# [技術展示] 可変パイプラインを持つ低消費電力アクセラレータ CCSOTB2によるストリーム処理

天野 英晴†

† 慶應義塾大学大学院 理工学研究科 223-8522 神奈川県横浜市港北区日吉 3-14-1 E-mail: †{tkojima,ando,matsushita,hayate,doan,hunga}@am.ics.keio.ac.jp

**あらまし** 粗粒度再構成可能アーキテクチャ (CGRA: Coarse-Grained Reconfigurable Architecture) は高い電力効率 とプログラマビリティを兼ねそなえるアーキテクチャとして注目を集めている。このような特徴から CGRA は汎用 の CPU に代わり、アプリケーションの compute-intensive な部分を実行するアクセラレータとしての利用が期待され ている。一般的な CGRA ではクロックサイクル単位の動的再構成をサポートしているが、画像処理や信号処理のよう なストリーム処理ではクロックサイクル単位の再構成を必要としないことが多い。そこで、再構成の頻度を下げ、動 的再構成の電力オーバーヘッドを削減した CMA(Cool Mega Array) アーキテクチャが提案されている。しかしなが ら、既存の CMA は大きなクリティカルパス遅延が原因でスループット向上に限界があった。そこで、PE アレイをパ イプライン化した VPCMA(Variable Pipelined CMA) が提案された。VPCMA ではパイプライン化による電力オー バーヘッドを最小限に抑えるために、要求性能に応じで可変的にパイプライン構造を変化させることができる。本稿 では、VPCMA アーキテクチャを SOTB 65nm プロセスで実チップ化した CCSOTB2 を用いて実画像処理を行い評 価を行う。

キーワード CGRA, 粗粒度再構成可能アーキテクチャ, SOTB, ストリーム処理

# [Technology Exhibit] Low Power Stream Processing on a Variable Pipelined Accelerator CCSOTB2

Takuya KOJIMA<sup>†</sup>, Naoki ANDO<sup>†</sup>, Yusuke MATSUSHITA<sup>†</sup>, Hayate OKUHARA<sup>†</sup>,

Ng. Doan Anh Vu<sup> $\dagger$ </sup>, and Hideharu AMANO<sup> $\dagger$ </sup>

† Graduate School of Science and Technology, Keio University Hiyoshi 3–14–1, Kohoku-ku, Yokohama, Kanagawa, 223–8522 Japan

# 1. はじめに

IoT 時代の到来により、インターネットに接続された組み込 み機器によって多種多様なサービスが実現されている。しかし、 これらの処理を全てクラウド側のサーバーに一任することで、 レイテンシや通信路負荷の増大、セキュリティの低下など問題 を招いてしまう。これらの問題を解消するために、ある程度の 処理を端末側で行うエッジコンピューティングが注目されてい る。エッジコンピューティング実現のためには、十分に高い処 理能力と低消費電力性の両方が必要とされ、汎用 CPU よりも 効率的に計算を行うアクセラレータの利用が期待される。

粗粒度再構成可能アーキテクチャ CGRA(Coarse-Grained Reconfigurable Architecture) はこれらの要求を満たす一つの デバイスである。一般に CGRA は多数の PE(Processing Element) が2次元のアレイ状に配置され、相互接続ネットワークで 互いに接続される。PE は単純な ALU、SE (Switch Element) を持ち、ALU で実行される演算と相互接続を再構成すること で効率的に処理を行うことができる。再構成粒度がビットレ ベルの FPGA (Field Programmable Gate Array) と比べ、柔 軟性は低下するものの ASIC(Application Specific Integrated



図 1: PE アレイの概要

Circuit) に近い電力効率を達成する。

クロックサイクル単位で PE アレイの再構成を行う汎用的 な CGRA として、学術界からは ADRES [1]、MorphoSys [2]、 FloRA [3]、DRRA [4]、また、産業界からは ADRES をベース とした Samsung Reconfigurable Processor (SRP) [5] などが提 案されている。

