(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7139426号
(P7139426)

(45)発行日　令和4年9月20日(2022.9.20)

(24)登録日　令和4年9月9日(2022.9.9)

(51)国際特許分類		FI		
G06N	*3/063(2006.01)*	G06N	3/063	
G11C	*11/54 (2006.01)*	G11C	11/54	
G06G	*7/60 (2006.01)*	G06G	7/60	
G11C	*13/00 (2006.01)*	G11C	13/00	270B

請求項の数　17　（全18頁）

(21)出願番号　特願2020-526371(P2020-526371)
(86)(22)出願日　平成30年11月19日(2018.11.19)
(65)公表番号　特表2021-503649(P2021-503649A)
(43)公表日　令和3年2月12日(2021.2.12)
(86)国際出願番号　PCT/IB2018/059102
(87)国際公開番号　WO2019/097494
(87)国際公開日　令和1年5月23日(2019.5.23)
審査請求日　令和3年4月23日(2021.4.23)
(31)優先権主張番号　15/818,521
(32)優先日　平成29年11月20日(2017.11.20)
(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】　人工神経回路網用のシナプス重みを構築するための方法、システム、およびプログラム

(57)【特許請求の範囲】

【請求項１】

人工神経回路網（ＡＮＮ）における固定デバイス非対称性によって誘発される重みへの不要な寄与を軽減する方法であって、前記ＡＮＮにおける各シナプスのシナプス重みは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例が前記ＡＮＮへその出力を確認しながらシリアルに入力され、前記方法は、

第１の訓練サイクルに相当する第１の転送間隔中に、低重要度のコンダクタンス対（ｇ－ｇｓｈａｒｅｄ）から高重要度のコンダクタンス対（Ｇ＋－Ｇ－）へ、前記シナプス重みの一部を転送することと、

前記第１の転送間隔後であり、また第２の訓練サイクルに相当する第２の転送間隔中に、極性が反転している前記低重要度のコンダクタンス対－（ｇ－ｇｓｈａｒｅｄ）から、高重要度のコンダクタンス対（Ｇ＋－Ｇ－）に、前記シナプス重みの一部を転送することと、を含み、

このＡＮＮにおける固定デバイス非対称性が、複数の前記第１および第２の転送間隔にわたって打ち消される、方法。

【請求項２】

前記ＡＮＮにおけるニューロンがＣＭＯＳ回路を用いて実装される、請求項１に記載の方法。

【請求項３】

各シナプスの前記シナプス重みが、アナログ・メモリ素子を使用して実装される、請求

項１に記載の方法。
【請求項４】
前記アナログ・メモリ素子が、不揮発性メモリ（ＮＶＭ）である、請求項３に記載の方法。
【請求項５】
前記ＮＶＭが、相変化メモリ、抵抗性ランダム・アクセス・メモリ、および導電性ブリッジ・ランダム・アクセス・メモリから成る群から選択される、請求項４に記載の方法。
【請求項６】
人工神経回路網（ＡＮＮ）において実施される方法であって、前記ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、前記シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、各シナプスの前記シナプス重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例が前記ＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力と前記所与の層からの所望の出力との間の差に応じて前記シナプス重みを更新し、前記方法は、
総シナプス重みが実質的に変わらないままであるように、低重要度のコンダクタンス対から高重要度のコンダクタンス対へ前記シナプス重みの一部を周期的に転送することと、
シナプス重みのこれらの周期的転送のうちの１つまたは複数を受けて、前記低重要度のコンダクタンス対の極性を反転させることと、
ネットワーク出力が所定の精度で前記所望の出力に近づくまで、前記訓練例の前記シリアル入力を繰り返すことと、を含む、方法。
【請求項７】
前記ＡＮＮにおける前記シナプス重みの一部が、下流ニューロンおよび上流ニューロンに関連付けられた信号に従って、コンデンサへ電荷を加え、また前記コンデンサから電荷を減じるために、プログラミング・トランジスタと共に、読み取りトランジスタのゲートに紐付けされたコンデンサを使用して実装される、請求項６に記載の方法。
【請求項８】
前記ＡＮＮが、パターン認識アプリケーション、画像認識アプリケーション、音声認識アプリケーション、および言語処理アプリケーションから成る群から選択される、請求項６に記載の方法。
【請求項９】
前記低重要度のコンダクタンス対の前記極性が、前記低重要度のコンダクタンス対において、２つのコンダクタンスの役割を入れ替えることによって反転される、請求項６に記載の方法。
【請求項１０】
前記低重要度のコンダクタンス対における前記２つのコンダクタンスのうちの１つが、第１の共用コンダクタンスから第２の共用コンダクタンスへ再構成される、請求項９に記載の方法。
【請求項１１】
人工神経回路網（ＡＮＮ）において実施される方法であって、前記ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、前記シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、前記シナプスのそれぞれの前記重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、前記複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例が前記ＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力と前記所与の層からの所

望の出力との間の差に応じて前記シナプス重みを更新し、前記方法は、
（ａ）前記ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させ、２つ以上のコンダクタンス対からの算術寄与によって、各シナプスに対して元のシナプス重みを計算することと、
（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、
（ｃ）前記低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、前記元のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、
（ｄ）後続の訓練作業に向けて、前記低重要度のコンダクタンス対の極性を反転させることと、を含む、方法。
【請求項１２】
人工神経回路網（ＡＮＮ）において実施される方法であって、前記ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、前記シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、前記シナプスのそれぞれの前記重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、前記複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例が前記ＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力と前記所与の層からの所望の出力との間の差に応じて前記シナプス重みを更新し、前記方法は、
（ａ）前記ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させることと、
（ｂ）所与のコンダクタンスにおいて、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、
（ｃ）前記低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、同様のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、
（ｄ）後続の訓練作業に向けて、前記低重要度のコンダクタンス対の極性を反転させることと、を含み、
前記ＡＮＮにおける前記シナプス重みの一部が、下流ニューロンおよび上流ニューロンに関連付けられた信号に従って、コンデンサへ電荷を加え、また前記コンデンサから電荷を減じるために、プログラミング・トランジスタと共に、読み取りトランジスタのゲートに紐付けされた前記コンデンサを使用して実装される、方法。
【請求項１３】
人工神経回路網（ＡＮＮ）における固定デバイス非対称性によって誘発される重みへの不要な寄与を軽減するシステムであって、前記ＡＮＮにおける各シナプスのシナプス重みは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例が前記ＡＮＮへその出力を確認しながらシリアルに入力され、前記システムは、
第１の訓練サイクルに相当する第１の転送間隔中に、低重要度のコンダクタンス対（ｇ－ｇｓｈａｒｅｄ）から高重要度のコンダクタンス対（Ｇ＋－Ｇ－）へ、前記シナプス重みの一部を転送することと、
前記第１の転送間隔後であり、また第２の訓練サイクルに相当する第２の転送間隔中に、極性が反転している前記低重要度のコンダクタンス対－（ｇ－ｇｓｈａｒｅｄ）から、高重要度のコンダクタンス対（Ｇ＋－Ｇ－）に、前記シナプス重みの一部を転送することと、を含む方法を行うように動作可能であり、
このＡＮＮにおける固定デバイス非対称性が、複数の前記第１および第２の転送間隔にわたって打ち消される、システム。
【請求項１４】

