

令和 6 年 5 月 28 日現在

機関番号：14401

研究種目：若手研究

研究期間：2020～2023

課題番号：20K19768

研究課題名（和文）ニアスレッシュヨルド回路の演算効率を最大化する近似コンピューティング基盤の創出

研究課題名（英文）Approximate Computing Platform Maximizing Computing Efficiency of Near-Threshold Circuits

研究代表者

塩見 準 (Shiomi, Jun)

大阪大学・大学院情報科学研究科・准教授

研究者番号：40809795

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：定格値の半分以下の電源電圧で集積回路を稼働させるニアスレッシュヨルドコンピューティング(NTC)と、多少の計算誤りを許容する近似コンピューティング(AC)を融合させた高効率コンピューティング基盤を開発した。NTCは省エネルギー動作を実現する有力手法だが、回路の部分的な動作故障を引き起こす欠点がある。この欠点をACで吸収するNTC回路設計手法を開発した。具体的には、脳型コンピュータ、低電圧メモリ、電圧最適化手法に分けて研究開発を行い、3件の国際論文誌掲載、3件の国際会議発表、2件の国内会議発表、3件の受賞があった。研究コンセプトのNTCとACの融合に成功し、研究の意義は大きい。

研究成果の学術的意義や社会的意義

あらゆるヒトやモノにIoT (Internet of Things)デバイスが組み込まれ、多種多様な情報を絶え間なく通信する次世代情報化社会が到来している。リアルタイム性、バッテリー駆動時間、CO2排出量の問題から、自動運転・ファクトリオートメーションなどの新サービスを支える集積回路の省エネルギー化・高性能化が喫緊の課題となっている。同時に、AC技術適用先の典型例である人工知能技術が急速に発展している。この社会トレンドをおさえ、NTCの欠点である動作技術をAC技術で克服し、NTC回路の利点である省エネルギー動作を享受できた意義は大きい。

研究成果の概要（英文）：This project developed a high-efficiency computing platform combining Near-Threshold Computing (NTC), where integrated circuits operate at less than half of the nominal supply voltage, and Approximate Computing (AC), which can tolerate small computational errors. However, NTC has a drawback that it can cause partial operation failures in the target circuit. This project developed NTC circuit design methods that absorb this drawback with AC. Specifically, this project was divided into brain-inspired computer design, low-voltage on-chip memory design, and dynamic voltage optimization methods. This project results in three publications in international journals, three presentations at international conferences, two presentations at domestic conferences, and three awards. The research concept of combining NTC and AC has been successfully demonstrated. Its impact is thus very significant.

研究分野：集積回路設計

キーワード：計算機システム 省エネルギー 低消費電力化 ニアスレッシュヨルドコンピューティング 近似コンピューティング

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

あらゆるヒトやモノに IoT (Internet of Things) デバイスが組み込まれ、多種多様な情報を絶え間なく通信する次世代情報化社会が到来している。リアルタイム性、バッテリー駆動時間、CO₂ 排出量の問題から、自動運転・ファクトリオートメーションなどの新サービスを支える集積回路の省エネルギー化・高性能化が喫緊の課題となっている。集積回路の低電圧動作が省エネルギー化の最有力手法である。例えば Intel 社は 2010 年代の 5 大重要技術として、トランジスタのしきい値電圧付近電圧で回路を動作させるニアスレッシュホールドコンピューティング (Near-Threshold Computing: NTC) 技術を挙げている。しかし、電源電圧を下げるとトランジスタの特性ばらつきに起因する回路特性ばらつきが大きくなり、最終的には回路の部分的な動作故障をもたらす。

本研究では、NTC の欠点をカバーする技術として、近似コンピューティング (Approximate Computing: AC) と融合した高効率 NTC 技術の解明を行う。AC について、例えば自動運転のキーテクノロジーである画像処理技術では、人間が知覚できない程度の誤差であれば、処理の過程に計算誤差が混入しても出力結果はほとんど変化しない。この特徴を逆手に取り、必要最低限の計算精度を保証しつつ、演算回路を簡略化することで、集積回路の消費エネルギーだけでなく、動作速度も大幅に改善させる AC が人工知能技術を中心としてホットトピックになっている。ニューラルネットワークのように、演算精度の低い“あいまいな”演算器を積極的に活用したアプリケーションが自動運転・ファクトリオートメーションなどの現代のキラーアプリケーションを担っている。本研究課題では NTC の動作故障を AC でカバーする技術開発を行う。

