科学研究費助成事業 研究成果報告書

科研費

 令和 6年 6月19日現在

 機関番号: 13901

 研究種目: 若手研究

 研究期間: 2022 ~ 2023

 課題番号: 22K17898

 研究課題名(和文)実アプリケーションの時空間ブロッキングによる高速化に関する研究

 研究課題名(英文)A Study on Acceleration by Temporal Blocking for Real-world Applications

 研究代表者 星野 哲也(Hoshino, Tetsuya)

 名古屋大学・情報基盤センター・准教授

 研究者番号: 40775946

 交付決定額(研究期間全体): (直接経費) 1,100,000円

研究成果の概要(和文):微分方程式を解析的に解く際に生じる時・空間の離散格子に対する特定の計算パター ンはステンシル計算と呼ばれ、様々な流体シミュレーションにおいて頻出する重要なカーネルである。ステンシ ル計算の高速化は盛んに研究されており、時空間ブロッキング手法はその一手法であるが、非常に煩雑なプログ ラミングを要求するため、実アプリケーションへの適用例はほとんどない。さらに、時空間ブロッキングの性能 は実行するプロセッサの性能パラメータに大きく依存するため、人手によって最適化することは現実的ではな い。そこで本研究では時空間ブロッキングの自動最適化に必要な性能モデリングを、最新のCPUを用いて行っ た。

研究成果の学術的意義や社会的意義

本研究では、主にHigh Bandwidth Memory (HBM)を搭載した最新のCPUである、富岳スパコンのA64FXや、Intel Xeon Sapphire Rapids世代のCPUを用いて、性能モデル化を進めた点に大きな価値がある。時空間ブロッキング 手法はその性質上、特にメインメモリの性能とラストレベルキャッシュの性能比に性能が大きく依存する。この 性能比はHBMの登場によって既存のCPUと大きく変化し、本研究では性能モデルによってその影響を明らかにした ことが、高性能計算分野において意義のある成果である。また当初想定していなかった、命令レイテンシの影響 を明らかにした点も意義がある。

研究成果の概要(英文): The specific calculation pattern for a discrete grid in time and space that arises when solving differential equations analytically is called a stencil calculation, and it is an important kernel that frequently appears in various fluid simulations. Acceleration of stencil calculations has been studied extensively, and the temporal blocking method is one such method, but has rarely been applied to real applications because it requires very complicated programming. Furthermore, since the performance of temporal blocking is highly dependent on the performance parameters of the processor executing the blocking, it is not realistic to optimize the blocking manually. Therefore, in this study, the performance modeling required for auto-tuning of temporal blocking was performed using state-of-the-art CPUs.

研究分野: 高性能計算

キーワード: 高性能計算 ステンシル計算 時空間ブロッキング 自動チューニング

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等に ついては、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。 様 式 C-19、F-19-1、Z-19(共通) **1. 研究開始当初の背景**

微分方程式を解析的に解く際に生じる時・空間の 離散格子に対する特定の計算パターンはステンシ ル計算と呼ばれ、流体などのシミュレーションに おいて頻出する重要なカーネルである。ステンシ ル計算では、ある時刻 T+1における空間格子点(x、 y)を更新するために、時刻 T における近傍点、例 えば(x、y)、(x+1、y)、(x-1、y)、(x、y+1)、(x、 y-1)の値を参照する。一般に空間格子点の集合は 配列としてメモリ上に保持され、時間ステップの 進行の度にこの配列を更新するため、ステンシル 計算の性能は多くの場合メモリアクセス速度に律 速される。これに対し時空間ブロッキングは、メ モリ性能による限界を超えてステンシル計算を高 速化する手法として盛んに研究されている。時・