人工神経回路網（ＡＮＮ）において実装されたシステムであって、前記ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、前記シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、各シナプスの前記シナプス重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例が前記ＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力と前記所与の層からの所望の出力との間の差に応じて前記シナプス重みを更新し、前記システムは、

総シナプス重みが実質的に変わらないままであるように、低重要度のコンダクタンス対から高重要度のコンダクタンス対へ前記シナプス重みの一部を周期的に転送することと、

シナプス重みのこれらの周期的転送のうちの１つまたは複数を受けて、前記低重要度のコンダクタンス対の極性を反転させることと、

ネットワーク出力が所定の精度内で前記所望の出力に近づくまで、前記訓練例の前記シリアル入力を繰り返すことと、を含む方法を行うように動作可能である、システム。

【請求項１５】

人工神経回路網（ＡＮＮ）において実装されたシステムであって、前記ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、前記シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、前記シナプスのそれぞれの前記重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、前記複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例が前記ＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力と前記所与の層からの所望の出力との間の差に応じて前記シナプス重みを更新し、方法であって、

（ａ）前記ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させ、２つ以上のコンダクタンス対からの算術寄与によって各シナプスに対して元のシナプス重みを計算することと、

（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、

（ｃ）前記低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、前記元のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、

（ｄ）後続の訓練作業に向けて、前記低重要度のコンダクタンス対の極性を反転させることと、を含む方法を行うように動作可能である、システム。

【請求項１６】

人工神経回路網（ＡＮＮ）において実装されたシステムであって、前記ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、前記シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、前記シナプスのそれぞれの前記重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、前記複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例が前記ＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力と前記所与の層からの所望の出力との間の差に応じて前記シナプス重みを更新し、前記システムは、

（ａ）前記ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させることと、

（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコ

ンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、
（ｃ）前記低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、同様のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、
（ｄ）後続の訓練作業に向けて、前記低重要度のコンダクタンス対の極性を反転させることと、を含む方法を行うように動作可能であり、
前記ＡＮＮにおける前記シナプス重みの一部が、下流ニューロンおよび上流ニューロンに関連付けられた信号に従って、コンデンサへ電荷を加え、また前記コンデンサから電荷を減じるために、プログラミング・トランジスタと共に、読み取りトランジスタのゲートに紐付けされたコンデンサを使用して実装される、システム。

【請求項１７】
人工神経回路網において実施されるコンピュータ・プログラムであって、
請求項１ないし１２のいずれか一項に記載の方法のステップを、前記人工神経回路網に実行させる、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、人工神経回路網（ＡＮＮ）の分野に関する。より詳細には、本発明は、低重要度のコンダクタンス対の極性が周期的に反転される、変動する重要度の符号付きアナログ・コンダクタンス対から人工神経回路網用のシナプス重みを構築するためのシステムおよび方法に関する。

【背景技術】
【０００２】
人工神経回路網（ＡＮＮ）は、シナプスと呼ばれる結合点を通して相互に連絡するいくつかのニューロンから成る分散型コンピューティング・システムである。各シナプスは、１つのニューロンの出力と別のニューロンの入力との結合の強度を符号化する。各ニューロンの出力は、それに結合されている他のニューロンから受信した入力の集計によって、したがって、これらの「上流の」結合されたニューロンの出力と、シナプス重みによって決まる結合の強度とによって、決まってくる。特定の入力クラスが所望の出力をもたらすようにシナプスの重みを調整することにより、特定の問題（例えば、パターン認識）を解くように、ＡＮＮが訓練される。この重み調整手順は、「学習」として知られている。ＡＮＮの文献には、画像認識、音声認識、言語処理などの様々なタスクに適している学習を行う際の多くのアルゴリズムがある。理想的には、これらのアルゴリズムは、学習プロセス中、所与の問題の最適解へ向かって収束するシナプス重みパターンをもたらす。
【０００３】
ＡＮＮの魅力的な実装形態は、いくつかの（例えば、ＣＭＯＳ）回路を使用してニューロンを表現し、その機能は、特定のニューロンが結合されている上流ニューロンから総入力を統合または総計すること、および、入力のいくつかの非線形関数を適用して、そのニューロンの出力を得ることである。一般的に、それぞれのニューロンは他のニューロンのいくらかの大きな断片に結合されているので、シナプス（結合）の数はニューロンの数よりずっと大きい。したがって、神経形態学的コンピューティング・チップ上で超高密度を達成できるシナプスのいくつかの実装形態を使用するのが有利である。１つの魅力的な選択は、抵抗性ランダム・アクセス・メモリ（ＲＲＡＭ：Resistive Random Access Memory）または相変化メモリ（ＰＣＭ：Phase-Change Memory）などのいくつかの不揮発性メモリ（ＮＶＭ：Non-Volatile Memory）技術である。正および負（すなわち、興奮性および抑制性）の重みが望ましいので、１つのスキームは、一対のＮＶＭを使用して、２つの間のコンダクタンスにおける差として重みを表現する（M.Suri et.al.,“Phase Change Memory as Synapse for Ultra-Dense Neuromorphic System:Application to Complex Visual Pattern Extraction”,IEDM Technical Digest,4.4,2011を参照）。このスキームは図１に示される。上流Ｎ¡ニューロンの出力は、下流Ｍ¡ニューロンの正

