(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7248667号
(P7248667)

(45)発行日　令和5年3月29日(2023.3.29)

(24)登録日　令和5年3月20日(2023.3.20)

(51)国際特許分類		FI		
G06N	*3/063(2023.01)*	G06N	3/063	
G06F	*15/173(2006.01)*	G06F	15/173	683C
G06F	*9/38 (2018.01)*	G06F	9/38	370C
G06F	*13/38 (2006.01)*	G06F	13/38	340A

請求項の数　24　（全20頁）

(21)出願番号　特願2020-520474(P2020-520474)
(86)(22)出願日　平成30年10月12日(2018.10.12)
(65)公表番号　特表2021-500648(P2021-500648A)
(43)公表日　令和3年1月7日(2021.1.7)
(86)国際出願番号　PCT/IB2018/057922
(87)国際公開番号　WO2019/077455
(87)国際公開日　平成31年4月25日(2019.4.25)
審査請求日　令和3年3月23日(2021.3.23)
(31)優先権主張番号　15/789,472
(32)優先日　平成29年10月20日(2017.10.20)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　390009531
インターナショナル・ビジネス・マシーンズ・コーポレーション
INTERNATIONAL BUSINESS MACHINES CORPORATION
アメリカ合衆国１０５０４ ニューヨーク州 アーモンク ニュー オーチャード ロード
New Orchard Road, Armonk, New York 10504, United States of America
(74)代理人　100112690
弁理士　太佐 種一

最終頁に続く

(54)【発明の名称】　メッセージ・パッシング計算システムのためのメモリ・マップド・インターフェースのシステム、方法及びコンピュータ・プログラム

(57)【特許請求の範囲】

【請求項１】

ネットワークによって相互接続された少なくとも１つのニューラル・ネットワーク処理コアを備える、ニューラル・ネットワーク・プロセッサ・システムと、

前記ネットワークへ作動的に接続されたインターフェースであって、前記インターフェースはメモリ・マップを備え、前記メモリ・マップは前記ニューラル・ネットワーク・プロセッサ・システムへの入力に対応する第１の領域および前記ニューラル・ネットワーク・プロセッサ・システムからの出力に対応する第２の領域を備える、前記インターフェースと

を備え、

前記インターフェースは、書き込みデータおよび書き込みアドレスを含む書き込みリクエストを受信するようになっており、前記書き込みアドレスは、前記メモリ・マップの前記第１の領域内のロケーションに対応し、

前記インターフェースは、前記書き込みデータを、前記ネットワークを介して、宛先ニューラル・ネットワーク処理コアの少なくとも１つの入力ロケーションへ送るようになっており、

前記インターフェースは、読み出しアドレスを含む読み出しリクエストを受信するようになっており、前記読み出しアドレスは、前記メモリ・マップの前記第２の領域内のロケーションに対応し、

前記インターフェースは、読み出しデータを前記読み出しリクエストの送信者へ返すよう

になっており、前記読み出しデータは、ソース・ニューラル・ネットワーク処理コアの出力を含む、
システム。
【請求項２】
前記インターフェースはアドレス変換器を備え、前記アドレス変換器は、
前記書き込みアドレスを目標アドレスへ変換するようになっており、前記目標アドレスは、前記宛先ニューラル・ネットワーク処理コアの前記少なくとも１つの入力ロケーションを特定する、
請求項１に記載のシステム。
【請求項３】
前記変換することは所定の関数を適用することを含む、請求項２に記載のシステム。
【請求項４】
前記変換することはルックアップ・テーブルにアクセスすることを含む、請求項２又は３に記載のシステム。
【請求項５】
前記インターフェースはアドレス変換器を備え、
前記インターフェースは、メッセージを、前記ネットワークを介して、前記ソース・ニューラル・ネットワーク処理コアから受信するようになっており、前記メッセージは、データおよび少なくとも１つのアドレスを含み、
前記アドレス変換器は、前記少なくとも１つのアドレスに基づいて前記メモリ・マップの前記第２の領域内のアドレスを決定するようになっており、
前記インターフェースは、前記第２の領域内の前記アドレスを含む読み出しリクエストに応答して前記データを提供するようになっている、
請求項１乃至４のいずれかに記載のシステム。
【請求項６】
前記インターフェースへ作動的に結合されたバスをさらに備え、
前記インターフェースは、前記バスを介して、前記書き込みリクエストを受信するようになっており、
前記インターフェースは、前記バスを介して、前記読み出しリクエストを受信するようになっている、
請求項１乃至５のいずれかに記載のシステム。
【請求項７】
前記少なくとも１つのニューラル・ネットワーク処理コアはニューロシナプティック・コアを備え、前記ネットワークはコア間ネットワークを備え、前記ニューロシナプティック・コアは複数の入力軸索および出力ニューロンを有する、請求項１乃至６のいずれかに記載のシステム。
【請求項８】
前記宛先ニューラル・ネットワーク処理コアはニューロシナプティック・コアであり、前記書き込みアドレスは前記ニューロシナプティック・コアの１つ以上の入力軸索に対応し、前記書き込みデータは前記ニューロシナプティック・コアへの複数の軸索入力を含む、請求項７に記載のシステム。
【請求項９】
前記ソース・ニューラル・ネットワーク処理コアはニューロシナプティック・コアであり、前記読み出しアドレスは前記ニューロシナプティック・コアの１つ以上の出力ニューロンに対応し、前記読み出しデータは前記ニューロシナプティック・コアの複数のニューロン出力を含む、請求項７又は８に記載のシステム。
【請求項１０】
前記インターフェースはバッファを備え、
前記インターフェースは、メッセージを、前記ネットワークを介して、前記ソース・ニューラル・ネットワーク処理コアから受信するようになっており、前記メッセージは、メ

ッセージ・データおよび少なくとも１つのアドレスを含み、
前記インターフェースは、前記メッセージ・データを前記バッファ中で１つのロケーションに記憶するようになっており、前記少なくとも１つのアドレスは、前記バッファ中の前記ロケーションに対応する、
請求項１乃至９のいずれかに記載のシステム。
【請求項１１】
前記インターフェースへ作動的に結合されたセンサをさらに備え、前記センサは、前記インターフェースを介して、センサ出力を前記ニューラル・ネットワーク・プロセッサ・システムへ送信するようになっている、
請求項１乃至１０のいずれかに記載のシステム。
【請求項１２】
前記インターフェースへ作動的に結合されたアクチュエータをさらに備え、前記アクチュエータは、前記インターフェースを介して、入力を前記ニューラル・ネットワーク・プロセッサ・システムから受信するようになっている、
請求項１乃至１１のいずれかに記載のシステム。
【請求項１３】
前記インターフェースへ作動的に結合された計算ノードをさらに備え、前記計算ノードは、直接メモリ・アクセスを介して、前記インターフェースにアクセスするようになっている、
請求項１乃至１２のいずれかに記載のシステム。
【請求項１４】
前記インターフェースへ作動的に結合された計算ノードをさらに備え、前記計算ノードは、バスを介して、前記インターフェースにアクセスするようになっている、請求項１乃至１３のいずれかに記載のシステム。
【請求項１５】
前記計算ノード、前記インターフェース、および前記ニューラル・ネットワーク・プロセッサ・システムは、単一のチップ上に位置する、請求項１４に記載のシステム。
【請求項１６】
メモリ・マップド・システムとメッセージ・パッシング・システムとを相互接続し、メモリ・マップを備えるインターフェースが実行する方法であって、
書き込みデータおよび書き込みアドレスを含む書き込みリクエストを受信するステップであって、前記書き込みアドレスは前記メモリ・マップ内のメモリ・アドレスである、前記受信するステップと、
前記書き込みアドレスをニューラル・ネットワーク・アドレスへ変換するステップであって、前記ニューラル・ネットワーク・アドレスは、宛先ニューラル・ネットワークの少なくとも１つの入力ロケーションを特定する、前記変換するステップと、
前記書き込みデータを、ネットワークを介して、前記ニューラル・ネットワーク・アドレスに従って前記宛先ニューラル・ネットワークの前記少なくとも１つの入力ロケーションへ送るステップと、
を含む、方法。
【請求項１７】
前記書き込みリクエストはバスを介して受信される、請求項１６に記載の方法。
【請求項１８】
前記ネットワークはコア間ネットワークである、請求項１６又は１７に記載の方法。
【請求項１９】
前記宛先ニューラル・ネットワークはニューロシナプティック・コアを備える、請求項１６乃至１８のいずれかに記載の方法。
【請求項２０】
前記変換するステップは所定の関数を適用するステップを含む、請求項１６乃至１９のいずれかに記載の方法。

