

高性能計算システム研究部門

筑波大学 計算科学研究センター
高性能計算システム研究部門主任
朴 泰祐



高性能計算システム研究部門の概要

- 計算科学を支える超高性能／超大規模／超大容量計算システムの研究開発
 - アーキテクチャ：プロセッサ、ネットワーク、システム
 - ソフトウェア：システム、コンパイラ、ライブラリ
 - アルゴリズム：並列アルゴリズム、数値解析
- 各種計算科学シミュレーションコードの開発／性能向上／大規模化に対する支援
 - 素粒子、宇宙、物性、物質生命、地球環境、進化等のCCSのあらゆるアプリケーション分野との共同研究
 - アルゴリズムレベル、ライブラリレベル、コーディング等における支援
- 外部機関との連携、共同研究推進
 - 次世代スパコン開発
 - e-Scienceプロジェクト、JST-CREST等



CCSにおける高性能計算システム研究開発

- 単なる計算機工学的発想からのシステム設計・開発ではなく、アプリケーションを見据えた研究開発
 - アプリケーションのニーズとテクノロジーのバランス
 - システム規模、コスト、性能のバランス
 - 計算量や時間のオーダー見積もりに基づくシステム設計
 - アルゴリズムレベルからのシステム設計
- 各種要素技術の研究から実用的研究まで
 - 競争的資金に基づく要素技術研究
 - 概算要求に基づく実応用研究
 - レンタルスパコン調達におけるシステム設計
- あらゆるレベルで、CCSの特色（アプリ+システム=実用的高性能システム）を活かす



高性能計算システム研究部門スタッフ

■ Faculty staff

- 朴泰祐（教授、部門主任）
- 佐藤三久（教授、センター長）
- 建部修見（准教授）
- 高橋大介（准教授）
- 多田野寛人（助教）
- 塙敏博（特任准教授）

■ システム情報工学系CS専攻内に研究室



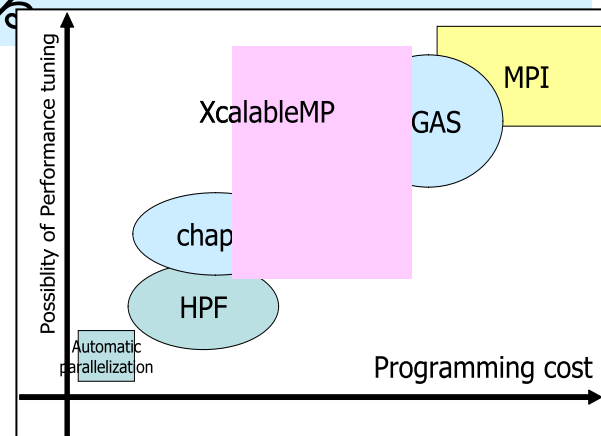
目標: 既存言語を指示文により拡張し、これからの大規模並列システム(分散メモリシステムと共有メモリノード)でのプログラミングを助け、生産性を向上させる並列プログラミング言語を設計・開発する

■ 現状と課題

- 並列プログラムの大半はMPI通信ライブラリによるプログラミング
 - 生産性が悪く、並列化のためのコストが高い。
- 並列プログラミングの教育のための簡便で標準的な言語がない (MPIでの教育にとどまっている)

● 並列言語 XcalableMPの策定・開発

- ✓ XcalableMP仕様策定ワーキンググループを組織し、言語仕様を策定
- ✓ ベース言語(C, Fortran95)を指示文で並列拡張
- ✓ グローバルビューでは、データ並列プログラミングモデルとワークシェアによって、典型的な並列化をサポート
- ✓ 通信モデルを明確化して、performance-awareな並列化を促進
- ✓ ローカルビューとしてPGAS (Partitioned Global Address Space)を提供



```
int array[YMAX][XMAX];
```

```
#pragma xmp nodes p(4)  
#pragma xmp template t(YMAX)  
#pragma xmp distribute t(block) on p  
#pragma xmp align array[i][*] to t(i)
```