NTC と AC の融合に際し、以下の学術的な問いがある。

問い 1: “正常動作” とは何か

NTC-AC 回路が最終的に出力する演算誤差が許容範囲内であることをモニタすることは重要である。集積回路が正常に動作していることを確認するため、通常の場合はチップ出荷前やフィールドでのテストが行われる。一般的には、テスト入力に対する出力と期待値の一致度でテストを行う。しかし、AC では計算誤差を許容するため出力の期待値は一意に定まらず、テスト(モニタ)手法そのものがオープンプロブレムである。

問い 2: 計算誤差のランタイム補償

AC が許容する精度は、処理アプリケーションに強く依存する。したがってアプリケーションに応じたランタイムなフィールドテストが必要である。NTC 回路の出力が許容する演算誤差を満たしていることを、フィールドにて効率よくテスト(モニタ)し、ランタイムにフィードバックする技術は未開拓である。

2. 研究の目的

次世代社会の持続的発展を目的とし、目的達成のために NTC 技術と AC 技術を融合した高効率集積回路設計技術を明らかにする。具体的には、画像処理プロセッサやニューラルネットワークなど、一定の計算誤差を許容するアプリケーションを NTC 回路で処理させ、NTC 回路の弱点である部分的な動作故障を克服しつつ、NTC 回路の利点である省エネルギー動作を享受する回路技術を開発する。また、NTC 回路の稼働状況をモニタし、ランタイムにフィードバックすることで性能を最大化する NTC 回路設計技術を開発する。

3. 研究の方法

以下のアプローチにわけて研究を実施した。

(a) 動作故障を考慮したディペンダブル回路設計手法、および動作故障をフィールドで効率良くセルフテスト・モニタする技術の開発

集積回路が低電圧領域で動作すると、トランジスタの特性ばらつきに起因する回路特性ばらつきが大きくなり、最終的には回路の部分的な動作故障をもたらす。部分的な動作故障に負けないディペンダブルな NTC 回路設計手法を開発する。部分的な動作故障が起きても、AC 技術がその影響を吸収する技術を開発する。具体的には、部分的な動作故障に対してロバストな人工知能回路の数値表現や学習方法、フィールドテスト・モニタ手法を開発する。

(b) テスト・モニタ結果に基づき、故障回路を自律補償する技術の開発

前述の(1)の技術で開発したテスト・モニタ結果に基づき、ベストな状態で回路を動作させる自律補償技術を開発する。

4. 研究成果

以下 C1, C2, C3 に成果を示す。NTC 回路と AC 回路という本研究の骨子を証明できた。例えば、C1 の成果では、脳型コンピュータを例に実データセットに対して、フィールドテスト技術等を用いて推論精度を落とすことなく NTC 動作を実現し、省エネルギー化を実現した。実応用への展開も期待でき、その社会的インパクトは大きい。

C1. NTC 回路向けディペンダブル脳型コンピュータの開発

集積回路において、低電圧動作に最も脆弱な素子の 1 つは SRAM (Static Random Access Memory) である。具体的には、SRAM のビットセルの一部が正しく情報を記憶できなくなる問題がある。エッジ向け脳型コンピューティング手法である HyperDimensional Computing (HDC) システムに対し予め適切な学習を施すことで、オンチップメモリの低電圧動作時の動作故障に起因する計算誤りを吸収できることを示した (提案システムを DependableHD と名付けた)。HDC は特徴量を高次元空間にマッピングする手法である。図 1 に HDC の全体像を示す。②~④で特徴量を高次元空間にマッピングし、クラス高次元ベクトルとの類似度計算を行うことで推論を行う。学習時には、推論結果が間違った場合、⑥と⑦に示す再学習を行う。HDC は本質的にノイズに強い特徴があるが、NTC 動作で生じる SRAM ビットセルの故障率は大きすぎて対応しきれない課題があった。ビットセルエラーの故障を想定した学習手法である Margin enhancement (図 1 の⑥と⑧)。学習時に推論の正誤だけでなく、「間違えそう」な推論結果も再学習の対象とする) と Random noise injection (図 1 の⑤)。ノイズを意図的に学習時に混入する) を提案し、深刻な動作故障が発生しても推論精度を維持することに成功した。また、オンチップメモリをランタイムにフィールドテストし、0/1 スタック故障したビットセルにそれぞれ 0/1 の高次元ベクトル要素を割り当てて、故障による数値変化の影響を排除する Dimension swapping 技術を提案した (図 2)。65-nm プロセスで設計した SRAM に対して、シミュレーションから幅広い電源電圧に対してビットセル故障率を求めた。この故障率評価結果をもとに HDC 推論精度の電源電圧依存性を図 3 の通り求めた。従来ベースの HDC 実装 (“Base”) と比較して、提案手法 (“M+R+S”) は推論精度を落とすことなく、電源電圧を 430 mV から 340 mV まで削減でき、従来比で推論精度をほぼ落とさず消費エネルギーを約 40%削減できた。本成果を通して、1 件の国際論文誌掲載、1 件の国際会議発表、1 件の国内会議発表、1 件の受賞があった。研究コンセプトの NTC と AC の融合に成功し、研究の意義は大きい。

