

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7684946号
(P7684946)

(45)発行日 令和7年5月28日(2025.5.28)

(24)登録日 令和7年5月20日(2025.5.20)

(51)国際特許分類		F I	
H 1 0 B	41/70 (2023.01)	H 1 0 B	41/70
H 1 0 B	12/00 (2023.01)	H 1 0 B	12/00 8 0 1
H 1 0 B	99/00 (2023.01)	H 1 0 B	12/00 6 2 1 Z
H 1 0 D	30/67 (2025.01)	H 1 0 B	12/00 6 7 1 C
H 1 0 D	30/68 (2025.01)	H 1 0 B	12/00 6 7 1 Z
請求項の数 8 (全60頁) 最終頁に続く			
(21)出願番号	特願2022-501383(P2022-501383)	(73)特許権者	000153878
(86)(22)出願日	令和3年2月8日(2021.2.8)		株式会社半導体エネルギー研究所
(86)国際出願番号	PCT/IB2021/050980		神奈川県厚木市長谷398番地
(87)国際公開番号	WO2021/165779	(72)発明者	岡本 佑樹
(87)国際公開日	令和3年8月26日(2021.8.26)		神奈川県厚木市長谷398番地 株式会
審査請求日	令和6年2月5日(2024.2.5)		社半導体エネルギー研究所内
(31)優先権主張番号	特願2020-27738(P2020-27738)	(72)発明者	上妻 宗広
(32)優先日	令和2年2月21日(2020.2.21)		神奈川県厚木市長谷398番地 株式会
(33)優先権主張国・地域又は機関			社半導体エネルギー研究所内
	日本国(JP)	(72)発明者	大貫 達也
(31)優先権主張番号	特願2020-36779(P2020-36779)		神奈川県厚木市長谷398番地 株式会
(32)優先日	令和2年3月4日(2020.3.4)		社半導体エネルギー研究所内
(33)優先権主張国・地域又は機関		審査官	宮本 博司
	日本国(JP)		
(31)優先権主張番号	特願2020-43999(P2020-43999)		
	最終頁に続く		最終頁に続く

(54)【発明の名称】 半導体装置

(57)【特許請求の範囲】

【請求項1】
複数のメモリ回路と、切替回路と、演算回路と、を有し、
複数の前記メモリ回路はそれぞれ、重みデータを保持する機能、および第1配線に前記重みデータを出力する機能を有し、
前記切替回路は、複数の前記第1配線のいずれかと、第2配線と、の導通状態を切り替える機能を有し、
前記演算回路は、入力データと、前記第2配線に与えられた前記重みデータと、を用いた演算処理を行う機能を有し、
複数の前記メモリ回路は、第1の層に設けられ、
前記切替回路および前記演算回路は、第2の層に設けられ、
前記第1の層は、前記第2の層とは異なる層であり、
前記重みデータは、第1のビット数のデータであり、
前記重みデータは、学習用データで最適化された第2のビット数の重みデータを変換して得られるデータであり、
前記第1のビット数は、前記第2のビット数より小さい、半導体装置。

【請求項2】
請求項1において、
前記第2配線は、基板表面に平行又は概略平行に設けられる配線を有する、半導体装置。

【請求項3】

請求項 1 又は請求項 2 において、

前記第 1 配線は、基板表面に概略垂直に設けられる配線を有する、半導体装置。

【請求項 4】

請求項 1 乃至請求項 3 のいずれかーにおいて、

前記第 1 の層は、第 1 トランジスタを有し、

前記第 1 トランジスタは、チャンネル形成領域に金属酸化物を有する半導体層を有する、半導体装置。

【請求項 5】

請求項 4 において、

前記金属酸化物は、I n と、G a と、Z n と、を含む、半導体装置。

10

【請求項 6】

請求項 1 乃至請求項 5 のいずれかーにおいて、

前記第 2 の層は、第 2 トランジスタを有し、

前記第 2 トランジスタは、チャンネル形成領域にシリコンを有する半導体層を有する、半導体装置。

【請求項 7】

請求項 1 乃至請求項 6 のいずれかーにおいて、

前記演算回路は、積和演算を行う回路である、半導体装置。

【請求項 8】

請求項 1 乃至請求項 7 のいずれかーにおいて、

前記第 1 の層は、前記第 2 の層の上に積層して設けられる、半導体装置。

20

【発明の詳細な説明】

【技術分野】

【0001】

本明細書は、半導体装置等について説明する。

【0002】

なお、本発明の一態様は、上記の技術分野に限定されない。本明細書等で開示する本発明の一態様の技術分野としては、半導体装置、撮像装置、表示装置、発光装置、蓄電装置、記憶装置、表示システム、電子機器、照明装置、入力装置、入出力装置、それらの駆動方法、又はそれらの製造方法、を一例として挙げることができる。

30

【背景技術】

【0003】

CPU (Central Processing Unit) 等を含む半導体装置を有する電子機器が普及している。このような電子機器では、大量のデータを高速に処理するため、半導体装置の性能向上に関する技術開発が活発である。高性能化を実現する技術としては、例えば、GPU (Graphics Processing Unit) 等のアクセラレータとCPUとを密結合させた、所謂SoC (System on Chip) 化がある。SoC化によって高性能化した半導体装置では、発熱、及び消費電力の増加が問題となってくる。

【0004】

40

AI (Artificial Intelligence) 技術では、計算量とパラメータ数が膨大になるため、演算量が増大する。演算量の増大は、発熱、および消費電力を増加させる要因となるため、演算量を低減するためのアーキテクチャが盛んに提案されている。代表的なアーキテクチャとして、Binary Neural Network (BNN)、およびTernary Neural Network (TNN) があり、回路規模縮小、および低消費電力化に対して特に有効となる（例えば特許文献1を参照）。

【先行技術文献】

【特許文献】

【0005】

【文献】国際公開第2019／078924号

50

【発明の概要】

【発明が解決しようとする課題】

【0006】

AI技術の演算では、重みデータと入力データを用いた積和演算を膨大な回数を繰り返すため、演算処理の高速化が求められる。メモリセルアレイでは、大量の重みデータや中間データを保持する必要がある。大量の重みデータや中間データを保持するメモリセルアレイでは、ビット線を介して演算回路に重みデータや中間データを読み出す。重みデータや中間データの読出しの頻度が多くなるため、メモリセルアレイと演算回路間のバンド幅が、動作速度の律速になることがある。

【0007】

メモリセルアレイと演算回路の間の配線の並列数を高めることで、高いバンド幅でメモリセルアレイと演算回路を接続することができるため、演算処理の高速化に有利となる。しかしながら、演算回路とメモリセルアレイの間の配線数が増えることになるため、周辺回路の面積が著しく増大する虞がある。

【0008】

またAI技術の演算では、ビット線の充放電エネルギーを如何にして低減するかが低消費電力化を図るうえで重要となる。

【0009】

ビット線の充放電エネルギーを低減するためには、ビット線を短くすることが有効である。しかしながら、演算回路とメモリセルアレイを交互に並べて配置することになるため、周辺回路の面積が著しく増大する虞がある。またビット線を短くすることを目的として、貼り合わせ技術などを用いて垂直方向にトランジスタを集積化する技術がある。しかしながら貼り合わせ技術では、電氣的に接続するための接続部の間隔が大きいため、却って寄生容量等が増えてしまい充放電エネルギーを低減できない虞がある。

【0010】

本発明の一態様は、小型化された半導体装置を提供することを課題の一とする。または、本発明の一態様は、低消費電力化された半導体装置を提供することを課題の一とする。または、本発明の一態様は、演算処理速度の向上が図られた半導体装置を提供することを課題の一とする。または、新規な構成の半導体装置を提供することを課題の一とする。

【0011】

なお、本発明の一態様は、必ずしも上記の課題の全てを解決する必要はなく、少なくとも一の課題を解決できるものであればよい。また、上記の課題の記載は、他の課題の存在を妨げるものではない。これら以外の課題は、明細書、特許請求の範囲、図面などの記載から、自ずと明らかとなるものであり、明細書、特許請求の範囲、図面などの記載から、これら以外の課題を抽出することが可能である。

【課題を解決するための手段】

【0012】

本発明の一態様は、複数のメモリ回路と、切替回路と、演算回路と、を有し、複数のメモリ回路はそれぞれ、重みデータを保持する機能を有し、切替回路は、メモリ回路のいずれか一と、演算回路と、の導通状態を切り替える機能を有し、複数のメモリ回路は、第1の層に設けられ、切替回路および演算回路は、第2の層に設けられ、第1の層は、第2の層とは異なる層である、半導体装置である。

【0013】

本発明の一態様は、複数のメモリ回路と、切替回路と、演算回路と、を有し、複数のメモリ回路はそれぞれ、重みデータを保持する機能、および第1配線に重みデータを出力する機能を有し、切替回路は、複数の第1配線のいずれか一と、演算回路と、の導通状態を切り替える機能を有し、複数のメモリ回路は、第1の層に設けられ、切替回路および演算回路は、第2の層に設けられ、第1の層は、第2の層とは異なる層である、半導体装置である。

【0014】

本発明の一態様は、複数のメモリ回路と、切替回路と、演算回路と、を有し、複数のメモリ回路はそれぞれ、重みデータを保持する機能、および第 1 配線に重みデータを出力する機能を有し、切替回路は、複数の第 1 配線のいずれかと、第 2 配線と、の導通状態を切り替える機能を有し、演算回路は、入力データと、第 2 配線に与えられた重みデータと、を用いた演算処理を行う機能を有し、複数のメモリ回路は、第 1 の層に設けられ、切替回路および演算回路は、第 2 の層に設けられ、第 1 の層は、第 2 の層とは異なる層である、半導体装置である。

【0015】

本発明の一態様において、第 2 配線は、基板表面に概略平行に設けられる配線を有する、半導体装置が好ましい。

10

【0016】

本発明の一態様において、第 1 配線は、基板表面に概略垂直に設けられる配線を有する、半導体装置が好ましい。

【0017】

本発明の一態様において、第 1 の層は、第 1 トランジスタを有し、第 1 トランジスタは、チャンネル形成領域に金属酸化物を有する半導体層を有する、半導体装置が好ましい。

【0018】

本発明の一態様において、金属酸化物は、In と、Ga と、Zn と、を含む、半導体装置が好ましい。

【0019】

本発明の一態様において、第 2 の層は、第 2 トランジスタを有し、第 2 トランジスタは、チャンネル形成領域にシリコンを有する半導体層を有する、半導体装置が好ましい。

20

【0020】

本発明の一態様において、演算回路は、積和演算を行う回路である、半導体装置が好ましい。

【0021】

本発明の一態様において、第 1 の層は、第 2 の層の上に積層して設けられる、半導体装置が好ましい。

【0022】

本発明の一態様において、重みデータは、第 1 のビット数のデータであり、重みデータは、学習用データで最適化された第 2 のビット数の重みデータを変換して得られるデータであり、第 1 のビット数は、第 2 のビット数より小さい、半導体装置が好ましい。

30

【0023】

なおその他の本発明の一態様については、以下で述べる実施の形態における説明、および図面に記載されている。

【発明の効果】

【0024】

本発明の一態様は、小型化された半導体装置を提供することができる。または、本発明の一態様は、低消費電力化された半導体装置を提供することができる。または、本発明の一態様は、演算処理速度の向上が図られた半導体装置を提供することができる。または、新規な構成の半導体装置を提供することができる。

40

【0025】

複数の効果の記載は、他の効果の存在を妨げるものではない。また、本発明の一形態は、必ずしも、例示した効果の全てを有する必要はない。また、本発明の一形態について、上記以外の課題、効果、および新規な特徴については、本明細書の記載および図面から自ずと明らかになるものである。

【図面の簡単な説明】

【0026】

図 1 A および図 1 B は、半導体装置の構成例を説明する図である。

図 2 A および図 2 B は、半導体装置の構成例を説明する図である。

50

図 3 A および図 3 B は、半導体装置の構成例を説明する図である。
図 4 は、半導体装置の構成例を説明する図である。
図 5 A および図 5 B は、半導体装置の構成例を説明する図である。
図 6 は、半導体装置の構成例を説明する図である。
図 7 A および図 7 B は、半導体装置の構成例を説明する図である。
図 8 A および図 8 B は、半導体装置の構成例を説明する図である。
図 9 A、図 9 B および図 9 C は、半導体装置の構成例を説明する図である。
図 10 は、半導体装置の構成例を説明する図である。
図 11 は、半導体装置の構成例を説明する図である。
図 12 A および図 12 B は、半導体装置の構成例を説明する図である。
図 13 A および図 13 B は、半導体装置の構成例を説明する図である。
図 14 A および図 14 B は、集積回路の構成例を示す図である。
図 15 は、トランジスタの構成例を示す図である。
図 16 は、演算処理システムの構成例を説明する図である。
図 17 は、C P U の構成例を説明する図である。
図 18 A および図 18 B は、C P U の構成例を説明する図である。
図 19 は、C P U の構成例を示す図である。
図 20 は、トランジスタの構成例を示す図である。
図 21 A および図 21 B は、トランジスタの構成例を示す図である。
図 22 A および図 22 B は、集積回路の構成例を説明する図である。
図 23 A および図 23 B は、集積回路の適用例を説明する図である。
図 24 A および図 24 B は、集積回路の適用例を説明する図である。
図 25 A、図 25 B および図 25 C は、集積回路の適用例を説明する図である。
図 26 は、集積回路の適用例を説明する図である。
図 27 A および図 27 B は、集積回路の適用例を説明する図である。
図 28 A および図 28 B は、重みデータを説明する図である。

10

20

【発明を実施するための形態】

【0027】

以下に、本発明の実施の形態を説明する。ただし、本発明の一形態は、以下の説明に限定されず、本発明の趣旨およびその範囲から逸脱することなくその形態および詳細を様々に変更し得ることは、当業者であれば容易に理解される。したがって、本発明の一形態は、以下に示す実施の形態の記載内容に限定して解釈されるものではない。

30

【0028】

なお本明細書等において、「第 1」、「第 2」、「第 3」という序数詞は、構成要素の混同を避けるために付したものである。従って、構成要素の数を限定するものではない。また、構成要素の順序を限定するものではない。また例えば、本明細書等の実施の形態の一において「第 1」に言及された構成要素が、他の実施の形態、あるいは特許請求の範囲において「第 2」に言及された構成要素とすることもありうる。また例えば、本明細書等の実施の形態の一において「第 1」に言及された構成要素を、他の実施の形態、あるいは特許請求の範囲において省略することもありうる。

40

【0029】

図面において、同一の要素または同様な機能を有する要素、同一の材質の要素、あるいは同時に形成される要素等には同一の符号を付す場合があり、その繰り返しの説明は省略する場合がある。

【0030】

本明細書において、例えば、電源電位 V D D を、電位 V D D、V D D 等と省略して記載する場合がある。これは、他の構成要素（例えば、信号、電圧、回路、素子、電極、配線等）についても同様である。

【0031】

また、複数の要素に同じ符号を用いる場合、特に、それらを区別する必要があるときに

50

は、符号に“__ 1 ”、“__ 2 ”、“[n] ”、“[m, n] ”等の識別用の符号を付記して記載する場合がある。例えば、２番目の配線 G L を配線 G L [2] と記載する。

【 0 0 3 2 】

(実施の形態 1)

本発明の一態様である半導体装置の構成、および動作等について説明する。

【 0 0 3 3 】

なお、本明細書等において半導体装置とは、半導体特性を利用することで機能し得る装置全般を指す。トランジスタなどの半導体素子をはじめ、半導体回路、演算装置、記憶装置は、半導体装置の一態様である。表示装置（液晶表示装置、発光表示装置など）、投影装置、照明装置、電気光学装置、蓄電装置、記憶装置、半導体回路、撮像装置、電子機器などは、半導体装置を有すると言える場合がある。

10

【 0 0 3 4 】

図 1 A は、本発明の一態様である半導体装置 1 0 を説明するための図である。

【 0 0 3 5 】

半導体装置 1 0 は、ホストプログラムから呼び出されたプログラム（カーネル、またはカーネルプログラムとも呼ばれる。）を実行する、アクセラレータとしての機能を有する。半導体装置 1 0 は、例えば、グラフィック処理における行列演算の並列処理、ニューラルネットワークの積和演算の並列処理、科学技術計算における浮動小数点演算の並列処理などを行うことができる。

【 0 0 3 6 】

半導体装置 1 0 は、メモリ回路部 2 0（メモリセルアレイともいう）、演算回路 3 0、および切替回路 4 0 を有する。演算回路 3 0 および切替回路 4 0 は、図中 x y 平面にトランジスタを有する層 1 1 に設けられる。メモリ回路部 2 0 は、図中 x y 平面にトランジスタを有する層 1 2 に設けられる。

20

【 0 0 3 7 】

層 1 1 は、チャンネル形成領域にシリコンを有するトランジスタ（S i トランジスタ）を有する。層 1 2 は、チャンネル形成領域に酸化物半導体を有するトランジスタ（O S トランジスタ）を有する。層 1 1 および層 1 2 は、x y 平面に対して概略垂直な方向（図 1 A 中、z 方向）で異なる層に設けられる。

【 0 0 3 8 】

あるいは層 1 2 は、S i トランジスタを有する構成とすることもできる。この場合、層 1 1 および層 1 2 は、貼り合わせ技術などを用いることで、x y 平面に対して概略垂直な方向（図 1 A 中、z 方向）で異なる層に設けることができる。貼り合わせ技術としては、プラズマ活性化接合技術、C u - C u ボンディング等による半導体基板を接合する技術などを用いることができる。

30

【 0 0 3 9 】

層 1 2 を O S トランジスタで構成する場合、メモリ回路部 2 0 は、S i トランジスタで構成することができる演算回路 3 0 および切替回路 4 0 と積層して設けることができる。つまりメモリ回路部 2 0 は、演算回路 3 0 および切替回路 4 0 が設けられる基板上に設けられる。そのため、回路面積の増加を招くことなく、メモリ回路部 2 0 を配置することができる。メモリ回路部 2 0 が設けられる領域を演算回路 3 0 および切替回路 4 0 が設けられる基板上とすることで、メモリ回路部 2 0 と、演算回路 3 0 及び切替回路 4 0 と、が同一層上に配置する場合と比較して、アクセラレータとして機能する半導体装置 1 0 における演算処理に必要な記憶容量を増やすことができる。記憶容量が増えることで、外部記憶装置から半導体装置への、演算処理に必要なデータの転送回数を削減することができるため、低消費電力化を図ることができる。

40

【 0 0 4 0 】

メモリ回路部 2 0 は、複数のメモリ回路部 2 0 __ 1 乃至 2 0 __ 4 を一例として図示している。各メモリ回路部は、複数のメモリ回路 2 1 を有する。複数のメモリ回路 2 1 は、メモリ回路部 2 0 __ 1 乃至 2 0 __ 4 のそれぞれにおいて、図 1 A に図示するように配線 L B

50

L₁乃至L₄（ローカルビット線、読出しビット線ともいう）を介して切替回路40に接続される。

【0041】

メモリ回路21は、NOSRAMの回路構成とすることができる。「NOSRAM（登録商標）」とは、「Nonvolatile Oxide Semiconductor RAM」の略称である。NOSRAMは、メモリセルが2トランジスタ型（2T）、又は3トランジスタ型（3T）ゲインセルであり、アクセストランジスタがOSTランジスタであるメモリのことをいう。メモリ回路21は、OSTランジスタで構成されるメモリである。メモリ回路21を有する層12は、演算回路30および切替回路40を有する層11上に積層して設けることができる。メモリ回路21を有するメモリ回路部20は、演算回路30および切替回路40を有する層11上に設けられるため、メモリ回路部20を有することによる面積オーバーヘッドを小さくすることが可能である。

10

【0042】

また、OSTランジスタはオフ状態でソースとドレインとの間を流れる電流、つまりリーク電流が極めて小さい。NOSRAMは、リーク電流が極めて小さい特性を用いてデータに応じた電荷をメモリ回路内に保持することで、不揮発性メモリとして用いることができる。特にNOSRAMは保持しているデータを破壊することなく読み出しすること（非破壊読み出し）が可能なため、データ読み出し動作を大量に繰り返す、ニューラルネットワークの積和演算の並列処理に適している。

【0043】

20

メモリ回路21は、NOSRAM、あるいはDOSRAMといったOSTランジスタを有するメモリ（以下、OSメモリともいう。）が好適である。酸化物半導体として機能する金属酸化物のバンドギャップは2.5 eV以上あるため、OSTランジスタは極小のオフ電流をもつ。一例として、ソースとドレイン間の電圧が3.5 V、室温（25℃）下において、チャネル幅1 μm当たりのオフ電流を 1×10^{-20} A未満、 1×10^{-22} A未満、あるいは 1×10^{-24} A未満とすることができる。そのため、OSメモリは、OSTランジスタを介して保持ノードからリークする電荷量が極めて少ない。従って、OSメモリは不揮発性のメモリ回路として機能できるため、半導体装置10のパワーゲーティングが可能となる。

【0044】

30

高密度でトランジスタが集積化された半導体装置は、回路の駆動による熱が発生する場合がある。この発熱により、トランジスタの温度が上がることで、当該トランジスタの特性が変化して、電界効果移動度の変化や動作周波数の低下などが起こることがある。OSTランジスタは、Siトランジスタよりも熱耐性が高いため、温度変化による電界効果移動度の変化が起こりにくく、また動作周波数の低下も起こりにくい。さらに、OSTランジスタは、温度が高くなっても、ドレイン電流がゲートソース間電圧に対して指数関数的に増大する特性を維持しやすい。そのため、OSTランジスタを用いることにより、高い温度環境下での安定した動作を行うことができる。

【0045】

40

OSTランジスタに適用される金属酸化物は、Zn酸化物、Zn-Sn酸化物、Ga-Sn酸化物、In-Ga酸化物、In-Zn酸化物、In-M-Zn酸化物（Mは、Ti、Ga、Y、Zr、La、Ce、Nd、SnまたはHf）などがある。特にMとしてGaを用いる金属酸化物をOSTランジスタに採用する場合、元素の比率を調整することで電界効果移動度等の電気特性に優れたトランジスタとすることができるため、好ましい。また、インジウムおよび亜鉛を含む酸化物に、アルミニウム、ガリウム、イットリウム、銅、バナジウム、ベリリウム、ホウ素、シリコン、チタン、鉄、ニッケル、ゲルマニウム、ジルコニウム、モリブデン、ランタン、セリウム、ネオジム、ハフニウム、タンタル、タングステン、マグネシウムなどから選ばれた一種、または複数種が含まれていてもよい。

【0046】

50

OSTランジスタの信頼性、電気特性の向上のため、半導体層に適用される金属酸化物

は、CAAC-OS、CAC-OS、nc-OSなどの結晶部を有する金属酸化物であることが好ましい。CAAC-OSとは、c-axis-aligned crystalline oxide semiconductorの略称である。CAC-OSとは、Cloud-Aligned Composite oxide semiconductorの略称である。nc-OSとは、nanocrystalline oxide semiconductorの略称である。

【0047】

CAAC-OSは、c軸配向性を有し、かつa-b面方向において複数のナノ結晶が連結し、歪みを有した結晶構造となっている。なお、歪みとは、複数のナノ結晶が連結する領域において、格子配列の揃った領域と、別の格子配列の揃った領域との間で格子配列の向きが変化している箇所を指す。

10

【0048】

CAC-OSは、キャリアとなる電子（または正孔）を流す機能と、キャリアとなる電子を流さない機能とを有する。電子を流す機能と、電子を流さない機能とを分離させることで、双方の機能を最大限に高めることができる。つまり、CAC-OSをOSトランジスタのチャネル形成領域に用いることで、高いオン電流と、極めて低いオフ電流との双方を実現できる。

【0049】

金属酸化物は、バンドギャップが大きく、電子が励起されにくいこと、ホールの有効質量が大きいことなどから、OSトランジスタは、一般的なSiトランジスタと比較して、アバランシェ崩壊等が生じにくい場合がある。従って、例えばアバランシェ崩壊に起因するホットキャリア劣化等を抑制できる。ホットキャリア劣化を抑制できることで、高いドレイン電圧でOSトランジスタを駆動することができる。

20

【0050】

OSトランジスタは、電子を多数キャリアとする蓄積型トランジスタである。そのため、pn接合を有する反転型トランジスタ（代表的には、Siトランジスタ）と比較して短チャネル効果の一つであるDIBL（Drain-Induced Barrier Lowering）の影響が小さい。つまり、OSトランジスタは、Siトランジスタよりも短チャネル効果に対する高い耐性を有する。

【0051】

OSトランジスタは、短チャネル効果に対する耐性が高いために、OSトランジスタの信頼性を劣化させずに、チャネル長を縮小できるので、OSトランジスタを用いることで回路の集積度を高めることができる。チャネル長が微細化するのに伴いドレイン電界が強まるが、上掲したように、OSトランジスタはSiトランジスタよりもアバランシェ崩壊が起きにくい。

30

【0052】

また、OSトランジスタは、短チャネル効果に対する耐性が高いために、Siトランジスタよりもゲート絶縁膜を厚くすることが可能となる。例えば、チャネル長及びチャネル幅が50nm以下の微細なトランジスタにおいても、10nm程度の厚いゲート絶縁膜を設けることが可能な場合がある。ゲート絶縁膜を厚くすることで、寄生容量を低減することができるので、回路の動作速度を向上できる。またゲート絶縁膜を厚くすることで、ゲート絶縁膜を介したリーク電流が低減されるため、静的消費電流の低減につながる。

