



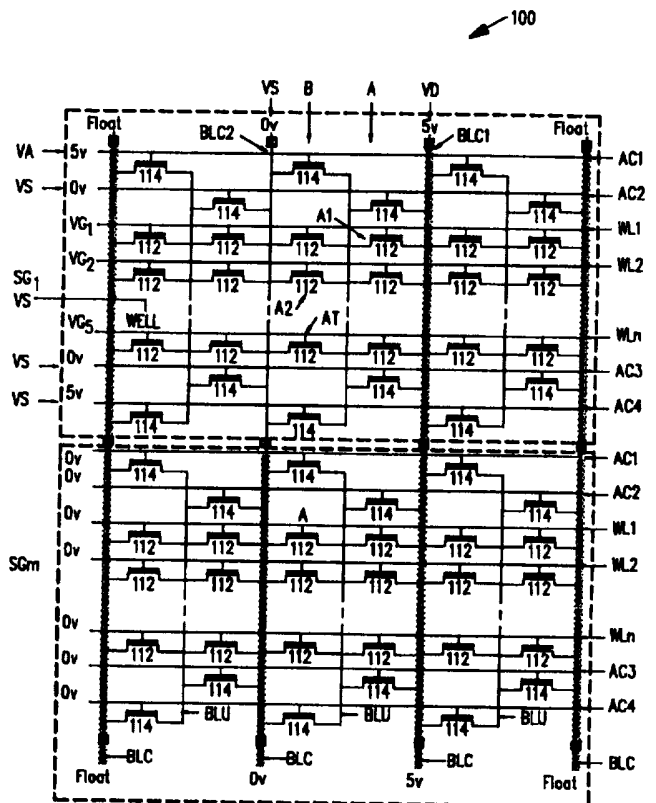
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification ⁶ : G11C 11/56</p>	<p>A1</p>	<p>(11) International Publication Number: WO 96/31883 (43) International Publication Date: 10 October 1996 (10.10.96)</p>
<p>(21) International Application Number: PCT/US96/04843 (22) International Filing Date: 8 April 1996 (08.04.96) (30) Priority Data: 08/417,938 6 April 1995 (06.04.95) US (71) Applicant: NATIONAL SEMICONDUCTOR CORPORATION [US/US]; 2900 Semiconductor Drive, M/S D3-589, Santa Clara, CA 95052 (US). (72) Inventors: BERGEMONT, Albert; 4149 Byron Street, Palo Alto, CA 94306 (US). CHI, Min-hwa; 320 Edlee Avenue, Palo Alto, CA 94306 (US). (74) Agent: CONSER, Eugene; National Semiconductor Corporation, 2900 Semiconductor Drive, M/S D3-579, Santa Clara, CA 95052 (US).</p>		<p>(81) Designated States: DE, KR, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: A METHOD FOR PROGRAMMING AN AMG EPROM OR FLASH MEMORY WHEN CELLS OF THE ARRAY ARE FORMED TO STORE MULTIPLE BITS OF DATA

(57) Abstract

Multiple logic levels can be simultaneously programmed into any combination of memory cells in a column of an alternate-metal virtual-ground (AMG) EPROM or flash memory array by applying one of a corresponding number of programming voltages to the word lines that correspond with the cells to be programmed. In the present invention, the memory cells in the array form a punchthrough current during programming which, in turn, leads to the formation of an increased number of substrate hot electrons. By utilizing the substrate hot electrons formed from the punchthrough current in addition to the channel hot electrons, much lower control gate voltages can be utilized during programming.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

A METHOD FOR PROGRAMMING
AN AMG EPROM OR FLASH MEMORY
WHEN CELLS OF THE ARRAY ARE FORMED
TO STORE MULTIPLE BITS OF DATA

5

BACKGROUND OF THE INVENTION

1. Field of the Invention.

The present invention relates to programming an alternate-metal virtual-ground (AMG) EPROM or flash memory and, more particularly, to a method for programming these memories when cells of the array are formed to store multiple bits of data, i.e., more than a logic "1" and a logic "0".

10

2. Description of the Related Art.

An alternate-metal virtual-ground (AMG) array architecture is a type of non-volatile memory architecture that is characterized by metal which only contacts every other bit line of the array. The AMG array architecture can be utilized with both U-V erasable EPROMs as well as flash memories.

15

FIG. 1 shows a portion of a conventional AMG array 10. As shown in FIG. 1, array 10 includes a plurality of contacted bit lines BLC, a plurality of non-contacted bit lines BLU formed so that one non-contacted bit line BLU is positioned between each pair of contacted bit lines BLC, a plurality of memory cells 12, and a plurality of access transistors 14. As shown, the contacted bit lines BLC directly contact a metal line ML, while the non-contacted bit lines BLU contact a metal line ML via one of the access transistors 14.

