OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

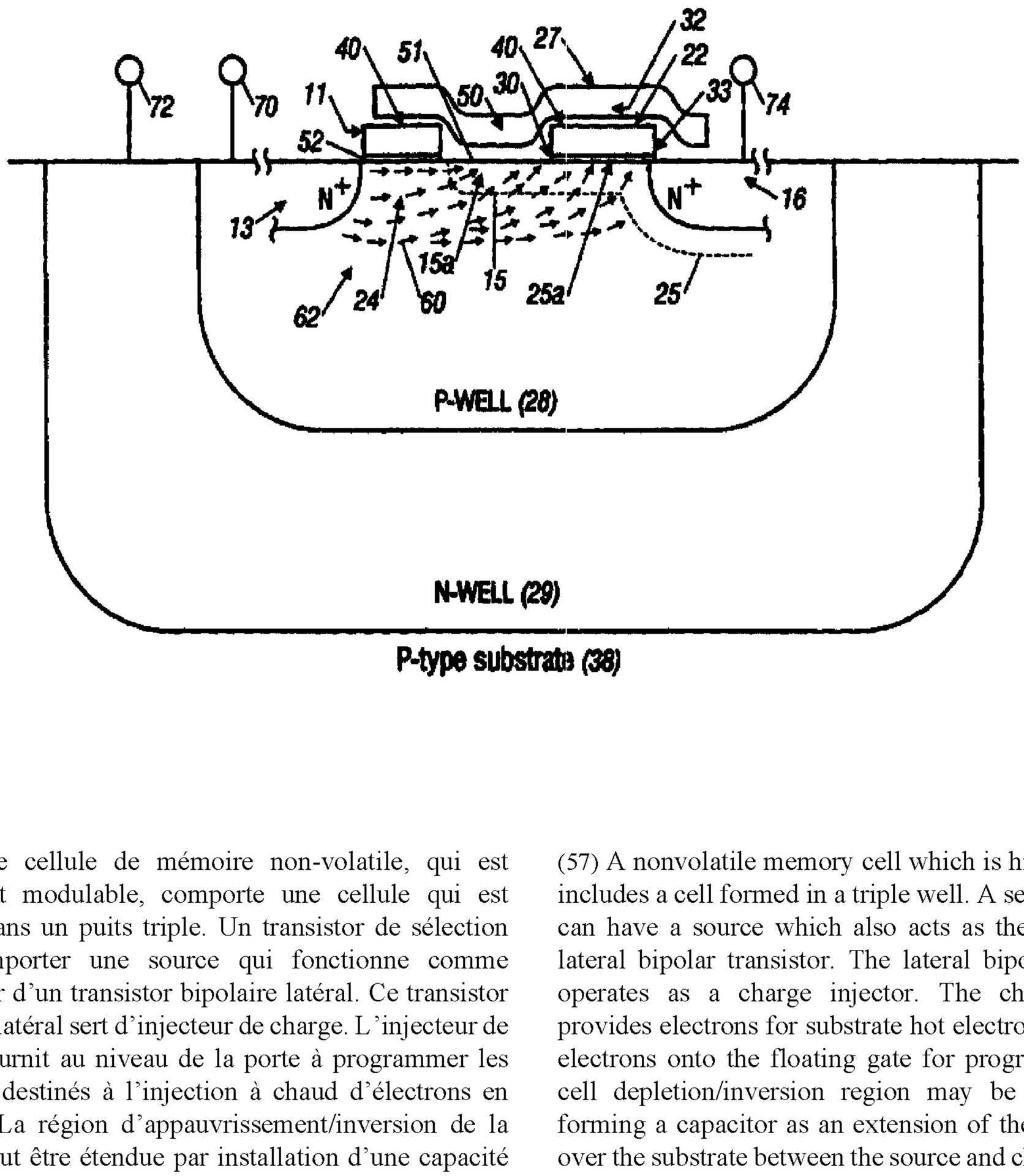
(12) (19) (CA) **Demande-Application**

(21) (A1) **2,286,180**
(86) 1998/04/09
(87) 1998/10/22

(72) WONG, TING-WAH, US
(71) PROGRAMMABLE SILICON SOLUTIONS, US

(51) Int.Cl.[6] H01L 29/788, G11C 16/04
(30) 1997/04/11 (08/838,854) US

(54) **MEMOIRE NON-VOLATILE A SEMI-CONDUCTEURS**
(54) **NONVOLATILE SEMICONDUCTOR MEMORY**



(57) Cette cellule de mémoire non-volatile, qui est hautement modulable, comporte une cellule qui est formée dans un puits triple. Un transistor de sélection peut comporter une source qui fonctionne comme l'émetteur d'un transistor bipolaire latéral. Ce transistor bipolaire latéral sert d'injecteur de charge. L'injecteur de charge fournit au niveau de la porte à programmer les électrons destinés à l'injection à chaud d'électrons en substrat. La région d'appauvrissement/inversion de la cellule peut être étendue par installation d'une capacité en extension de la grille de commande, sur le substrat, entre la source et le canal du transistor de lecture considéré.

(57) A nonvolatile memory cell which is highly scalable includes a cell formed in a triple well. A select transistor can have a source which also acts as the emitter of a lateral bipolar transistor. The lateral bipolar transistor operates as a charge injector. The charge injector provides electrons for substrate hot electron injection of electrons onto the floating gate for programming. The cell depletion/inversion region may be extended by forming a capacitor as an extension of the control gate over the substrate between the source and channel of said sense transistor.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
H01L 29/788, G11C 16/04

A1

(11) International Publication Number: **WO 98/47182**

(43) International Publication Date: 22 October 1998 (22.10.98)

(21) International Application Number: PCT/US98/07110

(22) International Filing Date: 9 April 1998 (09.04.98)

(30) Priority Data:
08/838,854 11 April 1997 (11.04.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/838,854 (CIP)
Filed on 11 April 1997 (11.04.97)

(71) Applicant *(for all designated States except US)*: PROGRAMMABLE SILICON SOLUTIONS [US/US]; 628A East Evelyn, Sunnyvale, CA 94086 (US).

