

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7281535号
(P7281535)

(45)発行日 令和5年5月25日(2023.5.25)

(24)登録日 令和5年5月17日(2023.5.17)

(51)国際特許分類 F I
G 0 6 N 3/063(2023.01) G 0 6 N 3/063

請求項の数 49 (全39頁)

(21)出願番号	特願2021-510952(P2021-510952)	(73)特許権者	500147506 シリコン ストージング テクノロジー インコーポレイテッド S I L I C O N S T O R A G E T E C H N O L O G Y , I N C .
(86)(22)出願日	令和1年7月23日(2019.7.23)		アメリカ合衆国 9 5 1 3 4 カリフォル ニア州 サンノゼ ホルガー ウェイ 4 5 0
(65)公表番号	特表2021-536623(P2021-536623 A)		
(43)公表日	令和3年12月27日(2021.12.27)	(74)代理人	110000626 弁理士法人英知国際特許商標事務所
(86)国際出願番号	PCT/US2019/043101	(72)発明者	トラン、ヒュー、バン アメリカ合衆国 9 5 1 3 5 カリフォル ニア州、サンノゼ、ゲイレイ プレイス 2 6 4 2
(87)国際公開番号	WO2020/046495	(72)発明者	ティワリ、ビピン アメリカ合衆国 9 4 5 6 8 カリフォル 最終頁に続く
(87)国際公開日	令和2年3月5日(2020.3.5)		
審査請求日	令和4年7月22日(2022.7.22)		
(31)優先権主張番号	62/723,398		
(32)優先日	平成30年8月27日(2018.8.27)		
(33)優先権主張国・地域又は機関	米国(US)		
(31)優先権主張番号	16/183,250		
(32)優先日	平成30年11月7日(2018.11.7)		
	最終頁に続く		

(54)【発明の名称】 深層学習ニューラルネットワークで使用されるアナログニューラルメモリシステムのメモリセルに対する温度補償及び漏れ補償

(57)【特許請求の範囲】

【請求項 1】

アナログニューロモーフィックメモリシステムであって、
メモリセルのアレイを備える、ベクトルマトリックス乗算アレイと、
前記メモリセルの前記アレイのメモリセルの動作温度が変化するにつれて、前記メモリセルの前記アレイの前記メモリセルの電流電圧特性曲線の傾きを連続的に修正するための温度補償ブロックであって、前記温度補償ブロックは、
前記動作温度を示す出力を生成するための温度センサと、
前記温度センサの前記出力にตอบสนองして1つ以上の制御信号を生成するためのコントローラと、

前記1つ以上の制御信号にตอบสนองして温度補償電圧を生成するための1つ以上の調整可能なデバイスと、を備える温度補償ブロックと、を備える、アナログニューロモーフィックメモリシステム。

【請求項 2】

前記ベクトルマトリックス乗算アレイにおける重みは、前記メモリセルによって実現される、請求項1に記載のシステム。

【請求項 3】

前記ベクトルマトリックス乗算アレイにおける重みは、差動セルによって実現される、請求項1に記載のシステム。

【請求項 4】

前記ベクトルマトリックス乗算アレイにおける重みは、1対のブレンドメモリセルによって実現される、請求項1に記載のシステム。

【請求項5】

前記メモリセルの前記アレイの前記メモリセルのそれぞれはフラッシュメモリセルであり、前記1つ以上の調整可能なデバイスは、1つ以上の調整可能な電流源を備え、前記1つ以上の調整可能な電流源の各々は、前記1つ以上の制御信号のうちの1つによって調整され、かつレジスタに結合されており、前記温度補償電圧は、前記1つ以上の調整可能な電流源と前記レジスタとの間のノードにおいて生成され、前記フラッシュメモリセルに印加されて、前記フラッシュメモリセルの前記電流電圧特性曲線の傾きを修正する、請求項1に記載のシステム。

10

【請求項6】

前記温度補償電圧は、前記フラッシュメモリセルの制御ゲート端子に印加される、請求項5に記載のシステム。

【請求項7】

前記温度補償電圧は、前記フラッシュメモリセルの消去ゲート端子に印加される、請求項5に記載のシステム。

【請求項8】

前記温度補償電圧は、前記フラッシュメモリセルのソース線端子に印加される、請求項5に記載のシステム。

【請求項9】

前記温度補償電圧は、前記フラッシュメモリセルのビット線端子に印加される、請求項5に記載のシステム。

20

【請求項10】

前記フラッシュメモリセルは、スプリットゲートフラッシュメモリセルである、請求項5に記載のシステム。

【請求項11】

前記フラッシュメモリセルは、積層ゲートフラッシュメモリセルである、請求項5に記載のシステム。

【請求項12】

前記メモリセルの前記アレイのメモリセルの前記動作温度が変化するにつれて漏れを低減させるために、前記メモリセルの前記アレイの前記メモリセルの端子に印加されたバイアス電圧を修正するための漏れ低減ブロックを更に備える、請求項1に記載のシステム。

30

【請求項13】

アナログニューロモーフィックメモリシステムであって、
メモリセルのアレイを備える、ベクトルマトリックス乗算システムと、
前記メモリセルの前記アレイのメモリセルの動作温度のレベルが変化するにつれて、前記メモリセルの前記アレイの前記メモリセルの電流電圧特性曲線を不連続的に修正するための温度補償ブロックと、を備える、アナログニューロモーフィックメモリシステム。

【請求項14】

前記ベクトルマトリックス乗算システムにおける重みは、前記メモリセルによって実現される、請求項13に記載のシステム。

40

【請求項15】

前記ベクトルマトリックス乗算システムにおける重みは、差動セルによって実現される、請求項13に記載のシステム。

【請求項16】

前記ベクトルマトリックス乗算システムにおける重みは、ブレンドメモリセルによって実現される、請求項13に記載のシステム。

【請求項17】

前記メモリセルの前記アレイの前記メモリセルの各々はフラッシュメモリセルであり、前記温度補償ブロックは、

50

前記動作温度を示す出力を生成するための温度センサと、
前記温度センサの前記出力に応答して1つ以上の制御ビットを生成するためのコントローラと、

複数の電流源であって、各電流源は、スイッチを介してレジスタに選択的に結合され、各スイッチは、前記制御ビットのうちの1つによって制御された、複数の電流源と、を備え、

前記レジスタの一端において生成された電圧は、前記フラッシュメモリセルに印加されて、前記フラッシュメモリセルの前記電流電圧特性曲線の傾きを修正する、請求項13に記載のシステム。

【請求項18】

前記修正することは、フラッシュメモリセルの電流電圧特性曲線の傾きを修正することを含む、請求項13に記載のシステム。

【請求項19】

前記メモリセルは、スプリットゲートフラッシュメモリセルである、請求項13に記載のシステム。

【請求項20】

前記メモリセルは、積層ゲートフラッシュメモリセルである、請求項13に記載のシステム。

【請求項21】

前記温度補償ブロックは、
前記動作温度を示す出力を生成するための温度センサと、
前記温度センサの前記出力に応答して1つ以上の制御ビットを生成するためのコントローラと、

複数のレジスタを備える増幅回路であって、各レジスタは、スイッチを介して前記増幅回路の出力に選択的に結合されており、各スイッチは、前記制御ビットのうちの1つによって制御される、複数の増幅回路と、を備え、

前記増幅回路の前記出力において生成された電圧は、前記メモリセルに印加されて、前記メモリセルの前記電流電圧特性曲線の傾きを修正する、請求項13に記載のシステム。

【請求項22】

前記修正することは、前記メモリセルの前記電流電圧特性曲線の傾きを修正することを含む、請求項21に記載のシステム。

【請求項23】

前記メモリセルは、スプリットゲートフラッシュメモリセルである、請求項21に記載のシステム。

【請求項24】

前記メモリセルは、積層ゲートフラッシュメモリセルである、請求項21に記載のシステム。

【請求項25】

前記メモリセルの前記アレイのメモリセルの前記動作温度が変化するにつれて漏れを低減させるために、前記メモリセルの前記アレイの前記メモリセルの端子に印加されたバイアス電圧を修正するための漏れ低減ブロック

を更に備える、請求項13に記載のシステム。

【請求項26】

複数のベクトルマトリックス乗算アレイを備えるアナログニューロモーフィックメモリシステムにおいて温度補償を実行する方法であって、各ベクトルマトリックス乗算アレイはメモリセルのアレイを備え、前記方法は、

前記メモリセルの前記アレイのメモリセルの動作温度が変化するにつれて、前記メモリセルの前記アレイの前記メモリセルの電流電圧特性曲線の傾きを、温度補償ブロックによって連続的に修正するステップであって、前記温度補償ブロックは、動作温度を示す出力を生成するための温度センサと、前記温度センサの前記出力に応答して1つ以上の制御信

10

20

30

40

50

号を生成するためのコントローラと、前記1つ以上の制御信号に応答して温度補償電圧を生成するための1つ以上の調整可能なデバイスと、を備える、修正するステップを含む、方法。

【請求項27】

前記ベクトルマトリックス乗算アレイにおける重みは、メモリセルによって実現される、請求項26に記載の方法。

【請求項28】

前記ベクトルマトリックス乗算アレイにおける重みは、差動セルによって実現される、請求項26に記載の方法。

【請求項29】

前記ベクトルマトリックス乗算アレイにおける重みは、1対のブレンドメモリセルによって実現される、請求項26に記載の方法。

【請求項30】

前記メモリセルの前記アレイの前記メモリセルのそれぞれはフラッシュメモリセルであり、前記1つ以上の調整可能なデバイスは、1つ以上の調整可能な電流源を備え、前記1つ以上の調整可能な電流源の各々は、前記1つ以上の制御信号のうちの1つによって調整され、かつレジスタに結合され、前記温度補償電圧は、前記1つ以上の調整可能な電流源と前記レジスタとの間のノードにおいて生成され、前記フラッシュメモリセルに印加されて、前記フラッシュメモリセルの前記電流電圧特性曲線の傾きを修正する、請求項26に記載の方法。

【請求項31】

前記温度補償電圧は、前記フラッシュメモリセルの制御ゲート端子に印加される、請求項30に記載の方法。

【請求項32】

前記温度補償電圧は、前記フラッシュメモリセルの消去ゲート端子に印加される、請求項30に記載の方法。

【請求項33】

前記温度補償電圧は、前記フラッシュメモリセルのソース線端子に印加される、請求項30に記載の方法。

【請求項34】

前記温度補償電圧は、前記フラッシュメモリセルのビット線端子に印加される、請求項30に記載の方法。

【請求項35】

前記フラッシュメモリセルは、スプリットゲートフラッシュメモリセルである、請求項30に記載の方法。

【請求項36】

前記フラッシュメモリセルは、積層ゲートフラッシュメモリセルである、請求項30に記載の方法。

【請求項37】

前記メモリセルの前記アレイの前記メモリセルの前記動作温度が変化するにつれて漏れを低減させるために、前記メモリセルの前記アレイの前記メモリセルの端子に印加されたバイアス電圧を修正するステップを更に含む、請求項26に記載の方法。

【請求項38】

複数のベクトルマトリックス乗算システムを備えるアナログニューロモフィックメモリシステムにおいて温度補償を実行する方法であって、各ベクトルマトリックス乗算システムはメモリセルのアレイを備え、前記方法は、

前記メモリセルの前記アレイのメモリセルの動作温度のレベルが変化するにつれて、前記メモリセルの前記アレイの前記メモリセルの電流電圧特性曲線のために温度補償ブロックによって不連続的に修正するステップを含む、方法。

【請求項39】

10

20

30

40

50

前記ベクトルマトリックス乗算システムにおける重みは、前記メモリセルによって実現される、請求項 38 に記載の方法。

【請求項 40】

前記ベクトルマトリックス乗算システムにおける重みは、差動セルによって実現される、請求項 38 に記載の方法。

【請求項 41】

前記ベクトルマトリックス乗算システムにおける重みは、ブレンドメモリセルによって実現される、請求項 38 に記載の方法。

【請求項 42】

前記メモリセルの前記アレイの前記メモリセルの各々はフラッシュメモリセルであり、前記温度補償ブロックは、前記動作温度を示す出力を生成するための温度センサと、前記温度センサの前記出力にตอบสนองして 1 つ以上の制御ビットを生成するためのコントローラと、複数の電流源であって、各電流源は、スイッチを介してレジスタに選択的に結合され、各スイッチは、前記制御ビットのうちの 1 つによって制御された、複数の電流源と、を備え、前記レジスタの一端において生成された電圧は、前記フラッシュメモリセルに印加されて、前記フラッシュメモリセルの前記電流電圧特性曲線の傾きを修正する、請求項 38 に記載の方法。

10

【請求項 43】

前記修正するステップは、フラッシュメモリセルの電流電圧特性曲線の傾きを修正するステップを含む、請求項 38 に記載のシステム。

20

【請求項 44】

前記メモリセルは、スプリットゲートフラッシュメモリセルである、請求項 38 に記載の方法。

【請求項 45】

前記メモリセルは、積層ゲートフラッシュメモリセルである、請求項 38 に記載の方法。

【請求項 46】

前記温度補償ブロックは、前記動作温度を示す出力を生成するための温度センサと、前記温度センサの前記出力にตอบสนองして 1 つ以上の制御ビットを生成するためのコントローラと、複数のレジスタを備える増幅回路であって、各レジスタは、スイッチを介して前記増幅回路の出力に選択的に結合され、各スイッチは、前記制御ビットのうちの 1 つによって制御された、複数の増幅回路と、を備え、前記増幅回路の前記出力において生成された電圧は、前記メモリセルに印加されて、前記メモリセルの前記電流電圧特性曲線の傾きを修正する、請求項 38 に記載の方法。

30

【請求項 47】

前記メモリセルは、スプリットゲートフラッシュメモリセルである、請求項 46 に記載の方法。

【請求項 48】

前記メモリセルは、積層ゲートフラッシュメモリセルである、請求項 46 に記載の方法。

【請求項 49】

前記メモリセルの前記アレイのメモリセルの前記動作温度が変化するにつれて漏れを低減させるために、前記メモリセルの前記アレイの前記メモリセルの端子に印加されたバイアス電圧を修正するための漏れ低減ブロック

40

を更に備える、請求項 38 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

(優先権の主張)

本出願は、2018年8月27日に出願された米国特許仮出願第62/723,398号、発明の名称「Temperature And Leakage Compensation For Memory Cells in an Analog Neural M

50

emory System Used in a Deep Learning Neural Network」及び2018年11月7日に出願された米国特許出願第16/183,250号、発明の名称「Temperature And Leakage Compensation For Memory Cells In An Analog Neural Memory System Used In A Deep Learning Neural Network」の優先権を主張する。

【0002】

(発明の分野)

深層学習ニューラルネットワークで使用されるアナログニューロモフィックメモリシステムのメモリセルに対する温度補償及び漏れ補償を提供するための多数の実施形態が開示される。

10

【背景技術】

【0003】

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(動物の中樞神経系、特に脳)を模倣しており、多数の入力に依存し得、かつ、一般的に未知である機能を推定する又は近似するために使用される。人工ニューラルネットワークは、概して、メッセージを交換する相互接続した「ニューロン」の層を含む。

【0004】

図1は、人工ニューラルネットワークを図示しており、ここで円は、入力又はニューロンの層を表す。接続(シナプスと呼ばれる)は、矢印によって表され、経験に基づいて調整され得る数値の重みを有する。これは、ニューラルネットワークが入力に適應できるようにし、学習できるようにする。典型的には、ニューラルネットワークは、複数の入力の層を含む。典型的には、1つ以上のニューロンの中間層、及びニューラルネットワークの出力を提供するニューロンの出力層が存在する。各レベルでニューロンは、シナプスから受信されたデータに基づいて個々に又は合わせて決定を行う。