20

Memory cells 12 are arranged in columns and rows so that a predetermined number of cells 12 are formed between each pair of contacted and non-contacted bit lines BLC and BLU. Access transistors 14, on the other hand, are arranged in columns and rows so that in each row only one transistor 14 is formed between every other pair of contacted and non-contacted bit lines BLC and BLU.

25

Array 10 further includes a series of word lines WL1-WLn which are formed so that one word line WL is formed over each of the memory cells 12 in a row of memory cells. As is well known, the portion of the word line 18 which is formed over each memory cell 12 functions as the control gate of that memory cell. Similarly, the access transistors 14 in a row of access transistors 14 share one of four access lines AC1-AC4.

30

A cell in array 10 is conventionally programmed to store one bit of data by selecting the cell to be programmed, and then applying a programming voltage to the word line that corresponds to the cell to be programmed. For example, to program cell A, the contacted bit line BLC that adjoins cell A is held at an intermediate voltage Vd (approximately 5-7V), while the contacted bit line BLC positioned on the opposite side of cell A is held at ground Vss. The remaining contacted bit lines BLC are allowed to float. In addition, access lines AC2 and AC3 are biased to the supply voltage Vcc (approximately 5V), while access lines AC1 and AC4 are held at ground Vss. This, in turn, pulls the non-contacted bit line BLU that contacts cell A down to ground Vss.

35

40

The programming voltage Vpp (approximately 12V) is then applied to word line WL1, while the remaining word lines WL2-WLn are grounded. These bias conditions result in current flow as shown by the arrow in FIG. 1, which results in electron injection from the drain of cell A to the floating gate of cell A, thus programming cell A.

One drawback to programming memory cells 12 as described above is that only one bit of data can be programmed into a cell, and only one cell in a column of cells can be programmed at any one time. Although it would appear that multiple cells in a column could be simultaneously programmed by applying the programming voltage Vpp to the word lines WL2-WLn that correspond to each cell in the

column to be programmed, the high current requirements of each cell during programming (approximately 400mA) preclude this. Thus, there is a need for a method of programming multiple cells in a column of an AMG array at the same time.

SUMMARY OF THE INVENTION

5 The present invention provides a method for simultaneously programming any combination of memory cells in a column of cells in an alternate-metal virtual-ground (AMG) EPROM or flash memory to each store one of three or more threshold voltages, i.e., logic levels. As a result, rather than storing either a logic "1" or a logic "0", any combination of memory cells in a column can be simultaneously programmed to store, for example, a logic "0-0", a "0-1", a "1-0", or a "1-1".

10 In the present invention, the array, which is formed in a well, includes a plurality of metal contacted bit lines, and a plurality of non-contacted bit lines which are formed so that one non-contacted bit line is formed between each adjacent pair of metal contacted bit lines. The array also includes a plurality of memory cells which are formed in columns and rows so that each column of memory cells is formed between and contacts adjacent pairs of metal contacted and non-contacted bit lines. A plurality of access transistors are also formed in columns and rows so that one access transistor is formed at each end of each column of memory cells. In addition, each row of access transistors is formed so that one access transistor is formed between and contacts every other pair of metal contacted and non-contacted bit lines. The array further includes a plurality of access lines and a plurality of word lines. The access lines are formed so that each row of access transistors is contacted by a corresponding access line, while the of word lines are formed so that each row of memory cells is contacted by a corresponding word line. In addition, each cell is formed to produce a punchthrough current during programming.

15 A method for simultaneously programming any combination of memory cells in a column to each store one of three or more threshold voltages includes the step of selecting a programming voltage from three or more programming voltages for each of a plurality of cells in a column that are to be programmed. In the present invention, the three or more programming voltages correspond to the three or more threshold voltages. In addition, each cell in the column that is to be programmed has a corresponding word line.

20 The method continues with the step of applying a first voltage to a first metal contacted bit line. The first metal contacted bit line contacts each memory cell and each access transistor in a first column of a pair of adjacent columns where the first column includes the cells to be programmed. In addition to the first voltage, a second voltage, which is less than the first voltage, is applied to a second metal contacted bit line. The second metal contacted bit line contacts each memory cell and access transistor in a second column of the pair of adjacent columns. The second voltage is also applied to the well.

25 The method further continues with the step of applying a third voltage to a pair of access lines where each access line of the pair of access lines contacts an access transistor located in the second column of the pair of adjacent columns. Each cell is programmed to store one of the multiple logic levels by applying the programming voltages selected for the cells to be programmed to the word lines that correspond to the cells to be programmed.

30 A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description and accompanying drawings which set forth an illustrative embodiment in which the principals of the invention are utilized.

