



東北大学



普通の人々のためのスーパーコンピュータ センターを目指して

東北大学サイバーサイエンスセンターの
高性能計算に関する研究開発活動

小林広明

センター長・教授

東北大学サイバーサイエンスセンター

koba@cc.tohoku.ac.jp

2015年3月20日

Cyber HPC Symposium

本日の内容

- ★ HPCに関する全国共同利用・共同研究拠点としてのサイバーサイエンスセンターの取り組み
- ★ なぜ、「普通の人々のためのスーパーコンピュータセンター」が必要なのか？
- ★ そのための取り組み：
 - 演算性能とメモリ性能がバランスした新スーパーコンピュータシステムの導入、ユーザコードの早期移行&性能評価
 - スーパーコンピュータシステムを活用したリアルタイム津波浸水予測システムの研究開発
 - 高メモリバンド幅アプリケーションのための将来のスーパーコンピュータシステムに関する研究開発

東北大学サイバーサイエンスセンターの歴史

★ 1969年に日本で2番目の全国共同利用型大型計算機センターとして設立

- 汎用大型計算機の運用

★ 1985年よりスーパーコンピュータセンターとして活動

- 大規模科学計算システムとしてベクトル型のスーパーコンピュータを運用

★ 2008年に情報シナジーセンターからサイバーサイエンスセンターへ改組

- 2010年より「学際大規模情報基盤共同利用・共同研究拠点」として文科省より認定を受け、2012年よりHPCIに資源提供を開始し、HPCに関する共同利用・共同研究拠点活動を強化

新システム設計



運用/
利用支援/
教育

研究開発



SENAC-1 in 1958



First Computer in 1969



ACOS 1000 in 1982



SX-1 in 1985



SX-4 in 1998



SX-2 in 1989



SX-7 in 2003



SX-3 in 1994



SX-9 in 2008

センター独自の取組

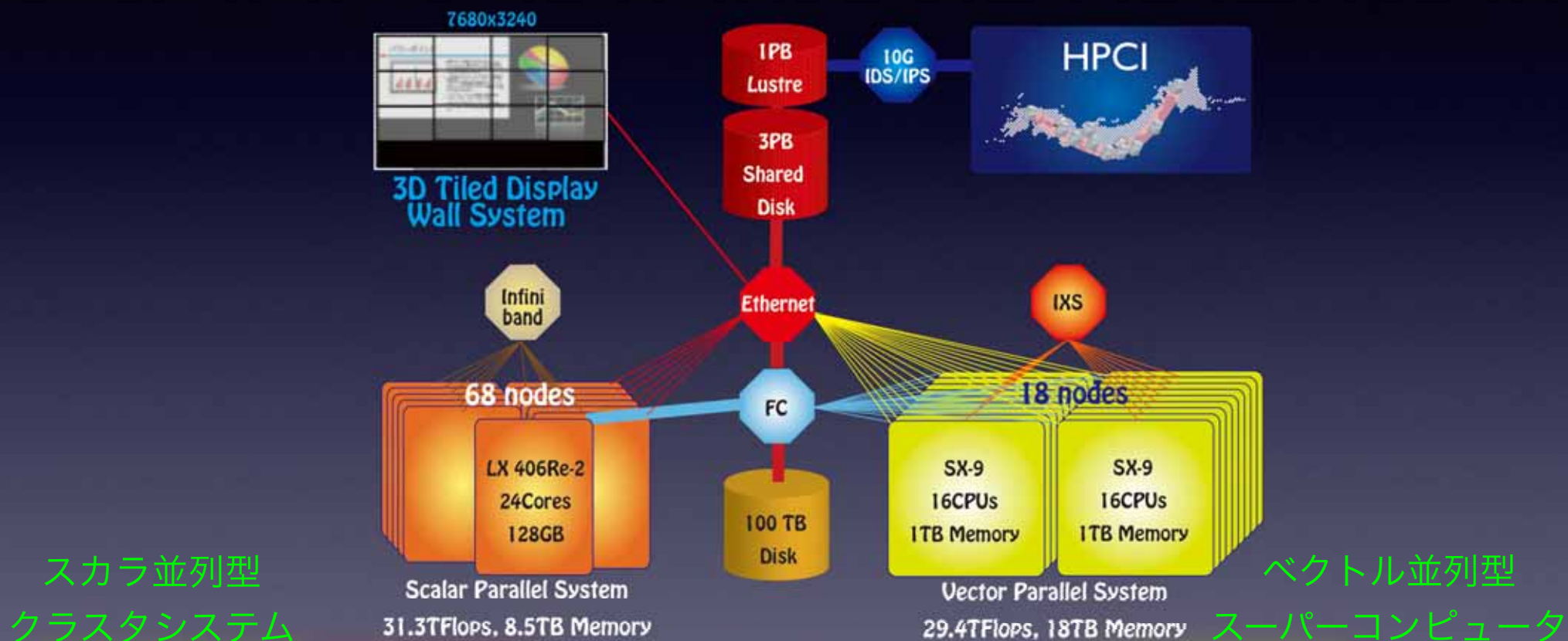
- 高性能計算に関する全国共同利用・共同研究拠点活動
 - ◆ 世界最高クラスの大規模科学計算システムの整備・運用
 - ✓ ベクトル並列型スーパーコンピュータ
 - ✓ スカラ並列型スーパーコンピュータ
 - ◆ 利用者支援・若手研究者育成
 - ✓ プログラム開発支援
 - ✓ 講習会の実施
 - ✓ 産業利用支援
 - ◆ 次世代高性能計算基盤の研究開発
 - ✓ センター教職員と利用者による共同研究の公募・実施
 - ✓ 先端情報基盤とその利用技術に関する研究の推進
 - 高性能計算基盤の研究開発に関する組織間協定の締結
 - 国内外の機関との共同研究の実施
 - ◆ 学部学生・大学院教育
 - ✓ 工学部, 大学院情報科学研究科・医工学研究科・工学研究科の協力講座として実践的・臨床的教育研究の推進

外部連携の中での中核的取組

- 学際大規模情報基盤共同利用・共同研究拠点活動
 - ◆ 7大学基盤センター＋東工大によるネットワーク型拠点形成
 - ◆ 共同研究の公募・実施
 - ◆ シンポジウム開催(年2回)
- 
-
- HPCIコンソーシアム活動 
 - ◆ 利用者視点で我が国のHPC環境の整備、利用者コミュニティの形成
 - ✓ HPCIシステム設計・運用
 - ✓ HPCI課題研究者支援

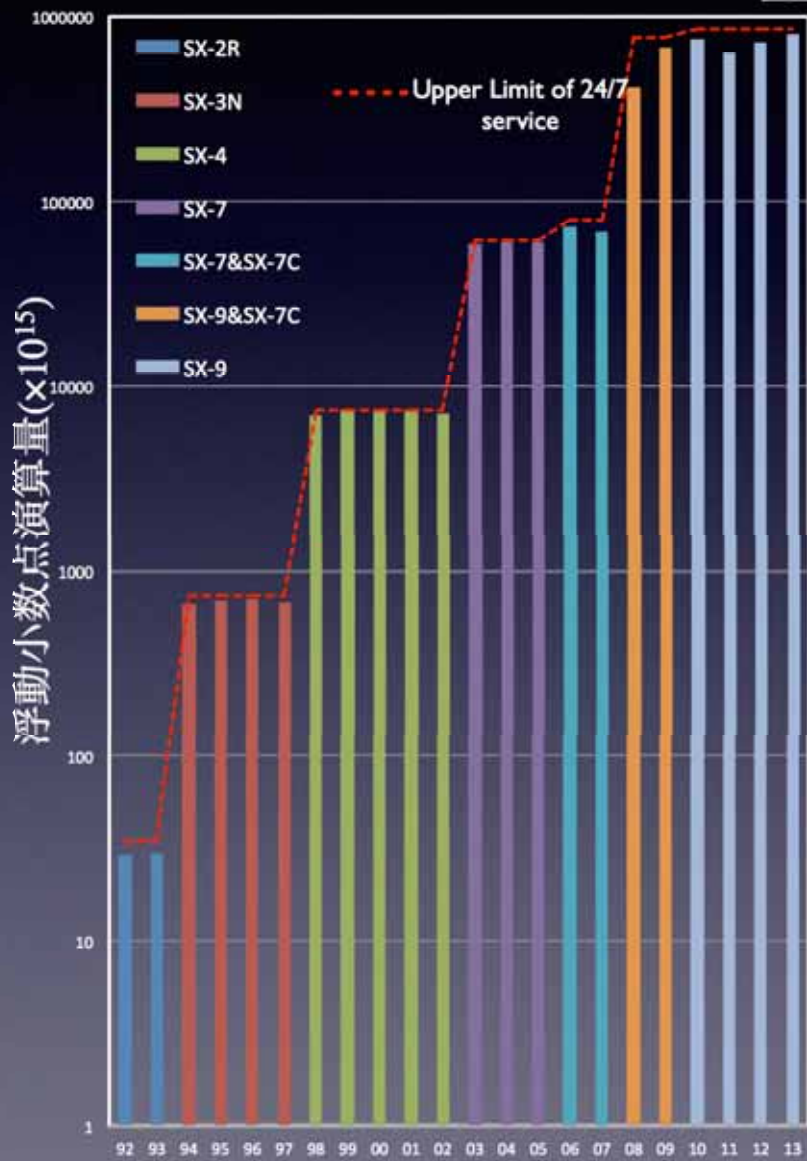
東北大学大規模科学計算システム (～H27.2.10)

最先端のスーパーコンピュータ設備の整備・運用

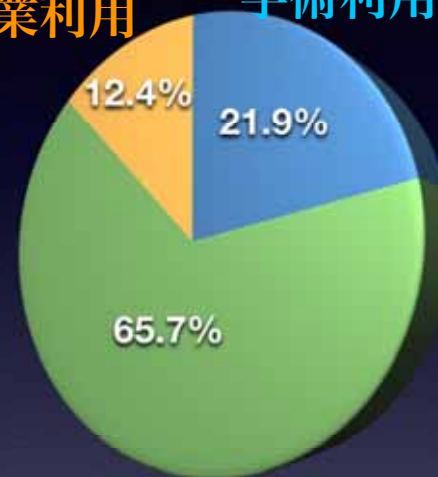


東北大学におけるスパコンサービス状況と 主な利活用分野

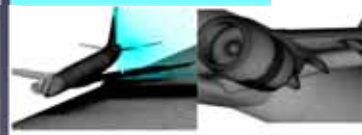
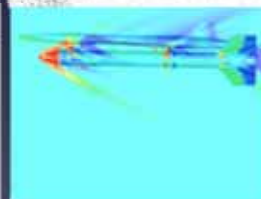
シミュレーションに活用された



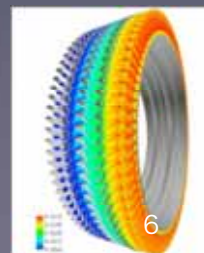
産業利用 学術利用 (学内)



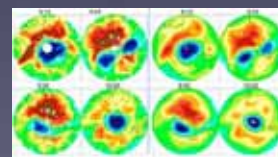
航空機設計



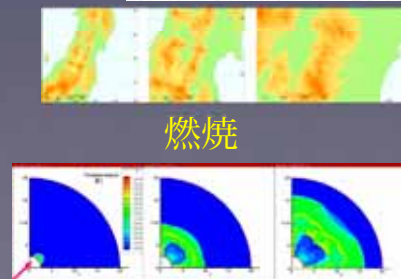
ターボ機械



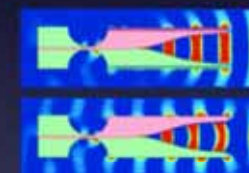
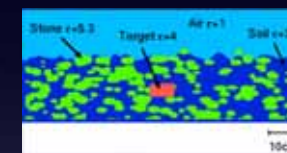
気象・気候



燃焼



アンテナ・電磁解析



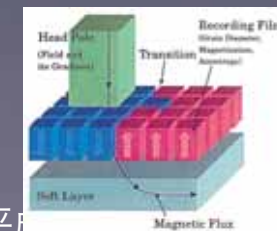
材料・物性



地震



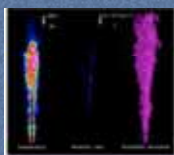
磁気デバイス



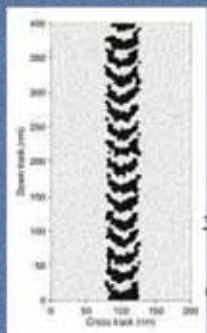


MRJ

MHI



Eco-wave



TDK

国際競争力のある製品開発へ貢献



毎年10%程度を
民間利用支援へ

国産小型民間航空機空力設計への大規模空力解析技術の適用 (三菱航空株式会社)

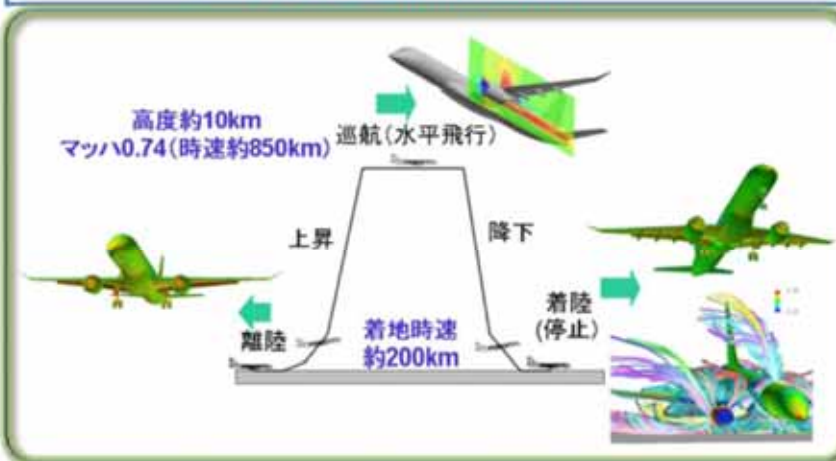
国内初のジェット小型航空機(MRJ)開発

環境負荷低減の為、同クラスで先行するジェット旅客機の燃費に対して、機体の軽量化・低抵抗化と新エンジンの搭載を含めて2割程度の燃費削減を目標とした機体MRJ (Mitsubishi Regional Jet)の開発

平成20年度より、東北大学のSX-9システムと東北大学工学研究科中橋教授が開発したTAS-Code (Tohoku Univ. Aerodynamic Simulation Code)を用いた全機全ミッションCFD解析を実施し、飛行試験前の最終版となる機体性能計算のための空力データを作成。MRJ開発スケジュールの維持に貢献

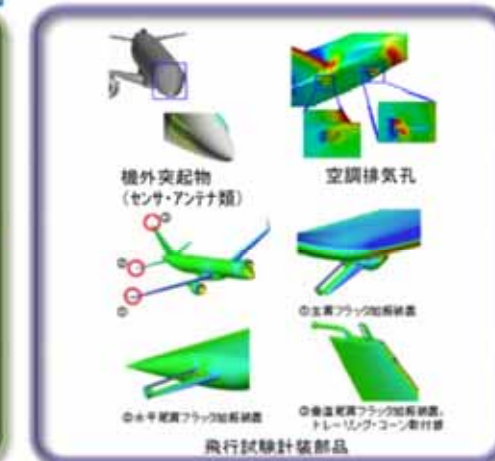


MRJ(参照: <http://www.mrj-japan.com/>)



全機全ミッションCFD解析

CFDの要素技術開発(非構造格子法の格子生成・解析技術の高度化、大規模並列解析等)によって、離陸⇒上昇⇒巡航⇒降下⇒着陸⇒停止という航空機の一連の運用ミッションに適用可能とし、全機全ミッションCFD解析技術を確立



詳細部品模擬CFD

従来は風洞試験模型でも模擬が困難で、従来は理論及び経験に基づく簡易推算に頼っていた機体装備品の詳細部品をCFD解析において模擬し、機体構造設計のための詳細局所空力荷重設定や機体抵抗(総抵抗)の高精度化に寄与

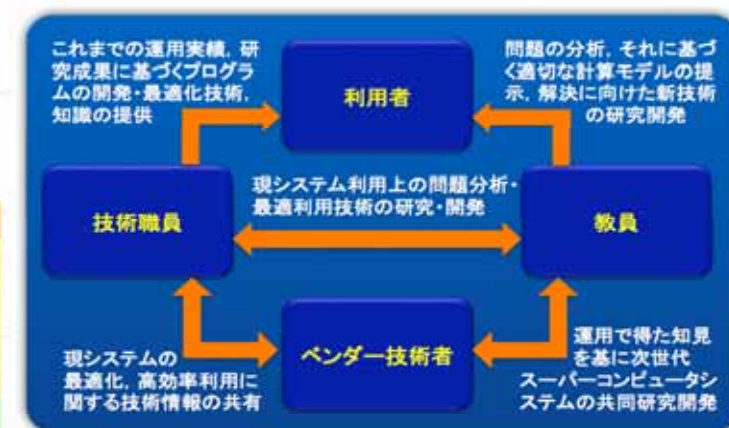
三菱航空機との航空機設計・開発に関する共同研究

TV program “Close-up GENDAI” by NHK (2013.1.8)



利用者支援機能の強化：プログラムの高速化に関する取り組み

- センター職員・教員・ベンダエンジニアによる高速化支援体制を平成9年に整備し，ユーザ支援，および計算科学・計算機科学的共同研究を実施



これまで約200件の共同研究を推進，高速化支援対象となるコード数も増加

年	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24	H25
件数	8	9	10	7	18	20	8	16	10	15	8	8	13	6	11
単体性能向上比	4.5	2.5	1.6	2.2	6.7	2.9	1.5	2.9	33	9.3	381	47	16.2	19.7	16.7
並列性能向上比	31.7	8.6	4.9	2.8	18.6	4.5	4.1	8.1	1.9	5.1	3.6	48	17.2	15.3	12.9

利用者プログラム開発支援、共同研究の取り組み

★平成9年よりセンター職員・教員・ベンダエンジニア
による高速化支援体制を構築し、ユーザ支援を強化

- ・特に若手技術者育成を目指して、職員あたり常時2,
3本のコード最適化を担当

大泉健治



小野敏



□ 高速化支援取りまとめ

山下毅



- ナノ粒子生成過程解析 (阪大茂田先生)
- 乱流解析 (工学研究科伊澤先生)
- 他

佐々木大輔



- 熱中症シミュレータ (名古屋工業大学平田先生)
- 第一原理材料シミュレータ (産総研石橋様)

森谷友映

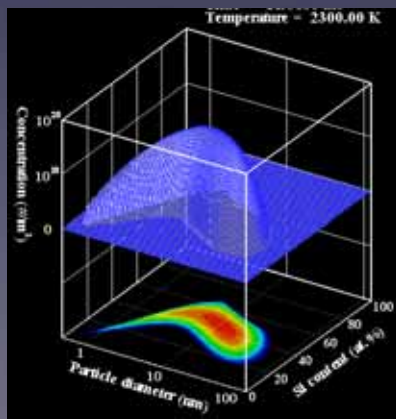


- 電子系・原子核系シミュレータ (産総研宮本様)
- 先進複合材料シミュレーション (工学研究科岡部先生)

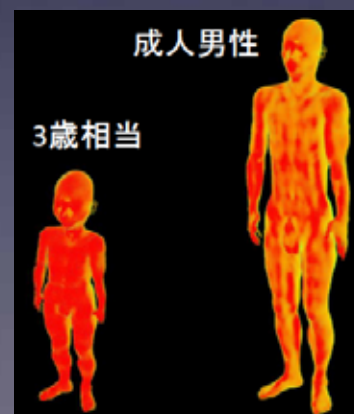
齋藤敦子



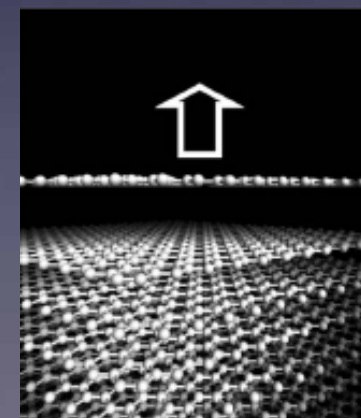
- 高分子流体解析 (理学研究科村島先生)
- 三次元可視化コンテンツ作成支援(大阪大学茂田先生)