data distribution

```
main(){  
  int i, j, res;  
  res = 0;
```

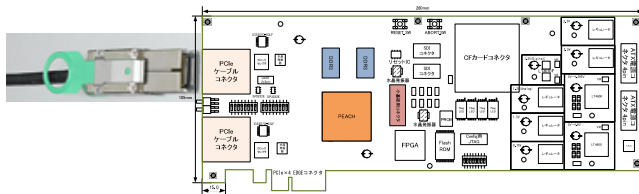
add to the serial code : incremental parallelization

```
#pragma xmp loop on t(i) reduction(+:res)  
  for(i = 0; i < 10; i++){  
    for(j = 0; j < 10; j++){  
      array[i][j] = func(i, j);  
      res += array[i][j];  
    }  
  }  
}
```

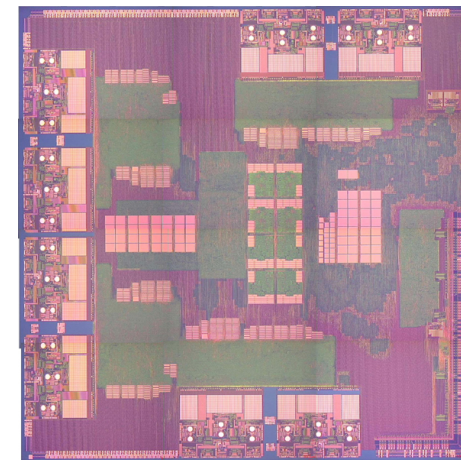
work sharing and data synchronization

高性能／低電力／高信頼通信リンク

- 通信リンク PEARL(PCI Express Adaptive and Reliable Link),
コミュニケーター PEACH(PCI Express Adaptive Communication Hub)
 - ルネサステクノロジ社と共同研究
 - PCIe Gen2 x4レーン (最大20Gbps)×4ポート, 4コアM32Rプロセッサ, パケットルータ, DDR3 I/Fを内蔵
 - PEACHチップ設計完了, 2010年2月末テープアウト
 - PEACH搭載PCIeボード設計完了, 子基板と合わせて3ポートのPCIe外部出力ポート
- 軽量コア間通信API MCAPI (Multicore Communications API)を
拡張した XMCAPIの提案
 - TCP/IPを用いた通信の実現



