

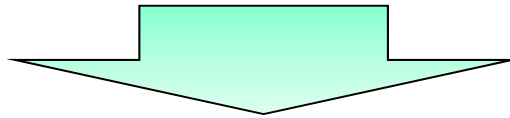
演算加速機構を持つ 将来のHPCIシステムに関する調査研究

主管事業実施機関：筑波大学
計算科学研究センター
佐藤三久

共同事業参画機関：東京工業大学、理化学研究所、
会津大学、日立製作所
協力機関：東京大学、広島大学、
高エネルギー加速器研究機構

概要

- ナノテクやライフサイエンスの進歩、気候気象予測や地震・防災への対応には計算科学は不可欠かつ有効な手段
 - そのためにはさらなる計算能力が要請されている。
 - 設置面積、消費電力等の制限からノード数の増加による並列システムの性能向上には限界
- ライフサイエンスの分子シミュレーション等、多様な分野で比較的小さい一定サイズの問題の高速化が望まれている(いわゆる強スケーリング)
 - 対応した研究開発の例: ANTON, MDGRAPE-4

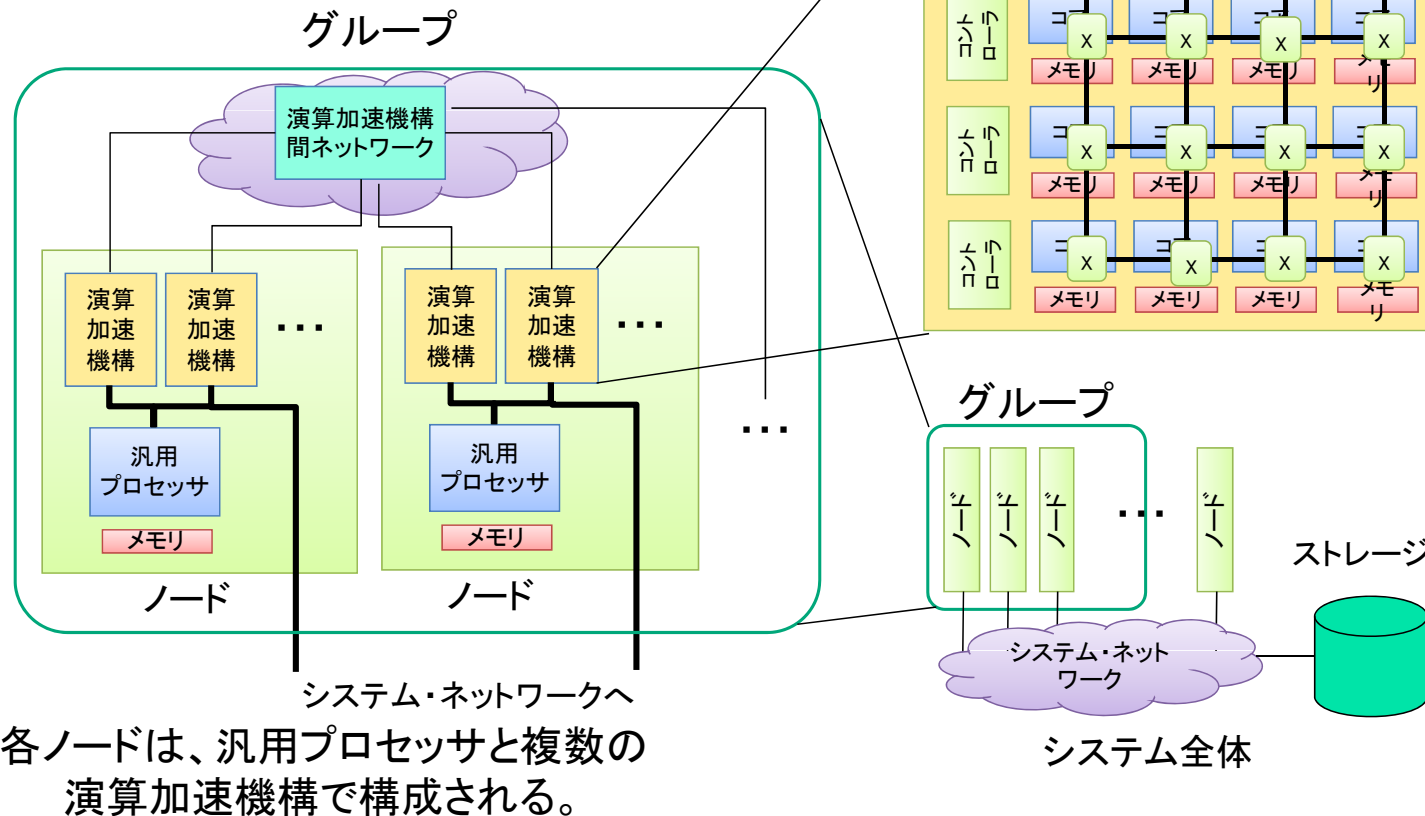


- 電力効率の大幅な効率化と強スケーリング問題の高速化による新たな計算科学の展開を目指して、演算加速機構による並列大規模システムについて調査研究を行う。
 - 平成23年度文部科学省アプリケーション&コンピュータアーキテクチャ・コンパイラ・システムソフトウェア合同作業部会において、まとめられた「今後のHPCI技術開発に関する報告書」の中で、分類されたシステム構成のうち「メモリ容量削減」および「演算重視」のシステムを主な調査研究の対象とする。

FS想定システム

- 多数の演算コアを内蔵したチップによる演算加速機構が汎用プロセッサで構成された並列システムの各ノードに接続もしくは内蔵されているヘテロジーニアスな並列システムを想定

