(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7442625号
(P7442625)

(45)発行日　令和6年3月4日(2024.3.4)

(24)登録日　令和6年2月22日(2024.2.22)

(51)国際特許分類		FI		
G11C 13/00 (2006.01)		G11C	13/00	340
G06N 3/063(2023.01)		G06N	3/063	
H10B 63/10 (2023.01)		G11C	13/00	210
		H10B	63/10	

請求項の数　12　（全20頁）

(21)出願番号　特願2022-513039(P2022-513039)
(86)(22)出願日　令和2年7月17日(2020.7.17)
(65)公表番号　特表2022-546032(P2022-546032A)
(43)公表日　令和4年11月2日(2022.11.2)
(86)国際出願番号　PCT/IB2020/056747
(87)国際公開番号　WO2021/038334
(87)国際公開日　令和3年3月4日(2021.3.4)
審査請求日　令和4年12月23日(2022.12.23)
(31)優先権主張番号　16/554,018
(32)優先日　令和1年8月28日(2019.8.28)
(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】　相変化メモリ・シナプスのプログラム中にドリフト係数外れ値を抑制すること

(57)【特許請求の範囲】

【請求項１】

相変化メモリ（ＰＣＭ）デバイスのドリフト外れ値を抑制するためのコンピュータ実装方法であって、

制御装置によって前記ＰＣＭデバイスのコンダクタンスをプログラムすることを含み、前記プログラムすることは、

第１の時点において、前記ＰＣＭデバイスの前記コンダクタンスを第１のコンダクタンス値となるように構成することであって、前記第１の時点はプログラミング時点である、前記構成することと、

第１の事前補償時点において、前記ＰＣＭデバイスの前記コンダクタンスが、目標コンダクタンス値との差が所定閾値以内である第２のコンダクタンス値に変化していると判定することと、

第２の事前補償時点において前記コンダクタンスを再び測定することを含めて、第２の時点において、前記ＰＣＭデバイスを前記第１のコンダクタンス値となるようにプログラムし直すことと

を含む、コンピュータ実装方法。

【請求項２】

前記第１の事前補償時点は前記プログラミング時点後の所定時間期間にある、請求項１に記載の方法。

【請求項３】

前記第１の事前補償時点における前記目標コンダクタンス値は、前記プログラミング時点後の第２の所定期間を有する目標時間窓に基づいて決定され、前記ＰＣＭデバイスの前記コンダクタンスは前記目標時間窓内で特定の範囲内にあることが望まれる、請求項２に記載の方法。

【請求項４】

プログラムすることは、（ｉ）前記第１の事前補償時点における前記第２のコンダクタンス値と（ｉｉ）前記目標コンダクタンス値との間の差が前記所定閾値未満になるまで継続される、請求項２に記載の方法。

【請求項５】

前記プログラムすることは、

前記第１の事前補償時点において、前記第１の事前補償時点における前記第２のコンダクタンス値と前記目標コンダクタンス値との間の差が前記所定閾値未満であると判定することと、

第１のチェックポイントにおいて、前記ＰＣＭデバイスの前記コンダクタンスが第２の目標コンダクタンス値との差が前記所定閾値以内である第３のコンダクタンス値に変化していると判定することと、

前記コンダクタンスを再び測定することを含めて、第３の時点において、前記ＰＣＭデバイスを前記第１のコンダクタンス値となるようにプログラムすることと、を更に含む、請求項１に記載の方法。

【請求項６】

前記第１のチェックポイントは前記プログラミング時点後の第２の所定時間期間にあたる、請求項５に記載の方法。

【請求項７】

前記第１のチェックポイントにおける前記第２の目標コンダクタンス値は、前記プログラミング時点以降の第２の所定期間にあたる目標時間窓に基づいて決定され、前記ＰＣＭデバイスの前記コンダクタンスは前記目標時間窓内で特定の範囲内に維持される、請求項５に記載の方法。

【請求項８】

前記ＰＣＭデバイスは人工ニューラル・ネットワーク・システムにおけるシナプスとして使用され、前記コンダクタンスは前記シナプスに割り当てられる重みである、請求項１に記載の方法。

【請求項９】

前記ＰＣＭデバイスは複数のＰＣＭデバイスを備え、前記複数のＰＣＭデバイスの各々は対応する目標コンダクタンス値と関連付けられている、請求項１に記載の方法。

【請求項１０】

前記ＰＣＭデバイスの前記コンダクタンスはコンダクタンス・ドリフトに起因して前記第２のコンダクタンス値に変化する、請求項１に記載の方法。

【請求項１１】

制御装置と、１つまたは複数のクロスバー・アレイを備えるコプロセッサと、を備えるシステムであって、前記制御装置は前記コプロセッサを使用して人工ニューラル・ネットワークの層をクロスバー・アレイとマッピングすることによって前記人工ニューラル・ネットワークを実装するように構成されており、前記人工ニューラル・ネットワークを実装することは請求項１ないし１０のいずれかに記載の方法を含む、システム。

【請求項１２】

相変化メモリ（ＰＣＭ）デバイスのドリフト外れ値を抑制するよう、コンピュータの処理回路に、請求項１ないし１０のいずれかに記載の方法を実行させる、コンピュータ・プログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

本発明は一般にコンピュータ技術に関し、より詳細には、相変化メモリ（ＰＣＭ）をプログラムすることに、およびＰＣＭのプログラム中にドリフト外れ値を抑制する方法に関する。

【背景技術】

【０００２】

ＰＣＭは、別個の電気特性を有する２つの相の間で切り替わる特質を有する材料の特徴を利用する。例えば、これらの材料は、非晶質の無秩序な相と、結晶質または多結晶質の秩序的な相との間で、切り替わることができる。これら２つの相は大きく異なる抵抗値と関連付けられる。更に、材料の一部しか非晶質相にまたは結晶相に切り替わっていない中間構成を、中間の抵抗値と関連付けることができる。

【発明の概要】

【０００３】

本発明の１つまたは複数の実施形態によれば、相変化メモリ（ＰＣＭ）デバイスのドリフト外れ値を抑制するためのコンピュータ実装方法は、制御装置によってＰＣＭデバイスのコンダクタンスをプログラムすることを含み、プログラムすることは、第１の時点において、ＰＣＭデバイスのコンダクタンスを第１のコンダクタンス値となるように構成することであって、第１の時点はプログラミング時点である、構成することを含む。更に、プログラムすることは、第１の事前補償時点において、ＰＣＭデバイスのコンダクタンスが、目標コンダクタンス値との差が所定閾値以内である第２のコンダクタンス値に変化していると判定することを含む。また更に、プログラムすることは、上記判定に基づいて、第２の事前補償時点において、上記事前補償を再び測定することを含めて、第２の時点において、ＰＣＭデバイスを第１のコンダクタンス値となるようにプログラムし直すことを含む。ＰＣＭデバイスのコンダクタンスは、コンダクタンス・ドリフトに起因して第２のコンダクタンス値に変化する。

【０００４】

１つまたは複数の例では、事前補償時点は、プログラミング時点後の所定時間期間にある。更に、１つまたは複数の例では、第１の事前補償時点における目標コンダクタンス値は、プログラミング時点後の第２の所定期間にあたる目標時間窓に基づいて決定される。ＰＣＭデバイスのコンダクタンスは、目標時間窓内で特定の範囲内に維持されることになる。

【０００５】

１つまたは複数の例では、プログラムすることは、（ｉ）事前補償時点における第２のコンダクタンス値と（ｉｉ）目標コンダクタンス値との間の差が所定閾値未満になるまで継続される。

【０００６】

事前補償時点において、事前補償時点における第２のコンダクタンス値と目標コンダクタンス値との間の差が所定閾値未満である場合、方法は、第１のチェックポイントにおいて、ＰＣＭデバイスのコンダクタンスが、第２の目標コンダクタンス値との差が所定閾値以内である第３のコンダクタンス値に変化していると判定することを更に含む。方法は、上記事前補償を再び測定することを含めて、第３の時点において、ＰＣＭデバイスを第１のコンダクタンス値となるようにプログラムすることを更に含む。

【０００７】

１つまたは複数の例では、第１のチェックポイントは、プログラミング時点後の第２の所定時間期間にある。第１のチェックポイントにおける第２の目標コンダクタンス値は、プログラミング時点以降の第２の所定期間にあたる目標時間窓に基づいて決定され、ＰＣＭデバイスのコンダクタンスは目標時間窓内で特定の範囲内に維持される。

【０００８】

１つまたは複数の例では、ＰＣＭデバイスは人工ニューラル・ネットワーク・システムにおけるシナプスとして使用され、コンダクタンスはシナプスに割り当てられる重みである。

