

09a-15, 09b-10

# 次世代スーパーコンピュータに向けた グランドチャレンジ・アプリケーション の開発

高橋大介

筑波大学大学院システム情報工学研究科

# メンバー

- 筑波大
  - 超高速計算システム分野: 佐藤, 朴, 高橋, 辻
  - 素粒子分野: 宇川, 吉江, 蔵増
  - 物質科学分野: 岩田
  - 数値解析分野: 櫻井, 多田野
- 東大: 物質科学分野: 押山, 古家, 小泉
- 広島大: 素粒子分野: 石川
- 兵庫県立大: 物質科学分野: 重田
- 理化学研究所  
次世代スーパーコンピュータ開発実施本部
  - システム開発チーム: 横川, 庄司
  - アプリケーション開発チーム: 南, 杉原, 黒田, 井上, 長谷川
  - 富士通, NEC情報システムズ, NECソフト, 高度情報科学技術研究機構: 井上, 浅見, 守田, 飯塚

# 目的

- 理化学研究所と筑波大学との間で、「大規模シミュレーションによる次世代スーパーコンピュータの性能評価に関する共同研究」を行っている。(H19.9~H23.3)
- 今後利用可能になる次世代スーパーコンピュータに向けたグランドチャレンジ・アプリケーションの開発および性能評価を行う。
- グランドチャレンジ・アプリケーションとしては、
  - 筑波大
    - QCD(素粒子分野)
    - RSDFT(物質科学分野)
    - FFTなど(数値計算ライブラリ)
  - 理研
    - ライフサイエンス, ナノサイエンス, 地球科学などを想定している。

# 平成21年度前期・後期の成果

- 実空間密度汎関数法 (RSDFT) による大規模第一原理計算
- 地球科学関連コード
- 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
- 次世代スーパーコンピュータに向けた全対全通信アルゴリズム

# 実空間第一原理計算コード

- 実空間の三次元格子を分割する空間分割とバンド方向の分割を実装した空間＋バンド並列版の性能測定を実施した。
- 実行コストのほとんどを占める
  - 規格直交化ルーチン
  - 部分対角化ルーチン
  - 共役勾配法ルーチン

について、演算部と通信部のスケーラビリティを調査した。

- スケーラビリティ悪化の要因として懸念される空間方向の通信コストを下げる効果によって全体処理時間が短縮されることを確認した。
  - 4096原子, 2048並列では, 空間並列のみの場合と比べて上記3ルーチンの全通信コストが70%から19%に削減され, 全体処理時間は1.87倍高速になった。
- OpenMP/MPIハイブリッド並列化とその性能評価も行った。

# 地球科学関連コード

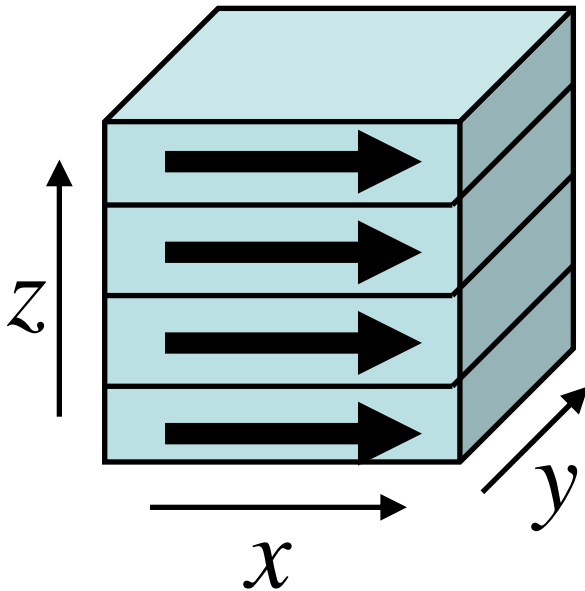
- 大気海洋結合モデルの大気部分の並列性能について、特にAMR (Adaptive Mesh Refinement) を用いた場合に、用いない場合に比べてどのような影響があるかについて並列数を変えた実測値を元に検討した。
- AMRなしの場合は、演算時間は良くスケールしていること、通信時間も演算時間に対して15%程度に収まっており、通信が大規模並列時のボトルネックにはならないものと予測されることが分かった。
- またAMRを用いる場合も、演算時間、通信時間の増加は演算量自体の増加でほぼ説明され、オーバーヘッドは問題にならないことが分かった。