図 1:1 次元ステンシル計算におけ る時空間ブロッキングの方式例。

空間格子をキャッシュメモリに乗るよう小ブロックに区分し、メモリに書き戻すことなく N 時 間ステップ分の計算を行うことで、メモリアクセスコストを最大 1/N まで低減する。しかしス テップ数 N が増えるに伴い、さらに外側の近傍点まで参照範囲が広がるため、ブロックサイズ が肥大化し必要なキャッシュ容量が増加する。とりわけ、3 次元的でかつ計算に必要な変数の多 い実アプリケーションでは、キャッシュ容量の問題はより深刻化する。またブロックは更新すべ き領域より大きくなるため、他のブロックと重なる冗長計算領域が生じ、その分だけ浮動小数点 演算数は増加する。近年では冗長計算が必須な overlapped 方式(図 1:左上)を避け、冗長計算 の必要がない wavefront、 trapezoid、 diamond(図 1:右上、左下、右下)などの方式が提案 されている。ただしこれらの方式はブロックの計算順序に依存関係があるために並列化が難し く、単純なステンシル計算であっても非常に煩雑なプログラミングを要求するため、実アプリケ ーションへの適用例はほとんどない。

一方で、Intel Xeon Ice Lake や AMD Milan など、スパコンに搭載される最新世代の CPU は、 高い浮動小数点演算性能に加え、数十~数百 MB の大きなラストレベルキャッシュを有する。 一部のメニーコアプロセッサや GPU が高いメモリ転送性能を持つ代わりに容量の小さい High Bandwidth Memory (HBM)を登載する一方、汎用 CPU には容量に優れる DDR メモリに加え 強力な共有キャッシュを持たせるのが近年の傾向である。この大きな共有キャッシュを始め、最 新世代 CPU のメモリや演算性能のバランスの変化により、並列化が容易で比較的単純なコード 変形によって実現し得る overlapped 方式の時空間ブロッキングであっても、十分に高速化可能 であることが研究代表者らの研究でわかってきているが、HBM を用いた CPU のように既存の 汎用 CPU と大きく性能バランスの異なる最新の CPU において有効となる性能パラメータは明 らかではない。

2. 研究の目的

本研究の最終的な目的は、既存のステンシルアプリケーションへの時空間ブロッキング手法の 自動適用および最適化である。本研究課題ではその前段階として、A)最新世代の CPU におけ る時空間ブロッキングの性能モデリング、B)性能モデルに基づく実アプリケーションへの時空 間ブロッキングの手動適用による高性能化、の2点を目的とする。時空間ブロッキングにおい て時・空間ブロックのサイズを最適化するためには、参照する近傍点の数や相対位置、近傍点が 保有する物理量の数、CPU のメモリ・キャッシュ・レジスタの容量や性能、といった多数のパ ラメータが関わっており、性能モデリングを行うことは容易ではなく、特に最新世代の CPU の 性能バランスを考慮した時空間ブロッキングの性能モデリングは例がない。

3.研究の方法

まず A64FX や Intel Xeon Sapphire Rapids などの最新世代 CPU について、各種ベンチマー クにより詳細な評価を行い、性能モデルを構築するためのデータを蓄積する。その上で時空間ブ ロッキング手法において有効なパラメータの特定、有効パラメータを推定するための性能モデ ルの構築を目指す。

4. 研究成果

近年のメニーコアプロセッサは、性能バランスが変わってきている。Intel の最新世代の CPU である Sapphire Rapids は、Intel 社の x86 製品としては初めて High Bandwidth Memory (HBM) が用いられている。同じく HBM を搭載した「富岳」のプロセッサである A64FX は、 一般的な DDR メモリを用いた Intel Xeon CPU と異なる特性を持つことが知られており、アプ リケーションの最適化戦略に影響を与えている。HBM を搭載した Sapphire Rapids について も、異なる最適化戦略が必要となる可能性があるが、Sapphire Rapids に関する評価は未だ十分 ではない。そこで本課題では、複数のマイクロベンチマークを用いて、Sapphire Rapids を評価 した。同じく HBM を採用する A64FX や、IceLake 世代の Intel Xeon CPU と比較を行った。 ① stream ベンチマーク

Sapphire Rapids with HBM のメモリ性能につ いて、指標としてよく用いられる Stream Triad (c = alpha * a + b)の結果を図 2 に示す。実験に 用いたのは京都大学の Camphor スーパーコン ピュータシステムの Sapphire Rapids であり、 コア数は56であるが、ハイパースレッディング が有効となっているため、それを考慮してスレ ッドとコアの割り付けを行う必要がある。また Intel コンパイラのオプションとして-qoptstreaming-stores があるが、これを-qoptstreaming-stores=always とすることで、ストリ ーミングストアの最適化が有効化される。スト リーミングストアとは、キャッシュへの書き込 みを行わずにメモリへ結果を書き込むストアを 行う方法で、キャッシュ書き込み分のオーバー ヘッドを削減できる。

