

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6764473号
(P6764473)

(45) 発行日 令和2年9月30日 (2020.9.30)

(24) 登録日 令和2年9月15日 (2020.9.15)

(51) Int. Cl.

F I

G 0 6 N 3/063 (2006.01)
G 1 1 C 13/00 (2006.01)G 0 6 N 3/063
G 1 1 C 13/00 2 0 0

請求項の数 9 (全 26 頁)

(21) 出願番号 特願2018-514890 (P2018-514890)
 (86) (22) 出願日 平成28年9月19日 (2016.9.19)
 (65) 公表番号 特表2019-502970 (P2019-502970A)
 (43) 公表日 平成31年1月31日 (2019.1.31)
 (86) 国際出願番号 PCT/IB2016/055572
 (87) 国際公開番号 WO2017/068443
 (87) 国際公開日 平成29年4月27日 (2017.4.27)
 審査請求日 平成31年2月25日 (2019.2.25)
 (31) 優先権主張番号 14/887,564
 (32) 優先日 平成27年10月20日 (2015.10.20)
 (33) 優先権主張国・地域又は機関
 米国 (US)

(73) 特許権者 390009531
 インターナショナル・ビジネス・マシー
 ズ・コーポレーション
 INTERNATIONAL BUSIN
 ESS MACHINES CORPOR
 ATION
 アメリカ合衆国10504 ニューヨーク
 州 アーモンク ニュー オーチャード
 ロード
 New Orchard Road, A
 rmonk, New York 105
 04, United States o
 f America
 (74) 代理人 100108501
 弁理士 上野 剛史

最終頁に続く

(54) 【発明の名称】 抵抗型処理ユニット

(57) 【特許請求の範囲】

【請求項 1】

2 端子抵抗型処理ユニット (R P U) であって、
 第 1 の端子と、
 第 2 の端子と、
 伝導状態を有する活性領域と
 を含み、

前記活性領域は、前記第 1 の端子に印加された少なくとも 1 つの第 1 のコード化信号及
 び前記第 2 の端子に印加された少なくとも 1 つの第 2 のコード化信号に基づいて前記伝導
 状態の非線形変化をもたらすように構成され、

前記活性領域は、前記伝導状態の前記非線形変化に少なくとも部分的に基づいて訓練方
 法論のデータストレージ動作を局所的に行うようにさらに構成され、

前記活性領域は、前記伝導状態の前記非線形変化に少なくとも部分的に基づいて前記訓
 練方法論のデータ処理動作を局所的に行うようにさらに構成され、

前記少なくとも 1 つの第 1 のコード化信号及び前記少なくとも 1 つの第 2 のコード化信
 号のコード化は、パルスの確率的シーケンスを含む、

2 端子 R P U。

【請求項 2】

前記少なくとも 1 つの第 1 のコード化信号及び前記少なくとも 1 つの第 2 のコード化信
 号のコード化は、大きさ変調を含む、請求項 1 に記載の 2 端子 R P U。

10

20

【請求項 3】

前記非線形変化は、整流的非線形変化又は飽和的非線形変化を含む、請求項 1 に記載の 2 端子 R P U。

【請求項 4】

前記非線形変化は、指数関数的非線形変化を含む、請求項 1 に記載の 2 端子 R P U。

【請求項 5】

訓練可能なクロスバーアレイであって

伝導性行ワイヤの組と、

前記伝導性行ワイヤの組との間の交差部に複数のクロスポイントを形成するように構成された伝導性列ワイヤの組と、

前記複数のクロスポイントの各々における 2 端子抵抗型処理ユニット (R P U) と、
を含み、

前記 2 端子 R P U は、第 1 の端子と、第 2 の端子と、伝導状態を有する活性領域とを含み、

前記活性領域は、前記第 1 の端子に印加された少なくとも 1 つの第 1 のコード化信号及び前記第 2 の端子に印加された少なくとも 1 つの第 2 のコード化信号に基づいて前記伝導状態の非線形変化をもたらすように構成され、

前記活性領域は、前記伝導状態の前記非線形変化に少なくとも部分的に基づいて訓練方法論のデータストレージ動作を局所的に行うようにさらに構成され、

前記活性領域は、前記伝導状態の前記非線形変化に少なくとも部分的に基づいて前記訓練方法論のデータ処理動作を局所的に行うようにさらに構成され、

前記少なくとも 1 つの第 1 のコード化信号及び前記少なくとも 1 つの第 2 のコード化信号のコード化は、パルスの確率的シーケンスを含む、

アレイ。

【請求項 6】

2 端子抵抗型処理ユニット (R P U) を形成する方法であって、

第 1 の端子を形成することと、

第 2 の端子を形成することと、

伝導状態を有する活性領域を形成することと、

前記活性領域を、前記第 1 の端子に印加された少なくとも 1 つの第 1 のコード化信号及び前記第 2 の端子に印加された少なくとも 1 つの第 2 のコード化信号に基づいて前記伝導状態の非線形変化をもたらすように構成することと、

前記活性領域を、前記伝導状態の前記非線形変化に少なくとも部分的に基づいて訓練方法論のデータストレージ動作を局所的に行うように構成することと、

前記活性領域を、前記伝導状態の前記非線形変化に少なくとも部分的に基づいて前記訓練方法論のデータ処理動作を局所的に行うように構成することと、

を含み、

前記少なくとも 1 つの第 1 のコード化信号及び前記少なくとも 1 つの第 2 のコード化信号のコード化は、パルスの確率的シーケンスを含む、

方法。

【請求項 7】

前記少なくとも 1 つの第 1 のコード化信号及び前記少なくとも 1 つの第 2 のコード化信号のコード化は、大きさ変調を含む、請求項 6 に記載の方法。

【請求項 8】

前記非線形変化は、整流的非線形変化又は飽和的非線形変化を含む、請求項 6 に記載の方法。

【請求項 9】

前記非線形変化は、指数関数的非線形変化を含む、請求項 7 に記載の方法。

【発明の詳細な説明】**【技術分野】**

【 0 0 0 1 】

本開示は、一般に、本明細書では抵抗型処理ユニット（RPU）と称する、訓練可能な抵抗型クロスポイントデバイスの新規構成に関する。より詳細には、本開示は、2端子RPU以外の付加的な処理要素を必要とせず、局所的データストレージ及び局所的データ処理を提供する2端子RPUのクロスバーアレイから形成される人工ニューラルネットワーク（ANN）に関し、これにより、ANNがオンライン・ニューラルネットワーク訓練、行列反転、行列分解などのようなアルゴリズムを学習し及び実装する能力が加速される。

【 背景技術 】

【 0 0 0 2 】

「機械学習」は、データから学習する電子システムの主機能を説明するのに広範に使用されている。機械学習及び認知科学において、ANNは、動物の特に脳における生物学的ニューラルネットワークに触発された、統計的学習モデルのファミリーである。ANNを使用して、多数の入力に依存する一般に未知のシステム及び機能を推定し、又は近似することができる。

【 0 0 0 3 】

ANNは、模擬「ニューロン」としての役割を果たして相互間で電子信号形態の「メッセージ」を交換する相互接続されたプロセッサ要素の、いわゆる「ニューロモフィック」システムとしてしばしば具体化される。生物学的ニューロン間でメッセージを運ぶシナプス神経伝達物質接続のいわゆる「可塑性」と同様に、模擬ニューロン間で電子メッセージを運ぶ、ANNにおける接続には、所与の接続の強さ又は弱さに対応する数値重みを与えられる。重みを経験に基づいて調整し及びチューニングして、ANNを入力に対して適応的なもの及び学習可能なものにすることができる。例えば、手書き文字認識用のANNは、入力イメージの画素によって活性化され得る入力ニューロンの組によって定められる。ネットワークの設計者によって決定された関数によって重み付けされ変換された後、これら入力ニューロンの活性化は、しばしば「隠れ」ニューロンと称されるその他の下流ニューロンへと渡される。このプロセスは、出力ニューロンが活性化されるまで繰り返される。活性化された出力ニューロンは、どの文字を読み取ったのかを定める。

【 0 0 0 4 】

クロスバーアレイは、クロスポイントアレイ又はクロスワイヤアレイとしても知られ、ANNアーキテクチャ、ニューロモフィック・マイクロチップ及び超高密度不揮発性メモリを含む多様な電子回路及びデバイスを形成するために使用される、高密度低コストの回路アーキテクチャである。基本的なクロスバーアレイ構成は、伝導性行ワイヤの組と、この伝導性行ワイヤの組と交差するように形成された伝導性列ワイヤの組とを含む。2つのワイヤの組の間の交差部は、薄膜材料から形成することができる、いわゆるクロスポイントデバイスによって分離される。

【 0 0 0 5 】

クロスポイントデバイスは、事実上、ANNのニューロン間の重み付き接続として機能する。高いエネルギー効率を有するシナプス可塑性をエミュレートするために、ナノスケールの2端子デバイス、例えば「理想的な」伝導状態のスイッチング特性を有するメモリスタ（memristor）が、しばしばクロスポイントデバイスとして使用される。理想的なメモリスタ材料の伝導状態（例えば抵抗）は、行ワイヤ及び列ワイヤの個々のワイヤ間に印加される電圧を制御することによって変更することができる。デジタルデータは、高伝導状態又は低伝導状態を達成する、交差部におけるメモリスタ材料の伝導状態の変更によって格納することができる。メモリスタ材料は、材料の伝導状態を選択的に設定することによって、2以上の相異なる伝導状態を維持するようにプログラムすることもできる。メモリスタ材料の伝導状態は、該材料の両端に電圧を印加し、対象のクロスポイントデバイスを通る電流を測定することによって読み出すことができる。

【 0 0 0 6 】

電力消費を制限するために、ANNチップアーキテクチャのクロスポイントデバイスは、オフライン学習技術を利用するように設計されることが多く、この場合、目的関数の近

10

20

30

40

50

似は、ひとたび初期訓練相が解決されると変化しない。オフライン学習は、クロスバー型ANNアーキテクチャのクロスポイントデバイスを、非常に少ない電力しか消費しないように単純化することを可能にする。

【0007】

電力消費をより低くする可能性があるにもかかわらず、オフライン学習を実行することは、訓練データに対して入力 - 出力ペアを適合させるために典型的には訓練中にANNモデル内の著しい数の調整可能なパラメータ（例えば重み）を修正することを必要とするので、困難かつリソース集約的なものになることがある。従って、電力節約型のオフライン学習技術を優先するためにANNアーキテクチャのクロスポイントデバイスを単純化することは、典型的には、訓練速度及び訓練効率は最適化されないことを意味する。

10

【先行技術文献】

【非特許文献】

【0008】

【非特許文献1】Anil K. Jain、Jianchang Mao及びK. M. Mohiuddin、「Artificial Neural Networks: A Tutorial」、IEEE、1996年3月

【非特許文献2】C. Lehmann他、「A Generic Systolic Array Building Block For Neural Networks with On-Chip Learning」、IEEE Transactions On Neural Networks、1993年5月、第4巻、第3号

20

【非特許文献3】V. K. Chippa他、「StoRM: A Stochastic Recognition and Mining Processor」、Proceedings of the 2014 International Symposium On Low power Electronics and Design

【非特許文献4】Chua, L. O., 「Resistance Switching Memories are Memristors」、Applied Physics A (2011年)、102(4): 765 - 783頁

【非特許文献5】D. Soudry他、「Memristor - Based Multilayer Neural Networks With Online Gradient Descent Training」、IEEE Transactions On Neural Networks and Learning Systems (2015年)

