

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第6130056号
(P6130056)

(45) 発行日 平成29年5月17日(2017.5.17)

(24) 登録日 平成29年4月21日(2017.4.21)

(51) Int.Cl.

F 1

G06N	3/04	(2006.01)
G06N	3/063	(2006.01)
G06N	3/08	(2006.01)

GO 6 N	3/04	GO 6 N	3/04	1 9 O
GO 6 N	3/063	GO 6 N	3/063	
GO 6 N	3/08	GO 6 N	3/08	

請求項の数 44 (全 33 頁)

(21) 出願番号	特願2016-515029 (P2016-515029)
(86) (22) 出願日	平成26年5月20日 (2014.5.20)
(65) 公表番号	特表2016-522495 (P2016-522495A)
(43) 公表日	平成28年7月28日 (2016.7.28)
(86) 國際出願番号	PCT/US2014/038841
(87) 國際公開番号	W02014/189970
(87) 國際公開日	平成26年11月27日 (2014.11.27)
審査請求日	平成29年2月3日 (2017.2.3)
(31) 優先権主張番号	61/825,657
(32) 優先日	平成25年5月21日 (2013.5.21)
(33) 優先権主張国	米国(US)
(31) 優先権主張番号	61/862,714
(32) 優先日	平成25年8月6日 (2013.8.6)
(33) 優先権主張国	米国(US)

(73) 特許権者	595020643 クアアルコム・インコーポレイテッド QUALCOMM INCORPORATED アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドライブ 5775
(74) 代理人	100108855 弁理士 蔵田 昌俊
(74) 代理人	100109830 弁理士 福原 淑弘
(74) 代理人	100158805 弁理士 井関 守三
(74) 代理人	100194814 弁理士 奥村 元宏

最終頁に続く

(54) 【発明の名称】スパイキングネットワークの効率的なハードウェア実装

(57) 【特許請求の範囲】

【請求項 1】

複数の人工ニューロンを伴う人工神経システムを動作させるための方法であって、前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、

前記スーパーニューロン処理ユニットをメモリとインターフェースすること、

前記メモリに記憶された前記人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスすること、ここにおいて、前記メモリの構成が連続メモリアクセスを可能にする、と、

を備える方法。

【請求項 2】

複数のスーパーニューロン処理ユニットを同時に動作させることは、入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化することを備える、請求項 1 に記載の方法。

【請求項 3】

前記スーパーニューロン処理ユニットのうちの 1 つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新することをさらに備える、請求項 1 に記載の方法。

【請求項 4】

10

20

前記人工ニューロンのいずれかに関連するシナプス後ファンアウトへのポインタを前記メモリに記憶することをさらに備える、請求項 1 に記載の方法。

【請求項 5】

複数のスーパーニューロン処理ユニットを同時に動作させることは、

前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施することを備える、請求項 1 に記載の方法。

【請求項 6】

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、請求項 1 に記載の方法。 10

【請求項 7】

前記メモリは、ダイナミックランダムアクセスメモリ (D R A M) を備え、

前記 D R A M の前記連続メモリアクセスは、複数のシナプスに関係するデータへの同時アクセスを備える、請求項 1 に記載の方法。

【請求項 8】

メモリの前記構成は、可塑性を実現するためのシナプス前ニューロンおよびシナプス後ニューロンのルックアップテーブルを備える、請求項 1 に記載の方法。

【請求項 9】

動的確率的方法での前記人工神経システムのスパイクの除去
をさらに備える、請求項 1 に記載の方法。 20

【請求項 10】

前記人工神経システムのシナプスに関連するデータを前記メモリに記憶することをさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、請求項 1 に記載の方法。

【請求項 11】

前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、前記人工神経システムに外部および内部シナプス活性化を統合することをさらに備える、請求項 1 に記載の方法。

【請求項 12】

複数の人工ニューロンを伴う人工神経システムを動作させるための装置であって、 30

前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させるように構成された第 1 の回路、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当たられる、と、

前記スーパーニューロン処理ユニットをメモリとインターフェースするように構成された第 2 の回路と、

前記メモリに記憶された前記人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするための第 3 の回路、ここにおいて、前記メモリの構成が連続メモリアクセスを可能にする、と、

を備える装置。 40

【請求項 13】

前記第 1 の回路はまた、

入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化するように構成される、請求項 1 2 に記載の装置。

【請求項 14】

前記スーパーニューロン処理ユニットのうちの 1 つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新するように構成された第 4 の回路をさらに備える、請求項 1 2 に記載の装置。

【請求項 15】

前記人工ニューロンのいずれかに関連するシナプス後ファンアウトへのポインタを前記 50

メモリに記憶するように構成された第4の回路をさらに備える、請求項12に記載の装置。

【請求項16】

前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施するように構成された第4の回路をさらに備える、請求項12に記載の装置。

【請求項17】

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、請求項12に記載の装置。

【請求項18】

10

前記メモリは、ダイナミックランダムアクセスメモリ(DRAM)を備え、

前記DRAMの前記連続メモリアクセスは、複数のシナップスに関するデータへの同時アクセスを備える、請求項12に記載の装置。

【請求項19】

メモリの前記構成は、可塑性を実現するためのシナップス前ニューロンおよびシナップス後ニューロンのルックアップテーブルを備える、請求項12に記載の装置。

【請求項20】

動的確率的方法で前記人工神経システムのスパイクを除去するように構成された第4の回路をさらに備える、請求項12に記載の装置。

【請求項21】

20

前記人工神経システムのシナップスに関するデータを前記メモリに記憶するように構成された第4の回路をさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、請求項12に記載の装置。

【請求項22】

前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、前記人工神経システムに外部および内部シナップス活性化を統合するように構成された第4の回路をさらに備える、請求項12に記載の装置。

【請求項23】

複数の人工ニューロンを伴う人工神経システムを動作させるための装置であって、

前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させるための手段、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、

前記スーパーニューロン処理ユニットをメモリとインターフェースするための手段と、

前記メモリに記憶された前記人工神経システムのシナップス重みおよび可塑性のパラメータにアクセスするための手段、ここにおいて、前記メモリの構成が連続メモリアクセスを可能にする、と、

を備える装置。

【請求項24】

複数のスーパーニューロン処理ユニットを同時に動作させるための前記手段は、

入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化するための手段をさらに備える、請求項23に記載の装置。

40

【請求項25】

前記スーパーニューロン処理ユニットのうちの1つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新するための手段をさらに備える、請求項23に記載の装置。

【請求項26】

前記人工ニューロンのいずれかに関連するシナップス後ファンアウトへのポインタを前記メモリに記憶するための手段をさらに備える、請求項23に記載の装置。

【請求項27】

50

複数のスーパーニューロン処理ユニットを同時に動作させるための前記手段は、前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施するための手段をさらに備える、請求項23に記載の装置。

【請求項28】

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、請求項23に記載の装置。

【請求項29】

前記メモリは、ダイナミックランダムアクセスメモリ(DRAM)を備え、前記DRAMの前記連続メモリアクセスは、複数のシナプスに関するデータへの同時アクセスを備える、請求項23に記載の装置。 10

【請求項30】

メモリの前記構成は、可塑性を実現するためのシナプス前ニューロンおよびシナプス後ニューロンのルックアップテーブルを備える、請求項23に記載の装置。

【請求項31】

動的確率的方法での前記人工神経システムのスパイクの除去のための手段をさらに備える、請求項23に記載の装置。

【請求項32】

前記人工神経システムのシナプスに関するデータを前記メモリに記憶するための手段をさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、請求項23に記載の装置。 20

【請求項33】

前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、前記人工神経システムに外部および内部シナプス活性化を統合するための手段をさらに備える、請求項23に記載の装置。

【請求項34】

複数の人工ニューロンを伴う人工神経システムを動作させるための非一時的コンピュータ可読媒体であって、前記非一時的コンピュータ可読媒体が、

前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、 30

前記スーパーニューロン処理ユニットをメモリとインターフェースすることと、

前記メモリに記憶された前記人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスすること、ここにおいて、前記メモリの構成が連続メモリアクセスを可能にする、と、

を行うように実行可能な命令を有する、非一時的コンピュータ可読媒体。

【請求項35】

入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化するためのコードをさらに備える、請求項34に記載の非一時的コンピュータ可読媒体。 40

【請求項36】

前記スーパーニューロン処理ユニットのうちの1つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新するためのコードをさらに備える、請求項34に記載の非一時的コンピュータ可読媒体。

【請求項37】

前記人工ニューロンのいずれかに関連するシナプス後ファンアウトへのポインタを前記メモリに記憶するためのコードをさらに備える、請求項34に記載の非一時的コンピュータ可読媒体。

【請求項38】

前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータ 50

を実施するためのコードをさらに備える、請求項34に記載の非一時的コンピュータ可読媒体。

【請求項39】

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、請求項34に記載の非一時的コンピュータ可読媒体。

【請求項40】

前記メモリは、ダイナミックランダムアクセスメモリ(DRAM)を備え、

前記DRAMの前記連続メモリアクセスは、複数のシナプスに関するデータへの同時アクセスを備える、請求項34に記載の非一時的コンピュータ可読媒体。

10

【請求項41】

メモリの前記構成は、可塑性を実現するためのシナプス前ニューロンおよびシナプス後ニューロンのルックアップテーブルを備える、請求項34に記載の非一時的コンピュータ可読媒体。

【請求項42】

動的確率的方法での前記人工神経システムのスパイクの除去

のためのコードをさらに備える、請求項34に記載の非一時的コンピュータ可読媒体。

【請求項43】

前記人工神経システムのシナプスに関するデータを前記メモリに記憶するためのコードをさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、請求項34に記載の非一時的コンピュータ可読媒体。

20

【請求項44】

前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、前記人工神経システムに外部および内部シナプス活性化を統合するためのコードをさらに備える、請求項34に記載の非一時的コンピュータ可読媒体。

【発明の詳細な説明】

【技術分野】

【0001】

米国特許法第119条による優先権の主張

[0001]本出願は、2013年5月21日に出願された「S p i k e T i m e W i n d o w i n g f o r I m p l e m e n t i n g S p i k e - T i m i n g D e p e n d e n t P l a s t i c i t y (S T D P) 」と題する米国仮特許出願第61/825,657号、2013年8月6日に出願された「S p i k e T i m e W i n d o w i n g f o r I m p l e m e n t i n g S p i k e - T i m i n g D e p e n d e n t P l a s t i c i t y (S T D P) 」と題する米国仮特許出願第61/862,714号、2013年8月6日に出願された「C o m p u t e d S y n a p s e s f o r N e u r o m o r p h i c S y s t e m s 」と題する米国仮特許出願第61/862,741号、2013年8月6日に出願された「I m p l e m e n t i n g D e l a y s b e t w e e n N e u r o n s i n a n A r t i f i c i a l N e r v o u s S y s t e m 」と題する米国仮特許出願第61/862,734号、2014年1月28日に出願された「E f f i c i e n t H a r d w a r e I m p l e m e n t a t i o n o f S p i k i n g N e t w o r k s 」と題する米国仮特許出願第61/932,364号の利益を主張し、これらの各々は、参照により全体が本明細書に組み込まれる。

30

【0002】

[0002]本開示のいくつかの態様は、一般に人工神経システムに関し、より詳細には、スパイキングニューラルネットワークの効率的なハードウェア実装のための方法および装置に関する。

【背景技術】

【0003】

40

50

[0003]人工ニューロン（すなわち、ニューラル処理ユニット）の相互結合されたグループを備え得る人工ニューラルネットワークは、計算デバイスであるか、または計算デバイスによって実行される方法を表す。人工ニューラルネットワークは、生物学的ニューラルネットワークにおける対応する構造および／または機能を有し得る。しかしながら、人工ニューラルネットワークは、従来の計算技法が厄介、実行不可能または不適切であるいくつかの適用例に革新的で有用な計算技法を提供することができる。人工ニューラルネットワークが観測から機能を推論することができるので、そのようなネットワークは、タスクまたはデータの複雑さが従来の技法による機能の設計を面倒にする適用例において、特に有用である。