【０００９】
１つまたは複数の例では、ＰＣＭデバイスは複数のＰＣＭデバイスを備え、各ＰＣＭデバイスは対応する目標コンダクタンス値と関連付けられている。
【００１０】
上記した特徴はまた、少なくともシステム、コンピュータ・プログラム製品、および機械によっても提供され得る。
【００１１】
本発明の１つまたは複数の実施形態によれば、コンピュータ実装方法は、クロスバー・アレイ中の複数の相変化メモリ（ＰＣＭ）デバイスから、あるＰＣＭデバイスのコンダクタンス値を、プログラミング時点である第１の時点において第１のコンダクタンス値となるようにＰＣＭデバイスのコンダクタンスを構成するための信号を受信することによって、構成することを含む。方法は、事前補償時点において、ＰＣＭデバイスのコンダクタンスが、目標コンダクタンス値との差が所定閾値以内である第２のコンダクタンス値に変化していると判定することを更に含む。方法は、上記事前補償を第２の事前補償時点において再び測定することを含めて、第２の時点において、ＰＣＭデバイスのコンダクタンスを第１のコンダクタンス値となるように構成するための別の信号を受信することを更に含む。
【００１２】
上記した特徴はまた、少なくともシステム、コンピュータ・プログラム製品、および機械によっても提供され得る。
【００１３】
本発明の１つまたは複数の実施形態はしたがって、特にＰＣＭデバイスをプログラムするおよびドリフト外れ値を抑制するための、コンピュータ技術の実際の応用および改善を容易にする。ＰＣＭデバイスの重みがどのようであるべきか知られている場合に本発明の実施形態を適用することができ、ＰＣＭデバイスはそれらが将来の適切な時間に適切な重み値に到達するような方法でプログラムされる。ＰＣＭデバイスは対数スケールで推移するので、（重みが使用される時点と比較して）比較的短い時間期間（例えば、１秒、１分、等）後であっても、ＰＣＭデバイスの重みの測定値を使用して、所望の目標時間窓（例えば、６ヶ月、１年、等）においてＰＣＭデバイスが所望の目標重みを有することになるかどうかの予測が可能になることが実現される。
【００１４】
事前補償時点は、プログラミング時点後の所定時間期間にあたる。事前補償時点における目標コンダクタンス値は、プログラミング時点後の第２の所定期間にあたる目標時間窓に基づいて決定される。ＰＣＭデバイスのコンダクタンスは、目標時間窓内で特定の範囲内に維持されることになる。
【００１５】
本発明の技術によって更なる技術的特徴および利益が実現される。本発明の実施形態および態様は本明細書に詳細に記載されており、特許請求される主題の一部と見なされる。よりよく理解するためには、詳細な説明および図面を参照されたい。
【００１６】
本明細書に記載される排他的権利の詳細は、本明細書の終結部に置く特許請求の範囲において具体的に指摘され、明確に特許請求される。本発明の実施形態の前述のおよび他の特徴および利点は、添付の図面と併せて解釈される続く詳細な説明から明らかである。
【図面の簡単な説明】
【００１７】
【図１】深層ニューラル・ネットワークのブロック図である。
【図２Ａ】本発明の１つまたは複数の実施形態に係る、アナログ・メモリのクロスバー・アレイを使用する深層ニューラル・ネットワークのブロック図である。
【図２Ｂ】本発明の１つまたは複数の実施形態に係る、アナログ・メモリのクロスバー・アレイを使用する深層ニューラル・ネットワークのブロック図である。
【図２Ｃ】本発明の１つまたは複数の実施形態に係る、アナログ・メモリのクロスバー・

アレイを使用する深層ニューラル・ネットワークのブロック図である。
【図３】本発明の１つまたは複数の実施形態に係る相変化メモリ・アレイ・デバイスのブロック図である。
【図４】本発明の１つまたは複数の実施形態に係る、複数のニューロンに結合されたクロスバー・アレイをネットワークとして使用して実装されたニューロモルフィック・システムの構造を表した図である。
【図５】本発明の１つまたは複数の実施形態に係る相変化メモリ・デバイスの抵抗／コンダクタンス値の例示のドリフトを表した図である。
【図６】本発明の１つまたは複数の実施形態に係る例示のデバイスに関するｖ係数を表した図である。
【図７】本発明の１つまたは複数の実施形態に係る、相変化メモリ・デバイスのプログラム中にドリフト係数外れ値を抑制するための方法を表したフローチャートである。
【図８】本発明の１つまたは複数の実施形態に従ってプログラムされている例示の相変化メモリ・デバイスを表した図である。
【図９】本発明の１つまたは複数の実施形態に係る、ドリフト外れ値の抑制を伴って相変化メモリ・デバイスをプログラムするためのシステムを表した図である。
【発明を実施するための形態】
【００１８】
本明細書に描かれている図は例示的なものである。これらの図またはそれらに記載されている動作には、本発明の思想から逸脱することなく多くの変形が存在し得る。例えば、アクションを異なる順序で実行することができるか、またはアクションを追加、削除、もしくは変更することができる。また、用語「結合された」およびその変形は２つの要素間に通信経路を有することを記述するものであり、それらの間に介在する要素／接続部を有さない要素間の直接的な接続は示唆していない。これらの変形は全て本明細書の一部であると見なされる。
【００１９】
添付の図および開示される実施形態の続く詳細な説明において、図に示される様々な要素には、２桁または３桁の参照番号が付されている。僅かな例外はあるが、各参照番号の最も左の数字は、その要素が最初に図示される図に対応している。
【００２０】
本発明の様々な実施形態が関連する図面を参照して本明細書に記載されている。本発明の範囲から逸脱することなく本発明の代替の実施形態を考案することができる。以下の説明および図面において、要素間の様々な接続および位置関係（例えば、上、下、隣、等）が説明される。これらの接続または位置関係あるいはその両方は、別に規定されていない限りは、直接的または間接的であり得、本発明はこの点に関して限定することは意図していない。したがって、エンティティの結合は直接的結合または間接的結合のいずれを指す場合もあり、エンティティ同士の位置関係は直接的な位置関係または間接的な位置関係であり得る。また更に、本明細書に記載する様々なタスクおよびプロセス・ステップは、本明細書に詳細に記載されていない追加のステップまたは機能性を有するより包括的な手順またはプロセスに組み込むことができる。
【００２１】
以下の定義および略語は、特許請求の範囲および明細書を解釈するために使用されるものである。本明細書で使用される場合、用語「備える（comprises）」、「備える（comprising）」、「含む（includes）」、「含む（including）」、「有する（has）」、「有する（having）」、「含有する（contains）」、もしくは「含有する（containing）」、またはその何らかの他の変形は、非排他的な包含を含むことが意図されている。例えば、列挙された要素を含む組成物、混合物、プロセス、方法、物品、または装置は必ずしもそれらの要素だけに限定されず、明示的に列挙されていないかまたはそのような組成物、混合物、プロセス、方法、物品、もしくは装置に固有の、他の要素を含み得る。
【００２２】

また加えて、用語「例示的」は本明細書では、「例、実例、または例示の役割を果たす」という意味で使用される。本明細書で「例示的」であるとして記載されるいかなる実施形態または設計も、必ずしも他の実施形態または設計に比べて好ましいかまたは有利であると解釈されるべきではない。用語「少なくとも１つの」および「１つまたは複数の」は、１よりも大きいかまたは１と等しい任意の整数、すなわち１、２、３、４、等を含むものと理解され得る。用語「複数」は、２よりも大きいかまたは２と等しい任意の整数、すなわち２、３、４、５、等を含むものと理解され得る。用語「接続」は、間接的な「接続」および直接的な「接続」の両方を含み得る。

【００２３】

用語「約（about）」、「実質的に」、「約（approximately）」、およびこれらの変形は、本願の出願の時点で利用可能な機器に基づく特定の量の測定に関連する誤差の程度を含むことが意図されている。例えば「約（about）」は、所与の値の±８％、または５％、または２％の範囲を含み得る。

【００２４】

簡潔にするために、本発明の態様の製作または使用に関連する従来の技術は、本明細書で詳細に説明する場合もあれば、しない場合もある。特に、本明細書に記載する様々な技術的特徴を実装するためのコンピューティング・システムおよび特定のコンピュータ・プログラムの様々な態様はよく知られている。したがって、簡潔にするために、本明細書では従来の実装の詳細の多くは簡潔にしか記述しないかまたは完全に省略し、よく知られているシステムまたはプロセスあるいはその両方の詳細は提供しない。

【００２５】

既に指摘したように、相変化メモリ（ＰＣＭ）素子は、別個の電気特性を有する２つの相の間で切り替わる特質を有する材料の特徴を利用する。現時点では、カルコゲニドまたはカルコゲニック材料と呼ばれる、ＴｅまたはＳｅなどの周期表の第６族の合金が、相変化セルにおいて有利に使用され得る。カルコゲニドでは、当該材料が非晶質相（抵抗がより高い）から結晶相（抵抗がより低い）へとおよびその逆に移行すると、抵抗値は２桁以上変化する。

【００２６】

そのような特性によって、ＰＣＭ素子は、デジタルおよびアナログのデータ・ストレージ用のメモリ・セルおよびメモリ・アレイを提供するのに非常に適したものとなる。特に、相変化メモリにおいて、カルコゲニド材料の一部は、第１の論理値および第２の論理値とそれぞれ関連付けられている高抵抗状態と低抵抗状態を切り替えるおよびその逆を行うように、制御された電流によって電気的に加熱可能な、プログラム可能な抵抗器として使用される。カルコゲニドの状態は、加熱が行われないような十分に低い電圧を印加することによって、およびそこに流れる電流を測定することによって、読み出すことができる。電流はカルコゲニド材料のコンダクタンスに比例するので、２つの状態を区別することが可能である。例えば、コンピュータおよび他のタイプのデジタル・ストレージ・デバイスにおいてなど、上書き可能なメモリ・デバイスに情報を記憶するために、Ｇｅ、Ｓｂ、およびＴｅの合金（Ｇｅ２Ｓｂ２Ｔｅ５、ＧＳＴ）によって形成されるカルコゲニドが広く使用されている。