【請求項２１】
前記変換するステップはルックアップ・テーブルにアクセスするステップを含む、請求項１６乃至２０のいずれかに記載の方法。
【請求項２２】
前記書き込みデータは複数の軸索入力を含む、請求項１６乃至２１のいずれかに記載の方法。
【請求項２３】
前記書き込みデータを前記宛先ニューラル・ネットワークの入力バッファにおいて受信するステップをさらに含む、請求項１６乃至２２のいずれかに記載の方法。
【請求項２４】
メモリ・マップド・システムをメッセージ・パッシング・システムとインターフェースするためのコンピュータ・プログラムであって、メモリ・マップと、バッファと、プロセッサとを備えたコンピュータにより実行されることで、前記プロセッサに
書き込みデータおよび書き込みアドレスを含む書き込みリクエストを受信するステップであって、前記書き込みアドレスは前記メモリ・マップ内のメモリ・アドレスである、前記受信するステップと、
前記書き込みアドレスをニューラル・ネットワーク・アドレスへ変換するステップであって、前記ニューラル・ネットワーク・アドレスは宛先ニューラル・ネットワークの少なくとも１つの入力ロケーションを特定する、前記変換するステップと、
前記書き込みデータを、ネットワークを介して、前記ニューラル・ネットワーク・アドレスに従って前記宛先ニューラル・ネットワークの前記少なくとも１つの入力ロケーションへ送るステップと、
メッセージを、前記ネットワークを介して、ソース・ニューラル・ネットワークから受信するステップであって、前記メッセージはデータおよび少なくとも１つのアドレスを含む、前記受信するステップと、
前記少なくとも１つのアドレスに基づいて前記バッファ中のロケーションを決定するステップと、
前記データを前記バッファ中の前記ロケーションに記憶するステップであって、前記バッファは前記メモリ・マップを介してアクセス可能である、前記記憶するステップと
を行わせる、
コンピュータ・プログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本開示の実施形態は、メッセージ・パッシング・システムへのインターフェースに関し、より具体的には、メッセージ・パッシング計算システムのためのメモリ・マップド・インターフェースに関する。
【発明の概要】
【発明が解決しようとする課題】
【０００２】
様々な従来の計算システムは、共有メモリ／メモリ・マップド（ＭＭ：ｍｅｍｏｒｙ ｍａｐｐｅｄ）パラダイムを介してシステム構成要素間で通信を行う。対照的に、ニューロシナプティック・システムのような、様々な並列・分散計算システムは、メッセージ・パッシング（ＭＰ：ｍｅｓｓａｇｅ ｐａｓｓｉｎｇ）パラダイムを介して相互に通信を行う。本開示は、これら２つのタイプのシステム間の効率的なインターフェースを提供する。
【課題を解決するための手段】
【０００３】
本開示の実施形態によれば、メモリ・マップド・システムをメッセージ・パッシング・システムとインターフェースする方法およびそのためのコンピュータ・プログラム製品が

提供される。書き込みリクエストが受信される。書き込みリクエストは、書き込みデータおよび書き込みアドレスを含む。書き込みアドレスは、メモリ・マップ内のメモリ・アドレスである。書き込みアドレスがニューラル・ネットワーク・アドレスへ変換される。ニューラル・ネットワーク・アドレスは、宛先ニューラル・ネットワークの少なくとも１つの入力ロケーションを特定する。書き込みデータが、ネットワークを介して、ニューラル・ネットワーク・アドレスに従って宛先ニューラル・ネットワークの少なくとも１つの入力ロケーションへ送られる。メッセージが、そのネットワークを介して、ソース・ニューラル・ネットワークから受信される。メッセージは、データおよび少なくとも１つのアドレスを含む。少なくとも１つのアドレスに基づいてバッファ中のロケーションが決定される。データがバッファ中のそのロケーションに記憶される。バッファは、メモリ・マップを介してアクセス可能である。

【０００４】

本開示の実施形態によれば、ニューラル・ネットワーク処理システムとインターフェースするためのシステムが提供される。ニューラル・ネットワーク処理システムは、ネットワークによって相互接続された少なくとも１つのニューラル・ネットワーク処理コアを備える。インターフェースがそのネットワークへ作動的に接続される。インターフェースは、メモリ・マップを備える。メモリ・マップは、ニューラル・ネットワーク・プロセッサ・システムへの入力に対応する第１の領域およびニューラル・ネットワーク・プロセッサ・システムからの出力に対応する第２の領域を備える。インターフェースは、書き込みデータおよび書き込みアドレスを含む書き込みリクエストを受信するようになっている。書き込みアドレスは、メモリ・マップの第１の領域内のロケーションに対応する。インターフェースは、書き込みデータを、そのネットワークを介して、宛先ニューラル・ネットワーク処理コアの少なくとも１つの入力ロケーションへ送るようになっている。インターフェースは、読み出しアドレスを含む読み出しリクエストを受信するようになっている。読み出しアドレスは、メモリ・マップの第２の領域内のロケーションに対応する。インターフェースは、読み出しデータを読み出しリクエストの送信者へ返すようになっている。読み出しデータは、ソース・ニューラル・ネットワーク処理コアの出力を含む。

【０００５】

次に、本発明の実施形態が、単に例として、添付図面を参照して記載される。

【図面の簡単な説明】

【０００６】

【図１】本開示の実施形態による例示的なメモリ・マップド（ＭＭ）システムを示す。

【図２】本開示の実施形態による例示的なメッセージ・パッシング（ＭＰ）システムを示す。

【図３】本開示の実施形態によるニューロシナプティック・コアを示す。

【図４】本開示の実施形態によるニューロシナプティック・コアおよびメッシュ・ルータを示す。

【図５】本開示の実施形態による組み合わされたメモリ・マップドおよびメッセージ・パッシング・システムを示す。

【図６】本開示の実施形態による例示的なコアの入力バッファおよび出力バッファを示す。

【図７】本開示の実施形態による例示的なメモリ・マップを示す。

【図８】本開示の実施形態によるメッセージ・パッシング・システムのための例示的なメモリ・マップド・インターフェースを示す。

【図９】本開示の実施形態によるメモリ・マップド・システムとメッセージ・パッシング・システムとをインターフェースする方法を示す。

【図１０】本開示の実施形態による計算ノードを示す。

【発明を実施するための形態】

【０００７】

様々な従来の計算システムは、共有メモリ／メモリ・マップド（ＭＭ）パラダイムを介してシステム構成要素間で通信を行う。対照的に、ニューロシナプティック・システムの

ような、様々な並列・分散計算システムは、メッセージ・パッシング（ＭＰ）パラダイムを介して相互に通信を行う。本開示は、これら２つのタイプのシステム間の効率的なインターフェースを提供する。

【０００８】

本明細書において提供される様々な例示的な実施形態は、ニューロシナプティック・スパイクベースのシステムの観点から記載される。しかしながら、本開示は、分散システム、ニューロシナプティック・システム、および他の非フォン・ノイマン・システムを含めて任意のメッセージ・パッシング・システムに適用可能であることが認識されよう。

【０００９】

様々なニューロシナプティック・システムにおいては、ニューロシナプティック・コアと称される、極めて低電力のニューロシナプティック処理ユニットのアレイがエクサスケールおよびビッグ・データ問題を解決するためのアーキテクチャを提供する。これらのコアは、情報をエンコードするためにスパイクを用いる。ニューロシナプティック・コアのネットワークでは、各コア上のニューロンが（それ自体を含めて）他の任意のニューロシナプティック・コアの任意の軸索へ接続できる。ニューロンは、スパイクを発するときに、宛先コア上の標的軸索へ送達されるスパイク・パケットを送る。

【００１０】

デジタル・スパイキング・ニューロモーフィック・システムでは、情報が複数のスパイクによって表され、かつ送達されて、各スパイクは、１つ以上のビットを運ぶ情報のデジタル・パケットである。例えば、ＩＢＭ（ＩＢＭ社の登録商標）ＴｒｕｅＮｏｒｔｈチップは、各スパイクが情報の単一ビット（バイナリ・スパイク）を運ぶ、デジタル・スパイキング・ニューロモーフィック・システムである。ＴｒｕｅＮｏｒｔｈのようなスパイキング・ニューラル・ネットワークは、情報のパケットを通信線切り替えを通じて送達することに基づいており、それにより、必要とされる配線を著しく削減する。スパイクの存在は、１を受信することとして取り扱われ、その非存在が０を表す。いくつかの異なるスパイク・コーディング・スキームを用いて、より多くの値をバイナリ・スパイクへコードすることができる。

【００１１】

ソース・コア上のソース・ニューロンから宛先コア上の標的軸索へのスパイク通信は、２Ｄグリッド中のルータを水平もしくは垂直のいずれかまたは両方の組み合わせで経由して、宛先コア上の標的軸索へ送達されるように、ある一定の数のホップをトラバースすることが実際上必要であろう。スパイク・パケットがトラバースする各ホップが電力およびエネルギーを消費する。

【００１２】

ＴｒｕｅＮｏｒｔｈのような例示的なニューロモーフィック・システム内では、スパイクがそのソース・ニューロンからその宛先軸索へ走行するために一定の時間が許容される。この固定ウィンドウは、ティックと呼ばれる。スパイクがその行程に必要とする時間は、スパイクが走行しなければならない距離、ならびにスパイクがそれらにわたって走行する２Ｄメッシュ・ルーティング、チップおよびボード・インターフェースの数に基づいて変化する。

【００１３】

各ティックの際には、最初のニューロンから開始し、最後のニューロンまで続いて、コア中のニューロンが順次に処理される。それに応じて、先に考察された送信遅延に加えて、各スパイクは、コア上のどのニューロンがそれを発生させたかに基づいていくらか追加的に一定時間さらに遅延される。例えば、コアごとに２５６個のニューロンを有するＴｒｕｅＮｏｒｔｈのような、例示的なニューロモーフィック・システムでは、先行する２５５個のニューロンが処理されるまで２５６番目のニューロンは処理されない。

