

ポストムーアに向けた 密行列／FFTアルゴリズムと 精度保証および 自動チューニング技術の新展開

片桐 孝洋

(東京大学情報基盤センター)

「ポストムーアに向けた計算機科学・計算科学の新展開」シンポジウム
東京大学 武田先端知ビル5階 武田ホール
2015年12月22日(火) セッション4: 15:25-16:10

密行列／FFTアルゴリズムと精度保証および 自動チューニング技術の新展開

• 共同研究者

- 片桐孝洋(東大・情報基盤センター)
- 尾崎克久(芝工大)
- 高橋大介(筑波大)
- 今村俊幸(理研)
- 山本有作(電通大)
- 田中輝雄(工学院大)
- 櫻井隆雄(日立)
- 黒田久泰(愛媛大)
- 須田礼仁(東大・情報理工学)
- 佐藤雅彦(核融合研)
- 長尾大道(東大・地震研)




背景

話のまとめ

- ハードウェアのB/Fが1を超えると、高性能な数値アルゴリズムが変わる！
 - キャッシュブロック化 → 古典アルゴリズムの再来？
※ただしベクトル計算機時代に戻るのではない(新技術開発が必須)
- 多様な計算機アーキテクチャでも性能が出るようにする
 - 自動チューニング(AT)
 - 複数の候補を自動生成し、網羅的、もしくは、有効そうな候補を自動で調べて最適化(高速化)
 - (探索時間削減のため)性能モデル化が重要
 - ポストムーア時代の計算機アーキテクチャに適合できるようにする

ポストムーア時代の密行列アルゴリズムとAT

- 2020年頃のエクサスケールスパコンを境に
ムーアの法則での速度向上が終焉
 - “One-time Speedup”のメニーコア化、配線微細化、による低電力化の終わり
→**ノード内の計算性能(FLOPS)は増加しなくなる**
- チップ内のメモリ量やメモリ帯域は
「3次元積層技術」により向上
- 3次元積層メモリ:
 - z方向(積層方向)のバンド幅は増加、アクセスレイテンシは維持(→高性能化)
 - x-y方向は、低アクセスレイテンシ
- ノード間は、アクセスレイテンシは(相対的に)悪化、だが光配線技術などでバンド幅は増加
- **演算量増加ではなく<データ移動量増加>の【アルゴリズム変更が求められる!】**
- 以上は2020年後半(→ポストムーア)

- 
- **メモリバンド幅増加**
 - **局所メモリ(キャッシュ量)増加**
 - **メモリアクセスレイテンシ非均一を考慮したアルゴリズム再構築**

HMC

Source:

<http://www.engadget.com/2011/12/06/the-big-memory-cube-gamble-ibm-and-micron-stack-their-chips/>

Intel 3D Xpoint

Source:

http://www.theregister.co.uk/2015/07/28/intel_micro_n_3d_xpoint/

AT技術の課題

- **階層型AT方式**
- **通信削減アルゴリズム**
- **高バンド幅活用の<新アルゴリズム探求>と<自動切替>(アルゴリズム選択)**
 - 古典アルゴリズム復活?
 - キャッシュブロック化なしアルゴリズム
 - 陽解法から陰解法へ
 - アウト・オブ・コア(OoC)アルゴリズム(主メモリ外)



研究の方向性

ポストムーアに向けた課題(1/2)

1. データ移動中心の階層型メモリ／非均質メモリ アクセス性能を考慮した:

<チャレンジングな新アルゴリズム開発>

- (a)密行列解法／FFT のアルゴリズム
- (b)精度保証方式、基本密行列演算(BLAS)演算再現性
- (c)自動チューニング(AT)方式
- (d)FPGAを利用した数値計算処理のハードウェア実装
(データフロー処理で実装)

2. 新アルゴリズムの探求

- 従来のFLOPS中心のブロック化アルゴリズムではなく、
高バンド幅／(従来より)より大きなメモリ容量を利用した
<非ブロック化> アルゴリズム
- 主メモリ外の大規模データ処理に対応できる
高速メモリ・デバイス I/O を用いた
<アウト・オブ・コア(OoC)アルゴリズム>

