

HPC システムにおける Approximate Computing のための 非同期な演算精度変更

Asynchronous Precision Adjustment for Approximate Computing on HPC Systems

和田 康孝
明星大学情報学部

1. 研究目的

限界を超えてシステムの実行性能を向上させるために、演算結果にある程度の誤りが含まれることを許容し、その誤差と引き換えに性能を向上させる Approximate Computing 手法が注目を集めているが、Approximate Computing の効果を最大限発揮するためには、単純にアプリケーション全体で演算精度を調整するのではなく、その構造に合わせて、実行中に動的に演算精度・データ型を変更することが望ましい。しかしながら、アプリケーションの箇所ごとに適切な演算精度を推定し、プログラムの書き換え等を行うことは、プログラム実行時のオーバーヘッドを増加させるのみならず、開発コストも増大してしまう。

このような背景から、本研究課題では、アプリケーション内で動的に演算精度を切り替える際のデータ型変換やコピーを非同期に行うことで、高速化あるいは隠蔽する手法について検討・評価を行い、その効果を検証することを目的とした。これにより、開発コストを抑えつつ、アプリケーションの高速化・高効率化をより簡単に実施できるようになるものと考えられる。効果が認められた手法においては、ライブラリ等として実装するとともに、プログラマから容易に利用可能となるように API を策定することも目指す。

2. 研究成果の内容

上述の目的・目標に向けた研究開発を実施するにあたり、Approximate Computing 適用時のオーバーヘッドを削減・隠蔽するため、特に通信コストを隠蔽するための非同期通信に関する性能評価と通信コストのモデル化を実施した。Approximate Computing 適用時には、データ精度変換や通信デバイス操作等のオーバーヘッドを考慮する必要はあるものの、演算処理とは異なり、データ精度（ビット幅）の縮小がそのまま通信対象のデータ量に反映される。そのため、メモリバンド幅を使い切らない範囲であれば、転送コストは比較的予測しやすいものと考えられた。しかし実際には、通信ライブラリの実装によっては、演算と通信が逐次的に行われる場合もあることが示唆される結果となった。そのため、単純に演算コストと転送データ量のみからオーバーラップ可能な通信の割合を直接的に見出すのは難しい。通信ライブラリの実装あるいは使用方法を改善することで、演算と通信が最大限オーバーラップされる状態を作り出すことが可能であることも明らかとなったため、これらの知見をライブラリの実装

や使用方法に反映した上で、演算精度、演算コスト、データ転送コスト、演算結果の正確性、を併せて考慮する必要がある。

併せて、NAS Parallel Benchmark に含まれるベンチマークプログラム等を用いて、実行時に演算精度を変更した際の実行性能および演算結果への影響を評価した。演算精度変更のタイミングによっては、計算結果が適切な値に収束せず、プログラムの実行が破綻してしまう場合もある。これは単純に演算精度を変更するだけでは解決できないため、ロールバック等の手法を用いて対策を取る必要があることがわかった。

3. 学際共同利用プログラムが果たした役割と意義

学際共同利用プログラムにおける取り組みを通じて、Pegasus, Cygnus, Wisteria-0 のみならず、Miyabi を先行利用にて活用する機会を提供いただき、様々な特徴を持ったシステムを対象に評価・開発を行うことができた。特に、本研究課題における成果を元に検討・応募した、科研費・基盤研究 (C) 研究課題「動的なデータ精度調整可能な通信ライブラリの開発」(2025 年度～2027 年度) が新たに採択となった。計算機環境の利活用に加え、研究者間の連携、さらには科研費研究課題の採択など、研究課題の発展という意味でも、本学際共同利用プログラムの果たした役割は大きい。

4. 今後の展望

HPC システムにおける性能向上は、ハードウェア、システム、アプリケーション等様々な要素が関係している。本研究課題が対象とする Approximate Computing 技術においてもそれは同様であり、今回特に評価の対象とした汎用プロセッサや通信に加え、ストレージやメモリ、アクセラレータデバイスなど、今後、より幅広い対象を考慮した手法・仕組みに発展させていく必要がある。新たに採択となった科研費研究課題における研究開発も含め、本研究課題の取り組みを継続し、より発展させていきたい。特に、ユーザの利便性等を考えると、ライブラリや API の策定から、コンパイラや開発フレームワークとしての発展に繋げていくことが必要になると考えられる。

5. 成果発表

(1) 学術論文

該当なし

(2) 学会発表

- [1] 和田 康孝, 小林 諒平, 森江 善之, 坂本 龍一, “イタレーションレベル Approximate Computing 手法の提案と予備評価”, 情報処理学会研究報告 ハイパフォーマンスコМПユーティング (HPC), Vol. 2025-HPC-199, No. 6, pp. 1-5, May, 2025.

- [2] 森江 善之, 和田 康孝, 小林 諒平, 坂本 龍一, 南里 豪志, “高スループ

ット非同期集団通信の性能モデル化に向けた予備評価” , 情報処理学会研究報告 ハイパフォーマンスコンピューティング (HPC) , Vol.2025-HPC-198, No. 49, pp.1-6, Mar. , 2025.

[3] Yasutaka WADA, “Toward Energy-Efficient HPC Systems with Dynamic Power-Performance Adjustment” , Keynote Speech, The 34th International Conference on Computer Theory and Applications (ICCTA 2024), Dec. , 2024.

(3) その他

[4] Yasutaka WADA, Yoshiyuki MORIE, Ryohei KOBAYASHI, and Ryuichi SAKAMOTO, “Toward Performance Analysis and Modeling of HPC Applications for Dynamic Approximate Computing” , Exhibition, The International Conference for High Performance Computing, Networking, Storage, and Analysis(SC24), Nov. , 2024.

使用計算機	使用計算機に ○	配分リソース※		
		当初配分	移行*	追加配分
Cygnus	○	6,048	0	0
Pegasus	○	6,048	0	0
Wisteria/BDEC-01	○	25,200		0
	※配分リソースについてはノード時間積をご記入ください。 *バジェット移行を行った場合、「+2000」「-1000」のように記入			