

地球シミュレータの アーキテクチャ・運用・成果

海洋研究開発機構
地球シミュレータセンター
板倉憲一

2012. 4. 27 京大セミナー

内容

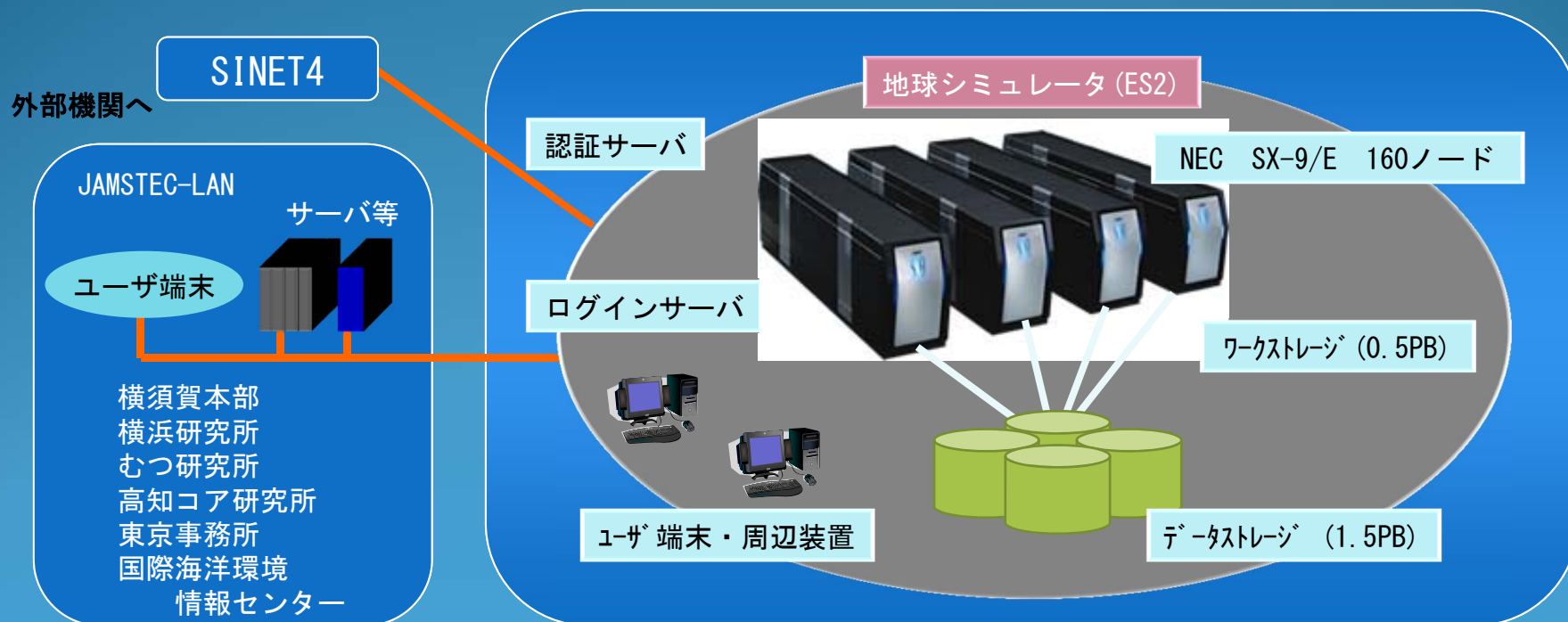
- 地球シミュレータ
 - 概要と運用状況など
- コンピュータアーキテクチャと高速化
 - スーパーコンピュータ(スパコン)とは
 - どうして高速に計算ができるのか
 - 地球シミュレータの特徴
- まとめ

地球シミュレータ (ES2) の概要(1)

地球シミュレータは、2002年3月に地球温暖化を始めとする気候変動の解析・将来予測、地震や地球内部変動の解明等、世界に類を見ない「人類的課題に挑戦できる世界最高速のスーパーコンピュータ」として運用を開始した。

地球シミュレータは運用開始から2年半もの間世界最速のスーパーコンピュータであったが、利用者からの膨大な演算処理要求によってフル稼働の状況が続いていた。そこで更なる性能向上を計るため、2009年3月に新システムへの更新を行い、計算能力の向上と消費電力量の低減に成功している。

■ システム概念図



地球シミュレータ (ES2) の概要(2)

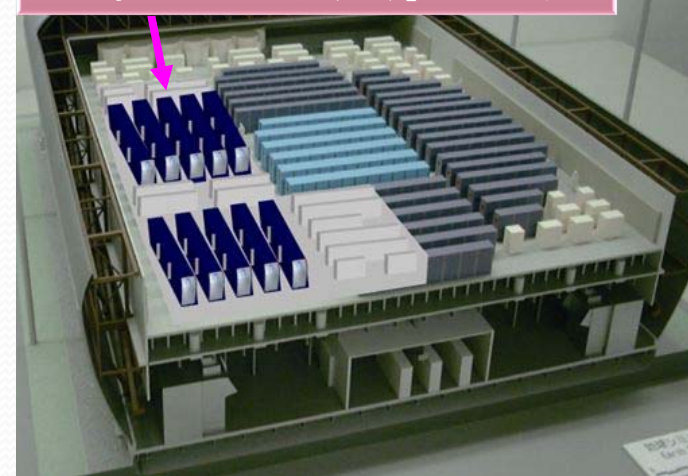
■ 更新スケジュール

	2007 (H19)	2008 (H20)	2009 (H21)	2010 (H22)	2011 (H23)
地球シミュレータ	運用 (2009. 3末まで)		システム一部撤去・データ移行		
地球シミュレータ (ES2)	仕様検討	調達	運用 (2009. 3より6年リース)		

■ ハードウェアスペック

		地球シミュレータ	地球シミュレータ(ES2)	性能比
CPU	クロック	1GHz	3.2GHz	3.2X
	ベクトル性能	8GF	102.4GF	12.8X
	メモリ転送性能	32GB/s	256GB/s	8X
ノード	CPU数	8	8	1X
	ベクトル性能	64GF	819.2GF	12.8X
	メモリ容量	16GB	128GB	8X
	ノード間転送性能	12.3GB/s x2	64GB/s x2	5.2X
システム	ノード数	640	160	1/4X
	演算性能	40TF	131TF	3.2X
	メモリ容量	10TB	20TB	2X
	NWトポロジ	フルクロスバ (回線交換方式)	2段ファットツリー (パケット交換方式)	—

「地球シミュレータ (ES2)」レイアウト



地球シミュレータ (ES2)

運用状況(利用プロジェクト)

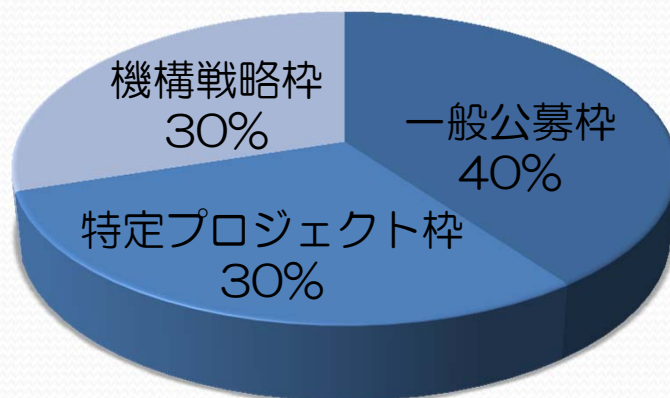
一般公募プロジェクト(*1)	29件
特定プロジェクト(*2)	19件
機構戦略枠によるプロジェクト	6件
産業界等からの利用プロジェクト(有償利用)	12件
外部(国内外)研究機関との共同研究	6件

(*1)

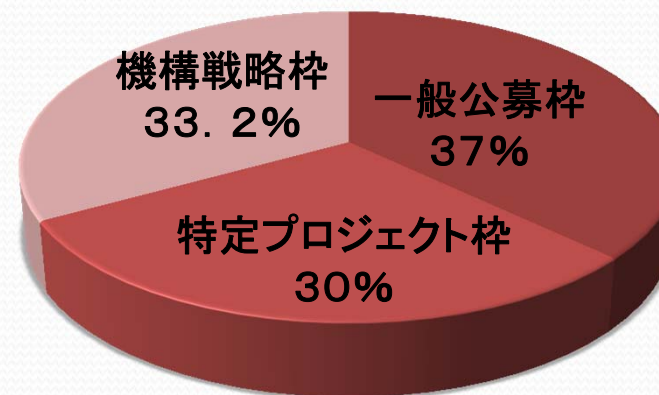
- ・地球科学分野 18件
- ・先進・創出分野 11件

(*2) 文科省等の委託・補助によるプロジェクト

- ・21世紀気候変動予測革新プログラム 5件
- ・地球シミュレータ産業戦略利用プログラム 13件
- ・戦略的創造研究推進事業(CRESTタイプ) 1件



2011年度リソース配分



2011年度使用実績

(2012年3月末現在)