幾つかのノードで、グループを構成し、グループ内の演算加速機構は演算加速機構間ネットワークで接続される。



演算加速機構は、多数のスループットコアにより構成。スループットコアは、チップ内ネットワークにより結合される。図に示したものは一つの例。

各ノードは、システム全体を結合するシステムネットワークで接続される。

最終レポートではこんな感じ(1)

2. サイエンスロードマップ「アプリケーションからの要求の概要」(9/9)

ネットワークレイテンシ

タンパクMD	1時間ステップがマイクロ秒程度。同期等がこれより十分短い必要あり
格子QCD	大域縮約をマイクロ秒程度
他の多く	もう少し余裕あり

ネットワークバンド幅

格子QCD	隣接ノードとの通信速度が B/F で 0.01 程度
大域FFT	バイセクションバンド幅で性能が決まる。普通の構成では効率 1% 以下 ハードウェアだけでなく、アルゴリズム面からの検討も重要

メモリ容量・バンド幅

地震波動解析 圧縮性流体計算 有限要素解析の 防災・工学応用	100ペタバイト前後のメモリ、高いメモリバンド幅(B/F 0.5 以上)が必要
タンパクMD 格子QCD	メモリ必要量は極めて小さい。大きなバンド幅(B/F 1以上) が必要
大規模粒子系計算 量子化学計算	バンド幅、メモリ量とも比較的要求小さい

ストレージ容量 速度

DNA	シーケンサデータ処理 50EB, 500TB/s 程度が必要
他の多く	1桁程度下の要求

多様な要求

- ・複数アーキテクチャも視野にいれる必要あり?
- ・メモリ・ネットワークバンド幅については新しいアルゴリズムの研究開発も重要

最終レポートではこんな感じ(2)

4. ロードマップ達成に向けて(アプリ要求性能)