30

【発明の概要】

【発明が解決しようとする課題】

【0009】

電力消費を許容可能な範囲内に保持するとともにANNアーキテクチャの訓練速度及び効率を加速する単純なクロスポイントデバイスを提供することは、全体としてのANN性能を改善し、より広範囲のANN用途を可能にするであろう。

【課題を解決するための手段】

【0010】

40

実施形態は、第1の端子、第2の端子及び活性領域を有する2端子抵抗型処理ユニット(RPU)に向けられる。活性領域は、RPUに適用される訓練方法論の重みを識別する伝導状態を有する。活性領域は、訓練方法論のデータストレージ動作を局所的に行うように構成され、かつ、活性領域は、訓練方法論のデータ処理動作を局所的に行うようにさらに構成される。

【0011】

1つ又は複数の上記実施形態において、データストレージ動作は、データ処理動作の結果に少なくとも部分的に基づく伝導状態の変化を含む。

【0012】

1つ又は複数の上記実施形態において、伝導状態の変化は、第1の端子に印加された少

50

なくとも1つの第1のコード化信号及び第2の端子に印加された少なくとも1つの第2のコード化信号に基づく非線形変化を含む。

【0013】

1つ又は複数の上記実施形態において、活性領域は、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータストレージ動作を局所的に行うようにさらに構成される。活性領域は、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論の前記データ処理動作を局所的に行うようにさらに構成することができる。

【0014】

1つ又は複数の上記実施形態において、訓練方法論は、オンライン・ニューラルネットワーク訓練、行列反転又は行列分解とすることができる。

10

【0015】

実施形態は、第1の端子、第2の端子及び活性領域を有する2端子RPUにさらに向けられる。活性領域は、第1の端子に印加された少なくとも1つの第1のコード化信号及び第2の端子に印加された少なくとも1つの第2のコード化信号に基づいて活性領域の伝導状態の非線形変化をもたらす。活性領域は、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータストレージ動作を局所的に行うように構成される。活性領域は、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータ処理動作を局所的に行うようにさらに構成される。

【0016】

1つ又は複数の上記実施形態において、少なくとも1つの第1のコード化信号及び少なくとも1つの第2のコード化信号のコード化は、パルスの確率的シーケンス又は大きさ(magnitude)変調を含む。

20

【0017】

1つ又は複数の上記実施形態において、非線形変化は、整流的非線形変化、飽和的非線形変化又は指数関数的非線形変化を含む。

【0018】

実施形態は、伝導性行ワイヤの組と、伝導性行ワイヤの組との間の交差部に複数のクロスポイントを形成するように構成された伝導性列ワイヤの組と、を有する訓練可能なクロスバーアレイにさらに向けられる。アレイは、複数のクロスポイントの各々における2端子抵抗型処理ユニット(RPU)を含む。RPUは、データストレージ動作を局所的に行うように、及びデータ処理動作を局所的に行うように構成され、局所的データストレージ動作及び局所的データ処理動作は両方とも、訓練可能なクロスバーアレイに適用される訓練方法論の一部である。

30

【0019】

1つ又は複数の上記実施形態において、2端子RPUは、第1の端子と、第2の端子と、伝導状態を有する活性領域とを含み、伝導状態は、RPUに適用された訓練方法論の重みを識別する。

【0020】

1つ又は複数の上記実施形態において、データストレージ動作は、データ処理動作の結果に少なくとも部分的に基づく伝導状態の変化を含む。伝導状態の変化は、第1の端子に印加された少なくとも1つの第1のコード化信号及び第2の端子に印加された少なくとも1つの第2のコード化信号に基づく非線形変化を含む。

40

【0021】

1つ又は複数の上記実施形態において、活性領域は、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータストレージ動作を局所的に行う。活性領域は、さらに、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータ処理動作を局所的に行う。訓練方法論は、オンライン・ニューラルネットワーク訓練、行列反転及び行列分解を含む。

【0022】

実施形態は、2端子RPUを形成する方法にさらに向けられる。方法は、第1の端子を

50

形成することと、第2の端子を形成することと、伝導状態を有する活性領域を形成することとを含む。方法は、活性領域を、第1の端子に印加された少なくとも1つの第1のコード化信号及び第2の端子に印加された少なくとも1つの第2のコード化信号に基づいて伝導状態の非線形変化をもたらすように構成することをさらに含む。方法は、活性領域を、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータストレージ動作を局所的に行うように構成することをさらに含む。方法は、活性領域を、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータ処理動作を局所的に行うように構成することをさらに含む。

【0023】

上記方法の1つ又は複数の実施形態において、少なくとも1つの第1のコード化信号及び少なくとも1つの第2のコード化信号のコード化は、パルスの確率的シーケンス又は大きさ変調を含む。

【0024】

上記方法の1つ又は複数の実施形態において、非線形変化は、整流的非線形変化、飽和的非線形変化又は指数関数的非線形変化を含む。

【0025】

実施形態は、訓練可能なクロスパーアレイを形成する方法にさらに向けられる。方法は、伝導性行ワイヤの組を形成することと、伝導性行ワイヤの組との間の交差部に複数のクロスポイントを形成するように構成された伝導性列ワイヤの組を形成することと、を含む。方法は、複数のクロスポイントの各々において2端子RPUを形成することをさらに含む。方法は、RPUを、訓練可能なクロスパーアレイに適用された訓練方法論のデータストレージ動作を局所的に行うように構成することと、RPUを、訓練方法論のデータ処理動作を局所的に行うように構成することと、をさらに含む。

【0026】

上記方法の1つ又は複数の実施形態において、2端子RPUを形成することは、第1の端子を形成することと、第2の端子を形成することと、伝導状態を有する活性領域を形成することとを含み、伝導状態は、RPUに適用された訓練方法論の重みを識別する。

【0027】

上記方法の1つ又は複数の実施形態において、データストレージ動作は、データ処理動作の結果に少なくとも部分的に基づく伝導状態の変化を含み、伝導状態の変化は、第1の端子に印加された少なくとも1つの第1のコード化信号及び第2の端子に印加された少なくとも1つの第2のコード化信号に基づく非線形変化を含む。

【0028】

上記方法の1つ又は複数の実施形態において、方法は、活性領域を、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータストレージ動作を局所的に行うように構成することをさらに含む。方法は、活性領域を、伝導状態の非線形変化に少なくとも部分的に基づいて訓練方法論のデータ処理動作を局所的に行うように構成することをさらに含む。

【0029】

上記方法の1つ又は複数の実施形態において、訓練方法論は、オンライン・ニューラルネットワーク訓練、行列反転又は行列分解を含む。

【0030】

さらなる特徴及び利点は、本明細書で説明する技術を通じて実現される。その他の実施形態および態様は、本明細書において詳細に説明される。より良く理解するために、説明及び図面を参照されたい。

【0031】

実施形態とみなされる主題は、特許請求の範囲及び本明細書の結論部において具体的に指摘され、明確に特許請求される。実施形態の上記及びその他の特徴及び利点は、添付の図面との関連で解釈される以下の詳細な説明から明らかである。

【図面の簡単な説明】

10

20

30

40

50

【 0 0 3 2 】

【図 1】生物学的ニューロンの入力及び出力接続の略図を示す。

【図 2】図 1 に示す生物学的ニューロンの公知の簡易モデルを示す。

【図 3】図 2 の生物学的ニューロンモデルを取り入れた A N N の公知の簡易モデルを示す。

【図 4】公知の重み更新方法論の略ブロック図を示す。

【図 5】1 つ又は複数の実施形態において使用可能な確率的計算方法論の略ブロック図を示す。

【図 6】受動的 2 端子メモリストの動作を支配する公知の式を示す。

【図 7】公知の 2 端子メモリストの線形スイッチング特性と本開示の 2 端子 R P U の非線形特性との間の図式的比較を示す。 10

【図 8】本開示による 2 端子非線形 R P U デバイスのクロスバーアレイを、R P U の動作を例証する電圧シーケンスと共に示す。

【図 9】1 つ又は複数の実施形態による確率的重み更新方法論の略ブロック図を示す。

【図 1 0】本開示の確率的重み更新方法論の異なるビット長 (B L) に対して達成することができる異なる重み更新精度を使用して M N I S T データセット上で訓練された A N N の分類精度を例証するグラフを示す。

【図 1 1】(A) ~ (D) は、1 つ又は複数の実施形態による 2 端子非線形 R P U を使用して確率的重み更新を実装するためのさらなる詳細を例証する略図及びグラフを示す。

【図 1 2】(A) ~ (B) は、1 つ又は複数の実施形態による正及び負の確率的重み更新サイクルを例証する電圧シーケンスを示す。 20

【図 1 3】1 つ又は複数の実施形態による 2 端子非線形 R P U を使用した確率的重み更新の実装についてのさらなる詳細を例証する略図及びグラフを示す。

【図 1 4】(A) ~ (B) は、公知の 2 端子メモリストの線形スイッチング特性と、本開示の 2 端子 R P U の非線形スイッチング特性との間の図式的比較を示す。

【図 1 5】1 つ又は複数の実施形態による指数関数的非線形性を有する 2 端子 R P U を使用した高さ変調重み更新方法論を例証するグラフ、図及び式を示す。

【図 1 6】1 つ又は複数の実施形態による 2 端子非線形 R P U を使用した高さ変調重み更新方法論をさらに例証するグラフ及び式を示す。

【図 1 7】本開示の 2 端子非線形 R P U のクロスバーアレイを含む A N N アーキテクチャの開発、訓練及び使用の態様を示す。 30

【図 1 8】本開示の 2 端子非線形 R P U のクロスバーアレイを含む A N N アーキテクチャの開発、訓練及び使用のさらなる態様を示す。

【図 1 9】本開示の 2 端子非線形 R P U のクロスバーアレイを含む A N N アーキテクチャの開発、訓練及び使用のさらなる態様を示す。

【図 2 0】1 つ又は複数の実施形態によるノード制御システムのブロック図を示す。

【図 2 1】本開示の 1 つ又は複数の実施形態による方法論を例証するフロー図を示す。

【図 2 2】本開示の 1 つ又は複数の実施形態による別の方法論を例証するフロー図を示す。

【発明を実施するための形態】 40

【 0 0 3 3 】

1 つ又は複数の実施形態は、特に脳の構造及び機能をモデル化することに重点を置いた生物学的ニューラルネットワークの文脈で開示されるが、本明細書で記述する教示の実施は特定の環境のモデル化に限定されないことがあらかじめ理解される。むしろ本開示の実施形態は、環境に対する様々な入力をベクトルに変えることができる限り、例えば気象パターン、インターネットから収集される任意のデータなどを含む、あらゆるタイプの環境をモデル化することが可能である。

【 0 0 3 4 】

本開示は、電子システムに向けられたものであるが、参照及び説明を容易にするために、本開示の電子システムの種々の態様を、例えばニューロン、可塑性及びシナプス等の神 50

経学的用語を使用して説明する。本明細書における電子システムのあらゆる議論又は例証について、神経学的用語又は神経学的略記法の使用は参照を容易にするためのものであり、記述した神経学的機能又は神経学的構成要素のニューロモーフィックなANN等価物に及ぶことを意味することが理解されるであろう。

【0035】

ANNは、ニューロモーフィック又はシナプトロニック・システムとしても知られており、例えば、生物学的神経システム、画像認識、音声認識などのようなヒトの脳及び脳様の機能を含むその他の機能又はシステムを推定又は近似することができる、計算システムである。ANNは、神経生理学、認知科学/心理学、物理学(統計力学)、制御理論、コンピュータ科学、人工知能、統計学/数学、パターン認識、計算機視覚、並列処理及びハードウェア(例えば、デジタル/アナログ/VLSI/光学)を含む多様な分野からの知識を取り入れている。