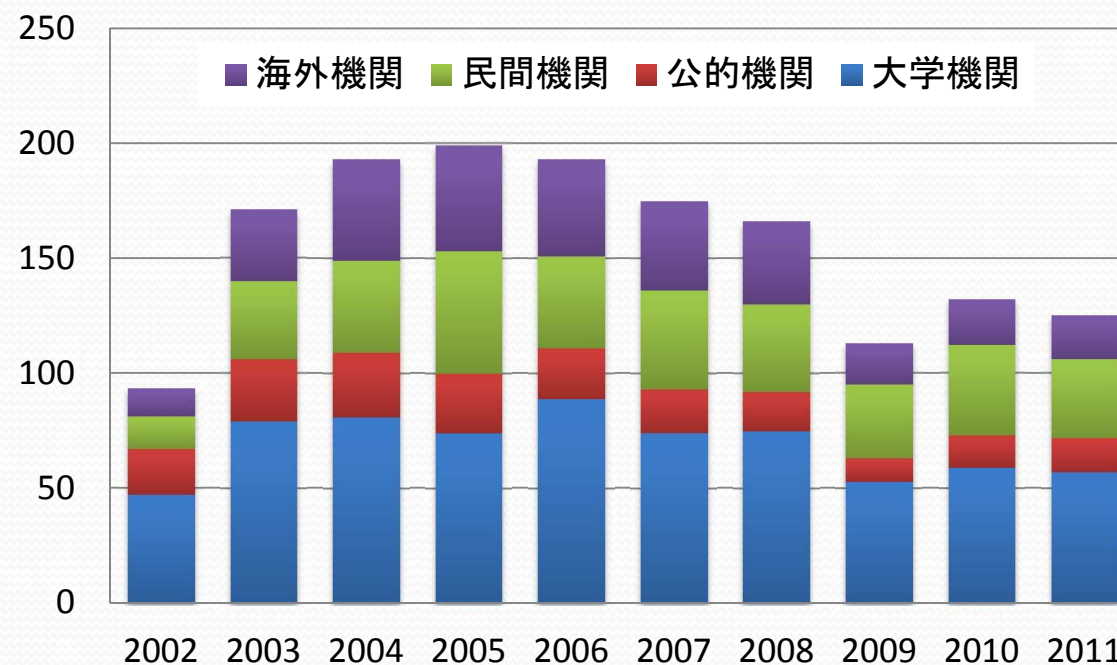
運用状況(登録利用者数)

- 利用者数 565ユーザ
- 利用機関数 125機関

※2012年3月末現在

大学機関:57	公的機関:15	民間機関:34	海外機関:19
---------	---------	---------	---------

地球シミュレータ利用機関の内訳



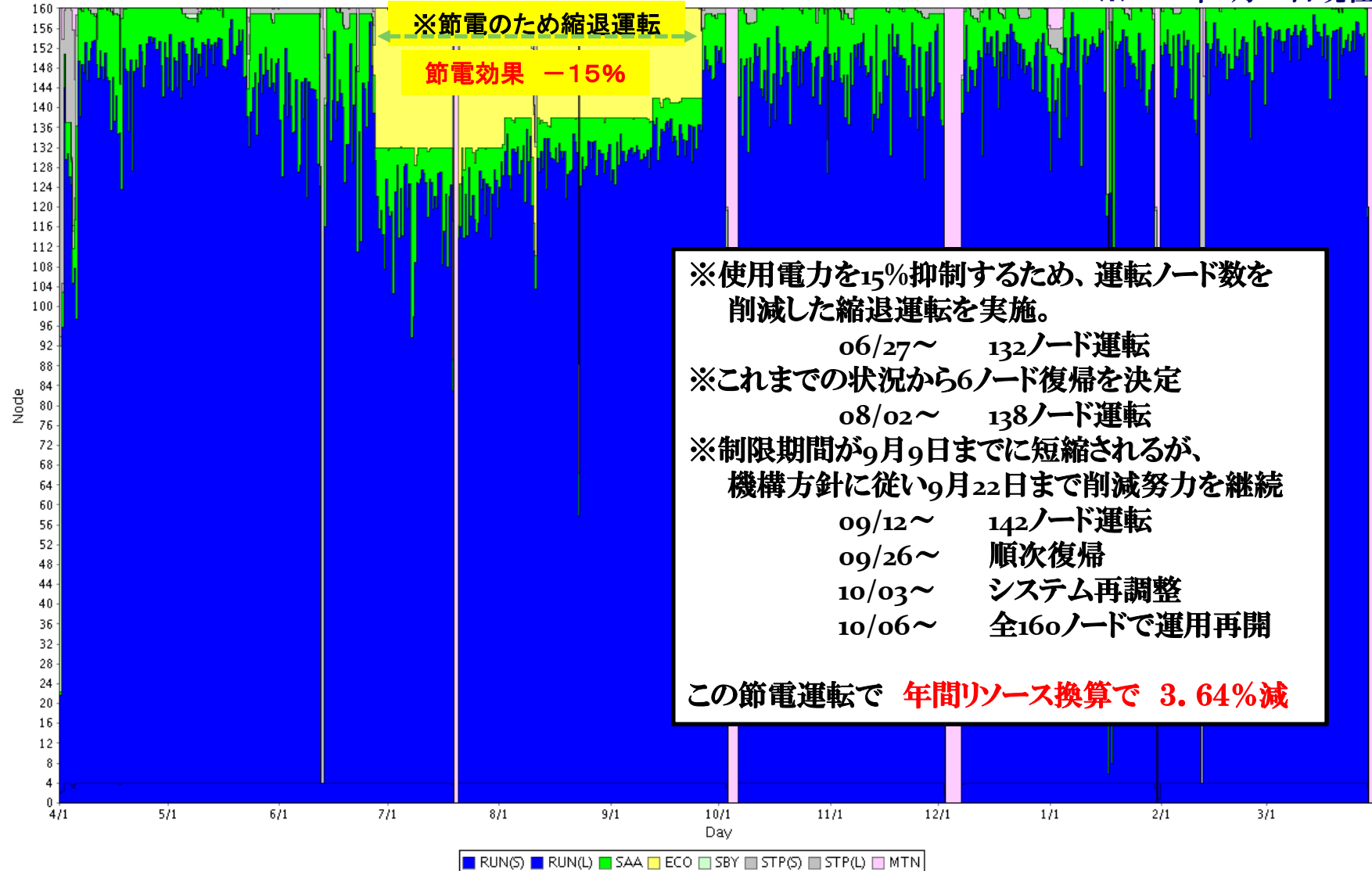
運用状況(ノード使用状況)

90%以上の
稼働実績を達成

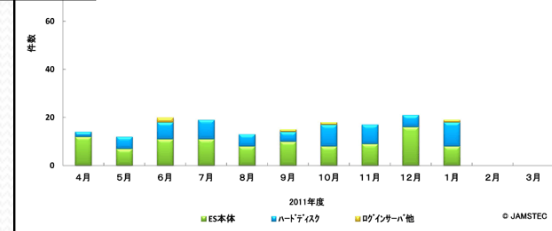
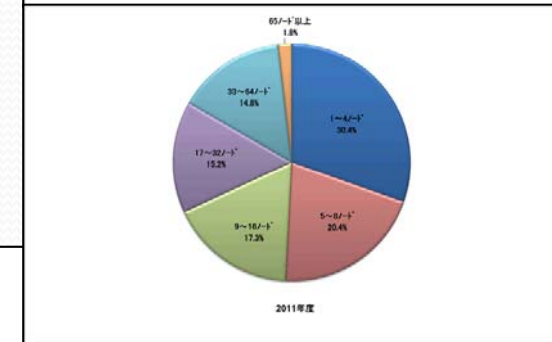
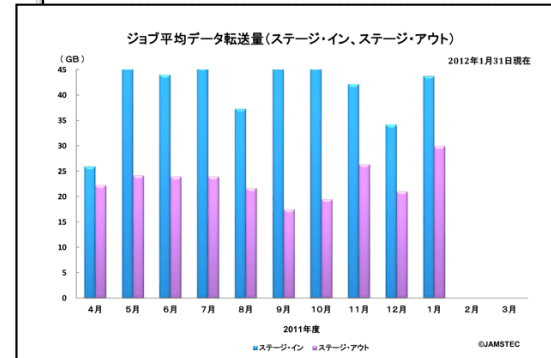
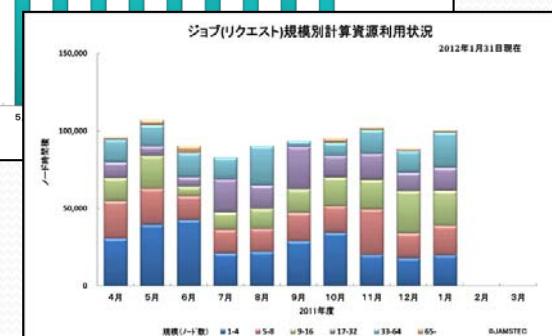
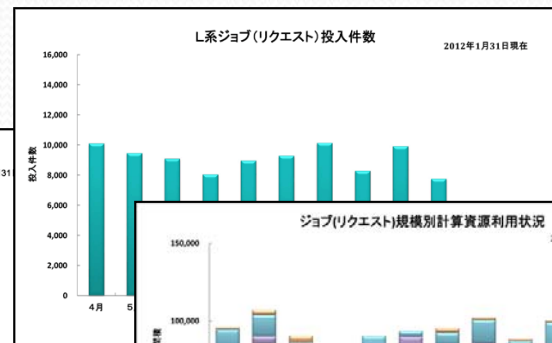
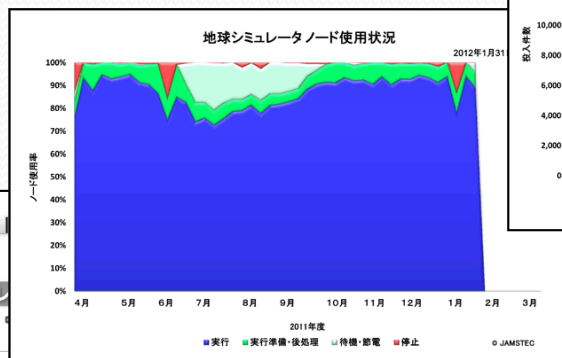
※2012年3月31日現在