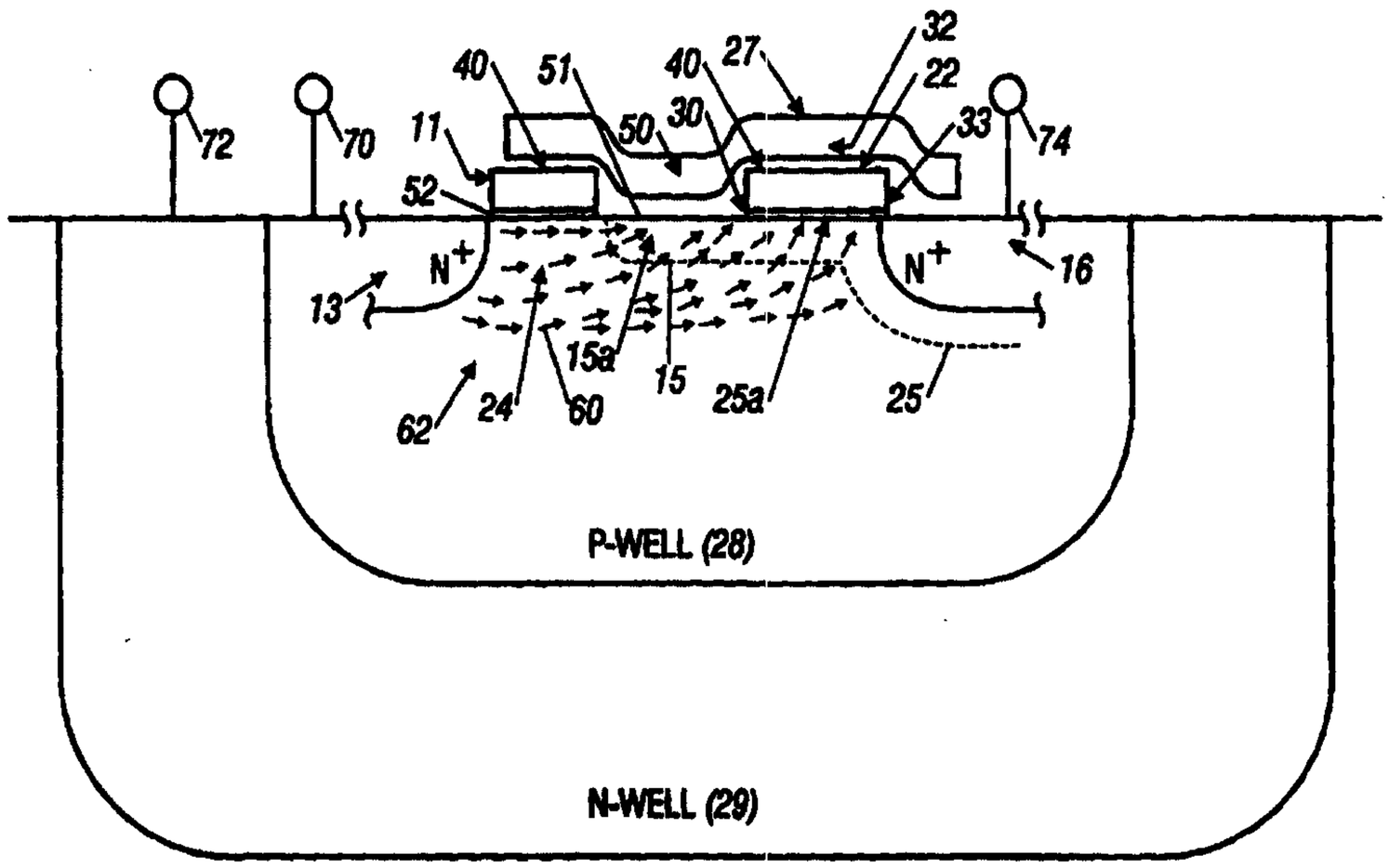
(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: WONG, Ting–wah [US/US]; 10996 Linda Vista Drive, Cupertino, CA 95014 (US).

(74) Agent: TROP, Timothy, N.; Trop, Pruner, Hu & Miles, P.C., Suite 128, 8550 Katy Freeway, Houston, TX 77024 (US).

(81) Designated States: CA, CN, KR, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: NONVOLATILE SEMICONDUCTOR MEMORY



(57) Abstract

A nonvolatile memory cell which is highly scalable includes a cell formed in a triple well. A select transistor can have a source which also acts as the emitter of a lateral bipolar transistor. The lateral bipolar transistor operates as a charge injector. The charge injector provides electrons for substrate hot electron injection of electrons onto the floating gate for programming. The cell depletion/inversion region may be extended by forming a capacitor as an extension of the control gate over the substrate between the source and channel of said sense transistor.

NONVOLATILE SEMICONDUCTOR MEMORY

Background of the Invention

This invention relates generally to nonvolatile memories and particularly to electrically erasable nonvolatile memories.

Nonvolatile memory cells are advantageous since they retain recorded information even when the power to the memory is turned off. There are several different types of nonvolatile memories including erasable programmable read only memories (EPROMs), electrically erasable and programmable read only memories (EEPROMs) and flash EEPROM memories. EPROMs are erasable through light exposure but are electrically programmable by channel hot electron injection onto a floating gate. Conventional EEPROMs have the same programming functionality, but instead of being light erasable they can be erased and programmed by electron tunneling. Thus, information may be stored in these memories, retained when the power is off, and the memories may be erased for reprogramming, as necessary, using appropriate techniques. Flash EEPROMs may be block erased, typically giving them better read access times than regular EEPROMs.

Currently, flash memories have gained considerable popularity. For example, flash memories are often utilized to provide on-chip memory for microcontrollers, modems and SMART cards and the like where it is desirable to store codes that may need fast updating.

While flash memories and EEPROMs are closely related, in many instances flash memories are preferred because their smaller cell size means that they can be made more economically. However, flash memories and EEPROMs often have very similar cell attributes.

Nonvolatile memory cells differ in certain respects from the transistors that are generally utilized in electronic components called logic devices, such as

microcontrollers, that work with the memory cells. Logic devices are formed of transistors that use a single gate electrode. Nonvolatile memories usually include two gate electrodes, known as the control and floating gate electrodes, situated one over the other. Because of this structural difference, nonvolatile memories and logic devices may be made by different processes. This may contribute to a substantial increase in process complexity and manufacturing cost.

Particularly with an EEPROM, the electrical programming of the cells normally requires substantial potentials to be applied to the cells. These potentials induce electron tunneling from an N+ region onto the floating gate. Additional complexity may arise from the need to provide substantially larger voltages to memory cells than are needed for normal transistor operation.

While the industry has come to accept the need for separate process technologies for logic and nonvolatile memories and while those in the industry have also come to appreciate that significant voltages are needed to program EEPROMs and significant currents to program flash EEPROMs, there would be a substantial demand for a nonvolatile memory which was both electrically erasable and programmable without the need for special process technologies or for relatively higher programming voltages and higher currents.

Furthermore, with the conventional FLASH EEPROMs, the electrical programming of the cells normally requires high current to be applied to the cells. A very minute amount of this electron current becomes injected from the drain depletion region onto the floating gate. This means that the injection efficiency of such devices is low (e.g., 1×10^{-6} to 1×10^{-9}). The requirement of high current adds additional complexity because of the design of the high current pump operated at low voltage.

Summary of the Invention

In accordance with one aspect of the present invention, an electrically erasable and programmable memory includes a sensing transistor having a floating gate, a channel, a source and a drain. A bipolar transistor is adapted to supply electrons for programming the floating gate by substrate hot electron injection of electrons onto the floating gate through the channel. The bipolar transistor is arranged such that its collector is the biased depletion region under the channel of the sensing transistor. A select transistor is formed adjacent to the sensing transistor. The source of the select transistor is the emitter of the bipolar transistor.

In accordance with another aspect, a memory cell includes a sense transistor having a control gate, a floating gate and a source and drain formed in a substrate. The control gate is situated over the floating gate. The sense transistor is adapted to form a depletion region in the substrate. A capacitor is formed adjacent the sense transistor floating gate, between the source and drain. The channel under the capacitor connects the channel regions of the sense and select transistors. The capacitor includes a plate formed from the control gate. The capacitor is situated so as to extend the inversion or depletion region formed by the sense transistor.