BRIEF DESCRIPTION OF THE DRAWINGS

35 FIG. 1 is a plan view illustrating a portion of a conventional AMG array 10.

FIG. 2 is a plan view illustrating a portion of an alternate-metal virtual-ground (AMG) array 100 in accordance with the present invention.

FIG. 3 is a cross-sectional view illustrating the structure of one of the memory cells 112 from array 100.

5 FIG. 4 is a graphical representation illustrating a cell programming characterization curve.

FIG. 5 is a graphical representation illustrating the amount of negative charge injected onto the floating gate for initial and final voltages V_1 and V_2 .

FIG. 6 is a graphical representation illustrating the amount of negative charge injected onto the floating gate for initial voltages V_{00} - V_{11} and final voltage V_4 .

10 FIG. 7 is a graphical representation illustrating a series of experimental results.

FIG. 8 is a plan view illustrating segment select transistors in a portion of an alternate-metal virtual-ground (AMG) array 100 in accordance with the present invention.

DETAILED DESCRIPTION

15 FIG. 2 shows a portion of an alternate-metal virtual-ground (AMG) array 100 in accordance with the present invention. As shown in FIG. 2, array 100, which can be utilized with both U-V erasable EPROMs as well as flash memories, is divided into a series of segments SG1-SG_m which each include a plurality of metal contacted bit lines BLC, and a plurality of non-contacted bit lines BLU which are formed so that one non-contacted bit line BLU is formed between each adjacent pair of metal contacted bit lines BLC.

20 Each segment SG also includes a plurality of memory cells 112 and a plurality of access transistors 114. The memory cells 112 are formed in columns and rows so that each column of memory cells 112 is formed between and contacts adjacent pairs of metal contacted and non-contacted bit lines BLC and BLU. The access transistors 114 are formed in columns and rows so that one access transistor 114 is formed at each end of each column of memory cells 112 in each segment SG. In addition, each row of access transistors 114 is formed so that one access transistor 114 is formed between and contacts every other pair of metal contacted and non-contacted bit lines BLC and BLU.

25 Further, each of the memory cells 112 in a row of cells 112 in a segment SG share one of a series of word lines WL1-WL_n while each of the access transistors 114 in a row of transistors 114 share one of a series of access lines AC1-AC₄. As is well known, the portion of the word line WL formed over each memory cell 112 functions as the control gate of that memory cell while the portion of the access line AC formed over each transistor 114 functions as the gate of that transistor.

30 FIG. 3 shows the structure of one of the memory cells 112 from array 100. As shown in FIG. 3, cell 112 is formed in a p-type well 120 which, in turn, is formed in an n-type substrate 118. Memory cell 112 includes an n-type source region 122, an n-type drain region 124, and a channel region 126 formed between the source and drain regions 122 and 124.

In accordance with the present invention, cell 112 is formed to produce a punchthrough current during programming. The channel lengths and doping levels, as well as the bias conditions, which are required to produce a punchthrough current through channel region 126 are well known in the art.

40 With 0.6 micron technology, for example, cell 112 preferably utilizes a channel length of 0.5-0.7 microns and a doping concentration of $1\text{-}2 \times 10^{17}$ p-type atoms. When smaller micron technologies are utilized, the channel length and the doping concentration can be reduced accordingly. Thus, for example, with 0.3 micron technology, cell 112 preferably utilizes a channel length of 0.2-0.4 microns and a doping concentration of 5×10^{17} p-type atoms.

45 As further shown in FIG. 3, memory cell 112 also includes a first insulation layer 130 formed over channel region 126, a floating gate 132 formed over insulation layer 130, a second insulation layer 134

formed over floating gate 132, and a control gate 136 (a portion of word line WL) formed over insulation layer 134.

In accordance with the present invention, memory cell 112 is programmed to store one of three or more logic levels by maintaining an equilibrium across the source-to-well junction, reverse-biasing the drain-to-well junction, and applying one of a corresponding three or more programming voltages to control gate 136 during programming.

With 0.6 micron technology, memory cell 112 preferably utilizes a drain voltage that is 4-7 volts greater than the well voltage. In addition, although other voltages can be utilized, well 120 and source 122 are preferably grounded.

As above, when smaller micron technologies are utilized, the preferred conditions can be reduced accordingly. Thus, for example, with 0.3 micron technology, cell 112 preferably utilizes a drain voltage which is 2-4 volts greater than the well voltage.

In operation, when one of the programming voltages is applied to control gate 136, a positive potential is induced on floating gate 132 which, in turn, attracts electrons from the doped p-type atoms in channel region 126 to the surface of well 120 to form a channel 140. This potential also repels holes from the doped impurity atoms and forms a depletion region 142.

