(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)特許出願公表番号
特表2017-509982
(P2017-509982A)
(43)公表日　平成29年4月6日(2017.4.6)

(51)Int.Cl.	F I	テーマコード（参考）
G06N 3/04 (2006.01)	G06N 3/04	

審査請求 未請求　予備審査請求 有　（全 41 頁）

(21)出願番号	特願2016-553381 (P2016-553381)
(86)(22)出願日	平成27年2月13日 (2015.2.13)
(85)翻訳文提出日	平成28年10月18日 (2016.10.18)
(86)国際出願番号	PCT/US2015/015917
(87)国際公開番号	WO2015/178977
(87)国際公開日	平成27年11月26日 (2015.11.26)
(31)優先権主張番号	61/943,155
(32)優先日	平成26年2月21日 (2014.2.21)
(33)優先権主張国	米国 (US)
(31)優先権主張番号	14/273,214
(32)優先日	平成26年5月8日 (2014.5.8)
(33)優先権主張国	米国 (US)

(71)出願人　595020643
クゥアルコム・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　100112807
弁理士　岡田　貴志

最終頁に続く

(54)【発明の名称】原位置ニューラルネットワークコプロセッシング

(57)【要約】
ニューラルネットワークにおいてコプロセッシングを実行する方法は、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングすることを含む。本方法はまた、第１の処理ノードでニューラルネットワークの一部を実行することを含む。さらに、本方法は、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返すことを含む。さらに、本方法は、第２の処理ノードでニューラルネットワークの一部を実行することを含む。
【選択図】図９

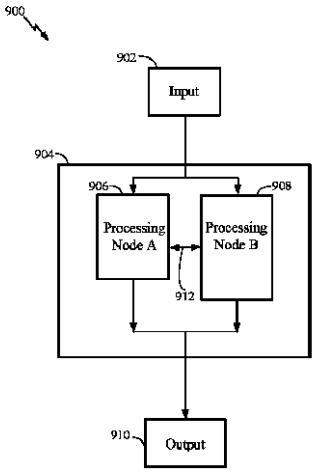

FIG. 9

【公報種別】特許法第１７条の２の規定による補正の掲載
【部門区分】第６部門第３区分
【発行日】平成30年3月1日(2018.3.1)

【公表番号】特表2017-509982(P2017-509982A)
【公表日】平成29年4月6日(2017.4.6)
【年通号数】公開・登録公報2017-014
【出願番号】特願2016-553381(P2016-553381)
【国際特許分類】
Ｇ０６Ｎ　3/04　(2006.01)
【ＦＩ】
Ｇ０６Ｎ　3/04

【手続補正書】
【提出日】平成30年1月18日(2018.1.18)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
人工ニューラルネットワークにおいてコプロセッシングを実行する、コンピュータ実装方法であって、
一定時間期間にわたって、前記ニューラルネットワークの一部を第２の処理ノードから第１の処理ノードにスワッピングすること、ここにおいて、前記第１の処理ノードおよび前記第２の処理ノードは、相互に前記ニューラルネットワークの機能的特徴の処理を包含するように構成され、前記第１の処理ノードは、強化学習を実装するように構成される、学習処理コアを備える、と、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行すること、ここにおいて、前記一部は、前記ニューラルネットワークに関する状態変数および接続性情報を含む、前記一部がそこから送られる前記ノードについての状態情報を備える、と
を備える、方法。
【請求項２】
前記第１の処理ノードは、第１のハードウェアコアに含まれ、前記第２の処理ノードは、第２のハードウェアコアに含まれ、前記第１のハードウェアコアは、前記第２のハードウェアコアとは別個である、
請求項１に記載の方法。
【請求項３】
前記学習処理コアは、前記第２の処理ノードよりも多くのリソースで構成される、
請求項１に記載の方法。
【請求項４】
前記第２の処理ノードは、前記ニューラルネットワークまたは前記一部を動作することに関連付けられる機能を実行するために構成された静的処理コアを備え、
スワッピングすることは、
前記静的処理コアの状態を前記学習処理コアにコピーすることと、
前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに

入力をルーティングすることと
を備え、
返すことは、
前記学習処理コアの状態を前記静的処理コアにコピーすることと、
前記静的処理コアに制御を返すことと
を備える、請求項１に記載の方法。
【請求項５】
前記人工ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、
請求項１に記載の方法。
【請求項６】
前記第１の処理ノードは、デバッギングコアを備える、
請求項１に記載の方法。
【請求項７】
人工ニューラルネットワークにおいてコプロセッシングを実行するための装置であって、
メモリと、
前記メモリに結合された少なくとも１つのプロセッサと
を備え、前記少なくとも１つのプロセッサは、
一定時間期間にわたって、前記ニューラルネットワークの一部を第２の処理ノードから第１の処理ノードにスワッピングすること、ここにおいて、前記第１の処理ノードおよび前記第２の処理ノードは、相互に前記ニューラルネットワークの機能的特徴の前記処理を包含するように構成され、前記第１の処理ノードは、強化学習を実装するように構成される、学習処理コアを備える、と、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行すること、ここにおいて、前記一部は、前記ニューラルネットワークに関する状態変数および接続性情報を含む、前記一部がそこから送られる前記ノードについての状態情報を備える、と
を行うように構成される、装置。
【請求項８】
前記第１の処理ノードは、第１のハードウェアコアに含まれ、前記第２の処理ノードは、第２のハードウェアコアに含まれ、前記第１のハードウェアコアは、前記第２のハードウェアコアとは別個である、
請求項７に記載の装置。
【請求項９】
前記学習処理コアは、前記第２の処理ノードよりも多くのリソースで構成される、
請求項７に記載の装置。
【請求項１０】
前記第２の処理ノードは、前記ニューラルネットワークまたは前記一部を動作することに関連付けられる機能を実行するために構成された静的処理コアを備え、前記少なくとも１つのプロセッサは、
前記静的処理コアの状態を前記学習処理コアにコピーすることと、
前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに入力をルーティングすることと、
前記学習処理コアの状態を前記静的処理コアにコピーすることと、
変更された静的処理コアに制御を返すことと
を行うようにさらに構成される、請求項７に記載の装置。
【請求項１１】

前記人工ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、
請求項７に記載の装置。
【請求項１２】
前記第１の処理ノードは、デバッギングコアを備える、
請求項７に記載の装置。
【請求項１３】
人工ニューラルネットワークにおいてコプロセッシングを実行するためのプログラムコードを符号化した非一時的コンピュータ可読媒体であって、前記プログラムコードは、プロセッサによって実行され、請求項１～請求項６に記載の方法のいずれかを実行するためのプログラムコードを備える、非一時的コンピュータ可読媒体。

【特許請求の範囲】
【請求項１】
ニューラルネットワークにおいてコプロセッシングを実行する方法であって、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングすることと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行することと
を備える、方法。
【請求項２】
前記第１の処理ノードは、別個のハードウェアコアを備える、
請求項１に記載の方法。
【請求項３】
前記第１の処理ノードは、学習処理コアを備える、
請求項１に記載の方法。
【請求項４】
前記学習処理コアは、前記第２の処理ノードよりも高いレベルのリソースで構成される、
請求項３に記載の方法。
【請求項５】
学習は、オフラインまたはオンラインで実装される、
請求項３に記載の方法。
【請求項６】
前記学習処理コアの入力および出力は、学習がオフラインで実装される場合、前記ニューラルネットワークの他のレイヤを備える、
請求項５に記載の方法。
【請求項７】
前記第１の処理ノードは、学習処理コアを備え、
前記第２の処理ノードは、静的処理コアを備え、
スワッピングすることは、
前記静的処理コアの状態を前記学習処理コアにコピーすることと、
前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに入力をルーティングすることと
を備え、
返すことは、
前記学習処理コアの状態を前記静的処理コアにコピーすることと、
変更された静的処理コアに制御を返すことと
を備える、請求項１に記載の方法。
【請求項８】
前記スワッピングすることは、前記第１の処理ノードから前記第２の処理ノードにリソースを割り振ることを備える、
請求項１に記載の方法。
【請求項９】
前記ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、
請求項１に記載の方法。
【請求項１０】
前記第１の処理ノードは、デバッギングコアを備える、
請求項１に記載の方法。
【請求項１１】

前記スワッピングすることは、システム性能がしきい値を下回る場合に発生する、
請求項１に記載の方法。
【請求項１２】
前記返すことは、システム性能がしきい値を上回る場合に発生する、
請求項１に記載の方法。
【請求項１３】
前記スワッピングすること、または返すことは、電力がシステムに適用されると発生する、
請求項１に記載の方法。
【請求項１４】
ニューラルネットワークにおいてコプロセッシングを実行するための装置であって、
メモリと、
前記メモリに結合された少なくとも１つのプロセッサと
を備え、前記少なくとも１つのプロセッサは、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングすることと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行することと
を行うように構成される、装置。
【請求項１５】
前記第１の処理ノードは、別個のハードウェアコアを備える、
請求項１４に記載の装置。
【請求項１６】
前記第１の処理ノードは、学習処理コアを備える、
請求項１４に記載の装置。
【請求項１７】
前記学習処理コアは、前記第２の処理ノードよりも高いレベルのリソースで構成される、
請求項１６に記載の装置。
【請求項１８】
学習は、オフラインまたはオンラインで実装される、
請求項１６に記載の装置。
【請求項１９】
前記学習処理コアの入力および出力は、学習がオフラインで実装される場合、前記ニューラルネットワークの他のレイヤを備える、
請求項１８に記載の装置。
【請求項２０】
前記第１の処理ノードは、学習処理コアを備え、前記第２の処理ノードは、静的処理コアを備え、前記少なくとも１つのプロセッサは、
前記静的処理コアの状態を前記学習処理コアにコピーすることと、
前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに入力をルーティングすることと、
前記学習処理コアの状態を前記静的処理コアにコピーすることと、
変更された静的処理コアに制御を返すことと
を行うようにさらに構成される、請求項１４に記載の装置。
【請求項２１】
前記少なくとも１つのプロセッサは、前記第１の処理ノードから前記第２の処理ノードにリソースを割り振ることを行うようにさらに構成される、

