



(19)  
Bundesrepublik Deutschland  
Deutsches Patent- und Markenamt

(10) **DE 694 32 568 T2** 2004.02.26

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 0 740 854 B1**

(21) Deutsches Aktenzeichen: **694 32 568.6**

(86) PCT-Aktenzeichen: **PCT/US94/00407**

(96) Europäisches Aktenzeichen: **94 906 603.9**

(87) PCT-Veröffentlichungs-Nr.: **WO 95/019047**

(86) PCT-Anmeldetag: **11.01.1994**

(87) Veröffentlichungstag

der PCT-Anmeldung: **13.07.1995**

(97) Erstveröffentlichung durch das EPA: **06.11.1996**

(97) Veröffentlichungstag

der Patenterteilung beim EPA: **23.04.2003**

(47) Veröffentlichungstag im Patentblatt: **26.02.2004**

(51) Int Cl.7: **H01L 29/68**

**H01L 29/78**

(73) Patentinhaber:

**Hyundai Electronics Industries Co., Ltd.,  
Seoul/Soul, KR**

(84) Benannte Vertragsstaaten:

**AT, CH, DE, DK, ES, FR, GB, IT, LI, NL, SE**

(74) Vertreter:

**Sparing · Röhl · Henseler, 40237 Düsseldorf**

(72) Erfinder:

**MA, Y., Yueh, Los Altos, US; CHANG, Kuo-Tung,  
San Jose, US**

(54) Bezeichnung: **SELBSTJUSTIERENDE FLASH-EEPROM-ZELLE MIT DOPPELBIT-GETEILTEM GAT**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

## Beschreibung

[0001] Diese Erfindung bezieht sich auf eine Dualbit-Flash-EEPROM-Zellenstruktur gemäß dem Oberbegriff des Anspruchs 1, eine Flash-EEPROM-Matrix, die mehrere derartige EEPROM-Zellenstrukturen umfaßt, und ein Verfahren für die Herstellung einer Flash-EEPROM-Struktur.

[0002] Eine Dreifach-Polysilicium-Flash-EEPROM-Zelle für nur 5 Volt mit einer geteilten Gate-Struktur ist in Naruke, u. a., "A New Flash-Erase EEPROM Cell With a Sidewall Select-Gate on Its Source Side", Technical Digest of IEEE Electron Device Meeting, 1988, offenbart. Wie darin offenbart ist, wird ein doppelter Polysilicium-Transistor mit einem Floating- oder schwebenden Gate durch erste und zweite Polysiliciumlagen unter Verwendung einer herkömmlichen Verarbeitung gebildet, wobei dann ein Auswahlgate-Transistor nach dem Abscheiden einer dritten Polysiliciumlage unter Verwendung einer Rückätz-Technologie gebildet wird. Die Länge des Auswahlgate-Transistors ist durch die Höhe des doppelten Polysilicium-Transistors mit schwebendem Gate definiert, die etwa 0,4 µm beträgt. Zurückzuführen auf die Rückätz-Technologie muß der Auswahlgate-Transistor zum Steuergate parallel verlaufen.

[0003] **Fig. 1** veranschaulicht die offenbarte Flash-EEPROM-Zelle. Die Flash-EEPROM-Zelle umfaßt einen N+-Source-Bereich **11a** und einen N+/N--Drain-Bereich **11b**, die durch den Kanalbereich **12** getrennt sind. Der Kanalbereich **12** besteht aus einem Abschnitt **12a** unter dem schwebenden Gate **13** und einem Abschnitt **12b** unter dem Auswahlgate **14**. Dem Kanalbereich **12** ist das Gate-Dielektrikum **16a** überlagert, auf dem das schwebende Gate **13** und das Auswahlgate **14** gebildet sind. Dem schwebenden Gate **13** ist eine Isolierung **16b** überlagert, typischerweise eine Zusammensetzungslage aus thermisch aufgewachsenem Oxid, abgeschiedenem Siliciumnitrid und thermisch aufgewachsenem Oxid und Nitrid. Ein Steuergate **17** ist oben auf der Isolierung **16b** gebildet. Typischerweise sind sowohl das schwebende Gate **13** als auch das Steuergate **17** aus Polysilicium gebildet. Eine Isolierung **16c** ist auf der Seitenwand des schwebenden Gates **13** und des Steuergates **17** gebildet. Das Auswahlgate **14** wird gebildet, indem die dritte Lage aus Polysilicium abgeschieden wird und dann das Polysilicium rückgeätzt wird, um einen Polysilicium-Abstandhalter zu bilden. Die Programmierung (die hohe Schwellenspannung) der Zelle wird erreicht, indem das Steuergate auf etwa 17 Volt, das Auswahlgate auf ein wenig über die Schwellenspannung des Auswahlgate-Transistors, die etwa 1,5 Volt beträgt, und der Drain auf 5 Volt angehoben werden und die Source auf Masse gelegt wird. Die Kanalelektronen werden durch einen Potentialabfall zwischen dem Auswahlgate und den Transistoren mit schwebendem Gate beschleunigt. Es ist bekannt, daß der Wirkungsgrad der Injektion der heißen Elektronen unter Verwendung dieses Verfahrens tausendmal höher als beim herkömmlichen seitlichen Beschleunigungsverfahren sein kann. Zurückzuführen auf den hohen Programmierungswirkungsgrad kann die flash-Zelle mit einer niedrigeren Drain-Spannung (5 Volt) und mit einem sehr niedrigen Programmierungsstrom (wenige Mikroampere) programmiert werden. Das Löschen der Zelle wird ausgeführt, indem der Drain-Bereich **11b** auf 14 Volt angehoben, das Steuergate auf Masse gelegt und die Source geöffnet wird (schwebende Source). Die hohe Löschespannung kann erhalten werden, indem eine Ladungspumpentechnik von einer 5-Volt-Stromversorgung verwendet wird. Diese Merkmale erlauben den Betrieb mit einer einzelnen 5-Volt-Stromversorgung.

[0004] Einige Hauptnachteile im Flash-EEPROM des Standes der Technik nach **Fig. 1** beziehen sich auf die Bildung des Auswahlgate-Transistors. Zuerst besteht die Zelle mit geteiltem Gate, wie sie im Stand der Technik offenbart ist, aus zwei Transistoren, die das schwebende Gate und das Auswahlgate sind. Es benötigt den Raum von zwei Transistoren, um ein einzelnes Speicherbit zu bilden, wenn es in einer Speichermatrix implementiert ist. Dies ist ein allgemeiner Nachteil für die herkömmliche Struktur mit geteiltem Gate, wobei es die Haupteinschränkung für das Implementieren einer Speichermatrix mit hoher Dichte auferlegt. Zweitens ist das Auswahlgate ein Polysilicium-Abstandhalter, der durch die Polysilicium-Rückätz-Technik gebildet wird, wobei die Länge des Auswahlgates durch die kombinierte Höhe des schwebenden Gates und des Steuergates bestimmt ist, die etwa 0,4 µm beträgt. Bei dieser kleinen Kanallänge ist der Durchschlag dieses Transistors sehr schwierig zu verhindern, insbesondere wenn das schwebende Gate auf eine negative Schwellenspannung überlöscht ist. Drittens ist das Auswahlgate aus Polysilicium gebildet, wobei es sehr schwierig ist, auf einen Polysilicium-Abstandhalter abgeschiedenes Polycid aufzubringen. Folglich kann im Stand der Technik nur Polysilicium mit einem Widerstand von etwa 20 bis 30 Ohm pro Quadrat für das Auswahlgate verwendet werden. Im Ergebnis ist die RC-Verzögerung der Wortleitung einer Speicherschaltung beträchtlich länger als in der herkömmlichen Speicherschaltung, in der das Polycid mit einem Widerstand im Bereich von 2 bis 4 Ohm pro Quadrat verwendet wird. Viertens kann, zurückzuführen auf das Rückätz-Verfahren im Stand der Technik, das Auswahlgate nur parallel zum Steuergate verlaufen. Folglich muß in einer Matrix mit hoher Dichte und mit virtueller Masse das Auswahlgate senkrecht zum Steuergate sein. Folglich kann die flash-EEPROM-Zelle im Stand der Technik nach **Fig. 1** nicht in einer Matrixstruktur mit höherer Dichte und mit virtueller Masse implementiert sein.