In accordance with still another aspect, a memory includes a floating gate over a substrate defining a channel in said substrate. A source of substrate electrons is spaced laterally away from the floating gate. A substrate electron path is provided for substrate electrons from the source to said channel free of any intervening N-type doped regions in the line between the source and channel. In one embodiment, a capacitor is formed between the source and the floating gate so as to form a depletion/inversion region adjacent to the floating gate

and in the substrate electron path. In another embodiment, the memory may include a single source and a single drain that are shared by the sense and select transistors. In another embodiment, there may be no heavily doped regions formed in the substrate between the source and the region of the substrate under the floating gate.

In accordance with yet another aspect of the present invention, a memory includes a source and a drain formed in a substrate. The floating gate is formed adjacent the drain over the substrate between the source and drain. A control gate is located over the floating gate and a portion of the control gate extends over the substrate between the floating gate and the source. A transistor gate is positioned adjacent the source and between the source and the control gate portion. The source and drain are arranged to act as a source and drain for both of the floating gate and the transistor.

In accordance with but another aspect of the present invention, a method of programming a memory includes the step of supplying substrate electrons from a source spaced from a sense transistor having a control gate and a floating gate. A depletion region is created under the floating gate and an underlying channel. Substrate electrons are supplied to the channel along a path from source to channel free of intervening N-type doped regions in the line from source to channel substantially parallel to the substrate surface. In one embodiment, the depletion region is extended laterally toward the substrate electron source beyond the region under the floating gate.

In accordance with still another aspect of the present invention, a memory cell includes a source and drain spaced from one another in a substrate. A floating gate and a transistor gate are arranged over the substrate between the source and drain. A device for forming an

inverted region in the substrate is located between the transistor gate and the floating gate.

Brief Description of the Drawing

Figure 1 is a schematic depiction of an array configuration for one embodiment;

Figure 2 is a greatly enlarged top plan view showing the layout of a semiconductor implementation of one cell of the embodiment shown in Figure 1;

Figure 3 is a cross-sectional view taken generally along the line 3-3 in Figure 2; and

Figure 4 is a partial three dimensional view of the cell shown in Figure 3.

Description of the Preferred Embodiment

Referring to the drawing wherein like reference characters are used for like parts throughout the several views, a memory cell 10, shown in Fig. 1, includes a sense transistor 12 and a select transistor 14. This structure is advantageously implemented on a semiconductor layer having situated thereon an electrically isolated floating gate 22.

For each cell 10a - 10d, the source 13 of the select transistor 14 is controlled by the source node 56. The gate of the select transistor 11 is controlled by the node 51. The control gate 27 of the sense transistor 12 is controlled by the control node 57. The drain 16 of the sense transistor 12 is connected to the drain node 55.

One layout for implementing a cell 10, shown in Fig. 2, includes the control gate 17. The control gate 17 extends across the active region 18 which is bordered by the drain 16 of the sense transistor 12 and the source 13 of the select transistor 14. The select gate 11 also extends parallel to and underneath the control gate 17, bordering the edge of the control gate 17 and the region

15a. The control gate 17 may be non-self-aligned to the select gate 11 and the sense gate 12. The floating gate 22 is also situated in isolation under the control gate 17, over the active region 18.

The drain 16 may include a contact 55 as illustrated in Fig. 2 which connects to the drain diffusion 16. The source node 56 may also be implemented by a contact.

The relationship of the sense transistor 12 and select transistor 14 is illustrated in Figure 3. The floating gate 22 forms portions of the transistor which has a drain 16 and a source 13. Similarly, the select gate 11 forms the other portion of the transistor between the source 13 and the drain 16. The sense transistor 12 includes a channel 25a while the select transistor 14 includes a channel 24. The control gate forms the plate of the capacitor whose channel is 15a. The select gate 14, the floating gate 22 and the control gate 27 form the gates of a transistor with a source 13 and drain 16.

In the illustrated embodiment, the channels 25a and 24 are P-type semiconductor material and are part of a P-well 28. The P-well 28 in turn is formed in an N-well 29. Finally, the N-well 29 is formed in a P-type substrate 38. The P-well 28 may be biased, as indicated at 70, and the N-well 29 may be biased, as indicated at 72.

A capacitor 50 is formed where the control gate 27 overlays the substrate region 15a between the select gate 11 and the floating gate 22. It controls the depletion/inversion of the region 15a to create an extension of the depletion/inversion region 25 formed by the sense transistor 12. It is operated during programming and read operations. In read operation, the capacitor 50 bridges the sense and select transistor channels by forming an inverted region. When the inverted region, under region 15a is formed, the sense and select transistors 12 and 14 are connected.

The floating gate 22 forms the tunneling capacitor 33 by its interaction with the channel 25a. A tunnel oxide 30 separates the floating gate 22 from the channel 25a. Similarly the interpoly dielectric oxide 40, which is part of a coupling capacitor 32, separates the floating gate 22 from the control gate 27. Finally, the control gate 27 is separated by the oxide 51 from the region 15a. Likewise the select transistor 14 includes a gate oxide 52, which may be of the same thickness as the tunnel oxide 30.

The overlaying of the control gate 27 over the select gate 11 is for processing convenience. Similarly, the control gate 27 is shown as overlaying the drain 16, but this too is merely for processing convenience. The control gate 27 need not be self-aligned to the sense or select transistors.

The cell 10 may be described as a flash EEPROM utilizing high efficiency substrate hot electron injection for programming and Fowler-Nordheim tunneling for erasure. The process of substrate hot electron injection is well described in T. H. Ning, C. M. Osburn, and H. W. Yu "Emission Probability of Hot Electrons from Silicon into Silicon Dioxide", J. Appl. Phys., vol. 48, p.286, (1977); Boaz Eitan, James L. McCreary, Daniel Amrany, Joseph Shappir, "Substrate Hot-electron Injection EPROM", IEEE Transactions on Electron Devices, Vol. ED-31, No. 7, p.934, (July 1984); I. C. Chen, C. Kaya, and J. Paterson, "Band-to-Band Tunneling Induced Substrate Hot-electron (BBISHE) Injection: A New Programming Mechanism for Nonvolatile Memory Devices", IEDM (1989) p.263; and C. Y. Hu, D. L. Kencke, S. K. Benerjee, "Substrate-current-induced Hot Electron (SCIHE) Injection: A New Convergence Scheme for FLASH Memory," IEDM (1995), p.283. Each of these articles is hereby expressly incorporated by reference herein.

Programming is achieved by high efficiency substrate hot electron injection. As indicated in Figure 3,

substrate electrons, indicated by the arrows at 60, are generated by forward biasing the source 13 which is separated from the sense transistor 12 channel 25a by the select transistor channel 24 and the region 15a under the capacitor 50. Some of the substrate electrons 60 diffuse through the region underneath the channel 24 to the channel region 25a underneath the sense transistor 12.

