

研究終了報告書

「エッジ型学習ハードウェア実現に向けたインバーティブルロジックの創成」

研究期間：2018年10月～2022年3月

研究者：鬼沢 直哉

1. 研究のねらい

近年ニューラルネットワークは、特にディープラーニングの飛躍的發展に伴い、物体認識などの一部のアプリケーションにおいて人間の能力を凌駕するまでに至っている。その適用範囲は自然言語処理・自動運転・医療応用など多岐にわたり、超スマート社会を支える基盤技術となりつつある(J. Schmidhuber, et al., *Neural Networks*, 61, 85/117, 2015)。ニューラルネットワークの計算量は膨大なことから、その処理には従来のCPUによるソフトウェア処理と比較して高速処理が可能なGPUや、電力効率を大幅に向上可能なField Programmable Gate Array (FPGA) や専用IC (Application Specific Integrated Circuit (ASIC)) 等が用いられている(G. Lacey, et al., arXiv:1602.04283, 2016)。

ニューラルネットワークの処理は、大きく推論と学習の2つに分けられる。近年の研究により、推論部分の処理は固定小数点の10ビット程度で、浮動小数点処理と同等の精度を達成できることが分かってきたため、FPGAやASIC等のハードウェアによる高速化・低電力化が可能である。その一方で、学習処理には一般的に誤差逆伝搬法(勾配法)が用いられており、高い精度を実現するためには依然として浮動小数点演算が必須であることが知られている(S. Zhou, et al., arxiv:1606.06160, 2016)。CMOS微細化がほぼ限界に近づきつつある現状において、学習処理の低電力化を実現するには、勾配法とはまったく異なるハードウェアフレンドリーな新概念の学習アルゴリズム及び基盤コンピューティング技術の実現が必要不可欠であると考えに至った。

本研究課題では、従来のクラウド側でのCPU+GPUサーバを用いたソフトウェアによるニューラルネットワークの学習処理を、エッジ側で置換え可能な「低電力エッジ型学習用ハードウェア」の実現をのため、ビット列を用いた確率による数値表現・演算を行う「ストカスティック演算」と双方向(順方向・逆方向)計算を可能にする「インバーティブルロジック」の融合による、新概念コンピューティング技術の創成を目指した。

2. 研究成果

(1) 概要

インバーティブルロジックは、確率的磁気抵抗素子モデルを活用した新たな計算手法として近年提案された(K. Y. Camsari, et al., *Phy. Rev. X*, 031014, 2017)。順方向計算だけでなく、通常演算で実現が難しい逆方向計算が可能である等、非常にユニークな特徴を持つ。しかしながら、ハードウェア実現には確率的磁気抵抗素子等の研究段階の特殊デバイスが必要であるだけでなく、素子モデルが実デバイス特性とかけ離れた理想的なものである等、実際のハードウェア化には様々な問題が存在した。

本研究課題では、インバーティブルロジックのユニークな特徴が因数分解処理だけでなく、様々な逆方向計算(線形回帰や非線形回帰)等にも応用可能であることを見出し、その発展としてニューラルネットワークの学習処理への適用を目指した。そのハードウェア実現として

は、特殊デバイスを用いずにデジタル CMOS で実現可能な手法であるストカスティック演算を活用することで、現在の技術で実装可能なインバーティブルロジックハードウェアによる学習処理の実現を図った。

インバーティブルロジックは任意の関数をハミルトニアンと呼ばれるエネルギー関数にマッピングをすることで双方向計算が実現される。一方、従来研究ではハミルトニアンへのマッピングは限定された小規模な関数に限定されているなど、学習処理への応用には課題があった。そこで、研究テーマ A において、任意の関数からハミルトニアンを生成する自動設計手法を確立することで、学習処理への応用を可能にした。また、研究テーマ B において、特殊デバイスモデルを CMOS 回路で効率的に近似する手法として、確率的演算の一種であるストカスティック演算に基づくインバーティブルロジック回路を考案した。上記で確立したハミルトニアン生成手法及び回路実現手法を活用することで、研究テーマ C において、従来の学習処理手法である誤差逆伝播法との比較を行った。その結果、2層のバイナリニューラルネットワークの学習処理を FPGA 実装した結果、従来手法と同等の学習精度を保ちつつ、134 倍のエネルギー効率を達成した。本研究課題の一連の成果は IEEE Nanotechnology Magazine(2022, in press)に招待論文として掲載予定である。

(2) 詳細

研究テーマ A「大規模なハミルトニアンの生成手法の考案」

学習に用いる任意の関数からハミルトニアンを生成し、インバーティブルロジックのシミュレーション及び動作検証を可能にする、自動設計検証環境の構築を行った。図 1 に示す提案フローでは、入力としてハードウェア記述言語(Verilog-HDL 等)を用いることで、任意の関数を双方向計算可能なインバーティブルロジックへの変換を可能にした。次に、VLSI 設計で一般的に用いる論理合成ツール(Synopsys 社 Design Compiler 等)を用いて、ゲートレベルネットリストへの変換を行う。

