

強スケーリング型の 超並列固有値・特異値計算手法

電気通信大学 情報理工学研究科
山本有作

(共同研究者: Marian Vajtersic, Gabriel Oksa, Martin Becka, 工藤周平)

統計数理研究所 数学協働プログラム研究集会
「計算物質科学における時空アップスケーリングと数理手法」
2016年11月29日

研究背景(1)

- ▶ 実対称行列 $A \in \mathbb{R}^{n \times n}$ の固有値問題

$$A = P \Lambda P^T$$

($P \in \mathbb{R}^{n \times n}$: 直交行列, $\Lambda \in \mathbb{R}^{n \times n}$: 対角行列)

- 全固有値・固有ベクトルを求める
 - $n \sim 10,000$ 程度の **中規模** 行列が対象
-
- ▶ 応用
 - 電子状態計算
 - 分子軌道法
 - 超大規模固有値問題の部分問題
 - フィルタ対角化法やCIRR (Contour Integral Rayleigh-Ritz) 法では, 射影により超大規模問題を小規模～中規模問題に帰着

研究背景(2)

- ▶ 実行列 $B \in \mathbb{R}^{n \times n}$ の特異値分解

$$B = U \Sigma V^T$$

($U, V \in \mathbb{R}^{n \times n}$: 直交行列, $\Sigma \in \mathbb{R}^{n \times n}$: 対角行列)

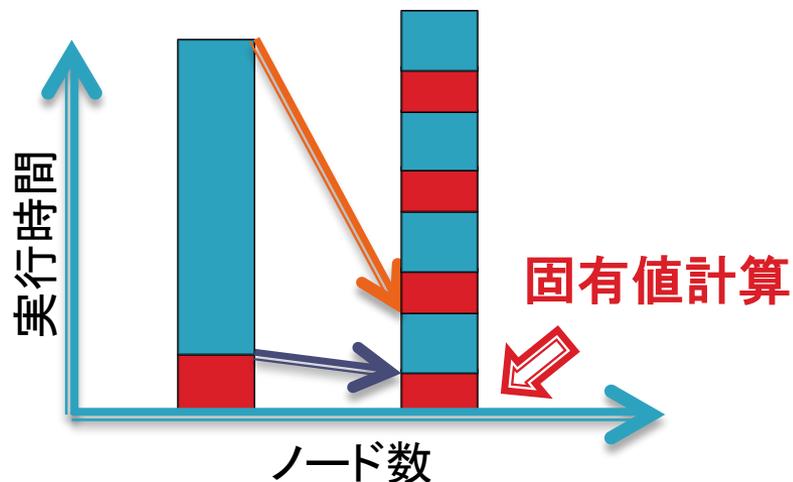
- 全特異値・特異ベクトルを求める
 - $n \sim 10,000$ 程度の**中規模**行列が対象
 - $C \in \mathbb{R}^{m \times n}$ ($m \geq n$) の特異値分解は, QR分解によりこの問題に帰着
 - $C = QR = (QU)\Sigma V^T$
-
- ▶ 応用
 - データ科学(最小二乗法, 情報圧縮, 情報検索, etc.)
 - 悪条件の行列の正則化
 - Proper Orthogonal Decomposition 法によるモデル縮減

中規模固有値問題の重要性

- ▶ 例：分子軌道法における固有値計算
 - サイズ： $n \sim 10,000$ 程度の需要大
 - 演算量で見ると，固有値計算の占める割合は小さい
 - 行列生成(多電子積分)： $O(n^4)$ 以上 (n : 基底数)
 - 固有値計算： $O(n^3)$
 - しかし，行列生成部分は一般に並列性が高く，**加速率大**

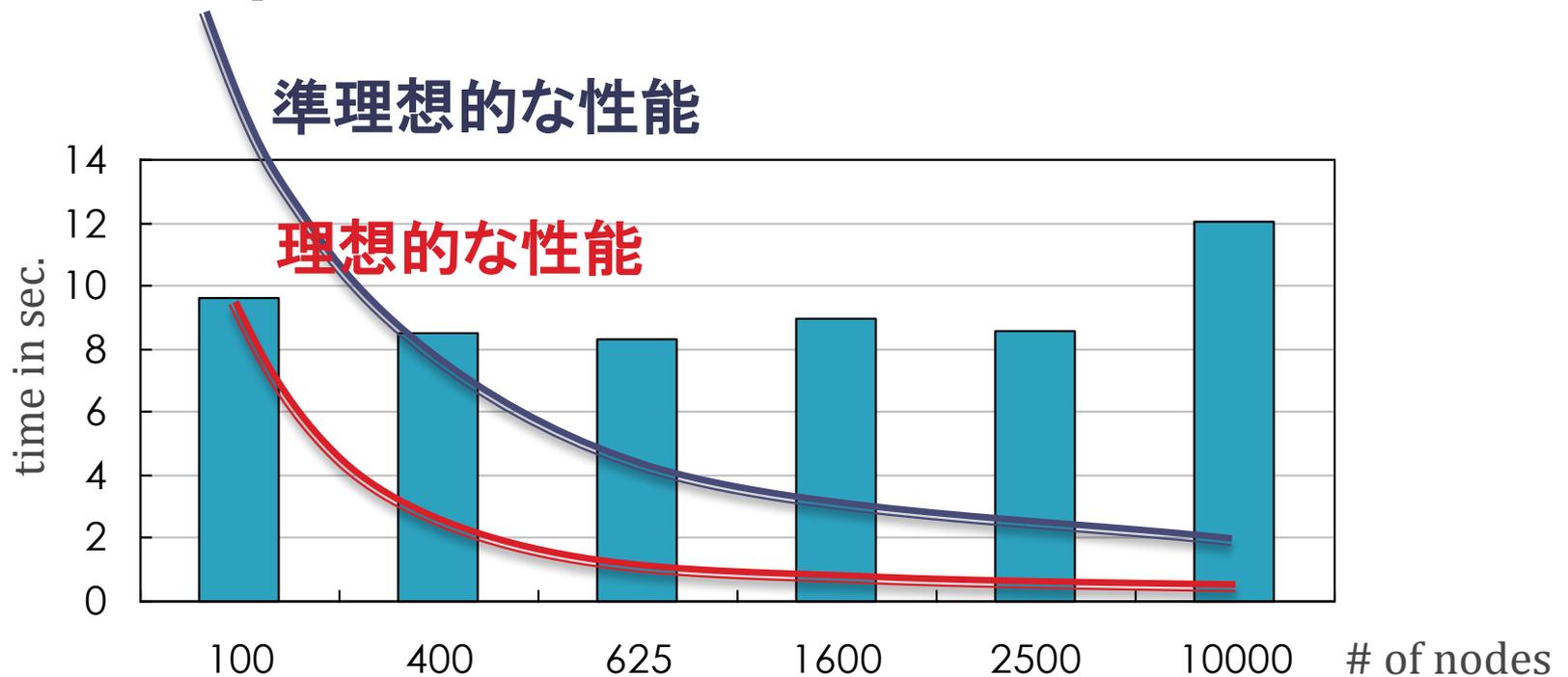


- 超並列環境では，固有値計算が並列化の**性能ネック**に
- 繰り返し計算では特に影響大
- **時間アップスケーリング**における重要課題



ScaLAPACK の性能

- ▶ 「京」での固有値計算の性能
 - 全固有値・固有ベクトルを計算 (PDSYEVD)
 - $n = 10000$, $p = 100 \sim 10000$ (強スケーリング)