[0005] Wie in der noch nicht veröffentlichten US-Patentanmeldung mit der Seriennummer 07/585.811, eingereicht am 20. September 1990, für eine "Triple Polysilicon Flash EPROM Device" beschrieben ist, wird eine herkömmliche Speichermatrix mit virtueller Masse, wie in **Fig. 1B** gezeigt ist, unter Verwendung der Flash-EEPROM-Zelle gebildet, wie sie in der obigen Anmeldung offenbart ist. Die Kanallänge des Auswahlgate-

te-Transistors ist nicht selbstausgerichtet, wobei sie durch die Verwendung eines Photoresist als eine Bitleitungsmaske definiert ist, wie in **Fig. 1C** gezeigt ist. Infolge dieser Fehlausrichtung zwischen der Bitleitungsmaske und dem Stapel aus einem schwebenden Gate und einem Steuergate wird die Kanallänge des nicht selbstausgerichteten Auswahlgates durch die Abmessung 19B vom Photoresist zur Kante des schwebenden Gates und der Fehlausrichtungstoleranz (MA) bestimmt. Typischerweise liegt die Fehlausrichtungstoleranz im Bereich von 0,3 bis 0,5  $\mu\text{m}$ , dies ist die Haupteinschränkung für eine Speichermatrix mit hoher Dichte. Wie in **Fig. 1B** gezeigt ist, ist die gesamte Abmessung pro Speicherbit gleich  $18 + 19A + 19B + MA$ , weil jedes Speicherbit eine Bitleitungs-Diffusion **18**, eine Länge 19A des schwebenden Gates und eine nicht selbstausgerichtete Auswahlgate-Länge  $19B + MA$  benötigt. Es ist ersichtlich, daß in der herkömmlichen Matrix mit virtueller Masse und mit geteiltem Gate, wie in **Fig. 1B** gezeigt ist, jeder Transistor mit schwebendem Gate einen Auswahlgate-Transistor benötigt, in der der Auswahlgate-Transistor nicht auf die Kante des schwebenden Gates selbstausgerichtet ist. Dies zeigt an, daß die herkömmliche virtuelle Masse mit geteiltem Gate ihre Grenze für das Implementieren einer Speichermatrix mit ultrahoher Dichte besitzen kann.

[0006] Aus DE-A-42 33 790 ist eine Dualbit-Flash-EEPROM-Zellenstruktur bekannt, die zwei beabstandete Drain-Bereiche, zwei beabstandete Speichertransistoren mit schwebendem Gate, die zwischen den zwei Drain-Bereichen angeordnet sind, und ein Auswahlgate zwischen den zwei Speichertransistoren mit schwebendem Gate besitzt. Das Auswahlgate ist mit einer Wortleitung verbunden, die sich zwischen den zwei Drain-Bereichen erstreckt, die selbst mit den Bitleitungen verbunden sind, die über die Zellenstruktur in der Richtung vom ersten zum zweiten Drain-Bereich und folglich senkrecht zur Wortleitung verlaufen.

[0007] Aus DE-A-42 33 790 ist ferner ein Verfahren für die Herstellung der Flash-EEPROM-Zellenstruktur bekannt, das die folgenden Schritte umfaßt: Vorsehen eines Halbleitersubstrats mit einem Oberflächenbereich eines ersten Leitfähigkeitstyps; Bilden einer ersten dielektrischen Lage auf dem Oberflächenbereich; Bilden einer ersten Polysiliciumlage auf der ersten dielektrischen Lage; Bilden einer zweiten dielektrischen Lage auf der ersten dotierten Polysiliciumlage; Bilden einer zweiten dotierten Polysiliciumlage auf der zweiten dielektrischen Lage; selektives Ätzen der zweiten dotierten Polysiliciumlage, der zweiten dielektrischen Lage, der ersten dotierten Polysiliciumlage und der ersten dielektrischen Lage, um einen ersten und einen zweiten Stapel aus, Steuergate und schwebendem Gate aus der zweiten dotierten Polysiliciumlage und der ersten dotierten Polysiliciumlage zu bilden, wobei der erste oder der zweite Stapel aus Steuergate und schwebendem Gate voneinander beabstandet sind;

selektives Implantieren von Dotierstoffen eines zweiten Leitfähigkeitstyps, der zu dem ersten Leitfähigkeitstyp entgegengesetzt ist, in den Oberflächenbereich unter Verwendung eines Photoresists und des ersten und des zweiten Stapels aus Steuergate und schwebendem Gate als Maske, wobei der Dotierstoff erste und zweite Drains für erste und zweite Transistoren mit schwebendem Gate bildet, die dazwischen angeordnet sind; Bilden einer dritten dielektrischen Lage über den Steuergates und auf dem Substrat zwischen dem ersten und dem zweiten Stapel aus Steuergate und schwebendem Gate,

Bilden einer dritten dotierten Polysiliciumlage auf der dritten dielektrischen Lage und

selektives Ätzen der dritten dotierten Polysiliciumlage, um eine Wortleitung zur bilden, die sich teilweise über die ersten und zweiten Stapel aus Steuergate und schwebendem Gate und zwischen den ersten und zweiten Stapeln aus Steuergate und schwebendem Gate erstreckt.

[0008] Aus EP-A-0 552 531 ist eine EEPROM-Zellenstruktur, bekannt, in der eine einzelne Source-Diffusion durch zwei Spalten aus Transistoren gemeinsam genutzt wird.

[0009] Einen Gegenstand der Erfindung bildet eine Dualbit-Flash-EEPROM-Zellenstruktur mit geteiltem Gate, die einen Auswahlgate-Transistor und zwei Transistoren mit schwebendem Gate verwendet, um zwei Speicherbits in einer Zelle zu bilden.

[0010] Einen weiteren Gegenstand der Erfindung bildet ein Verfahren; um eine völlig selbstausgerichtete Flash-EEPROM-Zellenstruktur mit geteiltem Gate unter Verwendung des selbstausgerichteten Ätzschritts, um die Kanallänge des Auswahlgate-Transistors genau zu definieren, herzustellen.

[0011] Einen nochmals weiteren Gegenstand der Erfindung bildet eine Dualbit-Flash-EEPROM-Matrix mit hoher Dichte, die eine Dualbit-Flash-EEPROM-Zelle mit geteiltem Gate verwendet.

[0012] Es ist ein Merkmal der Erfindung, daß in der Dualbit-Flash-EEPROM-Zelle mit geteiltem Gate ein einzelner Auswahlgate-Transistor durch zwei benachbarte Transistoren mit schwebendem Gate gemeinsam genutzt wird, d. h., ein einzelner Auswahlgate-Transistor wird für zwei Speicherbits verwendet.

[0013] Es ist ein weiteres Merkmal der Erfindung, daß in der Dualbit-Flash-EEPROM-Zelle mit geteiltem Gate der Auswahlgate-Transistor in der Struktur mit geteiltem Gate völlig selbstausgerichtet ist.

[0014] Es ist ein nochmals weiteres Merkmal, daß in der Dualbit-Flash-EEPROM-Matrix zwei Transistoren mit schwebendem Gate zwischen zwei Bitleitungen angebracht sind.

[0015] Es ist ein nochmals weiteres Merkmal der Erfindung, daß in der Dualbit-Flash-EEPROM-Zelle der Zugriff auf einen der zwei Transistoren mit schwebendem Gate durch das Durchsteuern des Auswahlgate-Transistors und des anderen Transistor mit schwebendem Gate erfolgt.

[0016] Es ist ein nochmals weiteres Merkmal der Erfindung, daß der Auswahltransistor in der Struktur mit ge-

teiltem Gate völlig selbstausgerichtet ist.

[0017] Die vorliegende Erfindung bezieht sich auf eine Flash-EEPROM-Zelle mit hoher Dichte, die durch einen Dreifach-Polysilicium-Prozeß mit einer geteilten Gate-Struktur und vier Anschlüssen (Steuergate, Auswahlgate, Drain und Source) hergestellt wird, in der ein Auswahlgate-Transistor durch zwei Transistoren mit schwebendem Gate gemeinsam genutzt wird. Unähnlich der im Stand der Technik offenbarten geteilten Gate-Struktur und der herkömmlichen Matrix mit virtueller Masse und mit geteiltem Gate, wie in der US-Patentanmeldung 07/585.811 beschrieben ist, in der jeder Transistor mit schwebendem Gate einen Auswahlgate-Transistor benötigt, benötigt die Dualbit-Zelle mit geteiltem Gate nur die Hälfte des Auswahlgate-Transistors für jeden Transistor mit schwebendem Gate. Außerdem ist unähnlich der herkömmlichen Zelle mit geteiltem Gate, in der die Kanallänge des Auswahlgate-Transistors nicht auf die Kante des Transistors mit schwebendem Gate selbstausgerichtet ist, die Kanallänge des Auswahlgate-Transistors in der Dualbit-Zelle mit geteiltem Gate genau definiert und völlig durch die Trennung der zwei Transistoren mit schwebendem Gate selbstausgerichtet, die durch den selbstausgerichteten Polysilicium-Ätzschritt gebildet wird.

[0018] In einer bevorzugten Ausführungsform wird eine 5-Volt-Flash-EEPROM-Zelle mit hoher Dichte und mit virtueller Masse unter Verwendung des Dreifach-Polysilicium-Prozesses mit einer Struktur mit geteiltem Gate und den vier Anschlüssen (Steuergate, Auswahlgate, Drain und Source) für jede Zelle hergestellt. Ein Auswahltransistor wird durch zwei Transistorzellen mit schwebendem Gate gemeinsam genutzt. Die Kanallänge des Auswahltransistors ist völlig auf die Transistoren mit schwebendem Gate ausgerichtet, wobei sie in einem Selbstausrichtungs-Ätzprozeß beim Bilden des Steuergates und des schwebenden Gates für jeden der zwei Transistoren mit schwebendem Gate definiert wird.

