

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2007-318135

(P2007-318135A)

(43) 公開日 平成19年12月6日(2007.12.6)

(51) Int. Cl.	F I	テーマコード (参考)
HO 1 L 21/8247 (2006.01)	HO 1 L 29/78 3 7 1	5 B 1 2 5
HO 1 L 29/788 (2006.01)	HO 1 L 27/10 4 3 4	5 F 0 8 3
HO 1 L 29/792 (2006.01)	G 1 1 C 17/00 6 2 2 Z	5 F 1 0 1
HO 1 L 27/115 (2006.01)	G 1 1 C 17/00 6 2 1 B	
G 1 1 C 16/04 (2006.01)	G 1 1 C 17/00 6 2 3 A	

審査請求 未請求 請求項の数 26 O L 外国語出願 (全 69 頁) 最終頁に続く

(21) 出願番号 特願2007-133919 (P2007-133919)
 (22) 出願日 平成19年5月21日 (2007.5.21)
 (31) 優先権主張番号 11/419,977
 (32) 優先日 平成18年5月23日 (2006.5.23)
 (33) 優先権主張国 米国 (US)

(71) 出願人 599129074
 旺宏電子股▲ふん▼有限公司
 台湾新竹科學工業園區力行路16號
 (74) 代理人 100104156
 弁理士 龍華 明裕
 (72) 発明者 呂 函庭
 台湾新竹科學工業園區力行路16號
 (72) 発明者 連 浩明
 台湾新竹科學工業園區力行路16號
 Fターム(参考) 5B125 BA05 CA06 CA08 CA14 CA27
 DC03 DC07 EA03 FA02 FA06

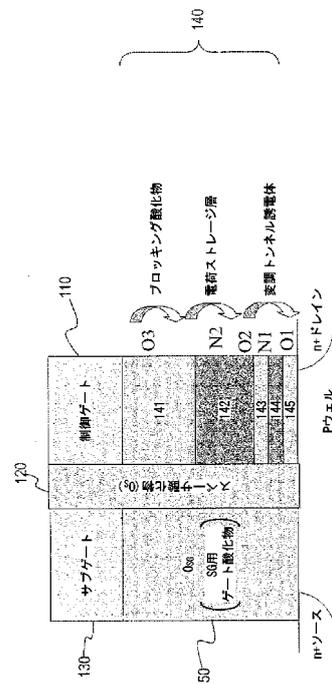
最終頁に続く

(54) 【発明の名称】 バンドギャップ操作SONOSデバイスを用いたサブゲートANDアーキテクチャの構造及び方法

(57) 【要約】 (修正有)

【課題】 スケーラビリティを有するAND型フローティングゲートフラッシュメモリを提供する。

【解決手段】 ソースサイドインジェクションプログラミング方法を実行するため、ANDアーキテクチャを有するバンドギャップ操作BE SONOSデバイス構造は、酸化物141 窒化物142 酸化物143 窒化物144 酸化物145スタックに重ねられる制御ゲート110と、ゲート酸化物150に重ねられるサブゲート130との間に配されるスペーサ酸化物120を含む。第1の形態においては、BE SONOSサブゲートANDアレイアーキテクチャは、サブゲートライン及び拡散ビットラインを有するSONOSデバイスの複数の列を含む。第2の形態においては、BE SONOSサブゲート反転ビットラインANDアーキテクチャは、サブゲート反転ビットラインを有して、拡散ビットラインは有さないSONOSデバイスの複数の列を含む。



【選択図】 図1

【特許請求の範囲】

【請求項 1】

半導体基板と、
前記半導体基板上の複数のメモリセルと、
前記半導体基板に配された N + 埋め込み拡散部とを含み、
それぞれの前記メモリセルは、ゲートとサブゲートとの間に配されたスペーサ誘電体層
を含み、

それぞれの前記ゲートは、ブロッキング酸化物電荷ストレージ層変調トンネル誘電体ス
タックに重ねられ、

それぞれの前記サブゲートは、ゲート酸化物に重ねられ、

前記 N + 埋め込み拡散部は、第 1 のゲート酸化物と、第 1 の拡散ビットラインとして働
く第 1 のブロッキング酸化物電荷ストレージ層変調トンネル誘電体スタックとの間の下に
配される集積回路デバイス。

10

【請求項 2】

前記変調トンネル誘電体スタックは、酸化物 窒化物 酸化物 (O 2 N 1 O 1) ス
タックを含む請求項 1 に記載の集積回路。

【請求項 3】

前記 O 2 N 1 O 1 スタックは、十分に薄い酸化物と窒化物層とを含む請求項 2 に記
載の集積回路。

【請求項 4】

半導体基板と、
前記半導体基板上の複数のメモリセルとを含み、
それぞれの前記メモリセルは、ゲートとサブゲートとの間に配されたスペーサ酸化物を
含み、

20

それぞれの前記ゲートは、ブロッキング酸化物電荷ストレージ層変調トンネル誘電体ス
タックに重ねられ、

それぞれの前記サブゲートは、ゲート酸化物に重ねられ、

第 1 のサブゲートは、前記第 1 のサブゲートが僅かにオン状態のときに、プログラミン
グ中にソースサイドインジェクションとして働き、

前記第 1 のサブゲートは、前記第 1 のサブゲートがオン状態のときに、第 1 のメタルビ
ットラインと接続するための反転層として働く集積回路デバイス。

30

【請求項 5】

前記変調トンネル誘電体スタックは、酸化物 窒化物 酸化物 (O 2 - N 1 O 1) ス
タックを含む請求項 4 に記載の集積回路。

【請求項 6】

前記 O 2 - N 1 O 1 スタックは、十分に薄い酸化物及び窒化物層を含む請求項 5 に記
載の集積回路。

【請求項 7】

並列に接続された S O N O N O S デバイスの複数の列を含むメモリアレイと、

複数のサブゲートラインと、

複数の拡散ビットラインとを含み、

それぞれの前記サブゲートラインは、S O N O N O S デバイスの対応する列と接続して

40

、
それぞれの前記拡散ビットラインは、S O N O N O S デバイスの対応する列と接続した
S G A N D アレイアーキテクチャ。

【請求項 8】

前記複数のサブゲートラインは、プログラミング中にソースサイドインジェクションの
ために用いられる請求項 7 に記載の S G A N D アレイアーキテクチャ。

【請求項 9】

奇数番号の前記サブゲートラインは、共に電氣的に接続した請求項 7 に記載の S G A

50

N D アレイアーキテクチャ。

【請求項 10】

偶数番号の前記サブゲートラインは、共に電氣的に接続した請求項 7 に記載の S G A N D アレイアーキテクチャ。

【請求項 11】

並列構造で接続された S O N O N O S デバイスの複数の列を含み、拡散ビットラインを含まないメモリアレイと、

複数のサブゲートラインとを含み、

それぞれの前記サブゲートラインは、S O N O N O S デバイスの対応する列と接続した S G I B A N D アレイ構造。

10

【請求項 12】

前記複数のサブゲートラインは、前記複数のサブゲートラインがオフ状態のときに、プログラミング中にソースサイドインジェクションのために用いられる請求項 11 に記載の S G I B A N D アレイアーキテクチャ。

【請求項 13】

前記複数のサブゲートラインのそれぞれは、前記複数のサブゲートラインがオン状態のときに、それぞれのメタルビットラインと接続するための反転層として働く請求項 11 に記載の S G I B A N D アレイアーキテクチャ。

【請求項 14】

前記複数のサブゲートラインにおいて、N 番目ごとの前記サブゲートラインが共に電氣的に接続した請求項 11 に記載の S G I B A N D アレイアーキテクチャ。

20

【請求項 15】

前記複数のサブゲートラインにおいて、4 番目ごとの前記サブゲートラインが共に電氣的に接続した請求項 11 に記載の S G I B A N D アレイアーキテクチャ。

【請求項 16】

第 1 の側壁と第 2 の側壁とを含むスペーサ酸化物と、

前記スペーサ酸化物の前記第 1 の側壁に対して水平に伸びるサブゲートと、

前記スペーサ酸化物の前記第 2 の側壁に対して水平に伸びる酸化物窒化物酸化物 (O 1 - N 1 O 2) スタックと、

第 3 の酸化物層 O 3 に重ねられ、前記スペーサ酸化物の前記第 2 の側壁に対して水平に伸びる制御ゲートとを含むメモリセル構造。

30

【請求項 17】

前記サブゲートに重ねられ、前記スペーサ酸化物の前記第 1 の側壁に対して水平に伸びるゲート酸化物を更に含む請求項 16 に記載のメモリ構造。

【請求項 18】

前記制御ゲートと前記第 2 の酸化物層 O 2 との間に配された第 2 の窒化物層 N 2 を更に含み、前記第 2 の窒化物層 N 2 は、前記スペーサ酸化物の前記第 2 の側壁に対して水平に伸び、電荷ストレージ層として働く請求項 17 に記載のメモリ構造。

【請求項 19】

前記制御ゲートと前記第 2 の窒化物層 N 2 との間に配された第 3 の酸化物層 O 3 を更に含み、前記第 3 の酸化物層は、前記スペーサ酸化物の第 2 の側壁に対して水平に伸び、ブロッキング酸化物として働く請求項 18 に記載のメモリ構造。

40

【請求項 20】

前記 O 1 - N 1 O 2 スタックは、十分に薄い層である酸化物及び窒化物を含む請求項 19 に記載のメモリ構造。

【請求項 21】

前記第 1 の酸化物層 O 1、前記第 1 の窒化物層 N 1 及び前記第 2 の酸化物層 O 2 のそれぞれは、3 nm より小さい請求項 20 に記載のメモリ構造。

【請求項 22】

それぞれの列がサブゲートラインとワードラインとに対応した、バンドギャップ操作 S

50

ONOS (BE SONOS) デバイスの列を含む AND メモリアレイを動作させる方法であって、

前記 AND メモリアレイの複数の BE SONOS デバイスをリセットする工程と、

前記 BE SONOS デバイスの第 1 の列の選択された BE SONOS デバイスを電氣的にプログラミングする工程と、

前記 AND メモリアレイの前記複数の BE SONOS デバイスを電氣的に消去する工程とを含み、

前記複数の BE SONOS デバイスは、閾値電圧 V_t 値をリセットする目的で、自己収束リセットを有し、

前記 BE SONOS デバイスの前記第 1 の列は、高電圧が印加される第 1 のワードラインに対応し、前記 BE SONOS デバイスの前記第 1 の列は、ソースサイドインジェクションを実行する目的で低電圧が印加される第 1 のサブゲートラインに対応し、

前記複数の BE SONOS デバイスは、前記閾値電圧 V_t 値をリセットする目的で、自己収束消去を有す方法。

10

【請求項 2 3】

前記選択された BE SONOS デバイスを読み出す工程を更に含み、

前記第 1 のワードラインの前記電圧は、消去状態レベルとプログラム状態レベルとの間の電圧である請求項 2 2 に記載の方法。

【請求項 2 4】

前記第 1 のサブゲートラインは、前記第 1 のサブゲートラインが僅かにオン状態であるときに、プログラミング中にソースサイドインジェクションとして働く請求項 2 2 に記載の方法。

20

【請求項 2 5】

前記第 1 のサブゲートラインは、前記第 1 のサブゲートラインがオン状態のときに、第 1 のメタルビットラインと接続するための反転層として働く請求項 2 2 に記載の方法。

【請求項 2 6】

前記第 1 のサブゲートラインに印加される前記低電圧は 1 ~ 2 V である請求項 2 2 に記載の方法。

【発明の詳細な説明】

【技術分野】

30

【0001】

本発明は概して不揮発性メモリデバイスに関する。より具体的には、窒化物ベースのトラッピングストレージフラッシュメモリに関する。

【背景技術】

【0002】

電氣的消去可能プログラマブル読み出し専用メモリ (EEPROM) 及びフラッシュメモリとして知られる、電荷ストレージ構造を基礎とした電氣的プログラマブル消去可能不揮発性メモリ技術は、さまざまな現代のアプリケーションにおいて用いられている。フラッシュメモリは、個別にプログラム及び読み出しができるメモリセルアレイを含む。不揮発性メモリ内に記憶された 1 つ又は複数のデータ値を判定する目的で、フラッシュメモリ内でセンス増幅器が用いられる。通常のセンシング体系においては、検知されるメモリセルを流れる電流は、電流センス増幅器によって、基準電流と比較される。

40

【0003】

数多くのメモリセル構造が、EEPROM 及びフラッシュメモリに用いられている。集積回路の寸法が縮小するにつれて、その製造プロセスにおける安定性及び単純性から、電荷トラッピング誘電体層を基礎としたメモリセル構造に対する関心がより増加している。電荷トラッピング誘電体層を基礎とするメモリセル構造は、例えば、業界名で窒化物読み出し専用メモリ、ポリシリコン 窒化物 窒化物 シリコン (SONOS) 及び PHINES 等として知られる構造を含んでいる。これらのメモリセル構造は、シリコン窒化物等の電荷トラッピング誘電体層に電化をトラッピングすることによってデータを記

50

憶する。負の電荷がトラップされると、メモリセルの閾値電圧値は増加する。メモリセルの閾値電圧値は、電荷トラッピング層から負の電荷を取り除くことによって減少する。

【0004】

窒化物読み出し専用メモリデバイスは、電荷損失を防ぐ目的で、例えば3nm以上、通常5~9nmの比較的厚みのある酸化物を底面に用いる。ダイレクトトンネリングの代わりに、バンド トゥ バンド・トンネリング誘導ホットホールインジェクション、いわゆるBTBTHHをセルの消去に用いることができる。しかしながら、ホットホールインジェクションは酸化物にダメージを与え、高閾値セルにおいて電荷損失を引き起こして、低閾値セルにおいては電荷増幅を引き起こす。更に、電荷トラッピング構造においては電荷蓄積を消去するのが困難なため、プログラム及び消去サイクル中の消去時間は次第に増加せざるを得ない。ホールインジェクションポイント及び電子インジェクションポイントが互いに一致せずに、電子の幾つかは消去パルス後にも残留するために、このような電子蓄積は発生する。更に、窒化物読み出し専用メモリフラッシュメモリデバイスのセクタ消去中において、プロセス差異(チャンネル長さの差異等)により、それぞれのセルの消去速度は異なる。消去速度のこの差異は、消去状態において広い V_t 分布を引き起こす。ここでは、セルの幾つかは消去が困難であり、セルの幾つかは過度に消去される。従って、多くのプログラム及び消去サイクルの後ではターゲット閾値 V_t の範囲が狭くなり、耐久性に難が生じる。この現象は、技術がスケールダウンを達成し続けると、より重大になってくるであろう。

10

【0005】

一般的なフラッシュメモリセル構造は、導電性ポリシリコントンネル酸化物層と結晶性シリコン半導体基板との間にトンネル酸化物層を配している。基板は、基礎をなすチャンネル領域によって分離されたソース領域とドレイン領域を含む。フラッシュメモリの読み出しは、ドレインセンシング又はソースセンシングによって実行することができる。ソース側のセンシングにおいては、メモリアレイの特定のメモリセルから電流を読み出す目的で、1つ以上のソースラインがメモリセルのソース領域に接続される。

20

【0006】

従来フローティングゲートデバイスは、導電性フローティングゲートに電荷1ビットを格納する。窒化物読み出し専用メモリセルが出現して、ここでは、それぞれのセルが、酸化物 窒化物 酸化物(ONO)誘電体中に電荷を格納する2ビットのフラッシュセルを有する。窒化物読み出し専用メモリセルの一般的な構造では、最上部の酸化物層と底面の酸化物層との間に位置するトラッピング物質として窒化物層が用いられる。ONO層構造は、フローティングゲートデバイスのゲート誘電体に効果的に取って代わっている。窒化物層を有するONO誘電体中の電荷は、NROMセルの左側又は右側どちらにトラップされても良い。

30

【0007】

フローティングゲートデバイスは、内部のフローティングゲート結合に起因する大きなスケールアップ課題に直面している。一方で、窒化物トラッピングデバイスは、そのような制限にとらわれない。窒化物トラッピングデバイスは主に2種類ある。局所的に電荷を格納するNROMと、チャンネルプログラム/消去を用いるSONOSである。これら2つのタイプのデバイスは欠点を備えている。窒化物読み出し専用メモリデバイスは、ホットホールが誘発するダメージの影響を受け易い。そして、SONOSデバイスは、薄膜トンネル酸化膜を介してダイレクトトンネリング漏れに起因する保持問題に悩まされている。

40

【0008】

従来AND型フローティングゲートフラッシュメモリは、多くの商業的アプリケーションに適している。なぜなら、このメモリデバイスは、高密度、低電力及び高速プログラミング特性を有すからである。しかしながら、内部フローティングゲート結合効果により、AND型フローティングゲートデバイスのスケールアップは制限される。フローティングゲートデバイスの空間パラメータが縮小すると、高フローティングゲートカップリング効

50

果により、望ましくない深刻な擾乱が引き起こされる。従来のAND型フローティングゲートデバイスはまた、トンネル酸化物スケーリング問題、及び、トンネル酸化物内の局所的欠陥又はトラップ電荷がフローティングゲート内の電荷の漏れを引き起こす異常ビットに悩まされている。

【0009】

フローティングゲートデバイスのスケーリング問題の解決において、SONOS、MNOS又はナノ結晶トラッピングデバイス等の電荷トラッピングデバイスが提案されている。しかしながら、これらのデバイスは全て、深刻な電荷保持問題に悩まされている。SONOSデバイスにおいては、超薄膜トンネル酸化物はストレージ電荷を適切に保持することができない。MNOSデバイスにおいては、その構造が電荷損失を防ぐための最上部の酸化物を備えていない。ナノ結晶デバイスにおいては、不規則に分布したナノ粒子が要因で制御性に難がある。

10

【0010】

従って、その保持問題を克服して、且つ有効なホールトンネリング消去を維持することで、スケラビリティを有するAND型フローティングゲートフラッシュメモリの考案が望まれている。

【発明の開示】

【発明が解決しようとする課題】

【0011】

本発明は、ソースサイドインジェクション(SOI)プログラミング方法を実行する目的で、さまざまなANDアーキテクチャを有する設計に用いられるバンドギャップ操作SONOS(「BE SONOS」又は「SONONOS」と称される)デバイス構造を提供する。BE SONOSデバイス構造は、酸化物窒化物酸化物窒化物酸化物(O₃ N₂ O₂ N₁ O₁)スタックに重ねられる制御ゲートと、ゲート酸化物に重ねられるサブゲートとの間に配置されるスペーサ酸化物を含む。第1の形態においては、BE SONOS SG AND(サブゲートAND)アレイアーキテクチャは、サブゲートライン及び拡散ビットラインを有するSONONOSデバイスの複数の列を含む。第2の形態においては、BE SONOS SGIB AND(サブゲート反転ビットラインAND)アーキテクチャは、サブゲート反転ビットラインを有して、拡散ビットラインは有さないSONONOSデバイスの複数の列を含む。

20

30

【0012】

大雑把に述べると、集積回路デバイスは、半導体基板、半導体基板上の複数のメモリセル、及びN⁺埋め込み拡散部を含む。それぞれのメモリセルは、ゲートとサブゲートの間に配置されたスペーサ酸化物を含み、それぞれのゲートはブロッキング酸化物電荷ストレージ層変調トンネル誘電体スタックに重ねられ、それぞれのサブゲートはゲート酸化物に重ねられる。N⁺埋め込み拡散部は、半導体基板に配置され、第1のゲート酸化物と、第1の拡散ビットラインとして働く第1のブロッキング酸化物電荷ストレージ層変調トンネル誘電体スタックとの間の下に置かれる。

【0013】

有利なことに、本発明のBE SONOS ANDアレイアーキテクチャは、メモリデバイスのフローティングゲート及びAND型に対して、より優れたスケラビリティを提供する。本発明はまた、均一且つ自己収束性のチャンネルホールトンネリング消去動作を有利に提供する。更に、本発明は内部フローティングゲート結合効果を取り除く。本発明は更に、優れた電荷保持による確かな特性、異常ビットがほとんど存在しない保証数、及びプログラム及び消去サイクル後に確かに劣化が少ない等の望ましい信頼性特性を提供する。

40

【0014】

本発明に関する構造及び方法は、以下の詳細の説明において記載される。本要約は本発明を定義することを意図していない。本発明は請求項によって定義される。本発明におけるこれら及び他の形態、特徴、側面、及び効果は以下の記載、添付の請求項及び添付図面

