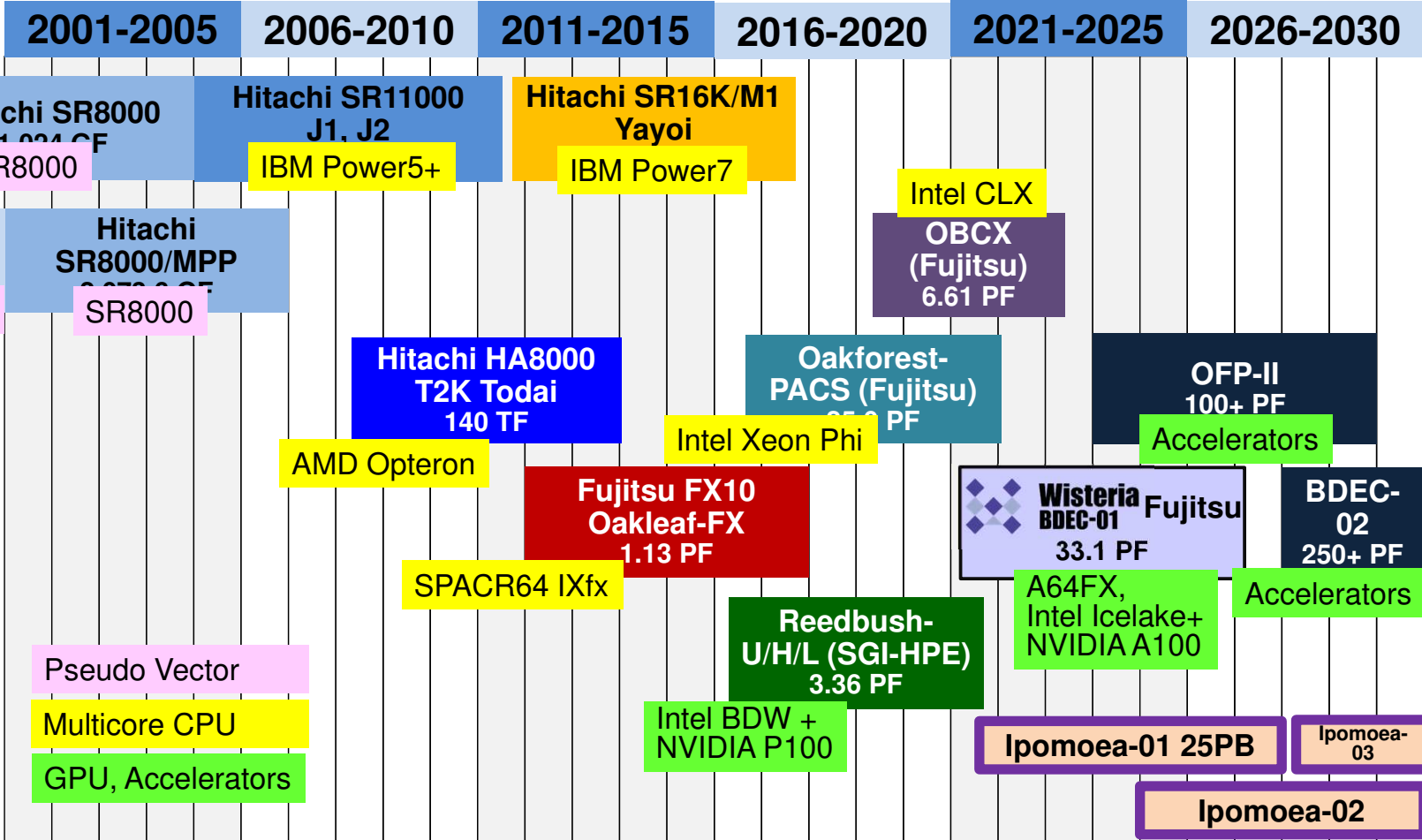
Reedbush-U/H/L (SGI-HPE)
3.36 PF

Mercury

Ipomoea-01 25PB

Ipomoea-03

Ipomoea-02



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
200+ 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