[0019] Vorteilhafterweise verläuft das Auswahlgate senkrecht zu den Bitleitungen, wie es in einem Flash-EEPROM mit virtueller Masse erforderlich ist. Das Auswahlgate kann eine Polycid-Lage besitzen, die darauf gebildet ist, um den Widerstand und die RC-Verzögerung der Wortleitung zu verringern.

[0020] Die Erfindung und ihre Aufgaben und Merkmale werden aus der folgenden ausführlichen Beschreibung und den beigefügten Ansprüchen zusammengefasst mit der Zeichnung leichter offensichtlich werden.

[0021] **Fig. 1A** ist eine Schnittansicht einer Dreifach-Polysilicium-EEPROM-Zelle gemäß dem Stand der Technik, wobei die **Fig. 1B** und **1C** Schnittansichten einer Dreifach-Polysilicium-Flash-EPRM-Vorrichtung gemäß der US-Patentanmeldung A7/585.811 sind.

[0022] **Fig. 2A, 2B** und **2C** sind eine Schnittansicht, ein Grundriß und ein elektrisches Schema einer, Dreifach-Polysilicium-EEPROM-Zelle gemäß einer Ausführungsform der vorliegenden Erfindung.

[0023] **Fig. 3** ist ein Grundriß eines Abschnitts einer Speichermatrix, die die EE-PROM-Zellen nach den **Fig. 2A-2C** verwendet.

[0024] **Fig. 4A-4C** sind Schnittansichten, die die Schritte beim Herstellen der Zellenstruktur nach den **Fig. 2A-2C** veranschaulichen.

[0025] **Fig. 5A** ist eine graphische Darstellung der Transistor-Schwellenspannung gegen die Programmierungszeit für die Zwei-Zellen-Struktur der Erfindung.

[0026] **Fig. 5B** ist eine graphische Darstellung der Schwellenspannung gegen die Auswahlgate-Spannung beim Programmieren der Zellenstruktur der Erfindung.

[0027] **Fig. 6** ist eine graphische Darstellung der Steuergate-Spannung gegen den Lesestrom für die Zellenstruktur der Erfindung.

[0028] **Fig. 7** ist eine graphische Darstellung der Transistor-Schwellenspannung gegen die Löschezit beim Löschen einer Zelle der Struktur der Erfindung.

[0029] In der Zeichnung sind die **Fig. 2A, 2B** und **2C** eine Schnittansicht, ein Grundriß und ein Schema einer Dualbit-Flash-EEPROM-Zelle gemäß einer Ausführungsform der Erfindung. Die Zellenstruktur umfaßt einen ersten Speichertransistor **20** mit schwebendem Gate, einen zweiten Speichertransistor **22** mit schwebendem Gate und einen Auswahlgate-Transistor **24**, der die Transistoren **20, 22** in Reihe schaltet. Wie in **Fig. 2C** schematisch veranschaulicht ist, ist eine erste Bitleitung BL1 mit dem Drain des Transistors **20** verbunden, eine zweite Bitleitung BL2 ist mit dem Drain des Transistors **22** verbunden und die Sources der zwei Transistoren sind durch das Auswahlgate **24** in Reihe geschaltet. Wie hierin im folgenden weiter beschrieben ist, werden, wenn der Speichertransistor **20** gelesen wird, die Transistoren **22** und **24** durchgesteuert, wobei der Drain des Transistors **22** praktisch die Source des Transistors **20** wird. Wenn der Transistor **22** gelesen wird, werden umgekehrt die Transistoren **20** und **24** durchgesteuert, wobei der Drain des Transistors **20** praktisch die Source des Transistors **22** wird.

[0030] In der Schnittansicht nach **Fig. 2A** ist die Zellenstruktur in einem P- -dotierten Substrat **26** gebildet, wobei der Drain **20a** des Transistors **20** und der Drain **22a** des Transistors **22** N+/N- -dotierte Bereiche in der Oberfläche des Substrats **26** umfassen. Die Drains sind mit den Drains in gemeinsamen Zeilen der Transistoren miteinander verbunden, wobei sie die Zellen-Bitleitungen bilden. Der Transistor **20** enthält ein schwebendes Gate **20b** und ein überlagertes Steuergate **20c**, wobei das schwebende Gate **20b** von der Oberfläche des Substrats **26** durch ein dünnes (z. B. 100 Ångström) Gate-Oxid **20d** getrennt ist, während das Steuergate **20c**

vom schwebenden Gate **20b** durch ein Dielektrikum **20e** getrennt ist, das Siliciumoxid, Siliciumnitrid oder eine Kombination daraus umfassen kann. Ähnlich enthält der Transistor **22** ein schwebendes Gate **22b**, ein Steuergate **22c**, wobei die Dielektrika **22d** und **22e** das schwebende Gate und das Steuergate voneinander und vom Substrat elektrisch isolieren. Die Dielektrika **20f** und **22f** bilden einen Teil einer dielektrischen Lage, die die Steuergates und die Oberfläche des Substrats **26** überlagert. Eine Wortleitung **28** erstreckt sich über die dielektrische Lage und bildet das Steuergate **24a** des Auswahltransistors zwischen den Transistoren **20**, **22** mit schwebendem Gate. Die Wortleitung **28** schaltet die Gates der Auswahltransistoren in einer Zeile einer Speichermatrix in Reihe, wobei sie senkrecht zu den Bitleitungen verläuft, die an die Drains der Transistoren in den Spalten angeschlossen sind, wie z. B. die Drains **20a** und **22a**, wie im Grundriß eines Abschnitts einer Flash-EEPROM-Matrix veranschaulicht ist, der in **Fig. 3** gezeigt ist.

[0031] **Fig. 2A** zeigt die Schnittansicht der vorliegenden Erfindung, in der eine Dualbit-Flash-EEPROM-Zelle mit geteiltem Gate zwei Speicherbits enthält. Unähnlich der im Stand der Technik offenbarten geteilten Gate-Struktur und der herkömmlichen Matrix mit virtueller Masse und mit geteiltem Gate, wie in der anhängigen Patentanmeldung offenbart ist, in der jeder Transistor mit schwebendem Gate einen Auswahlgate-Transistor benötigt, benötigt die Dualbit-Zelle mit geteiltem Gate nur die Hälfte des Auswahlgate-Transistors für jeden Transistor mit schwebendem Gate. Außerdem ist, unähnlich zur herkömmlichen Zelle mit geteiltem Gate, in der die Kanallänge des Auswahlgate-Transistors nicht auf die Kante des Transistors mit schwebendem Gate selbstausgerichtet ist, die Kanallänge des Auswahlgate-Transistors in der Dualbit-Zelle mit geteiltem Gate genau definiert, wobei sie durch die Trennung der zwei Transistoren mit schwebendem Gate völlig selbstausgerichtet ist, die durch den selbstausgerichteten Polysilicium-Ätzschritt gebildet wird. Wie in **Fig. 2A** gezeigt ist, umfaßt jedes Speicherbit eine Hälfte der Diffusion **20A**, eine Länge **20G** des schwebenden Gates und eine Hälfte der Auswahlgate-Länge **24G**, so daß die Gesamtabmessung eines Speicherbits gleich der Hälfte von  $20A + 20G +$  der Hälfte von  $24G$  ist. Das halbe Auswahlgate für jedes Speicherbit und die Eigenschaft der völligen Selbstausrichtung der Auswahlgate-Kanallänge machen das geteilte Dualbit-Gate für eine Speichermatrix mit ultrahoher Dichte in der Anwendung am skalierbarsten, wie z. B. einem Halbleiter-Plattenlaufwerk und einer IC-Speicherkarte. Außerdem kann im Vergleich zum Stand der Technik das Auswahlgate durch die Verwendung von Polycid implementiert sein, um die RC-Verzögerung der Arbeitsleitung zu verringern.

[0032] Die Herstellung der Zellenstruktur nach den **Fig. 2A–2C** ist in den Schnittansichten nach den **Fig. 4A–4C** veranschaulicht. Gleiche Elemente besitzen das gleiche Bezugszeichen. Zuerst wird ein dünnes Gate-Oxid von etwa 100 Angström thermisch auf die Oberfläche des P-dotierten Halbleitersubstrats **26** aufgewachsen; das einen spezifischen elektrischen Widerstand von 10–25 Ohm-Zentimeter besitzt. Eine erste Lage des dotierten Polysilicium wird dann oben auf dem dünnen Oxid abgeschieden und in einer Richtung mit einem Muster versehen, um die schwebende Gate-Lage teilweise zu bilden. Eine Isolationslage aus Siliciumoxid, Siliciumnitrid oder einer Kombination daraus wird dann oben auf der schwebenden Gate-Lage gebildet, wobei dann eine zweite dotierte Polysilicium-Lage für die Steuergates abgeschieden und in einer Richtung mit einem Muster versehen wird. Indem so verfahren wird, wird der doppelte Polysilicium-Stapel geätzt, um die Steuergate des **20c**, **22c** und die schwebenden Gates **20b**, **22b**, die selbstausgerichtet sind, wie angezeigt ist, mit der Isolierung dazwischen zu bilden. Ähnliche Strukturen werden für die benachbarten Transistoren gebildet, wobei sie bei **30** und **32** veranschaulicht sind.