しかし、これでサポートされているクロックサイクル単位で の再構成はダイナミック電力を多く消費することが問題であ る。そこで、再構成による電力オーバーヘッドを大幅に削減 した低消費電力を特徴とする CGRA として CMA(Cool Mega Array)が提案された [6]。CMA ではクロックサイクル単位で の再構成を行わず、静的な構成を利用する。また、PE はレジ スタファイルを持たず、PE アレイは完全な組み合わせ回路で 構成されるため、クロックツリーを必要としない。これによる PE アレイの柔軟性低下を補うために、小規模なマイクロコン トローラを備えており、PE アレイとデータメモリの間の柔軟 なデータ転送を行う。

最初のプロトタイプ CMA-1 [6] が試作された後、リーク電力 と性能のバランスを調整可能にするために FD-SOI (Fully Depleted Silicon On Insulator) の一種である 65nm の SOTB (Silicon on Thin Buried Oxide) プロセスを用いて CMA-SOTB [7] および CMA-SOTB2 [8] が開発された。CMA-SOTB2 では 743MOPS/mW の高いエネルギー効率が達成された。

さらに、巨大な組み合わせ回路である PE アレイの大きな 遅延時間を隠蔽するために、ウェーブパイプラインという機 能を備えた CMA-CUBE-SOTB(CC-SOTB) [9] が試作された。 ウェーブパイプラインはレジスタを利用せずに PE アレイへ一 定の間隔でデータを入力し、パイプライン処理を行う。これに よって、CMA-SOTB2 と比較して、1.9 倍の性能向上を達成 した。

しかしながら、ウェーブパイプラインはデータの入力間隔が

#### Chip Interface External Bus Config. Inst. DMAC Constant Controller Data Mem Register Mem Config. Registers Micro PE Array Controller

図 2: VPCMA の構成

短すぎると、先行して処理されていたデータと衝突し、正しく 計算ができない可能性があり、その入力間隔を事前に見積もる のも困難であるという問題があった。加えて、巨大な組み合わ せ回路ゆえにグリッチと呼ばれる不要なスイッチング電力の消 費が問題となっていた。

そこで、PE アレイをパイプラインレジスタを用いてパイプ ライン化した VPCMA (Variable Pipeline Cool Mega Array) が考案された [10]。これによって安全にパイプライン処理を行 うことができ、パイプラインレジスタの存在によりグリッチの 影響を抑制することができる。パイプライン化による電力オー バーヘッドを抑えるために再構成可能なパイプラインとなって おり、要求性能に応じてパイプライン段数を変化させることが 可能である。VPCMA はこれまでと同様の SOTB プロセスを 用いて実チップ CCSOTB2 が試作され、すでに動作が確認さ れている [11]。

本研究では CCSOTB2 を用いて比較的単純なストリーム処 理を実画像データを用いて行う。また、本稿ではソフトウェア 開発フローや実験環境の構築に関しても合わせて報告する。



# 2. アーキテクチャ

本節ではまず改良型の CMA である VPCMA アーキテクチャ を説明する [10]。図 2 に示すように、VPCMA アーキテクチャ はいくつかの機能ブロックに分けられる。各ブロックは共通の 外部データバスによって接続され、共通の 22bit 幅アドレス空 間にマップされている。オフチップのホストコントローラは チップインターフェースを通してこれらのブロックにアクセス することができる。

## 2.1 PE アレイとマイクロコントローラ

次に、演算処理部である PE アレイとその周辺要素を説明す る。図1はそれらのブロック図である。8×12の PE アレイ、 マイクロコントローラ、データマニピュレータ、そしてオン チップのデータメモリ (スクラッチパッドメモリ) が存在する。 パイプライン化されていない CMA と異なり、VPCMA は各 PE 行の間に計7本のパイプラインレジスタが挿入されている。