50

によって、より良く理解できるであろう。

【図面の簡単な説明】

【0015】

【図1】本発明に準じた、サブゲートを有すnチャンネルBE SONOSデバイスの1つのセル構造の断面図を示す構造図である。

【0016】

【図2A】本発明に準じた、拡散ビットラインを有すBE SONOS SG ANDアレイアーキテクチャの第1の形態を示す回路図である。

【0017】

【図2B】本発明に準じた、拡散ビットラインを有すBE SONOS SG ANDアレイアーキテクチャの第1の形態を示すレイアウト図である。 10

【0018】

【図3A】本発明に準じた、第1の形態における、BE SONOS SG ANDアレイアーキテクチャのチャンネル長さ方向の断面図を示すレイアウト図である。

【0019】

【図3B】本発明に準じた、第1の形態における、BE SONOS SG ANDアレイアーキテクチャのチャンネル幅方向の断面図を示すレイアウト図350である。

【0020】

【図4A】本発明に準じた、第1の形態における、SONONOS SG ANDアレイアーキテクチャの電気的リセットを示す回路図である。 20

【0021】

【図4B】本発明に準じた、第1の形態における、自己収束リセットの波形を示すグラフ図である。

【0022】

【図5A】本発明に準じた、第1の形態における、SONONOS SG ANDアレイアーキテクチャの電気的プログラムを示す回路図である。

【0023】

【図5B】本発明に準じた、第1の形態における、SONONOS SG ANDアレイアーキテクチャの電気的プログラムを示すレイアウト図である。

【0024】

【図6A】本発明に準じた、第1の形態のSONONOS SG ANDアレイアーキテクチャの電気的消去を示す回路図である。 30

【0025】

【図6B】本発明に準じた、自己収束消去の波形を示すグラフ図である。

【0026】

【図7】本発明に準じた、第1の形態における、SONONOS SG ANDアレイアーキテクチャの読み出し動作を示す回路図である。

【0027】

【図8A】本発明に準じた、BE SONOS SG IB ANDアレイアーキテクチャの第2の形態を示す回路図である。 40

【0028】

【図8B】本発明に準じた、BE SONOS SG IB ANDアレイアーキテクチャの第2の形態を示すレイアウト図である。

【0029】

【図9A】本発明に準じた、第2の形態における、SONONOS SG IB ANDアレイアーキテクチャのチャンネル長さ方向の断面図を示すレイアウト図である。

【0030】

【図9B】本発明に準じた、第2の形態における、SONONOS SG IB ANDアレイアーキテクチャのチャンネル幅方向の断面図を示すレイアウト図である。

【0031】

【図10A】本発明に準じた、第2の形態における、SONONOS SGIB AND アレイアーキテクチャの電気的リセットを示す回路図である。

【0032】

【図10B】本発明に準じた、自己収束リセットの波形を示すグラフ図である。

【0033】

【図11A】本発明に準じた、第2の形態における、SONONOS SGIB AND アレイアーキテクチャの電気的プログラムを示す回路図である。

【0034】

【図11B】本発明に準じた、第2の形態における、SONONOS SGIB AND アレイアーキテクチャの電気的プログラムを示すレイアウト図である。

10

【0035】

【図12A】本発明に準じた、第2の形態のBE SONONOS SGIB AND アレイアーキテクチャの電気的消去を示す回路図である。

【0036】

【図12B】本発明に準じた、第2の形態の自己収束消去の波形を示すグラフ図である。

【0037】

【図13A】本発明に準じた、第2の形態における、SONONOS SGIB AND アレイアーキテクチャの読み出し動作を示す回路図である。

【0038】

【図13B】本発明に準じた、第2の形態における、SONONOS SGIB AND アレイアーキテクチャの読み出し動作を示すレイアウト図である。

20

【発明を実施するための最良の形態】

【0039】

ここで、図1を参照すると、サブゲート(SG)を有するnチャンネルBE SONOS デバイス100の1つのセル構造の断面図を示す構造図が見られる。nチャンネルBE SONOS デバイス100は、制御ゲート110とサブゲート130の間に配置されたスペース酸化物(OS)120を含む。酸化物窒化物酸化物窒化物酸化物(O3 N2 O2 N1 O1)構造140は制御ゲート110の下に配置される。ゲート酸化物OSG150はサブゲート130の下に配置される。O3 N2 O2 N1 O1構造140はブロッキング酸化物O3141、電荷ストレージ層N2142、及び変調トンネル誘電体O2-N1 O1143~145を含む。底面のO2 N1 O1143~145層はホールトンネリング電流を提供して、優れたデータ保持をもたらす。

30

【0040】

nチャンネルBE SONOS デバイス100は、2つのゲート、制御ゲート110及びサブゲート130を有する、5つの端子を備えるデバイスである。制御ゲート110の下には、電荷ストレージ用のO3 N2 O2-N1 O1構造140が存在する。サブゲート130の下には、ノトラッピングゲート酸化物150が存在する。制御ゲート110は、プログラムを制御すること、及び、電荷ストレージ層を消去及び読み出しすることができる。サブゲート130はソースサイドインジェクション(SSI)プログラミング方法を提供する事ができる。ソースサイドインジェクションは、低電力且つ高速なプログラミング方法である。O1 N1 O2層は、ホールのダイレクトトンネリングを達成する目的で、通常3nm以内である超薄膜酸化物及び窒化物によって形成されても良い。N2層142は、より高いトラッピング効率を達成する目的で、5nmより厚くしても良い。層141の形成方法においては、1つの方法として、O3とN2との間の接触面における大きいトラップ密度を達成する目的で、湿式転換酸化物が最上層に用いられる。O3層は、最上層の酸化物における電荷損失を防ぐ目的で、通常6nmより厚い。O1 N1 O2層は、ホールトンネリング用のトンネリング誘電体として作用する。

40

【0041】

サブゲート130を有するnチャンネルBE SONOS デバイス100のデバイスパラメータの一例を以下に示す。

50

底面の酸化物 (O 1)	1 5 A
中間の窒化物 (N 1)	2 0 A
中間の酸化物 (O 2)	1 8 A
トラッピング窒化物 (N 2)	9 0 A
S G 用ゲート酸化物 (O S G)	1 5 0 A
スペーサ酸化物 (O S)	2 0 0 A
ゲート材料	N + ポリ又は P + ポリゲート

【 0 0 4 2 】

拡散ビットラインを有する BE SONOS SG AND アレイアーキテクチャ 2 0 0 の第 1 の形態を示す回路図を図 2 A に示す。複数の SONOS デバイスは、BE SONOS SG AND アレイアーキテクチャ 2 0 0 を形成する目的で、並列に接続される。BE SONOS アレイアーキテクチャ 2 0 0 は、複数のビットライン BL 0 2 2 0、BL 1 2 2 1、BL 2 2 2 2、BL 3 2 2 3、BL 4 2 2 4 及び BL n 2 2 5 と交差する複数のワードライン WL 0 2 1 0、WL 1 2 1 1、WL 2 2 1 2、及び W L m 2 1 3 を含む。対応するサブゲートラインは、ビットラインの近くに並列に配置される。サブゲート SG 1 2 3 0 は、ビットライン BL 0 2 2 0 の隣に配置される。サブゲート SG 2 2 3 1 は、ビットライン BL 1 2 2 1 の隣に配置される。サブゲート SG 3 2 3 2 は、ビットライン BL 2 2 2 2 の隣に配置される。サブゲート SG 4 2 3 3 は、ビットライン BL 3 2 2 3 の隣に配置される。サブゲート SG n 2 3 5 は、ビットライン BL 5 2 2 5 の隣に配置される。メモリセルとして機能する BE SONOS (又は SONOS) デバイス 2 4 0 の 1 サンプルを丸で囲んだ領域に示す。

【 0 0 4 3 】

拡散ビットラインを有する BE SONOS SG AND アレイアーキテクチャ 2 0 0 の第 1 の形態を示すレイアウト図 2 5 0 を図 2 B に示す。サブゲート SG のそれぞれは、対応するビットラインに並列しているが、それぞれの SG は対応するビットラインから僅かなオフセットを有する。例を挙げると、SG 1 2 3 0 は BL 0 2 2 0 の少し右側に位置する。全てのビットラインは、ソース又はドレインとしての機能を果たす事ができる。それぞれの SG は 2 つのビットラインの間に位置する。例を挙げると、SG 1 2 3 0 は BL 0 2 2 0 と BL 1 2 2 1 との間に位置する。BL 0 2 2 0、BL 1 2 2 1、BL 2 2 2 2、BL 3 2 2 3、BL 4 2 2 4 及び BL 5 2 2 5 を含むそれぞれのビットラインは、ソース領域又はドレイン領域としての機能を果たす事ができる。従って、SG 1 2 3 0 は、BL 0 2 2 0 のソース領域と BL 1 2 2 1 のドレイン領域との間に配置される。パラメータ W 2 4 0 及び W s 2 4 2 は、パラメータ F に略等しい。ここで、パラメータ F とは技術ノードにおける限界寸法を示す。例えば、パラメータ F は 5 0 n m ノードに対しては 5 0 n m となる。

【 0 0 4 4 】

図 3 A は、第 1 の形態における、BE SONOS SG AND アレイアーキテクチャのチャンネル長さ方向の断面図を示すレイアウト図 3 0 0 である。スペーサ酸化物 1 2 0 は、制御ゲート 1 1 0 とサブゲート 1 3 0 とを分離する。O 3 N 2 O 2 N 1 O 1 構造 1 4 0 は制御ゲート 1 1 0 の下に配置される。ゲート酸化物 O S G 1 5 0 は、ゲート 1 3 0 の下に配される。制御ゲート 1 1 0 の好ましい実施形態はポリ 1 であり、サブゲート 1 3 0 の好ましい実施形態はポリ 2 である。N + 埋め込み拡散 (B D) ウェル 3 3 0、3 3 2、3 3 4 及び 3 3 6 は、拡散ビットライン (B L) 用に形成される。レイアウト図 3 0 0 において、第 1 のセル構造は、制御ゲート 1 1 0 及びサブゲート 1 3 0 を含む。第 1 のセル構造に隣接した第 2 のセル構造は、ゲート 3 1 0 及びサブゲート 3 1 2 を含む。第 2 のセル構造に隣接した第 3 のセル構造は、ゲート 3 2 0 及びサブゲート 3 2 2 を含む。

【 0 0 4 5 】

図 3 B は、第 1 の形態における、BE SONOS SG AND アレイアーキテクチャ

ヤのチャンネル幅方向の断面図を示すレイアウト図350である。ゲート310とゲート320の間のギャップをパラメータ $W_s 360$ によって示す。ギャップにより、ゲート310とゲート320の間は絶縁される。2つのゲート間を絶縁する目的で、他の同様の絶縁部分が2つのゲート間に見られる。チャンネル幅方向のピッチは、拡散ビットラインに起因するチャンネル長さ方向のピッチ $3F$ よりも小さく、略 $2F$ である。従って、BE SONOS SG ANDアーキテクチャはセル当たり略 $6F^2$ となる。

【0046】

図4Aは、第1の形態における、SONONOS SG ANDアレイアーキテクチャの電気的リセット方法を示す回路図400である。電気的リセットの間、ワードライン(又は、ゲート)WL0210、WL1211、WL2212及びWLM213は、 $-10V$ にセットされ、ビットラインBL0220、BL1221、BL2222、BL3223、BL4224及びBL5225はフローティング状態にされ、サブゲートSG1230、SG2231、SG3232、SG4233、SG5234及びSGn235は $0V$ にセットされる。1つの形態では、SG1230、SG3232、及びSG5234を含む奇数番号のサブゲートが電気的に共に接続される一方で、SG2231、SG4233及びSGn235を含む偶数番号のサブゲートが電気的に共に接続される。動作前に、メモリ回路400は、 $V_{gb} = -15V$ を印加することによって(又は、ゲート電圧をそれぞれのWL及びPウェルに分配することによって)リセットされる。それにより、図4Bのグラフ450で示されるとおりの望ましい自己収束特性を実現できる。たとえBE SONOSデバイスが最初にさまざまな V_t をチャージしていたとしても、これらのリセット動作により、これら初期点をリセット/消去状態に収束させることができる。リセット時間は通常100ミリ秒ぐらいである。1つの例では、nチャンネルBE SONOSは以下のパラメータを有する。ONONO = $15 / 20 / 18 / 70 / 90$ オングストローム、 $L_g / W = 0.22 / 0.16 \mu m$ 、N+ポリゲート。

【0047】

別の言い方をすれば、動作前に V_t の分布を収束させる目的でリセット動作は実行される。自己収束消去を有さないフローティングゲートデバイスと異なり、BE SONOSは、自己収束消去リセット/消去方式を備えている。最初の V_t 分布は多くの場合、プラズマ帯電効果等のプロセス問題に起因して、広い分布となるので、この方式が必要となる。自己収束リセットは、最初の V_t 分布の収束を促進する。

【0048】

図5Aは、第1の形態におけるSONONOS SG ANDアレイアーキテクチャの電気的プログラムを示す回路図500である。一方、図5Bは、第1の形態におけるSONONOS SG ANDアレイアーキテクチャの電気的プログラムを示すレイアウト図550である。1つの例では、電気的プログラム中、ワードラインWL1211は $10V$ にセットされ、その他のワードラインWL0210、WL2212及びWLM213は $0V$ にセットされる。ビットラインBL1221は $5V$ にセットされ、ビットラインBL0220、BL2222、BL3223、BL4224及びBLn225は $0V$ にセットされる。奇数番号のサブゲートSG1230、SG3232及びSG5234は $1V$ にセットされ、偶数番号のサブゲートSG2231、SG4233及びSGn235は $0V$ にセットされる。ビットラインBL0220、BL1221、BL2222、BL3223、BL4224及びBL5225は、サブゲートSG1230、SG2231、SG3232、SG4233、SG5234及びSGn235よりもプログラミングにおいてより高い柔軟性を提供する。なぜなら、それぞれのビットラインは、個別にプログラムできるが、サブゲートはサブゲートの偶数番号又は奇数番号に基づいてプログラムされるからである。電気的プログラミングの1つの方法は、ソースサイドインジェクションである。ソースサイドインジェクションはセルを高電圧閾値 V_t 状態にプログラムする。例えば、ソースサイドインジェクションにより、選択されたWL1に $V_g = 10V$ が印加され、他のワードラインには $V_g = 0V$ が印加される。プログラミング時は $SG = 1V$ であり、プログラミング禁止時は $SG = 0V$ である。図に示すとおり、SG電圧の電圧設定値は $1V$ であるが、

これはSGゲート下の閾値電圧より通常0.5~2V高い値を意図している。

【0049】

セル A422がプログラミングのため選択されたとき、SGの下のチャンネルが僅かにオンになるようSGは1Vにセットされる。電圧閾値 V_t をPVより高くする目的で、ソースサイドインジェクション方法によって、セル A422に電子が注入される。セル B424用のSGは0Vにセットされ、これにより、SGはオフとなるので、セル B424に注入は行われない。セル C426に関しては、SGは1Vにセットされ、ここでは $WL = 0V$ となるので、セル C426はオフとなりセル C426にも注入は行われない。結果として、適切なプログラム禁止手法を用いてプログラミングを任意に選択できる。

10

【0050】

電氣的プログラムを実行する目的で、ソースサイドインジェクションを行うべく、選択されたワードラインに高電圧の10Vが印加され、サブゲートには1Vが印加される。ソースサイドインジェクションは、低電力且つ高速なプログラミング方法である。2kBのセルを平行に有するページプログラミング等の平行プログラミング手法によって、プログラミングのスループットを1秒当たり10MB以上も増加でき、一方でセル電流消費値は2mA以内に抑えられることは当業者であれば理解できるであろう。他のビットラインへのプログラム擾乱を防ぐ目的で、サブゲートSG2231は0Vにセットされ、禁止セルはオフにされる。

【0051】

図6Aに、第1の形態におけるSONONOS SG ANDアレイアーキテクチャの電氣的消去を示す回路図600を示す。消去動作はリセット動作と同様に実行される。電氣的消去中には、ワードライン $WL0210$ 、 $WL1211$ 、 $WL2212$ 及び $WLn213$ は-10Vにセットされ、ビットライン $BL0220$ 、 $BL1221$ 、 $BL2222$ 、 $BL3223$ 、 $BL4224$ 及び $BL5225$ はフローティング状態にされ、サブゲートSG1230、SG2231、SG3232、SG4233、SG5234及びSGn235は0Vにセットされる。消去動作は、セクタ又はブロック単位で実行される。図6Bのグラフ650で示されるとおり、BE SONOSデバイスは望ましい自己収束消去特性を備える。消去飽和値 V_t はパラメータ V_g に依存する。 V_g が高いほど、飽和値 V_t もより高くなる。収束時間は通常、10~100ミリ秒ぐらいである。

20

30

【0052】

図7は、第1の形態におけるSONONOS SG ANDアレイアーキテクチャの読み出し動作を示す回路図700である。1つの例では、セル A422の読み出し動作中には、ワードライン $WL1211$ は5Vにセットされ、その他のワードライン $WL0210$ 、 $WL2212$ 、 $WLn213$ は0Vにセットされる。ビットライン $BL1221$ は1Vにセットされ、ビットライン $BL0220$ 、 $BL2222$ 、 $BL3223$ 、 $BL4224$ 及び $BLn225$ は0Vにセットされる。奇数番号のサブゲートSG1230、SG3232及びSG5234は3Vにセットされ、偶数番号のサブゲートSG2231及びSG4233は0Vにセットされる。読み出し動作は、消去状態 $V_t(EV)$ とプログラム状態 $V_t(PV)$ との間のゲート電圧を印加することによって実行される。ゲート電圧は通常、5Vぐらいである。あるいは、高 V_t 値から低 V_t 値までの範囲であれば、ゲート電圧として5Vより大きい値又は小さい値を選択することもできる。セル A422の V_t が5Vより高い場合は、読み出し電流は微小値(例えば $< 0.1 \mu A$)になるであろう。セル A422の V_t が5Vより低い場合は、読み出し電流は大きい値(例えば $> 0.1 \mu A$)になるであろう。

40

【0053】

ビットライン(BL)への印加電圧は通常、1Vぐらいである。読み出し電圧が大きいほど電流も大きくなるが、読み出しによる擾乱もより大きくなるであろう。SG ANDストリングのWL数は通常64、128又は256である。SG ANDストリングの数が大きくなるほど、オーバーヘッドをより防ぐことができ、アレイ効率も増加するであろう

50

。しかしながら、プログラム分布はより大きくなるであろう。トレードオフを検討して、SG ANDストリングの数を適切に選択する必要がある。

【0054】

上記読み出し機能はランダムアクセス読み出し動作であるが、本発明の精神から逸脱しない範囲において、複数のセルのページ読み出しができることは当業者であれば理解できるであろう。

【0055】

図8Aに、第2の形態における、BE SONOS (又はSONONOS) SGIB AND (Sub gate Inversion Bitline AND) アレイアーキテクチャを示す回路図800を示す。一方、図8Bは、BE SONOS SGIB ANDアレイアーキテクチャの第2形態を示すレイアウト図850である。用語SGIBとは、サブゲートをオンすることによって反転層からビットラインを形成することを意味する。レイアウト図300で示した、BE SONOS構造がN+埋め込み拡散部を有している第1の形態とは異なり、図9に示すとおり、SONOS SGIB ANDセル構造はN+埋め込み拡散部を有さない。従って、ビットラインとサブゲートの間のオフセットは存在しない。SONONOSデバイスは、拡散ビットラインの存在しない、SGを有するANDアレイを形成する目的で、並列に接続される。

【0056】

BE SONOSアレイアーキテクチャ800は、複数のビットラインBL0820、BL1821、BL2822、BL3823及びBL4824と交差する複数のワードラインWL0810、WL1811、WL2812及びWlm813を含む。対応するサブゲートラインはそれぞれのビットラインに並列である。サブゲートSG0830はビットラインBL0820に並列に配置される。サブゲートSG1831はBL1821に並列に配置される。サブゲートSG2832はビットラインBL2822に並列に配置される。サブゲートSG3833はビットラインBL3823に並列に配置される。1つのメモリセルとして機能するBE SONOS (又はSONONOS) デバイス840の1サンプルを丸で囲んだ領域に示す。

【0057】

SGIB ANDアレイアーキテクチャ850においては、4番目ごとのサブゲートが共に電氣的に接続される。すなわち、SG0830、SG4834及びSG8等が共に電氣的に接続され、SG1831、SG5及びSG9等が共に電氣的に接続され、SG2832、SG6及びSG10が電氣的に共に接続され、SG3833、SG7及びSG11が共に電氣的に接続される。

