

パイプライン段数とボディバイアス電圧制御によるパイプライン型 CGRA の電力削減手法の検討

小島 拓也[†] 安藤 尚輝[†] 松下 悠亮[†] 奥原 颯[†] 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,hunga}@am.ics.keio.ac.jp

あらまし IoT時代においてウェアラブルデバイスなどのバッテリー駆動の高機能で小型なデバイスが求められている。こうしたデバイスは一定以上の性能を低消費電力で実現することが要求される。そのためエネルギー効率の高い粗粒度再構成可能アーキテクチャ (CGRA: Coarse-Grained Reconfigurable Architecture) が注目されている。この中でも CGRA の 1 種である VPCMA (Variable Pipelined Cool Mega Array) は可変パイプラインを最適化することによって高い性能向上とエネルギー効率が得られている。しかし、この VPCMA は極薄 BOX 層を持つ完全空乏型の Silicon On Insulator (SOI) を用いてボディバイアス制御を行うことでさらなる電力効率を向上させることができる。そこで本研究では VPCMA に対してパイプライン段数とボディバイアス制御を同時に行うことでさらなる電力効率を改善する。シミュレーションの結果、PE アレイに対して行毎のボディバイアス制御を行うと、ゼロバイアス時と比べて平均約 19% の電力削減率を得た。また、PE アレイ全体で一律のボディバイアス制御をする場合と比べると電力削減率は平均約 10% であった。

キーワード 電力削減, SOTB, 粗粒度再構成可能アクセラレータ, ボディバイアス制御, 可変パイプライン

Power Reduction for Pipelined CGRA with the Controlling Variable Pipeline and the Body Bias Voltage

Takuya KOJIMA[†], Naoki ANDO[†], Yusuke MATSUSHITA[†], Hayate OKUHARA[†], and Hideharu AMANO[†]

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

E-mail: †{tkojima,ando,matsushita,hayate,hunga}@am.ics.keio.ac.jp

1. はじめに

高度な処理を要求するウェアラブルデバイスでは高い演算性能が要求される一方で、長いバッテリー寿命で動作することが望ましい。そこでエネルギー効率の高さから粗粒度再構成可能アーキテクチャ (CGRA: Coarse-Grained Reconfigurable Architecture) が注目されている。CGRA は PE (Processing Element) と呼ばれる小さな演算処理部をアレイ状に多数持っており、PE 間の相互接続や演算の種類などの構成を適応的に変化させることで高いエネルギー効率を得ることができる。

我々は CGRA の 1 種として CMA (Cool Mega-Array) アーキテクチャを開発した [1]。CMA は PE アレイからクロックツリーを取り除き、動的な再構成を行わないことで消費電力の削

減を図っている。さらに CMA の低電力化を図るために FD-SOI^(注1) の 1 種である Silicon On Thin Buried oxide (SOTB) を使用した CMA-SOTB [2]、CMA-SOTB2 [3]、CMA-CUBE-SOTB (CC-SOTB) [4] を開発し、評価を行ってきた。SOTB プロセスではボディバイアス電圧を印加することで遅延とリーク電力のトレードオフを制御でき、CMA の性能と電力の最適化を行うことが可能である。

さらなるスループット向上を目指し PE アレイのパイプライン化を検討し、VPCMA (Variable Pipelined CMA) と呼ぶ新たな CMA アーキテクチャを提案し、評価を行った [5]。CC-SOTB

(注1): Fully Depleted Silicon On Insulator

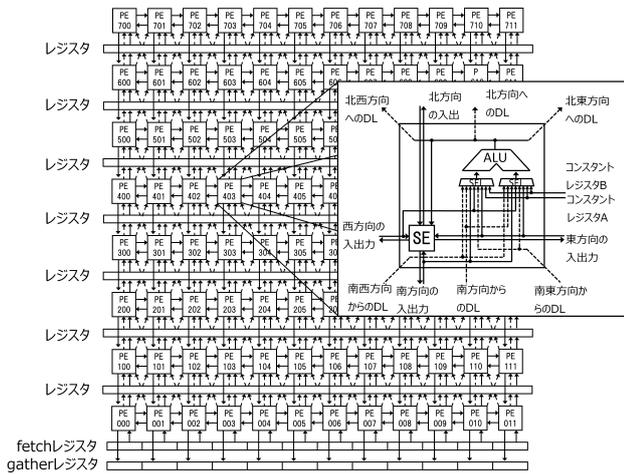


図 2 PE アレイの構成と PE の内部構造

へ送られる。ダイレクトリンクは北、北西、北東方向の PE へ接続されている。ALU への入力セレクタは東、西、南方向からの入力、南、南東、南西方向からのダイレクトリンク、および定数値レジスタの中から演算に使用するデータを選択する。図 2 において PE 間の相互接続を見ると行をまたぐ接続—3 つのダイレクトリンク、北方向への出力および南方向からの入力—はパイプラインレジスタを経由することがわかる。ただし、南方向への出力、北方向からの入力はパイプラインレジスタを経由していない。