For cells that need to be programmed, the channel region 25a is biased such that a depletion region 25 is formed. The depletion region 25 is extended, as indicated at 15, by the operation of the capacitor 50, under the capacitor 50 as well. When an electron gets to the depletion region 25, it is accelerated by an electric field, Vcs. The electric field Vcs is the difference between the channel 25a potential (potential of the surface inversion region) and the P-well 28 potential. Some of these electrons gain sufficient energy, in excess of the effective oxide barrier height potential, to be injected onto the floating gate 22.

For cells that are not to be programmed, the channel-to-P-well potential is less than the effective oxide barrier height. In such case, the electrons would not gain sufficient energy to overcome the barrier height and are not injected onto the floating gate 22.

The N+ doped region 13, the P-region 24 under the select transistor 14 and the sense transistor channel 25a, and the depletion region 15 under the capacitor 50 form a lateral bipolar transistor 62. The emitter (source 13) of the bipolar transistor 62 acts as a charge injector, injecting substrate electrons from the source diffusion to the biased depletion region under the floating gate 22. With the diffusion 13 as the emitter and the channel 24 as the base, the collector is the biased depletion region 25 (including the region 15). Since the channel region 25a acts as the channel for the sense transistor during read,

and the biased depletion region 25 under the sense transistor 12 acts as the collector of the bipolar transistor 62 during programming, a compact cell layout is achieved.

The efficiency of substrate hot electron injection is a function of a number of characteristics. Considering the depletion region 25, electrons scatter with lattice phonon scattering across the depletion region 25 with a certain electron mean free path. Some of these electrons, without much scattering, gain sufficient energy to overcome the effective barrier height and are injected onto the floating gate 22. Some electrons gain less energy than the effective barrier height and are not injected onto the floating gate 22. The injection efficiency is a strong function of the doping concentrations and the channel-to-P-well potential, Vcs.

Since the cell 10 is situated in a P-well 28 embedded in an N-well 29, during programming the floating gate 22 is capacitively coupled to a higher voltage through the coupling capacitor 32 by raising the control gate 27 to Vpp, which may be from 7 to 14 volts. The voltage that the floating gate 22 attains at low drain bias is approximately a function of the voltage on the floating gate when the control gate 27 and the P-well 28 and drain 16 are at ground, plus the coupling ratio times the voltage on the control gate 27. The coupling ratio, to a first order, is approximately equal to the capacitance of the capacitor 32 divided by the sum of the capacitances of the coupling capacitor 32 and the tunneling capacitor 33.

When the select transistor 14 is off, the sense transistor drain 16 potential can be forced close to the supply potential Vcc or higher. Since the select transistor 14 is off, the potential of node 51 follows the channel 25a potential. The channel 25a potential, which is the potential of the surface inversion region of the

channel region 25a, is set as follows. When the potential of the floating gate 22 (Vfg) is one sense transistor 12 threshold voltage higher than the drain 16 potential, the channel potential is the same as the drain potential. On the other hand, when the floating gate 22 potential is less than the drain 16 potential plus the sense transistor 12 threshold voltage, the channel potential is the difference between the floating gate 22 voltage and the sense transistor 12 threshold voltage.

The P-well potential is the voltage 70 applied to the P-well 28. Since the P-well 28 is embedded in an N-well 29, and the N-well is set at a voltage 72 approximately Vss or higher, the P-well potential Vp can be negative, typically negative one to negative two volts. Moreover, it is usually less than the effective oxide barrier height to avoid any potential disturb problem.

The potential difference between the channel 25a region and the P-well 28 potential (Vp) 70 is the voltage across the depletion region 25. For cells to be programmed, the drain 16 voltage is raised high, typically close to Vcc or higher. A depletion region 25 in the channels 25a and 24 underneath the sense transistor 12 and the capacitor 50 is formed with a voltage drop equal to the channel potential minus the P-well potential 70.

For those cells that are not to be programmed, the drain 16 voltage 74 is set to zero volts (Vss). The voltage drop across the depletion region 25 then is equal to the absolute value of Vp, which is typically less than the effective oxide barrier height.

Cell 10 erasure is achieved by Fowler-Nordheim tunneling of electrons from the floating gate 22 to the channel region 25a and the drain diffusion 16. During erasure, the control gate 27 is forced to a negative voltage from -7 to -14 volts, for example. As for the drain diffusion 16, the P-well 28, and the N-well 29, they

are biased to a positive potential close to Vcc or higher. Vcc is determined by the particular technology utilized. For example, it could be 5.0 to 2.5 volts with present technologies. This reduces the electric field across the junction between the N+ diffusion 16 and the P-well 28. The reduced field prevents acceleration of hot hole trapping in the gate oxide under the floating gate 22.

The drain 16 is preferably not biased to a voltage higher than the P-well 28 to such an extent that gate induced drain leakage (GIDL) becomes a problem. With current technologies, this means that the drain 16 bias can not be higher than the P-well 28 bias by about one volt. In addition, if the drain 16 bias significantly exceeds the P-well 28 bias, hot hole trapping may occur in the select gate oxide 52 due to the lateral junction field acceleration.

The ability to apply a positive voltage to the P-well arises because the P-well 28 is embedded in an N-well 29. The P-well voltage is preferably equal to or less than N-well potential to avoid P-well/N-well forward biasing. Thus, applying a positive voltage of Vcc or higher to the P-well, N-well and the drain 16 can eliminate hot hole trapping induced by GIDL while allowing the drain 16 voltage to be raised to Vcc or higher.

The voltage across the capacitor 33 is the difference between the floating gate 22 potential on the one hand and the diffusion 16 and P-well 28 potentials. When the difference exceeds 8 to 10 volts, sufficient tunneling current is generated and the floating gate 22 can be erased to a negative potential in the time frame of a few milliseconds to a few seconds, depending on the tunneling oxide 30 thickness.

Electrons tunnel to the drain region 16 (drain erase). The tunneling current depends on the voltage from the floating gate 22 to the drain 16.

Reading the programming state of the cell 10 may be accomplished as follows. The floating gate 22 is capacitively coupled to a higher potential by forcing the control gate 27 to a positive potential, for example, of 2.5 to 5 volts. The floating gate 22 is coupled to a potential Vfg which can be calculated as being equal to the sum of the floating gate potential when the control gate 27 is held at ground, plus the potential on the control gate 27 times the coupling ratio.

The drain 16 potential during reading is limited to a voltage of less than 2 volts. This is to avoid any read disturb.

For the selected cell to be read, the select gate 11 is forced to Vcc, and the source 13 is forced to ground. The unselected select gate 11 is also forced to ground.

When these potentials are applied to the selected cell 10, a current flows through the sense transistor 12. This current is then fed to a current sense amplifier (not shown). If the voltage on the floating gate 22 is greater than the threshold voltage on the sense transistor 12, a higher current, perhaps greater than 20 microamps, is detected as the conducting state. When the potential of the floating gate is less than the threshold voltage, a lower current, for example or less than one microamp flows, and a nonconducting state is detected.

A detected conducting state can be called a one state. A nonconducting state can be called the zero state.