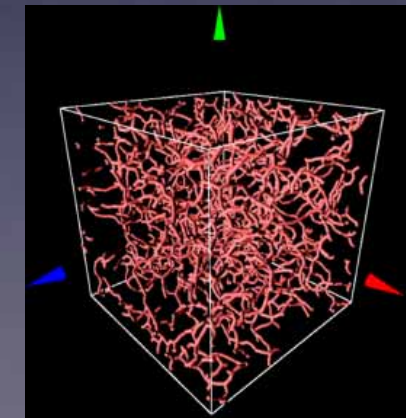
□ ナノ粒子生成過程解析



□ 熱中症シミュレータ



□ 電子系・原子核系シミュレーション



□ 高分子流体解析

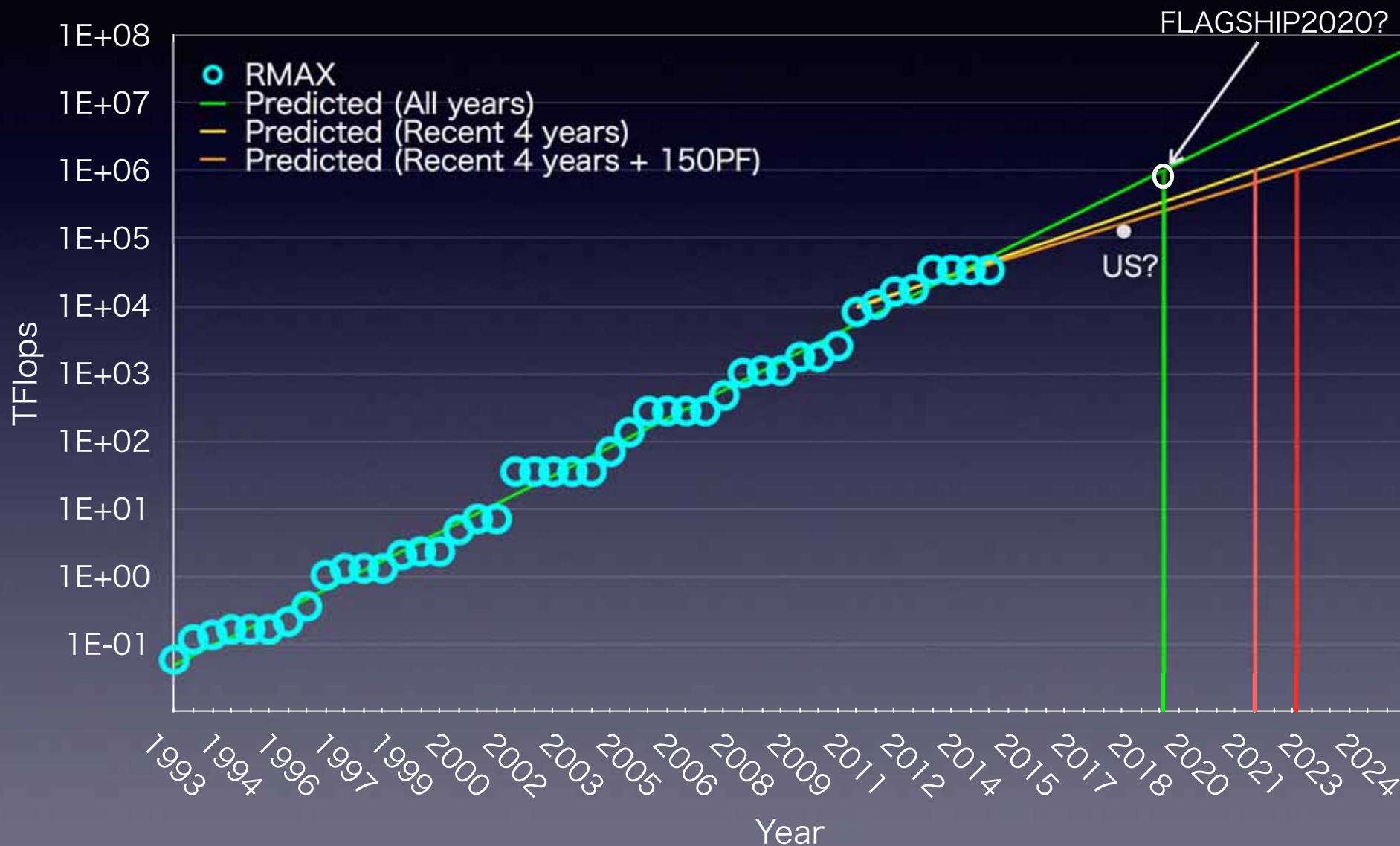


東北大学

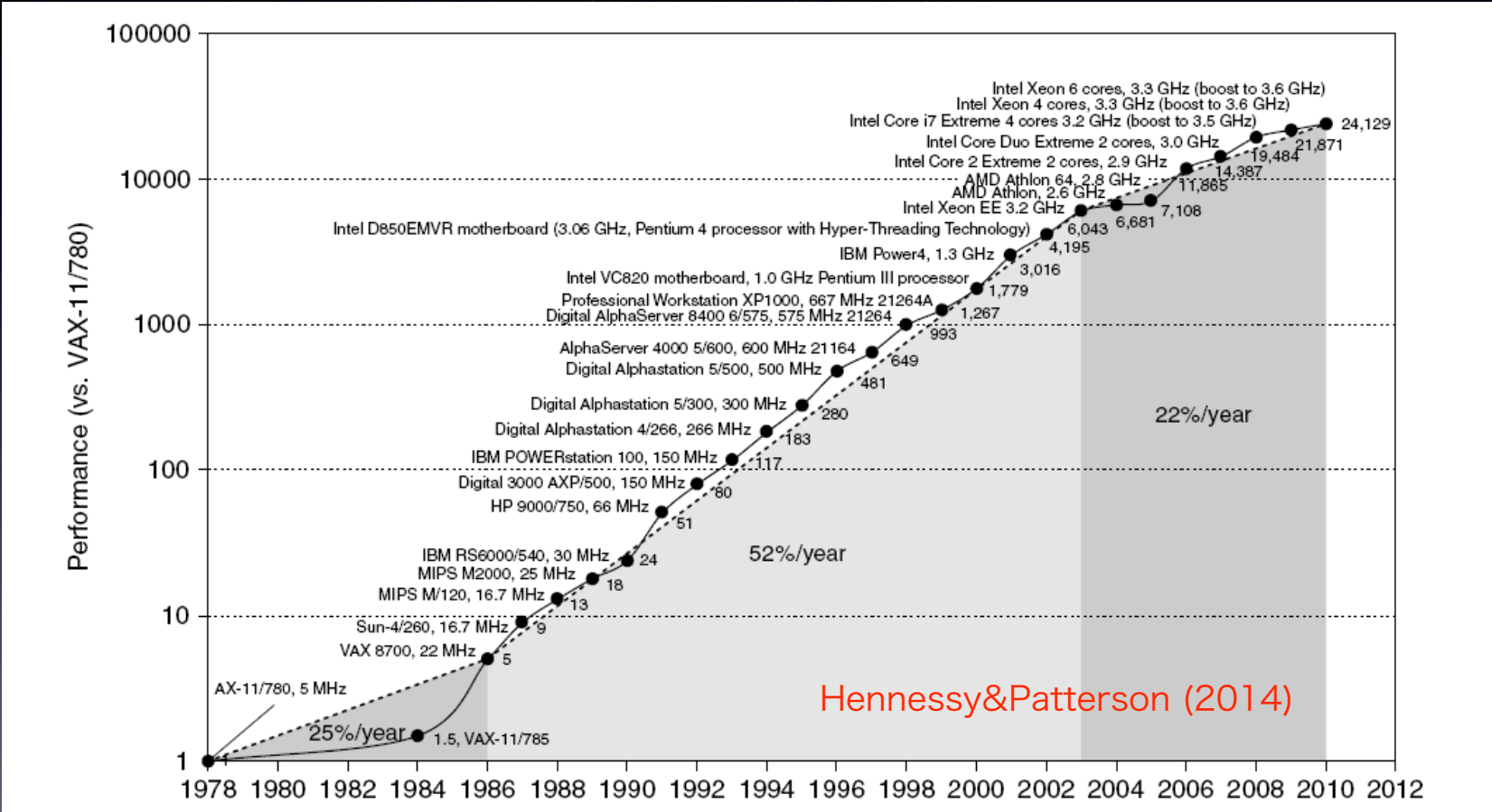


普通の人々のためのスーパーコンピュータセンターを目指して

エクサフロップを追い求めて・・・



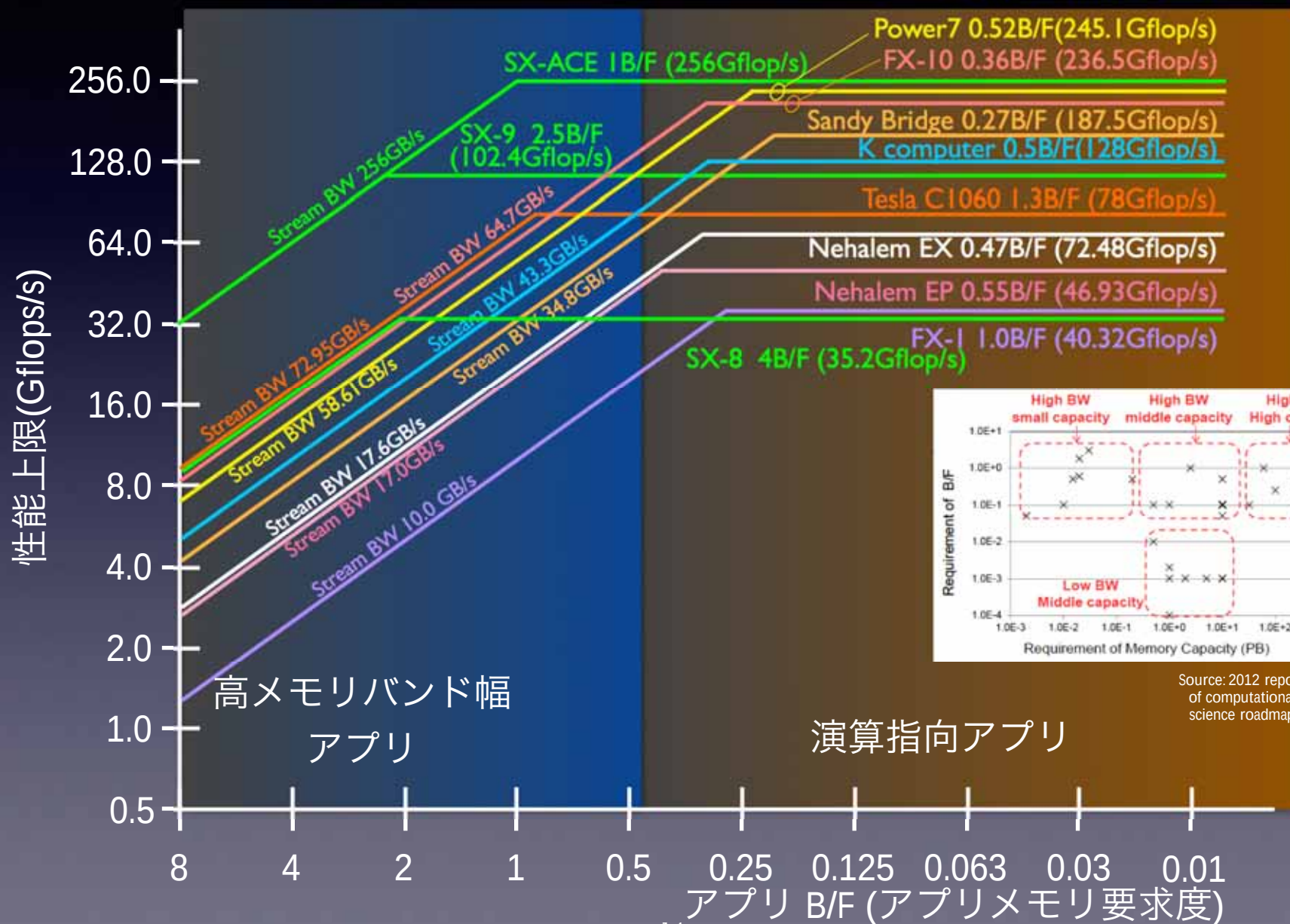
Flagshipにまつわる最悪のシナリオ！？



量によるbrute force approachから,質を重視したsmart approachへ

スパコンに関する誤信と落とし穴：その1

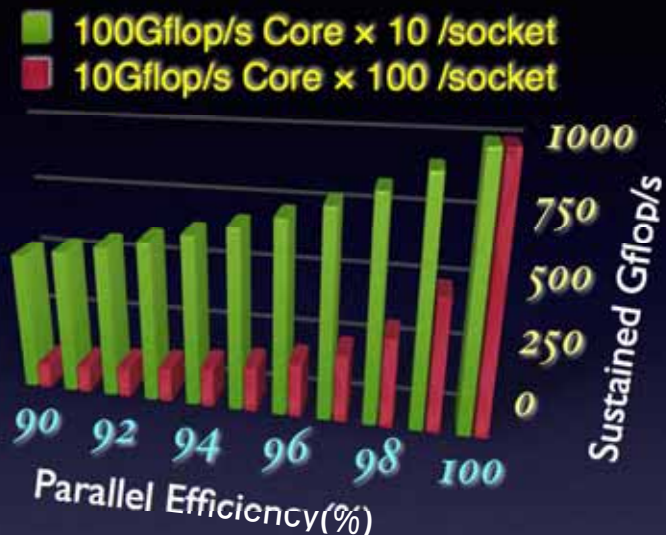
単にピーク性能が高ければ良いってモンじゃない



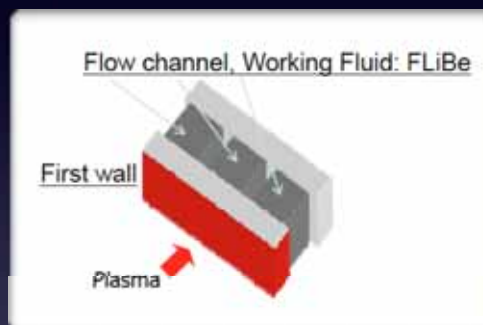
Source: 2012 report
of computational
science roadmap

スパコンに関する過信と落とし穴：その2

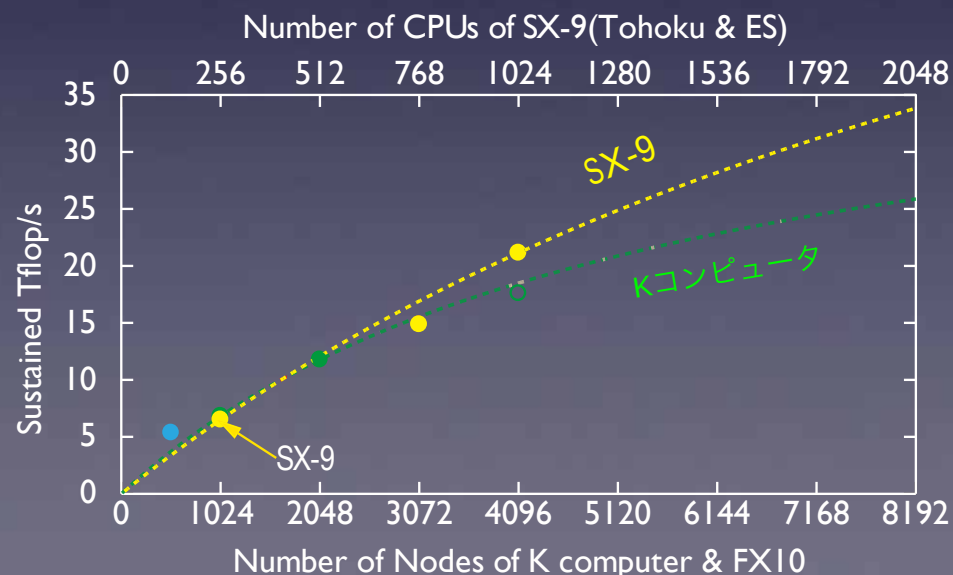
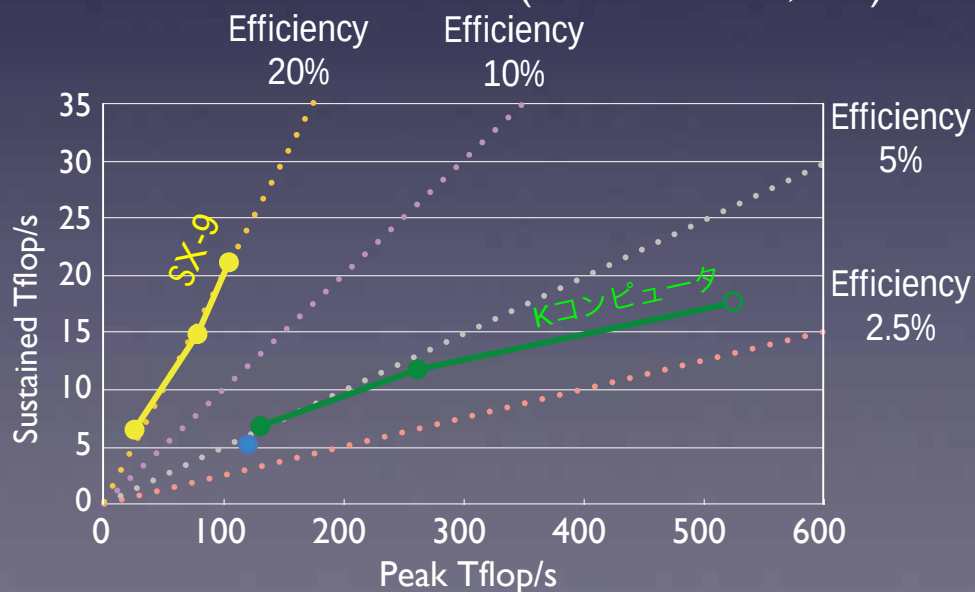
単に数を増やせば良いってもんじゃない



超大規模数値計算に基づく核融合炉先進ブランケットデザイン条件における高精度 MHD 熱伝達データベースの構築 (山梨大学 山本教授とのJHPCN共同研究)



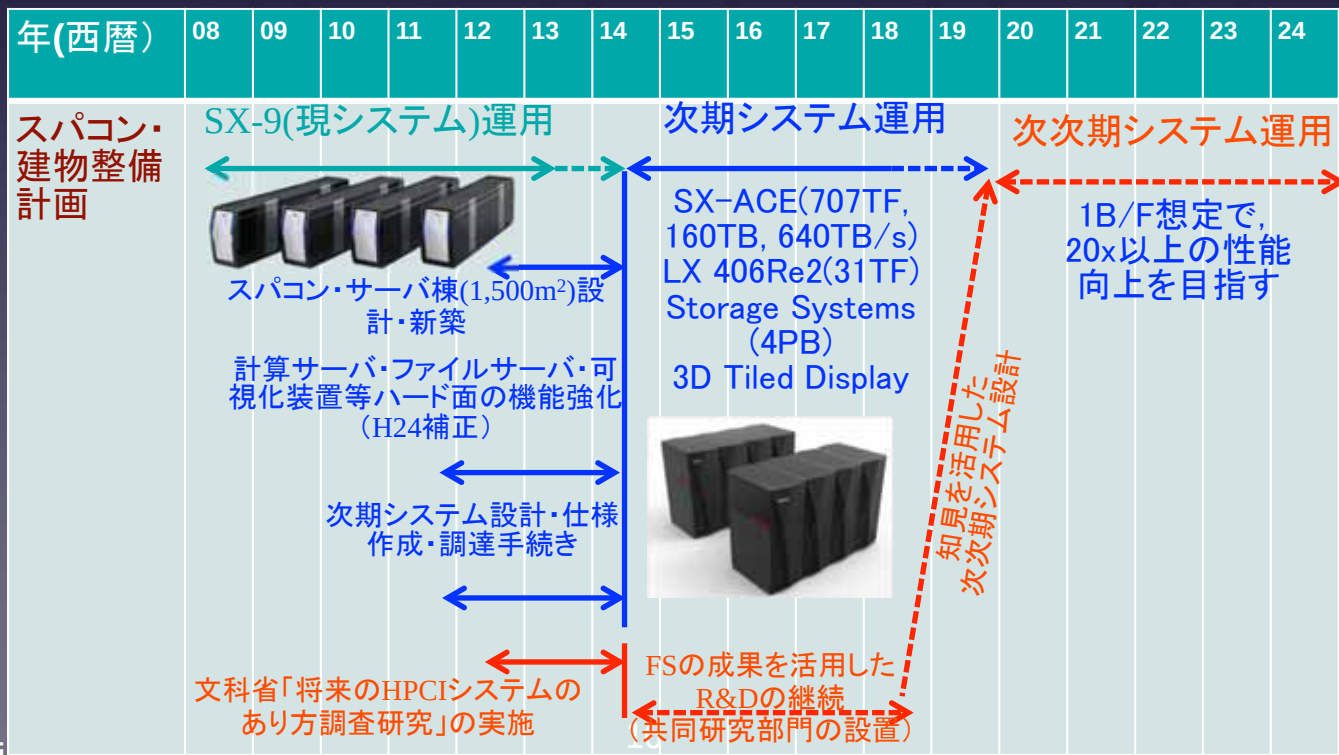
● SX-9 (Tohoku & ES; 1D) ● K computer (1D), ○ (2D) ● FX10 (1D)



東北大学サイバーサイエンスセンターの 今後のスーパーコンピュータ整備・運用・研究開発計画 ～次の10年間の活動計画～

● 基本戦略: Supercomputers for the rest of usの研究開発・整備・運用を目指して...