請求項１４に記載の装置。
【請求項２２】
前記ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、
請求項１４に記載の装置。
【請求項２３】
前記第１の処理ノードは、デバッギングコアを備える、
請求項１４に記載の装置。
【請求項２４】
前記少なくとも１つのプロセッサは、システム性能がしきい値を下回る場合に、前記ニューラルネットワークの前記一部を前記第１の処理ノードにスワッピングするようにさらに構成される、
請求項１４に記載の装置。
【請求項２５】
前記少なくとも１つのプロセッサは、システム性能がしきい値を上回る場合に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すようにさらに構成される、
請求項１４に記載の装置。
【請求項２６】
前記少なくとも１つのプロセッサは、電力がシステムに適用されると、前記ニューラルネットワークの前記一部を前記第１の処理ノードにスワッピングする、または前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すようにさらに構成される、
請求項１４に記載の装置。
【請求項２７】
ニューラルネットワークにおいてコプロセッシングを実行するための装置であって、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングするための手段と、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行するための手段と、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すための手段と、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行するための手段と
を備える、装置。
【請求項２８】
ニューラルネットワークにおいてコプロセッシングを実行するためのコンピュータプログラム製品であって、
プログラムコードを符号化した非一時的コンピュータ可読媒体を備え、前記プログラムコードは、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングするためのプログラムコードと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行するためのプログラムコードと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すためのプログラムコードと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行するためのプログラムコードと
を備える、コンピュータプログラム製品。
【発明の詳細な説明】
【関連出願の相互参照】
【０００１】

[0001]本出願は、２０１４年２月２１日に出願された「ＩＮ　ＳＩＴＵ　ＮＥＵＲＡＬ　ＮＥＴＷＯＲＫ　ＣＯ－ＰＲＯＣＥＳＳＩＮＧ」と題する米国仮特許出願第６１／９４３，１５５号の利益を主張し、その開示は、参照によりその全体が本明細書に明示的に組み込まれる。

【技術分野】

【０００２】

[0002]本開示のいくつかの態様は、一般にニューラルシステムエンジニアリングに関し、より詳細には、原位置ニューラルネットワークコプロセッシング（in situ neural network co-processing）のためのシステムおよび方法に関する。

【背景技術】

【０００３】

[0003]人工ニューロン（すなわち、ニューロンモデル）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は観測から関数を推測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関数の設計を煩わしくする用途において、特に有用である。

【発明の概要】

【０００４】

[0004]本開示のある態様では、ニューラルネットワークにおいてコプロセッシングを実行する方法が開示される。本方法は、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングすることを含む。本方法はまた、第１の処理ノードでニューラルネットワークの一部を実行することを含む。さらに、本方法は、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返すことを含む。本方法は、第２の処理ノードでニューラルネットワークの一部を実行することをさらに含む。

【０００５】

[0005]本開示の別の態様では、ニューラルネットワーク内でコプロセッシングを実行するための装置が開示される。本装置は、メモリと、メモリに結合された少なくとも１つのプロセッサとを含む。本プロセッサは、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングするように構成される。本プロセッサはまた、第１の処理ノードでニューラルネットワークの一部を実行するように構成される。さらに、本プロセッサは、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返すように構成される。本プロセッサは、第２の処理ノードでニューラルネットワークの一部を実行するようにさらに構成される。

【０００６】

[0006]本開示の別の態様では、ニューラルネットワーク内でコプロセッシングを実行するための装置が開示される。本装置は、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングするための手段を有する。本装置はまた、第１の処理ノードでニューラルネットワークの一部を実行するための手段を有する。さらに、本装置は、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返すための手段を有する。本装置は、第２の処理ノードでニューラルネットワークの一部を実行するための手段をさらに有する。

【０００７】

[0007]本開示の別の態様では、ニューラルネットワーク内でコプロセッシングを実行するためのコンピュータプログラム製品が開示される。本コンピュータプログラム製品は、プログラムコードを符号化した非一時的コンピュータ可読媒体を含む。本プログラムコードは、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにス

ワッピングするためのプログラムコードを含む。本プログラムコードはまた、第１の処理ノードでニューラルネットワークの一部を実行するためのプログラムコードを含む。さらに、本プログラムコードは、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返すためのプログラムコードを含む。本プログラムコードは、第２の処理ノードでニューラルネットワークの一部を実行するためのプログラムコードをさらに含む。

【０００８】

[0008]これは、以下の詳細な説明がより良く理解され得るために、本開示の特徴および技術的利点をかなり広く概説した。本開示の追加の特徴および利点は、以下で説明される。この開示は、本開示と同じ目的を実行するための他の構造を修正または設計するための基礎として容易に変更され得ることが、当業者によって理解されるべきである。また、添付の特許請求の範囲に記載されるように、そのような等価な構成は本開示の教示から逸脱しないことが、当業者によって理解されるべきである。本開示の特徴と考えられる新規な特徴は、その構成と動作の方法との両方に関して、さらなる目的および利点とともに、添付の図面と関連して考慮されるとき以下の説明からより良く理解されるであろう。しかしながら、図面の各々は単に例示および説明の目的のために提供されているにすぎず、本開示の制限の定義として意図されていないことが、明確には理解されるべきである。

【図面の簡単な説明】

【０００９】

[0009]本開示の特徴、性質、および利点は、同様の参照文字が全体を通して相応して識別する図面を考慮した場合、以下に示される詳細な説明から、より明らかになるだろう。

【図１】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】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）の処理ユニット（ニューロン）の一例を示す図。

【図３】本開示のいくつかの態様によるスパイクタイミング依存可塑性（ＳＴＤＰ）曲線の一例を示す図。

【図４】本開示のいくつかの態様による、ニューロンモデルの挙動を定義するための正レジームおよび負レジームの一例を示す図。

【図５】本開示のある態様による、汎用プロセッサを使用してニューラルネットワークを設計することの例示的な実装形態を示す図。

【図６】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ニューラルネットワークを設計する例示的な実装形態を示す図。

【図７】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ニューラルネットワークを設計する例示的な実装形態を示す図。

【図８】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。

【図９】本開示の態様による、ニューラルネットワークの例示的なアーキテクチャを示すブロック図。

【図１０Ａ】本開示の態様による、ニューラルネットワーク内の原位置コプロセッシングを示す例示的なブロック図。

【図１０Ｂ】本開示の態様による、ニューラルネットワーク内の原位置コプロセッシングを示す例示的なブロック図。

【図１０Ｃ】本開示の態様による、ニューラルネットワーク内の原位置コプロセッシングを示す例示的なブロック図。

【図１０Ｄ】本開示の態様による、ニューラルネットワーク内の原位置コプロセッシングを示す例示的なブロック図。

【図１０Ｅ】本開示の態様による、ニューラルネットワーク内の原位置コプロセッシングを示す例示的なブロック図。

【図１０Ｆ】本開示の態様による、ニューラルネットワーク内の原位置コプロセッシングを示す例示的なブロック図。

【図１１】本開示の態様による、ニューラルネットワーク内でコプロセッシングを実行す

るための方法を示すブロック図。
【図１２】本開示の態様による、ニューラルネットワーク内でコプロセッシングを実行するための方法を示すブロック図。
【発明を実施するための形態】
【００１０】
[0021]添付の図面に関連して以下に示される詳細な説明は、様々な構成の説明として意図されたものであり、本明細書において説明される概念が実現され得る唯一の構成を表すことを意図されるものではない。詳細な説明は、様々な概念の完全な理解を提供する目的で、具体的な詳細を含む。しかしながら、これらの概念がこれらの具体的な詳細なしで実施され得ることは、当業者にとっては明らかであろう。いくつかの事例では、よく知られている構造および構成要素が、そのような概念を曖昧にするのを避けるために、ブロック図形式で示される。
【００１１】
[0022]本教示に基づいて、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、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、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。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。
【００１２】
[0023]「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】
[0024]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。
例示的なニューラルシステム、トレーニングおよび動作
【００１４】
[0025]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人工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を有し得る。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、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レイヤ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レイヤのニューロンに戻る形で結合し得る。
【００１５】
[0026]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のニューロンの入力電流を表し得る。この電流は、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。膜電位がそれ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。いくつかのモデリング手法では、ニューロンは、信号をニューロンの次のレベルに継続的に転送し得る。この信号は、典型的には膜電位の関数である。そのような挙動は、以下で説明するものなどの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。

【００１６】

[0027]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、一時的な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実施形態（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得る。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、および他の１つまたは複数のスパイクに対するスパイクの時間によって決定され得る。以下で説明するように、スパイクの重要性は、ニューロン間の接続に適用される重みによって決定され得る。

【００１７】

[0028]図１に示されるように、ニューロンのあるレベルから別のレベルへのスパイクの転送は、シナプス結合（または、単純に「シナプス」）１０４のネットワークを介して達成され得る。シナプス１０４に関して、レベル１０２のニューロンはシナプス前ニューロンと考えられ得、レベル１０６のニューロンはシナプス後ニューロンと考えられ得る。シナプス１０４は、レベル１０２のニューロンから出力信号（すなわち、スパイク）を受信して、調整可能なシナプスの重み

【００１８】

【数１】

$$w_1^{(i,i+1)},\ldots,\ w_P^{(i,i+1)}$$

【００１９】

に応じてそれらの信号をスケーリングすることができ、上式で、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り、ｉはニューロンレベルの指標である。図１の例では、ｉはニューロンレベル１０２を表し、ｉ＋１は、ニューロンレベル１０６を表す。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、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。

【００２０】

[0029]生物学的シナプスは、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、ニューロン信号を増幅する役目を果たすことができる。興奮性信号は、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の時間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、自然にアクティブなニューロンに対して強力な制御を行うことができる。自然にアクティブなニューロンは、たとえば、それ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、一般にスカルプチャリングと呼ばれる、ニューロンの発火のパターンを形成することができる。様々なシナプス１０４は、望まれる挙動に応じて、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。

【００２１】

[0030]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、および似ているなど、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。