20

【0005】

高性能情報処理用の人工ニューラルネットワークの開発における主要な課題の1つは、適切なハードウェア技術の欠如である。実際には、実用ニューラルネットワークは、非常に多数のシナプスに頼り、ニューロン間の高い接続性、すなわち、非常に高度な計算的並行処理を可能にする。原理的には、このような複雑性は、デジタルスーパーコンピュータ又は専門化したグラフィック処理ユニットクラスタによって実現され得る。しかしながら、高いコストに加え、これらのアプローチはまた、主として低い精度のアナログ計算を実行するのには少ないエネルギーしか消費しない生物学的ネットワークに比べてあまり良くないエネルギー効率に悩まされている。CMOSアナログ回路は、人工ニューラルネットワークに使用されてきたが、ほとんどのCMOS実装シナプス(CMOS-implemented synapses)は、多数のニューロン及びシナプスを考えると嵩高すぎている。

30

【0006】

出願人は以前に、参照により組み込まれる米国特許出願第15/594,439号において、シナプスとして1つ以上の不揮発性メモリアレイを利用する人工(アナログ)ニューラルネットワークを開示した。不揮発性メモリアレイは、アナログニューロモフィックメモリとして動作する。ニューラルネットワークデバイスは、第1の複数の入力を受信するように、かつそれから第1の複数の出力を生成するように構成されている第1の複数のシナプス、及び第1の複数の出力を受信するように構成された第1の複数のニューロンを含む。第1の複数のシナプスは、半導体基板に形成された、間にチャネル領域が延在している離間したソース領域及びドレイン領域と、チャネル領域の第1の部分の上方に配設され、チャネル領域の第1の部分から絶縁された浮遊ゲートと、チャネル領域の第2の部分の上方に配設され、チャネル領域の第2の部分から絶縁された非浮遊ゲートと、を各々が含む複数のメモリセルを含む。複数のメモリセルの各々は、浮遊ゲートの多くの電子に対応する重み値を格納するように構成されている。複数のメモリセルは、第1の複数の入力に、格納された重み値を乗算して第1の複数の出力を生成するように構成されている。

40

50

【 0 0 0 7 】

アナログニューロモーフィックメモリシステムに使用される各不揮発性メモリセルは、消去され、浮遊ゲートに非常に特異的かつ正確な量の電荷、すなわち電子の数、を保持するようにプログラムされなければならない。例えば、各浮遊ゲートはN個の異なる値のうちの1つを保持しなければならない。ここで、Nは、各セルによって示され得る異なる重みの数である。Nの例としては、16、32、64、128及び256が挙げられる。

【 0 0 0 8 】

アナログニューロメモリシステムの実装の1つの課題は、各セルの消去動作、プログラム動作、及び読み出し動作に極めて高い精度が要求されることである。これは、各セルの各浮遊ゲートがN個の値のうちの1つを保持することを要求され得るためであり、Nは従来のフラッシュメモリシステムで使用される従来の値である2より大きい。しかしながら、各デバイスの電流電圧応答特性曲線など、各デバイスの特性は、その動作温度が変化することによって変化することになる。加えて、デバイスの漏れは、その特性に更に影響を及ぼす。

10

【 0 0 0 9 】

必要とされるのは、セルの動作温度が変化することによって各メモリセルの動作において精度及び正確度を確実に維持するために、深層学習ニューラルネットワークで使用されるアナログニューラルメモリシステムのメモリセルに対する温度補償及び漏れ補償を提供するためのシステムである。

【 発明の概要 】

20

【 0 0 1 0 】

深層学習ニューラルネットワークで使用されるアナログニューラルメモリシステムに対する温度補償及び漏れ補償を提供するための多数の実施形態が開示される。温度補償を提供するための実施形態は、慎重な又は連続的な適応的傾き補償及び繰り込みを含む。漏れ補償を提供するための実施形態は、適応的消去ゲート結合、制御ゲート端子への負バイアス、ワード線端子への負バイアス、及びソース線端子へのバイアスを含む。

【 0 0 1 1 】

【 0 0 1 2 】

【 0 0 1 3 】

【 0 0 1 4 】

30

【 0 0 1 5 】

【 0 0 1 6 】

【 0 0 1 7 】

【 0 0 1 8 】

【 0 0 1 9 】

【 0 0 2 0 】

【 0 0 2 1 】

【 0 0 2 2 】

【 0 0 2 3 】

【 0 0 2 4 】

40

【 0 0 2 5 】

【 0 0 2 6 】

【 0 0 2 7 】

【 0 0 2 8 】

【 0 0 2 9 】

【 0 0 3 0 】

【 0 0 3 1 】

【 0 0 3 2 】

【 0 0 3 3 】

【 0 0 3 4 】

50

【 0 0 3 5 】

【 0 0 3 6 】

【 0 0 3 7 】

【 0 0 3 8 】

【 0 0 3 9 】

【 0 0 4 0 】

【 0 0 4 1 】

【 0 0 4 2 】

【 0 0 4 3 】

【 0 0 4 4 】

【 0 0 4 5 】

【 0 0 4 6 】

【 0 0 4 7 】

【 図面の簡単な説明 】

【 0 0 4 8 】

【 図 1 】 先行技術の人工ニューラルネットワークを示す図である。

【 図 2 】 従来の 2 ゲート不揮発性メモリセルの横断面図である。

【 図 3 】 従来の 4 ゲート不揮発性メモリセルの横断面図である。

【 図 4 】 従来の 3 ゲート不揮発性メモリセルの横断面図である。

【 図 5 】 別の従来の 2 ゲート不揮発性メモリセルの横断面図である。

【 図 6 】 不揮発性メモリアレイを利用する異なるレベルの例示的な人工ニューラルネットワークを示す図である。

【 図 7 】 ベクトル乗算器マトリックスを示すブロック図である。

【 図 8 】 様々なレベルのベクトル乗算器マトリックスを示すブロック図である。

【 図 9 】 ベクトル乗算器マトリックスの別の実施形態を示す。

【 図 1 0 】 ベクトル乗算器マトリックスの別の実施形態を示す。

【 図 1 1 】 ベクトル乗算器マトリックスの別の実施形態を示す。

【 図 1 2 】 ベクトル乗算器マトリックスの別の実施形態を示す。

【 図 1 3 】 ベクトル乗算器マトリックスの別の実施形態を示す。

【 図 1 4 】 先行技術の長・短期メモリシステムを示す。

【 図 1 5 】 先行技術の長・短期メモリシステム中の例示的なセルを示す。

【 図 1 6 】 図 1 5 の長・短期メモリシステム中の例示的なセルの実装形態を示す。

【 図 1 7 】 図 1 5 の長・短期メモリシステム中の別の例示的なセルの実装形態を示す。

【 図 1 8 】 先行技術のゲート付き回帰型ユニットシステムを示す。

【 図 1 9 】 先行技術のゲート付き回帰型ユニットシステム中の例示的なセルを示す。

【 図 2 0 】 図 1 9 のゲート付き回帰型ユニットシステム中の例示的なセルの実装形態を示す。

【 図 2 1 】 図 1 9 のゲート付き回帰型ユニットシステム中の別の例示的なセルの実装形態を示す。

【 図 2 2 】 デバイスの動作温度が変化するにつれてデバイスの電流電圧特性曲線がどのように変化するかの一例を示すグラフを示す。

【 図 2 3 】 デバイスの電流電圧特性曲線の傾きを調整するための温度補償の一例を示すグラフを示す。

【 図 2 4 A 】 不連続的な温度補償システムの一実施形態を示す。

【 図 2 4 B 】 不連続的な温度補償システムの一実施形態を示す。

【 図 2 4 C 】 不連続的な温度補償システムの一実施形態を示す。

【 図 2 5 】 不連続的な温度補償システムの一実施形態を示す。

【 図 2 6 A 】 1 つ以上の調整可能な電流源を使用した温度補償の一例を示す。

【 図 2 6 B 】 1 つ以上の調整可能な電流源を使用した温度補償の一例を示す。

【 図 2 7 A 】 フラッシュメモリデバイスの制御ゲートにバイアス電圧を印加する、連続的

10

20

30

40

50

な温度補償システムの別の実施形態を示す。

【図 2 7 B】フラッシュメモリデバイスの制御ゲートにバイアス電圧を印加する、連続的な温度補償システムの別の実施形態を示す。

【図 2 8 A】フラッシュメモリデバイスの消去ゲートにバイアス電圧を印加する、連続的な温度補償システムの別の実施形態を示す。

【図 2 8 B】フラッシュメモリデバイスの消去ゲートにバイアス電圧を印加する、連続的な温度補償システムの別の実施形態を示す。

【図 2 9 A】フラッシュメモリデバイスのソース線端子にバイアス電圧を印加する、連続的な温度補償システムの一実施形態を示す。

【図 2 9 B】フラッシュメモリデバイスのソース線端子にバイアス電圧を印加する、連続的な温度補償システムの一実施形態を示す。

10

【図 3 0】特定の構成のトランジスタを示す。

【図 3 1】特定の構成のメモリセルを示す。

【図 3 2】特定の構成のメモリセルを示す。

【図 3 3】基準メモリセル及び選択されたメモリセルなど、システムの 2 つのデバイスの電流電圧特性曲線の傾きの差を補償するシステムを示す。

【図 3 4】傾き補償システムの一実施形態を示す。

【図 3 5】傾き補償システムの別の実施形態を示す。

【図 3 6】フラッシュメモリデバイスの制御ゲートにバイアス電圧を印加する、漏れ低減システムの一実施形態を示す。

20

【図 3 7】線形領域で動作するメモリセルに好適なアレイアーキテクチャを示す。

【図 3 8】傾き差を補償するために使用されるルックアップテーブルに格納されたデータを示す。

【発明を実施するための形態】

【0049】

本発明の人工ニューラルネットワークは、CMOS 技術及び不揮発性メモリアレイの組み合わせを利用する。

不揮発性メモリセル

【0050】

デジタル不揮発性メモリは、周知である。例えば、参照により本明細書に組み込まれる、米国特許第 5,029,130 号（「130 号特許」）は、フラッシュメモリセルの一種である、スプリットゲート不揮発性メモリセルのアレイを開示する。このようなメモリセル 210 を図 2 に示す。各メモリセル 210 は、半導体基板 12 に形成され、チャンネル領域 18 をその間に有するソース領域 14 とドレイン領域 16 と、を含む。浮遊ゲート 20 は、チャンネル領域 18 の第 1 の部分の上方に形成され、この第 1 の部分から絶縁されて（かつチャンネル領域 18 の第 1 の部分の伝導率を制御して）おり、またソース領域 14 の一部分の上方に形成されている。ワード線端子 22（典型的には、ワード線に結合される）は、チャンネル領域 18 の第 2 の部分の上方に配設され、チャンネル領域 18 の第 2 の部分から絶縁され、（かつチャンネル領域 18 の第 2 の部分の伝導率を制御する）第 1 の部分と、浮遊ゲート 20 の上方で上方向に延在する第 2 の部分と、を有する。浮遊ゲート 20 及びワード線端子 22 は、ゲート酸化物によって基板 12 から絶縁される。ビット線 24 はドレイン領域 16 に結合される。

30

【0051】

メモリセル 210 は、ワード線端子 22 に高い正電圧を印加することによって消去され（ここで電子は、浮遊ゲートから除去される）、それにより、ファウラーノルドハイムトンネリングによって浮遊ゲート 20 からワード線端子 22 まで中間絶縁体を通して浮遊ゲート 20 の電子をトンネリングさせる。

【0052】

メモリセル 210 は、ワード線端子 22 に正電圧、及びソース領域 14 に正電圧を印加することによってプログラムされる（ここで電子は、浮遊ゲートに加えられる）。電子電

50

流は、ソース領域 1 4 からドレイン領域 1 6 に向かって流れることになる。電子は、ワード線端子 2 2 と浮遊ゲート 2 0 との間隙に達すると、加速し発熱する。熱せられた電子の一部が、浮遊ゲート 2 0 からの静電引力に起因して、浮遊ゲート 2 0 にゲート酸化物を介して注入される。

【 0 0 5 3 】

ドレイン領域 1 6 及びワード線端子 2 2 に正の読み出し電圧を印加する（ワード線端子下のチャンネル領域 1 8 の部分をターンオンすることによって、メモリセル 2 1 0 が読み出される。浮遊ゲート 2 0 が正に帯電する（すなわち、電子を消去する）場合、浮遊ゲート 2 0 下のチャンネル領域 1 8 の部分は、次に同様にオンになり、電流は、チャンネル領域 1 8 を流れ、これは、消去された状態又は「 1 」の状態として検知される。浮遊ゲート 2 0 が負に帯電する（すなわち、電子でプログラムされる）場合、次に浮遊ゲート 2 0 下のチャンネル領域の部分は、ほとんど又は完全にオフになり、電流は、チャンネル領域 1 8 を流れず（又はわずかに流れる）、プログラムされた状態又は「 0 」の状態として検知される。

10

【 0 0 5 4 】

表 1 は、読み出し動作、消去動作、及びプログラム動作を実行するためのメモリセル 1 1 0 の端子に印加され得る典型的な電圧範囲を示す。

表 1：図 3 のフラッシュメモリセル 2 1 0 の動作

【表 1】

	WL	BL	SL
読み出し	2~3V	0.6~2V	0V
消去	約11~13V	0V	0V
プログラム	1~2V	1~3μA	9~10V

20

【 0 0 5 5 】

他の種類のフラッシュメモリセルである、他のスプリットゲートメモリセル構成が知られている。例えば、図 3 は、ソース領域 1 4 と、ドレイン領域 1 6 と、チャンネル領域 1 8 の第 1 の部分の上方にある浮遊ゲート 2 0 と、チャンネル領域 1 8 の第 2 の部分の上方にある選択ゲート 2 2（典型的には、ワード線、WL、に結合される）と、浮遊ゲート 2 0 の上方にある制御ゲート 2 8 と、及びソース領域 1 4 の上方にある消去ゲート 3 0 と、を備える 4 ゲートメモリセル 3 1 0 を示す。この構成は、あらゆる目的のため参照により本明細書に組み込まれる、米国特許第 6, 7 4 7, 3 1 0 号に記載されている。ここで、全てのゲートは、浮遊ゲート 2 0 を除いて、非浮遊ゲートであり、それらは電圧源に電氣的に接続されている又は接続可能であることを意味する。プログラミングは、浮遊ゲート 2 0 にそれ自体を注入するチャンネル領域 1 8 からの熱せられた電子によって実行される。消去は、浮遊ゲート 2 0 から消去ゲート 3 0 へトンネリングする電子によって実行される。

30

【 0 0 5 6 】

表 2 は、読み出し動作、消去動作、及びプログラム動作を実行するためのメモリセル 3 1 0 の端子に印加され得る典型的な電圧範囲を示す。

表 2：図 3 のフラッシュメモリセル 3 1 0 の動作

【表 2】

	WL/SG	BL	CG	EG	SL
読み出し	1.0~2V	0.6~2V	0~2.6V	0~2.6V	0V
消去	-0.5V/0V	0V	0V/-8V	8~12V	0V
プログラム	1V	1μA	8~11V	4.5~9V	4.5~5V

40

【 0 0 5 7 】

50

図 4 は、別の種類のフラッシュメモリセルである、3 ゲートメモリセル 4 1 0 を示す。メモリセル 4 1 0 は、メモリセル 4 1 0 が別個の制御ゲートを有しないことを除いて、図 3 のメモリセル 3 1 0 と同一である。消去動作（消去ゲートの使用を通して消去が起こる）及び読み出し動作は、制御ゲートバイアスが印加されないことを除いて、図 3 のものと同様である。プログラミング動作もまた、制御ゲートバイアスなしで行われるため、結果として、プログラム動作中は、制御ゲートバイアスの不足を補償するため、より高い電圧がソース線に印加されなければならない。

【 0 0 5 8 】

表 3 は、読み出し動作、消去動作、及びプログラム動作を実行するためのメモリセル 4 1 0 の端子に印加され得る典型的な電圧範囲を示す。

表 3：図 4 のフラッシュメモリセル 4 1 0 の動作

【表 3】

	WL/SG	BL	EG	SL
読み出し	0.7~2.2V	0.6~2V	0~2.6V	0V
消去	-0.5V/0V	0V	11.5V	0V
プログラム	1V	2~3 μ A	4.5V	7~9V