【００２７】

高抵抗の非晶質状態と高導電性の結晶状態の間の相転移は、適切な振幅および持続時間の電流パルスによって電気的に誘起することができる。特に、非晶質状態への転移（「リセット」）は、カルコゲニドをジュール効果によって融点を超えて加熱するのに十分な振幅の電流パルスを適用することによって実現される。使用される電流パルスは、カルコゲニドの冷却が結晶化を防止するのに十分な速さとなるような、急峻なエッジを有するもので、例えば矩形電流パルスである。結晶状態への転移（「セット」）は異なる技術を用いて誘起することができる。このように、ＰＣＭデバイスの相のコンダクタンスが異なるため、この現象を使用してビットを記憶することが可能である。ビットが中間のコンダクタンスを有する状態になるように内部温度およびその時間的推移を制御することも可能であ

る。このことは、ビットを記憶することに加えて、計算を行うために使用できるが、このことは、閾値未満の相変化の所定のグループ（または組）が次第に積み上がってビット反転が生じることによる。

【００２８】

このように計算を行うことの利点は２つある。すなわち、メモリ内で演算が行われるためメモリを行き来することが回避される、および、演算を並列で行うことができる。これらの違いはニューロンの集団の挙動と自然な対応関係を有し、このことにより、ＰＣＭデバイスは深層ニューラル・ネットワーク（ＤＮＮ）などの人工ニューラル・ネットワーク（ＡＮＮ）で使用するのに適したものとなる。

【００２９】

ＰＣＭは更に、ＤＮＮ（または他のタイプのニューラル・ネットワーク）での使用にも適している。このことはニューロン活動が二値的な全か無かの状態ではないことに起因しており、ニューロン活動はオンとオフの間のある範囲の中間挙動をとり得る。したがって、ＰＣＭデバイスが１と０の間の状態をとることができることによって、ニューロンの挙動を直接的にモデル化することが可能になる。

【００３０】

これを訓練に使用するために、ＰＣＭデバイス（ビット）のグリッド（またはアレイ）をＤＮＮの各層にマッピングすることができる。配線で構成した通信ネットワークによって、ニューロン同士の間で通信を行うことが可能になる。この通信の強度はメモリの状態によって設定され、このときメモリは完全にオンと完全にオフの間の変動範囲をとる。次にその状態がそこに投入される全てのビットによって設定される。通信ハードウェアは、相変化ビットからの可変強度信号を、デジタル通信ネットワークと適合性のある、様々な持続時間の信号へと変換する。

【００３１】

通常、ＤＮＮを実装するには前向き推論の計算は時間がかかりかつエネルギー負荷が高い可能性があるが、その理由は、ネットワーク用の重みデータを従来のデジタル・メモリ・チップとプロセッサ・チップの間で伝送する必要があるとともに、それらの重みを常にメモリに維持する必要があるからである。上記したように、ＰＣＭに基づくアナログ不揮発性メモリは、アナログ領域内の重みデータの場所において並列化された積和演算を行うことによって、前向き推論を加速させ、エネルギー消費を低減することができ、メモリ中の記憶されている重みデータをリフレッシュする必要性が低減される。そのようなＤＮＮの実際の応用には制限はなく、ＩｏＴデバイスに関するリアルタイムのセンサ・データ処理および推論などが含まれ得る。

【００３２】

前向き推論の精度は重みプログラミングの精度に強く依存する。しかしながら、ＰＣＭなどのアナログ・メモリを所望のアナログ・コンダクタンス値となるようにプログラムすることは、特にアナログ・メモリ・アレイにばらつきがある場合、簡単なことではない。特定の抵抗状態（例えばこれは、低抵抗またはＳＥＴ状態、高抵抗またはＲＥＳＥＴ状態、またはＲ１、Ｒ２、もしくはＲ３などのＳＥＴ／ＲＥＳＥＴ状態の間の何らかの中間状態であり得る）となるようにＰＣＭデバイスをプログラムした後で、その抵抗状態の特定の抵抗値が経時的にドリフトする場合がある。例えば、抵抗ドリフトは、ＰＣＭデバイスのＬＯＧ（抵抗）が、ＬＯＧ（時間）の関数として定常的に線形の増加（ＬＯＧ（コンダクタンス）の減少と等価である）を示し続ける物理的プロセスである。このことは、コンダクタンスが時間の一次関数としてまず急激に降下し、その後（時間の一次関数として）飽和を見せることを意味するが、少なくとも所定期間（例えば６週間、３ヶ月、１年、等）である時間間隔にわたってコンダクタンスは降下し続ける。そのようなドリフトは、特にメモリ・デバイスが多状態モードで操作されるとき、ある抵抗状態を別の抵抗状態と区別する際に問題となる場合がある。このことにより、ＰＣＭデバイスを使用するシステム、例えばＡＮＮは、ＰＣＭデバイスを繰り返しプログラムし直すことを余儀なくされるか、または（ＰＣＭデバイスからの様々な値に起因して）予想しない結果を生じる可能性が

ある。以降、本文書では、ドリフトがＰＣＭデバイスのコンダクタンス値を少なくとも所定値だけ変化させるまでの最短期間を、「ドリフト期間」と呼ぶ。

【００３３】

少なくとも過去にプログラムされたＰＣＭデバイスを使用して実装されるＡＮＮの前向き推論に関して、コンダクタンスの平均損失は読み出し電流をスケーリング・アップすることによって部分的に補償できるが、ドリフト係数「ニュー」のサイクル間のランダムな変動に起因するコンダクタンスの拡散は、この様式では容易には補正できない。シナプス荷重データが、平均スケーリング係数によって十分に補正されないＰＣＭに基づくシナプス荷重へと符号化されるのを回避する方法が、依然として必要とされている。

【００３４】

本発明の１つまたは複数の実施形態を使用する本明細書に記載する技術は、そのような技術的課題を克服する。本発明の１つまたは複数の実施形態はしたがって、特にＰＣＭデバイスをプログラムするおよびドリフト外れ値を抑制するための、コンピュータ技術の実際の応用および改善を容易にする。本発明の１つまたは複数の実施形態によって提供される他の利点および実際の応用は、本明細書の記載に基づいて当業者には明らかであろう。

【００３５】

本発明の１つまたは複数の実施形態は、コンダクタンスが最初に降下しその後飽和することに基づいて、本明細書に記載する技術的課題に対処する。しかしながら、ＰＣＭデバイスがプログラムされると（プログラミング事象）、コンダクタンスの推移は、時間がプログラミング事象以降の時間である限り、ＬＯＧ（コンダクタンス）対ＬＯＧ（時間）プロット上で常に直線となる。ｔ＝０である何らかの他の始点を使用すると、因果律に反するデータ点、または、時間軸の最初の大きさのオーダーに関して平坦であり後から初めてカーブして背後にある実際のニュー係数と整合する曲線、のいずれかが生じる。

【００３６】

ここで、ＡＮＮにおいてＰＣＭデバイスをシナプスとして使用することを含み、ＡＮＮ中のシナプスのプログラム中にドリフト係数外れ値を抑制するための技術を組み込むことによって技術的課題に対処する、本発明の１つまたは複数の実施形態について記載する。

【００３７】

図１は、深層ニューラル・ネットワークのブロック図を表す。描かれているＤＮＮ１００は、入力層１１０と隠れ層１２０と出力層１３０とを有し、各層はニューロン１０５を含む。ＤＮＮは生物のニューラル・ネットワークから大まかな発想を得ている。ニューロン１０５は、可塑性シナプスによって相互接続されている並列処理ユニットとして機能する。相互接続の重みを調整することによって、ＤＮＮ１００は分類問題などの特定の問題を効率的に解くことができる。ＤＮＮ１００の訓練は一般に、通常は逆伝播と呼ばれる、大域的教師あり学習アルゴリズムに基づいている。訓練中、入力データはニューロン層１１０、１２０、および１３０を通って順伝播され、シナプス・ネットワークは積和演算を行う。最終層（出力層１３０）の応答は入力データ・ラベルと比較され、誤差が逆伝播される。順伝播ステップおよび逆伝播ステップはいずれも、行列ベクトル乗算のシークエンスを含む。この結果、誤差が低減されるようにシナプス荷重が更新される。これらの操作を、非常に大規模なニューラル・ネットワークに対して非常に大きいデータセット（数ギガバイト）を使用して繰り返し行う必要があるため、この総当たり最適化手法は、ノイマン型計算機上の従来技術のネットワークを訓練するのに、数日または数週間を要する可能性がある。したがって、深層学習のそのようなステップを加速させるために、本発明の１つまたは複数の実施形態では、ＰＣＭデバイスの複数のクロスバー・アレイと、他のアナログ通信リンクおよび周辺回路構成とを含む、コプロセッサが使用される。

【００３８】

図２Ａ、図２Ｂ、および図２Ｃは、本発明の１つまたは複数の実施形態に係る、アナログ・メモリのクロスバー・アレイを使用する深層ニューラル・ネットワークのブロック図を表す。シナプス荷重は、クロスバー・アレイ２００内で組織化された不揮発性メモリ（ＮＶＭ）デバイス２１０のコンダクタンス値に関してＤＮＮ１００の各層（１１０、１２