An example of the operation of the cell in programming, reading and erasing is summarized in the following chart:

Cell Operation

	Erase	Program	Read
		(Selected)	(Unselected)
		(Unselected)	(Select)
	(Select)		
	(Unselected)		

	Drain (55)	≥Vcc	*0 or ≥Vcc	≥Vcc	*0 or ≥Vcc	~1.5V	* 0 or ~1.5V			
	Source (56)	Float	Float	Vs	Float	Vss	Vss			
		Vss	Vss	Vss to Vs	V	Vcc	Vss			
5	Select (51)	-7 to -14V	Vss		s	2 to 5V	Vss			
	Control (57)	≥Vcc	≥Vcc			Vcc	Vcc			
		≥Vcc	≥Vcc		s	Vss	Vss			
	N-well				t					
10	(29)				o					
	P-well (28)				V					
					s					
				Vpp	Vss					
				Vcc to Vss	V					
					c					
					c					
					t					
					o					
					V					
					s					
					s					
				Vbias	Vbias					

*0 is for unselected column.
Vpp ≈ 7 to 14 volts.

Vs is the node voltage set by the injection current level, ranging from a few tens of nanoamps to a few tens of microamps depending on the programming speed requirement. Typically it would be from tens of milliseconds to tens of microseconds. Vbias is the bias on the P-well 28 which can be Vss or it can be forced to -1 to -2 volts to enhance injection efficiency.

A suitable on-chip circuit for generating two negative bias potentials, one to bias the control gate 57 and the other to negatively bias the P-well 28, can be found in L.A. Glasser and D.W. Dobberpuhl, "The Design and Analysis of VLSI Circuits", (Dec. 1985), published by Addison-Wesley, at pages 301-329, hereby expressly incorporated by reference herein. Vss is the external ground potential.

While the cell 10 may be utilized as a single element, it can also be connected as an array as shown in Fig. 1. In the array, a plurality of cells 10, 10a, 10b, 10c, 10d are depicted. The source node 13 is formed by connecting all of the source nodes of cells in the same row as one single node 56. The control node 17 is formed by connecting all the control nodes of the individual cells in the same row together as a single node 57. The select gate node 11 is formed by connecting the select gate nodes for all the cells on the same row together as one node 51. Similarly, the drain node 16 is formed by connecting all the drain nodes for cells in the same column together as a single node 55. This node leads to the sense amplifier (not shown).

The cells in the array may be formed using conventional process technologies such as a double poly, single metal CMOS process. The illustrative parameters set forth

herein contemplate a .25μm or lower feature size with Vcc potentials of 1.8 volts. As the technology permits lowering voltages and smaller feature sizes, the parameters herein would scale accordingly.

The starting substrate material is typically P-type (100) silicon, for example having a resistivity in the range of 10-20 ohm-cm. The P-well 28 is embedded in an N-well 29 in the so-called triple well process. The P-well 28 has a typical well depth of, for example, 2 to 4 um with an average doping concentration, for example, in the range of 1×10^{16} to 5×10^{16} atoms per cubic centimeter.

The N-well 29 has a typical well depth of, for example, 4-8 um. The doping concentration may be from 4×10^{15} to 1×10^{16} atoms per cubic centimeter. The triple well is formed by the P-well 28 counterdoping the N-well 29.

The formation of the elements in the triple well is as follows. An N-well implant is done, for example, with phosphorous 31 with a typical dose of 1 to 1.5×10^{13} atoms per square centimeter and an energy of 160 to 100Kev. The N-well implant is driven using a high temperature step which may typically be 6 to 12 hours at 1125 to 1150° C. The N-well 29 is then counterdoped with a P-well implant. Typical dosages for the P-well implant could be 1.5 to 2.5 $\times 10^{13}$ atoms per square centimeter with an energy of 30Kev to 180Kev using a species such as boron 11. The N-well 29 and P-well 28 are then driven, typically 6 to 10 hours at 1125 to 1150°C. This sets the wells to the desired doping concentrations and depths.

After the well formation, standard logic field oxide formation and channel stop formation steps are applied. The field oxide thickness and implant doses are adjusted to achieve a field threshold of 7 to 14 volts, which is determined by the Vpp level for programming and erasing and by logic process capability. After this formation, a memory cell implant may be performed. For example, a B11

implant at 30 to 50Kev with a dose of 1.5 to 3 x 10^{13} atoms per square centimeter may be done through a sacrificial oxide. The gate oxide 52 and the tunnel oxide 30 are then formed. For example, an 85 to 100 Angstrom dry oxide may be grown across the wafer. A dry oxide is grown, for example, at 900°C in partial oxygen followed by a 975 to 1050°C anneal.

The floating gate 22 may then be formed of polysilicon, silicide or metals. If polysilicon is used, it can be 1600 Angstroms thick, and POCL3 doped at 870 to 1000°C. The interpoly dielectric is formed of an oxide-nitride-oxide sandwich (ONO) with the lower oxide being from 60 to 80 Angstroms, the nitride layer having a thickness of from 90 to 180 Angstroms and the upper oxide being from 30 to 40 Angstroms. A 125 to 200 Angstrom gate oxide may be grown for the oxide under the capacitor 50.

After floating gate and select gate definition, a N+ implant is implanted into the source of the select transistor 14 and the drain of the sense transistor 12. It is blocked between the two gates so that it does not enter the region under the plate of the capacitor 50, the N+ implant can be, for example, a phosphorous implant at 60Kev, at a dose of 1 to 3 x 10^{14} atoms per square centimeter may be followed by arsenic at 60Kev, 2.5 to 4.5 x 10^{15} atoms per square centimeter. It is also possible to form the source and drain using lightly doped drain (LDD) technology.

The polysilicon (poly 2) for the control gate may then be deposited and silicided if desired. The gates are patterned and defined using standard techniques. The control gate is not self-aligned to the sense and select gates.

With the completion of these capacitor and transistor structures, all subsequent processing for

contacts and interconnect layers follows standard logic rear end processing.

Referring to Fig. 4, the flow of electron current from the injector, which in the illustrated embodiment is the source 13, to the floating gate 22, is shown. The electrons may flow from the source 13 along the entire width "W" of the cell between adjacent columns. In addition, electrons flow from the portions 13a and 13b extending beyond the cell width. Namely, the portions of the cell 13a and 13b extending to either side of the cell in the width direction also contribute electrons which may be utilized to program the floating gate 22, as shown in Fig. 4. This may significantly increase the available electron current for programming and thus may expedite programming.

This arrangement is possible because the N+ source 13 runs parallel to the control gate 27. In other words, the source 13 runs parallel to the rows of the device. Because of this geometry, a portion of the source which is greater in length than the width of the cell may be utilized for programming purposes. Namely, additional electrons may flow from the extensions 13a and 13b between adjacent columns of cells.

In addition, the effect may be augmented by turning off both adjacent columns. This avoids the loss of current through the adjacent columns and allows available electrons in the adjacent source extensions 13a and 13b to flow towards the floating gate 22. The mode table below indicates how the system would operate during programming with adjacent columns turned off.

