様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月　 ７ 日現在

機関番号：８２６２６
研究種目：基盤研究(C)（一般）
研究期間：2020～2022
課題番号：２０Ｋ１１７４０

研究課題名（和文）アナログ常微分方程式ソルバーを用いた超低消費電力深層学習専用チップの開発

研究課題名（英文）Development of Ultra-low Power Deep Learning Accelerator Chip using Analog Circuit-based Ordinary Differential Equation Solver

研究代表者

更田　裕司（Fuketa, Hiroshi）

国立研究開発法人産業技術総合研究所・エレクトロニクス・製造領域・主任研究員

研究者番号：３０５８７４２３

交付決定額（研究期間全体）：（直接経費）　　3,400,000円

研究成果の概要（和文）：① キーワードスポッティング（事前に定義したキーワードを入力音声から検出）を対象として、常微分方程式に基づくアルゴリズム（Neural ODE）を適用する手法を新たに開発し、ネットワークのパラメータ数を68%削減可能である事を示した。
② キーワードスポッティングをハードウェアで実行する為、従来は入力されたアナログ音声信号をデジタルに変換し、特徴量を抽出するデジタル信号処理を行っていた。しかしこれには、デジタル変換の電力が大きいという課題があった。そこで本研究では、アナログ信号のまま演算を実行しデジタル変換を除去する回路技術を開発し、特徴量抽出に必要な電力を88%削減できることを示した。

研究成果の学術的意義や社会的意義
本研究は、バッテリー駆動のような電力制約が厳しい端末での常時深層学習処理を実現する為、低消費電力の（エネルギー効率の高い）集積回路技術の開発を目的としている。これまで、本研究で使用される常微分方程式に基づく深層学習アルゴリズム（Neural ODE）とエネルギー効率の関係について検討はなされておらず、本研究で示したNeural ODEがエネルギー効率向上に有効であることは学術的に意義深いと考える。さらに、エネルギー効率を高める為にアナログ回路を活用した演算回路技術を提案し、その有効性について実際にCMOSチップを製造し検証した点は、実用化に向けた重要な一歩であり、社会的な意義も高いと考える。

研究成果の概要（英文）：1) Techniques to utilize the Neural ODE for keyword spotting tasks were proposed, which can reduce the number of parameters in the network by 68%.
2) In the conventional circuits to perform keyword spotting tasks, analog speech signals were digitized, and then digital signal processors extracted features from the digital signals. However, there was a problem that the power consumption of the analog-to-digital conversion is significantly large. Thus, circuit techniques to extract features in an analog signal domain and remove the digital conversion were developed, which can reduce the power dissipation by 88%.

研究分野：集積回路設計

キーワード：ニューラルネットワーク　機械学習　深層学習　音声認識　キーワードスポッティング　常微分方程式　低消費電力　集積回路

１．研究開始当初の背景

深層学習とは、人間の脳神経回路をモデルとしたニューラルネットワークを多層構造にしたもので、近年の人工知能（AI）の進展の中核を成す技術である。深層学習の活用により、画像認識や音声認識などの分野で既存技術と比べて大幅な性能改善を実現できる一方、膨大な計算量が必要となる。この膨大な計算を効率的に実行する為に、深層学習の処理に特化した専用チップ（本報告書では「AI チップ」と呼ぶ）の開発が盛んに行われていた。

また、深層学習の処理はデータセンターなどクラウド側で行われる事が多いが、クラウド側にデータを送る為の電力・遅延時間・セキュリティなどの課題があり、エッジ（端末）側で AI 処理を行う試みが広がりつつあった。しかし、バッテリー駆動のような電力制約の厳しいエッジデバイス上で常時 AI 処理（例：監視カメラで顔認識）を実行しようとすると、100TOPS/W 以上のエネルギー効率が必要とされていた。一方、既存の AI チップは 1-10TOPS/W 程度であり、深層学習の計算のエネルギー効率を一桁以上改善する技術が求められていた。