ScaLAPACKでは、400ノードで性能が飽和

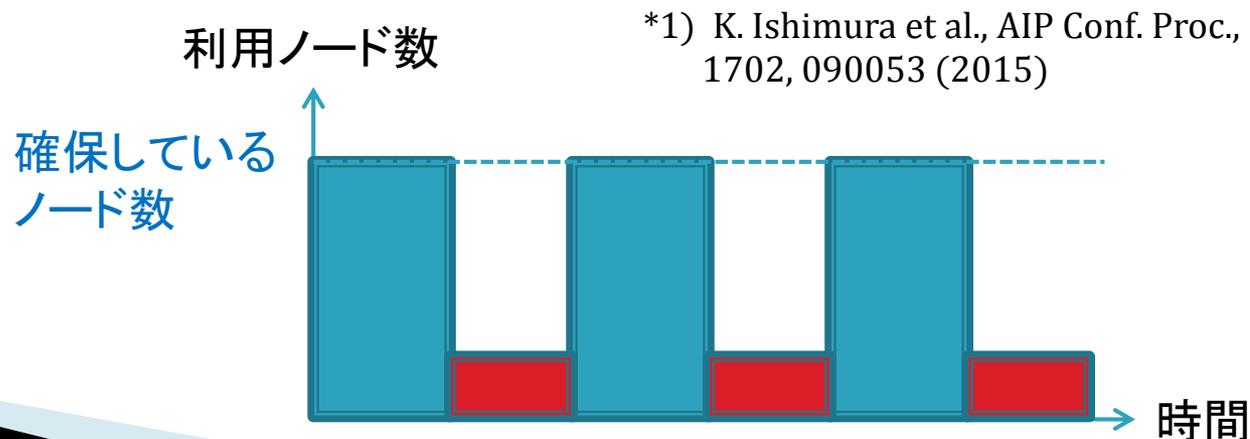
強スケーリング指向の固有値・特異値ソルバ

▶ 目標

- $n = 10,000$ 程度の中規模固有値・特異値問題をできるだけ高速に解く
- ノードはいくら使ってもよい

▶ 目標の妥当性：分子軌道法の場合

- 行列生成部分では、1万個以上のノードを使うことも*1)
- ScaLAPACKでの固有値計算時に、大部分のノードはアイドル状態
- 確保したノードをフルに利用することで、固有値計算の時間を短縮



ブロックヤコビ法に基づく強スケーリング型ソルバ

▶ ブロックヤコビ法の特徴

- 計算量は3重対角化／2重対角化に基づくソルバに比べて数倍多い
- 通信回数のオーダーが小さい
⇒ 強スケーリング環境に適する
- 様々な変種・最適化手法が存在
 - 片側(One-sided)ブロックヤコビ法 / 両側(Two-sided)ブロックヤコビ法
 - 前処理による収束性向上
 - ピボット戦略
- ある種の行列に対しては、高い相対精度で固有値・特異値を計算可能
 - 本発表では扱わない