ベンチマークの配列サイズは 100M 要素とし、 キャッシュに乗らないサイズに設定している。 コンパイラには Intel compiler 2021.7.1、オプ ションとしては-xCORE-AVX512、-qopenmp、 -O3 を用い、最大 56 コアを使用した。またコン パイラオプションの -qopt-streamingstores=always の有無、環境変数として KMP_AFFINITY を balanced または compact の設定に変更して実験した。実行時には numactl コマンドにより、コアに近いメモリを 利用するよう設定している。



図 2: Sapphire Rapids における Stream Triad。



図 3: Sapphire Rapids のコア構成(1 numa 分)と帯域幅。

図2はスレッド数を1から最大まで変化させ

た時の性能を示している。Sapphire Rapids は最大 56 コアだが、ハイパースレッディングを考慮し、112 スレッドまで計測している。ここで重要なのは、Sapphire Rapids は理論値で 1.6TB/sec のメモリバンド幅性能を持つが、何らかの理由でこの性能を引き出し切れていないことである。性能の最も高い部分を見ると、-qopt-streaming-stores=always とすることで 3%程度性能が向上しているが、それでも 880 GB/sec 程度で、理論性能の 55%しか得られていない。A64FX が HBM の 80%以上の性能を引き出していることと比較するとかなり低い。

この理由は、John D. McCalpin (ISC 2023)らの報告(https://www.ixpug.org/images/docs/ ISC23/McCalpin_SPR_BW_limits_2023-05-24_final.pdf))で示されているが、HBMの帯域幅 をコア間のメッシュのY方向の通信性能が下回るためである。

図3は、Sapphire Rapidsのコアの構成(1 numa分)を示したものである。コア間はX-Yの 二次元メッシュで接続されており、メッシュ間の1リンクの通信性能は51.2~76.8 GB/s であ る。なお幅があるのは、熱によりクロック周波数が変動するためであり、メモリバンド幅を使い 切るようなカーネル実行時には51.2 GB/s に近づくと考えられる。通信リンクは4系統× X-Y の2方向あるため、メッシュの総バンド幅はHBMと同等か上回るが、HBMへのアクセス時に は図3のようにY方向の通信に律速されるため、HBMの性能が活かしきれない。

以上の結果から、アプリケーションの性能を考える上では、メモリバンド幅は単純なメモリ性 能のみによって決まるわけではなく、コア間のメッシュ性能なども考慮する必要があることが 判明した。

キャッシュ性能に関するベンチマーク

近年の CPU は 100 MB を超える大きな共有キャッシュを備えるが、共有キャッシュの性能は コア間のメッシュの性能に依存するため、性能モデル構築の上で考慮する必要がある。図4は、 Sapphire Rapids のメモリ階層を示したものである。同様に HBM メモリを備える A64FX は、 L1 キャッシュがコアローカルのキャッシュで、L2 キャッシュが共有キャッシュであるが、

Sapphire Rapids では L1 及び L2 キャッシュがコアローカルのキャッシュで、L3 キャッシュが 共有キャッシュとなっている。

そこで A64FX と Sapphire Rapids の両者において、図 5 のコードを用いて評価を行なった。 なお、A64FX では SVE 命令に置き換えたものを使用した。

図5のコードをOpenMPを用いて並列化し、以下の3種を試した。

- normal:スレッドローカル領域でA(:) = A(:) + B(:)を繰り返す。
- normal_w/barrier:for 文終了後に#omp barrier を呼ぶ。
- round_w/barrier: 読み書きする領域をスレッド間でラウンドロビンに交換しながら A(:) = A(:) + B(:)を繰り返す。