# 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

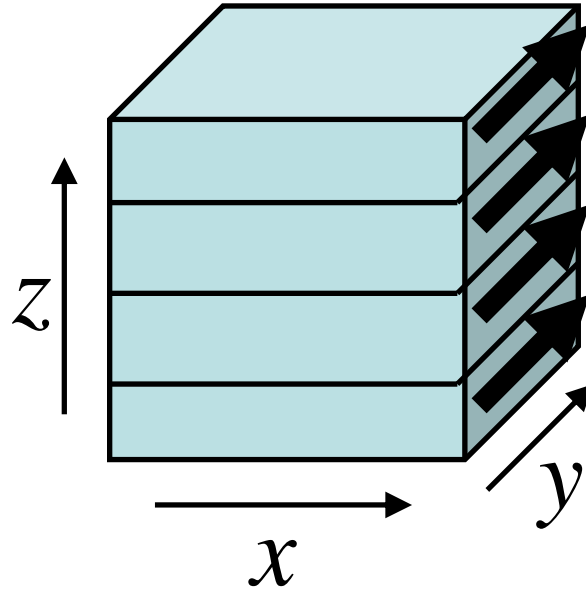
- ナノ分野グランドチャレンジアプリケーションの一つである3D-RISMにおいて、並列三次元FFTが律速となっている。
- 並列三次元FFTにおいて、二次元分割により通信時間を削減することで、MPIプロセス数が多い場合に性能を改善した。
- T2K筑波システムの4,096コアを用いて性能評価を行った結果、 $N=256^3$ 点FFTにおいて401GFlopsを超える性能が得られた。

