

Miyabi



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO

JCAHPC

JCAHPC Miyabi 利用説明会

(代表窓口)

東京大学情報基盤センター
スーパーコンピューティング研究部門

<http://www.cc.u-tokyo.ac.jp/>

問合せ先: uketsuke@cc.u-tokyo.ac.jp

本日のスケジュール



14:00-14:40 利用説明

- 最先端共同HPC基盤施設 (JCAHPC)の紹介
- Miyabiの概要
- Miyabi利用方法の紹介
 - リソースグループ、MIG
- Miyabi利用に向けたユーザーサポート
 - 講習会、GPUミニキャンプ、相談会
- Miyabi利用支援ポータル

14:40-14:55 チュートリアル

15:00-15:25 筑波大学の利用制度説明 or Miyabi見学(その1)

15:30-15:55 東京大学の利用制度説明 or Miyabi見学(その2)

16:00-16:30 Miyabi見学(その3)

本日のスケジュール

14:00-14:40 利用説明

- **最先端共同HPC基盤施設 (JCAHPC)の紹介**
- Miyabiの概要
- Miyabi利用方法の紹介
 - リソースグループ、MIG
- Miyabi利用に向けたユーザーサポート
 - 講習会、GPUミニキャンプ、相談会
- Miyabi利用支援ポータル

14:40-14:55 チュートリアル

15:00-15:25 筑波大学の利用制度説明 or Miyabi見学(その1)

15:30-15:55 東京大学の利用制度説明 or Miyabi見学(その2)

16:00-16:30 Miyabi見学(その3)

最先端共同HPC基盤施設

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



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO

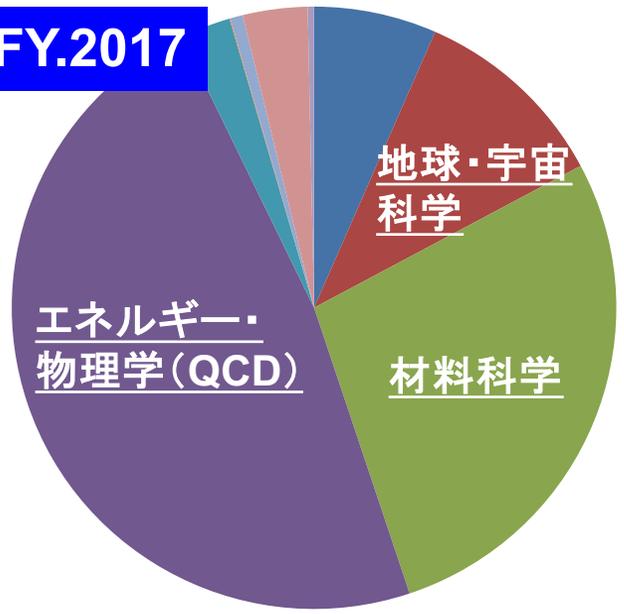
JCAHPC

- 2013年創立
 - JCAHPC (Joint Center for Advanced High Performance Computing)
 - 筑波大学計算科学研究センター・東京大学情報基盤センター
 - 最先端計算科学の推進
 - 大規模システムの設計・導入・運用を共同で実施する我が国でも初の試み: より大規模なシステムを効率的に導入可能
- Oakforest-PACS (OFP) : JCAHPC第一号機
 - Intel Xeon Phi 8,208ノード, 25PF (富士通)
 - 2016年11月時点でTop500の6位 (国内1位)
 - 2022年3月末で退役
 - 「京」の退役後, 「富岳」登場までの2019・2020年度は事実上の「National Flagship System」としての役割を担う
- OFP-II (OFP後継機) へ向けた試み

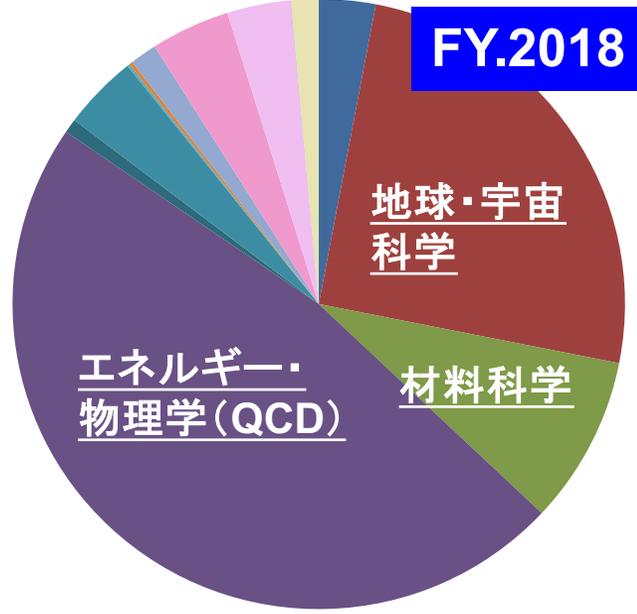


研究分野別CPU時間割合:OFP

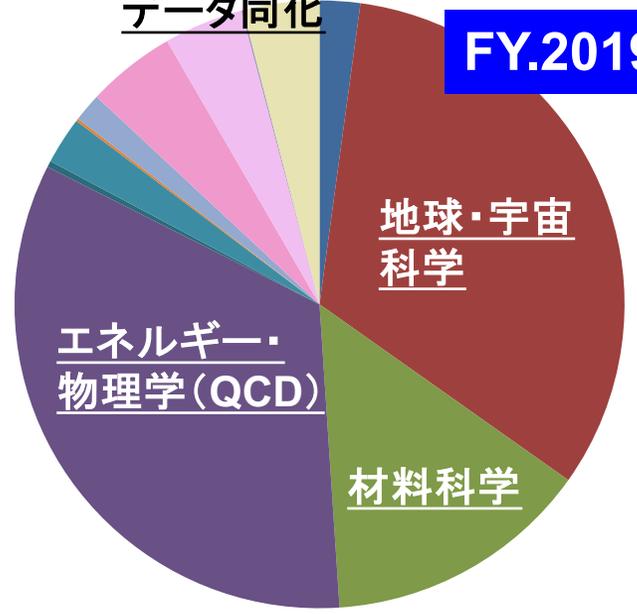
FY.2017



FY.2018



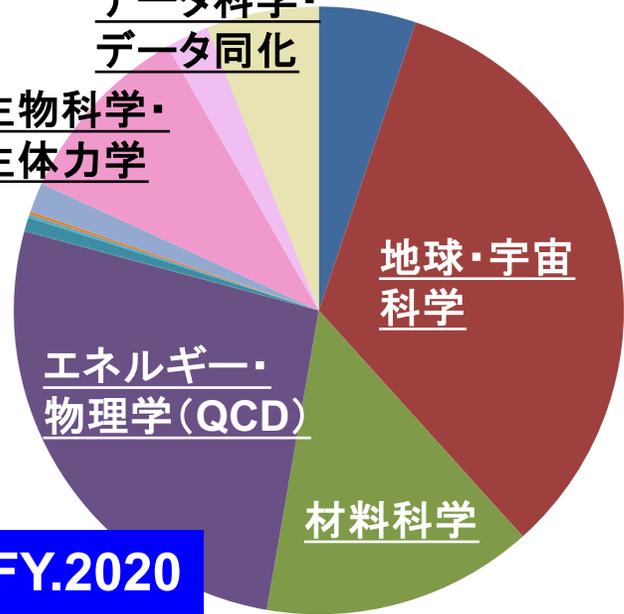
FY.2019



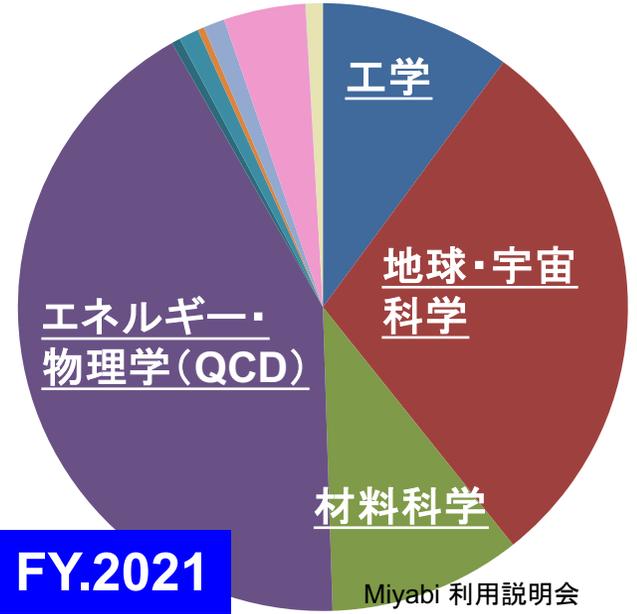
データ科学・データ同化

生物科学・生体力学

FY.2020



FY.2021



Miyabi 利用説明会

- 工学・ものづくり
- 教育
- 地球科学・宇宙科学
- 産業利用
- 材料科学
- 生物科学・生体力学
- エネルギー・物理学
- バイオ
- 情報科学:システム
- インフォマティクス
- 情報科学:アルゴリズム
- 社会科学・経済
- 情報科学:AI
- データ科学・データ同化