PE アレイで計算を開始するには入力データを Fetch Register へ書き込む。数クロック後 (パイプラインレイテンシ) に計算結 果が Gather Register へ格納される。

図1に示すパイプラインレジスタによって、VPCMAのパイ プライン構造は再構成可能なものになっている。コンフィギュ レーションデータにはパイプライン構造に関する情報も含まれ ている。各パイプラインレジスタはラッチとして利用するか、 または単純なデータパスとして利用するかが選択できる。これ によって、マルチプレクサはレジスタの値もしくは前段 PE か らの値のどちらかを後段 PE へ転送する。パイプラインレジス タがラッチとして利用されない場合、これらはクロックゲー ティングが施されダイナミック電力の消費を抑えることがで きる。

マイクロコントローラは Fetch Register と 12 のバンクに分 けられたデータメモリとの間のデータ転送を制御する。同様に、 Gather Register とデータメモリのとの間のデータ転送も制御 する。マイクロコントローラは小規模な RISC プロセッサであ り、単純なアセンブリで動作をプログラムすることができる。 データマニピュレータは 12 入力-12 出力のクロスバスイッチの ような置換ネットワークであり、転送テーブルの情報に従い、各 出力ポートは任意の入力ポートからのデータを転送することが できる。転送テーブルは Fetch Register 用と Gather Register 用で各々 16 個利用することができる。Fetch 命令及び Gather 命令はこのテーブル番号を指定するためのオペランドを持って おり、様々な転送パターンを容易に切り替えることができる。

## 表 1: CCSOTB2 の実装

設計	Verilog HDL		
プロセス	Renesas SOTB 65 nm		
ライブラリ	LPT-8		
標準電源電圧	$0.55 \mathrm{~V}$		
論理合成	Synopsys Design Compiler		
	2016.03-SP4		
配置配線	Synopsys IC Compiler		
	2016.03-SP4		
チップサイズ	$6$ mm $\times 3$ mm		



図 4: CCSOTB2 のチップ写真



図 5: システムの概要図

#### 2.2 実チップ CCSOTB2 の実装

我々は VPCMA アーキテクチャを SOTB 65nm プロセス を利用して実チップ CCSOTB2(CMA-Cube-SOTB2) を試作 した。SOTB は FD-SOI の一種であり、図 3 に示すように薄 い BOX 層の上にトランジスタが形成される。FD-SOI では ボディへ印加するする電圧を変化させることで性能とリーク 電力のトレードオフを広い範囲で調整することができる [12]。 CCSOTB2 ではパイプライン構造の制御だけでなくボディバイ アス制御によってさらに電力オーバーヘッドを削減することが 可能である。

CCSOTB2 の実装環境を表1にまとめる。また、完成した CCSOTB2 のチップ写真を図4に示す。黄色の枠で囲まれた部 分が PE 行であり、赤色の枠で囲まれた部分は誘導結合を用いた 無線接続のためのインターフェース TCI(ThruChip Interface) である。TCI を用いることで複数枚のチップを積層して利用す ることができるが、本研究では単一の CCSOTB2 チップを利 用することを想定する。



(a) 入力画像 (カラー)

(b) 20MHz, 0.55V の時の結果

(c) 20MHz, 0.45V の時の結果



(d) 20MHz, 0.40V の時の結果

(e) 35MHz, 0.55V の時の結果

図 6: 実験結果

表 2	2: シ	ス	テ	ム	構築環境
-----	------	---	---	---	------

利田ボード	Avnet 社 MicroZed		
ጥባ/11 4 V · · · · · ·	AES-Z7MB-7Z020-SOM		
搭載 FPGA	Xilinx Artex-7		
設計	Xilinx Vivado 2018.2		
Linux ビルド	Xilinx PetaLinux 2018.2		
ARM クロスコンパイラ	Linaro gcc 7.2.1		