40

【0053】

以上より、半導体装置10は、OSメモリであるメモリ回路21を有することで電源電圧の供給が停止してもデータを保持できる。そのため、半導体装置10のパワーゲーティングが可能となり、消費電力の大幅な低減を図ることができる。

【0054】

メモリ回路21が記憶するデータは、ニューラルネットワークの積和演算に用いられる重みパラメータに対応するデータ（重みデータ）である。重みデータは、デジタルデータとすることで、ノイズに強く、高速で演算可能な半導体装置とすることができる。また、

50

重みデータは、アナログデータでもよい。NOSRAMはアナログ値の電位を保持することができるため、当該データをデジタルデータと適宜変換して用いる構成とすることができる。アナログデータを保持可能なメモリ回路21は、高いビット数の重みデータを表す場合、メモリ回路を増やすことなく保持することができる。

【0055】

切替回路40の一例として図示する切替回路40__1乃至40__4は、複数のメモリ回路部20__1乃至20__4のそれぞれから延びる配線LBL__1乃至LBL__4の電位を選択して、配線GBL（グローバルビット線ともいう）に伝える機能を有する。配線GBLは切替回路40__1乃至40__4の出力端子が接続される。切替回路40は、選択された切替回路40と非選択の切替回路40の出力電位が同時に供給されて貫通電流が発生することを防ぐ必要がある。切替回路40は、例えば制御信号で出力電位の状態が制御されるスリーステートバッファを用いることができる。この構成例では、配線GBLは選択された切替回路が入力電位をバッファ出力し、非選択の切替回路の出力がハイインピーダンスとなるため、出力電位が同時に供給されることを回避できる。なお切替回路40は、Siトランジスタで構成されることが好ましい。当該構成とすることで高速で接続状態の切り替えを行う構成とすることができる。

【0056】

演算回路30の一例として図示する演算回路30__1乃至30__4は、積和演算といった同じ処理を繰り返し実行する機能を有する。演算回路30での積和演算のために入力される入力データおよび重みデータは、デジタルデータが好ましい。デジタルデータはノイズの影響を受けにくい。そのため演算回路30は、高い精度の演算結果が要求される演算処理を行うのに適している。なお演算回路30は、Siトランジスタで構成されることが好ましい。当該構成とすることでOSTランジスタと積層して設けることができる。

【0057】

演算回路30__1乃至30__4は、配線LBL__1乃至LBL__4および配線GBLを介して、メモリ回路21に保持された重みデータが与えられる。また、演算回路30__1乃至30__4は、外部から入力される入力データ（A₁、A₂、A₃、A₄）が与えられる。演算回路30__1乃至30__4は、メモリ回路21に保持された重みデータおよび外部から入力される入力データを用いて、積和演算の演算処理が行われる。

【0058】

演算回路30__1乃至30__4に与えられる重みデータは、複数のメモリ回路部20__1乃至20__4で選択された重みデータが、切替回路40__1乃至40__4で切り替えて配線GBLを介して与えられる重みデータである。つまり演算回路30__1乃至30__4では、同じ重みデータを用いた演算処理、例えば積和演算を行うことができる。そのため、本発明の一態様における半導体装置10は、畳み込みニューラルネットワークのように、同じ重みデータを用いた処理を効率的に行うことができる。

【0059】

また演算回路30__1乃至30__4に与えられる重みデータは、予め配線LBL__1乃至LBL__4に与えられていたデータを切替回路40__1乃至40__4で切り替えることで配線GBLに与えることができるため、配線GBLに与える重みデータは、Siトランジスタの電気特性に準ずる速度で切り替えることができる。そのため、メモリ回路部20__1乃至20__4から配線LBL__1乃至LBL__4に重みデータを読み出すための期間が長い場合であっても、予め、重みデータを配線LBL__1乃至LBL__4に読み出しておくことで、重みデータを高速で切り替えて演算処理することができる。

【0060】

なおメモリ回路部20から切替回路40に向けて延びる配線LBLは、図1Bに図示するように重みデータW_{data}を層12から層11に伝えるための配線となる。メモリ回路21から配線LBLへ重みデータW_{data}を高速に読み出すために、配線LBLは、短くすることが好ましい。また、配線LBLは、充放電に伴う消費エネルギーを小さくするために、短くすることが好ましい。つまり切替回路40は、z方向に延びて設けられる配線

10

20

30

40

50

L B L（図中、z 方向に延びる矢印）の近くになるよう、層 1 1 の x y 平面で分散して配置する構成とすることが好ましい。

【0061】

なお演算回路 3 0 __ 1 乃至 3 0 __ 4 は、メモリ回路 2 1 の読出し用のビット線である配線 L B L __ 1 乃至 L B L __ 4 毎、つまり一列（Column）毎に演算回路 3 0 __ 1 乃至 3 0 __ 4 を設ける構成とする（Column-Parallel Calculation）ことができる。当該構成とすることで、配線 L B L の列数分のデータを並列で演算処理することができる。C P U あるいは G P U を用いた積和演算に比べて、データバスサイズ（32ビット、など）に制限されないことから、Column-Parallel Calculation では、演算の並列度を大幅に上げることができるため、A I 技術であるディープニューラルネットワークの学習（深層学習）、浮動小数点演算を行う科学技術計算などの膨大な演算処理に係る演算効率の向上を図ることができる。加えて演算回路 3 0 から出力されるデータの演算を完了させて読み出すことができるため、メモリアクセス（演算回路とメモリ間のデータ転送など）で生じる電力を削減することができ、発熱および消費電力の増加を抑制することができる。さらに、演算回路 3 0 とメモリ回路部 2 0 の物理的な距離を近づけること、例えば積層によって配線距離が短くできることで、信号線に生じる寄生容量を削減できるため、低消費電力化が可能である。

10

【0062】

次いで図 2 A では、A I アクセラレータとして機能する半導体装置 1 0 を含む演算処理システム 1 0 0 の全体を示すブロック図について説明する。

20

【0063】

図 2 A では、図 1 A、図 1 B で説明した半導体装置 1 0 の他、C P U 1 1 0 およびバス 1 2 0 を図示している。C P U 1 1 0 は、C P U コア 2 0 0 およびバックアップ回路 2 2 2 を有する。アクセラレータとして機能する半導体装置 1 0 は、駆動回路 5 0、メモリ回路部 2 0 __ 1 乃至 2 0 __ N（N は 2 以上の自然数）、メモリ回路 2 1、切替回路 4 0、および演算回路 3 0 __ 1 乃至 3 0 __ N を図示している。

【0064】

C P U 1 1 0 は、オペレーティングシステムの実行、データの制御、各種演算やプログラムの実行など、汎用の処理を行う機能を有する。C P U 1 1 0 は、C P U コア 2 0 0 を有する。C P U コア 2 0 0 は、1 つまたは複数の C P U コアに相当する。また C P U 1 1 0 は、電源電圧の供給が停止しても C P U コア 2 0 0 内のデータを保持できるバックアップ回路 2 2 2 を有する。電源電圧の供給は、電源ドメイン（パワードメイン）からのパワースイッチ等による電氣的な切り離しによって制御することができる。なお電源電圧は、駆動電圧という場合がある。バックアップ回路 2 2 2 として、例えば、O S トランジスタを有する O S メモリが好適である。

30

【0065】

O S トランジスタで構成されるバックアップ回路 2 2 2 は、S i トランジスタで構成することができる C P U コア 2 0 0 と積層して設けることができる。バックアップ回路 2 2 2 の面積は C P U コア 2 0 0 の面積より小さいため、回路面積の増加を招くことなく、C P U コア 2 0 0 上にバックアップ回路 2 2 2 を配置することができる。バックアップ回路 2 2 2 は、C P U コア 2 0 0 が有するレジスタのデータを保持する機能を有する。バックアップ回路 2 2 2 は、データ保持回路ともいう。なお O S トランジスタを有するバックアップ回路 2 2 2 を備えた C P U コア 2 0 0 の構成の詳細については、実施の形態 4 でも説明する。

40

【0066】

メモリ回路部 2 0 __ 1 乃至 2 0 __ N は、それぞれメモリ回路 2 1 に保持された重みデータ W_1 乃至 W_N を、配線 L B L（図示せず）を介して切替回路 4 0 に出力する。切替回路 4 0 は、選択された重みデータを、配線 G B L（図示せず）を介して重みデータ W_{SEL} として各演算回路 3 0 __ 1 乃至 3 0 __ N に出力する。駆動回路 5 0 は、入力データ線を介して演算回路 3 0 __ 1 乃至 3 0 __ N に入力データ A_1 乃至 A_N を出力する。

50

【0067】

駆動回路50は、メモリ回路部20__1乃至20__Nにおける重みデータの書き込みおよび読み出しを制御するための信号を出力する機能を有する。また駆動回路50は、演算回路30__1乃至30__Nに入力データを与えてニューラルネットワークの積和演算等を実行させるための回路、およびニューラルネットワークの積和演算等で得られる出力データを保持する、などの機能を有する。

【0068】

バス120は、CPU110と半導体装置10とを電氣的に接続する。つまりCPU110と半導体装置10とは、バス120を介してデータ伝送を行うことができる。

【0069】

図2Bでは、図2Aに図示する半導体装置10において、Nを6とした場合の各構成の位置関係を説明するための図である。

【0070】

OSトランジスタで構成されるメモリ回路部20__1乃至20__6と、演算回路30__1乃至30__Nとは、駆動回路50、切替回路40および演算回路30__1乃至30__6が設けられる基板表面に対して概略垂直な方向に延在して設けられる配線LBL__1乃至LBL__6を介して電氣的に接続される。なお「概略垂直」とは、85度以上95度以下の角度で配置されている状態をいう。なお本明細書において図2B等に図示するX方向、Y方向、およびZ方向は、それぞれが互いに直交または交差する方向である。また、X方向およびY方向は基板表面に対して平行または概略平行であり、Z方向は基板表面に対して垂直または概略垂直である。

【0071】

メモリ回路部20__1乃至20__6はそれぞれ、メモリ回路21を有する。メモリ回路部20__1乃至20__6は、デバイスメモリ、共有メモリという場合がある。メモリ回路21は、トランジスタ22を有する。トランジスタ22が有する半導体層23は、酸化物半導体（金属酸化物）とすることで、上述したOSトランジスタで構成されるメモリ回路21とすることができる。

【0072】

メモリ回路部20__1乃至20__6が有する複数のメモリ回路21はそれぞれ、配線LBL__1乃至LBL__6に接続される。配線LBL__1乃至LBL__6は、Siトランジスタが設けられる基板表面に概略垂直、つまりz方向に伸びる配線を経由して、切替回路40に接続される。切替回路40は、配線LBL__1乃至LBL__6のいずれか一の電位を増幅して配線GBLに伝える構成とする。配線GBLは、Siトランジスタが設けられる基板表面に概略平行、つまりxy平面に伸びる配線となる。当該構成とすることで、切替回路40を制御することで配線GBLに与える重みデータを高速で切り替えることができる。

【0073】

演算回路30__1乃至30__6は、配線GBLを介して入力される重みデータと、駆動回路50から入力データ線を介して与えられる入力データA_{IN}と、に基づいて演算を行う。重みデータを保持するメモリ回路部20__1乃至20__6は、上層に配置することができるため、演算回路30__1乃至30__6を効率的に配置することができる。そのため、駆動回路50から延びる入力データ線を短くすることができ、半導体装置10の低消費電力化および高速化を図ることができる。

【0074】

次いで図2Bの構成とすることによる利点について説明する。図3Aは、説明のため、図2Bの各構成をブロック図で示したものである。なお6個のメモリ回路部20__1乃至20__6にあるメモリ回路21から重みデータW₁乃至W₆が配線LBL__1乃至LBL__6に読み出されるとして説明する。また切替回路40は、配線LBL__1乃至LBL__6に接続される切替回路40__1乃至40__6として説明する。また切替回路40で重みデータW₁乃至W₆から選択され、配線GBLに与えられる重みデータを重みデータW_{SE}

10

20

30

40

50

Lとして説明する。演算回路30__1乃至30__6にはそれぞれ入力データA₁乃至A₆が与えられ、出力データMAC₁乃至MAC₆を得るものとして説明する。

【0075】

配線LBL__1乃至LBL__6における上層と下層をつなぐ垂直方向（図2B参照）に延びる配線LBL_pは、水平方向に延びる配線と比べて短い。そのため、配線LBL__1乃至LBL__6の寄生容量を小さくでき、配線の充放電に要する電荷を削減でき、低消費電力化および演算効率の向上を図ることができる。また、メモリ回路21から配線LBL__1乃至LBL__6への読み出しを高速にできる。

【0076】

配線GBLを介して、演算回路30__1乃至30__6では同じ重みデータを用いた演算処理を行うことができる。当該構成は、同じ重みデータを用いた演算処理を行う畳み込みニューラルネットワークの演算処理に適している。

10

【0077】

図3Bは、図3Aに図示する切替回路40に適用可能な回路構成の一例である。図3Bに図示するスリーステートバッファは、配線LBLの電位を制御信号ENに応じて配線GBLに増幅して伝える機能を有する。切替回路40は、マルチプレクサと見做すことができる。複数の入力信号から、1つを選択する機能を有する。

【0078】

なお図3Aでは、切替回路40が複数の配線LBLから1つの配線を選択して重みデータWSELを配線GBLに与える構成について図示したが、他の構成でもよい。例えば図4に図示するように切替回路として切替回路40Aおよび切替回路40Bを設ける構成としてもよい。

20

【0079】

切替回路40Aは、切替回路40__1乃至40__12を有する。切替回路40Aが有する構成は、切替回路40と同様である。切替回路40__1乃至40__6と、切替回路40__7乃至40__12と、は、離れた位置に配置してもよい。切替回路40Aは、配線LBL__1乃至LBL__6からいずれかーを選択して重みデータW₁乃至W₆から選択される重みデータWSEL__Aを配線GBL__Aに与える。また切替回路40Aは、配線LBL__7乃至LBL__12からいずれかーを選択して重みデータW₇乃至W₁₂から選択される重みデータWSEL__Bを配線GBL__Bに与える。

30

【0080】

切替回路40Bは、切替回路40X乃至40Yを有する。切替回路40Bが有する構成は、切替回路40と同様である。切替回路40Bは、配線GBL__Aまたは配線GBL__Bを選択して重みデータWSEL__Aまたは重みデータWSEL__Bから選択される重みデータWSELを配線GBLに与える。配線GBLを介して、演算回路30__1乃至30__6、演算回路30__7乃至30__12では、それぞれ同じ重みデータを用いた演算処理を行うことができる。当該構成は、同じ重みデータを用いた演算処理を行う畳み込みニューラルネットワークの演算処理に適している。

【0081】

また図3Aでは、各メモリ回路21が1ビットのデータ（つまり‘1’か‘0’のデータ）を保持し、当該データを用いて演算処理を行う構成として説明したが、多ビットのデータを用いて演算処理を行う構成にも本発明の一態様は適用可能である。当該構成について図3Aと同様に図5Aに図示する。多ビット（例えばnビット）のデータの場合、図5Aに図示するように、ビット数に応じた本数の配線LBL__1乃至LBL__nに接続された切替回路40Mを用いて、配線GBLに与える多ビットの重みデータを選択する構成とすればよい。なお多ビットの重みデータがアナログ値の場合、切替回路40Mは、アナログスイッチ（トランスファークラップ）などで構成することができる。

40

【0082】

メモリ回路部20と演算回路30が別チップの場合、チップのピン数に従い、バス幅が制限される。一方で、本発明の一態様の構成のようにメモリ回路部20と演算回路30と

50

を積層する構成では、配線 L B L を設ける開口に応じて演算処理に必要なデータの並列数を高めることができるため、効率的な演算処理を行うことが可能である。

【0083】

図 5 B は、図 5 A に図示する切替回路 40 M に適用可能な回路構成の一例である。図 5 B に図示するスリーステートバッファは、n 本の配線 L B L の電位を n 本の制御信号 E N に応じて n 本の配線 G B L に増幅して伝える機能を有する。

【0084】

図 6 では、図 3 A で説明した構成の動作を説明するためのタイミングチャートを示す。半導体装置 10 では、クロック信号 C L K のトグル動作（例えば時刻 T 1 乃至 T 7）に応じて演算処理を行う。クロック信号 C L K の周波数を高める構成とすることで、演算処理の高速化を図ることができる。なお図 6 中、W_a 乃至 W_f、W₁ 乃至 W₁₇ は、重みデータである。

【0085】

入力データ A₁ 乃至 A₆ をそれぞれ図示するように A_{1a} 乃至 A₁₁₁、A_{2a} 乃至 A₂₁₁、A_{3a} 乃至 A₃₁₁、A_{4a} 乃至 A₄₁₁、A_{5a} 乃至 A₅₁₁、A_{6a} 乃至 A₆₁₁ とクロック信号 C L K に応じて高速で切り替える場合、重みデータを与える配線 G B L のデータを高速で切り替える必要がある。

【0086】

本発明の一態様の構成では、切替回路 40 で配線 L B L から配線 G B L に選択される重みデータをあらかじめ、配線 L B L₁ 乃至 L B L₆ に読み出ししておく構成とすることで、重みデータを与える配線 G B L のデータを高速で切り替えることができる。例えば時刻 T 1 で配線 L B L₁ に重みデータ W₁ を読み出しおき、時刻 T 6 で切替回路 40 を切り替えて配線 L B L₁ から配線 G B L に重みデータ W₁ を出力する構成とすることができる。時刻 T 2 乃至 T 7、および時刻 T 7 以降においても、配線 L B L への重みデータの読み出しと、配線 G B L での重みデータの選択と、の時刻を異ならせることで、クロック信号 C L K に応じた重みデータの切り替えを行う構成とすることができる。

【0087】

図 7 A では、演算回路の具体的な構成例を示す。図 7 A では、8 ビットの重みデータと 8 ビットの入力データとの積和演算を行うことができる演算回路 30 の構成例を図示している。図 7 A において、乗算回路 24、加算回路 25 およびレジスタ 26 を図示している。乗算回路 24 で乗算された 16 ビットのデータは加算回路 25 に入力する。加算回路 25 の出力がレジスタ 26 に保持され、乗算回路 24 で乗算されるデータと加算回路 25 で足しあわされることで積和演算が行われる。レジスタは、クロック信号 C L K およびリセット信号 r e s e t_B によって制御される。なお図中「17 + α」における「α」は、乗算データを加算することで生じる桁上りを示したものである。当該構成とすることで重みデータ W_{S E L} と入力データ A_{I N} との積和演算に相当する出力データ M A C を得ることができる。

【0088】

また図 7 A では、8 ビットのデータを用いた演算処理を行う構成として説明したが、1 ビットのデータを用いた構成にも本発明の一態様は適用可能である。当該構成について図 7 A と同様に図 7 B に図示する。1 ビットのデータの場合、図 7 B に図示するように、ビット数に応じた演算処理を行えばよい。

【0089】

図 8 A は、本発明の半導体装置 10 が有するメモリ回路部 20 に適用可能な回路構成例について説明する図である。図 8 A では、M 行 N 列（M、N は 2 以上の自然数）行列方向に並べて配置された書込用ワード線 W W L₁ 乃至 W W L_M、読出用ワード線 R W L₁ 乃至 R W L_M、書込用ビット線 W B L₁ 乃至 W B L_N、および配線 L B L₁ 乃至 L B L_N を図示している。また各ワード線およびビット線に接続されたメモリ回路 21 を図示している。

【0090】

10

20

30

40

50

図 8 B は、メモリ回路 2 1 に適用可能な回路構成例について説明する図である。メモリ回路 2 1 は、トランジスタ 6 1、トランジスタ 6 2、トランジスタ 6 3、容量素子 6 4（キャパシタともいう）を有する。

【0091】

トランジスタ 6 1 のソースまたはドレインの一方は、書込み用ビット線 W B L に接続される。トランジスタ 6 1 のゲートは、書込み用ワード線 W W L に接続される。トランジスタ 6 1 のソースまたはドレインの他方は、容量素子 6 4 の一方の電極およびトランジスタ 6 2 のゲートに接続される。トランジスタ 6 2 のソースまたはドレインの一方および容量素子 6 4 の他方の電極は、固定電位たとえばグラウンド電位を与える配線に接続される。トランジスタ 6 2 のソースまたはドレインの他方は、トランジスタ 6 3 のソースまたはドレインの一方に接続される。トランジスタ 6 3 のゲートは、読出し用ワード線 R W L に接続される。トランジスタ 6 3 のソースまたはドレインの他方は、配線 L B L に接続される。配線 L B L は切替回路 4 0 を介して配線 G B L に接続される。配線 L B L は、上述したように、演算回路 3 0 が設けられる基板表面に対して概略垂直な方向に延在して設けられる配線を介して切替回路 4 0 に接続される。

【0092】

図 8 B に示すメモリ回路 2 1 の回路構成は、3 トランジスタ型（3 T）ゲインセルの N O S R A M に相当する。トランジスタ 6 1 乃至トランジスタ 6 3 は、O S トランジスタである。O S トランジスタはオフ状態でソースとドレインとの間を流れる電流、つまりリーク電流が極めて小さい。N O S R A M は、リーク電流が極めて小さい特性を用いてデータに応じた電荷をメモリ回路内に保持することで、不揮発性メモリとして用いることができる。なお図 8 B に示すトランジスタ 6 1 を S i トランジスタとする場合、オフ状態でソースとドレインとの間を流れる電流、つまりリーク電流が極めて小さくなるよう設計する。例えば、チャネル長をチャネル幅に対して十分長くなるよう設計する。

【0093】

図 8 A のメモリ回路 2 1 に適用可能な回路構成は、図 8 B の 3 T 型の N O S R A M に限らない。例えば、図 9 A に図示する D O S R A M に相当する回路でもよい。図 9 A では、トランジスタ 6 1 A および容量素子 6 4 A を有するメモリ回路 2 1 A を図示している。トランジスタ 6 1 A は、O S トランジスタである。メモリ回路 2 1 A は、ビット線 B L、ワード線 W L およびバックゲート線 B G L に接続される例を図示している。

【0094】

図 8 A のメモリ回路 2 1 に適用可能な回路構成は、図 9 B に図示する 2 T 型の N O S R A M に相当する回路でもよい。図 9 B では、トランジスタ 6 1 B、トランジスタ 6 2 B および容量素子 6 4 B を有するメモリ回路 2 1 B を図示している。トランジスタ 6 1 B およびトランジスタ 6 2 B は、O S トランジスタである。トランジスタ 6 1 B およびトランジスタ 6 2 B は、異なる層に半導体層が配置される O S トランジスタもよいし、同じ層に半導体層が配置される O S トランジスタでもよい。メモリ回路 2 1 B は、書込み用ビット線 W B L、読出し用ビット線として機能する配線 L B L、書込み用ワード線 W W L、読出し用ワード線 R W L、ソース線 S L およびバックゲート線 B G L に接続される例を図示している。

【0095】

図 8 A のメモリ回路 2 1 に適用可能な回路構成は、図 9 C に図示する 3 T 型の N O S R A M を組み合わせた回路でもよい。図 9 C では、論理の異なるデータを保持できるメモリ回路 2 1 __ P と、メモリ回路 2 1 __ N と、を有するメモリ回路 2 1 C を図示している。図 9 C では、トランジスタ 6 1 __ P、トランジスタ 6 2 __ P、トランジスタ 6 3 __ P および容量素子 6 4 __ P を有するメモリ回路 2 1 __ P と、トランジスタ 6 1 __ N、トランジスタ 6 2 __ N、トランジスタ 6 3 __ N および容量素子 6 4 __ N を有するメモリ回路 2 1 __ N と、を図示している。メモリ回路 2 1 __ P およびメモリ回路 2 1 __ N が有する各トランジスタは、O S トランジスタである。メモリ回路 2 1 __ P およびメモリ回路 2 1 __ N が有する各トランジスタは、異なる層に半導体層が配置される O S トランジスタもよいし、同じ層

に半導体層が配置されるＯＳトランジスタでもよい。メモリ回路２１Ｃは、書込み用ビット線ＷＢＬ＿Ｐ、配線ＬＢＬ＿Ｐ、書込み用ビット線ＷＢＬ＿Ｎ、配線ＬＢＬ＿Ｎ、書込み用ワード線ＷＷＬ、読出し用ワード線ＲＷＬに接続される例を図示している。メモリ回路２１Ｃは、論理の異なるデータを保持し、論理の異なるデータを配線ＬＢＬ＿Ｐおよび配線ＬＢＬ＿Ｎに読出し、図３などと同様に、切替回路４０を介して配線ＧＢＬに出力することができる。

【００９６】

なお図９Ｃの構成において、メモリ回路２１＿Ｐと、メモリ回路２１＿Ｎとに保持するデータの乗算に相当するデータが配線ＬＢＬに出力されるように排他的論理和回路（ＸＯＲ回路）を設けてもよい。当該構成とすることで、演算回路３０における乗算に相当する演算を省略できるため、低消費電力化を図ることができる。

10

【００９７】

図１０には、畳み込みニューラルネットワークの演算処理の流れを図示する。図１０では、入力層９０Ａ、中間層９０Ｂ（隠れ層ともいう）、出力層９０Ｃを図示している。入力層９０Ａでは、入力データの入力処理９１（図中、Ｉｎｐｕｔと図示）を図示している。中間層９０Ｂでは、畳み込み演算処理９２、９３、９５（図中、Ｃｏｎｖ．と図示）、複数のプーリング演算処理９４、９６（図中、Ｐｏｏｌ．と図示）を図示している。出力層９０Ｃでは、全結合演算処理９７（図中、Ｆｕｌｌと図示）を図示している。入力層９０Ａ、中間層９０Ｂ、出力層９０Ｃにおける演算処理の流れは一例であり、実際の畳み込みニューラルネットワークの演算処理では、ソフトマックス演算などの他の演算処理を行うことがあり得る。