【0058】

図8Bのレイアウト図850に示すとおり、SGIB ANDアーキテクチャ800のメモリアレイにおいては、SG0830、SG1831、SG2832、SG3833、SG4834及びSG835のそれぞれの下にはN+領域は存在しない。SGIB ANDアーキテクチャ800のセルサイズの全体寸法を、図2Aで示したSG ANDアレイアーキテクチャ200におけるセルサイズに比べて減らせる。

【0059】

SG0830、SG1831、SG2832、SG3833、SG4834及びSG835がオンになると、それぞれのSGは、ソース/ドレイン860、ソース/ドレイン861、ソース/ドレイン862、ソース/ドレイン863、ソース/ドレイン864及びソース/ドレイン865としてそれぞれ機能するバリア拡散層として有効に働くNチャンネル反転層を形成する。それゆえ、SG0830、SG1831、SG2832、SG3833、SG4834及びSG835のそれぞれのサブゲートは、二重の機能を果たす。SG0830、SG1831、SG2832、SG3833、SG4834及びSG835のそれぞれのサブゲートが果たす第1の機能は、ソースサイドインジェクションプログラミングのためのサブゲートである。SG0830、SG1831、SG2832、SG3833、SG4834及びSG835のそれぞれのサブゲートが果たす第2の機能は、

サブゲートがオンになったときの反転ビットラインである。ソース/ドレイン 860、ソース/ドレイン 861、ソース/ドレイン 862、ソース/ドレイン 863、ソース/ドレイン 864 及びソース/ドレイン 865 を含むそれぞれのソース/ドレインは、メタルビットラインに接続するためのものである。記号 Lg_{870} は図で示すとおりチャンネル長を表す。記号 W_{874} はチャンネル幅を表す。通常、パラメータ W_{874} 、 W_{s876} 、 Lg_{870} 及び L_{s872} はパラメータ F に略等しい。ここでパラメータ F は、技術ノードにおける限界寸法を表す。例えば、パラメータ F は 50 nm ノードに対して 50 nm となる。

【0060】

図 9A に、第 2 の形態における、SONONOS SGIB AND アレイアーキテクチャのチャンネル長さ方向の断面図を示すレイアウト図 900 を示す。一方で、図 9B に、第 2 の形態における、SONONOS SGIB AND アレイアーキテクチャのチャンネル幅方向の断面図を示すレイアウト図を示す。SG910、SG912、SG914、SG916 及び SG918 を含むそれぞれのサブゲートは、ソース領域又はドレイン領域の埋め込み部分を持たない。つまり、SG910、SG912、SG914、SG916 及び SG918 を含むそれぞれのサブゲートの下には N^+ 領域が存在しない。チャンネル幅方向のピッチは、チャンネル長さ方向のピッチ $2F$ に略等しい。従って、BE SONOS SGIB AND アレイアーキテクチャはセル当たり略 $4F^2$ となる。

【0061】

図 10A に、第 2 の形態における、SONONOS SGIB AND アレイアーキテクチャ 1000 の電気的リセットを示す回路図を示す。電気的リセット中、ワードライン（又はゲート） WL_{0810} 、 WL_{1811} 、 WL_{2812} 及び W_{lm813} は -10 V にセットされ、ビットライン BL_{0820} 、 BL_{1821} 、 BL_{2822} 、 BL_{3823} 、 BL_{4824} 、 BL_{5825} はフローティング状態にされ、サブゲート SG_{0830} 、 SG_{1831} 、 SG_{2832} 、 SG_{3833} 、 SG_{4834} 及び SG_n835 は 0 V にセットされ、P ウェル 1010 は 5 V にセットされる。サブゲート SG が 0 V であるとき、 SG はオン状態でないため反転ビットラインは存在しない。1 つの形態では、例えば SG_{0830} が SG_4 と接続され、 SG_{1831} が SG_5 と接続されるように、4 番目ごとのサブゲートが共に接続される。

【0062】

動作前に、 $V_{gb} = -15 V$ を印加することによって（又は、ゲート電圧をそれぞれの WL 及び P ウェルに分配することによって）、メモリ回路 1000 はリセットされる。それにより、図 10B のグラフ 1050 で示すとおり、望ましい自己収束特性を実現できる。グラフ 1050 内のさまざまな丸及び三角は、広分布に渡る異なる初期値を表している。そして、これらの初期値は閾値電圧 V_t に収束している。たとえ BE SONOS デバイスが最初にさまざまな V_t をチャージしていたとしても、これらのリセット動作により、これら初期値をリセット/消去状態に収束させることができる。リセット時間は通常 100 ミリ秒ぐらいである。1 つの例では、 n チャンネル BE SONOS デバイスは以下のパラメータを有する。 $ONONO = 15 / 20 / 18 / 70 / 90$ オングストローム、 $Lg / W = 0.22 / 0.16 \mu m$ 、 N^+ ポリゲート。

【0063】

別の言い方をすれば、動作前に V_t の分布を収束させる目的でリセット動作は実行される。自己収束消去を有さないフローティングゲートデバイスと異なり、BE SONOS は、自己収束消去リセット/消去方式を備えている。最初の V_t 分布は多くの場合、プラズマ帯電効果等のプロセス問題に起因して、広い分布となるので、この方式が必要となる。自己収束リセットは、最初の V_t 分布の収束を促進する。

【0064】

図 11A は、第 2 の形態における SONONOS SGIB AND アレイアーキテクチャ 1100 の電気的プログラムを示す回路図である。一方、図 11B は、第 2 の形態における SONONOS SGIB AND アレイアーキテクチャの電気的プログラムを示

10

20

30

40

50

すレイアウト図 1 1 5 0 である。1つの例では、セル A 1 1 1 0 の電氣的プログラム中、ワードライン W L 1 8 1 1 は 1 0 V にセットされ、その他のワードライン W L 0 8 1 0、W L 2 8 1 2 及び W L m 8 1 3 は 0 V にセットされる。ビットライン B L 0 8 2 0、B L 1 8 2 1 及び B L 3 8 2 3 はフローティング状態にされる。ビットライン B L 2 は 0 V にセットされ、ビットライン B L 4 は 5 V にセットされる。サブゲート S G 0 8 3 0 及び S G 4 8 3 4 は 8 V にセットされる。サブゲート S G 1 は 0 V にセットされ、サブゲート S G 2 は 5 V にセットされ、サブゲート S G 3 は 1 V にセットされる。

【 0 0 6 5 】

ビットライン B L 0 8 2 0、B L 1 8 2 1、B L 2 8 2 2、B L 3 8 2 3 及び B L 4 8 2 4 は、サブゲート S G 0 8 3 0、S G 1 8 3 1、S G 2 8 3 2、S G 3 8 3 3 及び S G 4 8 3 4 よりもプログラミングにおいてより高い柔軟性を提供する。なぜなら、それぞれのビットラインは、個別にプログラムできるからである。電氣的プログラミングの1つの方法は、ソースサイドインジェクションである。ソースサイドインジェクションはセルを高電圧閾値 V_t 状態にプログラムする。例えば、ソースインジェクションにより、選択された W L 1 に $V_g = 10 V$ が印加され、他のワードラインには $V_g = 0 V$ が印加される。プログラミング時は $S G = 1 V$ であり、プログラミング禁止時は $S G = 0 V$ である。S G 電圧の電圧設定値は 1 V であるが、これは S G ゲート下の閾値電圧より通常 0.5 ~ 2 V 高い値を意図している。

【 0 0 6 6 】

電氣的プログラムを実行する目的で、ターゲットのセルをプログラミングするためのソースサイドインジェクションを行うべく、選択されたワードラインに高電圧の 10 V が印加され、サブゲート S G 3 8 3 3 には 1 V が印加される。S G 1 8 3 1 は、プログラムを禁止するため 0 V にセットされ、S G 4 8 3 4 は、ビットライン抵抗を減らすべく十分なオーバドライブを提供する目的で、8 V にセットされる。2 k B のセルをパラレルに有するページプログラミング等のパラレルプログラミング手法によって、プログラミングのスループットを 1 秒当たり 10 M B 以上も増加でき、一方でセル電流消費値は 2 m A 以内に抑えられることは当業者であれば理解できるであろう。他のビットラインへのプログラム擾乱を防ぐ目的で、サブゲート S G 2 2 3 1 は 0 V にセットされ、禁止セルはオフにされる。

【 0 0 6 7 】

電氣的プログラミングとは、セルを高電圧閾値 V_t 状態にプログラムする目的で、ソースサイドインジェクションを実行することである。例えば、セル A 1 1 1 0 の電氣的プログラミングにおいて、選択された W L 1 8 1 1 に $V_g = 10 V$ を印加して、W L 0 8 1 0、W L 2 8 1 2 及び W L m 8 1 3 を含む他のワードラインに $V_g = 0 V$ を印加する。プログラミング時は $S G 3 = 1 V$ に、プログラミング禁止時は $S G 1 = 0 V$ に、そしてパステート時には $S G 2 = 5 V$ にセットされる。サブゲート S G 4 8 3 4 を高いオン状態にして、反転層の電位を 5 V まで引き上げる目的で、サブゲート S G 4 8 3 4 は 8 V にセットされる。セル A 1 1 1 0 のプログラミングをするときには、ソースサイドインジェクションが行われるようサブゲート S G 3 8 3 3 は 1 V にセットされる。閾値電圧 V_t はプログラミング電圧 P V よりも大きくなる。セル B 1 0 1 2、セル C 1 0 1 4 及びセル D 1 0 1 6 はプログラム禁止状態となる。

【 0 0 6 8 】

図 1 2 A は、第 2 の形態における B E S O N O N O S S G I B A N D アレイアーキテクチャの電氣的消去を示す回路図 1 2 0 0 である。一方で、図 1 2 B は、第 2 の形態に関する望ましい自己収束消去特性を示すグラフ図 1 2 5 0 である。消去動作はリセット動作と同様である。電氣的消去時には、ワードライン W L 0 8 1 0、W L 1 8 1 1、W L 2 8 1 2 及び W L m 8 1 3 は - 1 0 V にセットされ、ビットライン B L 0 8 2 0、B L 1 8 2 1、B L 2 8 2 2、B L 3 8 2 3、B L 4 8 2 4 及び B L 5 8 2 5 はフローティング状態にされ、サブゲート S G 2 8 3 0、S G 1 8 3 1、S G 2 8 3 2、S G 3 8 3 3 及び S G 4 8 3 4 は 0 V にセットされる。電氣的消去動作は、セクタ又はブロック単位で実行

10

20

30

40

50

される。図12Bのグラフ1250で示されるとおり、BE SONOSデバイスは望ましい自己収束消去特性を備える。消去飽和値 V_t は V_g に依存する。 V_g が高いほど、飽和値 V_t もより高くなる。収束時間は通常、10～100ミリ秒ぐらいである。

【0069】

図13Aは、第2の形態におけるSONOS SGIB ANDアレイアーキテクチャの読み出し動作を示す回路図1300である。一方で、図13Bは第2の形態におけるSONOS SGIB ANDアレイアーキテクチャの読み出し動作を示すレイアウト図1350である。1つの例では、セルA1110の読み出し動作中には、ワードラインWL1811は5Vにセットされ、その他のワードラインWL0810、WL2812及びWLM813は0Vにセットされる。ビットラインBL0820、BL1821及びBL4824はフローティング状態にされる。ビットラインBL2822は0Vにセットされ、ビットラインBL3823は1Vにセットされる。サブゲートSG0830、SG1831及びSG4834は0Vにセットされ、SG2832及びSG3833は5Vにセットされる。読み出し動作は、消去状態 V_t (EV)とプログラム状態 V_t (PV)との間のゲート電圧を印加することによって実行される。ゲート電圧は通常、5Vぐらいである。あるいは、高 V_t 値から低 V_t 値までの範囲であれば、ゲート電圧として5Vより大きい値又は小さい値を選択することもできる。セルA1110の V_t が5Vより高い場合は、読み出し電流は微小値(例えば $< 0.1 \mu A$)になるであろう。セルA1110の V_t が5Vより低い場合は、読み出し電流は大きい値(例えば $> 0.1 \mu A$)になるであろう。それから、メモリの状態が特定される。

10

20

【0070】

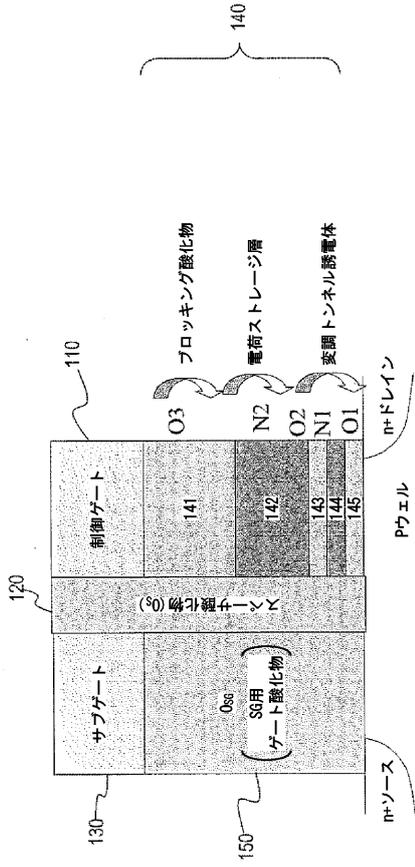
ビットライン(BL)への印加電圧は通常、1Vぐらいである。読み出し電圧が大きいほど電流も大きくなるが、読み出しによる擾乱もより大きくなるであろう。SG ANDストリングのWL数は通常64、128又は256である。SG ANDストリングの数が大きくなるほど、オーバーヘッドをより防ぐことができ、アレイ効率も増加するであろう。しかしながら、プログラム分布はより大きくなるであろう。トレードオフを検討して、SGIB ANDストリングの数を適切に選択する必要がある。

【0071】

上記読み出し機能はランダムアクセス読み出し動作であるが、本発明の精神から逸脱しない範囲において、複数のセルのページ読み出しができることは当業者であれば理解できるであろう。本発明は特定の実施形態に基づいて説明してきたが、本発明の精神及び範囲を逸脱しない範囲においてさまざまな改良、適合及び変更を行っても良い。従って、本明細書及び図面は本発明の原理を説明するためのものであり、本発明は以下の請求項によって定義される。