Optional Mode to Enhance Programming

	Select	Adjacent Select

Drain	≥ Vcc	Vss or Float
Source	Vs	Vs
Select	Vss to Vs	Vss to Vs
Control	Vpp	Vpp
N-Well	Vcc to Vss	Vcc to Vss
P-Well	Vbias	Vbias

While a number of parameters and levels were provided in the foregoing description, those skilled in the art will appreciate that these parameters and levels are merely for illustration purposes. For example, by reversing the conductivity types of the doped junctions and the bias polarities, a cell structure using substrate hot hole injection may be implemented. It is intended that the appended claims cover all modifications and variations as fall within the true spirit and scope of the present invention.

What is claimed is:

1. An electrically erasable and programmable read only memory comprising:

a sensing transistor including a floating gate, a channel, a source and a drain;

a bipolar transistor adapted to supply electrons for programming said floating gate by substrate hot electron injection of electrons onto said floating gate through said channel, said bipolar transistor arranged such that its collector is also the biased depletion region under the channel of said sensing transistor; and

a select transistor formed adjacent to said sensing transistor, said source of said select transistor being the emitter of said bipolar transistor.

2. The memory of claim 1 wherein said control gate extends over the gates of both of said select transistor and the floating gate of said sensing transistor.

3. The memory of claim 1 wherein said control gate forms a capacitor adjacent said sensing transistor.

4. The memory of claim 3 wherein said third transistor is between said select transistor and said sensing transistor.

5. A memory cell comprising:

a sense transistor having a control gate, a floating gate and a source and drain formed in a substrate, said control gate situated over said floating gate, said sense transistor adapted to form a depletion region in said substrate; and

a capacitor formed adjacent said sense transistor floating gate, between said channel and said drain, said capacitor including a plate formed from said control gate, said capacitor situated so as to extend the

depletion/inversion region formed by said sense transistor.

6. The cell of claim 5 including a select transistor, said capacitor formed between said select transistor and said sense transistor.

7. The cell of claim 6 including a bipolar transistor formed from the source of said select transistor, the channel of said select transistor and the biased depletion region under the channel of said sense transistor.

8. The cell of claim 7 wherein said P-well is negatively biased and said sense and select transistors are n-channel transistors.

9. The cell of claim 5 including a select transistor, said sense and select transistors having a single source and a single drain, said capacitor arranged to bridge the channels of said select and sense transistors.

10. A memory comprising:

a floating gate over a substrate defining a channel in said substrate;

a source of substrate electrons spaced laterally away from said floating gate; and

a substrate electron path for substrate electrons from said source to said channel free of any intervening N-type doped region in the line between the source and channel.

11. The memory of claim 10 including a capacitor formed between said source and said floating gate so as to

form a depletion/inversion region adjacent to said floating gate and in said substrate electron path.

12. The memory of claim 11 wherein said capacitor is formed from a control gate electrode which extends over said substrate and over said floating gate.

13. The memory of claim 12 including a select transistor having a gate arranged between said capacitor and said source.

14. The memory of claim 13 wherein said source of substrate electrons is the select transistor source.

15. The memory of claim 14 wherein there are no heavily doped N-type regions formed in said substrate between said source and the region of said substrate under said floating gate.

16. The memory of claim 15 including a single source and a single drain shared by said sense and select transistors.

17. A memory comprising:

a source and a drain formed in a substrate;

a floating gate formed adjacent said drain over said substrate between said source and drain;

a control gate over said floating gate and a portion of said control gate extending over said substrate between said floating gate and said source; and

a transistor gate adjacent said source and between said source and said control gate portion, said source and drain arranged to act as the source and drain for both of said floating gate and said transistor gate.

18. A method of programming a memory comprising the steps of:

supplying substrate electrons from a source spaced from a sense transistor having a control gate and a floating gate and an underlying channel;

creating a depletion region under said floating gate; and

supplying substrate electrons to said channel along a path from source to channel free of intervening N-type doped regions in the line from source to channel substantially parallel to said substrate surface.

19. The method of claim 18 including the step of extending said depletion region laterally toward said substrate electron source beyond the region under said floating gate.

20. The method of claim 18 including the step of causing said substrate electrons to be injected onto said floating gate.

21. The method of claim 18 including the step of creating a conduction path between said source and said floating gate by extending said depletion region into the space between said floating gate and said source.

22. A memory cell comprising:

a source and drain spaced from one another in a substrate;

a floating gate and a transistor gate arranged over said substrate between said source and said drain; and

a device for forming an inverted region in said substrate between said transistor gate and said floating gate.

23. The cell of claim 22 wherein said device is a capacitor.

24. The cell of claim 22 wherein said device includes a portion of said control gate situated over said substrate and extending beyond said floating gate.

25. The cell of claim 22 wherein said transistor gate forms part of a select transistor.

26. The cell of claim 22 wherein said gates control conduction between the same source and drain.

27. A method of programming a memory comprising:

supplying substrate electrons from a source spaced from a sense transistor having a control gate and a floating gate and an underlying channel; and

supplying the substrate electrons to the channel along a path from the source to the channel which is wider than the width of the cell.

28. The method of claim 27 including turning off at least one adjacent column during programming.

29. The method of claim 28 including turning off both adjacent columns during programming.

30. A memory cell comprising:

a source and drain spaced from one another in the substrate;

a floating gate and a transistor gate arranged over said substrate between the source and the drain; and

a source extending along the width direction of the cell and extending beyond the width of the cell in both directions such that said source can supply electrons to

said floating gate from a region which is greater than the width of the cell.