【 0 0 5 9 】

図 5 は、別の種類のフラッシュメモリセルである、積層ゲートメモリセル 5 1 0 を示す。メモリセル 5 1 0 は、浮遊ゲート 2 0 がチャネル領域 1 8 全体にわたって延在し、制御ゲート 2 2（ここではワード線に結合される）が絶縁層（図示せず）によって分離された浮遊ゲート 2 0 の上方に延在することを除いて、図 2 のメモリセル 2 1 0 と同様である。消去動作、プログラミング動作、及び読み出し動作は、メモリセル 2 1 0 について前述したものと同様の方法で動作する。

【 0 0 6 0 】

表 4 は、読み出し動作、消去動作、及びプログラム動作を実行するためのメモリセル 5 1 0 及び基板 1 2 の端子に印加され得る典型的な電圧範囲を示す。

表 4：図 5 のフラッシュメモリセル 5 1 0 の動作

【表 4】

	CG	BL	SL	基板
読み出し	2~5V	0.6~2V	0V	0V
消去	-8~-10V/0V	FLT	FLT	8~10V/15~20V
プログラム	8~12V	3~5V	0V	0V

【 0 0 6 1 】

上記の人工ニューラルネットワークにおける不揮発性メモリセルの種類の中の 1 つを含むメモリアレイを利用するために、2 つの修正が行われる。第 1 に、以下に更に説明されるように、アレイの他のメモリセルのメモリ状態に悪影響を与えずに各メモリセルを個々にプログラム、消去、及び読み出しすることができるよう線を構成する。第 2 に、メモリセルの連続（アナログ）プログラミングを提供する。

【 0 0 6 2 】

具体的には、アレイの各メモリセルのメモリ状態（すなわち、浮遊ゲートの電荷）を、完全に消去された状態から完全にプログラムされた状態へ、独立して、かつ他のメモリセルの異常が最小で連続的に変えることができる。別の実施形態では、アレイの各メモリセルのメモリ状態（すなわち、浮遊ゲートの電荷）を、完全にプログラムされた状態から完全に消去された状態へ、及び逆もまた同様に、独立して、かつ他のメモリセルの異常が最

10

20

30

40

50

小で連続的に変えることができる。このことは、セル格納がアナログであるか、又は多数の不連続値（16又は64の異なる値など）のうちの1つを最低限格納することができ、これにより、メモリアレイの全てのセルの非常に正確、かつ個々の調整を可能にし、またメモリアレイを格納に対して理想的にし、ニューラルネットワークのシナプシスの重みに微調整を加えることを意味する。

不揮発性メモリアレイを使用するニューラルネットワーク

【0063】

図6は、本実施形態の不揮発性メモリアレイを利用するニューラルネットワークの非限定実施例を概念的に示す。この例は、顔認識アプリケーションのために不揮発性メモリアレイニューラルネットワークを使用するが、不揮発性メモリアレイベースのニューラルネットワークを使用して他の任意の適切なアプリケーションを実装することができる。

10

【0064】

S0は入力層であり、この例では、5ビット精度の 32×32 ピクセルRGB画像である（すなわち、各色R、G、及びBにつき1つずつの3つの 32×32 ピクセルアレイ、各ピクセルが5ビット精度である）。入力層S0から層C1に行くシナプスCB1は、ある事例では、異なる重みのセットが適用され、別の事例では、共有される重みが適用され、入力画像を 3×3 ピクセルの重なり合うフィルタでスキャンし（カーネル）、1ピクセル（又はモデルによって決まるように2ピクセル以上）ずつフィルタをシフトする。具体的には、画像の 3×3 部分における9ピクセルの値（すなわち、フィルタ又はカーネルと呼ばれる）は、シナプスCB1に提供され、これらの9個の入力値に適切な重みを乗算し、その乗算の出力の合計後、単一の出力値が決定され、特徴マップC1の層の1つのピクセルを生成するためにCB1の第1のシナプスによって与えられる。 3×3 フィルタは次に、入力層S0内部の右側に1ピクセルだけシフトされ（すなわち、3ピクセルの列を右側に追加し、左側で3ピクセルの列をドロップする）、この新しく位置づけられたフィルタの9ピクセル値が、シナプスCB1に提供されるため、それらに同じ重みを乗算し、関連するシナプスによって第2の単一の出力値を決定する。このプロセスを、 3×3 フィルタが入力層S0の 32×32 ピクセル画像全体にわたって3色全て及び全てのビットについてスキャンするまで続ける（精度値）。プロセスは次に、層C1の特徴マップ全てが計算されるまで、異なる重みのセットを使用して繰り返されて、C1の異なる特徴マップを生成する。

20

30

【0065】

層C1において、本例では、各々 30×30 ピクセルを有する16個の特徴マップが存在する。各ピクセルは、入力とカーネルとの乗算から抽出された新しい特徴ピクセルであり、したがって、各特徴マップは、2次元アレイであり、したがってこの例では、層C1は、2次元アレイの16層を構成する（本明細書で言及される層及びアレイは、必ずしも物理的な関係ではなく論理関係であり、すなわち、アレイは必ずしも物理的な2次元アレイに配向されないことに留意する）。層C1の16個の特徴マップの各々は、フィルタスキャンに適用される16個の異なるシナプス重みのセットの1つによって生成される。C1特徴マップは全て、境界同定など同じ画像特徴の異なる態様を対象とすることができる。例えば、第1のマップ（この第1のマップを生成するのに使用される全てのスキャンに共有される第1の重みセットを使用して生成される）は、円形エッジを識別することができ、第2のマップ（第1の重みセットと異なる第2の重みセットを使用して生成される）は、方形エッジ又は特定の特徴のアスペクト比などを識別することができる。

40

【0066】

各特徴マップ中の連続する、重なり合わない 2×2 領域からの値をプールする活性化関数P1（プーリング）は、層C1から層S1へ行く前に適用される。プーリング機能の目的は、平均して近隣の位置にすること（又は $m \times n$ 関数もまた使用され得る）、例えばエッジ位置の依存を低減すること、及び次の段階に行く前にデータサイズを低減することである。層S1において、16個の 15×15 特徴マップ（すなわち、各々 15×15 ピクセルの16個異なるアレイ）が存在する。層S1から層C2に行くシナプスCB2は、S

50

1のマップを1ピクセルのフィルタシフトを使用して4×4フィルタでスキャンする。層C2において、22個の12×12特徴マップが存在する。各特徴マップ中の連続する、重なり合わない2×2領域からの値をプールの活性化関数P2(プーリング)は、層C2から層S2へ行く前に適用される。層S2において、22個の6×6特徴マップが存在する。活性化関数(プーリング)は、層S2から層C3へ行くシナプスCB3で適用され、ここで層C3の全てのニューロンは、CB3のそれぞれのシナプスを介して層S2の全てのマップに接続する。層C3において、64個のニューロンが存在する。層C3から出力層S3へと行くシナプスCB4は、C3~S3を完全に接続し、すなわち、層C3の全てのニューロンは、層S3の全てのニューロンに接続される。S3における出力は、10個のニューロンを含み、ここで最も高い出力ニューロンは、クラスを決定する。この出力は、例えば、元の画像の内容の同定又は分類を示すことができる。

10

【0067】

シナプスの各層は、不揮発性メモリセルのレイ又はレイの一部を使用して実装される。

【0068】

図7は、その目的のために使用され得るレイのブロック図である。ベクトルマトリックス乗算(Vector-by-matrix multiplication、VMM)レイ32は、不揮発性メモリセルを含み、ある層と次の層との間のシナプス(図6のCB1、CB2、CB3、及びCB4など)として利用される。具体的には、VMMレイ32は、不揮発性メモリセルのレイ33、消去ゲート及びワード線ゲートデコーダ34、制御ゲートデコーダ35、ビット線デコーダ36、並びにソース線デコーダ37を含み、それらのデコーダは不揮発性メモリセルレイ33に対するそれぞれの入力をデコードする。VMMレイ32への入力、消去ゲート及びワード線ゲートデコーダ34から、又は制御ゲートデコーダ35からとされ得る。この例におけるソース線デコーダ37はまた、不揮発性メモリセルレイ33の出力をデコードする。あるいは、ビット線デコーダ36が、不揮発性メモリセルレイ33の出力をデコードすることができる。

20

【0069】

不揮発性メモリセルレイ33は、2つの目的を果たす。第1に、VMMレイ32により使用される重みを格納する。第2に、不揮発性メモリセルレイ33は、不揮発性メモリセルレイ33に格納された重みによって、入力を有効に乗算して、それらを出力線(ソース線又はビット線)ごとに加算して、次の層への入力又は最後の層への入力になる、出力を生成する。乗算及び加算の関数を実行することによって、不揮発性メモリセルレイ33は、別個の乗算及び加算の論理回路の必要性をなくし、また、その現場でのメモリ計算により電力効率も良い。

30

【0070】

不揮発性メモリセルレイ33の出力は、不揮発性メモリセルレイ33の出力を合計してその畳み込み用の単一の値を作成する、差動加算器(加算オペアンプ又は加算電流ミラーなど)38に供給される。差動加算器38は、正の重み及び負の重みの総和を実行するように配置される。

【0071】

差動加算器38の合計された出力値は、次に出力を整流する活性化関数回路39に供給される。活性化関数回路39は、シグモイド、tanh、又はReLU関数を提供し得る。活性化関数回路39の整流された出力値は、次の層(例えば図6のC1)として特徴マップの要素になり、次いで、次のシナプスに適用されて次の特徴マップ層又は最後の層を生成する。したがって、この例では、不揮発性メモリセルレイ33は、複数のシナプスを構成し(ニューロンの前の層から、又は画像データベースなどの入力層から入力を受信する)、加算オペアンプ38及び活性化関数回路39は、複数のニューロンを構成する。

40

【0072】

図7のVMMレイ32への入力(WLx、EGx、CGx、及び任意選択的にBLx及びSLx)は、アナログレベル、バイナリレベル、又はデジタルビット(この場合、D

50

A C は、デジタルビットを適切な入力アナログレベルに変換するために提供される)であり得、出力は、アナログレベル、バイナリレベル、又はデジタルビットであり得る(この場合、出力A D Cは出力アナログレベルをデジタルビットに変換するために提供される)。

【0073】

図8は、ここでVMMアレイ32a、32b、32c、32d及び32eとして標示されたVMMアレイ32の多数の層の使用を示すブロック図である。図8に示されるように、Input xで示される、入力は、デジタルアナログ変換器31によってデジタルからアナログに変換され、入力VMMアレイ32aに提供される。変換されたアナログ入力は、電圧又は電流であり得る。第1の層の入力D/A変換は、入力VMMアレイ32aのマトリクス乗算器の適切なアナログレベルに入力Input xをマッピングする関数又はLUT(ルックアップテーブル)を使用することによって行うことができる。入力変換はまた、外部アナログ入力を入力VMMアレイ32aへのマッピングされたアナログ入力に変換するために、アナログ-アナログ(A/A)変換器によって行うこともできる。

10

【0074】

入力VMMアレイ32aによって生成された出力は、次に、入力VMMアレイ(隠しレベル2)32cへの入力として提供される出力を生成する次のVMMアレイ(隠しレベル1)32bへの入力として提供される、などとなる。VMMアレイ32の様々な層は、畳み込みニューラルネットワーク(convolutional neural network、CNN)のシナプス及びニューロンの異なる層として機能する。各VMMアレイ32a、32b、32c、32d及び32eは、スタンドアローンの物理的不揮発性メモリアレイであり得、又は複数のVMMアレイは、同じ物理的不揮発性メモリアレイの異なる部分を利用することができ、又は複数のVMMアレイは、同じ物理的不揮発性メモリアレイの重なり合う部分を利用することができる。図8に示される例は、5つの層(32a、32b、32c、32d、32e)、すなわち、1つの入力層(32a)、2つの隠れ層(32b、32c)、及び2つの完全に接続された層(32d、32e)を含む。当業者であれば、これは単なる例示であり、代わりにシステムが2つを超える隠れ層及び2つを超える完全に接続された層を含み得ることを理解するであろう。

20

ベクトルマトリクス乗算(VMM)アレイ

【0075】

図9は、図3に示されるメモリセル310に特に適したニューロンVMMアレイ900を示し、入力層と次の層との間でシナプス及びニューロンの一部として利用される。VMMアレイ900は、不揮発性メモリセルのメモリアレイ901、及び不揮発性基準メモリセルの基準アレイ902(アレイの頂部で)を備える。あるいは、別の基準アレイが底部に位置することができる。

30

【0076】

VMMアレイ900では、制御ゲート線903などの制御ゲート線が垂直方向に延びており(したがって、行方向の基準アレイ902が、制御ゲート線903に直交する)、消去ゲート線904などの消去ゲート線が水平方向に延びている。ここで、VMMアレイ900への入力、制御ゲート線(CG0、CG1、CG2、CG3)に提供され、VMMアレイ900の出力はソース線(SL0、SL1)に現れる。一実施形態では、偶数行のみが使用され、別の実施形態では、奇数行のみが使用される。各ソース線(SL0、SL1、それぞれに)の電流は、その特定のソース線に接続されたメモリセルからの全ての電流の合計関数を実行する。

40

【0077】

ニューラルネットワークについて本明細書に記載されているように、VMMアレイ900の不揮発性メモリセル、すなわちVMMアレイ900のフラッシュメモリは、サブスレッショルド領域で動作するように構成されることが好ましい。

【0078】

本明細書に記載される不揮発性基準メモリセル及び不揮発性メモリセルは、以下のように弱い反転でバイアスされ：

50

$$I_{ds} = I_o * e^{(V_g - V_{th})/kV_t} = w * I_o * e^{(V_g)/kV_t}$$

ここで、 $w = e^{(-V_{th})/kV_t}$ である。

【0079】

メモリセル（基準メモリセル又は周辺メモリセルなど）又はトランジスタを使用して入力電流を入力電圧に変換するI - Vログ変換器を使用した場合：

$$V_g = k * V_t * \log [I_{ds} / w_p * I_o]$$

ここで、 w_p は、基準又は周辺メモリセルの w である。

【0080】

ベクトルマトリックス乗算器VMMアレイとして使用されるメモリアレイについて、出力電流は以下である：

$$I_{out} = w_a * I_o * e^{(V_g)/kV_t}、すなわち$$

$$I_{out} = (w_a / w_p) * I_{in} = W * I_{in}$$

$$W = e^{(V_{thp} - V_{tha})/kV_t}$$

ここで、メモリアレイの各メモリセルの $w_a = w$ である。

【0081】

ワード線又は制御ゲートは、入力電圧のためのメモリセルの入力として使用することができる。

【0082】

あるいは、本明細書に記載されたVMMアレイのフラッシュメモリセルは、線形領域で動作するように構成することができる。

$$I_{ds} = \mu * (V_{gs} - V_{th}) * V_{ds} ; \quad C_{ox} = \epsilon_0 * \epsilon_r * q / L$$

$$W = \mu * (V_{gs} - V_{th}) * L$$

【0083】

ワード線又は制御ゲート又はビット線又はソース線は、入力電圧のための線形領域内で動作するメモリセルの入力として使用することができる。

【0084】

I - V線形変換器では、線形領域で動作するメモリセル（基準メモリセル又は周辺メモリセルなど）又はトランジスタを、入出力電流を入出力電圧に線形変換するために使用することができる。

【0085】

図7のVMMアレイ32のための他の実施形態は、参照により本明細書に組み込まれる米国特許出願第15/826,345号に記載されているとおりである。上記出願に記載されているように、ソース線又はビット線は、ニューロン出力（電流和出力）として使用することができる。

【0086】

図10は、図2に示されるメモリセル210に特に適したニューロンVMMアレイ1000を示し、入力層と次の層との間のシナプスとして利用される。VMMアレイ1000は、不揮発性メモリセルのメモリアレイ1003、第1の不揮発性基準メモリセルの基準アレイ1001、及び第2の不揮発性基準メモリセルの基準アレイ1002を備える。アレイの列方向に配置された基準アレイ1001及び1002は、端子BLR0、BLR1、BLR2、及びBLR3に流入する電流入力を電圧入力WL0、WL1、WL2、及びWL3に変換するように機能する。実際には、第1及び第2の不揮発性基準メモリセルは、電流入力が流入する状態で、マルチプレクサ1014（一部のみ示す）を通してダイオード接続される。基準セルは、標的基準レベルに調整（例えば、プログラム）される。標的基準レベルは、基準ミアレイマトリックス（図示せず）によって提供される。

