

スーパーコンピュータ「不老」 導入 と 産業利用の新展開

名古屋大学情報基盤センター 片桐孝洋



概要

1. スーパーコンピュータ「不老」導入
2. 名古屋大学情報基盤センターの
民間利用制度
3. センター独自開発ソフトウェア
4. おわりに

概要

1. スーパーコンピュータ「不老」導入
2. 名古屋大学情報基盤センターの
民間利用制度
3. センター独自開発ソフトウェア
4. おわりに

スーパーコンピュータ「不老」導入

スーパーコンピュータ「不老」の役割

1. 全国共同利用・共同研究拠点として学内外へ計算資源提供

- ▶ 全国共同利用・共同研究拠点として国が位置づけ
 - ▶ 全国の研究の世界トップレベル研究を強力に支援

世界トップレベル
研究の支援

国策スパコン
利用支援

HPCI (High Performance
Computing Infrastructure) 利用者

JHPCN利用者

名大拠点利用者

名大「不老」
Type I システム
(富岳型ノード)



簡便な
移行支援

スーパー
コンピュータ
「富岳」

導入支援／高性
能化／特殊処理
／長時間実行

世界トップレベル
研究成果創出

2. ものづくり企業支援(地域イノベーションコア形成)

- ▶ 産業利用制度(公開、非公開)
- ▶ 計算機利用型講習会による並列処理・大規模計算普及(地域特有の中小企業支援)

3. 新しい計算需要に向けたサービス開拓

- ▶ データサイエンス(ビッグデータ)、AI基盤の提供による新サービスの開拓

4. 指定国立大学として重要な役割

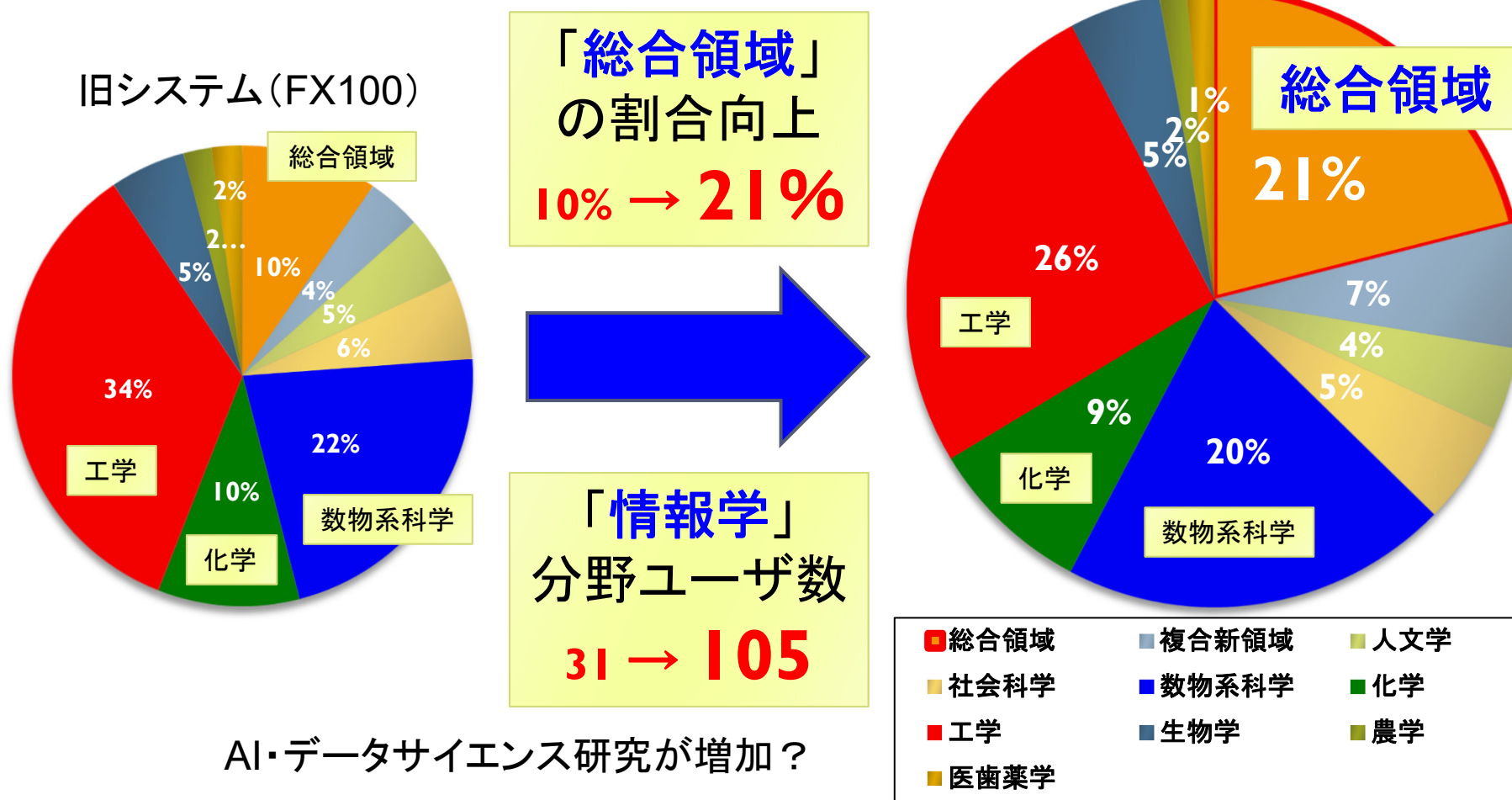
- ▶ 数理・データ科学教育
- ▶ 人材育成・研究力強化・社会との連携

- ・数理データ科学分野の人材育成
- ・AI技術基盤提供(大規模AI計算基盤、大規模ストレージ、サイバーフィジカル)
- ・技術コンサルティング...等による社会貢献

ユーザの研究分野の変遷

※2020年10月現在

スーパーコンピュータ「不老」



スーパーコンピュータ「不老」の特徴

1. 数値計算・データサイエンス融合スパコン
 - ▶ 超並列型のType I サブシステム と AI用Type II サブシステム
 - ▶ 双方でシームレスなデータ移動を実現(共有ファイルシステム)
 - ▶ 大規模・堅牢なデータ蓄積(コールドストレージ)
2. オープンソースソフトウェア(OSS)の活用推進
 - ▶ オープンソース・国費開発ソフトウェアをプリインストール
 - ▶ 持続性(ライセンス無料、ユーザのライセンス費用削減)
3. 民間利用の推進
 - ▶ 公開制度・非公開制度による社会貢献
4. 国策スパコンへのシームレスな移行支援

スーパーコンピュータ「富岳」との連携

1. 同型計算機・同一ソフトウェアスタック

- ▶ 「不老」Type I サブシステムはCPUが「富岳」と同型
- ▶ OS・コンパイラ等のソフトウェアスタックも同一と想定
 - ▶ ☞「不老」のほうがより進んだバージョンが提供される可能性があります

2. 国策ソフトウェアの「不老」Type I サブシステムの プリインストール・講習会実施

- ▶ 国費開発ソフトウェアで「富岳」で動作するソフトウェアのいくつかを、一般財団法人 高度情報科学技術研究機構(RIST)の協力のもと、「不老」Type I で提供、ハンズオン講習会も実施
 - ▶ ☞名古屋大学情報基盤センターHPの「講習会予定」をご覧ください

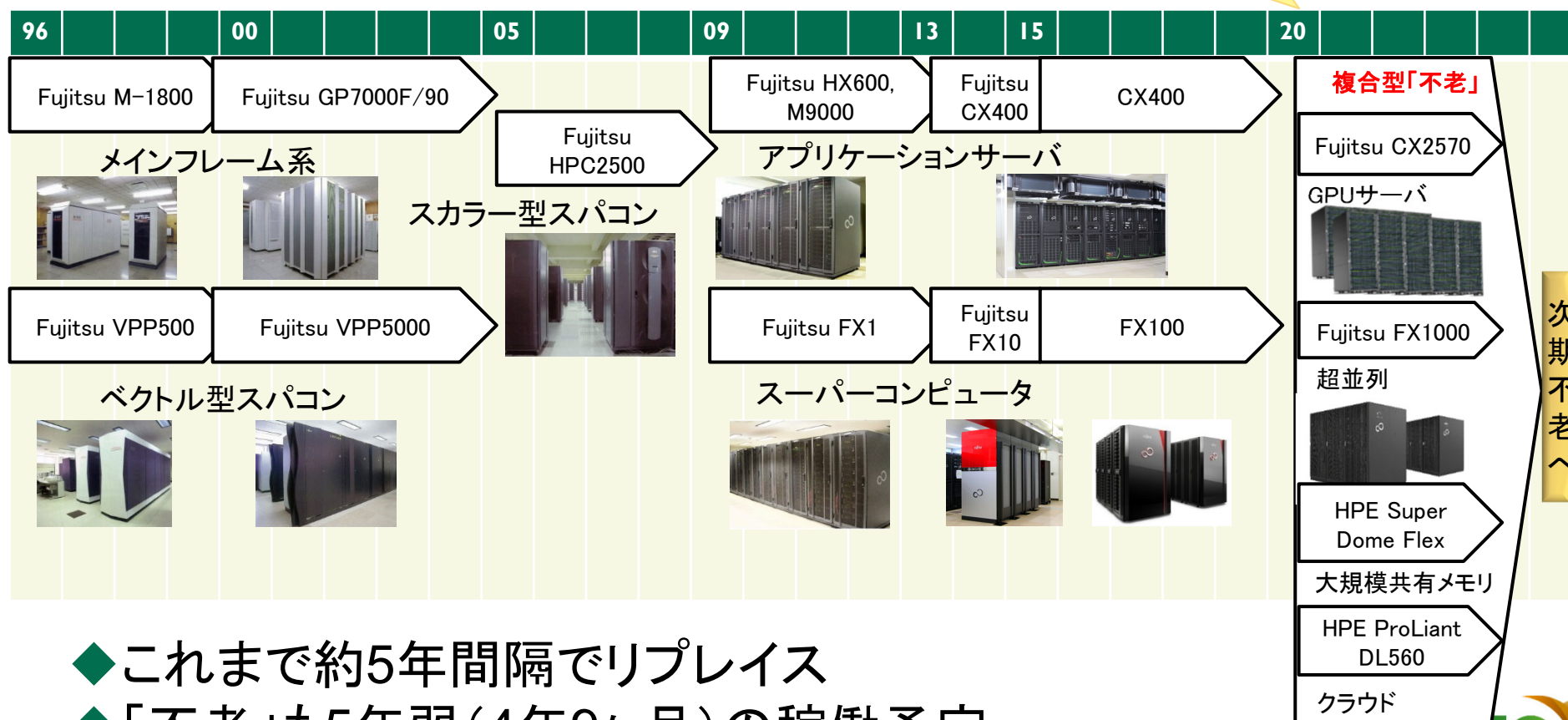
3. 「富岳」向けチューニングと「富岳」への移行支援

- ▶ コンサルティング・教員との共同研究で「富岳」向けと想定されるコードチューニング(☞「不老」Type I 向けと等価)を支援します



名古屋大学情報基盤センターの スパコンの歴史

スーパーコンピュータ
「不老」導入



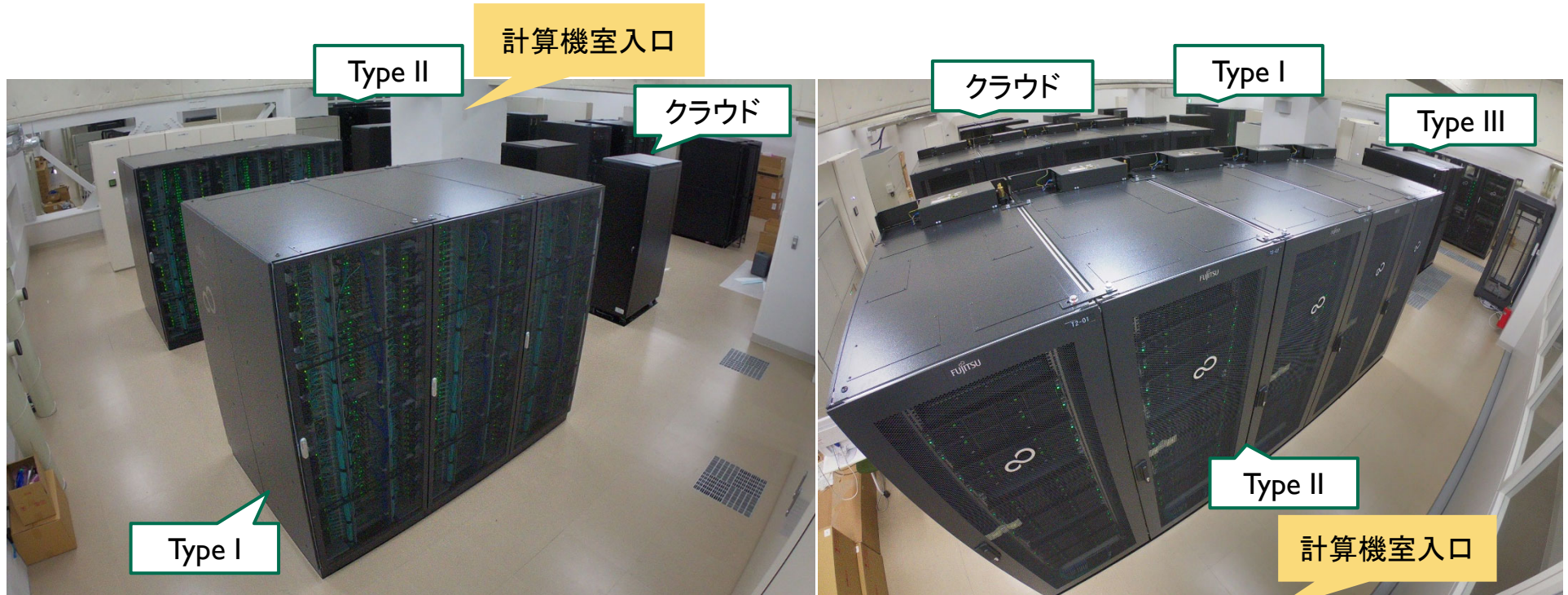
- ◆これまで約5年間隔でリプレイス
- ◆「不老」も5年弱(4年9ヶ月)の稼働予定

導入の背景

- ▶ **研究のデジタル化 (デジタルサイエンス)**
 - ▶ コンピューティングを活用した研究の広まり
- ▶ **AI/機械学習研究の増大**
 - ▶ 自動運転、医療、創薬
- ▶ **シミュレーション研究の増加**
 - ▶ 異常気象、津波など国民の安全に密接にかかわる現象
 - ▶ 生命・宇宙などの基礎科学
- ▶ **データの爆発的増大**
 - ▶ 元データ、解析結果、AI学習結果など
- ▶ **従来のスパコンでは明らかな能力不足**

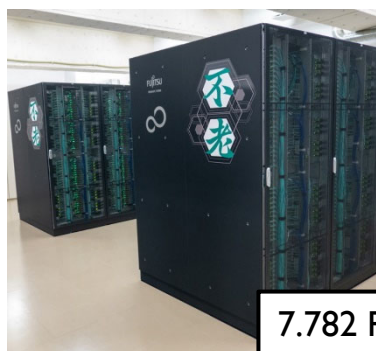
設置状況

- ▶ 2020年7月1日、スーパーコンピュータ「不老」が稼働開始しました。現在も順調に稼働中です。
- ▶ 名古屋大学 情報基盤センター 本館地下1階の様子



実際に入ったもの（主な構成要素）

**Type I, II, III, クラウドの合計で15.8PFLOPS
(旧システムの約4倍)**



7.782 PF

Type Iサブシステム

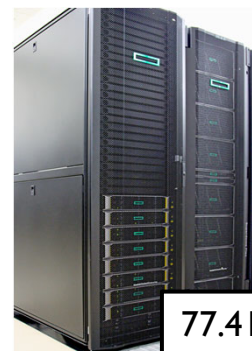
FUJITSU Supercomputer FX1000
「富岳」型



7.489 PF

Type IIサブシステム

FUJITSU Server PRIMERGY CX2570 M5
GPUスパコン



77.414 TF

Type IIIサブシステム

HPE Superdome Flex
大容量メモリ・可視化



537.6 TF

クラウドシステム

HPE ProLiant DL560
バッチ & インタラクティブ



30 PB

ホットストレージ

FUJITSU PRIMERGY RX2540 M5
FUJITSU ETERNUS AF250 S2
DDN SFA18KE
DDN SS9012

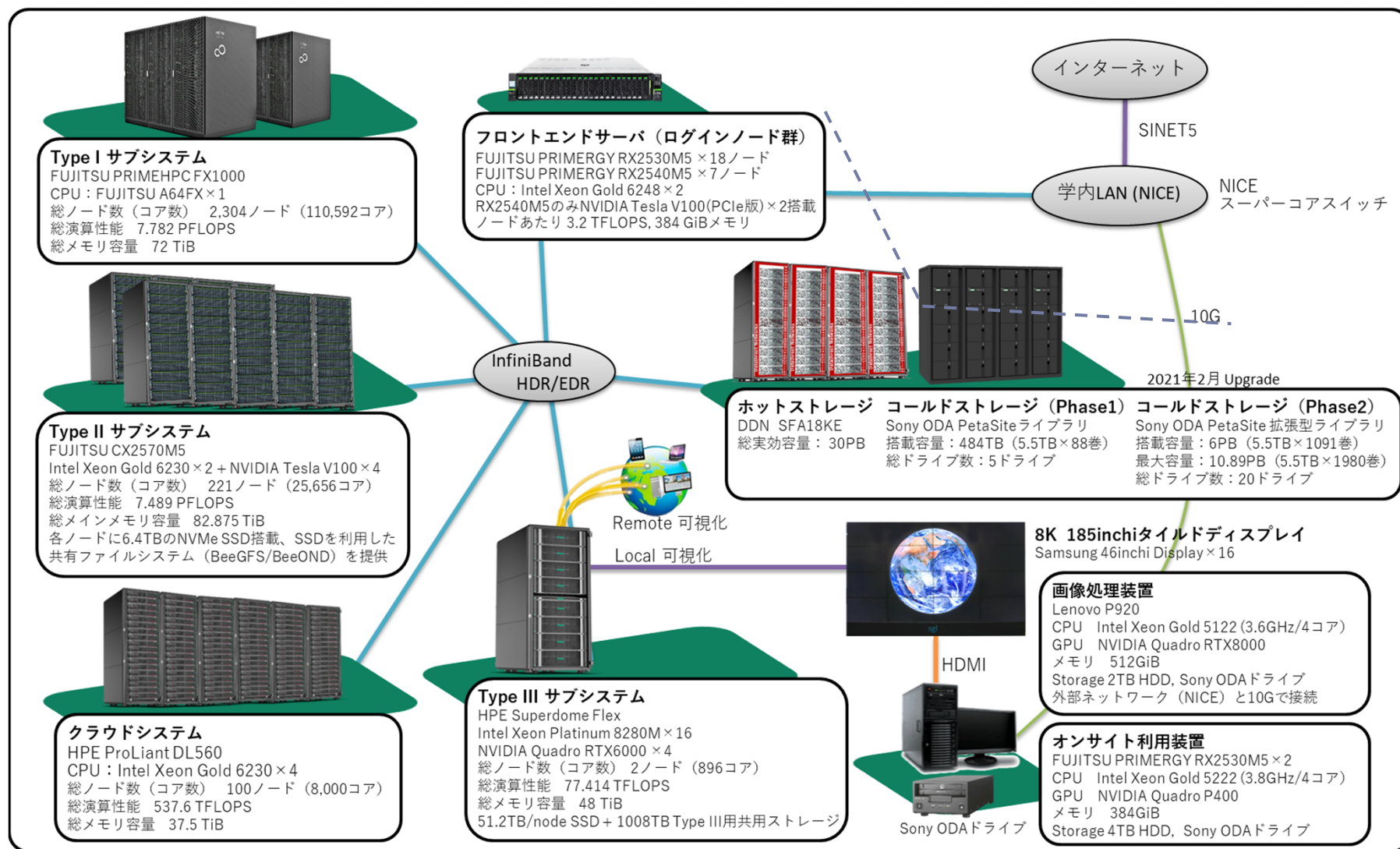


484 TB → 6 PB

コールドストレージ

SONY PetaSite 拡張型 Library
↓ 2020年2月更新
SONY PetaSite 拡張型 Library

全体システム構成

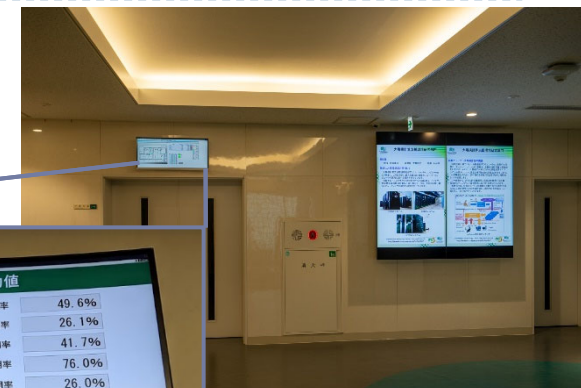
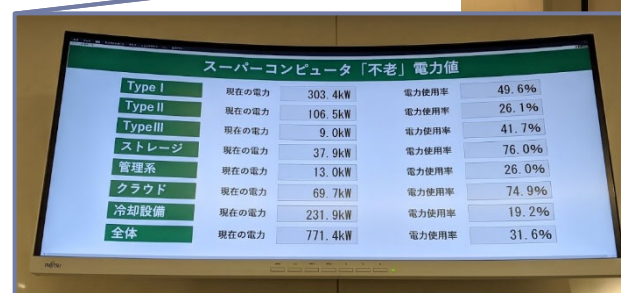