【００１４】

本開示の様々な実施形態によれば、ニューロシナプティック・プログラムは、ニューロシナプティック・ネットワークを表す。ニューロシナプティック・プログラムは、ニュー

ロシナプティック・ネットワークに関する情報を含む。いくつかの実施形態において、情報は、入力スパイクへの電子ニューロンの応答を決定するニューロン特性およびダイナミクスを含む。例えば、ニューロン特性およびダイナミクスは、閾値パラメータ、リーク・パラメータ、遅延パラメータ、またはリセット・パラメータを含むことができる。いくつかの実施形態において、ニューロシナプティック・プログラム情報は、ニューロンのシナプス接続（例えば、シナプス・クロスバーを介してなされるシナプス接続）を含む。いくつかの実施形態において、ニューロシナプティック・プログラム情報は、軸索特性（例えば、軸索タイプ）を含む。いくつかの実施形態において、ニューロシナプティック・プログラム情報は、ニューロンの出力スパイクが送達されるべき１つ以上の宛先（例えば、標的軸索）を含む。

【００１５】

様々な実施形態によれば、ニューロシナプティック・ネットワークは、ニューロシナプティック・プログラムのインスタンス化を表す。ニューロシナプティック・ネットワークは、ハードウェアで、シミュレーションでまたは両方でインスタンス化されてよい。例えば、ニューロシナプティック・プログラムがニューロシナプティック・ネットワークの１つ以上のインスタンスを生じさせてよく、これらのインスタンスは、単一のコア、複数のコア、または複数のチップ上にあってよい。

【００１６】

様々な実施形態によれば、ニューロモーフィック・コア回路は、ニューロシナプティック・プログラムによって記述されるニューロシナプティック・ネットワークの例を表す。

【００１７】

様々な実施形態によれば、コアレットまたはＣｏｒｅｌｅｔ　Ｐｒｏｇｒａｍｍｉｎｇ　Ｌａｎｇｕａｇｅは、ニューロシナプティック・プログラムの抽象化を提供するソフトウェアを表す。ニューロシナプティック・プログラムの構成は、コアレットを構成することによって作成されてよい。

【００１８】

ＴｒｕｅＮｏｒｔｈプログラムは、その外部入力および出力と併せて、ニューロシナプティック・コアのネットワークの完全な仕様である。様々な実施形態においては、分割統治アプローチが採用されて、それにより、ニューロシナプティック・コアの大きいネットワークが、ニューロシナプティック・コアの一組のより小さいネットワークを相互接続することによって構築され、より小さいネットワークの各々は、分割不可能な基本的ビルディング・ブロックである単一のニューロシナプティック・コアからなるネットワークまで、一組のさらにより小さいネットワークを相互接続することなどによって、順次、構築されうるであろう。このプログラミング・パラダイムがＣｏｒｅｌｅｔ　Ｐｒｏｇｒａｍｍｉｎｇと呼ばれる。

【００１９】

図１を参照すると、例示的なメモリ・マップド・システム１００が示される。メモリ・マップ１０１がセグメント化されて、領域１０２．．．１０５は、様々なシステム構成要素へ配分される。計算コア１０６．．．１０９、例えば、１つ以上のチップ上のプロセッサ・コアは、バス１１０へ接続される。各コア１０６．．．１０９は、バス１１０へ接続されて、メモリ・マップ１０２．．．１０３のアドレス可能な領域に対応する共有メモリ１１１．．．１１２を通して相互に通信できる。各コア１０６．．．１０９は、メモリ・マップ１０１のアドレス可能な領域１０４を通してサブシステム１１３と通信できる。同様に、各コア１０６．．．１０９は、メモリ・マップ１０１のアドレス可能な領域１０５を通して外部システム１１４と通信できる。

【００２０】

メモリ・マップ（ＭＭ）アドレスは、グローバル・メモリ・マップに関連し、この例では０ｘ００００００００～０ｘＦＦＦＦＦＦＦＦに及ぶ。

【００２１】

図２を参照すると、例示的なメッセージ・パッシング（ＭＰ）システム２００が示され

る。複数のコア２０１．．．２０９は、各々が計算コア２１０、メモリ２１１、および通信インターフェース２１２を備える。コア２０１．．．２０９の各々がネットワーク２１３によって接続される。通信インターフェース２１２は、パケットをネットワーク２１３へ注入およびネットワーク２１３から受信するために入力バッファ２１４および出力バッファ２１５を備える。このように、コア２０１．．．２０９は、メッセージをパスすることによって相互に通信してよい。

【００２２】

同様に、サブシステム２１６は、入力バッファ２１８および出力バッファ２１９を有する通信インターフェース２１７を介してネットワーク２１３へ接続されてよい。外部システムは、インターフェース２２０を介してネットワーク２１３へ接続されてよい。このように、コア２０１．．．２０９は、メッセージをパスすることによってサブシステムおよび外部システムと通信してよい。

【００２３】

メッセージ・パッシング（ＭＰ）アドレスは、コアに対してローカルなネットワーク・アドレスに関連する。例えば、個々のコアがチップ上のそのＸ、Ｙ位置によって特定されてよく、一方でローカル・アドレスが個々のコアにローカルなバッファまたはメモリのために用いられてよい。

【００２４】

次に図３を参照すると、本開示の実施形態によるニューロシナプティック・コアが示される。いくつかの実施形態において、ニューロシナプティック・コア３００は、行として表される軸索３０１、列として表される樹状突起３０２、行－列接合部として表されるシナプス３０３、および入力を樹状突起から受信するニューロン３０４を含む。いくつかの実施形態では、２５６個の軸索、および２５６個のニューロンがある。かかる実施形態では、２５６×２５６＝６５，５３６個のシナプスがある。情報は、軸索３０１からニューロン３０４へ流れ、シナプス３０３によって変調される。様々な実施形態において、シナプスは、バイナリであってよく、シナプスの重みと関連付けられてよい。

【００２５】

いくつかの実施形態では、複数のニューロシナプティック・コアがチップ上にタイルされる。例示的な実施形態において、合計１，０４８，５７６個のニューロンおよび２６８，４３５，４５６個のシナプスに対して、６４×６４グリッドのコアがタイルされて、４，０９６個のコアを生じさせる。かかる実施形態では、ニューロン、シナプス、および短距離接続がコア回路によって実装される。長距離接続は、論理的である。例示的な実施形態が図４に示される。メッシュ・ルータ４０１は、コア間の通信を提供する。所与のコア上に、ニューロンからコア４０２へおよびコアから軸索４０３への通信リンクも設けられる。

【００２６】

次に図５を参照すると、本開示の様々な実施形態による組み合わされたメモリ・マップドおよびメッセージ・パッシング・システム５００が示される。グローバル・メモリ・マップ１０１内のメモリ・スペース５０１がメッセージ・パッシング・システムとインターフェースするために提供される。インターフェース２２０は、アドレスをメモリ・マップド・システムおよびメッセージ・パッシング・システムのドメイン間で変換する。このように、メッセージ・パッシング・システム中の各コアのメッセージ・パッシングＩ／Ｏバッファは、メモリ・マップ１０１中に現れ、それを通してアクセス可能である。

【００２７】

次に図６を参照すると、本開示の様々な実施形態による例示的なコア２０１の入力バッファ２１４および出力バッファ２１５が示される。この例示的なコアでは、計算コア２１０が、例えば、図３～４に関して先に記載されたニューロシナプティック・システムにおけるような、ニューロン計算を提供する。通信インターフェース２１２は、ニューロン計算２１０からのニューロン・スパイク事象を提供する。先に記載されたようなニューロシナプティック・システムでは、ローカル・メモリ２１１がシナプス・クロスバーおよびコ

アに対するローカル・パラメータを含む。

【００２８】

この例示的な実施形態において、出力バッファ２１５は、複数の出力時間（T_o）における各出力ニューロン（N_o）からのデータを含む。入力バッファ２１４は、同様に、複数の入力時間（T_i）において各入力軸索（A_i）へ注入されるためのデータを含む。

【００２９】

次に図７を参照すると、本開示の様々な実施形態による例示的なメモリ・マップ１０１が示される。メモリ・マップ１０１のメッセージ・パッシング部分は、入力（軸索）領域７０２および出力（ニューロン）領域７０１へセグメント化される。C_o個の出力コアに対して、出力メモリ領域７０１のサイズは、$C_o * T_o * N_o * B_o$であり、ここでB_oは軸索当たりの出力データのビット数である。同様に、C_i個の入力コアに対して、入力メモリ領域のサイズは、$C_i * T_i * A_i * B_i$であり、ここでB_iは、ニューロン当たりの入力データのビット数である。

【００３０】

次に図８を参照すると、本開示の実施形態によるメッセージ・パッシング・システムのための例示的なメモリ・マップド・インターフェースが示される。インターフェース８０１は、バス８０２へ結合されて、それを通してアクセス可能である。メモリ・マップを介してアドレス可能な、インターフェース８０１において書き込みリクエストが受信される。書き込みリクエストは、メモリ・アドレスへ関する。インターフェース８０１のアドレス変換ユニット８０３がメモリ・アドレスをメッセージ・コア・アドレスへ変換する。上述のように、メモリ・アドレスは、メモリ内のオフセットの観点からであってよく、一方でコア・アドレスは、コア・ネーム、またはＸ、Ｙコア・アドレスを含んでよい。書き込みデータが、次に、アドレス変換ユニット８０３からのメッセージ・アドレス出力を用いて宛先コアへ送達される。先に考察されたように、そのデータは、ネットワーク８０４を通じて送信されて、次に、コア８０６に含まれる入力バッファ８０５へ加えられてよい。