20

【００９８】

図１０に図示する畳み込みニューラルネットワークでは、複数回の畳み込み演算処理９２、９３、９５を行う。畳み込み演算処理では、同じ重みデータを用いた演算処理を行う。そのため、同じ重みデータを用いる演算処理を行う本実施の一態様の構成を適用することで動作速度と、低消費電力化との両立を図ることができる。

【００９９】

次に、半導体装置１０の詳細なブロック図について図１１に示す。

【０１００】

図１１では、図１Ａおよび図１Ｂ、並びに図２Ａおよび図２Ｂで説明した、メモリ回路部２０、メモリ回路２１、演算回路３０、切替回路４０、層１１、層１２に相当する構成の他、図２Ａおよび図２Ｂで図示する駆動回路５０の構成例について図示している。

30

【０１０１】

図１１では、図２Ａおよび図２Ｂで説明した駆動回路５０に対応する構成として、コントローラ７１、ロウデコーダ７２、ワード線ドライバ７３、カラムデコーダ７４、書き込みドライバ７５、プリチャージ回路７６、入出力バッファ８１および演算制御回路８２を図示している。

【０１０２】

図１２Ａは、図１１に図示する各構成について、メモリ回路部２０を制御するブロックを抜き出した図である。図１２Ａでは、コントローラ７１、ロウデコーダ７２、ワード線ドライバ７３、カラムデコーダ７４、書き込みドライバ７５、プリチャージ回路７６を抜き出して図示している。

40

【０１０３】

コントローラ７１は、外部からの入力信号を処理して、ロウデコーダ７２およびカラムデコーダ７４の制御信号を生成する。外部からの入力信号は、書き込みイネーブル信号や読み出しイネーブル信号などのメモリ回路部２０を制御するための制御信号である。またコントローラ７１は、ＣＰＵ１１０と半導体装置１０の間でバス１２０を介してデータの入出力が行われる。

【０１０４】

ロウデコーダ７２は、ワード線ドライバ７３を駆動するための信号を生成する。ワード

50

線ドライバ73は、書込み用ワード線WWL、および読出し用ワード線RWLに与える信号を生成する。カラムデコーダ74は、書き込みドライバ75を駆動するための信号を生成する。書き込みドライバ75は、メモリ回路21に与える重みデータを生成する。プリチャージ回路76は、配線LBLなどをプリチャージする機能を有する。メモリ回路部20のメモリ回路21から読み出される重みデータに応じた信号は、図2Aおよび図2B等で説明したように、配線LBLを介して切替回路40に入力される。

【0105】

図12Bは、図11に図示する各構成について、演算回路30および切替回路40を制御するブロックを抜き出した図である。

【0106】

コントローラ71は、外部からの入力信号を処理して、演算制御回路82の制御信号を生成する。またコントローラ71は、演算回路30を制御するためのアドレス信号、およびクロック信号などの各種信号を生成する。演算制御回路82は、コントローラ71の制御および入出力バッファ81の出力に応じて、データ入力線に与えられる入力データA₁乃至A_Nを生成する。演算制御回路82は、切替回路40を制御する制御信号を出力する。切替回路40は、図2Aおよび図2B等で説明したように、複数の配線LBLの与えられる重みデータのいずれかを、配線GBLを介して複数の演算回路30に与える。演算回路30は、与えられる重みデータおよび入力データを切り替えることで、積和演算に応じた出力データMACを生成する。生成された出力データMACは、中間データとして入出力バッファ81を介して演算制御回路82内のSRAMあるいはレジスタなどのメモリに一時的に保持される。保持された中間データは、演算回路30に再入力される。

【0107】

なお本発明の一態様における半導体装置10は、並列数が高められた並列計算を可能にするため、複数組み合わせる構成が好ましい。この場合の構成例について図13A、図13Bを用いて説明する。

【0108】

図13Aでは、上述した半導体装置10に対応する構成として、半導体装置10__1乃至10__n（nは2以上の数）と、半導体装置10__1乃至10__nとの間でデータの入出力および制御を行うコントローラ71Gを図示している。コントローラ71Gは、内部にSRAM等のメモリ回路60を有する。コントローラ71Gは、複数の半導体装置10__1乃至10__nで得られる出力データMACをメモリ回路60に保持する。そしてメモリ回路60に保持した出力データMACを複数の半導体装置10__1乃至10__nにおける入力データA_{IN}として出力する構成とする。当該構成とすることで複数の半導体装置を用いた、並列数が高められた並列計算を行うことができる。

【0109】

また図13Aとは別の構成例である図13Bでは、コントローラ71Gにおいて、メモリ回路60に保持した出力データに対し、別の演算処理を施した入力データを複数の半導体装置10__1乃至10__nにおける入力データA_{IN}__1乃至A_{IN}__nをとって出力する構成とする。当該構成の場合、例えばコントローラ71Gでは、メモリ回路60に保持した出力データに対し、活性化関数に基づく演算処理、プーリング処理、規格化演算処理（ノーマライゼーション）などを行う構成とする。当該構成とすることで複数の半導体装置を用いた、並列数が高められた並列計算に加え、畳み込み演算処理以外の演算処理を効率よく行うことができる。

【0110】

半導体装置10では、入出力バッファ81におけるバッファメモリを利用して演算回路30の演算結果に応じた出力データMACを中間データとして演算制御回路82に入力する。演算制御回路82がこの中間データを再度演算回路30への入力データとして出力できる。そのため、演算途中のデータを半導体装置10の外部にあるメインメモリなどに読み出すことなく、演算処理を実行可能である。また半導体装置10では、メモリ回路部と、演算回路と、の間の電氣的な接続を、絶縁膜等に設ける開口部の配線を介して行うこと

10

20

30

40

50

ができるため、配線数をふやすことで並列数を増やすことが可能である。そのため半導体装置 10 では、CPU 110 のデータバス幅以上のビット数の並列計算が可能となる。また膨大な数の重みデータを CPU 110 との間で転送する回数を削減できるため、低消費電力化を図ることができる。

【0111】

以上説明したように、本発明の一態様は、小型化が図られた、アクセラレータとして機能する半導体装置を提供することができる。または、本発明の一態様は、低消費電力化が図られた、アクセラレータとして機能する半導体装置を提供することができる。または、新規な構成の、アクセラレータとして機能する半導体装置を提供することができる。

【0112】

(実施の形態 2)

本実施の形態では、半導体装置 10 として説明したアクセラレータに適用可能な Si トランジスタを有する集積回路の構成について説明する。当該構成とすることで、半導体装置の設計自由度を高めるとともに、半導体装置の集積度を高めることができる。

【0113】

図 14 A は、集積回路 390 を説明するための断面模式図の一例である。集積回路 390 において、パッケージ基板 400 上には上記実施の形態で説明した半導体装置 10 が設けられる。パッケージ基板 400 には、別のプリント基板等と接続するためのソルダーボール 401 が設けられている。半導体装置 10 は、パッケージ基板 400 とインターポーザ等を介して接続される。パッケージ基板 400 は、セラミック基板、プラスチック基板、またはガラスエポキシ基板などを用いることができる。

【0114】

図 14 A に図示する集積回路 390 の断面模式図は、層 11 側において、半導体基板 402、半導体基板 402 に設けられる複数のトランジスタ 403、配線 404、および電極 405 を図示している。また層 12 側において、半導体基板 412、半導体基板 412 に設けられる複数のトランジスタ 413、配線 414、および電極 415 を図示している。図 14 A に図示する領域 420 の構成について、図 14 B を参照して説明する。

【0115】

図 14 B では、図 14 A に図示した半導体基板 402、トランジスタ 403、配線 404、および電極 405 を図示している。また図 14 B では、図 14 A に図示した半導体基板 412、半導体基板 412 に設けられる複数のトランジスタ 413、配線 414、および電極 415 を図示している。

【0116】

層 11 と層 12 との貼り合わせをする場合、それぞれの半導体基板に設けられたトランジスタ 403 およびトランジスタ 413 は、配線 404 および配線 414 を介して、電極 405 および電極 415 で接続される。電極 405 および電極 415 は、Cu-Cu 接合やマイクロバンプなどの接合技術により貼り合わせる。なお、Cu-Cu 接合は、Cu (銅) のパッド同士を接続することで電氣的導通を図る技術である。なお半導体基板 402、412 に対し、Si 貫通電極 (TSV: through-silicon via) を形成し、電極 405 および電極 415 に接続する構成としてもよい。また、半導体基板 402、412 の厚さは 100 μm 乃至 300 μm であるが、研磨により、10 μm 乃至 100 μm に薄片化しても良い。

【0117】

図 15 を用いて層 11 における半導体基板 402、トランジスタ 403、配線 404、電極 405、および層 12 における半導体基板 412、トランジスタ 413、配線 414、電極 415 を説明する。なお繰り返しの説明を避けるため、層 11 にある半導体基板 402、トランジスタ 403、配線 404、電極 405 に対応する、層 12 の構成である半導体基板 412、トランジスタ 413、配線 414、電極 415 については説明を簡略化する。

【0118】

トランジスタ403は、半導体基板402上に設けられ、ゲートとして機能する導電体430、ゲート絶縁体として機能する絶縁体431、半導体基板402の一部からなる半導体領域432、およびソース領域またはドレイン領域として機能する低抵抗領域433a、および低抵抗領域433bを有する。トランジスタ403は、pチャネル型、あるいはnチャネル型のいずれでもよい。

【0119】

半導体領域432、低抵抗領域433a、および低抵抗領域433bを有する半導体基板402は、シリコン系半導体などの半導体を含むことが好ましく、単結晶シリコンを含むことが好ましい。または、Ge（ゲルマニウム）、SiGe（シリコンゲルマニウム）、GaAs（ガリウムヒ素）、GaAlAs（ガリウムアルミニウムヒ素）などを有する材料で形成してもよい。結晶格子に応力を与え、格子間隔を変化させることで有効質量を制御したシリコンを用いた構成としてもよい。またはGaAsとGaAlAs等を用いることで、トランジスタ403をHEMT（High Electron Mobility Transistor）としてもよい。

【0120】

半導体領域432、低抵抗領域433a、および低抵抗領域433bに適用される半導体材料に加え、ヒ素、リンなどのn型の導電性を付与する元素、またはホウ素などのp型の導電性を付与する元素を含む。

【0121】

ゲート電極として機能する導電体430は、ヒ素、リンなどのn型の導電性を付与する元素、もしくはホウ素などのp型の導電性を付与する元素を含むシリコンなどの半導体材料、金属材料、合金材料、または金属酸化物材料などの導電性材料を用いることができる。

【0122】

なお、導電体の材料により、仕事関数が定まるため、導電体の材料を変更することで、しきい値電圧を調整することができる。具体的には、導電体に窒化チタンや窒化タンタルなどの材料を用いることが好ましい。さらに導電性と埋め込み性を両立するために導電体にタングステンやアルミニウムなどの金属材料を積層として用いることが好ましく、特にタングステンをを用いることが耐熱性の点で好ましい。

【0123】

なお、図15に示すトランジスタ403は一例であり、その構造に限定されず、回路構成や駆動方法に応じて適切なトランジスタを用いればよい。

【0124】

トランジスタ403を覆って、絶縁体440、絶縁体442、絶縁体444、および絶縁体446が順に積層して設けられている。

【0125】

絶縁体440、絶縁体442、絶縁体444、および絶縁体446として、例えば、酸化シリコン、酸化窒化シリコン、窒化酸化シリコン、窒化シリコン、酸化アルミニウム、酸化窒化アルミニウム、窒化酸化アルミニウム、窒化アルミニウムなどを用いればよい。

【0126】

絶縁体442は、その下方に設けられるトランジスタ403などによって生じる段差を平坦化する平坦化膜としての機能を有していてもよい。例えば、絶縁体442の上面は、平坦性を高めるために化学機械研磨（CMP）法等を用いた平坦化処理により平坦化されていてもよい。

【0127】

なお、絶縁体446は、絶縁体444よりも誘電率が低いことが好ましい。例えば、絶縁体446の比誘電率は4未満が好ましく、3未満がより好ましい。また例えば、絶縁体446の比誘電率は、絶縁体444の比誘電率の0.7倍以下が好ましく、0.6倍以下がより好ましい。誘電率が低い材料を層間膜とすることで、配線間に生じる寄生容量を低減することができる。

【0128】

10

20

30

40

50

また、絶縁体 440、絶縁体 442、絶縁体 444、および絶縁体 446 にはトランジスタ 403 と電氣的に接続する導電体 448、および配線 404 として機能する導電体等が埋め込まれている。なお、導電体 448 はプラグ、または配線として機能する。また、プラグまたは配線として機能する導電体は、複数の構造をまとめて同一の符号を付与する場合がある。また、本明細書等において、配線と、配線と電氣的に接続するプラグとが一体物であってもよい。すなわち、導電体の一部が配線として機能する場合、および導電体の一部がプラグとして機能する場合もある。

【0129】

各プラグ、および配線（導電体 448、および配線 404 等）の材料としては、金属材料、合金材料、金属窒化物材料、または金属酸化物材料などの導電性材料を、単層または積層して用いることができる。耐熱性と導電性を両立するタングステンやモリブデンなどの高融点材料を用いることが好ましく、タングステンを用いることが好ましい。または、アルミニウムや銅などの低抵抗導電性材料で形成することが好ましい。低抵抗導電性材料を用いることで配線抵抗を低くすることができる。

【0130】

電極 405 は、絶縁体 446、および配線 404 上に設けることができる。例えば、図 15 において、絶縁体 450、絶縁体 452、および絶縁体 454 が順に積層して設けられている。電極 405 は、絶縁体 450、絶縁体 452、および絶縁体 454 を形成した後に関口部を設け、該開口部を埋めるように導電層を埋め込むように設け、表面に CMP 法による研磨を行って形成すればよい。

【0131】

電極 405 としては、例えば、Al、Cr、Cu、Ta、Ti、Mo、W から選ばれた元素を含む金属膜、又は上述した元素を成分とする金属窒化物膜（窒化チタン膜、窒化モリブデン膜、窒化タングステン膜）等を用いることができる。なお電極 405 として、導電性を有するバンプ（以降、バンプ）を用いることで、Cu-Cu（銅・銅）直接接合などとして行うことができる。なお、Cu-Cu 直接接合は、Cu（銅）のパッド同士を接続することで電氣的導通を図る技術である。電極 405 は、プラグ、または配線として機能する。なお電極 405 は、導電体 448、および配線 404 等と同様の材料を用いて設けることができる。

【0132】

本実施の形態は、他の実施の形態の記載と適宜組み合わせることができる。

【0133】

（実施の形態 3）

本実施の形態では、上記実施の形態で説明した CPU 110 で実行するプログラムの演算の一部を半導体装置 10 として説明したアクセラレータで実行する場合の、動作の一例を説明する。

【0134】

図 16 は、CPU で実行するプログラムの演算の一部をアクセラレータで実行する場合の、動作の一例を説明する図である。

【0135】

CPU にて、ホストプログラムが実行される（ホストプログラム実行；ステップ S1）。

【0136】

CPU は、アクセラレータを用いて演算を行う際に必要とされるデータ用領域を、メモリ回路部に確保するとの命令を確認した場合（メモリ確保命令；ステップ S2）、該データ用領域を、メモリ回路部に確保する（メモリ確保；ステップ S3）。

【0137】

次に、CPU は、メインメモリあるいは外部記憶装置から上記メモリ回路部へ入力データである重みデータを送信する（データ送信；ステップ S4）。上記メモリ回路部は該重みデータを受信し、該重みデータを、ステップ S2 で確保された領域に格納する（データ受信；ステップ S5）。

【0138】

CPUは、カーネルプログラムを起動するとの命令を確認した場合（カーネルプログラムの起動；ステップS6）、アクセラレータは、カーネルプログラムの実行を開始する（演算開始；ステップS7）。

【0139】

アクセラレータがカーネルプログラムの実行を開始した直後、CPUを、演算を行う状態からPG（パワーゲーティング）状態へと切り替えてもよい（PG状態移行；ステップS8）。その場合、アクセラレータがカーネルプログラムの実行を終了する直前に、CPUは、PG状態から演算を行う状態へ切り替えられる（PG状態停止ステップS9）。ステップS8からステップS9までの期間、CPUをPG状態にすることで、演算処理システム全体として消費電力および発熱を抑制することができる。

10

【0140】

アクセラレータがカーネルプログラムの実行を終了すると、出力データがアクセラレータ内の演算結果を保持する記憶部に格納される（演算終了；ステップS10）。

【0141】

カーネルプログラムの実行が終了した後、CPUは、記憶部に格納された出力データをメインメモリあるいは外部記憶装置へ送信するとの命令を確認した場合（データ送信リクエスト；ステップS11）、上記の出力データがメインメモリあるいは外部記憶装置へ送信され、メインメモリあるいは外部記憶装置に格納される（データ送信；ステップS12）。

20

【0142】

以上のステップS1からステップS14までの動作を繰り返すことにより、CPUおよびアクセラレータの消費電力および発熱を抑制しつつ、CPUで実行する演算の一部をアクセラレータで実行することができる。本発明の一態様の半導体装置は、非ノイマン型アーキテクチャを有し、処理速度の増加に伴って消費電力が大きくなるノイマン型アーキテクチャと比較して、極めて少ない消費電力で演算処理を行うことができる。

【0143】

本実施の形態は、他の実施の形態の記載と適宜組み合わせることができる。

【0144】

（実施の形態4）

本実施の形態では、パワーゲーティングが可能なCPUコアを有するCPUの一例について説明する。

30

【0145】

図17に、CPU110の構成例を示す。CPU110は、CPUコア（CPU Core）200、L1（レベル1）キャッシュメモリ装置（L1 Cache）202、L2キャッシュメモリ装置（L2 Cache）203、バスインターフェース部（Bus I/F）205、パワースイッチ210～212、レベルシフタ（LS）214を有する。CPUコア200はフリップフロップ220を有する。

【0146】

バスインターフェース部205によって、CPUコア200、L1キャッシュメモリ装置202、L2キャッシュメモリ装置203が相互に接続される。

40

【0147】

外部から入力される割り込み信号（Interrupts）、CPU110が発行する信号SLEEP1等の信号に応じて、PMU193はクロック信号GCLK1、各種のPG（パワーゲーティング）制御信号（PG control signals）の生成を行う。クロック信号GCLK1、PG制御信号はCPU110に入力される。PG制御信号は、パワースイッチ210～212、フリップフロップ220を制御する。

【0148】

パワースイッチ210、211は、仮想電源線V__VDD（以下、V__VDD線と呼ぶ）への電圧VDDD、VDD1の供給をそれぞれ制御する。パワースイッチ212は、レ

50

ベルシフタ（LS）214への電圧VDDHの供給を制御する。CPU110およびPMU193には、パワースイッチを介さずに電圧VSSSが入力される。PMU193には、パワースイッチを介さずに電圧VDDDが入力される。

【0149】

電圧VDDD、VDD1はCMOS回路用の駆動電圧である。電圧VDD1は電圧VDDDよりも低く、スリープ状態での駆動電圧である。電圧VDDHはOストランジスタ用の駆動電圧であり、電圧VDDDよりも高い。

【0150】

L1キャッシュメモリ装置202、L2キャッシュメモリ装置203、バスインターフェース部205それぞれは、少なくとも1つパワーゲーティング可能なパワードメインを有する。パワーゲーティング可能なパワードメインには、1または複数のパワースイッチが設けられている。これらのパワースイッチは、PG制御信号によって制御される。

10

【0151】

フリップフロップ220は、レジスタに用いられる。フリップフロップ220には、バックアップ回路が設けられている。以下、フリップフロップ220について説明する。

【0152】

図18にフリップフロップ220（Flip-flop）の回路構成例を示す。フリップフロップ220はスキャンフリップフロップ（Scan Flip-flop）221、バックアップ回路（Backup Circuit）222を有する。

【0153】

スキャンフリップフロップ221は、ノードD1、Q1、SD、SE、RT、CK、クロックバッファ回路221Aを有する。

20

【0154】

ノードD1はデータ（data）入力ノードであり、ノードQ1はデータ出力ノードであり、ノードSDはスキャンテスト用データの入力ノードである。ノードSEは信号SCEの入力ノードである。ノードCKはクロック信号GCLK1の入力ノードである。クロック信号GCLK1はクロックバッファ回路221Aに入力される。スキャンフリップフロップ221のアナログスイッチは、クロックバッファ回路221AのノードCK1、CKB1に接続される。ノードRTはリセット信号（reset signal）の入力ノードである。

30

【0155】

信号SCEは、スキャンイネーブル信号であり、PMU193で生成される。PMU193は信号BK、RCを生成する。レベルシフタ214は信号BK、RCをレベルシフトし、信号BKH、RCHを生成する。信号BKはバックアップ信号、信号RCはリカバリ信号である。

【0156】

スキャンフリップフロップ221の回路構成は、図18に限定されない。標準的な回路ライブラリに用意されているフリップフロップを適用することができる。

【0157】

バックアップ回路222は、ノードSD__IN、SN11、トランジスタM11～M13、容量素子C11を有する。

40

【0158】

ノードSD__INは、スキャンテストデータの入力ノードであり、スキャンフリップフロップ221のノードQ1に接続される。ノードSN11は、バックアップ回路222の保持ノードである。容量素子C11はノードSN11の電圧を保持するための保持容量である。

【0159】

トランジスタM11はノードQ1とノードSN11間の導通状態を制御する。トランジスタM12はノードSN11とノードSD間の導通状態を制御する。トランジスタM13はノードSD__INとノードSD間の導通状態を制御する。トランジスタM11、M13

50

のオンオフは信号 B K H で制御され、トランジスタ M 1 2 のオンオフは信号 R C H で制御される。

【 0 1 6 0 】

トランジスタ M 1 1 ~ M 1 3 は、上述したメモリ回路 2 1 が有するトランジスタ 6 1 乃至 6 3 と同様に、O S トランジスタである。トランジスタ M 1 1 ~ M 1 3 はバックゲート有する構成を図示している。トランジスタ M 1 1 ~ M 1 3 のバックゲートは、電圧 V B G 1 を供給する電源線に接続されている。

【 0 1 6 1 】

少なくともトランジスタ M 1 1、M 1 2 が O S トランジスタであることが好ましい。オフ電流が極めて小さいという O S トランジスタの特長によって、ノード S N 1 1 の電圧の低下を抑えることができること、データの保持に電力を殆んど消費しないことから、バックアップ回路 2 2 2 は不揮発性の特性をもつ。容量素子 C 1 1 の充放電によってデータを書き換えるため、バックアップ回路 2 2 2 は原理的には書き換え回数に制約はなく、低エネルギーで、データの書き込みおよび読み出しが可能である。

10

【 0 1 6 2 】

バックアップ回路 2 2 2 の全てのトランジスタは O S トランジスタであることが非常に好ましい。図 1 8 B に示すように、シリコン C M O S 回路で構成されるスキャンフリップフロップ 2 2 1 上にバックアップ回路 2 2 2 を積層することができる。

【 0 1 6 3 】

バックアップ回路 2 2 2 は、スキャンフリップフロップ 2 2 1 と比較して素子数が非常に少ないので、バックアップ回路 2 2 2 を積層するためにスキャンフリップフロップ 2 2 1 の回路構成およびレイアウトの変更が必要ない。つまり、バックアップ回路 2 2 2 は、汎用性が非常に高いバックアップ回路である。また、スキャンフリップフロップ 2 2 1 が形成されている領域内にバックアップ回路 2 2 2 を設けることができるので、バックアップ回路 2 2 2 を組み込んでも、フリップフロップ 2 2 0 の面積オーバーヘッドはゼロにすることが可能である。よって、バックアップ回路 2 2 2 をフリップフロップ 2 2 0 に設けることで、C P U コア 2 0 0 のパワーゲーティングが可能となる。パワーゲーティングに必要なエネルギーが少ないため、C P U コア 2 0 0 を高効率にパワーゲーティングすることが可能である。

20

【 0 1 6 4 】

バックアップ回路 2 2 2 を設けることによって、トランジスタ M 1 1 による寄生容量がノード Q 1 に付加されることになるが、ノード Q 1 に接続される論理回路による寄生容量と比較して小さいので、スキャンフリップフロップ 2 2 1 の動作に影響はない。つまり、バックアップ回路 2 2 2 を設けても、フリップフロップ 2 2 0 の性能は実質的に低下しない。

30

【 0 1 6 5 】

C P U コア 2 0 0 の低消費電力状態として、例えば、クロックゲーティング状態、パワーゲーティング状態、休止状態を設定することができる。P M U 1 9 3 は、割り込み信号、信号 S L E E P 1 等に基づき、C P U コア 2 0 0 の低消費電力モードを選択する。例えば、通常動作状態からクロックゲーティング状態に移行する場合、P M U 1 9 3 はクロック信号 G C L K 1 の生成を停止する。

40

【 0 1 6 6 】

例えば、通常動作状態から休止状態に移行する場合は、P M U 1 9 3 は、電圧および／または周波数スケーリングを行う。例えば、電圧スケーリングを行う場合、P M U 1 9 3 は、電圧 V D D 1 を C P U コア 2 0 0 に入力するため、パワースイッチ 2 1 0 をオフにし、パワースイッチ 2 1 1 をオンにする。電圧 V D D 1 は、スキャンフリップフロップ 2 2 1 のデータを消失させない電圧である。周波数スケーリングを行う場合、P M U 1 9 3 はクロック信号 G C L K 1 の周波数を低下させる。