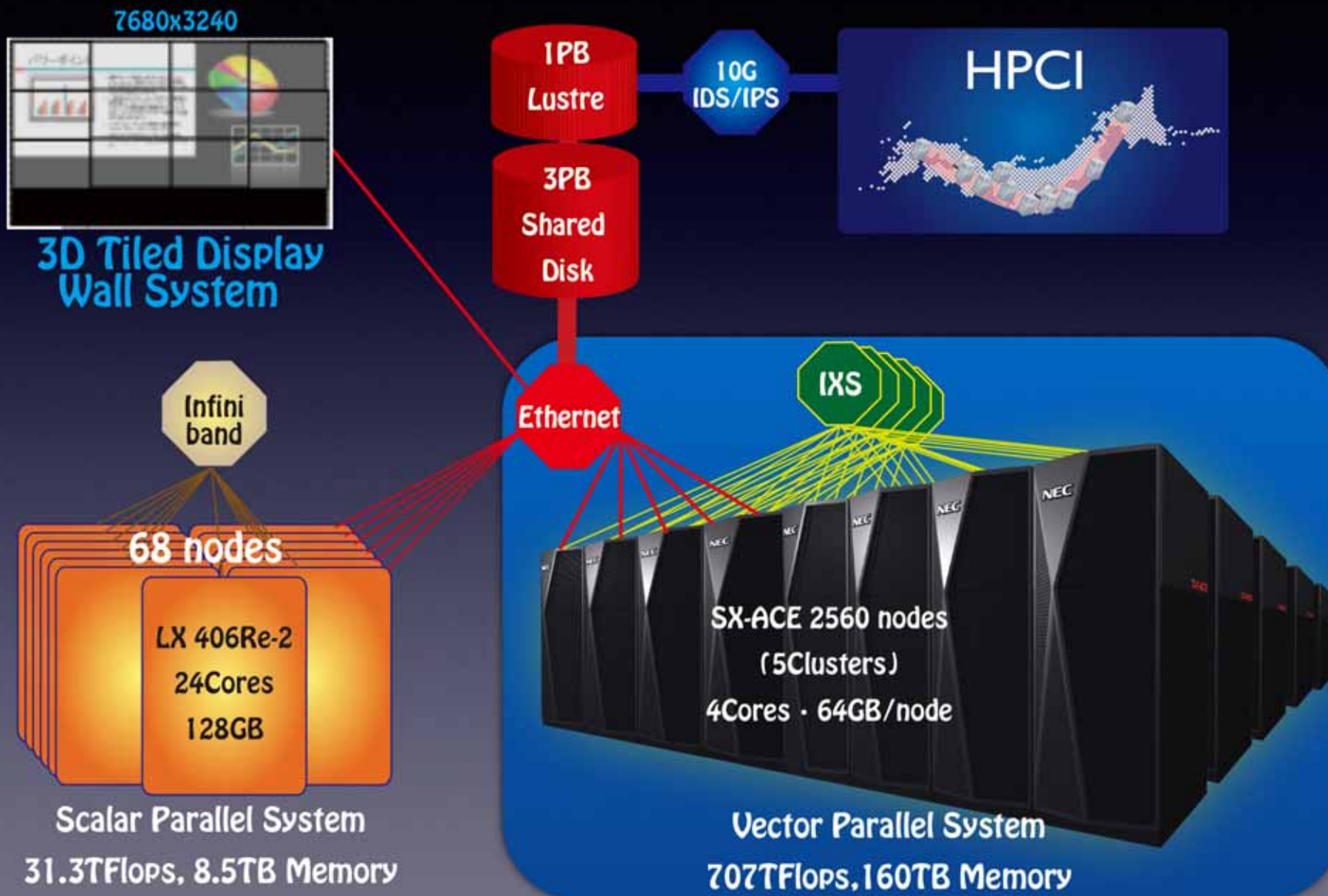
- ★ 第2階層に位置するNISとして、特にflop/s指向のフラグシップマシンを補完すべく、flop/sとバランスしたメモリ性能を有するマシンの研究・開発に組み、導入時にリーディングマシンの1/10程度のシステムでも高い生産性 (Short time to Solution)を提供できるシステムの整備・運用を目指す
 - ✓ システム運用と開発を両輪に、ユーザ支援・共同研究で得た知見を、次のシステムの設計・開発に生かす
- ★ 教員・技術系職員・ベンダー技術者が一体となった現ユーザ支援体制のさらなる強化
- ★ 学部・研究科（リーディング大学院等）と連携したHPC教育プログラムへの参画
- ★ 社会貢献として産業利用支援の強化、復興支援基盤としてスパコンの活用
- ★ 全国共同利用型の学内スパコン連携体制の構築、およびHPCI基盤を活用した学外スパコン機関との連携のさらなる推進



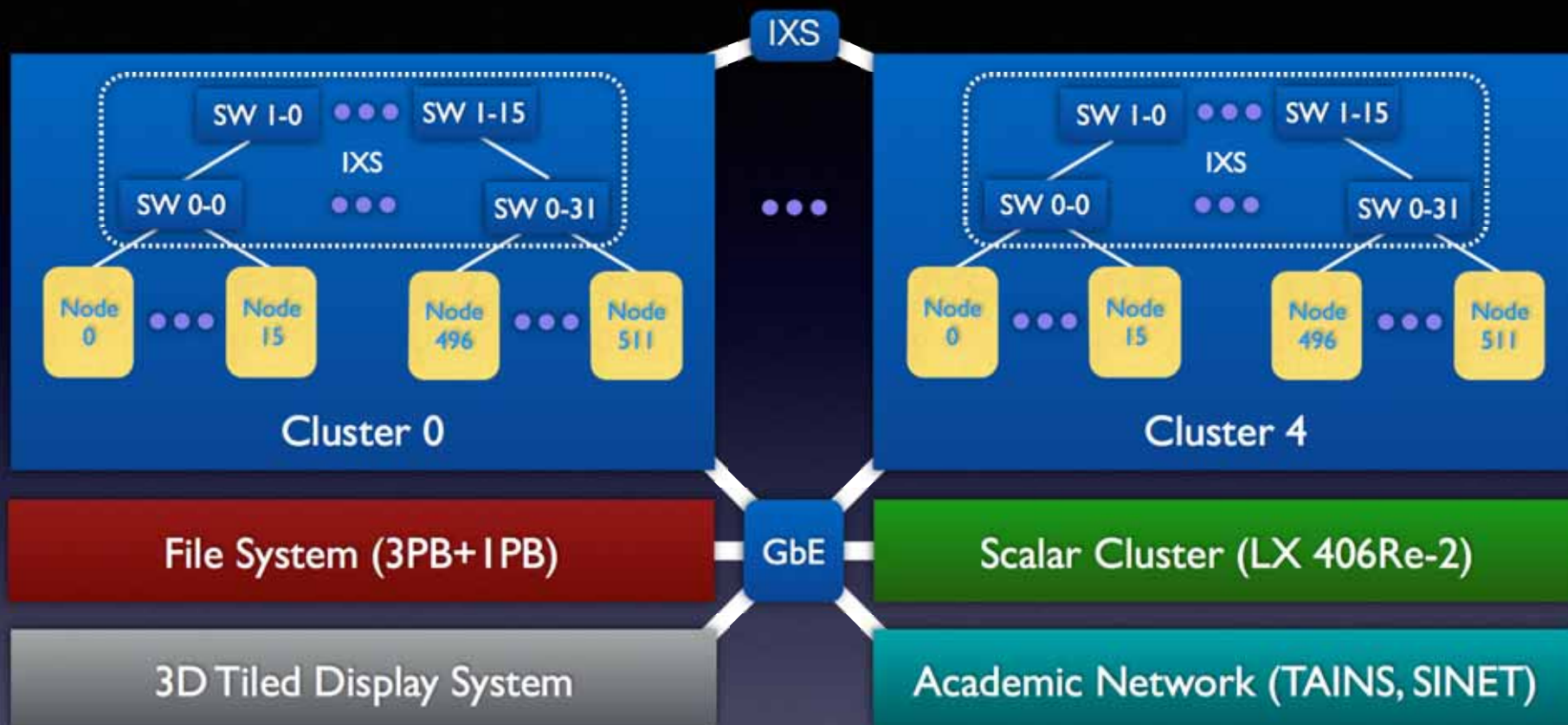
新スパコン棟の建設と新スパコンの導入



平成26年度システム強化計画(H26.10.1~)



東北大学新スーパーコンピュータシステムの構成



	Core	CPU(Socket)	Node	Cluster	Total System
Size	1	4 Cores	1CPU	512 Nodes	5 Clusters
Performance (VPU+SPU)	69GFlop/s (68GF+1GF)	276GFlop/s (272GF+4GF)		141Tflop/s (139TF+ 2TF)	707Tflop/s (697TF+10TF)
Mem. BW	256GB/s			131TB/s	655TB/s
Memory Cap.	64GB			32TB	160TB
IXS Node BW	-		4GB/s x2		-

東北大学新スーパーコンピュータシステムの特徴

大幅な性能強化を省スペース・省電力で実現				
		SX-9 (2008)	SX-ACE (2014)	性能向上比
	コア数	1	4	4x
CPU性能	理論最大演算性能	118.4Gflop/s	276Gflop/s	2.3x
	メモリバンド幅	256GB/sec	256GB/sec	1
	ADB容量	256KB	4MB	16x
システム性能・規模・消費電力	理論最大演算性能	34.1Tfop/s	706.6Tflop/s	20.7x
	総メモリバンド幅	73.7TB/s	655TB/s	8.9x
	総メモリ容量	18TB	160TB	8.9x
	最大消費電力	590kVA	1,080kVA	1.8x
	計算機室床面積	293平米	430平米	1.5x

量より質へのこだわり				
		SX-ACE(2014)	京(2011)	性能比
CPU (ノード) 性能	クロック周波数	1GHz	2GHz	0.5x
	コア性能	64Gflop/s	16Gflop/s	4x
	コア数/CPU	4	8	0.5x
	積和演算性能	256Gflop/s	128Gflop/s	2x
	メモリバンド幅	256GB/s	64GB/s	4x
	メモリ容量	64GB	16GB	4x

計算科学の多様な分野において萌芽研究から実用研究に対し
生産性の高いシミュレーション実行環境の構築をめざし、全国の利用者に提供

SX-ACEプロセッサの特徴

4つの高性能ベクトルコア&オンチップメモリ機構の強化

272Gflop/s of VPU + 4Gflop/s of SPU per socket

4-core configuration, 68Gflop/s + 1Gflop/s per core **SX-ACE Processor Architecture**

256 GB/s memory bandwidth

1B/F in 4-core Multiply-Add operations
 ~ 4B/F in 1-core Multiply-Add operations

1MB private ADB per core (4MB per socket)

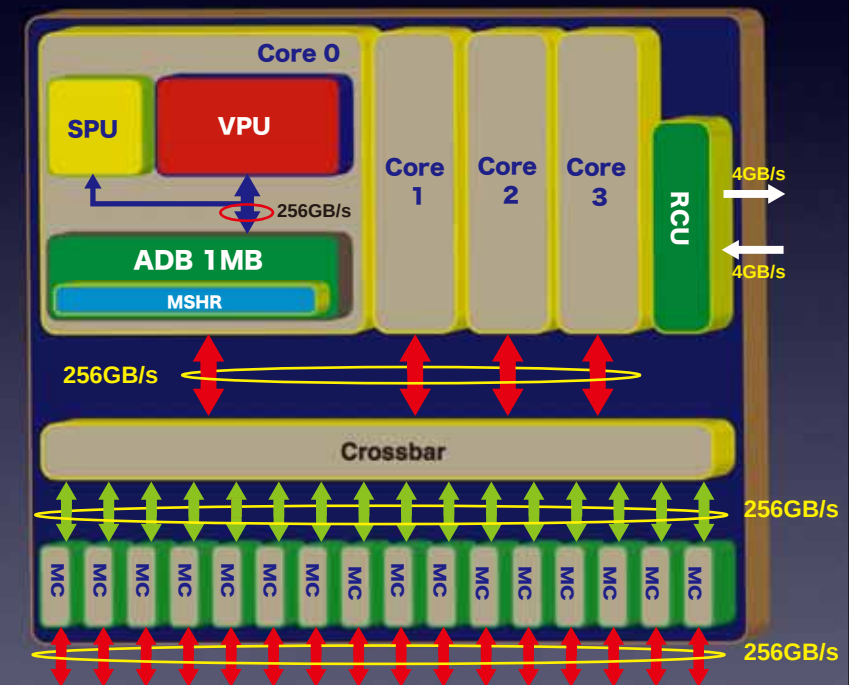
- 4x compared with SX-9
- 4-way set-associative
- MSHR with 512 entries (address+data)
- 256GB/s to/from Vec. Reg.
 - 4B/F for Multiply-Add operations

短ベクトル処理および間接メモリ参照処理の強化

Out of Order execution for vector load/store operations

Advanced data forwarding in vector pipes chaining

Shorter memory latency than SX-9



Source: NEC

センター利用者の実コードを用いたSX-ACEの評価

★ 評価のポイント

● ソケット性能の向上&マルチベクトルコア化の効果は？

- Is there any negative effect of the reduction in B/F from 2.5(SX-9) to 1.0(SX-ACE) on the sustained performance?

● ADB強化の効果は？

- Exploitation of locality in the inner-most loop
- Enlarged Capacity (4x compared with SX-9)

● 間接参照, 短ベクトル処理に対する効果は？

- OoO Vector Load/Store Operations
- Shorter Memory Latency (1/2 of SX-9)
- Direct Data Forwarding in Vector Pipe Chaining

評価対象システム一覧

System	Perf./ System (Tflop/s)	No. of nodes	No. of Sockets /Node	Perf./ Socket (Gflop/s)	No. of Cores	Perf. / core (Gflop/s)	Mem. BW GB/sec	On-chip mem	NW BW (GB/sec)	Sys. B/F
SX-ACE	131.1	512	1	256	4	64	256	1MB ADB /core	2 x 4 IXS	1.0
SX-9	26.2	16	16	102.4	1	102.4	256	256KB ADB/core	2 x 128 IXS	2.5
ES2	104.2	128	8	102.4	1	102.4	256	256KB ADB/ core	2 x 64IXS	2.5
LX 406	29.4	64	2	230.4	12	19.2	59.7	256KB L2/core 30MB Shared L3	5 IB	0.26
FX10	22.7	96	1	236.5	16	14.78	85	12MB shared L2	5 - 50 Tofu NW	0.36
K	524.2	4096	1	128	8	16	64	6MB Shared L2	5 - 50 Tofu NW	0.5
SR16K M1	62.7	64	4	245.1	8	30.6	128	256KB L2/core 32MB shared L3	2 x 24 - 96 custom NW	0.52

Remarks: Listed performances are obtained based on total Multiply-Add performances of individual systems

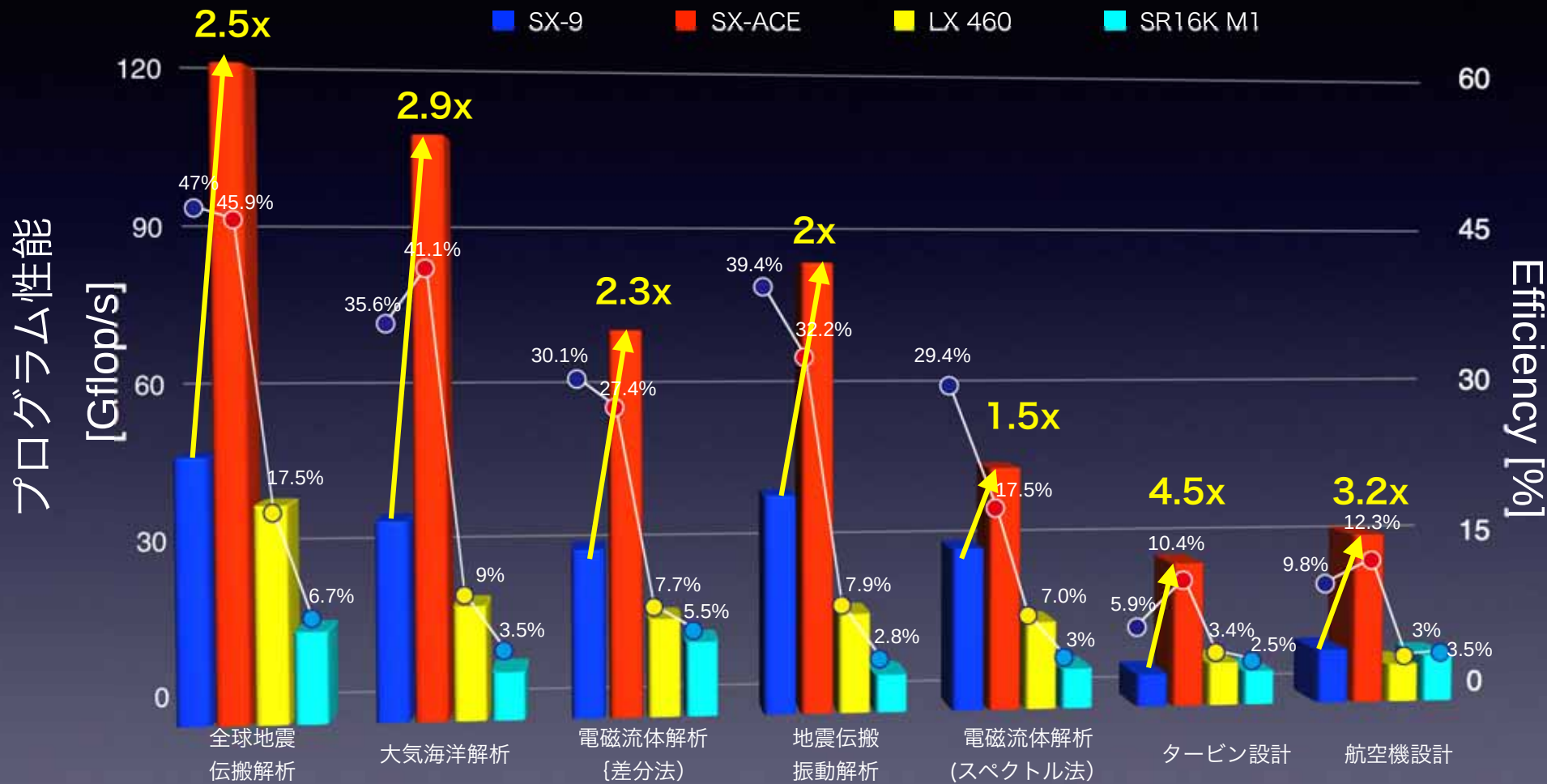
評価に用いたアプリケーション一覧

Applications	Fields	Methods	Mem Access Characteristics	Mesh Size	Code B/F	Actual B/F on ACE
QSFDM GLOBE	Seismology	Spherical 2.5D FDM	Stencil with sequential memory accesses	4.3×10^7 grids	2.16	0.78
Barotropic ocean	OGCM (Ocean General Circulation Model)	Shallow water model	Stencil with sequential memory accesses	4322×216	1.97	1.11
MHD (FDM)	MHD	Finite Difference Method	Stencil with sequential memory accesses	$200 \times 1920 \times 32$	3.04	1.41
Seism 3D	Seismology	Finite Difference Model	Stencil with sequential memory accesses	$1024 \times 512 \times 512$ † $4096 \times 2048 \times 2048$ ‡	2.15	1.68
MHD (Spectral)	MHD	Pseudo spectral Method	Stride memory access	$900 \times 768 \times 96$ † $3600 \times 3072 \times 2048$ ‡	2.21	2.18
TURBINE	CFD	DNS	Indirect memory access with short vectors	$91 \times 91 \times 91 \times 13$	1.78	5.47
BCM	CFD	Navier Stokes Equation	Stencil and Indirect memory access	$(128 \times 128 \times 128 \text{ cells}) \times 64 \text{ Cubes}$	7.01	5.86

