

多様な演算加速装置を用いる高性能システムのためのプログラミング環境の開発

Development of Programming Environment for High-performance Systems using Various Accelerators

藤田典久

筑波大学計算科学研究センター

1. 研究目的

科学技術計算では、スーパーコンピュータによる大規模計算が不可欠である一方、性能向上に伴う消費電力の増大が大きな課題となっている。そのため、単なる高性能化だけでなく、電力あたり性能の高い計算環境の実現が求められている。また、ムーアの法則に基づく従来型の性能向上が限界に近づく中、CPU の性能向上だけに依存した計算機アーキテクチャでは、今後の科学技術計算の要求を十分に満たすことが困難になりつつある。このような背景から、GPU をはじめとする演算加速装置の利用が広がっている。しかし、GPU はすべての計算に有効なわけではなく、加速できない部分がアプリケーション全体の性能を制限する場合がある。そこで本研究では、複数種類の演算加速装置を計算内容に応じて適材適所に利用し、アプリケーション全体の性能および電力効率を高めることを目指す。一方、複数の演算加速装置を扱うプログラミングは複雑であり、開発者は異なる言語や開発環境、データ転送、実行順序制御などを個別に扱う必要がある。さらに GPU ベンダの多様化や HPC-QC の進展により、アプリケーションには多様な環境で効率よく動作する性能可搬性が求められている。以上を踏まえ、本研究では、単一計算機内の複数アクセラレータ利用と、異なるアクセラレータを備えた複数計算機の横断利用の双方に対応できる、統一的なプログラミング環境の実現を目的とする。

2. 研究成果の内容

本年度は、複数種類の演算加速装置を統一的に扱う開発環境「UniSYCL」に関する研究開発を継続した。UniSYCL は、異なるアクセラレータや計算機環境に対して、可能な限り共通の記述でアプリケーションを実行可能にすることを目指すものである。現在、複数のアプリケーションを用いた性能評価を進めており、異種環境における実行性能、プログラミング容易性、性能可搬性に関する知見を蓄積している。また、CPU と GPU が密に結合した近年の計算機アーキテクチャに着目し、NVIDIA GH200 および AMD MI300A (MCRP2025 の一部として提供された Sirius テスト利用による) を対象として、共有メモリ機構の性能評価を行った。これらのシステムでは CPU と GPU の間でメモリ空間を共有できるため、明示的なデータ転送の負担を軽減できる可能性がある。一方で、実際の性能はメモリアクセスパターンやデータ配置に依存するため、プログラミング環境がメモリ階層をどのように扱うべきかを検討する上で、基礎的な評価が重要となる。本研究で

は、共有メモリ利用時の基本的な性能特性を調査し、今後の UniSYCL の設計に反映するための知見を得た。さらに、ヘテロジニアスなメモリ構成を持つ Pegasus を用い、Intel Optane Persistent Memory を活用した Vlasov シミュレーションコードの大規模計算を実施した。その結果、異なる特性を持つメモリを組み合わせた環境でも、大規模な科学技術計算を実現できることを示した。本成果は UniSYCL を直接用いたものではないが、アクセラレータだけでなくメモリ階層の多様化も性能可搬性の重要な課題であり、統一的なプログラミング環境の設計に資する予備的知見である。

3. 学際共同利用プログラムが果たした役割と意義

本研究において、学際共同利用プログラムは、多様な計算機アーキテクチャに対する実機評価の機会を提供した点で大きな役割を果たした。特に、Pegasus、Miyabi-G、Sirius という、それぞれ異なる特徴を持つシステムを利用できたことにより、単一の環境では得られない比較評価が可能となった。

4. 今後の展望

今後は、CPU-GPU 共有メモリシステムに関する性能評価を継続する。現時点では基本的なメモリアクセスパターンを対象とした評価にとどまっているため、今後は実アプリケーションを用いた評価を進め、共有メモリ機構がアプリケーション全体の性能に与える影響を明らかにする。得られた知見は、UniSYCL におけるデータ配置、メモリ管理、実行制御の設計にフィードバックする。また、UniSYCL の研究開発を引き続き進める。複数のアクセラレータを統一的に扱うという基本方針の下、対象アーキテクチャやアプリケーションを拡大し、性能可搬性と実用性の両面から評価を深める。一方で、フロントエンド言語として SYCL を基盤とする方針が最適であるかについても検討を行う。最終的には、アクセラレータやメモリ階層の違いを開発者が過度に意識することなく、多様なスーパーコンピュータ上で高性能な科学技術計算を実行できるプログラミング環境の実現を目指す。

5. 成果発表

(1) 学術論文

- ① Norihisa Fujita, Keita Ito, Kohji Yoshikawa, Kohei Hiraga, Osamu Tatebe, Akira Nukada, and Taisuke Boku. 2026. Large-Scale Vlasov Simulations for Astrophysics using Non-volatile Memory as Large Memory. In SCA/HPCAsia 2026 Workshops: Supercomputing Asia and International Conference on High Performance Computing in Asia Pacific Region Workshops (SCA/HPCAsiaWS2026), January 26–29, 2026, Osaka, Japan. ACM, New York, NY, USA, 9 pages.
<https://doi.org/10.1145/3784828.3785352>

(2) 学会発表

- ① 藤田 典久, 吉川 耕司, 辻 美和子, 朴 泰祐, 建部 修見, AMD MI300A APU における共有メモリシステムの性能評価, 研究報告ハイパフォーマンスコン

ピューティング (HPC) , 巻 2026-HPC-203, 号 2, p. 1-8, 発行日 2026-03-09 (査読無し) .

- ② 藤田 典久, 伊藤 圭汰, 吉川 耕司, 額田 彰, 朴 泰祐, 不揮発性メモリを用いた Vlasov シミュレーションの大規模化, 研究報告ハイパフォーマンスコンピューティング (HPC) , 巻 2025-HPC-201, 号 7, p. 1-8, 発行日 2025-09-22 (査読無し) .
- ③ 吉田 智, 藤田 典久, 白井 拓翔, 朴 泰祐, 辻 美和子, NVIDIA GH200 における System-Allocated Memory の性能評価, 研究報告ハイパフォーマンスコンピューティング (HPC) , 巻 2025-HPC-199, 号 7, p. 1-12, 発行日 2025-05-05 (査読無し) .
- ④ 阿部 崇人, 佐藤 拓人, 朴 泰祐, 藤田 典久, 日下 博幸, ドライミスト効果を持つ都市気象コードの GH200 vs Xeon+H100 上の性能比較, 研究報告ハイパフォーマンスコンピューティング (HPC) , 巻 2025-HPC-199, 号 5, p. 1-11, 発行日 2025-05-05 (査読無し) .

(3) その他

使用計算機	使用計算機に○	配分リソース※		
		当初配分	移行*	一般利用による追加
Pegasus	○	2800		
Miyabi-G	○	9450		
Miyabi-C	○	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				