Sapphire Rapids 及び A64FX における結果をそれぞ れ図 6 及び図 7 に示す。A64FX では normal_w/barrier と round_w/barrier の性能がほぼ 一致しているが、Sapphire Rapids では乖離してい る。これは A64FX が両者で L2 キャッシュを使う一 方で、Sapphire Rapids は normal_w/barrier を L2 キャッシュ、round_w/barrier を L3 キャッシュを使 って実行するためである。グラフの Y 軸はコアあた りの性能であるが、これを全コアでの性能に換算す ると、Sapphire Rapids の round_w/barrier の性能 は Stream triad のスループットの 2 倍強となる。こ れは、ラウンドロビンによるメッシュ間のデータ交 換では、図 3 に示した X-Y 両方向のリンクを利用で きているためと考えられる。

上述の結果から、アプリケーションの性能モデルを 考える上では、コア間の通信方法、共有キャッシュの 性能、コアローカルなキャッシュの性能など、考慮す べきパラメータがさらに増加傾向であることがわか る。

③ 命令レイテンシに関するベンチマーク

上述の実験では、近年の CPU の性能モデルを考え る上では、単にメモリの性能だけではなく、コア間の メッシュ性能など、様々な要因がアプリケーション の性能に影響することを示した。CPU のコア性能に 関しても、単にピーク性能だけでなく、命令レイテン シやレジスタ容量についても考える必要がある。 A64FX、Sapphire Rapids 及び Intel Xeon IceLake を用いて、図 8 のコードによる評価を行なった。な お、A64FX では SVE 命令に置き換えたものを使用 した。

図 8 のコードは、c=c+a*b を繰り返すループを変 形したものである。通常このループはコンパイラの 最適化により高速に実行されてしまうが、intrinsics を用いて実装することで敢えてパイプライン並列 化を阻害するコードとなっている。ループイテレー ション間で命令の依存性がある(右辺に現れる変数 veccN が前ループイテレーションの左辺変数として 現れている)ためであるが、ループボディ内で別変 数に書き込む statement を増加させると、その分だ けパイプライン並列化が可能となる。

各プロセッサにおいて、ループボディ内の独立な statement 数を 1~40 まで変化させた際の性能を、 図9に示す。結果から、ピークに達する(つまり FMA 命令のレイテンシを隠し切る)までに必要な独立な statement 数(つまりパイプライン並列数)は、

Sapphire Rapids 及び IceLake では 10、A64FX で は 20 となることがわかる。これは FMA 命令+代入 命令のレイテンシ (Sapphire Rapids と IceLake は 4 + 1 cycle、A64FX は 9 + 1 cycle) と FMA 演算器 の数を (全プロセッサで 2 FMA) を考慮すると妥当 な数値である。また A64FX は statement 数 31 で性 能が大幅に低下している一方で、Sapphire Rapids と IceLake では性能低下が緩やかである。Statement 数 31 の時に必要なベクトルレジスタ数が 33 であり、 32 を超えたことによる性能低下は妥当であるが、プ ロセッサ (あるいはコンパイラ)によってその度合い は大きく異なることがわかった。このように A64FX



図 4: Sapphire Rapids のメモリ階層。



図 5: キャッシュ性能を測るためのマイ クロベンチマークコード。



図 6: 図 5 のコードの Sapphire Rapids による評価結果



図 7: 図 5 のコードの A64FX による 評価結果



図 8:ピーク性能を達成する上で必要 なパイプライン並列数、レジスタスピ ルの影響を測る目的のマイクロベン チマークコード。

のように命令レイテンシの大きいプロセッサでは、ピーク性能を得られるスイートスポットが 狭いことなども考慮して性能モデル化する必要性があることがわかる。

④ ステンシル計算の時空間ブロッキングを対象とした性能モデリング

アプリケーションに近いベンチマークとして、7 点ステンシルと呼ばれる、流体計算に頻出する3次元の拡散方程式のカーネル(図10)を利用する。特に、時空間ブロッキングと呼ばれる

高速化手法に着目し、性能モデル化を目指す。7 点 ステンシル計算に、時空間ブロッキングを含む以 下の最適化を適用し、Sapphire Rapids、A64FX、 Intel Xeon CascadeLake を用いて評価した。

