

中間報告

2013.03.27

東京大学
石川裕



Feasibility Study on
Advanced and Efficient Latency Core-based
Architecture for Future HPCI R&D

2013/3/27

1

概念設計中のシステムの概要

▶ デザインコンセプト

- 幅広いアプリケーションユーザのニーズに合致した問題サイズおよび要求実行時間内で終了
- 「京」のシステムアーキテクチャを踏襲し、「京」で開発されたアルゴリズムやチューニング技術などを継承, 発展可能なアーキテクチャ
- コモディティ(PCクラスタ等)と技術(アプリケーション, システムソフト)の相互移植性が高いこと

▶ 取り組みと特徴

- 「京」およびFX10を起点に, 資産の継承とエクサへ向けての強化・革新
- 単体高B/F値でなく, システム全体で高演算性能&高メモリバンド幅を提供
- アプリケーション&システムソフトウェア&アーキテクチャのコデザイン
 - 設計&予測&評価&チューニング&再設計ループ
 - 要求B/F値高いアプリから低いアプリを対象

「京」アプリと社会的・科学的課題の解決

「京」のシステム
アーキテクチャを踏襲

汎用マルチコアCPU
・高いスカラー性能
・アプリ効率向上
スケーラビリティ
・~数十万ノード

ターゲットシステム
(評価・改善中)

大規模・詳細化

コモディティとも協調

汎用性維持し高速化

スケーラビリティと
堅牢性

省電力性

「京」を進化させる要素技術開発

アプリケーション

システムソフトウェア

高信頼・高性能CPU

高並列インターコネクト

ストレージシステム

「京」の資産



2013/3/27

2

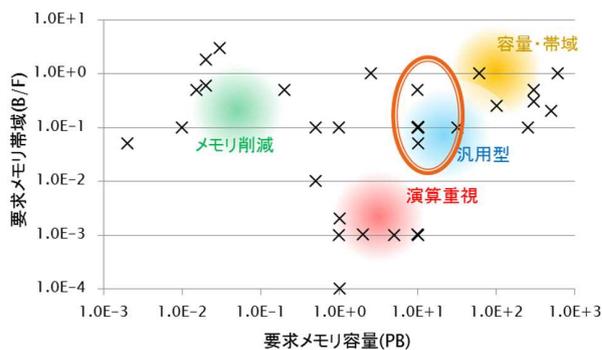
科学的課題と解決法

アプリ名	科学的課題	問題サイズ	要求実行時間	予測実行時間
ALPS	従来実行できなかった現実的なナノ磁性体(1辺~10nm)の極低温における量子相転移現象や、量子液体状態などの新奇量子相の発現機構が明らかになる。将来のスイッチング素子実現に向けての基礎研究段階	ナノ磁性体(1辺~10nm)の極低温におけるシミュレーション(L=262144、T=0.00083)	1日	2時間以内
RSDFT	従来困難である、Siナノワイヤトランジスタにおける実物サイズ(1万~10万原子)でのチャンネル特性の予測を可能とする	30万原子レベルのシミュレーション	数日	予測中
NICAM	現在のスパコンは非現実的な実行時間のために行われていない、高解像度での50年を越える気候シミュレーションを可能とする	問題レベル、glevel=14, vlayer=100(水平400m、鉛直250m相当)の50年間分のシミュレーション	約1か月	16日以内
COCO	(日本)沿岸の小規模な湾の流れを(太平洋スケールの)外洋の海流とリンクさせたシミュレーションシステムをリアルタイムで動かす	40,000 x 40,000 x 1,000の問題規模(格子数)について、10年間分のシミュレーション	約1か月	21日以内