【0087】

メモリアレイ1003は、2つの目的を果たす。第1に、VMMアレイ1000により使用される重みを、それぞれのメモリセルに格納する。第2に、メモリアレイ1003は、メモリアレイ1003に格納された重みによって、入力を（すなわち、端子BLR0、BLR1、BLR2、及びBLR3に提供された電流入力であり、これにより基準アレイ

10

20

30

40

50

1001及び1002がワード線WL0、WL1、WL2、及びWL3に供給するための入力電圧に変換される)有効に乗算して、次いで、全ての結果(メモリセル電流)を加算して、次の層への入力又は最後の層への入力となる、それぞれのビット線(BL0~BLN)の出力を生成する。乗算及び加算の関数を実行することによって、メモリアレイ1003は、別個の乗算及び加算の論理回路の必要性をなくし、また、電力効率も良い。ここで、電圧入力はワード線WL0、WL1、WL2、及びWL3に提供され、出力は、読み出し(推論)動作中にそれぞれのビット線BL0~BLNに現れる。ビット線BL0~BLNの各々の電流は、その特定のビット線に接続された全ての不揮発性メモリセルからの電流の合計関数を実行する。

【0088】

表5は、VMMアレイ1000の動作電圧を示す。表中の列は、選択セルのワード線、非選択セルのワード線、選択セルのビット線、非選択セルのビット線、選択セルのソース線、及び非選択セルのソース線の電圧を示す。行は、読み出し、消去、及びプログラムの動作を示す。

表5：図10のVMMアレイ1000の動作

【表5】

	WL	WL-非選択	BL	BL-非選択	SL	SL-非選択
読み出し	1~3.5V	-0.5V/0V	0.6~2V (1ニューロン)	0.6V~2V/0V	0V	0V
消去	約5~13V	0V	0V	0V	0V	0V
プログラム	1~2V	-0.5V/0V	0.1~3uA	Vinh約2.5V	4~10V	0~1V/FLT

【0089】

図11は、図2に示されるメモリセル210に特に適したニューロンVMMアレイ1100を示し、入力層と次の層との間でシナプス及びニューロンの一部として利用される。VMMアレイ1100は、不揮発性メモリセルのメモリアレイ1103、第1の不揮発性基準メモリセルの基準アレイ1101、及び第2の不揮発性基準メモリセルの基準アレイ1102を備える。基準アレイ1101及び1102は、VMMアレイ1100の行方向に延びる。VMMアレイは、VMMアレイ1100においてワード線が垂直方向に延びることを除いて、VMM1000と同様である。ここで、入力はワード線(WLA0、WLB0、WLA1、WLB2、WLA2、WLB2、WLA3、WLB3)に提供され、出力は、読み出し動作中にソース線(SL0、SL1)に現れる。各ソース線の電流は、その特定のソース線に接続されたメモリセルからの全ての電流の合計関数を実行する。

【0090】

表6は、VMMアレイ1100の動作電圧を示す。表中の列は、選択セルのワード線、非選択セルのワード線、選択セルのビット線、非選択セルのビット線、選択セルのソース線、及び非選択セルのソース線の電圧を示す。行は、読み出し、消去、及びプログラムの動作を示す。

表6：図11のVMMアレイ1100の動作

10

20

30

40

50

【表 6】

	WL	WL-非選択	BL	BL-非選択	SL	SL-非選択
読み出し	1~3.5V	-0.5V/0V	0.6~2V	0.6V~2V/0V	約0.3~1V (Iニューロン)	0V
消去	約5~13V	0V	0V	0V	0V	SL-inhibit (約4~8V)
プログラム	1~2V	-0.5V/0V	0.1~3uA	Vinh約2.5V	4~10V	0~1V/FLT

10

【0091】

図12は、図3に示されるメモリセル310に特に適したニューロンVMMアレイ1200を示し、入力層と次の層との間でシナプス及びニューロンの一部として利用される。VMMアレイ1200は、不揮発性メモリセルのメモリアレイ1203、第1の不揮発性基準メモリセルの基準アレイ1201、及び第2の不揮発性基準メモリセルの基準アレイ1202を備える。基準アレイ1201及び1202は、端子BLR0、BLR1、BLR2、及びBLR3に流入する電流入力を電圧入力CG0、CG1、CG2、及びCG3に変換するように機能する。実際には、第1及び第2の不揮発性基準メモリセルは、電流入力がBLR0、BLR1、BLR2、及びBLR3を通して流入する状態で、マルチプレクサ1212（一部のみ示す）を通してダイオード接続される。マルチプレクサ1212は、読み出し動作中に第1及び第2の不揮発性基準メモリセルの各々のビット線（BLR0など）の一定電圧を確実にするために、それぞれのマルチプレクサ1205及びカスコーディングトランジスタ1204を各々含む。基準セルは、標的基準レベルに調整される。

20

【0092】

メモリアレイ1203は、2つの目的を果たす。第1に、VMMアレイ1200により使用される重みを格納する。第2に、メモリアレイ1203は、メモリアレイに格納された重みによって、入力を（端子BLR0、BLR1、BLR2、及びBLR3へ提供された電流入力であり、基準アレイ1201及び1202は、これらの電流入力を制御ゲート（CG0、CG1、CG2、及びCG3）に供給するための入力電圧に変換する）有効に乗算して、次いで、全ての結果（セル電流）を加算して、BL0~BLNに出現し、かつ次の層への入力又は最後の層への入力となる、出力を生成する。乗算及び加算の関数を実行することによって、メモリアレイは、別個の乗算及び加算の論理回路の必要性をなくし、また、電力効率も良い。ここで、入力は制御ゲート線（CG0、CG1、CG2、及びCG3）に提供され、出力は、読み出し動作中にビット線（BL0~BLN）に現れる。各ビット線の電流は、その特定のビット線に接続されたメモリセルからの全ての電流の合計関数を実行する。

30

【0093】

VMMアレイ1200は、メモリアレイ1203の不揮発性メモリセルの一方調整を実装する。すなわち、各不揮発性メモリセルは消去され、次いで、浮遊ゲートの所望の電荷に達するまで部分的にプログラムされる。これは、例えば、以下に記載される新規の精密プログラミング技術を使用して実行することができる。過度に多くの電荷が浮遊ゲートに加わる場合（誤った値がセルに格納されるなど）、セルは消去されなければならない。示されるように、同じ消去ゲート（EG0又はEG1など）を共有する2つの行は、一緒に消去される必要があり（ページ消去として知られる）、その後、各セルは、浮遊ゲートの所望の電荷に達するまで部分的にプログラムされる。

40

【0094】

表7は、VMMアレイ1200の動作電圧を示す。表中の列は、選択セルのワード線、

50

非選択セルのワード線、選択セルのビット線、非選択セルのビット線、選択セルの制御ゲート、選択セルと同じセクタの非選択セルの制御ゲート、選択セルとは異なるセクタの非選択セルの制御ゲート、選択セルの消去ゲート、非選択セルの消去ゲート、選択セルのソース線、及び非選択セルのソース線の電圧を示す。行は、読み出し、消去、及びプログラムの動作を示す。

表 7：図 1 2 の VMM アレイ 1 2 0 0 の動作

	WL	WL-非選択	BL	BL-非選択	CG	CG-同じセクタを非選択	CG-非選択	EG	EG-非選択	SL	SL-非選択
読み出し	1. 0~2V	-0. 5V/0V	0. 6~2V (ニューロン)	0V	0~2. 6V	0~2. 6V	0~2. 6V	0~2. 6V	0~2. 6V	0V	0V
消去	0V	0V	0V	0V	0V	0~2. 6V	0~2. 6V	5~12V	0~2. 6V	0V	0V
プログラム	0. 7~1V	-0. 5V/0V	0. 1~1uA	Vinh (1~2V)	4~11V	0~2. 6V	0~2. 6V	4. 5~5V	0~2. 6V	4. 5~5V	0~1V

10

【 0 0 9 5 】

図 1 3 は、図 3 に示されるメモリセル 3 1 0 に特に適したニューロン VMM アレイ 1 3 0 0 を示し、入力層と次の層との間でシナプス及びニューロンの一部として利用される。VMM アレイ 1 3 0 0 は、不揮発性メモリセルのメモリアレイ 1 3 0 3、基準アレイ 1 3 0 1 又は第 1 の不揮発性基準メモリセル、及び第 2 の不揮発性基準メモリセルの基準アレイ 1 3 0 2 を備える。EG 線 EGR 0、EG 0、EG 1、及び EGR 1 は垂直に延び、CG 線 CG 0、CG 1、CG 2、及び CG 3 並びに SL 線 WL 0、WL 1、WL 2、及び WL 3 は水平に延びる。VMM アレイ 1 3 0 0 は、VMM アレイ 1 3 0 0 が双方向調整を実装することを除いて VMM アレイ 1 4 0 0 と同様であり、各個々のセルは、個別の EG 線の使用により、浮遊ゲートの所望の電荷量に達するために、完全に消去され、部分的にプログラムされ、必要に応じて部分的に消去され得る。示されるように、基準アレイ 1 3 0 1 及び 1 3 0 2 は、端子 BLR 0、BLR 1、BLR 2 及び BLR 3 の入力電流を制御ゲート電圧 CG 0、CG 1、CG 2 及び CG 3 に変換し (マルチプレクサ 1 3 1 4 を介したダイオード接続された基準セルのアクションを通じて)、行方向でメモリセルに印加される。電流出力 (ニューロン) は、ビット線 BL 0 ~ BL N 中にあり、各ビット線は、その特定のビット線に接続された不揮発性メモリセルからの全ての電流を合計する。

20

【 0 0 9 6 】

図 8 は、VMM アレイ 1 3 0 0 の動作電圧を示す。表中の列は、選択セルのワード線、非選択セルのワード線、選択セルのビット線、非選択セルのビット線、選択セルの制御ゲート、選択セルと同じセクタの非選択セルの制御ゲート、選択セルとは異なるセクタの非選択セルの制御ゲート、選択セルの消去ゲート、非選択セルの消去ゲート、選択セルのソース線、及び非選択セルのソース線の電圧を示す。行は、読み出し、消去、及びプログラムの動作を示す。

30

表 8：図 1 3 の VMM アレイ 1 3 0 0 の動作

【表 8】

	WL	WL-非選択	BL	BL-非選択	CG	CG-同じセクタを非選択	CG-非選択	EG	EG-非選択	SL	SL-非選択
読み出し	1. 0~2V	-0. 5V/0V	0. 6~2V (ニューロン)	0V	0~2. 6V	0~2. 6V	0~2. 6V	0~2. 6V	0~2. 6V	0V	0V
消去	0V	0V	0V	0V	0V	4~9V	0~2. 6V	5~12V	0~2. 6V	0V	0V
プログラム	0. 7~1V	-0. 5V/0V	0. 1~1uA	Vinh (1~2V)	4~11V	0~2. 6V	0~2. 6V	4. 5~5V	0~2. 6V	4. 5~5V	0~1V

40

長・短期メモリ

【 0 0 9 7 】

先行技術は、長・短期メモリ (long short-term memory、LSTM) として知られる概念を含む。LSTM ユニットは、しばしば、ニューラルネットワークで使用される。LSTM は、ニューラルネットワークが、所定の任意の時間間隔にわたって情報を記憶する

50

ことを可能にし、後続の動作においてその情報を使用することを可能にする。従来の LSTM ユニットは、セル、入力ゲート、出力ゲート、及びフォワードゲートを含む。3つのゲートは、セル内及びセル外への情報の流れ、及び情報が LSTM で記憶される時間間隔を調整する。VMM は、LSTM ユニットにおいて特に有用である。

【0098】

図14は、例示的な LSTM 1400 を示す。この例における LSTM 1400 は、セル 1401、1402、1403 及び 1404 を含む。セル 1401 は入力ベクトル x_0 を受け取り、出力ベクトル h_0 及びセル状態ベクトル c_0 を生成する。セル 1402 は、入力ベクトル x_1 と、セル 1401 から出力ベクトル (隠れた状態) h_0 と、セル 1401 からセル状態ベクトル c_0 とを受け取り、出力ベクトル h_1 とセル状態ベクトル c_1 とを生成する。セル 1403 は、入力ベクトル x_2 と、セル 1402 から出力ベクトル (隠れた状態) h_1 と、セル 1402 からセル状態ベクトル c_1 とを受け取り、出力ベクトル h_2 とセル状態ベクトル c_2 とを生成する。セル 1404 は、入力ベクトル x_3 と、セル 1403 から出力ベクトル (隠れた状態) h_2 と、セル 1403 からセル状態ベクトル c_2 とを受け取り、出力ベクトル h_3 を生成する。追加のセルを使用することができ、4つのセルを有する LSTM は単なる例である。

10

【0099】

図15は、図14のセル 1401、1402、1403 及び 1404 に使用することができる、LSTM セル 1500 の例示的な実装形態を示す。LSTM セル 1500 は、先行するセルから入力ベクトル $x(t)$ 、セル状態ベクトル $c(t-1)$ を受け取り、先行するセルから出力ベクトル $h(t-1)$ を受け取り、セル状態ベクトル $c(t)$ 及び出力ベクトル $h(t)$ を生成する。

20

【0100】

LSTM セル 1500 は、シグモイド関数デバイス 1501、1502、1503 を備え、各々が、入力ベクトルの各構成要素のうちどれだけの量が出力ベクトルを通して許容されるかを制御するために、0~1の数を用いる。LSTM セル 1500 はまた、入力ベクトルに双曲線 \tanh 関数を適用するための \tanh デバイス 1504 及び 1505 と、2つのベクトルを一緒に乗算するための乗算器デバイス 1506、1507 及び 1508 と、2つのベクトルを一緒に加算するための加算器デバイス 1509 とを備える。出力ベクトル $h(t)$ は、システムの次の LSTM セルに提供することができ、又は他の目的でアクセスすることができる。

30

【0101】

図16は、LSTM セル 1500 の一実装形態である LSTM セル 1600 を示す。読者の便宜のために、LSTM セル 1500 からの同じ採番方法が、LSTM セル 1600 で使用される。シグモイド関数デバイス 1501、1502 及び 1503 及び \tanh デバイス 1504 は各々、複数の VMM アレイ 1601 及び活性化回路ブロック 1602 を備える。したがって、VMM アレイは、特定のニューラルネットワークシステムで使用される LSTM セルにおいて特に有用であることが分かり得る。

【0102】

LSTM セル 1600 の代替例 (及び LSTM セル 1500 の実装形態の別の例) を図17に示す。図17では、シグモイド関数デバイス 1501、1502、及び 1503、並びに \tanh デバイス 1504 は、同じ物理ハードウェア (VMM アレイ 1701 及び活性化関数ブロック 1702) を、時分割多重化された方式で共有する。LSTM セル 1700 はまた、2つのベクトルを一緒に乗算するための乗算器デバイス 1703 と、2つのベクトルを一緒に加算するための加算器デバイス 1708 と、(活性化回路ブロック 1702 を備える) \tanh デバイス 1505 と、シグモイド関数ブロック 1702 から $i(t)$ が出力されたときの値 $i(t)$ を格納する、レジスタ 1707 と、上記値がマルチプレクサ 1710 を介して乗算器デバイス 1703 から出力されるとききの値 $f(t) * c(t-1)$ を格納する、レジスタ 1704 と、上記値がマルチプレクサ 1710 を介して乗算器デバイス 1703 から出力されるとききの値 $i(t) * u(t)$ を格納するレジスタ 17

40

50

05と、上記値がマルチプレクサ1710とマルチプレクサ1709を介して乗算器デバイス1703から出力されるときにの値 $o(t) * c \sim (t)$ を格納する、レジスタ1706とを備える。