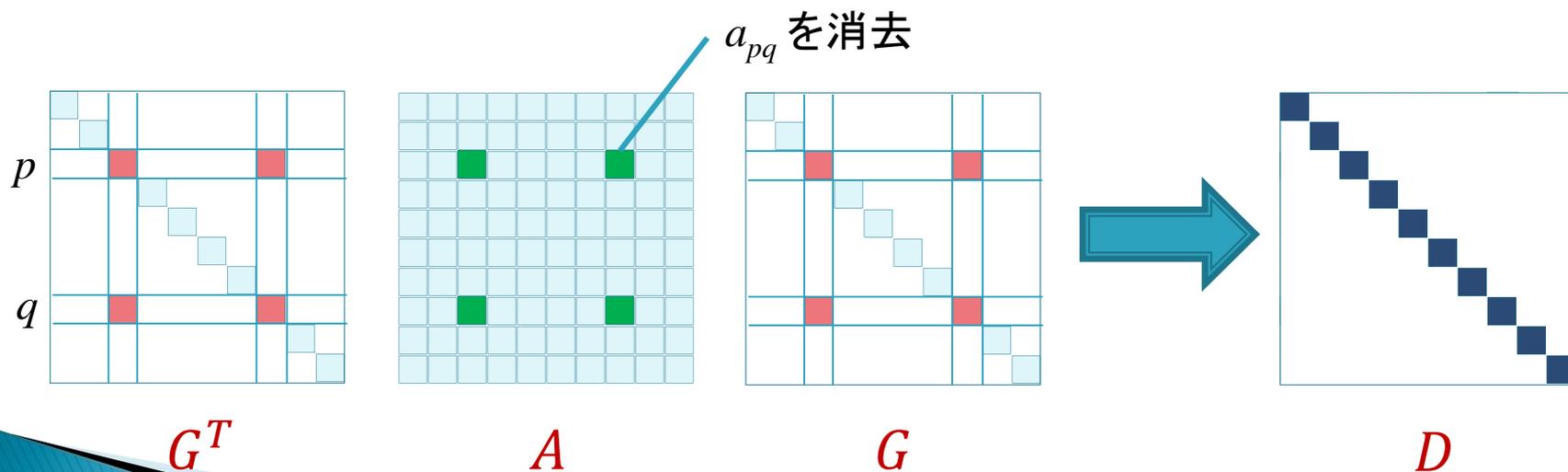
発表の概要

- ▶ 研究背景
- ▶ ブロックヤコビ法のアルゴリズム
- ▶ 実装と性能評価
- ▶ おわりに

ブロックヤコビ法のアルゴリズム

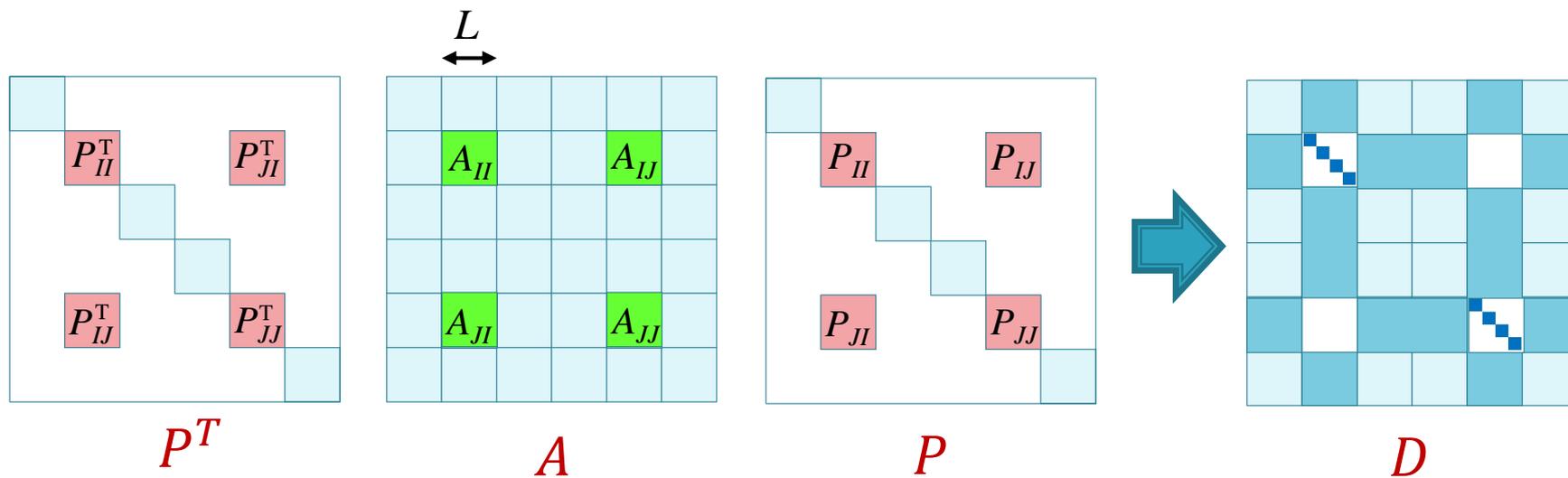
ヤコビ法による実対称行列の固有値計算

- ▶ すべての非対角要素が十分小さくなるまで、以下を繰り返す
 - 消去する要素 a_{pq} を決定
 - 第 i 行・第 j 行に作用するギブンス回転行列 G を決定
$$\cot 2\theta = (a_{qq} - a_{pp}) / (2 a_{pq}), \text{ ただし } -\pi/4 \leq \theta \leq \pi/4$$
 - 行列に左から G^T , 右から G をかける



ブロックヤコビ法

- ▶ n 次実対称行列 A を $L \times L$ のブロックに分割 (ブロック数 $W = n/L$)
- ▶ ブロックを要素と見て, 直交変換により非対角ブロック A_{IJ} を消去



- ▶ 直交変換 P の決定: $2L \times 2L$ 行列の対角化

$$\begin{array}{ccc}
 \begin{array}{|c|c|} \hline P_{II}^T & P_{JI}^T \\ \hline P_{IJ}^T & P_{JJ}^T \\ \hline \end{array} &
 \begin{array}{|c|c|} \hline A_{II} & A_{IJ} \\ \hline A_{JI} & A_{JJ} \\ \hline \end{array} &
 \begin{array}{|c|c|} \hline P_{II} & P_{IJ} \\ \hline P_{JI} & P_{JJ} \\ \hline \end{array} &
 = &
 \begin{array}{|c|c|} \hline D_{II} & \\ \hline & D_{JJ} \\ \hline \end{array} \\
 \tilde{P}^T & \tilde{A} & \tilde{P} & & \tilde{D}
 \end{array}$$

ブロックヤコビ法のアルゴリズム

▶ アルゴリズム

Algorithm 1 Block Jacobi method for the symmetric eigenvalue problem

- 1: $r = 0$; $A^{(0)} = A$
 - 2: Diagonalize $A_{II}^{(0)}$ for **for** $I = 1, 2, \dots, W$.
 - 3: **while** $w_{\max} > \epsilon$ **do**
 - 4: Choose the off-diagonal block $A_{IJ}^{(r)}$ to be eliminated.
 - 5: Diagonalize the $2L \times 2L$ submatrix $\tilde{A}^{(r)} = \begin{bmatrix} A_{II}^{(r)} & A_{IJ}^{(r)} \\ A_{JI}^{(r)} & A_{JJ}^{(r)} \end{bmatrix}$; $\tilde{P}^\top \tilde{A}^{(r)} \tilde{P} = \tilde{D}$.
 - 6: Update $A_{I,*}^{(r)}$ and $A_{J,*}^{(r)}$ by multiplying \tilde{P}^\top from left.
 - 7: Update $A_{*,I}^{(r)}$ and $A_{*,J}^{(r)}$ by multiplying \tilde{P} from right.
 - 8: $w_{\max} = \max_{I < J} \|A_{IJ}^{(r)}\|_F$.
 - 9: $r := r + 1$
 - 10: **end while**
-

▶ 特徴

- アルゴリズムが単純・均質 ⇒ ハードウェア実装向き
- 計算の大部分が行列乗算

ピボット戦略

- ▶ 各ステップにおいて消去すべき非対角ブロック $A_{IJ}^{(r)}$ としてどれを選ぶかという戦略をピボット戦略という
 - 対称性より $I < J$ としてよい
- ▶ ピボット戦略は、収束性、並列性、並列実行時の通信パターン等に大きな影響を与える
- ▶ 点ヤコビ法(ブロック版でないヤコビ法)においては様々なピボット戦略が提案されており、それらの大部分はブロックヤコビ法でも利用可能である
 - 古典的ヤコビ法 ⇒ 古典的ブロックヤコビ法
 - 巡回型ヤコビ法 ⇒ 巡回型ブロックヤコビ法

古典的ブロックヤコビ法

▶ 戦略

- 非対角ブロック $A_{IJ}^{(r)}$ としてフロベニウスノルムが最大のものを選ぶ

$$(I, J) = \operatorname{argmax}_{P < Q} \|A_{PQ}^{(r)}\|_F$$

▶ 特徴

- 収束が速い
 - 非対角ノルム $\sum_{I < J} \|A_{IJ}\|_F^2$ の減少分 = 消去したブロックの F ノルムの2乗
- 大域的収束性の証明が容易
- フロベニウスノルム最大の非対角ブロックを探す手間が掛かる
 - ブロックサイズが十分大きければ, この手間は無視できる

巡回型ブロックヤコビ法

▶ 戦略

- $W(W-1)/2$ 個の非対角ブロックを決まった順番で巡回的に消去する
- 順番の決め方により, $(W(W-1)/2-1)!$ 通り(円順列)の戦略がある
 - 行巡回型
 - 列巡回型
 - ラウンドロビン型
 - リング型

▶ 特徴

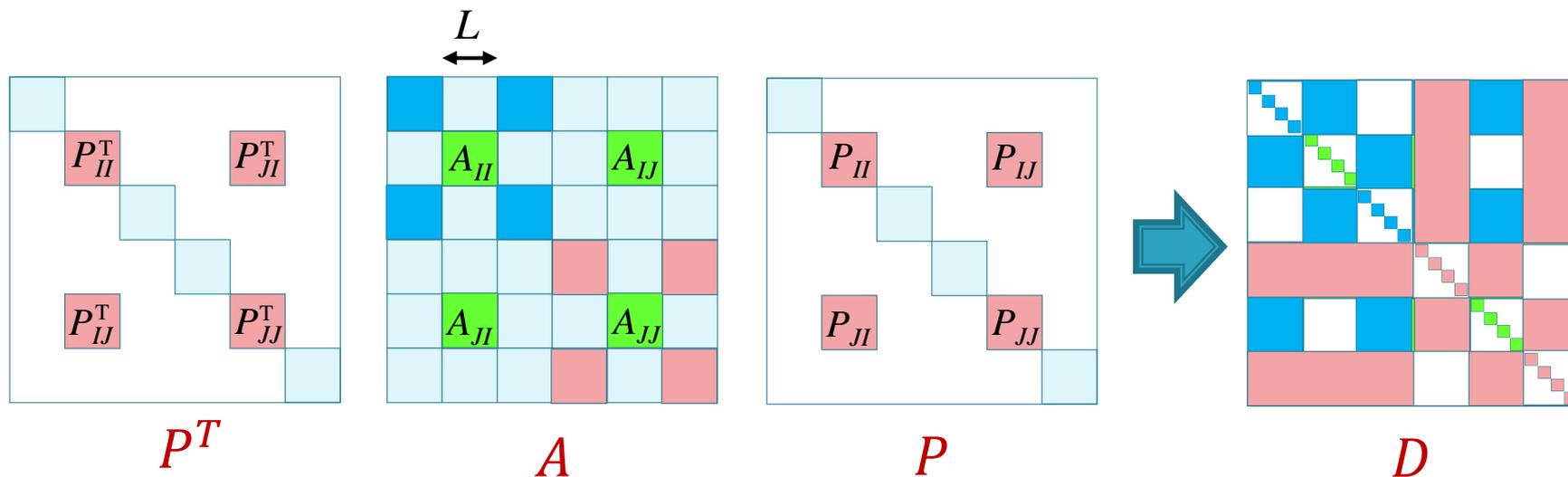
- 消去順により, 収束性, 並列性, 通信パターンが大きく異なる
- 消去順が固定されているため, 並列化・最適化がしやすい
- フロベニウスノルム最大のブロックを探索する必要がない