Node usage in fiscal 2011

RUN: 84.62 % SAA: 6.45 % ECO: 3.64 % SBY: 0.30 % STP: 1.41 % MTN: 3.58 %



情報提供等：一般向け情報公開ウェブページ



■ 使用状況、ジョブ数、障害件数等、システム運用状況を引き続きウェブで一般に公開

成果の発信:ES研究成果リポジトリ

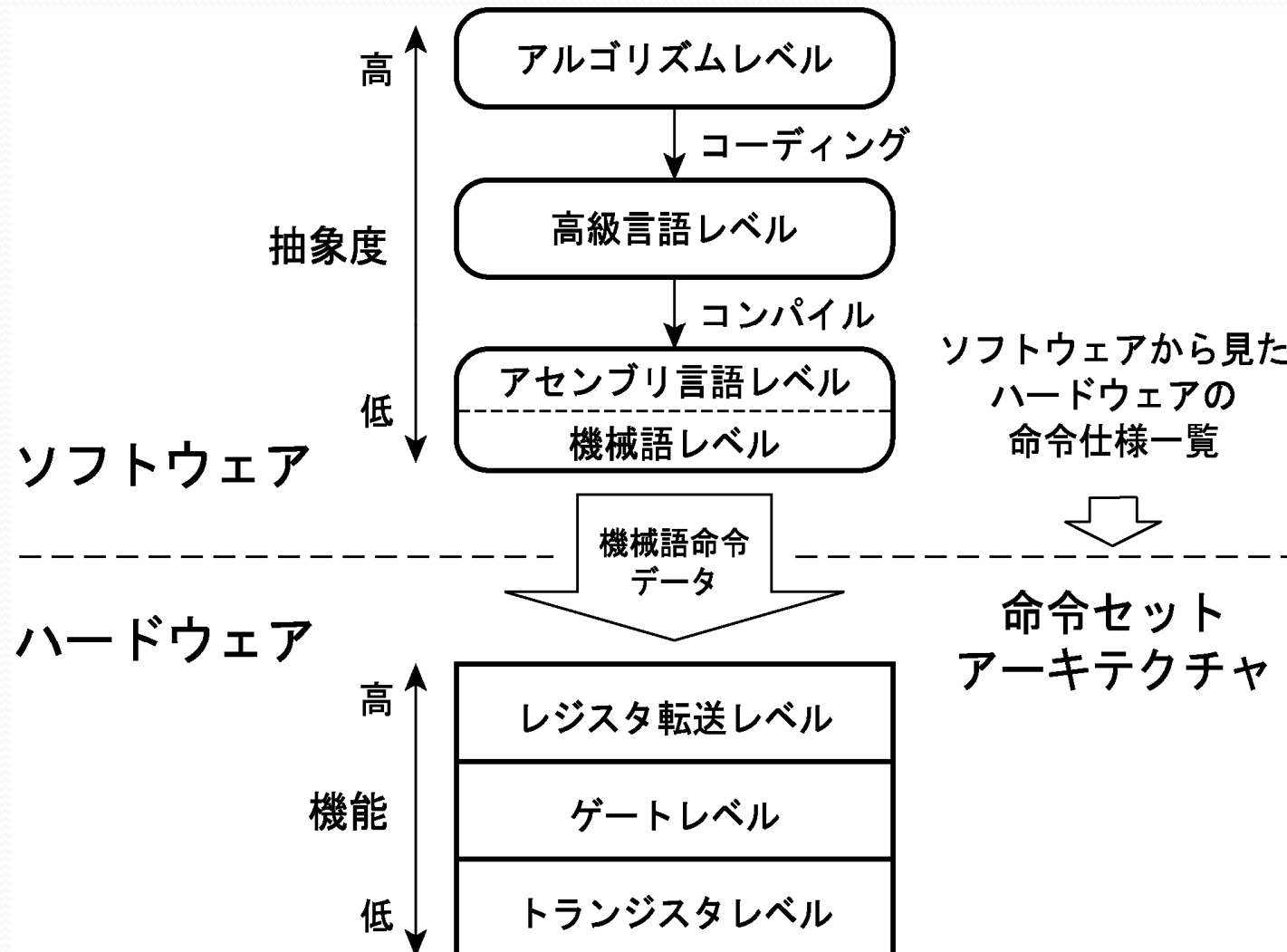


- ES研究成果を機関リポジトリとして公開
2012年3月末までの登録数 5095件
- 利用者の実態に合わせたデータ属性の追加や検索結果表示の強化等を行った。

コンピュータアーキテクチャ

- アーキテクチャ(Architecture):
建築, 建築術, 建築様式, 構造, 構成, 設計, 体系 等
- コンピュータアーキテクチャ:
計算機的设计方式、構造
- コンピュータのソフトウェアとハードウェアの境界仕様. ソフトウェアから見たハードウェアの仕様

コンピュータの階層レベル



ハードウェアの発展と設計技術(1)

- ハードウェアの性能向上
 - 集積回路技術の向上
 - 微細加工技術, 高集積度
 - 動作の高速化を牽引
 - ムーアの法則
 - 集積回路におけるトランジスタの集積密度は、18～24か月ごとに倍になる、という経験則
 - システム LSI へ
 - 従来個別にボード上に載せていた複数のLSI(CPUや各種ASSPおよびメモリ)を1チップ化

ハードウェアの発展と設計技術(2)

年	プロセッサ	プロセスルール	動作周波数
1971	4004	10 μ m	0.75MHz
1974	8080	6 μ m	2MHz
1978	8086	3 μ m	5MHz
1982	i80286	1.5 μ m	8MHz
1985	i80386	1.0 μ m	16MHz
1989	i80486	1.0 μ m	33MHz
1993	Pentium	0.8 μ m	66MHz
1997	Pentium II	0.35 μ m	300MHz
1999	Pentium III	0.18 μ m	500MHz
2001	Pentium4	0.13 μ m	1.3GHz

マイクロプロセッサ？スパコン用プロセッサ？ スカラ？ベクトル？

- WindowsパソコンやMacで使われているマイクロプロセッサはスカラ型のCPU
- 地球シミュレータで使われているプロセッサはベクトル型のCPU
- 初期のCPUは沢山のトランジスタで構成
- トランジスタのIC化、LSI化が進む。
- 世界初のマイクロプロセッサ＝1つのLSIのCPUは1970年のIntel4004
- 1974年の8bitプロセッサ Intel 8080 は広く普及した。
- 1970年代のスパコンは？
- 1976年 ベクトル型Cray-1 250MFLOPS ロスアラモス国立研究所, ニューメキシコ州, アメリカ
- ICは5層のプリント基板上に実装され、基板1枚に144個のICが搭載されている。基板は冷却のために背中合わせに実装され、24台の28インチ高のラックに72組の背中合わせの基板がセットされた。モジュール(個別の演算ユニット)は一枚か二枚の基板で構成される。マシン全体では 113種類 1,662モジュールが使われた。

Cray-1



高速化のカギは並列化

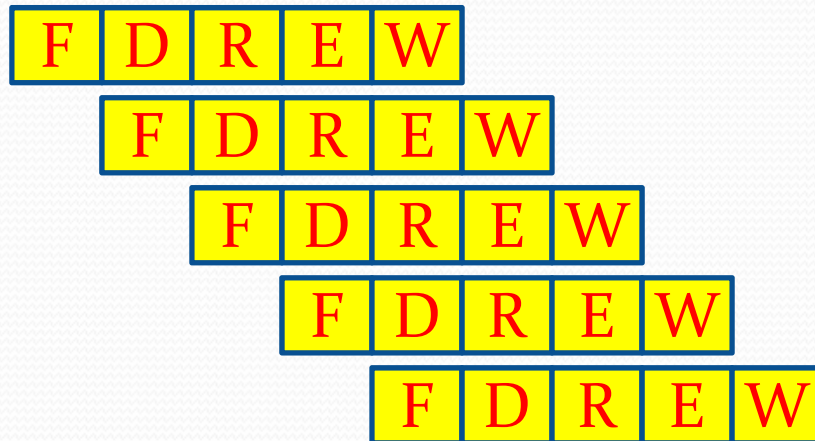
- プロセッサの内部での並列化＝高速化
 - パイプライン並列化
 - 同時命令演算並列化(スーパースカラ並列化)
- 沢山のプロセッサを繋いで並列化＝高速化
 - 密結合の並列化
 - 疎結合の並列化

パイプライン並列化

- 非パイプイン化 5クロックに1回演算



- パイプライン化



1クロックに1回演算

- 1命令の処理
 - F: 命令取り出し
 - D: 解析
 - R: データ取り出し
 - E: 演算
 - W: データ書き戻し

同時命令実行並列化

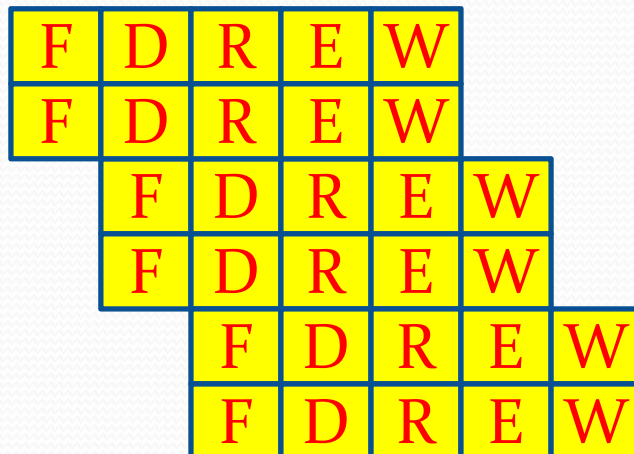
- 非並列化 5クロックに1回演算



- 2命令並列化 5クロックに2回演算



- さらに、パイプライン化 1クロックに2回演算



ベクトルプロセッサ

- ベクトル演算命令
(イメージ)