3.1.2 μ コントローラ

μ コントローラは固定長 16bit の命令を実行する小さなプロセッサであり、PE アレイとデータメモリの間のデータ転送などを行う。データマニピュレータと呼ぶ機構を持っていて PE アレイとの複雑なデータ転送を行うことができる。

3.1.3 コンフィギュレーションデータ

コンフィギュレーションデータとは PE アレイ内にある各 PE とパイプラインレジスタの構成情報を指定するものである。PE アレイの外部にコンフィギュレーションレジスタが配置されており、そこから PE アレイへ送信される。使用するパイプラインレジスタの情報をコンフィギュレーションデータに含めることでアプリケーションごとに可変なパイプライン段数を使用することが可能となる。すでに述べているようにアプリケーション実行中、動的にコンフィギュレーションデータは変化しない。

3.2 パイプライン型 CGRA

VPCMA は以下のような特徴を持つパイプライン型 CGRA に分類することができる。また、パイプライン型 CGRA のアーキテクチャの概念図を図 3 に示す。

- 単純なデータフローのみを扱う
- パイプライン的な動作が可能である
- 再構成の頻度が低い

PE 間の相互接続のパターンを制限することにより、データフローが特定の方向に流れるように単純化されている。このため PE アレイの入出力は PE アレイの端に配置され柔軟性が低下してしまう。こうした柔軟性低下に対処するために、データ

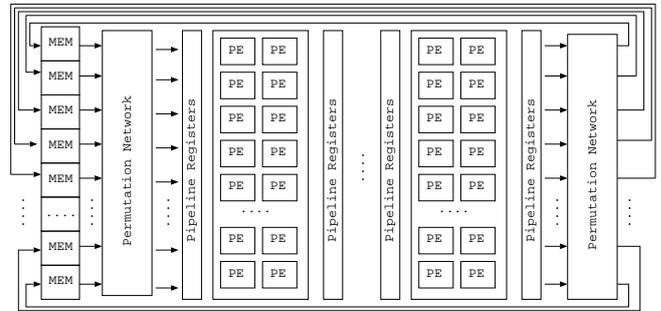


図 3 パイプライン型 CGRA の概念図

メモリと PE アレイとの間にクロスバのような機構を備えていることが多い。

VPCMA 以外のアーキテクチャでパイプライン型 CGRA に分類されるアーキテクチャは PipeRench [8], S5 Engine [9], XPP [10], DT-CGRA [11] が提案されている。

マッピングされる演算によって各パイプラインステージには遅延のばらつきが発生する。パイプラインステージ単位でボディバイアス制御を行うことでこのばらつきを調整することができる。さらに、要求される性能が低い場合はパイプラインステージの遅延が増加するような構成も許される。ゆえに、演算のマッピングや相互接続などの構成情報を変化させてパイプラインステージにおける遅延を調節することと、同時にボディバイアス電圧を調節することが可能である。本研究で対象とした VPCMA ではパイプラインステージの遅延を調節する手法としてパイプライン段数を変化させている。

4. 電力効率改善手法

VPCMA においてはボディバイアス電圧、パイプライン段数を変化させることで表 1 に示すトレードオフが存在する。ボディバイアス電圧のトレードオフは 2. 節で述べた通りである。VPCMA ではパイプライン段数を増やすとパイプラインレジスタの電力が増加する一方で、グリッチによる電力は低下する。これは PE の行で発生したグリッチによるスイッチングが次の行へ伝搬されなくなるためである。

表 1 トレードオフのまとめ

ボディバイアス電圧		
	VBB > 0	VBB < 0
動作周波数	向上	低下
リーク電力	増加	減少
パイプライン段数		
	大	小
動作周波数	向上	低下
スループット	向上	低下
グリッチの影響	低下	向上
パイプラインレジスタの電力	増加	減少