【0004】

10

[0004]1つのタイプの人工ニューラルネットワークはスパイキングニューラルネットワークであり、これは、その動作モデルならびにニューロンおよびシナプスの状態に時間の概念を組み込み、それによって、ニューラルネットワークにおける計算機能の発生元となり得る挙動の豊かなセットを提供する。スパイキングニューラルネットワークは、ニューロンがニューロンの状態に基づいて特定の（1つまたは複数の）時間に発火する、または「スパイクする」という概念、および時間がニューロン機能にとって重要であるという概念に基づく。ニューロンが発火するとき、そのニューロンは、他のニューロンに進むスパイクを生成し、他のニューロンは、今度は、このスパイクが受信された時間に基づいてそれらの状態を調整することができる。言い換えれば、ニューラルネットワークにおけるスパイクの相対的タイミングまたは絶対的タイミングで情報が符号化され得る。

20

【発明の概要】

【0005】

[0005]本開示のいくつかの態様は、複数の人工ニューロンを伴う人工神経システム動作のための方法を提供する。本方法は一般に、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てる、と、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにスーパーニューロン処理ユニットをメモリとインターフェースすること、ここにおいて、メモリの構成が連続メモリアクセスを可能にする、とを含む。

【0006】

30

[0006]本開示のいくつかの態様は、人工神経システムを動作させるための装置を提供する。本装置は一般に、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させるように構成された第1の回路、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てる、と、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにスーパーニューロン処理ユニットをメモリとインターフェースするように構成された第2の回路、ここにおいて、メモリの構成が連続メモリアクセスを可能にする、と、を含む。

【0007】

40

[0007]本開示のいくつかの態様は、人工神経システムを動作させるための装置を提供する。本装置は一般に、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させるための手段、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てる、と、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにスーパーニューロン処理ユニットをメモリとインターフェースするための手段、ここにおいて、メモリの構成が連続メモリアクセスを可能にする、と、を含む。

【0008】

[0008]本開示のいくつかの態様は、人工神経システムを動作させるためのコンピュータプログラム製品を提供する。本コンピュータプログラム製品は一般に、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニット

50

を同時に動作させること、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにスーパーニューロン処理ユニットをメモリとインターフェースすること、ここにおいて、メモリの構成が連続メモリアクセスを可能にする、と、を行うように実行可能な命令を有する非一時的コンピュータ可読媒体（たとえば、記憶デバイス）含む。

【0009】

[0009]本開示の上述の特徴が詳細に理解され得るように、添付の図面にその一部が示される態様を参照することによって、上記で簡単に要約された内容のより具体的な説明が得られ得る。ただし、その説明は他の等しく有効な態様に通じ得るので、添付の図面は、本開示のいくつかの典型的な態様のみを示し、したがって、本開示の範囲を限定するものと見なされるべきではないことに留意されたい。10

【図面の簡単な説明】

【0010】

【図1】[0010]本開示のいくつかの態様によるニューロンの例示的なネットワークを示す図。

【図2】[0011]本開示のいくつかの態様による、計算ネットワーク（ニューラルシステムまたはニューラルネットワーク）の例示的な処理ユニット（ニューロン）を示す図。

【図3】[0012]本開示のいくつかの態様による例示的なスパイクタイミング依存可塑性（STDP）曲線を示す図。20

【図4】[0013]本開示のいくつかの態様による、ニューロンの挙動を定義するための正レジームと負レジームとを示す、人工ニューロンに関する状態の例示的なグラフ。

【図5】[0014]本開示のいくつかの態様による、人工神経システムの例示的なハードウェア実装形態を示す図。

【図6】[0015]本開示のいくつかの態様による、人工神経システムを動作させるための例示的な動作の流れ図。

【図6A】[0016]図6に示す動作を実行することが可能な例示的な手段を示す図。

【図7】[0017]本開示のいくつかの態様による、汎用プロセッサを使用して人工神経システムを動作させるための例示的な実装形態を示す図。

【図8】[0018]本開示のいくつかの態様による、メモリが個々の分散処理ユニットとインターフェースされ得る人工神経システムを動作させるための例示的な実装形態を示す図。30

【図9】[0019]本開示のいくつかの態様による、分散メモリおよび分散処理ユニットに基づいて人工神経システムを動作させるための例示的な実装形態を示す図。

【図10】[0020]本開示のいくつかの態様による、ニューラルネットワークの例示的な実装形態を示す図。

【発明を実施するための形態】

【0011】

[0021]添付の図面を参照しながら本開示の様々な態様について以下でより十分に説明する。ただし、本開示は、多くの異なる形態で実施され得、本開示全体にわたって提示される任意の特定の構造または機能に限定されるものと解釈されるべきではない。むしろ、これらの態様は、本開示が周到で完全になり、本開示の範囲を当業者に十分に伝えるように与えられる。本明細書の教示に基づいて、本開示の範囲は、本開示の任意の他の態様とは無関係に実装されるにせよ、本開示の任意の他の態様と組み合わされるにせよ、本明細書で開示する本開示のいかなる態様をもカバーするものであることを、当業者なら諒解されたい。たとえば、本明細書に記載される態様をいくつ使用しても、装置は実装され得、または方法は実施され得る。さらに、本開示の範囲は、本明細書に記載される本開示の様々な態様に加えてまたはそれらの態様以外に、他の構造、機能、または構造および機能を使用して実施されるそのような装置または方法をカバーするものとする。本明細書で開示する本開示のいずれの態様も、請求項の1つまたは複数の要素によって実施され得ることを理解されたい。40

【0012】

[0022]「例示的」という単語は、本明細書では「例、事例、または例示の働きをすること」を意味するために使用される。「例示的」として本明細書で説明するいかなる態様も、必ずしも他の態様よりも好ましいまたは有利であると解釈されるべきであるとは限らない。

【0013】

[0023]本明細書では特定の態様について説明するが、これらの態様の多くの変形および置換は本開示の範囲内に入る。好ましい態様のいくつかの利益および利点が説明されるが、本開示の範囲は特定の利益、使用、または目的に限定されるものではない。むしろ、本開示の態様は、様々な技術、システム構成、ネットワーク、およびプロトコルに広く適用可能であるものとし、そのうちのいくつかを例として図および好ましい態様についての以下の説明で示す。発明を実施するための形態および図面は、本開示を限定するものではなく説明するものにすぎず、本開示の範囲は添付の特許請求の範囲およびそれの均等物によって定義される。10

【0014】

例示的なニューラルシステム

[0024]図1は、本開示のいくつかの態様による、複数のレベルのニューロンをもつ例示的なニューラルシステム100を示す。ニューラルシステム100は、シナプス結合のネットワーク104（すなわち、フィードフォワード結合）を介してニューロンの別のレベル106に結合されたニューロンのあるレベル102を備え得る。簡単のために、図1には2つのレベルのニューロンのみが示されているが、典型的なニューラルシステムには、より少ないまたはより多くのレベルのニューロンが存在し得る。ニューロンのいくつかは、ラテラル結合を介して同じ層の他のニューロンに結合し得ることに留意されたい。さらに、ニューロンのいくつかは、フィードバック結合を介して前の層のニューロンに戻る形で結合し得る。20

【0015】

[0025]図1に示すように、レベル102における各ニューロンは、前のレベル（図1に図示せず）の複数のニューロンによって生成され得る入力信号108を受信し得る。信号108は、レベル102のニューロンへの入力（たとえば、入力電流）を表し得る。そのような入力は、膜電位を充電するためにニューロン膜上に蓄積され得る。膜電位がそのしきい値に達すると、ニューロンは、発火し、ニューロンの次のレベル（たとえば、レベル106）に転送されるべき出力スパイクを生成し得る。そのような挙動は、アナログおよびデジタル実装形態を含むハードウェアおよび/またはソフトウェアでエミュレートまたはシミュレートされ得る。30

【0016】

[0026]生物学的ニューロンでは、ニューロンが発火するときに生成される出力スパイクは、活動電位と呼ばれる。電気信号は、約100mVの振幅と約1msの持続時間とを有する比較的急速で、一時的で、一か零かの神経インパルスである。一連の結合されたニューロンを有するニューラルシステムの特定の態様（たとえば、図1におけるあるレベルのニューロンから別のレベルのニューロンへのスパイクの転送）では、あらゆる活動電位が基本的に同じ振幅と持続時間とを有するので、信号における情報は、振幅によってではなく、スパイクの周波数および数（またはスパイクの時間）によってのみ表される。活動電位によって搬送される情報は、スパイク、スパイクしたニューロン、および1つまたは複数の他のスパイクに対するスパイクの時間によって決定される。40

【0017】

[0027]図1に示すように、あるレベルのニューロンから別のレベルのニューロンへのスパイクの転送は、シナプス結合（または単に「シナプス」）のネットワーク104によって達成され得る。シナプス104は、レベル102のニューロン（シナプス104に対するシナプス前ニューロン（pre-synaptic neurons））から出力信号（すなわち、スパイク）を受信し得る。いくつかの態様では、これらの信号は、調整可能なシナプス重み50

【数1】

$$w_1^{(i,i+1)}, \dots, w_p^{(i,i+1)}$$

(ここでPは、レベル102のニューロンとレベル106のニューロンとの間のシナプス結合の総数である)に従ってスケーリングされ得る。他の態様では、シナプス104は、いかなるシナプス重みをも適用しないことがある。さらに、(スケーリングされた)信号は、レベル106における各ニューロン(シナプス104に対するシナプス後ニューロン(post-synaptic neurons))の入力信号として合成され得る。レベル106におけるあらゆるニューロンは、対応する合成された入力信号に基づいて、出力スパイク110を生成し得る。出力スパイク110は、次いで、シナプス結合の別のネットワーク(図1には図示せず)を使用して、別のレベルのニューロンに転送され得る。

【0018】

[0028]生物学的シナプスは、電気シナプスまたは化学シナプスのいずれに分類され得る。電気シナプスは、興奮性信号を送るために主に使用される一方、化学シナプスは、シナプス後ニューロンにおける興奮性活動または抑制性(過分極化)活動のいずれかを調停することができ、ニューロン信号を增幅する役目を果たすこともできる。興奮性信号は通常、膜電位を脱分極する(すなわち、静止電位に対して膜電位を増加させる)。しきい値を超えて膜電位を脱分極するために十分な興奮性信号が一定期間内に受信された場合、シナプス後ニューロンに活動電位が生じる。対照的に、抑制性信号は一般に、膜電位を過分極する(すなわち、低下させる)。抑制性信号は、十分に強い場合、興奮性信号のすべてを相殺し、膜電位がしきい値に達するのを防止することができる。シナプス興奮を相殺することに加えて、シナプス抑制は、自然に活発なニューロンに対して強力な制御を行うことができる。自然に活発なニューロンは、たとえば、そのダイナミクスまたはフィードバックに起因するさらなる入力なしにスパイクするニューロンを指す。これらのニューロンにおける活動電位の自然な生成を抑圧することによって、シナプス抑制は、ニューロンの発火のパターンを形成することができ、それは一般にスカルプチャリング(sculpturing)と呼ばれる。様々なシナプス104は、望まれる拳動に応じて、興奮性シナプスまたは抑制性シナプスの任意の組合せとして働き得る。

【0019】

[0029]ニューラルシステム100は、汎用プロセッサ、デジタル信号プロセッサ(DSP)、特定用途向け集積回路(ASIC)、フィールドプログラマブルゲートアレイ(FPGA)もしくは他のプログラマブル論理デバイス(PLD)、個別ゲートもしくはトランジスタ論理、個別ハードウェア構成要素、プロセッサによって実行されるソフトウェアモジュール、またはそれらの任意の組合せによってエミュレートされ得る。ニューラルシステム100は、たとえば画像およびパターン認識、機械学習、モータ制御など、かなりの適用範囲において利用され得る。ニューラルシステム100における各ニューロンは、ニューロン回路として実装され得る。出力スパイクを開始するしきい値まで充電されるニューロン膜は、たとえば、そこを通って流れる電流を積分するキャパシタとして実装され得る。

