2025 年度年次報告書

先端国際共同研究推進事業
【2023 年度採択　次世代のための ASPIRE　半導体分野】

【研究課題名】
次世代の高効率計算基盤を実現する適応型データ圧縮ハードウェアの探求

研究代表者：
上野 知洋（理化学研究所　研究員）

主たる共同研究者：
長名 保範（熊本大学　准教授）
宮島 敬明（明治大学　専任講師）
小林 諒平（東京科学大学　准教授）
高瀬 英希（東京大学　准教授）
藤田 典久（筑波大学　助教）

相手側研究代表者：
Franck Cappello（Senior Computer Scientist/ R&D Leader, Argonne National Laboratory）
Jason Anderson（Professor, University of Toronto）

1. 研究進捗状況と成果

① 研究構想

本研究では、適応型データ圧縮ハードウェアの設計及び実現に関する研究と、CGRA アーキテクチャとその自動再構成を行う研究が主要なテーマであり、それぞれの進捗について以下に述べる。

本年度の適応型データ圧縮ハードウェアプラットフォームに関する研究進捗として、設計空間探索およびハードウェア設計の最適化が重要な成果である。前者では、データ圧縮ハードウェアプラットフォーム設計に量子化パラメータを導入して設計空間探索を行い、最適な圧縮性能と回路規模との関係を持つ設計を可能にした。後者では、FPGA ベースのシステムにおけるハードウェア設計の最適化として、大量の計算資源を必要とするバレルシフタの機能を、より小規模なバレルシフタとオンチップメモリへのアクセスに分散させ、必要な回路資源の大幅な削減と動作周波数向上によるスループットの向上にも成功した。その他にも、マルチチャネルデータ圧縮のための出力制御アルゴリズムの探索や、Smart NIC によるエッジデータ圧縮の検討、データ圧縮対象となるデータセットの探索、およびハードウェア化を考慮したデータ圧縮アルゴリズムの探索と評価を行った。

CGRA に関連する研究進捗としては、不均衡なデータフローグラフを CGRA にマッピングする際の性能低下を抑える研究と、マルチカーネル CGRA におけるカーネルの自動選択メカニズムについての研究を主に行った。前者では、CGRA アーキテクチャにおけるバッファの配置について、それによる全体のスループットの変化について、定量的に評価した。後者では、CGRA のデータフローファブリックとは別に制御用のインターコネクトを追加したマルチレイヤデザインを提案し、それらの性能比較を行った。これらの研究により、CGRA 上で自動的に最適なカーネルを選択する手法の実現に一歩近づいた。この他にも、CGRA バッファの柔軟な活用のためのプログラマブルバッファに関する研究や、デカップルド CGRA アーキテクチャの提案等も行った。

② 国際頭脳循環の促進

本年度も前年度に引き続き、大学院生および研究者の渡航および招聘を積極的に実施した。1 か月以上の中長期の渡航が 3 名、招聘が 2 名で、いくつかのケースでは滞在期間中に投稿論文が採択される等、大きな研究成果につながった。加えて、昨年度以上にワークショップ等の開催にも力を入れ、国際ワークショップを計 3 件、国際会議における BoF セッションを 1 件、その他招聘を伴うミーティングを 2 件実施した。これらの活動により、参加者間にとどまらない国際的なコミュニティの形成や研究者間の連携強化につながることが期待できる。

2. 代表的な成果事例

1. T. Ueno et al., "Evaluation of Trade-Off Between Compression Ratio and Hardware Cost for Adaptive Bandwidth Compression Hardware Platform," 2025 IEEE Symposium on Low-Power and High-Speed Chips and Systems (COOL CHIPS), Tokyo, Japan, 2025, pp. 1-6
   概要：量子化パラメータの導入によるデータ圧縮ハードウェアの設計空間探索と性能評価
2. Linyi Li et al., 2026. Performance and Area Optimization of Lossless Hardware Compression of Floating-Point Data Streams. In Proceedings of the Supercomputing Asia and International Conference on High Performance Computing in Asia Pacific Region Workshops (SCA/HPCAsiaWS '26). Association for Computing Machinery, New York, NY, USA, 284–292.

概要：データ圧縮ハードウェアの設計最適化による回路規模削減とスループット向上