および負の入力に、一対のＮＶＭコンダクタンスを通して同時に総計される。この並列性は、効率的な計算に非常に有利である。

【０００４】

学習中、ＮＶＭ素子のコンダクタンスは、学習規則により、コンダクタンスを増やすか減らすかいずれかができるパルスをそれらに送信することによってプログラムされる。研究された１つの共通の学習規則は、逆伝播であり（Rumelhart et.al.,“Learning Internal Representations by Error Propagation”,Parallel Distributed Processing:Explorations in the Microstructure of Cognition,Vol1,1986を参照）、これは、ビデオ・ゲームなどをするのを学習する、画像認識のためのグラフィック処理ユニット（ＧＰＵ：Graphical Processing Unit）上に現在実装されている、深層学習ネットワークにおいて広範囲に使用されている。逆伝播アルゴリズムは、学習率として知られる比例定数ηと、上流ニューロンの出力χ_i、および下流ニューロンからの誤差寄与δ_jの積に比例する、重み更新$\Delta\omega_{ij}=\eta\cdot\chi_i\cdot\delta_j$を要求する。この「クロスバー互換性」学習規則は、従来の逆伝播規則と同等に効果的であることが以前に示されていた（Burr et.al.,“Experimental demonstration and tolerancing of a large scale neural network(165,000synapses),using phase-change memory as the synaptic weight element”,IEDM Technical Digest,2014を参照）。

【０００５】

任意の実際のＮＶＭ素子は、非理想的反応を有する。それは、非線形であり、それが達成できる最大コンダクタンスを制限する。コンダクタンスを上げるように設計されたパルスに対するコンダクタンス変化は、コンダクタンスを下げるように設計されたパルスに対するコンダクタンス変化とは異なり、すなわち、その反応は非対称である。デバイス間にはバラツキがあり、いくつかのデバイスは作動不能となり、高コンダクタンス状態で止まるか低コンダクタンス状態で止まるかのいずれかとなる。これらの欠点の多くは、ＡＮＮ性能の極めて小さい低下を引き起こすことが本発明者らの研究によって示された。しかし、非線形性、有界コンダクタンス、および非対称性応答は、訓練中の９９＋％精度から、８０％～８５％のいずれかまで、ＭＮＩＳＴ数字認識問題に対する精度の低下を引き起こす。

【０００６】

訓練中、多くの様々な入力は、ネットワークに提示され、逆伝播学習規則が、それぞれの訓練の後に（またはミニバッチと呼ばれる、わずかな入力の後に）ＮＶＭコンダクタンスを更新するのに使用される。ネットワークにおけるいくつかの重みは、いくつかの安定した値の方に着実に進展するが、他は上下に揺れ動く（dither）傾向があり、時には上がり、他の時には下がる。ＮＶＭ反応が非線形性または非対称性である場合、重み値を下げるように意図されたパルスに対する反応は、重みを上げるように意図された反応より強くなる。これは、多くのこれらの重みをゼロの方に向かわせる傾向があり、逆伝播学習規則を無効にし、ネットワーク性能を低下させる。

【０００７】

先行技術の研究は、２つのコンダクタンスに基づいた符号付きシナプス重みの導入である。しかし、相変化メモリ（ＰＣＭ）、フィラメント系ＲＲＡＭ（ＨｆＯｘまたはＴａＯｘを使用するなど）、または金属フィラメントに基づいた導電性ブリッジＲＡＭなどの、いくつかのＮＶＭデバイスに対して、小さなコンダクタンス変化は、１つの方向にのみ実施され得る。この直接的結果として、大きいシナプス重みは、極めて脆弱である傾向があり、より小さなコンダクタンス（重みの大きさを下げる）におけるステップにはよく反応するが、より大きなコンダクタンス（重みの大きさを上げる）におけるステップには不十分にしか反応しない。したがって、ネットワークが大きくしたい重みが、大きいままでいることが難しいので、ネットワーク性能は低下する。

【先行技術文献】

【特許文献】

【０００８】

【文献】ＵＳＰＴＯ第１５／６６０，９４５号
【非特許文献】
【０００９】
【文献】M.Suri et.al.,“Phase Change Memory as Synapse for Ultra-Dense Neuromorphic System:Application to Complex Visual Pattern Extraction”,IEDM Technical Digest,4.4,2011
Rumelhart et.al.,“Learning Internal Representations by Error Propagation”,Parallel Distributed Processing:Explorations in the Microstructure of Cognition, Vol1,1986
Burr et.al.,“Experimental demonstration and tolerancing of a large scale neural network(165,000synapses),using phase-change memory as thesynaptic weight element”,IEDM Technical Digest,2014
【発明の概要】
【発明が解決しようとする課題】
【００１０】
本発明の実施形態は、先行技術と先行発明の両方のシステムおよび方法にわたる改良である。本発明は、人工神経回路網用のシナプス重みを構築するための方法、システム、およびプログラムを、より詳細には、人工神経回路網における固定デバイス非対称性によって誘発される重みへの不要な寄与を軽減する方法、システム、およびプログラムを提供する。
【課題を解決するための手段】
【００１１】
一実施形態は、人工神経回路網における固定デバイス非対称性によって誘発される重みへの不要な寄与を軽減する方法である。ＡＮＮにおける各シナプスのシナプス重みは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例がＡＮＮへその出力を確認しながらシリアルに入力され、この方法は、第１の訓練サイクルに相当する第１の転送間隔中に、低重要度のコンダクタンス対（$g-g_{shared}$）から高重要度のコンダクタンス対（$G^{+}-G^{-}$）へ、シナプス重みの一部を転送することと、第１の転送間隔後であり、また第２の訓練サイクルに相当する第２の転送間隔中に、極性が反転している低重要度のコンダクタンス対－（$g-g_{shared}$）から、高重要度のコンダクタンス対（$G^{+}-G^{-}$）に、シナプス重みの一部を転送することと、を含み、このＡＮＮにおける固定デバイス非対称性が、複数のこのような第１および第２の転送間隔にわたって打ち消される。
【００１２】
別の実施形態において、本発明は、人工神経回路網（ＡＮＮ）において実施される方法を提供し、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれが、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、各シナプスのシナプス重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じてシナプス重みを更新し、この方法は、（ａ）総シナプス重みが実質的に変わらないままであるように、低重要度のコンダクタンス対から高重要度のコンダクタンス対へシナプス重みの一部を周期的に転送することと、（ｂ）シナプス重みのこれらの周期的転送のうちの１つまたは複数を受けて、低重要度のコンダクタンス対の極性を反転させることと、（ｃ）ネットワーク出力が所定の精度内で所望の出力に近づくまで、訓練例のシリアル入力を繰り返すことと、を含む。
【００１３】
また別の実施形態において、本発明は、人工神経回路網（ＡＮＮ）において実施される

方法を提供し、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、シナプスのそれぞれの重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、複数のコンダクタンス対のそれぞれは、ジョイント寄与（joint contribution）に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じてシナプス重みを更新し、この方法は、（ａ）ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させ、２つ以上のコンダクタンス対からの算術寄与によって各シナプスに対して元のシナプス重みを計算することと、（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、（ｃ）低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、その元のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、（ｄ）後続の訓練作業に向けて、低重要度のコンダクタンス対の極性を反転させることと、を含む。