▶サイエンスロードマップに基づいて2018年ごろのアプリケーションに必要な性能を調査

▶演算性能要求・総メモリ容量・演算性能あたりメモリ帯域・ネットワーク要求を調査

要求性能の解析結果

▶演算性能・メモリ容量・メモリ帯域に関する要求

▶演算性能は800PFLOPS~2500PFLOPS

▶メモリ容量は10TB~500PBの幅があり、帯域も1000倍程度の差がある

▶特徴的なもの

▶メモリ容量が少なくても良い: MD・気候・宇宙物理・素粒子物理

▶メモリ帯域が少なくても良い: 量子化学・原子核物理

▶メモリ容量・帯域が両方必要: 構造解析・非圧縮流体解析など

▶ネットワークに対する要求 ※トポロジに依存する部分もあり継続検討が必要

▶レイテンシ・帯域とも強い要求はないが、性能必要なアプリもあった

▶タンパク質の構造解析などでは1us以下での通信が必要な見込み

▶物質化学分野ではBisection帯域が必要なアプリもある

▶1us以下での高速な同期・放送・縮約などが要求されるアプリケーションもあり、専用のハードウェアによるサポートが必要になる可能性もある

▶ストレージに対する要求

▶要求容量に対して他の課題に比べて大きな課題はない

▶性能要求に対しては今後のストレージデバイス技術に応じて構成方法を検討

▶要求性能をトレンドから予想される性能にマッピングした(図1)