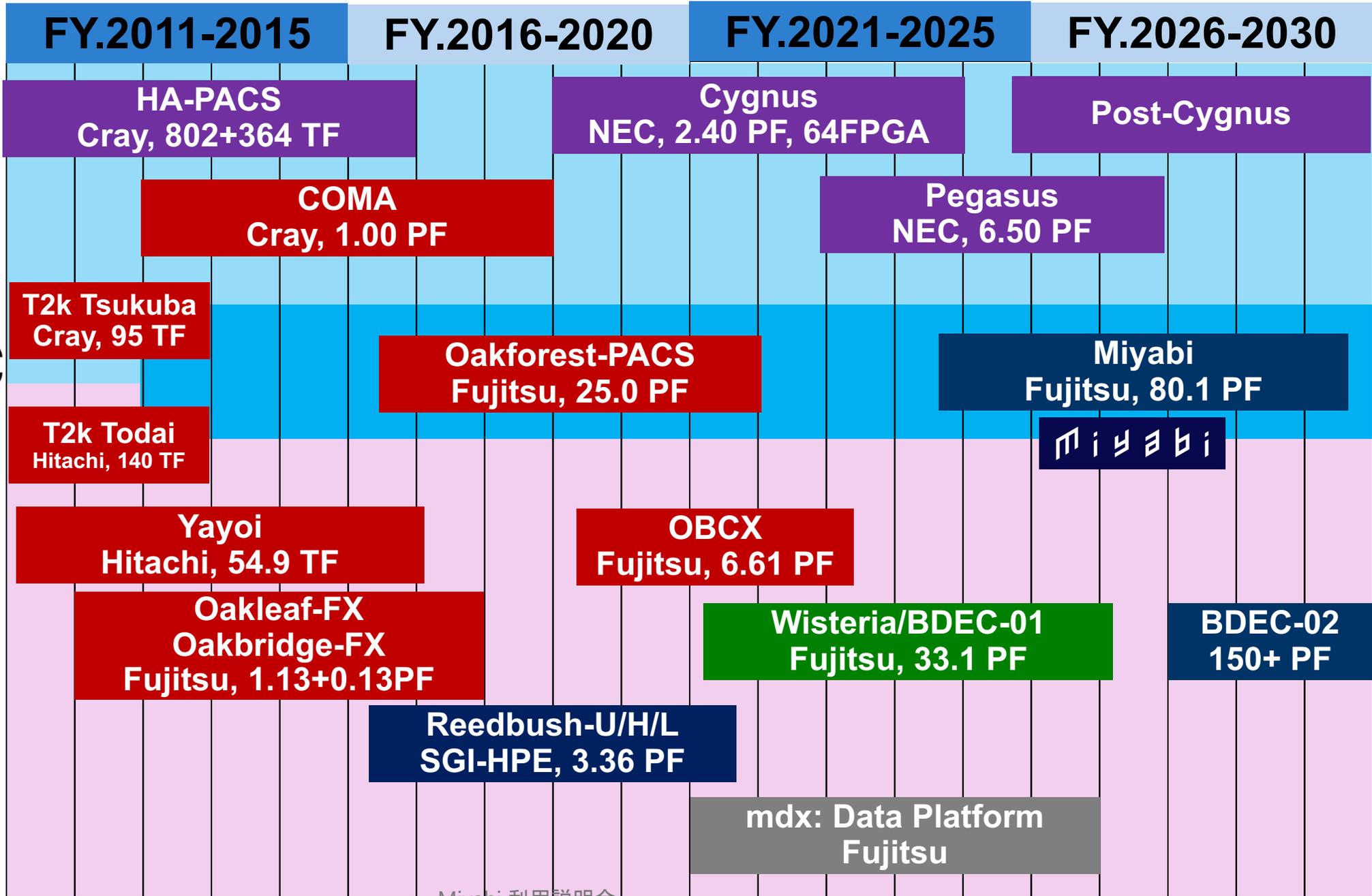
筑波大学
University of Tsukuba



JCAHPC



東京大学
THE UNIVERSITY OF TOKYO



筑波大：Pegasusビッグメモリスーパーコンピュータ



- Installed in Q4 2022
- Total Performance
 - 150 nodes, 8.1 PFlops, 300 TiB Pmem
- Node specification
 - 3.2 TFlops Intel Platinum 8468 (**Sapphire Rapids**)
 - 51 TFlops NVIDIA **H100** PCIe GPU
 - 128 GiB **DDR5** DRAM (282 GB/s)
 - 2 TiB **Optane PM 300** series (Crow Pass)
 - 6 TB NVMe SSD (7 GB/s)
- Interconnection Network
 - NVIDIA Quantum-2 InfiniBand platform (200 Gbps) full bisection (**InfinBand NDR200**)
- Parallel File System
 - 7.1 PByte DDN EXAScaler (40 GB/s)



NEC LX B1000E Blade Enclosure

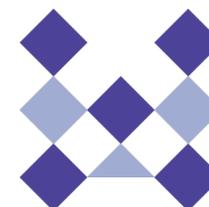
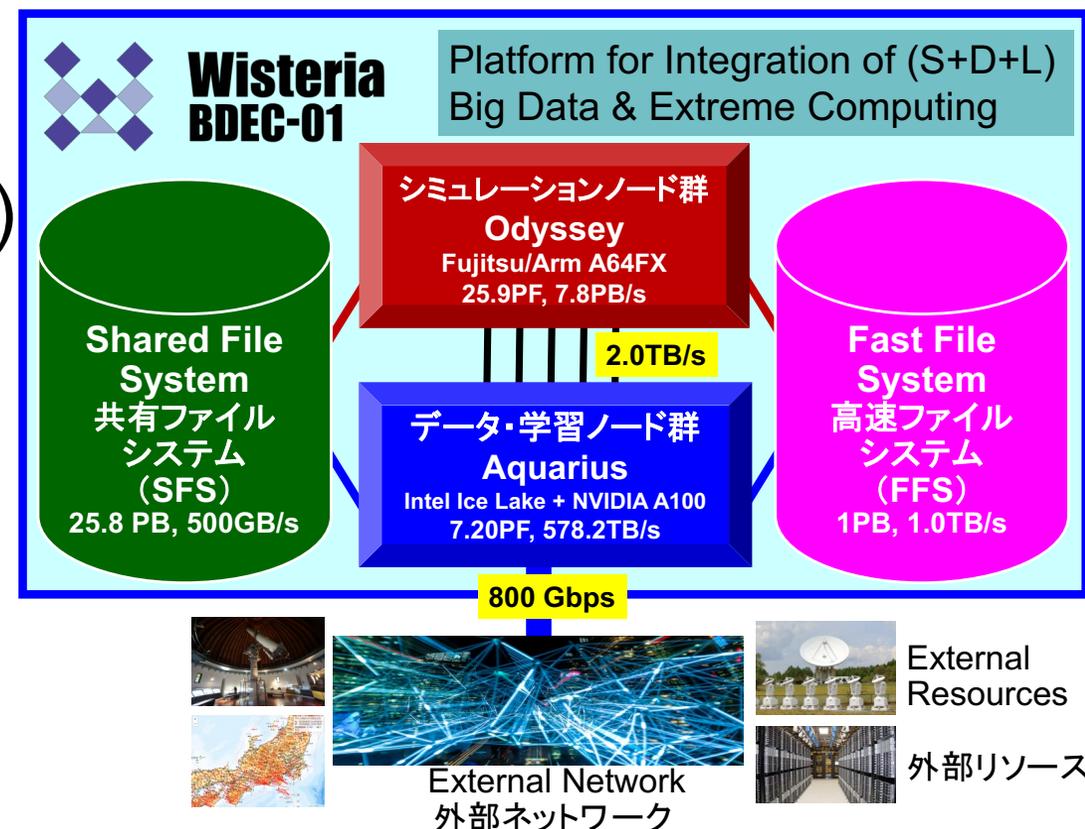


NEC LX 102Bk-6

東大: Wisteria/BDEC-01

- 2021年5月14日運用開始
 - 東京大学柏Ⅱキャンパス
- 33.1 PF, 8.38 PB/sec., **富士通製**
 - ~4.5 MVA(空調込み), ~360m²
- Hierarchical, Hybrid, Heterogeneous (h3)
- 2種類のノード群**
 - シミュレーションノード群 (S, SIM) : **Odyssey**
 - 従来のスパコン
 - **Fujitsu PRIMEHPC FX1000 (A64FX), 25.9 PF**
 - 7,680ノード(368,640 コア), 20ラック, Tofu-D
 - データ・学習ノード群 (D/L, DL) : **Aquarius**
 - データ解析, 機械学習
 - **Intel Xeon Ice Lake + NVIDIA A100, 7.2 PF**
 - 45ノード(Ice Lake:90基, A100:360基), IB-HDR
 - 外部リソース(ストレージ, サーバー, センサーネットワーク他)に直接接続
 - ファイルシステム: 共有(大容量) + 高速

BDEC:「計算・データ・学習 (S+D+L)」
融合のためのプラットフォーム
(Big Data & Extreme Computing)



**Wisteria
BDEC-01**

本日のスケジュール



14:00-14:40 利用説明

- 最先端共同HPC基盤施設 (JCAHPC)の紹介
- **Miyabiの概要**
- Miyabi利用方法の紹介
 - リソースグループ、MIG
- Miyabi利用に向けたユーザーサポート
 - 講習会、GPUミニキャンプ、相談会
- Miyabi利用支援ポータル

14:40-14:55 チュートリアル

15:00-15:25 筑波大学の利用制度説明 or Miyabi見学(その1)

15:30-15:55 東京大学の利用制度説明 or Miyabi見学(その2)

16:00-16:30 Miyabi見学(その3)

Miyabiの概要 (1/3)

2025年1月運用開始 80.1 PFLOPS



筑波大学
University of Tsukuba



東京大学
THE UNIVERSITY OF TOKYO

Miyabi-G : 78.8 PFLOPS, 5.07 PB/s

(演算加速ノード)

Supermicro

CPU+GPU: NVIDIA GH200 Superchip

CPU: NVIDIA Grace

(72 コア, 3.0 GHz, 117MB L3 Cache)

Mem: 120 GB (LPDDR5X, 512 GB/sec)

GPU: NVIDIA H100

(66.9 TFLOPS, NVLink-C2C 450 GB/sec(片方向))

Mem: 96 GB (HBM3, 4.022 TB/sec)

× 1,120

Miyabi-C : 1.3 PFLOPS, 608 TB/s

(汎用CPUノード)

富士通 PRIMERGY Server

CPU: Intel Xeon CPU Max 9480 x2 ソケット

(56 コア, 1.9GHz, 112.5MB L3 Cache) x2

Mem: 128 GiB (HBM2E, 3.2 TB/sec)

× 190

InfiniBand NDR200
(200 Gbps)

InfiniBand NDR200
(200 Gbps)

InfiniBand NDR (400 Gbps), Full-bisection Fat-Tree

1.0 TB/s

共有ファイルシステム
Lustre FS

11.3 PB
All Flash

ログインノードプリポスト



汎用CPUノード
+プリポスト用

Intel Xeon 8480+ x2



演算加速ノード用

NVIDIA Grace
CPU Superchip

外部接続
ルータ



Ethernet
RDMA

大規模共通ファイル
システム(東大)
Ipomoea-01
Lustre FS
25.9 PB

設置・運用:
富士通

2025/1/16

DDN ES400 NVX2 x10

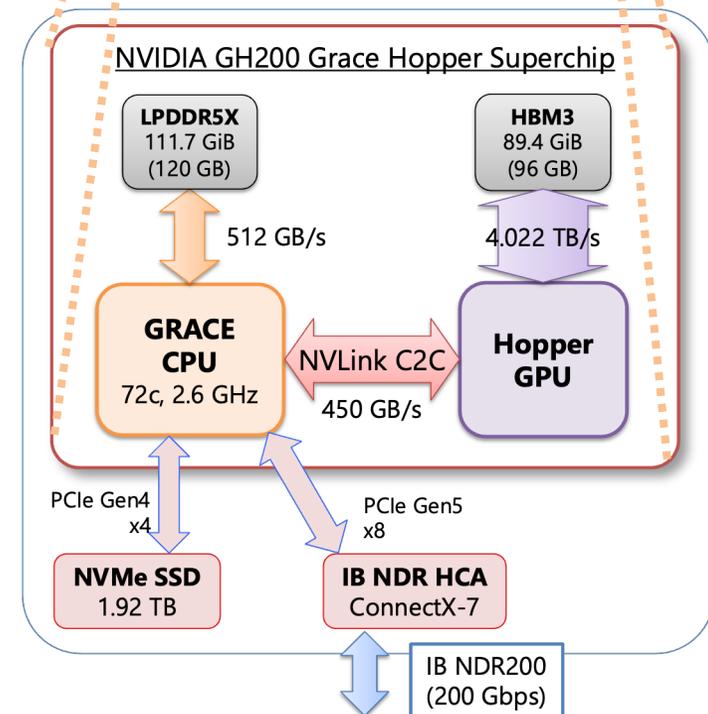
Miyabiの概要 (2/3)