【００１４】

別の実施形態において、本発明は、人工神経回路網（ＡＮＮ）において実施される方法を提供し、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、シナプスのそれぞれの重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じてシナプス重みを更新し、この方法は、（ａ）ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させ、２つ以上のコンダクタンス対からの算術寄与によって各シナプスに対して元のシナプス重みを計算することと、（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、（ｃ）低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、その元のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、（ｄ）後続の訓練作業に向けて、低重要度のコンダクタンス対の極性を反転させることと、を含み、ＡＮＮにおけるシナプス重みの一部が、下流ニューロンおよび上流ニューロンに関連付けられた信号に従って、コンデンサへ電荷を加え、またコンデンサから電荷を減じるために、プログラミング・トランジスタと共に、読み取りトランジスタのゲートに紐付けされたコンデンサを使用して実装される。

【００１５】

一実施形態により、人工神経回路網（ＡＮＮ）において実施される方法が提供され、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、シナプスのそれぞれの重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入力さ

れ、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じて、シナプス重みを更新し、この方法は、（ａ）ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させることと、（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、（ｃ）低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、同様のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、（ｄ）後続の訓練作業に向けて、低重要度のコンダクタンス対の極性を反転させることと、を含み、ＡＮＮにおけるシナプス重みの一部が、下流ニューロンおよび上流ニューロンに関連付けられた信号に従って、コンデンサへ電荷を加え、またコンデンサから電荷を減じるために、プログラミング・トランジスタと共に、読み取りトランジスタのゲートに紐付けされたコンデンサを使用して実装される。

【００１６】

一実施形態により、人工神経回路網（ＡＮＮ）における固定デバイス非対称性によって誘発される重みへの不要な寄与を軽減するシステムが提供され、ＡＮＮにおける各シナプスのシナプス重みは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例がＡＮＮへその出力を確認しながらシリアルに入力され、システムは、第１の訓練サイクルに相当する第１の転送間隔中に、低重要度のコンダクタンス対（ｇ－ｇ$_{shared}$）から高重要度のコンダクタンス対（$G^{+}-G^{-}$）へ、シナプス重みの一部を転送することと、第１の転送間隔後であり、また第２の訓練サイクルに相当する第２の転送間隔中に、極性が反転している前記低重要度のコンダクタンス対－（ｇ－ｇ$_{shared}$）から、高重要度のコンダクタンス対（$G^{+}-G^{-}$）に、前記シナプス重みの一部を転送することと、を含む方法を行うように動作可能であり、このＡＮＮにおける固定デバイス非対称性が、複数のこのような第１および第２の転送間隔にわたって打ち消される。

【００１７】

一実施形態により、人工神経回路網（ＡＮＮ）において実装されたシステムが提供され、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれが、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、各シナプスのシナプス重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じて、シナプス重みを更新し、このシステムは、総シナプス重みが実質的に変わらないままであるように、低重要度のコンダクタンス対から高重要度のコンダクタンス対へ、シナプス重みの一部を周期的に転送することと、シナプス重みのこれらの周期的転送のうちの１つまたは複数を受けて、低重要度のコンダクタンス対の極性を反転させることと、ネットワーク出力が所定の精度内で所望の出力に近づくまで、訓練例のシリアル入力を繰り返すことと、を含む方法を行うように動作可能である。

【００１８】

一実施形態により、人工神経回路網（ＡＮＮ）において実装されたシステムが提供され、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、シナプスのそれぞれの重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入

力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じて、シナプス重みを更新し、このシステムは、（ａ）ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させ、２つ以上のコンダクタンス対からの算術寄与によって各シナプスに対して元のシナプス重みを計算することと、（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、（ｃ）低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、その元のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、（ｄ）後続の訓練作業に向けて、低重要度のコンダクタンス対の極性を反転させることと、を含む方法を行うように動作可能である。

【００１９】

一実施形態により、人工神経回路網（ＡＮＮ）において実装されたシステムが提供され、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、シナプスのそれぞれの重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じて、シナプス重みを更新し、このシステムは、（ａ）ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させることと、（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、（ｃ）低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、同様のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、（ｄ）後続の訓練作業に向けて、低重要度のコンダクタンス対の極性を反転させることと、を含む方法を行うように動作可能であり、ＡＮＮにおけるシナプス重みの一部が、下流ニューロンおよび上流ニューロンに関連付けられた信号に従って、コンデンサへ電荷を加え、またコンデンサから電荷を減じるために、プログラミング・トランジスタと共に、読み取りトランジスタのゲートに紐付けされたコンデンサを使用して実装される。

【００２０】

本発明がコンピュータ・プログラムとして実施される可能性のあることが分かるであろう。

【００２１】

本開示は、１つまたは複数の様々な例により、以下の図を参照して詳細に記述される。図面は、例示のみを目的として提供されており、本開示の例を示しているにすぎない。これらの図面は、本開示の読者の理解を容易にするために提供され、本開示の幅、範囲、または適用性を限定するものとみなされるべきではない。説明を簡単で分かり易くするために、これらの図面は必ずしも正しい縮尺とは限らないことに留意されたい。