【0020】

[0030]一態様では、ニューロン回路の電流積分デバイスとしてのキャパシタは除去され得、その代わりにより小さいメモリストラ(memristor)要素が使用され得る。この手法は、ニューロン回路において、かさばるキャパシタが電流積分器として利用される様々な他の適用例と同様に適用され得る。さらに、シナプス104の各々は、メモリストラ要素に基づいて実装され得、シナプス重みの変化は、メモリストラ抵抗の変化に関係し得る。ナノメートルの特徴サイズのメモリストラを用いると、ニューロン回路およびシナプスの面積が大幅に低減され得、それによって、非常に大規模なニューラルシステムハードウェア実装形態の実装が実用的になり得る。

10

20

30

40

50

【0021】

[0031]ニューラルシステム 100 をエミュレートするニューラルプロセッサの機能は、シナプス結合の重みに依存し得、それはニューロン間の結合の強さを制御し得る。シナプス重みは、パワーダウン後にプロセッサの機能を維持するために、不揮発性メモリに記憶され得る。一様では、シナプス重みメモリは、主たるニューラルプロセッサチップとは別個の外部チップ上に実装され得る。シナプス重みメモリは、交換可能メモリカードとしてニューラルプロセッサチップとは別個にパッケージ化され得る。これは、ニューラルプロセッサに多様な機能を提供することができ、特定の機能は、ニューラルプロセッサに現在取り付けられているメモリカードに記憶されたシナプス重みに基づき得る。

【0022】

[0032]図 2 は、本開示のいくつかの態様による、計算ネットワーク（たとえば、ニューラルシステムまたはニューラルネットワーク）の処理ユニット（たとえば、人工ニューロン 202）の一例 200 を示す。たとえば、ニューロン 202 は、図 1 のレベル 102 のニューロンおよび 106 のニューロンのうちのいずれかに対応し得る。ニューロン 202 は、複数の入力信号 $204_1 \sim 204_N$ ($x_1 \sim x_N$) を受信し得、それらは、ニューラルシステムの外部にある信号、または同じニューラルシステムの他のニューロンによって生成された信号、またはその両方であり得る。入力信号は、電流または電圧、実数値または複素数値であり得る。入力信号は、固定小数点表現または浮動小数点表現をもつ数値を備え得る。これらの入力信号は、調整可能なシナプス重み $206_1 \sim 206_N$ ($w_1 \sim w_N$) に従って信号をスケーリングするシナプス結合を通してニューロン 202 に伝えられ得、N はニューロン 202 の入力接続の総数であり得る。

【0023】

[0033]ニューロン 202 は、スケーリングされた入力信号を合成し、合成された、スケーリングされた入力を使用して、出力信号 208（すなわち、信号 y ）を生成し得る。出力信号 208 は、電流または電圧、実数値または複素数値であり得る。出力信号は、固定小数点表現または浮動小数点表現をもつ数値を備え得る。出力信号 208 は、次いで、同じニューラルシステムの他のニューロンへの入力信号として、または同じニューロン 202 への入力信号として、またはニューラルシステムの出力として伝達され得る。

【0024】

[0034]処理ユニット（ニューロン 202）は電気回路によってエミュレートされ得、その入力接続および出力接続は、シナプス回路をもつ配線によってエミュレートされ得る。処理ユニット、その入力接続および出力接続はまた、ソフトウェアコードによってエミュレートされ得る。処理ユニットはまた、電気回路によってエミュレートされ得るが、その入力接続および出力接続はソフトウェアコードによってエミュレートされ得る。一様では、計算ネットワーク中の処理ユニットはアナログ電気回路を備え得る。別の態様では、処理ユニットはデジタル電気回路を備え得る。さらに別の態様では、処理ユニットは、アナログ構成要素とデジタル構成要素の両方をもつ混合信号電気回路を備え得る。計算ネットワークは、上述の形態のいずれかにおける処理ユニットを備え得る。そのような処理ユニットを使用した計算ネットワーク（ニューラルシステムまたはニューラルネットワーク）は、たとえば画像およびパターン認識、機械学習、モータ制御など、かなりの適用範囲において利用され得る。

【0025】

[0035]ニューラルネットワークをトレーニングする過程で、シナプス重み（たとえば、図 1 の重み

【数 2】

$$w_1^{(i,i+1)}, \dots, w_p^{(i,i+1)}$$

および / または図 2 の重み $206_1 \sim 206_N$ ）がランダム値により初期化され得、学習ルールに従って増加または減少し得る。学習ルールのいくつかの例として、スパイクタイミ

10

20

30

40

50

ング依存可塑性（S T D P）学習ルール、Hebb則、Oja則、Biernestock-Copper-Munro（BCM）則などがある。非常に多くの場合、重みは、2つの値のうちの1つに安定し得る（すなわち、重みの双峰分布）。この効果が利用されて、シナプス重み当たりのビット数を低減し、シナプス重みを記憶するメモリとの間の読み取りおよび書き込みの速度を上げ、シナプスメモリの電力消費量を低減し得る。

【0026】

シナプスタイプ

[0036] ニューラルネットワークのハードウェアおよびソフトウェアモデルでは、シナプス関係機能の処理がシナプスタイプに基づき得る。シナプスタイプは、非塑性シナプス（non-plastic synapse）（重みおよび遅延の変化がない）と、可塑性シナプス（重みが変化し得る）と、構造遅延可塑性シナプス（重みおよび遅延が変化し得る）と、完全可塑性シナプス（重み、遅延および結合性が変化し得る）と、その変形（たとえば、遅延は変化し得るが、重みまたは結合性の変化はない）とを備え得る。これの利点は、処理が再分割（subdivide）され得ることである。たとえば、非塑性シナプスは、可塑性機能を実行すること（またはそのような機能が完了するのを待つこと）を必要としない。同様に、遅延および重み可塑性は、一緒にまたは別々に、順にまたは並列に動作し得る動作に再分割され得る。異なるタイプのシナプスは、適用される異なる可塑性タイプの各々の異なるルックアップテーブルまたは式およびパラメータを有し得る。したがって、本方法は、シナプスのタイプの関係するテーブルにアクセスすることになる。

【0027】

[0037] また、スパイクタイミング依存構造可塑性がシナプス可塑性とは無関係に実行され得るという事実のさらなる含意がある。構造可塑性（すなわち、遅延量の変化）は前後スパイク時間差（pre-post spike time difference）の直接の関数であり得るので、構造可塑性は、重みの大きさに変化がない場合（たとえば、重みが最小値または最大値に達したか、あるいはそれが何らかの他の理由により変更されない場合）でも実行され得る。代替的に、それは、重み変化量に応じて、または重みもしくは重み変化の限界に関する条件に基づいて設定され得る。たとえば、重み変化が生じたとき、または重みが最大限に達するのではなく、重みがゼロに達した場合のみ、シナプス遅延が変化し得る。しかしながら、これらのプロセスが並列化され、メモリアクセスの数および重複を低減し得るように、独立した機能を有することが有利であり得る。

【0028】

シナプス可塑性の決定

[0038] 神経可塑性（または単に「可塑性」）は、脳内のニューロンおよびニューラルネットワークがそれらのシナプス結合と挙動とを新しい情報、感覚上の刺激、発展、損傷または機能不全に応答して変える能力である。可塑性は、生物学における学習および記憶にとって、また計算論的神経科学およびニューラルネットワークにとって重要である。（たとえば、Hebb則理論による）シナプス可塑性、スパイクタイミング依存可塑性（S T D P）、非シナプス可塑性、活性依存可塑性、構造可塑性および恒常的可塑性など、様々な形の可塑性が研究されている。

【0029】

[0039] S T D Pは、脳内の場合のようなニューロン間のシナプス結合の強さを調整する学習プロセスである。結合強度は、特定のニューロンの出力スパイクおよび受信入力スパイク（すなわち、活動電位）の相対的タイミングに基づいて調整される。S T D Pプロセスの下で、あるニューロンに対する入力スパイクが、平均して、そのニューロンの出力スパイクの直前に生じる傾向がある場合、長期増強（LTP）が生じ得る。その場合、その特定の入力はいくらか強くなる。対照的に、入力スパイクが、平均して、出力スパイクの直後に生じる傾向がある場合、長期抑圧（LTD）が生じ得る。その場合、その特定の入力はいくらか弱くなるので、「スパイクタイミング依存可塑性」と呼ばれる。したがって、シナプス後ニューロンの興奮の原因であり得る入力は、将来的に寄与する可能性がさらに高くなる一方、シナプス後スパイクの原因ではない入力は、将来的に寄与する可能性が

10

20

30

40

50

低くなる。結合の初期セットのサブセットが残る一方で、その他の部分の影響がゼロまたはゼロの近くまで低減されるまで、このプロセスは続く。

【0030】

[0040]ニューロンは、一般に、入力の多くが短い期間内に生じる（すなわち、出力をもたらすのに十分に累積している）ときに出力スパイクを生成するので、通常残っている入力のサブセットは、時間的に相関する傾向のあった入力を含む。さらに、出力スパイクの前に生じる入力は強化されるので、最も早い十分に累積的な相関の示度を与える入力は結局、ニューロンへの最終入力となる。

【0031】

[0041] S T D P 学習ルールは、シナプス前ニューロンのスパイク時間 t_{pre} とシナプス後ニューロンのスパイク時間 t_{post} との間の時間差（すなわち、 $t = t_{post} - t_{pre}$ ）の関数として、シナプス前ニューロンをシナプス後ニューロンに結合するシナプスのシナプス重みを効果的に適合させ得る。¹⁰ S T D P の通常の公式化は、時間差が正である（シナプス前ニューロンがシナプス後ニューロンの前に発火する）場合にシナプス重みを増加させ（すなわち、シナプスを増強し）、時間差が負である（シナプス後ニューロンがシナプス前ニューロンの前に発火する）場合にシナプス重みを減少させる（すなわち、シナプスを抑制する）ことである。

【0032】

[0042] S T D P プロセスでは、経時的なシナプス重みの変化は通常、指數関数的減衰を使用して達成され得、以下のように与えられる。²⁰

【数3】

$$\Delta w(t) = \begin{cases} a_+ e^{-t/k_+} + \mu, & t > 0 \\ a_- e^{t/k_-}, & t < 0 \end{cases}, \quad (1)$$

ここで、 k_+ および k_- はそれぞれ、正の時間差および負の時間差の時間定数であり、 a_+ および a_- は対応するスケーリングの大きさであり、 μ は正の時間差および / または負の時間差に適用され得るオフセットである。

【0033】

[0043] 図 3 は、S T D P による、シナプス前スパイクおよびシナプス後スパイクの相対的タイミングの関数としてのシナプス重み変化の例示的なグラフ 3 0 0 を示す。シナプス前ニューロンがシナプス後ニューロンの前に発火する場合、グラフ 3 0 0 の部分 3 0 2 に示すように、対応するシナプス重みは増加し得る。この重み増加は、シナプスの L T P と呼ばれる。グラフ部分 3 0 2 から、シナプス前スパイク時間とシナプス後スパイク時間との間の時間差に応じて、L T P の量がほぼ指數関数的に減少し得ることが観測され得る。グラフ 3 0 0 の部分 3 0 4 に示すように、発火の逆の順序は、シナプス重みを減少させ、シナプスの L T D をもたらし得る。³⁰

【0034】

[0044] 図 3 のグラフ 3 0 0 に示すように、S T D P グラフの L T P (原因) 部分 3 0 2 に負のオフセット μ が適用され得る。x 軸の交差 3 0 6 のポイント ($y = 0$) は、層 $i - 1$ (シナプス前層) からの原因入力の相関を考慮して、最大タイムラグと一致するように構成され得る。フレームベースの入力（すなわち、入力は、スパイクまたはパルスを備える特定の持続時間のフレーム (frame) の形をとる）の場合、オフセット値 μ は、フレームの境界を反映するように計算され得る。直接的にシナプス後電位によってモデル化されるように、またはニューラル状態に対する影響の点で、フレームにおける第 1 の入力スパイク (パルス) が経時に減衰することが考慮され得る。フレームにおける第 2 の入力スパイク (パルス) が特定の時間フレームの相関したまたは関連したものと考えられる場合、フレームの前および後の関連する時間は、その時間フレームの境界で分離され、関連する時間の値が異なり得る（たとえば、1 つのフレームよりも大きい場合は負、1 つのフレームよりも小さい場合は正）⁴⁰