## 3. ソフトウェア開発環境

CCSOTB2 はプログラム可能なマイクロコントローラを備 えているものの、あくまでアクセラレータであり、ホストコン トローラが外部から制御する必要がある。そこで、本研究では CPU として ARM を搭載した Zynq-7000 の FPGA をホスト コントローラとして利用する。構築したシステムの概要を図 5 に示す。

Zynq の PL 部分 (FPGA) では CCSOTB2 への各種制御信 号やデータの入出力を行うためのインターフェースとして利用 する。CCSOTB2 は試作チップであり、利用可能な IO ピンの 数に限りがあったことからデータの入出力はシリアルインター フェースを通して行われる。また、CCSOTB2 とホストコント ローラは 35bit を 1flit とするパケット形式でデータを交換す る。そのため、FPGA 上のインターフェースは SERDES およ び入出力バッファを備えている。

PS 部分では Xilinx 社の提供する PetaLinux を用いてビル ドした Linux OS を稼働させ、前述のインターフェースを利用 するためのデバイスドライバなどを API としてまとめた。こ の API を利用することで、C 言語で記述したプログラムから



図 7: 実験環境

CCSOTB2を容易に制御、利用することが可能となる。

CCSOTB2 の PE アレイで実行されるアプリケーションは dot 形式のデータフローグラフで記述される。このデータフ ローグラフは専用のコンパイラによって最適化される[11]。こ の最適化では膨大な計算を必要とするため、これらの処理はオ フチップのハイエンドサーバーなどで行われる。この最適化 ツールはコンフィギュレーションデータなどの必要な情報を C 言語で利用可能な形でエクスポートすることが可能である。

### 4. 実 験

#### 4.1 実験環境

3. 節で説明したシステムを実現するために、CCSOTB2 と Zynq-7000 の FPGA を接続するためのマザーボードを開発し た。図7にこのマザーボードを含む実験環境を示す。

## 4.2 実画像処理による評価

本研究では構築した実験環境の上で、簡単なストリーム処理 を行う。そこで、画像処理として RGB 画像をグレースケール 化するアプリケーションをケーススタディとして利用する。い くつかの動作周波数、電源電圧でグレースケールのアプリケー ションを実行し、出力結果を記録した。得られた結果を図 6 に 示す。

入力画像 (図 6a) は 300x400 サイズのカラー画像であり、出 力も同じサイズである。図 6b は標準電源電圧の 0.55V を与え、 クロック周波数 20MHz にした場合の実行結果である。この場 合、正しく処理を行うことができていることがわかる。次に、同 じ周波数 20MHz で電源電圧を 0.45V に下げた場合の結果が図 6c である。PE アレイで動作中のアプリケーションは 20MHz、 標準電圧で動作させる想定で最適化されているものであるた め、PE アレイでタイミング違反が発生し、一部の画素データ に不備が見られる。さらに、電源電圧を 0.40V まで下げると、 図 6d のような結果が得られた。図 6c と比較するとより多くの 画素データが正しくない結果となっていることがわかる。

また、[11] で報告されているように、VPCMA の現在の設計 ではマイクロコントローラが動作周波数のボトルネックとなっ ており、実チップ CCSOTB2 で確認されている最大動作周波 数は 30MHz である。そのため、図 6e に示すようにクロック周 波数を 35MHz に設定すると、タイミング違反によって PE ア レイの出力結果を取りこぼしている時がある。図中の黒い部分 は計算結果を Gather Register から正しくデータメモリへ下記 戻すことができていない部分である。

### 4.3 消費電力

図 6b の出力結果が得られた電源電圧 0.55V, 動作周波数 20MHz において計算時の消費電力は約 2.6mW であった。

## 謝 辞

本研究は、JSPS 科研費 (S) 誘導結合を用いたビルディング ブロック型計算システムの研究 (25220002)、および JSPS 科研 費 (B) ビルディングブロック型計算システムにおけるチップブ リッジを用いた積層方式 (18H03125) の助成を受けたものであ る。また、東京大学大規模集積システム設計教育研究センター を通し、シノプシス株式会社の協力で行われたものです。関係 者の皆様に感謝致します。