† for single-node evaluation

‡ for multi-node evaluation

SX-ACEの性能評価結果:単一CPU性能



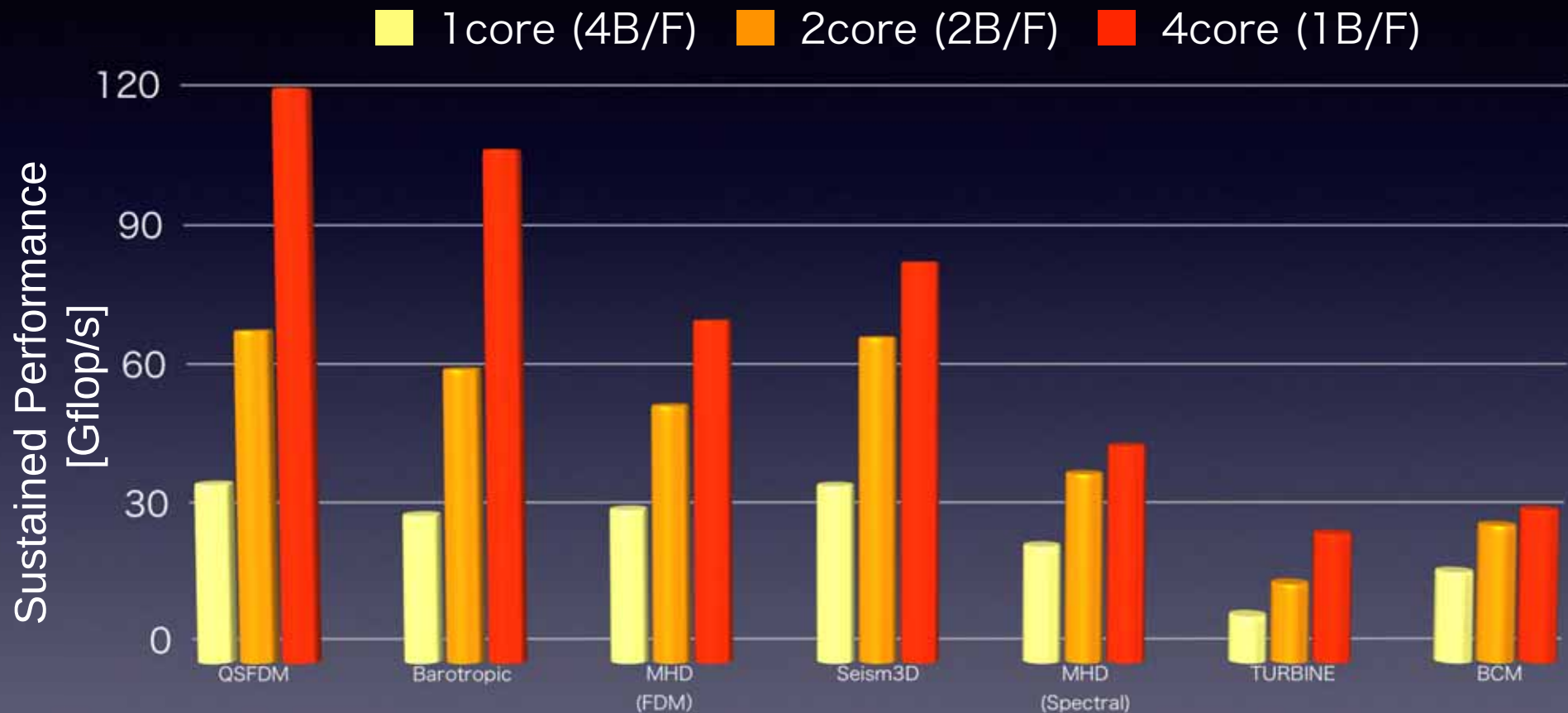
	QSFD	Barotropic	MHD (FDM)	Seism3D	MHD (Spectral)	TURBINE	BCM
Code B/F	2.16	1.97	3.04	2.15	2.21	1.78	7.01
Actual B/F	0.78	1.11	1.41	1.68	2.18	5.47	5.86

Stream Memory Bandwidth

- STREAM (TRIAD)

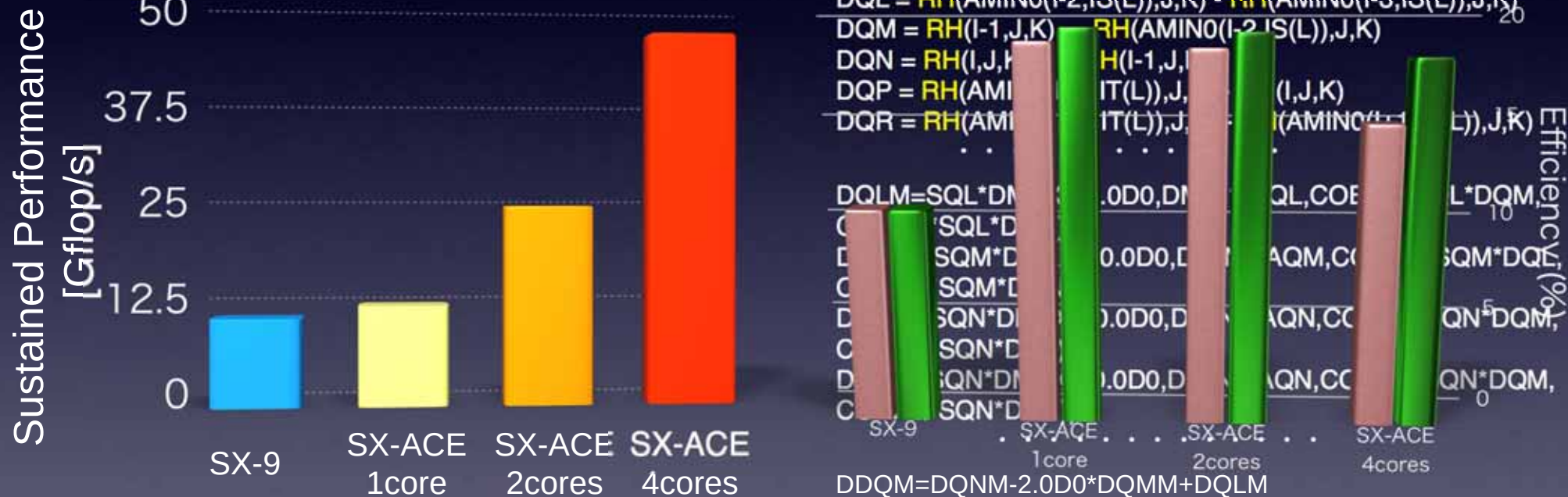


Sustained Performance of SX-ACE Processor



	QSFDM	Barotropic	MHD (FDM)	Seism3D	MHD (Spectral)	TURBINE	BCM
Code B/F	2.16	1.97	3.04	2.15	2.21	1.78	7.01
Actual B/F	0.78	1.11	1.41	1.68	2.18	5.47	5.86

Performance of Indirect Memory Accesses in TURBINE



```
DO 200 K=1,KF;
DO 200 J=1,JF;
DO 200 L=Istart,lend;
DO 200 I=IS(L),IT(L), ADB OFF
```

ADB ON

```
DQL = RH(AMINO(I-2,IS(L)),J,K) - RH(AMINO(I-3,IS(L)),J,K)
```

```
DQM = RH(I-1,J,K) - RH(AMINO(I-2,IS(L)),J,K)
```

```
DQN = RH(I,J,K) - RH(I-1,J,K)
```

```
DQP = RH(AMINO(IT(L)),J,K) - RH(AMINO(I-1,IS(L)),J,K)
```

```
DQR = RH(AMINO(IT(L)),J,K) - RH(AMINO(I-1,IS(L)),J,K)
```

```
...
```

```
DQLM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

```
DDQM=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

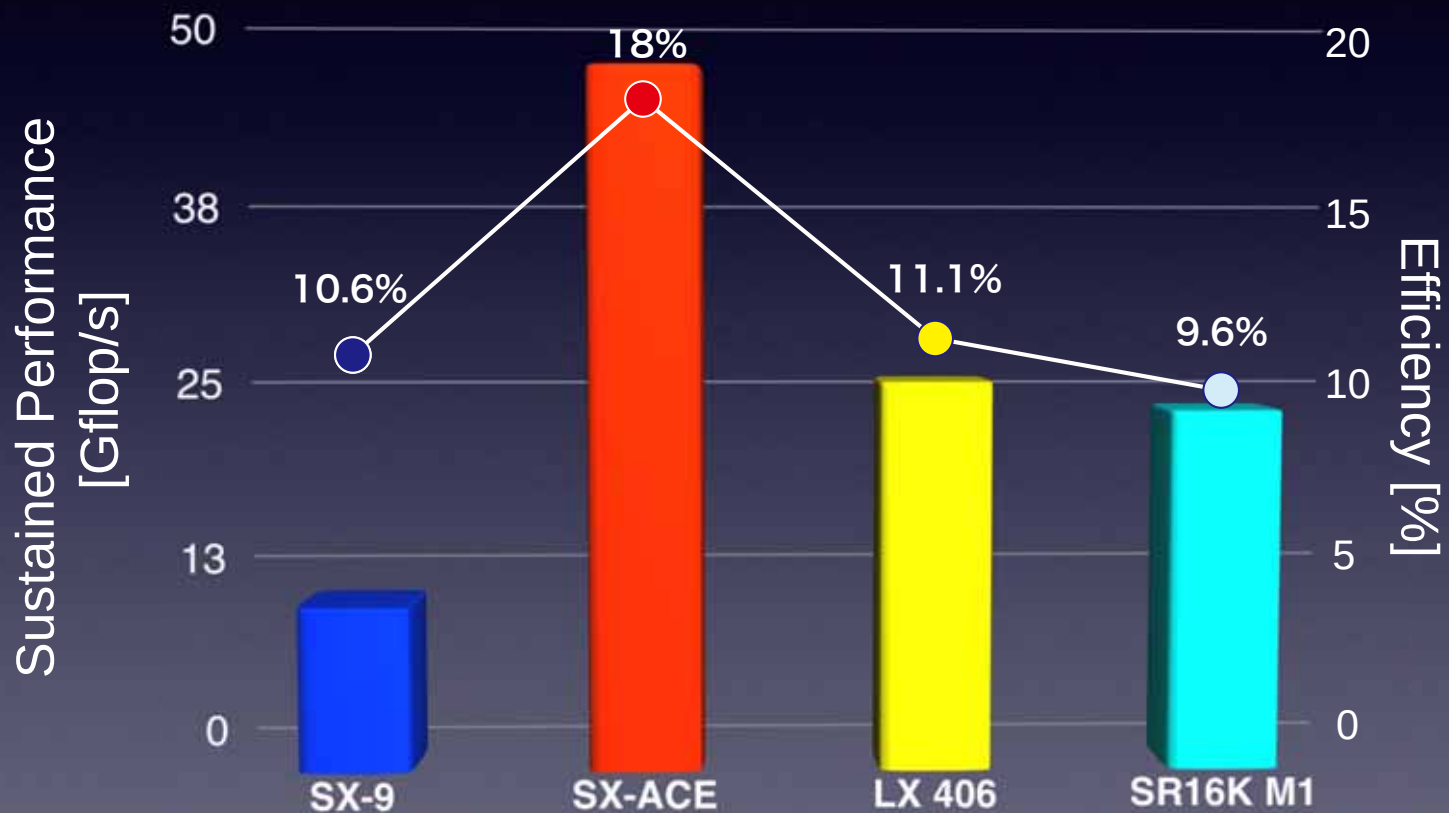
```
DDQN=SQL*DM0-2.0D0,DM0,SQL,COEFB*DPQM0+COEFC*DMQN0
```

200 CONTINUE

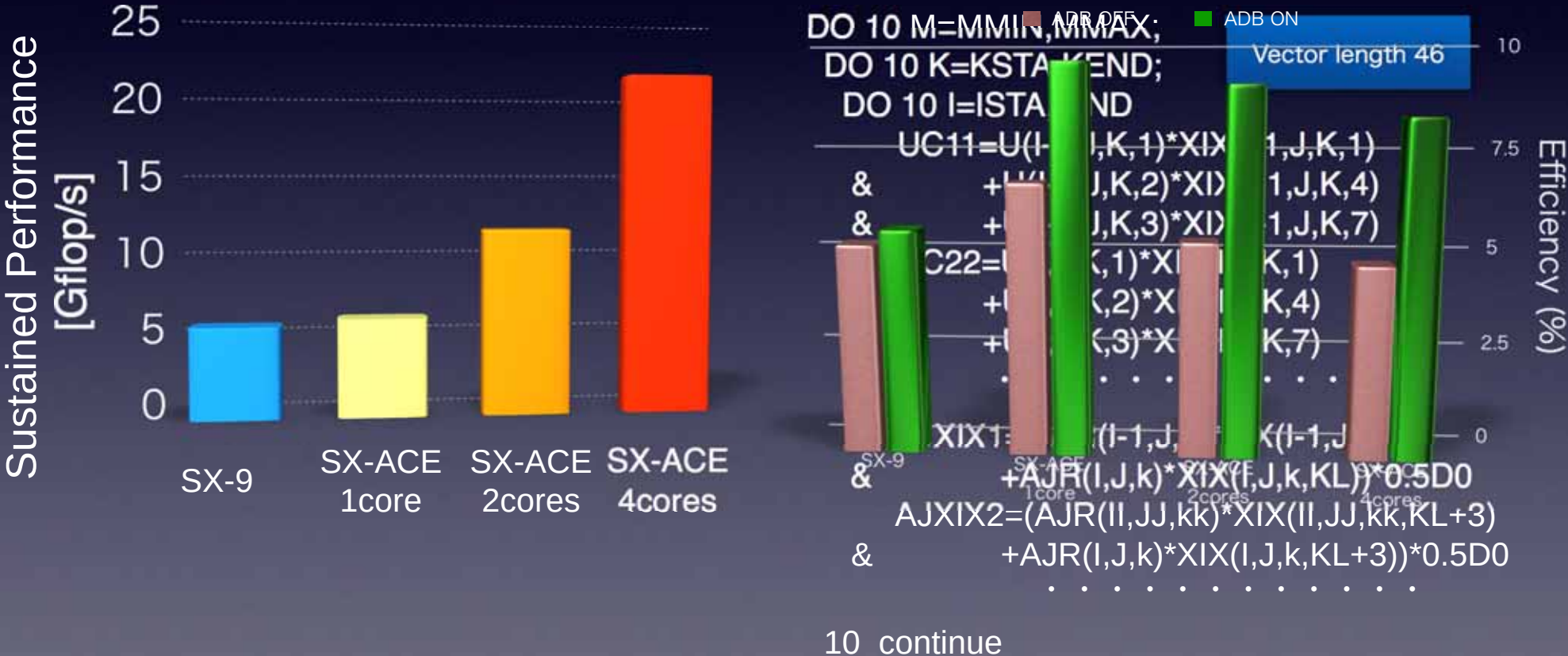
```
QLL(I,J,K,M)=DQL0+COEFB*DPQM0+COEFC*DMQN0
```

```
QRR(I,J,K,M)=DQR0-COEFB*DMQP0-COEFC*DPQN0
```

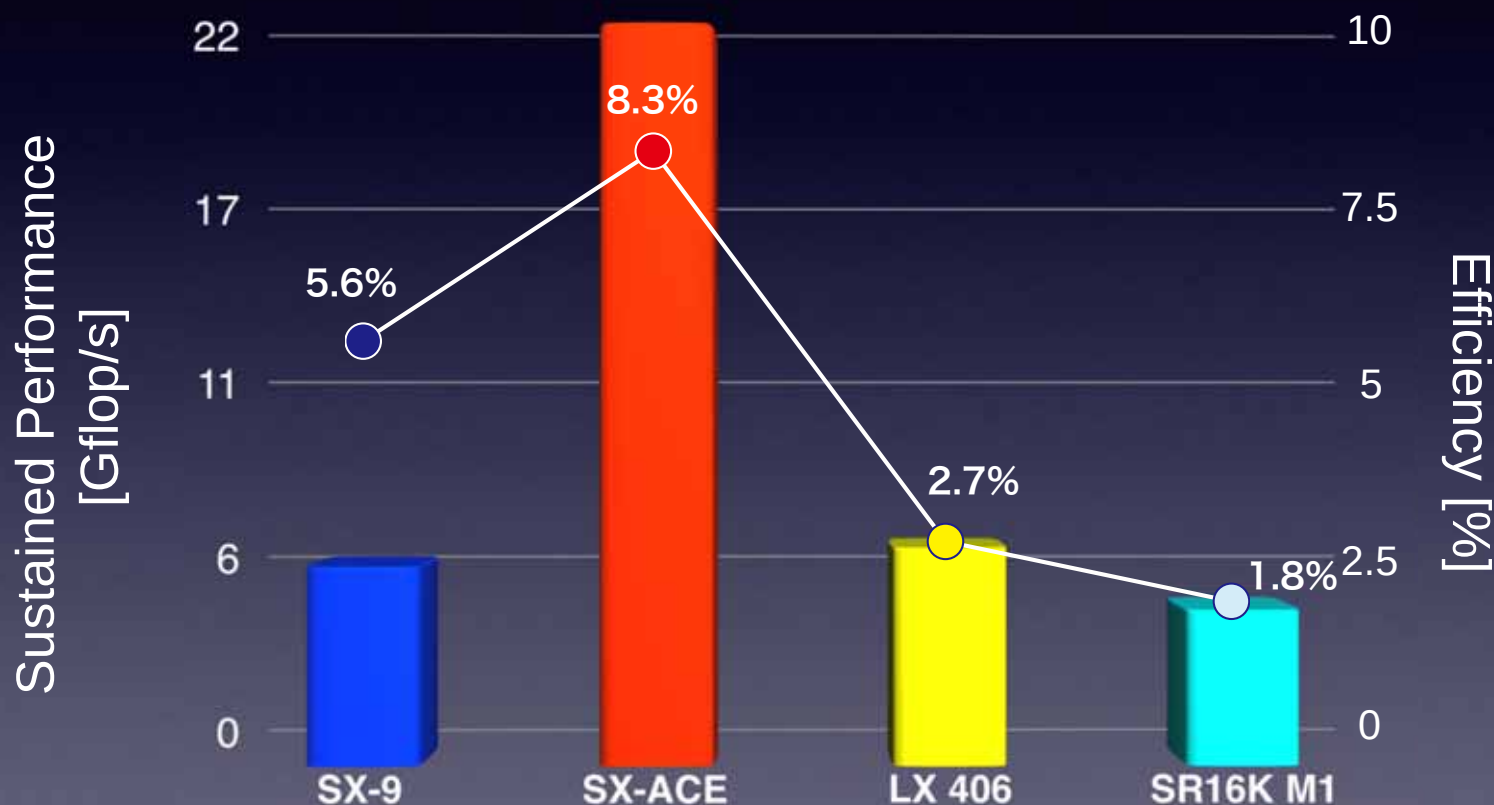
Performance of Indirect Memory Accesses in TURBINE on Modern HPC Processors