• Miyabi-G: 演算加速ノード: NVIDIA GH200

- 計算ノード: NVIDIA GH200 Grace-Hopper Superchip
 - Grace: 72c, 3.45 TF, 120 GB, 512 GB/sec (LPDDR5X)
 - H100: 66.9 TF DP-Tensor Core, 96 GB, 4,022 GB/sec (HBM3)
 - CPU-GPU間はキャッシュコヒーレント
 - NVMe SSD for each GPU: 1.9TB, 8.0GB/sec, GPUDirect Storage
- **合計 (CPU+GPUの合計値)**
 - **1,120 ノード, 78.8 PF, 5.07 PB/sec, IB-NDR 200**

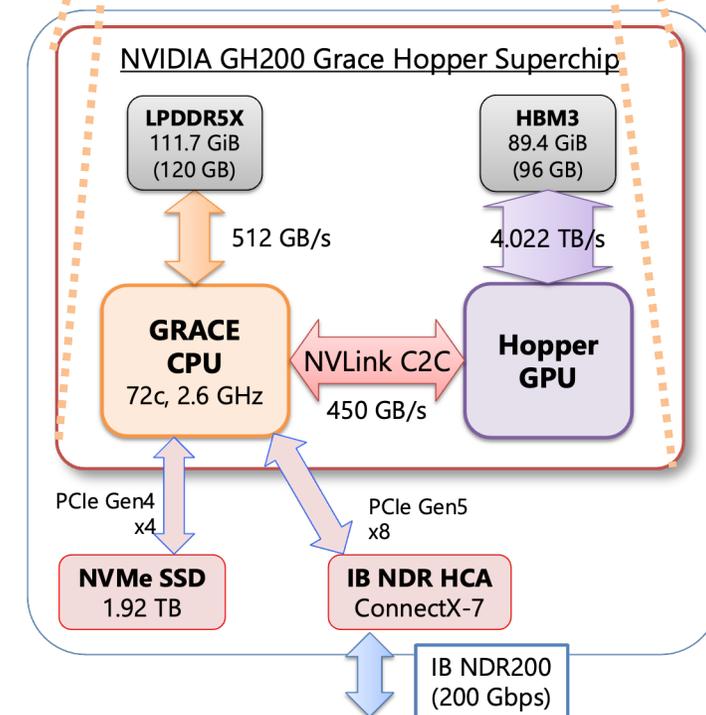
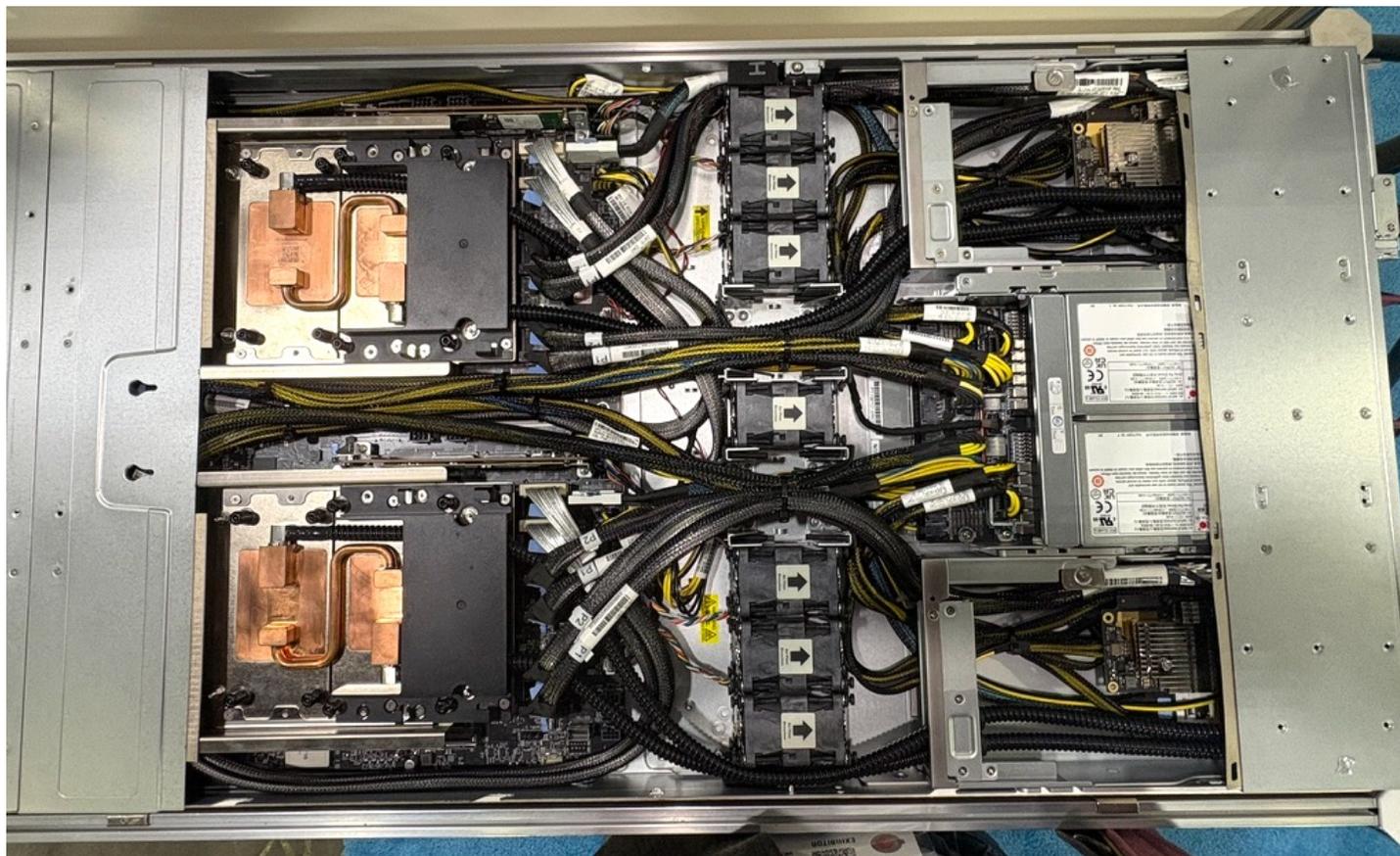
• Miyabi-C: 汎用CPUノード: Intel Xeon Max 9480 (SPR)

- 計算ノード: Intel Xeon Max 9480 (1.9 GHz, 56c) x 2
 - 6.8 TF, 128 GiB, 3,200 GB/sec (HBM2e only)
- **合計**
 - **190 ノード, 1.3 PF, IB-NDR 200**
 - **372 TB/sec for STREAM Triad (Peak: 608 TB/sec)**



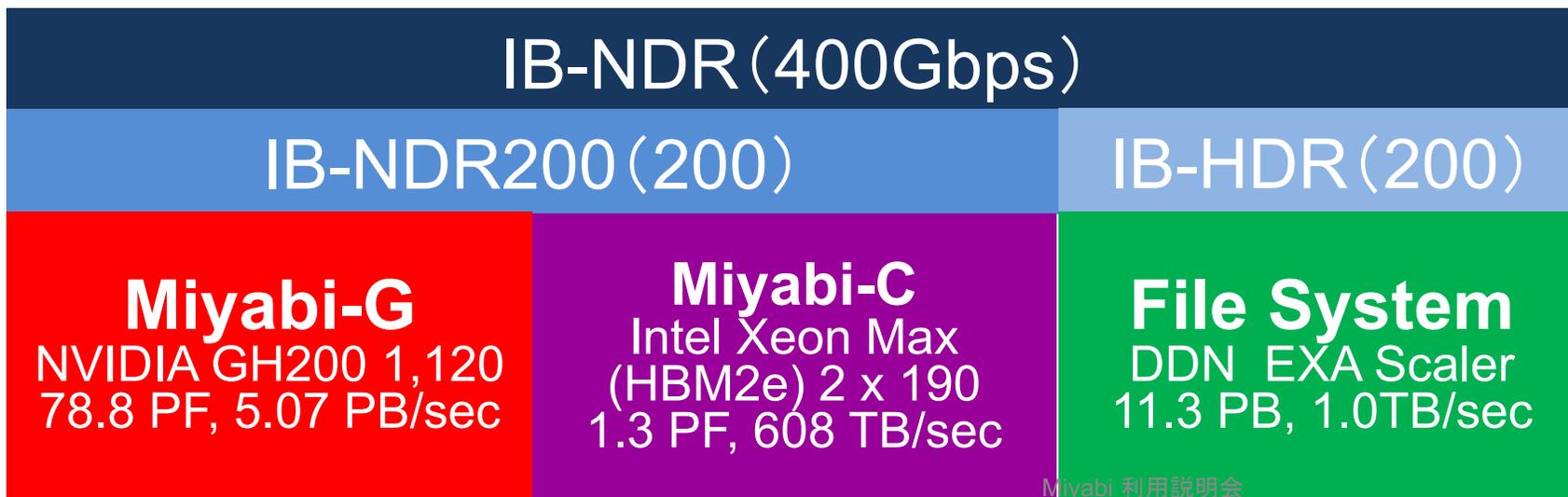
Miyabiの概要 (2/3)

- Supermicro ARS-111GL-DNHR-LCC
– 1U 2ノード、直接水冷



Miyabiの概要 (3/3)

- **ファイルシステム: DDN EXAScalar, Lustre FS**
 - 11.3 PB (NVMe SSD) 1.0TB/sec, "Ipomoea-01" (26 PB) も利用可能
- **Miyabi-G/C の全ノードはフルバイセクションバンド幅Fat Treeで接続**
 - $(400\text{Gbps}/8) \times (32 \times 20 + 16 \times 1) = 32.8 \text{ TB/sec}$
- **2025年1月運用開始、Miyabi-G/C間の通信はh3-Open-SYS/WaitIOにより実現**



