

# 東大FSアプリWG 進捗報告

片桐 孝洋, 大島 聡史, 中島 研吾(東大)

米村 崇, 熊洞 宏樹, 樋口 清隆, 橋本 昌人, 高山 恒一(日立情報・通信システム社),  
藤堂 眞治, 岩田 潤一, 内田 和之, 佐藤正樹, 羽角博康(東大),  
黒木聖夫(海洋研究開発機構)

【将来HPCI調査研究「アプリ分野」第9回全体ミーティング】

日時: 9月17日(火) 13:00-16:45

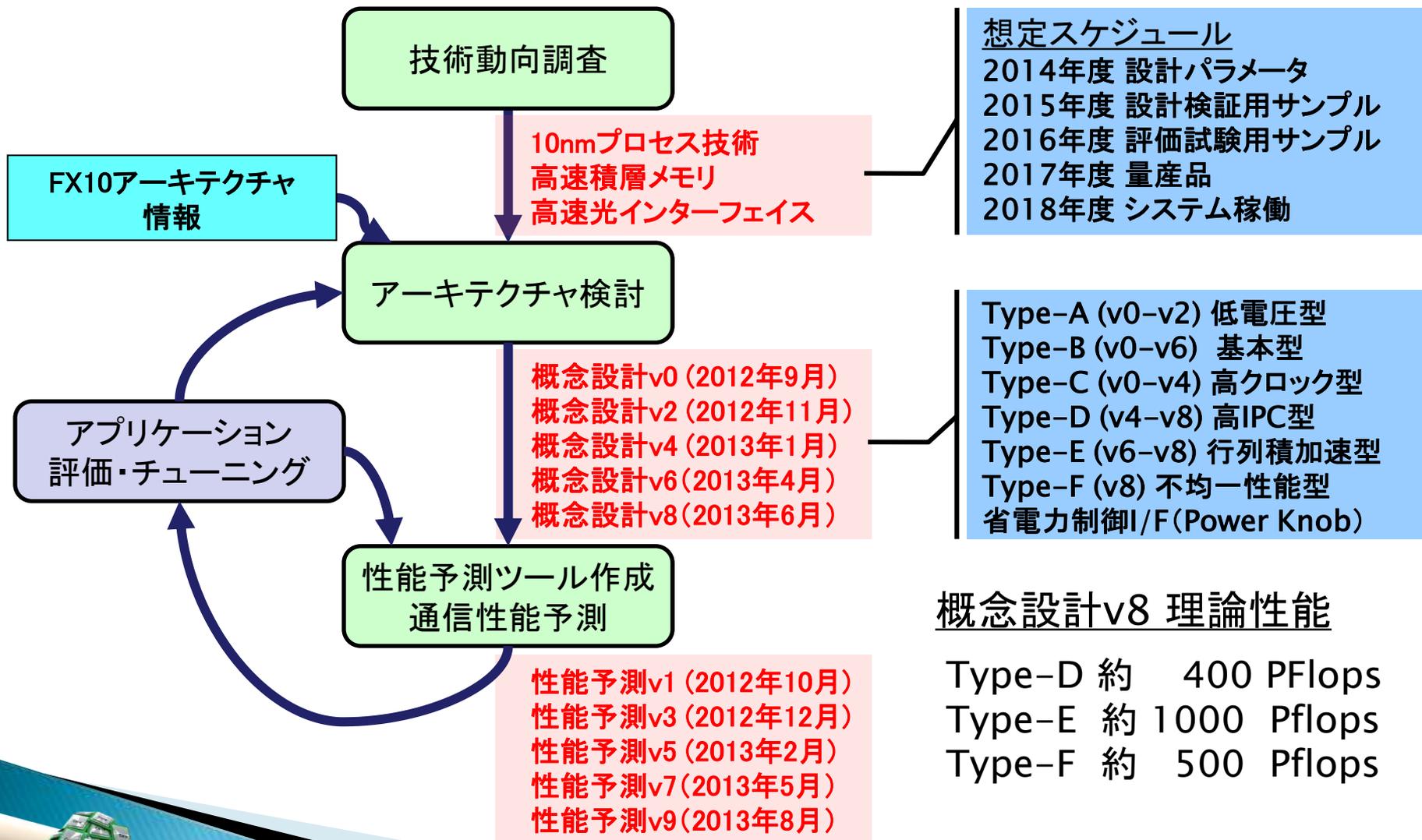
場所: TKP大手町ビジネスセンター 6階ホール6A

13:40-14:40 システムFSアップデート報告(東北大、東大、筑波大)