[0033] Als nächstes wird eine Photoresist-Lage **34** gebildet und mit einem Muster versehen, wie in **Fig. 4B** veranschaulicht ist, um einen Abschnitt des doppelten Polysilicium-Stapels teilweise abzudecken, wobei ein Ionenimplantationsschritt die Bitleitungs-Bereiche **20a** und **22a** bildet. Durch das Implantieren zweier verschiedener n-Dotierstoffe, wie z. B. Arsen und Phosphor, und dann das Glühen der Struktur ergibt sich das N+/N-Dotierstoffprofil. Dieser Prozeß ist in der vorgenannten US-Patentanmeldung mit der Seriennummer 07/585.811 beschrieben. Der Substratbereich, der durch den Photoresist zwischen dem Stapel aus einer Struktur aus einem Steuergate/einem schwebenden Gate abgedeckt ist, nimmt die Ionenimplantation nicht auf, wobei er verwendet wird, um die Kanallänge des Auswahlgate-Transistors zu definieren.

[0034] Der Photoresist wird abgestreift, wobei eine Siliciumoxid-Isolationslage oben auf den doppelten Polysilicium-Stapeln und auf den Seitenwänden der doppelten Polysilicium-Stapel gebildet wird. Eine dritte Lage aus dotiertem Polysilicium wird dann abgeschieden und mit einem Muster versehen, um eine Wortleitung und das Auswahlgate **24** zu bilden, wie in **Fig. 4C** gezeigt ist. Die Wortleitung ist senkrecht zu den Steuergates, wobei sie längs der Wortleitung zu benachbarten Auswahlgates verläuft. Das Auswahlgate ist durch die Trennung der zwei Transistoren mit schwebendem Gate genau definiert, die durch den selbstausgerichteten Ätzschritt gebildet werden. Die Kanallänge des Auswahltransistors beträgt etwa 0,6 bis 0,8 Mikrometer, wobei sie völlig auf die Transistoren mit schwebendem Gate ausgerichtet ist. Es wird angemerkt, daß der Auswahltransistor-Kanal **24g** zwischen den Kanälen **20g** und **22g** der Transistoren **20** bzw. **22** mit schwebendem Gate positioniert ist, wie in **Fig. 2A** gezeigt ist. Eine Polycid-Beschichtung kann auf dem Auswahlgate und der Wortleitung aufgebracht werden, um die RC-Verzögerung der Wortleitung zu verringern. Außerdem kann, indem das Auswahlgate senkrecht zum Steuergate verläuft, die Flash-EEPROM-Zelle in einer Matrix mit hoher Dichte und mit virtueller Masse implementiert werden.

[0035] Das Programmieren des Transistors **20** mit schwebendem Gate wird ausgeführt, indem die Steuergates **20c** und **22c** auf 12 Volt, das Auswahlgate **24a** auf ein wenig über die Schwellenspannung des Auswahlgate-Transistors (etwa 2 Volt) und die Bitleitung (Drain) **20a** auf 5 Volt angehoben werden und die Bitleitung (Drain) **22a** auf Masse gelegt wird. Die Transistoren mit schwebendem Gate sind, zurückzuführen auf die 12 Volt an den Steuergates, vollständig durchgesteuert, während der Auswahlgate-Transistor ein wenig durchgesteuert ist. In diesem Zustand wird der Kanalstrom während der Programmierung durch den Auswahlgate-Transistor gesteuert, wobei er im Bereich von 10–20 Mikroampere liegt. Die Kanalelektronen werden am Potentialabfall zwischen dem Auswahlgate-Bereich 24g und dem schwebenden Gate-Bereich **20a** beschleunigt, wobei sie durch die Barriere des dünnen Gate-Oxids **20d** durch Heißelektroneninjektion in das schwebende Gate **20b** injiziert werden.

[0036] Eine Leseoperation des Transistors **20** mit schwebendem Gate wird ausgeführt, indem 12 Volt an das Steuergate des Transistor **22**, um den Transistor **22** durchzusteuern, 5 Volt an das Auswahlgate **24** und das Steuergate **20c** und 1–2 Volt an die Bitleitung **20a** angelegt werden und die Bitleitung **22a** auf Masse gelegt wird. Die Sperrvorspannungswirkung vom Transistor **22** mit schwebendem Gate beträgt etwa 0,17 bis 0,24 Volt, wenn seine Schwellenspannung –0,25 bzw. 6,6 beträgt.

[0037] Die Löschoption des Transistors **20** mit schwebendem Gate wird ausgeführt, indem die Bitleitung, **20a** auf 12 Volt angehoben wird, die Steuergates **20c** und **22c** auf Masse gelegt werden und die Bitleitung **22a** geöffnet wird. Die Elektronen werden durch den Fowler-Nordheim-Tunneleffekt aus dem schwebenden Gate **20b** zur Bitleitung **20a** entladen.

[0038] Die Lese-, Programmier- und Löschoptionsbedingungen der Flash-EEPROM-Zelle sind in der folgenden Tabelle zusammengefaßt:

<u>LESEN ZELLE 1</u>	<u>PROGRAMMIEREN ZELLE 1</u>	<u>LÖSCHEN ZELLE 1</u>	
$V_{sg}$	5	1,8 ~ 2	0
$V_{cg1}$	5	12	0
$V_{cg2}$	12	12	0
$V_{BL1}$	1 ~ 2	5	12
$V_{BL2}$	0	0	schwebend

[0039] Es wird angemerkt, daß in der obigen Tabelle angenommen worden ist, daß die Zelle mit einer 5-Volt-Stromversorgung arbeitet. Es ist eine Ladungspumpe erforderlich, um die 12 Volt zu erzeugen, die für das Programmieren, das Lesen und das Löschen der Zelle notwendig sind. Falls jedoch eine 12-Volt-Quelle verfügbar ist, ist die Verwendung einer Ladungspumpe überflüssig, wobei die Spannung an der Bitleitung **1** während der Programmierung auf 6,5 Volt vergrößert werden kann, um die Geschwindigkeit der Programmierung zu vergrößern.

[0040] **Fig. 5A** ist eine graphische Darstellung, die die Änderung der Schwellenspannung während der Programmierung gegen die Programmierungszeit veranschaulicht, wobei angemerkt wird, daß die Schwellenspannung eines Transistors um mehr als 4,5 Volt in weniger als 10 Mikrosekunden geändert werden kann. **Fig. 5B** ist eine graphische Darstellung, die die Wirkung der Auswahlgate-Spannung VSG auf die Programmierung von einem Transistor mit schwebendem Gate unter Verwendung von 10-Mikrosekunden-Impulsen veranschaulicht. Es ist ersichtlich, daß es eine sehr kleine Differenz gibt, wenn der andere Transistor mit schwebendem Gate sich in einem Zustand der hohen oder niedrigen Schwellenspannung befindet.

[0041] **Fig. 6** ist eine graphische Darstellung der Steuergate-Spannung gegen den Lesestrom, wenn sich die Transistoren mit schwebendem Gate in einem Zustand der hohen oder niedrigen Schwellenspannung befinden.

[0042] **Fig. 7** ist eine graphische Darstellung, die die Löschoption gegen die Schwellenspannung für einen Transistor während einer Löschoption veranschaulicht. Die Zelle kann die etwa 100 Millisekunden auf die Null-Schwellenspannung gelöscht werden.

[0043] Es ist eine neuartige Dualbit-Flash-EEPROM-Zellenstruktur beschrieben worden, in der ein Paar Speichertransistoren mit schwebendem Gate einen einzelnen Auswahlgate-Transistor gemeinsam nutzen. Die Wortleitungen, die die Auswahlgates miteinander verbinden, verlaufen senkrecht zu den Steuergate-Leitungen, wobei die Bitleitungen, die die Drains der Transistoren mit schwebendem Gate miteinander verbinden, folglich eine Flash-Matrix mit virtueller Masse erlauben. Ferner können die Wortleitungen eine Silicid-Beschichtung besitzen, wobei folglich der Widerstand und die RC-Verzögerung verringert werden, die den Wortleitungen zugeordnet sind. Unähnlich dem Abstandhalter-Auswahlgate des Standes der Technik ist die Kanallänge

des Auswahlgate-Transistors durch das selbstausgerichtete Ätzen der Transistoren mit schwebendem Gate leicht definiert, wobei folglich die Optimierung der Durchschlags-Eigenschaften des Auswahlgate-Transistors unterstützt wird. Die völlig selbstausgerichtete Struktur des Auswahltransistors macht diese Zelle in den Strukturen mit geteiltem Gate zur skalierbarsten Zelle.