０、および１３０）と関連付けられている。ＮＶＭデバイス２１０は、ＰＣＭデバイス、抵抗ランダム・アクセス・メモリ（ＲＲＡＭ）・デバイスなどであり得る。図２Ａ、図２Ｂ、および図２ＣにはＤＮＮ１００を実装する様々な段階が、すなわち、図２Ａには順伝播が、図２Ｂには逆伝播が、および図２Ｃには重みの更新が描かれている。

【００３９】

１つまたは複数の例では、ＤＮＮ１００の複数の層（１１０、１２０、および１３０）の各々にそれぞれ対応する、複数のそのようなクロスバー・アレイが存在する。本発明の１つまたは複数の実施形態によれば、コプロセッサ／チップ／システムは、そのようなクロスバー・アレイ２００を、ニューロン活性化関数とクロスバー・アレイ２００間の通信とを実施するための追加の周辺回路構成と共に含む。

【００４０】

図３は、本発明の１つまたは複数の実施形態に係る相変化メモリ・アレイ・デバイスのブロック図を表す。アーキテクチャには、チップ／システム３００上の融通性のあるルーティング・ネットワークによって接続されている、複数の同一のアレイ・ブロック３１０が描かれている。各アレイ・ブロック３１０は、ＮＶＭデバイス２１０のクロスバー・アレイ２００を表す。融通性のあるルーティング・ネットワークは、少なくとも３つのタスクを有する：（１）チップ３００のエッジからデバイス・アレイ３１０へとチップ入力（例えば例データ、例ラベル、および重み上書き）を伝達すること、（２）チップ出力（例えば推論された分類および更新された重み）をアレイ３１０からチップ３００のエッジへと伝搬すること、ならびに（３）多層ニューラル・ネットワークを実装するために様々なアレイ３１０を相互接続すること。各アレイ３１０は、シナプス接続部１１５の密なグリッドで接続されている、入力ニューロン１１０（ここでは各アレイの「西」側に示されている）および出力ニューロン１３０（「南」側）を有する。周辺回路構成は、個々の行（行回路構成３２０）および列（列回路構成３３０）に割り当てられた回路構成へと分割されており、回路構成はいくつかの隣り合う行および列の間で共有される。

【００４１】

本発明の１つまたは複数の実施形態によれば、クロスバー・アレイ２００におけるＮＶＭデバイス２１０（例えば、ＰＣＭデバイス）は、事前に訓練された重みに従ってプログラムされる。事前に訓練された重みは、ニューラル・ネットワーク・コンピューテーションにおいてＮＶＭデバイス２１０が、ＮＶＭデバイス２１０がプログラミング時間と将来の時点である実際の推論使用時間との間で経験することになるドリフトのいかなる不確実性／分布とも無関係に実施すべき重みである。

【００４２】

図４は、本発明の１つまたは複数の実施形態に係る、複数のニューロンに結合されたクロスバー・アレイをネットワークとして使用して実装されたニューロモルフィック・システムの構造を表す。描かれているニューロモルフィック・システム４００は、クロスバー・アレイ２００を使用して相互接続された複数のニューロン４１４、４１６、４１８、および４２０を含む。一例では、クロスバー・アレイ２００は、約０．１ｎｍから１０μｍの範囲内のピッチを有する。システム４００は、クロスバー・アレイ４１２の交点接合部において可変状態抵抗器として使用されるＮＶＭデバイス２１０を含むシナプス・デバイス４２２を更に含む。シナプス・デバイス４２２は、軸索経路４２４、樹状突起経路４２６、および膜経路４２７に接続され、このとき軸索経路４２４および膜経路４２７は樹状突起４２６と直交する。用語「軸索経路」、「樹状突起経路」、および「膜経路」を、それぞれ「軸索」、「樹状突起」、および「膜」と呼ぶ場合もある。

【００４３】

交点接合部にＮＶＭデバイス２１０を含むナノスケールのクロスバー・アレイであり得る、クロスバー・アレイ２００を利用して、上記電子ニューロン間の任意の可塑性接続性が実装される。各シナプス・デバイス４２２は、信号通信（ニューロン発火事象）中のクロストークを防止するために、ならびに漏れ電流および電力消費を最小化するために、クロスバー接合部ごとに、ダイオードとして配線されない電界効果トランジスタ（ＦＥＴ）

を含み得るアクセスまたは制御デバイス４２５を更に含む。他の実施形態では制御デバイス４２５として他のタイプの回路を使用することができ、ＦＥＴは本明細書の説明では１つの可能な例として使用されていることに留意すべきである。

【００４４】

電子ニューロン４１４、４１６、４１８、および４２０は、クロスバー・アレイ２００の周辺にある回路として構成される。設計および製作が簡単であることに加えて、クロスバー・アーキテクチャは、利用可能なスペースの効率的な使用を実現する。完全なクロスバー・アレイに固有の完全なニューロン接続性は、製作中の電気的初期化または望まれない場所でのマスク・ステップの省略によって、どのような任意の接続性にも変換され得る。クロスバー・アレイ２００はニューロン間の通信をカスタマイズするように構成され得る（例えば、あるニューロンは別のニューロンと決して通信しない）。製作レベルで特定のシナプスをブロックすることによって、任意の接続が得られ得る。このため、システム４００のアーキテクチャ原理は、生物のニューロモルフィック・ネットワークにおいて観察される全ての直接的配線組合せを模倣することができる。

【００４５】

クロスバー・アレイ２００は図４に示すように、ドライバ・デバイスX_2、X_3、およびX_4を更に含む。デバイスX_2、X_3、およびX_4は、インターフェース・ドライバ・デバイスを含み得る。詳細には、樹状突起４２６は、クロスバー・アレイ２００の一方側にドライバ・デバイスX_2を、およびクロスバー・アレイの他方側にレベル変換器デバイス（例えば、センス・アンプ）X_4を有する。軸索４２４は、クロスバー・アレイ２００の一方側にドライバ・デバイスX_3を有する。ドライバ・デバイスは、「西」側および「南」側（図３）などの本明細書に記載する機能を実施するＣＭＯＳ論理回路を含み得る。

【００４６】

これらのシグナリング技術は、図２Ａ、図２Ｂ、および図２Ｃに描かれているようなクロスバー・アレイ２００を使用するＤＮＮ１００を実装するための操作に使用される。図２Ａを参照すると、順伝播には、シナプス・ネットワークが積和演算を行うニューロン層（１１０、１２０、および１３０）を通してデータを処理することが含まれる。順方向パスと関連付けられる行列ベクトル乗算は、描かれているクロスバー・アレイ２００を使用してＯ（１）複雑性で実施され得る。例えば、Ａが行列でありｘおよびｂがベクトルであるときのＡｘ＝ｂを実行するために、Ａの要素は、クロスバー２００内で組織化されたＰＣＭデバイス２１０のコンダクタンス値に線形にマッピングされる。ｘ値は、行に沿って印加される読み出し電圧の振幅または持続時間へと符号化される。Ａの正および負の要素は、個別のデバイス上で減算回路と一緒にコーディングされる。別法として、１つまたは複数の例では、負のベクトル要素が負の電圧として加えられる。列に沿った結果的な電流は、結果ｂに比例している。入力が持続時間へと符号化される場合、結果ｂは合計電荷（例えば、時間で積分された電流）である。使用されるＮＶＭデバイス２１０の特質は、多レベルの記憶能力、ならびに、キルヒホッフの回路法則、すなわちオームの法則およびキルヒホッフの電流法則である。

【００４７】

クロスバー・アレイ２１０のおよびニューロモルフィック・システム４００の説明は１つの可能な例示の実装形態であること、ならびに、本発明の１つまたは複数の実施形態を他のタイプの実装形態において使用できることに留意すべきである。

【００４８】

典型的には、ＮＶＭデバイス２１０をプログラムすること（重みを更新すること）は、コンプライアンス電流を定常的に大きくしながらＳＥＴパルスを反復的に適用することによって行われる。より低いコンダクタンス値への急激な移行を引き起こし得るＲＥＳＥＴパルスとは異なり、部分的ＳＥＴパルスの連続的な適用は、結果的にＮＶＭデバイス２１０のコンダクタンス値のより漸進的な増大をもたらすと考えられる。このため、ＮＶＭデバイス２１０を使用するニューロモルフィック・システム４００に対して、既存のソリューションでは、ＰＣＭデバイスを所望のシナプス荷重へと調整するために、部分的ＳＥＴ

パルスが使用される。

【００４９】

例えば、ＮＶＭデバイス２１０のアナログ・コンダクタンス調整のための典型的なプログラミング戦略は、目標アナログ・コンダクタンス値に達するまで、コンプライアンス電流を定常的に大きくしながら（同時に場合によってはパルス持続時間も長くしながら）、ＳＥＴパルスを反復的に適用することである。ＮＶＭデバイス２１０は平均スケーリング係数によって補正される。しかしながら、ＰＣＭに基づくシナプス荷重へのシナプス荷重データの符号化は、ドリフト期間にわたるそのような平均スケーリング係数によっては補正されない。

【００５０】

図５は、本発明の１つまたは複数の実施形態に係るＮＶＭデバイスの抵抗／コンダクタンス値の例示のドリフトを表した図である。表されているのは例示的な値についてのものであり、他の例では任意の他の抵抗／コンダクタンス値およびドリフトが可能であることが理解される。図５は、個別のＮＶＭデバイス２１０の経時的な抵抗／コンダクタンスのドリフトを示している。見てとれるように、デバイスは互いに異なってドリフトし得る。デバイスごとの異なるドリフトは、対応するＮＶＭデバイス２１０のドリフト係数（「ν」（ニュー）係数）として表現できる。ＮＶＭデバイス２１０のν係数はＮＶＭデバイス２１０に関するコンダクタンス推移の軌跡を示し、ｌｏｇ（コンダクタンス）対ｌｏｇ（時間）のプロット上の直線の傾きを表す。

