

令和 5 年 6 月 26 日現在

機関番号：17104

研究種目：基盤研究(C)（一般）

研究期間：2018～2022

課題番号：18K04266

研究課題名（和文）レシオレス型3値連想メモリの連続駆動による脳型ハードウェアの実現に関する研究

研究課題名（英文）Study on Realization of Brain-like Hardware by Sequential Driving of Ratio-less Tri-level Associative Memory

研究代表者

中村 和之（Nakamura, Kazuyuki）

九州工業大学・マイクロ化総合技術センター・教授

研究者番号：60336097

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）：提案したNN-TCAM(Neural Network-TCAM)の基本アイデアの実証を行うために、機能設計と方式設計を行い、アーキテクチャ的な動作を確認した。まずC言語によるNN-TCAMシミュレータを開発し、基本方式を検証することに成功した。さらに、レシオレスTCAMチップを修正し、基本コア部分の回路設計を進めた。その過程で、消費電力が大きな問題となり、メモリの多段アクセスをベースとした新たな構成の検討を行い、次の研究課題「21K04179；メモリカスケード構成による記憶駆動型人工知能LSIの実現に関する研究」を提案するに至った。

研究成果の学術的意義や社会的意義

TCAM上にニューラルネットワークをエントリとして直接マッピングし、駆動演算する本提案は、学術的に全く独創的で、従来のデジタル方式とアナログ方式の中間に位置付けられ、エッジにおいて使用環境への適応化のための学習機能にも対応可能であり、高い産業的価値を持つ。長年我が国の半導体設計技術の中核であったメモリ技術の延長にあり、第3の新たな脳型ハードウェアの実現候補として、新たなメモリの応用分野を切り開くものである。

研究成果の概要（英文）：To demonstrate the basic idea of the proposed NN-TCAM (Neural Network-TCAM), I conducted functional design to confirm its architectural behavior. First, we developed an NN-TCAM simulator by C-program and successfully verified the basic scheme. Furthermore, we modified the ratio-less TCAM chip and proceeded with the circuit design of the basic core part. In the research process, we investigated a new configuration based on multi-level accesses of memory, which is a major problem for power consumption, and came to propose the following research topic, "21K04179: Research on realization of memory-driven artificial intelligence LSI with memory cascade configuration."

研究分野：工学

キーワード：ニューラルネットワーク 連想メモリ SRAM 人工知能 推論 LUTカスケード

### 1. 研究開始当初の背景

提案者らは、素子性能のばらつきや経年劣化の影響を受けず、超低電源電圧での動作を可能にしたレシオレス SRAM 技術を開発した[1]。さらにこの技術を TCAM 回路へ応用し、従来に対して1桁以上の低消費電力化を達成したレシオレス TCAM の開発に成功した[2]。近年特に注目を集めている AI 応用の脳型ハードウェアに関して、「TCAM の部分一致検索機能+ごくわずかな演算機能」をメモリ内に集積し、これを「連続参照(メモリ駆動)」することで、ニューラルネットワークに相当する処理が、メモリ LSI 内で実現できるのではないかと「問い」が提起されてきた。脳型ハードウェアとして提案・開発されているものは、主として、デジタル型とアナログ型の2つに大別することができる[3][4]。表1に示すように、デジタル型は、FPGA あるいは GP-GPU を用いたプロセッサ+メモリの構造であり、並列化と通信により学習能力を高めたもので、これらは、主としてクラウド上のサーバで、高速な学習を行うために開発された専用 LSI であるが処理の主体はソフトウェアである。一方、アナログ方式は、主としてエッジ側での推論のみに使用する目的で、例えば、ReRAM 等の不揮発素子(メモリスタ)にアナログ値を記憶させ、クロスバー構造に配置し、アナログ電流加算回路により、物理的にニューラルネットを実現するものである。

[1] T. Kondo, H. Yamamoto, S. Hoketsu, H. Imi, H. Okamura, K. Nakamura, "Ratioless full-complementary 12-transistor static random access memory for ultra low supply voltage operation", Japanese Journal of Applied Physics, Vol.54 No.4S, pp04DD11-1,04DD11-6, April, 2015

[2] D. Nishikata, M. A. Bin Mohd Ali, K. Hosoda, H. Matsumoto, K. Nakamura, " Design and measurement of fully digital ternary content addressable memory using ratioless static random access memory cells and hierarchical-AND matching comparator ", Japanese Journal of Applied Physics 57 (4S) 04FF11-1-04FF11-5, Mar 2018.