### Patentansprüche

1. Dualbit-Flash-EEPROM-Zellenstruktur, die ein Halbleitersubstrat (**26**) umfaßt, das versehen ist mit einem Oberflächenbereich eines ersten Leitfähigkeitstyps, einem ersten Drain-Bereich (**20A**) und einem zweiten Drain-Bereich (**22A**), die in dem Oberflächenbereich gebildet sind und von einem zu dem einen Leitfähigkeitstyp entgegengesetzten zweiten Leitfähigkeitstyp sind, einem ersten Stapel aus einem schwebenden Gate (**20B**) und einem Steuergate (**20C**) und einem zweiten Stapel aus einem schwebenden Gate (**22B**) und einem Steuergate (**22C**) auf dem Oberflächenbereich zwischen dem ersten Drain-Bereich (**20A**) und dem zweiten Drain-Bereich (**22A**), wobei die ersten und zweiten Stapel aus schwebendem Gate (**20B, 22B**) und Steuergate (**20C, 22C**) beabstandet sind, einem Auswahlgate (**24A**) auf dem Oberflächenbereich zwischen dem ersten und dem zweiten Stapel aus schwebendem Gate (**20B, 22B**) und Steuergate (**20C, 22C**), einer ersten Bitleitung (BL1), die mit dem ersten Drain-Bereich (**20A**) in Kontakt ist, einer zweiten Bitleitung (BL2), die mit dem zweiten Drain-Bereich (**22A**) in Kontakt ist, und einer Wortleitung (**28**), die im allgemeinen senkrecht zu der ersten Bitleitung (BL1) und zu der zweiten Bitleitung (BL2) orientiert ist,

**dadurch gekennzeichnet**, daß

das Auswahlgate (**24A**) ein Abschnitt der Wortleitung (**28**) ist, der sich über die Drain-Bereiche (**20A, 22A**) und die Stapel aus schwebenden und Steuergates (**20B, 20C, 22B, 22C**) erstreckt, die Gates der Auswahltransistoren in einer Zeile einer durch eine solche Zellenstruktur gebildeten Speichermatrix in Reihe schaltet und senkrecht zu der Bitleitung verläuft.

2. Dualbit-Flash-EEPROM-Zellenstruktur nach Anspruch 4, dadurch gekennzeichnet, daß das Auswahlgate (**24A**) von dem ersten und dem zweiten Stapel aus schwebendem Gate (**20B, 22B**) und Steuergate (**20C, 22C**) gemeinsam genutzt wird.

3. Dualbit-Flash-EEPROM-Zellenstruktur nach Anspruch 1 oder 2, dadurch gekennzeichnet, daß die schwebenden Gates (**20B, 22B**) aus einer ersten Polysiliciumlage gebildet sind, die Steuergates (**20C, 22C**) aus einer zweiten Polysiliciumlage gebildet sind und das Auswahlgate (**24A**) aus einer dritten Polysiliciumlage gebildet ist.

4. Dualbit-Flash-EEPROM-Zellenstruktur nach Anspruch 3, dadurch gekennzeichnet, daß die Wortleitung (**28**) eine Polycidlage enthält.

5. Dualbit-Flash-EEPROM-Zellenstruktur nach Anspruch 3, gekennzeichnet durch eine erste dielektrische Lage (**36A**), die auf dem Oberflächenbereich gebildet ist und das erste schwebende Gate (**20B**) sowie das zweite schwebende Gate (**22B**) vom Substrat (**26**) trennt, eine zweite dielektrische Lage (**36B**) auf den schwebenden Gates (**20B, 22B**), die die Steuergates (**20C, 22C**) von den schwebenden Gates (**20B, 22B**) trennt, und eine dritte dielektrische Lage (**36C**) auf den Steuergates (**20C, 22C**) und dem Substrat (**26**) zwischen dem ersten und dem zweiten Stapel aus schwebendem Gate (**20B, 22B**) und Steuergate (**20C, 22C**), wobei die dritte dielektrische Lage (**36C**) die Wortleitung (**28**) und das Auswahlgate (**24A**) vom Substrat (**26**) trennt.

6. Flash-EEPROM-Matrix, die mehrere Dualbit-Flash-EEPROM-Zellenstrukturen nach einem der Ansprüche 3 bis 5, die in Zeilen und Spalten angeordnet sind, umfaßt, wobei mehrere der ersten und zweiten Bitleitungen (BL1, BL2), die die Drains von Transistoren mit schwebendem Gate der Zellenstrukturen miteinander verbinden, vertikal ausgerichtet sind und mehrere der Wortleitungen (**28**), die die Auswahlgates von Zellenstrukturen miteinander verbinden, horizontal ausgerichtet sind.

7. Verfahren für die Herstellung einer Flash-EEPROM-Zellenstruktur, das die folgenden Schritte umfaßt:

- Vorsehen eines Halbleitersubstrats (**26**) mit einem Oberflächenbereich eines ersten Leitfähigkeitstyps (P-),
- Bilden einer ersten dielektrischen Lage (**36A**) auf dem Oberflächenbereich,
- Bilden einer ersten dotierten Polysiliciumlage (**33**) auf der ersten dielektrischen Lage (**36A**),
- selektives Ätzen der ersten dotierten Polysiliciumlage (**33**),
- Bilden einer zweiten dielektrischen Lage (**36B**) auf der ersten dotierten Polysiliciumlage (**33**),

- f) Bilden einer zweiten dotierten Polysiliciumlage auf der zweiten dielektrischen Lage (**36B**),
- g) selektives Ätzen der zweiten dotierten Polysiliciumlage, der zweiten dielektrischen Lage (**36B**), der ersten dotierten, Polysiliciumlage (**33**) und der ersten dielektrischen Lage (**36A**), um einen ersten und einen zweiten Stapel aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**) aus der zweiten dotierten Polysiliciumlage und der ersten dotierten Polysiliciumlage (**33**) zu bilden, wobei der erste und der zweite Stapel aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**) voneinander beabstandet sind,
- h) selektives Implantieren von Dotierstoffen eines zweiten Leitfähigkeitstyps (N+/N-), der zu dem ersten Leitfähigkeitstyp (P-) entgegengesetzt ist, in den Oberflächenbereich unter Verwendung eines Photoresists (**34**) und des ersten und des zweiten Stapels aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**) als Maske, wobei der Dotierstoff erste und zweite Drains (**20A, 22A**) für erste und zweite Transistoren (**20, 22**) mit schwebendem Gate bildet, wobei jeder der Drains (**20A, 22A**) auf einen der Stapel aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**) selbstausgerichtet ist,
- i) Bilden einer dritten dielektrischen Lage (**36C**) über den Steuergates (**20C, 22C**) und auf dem Substrat (**26**) zwischen dem ersten und dem zweiten Stapel aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**),
- j) Bilden einer dritten dotierten Polysiliciumlage auf der dritten dielektrischen Lage (**36C**) und
- k) selektives Ätzen der dritten dotierten Polysiliciumlage, um eine Wortleitung (**28**) zu bilden, die sich über die Drains, die ersten und zweiten Stapel aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**) und zwischen den ersten und zweiten Stapeln aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**) erstreckt, um ein Auswahlgate (**24**) zu bilden.

8. Verfahren nach Anspruch 7, bei dem der Schritt g) den Kanal eines Auswahltransistors (**24**) bildet, der auf den ersten und den zweiten Stapel aus Steuergate (**20C, 22C**) und schwebendem Gate (**20B, 22B**) selbstausgerichtet ist.

9. Verfahren nach Anspruch 7, bei dem der Schritt j) ferner das Bilden einer Polycidlage auf der Wortleitung (**28**) umfaßt.