10

【0036】

0及び1を操作する伝統的なデジタルモデルを利用する代わりに、ANNは、推定又は近似されるコアシステム機能の実質的な機能的等価物である処理要素間に接続を生成する。例えば、IBM(IBM社の登録商標)のSyNapseコンピュータチップは、哺乳類の脳に類似した形態、機能及びアーキテクチャを提供することを企図した電子的ニューロモーフィックマシンの中心的コンポーネントである。IBM(IBM社の登録商標) SyNapseコンピュータチップは、従来のコンピュータチップと同じ基本トランジスタコンポーネントを使用しているが、そのトランジスタは、ニューロン及びそのシナプス接続の挙動を模倣するように構成されている。IBM(IBM社の登録商標) SyNapseコンピュータチップは、生物学的ニューロン間のシナプス通信と類似した電気スパイクを用いて互いに通信する、百万個を超えるほどの模擬「ニューロン」のネットワークを使用して情報を処理する。IBM(IBM社の登録商標) SyNapseアーキテクチャは、メモリ(すなわち模擬「シナプス」)を読み出して単純な動作を行う、プロセッサ(すなわち模擬「ニューロン」)の構成を含む。典型的には異なるコア内に位置するこれらのプロセッサ間の通信は、オンチップ・ネットワークルータによって行われる。

20

【0037】

背景として、典型的なANNがどのように動作するかの一般的な説明を、ここで図1、図2及び図3を参照して提供する。本明細書で前述したように、典型的なANNは、約1000億もの相互接続されたニューロンと呼ばれる細胞を含むヒトの脳をモデル化する。図1は、上流の入力112、114、下流の出力116及び下流の「他の」ニューロン118に接続する経路104、106、108、110を有する、生物学的ニューロン102の略図を示す。各生物学的ニューロン102は、経路104、106、108、110を通して電気インパルスを送り及び受け取る。これらの電気インパルスの性質及びそれらが生物学的ニューロン102内でどのように処理されるかということが、主として脳全体の機能を担う。生物学的ニューロン間の経路接続は、強い場合もあり又は弱い場合もある。所与のニューロンが入力インパルスを受け取ると、そのニューロンは、ニューロンの機能に従って入力処理し、その機能の結果を下流の出力及び/又は下流の「他の」ニューロンへ送る。

30

40

【0038】

生物学的ニューロン102は、図2において、図2に示す式で表される数学的関数 $f(x)$ を有するノード202としてモデル化される。ノード202は、入力212、214から電気信号を取得し、各入力212、214に対してそれぞれの接続経路204、206の強さを乗算し、入力の和を取り、その和を関数 $f(x)$ に代入して結果216を生成し、これは、最終出力、若しくは別のノードへの入力、又はその両方とすることができる。本開示において、アスタリスク(*)は、乗算を表すために使用される。弱い入力信号には非常に小さい接続強さ数が乗算されるので、弱い入力信号が関数に与える影響は非常に低い。同様に、強い入力信号にはより高い接続強さ数が乗算されるので、強い入力信号が関数に与える影響はより大きくなる。関数 $f(x)$ は、設計上の選択であり、様々

50

な関数を使用することができる。 $f(x)$ についての典型的な設計上の選択は双曲線正接関数であり、これは、前の和をとり、 -1 と $+1$ との間の数を入力する関数である。

【0039】

図3は、重み付き有向グラフとして編成された単純化ANNモデル300を示し、ここで人工ニューロンはノード（例えば302、308、316）であり、重み付き有向エッジ（例えば、 m_1 から m_{20} まで）がノードを接続する。ANNモデル300は、ノード302、304、306が入力層ノードであり、ノード308、310、312、314が隠れ層ノードであり、ノード316、318が出力層ノードとなるように編成されている。各ノードは、隣接する層のあらゆるノードに接続経路によって接続され、この接続経路は図3において接続強さ m_1 から m_{20} までを有する有向矢印として描かれている。1つの入力層、1つの隠れ層及び1つの出力層のみが示されているが、実際には、複数の入力層、隠れ層及び出力層を設けることができる。

10

【0040】

ヒトの脳の機能と同様に、ANN300の各入力層ノード302、304、306は、入力 x_1 、 x_2 、 x_3 をソース（図示せず）から直接、接続強さの調整及びノード加算を伴わずに受け取る。従って、図3の下部に挙げた式で示されるように、 $y_1 = f(x_1)$ 、 $y_2 = f(x_2)$ 及び $y_3 = f(x_3)$ である。各隠れ層ノード308、310、312、314は、その入力を、すべての入力層ノード302、304、306から、該当する接続経路に関連付けられた接続強さに従って受け取る。それゆえ、隠れ層ノード308において、 $y_4 = f(m_1 * y_1 + m_5 * y_2 + m_9 * y_3)$ であり、ここで $*$ は乗算を表す。図3の下部に示した関数 y_5 から y_9 までを定義する式によって示すように、隠れ層ノード310、312、314及び出力層ノード316、318について、同様の接続強さ乗算及びノード加算が行われる。

20

【0041】

ANNモデル300は、データレコードを一度に1つずつ処理するものであり、初期の任意のレコード分類を既知の実際のレコード分類と比較することによって「学習」する。「バックプロパゲーション（誤差逆伝播法）」（すなわち「後方への誤差伝播」）として知られる訓練方法論を使用して、第1のレコードの初期分類に由来する誤りをネットワークへフィードバックし、これを使用して2回目にネットワークの重み付き接続を修正し、このフィードバックプロセスを多数回の反復にわたって続ける。ANNの訓練相においては、各レコードについての正しい分類は既知であり、したがって出力ノードに「正しい」値を割り当てることができる。例えば、正しいクラスに対応するノードに対してノード値「1」（又は0.9）、その他のノードに対してノード値「0」（又は0.1）とする。それゆえ、その出力ノードに対してネットワークが計算した値をこれらの「正しい」値と比較して、各ノードについて誤差項を計算することが可能である（すなわち「デルタ」則）。次にこれらの誤差項を使用して隠れ層における重みを調整して、次の反復において出力値が「正しい」値により近づくようにする。

30

【0042】

多くのタイプのニューラルネットワークが存在するが、2つの最大カテゴリはフィードフォワード型（順伝播型）及びフィードバック型（逆伝播型）/リカレント型（回帰型）ネットワークである。ANNモデル300は、入力層、出力層及び隠れ層を有する非リカレント型のフィードフォワード型ネットワークである。信号は一方向のみに移動することができる。入力データは、計算を行う処理要素の層へ渡される。各処理要素は、その入力の重み付き和に基づいてその計算を行う。新たに計算された値が、次いで次層へ送られる新たな入力値になる。このプロセスは、すべての層を通過し終えて出力が決定されるまで続く。出力層内のニューロンの出力を定量するために、時には閾値伝達関数が使用される。

40

【0043】

フィードバック型/リカレント型ネットワークはフィードバック経路を含むものであり、これは信号がループを使用して両方向に移動できることを意味する。ノード間のすべて

50

の可能な接続が許容される。このタイプのネットワーク内にはループが存在するので、特定の動作の下で、これは、平衡状態に達するまで連続的に変化する非線形動的システムとなり得る。フィードバック型ネットワークは、連合記憶及び最適化問題においてしばしば使用され、このときネットワークは、相互接続した因子の最良配置を探し求める。

【0044】

フィードフォワード型及びリカレント型のANNアーキテクチャにおける機械学習の速度及び効率は、ANNクロスバーレイのクロスポイントデバイスが、典型的な機械学習アルゴリズムのコア動作をどれほど効率的に行うかに依存する。機械学習の正確な定義を公式化することは難しいが、ANN文脈における学習プロセスは、クロスポイントデバイスの接続重みを、ネットワークが特定のタスクを効率的に行うことができるように更新する問題として捉えることができる。クロスポイントデバイスは、典型的には、利用可能な訓練パターンから必要な接続重みを学習する。性能は、ネットワーク内の重みを反復的に更新することによって経時的に向上する。専門家によって指定された規則の組に従う代わりに、ANNは、根底にある（入力 - 出力関係の様な）規則を所与の代表例の集まりから「学習」する。従って、学習アルゴリズムは、一般に、学習規則を使用して該当する重みを更新する及び/又は調整する手順として定義することができる。

【0045】

3つの主要な学習アルゴリズムのパラダイムは、教師付き(supervised)、教師なし(unsupervised)及びハイブリッドである。教師付き学習、すなわち「教師(teacher)」を伴う学習において、ネットワークには、すべての入力パターンに対して正解(出力)が与えられる。重みは、ネットワークが既知の正解にできる限り近い解答を生成することを可能にするように決定される。強化学習は、教師付き学習の変形であり、ネットワークには、正解そのものではなく、ネットワーク出力の正しさに対する批評のみが与えられる。対照的に、教師なし学習、すなわち教師を伴わない学習は、訓練データセット内の各入力パターンに関連付けられた正解を必要としない。これは、データにおける根底にある構造、又はデータにおけるパターン間の相関を探求し、これらの相関からパターンをカテゴリに編成する。ハイブリッド学習は、教師付き学習と教師なし学習とを組み合わせたものである。重みの一部は、一般に教師付き学習を通じて決定され、その他は、教師なし学習を通じて得られる。ANN及び学習規則のさらなる詳細は、非特許文献1に記載されている。

【0046】

本明細書において前述したように、電力消費を制限するために、ANNチップアーキテクチャのクロスポイントデバイスは、オフライン学習技術を利用するように設計されることが多く、この場合、目的関数の近似は、ひとたび初期訓練相が解決されると変化しない。オフライン学習は、クロスバー型ANNアーキテクチャのクロスポイントデバイスを、非常に少ない電力しか消費しないように単純化することを可能にする。

【0047】

電力消費をより低くする可能性があるにもかかわらず、オフライン学習を実行することは、訓練データに対して入力 - 出力ペアを適合させるために典型的には訓練中にANNモデル内の著しい数の調整可能なパラメータ(例えば重み)を修正することを必要とするので、困難かつリソース集約的なものとなることがある。図4は、典型的な読出し - 処理 - 書込みの重み更新動作の略図を示したものであり、CPU/GPUコア(すなわち模擬「ニューロン」)は、メモリ(すなわち模擬「シナプス」)を読み出し、重み更新処理動作を行い、次いで更新された重みを再びメモリに書き込む。従って、電力節約型のオフライン学習技術を優先するためにANNアーキテクチャのクロスポイントデバイスを単純化することは、典型的には、訓練速度及び訓練効率は最適化されないことを意味する。

【0048】

電力消費を許容可能な範囲内に保持するとともにANNアーキテクチャの訓練速度及び効率を加速する単純なクロスポイントデバイスを提供することは、全体としてのANN性能を改善し、より広範囲のANN用途を可能にするであろう。