PEACH test PCI-E board

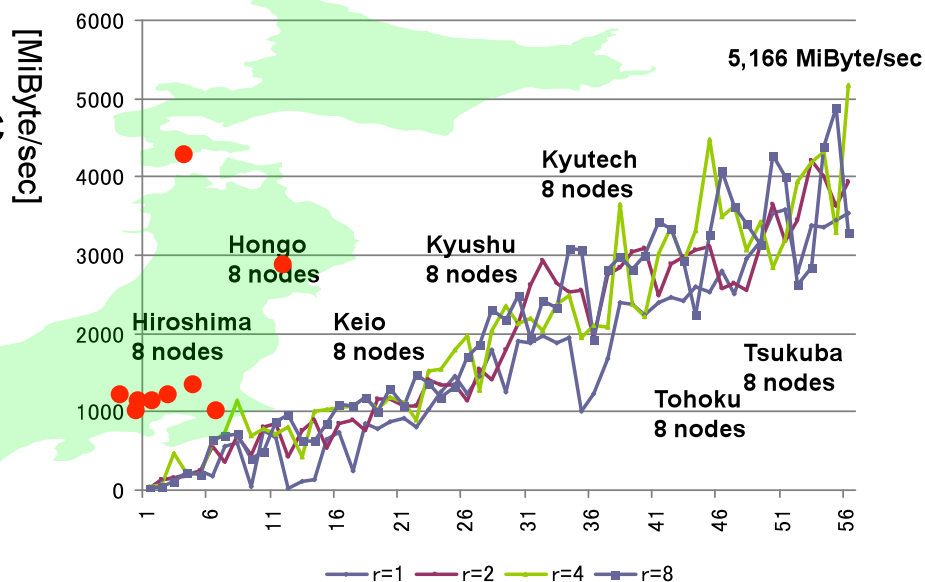
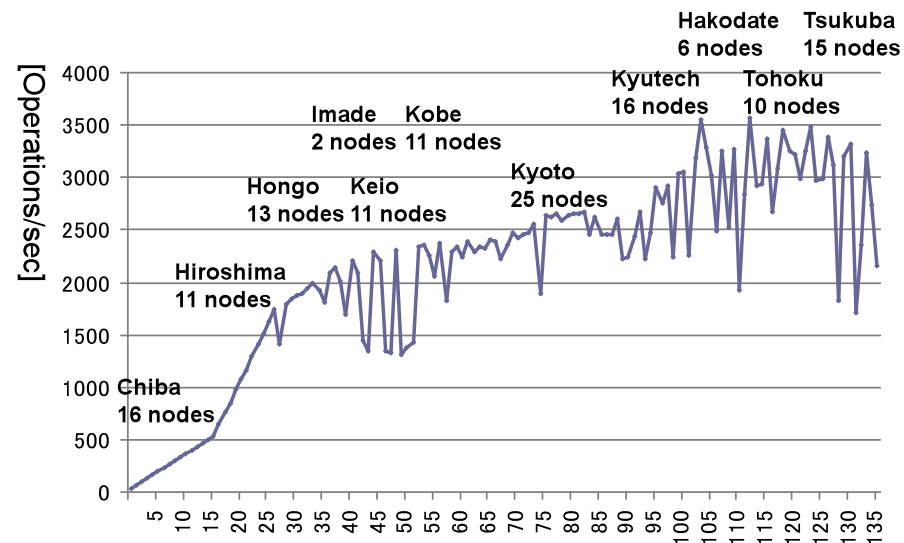
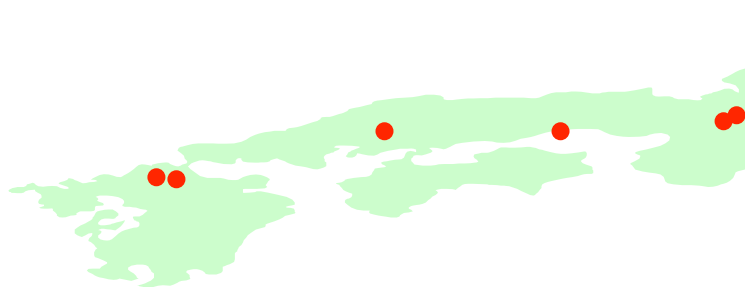


PEACH prototype chip

大規模環境における広域ファイルシステムの評価

■ Gfarm広域ファイルシステム

- オープンソースで開発
<http://sf.net/projects/gfarm/>
- InTrigger上で広域ファイルシステムの評価
 - 147 TByte (239ノード, 国内13拠点, 函館~福岡)
 - メタデータ操作3,570 ops/sec
 - 共有データ読込5,133 MiB/sec



PCクラスタ向け高性能FFT

- T2K筑波システムによる円周率2兆5769億8700万桁計算
 - 640ノードを用いて主計算が29時間5分，検証計算が44時間30分で完了した
 - その間一度もシステムダウンはなく，T2K筑波システムの信頼性を実証することができた
 - 現在，論文を投稿中
- 次世代スーパーコンピュータに向けた高速フーリエ変換（FFT）のアルゴリズムに関する研究
 - 並列三次元FFTにおいて，二次元分割により通信時間を削減することで，MPIプロセス数が多い場合に性能を改善した
 - T2K筑波システムの4,096コアを用いて性能評価を行った結果， 256^3 点FFTにおいて401GFlopsを超える性能が得られた