消去演算の並列性

- ▶ A_{IJ} の消去では、第*i*ブロック行(列)と第*j*ブロック行(列)のみが更新される



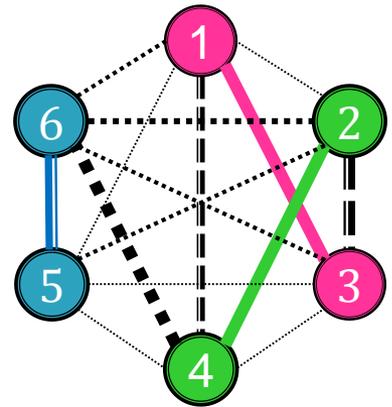
これらの行・列に属さない非対角ブロックを、同時に消去可能



一般に、 $\{I_1, J_1, I_2, J_2, \dots, I_{W/2}, J_{W/2}\} = \{1, 2, \dots, W\}$ であれば、 $W/2$ 個のブロック $A_{I_1 J_1}, A_{I_2 J_2}, \dots, A_{I_{W/2} J_{W/2}}$ を同時に消去可能

並列動的オーダリング

- ▶ アイディア (Becka et al., 2002)
 - Fノルムの2乗和 $\|A_{I_1 J_1}\|_F^2 + \dots + \|A_{I_{W/2} J_{W/2}}\|_F^2$ が最大になるように、消去すべき $W/2$ 個の非対角ブロックを選ぶ
 - 各行(列)を頂点, 枝 IJ の重みを $\|A_{IJ}\|_F^2$ とする完全グラフ上での**最大重みマッチング問題**と等価
 - 古典的ブロックヤコビ法の並列向け拡張と見なせる
- ▶ **最大重みマッチング問題の解法**
 - Blossom のアルゴリズム (Edmonds, 1964)
 - 厳密解を求められるが, $O(W^3)$ の演算量
 - 貪欲算法
 - 最も重い枝を選んで両端の頂点を除去. 以下同様に続ける
 - $O(W^2 \log W)$ の演算量で近似解が得られる



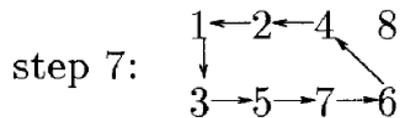
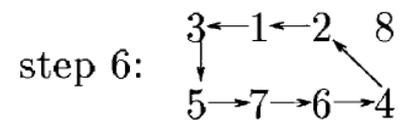
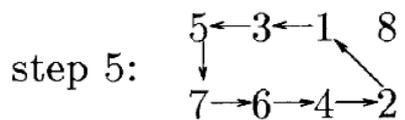
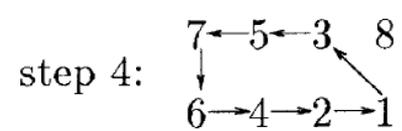
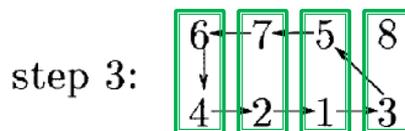
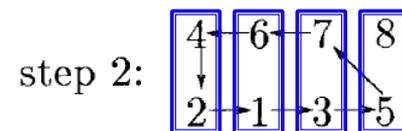
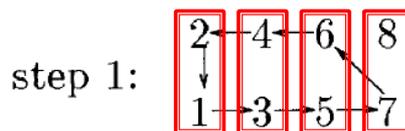
巡回型ブロックヤコビ法の並列化

方針

- 各ステップで $W/2$ 個の非対角ブロックを消去
- $W-1$ ステップで, 全非対角ブロックを1回ずつ消去するようにする

例1: ラウンドロビン順序

- 上下に並んだ数字が, 消去すべきブロックの行/列番号
- 数字の列が1回転すると1巡回
- ブロックの交換は隣接通信のみ
- 行巡回型(並列性なし)に比べて収束性が劣る



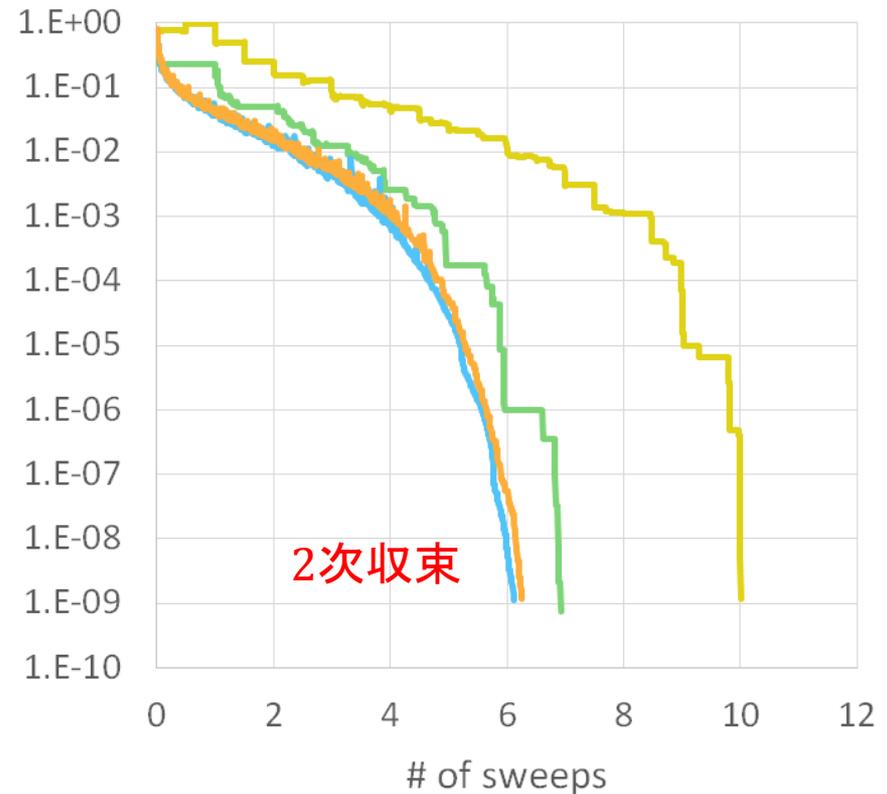
ピボット戦略による収束性の違い

▶ 実験条件

- 片側 (one-sided) ブロックヤコビ法による特異値分解 (後述)
- $n = 1200, L = 240$
- 乱数行列 (特異値が正規分布)
- 4種類の戦略を比較

