「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」中間報告

主管事業実施機関：　筑波大学
計算科学研究センター

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

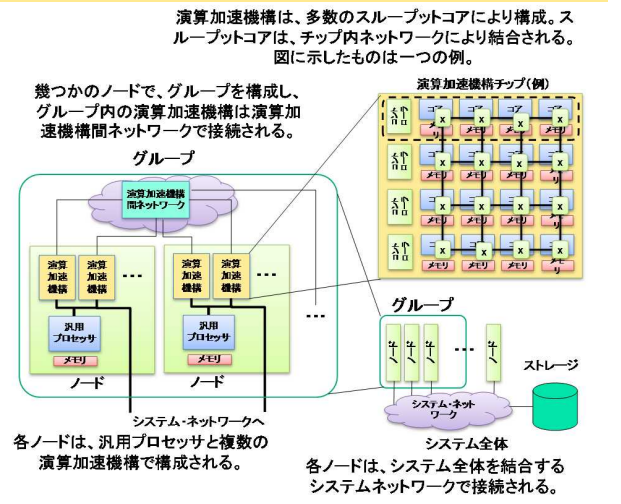
- ナノテクやライフサイエンスの進歩、気候気象予測や地震・防災への対処には計算科学は不可欠かつ有効な手段
  - そのためにはさらなる計算能力が要請されている。
  - 設置面積、消費電力等の制限からノード数の増加による並列システムの性能向上には限界
- ライフサイエンスの分子シミュレーション等、多様な分野で比較的小さい一定サイズの問題の高速化が望まれている(いわゆる強スケーリング)
  - 対応した研究開発の例: ANTON, MDGRAPE-4

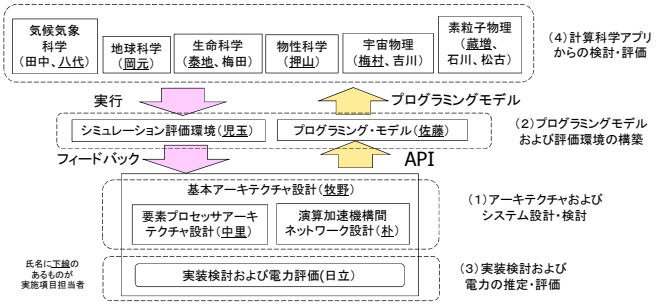
電力効率の大幅な効率化と**強スケーリング問題の高速化**による新たな計算科学の展開を目指して、演算加速機構による並列大規模システムについて調査研究を行う。

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。

- 主管事業実施機関：　筑波大学　計算科学研究センター
- 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
- 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

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





①社会的・科学的課題の達成可能性 (1)

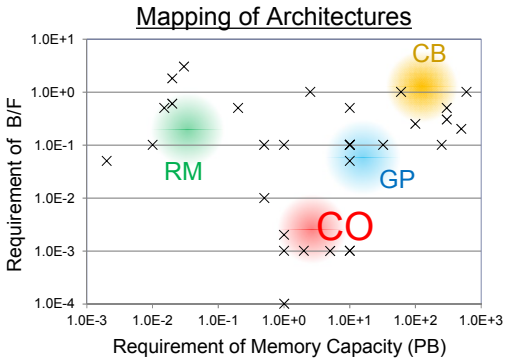
- 計算科学に対する**社会的・科学的課題の達成のために必要なアプリケーション**のうち、本調査研究で対象とするメモリ削減型(RM)および演算重視型(CO)で、ある程度の実行効率が期待できるものの洗い出しを進めている。
  - 生命科学、物性科学における分子動力学計算、生命科学、物性科学、ものづくり分野における第一原理計算、素粒子物理における格子QCD、原子核物理における様々な手法、宇宙物理における粒子シミュレーション、流体計算等(合同作業部会報告より)

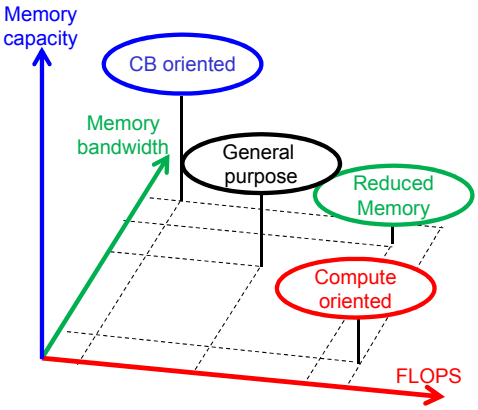
- 以下の5つのアプリのカーネルをターゲットとしてアーキテクチャとのco-designを進めている(今年度)
  - 格子QCD　(素粒子分野)
  - 重力多体計算　N-body (宇宙物理分野)
  - 磁気流体コード HMD　(宇宙物理分野)
  - 分子動力学　MD (生命科学)
  - 地震波計算コード(地球物理)

➢ **強スケーリング**による分子動力学アプリケーションの実時間の大幅な高速化

➢ **電力効率の大幅な効率化**による格子QCD等のメモリ削減型アプリケーションの大規模・効率的実行

(合同作業部会報告より抜粋)