【００２２】

[0031]一態様では、キャパシタは、ニューロン回路の電流積分デバイスとして除去され得、その代わりにより小さいメモリスタ（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、ならびに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において適用され得る。さらに、シナプス１０４の各々は、メモリスタ要素に基づいて実装され得、シナプス重みの変化は、メモリスタ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タを用いると、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、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より実用的になり得る。

【００２３】

[0032]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、ニューロン間の結合の強さを制御し得る、シナプス結合の重みに依存し得る。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。一態様では、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。

【００２４】

[0033]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、ニューロンまたはニューロン回路）２０２の例示的な図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、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、複数の入力信号２０４_{1}～２０４_{N}を受信し得る。入力信号は、電流、コンダクタンス、電圧、実数値および／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}～２０６_{N}（w_{1}～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。

【００２５】

[0034]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、コンダクタンス、電圧、実数値および／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であり得る。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。

【００２６】

[0035]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、それの入力接続および出力接続は、シナプス回路をもつ電気接続によってエミュレートされ得る。処理ユニット２０２ならびにそれ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２０２はまた、電気回路によってエミュレートされ得るが、それ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。一態様では、計算ネットワーク中の処理ユニット２０２はアナログ電気回路であり得る。別の態様では、処理ユニット２０２はデジタル電気回路であり得る。さらに別の態様では、処理ユニット２０２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であり得る。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含み得る。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。

【００２７】

[0036]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

【００２８】

【数２】

$$w_1^{(i,i+1)},\ldots,\ w_P^{(i,i+1)}$$

【００２９】

および／または図２の重み２０６$_1$～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。学習ルールの例は、これに限定されないが、スパイクタイミング依存可塑性（ＳＴＤＰ）学習ルール、Ｈｅｂｂ則、Ｏｊａ則、Ｂｉｅｎｅｎｓｔｏｃｋ－Ｃｏｐｐｅｒ－Ｍｕｎｒｏ（ＢＣＭ）則等を含むことを当業者は理解するだろう。いくつかの態様では、重みは、２つの値のうちの１つに安定または収束し得る（すなわち、重みの双峰分布）。この効果が利用されて、シナプス重みごと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取りおよび書込みの速度を上げ、シナプスメモリの電力および／またはプロセッサ消費量を低減し得る。

シナプスタイプ

【００３０】

[0037]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）、可塑性シナプス（重みが変化し得る）、構造遅延可塑性シナプス（重みおよび遅延が変化し得る）、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）、およびそれの変形（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）であり得る。複数のタイプの利点は、処理が再分割され得ることである。たとえば、非塑性シナプスは、可塑性機能の実行を含まない場合がある（またはそのような機能が完了するのを待つ）。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、シナプスのタイプについての関連する表、式、またはパラメータにアクセスする。

【００３１】

[0038]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）ｓ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）の直接関数であり得ても実行され得る。代替的に、構造可塑性は、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係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、または重みが最大値になるのではなく、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。

シナプス可塑性の決定

【００３２】

[0039]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（ＳＴＤＰ）、非シナプス可塑性、アクティビティ依存可塑性、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。

【００３３】

[0040]ＳＴＤＰは、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。ＳＴＤＰ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。一方、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わずかなレベルまで低減されるまで、このプロセスは続く。

【００３４】

[0041]ニューロンは一般に出力スパイクを、それの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な累積がある）ときに生成するので、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指示を提供する入力は結局、ニューロンへの最終入力となる。

【００３５】

[0042]ＳＴＤＰ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、$t = t_{post} - t_{pre}$）に応じて、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。ＳＴＤＰ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。

【００３６】

[0043]ＳＴＤＰプロセスでは、経時的なシナプス重みの変化は通常、以下の式によって与えられるように、指数関数的減衰を使用して達成され得る。

【００３７】

【数３】

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

【００３８】

ここで、k_{+}および$k_{-}\tau_{sign(\Delta t)}$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。

【００３９】

[0044]図３は、ＳＴＤＰ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に応じたシナプス重み変化の例示的な図３００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３００の部分３０２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ＬＴＰと呼ばれ得る。グラフ部分３０２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、ＬＴＰの量がほぼ指数関数的に減少し得ることが観測され得る。グラフ３００の部分３０４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ＬＴＤをもたらし得る。

【００４０】

[0045]図３のグラフ３００に示すように、ＳＴＤＰグラフのＬＴＰ（原因）部分３０２に負のオフセットμが適用され得る。ｘ軸の交差３０６のポイント（ｙ＝０）は、レイヤｉ－１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、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の形態である入力）の場合、オフセット値μは、フレーム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）が経時的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と相関したまたは特定の時間フレームに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）ように、ＳＴＤＰ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。

ニューロンモデルおよび演算

【００４１】

[0046]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：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ィンドウを有し得る。最終的に、計算上魅力的であるために、良いニューロンモデルは、連続時間に閉形式解と、ニアアトラクター（near attractor）と鞍点とを含む安定した挙動とを有し得る。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でき、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が可能なニューロンモデルである。

【００４２】

[0047]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。

【００４３】

[0048]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）によるスパイキングリーキー積分発火ニューロンとしてモデル化され得る。

【００４４】

【数４】

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

【００４５】

ここでαおよびβ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パラメータ、$w_{m,n}$は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$\Delta t_{m,n}$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。

【００４６】

[0049]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、電圧および復元のための１対の微分方程式、すなわち、

【００４７】

【数５】

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

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

【００４８】

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のしきい値下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。

Ｈｕｎｚｉｎｇｅｒ　Ｃｏｌｄモデル

【００４９】

[0050]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最小二重レジームスパイキング線形動的モデルである。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、時間定数（および結合）はレジームに依存し得る。しきい値下レジームでは、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。しきい値上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。

【００５０】

[0051]図４に示すように、モデル４００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、負のレジーム（ｎｅｇａｔｉｖｅ　ｒｅｇｉｍｅ）４０２（ｌｅａｋｙ－ｉｎｔｅｇｒａｔｅ－ａｎｄ－ｆｉｒｅ（ＬＩＦ）ニューロンモデルと混同されないように、交換可能にＬＩＦレジームとも呼ばれる）、および正のレジーム（ｐｏｓｉｔｉｖｅ　ｒｅｇｉｍｅ）４０４（ａｎｔｉ－ｌｅａｋｙ－ｉｎｔｅｇｒａｔｅ－ａｎｄ－ｆｉｒｅ（ＡＬＩＦ）ニューロンモデルと混同されないように、交換可能に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ける静止（v_-）の傾向がある。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しきい値下挙動とを示す。正レジーム４０４では、状態はスパイキング事象（v_s）の傾向がある。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。

【００５１】

[0052]線形二重レジーム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。

【００５２】

【数６】

$$\tau_\rho \frac{dv}{dt} = v + q_\rho \qquad (5)$$

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

【００５３】

ここでq_ρおよびrは、結合のための線形変換変数である。

【００５４】

[0053]シンボルρ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ρを置き換える慣例がある。

【００５５】

[0054]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。

【００５６】

[0055]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_+とを含む。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。

【００５７】

[0056]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、

【００５８】

【数７】

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

$$r = \delta(v + \varepsilon) \qquad (8)$$

【００５９】
であり、δ、ε、βおよびν$_-$、ν$_+$はパラメータである。ν$_\rho$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ν$_-$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ν$_-$に減衰する。パラメータν$_+$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ν$_+$から離れる傾向となる。
【００６０】
[0057]νおよびｕのためのヌルクラインは、それぞれ変換変数ｑ$_\rho$およびｒの負によって与えられる。パラメータδは，ｕヌルクラインの傾きを制御するスケール係数である。パラメータεは通常、－ν$_-$に等しく設定される。パラメータβは、両方のレジームにおいてνヌルクラインの傾きを制御する抵抗値である。τ$_\rho$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ンの傾きを制御する。
【００６１】
[0058]モデルは、電圧νが値ν$_s$に達したときにスパイクするように定義され得る。続いて、状態は（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得る。
【００６２】
【数８】

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

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

【００６３】
ここで、
【００６４】
【数９】

$$\hat{v}_-$$

【００６５】
およびΔｕはパラメータである。リセット電圧
【００６６】
【数１０】

$$\hat{v}_-$$

【００６７】
は通常、ν$_-$にセットされる。
【００６８】
[0059]瞬時結合の原理によって、状態について（また、単一の指数項による）だけでは

なく、特定の状態に到達するための時間についても、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。

【００６９】

【数１１】

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

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

【００７０】

[0060]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。

【００７１】

[0061]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。

【００７２】

【数１２】

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

【００７３】

[0062]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。

【００７４】

【数１３】

$$\Delta t_s=\begin{cases}\tau_+\log\dfrac{v_s+q_+}{v+q_+} & \text{もし、}\ v>\hat{v}_+\ \text{であれば}\\ \infty & \text{そうでなければ}\end{cases} \qquad (14)$$

【００７５】

ここで、

【００７６】

【数１４】

$$\hat{v}_+$$

【００７７】

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。

【００７８】

[0063]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ρ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。

【００７９】

[0064]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、１ｍｓ）を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含むとは限らない。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。

原位置ニューラルネットワークコプロセッシング

【００８０】

[0065]本開示の態様は、ニューラルネットワークシミュレータを対象とし、より詳細には、原位置ニューラルコプロセッシングを対象とする。

【００８１】

[0066]一般的に、ニューラルネットワークシミュレータは、柔軟性と性能（たとえば、シミュレータの電力）との間でトレードオフを行う。たとえば、設計者は、しばしば、学習を可能にするチップを作成するか、より高速に実行するチップを作成するか、または消費電力がより少ないチップを作成するかを決定しなければならない場合がある。したがって、学習がオフラインで実装されている場合、学習をサポートしないシミュレータ上に実装されるトレーニングされたニューラルネットワークは、学習をサポートするシミュレータ上に実装されたネットワークと同じ入力を経験しない場合がある。これは、学習に関連付けられるネットワークへのリアルタイムの変化が、ニューラルネットワークの環境に影響を与える（ニューラルネットワークに関連付けられるエフェクタを介して）可能性があるためであり得、それは、今度は環境を表現してネットワークへの入力を提供するセンサを介して、ニューラルネットワークへの入力に影響を与え得る。ニューラルネットワークの環境はまた、フィードバック接続または非ローカル信号を通じてニューラルネットワークへの変化を同様に引き起こす可能性がある下流ニューラルネットワークを示す可能性がある。