【0103】

LSTMセル1600が複数のVMMアレイ1601及びそれぞれの活性化関数ブロック1602の複数の組を含むのに対し、LSTMセル1700は、LSTMセル1700の実施形態において複数の層を表すために使用されるVMMアレイ1701及び活性化関数ブロック1702の1つの組のみを含む。LSTMセル1700は、LSTMセル1600と比較して、VMM及び活性化関数ブロックのために1/4のスペースを必要とするので、LSTMセル1700は、LSTM1600より必要なスペースが少ない。

10

【0104】

LSTMユニットは、典型的には、複数のVMMアレイを備え、これらは各々、加算器及び活性化回路ブロック及び高電圧生成ブロックなどの、VMMアレイの外側の特定の回路ブロックによって提供される機能を必要とすることが更に理解され得る。各VMMアレイに別個の回路ブロックを提供することは、半導体デバイスにかなりの量のスペースを必要とし、幾分非効率的であろう。したがって、以下に記載される実施形態は、VMMアレイ自体の外側に必要とされる回路を最小化することを試みる。

ゲート付き回帰型ユニット

【0105】

アナログVMM実装形態をゲート付き回帰型ユニット(gated recurrent unit、GRU)システムに利用することができる。GRUは、回帰型ニューラルネットワークのゲーティング機構である。GRUは、GRUセルが一般にLSTMセルより少ない構成要素を含むことを除いて、LSTMに類似している。

20

【0106】

図18は、例示的なGRU1800を示す。この例におけるGRU1800は、セル1801、1802、1803及び1804を含む。セル1801は入力ベクトル x_0 を受け取り、出力ベクトル h_0 を生成する。セル1802は、入力ベクトル x_1 と、セル1801から出力ベクトル h_0 とを受け取り、出力ベクトル h_1 を生成する。セル1803は、入力ベクトル x_2 と、セル1802から出力ベクトル(隠れた状態) h_1 とを受け取り、出力ベクトル h_2 を生成する。セル1804は、入力ベクトル x_3 と、セル1803から出力ベクトル(隠れた状態) h_2 とを受け取り、出力ベクトル h_3 を生成する。追加のセルを使用することができ、4つのセルを有するGRUは単なる例である。

30

【0107】

図19は、図18のセル1801、1802、1803及び1804に使用することができる、GRUセル1900の例示的な実装形態を示す。GRUセル1900は、先行するGRUセルから入力ベクトル $x(t)$ 及び出力ベクトル $h(t-1)$ を受け取り、出力ベクトル $h(t)$ を生成する。GRUセル1900は、シグモイド関数デバイス1901及び1902を備え、各々が、出力ベクトル $h(t-1)$ 及び入力ベクトル $x(t)$ からの構成要素に0~1の数を適用する。GRUセル1900はまた、入力ベクトルに双曲線 \tanh 関数を適用するための \tanh デバイス1903と、2つのベクトルを一緒に乗算するための乗算器デバイス1904、1905及び1906と、2つのベクトルを一緒に加算するための加算器デバイス1907と、1からの入力を減算して出力を生成する補完デバイス1908とを備える。

40

【0108】

図20は、GRUセル1900の一実装形態であるGRUセル2000を示す。読者の便宜のために、GRUセル1900からの同じ採番方法が、GRUセル2000で使用される。図20から分かり得るように、シグモイド関数デバイス1901及び1902、並びに \tanh デバイス1903は各々、複数のVMMアレイ2001及び活性化関数ブロック2002を備える。したがって、VMMアレイは、特定のニューラルネットワークシステムで使用されるGRUセルにおいて特に使用されることが分かり得る。

50

【 0 1 0 9 】

GRUセル2000の代替例（及びGRUセル1900の実装形態の別の例）を図21に示す。図21において、GRUセル2100は、VMMアレイ2101及び活性化関数ブロック2102を使用しており、シグモイド関数として構成された場合には、入力ベクトルの各構成要素のうちどれだけの量が出力ベクトルを通して許容されるかを制御するために、0～1の数を適用する。図21では、シグモイド関数デバイス1901及び1902、並びにtanhデバイス1903は、同じ物理ハードウェア（VMMアレイ2101及び活性化関数ブロック2102）を、時分割多重化された方式で共有する。GRUセル2100はまた、2つのベクトルと一緒に乗算するための乗算器デバイス2103と、2つのベクトルと一緒に加算するための加算器デバイス2105と、1からの入力を減算して出力を生成するための補完デバイス2109と、マルチプレクサ2104と、上記値がマルチプレクサ2104を介して乗算器デバイス2103から出力されるとき値 $h(t-1) * r(t)$ を保持するレジスタ2106と、上記値がマルチプレクサ2104を介して乗算器デバイス2103から出力されるとき値 $h(t-1) * z(t)$ を保持するレジスタ2107と、上記値がマルチプレクサ2104を介して乗算器デバイス2103から出力されるとき値 $h(t) * (1 - z(t))$ を保持するレジスタ2108と、を備える。

10

【 0 1 1 0 】

GRUセル2000が複数のVMMアレイ2001及び活性化関数ブロック2002の複数の組を含むのに対し、GRUセル2100は、GRUセル2100の実施形態において複数の層を表すために使用されるVMMアレイ2101及び活性化関数ブロック2102の1つの組のみを含む。GRUセル2100は、GRUセル2000と比較して、VMM及び活性化関数ブロックのために1/3のスペースを必要とするので、GRUセル2100は、GRUセル2000より必要なスペースが少ないであろう。

20

【 0 1 1 1 】

GRUシステムは、典型的には、複数のVMMアレイを備え、これらは各々、加算器及び活性化回路ブロック及び高電圧生成ブロックなどの、VMMアレイの外側の特定の回路ブロックによって提供される機能を必要とすることが更に理解され得る。各VMMアレイに別個の回路ブロックを提供することは、半導体デバイスにかなりの量のスペースを必要とし、幾分非効率的であろう。したがって、以下に記載される実施形態は、VMMアレイ自体の外側に必要とされる回路を最小化することを試みる。

30

【 0 1 1 2 】

VMMアレイへの入力、アナログレベル、バイナリレベル、又はデジタルビット（この場合、デジタルビットを適切な入力アナログレベルに変換するためにDACが必要とされる）であり、出力はアナログレベル、バイナリレベル、又はデジタルビットであり得る（この場合、出力アナログレベルをデジタルビットに変換するために出力ADCが必要とされる）。

【 0 1 1 3 】

VMMアレイの各メモリセルに関して、各重みwは、単一のメモリセルによって、又は差動セルによって、又は2つのブレンドメモリセル（2つのセルの平均）によって実装することができる。差動セルの場合では、重みwを差動重み（ $w = w_+ - w_-$ ）として実装するために、2つのメモリセルが必要とされる。2つのブレンドメモリセルの場合、2つのセルの平均として重みwを実装するために2つのメモリセルが必要とされる。

40

デバイスの動作温度及び電流電圧特性

【 0 1 1 4 】

図22は、例示的なトランジスタ又はメモリセルの3つの例示的な電流電圧特性曲線を示すグラフを示す。図22に示される現象は、フラッシュメモリシステムにおける典型的な基準トランジスタ、基準メモリセル、及び選択されたメモリセルの挙動を示すことが理解され得る。各曲線は、ゲートとソースとの間の電圧Vgsが変化するにつれて、デバイスを通る電流Idsがどのように変化するかを示す。各曲線は、特定の動作温度における

50

デバイスの電流電圧特性を表す。したがって、デバイスの動作温度が変化するにつれて、同一の V_{gs} 値の対応する I_{ds} 値は変化すると考えられ得る。温度が変化するにつれてデバイス特性がこのように変動することは問題である。これは、アナログニューロモフィックメモリシステムが、通常のフラッシュメモリシステムと比較して、より高い精度及び正確度を必要とするためである。

【0115】

本明細書に記載される実施形態は、デバイスの電流電圧特性曲線に対する傾き補償（不連続的又は連続的のいずれか）、絶対スケーリング（正規化）、及び漏れ低減を提供することによって、デバイスの動作温度の変化につれてのデバイスの挙動の変化を補償する。

【0116】

図23は、温度の変化を補償するためにデバイスの電流電圧特性に適用される傾き補償の一例を示す。同一デバイスが、異なる温度において異なるサブスレッショルド電流電圧特性曲線を有するであろうことが理解され得る。

【0117】

より一般的には、温度の関数として変化する変数 a を導入することにより、以下の式に基づいて、傾き補償をデバイスに適用できることが理解され得る。

$I_{ds} = I_{ds0} \cdot \exp(a \cdot V_{gs1} - V_t) / (k \cdot V_t)$ 、式中、 V_t は熱電圧である。

【0118】

図24A及び図24Bは、不連続的な温度補償を示す。図24Aは、温度の関数として a の例示的な値を示す。温度が上昇するにつれて、 a は不連続的な段階で増加することが分かり得る。あるいは、線形領域又は飽和領域で作動するメモリセルなどについては、等価スケーリング定数が、温度に対して慎重な段階で減少することになる。図24Bは、例示的な温度補償ブロック2400を示す。温度センサ2401は、デバイスの温度を測定する。次いで、温度センサ2401の出力がコントローラ2402（個別論理又はソフトウェアを実行するマイクロコントローラであり得る）に提供され、コントローラ2402は、制御ビット2403a、...、2403iを生成する。範囲 $a \sim i$ は、 k の例示的な個別レベルの数を表す。より広い範囲又はより小さい範囲を使用できることが理解されるであろう。制御ビット2403の各々は、当該制御ビットに割り当てられた特定の温度範囲の間にアサートされる。したがって、温度が上昇するにつれて、各個別レベルの温度の達すると、異なる制御ビット2403がアサートされる。次いで、制御ビット2403は、スイッチ2405に（同じく、2405a、...、2405iの番号）適用される。各スイッチ2405は、閉じたときに、電流源2404（同じく、2404a、...、2404iの番号）をレジスタ2406の一端に適用する。ここで、各電流源2404は、前述の温度範囲でスイッチによって起動された電流源2404とは異なる量の電流を生成する。その結果、当該ノードにおける電圧 V_{tc} は、温度が上昇するにつれて変化する。 V_{tc} は、トランジスタのゲート、メモリセル210として図2に示される種類のメモリセルのワード線、メモリセル310として図3に示される種類のメモリセルの制御ゲート、メモリセル410として図4に示される種類のメモリセルのソースゲート、又はメモリセル510として図5に示される種類のメモリセルの制御ゲートに印加され得る、温度補償バイアス電圧である。このように V_{tc} を加えることにより、各ゲート電圧値が本質的に量 V_{tc} だけシフトされるため、デバイスの電圧特性曲線は修正されることになる。

【0119】

図25は、温度補償ブロック2500を示す。温度補償ブロック2500は、演算増幅器2501及び2505と、調整可能なレジスタブロック2506と、温度センサ2401と、コントローラ2502と、を備える。調整可能なレジスタブロック2506は、スイッチ2508a、...、2508nのうちの1つにそれぞれ結合されたレジスタ2507a、...、2507nを備える。

【0120】

図24のように、温度センサ2401は、デバイスの温度を測定する。次いで、温度セ

10

20

30

40

50

ンサ 2401 の出力がコントローラ 2502 (個別論理又はソフトウェアを実行するマイクロコントローラであり得る) に提供され、コントローラ 2502 は、制御ビット 2509a、. . .、2509n を生成する。範囲 a ~ n は、a の例示的な個別レベルの数を表す。より広い範囲又はより小さい範囲を使用できることが理解されるであろう。温度センサ 2301 によって検知された温度に応じて、制御ビット 2509a、. . .、2509n の間で異なる制御ビットがアサートされる。次いで、制御ビット 2509 は、スイッチ 2508 に (同じく、2508a、. . .、2508n の番号) 適用される。各スイッチ 2508 は、閉じたときに、それぞれのレジスタ 2507 (同じく、2507a、. . .、2507n の番号) をフィードバックレジスタとして演算増幅器 2505 に適用する。その結果、演算増幅器 2505 の出力電圧 V_{gs_tc} は、温度が上昇するにつれて増加する。

10

【0121】

V_{gs_tc} は、トランジスタのゲート、メモリセル 210 / 310 / 410 として図 2 / 3 / 4 に示される種類のメモリセルのワード線、メモリセル 310 / 510 として図 3 / 5 に示される種類のメモリセルの制御ゲート、メモリセル 210 / 310 / 410 / 510 として図 2 / 3 / 4 / 5 に示される種類のメモリセルのソースゲート、又はメモリセル 310 / 510 として図 3 / 5 に示される種類のメモリセルの制御ゲートに印加され得る、温度補償バイアス電圧である。このように V_{gs_tc} を修正することにより、各 V_{gs} 値が本質的に V_{gs_tc} の量だけシフトされるため、デバイスの電圧特性曲線は修正されることになる。

20

【0122】

図 26A 及び図 26B は、正の温度補償電流の場合は補償電流 i_{tcpn} 、 i_{tcp} を、負の温度補償電流の場合は補償電流 i_{tcn} を加えることによって、デバイスの電流電圧特性曲線を調整するための実施形態を示す。図 26A は、各種類のデバイスの温度変化につれての補完デバイスの 3 つ (4 つ以上であってもよい) の例示的な曲線を示す。補償電流 i_{tcpn} を注入することにより、デバイスの電流電圧特性曲線の傾きは修正され得る。

【0123】

図 26B は、温度補償ブロック 2600 の一実施形態を示す。温度補償ブロック 2600 は、デバイスの温度を測定する温度センサ 2401 と、コントローラ 2605 (論理又はソフトウェアを実行しているマイクロコントローラであり得る) とを備える。コントローラ 2605 は、1 つ以上の制御信号 2606 を生成する。

30

【0124】

温度補償ブロック 2600 は、補償電流 i_{tcpn} を提供する電流源 2603 によって表される電流を一緒に生成する、調整可能な電流源 2601 I_{tcp} 及び 2602 I_{tcn} を更に備える。調整可能な電流源 2601 及び 2602 は、制御信号 2606 によって調整される。調整可能な電流源 2601 及び 2602 の一方又は両方は、制御信号 2606 によってデバイスの温度が変化するにつれて、提供される補償電流 i_{tcpn} の量を変化させるように調整され得る。電流 i_{tcpn} はレジスタ 2604 に印加され、電流源 2603 とレジスタ 2604 との間のノードにおいて電圧 V_{tc} を生じさせる。

40

【0125】

V_{tc} は、トランジスタのゲート、メモリセル 210 / 310 / 410 / 510 として図 2 / 3 / 4 / 5 に示される種類のメモリセルのワード線、メモリセル 310 / 510 として図 3 / 5 に示される種類のメモリセルの制御ゲート、メモリセル 210 / 310 / 410 / 510 として図 2 / 3 / 4 / 5 に示される種類のメモリセルのソースゲート、又はメモリセル 310 / 410 として図 3 / 4 に示される種類のメモリセルの消去ゲートに印加され得る、温度補償バイアス電圧である。このように V_{tc} を加えることにより、各 V_{gs} 値が本質的に量 V_{tc} だけシフトされるため、デバイスの電圧特性曲線は修正されることになる。

【0126】

50

図 27A 及び図 27B は、電流 i_{tcpn} によって生成されたバイアス電圧をデバイスの制御ゲートで加えることによって、デバイスの電流電圧特性曲線を調整するための一実施形態を示す。図 27A は、デバイスの動作温度が上昇するにつれて特定の動作を実行するために制御ゲートに印加される必要がある、必須の制御ゲート電圧 V_{cg} の 3 つの例示的な曲線を示す。

【0127】

図 27B は、温度補償ブロック 2700 の一実施形態を示す。温度補償ブロック 2700 は、デバイスの温度を測定する温度センサ 2401 と、コントローラ 2705（論理又はソフトウェアを実行しているマイクロコントローラであり得る）とを備える。コントローラ 2705 は、1 つ以上の制御信号 2706 を生成する。

10

【0128】

温度補償ブロック 2700 は、補償電流 i_{pcn} を生成する、調整可能な電流源 2701 と、調整可能なレジスタ 2702 とを更に備える。調整可能な電流源 2701 及び調整可能なレジスタ 2702 は、制御信号 2706 によって調整される。電圧 V_{tc} は、調整可能な電流源 2701 と調整可能なレジスタ 2702 との間のノードにおいて生成される。