▶ 結果

- 並列動的オーダリングを用いた場合が最も高速
- 最大重みマッチングを貪欲法で求めても厳密解と同程度の性能
- ラウンドロビンは行巡回より遅い



— Row Cyclic — Round-Robin
— Dynamic (Greedy) — Dynamic (Best Match)

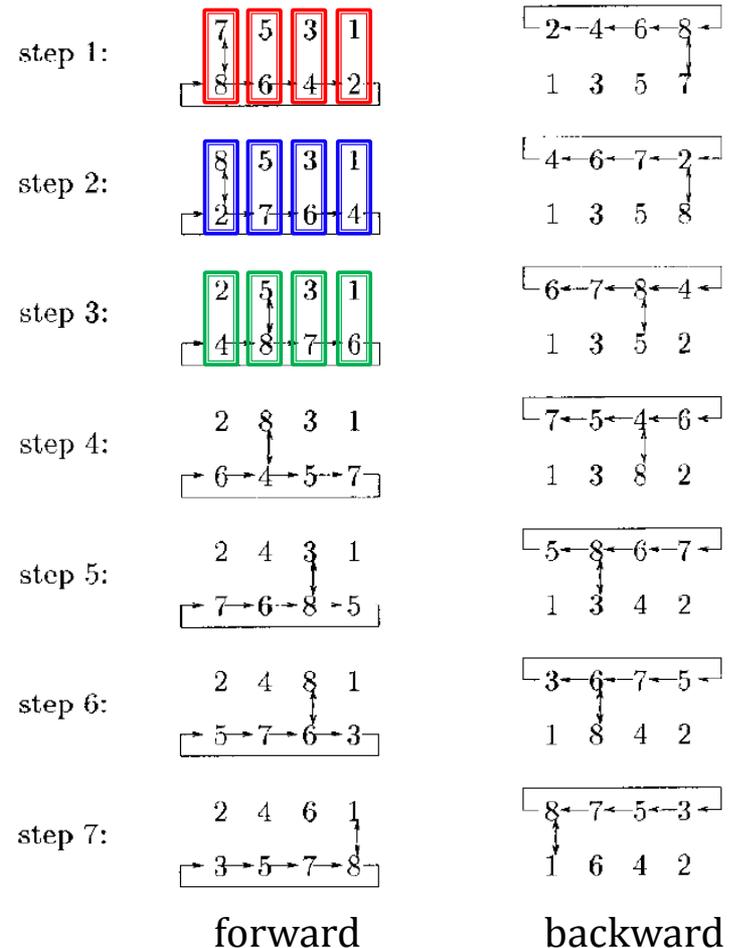
並列動的オーダリング + 貪欲算法

巡回型ブロックヤコビ法の並列化(続き)

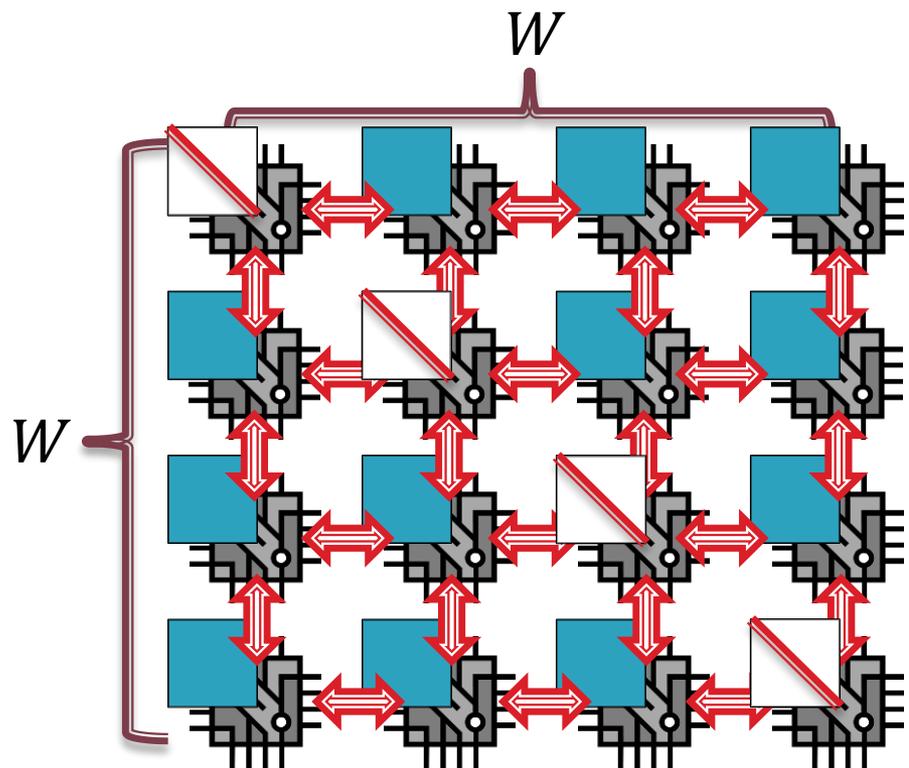
▶ 例2: リング型順序 (Zhou & Brent, 1997; 点ヤコビ向けに提案)

- 行番号を2列に並べ, 片側を回す
- 毎回決まった位置の行について, 上下を入れ替える
- 上下に並んだ数字が, 消去すべきブロックの行/列番号
- forward と backward とで1巡回
- ブロックの交換は隣接通信のみ
- 対角要素のソートと組み合わせることで, ラウンドロビン順序に比べて収束性が向上

➡ 我々のソルバで採用



並列実装手法 - 1ステップ内の処理



ノード形状とAの分割を
合わせる



データの交換

対角ブロックの
対角化

ブロック行列積
による行列更新

処理がブロック単位で行
われる(**大粒度並列性**)

3重対角化に基づく手法との比較

	3重対角化に基づく手法	ブロックヤコビ法
クリティカルパス上の計算量	$\frac{4n^3}{p}$	$\frac{20Kn^3}{p}$
通信回数	$O(n \log p)$	$O(K\sqrt{p} \log p)$

ノード数: p

反復回数: $K = (2L - 1)\text{step} \sim 10$

ブロックヤコビ法は通信回数が少ない



通信OHが大きい強スケーリング環境下での優位性

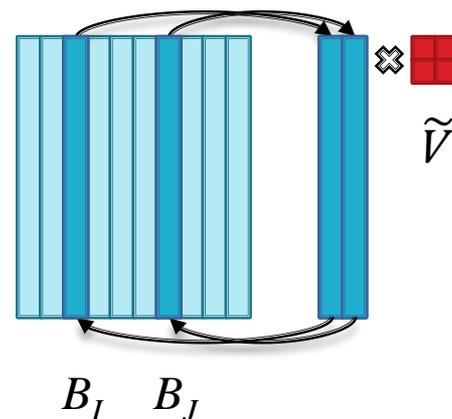
特異値分解：片側ブロックヤコビ法

▶ アイディア

- $B = U\Sigma V^T \Leftrightarrow BV = U\Sigma$
- すなわち, B に直交行列 V をかけて, 列が直交する行列 $U\Sigma$ に変形できれば, 特異値分解が求まる