【００８２】

[0067]本開示の態様によれば、複数のシミュレーションプラットフォームは、シミュレータの通常動作中にトレードオフが行われ得るように組み合わせられ得る。たとえば、学習を利用しないシミュレーションは、この機能を提供しないシミュレーションプラットフォーム上で実行され得る。これは、たとえば、第２のシミュレーションプラットフォームが消費する電力が、第１のシミュレーションプラットフォームが消費する電力よりも少ない場合に有益であり得る。

【００８３】

[0068]本開示のいくつかの態様では、相互にスワッピングし得るニューラルコプロセッサが提供され得る。いくつかの態様では、ニューラルコプロセッサは、異なる機能を備えたニューラル処理ユニットまたはノードであり得る。たとえば、あるニューラル処理ノードは学習動作を実行するように構成され得、他の処理コアは静的重みで構成される。

【００８４】

[0069]１つの例示的な態様では、より多くの機能を備えたコア（すなわち、コアより多くの機能（たとえば、メモリまたはプロセッサを有するコア）は、より少ない機能を備えたコア（すなわち、より少ない機能を有するコア）の機能を引き継ぐ、または包含するこ

とができる。機能の包含は、処理ノードの「ホットスワップ（ｈｏｔ　ｓｗａｐ）」の形で行われ得る。この「ホットスワップ」を行うことにより、柔軟性と性能が向上され得る。

【００８５】

[0070]図５は、本開示の特定の態様による、汎用プロセッサ５０２を使用して、ニューラルネットワークにおける上述の実行しているコプロセッシングの例示的な実装形態５００を示す。変数（ニューラル信号）、シナプス重み、計算ネットワーク（ニューラルネットワーク）に関連付けられるシステムパラメータ、遅延、周波数ビン情報、性能メトリック、およびシステム状態情報は、メモリブロック５０４に記憶され得、汎用プロセッサ５０２で実行される命令はプログラムメモリ５０６からロードされ得る。本開示のある態様では、汎用プロセッサ５０２にロードされる命令は、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングして、第１の処理ノードでニューラルネットワークの一部を実行して、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返して、および／または、第２の処理ノードでニューラルネットワークの一部を実行するためのコードを備え得る。

【００８６】

[0071]図６は、本開示のいくつかの態様による、メモリ６０２が相互接続ネットワーク６０４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）６０６とインターフェースされ得る、ニューラルネットワーク内の上述したコプロセッシングの実行の例示的な実装形態６００を示している。計算ネットワーク（ニューラルネットワーク）、遅延、周波数ビン情報、性能メトリック、およびシステム状態情報に関連付けられる、変数（ニューラル信号）、シナプス重み、システムパラメータはメモリ６０２に記憶され得、相互接続ネットワーク６０４の接続を介してメモリ６０２から各処理ユニット（ニューラルプロセッサ）６０６にロードされ得る。本開示のある態様では、処理ユニット６０６は、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングして、第１の処理ノードでニューラルネットワークの一部を実行して、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返して、および／または、第２の処理ノードでニューラルネットワークの一部を実行するように構成され得る。

【００８７】

[0072]図７は、ニューラルネットワーク内の上述したコプロセッシングの実行の例示的な実装形態７００を示している。図７に示されるように、１つのメモリバンク７０２は、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７０４に直接インターフェースされ得る。各メモリバンク７０２は、対応する処理ユニット（ニューラルプロセッサ）７０４、遅延、周波数ビン情報、性能メトリック、およびシステム状態情報に関連付けられる変数（ニューラル信号）、シナプス重み、および／またはシナプスパラメータを記憶し得る。本開示のある態様では、処理ユニット７０４は、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングして、第１の処理ノードでニューラルネットワークの一部を実行して、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返して、および／または、第２の処理ノードでニューラルネットワークの一部を実行するように構成され得る。

【００８８】

[0073]図８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８００の例示的な実装形態を示す。図８に示すように、ニューラルネットワーク８００は、本明細書に記載される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８０２を有することができる。各ローカル処理ユニット８０２は、ニューラルネットワークのパラメータを記憶する、ローカルステートメモリ８０４およびローカルパラメータメモリ８０６を備え得る。また、ローカル処理ユニット８０２は、ローカルモデルプログラムを記憶するためのローカル（ニューロン）モデルプログラム（ＬＭＰ）メモリ８０８、ローカル学習プログラムを記憶するためのローカル学習プログラム（ＬＬＰ）メモリ８１０、およびローカ

ル接続メモリ８１２を有し得る。さらに、図８に示されるように、各ローカル処理ユニット８０２は、ローカル処理ユニットのローカルメモリの構成を提供するための構成処理ユニット８１４と、またローカル処理ユニット８０２間のルーティングを提供するルーティング接続処理ユニット８１６とインターフェースされ得る。

【００８９】

[0074]一構成では、ニューロンモデルは、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングして、第１の処理ノードでニューラルネットワークの一部を実行して、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返して、および／または、第２の処理ノードでニューラルネットワークの一部を実行するために構成される。ニューロンモデルは、スワッピング手段と、第１の処理ノードでニューラルネットワークの一部を実行するための手段と、返す手段と、第２の処理ノードでニューラルネットワークの一部を実行するための手段とを含む。一態様では、スワッピング手段、第１の処理ノードでニューラルネットワークの一部を実行するための手段、返す手段、および／または、第２の処理ノードでニューラルネットワークの一部を実行するための手段は、記載された機能を実行するように構成された汎用プロセッサ５０２、プログラムメモリ５０６、メモリブロック５０４、メモリ６０２、相互接続ネットワーク６０４、処理ユニット６０６、処理ユニット７０４、ローカル処理ユニット８０２、およびまたはルーティング接続処理ユニット８１６であり得る。別の構成では、上述の手段は、上記の手段によって記載された機能を実行するように構成された任意のモジュールまたは任意の装置であり得る。

【００９０】

[0075]別の構成では、ニューロンモデルは、まず第１の処理コアでニューラルネットワークの一部を実行することによって、および／または、さらなる実行のためにニューラルネットワークの一部を第２の処理コアに移動させることによって、オフライン学習を共同設置するために構成される。ニューロンモデルは、共同設置手段と移動手段とを含む。一態様では、共同設置手段および／または移動手段は、記載された機能を実行するように構成された汎用プロセッサ５０２、プログラムメモリ５０６、メモリブロック５０４、メモリ６０２、相互接続ネットワーク６０４、処理ユニット６０６、処理ユニット７０４、ローカル処理ユニット８０２、およびまたはルーティング接続処理ユニット８１６であり得る。別の構成では、上述の手段は、上記の手段によって記載された機能を実行するように構成された任意のモジュールまたは任意の装置であり得る。

【００９１】

[0076]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８０２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て、決定されたパラメータがさらに適応され、同調され、更新されるにつれて、所望の機能的特徴に向けて１つまたは複数の機能的特徴を開発するように構成され得る。

【００９２】

[0077]図９は、本開示の態様による、ニューラルネットワークの例示的なアーキテクチャ９００を示すブロック図である。アーキテクチャ９００は、処理ノードＡ９０６と処理ノードＢ９０８とを含み得るコプロセッサ９０４を備え得る。いくつかの態様では、処理ノードＡ９０６と処理ノードＢ９０８とは、同じハードウェアコア内に含まれ得る。しかしながら、これは単なる例示であり、処理ノードＡ９０６と処理ノードＢ９０８とは、代替で、別個のハードウェアコアにおいて提供され得る。

【００９３】

[0078]処理ノードＡ９０６と処理ノードＢ９０８とは、異なるように構成され得る。すなわち、いくつかの態様では、処理ノードＡ９０６と処理ノードＢ９０８とは、ニューラルネットワークの機能的特徴を効率的に実行することに適した異なる構成を有し得る。いくつかの構成では、処理ノードＡ９０６は、処理ノードＢよりも大きいリソースで構成され得る。たとえば、処理ノードＡ９０６は、処理ノードＢ９０８よりも速いおよび／また

は高い処理能力（たとえば、複数のプロセッサ、またはより速い処理速度）で構成され得る。第２の例では、処理ノードＢ９０８は、より多数および／またはより速いメモリで構成され得る。

【００９４】

[0079]処理ノードＡ９０６と処理ノードＢ９０８とは、入力ノード９０２を介して入力を受信するように構成され得る。処理ノードＡ９０６と処理ノードＢ９０８とはまた、出力ノード９１０に出力を供給するように構成され得る。入力９０２と出力９１０とは、センサ、アクチュエータ、および他の入力／出力デバイスを備え得る。

【００９５】

[0080]さらに、処理ノードＡ９０６と処理ノードＢ９０８とは、処理ノード間でニューラルネットワークの機能的特徴を実行することのホットスワッピングを可能にするために、相互に通信可能に結合され得る。すなわち、実行時に、より多くの機能を持つ処理ノード（たとえば、９０６、９０８）が、より少ない特徴を有するコアの機能の処理を包含または引き継ぎ得る。

【００９６】

[0081]いくつかの態様では、処理ノードＡ９０６の状態がコピーされて、通信経路９１２または任意の他の通信経路を介して処理ノードＢ９０８に供給され得る。処理ノードＡ９０６の状態は、たとえば、状態変数、接続性情報、および他の状態情報を含み得る。

【００９７】

[0082]処理ノードＢ９０８のリソースは、処理ノードＡ９０６からニューラルネットワークの機能的特徴の処理を引き継ぐために割り振られ得る。さらに、入力ノード９０２を介して供給された入力は、処理ノードＢ９０８にルーティングされ得る。処理ノードＡ９０６からの状態情報および入力に基づいて、処理ノードＢ９０８は、以前は処理ノードＡ９０６によって処理されていたニューラルネットワークの機能的特徴を処理することを引き継ぎ得る。