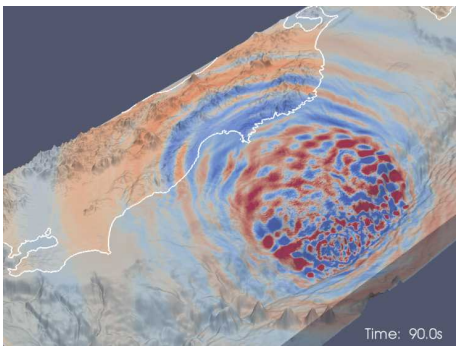


①社会的・科学的課題の達成可能性 (2)

- 本調査研究では科学技術計算における**典型的な計算の一つであるステンシル型の並列アプリケーション**について適用可能の検討を進めている
  - 地震シミュレーション　地震波計算コード　FDM
  - 気象シミュレーション NICAM

- 本システムは、汎用ホストに接続されて利用されることを想定。
  - コードのカーネル部分を演算加速機構部(グループ)にオフロードして、アプリケーションの実行を加速する。そのためには、どの部分がオフロードできるかを見極めることが重要
  - 社会的課題に関するアプリケーションについても(間接的に)一部を加速できる可能性がある。

- 来年度の予定:NICAM(気象)、RS-DFT(物性)、FMO(化学)



- 地震波計算コードの検討状況
  - オンチップメモリでの計算では、地震波の実時間より高速なシミュレーションが可能になる。
  - より大規模・高精度なシミュレーションが必要な場合、チップ内のメモリでは足りない場合もあるため、チップ内積層メモリやメモリを効率的に利用する新しい解法の検討が必要

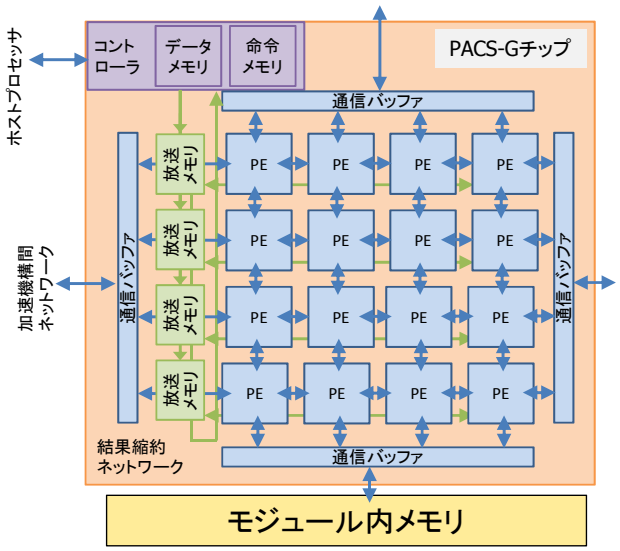
PACS-G アーキテクチャの概要： ノード（チップ）

- 以下のアーキテクチャを、**Straw man（たたき台）アーキテクチャ**として設定

- 演算集約型とメモリ削減型の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）をターゲットに設定
- 2018～2020年のLSIテクノロジーとして、14nmを想定。チップサイズを20mm^{2}として、メモリ(SRAM)換算で1GB/チップを想定
- チップの基本アーキテクチャは、SIMD
- チップ内は、2次元のメッシュ・ネットワークを（当面）想定 （＋ブロードキャスト・リダクションネットワークを検討）、コア間 16GB/s （双方向）

- コアとメモリの比を1:1 として、チップあたり4096 コア（PE）= 64 x 64
- チップ内メモリ 512MB/チップ, 128KB/コア
- コアの基本性能は2FMA@1GHz, したがって、4GFlopsx4096 = 16TFlops/チップ (64Kチップ/1EF)
- TSV 2.5次元実装によるモジュール内メモリを想定。HMC もしくはWide IO DRAM で、
  - バンド幅は1000-1500GB/s
  - サイズは、16-32GB/chip程度
- チップ外付けメモリ(DDR/DIM)は、想定しない

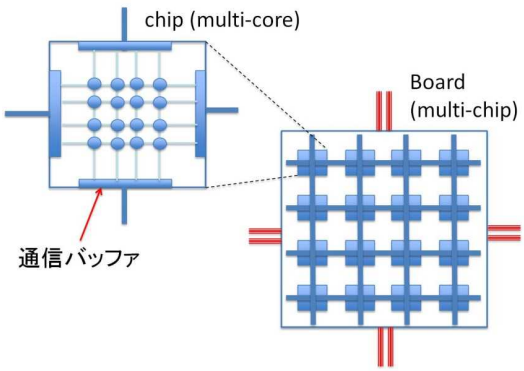
- 電力は250W/チップを目標 （16MW/1EF）
- 2048 チップ/group, Group内のチップは演算加速機構ネットワークで結合