Block Krylov 部分空間反復法のBiCGGR法への応用

平成21年度の研究成果

連立一次方程式

$$AX = B, A \in \mathbb{C}^{n \times n}, X, B \in \mathbb{C}^{n \times L}$$



Block Krylov 部分空間反復法

Block Krylov の特長

長所：反復回数が減少する
 短所： L が増えると収束性が悪化する
 近似解の精度が悪化する

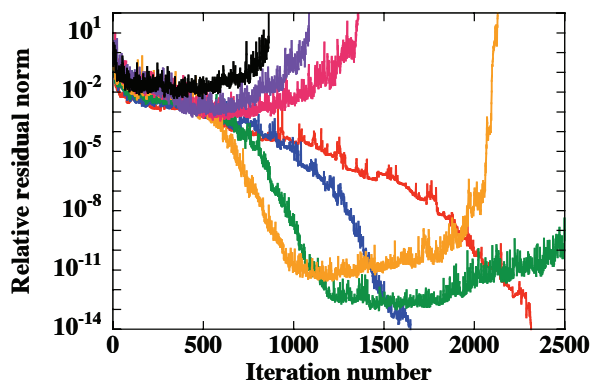


Block BiCGGR法を開発
 高精度近似解が生成可能！

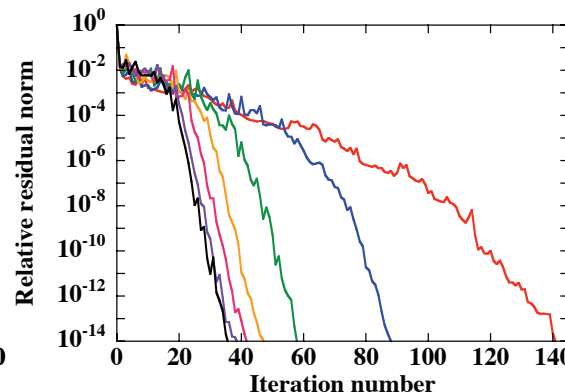
Jacobi前処理つきBlock BiCGGR法の1本あたりの計算時間. 但し, J : Jacobi法の反復回数.

L	$J = 0$	$J = 12$	$J = 24$	$J = 30$
1	1435.5	1599.3	1721.3	1753.0
2	721.8	642.3	689.6	694.1
4	—	304.9	322.1	328.6
6	—	216.3	224.0	231.7
8	—	178.2	185.5	190.9
10	—	174.8	166.3	169.5
12	—	146.8	151.5	151.1

1本あたりの求解時間が減少！



(a) 前処理無し.



(b) Jacobi法($J = 30$)による前処理.

Block BiCGGR 法の中で
 Jacobi 法を用いている



Jacobi 前処理により
 L が大きい場合も残差
 が収束するようになった

格子 QCD における Block BiCGGR法の相対残差履歴.

■ : $L = 1$, ■ : $L = 2$, ■ : $L = 4$, ■ : $L = 6$, ■ : $L = 8$, ■ : $L = 10$, ■ : $L = 12$.

高性能計算システム研究部門の今後

■ H22年度計画

- XcalableMPコンパイラのリリースとHPCC class2でのaward
- PEACH/PEARLによる省電力ディペンダブルクラスタの実装と性能機能評価、通信レイヤMCAPI実装
- multi-core/multi-railクラスタ向け自動プロファイリング+core数最適化システムの実装と評価
- 大規模広域分散解析の性能評価、JLDGにおける自動複製配置機能
- GPUによる大規模科学技術計算高速化
- 高精度近似解生成と高収束性を併せ持つblock Krylov部分空間反復法の開発



今後の研究活動

- CCSの応用分野研究との密接な連携によるアプリケーションの高性能化・大規模化支援、特に次世代スパコン向けの素粒子・物性・宇宙等のPFLOPS級アプリケーションの完成
- GPUに代表されるアクセラレータ技術等を含めたpost PACS-CSシステムの開発と実応用プログラムへの適用
- 省電力技術・高信頼技術等、超並列処理向け言語等、今後の大規模高性能システムの要素技術に関する研究開発
- グリッド/クラウドコンピューティング等の分散リソース利用に関する研究の拡充