【 0 0 4 9 】

ここで本開示の概要に目を向けると、1つ又は複数の実施形態は、本明細書において抵抗型処理ユニット(RPU)と称される2端子のプログラム可能な抵抗型クロスポイントコンポーネントに向けられ、これは局所的データストレージ機能及び局所的データ処理機能を提供する。換言すれば、データ処理を行うとき、各RPUに格納された値は、並列的かつ局所的に更新され、そのことにより、該当データをプロセッサ及び別個のストレージ要素から出し入れして移動させる必要性が解消される。加えて、本開示の2端子RPUによって提供される局所的データストレージ及び局所的データ処理は、ANNが、バックプロパゲーション型オンライン・ニューラルネットワーク訓練、行列反転、行列分解などのようなアルゴリズムを学習し及び実装する能力を加速する。従って、本開示のRPUを有する機械学習ANNアーキテクチャを実装することは、学習の速度、効率及び電力消費を最適化するオンライン機械学習能力の実装を可能にする。本開示のRPU、及び結果として得られるANNアーキテクチャは、全体としてのANN性能を改善し、より広範囲の実用的なANN用途を可能にする。

10

【 0 0 5 0 】

本開示のRPUは、2端子抵抗型クロスポイントデバイスとして実装することができ、それらのスイッチング特性は、データを処理するために使用することができる非線形性を有する。それゆえ、本開示のRPUは、ANNにおいて計算を行うために使用することができる適切な非線形特性を有する、任意の2端子デバイスによって実装することができる。例えば、本開示のRPUデバイスは、抵抗変化型ランダムアクセスメモリ(RRAM)、相変化メモリ(PCM)、プログラム可能金属化セル(PMC)メモリ、非線形メモリスタ・システム、又はその他の非線形抵抗型スイッチング特性を有する任意の2端子デバイスで実装することができる。

20

【 0 0 5 1 】

本開示のRPUのより詳細な説明に先立って、本開示に関連する、シストリックアレイ、確率的計算並びに線形及び非線形メモリスタ・デバイスの概要をここで提供する。シストリックアレイは、特定の大きいに用いられているアルゴリズムの学習を加速することを企図した並列処理要素(PE)で構成される。シストリックアレイは、しばしば「積和(multiply and accumulate)」等の特定の演算用にハードワイヤードされており、積分、たたみこみ、相関、行列乗算又はデータ格納のタスクを大規模に並列に行う。非特許文献2において、シストリックアレイをオンライン学習ニューラルネットワーク用の構成ブロックとして使用することが提唱されており、ここでシストリックアレイ内の各PEは、単一の重み値を格納するための局所ストレージを有し、行列乗算及び重み更新のために必要な計算を行うことが可能である。非特許文献2に記載のPEの超大規模集積(VLSI)実装は、PE当たりおよそ1800個のトランジスタを要し、これは電力消費を増大させ、スケーラビリティを低減する。したがって、PE当りに必要とされるトランジスタができるだけ少ないPEを提供することが望ましい。

30

【 0 0 5 2 】

確率的計算は、ランダムビットのストリームによって連続的な値を表す技術の集まりであり、複雑な計算をストリーム上の単純なビット単位の演算で行うことができる。詳細には、確率的数(stochastic number)と呼ばれる2つのランダムかつ独立したビットストリーム S_1 、 S_2 が存在するとして(すなわちベルヌーイ過程)、ここで第1のストリームにおける「1」の確率を p 、第2のストリームにおける「1」の確率を q とすると、2つのストリームの論理ANDを図5に示すように得ることができる。出力ストリームにおける「1」の確率は、 pq である。十分な出力ビットを観測し、「1」の頻度を測定することによって、 pq を任意の精度まで推定することが可能である。少ない論理ゲート/トランジスタで実装することができるこれらのいわゆる「積和」演算の設計の単純さゆえに、確率的計算は、しばしばニューラルネットワーク用のハードウェア設計において使用される。非特許文献3は、ニューラルネットワーク訓練アルゴリズムのためのハードウェアアクセラレータとして使用することができる2次元(2D)シストリッ

40

50

クアレイに対する確率的計算の適用を示す。

【 0 0 5 3 】

しかしながら、非特許文献 3 においては、計算に必要な重みは、外部位置からシストリックアレイに供給され、重みに対する更新はアレイによって行われるわけではない。非特許文献 3 は、ニューラルネットワーク訓練中に大量に使用されるベクトル - 行列乗算又は行列 - 行列乗算演算の加速のみを扱うものである。しかしながら、局所ストレージを伴わないシストリックアレイでは、重みは外部メモリ位置に格納されているので、重み更新を並列に行うことはできない。全体としての学習アルゴリズムを加速するためには、非特許文献 3 には開示されていない重み更新の加速が必要である。

【 0 0 5 4 】

「メモリスト」いう用語は、デバイスの抵抗値がそのデバイスに以前に印加された電圧の履歴に依存する受動的 2 端子電気コンポーネントを記述するために使用される。メモリストの動作は、図 6 に示す式 [1] 及び式 [2] によって支配され、式中、 i はデバイスを通る電流、 v はデバイスに印加された電圧、 g はデバイスのコンダクタンス値（抵抗の逆数）、 s はコンダクタンス値を制御するデバイスの内部状態変数、 f は内部状態変数 s の時間発展を示す関数である。非特許文献 4 において、メモリスト機能が、抵抗変化型ランダムアクセスメモリ（RRAM）、相変化メモリ（PCM）及び導電性ブリッジランダムアクセスメモリ（CBRAM）等の抵抗型メモリデバイスの動作に対して提案されている。メモリスト・デバイスは、その履歴を記憶する（すなわち、いわゆる「不揮発性」）ので、非特許文献 4 は、このようなデバイスを不揮発性メモリ技術の可能な代替として提案している。

【 0 0 5 5 】

非特許文献 5 は、バックプロパゲーション型ニューラルネットワーク訓練ハードウェアに対するメモリストの使用を提案している。しかしながら、非特許文献 5 は、デバイスに印加された電圧に対して抵抗の変化が線形である、理想的なメモリスト動作を仮定している。非特許文献 5 の設計は、図 6 の式 [2] における関数 $f(s, v)$ が関係 $f(s, v) = v$ で与えられる単順な関数であると仮定したものである。非特許文献 5 は、上述のような 2 D シストリックアレイに類似したアーキテクチャを提案しており、各クロスポイントは、理想的メモリスト及び一対のトランジスタで実装されている。非特許文献 5 において、メモリストは、事実上、重み値を格納するのに使用され、一対のトランジスタは、重み更新のために必要な局所的乗算演算を計算するのに使用され、重み更新の結果がメモリストの伝導状態を修正する。非特許文献 5 は、事実上、メモリスト及び 2 つのトランジスタで構成された 4 端子デバイスを開示し、これはニューラルネットワーク・ハードウェアのバックプロパゲーション型訓練を実装するために 4 端子デバイスの 2 D アレイを作製するのに使用される。

【 0 0 5 6 】

ここで本開示のより詳細な説明に目を向けると、1 つ又は複数の実施形態は、2 端子 RPU に向けられ、これは、トランジスタ並びにオフチップ・ストレージ及び / 又は処理コンポーネント等の余分な回路要素を必要とすることなく、局所的データストレージ機能及び局所的データ処理機能を提供する。本開示の RPU は、RRAM、PCM、CBRAM、非線形メモリスト・システムなどを含むがこれらに限定されない、非線形スイッチング特性を有する任意のデバイスとして実装することができる。

【 0 0 5 7 】

本開示の RPU デバイスをメモリスト・システムとして実装するとき、メモリストは、非線形スイッチング特性を有する非理想的なものであることが重要である。理想的メモリストにおいては、抵抗の変化は、理想的メモリスト・デバイスに印加された電圧に対して線形に比例する。従って、理想的メモリスト・デバイスは、なんらかの電圧を経験した途端に、その抵抗状態が変化する。しかしながら、非理想的な非線形メモリスト・システムとして実装された本開示の RPU の場合、明確な「セット」電圧 V_{SET} （これはデバイスの特性である）が存在し、メモリスト RPU デバイスは、その内部抵抗状態を変化させ

10

20

30

40

50

るためには電圧 $V > V_{SET}$ を経験することを必要とする。このようなデバイスの場合、電圧 $V = 0.5 V_{SET}$ のハーフバイアス印加は、内部抵抗状態になんら変化を生じさせない。

【0058】

理想的メモリストと、本開示の RPU を実装するのに使用することができる非理想的な非線形メモリストとの間の差異をさらに例証するために、図 7 は、理想的メモリストと本開示による RPU との間での電圧スイッチング挙動比較を示すグラフである。縦軸は特定の電圧におけるデバイス状態変化を表し、横軸は印加された電圧を表す。理想的メモリスト動作においては、抵抗の変化は、デバイスに印加された電圧に線形に比例する。それゆえ、メモリストが電圧に遭遇した途端に、その抵抗状態が変化する。このことは曲線 702 によって示され、これは、低電圧であっても状態の変化が劇的であることを示す。

10

【0059】

曲線 704 によって示されるような非線形 RPU デバイスの場合、デバイスがその内部抵抗状態を変化させるために経験する必要がある明確なセット電圧 V_{SET} が存在する。 $0.5 V_{SET}$ のバイアス電圧は、内部抵抗状態を変化させない。本開示においては、この RPU デバイスの非線形特性を活用して、局所的に乗算を行う。 f_1 が小さいと仮定すると、デバイスは、 $0.5 V_{SET}$ が印加されただけではその内部状態を変化させない。特に、図 7 は、正電圧と、RPU デバイスの抵抗状態に対する正の変化とを例証するが、負電圧と負の抵抗変化との間にも同様の関係が存在する。

【0060】

20

図 8 は、本開示による、順方向行列乗算、逆方向行列乗算及び重み更新を行う 2 次元 (2D) クロスバーアレイ 800 の図である。クロスバーアレイ 800 は、伝導性行ワイヤ 802、804、806 の組と、伝導性行ワイヤ 802、804、806 の組と交差する伝導性列ワイヤ 808、810、812、814 の組とから形成される。行ワイヤの組と列ワイヤの組との間の交差部は、RPU によって分離され、それらは図 8 において、各自がそれぞれ $_{11}$ 、 $_{21}$ 、 $_{31}$ 、 $_{41}$ 、 $_{12}$ 、 $_{22}$ 、 $_{32}$ 、 $_{42}$ 、 $_{13}$ 、 $_{23}$ 、 $_{33}$ 及び $_{43}$ で表されるそれ自体の調整可能 / 更新可能な抵抗重みを有する抵抗要素として示される。図示を簡単にするために、図 8 においては 1 つの RPU 820 にのみ符号を付した。順方向行列乗算において、RPU の伝導状態 (すなわち格納された重み) は、RPU の両端間に電圧を印加し、RPU を通る電流を測定することによって読み出すことができる。

30

【0061】

入力電圧 V_1 、 V_2 、 V_3 がそれぞれ行ワイヤ 802、804、806 に印加される。各列ワイヤ 808、810、812、814 は、特定の列ワイヤに沿って各 RPU によって生成された電流 I_1 、 I_2 、 I_3 、 I_4 を合計する。例えば、図 8 に示すように、列ワイヤ 814 によって生成された電流 I_4 は、式 $I_4 = V_1 \cdot \text{}_{41} + V_2 \cdot \text{}_{42} + V_3 \cdot \text{}_{43}$ に従う。それゆえ、アレイ 800 は、RPU に格納された値に、電圧 V_1 、 V_2 、 V_3 で定められる行ワイヤ入力を乗ずることによって、順方向行列乗算を計算する。逆方向行列乗算は、全く同様である。逆方向行列乗算においては、電圧は列ワイヤ 808、810、812、814 に印加され、行ワイヤ 802、804、806 から読み出される。より詳細に後述する重み更新の場合、電圧は列ワイヤ及び行ワイヤに同時に印加され、該当する RPU デバイスに格納されたコンダクタンス値がすべて並列に更新される。従って、重み更新を行うために必要な積和演算は、RPU デバイス自体に加えて、アレイ 800 の該当する行又は列ワイヤを使用して、アレイ 800 の各 RPU 820 において局所的に行われる。それゆえ、本開示によれば、アレイ 800 において読出し - 更新 - 書込みサイクル (図 4 に示す) は必要とされない。