２．研究の目的

実用化されている AI チップの大半はデジタル回路で実現されている。そこで、エネルギー効率を向上させる為、深層学習処理の大半を占める積和算をアナログ回路で実現する研究が盛んに行われている。この方式の課題は、アナログ演算の前後で、デジタルとアナログの変換が必要で、その電力オーバヘッドが発生する点である。

本研究の目的は、さらなるエネルギー効率向上を目指して、従来のニューラルネットワークと計算原理が異なる、常微分方程式（ODE; Ordinary Differential Equation）に基づく計算アルゴリズム（Neural ODE[1]）の導入とそれをアナログ回路で実装することで、既存 AI チップと比べてエネルギー効率を 10 倍改善する事である。特に、デジタルとアナログの変換を軽量化（除去）する技術を確立が本目的の達成に重要となる。

３．研究の方法

(1) アプリケーションの決定とアルゴリズムの検討

まず、提案のアナログ回路ベースの Neural ODE が有効となるアプリケーションの検討を行い、対象とするアプリケーションを確定する。次に、そのアプリケーションに応じた、ニューラルネットワーク構造などアルゴリズムを検討する。

(2) 低消費電力を実現する、デジタル・アナログ変換の軽量化技術の開発

デジタルとアナログの変換を軽量化（除去）する回路技術を確立し、(1)で開発したアルゴリズムを低電力で実現する回路設計を行う。さらに、チップ試作を行い、実チップでその回路技術の有効性を検証する。

４．研究成果

(1) アプリケーションの決定とアルゴリズムの検討

提案のアナログ回路ベースの Neural ODE が有効となるアプリケーションの検討を行い、まず提案技術が有効となるアプリケーションの検討を行い、その結果、音声認識を対象とすることに決定した。この理由は次の 2 点である。1) 提案技術は、アナログ回路で計算を実行する技術であり、センサと直結できるようなアプリケーションと親和性が高く、マイクというセンサを使用する音声認識は最適である。2) 音声認識は、モバイル機器への搭載が進んでおり、低電力化への要望が強い。

一方、Neural ODE は、これまで音声認識の分野に適用されていなかった為、ハードウェア化の検討の前にまず、Neural ODE を用いて音声認識を行う為のニューラルネットワーク構造などアルゴリズム検討を行った。対象とする音声認識のタスクとしては、キーワードスポッティング（事前に定義したキーワードを音声中から検出）とした。キーワードスポッティングはスマートフォンやスマートスピーカーなどエッジデバイスで実行される事が通常で、低消費電力動作が求められる。従って、ニューラルネットワークのパラメータ数や計算量の削減が重要である。

本研究では、Neural ODE に適用するアルゴリズムとして、Temporal Convolutional Neural Network（TCNN）[2]と Time Delay Neural Network（TDNN）[3]の 2 種類を採用することにした。これらのアルゴリズムの詳細を図 1 に示す。音声特徴量として、メル周波数ケプストラム係数（MFCC）を入力とする（チャネル数は 40 とする）。これを初期値として常微分方程式を解き、その結果を全結合ニューラルネットワークに入力することで、入力音声を分類する。この「常微分方程式を解く」という処理は、「多層の Residual Neural Network（ResNet）[4]」の処理と等価とされている一方、パラメータを使いまわせる事から、提案技術を用いることで、大幅なパラメータ数削減が期待できる。

表 1 は、提案のネットワークモデルの各層の設定と、パラメータ数・計算量（乗算数）をまとめたものである。本研究の対象アプリケーションである、キーワードスポッティングの性能を評

価する為、本研究では、Google Speech Commands Dataset をデータセットとして用いて、既存研究[2,4]と同様に、入力音声から 12 種の分類（10 種のキーワード、それ以外のキーワード、非音声の計 12 種の分類）を行うタスクとする。