F	D	R	E	W
		R	E	W
		R	E	W
		R	E	W
		R	E	W
			⋮	
		R	E	W

(実際)

F	D	R	E	W		
		R	E	W		
		R	E	W		
		R	E	W		
			R	E	W	
			R	E	W	
			R	E	W	
			R	E	W	
				R	E	W
				R	E	W
				R	E	W
				R	E	W

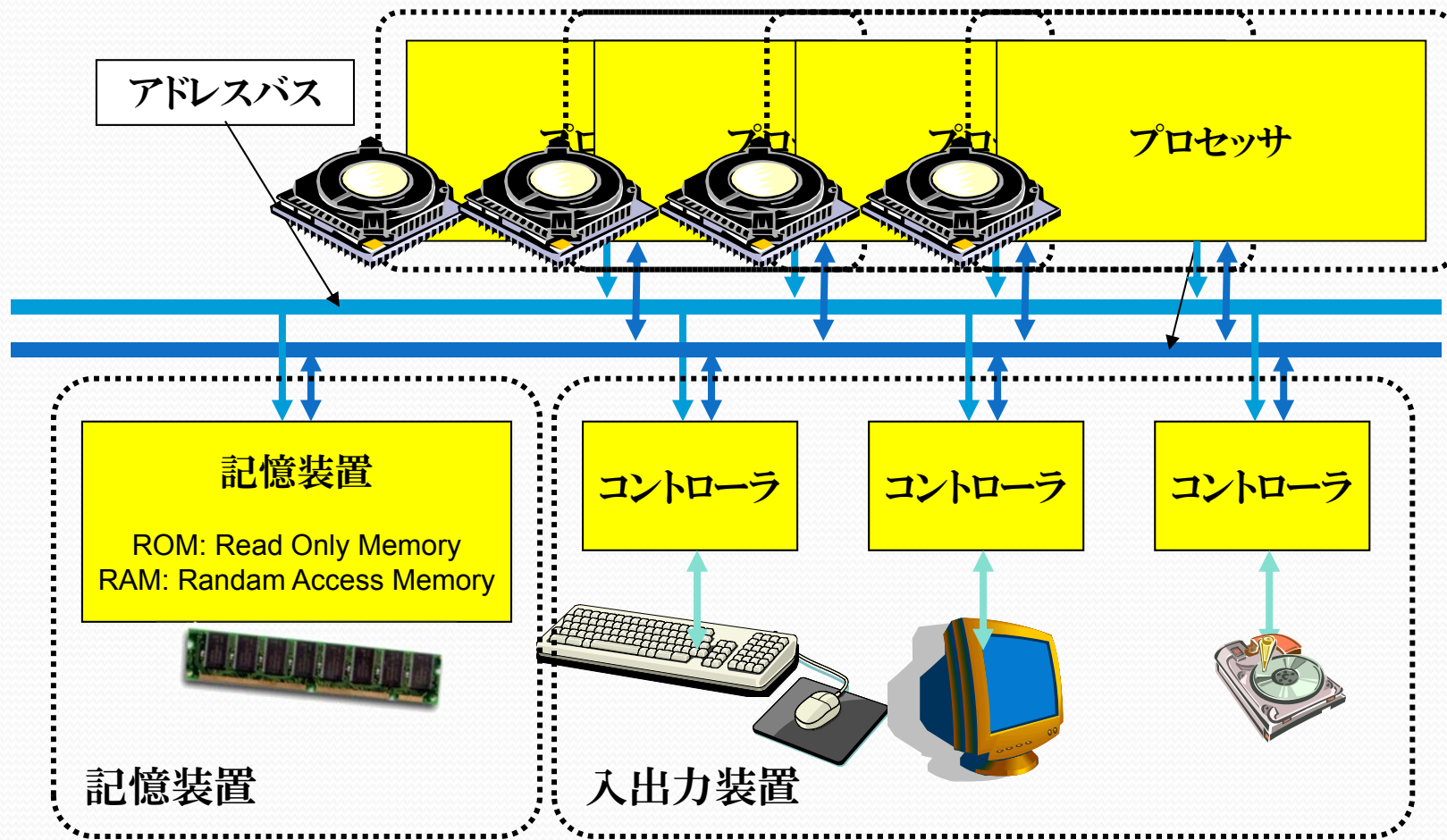
演算並列性

応用された。

パイプライン並列性と同時演算並列性
を自然に利用
⇒スカラプロセッサに技術応用された。

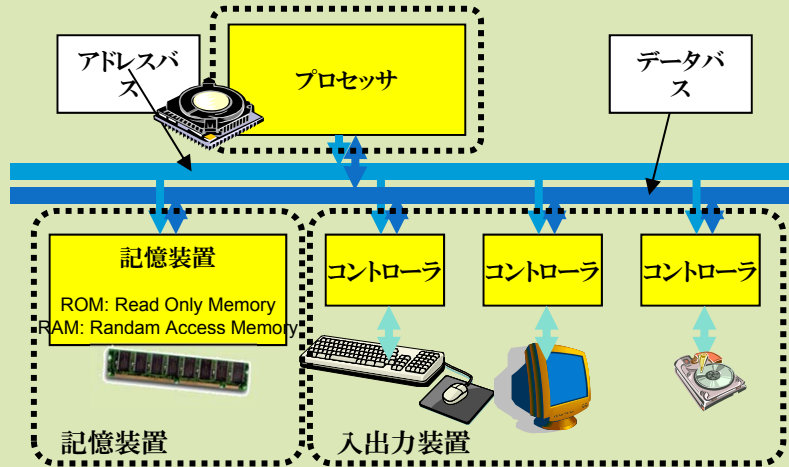
密結合の並列化

ノイマン型計算機の実例の例

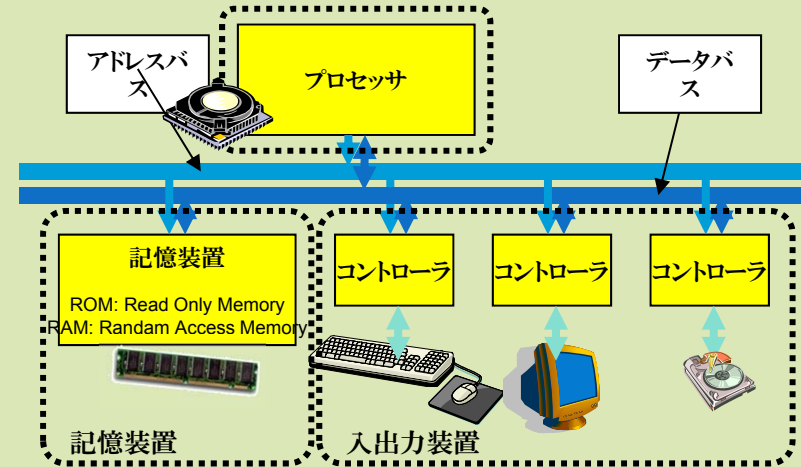


疎結合の並列化

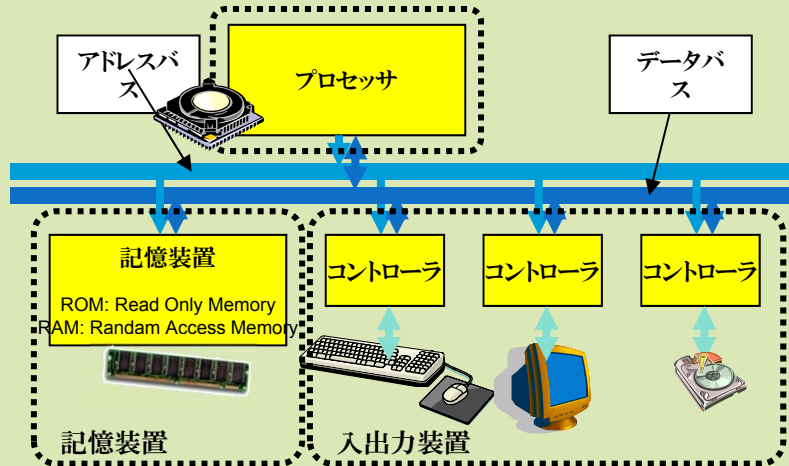
ノイマン型計算機の実例の例



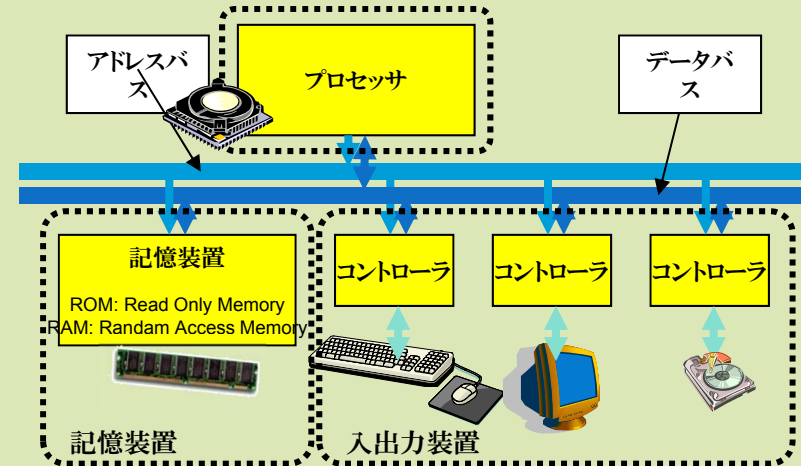
ノイマン型計算機の実例の例



ノイマン型計算機の実例の例



ノイマン型計算機の実例の例



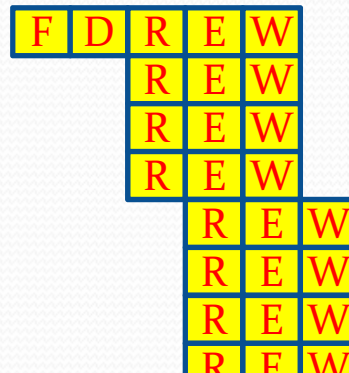
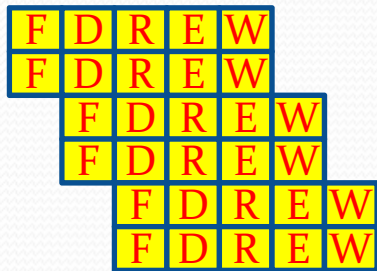
演算性能

