15-NA24

遠藤敏夫（東京工業大学）

超大規模シミュレーションのためのアーキテクチャ特性を考慮した通信削減技術

副代表者：中島研吾(東大)
共同研究者：松岡聡、額田彰、長坂侑亮(東工大)、片桐孝洋、大島聡史(東大)、岩下武史(北大)

将来のスパコンアーキテクチャ上では、メモリウォール問題のさらなる悪化が予想され、様々なシミュレーションのさらなる大規模化と高速化の両立の妨げになる。

本プロジェクトの目的：
アルゴリズムの特性を考慮した、データ通信（ノード間・ノード内メモリ階層間）の削減技術の研究・異種アーキテクチャ上での評価
- **研究項目(A)** 疎行列を主な対象とした、**メモリレイアウト・通信の最適化**
- **研究項目(B)** ステンシルを主な対象とした、**アルゴリズム局所性向上**

疎行列計算向け メモリレイアウト・通信の最適化

大規模並列多重格子法の通信最適化

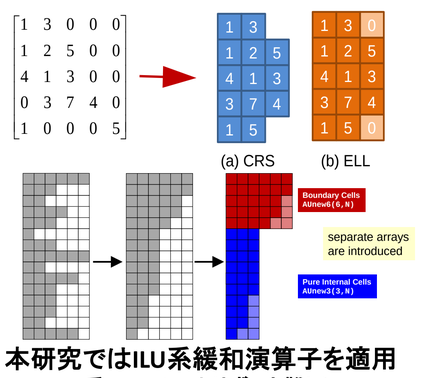
多重格子法（Multigrid）は大規模計算向けスケーラブルな解法であり，次世代ベンチマークHPCGにも採用されている。ノード内・ノード間双方の最適化を行う。