【００５１】

図６は、本発明の１つまたは複数の実施形態に係る例示のデバイスに関するν係数を表した図である。図６のプロットは、個別のＮＶＭデバイス２１０に関するコンダクタンスの経時的な変化の軌跡を示す。それぞれのＮＶＭデバイスのν係数も示されている。コンダクタンスが最後に読み出されたときのコンダクタンスおよび時間はいずれも、ＬＯＧスケールを使用してプロットされる。ν係数はＮＶＭデバイス２１０に割り当て可能な値ではない。むしろν係数は、ＮＶＭデバイスがプログラムされるときに、ＮＶＭデバイス２１０によってランダムに採択されているように見える。これらの現象は、各プログラミング事象後のＮＶＭデバイス中の多結晶粒子のランダム分布に起因すると考えられる。このため、ＮＶＭデバイス２１０のドリフトは予測不可能であり、本明細書で検討する技術的課題をもたらす。

【００５２】

本発明の１つまたは複数の実施形態は、ＮＶＭデバイス２１０（シナプス）における重み値をプログラムすること、およびＮＶＭデバイス２１０における変化した重み値を所定期間後にチェックすることによって、そのような技術的課題に対処する。例えば、変化した重み値は、１～６０秒待機した後で、ν係数に起因する軌跡が明らかになるまでチェックされる。所定期間（例えば１～６０秒）後の変化した重み値と目標値との差が少なくとも所定閾値である場合、ＮＶＭデバイス２１０をプログラムし直す。

【００５３】

このサイクルは、ＮＶＭデバイス２１０がν係数の予想される分布の中央値に十分に近いν係数を選択するまで継続される。このことは、１秒から１分までの合理的な間隔において、その時間におけるデバイスのコンダクタンスが、その時間間隔にわたるν係数中央値でドリフトするデバイスのコンダクタンスの変化に対応する比によって元のバウンディング・ボックスと関連付けられる、所与のバウンディング・ボックス内にあることをチェックすることによってチェックされ得る。各チェックには１回の試行あたり要するのは１分未満なので、将来への長期間にわたって「外れ値」のニュー係数からニューラル・ネットワークを保護するために、複数回のプログラミング試行を無理なく実施できる。言い換えれば、所定閾値を超えてドリフトすることになる重み値を抑制するようにＮＶＭデバイス２１０が結果的にプログラムされることを考慮すれば、ＮＶＭデバイス２１０を複数回プログラムするために「余分な」時間を費すことを許容できる。その理由は、ドリフトのためにＮＶＭデバイス２１０をプログラムし直す必要があり得るか、またはＮＶＭデバイ

ス２１０から誤った結果が生じ得るからである。
【００５４】
図７は、本発明の１つまたは複数の実施形態に係る、ＮＶＭデバイスのプログラム中にドリフト係数外れ値を抑制するための方法のフローチャートを表す。図８は、図７に示されている方法７００に従ってプログラムされている例示のＮＶＭデバイスを表す。
【００５５】
方法７００は、７０２において、事前補償時点８１０（例えば、２０ｎｓ）におけるＮＶＭデバイス２１０のコンダクタンス目標を決定することを含み、これは、ＮＶＭデバイス２１０のプログラミング後のその時点８１０と、プログラミング時点８０５後の所望の目標窓８５０（例えば７時間、２週間、６ヶ月等）との間の、予想される平均ドリフトを事前補償する。事前補償時点８１０は、ＮＶＭデバイス２１０のコンダクタンス値の変化がチェックされる所定期間（例えば２０ｎｓ）を表す。
【００５６】
図８を参照すると、ＮＶＭデバイス２１０は時点８０５においてプログラムされる。所望の（目標）時間窓８５０は、プログラミング時点８０５以降の約６時間～６ヶ月で表されている。他の例では所望の時間窓８５０は異なり得ることが理解される。所望の時間窓８５０は、ＮＶＭデバイス２１０にプログラムされるコンダクタンス値が、本明細書に記載するように、ドリフト係数中央値により、その期間中に予想されるドリフト周辺の所定の間隔の外に、所定閾値を超えて変化しない期間を表す。事前補償時点８１０は、プログラミング時点８０５から２０ｎｓの所定期間のところに示されている。他の例では事前補償時点８１０は異なる期間においてであり得ることが理解される。
【００５７】
したがって、事前補償時点８１０におけるＮＶＭデバイス２１０のコンダクタンス目標を決定することは、所定期間後にＮＶＭデバイス２１０が有するべきコンダクタンス値を計算することを含む。ＮＶＭデバイス２１０のコンダクタンス値がコンダクタンス目標から所定閾値内にある場合、ｖ係数は、ＮＶＭデバイス２１０のコンダクタンス値が目標窓８５０内で所定閾値を超えてドリフトしないような、許容可能な範囲内にあると判断される。ＮＶＭデバイス２１０のコンダクタンス値が事前補償時点８１０においてコンダクタンス目標から所定閾値内にない場合、ｖ係数は許容可能な範囲外にあると判断される。言い換えれば、ＮＶＭデバイス２１０のコンダクタンスは、目標窓８５０内で所定閾値を超えてドリフトすると予測される。
【００５８】
１つまたは複数の例では、方法７００は、異なる時点において複数のコンダクタンス目標値を決定することを含む。例えば、事前補償時点８１０について第１の目標コンダクタンス値が算出され、チェックポイント８１５について第２の目標コンダクタンス値が算出され、以下同様である。実際の実装の観点からは、チェックされ得るコンダクタンス目標値の数を、例えば２つ、４つ、または、限定された時間、例えば１分、５分、もしくはそのようなプログラミング時間内での、ＮＶＭデバイスの複数回のプログラミングおよびチェックを容易にし得るような任意の他の数に、限定しなければならない。
【００５９】
ここで図７を参照すると、ＮＶＭデバイス２１０は、７０４において、ＮＶＭデバイス２１０によって表されるシナプスに記憶されることになる重みに対応するコンダクタンス値を用いてプログラムされる。プログラムすることは、既に指摘したように、時点８０５において行われる。方法７００は、７０６において、事前補償時点（８１０）においてＮＶＭデバイスのコンダクタンス値を測定することを更に含む。７０８において、測定されたコンダクタンス値は目標コンダクタンス値と比較される。事前補償時点８１０において測定されたコンダクタンス値が、事前補償時点８１０に対応するコンダクタンス目標値の第１の所定閾値内にない場合、ＮＶＭデバイス２１０はプログラムし直され、このチェックが継続される（７０４、７０６、７０８）。
【００６０】

測定されたコンダクタンス値が目標コンダクタンス値の第１の所定閾値内にある場合、方法７００は、７１０において、第２のチェックポイント８１５においてコンダクタンス値を再び測定する。図８に表されている例では、チェックポイント８１５はプログラミング時点８０５から１秒のところに示されている。しかしながら、他の例では他のチェックポイントを使用できることが理解される。

【００６１】

第２の測定されたコンダクタンス値は、７１２において、チェックポイント８１５に対応する第２の目標コンダクタンス値と比較される。チェックポイント８１５における第２の測定されたコンダクタンス値が、チェックポイント８１５に対応する第２のコンダクタンス目標値の所定閾値内にない場合、ＮＶＭデバイス２１０はプログラムし直され、方法の動作が繰り返される（７０４、７０６、７０８、７１０）。第２の測定されたコンダクタンス値が第２の目標コンダクタンス値の所定閾値内にある場合、ＮＶＭデバイス２１０は、ドリフト外れ値が抑制された状態でプログラムされていると考えられる。

【００６２】

図８には、ｖ係数が変化して、事前補償点８１０またはチェックポイント８１５あるいはその両方において様々な量のコンダクタンス値のドリフトを引き起こす、様々なケースが表されている。図８の軌跡８６０および８６５は、目標窓８５０内でのコンダクタンス値が所望したようになるようにプログラムされているＮＶＭデバイスを表す。他の軌跡は、測定されたコンダクタンス値が目標コンダクタンス値を満たさない場合に、ＮＶＭデバイス２１０が目標窓８５０内で所望されるコンダクタンス値とは異なるコンダクタンス値を有することになることを例示するために描かれている。

【００６３】

ｖ係数にはプログラミング事象の大部分についてＮＶＭデバイスのプログラミングの時間においてｖ係数のランダム分布の中央領域に対応するランダム性が関連付けられているので、コンダクタンスは計算される目標範囲内にあることになる。測定されたコンダクタンス値が目標値の閾値内である場合、ＮＶＭデバイスは、目標時間窓８５０によって表されるその適用の目的の時間スパン（ここでは６時間から６ヶ月）において、所望のコンダクタンス値を有する可能性が非常に高い。代わりに、プログラミング後の事前補償時点８１０（２０ｎｓ）において首尾よく第１の目標範囲内にあることが知られていたコンダクタンスが、その後のチェックポイント８１５において第２の目標コンダクタンス範囲外にある場合、このことは、そのｖ係数が外れ値であることを示している。したがってこの場合、適用目的の時間スパンにおいて予想されるコンダクタンスは、所望の値よりも高いまたは低いと予測することができる。