ームよりも小さい場合は正) ように、STD P 曲線の 1 つまたは複数の部分をオフセットすることによって、可塑性の点で別様に扱われ得る。たとえば、曲線が、フレーム時間よりも大きい前後の時間で実際にゼロよりも下になり、結果的に LTP の代わりに LTD の一部であるように LTP をオフセットするために負のオフセット μ が設定され得る。

【0035】

ニューロンモデルおよび演算

[0045] 有用なスパイキングニューロンモデルを設計するための一般的原理がいくつかある。良いニューロンモデルは、2つの計算レジーム、すなわち、一致検出および関数計算の点で豊かな潜在的挙動を有し得る。その上、良いニューロンモデルは、時間コーディングを可能にするための2つの要素を有する必要がある。すなわち、入力の到着時間は出力時間に影響を与え、一致検出は狭い時間ウインドウを有し得る。最後に、計算上魅力的であるために、良いニューロンモデルは、連続時間において閉形式解を有し得、ニアアトラクター (near attractor) と鞍点とを含む安定した挙動を有し得る。言い換えれば、有用なニューロンモデルは、実用的なニューロンモデルであり、豊かで、現実的で、生物学的に一貫した挙動をモデル化するために使用され得、神経回路のエンジニアリングとリバースエンジニアリングの両方を行うために使用され得るニューロンモデルである。

【0036】

[0046] ニューロンモデルは事象、たとえば入力の到着、出力スパイク、または内部的であるか外部的であるかを問わず他の事象に依存し得る。豊かな挙動レパートリーを実現するために、複雑な挙動を示すことができる状態機械が望まれ得る。入力寄与 (ある場合) とは別個の事象の発生自体が状態機械に影響を与え、事象の後のダイナミクスを制限し得る場合、システムの将来の状態は、単なる状態および入力の関数ではなく、むしろ状態、事象および入力の関数である。

【0037】

[0047] 一態様では、ニューロン n は、下記のダイナミクスによって決定される膜電圧 $v_n(t)$ についてのスパイキングリーキー積分発火ニューロン (spiking leaky-integrate-and-fire neuron) としてモデル化され得る。

【数4】

$$\frac{dv_n(t)}{dt} = \alpha v_n(t) + \beta \sum_m w_{m,n} y_m(t - \Delta t_{m,n}), \quad (2)$$

ここで α および β はパラメータであり、 $w_{m,n}$ は、シナプス前ニューロン m をシナプス後ニューロン n に結合するシナプスのシナプス重みであり、 $y_m(t)$ は、ニューロン n の細胞体に到着するまで $t_{m,n}$ に従って樹状遅延または軸索遅延によって遅延し得るニューロン m のスパイキング出力である。

【0038】

[0048] シナプス後ニューロンへの十分な入力が達成された時間からシナプス後ニューロンが実際に発火する時間までの遅延があることに留意されたい。イジケヴィッチの単純モデル (Izhikevich's simple model) など、動的スパイキングニューロンモデルでは、脱分極しきい値 v_t とピークスパイク電圧 v_{peak} との間に差がある場合、時間遅延が生じ得る。たとえば、単純モデルでは、電圧および復元に関する 1 対の微分方程式、すなわち、

【数5】

$$\frac{dv}{dt} = (k(v - v_t)(v - v_r) - u + I)/C, \quad (3)$$

【数6】

$$\frac{du}{dt} = a(b(v - v_r) - u). \quad (4)$$

によってニューロン細胞体ダイナミクス(neuron soma dynamics)が決定され得る。ここで v は膜電位であり、 u は、膜復元変数であり、 k は、膜電位 v の時間スケールを記述するパラメータであり、 a は、復元変数 u の時間スケールを記述するパラメータであり、 b は、膜電位 v の閾値下(sub-threshold)変動に対する復元変数 u の感度を記述するパラメータであり、 v_r は、膜静止電位であり、 I は、シナプス電流であり、 C は、膜のキャパシタンスである。このモデルによれば、ニューロンは $v > v_{peak}$ のときにスパイクすると定義される。

【0039】

Hunzinger Cold モデル

[0049] Hunzinger Cold ニューロンモデルは、豊かな様々な神経挙動を再生し得るミニマル・デュアルレジーム・スパイキング線形動的モデル(minimal dual-regime spiking linear dynamical model)である。モデルの1次元または2次元の線形ダイナミクスは2つのレジームを有することができ、ここで、時間定数(および結合)はレジームに依存し得る。閾値下レジームでは、時間定数は、慣例により負であり、一般に生物学的に一貫した線形方式で静止状態に細胞を戻す役目を果たすリーキーチャネルダイナミクスを表す。閾値上(supra-threshold)レジームにおける時間定数は、慣例により正であり、一般にスパイク生成のレイテンシを生じさせる一方でスパイク状態に細胞を駆り立てる反リーキーチャネルダイナミクスを反映する。

【0040】

[0050] 図4に示すように、モデルのダイナミクスは2つの(またはそれよりも多くの)レジームに分割され得る。これらのレジームは、負レジーム402(リーキー積分発火(LIF)ニューロンモデルと混同されないように互換的にLIFレジームとも呼ばれる)および正レジーム404(反リーキー積分発火(ALIF)ニューロンモデルと混同されないように互換的にALIFレジームとも呼ばれる)と呼ばれ得る。負レジーム402では、状態は将来の事象の時点において静止(v_-)である傾向がある。この負レジームでは、モデルは一般に、時間的入力検出特性と他の閾値下の挙動とを示す。正レジーム404では、状態はスパイキング事象(v_s)である傾向がある。この正レジームでは、モデルは、後続の入力事象に応じてスパイクにレイテンシを生じさせるなどの計算特性を示す。事象の点からのダイナミクスの公式化およびこれら2つのレジームへのダイナミクスの分離は、モデルの基本的特性である。

【0041】

[0051] 線形デュアルレジーム2次元ダイナミクス(状態 v および u の場合)は、慣例により次のように定義され得る。

【数7】

$$\tau_p \frac{dv}{dt} = v + q_p \quad (5)$$

【数8】

$$-\tau_u \frac{du}{dt} = u + r \quad (6)$$

ここで q および r は、結合のための線形変換変数である。

【0042】

10

20

30

40

50

[0052]シンボル v は、ダイナミクスレジームを示すためにここで使用され、特定のレジームの関係を論述または表現するときに、それぞれ負レジームおよび正レジームについて符号「-」または「+」にシンボル v を置き換える慣例がある。

【0043】

[0053]モデル状態は、膜電位（電圧） v および復元電流 u によって定義される。基本形態では、レジームは基本的にモデル状態によって決定される。正確で一般的な定義の微妙だが重要な側面があるが、差し当たり、モデルが、電圧 v がしきい値 (v_+) を上回る場合に正レジーム 404 にあり、そうでない場合に負レジーム 402 にあると考える。

【0044】

[0054]レジーム依存時間定数は、負レジーム時間定数である τ_- と正レジーム時間定数である τ_+ を含む。復元電流時間定数 τ_u は通常、レジームから独立している。便宜上、 τ_u と同様に、指数および τ_+ が一般に正となる正レジームの場合に、電圧発展 (voltage evolution) に関する同じ表現が使用され得るように、減衰を反映するために負の量として負レジーム時間定数 τ_- が一般に指定される。

【0045】

[0055]2つの状態要素のダイナミクスは、事象において、ヌルクライン (null-cline) から状態をオフセットする変換によって結合され得、ここで変換変数は、

【数9】

$$q_\rho = -\tau_\rho \beta u - v_\rho \quad (7)$$

10

20

【数10】

$$r = \delta(v + \epsilon) \quad (8)$$

であり、 v_- 、 v_+ および v_- 、 v_+ はパラメータである。 v のための 2 つの値は、2つのレジームのための参照電圧のベースである。パラメータ v_- は、負レジームのためのベース電圧であり、膜電位は一般に、負レジームにおいて v_- に向かって減衰することになる。パラメータ v_+ は、正レジームのためのベース電圧であり、膜電位は一般に、正レジームにおいて v_+ から離れる傾向となる。

30

【0046】

[0056] v および u のためのヌルクライインは、それぞれ変換変数 q および r の負によって与えられる。パラメータ β は、 u ヌルクライインの傾きを制御するスケール係数である。パラメータ δ は通常、 $-v_-$ に等しく設定される。パラメータ ϵ は、両方のレジームにおいて v ヌルクライインの傾きを制御する抵抗値である。時間定数パラメータは、指数関数的減衰だけでなく、各レジームにおいて別個にヌルクライインの傾きを制御する。

【0047】

[0057]モデルは、電圧 v が値 v_s に達したときにスパイクするように定義される。続いて、状態は通常、（技術的に、スパイク事象と同じ 1 つのものであり得る）リセット事象でリセットされる。

40

【数11】

$$v = \hat{v}_- \quad (9)$$

【数12】

$$u = u + \Delta u \quad (10)$$

ここで、

【数13】

$$\hat{v}_-$$

および u はパラメータである。リセット電圧

【数14】

$$\hat{v}_+$$

は通常、 v_+ にセットされる。

10

【0048】

[0058] 瞬時結合の原理によって、状態について（また、単一の指指数項による）だけではなく、特定の状態に到達するために必要とされる時間についても、閉形式解が可能である。閉形式状態解は、次のとおりである。

【数15】

$$v(t + \Delta t) = (v(t) + q_\rho) e^{\frac{\Delta t}{\tau_\rho}} - q_\rho \quad (11)$$

【数16】

20

$$u(t + \Delta t) = (u(t) + r) e^{\frac{\Delta t}{\tau_u}} - r \quad (12)$$

【0049】

[0059] したがって、モデル状態は、入力（シナプス前スパイク）または出力（シナプス後スパイク）などの事象に伴ってのみ更新され得る。また、演算が（入力があるか、出力があるかを問わず）任意の特定の時間に実行され得る。

【0050】

[0060] その上、瞬時結合原理によって、反復的技法または数値解法（たとえば、オイラー数値解法）なしに、特定の状態に到達する時間が事前に決定され得るように、シナプス後スパイクの時間が予想され得る。前の電圧状態 v_0 を踏まえ、電圧状態 v_f に到達するまでの時間遅延は、次の式によって与えられる。

30

【数17】

$$\Delta t = \tau_\rho \log \frac{v_f + q_\rho}{v_0 + q_\rho} \quad (13)$$

【0051】

[0061] スパイクが、電圧状態 v が v_s に到達する時間に生じると定義される場合、電圧が所与の状態 v にある時間から測定されたスパイクが生じるまでの時間量、または相対的遅延に関する閉形式解は、次のとおりである。

40

【数18】

$$\Delta t_s = \begin{cases} \tau_+ \log \frac{v_s + q_+}{v + q_+} & \text{if } v > \hat{v}_+ \\ \infty & \text{otherwise} \end{cases} \quad (14)$$

ここで、

50

【数19】

 \hat{v}_+

は通常、パラメータ v_+ にセットされるが、他の変形も可能であり得る。

【0052】

[0062] モデルダイナミクスの上記の定義は、モデルが正レジームにあるか、それとも負レジームにあるかに依存する。上述のように、結合およびレジーム は、事象に伴って計算され得る。状態の伝搬のために、レジームおよび結合(変換)変数は、最後の(前の)事象の時間における状態に基づいて定義され得る。続いてスパイク出力時間を予想するために、レジームおよび結合変数は、次の(最新の)事象の時間における状態に基づいて定義され得る。10

【0053】

[0063] Cold モデルの、適時にシミュレーション、エミュレーションまたはモデルを実行するいくつかの可能な実装形態がある。これは、たとえば、事象更新モード、ステップ事象更新モード、およびステップ更新モードを含む。事象更新は、(特定の瞬間ににおける)事象または「事象更新」に基づいて状態が更新される更新である。ステップ更新は、間隔(たとえば、1 ms)においてモデルが更新される更新である。これは必ずしも、反復的技法または数値解法を必要とするとは限らない。また、事象がステップもしくはステップ間で生じる場合または「ステップ事象」更新によってモデルを更新するのみによって、ステップベースのシミュレータにおいて限られた時間分解能で事象ベースの実装形態が可能である。20