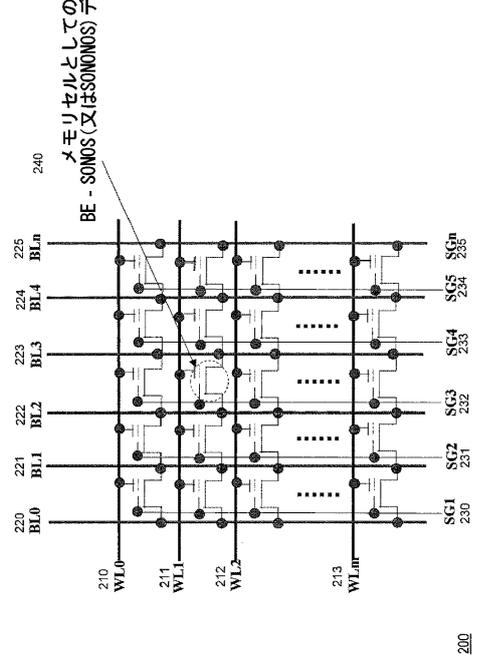
30

【図1】



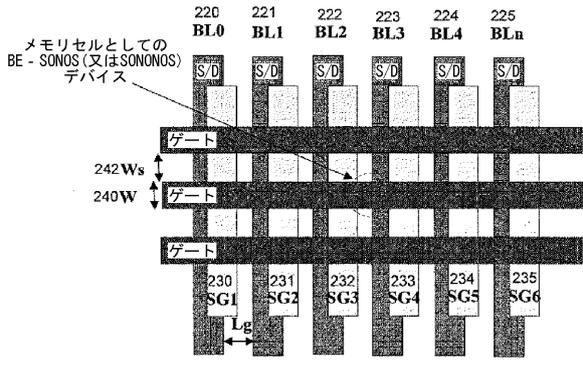
100

【図2A】



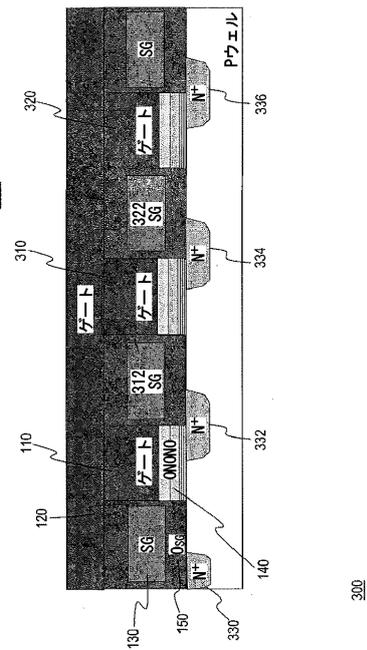
200

【図2B】



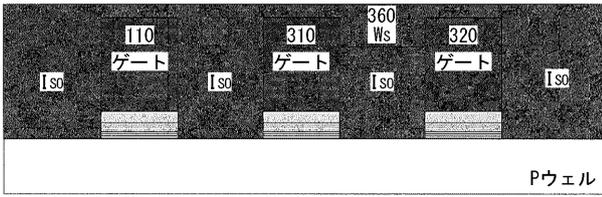
250

【図3A】



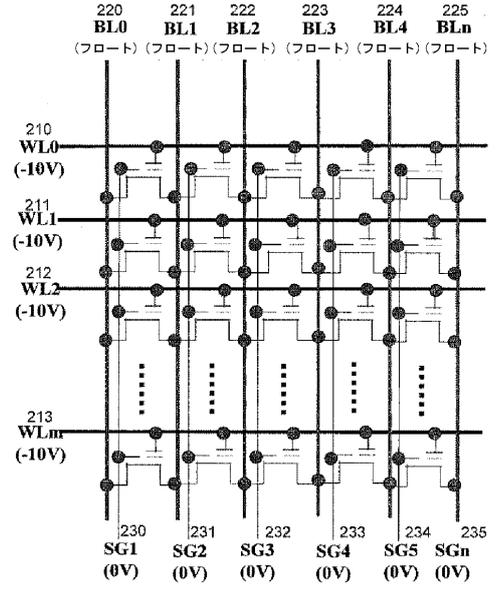
300

【 図 3 B 】



350

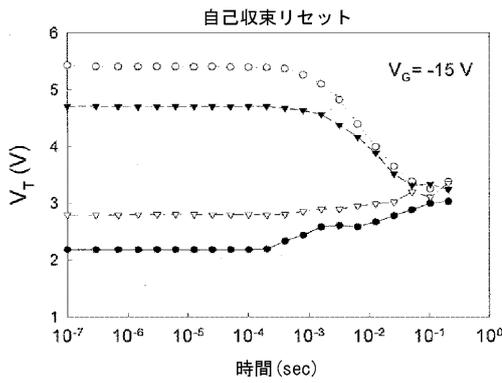
【 図 4 A 】



400

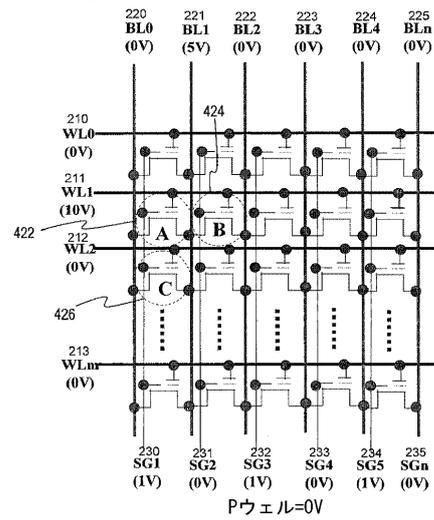
Pウェル=5V

【 図 4 B 】



450

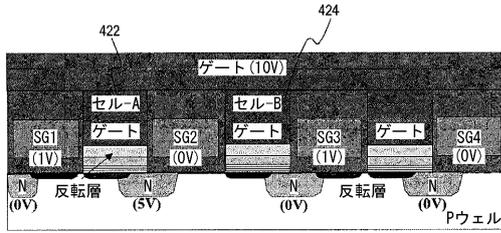
【 図 5 A 】



500

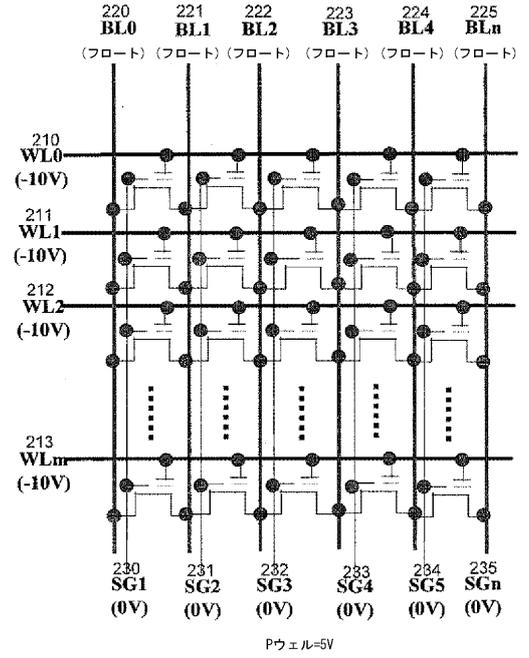
Pウェル=0V

【図5B】



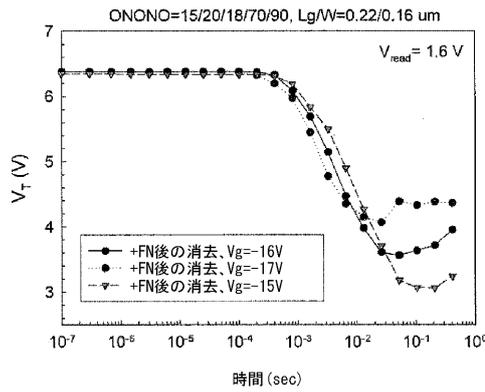
550

【図6A】



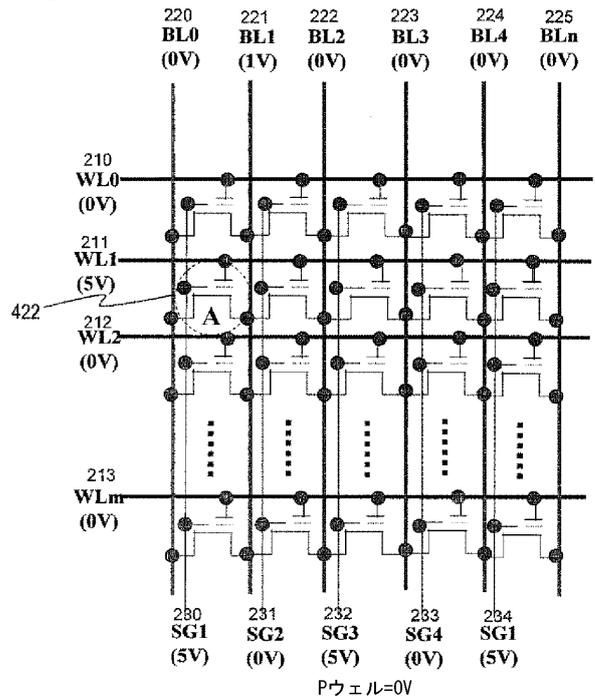
600

【図6B】



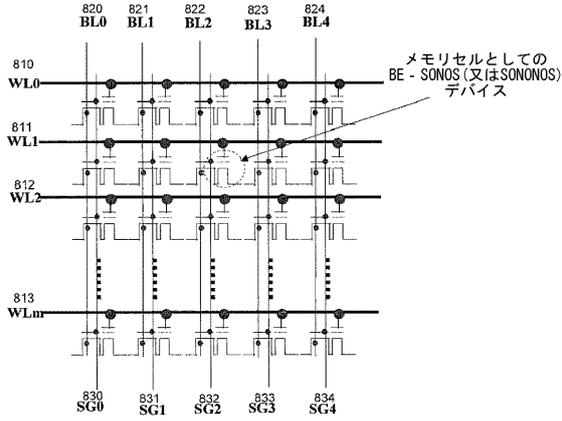
650

【図7】



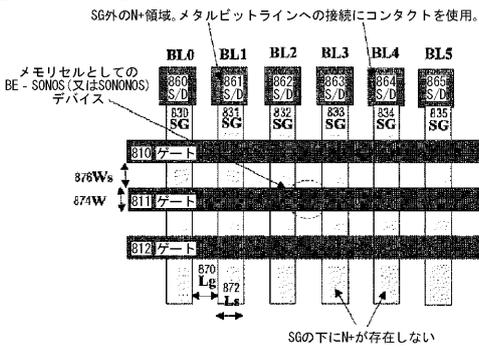
700

【図8A】



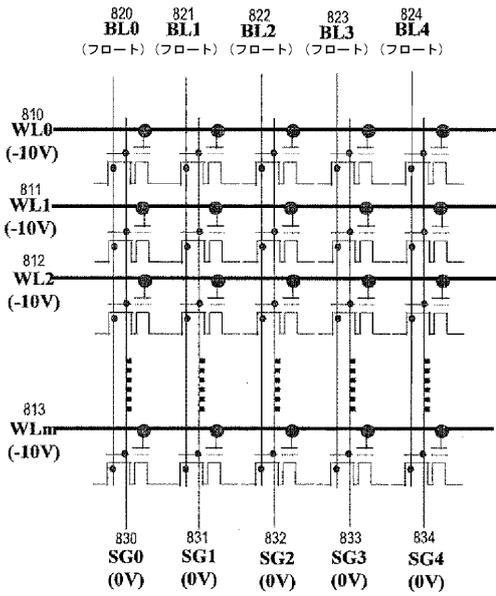
800

【図8B】



850

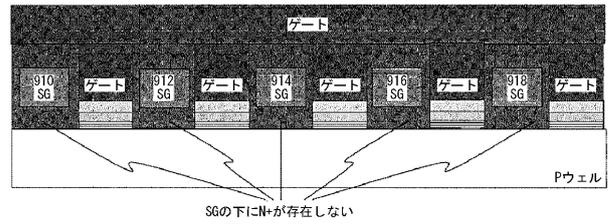
【図10A】



1000

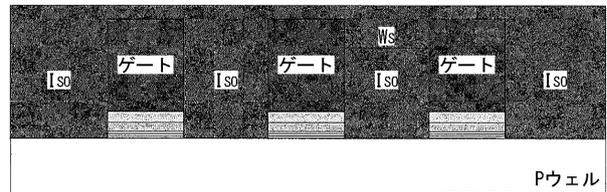
1010
Pウェル=5V

【図9A】



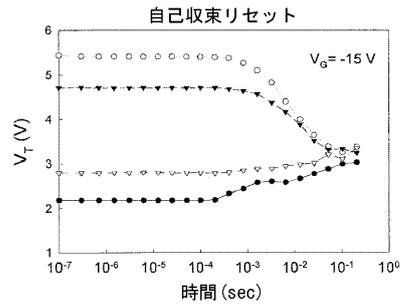
900

【図9B】



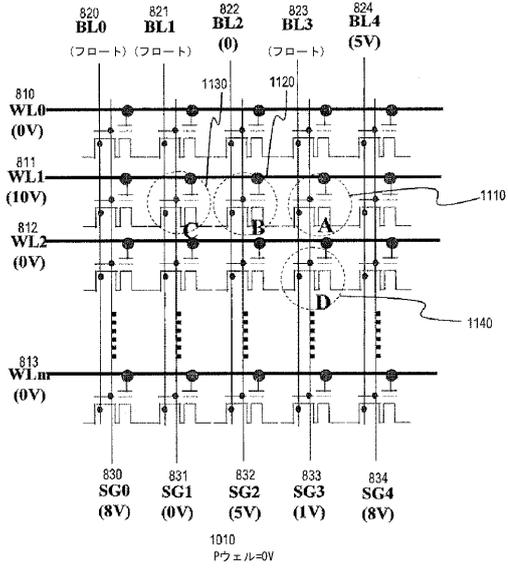
950

【図10B】



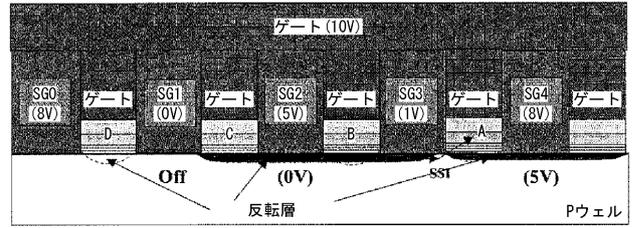
1050

【図11A】

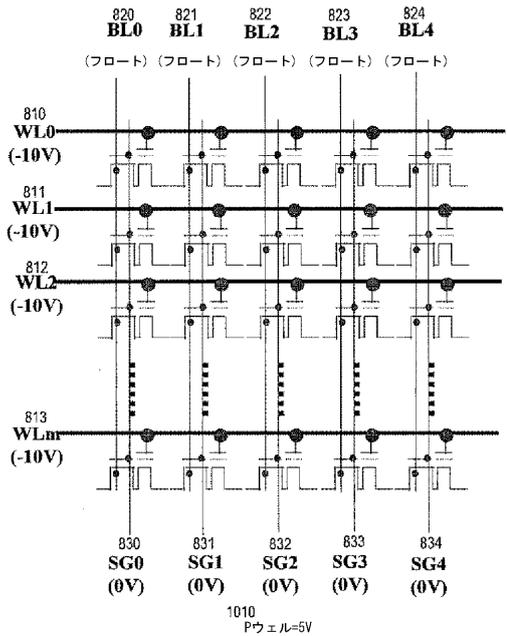


1100

【図11B】

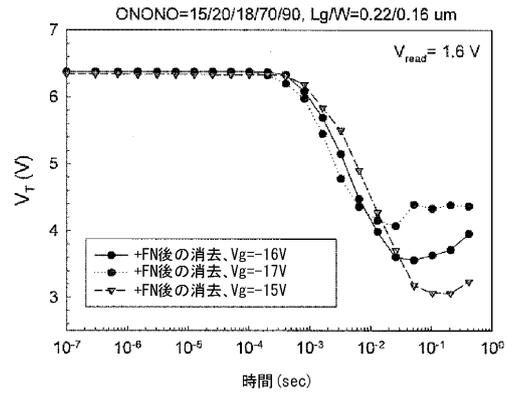


【図12A】

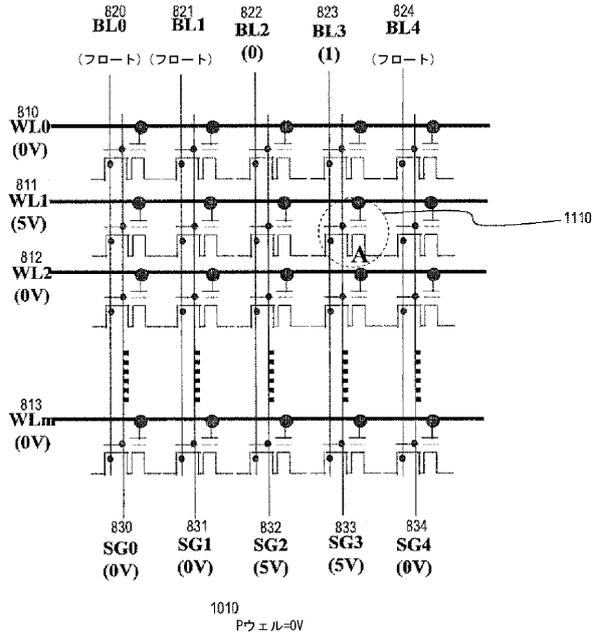


1200

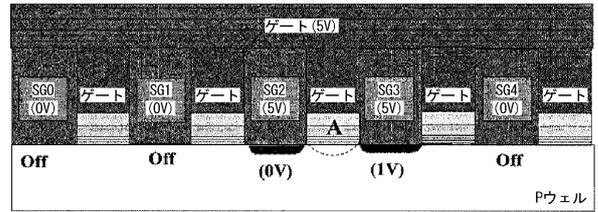
【図12B】



【図 13 A】



【図 13 B】



1300

1350

フロントページの続き

(51) Int.Cl.

G 1 1 C 16/02 (2006.01)

F I

G 1 1 C 17/00 6 1 1 G

G 1 1 C 17/00 6 1 2 F

テーマコード(参考)

Fターム(参考) 5F083 EP17 EP18 EP22 EP30 EP43 EP44 EP79 ER01 ER03 ER06
ER11 ER23 ER30 GA09 GA11 JA04 KA01 KA08 KA17 LA02
LA11 LA12 LA16 ZA21
5F101 BA45 BA54 BB02 BB09 BC01 BD02 BD10 BD34 BE02 BE05
BE07 BF02 BF05

【外国語明細書】

Structure and Method of Sub-Gate AND Architectures Employing Bandgap Engineered SONOS Devices

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The present invention relates generally to non-volatile memory devices, and more particularly, to nitride-based trapping-storage flash memories.

Description of Related Art

[0002] Electrically programmable and erasable non-volatile memory technologies based on charge storage structures known as Electrically Erasable Programmable Read-Only Memory (EEPROM) and flash memory are used in a variety of modern applications. A flash memory is designed with an array of memory cells that can be independently programmed and read. Sense amplifiers in a flash memory are used to determine the data value or values stored in a non-volatile memory. In a typical sensing scheme, an electrical current through the memory cell being sensed is compared to a reference current by a current sense amplifier.

[0003] A number of memory cell structures are used for EEPROM and flash memory. As the dimensions of integrated circuits shrink, greater interest is arising for memory cell structures based on charge trapping dielectric layers, because of the scalability and simplicity of the manufacturing processes. Memory cell structures based on charge trapping dielectric layers include structures known by the industry names Nitride Read-Only Memory, Poly-Silicon-Oxide-Nitride-Oxide-Silicon (SONOS), and PHINES, for example. These memory cell structures store data by trapping charge in a charge trapping dielectric layer, such as silicon nitride. As negative charge is trapped, the threshold voltage of the memory cell increases. The threshold voltage of the memory cell is reduced by removing negative charge from the charge trapping layer.

[0004] Nitride Read-Only Memory devices use a relatively thick bottom oxide, e.g. greater than 3 nanometers, and typically about 5 to 9 nanometers, to prevent charge

loss. Instead of direct tunneling, band-to-band tunneling induced hot hole injection BTBTHH can be used to erase the cell. However, the hot hole injection causes oxide damage, leading to charge loss in the high threshold cell and charge gain in the low threshold cell. Moreover, the erase time must be increased gradually during program and erase cycling due to the hard-to-erase accumulation of charge in the charge trapping structure. This accumulation of charge occurs because the hole injection point and electron injection point do not coincide with each other, and some electrons remain after the erase pulse. In addition, during the sector erase of a Nitride Read-Only Memory flash memory device, the erase speed for each cell is different because of process variations (such as channel length variation). This difference in erase speed results in a large V_t distribution of the erase state, where some of the cells become hard to erase and some of them are over-erased. Thus the target threshold V_t window is closed after many program and erase cycles and poor endurance is observed. This phenomenon will become more serious when the technology keeps scaling down.

[0005] A typical flash memory cell structure positions a tunnel oxide layer between a conducting polysilicon tunnel oxide layer and a crystalline silicon semiconductor substrate. The substrate refers to a source region and a drain region separated by an underlying channel region. A flash memory read can be executed by a drain sensing or a source sensing. For source side sensing, one or more source lines are coupled to source regions of memory cells for reading current from a particular memory cell in a memory array.

[0006] A traditional floating gate device stores 1 bit of charge in a conductive floating gate. The advent of Nitride Read-Only Memory cells in which each cell provides 2 bits of flash cells that store charge in an Oxide-Nitride-Oxide (ONO) dielectric. In a typical structure of a Nitride Read-Only Memory memory cell, a nitride layer is used as a trapping material positioned between a top oxide layer and a bottom oxide layer. The ONO layer structure effectively replaces the gate dielectric in floating gate devices. The charge in the ONO dielectric with a nitride layer may be either trapped on the left side or the right side of a NROM cell.

[0007] Floating gate devices encounter substantial scaling challenges due to inter-floating gate coupling, while nitride trapping device is flexible from such limitations.

There are two main types of nitride trapping device: NROM that stores charges locally and SONOS that uses channel program/erase. These two types of devices have drawbacks. A Nitride Read-Only Memory device is sensitive to hot-hole induced damages, and a SONOS device suffers from retention problems caused by direct tunneling leakage through the thin tunnel oxide.

[0008] A conventional AND-type floating gate flash memory is suitable for many commercial applications because the memory device possesses the characteristics of high-density, low-power and fast speed programming. However, due to the inter-floating gate coupling effect, the scaling of AND-type floating gate devices is limited. When the space parameter for the floating gate device is shrunk, a high floating gate coupling effect may cause undesirable and severe disturbance. The conventional AND-type floating gate device also suffers from tunnel oxide scaling issues and erratic bits where a local defect, or trapped charge, in a tunnel oxide can result in the leakage of the charge in the floating gate.

[0009] To address the scaling issue in floating gate devices, charge trapping devices such as SONOS, MNOS or nano-crystal trapping devices are suggested. However, these devices all suffer serious charge retention problems. For a SONOS device, the ultra-thin tunnel oxide is unable to properly preserve a charge storage. For a MNOS device, the structure does not provide a top oxide to block the charge loss. A nano-crystal device cannot be well-controlled because of the randomly distributed nano particles.

[0010] Accordingly, it is desirable to design AND-type floating gate flash memories that provide scalability while overcoming the retention problems as well as maintaining efficient hole tunneling erase.

SUMMARY OF THE INVENTION

[0011] The present invention provides a bandgap engineered SONOS (referred to as "BE-SONOS" or "SONONOS") device structure for design with various AND architectures to perform a source side injection (SSI) programming method. The BE-SONOS device structure comprises a spacer oxide disposed between a control gate overlaying an oxide-nitride-oxide-nitride-oxide ($O_3-N_2-O_2-N_1-O_1$) stack and a sub-gate overlaying a gate oxide. In a first embodiment, a BE-SONOS SG-AND (sub-gate-AND)

array architecture is constructed multiple columns of SONOS devices with sub-gate lines and diffusion bitlines. In a second embodiment, a BE-SONOS SGIB-AND (sub-gate-inversion-bitline-AND) architecture is constructed multiple columns of SONOS devices with sub-gate inversion bitlines and with no diffusion bitlines.

[0012] Broadly state, an integrated circuit device comprises a semiconductor substrate; a plurality of memory cells on the semiconductor substrate, each memory cell having a spacer oxide disposed between a gate and a sub-gate, each gate overlaying a blocking oxide-charge storage layer-modulated tunnel dielectric stack, each sub-gate overlaying a gate oxide; and an N⁺ buried diffusion disposed in the semiconductor substrate and positioned underneath between a first gate oxide and a first blocking oxide-charge storage layer-modulated tunnel dielectric stack that serves as a first diffusion bitline.

[0013] Advantageously, the BE-SONOS AND array architectures of the present invention provides greater scalability over floating gate and AND type of memory devices. The present invention also advantageously provides a uniform and self-converging channel hole tunneling erase operation. Moreover, the present invention eliminates the inter-floating gate coupling effect. The present invention further provides desirable reliability properties including predictable characteristic of excellent charge retention, predictable number of nearly no erratic bits, and predicable small degradation after program and erase cycles.

[0014] The structures and methods regarding to the present invention are disclosed in the detailed description below. This summary does not purport to define the invention. The invention is defined by the claims. These and other embodiments, features, aspects, and advantages of the invention will become better understood with regard to the following description, appended claims and accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] Fig. 1 is a structural diagram illustrating a cross-sectional view of a single cell structure of an n-channel BE-SONOS device with sub-gate in accordance with the present invention.

[0016] Fig. 2A is a circuit diagram illustrating a first embodiment of a BE-SONOS SG-AND array architecture with diffusion bitlines in accordance with the present invention; Fig. 2B is a layout diagram illustrating the first embodiment of the BE-SONOS SG-AND array architecture with diffusion bitlines in accordance with the present invention.

[0017] Fig. 3A is a layout diagram illustrating a cross-sectional view in channel length direction of the BE-SONOS SG-AND array architecture in the first embodiment in accordance with the present invention; Fig. 3B is a layout diagram 350 illustrating a cross-sectional view in channel width direction of the BE-SONOS SG-AND array architecture in the first embodiment in accordance with the present invention.

[0018] Fig. 4A is a circuit diagram illustrating an electrical reset for the SONOS SG-AND array architecture in the first embodiment in accordance with the present invention; Fig. 4B is a graphical diagram illustrating a waveform of a self-converging reset in the first embodiment in accordance with the present invention.

[0019] Fig. 5A is a circuit diagram illustrating an electrical program for the SONOS SG-AND array architecture in the first embodiment in accordance with the present invention; Fig. 5B is a layout diagram illustrating the electrical program for the SONOS SG-AND array architecture in the first embodiment in accordance with the present invention.

[0020] Fig. 6A is a circuit diagram illustrating an electrical erase for the SONOS SG-AND array architecture of the first embodiment in accordance with the present invention; Fig. 6B is a graphical diagram illustrating a waveform of a self-converging erase in accordance with the present invention.