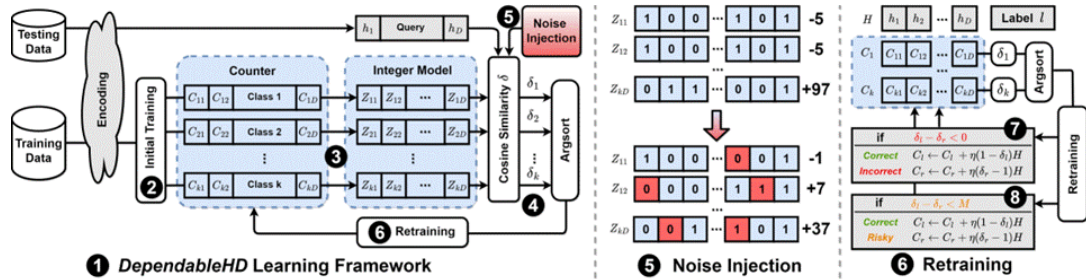


図 1: DependableHD の概要。

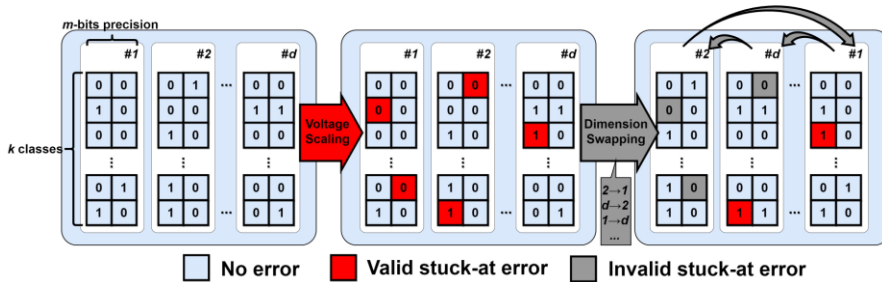


図 2: Dimension swapping の概要。

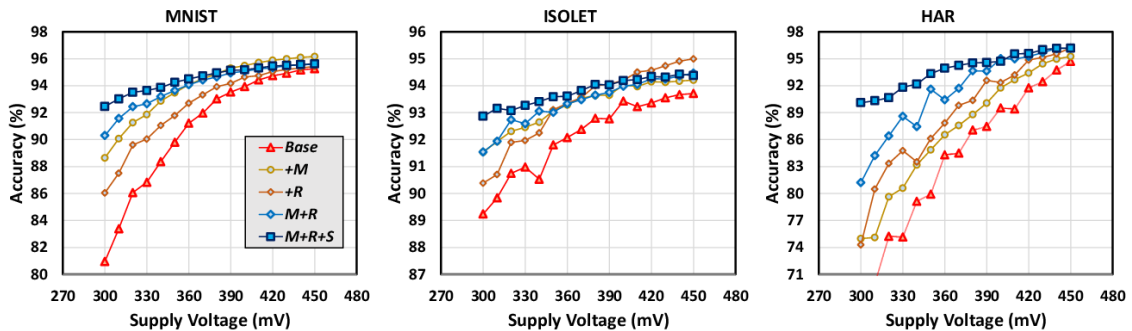


図 3: データセット (MNIST, ISOLET, HAR) の推論結果。

M: Margin enhancement. R: Random noise injection. S: Dimension swapping.