【0054】

ニューラルコーディング

[0064] 図1の人工ニューロン102、106から構成されるニューラルネットワークモデルなどの有用なニューラルネットワークモデルは、一致コーディング、時間コーディングまたはレートコーディングなど、様々な好適なニューラルコーディング方式のうちのいずれかを介して情報を符号化することができる。一致コーディングでは、情報は、ニューロン集団の活動電位(スパイキング活動)の一一致(または時間的近接度)で符号化される。時間コーディングでは、ニューロンは、絶対時間であるか相対時間であるかを問わず、活動電位(すなわち、スパイク)の正確なタイミングを通して情報を符号化する。したがって、情報は、ニューロン集団の間でスパイクの相対的タイミングで符号化され得る。対照的に、レートコーディングは、発火レートまたは集団発火レートでニューラル情報をコーディングすることを伴う。30

【0055】

[0065] ニューロンモデルは、時間コーディングを実行し得る場合、(レートは単に、タイミングまたはスパイク間の間隔の関数であるので)レートコーディングも実行し得る。時間コーディングを行うために、良いニューロンモデルは2つの要素を有する必要がある。すなわち、(1)入力の到着時間は出力時間に影響を与え、(2)一致検出は狭い時間ウィンドウを有し得る。時間パターンの要素を適切に遅延させることによって、要素はタイミング一致に組み込まれ得るので、結合遅延は、一致検出を時間パターン復号に拡大するための1つの手段を提供する。40

【0056】

到着時間

[0066] 良いニューロンモデルでは、入力の到着の時間は、出力の時間に影響を与えるはずである。シナプス入力は、ディラックのデルタ関数であるか、成形シナプス後電位(PSP: shaped post-synaptic potential)であるかを問わず、興奮性(EPSP)であるか、抑制性(IPSP)であるかを問わず、到着時間(たとえば、デルタ関数またはステップもしくは他の入力関数の開始もしくはピークの時間)を有し、これは入力時間と呼ばれ得る。ニューロン出力(すなわち、スパイク)は、(細胞体、軸索に沿ったポイント、50

または軸索の端部など、どこで測定される場合でも)発生の時間を有し、これは出力時間と呼ばれ得る。出力時間は、スパイクのピークの時間、スパイクの開始の時間、または出力波形に關係する任意の他の時間であり得る。支配的原理は、出力時間が入力時間に依存することである。

【0057】

[0067]一見したところ、すべてのニューロンモデルがこの原理に従うと思われるかも知れないが、これは一般には当てはまらない。たとえば、レートベースのモデルは、この特徴を有しない。多くのスパイキングモデルも、一般には適合しない。リーキー積分発火(LIF)モデルは、(しきい値を越えて)追加の入力がある場合にさらに速く発火することはない。その上、非常に高いタイミング分解能でモデル化された場合に適合する可能性があるモデルは多くの場合、タイミング分解能がたとえば1 m sのステップに限定されているときに適合しない。10

【0058】

入力

[0068]ニューロンモデルへの入力はディラックのデルタ関数、たとえば電流としての入力または伝導性ベースの入力を含み得る。後者の場合、ニューロン状態への寄与は連續的または状況依存的であり得る。

【0059】

スパイキングネットワークの効率的なハードウェア実装

[0069]本開示のいくつかの態様は、スパイキングニューラルネットワークの効率的なハードウェア実装形態を提供する。スパイクベースの計算は、ハードウェアで実装されるときに特別な考慮を必要とし得る。とても細かい粒度の並列は、しばしばスパイクベースの計算を伴う人工神経システム固有のものである。たとえば、ニューロンの「ファンアウト」が、スパイキングニューラルネットワーク(人工神経システム、たとえば、図1によるシステム100)における多数の人工ニューロンに関して効率的に計算される必要があり得る。この問題に対する解決策は、ダイナミックランダムアクセスメモリ(DRAM)を利用することに関係し得、それは、最高密度を有する現在の技術を表し、最高のメモリスループットのために連続アクセスを必要とする。さらに、スパイクベースの計算は、人工神経システムのシナプス前人工ニューロンおよびシナプス後人工ニューロンに対する可塑性照会を必要とし得る。20

【0060】

[0070]図5は、本開示のいくつかの態様による、人工神経システムの例示的なハードウェア実装形態500を示す。上述したようなSTD P学習ルールに基づくシナプス重みの更新(学習)は、結果可塑性の更新および再アセンブルブロック(Effect Plasticity Updates and Reassemble block)502において発生し得る。本開示のいくつかの態様では、更新されたシナプス重みおよび可塑性のパラメータが、キャッシュラインインターフェース504を介してオフチップメモリ(たとえば、DRAM506)に記憶され得る。30

【0061】

[0071]典型的な人工神経システムでは、人工ニューロンよりもはるかに多くのシナプス(シナプス結合)がある。したがって、大きいニューラルネットワーク(人工神経システム)の場合、効率的な方法でシナプス重み更新を処理することがしばしば望まれる。多数のシナプスは、シナプス重みおよび他のパラメータをメモリ(たとえば、DRAM506)に記憶することを示唆し得る。人工神経システムの人工ニューロンがいわゆる「スーパーニューロン」(SN)(たとえば、図5におけるSN508)においてスパイクを生成するとき、人工ニューロンは、シナプス後人工ニューロンと対応するニューラル重みとを決定するために、DRAMルックアップを通して、シナプス後人工ニューロンにそれらのスパイクを転送し得る。40

【0062】

[0072]高速で効率的なルックアップを可能にするために、シナプス配列が、たとえば、人工ニューロンからのファンアウトに基づいて、メモリで連続的に維持され得る。後に、50

図5の結果可塑性の更新および再アセンブルブロック502においてSTD P更新を処理するときに、LTP更新についての逆マッピングを決定するためにDRAMまたは大きいルックアップテーブルが検索される必要がないことから、このメモリレイアウトに与えられた転送ファンアウト(forward fan-out)に基づいてSTD P更新を処理することを効率性が規定し得る。図5に示す手法は、これを可能にする。結果可塑性の更新および再アセンブルブロック502は、シナプス前スパイク時間とシナプス後スパイク時間とを取得しようとして、スーパーニューロン508に照会し得、ここでも、関係する状態メモリの量を低減し得る。

【0063】

[0073]図5はさらに、入力電流アクチュエータおよびニューラルモデルのための時間多重化ハードウェアに基づくスーパーニューロンの概念を示す。図5に示すように、効率性、すなわち実行速度を改善するために、複数のスーパーニューロン508が並列に動作し得る。本開示のいくつかの態様によれば、スーパーニューロン508における計算ユニットの数は、そのスーパーニューロンに割り当てられた人工ニューロンの数よりも少くなり得る、すなわち、1つの計算ユニットが複数の人工ニューロンに関連付けられ得る。

【0064】

[0074]本開示の一態様では、人工ニューロンの状態は、定められた時間期間、すなわち「タウ」ステップごとに更新され得る。一態様では、人工ニューロンはメモリに、たとえばDRAM506に、シナプス後ファンアウトへのポインタを記憶し得る。一態様では、各スーパーニューロン508は、ニューラル入力の蓄積を実施し得る。

【0065】

[0075]本開示のいくつかの態様によれば、ファンアウトベースのシナプスマモリ構成は、DRAMに適した、高スループットおよび長い「バースト的」読み取りをもたらし得る。たとえば、単一の512ビットの「キャッシュライン」(たとえば、図5におけるキャッシュラインインターフェース504)は、複数のシナプス結合を備え得る。512ビットのアクセスは現在、DRAMにとって最高の効率性を表すことに留意されたい。

【0066】

[0076]本開示のいくつかの態様によれば、シナプス前ニューロンおよびシナプス後ニューロンのルックアップが、可塑性学習ルールを実施するために利用され得る。さらに、本開示のいくつかの態様は、スパイクの動的確率的除去(dynamic probabilistic dropping)をサポートし、ここにおいて、従来のプロセッサがその構造にしつかり統合され得る。本開示の一態様では、DRAMメモリ506に記憶されたシナプスに関する情報(たとえば、重みおよび他のパラメータ)は、宛先ルーティング情報を備え得る。これは、メモリのマッピングおよび構成によって達成され得る。さらに、外部および内部シナプス活性化が、他の処理ノードとの同期と同様に、例示的なハードウェア実装形態500に統合され得る。本開示で提供されるメモリ構成は、各シナプス前ニューロンがそのシナプスのすべてと一緒に備え得ることをサポートする。

【0067】

[0077]本開示のいくつかの態様によれば、可塑性照会は、シナプス後ニューロンにルート要求を提供し得る。シナプス前(攻撃者)ニューロンからシナプス後(犠牲者)ニューロンにメッセージが送られ得、次いでシナプス後(犠牲者)ニューロンは、(戻りアドレスを含む)情報を返信し得る。別個のシナプスブロックがこれを可能にし得る。

【0068】

[0078]図6は、本開示のいくつかの態様による、複数の人工ニューロンを伴う人工神経システムを動作させるための例示的な動作600の流れ図である。動作600は、ハードウェアで(たとえば、神経形態学的プロセッサなどの1つもしくは複数のニューラル処理ユニットによって)、ソフトウェアで、またはファームウェアで実行され得る。人工神経システムは、視覚神経系、聴覚神経系、海馬などのような様々な生物学上または想像上の神経系のうちのいずれかでモデル化され得る。

【0069】

10

20

30

40

50

[0079]動作 600 は、602において、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させることによって始まり得、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられ得る。604において、スーパーニューロン処理ユニットは、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにメモリとインターフェースされ得、ここにおいて、メモリの構成が連続メモリアクセスを可能にする。

【0070】

[0080]本開示の一態様では、複数のスーパーニューロン処理ユニットを同時に動作させることは、入力電流アキュムレータおよび人工ニューロンのモデルのためにスーパーニューロン処理ユニットを時間多重化することを備え得る。スーパーニューロン処理ユニットのうちの1つに割り当てられた人工ニューロンの状態は、定められたステップ（たとえば、「タウ」ステップ）ごとに更新され得る。一態様では、人工ニューロンのいずれかに関連するシナプス後ファンアウトへのポインタが、メモリに記憶され得る。一態様では、複数のスーパーニューロン処理ユニットを同時に動作させることは、スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施することを備え得る。

10

【0071】

[0081]本開示の一態様では、各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた人工ニューロンの数よりも少なくなり得る。メモリは、たとえば、DRAMを備え得、DRAMの連続メモリアクセスは、複数のシナプスに関するデータへの同時アクセス（たとえば、最大512ビット）を備え得る。

20

【0072】

[0082]一態様では、メモリの構成は、可塑性を実現するためのシナプス前ニューロンおよびシナプス後ニューロンのルックアップを備え得る。一態様では、人工神経システムのスパイクの除去が、動的確率的方法で実行され得る。一態様では、人工神経システムのシナプスに関するデータがメモリに記憶され得、ここにおいて、データは宛先ルーティング情報を備え得る。一態様では、外部および内部シナプス活性化が、スーパーニューロン処理ユニットとメモリの構成とを利用することによって人工神経システムに統合され得る。

30

【0073】

[0083]図7は、本開示のいくつかの態様による、汎用プロセッサ702を使用して複数の人工ニューロンを伴う人工神経システムを動作させるための上述した方法の例示的なブロック図700を示す。計算ネットワーク（ニューラルネットワーク）に関連する変数（ニューラル信号）、シナプス重み、および／またはシステムパラメータは、メモリプロック704に記憶され得る一方、汎用プロセッサ702において実行される関連した命令は、プログラムメモリ706からロードされ得る。本開示の一態様では、汎用プロセッサ702にロードされる命令は、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにスーパーニューロン処理ユニットをメモリとインターフェースすること、ここにおいて、メモリの構成が連続メモリアクセスを可能にする、とを行うためのコードを備え得る。