▶サイエンスロードマップの達成には、前スライドの4分類とも、技術トレンドから予想される性能よりも高い数値が要求されている

▶ロードマップ達成のためにアプリケーションの特性をさらに詳細化・定量化し、将来のスーパーコンピュータの設計目標を提示していくことが必要である

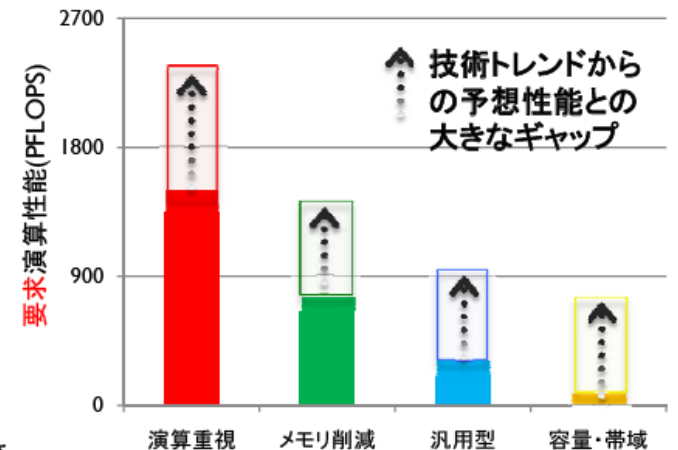
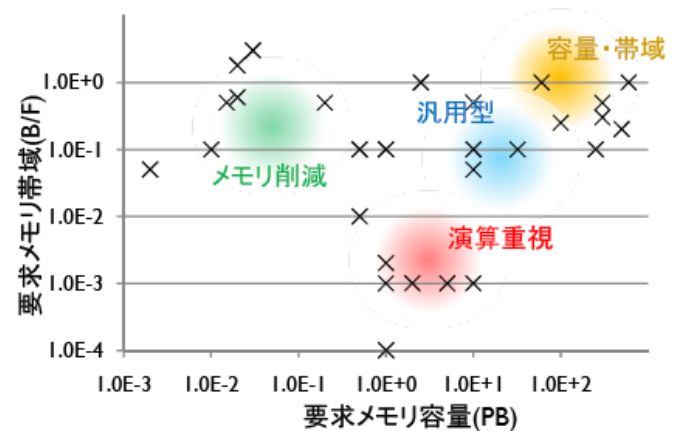
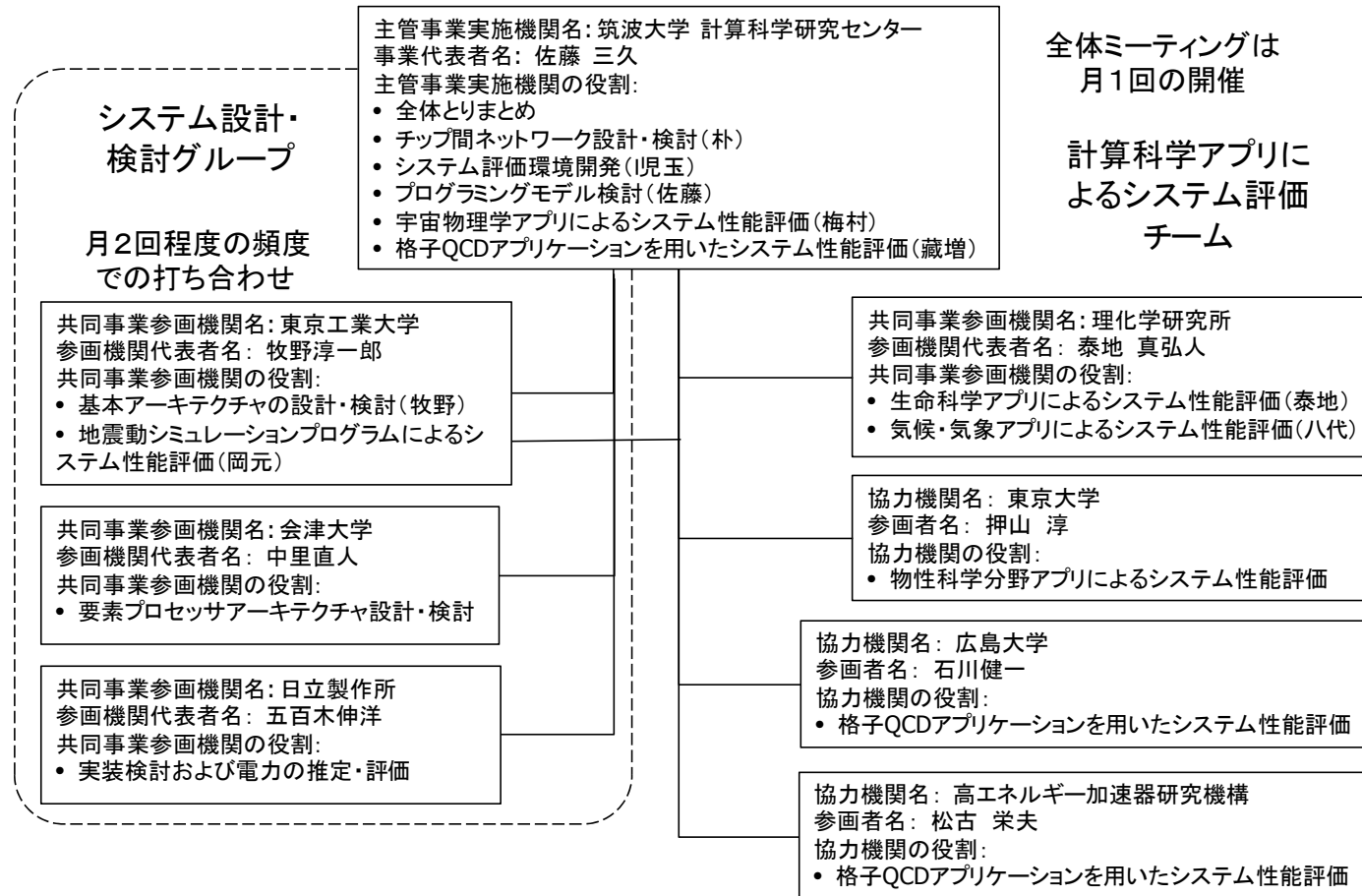