[3] 浅井, "ニューロモルフィック工学・脳型機械学習ハードウェアの行方" 日本神経回路学会誌 Vol.22, No.4, pp162-169 (2015)

[4] 森江, "脳型人工知能のためのアナログ VLSI の回路とデバイス" 日本学術振興会シリコン超集積システム第165委員会, 2017.8.25

表1 脳型ハードウェア実装方式の比較

	従来方式		提案方式
	デジタル方式	アナログ方式	メモリ方式
動作原理	FPGA/GPU+メモリ	電流加算+オペアンプ	TCAM駆動
ネットワーク構成	ソフトウェア処理	ハードワイヤード	メモリマップド
シナプス記憶	DRAM/SRAM	クロスバー型メモリスタ	連続メモリ
推論	○		○
学習		-	○
処理速度			○
消費電力	X		(レシオレス駆)
運用	クラウド・サーバ	エッジ	エッジ

: 最適、○: 適、: 可、X: 適さない

### 2. 研究の目的

我々の持つ超低電圧で動作可能な3値連想メモリ技術を基盤とした、それを連続駆動する新しいタイプのニューラルネットワーク(脳型)ハードウェアの実現性を検証することを本研究の目的とする。本研究では、申請者オリジナルのレシオレス型 SRAM 技術と、それを適用した3値連想メモリ: TCAM (Ternary Content Addressable Memory) 技術をベースとした「記憶駆動型」の全く新しい脳型ハードウェア: NN-TCAM (Neural Network-TCAM) の実現を目指す。シナプスとニューロンに相当する情報は、TCAM 内にエン트리情報としてマッピングされ、TCAM の持つ部分一致検索機能を拡張することで、記憶情報に柔軟にアクセスし、特定のニューロンへ接続する全てのシナプスの同時抽出と重み演算を行うことで前方検索(推論)を NN-TCAM 単独チップ内で行う。TCAM 上にニューラルネットワークをエン트리として直接マッピングし、駆動演算する本提案は、学術的に全く独創的で、デジタル方式とアナログ方式の中間に位置付けられ、エッジにおいて使用環境への適応化のための学習機能にも対応可能であり、高い産業的価値を持つ。長年我が国の半導体設計技術の中核であったメモリ技術の延長にあり、第3の新たな脳型ハードウェアの実現候補として、新たなメモリの応用分野を切り開くものである。

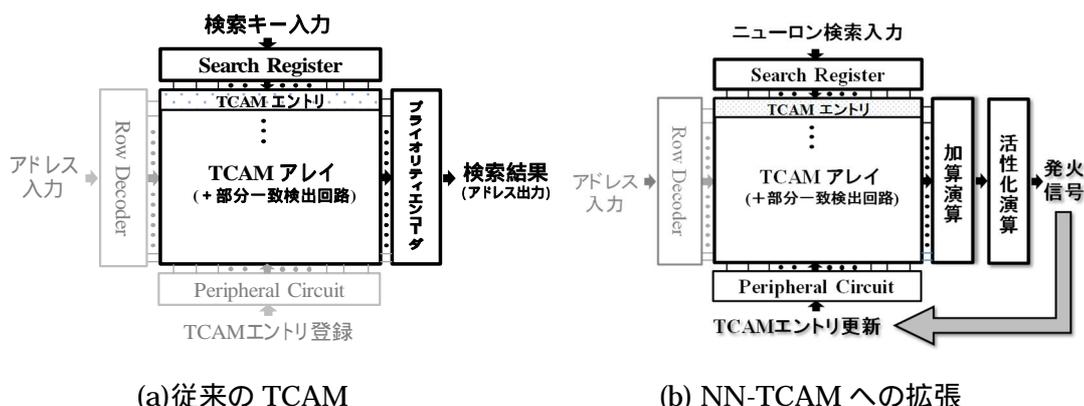
### 3. 研究の方法

NN-TCAM の基本アイデアの実証を行うために、まず機能設計と方式設計を行い、アーキテクチャ的な動作を確認する。さらに LSI としての実装可能性を検証するためには、回路・レイアウトの見地からの検討が必須である。まず、前研究で開発済みの TCAM シミュレータを拡張することで、NN-TCAM シミュレータを開発し、基本方式や追加機能ブロックについて検討する。次に 180nm で開発済みのレシオレス TCAM チップを設計変更することで、NN-TCAM の基本コア部分の具体的な回路設計を行う。