消費電力・省電力対策

▶ 最大消費電力

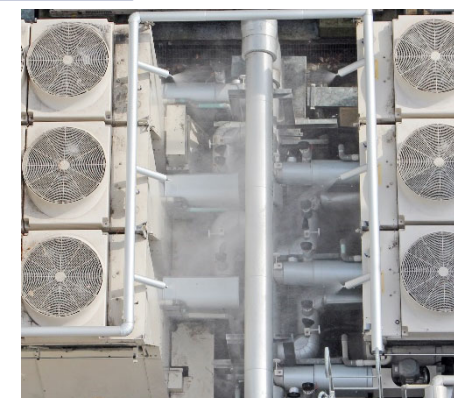
サブシステム名	消費電力
TypeIサブシステム	628.1kVA
TypeIIサブシステム	393.5kVA
TypeIIIサブシステム	21.6kVA
クラウドシステム	93.0kVA
ストレージ	49.9kVA
フロントエンド	19.6kVA
運用管理システム他	52.3kVA
冷却設備	641.9kVA
合 計	1,899.9kVA

▶ 電力可視化



▶ 湧水を用いた冷却

- ▶ 地下の湧水を活用したら総合評価時加点
- ▶ 屋外チラーに散水して冷却

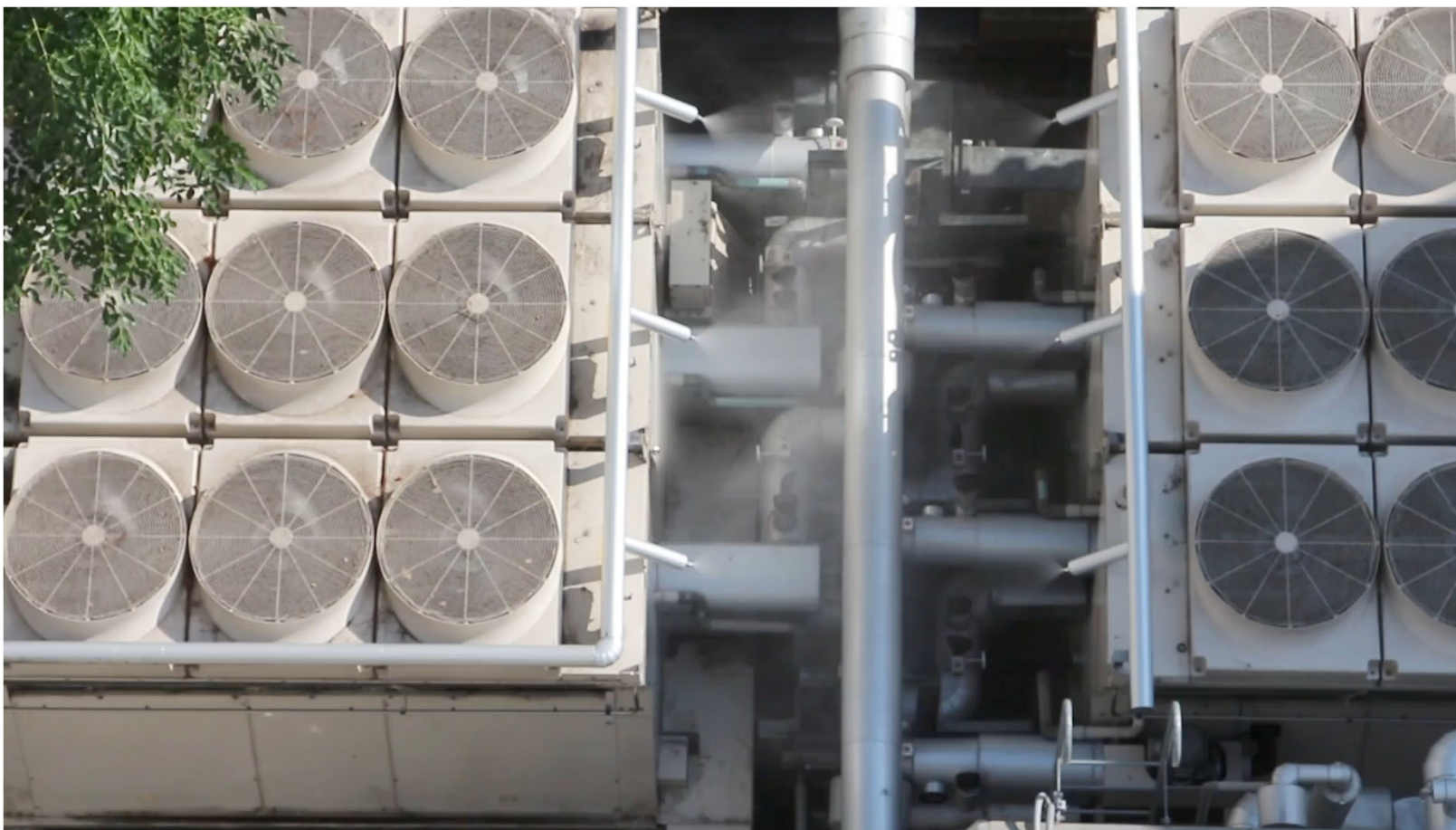


湧水による冷却システム

- ▶ 情報基盤センターの地下は
夏季でも18°C程度の
湧き水が毎分30L程度湧く
- ▶ この湧き水は、地下からポンプで吸い上げて雨水扱いで捨てていた
- ▶ 今回の仕様で、湧き水を冷熱源として使用する場合は加点
 - ▶ 冷却水としての利用許可・水質検査済み

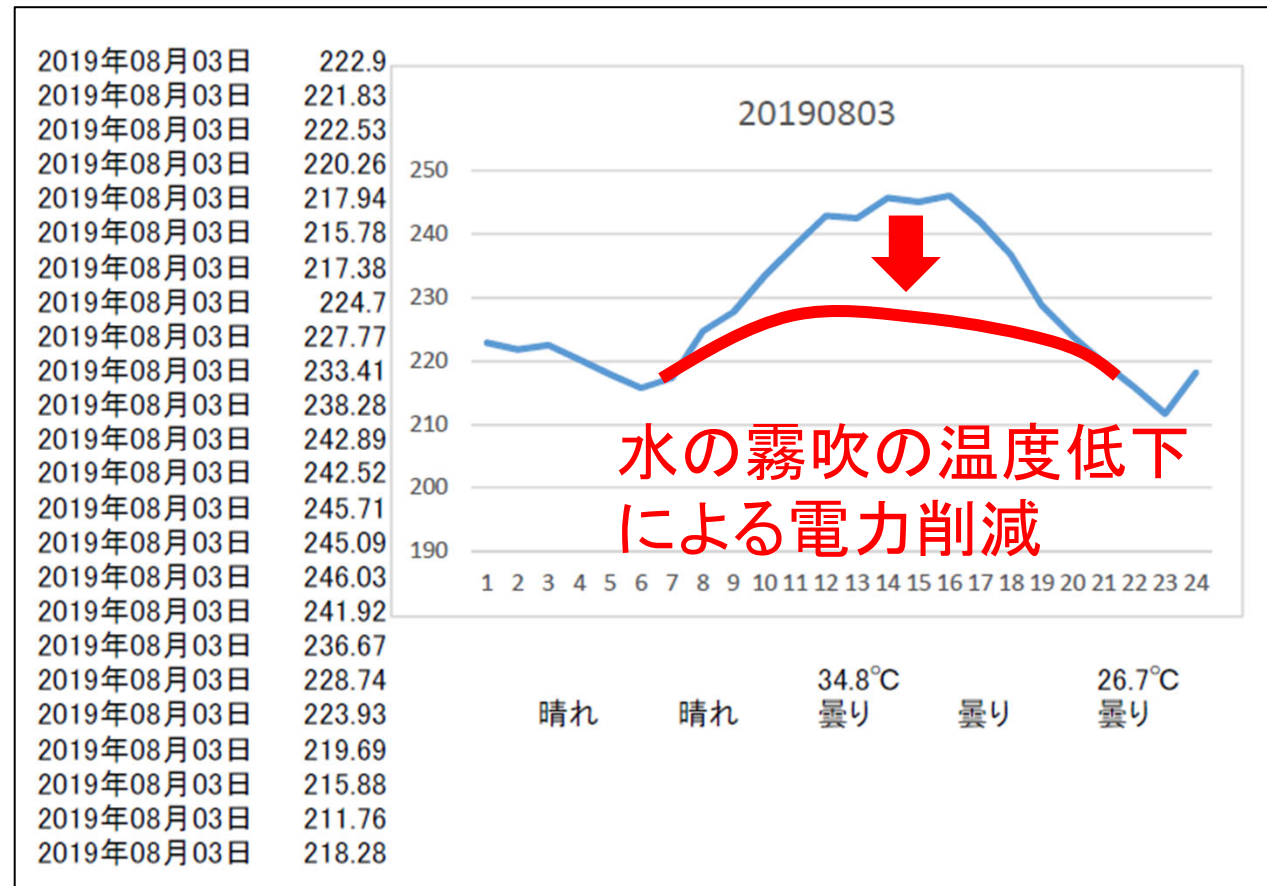


湧水による冷却システム



湧水による冷却システム

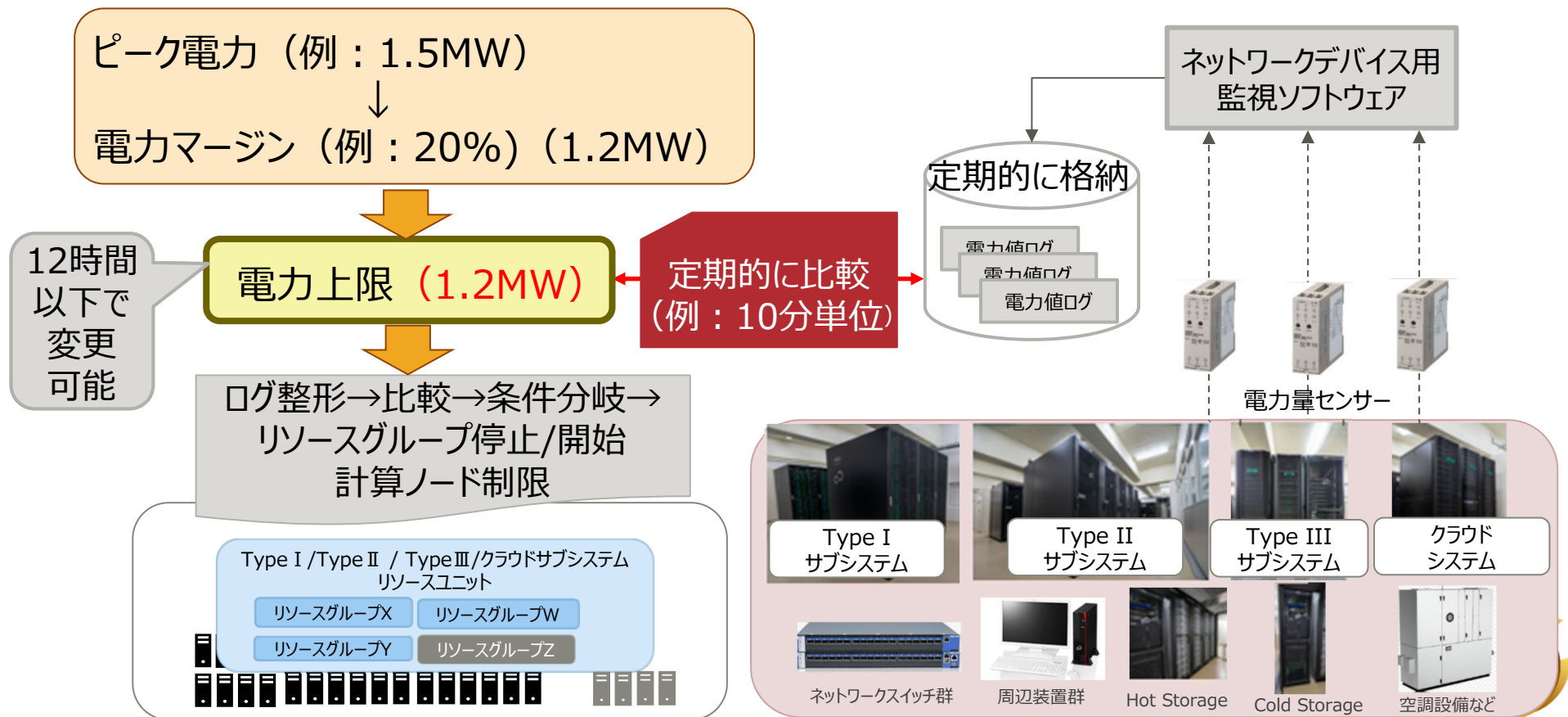
- ▶ 気温の高い
4月から11月
の間で利用
- ▶ 夏季の1日
(2019年8月3日)
の(旧)FX100
システムの
水冷チラーの
電気使用量
(KW)



年間数百万円程度の電気代削減を予想

使用最大電力の動的制御機構

- 監視ソフトウェアから一定時間毎に電力値を取得
- 出力された電力値と、あらかじめ規定したシステム全体の使用最大電力の上限値を比較し、最大電力の上限を超えないよう、計算ノードやジョブ実行可能範囲を制限



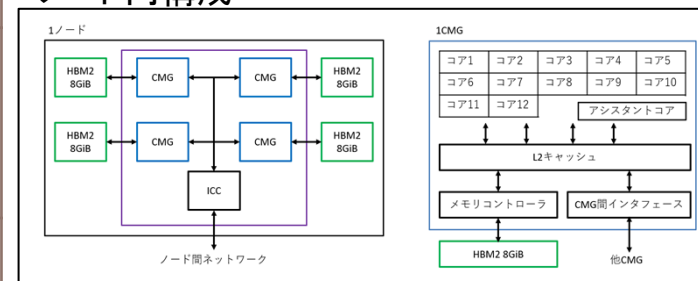
Type I サブシステム

機種名		FUJITSU Supercomputer PRIMEHPC FX1000
計算ノード	CPU	A64FX (Armv8.2-A + SVE), 48コア+2アシスタントコア(I/O兼計算ノードは48コア+ 4アシスタントコア), 2.2GHz, 4ソケット相当
	メインメモリ	HBM2, 32GiB
	理論演算性能	倍精度 3.3792 TFLOPS, 単精度 6.7584 TFLOPS, 半精度 13.5168 TFLOPS
	メモリバンド幅	1,024 GB/s (ICMG=12コアあたり256 GB/s, 1CPU=4CMG)
ノード数、総コア数		2,304ノード, 110,592コア (+4,800アシスタントコア)
総理論演算性能		7.782 PFLOPS
総メモリ容量		72 TiB
ノード間インターコネクト		TofuインターコネクトD 各ノードは周囲の隣接ノードへ同時に合計 40.8 GB/s × 双方向 で通信可能(1リンク当たり 6.8 GB/s × 双方向, 6リンク同時通信可能)
ユーザ用ローカルストレージ		なし
冷却方式		水冷



- 世界初正式運用(2020年7月1日)のスーパーコンピュータ「富岳」型システム
- 自己開発のMPIプログラム向き
- 超並列処理用
- AIツールも提供

ノード内構成



Type II サブシステム

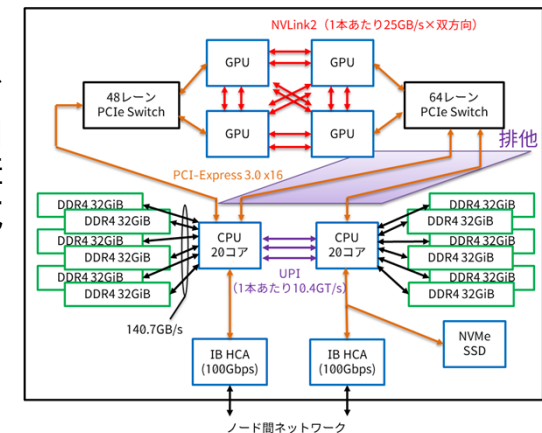
GPU (NVIDIA Volta) を 884基 搭載(221ノード)



機種名	FUJITSU Server PRIMERGY CX2570 M5
CPU	Intel Xeon Gold 6230, 20コア, 2.10 - 3.90 GHz × 2 ソケット
GPU	NVIDIA Tesla V100 (Volta) SXM2, 2,560 FP64コア, up to 1,530 MHz × 4ソケット
メモリ	メインメモリ(DDR4 2933 MHz): 384 GiB (32 GiB × 6 枚 × 2 ソケット) デバイスメモリ(HBM2): 32 GiB × 4 ソケット
理論演算性能	倍精度 33.888 TFLOPS (CPU 1.344 TFLOPS × 2 ソケット, GPU 7.8 TFLOPS × 4 ソケット)
メモリバンド幅	メインメモリ 281.5 GB/s (23.464 GB/s × 6 枚 × 2 ソケット) デバイスメモリ 900 GB/s × 4 ソケット
GPU間接続	NVLINK2 (1GPUから他の3GPUに対してそれぞれ50GB/s × 双方方向)
CPU-GPU間接続	PCI-Express 3.0 (x16)
ノード数、総コア数	221ノード、8,840 CPUコア + 2,263,040 FP64 GPUコア
総理論演算性能	7.489 PFLOPS (CPU 0.594 PFLOPS, GPU 6.895 PFLOPS)
総メモリ容量	メインメモリ 82.875 TiB、デバイスメモリ 28.288 TiB
ノード間インターコネクト	InfiniBand EDR 100 Gbps × 2, 200 Gbps
ユーザ用ローカルストレージ	NVMe SSD 6.4TB, 一部ノードにて BeeGFS/BeeOND/NVMesh (ローカルストレージを使用した共有ファイルシステム) を提供
冷却方式	水冷

- データサイエンス研究、機械学習用のGPUクラスタ型
- 最新GPU (Volta) 4台／ノード
- 充実したAIツール
- 高速SSDローカルディスク

ノード内構成



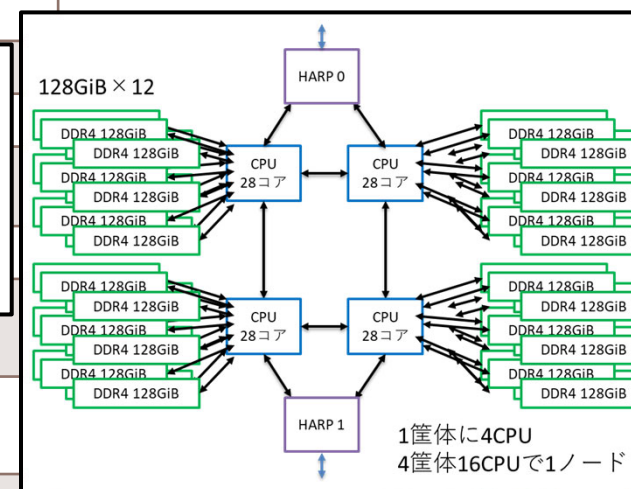
Type III サブシステム



機種名	HPE Superdome Flex	
計算ノード	CPU	Intel Xeon Platinum 8280M, 28コア , 2.70 - 4.00 GHz × 16 ソケット
	GPU	NVIDIA Quadro RTX6000 × 4
	メモリ	メインメモリ(DDR4 2933 MHz): 24 TiB (128 GiB × 12枚 × 16ソケット) デバイスメモリ(GDDR6): 24 GiB × 4
	理論演算性能	倍精度 38.7072 TFLOPS (CPU 2419.2 TFLOPS × 16 ソケット)
	メモリバンド幅	メインメモリ 2252.544 GB/s (23.464 GB/s × 12枚 (6チャンネル) × 16ソケット)
	CPU-GPU間接続	PCI-Express 3.0 (x16)
ノード数	2	
総理論演算性能	77.414 TFLOPS (38.7072 TFLOPS × 2 ノード)	
総メインメモリ容量	48 TiB	
ノード間インターコネクト	InfiniBand EDR 100 Gbps	
ユーザ用ローカルストレージ	一方のノードに102.4 TB SSD、 もう一方のノードに1008 TB 共有ストレージを接続	
冷却方式	空冷	