【００３１】

インターフェース８０１は、出力バッファ８０７も含む。バッファ８０７は、メッセージ・パッシング・システムからのメッセージを記憶して、先に提示されたようにメモリ・マップド・システムによって直接に読み出されてよい。コア８０８は、ネットワーク８０４を介して、出力メッセージを出力バッファ８０７へプッシュする。いくつかの実施形態では、各コアが出力バッファを含み、一方でいくつかの実施形態では、コア上でのバッファリングなしに出力がネットワークへ直接に送られる。この例示的な実施形態では、軸索への取り込みのためにデータがコア入力バッファへ書き込まれ、出力ニューロンからそれへデータが読み出されるユニファイド出力バッファが設けられる。ユニファイド・バッファから、コア出力データがメモリ・マップを介して読み出し可能である。図示されるように、いくつかの実施形態では、出力バッファは、ユニファイドまたはセパレートに係わらず、個々のコア８０５、８０８中ではなくインターフェース・ブロック８０１中に物理的に位置する。ユニファイド出力バッファは、$C_o * T_o * N_o * B_o$のサイズを有する。

【００３２】

先に提示されたように、様々な実施形態がメモリ・マップド・ドメインからメッセージ・パッシング・ドメインへのアドレス変換を提供する。メモリ・マップド書き込みリクエストが受信される。メモリ・アドレスは、メッセージ・パケットをルーティングするために用いられるアドレスへ変換される。例えば、いくつかの実施形態において、Ｘ、Ｙコア・アドレスが提供される。いくつかの実施形態では、このアドレス変換がアルゴリズムで行われる。いくつかの実施形態では、このアドレス変換がルックアップ・テーブル（ＬＵＴ：ｌｏｏｋ－ｕｐ　ｔａｂｌｅ）を通して提供される。いくつかの実施形態では、アルゴリズムによるアドレス変換とルックアップ・テーブルとの組み合わせが用いられる。

【００３３】

先に提示されたように、様々な実施形態がメッセージ・パッシング・ドメインからメモリ・マップド・ドメインへのアドレス変換を提供する。メッセージ・パケット・アドレス

は、所与のコアの出力メッセージを出力バッファへルーティングするため、および出力バッファ中の適切なロケーション中へインデックスするために用いられる。出力バッファは、グローバル・メモリ・マップの一部であり、読み出しオペレーションによりメモリ・マップド・ドメイン中のバスを通じて読み出されてよい。出力バッファのセグメントが読み出された後に、メッセージ・パッシング・システムからのその後の書き込みに備えて、そのセグメントをクリアできる。

【００３４】

様々な実施形態において、インターフェース・ユニットは、メモリ・マップド・システムへ接続する標準バス・インターフェースを有する。様々な実施形態において、バスは、アドバンスト・エクステンシブル・インターフェース（ＡＸＩ：Ａｄｖａｎｃｅｄ ｅＸｔｅｎｓｉｂｌｅ Ｉｎｔｅｒｆａｃｅ）または他のアドバンスト・マイクロコントローラ・バス・アーキテクチャ（ＡＭＢＡ：Ａｄｖａｎｃｅｄ Ｍｉｃｒｏｃｏｎｔｒｏｌｌｅｒ Ｂｕｓ Ａｒｃｈｉｔｅｃｔｕｒｅ）規格に適合する。様々な実施形態において、インターフェース・ユニットは、メッセージ・パッシング・システムへ接続するインターフェースを有する。様々な実施形態において、インターフェースは、アドレス・イベント・リプレゼンテーション（ＡＥＲ：Ａｄｄｒｅｓｓ Ｅｖｅｎｔ Ｒｅｐｒｅｓｅｎｔａｔｉｏｎ）またはメッセージ・パッシング・インターフェース（ＭＰＩ：Ｍｅｓｓａｇｅ Ｐａｓｓｉｎｇ Ｉｎｔｅｒｆａｃｅ）のような、標準通信プロトコルに適合する。

【００３５】

本開示のシステムは、任意のバス・プロトコールをサポートする。それに応じて、いくつかの実施形態が直接メモリ・アクセス（ＤＭＡ：Ｄｉｒｅｃｔ Ｍｅｍｏｒｙ Ａｃｃｅｓｓ）を含み、一方でいくつかはそれを含まない。例えば、いくつかの実施形態は、単一のバス・トランザクションを実装する。いくつかの実施形態は、メモリ・マップドＩ／Ｏ（ＭＭＩＯ：Ｍｅｍｏｒｙ－Ｍａｐｐｅｄ Ｉ／Ｏ）またはポート・マップドＩ／Ｏ（ＰＭＩＯ：Ｐｏｒｔ－Ｍａｐｐｅｄ Ｉ／Ｏ）を実装する。いくつかの実施形態は、バースト・バス・トランザクション（例えば、ＤＭＡ）を実装する。ＤＭＡは、書き込みアドレス変換ユニットへ直接にブロック転送するか、またはフル・フレームを出力バッファから直接に読み出し転送することができて、それにより、ある一定のユースケースについて性能を向上させる。ＤＭＡを用いたいくつかの実施形態では、１つ以上のＤＭＡバス・マスタが設けられる。

【００３６】

本開示は、メモリ・マップド・システムおよびメッセージ・パッシング・システムをシステムオンチップ（ＳｏＣ：Ｓｙｓｔｅｍ－ｏｎ－Ｃｈｉｐ）中に集積化するのに役立つ。いくつかのかかる実施形態において、メッセージ・パッシング・システムは、標準バス（例えば、ＡＭＢＡまたはＡＸＩ）上に存在するＩＰブロックであってよい。

【００３７】

先に記載されたいくつかの実施形態では、インターフェース・ユニットを介して変換される、メモリ・マップド・ドメイン中のアドレスとメッセージ・パッシング・ドメイン中のアドレスとの間に１：１マッピングがある。しかしながら、いくつかの実施形態では、宛先ドメインがソース・ドメイン中のアドレスおよびデータの組み合わせを用いてアドレスされる。かかる実施形態では、メッセージそれ自体がアドレシング情報を含む。かかる実施形態では、アドレスを出力データ中に挿入する（または除去する）ために、ソース・ドメイン中のソース・プロセッサがそれらを明示的に計算する必要がある。これは、データ中のアドレスのいくつかをエンコードするのに必要なメモリ・スペースのサイズを削減する。

【００３８】

例えば、時間／ニューロン・アレイのメモリ・スペースへの１：１マッピングは、プロセッサがメッセージ・パッシング・ドメイン中のそれぞれのローカル・アドレスを個別かつ直接にアドレスすることを許容する。これは、大きいメモリ・スペースをもたらすが、プロセッサによるアドレス計算を必要としない。代わりに、すべてのアドレスをデータ中

にエンコードすることができる。このケースでは、メモリ・マップ・アドレス・スペースは、その後ろにＦＩＦＯを伴う単一のアドレスへ削減される。これは、小さいメモリ・スペースをもたらし、すべてのアドレスがプロセッサによって計算される。これらのアプローチを組み合わせると、メモリ・スペースのサイズがこれらの両極端間のどこかであることを許容し、２つのインターフェース間のデコーディング・ロジック（および計算）における複雑さと、どのくらいのメモリ・スペースが消費されるかとの間のトレードオフにつながる。

【００３９】

本開示のシステムおよび方法は、様々な周辺機器とのインテグレーションに適することが認識されよう。本開示によるバスは、例えば、センサ、アクチュエータ、ニューロシナプティック・システム、プロセッサ、または周辺機器のための複数のバス・マスタを有してよい。各システム構成要素にメモリ・スペースを割り当てられて、それがバスを制御しているときに、他の任意のシステム構成要素のメモリ・スペースへ送ることができる。アービタは、ある時点でどのバス・マスタがバスを有するかを制御する。いくつかの実施形態では、データを構成要素のメモリ・スペース間で転送するためにＤＭＡエンジンが用いられる。このように、従来のプロセッサ・システム、メッセージ・パッシング・システム、ならびにセンサおよびアクチュエータのような周辺デバイスの間でシームレスなシステム・インテグレーションが提供される。

【００４０】

本開示による組み合わされたメモリ・マップド／メッセージ・パッシング・システムは、（データ走行距離を最小化することによって）電力消費およびレイテンシを最小限に抑えて、一方ではスループットを維持するために同じダイ上に実装されてよい。必要に応じてスループット目標を満たすために、様々な実施形態においてインターフェースにおけるパイプライニングが提供される。かかるシステムは、ＰＣＭ、メモリスタ、ナノチューブ、またはオプティックスのような様々なプロセス技術の使用を許容するために、別々のダイまたはチップとして実装されてもよい。いくつかの実施形態では、より長い通信距離に対するインターフェース・ブロック性能を向上させるためにシリアライザ／デシリアライザ（ＳｅｒＤｅｓ：Ｓｅｒｉａｌｉｚｅｒ／Ｄｅｓｅｒｉａｌｉｚｅｒ）が用いられる。様々な実施形態によれば、Ｅｔｈｅｒｎｅｔ（Ｒ）およびＰＣＩｅ（Ｒ）のようなインターフェースを用いて物理的通信プリミティブが実装されてよい。