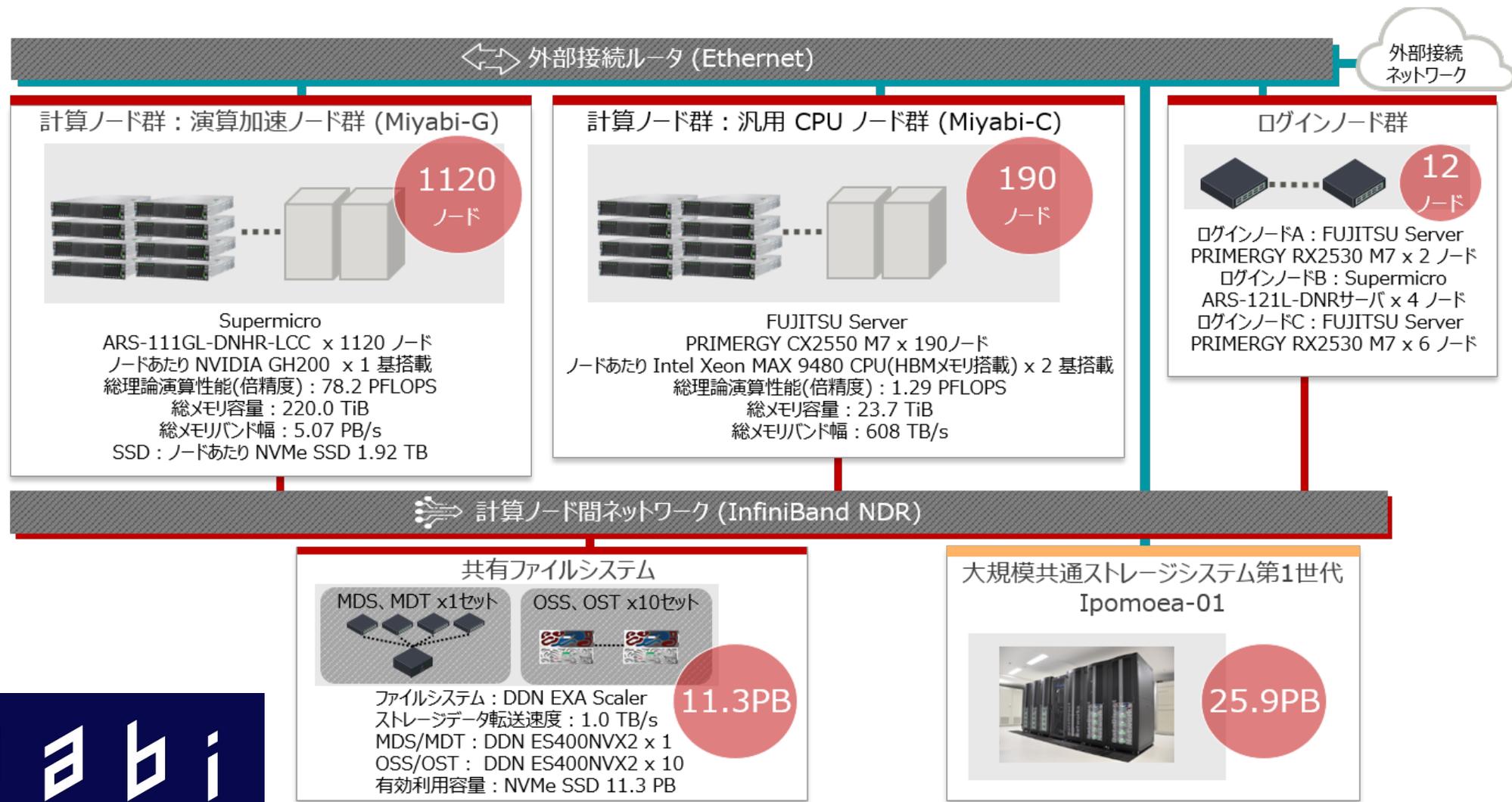
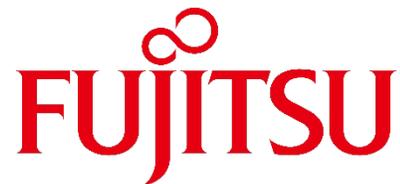
Miyabi 仕様まとめ

	Miyabi-G (演算加速ノード)	Miyabi-C (汎用CPUノード)
理論ピーク性能	78.8 PFLOPS	1.29 PFLOPS
ノード数	1,120	190
合計メモリ容量	241.9 TB	23.75 TiB
合計メモリバンド幅	5.07 PB/sec	608 TB/sec
インタコネク トポロジ	InfiniBand NDR200 (200 Gbps) Full-bisection Fat Tree	

項目	Miyabi-G (演算加速ノード)	Miyabi-C (汎用CPUノード)
サーバ	Supermicro ARS-111GL-DNHR-LCC	FUJITSU Server PRIMERGY CX2550 M7
プロセッサ名	NVIDIA GH200 Grace Hopper Superchip, NVIDIA Grace	Intel Xeon CPU Max 9480 (Sapphire Rapids)
プロセッサ数 (コア 数)	1 (72)	2 (56+56)
周波数	3.0 GHz	1.9 GHz
理論演算性能	3.456 TFLOPS	6.8096 TFLOPS
メモリ	LPDDR5X	HBM2E
メモリ容量	120 GB	128 GiB
メモリ帯域幅	512 GB/s	3.2 TB/s
プロセッサ名	NVIDIA Hopper	
プロセッサ数	1	
SM数	132	
理論演算性能	66.9 TFLOPS	
メモリ	HBM3	-
メモリ容量	96 GB	
メモリ帯域幅	4.02 TB/s	
CPU-GPU間接続	NVLink-C2C 450 GB/sec (片方 向) キャッシュコヒーレント	
NVMe SSD	1.92 TB, PCIe Gen4 x4	-

共有ファイルシステム		Lustre FS
M DS	サーバ	DDN ES400NVX2
	サーバ数(VM)	1 (4)
	inode数	appx. 23.5 B
O SS	サーバ	DDN ES400NVX2
	サーバ数	10 set
	容量	11.3 PB (All Flash)
	理論バンド幅	1.0 TB/sec

Miyabi システム構成図





Miyabi ソフトウェア構成 (1/2)




	汎用CPUノード群	演算加速ノード群
OS	Rocky Linux 9 (ログインノードはRed Hat Enterprise Linux 9)	
ジョブスケジューラ	PBS Professional	
コンパイラ	GNU コンパイラ	
	Intelコンパイラ Fortran77/90/95/2003/2008 C、C++	NVIDIA HPC SDK Fortran77/90/95/2003/2008 C、C++ OpenMP、OpenACC NVIDIA CUDA Toolkit CUDA C CUDA C++
メッセージ通信ライブラリ	Intel MPI	Open MPI
ライブラリ	—	
	BLAS、CBLAS、LAPACK、ScaLAPACK、SuperLU、SuperLU MT、SuperLU DIST、METIS、MT-METIS、ParMETIS、Scotch、PT-Scotch、PETSc、Trillinos、FFTW、GNU Scientific Library、NetCDF、Parallel netCDF、HDF5、Parallel HDF5、OpenCV、Xabclib、ppOpen-HPC、MassiveThreads、Standard Template Library (STL)、Boost C++	

Miyabi ソフトウェア構成 (2/2)




	汎用CPUノード群	演算加速ノード群
アプリケーション	OpenFOAM、ABINIT-MP、PHASE、FrontFlow/blue、FrontISTR、REVOCAP-Coupler、REVOCAP-Refiner、OpenMX、MODYLAS、GROMACS、BLAST、R packages、bioconductor、BioPerl、BioRuby、BWA、GATK、SAMtools、Quantum ESPRESSO、Xcrypt、ROOT、Geant4、LAMMPS、CP2K、NWChem、DeepVariant、Paraview、VisIt、POV-Ray、TensorFlow、PyTorch、JAX、Keras、Horovod、MXNet、Miniforge、Kokkos	
フリーソフトウェア	autoconf、automake、bash、bzip2、cvs、emacs、findutils、gawk、gdb、make、grep、gnuplot、gzip、less、m4、python、perl、ruby、screen、sed、subversion、tar、tcsh、tcl、vim、zsh、gitなど	
	Julia、CMake、Ninja、Java JDK	
	Grid Community Toolkit、Gfarm、FUSE	
コンテナ仮想化	Apptainer、Singularity Community Edition、	

64th TOP500 List (Nov, 2024)

R_{max}: Performance of Linpack (TFLOPS) <http://www.top500.org/>R_{peak}: Peak Performance (TFLOPS), Power: kW

	Site	Computer/Year Vendor	Cores	R _{max} (PFLOPS)	R _{peak} (PFLOPS)	GFLOPS/W	Power (kW)
1	<u>El Capitan, 2024, USA</u> DOE/NNSA/LLNL	HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS	11,039,616	1,742.00 (=1.742 EF)	2,746.38 63.4 %	58.99	29,581
2	<u>Frontier, 2021, USA</u> DOE/SC/Oak Ridge National Laboratory	HPE Cray EX235a, AMD Optimized 3 rd Gen. EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11	9,066,176	1.353.00	2,055.72 65.8 %	54.98	24,607
3	<u>Aurora, 2023, USA</u> DOE/SC/Argonne National Laboratory	HPE Cray EX - Intel Exascale Compute Blade, Xeon CPU Max 9470 52C 2.4GHz, Intel Data Center GPU Max, Slingshot-11, Intel	9,264,128	1,012.00	1,980.01 51.1 %	26.15	38,698
4	<u>Eagle, 2023, USA</u> Microsoft	Microsoft NDv5, Xeon Platinum 8480C 48C 2GHz, NVIDIA H100, NVIDIA Infiniband NDR	2,073,600	561.20	846.84 66.3 %		
5	<u>HPC 6, 2024, Italy</u> Eni S.p.A.	HPE Cray EX235a, AMD Optimized 3rd Generation EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11, RHEL 8.9	3,143,520	477.90	606.97 66.3 %	56.48	8,461
6	<u>Fugaku, 2020, Japan</u> R-CCS, RIKEN	Fujitsu PRIMEHPC FX1000, Fujitsu A64FX 48C 2.2GHz, Tofu-D	7,630,848	442.01	537.21 82.3 %	14.78	29,899
7	<u>Alps, 2024, Switzerland</u> Swiss Natl. SC Centre (CSCS)	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11	2,121,600	434.90	574.84 75.7 %	61.05	7,124
8	<u>LUMI, 2023, Finland</u> EuroHPC/CSC	HPE Cray EX235a, AMD Optimized 3 rd Gen. EPYC 64C 2GHz, AMD Instinct MI250X, Slingshot-11	2,752,704	379.70	531.51 71.4 %	53.43	7,107
9	<u>Leonard, 2023, Italy</u> EuroHPC/Cineca	BullSequana XH2000, Xeon Platinum 8358 32C 2.6GHz, NVIDIA A100 SXM4 64GB, Quad-rail NVIDIA HDR100	1,824,768	241.20	306.31 78.7 %	32.19	7,494
10	<u>Tuolumne, 2024, USA</u> DOE/NNSA/LLNL	HPE Cray EX255a, AMD 4th Gen EPYC 24C 1.8GHz, AMD Instinct MI300A, Slingshot-11, TOSS	1,161,216	208.10	288.88 72.0 %	61.45	3,387
13	<u>Venado, 2024, USA</u> DOE/NNSA/LANL	HPE Cray EX254n, NVIDIA Grace 72C 3.1GHz, NVIDIA GH200 Superchip, Slingshot-11	481,440	98.51	130.44 75.5 %	59.29	1,662
16	<u>CHIE-3, 2024, Japan</u> SoftBank, Corp.	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband NDR400, Ubuntu 22.04.4 LTS	163,200	91.94	138.32 66.5 %		
17	<u>CHIE-2, 2024, Japan</u> SoftBank, Corp.	NVIDIA DGX H100, Xeon Platinum 8480C 56C 2GHz, NVIDIA H100, Infiniband NDR400, Ubuntu 22.04.4 LTS	163,200	89.78	138.32 64.9 %		
18	<u>JETI, 2024, Germany</u> EuroHPC/FZJ	BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail NVIDIA InfiniBand NDR200, RedHat Linux, Modular OS	391,680	83.14	94.00 88.4 %	63.43	1,311
22	<u>CEA-HE, 2024, France</u> CEA	BullSequana XH3000, Grace Hopper Superchip 72C 3GHz, NVIDIA GH200 Superchip, Quad-Rail BXI v2, EVIDEN	389,232	64.32	103.48 62.2 %	52.17	1,233
28	<u>Miyabi-G, 2024, Japan</u> JCAHPC	Fujitsu, Supermicro ARS 111GL DNHR LCC, Grace Hopper Superchip 72C 3GHz, Infiniband NDR200, Rocky Linux	80,640	46.80	72.80 64.3 %	47.59	983
36	<u>TSUBAME 4.0, 2024, Japan</u> Institute of Science Tokyo	HPE Cray XD665, AMD EPYC 9654 96C 2.4GHz, NVIDIA H100 SXM5 94 GB, Infiniband NDR200	172,800	39.62	61.60 64.3 %	48.55	816
58	<u>Wisteria/BDEC-01 (Odyssey), 2021, Japan</u> U.Tokyo	Fujitsu PRIMEHPC FX1000, A64FX 48C 2.2GHz, Tofu D	368,640	22.12	25.95 85.2 %	15.07	1,468