Serial通信改善

ELL (Ellpack-Itpack), Sliced-ELL（S-ELL）

X-ELL-Y-Z：ELLの多くの亜種

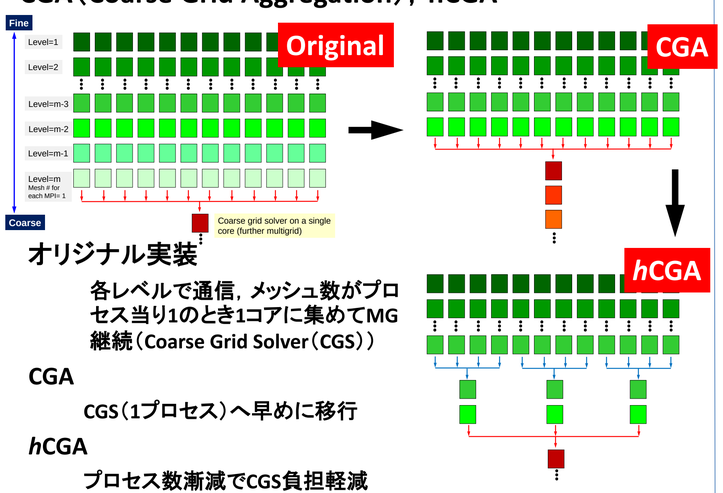
- **行列ベクトル積（SpMV）に注目**



本研究ではILU系緩和演算子を適用
GS系smootherよりずっと難しい

Parallel通信改善

CGA（Coarse Grid Aggregation），hCGA

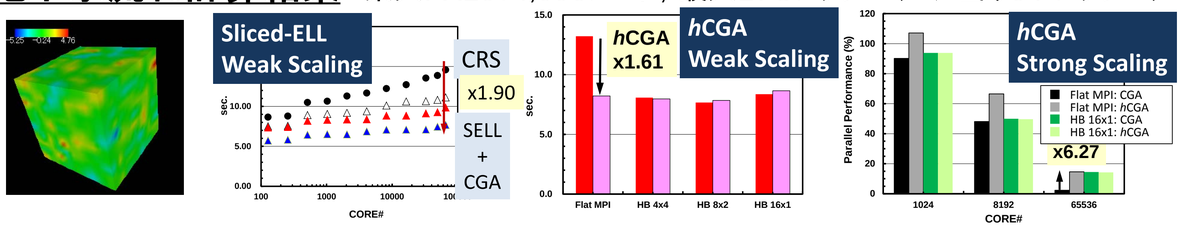


オリジナル実装
各レベルで通信，メッシュ数がプロセス当り1のとき1コアに集めてMG継続（Coarse Grid Solver（CGS））

CGA
CGS（1プロセス）へ早めに移行

***h*CGA**
プロセス数漸減でCGS負担軽減

地下水流れ計算結果 東大FX10：4,096ノード，最大17Gメッシュ（コア当り64^3メッシュ）

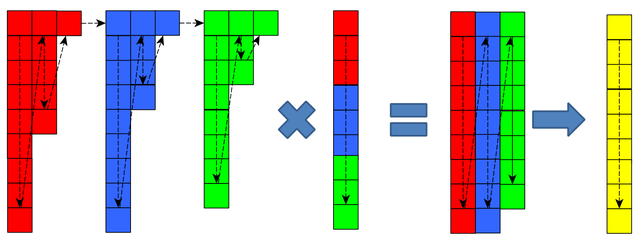


NUS疎行列フォーマットによるメモリ最適化

疎行列演算の重要カーネルである疎行列ベクトル積計算（SpMV）の性能はデータフォーマットの選択に大きく影響される。さらに入力ベクトル要素アクセス時のキャッシュヒット率を改善するため、NUSフォーマットを提案している。

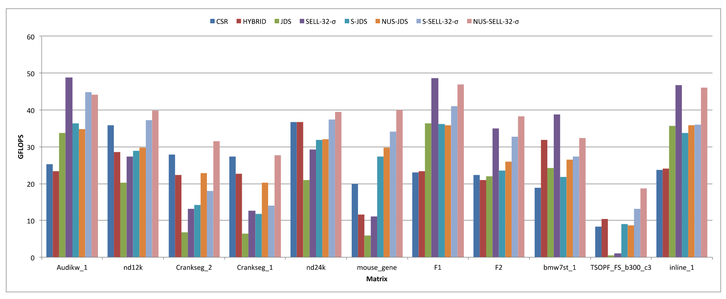
提案手法 [Nagasaka, ICPADS2014]

◯Segmented Formats
キャッシュヒット率改善のために行列を列方向に均等分割
- 各セグメントはJDSもしくはSELL-C-σに変換
- セグメントごとに演算を行い，合算



◯Non-Uniformly Segmented (NUS) Formats
リオーダリングによって再利用性が高いベクトル要素を抽出し，部分的な分割を行う
➔ 分割によって生じるオーバーヘッドを削減

東工大TSUBAME Tesla K20X GPUでの性能評価（単精度）



今後の方向性
東北大学のNEC社ベクトルマシンSX-ACEなど、異種アーキテクチャでの評価
- コア毎Assignable Data Buffer（ADB）の効率利用など

[Nagasaka, ICPADS2014] Yusuke Nagasaka, Akira Nukada, Satoshi Matsuoka, “Cache-aware Sparse Matrix Formats for Kepler GPU”, International Conference on Parallel and Distributed Systems (ICPADS 2014)

ステンシル計算向け アルゴリズム局所性向上

ステンシル計算の一般的な実装は局所性の利用が不十分である。
これを改善するため、時間(時空間)ブロッキング(TB)という手法が知られている：一部領域を取り上げたら、複数のタイムステップを一時に更新 → 局所性向上

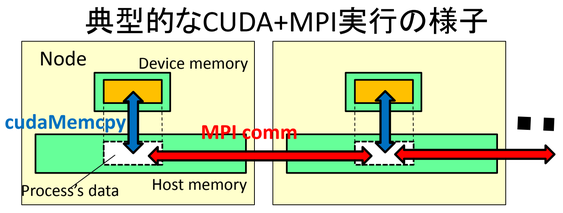
異種メモリを活用する時間ブロッキングと応用

GPUなどにより高速ステンシル計算が可能であるが、小さいデバイスメモリ容量が大規模化を妨げる。ホストメモリの大容量を利用しつつ、時間ブロッキングにより性能を維持し、さらに開発中のランタイム利用により生産性を維持する。