- スピードはどうやって測るのか？
 - 1秒間に何回演算できるか 1秒間に10回で 10FLOPS(フロップス)
 - | | | | |
|------------------------|-----|----|------|
| 1,000回 | 千回 | 1K | (キロ) |
| 1,000,000回 | 百万回 | 1M | (メガ) |
| 1,000,000,000回 | 十億回 | 1G | (ギガ) |
| 1,000,000,000,000回 | 一兆回 | 1T | (テラ) |
| 1,000,000,000,000,000回 | 千兆回 | 1P | (ペタ) |
- 理論ピーク性能
- 実効性能
 - アプリケーションプログラム
 - ベンチマークプログラム
 - TOP500リストではLINPACK

計算性能とメモリ性能

- データはどこにあるのか？
CPU内のレジスタのデータしか演算できない。
- レジスタとメインメモリでデータの入れ替えが必要
load/store 命令
- どんなに演算器を用意しても
データが無ければストールしてしまう。

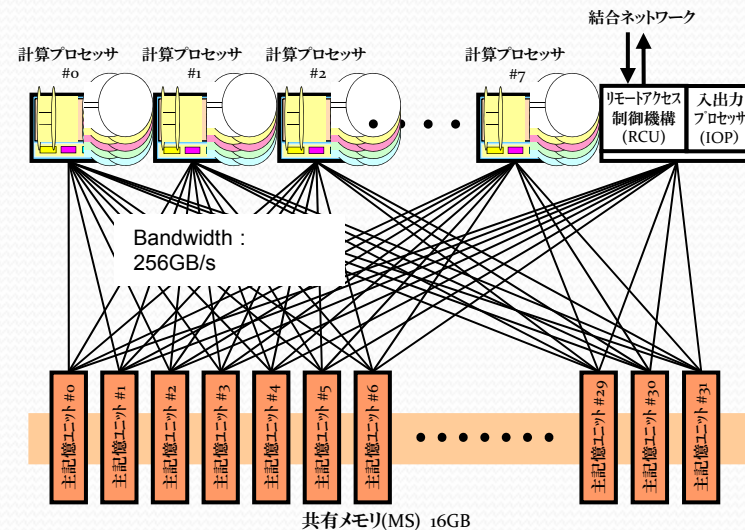
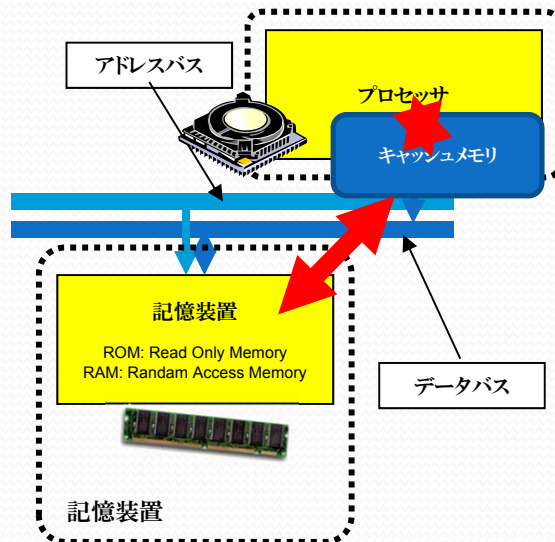
- 1命令の処理
 - F: 命令取り出し
 - D: 解析
 - R: データ取り出し
 - E: 演算
 - W: データ書き戻し



メモリアクセス時間

- プロセッサの高速化にメモリシステムの高速化は追いついていない。
- スカラプロセッサ: キャッシュメモリ(小さいが高速な一時保管場所)
- ベクトルプロセッサ: バンクメモリ(多数の独立したメモリシステム)

ノイマン型計算機の実例の例



演算性能とメモリ性能のバランス

- 一定時間内の
 - Load/store のバイト量 $b[B]$
 - 演算数 $n[FLOP]$
- ⇒ $b/n [B/FLOP]$ (B/F , BF 等と略される。)

$$a[i] = b[i] * s + c[i] \Rightarrow 2\text{load}, 1\text{store} / 2\text{演算} = 8B * 3 / 2FLOP = 12BF$$

$$a[i] = b[i-2] * n_2 + b[i-1] * n_1 + b[i] * n_0$$
$$\Rightarrow 1\text{load}, 1\text{store} / 5\text{演算} = 8B * 2 / 5FLOP = 3.2BF$$