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

# アーキテクチャ概念設計



# スケジュール確認

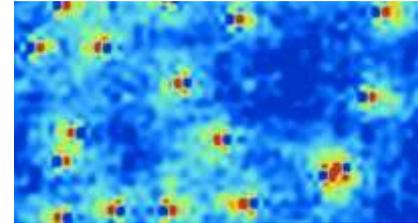


Many many  
More to come

# ターゲットアプリケーション群

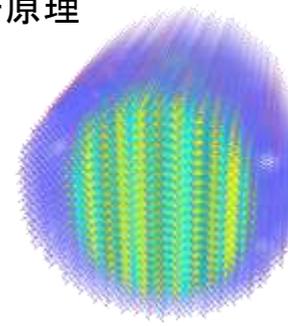
## ALPS/looper

- 新機能を持った強相関・磁性材料の物性予測・解明。虚時間経路積分にもとづく量子モンテカルロ法と厳密対角化
- **総メモリ**: 10~100PB
- **整数演算**、低レイテンシ、高次元のネットワーク
- **利用シナリオ**: 1ジョブ当たり24時間、生成ファイル: 10GB. 同時実行1000ジョブ、合計生成ファイル: 10TB.



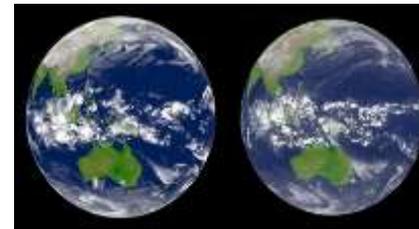
## RSDFT

- Siナノワイヤ等、次世代デバイスの根幹材料の量子力学的第一原理シミュレーション。実空間差分法
- **総メモリ**: 1PB
- **演算性能**: 1EFLOPS (B/F = 0.1以上)
- **利用シナリオ**: 1ジョブ当たり10時間、生成ファイル: 500TB. 同時実行10ジョブ、合計生成ファイル5 PB.



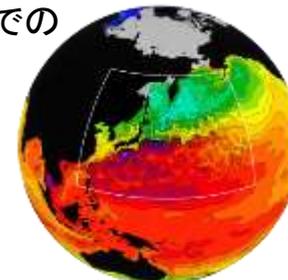
## NICAM

- 長期天気予報の実現、温暖化時の台風・豪雨等の予測
- 正20面体分割格子非静力学大気モデル。水平格子数kmで全球を覆い、積雲群の挙動までを直接シミュレーション
- **総メモリ**: 1PB、**メモリ帯域**: 300 PB/sec
- **演算性能**: 100 PFLOPS (B/F = 3)
- **利用シナリオ**: 1ジョブ当たり240時間、生成ファイル: 8PB. 同時実行10ジョブ、合計生成ファイル: 80 PB.



## COCO

- 海況変動予測、水産環境予測
- 外洋から沿岸域までの海洋現象を高精度に再現し、気候変動下での海洋変動を詳細にシミュレーション
- **総メモリ**: 320 TB、**メモリ帯域**: 150 PB/sec
- **演算性能**: 50 PFLOPS (B/F = 3)
- **利用シナリオ**: 1ジョブ当たり720時間、生成ファイル: 10TB. 同時実行100ジョブ、合計生成ファイル: 1 PB.