既存研究と提案の Neural ODE を用いたニューラルネットワークの推論精度とパラメータ数、及び、計算量の関係を図 2(a)と(b)それぞれに示す。提案の TDNN を用いたモデル（ode-tdnn29）は、従来技術（res8-narrow）とほぼ同じ精度を維持しつつ、パラメータ数を 68%削減できる事が分かる。一方で、提案の Neural ODE を用いたモデルの計算量については、TCNN を用いた従来モデル[2]と同程度にとどまる。これは、提案モデルでは常微分方程式を解く為に、計算を繰り返す必要がある（この繰り返す回数は表 1 の NFE として表される）からである。

以上から、提案技術を用いることで、従来と比べて、精度を維持しつつパラメータ数を削減でき、その結果低電力化が期待できる。一方、計算量については従来と同程度であり、この計算量の削減が今後の課題である。

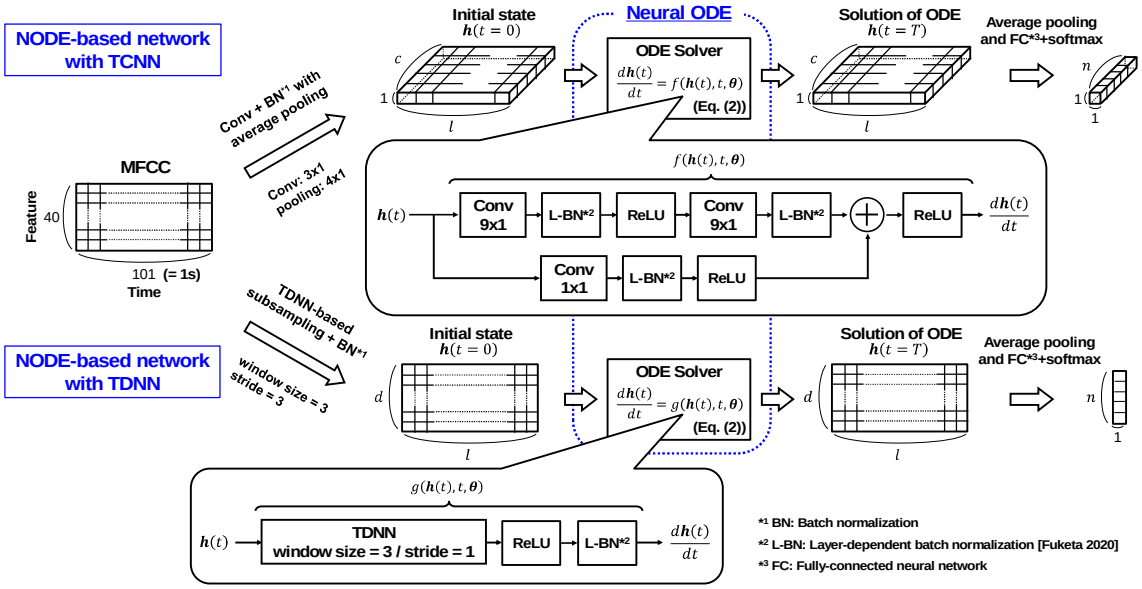

図 1．キーワードスポッティング向け Neural ODE を用いたネットワークモデル（提案手法）

表 1．提案モデルの各層の設定

(a) ode-tcnn20

Layer	m	r	c	l	# param	# mult.	
Input			40	101			
Conv	3	1	20	101	2.4k	242k	
Avg. pool	4	1	20	25			
Conv	9	1	20	25	3.6k	90k	ODE
Conv	9	1	20	25	3.6k	90k	
Conv	1	1	20	25	0.4k	10k	
Avg. pool			20	1			
FC			12	1	0.24k	240	
Total					10k	242k + 190k x *NFE**	

(b) ode-tdnn32

Layer	w	s	d	l	# param	# mult.	
Input			40	101			
TDNN-SUB	3	3	32	34	3.9k	131k	
TDNN	3	1	32	34	3.1k	104k	ODE
Avg. pool			32	1			
FC			12	1	0.4k	384	
Total					7.4k	131k + 104k x *NFE**	