### 4. 研究成果

TCAM は、図1(a)に示すように、記憶ワード(エン트리)単位に、部分検索一致機能を設けた

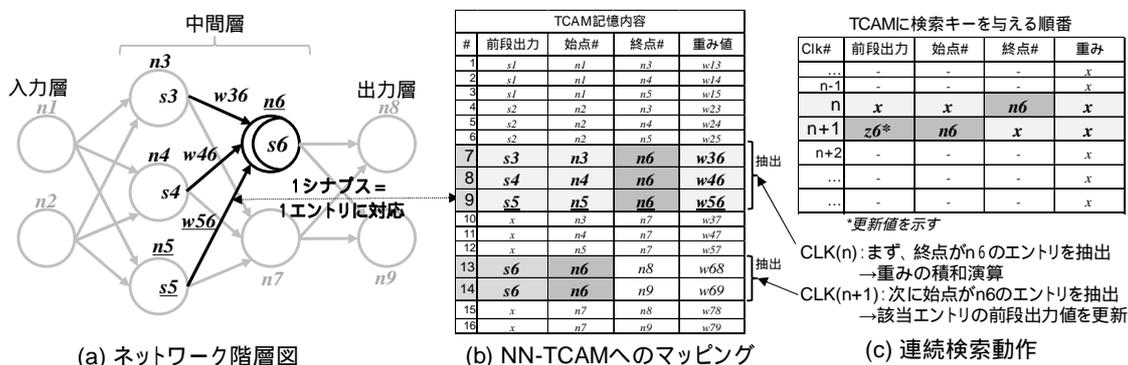
構成となっており、同時に複数のエントリが合致する場合には、プライオリティエンコーダにより優位候補を選び出力する。提案する NN-TCAM (図 1 (b)) は、従来のプライオリティエンコーダに代えて、少ビット幅の積和演算と活性化判定回路、発火情報のフィードバック更新回路等を追加した構成になる。図 2 に、NN-TCAM による NN 処理の概要例を示す。図 2 (a) に示す NN において、TCAM の 1 エントリを 1 シナプスに相当させ、図 2 (b) に示すように NN-TCAM エントリへマッピングする。図 2 (c) に示す手順で、まず終点ニューロン番号で部分一致検索をかけることで、特定のニューロンに接続されるすべてのシナプスを抽出し、その重み加算演算を 1 クロックで完了する。次のクロックで発火判定の結果を、始点ニューロン番号で部分一致検索をかけて、該当エントリの発火情報として更新する。



(a)従来の TCAM

(b) NN-TCAM への拡張

図 1 従来 TCAM から NN-TCAM への拡張



(a) ネットワーク階層図

(b) NN-TCAM へのマッピング

(c) 連続検索動作

図 2 NN-TCAM へのニューラルネットワークのマッピングと基本動作原理

C 言語を使用して NN-TCAM シミュレータを開発し、4.1 節の方法で NN 処理が実現可能か確認を行った。手書き文字認識を例題に、NN-TCAM シミュレータの記憶内容と出力結果を図 3 に示す。これにより NN-TCAM シミュレータにより、本方式はプログラムと同様の推論結果が得られることが確認できた。また、RL-TCAM に追加する積和演算器、活性化演算器の設計を行い、図 4 に示す回路構成により、活性化関数に 3 値関数を採用することで乗算回路部分が単純化でき、活性化関数も積和演算結果の一部のビットを判定するだけで発火情報を得ることができることが分かった。

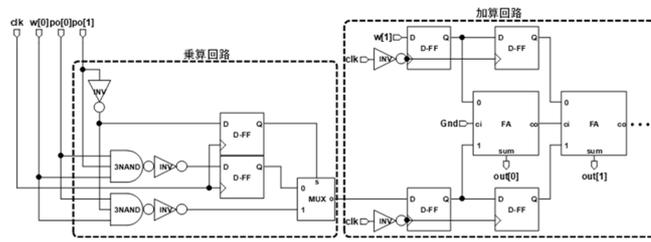
エントリ #	始点 #	終点 #	前段出力	重み
0	64ビット			
1	8ビット	8ビット	24ビット	24ビット
...				
1982				