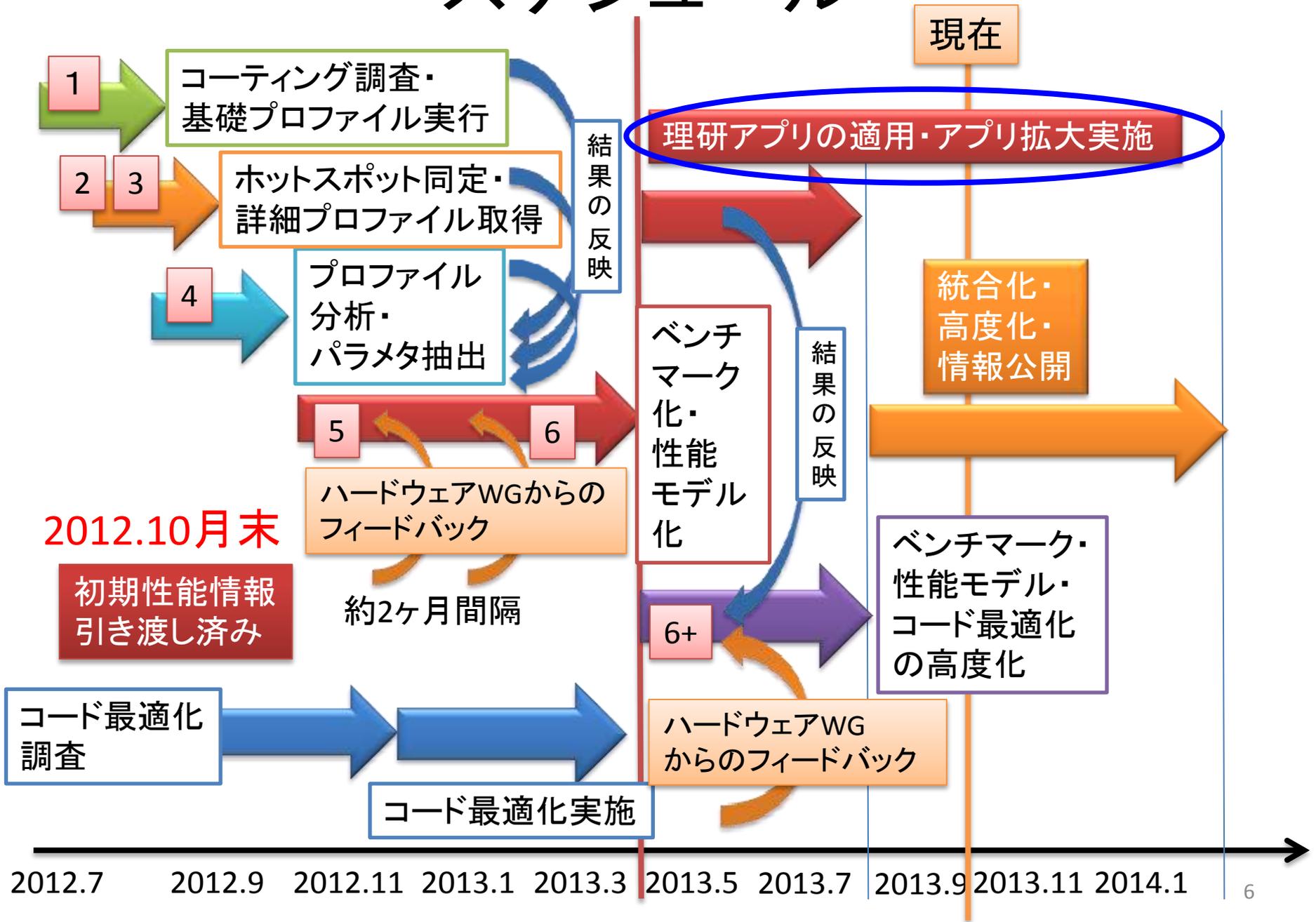
**利用シナリオ**  
**アンサンブル型**  
全系の1/10~  
1/100資源を用  
いた1ジョブを、  
複数同時実行  
することで、全  
資源を使い切  
る形態。

要求性能は  
「計算科学  
ロードマップ白  
書」(2012年3月)  
の見積値からの  
抜粋、  
および、  
開発者による  
新見積値である

# 性能モデル化手法

- 1. ホットスポット同定: 富士通社の**基本プロファイラで複数の**ホットスポット**(ループレベル)を同定、全体性能の予測をホットスポットのみで行う
  - ホットスポットの部品化
  - 数理レベル(支配方程式、離散化方法)の処理ブロックとの対応を検討
- 2. カーネル分離:**(目視により)計算部分、通信部分、I/O部分の分離
  - 計算部分: 演算カーネル
  - 通信部分: 通信カーネル
  - I/O部分: I/Oカーネル
- 3. 通信パターン確認**
- 4. 詳細プロファイルと分析: 富士通社の**詳細プロファイラを用い、ホットスポットごとにハードウェア性能情報(=性能パラメタ)を取得し分析
  - 演算カーネルの 演算効率/命令発行量/キャッシュ利用効率 など
  - 通信カーネルの 通信回数/量/通信待ち時間 など
  - I/Oカーネルの データ読み書き 量/頻度 など
- 5. ベンチマーク化:**ホットスポットのみで動作するようにコードを再構成
  - マシン特化の書き方、および、汎用的な書き方、の2種を区別
  - 演算カーネル、通信カーネル、I/Oカーネルの分類
- 6. 詳細モデル化と予測:**ハードウェア因子を用いた数式による実行時間を近似。
  - 富士通社の性能予測ツールにより、概念設計マシンの実行時間を予測

