

戦略的高性能計算システム開発に関するワークショップ

**5年後(2015年)にセンター運用可能な  
高並列計算機システム**

2010年 8月 2日

株式会社日立製作所 中央研究所  
吉村 地尋



# 自己紹介

## ●名前と所属

- 吉村 地尋 (ヨシムラ チヒロ)
- 2007年 株式会社日立製作所 入社
- 以来, 中央研究所において計算機アーキテクチャの研究に従事。

## ●研究テーマ

- スーパーコンピュータや基幹サーバ向けのプロセッサ, I/O処理技術の研究



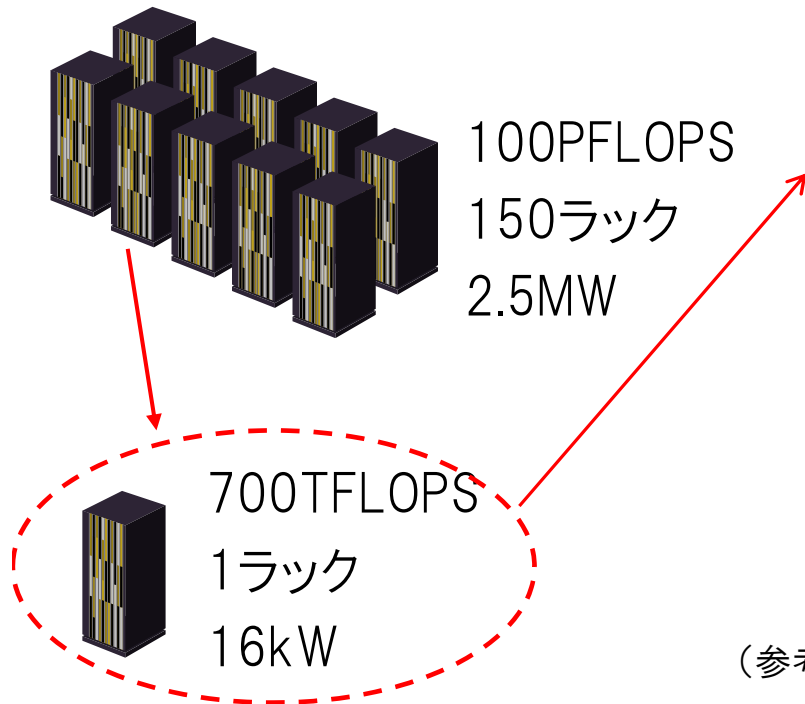
## 本発表の目的:5年後(2015年)

- 5年後(2015年)に、センター運用可能な高並列計算機システムは、どのようなシステム仕様が考えられ、そのために、今後どういう研究開発をしていくべきなのか？
- センター運用で求められる制約を、以下のように想定して検討する。
  - 設置面積(大学センターなので最大でも150ラック程度と想定)
    - 次世代スパコン「京」は約800ラック, 東大T2Kは60ラック程度？
  - 消費電力(空調含めて5MW, IT装置に2.5MW程度と想定)
    - 今の東大機は約0.8MW(TOP500登録値)
  - 汎用的に利用可能なアーキテクチャ



# 目標性能

- 2015年にピーク性能50PFLOPS～100PFLOPSが必要。
- 100PFLOPSを150ラックで実現するためには0.7PFLOPS/ラックが必要  
(参考:2020年米国ExaFLOPS国プロは1PFLOPS/ラック目標)。



#	ノード構成	単体性能	ノード数
1	コモディティCPU	8FLOP/cyc * 32core * 3GHz*4S=3TFLOPS	233ノード/ラック (34950ノード)
2	メニーコア混在 コモディティCPU	8FLOP/cyc * 4core + 32FLOP/cyc * 16core * 2.3GHz*4S=5TFLOPS	140ノード/ラック (21000ノード)
3	コモディティCPU + GPU併用	3TFLOPS(CPU) + 5TFLOPS(GPU) = 8TFLOPS	88ノード/ラック (13200ノード)
4	ワンチップノード (FPU強化 組込みCPU)	16FLOPS/cyc * 32core * 2GHz = 1TFLOPS	700ノード/ラック (105000ノード)

(参考) T2K東大機(140TFLOPS)は147GFLOPS/ノード, 16ノード/ラック  
(全60ラック程度)ぐらい?