その後、予め準備されたハミルトニアンライブラリを用いることにより、インバーティブルロジックネットリスト(ハミルトニアン)への変換を可能にするツールを作成した。 証環境

ハミルトニアンライブラリには AND 等の個々のゲートに対応するハミルトニアンが含まれており、そのハミルトニアンを線形計画法により導出する手法を考案した(代表的論文 3)。作成されたインバーティブルロジックネットリストを用いることでシミュレーションによる動作検証を行うことで、大規模なハミルトニアンに基づくインバーティブルロジックの設計を可能になった (Journal of Applied Logics, 8(5), 1311/1333, June 2021)。

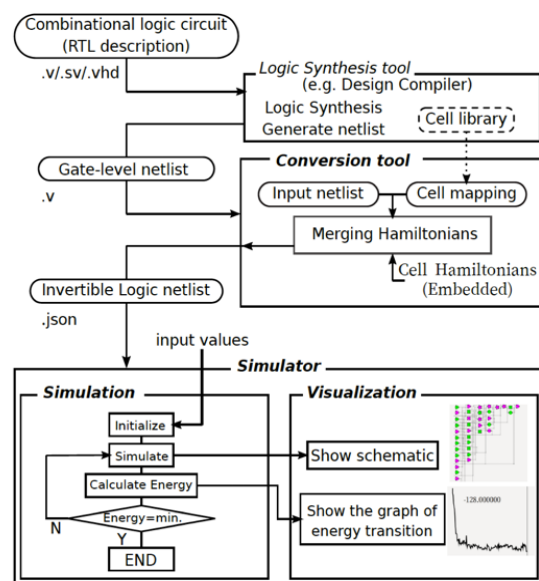


図 1 インバーティブルロジックの自動設計検証環境

研究テーマ B「ストカスティック演算の最適化によるインバーティブルロジックの高速化」

従来手法で用いられていた確率的磁気抵抗素子モデルを、ストカスティック演算に基づく CMOS 回路で近似する手法を考案した(図 2)。この研究は、共同研究先であるカナダ McGill 大学の Gross 教授グループと研究を進めた成果である。ストカスティック演算は、通常のデジタル回路では複雑化してしまう tanh などの非線形関数を効率的に回路実現可能であるため、コンパクトなハードウェアが実現された(代表的論文 1)。このような確率的演算を活用した新概念計算技術に関して、令和 2 年度科学技術分野の文部科学大臣表彰・若手科学者賞を受賞した。

インバーティブルロジック回路により双方向計算を実現するには、まず、研究テーマ A で決定したハミルトニアン係数(h, J)を近似回路にアサインする必要がある。次に、近似回路にノイズ信号を付加することにより状態 $m(t)$ を変化させ、エネルギー関数を最小化させる。シミュレーションにより、インバーティブルロジック回路の速度は、ノイズ信号の制御に強く依存することがわかったため、高速に収束可能なノイズ制御手法を考案した。考案したスパースノイズ制御方式では、従来手法のように常に一定のノイズ信号を付加するのではなく、インターバルを与えてノイズ信号を付加するものである。インターバルを与えることで、ハミルトニアンエネルギーの増加確率を減少させることが可能となり、一桁程度の収束時間の削減に成功した(IEEE Access, 9, 62890/62898, 2021)。

研究テーマ C「従来誤差逆伝播法との学習精度及び消費エネルギーの比較」

研究テーマ A, B で考案した自動設計検証環境及び回路実現法を活用することで、ニューラルネットワークの学習処理への適用を行った。提案手法の有効性を確認するために、小規模なニューラルネットワークとして、パーセプトロン (Journal of Applied Logics, 7(1), 41/58, 2020)及び図 3に示す 2 層ニューラルネットワーク (IEEE Access, 8 188004/188014, 2021)において、従来誤差逆伝播法との比較を行った。その結果、約 1/100 のデータセットで従来手法と同等の学習