図1. 要求性能と予測性能の相関
上: 各アプリの要求するB/F値とメモリ容量
下: 各分類に対する要求性能値と予想性能

実施体制：実施機関

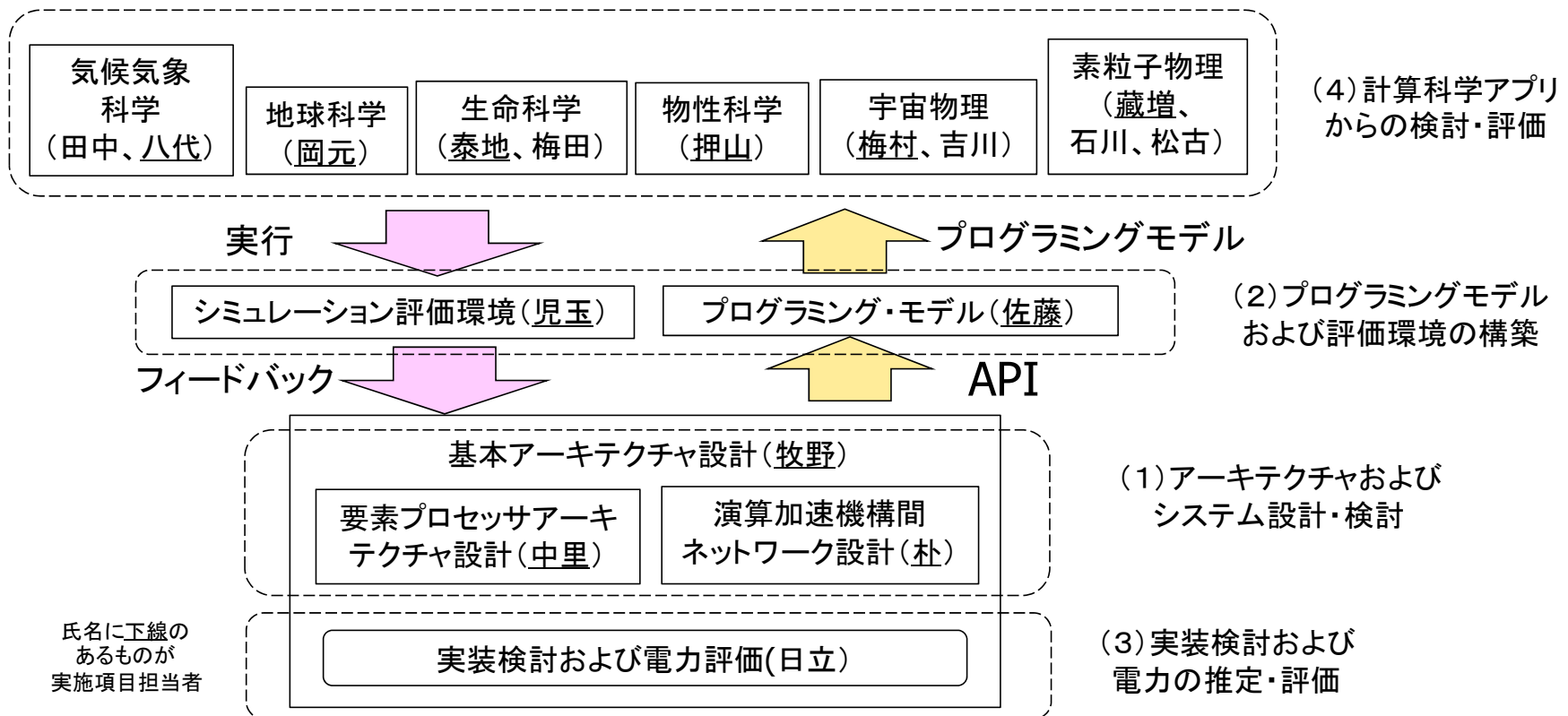
■ 主管事業実施機関：筑波大学計算科学研究センター

- 最先端の計算科学を推進する、計算科学と計算機科学が連携・協働を行う学際的な組織。CP-PACS(1996年top500 1位)など、多数のマシンを開発
- 共同利用・共同研究拠点「先端学際計算科学共同研究拠点」に認定
- 文部科学省の特別経費によるHA-PACSプロジェクト「エクサスケール計算技術開拓による先端学際計算科学教育研究拠点の充実」(2011-2013)