こうした複雑なトレードオフが存在するため、本研究では以下のような制約で電力最小化を行う。

- ブルートフォース探索
- ボディバイアス制御のドメインは最小粒度のパイプラインステージ (PE アレイの行単位)

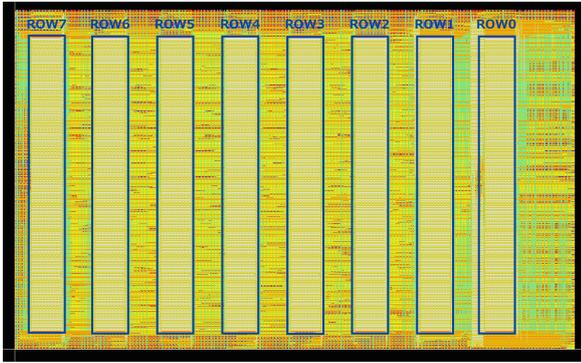


図 4 PE アレイのレイアウト図

表 2 予備評価の環境

設計	verilogHDL
プロセス	LEAP65nm/LPT-8
論理合成	Synopsys Design Compiler 2016.03-SP4
配置配線	Synopsys IC Compiler 2016.03-SP4
電源電圧	0.55V
温度	25°C
遅延、リーク電力 シミュレーション	Synopsys HSIM 2012.06-SP2
ダイナミック電力	Synopsys Prime Time 2012.12-SP3
動作シミュレーション	Cadence NC-Verilog 10.20-s131

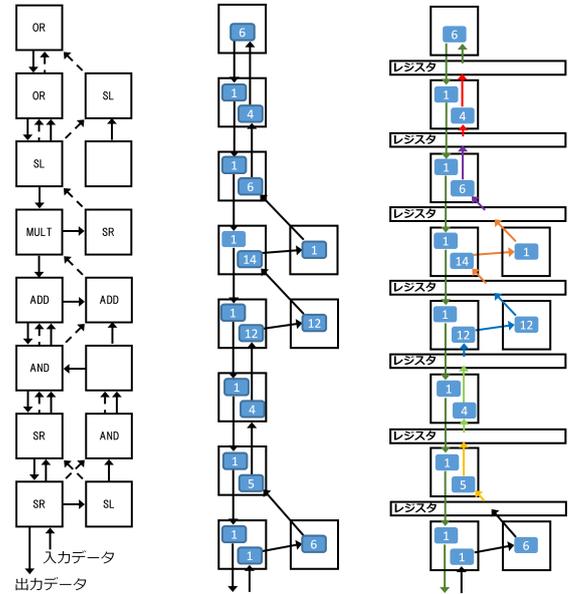
• パイプライン段数は 1 段, 2 段, 4 段, 全段の 4 パターン
ボディアバイアス制御のドメインとは同一のボディアバイアス電圧を印加する領域のことである。つまり、PE アレイにおいて同一行の PE は同じボディアバイアス電圧になるが、行単位で異なるボディアバイアス電圧を与えることができる。パイプラインレジスタは 7 本あるため分割のパターンは $2^7 = 128$ 通りある。しかしながら、計算量の関係から上記の 4 つのパターンのみを検討することにした。全段とは最小粒度にパイプライン分割したときのことであり、使用している行全てを 1 行ずつに分割したときのことである。アプリケーションによって使用している行数が異なるため 8 段とは限らない。例えばアプリケーションに *sepia* を選んだ場合、6 行しか使用していないため全段とは 6 段を意味する。

4.1 予備評価

各パイプライン分割、ボディアバイアス電圧のパターンで達成可能な性能と消費電力の 2 つをシミュレーションする必要がある。したがって、PE アレイのレイアウトを行い予備評価として以下のシミュレーションを行った。予備評価のための評価環境を表 2 に示す。

- (1) PE アレイの行単位でのリーク電力
- (2) PE での遅延時間
- (3) PE アレイでのダイナミック電力

1、2 はボディアバイアス電圧を -2.0~0.4V の範囲で 0.2V 毎にシミュレーションを行った。3 は実行するアプリケーション、パイプライン分割のパターン毎にシミュレーションを行った。



(a) マッピング例 (b) 非パイプライン (c) 8 段分割

図 5 演算のマッピングと遅延の例

実行したアプリケーションを表 3 に示す。

表 3 シミュレーションしたアプリケーション

アプリケーション	内容	PE アレイ使用行数
af	24bit(RGB) アルファブレンダ	7 行
sf	8bit セピアフィルタ	8 行
sepia	24bit(RGB) セピアフィルタ	6 行
gray	24bit(RGB) グレースケール	8 行