# z方向に一次元ブロック分割した 場合の並列三次元FFT

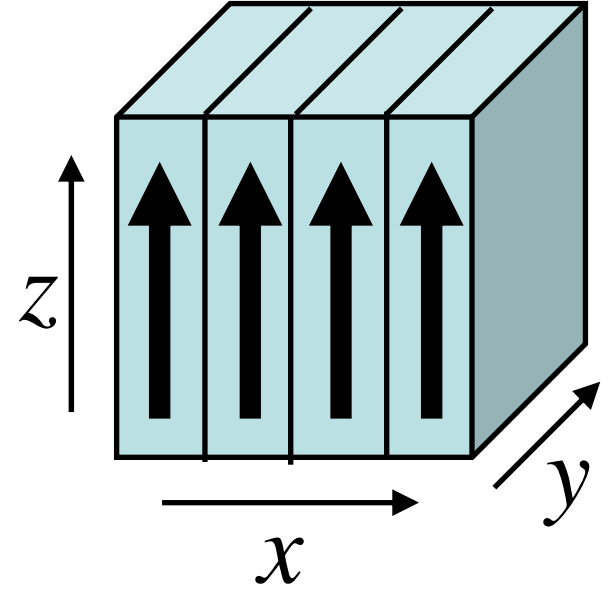
1. x方向FFT



2. y方向FFT



3. z方向FFT

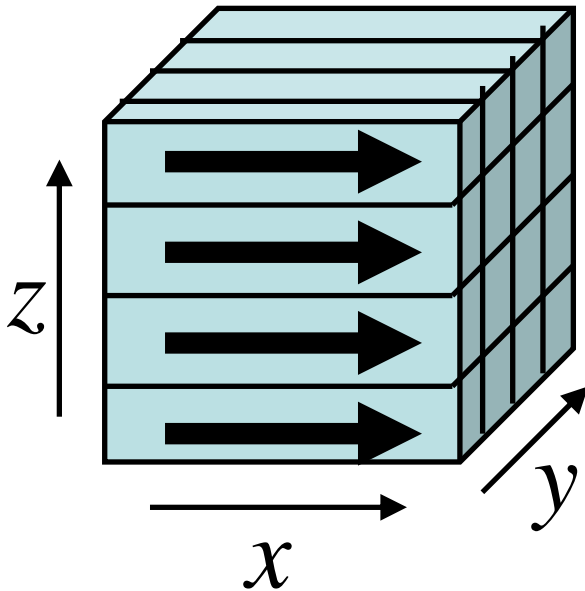


各プロセッサでslab形状に分割

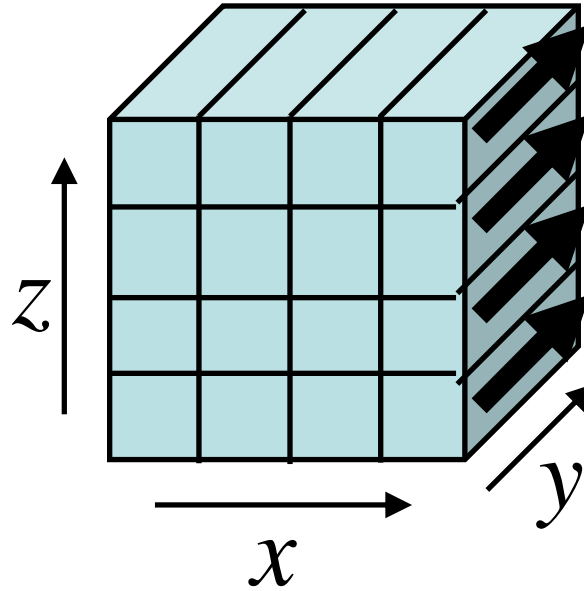


# y, z方向に二次元ブロック分割 した場合の並列三次元FFT

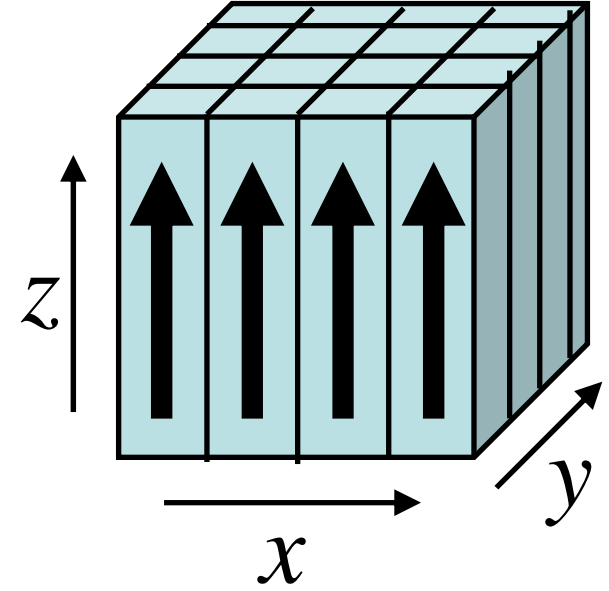
1. x方向FFT



2. y方向FFT