【図面の簡単な説明】

【００２２】

【図１】シナプス当たり１つのコンダクタンス対を有する先行技術のアレイを示す図である。

【図２】一実施形態による、１組の対のその読み取り電流が他に対してＦ倍増幅された、シナプス当たり２つ以上のコンダクタンス対を有するアレイを描写する図である。

【図３】一実施形態による、１組の対のその読み取り電流が他に対してＦ倍増幅され、そこでは、低重要度対の極性が反転されている、シナプス当たり２つ以上のコンダクタンス

対を有するアレイを描写する図である。
【図４】一実施形態による、低重要度コンダクタンス対の極性を周期的に反転させる技法の利益を示す、シミュレーション結果を描写する図である。
【図５】一実施形態による、ニューロンM_1と同じ列におけるシナプスの転送直後の状況を示す図であり、ニューロンM_1は左側と共用しており、かつそのｇ値の極性がちょうど反転されており（正味の重みからｇを減じる）、一方ニューロンM_mは、正味の重みに加えられたｇ値を持ち続けており、その右に位置するg_{shared}デバイスと共用している。
【図６】一実施形態による、ｇ値がここでも反転されている、列M_mの転送後の状況を示す図であり、ここでは、g_{shared}デバイスからの共用（shared）バイアス読み取り電流が、共用デバイスから、（最初の共用列を得ると同時に）初期値に最近リセットされた左側に送られている。
【発明を実施するための形態】
【００２３】
好適な実施形態の米国記載
本発明は好適な実施形態において説明かつ記述されるが、本発明は多くの様々な構成において作り出され得る。本発明の好ましい実施形態は、本開示が本発明の原理および関連するその構造に関する機能上の仕様を例示するものとみなされるべきであり、本発明が例示された実施形態に限定されることを意図するものでないことを理解した上で、図面に記載され、本明細書で詳細に説明される。当業者であれば、本発明の範囲内で多くの他の考えられ得る変形形態を想到するであろう。
【００２４】
ここで記述された本発明は、それぞれが重要度にバラツキがある、複数のコンダクタンス対にわたるシナプス重みを分散させることによって、コンダクタンス応答において非線形性および非対称性の影響を低減するための技法の先の開示に基づいている。それぞれの対は、通常通り、興奮性コンダクタンスＧ＋と抑制性コンダクタンスＧ－との間の差を通してシナプス重みに寄与する。しかし、それぞれの対の寄与は、それらの対が合計されるビット線の端において実装された増幅率において相違する。例ごとのプログラミングが、様々な態様で実装され得る（１つ明らかなことにおいて、最低重要度のコンダクタンス対が、ネットワーク訓練中に更新される）。ほとんどの重み値は通常、より重要度の高いコンダクタンス対に含まれるので、より重要度の低いコンダクタンス対のうちの１つまたは複数は、読み取りトランジスタのゲートに紐付けされたコンデンサなどの、揮発性アナログ・メモリ素子から構成され得る。この場合、コンデンサへの電荷を追加または差し引くことにより、読み取りトランジスタのゲート電圧、延いては、その実効コンダクタンスがことができる、コンダクタンス値の場合、それぞれのコンダクタンス対におけるコンダクタンスの１つは、すべての個々のシナプス・プログラミングが非共用コンダクタンスを調整することによって起こる状態で、多くの様々なシナプスにわたって共用され得る。
【００２５】
出願者の先の発明（ＵＳＰＴＯ第１５／６６０，９４５号（参照：ＡＲＣ９２０１６００３１ＵＳ１）、このコピーはファイル上にある）において、より重要度の低い符号付きアナログ・コンダクタンス対からより重要度の高いアナログ・コンダクタンス対への重み転送を実装する、重要度を変える複数のコンダクタンスにわたってシナプス重みを分散させることによって性能を向上させる方法が導入された。図２は、１組の対のその読み取り電流が他に対してＦ倍増幅された、シナプス当たり２つ以上のコンダクタンス対を有する、先に開示されたアレイ（すなわち、ＵＳＰＴＯ第１５／６６０，９４５号に開示される）を示す。この図において、それぞれのシナプス重みは、２つのコンダクタンス対から成る。第１の対は、第２の対に対してＦのゲイン係数でアレイの底で、その下流ニューロンの合計電流に寄与する。このことは、第２の対における大きなコンダクタンス差は、総シナプス重みに対して、第１の対における小さなコンダクタンス差と同じ寄与を有することを意味する。したがって、この技法は、総シナプスの全ダイナミック・レンジを増加させ

、また、累積したシナプス重み変化が第１のコンダクタンス対に時折慎重に転送される状態で、より線形で対称なコンダクタンス変化を示す、高速であるが揮発性のアナログ・メモリ素子が、ＡＮＮの例ごとのプログラミングに対する、第２または低重要度のコンダクタンス対として使用されることを可能にする。これは、ここには同じ総シナプス重みを正確に表す複数のコンダクタンスの複数のアナログ構成があるので、デジタル・ワードの有効ビットに多かれ少なかれ直接一致しないことに注意されたい。同様に、低重要度から高重要度のコンダクタンス対への転送演算は、デジタル加算器においてデジタル「キャリー」演算と同じではなく、総シナプス重みへのアナログ構成のマッピングにおける冗長性のために、必要に応じて遅延され得る。図は２つのコンダクタンス対を示しているが、３つ以上の対への拡張が可能である。

【００２６】

大きなシナプス重みが展開された場合、この情報は、低重要度のコンダクタンス対における大きなコンダクタンス差から、次に重要度の高いコンダクタンス対における小さな（したがって、より簡単に維持される）コンダクタンス差に転送される。より重要度の高いコンダクタンス対は低い頻度で更新されるので、この大きなシナプス重みは、コンダクタンスの非線形性および非対称性を通して失われることからより良く保護されるが、ネットワークは、そうすることを選択する場合、依然としてこの重みを軽減することもできる。ある程度まで、ネットワークは、高重要度のコンダクタンス対にこの重みを「積み上げ」ている。

【００２７】

しかし、ネットワークが、重み減少とほとんど同じ回数だけ重み増加を要求し、これらの反対符号の寄与を正確に取り消させるように、低重要度のコンダクタンス対の重み更新ができるだけ対称であることは依然として極めて重要である。非対称性の主な原因は２つある。すなわち、コンダクタンス応答（通常、その後に飽和する急激な応答）における非線形性、および固定デバイスの非対称性（１つのデバイスは、他の対になったコンダクタンスよりも、コンダクタンスを変えることについて単に常に強力であるかまたはより効果的である）である。

【００２８】

より線形なコンダクタンス更新のための１つの選択肢は、ＣＭＯＳベースの揮発性コンダクタンス、つまりコンデンサに関連付けられたそのゲートを有するトランジスタを使用することである。トランジスタの対を使用して、電荷は、極めて線形にコンデンサに加算および減算され得る。しかし、固有のＣＭＯＳ製作の可変性により、（充電／放電トランジスタのこの対の両部材を含んでいる）ウェーハー上のそれぞれのトランジスタは、閾値電圧、延いては、それらのコンデンサに加えられるかまたは差し引かれる充電量を変化させる。これは、コンダクタンス増加のサイズとコンダクタンス減少のサイズとの間に強力に固定された非対称性を再導入する可能性がある。同様に、相変化メモリ（ＰＣＭ）デバイスまたは他のタイプの不揮発性メモリ（ＮＶＭ）デバイスは、所与の部分ＳＥＴパルスの反復印加が誘発される、最大飽和コンダクタンスにおける差のために、固定された非対称性も示す可能性がある。

【００２９】

したがって、固定デバイス非対称性によって導入されたコンダクタンス更新の非対称性を抑える技法が当技術分野において必要とされる。

【００３０】

「バラツキのある重要度の複数のコンダクタンス」を使用する場合の訓練中、低重要度コンダクタンスｇ（g^{shared}からのその基準電流に関して）のみは、重み転送が行われるまで更新される。チップ実装形態において、転送は一度に１つの列で行われる。重み転送を開始する前、デバイスの列全体の現在の重みは、Ｆ×（$G^{+}-G^{-}$）＋（$g-g^{shared}$）であり、ここで、Ｆは約２～５の小さなゲイン係数である。現在の重み値は、ｇとg^{shared}の両方がそれらの初期状態にリセットされた後に、（列全体用の）周辺回路に記録されることになり、こうして、アレイ上に格納された重みを変える。例えばＰＣ