4.2 シミュレーション方法

予備評価の結果をもとに動作周波数 f を決定するには PE アレイ内の全データフローの中で最も大きい遅延時間 D_{max} を計算し、 $f = 1/D_{max}$ とする。最大遅延時間 D_{max} の求め方の手順を以下に示す。

- (1) PE アレイ中のデータパスを抽出
- (2) 各データパス、各ステージの遅延時間を求める
- (3) 求めた遅延時間のなかで最大のものを決定する

例えば、図 5(a) のような演算のマッピングを考える。このマッピングにおいては複数のデータパスが考えられるが一例として図 5(b)(c) のようなデータパスを用いる。図 5(b)(c) の数値は各 PE における遅延時間を表している。PE 内に遅延時間が 2 つ示されているが、右下にある遅延時間は ALU を使用して演算を行う場合のもので、左上にある遅延時間は出力データがスイッチエレメントのみを経由するときのものである。図中の矢印はデータの流れを示している。

パイプライン分割しない場合の遅延時間を D_1 とし、8 段パイプラインの n 番目のステージでの遅延を $D_8(n)$ とする。最大遅延は以下のように計算される。

$$D_1 = 78$$

$$D_8(4) = 12 + 12 = 24$$

このようにパイプライン分割しない 1 段の場合遅延が 78 であり、8 段の場合最大遅延が 24 であると計算できる。

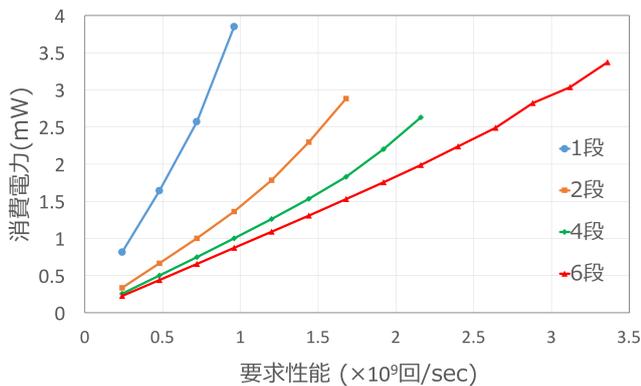


図 6 sepia: 電力最小化の結果

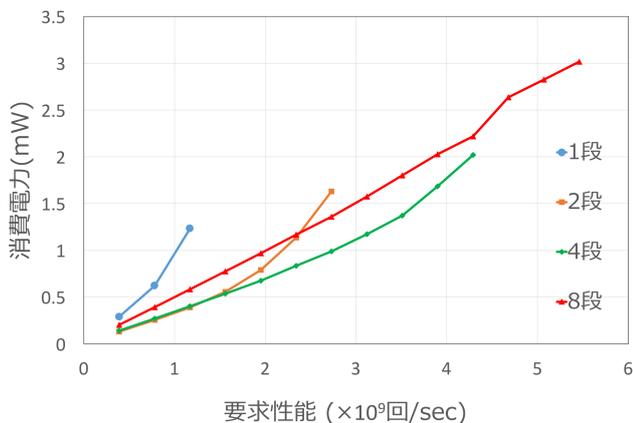


図 7 gray: 電力最小化の結果

5. 評価

5.1 パイプライン分割の段数と電力最小化の関係

4つのパイプライン分割パターン(1段, 2段, 4段, 全段)でボディバイアスを制御し電力を最小化した結果を図6, 7に示す。横軸の要求性能はPEアレイで1秒間に実行できる演算の回数であり、動作周波数 f と演算がマッピングされているPEの数との積で計算される。縦軸はPEアレイにおける消費電力である。

図6に示すsepiaでは6段にパイプライン分割する場合が常に電力最小となっている。図7に示すgrayでは要求性能に応じて電力が最小となるパイプライン段数が複雑に変化している。差は小さいものの要求性能が 1.17×10^9 回/sec以下の場合には2段に分割した時が電力最小となっている。一方でそれ以上の性能が要求される場合は4段が電力最小となっている。 4.29×10^9 回/sec以降は4段では達成できない性能であり、8段のみが達成可能である。