【００４１】

様々な代わりの実施形態では、軸索への取り込みのためにコア入力バッファへ直接に書き込みが行われ、一方でニューロン出力は、コア出力バッファから読み出される。かかる実施形態では、読み出しオペレーションが読み出しリクエスト・メッセージを生成し、アドレスを保存して、読み出しリクエスト・メッセージをコアへ送る。コアは、読み出しアドレスに相関付けられた、データ・メッセージで応答する。そのデータが、次に、バス上で返される。これは、一般に、単一のインターフェースの実施形態において出力バッファを維持するよりも多くのオーバーヘッドをもたらし、従って、より高いレイテンシを有する。

【００４２】

様々な代わりの実施形態では、軸索への取り込みのためにコア入力バッファへ書き込みが行われ、一方でニューロン出力は、主メモリ中の出力バッファから読み出される。かかる実施形態では、ＤＭＡを用いて、メッセージ・パッシング出力が主メモリへプッシュされる。これは、出力バッファにとって、特に、空間的にスパースな出力にとって必要とされるメモリのサイズを最小限に抑える。しかしながら、かかるＤＭＡエンジンは、任意のメモリ・アクセス（スキャッタ・ライト）を行うことが可能でなければならない。このオペレーションは、一般に、非効率的であり、従って、かかる実施形態は、一般に、インターフェース中に単一の出力バッファを含むものよりも性能が低い。

【００４３】

次に図９を参照すると、本開示の実施形態によるメモリ・マップド・システムとメッセ

ージ・パッシング・システムとをインターフェースする方法が示される。９０１では、書き込みリクエストが受信される。書き込みリクエストは、第１のデータおよび第１の目標アドレスを含む。第１の目標アドレスは、メモリ・マップ内のメモリ・アドレスである。９０２では、第１の目標アドレスが第２の目標アドレスへ変換される。第２の目標アドレスは、宛先コアを特定する。９０３では、第１のデータが第２の目標アドレスに従って、メッセージ・パッシング・システムを介して、宛先コアへ送られる。

【００４４】

９０４では、メッセージが、メッセージ・パッシング・システムを介して、ソース・コアから受信される。メッセージは、第２のデータを含む。９０５では、バッファ中のロケーションがソース・コアに基づいて決定される。９０６では、第２のデータがバッファ中のその第１のロケーションに記憶される。バッファは、メモリ・マップを介してアクセス可能である。任意の実施形態において、ステップ９０１～９０３は、ステップ９０４～９０６とは別に実施されてよく、同様に、所与の実施形態は、メッセージがメッセージ・パッシング・システムから、またはそれへ流れているかどうかに従って、ステップ９０１～９０３のみ、または９０４～９０６のみを含んでよいことが認識されよう。

【００４５】

様々な実施形態によれば、メモリ・マップド（ＭＭ）システムとメッセージ・パッシング（ＭＰ）システムとをシームレスに相互接続するための方法およびシステムが提供される。いくつかの実施形態において、メッセージ・パッシング・システムは、ニューロシナプティック・システムである。いくつかの実施形態において、メモリ・マップド・システムおよびメッセージ・パッシング・システムは、システムオンチップ（ＳｏＣ）上に集積化される。いくつかの実施形態において、インターフェース・ユニットは、メモリ・マップド・アドレスとメッセージ・パッシング・アドレスとの間で変換を行う。いくつかの実施形態において、ＭＭからＭＰへのアドレス変換は、アルゴリズムで、ルックアップ・テーブルによって、またはそれらの組み合わせによって行われる。いくつかの実施形態において、ＭＰからＭＭへのアドレス変換は、出力バッファを用いて行われる。いくつかの実施形態において、システム間のアドレス・マッピングは、アドレスとデータとの組み合わせによって提供される。いくつかの実施形態において、メモリ・マップド・システムは、複数のバス・マスタまたは周辺デバイスを含む。

【００４６】

次に図１０を参照すると、計算ノードの例の概略図が示される。計算ノード１０は、適切な計算ノードの一例に過ぎず、本明細書に記載される本発明の実施形態の使用の範囲または機能性についていずれかの限定を示唆することは意図されない。いずれにしても、計算ノード１０は、先に提示された機能性のいずれかが実装されるかもしくはそれを行うこと、またはその両方が可能である。

【００４７】

計算ノード１０には多くの他の汎用もしくは専用計算システム環境または構成によって動作可能なコンピュータ・システム／サーバ１２がある。コンピュータ・システム／サーバ１２とともに用いるのに適しうるよく知られた計算システム、環境、もしくは構成またはそれらの組み合わせの例は、以下には限定されないが、パーソナル・コンピュータ・システム、サーバ・コンピュータ・システム、シン・クライアント、シック・クライアント、ハンドヘルドまたはラップトップ・デバイス、マルチプロセッサ・システム、マイクロプロセッサベース・システム、セット・トップ・ボックス、プログラマブルな民生用電子機器、ネットワークＰＣ、ミニコンピュータ・システム、メインフレーム・コンピュータ・システム、および上記のシステムまたはデバイスのいずれかを含む分散型クラウド計算環境などを含む。

【００４８】

コンピュータ・システム／サーバ１２は、コンピュータ・システムによって実行される、プログラム・モジュールなど、コンピュータ・システムが実行可能な命令の一般的なコンテキストで記述されてよい。一般に、プログラム・モジュールは、特定のタスクを行う

か、または特定の抽象データ型を実装する、ルーチン、プログラム、オブジェクト、コンポーネント、ロジック、データ構造などを含んでよい。コンピュータ・システム／サーバ１２は、通信ネットワークを通してリンクされたリモート処理デバイスによってタスクが行われる、分散型クラウド計算環境において実行されてよい。分散型クラウド計算環境において、プログラム・モジュールは、メモリ記憶デバイスを含んだローカルおよびリモート両方のコンピュータ・システム記憶媒体中に位置してよい。

【００４９】

図１０に示されるように、計算ノード１０中のコンピュータ・システム／サーバ１２は、汎用計算デバイスの形態で示される。コンピュータ・システム／サーバ１２の構成要素は、以下には限定されないが、１つ以上のプロセッサまたは処理ユニット１６、システム・メモリ２８、およびシステム・メモリ２８を含めて様々なシステム構成要素をプロセッサ１６へ結合するバス１８を含んでよい。

【００５０】

バス１８は、メモリ・バスもしくはメモリ・コントローラ、周辺機器用バス、アクセラレイティッド・グラフィックス・ポート、および様々なバス・アーキテクチャのいずれかを用いたプロセッサまたはローカル・バスを含めて、いくつかのタイプのバス構造のいずれか１つ以上を表す。限定ではなく、例として、かかるアーキテクチャは、インダストリ・スタンダード・アーキテクチャ（ＩＳＡ：Ｉｎｄｕｓｔｒｙ Ｓｔａｎｄａｒｄ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、マイクロ・チャネル・アーキテクチャ（ＭＣＡ：Ｍｉｃｒｏ Ｃｈａｎｎｅｌ Ａｒｃｈｉｔｅｃｔｕｒｅ）バス、拡張ＩＳＡ（ＥＩＳＡ：Ｅｎｈａｎｃｅｄ ＩＳＡ）バス、ビデオ・エレクトロニクス・スタンダーズ・アソシエーション（ＶＥＳＡ：Ｖｉｄｅｏ Ｅｌｅｃｔｒｏｎｉｃｓ Ｓｔａｎｄａｒｄｓ Ａｓｓｏｃｉａｔｉｏｎ）ローカル・バス、およびペリフェラル・コンポーネント・インターコネクト（ＰＣＩ：Ｐｅｒｉｐｈｅｒａｌ Ｃｏｍｐｏｎｅｎｔ Ｉｎｔｅｒｃｏｎｎｅｃｔ）バスを含む。

【００５１】

コンピュータ・システム／サーバ１２は、典型的に、様々なコンピュータ・システム可読媒体を含む。かかる媒体は、コンピュータ・システム／サーバ１２によってアクセス可能ないずれかの利用可能な媒体であってよく、揮発性および不揮発性媒体、リムーバブルおよびノンリムーバブル媒体の両方を含む。

【００５２】

システム・メモリ２８は、ランダム・アクセス・メモリ（ＲＡＭ：ｒａｎｄｏｍ ａｃｃｅｓｓ ｍｅｍｏｒｙ）３０もしくはキャッシュ・メモリ３２またはその両方など、揮発性メモリの形態のコンピュータ・システム可読媒体を含むことができる。コンピュータ・システム／サーバ１２は、さらに、他のリムーバブル／ノンリムーバブル、揮発性／不揮発性コンピュータ・システム記憶媒体を含んでよい。単に例として、（示されず、典型的に「ハード・ドライブ」と呼ばれる）ノンリムーバブル不揮発性磁気媒体から読み出し、それに書き込むための記憶システム３４を設けることができる。示されないが、リムーバブル不揮発性磁気ディスク（例えば、「フレキシブル・ディスク」）から読み出し、それに書き込むための磁気ディスク・ドライブ、およびＣＤ－ＲＯＭ、ＤＶＤ‐ＲＯＭまたは他の光媒体などのリムーバブル不揮発性光ディスクから読み出し、またはそれに書き込むための光ディスク・ドライブを設けることができる。かかる事例では、１つ以上のデータ媒体インターフェースによって各々をバス１８へ接続できる。以下にさらに描かれ、記載されるように、メモリ２８は、本発明の実施形態の機能を実施するように構成された一組の（例えば、少なくとも１つの）プログラム・モジュールを有する少なくとも１つのプログラム製品を含んでよい。

【００５３】

一組の（少なくとも１つの）プログラム・モジュール４２、ならびにオペレーティング・システム、１つ以上のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データを有する、プログラム／ユーティリティ４０が、限定ではなく