ラック内のノード数を考えると, どの構成でもノード数が多すぎる



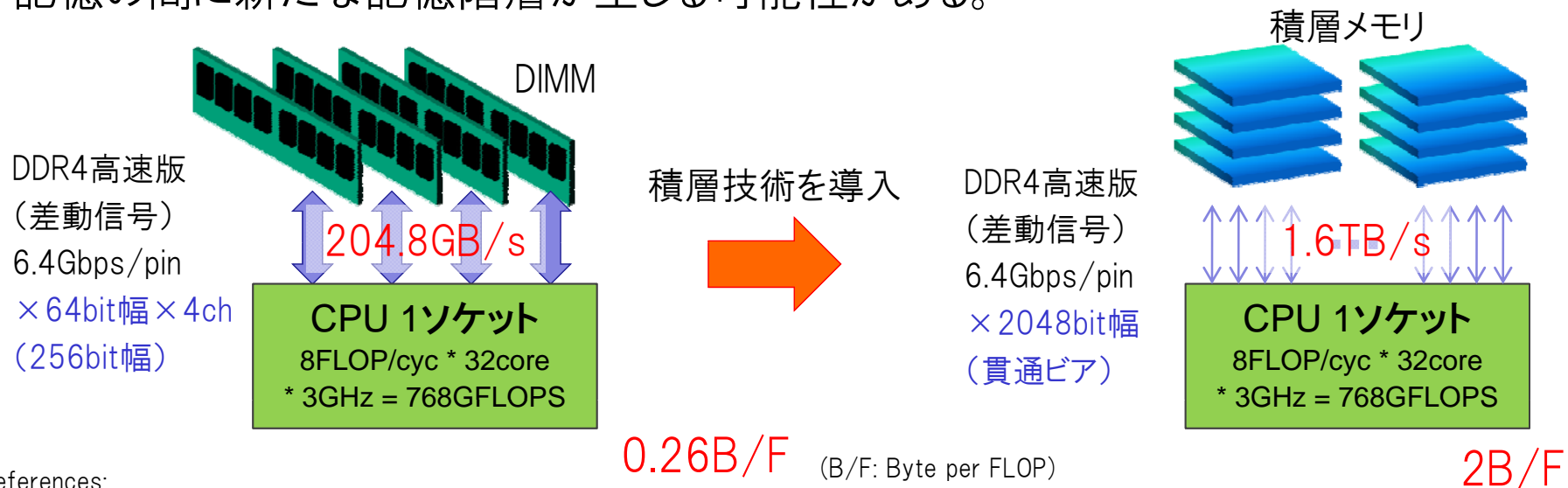
## 課題

- センター運用可能な100PFLOPSスパコンは、主にラック数制約でコモディティCPUのみでは困難そうに見える。性能を落とすか、ラック数を増やすか、非コモディティ路線での解決を図る必要がある。
- ノード数が莫大になることからPower Wall問題がより一層顕在化する。既に消費電力削減の取り組みはあるが、さらなる消費電力削減が必要。
- 演算性能を高めることで、メモリ、ネットワーク、ストレージなどのI/O系との性能乖離がより一層顕在化する。性能乖離を埋めるために、デバイスのレイヤーから見直しが必要。  
(アーキテクチャとしては、デバイス使いこなし技術の深化要)



# 積層技術によるメモリB/F改善

- DRAMのチップ単体容量トレンドが鈍化しているが、既に積層DRAMで容量増加トレンドを維持する方向にある(メモリーメモリ積層)。
- 演算性能とメモリ/F性能の乖離(メモリB/Fの低下)も進んでいるが、2015年にはプロセスサ-メモリ積層で対処する。
- 演算性能に見合ったメモリ全量を積層化できるかは微妙。オンチップキャッシュと主記憶の間に新たな記憶階層が生じる可能性がある。



References:

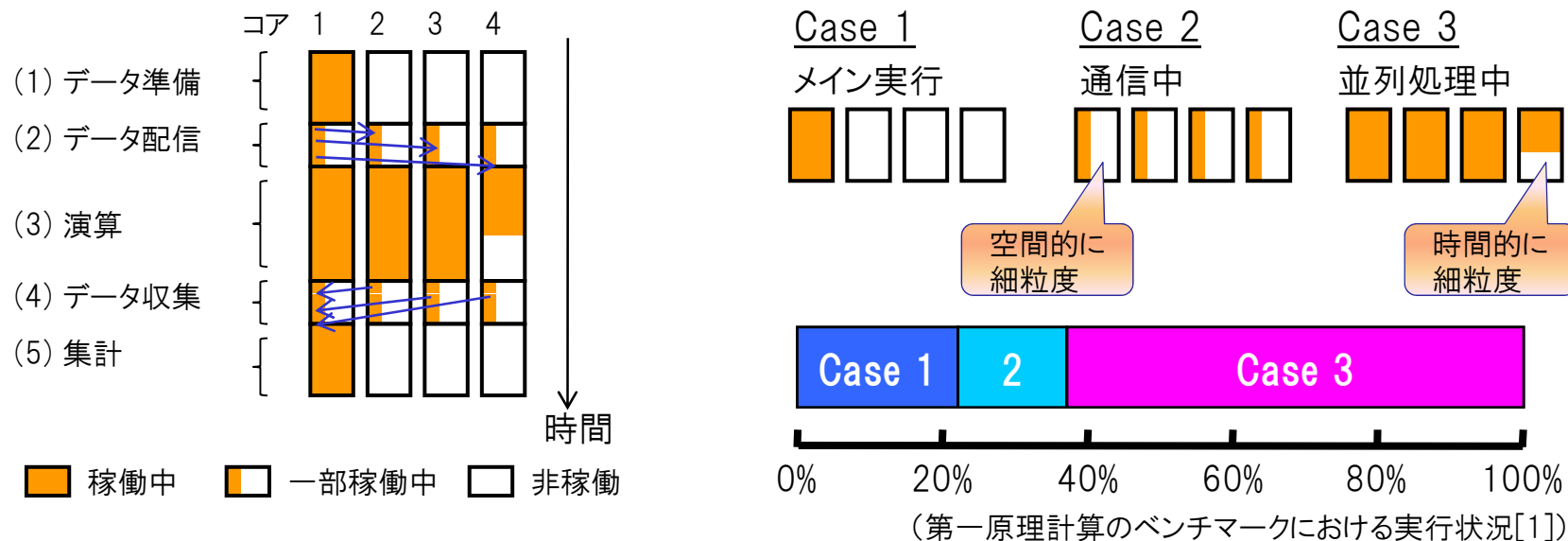
[1] 安達隆郎, TSVを用いた三次元積層DRAM技術の展望, ドリームチップ平成21年度研究成果報告会, 2010.

[2] K. Ono et al., 1-Tbyte/s 1-Gbit DRAM Architecture with Micro-pipelined 16-DRAM Cores, 8-ns Cycle Array and 16-Gbit/s 3D Interconnect for High Throughput Computing, Symposium on VLSI Circuits Dig. Tech. Papers. pp.187-188, 2010.



# 動的電力制御による省電力化

- 低周波数化，低電圧化のような静的な省電力化技術に加えて，アプリ実行状況に応じた**時間的・空間的に細粒度な動的電力制御**を導入。
- 高並列化に伴い，実効効率が低下する傾向にあるので，電力制御で省電力化可能なシチュエーションは増える。



References:

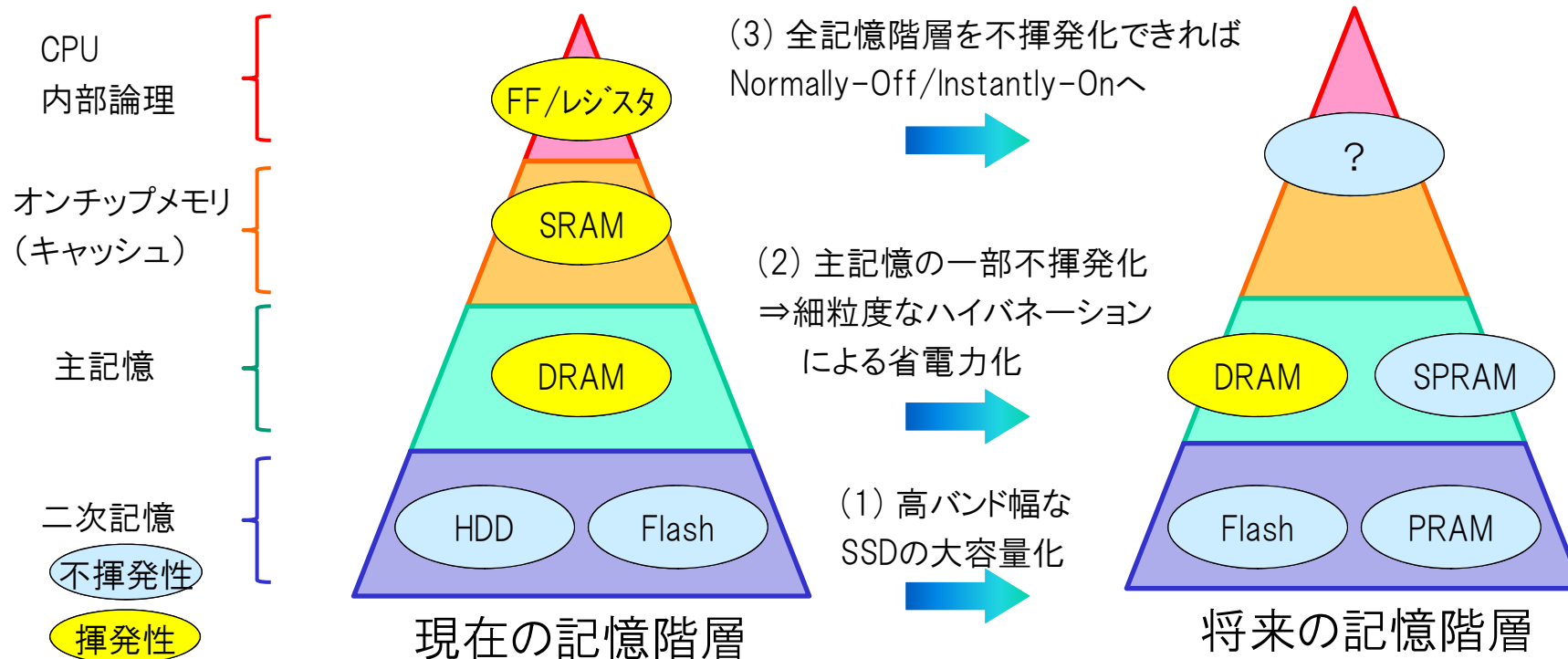
[1] H. Aoki et al., A Powerful yet Ecological Parallel Processing System using Execution-based Adaptive Power-down Control and Compact Quadruple-Precision Assist FPU's, Symposium on VLSI Circuits Dig. Tech. Papers. pp.186-187, 2008.

[2] 中村宏, Power Wall問題へのブレークスルーを目指して, 情報処理, Vol.51, No.7, pp. 855-860, 2010.



# 不揮発メモリによるI/O強化・省電力化

- 2015年～2020年に向けて、PRAM(相変化メモリ)やSPRAM(スピンRAM)など、新概念に基づく不揮発メモリの実用化が始まる。
- PRAMはFlashより記録密度で優位なため、SSDの大容量化につながる(I/O強化)。SPRAMは記録密度では劣るが、DRAMと同等の速度で、書換え回数無制限であるため、主記憶の一部不揮発化が実現できる(ノードの電源制御による省電力化)。







# 10年後(2020年)に向けて

- 2020年の1ExaFLOPS級NLS(National Leading System)は、1000ラック構成で、1PFLOPS/ラック(米国UHPCの構成案)。
- 1PFLOPS/ラック×1000ラックでも、ラック内の実装密度が問題になる。また、ネットワークB/Fを0.01～0.1B/F確保しようとする、ラック間のネットワークが10TB/s～100TB/sにもなる。  
⇒ネットワークの配線物量がネックになる。
- 100PFLOPSのセンター運用可能マシンの実現は、1ExaFLOPSへの挑戦に向けた第一歩。