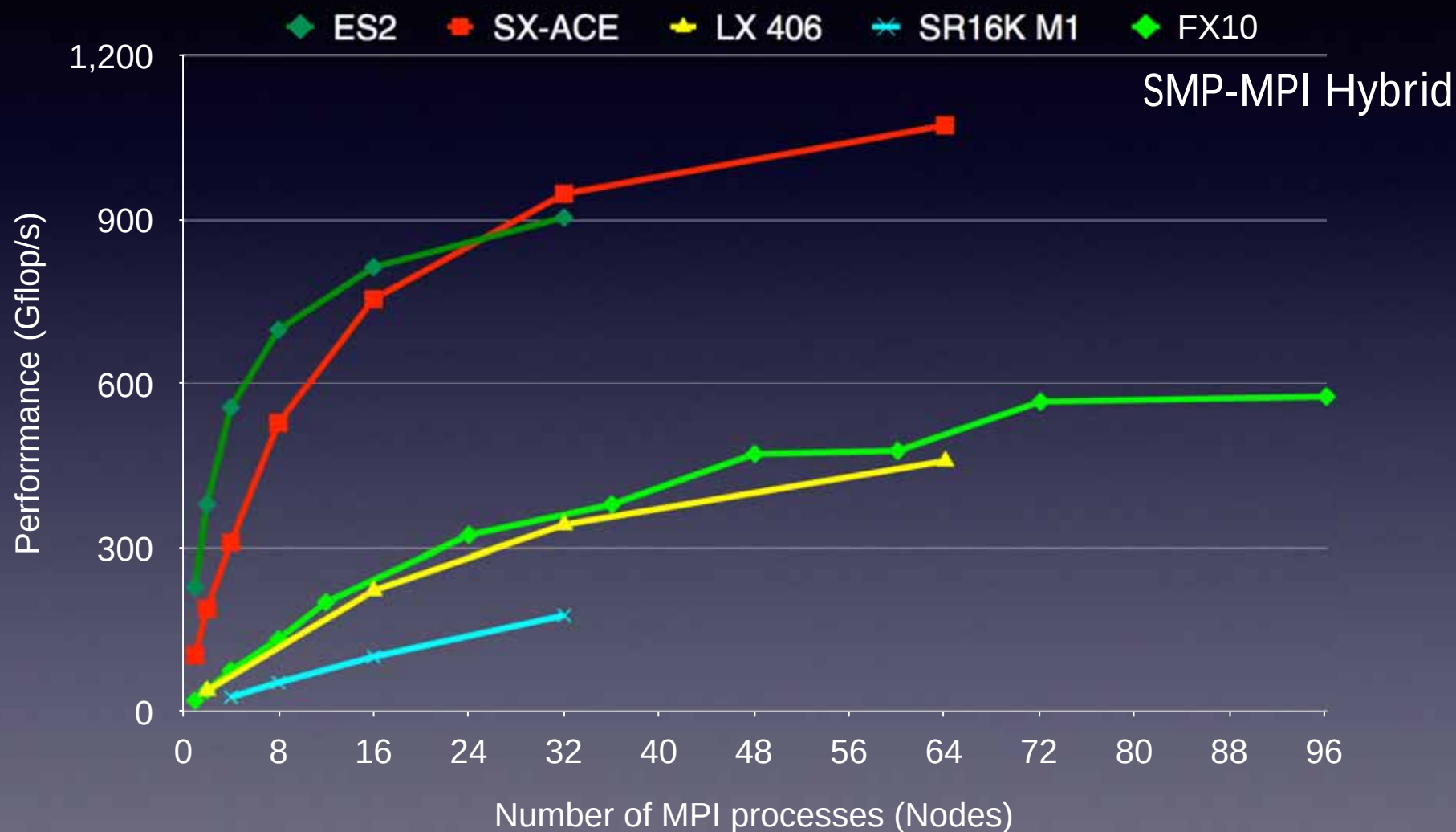
Performance of Short-Vector Processing in TURBINE



Performance of Short-Vector Processing in TURBINE on Modern HPC Processors



Sustained Performance of Barotropic Ocean on Multi-Node Systems



スパコン評価指標HPCGによる評価の取り組み

★ HPLの性能評価指標としての限界

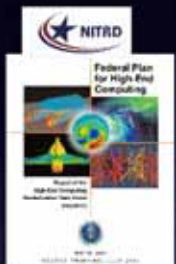
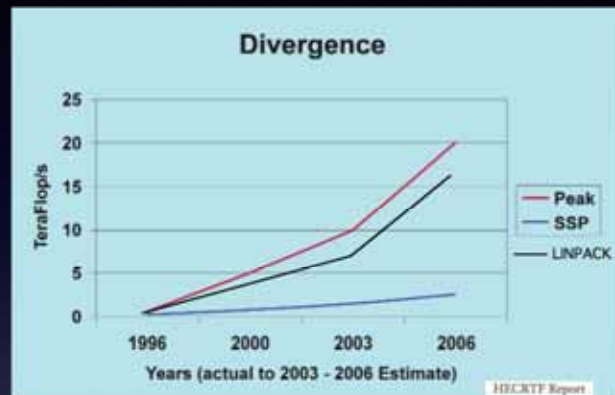
- 実アプリケーションとの性能乖離

＊HPCシステムの研究開発をゆがめる？

- 長時間実行が必要

＊耐久試験, 信頼性評価にはよいかも？

✓ 実アプリケーションの挙動に近い新しい評価指標が求められてきた



★ HPCG ベンチマークの誕生

- 実アプリケーションで頻繁に使用される疎行列計算と通信の性能を重視

＊疎行列ベクトル積, 内積演算, 総和通信, 隣接間通信等

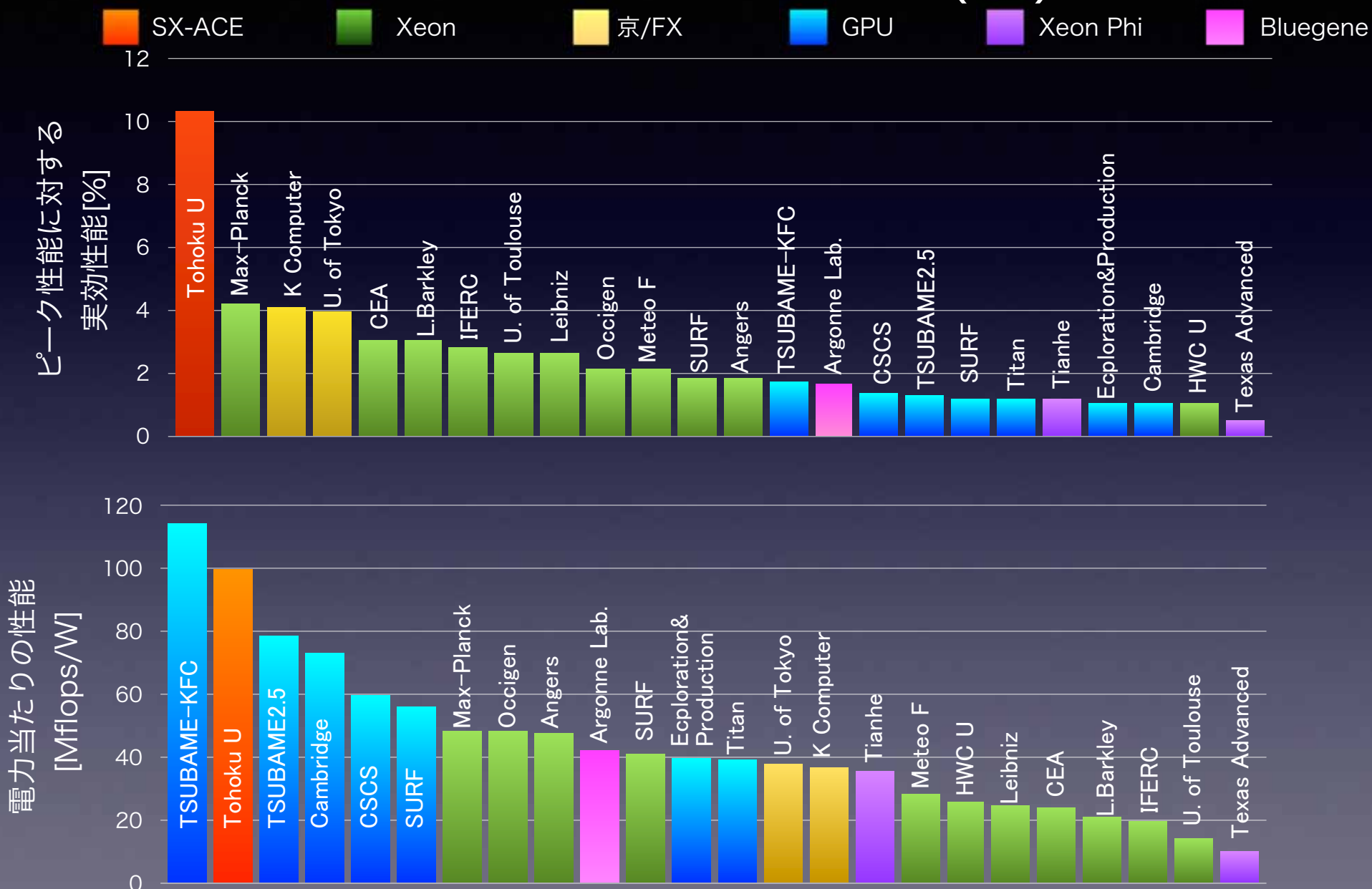
- アプリケーション最適化技術も評価の対象

スパコン評価指標HPCGでのランキング(2014.11版)

Rank	Site	Computer	cores	HPL (PFlops)	HPL Rank	HPCG (PFlops)	HPCG/HPL	HPCG/Peak
1	NSCC/Guangzhou	Tianhe-2 NUDT, Xeon 12C, 2.2 GHz + Xeon Phi 57C +Custom	3,120,000	33.9	1	0.632	1.5	1.1
2	RIKEN AICS	K Computer, SPARC64 VIIIfx 8C + Custom	705,024	10.5	4	0.461	4.4	4.1
3	DOE/OS Oak Ridge Nat Lab.	Titan, Cray XK7 AMD 16C + Kepler 14C + Custom	560,640	17.6	2	0.322	1.8	1.19
4	DOE/OS Argonne Lab.	Mira Bluegene/Q, Power BQC 16C 1.6GHz +Custom	786,432	8.59	5	0.167	1.9	1.7
5	Sewiss CSCS	Piz Daint Cray XC30, Xeon 8C + Kepler 14C +Custom	115,984	6.27	6	0.105	1.7	1.3
6	Leibniz Rechenzentrum	SuperMUC, Intel8C +IB	147,456	2.90	14	0.0833	2.9	2.61
7	DOE/OS L Barkley Nat Lab.	Edison XC30, Xeon 12C +Custom	133,824	1.65	24	0.0786	4.8	3.1
8	GSIC Center TiTech	Tsubame 2.5 Xeon 6C+Keplar 20x +IB	76,032	2.78	15	0.073	2.6	1.3
9	MaxPlanck	iDataPlex Xeon10C+IB	65,320	1.28	34	0.061	4.8	4.2
10	CEA/TGCC-GENCI	Curie fine nodes Bullx B510 Xeon 8C +IB	77,184	1.36	33	0.051	3.8	3.1
11	Exploration and Production Eni S.p.A	HPC2, Xeon 10C + Kepler 14C +IB	62,640	3	12	0.0489	1.6	1.22
12	Grand Equipment National de Calcul	Occigen Xeon 12C	50,554	2.07		0.0448	2.8	2.2
13	U of Tokyo	Oakleaf-FX SPARK64 16C +Tofu	76,800	1.043	36	0.0448	4.3	3.9
14	TACC	Stampede, Dell Xeon8C + Xeon Phi(61c) + IB	462,462	5.168	7	0.044	0.9	0.5
15	IFERC	Helios Xeon 8C+IB	70,560	1.24	30	0.0426	3.4	2.8
16	HWC U of Stuttgart	Hornet Cray Xeon	94,656	2.763	-	0.0391	1.4	1.0
17	Surf Sara	Cartesius Xeon	25,920	0.848	-	0.0195	2.3	1.8
18	Tohoku U Cyberscience Centr	SX-ACE, 4C + IXS	2,048	0.123	-	0.0134	10.8	10.3

スパコン評価指標HPCGによる

計算能力の質の評価(1/2)

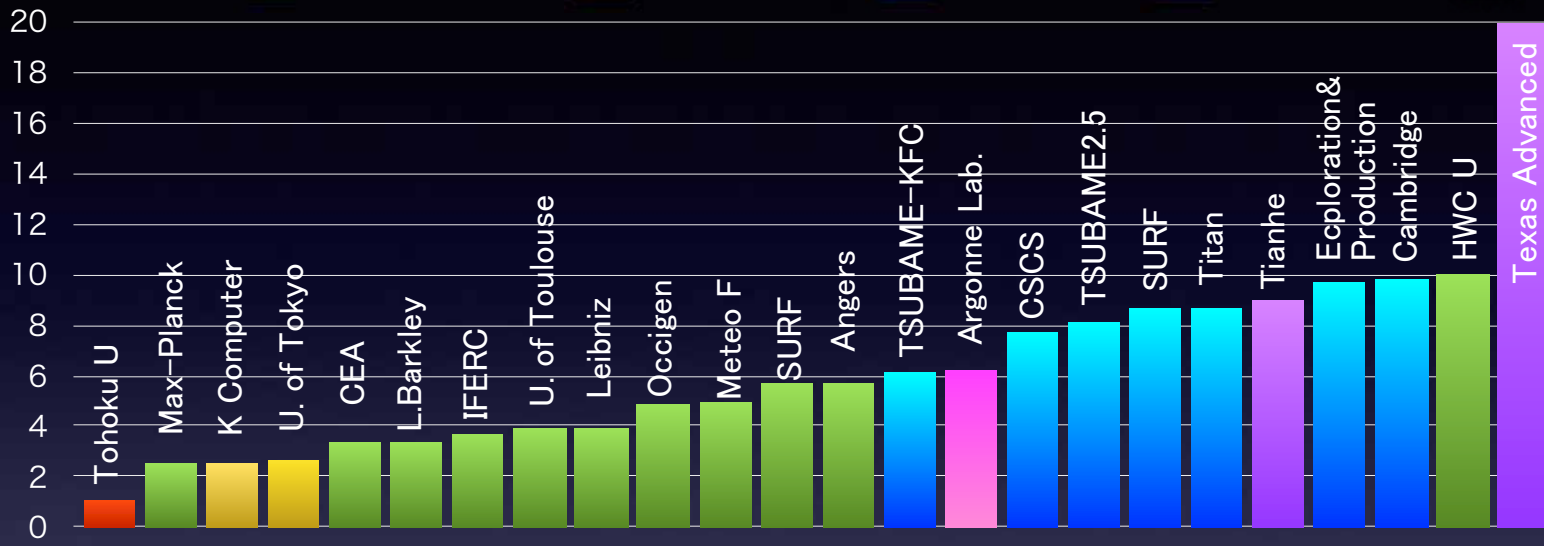


スパコン評価指標HPCG HPCGによる

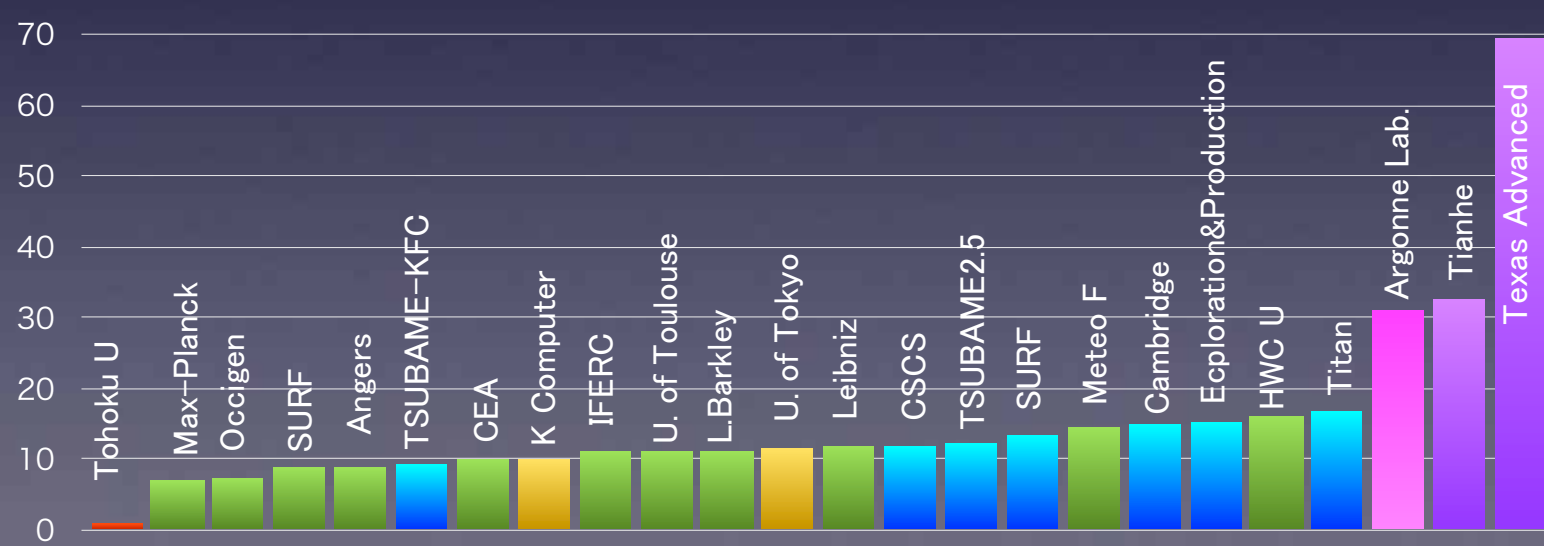
計算能力の質の評価(2/2)

■ SX-ACE
 ■ Xeon
 ■ 京・FX
 ■ GPU
 ■ Xeon Phi
 ■ Bluegene

同じHPCG性能を得るのに
 必要なピーク性能
 (SX-ACEで正規化)



同じHPCG性能を得るのに
 必要なコア数
 (SX-ACEで正規化)





東北大学



Cyberscience
Center

スパコンを活用した学際的R&Dの取り組み: リアルタイム津波浸水・被害予測・災害情報 配信による自治体の減災力強化の実証事業



IRIDeS
International Research Institute
for Disaster Science
災害科学国際研究所



KOKUSAI KOGYO CO., LTD.



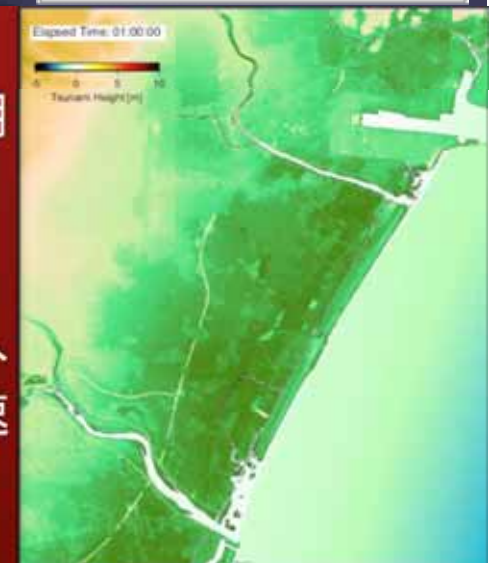
本事業の目的

- 東日本大震災における津波被害の教訓を踏まえ、世界最先端のシミュレーション・センシング・ICTを融合し、迅速かつきめ細やかな被害情報の推計・把握と配信によるG空間防災モデルを構築
- 沿岸部10mメッシュの分解能での「世界初」のリアルタイム津波浸水予測とG空間情報の活用による建物被害予測を、地震発生から20分以内に完了させ、実証自治体での災害対応の基盤情報に組み込み、準天頂衛星からのメッセージ送信や災害に強いワイヤレスネットワークを活用し、ユーザに対して確実に情報を配信