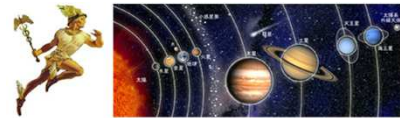
Future Perspective of JCAHPC



- JCAHPC is shifting to GPUs
 - Maximum performance under constraint of power consumption

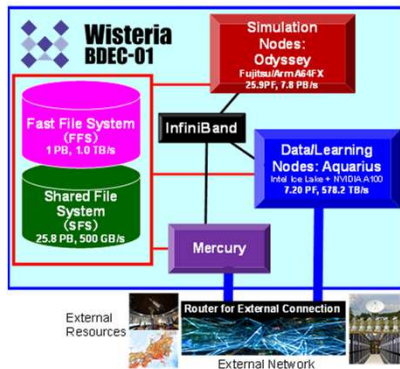
Wisteria-Mercury (June 2023)

- GPU Cluster, for supporting “Aquarius”@ITC/U.Tokyo
- Prototype of OFP-II (128+ GPU’s, 32+ nodes)



OFP-II (April 2024)

- Successor of OFP (JCAHPC, U.Tsukuba & U.Tokyo), 200+PF
- Group-A (CPU+GPU), Group-B (Only CPU)
 - Same GPUs as those of Mercury
 - CPUs in Group-A and Group-B could be different



- Porting codes of 3,000+ users of OFP to GPU is the most critical issue
 - Starting this Fall

- Integration of (Simulation + Data + Learning): Both Systems focusing of “AI for Science”

Summary of Schedule (1/2)

Mercury & OFP-II



- **The work for porting to GPUs should begin in the fall of 2022 (NOW) at the latest, because the operation of OFP-II starts in April 2024.**
- **The GPUs for both Mercury and OFP-II systems (they are the same) must be determined before then.**

Summary of Schedule (2/2)



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO

Mercury & OFP-II

- February-May 2022
 - Pre-Benchmarks for selection of the target GPU on Mercury and OFP-II
 - Seven benchmarks in CSE (Fortran, C)
- June 2022 (Mid-Late)
 - GPU for Mercury and OFP-II decided (already decided)
 - Based on Performance, Ease of Porting, Support System, Fortran Support
- Fall 2022
 - Porting of applications starts using Aquarius etc. (probably in December)
- June 2023
 - Operation of Mercury starts
 - Performance evaluation and further optimization of ported codes on Mercury
- April 2024
 - Operation of OFP-II starts

How to port your codes

Mercury & OFP-II



- If your code is already parallelized by (OpenMP+MPI), porting to GPU is not so difficult.
 - Sometimes, you need to change data/loop structure (e.g. AoS, SoA ...)
 - OpenACC, OpenMP, CUDA, Standard Language etc.
- We start to host “Online Mini-Camp’s” frequently from this December
 - You can bring your code(s)
 - Members of JCAHPC and the GPU Vendor will support porting

Outline of Today's Lecture

70+ Registrations !!



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO

10

- Varieties of environments for programming are available on our target GPU (NVIDIA H100 Tensor Core GPU)
 - OpenACC, OpenMP, CUDA, Standard Language etc.
- The speaker (Jeff Larkin (NVIDIA)) will introduce some of those with code examples and performance evaluations
- We hope this lecture will be helpful for you to choose appropriate programming environment on the GPU for your code(s)
- The lecture and Q/A session are video-recorded, and it will be open to public through our HP (with this file).
- You can also watch the following video (in Japanese, OpenACC & CUDA)
 - <https://www.youtube.com/watch?v=3Q7TDodhbho&t=2s>

Speaker: Jeff Larkin (NVIDIA)

<https://developer.nvidia.com/blog/author/jlarkin/>



- Jeff is a Principal HPC Application Architect in NVIDIA's HPC Software team.
- **He is passionate about the advancement and adoption of parallel programming models for High Performance Computing.**
- He was previously a member of NVIDIA's Developer Technology group, specializing in performance analysis and optimization of high performance computing applications.
- Jeff is also the chair of the OpenACC technical committee and has worked in both the OpenACC and OpenMP standards bodies.
- Before joining NVIDIA, Jeff worked in the Cray Supercomputing Center of Excellence, located at Oak Ridge National Laboratory.