[0021] Fig. 7 is a circuit diagram illustrating a read operation for the SONOS SG-AND array architecture in the first embodiment in accordance with the present invention.

[0022] Fig. 8A is a circuit diagram illustrating a second embodiment of a BE-SONOS SGIB-AND array architecture in accordance with the present invention; Fig. 8B is a layout diagram illustrating the second embodiment of the BE-SONOS SGIB-AND array architecture in accordance with the present invention.

[0023] Fig. 9A is a layout diagram illustrating the cross sectional view in channel length direction of the SONONOS SGIB-AND array architecture in the second embodiment in accordance with the present invention; Fig. 9B is a layout diagram illustrating the cross sectional view in channel width direction of the SONONOS SGIB-AND array architecture in the second embodiment in accordance with the present invention.

[0024] Fig. 10A is a circuit diagram illustrating an electrical reset of the SONONOS SGIB-AND array architecture in the second embodiment in accordance with the present invention; Fig. 10B is a graphical diagram illustrating a waveform of a self-converging reset in accordance with the present invention.

[0025] Fig. 11A is a circuit diagram illustrating an electrical program of the SONONOS SGIB-AND array architecture in the second embodiment in accordance with the present invention; Fig. 11B is a layout diagram illustrating the electrical program for the SONONOS SGIB-AND array architecture in the second embodiment in accordance with the present invention.

[0026] Fig. 12A is a circuit diagram illustrating an electrical erase of the BE-SONONS SGIB-AND array architecture of the second embodiment in accordance with the present invention; Fig. 12B is a graph diagram illustrating a waveform of a self-converging erase of the second embodiment in accordance with the present invention.

[0027] Fig. 13A is a circuit diagram illustrating a read operation of the SONONOS SGIB-AND array architecture in the second embodiment in accordance with the present invention; Fig. 13B is a layout diagram illustrating the read operation of the SONONOS SGIB-AND array architecture of the second embodiment in accordance with the present invention.

DETAILED DESCRIPTION OF EMBODIMENTS OF THE INVENTION

[0028] Referring now to Fig. 1, there is shown a structural diagram illustrating a cross-sectional view of a single cell structure of an n-channel BE-SONOS device 100 with sub-gate (SG). The n-channel BE-SONOS device 100 includes a spacer oxide (Os) 120 that is disposed between a control gate 110 and a sub-gate 130. An oxide-nitride-oxide-nitride-oxide ($O_3-N_2-O_2-N_1-O_1$) structure 140 is disposed underneath the control

gate 110. A gate oxide O_{SG} 150 is disposed underneath the sub-gate 130. The O_3 - N_2 - O_2 - N_1 - O_1 structure 140 includes a blocking oxide O_3 141, a charge storage layer N_2 142, and a modulated tunnel dielectric O_2 - N_1 - O_1 143-145. The bottom O_2 - N_1 - O_1 143-145 layers provide hole tunneling current and good data retention.

[0029] The n-channel BE-SONOS device 100 is a five-terminal device with two gates, the control gate 110 and the sub-gate 130. Underneath the control gate 110, there is the O_3 - N_2 - O_2 - N_1 - O_1 structure 140 for the charge storage. Underneath the sub-gate 130, there is the non-trapping gate oxide 150. The control gate 110 can control program, erase, and read the charge storage layer. The sub-gate 130 can provide source side injection (SSI) programming method. The source side injection is a low-power and high-speed programming method. The O_1 - N_1 - O_2 layer can be implemented with ultra-thin oxide and nitride, typically within 3 nm to provide hole direct tunneling. The N_2 layer 142 is thicker than 5 nm to provide higher trapping efficiency. In the layer 141 formation method, one technique is to use a wet converted top oxide to provide a large density of traps at the interface between O_3 and N_2 . The O_3 layer is typically thicker than 6 nm in order to prevent charge loss from top oxide. The O_1 - N_1 - O_2 layers serve as a tunneling dielectric for the hole tunneling.

[0030] An exemplary set of device parameters for the n-channel BE-SONOS device 100 with the sub-gate 130 is shown below.

Bottom Oxide (O_1)	15 Å
Inter Nitride (N_1)	20 Å
Inter Oxide (O_2)	18 Å
Trapping Nitride (N_2)	90 Å
Gate Oxide for SG (O_{SG})	150 Å
Spacer Oxide (O_S)	200 Å
Gate material	N+ - poly or P+ - poly gate

[0031] In Fig. 2A, there is shown a circuit diagram illustrating a first embodiment of a BE-SONOS SG-AND array architecture 200 with diffusion bitlines. A plurality of SONONOS devices are connected in parallel to form the BE-SONOS SG-AND array

architecture 200. The BE-SONOS array architecture 200 comprises a plurality of wordlines WL0 210, WL1 211, WL2 212, WLm 213 intersecting a plurality of bitlines BL0 220, BL1 221, BL2 222, BL3 223, BL4 224 and BLn 225. A corresponding sub-gate line is parallel and located nearby a bitline. A sub-gate SG1 230 is located adjacent to the bitline BL0 220. A sub-gate SG2 231 is located adjacent to the bitline BL1 221. A sub-gate SG3 232 is located adjacent to the bitline BL2 222. A sub-gate SG4 233 is located adjacent to the bitline BL3 223. A sub-gate SG5 234 is located adjacent to the bitline BL4 224. A sub-gate SGn 235 is located adjacent to the bitline BL5 225. A sample BE-SONOS (or SONONOS) device 240 that functions as a memory cell is shown in a circled area.

[0032] As shown in Fig. 2B, there is a layout diagram 250 illustrating the first embodiment of the BE-SONOS SG-AND array architecture 200 with diffusion bitlines. Although each sub-gate SG is in parallel with a corresponding bitline, each SG has a slight offset from the corresponding bitline, e.g. the SG1 230 is positioned slightly to the right of BL0 220. Every bitline can either serve as a source or a drain. Each SG is positioned between two bitlines, e.g. the SG1 230 between BL0 220 and BL1 221. Each bitline in the BL0 220, BL1 221, BL2 222, BL3 223, BL4 224, and BL5 225 can function as a source region or a drain region. Therefore, the SG1 230 is disposed between the source region in the BL0 220 and the drain region in the BL1 221. The parameters W 240 and Ws 242 are approximately equal to the parameter F, where the parameter F denotes the critical dimension in a technology node. For example, the parameter F is equal to 50 nm for a 50 nm node.

[0033] Fig. 3A is a layout diagram 300 illustrating a cross-sectional view in channel length direction of the BE-SONOS SG-AND array architecture in the first embodiment. The spacer oxide 120 separates the control gate 110 and the sub-gate 130. The O₃-N₂-O₂-N₁-O₁ structure 140 is disposed underneath the control gate 110. The gate oxide O_{SG} 150 is disposed underneath the gate 130. A suitable implementation of the control gate 110 is poly-1, and a suitable implementation of the sub-gate 130 is poly-2. N⁺ buried diffusion (BD) wells 330, 332, 334 and 336 are implemented for diffusion bitlines (BLs). In the layout diagram 300, a first cell structure comprises the control gate 110 and the sub-gate 130, with an adjacent and second cell structure that comprises a

gate 310 and a sub-gate 312, with an adjacent and third cell structure that comprises a gate 320 and a sub-gate 322.

[0034] In Fig. 3B, there is shown a layout diagram 350 illustrating a cross-sectional view in channel width direction of the BE-SONOS SG-AND array architecture in the first embodiment. The gap between the gate 310 and the gate 320 is denoted by a parameter W_s 360, which provides isolation between the gate 310 and the gate 320. Other similar isolations are shown between two gates to provide an isolation between two gates. The pitch in the channel width direction is approximately equal to $2F$ smaller than that in channel length direction $3F$ caused by a diffusion bitline. Therefore, the BE-SONOS SG-AND architecture is approximately equal to $6F^2$ per cell.

[0035] Fig. 4A is a circuit diagram 400 illustrating an electrical reset for the SONONOS SG-AND array architecture of the first embodiment. During the electrical reset, the wordlines (or gates) WL0 210, WL1 211, WL2 212, and WLm 213 are set to -10 volts, the bitlines BL0 220, BL1 221, BL2 222, BL3 223, BL4 224 and BL5 225 are left floating, and the sub-gates SG1 230, SG2 231, SG3 232, SG4 233, SG5 234, and SGn 235 are set to 0 volt. In one embodiment, the odd number sub-gates are electrically connected together, including SG1 230, SG3 232 and SG5 234, while the even number sub-gates are electrically connected together, including SG2 231, SG4 233 and SGn 235. Before operations, the memory circuit 400 is reset by applying $V_{gb} = -15V$ (or partition the gate voltage into each WL and p-well), which produces a desirable self-converging property, as shown in a graph 450 in Fig. 4B. Even if the BE-SONOS device is initially charged to various V_t , the reset operation can tighten these initial points to the reset/erase state. A typical reset time is around 100 msec. In one example, the n-channel BE-SONOS with ONONO = 15/20/18/70/90 Angstrom, and a N+-poly gate. $L_g/W = 0.22/0.16 \mu m$.

[0036] To state in another way, a reset operation is carried out to tighten the V_t distribution before operations. In contrast to a floating gate device where there is no self-converging erase, the BE-SONOS provides a self-converging erase reset/erase methods, which is necessary because the initial V_t distribution is often widely distributed due to the process issues, such as plasma charging effect. The self-converging reset assists in tighten the initial V_t distribution.

[0037] Fig. 5A is a circuit diagram 500 illustrating an electrical program for the SONONOS SG-AND array architecture in the first embodiment, while Fig. 5B is a layout diagram 550 illustrating the electrical program for the SONONOS SG-AND array architecture in the first embodiment. In one example during electrical program, the wordline WL1 211 is set to 10 volts while the other wordlines WL0 210, WL2 212, and WLm 213 are set 0 volt. The bitline BL1 221 is set to 5 volts, and the bitlines BL0 220, BL2 222, BL3 223, BL4 224, and BLn 225 are set to 0 volt. The odd number sub-gates SG1 230, SG3 232 and SG5 234 are set to 1 volt, while the even number sub-gates SG2 231, SG4 233, and SGn 235 are set to 0 volt. The bitlines BL0 220, BL1 221, BL2 222, BL3 223, BL4 224 and BL5 225 provide a greater flexibility in programming than sub-gates SG1 230, SG2 231, SG3 232, SG4 233, SG5 234, and SGn 235 because each bitline can be independently programmed, while the sub-gates are programmed based on the even number or odd number of sub-gates. One type of electrical programming method is a source side injection. The source side injection programs a cell to a high voltage threshold V_t state. For example, the source injection applies $V_g = 10V$ to the selected WL1, $V_g = 0V$ to other wordlines, $SG = 1V$ for programming, and $SG = 0V$ for inhibition. The voltage setting of the SG voltage at 1 volt is intended as an illustration such that in general it is typically 0.5 to 2 volts higher than the threshold voltage under SG gate.

[0038] When a cell-A 422 is selected programming, the SG is set to 1 volt so that the channel underneath SG is slightly turned on. Electrons are injected into the cell-A 422 by source side injection method to make the voltage threshold, V_t , higher than PV . The SG for a cell-B 424 is set to 0 volt, which turns SG off so that there is no injection into the cell-B 424. As for a cell-C 426, the SG is set to 1 volt where the WL = 0 volt which turns off cell-C 426 so that there is also no injection into the cell-C 426. As a result, programming can be randomly selected with adequate program inhibit technique.

[0039] To carry out an electrical program, the selected wordline is applied a high voltage, 10 volts, and the sub-gate is applied 1 volt to perform a source side injection. The source side injection is a low-power and high-speed programming method. One of skill in the art should recognize that parallel programming methods such as page programming with 2kB cells in parallel can burst the programming throughput to more

than 10 MB per second while the cell current consumption can be controlled within 2 mA. To avoid program disturbance to other bitlines, the sub-gate SG2 231 is set to 0 volt and turns off the inhibit cell.

[0040] As illustrated in Fig. 6A, there is a circuit diagram 600 showing an electrical erase for the SONOS SG-AND array architecture in the first embodiment. The erase operation is executed similar to the reset operation. During electrical erase, the wordlines WL0 210, WL1 211, WL2 212, and WLm 213 are set to -10 volts, the bitlines BL0 220, BL1 221, BL2 222, BL3 223, BL4 224 and BL5 225 are left floating, and the sub-gates SG1 230, SG2 231, SG3 232, SG4 233, SG5 234, and SGn 235 are set to 0 volt. The erase operation is performed in the unit of a sector or block. The BE-SONOS device produces desirable self-converging erase property, as shown in a graph 650 in Fig. 6B. The erase saturation V_t is dependent on the parameter V_g . A higher V_g causes a higher saturated V_t . The convergent time is typically around 10 to 100 msec.

[0041] Fig. 7 is a circuit diagram 700 illustrating a read operation for the SONOS SG-AND array architecture in the first embodiment. In one example during a read operation of the cell-A 422, the wordline WL1 211 is set to 5 volts while the other wordlines WL0 210, WL2 212, and WLm 213 are set 0 volt. The bitline BL1 221 is set to 1 volt, and the bitlines BL0 220, BL2 222, BL3 223, BL4 224, and BLn 225 are set to 0 volt. The odd number sub-gates SG1 230, SG3 232 and SG5 234 are set to 3 volts, while the even number sub-gates SG2 231 and SG4 233 are set to 0 volt. A read operation is performed by applying a gate voltage that is between an erased state V_t (EV) and a programmed state V_t (PV). The gate voltage is typically around 5 volts. Alternatively, the gate voltage can be selected to be more than 5 volts or less than 5 volts, provided the gate voltage falls in the range of a high V_t value and a low V_t value. If V_t of the cell-A 422 is higher than 5 volts, then the read current is likely to be a very small value (e.g., $<0.1 \mu\text{A}$). If V_t of the cell-A 422 is less than 5 volts, the read current is likely to be a high value (e.g., $>0.1 \mu\text{A}$).

[0042] The applied voltage at a bitline (BL) is typically around 1 volt. A larger read voltage will induce more current, but the read disturbance may be larger. The WL number of SG-AND string is typically 64, 128, or 256. A larger number of SG-AND string may save more overhead and increase the array efficiency. However, the

program distribution may be larger. A trade-off is weighed in choosing an adequate number of SG-AND string.

[0043] Although the above read function describes a random access read operation, one of ordinary skill in the art should recognize that a page read of multiple cells are possible without departing from the spirits of the present invention.

[0044] Turning now to Fig. 8A, there is shown a circuit diagram 800 illustrating a second embodiment a BE-SONOS (or SONONOS) SGIB-AND (Sub-gate Inversion Bitline-AND) array architecture, while Fig. 8B is a layout diagram 850 illustrating the second embodiment of the BE-SONOS SGIB-AND array architecture. The term SGIB means that a bitline is formed from an inversion layer by turning on a sub-gate. Unlike the first embodiment where the BE-SONOS structure as shown in the layout diagram 300 has a N+ buried diffusion, the SONOS SGIB-AND cell structure has no N+ buried diffusion, as shown in Fig. 9, and therefore, there is no offset between a bitline and a sub-gate. The SONONOS devices are connected in parallel to form an AND array with SG in which there are no diffusion bitlines.

[0045] The BE-SONOS array architecture 800 comprises a plurality of wordlines WL0 810, WL1 811, WL2 812, WLm 813 intersecting a plurality of bitlines BL0 820, BL1 821, BL2 822, BL3 823 and BL4 824. A corresponding sub-gate line is parallel to each bitline. A sub-gate SG0 830 is placed in parallel to the bitline BL0 820. A sub-gate SG1 831 is placed in parallel to the bitline BL1 821. A sub-gate SG2 832 is placed in parallel to the bitline BL2 822. A sub-gate SG3 833 is placed in parallel to the bitline BL3 823. A sample BE-SONOS (or SONONOS) device 840 that functions as a memory cell is shown in a circled area.

[0046] In the SGIB-AND array architecture 850, every fourth sub-gates are commonly electrically connected, i.e. the SG0 830, SG4 834, SG8, etc. are electrically connected together, the SG1 831, SG5, SG9, etc are electrically connected together, the SG2 832, SG6, SG10, etc are electrically connected together, and the SG3 833, SG7, SG11, etc are electrically connected together.

[0047] As shown in the layout diagram 850 in Fig. 8B, there are no N+ region underneath each of the SG0 830, the SG1 831, the SG2 832, the SG3 833, the SG4 834, and the SG 835 in the memory array of the SGIB-AND architecture 800. The overall

dimension of a cell size in the SGIB-AND architecture 800 is reduced relative to a cell size as shown in the SG-AND array architecture 200 in Fig. 2A.

[0048] When the SG0 830, the SG1 831, the SG2 832, the SG3 833, the SG4 834, and the SG 835 are turned on, each creates an N-channel inversion layer, which effectively serves as a barrier diffusion layer to function, respectively, as a source/drain 860, a source/drain 861, a source/drain 862, a source/drain 863, a source/drain 864 and a source/drain 865. Each sub-gate in the SG0 830, the SG1 831, the SG2 832, the SG3 833, the SG4 834, and the SG 835 therefore serves dual functions. The first function that each sub-gate in the SG0 830, the SG1 831, the SG2 832, the SG3 833, the SG4 834, and the SG 835 serves is a sub-gate for a source side injection programming. The second function that each sub-gate in the SG0 830, the SG1 831, the SG2 832, the SG3 833, the SG4 834, and the SG 835 serves is an inversion bitline when a subgate is turned on. Each source/drain in the source/drain 860, the source/drain 861, the source/drain 862, the source/drain 863, the source/drain 864, and the source/drain 865 is for connecting to a metal bitline. The symbol L_g 870 denotes the drawn channel length. The symbol W 874 denotes the channel width. Typically, the parameters W 874, W_s 876, L_g 870, L_s 872 are approximately equal to the parameter F , where the parameter F represents the critical dimension in a technology node. For example, the parameter F is equal to 50 nm for a 50 nm node.

[0049] As shown in Fig. 9A, there is a layout diagram 900 illustrating the cross sectional view in channel length direction of the SONONOS SGIB-AND array architecture in the second embodiment, while Fig. 9B is a layout diagram illustrating the cross sectional view in channel width direction of the SONONOS SGIB-AND array architecture in the second embodiment. Each subgate SG in SG 910, SG 912, SG 914, SG 916 and SG918 has no implant of source region or drain region, which means that there is no N+ region underneath each subgate SG in SG 910, SG 912, SG 914, SG 916 and SG 918. The pitch in a channel width direction is approximately equal to that in a channel length direction $2F$. Therefore, the BE-SONOS SGIB-AND architecture is approximately equal to $4F^2$ per cell.

[0050] In Fig. 10A, there is a circuit diagram illustrating an electrical reset of the SONONOS SGIB-AND array architecture 1000 in the second embodiment. During the

electrical reset, the wordlines (or gates) WL0 810, WL1 811, WL2 812, and WLm 813 are set to -10 volts, the bitlines BL0 820, BL1 821, BL2 822, BL3 823, BL4 824 and BL5 825 are left floating, the sub-gates SG0 830, SG1 831, SG2 832, SG3 833, SG4 834, and SGn 835 are set to 0 volt, and P-well 1010 is set to 5 volts. When a sub-gate SG = is equal to zero volt, that means the SG is not turned on so that there is no inversion bitline. In one embodiment, every fourth sub-gates are connected together such that SG0 830 is connected to SG4, SG1 831 is connected to SG5, and so on.

[0051] Before operations, the memory circuit 1000 is reset by applying $V_{gb} = -15V$ (or partition the gate voltage into each WL and p-well), which produces a desirable self-converging property, as shown in the graph 1050 in Fig. 10B. Various circles and triangles in the graph 1050 represent different initial points over a wide distribution where these points converges to a threshold voltage V_t . Even if the BE-SONOS device is initially charged to various V_t , the reset operation can tighten these initial points to the reset/erase state. A typical reset time is around 100 msec. In one example, the n-channel BE-SONOS device with ONONO = 15/20/18/70/90 Angstrom, and a N+-poly gate. $L_g/W = 0.22/0.16 \mu m$.

[0052] To phrase in another way, a reset operation is carried out to tighten the V_t distribution before operations. In contrast to a floating gate device where there is no self-converging erase, the BE-SONOS provides a self-converging erase reset/erase methods, which is necessary because the initial V_t distribution is often widely distributed due to the process issues, such as plasma charging effect. The self-converging reset assists in tighten the initial V_t distribution.