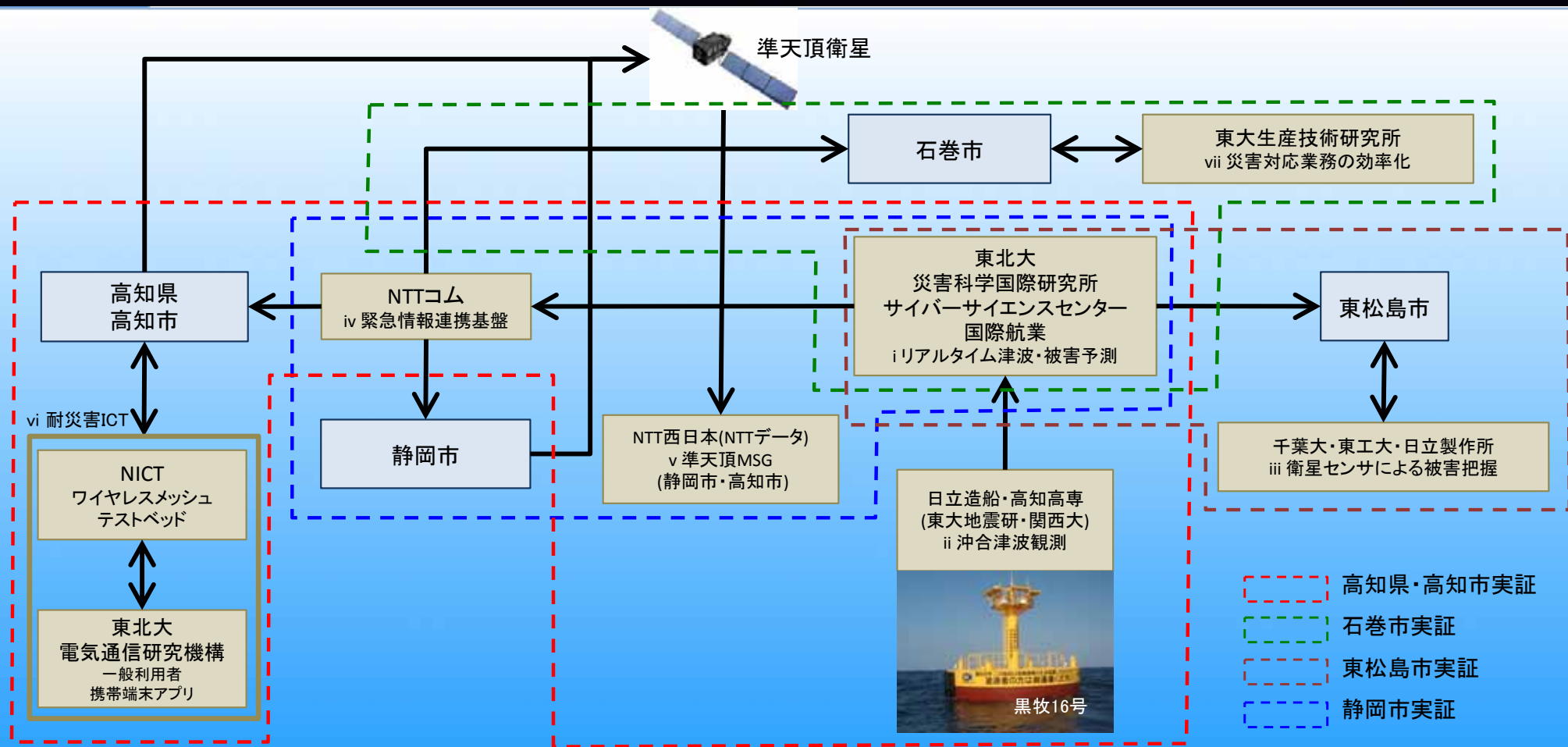


本事業の達成により

- ★ **平常時**には自治体の津波想定や組織内訓練システムとして運用し、様々な津波被害発生シナリオ構築、より現実的な避難訓練、人員・資材の効果的な配分の検討に活用
- ★ **災害発生時**には、リアルタイム情報提供に加え、社会情報（人口、建物、交通）を活用してリアルタイムで被害を推計し、迅速な初動対応・意志決定上の基本情報を提供



リアルタイム津波浸水・被害予測・災害情報配信 による自治体の減災力強化の実証事業体制図



事業代表者： 東北大学災害科学国際研究所 越村俊一教授

社会に役立つスーパーコンピュータ技術の創出： リアルタイム津波浸水・被害予測システムの研究開発

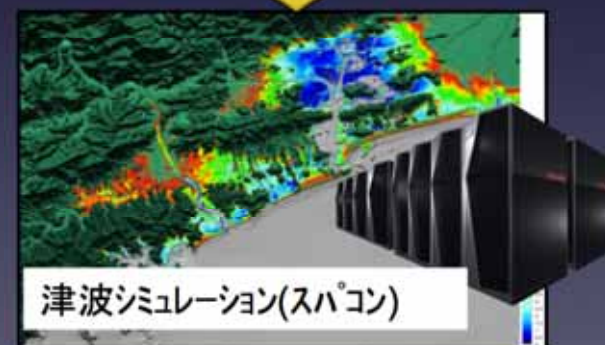
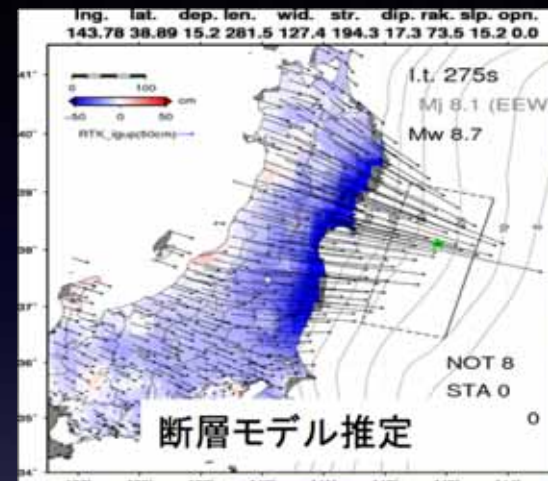
地震発生後20分以内に

- ★地震断層モデルに基づく津波発生予測 (10分)
- ★SX-ACEによる10mメッシュの津波浸水予測および被害推計 (10分)



運用システム
として世界初

- 10メートルメッシュの精度で、6時間分の浸水域・建物被害予測情報を自治体に提供
- 将来はスマホなどへlocation-awareな災害情報配信技術の確立を目指す



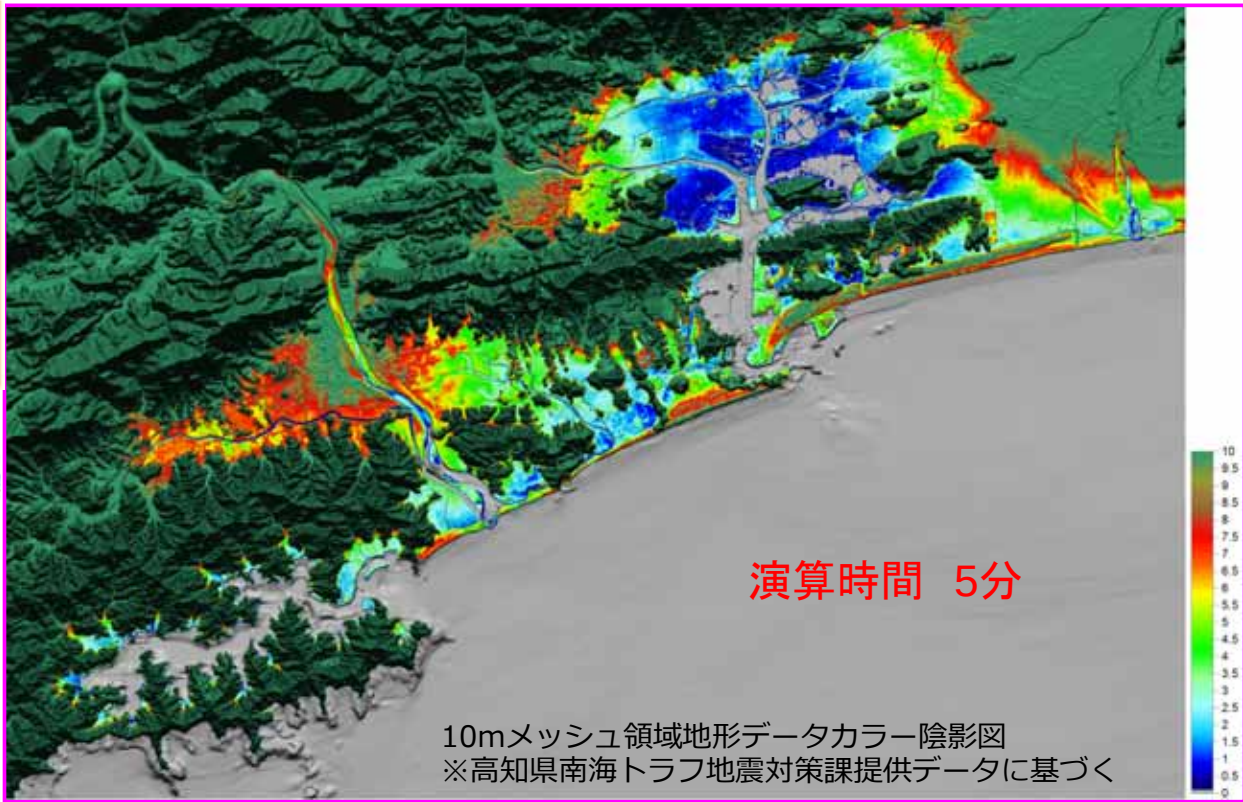
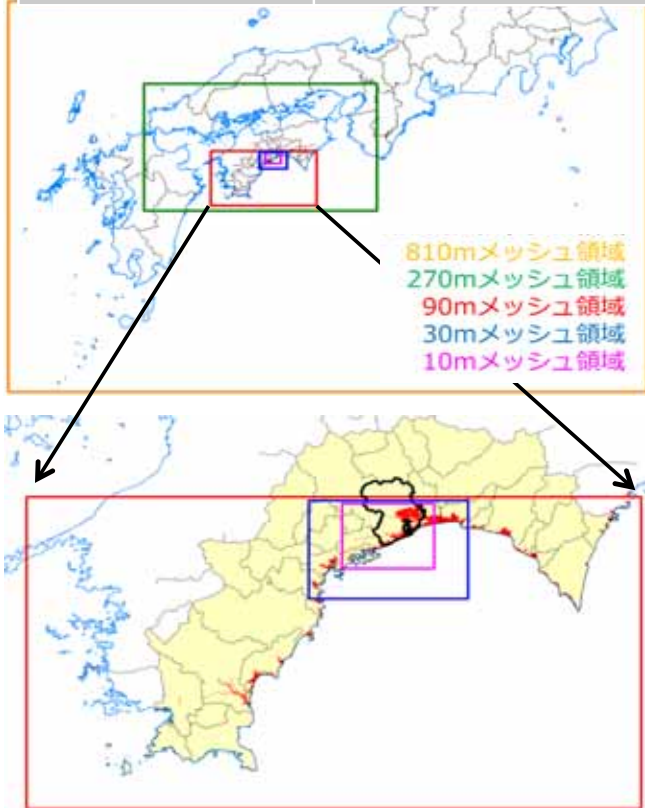
リアルタイムシミュレーション

Items	Condition
理論式	Non-linear Shallow Water Equations
計算スキーム	Staggered Leap-frog FDM
dx	Nested Grid (810m, 270m, 90m, 30m, 10m)
dt	0.25 sec. (Satisfying stability condition)
潮位	HWL

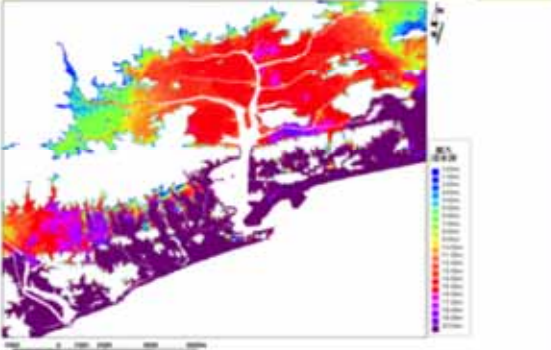
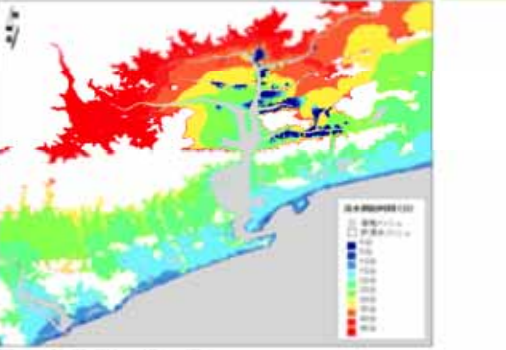
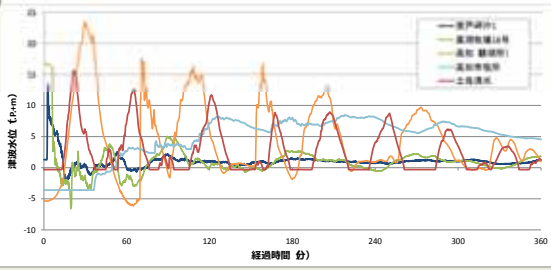
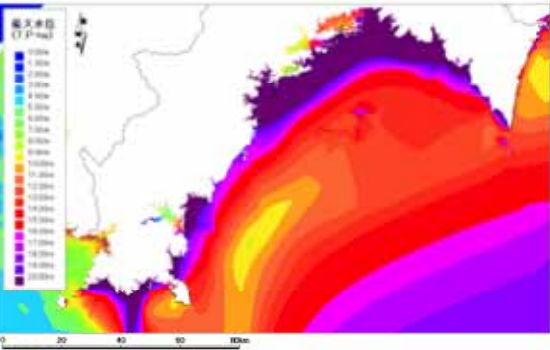
TUNAMI-Code

Tohoku University's Numerical
Analysis Model for Investigating
Tsunami

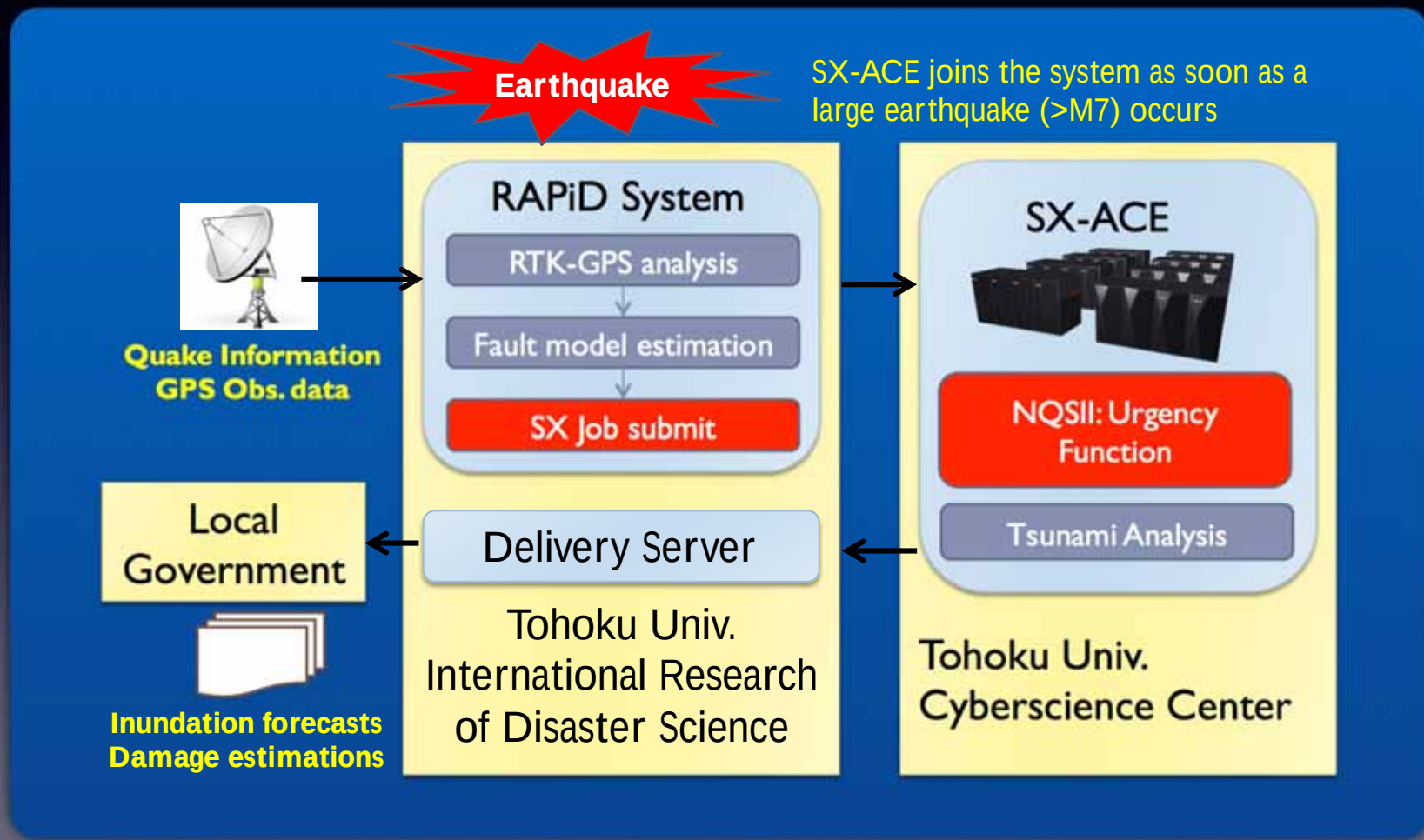
(高知市での事例)



リアルタイム津波・浸水被害予測システムにより得られる情報

優先順位	項目	イメージ	優先順位	項目	イメージ
1	到達時間		4	被害推定	
2	最大浸水深		5	浸水開始時間	
3	水位時系列		6	最大水位	

計測・シミュレーション・可視化・配信まで 運用システムへの実装



リアルタイム津波浸水シミュレーションの実現に向けて

リアルタイムに津波浸水被害予測を行うために
10分間のシミュレーションで**6時間分の津波浸水現象の再現**を目指す

例えば、静岡市の場合：

太平洋沖 1404Km×845km

741万格子 (810mから10m)

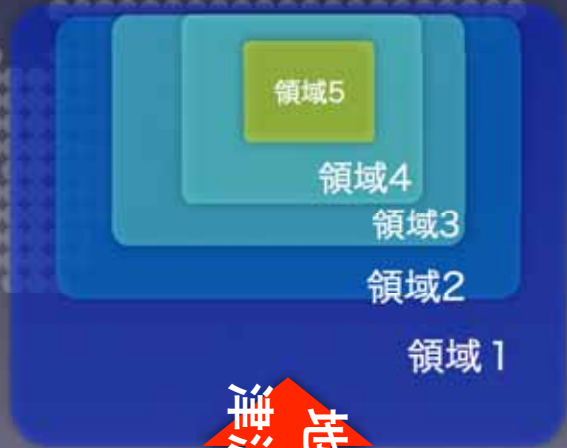
$\Delta T = 0.05$ 432,000ステップ (6時間)

浮動小数点演算数 2.5×10^{15} 回



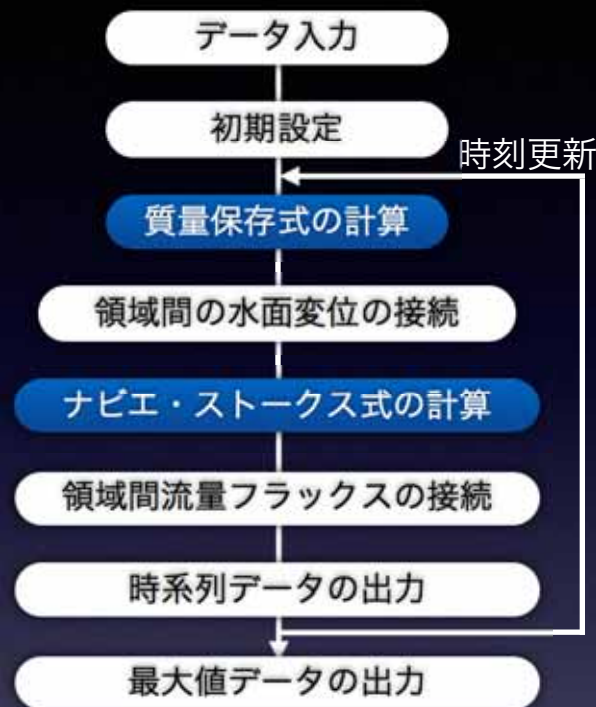
現在 研究室レベルでのPCサーバでは
約87時間必要
(IvyBridge, コア性能：19.1Gflop/s)

シミュレーション領域



津波浸水予測プログラム

プログラムフロー



★地震による海面水位の変動を入力として、津波が到達する時刻、津波の高さ等を計算

- 計算モデル：非線形長波理論
- 計算スキーム：Staggered leap-frog法

★プログラムの特徴

- Xeonサーバ用に開発されたコードがベース
- 質量保存の式とナビエ・ストークスの式において、計算負荷とメモリ負荷が非常に高い
- メモリバンド幅制約：B/F = 1.82 (単精度)