Ｍを実装した、不揮発性高重要度のコンダクタンス対（$G^{+}-G^{-}$）のバラツキおよび非線形応答は、アレイ上に格納された重みが所望のターゲット値（周辺回路に一時的に保存された値と比較することによって）に返るように、閉ループ重みチューニングによって補償され得る。

【００３１】

重み情報が、$g-g_{shared}$コンダクタンス対から$G^{+}-G^{-}$ＰＣＭ対まで転送されると、任意の残余の重みチューニング誤差は、重み総数が、周辺回路に初めに記録されていた値に再び一致するまで、ｇコンダクタンスをプログラミングすることによって対処され得る。その時点では、周辺回路に一時的に保持されていた、現在の重み値は処分される。

【００３２】

しかし、問題が３Ｔ１Ｃデバイス（例えば、ＵＳＰＴＯ第１５／６６０，９４５号の図６を参照）に残っており、そのデバイスのＰＦＥＴは、この固定非対称性のために、転送間隔が終わる度に、重み増加が報告されることがある、そのＮＦＥＴより効果的である。すべての重みが、$g-g_{shared}$コンダクタンス対から転送されたので、このコンダクタンス対の有効な極性は、次の転送間隔中に反転される。図３は、１組の対のその読み取り電流が他に対してＦ倍増幅された状態で、シナプス当たり２つ以上のコンダクタンス対を有するアレイを描写し、その図において、低重要度対の極性が反転されている。これは、（前向き推論と逆伝播の両方に対して）デバイス電流を読み取る間には方程式$F\times(G^{+}-G^{-})-(g-g_{shared})$を使用し、重みを減らすときには電荷を加えるようにＰＦＥＴを使用し、かつ、重みを増やすときには電荷を減らすようにＮＦＥＴを使用することを意味する。極性反転は、ｇおよびg_{shared}が同じ電圧にリセットされ、ｇの値を調節することによって$G^{+}-G^{-}$の値における任意の残余の重みチューニング誤差を修正することを開始する前の瞬間に起こらなければならない。このｇに対する調節は、このデバイス上で次の重み転送演算までずっと、すべての相互作用のためのこのデバイスに対して使用されることになる、極性の新しい値で行われなければならない。基準デバイスの共用によってもたらされるさらなる厄介な問題があるが、これらも対処され得る。

【００３３】

それぞれの転送間隔の後、$g-g_{shared}$のために使用された極性は、後続の訓練サイクルの間中、反転される。その結果として、固定デバイス非対称性によって誘発される、重みへの不要な寄与は、複数の転送間隔にわたって打ち消される傾向にある。図４は、この極性反転技法（２０個のエポックに延びる、上の実線および点線グラフ）は、ＳＰＩＣＥ回路シミュレーションによってモデル化されるような、ＣＭＯＳトランジスタの予測される可変性を含む、シミュレーション条件の下で性能を著しく向上させる。極性反転なしで、不一致ＣＭＯＳデバイスに起因する、電荷加算と電荷減算との間の固定非対称性は、訓練精度を著しく低下させる。しかし、それぞれの重み転送演算に対する極性を反転させることによって、訓練精度は高く、非常に不完全なデバイスの存在にもかかわらず、純粋なソフトウェア・ベースの訓練の予測に一致する。

【００３４】

極性ビットは、どの電流が前向き推論に結合されているか、どのデバイスが逆読みに対して起動されるか、および、２つのトランジスタのうちのどちらがコンデンサに電荷を加える／差し引くように起動されるかを、どのように制御するのかを下流ニューロンが知るように、アレイの列の端に格納されなければならない。したがって、転送演算は、列単位で行われる必要があり、列の端（Ｓ側かＮ側のいずれか）におけるこの極性ビットは、前に進むこの列に対して、電流極性反転を記録するように反転される。

【００３５】

したがって、唯一の付加的な詳細は、g_{shared}をどのように巧みに処理するかである。１つの手法は、極性ビットを使用して、列－ニューロンが、共用デバイスに対するｇからの流入電流を、左側（より小さい列番号）と比較するべきか、または右側（より大きい列番号）と比較すべきかを信号で知らせることである。この考え方としては、列が、ｇデバイスのその専用の列とどのように相互に作用するかにおいて、その極性を反転させる

ことを選ぶと同時に、列が、すでに初期条件に最近リセットされているg^{shared}デバイスと共用することも始める、ということである。これは、転送に向けて列を扱うのに左から右へ移動するとき、極性ビットの反転が、それまでは右側と共用していたことから、今度は左側と共用することに変わる必要があることを意味する。これはまた、アレイにわたる転送事象の次の列方向の掃引が、今度は、右から左の反対方向において起こり、極性ビットが元の値に入れ替わると同時に、「左側と共用」から今度は「右側と共用」の変更を誘引する必要があることも意味する。図５は、ニューロンM_1と同じ列におけるシナプスの転送直後の状況を示しており、ニューロンM_1は左側と共用しており、かつそのｇ値がちょうど反転されている（正味の重みからｇを減じる）。一方、ニューロンM_mは、正味の重みに加えられたｇ値を持ち続けており、その右に位置するg^{shared}デバイスと共用している。図６は、ｇ値がここでも反転されている、列M_mの転送後の状況を示しており、ここでは、g^{shared}デバイスからの共用バイアス読み取り電流が、共用デバイスから、（最初の共用列を得ると同時に）初期値に最近リセットされた左側に送られている。

【００３６】

同じ列において２つの連続する転送を行う（例えば、左から右への掃引が終了し、右から左の同じデバイスの即座の掃引がそれに続くのにつれて）ことを避けるために、四文区間ベースまたは、区間ベースの手法があり得る。ここでは、ある集合、例えば３２列集合が、左から右、列１から列３２に掃引され、正から負の極性に反転される。次の掃引は、右から左、列６４から列３３に下がり、ここで元の負から正の極性に反転される。その後、３２から１に掃引し、次に３３から６４まで掃引することができ、それにより、この区間を、それらが当初持っていたのと同じ極性に戻す。区間サイズと区間数との様々な組合せを使用して、各転送間隔の長さをできる限り等しくしようとしてよい。