C2. 細粒度パワーゲーティングが可能な NTC 回路向けオンチップメモリの開発

成果 C1 と同じく、オンチップメモリに注目して開発を行った。SCM (Standard-Cell Memory) と呼ばれるセルベースの低電圧メモリ (面積効率が SRAM より悪いが低電圧で安定動作する。小容量メモリ向け) を対象に、メモリに記憶する精度を動的に変更可能で、精度を切り落としたビット部分にパワーゲーティングを施す独自の SCM を開発した。例えば 32 ビット精度ではなく 16 ビット精度のみ必要なアプリケーションを稼働させる場合は、半数のビットセルにパワーゲーティングを施せる。図 4 にコンセプトを示す。スタンダードセルの電源レールを取り除くことで、隣り合うセルが別の電源ドメインで稼働することができ、細粒度にパワーゲーティングを施せることを提案した (図 4 右側のレイアウトに書かれた Bank1 から Bank8 は異なる電力ドメインで稼働する)。この結果、メモリにストアされたデータの表現精度を削った場合、削られたビット部分にパワーゲーティングを施すことができリーク電力を削減できる。商用 55-nm プロセスを用いた実験では、ベースライン回路と比較して最大 59% のリーク電力を削減できた (図 5 左)。また、ゼロだけを記憶しているメモリバンクを検出し、動的にパワーゲーティングを施すスタンダードセル構造も提案した。55-nm プロセスを用いた実験結果では、ほぼゼロを記憶するメモリのリーク電力をベースライン回路と比較して 33% 削減できることを明らかにした (図 5 右)。一般的な AC 技術では、数値の精度を大幅に削っても (ゼロ埋めしても) 正常に動作する。また、ほぼゼロのデータしか記憶しないアプリケーションも存在する。NTC 回路ではリーク電流に起因する電力消費が一大問題となっており、ゼロを記憶するビットセルにパワーゲーティングを積極的に適用することで、NTC 回路の省エネルギー動作に貢献する。本研究成果をまとめて国際会議で 1 件発表した。