When the source and drain voltages are applied, an electric field is established between the source and drain regions 122 and 124. The electric field, as with conventional programming, accelerates the electrons in channel 140 which, in turn, have ionizing collisions that form channel hot electrons. The positive potential on floating gate 132 attracts these channel hot electrons which penetrate insulation layer 130 and begin accumulating on floating gate 132.

In addition to the formation of channel hot electrons, the present invention utilizes the punchthrough current to form substrate hot electrons which also collect on floating gate 132. In operation, due to the relatively short channel length, i.e., 0.5 microns in a 0.6 micron process, the electric field reduces the potential energy barrier at the source-to-well junction. The reduced potential energy barrier allows more majority carriers in source 122 to overcome the barrier which, in turn, produces the punchthrough current across the source-to-well junction.

As the electrons associated with the punchthrough current near drain 124, the electric field accelerates the electrons which, in turn, also have ionizing collisions that form substrate hot electrons. The positive potential on floating gate 132 also attracts these substrate hot electrons which penetrate insulation layer 130 and begin accumulating on floating gate 132.

However, unlike conventional floating-gate programming, the flow of electrons generated by the punchthrough current does not depend on the existence of a channel or the relative positive charge on floating gate 132. As a result, the electrons associated with the punchthrough current continue to accumulate on floating gate 132 after channel 130 has been turned off.

Thus, cell 112 is programmed by utilizing both channel hot electrons and substrate hot electrons to change the potential on floating gate 132. As a result of utilizing two sources of hot electrons, a significantly lower control gate voltage can be used during programming than is conventionally used to program a cell because fewer channel hot electrons need to be attracted to floating gate 132.

FIG. 4 shows a graphical representation that illustrates a cell programming characterization curve. As shown by lines L_1 and L_2 in FIG. 4, the voltage (V) on the floating gate influences the number of hot electrons (I) that are injected onto the floating gate.

Conventionally, the primary consideration in programming memory cells is the time required to place a defined amount of negative charge on the floating gate of the cell. As a result, the typical memory cell is designed to utilize an initial floating gate voltage V_1 and a final floating gate voltage V_2 that are positioned on opposite sides of the peak of the curve shown in FIG. 4, thereby taking advantage of the maximum injection of hot electrons onto the floating gate. As described, the initial floating gate voltage V_1 represents the voltage capacitively coupled to the floating gate from the control gate, while

the final floating gate voltage V_2 represents the initial voltage V_1 reduced by the accumulated negative charge.

FIG. 5 shows a graphical representation that illustrates the amount of negative charge injected onto the floating gate for initial and final voltages V_1 and V_2 . As shown in FIG. 5, the amount of charge injected on the floating gate can be determined by integrating under the curve from the initial voltage V_1 at time t_0 to the final voltage V_2 at time t_1 .

More importantly, however, FIG. 5 illustrates that any variation in the timing will cause a greater or lesser amount of negative charge to be injected onto the floating gate. Thus, for example, if the programming is terminated at time t_2 rather than time t_1 , a greater amount of charge will be injected.

With conventional programming, this additional (or lesser) amount of negative charge does not present any problems because the cell is only being programmed to one of two logic levels. Thus, as long as the cell is programmed to have a minimum amount of charge, any additional charge is acceptable.

However, with multi-level programming, the accumulation of additional negative charge makes it difficult to determine which logic level is represented by the charge. Thus, to insure that the charge is within a defined range, the timing must be precisely controlled. This timing, however, is very difficult to control.

The present invention achieves multiple levels of injected charge by utilizing one of a plurality of initial voltages. Since the initial voltages are defined by the voltage capacitively coupled to the floating gate from the control gate, the initial voltages are selected by selecting one of a plurality of control voltages.

For example, referring again to FIG. 4, voltage V_{00} , V_{01} , V_{10} , or V_{11} may be selected as the initial voltage by selecting a corresponding control gate voltage. As further shown in FIG. 4, the present invention utilizes voltage V_4 as the final voltage. The significance of utilizing voltage V_4 as the final voltage can be seen in FIG. 6.

FIG. 6 shows a graphical representation that illustrates the amount of negative charge injected onto the floating gate for initial voltages V_{00} - V_{11} and final voltage V_4 . As shown in FIG. 6, the amount of negative charge injected on the floating gate can be determined by integrating under the curve from each of the initial voltages V_{00} - V_{11} at time t_0 to the final voltage V_4 at time t_4 .

FIG. 6, however, also illustrates that because the injected charge converges towards zero for each initial voltage V_{00} - V_{11} , the injected charge is largely insensitive to timing variations. Thus, the floating gate can still have one of a plurality of discrete levels of injected charge if programming is terminated anytime between time t_4 and t_5 because the amount of additional charge during this time is so small.

One drawback of the approach, as described, is that the magnitude of the injected current drops as the voltage on the floating gate approaches the final voltage V_4 (see FIG. 4). As a result, it takes a greater amount of time to program the cell.