40

【0062】

引き続き図 8 の図を用いて、1 つ又は複数の実施形態により、ここで、正の重み更新方法論の動作を、RPU 820 及び伝導性行ワイヤ 806 と伝導性列ワイヤ 814 との交差部におけるその対応する重み $_{43}$ について提供する。RPU 820 の非線形特性を用い

50

て、後述のように確率的計算を使用してRPU820において局所的に乗算演算を行う。より詳細には、本開示の方法論は、RPU820の非線形スイッチング特性及び確率的ビットストリーム830、832を使用して、他の回路要素を必要とすることなく、乗算演算及び必要な重み更新をRPU820において局所的に行う。更新発生器回路（図示せず）がクロスバーアレイ800の周辺部に設けられ、2Dクロスバーアレイ800のすべてのRPUに並列に印加される確率的ビットストリーム（例えば830、832）形態の必要な電圧パルスを発生するための、周辺「翻訳機」として使用される。

【0063】

図7の図を一時的に参照すると、図8に示すRPU820に対する f_1 は、非常に小さい（例えば $f_1 = 0$ ）と仮定され、これは $0.5 V_{SET}$ が印加されただけではRPU820がその内部状態を変化させないことを意味する。行電圧シーケンス又はビットストリーム830は、行ワイヤ806に印加され、ゼロ電圧又は $+0.5 V_{SET}$ 電圧を有する、重み更新を表す電圧パルスのシーケンスとして示される。列電圧シーケンス又はビットストリーム832は、列ワイヤ814に印加され、ゼロ電圧又は $-0.5 V_{SET}$ 電圧のいずれかを有する、同じく重み更新を表す電圧パルスのシーケンスとして示される。図8の例においては、4/8が行電圧シーケンス830によってコード化され、6/8が列電圧シーケンス832によってコード化されている。例示的な電圧シーケンス830、832は、重み更新の正の抵抗変化相を表す。正の重み更新が行われた後、それぞれの電圧の極性を反転させた別個のシーケンスの組を使用して、負の方向に、そのような訂正が必要な重みについて重みを更新することができる。

【0064】

電圧シーケンス834は、行電圧シーケンス830と列電圧シーケンス832との間の差分から得られる、RPU820に印加される電圧である。電圧シーケンス834は、 $0V$ 、 $0.5 V_{SET}$ 及び V_{SET} の3つの電圧段階を有する。しかしながら、RPU820の抵抗 R_{43} はデバイス電圧が V_{SET} に達したときにのみ変化するので、列ワイヤ又は行ワイヤのどちらかを通じて送られる単一のパルスでは、RPU820の抵抗状態を変化させるのに十分ではない。列ワイヤが $0.5 V_{SET}$ の電圧を送り、行ワイヤが $-0.5 V_{SET}$ の電圧を送るとき、結果として生じる、該当のRPUに印加される V_{SET} パルスが、デバイスの抵抗のインクリメンタルな変化を引き起こす。従って、RPU820に印加される電圧パルスは、（図5に示すような）ビット単位の確率的AND演算をRPU820において局所的に行うために、RPU820の非線形スイッチング特性を利用することになる。それゆえ、結果として得られる、RPUの格納された重み（例えば R_{43} ）の変化は、クロスバーアレイ800の周辺の更新発生器回路によって「翻訳」された2つの数の積（ $4/8 * 6/8 = 3/8$ ）に比例する。

【0065】

図9及び図10は、例示的な公知の浮動小数点（FP）重み更新規則の分類誤り率の性能と、本開示の確率的RPU（SRPU）更新規則との比較を示す。FP重み更新規則は、乗算演算と等価であるベクトル-ベクトル外積を計算することを必要とし、インクリメンタルな重み更新を図9に示すように各クロスポイントにおいて局所的に行わなければならない。FP重み更新規則は、 $w_{ij} = w_{ij} + \eta x_i y_j$ として表すことができ、ここで w_{ij} は、 i 番目の行及び j 番目の列に対する重み値を表し、 x_i は、入力ニューロンにおける活性であり、 y_j は、出力ニューロンによって計算された誤りであり、 η は大域的学習率である。

【0066】

図9に示すように、FP重み更新規則は、クロスバーアレイ（図示せず）の行ワイヤ904と列ワイヤ906との間のクロスポイントに位置するFPクロスポイントデバイス902を使用する。FPクロスポイントデバイス902は、2つより多い端子を有する処理コンポーネント（図示せず）を含み、処理コンポーネントは、それぞれ行ワイヤ904及び列ワイヤ906に印加される信号 x_i 及び y_j を介して、更新データをFP数として受け取る。FPクロスポイントデバイス902は、乗算演算908及び加算演算910を使

用して重み更新 W_{ij} を計算し、次いで重み更新 W_{ij} を格納する。FP重み更新規則は、精度をもたらすが、読出し・書込み・処理の更新動作（例えば図4に示す）、又は、2つより多くの端子を有する比較的複雑で電力を消費する局所処理コンポーネントのいずれかを必要とする。例えば、FPクロスポイントデバイス902の処理コンポーネントは、非特許文献5に記載されているように一对のトランジスタとタンデムで動作する2端子の理想的メモリスタとして実装することができる。メモリスタは、事実上、重み値を格納するために使用され、一对のトランジスタは、重み更新のために必要な局所的乗算演算を計算するのに使用され、重み更新の結果がメモリスタの伝導状態を修正する。

【0067】

本開示のSRPU重み更新規則の実施形態は、図9の式3で示され、式中、 BL は更新サイクル中に使用されるビットストリームの長さであり、 w_{min} は単一の一致事象に起因する重み値の有効変化であり、 A_i^n 及び B_j^n は、ベルヌーイ過程によって特徴づけられるランダム変数であり、それらの成功率はそれぞれ x_i 及び j によって特徴づけられ、上付き文字 n はパルス指数を表す。図9に示すSRPU重み更新規則は、単一のアレイ内のすべてのRPUデバイスが並列に動作すること、及び、局所的乗算演算を単一致事象の統計に対して応答することによって行うことを可能にする。しかしながら、更新の確率的(stochastic)性質ゆえに、結果は常に確率的(probabilistic)であり、乗算演算の分散対平均比は、 BL の平方根の逆数に比例する。

【0068】

図9に示すように、SRPU重み更新規則は、RPUクロスポイントデバイス820Aを使用し、これは図8に示しかつ上で詳述したRPU820と実質的に同じように動作する。RPU820Aは、クロスバーアレイ(図示せず)の行ワイヤ910と列ワイヤ912との間のクロスポイントに位置する。RPU820Aは、2つの端子を有する処理コンポーネント(図8で820において示す)を含み、処理コンポーネントは、入力データ x_i 及び j を受け取り、それぞれ行ワイヤ910及び列ワイヤ912に確率的電圧シーケンスを印加する、確率的翻訳機914、916によって生成された確率的信号を介して、更新データを確率的数として受け取る。RPU820Aは、 W_{ij} の新たな値を、確率的ビットストリーム、RPU820Aの非線形特性、AND演算918及び加算演算920を使用して計算する。より詳細には、RPU820Aは、一致事象毎に、重み変化 w_{min} と等価なインクリメンタルなコンダクタンス変化を生じさせ、 w_{min} を格納された重み値に加算して、更新された重み値 W_{ij} に到達する。

【0069】

図10のグラフは、バックプロパゲーションアルゴリズムを使用した深層ニューラルネットワークの訓練問題を上述のSRPU重み更新規則として定義した確率的更新規則を用いて解析し、その結果を上述のFP重み更新規則を使用したベースラインと比較することによって、SRPU重み更新規則の分類誤り率とFP重み更新の分類誤り率との間の比較の結果を示す。MNISTデータセットに対して数字分類(digit classification)タスクを行う。深層ニューラルネットワークは、入力から出力まで各層がそれぞれ784、256、128及び10ニューロンを含む、完全に接続された層によって構成した。データの事前処理は行わず、未処理の画素値を入力として与えた。シグモイド・ニューロンを隠れ層において使用し、ソフトマックス・ニューロンを出力において使用した。クロスエントロピー目的関数を使用した。ベースラインFP重み更新については、浮動小数点算術を使用して、バッチサイズ1(unit)、並びに、0~10、11~20及び21~30のエポックについてそれぞれ学習率0.01、0.005及び0.0025で訓練を行った。ベースラインは、図10に示すグラフ上の白丸で示すように、30エポックの終了時にテストデータに対して約2.0%の分類精度に達した。

【0070】

学習率は、ニューラルネットワークの許容可能な収束特性を達成するためにある程度まで調整しなければならない重要なハイパーパラメータである。したがって、本開示のSRPU重み更新については、FP重み更新ベースラインに対して使用した値と平均して同一

10

20

30

40

50

となる学習率を使用する。図10に示すシミュレーションは、 x_i 及び j がスケールリングなしで確率的ストリームに直接翻訳されること、及び、 $P(A_i^n = \text{成功}) = x_i$ 及び $P(B_j^n = \text{成功}) = j$ であることを仮定した。したがって、重み値の平均変化は、 $E(w_{ij}) = BL \times w_{min} \times x_i \times j$ と記述することができる。第1のシミュレーションの組において、FP重み更新ベースラインに対して使用した学習率と一致させるために、BLを固定し、 $w_{min} = 1/BL$ とした。異なるBLを使用して訓練を行い、SRPU重み更新に対するアルゴリズムの許容差を観測した。図10のグラフに示すように、SRPU重み更新規則はBLに対して非常にロバストであり、実際、BL 10について、SRPU重み更新規則の精度は、ベースラインのFP重み更新規則と区別できなくなる。それゆえ、図10で実証されるように、SRPU重み更新は、単にBL 10サイクルを使用することによって、FP重み更新と同様に行うことができ、アレイサイズに依存せずに並列に動作するRPUデバイスによって実装することができる。

【0071】

図11(A)～図11(D)は、1つ又は複数の実施形態による2端子非線形RPUを使用する本開示の確率的重み更新について、RPUの線形性(r)をどのように計算するかを例証する略図及びグラフを示す。図11(A)は、前述の確率的電圧シーケンスを示し、図11(B)は、確率的電圧シーケンスを受け取って応答するRPU820を示す。図11(A)は、 x_i 及び j に印加された電圧シーケンスに応答したRPUの伝導値の変化である、 g も示す。より詳細には、 g は、 V_{DD} に達した電圧パルスに対するRPU820の応答である。従って、デバイス線形率 r は、図11(C)に示す式に従って決定することができる。図11(C)は、図7と同様のスイッチング図であり、これは g 対 V_{DD} をプロットし、理想的メモリストの応答を非理想的RPU及び理想的RPUと比較したものである。図11(D)は、異なる非線形率について、分類誤りを示す。図11(C)及び図11(D)は、関連したパラメータを変化させることによって異なる非線形性レベルでRPU820を設計し及び動作させることができること、及び、RPU820はたとえ非理想的であってもなお、許容可能な分類誤り率で動作することができることを実証する。

【0072】

図11(A)は、正の更新サイクルパルスを示す。図12(A)及び図12(B)は、パルスの極性を変更することによって、正及び負の更新サイクルの両方を実装することができることを示す。