【００９８】

[0083]いくつかの態様では、処理ノードＡ９０６は、入力ノード９０２を介して、処理ノードＢ９０８に供給されたものと同じ入力を受信することを継続し得る。したがって、整合性チェックを提供するために、処理ノードＡ９０６の出力が処理ノードＢ９０８の出力と比較され得る。一例では、処理ノードＢ９０８は、処理ノードＡ９０６内の欠陥またはバグを識別して減少させるためのデバッギングコアとして構成され得る。本開示の他の態様では、処理ノードＡ９０６は、ニューラルネットワークの他の機能的特徴を処理し得る。

【００９９】

[0084]処理ノードＢ９０８は、あらかじめ定められた時間期間にわたって、または、いくつかの態様では、特定のタスクまたはタスクのセットの完了まで、処理ノードＡ９０６から包含されたニューラルネットワークの一部の処理を継続し得る。たとえば、処理ノードＢ９０８は、学習を実装するように構成され得、また、学習が達成されるまで処理ノードＡ９０６から包含されたニューラルネットワークの一部を処理することを継続し得る。別の例では、処理ノードＢ９０８は、スパイクタイミング依存可塑性を実装するように構成され得る。したがって、処理ノードＢは、受信された状態情報を処理して、状態情報の更新（たとえば、重み更新）が決定されるまで可塑性ルールを適用し得る。

【０１００】

[0085]いくつかの態様では、より多くの機能を備えた処理ノード（たとえば、９０６、９０８）は、システム性能メトリックに基づいて処理を引き継ぎ得る。たとえば、より多くの機能を備えた処理ノードは、より少ない機能を備えた処理ノードのシステム性能がしきい値レベルを下回る場合、処理を包含し得る。他の態様では、電力がシステムに適用されるとスワッピングが実行され得る。もちろん、これらは単に例示的な基礎であり、他のシステムおよびネットワーク性能メトリックは、より少ない機能を備えた処理ノードから、より多くの機能を備えた処理ノードに、スワッピング処理のための基礎を提供し得る。

【０１０１】
[0086]タスクが完了すると、または時間期間が満了すると、処理ノードＢ９０８の状態がコピーされて、変更されたコアとして処理ノードＡ９０６に供給され得る。いくつかの態様では、ニューラルネットワークの一部を返すことは、システム性能メトリックに基づいて実行され得る。たとえば、システム性能がしきい値を上回る場合、処理ノードＢ９０８の状態がコピーされて、処理ノードＡ９０６に供給され得る。第２の例では、返すことは、電力がシステム（たとえば、プラグインシステム）に適用されると発生し得る。いくつかの態様では、入力ノード９０２を介して提供される入力は、処理ノードＢ９０８からの状態情報を含む変更されたコアを使用してニューラルネットワークの機能的特徴を処理することを再開するために処理ノードＡ９０６にルーティングされ得る。

【０１０２】
[0087]図１０Ａ～図１０Ｆは、本開示の態様による、ニューラルネットワークにおける原位置コプロセッシングを示す例示的なブロック図１０００である。例示的なブロック図の各々は、静的コア１００８と学習コア１００６とを含むコプロセッサ１００４を示す。静的コア１００８は、ニューラルネットワークまたはその一部を動作することに関連付けられる機能を実行するための静的重みで構成され得る。学習コア１００６は、学習を実装して、学習動作を実行するように構成され得る。たとえば、いくつかの態様では、学習コア１００６は、強化学習または他の学習モデルを実装するように構成され得る。

【０１０３】
[0088]いくつかの態様では、学習コア１００６は、静的コア１００８よりも大きいリソースで構成され得る。たとえば、学習コア１００６は、静的コア１００８よりも速いおよび／または多数の処理能力（たとえば、複数のプロセッサ、またはより速い処理速度）で構成され得る。別の例では、学習コア１００６は、静的コア１００８とは異なるメモリリソース（たとえば、より多数および／またはより速いメモリ）で構成され得る。異なるタイプのメモリリソースは、たとえば、パラメータ（たとえば、重み）に関してより高い（または、より低い）精度を可能にしてもよく、スパイク履歴をキャプチャするためのより多くのリソースを提供してもよく、学習ルールへのアクセス、ならびにスパイクタイミング依存可塑性および／またはビット割振りの実装を可能にしてもよい。もちろん、これらの処理および性能関連機能は単なる例示であり、他の処理および性能関連機能または強化は、学習コア１００６および静的コア１００８に異なるように含まれ得る。

【０１０４】
[0089]図１０Ａ～図１０Ｆに含まれるブロック図の各々は１つだけの静的コア１００８および学習コア１００６を示すが、これは単なる例示であり、説明を容易にするためである。代わりに、たとえば設計効率の目的のために、任意の数の静的コア１００８および学習コア１００６が含まれ得る。さらに、静的コア１００８および学習コア１００６は、同一の処理コア内に含まれてもよく、代替で、別個の処理コアにおいて提供されてもよい。

【０１０５】
[0090]静的コア１００８および学習コア１００６は、入力ノード１００２を介して入力を選択的に受信して、出力ノード１０１０に出力を供給し得る。いくつかの態様では、静的コア１００８と学習コア１００６との両方は、入力ノード１００２を介して入力を受信し得る。同様に、静的コア１００８と学習コア１００６との両方は、整合性チェックまたは処理検証を可能にするために、出力を出力ノード１０１０に供給し得る。

【０１０６】
[0091]図１０Ａで、入力ノード１００２からの入力が静的コア１００８に提供されるが、学習コア１００６には提供されない。この例示的な態様では、ニューラルネットワークの動作は、静的コア１００８を介して実行のために合理化され得る。いくつかの態様では、学習は実装され得ない。

【０１０７】
[0092]図１０Ｂで、静的コア１００８の状態情報がコピーされて、通信経路１０１２を介して学習コア１００６に提供され得る。状態情報は、たとえば、ニューロン状態変数、

シナプス状態情報、接続性情報（たとえば、図または表）、および重み情報を含み得る。
【０１０８】
[0093]図１０Ｃで、入力ノード１００２を介する入力が学習コア１００６にルーティングされ得る。いくつかの態様では、入力は学習コア１００６だけに提供され得る。もちろん、入力は、代替で学習コア１００６と静的コア１００８との両方に提供され得る。この構成では、たとえば、検証技法は、静的コア１００８からの出力と学習コア１００６からの出力とが一致している（たとえば、同一である）ことを保証するために実行され得る。
【０１０９】
[0094]図１０Ｄで、学習コア１００６は、静的コア１００８によって実行されていたニューラルネットワーク（または、その一部）に関連付けられる処理機能を包含する、または引き継ぐ。学習コア１００６は、あらかじめ定められた時間期間にわたって、あるいは特定のタスクまたは機能の実行の間に、処理を引き継ぎ得る。たとえば、いくつかの態様では、学習コア１００６は、ＳＴＤＰ、あるいはニューラルネットワークまたはその一部に関連する強化学習などの学習モデルを実装するために、より少ない機能を備えた静的コア１００８から処理を引き継ぎ得る。
【０１１０】
[0095]別の例では、学習コア１００６によって処理が包含されるニューラルネットワークの一部は、深層信念ネットワークのレイヤであり得る。深層信念ネットワークは、確率的潜在変数の複数のレイヤからなる確率的生成モデルである。深層信念ネットワークでは、学習は、たとえば、トップダウン様式で、レイヤごとに実装され得る。
【０１１１】
[0096]学習はオンラインで実装されてもよく、オフラインで実装されてもよい。オフライン学習が発生すると、学習コア１００６の入力（たとえば、１００２）および出力（たとえば、１０１０）は、ニューラルネットワークの他のレイヤを備え得る。さらに、学習コア１００６の入力（たとえば、１００２）および出力（たとえば、１０１０）はまた、センサ、アクチュエータ等を備え得る。
【０１１２】
[0097]いくつかの態様では、静的コア１００８は、入力の受信を継続し得る。たとえば、静的コア１００８は、教師付き学習を可能にするために監視コアとして動作され得る。したがって、静的コア１００８の出力は、学習コア１００６をトレーニングし得る。他の態様では、静的コア１００８は入力の受信を継続し得、また、ニューラルネットワークまたはその一部の動作に関連付けられる他のタスクを実行するよう割り当てられ得る。他の態様では、静的コア１００８は、入力の受信を停止し得る。
【０１１３】
[0098]図１０Ｅで、あらかじめ定められた時間期間の満了後、あるいはタスクまたは実行された機能が完了する（たとえば、学習が達成される）と、学習コア１００６は処理制御の静的コア１００８への返却を開始し得る。学習コア１００６の状態情報がコピーされて、通信経路１０１２を介して静的コア１００８に供給され得る。いくつかの態様では、学習コア１００６の状態情報は、静的コア１００８の異なるインスタンスを備え得る。たとえば、異なるインスタンスは、達成された学習に基づいて拡張された、変更された静的コア１００８であり得る。別の例では、変更された静的コア１００８は、ＳＴＤＰルールの実装形態に基づいて静的重みの更新を含み得る。
【０１１４】
[0099]図１０Ｆで、学習コア１００６は、学習コア１００６からの状態情報に基づいて、ニューラルネットワークまたはその一部の動作に関連付けられる機能の実行を再開するために、制御を静的コア１００８に返す。
【０１１５】
[00100]図１１は、ニューラルネットワークにおいてコプロセッシングを実行するための方法１１００を示す。ブロック１１０２で、ニューロンモデルは、一定時間期間にわたって、ニューラルネットワークの一部を第１の処理ノードにスワッピングする。ブロック

１１０４で、ニューロンモデルは、第１の処理ノードでニューラルネットワークの一部を実行する。ブロック１１０６で、ニューロンモデルは、一定時間期間後に、ニューラルネットワークの一部を第２の処理ノードに返す。さらに、ブロック１１０８で、ニューロンモデルは、第２の処理ノードでニューラルネットワークの一部を実行する。