初代ESは 4BF, ES2は2.5BF
スカラプロセッサは 1BF以下

計算プロセッサ(AP)の構成

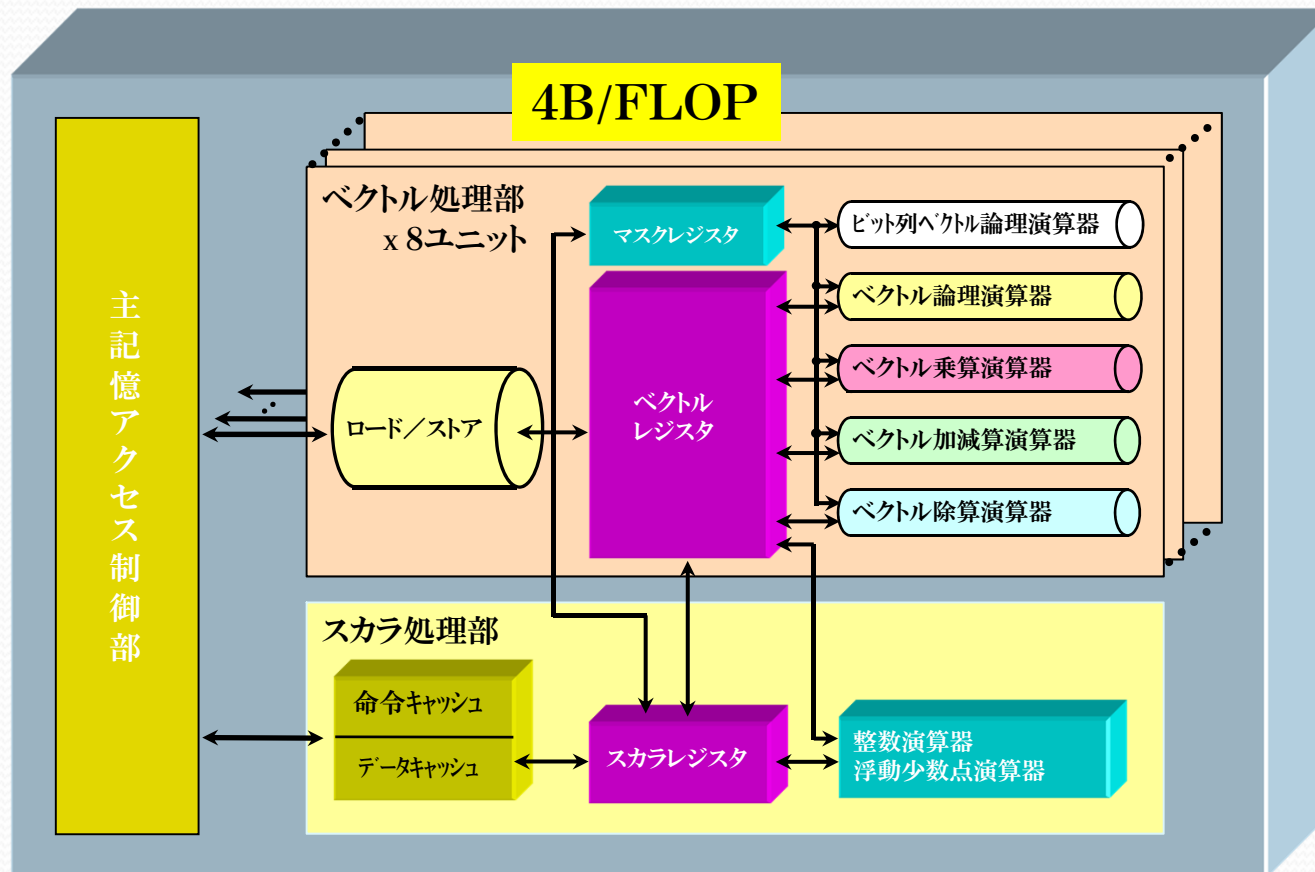
○ ベクトルユニット:8セット

- ◆ 6種のベクトルパイプライン
- ◆ 256要素のベクトルレジスタ: 72個
- ◆ 256ビットのマスクレジスタ: 17個

○ 主記憶アクセス制御部

○ スカラユニット

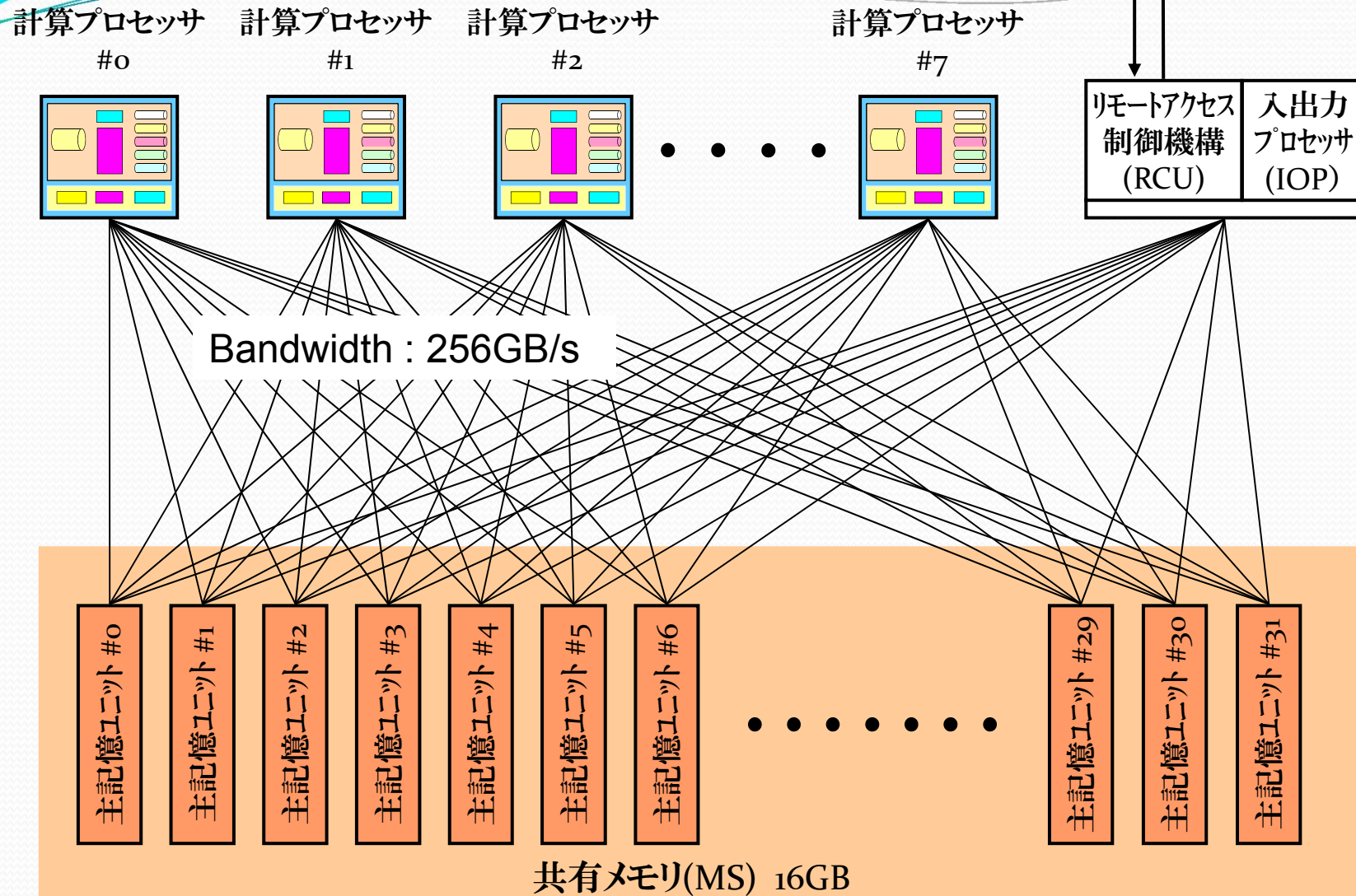
- ◆ 4-ウェイ スーパースカラ
- ◆ 64KB 命令キャッシュ
- ◆ 64KB データキャッシュ
- ◆ 128個の汎用レジスタ



1チップLSI: 8Gflops

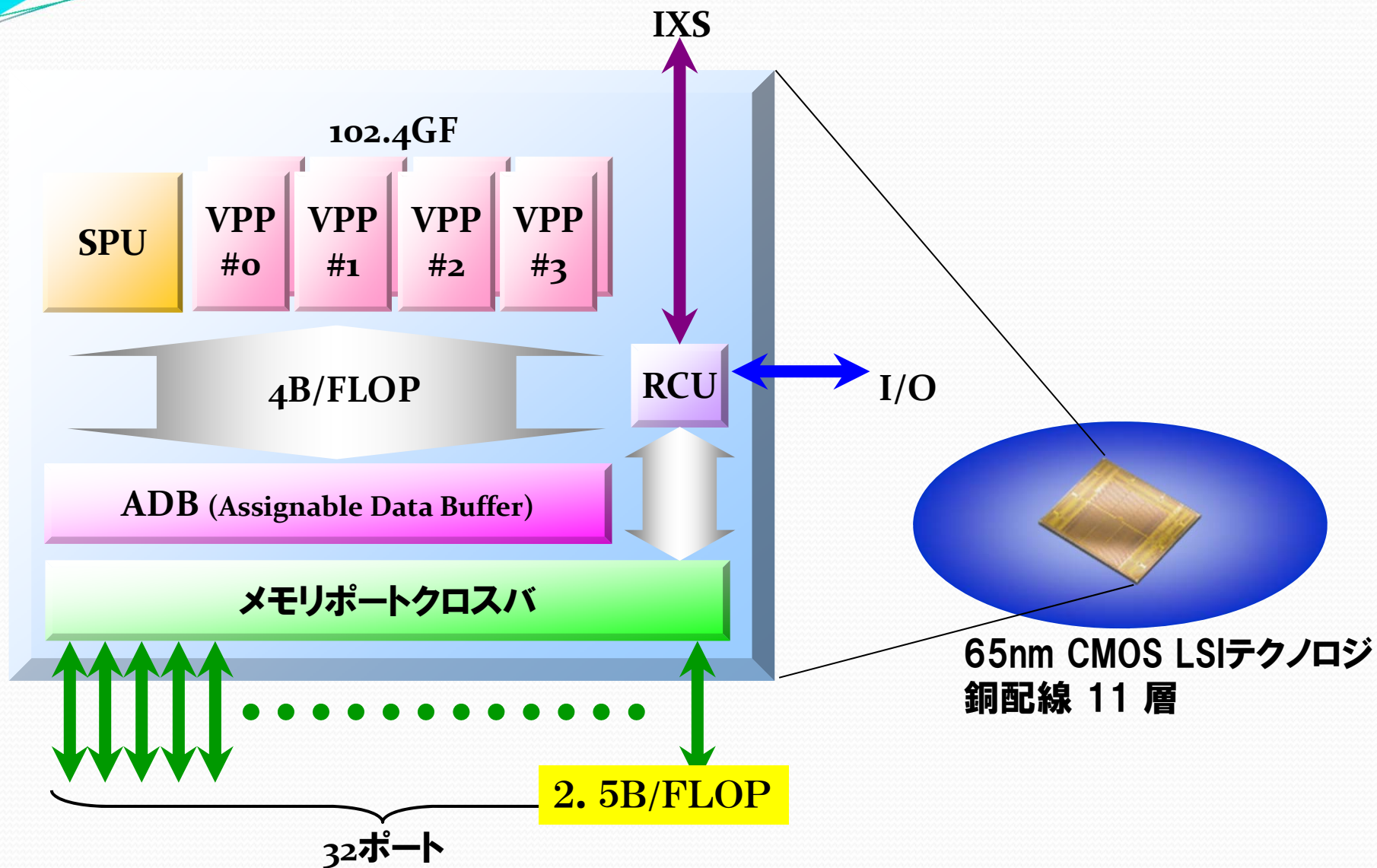
- ◆ 0.15 μ m CMOSテクノロジー + 銅配線
- ◆ 20.79mm x 20.79mm
- ◆ 5,700万トランジスタ
- ◆ 5185 ピン
- ◆ クロック周波数
500MHz
- ◆ 消費電力
140W

計算ノード(PN)の構成

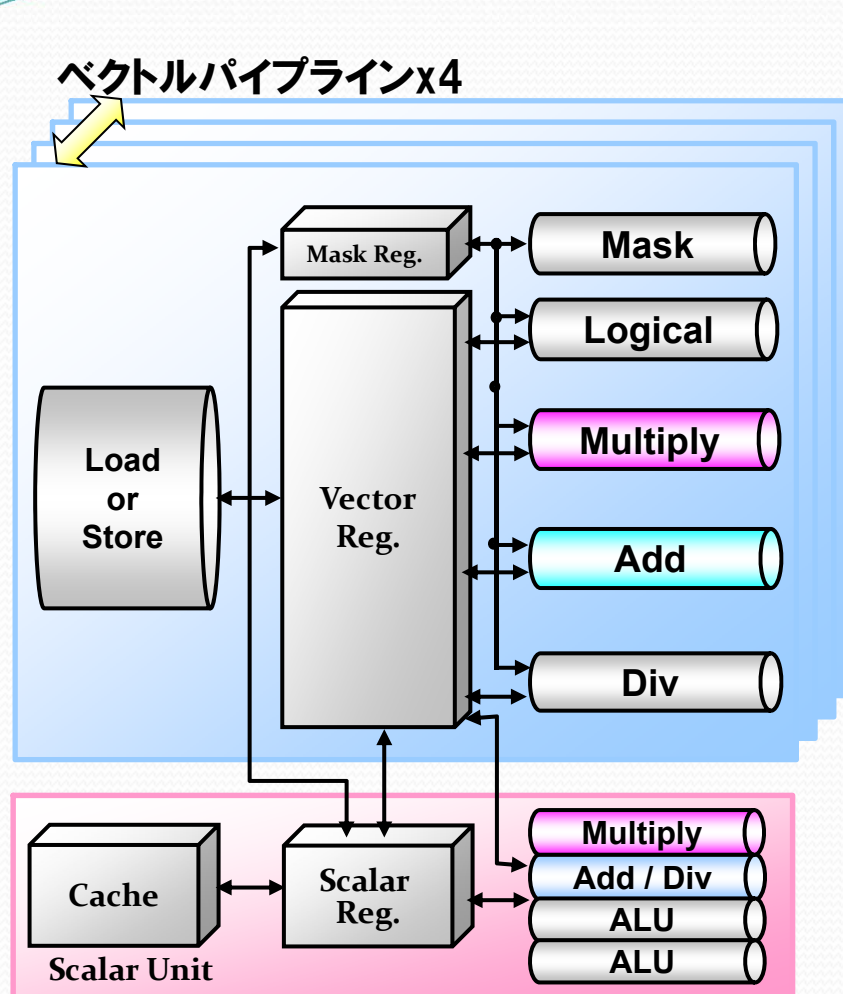


- ES向けに開発された128M bit 高速パイプラインRAM (FPLRAM)が使われている。
(バンクサイクル時間は24nsec)
- 各MMUは64バンクで構成されており、全体で2048バンクを実現している。