* NFE: The number of function evaluations [Chen 2018]

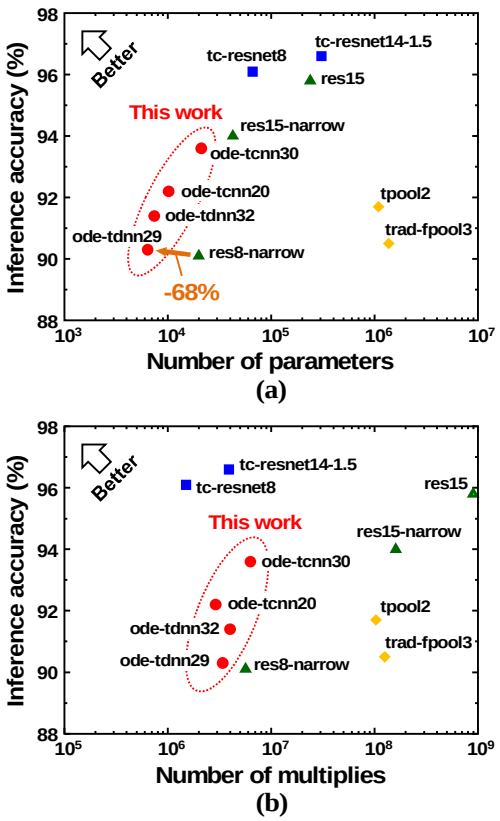

図 2．推論精度と(a) パラメータ数、(b) 計算量の関係

(2) 低消費電力を実現する、デジタル・アナログ変換の軽量化技術の開発

キーワードスポッティングでは、図 1 でも示した通り、音声信号を MFCC 特徴量に変換して、それをニューラルネットワークで構成された識別器によりキーワードの検出を行う。近年、キーワードスポッティングを超低電力で実行する AI チップが提案されているが、本 AI チップでは、マイクから入力された音声信号を ADC（アナログ・デジタル変換）でデジタル信号に変換し、MFCC 抽出と識別器をデジタル回路で実装している[5]。一方で、この方式には、ADC と MFCC 抽出に必要な消費電力が大きいという課題があった。そこで本研究では、次の 3 点を特徴とする回路の提案を行った。

- MFCC を複数のバンドパスフィルター（BPF）に置き換え
- 深層学習アルゴリズムを用いた特徴量抽出手法を提案
- これらをアナログ回路で実現することで ADC を除去

以上で述べた提案手法の概略を図 3 に示す。なお、提案技術の有効性検証の為、ここでは、ニューラルネットワーク部分は、Neural ODE を使用せず、一般的なニューラルネットワークを使用することを前提としているが、Neural ODE にも適用可能であると考えている。

重要な点は、図 3(b)に示す通り、深層学習アルゴリズムである TDNN[3]を用いた特徴量抽出手法、及び、これをアナログ回路で実現する手法の提案である。TDNN は、アルゴリズムが単純で、アナログ回路でもハードウェア化が容易である為、本研究に採用することにした。また、音声認識に従来使用されていた特徴量である MFCC はフーリエ変換など高度な信号処理が必要で通常デジタル回路で実装されてきたが、本研究では MFCC の代わりに、BPF を用いて周波数帯域毎の信号強度を求めることにした。BPF は MFCC に比べて、非常に簡単なアナログ回路で実現可能である。これにより、従来アナログの入力音声信号を ADC でデジタル化し特徴量を求めていたものを、全てアナログ回路に置き換え ADC を除去する事が可能となる。