【0129】

V_{tc} は、メモリセル 310 として図 3 に示される種類のメモリセルの制御ゲート、又はメモリセル 510 として図 5 に示される種類のメモリセルの制御ゲートに印加され得る温度補償バイアス電圧である。このように V_{tc} を加えることにより、各 V_{gs} 値が本質的に量 V_{tc} だけシフトされるため、デバイスの電圧特性曲線は修正されることになる。

20

【0130】

図 28A 及び図 28B は、電流 i_{tcpn} によって生成されたバイアス電圧をデバイスの消去ゲートで加えることによって、デバイスの電流電圧特性曲線を調整するための一実施形態を示す。図 28A は、デバイスの動作温度が上昇するにつれて特定の動作を実行するために消去ゲートに印加される必要がある、必須の制御ゲート電圧 V_{eg} の 3 つの例示的な曲線を示す。

【0131】

図 28B は、多数の特性曲線を示しており、各曲線は、アレイ読み取り（ベクトルマトリックス乗算器）動作を実行するために V_{eg} の各値に必要な V_{eg} を示し、各曲線は、特定の温度又は温度範囲におけるデバイスの特性を表す。

30

【0132】

図 28C は、温度補償ブロック 2800 の一実施形態を示す。温度補償ブロック 2800 は、デバイスの温度を測定する温度センサ 2401 と、コントローラ 2805（論理又はソフトウェアを実行しているマイクロコントローラであり得る）とを備える。コントローラ 2805 は、1 つ以上の制御信号 2806 を生成する。

【0133】

温度補償ブロック 2800 は、補償電流 i_{tcpn} を生成する、調整可能な電流源 2801 と、調整可能なレジスタ 2802 とを備える。調整可能な電流源 2801 及び調整可能な電流レジスタ 2802 は、デバイスの温度が変化するにつれて、制御信号 2806 によって調整される。電圧 V_{tc} は、電流源 2803 とレジスタ 2804 との間のノードにおいて生成される。

40

【0134】

V_{tc} は、メモリセル 310 として図 3 に示される種類のメモリセルの消去ゲート、又はメモリセル 410 として図 4 に示される種類のメモリセルの消去ゲートに印加され得る温度補償バイアス電圧である。このように V_{tc} を加えることにより、各 V_{gs} 値が本質的に量 V_{tc} だけシフトされるため、デバイスの電圧特性曲線は修正されることになる。

【0135】

図 29A 及び図 29B は、電流 i_{tcpn} によって生成されたバイアス電圧をデバイスのソース線端子で加えることによって、デバイスの電流電圧特性曲線を調整するための一実施形態を示す。図 29A は、温度が上昇するにつれて特定の動作を実行するために必要

50

とされる、必須のソース線端子電圧の3つの例示的な曲線を示す。

【0136】

図29Bは、温度補償ブロック2900の一実施形態を示す。温度補償ブロック2900は、デバイスの温度を測定する温度センサ2401と、コントローラ2905（論理又はソフトウェアを実行しているマイクロコントローラであり得る）とを備える。コントローラ2905は、1つ以上の制御信号2906を生成する。

【0137】

温度補償ブロック2900は、補償電流 i_{tcpn} を生成する、調整可能な電流源2901と、調整可能なレジスタ2902とを更に備える。調整可能な電流源2901及び調整可能なレジスタ2902は、デバイスの温度が変化するにつれて、デバイスの制御ゲートに印加される負バイアスの量を変化させるように制御信号2906によって調整され得る。電圧 V_{tc} は、電流源2901とレジスタ2902との間のノードにおいて生成される。

10

【0138】

V_{tc} は、メモリセル210として図2に示される種類のメモリセルのソース線端子、メモリセル310として図3に示される種類のメモリセルのソース線端子、メモリセル410として図4に示される種類のメモリセルのソース線端子、又はメモリセル410として図4に示される種類のメモリセルの消去ゲート、又はメモリセル510として図5に示される種類のメモリセルのソース線端子に印加され得る温度補償バイアス電圧である。このように V_{tc} を加えることにより、各 V_{gs} 値が本質的に量 V_{tc} だけシフトされるため、デバイスの電圧特性曲線は修正されることになる。

20

【0139】

電流電圧特性曲線の差は、異なるデバイス及び異なる構成の類似のデバイスに対して存在することが更に理解され得る。

【0140】

図30は、特定の構成を有する、例示的な基準トランジスタ3000を示す。

【0141】

図31は、別の構成（ワード線がビット線に結合されている）を有する、例示的な基準メモリセル3100を示す。

【0142】

図32は、別の構成（浮遊ゲートFGがビット線に結合されている）を有する、別の例示的な基準メモリセル3200を示す。これらのデバイスの各々は、異なる電流電圧特性曲線を有し得ることが理解され得る。

30

【0143】

本明細書に記載される実施形態は、基準トランジスタ、基準メモリセル、及び/又は選択されたメモリセルの電流電圧特性曲線の傾きにおける差を補償する。

【0144】

異なるサブスレッショルド電流電圧特性曲線を有する2つのデバイスを有するシステムでは、第1のデバイスを通るドレインソース電流は、

$$I_{ds1} = I_{ds0} \cdot \exp(V_{gs1} - V_t) / k_1 \cdot U_t \text{ である。}$$

40

【0145】

第2のデバイスを通るドレインソース電流は、

$$I_{ds2} = I_{ds0} \cdot \exp(V_{gs2} - V_t) / k_2 \cdot U_t \text{ である。}$$

【0146】

各例では、傾きは $\sim 1/k$ に比例することが分かり得る。

【0147】

以下の実施形態のうちいくつかにおいて、傾き正規化は、

$V_{gs1} = k \cdot V_{gs2}$ の第1のデバイスのゲートソース電圧を使用することによって実装される。

これは、 I_{ds1} 及び I_{ds} が、傾き正規化後に同じ傾きを有することを意味する。

50

【0148】

これは、図33に図式的に示され、 $V_{gs1} = a * V_{gs2}$ の電圧がデバイス1に印加され、これにより、第1のデバイスの電流電圧特性曲線の傾きを第2のデバイスの電流電圧特性曲線の傾きに近似させる。

【0149】

ここで、異なるデバイス間で傾き正規化を実行するための実施形態について説明する。

【0150】

図34は、基準トランジスタ3401と、選択されたメモリセル3402と、ゲートドライバ3403と、絶対正規化回路3404とを備える傾き正規化システム3400を示す。ゲートドライバ3403は、入力電圧 V_{gs} を受信し、その入力電圧に a を乗算して、基準トランジスタ3401のゲートに印加される出力電圧 V_{gsint} を生成する。絶対正規化回路3404は、トリミング可能な電流ミラー（基準トランジスタ3401からの電流とメモリセル3402からの出力電流との間の比を調整するための電流ミラー回路）であり得、トリミングプロセスは、基準トランジスタ若しくはアレイトランジスタによって、又は $I-V$ 傾きの不整合によって引き起こされる不一致を調整することができる。選択されたメモリセル3402は、メモリセルのアレイのメモリセルのうちの一つである。

10

【0151】

図35は、基準トランジスタ3501と、選択されたメモリセル3502と、ドライバ3503とを備える傾き正規化システム3500を示す。ドライバ3503は、電圧 V_{gs} を受信し、それに a を乗算し、 $V_{gs'}$ の出力電圧をもたらす。このため、基準トランジスタ3501及び選択されたメモリセル3502は、異なる電圧を受信することになり、差は、傾きの差に相当する。

20

【0152】

図38は、動作温度の変化によるデバイスの電流電圧特性曲線3800の変化を補償するためにルックアップテーブル3801を利用する一実施形態を示す。これらのデバイスは、トランジスタ、基準メモリセル、及び選択されたメモリセルを含み得る。この実施形態では、曲線3800のデータは、ルックアップテーブル3801に格納され、動作中、 k は、数式を介するのではなく、ルックアップテーブル3801から判定される。ルックアップテーブル3801は、様々な動作温度における各入力電圧に対して異なる所望の出力電流を含むことができる。このデータは、製造プロセス又は試験プロセス中にルックアップテーブル3801に追加され得る。

30

【0153】

実施形態はまた、漏れ低減を提供することができる。具体的には、トランジスタ又はメモリセルの漏れは、制御ゲート端子に負電圧（ $-0.1V \sim -0.4V$ の電圧など）を、ワード線端子に負電圧（ $-0.1V \sim -0.4V$ の電圧など）を、消去ゲート端子に負電圧（ $-0.1V \sim -0.4V$ の電圧など）を、より高い温度においてビット線により小さい電圧を、より高い温度においてバルク端子（基板）により小さい電圧を、又はソース線端子にバイアス電圧を印加することによって低減され得る。

【0154】

図36は、デバイスの制御ゲート端子への電圧の印加による漏れ電流の低減を示す図を提供する。この実施形態では、制御ゲートバイアス電圧は、 $I-1k = n * I-min / 列数$ となるように選択される。アレイが $256行 \times 512列$ を含む場合、 $256 * I-1kg < I-min$ の例となるだろう。例えば、 $I-min$ は約 $500pA$ であってよく、 $I-1kg$ は約 $2pA$ であってよい。これは、制御ゲートバイアス電圧が約 $2V$ 超であろうことを意味する。

40

【0155】

温度補償のための別の実施形態では、逆の基板電圧（例えば、負電圧）が印加されるか、又はソース線電圧がバイアスされて、浮遊ゲートスレッショルド電圧 V_{tFG} を上昇させることによってサブスレッショルド動作電圧範囲を拡大する。

【0156】

50

図37は、線形領域で動作するメモリセルに好適なアレイアーキテクチャを示す。システム3700は、入力ブロック3701と、出力ブロック3702と、メモリセルのアレイ3703とを備える。入力ブロック3701は、アレイ3703のメモリセルのドレイン（ソース線）に結合され、出力ブロック3702は、アレイ3703のメモリセルのビット線に結合される。あるいは、入力ブロック3701は、アレイ3703のメモリセルのワード線に結合され、出力ブロック3702は、アレイ3703のメモリセルのビット線に結合される。

【0157】

システム3700がLSTM又はGRUを実装するために使用される場合、出力ブロック3702及び/又は入力ブロック3701は、LSTM/GRUアーキテクチャに必要とされるように、乗算器ブロック、加算器ブロック、減算器（出力 = 1 - 入力）ブロックを含み得、任意選択的に、アナログサンプルアンドホールド回路又はデジタルサンプルアンドホールド回路（例えば、レジスタ又はSRAM）を必要に応じて含むことができる。

【0158】

本明細書で使用される場合、「の上方に」及び「に」という用語は両方とも、「に直接」（中間材料、要素、又は空間がそれらの間に何ら配設されない）、及び「に間接的に」（中間材料、要素、又は空間がそれらの間に配設される）を包括的に含むことに留意すべきである。同様に、「隣接した」という用語は、「直接隣接した」（中間材料、要素、又は空間がそれらの間に何ら配設されない）、及び「間接的に隣接した」（中間材料、要素、又は空間がそれらの間に配設される）を含み、「に取り付けられた」は、「に直接取り付けられた」（中間材料、要素、又は空間がそれらの間に何ら配設されない）、及び「に間接的に取り付けられた」（中間材料、要素、又は空間がそれらの間に配設される）を含み、「電氣的に結合された」は、「に直接電氣的に結合された」（要素と一緒に電氣的に接続する中間材料又は要素がそれらの間にない）、及び「間接的に電氣的に結合された」（要素と一緒に電氣的に接続する中間材料又は要素がそれらの間にある）を含む。例えば、要素を「基板の上方に」形成することは、その要素を基板に直接、中間材料/要素をそれらの間に何ら伴わずに、形成すること、並びにその要素を基板に間接的に、1つ以上の中間材料/要素をそれらの間に伴って、形成することを含み得る。

10

20

30

40

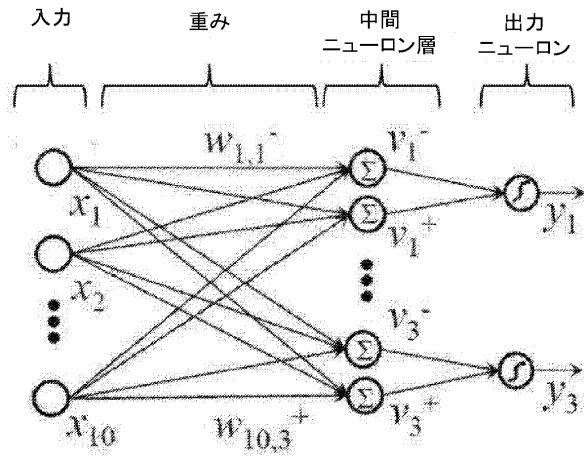
50

【図面】

【図 1】

(先行技術)

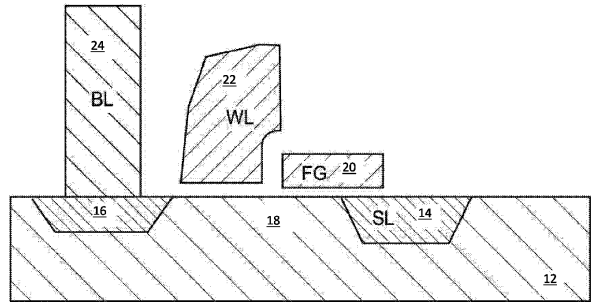
100



【図 2】

(先行技術)

メモリセル
210

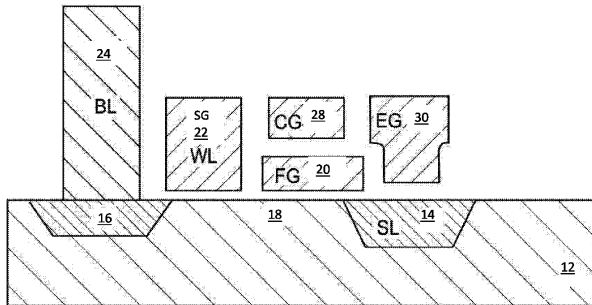


10

【図 3】

(先行技術)

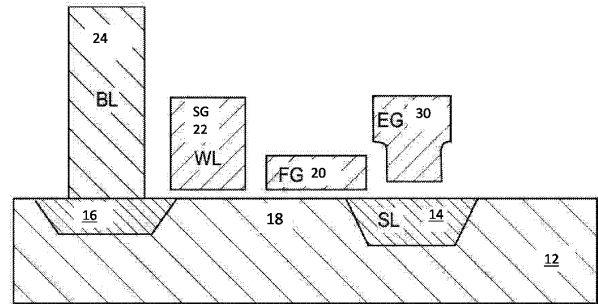
メモリセル
310



【図 4】

(先行技術)

メモリセル
410



20

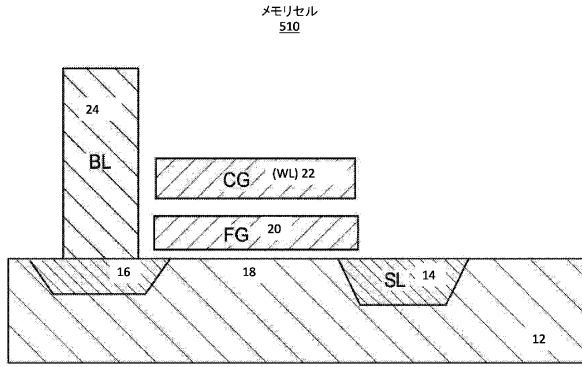
30

40

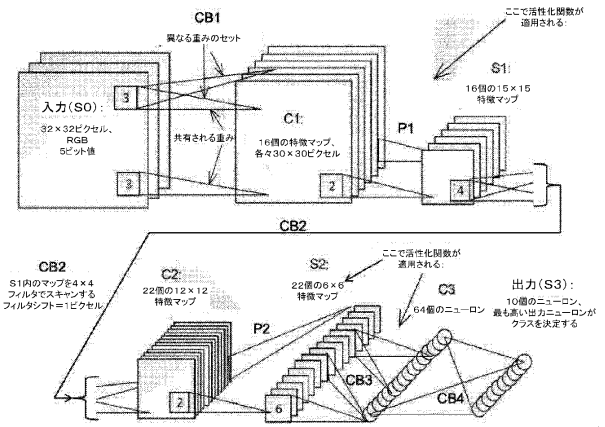
50

【図5】

(先行技術)