[0053] As illustrated in Fig. 11A, there is a circuit diagram illustrating an electrical program of the SONONOS SGIB-AND array architecture 1100 in the second embodiment, while Fig. 11B is a layout diagram 1150 illustrating the electrical program for the SONONOS SGIB-AND array architecture in the second embodiment. In one example during electrical program of a cell A 1110, the wordline WL1 811 is set to 10 volts while the other wordlines WL0 810, WL2 812, and WLm 813 are set 0 volt. The bitlines BL0 820, BL1 821, and BL3 823 are left floating. The bitline BL2 is set to 0 volt, and the bitline BL4 is set to 5 volts. The sub-gates SG0 830 and SG4 834 are set to

8 volts. The sub-gate SG1 is set to 0 volt, the sub-gate SG2 is set to 5 volts, and the sub-gate SG3 is set to 1 volt.

[0054] The bitlines BL0 820, BL1 821, BL2 822, BL3 823, and BL4 824 provide a greater flexibility in programming than sub-gates SG0 830, SG1 831, SG2 832, SG3 833 and SG4 834 because each bitline can be independently programmed. One type of electrical programming method is a source side injection. The source side injection programs a cell to a high voltage threshold V_t state. For example, the source injection applies $V_g = 10\text{ V}$ to the selected WL1, $V_g = 0\text{ V}$ to other WL's, $SG = 1\text{ V}$ for programming and $SG = 0\text{ V}$ for inhibition. The voltage setting of the SG voltage at 1 volt is intended as an illustration such that in general it is 0.5 to 2 volts higher than the threshold voltage under SG gate.

[0055] To carry out an electrical program, the selected wordline is applied a high voltage, 10 volts, and the sub-gate SG3 833 is applied 1 volt to perform a source side injection to program a target cell. The SG1 831 is set to 0 volt for program inhibit and the SG4 834 is set to 8 volts to provide sufficient overdrive to reduce a bitline resistance. One of skill in the art should recognize that parallel programming methods such as page programming with 2kB cells in parallel can burst the programming throughput to more than 10 MB per second while the cell current consumption can be controlled within 2 mA. To avoid program disturbance to other bitlines, the sub-gate SG2 231 is set to 0 volt and turns off the inhibit cell.

[0056] The electrical programming is to conduct a source side injection to program a cell to a high voltage threshold, V_t , state. For example in the electrical programming of the cell A 1110, the operations apply $V_g = 10\text{ V}$ to the selected WL₁ 811, apply $V_g = 0\text{ V}$ to other wordlines including WL0 810, WL2 812 and WL_m 813, set $SG3 = 1\text{ V}$ for programming, set $SG1 = 0\text{ V}$ for program inhibition, and set $SG2 = 5\text{ V}$ for pass gate. The sub-gate SG4 834 is set to 8 volt to highly turn on the sub-gate SG4 834 so that the inversion layer potential can be raised up to 5 volts. In programming the cell A 1110, the sub-gate SG3 833 is set to 1 volt so that the source side injection occurs. The threshold voltage V_t is raised to above the programming voltage, PV. A program inhibition is provided to a cell-B 1012, a cell-C 1014 and a cell-D 1016.

[0057] Fig. 12A is a circuit diagram 1200 illustrating an electrical erase of the BE-SONOS SGIB-AND array architecture in the second embodiment, while Fig. 12B is a graph diagram 1250 illustrating a desirable self-converging erase property with respect to the second embodiment. The erase operation is similar to the reset operation. During electrical erase, the wordlines WL0 810, WL1 811, WL2 812, and WLn 813 are set to -10 volts, the bitlines BL0 820, BL1 821, BL2 822, BL3 823, BL4 824 and BL5 825 are left floating, and the sub-gates SG0 830, SG1 831, SG2 832, SG3 833, SG4 834 are set to 0 volt. The electrical erase is performed in the unit of a sector or block. The BE-SONOS device produces a desirable self-converging erase property, as shown in a graph diagram 1250 in Fig. 6B. The erase saturation V_t is dependent on V_g . A higher V_g causes a higher saturated V_t . The convergent time is typically around 10 to 100 msec.

[0058] Fig. 13A is a circuit diagram 1300 illustrating a read operation of the SONOS SGIB-AND array architecture in the second embodiment, while Fig. 13B is a layout diagram 1350 illustrating the read operation of the SONOS SGIB-AND array architecture in the second embodiment. In one example during a read operation of the cell A 1110, the wordline WL1 811 is set to 5 volts while the other wordlines WL0 810, WL2 812, and WLn 813 are set 0 volt. The bitlines BL0 820, BL1 821, and BL4 824 are left floating. The bitline BL2 822 is set to 0 volt and the bitline BL3 823 is set to 1 volt. The sub-gates SG0 830, SG1 831, and SG4 834 are set to 0 volts, while SG2 832 and SG3 833 are set to 5 volts. A read operation is performed by applying a gate voltage that is between an erased state V_t (EV) and a programmed state V_t (PV). The gate voltage is typically around 5 volts. Alternatively, the gate voltage can be selected to be more than 5 volts or less than 5 volts, provided the gate voltage falls in the range of a high V_t value and a low V_t value. If V_t of the cell-A 422 is higher than 5 volts, then the read current is likely to be a very small value (e.g., $<0.1 \mu A$). If V_t of the cell-A 422 is less than 5 volts, the read current is likely to be a high value (e.g., $>0.1 \mu A$). The state of a memory can then be identified.

[0059] The applied voltage at a bitline (BL) is typically around 1 volt. A larger read voltage will induce more current, but the read disturbance may be larger. The WL number of SG-AND string is typically 64, 128, or 256. A larger number of SG-AND string may save more overhead and increase the array efficiency. However, the

program distribution may be larger. A trade-off is weighed in choosing an adequate number of SGIB-AND string.

[0060] Although the above read function describes a random access read operation, one of ordinary skill in the art should recognize that a page read of multiple cells are possible without departing from the spirits of the present invention. The invention has been described with reference to specific exemplary embodiments. Various modifications, adaptations, and changes may be made without departing from the spirit and scope of the invention. Accordingly, the specification and drawings are to be regarded as illustrative of the principles of this invention rather than restrictive, the invention is defined by the following appended claims.

CLAIMS

We Claim:

1. An integrated circuit device comprising:
a semiconductor substrate;
a plurality of memory cells on the semiconductor substrate, each memory cell having a spacer dielectric layer disposed between a gate and a sub-gate, each gate overlaying a blocking oxide-charge storage layer-modulated tunnel dielectric stack, each sub-gate overlaying a gate oxide; and
an N+ buried diffusion disposed in the semiconductor substrate and positioned underneath between a first gate oxide and a first blocking oxide-charge storage layer-modulated tunnel dielectric stack that serves as a first diffusion bitline.
2. The integrated circuit of claim 1 wherein the modulated tunnel dielectric stack comprises an oxide-nitride-oxide (O₂-N₁-O₁) stack.
3. The integrated circuit of claim 2 wherein the O₂-N₁-O₁ stack comprises a substantially thin oxide and nitride layers.
4. An integrated circuit device comprising:
a semiconductor substrate; and
a plurality of memory cells on the semiconductor substrate, each memory cell having a spacer oxide disposed between a gate and a sub-gate, each gate overlaying a blocking oxide-charge storage layer-modulated tunnel dielectric stack, each sub-gate overlaying a gate oxide;
wherein a first sub-gate serving as source side injection during programming when the first sub-gate is a slightly turn-on state; and
wherein the first sub-gate serving as an inversion layer for connection to a first metal bitline when the first sub-gate is an ON state.

5. The integrated circuit of claim 4 wherein the modulated tunnel dielectric stack comprises an oxide-nitride-oxide ($O_2-N_1-O_1$) stack.

6. The integrated circuit of claim 5 wherein the $O_2-N_1-O_1$ stack comprises a substantially thin oxide and nitride layers.

7. A SG-AND array architecture comprising:
a memory array having a plurality of columns of SONONOS devices that are connected in parallel;
a plurality of sub-gate lines, each sub-gate line connecting to a corresponding column of SONONOS devices; and
a plurality of diffusion bitlines, each diffusion bitline connecting to a corresponding column of SONONOS devices.

8. The SG-AND array architecture of claim 7 wherein the plurality of sub-gate lines are used for source side injection during programming.

9. The SG-AND array architecture of claim 7 wherein the odd number of sub-gate lines are commonly electrically connected.

10. The SG-AND array architecture of claim 7 wherein the even number of sub-gate lines are commonly electrically connected

11. A SGIB-AND array structure comprising:
a memory array having a plurality of columns of SONONOS devices that are connected in parallel structured, the memory array having no diffusion bitlines; and
a plurality of sub-gate lines, each sub-gate line connecting to a corresponding column of SONONOS devices.

12. The SGIB-And array architecture of claim 11 wherein the plurality of sub-gate lines are used for source side injection during programming when the plurality of sub-gate lines are in an off state.

13. The SGIB-And array architecture of claim 11 wherein each of the plurality of sub-gate lines functions as inversion layers for connection to a respective metal bitline when the plurality of sub-gate lines are in an ON state.

14. The SGIB-AND array architecture of claim 11 wherein the plurality of sub-gate lines in which every Nth sub-gate lines are commonly electrically connected.

15. The SGIB-AND array architecture of claim 11 wherein the plurality of sub-gate lines in which every 4th sub-gate lines are commonly electrically connected.

16. A memory cell structure comprising:
a spacer oxide having a first sidewall and a second sidewall;
a sub-gate extending horizontally to the first sidewall of the spacer oxide;
an oxide-nitride-oxide ($O_1-N_1-O_2$) stack extending horizontally to the second sidewall of the spacer oxide; and
a control gate overlaying a third oxide layer O_3 and extending horizontally to the second side wall of the spacer oxide.

17. The memory structure of claim 16 further comprising a gate oxide underlying the sub-gate and extending horizontally to the first sidewall of the spacer oxide.

18. The memory structure of claim 17 further comprising a second nitride layer N_2 disposed between the control gate and the second oxide layer O_2 , the second nitride layer N_2 extending horizontally to the second sidewall of the spacer oxide, the second nitride layer N_2 serving as a charge storage layer.

19. The memory structure of claim 18 further comprising a third oxide layer O₃ disposed between the control gate and the second nitride layer N₂, the third oxide layer extending horizontally to the second side wall of the spacer oxide, the third oxide layer O₃ serving as a blocking oxide.

20. The memory structure of claim 19 wherein the O₁-N₁-O₂ stack comprises substantially thin layers of oxide and nitride.

21. The memory structure of claim 20 wherein the each of the first oxide layer O₁, the first nitride layer N₁, and the second oxide layer O₂ is less than 3 nm.

22. A method for operating an AND memory array having columns of bandgap engineered SONOS (BE-SONOS) devices, each column of BE-SONOS devices corresponding to a sub-gate line and a wordline, comprising:

resetting a plurality of BE-SONOS devices in the AND memory array, the plurality of BE-SONOS devices producing a self-converging reset to reset a threshold voltage V_t value;

programming electrically a selected BE-SONOS device in a first column of the BE-SONOS devices, the first column of the BE-SONOS devices corresponding to a first wordline which is applied a high voltage, the first column of the BE-SONOS devices corresponding to a first sub-gate line which is applied a low voltage to perform source side injection; and

erasing electrically the plurality of BE-SONOS devices in the AND memory array, the plurality of BE-SONOS devices producing a self-converging erase to the reset voltage threshold V_t value.

23. The method of claim 22 further comprising reading the selected BE-SONOS device, wherein the voltage of the first wordline is raised to a voltage that is between an erases state level and a program state level.

24. The method of claim 22 wherein the first sub-gate line serving as source side injection during programming when the first sub-gate line is a slightly turn-on state.

25. The method of claim 22 wherein the first sub-gate line serving as an inversion layer for connection to a first metal bitline when the first sub-gate line is an ON state.

26. The method of claim 22 wherein the low voltage applied to the first sub-gate line is 1-2 volt.

ABSTRACT

[0061] A bandgap engineered SONOS device structure for design with various AND architectures to perform a source side injection programming method. The BE-SONOS device structure comprises a spacer oxide disposed between a control gate overlaying an oxide-nitride-oxide-nitride-oxide stack and a sub-gate overlaying a gate oxide. In a first embodiment, a BE-SONOS sub-gate-AND array architecture is constructed multiple columns of SONONOS devices with sub-gate lines and diffusion bitlines. In a second embodiment, a BE-SONOS sub-gate-inversion-bitline-AND architecture is constructed multiple columns of SONONOS devices with sub-gate inversion bitlines and with no diffusion bitlines.