- **大規模共有メモリ(合計48TiB)**
- プリポスト処理用・可視化処理用
- NICE DCVを用いたリモート可視化

ノード内構成



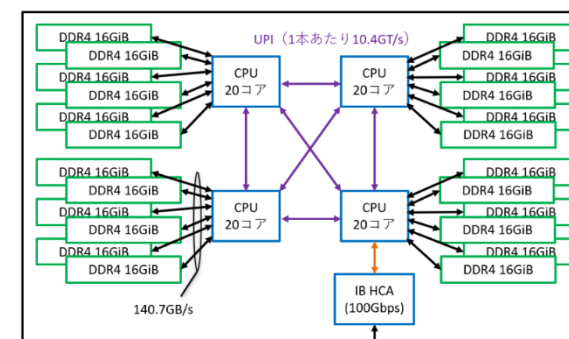
クラウドシステム

機種名		HPE ProLiant DL560
計算ノード	CPU	Intel Xeon Gold 6230 , 20コア, 2.10 - 3.90 GHz × 4 ソケット
	メモリ	メインメモリ(DDR4 2933 MHz) 384 GiB (16 GiB × 6 枚 × 4 ソケット)
	理論演算性能	倍精度 5.376 TFLOPS (1.344 TFLOPS × 4 ソケット)
	メモリバンド幅	メインメモリ 563.136 GB/s (23.464 GB/s × 6枚 × 4 ソケット)
ノード数		100
総理論演算性能		537.6 TFLOPS (5.376 TFLOPS × 100 ノード)
総メインメモリ容量		37.5 TiB
ノード間インターコネク ト		InfiniBand EDR 100 Gbps
ユーザ用ローカルストレ ージ		なし
冷却方式		空冷



- 研究室クラスタから移行しやすい
Intel CPU搭載システム
- 高いノードあたりCPU性能
(4ソケット)
- 時刻を指定してのバッチジョブ・
インタラクティブ利用が可能

ノード
内構成



ノード間ネットワーク

ホットストレージ

メタデータサーバ(MDS)	
機種名	FUJITSU PRIMERGY RX2540 M5
CPU	Intel Xeon Gold 5222 (3.80GHz, 4コア) × 2
メインメモリ	DDR4 192 GiB
HDD	SAS 900 GB 10krpm × 2 (RAID1)
Interconnect	InfiniBand EDR × 2
SAN	FibreChannel 32 Gbps × 2
OS	RedHat Enterprise Linux
ノード数	4台
メタデータストレージサーバ(MDT)	
機種名	FUJITSU ETERNUS AF250 S2
SSD	RAID1+0 [4D+4M] × 2 + 2HS RAID1+0 [3D+3M] × 1 + 2HS
ノード数	1台

データストレージ(OSS/OST)	
機種名	DDN SFA18KE × 1台 DDN SS9012 × 10 台
HDD	NL-SAS 14TB 7.2krpm × 730、RAID6 [8D+2P] 30 Device × 24 DCR Pool + 10HS
Interconnect	InfiniBand EDR × 8
搭載セット数	4
総容量	
物理容量	40.32 PB (Global Spareを除く)
実効容量	約 30.44PB

- HDD RAID
- 大容量: 30.44 PB (実効容量)
- 超高速アクセス性能: 384 GB/s



コールドストレージ

フェーズ1: 2020年7月1日より稼働

機種名	PetaSite拡張型 Library
総スロット数 (最大搭載可能カートリッジ数)	88巻
総物理容量 / 最大搭載可能容量	484 TB / 484 TB
総ドライブ数	6
ODAサーバ数	1



フェーズ2: 2021年2月1日より稼働

機種名	PetaSite拡張型 Library
総スロット数 (最大搭載可能カートリッジ数)	1,980巻
総物理容量 / 最大搭載可能容量	6 PB / 10.89 PB
総ドライブ数	20
ODAサーバ数	4

- 1度書き込み(追記)のみの光ディスクストレージ
- **実験データ等の長期データ保存用**
- 理論上100年データ保持可能
- 水にぬれても読み出せる
- サービス終了後ユーザに光ディスクを返却

高精細可視化システム

Type IIIサブシステム(HPE Superdome Flex)

77.4TFLOPS/48TiB MEM

Intel Xeon Platinum 8280M(2.7GHz,28Core)×16CPU×2 24TiB×2
NVIDIA Quadro RTX6000×4×2
HDD:実効容量500TB(RAID6)×2, NVMe:51.2TB×2



Quadro RTX6000



光ケーブル

SINET



リモート
可視化

大規模共有メモリ:
24TiB

スーパーコンピュータ共有ストレージ



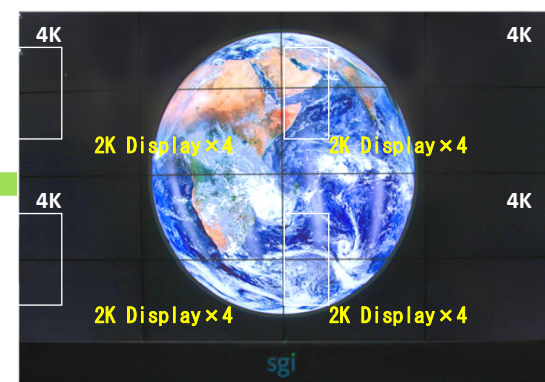
ホットストレージ
DDN SFA18KE
総実行容量: 30PB



コールドストレージ
Sony ODA PetaSite 拡張型ライブラリ
搭載容量: 6PB (5.5TB×1091巻)
最大容量: 10.89PB (5.5TB×1980巻)

8Kタイルドディスプレイ

185inchi 8K高精細大画面タイルドディスプレイ
(総解像度: 7680×4320)
Samsung 46inchi Display×16



HDMI

画像処理装置
(Windows
可視化サーバ)



Intel Xeon Gold 5122
3.6GHz 4 Cores, 512 GiB MEM
NVIDIA Quadro RTX8000
SATA SSD 2TB
光Disk 5.5TBドライブ ODS-D380U



オープンソースソフトウェア（OSS）の 利用促進

- ▶ 民間利用に限らず、アカデミック利用でも、ISVソフトのライセンス料の問題がある
 - ▶ 特に、並列利用時に、膨大なライセンス料がかかることがある
- ▶ 国策プロジェクトにより開発されたOSSもある
- ▶ スーパーコンピュータ「不老」では、上記のOSSをプリインストールとすることで、利用者のOSSへの移行を促進する
- ▶ 国策プロジェクトのOSSのインストールは、一部、一般財団法人 高度情報科学研究機構（RIST）の協力の上、講習会を実施

スーパーコンピュータ「不老」 ソフトウェア利用環境（1／4）

プログラム開発環境など

		フロントエンドシステム	Type I サブシステム	Type II サブシステム	Type III サブシステム	クラウドシステム	画像処理サーバ	オンサイト利用装置	企業利用
Intel Parallel Studio Computing Suite	コンパイラ (Fortran, C/C++), プロファイラ/デバッガ, MPI, 数値計算ライブラリ	○		○	○	○			●
PGI Professional Edition	コンパイラ (Fortran, C/C++, OpenACC, CUDA Fortran), プロファイラ/デバッガ, MPI, 数値計算ライブラリ	○		○					●
FUJITSU Technical Computing Suite	コンパイラ (Fortran, C/C++), プロファイラ/デバッガ, MPI, 数値計算ライブラリ	○	○						●
Arm Forge Proffesional	プロファイラ/デバッガ/最適化	○		○	○				●
NVIDIA CUDA SDK	GPU 統合開発環境	○		○					●
Singularity	コンテナ環境	○		○					●
その他	GV, Gfortran, GCC, perl, Python, Ruby, R, Emacs, vi, nkf, etc.	○	○	○	○	○		○	●
	OpenGL	○		○	○	○		○	●

スーパーコンピュータ「不老」 ソフトウェア利用環境（2 / 4）

ライブラリ利用環境

		フロントエンド システム	Type I サブシ ステム	Type II サブシ ステム	Type III サブシ ステム	クラウド サブシ ステム	画像処理 サー バ	オンサイト 利用 装置	企業利用
数値計算 ライブラリ	FFTW, SuperLU, SuperLU MT, SuperLU DIST, METIS, MT-METIS, ParMETIS, Scotch, PT- Scotch, PETSc, MUMPUS, Xabclib								
(含、セン ター独自 開発ソフ ト)	ppOpen-HPCライブ ラリ: ppOpen- APPL, ppOpen- AT, ppOpen- MATH 精度保障ライブラリ: LINSYS_V, DHPMM_F	○	○	○	○	○			●
入出力 フォーマッ トライブラ リ	NetCDF, Parallel netCDF, HDF5, JHPCN-DF	○	○	○	○	○			●
画像処理 ソフトウェ ア	OpenCV, Geant4	○	○	○	○	○			●
機械学習 ソフトウェ ア	Caffe, Chainer, Keras, PyTorch, TensorFlow, Theano, Mxnet, ONNX パッケージ: conda, Numpy, Scipy, scikit- image, pillow, matplotlib, jupyterlab	○	○	○	○	○			●

スーパーコンピュータ「不老」 ソフトウェア利用環境（3 / 4）

解析ソフトウェア利用環境

- *1 Type III サブシステムの会話型ノード(lm01)で利用可。
- *2 CPU並列版の他にGPU対応版が利用可。

名大では**Gaussian**も企業利用可能
(ライセンス保有)

		フロントエンドシステム	Type I サブシステム	Type II サブシステム	Type III サブシステム	クラウド サブシステム	画像処理サーバ	オンサイト利用装置	企業利用
流体解析	OpenFOAM, FrontFlow blue/red		○	○	○	○			●
構造解析	LS-Dyna			○					
	FrontISTR		○	○	○	○			●
計算化学解析	AMBER		○	○ *2		○			
	Gaussian, Gamess, Gromacs, LAMMPS, NAMD		○	○ *2		○			●
	Modylas		○	○		○			●
メッシャー	Pointwise	○			○ *1				
統合ソフトウェア	HyperWorks			○ *2		○			

スーパーコンピュータ「不老」 ソフトウェア利用環境（4 / 4）

可視化ソフトウェア利用環境

•*1 Type III サブシステムの会話型ノード(lm01)で利用可。

		フロント エンド システム	Type I サブシス テム	Type II サブシス テム	Type III サブシス テム	クラウド サブシス テム	画像処理 サーバ	オンサイト 利用装置	企業利用
リモート可 視化	NICE DCV	○			○*1				●
可視化ソ フトウェア	FieldView	○			○		○	○	
	AVS/Expre ss, Paraview, POV-Ray, VMD	○			○		○	○	●
	3D AVS Player, ffmpeg, ffplay	○			○*1		○	○	●
	IDL, ENVI	○			○		○		
	MicroAVS						○		●
	3dsMax, Visual Studio Pro						○		

ベンチマーク結果

スーパーコンピュータ「不老」ベンチマーク結果

ベンチマーク名	性能	旧システムからの 速度向上
TOP500 (HPL) 連立一次方程式の求解	6.617 PFLOPS (世界 36位)(国内5位)→41位 (2020年6月: Type I サブシステム)	2.27倍 2.910 PFLOPS (FX100)
HPCG 産業利用で多い疎行列反復解法	0.231 PFLOPS (世界 16位)(国内4位)→21位 (2020年6月: Type I サブシステム)	2.65倍 0.087 PFLOPS (FX100)
GKVカーネルベンチ 名大独自開発のプラズマ シミュレーションベンチマーク	0.258 [秒] (kernel2_intgrl) Type I サブシステム	9.16倍 2.36[秒](FX100)
Modylas 分子動力学ソフトウェア	20.72 [秒] Type I サブシステム	2.99倍 61.9[秒] (FX100)
VOLR ボリュームレンダリング	1.29 [秒] Type III サブシステム	9.79倍 12.6[秒](UV2000)
HPL-AI 人工知能で必要な演算	30.1 PFLOPS (96.6%) (世界 5位)(国内2位) (2020年11月: Type I サブシステム)	--倍

スーパーコンピュータ「不老」 アプリケーション

スーパーコンピュータ「不老」で動かす アプリケーション例

■ 大規模数値計算 (Type I サブシステム)

- ▶ 名大 坪木和久 教授：スーパー台風解析 (雲解像モデル CReSS)
- ▶ 名大 渡邊智彦 教授：プラズマシミュレーションGKV

■ AI/GPUコンピューティング (Type II サブシステム)

- ▶ 名大 森健策 教授 (GPU大規模AI計算)：AIによる医用画像診断支援技術
- ▶ 名工大 本谷秀堅 教授 (規模GPU計算)：医用画像処理 LDDMM

■ 大規模数値計算+AI (Type I サブシステム+Type II サブシステム)

- ▶ 立教大 望月祐志 教授：COVID-19解析、フラグメント分子軌道計算
(ABINIT-MP) 理研R-CCS：新型コロナウイルス対策を目的としたスーパーコンピュータ「富岳」の優先的な試行的利用採択課題
- ▶ 山梨大 相馬一義 准教授：機械学習による気象予測

■ 可視化・ストレージ (Type III サブシステム+ホット/コールド ストレージ)

- ▶ 名大 高橋一郎 特任主任技師：
可視化ツール VisPlus、コールドストレージ操作ツールODAPLUS

■ 大規模数値計算（Type I サブシステム）アプリケーション

スーパー台風のメカニズム解析

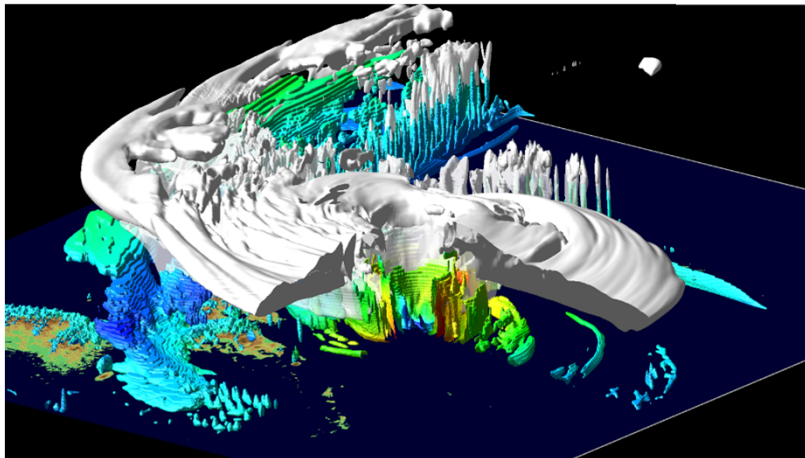
- **名古屋大学 坪木和久教授 開発による 雲解像モデルCReSS**

(Cloud Resolving Storm Simulator)によるシミュレーション

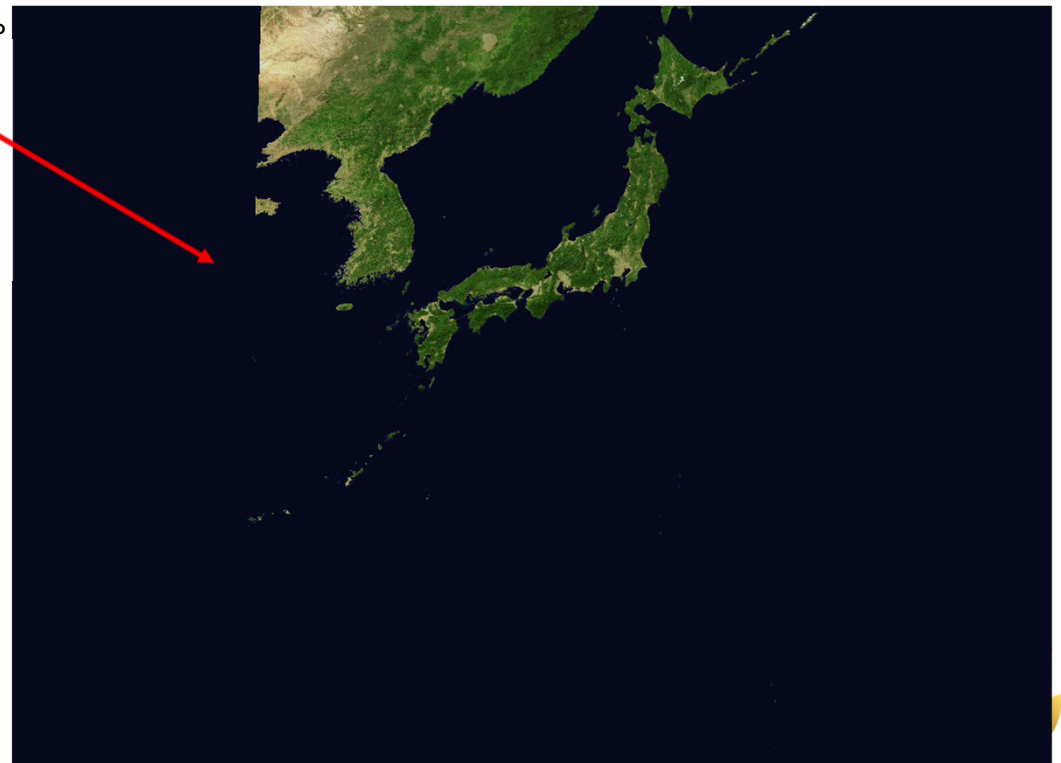
- **未来の台風（2076年09月に発生）**に伴う雲を立体的表示

この台風は太平洋上を北上し、日本に上陸する直前でも中心気圧880 hPa以下を維持。

台風が太平洋上にあるとき、中心気圧870～860 hPa、最大地上風速70～80 m/sを4日間維持し、ほぼその強度のまま関東地方に上陸する。



伊勢湾台風のシミュレーション結果



台風のメカニズム解析

：スーパーコンピュータ「不老」でのターゲット

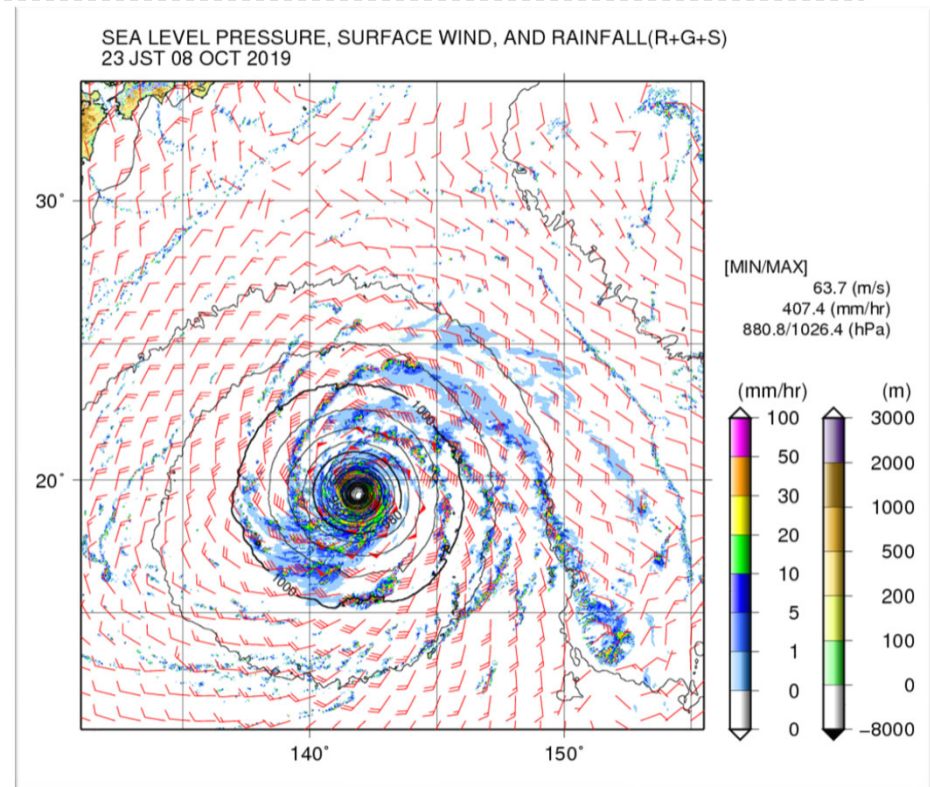
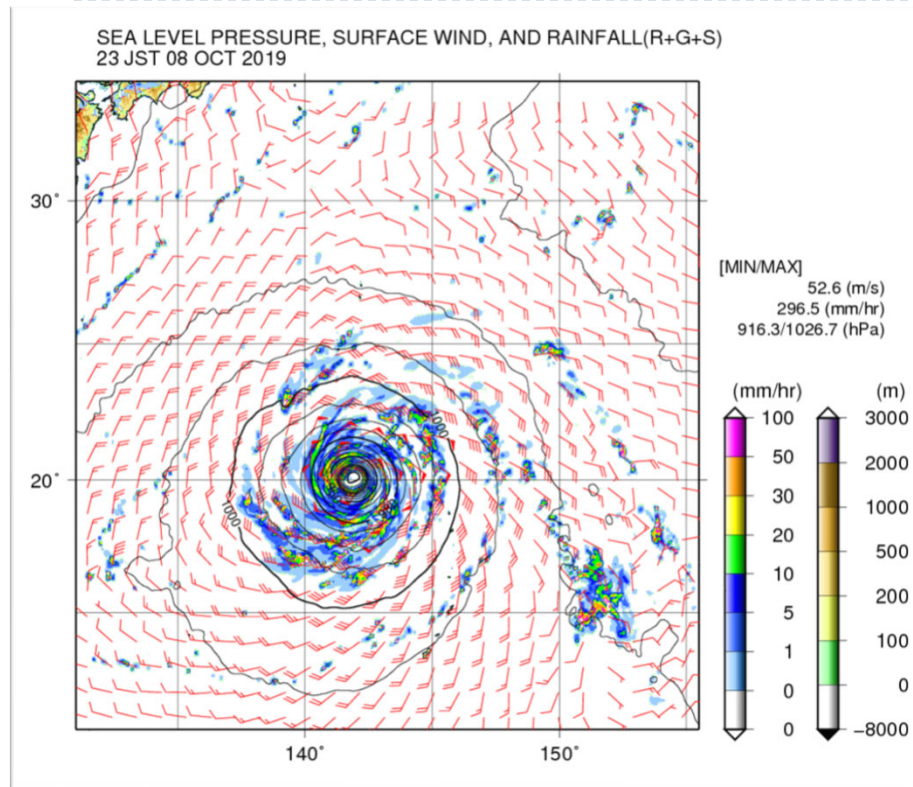
▶ 旧システム(FX100)