Es folgen 6 Blatt Zeichnungen

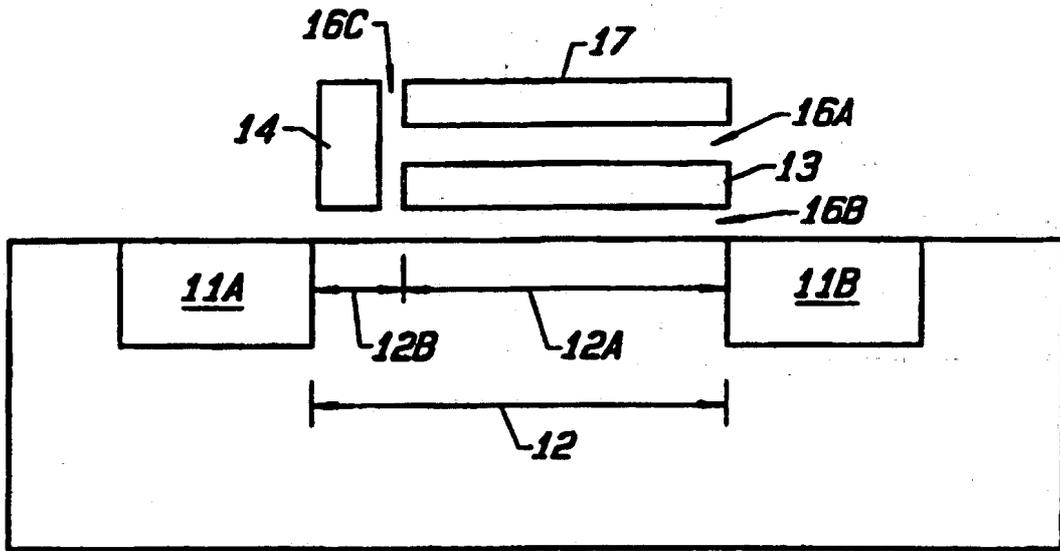


FIG. 1A

Stand der Technik

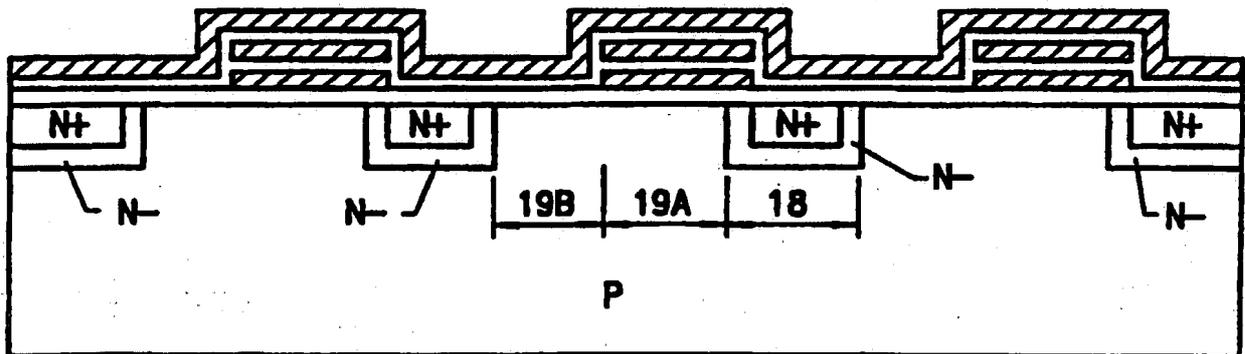


FIG. 1B

Bitleitungs-Resist

Bitleitungs-Resist

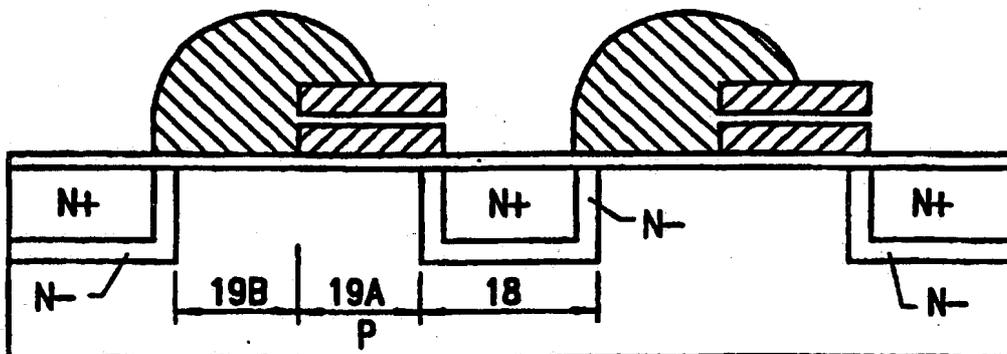


FIG. 1C

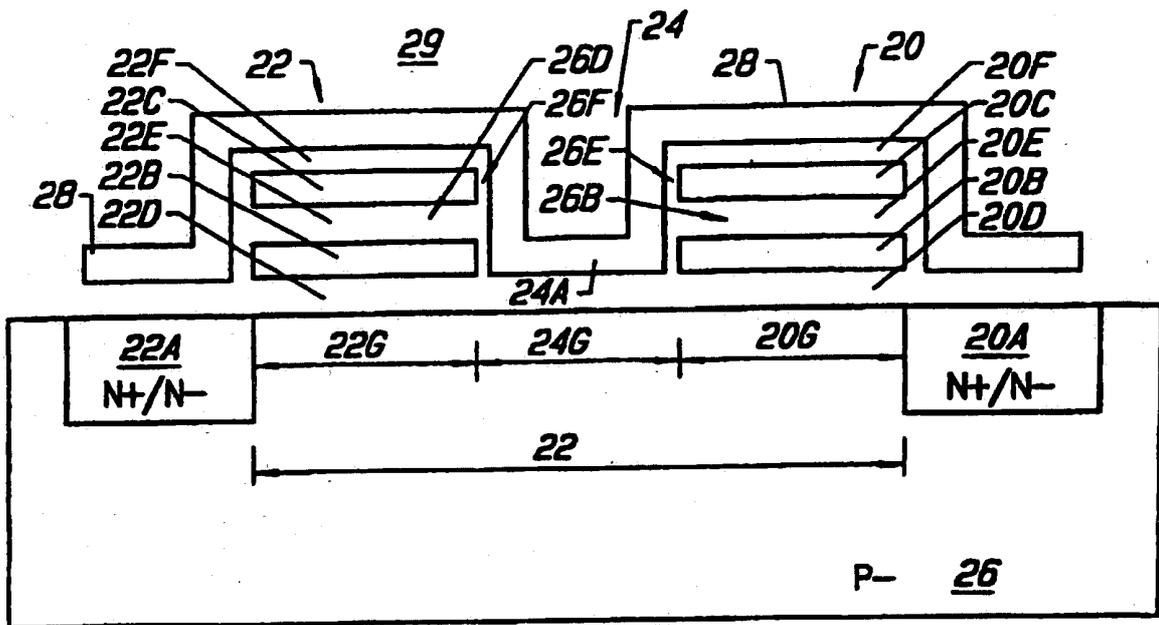