【 0 1 6 7 】

C P U コア 2 0 0 を通常動作状態からパワーゲーティング状態に移行する場合には、ス

50

キャンフリップフロップ 221 のデータをバックアップ回路 222 にバックアップする動作が行われる。CPU コア 200 をパワーゲーティング状態から通常動作状態に復帰する際には、バックアップ回路 222 のデータをキャンフリップフロップ 221 にリカバリする動作が行われる。

【0168】

図 19 に、CPU コア 200 のパワーゲーティングシーケンスの一例を示す。なお、図 19 において、 $t_1 \sim t_7$ は時刻を表している。信号 $PSE0 \sim PSE2$ は、パワースイッチ 210 ~ 212 の制御信号であり、PMU 193 で生成される。信号 $PSE0$ が “H” / “L” のとき、パワースイッチ 210 はオン / オフである。信号 $PSE1$ 、 $PSE2$ についても同様である。

10

【0169】

時刻 t_1 以前は、通常動作状態 (Normal Operation) である。パワースイッチ 210 はオンであり、CPU コア 200 には電圧 V_{DDD} が入力される。キャンフリップフロップ 221 は通常動作を行う。このとき、レベルシフタ 214 は動作させる必要がないため、パワースイッチ 212 はオフであり、信号 SCE 、 BK 、 RC は “L” である。ノード SE が “L” であるため、キャンフリップフロップ 221 はノード $D1$ のデータを記憶する。なお、図 19 の例では、時刻 t_1 において、バックアップ回路 222 のノード $SN11$ は “L” である。

【0170】

バックアップ (Backup) 時の動作を説明する。動作時刻 t_1 で、PMU 193 はクロック信号 $GCLK1$ を停止し、信号 $PSE2$ 、 BK を “H” にする。レベルシフタ 214 はアクティブになり、“H” の信号 BKH をバックアップ回路 222 に出力する。

20

【0171】

バックアップ回路 222 のトランジスタ $M11$ がオンになり、キャンフリップフロップ 221 のノード $Q1$ のデータがバックアップ回路 222 のノード $SN11$ に書き込まれる。キャンフリップフロップ 221 のノード $Q1$ が “L” であれば、ノード $SN11$ は “L” のままであり、ノード $Q1$ が “H” であれば、ノード $SN11$ は “H” になる。

【0172】

PMU 193 は、時刻 t_2 で信号 $PSE2$ 、 BK を “L” にし、時刻 t_3 で信号 $PSE0$ を “L” にする。時刻 t_3 で、CPU コア 200 の状態はパワーゲーティング状態に移行する。なお、信号 BK を立ち下げるタイミングで信号 $PSE0$ を立ち下げてもよい。

30

【0173】

パワーゲーティング (Power-gating) 時の動作を説明する。信号 $PSE0$ が “L” になることで、 V_{VDD} 線の電圧が低下するため、ノード $Q1$ のデータは失われる。ノード $SN11$ は、時刻 t_3 でのノード $Q1$ のデータを保持し続ける。

【0174】

リカバリ (Recovery) 時の動作を説明する。時刻 t_4 で、PMU 193 が信号 $PSE0$ を “H” にすることで、パワーゲーティング状態からリカバリ状態に移行する。 V_{VDD} 線の充電が開始され、 V_{VDD} 線の電圧が V_{DDD} になった状態 (時刻 t_5) で、PMU 193 は信号 $PSE2$ 、 RC 、 SCE を “H” にする。

40

【0175】

トランジスタ $M12$ はオンになり、容量素子 $C11$ の電荷がノード $SN11$ とノード SD とに分配される。ノード $SN11$ が “H” であれば、ノード SD の電圧は上昇する。ノード SE は “H” であるので、キャンフリップフロップ 221 の入力側ラッチ回路にノード SD のデータが書き込まれる。時刻 t_6 でノード CK にクロック信号 $GCLK1$ が入力されると、入力側ラッチ回路のデータがノード $Q1$ に書き込まれる。つまり、ノード $SN11$ のデータがノード $Q1$ に書き込まれたことになる。

【0176】

時刻 t_7 で、PMU 193 は信号 $PSE2$ 、 SCE 、 RC を “L” にし、リカバリ動作が終了する。

50

【0177】

ＯＳトランジスタを用いたバックアップ回路２２２は、動的および静的低消費電力双方が小さいため、ノーマリオフ・コンピューティングに非常に好適である。なお、ＯＳトランジスタを用いたバックアップ回路２２２を有するＣＰＵコア２００を含むＣＰＵ１１０は、ＮｏｆｆＣＰＵ（登録商標）と呼称することができる。ＮｏｆｆＣＰＵは、不揮発性メモリを有し、動作が必要ない場合には、電力供給を停止することができる。フリップフロップ２２０を搭載しても、ＣＰＵコア２００の性能低下、動的電力の増加をほとんど発生させないようにできる。

【0178】

なお、ＣＰＵコア２００は複数のパワーゲーティング可能なパワードメインを有してもよい。複数のパワードメインには、電圧の入力を制御するための１または複数のパワースイッチが設けられる。また、ＣＰＵコア２００は、１または複数のパワーゲーティングが行われないパワードメインを有していてもよい。例えば、パワーゲーティングが行われないパワードメインに、フリップフロップ２２０、パワースイッチ２１０～２１２の制御を行うためのパワーゲーティング制御回路を設けてもよい。

【0179】

なお、フリップフロップ２２０の適用はＣＰＵ１１０に限定されない。ＣＰＵ１１０において、パワーゲーティング可能なパワードメインに設けられるレジスタに、フリップフロップ２２０を適用できる。

【0180】

本実施の形態は、他の実施の形態の記載と適宜組み合わせることができる。

【0181】

(実施の形態５)

本実施の形態では、上記実施の形態で説明したＣＰＵ１１０、および半導体装置１０として説明したアクセラレータに適用可能なトランジスタの構成の一例について説明する。一例として、異なる電気特性を有するトランジスタを積層して設ける構成について説明する。当該構成とすることで、半導体装置の設計自由度を高めることができる。また、異なる電気特性を有するトランジスタを積層して設けることで、半導体装置の集積度を高めることができる。

【0182】

半導体装置の断面構造の一部を図２０に示す。図２０に示す半導体装置は、トランジスタ５５０と、トランジスタ５００と、容量素子６００と、を有している。図２１Ａはトランジスタ５００のチャネル長方向の断面図であり、図２１Ｂはトランジスタ５００のチャネル幅方向の断面図である。例えば、トランジスタ５００は上記実施の形態に示したメモリ回路２１が有するＯＳトランジスタ、つまりチャネル形成領域に酸化物半導体を有するトランジスタに相当する。また、トランジスタ５５０は上記実施の形態に示した演算回路３０が有するＳｉトランジスタ、つまりチャネル形成領域にシリコンを有するトランジスタに相当する。また、容量素子６００はメモリ回路２１が有する容量素子に相当する。

【0183】

トランジスタ５００は、ＯＳトランジスタである。ＯＳトランジスタは、オフ電流が極めて少ない。よって、トランジスタ５００を介して記憶ノードに書き込んだデータ電圧あるいは電荷を長期間保持することが可能である。つまり、記憶ノードのリフレッシュ動作頻度を低減、あるいは、リフレッシュ動作を必要としないため、半導体装置の消費電力を低減することができる。

【0184】

図２０では、トランジスタ５００はトランジスタ５５０の上方に設けられ、容量素子６００はトランジスタ５５０、およびトランジスタ５００の上方に設けられている。

【0185】

トランジスタ５５０は、基板３１１に設けられる。基板３１１は、例えば、ｐ型のシリコン基板である。基板３１１は、ｎ型のシリコン基板でもよい。酸化物層３１４は、基板

10

20

30

40

50

311に埋め込み酸化(Burried oxide)によって形成された絶縁層(BOX層ともいう)、例えば酸化シリコンであることが好ましい。トランジスタ550は、基板311に酸化物層314を介して設けられた単結晶シリコン、いわゆるSOI(Silicon On Insulator)基板に設けられる。

【0186】

SOI基板における基板311は、素子分離層として機能する絶縁体313が設けられる。また基板311は、ウェル領域312を有する。ウェル領域312は、トランジスタ550の導電型に応じてn型またはp型の導電性が付与された領域である。SOI基板における単結晶シリコンには、半導体領域315、ソース領域またはドレイン領域として機能する低抵抗領域316a、低抵抗領域316bが設けられる。またウェル領域312上には、低抵抗領域316cを有する。

10

【0187】

トランジスタ550は、導電性を付与する不純物元素が付加されたウェル領域312に重ねて設けることができる。ウェル領域312は、低抵抗領域316cを介して電位を独立して変化させることで、トランジスタ550のボトムゲート電極として機能させることができる。そのため、トランジスタ550のしきい値電圧を制御することができる。特に、ウェル領域312に負の電位を印加することにより、トランジスタ550のしきい値電圧をより大きくし、オフ電流を低減することが可能となる。したがって、ウェル領域312に負の電位を印加することで、Siトランジスタのゲート電極に印加する電位が0Vのときのドレイン電流を小さくすることができる。その結果、トランジスタ550を有する演算回路30における貫通電流等に基づく消費電力を低減でき、演算効率の向上を図ることができる。

20

【0188】

トランジスタ550は、半導体層の上面およびチャネル幅方向の側面が絶縁体317を介して導電体318に覆われている、いわゆるFin型とすることが好ましい。トランジスタ550をFin型とすることにより、実効上のチャネル幅が増大することによりトランジスタ550のオン特性を向上させることができる。また、ゲート電極の電界の寄与を高くすることができるため、トランジスタ550のオフ特性を向上させることができる。

【0189】

なお、トランジスタ550は、pチャネル型のトランジスタ、あるいはnチャネル型のトランジスタのいずれでもよい。

30

【0190】

導電体318は、第1ゲート(トップゲートともいう)電極として機能する場合がある。また、ウェル領域312は、第2ゲート(ボトムゲートともいう)電極として機能する場合がある。その場合、ウェル領域312に印加する電位は、低抵抗領域316cを介して制御することができる。

【0191】

半導体領域315のチャネルが形成される領域、その近傍の領域、ソース領域、またはドレイン領域となる低抵抗領域316a、および低抵抗領域316b、ウェル領域312の電位を制御する電極に接続される低抵抗領域316cなどにおいて、シリコン系半導体などの半導体を含むことが好ましく、単結晶シリコンを含むことが好ましい。または、Ge(ゲルマニウム)、SiGe(シリコンゲルマニウム)、GaAs(ガリウムヒ素)、GaAlAs(ガリウムアルミニウムヒ素)などを有する材料で形成してもよい。結晶格子に応力を与え、格子間隔を変化させることで有効質量を制御したシリコンを用いた構成としてもよい。またはGaAsとGaAlAs等を用いることで、トランジスタ550をHEMT(High Electron Mobility Transistor)としてもよい。

40

【0192】

ウェル領域312、低抵抗領域316a、低抵抗領域316b、および低抵抗領域316cは、半導体領域315に適用される半導体材料に加え、ヒ素、リンなどのn型の導電

50

性を付与する元素、またはホウ素などの p 型の導電性を付与する元素を含む。

【0193】

ゲート電極として機能する導電体 318 は、ヒ素、リンなどの n 型の導電性を付与する元素、もしくはホウ素などの p 型の導電性を付与する元素を含むシリコンなどの半導体材料、金属材料、合金材料、または金属酸化物材料などの導電性材料を用いることができる。また導電体 318 は、ニッケルシリサイド等のシリサイドを用いてもよい。

【0194】

なお、導電体の材料によって仕事関数が決まるため、当該導電体の材料を選択することで、トランジスタのしきい値電圧を調整することができる。具体的には、導電体に窒化チタンや窒化 tantalum などの材料を用いることが好ましい。さらに導電性と埋め込み性を両立するために導電体にタングステンやアルミニウムなどの金属材料を積層として用いることが好ましく、特にタングステンをを用いることが耐熱性の点で好ましい。

【0195】

低抵抗領域 316 a、低抵抗領域 316 b、および低抵抗領域 316 c は、別の導電体、例えばニッケルシリサイド等のシリサイドを積層して設ける構成としてもよい。当該構成とすることで、電極として機能する領域の導電性を高めることができる。またこのとき、ゲート電極として機能する導電体 318 の側面、およびゲート絶縁膜として機能する絶縁体の側面には、サイドウォールスペーサ（側壁絶縁層ともいう）として機能する絶縁体を設ける構成としてもよい。当該構成とすることで、導電体 318 と、低抵抗領域 316 a および低抵抗領域 316 b と、が導通状態となることを防ぐことができる。

【0196】

トランジスタ 550 を覆って、絶縁体 320、絶縁体 322、絶縁体 324、および絶縁体 326 が順に積層して設けられている。

【0197】

絶縁体 320、絶縁体 322、絶縁体 324、および絶縁体 326 として、例えば、酸化シリコン、酸化窒化シリコン、窒化酸化シリコン、窒化シリコン、酸化アルミニウム、酸化窒化アルミニウム、窒化酸化アルミニウム、窒化アルミニウムなどを用いればよい。

【0198】

なお、本明細書中において、酸化窒化シリコンとは、その組成として窒素よりも酸素の含有量が多い材料を指し、窒化酸化シリコンとは、その組成として、酸素よりも窒素の含有量が多い材料を示す。また、本明細書中において、酸化窒化アルミニウムとは、その組成として窒素よりも酸素の含有量が多い材料を指し、窒化酸化アルミニウムとは、その組成として、酸素よりも窒素の含有量が多い材料を示す。

【0199】

絶縁体 322 は、その下方に設けられるトランジスタ 550 などによって生じる段差を平坦化する平坦化膜としての機能を有していてもよい。例えば、絶縁体 322 の上面は、平坦性を高めるために化学機械研磨（CMP）法等を用いた平坦化処理により平坦化されていてもよい。

【0200】

また、絶縁体 324 には、基板 311、またはトランジスタ 550 などから、トランジスタ 500 が設けられる領域に、水素や不純物が拡散しないようなバリア性を有する膜を用いることが好ましい。

【0201】

水素に対するバリア性を有する膜の一例として、例えば、CVD 法で形成した窒化シリコンを用いることができる。ここで、トランジスタ 500 等の酸化物半導体を有する半導体素子に、水素が拡散することで、当該半導体素子の特性が低下する場合がある。したがって、トランジスタ 500 と、トランジスタ 550 との間に、水素の拡散を抑制する膜を用いることが好ましい。水素の拡散を抑制する膜とは、具体的には、水素の脱離量が少ない膜とする。

【0202】

水素の脱離量は、例えば、昇温脱離ガス分析法（TDS）などを用いて分析することができる。例えば、絶縁体324の水素の脱離量は、TDS分析において、膜の表面温度が50℃から500℃の範囲において、水素原子に換算した脱離量が、絶縁体324の面積あたりに換算して、 $10 \times 10^{15} \text{ atoms/cm}^2$ 以下、好ましくは $5 \times 10^{15} \text{ atoms/cm}^2$ 以下であればよい。

【0203】

なお、絶縁体326は、絶縁体324よりも誘電率が低いことが好ましい。例えば、絶縁体326の比誘電率は4未満が好ましく、3未満がより好ましい。また例えば、絶縁体326の比誘電率は、絶縁体324の比誘電率の0.7倍以下が好ましく、0.6倍以下がより好ましい。誘電率が低い材料を層間膜とすることで、配線間に生じる寄生容量を低減することができる。

10

【0204】

また、絶縁体320、絶縁体322、絶縁体324、および絶縁体326には容量素子600、またはトランジスタ500と接続する導電体328、および導電体330等が埋め込まれている。なお、導電体328、および導電体330は、プラグまたは配線としての機能を有する。また、プラグまたは配線としての機能を有する導電体は、複数の構成をまとめて同一の符号を付与する場合がある。また、本明細書等において、配線と、配線と接続するプラグとが一体物であってもよい。すなわち、導電体の一部が配線として機能する場合、および導電体の一部がプラグとして機能する場合もある。

【0205】

20

各プラグ、および配線（導電体328、導電体330等）の材料としては、金属材料、合金材料、金属窒化物材料、または金属酸化物材料などの導電性材料を、単層または積層して用いることができる。耐熱性と導電性を両立するタングステンやモリブデンなどの高融点材料を用いることが好ましく、タングステンを用いることが好ましい。または、アルミニウムや銅などの低抵抗導電性材料で形成することが好ましい。低抵抗導電性材料を用いることで配線抵抗を低くすることができる。

【0206】

絶縁体326、および導電体330上に、配線層を設けてもよい。例えば、図20では、絶縁体350、絶縁体352、および絶縁体354が順に積層して設けられている。また、絶縁体350、絶縁体352、および絶縁体354には、導電体356が形成されている。導電体356は、トランジスタ550と接続するプラグ、または配線としての機能を有する。なお導電体356は、導電体328、および導電体330と同様の材料を用いて設けることができる。

30

【0207】

なお、例えば、絶縁体350は、絶縁体324と同様に、水素に対するバリア性を有する絶縁体を用いることが好ましい。また、導電体356は、水素に対するバリア性を有する導電体を含むことが好ましい。特に、水素に対するバリア性を有する絶縁体350が有する開口部に、水素に対するバリア性を有する導電体が形成される。当該構成により、トランジスタ550とトランジスタ500とは、バリア層により分離することができ、トランジスタ550からトランジスタ500への水素の拡散を抑制することができる。

40

【0208】

なお、水素に対するバリア性を有する導電体としては、例えば、窒化タンタル等を用いるとよい。また、窒化タンタルと導電性が高いタングステンを積層することで、配線としての導電性を保持したまま、トランジスタ550からの水素の拡散を抑制することができる。この場合、水素に対するバリア性を有する窒化タンタル層が、水素に対するバリア性を有する絶縁体350と接する構成であることが好ましい。

【0209】

絶縁体354、および導電体356上に、配線層を設けてもよい。例えば、図20では、絶縁体360、絶縁体362、および絶縁体364が順に積層して設けられている。また、絶縁体360、絶縁体362、および絶縁体364には、導電体366が形成されて

50

いる。導電体 366 は、プラグまたは配線としての機能を有する。なお導電体 366 は、導電体 328、および導電体 330 と同様の材料を用いて設けることができる。

【0210】

なお、例えば、絶縁体 360 は、絶縁体 324 と同様に、水素に対するバリア性を有する絶縁体を用いることが好ましい。また、導電体 366 は、水素に対するバリア性を有する導電体を含むことが好ましい。特に、水素に対するバリア性を有する絶縁体 360 が有する開口部に、水素に対するバリア性を有する導電体が形成される。当該構成により、トランジスタ 550 とトランジスタ 500 とは、バリア層により分離することができ、トランジスタ 550 からトランジスタ 500 への水素の拡散を抑制することができる。

【0211】

絶縁体 364、および導電体 366 上に、配線層を設けてもよい。例えば、図 20 では、絶縁体 370、絶縁体 372、および絶縁体 374 が順に積層して設けられている。また、絶縁体 370、絶縁体 372、および絶縁体 374 には、導電体 376 が形成されている。導電体 376 は、プラグまたは配線としての機能を有する。なお導電体 376 は、導電体 328、および導電体 330 と同様の材料を用いて設けることができる。

【0212】

なお、例えば、絶縁体 370 は、絶縁体 324 と同様に、水素に対するバリア性を有する絶縁体を用いることが好ましい。また、導電体 376 は、水素に対するバリア性を有する導電体を含むことが好ましい。特に、水素に対するバリア性を有する絶縁体 370 が有する開口部に、水素に対するバリア性を有する導電体が形成される。当該構成により、トランジスタ 550 とトランジスタ 500 とは、バリア層により分離することができ、トランジスタ 550 からトランジスタ 500 への水素の拡散を抑制することができる。

【0213】

絶縁体 374、および導電体 376 上に、配線層を設けてもよい。例えば、図 20 では、絶縁体 380、絶縁体 382、および絶縁体 384 が順に積層して設けられている。また、絶縁体 380、絶縁体 382、および絶縁体 384 には、導電体 386 が形成されている。導電体 386 は、プラグまたは配線としての機能を有する。なお導電体 386 は、導電体 328、および導電体 330 と同様の材料を用いて設けることができる。

【0214】

なお、例えば、絶縁体 380 は、絶縁体 324 と同様に、水素に対するバリア性を有する絶縁体を用いることが好ましい。また、導電体 386 は、水素に対するバリア性を有する導電体を含むことが好ましい。特に、水素に対するバリア性を有する絶縁体 380 が有する開口部に、水素に対するバリア性を有する導電体が形成される。当該構成により、トランジスタ 550 とトランジスタ 500 とは、バリア層により分離することができ、トランジスタ 550 からトランジスタ 500 への水素の拡散を抑制することができる。

【0215】

上記において、導電体 356 を含む配線層、導電体 366 を含む配線層、導電体 376 を含む配線層、および導電体 386 を含む配線層、について説明したが、本実施の形態に係る半導体装置はこれに限られるものではない。導電体 356 を含む配線層と同様の配線層を 3 層以下にしてもよいし、導電体 356 を含む配線層と同様の配線層を 5 層以上にしてもよい。

【0216】

絶縁体 384 上には絶縁体 510、絶縁体 512、絶縁体 514、および絶縁体 516 が、順に積層して設けられている。絶縁体 510、絶縁体 512、絶縁体 514、および絶縁体 516 のいずれかは、酸素や水素に対してバリア性のある物質を用いることが好ましい。

【0217】

例えば、絶縁体 510、および絶縁体 514 には、例えば、基板 311、またはトランジスタ 550 を設ける領域などから、トランジスタ 500 を設ける領域に、水素や不純物に対するバリア性を有する膜を用いることが好ましい。したがって、絶縁体 324 と同様

10

20

30

40

50

の材料を用いることができる。

【0218】

水素に対するバリア性を有する膜の一例として、CVD法で形成した窒化シリコンを用いることができる。ここで、トランジスタ500等の酸化物半導体を有する半導体素子に、水素が拡散することで、当該半導体素子の特性が低下する場合がある。したがって、トランジスタ500と、トランジスタ550との間に、水素の拡散を抑制する膜を用いることが好ましい。

【0219】

また、水素に対するバリア性を有する膜として、例えば、絶縁体510、および絶縁体514には、酸化アルミニウム、酸化ハフニウム、酸化タンタルなどの金属酸化物を用いることが好ましい。

【0220】

特に、酸化アルミニウムは、酸素、およびトランジスタの電気特性の変動要因となる水素、水分などの不純物、の両方に対して膜を透過させない遮断効果が高い。したがって、酸化アルミニウムは、トランジスタの作製工程中および作製後において、水素、水分などの不純物のトランジスタ500への混入を防止することができる。また、トランジスタ500を構成する酸化物からの酸素の放出を抑制することができる。そのため、トランジスタ500に対する保護膜として用いることに適している。

【0221】

また、例えば、絶縁体512、および絶縁体516には、絶縁体320と同様の材料を用いることができる。また、これらの絶縁体に、比較的誘電率が低い材料を適用することで、配線間に生じる寄生容量を低減することができる。例えば、絶縁体512、および絶縁体516として、酸化シリコン膜や酸化窒化シリコン膜などを用いることができる。

【0222】

また、絶縁体510、絶縁体512、絶縁体514、および絶縁体516には、導電体518、およびトランジスタ500を構成する導電体（例えば、導電体503）等が埋め込まれている。なお、導電体518は、容量素子600、またはトランジスタ550と接続するプラグ、または配線としての機能を有する。導電体518は、導電体328、および導電体330と同様の材料を用いて設けることができる。

【0223】

特に、絶縁体510、および絶縁体514と接する領域の導電体518は、酸素、水素、および水に対するバリア性を有する導電体であることが好ましい。当該構成により、トランジスタ550とトランジスタ500とは、酸素、水素、および水に対するバリア性を有する層で、分離することができ、トランジスタ550からトランジスタ500への水素の拡散を抑制することができる。

【0224】

絶縁体516の上方には、トランジスタ500が設けられている。

【0225】

図21Aおよび図21Bに示すように、トランジスタ500は、絶縁体514および絶縁体516に埋め込まれるように配置された導電体503と、絶縁体516および導電体503の上に配置された絶縁体522と、絶縁体522の上に配置された絶縁体524と、絶縁体524の上に配置された酸化物530aと、酸化物530aの上に配置された酸化物530bと、酸化物530b上に互いに離れて配置された導電体542aおよび導電体542bと、導電体542aおよび導電体542b上に配置され、導電体542aと導電体542bの間に重畳して開口が形成された絶縁体580と、開口の底面および側面に配置された絶縁体545と、絶縁体545の形成面に配置された導電体560と、を有する。

【0226】

また、図21Aおよび図21Bに示すように、酸化物530a、酸化物530b、導電体542a、および導電体542bと、絶縁体580の間に絶縁体544が配置されるこ

10

20

30

40

50

とが好ましい。また、図 2 1 A および図 2 1 B に示すように、導電体 5 6 0 は、絶縁体 5 4 5 の内側に設けられた導電体 5 6 0 a と、導電体 5 6 0 a の内側に埋め込まれるように設けられた導電体 5 6 0 b と、を有することが好ましい。また、図 2 1 A および図 2 1 B に示すように、絶縁体 5 8 0、導電体 5 6 0、および絶縁体 5 4 5 の上に絶縁体 5 7 4 が配置されることが好ましい。