In accordance with the present invention, however, line L1 of FIG. 4 can be altered, as shown by line L2, by increasing the formation of substrate hot electrons as described above. Thus, although the time required to program a memory cell in accordance with the present invention remains longer than conventional programming, the formation of substrate hot electrons substantially narrows the time difference.

FIG. 7 shows a graphical representation that illustrates a series of experimental results. As shown in FIG. 7, when one volt was applied to the control gate, the charge on the floating gate converged to a threshold voltage of approximately three volts from an initial threshold voltage of 1.5 volts within 50 milliseconds or less. Similarly, when two, three, and four volts were applied to the control gate, the charge on the floating gate converged to threshold voltages of approximately four, five, and six volts, respectively, within 50 milliseconds or less. Although an initial threshold voltage of 1.5 volts was utilized in the above experiment, any initial threshold voltage after erase may be utilized.

Since the charge on the floating gate converges to a stable value which corresponds to a defined threshold voltage within 50 milliseconds or less, a single floating gate memory cell can be programmed to have one of a plurality of threshold voltages by applying the corresponding voltage to the control gate during programming. As a result, a single floating gate memory cell can be utilized to store two or more bits of data.

For example, a 0-0 could be represented by a threshold voltage of 3 volts, while a 0-1 could be represented by a threshold voltage of 4 volts. Similarly, a 1-0 could be represented by a threshold voltage of 5 volts, while a 1-1 could be represented by a threshold voltage of 6 volts.

As further shown in FIG. 7, the experimental results also show that changes in the control gate voltage are linearly related to changes in the threshold voltage, i.e., a one volt increase in the control gate voltage increases the threshold voltage by one volt. As a result, memory cell 112 is not limited to representing two bits, but can represent any number of bits depending on the sensitivity of the current sense detectors utilized to discriminate one threshold voltage from another. Furthermore, even a continuous analog level can be stored in a cell as a threshold voltage.

For example, a 0-0-0 could be represented by a threshold voltage of 3 volts, while a 0-0-1 could be represented by a threshold voltage of 3.5 volts. Similarly, a 0-0-0-0 could be represented by a threshold voltage of 3 volts, while a 0-0-0-1 could be represented by a threshold voltage of 3.25 volts.

Thus, as described above, when an equilibrium condition is maintained across the source-to-well junction, and the drain-to-well junction is reverse-biased, cell 112 can be programmed to store three or more logic levels by applying one of a corresponding three or more programming voltages to the control gate (word line).

The equilibrium and reverse-bias conditions for the cells to be programmed are established in array 100 by identifying the column of cells to be programmed, and applying a voltage V_D to the metal contacted bit line BLC that contacts the memory cells 112 and access transistors 114 in that column.

For example, referring again to FIG. 2, if any combination of cells in column A are to be programmed, voltage V_D is applied to metal contacted bit line BLC1. The application of voltage V_D to metal contacted bit line BLC1 is equivalent to applying a voltage to the drain of an individual cell as described above.

In addition to voltage V_D , a voltage V_S , which is less than voltage V_D , is applied to the metal contacted bit line BLC that contacts the memory cells 112 and access transistors 114 in a column adjacent to the column containing the cells to be programmed. The remaining metal contacted bit lines BLC are allowed to float.

Thus, continuing the above example, since column B is the only column adjacent to column A which contacts one of the metal contacted bit lines BLC, voltage V_S is applied to metal contacted bit line BLC2. The application of voltage V_S to metal contacted bit line BLC2 is, as a result of the access transistor, approximately equivalent to applying a voltage to the source of an individual cell as described above. The basic bias conditions are completely established by applying voltage V_S to the well.

The preferred voltages depend on the particular technology used to form the array. With 0.6 micron technology, voltage V_D is preferably 4-7 volts greater than voltage V_S . In addition, voltage V_S is preferably ground. When smaller micron technologies are utilized, the preferred conditions can be reduced accordingly. Thus, for example, with 0.3 micron technology, voltage V_D is preferably 2-4 volts greater than voltage V_S .

The column of cells to be programmed is selected by applying a voltage V_A to the access lines AC that contact the access transistors 114 that are in the adjacent column, i.e., column B. The remaining access lines AC are held at ground. For example, as shown in FIG. 2, the cells 112 in column A are selected by applying voltage V_A to access lines AC1 and AC4, while access lines AC2 and AC3 are held at ground. Since voltage V_A is used to turn on the appropriate access transistors 114, any voltage which turns on the transistors can be used.

Programming is accomplished by selecting a programming voltage VG from three or more programming voltages VG1-VGs for each cell to be programmed where the three or more programming voltages correspond to three or more threshold voltages, and by applying the programming voltages VG1-VGs to the word lines WL that correspond to the cells 112 to be programmed for a predetermined time.