★チューニングのポイント

- ADBの効果的な活用
 - 差分式での局所性活用
- 高ベクトル化率：99.6%, 平均ベクトル長:235
- ループ内条件式の削減, インライン展開等

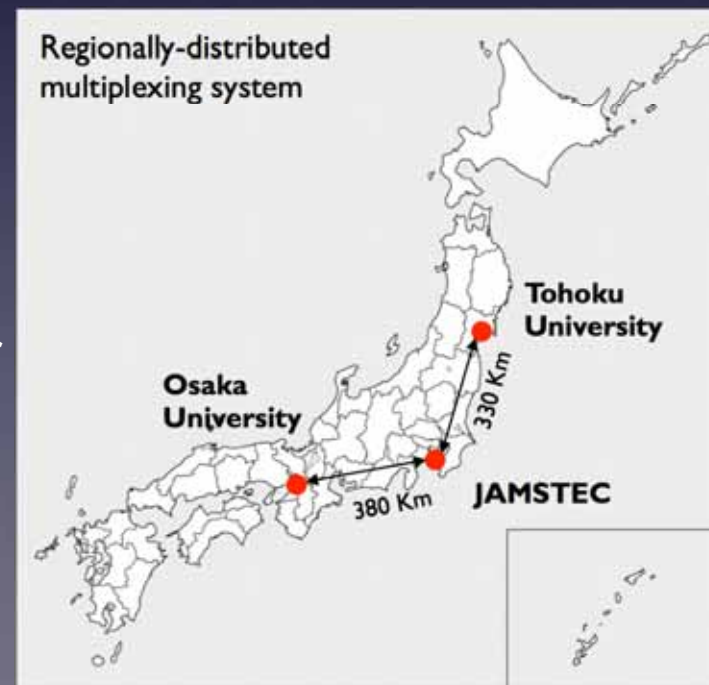
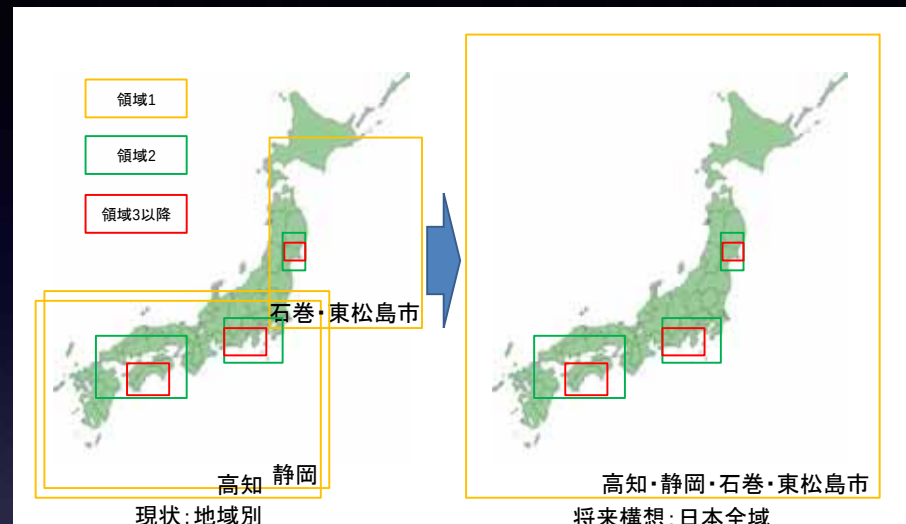


津波浸水予測プログラムのスケーラビリティ評価 (高知市の例)



今後の展開

- 3月16~18日 国連防災会議で成果発表
- 3月末 運用システムとして整備完了
- 4月以降
 - 対象エリア・対象自治体の拡大
 - 日本海側を含む全国展開へ
 - 阪大, JAMSTEC との連携
 - 相補的, かつ重層的にシステムを強化へ





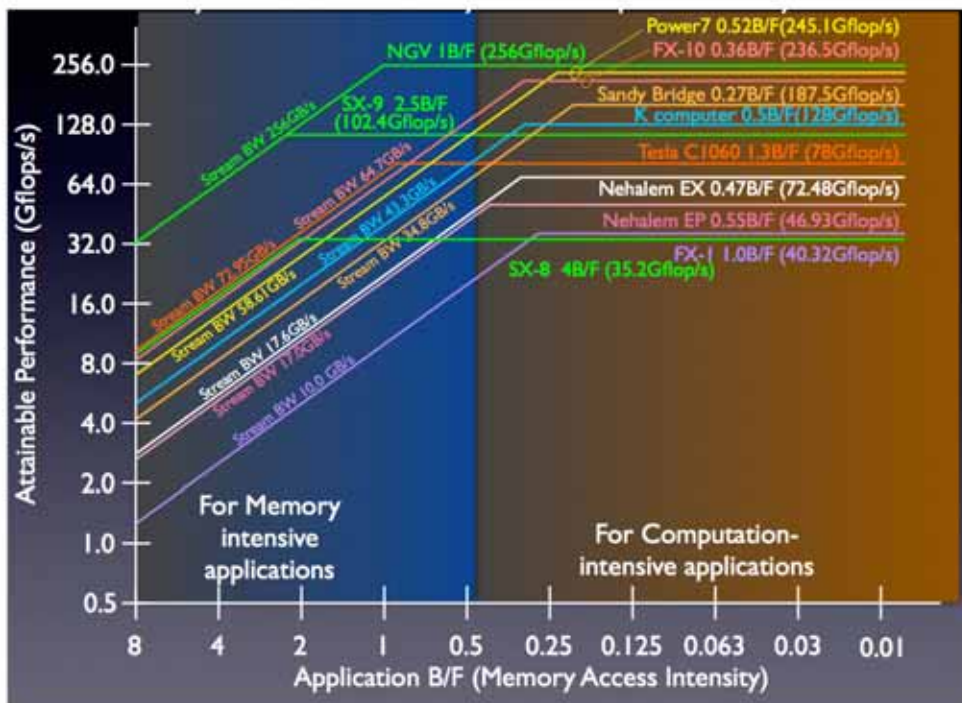
東北大学



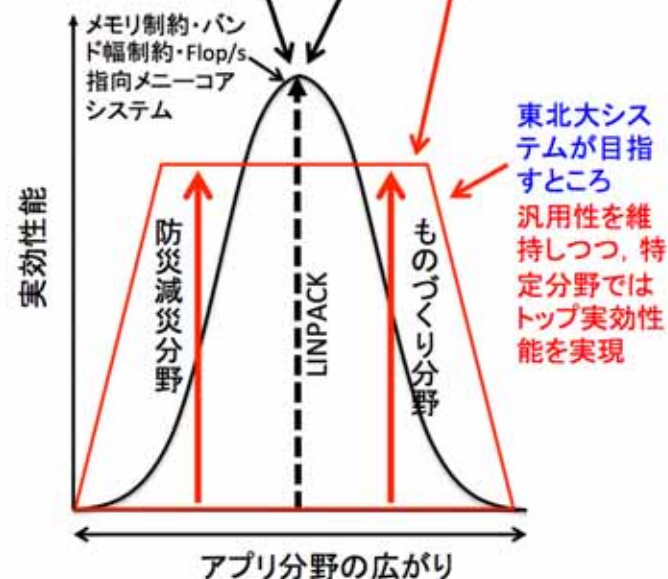
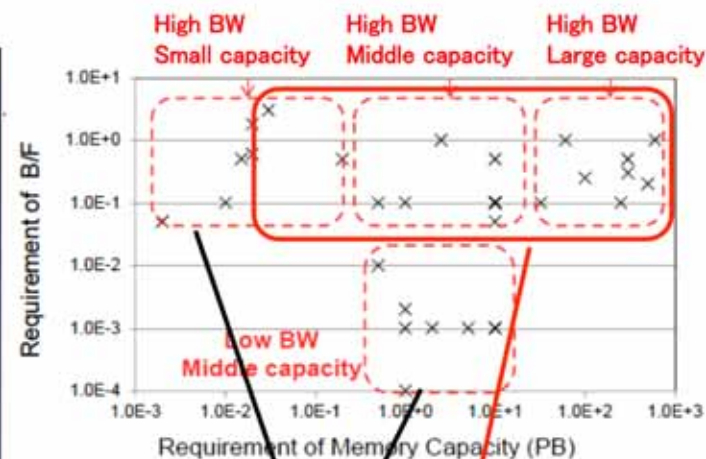
高メモリバンド幅アプリケーションのための 次世代スーパーコンピュータR&D

高メモリバンド幅アプリケーションのための次世代スーパーコンピュータに関する取り組み ～文科省「将来のHPCIシステムのあり方に関する調査研究」～

HPCI調査研究の背景と東北大チームがめざすところ



Flop/s性能向上にあわせてメモリバンド幅を高めることにより、可能な限り眠っている演算器を活用し、ノードレベルの超並列性に頼ることをできるだけおさえ、広汎なアプリで高い実効性能を実現する



高メモリバンド幅アプリの処理では、京の10倍のピーク性能で100倍のプログラム実行性能の実現をめざす



TOHOKU
UNIVERSITY



ポストペタ時代こそバランスの良いシステムが重要

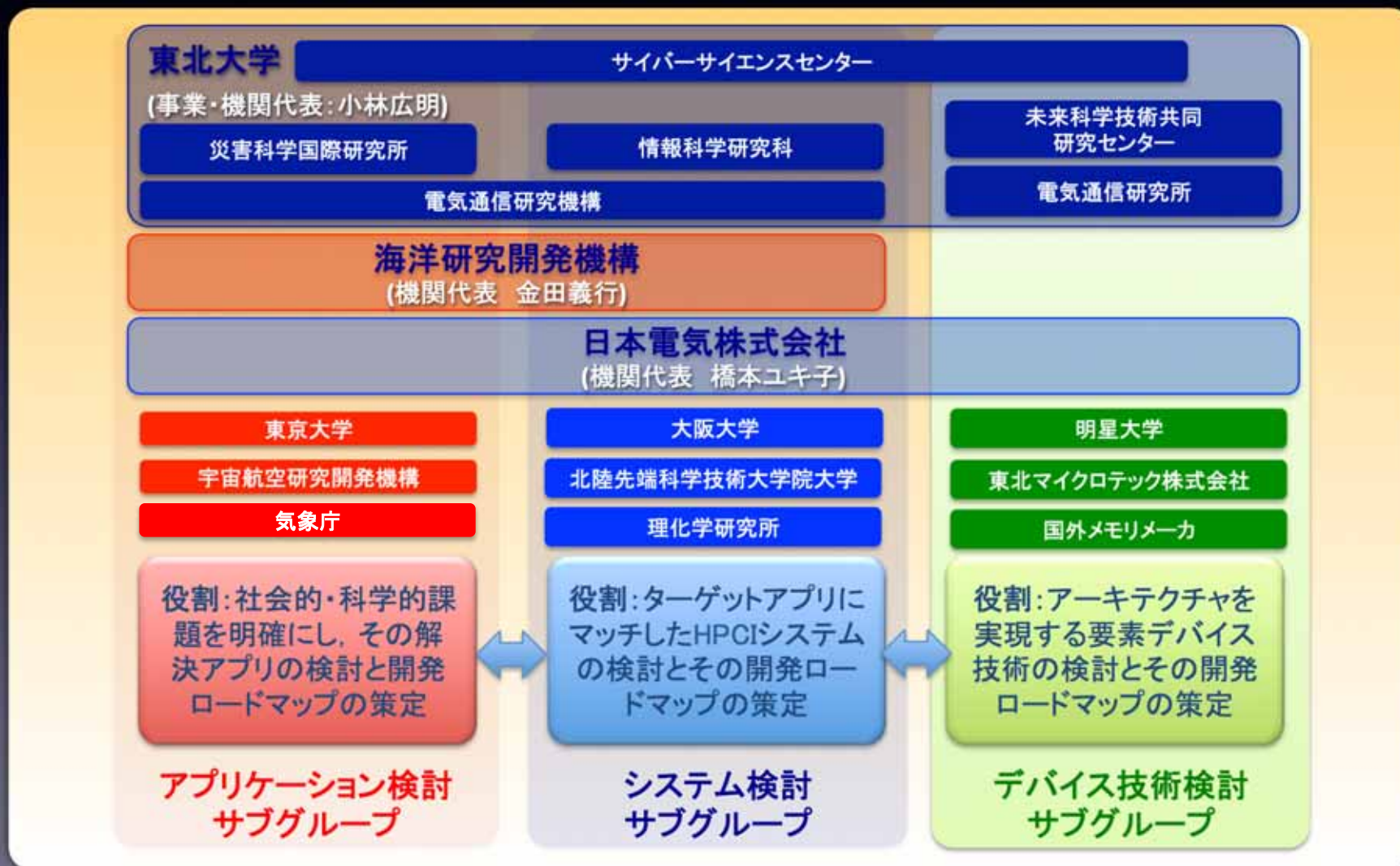
～ハイエンドHPCシステムの「脳梗塞」状態からの脱却～

- ★ データ転送隘路（メモリ制約）で膨大な数の演算器に十分データが届いていないため、演算器が無駄になっている
 - 例外としてLINPACK：90%以上の実行効率を達成できるので、CPU単体とシステム規模のピーク性能が意味を持つが．．．
 - 数%の効率を前提にシステム規模で必要性能をリカバーするまえに、チップ内に眠っている演算器を有効活用することを考えるべき。そのためには
- * メモリサブシステム設計（先進デバイスとアーキテクチャ技術の融合）と利用技術（ソフトウェアによるデータ管理）のコデザインによるシステム再設計をする時期
 - ✓ SIMDをワイド化してベクトルユニットを強化する方向性はIBM/AMD/INTELなども目指すところ。我々のFSはそれをすでに先取りしており、重要なポイントはそれを生かすメモリサブシステムをいかに他に先駆けて実現するかである
 - ★ メモリから撤退したといわれているインテルが次期プロセッサ向けメモリサブシステム設計/開発に取り組んでいる
 - ★ IBMはメモリインタフェースを強化し大容量・高バンド幅のpower8プロセッサを開発

東北大学チームは、チップ内ハイブリッドヘテロアーキテクチャ化が進むと思われる2018～2020頃に解決が求められている社会的・科学的課題として、高実効メモリバンド幅・低消費電力メモリサブシステムのアーキテクチャの確立も目標に掲げている

研究開発体制

アプリ・システム・デバイス三位一体でスパコンをバランス良く再設計!



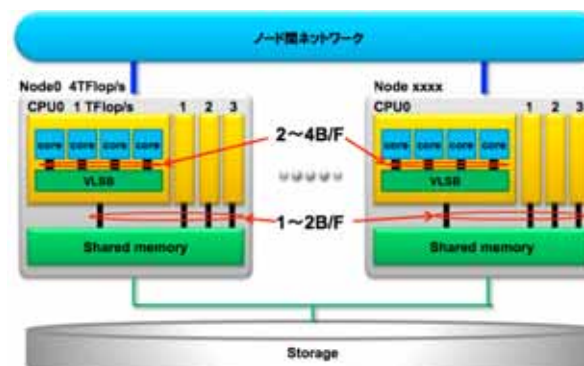
HPCI FSで得られたシステム概念設計結果

システム・ノードアーキテクチャ

■ 設計方針, 目標

アプリ要求性能に基づくシステム設計

- 限られた計算資源を最大限に活用できるシステム
- ・ システムピーク性能 100PF
- ・ システムメモリ性能 100 ~ 200 PB/s
- ・ システムメモリ容量 3.2PB ~ 6.4PB



■ 検討内容と結果～システム特徴～

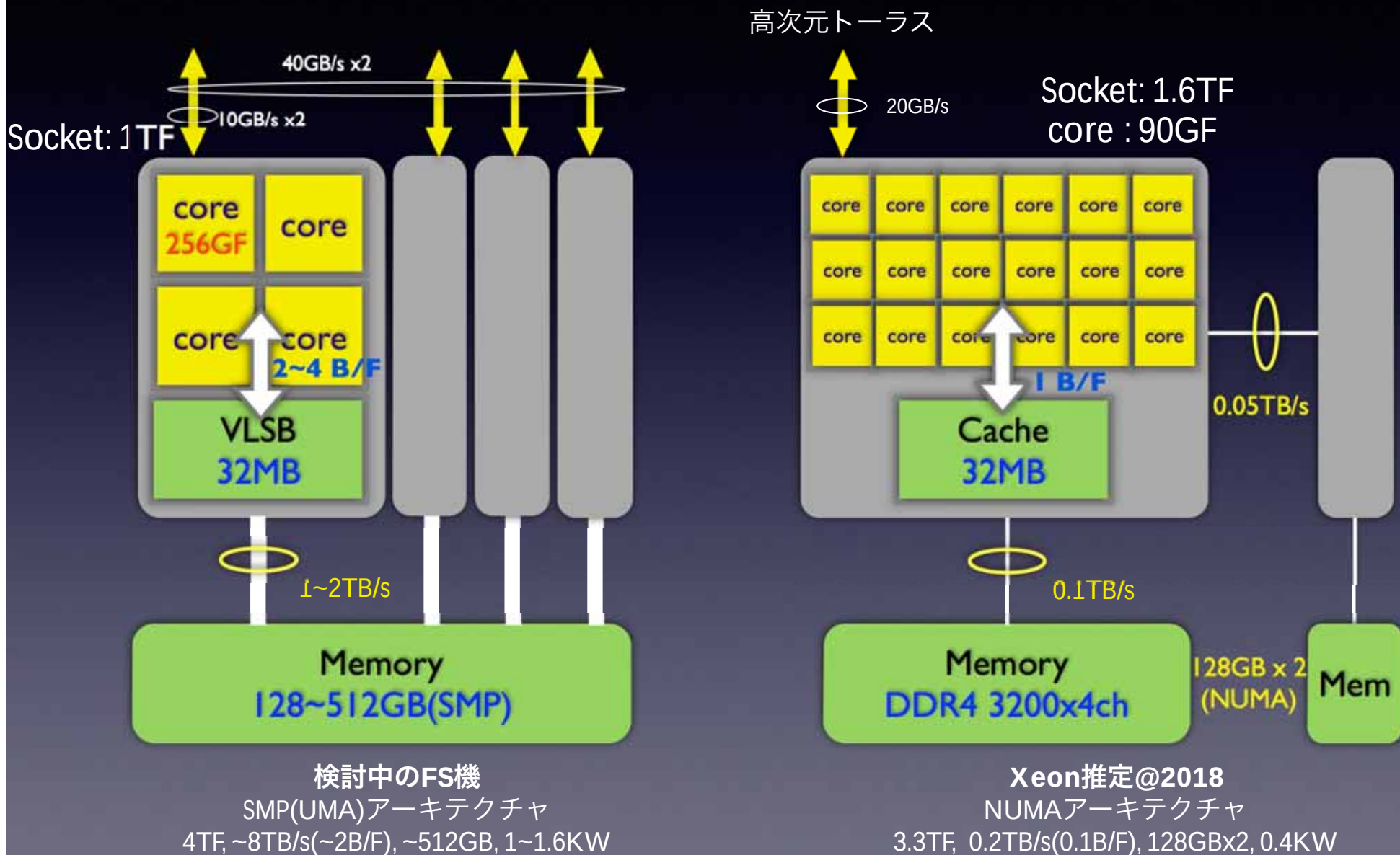
- ノード演算性能: 4TFlop/s
- UMAアーキテクチャ
- 高速信号伝送技術を用いた低消費電力ノード内通信
- 4CPU(16コア)によるSMPノード構成
- 各CPUは実行効率の高いマルチコア・ベクトルプロセッサ
- 演算器へのデータ供給性能重視設計

■ 広帯域・大容量共有メモリの実現

- ノード当たりのメモリ容量: 128~256GB
- メモリバンド幅: 4~8TB/s
- 2.5次元, 3次元積層技術など新規デバイス技術を用いた専用メモリモジュール実装による高メモリバンド幅の実現
- 汎用モジュール, 専用モジュールの利用も検討
- システム・デバイス協調設計