【0 2 2 7】

なお、本明細書などにおいて、酸化物 5 3 0 a、および酸化物 5 3 0 b をまとめて酸化物 5 3 0 という場合がある。

【0 2 2 8】

なお、トランジスタ 5 0 0 では、チャンネルが形成される領域と、その近傍において、酸化物 5 3 0 a、および酸化物 5 3 0 b の 2 層を積層する構成について示しているが、本発明はこれに限られるものではない。例えば、酸化物 5 3 0 b の単層、または 3 層以上の積層構成を設ける構成にしてもよい。

【0 2 2 9】

また、トランジスタ 5 0 0 では、導電体 5 6 0 を 2 層の積層構成として示しているが、本発明はこれに限られるものではない。例えば、導電体 5 6 0 が、単層構成であってもよいし、3 層以上の積層構成であってもよい。また、図 2 0、図 2 1 A、および図 2 1 B に示すトランジスタ 5 0 0 は一例であり、その構成に限定されず、回路構成や駆動方法などに応じて適切なトランジスタを用いればよい。

【0 2 3 0】

ここで、導電体 5 6 0 は、トランジスタのゲート電極として機能し、導電体 5 4 2 a および導電体 5 4 2 b は、それぞれソース電極またはドレイン電極として機能する。上記のように、導電体 5 6 0 は、絶縁体 5 8 0 の開口、および導電体 5 4 2 a と導電体 5 4 2 b に挟まれた領域に埋め込まれるように形成される。導電体 5 6 0、導電体 5 4 2 a および導電体 5 4 2 b の配置は、絶縁体 5 8 0 の開口に対して、自己整合的に選択される。つまり、トランジスタ 5 0 0 において、ゲート電極を、ソース電極とドレイン電極の間に、自己整合的に配置させることができる。よって、導電体 5 6 0 を位置合わせのマージンを設けることなく形成することができるので、トランジスタ 5 0 0 の占有面積の縮小を図ることができる。これにより、半導体装置の微細化、高集積化を図ることができる。

【0 2 3 1】

さらに、導電体 5 6 0 が、導電体 5 4 2 a と導電体 5 4 2 b の間の領域に自己整合的に形成されるので、導電体 5 6 0 は、導電体 5 4 2 a または導電体 5 4 2 b と重畳する領域を有さない。これにより、導電体 5 6 0 と導電体 5 4 2 a および導電体 5 4 2 b との間に形成される寄生容量を低減することができる。よって、トランジスタ 5 0 0 のスイッチング速度を向上させ、高い周波数特性を有せしめることができる。

【0 2 3 2】

導電体 5 6 0 は、第 1 ゲート（トップゲートともいう）電極として機能する場合がある。また、導電体 5 0 3 は、第 2 ゲート（ボトムゲートともいう）電極として機能する場合がある。その場合、導電体 5 0 3 に印加する電位を、導電体 5 6 0 に印加する電位と、連動させず、独立して変化させることで、トランジスタ 5 0 0 のしきい値電圧を制御することができる。特に、導電体 5 0 3 に負の電位を印加することにより、トランジスタ 5 0 0 のしきい値電圧をより大きくし、オフ電流を低減することが可能となる。したがって、導電体 5 0 3 に負の電位を印加したほうが、印加しない場合よりも、導電体 5 6 0 に印加する電位が 0 V のときのドレイン電流を小さくすることができる。

【0 2 3 3】

導電体 5 0 3 は、酸化物 5 3 0、および導電体 5 6 0 と、重なるように配置する。これにより、導電体 5 6 0、および導電体 5 0 3 に電位を印加した場合、導電体 5 6 0 から生じる電界と、導電体 5 0 3 から生じる電界と、がつながり、酸化物 5 3 0 に形成されるチャンネル形成領域を覆うことができる。

【0 2 3 4】

10

20

30

40

50

本明細書等において、一対のゲート電極（第1のゲート電極、および第2のゲート電極）の電界によって、チャンネル形成領域を電氣的に取り囲むトランジスタの構成を、*surrounded channel*（*S-channel*）構成とよぶ。また、本明細書等で開示する*S-channel*構成は、*Fin*型構成およびプレーナ型構成とは異なる。*S-channel*構成を採用することで、短チャンネル効果に対する耐性を高める、別言すると短チャンネル効果が発生し難いトランジスタとすることができる。

【0235】

また、導電体503は、導電体518と同様の構成であり、絶縁体514および絶縁体516の開口の内壁に接して導電体503aが形成され、さらに内側に導電体503bが形成されている。なお、トランジスタ500では、導電体503aおよび導電体503bを積層する構成について示しているが、本発明はこれに限られるものではない。例えば、導電体503は、単層、または3層以上の積層構成として設ける構成にしてもよい。

【0236】

ここで、導電体503aは、水素原子、水素分子、水分子、銅原子などの不純物の拡散を抑制する機能を有する（上記不純物が透過しにくい。）導電性材料を用いることが好ましい。または、酸素（例えば、酸素原子、酸素分子などの少なくとも一）の拡散を抑制する機能を有する（上記酸素が透過しにくい。）導電性材料を用いることが好ましい。なお、本明細書において、不純物、または酸素の拡散を抑制する機能とは、上記不純物、または上記酸素のいずれか一または、すべての拡散を抑制する機能とする。

【0237】

例えば、導電体503aが酸素の拡散を抑制する機能を持つことにより、導電体503bが酸化して導電率が低下することを抑制することができる。

【0238】

また、導電体503が配線の機能を兼ねる場合、導電体503bは、タングステン、銅、またはアルミニウムを主成分とする、導電性が高い導電性材料を用いることが好ましい。なお、本実施の形態では導電体503を導電体503aと導電体503bの積層で図示したが、導電体503は単層構成であってもよい。

【0239】

絶縁体522、および絶縁体524は、第2のゲート絶縁膜としての機能を有する。

【0240】

ここで、酸化物530と接する絶縁体524は、化学量論的組成を満たす酸素よりも多くの酸素を含む絶縁体を用いることが好ましい。当該酸素は、加熱により膜中から放出されやすい。本明細書などでは、加熱により放出される酸素を「過剰酸素」と呼ぶ場合がある。つまり、絶縁体524には、過剰酸素を含む領域（「過剰酸素領域」ともいう。）が形成されていることが好ましい。このような過剰酸素を含む絶縁体を酸化物530に接して設けることにより、酸化物530中の酸素欠損（*VO: oxygen vacancy*ともいう）を低減し、トランジスタ500の信頼性を向上させることができる。なお、酸化物530中の酸素欠損に水素が入った場合、当該欠陥（以下、*VOH*と呼ぶ場合がある。）はドナーとして機能し、キャリアである電子が生成されることがある。また、水素の一部が金属原子と結合する酸素と結合して、キャリアである電子を生成する場合がある。従って、水素が多く含まれている酸化物半導体を用いたトランジスタは、ノーマリーオン特性となりやすい。また、酸化物半導体中の水素は、熱、電界などのストレスによって動きやすいため、酸化物半導体に多くの水素が含まれると、トランジスタの信頼性が悪化する恐れもある。本発明の一態様においては、酸化物530中の*VOH*をできる限り低減し、高純度真性または実質的に高純度真性にするのが好ましい。このように、*VOH*が十分低減された酸化物半導体を得るには、酸化物半導体中の水分、水素などの不純物を除去すること（「脱水」または「脱水素化处理」ともいう。）と、酸化物半導体に酸素を供給して酸素欠損を補填すること（「加酸素化处理」ともいう。）が重要である。*VOH*などが十分に低減された酸化物半導体をトランジスタのチャンネル形成領域に用いることで、安定した電気特性を付与することができる。

10

20

30

40

50

【0241】

過剰酸素領域を有する絶縁体として、具体的には、加熱により一部の酸素が脱離する酸化物材料を用いることが好ましい。加熱により酸素を脱離する酸化物とは、TDS (Thermal Desorption Spectroscopy) 分析にて、酸素原子に換算しての酸素の脱離量が $1.0 \times 10^{18} \text{ atoms/cm}^3$ 以上、好ましくは $1.0 \times 10^{19} \text{ atoms/cm}^3$ 以上、さらに好ましくは $2.0 \times 10^{19} \text{ atoms/cm}^3$ 以上、または $3.0 \times 10^{20} \text{ atoms/cm}^3$ 以上である酸化物膜である。なお、上記TDS分析時における膜の表面温度としては 100°C 以上 700°C 以下、または 100°C 以上 400°C 以下の範囲が好ましい。

【0242】

また、上記過剰酸素領域を有する絶縁体と、酸化物530と、を接して加熱処理、マイクロ波処理、またはRF処理のいずれか一または複数の処理を行っても良い。当該処理を行うことで、酸化物530中の水、または水素を除去することができる。例えば、酸化物530において、V-O-Hの結合が切断される反応が起きる、別言すると「 $\text{V-O-H} \rightarrow \text{V-O} + \text{H}$ 」という反応が起きて、脱水素化することができる。このとき発生した水素の一部は、酸素と結合して H_2O として、酸化物530、または酸化物530近傍の絶縁体から除去される場合がある。また、水素の一部は、導電体542にゲッターリングされる場合がある。

【0243】

また、上記マイクロ波処理は、例えば、高密度プラズマを発生させる電源を有する装置、または、基板側にRFを印加する電源を有する装置を用いると好適である。例えば、酸素を含むガスを用い、且つ高密度プラズマを用いることより、高密度の酸素ラジカルを生成することができる。また、基板側にRFを印加することで、高密度プラズマによって生成された酸素ラジカルを、効率よく酸化物530、または酸化物530近傍の絶縁体中に導入することができる。また、上記マイクロ波処理は、圧力を 133 Pa 以上、好ましくは 200 Pa 以上、さらに好ましくは 400 Pa 以上とすればよい。また、マイクロ波処理を行う装置内に導入するガスとしては、例えば、酸素と、アルゴンとを用い、酸素流量比 ($\text{O}_2 / (\text{O}_2 + \text{Ar})$) が 50% 以下、好ましくは 10% 以上 30% 以下で行うとよい。

【0244】

また、トランジスタ500の作製工程中において、酸化物530の表面が露出した状態で、加熱処理を行うと好適である。当該加熱処理は、例えば、 100°C 以上 450°C 以下、より好ましくは 350°C 以上 400°C 以下で行えばよい。なお、加熱処理は、窒素ガスもしくは不活性ガスの雰囲気、または酸化性ガスを 10 ppm 以上、 1% 以上、もしくは 10% 以上含む雰囲気で行う。例えば、加熱処理は酸素雰囲気で行うことが好ましい。これにより、酸化物530に酸素を供給して、酸素欠損 (V-O) の低減を図ることができる。また、加熱処理は減圧状態で行ってもよい。または、加熱処理は、窒素ガスもしくは不活性ガスの雰囲気で行った後に、脱離した酸素を補うために、酸化性ガスを 10 ppm 以上、 1% 以上、または 10% 以上含む雰囲気で行ってもよい。または、酸化性ガスを 10 ppm 以上、 1% 以上、または 10% 以上含む雰囲気で行った後に、連続して窒素ガスもしくは不活性ガスの雰囲気で行った加熱処理を行ってもよい。

【0245】

なお、酸化物530に加酸素化処理を行うことで、酸化物530中の酸素欠損を、供給された酸素により修復させる、別言すると「 $\text{V-O} + \text{O} \rightarrow \text{null}$ 」という反応を促進させることができる。さらに、酸化物530中に残存した水素に供給された酸素が反応することで、当該水素を H_2O として除去する (脱水化する) ことができる。これにより、酸化物530中に残存していた水素が酸素欠損に再結合してV-O-Hが形成されるのを抑制することができる。

【0246】

また、絶縁体524が、過剰酸素領域を有する場合、絶縁体522は、酸素 (例えば、酸素原子、酸素分子など) の拡散を抑制する機能を有する (上記酸素が透過しにくい) ことが好ましい。

【0247】

絶縁体522が、酸素や不純物の拡散を抑制する機能を有することで、酸化物530が有する酸素は、導電体503側へ拡散することがなく、好ましい。また、導電体503が、絶縁体524や、酸化物530が有する酸素と反応することを抑制することができる。

【0248】

絶縁体522は、例えば、酸化アルミニウム、酸化ハフニウム、アルミニウムおよびハフニウムを含む酸化物（ハフニウムアルミネート）、酸化タンタル、酸化ジルコニウム、チタン酸ジルコン酸鉛（PZT）、チタン酸ストロンチウム（ SrTiO_3 ）、または（Ba, Sr） TiO_3 （BST）などのいわゆるh i g h - k材料を含む絶縁体を単層または積層で用いることが好ましい。トランジスタの微細化、および高集積化が進むと、ゲート絶縁膜の薄膜化により、リーク電流などの問題が生じる場合がある。ゲート絶縁膜として機能する絶縁体にh i g h - k材料を用いることで、物理膜厚を保ちながら、トランジスタ動作時のゲート電位の低減が可能となる。

【0249】

特に、不純物、および酸素などの拡散を抑制する機能を有する（上記酸素が透過しにくい）絶縁性材料であるアルミニウム、ハフニウム的一方または双方の酸化物を含む絶縁体を用いるとよい。アルミニウム、ハフニウム的一方または双方の酸化物を含む絶縁体として、酸化アルミニウム、酸化ハフニウム、アルミニウムおよびハフニウムを含む酸化物（ハフニウムアルミネート）などを用いることが好ましい。このような材料を用いて絶縁体522を形成した場合、絶縁体522は、酸化物530からの酸素の放出や、トランジスタ500の周辺部から酸化物530への水素等の不純物の混入を抑制する層として機能する。

【0250】

または、これらの絶縁体に、例えば、酸化アルミニウム、酸化ビスマス、酸化ゲルマニウム、酸化ニオブ、酸化シリコン、酸化チタン、酸化タングステン、酸化イットリウム、酸化ジルコニウムを添加してもよい。またはこれらの絶縁体を窒化処理してもよい。上記の絶縁体に酸化シリコン、酸化窒化シリコンまたは窒化シリコンを積層して用いてもよい。

【0251】

なお、図21Aおよび図21Bのトランジスタ500では、3層の積層構成からなる第2のゲート絶縁膜として、絶縁体522、および絶縁体524が図示されているが、第2のゲート絶縁膜は、単層、2層、または4層以上の積層構成を有していてもよい。その場合、同じ材料からなる積層構成に限定されず、異なる材料からなる積層構成でもよい。

【0252】

トランジスタ500は、チャネル形成領域を含む酸化物530に、酸化物半導体として機能する金属酸化物を用いる。例えば、酸化物530として、In-M-Zn酸化物（元素Mは、アルミニウム、ガリウム、イットリウム、銅、バナジウム、ベリリウム、ホウ素、チタン、鉄、ニッケル、ゲルマニウム、ジルコニウム、モリブデン、ランタン、セリウム、ネオジム、ハフニウム、タンタル、タングステン、またはマグネシウムなどから選ばれた一種、または複数種）等の金属酸化物を用いるとよい。

【0253】

酸化物半導体として機能する金属酸化物の形成は、スパッタリング法で行なってもよいし、ALD（Atomic Layer Deposition）法で行なってもよい。なお、酸化物半導体として機能する金属酸化物については、他の実施の形態で詳細に説明する。

【0254】

また、酸化物530においてチャネル形成領域にとして機能する金属酸化物は、バンドギャップが2 eV以上、好ましくは2.5 eV以上のものを用いることが好ましい。このように、バンドギャップの大きい金属酸化物を用いることで、トランジスタのオフ電流を低減することができる。

【0255】

酸化物 530 は、酸化物 530 b 下に酸化物 530 a を有することで、酸化物 530 a よりも下方に形成された構成物から、酸化物 530 b への不純物の拡散を抑制することができる。

【0256】

なお、酸化物 530 は、各金属原子の原子数比が異なる複数の酸化物層の積層構成を有することが好ましい。具体的には、酸化物 530 a に用いる金属酸化物において、構成元素中の元素 M の原子数比が、酸化物 530 b に用いる金属酸化物における、構成元素中の元素 M の原子数比より、大きいことが好ましい。また、酸化物 530 a に用いる金属酸化物において、I n に対する元素 M の原子数比が、酸化物 530 b に用いる金属酸化物における、I n に対する元素 M の原子数比より大きいことが好ましい。また、酸化物 530 b に用いる金属酸化物において、元素 M に対する I n の原子数比が、酸化物 530 a に用いる金属酸化物における、元素 M に対する I n の原子数比より大きいことが好ましい。

10

【0257】

また、酸化物 530 a の伝導帯下端のエネルギーが、酸化物 530 b の伝導帯下端のエネルギーより高くなることが好ましい。また、言い換えると、酸化物 530 a 電子親和力が、酸化物 530 b の電子親和力より小さいことが好ましい。

【0258】

ここで、酸化物 530 a および酸化物 530 b の接合部において、伝導帯下端のエネルギー準位はなだらかに変化する。換言すると、酸化物 530 a および酸化物 530 b の接合部における伝導帯下端のエネルギー準位は、連続的に変化または連続接合するともいうことができる。このようにするためには、酸化物 530 a と酸化物 530 b との界面において形成される混合層の欠陥準位密度を低くするとよい。

20

【0259】

具体的には、酸化物 530 a と酸化物 530 b が、酸素以外に共通の元素を有する（主成分とする）ことで、欠陥準位密度が低い混合層を形成することができる。例えば、酸化物 530 b が I n - G a - Z n 酸化物の場合、酸化物 530 a として、I n - G a - Z n 酸化物、G a - Z n 酸化物、酸化ガリウムなどを用いるとよい。

【0260】

このとき、キャリアの主たる経路は酸化物 530 b となる。酸化物 530 a を上述の構成とすることで、酸化物 530 a と酸化物 530 b との界面における欠陥準位密度を低くすることができる。そのため、界面散乱によるキャリア伝導への影響が小さくなり、トランジスタ 500 は高いオン電流を得られる。

30

【0261】

酸化物 530 b 上には、ソース電極、およびドレイン電極として機能する導電体 542 a、および導電体 542 b が設けられる。導電体 542 a、および導電体 542 b としては、アルミニウム、クロム、銅、銀、金、白金、タンタル、ニッケル、チタン、モリブデン、タングステン、ハフニウム、バナジウム、ニオブ、マンガン、マグネシウム、ジルコニウム、ベリリウム、インジウム、ルテニウム、イリジウム、ストロンチウム、ランタンから選ばれた金属元素、または上述した金属元素を成分とする合金か、上述した金属元素を組み合わせた合金等を用いることが好ましい。例えば、窒化タンタル、窒化チタン、タングステン、チタンとアルミニウムを含む窒化物、タンタルとアルミニウムを含む窒化物、酸化ルテニウム、窒化ルテニウム、ストロンチウムとルテニウムを含む酸化物、ランタンとニッケルを含む酸化物などを用いることが好ましい。また、窒化タンタル、窒化チタン、チタンとアルミニウムを含む窒化物、タンタルとアルミニウムを含む窒化物、酸化ルテニウム、窒化ルテニウム、ストロンチウムとルテニウムを含む酸化物、ランタンとニッケルを含む酸化物は、酸化しにくい導電性材料、または、酸素を吸収しても導電性を維持する材料であるため、好ましい。更に、窒化タンタルなどの金属窒化物膜は、水素または酸素に対するバリア性があるため好ましい。

40

【0262】

また、図 21 A では、導電体 542 a、および導電体 542 b を単層構成として示した

50

が、２層以上の積層構成としてもよい。例えば、窒化タンタル膜とタングステン膜を積層するとよい。また、チタン膜とアルミニウム膜を積層してもよい。また、タングステン膜上にアルミニウム膜を積層する二層構成、銅－マグネシウム－アルミニウム合金膜上に銅膜を積層する二層構成、チタン膜上に銅膜を積層する二層構成、タングステン膜上に銅膜を積層する二層構成としてもよい。

【０２６３】

また、チタン膜または窒化チタン膜と、そのチタン膜または窒化チタン膜上に重ねてアルミニウム膜または銅膜を積層し、さらにその上にチタン膜または窒化チタン膜を形成する三層構成、モリブデン膜または窒化モリブデン膜と、そのモリブデン膜または窒化モリブデン膜上に重ねてアルミニウム膜または銅膜を積層し、さらにその上にモリブデン膜または窒化モリブデン膜を形成する三層構成等がある。なお、酸化インジウム、酸化錫または酸化亜鉛を含む透明導電材料を用いてもよい。

【０２６４】

また、図２１Ａに示すように、酸化物５３０の、導電体５４２ａ（導電体５４２ｂ）との界面とその近傍には、低抵抗領域として、領域５４３ａ、および領域５４３ｂが形成される場合がある。このとき、領域５４３ａはソース領域またはドレイン領域の一方として機能し、領域５４３ｂはソース領域またはドレイン領域の他方として機能する。また、領域５４３ａと領域５４３ｂに挟まれる領域にチャンネル形成領域が形成される。

【０２６５】

酸化物５３０と接するように上記導電体５４２ａ（導電体５４２ｂ）を設けることで、領域５４３ａ（領域５４３ｂ）の酸素濃度が低減する場合がある。また、領域５４３ａ（領域５４３ｂ）に導電体５４２ａ（導電体５４２ｂ）に含まれる金属と、酸化物５３０の成分とを含む金属化合物層が形成される場合がある。このような場合、領域５４３ａ（領域５４３ｂ）のキャリア密度が増加し、領域５４３ａ（領域５４３ｂ）は、低抵抗領域となる。

【０２６６】

絶縁体５４４は、導電体５４２ａ、および導電体５４２ｂを覆うように設けられ、導電体５４２ａ、および導電体５４２ｂの酸化を抑制する。このとき、絶縁体５４４は、酸化物５３０の側面を覆い、絶縁体５２４と接するように設けられてもよい。

【０２６７】

絶縁体５４４として、ハフニウム、アルミニウム、ガリウム、イットリウム、ジルコニウム、タングステン、チタン、タンタル、ニッケル、ゲルマニウム、ネオジム、ランタンまたは、マグネシウムなどから選ばれた一種、または二種以上が含まれた金属酸化物を用いることができる。また、絶縁体５４４として、窒化酸化シリコンまたは窒化シリコンなども用いることができる。

【０２６８】

特に、絶縁体５４４として、アルミニウム、またはハフニウムの一方または双方の酸化物を含む絶縁体である、酸化アルミニウム、酸化ハフニウム、アルミニウム、およびハフニウムを含む酸化物（ハフニウムアルミネート）などを用いることが好ましい。特に、ハフニウムアルミネートは、酸化ハフニウム膜よりも、耐熱性が高い。そのため、後の工程での熱処理において、結晶化しにくいため好ましい。なお、導電体５４２ａ、および導電体５４２ｂが耐酸化性を有する材料、または、酸素を吸収しても著しく導電性が低下しない場合、絶縁体５４４は、必須の構成ではない。求めるトランジスタ特性により、適宜設計すればよい。

【０２６９】

絶縁体５４４を有することで、絶縁体５８０に含まれる水、および水素などの不純物が絶縁体５４５を介して、酸化物５３０ｂに拡散することを抑制することができる。また、絶縁体５８０が有する過剰酸素により、導電体５６０が酸化するのを抑制することができる。

【０２７０】

絶縁体 5 4 5 は、第 1 のゲート絶縁膜として機能する。絶縁体 5 4 5 は、上述した絶縁体 5 2 4 と同様に、過剰に酸素を含み、かつ加熱により酸素が放出される絶縁体を用いて形成することが好ましい。

【0271】

具体的には、過剰酸素を有する酸化シリコン、酸化窒化シリコン、窒化酸化シリコン、窒化シリコン、フッ素を添加した酸化シリコン、炭素を添加した酸化シリコン、炭素、および窒素を添加した酸化シリコン、空孔を有する酸化シリコンを用いることができる。特に、酸化シリコン、および酸化窒化シリコンは熱に対し安定であるため好ましい。

【0272】

過剰酸素を含む絶縁体を絶縁体 5 4 5 として設けることにより、絶縁体 5 4 5 から、酸化物 5 3 0 b のチャンネル形成領域に効果的に酸素を供給することができる。また、絶縁体 5 2 4 と同様に、絶縁体 5 4 5 中の水または水素などの不純物濃度が低減されていることが好ましい。絶縁体 5 4 5 の膜厚は、1 nm 以上 20 nm 以下とするのが好ましい。また、絶縁体 5 4 5 の形成前および／または形成後に、前述したマイクロ波処理を行なってもよい。

【0273】

また、絶縁体 5 4 5 が有する過剰酸素を、効率的に酸化物 5 3 0 へ供給するために、絶縁体 5 4 5 と導電体 5 6 0 との間に金属酸化物を設けてもよい。当該金属酸化物は、絶縁体 5 4 5 から導電体 5 6 0 への酸素拡散を抑制することが好ましい。酸素の拡散を抑制する金属酸化物を設けることで、絶縁体 5 4 5 から導電体 5 6 0 への過剰酸素の拡散が抑制される。つまり、酸化物 5 3 0 へ供給する過剰酸素量の減少を抑制することができる。また、過剰酸素による導電体 5 6 0 の酸化を抑制することができる。当該金属酸化物としては、絶縁体 5 4 4 に用いることができる材料を用いればよい。

【0274】

なお、絶縁体 5 4 5 は、第 2 のゲート絶縁膜と同様に、積層構成としてもよい。トランジスタの微細化、および高集積化が進むと、ゲート絶縁膜の薄膜化により、リーク電流などの問題が生じる場合があるため、ゲート絶縁膜として機能する絶縁体を、high-k 材料と、熱的に安定している材料との積層構成とすることで、物理膜厚を保ちながら、トランジスタ動作時のゲート電位の低減が可能となる。また、熱的に安定かつ比誘電率の高い積層構成とすることができる。