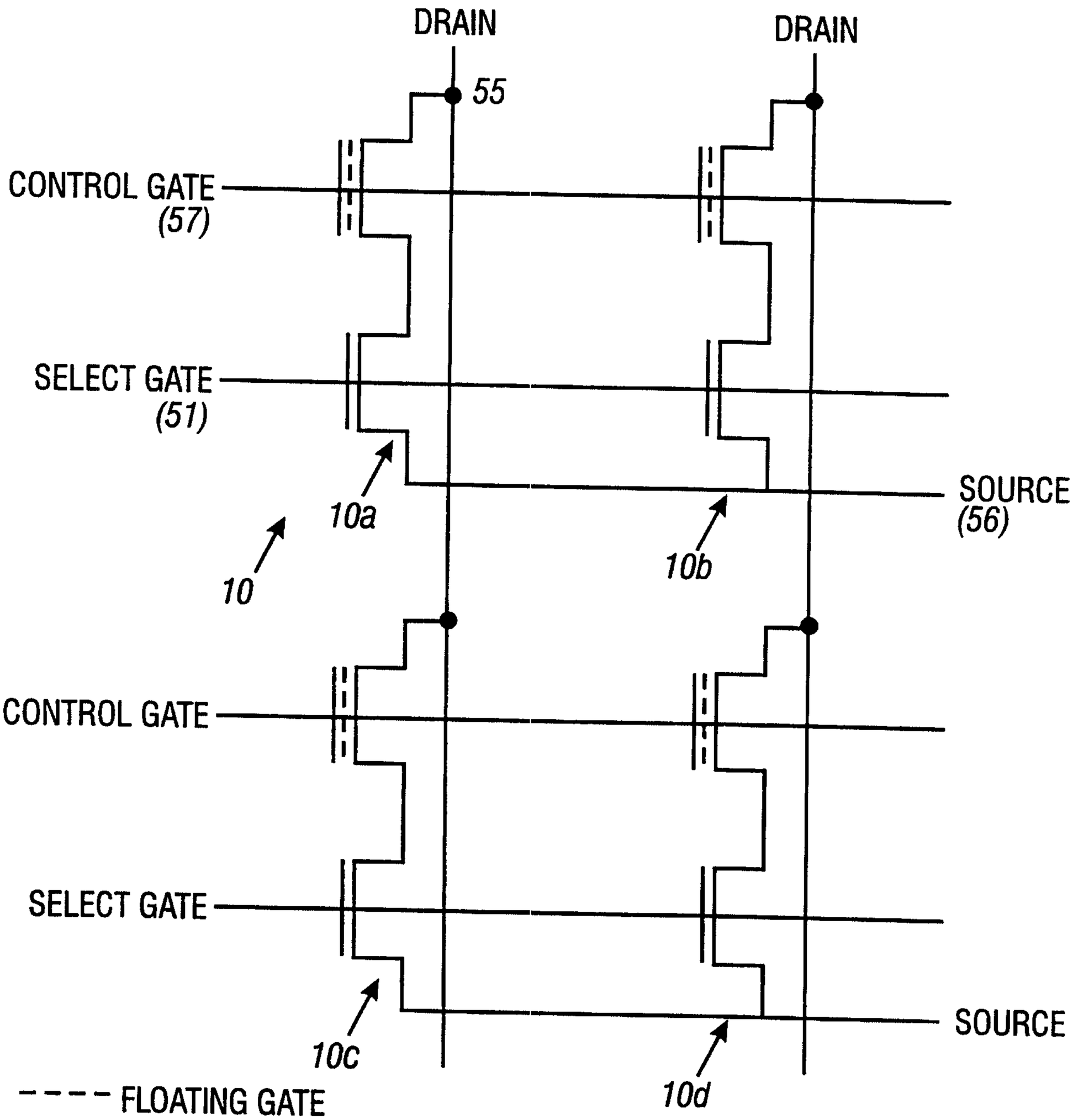

DRAIN
55
CONTROL GATE
(57)
SELECT GATE
(51)
10a
10
10b
SOURCE
(56)
CONTROL GATE
SELECT GATE
10c
10d
SOURCE
FLOATING GATE

FIG. 1

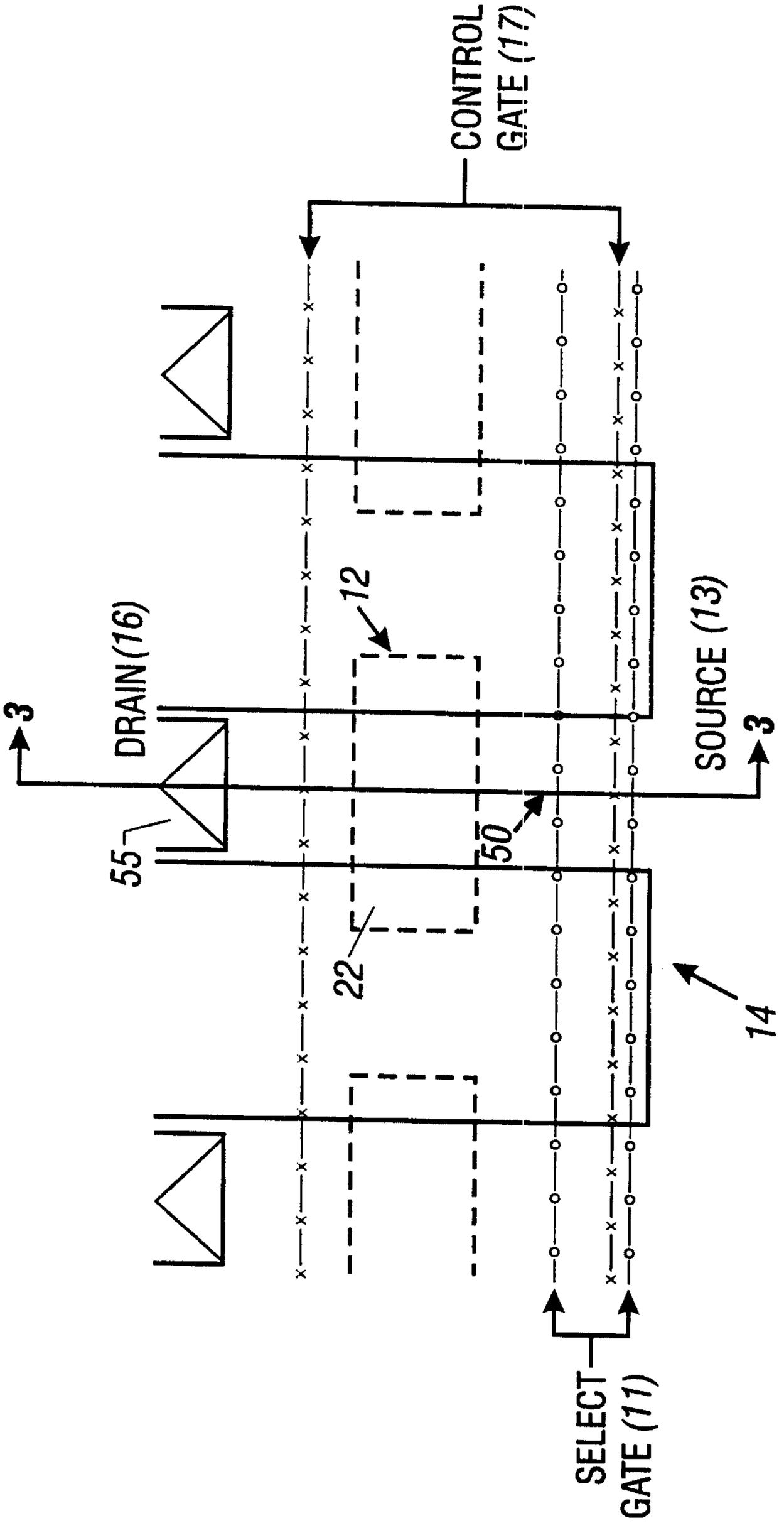

CONTROL GATE (17)
POLY 1
POLY 2
3
DRAIN (16)
55
12
22
50
SOURCE (13)
3
14
SELECT GATE (11)