- ▶ 水平間隔 2 km 格子、東西・南北それぞれ1000～2000格子、鉛直は地上から上空約20 km まで 100層、積分時間数日程度の計算
- ▶ 数日の実行時間を必要

▶ スーパーコンピュータ「不老」によって

- ▶ より細かく短時間に
- ▶ 水平1 kmあるいは 500 m 、かつより広い領域で、台風の急発達のメカニズムや詳細な構造が明らかになる
- ▶ 数100 m 格子でより詳細な地形を考慮したシミュレーションにより豪雨の詳細な構造を明らかにし、豪雨に伴う被害の詳細な予測が可能に

台風メカニズム解析 ：スーパーコンピュータ「不老」での計算結果



従来：水平格子間隔約2.5km

水平格子間隔約1.0km

スーパーコンピュータ「不老」で計算

1600ノード、24時間程度

■ AI／GPUコンピューティング
(Type II サブシステム)
アプリケーション

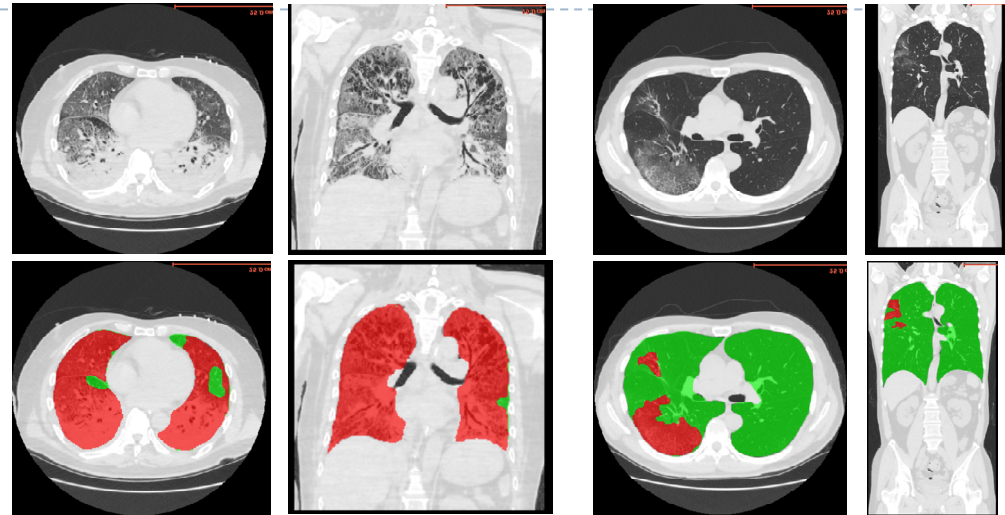
新システムで想定されるアプリケーション

COVID19症例CT画像のAI解析

名大 森健策研究室

- ▶ AIによるCT像からの肺野セグメンテーション

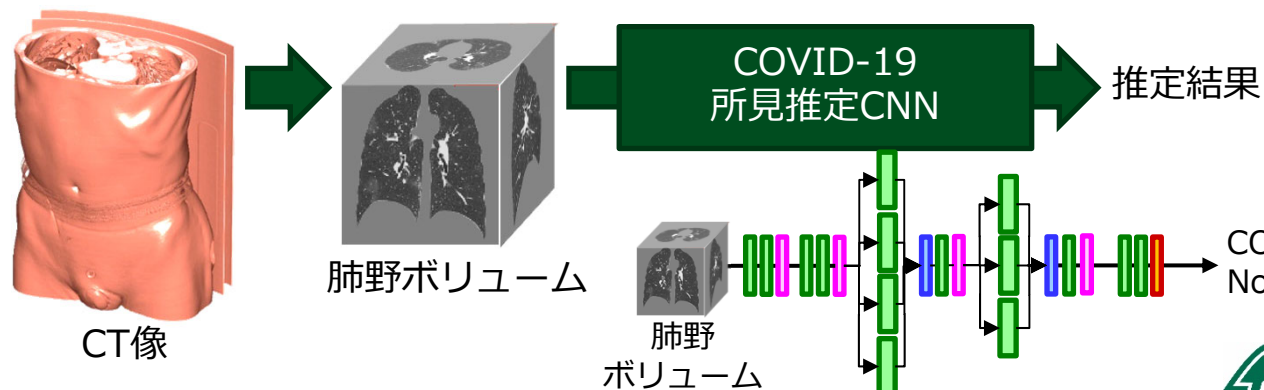
- ▶ 正常と炎症領域を3次元的に自動識別



肺野自動セグメンテーション結果

- ▶ AIによるCOVID-19画像所見推定

- ▶ 医師の所見をAIが推定し迅速な診断を支援



▶ 40

画像はNII J-MIDとの共同研究成果

新システムで想定されるアプリケーション

COVID19症例CT画像のAI解析

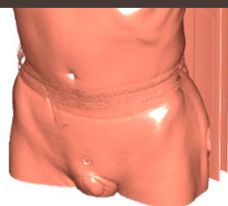
名大 森健策研究室

- ▶ AIによるCT像からの肺野セグメンテーション

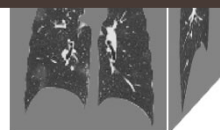
- ▶ 正常と炎症領域を3次元的に自動識別



約10万枚の画像を横断的に解析
「不老」で解析スピードが約800倍に



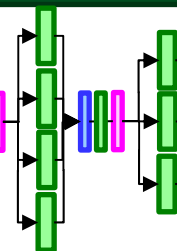
CT像



肺野ボリューム



肺野
ボリューム



COVID-19 typical
Not typical



▶ 41

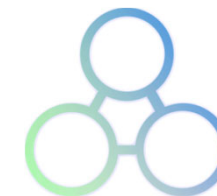
画像はNII J-MIDとの共同研究成果



名古屋大学
NAGOYA UNIVERSITY

■ 大規模数値計算＋AI
(Type I サブシステム＋Type II サブ
システム)

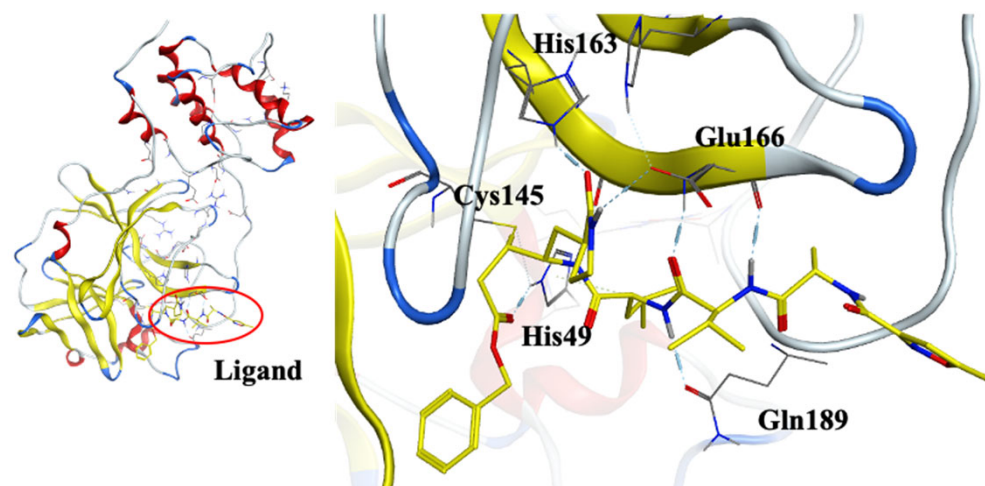
コロナウイルスのメインプロテアーゼとN3阻害剤の複合構造に関するフラグメント分子軌道計算



畑田峻, 奥脇弘次, ○望月祐志 (立教大学), 福澤薫 (星薬科大学)
古明地勇人 (産業技術総合研究所), 沖山佳生 (国立医薬品食品衛生研究所),
田中成典 (神戸大学院)

●SARS-CoV-2 メインプロテアーゼとN3阻害剤の結晶構造

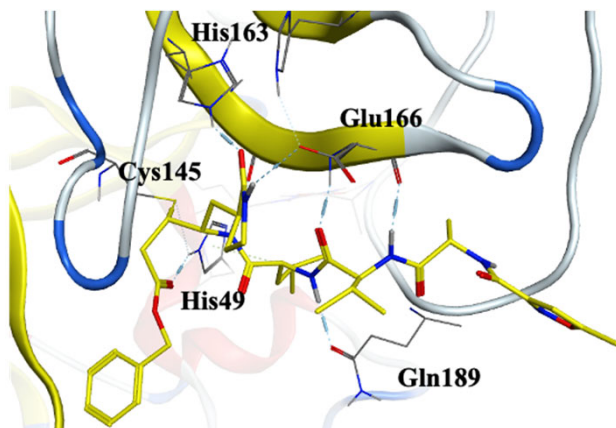
- 2020年2月に、LiuらによってSARS-CoV-2のメインプロテアーゼと、阻害能を有する化合物の複合体の結晶構造が発表された
(PDB ID: 6LU7)



本研究では、フラグメント分子軌道法と6LU7結晶構造を用いて
メインプロテアーゼ-N3阻害剤間の相互作用解析を行った

- 研究内容はChemRxivで2020/3/17に公開 (<https://bit.ly/3f67n1c>)
- アメリカ化学会の専門誌に採録：Fragment molecular orbital based interaction analyses on COVID-19 main protease - inhibitor N3 complex (PDB ID:6LU7), J. Chem. Inf. Model. 2020, June 15, 2020 (<https://doi.org/10.1021/acs.jcim.0c00283>)

フラグメント分子軌道法を用いた相互作用解析



● 解析方法

N3阻害剤を5つの部位に、タンパク質をアミノ酸単位にフラグメント分割し、フラグメント間の相互作用エネルギー(IFIE)を算出
気相条件・溶媒条件下で各々MP2/6-31G*レベルで計算

● 計算環境：ABINIT-MPプログラム @ 名大FX-100

気相条件：128ノード（16スレッド - 256プロセス） 1.4時間

溶媒条件：192ノード（16スレッド - 384プロセス） 30.5時間

● 解析結果

分割部位
BDA→BAA

Fragment5
-21.32 kcal/mol (気相)
-17.77kcal/mol (溶媒)
大きな相互作用はなし

Fragment5

Fragment4
-84.85 kcal/mol(気相), -79.58 kcal/mol(溶媒)
主にHis163, 164と相互作用

Fragment3
-47.50 kcal/mol(気相)
-49.45 kcal/mol(溶媒)
主にGlu166, Gln189と相互作用

Fragment3

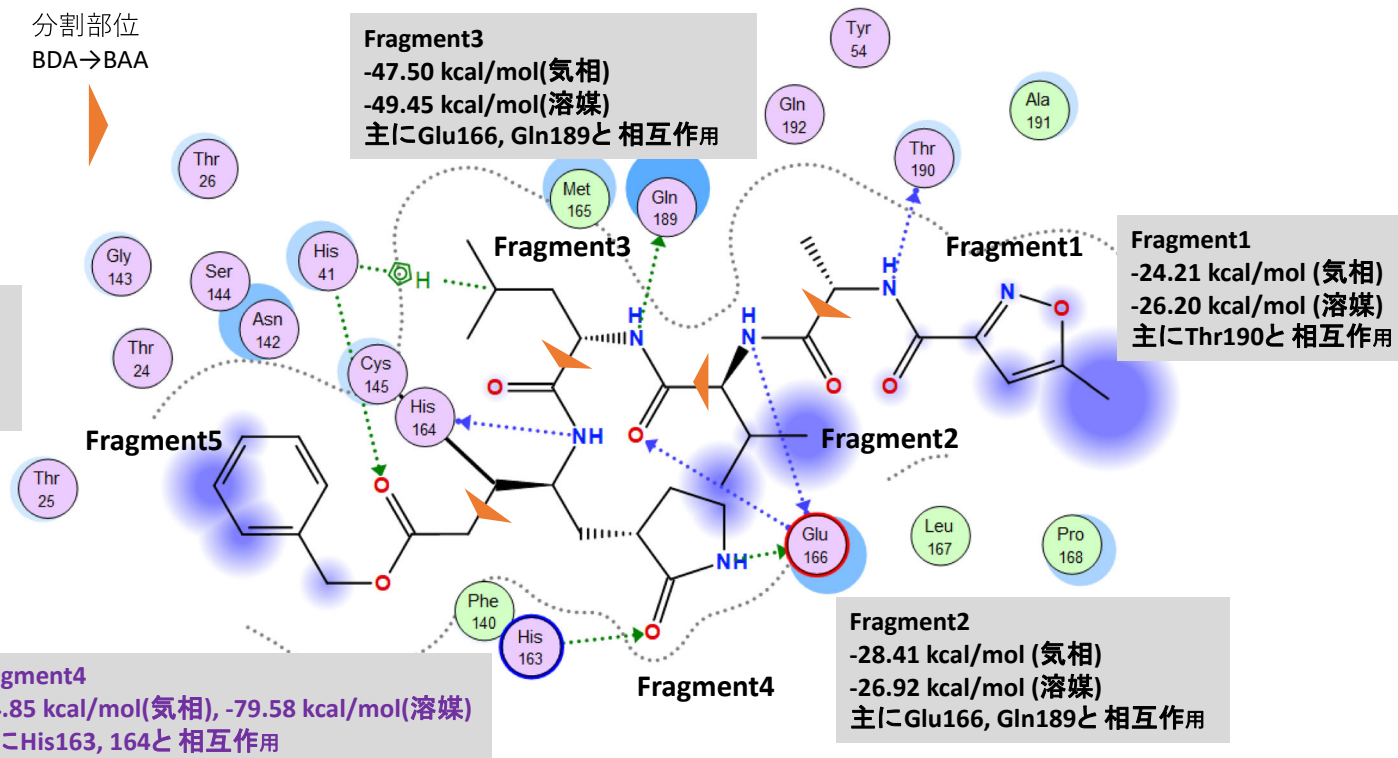
Fragment4

Fragment2
-28.41 kcal/mol (気相)
-26.92 kcal/mol (溶媒)
主にGlu166, Gln189と相互作用

Fragment2

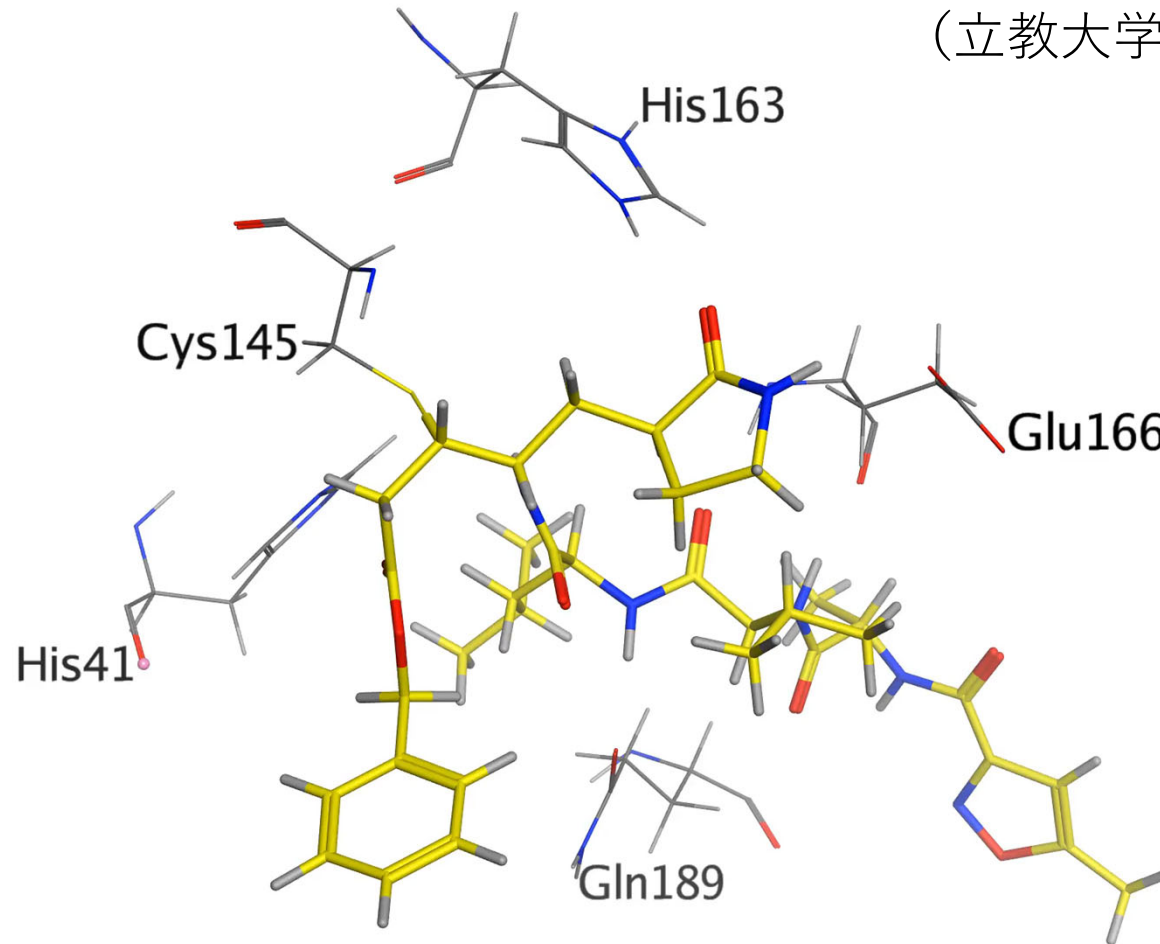
Fragment1
-24.21 kcal/mol (気相)
-26.20 kcal/mol (溶媒)
主にThr190と相互作用

Fragment1



新型コロナウイルスのメインプロテアーゼと 結合した阻害剤N3の重要部位の構造ゆらぎ

(立教大学望月研究室提供)



スーパーコンピュータ「不老」で以下の研究開発中：

- **Type I サブシステム**：SIMD強化等のコードチューニング
- **Type II サブシステム**：AI処理連携

課金体系

スーパーコンピュータ「不老」 課金体系 (1/8)

- ▶ 前払い定額制(プリペイド形式)
 - ▶ 利用すべき資源の料金を前払いして利用
 - ▶ 利用ポイントに変換して利用
- ▶ 単年度会計(4月1日～翌年3月31日)
 - ▶ 年度途中で申込み可能だが、利用終了は年度末
 - ▶ 年度末に余った利用ポイントは没収
- ▶ 一度の申込みで全てのサブシステムと可視化システムを利用可能
 - ▶ Type I、Type II、Type III、クラウド 全て共通

スーパーコンピュータ「不老」 課金体系 (2/8)

- ▶ アカデミックユーザ(大学、研究機関など所属)向けプラン
 - ▶ 基本負担金
 - ▶ 利用登録1名につき年額10,000円(登録料)
 - ▶ 10,000利用ポイントを付加(他ユーザへの譲渡不可)
 - ▶ 追加負担金
 - ▶ 1,000円単位で追加が可能
 - ▶ 50万円未満: 1円あたり1ポイント付加
 - ▶ 50万円以上: 1円あたり1.25ポイント付加

スーパーコンピュータ「不老」 課金体系 (3/8)

- ▶ Type1サブシステム(「富岳」型ノード)消費ポイント
 - ▶ 計算課金: $\text{利用ノード数} \times \text{経過時間[s]} \times 0.0056$
 - ▶ 基本負担金1万円=1万ポイント付加で利用可能な目安
 - 1ノードを 約21日
 - 4ノードを 約 5日
 - ▶ 10万円(基本利用料金1万円、追加料金9万円)
=10万ポイント付加で利用可能な目安
 - 1ノードを 約207日
 - 4ノードを 約 52日
 - 8ノードを 約 26日
 - ▶ 1ノードの年間利用額: 約16万9000円
 - 保守日等を考慮し年間350日利用できると仮定、以下同様

スーパーコンピュータ「不老」 課金体系 (4/8)

▶ Typellサブシステム(GPUノード)消費ポイント

- ▶ 計算課金: $\text{利用GPU数} \times \text{経過時間[s]} \times 0.007$
 - ▶ 基本負担金1万円
= 1万ポイント付加で利用可能な目安
 - 1ノード(1GPU)を 約17日
 - 1ノード(4GPU)を 約 4日
 - ▶ 10万円(基本利用料金1万円、追加料金9万円)
= 10万ポイント付加で利用可能な目安
 - 1ノード(1GPU)を 約165日
 - 1ノード(4GPU)を 約 41日
 - 4ノード(16GPU)を 約 10日
 - ▶ 1ノード(4GPU)の年間利用額: 約84万7000円