【0073】

図13は、1つ又は複数の実施形態による2端子非線形RPUを使用した確率的重み更新の実装についてのさらなる詳細を例証する略図及びグラフを示す。図13は、図12(A)及び図12(B)に示す正及び負のパルシング・スキームによって、平均で、どの程度の重み変化 w が得られるかを示す。従って、 x_i 及び j をそれぞれのパルス符号に基づいて正の更新サイクル及び負の更新サイクルの両方に代入する。これは、RPU上に重なりがあるときにRPU上の電圧が g の符号に基づいて正方向又は負方向のいずれかで閾値を通ることが可能であることを保証する。パルスシーケンスの下に示す図は、パルスシーケンス内のパルス x 、 $+$ 及び $-$ がどのように相互作用するかの単なる図式的表現である。半パルス及び全パルスからの寄与に基づいて w についての式によって示されるように期待値を合計することによって、最終的な期待更新規則 w の2つの形態を導出することができる。

【0074】

図14(A)及び図14(B)は、本開示のRPUの整流的(rectifying)非線形性(図11(C)にも示す)及び本開示のRPUの飽和的(saturating)非線形性を横に並べて比較して示す。飽和的非線形性については、 x をコード化してRPUへ入れる代わりに、飽和的非線形RPUが必要な乗算を行うために $1-x$ をコード化することが必要である。

【0075】

10

20

30

40

50

図 1 5 及び図 1 6 は、指数関数的非線形性を有する 2 端子 R P U と共に高さ変調パルシング・スキームを使用したときの期待重み更新規則 w を例証するグラフ、図及び式を示す。この例においては、確率的コード化を電圧高さ変調によって置き換える。図 1 6 に示すように、テイラー展開を用いて、期待重み更新規則を単純化して $w = x$ にすることができる。従って、図 1 5 及び図 1 6 に示す導出は、図 1 5 及び図 1 6 に示す高次項が問題にならないこと、及び、指数関数的高さ変調 R P U は、図 1 6 の右下に示すグラフによって示されるように前述の実施形態と同じくらい良く機能することを実証する。

【 0 0 7 6 】

図 1 7 ~ 図 1 9 は、本開示による 2 端子非線形 R P U のクロスパーアレイを含む A N N アーキテクチャの開発、訓練及び使用の態様を示す。図 1 7 は、ニューラルネットワークを設計するための開始点を示す。事実上、図 1 7 は、図 3 に示すニューラルネットワーク図の代替的表現である。図 1 7 に示すように、 x_1 、 x_2 及び x_3 である入力ニューロンは、シグマ () で示される隠れニューロンに接続される。重みは、接続の強さを表し、入力ニューロン / ノードと隠れニューロン / ノードとの間、並びに、隠れニューロン / ノードと出力ニューロン / ノードとの間の接続に適用される。重みは、行列の形態である。データがネットワークを通過して順方向に移動するとき、ベクトル行列乗算が行われ、隠れニューロン / ノードは、入力を取得し、非線形変換を行い、次いで結果を次の重み行列へ送る。このプロセスは、データが出力ニューロン / ノードに到達するまで続く。出力ニューロン / ノードは、分類誤りを評価し、次いでこの分類誤りを順方向経路と同様の方式で逆方向に伝播し、その結果、ベクトル行列乗算が逆方向に行われる。各データセットについて、順方向経路及び逆方向経路が完了すると、重み更新が行われる。基本的に、各重みは、入力ニューロン / ノードによって定められるその重みに対する入力、及び、それが接続されたニューロン / ノードによって計算された誤りに比例して、更新される。

【 0 0 7 7 】

図 1 8 は、どのようにしてネットワークを中継器によって接続される構成ブロックに分割することができるかを示し、それゆえ、ニューラルネットワークを任意のサイズに拡張することが可能になる。ネットワークは本開示の R P U を使用するもので、情報の流れは局所的であり、そのことにより、構成ブロックを反復することが容易になり、また、スケーラビリティを制限するボトルネックとなり得るルータの必要性が排除される。

【 0 0 7 8 】

図 1 9 は、R P U 8 2 0 (図 8 に示す) から形成されるネットワーク構成ブロックを示す。R P U 8 2 0 には、図 7、図 1 1 (C)、図 1 4 (A)、図 1 4 (B) 又は図 1 5 に示す任意の非線形プロファイルを与えることができる。

【 0 0 7 9 】

ここで図 2 0 を参照して、ノード / ニューロン制御システム 1 9 0 0 を示す。ニューロン制御システム 1 9 0 0 は、ハードウェアプロセッサ 1 9 0 2 及びメモリ 1 9 0 4 を含む。A N N 用の訓練データ 1 9 0 6 は、メモリ 1 9 0 4 内に格納され、A N N の重みを訓練するために使用される。ニューロンインタフェース 1 9 0 8 は、A N N 上のニューロンを制御し、ニューロンがフィードフォワードモード、バックプロパゲーションモード、又は重み更新モードのいずれであるかを決定する。ニューロンインタフェース 1 9 0 8 は、さらに、入力ニューロンに入力を与え、出力ニューロンから出力を受け取る。誤り計算モジュール 1 9 1 0 は、ニューロンからの出力を訓練データ 1 9 0 6 と比較して、誤り信号を決定する。ニューロンインタフェース 1 9 0 8 は、誤り信号をバックプロパゲーションモード中に出力ニューロンに印加し、次いで重み更新モードをトリガして、A N N の重みをそれに応じて訓練する。

【 0 0 8 0 】

図 2 1 は、1 つ又は複数の実施形態による方法論 2 0 0 0 を例証するフロー図を示す。方法論 2 0 0 0 は、ブロック 2 0 0 2 において、伝導性並列ワイヤの第 1 の組を形成することによって開始する。ブロック 2 0 0 4 は、第 1 の組と交差する伝導性並列ワイヤの第 2 の組を形成して、伝導性並列ワイヤの第 1 の組と伝導性並列ワイヤの第 2 の組との間の

交差部に複数のクロスポイントを形成する。ブロック 2006 は、伝導性並列ワイヤの第 1 の組と伝導性並列ワイヤの第 2 の組との間の複数のクロスポイントの各々において 2 端子 RPU を形成する。ブロック 2008 は、各 RPU に、データを局所的に格納するための可変の伝導特性を与え、ブロック 2010 は、各 RPU に、データを局所的に処理するための非線形スイッチング特性を与える。

【0081】

図 22 は、1 つ又は複数の実施形態による方法論 2100 を例証する別のフロー図を示す。方法論 2100 は、ブロック 2102 において、複数の行、複数の列、及びクロスバーアレイの各クロスポイントにおける複数の 2 端子非線形 RPU を有する、クロスバーアレイを形成することによって開始する。ブロック 2104 は、クロスバーアレイの行及び列を介して、確率的数の独立したビットストリームを RPU に並列に適用する。ブロック 2106 は、各 RPU の非線形スイッチング特性及びそれに関連付けられたビットストリーム（行及び列）を利用して、その RPU において局所的に計算を行う。

【0082】

従って、本開示は、多くの技術的利益を提供する。単一の 2 端子デバイスは、コード化された入力ビットストリームとともに、そのデバイスの非線形スイッチング特性を使用して、重み更新を局所的に行い、これは、消費電力が低く、かつ、非線形 RPU のクロスバーアレイから形成された ANN における機械学習を加速する。本開示の RPU は、順方向経路の行列乗算、逆方向経路の行列乗算及び重み更新を行うために必要な処理及びストレージステップを局所的に行うことができる。この RPU は、オンライン・ニューラルネットワーク訓練、行列反転及び行列分解を含む様々な訓練方法論を加速することができる。

【0083】

1 つ又は複数の実施形態の一部は、デバイス、システム、方法及び / 又はコンピュータプログラム製品とすることができる。1 つ又は複数の実施形態によるコンピュータプログラム製品は、1 つ又は複数の実施形態の態様をプロセッサに実行させるためのプログラム命令を有する 1 つ又は複数のコンピュータ可読ストレージ媒体を含む。

【0084】

コンピュータ可読ストレージ媒体は、命令実行デバイスによる使用のために命令を保持及び格納することができる有形デバイスとすることができる。コンピュータ可読ストレージ媒体は、例えば、電子ストレージデバイス、磁気ストレージデバイス、光ストレージデバイス、電磁気ストレージデバイス、半導体ストレージデバイス、又は上記のものの任意の適切な組合せとすることができるがこれらに限定されない。コンピュータ可読ストレージ媒体のより具体的な例の非網羅的なリストは、ポータブル・コンピュータ・ディスクレット、ハードディスク、ランダムアクセスメモリ (RAM)、読み出し専用メモリ (ROM)、消去可能プログラム可能読み出し専用メモリ (EPROM 又はフラッシュメモリ)、静的ランダムアクセスメモリ (SRAM)、ポータブル・コンパクトディスク読み出し専用メモリ (CD-ROM)、デジタル多目的ディスク (DVD)、メモリスティック、フロッピーディスク、記録された命令を有するパンチカード若しくは溝内に隆起した構造等の機械式コード化デバイス、及び上記のものの任意の適切な組合せを含む。コンピュータ可読ストレージ媒体は、本明細書で用いられる場合、無線波若しくは他の自由に伝搬する電磁波、導波路若しくは他の伝送媒体を通して伝搬する電磁波（例えば光ファイバケーブルを通る光パルス）、又は電線を通して伝送される電気信号のような一時的な信号自体と解釈されるべきではない。

【0085】

本明細書で説明されるコンピュータ可読プログラム命令は、コンピュータ可読ストレージ媒体からそれぞれのコンピューティング / 処理デバイスにダウンロードすることもでき、又は、例えばインターネット、ローカル・エリア・ネットワーク、広域ネットワーク及び / 若しくは無線ネットワークを経由して、外部コンピュータ若しくは外部ストレージデバイスにダウンロードすることもできる。ネットワークは、銅伝送ケーブル、光伝送ファイバ、無線伝送、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータ及

び／又はエッジサーバを含むことができる。各コンピューティング／処理デバイス内のネットワーク・アダプタ・カード又はネットワーク・インタフェースは、ネットワークからコンピュータ可読プログラム命令を受け取り、そのコンピュータ可読プログラム命令をそれぞれのコンピューティング／処理デバイス内のコンピュータ可読ストレージ媒体におけるストレージのために転送する。

【0086】

本開示の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セット・アーキテクチャ（ISA）命令、機械語命令、機械依存命令、マイクロコード、ファームウェア命令、状態設定データ、又は、Smalltalk、若しくはC++などのオブジェクト指向プログラミング言語及び「C」プログラミング言語若しくは類似のプログラミング言語のような従来の手続き型プログラミング言語を含む1つ若しくは複数のプログラミング言語の任意の組合せで記述されたソースコード若しくはオブジェクトコードのいずれかとすることができる。コンピュータ可読プログラム命令は、完全にユーザのコンピュータ上で実行される場合もあり、一部がユーザのコンピュータ上で独立型ソフトウェア・パッケージとして実行される場合もあり、一部がユーザのコンピュータ上で実行され、一部が遠隔コンピュータ上で実行される場合もあり、又は完全に遠隔コンピュータ若しくはサーバ上で実行される場合もある。後者のシナリオにおいては、遠隔コンピュータは、ローカル・エリア・ネットワーク（LAN）若しくは広域ネットワーク（WAN）を含むいずれかのタイプのネットワークを通じてユーザのコンピュータに接続される場合もあり、又は外部コンピュータへの接続が行われる場合もある（例えば、インターネット・サービス・プロバイダを用いたインターネットを通じて）。幾つかの実施形態において、例えばプログラム可能論理回路、フィールドプログラム可能ゲートアレイ（FPGA）、又はプログラム可能論理アレイ（PLA）を含む電子回路は、本開示の態様を実施するために、コンピュータ可読プログラム命令の状態情報を利用して電子回路を個別化することにより、コンピュータ可読プログラム命令を実行することができる。