5 For example, if cells A1, A2, and At are to be programmed with voltages VG1, VG2, and VGs, respectively, voltages VG1, VG2, and VGs are simultaneously applied to word lines WL1, WL2, and WLn, respectively.

The preferred voltages for programming voltages VG1-VGs depend on the number of logic levels to be programmed and generally fall within 0-5 volts. Although negative voltages can also be used, these
10 voltages are not as desirable because of the additional circuitry needed to generate negative voltages.

In contrast to conventional programming, where the current flowing through a cell is approximately 400mA, the current flowing through a cell during programming under the present invention is approximately 10 μ A. As a result, as shown in FIG. 2, cells A1-At, or any combination thereof in column A of segment SG1, can be programmed at the same time to each have any one of the
15 three or more logic levels by applying the corresponding programming voltages VG1-VGs to corresponding word lines WL1-WLn.

As further shown in FIG. 2, all of the access lines AC1-AC4 and word lines WL1-WLn in the remaining segments SG are held at ground. However, the same column in multiple segments can also be programmed at the same time.

20 One of the principle advantages of being able to program an entire column at the same time is that each cell in an entire segment can be programmed to store multiple bits of data in less time than is required for each cell to be conventionally programmed to store one bit of data. For example, in a segment that is eight bits wide by 32 bytes long, it takes one programming cycle to conventionally program each cell in a segment for a total of 256 (8x32) programming cycles.

25 On the other hand, since an entire column can be programmed at the same time, only eight programming cycles are required to program a segment in accordance with the present invention. Thus, in applications where an entire segment is frequently rewritten, the present invention provides a faster overall programming speed than is conventionally available.

In the array illustrated in FIG. 2, the entire array is formed in a single well. In an alternative embodiment, each segment SG of array 100 can be formed in a separate well. Further, as shown in FIG. 8, whether formed in a single well or separate wells, each segment can be individually accessed via segment select transistors.

30 As a result, each segment SG can be operated on independently. Thus, for example, one column in segment SG1 can be programmed at the same time that a separate column in segment SG2 is being programmed which both can occur at the same time that a row of data is being read from segment SG3.

35 As stated above, only approximately 10 μ A is required to program a single cell. As a result, multiple cells in a single row can be programmed at the same time if the cells are being programmed to the same logic level.

40 As shown in FIG. 2, cell C1 can not be conventionally programmed at the same time that cell A1 is being programmed due to the high programming currents that are required, i.e., 400mA. In the present invention, however, cells A1 and C1 can be programmed at the same time, if both cells are being programmed to the same logic level, by also grounding bit line BLC3.

45 As stated above, one advantage of the present invention is that the memory cells 112 of array 100 can be programmed to store multiple levels by utilizing a programming voltage that is considerably less than the programming voltage typically used, i.e., less than five volts in the present invention compared to the approximately 12 volts that are typically used. In addition to providing a substantial power savings for low power applications, such as notebook computers, the present invention eliminates the need to form charge pumps on memory chips to produce the programming voltage, i.e., the 12 volts.

As is well known, charge pumps can consume a significant area, i.e., up to 30% of the total die area of a memory chip. Thus, by eliminating the need for charge pumps, the present invention significantly reduces the area required for a memory cell, and therefore the cost of a memory.

5 The elimination of high programming voltages also leads to an increase in the density of array 100 because less isolation is required between both memory cells and the peripheral circuitry. As a result, the present invention substantially reduces the size of array 100.

10 In another embodiment of the present invention, array 100 can be configured to support both conventional as well as multilevel programming by utilizing a high programming voltage. In this embodiment, the high programming voltage can be provided either externally or on-chip, i.e., via the charge pumps.

As a result, a single memory chip can utilize a first portion of the array to store one bit per cell when the data must be stored quickly, and a second portion of the array to store multiple bits per cell when more time can be taken to store the data.

15 It should be understood that various alternatives to the embodiment of the invention described herein may be employed in practicing the invention. For example, although the present invention has been described in terms of an n-channel, floating-gate memory cell formed in a p-well, the present invention equally applies to a p-channel, floating-gate memory cell formed in an n-well.