【０１１６】

[00101]図１２は、ニューラルネットワークにおいてコプロセッシングを実行するための方法１２００を示す。ブロック１２０２で、ニューロンモデルは、まず第１の処理コアでニューラルネットワークの一部を実行することによって、オフライン学習を共同設置する。ブロック１２０４で、ニューロンモデルは、さらなる実行のためにニューラルネットワークの一部を第２の処理コアに移動させる。

【０１１７】

[00102]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（ＡＳＩＣ）、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。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。

【０１１８】

[00103]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。さらに、「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。さらに、「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。

【０１１９】

[00104]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。

【０１２０】

[00105]本開示および付録Ａ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。

【０１２１】

[00106]本開示および付録Ａ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式の記憶媒体で存在し得る。使用され得る記憶媒体のいくつかの例は、ランダムアクセスメモリ（ＲＡＭ）、読出し専用メモリ（ＲＯＭ）、フラッシュメモリ、消去可能プログラマブル読出し専用メモリ（ＥＰＲＯＭ）、電気的消去可能プログラマブル読出し専用メモリ（ＥＥＰＲＯＭ（登録商標））、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に、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。

【０１２２】

[00107]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。

【０１２３】

[00108]本明細書で説明した機能は、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し得る。ネットワークアダプタは、信号処理機能を実装し得る。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。

【０１２４】

[00109]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、一例として、ランダムアクセスメモリ（ＲＡＭ）、フラッシュメモリ、読出し専用メモリ（ＲＯＭ）、プログラマブル読出し専用メモリ（ＰＲＯＭ）、消去可能プログラマブル読出し専用メモリ（ＥＰＲＯＭ）、電気的消去可能プログラム可能読出し専用メモリ（ＥＥＰＲＯＭ）、レジスタ、磁気ディスク、光ディスク、ハードドライブ、または他の任意の適切な記憶媒体、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。

【０１２５】

[00110]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。論じた様々な構成要素は、ローカル構成要素などの特定の位置を有するものとして説明され得るが、それらはまた、分散コンピューティングシステムの一部として構成されているいくつかの構成要素などの様々な方法で構成され得る。

【０１２６】
[00111]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。あるいは、処理システムは、本明細書に記載のニューロンモデルとニューラルシステムのモデルとを実装するための１つまたは複数のニューロモルフィックプロセッサを備え得る。別の代替として、処理システムは、プロセッサを有する特定用途向け集積回路（ＡＳＩＣ）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とを用いて、あるいは１つまたは複数のフィールドプログラマブルゲートアレイ（ＦＰＧＡ）、プログラマブル論理デバイス（ＰＬＤ）、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、または他の任意の適切な回路、あるいは本開示全体を通じ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。

【０１２７】
[00112]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。

【０１２８】
[00113]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し得、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。さらに、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。

【０１２９】

[00114]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。

【０１３０】

[00115]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、適用可能な場合にユーザ端末および／または基地局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、ユーザ端末および／または基地局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。