ポストムーアに向けた課題(2/2)

3. 精度保証方式の探求

- 演算精度を落としても計算結果を保証する
精度保証の枠組み(反復改良アルゴリズム)
- 演算結果の保証、および演算数の削減(Approximation)
- ポストムーア計算機を想定した高性能実装方式開発
(データ移動中心アルゴリズムの積極活用)

4. ハードウェア研究と連携し、ATで必要となる性能モデル開発

- 基本演算レベル、アプリレベルの性能モデル
- <汎用型> と <特化型>

5. <データ移動中心のアルゴリズムを用いた> 実用アプリでのAT機能検証

- データ同化処理、プラズマコード(MHD)
- 大気海洋コード、物性物理コード、ほか



具体例

具体的なターゲット(1/2)

- 以下に限定しないが、実例を紹介
- データ移動中心アルゴリズムの再評価
 - 密行列アルゴリズム
 - 非ブロック化Householder三重対角化
 - 三重対角化は、ブロック化すると演算量が増える！
 - ブロックヤコビ法(B/F値大、演算量10倍)
 - 非ブロック化／多階層メモリ向け FFTアルゴリズム
- アウト・オブ・コア(OoC)アルゴリズム
 - キャッシュ＋3D積層メモリ＋新技術(MCDRAM + Intel 3DXpoint)の技術展開を想定した、主記憶外超高バンド幅、超多階層メモリ向けOoCアルゴリズム
 - OoC LU分解、OoC QR分解

具体的なターゲット(2/2)

• 特定部分をFPGA化

– 固有値計算

- スツルムカウント
- ある値以下の場合数をカウントする
- **必ずIF文が最内側**

- 先進CPU/GPUで実行効率が悪い^{ため}、FPGA化による高速化に期待
- 最外側ループ(固有値存在区間)の並列性はある(多分法と併用)

```

NEG=0;
S=0.0;
for (i=0; i<m; i++) {
  T = S - SIGMA;
  DPLAS = A[i] + T;
  S = T * B[i] / DPLAS;
  if (DPLAS < 0.0) NEG ++
}

```


• 実アプリケーション適用評価

– MHDコード

- 従来使っている陽解法から <陰解法の適用可能性>
- 陰解法: GMRES、BiCGStab法の利用
(データ移動の多い<疎行列-ベクトル積>ベースの演算)
- **データ移動の多い前処理手法**の適用可能性

– データ同化処理

- データ同化の主要演算、シミュレーション部分(陽解法)の高速化



短期・中期課題

今後2、3年で行うべき課題

- 階層メモリ・高バンド幅対応アルゴリズム開発と評価
 - 階層化FFT
 - 固有値アルゴリズム
 - 非ブロック化、高B/F値、の古典アルゴリズム
 - B/F=1程度のハードウェアを想定(中期的には、 $B/F > 4$)
- 非均質レイテンシアルゴリズム開発
 - 通信回避・削減する反復解法アルゴリズム(CA-CG)
- 精度保証アルゴリズム開発
 - 行列一行列積、連立一次方程式
- 性能モデル化(AT方式開発)
 - 階層メモリ対応
- FPGA化調査
 - 固有値アルゴリズム

中期的に取り組むべき課題

- **アウト・オブ・コア(OoC)アルゴリズム開発**
 - 基本行列分解(LU、QR)
- **精度保証アルゴリズム開発**
 - 高速な疑似多倍長計算、最高品質の追及
 - BLAS演算再現性
- **ポストムーア環境を想定した性能モデル化**
 - ハードウェア研究、システム研究と連携
- **FPGA化実証**
 - 固有値ソルバ、精度保証の基本演算、その他
- **アプリ実証**
 - MHD、データ同化、大気海洋、物性物理
 - アプリケーション研究と連携
- **電力最適化**
 - 数値計算ライブラリ、AT、および精度保証(Approximation)の観点
 - ハードウェア研究と連携