Green 500 Ranking (Nov, 2024)

	TOP 500 Rank	System	Accelerator	Cores	HPL Rmax (Pflop/s)	Power (kW)	GFLOPS/W	Level
1	224	JEDI, EuroHPC/Julich, Germany	NVIDIA GH200	19,584	4.50	67	72.733	1
2	122	ROMEO-2025, ROMEO HPC Center - Champagne-Ardenne, France	NVIDIA GH200	47,328	9.86	160	70.912	1
3	442	Adastra2, GENCI-CINES, France	AMD Instinct MI300A	16,128	2.53	37	69.098	1
4	155	Isambard-AI phase1, U. Bristol, UK	NVIDIA GH200	34,272	7.42	117	68.835	1
5	51	Capella, TU Dresden ZIH, Germany	NVIDIA H100 94GB	85,248	24.06	445	68.053	3
6	18	JETI, EuroHPC/Julich, Germany	NVIDIA GH200	391,680	83.14	1,311	67.963	1
7	69	Helios GPU, Cyfronet, Poland	NVIDIA GH200	89,760	19.14	317	66.948	2
8	371	Henri, Flatiron Institute, USA	NVIDIA H100 80GB	8,288	2.88	44	65.396	?
9	340	HoreKa-Teal, KIT, Germany	NVIDIA H100 94GB SXM5	13,616	3.12	50	62.964	1
10	49	rzAdams, DoE LLNL, US	AMD Instinct MI300A	129,024	24.38	388	62.803	2
16	13	Venado, DoE LANL, US	NVIDIA GH200	481,440	98.51	1,662	59.287	1
30	36	TSUBAME 4.0, Science Tokyo	NVIDIA H100 94GB SXM5	172,800	39.62	816	48.565	3
33	28	Miyabi-G, JCAHPC	NVIDIA GH200	221,952	46.80	983	47.588	3
48	230	Pegasus, University of Tsukuba, Japan	NVIDIA H100 80GB	27,000	4.34	130	41.123	2
49	191	Wisteria/BDEC-01 (Aquarius), The University of Tokyo, Japan	NVIDIA A100 40GB	42,120	4.425	183	24.06	2

HPCG Ranking (Nov, 2024)

	Computer	Cores	HPL Rmax (Pflop/s)	TOP500 Rank	HPCG (Pflop/s)
1	Fugaku	7,630,848	442.01	4	16.00
2	Frontier	8,699,904	1,206.00	1	14.05
3	Aurora	9,264,128	1,012.00	2	5.61
4	LUMI	2,752,704	379.70	5	4.59
5	Alps	2,121,600	434.90	7	3.67
6	Leonardo	1,824,768	241.20	9	3.11
7	Perlmutter	888,832	79.23	19	1.91
8	Sierra	1,572,480	94.64	14	1.80
9	Selene	555,520	63.46	23	1.62
10	JUWELS Booster	449,228	44.12	33	1.28
12	AOBA-S	64,512	17.22	76	1.09
17	Wisteria/Odyssey	348,640	22.12	58	0.818
19	ES4-SX-Aurora Tsubasa	43,776	9.99	119	0.748
21	Miyabi-G	221,952	46.80	28	0.645

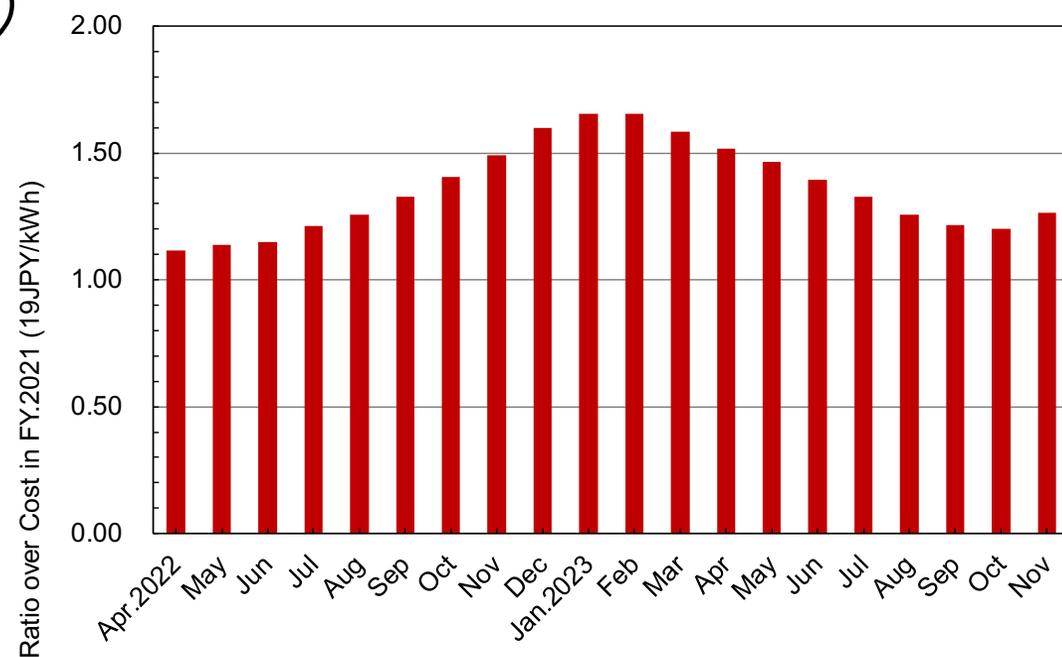
Miyabi (OFP-II) への道 (1/2)



- OFP (Oakforest-PACS) 後継機種 (OFP-II)
 - JCAHPC (筑波大学と共同)
 - 2025年1月運用開始予定
- スパコンへの性能要求, 省電力, 脱炭素化⇒演算加速器搭載は不可避 (電気代も高騰)
 - 2021年秋には方針決定



System (Top/Green 500)	HW	GF/W
Henri (255,1)	NVIDIA H100	65.4
Frontier (1,6)	AMD MI250X	52.6
Leonardo (4,15)	NVIDIA A100	32.2
Fugaku (2, 49)	A64FX	15.4



電力単価推移 (対2021年度比)

Miyabi (OFP-II) への道 (2/2)



- OFP-II (2025年1月運用開始)
 - 汎用CPUクラスタ(CPU-Group) + GPUクラスタ(Acc-Group)
 - 「計算・データ・学習」融合路線は継続
- GPUの選定は2022年6月に終了
 - OFPユーザー(3,000人以上)のGPUへの移行には18-30ヶ月必要
 - 7種類のベンチマーク
- NVIDIA社製GPU採用に決定(2022年6月)
 - H100もしくははその後継機
 - 決め手
 - 性能そのもの
 - Fortranで記述されたアプリケーションのポータビリティ
 - OpenACC/StdPar(Standard Parallelism)によるGPU化は比較的簡単, OpenMP/MPIハイブリッドによって並列化されたプログラムに適している

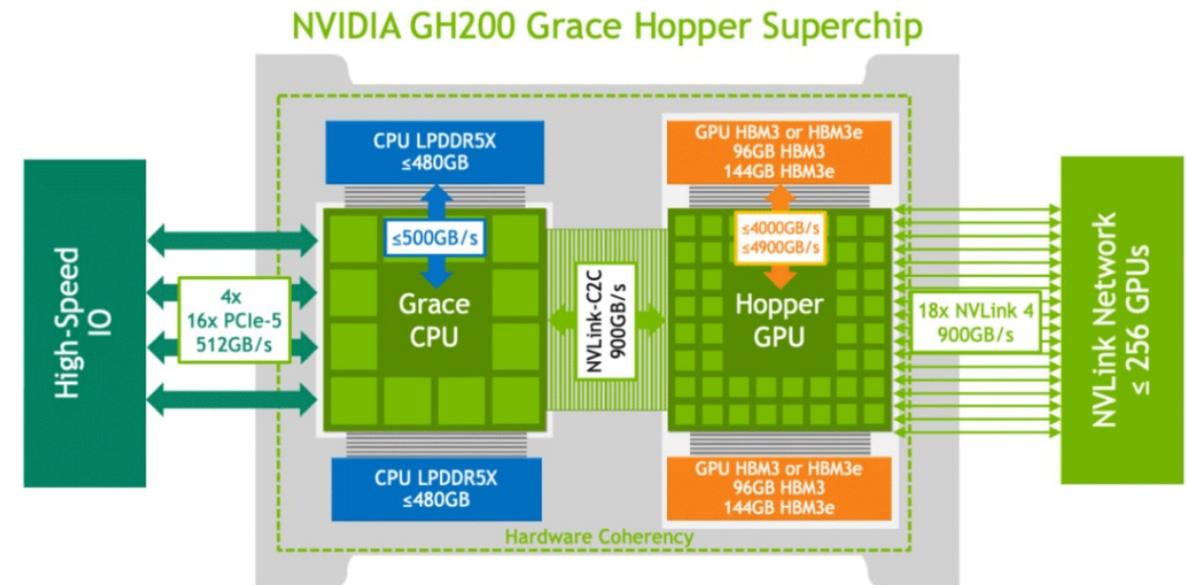
CPU-Group
CPU only

Acc-Group
CPU+GPU

	ノード年 GPU年 (千円)	ピークメモリバンド 幅 (GB/sec) GPUのみ*	Stream (Triad) (GB/sec)	定格消費電力 (W/node)	GF/W Green 500
Miyabi-G	300	4,022*	3,755	1,250	72.73 (JEDI, #1 in Green500, June 2024)
Miyabi-C	240	3,200	1,976	920	6.53 (Kyoto U., #97)
Odyssey	90	1,024	830	260	15.07
Aquarius	270	1,555*	1,377	679 (1GPU+0.25CP U)	24.06
OBCX	150	281.6	237.1	623	5.08
OFP	50	115.2 (DDR4) 921.6 (MC-DRAM)	90 490	215	4.99