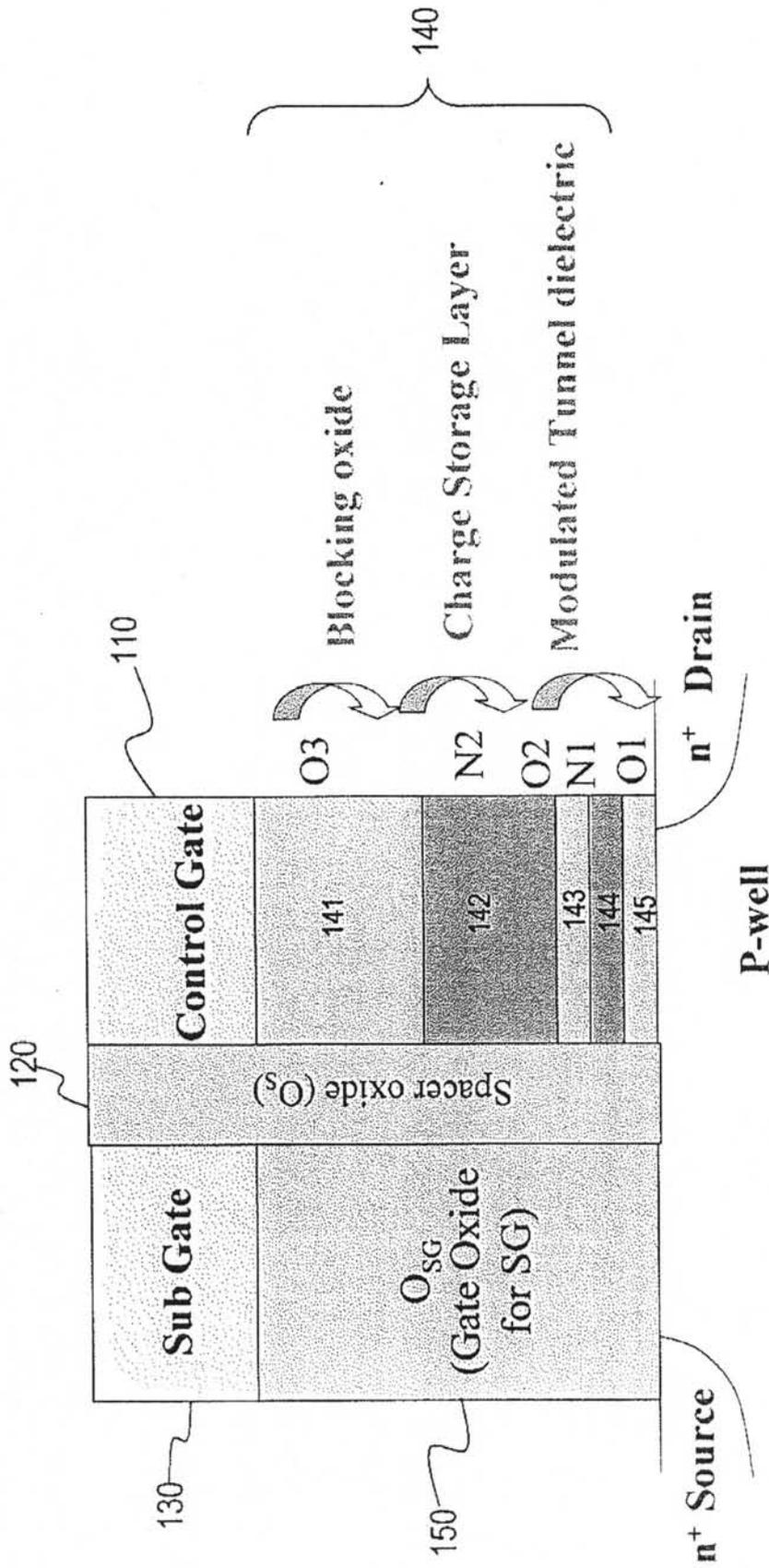


Fig. 1

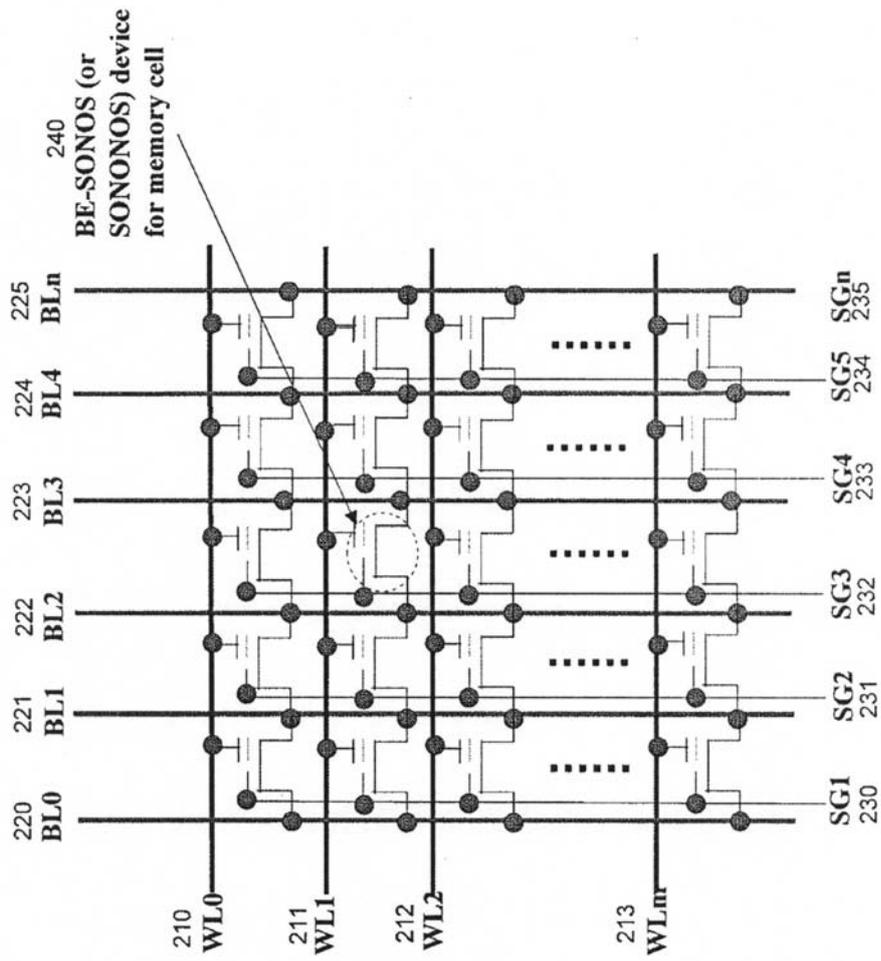


Fig. 2A

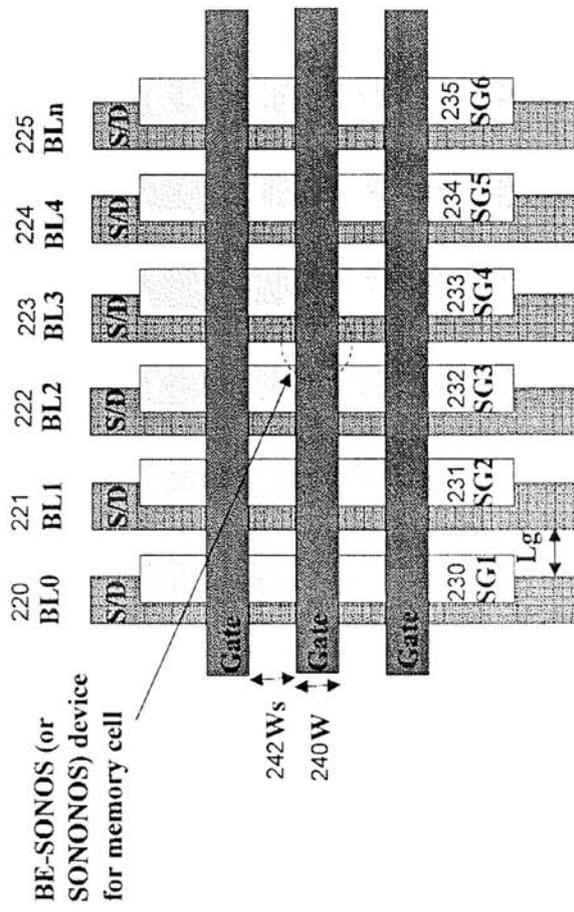


Fig. 2B

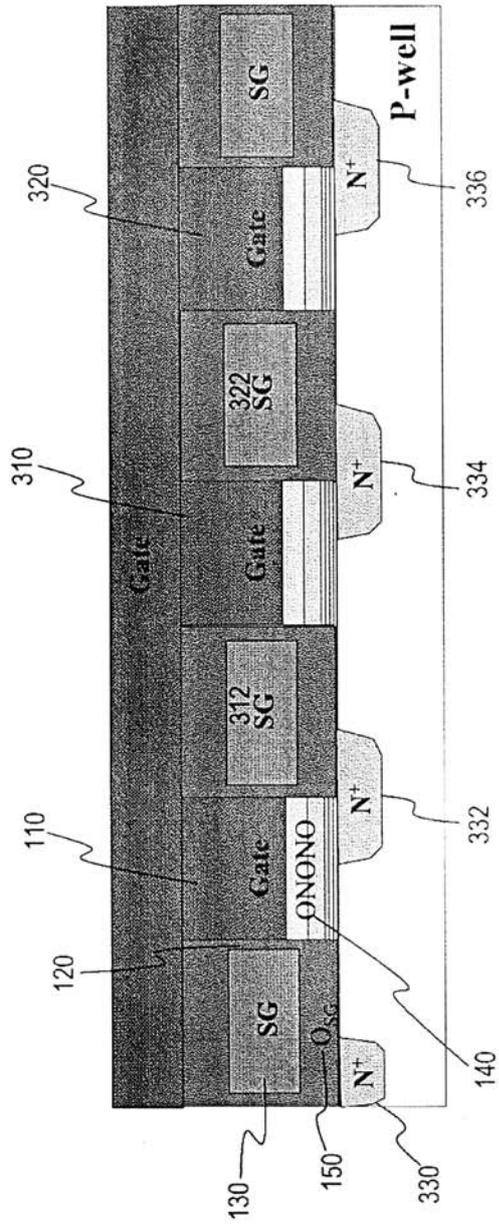
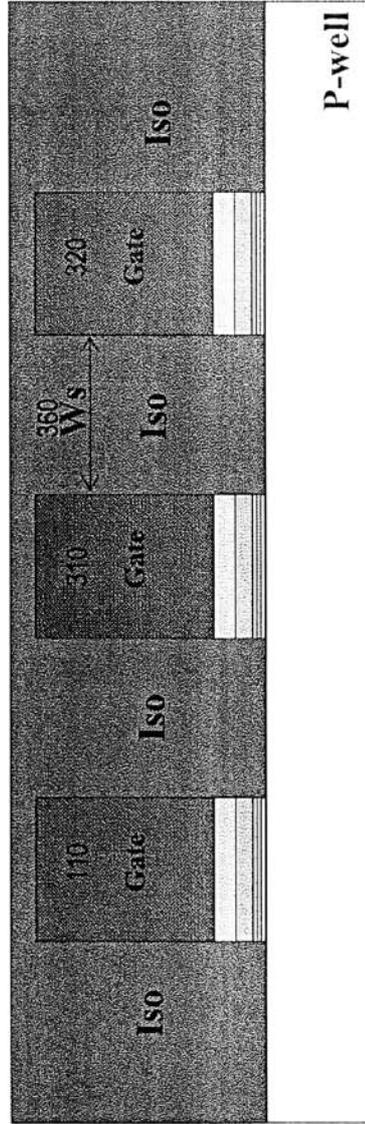


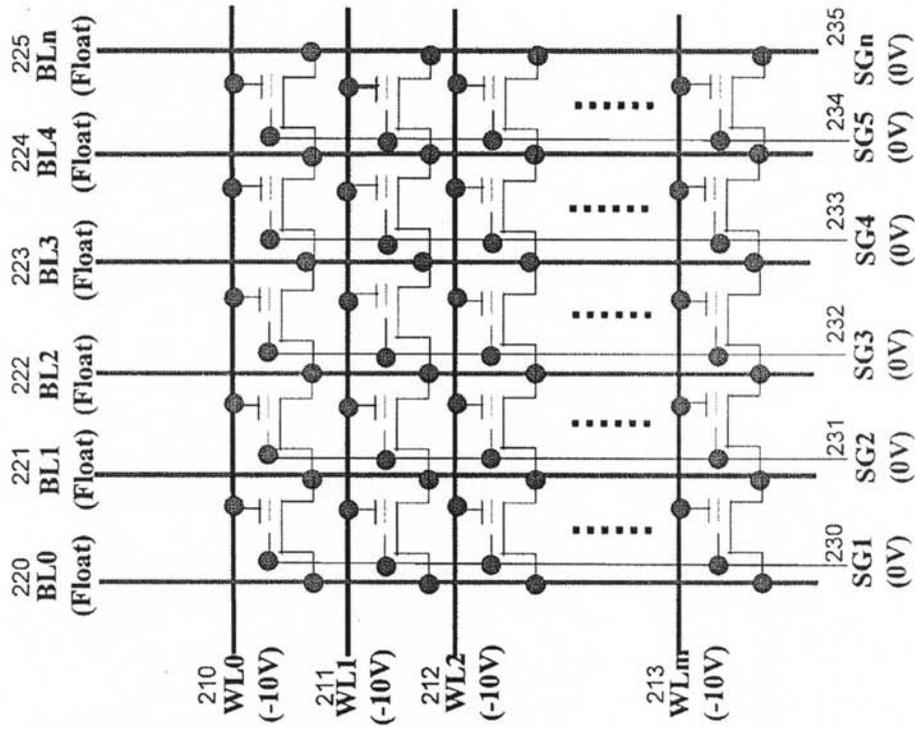
Fig. 3A

300



350

Fig. 3B



P-well= 5V

Fig. 4A

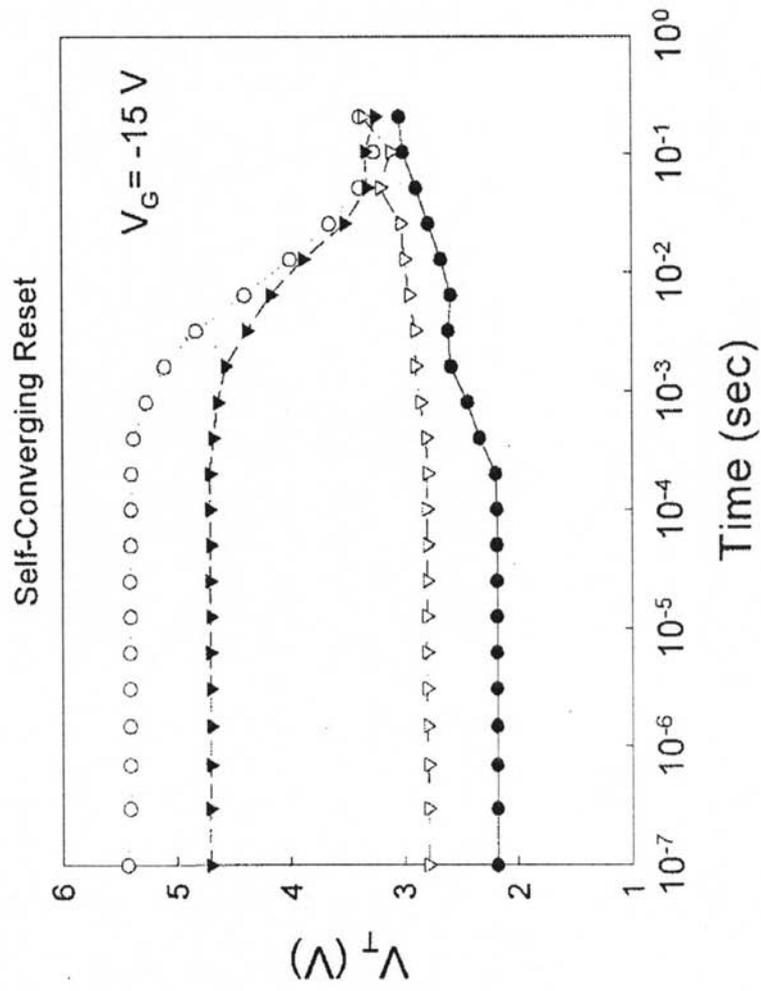


Fig. 4B

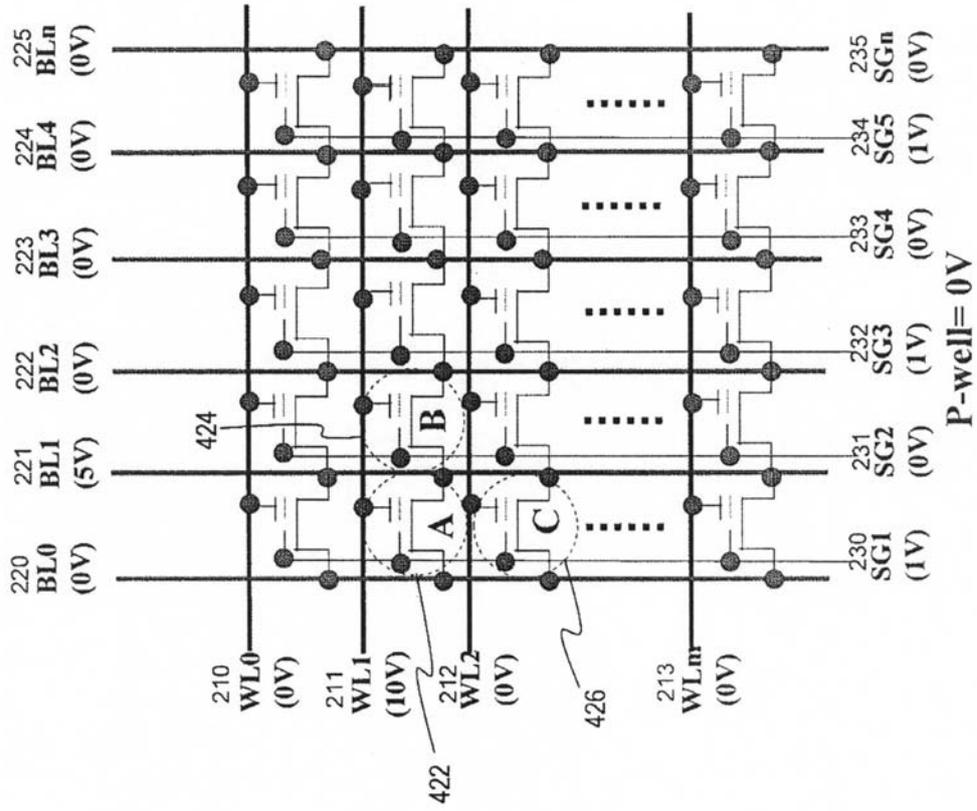


Fig. 5A

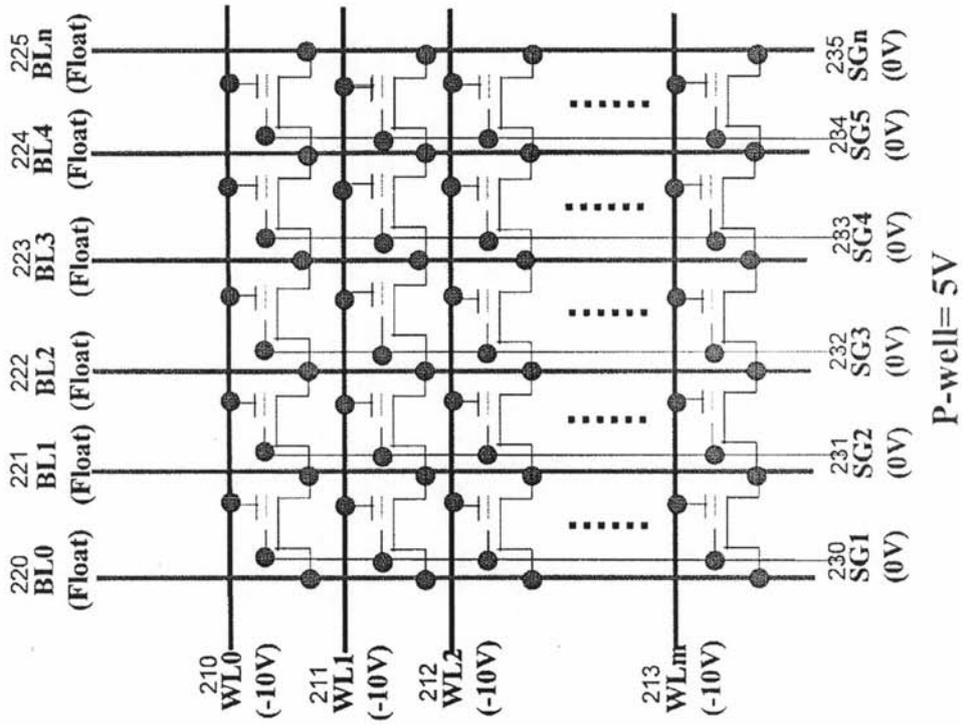


Fig. 6A

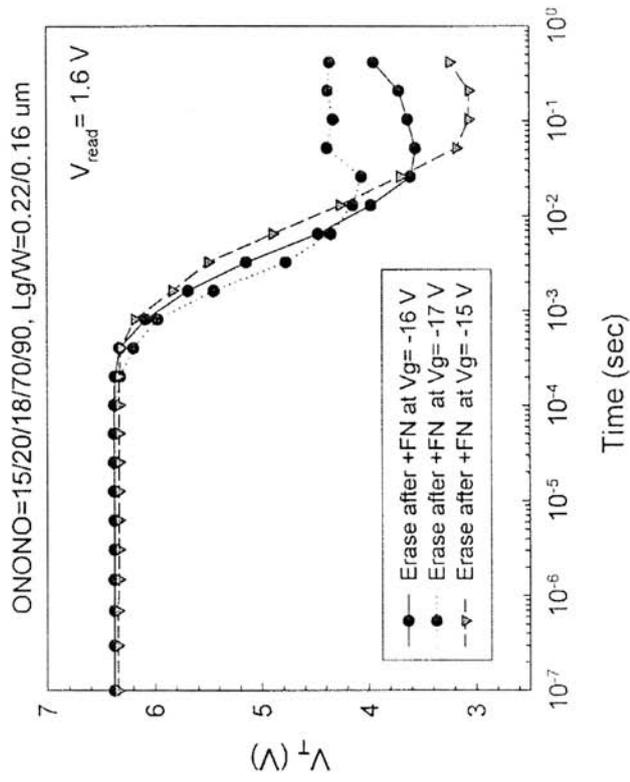


Fig. 6B

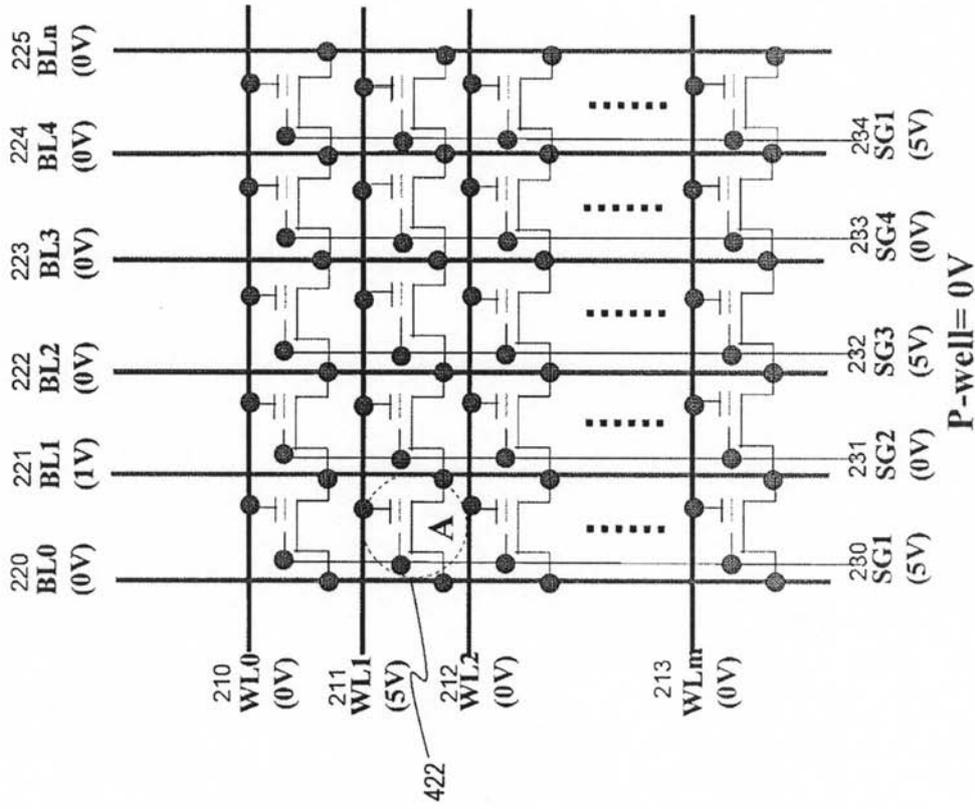


Fig. 7

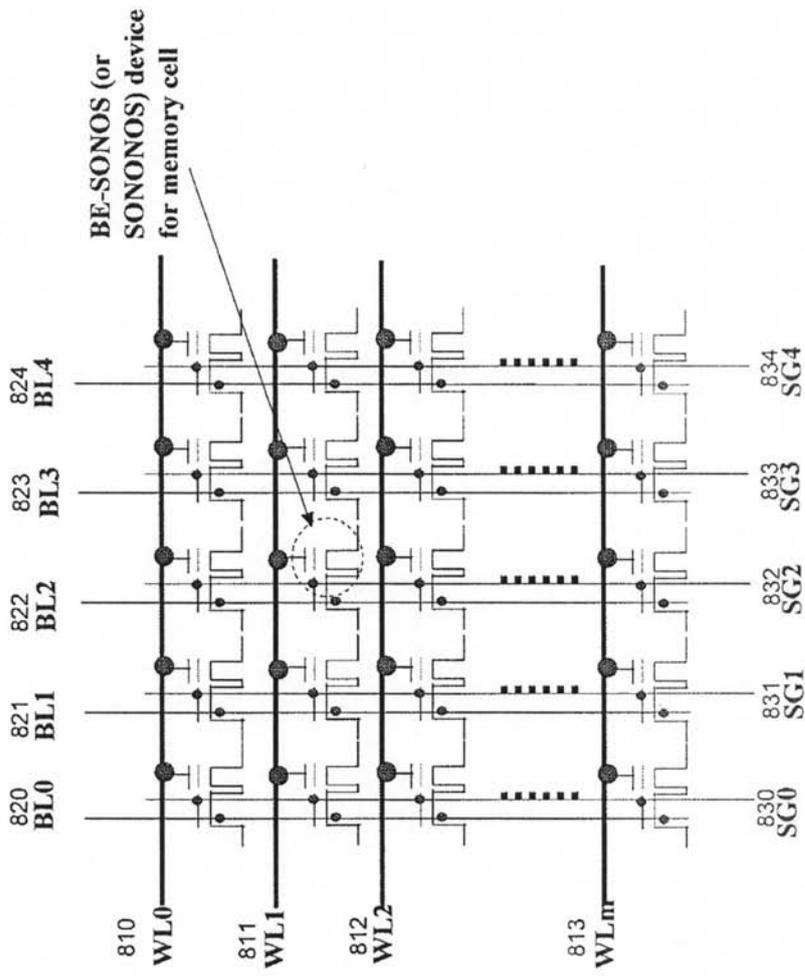
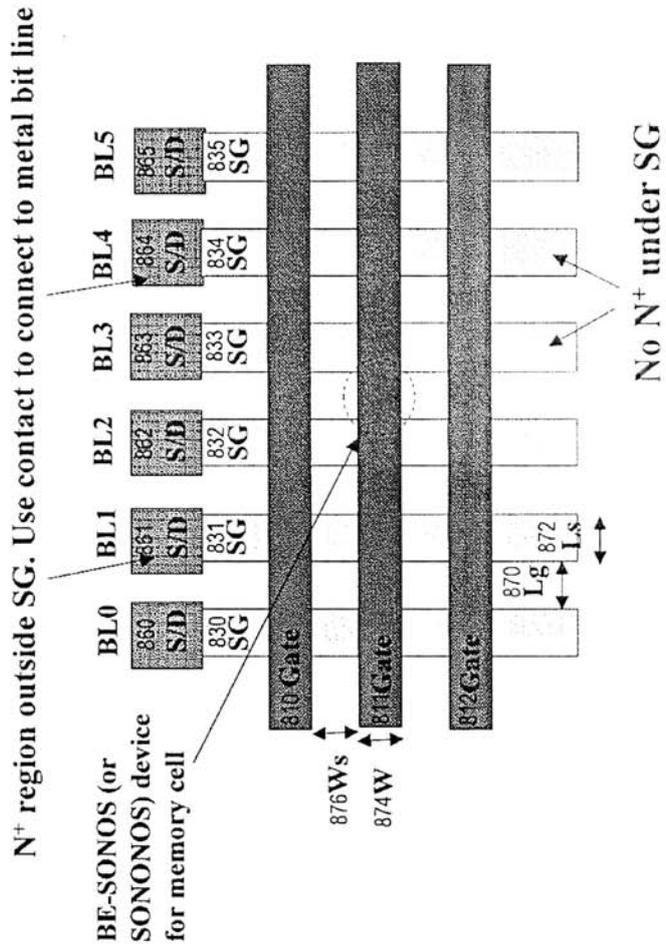
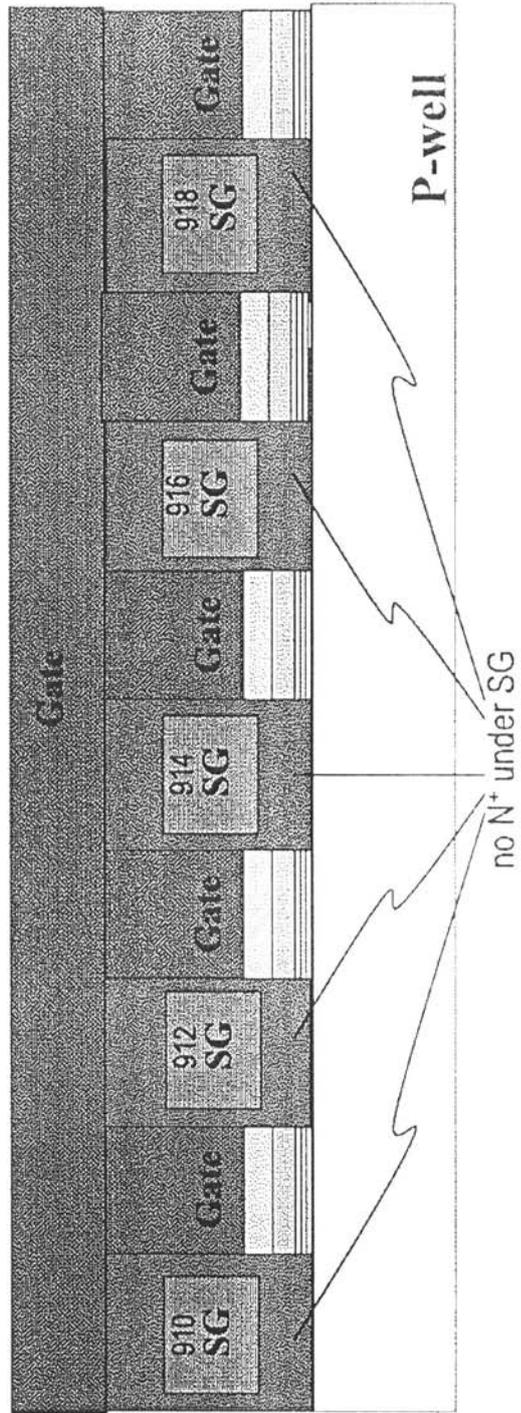