【0275】

第 1 のゲート電極として機能する導電体 5 6 0 は、図 2 1 A および図 2 1 B では 2 層構成として示しているが、単層構成でもよいし、3 層以上の積層構成であってもよい。

【0276】

導電体 5 6 0 a は、水素原子、水素分子、水分子、窒素原子、窒素分子、酸化窒素分子 (N_2O 、 NO 、 NO_2 など)、銅原子などの不純物の拡散を抑制する機能を有する導電性材料を用いることが好ましい。または、酸素（例えば、酸素原子、酸素分子などの少なくとも一）の拡散を抑制する機能を有する導電性材料を用いることが好ましい。導電体 5 6 0 a が酸素の拡散を抑制する機能を持つことにより、絶縁体 5 4 5 に含まれる酸素により、導電体 5 6 0 b が酸化して導電率が低下することを抑制することができる。酸素の拡散を抑制する機能を有する導電性材料としては、例えば、タンタル、窒化タンタル、ルテニウム、または酸化ルテニウムなどを用いることが好ましい。また、導電体 5 6 0 a として、酸化物 5 3 0 に適用できる酸化物半導体を用いることができる。その場合、導電体 5 6 0 b をスパッタリング法で成膜することで、導電体 5 6 0 a の電気抵抗値を低下させて導電体にすることができる。これを OC (Oxide Conductor) 電極と呼ぶことができる。

【0277】

また、導電体 5 6 0 b は、タングステン、銅、またはアルミニウムを主成分とする導電性材料を用いることが好ましい。また、導電体 5 6 0 b は、配線としても機能するため、導電性が高い導電体を用いることが好ましい。例えば、タングステン、銅、またはアルミ

ニウムを主成分とする導電性材料を用いることができる。また、導電体 560b は積層構成としてもよく、例えば、チタン又は窒化チタンと上記導電性材料との積層構成としてもよい。

【0278】

絶縁体 580 は、絶縁体 544 を介して、導電体 542a、および導電体 542b 上に設けられる。絶縁体 580 は、過剰酸素領域を有することが好ましい。例えば、絶縁体 580 として、酸化シリコン、酸化窒化シリコン、窒化酸化シリコン、窒化シリコン、フッ素を添加した酸化シリコン、炭素を添加した酸化シリコン、炭素、および窒素を添加した酸化シリコン、空孔を有する酸化シリコン、または樹脂などを有することが好ましい。特に、酸化シリコン、および酸化窒化シリコンは、熱的に安定であるため好ましい。特に、酸化シリコン、空孔を有する酸化シリコンは、後の工程で、容易に過剰酸素領域を形成することができるため好ましい。

10

【0279】

絶縁体 580 は、過剰酸素領域を有することが好ましい。加熱により酸素が放出される絶縁体 580 を設けることで、絶縁体 580 中の酸素を酸化物 530 へと効率良く供給することができる。なお、絶縁体 580 中の水または水素などの不純物濃度が低減されていることが好ましい。

【0280】

絶縁体 580 の開口は、導電体 542a と導電体 542b の間の領域に重畳して形成される。これにより、導電体 560 は、絶縁体 580 の開口、および導電体 542a と導電体 542b に挟まれた領域に、埋め込まれるように形成される。

20

【0281】

半導体装置を微細化するに当たり、ゲート長を短くすることが求められるが、導電体 560 の導電性が下がらないようにする必要がある。そのために導電体 560 の膜厚を大きくすると、導電体 560 はアスペクト比が高い形状となりうる。本実施の形態では、導電体 560 を絶縁体 580 の開口に埋め込むように設けるため、導電体 560 をアスペクト比の高い形状にしても、工程中に導電体 560 を倒壊させることなく、形成することができる。

【0282】

絶縁体 574 は、絶縁体 580 の上面、導電体 560 の上面、および絶縁体 545 の上面に接して設けられることが好ましい。絶縁体 574 をスパッタリング法で成膜することで、絶縁体 545、および絶縁体 580 へ過剰酸素領域を設けることができる。これにより、当該過剰酸素領域から、酸化物 530 中に酸素を供給することができる。

30

【0283】

例えば、絶縁体 574 として、ハフニウム、アルミニウム、ガリウム、イットリウム、ジルコニウム、タングステン、チタン、タンタル、ニッケル、ゲルマニウム、またはマグネシウムなどから選ばれた一種、または二種以上が含まれた金属酸化物を用いることができる。

【0284】

特に、酸化アルミニウムはバリア性が高く、0.5nm 以上 3.0nm 以下の薄膜であっても、水素、および窒素の拡散を抑制することができる。したがって、スパッタリング法で成膜した酸化アルミニウムは、酸素供給源であるとともに、水素などの不純物のバリア膜としての機能も有することができる。

40

【0285】

また、絶縁体 574 の上に、層間膜として機能する絶縁体 581 を設けることが好ましい。絶縁体 581 は、絶縁体 524 などと同様に、膜中の水または水素などの不純物濃度が低減されていることが好ましい。

【0286】

また、絶縁体 581、絶縁体 574、絶縁体 580、および絶縁体 544 に形成された開口に、導電体 540a、および導電体 540b を配置する。導電体 540a および導電

50

体 5 4 0 b は、導電体 5 6 0 を挟んで対向して設ける。導電体 5 4 0 a および導電体 5 4 0 b は、後述する導電体 5 4 6、および導電体 5 4 8 と同様の構成である。

【0287】

絶縁体 5 8 1 上には、絶縁体 5 8 2 が設けられている。絶縁体 5 8 2 は、酸素や水素に対してバリア性のある物質を用いることが好ましい。したがって、絶縁体 5 8 2 には、絶縁体 5 1 4 と同様の材料を用いることができる。例えば、絶縁体 5 8 2 には、酸化アルミニウム、酸化ハフニウム、酸化タンタルなどの金属酸化物を用いることが好ましい。

【0288】

特に、酸化アルミニウムは、酸素、およびトランジスタの電気特性の変動要因となる水素、水分などの不純物、の両方に対して膜を透過させない遮断効果が高い。したがって、酸化アルミニウムは、トランジスタの作製工程中および作製後において、水素、水分などの不純物のトランジスタ 5 0 0 への混入を防止することができる。また、トランジスタ 5 0 0 を構成する酸化物からの酸素の放出を抑制することができる。そのため、トランジスタ 5 0 0 に対する保護膜として用いることに適している。

【0289】

また、絶縁体 5 8 2 上には、絶縁体 5 8 6 が設けられている。絶縁体 5 8 6 は、絶縁体 3 2 0 と同様の材料を用いることができる。また、これらの絶縁体に、比較的誘電率が低い材料を適用することで、配線間に生じる寄生容量を低減することができる。例えば、絶縁体 5 8 6 として、酸化シリコン膜や酸化窒化シリコン膜などを用いることができる。

【0290】

また、絶縁体 5 2 2、絶縁体 5 2 4、絶縁体 5 4 4、絶縁体 5 8 0、絶縁体 5 7 4、絶縁体 5 8 1、絶縁体 5 8 2、および絶縁体 5 8 6 には、導電体 5 4 6、および導電体 5 4 8 等が埋め込まれている。

【0291】

導電体 5 4 6、および導電体 5 4 8 は、容量素子 6 0 0、トランジスタ 5 0 0、またはトランジスタ 5 5 0 と接続するプラグ、または配線としての機能を有する。導電体 5 4 6、および導電体 5 4 8 は、導電体 3 2 8、および導電体 3 3 0 と同様の材料を用いて設けることができる。

【0292】

また、トランジスタ 5 0 0 の形成後、トランジスタ 5 0 0 を囲むように開口を形成し、当該開口を覆うように、水素、または水に対するバリア性が高い絶縁体を形成してもよい。上述のバリア性の高い絶縁体でトランジスタ 5 0 0 を包み込むことで、外部から水分、および水素が侵入するのを防止することができる。または、複数のトランジスタ 5 0 0 をまとめて、水素、または水に対するバリア性が高い絶縁体で包み込んでもよい。なお、トランジスタ 5 0 0 を囲むように開口を形成する場合、例えば、絶縁体 5 2 2 または絶縁体 5 1 4 に達する開口を形成し、絶縁体 5 2 2 または絶縁体 5 1 4 に接するように上述のバリア性の高い絶縁体を形成すると、トランジスタ 5 0 0 の作製工程の一部を兼ねられるため、好適である。なお、水素、または水に対するバリア性が高い絶縁体としては、例えば、絶縁体 5 2 2 または絶縁体 5 1 4 と同様の材料を用いればよい。

【0293】

続いて、トランジスタ 5 0 0 の上方には、容量素子 6 0 0 が設けられている。容量素子 6 0 0 は、導電体 6 1 0 と、導電体 6 2 0 と、絶縁体 6 3 0 とを有する。

【0294】

また、導電体 5 4 6、および導電体 5 4 8 上に、導電体 6 1 2 を設けてもよい。導電体 6 1 2 は、トランジスタ 5 0 0 と接続するプラグ、または配線としての機能を有する。導電体 6 1 0 は、容量素子 6 0 0 の電極としての機能を有する。なお、導電体 6 1 2、および導電体 6 1 0 は、同時に形成することができる。

【0295】

導電体 6 1 2、および導電体 6 1 0 には、モリブデン、チタン、タンタル、タングステン、アルミニウム、銅、クロム、ネオジム、スカンジウムから選ばれた元素を含む金属膜

10

20

30

40

50

、または上述した元素を成分とする金属窒化物膜（窒化タンタル膜、窒化チタン膜、窒化モリブデン膜、窒化タングステン膜）等を用いることができる。または、インジウム錫酸化物、酸化タングステンを含むインジウム酸化物、酸化タングステンを含むインジウム亜鉛酸化物、酸化チタンを含むインジウム酸化物、酸化チタンを含むインジウム錫酸化物、インジウム亜鉛酸化物、酸化ケイ素を添加したインジウム錫酸化物などの導電性材料を適用することもできる。

【0296】

本実施の形態では、導電体612、および導電体610を単層構成で示したが、当該構成に限定されず、2層以上の積層構成でもよい。例えば、バリア性を有する導電体と導電性が高い導電体との間に、バリア性を有する導電体、および導電性が高い導電体に対して密着性が高い導電体を形成してもよい。

10

【0297】

絶縁体630を介して、導電体610と重畳するように、導電体620を設ける。なお、導電体620は、金属材料、合金材料、または金属酸化物材料などの導電性材料を用いることができる。耐熱性と導電性を両立するタングステンやモリブデンなどの高融点材料を用いることが好ましく、特にタングステンを用いることが好ましい。また、導電体などの他の構成と同時に形成する場合は、低抵抗金属材料であるCu（銅）やAl（アルミニウム）等を用いればよい。

【0298】

導電体620、および絶縁体630上には、絶縁体640が設けられている。絶縁体640は、絶縁体320と同様の材料を用いて設けることができる。また、絶縁体640は、その下方の凹凸形状を被覆する平坦化膜として機能してもよい。

20

【0299】

本構成を用いることで、酸化物半導体を有するトランジスタを用いた半導体装置において、微細化または高集積化を図ることができる。

【0300】

本実施の形態に示す構成、構造、方法などは、他の実施の形態および実施例などに示す構成、構造、方法などと適宜組み合わせる用いることができる。

【0301】

（実施の形態6）

本実施の形態では、上記実施の形態で説明した演算処理システム100が有する各構成を含む集積回路の構成について図22A、図22Bを参照しながら説明する。

30

【0302】

図22Aは、演算処理システム100が有する各構成を含む集積回路を説明するための模式図の一例である。図22Aに図示する集積回路390は、CPU110および半導体装置10として説明したアクセラレータが有する回路の一部をOSTランジスタで構成することで、各回路を一体化した1つの集積回路とすることができる。

【0303】

図22Aに図示するように、CPU110において、CPUコア200の上層にあるOSTランジスタを有する層にバックアップ回路222を設ける構成とすることができる。また図22Aに図示するように、半導体装置10として説明したアクセラレータにおいて、演算回路30および切替回路40を構成するSiトランジスタを有する層の上層には、OSTランジスタを有する層にメモリ回路部20を設ける構成とすることができる。その他、Siトランジスタを有する層には駆動回路50、OSTランジスタを有する層にはOSメモリ300N等を設ける構成とすることができる。OSメモリ300Nとしては、上記実施の形態で説明したNOSRAMの他、DOSRAMを適用することができる。またOSメモリ300Nでは、Siトランジスタを有する層に設けられる駆動回路上にOSTランジスタを有する層を積層することで、メモリ密度の向上を図ることができる。

40

【0304】

図22Aに図示するように、CPU110、半導体装置10として説明したアクセラレ

50

ータおよびOSメモリ300N等の各回路を密結合させたSoCの場合、発熱の問題があるが、OSTランジスタは熱による電気特性の変動量がSitランジスタと比べて小さいため、好適である。また、図22Aに図示するように三次元方向において回路を集積化することによって、シリコン貫通電極(Through Silicon Via:TSV)などを用いた積層構造などと比較して寄生容量を小さくすることができる。各配線の充放電に要する消費電力を削減することができる。そのため、演算処理効率の向上を図ることができる。

【0305】

図22Bに、集積回路390を組み込んだ半導体チップの一例を示す。図22Bに示す半導体チップ391は、リード392及び集積回路390を有する。集積回路390は、図22Aで説明したように、上記実施の形態で示した各種の回路が1のダイに設けられている。集積回路390は積層構造をもち、Sitランジスタを有する層(Sitランジスタ層393)、配線層394、OSTランジスタを有する層(OSTランジスタ層395)に大別される。OSTランジスタ層395は、Sitランジスタ層393上に積層して設けることができるため、半導体チップ391の小型化が容易である。

【0306】

図22Bでは、半導体チップ391のパッケージにQFP(Quad Flat Package)を適用しているが、パッケージの態様はこれに限定されない。その他の構成例としては、挿入実装型であるDIP(Dual In-line Package)、PGA(Pin Grid Array)、表面実装型であるSOP(Small Outline Package)、SSOP(Shrink Small Outline Package)、TSOP(Thin-Small Outline Package)、LCC(Leaded Chip Carrier)、QFN(Quad Flat Non-leaded package)、BGA(Ball Grid Array)、FBGA(Fine pitch Ball Grid Array)、接触実装型であるDTP(Dual Tape carrier Package)、QTP(Quad Tape-carrier Package)等の構造を適宜用いることができる。

【0307】

Sitランジスタを有する演算回路および切替回路と、OSTランジスタを有するメモリ回路は、全て、Sitランジスタ層393、配線層394およびOSTランジスタ層395に形成することができる。すなわち、上記半導体装置を構成する素子は、同一の製造プロセスで形成することが可能である。そのため、図22Bに示すICは、構成する素子が増えても製造プロセスを増やす必要がなく、上記半導体装置を低コストで組み込むことができる。

【0308】

以上説明した本発明の一態様により、新規な半導体装置および電子機器を提供することができる。又は、本発明の一態様により、消費電力の小さい半導体装置および電子機器を提供することができる。又は、本発明の一態様により、発熱の抑制が可能な半導体装置および電子機器を提供することができる。

【0309】

本実施の形態は、他の実施の形態の記載と適宜組み合わせることができる。

【0310】

(実施の形態7)

本実施の形態では、上記実施の形態で説明した集積回路390を適用することが可能な電子機器、移動体、演算システムについて、図23乃至図26を参照しながら説明する。

【0311】

図23Aは、移動体の一例として自動車の外観図を図示している。図23Bは、自動車内でのデータのやり取りを簡略化した図である。自動車590は、複数のカメラ591等を有する。また、自動車590は、赤外線レーダー、ミリ波レーダー、レーザーレーダーなど各種センサ(図示せず)などを備える。

【0312】

自動車590において、カメラ591等の上記集積回路390（あるいは上記集積回路390を組み込んだ半導体チップ391）を用いることができる。自動車590は、カメラ591が複数の撮像方向592で得られた複数の画像を上記実施の形態で説明した集積回路390で処理し、バス593等を介してホストコントローラ594等により複数の画像をまとめて解析することで、ガードレールや歩行者の有無など、周囲の交通状況を判断し、自動運転を行うことができる。また、道路案内、危険予測などを行うシステムに用いることができる。

【0313】

集積回路390では、得られた画像データをニューラルネットワークなどの演算処理を行うことで、例えば、画像の高解像度化、画像ノイズの低減、顔認識（防犯目的など）、物体認識（自動運転の目的など）、画像圧縮、画像補正（広ダイナミックレンジ化）、レンズレスイメージセンサの画像復元、位置決め、文字認識、反射映り込み低減などの処理を行うことができる。

10

【0314】

なお、上述では、移動体の一例として自動車について説明しているが、移動体は自動車に限定されない。例えば、移動体としては、電車、モノレール、船、飛行体（ヘリコプター、無人航空機（ドローン）、飛行機、ロケット）なども挙げることができ、これらの移動体に本発明の一態様のコンピュータを適用して、人工知能を利用したシステムを付与することができる。

20

【0315】

図24Aは、携帯型電子機器の一例を示す外観図である。図24Bは、携帯型電子機器内でのデータのやり取りを簡略化した図である。携帯型電子機器595は、プリント配線基板596、スピーカー597、カメラ598、マイクロフォン599等を有する。

【0316】

携帯型電子機器595において、プリント配線基板596に上記集積回路390を設けることができる。携帯型電子機器595は、スピーカー597、カメラ598、マイクロフォン599等で得られる複数のデータを上記実施の形態で説明した集積回路390を用いて処理・解析することで、ユーザの利便性を向上させることができる。また、音声案内、画像検索などを行うシステムに用いることができる。

30

【0317】

集積回路390では、得られた画像データをニューラルネットワークなどの演算処理を行うことで、例えば、画像の高解像度化、画像ノイズの低減、顔認識（防犯目的など）、物体認識（自動運転の目的など）、画像圧縮、画像補正（広ダイナミックレンジ化）、レンズレスイメージセンサの画像復元、位置決め、文字認識、反射映り込み低減などの処理を行うことができる。

【0318】

図25Aに示す携帯型ゲーム機1100は、筐体1101、筐体1102、筐体1103、表示部1104、接続部1105、操作キー1107等を有する。筐体1101、筐体1102および筐体1103は、取り外すことが可能である。筐体1101に設けられている接続部1105を筐体1108に取り付けることで、表示部1104に出力される映像を、別の映像機器に出力することができる。他方、筐体1102および筐体1103を筐体1109に取り付けることで、筐体1102および筐体1103を一体化し、操作部として機能させる。筐体1102および筐体1103の基板に設けられているチップなどに先の実施の形態に示す集積回路390を組み込むことができる。

40

【0319】

図25BはUSB接続タイプのスティック型の電子機器1120である。電子機器1120は、筐体1121、キャップ1122、USBコネクタ1123および基板1124を有する。基板1124は、筐体1121に収納されている。例えば、基板1124には、メモリチップ1125、コントローラチップ1126が取り付けられている。基板11

50

24のコントローラチップ1126などに先の実施の形態に示す集積回路390を組み込むことができる。

【0320】

図25Cは人型のロボット1130である。ロボット1130は、センサ2101乃至2106、および制御回路2110を有する。例えば、制御回路2110には、先の実施の形態に示す集積回路390を組み込むことができる。

【0321】

上記実施の形態で説明した集積回路390は、電子機器に内蔵する代わりに、電子機器と通信を行うサーバーに用いることもできる。この場合、電子機器とサーバーによって演算システムが構成される。図26に、システム3000の構成例を示す。

【0322】

システム3000は、電子機器3001と、サーバー3002によって構成される。電子機器3001とサーバー3002間の通信は、インターネット回線3003を介して行うことができる。

【0323】

サーバー3002には、複数のラック3004を有する。複数のラックには、複数の基板3005が設けられ、当該基板3005上に上記実施の形態で説明した集積回路390を搭載することができる。これにより、サーバー3002にニューラルネットワークが構成される。そして、サーバー3002は、電子機器3001からインターネット回線3003を介して入力されたデータを用いて、ニューラルネットワークの演算を行うことができる。サーバー3002による演算の結果は必要に応じて、インターネット回線3003を介して電子機器3001に送信することができる。これにより、電子機器3001における演算の負担を低減することができる。

【0324】

本実施の形態は、他の実施の形態の記載と適宜組み合わせることができる。

【0325】

(実施の形態8)

本実施の形態では、半導体装置10を有する集積回路390において、畳み込みニューラルネットワーク(Convolutional neural network; 以下、CNN)などにおける畳み込み演算処理において用いる重みデータの構成例について図27および図28を参照して説明する。

【0326】

図27Aでは、学習(訓練)用データの入力によって、CNNの結合パラメータである重みデータが生成される様子を示す概念図である。図27Aには、サーバー31に記憶された学習用データ D_{TR} 、学習用データ D_{TR} が入力されるコンピュータ装置32を図示している。また、図27Aでは、学習用データ D_{TR} に対し、重みデータ34(W_{TR})を用いて行われる、積和演算などの処理33Aおよび活性化関数などの処理33Bを経て得られる学習用の畳み込みデータ D_{CT} を図示している。

【0327】

学習用データ D_{TR} は、音声データ、画像データ、あるいはテキストデータなどに相当する。それぞれのデータは、コンピュータ装置32内で処理が容易になるよう機械学習の内容に適したデータサイズやフォーマットに規格化されたデータであることが好ましい。重みデータ34(W_{TR})は、学習用データ D_{TR} の誤差逆伝播法(バックプロパゲーション)などによる演算処理によって生成される。学習用データ D_{TR} を処理するコンピュータ装置32は、安定した電力供給が可能な据え置き型であるため、膨大なメモリ、および演算性能が高い演算装置を用いた、消費電力の大きい演算処理を実行することができる。そのため、学習用データ D_{TR} のビット数は、16ビット乃至64ビットといった高いビット数のデータを用いて高精度な重みデータ34(W_{TR})の最適化を図ることができる。また、計算アルゴリズムによってはデータのビット精度が計算の収束性に影響することもあるため、幅広いビット数で演算できることが好ましい。

10

20

30

40

50

【0328】

図27Bでは、推論用データの入力によって、推論されたデータの出力を行う、CNNの演算処理が行われる様子を示す概念図である。図27Bでは、電子デバイス35などヘユーザが発話した音声データや、自動車36に搭載される撮像装置が取得する画像データなどを、推論用データ D_{IN} としている。推論用データ D_{IN} は、上記実施の形態で説明した半導体装置10を有する集積回路390に入力される。集積回路390では、推論用データ D_{IN} を入力データとして、メモリ回路に保持される重みデータ37 (W_{INF})を用いた畳み込み演算などの演算処理を行う。また、図27Bでは、推論用データ D_{IN} に対し、重みデータ37 (W_{INF})を用いて行われる、積和演算などの処理38Aおよび活性化関数などの処理38Bを経て得られる推論用の畳み込みデータ D_{CI} を図示している。集積回路390では、畳み込み演算処理等を含む演算処理を行うことで、推論された出力データ D_{JD} を出力する。

10

【0329】

推論用データ D_{IN} を処理する集積回路390は、処理能力の限られた環境で演算処理を行う。図27Aのコンピュータ装置32と比べ、回路リソースが少なくても済む演算処理だけを行う。集積回路390では、処理能力の限られた環境において、演算処理の高速化および低消費電力化が求められる。本発明の一態様の半導体装置10は、小型化、低消費電力化あるいは高速化に優れたアクセラレータとして機能する半導体装置とすることができる。そのため、エッジデバイスのように処理能力の限られた環境で用いられることに適している。

20

【0330】

なお推論用データ D_{IN} のビット数は、学習用データ D_{TR} のビット数より小さいことが好ましい。例えば、学習用データ D_{TR} を8ビット乃至64ビットといった高いビット数とする場合、集積回路390に入力される推論用データ D_{IN} は、16ビット以下、好ましくは8ビット以下、好ましくは4ビット以下、好ましくは2ビット以下のようなビット数（第1のビット数）の低いデータとする。つまり推論用のビット数は、学習用データ D_{TR} の高いビット数（第2のビット数）と比べて小さいことが好適である。

【0331】

同様に、集積回路390に保持される重みデータ37 (W_{INF})は、重みデータ34 (W_{TR})と比べて、16ビット以下、好ましくは8ビット以下、好ましくは4ビット以下、好ましくは2ビット以下のようなビット数の低いデータとすることが好ましい。当該構成とすることで、演算処理において限られたメモリ容量および演算性能しか実現できないような回路リソースが乏しい環境でも、精度の劣化の小さい演算を行うことができる。このような構成においては、ニューラルネットワークモデルに応じた、推論精度の劣化が小さい条件内でビット数の設定をすることが望ましい。

30

【0332】

重みデータ34 (W_{TR})から重みデータ37 (W_{INF})への変換は、各重みデータの相対的な関係を維持するよう規格化された処理によって、ビット数の削減によって行われる。たとえば、重みデータ34 (W_{TR})から重みデータ37 (W_{INF})へのビット数の削減は、指数部および／または仮数部のビット数を落とすことで実現できる。例えば、図28Aに図示する重みデータ W_{TR} から重みデータ W_{INF} への変換では、符号部39Aをそのままとして、指数部39Bおよび仮数部39Cのビット数を削減して、ビット数の低減された重みデータ W_{INF} としている。