GH200の特性

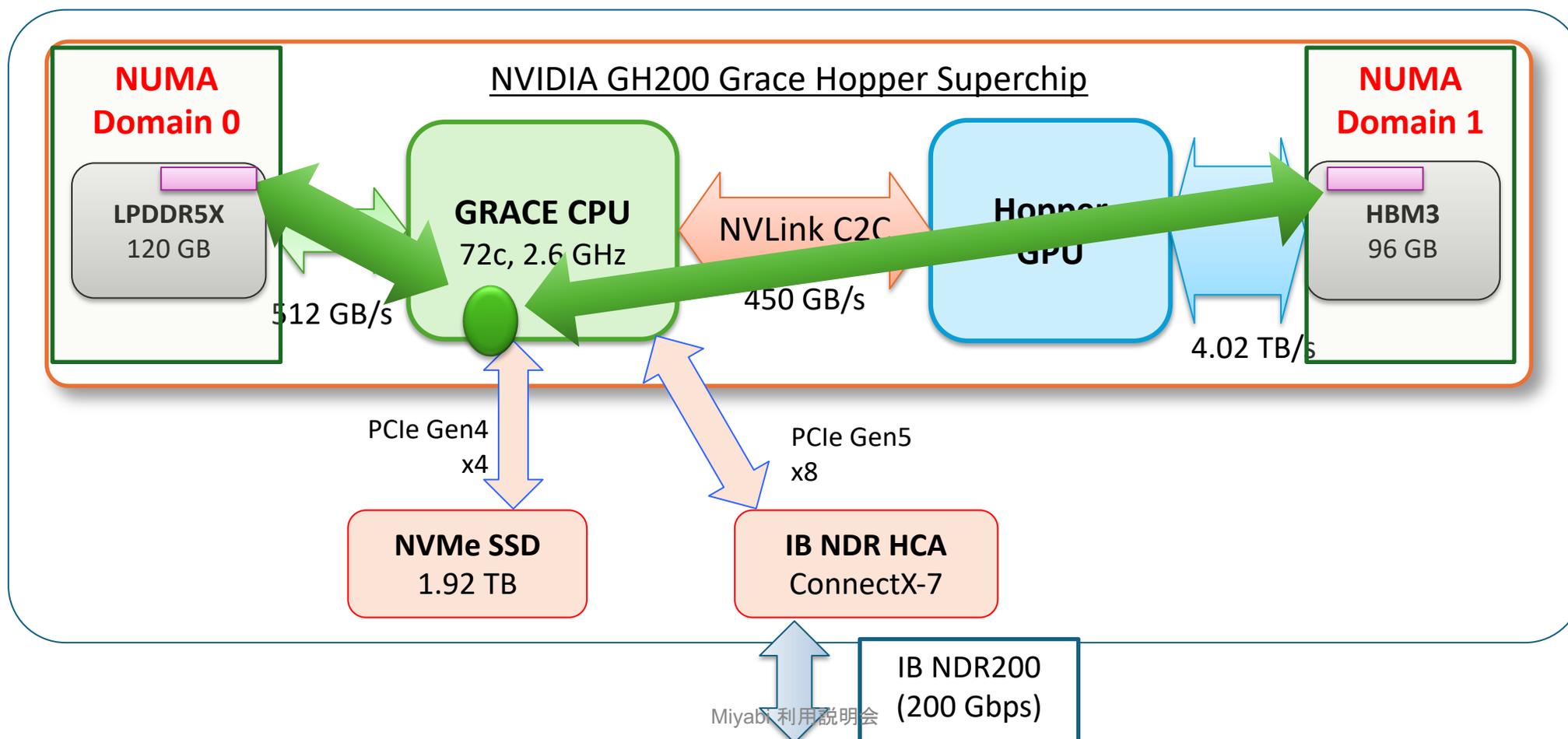
- Grace-Hopper相互のメモリ空間を直接参照可, NUMA的な扱い
- CPU-GPU間: コヒーレントインタフェース (NVLink-C2C, 450GB/sec/dir)
 - PCIe Gen 5の7倍以上の帯域
 - CPU-GPUの効率的使い分け可能
 - 従来は転送がボトルネック
 - プログラミングも容易
 - AMD MI300Aも同じ方向性



- 小規模問題, GPUが不得意な計算をCPUが柔軟に処理することも可能

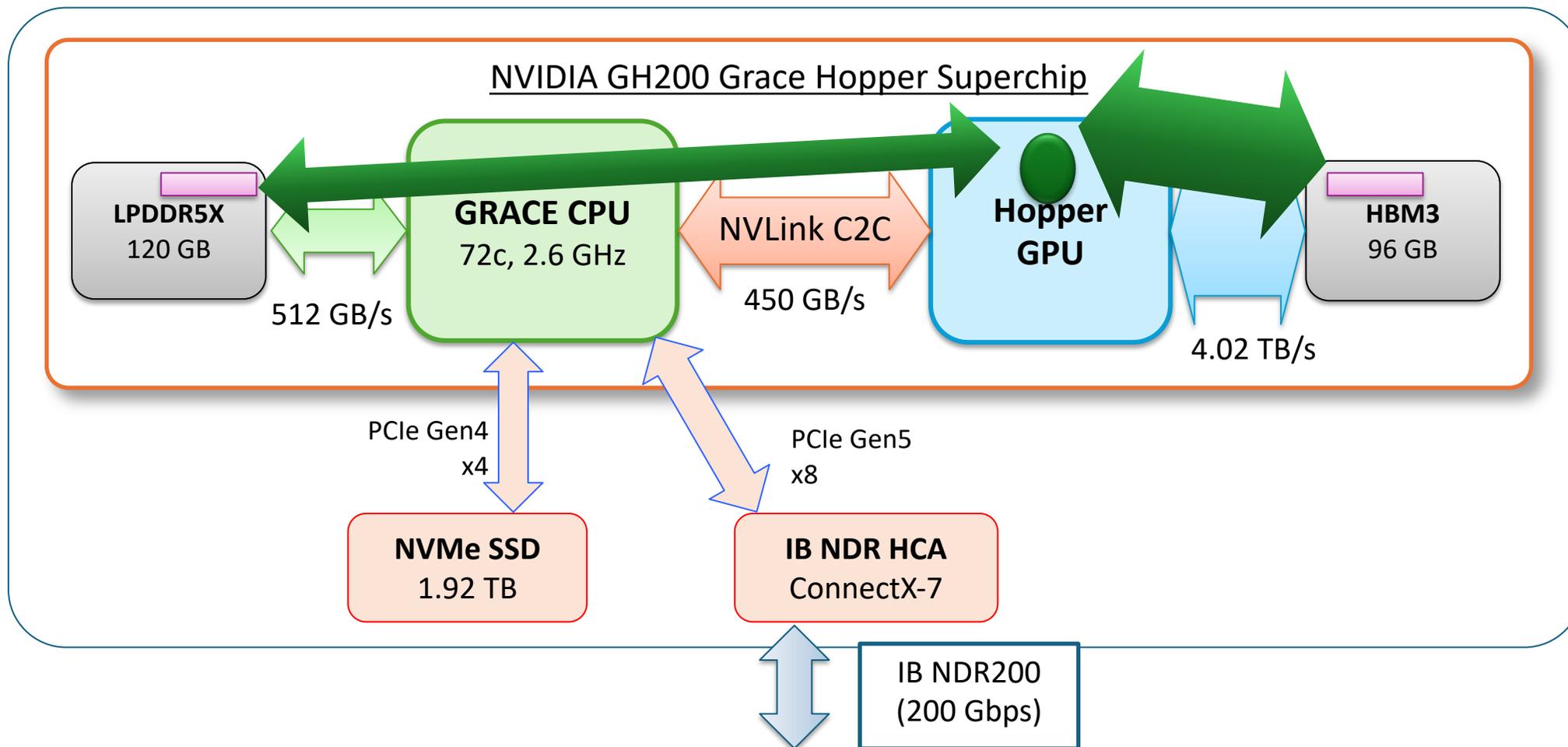
Graceからのメモリビュー

- NUMAとして見える
 - malloc()ではFirst Touchが大事
- 従来のCUDAのメモリモデルも使えて普通に動く + 転送が速い
 - cudaMalloc() + cudaMemcpy()
 - cudaMalloc() した領域はGraceからアクセス不可



Hopperからのメモリビュー

- Graceのアドレスを使って直接アクセス可能
- 従来のCUDAのコードはH100とほぼ挙動は変わらない（メモリBW比相当）



GH200のメモリアクセス

- CUDA

メモリ	メモリモード	確保される場所	Access-based Migration	CPUからアクセス	GPUからアクセス
System-allocated (malloc)	Unified相当	First-touch (GPU または CPU)	○	○	○
CUDA managed (cudaMallocManaged)	Managed相当	First-touch (GPU または CPU)	○	○	○
CUDA device memory (cudaMalloc)	Separate相当 (Device)	GPU			○
CUDA host memory (cudaMallocHost)		CPU		○	○

- OpenACC, OpenMP, stdpar

メモリモード	コンパイルフラグ	デフォルトとなる環境	
Separate	-gpu=mem:separate	OpenACC OpenMP	GPU上のデータはGPUからのみアクセス可能。 GPU-CPU間の明示的なデータ移動が必要。
Managed	-gpu=mem:managed	stdpar (Managed Memory のみの環境)	動的メモリ確保されたデータはGPU, CPU どちらからもアクセス可能。
Unified	-gpu=mem:unified	stdpar (Unified Memory 対応の環境)	全てのデータはGPU, CPU どちらからもアクセス可能。