40

【0074】

[0084]図8は、本開示のいくつかの態様による、メモリ802が相互接続ネットワーク804を介して計算ネットワーク（ニューラルネットワーク）の個々の（分散型）処理ユニット（ニューラルプロセッサ）806とインターフェースされ得る複数の人工ニューロンを伴う人工神経システムを動作させるための上述した方法の例示的なブロック図800を示す。計算ネットワーク（ニューラルネットワーク）に関連する変数（ニューラル信号）、シナプス重み、および／またはシステムパラメータは、メモリ802に記憶され得、

50

相互接続ネットワーク 804 の接続を介してメモリ 802 から各処理ユニット（ニューラルプロセッサ）806 にロードされ得る。本開示の一態様では、処理ユニット 806 は、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにスーパーニューロン処理ユニットをメモリとインターフェースすること、ここにおいて、メモリの構成が連続メモリアクセスを可能にする、とを行うように構成され得る。

【0075】

[0085]図 9 は、本開示のいくつかの態様による、分散型重みメモリ 902 および分散型処理ユニット（ニューラルプロセッサ）904 に基づいて複数の人工ニューロンを伴う人工神経システムを動作させるための上述した方法の例示的なプロック図 900 を示す。図 9 に示すように、1 つのメモリバンク 902 が、計算ネットワーク（ニューラルネットワーク）の 1 つの処理ユニット 904 と直接インターフェースされてよく、メモリバンク 902 は、その処理ユニット（ニューラルプロセッサ）904 に関連する変数（ニューラル信号）、シナプス重み、および / またはシステムパラメータを記憶することができる。本開示の一態様では、処理ユニット 904 は、人工ニューロンの状態を更新するために人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするためにスーパーニューロン処理ユニットをメモリとインターフェースすること、ここにおいて、メモリの構成が連続メモリアクセスを可能にする、とを行うように構成され得る。

【0076】

[0086]図 10 は、本開示のいくつかの態様による、ニューラルネットワーク 1000 の例示的な実装形態を示す。図 10 に示すように、ニューラルネットワーク 1000 は、上述した方法の様々な動作を実行し得る複数のローカル処理ユニット 1002 を備えることができる。各処理ユニット 1002 は、ローカル状態メモリ 1004 と、ニューラルネットワークのパラメータを記憶するローカルパラメータメモリ 1006 とを備えることができる。さらに、処理ユニット 1002 は、ローカル（ニューロン）モデルプログラムを有するメモリ 1008 と、ローカル学習プログラムを有するメモリ 1010 と、ローカル接続メモリ 1012 とを備えることができる。さらに、図 10 に示すように、各ローカル処理ユニット 1002 は、ローカル処理ユニットのローカルメモリのための設定を提供し得る設定処理のためのユニット 1014 と、またローカル処理ユニット 1002 間のルーティングを提供するルーティング接続処理要素 1016 とインターフェースされ得る。

【0077】

[0087]本開示のいくつかの態様によれば、各ローカル処理ユニット 1002 は、ニューラルネットワークの所望の 1 つまたは複数の機能的特徴に基づいてニューラルネットワークのパラメータを決定し、決定されたパラメータがさらに適合、調整および更新されることで、1 つまたは複数の機能的特徴を所望の機能的特徴に向けて発展させるように構成され得る。

【0078】

[0088]上述した方法の様々な動作は、対応する機能を実行することが可能な任意の好適な手段によって実行され得る。それらの手段は、限定はしないが、回路、特定用途向け集積回路（A S I C）、またはプロセッサを含む、様々なハードウェアおよび / またはソフトウェア構成要素および / またはモジュールを含み得る。たとえば、様々な動作は、図 5、図 7 ~ 図 10 に示す様々なプロセッサのうちの 1 つまたは複数によって実行され得る。概して、図に示されている動作がある場合、それらの動作は、同様の番号をもつ対応するカウンターパートのミーンズプラスファンクション構成要素を有し得る。たとえば、図 6 に示す動作 600 は、図 6 A に示す手段 600 A に対応する。

10

20

30

40

50

【0079】

[0089]たとえば、表示するための手段は、ディスプレイ（たとえば、モニタ、フラットスクリーン、タッチスクリーンなど）、プリンタ、または視覚的描写（たとえば、表、チャートもしくはグラフ）のためのデータを出力するための任意の他の好適な手段を含み得る。処理するための手段、受信するための手段、追跡するための手段、調整するための手段、更新するための手段、または決定するための手段は、1つまたは複数のプロセッサまたは処理ユニットを含み得る、処理システムを備え得る。感知するための手段は、センサーを含み得る。記憶するための手段は、処理システムによってアクセスされ得る、メモリまたは任意の他の好適な記憶デバイス（たとえば、RAM）を含み得る。

【0080】

10

[0090]本明細書で使用する「決定」という用語は、多種多様なアクションを包含する。たとえば、「決定」は、計算すること、算出すること、処理すること、導出すること、調査すること、ルックアップすること（たとえば、テーブル、データベースまたは別のデータ構造においてルックアップすること）、確認することなどを含み得る。また、「決定」は、受信すること（たとえば、情報を受信すること）、アクセスすること（たとえば、メモリ中のデータにアクセスすること）などを含み得る。また、「決定」は、解決すること、選択すること、選定すること、確立することなどを含み得る。

【0081】

[0091]本明細書で使用する、項目のリスト「のうちの少なくとも1つ」を指す句は、単一のメンバーを含む、それらの項目の任意の組合せを指す。一例として、「a、b、またはcのうちの少なくとも1つ」は、a、b、c、a - b、a - c、b - c、およびa - b - cを包含するものとする。

20

【0082】

[0092]本開示に関連して説明した様々な例示的な論理ブロック、モジュール、および回路は、汎用プロセッサ、デジタル信号プロセッサ（DSP）、特定用途向け集積回路（ASIC）、フィールドプログラマブルゲートアレイ信号（FPGA）または他のプログラマブル論理デバイス（PLD）、個別ゲートまたはトランジスタ論理、個別ハードウェア構成要素、あるいは本明細書で説明した機能を実行するように設計されたそれらの任意の組合せを用いて実装または実行され得る。汎用プロセッサはマイクロプロセッサであり得るが、代替として、プロセッサは、任意の市販のプロセッサ、コントローラ、マイクロコントローラまたは状態機械であり得る。プロセッサはまた、コンピューティングデバイスの組合せ、たとえば、DSPとマイクロプロセッサとの組合せ、複数のマイクロプロセッサ、DSPコアと連携する1つまたは複数のマイクロプロセッサ、あるいは任意の他のそのような構成として実装され得る。

30

【0083】

[0093]本開示に関連して説明した方法またはアルゴリズムのステップは、ハードウェアで直接実施されるか、プロセッサによって実行されるソフトウェアモジュールで実施されるか、またはその2つの組合せで実施され得る。ソフトウェアモジュールは、当技術分野で知られている任意の形態の記憶媒体中に常駐し得る。使用され得る記憶媒体のいくつかの例としては、ランダムアクセスメモリ（RAM）、読み取り専用メモリ（ROM）、フラッシュメモリ、EPROMメモリ、EEPROM（登録商標）メモリ、レジスタ、ハードディスク、リムーバブルディスク、CD-ROMなどを含む。ソフトウェアモジュールは、単一の命令、または多数の命令を備えることができ、いくつかの異なるコードセグメント上で、異なるプログラム間で、複数の記憶媒体にわたって分散され得る。記憶媒体は、プロセッサがその記憶媒体から情報を読み取ることができ、その記憶媒体に情報を書き込むことができるよう、プロセッサに結合され得る。代替として、記憶媒体はプロセッサと一体化され得る。

40

【0084】

[0094]本明細書で開示する方法は、説明した方法を達成するための1つまたは複数のステップまたはアクションを備える。本方法のステップおよび／またはアクションは、特許

50

請求の範囲から逸脱することなく互いに交換され得る。言い換えれば、ステップまたはアクションの特定の順序が指定されない限り、特定のステップおよび／またはアクションの順序および／または使用は、特許請求の範囲から逸脱することなく変更され得る。

【 0 0 8 5 】

[0095] 説明した機能は、ハードウェア、ソフトウェア、ファームウェア、またはそれらの任意の組合せで実装され得る。ハードウェアで実装される場合、例示的なハードウェア構成はデバイス中に処理システムを備え得る。処理システムは、バスアーキテクチャを用いて実装され得る。バスは、処理システムの特定の適用例および全体的な設計制約に応じて、任意の数の相互接続バスとブリッジとを含み得る。バスは、プロセッサと、機械可読媒体と、バスインターフェースとを含む様々な回路を互いにリンクし得る。バスインターフェースは、ネットワークアダプタを、特に、バスを介して処理システムに接続するため10に使用され得る。ネットワークアダプタは、信号処理機能を実装するために使用され得る。いくつかの態様では、ユーザインターフェース（たとえば、キーパッド、ディスプレイ、マウス、ジョイスティックなど）もバスに接続され得る。バスはまた、タイミングソース、周辺機器、電圧調整器、電力管理回路などの様々な他の回路にリンクし得るが、それらは当技術分野でよく知られており、したがってこれ以上は説明されない。

【 0 0 8 6 】

[0096] プロセッサは、機械可読媒体に記憶されたソフトウェアの実行を含む、バスおよび一般的な処理を管理することを担当し得る。プロセッサは、1つまたは複数の汎用および／または専用プロセッサを用いて実装され得る。例としては、マイクロプロセッサ、マイクロコントローラ、DSPプロセッサ、およびソフトウェアを実行し得る他の回路を含む。ソフトウェアは、ソフトウェア、ファームウェア、ミドルウェア、マイクロコード、ハードウェア記述言語などの名称にかかわらず、命令、データ、またはそれらの任意の組合せを意味すると広く解釈されたい。機械可読媒体は、例として、RAM（ランダムアクセスメモリ）、フラッシュメモリ、ROM（読み取り専用メモリ）、PROM（プログラマブル読み取り専用メモリ）、EPROM（消去可能プログラマブル読み取り専用メモリ）、EEPROM（電気消去可能プログラマブル読み取り専用メモリ）、レジスタ、磁気ディスク、光ディスク、ハードドライブ、または任意の他の好適な記憶媒体、あるいはそれらの任意の組合せを含み得る。機械可読媒体はコンピュータプログラム製品において実施され得る。コンピュータプログラム製品はパッケージング材料を備え得る。
30

【 0 0 8 7 】

[0097] ハードウェア実装形態では、機械可読媒体は、プロセッサとは別個の処理システムの一部であり得る。しかしながら、当業者なら容易に理解するように、機械可読媒体またはその任意の部分は処理システムの外部にあり得る。例として、機械可読媒体は、すべてバスインターフェースを介してプロセッサによってアクセスされ得る、伝送線路、データによって変調された搬送波、および／またはデバイスとは別個のコンピュータ製品を含み得る。代替的に、または追加で、機械可読媒体またはその任意の部分は、キャッシュおよび／または汎用レジスタファイルがそうであり得るように、プロセッサに統合され得る。

【 0 0 8 8 】

[0098] 処理システムは、すべて外部バスアーキテクチャを介して他のサポート回路と互いにリンクされる、プロセッサ機能を提供する1つまたは複数のマイクロプロセッサと、機械可読媒体の少なくとも一部分を提供する外部メモリとをもつ汎用処理システムとして構成され得る。代替的に、処理システムは、プロセッサをもつASIC（特定用途向け集積回路）と、バスインターフェースと、ユーザインターフェースと、サポート回路と、単一のチップに統合された機械可読媒体の少なくとも一部分とを用いて、あるいは1つまたは複数のFPGA（フィールドプログラマブルゲートアレイ）、PLD（プログラマブル論理デバイス）、コントローラ、状態機械、ゲート論理、個別ハードウェア構成要素、もしくは他の好適な回路、または本開示全体にわたって説明した様々な機能を実行し得る回路の任意の組合せを用いて、実装され得る。当業者なら、特定の適用例と、全体的なシス
40