PACS-G アーキテクチャの概要： ノード間、ホスト間

- 1024～2048チップ（グループ）ごとに演算加速機構間ネットワークで結合
- チップの2次元メッシュネットワークをボード上のチップ間ネットワークに展開する際、ボード上のチップ（例えば4x4=16個）を同様に2次元メッシュ結合すると、隣接チップ間（数cm～10cm程度）接続のバンド幅は、チップ内隣接ネットワークの20～40%程度で実現可能（電気配線）
- さらに、ボード間ネットワークまでも2次元（あるいはより高位の多次元）メッシュ展開とすると、インタフェースチップからの光コネクションが使えればチップ間バンド幅と同等（=チップ内ネットの40%程度）が実現可能。

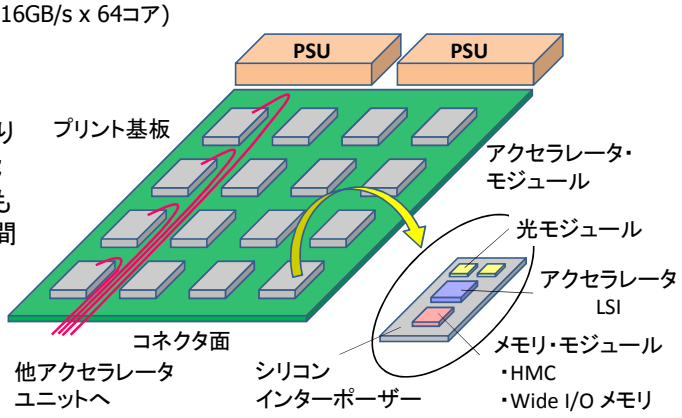
チップ内コア間：
16GB/s (= 1GHz x 8B x 2(双方向)@コア間)→1024GB/s (= 16GB/s x 64コア)
チップ間（ボード内、ボード外）：
200～400GB/s (= 32ch. x 25～50Gbps x 2(双方向))



チップ間ネットワークの検討例

- QCDのような隣接通信のアプリであれば演算のB/F値よりラックサイズのシステムまではメッシュのままで対応可能
- 当面、2次元メッシュで考えるが、もう少し高次元の実装も検討。また、メッシュをトーラスに変更することは、チップ間配線の実装で対応可

- 検討しているシステムは、汎用CPUを基本とした超並列システムにアタッチされることを想定
- ホストとのインタフェースは、PCI Express Gen4 x 16 相当の性能を期待



システムの実装イメージの検討例

⑤ システムの評価アプリによる性能評価(性能概算)

- 計算の1ステップでの演算、メモリアクセス、通信のパターンと発生量を分析し、想定したプロセッサアーキテクチャ、ネットワークアーキテクチャから期待できる性能を推定。現時点では、グループ単位(～2048チップ,32PF)に限定。
  - 全部データがオンチップに乗る場合:演算・メモリ性能　4 B/F, メモリ1TB/group
  - データが積層メモリに乗る場合:演算・メモリ性能　0.05B/F, メモリ32TB/group

アプリケーション	想定問題サイズ	効率・性能	コメント・比較
格子QCD (素粒子物理)	物理体積(12fm)^4 ハドロン多体系 128^{4}格子	12%～53% 7.9 ～34.7PF　2048 チップ(単精度 peak 65.5 PF)	• 評価対象アルゴリズム:領域分割前処理単精度クォークソルバー(ウィルソンクォーク型、BiCGStab法) • オンチップのメモリのみを利用 • 通信レイテンシパラメータの範囲で性能に幅がでる。 • 京では、効率が26%,　32768ノード、1.1 PF
磁気流体コード (宇宙物理)	セル数　1984^3	1.89 PF, 22.5% 512チップ(8PF)	• HLL近似リーマン解法、磁場をflux-CT法よる有限体積法。時間積分を2次精度のTVD Runge-Kutta法 • グローバルメモリを利用 • 210-220ms/step,　Intel Core i7 4096コアで4.5s/step
重力多体計算 (宇宙物理)	814G interaction/sec/chip (単精度、無衝突系)		• 重力計算を演算加速機構で加速。粒子の軌道計算はホスト計算機で行う。オンチップメモリのみ使用 • Intel Xeon E5-2670 の 66.7倍
分子動力学 (MD)カーネル (生命科学)	1セル/コア、1セル (5Å)^3, カットオフ 半径12Åを仮定 2580原子/コア	3.67PF、 最大15M原子 /256チップ, 784.4us/ステップ	• 近距離相互作用の直接和計算を計算。遠距離相互作用計算、結合力計算は未評価 • セルインデックス法(空間座標分割)とハーフシェルスキームを仮定 • 通信ネックになっていないため小規模問題ではさらに高速化可能? • 京では、全ノードで500M原子、4.6PF, 114ms/ステップ
地震波 計算コード (地球科学)	格子サイズ 2048x2048x512	3.5 PF /1024チップ	• 3次元時間領域差分法(FDTD)、空間差分4次精度、時間差分2次精度、弾性体、速度と応力を変数とするスキーム • オンチップのみ。今後、グローバルメモリも検討。 • 格子間隔 50 m, 最小横波速度 300 m/s を想定した場合 (100 × 100 × 25 km, Δt～0.001s), 実時間の10倍程度の速さ • 現状のGPUクラスタでは実時間の1/20 程度の計算速度(1/200)

