

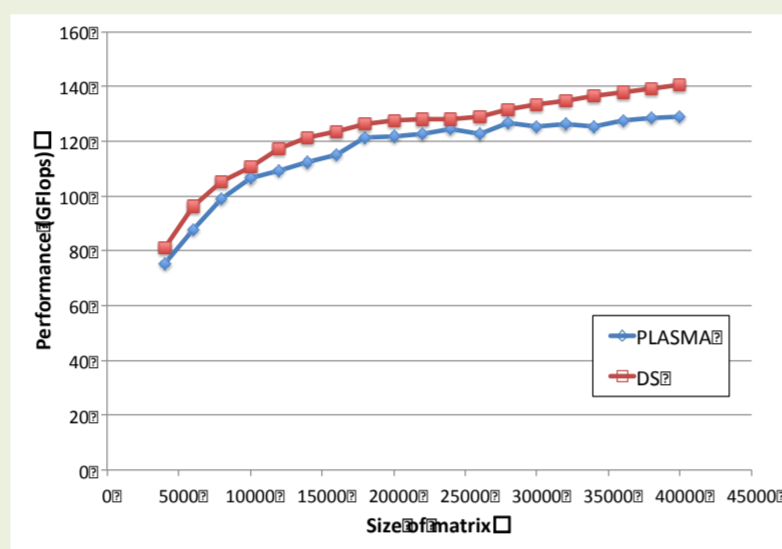
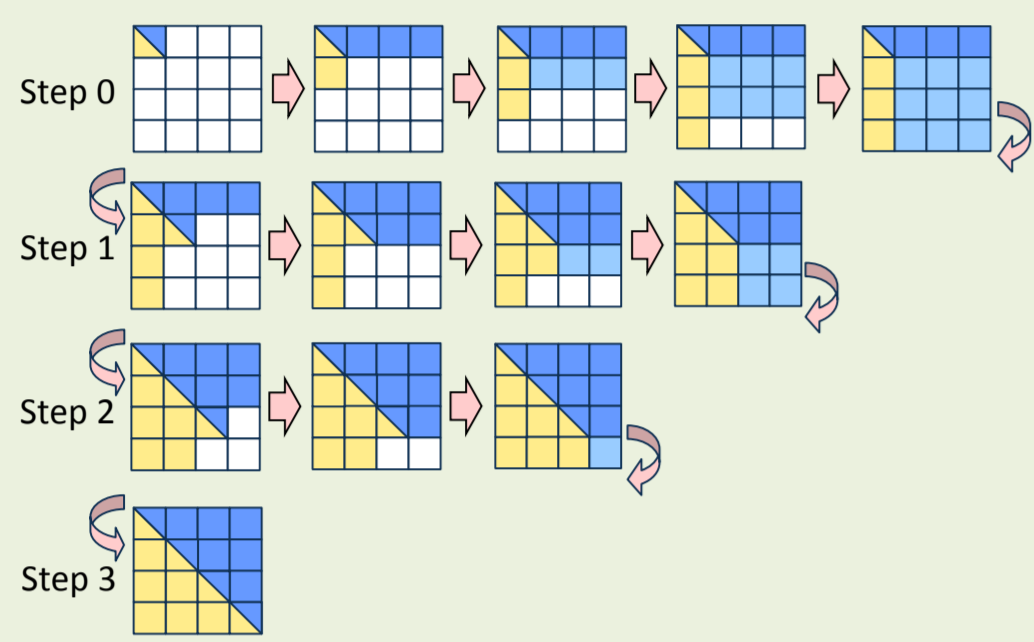
タイルアルゴリズムの大規模適用時の通信最適化