# スケジュール



# 今後の予定



*Many many  
More to come*

# H25年度のアプリ評価

## ● H25年度のアプリ評価

ALPS、RSDFT、NICAM、COCOの性能予測ツールVer.9以降による演算時間予測、通信性能予測Ver.9以降による通信時間予測、これらを基にした実行時間推定を完了

## ● H25年度追加アプリの評価 (H25.9.11現在)

アプリケーション	実行シナリオ確定	演算時間推定	通信時間推定	実行時間推定
QCD	◎	◎	◎	◎
Modylas	◎	◎	○(分析済)	実施中
流体	◎	◎	○(分析済)	実施中
NGS Analyzer	アプリFSで準備中			



Many many  
More to come

# H25年度アプリ評価工程表

実施項目	～H25.10月末	～H25.12月末	～H26.2月末	～H26.3月末
アプリチューニング	新規: QCD、流体、Modylas	NGS Analyzer?	1月末	最終評価: 全アプリ
	高度化: ALPS、COCO、NICAM	高度化: QCD、流体、Modylas		
性能予測 (実行時間推定)	RSDFT(9月終了)、QCD	流体、Modylas	NGS Analyzer?	最終評価: 全アプリ
詳細シミュレータ	NICAM、COCO、RSDFT	QCD、流体、Modylas	NGS Analyzer? 最終評価	
異機種評価	ALPS、RSDFT、NICAM、COCO、QCD (Intel系、Power系、Xeon Phi)	12月末	最終評価	
その他	RSDFT通信回避アルゴリズム評価	11月末	最終評価	

# 詳細シミュレータ用カーネルベンチ作成

- ▶ 詳細シミュレータで評価するため、カーネルのみで動作するプログラム(カーネルベンチ)を作成
- ▶ ALPSはC++なので対象外
- ▶ RSDFTはdgemmがカーネルのため、別途評価可能な理由からカーネルベンチ作成は除外
- ▶ NICAM、COCOは作成済み
- ▶ QCD、流体、Modylasは作成中



Many many  
More to come

# コード最適化と異機種評価

## ▶ 新規4アプリのコードチューニングを実施

- QCD: 前回報告済み
  - Clover部のSMP化
    - OMPプロセスの演算負荷均等化
    - FX10の1ノードで、BiCG: 1.11倍、Clover: 2.53倍
  - その他も解析中、実施予定

## ▶ 異機種環境での性能評価

- FX10上の性能が妥当であるか検証することが目的
  - Intel系、Power系で性能評価
  - **メニーコア(Intel MIC (Xeon Phi))**で性能評価
  - 上記の異機種環境でのコードチューニングも一部実施予定



# アルゴリズムの再検討(1/2)

## ▶ RSDFTの通信

- 固有値ソルバ部(Eigen) と直交化部(理研版ブロック化MGS)
- 通信時間の内訳予想:
  - 固有値ソルバ部: 7割、直交化部: 3割
- 現在の直交化部は、通信回数／量を減らせる代わりに、メモリ量が多い方式が実装されている
  - MPI\_AllgatherVで分散した固有ベクトルを集める
  - ⇒ エクサ環境ではメモリ制約から、現在の実装では動かない可能性が高い
  - ⇒ メモリ量を減らす実装に変更すると、通信回数が増加し、通信時間も増大していくと予想



# アルゴリズムの再検討(2/2)

- ▶ 通信回数が少ない新アルゴリズムを適用
  - 通信回避アルゴリズム
  - 通信回避QR法、CAQR  
(Communication Avoidance QR)
  - 米澤CREST採択課題の筑波大櫻井グループと連携し、CAQRをRSDFTの直交化部分に実装中