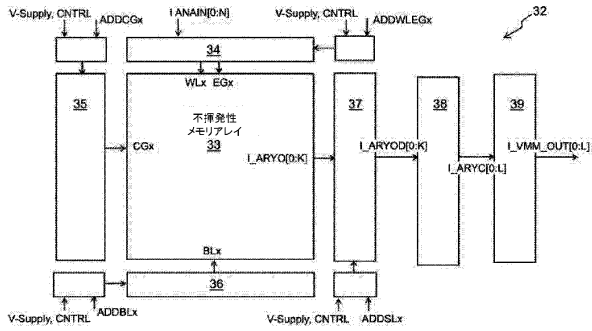


【図6】



10

【図7】



【図8】

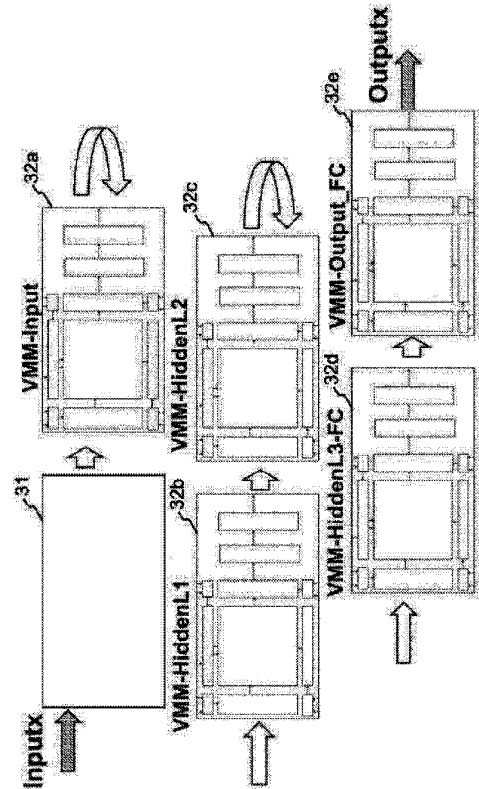


FIGURE 8

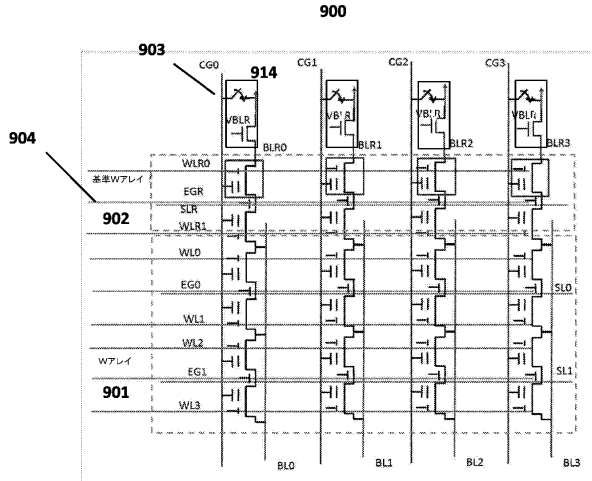
20

30

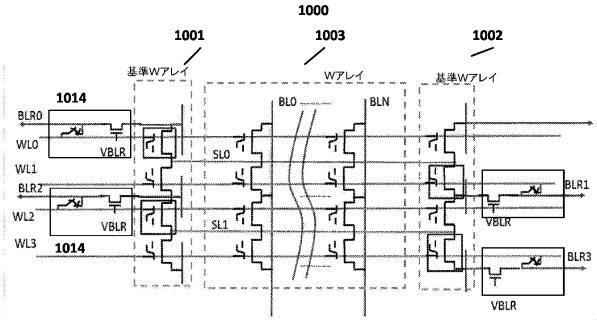
40

50

【図 9】

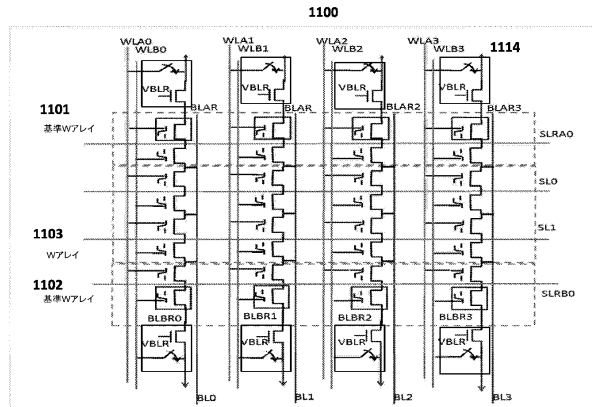


【図 10】

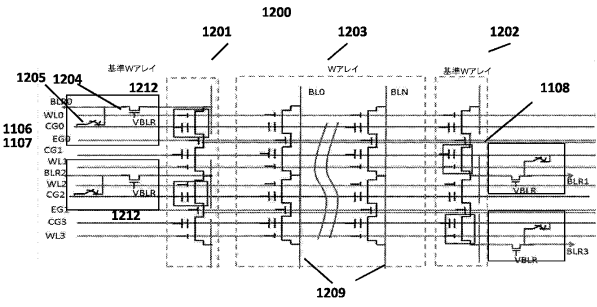


10

【図 11】



【図 12】



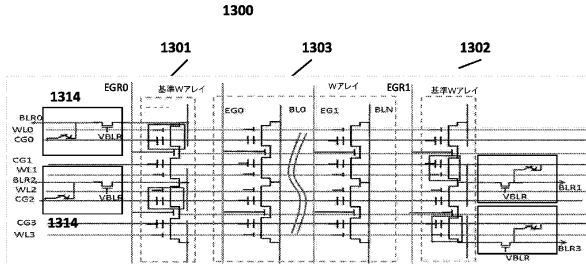
20

30

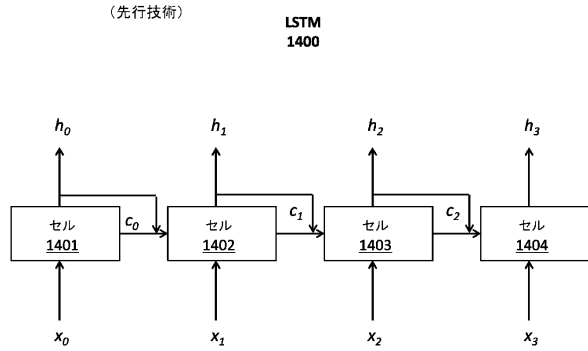
40

50

【図13】

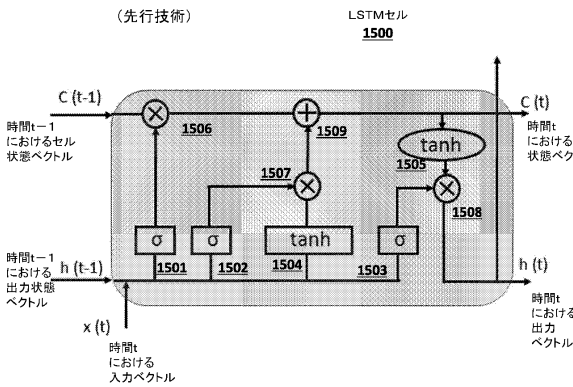


【図14】

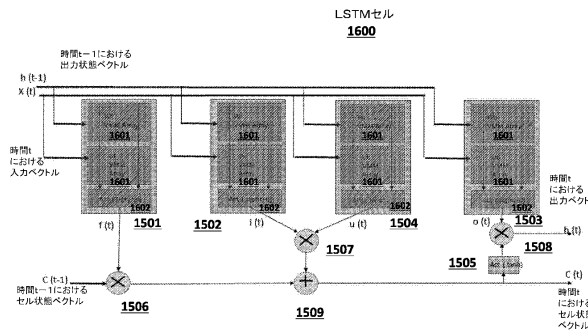


10

【図15】

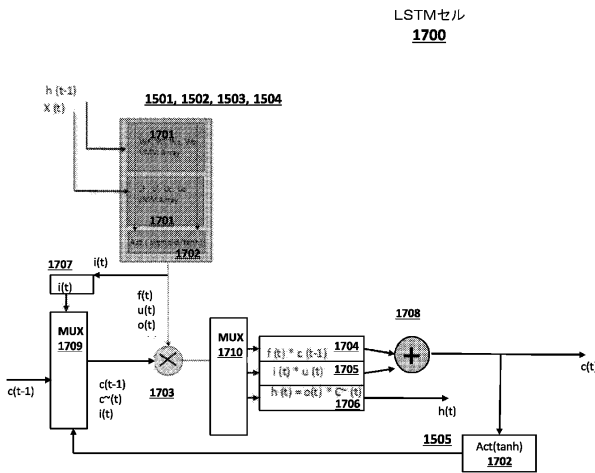


【図16】

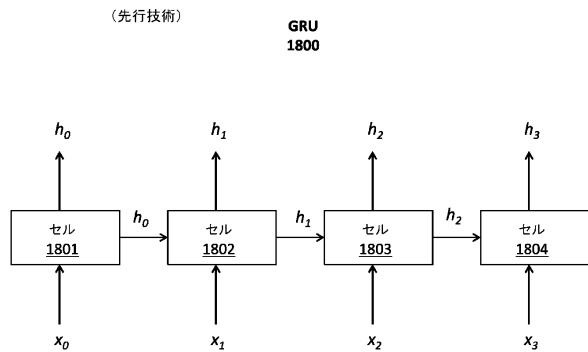


20

【図17】



【図18】

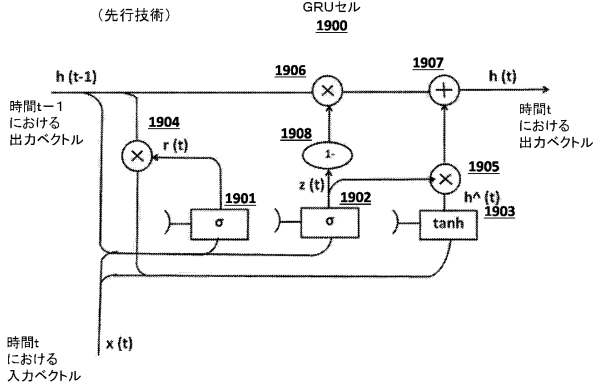


30

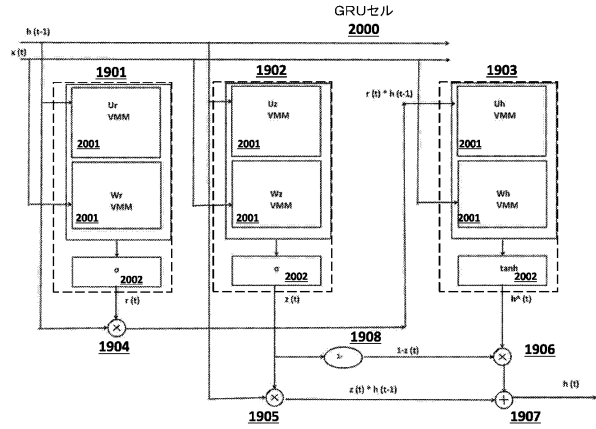
40

50

【図 19】

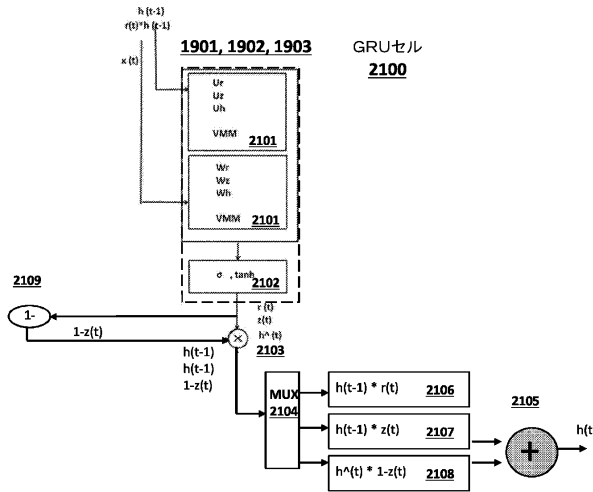


【図 20】

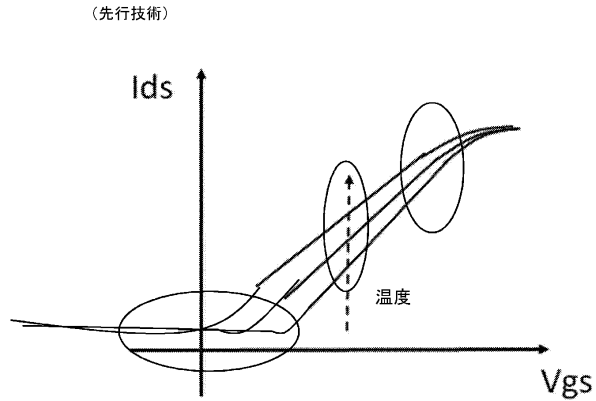


10

【図 21】



【図 22】



20

30

40

50

【図 2 3】

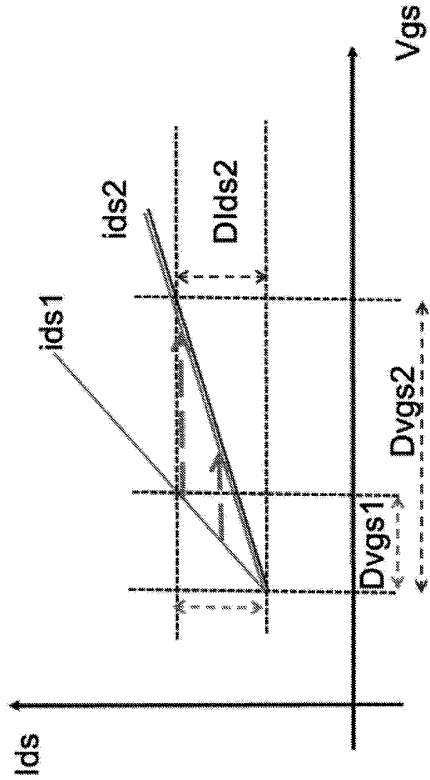


FIGURE 23

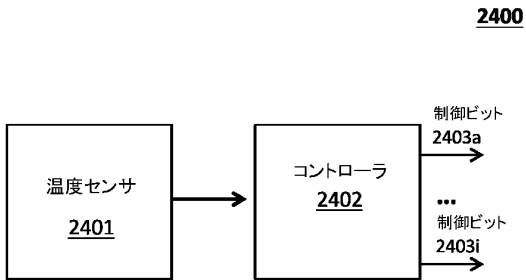
【図 2 4 A】



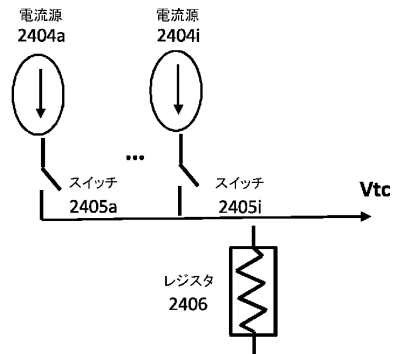
10

20

【図 2 4 B】



【図 2 4 C】

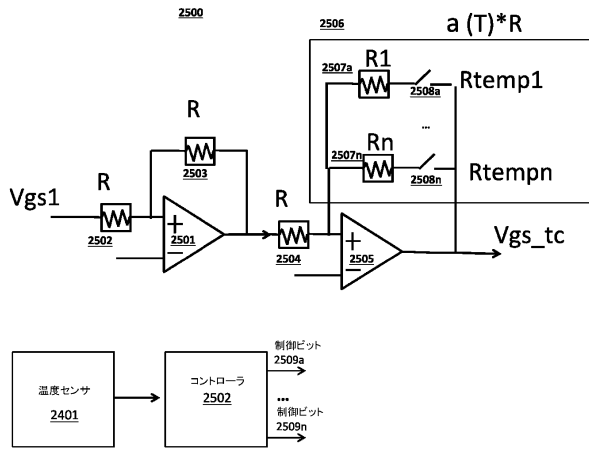


30

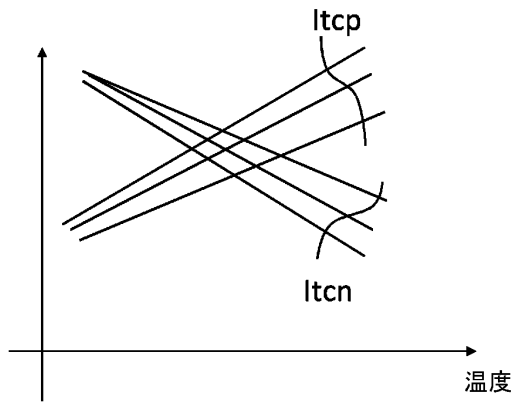
40

50

【図 2 5】

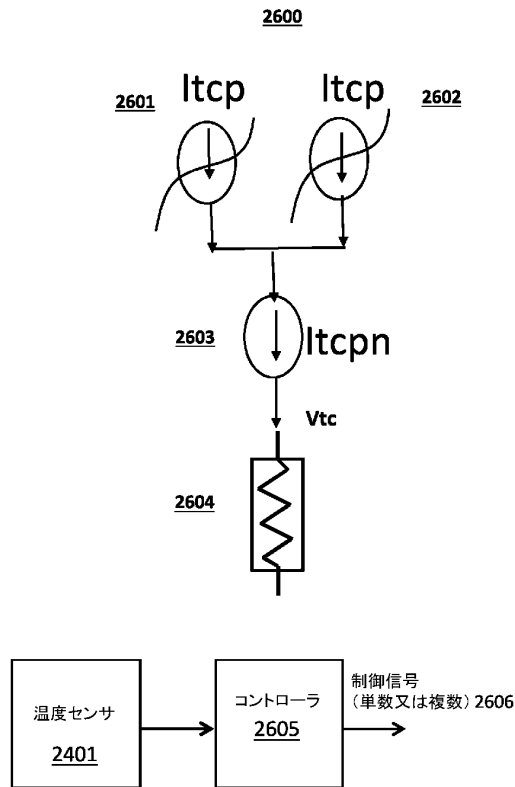


【図 2 6 A】

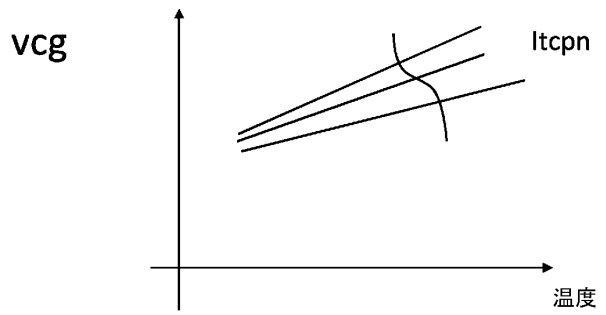


10

【図 2 6 B】



【図 2 7 A】



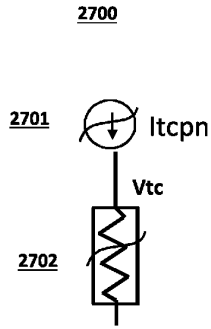
20

30

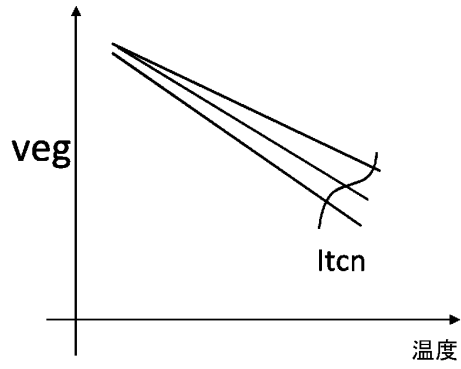
40

50

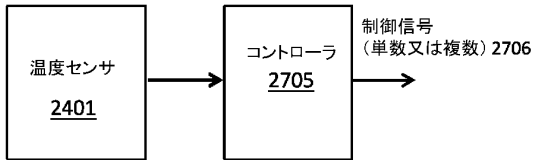
【図 27B】