▶ アルゴリズム

- 行列 B を幅 L の列ブロックに分割 (ブロック数 $W = n/L$)
- 2本の列ブロック B_I, B_J を取り出し, 右側から $2L \times 2L$ の直交行列 \tilde{V} をかけることで, $2L$ 本の列を直交化



固有値計算用のブロックヤコビ法との関係

▶ 対応関係

- ブロック分割された行列 $A = (A_{IJ})_{I,J=1}^W$ と $B = [B_1, B_2, \dots, B_W]$ について $A = B^T B$ という関係が成り立つとする
- このとき、次の対応関係が成り立つ

$$\begin{bmatrix} A_{II} & A_{IJ} \\ A_{JI} & A_{JJ} \end{bmatrix} = [B_I, B_J]^T [B_I, B_J]$$
$$\tilde{V}^T \begin{bmatrix} A_{II} & A_{IJ} \\ A_{JI} & A_{JJ} \end{bmatrix} \tilde{V} \text{ が対角行列} \Leftrightarrow [B_I, B_J] \tilde{V} \text{ の列が直交}$$

- この関係により、固有値計算用のブロックヤコビ法の収束定理の多くは片側ブロックヤコビ法の収束定理に翻訳できる
 - ただし、**有限精度での計算**では、両者の性質は大きく異なる

ここまでのまとめ

- ▶ 固有値計算のためのブロックヤコビ法
 - ピボット戦略
 - 古典的, 巡回型(多種類)
 - 並列化向きのピボット戦略
 - 並列動的オーダリング, 巡回型(ラウンドロビン順序, リング順序)
 - ピボット戦略による収束性の違い
 - 3重対角化法との比較: 通信回数が少ないことが特徴
- ▶ 特異値分解のための片側ブロックヤコビ法
 - 固有値計算のためのブロックヤコビ法との等価性

ブロックヤコビ法の収束性

▶ 大域的収束性

- 任意の初期行列 $A = A^{(0)}$ から出発したとき、非対角ノルム $\sum_{i < j} \|a_{ij}\|_F^2$ が 0 に収束するという性質
 - 古典的ブロックヤコビ法, 並列動的オーダリングでは成り立つ
 - 巡回型ブロックヤコビ法では, 一部の巡回順序については証明済み

▶ 局所的収束性

- 非対角ノルムが 0 に収束する場合に, 2次収束するという性質
 - 固有値に縮退がない場合については証明済み

実装と性能評価

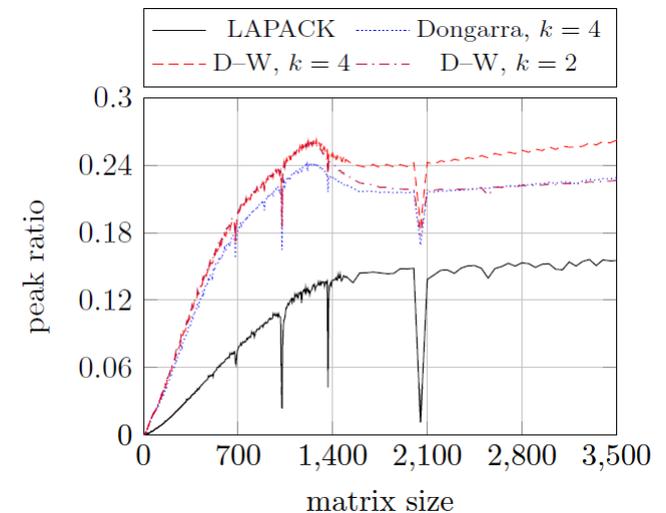
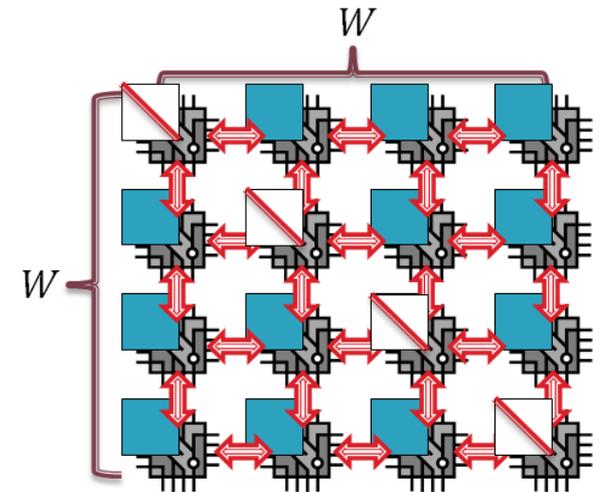
固有値計算のためのブロックヤコビ法

▶ 基本的なアルゴリズム・並列化手法

- 巡回型ブロックヤコビ法
- リング順序 + 対角要素のソート
- 2次元ブロック割り当て
 - 2×2 個のブロックを1ノードに割り当て

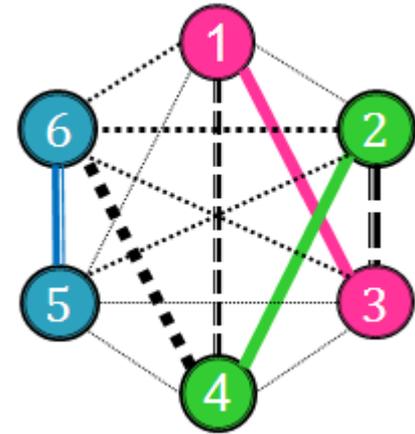
▶ 高性能実装手法

- $2L \times 2L$ 行列の固有値計算の最適化
 - Dongarra-Wilkinson法による3重対角化
(工藤他, HPCS2015)
- 隣接通信の最適化
- 演算と通信のオーバーラップ



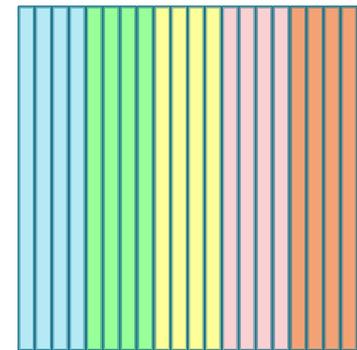
特異値分解のための片側ブロックヤコビ法

- ▶ 基本的なアルゴリズム・並列化手法
 - 並列動的オーダリング + 貪欲算法を用いた片側ブロックヤコビ法
 - 内積を $\|B_I^T B_J\|_F^2 = \|B_I^T B_J e\|_2^2$ と近似
 - ブロック列分割
 - 1本のブロック列を複数のノードに割り当て



- ▶ 高性能実装手法
 - 直交化演算 $[B_I^{(r+1)} B_J^{(r+1)}] := [B_I^{(r)} B_J^{(r)}] V^{(r)}$ における直交行列 $V^{(r)}$ の計算法
 - $[B_I^{(r)} B_J^{(r)}]$ が良条件: グラム行列の固有値分解
 - $[B_I^{(r)} B_J^{(r)}]$ が悪条件: 特異値分解
 - $\|B_I^T B_J e\|_2^2$ の計算の行列乗算化