10

20

30

40

50

テムに課される全体的な設計制約とに応じて、どのようにしたら処理システムについて説明した機能を最も良く実装し得るかを理解されよう。

【0089】

[0099] 機械可読媒体はいくつかのソフトウェアモジュールを備え得る。ソフトウェアモジュールは、プロセッサによって実行されたときに、処理システムに様々な機能を実行させる命令を含む。ソフトウェアモジュールは、送信モジュールと受信モジュールとを含み得る。各ソフトウェアモジュールは、単一の記憶デバイス中に常駐するか、または複数の記憶デバイスにわたって分散され得る。例として、トリガイベントが発生したとき、ソフトウェアモジュールがハードドライブから RAM にロードされ得る。ソフトウェアモジュールの実行中、プロセッサは、アクセス速度を高めるために、命令のいくつかをキャッシュにロードし得る。次いで、1つまたは複数のキャッシュラインが、プロセッサによる実行のために汎用レジスタファイルにロードされ得る。以下でソフトウェアモジュールの機能に言及する場合、そのような機能は、そのソフトウェアモジュールからの命令を実行したときにプロセッサによって実装されることが理解されよう。

【0090】

[0100] ソフトウェアで実装される場合、機能は、1つまたは複数の命令またはコードとしてコンピュータ可読媒体上に記憶されるか、あるいはコンピュータ可読媒体を介して送信され得る。コンピュータ可読媒体は、ある場所から別の場所へのコンピュータプログラムの転送を可能にする任意の媒体を含む、コンピュータ記憶媒体と通信媒体の両方を含む。記憶媒体は、コンピュータによってアクセスされ得る任意の利用可能な媒体であり得る。限定ではなく例として、そのようなコンピュータ可読媒体は、RAM、ROM、EEPROM、CD-ROM または他の光ディスクストレージ、磁気ディスクストレージまたは他の磁気記憶デバイス、あるいは命令またはデータ構造の形態の所望のプログラムコードを搬送または記憶するために使用され得、コンピュータによってアクセスされ得る、任意の他の媒体を備えることができる。また、いかなる接続もコンピュータ可読媒体を適切に名づけられる。たとえば、ソフトウェアが、同軸ケーブル、光ファイバーケーブル、ツイストペア、デジタル加入者回線 (DSL)、または赤外線 (IR)、無線、およびマイクロ波などのワイヤレス技術を使用して、ウェブサイト、サーバ、または他のリモートソースから送信される場合、同軸ケーブル、光ファイバーケーブル、ツイストペア、DSL、または赤外線、無線、およびマイクロ波などのワイヤレス技術は、媒体の定義に含まれる。本明細書で使用するディスク (disk) およびディスク (disc) は、コンパクトディスク (disc) (CD)、レーザーディスク (登録商標) (disc)、光ディスク (disc)、デジタル多用途ディスク (disc) (DVD)、フロッピー (登録商標) ディスク (disk)、および Blu-ray (登録商標) ディスク (disc) を含み、ディスク (disk) は、通常、データを磁気的に再生し、ディスク (disc) は、データをレーザーで光学的に再生する。したがって、いくつかの態様では、コンピュータ可読媒体は非一時的コンピュータ可読媒体（たとえば、有形媒体）を備え得る。さらに、他の態様では、コンピュータ可読媒体は一時的コンピュータ可読媒体（たとえば、信号）を備え得る。上記の組合せもコンピュータ可読媒体の範囲内に含まれるべきである。

【0091】

[0101] したがって、いくつかの態様は、本明細書で提示する動作を実行するためのコンピュータプログラム製品を備え得る。たとえば、そのようなコンピュータプログラム製品は、本明細書で説明する動作を実行するために1つまたは複数のプロセッサによって実行可能である命令を記憶した（および / または符号化した）コンピュータ可読媒体を備え得る。いくつかの態様では、コンピュータプログラム製品はパッケージング材料を含み得る。

【0092】

[0102] さらに、本明細書で説明した方法および技法を実行するためのモジュールおよび / または他の適切な手段は、適用可能な場合にデバイスによってダウンロードされ、および / または他の方法で取得され得ることを諒解されたい。たとえば、そのようなデバイス

10

20

30

40

50

は、本明細書で説明した方法を実施するための手段の転送を可能にするためにサーバに結合され得る。代替的に、本明細書で説明した様々な方法は、デバイスが記憶手段をデバイスに結合または提供すると様々な方法を得ることができるように、記憶手段（たとえば、RAM、ROM、コンパクトディスク（CD）またはフロッピーディスクなどの物理記憶媒体など）によって提供され得る。その上、本明細書で説明した方法および技法をデバイスに与えるための任意の他の好適な技法が利用され得る。

【0093】

[0103]特許請求の範囲は、上記で示した厳密な構成および構成要素に限定されないことを理解されたい。上記で説明した方法および装置の構成、動作および詳細において、特許請求の範囲から逸脱することなく、様々な改変、変更および変形が行われ得る。 10

以下に、本願出願の当初の特許請求の範囲に記載された発明を付記する。

[C 1]

複数の人工ニューロンを伴う人工神経システムを動作させるための方法であって、前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、前記人工神経システムのシナップス重みおよび可塑性のパラメータにアクセスするために、前記スーパーニューロン処理ユニットをメモリとインターフェースすること、ここにおいて、前記メモリの構成が連続メモリアクセスを可能にする、と、を備える方法。 20

[C 2]

複数のスーパーニューロン処理ユニットを同時に動作させることは、入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化することを備える、C 1に記載の方法。

[C 3]

前記スーパーニューロン処理ユニットのうちの1つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新することをさらに備える、C 1に記載の方法。

[C 4]

前記人工ニューロンのいずれかに関連するシナップス後ファンアウトへのポインタを前記メモリに記憶することをさらに備える、C 1に記載の方法。

[C 5]

複数のスーパーニューロン処理ユニットを同時に動作させることは、前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施することを備える、C 1に記載の方法。

[C 6]

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、C 1に記載の方法。 40

[C 7]

前記メモリは、ダイナミックランダムアクセスメモリ（DRAM）を備え、前記DRAMの前記連続メモリアクセスは、複数のシナップスに関するデータへの同時アクセスを備える、C 1に記載の方法。

[C 8]

メモリの前記構成は、可塑性を実現するためのシナップス前ニューロンおよびシナップス後ニューロンのルックアップテーブルを備える、C 1に記載の方法。

[C 9]

10

20

30

40

50

動的確率的方法での前記人工神経システムのスパイクの除去
をさらに備える、C 1に記載の方法。

[C 1 0]

前記人工神経システムのシナップスに関連するデータを前記メモリに記憶することをさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、C 1に記載の方法。

[C 1 1]

前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、前記人工神経システムに外部および内部シナップス活性化を統合することをさらに備える、C 1に記載の方法。

10

[C 1 2]

複数の人工ニューロンを伴う人工神経システムを動作させるための装置であって、前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させるように構成された第1の回路、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当たられる、と、

前記人工神経システムのシナップス重みおよび可塑性のパラメータにアクセスするために、前記スーパーニューロン処理ユニットをメモリとインターフェースするように構成された第2の回路、ここにおいて、前記メモリの構成が連続メモリアクセスを可能にする、と、

20

を備える装置。

[C 1 3]

前記第1の回路はまた、
入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化する
ように構成される、C 1 2に記載の装置。

[C 1 4]

前記スーパーニューロン処理ユニットのうちの1つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新するように構成された第3の回路をさらに備える、C 1 2に記載の装置。

30

[C 1 5]

前記人工ニューロンのいずれかに関連するシナップス後ファンアウトへのポインタを前記メモリに記憶するように構成された第3の回路
をさらに備える、C 1 2に記載の装置。

[C 1 6]

前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施するように構成された第3の回路
をさらに備える、C 1 2に記載の装置。

[C 1 7]

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、C 1 2に記載の装置。

40

[C 1 8]

前記メモリは、ダイナミックランダムアクセスメモリ(D R A M)を備え、
前記D R A Mの前記連続メモリアクセスは、複数のシナップスに関するデータへの同時アクセスを備える、C 1 2に記載の装置。

[C 1 9]

メモリの前記構成は、可塑性を実現するためのシナップス前ニューロンおよびシナップス後ニューロンのルックアップテーブルを備える、C 1 2に記載の装置。

[C 2 0]

50

動的確率的方法で前記人工神経システムのスパイクを除去するように構成された第3の回路

をさらに備える、C 1 2 に記載の装置。

[C 2 1]

前記人工神経システムのシナプスに関連するデータを前記メモリに記憶するように構成された第3の回路をさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、C 1 2 に記載の装置。

[C 2 2]

前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、前記人工神経システムに外部および内部シナプス活性化を統合するように構成された第3の回路

をさらに備える、C 1 2 に記載の装置。

[C 2 3]

複数の人工ニューロンを伴う人工神経システムを動作させるための装置であって、前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させるための手段、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、

前記人工神経システムのシナプス重みおよび可塑性のパラメータにアクセスするために、前記スーパーニューロン処理ユニットをメモリとインターフェースするための手段、ここにおいて、前記メモリの構成が連續メモリアクセスを可能にする、と、

を備える装置。

[C 2 4]

複数のスーパーニューロン処理ユニットを同時に動作させるための前記手段は、入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化するための手段

をさらに備える、C 2 3 に記載の装置。

[C 2 5]

前記スーパーニューロン処理ユニットのうちの1つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新するための手段

をさらに備える、C 2 3 に記載の装置。

30

[C 2 6]

前記人工ニューロンのいずれかに関連するシナプス後ファンアウトへのポインタを前記メモリに記憶するための手段

をさらに備える、C 2 3 に記載の装置。

[C 2 7]

複数のスーパーニューロン処理ユニットを同時に動作させるための前記手段は、前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施するための手段

をさらに備える、C 2 3 に記載の装置。

[C 2 8]

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、C 2 3 に記載の装置。

40

[C 2 9]

前記メモリは、ダイナミックランダムアクセスメモリ(D R A M)を備え、

前記D R A Mの前記連続メモリアクセスは、複数のシナプスに関係するデータへの同時アクセスを備える、C 2 3 に記載の装置。

[C 3 0]

メモリの前記構成は、可塑性を実現するためのシナプス前ニューロンおよびシナプス後ニューロンのルックアップテーブルを備える、C 2 3 に記載の装置。

50

[C 3 1]

動的確率的方法での前記人工神経システムのスパイクの除去のための手段をさらに備える、C 2 3 に記載の装置。

[C 3 2]

前記人工神経システムのシナップスに関連するデータを前記メモリに記憶するための手段をさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、C 2 3 に記載の装置。

[C 3 3]

前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、前記人工神経システムに外部および内部シナップス活性化を統合するための手段をさらに備える、C 2 3 に記載の装置。

10

[C 3 4]

非一時的コンピュータ可読媒体を備える、複数の人工ニューロンを伴う人工神経システムを動作させるためのコンピュータプログラム製品であって、前記非一時的コンピュータ可読媒体が、

前記人工ニューロンの状態を更新するために前記人工神経システムにおける複数のスーパーニューロン処理ユニットを同時に動作させること、ここにおいて、前記人工ニューロンのサブセットが各スーパーニューロン処理ユニットに割り当てられる、と、

前記人工神経システムのシナップス重みおよび可塑性のパラメータにアクセスするために、前記スーパーニューロン処理ユニットをメモリとインターフェースすること、ここにおいて、前記メモリの構成が連続メモリアクセスを可能にする、と、

20

[C 3 5]

前記コンピュータ可読媒体は、

入力電流アキュムレータおよび前記人工ニューロンのモデルのために前記スーパーニューロン処理ユニットを時間多重化する

ためのコードをさらに備える、C 3 4 に記載のコンピュータプログラム製品。

[C 3 6]

前記コンピュータ可読媒体は、

前記スーパーニューロン処理ユニットのうちの 1 つに割り当てられた任意の人工ニューロンの状態を、定められたステップごとに更新する

30

ためのコードをさらに備える、C 3 4 に記載のコンピュータプログラム製品。

[C 3 7]

前記コンピュータ可読媒体は、

前記人工ニューロンのいずれかに関連するシナップス後ファンアウトへのポインタを前記メモリに記憶する

ためのコードをさらに備える、C 3 4 に記載のコンピュータプログラム製品。

[C 3 8]