、例として、メモリ２８中に記憶されてよい。オペレーティング・システム、１つ以上のアプリケーション・プログラム、他のプログラム・モジュール、およびプログラム・データの各々またはそれらの何らかの組み合わせは、ネットワーキング環境の実装を含んでよい。プログラム・モジュール４２は、一般に、本明細書に記載される本発明の実施形態の機能もしくは方法論またはその両方を実施する。

【００５４】

コンピュータ・システム／サーバ１２は、さらに、キーボード、ポインティング・デバイス、ディスプレイ２４など、１つ以上の外部デバイス１４、ユーザがコンピュータ・システム／サーバ１２と対話することを可能にする１つ以上のデバイス、もしくはコンピュータ・システム／サーバ１２が１つ以上の他の計算デバイスと通信することを可能にするいずれかのデバイス（例えば、ネットワーク・カード、モデムなど）、またはそれらの組み合わせと通信してよい。かかる通信は、入力／出力（Ｉ／Ｏ）インターフェース２２を介して発生しうる。さらにまた、コンピュータ・システム／サーバ１２は、ネットワーク・アダプタ２０を介して、１つ以上のネットワーク、例えば、ローカル・エリア・ネットワーク（ＬＡＮ：ｌｏｃａｌ　ａｒｅａ　ｎｅｔｗｏｒｋ）、一般的なワイド・エリア・ネットワーク（ＷＡＮ：ｗｉｄｅ　ａｒｅａ　ｎｅｔｗｏｒｋ）、もしくは公衆ネットワーク（例えば、インターネット）またはそれらの組み合わせと通信できる。図示のように、ネットワーク・アダプタ２０は、バス１８を介して、コンピュータ・システム／サーバ１２の他の構成要素と通信する。示されないが、コンピュータ・システム／サーバ１２と併せて、他のハードウェアもしくはソフトウェア構成要素またはその両方を用いてよいだろうということが理解されるべきである。例は、以下には限定されないが、マイクロコード、デバイス・ドライバ、冗長処理ユニット、外部ディスク・ドライブ・アレイ、ＲＡＩＤシステム、テープ・ドライブ、およびデータ・アーカイブ記憶システムなどを含む。

【００５５】

本発明は、システム、方法、もしくはコンピュータ・プログラム製品またはそれらの組み合わせであってよい。コンピュータ・プログラム製品は、プロセッサに本発明の様態を実施させるためのコンピュータ可読プログラム命令をその上に有する１つ（または複数）のコンピュータ可読記憶媒体を含んでよい。

【００５６】

コンピュータ可読記憶媒体は、命令実行デバイスによる使用のための命令を保持して記憶できる有形のデバイスとすることができる。コンピュータ可読記憶媒体は、例えば、以下には限定されないが、電子記憶デバイス、磁気記憶デバイス、光記憶デバイス、電磁記憶デバイス、半導体記憶デバイス、またはそれらのいずれか適切な組み合わせであってよい。コンピュータ可読記憶媒体のより具体的な例の非網羅的リストは、ポータブル・コンピュータ・ディスケット、ハード・ディスク、ランダム・アクセス・メモリ（ＲＡＭ）、リードオンリ・メモリ（ＲＯＭ：ｒｅａｄ‐ｏｎｌｙ　ｍｅｍｏｒｙ）、消去可能なプログラマブル・リードオンリ・メモリ（ＥＰＲＯＭ：ｅｒａｓａｂｌｅ　ｐｒｏｇｒａｍｍａｂｌｅ　ｒｅａｄ‐ｏｎｌｙ　ｍｅｍｏｒｙまたはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（ＳＲＡＭ：ｓｔａｔｉｃ　ｒａｎｄｏｍ　ａｃｃｅｓｓ　ｍｅｍｏｒｙ）、ポータブル・コンパクト・ディスク・リードオンリ・メモリ（ＣＤ－ＲＯＭ：ｃｏｍｐａｃｔ　ｄｉｓｃ　ｒｅａｄ‐ｏｎｌｙ　ｍｅｍｏｒｙ）、デジタル・バーサタイル・ディスク（ＤＶＤ：ｄｉｇｉｔａｌ　ｖｅｒｓａｔｉｌｅ　ｄｉｓｋ）、メモリ・スティック、フレキシブル・ディスク、機械的にエンコードされたデバイス、例えば、パンチカードもしくはその上に記録された命令を有する溝中の凸構造、ならびにそれらのいずれかの適切な組み合わせを含む。コンピュータ可読記憶媒体は、本明細書では、それ自体が、電波または他の自由に伝搬する電磁波、導波路もしくは他の伝送媒体を通して伝搬する電磁波（例えば、光ファイバ・ケーブルを通過する光パルス）、または電線を通して伝送される電気信号のような、一時的な信号であるとして解釈されるべきではない。

【００５７】

本明細書に記載されるコンピュータ可読プログラム命令をコンピュータ可読記憶媒体か

らそれぞれの計算／処理デバイスへ、あるいは、ネットワーク、例えば、インターネット、ローカル・エリア・ネットワーク、ワイド・エリア・ネットワークもしくは無線ネットワークまたはそれらの組み合わせを介して、外部コンピュータもしくは外部記憶デバイスへダウンロードできる。ネットワークは、銅伝送ケーブル、光伝送ファイバ、無線伝送、ルータ、ファイアウォール、スイッチ、ゲートウェイ・コンピュータもしくはエッジ・サーバまたはそれらの組み合わせを備えてよい。各計算／処理デバイス中のネットワーク・アダプタ・カードもしくはネットワーク・インターフェースは、コンピュータ可読プログラム命令をネットワークから受信して、それらのコンピュータ可読プログラム命令をそれぞれの計算／処理デバイス内のコンピュータ可読記憶媒体中に記憶するために転送する。

【００５８】

本発明のオペレーションを実施するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セットアーキテクチャ（ＩＳＡ：ｉｎｓｔｒｕｃｔｉｏｎ‐ｓｅｔ‐ａｒｃｈｉｔｅｃｔｕｒｅ）命令、機械命令、機械依存命令、マイクロコード、ファームウェア命令、状態設定データ、あるいは、Ｓｍａｌｌｔａｌｋ（Ｒ）、Ｃ＋＋もしくは同様のものなどのオブジェクト指向プログラミング言語、および「Ｃ」プログラミング言語のような従来の手続き型プログラミング言語または同様のプログラミング言語を含めて、１つ以上のプログラミング言語のいずれかの組み合わせで書かれたソース・コードもしくはオブジェクト・コードのいずれかであってよい。コンピュータ可読プログラム命令は、専らユーザのコンピュータ上で、部分的にユーザのコンピュータ上で、スタンドアロン・ソフトウェア・パッケージとして、実行してもよく、部分的にユーザのコンピュータ上および部分的にリモート・コンピュータ上で、あるいは専らリモート・コンピュータまたはサーバ上で実行してもよい。後者のシナリオでは、リモート・コンピュータは、ローカル・エリア・ネットワーク（ＬＡＮ）もしくはワイド・エリア・ネットワーク（ＷＡＮ）を含めて、いずれかのタイプのネットワークを通してユーザのコンピュータへ接続されてもよく、または（例えば、インターネット・サービス・プロバイダを用いたインターネットを通して）外部コンピュータへの接続が行われてもよい。いくつかの実施形態では、例えば、プログラマブル・ロジック回路素子、フィールドプログラマブル・ゲート・アレイ（ＦＰＧＡ：ｆｉｅｌｄ－ｐｒｏｇｒａｍｍａｂｌｅ　ｇａｔｅ　ａｒｒａｙ）、またはプログラマブル・ロジック・アレイ（ＰＬＡ：ｐｒｏｇｒａｍｍａｂｌｅ　ｌｏｇｉｃ　ａｒｒａｙ）を含む電子回路素子が、本発明の態様を行うために、コンピュータ可読プログラム命令の状態情報を利用してその電子回路素子をパーソナライズすることによって、コンピュータ可読プログラム命令を実行してよい。

【００５９】

本発明の態様は、本明細書において本発明の実施形態による方法、装置（システム）、およびコンピュータ・プログラム製品のフローチャート説明図もしくはブロック図またはその両方を参照して記載される。フローチャート説明図もしくはブロック図またはその両方の各ブロック、およびフローチャート説明図もしくはブロック図またはその両方におけるブロックの組み合わせをコンピュータ可読プログラム命令によって実装できることが理解されよう。

【００６０】

これらのコンピュータ可読プログラム命令は、コンピュータまたは他のプログラマブル・データ処理装置のプロセッサを介して実行するそれらの命令がフローチャートもしくはブロック図またはその両方の１つまたは複数のブロックにおいて指定された機能／動作を実装するための手段を生み出すような、マシンを作り出すように、汎用コンピュータ、専用コンピュータ、もしくは他のプログラマブル・データ処理装置のプロセッサへ提供されてよい。これらのコンピュータ可読プログラム命令は、さらに、コンピュータ可読ストレージ媒体中に記憶されてよく、これらのプログラム命令は、その中に記憶された命令を有するコンピュータ可読ストレージ媒体がフローチャートもしくはブロック図またはその両方の１つもしくは複数のブロックにおいて指定された機能／動作の態様を実装する命令を含んだ製造品を備えるような、特定の仕方で機能するように、コンピュータ、プログラマ