図 4 に具体的な回路図を示す。BPF は Super-source-follower[6]をベースとしたもので、キャパシタの容量 C_1 と C_2 の大きさを変えることで、様々な周波数帯域の BPF を実現できる。BPF の出力をアンプで増幅し、整流器をベースとした回路（MAX 回路）を用いてその最大値を取得する。最後に、重みがバイナリ（つまり、重みが 0 か 1）の TDNN により特徴量を計算する。TDNN の計算は積和算（MAC）で構成され、ここでは、スイッチトキャパシタを用いたアナログ回路ベースの積和算回路を使用する。

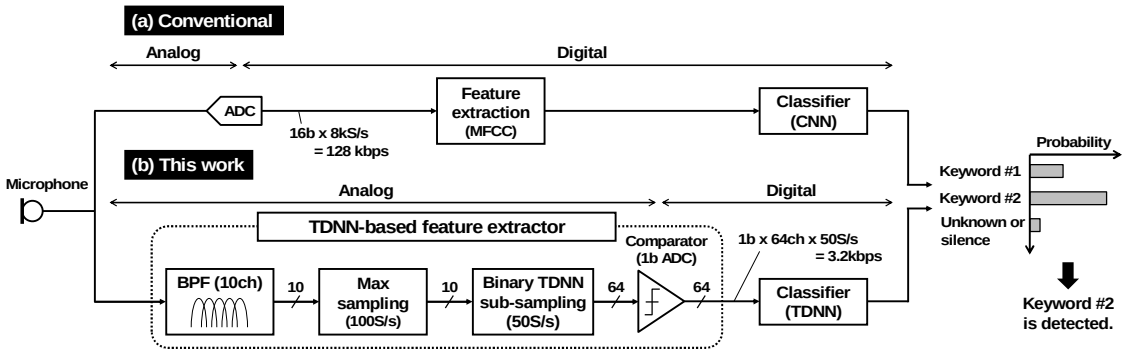

図 3．キーワードスポッティング向けアルゴリズム、(a) 従来手法 (b) 提案手法

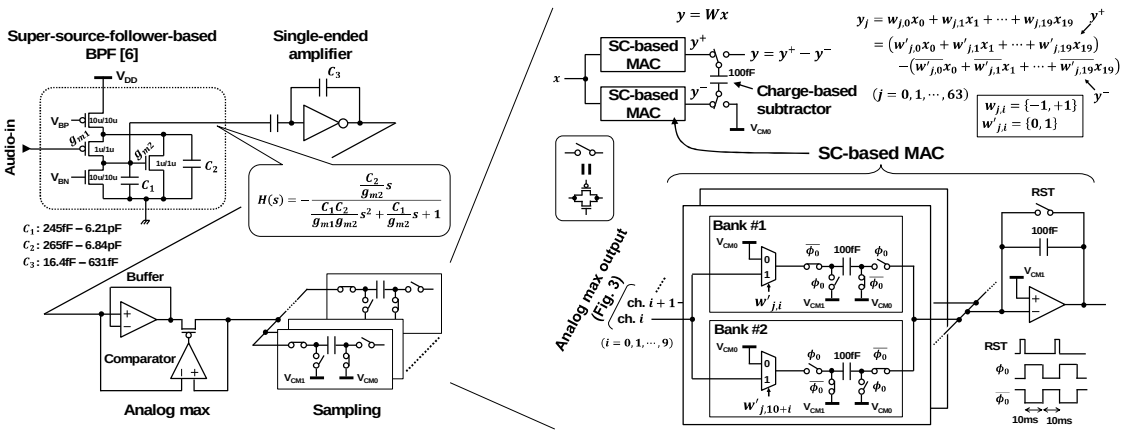

図 4．アナログ回路で実現された特徴量抽出の回路図（提案手法）

図 5 に提案手法による電力削減効果のシミュレーション結果を示す。従来の特徴量抽出では、ADC の電力が大半を占めていたが、提案手法ではこれが不要となる。加えて、従来の MFCC の代わりに、提案の BPF と TDNN ベースの特徴量を用いる事で、特徴量抽出に必要な電力も削減可能である。その結果、消費電力を 88%削減（エネルギー効率 10 倍改善）できることが分かった。さらに、本研究では、65nm プロセスでチップの試作を行い、実チップで提案手法の有効性を検証した。試作したチップの写真を図 6 に示す。