FIG. 2

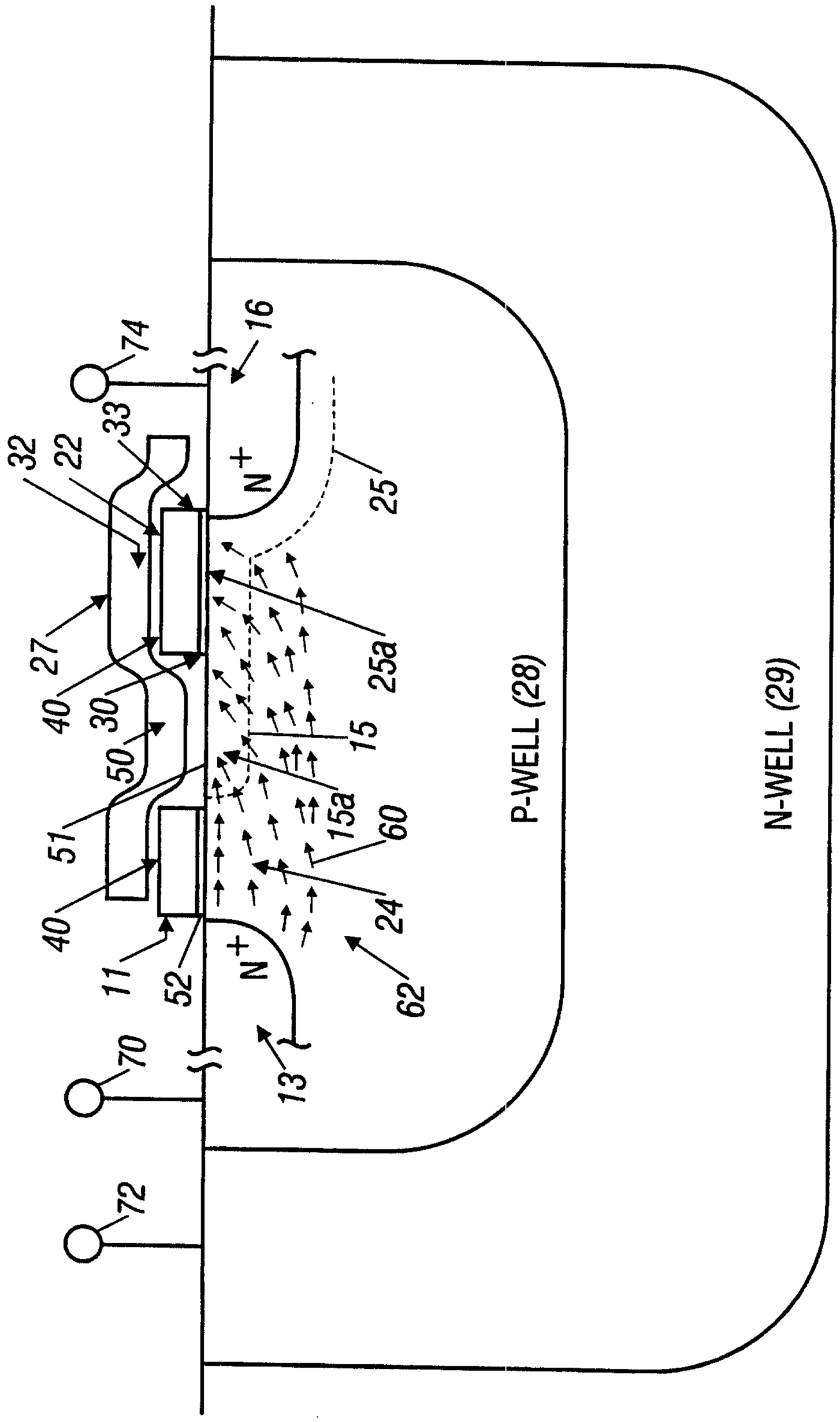

74
16
33
32
22
N+
25
27
40
30
50
51
15
25a
15a
60
40
24
11
52
N+
62
70
13
P-WELL (28)
N-WELL (29)
P-type substrate (38)
72

FIG. 3

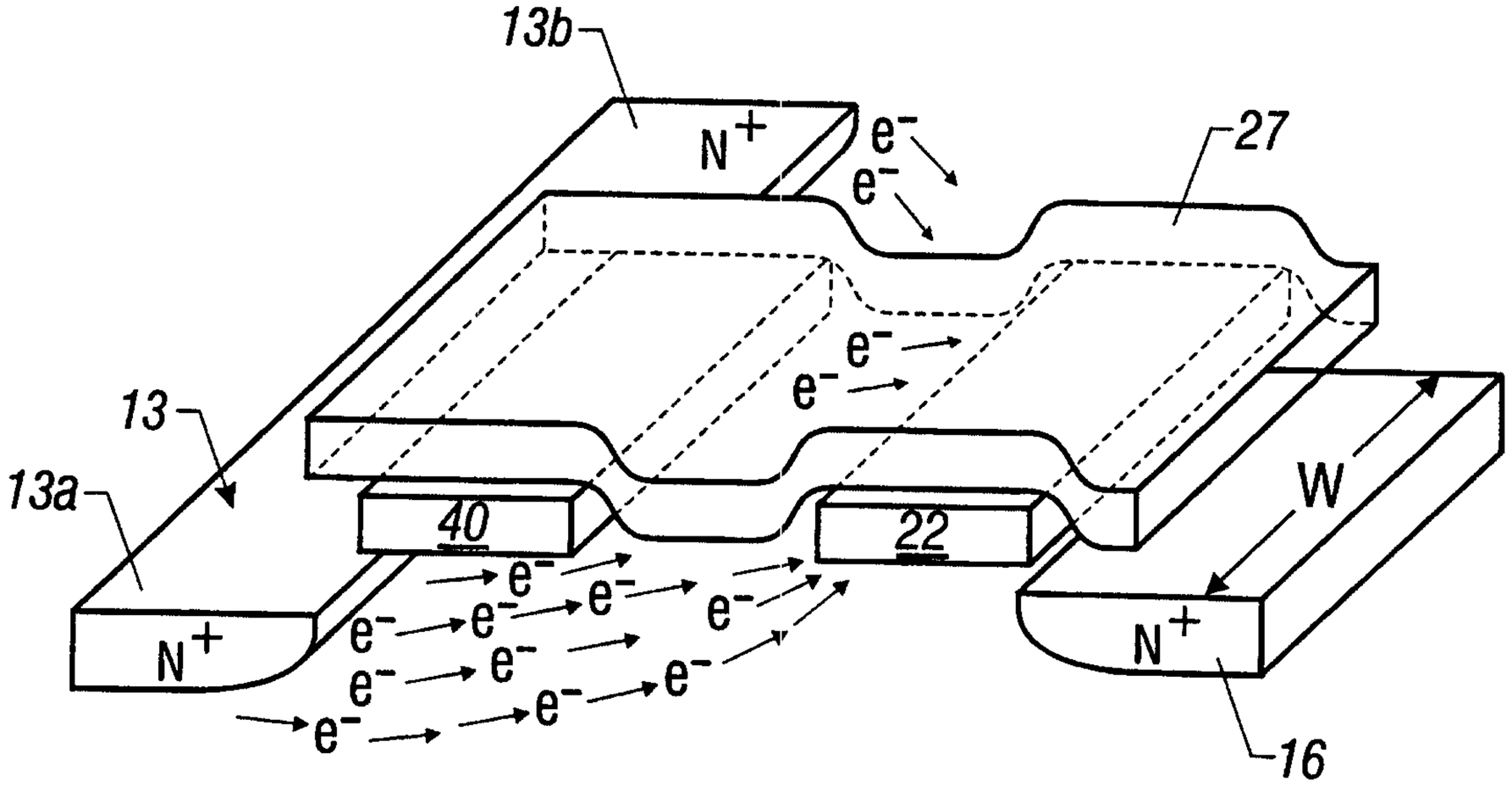


FIG. 4