ブル・データ処理装置、もしくは他のデバイスまたはそれらの組み合わせに命令できる。

【００６１】

コンピュータ可読プログラム命令は、さらに、コンピュータ、他のプログラマブル・データ処理装置、または他のデバイス上で実行する命令がフローチャートもしくはブロック図またはその両方の１つもしくは複数のブロックにおいて指定された機能／動作を実装するような、コンピュータ実装プロセスを作り出すための一連のオペレーション・ステップがコンピュータ、他のプログラマブル・データ処理装置、または他のデバイス上で行われるようにするために、コンピュータ、他のプログラマブル・データ処理装置、または他のデバイス上へロードされてよい。

【００６２】

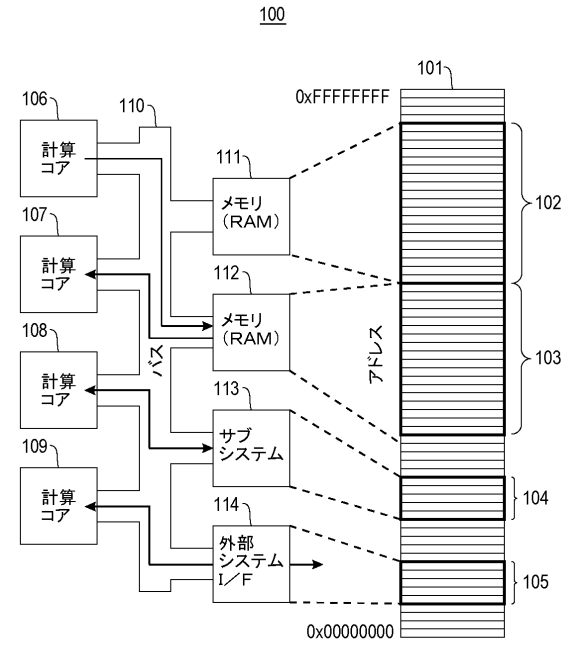
図中のフローチャートおよびブロック図は、本発明の様々な実施形態によるシステム、方法、およびコンピュータ・プログラム製品の可能な実装のアーキテクチャ、機能性、およびオペレーションを示す。この点で、フローチャートまたはブロック図中の各ブロックは、指定された論理機能（単数または複数）を実装するための１つ以上の実行可能な命令を含む、モジュール、セグメント、または命令の一部を表してよい。いくつかの代わりの実装において、ブロック中に記された機能は、図中に記された順序以外で発生してもよい。例えば、連続して示される２つのブロックが、実際には、実質的に同時に実行されてもよく、または関与する機能性に依存して、複数のブロックがときには逆の順序で実行されてもよい。ブロック図もしくはフローチャート説明図またはその両方の各ブロック、およびブロック図もしくはフローチャート説明図またはその両方におけるブロックの組み合わせを、指定された機能または動作を行うかあるいは専用ハードウェアとコンピュータ命令との組み合わせを実施する専用ハードウェア・ベースのシステムによって実装できることにも気付くであろう。

【００６３】

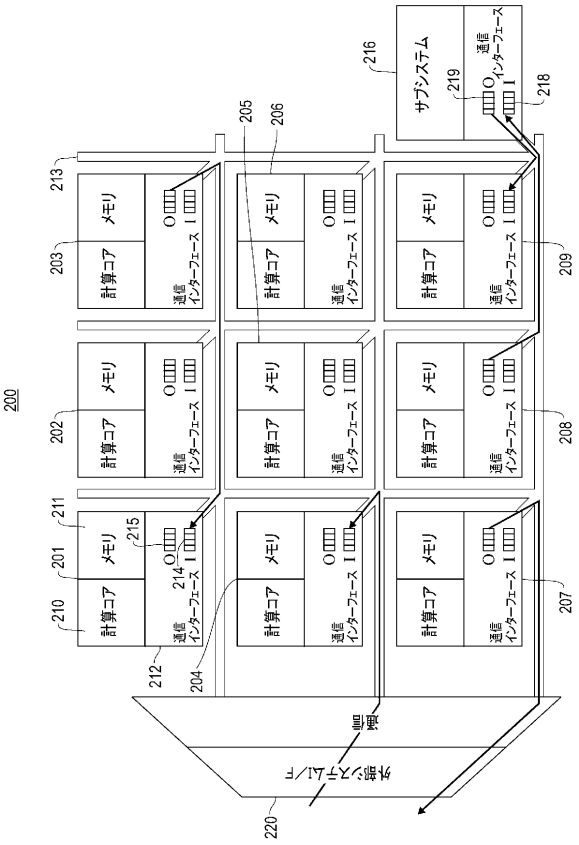
本発明の様々な実施形態の記載が説明のために提示されたが、これらの記載が網羅的であり、または開示される実施形態に限定されることは意図されない。記載された実施形態の範囲および趣旨から逸脱することなく、多くの修正および変更が当業者には明らかであろう。本明細書に用いられる用語法は、実施形態の原理、実用用途または市場に見られる技術を超える技術的改良を最もよく説明するために、あるいは本明細書に開示される実施形態を当業者が理解できるようにするために選ばれた。

【図面】

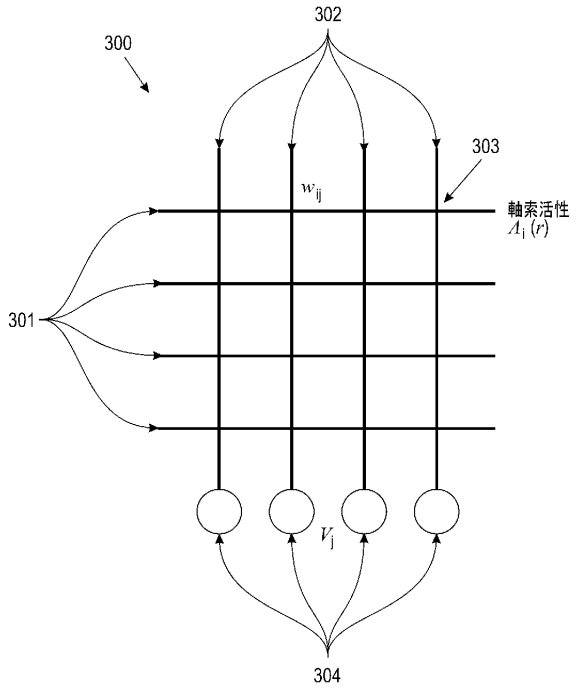
【図１】


100
101
0xFFFFFFFF
0x00000000
106
107
108
109
110
111
112
113
114
計算コア
メモリ（RAM）
サブシステム
外部システムI／F
バス
アドレス
102
103
104
105

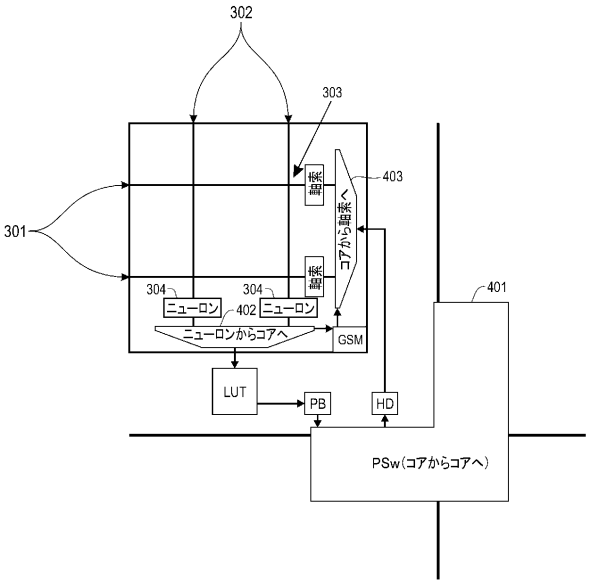
【図２】


200
216
サブシステム
219
218
通信インターフェース
205
206
213
203
209
202
208
211
201
215
214
207
210
212
204
メモリ
計算コア
バス
外部システムI／F
220