【0087】

本開示の態様は、本明細書において、本開示の実施形態による方法、装置（システム）、及びコンピュータプログラム製品のフローチャート図及び／又はブロック図を参照して説明される。フローチャート図及び／又はブロック図の各ブロック、並びにフローチャート図及び／又はブロック図のブロックの組合せは、コンピュータ可読プログラム命令によって実装することができることが理解されるであろう。

【0088】

これらのコンピュータ可読プログラム命令を、汎用コンピュータ、専用コンピュータ、又は他のプログラム可能データ処理装置のプロセッサに与えてマシンを製造し、それにより、コンピュータ又は他のプログラム可能データ処理装置のプロセッサによって実行される命令が、フローチャート及び／又はブロック図の1つ又は複数のブロック内で指定された機能／動作を実装するための手段を作り出すようにすることができる。これらのコンピュータプログラム命令を、コンピュータ、プログラム可能データ処理装置、及び／又は他のデバイスを特定の方式で機能させるように指示することができるコンピュータ可読ストレージ媒体内に格納し、それにより、その中に格納された命令を有するコンピュータ可読媒体が、フローチャート及び／又はブロック図の1つ又は複数のブロックにおいて指定された機能／動作の態様を実装する命令を含む製品を含むようにすることもできる。

【0089】

コンピュータ可読プログラム命令を、コンピュータ、他のプログラム可能データ処理装置又は他のデバイス上にロードして、一連の動作ステップをコンピュータ、他のプログラム可能データ処理装置又は他のデバイス上で行わせてコンピュータ実装のプロセスを生成し、それにより、コンピュータ、他のプログラム可能装置又は他のデバイス上で実行される命令が、フローチャート及び／又はブロック図の1つ又は複数のブロックにおいて指定された機能／動作を実装するようにすることもできる。

【0090】

図面内のフローチャート及びブロック図は、本開示の種々の実施形態による、システム、方法、及びコンピュータプログラム製品の可能な実装の、アーキテクチャ、機能及び動作を示す。この点に関して、フローチャート又はブロック図内の各ブロックは、指定された論理機能を実装するための1つ又は複数の実行可能命令を含む、モジュール、セグメント、又は命令の一部を表すことができる。幾つかの代替的な実装において、ブロック内に記された機能は、図中に記された順序とは異なる順序で行われることがある。例えば、連続して示された2つのブロックは、関与する機能に応じて、実際には実質的に同時に実行されることもあり、又はこれらのブロックはときとして逆順で実行されることもある。ブロック図及び/又はフローチャート図の各ブロック、及びブロック図及び/又はフローチャート図中のブロックの組合せは、指定された機能又は動作を実行する専用ハードウェア・ベースのシステムによって実装することもでき、又は専用ハードウェアとコンピュータ命令との組合せを実行することもできることに留意されたい。

10

【0091】

本明細書で用いられる用語は、特定の実施形態を説明のみを目的としたものであり、本開示を限定することを意図したものではない。本明細書で使用される単数形「a」、「an」及び「the」は、前後関係から明らかに別の意味を示さない限り、複数形態も含むことを意図する。さらに、本明細書内で使用する場合に、「備える、含む」及び/又は「備えている、含んでいる」という用語は、そこに述べた特徴、整数、ステップ、動作、要素及び/又はコンポーネントの存在を明示しているが、1つ又は複数の他の特徴、整数、ステップ、動作、要素、コンポーネント及び/又はそれらの群の存在又は付加を排除するものではないことが理解されるであろう。

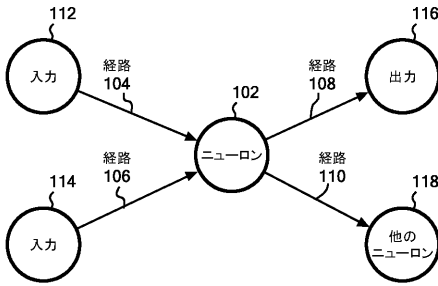
20

【0092】

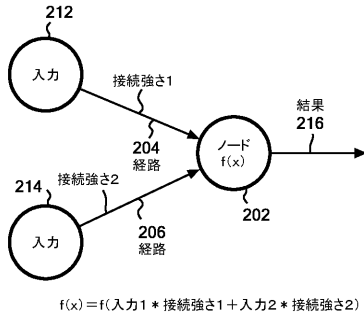
以下の特許請求の範囲における全ての「手段又はステップと機能との組み合わせ（ミーンズ又はステップ・プラス・ファンクション）」要素の対応する構造、材料、動作、及び均等物は、その機能を、明確に特許請求された他の請求要素との組み合わせで実行するためのあらゆる構造、材料、又は動作を含むことが意図されている。本開示は、例証及び説明を目的として提示されたものであるが、網羅的であること又は開示された形態に限定することを意図したものではない。本開示の範囲及び思想から逸脱しない多くの修正及び変形が当業者には明らかであろう。実施形態は、本開示の原理及び実際の用途を最も良く説明するように、また企図される特定の用途に適するような種々の修正を伴う種々の実施形態に関して本開示を当業者が理解するのを可能にするように、選択しかつ説明したものである。

30

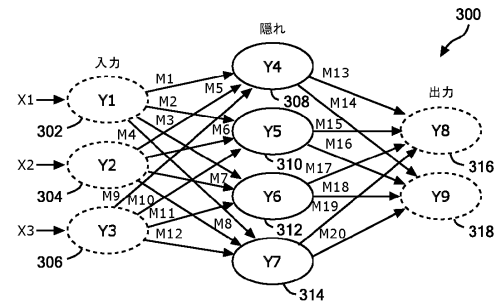
【図 1】



【図 2】

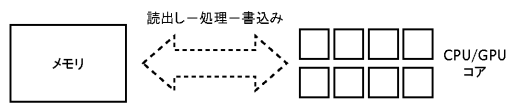


【図 3】

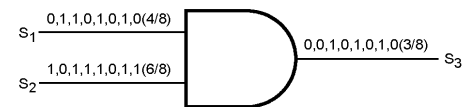


$$\begin{aligned} Y1 &= F(X1) & Y4 &= F(M1*Y1 + M5*Y2 + M9*Y3) & Y8 &= F(M13*Y4 + M15*Y5 + M17*Y6 + M19*Y7) \\ Y2 &= F(X2) & Y5 &= F(M2*Y1 + M6*Y2 + M10*Y3) & Y9 &= F(M14*Y4 + M16*Y5 + M18*Y6 + M20*Y7) \\ Y3 &= F(X3) & Y6 &= F(M3*Y1 + M7*Y2 + M11*Y3) & & \\ & & Y7 &= F(M4*Y1 + M8*Y2 + M12*Y3) & & \end{aligned}$$