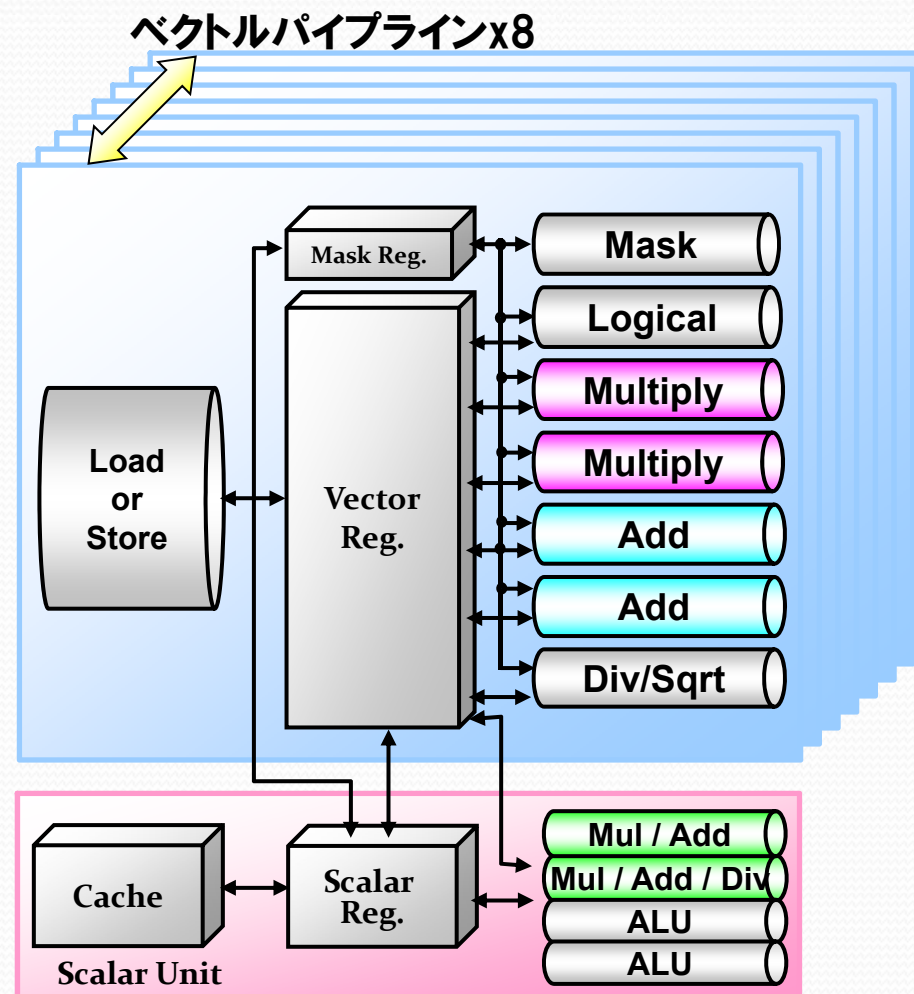
SX-9/E CPU構成（資料 NEC提供）



CPU構成（資料 NEC提供）

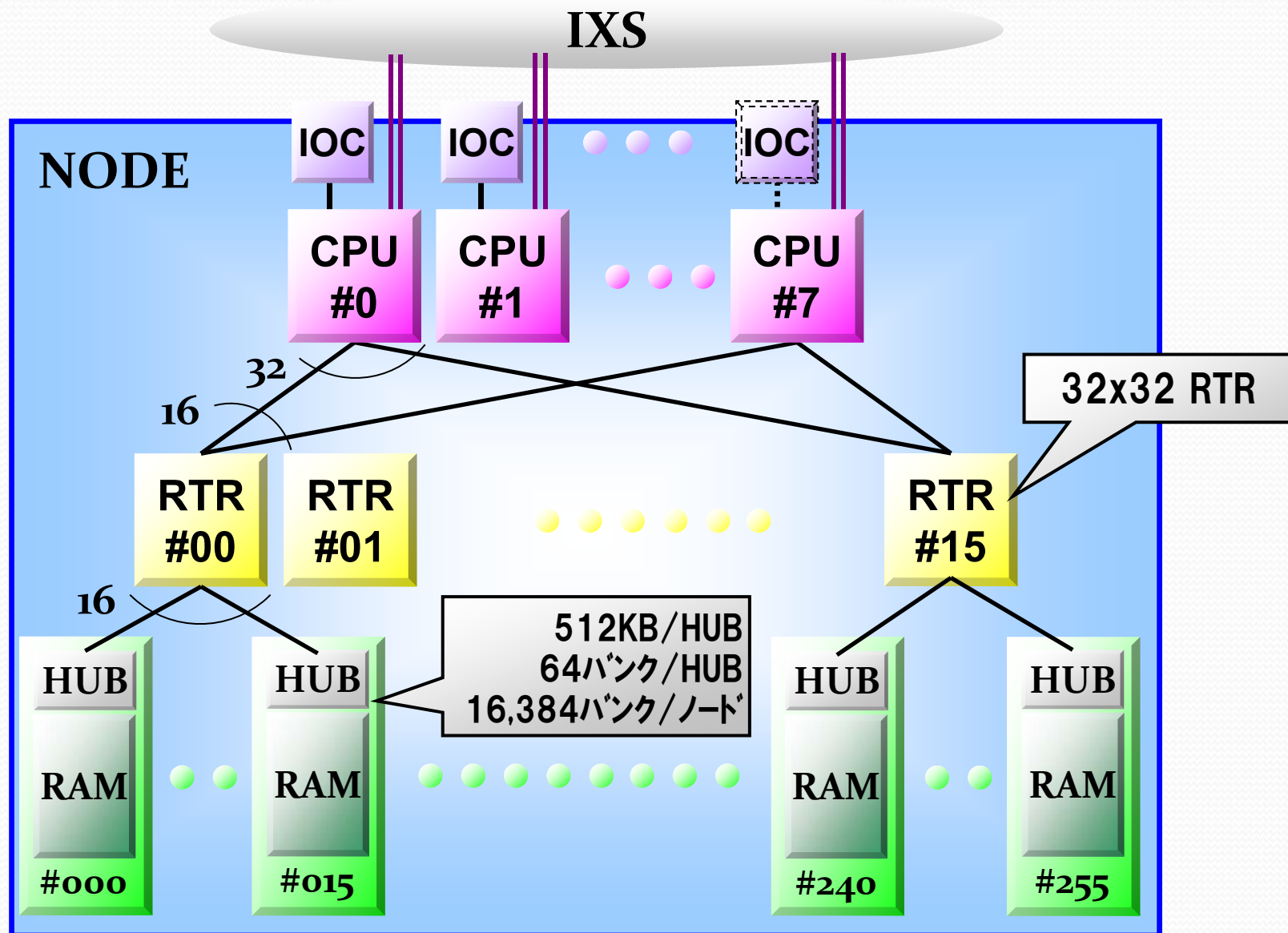


ES CPU
8GF

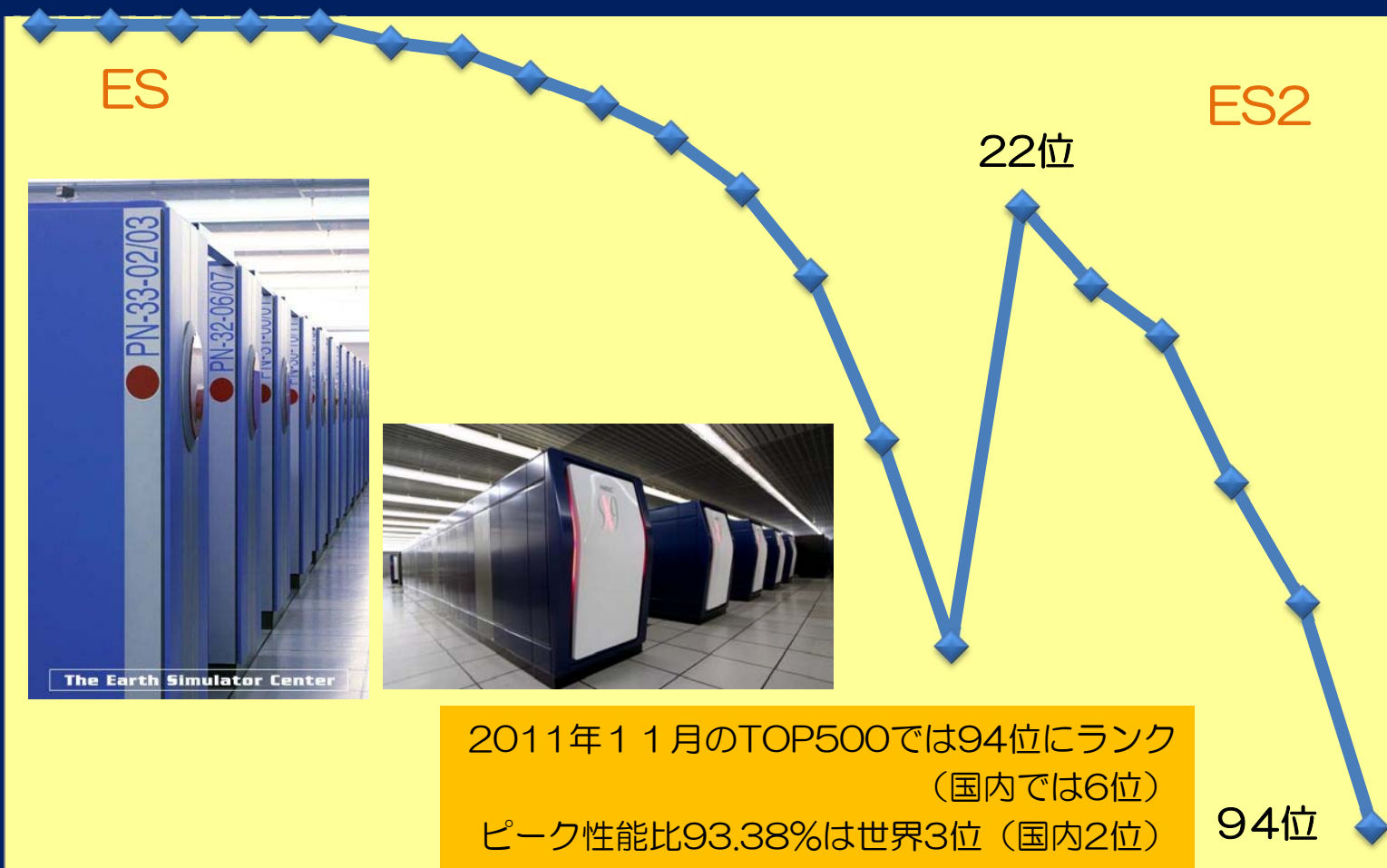


SX-9/E CPU
102.4GF

SX-9/E ノード構成 (資料 NEC提供)



地球シミュレータのTOP500ランキング



順位	2002 Jun	2002 Nov	2003 Jun	2003 Nov	2004 Jun	2004 Nov	2005 Jun	2005 Nov	2006 Jun	2006 Nov	2007 Jun	2007 Nov	2008 Jun	2008 Nov	2009 Jun	2009 Nov	2010 Jun	2010 Nov	2011 Jun	2011 Nov
	1	1	1	1	1	3	4	7	10	14	20	30	49	73	22	31	37	54	68	94

もう1つの性能指標「HPCチャレンジ」のG-FFT世界2位



G-HPL	Achieved	System	Affiliation	Submitter
1st place	2,118 Tflop/s	K computer	RIKEN AICS	Mitsuo Yokokawa
1st runner up	1,533 Tflop/s	Cray XT5	ORNL	Buddy Bland
2nd runner up	736 Tflop/s	Cray XT5	UTK	Steve Whalen
G-RandomAccess	Achieved	System	Affiliation	Submitter
1st place	121 GUPS	K computer	RIKEN AICS	Mitsuo Yokokawa
1st runner up	117 GUPS	IBM BG/P	LLNL	Tom Spelce
2nd runner up	103 GUPS	IBM BG/P	ANL	Scott Parker
G-FFT	Achieved	System	Affiliation	Submitter
1st place	34.7 Tflop/s	K computer	RIKEN AICS	Mitsuo Yokokawa