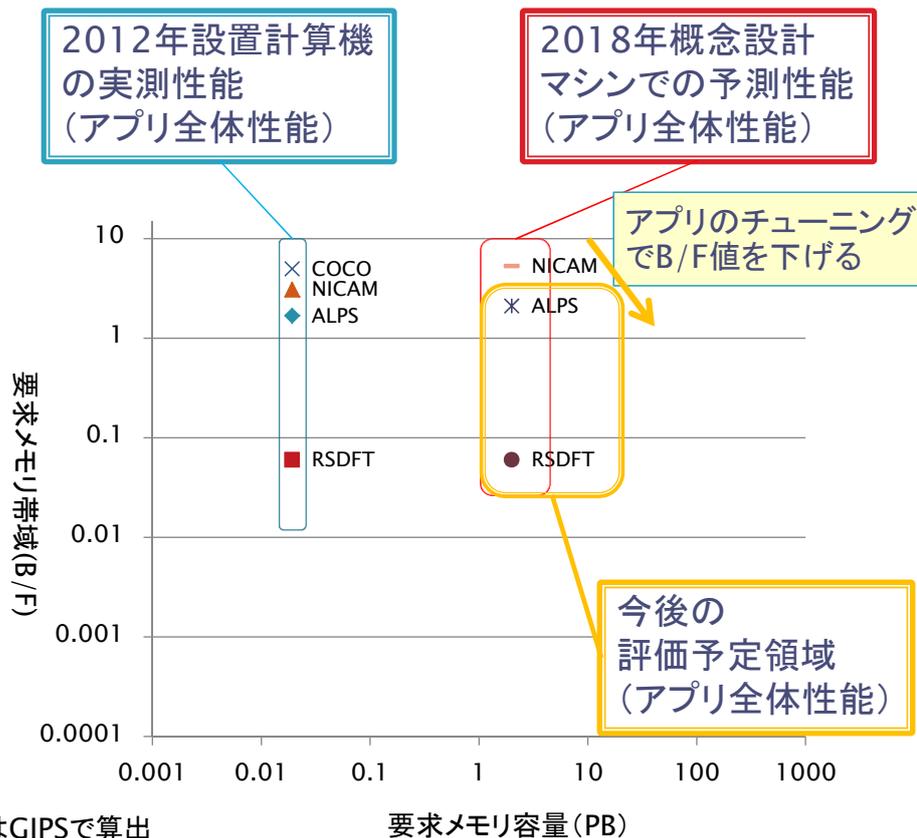


DO NOT DISTRIBUTE

東大FSのアプリ要求性能と方向性



出典: 今後のHPCI技術開発に関する報告書p.13から



DO NOT DISTRIBUTE

東大FSアプリの性能と方向性

	●FX10プロフィール結果 東大FX10全系実行:4GB/node * 4800 nodes = 0.0192PB		●予測性能 (今後の検討により変更有)	
アプリ名	全体B/F	使用メモリ量[PB]	全体B/F	使用メモリ量[PB]
ALPS*	1.68	0.0192	2.10	2
RSDFT	0.06	0.0192	0.06	2
NICAM	3.08	0.0192	5.33	2
COCO	4.98	0.0192	予測中	2

※ALPSはGIPSで算出

$$\text{全体B/F} = \frac{\text{アプリケーション実行時のメモリバンド幅}}{\text{アプリケーション実行時のFlops値}}$$



DO NOT DISTRIBUTE

5

要素技術の実現可能性

要素技術	実現可能性	開発期間	展開可能性
10nm半導体プロセス技術	中	※	他社汎用技術であるが、次期スパコンやサーバ全般に応用可能
高速積層メモリ技術	中		
高速光インタフェース技術	高		
高信頼・高性能CPU技術	高	H26-H29年度	サーバ全般
高並列インターコネクト技術	高	H26-H29年度	クラウドサーバ、ビッグデータ処理
システムソフトウェア (OSカーネル、通信ライブラリ、並列I/O、階層型ファイルシステム、バッチジョブシステム、デバッグ・チューニング、可視化システム)	高	H26-H30	コモディティPCクラスタデータセンター ビッグデータ処理

- ▶ ※汎用技術は、以下の入手期限をリスク要因とし、実現可能性を推定
 - H26(2014)年度末:設計パラメータ
 - H27(2015)年度末:設計検証用サンプル
 - H28(2016)年度末:評価試験用サンプル
 - H29(2017)年度末:量産品
- ▶ ハードウェア要素技術:H26~H29年度で開発可能の見込み
- ▶ ソフトウェア要素技術:全系設置後、調整および性能チューニングが必要



2013/3/27

6

システムの消費電力、耐故障性、信頼性、 設置面積

▶ 性能と消費電力

- 200PF～300PF
- 20MW～30MW

▶ 信頼性 & 耐故障性

- CPUの高信頼性/高可用性機能、インターコネクタの高可用性機能により耐故障性を確保
- ノードの構成は「京」に類似し、同程度のノード信頼性確保が可能

▶ 目標MTBF

- 「京」と同程度
- システムのノード数のオーダを「京」と同程度に抑えることにより実現可能

▶ 設置面積

- ストレージを除き「京」の1/2程度
- 高速積層メモリの採用・全面水冷化によりノード実装密度を「京」の2倍程度に向上