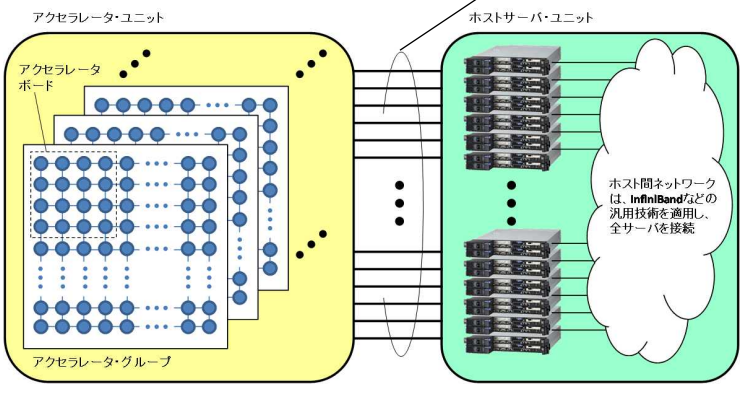
汎用システム(ホスト)との統合イメージ

- 電力・性能等を勘案して、次の2つの構成を想定 (目標全体性能 1EF)
  - ケース1:　ホスト　100PF,　演算加速機構部　900PF
  - ケース2:　ホスト　300PF,　演算加速機構部　700PF

- システム(論理)構成イメージ
  - 1,024～2,048のアクセラレータ・チップで、グループを構成(図は、2次元メッシュ)
  - 20～50セットのアクセラレータ・グループにより、アクセラレータ・ユニットを構成　(グループ数は、想定性能、グループあたりのチップ数による)
  - ホストサーバーは、Xeon Phi相当のサーバを想定

- 演算加速機構の利用イメージ
  - 個々のホストの一部分のコードを演算加速機構にオフロードする。または、ライブラリとして呼び出す(現在のGPUと同様)
  - 並列実行する一部分のコードを演算加速機構にオフロードする。オフロードされ演算加速機構で実行される部分は、演算加速機構ネットワークで通信する。(XcalableMP +OpenACCで記述)
  - ある程度のプロセッサを演算加速機構に付加することも検討。これによりオフロードできる部分を増やす

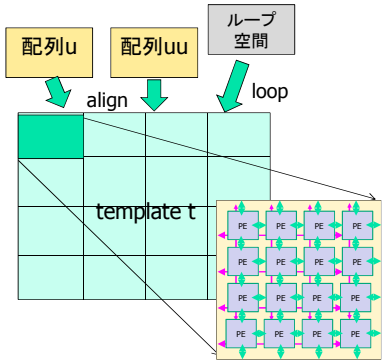
- ホストサーバとアクセラレータ・ボード間をPCI Express、もしくは、専用の高速伝送路で接続
- ホストサーバ 5～12台に対して、アクセラレータ・ボード 1台の割合で接続
- ホストサーバは、Xeon Phi サーバを想定時(2018年)、18,000～33,000台で構成
- 筐体の中でのアクセラレターボードとサーバの混在の案もあり。



プログラミング・モデルの検討および評価環境の構築の状況

- 性能評価のためのクロックレベルのシミュレータと、ソフトウエア開発用命令レベルシミュレータを開発し、評価を進めている。
  - 命令セットの設計、ネットワーク構成などの設計・評価、モジュールメモリ内の利用、
  - 具体的コードによる定量的評価

- プログラミングモデルの検討
  - アセンブラレベルのSIMDプログラミングをするためのCのsubsetのような言語
  - ユーザに提供するための言語として、XcalableMPの拡張を検討
  - （C＊などのデータ並列言語）



- XcalableMP + OpenACCによるプログラミングモデル
  - ホストからチップにオフロードするために、OpenACCの指示文を用いる。
  - チップの中のプログラミングに、XcalableMPのtemplateを利用
    - Templateは、データやindex空間をマップするための仮想格子
  - ループのプロセッサと配列を整合させることで、PEあたりのコードを生成することができる。
  - データを積層メモリに置く場合は、仮想プロセッサという形でマッピング（？）

- Template directiveで、宣言
- (distribute directiveで、templateをpeにmapping)
- align directiveで、配列を整列
- loop directiveで、ループの実行プロセッサを割り当て

```
#pragma xmp template t(0:XSIZE+2, 0:YSIZE+2)

double u[XSIZE+2][YSIZE+2],uu[XSIZE+2][YSIZE+2];
#pragma xmp align u[i][j] with t(i,j)
#pragma xmp align uu[i][j] with t(i,j)

#pragma xmp loop on t(x,y)
  for(x = 1; x <= XSIZE; x++)
   for(y = 1; y <= YSIZE; y++)
      u[x][y] = (uu[x-1][y] + uu[x+1][y]
              + uu[x][y-1] + uu[x][y+1])/4.0;
```