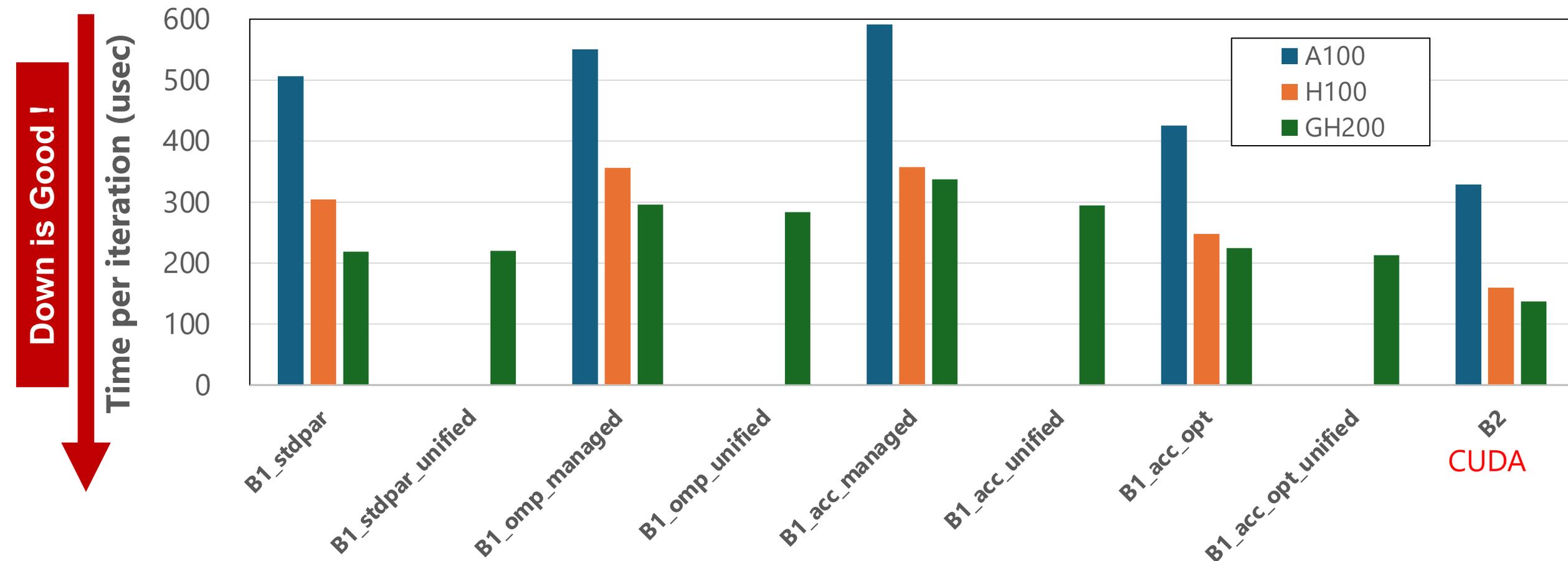
NVIDIAのGPUの歴史 (2017~)

	Memory (GB)	Memory Bandwidth (GB/sec)	FP64 (GFLOPS) 64bit 倍精度	FP32 (GFLOPS) 32bit 単精度	FP16 Tensor Core (GFLOPS) 16bit 半精度	TDP (Watt) 熱設計電力
P100	16	720	4,700	9,300	18,700	250
V100	32	1,134	8,200	16,400	130,000	300
A100	40	1,555 (1.00)	9,700	19,500	312,000	400
H100	80	3,352 (2.16)	33,500	66,900	1,979,000	700
GH200 GPU	96	4,022 (2.59)	33,500 1	66,900 x2	1,979,000 x60	~1,000 (CPU含む)

Poisson 3D: small (128x128x128)

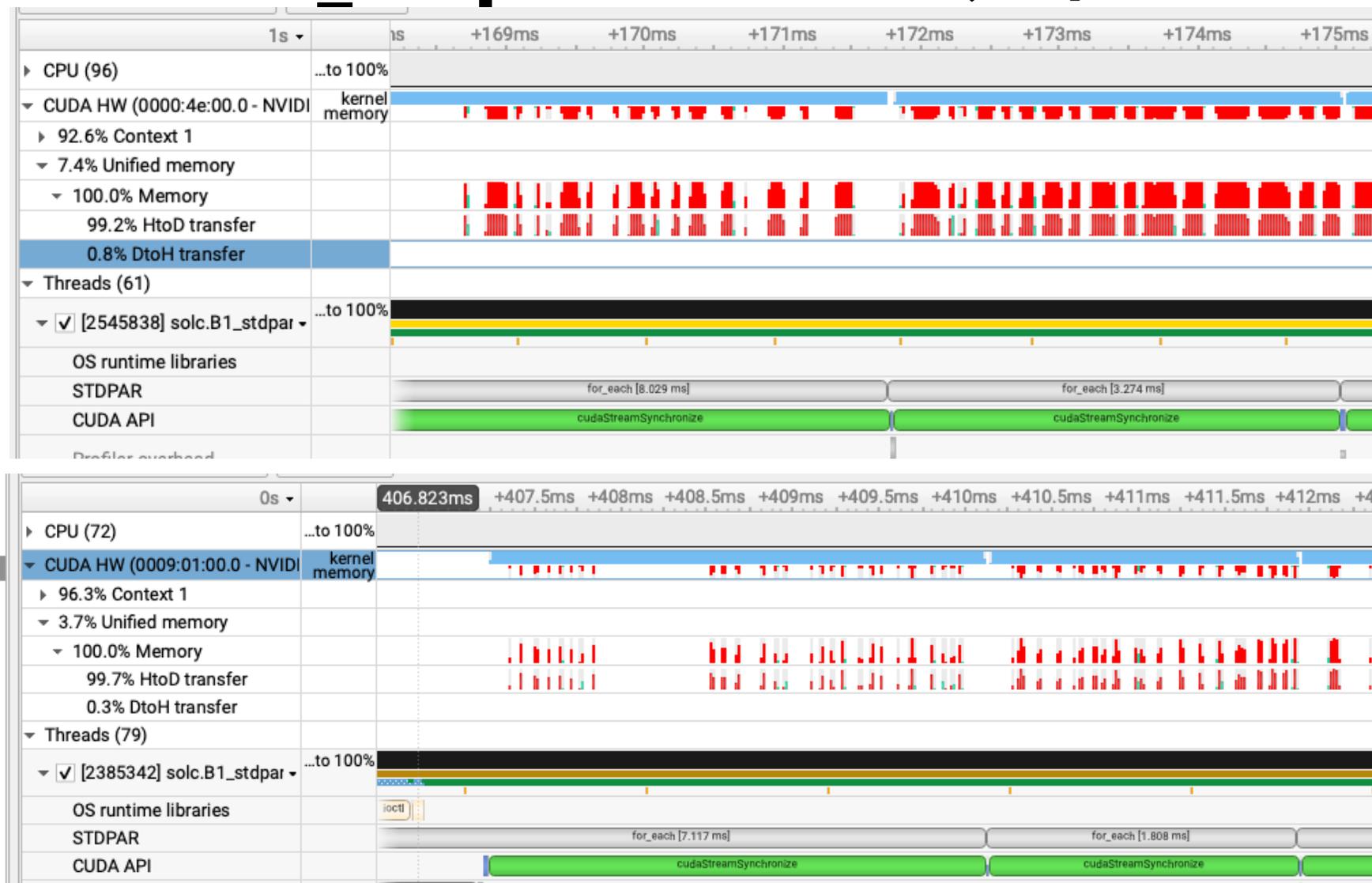
[基他, 第195回HPC研究会,
2024]

	メモリバンド幅 (GB/sec)	FP64 (GFLOPS)
A100	1,555 (1.00)	9,700
H100	3,352 (2.16)	33,500
GH200 : GPU	4,022 (2.59)	33,500



Poisson 3D small B1_stdparのプロファイル

- 上: H100 SXM5
下: GH200
 - 赤: データ転送
- データ転送の総量はさほど変わらない
- NVLink-C2C vs PCIeは約7~10倍の性能
- 64Kページ@GH200の効果
 - ページフォルトは1/4
(ページフォルトをトリガにしてデータ転送が起こる)



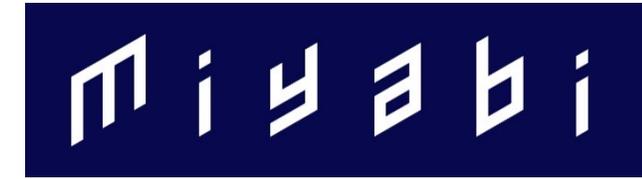
Miyabiによって何をするか，何をできるか？

- 中心は「計算科学」，「シミュレーション」
 - GH200の特性を活用すれば，より複雑な処理を効率的にかつ容易に実施できそう
- 「計算・データ・学習」融合の更なる促進
- 更に進めて「AI for Science」の推進
 - AIを活用した科学研究の革新
 - 従来の「AI for Science」: Data Driven Science
 - 生成AI・基盤モデル
 - LLM等に使用される「基盤モデル」の「トークン」の入れ替え，学習によって様々な分野へ適用可能
 - ANLの「Stormer (Short-Term Weather Forecasting)」
 - <https://www.anl.gov/cels/development-of-predictive-models-shortterm-forecasting>
- 「量子コンピューティング」との融合（次頁）
 - 2025年度中には，Miyabi, Wisteria/BDEC-01とJHPC-quantumの量子コンピュータ群によるハイブリッド環境での利用が可能となる

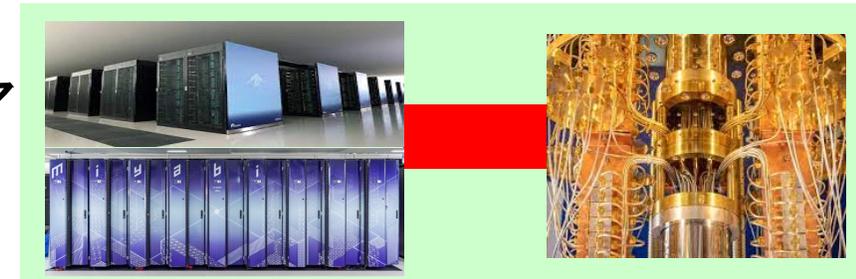
量子・スパコン連携プラットフォーム (JHPC-Quantum)

<http://jhpc-quantum.org/>

- JHPC-quantum (FY.2023-FY.2028)
 - 理研R-CCS, ソフトバンク, 東大, 阪大
 - NEDOポスト5G関連事業, 産業応用への期待
- 2基のリアルな量子コンピュータを導入
 - IBM (超伝導型), 100+Qubit, 理研神戸
 - Quantinuum (Ion-Trap型), 20+Qubit, 理研和光
- アプリケーション
 - 量子物理, 誤り訂正, 量子機械学習 + シミュレーション
- 東大の役割
 - QC-HPCハイブリッド連携環境向けシステムソフトウェア
 - QC as Accelerators
 - h3-Open-BDECの拡張
- 2025年度末: 富岳・東大スパコンとQCの連携開始