40

【0333】

また図28Bに図示する重みデータ W_{TR} から重みデータ W_{INF} への変換では、符号部39Aおよび指数部39Bをそのままとして、仮数部39Cのビット数を大幅に削減して、ビット数の低減された重みデータ W_{INF} としている。

【0334】

また図28Aおよび図28B以外の構成としては、FP32などの浮動小数点形式を、INT8などの整数形式に変換することでビット数の削減を行うこともできる。

50

【0335】

ビット数の低減された重みデータ W_{INF} では、ビット数の低下による数値の丸め誤差が発生することや、表現可能な数値表現範囲が狭まる。一方で、ビット数を低下させても重みデータ同士の大小関係（相対関係）は維持することができるため、畳み込み演算処理による出力値の大小関係を維持される。そのため、ニューラルネットワークモデル次第で演算精度の低下が小さい演算処理の実行が可能である。また、エッジデバイスのように処理能力の限られた環境では、ビット数の低減された重みデータ W_{INF} を用いた推論処理が適している。

【0336】

なおニューラルネットワークモデルにおいて、層毎にビット幅の最適化を行う構成、あるいは重要性の低いニューロンを削減するといった最適化を行う構成、とすることも好ましい。当該構成とすることで、演算精度の低下を抑えつつ、演算量も低減することができる。

10

【0337】

（本明細書等の記載に関する付記）

以上の実施の形態、および実施の形態における各構成の説明について、以下に付記する。

【0338】

各実施の形態に示す構成は、他の実施の形態あるいは実施例に示す構成と適宜組み合わせ、本発明の一態様とすることができる。また、1つの実施の形態の中に、複数の構成例が示される場合は、構成例を適宜組み合わせることが可能である。

20

【0339】

なお、ある一つの実施の形態の中で述べる内容（一部の内容でもよい）は、その実施の形態で述べる別の内容（一部の内容でもよい）、および／または、一つ若しくは複数の別の実施の形態で述べる内容（一部の内容でもよい）に対して、適用、組み合わせ、または置き換えなどを行うことが出来る。

【0340】

なお、実施の形態の中で述べる内容とは、各々の実施の形態において、様々な図を用いて述べる内容、または明細書に記載される文章を用いて述べる内容のことである。

【0341】

なお、ある一つの実施の形態において述べる図（一部でもよい）は、その図の別の部分、その実施の形態において述べる別の図（一部でもよい）、および／または、一つ若しくは複数の別の実施の形態において述べる図（一部でもよい）に対して、組み合わせることにより、さらに多くの図を構成させることが出来る。

30

【0342】

また本明細書等において、ブロック図では、構成要素を機能毎に分類し、互いに独立したブロックとして示している。しかしながら実際の回路等においては、構成要素を機能毎に切り分けることが難しく、一つの回路に複数の機能が係わる場合や、複数の回路にわたって一つの機能が関わる場合があり得る。そのため、ブロック図のブロックは、明細書で説明した構成要素に限定されず、状況に応じて適切に言い換えることができる。

【0343】

また、図面において、大きさ、層の厚さ、または領域は、説明の便宜上任意の大きさに示したものである。よって、必ずしもそのスケールに限定されない。なお図面は明確性を期すために模式的に示したものであり、図面に示す形状または値などに限定されない。例えば、ノイズによる信号、電圧、若しくは電流のばらつき、または、タイミングのずれによる信号、電圧、若しくは電流のばらつきなどを含むことが可能である。

40

【0344】

また、図面等において図示する構成要素の位置関係は、相対的である。従って、図面を参照して構成要素を説明する場合、位置関係を示す「上に」、「下に」等の語句は便宜的に用いられる場合がある。構成要素の位置関係は、本明細書の記載内容に限定されず、状況に応じて適切に言い換えることができる。

50

【0345】

本明細書等において、トランジスタの接続関係を説明する際、「ソースまたはドレインの一方」（または第1電極、または第1端子）、ソースとドレインとの他方を「ソースまたはドレインの他方」（または第2電極、または第2端子）という表記を用いる。これは、トランジスタのソースとドレインは、トランジスタの構造または動作条件等によって変わるためである。なおトランジスタのソースとドレインの呼称については、ソース（ドレイン）端子や、ソース（ドレイン）電極等、状況に応じて適切に言い換えることができる。

【0346】

また、本明細書等において「電極」や「配線」の用語は、これらの構成要素を機能的に限定するものではない。例えば、「電極」は「配線」の一部として用いられることがあり、その逆もまた同様である。さらに、「電極」や「配線」の用語は、複数の「電極」や「配線」が一体となって形成されている場合なども含む。

10

【0347】

また、本明細書等において、電圧と電位は、適宜言い換えることができる。電圧は、基準となる電位からの電位差のことであり、例えば基準となる電位をグラウンド電圧（接地電圧）とすると、電圧を電位に言い換えることができる。グラウンド電位は必ずしも0Vを意味するとは限らない。なお電位は相対的なものであり、基準となる電位によっては、配線等に与える電位を変化させる場合がある。

【0348】

また本明細書等において、ノードは、回路構成やデバイス構造等に応じて、端子、配線、電極、導電層、導電体、不純物領域等と言い換えることが可能である。また、端子、配線等をノードと言い換えることが可能である。

20

【0349】

本明細書等において、AとBとが接続されている、とは、AとBとが電氣的に接続されているものをいう。ここで、AとBとが電氣的に接続されているとは、AとBとの間で対象物（スイッチ、トランジスタ素子、またはダイオード等の素子、あるいは当該素子および配線を含む回路等を指す）が存在する場合にAとBとの電気信号の伝達が可能である接続をいう。なおAとBとが電氣的に接続されている場合には、AとBとが直接接続されている場合を含む。ここで、AとBとが直接接続されているとは、上記対象物を介することなく、AとBとの間で配線（または電極）等を介してAとBとの電気信号の伝達が可能である接続をいう。換言すれば、直接接続とは、等価回路で表した際に同じ回路図として見なせる接続をいう。

30

【0350】

本明細書等において、スイッチとは、導通状態（オン状態）、または、非導通状態（オフ状態）になり、電流を流すか流さないかを制御する機能を有するものをいう。または、スイッチとは、電流を流す経路を選択して切り替える機能を有するものをいう。

【0351】

本明細書等において、チャンネル長とは、例えば、トランジスタの上面図において、半導体（またはトランジスタがオン状態のときに半導体の中で電流の流れる部分）とゲートとが重なる領域、またはチャンネルが形成される領域における、ソースとドレインとの間の距離をいう。

40

【0352】

本明細書等において、チャンネル幅とは、例えば、半導体（またはトランジスタがオン状態のときに半導体の中で電流の流れる部分）とゲート電極とが重なる領域、またはチャンネルが形成される領域における、ソースとドレインとが向かい合っている部分の長さをいう。

【0353】

なお本明細書等において、「膜」、「層」などの語句は、場合によっては、または、状況に応じて、互いに入れ替えることが可能である。例えば、「導電層」という用語を、「導電膜」という用語に変更することが可能な場合がある。または、例えば、「絶縁膜」という用語を、「絶縁層」という用語に変更することが可能な場合がある。

50

【符号の説明】

【0354】

A I N _ 1 : 入力データ、A I N : 入力データ、B G L : バックゲート線、B K : 信号、
 B K H : 信号、B L : ビット線、C 1 1 : 容量素子、C K : ノード、C L K : クロック信
 号、D I N : 推論用データ、D J D : 出力データ、D T R : 学習用データ、E N : 制御信
 号、G B L _ A : 配線、G B L _ B : 配線、G B L _ N : 配線、G B L _ P : 配線、G B
 L : 配線、G L [2] : 配線、G L : 配線、L B L _ 1 : 配線、L B L _ 7 : 配線、L B
 L _ N : 配線、L B L _ P : 配線、L B L : 配線、L B L P : 配線、M 1 1 : トランジス
 タ、M 1 2 : トランジスタ、M 1 3 : トランジスタ、M A C : 出力データ、R C : 信号、
 R C H : 信号、R T : ノード、R W L _ 1 : 読出用ワード線、R W L : 読出し用ワード線
 、S C E : 信号、S D _ I N : ノード、S D : ノード、S E : ノード、S L : ソース戦、
 S N 1 1 : ノード、W B L _ N : 書込み用ビット線、W B L _ P : 書込み用ビット線、W
 B L : 書込み用ビット線、W d a t a : 重みデータ、W I N F : 重みデータ、W L : ワー
 ド線、W S E L _ A : 重みデータ、W S E L _ B : 重みデータ、W S E L : 重みデータ、
 W T R : 重みデータ、W W L _ 1 : 書込用ワード線、W W L : 書込み用ワード線、1 0 _
 1 : 半導体装置、1 0 _ n : 半導体装置、1 0 : 半導体装置、1 1 : 層、1 2 : 層、2 0
 _ 1 : メモリ回路部、2 0 _ 4 : メモリ回路部、2 0 _ 6 : メモリ回路部、2 0 _ N : メ
 モリ回路部、2 0 _ N (N : メモリ回路部、2 0 : メモリ回路部、2 1 _ N : メモリ回路
 、2 1 _ P : メモリ回路、2 1 A : メモリ回路、2 1 B : メモリ回路、2 1 C : メモリ回
 路、2 1 : メモリ回路、2 2 : トランジスタ、2 3 : 半導体層、2 4 : 乗算回路、2 5 :
 加算回路、2 6 : レジスタ、3 0 _ 1 : 演算回路、3 0 _ 1 2 : 演算回路、3 0 _ 4 : 演
 算回路、3 0 _ 6 : 演算回路、3 0 _ 7 : 演算回路、3 0 _ N : 演算回路、3 0 : 演算回
 路、3 1 : サーバ、3 2 : コンピュータ装置、3 3 A : 処理、3 3 B : 処理、3 4 : 重
 みデータ、3 5 : 電子デバイス、3 6 : 自動車、3 7 : 重みデータ、3 8 A : 処理、3 8
 B : 処理、3 9 A : 符号部、3 9 B : 指数部、3 9 C : 仮数部、4 0 _ 1 : 切替回路、4
 0 _ 1 2 : 切替回路、4 0 _ 4 : 切替回路、4 0 _ 6 : 切替回路、4 0 _ 7 : 切替回路、
 4 0 A : 切替回路、4 0 B : 切替回路、4 0 M : 切替回路、4 0 X : 切替回路、4 0 Y :
 切替回路、4 0 : 切替回路、5 0 : 駆動回路、6 0 : メモリ回路、6 1 _ N : トランジス
 タ、6 1 _ P : トランジスタ、6 1 A : トランジスタ、6 1 B : トランジスタ、6 1 : ト
 ランジスタ、6 2 _ N : トランジスタ、6 2 _ P : トランジスタ、6 2 B : トランジスタ
 、6 2 : トランジスタ、6 3 _ N : トランジスタ、6 3 _ P : トランジスタ、6 3 : トラ
 ンジスタ、6 4 _ N : 容量素子、6 4 _ P : 容量素子、6 4 A : 容量素子、6 4 B : 容量
 素子、6 4 : 容量素子、7 1 G : コントローラ、7 1 : コントローラ、7 2 : ロウデコー
 ダ、7 3 : ワード線ドライバ、7 4 : カラムデコーダ、7 5 : 書き込みドライバ、7 6 :
 プリチャージ回路、8 1 : 入出力バッファ、8 2 : 演算制御回路、9 0 A : 入力層、9 0
 B : 中間層、9 0 C : 出力層、9 2 : 畳み込み演算処理、9 3 : 畳み込み演算処理、9 4
 : プーリング演算処理、9 5 : 畳み込み演算処理、9 6 : プーリング演算処理、1 0 0 :
 演算処理システム、1 1 0 : C P U、1 2 0 : バス、1 9 3 : P M U、2 0 0 : C P U コ
 ア、2 0 2 : L 1 キャッシュメモリ装置、2 0 3 : L 2 キャッシュメモリ装置、2 0 5 :
 バスインターフェース部、2 1 0 : パワースイッチ、2 1 1 : パワースイッチ、2 1 2 :
 パワースイッチ、2 1 4 : レベルシフタ、2 2 0 : フリップフロップ、2 2 1 A : クロッ
 クバッファ回路、2 2 1 : スキャンフリップフロップ、2 2 2 : バックアップ回路、3 0
 0 N : O S メモリ、3 1 1 : 基板、3 1 2 : ウェル領域、3 1 3 : 絶縁体、3 1 4 : 酸化
 物層、3 1 5 : 半導体領域、3 1 6 a : 低抵抗領域、3 1 6 b : 低抵抗領域、3 1 6 c :
 低抵抗領域、3 1 7 : 絶縁体、3 1 8 : 導電体、3 2 0 : 絶縁体、3 2 2 : 絶縁体、3 2
 4 : 絶縁体、3 2 6 : 絶縁体、3 2 8 : 導電体、3 3 0 : 導電体、3 5 0 : 絶縁体、3 5
 2 : 絶縁体、3 5 4 : 絶縁体、3 5 6 : 導電体、3 6 0 : 絶縁体、3 6 2 : 絶縁体、3 6
 4 : 絶縁体、3 6 6 : 導電体、3 7 0 : 絶縁体、3 7 2 : 絶縁体、3 7 4 : 絶縁体、3 7
 6 : 導電体、3 8 0 : 絶縁体、3 8 2 : 絶縁体、3 8 4 : 絶縁体、3 8 6 : 導電体、3 9
 0 : 集積回路、3 9 1 : 半導体チップ、3 9 2 : リード、3 9 3 : S i トランジスタ層、

10

20

30

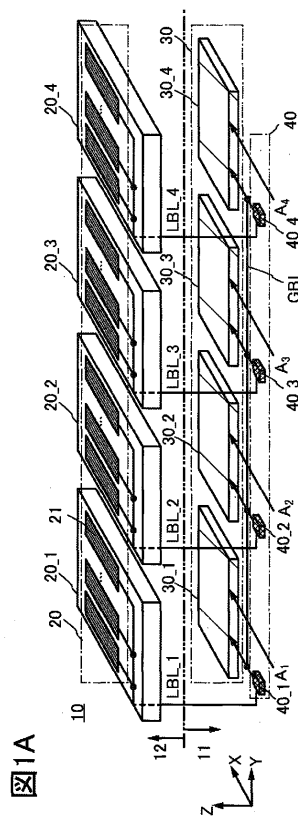
40

50

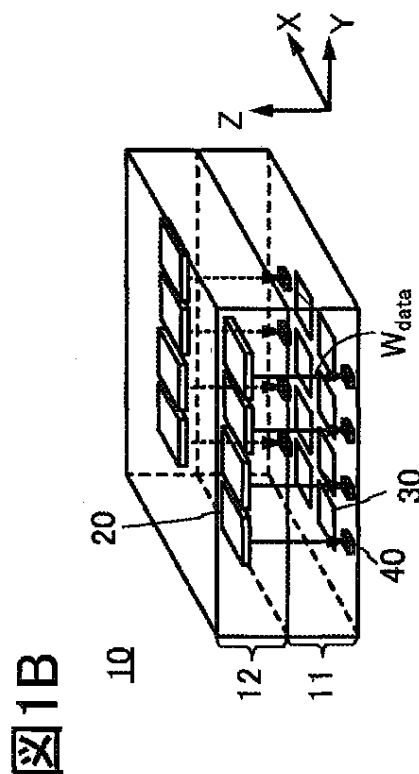
394：配線層、395：OSトランジスタ層、400：パッケージ基板、401：ソル
 ダーボール、402：半導体基板、403：トランジスタ、404：配線、405：電極
 、412：半導体基板、413：トランジスタ、414：配線、415：電極、420：
 領域、430：導電体、431：絶縁体、432：半導体領域、433a：低抵抗領域、
 433b：低抵抗領域、440：絶縁体、442：絶縁体、444：絶縁体、446：絶
 縁体、448：導電体、450：絶縁体、452：絶縁体、454：絶縁体、500：ト
 ランジスタ、503a：導電体、503b：導電体、503：導電体、510：絶縁体、
 512：絶縁体、514：絶縁体、516：絶縁体、518：導電体、522：絶縁体、
 524：絶縁体、530a：酸化物、530b：酸化物、530：酸化物、540a：導
 電体、540b：導電体、542a：導電体、542b：導電体、542：導電体、54
 3a：領域、543b：領域、544：絶縁体、545：絶縁体、546：導電体、54
 8：導電体、550：トランジスタ、560a：導電体、560b：導電体、560：導
 電体、574：絶縁体、580：絶縁体、581：絶縁体、582：絶縁体、586：絶
 縁体、590：自動車、591：カメラ、592：撮像方向、593：バス、594：ホ
 ストコントローラ、595：携帯型電子機器、596：プリント配線基板、597：スピー
 ーカー、598：カメラ、599：マイクロフォン、600：容量素子、610：導電体
 、612：導電体、620：導電体、630：絶縁体、640：絶縁体、1100：携帯
 型ゲーム機、1101：筐体、1102：筐体、1103：筐体、1104：表示部、1
 105：接続部、1107：操作キー、1108：筐体、1109：筐体、1120：電子
 機器、1121：筐体、1122：キャップ、1123：USBコネクタ、1124：基
 板、1125：メモリチップ、1126：コントローラチップ、1130：ロボット、
 2101：センサ、2106：センサ、2110：制御回路、3000：システム、30
 01：電子機器、3002：サーバー、3003：インターネット回線、3004：ラッ
 ク、3005：基板