(a) NN-TCAM シミュレータの記憶内容

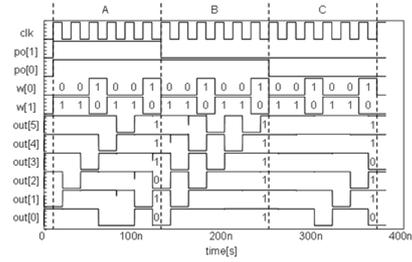
テストデータ	出力値		誤差 [%]
	プログラム	シミュレータ	
A	1.00069	1.01364	1.29
B	1.75229	1.75360	0.07
C	2.74436	2.75228	0.29

(b) 手書き文字認識結果

図 3 NN-TCAM シミュレータによる推論動作の実証



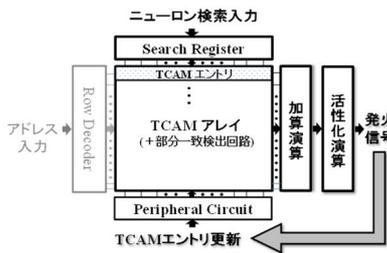
(a) 回路図



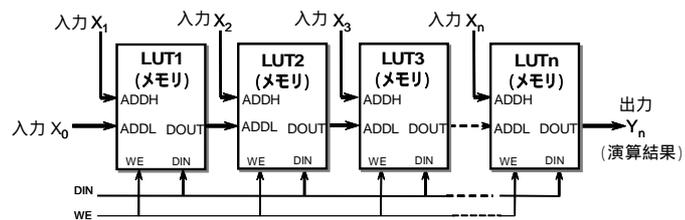
(b) シミュレーション波形

図 4 積和演算部の回路設計

さらに論理設計により動作原理の実証できたが、より具体的な回路設計の前段階で、CAM の全検索時の消費電力がかなり大きく、1 ニューロンに相当する演算を行うために、それに接続されるシナプス数の 2 倍の全検索動作が必要で、これが実用上の大きなネックとなることが分かった。低電力化を図る回路・方式等の検討を行ったが、過去に研究を行った LUT-Cascade 技術を拡張・改良することにより、NN の動作が通常のメモリの多段連続アクセスに代替可能にできるという着想を得た。その基本的な構成を図 5 (b) に示す。ハードウェア構成としては、メモリアレイが多段配置され、前段のデータ出力がそのまま次段のアドレス入力の一部になるだけの非常に単純な構成である。論理関数をメモリ内に情報としてマッピングし、任意の論理式を実現する。本提案では、4 段の LUT-Cascade ライクな構成で、1 段目のメモリアレイに重みと接続先ニューロン番号を、2 段目にニューロン保持値、3 段目には重みとニューロン値の積、4 段目には活性化関数を格納する。これらの機能拡張と改造で、汎用メモリアレイを多段連続的にアクセスするだけで、NN 処理のほとんどがメモリアクセスのみで実現可能となる。NN-TCAM と比較しても、記憶部は CAM セルから汎用のメモリセルにすることが可能となり、面積・電力・速度・コストの点で大幅な性能改善が図れることが分かり、この内容を次の基盤研究 C (21K04179)「メモリアレイ構成による記憶駆動型人工知能 LSI の実現に関する研究」として提案するに至った。



(a) NN-TCAM



(b) CMM-NN(Cascaded Memory Mapped NeuralNetwork)

図 5 新メモリ連続駆動型 NN LSI の基本構成

5. 主な発表論文等

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

1. 著者名 hikata, D.Email Author, Bin Mohd Ali, M.A., Hosoda, K., Matsumoto, H., Nakamura, K.	4. 巻 57
2. 論文標題 Design and measurement of fully digital ternary content addressable memory using ratioless static random access memory cells and hierarchical-AND matching comparator	5. 発行年 2018年
3. 雑誌名 Japanese Journal of Applied Physics	6. 最初と最後の頁 04FF11
掲載論文のDOI（デジタルオブジェクト識別子） 10.7567/JJAP.57.04FF11	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 該当する

〔学会発表〕 計0件

〔図書〕 計0件

〔出願〕 計1件

産業財産権の名称 半導体装置及びニューラルネットワークの構成方法	発明者 中村和之	権利者 九州工業大学
産業財産権の種類、番号 特許、特願2019-069485	出願年 2019年	国内・外国の別 国内

〔取得〕 計0件

〔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
---------	---------