目的：大規模並列環境向け高速行列分解ルーチン

共有メモリ環境

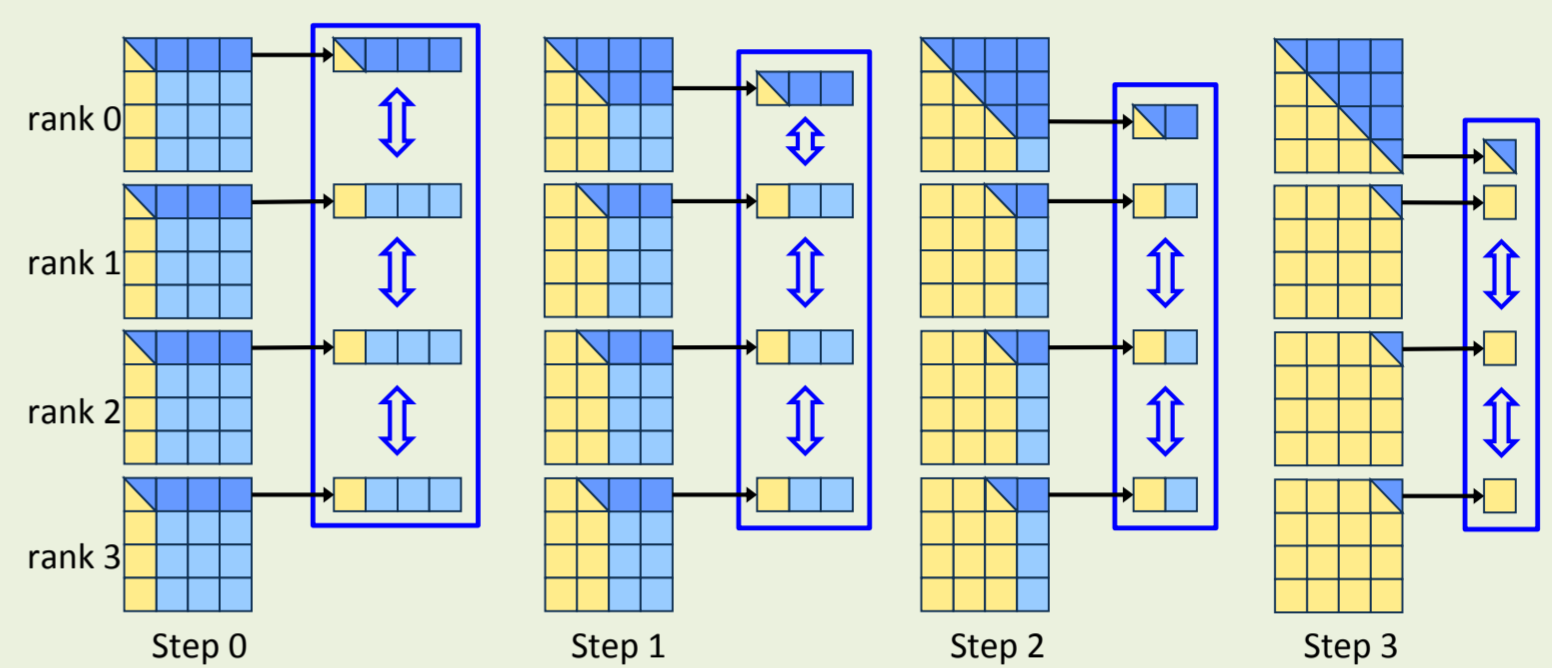
- **タイルアルゴリズム** → 行列を小行列（タイル）に分割
- 1 or 2タイル毎の処理 → 大量の細粒度タスク
- タスク並列プログラミングモデル
- 非同期タスク実行
- 共有メモリ環境においては十分に高速な実装が得られている



pdgeqrfとの比較 (FX10 1ノード)

分散メモリ環境

- 分散メモリ環境においては十分な性能を発揮する実装が得られていない
- タイルCAQR
- タスク間のデータ依存
- ノード間でタスクの実行状況を共有する必要あり
- 比較的小さいデータ量のノード間通信が多く発生



ノード間通信の最適化が必要

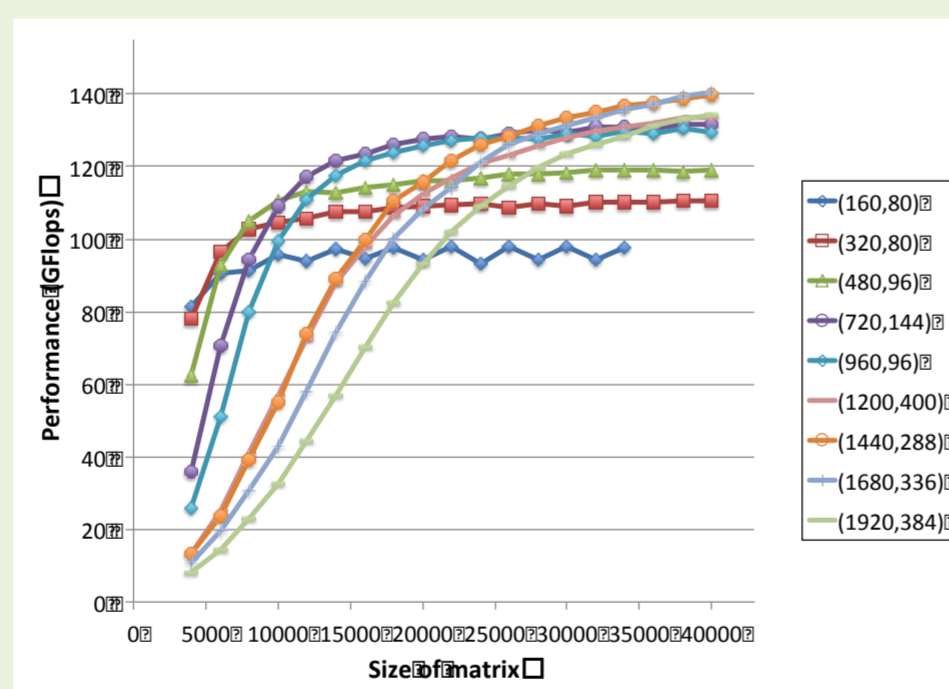
これまでの研究成果

タスクスケジューリング法の検討

- タスク非同期実行のための動的タスクスケジューリング
- タスクキュー&プログレステーブル
- task構文depend節 (OpenMP 4.0 以降)

パラメータチューニングにおける探索時間の削減

- 枝刈り探索
- タイルQRの主要タスク SSRFB のチューニング
- 最適パラメータ候補 (タイルサイズ、内部ブロック幅) の選出
- 候補についてのみQR分解を実行
- タスク数の指標を導入 → パラメータ候補数削減