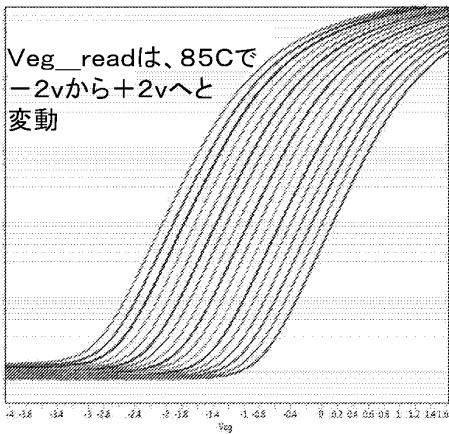
【図 28A】



10

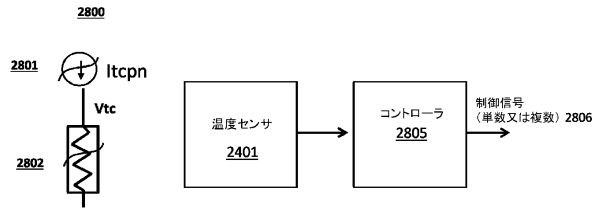


【図 28B】



20

【図 28C】

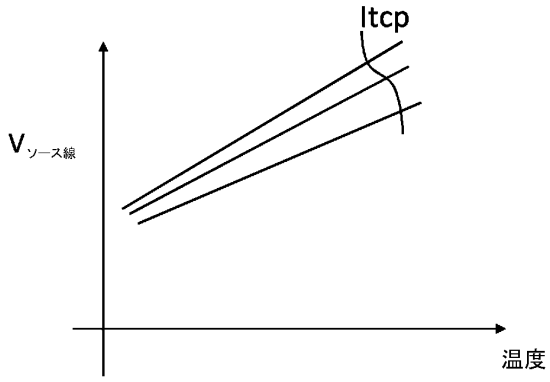


30

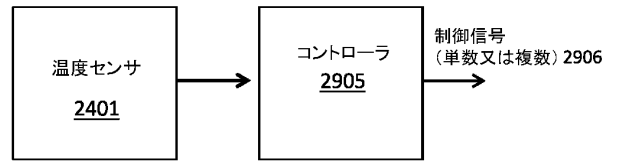
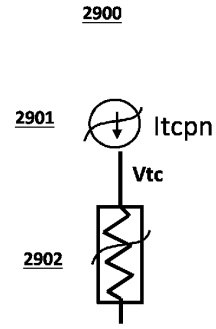
40

50

【図 29 A】



【図 29 B】



10

【図 30】

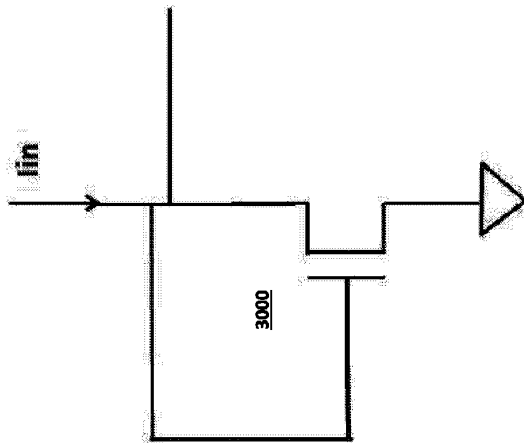


FIGURE 30

【図 31】

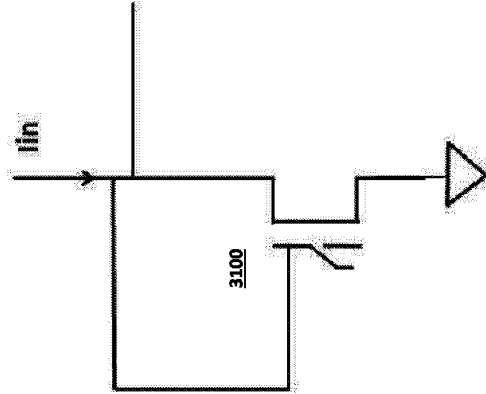


FIGURE 31

20

30

40

50

【 3 2 】

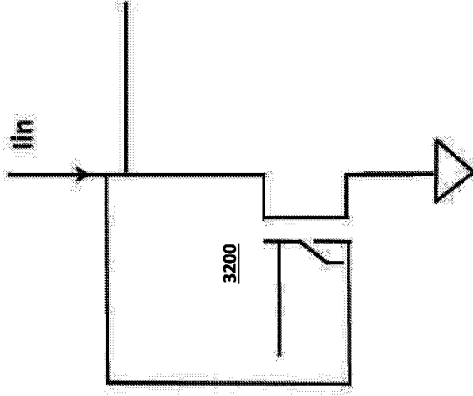


FIGURE 32

【 3 3 】

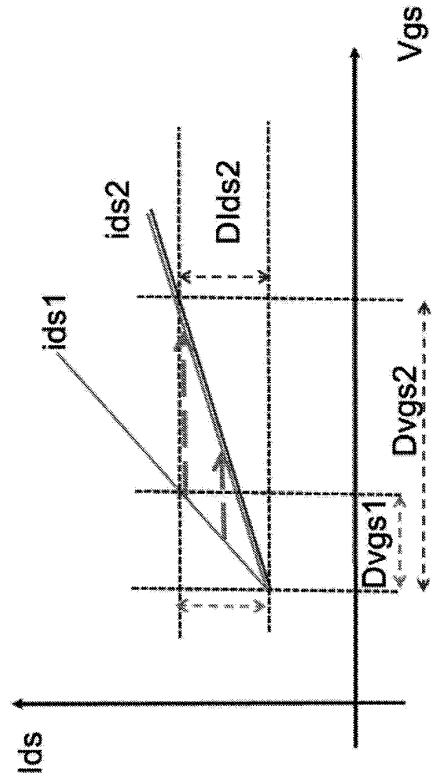
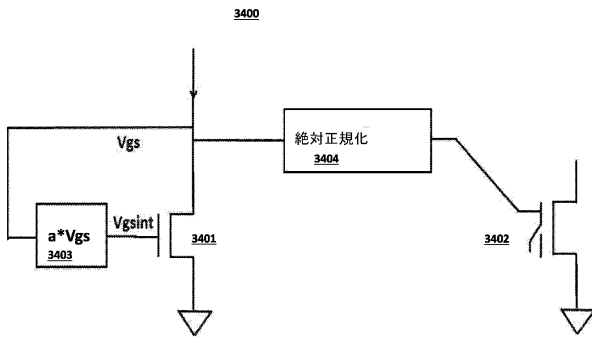


FIGURE 33

【 3 4 】



【 3 5 】

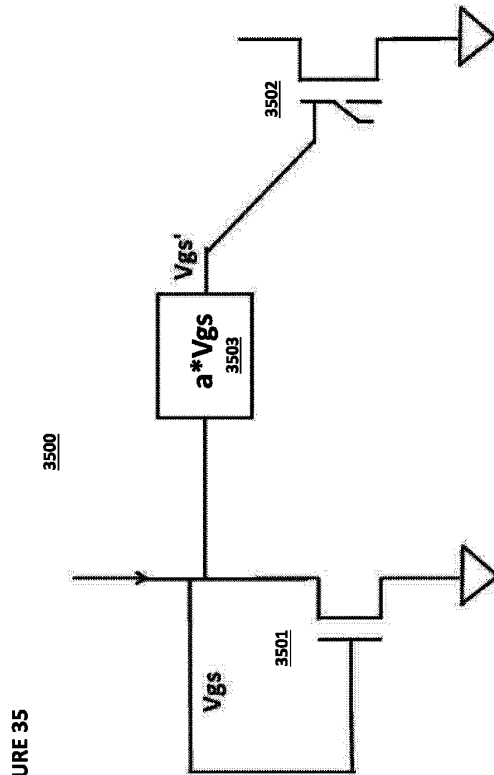


FIGURE 35

10

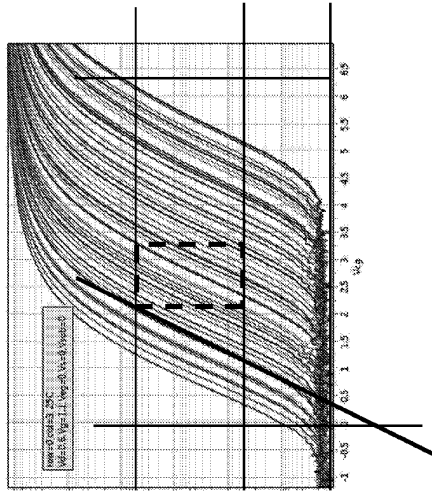
20

30

40

50

【図 36】



【図 37】

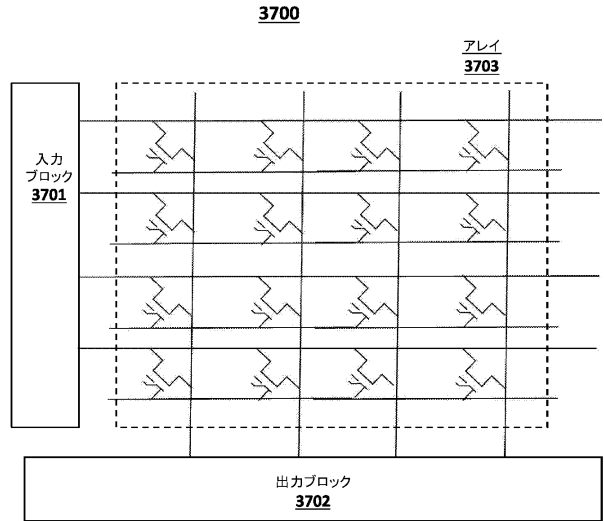
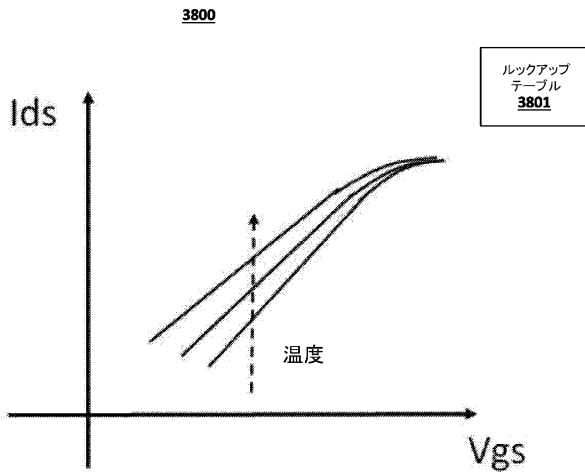


FIGURE 36

【図 38】



10

20

30

40

50

フロントページの続き

(33)優先権主張国・地域又は機関
米国(US)

早期審査対象出願

ニア州、ダブリン、アスターウッド ドライブ 5 5 9 9

(72)発明者 レイテン、マーク

アメリカ合衆国 9 4 5 0 7 カリフォルニア州、アラモ、ラウンドヒル ロード 2 9 4 6

(72)発明者 ドー、ナン

アメリカ合衆国 9 5 0 7 0 カリフォルニア州、サラトガ、ウォールナット アベニュー 2 0 4 5 1

(72)発明者 レムケ、スティーブン

アメリカ合衆国 9 5 0 0 6 カリフォルニア州、ボルダー クリーク、エイコーン ドライブ 5 5 4

審査官 渡辺 順哉

(56)参考文献 特表2021-531609 (JP, A)

GUO, X ほか, Temperature-Insensitive Analog Vector-by-Matrix Multiplier Based on 55 nm NOR Flash Memory Cells, arXiv[online], arXiv, 2016年11月10日, Retrieved from the Internet: URL: <https://arxiv.org/ftp/arxiv/papers/1611/1611.03379.pdf>

(58)調査した分野 (Int.Cl., DB名)

G 0 6 N 3 / 0 0 - 9 9 / 0 0