②システム開発に必要な要素技術の実現可能性
開発に必要な期間、展開可能性

要素技術	概要・現状	見込み
(1) 14nm LSI テクノロジ	• 現在、利用可能なテクノロジーは、22-28nm • 演算加速機構のチップ製造のテクノロジーとして、14nmを前提として、検討を進めている	2018年までには、14nmテクノロジーがTSMC等で、利用可能になると思われる
(2) チップ内積層メモリ技術	• モジュール内に大容量メモリを付加することによってカバーできるアプリケーションを幅を広げる • メモリをシリコンインターポーザによる2.5次元実装技術 • HMC (Hybrid Memory Cube)、もしくWide I/O DRAMをシリコンインターポーザで同一パッケージに実装	2018年にはこれらのテクノロジが利用可能になると見込まれる。
(3) LSI間伝送技術および光伝送技術	• 演算加速機構のチップ間の伝送について、50Gbpsをターゲットに設定 • InfiniBand, 100Gbpsイーサネット規格技術では、25Gbps • 50Gbpsでは、プリント基板内の配線は数十センチ。ボード端面のコネクタまで配線が困難 • アクセラレータLSIのシリコンインターポーザに光変換デバイスも搭載し、マルチチップモジュールから直接、光信号を引き出す	伝送速度が25Gbpsになった場合に、システム性能へ与える影響について検討も進めている 50Gbps伝送, チップ間光通信については、2018年に採用できるか、今後、継続して技術動向の確認が必要
(4) ホストとのインタフェース	演算加速機構は、汎用ホストに接続されて、利用されることを想定	現在のところ、標準の規格であるPCI Express Gen3の次の規格を想定
(5)コンパイラ・プログラミング環境	演算加速機構を幅広く利用するためには、そのためのコンパイラと関連するプログラミング環境が必須	さらに検討・改善に、数年にわたる研究開発を行う必要

- 技術展開可能性については、(1)(2)によるチップを単体ボードに実装し、として、現在のGPUのように演算加速機構ボードとして展開することが考えられる。

電力評価の状況

- GRAPE-DRのアーキテクチャをベースに28nmでの電力の見積もりを進めている
  - GRAPE-DR プロセッサコアは、1サイクルで倍精度加減算1つ、単精度乗算1つを実行でき、2サイクル毎に倍精度乗算を実行
  - 消費電力は 512 コア、500MHz 動作で65W であり倍精度で 4Gflops/W、単精度で 8Gflops/W（システム全体では、1.5GF/W）

- GRAPE-DR コアに、改良をおこなった RTL レベル設計を用いて、予備的な電力評価を 28nm プロセスを用いておこなった。改良点のうち電力消費削減に寄与するのは以下。
  - 倍精度演算に最適化した乗算器+単精度専用の演算の構成に変更し、消費電力の若干の増加でスループットを2倍にした
  - レジスタファイルについて、IPマクロを使う形に変更した
  - ターゲットクロックを低め(400-600MHzで評価)に設定し、低しきい値 トランジスタの利用を避ける
  - この他、メモリ容量の増加、PE間ネットワークの強化等の変更を行っている。

  - 消費電力見積もりはまだ予備的なものだが、いわゆる "Typical" 値では44Gflops/W(最悪値では6割以下だが、これは供給電圧の寄与が大きく非現実的)であり、35-45 Gflops/W を実現できる見込みを得た

- 現時点の見込み： 14nm プロセス で 70-90 Gflops/W であり、外部メモリインタフェース等を含めて 60-80GF/W 程度と予想される

- proof-of-conceptとして演算加速機構チップの一部を具体的に論理設計し、テストチップを試作し、消費電力見積もりの精度を上げる予定
  - TSMC のシャトルで試作できる最小単位である 6平方ミリのチップ
  - 90nmのGARAPE-DRからの外挿では精度に限界がある。