※50万円以上を一度に入金すると
ポイントが1.25倍
→約67万7600円に！

スーパーコンピュータ「不老」 課金体系 (5/8)

◆ Type III: 1ソケット当たり28コア
◆ クラウド: 1ソケット当たり20コア

- ▶ TypeIIIサブシステム(大規模共有メモリノード)、
およびクラウドシステムの消費ポイント
 - ▶ 計算課金: **利用ソケット数 × 経過時間[s] × 0.002**
 - ▶ 基本負担金1万円
= 1万ポイント付加で利用可能な目安
 - 1ソケットを 約58日
 - 4ソケットを 約14日
 - ▶ 10万円(基本利用料金1万円、追加料金9万円)
= 10万ポイント付加で利用可能な目安
 - 4ソケットを 約145日
 - 32ソケットを 約 18日
 - ▶ 2ソケットの年間利用額: 約12万1000円

スーパーコンピュータ「不老」 課金体系 (6/8)

- ▶ 会話型利用: ログインノード上での処理、および Type III サブシステム (会話型処理) の消費ポイント
 - ▶ 計算課金: $\text{利用CPUコア数} \times \text{経過時間[s]} \times 0.002$
 - ▶ ログインノードでの処理は割高に設定しています
 - ▶ ログインノードは限りある資源のため、計算ノードの利用にご協力ください
 - ▶ 各計算ノードの備える interactive ジョブクラスは バッチジョブ扱いの課金
です (fx-interactive, cx-interactive, cl-interactive)

スーパーコンピュータ「不老」 課金体系 (7/8)

▶ ホットストレージ

▶ ファイル課金

- ▶ 1TB 以下の場合(Home + Large): 徴収しない
- ▶ ファイルの使用容量が1TB を超えた場合:
超えた容量について、1GBにつき 1日当たり 0.01 ポイント
- ▶ 例) 2TB(2000GB)利用: 1000GBが課金対象
⇒ 10ポイント/日 ⇒ 300円/30日、3,500円/350日
(保守などで停止する日については徴収しない)

※128TBを超える場合は、全体容量を考慮して、削除依頼をさせていただきます【予定】。

※128TBを超える容量が必要な場合は、事前に相談ください。

スーパーコンピュータ「不老」 課金体系 (8/8)

▶ コールドストレージ

▶ ファイル課金

▶ 1口: 50TB

□ 1回だけ書き込める(追記可能)の
光ディスク×10枚(1枚約5TB)

▶ ファイル負担経費(初回利用時のみ必要):

1口 190,000円

▶ ファイル管理経費(毎年必要、基本負担金とは別):

1口 10,000円

※ユーザの利用終了時、もしくは、スーパーコンピュータ
「不老」運用終了時に、光ディスクを持ち帰りいただけます。

スーパーコンピュータ「不老」 新サービス：ノード準占有利用

- ▶ ノード準占有利用
 - ▶ 1時間以内のジョブ実行開始を保証
 - ▶ バッチ利用のみ
- ▶ 1ノード、1ヶ月間の利用負担金
 - ▶ Type IIサブシステム：
210,000円（通常価格の約2.8倍）
 - ▶ クラウドシステム：
62,000円（通常価格の約2.8倍）

スーパーコンピュータ「不老」 新サービス：クラウドノード予約利用

▶ クラウドノード予約利用

- ▶ 専用の予約システム「UNCAI」(Webブラウザで操作)でノードを予約して利用する

▶ 利用料金

- ▶ 計算課金：**利用コア数 × 経過時間[s] × 0.0001**
(ソケット当たりコア数を考えればバッチ実行と同等)
 - ▶ 1ソケット当たり20コア、10コア(0.5ソケット)から利用可能
 - ▶ 利用可能なメモリ容量もコア数に比例
 - ▶ **基本負担金1万円でXeon Gold 20コア1ソケットを約58日間使用可能**

スーパーコンピュータ「不老」 新サービス：グループ利用

▶ グループ利用

- ▶ 1口20人まで、20万円で200,000ポイント付与
- ▶ 登録料なし
- ▶ 個人利用(個別に1万円×20人が個別に基本負担金を払う)
との違い
 - ▶ 200,000ポイントを20人で共有利用できます
 - ▶ 個人利用では、購入した10,000ポイントを他者と共有できません

お試し利用、リテラシー利用

▶ トライアルユース

- ▶ ソフトウェアの動作確認などを、無料で行える制度です。
- ▶ お1人様1回限りで申請できます。
- ▶ 企業においては、同一の課で1回のみです。
- ▶ アカデミックユーザ(無審査)、企業ユーザ(書類審査)
- ▶ 10,000ポイント付加
- ▶ 有効期限1ヶ月

▶ リテラシー利用(アカデミックユーザのみ)

- ▶ 名大学内外の学部・大学院等の講義や演習で利用いただける制度です。
- ▶ 利用登録25件につき10,000円、50,000ポイント付与
- ▶ 有効期限:上限6ヶ月(講義・演習実施期間に依存)

優先ジョブクラス（アカデミック・民間）

- ▶ Type1、Type11、クラウドの各サブシステム
 - ▶ ポイントを通常^の2倍消費することで利用可能なジョブクラス
 - ▶ 専用のキューにジョブを投げることで利用
 - ▶ 通常のジョブクラスが混んでいるときでも早く実行したい、
というユーザの利用を想定
- ▶ （優先ジョブクラスも混んでしまったらすいません）

コンサルティング

- ▶ 並列化、利用高度化、ISVアプリの利用方法などに関するコンサルティングを行っています。
- ▶ 本センター教職員や学内外の専門家で構成される専門分野相談員によるコンサルティング（面談）ができます。
 - ▶ Web受付 Q&A SYSTEM
 - ▶ 各種ご質問、ご相談等は下記Webサイトからお問合せください。
 - ▶ <https://qa.icts.nagoya-u.ac.jp/>
 - ▶ 面談相談（※ZOOMによる遠隔面談も現在計画中）
 - ▶ 実際に画面を見ながらなど、電話やメールでは伝えにくいご質問やご相談は面談でも受け付けています。
 - ▶ 事前にお約束の上、本センター3階図書室内のIT相談コーナーにお越しく下さい。または、相談員が訪問させていただくことも可能です。
 - 連絡先：052-789-4366（IT相談コーナー直通）、または上記のQ&A SYSTEMから

可視化設備

- ▶ 情報基盤センター可視化室（本館1階）
 - ▶ 可視化室の利用は予約制となります

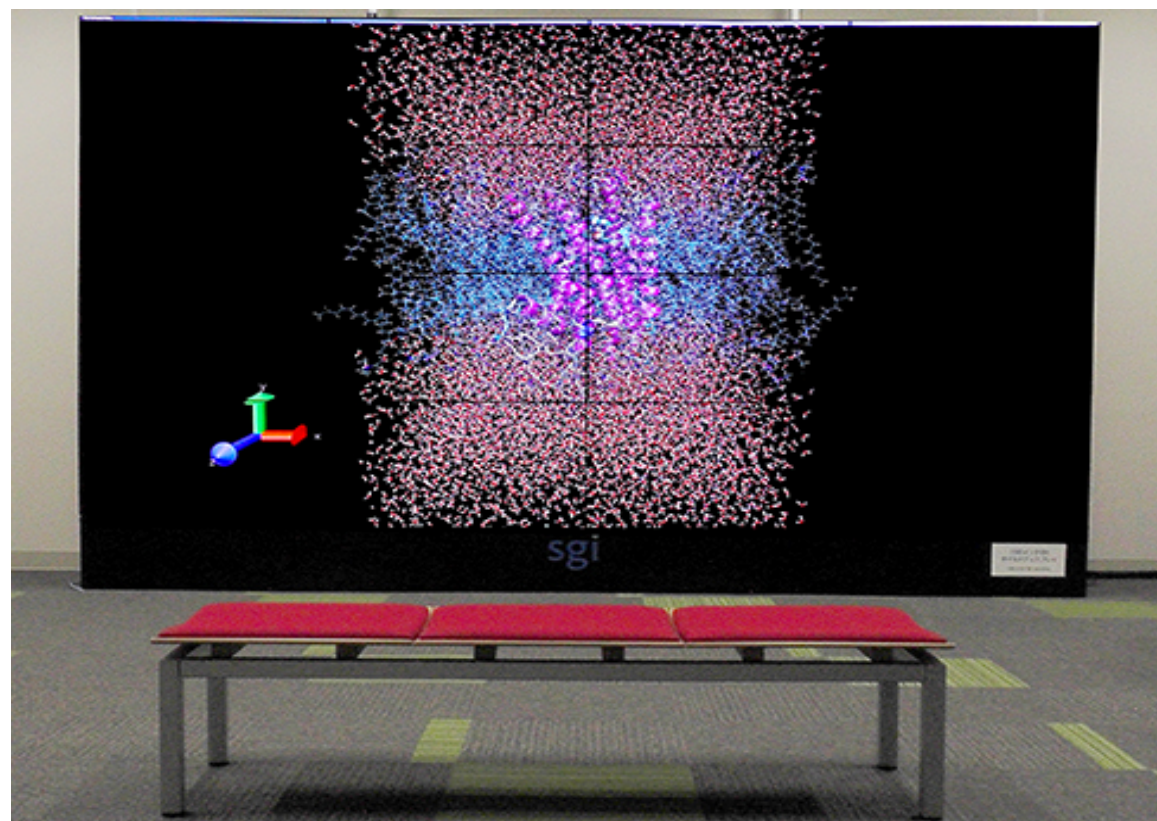


- 8K 185型タイルドディスプレイ
- 全天周映像視聴システム
- 円偏光立体視システム など

可視化設備

- ▶ 情報基盤センター可視化室（本館1階）
 - ▶ 可視化室の利用は予約制となります

8K 185型
タイルド
ディスプレイ



利用者支援室（2 部屋）

- ▶ 情報基盤センター利用者支援室（本館3階）
 - ▶ 利用者支援室の利用は予約制となります



訪問者が現地で利用可能な機材

▶ 画像処理装置 (1F可視化室)

- Windows10が動作するデスクトップパソコン(1台)
- 対話型のプリ・ポスト処理や共有ファイルストレージの大容量ファイル取り扱い用
- 10Gbps SINETによる高速インターネットが利用可能
- USB外付けHDDやディスクメディア (Blu-rayディスク) を持ち込んでホットストレージに対するデータの読み書きが可能
- 可視化室に設置された185インチ8K高精細ディスプレイに接続、大規模データの高品位なプリポスト処理やコンテンツ生成に活用可能

画像処理装置



▶ オンサイト利用装置 (3F利用者支援室)

- Linuxが動作する計算サーバ(2台)
- 対話型のプリ・ポスト処理、スーパーコンピュータシステムとの大容量ファイル取り扱い用
- USB外付けHDDや光ディスクメディア (Blu-rayディスク) を持ち込んでホットストレージに対するデータの読み書きが可能
- 予約制の部屋貸し切りで、ハードディスク持ち込みで大規模データの入力、回収が可能

オンサイト利用装置



ODA単体ドライブ



- コールドストレージの光ディスクアーカイブと互換性のあるUSB接続ドライブ
- 画像処理装置・オンサイト利用装置に接続して利用可能

概要

1. スーパーコンピュータ「不老」導入
2. 名古屋大学情報基盤センターの
民間利用制度
3. センター独自開発ソフトウェア
4. おわりに

名古屋大学情報基盤センターの 民間利用制度

名古屋大学情報基盤センターにおける スパコン民間利用制度

- ▶ 近年、スーパーコンピュータは学術研究の発展だけでなく、産業競争力の強化にとっても重要な基盤
- ▶ 名古屋大学情報基盤センターではこれまで社会貢献の一環として、民間企業のスパコン利用について、以下の事業を行ってきた
 - ▶ 先端的大規模計算シミュレーションプログラム利用サービス（文部科学省先端研究施設共用イノベーション創出事業・平成19～20年）
 - ▶ 先端的大規模計算利用サービス（文科省先端研究施設共用促進事業・平成21年～22年）
- ▶ これらの取り組みを継続し、自主事業として民間利用サービスを推進

スーパーコンピュータ「不老」 民間利用制度

- ▶ 書類での審査があります。追加負担金も同額です。

- ▶ 公開型

- ▶ 10アカウントまで**20万円**
 - ▶ アカデミック利用の料金の**2倍(20万円当たり100,000ポイント)**
 - ▶ 企業名、課題名、報告書をWebで公開(延期制度あり)

申込み金額に応じたポイント優遇
はございません。
詳しくは、産業利用のパンフレットを
ご参照ください。

- ▶ 非公開型

- ▶ 10アカウントまで**40万円**
 - ▶ アカデミック利用の料金の**4倍(40万円当たり100,000ポイント)**
 - ▶ 外部に情報は非公開(ただし内部会議では情報が出ます)

民間利用制度の概要

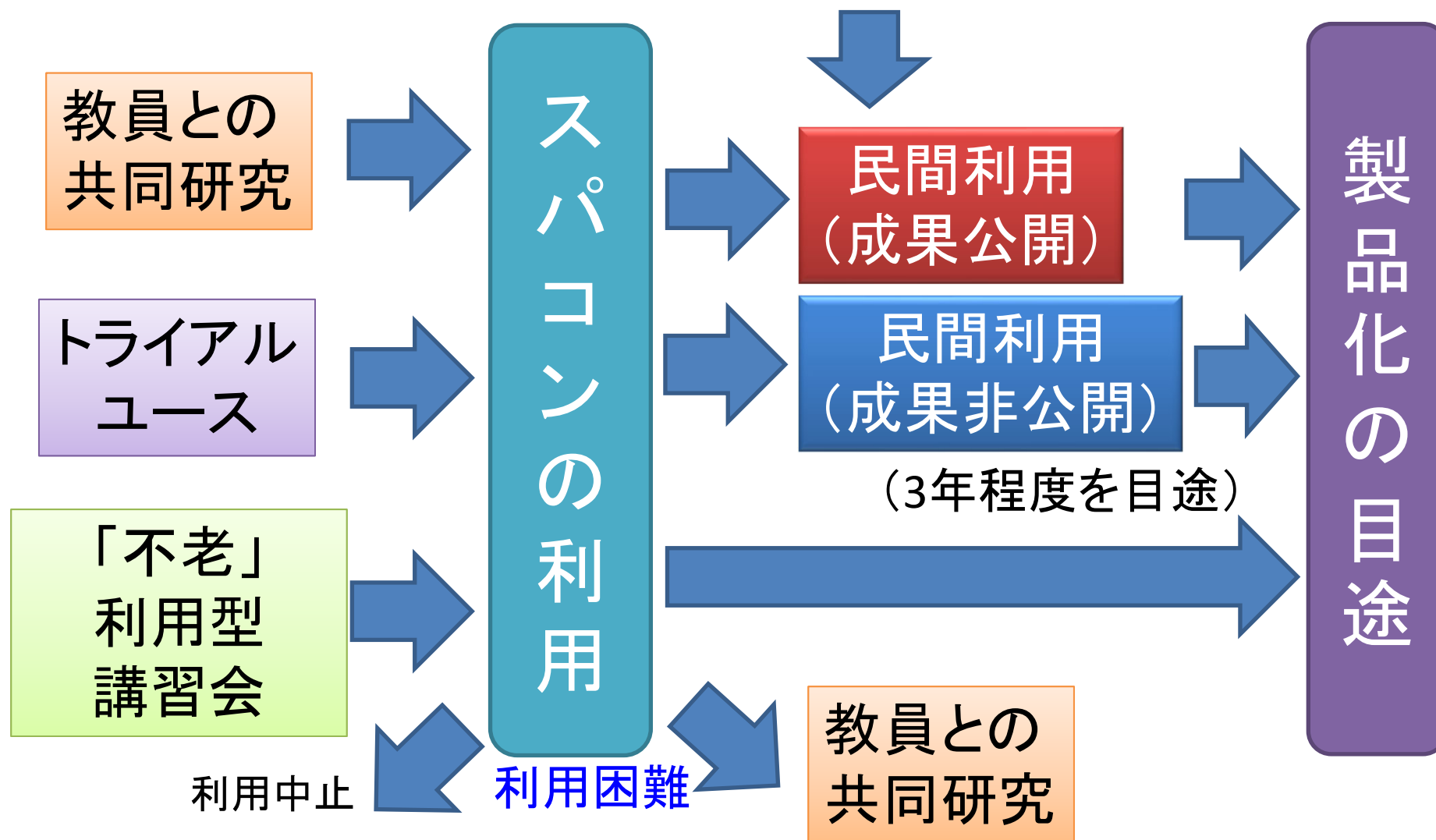
- ▶ 書類審査の上、採否を決定します
- ▶ 公開制度、非公開制度があります
- ▶ 双方、審査制度、課金設定が異なる
- ▶ 一般利用者と、計算機環境の差は無い
 - ▶ 一部、ライセンス問題から、企業ユーザの方は利用できないソフトウェアがあります。

民間利用制度の利用ケース



名大における民間利用制度の 利用ケース

71



民間利用の状況

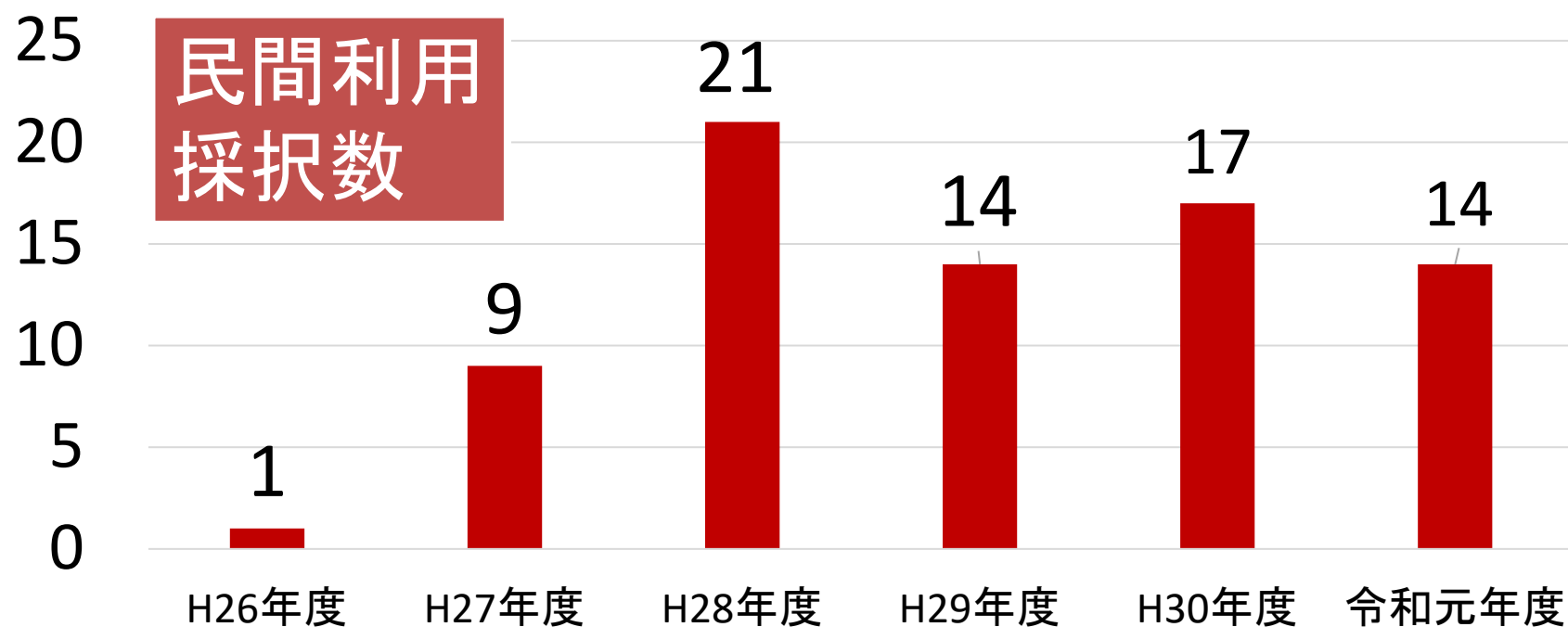


民間利用の状況(旧システム)