#### 献

文

- B. Mei, F.-J. Veredas and B. Masschelein: "Mapping an H. 264/AVC decoder onto the ADRES reconfigurable architecture", Field Programmable Logic and Applications, 2005. International Conference onIEEE, pp. 622–625 (2005).
- [2] H. Singh, M.-H. Lee, G. Lu, F. J. Kurdahi, N. Bagherzadeh and E. M. Chaves Filho: "MorphoSys: an integrated

reconfigurable system for data-parallel and computationintensive applications", IEEE transactions on computers, 5, pp. 465–481 (2000).

- [3] D. Lee, M. Jo, K. Han and K. Choi: "FloRA: Coarsegrained reconfigurable architecture with floating-point operation capability", Field-Programmable Technology, 2009. FPT 2009. International Conference onIEEE, pp. 376–379 (2009).
- [4] M. A. Shami and A. Hemani: "Partially reconfigurable interconnection network for dynamically reprogrammable resource array", ASIC, 2009. ASICON'09. IEEE 8th International Conference onIEEE, pp. 122–125 (2009).
- [5] M. Konijnenburg, Y. Cho, M. Ashouei, T. Gemmeke, C. Kim, J. Hulzink, J. Stuyt, M. Jung, J. Huisken, S. Ryu, et al.: "Reliable and Energy-Efficient 1MHz 0.4V Dynamically Reconfigurable SoC for ExG Applications in 40nm LP CMOS ", Proc. of ISSCC, p. 24.6 (2013).
- [6] N. Ozaki, Y. Yasuda, M. Izawa, Y. Saito, D. Ikebuchi, H. Amano, H. Nakamura, K. Usami, M. Namiki and M. Kondo: "Cool Mega-Arrays: Ultralow-Power Reconfigurable Accelerator Chips", IEEE Micro, **31**, 6, pp. 6–18 (2011).
- [7] H. Su, Y. Fujita and H. Amano: "Body bias control for a coarse grained reconfigurable accelerator implemented with Silicon on Thin BOX technology", Field Programmable Logic and Applications (FPL), 2014 24th International Conference onIEEE, pp. 1–6 (2014).
- [8] K. Masuyama, Y. Fujita, H. Okuhara and H. Amano: "A 297mops/0.4mw ultra low power coarse-grained reconfigurable accelerator CMA-SOTB-2", 2015 International Conference on ReConFigurable Computing and FPGAs (Re-ConFig), pp. 1–6 (2015).
- [9] Y. Matsushita, H. Okuhara, K. Masuyama, Y. Fujita, R. Kawano and H. Amano: "Body bias grain size exploration for a coarse grained reconfigurable accelerator", 2016 26th International Conference on Field Programmable Logic and Applications (FPL), pp. 1–4 (2016).
- [10] N. Ando, K. Masuyama, H. Okuhara and H. Amano: "Variable Pipeline Structure for Coarse Grained Reconfigurable Array CMA", 2016 International Conference on Field-Programmable Technology, pp. 231–238 (2016).
- [11] T. Kojima, N. Ando, Y. Matshushita, H. Okuhara, N. A. V. Doan and H. Amano: "Real Chip Evaluation of a Low Power CGRA with Optimized Application Mapping", Proceedings of the 9th International Symposium on Highly-Efficient Accelerators and Reconfigurable TechnologiesACM, p. 13 (2018).
- [12] Ishigaki, Takashi and Tsuchiya, Ryuta and Morita, Yusuke and Sugii, Nobuyuki and Kimura, Shin' ichiro: "Ultralowpower LSI Technology with Silicon on Thin Buried Oxide (SOTB) CMOSFET", Solid State Circuits Technologies, Jacobus W. Swart (Ed.), ISBN: 978-953-307-045-2, InTech, pp. 146–156 (2010).