このようにアプリケーションによっては電力を最小化する段数が変化するものとそうでないものが存在することが明らかとなった。

表3のPEの使用率を見るとgrayとsfはともに高いPE使用率で、使用している行数も最大である。しかし、sfとgrayではすでに述べたようにパイプライン段数と電力の関係が異なる。

5.1.1 ボディバイアス制御の違いによる効果

次にボディバイアス制御の効果について議論する。以下の3パターンで比較を行った。

- (1) ボディバイアス制御をしない(ゼロバイアス)
- (2) 全ての行に一律のボディバイアス電圧を与える
- (3) 行単位でボディバイアスを制御する

gray実行時において上記の3パターンでの消費電力を図8に示す。また、各アプリケーション実行時の行単位ボディバイアス制御による電力削減率を図9に示す。

図8を見ると行単位でボディバイアス制御を行うと、ボディバイアス制御をせずにゼロバイアスを与えた場合と比較すると電力を削減しているのに加えて、達成可能な性能を向上させていることがわかる。これはフォワードバイアスを与えることでさらに高い要求性能を満たすことができるからである。電力削減率は図9に示すようにgrayが最も高く25%を超える。ゼロバイアス時と比較すると電力削減効果は要求性能が低いほど高い。なぜならば、ゼロバイアスを与える場合は遅延時間に余裕がある低い要求性能時でもリーク電力を削減することができないが、ボディバイアス制御を行えばリーク電力を削減できるからである。また、gray実行時にゼロバイアスの場合に 3.12×10^9 回/secにおいて電力が急激に増加しているのは要求性能を満たすために8段のパイプライン分割に変更する必要がある、その分パイプラインレジスタでの電力が増加しているからである。

一律制御の場合と比較するとゼロバイアスと比較した場合のように達成可能な性能を向上する効果はない。一方で要求性能が達成可能な性能に近づくほど電力の差が大きくなっている。この理由は低い要求性能の場合、どのステージにも遅延時間に余裕があるため行単位の制御をしたとしても行毎でボディバイアス電圧の違いが少なく、どの行も近い値のボディバイアス電圧を与えていた。そのため一律制御と差がほとんどなかった。一方で要求性能が高くなると行単位の制御ではボトルネックとなる行にのみフォワードバイアスを与え、遅延時間に余裕のある行にはリバースバイアスを与えリーク電力の増加を最小限に抑えることができるが、一律制御の場合はPEアレイ全体にフォワードバイアスを与えなくてはならないため差が大きくなっている。一律制御に対する電力削減率は平均して10%前後であるがgrayにおいて特定の要求性能の時に最大28.87%の削減率を示している。

図9を見ると、ゼロバイアス時と比較した場合高い削減率を示したgray以外のアプリケーションでも15%を超える削減率が得られており、全アプリケーションで平均して19.64%の削減率が得られた。一律制御の場合と比較するとafでは10%を下回る削減率となっているが全アプリケーションで平均して10.71%の削減率を示している。これらの削減率の違いはアプリケーションによって行毎の遅延時間のばらつきが異なるからである。このように削減率は要求性能とアプリケーションに依存するため、本手法を用いて要求性能に応じてパイプライン段数とボディバイアス電圧を決定するのは有効である。