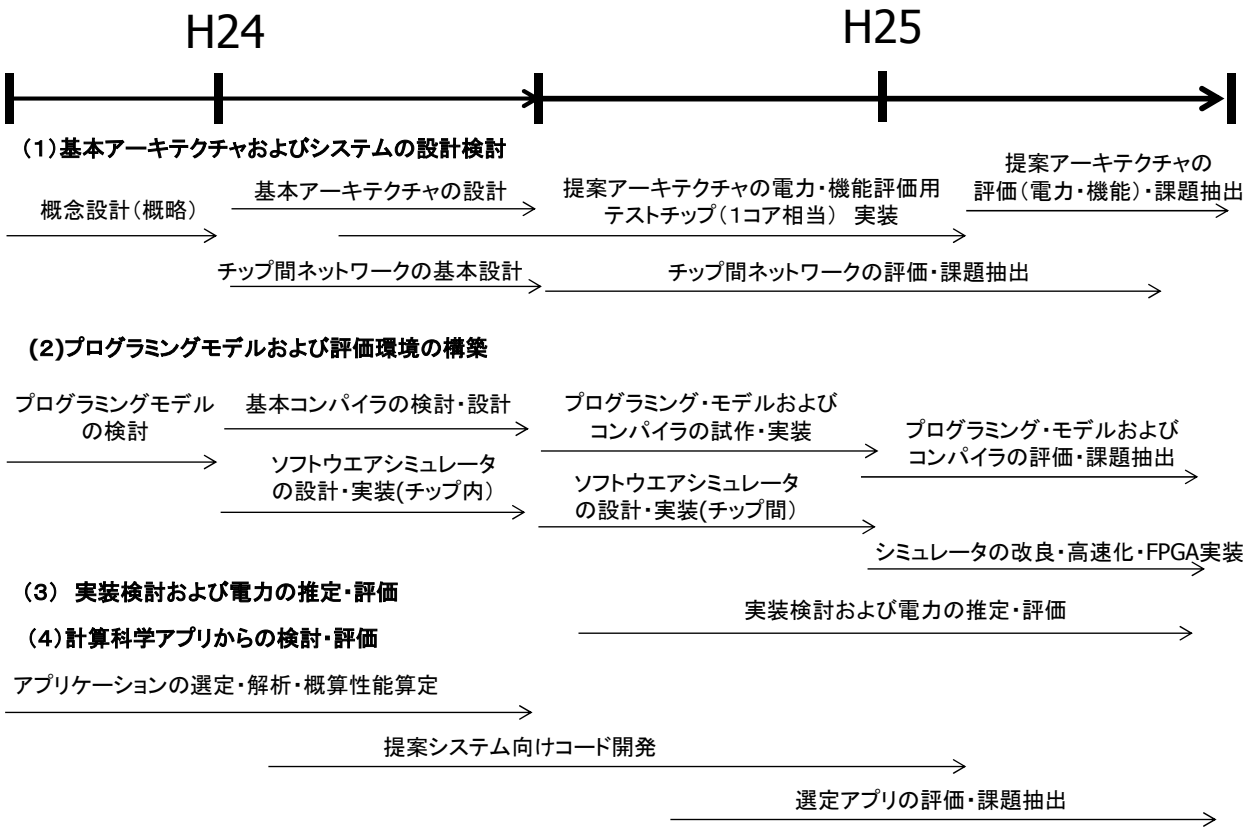
③システムの消費電力、耐故障性、信頼性

- 消費電力

項目	概要	ケース1	ケース2
構成	演算加速機構は汎用ホストに接続される。全体で1EF (peak)	演算加速機構部 900PFLOPSの場合	演算加速機構部 700PFLOPSの場合
チップ数	1EFでは、62,500チップ	56,250	43,750
チップ全体(MW)	チップあたりの電力は、320～200W(50～80GF/W)	18.0～11.3	14～8.8
メモリ全体(MW)	1チップあたりのモジュール内メモリは20～10W	1.1～0.56	0.8～0.4
周辺 (MW)	ボードあたりの電力100～50W	0.36～0.18	0.3～0.1
加速機構部分合計		19.5～12.0	15.1～9.3
ホスト消費電力	京、およびメニーコアの動向から推定	10MW（10GF/W（京の10倍）を想定）	15MW（20GF/W（京の20倍）を想定）
システム消費電力		29.5～22.0	30.1～24.3

- 耐故障性・信頼性
  - ①予備コアの列と行を用意しておき、障害コアが検出された時には、障害コアの列または列をスルーモードとする。（2次元メッシュの場合）
  - ②メモリ信頼向上として、ECCを付加するほか、ネットワークについても、ECCを付加して信頼性向上を図る。
  - ③システムは、ある程度の規模で演算加速機構チップをネットワークで接続する構成になる（現在のところ、2048チップを1グループとして結合）。障害が起きた場合は、このグループごとに切り離して、運用

実施スケジュール



課題

- 次のステップはより詳細なプロセッサアーキテクチャの決定、命令レベルシミュレータの開発と、それによるアプリケーション(システム全体)の性能評価である
  - ネットワークトポロジーの検討(2次元メッシュでいいか?)
  - ある程度のプロセッサを演算加速機構に付加することも検討
  - プログラミングモデルとコンパイラの実装
  - 実際のコードを用いた定量的評価
  - 来年度の予定:NICAM(気象)、RS-DFT(物性)、FMO(化学)
- 全体システムの検討
  - ホストとの接続形態の検討
  - システム全体としての耐故障機能
  - I/O等
- 詳細な電力評価