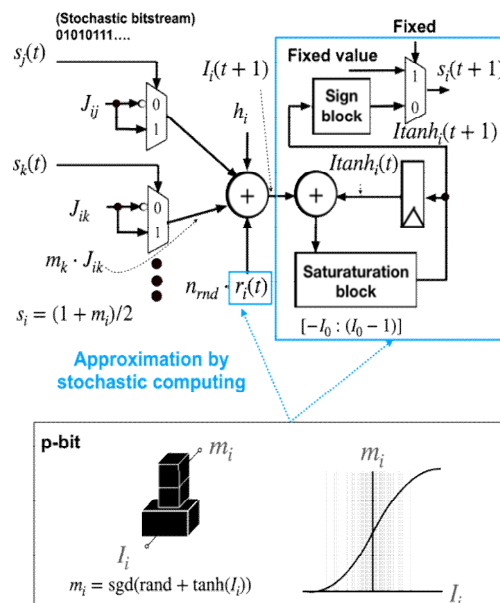


図 2 ストカスティック演算に基づく確率的磁気抵抗素子モデル(p-bit)の近似回路

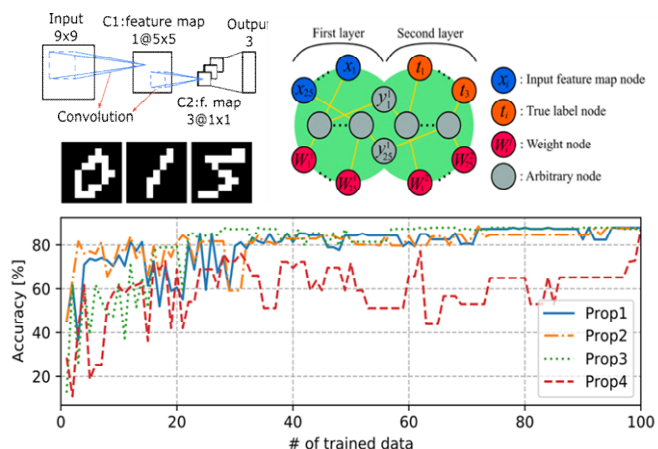


図 3 インバーティブルロジックに基づく2層ニューラルネットワークの学習処理

精度を達成することができ、ハードウェアにおけるレイテンシ及び消費エネルギーの削減効果の見込みが得られた。

提案手法の低消費電力性を実証するために、小規模なニューラルネットワークの学習処理を行うインバーティブルロジック回路をTSMC 65nm CMOSプロセスにより評価を行った(表 1)。その結果、従来の CPU 及び GPU を用いた学習処理と比較して、約 3 桁のレイテンシと消費電力削減に成功した(代表的論文 2)。従来より少ないデータセットで学習が可能となる要因は、学習アルゴリズムの違いによるものであると考えられるが、実用規模のニューラルネットワークへの適用も見据え

て詳細については現在も評価中である。また、大規模なニューラルネットワークを学習可能なインバーティブルロジック回路の実装に向けて、効率的なスパース行列に基づくハミルトニアン係数の表現手法とハードウェアアーキテクチャの考案した(IEEE Open Journal of Circuits and Systems, 2, 782/791, 2021)。

表 1 従来誤差逆伝播法との学習精度とハードウェア性能比較

	CPU	GPU	This work
Algorithm	Gradient descent (back propagation)		Invertible logic
Computing	Floating-point arithmetic		Stochastic computing
Latency	0.11 s	2.26 s	53 us
Power dissipation	180 W	225 W	32.8 mW
Energy dissipation	19.8 J	508 J	1.74 uJ
Device	Ryuen 1950X (14-nm CMOS)	NVIDIA Tesla K20 (45-nm CMOS)	ASIC (65-nm CMOS)

3. 今後の展開

本研究課題では、省エネルギーニューラルネットワークの学習処理ハードウェアを実現するための、新概念の双方向計算基盤(インバーティブルロジック)の開発を行った。開発した双方向計算基盤は、小規模な学習ハードウェアにおいて従来手法と比較して省エネルギー化の可能性を示した一方で、スケーラビリティが今後の課題である。

現状技術では大規模化するに従って、収束速度や収束確率の点で大きな問題があり、研究テーマ A,B で述べたハミルトニアン設計・ノイズ最適化に関してさらなる技術発展が必須であり、今後3年程度でスケーラビリティの課題に取り組んでいく予定である。スケーラビリティの問題が解決できれば、インバーティブルロジックが持つ双方向計算技術を活かすことで、ニューラルネットワークの学習処理に留まらず、従来計算手法では計算コストが膨大な他の応用先への適用も考慮しつつ、本提案技術の社会実装に向けて取り組んでいきたい。

4. 自己評価

研究目的の達成状況に関して、これまでの計算手法とは全く異なる新しい計算基盤であるインバーティブルロジックの創成という点においては、当初の目的通り、設計手法の確立や CMOS 回路によるハードウェア実現など概ね達成された。一方で、その応用先であるニューラルネットワークの学習処理への適用に関しては、当初の目的である MNIST の学習処理の実現までに至らず、より小規模な学習処理への適用に留まっている。課題に関しては「3. 今後の展開」にまとめた通りであり、スケーラビリティの問題解決が必須である。

