

大量データ転送に向けたデータ転送方式

“ベクトルレジスタ共有機構”



Vector Register Sharing Mechanism: High Bandwidth Data Transfer Method for Accelerators

加藤倫也¹ 東良輔² 田中友章² 田中清史³ 長名保範⁴ 三好健文⁵ 多田十兵衛⁶ 中條拓伯⁷

¹東京農工大学大学院工学府知能情報システム工学専攻 ²東京農工大学大学院工学府情報工学専攻 ³北陸先端科学技術大学院大学先端科学技術研究科
⁴熊本大学半導体・デジタル研究教育機構 ⁵わさらぼ合同会社 ⁶山形大学大学院理工学研究科 ⁷東京農工大学大学院工学研究科先端情報科学部門

背景

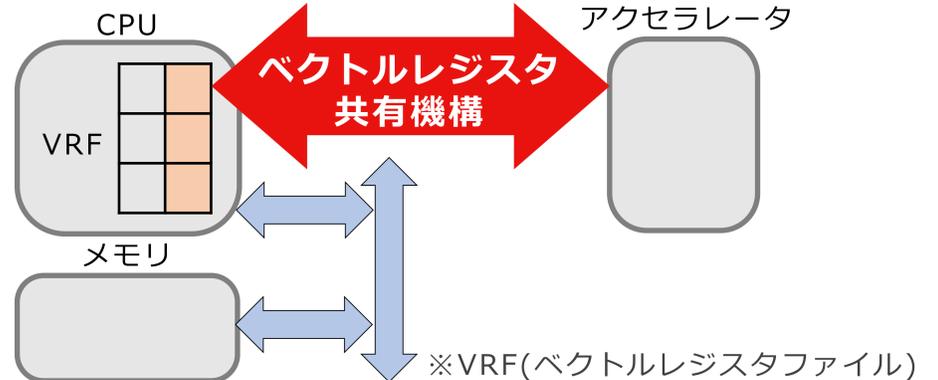
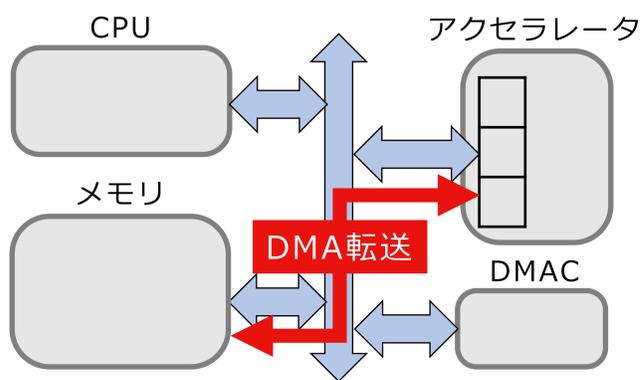
近年, AI処理をはじめとするデータレベル並列性が高い処理を高速化の対象としたアクセラレータが増加

➡ アクセラレータへ大量のデータの供給が必要



提案: ベクトルレジスタ共有機構

ベクトルプロセッサ中のベクトルレジスタをデュアルポート化し,
ベクトルレジスタ - アクセラレータ間で直接大量のデータのやり取りを可能に



実装

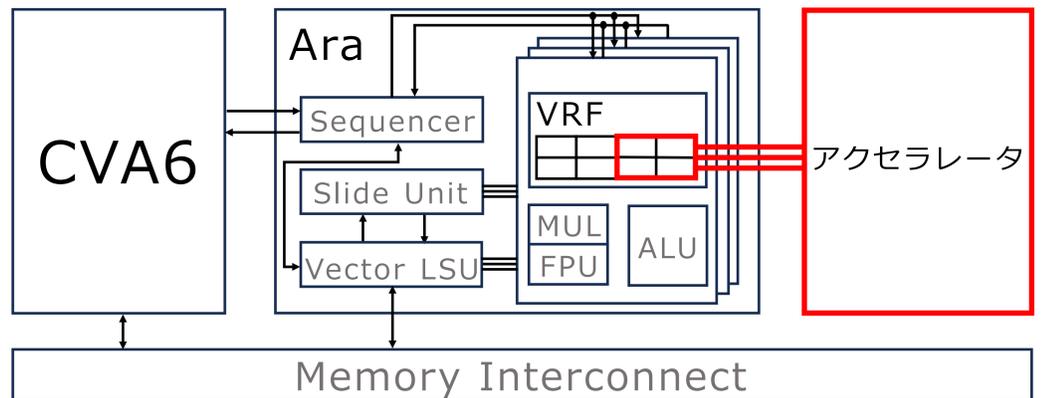
PULPが開発したCVA6+Araを使用してベクトルレジスタ共有機構を実現

CVA6: RV64GCインオーダープロセッサ

Ara: CVA6のコプロセッサ

RISC-VのV拡張(RVV)v1.0をサポート

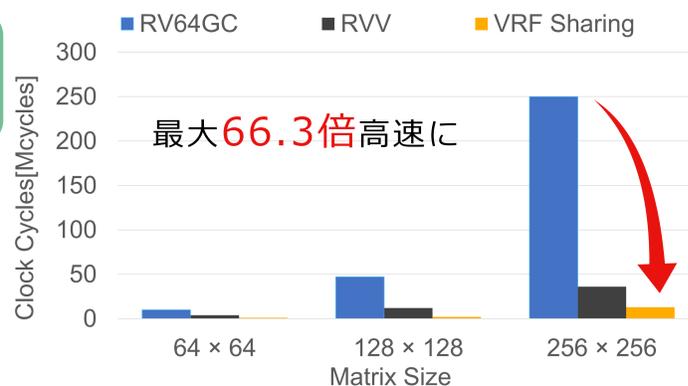
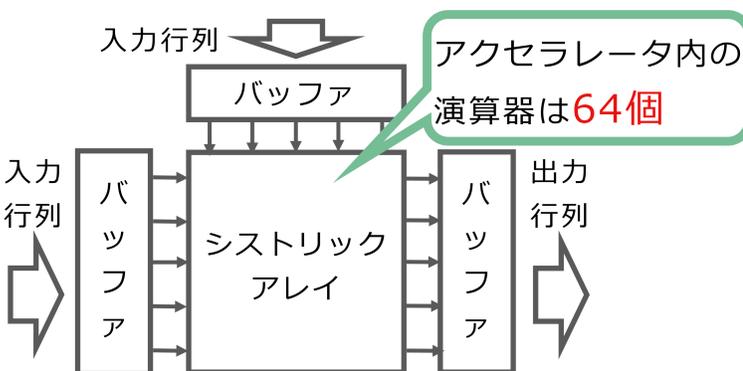
※今回用いたベクトル長: 4096bit



評価

RTLシミュレーションにてMATMUL実行時を例としてクロック数を計測して, 高速化の割合を比較

比較対象: **RV64G** vs **RVV** vs **ベクトルレジスタ共有機構**



アクセラレータの実行に十分なデータ転送量を実現

今後の展望

- 現在アクセラレータへの転送に広く用いられるDMAの実装, および比較評価の実施
- 比較評価に用いるテストベンチの拡充
 - データレベル並列性が高いアプリケーション
 - AI処理をはじめとした前処理・後処理を含んだプログラム