Fig. 8A



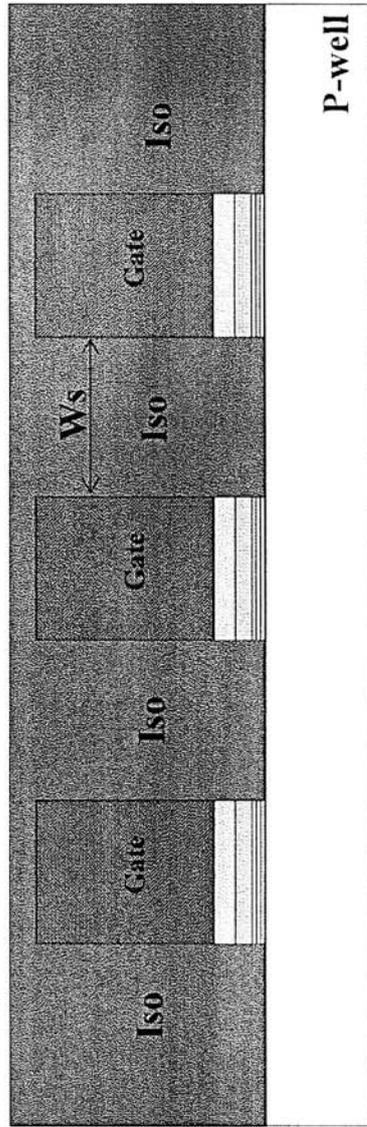
850

Fig. 8B



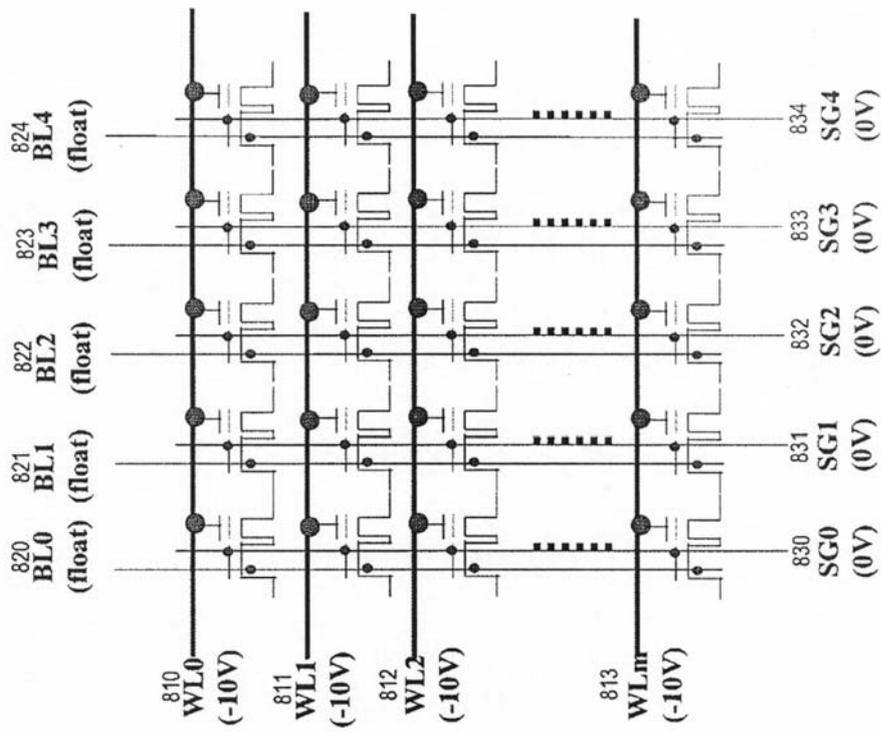
900

Fig. 9A



950

Fig. 9B



1000

¹⁰¹⁰
P-well= 5V
Fig. 10A

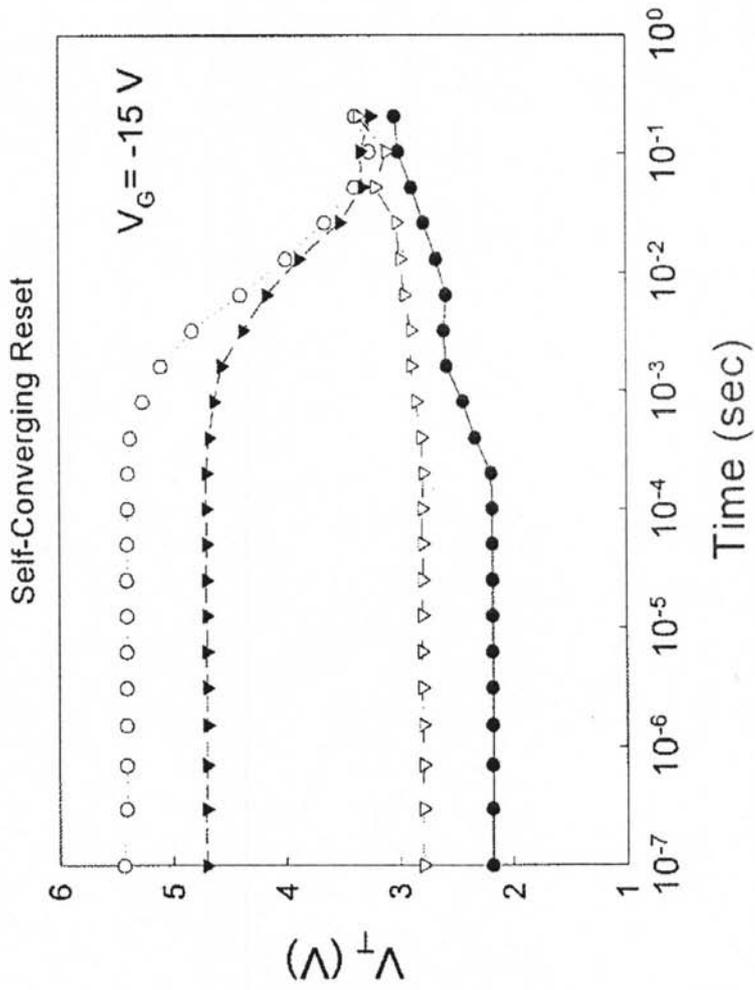
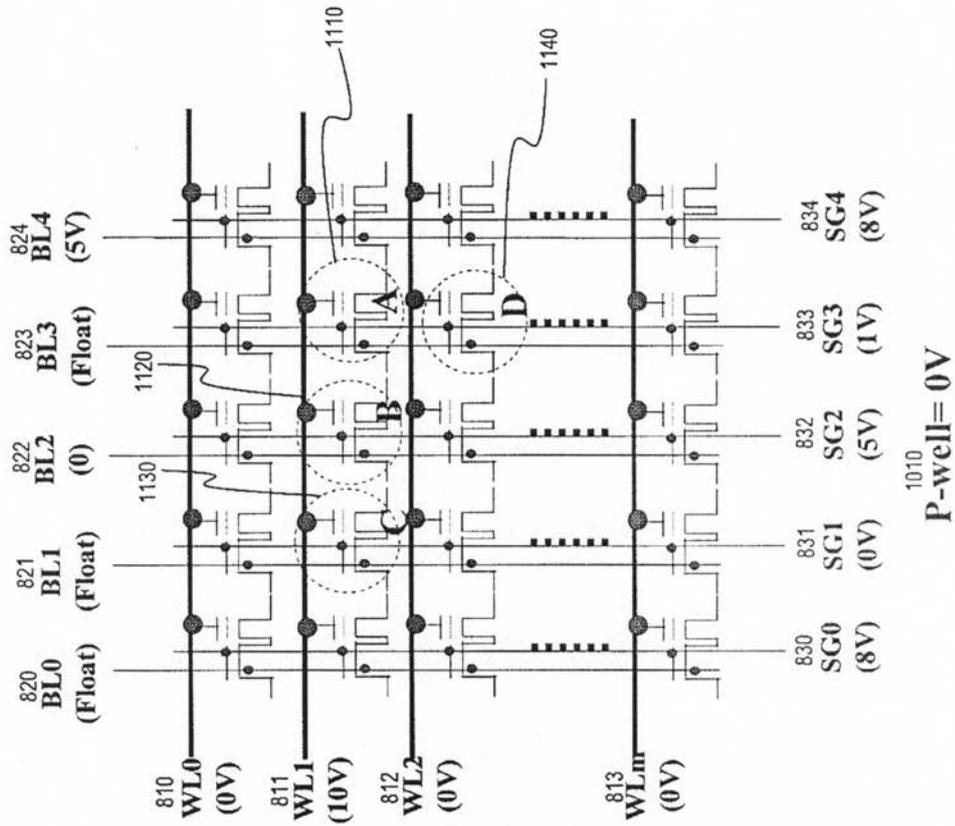


Fig. 10B



1100

Fig. 11A

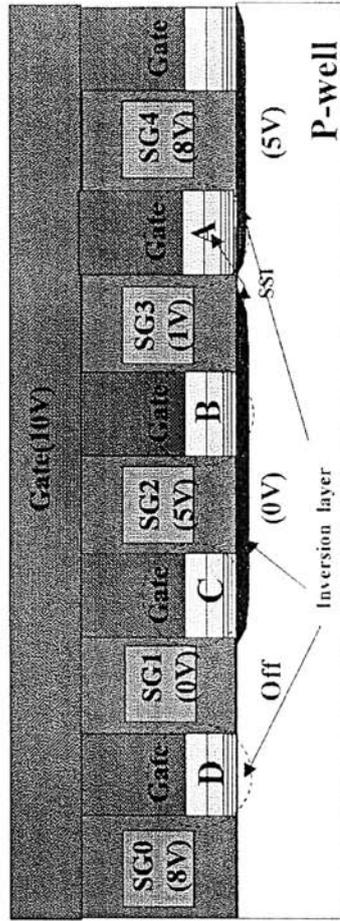
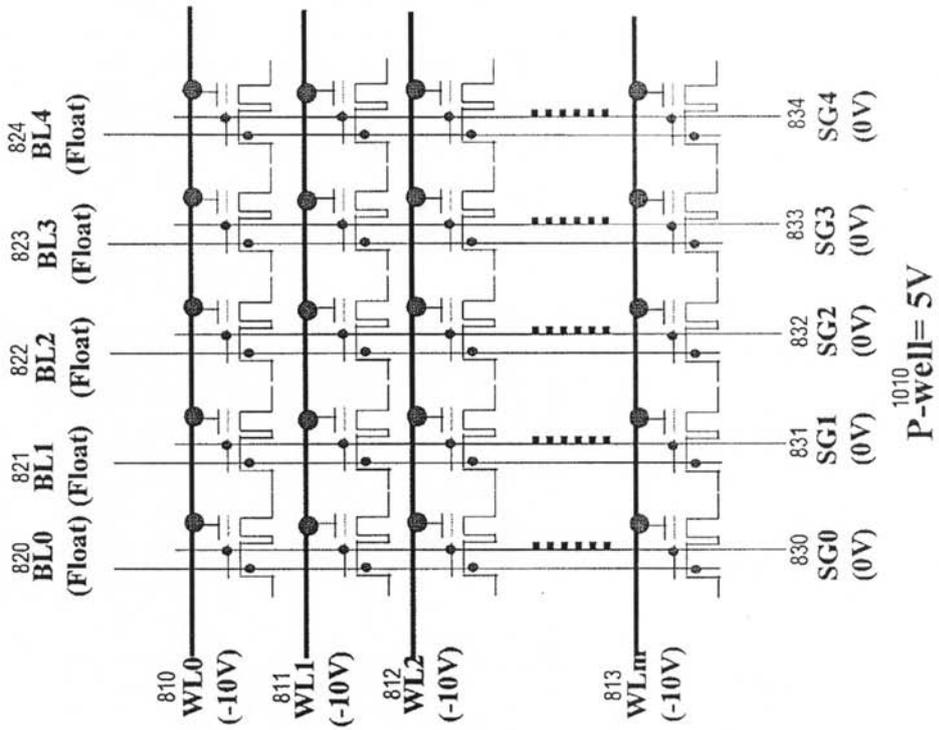


Fig. 11B



1200

Fig. 12A

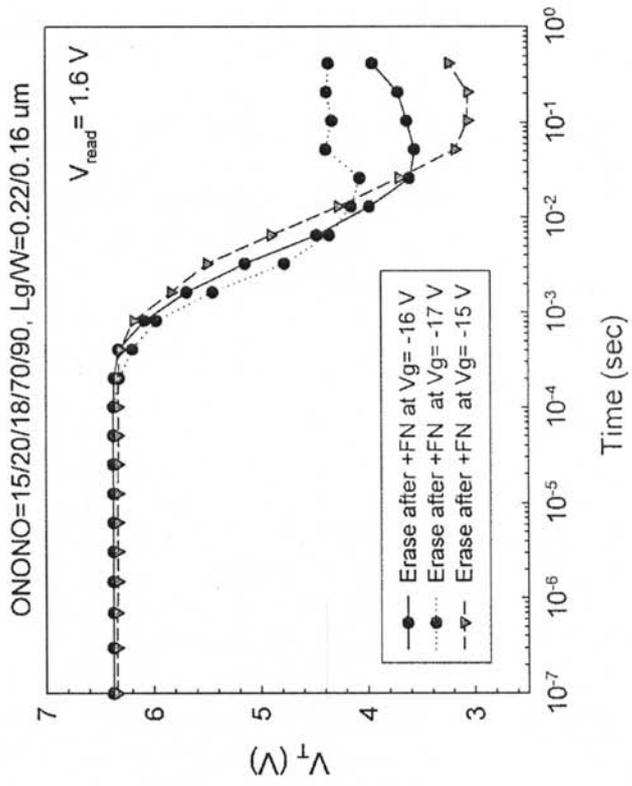
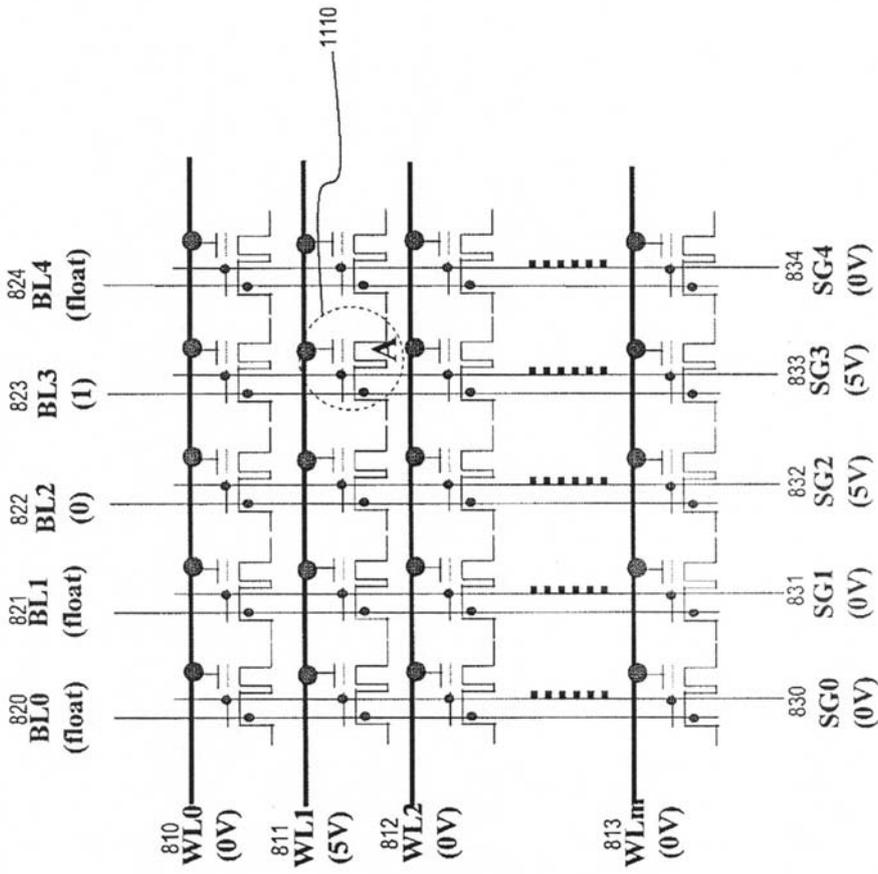


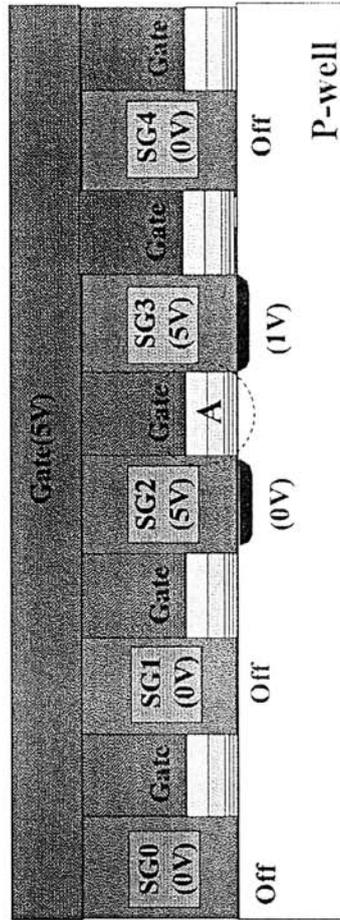
Fig. 12B



1300

¹⁰¹⁰
P-well=0V

Fig. 13A



1350

Fig. 13B