【０１３１】

[00116]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。

【図１】

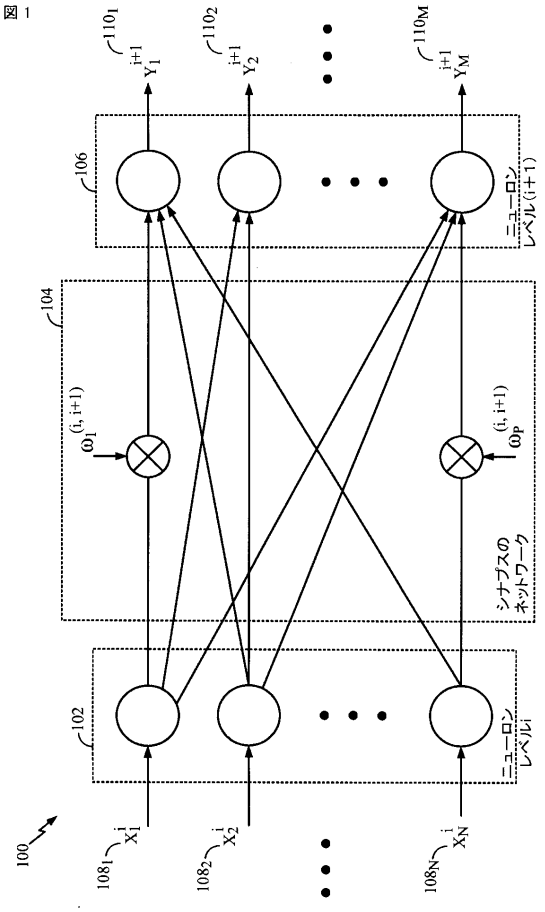

FIG. 1

【図２】

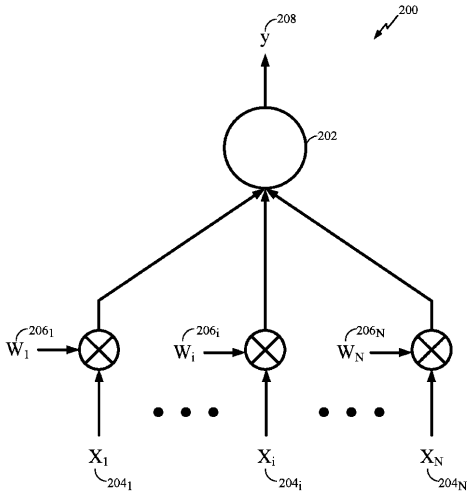

FIG. 2

【図３】

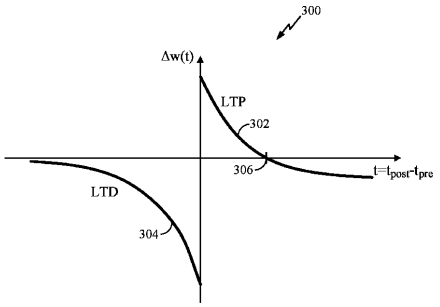

FIG. 3

【図４】

図 4

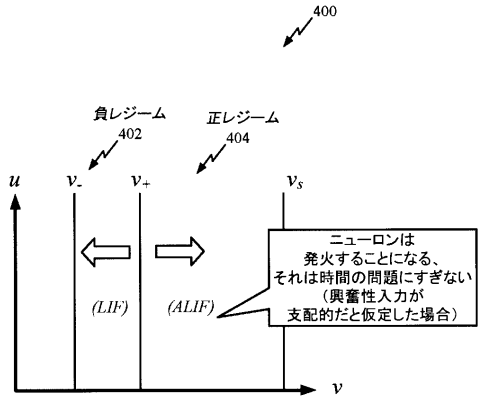

FIG. 4

【図５】

図 5

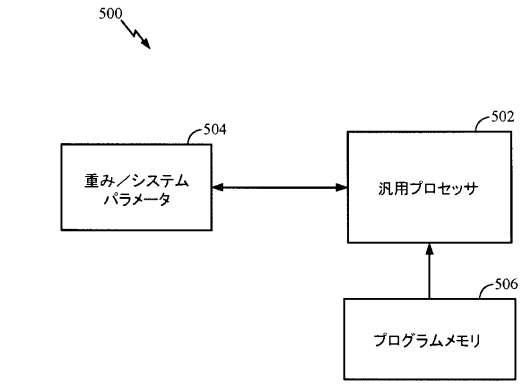

FIG. 5

【図６】

図 6

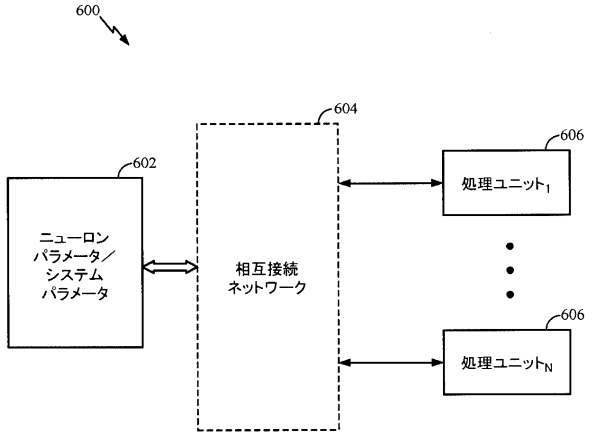

FIG. 6

【図７】

図 7

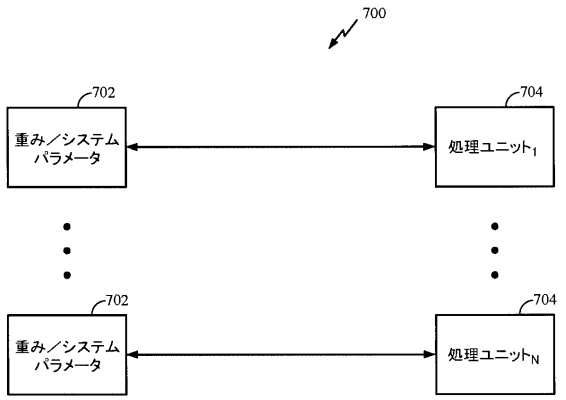

FIG. 7

【図８】

図 8

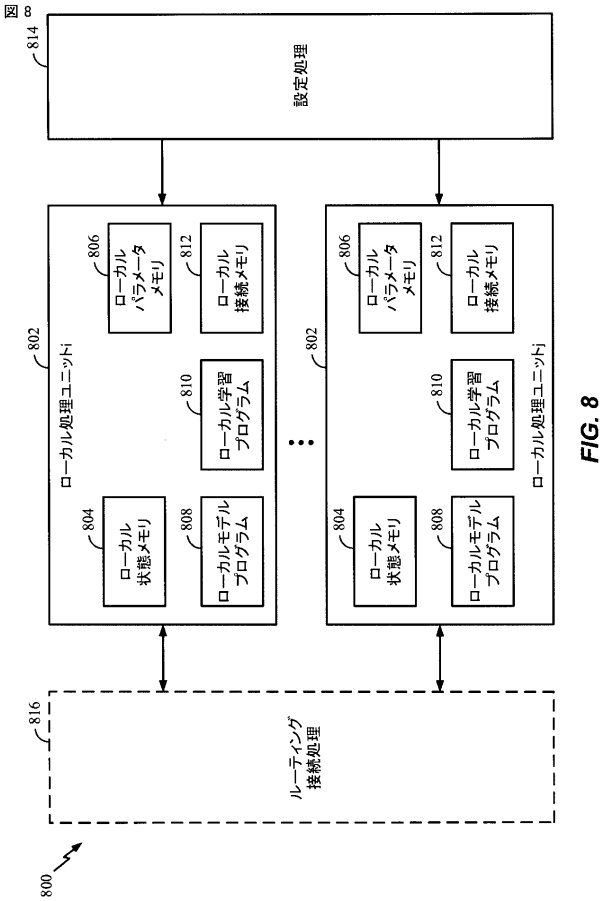

FIG. 8

【図９】

図 9

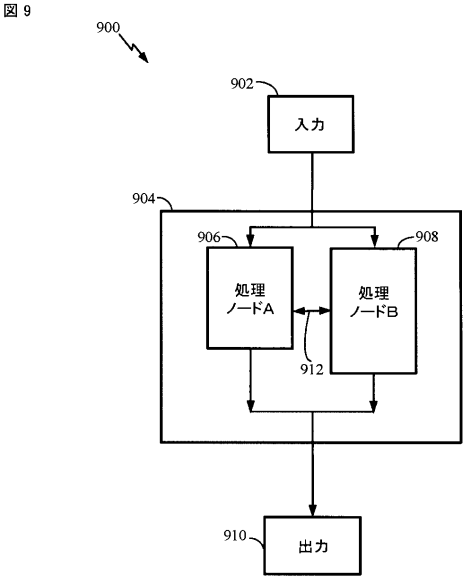

FIG. 9

【図１０Ａ】

図 10A

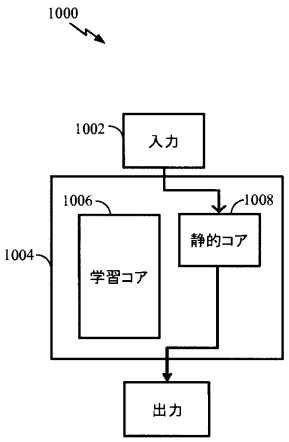

FIG. 10A

【図１０Ｂ】

図 10B

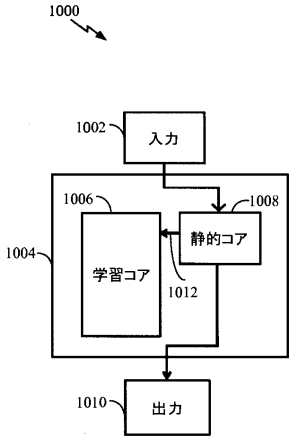

FIG. 10B

【図１０Ｃ】

図 10C

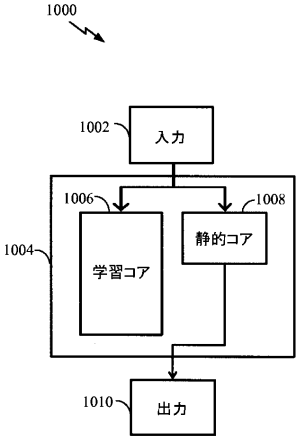

FIG. 10C

【図１０Ｄ】

図 10D

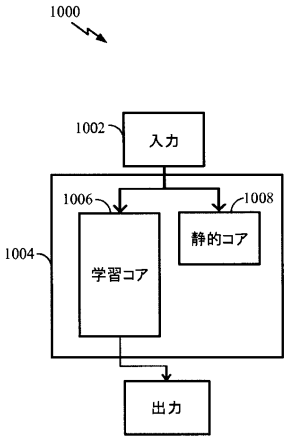

FIG. 10D

【図１０Ｅ】

図 10E

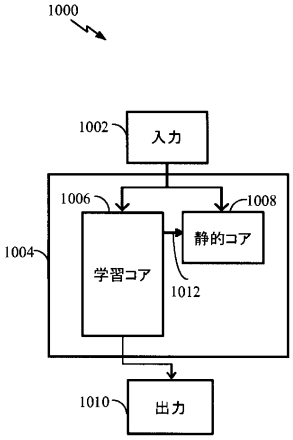

FIG. 10E

【図１０Ｆ】

図 10F

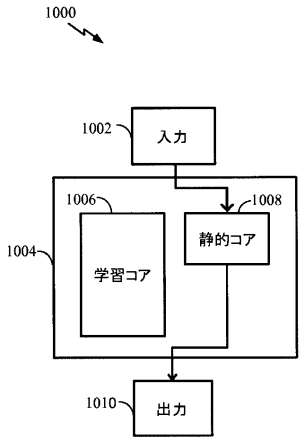

FIG. 10F

【図１１】

図 11

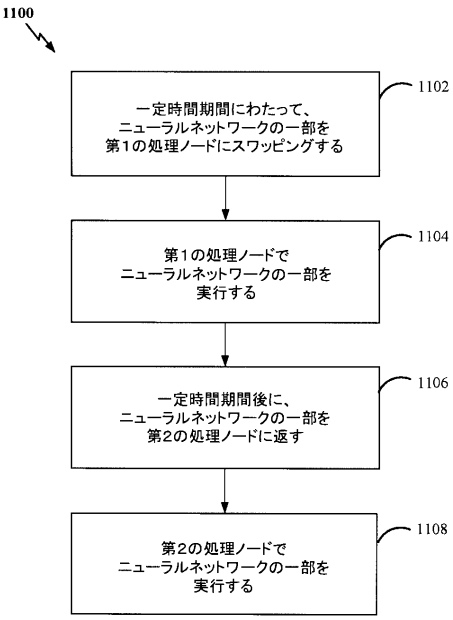

FIG. 11

【図１２】

図 12

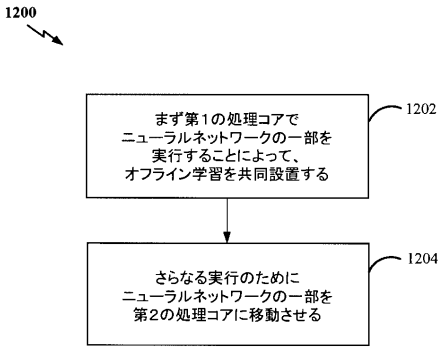

FIG. 12

【手続補正書】
【提出日】平成28年10月20日(2016.10.20)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
人工ニューラルネットワークにおいてコプロセッシングを実行する方法であって、
一定時間期間にわたって、前記ニューラルネットワークの一部を第２の処理ノードから第１の処理ノードにスワッピングすることと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行することと
を備える、方法。
【請求項２】
前記第１の処理ノードは、前記第２の処理ノードとは別個のハードウェアコアを備える、
請求項１に記載の方法。
【請求項３】
前記第１の処理ノードは、学習処理コアを備える、
請求項１に記載の方法。
【請求項４】
前記学習処理コアは、前記第２の処理ノードよりも多くのリソースで構成される、
請求項３に記載の方法。
【請求項５】
学習は、オフラインで実装される、
請求項３に記載の方法。
【請求項６】
前記第１の処理ノードは、学習処理コアを備え、
前記第２の処理ノードは、静的処理コアを備え、
スワッピングすることは、
前記静的処理コアの状態を前記学習処理コアにコピーすることと、
前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに入力をルーティングすることと
を備え、
返すことは、
前記学習処理コアの状態を前記静的処理コアにコピーすることと、
変更された静的処理コアに制御を返すことと
を備える、請求項１に記載の方法。
【請求項７】
前記スワッピングすることは、前記第２の処理ノードから前記人工ニューラルネットワークの前記一部を処理するための前記第１の処理ノードのリソースを割り振ることを備える、
請求項１に記載の方法。
【請求項８】
前記人工ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、

請求項１に記載の方法。
【請求項９】
前記第１の処理ノードは、デバッギングコアを備える、
請求項１に記載の方法。
【請求項１０】
前記スワッピングすることは、システム性能がしきい値を下回る場合にトリガされる、
請求項１に記載の方法。
【請求項１１】
前記返すことは、システム性能がしきい値を上回る場合にトリガされる、
請求項１に記載の方法。
【請求項１２】
前記スワッピングすること、または返すことは、電力がシステムに適用されるとトリガされる、
請求項１に記載の方法。
【請求項１３】
人工ニューラルネットワークにおいてコプロセッシングを実行するための装置であって、
メモリと、
前記メモリに結合された少なくとも１つのプロセッサと
を備え、前記少なくとも１つのプロセッサは、
一定時間期間にわたって、前記ニューラルネットワークの一部を第２の処理ノードから第１の処理ノードにスワッピングすることと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行することと
を行うように構成される、装置。
【請求項１４】
前記第１の処理ノードは、前記第２の処理ノードとは別個のハードウェアコアを備える、
請求項１３に記載の装置。
【請求項１５】
前記第１の処理ノードは、学習処理コアを備える、
請求項１３に記載の装置。
【請求項１６】
前記学習処理コアは、前記第２の処理ノードよりも多くのリソースで構成される、
請求項１５に記載の装置。
【請求項１７】
学習は、オフラインで実装される、
請求項１５に記載の装置。
【請求項１８】
前記第１の処理ノードは、学習処理コアを備え、前記第２の処理ノードは、静的処理コアを備え、前記少なくとも１つのプロセッサは、
前記静的処理コアの状態を前記学習処理コアにコピーすることと、
前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに入力をルーティングすることと、
前記学習処理コアの状態を前記静的処理コアにコピーすることと、
変更された静的処理コアに制御を返すことと
を行うようにさらに構成される、請求項１３に記載の装置。
【請求項１９】