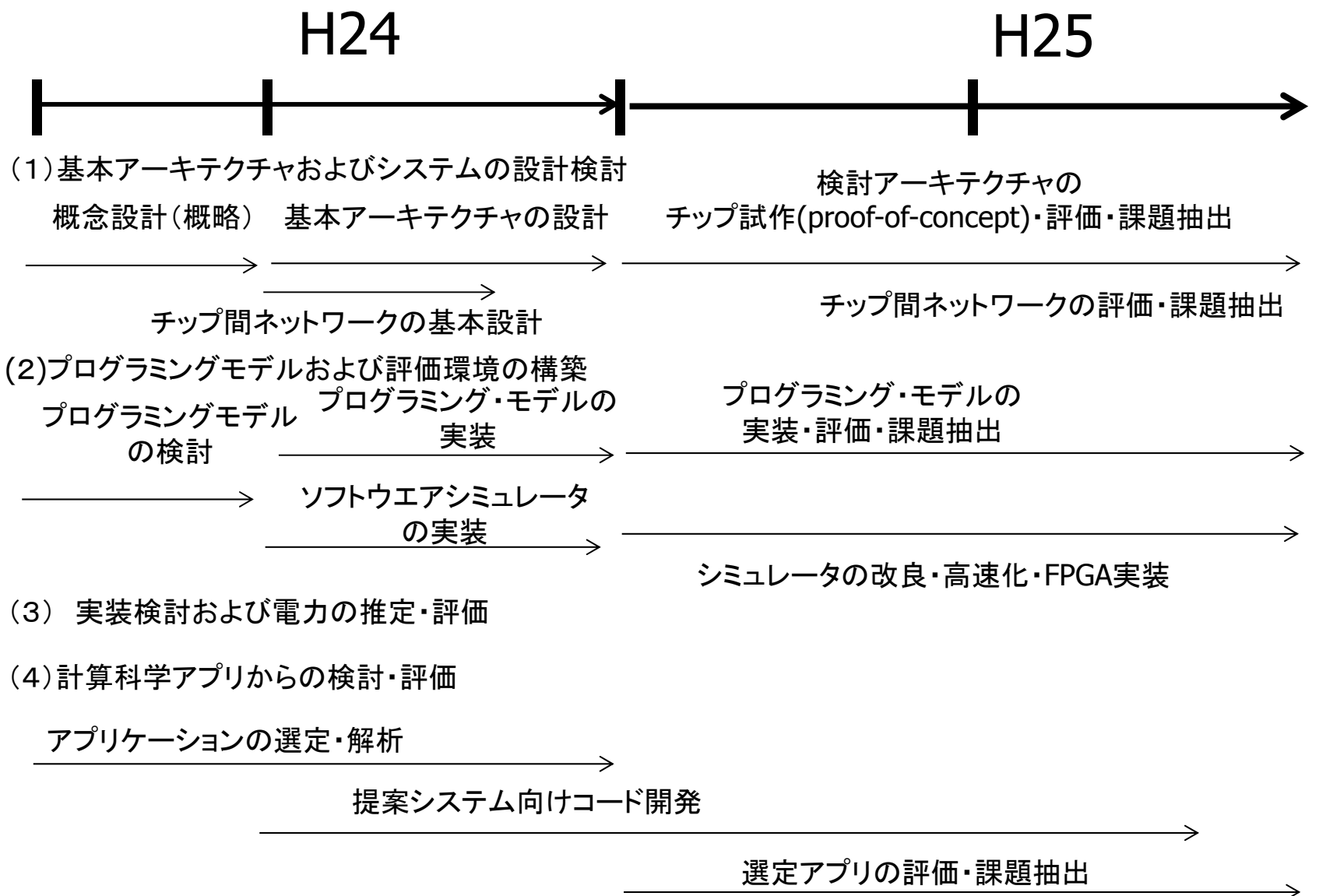


実施体制：実施項目

- 調査研究を、以下の4つの項目に分けて実施
 - (1)アーキテクチャおよびシステムの設計・検討
 - (2)プログラミング・モデルおよび評価環境の構築
 - (3)実装検討および電力の推定・評価
 - (4)計算科学アプリからの検討・評価