【００６４】

ＮＶＭデバイス２１０をプログラムし直しプログラム後にチェックポイントにおいてコンダクタンス測定値をチェックするこの方法は、実際上許容可能なプログラミング時間限度内で実行され得る。これに応じて、許容可能な「平均」ｖ係数を用いたＮＶＭデバイスのプログラミングの複数回の試行後に、ＮＶＭデバイスのドリフト外れ値が抑制されていると見なすことができ、コンダクタンス値が目標時間窓中に所望の範囲内にあると予測することができる。

【００６５】

図９は、本発明の１つまたは複数の実施形態に係る、ドリフト外れ値の抑制を伴ってＮＶＭデバイスをプログラムするためのシステムを表す。描かれているシステム９００は、制御装置９１０と、１つまたは複数のＮＶＭデバイス２１０を含むニューラル・ネットワーク・システムなどの、ＮＶＭデバイス・システム９２０と、を含む。制御装置９１０は、メモリ・デバイスに記憶されている１つまたは複数のコンピュータ実行可能命令を実行できる、１つまたは複数の処理ユニットを含むコンピューティング・デバイスであり得る。例えば、制御装置９１０は、デスクトップ・コンピュータ、タブレット・コンピュータ、電話機、サーバ・コンピュータ、または任意の他のコンピューティング・デバイスであり得る。制御装置９１０およびＮＶＭデバイス・システム９２０は、制御装置９１０がＮ

ＶＭデバイス・システム９２０中のＮＶＭデバイス２１０の抵抗／コンダクタンス（すなわち重み）を構成できるように互いに結合される。制御装置９１０は更に、ＮＶＭデバイス２１０の抵抗／コンダクタンスを読み出す／測定することができる。１つまたは複数の例では、制御装置９１０は、本明細書に記載される１つまたは複数の方法を実施する。図９では制御装置９１０およびＮＶＭデバイス・システム９２０は別個のブロックとして示されているが、本発明の１つまたは複数の実施形態では、制御装置９１０およびＮＶＭデバイス２１０を単一のブロックの一部として統合することができる。

【００６６】

本発明は、任意の可能な技術詳細レベルで組み込まれた、システム、方法、またはコンピュータ・プログラム製品あるいはそれらの組合せであり得る。コンピュータ・プログラム製品は、プロセッサに本発明の態様を実行させるためのコンピュータ可読プログラム命令を有する、コンピュータ可読記憶媒体を含んでもよい。

【００６７】

コンピュータ可読記憶媒体は、命令実行デバイスによって使用される命令を保持および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、電子ストレージ・デバイス、磁気ストレージ・デバイス、光ストレージ・デバイス、電磁ストレージ・デバイス、半導体ストレージ・デバイス、または以上の任意の好適な組合せであり得るが、これらに限定されない。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なリストには、以下、すなわち、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）、読取り専用メモリ（ＲＯＭ）、消去可能プログラマブル読取り専用メモリ（ＥＰＲＯＭもしくはフラッシュ・メモリ）、スタティック・ランダム・アクセス・メモリ（ＳＲＡＭ）、ポータブル・コンパクト・ディスク読取り専用メモリ（ＣＤ－ＲＯＭ）、デジタル・バーサタイル・ディスク（ＤＶＤ）、メモリ・スティック、フロッピー（Ｒ）・ディスク、命令が記録されているパンチ・カードもしくは溝の中の隆起構造などの機械的に符号化されたデバイス、および以上の任意の好適な組合せが含まれる。本明細書において使用されるコンピュータ可読記憶媒体は、電波もしくは他の自由に伝播する電磁波、導波路もしくは他の伝送媒体を通じて伝播する電磁波（例えば、光ファイバ・ケーブルを通過する光パルス）、または配線を介して伝送される電気信号などの、一過性の信号そのものであると解釈されるべきではない。

【００６８】

本明細書に記載するコンピュータ可読プログラム命令は、コンピュータ可読記憶媒体からそれぞれのコンピューティング／処理デバイスに、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、もしくはワイヤレス・ネットワーク、またはその組合せを経由して外部のコンピュータまたは外部ストレージ・デバイスに、ダウンロードされ得る。ネットワークは、銅伝送ケーブル、光伝送ファイバ、ワイヤレス伝送、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、またはエッジ・サーバ、あるいはそれらの組合せを備え得る。各コンピューティング／処理デバイス内のネットワーク・アダプタ・カードまたはネットワーク・インターフェースが、ネットワークからコンピュータ可読プログラム命令を受信し、それらのコンピュータ可読プログラム命令を、それぞれのコンピューティング／処理デバイス内のコンピュータ可読記憶媒体に記憶されるように転送する。

【００６９】

本発明の動作を実行するためのコンピュータ可読プログラム命令は、アセンブラ命令、命令セット・アーキテクチャ（ＩＳＡ）命令、マシン命令、マシン依存型命令、マイクロコード、ファームウェア命令、状態設定データ、集積回路用の構成データ、または、Ｓｍａｌｌｔａｌｋ（Ｒ）、Ｃ＋＋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もしくは類似のプログラミング言語などの手続き型プログラミング言語を含む１つもしくは複数のプログラミング言語の任意の組合せで書かれた、ソース・コードもしくはオブジェクト・コードのいずれか、であり得る。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、スタンド・アロン・ソフトウェア・パッケー

ジとして部分的に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かつ部分的に遠隔のコンピュータ上で、または専ら遠隔のコンピュータもしくはサーバ上で、実行することができる。後者のシナリオでは、遠隔のコンピュータを、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）を含む任意の種類のネットワークを介してユーザのコンピュータに接続してもよく、または、外部のコンピュータへの接続を（例えば、インターネット・サービス・プロバイダを使用してインターネットを介して）行ってもよい。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、フィールド・プログラマブル・ゲート・アレイ（ＦＰＧＡ）、またはプログラマブル・ロジック・アレイ（ＰＬＡ）を含む電子回路は、本発明の態様を実施するために、コンピュータ可読プログラム命令の状態情報を利用することによって、コンピュータ可読プログラム命令を実行して電子回路を個人化することができる。

【００７０】

本明細書には、本発明の実施形態に係る方法、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図またはブロック図あるいはその両方を参照して、本発明の態様が記載されている。フローチャート図またはブロック図あるいはその両方の各ブロック、およびフローチャート図またはブロック図あるいはその両方におけるブロックの組合せを、コンピュータ可読プログラム命令によって実施できることが、理解されるであろう。

【００７１】

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ム可能データ処理装置のプロセッサを介して実行される命令が、フローチャートまたはブロック図あるいはその両方の１つまたは複数のブロックに指定されている機能／動作を実施するための手段を作り出すべく、汎用コンピュータ、専用コンピュータ、または他のプログラム可能データ処理装置のプロセッサに提供されてマシンを作り出すものであってよい。これらのコンピュータ可読プログラム命令はまた、コンピュータ、プログラム可能データ処理装置、または他のデバイス、あるいはそれらの組合せに特定の方式で機能するように指示できるコンピュータ可読記憶媒体に保存されてもよく、この結果、命令が保存されたコンピュータ可読記憶媒体は、フローチャートまたはブロック図あるいはその両方の１つまたは複数のブロックに指定されている機能／動作の態様を実施する命令を含んだ製造品を構成することになる。

【００７２】

コンピュータ可読プログラム命令はまた、コンピュータ、他のプログラム可能装置、または他のデバイスで実行される命令が、フローチャートまたはブロック図あるいはその両方の１つまたは複数のブロックに指定される機能／動作を実施するように、コンピュータによって実行されるプロセスを作り出すべく、コンピュータ、他のプログラム可能データ処理装置、または他のデバイスにロードされ、コンピュータ、他のプログラム可能装置、または他のデバイス上で一連の動作ステップを実行させるものであってもよい。

【００７３】

図中のフローチャートおよびブロック図には、本発明の様々な実施形態に係るシステム、方法、およびコンピュータ・プログラム製品の、可能な実装形態のアーキテクチャ、機能性、および動作が説明されている。この関連において、フローチャートまたはブロック図内の各ブロックは、指定された論理機能を実施するための１つまたは複数の実行可能命令を備える、モジュール、セグメント、または命令の一部分を表すことができる。いくつかの代替的実装形態において、ブロック内に記された機能は、図に記されたものとは異なる順序で行われ得る。例えば、連続して示される２つのブロックは、実際は実質的に並行して実行され得、またはこれらのブロックは時には、関わる機能性に応じて逆の順序で実行され得る。また、ブロック図またはフローチャート図あるいはその両方の各ブロック、およびブロック図またはフローチャート図あるいはその両方におけるブロックの組合せは、指定された機能もしくは動作を行う、または専用ハードウェアとコンピュータ命令の組合せを実行する、専用ハードウェア・ベースのシステムによって実施され得ることも、留

意されるであろう。

【００７４】

本発明の様々な実施形態の説明を例示の目的で提示してきたが、それらは網羅的であること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ていない。当業者には記載される実施形態の範囲から逸脱することなく多くの修正および変更が明らかであろう。本明細書で用いられる専門用語は、実施形態の原理、実際の応用、もしくは市場で見られる技術に対する技術的な改善を最もよく説明するように、または、他の当業者が本明細書に記載される実施形態を理解できるように、選択された。