■ 民間利用制度の導入

◆ H27年9月より民間利用(非公開)を開始

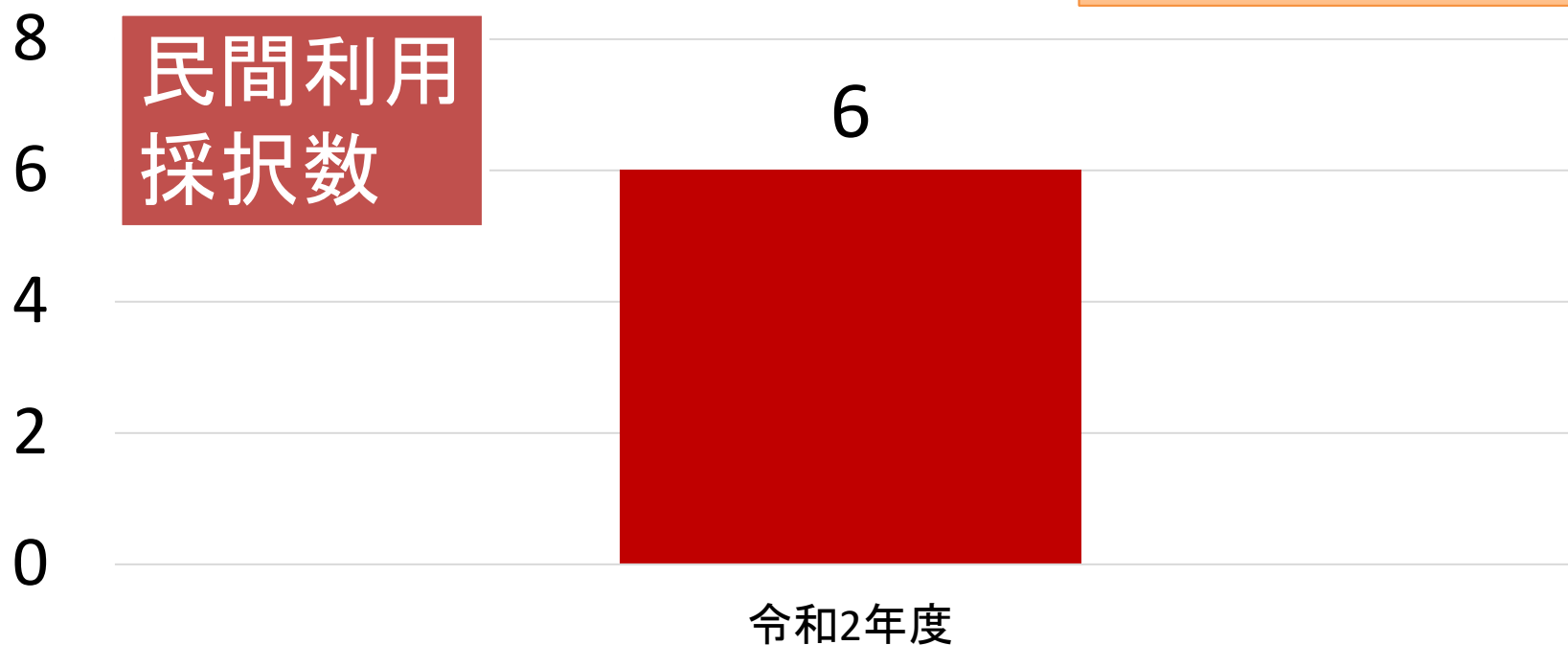
◆ H28年4月より本格利用が始動



民間利用の状況(「不老」)

■ 民間利用制度(スーパーコンピュータ「不老」) 2020年7月1日～

2021年3月9日現在



名古屋大学情報基盤センター 民間利用のまとめ

75

- 総じて、民間利用の周知不足
- 従来のPCクラスタ型（クラウドシステム）の利用は、例年通りの利用が期待できる
- AI関連利用で、Type II サブシステム（GPUクラスタ）の利用を期待
 - ◆ 名大ではGPUクラスタの導入経験がないので、周知がまだ十分ではない
 - ◆ 名大ベンチャー企業を中心に利用斡旋中



概要

1. スーパーコンピュータ「不老」導入
2. 名古屋大学情報基盤センターの
民間利用制度
3. センター独自開発ソフトウェア
4. おわりに

センター独自開発ソフトウェア



コールドストレージへの展開

プログラム名: VisPlus

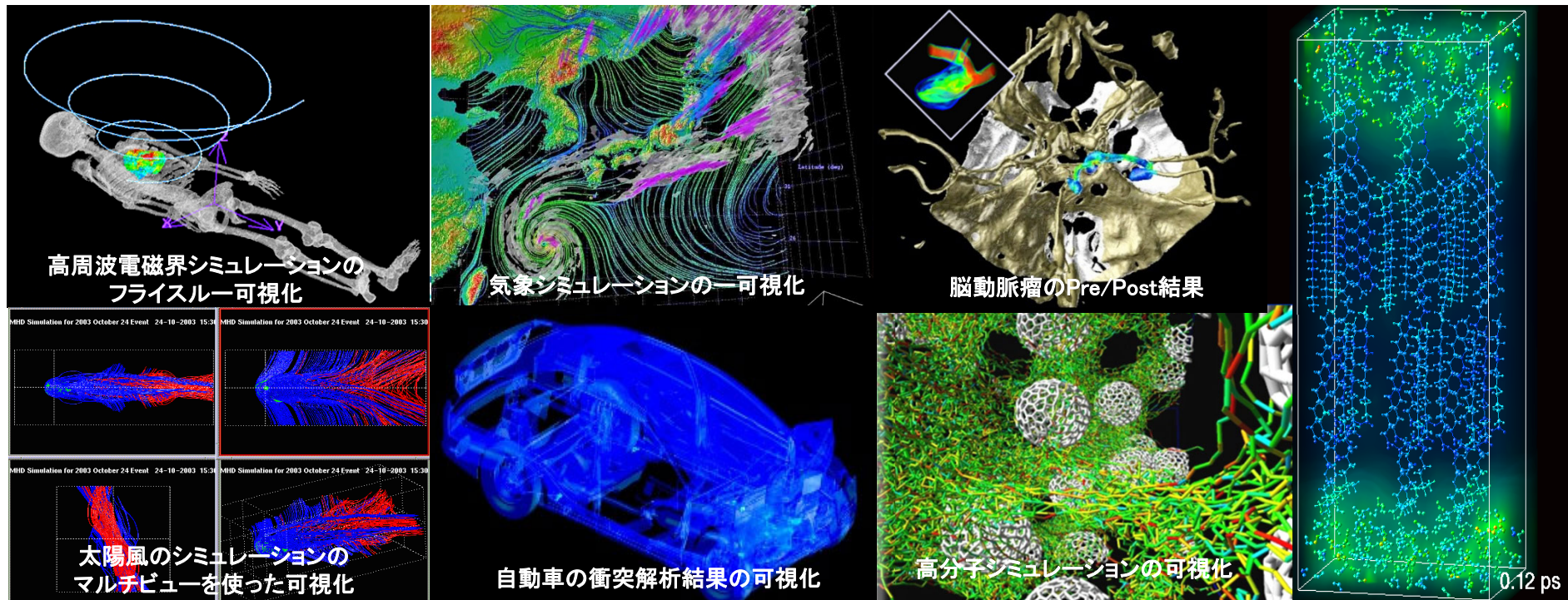
開発者: 名古屋大学 高橋一郎

概要 : 多くの大学や研究機関で利用されている可視化アプリケーション開発ツール
”*AVS Express*”を使って開発した可視化アプリケーションプログラムと
ユーティリティプログラムの集合体である。

可視化アプリケーションプログラムは、AVSのNetworkエディタを使って
カスタマイズしてオーブンプラットフォームで利用することができる。

また、リモート可視化、ローカル可視化にも対応している。

動作環境: Linux, Windows, Mac



プログラム名： ODAPLUS

開発者： Sony, 高橋一郎

概要： Sony社製コールドストレージ(光DiskライブラリODA)システムの
管理・運用を行うプログラム(現在開発中)

利用者端末



— ODA単体ドライブ&ライブラリ装置の利用方法 —

学内Network, SINET

情報基盤センター

スーパーコンピュータ「不老」

フロントエンド・サーバ

スパコン用ログインノード

データ転送用ODAログインノード

スーパーコンピュータ

Type I Type II Type III クラウド
サブシステム サブシステム サブシステム サブシステム

利用者支援室

ODA単体ドライブ装置

Hot Storage



DDN 共有ストレージ(/home,/largeファイルシステム)

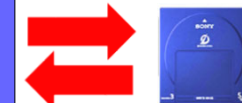
データ転送
アーカイブ
リコール

Cold Storage



ODAライブラリ装置

センター内・センター外での
利用、研究室で保管



カートリッジの
収容 / 搬出

オプティカルディスク・アーカイブとは

オプティカルディスク・アーカイブ(ODAと呼ぶ)は、デジタルデータの長期保存(アーカイブ)を目的とした、大容量光ディスクストレージシステムです。

データを格納するカートリッジに複数枚の光ディスクを格納し、1つのボリュームとして光ディスクを大容量に利用することができます。

- ① Archival Disc(業務用次世代光ディスク)
- ② ODAドライブユニット
PCやWSに、USB接続して利用する。
- ③ PetaSite拡張型ライブラリー

資料) オプティカルディスク・アーカイブのご紹介

https://www.youtube.com/watch?v=TAYQ_FiJrAk

Introduction of PetaSite Library Systems

<https://www.youtube.com/watch?v=pPbt0sARido>

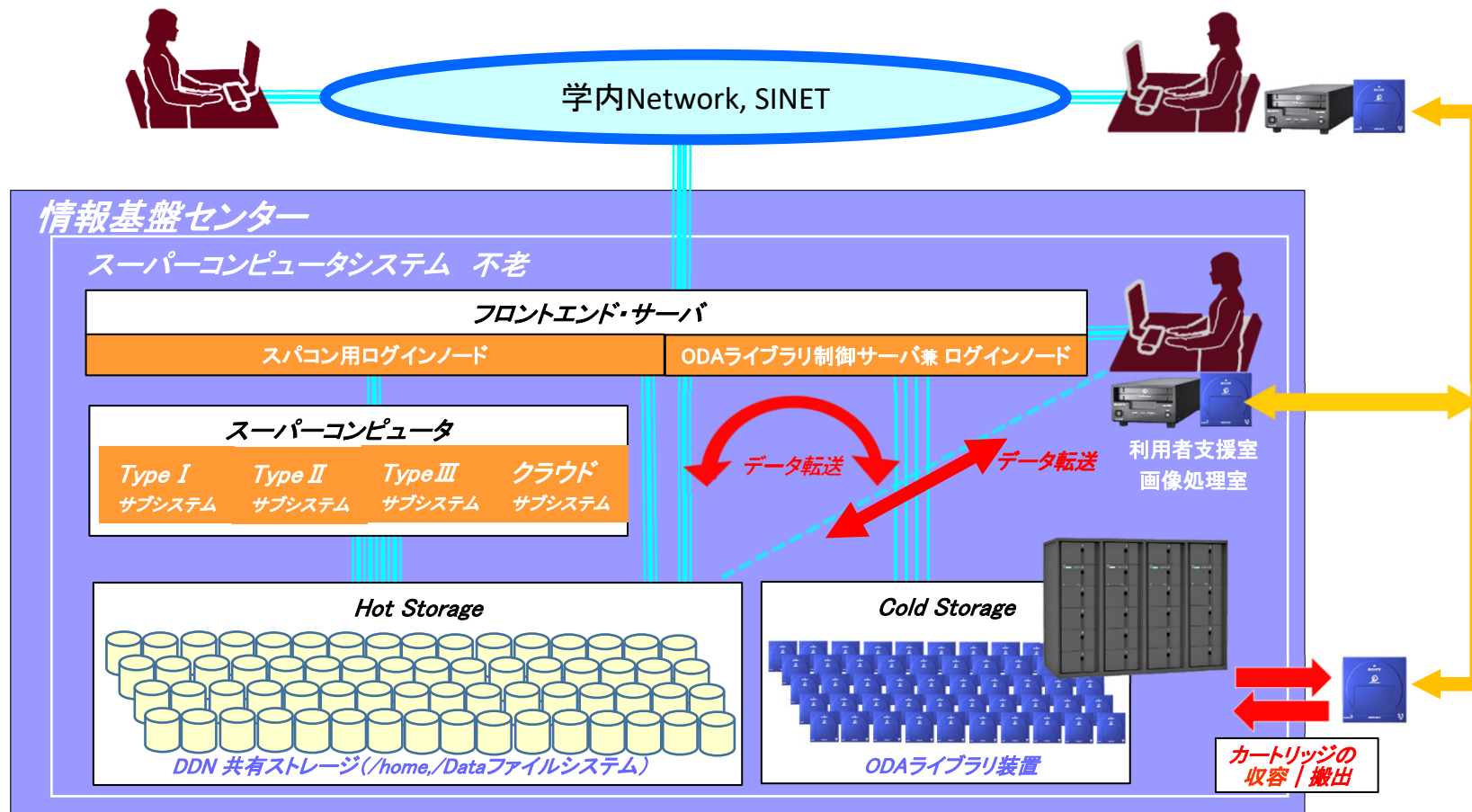
オプティカルディスク・アーカイブ・カートリッジ耐久性実証

<https://www.youtube.com/watch?v=UnlrreBvndg>

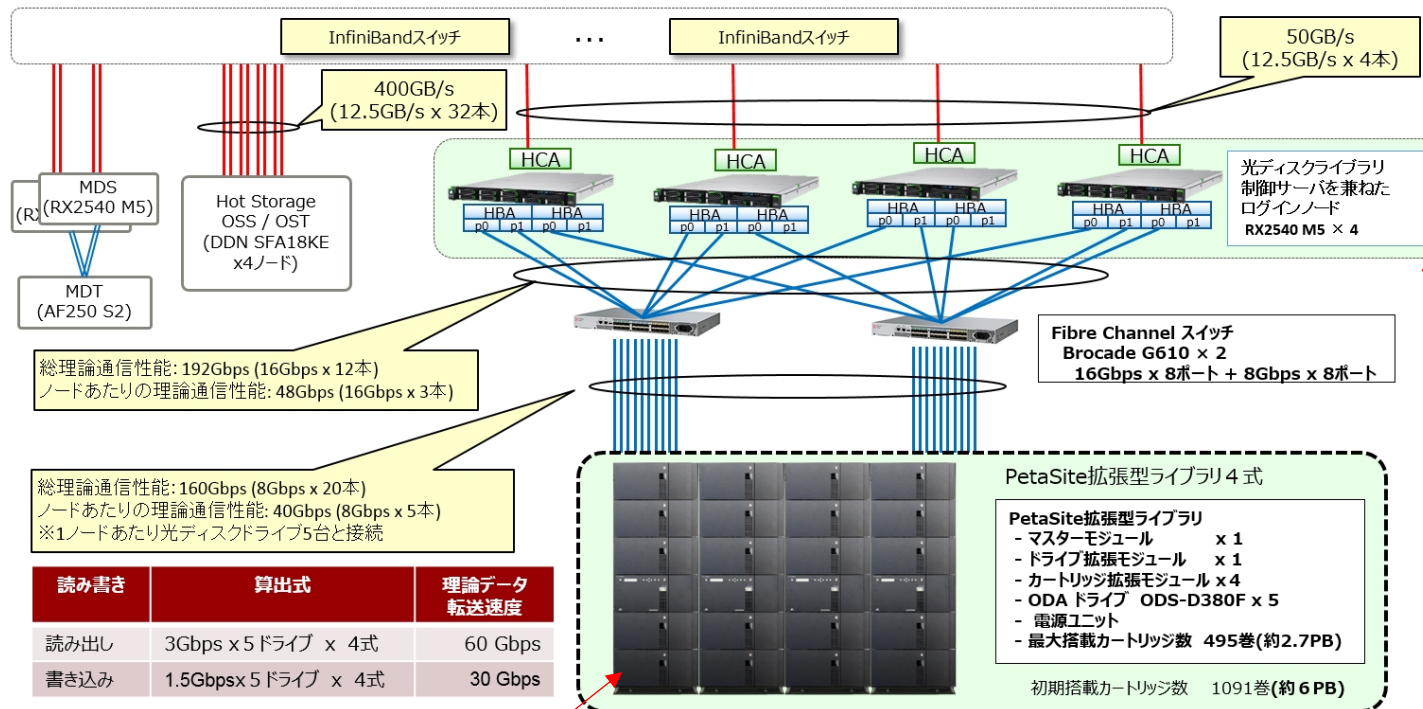
<https://www.youtube.com/watch?v=nGzrGgDMupw>



ODAカートリッジとデータの流れ



コールドストレージシステム構成(Phase2)



ODAライブラリ装置(PetaSite拡張型ライブラリ)は、ODAカートリッジを格納するキャビネット、ドライブ、キャビネット内のODAカートリッジを識別してドライブにセットするロボット機構、ODAカートリッジをライブラリ内に収容および搬出する機構から成る。

このODAライブラリ装置は、Fibre Channelでフロントエンドサーバを兼ねているODAライブラリ制御サーバに接続されている。

単体ドライブを利用する方法

情報基盤センター内の利用者支援室と画像処理室
には、端末としての利用やファイル転送を行うために
ODA単体ドライブがUSB接続されている。
Linux環境とWindows環境で利用できる。

ODS-D380U



Windows環境では、ベンダー提供のWindows Explorer相当のGUIを使った
ファイル操が行えるソフトウェアが利用できる。

Linux環境では、ワークステーションにスパコンのHot Storageの
ファイルシステムが富士通のFEFSでマウントされているため、
Hot Storageとの高速ファイル転送に利用できる。

ODAライブラリ装置を利用する方法

- ① ターミナルソフト(PuTTYやTeraTerm等)を使って、
公開鍵認証で
「指定されたODAのフロントエンドサーバ」にログインします。
- ② ODAライブラリ内のカートリッジへのアクセスは、
センターが提供する次の2種類のバッチジョブ
投入コマンドを使ってジョブを投入して、
ファイル操作を行います。

- インターラクティブ・ジョブで実行(@isubコマンド)
- バッチジョブで実行(@qsubコマンド)



1) バッチジョブ

まず、ファイル操作を行うコマンド列を記載したテキストファイル(バッチファイル)を作成します。そして、そのファイルの名前をジョブ投入コマンドのオペランドに指定してバッチジョブを投入してカートリッジへのアクセスを行います。

2) インターラクティブ・バッチジョブ

ジョブ投入後、一定時間内に対話型でコマンドを入力してカートリッジへのアクセス※1を行います。

※1) センター提供の`odaalloc`コマンドを使って、使用するカートリッジをドライブに取付けてマウントします。

そして、Hot Storage間のファイルコピーやファイル操作を行います。最後に、`odafree`コマンドを使って、アンマウント後カートリッジをキャビネットに戻します。

ODAライブラリ装置およびカートリッジの利用上の留意点

- ① 記録メディアは、Write Onceです。
- ② ODAサーバー当りのドライブユニットの数は、5ドライブです。
サーバーのドライブ数が少ないため、カートリッジへのアクセスは、
バッチジョブを使って利用します。
- ③ 書き出し操作は、シーケンシャル記録のみ。
同一メディアに対しての多重アクセス(メディア内のコピーを含む)は行えない。
- ④ カートリッジ当りの記憶容量は、次のとおりです。
記録容量 : 5TB 程度
作成可能ファイル数 : 80万ファイル程度 ← iノード制限値
ディレクトリおよびファイル階層: 最大64 階層まで(ルートディレクトリは、1階層とカウント)
- ⑤ 使用できるLinuxコマンドに制限があります。

- ⑥ ファイル名およびディレクトリ名の命名規約
- Unicode 2.0 で表現可能な1 文字以上、最大127 文字です。
 - 同一ディレクトリ内に、大文字小文字でファイル名を重複させることはできません。
 - 次の文字は、使用できません。
“ * / : < > ? | ¥ (back slash or Yen) (DEL)
 - .(dot)で始まる文字列は利用できません。
- ⑦ ODAのRead/Write性能は、ファイルサイズに依存します。
ファイルサイズが大きいほど高速になります。ファイルサイズは、50MB以上を推奨します。

ファイルサイズの小さなデータを大量に取り扱う場合は、iノード不足になったりカートリッジへのReadおよびWrite処理に時間がかかります。

このような場合は、複数ファイルをtarコマンドを使ってアーカイブファイルに取りまとめてからカートリッジに記録してください。

※ tarコマンドを使ったアーカイブ & 圧縮 | 解凍コマンドやカートリッジ内のファイル情報を管理するソフトウェアは、ご提供いたします。

- ⑧ マルチボリュームは利用できません。事前にLinuxのsplitコマンドでファイルを分割してから、カートリッジにデータを記録してください。

コールドストレージシステム利用申請区分と利用負担金

利用申請書 (区分)	ファイル利用負担経費 (年間)	内容	備 考
新規申請	管理費: 1口 1万円 ファイル負担経費(光ディスクの生ディスク代金相当): 1口 19万円	カートリッジはセンター提供 1口: 10カートリッジ (50TB相当)	管理費は初年度のみ 利用終了にカートリッジを利用者に返却
継続申請	管理費: 1口 1万円	1口: 10カートリッジ	次年度以降
変更申請	な し	利用メンバーの変更 カートリッジの持ち出し利用 カートリッジの持ち込み利用 利用停止	
カートリッジ 持ち込み申請	管理費: 1口 1万円	カートリッジは利用者持込み 1口: 10カートリッジ(上限)	

ODAライブラリ装置とODAライブラリ制御サーバの諸元

	Phase1	Phase2
ODA制御サーバ数	1	4
ODAライブラリ数	1	4
機種名	PetaSite拡張型ライブラリシステム	
総物理容量	484TB	10.89PB (2.72PB × 4ライブラリ)
総スロット数	88巻	1980巻 (495 × 4ライブラリ)
総ドライブ数	5	20 (5 × 4ライブラリ)
ロボット機構	1	4 (1 × 4ライブラリ)