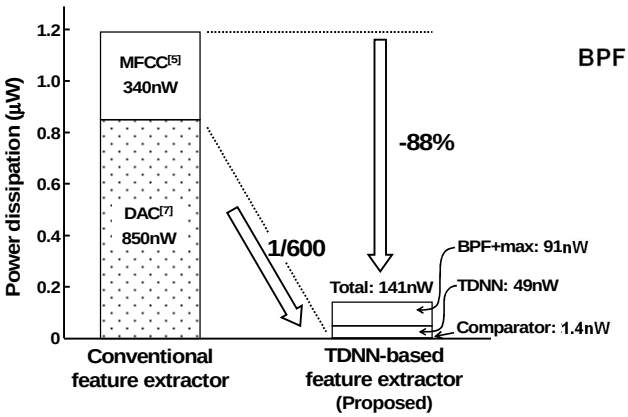

図 5．提案手法による特徴量抽出の電力削減

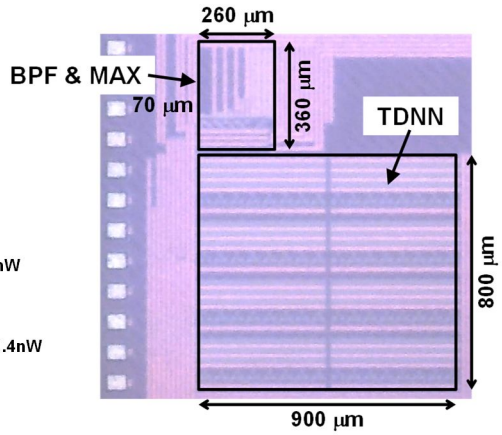

図 6．試作チップ写真

参考文献

[1] R.T.Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud, “Neural Ordinary Differential Equations,” Proc. NIPS, 2018.

[2] S. Choi, S. Seo, B. Shin, H. Byun, M. Kersner, B. Kim, D. Kim, and S. Hay, “Temporal Convolution for Real-time Keyword Spotting on Mobile Devices,” Proc. INTERSPEECH, 2019.

[3] V. Peddinti, D. Povey, and S. Khudanpur, “A Time Delay Neural Network Architecture for Efficient Modeling of Long Temporal Contexts,” Proc. INTERSPEECH, 2015.

[4] R. Tang and J. Lin, “Deep Residual Learning for Small-Footprint Keyword Spotting,” Proc. ICASSP, 2018.

[5] W. Shan, et.al., “A 510nW 0.41V Low-Memory Low-Computation Keyword-Spotting Chip Using Serial FFT-Based MFCC and Binarized Depthwise Separable Convolutional Neural Network in 28nm CMOS,” ISSCC, 2020, pp. 230-231.

[6] M. Yang, et al., “A 1μW Voice Activity Detector Using Analog Feature Extraction and Digital Deep Neural Network,” ISSCC, 2018, pp. 346-347.

[7] D. Venuto, et al., “0.8 μW 12-bit SAR ADC Sensors Interface for RFID Applications,” Microelectronics J., vol. 41, no. 11, pp. 746-751, Nov. 2010.