補足説明資料

各アプリの性能概算について

格子QCD（素粒子物理） 性能概算

- ターゲット問題： 物理体積$(12\text{fm})^4$ ハドロン多体系
- 問題サイズ：格子間隔 0.1fm => 128^4格子
- 性能評価対象アルゴリズム：領域分割前処理単精度クォークソルバー（ウィルソンクォーク型、BiCGStab法）
- 必要性能：収束までの反復回数10000回を仮定。1年で10^6回解く必要があると仮定。一収束までに32秒以内で解ける性能が必要。
- 評価方法：本アーキテクチャ向けに試作したソルバー(SIMD,マルチコア考慮)をもとに、計算量、メモリアクセス量、通信量を数え上げて、アーキテクチャパラメータに基づき計算時間を評価
- 2048チップ評価 （単精度ピーク65.5 PFlops）
  - 効率：12%～53%、実行時間 ：3秒～12秒
  - 通信レイテンシパラメータの範囲で性能に幅がでる。
  - ソルバー部メモリ容量はオンチップメモリに載る。
  - HMC法全体の計算時間のうち90%は本ソルバーで実行される。HMC法全体評価のためのホスト部とアクセラレータ部のデータの出し入れやソルバー以外の計算については今後の評価対象
  - 5次元カイラル型クォークについての性能効率については本評価と同じになる見込み。

磁気流体コード(宇宙物理)性能概算

概要

理想磁気流体計算コードをターゲット計算機システム上で実行性能を行った。結果は次の通りである:
1チップ実行
　ローカルメモリのみを使用する場合についてのみ算出。
　この場合、セル数124^3に対し、最大で、実行性能は4.475[TFLOPS]、実行効率は27%であった。
複数チップ実行
　グローバルメモリを使用してセル数1984^3の場合について算出。チップ数512(ボード数にして32)で実行した場合、最大で、実行性能は1.89[PFLOPS]、実行効率は22.5%であった。2次精度の時間積分法を用いる場合、1ステップに要する時間は210-220[ms]であり、1ヶ月で約10^7ステップの計算を行うことが可能である。

1. 計算コード及びターゲット計算機システム

性能評価に用いた磁気流体計算コードは、流体をHLL近似リーマン解法、磁場をflux-CT法によって解く有限体積法に基づいたコードである。時間積分を2次精度のTVD Runge-Kutta法によって行う。必要な変数の数は、1セルあたり倍精度19個である。

性能評価を行う演算加速システム(ターゲット計算機システム)の詳細は、PACS-Gアーキテクチャの説明ページ(p.2)を参照。

2. 計算アルゴリズム

各チップは袖領域を含めて128^3個のセルの情報を保持し、内124^3個のセルの更新を行う。袖領域の情報は事前に取得する。グローバルメモリを使用する場合、グローバルメモリ上から124^3の単連結な計算領域を袖領域を含めてローカルメモリにロードして計算を行う。この計算方法の下、通信時間、実行性能等を見積もる。

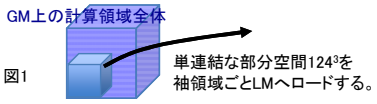


図1

3. 性能評価　※紙面の都合から1チップ実行の場合の評価を、資料として載せる。

セル1個の流束計算に要する演算数は16062、メモリ読み出し回数は1836、メモリ書き込み回数は336である。1サイクル1命令で実行する場合、実行時間は16[μs]、メモリIOは1[μs]。

今、1つのセルは8^3個のセルを計算する。この内、z方向に関して上下それぞれ2層(計4層;セル数にして1/2)は、z方向の袖領域通信(ホップ数16)が発生する。この通信時間をTとすると、実行時間T_{exec}は、

$$T_{exec} = 8^3 \times [17[\mu s] \times (1/2) + (16 + \overline{T})[\mu s] \times (1/2)]$$

と書ける。通信時間をホップ数を考慮して推定すると、15.6[μs]となるので、1セルの平均実行時間は24.3[μs]で、1ステップの実行時間は、12.4[ms]。実行性能は0.66[GFLOPS/コア]。SIMDとFMAにより演算時間を6割削減できた場合で、1.1[GFLOPS/コア]。このとき、チップあたりの実行性能と実行効率は、4.475[TFLOPS]、27%となる。

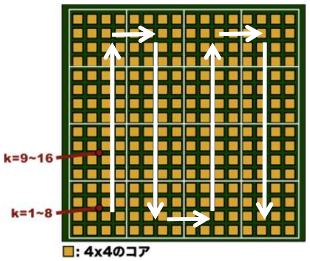


図2: 計算領域124^3のチップ内へのマッピング。
■は4×4コアのグループを示す。124^3をz分割して、蛇の字状にマッピングする。

重力多体計算(宇宙物理)　性能概算

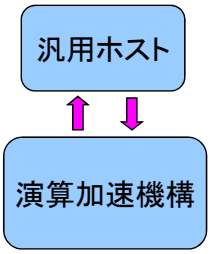
- 計算対象

無衝突系重力多体シミュレーション：　銀河、銀河団、宇宙大規模構造

衝突系重力多体シミュレーション：　星団、銀河中心ブラックホール周辺

(衝突系のシミュレーションは重力も軌道計算も高い精度が必要)