CPU		GPU			Others	
A64FX Arm	X86	NVIDIA Intel AMD Arm	AMD	Intel	Sambanova Cerebras Graphcore etc.	Quantum
h3-Open-SYS/WaitIO						



本日のスケジュール

14:00-14:40 利用説明

- 最先端共同HPC基盤施設 (JCAHPC)の紹介
- Miyabiの概要
- **Miyabi利用方法の紹介**
 - リソースグループ、MIG
- Miyabi利用に向けたユーザーサポート
 - 講習会、GPUミニキャンプ、相談会
- Miyabi利用支援ポータル

14:40-14:55 チュートリアル

15:00-15:25 筑波大学の利用制度説明 or Miyabi見学(その1)

15:30-15:55 東京大学の利用制度説明 or Miyabi見学(その2)

16:00-16:30 Miyabi見学(その3)

Miyabiの運用

- 基本的にはこれまでのスパコンの運用を踏襲
 - ジョブクラス
 - <https://www.cc.u-tokyo.ac.jp/supercomputer/miyabi/service/job.php>
 - OpenOnDemandによる利用についても検討中
 - <https://openondemand.org/>
 - Wisteria/BDEC-01ではテスト済み
 - 大規模HPCチャレンジ実施, ノード固定は実施しない
 - Miyabi-GではGPUを分割して利用するMIG (Multi-Instance GPU) の使用を可能とする
 - 4分割, 負担金はGPUの4分の1, CPU (Grace) の資源も4分の1割り当てる
 - 講習会, 講義, デバッグなど少量の資源で良い場合に活用
- Miyabi-G (NVIDIA GH200), Miyabi-C (Intel Max) は分離して運用するが, 技術的に連携して動かすことは可能 (WaitIO等の利用)
 - 必要に応じて, 使えるようにする



MIG (Multi-Instance GPU)

- GPUリソース (SM (演算コア), メモリ) を複数のインスタンスに分割する機能
 - <https://www.nvidia.com/ja-jp/technologies/multi-instance-gpu/>
- 想定される活用シーン
 - 1基のNVIDIA GH200の演算リソースを使い切れない場合
 - 分割されたGPUを使えば, トークン消費量を抑えられる
 - 実装・開発中のコードのデバッグ・動作テスト・機能テスト
 - (見かけの) GPU数が増えるので, ジョブ投入後の待ち時間が短縮される
 - 教育利用 (主にGPUプログラミングの初心者・初級者向け)
 - (見かけの) GPU数が増えるので, 多数の受講者のジョブが同時に実行される
- Miyabi-GでのMIG利用形態 (NVIDIA GH200を4分割)
 - MIG利用キュー: debug-mig, short-mig, regular-mig

- トークン消費量は通常のノード占有キュー (debug-g, short-g, regular-g など) の 1/4

	MIG#0	MIG#1	MIG#2	MIG#3
GPUリソース	32 SMs, 24 GB	32 SMs, 24 GB	32 SMs, 24 GB	26 SMs, 24 GB
CPUリソース	18コア, 25 GiB	18コア, 25 GiB	18コア, 25 GiB	18コア, 25 GiB

バッチジョブ: Miyabi-G(演算加速ノード) 1,120ノード

キュー名	ノード数	制限時間 (Elapsed)	メモリ容量 (GiB)
debug-g	1 ~ 16	30 分	100
short-g	1 ~ 8	8 時間	100
regular-g			
(small-g)	1 ~ 16	48 時間	100
(medium-g)	17 ~ 64	//	//
(large-g)	65 ~ 128	//	//
(x-large-g)	129 ~ 256	24 時間	//

キュー名	MIG数	制限時間 (Elapsed)	メモリ容量 (GiB)
debug-mig	1, 2, 4	30 分	25
short-mig	1, 2, 4	8 時間	25
regular-mig	1, 2, 4	48 時間	25

バッチジョブ: Miyabi-C(汎用CPUノード) 190ノード

キュー名	ノード数	制限時間 (Elapsed)	メモリ容 量 (GiB)
debug-c	1 ~ 4	30 分	118
short-c	1 ~ 2	8 時間	118
regular-c			
(small-c)	1 ~ 16	48 時間	118
(medium-c)	17 ~ 32	//	//
(large-c)	33 ~ 64	24 時間	//

インタラクティブジョブ（トークン消費無し）

- Miyabi-G (演算加速ノード群)

キュー名	ノード数 MIG数	制限時間 (Elapsed)	メモリ容量 (GiB)
interact-g (interact-g_n1)	1 ノード	2時間	100
(interact-g_n8)	2 ~8 ノード	10分	100
interact-mig (interact-mig_n1)	1 MIG	2時間	25

- Miyabi-C (汎用CPUノード群)

キュー名	ノード数	制限時間 (Elapsed)	メモリ容量 (GiB)
interact-c (interact-c_n1)	1	2時間	118
(interact-c_n2)	2 ~2	10分	118

教育利用、講習会: Miyabi-G (演算加速ノード)

キュー名	ノード数	制限時間 (Elapsed)	メモリ容量 (GiB)
lecture-g (教育利用全体で共有)	1 ~ 4	15 分	100
tutorial-g, lecture0-g, lecture1-g, ... (時間中のみ使用可能)	1 ~ 4	15 分	100

キュー名	MIG数	制限時間 (Elapsed)	メモリ容量 (GiB)
lecture-mig (教育利用全体で共有)	1, 2, 4	15 分	25
tutorial-mig, lecture0-mig, lecture1-mig, ... (時間中のみ使用可能)	1, 2, 4	15 分	25

教育利用、講習会: Miyabi-C (汎用CPUノード)

キュー名	ノード数	制限時間 (Elapsed)	メモリ容量 (GiB)
lecture-c (教育利用全体で共有)	1 ~ 2	15 分	118
tutorial-c, lecture0-c, lecture1-c, ... (時間中のみ使用可能)	1 ~ 2	15 分	118

本日のスケジュール

14:00-14:40 利用説明

- 最先端共同HPC基盤施設 (JCAHPC)の紹介
- Miyabiの概要
- Miyabi利用方法の紹介
 - リソースグループ、MIG
- Miyabi利用に向けたユーザーサポート
 - 講習会、GPUミニキャンプ、相談会
- Miyabi利用支援ポータル

14:40-14:55 チュートリアル

15:00-15:25 筑波大学の利用制度説明 or Miyabi見学(その1)

15:30-15:55 東京大学の利用制度説明 or Miyabi見学(その2)

16:00-16:30 Miyabi見学(その3)

Miyabi利用に向けたユーザーサポート

- GPU移行
 - 基本的には各自によるポーティング
- 利用手引書
 - 利用支援ポータル(後述)の「ドキュメント閲覧」にあります
- ユーザーサポート
 - 講習会、ミニキャンプ、ハッカソン(後述)
 - 相談会(毎月1回)(後述)
 - 移行ポータルサイト
 - GPUへ移植するために、役立つ情報を集約したサイト
 - 移行方法の特徴のまとめ
 - https://www.cc.u-tokyo.ac.jp/supercomputer/gpu_porting.php

お試しアカウント付き並列プログラミング講習会

<https://www.cc.u-tokyo.ac.jp/events/lectures/>

<https://www.cc.u-tokyo.ac.jp/events/lectures/schedule.php>

- Miyabi, Wisteria/BDEC-01を利用した講習会
- 2020年度から完全オンラインで実施：年20回程度，概して好評（移動不要）
- 既存利用者に限定せず，企業の技術者・研究者も受講可能
 - ✓ 受講者の1/2~2/3は企業から受講：裾野拡大に大きな貢献
 - ✓ PCクラスタコンソーシアム（実用アプリケーション部会）と共催
- 1~2日間の講習，1ヶ月有効な「お試しアカウント」
 - ✓ スパコン超入門：Linuxの使い方
 - ✓ MPI基礎，MPI応用（並列有限要素法）（4日間），マルチコアプログラミング
 - ✓ GPUプログラミング，GPUミニキャンプ（ハッカソン）
 - ✓ ライブラリ利用（センター教員開発のライブラリ普及）
 - ✓ OpenFOAM（初級，中級），Altair HyperWorks，MATLAB
 - ✓ 利用ノード数，実行時間に制限あり，Miyabi 利用説明会 スパコンを使用しない講義もある

GPUハッカソン・ミニキャンプ

- ハッカソン・ミニキャンプとは？

- 参加者がコードやデータセットを持ち込み、CUDA、OpenACC、Deep Learning など、GPUに関連した課題に対して、メンターからの助言を受けながら、その課題解決に取り組みます。
- ミニキャンプは8日間、ハッカソンは15日間（講習会実施日はそれぞれ2日、3日）
- ハイブリッドまたはオンライン

- メンター

- 課題解決に協力
- JCAHPC（筑波大、東大）、北大、科学大、名大、九州大、NVIDIA、富士通、プロメテック・ソフトウェアなどから参加

- 次回「JCAHPC Open Hackathon」（Miyabiを利用）

- 2月3日（月）-17日（月）でハイブリッド実施（各週（月）が講習会日、中日は各チームで実践）
- 1月31日（金）にプログラミング、プロファイラ講座
- 募集チーム数：約10チーム。応募チーム多数のため、申込を締め切りました。
- <https://www.cc.u-tokyo.ac.jp/events/lectures/239/>



GPU移行相談会

- GPU移行などGPU関連の疑問を相談できる会
- チューター
 - JCAHPC(筑波大、東大)、NVIDIAなどの研究者や技術者
- 定期的に開催(毎月一回程度)
- オンライン開催、事前申込制、参加費無料
- 相談事例:
 - 既存のCPUコードをGPU化する適切な方法がわからない。
 - GPU向けライブラリを紹介してほしい。
 - 性能を上げるためのプロファイラの利用方法を聞きたい。
 - 複数のGPUを利用した計算方法について相談したい。
 - そもそもGPU計算がよくわからない。
- 次回
 - 1月24日(金) 13:00-14:00

不具合やご要望などありましたら、ご連絡ください。

- 1月14日(火)より正式運用を開始しました。
- 不具合やご要望などございましたら、下記よりお問い合わせいただければ幸いです。
 - <https://www.cc.u-tokyo.ac.jp/supports/contact/index.php>

本日のスケジュール

14:00-14:40 利用説明

- 最先端共同HPC基盤施設 (JCAHPC)の紹介
- Miyabiの概要
- Miyabi利用方法の紹介
 - リソースグループ、MIG
- Miyabi利用に向けたユーザーサポート
 - 講習会、GPUミニキャンプ、相談会
- **Miyabi利用支援ポータル**

14:40-14:55 チュートリアル

15:00-15:25 筑波大学の利用制度説明 or Miyabi見学(その1)

15:30-15:55 東京大学の利用制度説明 or Miyabi見学(その2)

16:00-16:30 Miyabi見学(その3)