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　2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 Fuketa Hiroshi	４．巻 5
２．論文標題 Ultralow Power Feature Extractor Using Switched-Capacitor-Based Bandpass Filter, Max Operator, and Neural Network Processor for Keyword Spotting	５．発行年 2022年
３．雑誌名 IEEE Solid-State Circuits Letters	６．最初と最後の頁 82～85
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/LSSC.2022.3164573	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 Fuketa Hiroshi	４．巻 69
２．論文標題 Time-Delay-Neural-Network-Based Audio Feature Extractor for Ultra-Low Power Keyword Spotting	５．発行年 2022年
３．雑誌名 IEEE Transactions on Circuits and Systems II: Express Briefs	６．最初と最後の頁 334～338
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TCSII.2021.3098813	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 更田 裕司、森田 行則	４．巻 JSAI2021
２．論文標題 キーワードスポッティング向けNeural ODEを用いたパラメータ削減手法の提案	５．発行年 2021年
３．雑誌名 人工知能学会全国大会論文集	６．最初と最後の頁 4I2GS7c03
掲載論文のDOI（デジタルオブジェクト識別子） 10.11517/pjsai.JSAI2021.0_4I2GS7c03	査読の有無 無
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 Fuketa Hiroshi、Uchiyama Kunio	４．巻 54
２．論文標題 Edge Artificial Intelligence Chips for the Cyberphysical Systems Era	５．発行年 2021年
３．雑誌名 Computer	６．最初と最後の頁 84～88
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/MC.2020.3034951	査読の有無 無
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 Fuketa Hiroshi、Morita Yukinori	４．巻 1
２．論文標題 Neural ODE with Temporal Convolution and Time Delay Neural Networks for Small-Footprint Keyword Spotting	５．発行年 2020年
３．雑誌名 arXiv	６．最初と最後の頁 1〜5
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
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 －

〔学会発表〕　計12件（うち招待講演　3件／うち国際学会　0件）

１．発表者名 更田 裕司
２．発表標題 ISSCC2023におけるAIチップ研究動向
３．学会等名 第45回AIチップ設計拠点フォーラム
４．発表年 2023年

１．発表者名 更田 裕司
２．発表標題 Symposium on VLSI Technology and Circuits 2022 におけるAIチップ向け回路設計技術の研究動向
３．学会等名 第36回AIチップ設計拠点フォーラム
４．発表年 2022年

１．発表者名 更田 裕司
２．発表標題 ISSCC2022におけるAIチップ研究動向
３．学会等名 第32回AIチップ設計拠点フォーラム
４．発表年 2022年

１．発表者名
更田 裕司

２．発表標題
キーワードスポッティング向けNeural ODEを用いたパラメータ削減手法の提案

３．学会等名
2021年度人工知能学会全国大会（第35回）

４．発表年
2021年

１．発表者名
更田 裕司

２．発表標題
2021 Symposium on VLSI Circuits での AI チップ研究動向

３．学会等名
第25回AIチップ設計拠点フォーラム

４．発表年
2021年

１．発表者名
更田 裕司

２．発表標題
国際会議IEDM2020 およびISSCC2021 におけるAI チップ研究開発動向

３．学会等名
IMPULSEコンソーシアム 2020 第6回セミナー（招待講演）

４．発表年
2021年

１．発表者名
更田 裕司

２．発表標題
ISSCC2021におけるAIチップ研究動向

３．学会等名
第20回AIチップ設計拠点フォーラム

４．発表年
2021年

１．発表者名
更田 裕司

２．発表標題
エッジAIを実現する低電力AIチップの研究開発動向

３．学会等名
株式会社アドダイス ウェビナー（招待講演）

４．発表年
2020年

１．発表者名
更田 裕司

２．発表標題
IEDM2020に見る半導体技術研究動向

３．学会等名
第18回AIチップ設計拠点フォーラム

４．発表年
2020年

１．発表者名
更田 裕司

２．発表標題
A-SSCCの研究動向

３．学会等名
第18回AIチップ設計拠点フォーラム

４．発表年
2020年

１．発表者名
更田 裕司

２．発表標題
エッジAIチップ研究開発動向

３．学会等名
ITmedia Virtual EXPO 2020秋（招待講演）

４．発表年
2020年

１．発表者名 更田 裕司
２．発表標題 2020 Symposia on VLSI Technology & CircuitsでのAIチップ研究動向
３．学会等名 第12回AIチップ設計拠点フォーラム
４．発表年 2020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