ODAライブラリ装置の諸元



項 目	仕 様
CPU	Intel Xeon Gold6248(2.5GHz/20コア) × 2 理論演算性能 3.2TFLOPS
主記憶	384GiB (32GiB DDR4 × 12)
内蔵ストレージ	2TB SSD, 19.2 SSD
インターコネクト	InfiniBand EDR × 1
SANインターフェース	FC16Gbps × 4
ネットワークインターフェース	10GTwinax × 2, 1000BASE-T × 2

ODAライブラリ制御サーバ兼ODAフロントエンドサーバの諸元



Phase1は1台、Phase2は4台で構成

光ディスクドライブの仕様

■ ODAライブラリ装置の光ディスクドライブユニットとして「ODS-D380F」

■ ODA単体ドライブの光ディスクドライブユニットとして「ODS-D380U」



仕様		ODS-D380F	ODS-D380U
入出力インターフェース		Fibre Channel 8Gbps	SuperSpeed USB 10 Gbps (USB 3.2 / USB3.0互換)
タイプ	読み出し再生	3Gbps	
	ベリファイ記録	1.5Gbps	
電源		DC 19.5V	
消費電力		約115W	約105W
動作温度		5℃ to 40℃	
動作湿度		20% to 90%	
質量		約4.9Kg	約4.8kg
外形寸法		146x94.2x401.8mm	146x95.5x414.4mm
その他		各ドライブは5.5TBの光ディスクに対応しており、双方で記録したデータは、双方で読み取ることが可能（完全互換）	

ODA ライブラリ操作コマンドの利用方法（センター開発）

コマンド名	機 能
<code>odaalloc</code>	ライブラリ内のカートリッジのマウント処理を行う。
<code>odafree</code>	ライブラリ内のカートリッジのアンマウント処理とDBのファイル情報を更新を行う。
<code>@isub</code>	ライブラリ内のカートリッジにアクセスする インターラクティブ・ジョブ を投入する。
<code>@qsub</code>	ライブラリ内のカートリッジにアクセスする バッチジョブ を投入する。
<code>qstat</code>	ジョブの状態を表示する（ジョブスケジューラ PBSpro のコマンド）。
<code>qdel</code>	ジョブをキャンセルする（ジョブスケジューラ PBSpro のコマンド）。
<code>@vlist</code>	利用可能なライブラリ内のカートリッジの一覧情報を表示する。
<code>@flist</code>	指定したカートリッジのDBに記録しているファイル情報を表示する。 ※ライブラリー内のカートリッジに限る。
<code>@find</code>	利用可能なカートリッジに記録しているファイル情報を、DBを参照してキーワードやワイルドカードで検索する。
<code>@update</code>	指定したカートリッジの最新のファイル情報で、DBのファイル情報を更新を行う。
<code>@tree</code>	<code>@ls</code> コマンドで格納したファイル情報をもとに、ディレクトリ構造のツリー表示を行う。
<code>@ls</code>	指定したディレクトリ配下のディレクトリとファイル情報を、利用者が指定するファイルに格納する。 DBとは別に、利用者が独自にファイル情報を管理することができます。 Hotストレージでも利用できます。
<code>grep</code>	Linuxのgrepコマンド。 <code>@ls</code> コマンドで格納したファイル情報を、キーワードやワイルドカードを使って検索する。

【例題】

ボリューム通番: x49999a001に、saveというディレクトリを作成して、/home上のdata01ディレクトリ全体をsaveにコピーする手順を以下に示します。

【@qsubコマンドを使ったバッチジョブの利用方法】

```
[oda01~]$ @qsub ex1.bat
```

@qsubコマンドのオペランドにファイル操作を記載したバッチファイルの名前を指定してジョブを投入します。
※@subコマンドで投入するジョブの最大実行時間は、24Hです。

バッチファイル ファイル名: ex1.bat

```
odaalloc x49999a001      ← ボリュームのマウント
cd /oda/x49999a/x49999a001 ← ディレクトリの移動
mkdir save              ← ディレクトリ作成
cp -ar /home/x49999a/data01 save/. ← ファイルのコピー
ls -al save             ← lsコマンド
cd .                    ← カレントディレクトリをボリューム外に移動
odafree x49999a001      ← ボリュームのアンマウント
```

【@isubコマンドを使ったインタラクティブジョブの利用方法】

```
[oda01~]$ @isub
```

```
*****
* interactive_job submiied.
*****
qsub: waiting for job 60.oda01 to start
qsub: job 60.oda01 ready
[oda01~]$
[oda01~]$ odaalloc x49999a001
[oda01~]$ cd /oda/x49999a/x49999a001
[oda01~]$ mkdir save
[oda01~]$ cp -ar /home/x49999a/data01 save/.
[oda01~]$ ls -al save
[oda01~]$ odafree x49999a001
[oda01~]$ exit
```

@isubコマンドのジョブが正常に実行されると、qsub: to start、qsub:..... readyメッセージの後に、\$プロンプトが表示され、コマンドが入力できる状態になります。
対話形式でコマンドを入力して、ファイル操作を行います。最後に、exitコマンドを入力してジョブを終了します。

※@isubコマンドで投入するジョブの最大実行時間は、60分です。

※1) 計算依頼したのジョブの操作コマンド

qstatコマンド: ジョブの状態を表示する。

qdelコマンド: 指定したジョブIDのジョブを削除する。

1) odaallocコマンド

コマンドの形式：
`odaalloc` ボリューム名

マウントポイントの形式：
`/oda/ログインID/ボリューム名`

機能：

odaallocコマンドは、マウントポイントを作成して指定したボリュームをマウントします。

次の一連の処理が順に実行され、取付けたドライブ名とマウントポイントが表示されます。

- ・ 指定したカートリッジの所有者の認証
- ・ ドライブの状態確認
- ・ キャビネットからドライブにカートリッジを搬送
- ・ マウントポイントの作成
- ・ カートリッジのマウント処理

【使用例】

[oda01~]\$ `odaalloc x49999a001`

allocate start -> check -> move -> drive -> mount -> termination
oda drive: /dev/sdh … メディアを取付けたドライブ名
mount point: /oda/x49999a/x49999a001 … マウントポイント

【使用例】

ログインIDが「`x49999a`」、マウントするカートリッジのボリューム名「`x49999a009`」の場合のマウントポイント。

`/oda/x49999a/x49999a009a`

2) odafreeコマンド

コマンドの形式：

odafree ボリューム通番

機能：

odafreeコマンドは、指定したカートリッジのアンマウントを行います。

次の一連の処理が順に実行されます。

- ・ 指定したメディアの所有者の認証
- ・ ドライブの状態確認
- ・ アンマウント処理（カートリッジの記録情報をDBに反映）
- ・ ドライブの開放
- ・ カートリッジをキャビネットに搬送
- ・ マウントポイントの消去

odafreeコマンドを実行する前に、必ず、マウントポイント配下のファイルは使用していないことを確認してください。使用しているとBusy状態となり、エラーが発生します。その場合は、原因を対処した後に、再度、odafreeコマンドを入力してください。

【使用例】

[oda01~]\$ [odafree x49999a001](#)

free start-> unmount-> eject-> mount point delete->
termination

3) @vlistコマンド

コマンドの形式 :
[@vlist](#)

機能 :

割当てられているライブラリ内のカートリッジの一覧と空きスペース量を表示します。

【使用例】

[oda01~]\$ [@vlist](#)

Label	Barcode	FreeSize(byte)	iFree
w49000a001		5204767932416	799998
w49001a001		4898986655744	799998
w49001a002		4346016694272	796380
w49001a003		4895680233472	799998

4) @flistコマンド

コマンドの形式：
@flist ボリューム名

機能：

指定したカートリッジのファイル情報を、ODAライブラリのDBを参照して表示します。

【使用例】

[oda01~]\$ @flist w49001a001

```
FilePath
/odatest/500mb
/odatest/100mb
/odatest1/01x10000
/odatest1/1000x5
/odatest2/1mb
/odatest2/10mb
/odatest3/50mb
```

5) @driveコマンド

コマンドの形式 :
`@drive`

機能 :

ODAライブラリの空きドライブ数と、ドライブの利用状況を表示します。

【使用例】

`[@oda01 ~]$ @drive`

空きドライブ数		メディアの固体番号	所有者	ボリューム名	
DriveID	Filesystem	Mounted on	MediaSN	Owner	Label
Empty Drives Count: 4					
1	/dev/sdh		51AA2013WR70416	x49999a	x49999a001
2	/dev/sdi				
3	/dev/sdj				
4	/dev/sdk				
5	/dev/sdg				

6) @findコマンド

コマンドの形式：
`@find キーワード`

機能：

利用可能なカートリッジに記録しているファイル情報を、DBを参照して
キーワードやワイルドカードで検索する。

`[@oda01 ~]$ @find *xy*`

Label	Barcode	FilePath
w40001a001		/test999/Demo/aaa/rxy02.dat
w40001a002		/test999/Demo/bbb/rxy02.001.dat
w40001a003		/test999/Demo/ccc/rxy02.002.data

7) @updateコマンド

コマンドの形式 :
@update ボリューム名

機能 :

マウント状態の指定したカートリッジの現在のファイル情報をDBに反映する。

【使用例】

[oda01~]\$ @update w49001a001

8) @lsコマンド

コマンドの形式：

@ls ディレクトリ名 [出力ファイル名]

機能：

@lsコマンドは、指定したディレクトリ配下のすべての「ディレクトリ情報」と「ファイル情報」を、フルパスの形式で出力ファイルに保存します。
gzip形式で圧縮保存する場合は、出力ファイル名のサフィックスに「.gz」を付けます。
このコマンドは、ODAライブラリシステムのDBとは別に、利用者が独自にファイル情報を管理することができます。 Hotストレージでも利用できます。
次のLinuxコマンドを使って検索および表示を行うことができます。

cat, less, more		zcat, zless, zmore	→	内容を表示
find, grep		zgrep	→	パスを検索
diff, cmp		zdiff, zcmp	→	比較
@tree			→	ディレクトリ構造の表示

【使用例】

```
[@oda01 ~]$ @ls /oda/x49999a/x49999a009 009.20200901.gz
```

```
[@oda01 ~]$ @ls /oda/x49999a/x49999a009 009.20200901.gz | @tree -d
```

9) @treeコマンド

コマンドの形式：

@tree @lsコマンドの出力ファイル名 [-d]

機能：

@treeコマンドは、@lsコマンド又はodafilereadコマンドで作成したファイルを、tree構造で表示します。「-d」オプションを指定すると、ディレクトリ情報のみ表示します。

【使用例】

[oda01~]\$ @tree out.gz

「不老」プリインストールソフト： 精度保証数値計算ライブラリUNC-HPC

「不老」プリインストール: UNC-HPCライブラリ

- ポスト「京」萌芽的課題1 基礎科学のフロンティアー極限への挑戦ー:「極限の探究に資する精度保証付き数値計算学の展開と超高性能計算環境の創成」(代表:荻田武史 東京女子大学 教授、~2019年度終了)で開発

- 精度が保証された計算結果を実用的に得られるような超高性能計算環境を構築

- 以下のライブラリのソースコードを公開中

1. **OzBLAS: Accurate and Reproducible BLAS based on Ozaki scheme** (尾崎スキームに基づく高精度基本線形計算ライブラリ) [for PC, GPU]
2. **GEMMTC: GEMM using Tensor Cores** (Tensorコアを用いた行列積プログラム) [for GPU]
3. **DHPMM_F for GPU: High-precision Matrix Multiplication with Faithful Rounding** (高精度行列積プログラム) [for GPU]
4. **PDDOTK: K-fold Precision Dot Product** (分散並列版高精度内積計算プログラム) [for PC, FX100]
5. **BLAS-DOT2: Higher-precision BLAS based on Dot2** (高精度内積計算アルゴリズムDot2に基づく基本線形計算ライブラリ) [for GPU]
6. **LINSYS_VR: Verified Solution of Linear Systems with Directed Rounding** (連立一次方程式に対する精度保証プログラム(丸めモードの変更機能付き)) [for K Computer, FX100]
7. **LINSYS_V: Verified Solution of Linear Systems** (連立一次方程式に対する精度保証プログラム) [for PC, K Computer, FX100]
8. **DHPMM_F: High-precision Matrix Multiplication with Faithful Rounding** (高精度行列積プログラム) [for PC, K Computer, FX100]

HP: <http://www.math.twcu.ac.jp/ogita/post-k/index.html>

Verified Numerical Computations

UNC-HPC

High-Performance Computing

ポスト「京」萌芽的課題1
基礎科学のフロンティアー極限への挑戦
極限の探究に資する精度保証付き数値計算学の展開と超高性能計算環境の創成

トップページ Top	課題概要 Abstract	研究体制 System	研究内容 Contents	成果公開 Results
---------------	------------------	----------------	------------------	-----------------

リンク

- ・ トップページ
- ・ 課題概要
- ・ 研究体制
- ・ 研究内容
- ・ 成果公開

課題概要

本研究課題の目的は、ポスト「京」において、精度が保証された計算結果を実用的に得られるような超高性能計算環境を構築することです。具体的には、ポスト「京」で実行される様々な数値シミュレーションにおいて、数値計算による計算誤差の問題が解消されることによって、シミュレーションサイエンスの品質を向上させ、さらに想定外の現象が発生する可能性を低減することが可能となります。すなわち、本研究課題の遂行は、人が安心して生活できる社会基盤の構築に直結します。

たとえば、産業界における製品開発や非破壊検査等のための計算工学シミュレーションや、地震や津波などの災害シミュレーションは、日本に限らず世界中で盛んに行われていますが、計算誤差の観点から高精度なシミュレーションの実現をしている例は皆無です。これは、計算機によって問題の近似的な解を得ることよりも、その近似解の検証のほうがはるかに困難かつ計算資源を必要とする、と考えられており、それが高性能計算分野の常識であるからです。実際、これは1990年代までは事実でしたが、本研究グループが2000年代から切り拓いてきた高速で実用的な精度保証付き数値計算法やエラーフリー変換に基づく高精度数値計算法をベースとして、今、ポスト「京」によって、この常識を打ち破る時期が到来しています。すなわち、ポスト「京」において、高速性と高精度性を融合した超高性能計算環境の創成を世界に先駆けて達成することは、スーパーコンピュータに質的転換をもたらし、我が国の高い科学技術力を国内外に示すこととなります。

精度
(高信頼性)

エネルギー効率(省エネ)
Green500,
Green Graph500, ...

速度(高速性)
Top500, Graph500, ...

性能 = 速度 × 精度

大規模数値計算を行う高性能計算分野においては、問題の大規模化に伴って計算誤差が累積しやすくなり、これが今後、大きな問題となってきます。たとえば、計算機上で標準的に用いられる32ビットの浮動小数点演算では、100万次元程度の密行列系線形問題に対して数値計算を行うと、問題自身は比較的良好条件で解きやすい問題であったとしても、誤差解析の結果から理論的には1桁も正しくないような計算結果が得られることが分かっています。また、計算誤差の単なる累積だけでなく、問題の困難さが条件数として

尾崎の方法の概要

- 行列 A と行列 B を分解する:

$$A = A^{(1)} + A^{(2)} + A^{(3)} + \dots + A^{(p)}$$

$$B = B^{(1)} + B^{(2)} + B^{(3)} + \dots + B^{(q)}$$

- インデックスが若いほうが、高いビットを持つ
- 行列積 AB は下記のように書ける

$$AB = (A^{(1)} + A^{(2)} + \dots + A^{(p)})(B^{(1)} + B^{(2)} + \dots + B^{(q)})$$

尾崎の方法の概要

- 以下、単純に展開をする

$$\begin{aligned} AB &= (A^{(1)} + A^{(2)} + \dots + A^{(p)})(B^{(1)} + B^{(2)} + \dots + B^{(q)}) \\ &= A^{(1)} B^{(1)} + A^{(1)} B^{(2)} + A^{(2)} B^{(1)} + \dots + A^{(p)} B^{(q)} \end{aligned}$$

ここで

$$fl(A^{(i)} B^{(j)}) = A^{(i)} B^{(j)} \text{ for } 1 \leq i \leq p, 1 \leq j \leq q.$$

$$\begin{aligned} &= fl(A^{(1)} B^{(1)}) + fl(A^{(1)} B^{(2)}) + fl(A^{(2)} B^{(1)}) + \dots + fl(A^{(p)} \\ &B^{(q)}) \end{aligned}$$

$$= C_1 + C_2 + \dots + C_{pq}$$

- 行列積の回数は pq 回

例題

以下，倍精度浮動小数点数を用いるとする.

$$A = \begin{bmatrix} 2^0 + 2^{-30} & 2^{-20} + 2^{-40} & 2^{-10} + 2^{-15} \\ 2^{-10} + 2^{-30} & 2^{-15} + 2^{-35} & 2^1 + 2^{-15} \\ 2^{-5} + 2^{-25} & 2^0 + 2^{-30} & 2^5 + 2^{-30} \end{bmatrix},$$

$$B = \begin{bmatrix} 2^0 + 2^{-30} & 2^{-35} + 2^{-60} & 2^5 + 2^{-15} \\ 2^{-5} + 2^{-10} & 2^{-30} + 2^{-40} & 2^{-10} + 2^{-30} \\ 2^{-3} + 2^{-15} & 2^{-40} + 2^{-50} & 2^{-8} + 2^{-17} \end{bmatrix}$$

丸め誤差

- 例えばAの1行とBの1列の内積を考え、
A(1,1)とB(1,1)の積を計算すれば
 $(2^0 + 2^{-30})(2^0 + 2^{-30}) = 2^0 + 2^{-29} + 2^{-60}$
であり、赤字の部分は誤差として捨てられる.
- 他の部分も同様に、53ビットで収まらない情報は捨てられる

無誤差変換例

$$\bullet \quad A = \begin{bmatrix} 2^0 + 2^{-30} & 2^{-20} + 2^{-40} & 2^{-10} + 2^{-15} \\ 2^{-10} + 2^{-30} & 2^{-15} + 2^{-35} & 2^1 + 2^{-15} \\ 2^{-5} + 2^{-25} & 2^0 + 2^{-30} & 2^5 + 2^{-30} \end{bmatrix}$$

を2つの行列に分ける

$$\bullet \quad A^{(1)} = \begin{bmatrix} 2^0 & 2^{-20} & 2^{-10} + 2^{-15} \\ 2^{-10} & 2^{-15} & 2^1 + 2^{-15} \\ 2^{-5} & 2^0 & 2^5 \end{bmatrix}, \quad A^{(2)} = \begin{bmatrix} 2^{-30} & 2^{-40} & 0 \\ 2^{-30} & 2^{-35} & 0 \\ 2^{-25} & 2^{-30} & 2^{-30} \end{bmatrix}$$

無誤差変換例

- $$B = \begin{bmatrix} 2^0 + 2^{-30} & 2^{-35} + 2^{-60} & 2^5 + 2^{-15} \\ 2^{-5} + 2^{-10} & 2^{-30} + 2^{-40} & 2^{-10} + 2^{-30} \\ 2^{-3} + 2^{-15} & 2^{-40} + 2^{-50} & 2^{-8} + 2^{-17} \end{bmatrix}$$

を2つの行列に分ける

- $$B^{(1)} = \begin{bmatrix} 2^0 & 2^{-35} & 2^5 + 2^{-15} \\ 2^{-5} + 2^{-10} & 2^{-30} + 2^{-40} & 2^{-10} \\ 2^{-3} + 2^{-15} & 2^{-40} + 2^{-50} & 2^{-8} + 2^{-17} \end{bmatrix}$$

- $$B^{(2)} = \begin{bmatrix} 2^{-30} & 2^{-60} & 0 \\ 0 & 0 & 2^{-30} \\ 0 & 0 & 0 \end{bmatrix}$$

精度保証演算例

$$\begin{aligned} \bullet \quad A^{(1)} &= \begin{bmatrix} 2^0 & 2^{-20} & 2^{-10} + 2^{-15} \\ 2^{-10} & 2^{-15} & 2^1 + 2^{-15} \\ \color{red}{2^{-5}} & \color{red}{2^0} & \color{red}{2^5} \end{bmatrix} \text{と} \\ \bullet \quad B^{(1)} &= \begin{bmatrix} \color{red}{2^0} & 2^{-35} & 2^5 + 2^{-15} \\ \color{red}{2^{-5} + 2^{-10}} & 2^{-30} + 2^{-40} & 2^{-10} \\ \color{red}{2^{-3} + 2^{-15}} & 2^{-40} + 2^{-50} & 2^{-8} + 2^{-17} \end{bmatrix} \end{aligned}$$

この行列の積(例えば3行と1列の積)は

$$2^{-5} * 2^0 + 2^0(2^{-5} + 2^{-10}) + 2^5(2^{-3} + 2^{-15}) = 2^2 + 2^{-4} + 2^{-9}$$

先頭から53ビット以内の計算であり, 誤差はない.

(他も内積も同様であり, $A^{(1)}B^{(2)}$, $A^{(2)}B^{(1)}$, $A^{(2)}B^{(2)}$ も同様)

尾崎の方法の実装

- AB は $A^{(1)}B^{(2)} + A^{(1)}B^{(2)} + A^{(2)}B^{(1)} + A^{(2)}B^{(2)}$

この4つの積を浮動小数点演算で計算しても誤差は発生しない.