1st runner up	11.9 Tflop/s	NEC SX-9	JAMSTEC	Kenichi Itakura
2nd runner up	10.7 Tflop/s	Cray XT5	ORNL	Buddy Bland
EP-STREAM-Triad (system)	Achieved	System	Affiliation	Submitter
1st place	812 TB/s	K computer	RIKEN AICS	Mitsuo Yokokawa
1st runner up	398 TB/s	Cray XT5	ORNL	Buddy Bland
2nd runner up	267 TB/s	IBM BG/P	LLNL	Tom Spelce

HPC Challenge Awards

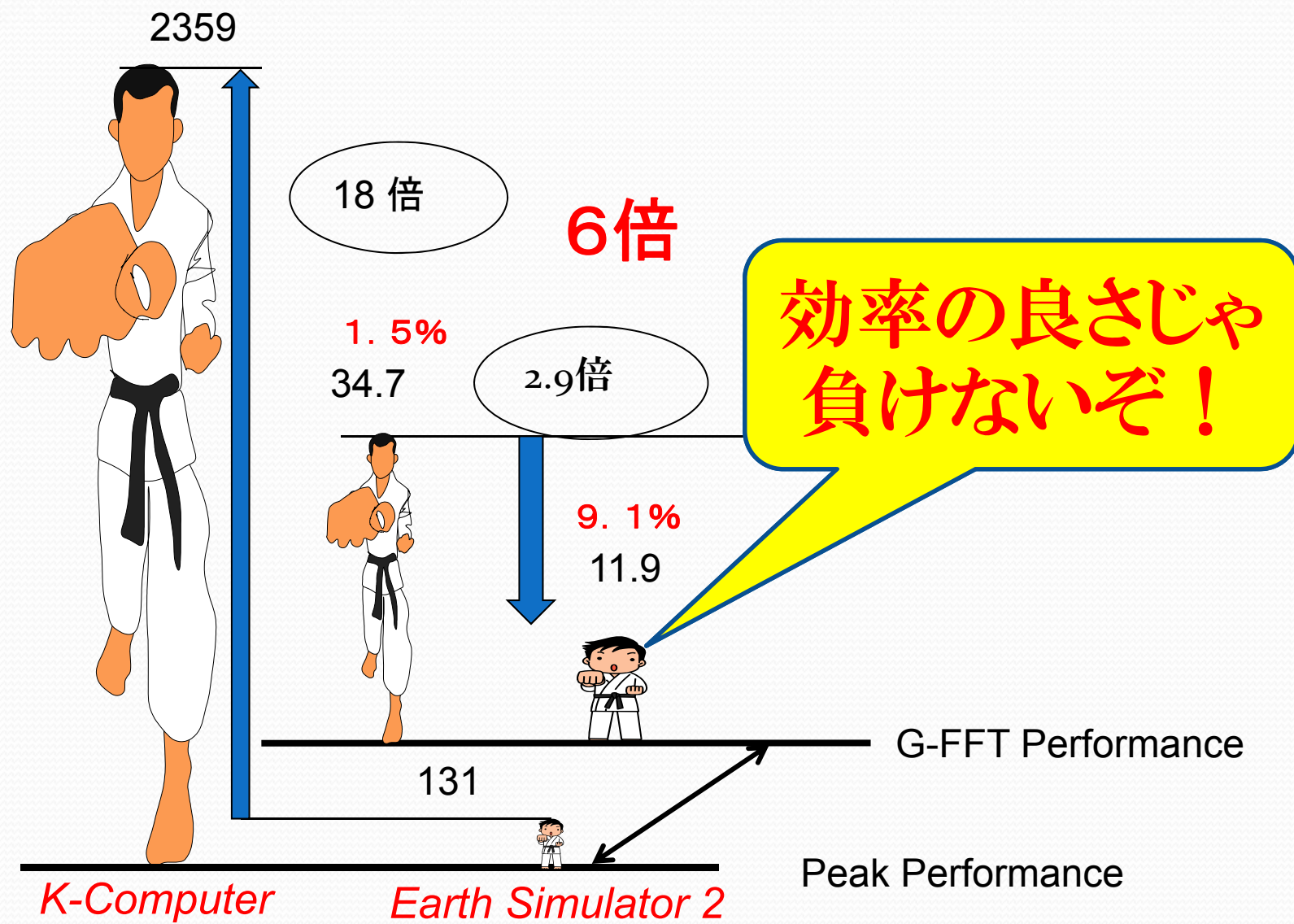
実際の科学技術計算でよく使われる計算パターンから抽出した4つの指標を使って多面的な観点から性能を評価するもの。

<http://www.hpcchallenge.org/>

G-FFT(高速フーリエ変換の総合性能)で2位を受賞。FFTは気象予測や気候変動予測などの球面上の流れを解析する分野で使われる重要な手法。

システム	TOP500(Linpack)		使用コア数	G-FFT		ピーク性能比 K下段:G-FFT使用コア比	使用コア数
JAMSTEC ES2 (SX-9)	94位	122 TF	1,280個	2位	11.9 TF	9.1 %	1,280個
RIKEN AICS (K Computer)	1位	10,510 TF	705,024個	1位	34.7 TF	1.47 %	147,456個
比較	K が86倍高速		Kは550倍の コア数	Kが2.9倍高速		ESが6.2倍高効率	ESは1/115の コア数

HPC Challenge Awards 2011



まとめ

- 地球シミュレータ
 - 初代:2002年から2009年まで 40TFで世界1位を2年半
 - ES2:2009年から2015年まで ピーク性能3.2倍(実行性能で2倍程度)
 - 90%以上の稼働実績で運用
- コンピュータアーキテクチャと高速化
 - スパコン:マルチボードプロセッサから1チップ化
 - 高速化は並列化:プロセッサ内、ノード内、ネットワーク接続
 - プロセッサ性能とメモリ性能のバランス(BF)が大事
 - 地球シミュレータの高いBF値⇒高い実効性能を維持