【００３７】

一実施形態において、本発明は、人工神経回路網（ＡＮＮ）における固定デバイス非対称性によって誘発される重みへの不要な寄与を軽減する方法を提供し、ＡＮＮにおける各シナプスのシナプス重みは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例がＡＮＮへその出力を確認しながらシリアルに入力され、そこでは、第１の訓練サイクルに相当する第１の転送間隔中に、低重要度のコンダクタンス対（$g-g^{shared}$）から高重要度のコンダクタンス対（$G^{+}-G^{-}$）へ、シナプス重みの一部を転送することと、第１の転送間隔後であり、また第２の訓練サイクルに相当する第２の転送間隔中に、極性が反転している低重要度のコンダクタンス対－（$g-g^{shared}$）から、高重要度のコンダクタンス対（$G^{+}-G^{-}$）に、シナプス重みの一部を転送することと、を含み、このＡＮＮにおける固定デバイス非対称性が、複数のこのような第１および第２の転送間隔にわたって打ち消される。

【００３８】

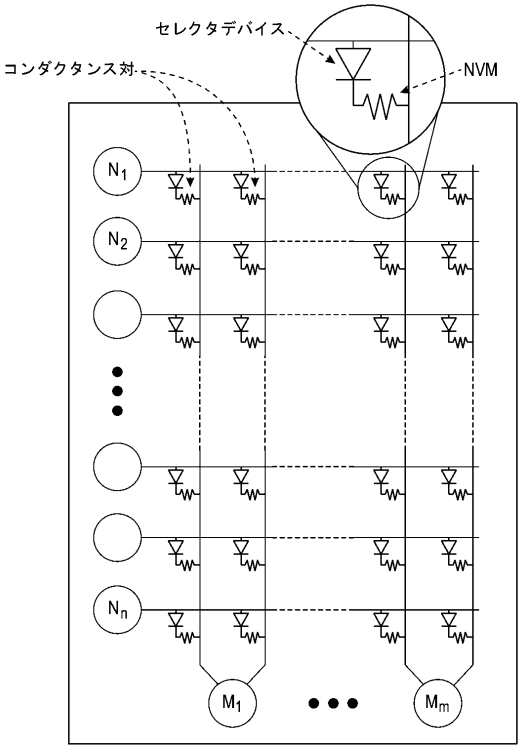
別の実施形態において、本発明は、人工神経回路網（ＡＮＮ）において実施される方法を提供し、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれが、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、各シナプスのシナプス重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、そこでは、複数の訓練例がＡＮＮにその出力を確認しながらシリアルに入力され、そこでは、逆伝播アルゴリズムが、所与の層からの出力とその所与の層からの所望の出力との間の差に応じて、シナプス重みを更新し、この方法は、（ａ）総シナプス重みが実質的に変わらないままであるように、低重要度のコンダクタンス対から高重要度のコンダクタンス対へシナプス重みの一部を周期的に転送することと、（ｂ）シナプス重みのこれらの周期的転送のうちの１つまたは複数を受けて、低重要度のコンダクタンス対の極性を反転させることと、（ｃ）ネットワーク出力が所定の精度内で所望の出力に近づくまで、訓練例のシリアル入力を繰り返すことと、を含む。

【００３９】
また別の実施形態において、本発明は、人工神経回路網（ＡＮＮ）において実施される方法を提供し、ＡＮＮは、１つの層の出力が後続の層の複数のニューロンの入力に結合された状態で層状に配置された複数のニューロンを備え、そこでは、ニューロンが複数のシナプスを介して互いに結合され、シナプスのそれぞれは、２つの結合されたニューロン間の結合強度を符号化するシナプス重みを有し、シナプスのそれぞれの重みの大きさは、複数のコンダクタンス対からの重み付き電流フローによって表され、複数のコンダクタンス対のそれぞれは、ジョイント寄与に相当し、かつ高重要度コンダクタンス対および低重要度コンダクタンス対を有し、そこでは、複数の訓練例がＡＮＮに所与の層からのその出力を確認しながらシリアルに入力され、この方法は、（ａ）ＡＮＮのネットワークにおけるアナログ・メモリ素子にわたるコンダクタンスの訓練および測定を中断させ、２つ以上のコンダクタンス対からの算術寄与によって各シナプスに対して元のシナプス重みを計算することと、（ｂ）所与のコンダクタンス対において、その絶対値が所定の値だけその対になったコンダクタンスよりも大きい少なくとも１つの測定済みコンダクタンスを特定することと、（ｃ）低重要度コンダクタンス対を、総シナプス重みへのゼロ・ジョイント寄与に対応して、実質的に等しくなるように再構成し、その元のシナプス重み値が得られるまで、より重要なコンダクタンス対のうちの１つを再構成することと、（ｄ）後続の訓練作業に向けて、低重要度のコンダクタンス対の極性を反転させることと、を含む。

【００４０】
結論
重要度にバラツキのある符号付きアナログ・コンダクタンス対から構築された人工神経回路網に対し、シナプス重み内で、低重要度コンダクタンス対の極性を周期的に反転させるためのシステムおよび方法の有効な実施に関して、上記の実施形態において、システムおよび方法を示した。様々な好適な実施形態を示し、説明したが、本発明をこのような開示によって限定する意図はなく、むしろ、添付の特許請求の範囲に定められているような、本発明の思想および範囲に入るすべての変更形態に及ぶことを意図していることが理解されるであろう。