【００７５】

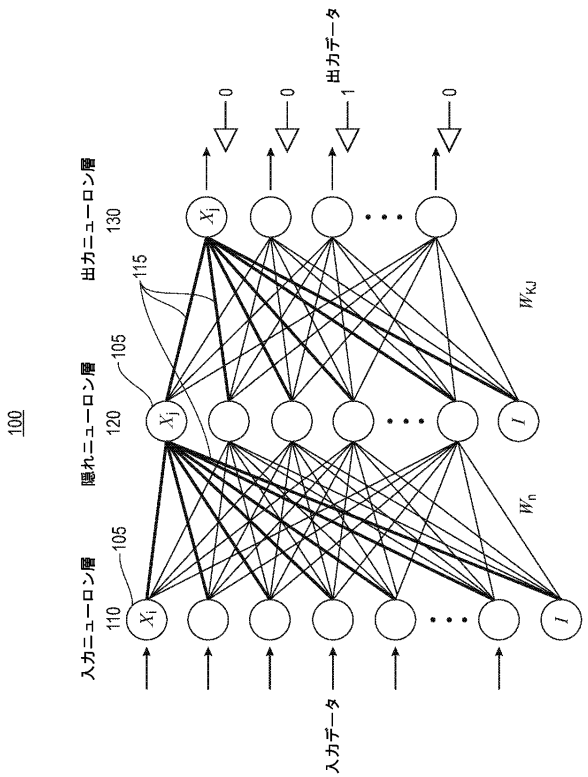
本明細書に記載する本発明の好ましい実施形態では、相変化メモリ（ＰＣＭ）がクロスバー・アレイにおける複数の交点の各々にある、複数のＰＣＭデバイスを備えるクロスバー・アレイであって、１つまたは複数のＰＣＭデバイスは人工ニューラル・ネットワークにおけるアナログ・シナプスを表し、クロスバー・アレイのＰＣＭデバイスのコンダクタンスをプログラムすることは、プログラミング時点である第１の時点において、ＰＣＭデバイスのコンダクタンスを第１のコンダクタンス値となるように構成するための信号を受信することと、事前補償時点において、ＰＣＭデバイスのコンダクタンスが、目標コンダクタンス値との差が所定閾値以内である第２のコンダクタンス値に変化していると判定することと、第２の事前補償時点において前記事前補償を再び測定することを含めて、第２の時点において、ＰＣＭデバイスのコンダクタンスを第１のコンダクタンス値となるように構成するための別の信号を受信することと、を含む、クロスバー・アレイ、が提供される。好ましくは、事前補償時点はプログラミング時点後の所定時間期間にあり、事前補償時点における目標コンダクタンス値は、プログラミング時点後の第２の所定期間にあたる目標時間窓に基づいて決定され、ＰＣＭデバイスのコンダクタンスは目標時間窓内で特定の範囲内にある。

【００７６】

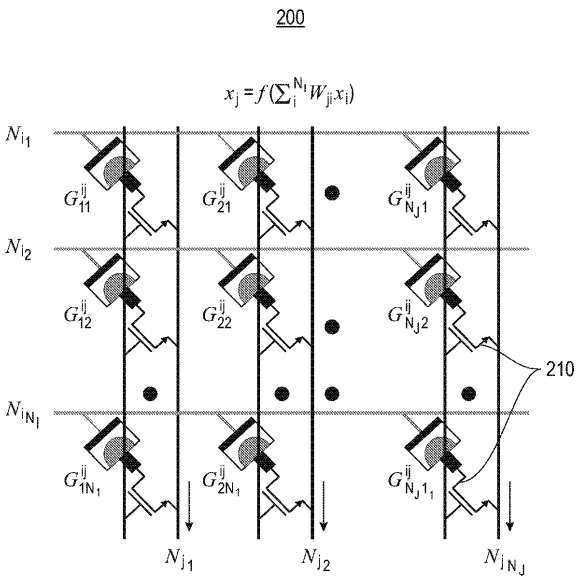
本明細書に記載する本発明の別の好ましい実施形態では、コンピュータ実装方法であって、プログラミング時点である第１の時点において、相変化メモリ（ＰＣＭ）デバイスのコンダクタンスを第１のコンダクタンス値となるように構成するための信号を受信することと、事前補償時点において、ＰＣＭデバイスのコンダクタンスが目標コンダクタンス値との差が所定閾値以内である第２のコンダクタンス値に変化していると判定することと、第２の事前補償時点において前記事前補償を再び測定することを含めて、第２の時点において、ＰＣＭデバイスのコンダクタンスを第１のコンダクタンス値となるように構成するための別の信号を受信することと、によってクロスバー・アレイ中の複数のＰＣＭデバイスから、あるＰＣＭデバイスのコンダクタンス値を構成することを含む、コンピュータ実装方法、が提供される。