Thus, it is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

WHAT IS CLAIMED IS:

1. A method for simultaneously programming any combination of memory cells in a column of an alternate-metal virtual-ground (AMG) electrically-programmable read-only-memory (EPROM) or flash memory array to each store one of three or more threshold voltages,

5 the array having a plurality of metal contacted bit lines, a plurality of non-contacted bit lines formed so that one non-contacted bit line is formed between each adjacent pair of metal contacted bit lines, a plurality of memory cells formed in columns and rows so that each column of memory cells is formed between and contacts adjacent pairs of metal contacted and non-contacted bit lines, a plurality of access transistors formed in columns and rows so that one access transistor is formed at each end of each column of memory cells, each row of access transistors being formed so that one access transistor is formed between and contacts every other pair of metal contacted and non-contacted bit lines, a plurality of access lines formed so that each row of access transistors is contacted by a corresponding access line, and a plurality of word lines formed so that each row of memory cells is contacted by a corresponding word line, each cell being formed to produce a punchthrough current during programming, the array being formed in a well,

10

15 the method comprising the steps of:

selecting a programming voltage from three or more programming voltages for each of a plurality of cells to be programmed in a column, each cell to be programmed having a corresponding word line, the three or more programming voltages corresponding to said three or more threshold voltages;

20 applying a first voltage to a first metal contacted bit line, the first metal contacted bit line contacting each memory cell and each access transistor in a first column of a pair of adjacent columns, the first column including the cells to be programmed;

applying a second voltage to a second metal contacted bit line, the second voltage being less than the first voltage, the second metal contacted bit line contacting each memory cell and access transistor in a second column of the pair of adjacent columns;

25 applying the second voltage to the well;

applying a third voltage to a pair of access lines, each access line of said pair of access lines contacting an access transistor located in the second column of the pair of adjacent columns; and

30 applying the programming voltages selected for the cells to be programmed to the word lines that correspond to the cells to be programmed.

2. The method of claim 1 wherein the second voltage is equal to ground.

3. The method of claim 1 wherein the second voltage is equal to the third voltage.

4. The method of claim 1 wherein the second voltage is less than the first voltage by a difference voltage within the range of two to seven volts.

35 5. A method for simultaneously programming any combination of memory cells in a column of a plurality of segments of an alternate-metal virtual-ground (AMG) electrically-programmable read-only-memory (EPROM) or flash memory array to each store one of three or more threshold voltages,

40 at least two of the segments having a plurality of metal contacted bit lines, a plurality of non-contacted bit lines formed so that one non-contacted bit line is formed between each adjacent pair of metal contacted bit lines, a plurality of memory cells formed in columns and rows so that each column of memory cells is formed between and contacts adjacent pairs of metal contacted and non-contacted bit lines, a plurality of access transistors formed in columns and rows so that one access transistor is formed at each end of each column of memory cells, each row of access

transistors being formed so that one access transistor is formed between and contacts every other pair of metal contacted and non-contacted bit lines, a plurality of access lines formed so that each row of access transistors is contacted by a corresponding access line, and a plurality of word lines formed so that each row of memory cells is contacted by a corresponding word line, each cell being formed to produce a punchthrough current during programming, at least two of the segments being formed in a different well,

the method comprising the steps of:

selecting a programming voltage from three or more programming voltages for each of a plurality of cells to be programmed in a column of each segment to be programmed, each cell to be programmed having a corresponding word line, the three or more programming voltages corresponding to said three or more threshold voltages;

applying a first voltage to a first metal contacted bit line in each segment to be programmed, the first metal contacted bit line contacting each memory cell and each access transistor in a first column of a pair of adjacent columns, the first column including the cells to be programmed;

applying a second voltage to a second metal contacted bit line in each segment to be programmed, the second voltage being less than the first voltage, the second metal contacted bit line contacting each memory cell and access transistor in a second column of the pair of adjacent columns;

applying the second voltage to the well in each segment to be programmed;

applying a third voltage to a pair of access lines in each segment to be programmed, each access line of said pair of access lines contacting an access transistor located in the second column of the pair of adjacent columns; and

applying the programming voltages selected for the cells to be programmed to the word lines that correspond to the cells to be programmed in each segment to be programmed.

6. A method for simultaneously programming multiple memory cells in a row of an alternate-metal virtual-ground (AMG) electrically-programmable read-only-memory (EPROM) or flash memory array to each store one of three or more threshold voltages,

the array having a plurality of metal contacted bit lines, a plurality of non-contacted bit lines formed so that one non-contacted bit line is formed between each adjacent pair of metal contacted bit lines, a plurality of memory cells formed in columns and rows so that each column of memory cells is formed between and contacts adjacent pairs of metal contacted and non-contacted bit lines, a plurality of access transistors formed in columns and rows so that one access transistor is formed at each end of each column of memory cells, each row of access transistors being formed so that one access transistor is formed between and contacts every other pair of metal contacted and non-contacted bit lines, a plurality of access lines formed so that each row of access transistors is contacted by a corresponding access line, and a plurality of word lines formed so that each row of memory cells is contacted by a corresponding word line, each cell being formed to produce a punchthrough current during programming, the array being formed in a well,

the method comprising the steps of:

selecting a programming voltage from three or more programming voltages for the memory cells to be programmed, the cells to be programmed having a corresponding word line, the three or more programming voltages corresponding to said three or more threshold voltages;

applying a first voltage to a first metal contacted bit line, the first metal contacted bit line contacting each memory cell and each access transistor in a first column of a pair of adjacent columns, the first column including one of the cells to be programmed;