3次元実装技術など先端デバイス技術と先進ベクトル&メモリアーキテクチャをベースにアプリケーションとのコデザインでシステム設計を実現

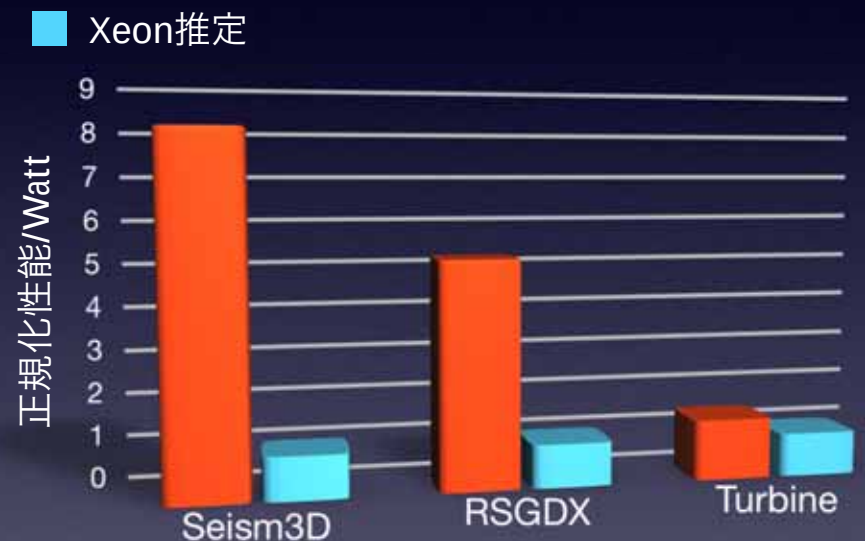
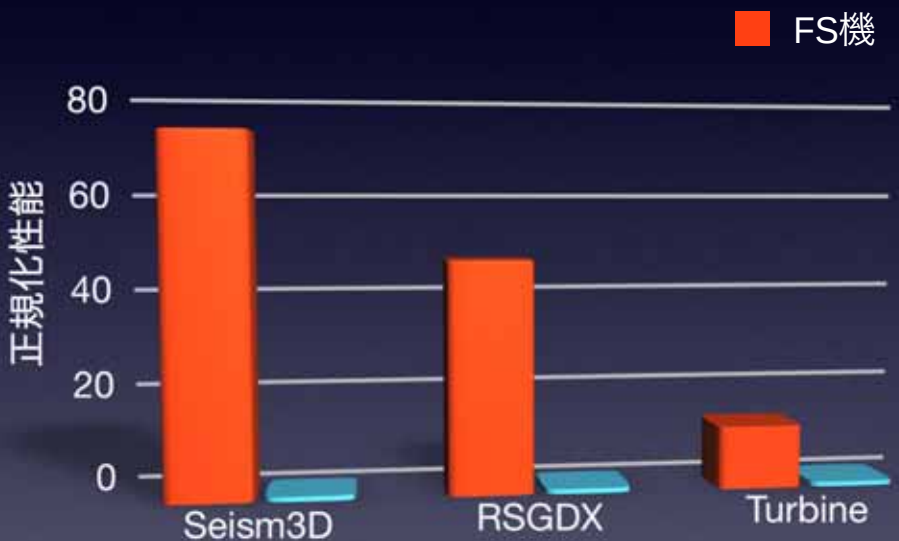
FS検討システムとXeonベース推定システムの比較



高メモリバンド幅システムの優位性(1/2)

～INTEL Xeonベースのシステムとの性能比較～

- 同一プロセス数(100,000proc)で実行した場合のFS機とXeon(推定) 機の性能比較
- 全体実行時間 (=通信時間+演算処理時間)を評価
- Xeon(推定)で正規化



	FS機	Xeon(推定)	比率
Tread/proc	4	4	
PeakPF/proc	1TF	0.36TF	2.79
PeakPF/100,000proc	100PF	35.84PF	2.7
Node数 /100,000proc	25,000	11,111	2.25
Total MemBW	200PB/s	3.4PB/s	58.6
Total キャッシュ容量	3.2TB	0.8TB	4

◆ 要求B/Fが高いSeism3D,RSGDXの消費電力あたりの性能は、FS機の方が4～8倍高い

◆ 要求B/Fが低い数値タービンにおいてもFS機はXeonと同等の性能/消費電力

高メモリバンド幅システムの優位性

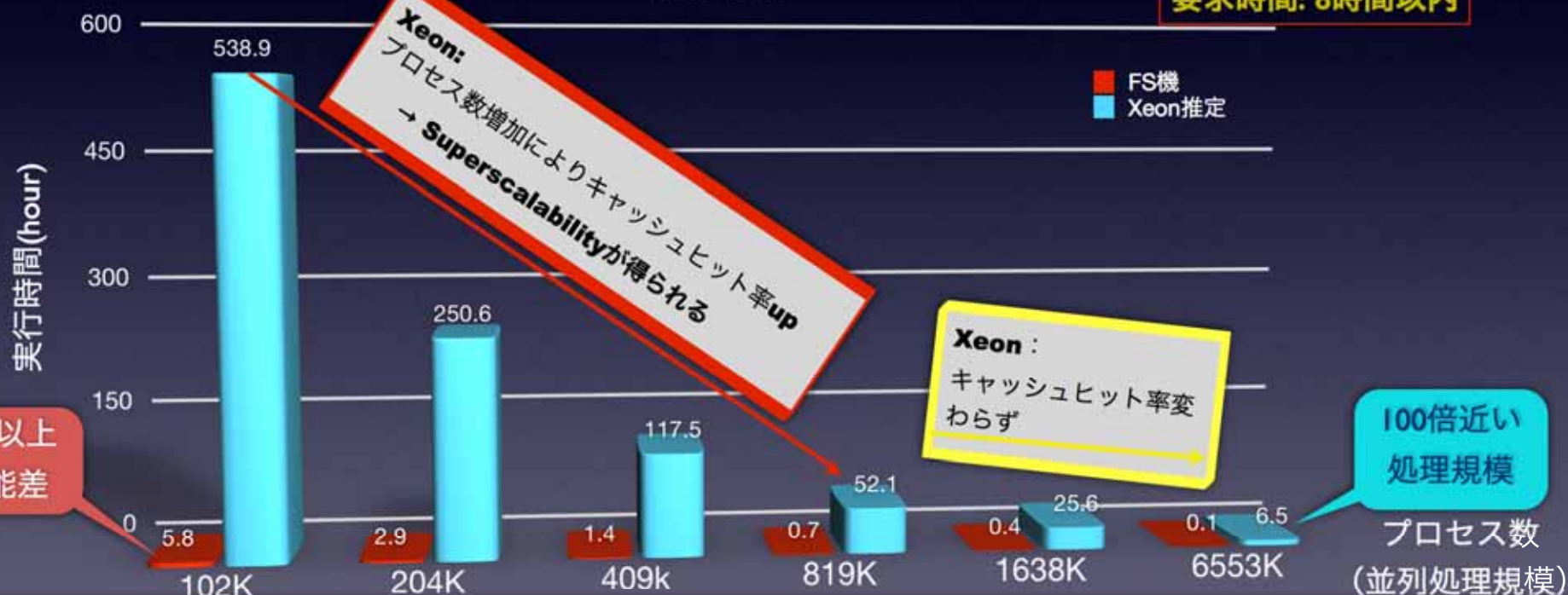
～INTEL Xeonベースのシステム(2018年頃)との性能比較～

●両システムが同等の性能を実現するためのプロセス数を評価

- ★全体実行時間（＝通信時間＋演算処理時間）を評価
- ★Xeon推定機の通信時間はFS機の通信時間と同一と想定

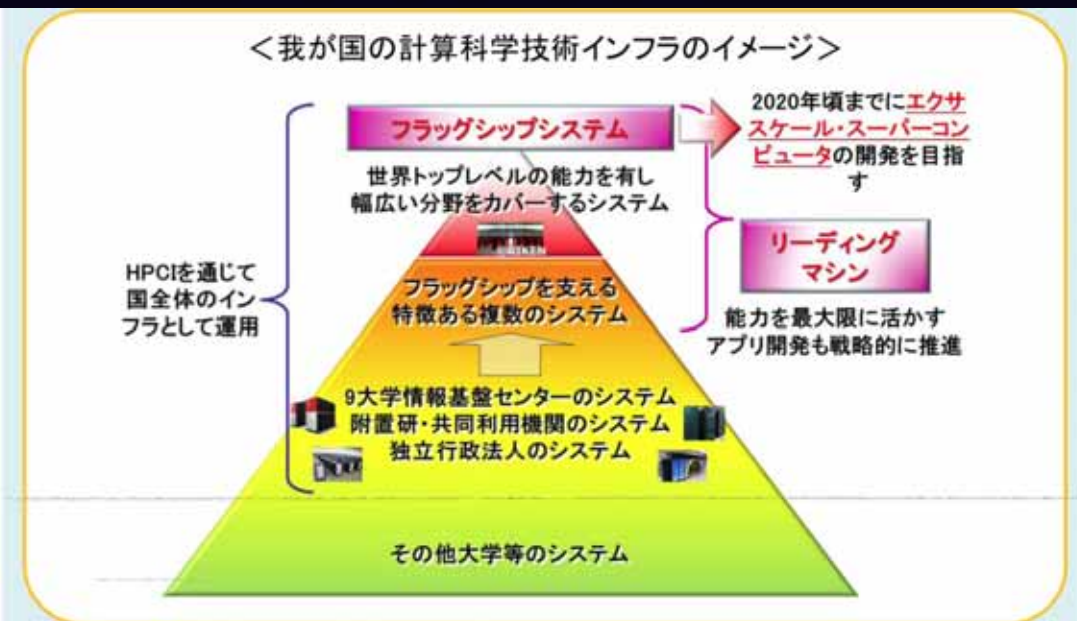
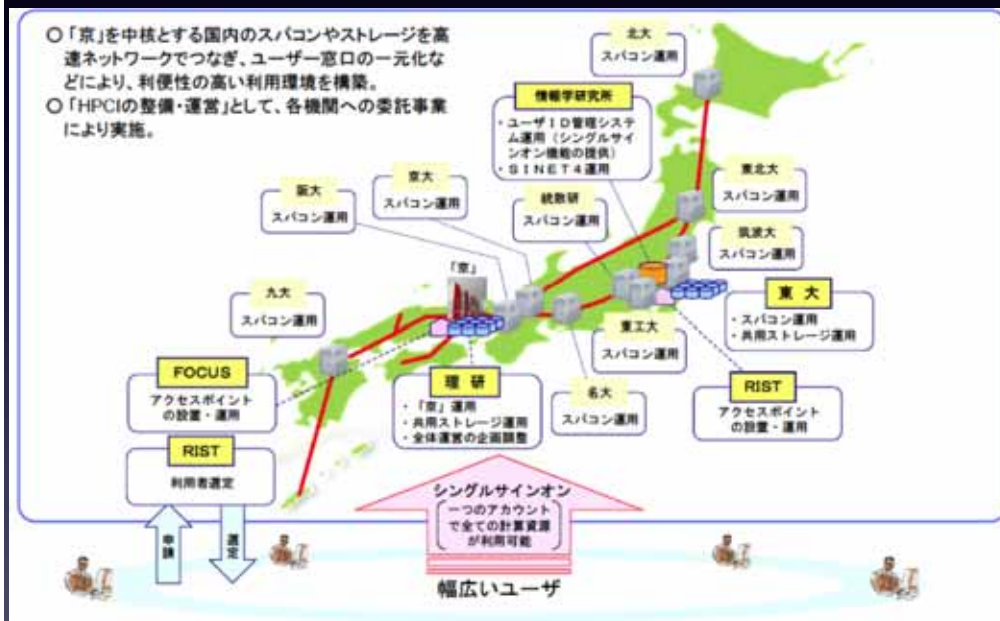
Seism3D

要求時間: 8時間以内



- プロセス数を増やす事でキャッシュヒット率が向上し、実行時間が短縮されるが、最終的にはキャッシュヒットしないリクエストのメモリアクセス時間が性能を決める **HWの持つメモリ帯域が大きいシステムが有効**
- Xeon推定機を用いてFS機(100Pflo/s)と同等の実行時間を実現するためには、**FS機の64倍のプロセス数、すなわち640万プロセスを処理できる640万ノードのシステム、ピーク性能3.2EFlop/sの性能が必要**
：そのときの消費電力は291MW(14nmを想定) → 非現実的

我が国のHPC基盤の安定維持・高度化を目指して ～東北大学サイバーサイエンスセンターに求められる役割～



現在のHPCIシステムの安定運用・高度化への貢献

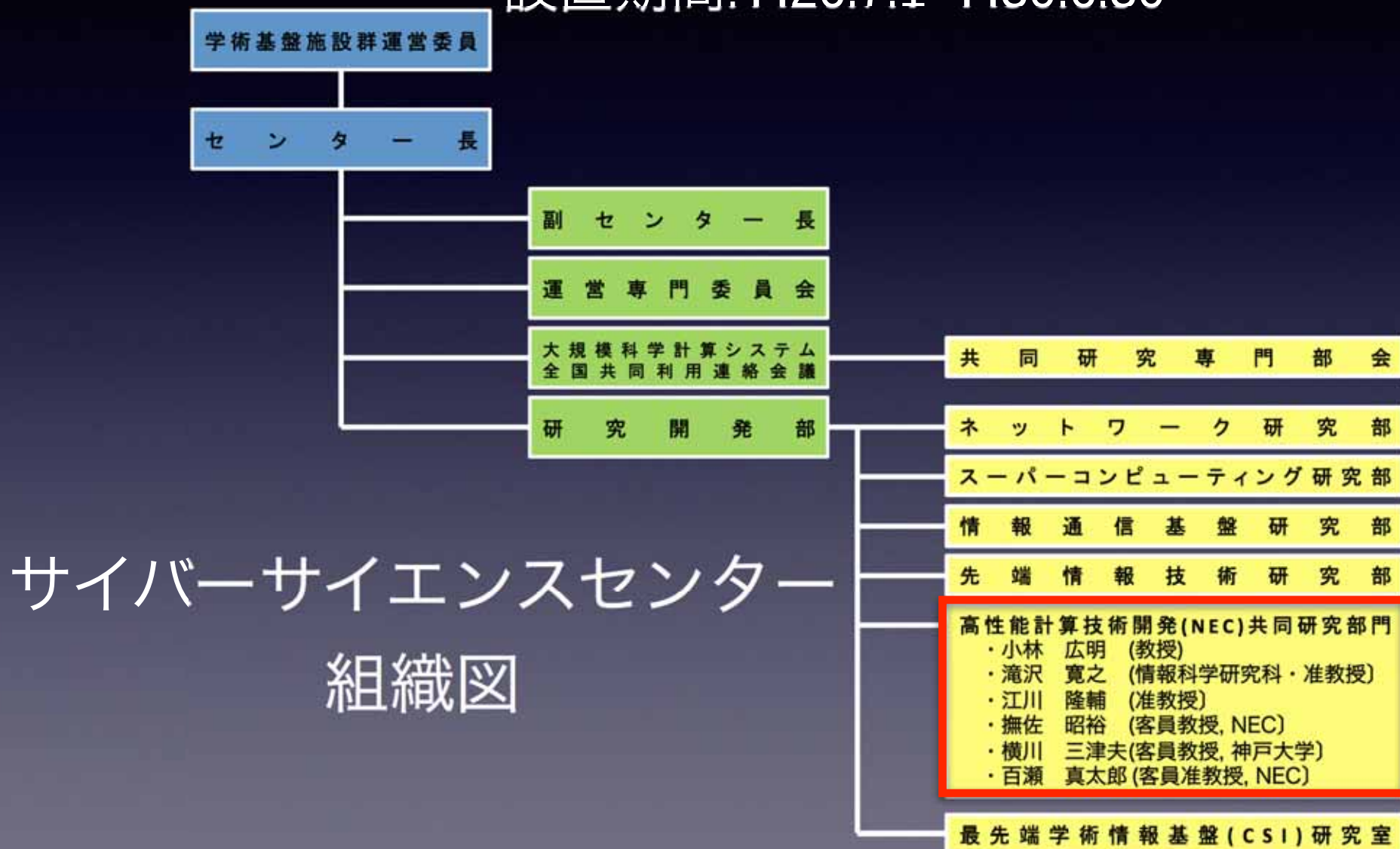
将来のHPCIシステムの研究開発への貢献

これらの取り組みを通じて、計算機科学・計算科学の両分野に精通した分野融合・横断型の人材育成

HPCに関する研究・人材育成機能の強化

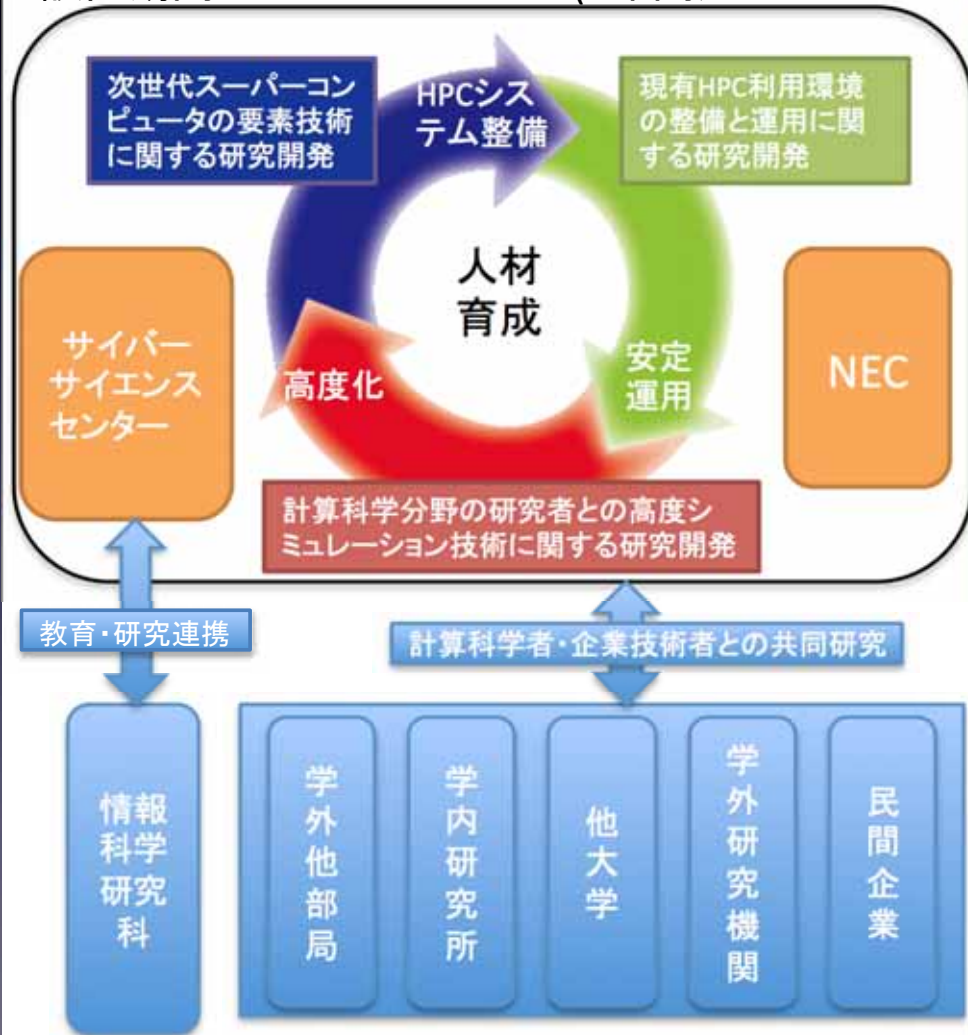
高性能計算技術開発(NEC)共同研究部門の新設

設置期間: H26.7.1~H30.6.30



高性能計算技術開発(NEC)共同研究部門の 取り組みとその期待される効果

設置期間:H26.7.1~H30.6.30(4年間)



- **我が国のHPC基盤の強化, およびリーディングマシンの研究開発への貢献**
 - HPCIシステムの高度化に関する研究開発の実施
 - ポスト京及びそれを補完するシステムが位置付けられるリーディングマシンの研究開発への貢献
- **学内外の計算科学者との共同研究の受け皿として, シミュレーション科学の高度化に貢献**
 - G空間を活用したリアルタイム津波シミュレーション(災害科学国際研究所との共同研究)
 - 防災・減災に資するスーパーコンピュータシステムの研究開発(JAMSTEC・東北大組織的連携協定, 電気通信研究機構との共同研究, その他)
- **部門教員が担当する情報科学研究科の基幹・協力講座において, 計算科学・計算機科学の両分野に精通した学際的, かつ実践的人材育成**
 - NECおよびその他外部協力機関から客員教員の招聘し、研究・教育に従事
 - H26年度情報科学研究科プロジェクト「ビッグデータ解析利活用ソリューションラボの構築」への参画
- **HPCを通じて産学連携研究拠点の構築**
 - NECとの次期スパコンに関する共同研究
 - これまでに, MRJに対する航空機開発支援を含め7社へスパコン提供・利用支援を実施

さいごに

高性能計算に関する共同利用・共同研究拠点として、東北大学サイバーサイエンスセンターのミッション

- ★メモリ性能と演算性能がバランスした最先端のスーパーコンピュータシステムの導入・提供
- ★使い易い利用環境の整備・提供、利用者支援
- ★現有スパコンの利用技術、および次世代スパコン実現のための要素技術に関する研究開発
- ★研究基盤としてだけでなく、社会基盤、産業基盤としてのスパコンのあり方を追求

を果たし、

- ★普通の人々に役立つスーパーコンピュータセンター

Supercomputer Center for the Rest of US!

をめざします