前記少なくとも１つのプロセッサは、前記第２の処理ノードから前記人工ニューラルネットワークの前記一部を処理するための前記第１の処理ノードのリソースを割り振ることを行うようにさらに構成される、
請求項１３に記載の装置。
【請求項２０】
前記人工ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、
請求項１３に記載の装置。
【請求項２１】
前記第１の処理ノードは、デバッギングコアを備える、
請求項１３に記載の装置。
【請求項２２】
前記少なくとも１つのプロセッサは、システム性能がしきい値を下回る場合に、前記ニューラルネットワークの前記一部を前記第１の処理ノードにスワッピングするようにさらに構成される、
請求項１３に記載の装置。
【請求項２３】
前記少なくとも１つのプロセッサは、システム性能がしきい値を上回る場合に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すようにさらに構成される、
請求項１３に記載の装置。
【請求項２４】
前記少なくとも１つのプロセッサは、電力がシステムに適用されると、前記ニューラルネットワークの前記一部を前記第１の処理ノードにスワッピングする、または前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すようにさらに構成される、
請求項１３に記載の装置。
【請求項２５】
人工ニューラルネットワークにおいてコプロセッシングを実行するための装置であって、
一定時間期間にわたって、前記ニューラルネットワークの一部を第２の処理ノードから第１の処理ノードにスワッピングするための手段と、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行するための手段と、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すための手段と、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行するための手段と
を備える、装置。
【請求項２６】
人工ニューラルネットワークにおいてコプロセッシングを実行するためのプログラムコードを符号化した非一時的コンピュータ可読媒体であって、前記プログラムコードは、プロセッサによって実行され、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングするためのプログラムコードと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行するためのプログラムコードと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すためのプログラムコードと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行するためのプログラムコードと

を備える、非一時的コンピュータプログラム可読媒体。
【手続補正２】
【補正対象書類名】明細書
【補正対象項目名】０１３１
【補正方法】変更
【補正の内容】
【０１３１】
[00116]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。
以下に、出願当初の特許請求の範囲に記載された発明を付記する。
［Ｃ１］
ニューラルネットワークにおいてコプロセッシングを実行する方法であって、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングすることと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行することと
を備える、方法。
［Ｃ２］
前記第１の処理ノードは、別個のハードウェアコアを備える、
Ｃ１に記載の方法。
［Ｃ３］
前記第１の処理ノードは、学習処理コアを備える、
Ｃ１に記載の方法。
［Ｃ４］
前記学習処理コアは、前記第２の処理ノードよりも高いレベルのリソースで構成される、
Ｃ３に記載の方法。
［Ｃ５］
学習は、オフラインまたはオンラインで実装される、
Ｃ３に記載の方法。
［Ｃ６］
前記学習処理コアの入力および出力は、学習がオフラインで実装される場合、前記ニューラルネットワークの他のレイヤを備える、
Ｃ５に記載の方法。
［Ｃ７］
前記第１の処理ノードは、学習処理コアを備え、
前記第２の処理ノードは、静的処理コアを備え、
スワッピングすることは、
前記静的処理コアの状態を前記学習処理コアにコピーすることと、
前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに入力をルーティングすることと
を備え、
返すことは、
前記学習処理コアの状態を前記静的処理コアにコピーすることと、
変更された静的処理コアに制御を返すことと
を備える、Ｃ１に記載の方法。
［Ｃ８］

前記スワッピングすることは、前記第１の処理ノードから前記第２の処理ノードにリソースを割り振ることを備える、
Ｃ１に記載の方法。
［Ｃ９］
前記ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、
Ｃ１に記載の方法。
［Ｃ１０］
前記第１の処理ノードは、デバッギングコアを備える、
Ｃ１に記載の方法。
［Ｃ１１］
前記スワッピングすることは、システム性能がしきい値を下回る場合に発生する、
Ｃ１に記載の方法。
［Ｃ１２］
前記返すことは、システム性能がしきい値を上回る場合に発生する、
Ｃ１に記載の方法。
［Ｃ１３］
前記スワッピングすること、または返すことは、電力がシステムに適用されると発生する、
Ｃ１に記載の方法。
［Ｃ１４］
ニューラルネットワークにおいてコプロセッシングを実行するための装置であって、
メモリと、
前記メモリに結合された少なくとも１つのプロセッサと
を備え、前記少なくとも１つのプロセッサは、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングすることと、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行することと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すことと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行することと
を行うように構成される、装置。
［Ｃ１５］
前記第１の処理ノードは、別個のハードウェアコアを備える、
Ｃ１４に記載の装置。
［Ｃ１６］
前記第１の処理ノードは、学習処理コアを備える、
Ｃ１４に記載の装置。
［Ｃ１７］
前記学習処理コアは、前記第２の処理ノードよりも高いレベルのリソースで構成される、
Ｃ１６に記載の装置。
［Ｃ１８］
学習は、オフラインまたはオンラインで実装される、
Ｃ１６に記載の装置。
［Ｃ１９］
前記学習処理コアの入力および出力は、学習がオフラインで実装される場合、前記ニューラルネットワークの他のレイヤを備える、
Ｃ１８に記載の装置。
［Ｃ２０］
前記第１の処理ノードは、学習処理コアを備え、前記第２の処理ノードは、静的処理コ

アを備え、前記少なくとも１つのプロセッサは、
　前記静的処理コアの状態を前記学習処理コアにコピーすることと、
　前記学習処理コアが前記静的処理コアの機能を包含するように、前記学習処理コアに入力をルーティングすることと、
　前記学習処理コアの状態を前記静的処理コアにコピーすることと、
　変更された静的処理コアに制御を返すことと
を行うようにさらに構成される、Ｃ１４に記載の装置。
［Ｃ２１］
前記少なくとも１つのプロセッサは、前記第１の処理ノードから前記第２の処理ノードにリソースを割り振ることを行うようにさらに構成される、
Ｃ１４に記載の装置。
［Ｃ２２］
前記ニューラルネットワークの前記一部は、深層信念ネットワークのレイヤを備える、
Ｃ１４に記載の装置。
［Ｃ２３］
前記第１の処理ノードは、デバッギングコアを備える、
Ｃ１４に記載の装置。
［Ｃ２４］
前記少なくとも１つのプロセッサは、システム性能がしきい値を下回る場合に、前記ニューラルネットワークの前記一部を前記第１の処理ノードにスワッピングするようにさらに構成される、
Ｃ１４に記載の装置。
［Ｃ２５］
前記少なくとも１つのプロセッサは、システム性能がしきい値を上回る場合に、前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すようにさらに構成される、
Ｃ１４に記載の装置。
［Ｃ２６］
前記少なくとも１つのプロセッサは、電力がシステムに適用されると、前記ニューラルネットワークの前記一部を前記第１の処理ノードにスワッピングする、または前記ニューラルネットワークの前記一部を前記第２の処理ノードに返すようにさらに構成される、
Ｃ１４に記載の装置。
［Ｃ２７］
ニューラルネットワークにおいてコプロセッシングを実行するための装置であって、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングするための手段と、
前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行するための手段と、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すための手段と、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行するための手段と
を備える、装置。
［Ｃ２８］
ニューラルネットワークにおいてコプロセッシングを実行するためのコンピュータプログラム製品であって、
プログラムコードを符号化した非一時的コンピュータ可読媒体を備え、前記プログラムコードは、
一定時間期間にわたって、前記ニューラルネットワークの一部を第１の処理ノードにスワッピングするためのプログラムコードと、

前記第１の処理ノードで前記ニューラルネットワークの前記一部を実行するためのプログラムコードと、
前記一定時間期間後に、前記ニューラルネットワークの前記一部を第２の処理ノードに返すためのプログラムコードと、
前記第２の処理ノードで前記ニューラルネットワークの前記一部を実行するためのプログラムコードと
を備える、コンピュータプログラム製品。

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/015917

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06N3/08 G06N3/10
ADD. G06N3/04

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, WPI Data, INSPEC

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 6 418 423 B1 (KAMBHATLA NANDAKISHORE [US] ET AL) 9 July 2002 (2002-07-09) columns 4-10	1-28
X	D. CALVERT ET AL: "Distributed artificial neural network architectures", PROCEEDINGS OF THE 19TH INTERNATIONAL SYMPOSIUM ON HIGH PERFORMANCE COMPUTING SYSTEMS AND APPLICATIONS (HPCS'05), 15 May 2005 (2005-05-15), pages 2-10, XP010800324, DOI: 10.1109/HPCS.2005.24 section II	1-28

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
1 December 2015	14/12/2015

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Douarche, Nicolas

3

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/015917

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	C. ROSSANT ET AL: "Playdoh: a lightweight Python library for distributed computing and optimisation", JOURNAL OF COMPUTATIONAL SCIENCE, vol. 4, no. 5, 6 July 2011 (2011-07-06), pages 352-359, XP055183262, DOI: 10.1016/j.jocs.2011.06.002 sections 2 and 4.1	1-28
X	F. GALLUPPI ET AL: "A hierachical configuration system for a massively parallel neural hardware platform", PROCEEDINGS OF THE 9TH CONFERENCE ON COMPUTING FRONTIERS (CF'12), 15 May 2012 (2012-05-15), pages 183-192, XP055179253, DOI: 10.1145/2212908.2212934 section 4.1	1-28
X	X. CHEN ET AL: "Pipelined back-propagation for context-dependent deep neural networks", PROCEEDINGS OF THE 13TH ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION (INTERSPEECH'12), 9 September 2012 (2012-09-09), pages 26-29, XP055113689, ISSN: 1990-9770 section 3	1-28
X	E. SCHIKUTA, E. MANN: "N2Sky - neural networks as services in the clouds", PROCEEDINGS OF THE 2013 INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS (IJCNN'13), 4 August 2013 (2013-08-04), XP032542139, DOI: 10.1109/IJCNN.2013.6707113 the whole document	1-28
X	K. MINKOVICH ET AL: "HRLSim: a high performance spiking neural network simulator for GPGPU clusters", IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, vol. 25, no. 2, 15 August 2013 (2013-08-15), pages 316-331, XP011536922, DOI: 10.1109/TNNLS.2013.2276056 section III	1-28

3

Form PCT/ISA/210 (continuation of second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2015/015917

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 6418423 B1	09-07-2002	NONE	

フロントページの続き

(81)指定国 AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CZ,DE,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,JP,KE,KG,KN,KP,KR,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,SM,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US

(72)発明者 カンポス、マイケル
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 レウィス、アンソニー
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 ラオ、ナビーン・ガンドハム
アメリカ合衆国、カリフォルニア州 ９２１３１、サン・ディエゴ、カミニート・アレグラ １１０８７