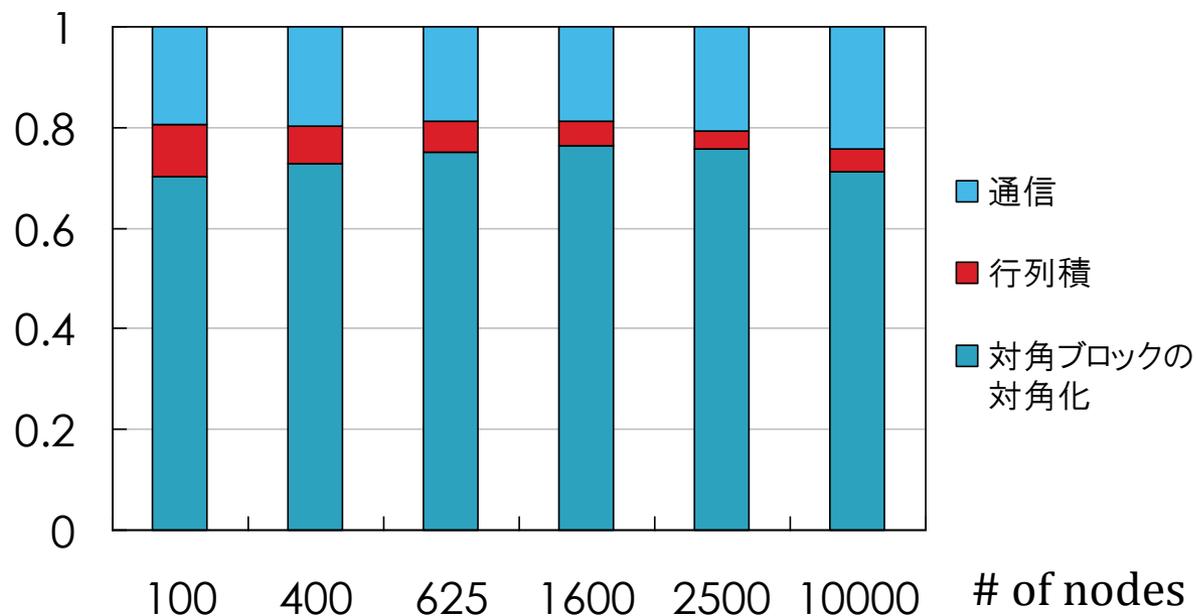
幅 L (ブロック列)



1つのブロック列を複数ノードに割り当て

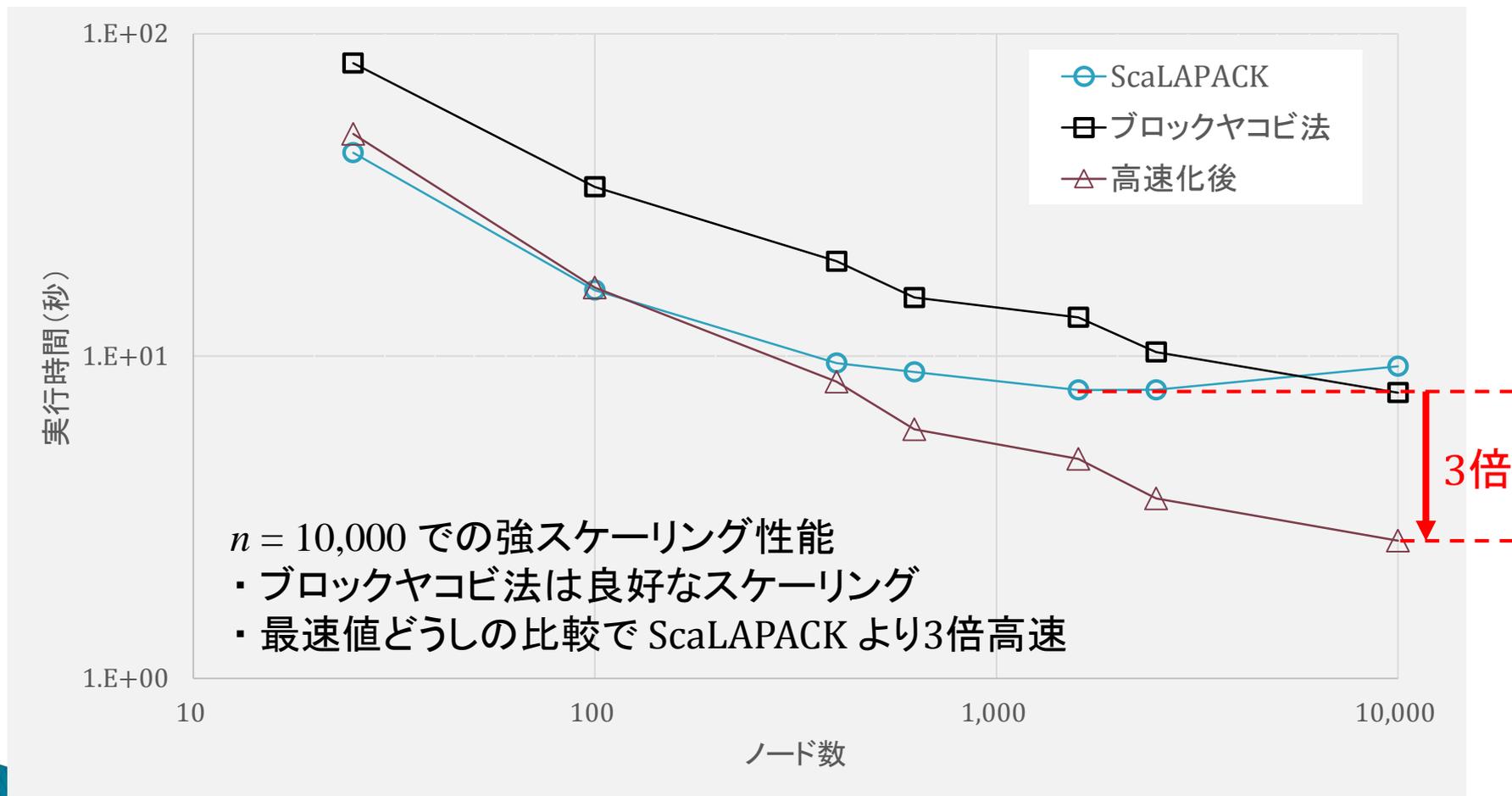
「京」上での固有値計算の性能(1)