研究の進め方に関しては、COVID-19以降、以前のような研究者との交流機会が激減し、なかなか新規の研究者との交流や共同研究を思うように進めることが困難であった。また、以前から交流のあった海外大学の研究グループとの student exchange も行えず、後半2年間に限っては研究費(旅費)の執行も含めて試行錯誤しつつ研究を進めるしかなかった。

研究成果の科学技術への波及効果においては、これまでにない新概念の計算基盤を確立したことで、Unconventional computing の一種として当該研究分野において一定の影響を与えており(L. Sousa, IEEE CAS Magazine, 21(1), 6/40, 2021), インバーティブルロジックの解説論文が2編(IEEE Nanotechnology Magazine 及び電子情報学会会報「講座」)掲載予定である。経済・社会への波及効果に関しては、「3. 今後の展開」で述べた課題が解決できれば、ニューラルネットワークの学習処理の大幅な省エネルギー化及び CO2 排出量の大幅な削減が期待される。人工知能の発展に伴う CO2 排出量は今後深刻になることが予測されており、本提案技術は社会・経済の変革をもたらす科学イノベーションの源泉となりうるものである。

5. 主な研究成果リスト

(1) 代表的な論文(原著論文)発表

研究期間累積件数: 9件

1. S. C. Smithson, N. Onizawa, B. H. Meyer, W. J. Gross, and T. Hanyu, "Efficient CMOS Invertible Logic Using Stochastic Computing," *IEEE Trans. on Circuits and Syst. I Reg. Papers*, vol. 66, no. 6, pp. 2263–2274, June 2019.

本論文はインバーティブルロジックをストカスティック演算に基づく近似回路で実現し、世界で初めて CMOS デジタル回路上で双方向計算を実現した研究成果をまとめたものである。これまでの確率的磁気抵抗デバイスモデルに基づくソフトウェアベースの小規模実現とは異なり、インバーティブルロジックの大規模化・ハードウェア化を可能にした。TSMC 65nm CMOS プロセス上でインバーティブルロジックの典型例である素因数分解処理を実証した。

2. N. Onizawa, S. C. Smithson, B. H. Meyer, W. J. Gross, and T. Hanyu, "In-Hardware Training Chip Based on CMOS Invertible Logic for Machine Learning," *IEEE Trans. on Circuits and Syst. I Reg. Papers*, vol. 67, no.5, pp. 1541–1550, May 2020.

本論文はインバーティブルロジックをニューラルネットワークの学習処理に適用した研究で、小規模なニューラルネットワークにおいて大幅な省エネルギー化の実証に成功した。従来の学習処理である誤差逆伝播法を用いずに、インバーティブルロジックによる学習データからの順方向及びラベルからの逆方向を含む双方向計算を活用することで、複雑な計算が必要とされる浮動小数点演算が不要な計算効率の高い学習処理を実現した。

3. N. Onizawa, K. Nishino, S. C. Smithson, B. H. Meyer, W. J. Gross, H. Yamagata, H. Fujita, and T. Hanyu, "A Design Framework for Invertible Logic," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 40, no. 4, pp. 655–665, Apr. 2021.

本論文はインバーティブルロジックの自動設計検証環境を構築した研究成果をまとめたものである。インバーティブルロジックの設計手法は仕様から実装まで複数のステップが必要である一方で、それぞれのステップは手設計かつ一貫したフローが存在せず、大規模設計が

困難であった。本設計検証環境を活用することにより、任意の関数からインバーティブルロジック回路の設計が可能となり、さまざまな応用先への適用が可能になった。

(2) 特許出願

研究期間全出願件数: 0 件 (特許公開前のものも含む)

(3) その他の成果 (主要な学会発表、受賞、著作物、プレスリリース等)

[著書] N. Onizawa, W. J. Gross and T. Hanyu, “Brain-Inspired Computing,” Chapter 9 in Book: Stochastic Computing: Techniques and Applications, pp. 185–199, Editors: V. C. Gaudet and W. J. Gross, Springer, ISBN: 978-3-030-03729-1, 2019

[受賞] 令和 2 年度科学技術分野の文部科学大臣表彰・若手科学者賞, 受賞日: 2020 年 4 月 7 日, 受賞内容: 「確率的コンピューティングに基づく脳型情報処理システム研究」

[受賞] 東北大学ディスティングイッシュュトリサーチャー, 2021 年 5 月

[招待講演] N. Onizawa, W. J. Gross, and T. Hanyu, “Stochastic-Computing Based Brainware LSI Towards an Intelligence Edge,” Special session of Machine Learning in *26th IEEE International Conference on Electronics Circuits and Systems*, Nov. 2019

[招待講演] 鬼沢直哉, “CMOS インバーティブルロジックとその学習ハードウェアへの応用展開,” 第 42 回 IBISML 研究会, 2021 年 3 月