実施スケジュール



おわりに

- 本調査研究では、以下の2つが特に重要
 - 電力効率: 現在の汎用プロセッサだけでは電力規模に限界
 - プログラマビリティ: プログラムの書き換えは大きなコスト
- これに対し、
 - Proof-of-conceptのチップを制作、電力の見積もりを精緻化
 - プログラミングシステム、シミュレータを開発し、実際のコードで推計できる環境を構築
- 今後の展開可能性・波及効果
 - 演算加速機構は汎用プロセッサに付加されることを前提としており、エクサフロップスのシステムだけでなく、現在のGPUのように研究室レベルのクラスタやPC、組み込み機器などにも付加し、加速が可能
 - 強スケーリング問題の高速化を可能とするアーキテクチャにより小規模のクラスタにおいても、適当な規模の問題を加速することが可能。

演算重視 / アクセラレータ

- メモリ帯域が少なくても良いとなったアプリケーションの大半が量子系 (密行列の直交化や対角化が計算量のほとんどを占める)
- 後は大規模な粒子系
- GPGPU 的なものでいいが、アクセラレータ側に外部メモリはあまりなくてもいい (そちらにメモリあるならホストはなんのため? という問題も)
- GRAPE-DR ベース?

メモリ削減/SoC

- 想定アプリケーション: 小サイズMD、流体、QCD等
- 大サイズ差分法を out-of-core でできるかどうかは要検討
- 外付けメモリがないか、極端に低バンド幅
- 軽量コアを非常に多数集積して、電力当り性能を上げる
- オンチップメモリに対しては 高B/F
- メモリはチップ当り 1GB以下程度?
- ネットワークは 100GB/s 程度?

もうちょっと具体的なイメージ

- 大雑把には: 70-80年代の大規模SIMDマシンを1チップ化。Goodyear MPP, CM, MasPar 等
- 例: CM-2。 2048 FPU, トータル 512MB メモリ
- 14nm だと 16384 FPU, 256-512MB メモリくらいが入るかも
- 考えるべきことは
 - コア内部アーキテクチャ
 - コア間接続
 - チップ間接続

まとめ

- 「演算加速機構を持つ将来の HPCI システムのあり方に関する調査研究」では、演算重視＋メモリ削減の2タイプのアーキテクチャを1種類のプロセッサでカバーすることで、伝統的なスカラー超並列やベクトル並列では実現困難である
 - － 小規模問題 (強スケールリング) の高速化
 - － 非常に高い電力あたり性能
- を両立させる
- 大メモリ・大メモリバンド幅の両方が必要なアプリケーションは対象にしない。
 - 小規模なテストチップを試作し、実現性を評価する計画である。

FS検討課題のポイント

- 適切なサイズの規模の演算加速機構間の直接通信を可能とするネットワークリンクによる演算加速を検討
 - 現在の市販のGPUでも問題
- 演算加速機構チップに内蔵されているコアに内蔵のメモリを付加したり、特定アプリケーションを加速する演算器、コアで共有するメモリも検討項目
 - 例に示した演算加速機構の構成ではコアの数を増やして演算器の数を増やすことにより「演算重視」のいわゆるアクセラレータになり、ある程度のローカルメモリをコアに付加してチップ内での計算に特化させることにより「メモリ容量削減」のSoC(System-On-Chip)タイプのシステムになる
- システム全体を構成する汎用並列部に必要なスペックの検討
- 現在のGPUと比較して、2010年代後半のテクノロジーを用いることを前提に、演算加速機構については現状の10倍以上の電力効率化を検討
 - 現在、電力効率が良いとされるGPGPUにおいて2TFlops/200W程度(単精度、テクノロジーは45nm)であり、半導体テクノロジーの進歩によって、3-5TF/200W(倍精度、22nm)程度にはなると予想
 - 想定システムでは、総電力は10MW~20MW程度と見込まれるために、エクサフロップスの性能を達成するには、少なくとも、この2倍以上の電力効率を達成が必要。汎用プロセッサ部分を含めるとさらに効率化が必要。
- 想定システムに対するプログラミングモデル、すなわちどうやって既存のアプリを対応させるのかは重要な課題
 - OpenACC, XMP-dev, ... DSL