- BASE: 図 10 のベースライン実装。
- FT: ファーストタッチを行った実装。
- PEEL: ループピーリングや branch hoisting と呼ばれる、最内ループ中の分岐を ループの外に出す手法。例外処理の発生する ループの端の処理のみ分離する。
- 図 10 の 2 行目の omp parallel for 指示文を、 nowait 指示節を付与した上で 4 行目のループ に適用し、y 軸分割を行う実装。
- Y-Zdim: omp parallel for 指示文を用いず、 スレッド番号を利用し、y-zのループを2次 元分割する。この際 z 軸ループは NUMA 単 位で4分割し、y 軸ループはコア単位で分割 する。
- INT: AVX512 や SVE の Intirnsics を利用し た実装。
- UNR: ループアンローリングにより、ループ 内で実行される独立な命令数を増やすこと で、パイプライン実行の効率化を図った実 装。
- REG: x 軸方向のレジスタブロッキングを行った実装。また AVX512 の_mm512 _alignr_epi640命令を利用することで、非ア ラインドなメモリアクセスを回避する。
- TILE: タイリングによるキャッシュの利用 効率化を行った実装。
- TB:時空間ブロッキング(別名:テンポラルブ ロッキング)を適用した実装。

図11に、上記最適化を順次適用した結果を示す。 全体的な傾向として、Sapphire Rapidsの結果は CascadeLake よりも A64FX の性能の傾向と似て いる。上記最適化のうち、Y-Zdim が効果的である ことがわかる。これは、NUMA 間で発生するキャ ッシュコヒーレントのための通信を最小化するた めの分割手法である。つまり、Sapphire Rapids で は、NUMA 間通信が性能の足を引っ張ることが考 えられるため、その点に気を付ける必要がある。 UNR で表されるループアンローリングは、A64FX で有効である一方、Sapphire Rapids では性能を低 下させる要因となった。A64FX で有効だった理由 は、上述の命令レイテンシによるものと考えられ

4000 ピーク性能を得ら れる範囲 3500 3000 2500 シスタスビルに る性能低下 og 2000 1500 1000 500 Sapphire Rapids A64FX IceLake 0 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 独立なStatement数



```
do {
#pragma omp parallel for
   for (int z = 0; z < nz; z++) {</pre>
     for (int y = 0; y < ny; y++)
       for (int x = 0; x < nx; x++) {</pre>
         int c = x + y * nx + z * nx * ny;
          int w = (x == 0)
                                  c : c - 1;
         int e = (x == nx-1) ? c : c + 1;
         int n = (y == 0)
                                  c : c - nx;
         int s = (y == ny-1) ? c : c + nx;
         int b = (z == 0)
                                ? c : c - nx * ny;
         int t = (z == nz-1) ? c : c + nx * ny;
         f2_t[c] = cc * f1_t[c]
             + cw * f1_t[w] + ce * f1_t[e]
             + cs * f1_t[s] + cn * f1_t[n]
+ cb * f1_t[b] + ct * f1_t[t];
       }
     }
   }
   double *tmp = f1_t;
   f1_t = f2_t;
   f2_t = tmp;
   time += dt;
} while (time + 0.5*dt < 0.1);</pre>
```







る。ステンシル計算は一変数への足し込みを行うため、命令間に依存が生じる。ループアンロー リングはループボディ内の独立な statement 数を増やす効果があるためである。一方 Sapphire Rapids で性能が低下した原因は性能モデルからは説明がつかず、他の要因を考える必要があり そうだが、原因の究明には至っていない。また TB で表される時空間ブロッキングは、DDR メ モリを搭載する従来型の CPU である CascadeLake で非常に有効であるものの、Sapphire Rapids では有効ではなかった。時空間ブロッキングはメインメモリへの負荷を抑え、キャッシ ュの負荷を増加させる手法である。今回適用した時空間ブロッキング手法は、ラストレベルの共 有キャッシュを利用する手法であったため、上述した通り、ラストレベルキャッシュが負荷の増 加に耐えられなかったと考えられる。