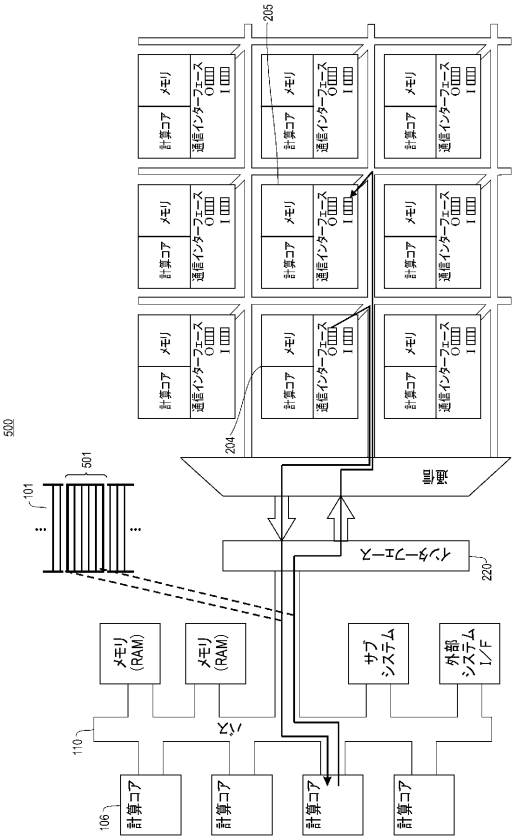
【図３】


300
302
303
w_{ij}
軸索活性
$A_i(r)$
301
V_j
304

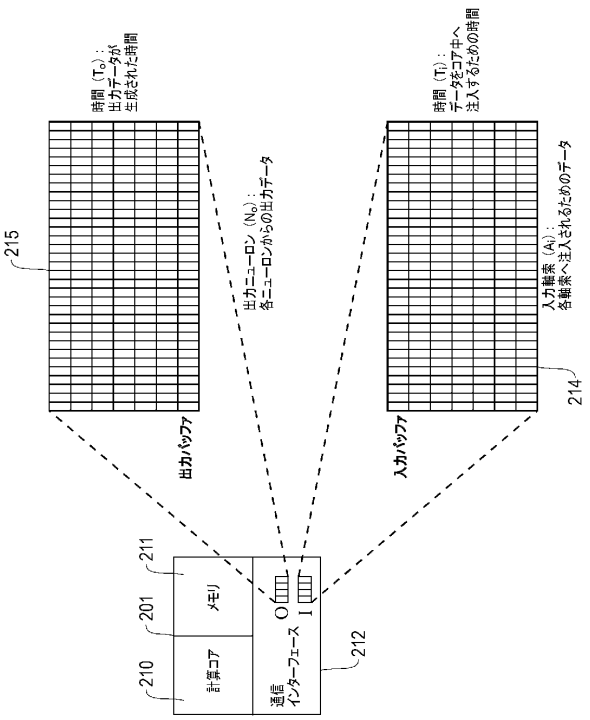
【図４】


302
303
403
301
樹素
コアから樹素へ
304
ニューロン
402
ニューロンからコアへ
GSM
401
LUT
PB
HD
PSw（コアからコアへ）

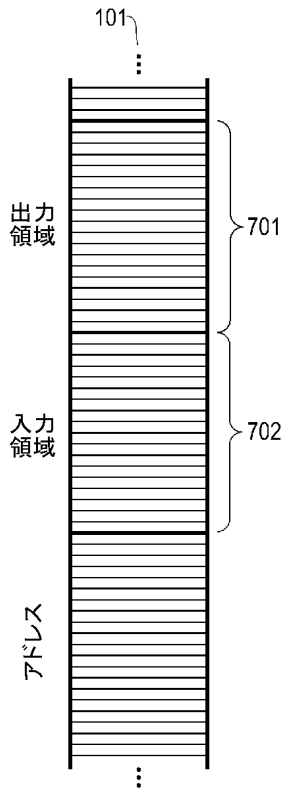
【図５】


500
101
501
205
204
メモリ
計算コア
通信インターフェース
O
I
バス
インターフェース
220
メモリ (RAM)
サブシステム
外部システム I/F
110
106
計算コア

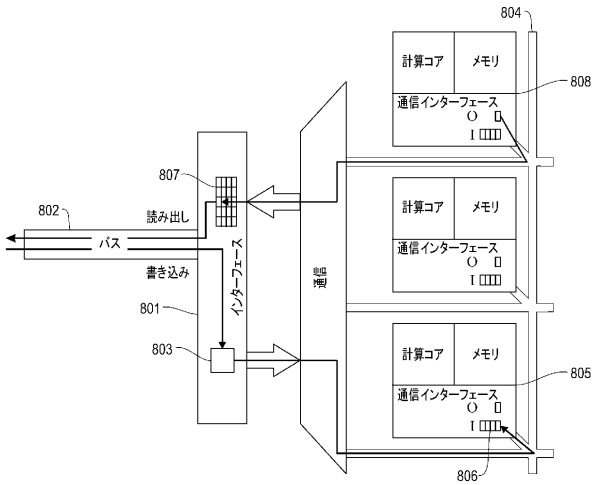
【図６】


時間（To）：出力データが生成された時間
時間（Ti）：データをコア中へ注入するための時間
出力ニューロン（No）：各ニューロンからの出力データ
入力素子（Ai）：各素子へ注入されるためのデータ
215
214
出力バッファ
入力バッファ
211
201
210
212
メモリ
計算コア
通信インターフェース
O
I

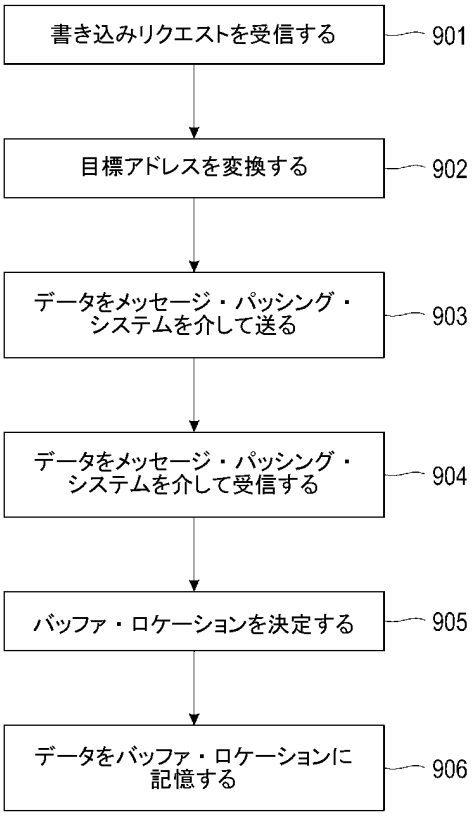
【図７】


101
出力領域
入力領域
アドレス
701
702

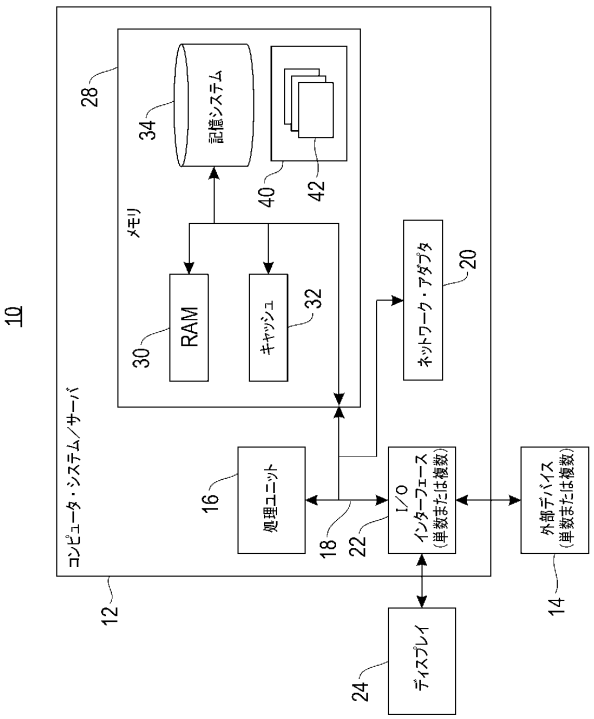
【図８】


804
計算コア
メモリ
808
通信インターフェース
O
I
807
802
読み出し
バス
書き込み
801
803
インターフェース
通信
805
806

【図９】


書き込みリクエストを受信する
901
目標アドレスを変換する
902
データをメッセージ・パッシング・システムを介して送る
903
データをメッセージ・パッシング・システムを介して受信する
904
バッファ・ロケーションを決定する
905
データをバッファ・ロケーションに記憶する
906

【図１０】


10
コンピュータ・システム／サーバ
12
メモリ
28
記憶システム
34
40
42
RAM
30
キャッシュ
32
ネットワーク・アダプタ
20
処理ユニット
16
I/O インターフェース（単数または複数）
18
22
外部デバイス（単数または複数）
14
ディスプレイ
24

フロントページの続き

(72)発明者　キャシディ、アンドリュー、ステファン
アメリカ合衆国７８７５８－３４００ テキサス州 オースティン バーネット・ロード １１５０１
(72)発明者　アコピャン、フィリップ
アメリカ合衆国１０５９８ ニューヨーク州 ヨークタウン・ハイツ キッチャワン・ロード １１０１
(72)発明者　アーサー、ジョン、ヴァーノン
アメリカ合衆国９５１２０－６０９９ カリフォルニア州 サンノゼ ハリー・ロード ６５０
(72)発明者　デーボレ、マイケル、ヴィンセント
アメリカ合衆国９５１２０－６０９９ カリフォルニア州 サンノゼ ハリー・ロード ６５０
(72)発明者　メローラ、ポール
アメリカ合衆国９５１２０－６０９９ カリフォルニア州 サンノゼ ハリー・ロード ６５０
(72)発明者　モダ、ダルメーンドラ
アメリカ合衆国９５１２０－６０９９ カリフォルニア州 サンノゼ ハリー・ロード ６５０
(72)発明者　澤田 潤
アメリカ合衆国７８７５８－３４００ テキサス州 オースティン バーネット・ロード １１５０１

審査官　関口 明紀

(56)参考文献　特開平０５－２４２０６５（ＪＰ，Ａ）
米国特許出願公開第２０１５／０３２４６８４（ＵＳ，Ａ１）
米国特許出願公開第２０１６／０１９６４８８（ＵＳ，Ａ１）
Martin J. Pearson et al.，Implementing Spiking Neural Networks for Real-Time Signal-Processing and Control Applications: A Model-Validated FPGA Approach，IEEE Transactions on Neural Networks，IEEE，2007年09月04日，Vol.18, No.5，pages 1472-1487

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｎ　３／０６３
Ｇ０６Ｆ　１５／１７３
Ｇ０６Ｆ　９／３８
Ｇ０６Ｆ　１３／３８