【図 4】



【図 5】

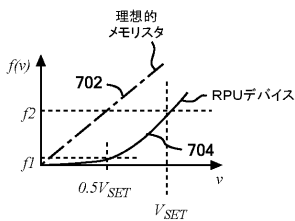


【図 6】

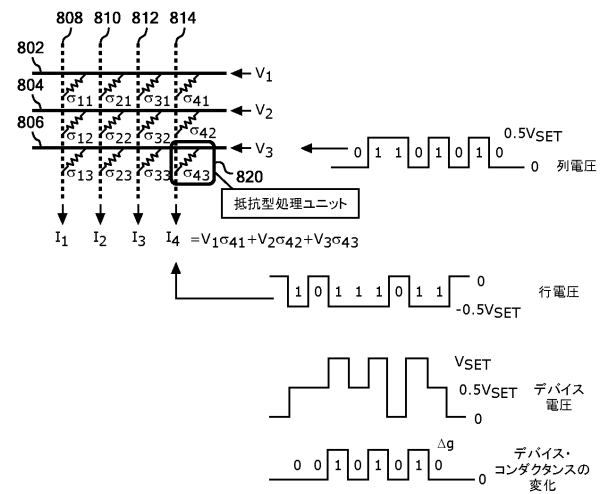
$$i = g(s, v)v \quad \text{式[1]}$$

$$\frac{\partial s(t)}{\partial t} = f(s, v) \quad \text{式[2]}$$

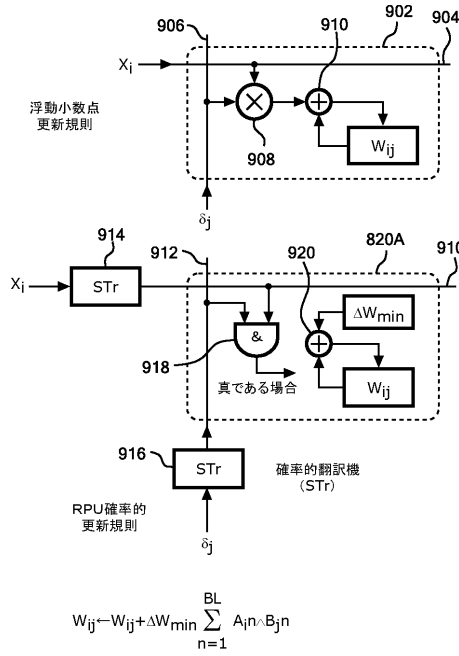
【図 7】



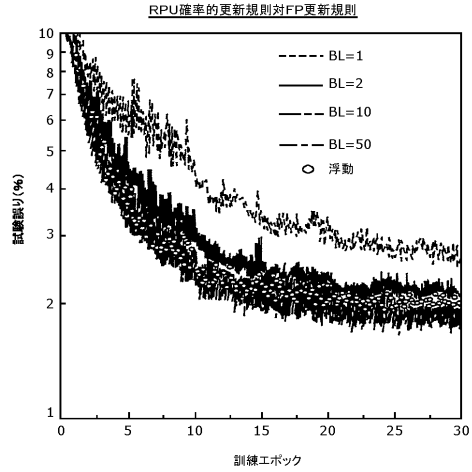
【図 8】



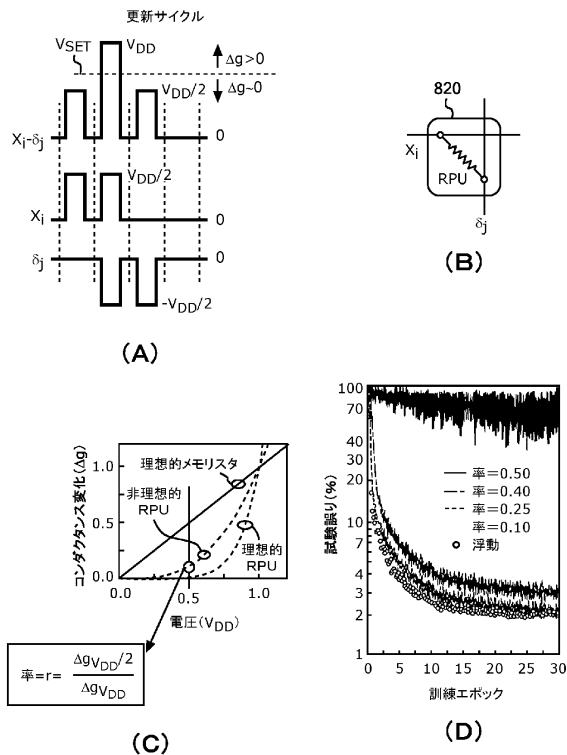
【図 9】



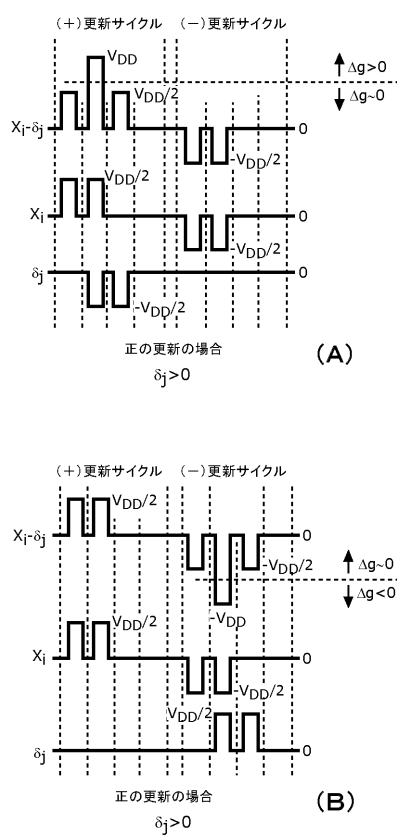
【図 10】



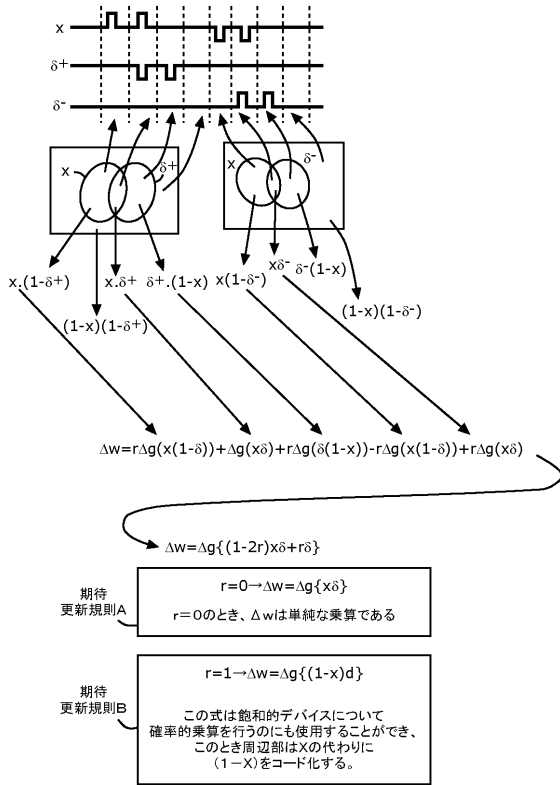
【図 11】



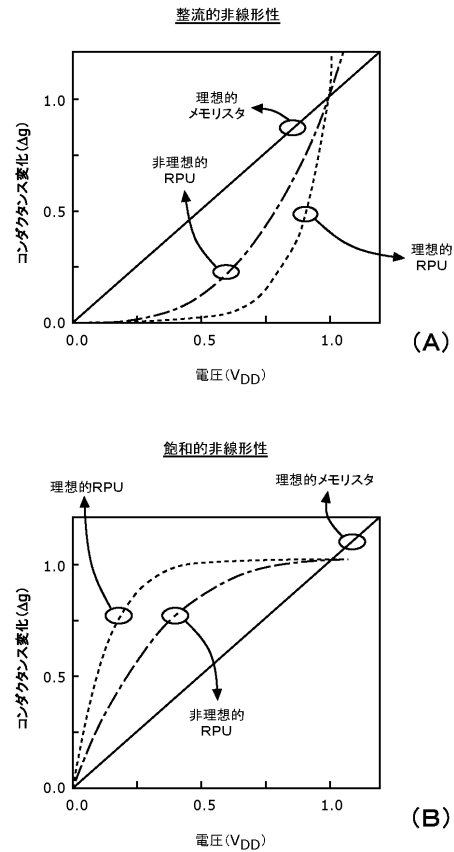
【図 12】



【図 13】

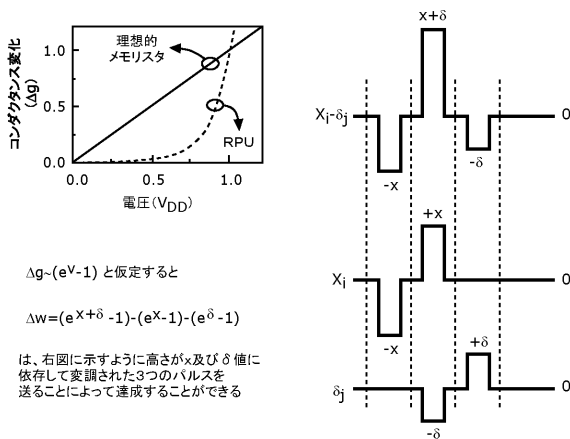


【図 14】



【図 15】

局所的乗算のための電圧高さ変調



【図 16】

局所的乗算のための電圧高さ変調・続き

テイラー展開の最初の2項は以下の結果を与える

指数関数:

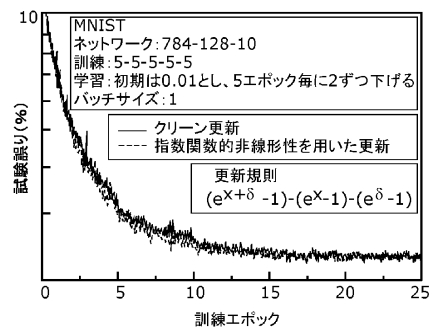
$$\Delta w = (e^{x+\delta} - 1) - (e^x - 1) - (e^\delta - 1) \text{ すべての } x \text{ について}$$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

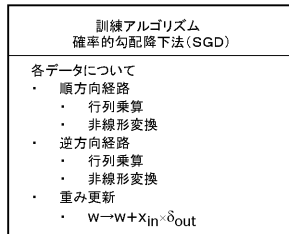
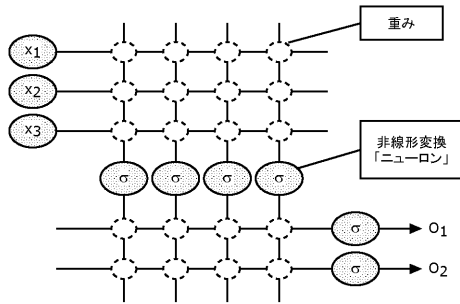
$$\Delta w = (1 + (x+\delta) + \frac{(x+\delta)^2}{2} - 1) - (1 + x + \frac{x^2}{2} - 1) - (1 + \delta + \frac{\delta^2}{2} - 1)$$

$$\Delta w = x\delta$$

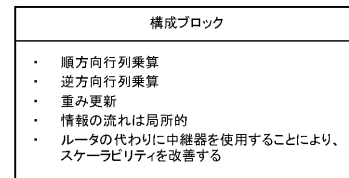
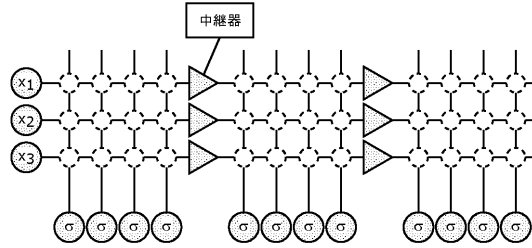
これは、高次項が問題にならないこと、及び、本開示の更新方法論が、右図に示すように完全な乗算の場合と同じくらい良く機能することを実証する



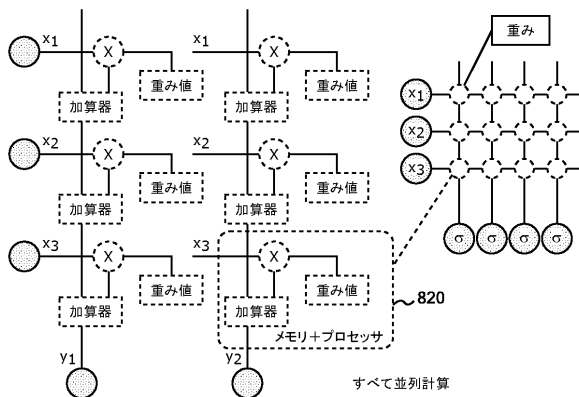
【図 17】



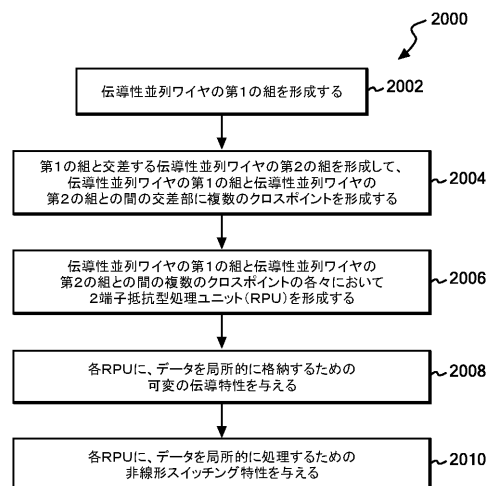
【図 18】



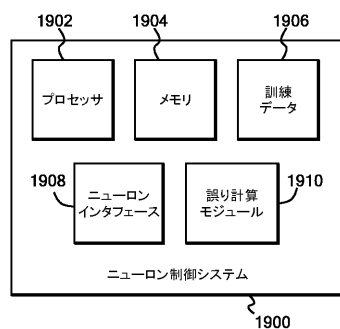
【図 19】



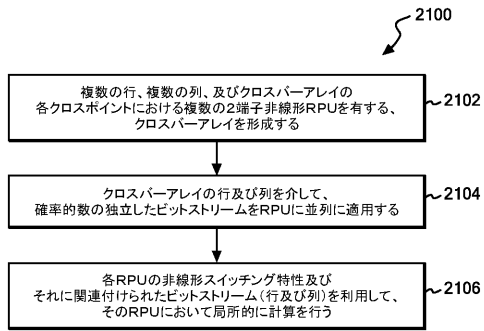
【図 21】



【図 20】



【図 22】



フロントページの続き

(74)代理人 100112690

弁理士 太佐 種一

(72)発明者 ゴクメン、タイフン

アメリカ合衆国 1 0 5 9 8 ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード
1 1 0 1

(72)発明者 ウラソフ、ユーリ

アメリカ合衆国 1 0 5 9 8 ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード
1 1 0 1

(72)発明者 キム、セヨン

アメリカ合衆国 1 0 5 9 8 ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード
1 1 0 1

審査官 三坂 敏夫

(56)参考文献 特開平 0 7 - 3 1 9 5 0 7 (J P , A)

特開平 0 5 - 2 8 2 2 7 2 (J P , A)

特開平 0 7 - 1 2 1 4 9 8 (J P , A)

米国特許出願公開第 2 0 1 5 / 0 1 7 0 0 2 5 (U S , A 1)

特開平 0 5 - 1 8 1 9 9 3 (J P , A)

特表 2 0 1 3 - 5 4 6 0 6 4 (J P , A)

ALIBART, Fabien et al., Pattern classification by memristive crossbar circuits using ex
situ and in situ training, nature COMMUNICATIONS, Macmillan Publishers Limited., 2 0
1 3 年 6 月 2 5 日, pages:1-7, [令和2年5月11日検索], インターネット<URL : <https://www.nature.com/articles/ncomms3072.pdf>>

(58)調査した分野(Int.Cl., D B 名)

G 0 6 N 3 / 0 0 - 3 / 1 2

1 0 / 0 0 - 2 0 / 2 0

G 1 1 C 1 3 / 0 0