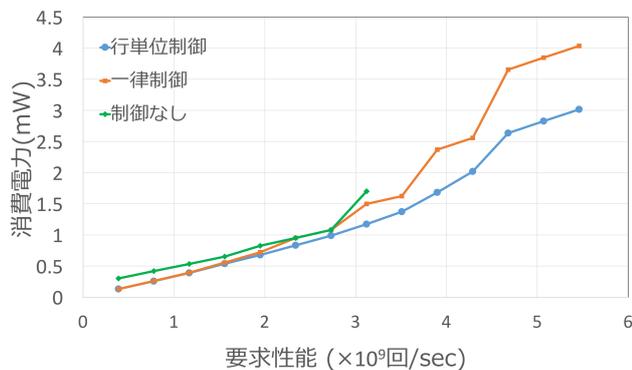


図8 grayにおける電力最小化結果の比較

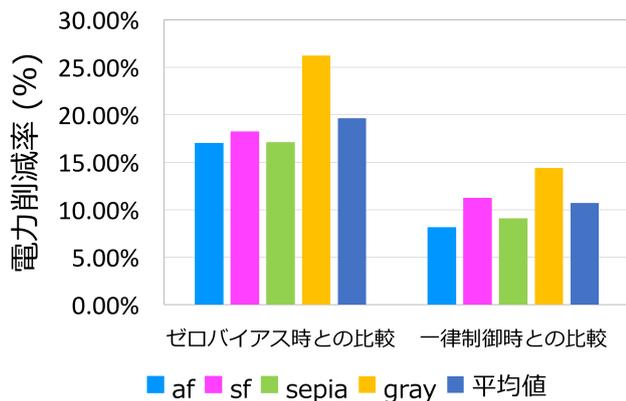


図9 行単位の制御による平均電力削減率

6. 結 論

VPCMAにおいて、実行するアプリケーションと要求性能に応じて電力を最小化するパイプライン段数とボディバイアス電圧を決定する手法を検討した。トレードオフの複雑さから単純に計算することが困難であったため探索はブルートフォースで行った。

電力を最小化するパイプライン段数とボディバイアス電圧には要求性能とアプリケーションに依存して複雑に変化するが本手法を用いることでそれを決定することができた。ボディバイアス制御の粒度を行単位とするとした場合ボディバイアス制御をしない場合と比べて平均19.64%の削減率が得られ、達成可能な性能が向上した。また、一律制御と比べると電力削減率は平均10.71%であった。しかし、要求性能が高くなるにつれて行単位の制御では一律制御に対して高い電力削減率を示し、最大で約28%を超える削減率が得られた。

今後の展望としては本研究ではパイプライン分割パターンに制限を設けていたがパイプライン分割パターンをさらに細粒度に変化させて探索を行う必要がある。それに伴い探索方法にも工夫が必要である。また、本研究では演算のマッピングは固定していたがパイプライン分割パターンによって最適なマッピングは異なると考えられ、マッピングも同時に検討することも必要である。

謝 辞

本研究は、JSPS 科研費 (S) ビルディングブロック型計算シス

テムに関する研究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。

文 献

- [1] 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).
- [2] H. Su, Y. Fujita and H. Amano: "Body bias control for a coarse grained reconfigurable accelerator implemented with silicon on thin box technology", 2014 24th International Conference on Field Programmable Logic and Applications (FPL), pp. 1–6 (2014).
- [3] 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 (ReConFig), pp. 1–6 (2015).
- [4] 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).
- [5] V. P. S. for Coarse Grained Reconfigurable Array CMA: "Naoki ando, koichiro masuyama, hayate okuhara, hideharu amano", 2016 INTERNATIONAL CONFERENCE ON FIELD-PROGRAMMABLE TECHNOLOGY, pp. 231–238 (2016).
- [6] Y. Morita, R. Tsuchiya, T. Ishigaki, N. Sugii, T. Iwamatsu, T. Ipposhi, H. Oda, Y. Inoue, K. Torii and S. Kimura: "Smallest vth variability achieved by intrinsic silicon on thin box (sotb) cmos with single metal gate", 2008 Symposium on VLSI Technology, pp. 166–167 (2008).
- [7] T. Ishigaki, N. Sugii, R. Tsuchiya, S. Kimura and Y. Morita: "Ultralow-power LSI Technology with Silicon on Thin Buried Oxide (SOTB) CMOSFET.", chapter 7, pp. 145–156, INTECH Open Access Publisher (2010).
- [8] H. Schmit, D. Whelihan, A. Tsai, M. Moe, B. Levine and R. R. Taylor: "Piperench: A virtualized programmable datapath in 0.18 micron technology", Custom Integrated Circuits Conference, 2002. Proceedings of the IEEE 2002IEEE, pp. 63–66 (2002).
- [9] J. M. Arnold: "S5: the architecture and development flow of a software configurable processor", Proceedings. 2005 IEEE International Conference on Field-Programmable Technology, 2005.IEEE, pp. 121–128 (2005).
- [10] M. Petrov, T. Murgan, F. May, M. Vorbach, P. Zipf and M. Glesner: "The xpp architecture and its co-simulation within the simulink environment", International Conference on Field Programmable Logic and ApplicationsSpringer, pp. 761–770 (2004).
- [11] X. Fan, H. Li, W. Cao and L. Wang: "Dt-cgra: Dual-track coarse-grained reconfigurable architecture for stream applications", Field Programmable Logic and Applications (FPL), 2016 26th International Conference onIEEE, pp. 1–9 (2016).
- [12] N. Weste and D. Harris: "CMOS VLSI Design: A Circuits and Systems Perspective", Addison-Wesley Publishing Company, 4th edition (2010).