FIG. 2A

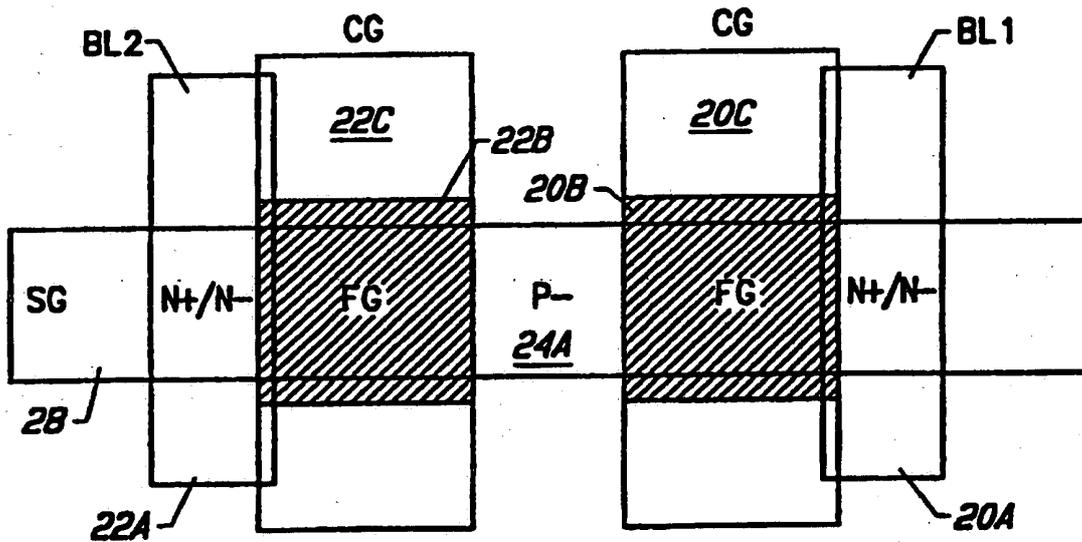


FIG. 2B

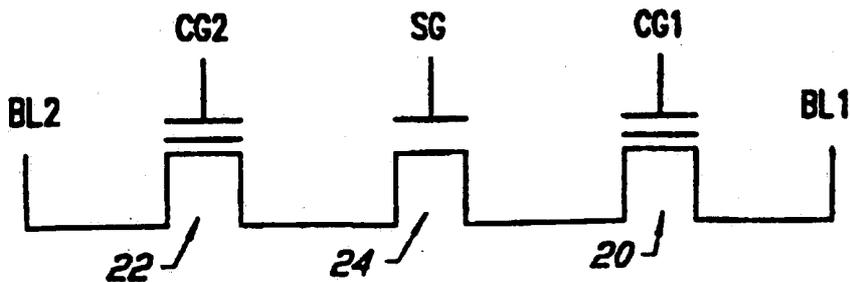


FIG. 2C

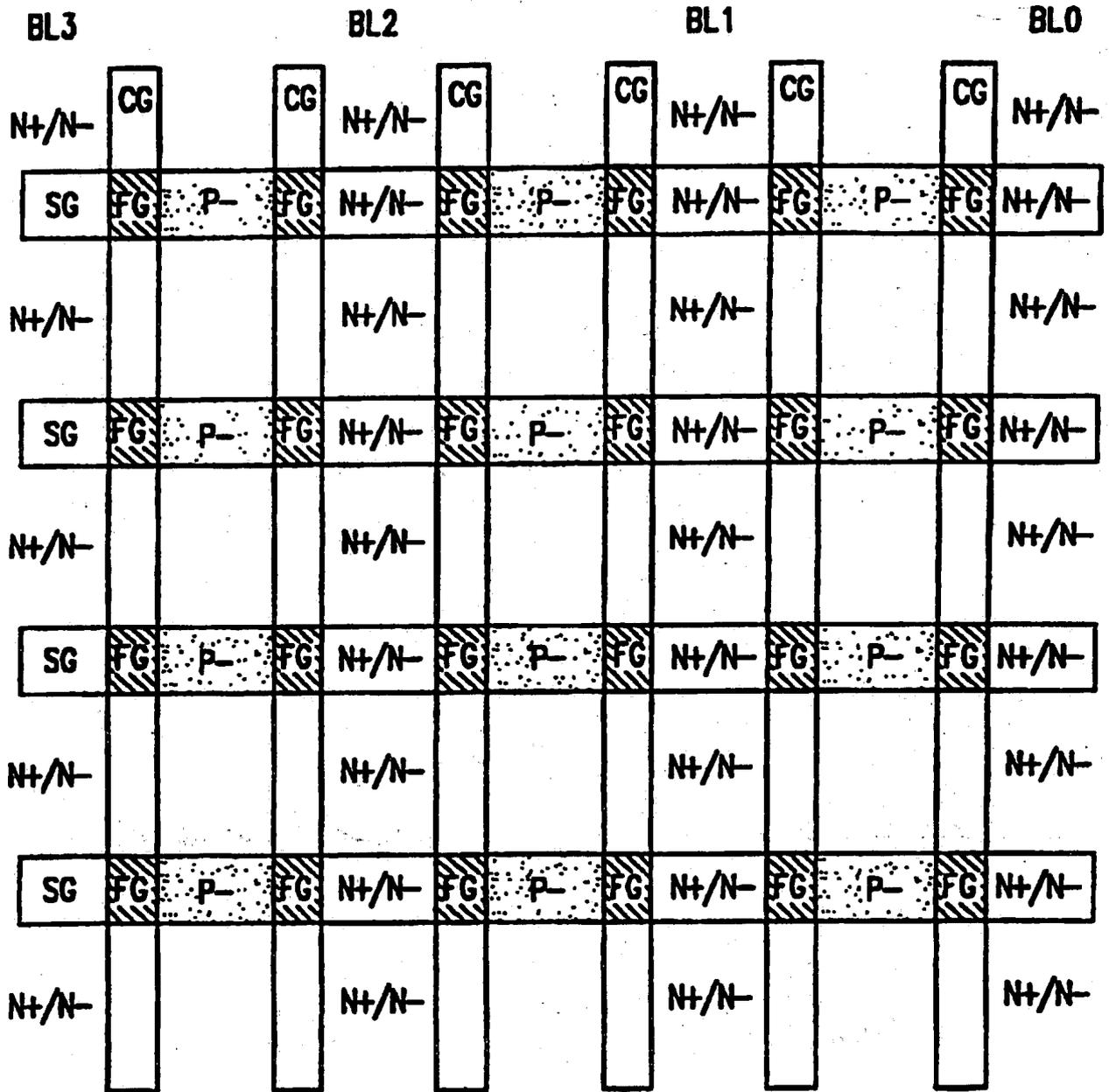


FIG. 3

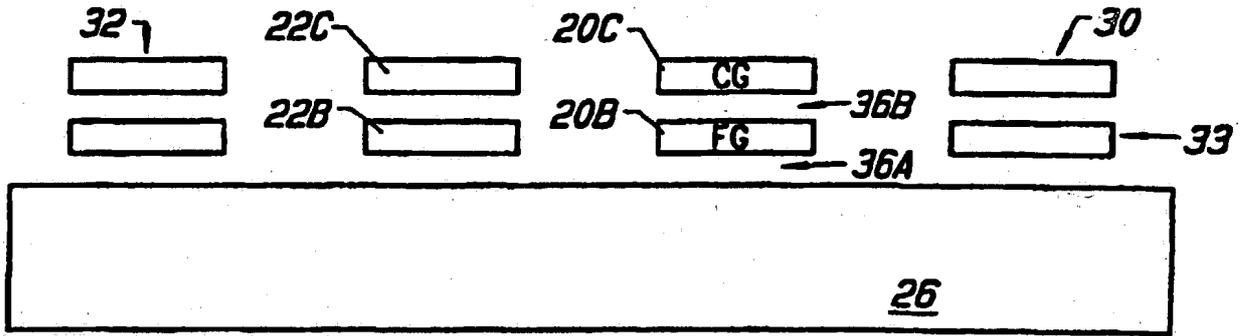


FIG. 4A

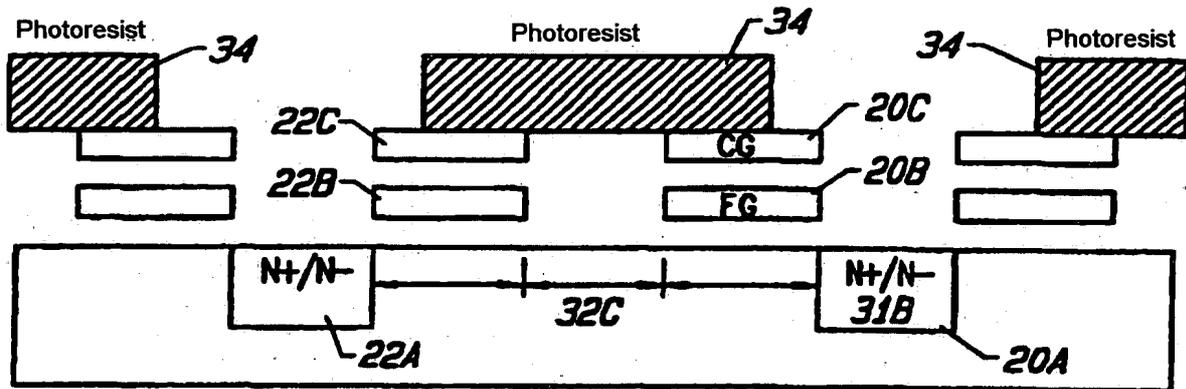


