

2025 年度年次報告書

次世代エッジ AI 半導体研究開発事業・テーマ①「高効率自動設計による次世代 AI 回路・システム」

AI for Science のためのエッジの知能化加速

研究開発代表者：

泰地 真弘人

（理化学研究所 最先端研究プラットフォーム連携(TRIP)事業本部 プログラムディレクター）

主たる研究分担者：

佐野 健太郎（理化学研究所 計算科学研究センター チームプリンシパル）

西田 圭吾（理化学研究所 革新知能統合研究センター 基礎科学特別研究員）

新稲 亮（理化学研究所 生命機能科学研究センター 研究員）

近藤 正章（慶應義塾大学 理工学部 教授）

高山 順（東北大学 大学院医学系研究科 准教授）

1. 研究進捗状況と成果

コ・デザイン G では、1-A 分子動力学計算専用機の開発としてインテリジェント NoC のハードウェア開発を開始し、性能評価用ソフトウェアモデルにより通信量削減効果として水で 20 倍以上、タンパク質を含む系でも 4 倍以上を確認、コア技術の特許申請も行った。1-B では全ゲノム解析加速 ASIC に向け BWA-MEM のボトルネック解析を実施、1-C では低精度バイズ学習法 LMD の評価を進め ICML 2026 に投稿した。CGRA G では、コンフィギャラブル CGRA 設計の確認、AI 支援型設計空間探索フレームワークの提案、SoC 設計とエミュレーション環境構築、コンパイラを含むエコシステム整備と基本ベンチマーク開発を達成度◎で実施。3DLSI G では、3 次元実装メモリ向け高効率行列積演算アクセラレーションユニットを設計し、MD Cube テープアウトに向けた検証を進めた。

2. 代表的な成果事例

[1] Jiaheng Liu, Chenlin Shi, Boma Adhi, Shinobu Miwa, Kentaro Sano, "RCM-SPM: Reduced-Connectivity Multibank Scratchpad Memory for RIKEN CGRA," 組込み技術とネットワークに関するワークショップ ETNET2026 信学技報, vol. 125, no. 401, CPSY2025-58, pp. 35-40, 2026 年 3 月.

概要：エラスティックな粗粒度再構成可能アレイ (CGRA) は、バックプレッシャー機構により静的スケジューリングやタイル単位の厳密なレイテンシ仮定を排し、自律・分散型のフロー制御を実現する。このエラスティック特性は、動的かつデータ依存の可変遅延を持つ外部モジュールとの結合を容易にし、オンチップスクラッチパッドメモリ (SPM) の設計空間を広げる。本研究では、RIKEN CGRA を対象に複数の SPM 設計を評価し、面積・消費電力に加え、代表的なカーネルにおける演算スループットおよびデータ重複への影響を分析した。評価の結果、完全接続マルチバンク SPM は相互結合網に大きな電力・面積を要する一方、局所接続設計ではワーキングセットを共有するカーネルの場合に異なる SPM でより多くのデータが重複するというトレードオフが明らかになった。そこで本稿では、各ポートがバンクの一部にのみアクセスする削減接続型マルチバンク SPM (RCM-SPM) を提案する。RCM-SPM は、ポートあたりのバンクアクセス範囲を制限することで相互結合のオーバーヘッドとデータ重複を調整しつつ、対象ワークロードのスループットを概ね維持可能である。