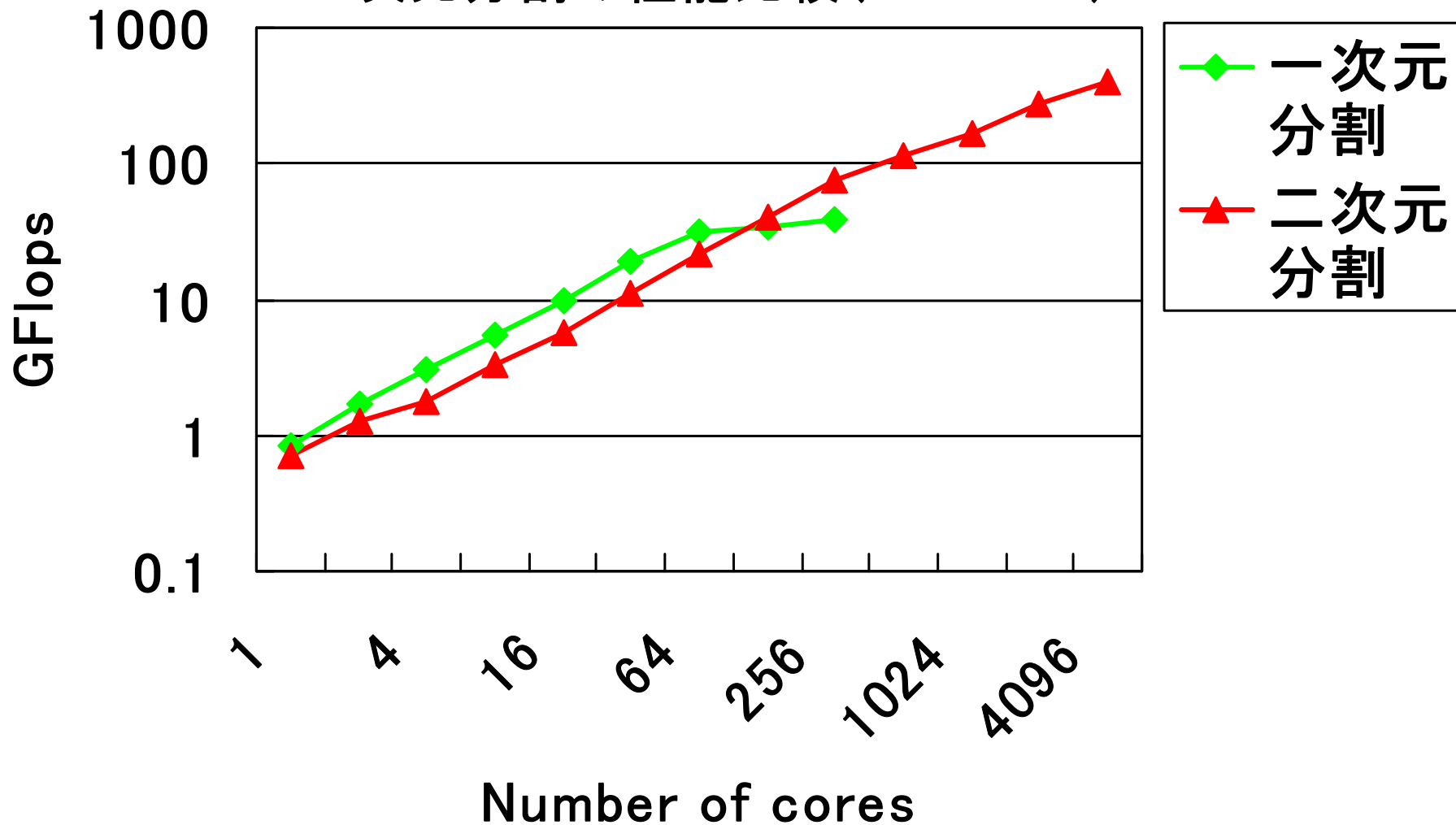


3. z方向FFT

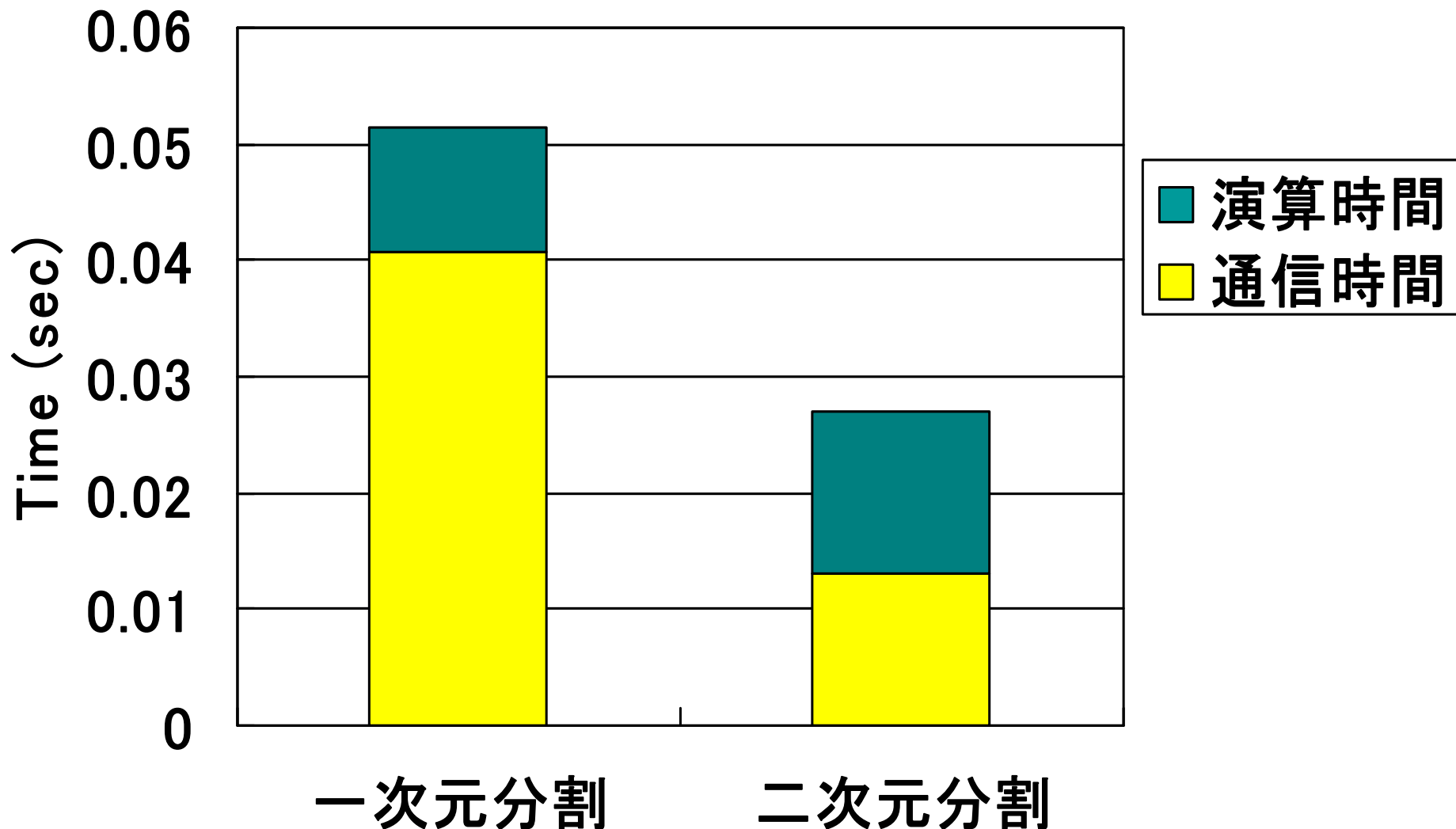


各プロセッサで直方体形状に分割

# 256<sup>3</sup>点FFTにおける一次元分割と 二次元分割の性能比較 (FFTE 4.1)



# 並列三次元FFTの実行時間の内訳 (FFTE 4.1, 256cores, 256<sup>3</sup>点FFT)



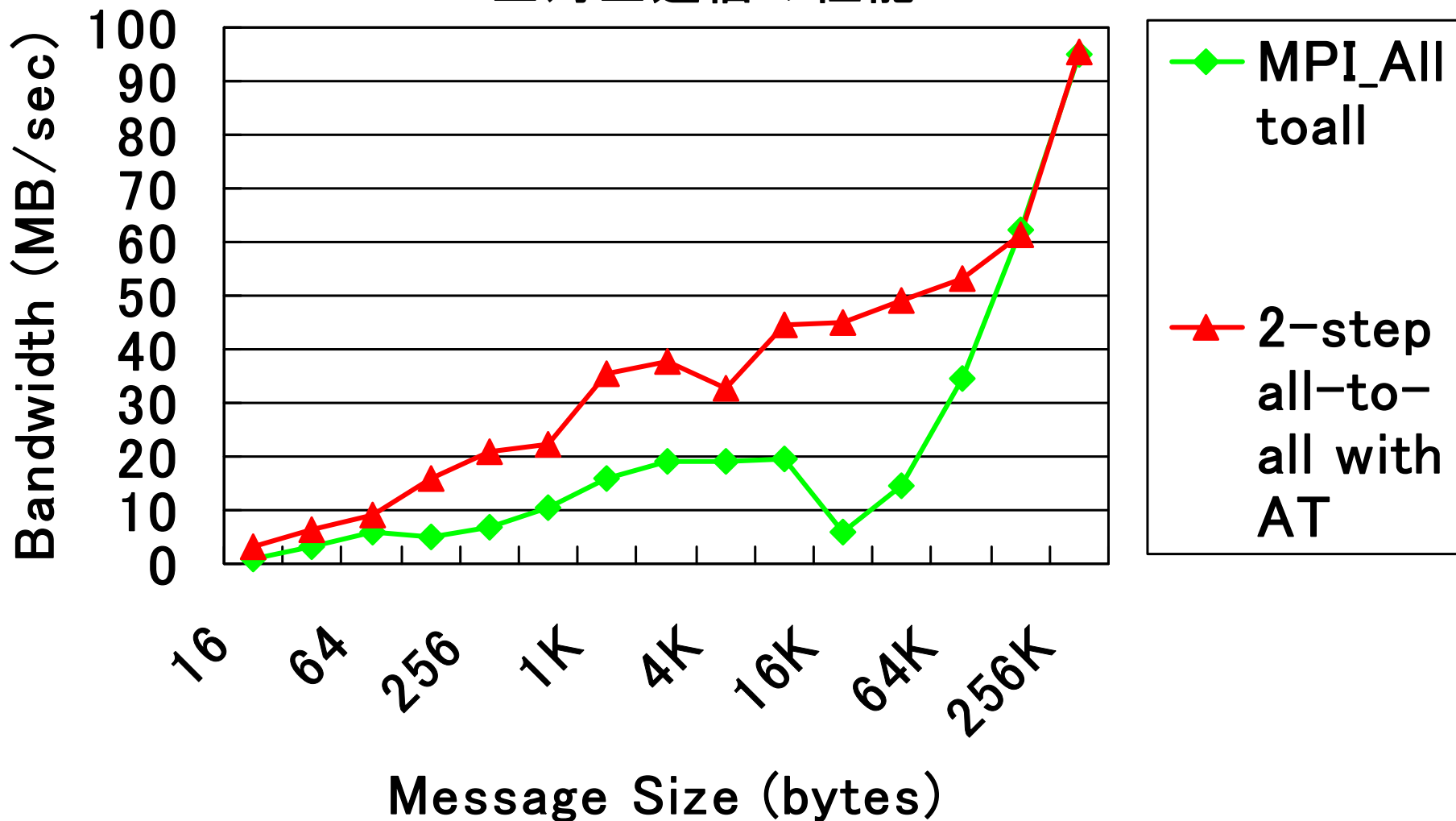
2010/5/7

第1回「学際計算科学による新たな  