【図面】

【図1A】



【図1B】



10

20

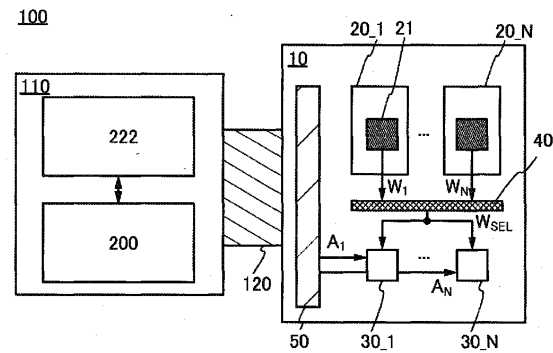
30

40

50

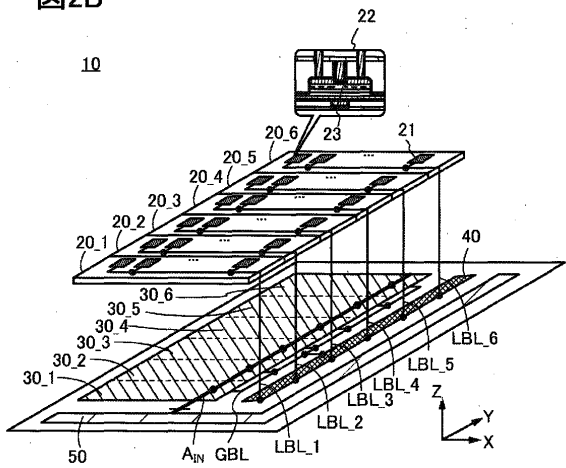
【図 2 A】

図2A



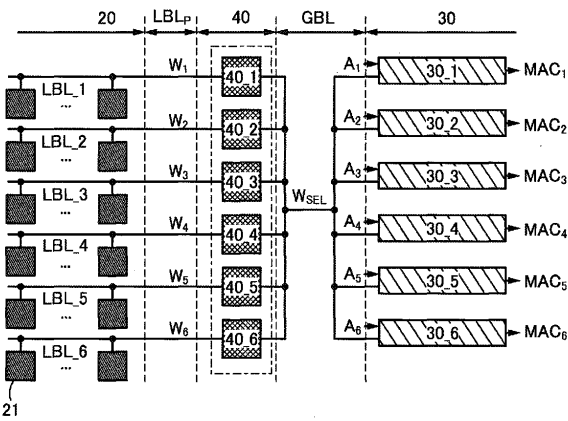
【図 2 B】

図2B



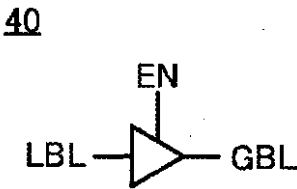
【図 3 A】

図3A



【図 3 B】

図3B



10

20

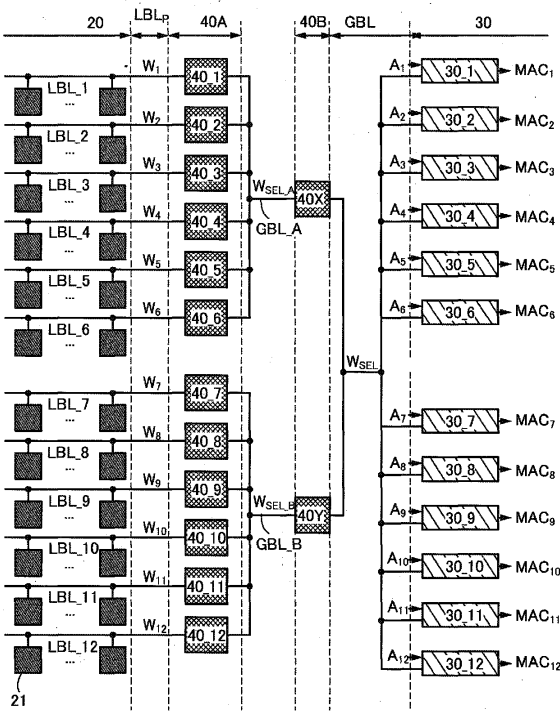
30

40

50

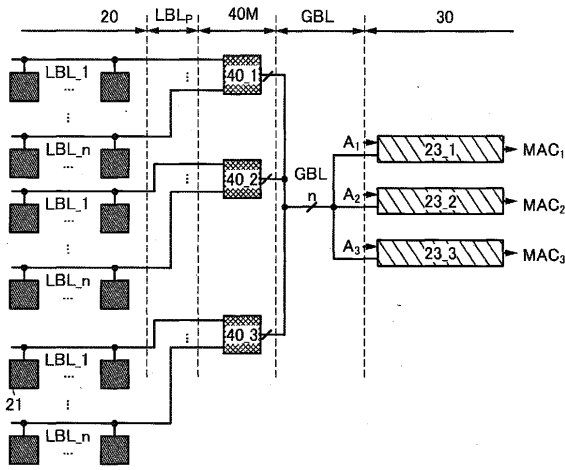
【図 4】

図4



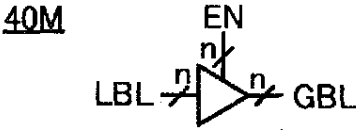
【図 5 A】

図5A

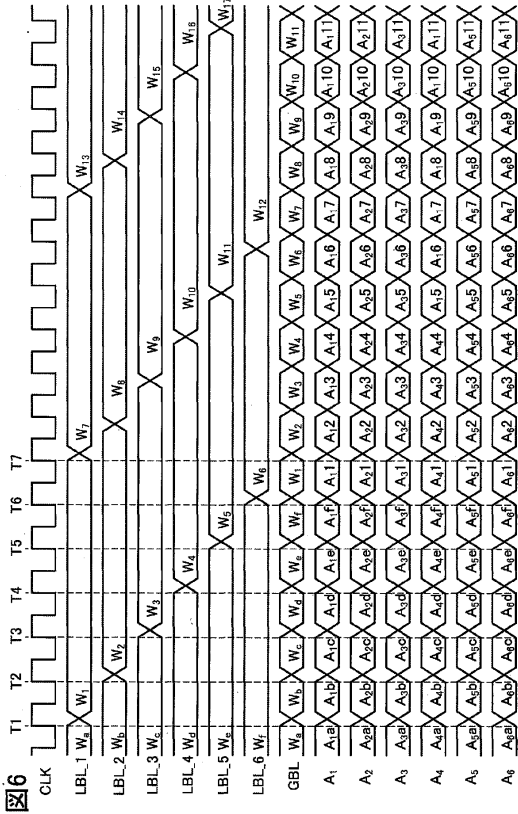


【図 5 B】

図5B



【図 6】



10

20

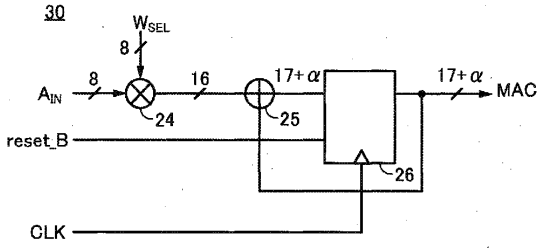
30

40

50

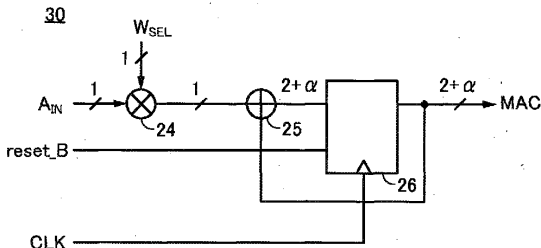
【図 7 A】

図 7A



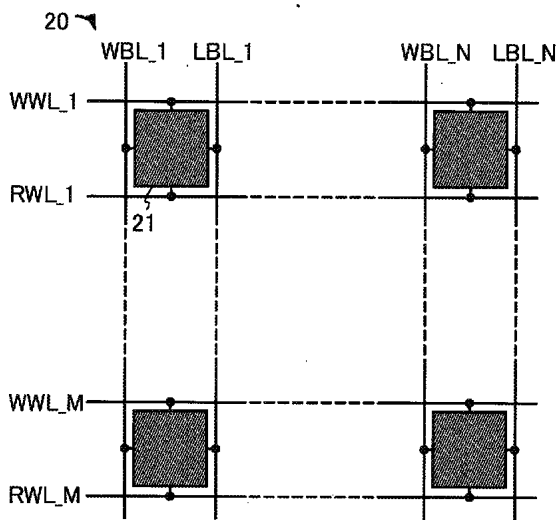
【図 7 B】

図 7B



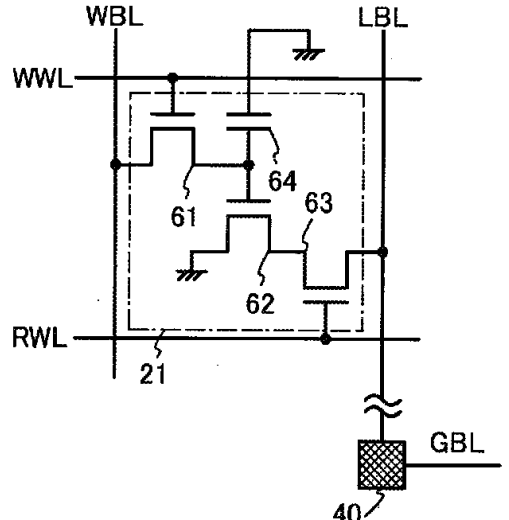
【図 8 A】

図 8A



【図 8 B】

図 8B



10

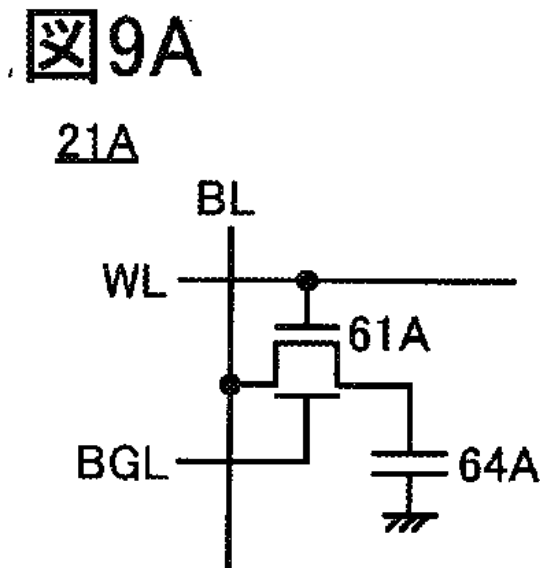
20

30

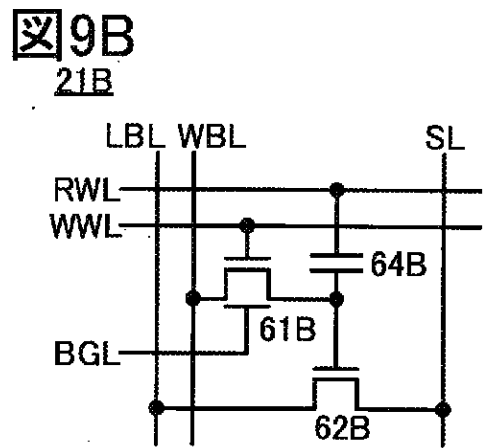
40

50

【図 9 A】

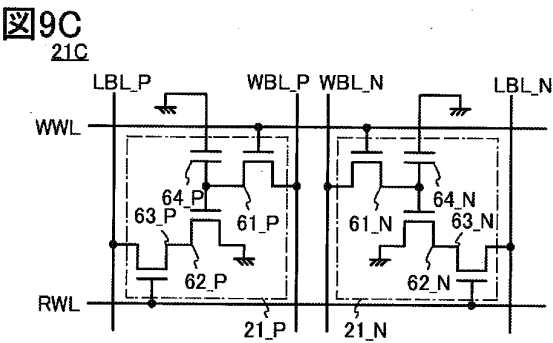


【図 9 B】

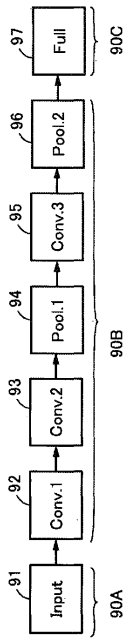


10

【図 9 C】



【図 1 0】



20

30

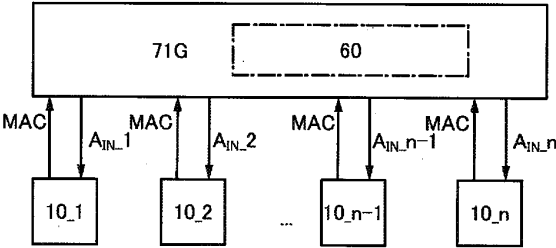
40

図 10

50

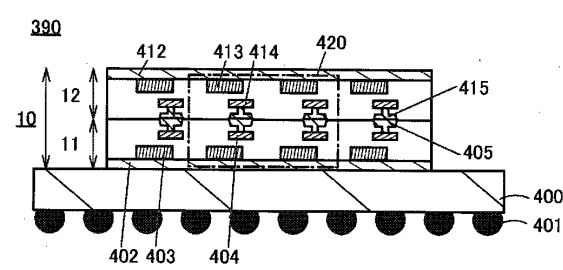
【図 1 3 B】

図13B



【図 1 4 A】

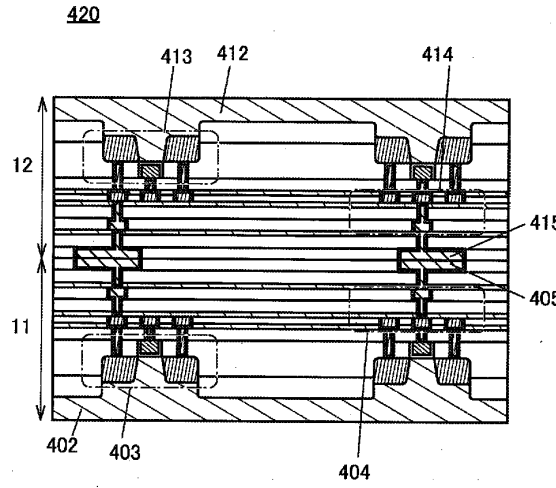
図14A



10

【図 1 4 B】

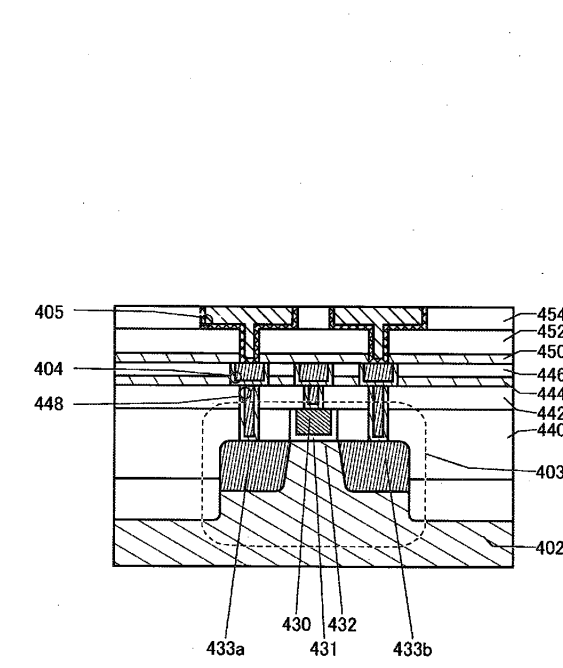
図14B



20

【図 1 5】

図15



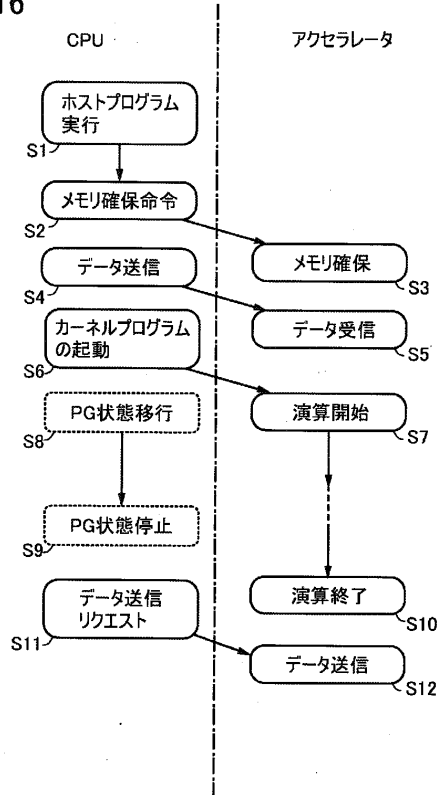
30

40

50

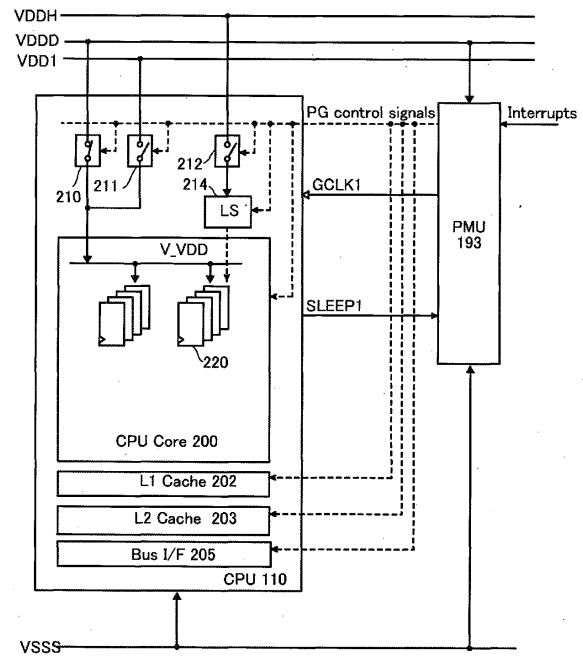
【図 16】

図16



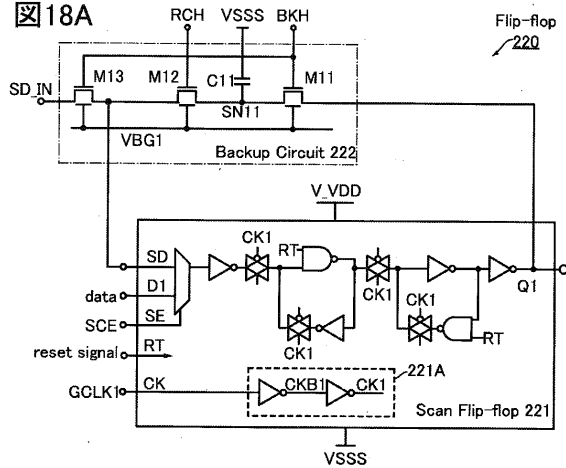
【図 17】

図17



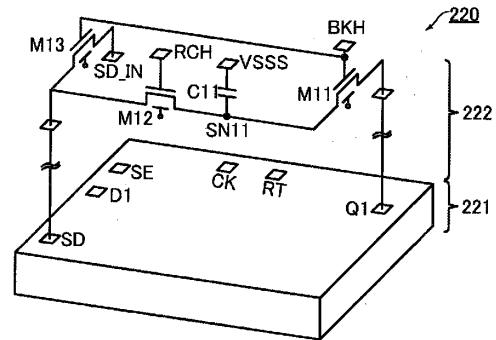
【図 18 A】

図18A



【図 18 B】

図18B



10

20

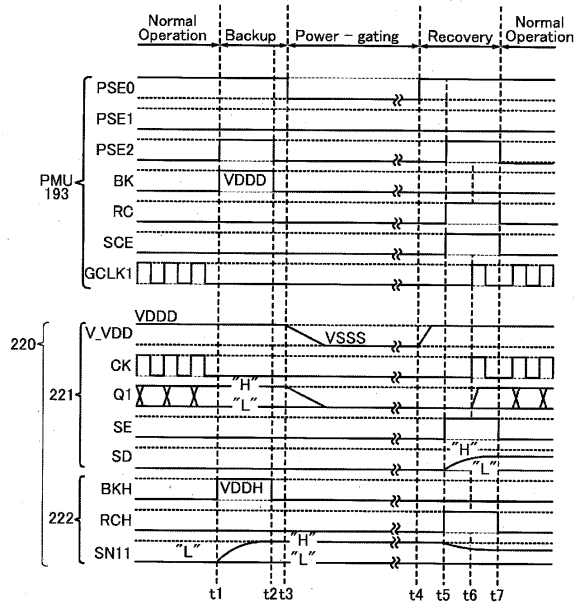
30

40

50

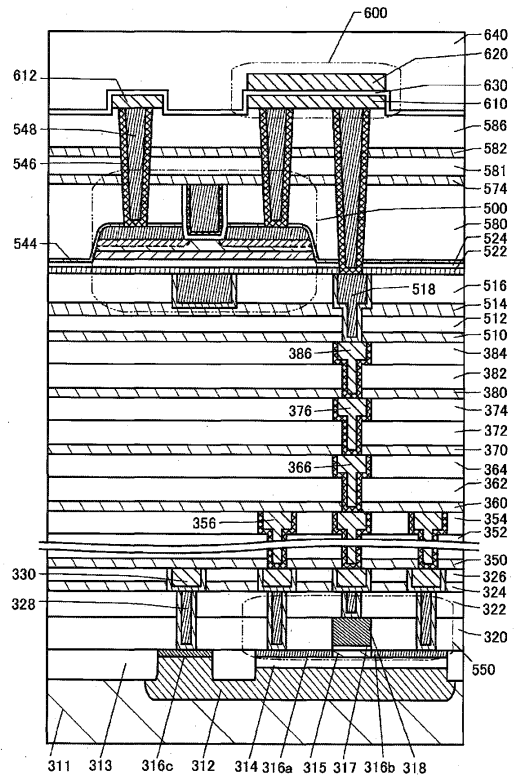
【図 19】

図19



【図 20】

図20

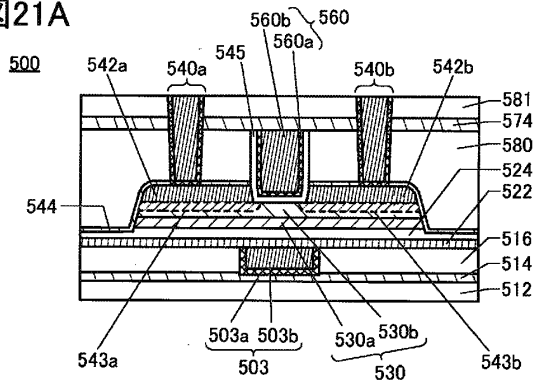


10

20

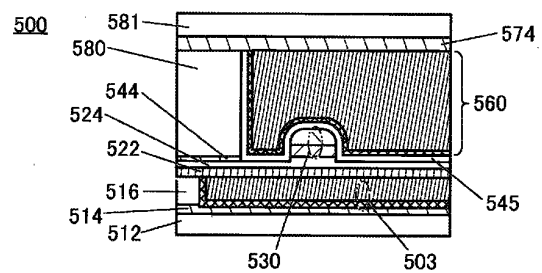
【図 21 A】

図21A



【図 21 B】

図21B



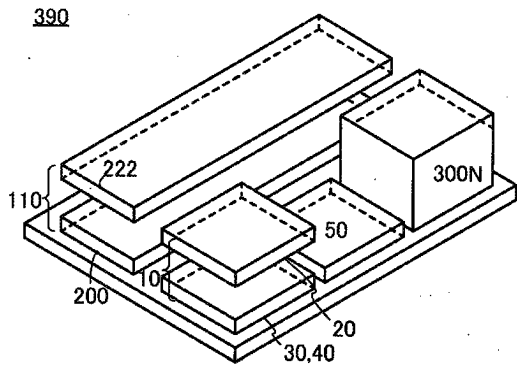
30

40

50

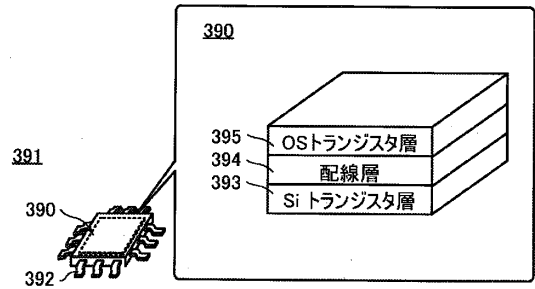
【図 2 2 A】

図22A



【図 2 2 B】

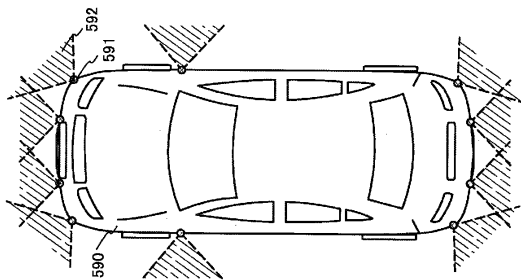
図22B



10

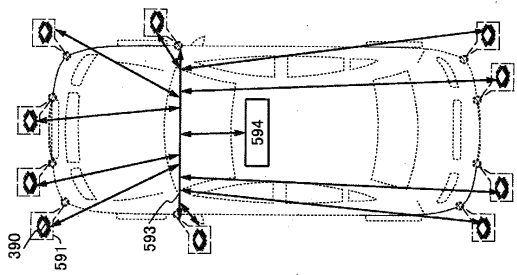
【図 2 3 A】

図23A



【図 2 3 B】

図23B



20

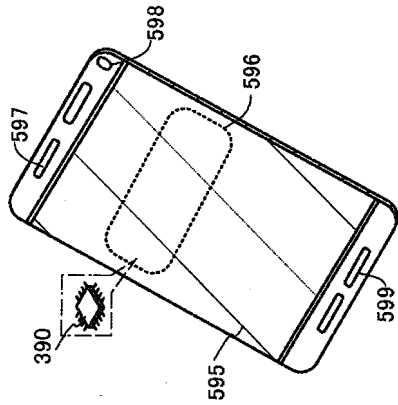
30

40

50

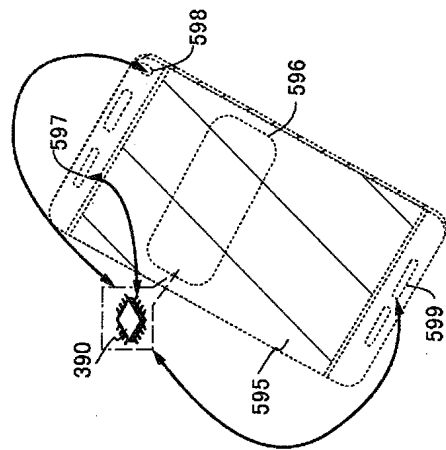
【図 2 4 A】

図24A



【図 2 4 B】

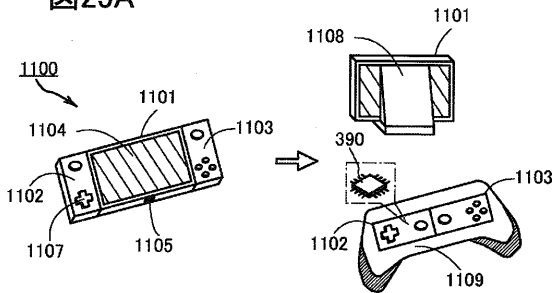
図24B



10

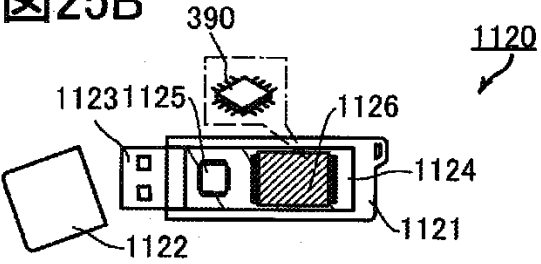
【図 2 5 A】

図25A



【図 2 5 B】

図25B



20

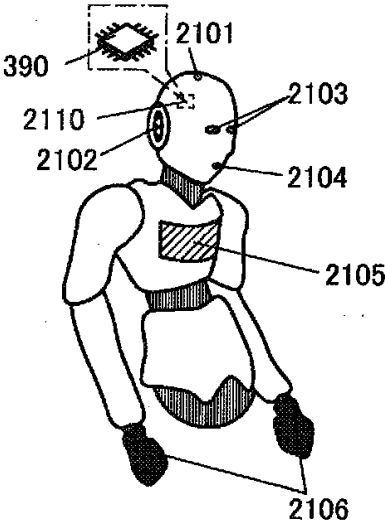
30

40

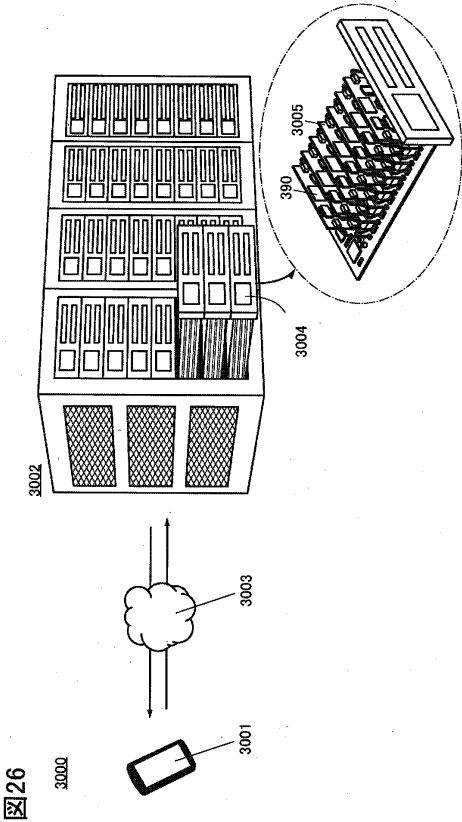
50

【図 25 C】

図25C 1130



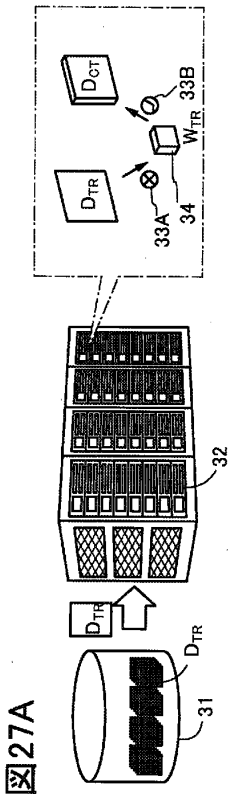
【図 26】



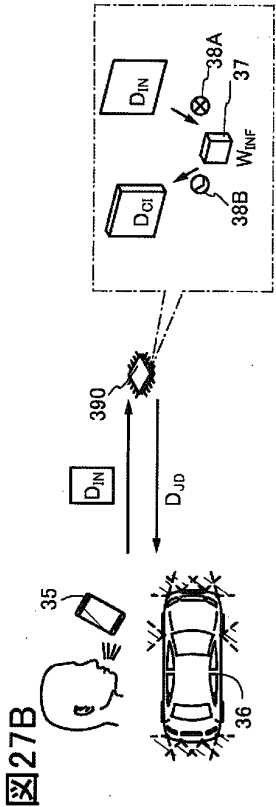
10

20

【図 27 A】



【図 27 B】



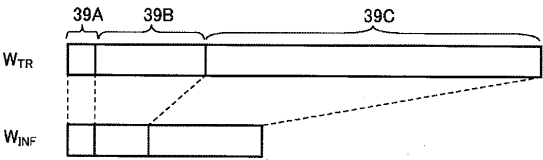
30

40

50

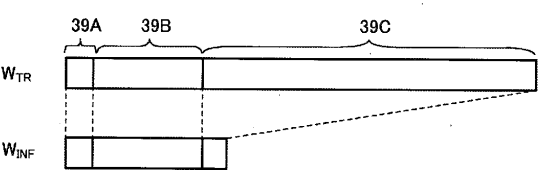
【図 28 A】

図28A



【図 28 B】

図28B



10

20

30

40

50

フロントページの続き

(51)国際特許分類

F I

H 1 0 D	84/80	(2025.01)	H 1 0 B	99/00	4 4 1
H 1 0 D	84/83	(2025.01)	H 1 0 D	30/67	1 0 3 B
H 1 0 D	86/40	(2025.01)	H 1 0 D	30/68	
H 1 0 D	86/60	(2025.01)	H 1 0 D	84/80	1 0 1 A
H 1 0 D	89/00	(2025.01)	H 1 0 D	84/83	E
G 0 6 F	12/00	(2006.01)	H 1 0 D	84/83	1 0 1 E
G 0 6 F	17/10	(2006.01)	H 1 0 D	86/40	1 0 1 B
G 1 1 C	5/02	(2006.01)	H 1 0 D	86/60	C
G 1 1 C	5/04	(2006.01)	H 1 0 D	89/00	1 0 1 U
			G 0 6 F	12/00	5 6 0 F
			G 0 6 F	17/10	S
			G 1 1 C	5/02	1 0 0
			G 1 1 C	5/04	2 2 0

(32)優先日 令和2年3月13日(2020.3.13)

(33)優先権主張国・地域又は機関

日本国(JP)

(56)参考文献 国際公開第2018/211349 (WO, A1)
 特開2018-133016 (JP, A)
 国際公開第2018/189620 (WO, A1)
 特開2019-036280 (JP, A)
 特開平04-067259 (JP, A)

(58)調査した分野 (Int.Cl., DB名)

H 1 0 B 41/70
 H 1 0 B 12/00
 H 1 0 B 99/00
 H 1 0 D 30/67
 H 1 0 D 30/68
 H 1 0 D 84/80
 H 1 0 D 84/83
 H 1 0 D 86/40
 H 1 0 D 86/60
 H 1 0 D 89/00
 G 0 6 F 12/00
 G 0 6 F 17/10
 G 1 1 C 5/02
 G 1 1 C 5/04