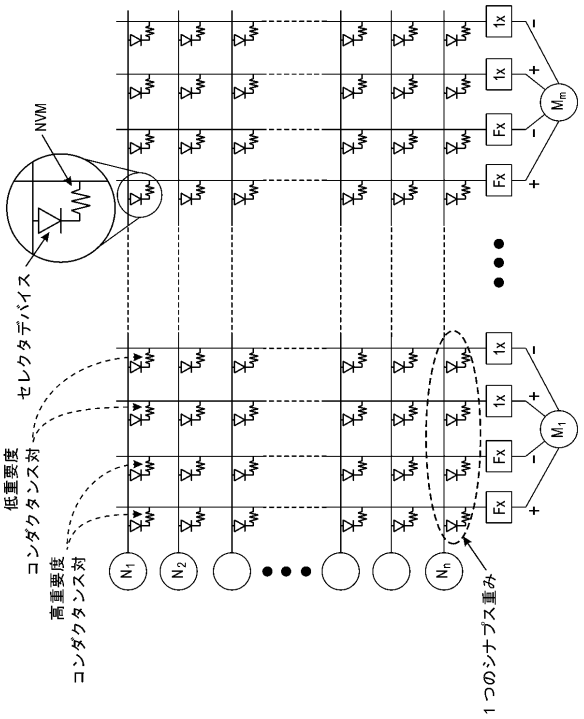
【図面】

【図１】

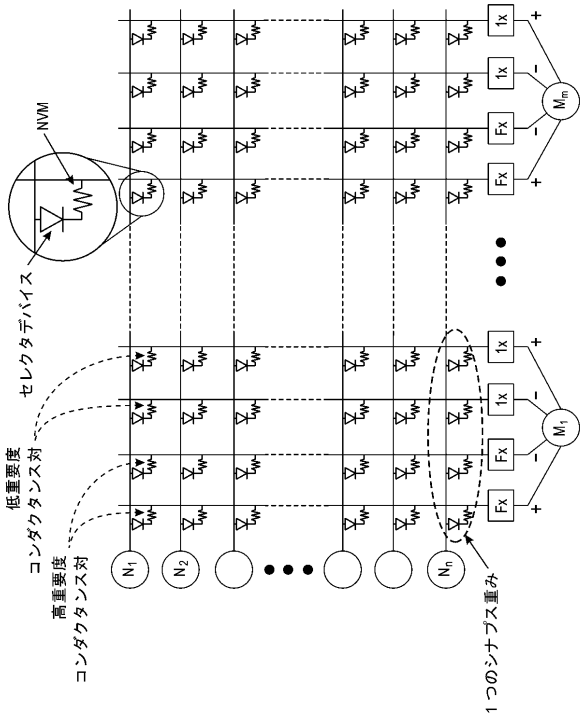


先行技術

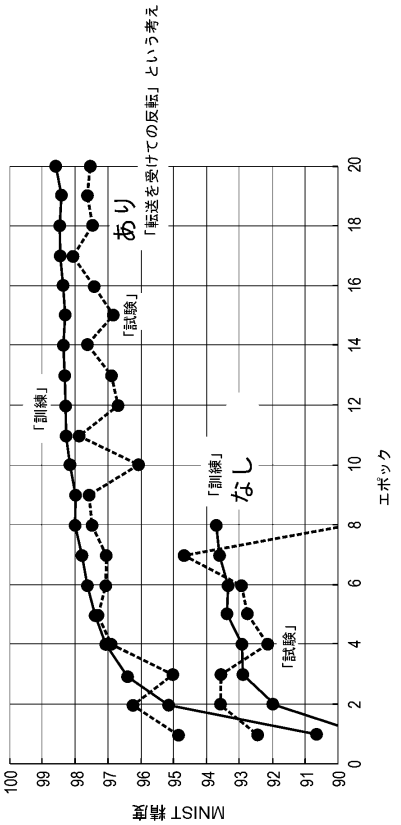
【図２】



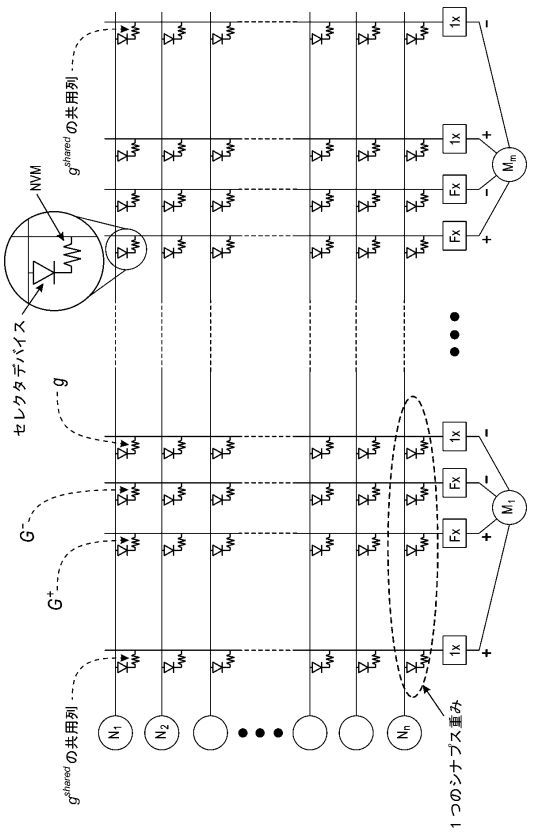
【図３】



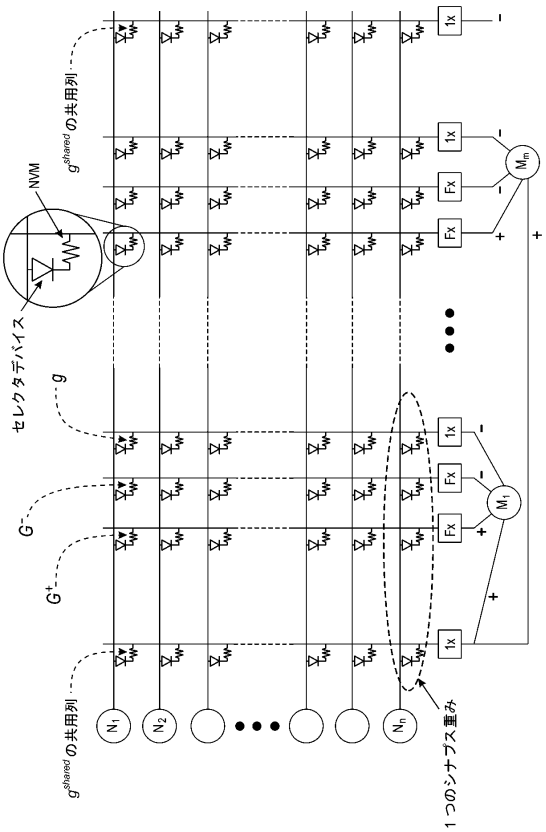
【図４】



【図５】


NVM
セレクタデバイス
g
G⁻
G⁺
gshared の共用列
1x
Fx
Mm
M1
N1
N2
Nn
1つのシナプス重み

【図６】


NVM
セレクタデバイス
g
G⁻
G⁺
gshared の共用列
1x
Fx
Mm
M1
N1
N2
Nn
1つのシナプス重み

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 ハリー・ロード６５０

審査官　石川 亮

(56)参考文献　特表２０１３－５３４６７７（ＪＰ，Ａ）
国際公開第２０１３／１０８２９９（WO，Ａ１）
米国特許第０９７１５６５６（ＵＳ，Ｂ１）
特表２０２０－５２８６１１（ＪＰ，Ａ）

(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｎ　３／０２－３／１０
Ｇ１１Ｃ　１１／５４
Ｇ０６Ｇ　７／６０
Ｇ１１Ｃ　１３／００