A64FX においても時空間ブロッキングは有効ではなかったが、こちらはプロファイリングの 結果を見るに、時空間ブロッキングで冗長な計算が増えたことによって、キャッシュではなくコ アの負担増によって性能が低下したと考えられる。上記のように、プロセッサのパラメータによ って最適化方針が全く異なることがわかり、また性能モデルから律速原因の説明が可能である ことがわかった。A64FX 及び Sapphire Rapids での高速化のためには、単なるパラメータの調 整ではなく、実装方針を切り替える必要があり、性能モデルに基づく自動最適化は今後の課題で ある。

5.主な発表論文等

〔雑誌論文〕 計3件(うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 0件)

1.著者名 Dan Yungahangha Kawai Magataghi Haghing Tatauwa Katagiri Takahira Nagai Taru	4.巻
Ken Auanzhengbo, Kawai wasatoshi, Hoshino retsuya, Katagiri rakaniro, Wagai foru	none
2.論文標題	5 . 発行年
Auto-tuning Mixed-precision Computation by Specifying Multiple Regions	2023年
3.雑誌名	6.最初と最後の頁
2023 Eleventh International Symposium on Computing and Networking (CANDAR)	175-181
「掲載論文のDOI(デジタルオブジェクト識別子)	査読の有無
10.1109/CANDAR60563.2023.00031	有
オープンアクセス	国際共著
オープンアクセスではない、又はオープンアクセスが困難	-

1.著者名	4.巻
Masatoshi Kawai, Akihiro Ida, Toshihiro Hanawa, Tetsuya Hoshino	none
2 . 論文標題	5.発行年
Optimize Efficiency of Utilizing Systems by Dynamic Core Binding	2024年
3.雑誌名 HPCAsia '24 Workshops: Proceedings of the International Conference on High Performance Computing in Asia-Pacific Region Workshops	6 . 最初と最後の頁 77-82
掲載論文のDOI(デジタルオプジェクト識別子)	査読の有無
10.1145/3636480	有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著

1.著者名	4.巻
Hoshino Tetsuya、Ida Akihiro、Hanawa Toshihiro	2022
2.論文標題	5 . 発行年
Optimizations of H-matrix-vector Multiplication for Modern Multi-core Processors	2022年
3.雑誌名	6.最初と最後の頁
2022 IEEE International Conference on Cluster Computing (CLUSTER)	462,472
掲載論文のD01(デジタルオプジェクト識別子)	査読の有無
10.1109/CLUSTER51413.2022.00056	有
オープンアクセス	国際共著
オープンアクセスではない、又はオープンアクセスが困難	-

<u>〔学会発表〕 計5件(うち招待講演 1件 / うち国際学会 0件)</u> 1.発表者名

百武尚輝,星野哲也,小澤創,伊田明弘,安藤亮輔,河合直聡,永井 亨,片桐孝洋

2.発表標題

OpenACCを用いた地震シミュレーションのGPU並列化

3 . 学会等名

情報処理学会全国大会

4.発表年

2024年

1.発表者名

Tetsuya Hoshino, Akihiro Ida, Toshihiro Hanawa

2.発表標題

Optimizations of H-matrix-vector Multiplication for Modern Multi-core Processors

3 . 学会等名

Japan Geoscience Union Meeting 2023(招待講演)

4.発表年 2023年

1.発表者名

Tetsuya Hoshino, Akihiro Ida, Toshihiro Hanawa

2.発表標題

Optimizations of H-matrix-vector Multiplication for Modern Multi-core Processors

3 . 学会等名

ICIAM

4 . 発表年 2023年

1.発表者名

星野 哲也 ,河合 直聡 ,伊田 明弘 ,塙 敏博 ,片桐 孝洋

2.発表標題

HPCカーネルベンチマークによるSapphire Rapids HBMの性能評価

3 . 学会等名

研究報告ハイパフォーマンスコンピューティング(HPC)

4.発表年 2024年

1.発表者名

満田 晴紀,星野 哲也,望月 祐志,坂倉 耕太,片桐 孝洋,大島 聡史,永井 亨,河合 直聡

2.発表標題

分子軌道計算プログラムの性能評価と自動チューニング適用の検討

3 . 学会等名

研究報告ハイパフォーマンスコンピューティング(HPC)

4.発表年 2023年 〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6	研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7.科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8.本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------