【図面】

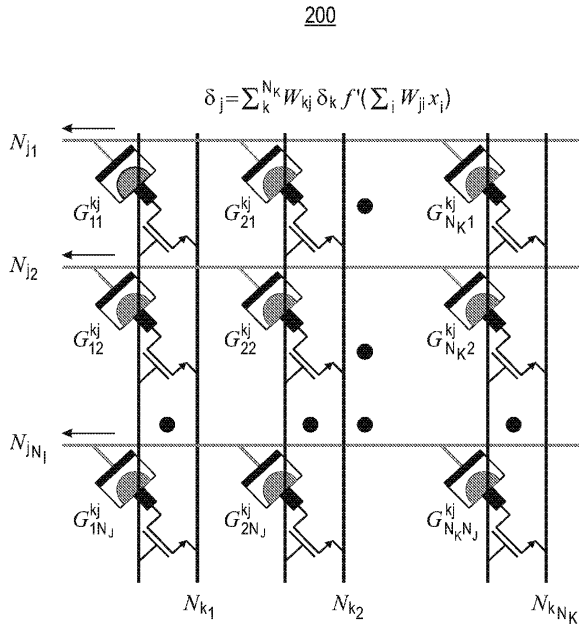
【図１】



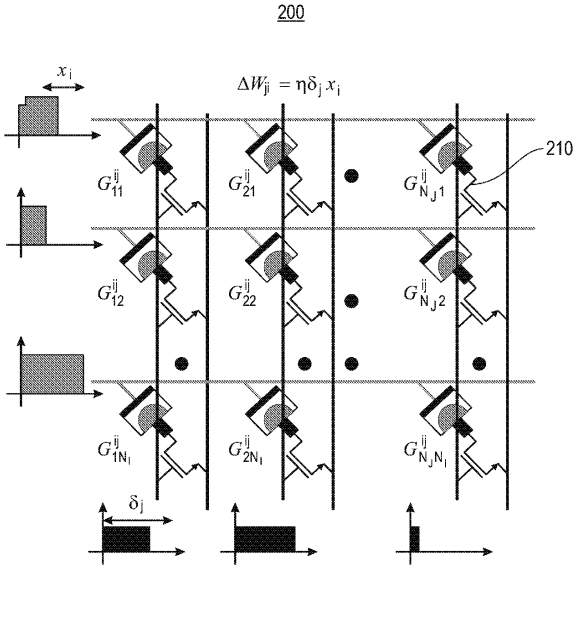
【図２Ａ】



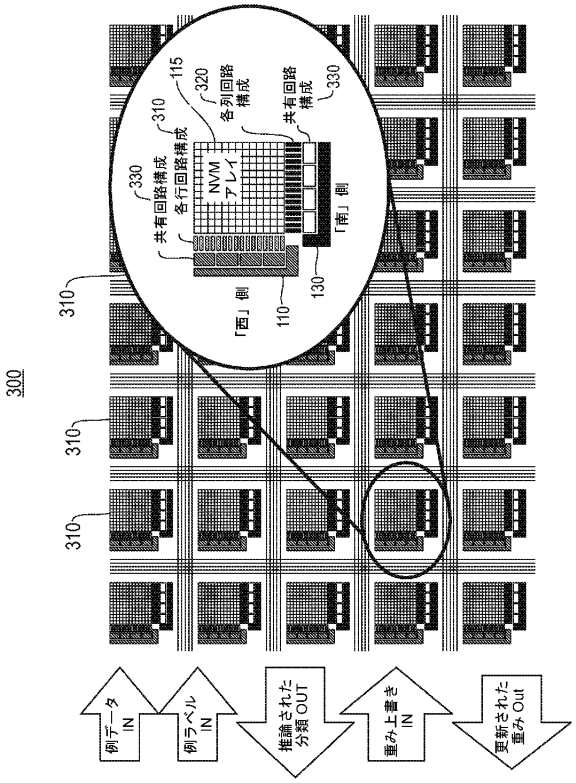
【図２Ｂ】



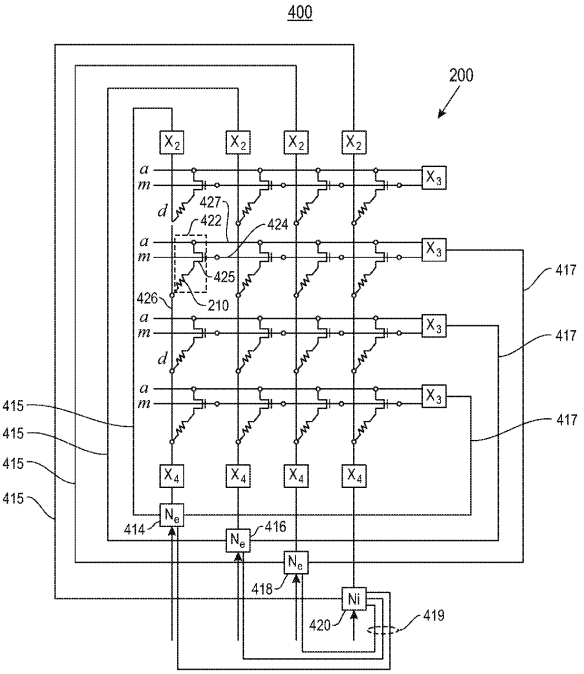
【図２Ｃ】



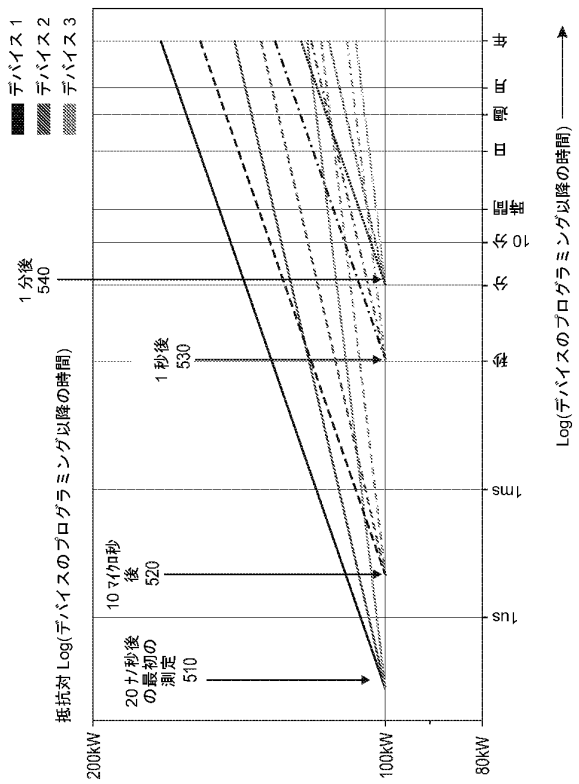
【図３】


300
310
115
320
各列回路構成
310
共有回路構成
330
NVM アレイ
共有回路構成
330
各行回路構成
「南」側
「西」側
110
130
例データ IN
例ラベル IN
推論された分類 OUT
重み上書き IN
更新された重み Out

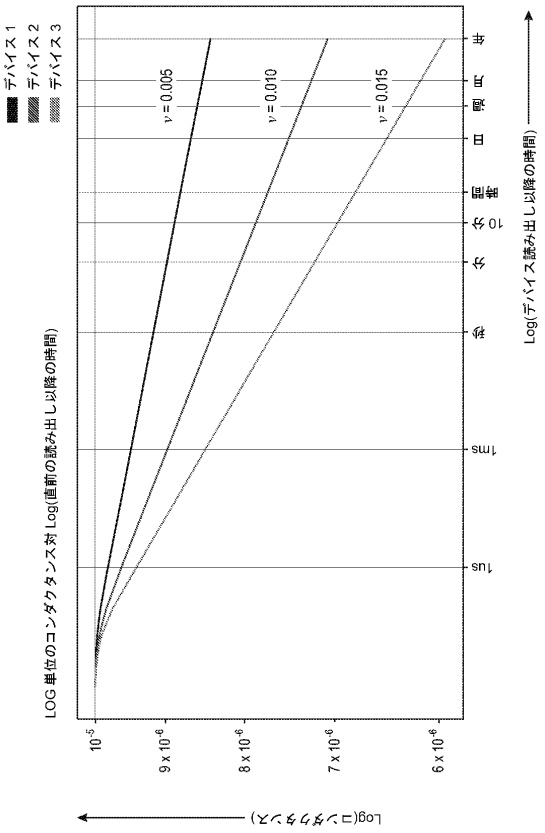
【図４】


400
200
X2
X3
X4
a
m
d
427
422
424
425
426
210
417
415
414
Ne
416
418
Ni
420
419

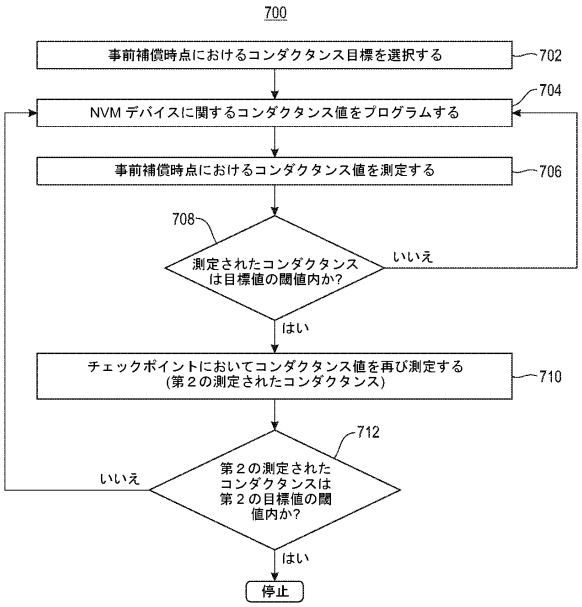
【図５】


デバイス 1
デバイス 2
デバイス 3
抵抗対 Log(デバイスのプログラミング以降の時間)
1 分後 540
1 秒後 530
10 マイクロ秒後 520
20 ナノ秒後の最初の測定 510
200kW
100kW
80kW
1us
1ms
秒
分
10 分
時間
日
週
年
Log(デバイスのプログラミング以降の時間)

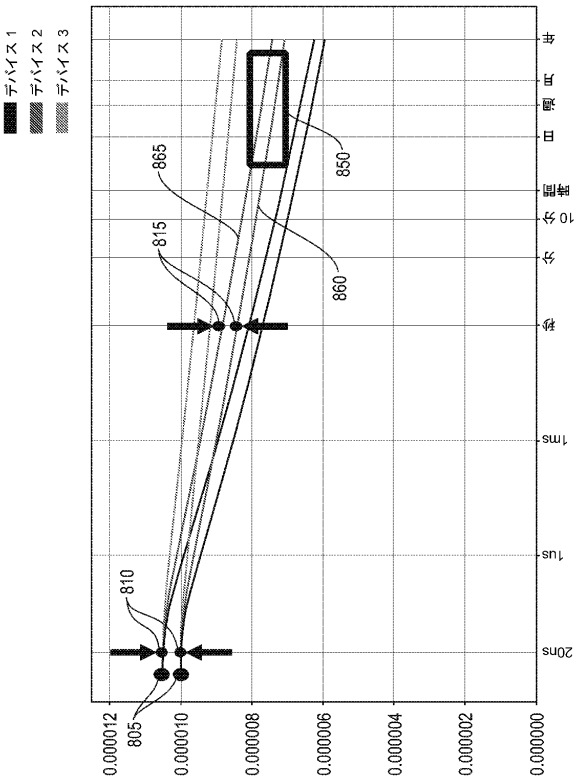
【図６】


デバイス 1
デバイス 2
デバイス 3
ν = 0.005
ν = 0.010
ν = 0.015
LOG 単位のコンダクタンス対 Log(直前の読み出し以降の時間)
Log(コンダクタンス)
10^-5
9 x 10^-6
8 x 10^-6
7 x 10^-6
6 x 10^-6
1us
1ms
秒
分
10 分
時間
日
週
年
Log(デバイス読み出し以降の時間)

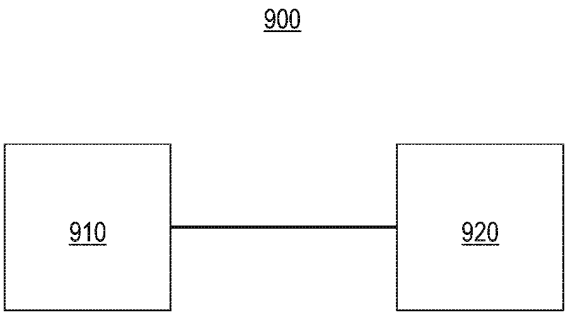
【図７】


700
事前補償時点におけるコンダクタンス目標を選択する
702
NVM デバイスに関するコンダクタンス値をプログラムする
704
事前補償時点におけるコンダクタンス値を測定する
706
708
測定されたコンダクタンス
は目標値の閾値内か?
いいえ
はい
チェックポイントにおいてコンダクタンス値を再び測定する
(第２の測定されたコンダクタンス)
710
712
第２の測定された
コンダクタンスは
第２の目標値の閾
値内か?
いいえ
はい
停止

【図８】


デバイス１
デバイス２
デバイス３
年
月
週
日
時間
10分
分
秒
1ms
1us
20ns
865
850
815
860
810
805
0.000012
0.000010
0.000008
0.000006
0.000004
0.000002
0.000000

【図９】


900
910
920

フロントページの続き

(72)発明者　バール、ジェフリー
アメリカ合衆国９５１２０－６０９９ カリフォルニア州サンノゼ アルマデン・リサーチ・センター ハリー・ロード６５０

審査官　小林 紀和

(56)参考文献　国際公開第２０１６／０４６９８０（WO，Ａ１）
国際公開第２０１２／１４０９０３（WO，Ａ１）
特表２０１２－５２７６６９（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ１１Ｃ　１３／００
Ｇ０６Ｎ　３／０６３
Ｈ１０Ｂ　６３／１０