$n = 10,000$ の乱数行列
各部分の実行時間の割合



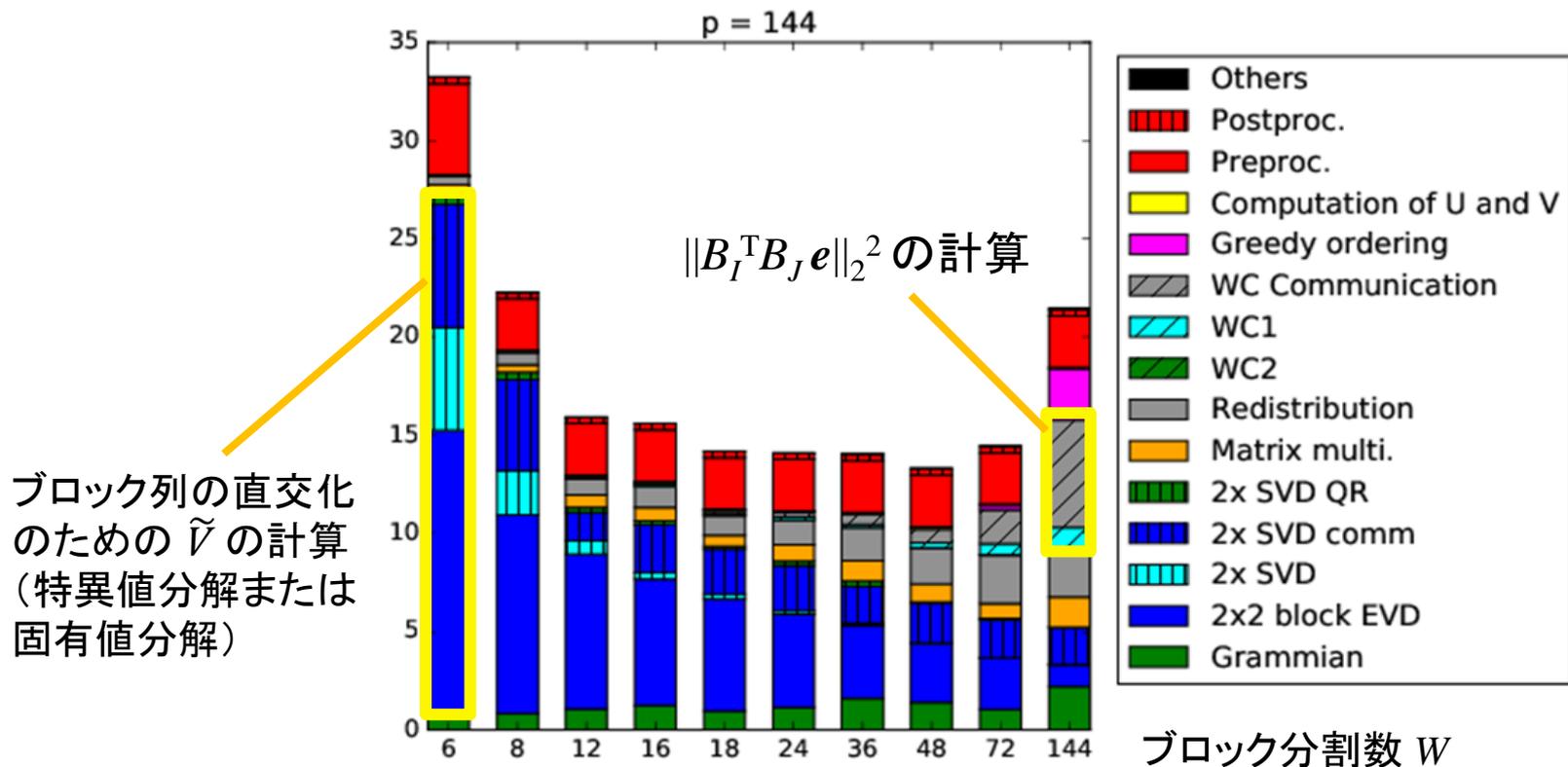
- ▶ 通信時間の割合が少ないことが特徴
 - ノード数をさらに増やしても加速する可能性
 - 一方, 3重対角化では, 実行時間の70%以上が通信
- ▶ 対角ブロックの対角化が大部分を占める

「京」上での固有値計算の性能(2)



特異値分解の性能(1)

- ▶ FX10(144ノード), $n = 4320$ の実行時間の内訳

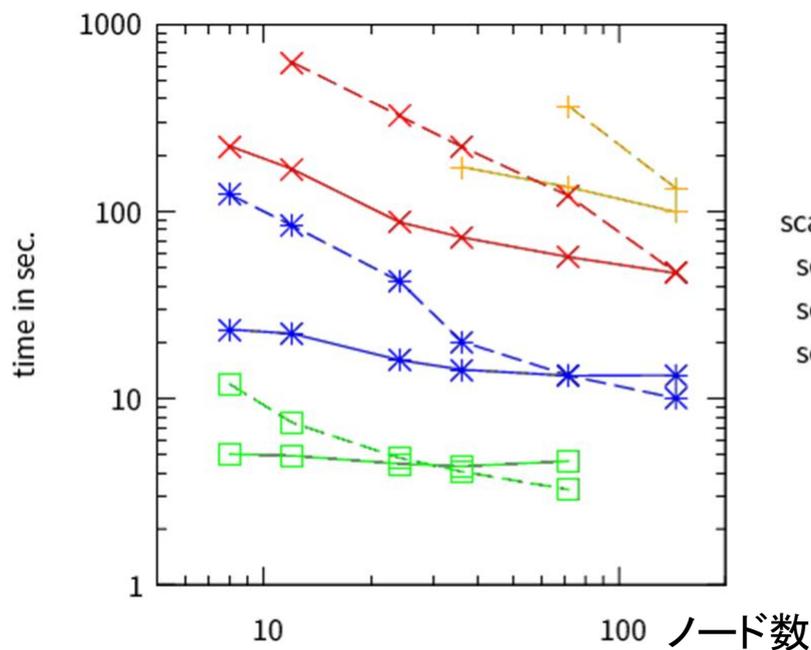


各部分の計算時間のトレードオフにより, 最適な分割数 W を決める必要あり

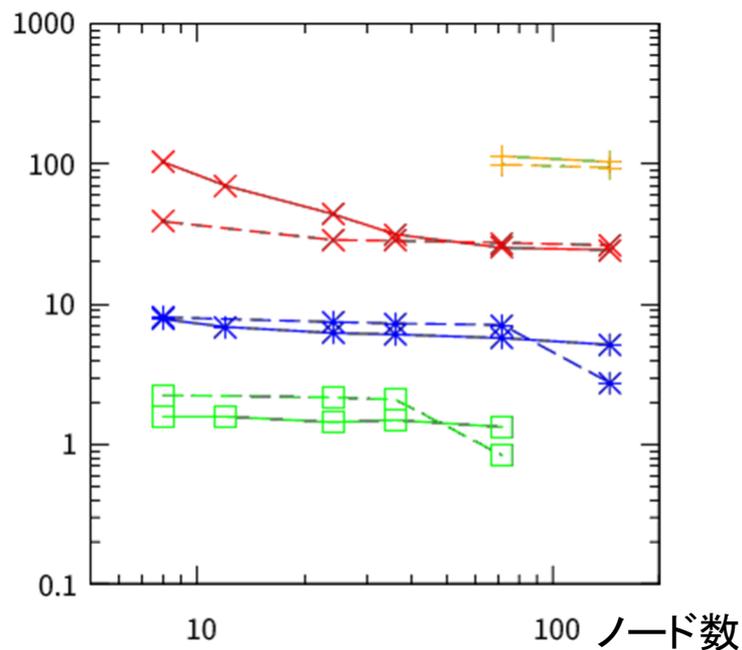
Kudo, Y., Becka, Vajtersic, Concurrency & Computation, Practice & Experience, to appear.

特異値分解の性能(2)

▶ ScaLAPACK との性能比較



FX10



Altix ICE (Xeon クラスタ)

ノード数が小さいときは片側ブロックヤコビ法が高速

- ・ 強スケーリング性能についてはまだ改善要

おわりに

まとめと今後の課題

▶ 本発表のまとめ

- 固有値・特異値計算のためのブロックヤコビ法について、アルゴリズムを紹介し、高性能実装手法と性能評価結果を示した
- ブロックヤコビ法は通信回数のオーダーが小さく、強スケーリング環境下では優位になり得ることを示した

▶ 今後の課題

- 巡回型ブロックヤコビ法の大域的収束性の解析
- 精度に関する理論的解析
- 実問題での評価
- FPGA などのハードウェアを用いた実装