知の発見・統合・創出」シンポジウム

# 次世代スーパーコンピュータに向けた 全対全通信アルゴリズムの検討

- 全対全通信は, FFTなどで広く用いられている集合通信.
- 全対全通信における, 通信の立ち上がり時間(レイテンシ)の影響はMPIプロセス数に比例する.
- MPIプロセス数が  $P \times Q$  と分解できる場合, 以下のように2段階に分けて全対全通信を行うことで, 転送量は2倍になるが, レイテンシの影響をMPIプロセス数の平方根のオーダーにできる
  - 縦方向の  $P$  プロセスで  $Q$  組の全対全通信
  - 横方向の  $Q$  プロセスで  $P$  組の全対全通信
- 最適な  $P$  と  $Q$  の組み合わせが存在する.

# T2K-Tsukuba(64ノード, 1024コア, flatMPI)における 全対全通信の性能



# まとめ

- 性能を引き出す上で最も重要であると考えられるのは、スケーラビリティ。
- これまで使われてきたアルゴリズムがもともと想定していない並列度が要求される。
- 数万ノードの超並列システムにおいては、全対全通信を伴うアルゴリズムは現実的でないと思われていたが、2段階全対全通信アルゴリズムを用いることで、高いスケーラビリティを維持できる見込みが見ついた。
- 今年度も学際共同利用に採択して頂いたので、引き続き研究を行っていく予定。