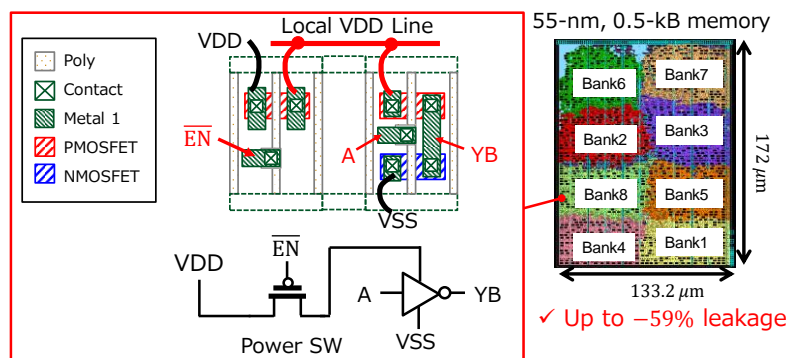


図 4: 細粒度パワーゲーティング SCM の概要。

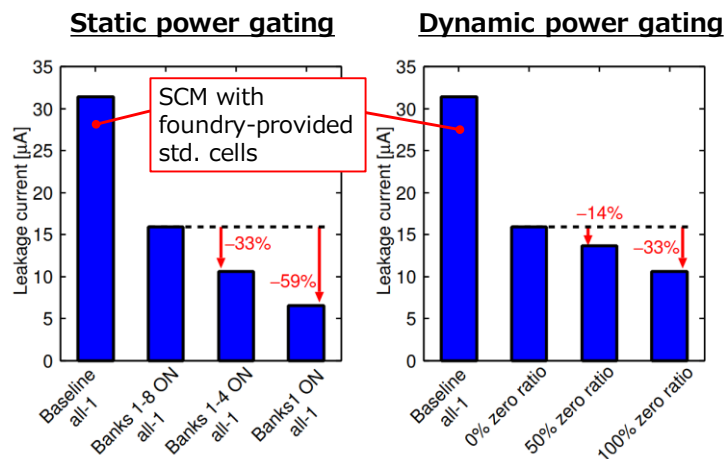


図 5: リーク電力の削減率。(左) 電源レールレススタンダードセルによる細粒度パワーゲーティング。(右) ゼロ検出機構による動的なパワーゲーティング。

C3. 動的な電圧フィードバック技術の開発

集積回路に与えられた要求動作速度を満たしながら、最小のエネルギーを実現する電圧組 (電源電圧 V_{dd} とバックゲート電圧 V_{bb} の組。Minimum Energy Point: MEP。) をランタイムに設定する (MEP Tracking: MEPT) ことは、集積回路の省エネルギー動作に重要である。図 6 左に MEPT システムの実装例を示す。制御対象回路の動的/静的消費エネルギー (E_d および E_s)、回路遅延 (D)、およびチップ温度 (T) という稼働状況をモニタすることで、“MEPT Controller” にてランタイムに MEP を決定、“DC/DC コンバータ” でフィードバックすることができるが、MEP 計算のコストが大きい課題があった。最適値から電圧組が少しずれても、消費エネルギーは最適値からあまり増加しない (ロスが小さい) 事実に着目し、最適な電源電圧とバックゲート電圧の導出に必要な計算に大胆な近似を適用した。

具体的には、計算コストの大きい指数関数と対数関数を、ハードウェア実装効率の良い Mitchell 近似対数関数で置き換えた。このコンセプトのもと、MEPT システムのプロトタイプを 65-nm プロセステクノロジーで試作し、実測でその効果を検証した (図 6 右)。その結果、その計算の負荷をソフトウェア実装時に約 75 分の 1 の遅延に、ハードウェア実装時に約 4 分の 1 の面積に削減する一方、消費エネルギーのロス を 1%未満に抑える手法を提案した。図 7 に MEPT の例を示す。横軸が組込プロセッサに印加したバックゲート電圧 V_{bb} で、縦軸が電源電圧 V_{dd} である。ダイヤモンド印の点が正確に MEP を推定した結果、丸い点が近似 MEP である。前述の通り消費エネルギーのロス を 1%未満に抑えられることを実プロセッサで確認した。電圧最適化技術に関し、1 件の国際会議発表、1 件の国内会議発表、2 件の論文誌発表を行った。2 件の受賞があった。