FIG. 4B

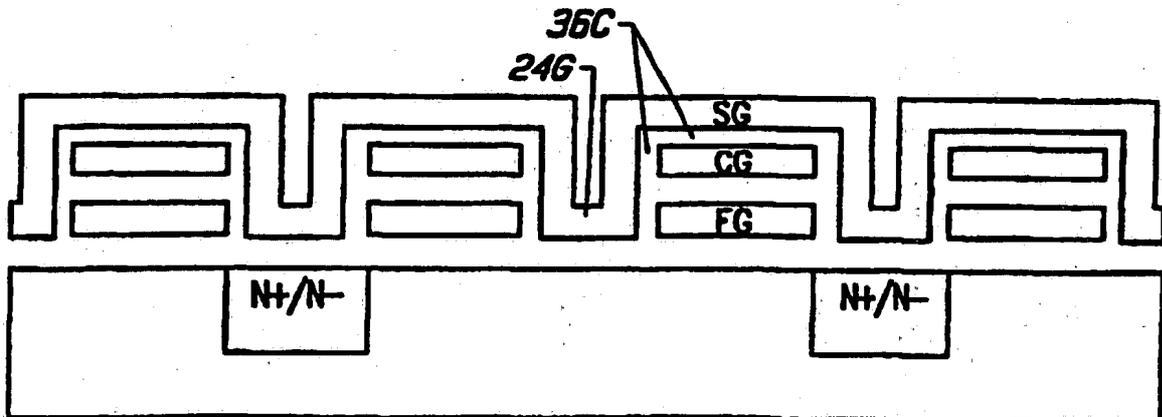


FIG. 4C

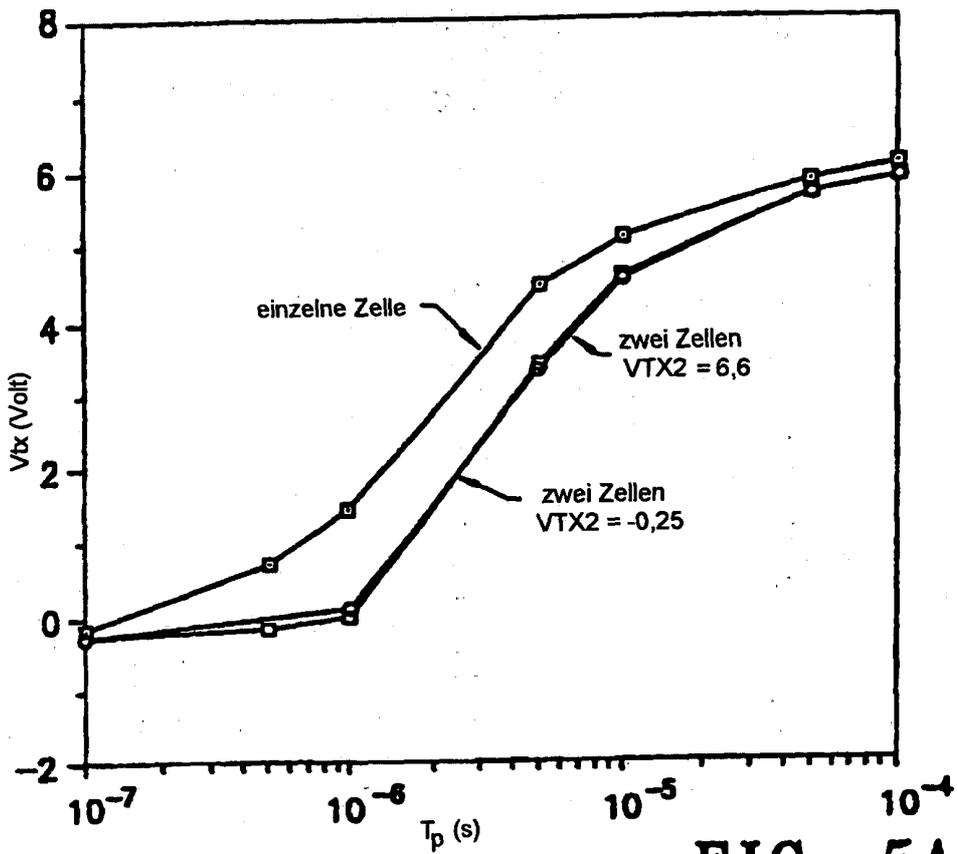


FIG. 5A

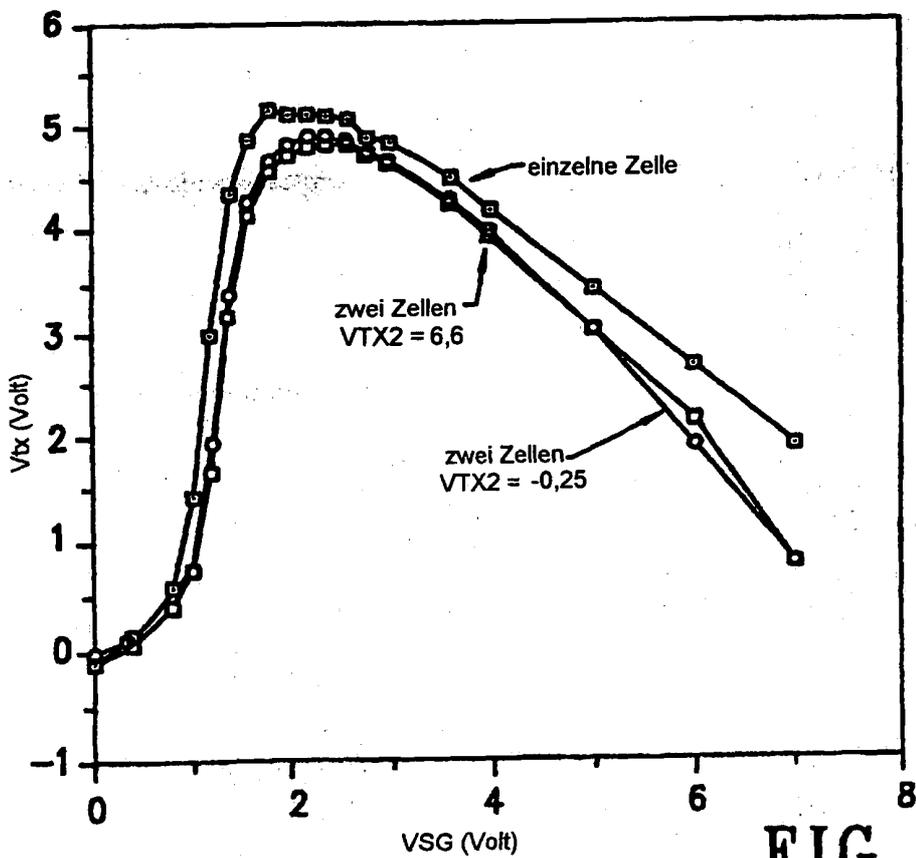


FIG. 5B

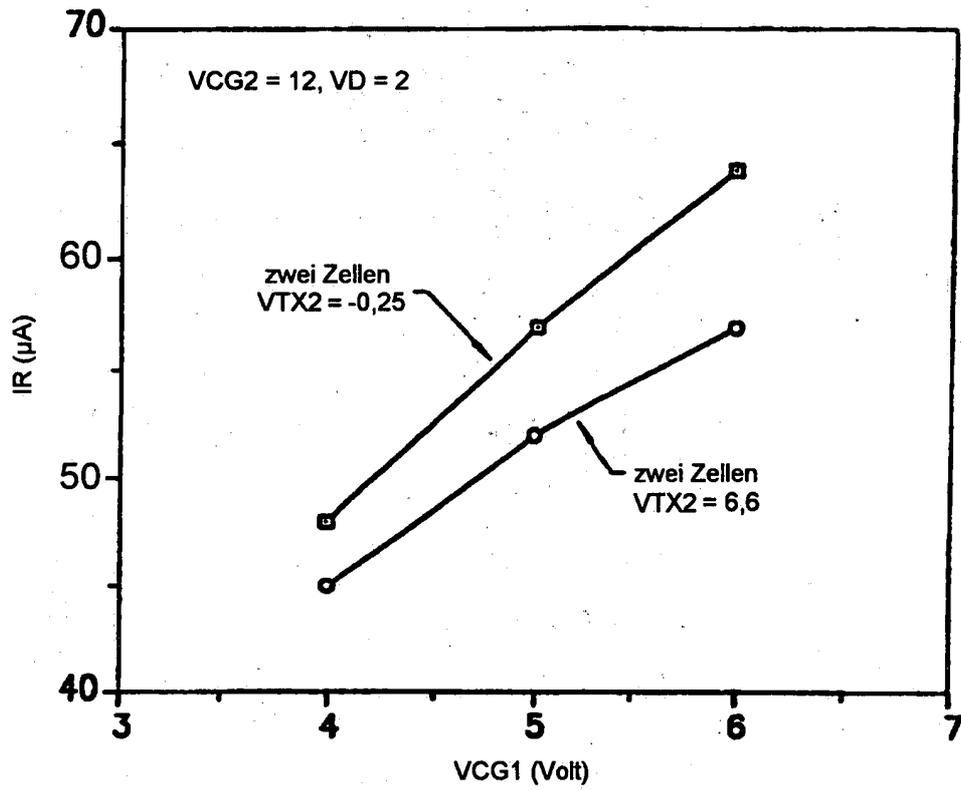


FIG. 6

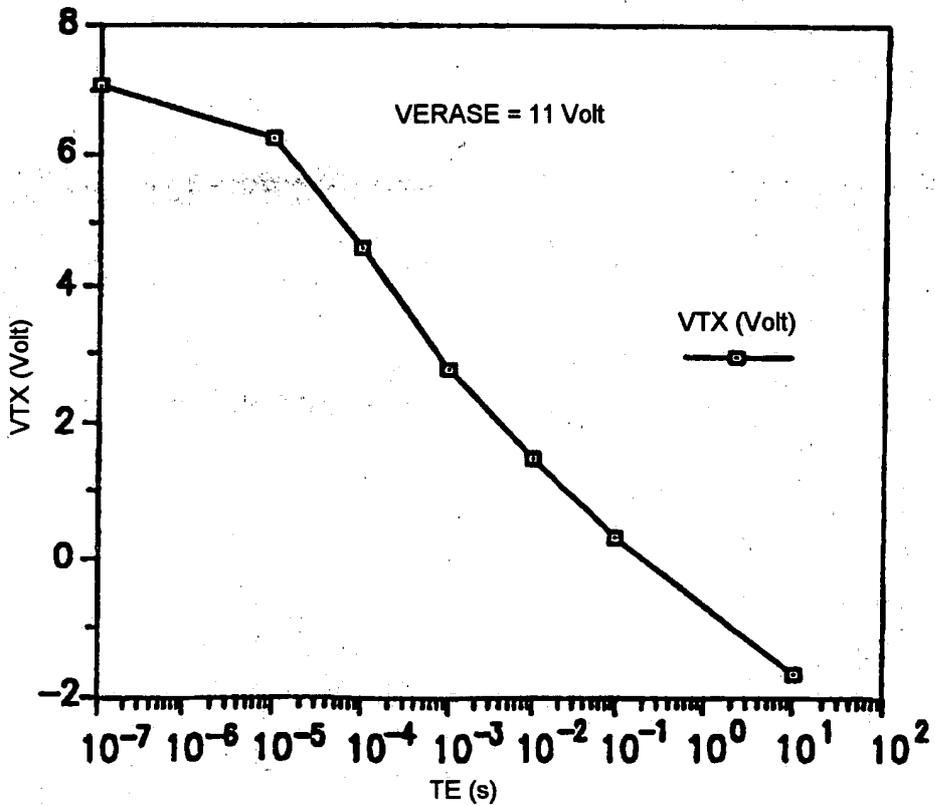


FIG. 7