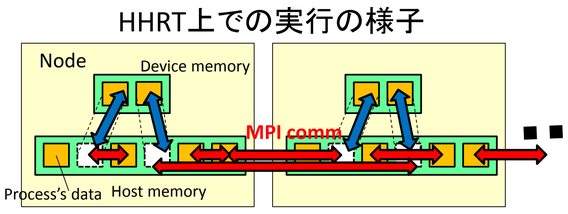
メモリ階層利用ランタイムHHRT

ユーザプログラムに対してほぼ透過的に、プロセス単位のスワップ機構を提供

典型的なCUDA+MPI実行の様子

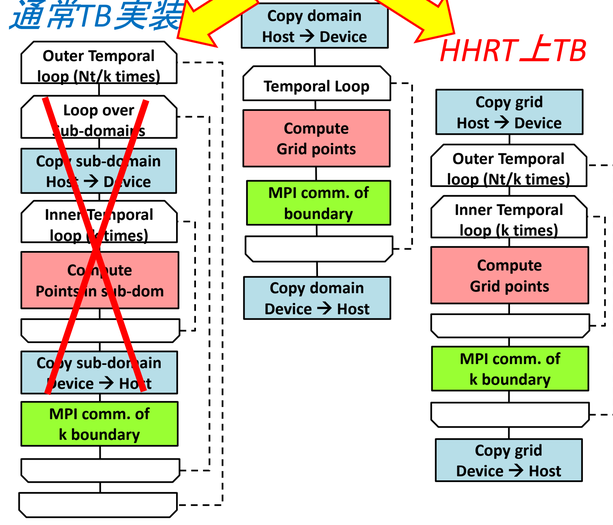


HHRT上での実行の様子



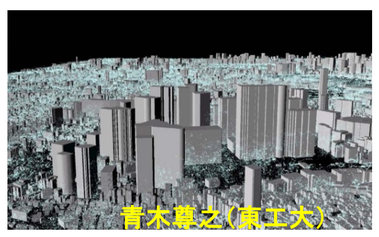
1GPUが複数MPIプロセスに共有され実行

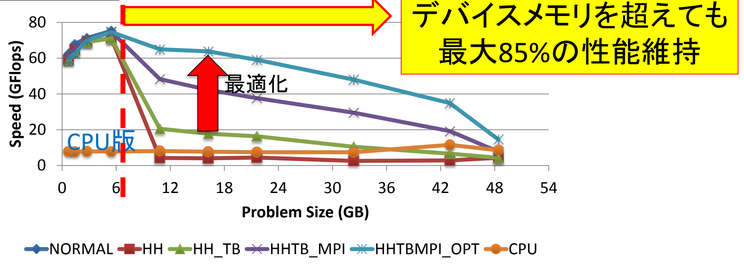
HHRT上の時間ブロッキング実装



Endo, Jin: Software Technologies Coping with Memory Hierarchy of GPGPU Clusters for Stencil Computations, IEEE Cluster 2014

都市気流シミュレーションへの適用 東工大TSUBAME2.5, K20X GPU



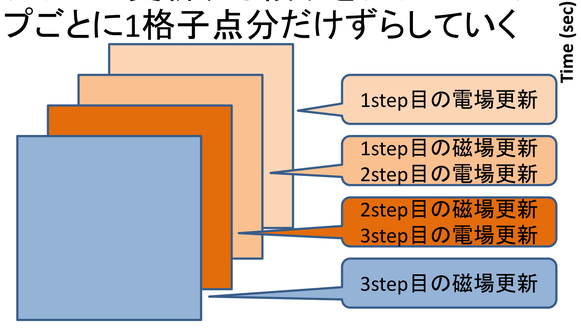


3次元FDTD法における時空間ブロッキング

FDTD法(Finite Difference Time Domain 法)は高周波電磁場解析の標準解法の一つであり、アンテナや電子デバイスの設計に広く用いられる。
時空間ブロッキングの導入により、時間局所性を高めキャッシュヒット率を向上

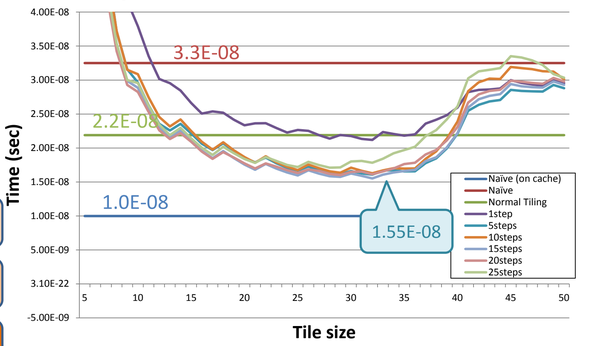
冗長計算を伴わない時空間ブロッキング

Skewアルゴリズムを3次元FDTD法で利用可能であることを証明
タイルの更新する領域をタイムステップごとに1格子点分だけずらしていく



時空間タイリングの効果

T2K-Kyodai, 4 thread, 1socketによる実行



Naïveな実装手法に対して計算時間は半分以下に