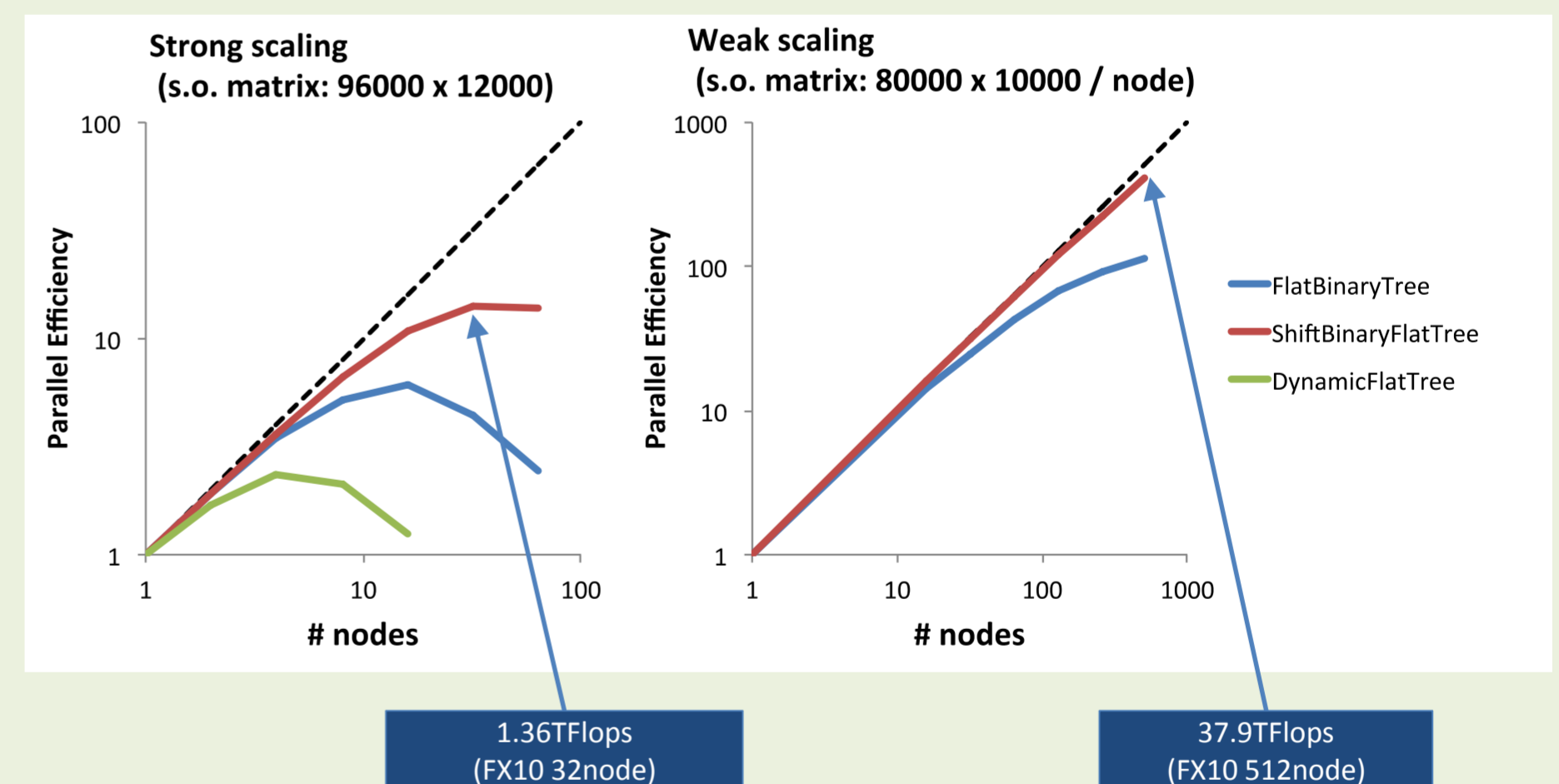


パラメータチューニング (FX10 1ノード)

現在のクラスタシステム向け実装

- MPI-OpenMPハイブリッド
- 1Dブロックデータ分散 (縦方向)
- ノード間通信は同期通信
- 負荷分散 (シフト)
- 1から32ノード程度の小規模な環境では強、弱両方の流儀でスケール

ノード間通信の違いによる性能差



1.36TFlops (FX10 32node)

37.9TFlops (FX10 512node)

今後の研究計画

クラスタシステム向け実装の性能評価 (つづき)

- FX10での性能評価
 - Strong scaling
- トレース、プロファイル取得
 - ノード間通信部分

分散メモリ環境におけるタイルサイズチューニング法の検討

- 実行スレッド数の指標
- ノード間通信のデータ量、回数の指標
- 性能モデルの構築

実装の改良

- データ分散の変更 (サイクリック分散)
 - (更なる) 負荷分散
 - 大規模データ対応
- ノード間通信削減アルゴリズム
 - ノード間通信のデータ量、回数を削減 ⇔ 全体計算量増加