前記コンピュータ可読媒体は、

前記スーパーニューロン処理ユニットの各々によって、ニューラル入力アキュムレータを実施する

40

ためのコードをさらに備える、C 3 4 に記載のコンピュータプログラム製品。

[C 3 9]

各スーパーニューロン処理ユニットにおける計算ユニットの数は、そのスーパーニューロン処理ユニットに割り当てられた前記人工ニューロンの数よりも少ない、C 3 4 に記載のコンピュータプログラム製品。

[C 4 0]

前記メモリは、ダイナミックランダムアクセスメモリ(D R A M)を備え、

前記 D R A M の前記連続メモリアクセスは、複数のシナップスに関するデータへの同時アクセスを備える、C 3 4 に記載のコンピュータプログラム製品。

50

[C 4 1]

メモリの前記構成は、可塑性を実現するためのシナプス前ニューロンおよびシナプス後ニューロンのルックアップテーブルを備える、C 3 4 に記載のコンピュータプログラム製品。

[C 4 2]

前記コンピュータ可読媒体は、
動的確率的方法での前記人工神経システムのスパイクの除去
のためのコードをさらに備える、C 3 4 に記載のコンピュータプログラム製品。

[C 4 3]

前記コンピュータ可読媒体は、
前記人工神経システムのシナプスに関連するデータを前記メモリに記憶するためのコードをさらに備え、ここにおいて、前記データが宛先ルーティング情報を備える、C 3 4 に記載のコンピュータプログラム製品。

[C 4 4]

前記コンピュータ可読媒体は、
前記スーパーニューロン処理ユニットと前記メモリの構成とを利用することによって、
前記人工神経システムに外部および内部シナプス活性化を統合する
ためのコードをさらに備える、C 3 4 に記載のコンピュータプログラム製品。

10

【図 1】

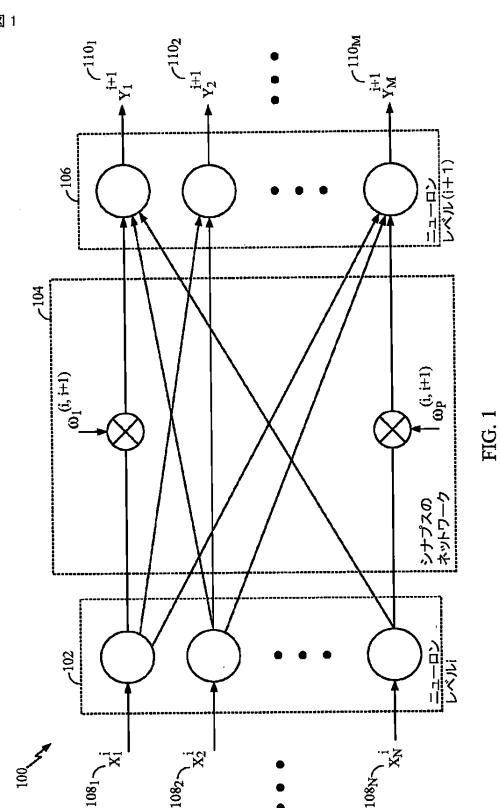


FIG. 1

【図 2】

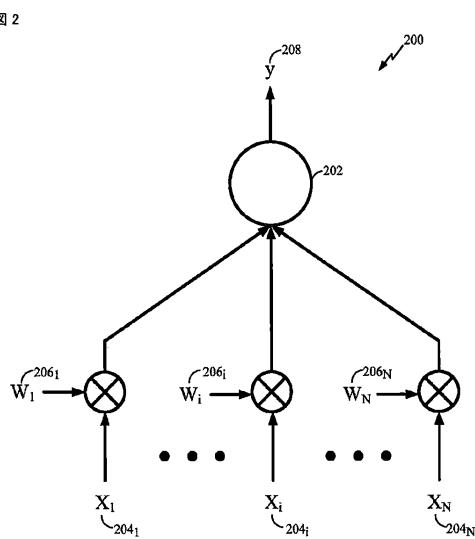


FIG. 2

【図3】

図3

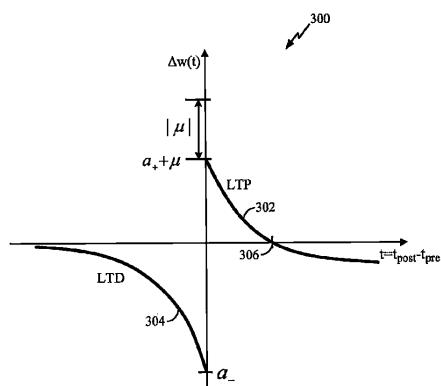


FIG. 3

【図4】

図4

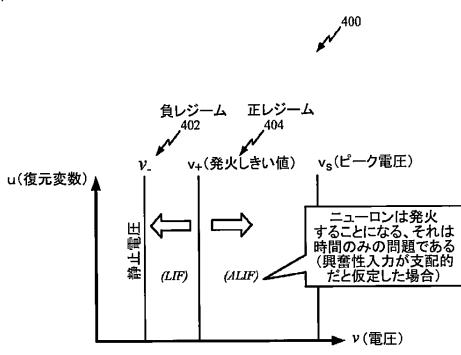


FIG. 4

【図5】

図5

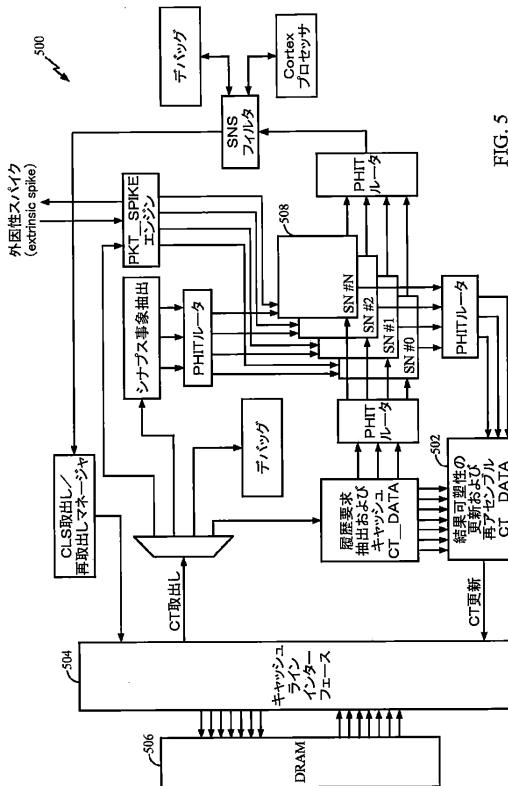


FIG. 5

【図6】

図6

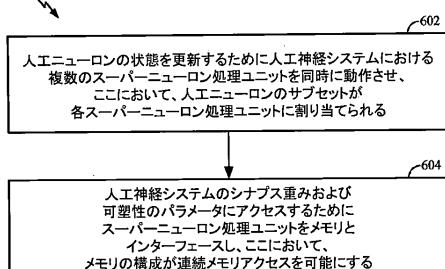


FIG. 6

【図 6 A】

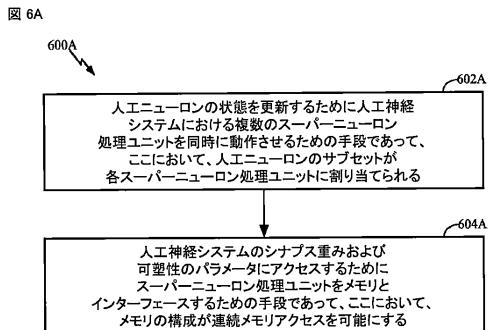


FIG. 6A

【図 7】

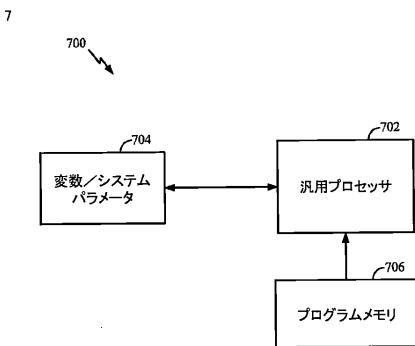


FIG. 7

【図 8】

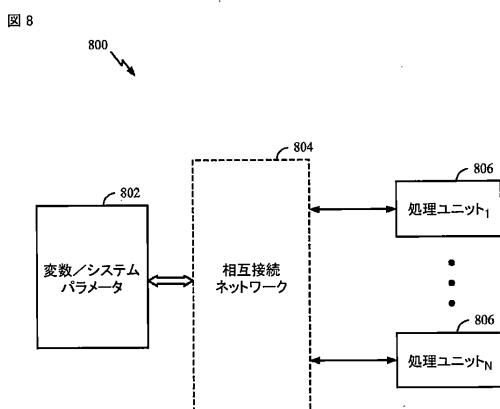


FIG. 8

【図 9】

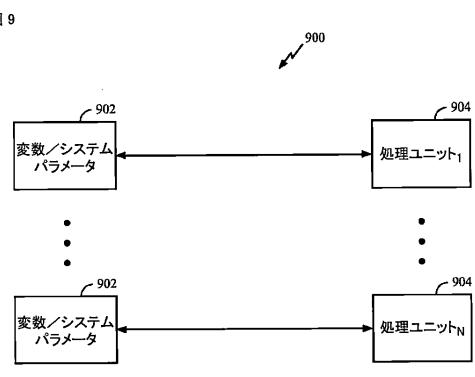


FIG. 9

【図 10】

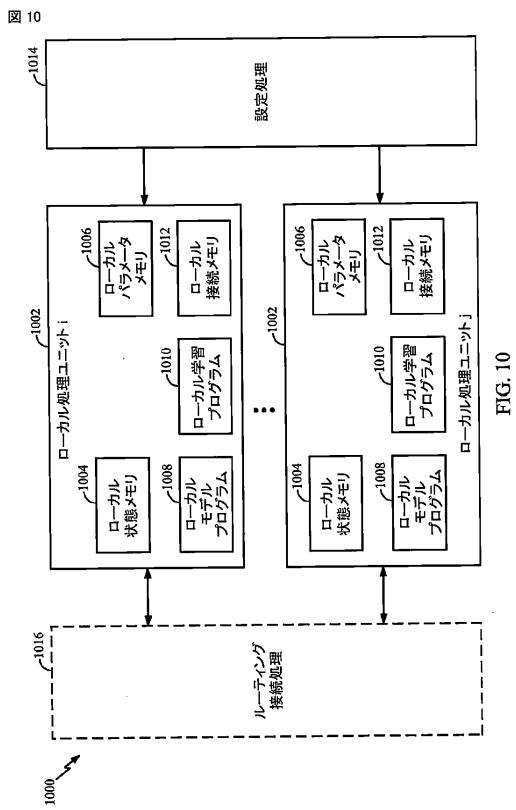


FIG. 10

フロントページの続き

(31)優先権主張番号 61/862,741
(32)優先日 平成25年8月6日(2013.8.6)
(33)優先権主張国 米国(US)
(31)優先権主張番号 61/862,734
(32)優先日 平成25年8月6日(2013.8.6)
(33)優先権主張国 米国(US)
(31)優先権主張番号 61/932,364
(32)優先日 平成26年1月28日(2014.1.28)
(33)優先権主張国 米国(US)
(31)優先権主張番号 14/267,005
(32)優先日 平成26年5月1日(2014.5.1)
(33)優先権主張国 米国(US)

早期審査対象出願

(72)発明者 レビン、ジェフリー・アレキサンダー
アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドラ
イブ 5775
(72)発明者 ランガン、ベンカト
アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドラ
イブ 5775
(72)発明者 マロン、エリック・クリストファー
アメリカ合衆国、カリフォルニア州 92121-1714、サン・ディエゴ、モアハウス・ドラ
イブ 5775

審査官 石川 亮

(56)参考文献 特表2015-501972(JP,A)
特表2014-530428(JP,A)
米国特許出願公開第2013/0073497(US,A1)

(58)調査した分野(Int.Cl., DB名)

G 06 N 3 / 02 - 3 / 10