- 重力計算を演算加速機構で加速。粒子の軌道計算はホスト計算機で行う。



従来のGRAPEシリーズと同様の手法

各演算コアで粒子4個から1個の粒子に及ぼされる重力を計算

高速な逆数平方根命令を利用することを想定

- 性能の見積もり

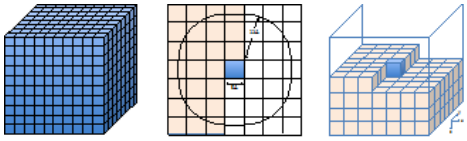
無衝突系の場合：　1コアで4 相互作用に44サイクル

⟺ 370G interaction/sec/chip = 5.9T interaction/sec/board

衝突系の場合：　ざっくり見積もって 180G interaction/sec/chipくらい

分子動力学(MD)カーネル(生命科学) 性能概算

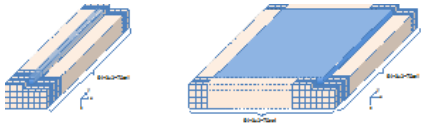
- 評価の概要
  - 最大負荷部分である二体相互作用計算を近or遠距離二つの相互作用計算に分けて計算する. この内, 近距離相互作用の直接和計算が演算加速機構へのオフロードに適する.
  - セルインデックス法(空間座標分割)とハーフシェルスキームを仮定
  - 1コア1セルを仮定(強scaling重視)
  - 遠距離相互作用計算, 結合力計算は未評価



セルインデックス法とハーフシェルスキーム

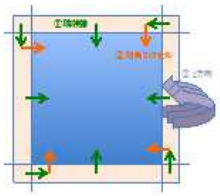
- コアあたりの問題サイズ
  - 1原子は{座標, 電荷, 原子種}の5単精度語
  - L-Jパラメタσとvεの2単精度語は原子種から間接参照
  - 1セル/コアを仮定(強scaling重視)
  - 1セル(5 Å)^3, カットオフ半径12 Åを仮定→3セル先まで必要
  - ハーフシェルは1辺7セルboxの半分で(7^3-1)/2=171のjセル
  - 生体分子シミュレーションの一般的な原子密度はおよそ120原子/(10 Å)^3
  - 15原子/セル, 172セル/コアなので2580原子/コア
  - 38Kペア/コア

- 演算加速機構チップ内実行評価
  - コア内実行
    - 1原子ペアの力の計算はほぼ80FLOP, 3MFLOP/コア
    - Read: 7単精度後=28Byte/ペア, 1MByte/コア
    - Write: 座標3単精度語=12Byte/ペア, 0.5MByte/コア
    - 実行時間は演算ネックで3MFLOP/4GFLOPS=732us
  - コア間通信
    - チップ内コア64 x 64をハーフシェルのx-y平面, z軸方向はチップ間通信と仮定
    - jセルの座標データ301KByte/列を64並列で放送すると39us
    - 計算した力の集約にReduction通信可の場合は0.23us, 不可の場合は0.81us
  - チップ内実行性能
    - 772us/ステップ, 14.9TFLOPS/チップ



1列(64コア分)の放送データと隣接チップからの受信セル

- 演算加速機構グループでの評価
  - チップ間通信
    - x-y平面での通信に2段階で82.3ns, z軸方向の通信に6.1us
    - 座標配信と力集約で同一通信パターンを2回行うので, 12.4us
  - グループ内実行性能
    - 最大15M原子/256チップ
    - 3.67PFLOPS/256チップ, 784.4us/ステップ



チップ間通信パターン

参考 京の事例

- カットオフ28 Å (原子あたりのカットオフ計算は12 Å の場合の13倍)
- 1セル(10 Å)^3
- 64セル/ノード (6500原子/ノード)
- 全ノードで、542,644,272原子, 500M原子
- 全ノード 実効性能 4.6 PFLOPS

時間消費 114 ms/ステップ

- カの計算 92 (うちカットオフ計算89)
- 通信 16
- その他 6

強scaling

- 粒子数1/8 (800原子/ノード)で計算時間は1/4 (性能1/2)
- 粒子数1/64 (100原子/ノード)で計算時間は1/10 (性能1/6)

地震波計算(差分法)(地球物理)性能概算

- 地震波計算の中心部分の性能見積もり
  - 3次元時間領域差分法(FDTD)
  - 空間差分4次精度、時間差分2次精度
  - 弾性体、速度と応力を変数とするスキーム
  - 変数は1セルあたり12個、速度・応力型スキーム
  - コアあたり0.8 GFLOPS、チップ全体で3.4 TFLOPS
  - 256チップの場合に 2048x2048x128、861 TFLOPS

- 1グループ1024チップ(オンチップメモリ 512GB)の場合
  - 格子サイズ 2048x2048x512
  - 性能 3482 TFLOPS
  - 格子間隔 50 m、最小横波速度 300 m/s を想定した場合
    - 100×100×25 km の領域、最高周波数 10Hz、Δt～0.001s
    - 弾性体版では実時間の10倍程度の速さになることが見込まれる
      - 現状:GPUを利用した場合に実時間の 1/20 程度の計算速度
      - サイエンス:地球構造モデルや震源モデルの逆推定の高速化、高分解能化
      - 防災面:地震波計算によるリアルタイム地震動警報の可能性