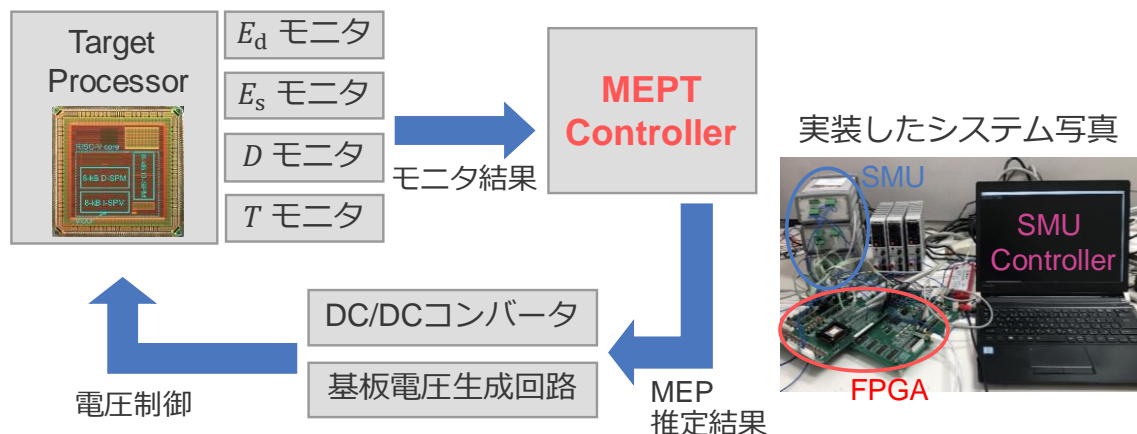


図 6: (左) MEPT システムの例。(右) 提案手法を実装した近似 MEPT システム。

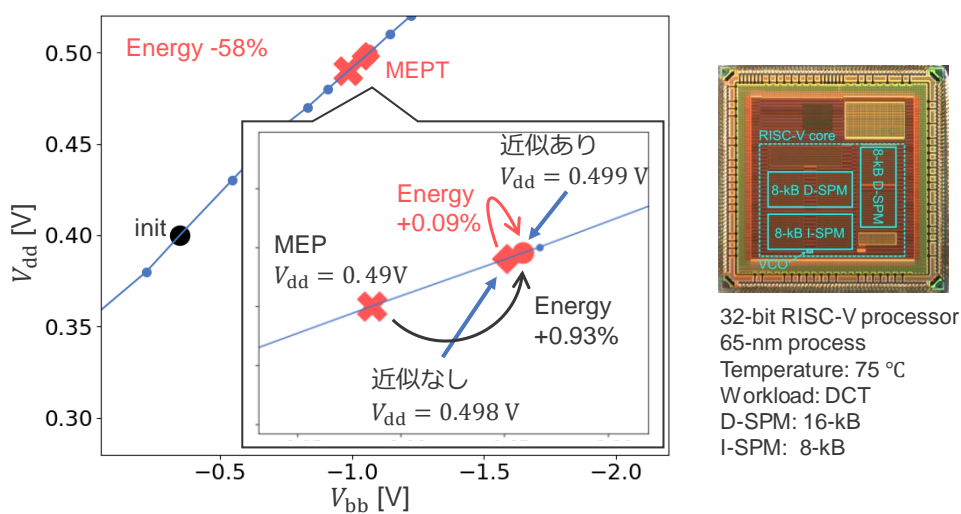


図 7: 近似 MEPT システムの MEP 推定結果。

5. 主な発表論文等

〔雑誌論文〕 計3件（うち査読付論文 3件/うち国際共著 0件/うちオープンアクセス 1件）

1. 著者名 SONODA Shoya, SHIOMI Jun, ONODERA Hidetoshi	4. 巻 E106.A
2. 論文標題 Approximation-Based System Implementation for Real-Time Minimum Energy Point Tracking over a Wide Operating Performance Region	5. 発行年 2023年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 542 ~ 550
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.2022VLP0006	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 SONODA Shoya, SHIOMI Jun, ONODERA Hidetoshi	4. 巻 E104.A
2. 論文標題 Supply and Threshold Voltage Scaling for Minimum Energy Operation over a Wide Operating Performance Region	5. 発行年 2021年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	6. 最初と最後の頁 1566 ~ 1576
掲載論文のDOI（デジタルオブジェクト識別子） 10.1587/transfun.2020kep0013	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Liang Dehua, Awano Hiromitsu, Miura Noriyuki, Shiomi Jun	4. 巻 -
2. 論文標題 A Robust and Energy Efficient Hyperdimensional Computing System for Voltage-scaled Circuits	5. 発行年 2023年
3. 雑誌名 ACM Transactions on Embedded Computing Systems	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3620671	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計5件（うち招待講演 1件/うち国際学会 3件）

1. 発表者名 Dehua Liang, Hiromitsu Awano, Noriyuki Miura, Jun Shiomi
2. 発表標題 DependableHD: A Hyperdimensional Learning Framework for Edge-Oriented Voltage-Scaled Circuits
3. 学会等名 28th Asia and South Pacific Design Automation Conference (ASP-DAC)（国際学会）
4. 発表年 2023年

1. 発表者名 Jun Shiomi, Shogo Terada, Tohru Ishihara, Hidetoshi Onodera
2. 発表標題 Zero-Aware Fine-Grained Power Gating for Standard-Cell Memories in Voltage-Scaled Circuits
3. 学会等名 IEEE 35th International System-on-Chip Conference (SOCC) (国際学会)
4. 発表年 2022年

1. 発表者名 Dehua Liang, Hiromitsu Awano, Noriyuki Miura, Jun Shiomi
2. 発表標題 [記念講演] DependableHD: A Hyperdimensional Learning Framework for Edge-oriented Voltage-scaled Circuits
3. 学会等名 信学技報 (招待講演)
4. 発表年 2023年

1. 発表者名 Shoya Sonoda, Jun Shiomi, Hidetoshi Onodera
2. 発表標題 Approximation-Based Implementation for a Minimum Energy Point Tracking Algorithm over a Wide Operating Performance Region
3. 学会等名 13th Latin American Symposium on Circuits and Systems (LASCAS) (国際学会)
4. 発表年 2022年

1. 発表者名 園田翔也, 塩見準, 小野寺秀俊
2. 発表標題 幅広い動作性能領域で最小エネルギー動作点を追跡するアルゴリズムの近似演算に基づく高効率実装
3. 学会等名 DAシンポジウム2021
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------