実施内容 (1/3)

- (1)アーキテクチャおよびシステムの設計・検討
 - プロジェクト全体の基本アーキテクチャの検討後、演算加速機構チップ、演算加速機構間ネットワークの検討
 - 演算加速機構チップについては、proof-of-conceptとして一部を具体的に論理設計し、LSIを制作して消費電力見積もりの精度を上げる。
- (2)プログラミング・モデルおよび評価環境の構築
 - 汎用プロセッサからのAPIやシステム全体を用いるプログラミング・モデルを検討する。プログラミングについては、できるだけ指示文ベースでプログラミングできるようにし、既存のプログラムからの移行を容易にすることを目指す。
 - 評価のために、命令レベルシミュレータを開発する。
 - FPGAを用いたシミュレータの加速装置を開発し、より高速にシミュレーションできる環境を構築する。なお、FPGAについては、現在、筑波大で開発中のHA-PACSシステムの通信部分であるPEACH2(FPGAで実装)の一部を活用して行う。

実施内容(2/3)

- (4) 計算科学アプリからの検討・評価
 - それぞれの計算科学の分野において課題とされる計算科学プログラムを取り上げ、システム設計側と議論し、co-designによるアーキテクチャの設計を協働して行う。その過程において、対象とする分野の計算科学アプリケーションの選定、解析を行う
 - 本調査研究の対象とする演算加速機構アーキテクチャに想定されるアプリケーションとして、以下の2つが考えられる。
 - 行列積などに代表されるように演算密度が高い計算、もしくは特定の演算の頻度が高い部分があり、その部分を特定の演算加速機構にオフロードすることによって高い効果が見込めるもの。(これは、「演算重視」のシステム)
 - 演算あたりのメモリアクセス要求が高く、オンチップメモリや特定のメモリアーキテクチャの改善によって、高い効果が認められるもの。(これは、「メモリ容量削減」のシステム)
 - 現在のGPUによるプログラミングと共通の部分もあることから、HA-PACSのGPUを活用してコード開発を始め、本調査研究で検討するアーキテクチャの評価環境が利用できるようになり次第、評価、問題点や改良点をフィードバックする

実施内容 (3/3)

- (3) 実装検討および電力の推定・評価
 - 概念設計レベルでの実装可能性及び推定電力の評価を行うとともに、冷却シミュレーションなどを含む、システム構築の観点からの検討を行う。
 - なお、本調査研究においては演算加速機構を付加する汎用プロセッサやシステムネットワーク、ストレージ等については主な検討項目とはしていない。これについては、コモデティ製品、もしくは他の調査研究で行われるであろう汎用のシステムを想定する。
 - 以上を踏まえて、最終的な研究開発ロードマップとコスト見積もりを得る。

検討状況

- これまで、3回のMeetingを開催し、アーキテクチャの方向性、対象アプリケーションの説明を行った。
- 演算集約型 (GRAPEのような?) とステンシル計算を両立させるアーキテクチャ (プロセッサ、ネットワーク)
- どの程度のネットワーク (演算加速チップ間) が実現可能とみるか。
 - 筐体程度?、それ以上? (全部つなぐことが実現可能か、「汎用」と同じくらいのネットワークをつくることができるか)
- 外付けのメモリは、必要か
 - 他の「調査研究」とのすみわけ (例えば、外付けのDRAMをつけると、「汎用」とどちらがうのか、わからなくなる)
 - どのアプリを捨てるべきか
 - 積層メモリ等をつかうべきか
- ホストとの関係