- AB は4つの行列の積にBLASのdgemmを使用して計算可能である.
- その後、和を高精度に計算する
(高精度和アルゴリズムを利用)

試験行列

- A : 単位行列と, ある疎度分の要素に対して $0 \sim 1$ の範囲で生成した値で置き換えた行列
- B : A の逆行列
- 行列サイズ: 10, 50, 100, 500, 1000, 5000
- 疎度: 90%

試験行列における演算精度 (提案: batched BLASによる実装)

サイズ	10	50	100	500	1000	5000
DGEMM(絶対誤差)	8.77e-17	8.18e-16	1.70e-15	2.94e-15	5.33e-15	5.14e-15
DGEMM(相対誤差)	2.11e-10	1.82e-8	7.04e-7	2.71e-5	1.92e-5	3.42
従来(絶対誤差)	8.77e-17	1.11e-16	1.10e-16	1.11e-16	1.11e-16	1.11e-16
従来(相対誤差)	8.77e-17	1.11e-16	1.10e-16	1.11e-16	1.11e-16	1.11e-16
提案(絶対誤差)	8.77e-17	1.11e-16	1.10e-16	1.11e-16	1.11e-16	1.11e-16
提案(相対誤差)	8.77e-17	1.11e-16	1.10e-16	1.11e-16	1.11e-16	1.11e-16

Performance Evaluation of Accurate Matrix–Matrix Multiplication on GPU Using Sparse Matrix Multiplications

Fumiya Ishiguro¹, Takahiro Katagiri², Satoshi Ohshima²,
Toru Nagai²

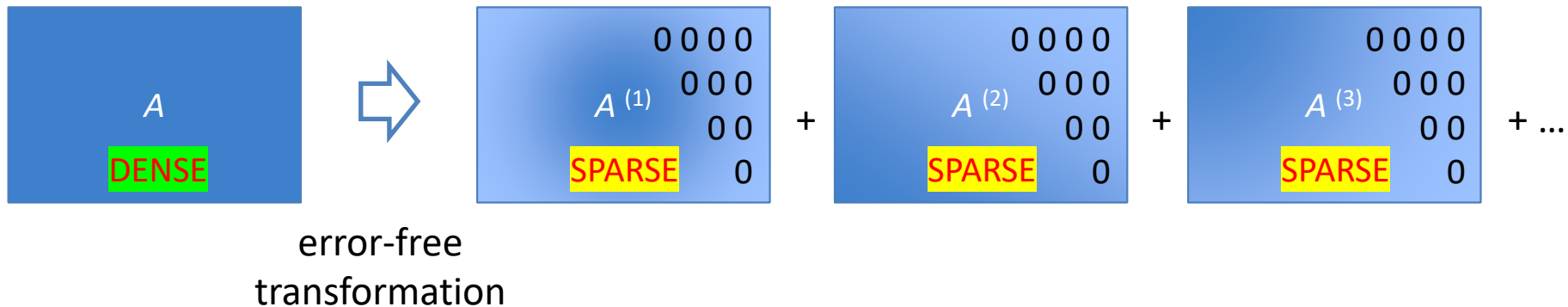
1: Graduate School of Informatics, Nagoya University

2: Information Technology Center, Nagoya University

7th International Workshop on Large-scale HPC Application
Modernization (one of co-located workshop of CANDAR'20)
2020.11.25 online

Previous work and Proposed method

- When the value ranges of the input matrix elements are large, error-free transformation generates many sparse matrices.



- In this case, many double precision general matrix - matrix multiplication (dgemm) are performed. Transforming dense matrices into sparse matrices and **performing sparse matrix operations will require shorter calculation time than dense matrix operations.**
- Therefore, our previous work proposed to transform the target matrices into sparse matrices and calculate sparse matrix computations **on CPU.**
 - Considering the performance, sparse matrix - vector multiplications (SpMV) are used.
- In this study, we propose to calculate these sparse matrix computations **on GPU.**

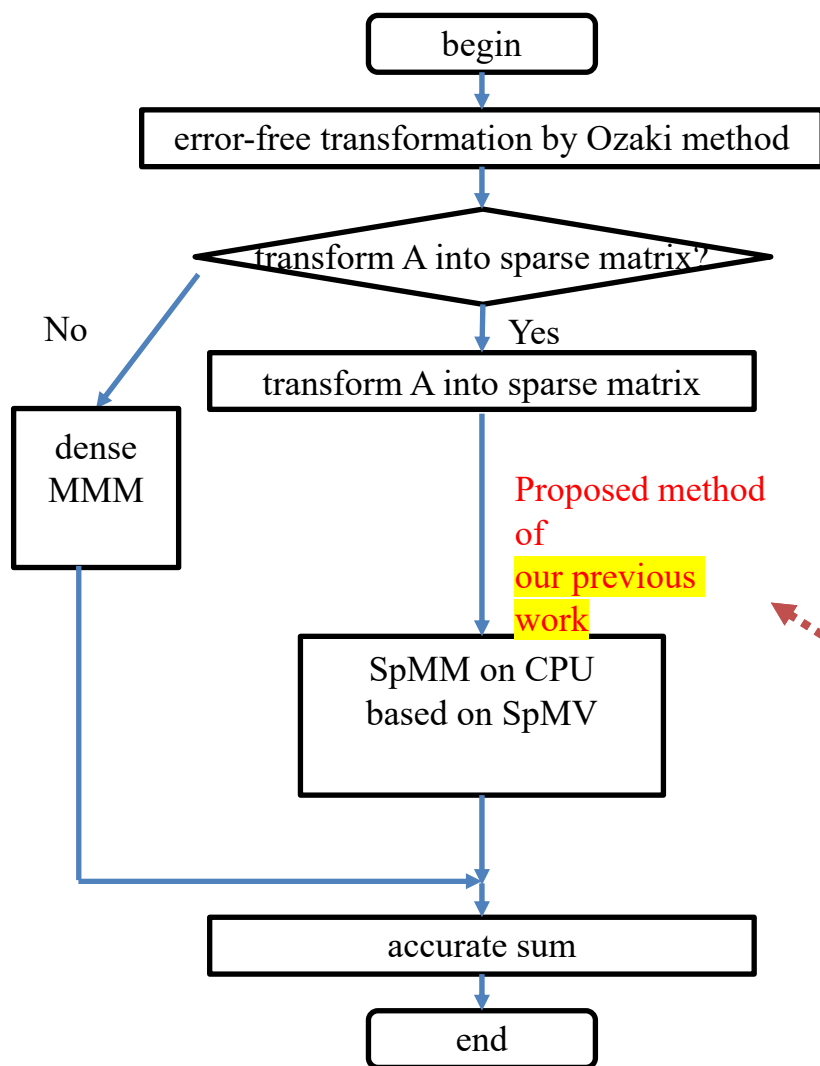
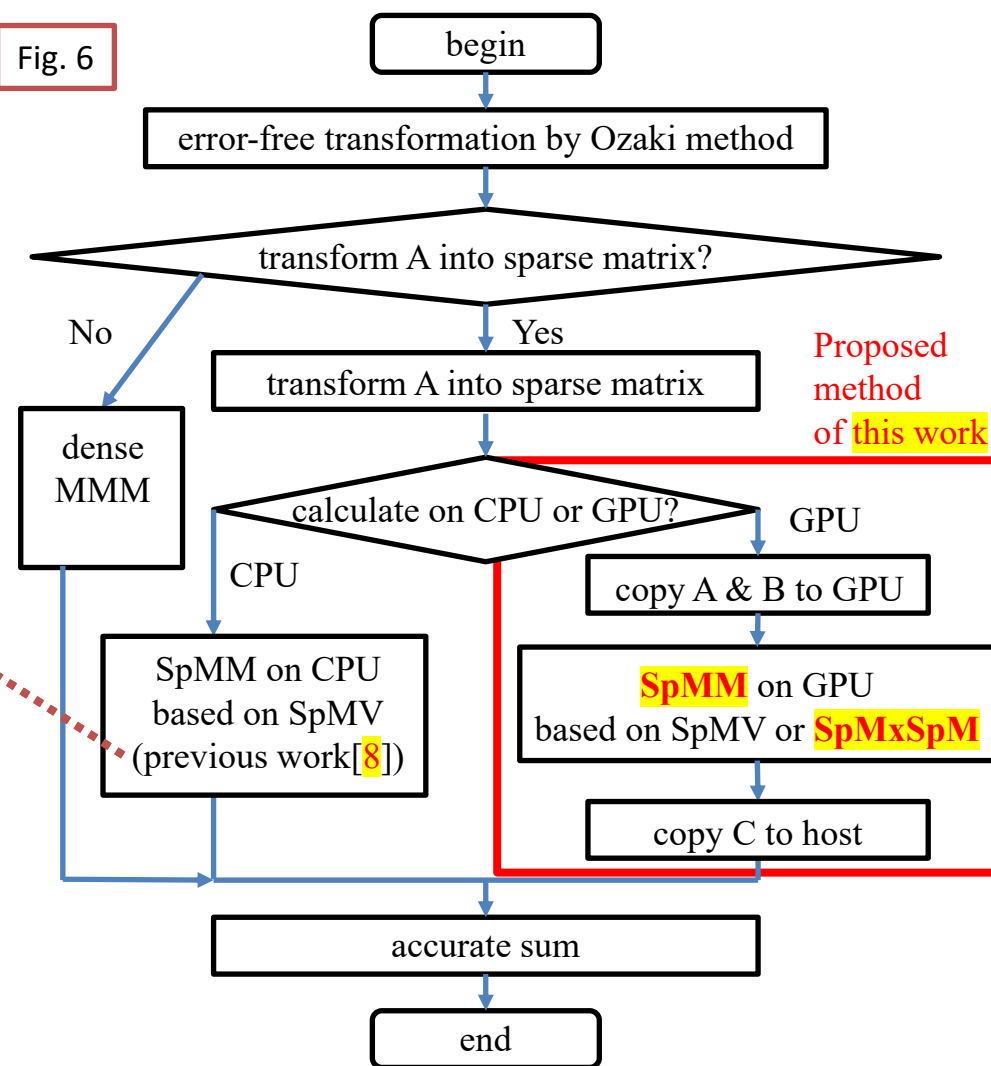
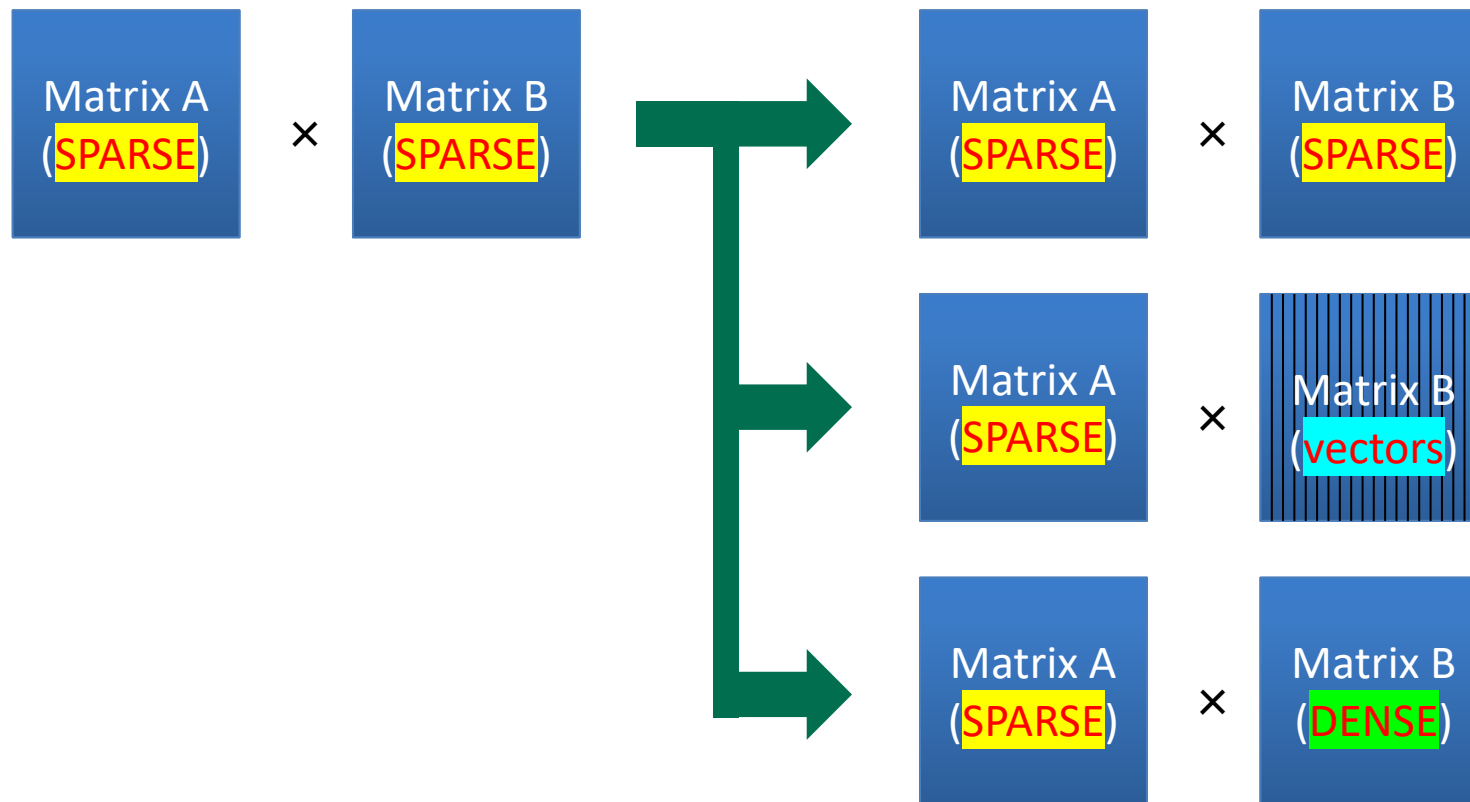


Fig. 6



How to calculate Sparse Matrix A - Matrix B multiplication?

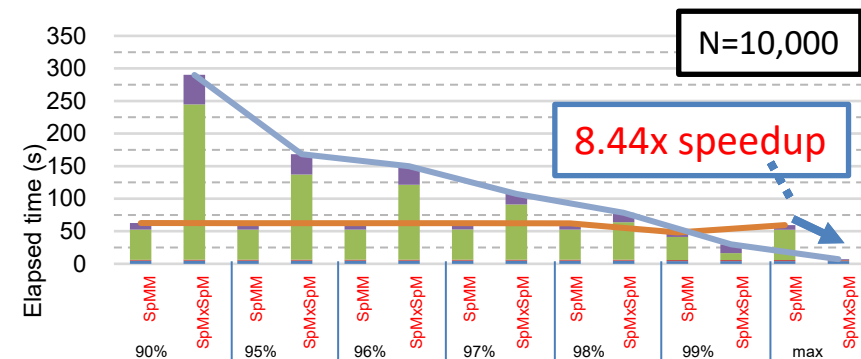
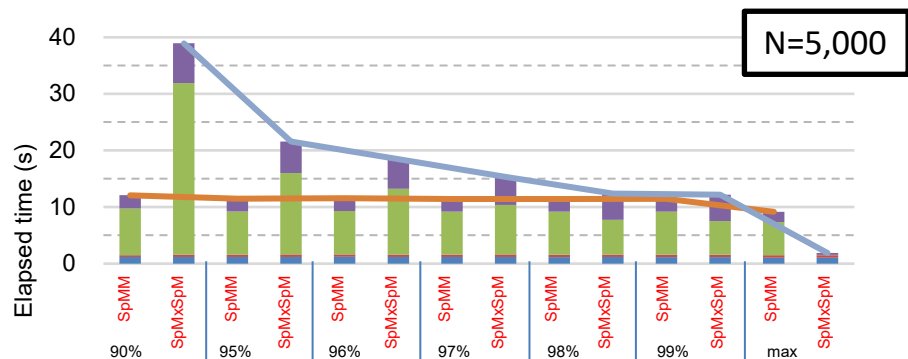
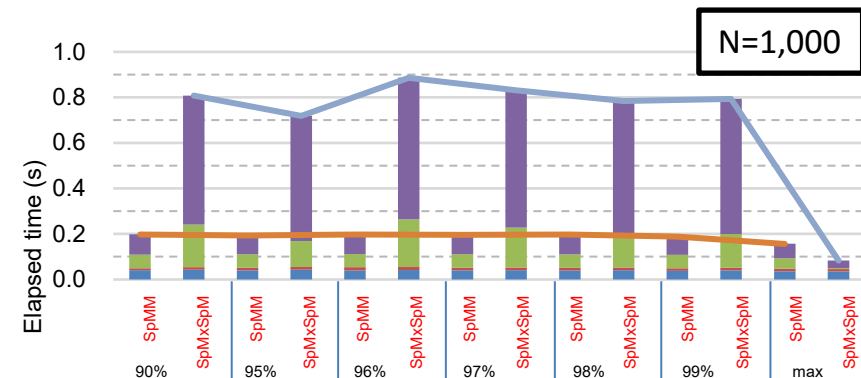
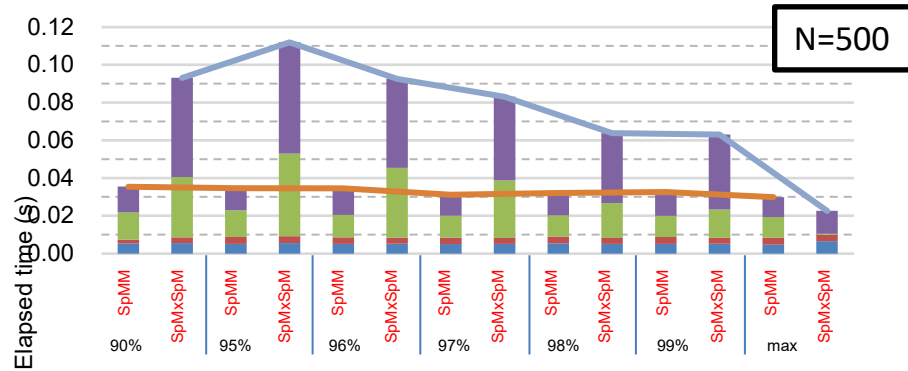


Performance evaluation

- Evaluation environment
 - [Reedbush-H supercomputer](#), 1 node
 - installed at the Information Technology Center, The University of Tokyo
 - each nodes has
Intel Xeon E5-2695 v4 ([Broadwell-EP, 18cores](#)) * 2 + [NVIDIA Tesla P100](#) * 2
 - Intel C++ Compiler ver.18.1.163 + CUDA 9.2.148
- Test conditions
 - All matrix–matrix multiplications are performed with double precision.
 - On the CPU, 36 threads per node are used.
 - "export KMP AFFINITY=granularity=fine,compact,1,0" is specified in the job script for NUMA Affinity and the default values are chosen for NUMA policies.
 - On the GPU, cuSPARSE is used for matrix computations.
 - [SpMV](#) (Sparse Matrix * Vector) / CRS format: [cusparsedcsmv](#)
 - [SpMV](#) (Sparse Matrix * Vector) / ELL format: [cusparsedhybm](#)
 - [SpMM](#) (Sparse Matrix * Dense Matrix) / CRS format: [cusparsedcsmm2](#)
 - [SpMxSpM](#) (Sparse Matrix * Sparse Matrix) / CRS format: [cusparsedcsmm](#)

Result2: Execution time of accurate MMM using SpMM and SpMxSpM in CRS on GPU for Matrix set 2 (2/2)

■ error-free conversion
 ■ conversion matrix A and CPU->GPU
 ■ GPU kernel
 ■ memory management of matrix B, C
 ■ others
 — SpMM — SpMxSpM



概要

1. スーパーコンピュータ「不老」導入
2. 名古屋大学情報基盤センターの
民間利用制度
3. センター独自開発ソフトウェア
4. **おわりに**

おわりに

- ▶ 名古屋大学情報基盤センター **スーパーコンピュータ「不老」**
- ▶ **数値シミュレーションとデータサイエンスを融合可能な設計と運用**
 - 1. Type I サブシステム (「富岳」型ノード) ⇒ 超並列処理用、国策スパコン連結
 - 2. Type II サブシステム ⇒ 機械学習、高速ローカルディスク
 - 3. Type III サブシステム ⇒ 48TBの大規模共有メモリによるプレ／ポスト／可視化処理
 - 4. クラウド ⇒ PCクラスタ利用、時刻指定予約
 - 5. 可視化システム ⇒ 詳細可視化ディスプレイ連結、リモート可視化
 - 6. 大規模共有ホットストレージ ⇒ 1～5に連結、シームレスなデータ移動
 - 7. コールドストレージ ⇒ データを100年保存可能、簡便な利用コマンドをセンターで開発、光ディスク持ち込みも可能、(業界初) サービス終了時に光ディスクを持ち帰り可能
- ▶ **オープンソースソフトウェア(OSS)・国策開発ソフトの利用促進**
- ▶ **民間利用制度(公開、非公開)による利用促進**
 - ▶ AI分野で利用促進を期待(GPU、大規模データ(ホット／コールド ストレージ)利活用)
- ▶ **センター独自開発ソフトウェアの開発・普及**
 - ▶ ツール(可視化、コールドストレージ操作)、数値計算ライブラリ