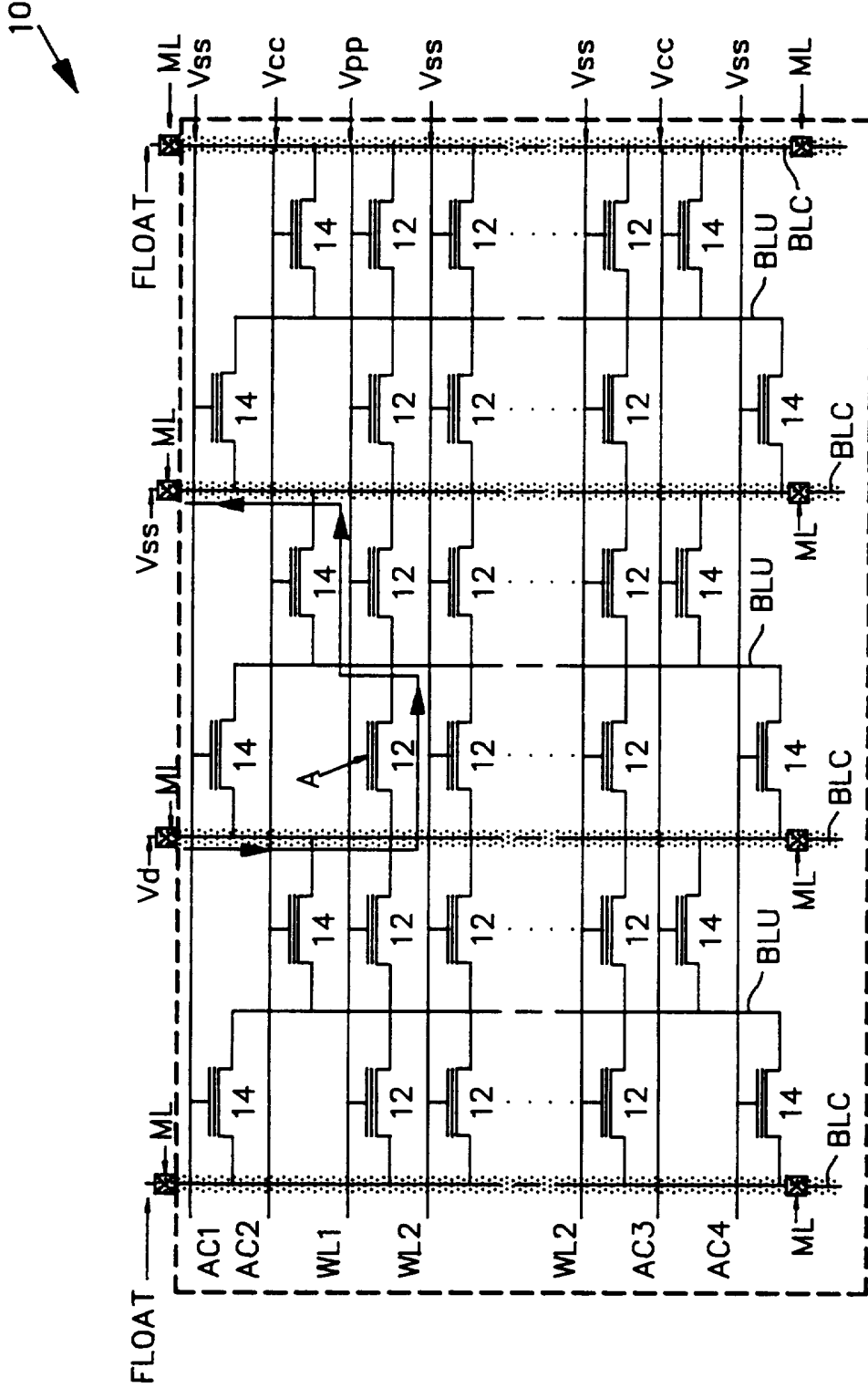
applying a second voltage to a second metal contacted bit line, the second voltage being less than the first voltage, the second metal contacted bit line contacting each memory cell and access transistor in a second column of the pair of adjacent columns;

applying the second voltage to a third metal contacted bit line, the third metal contacted bit line contacting each memory cell and access transistor in a third column one column removed from the first column;

applying the second voltage to the well;

5 applying a third voltage to a pair of access lines, each access line of said pair of access lines contacting an access transistor located in the second column of the pair of adjacent columns, and a fourth column positioned between the first and third columns; and

applying the programming voltage selected for the cells to be programmed to the word line that correspond to the cells to be programmed.



PRIOR ART

FIG. 1

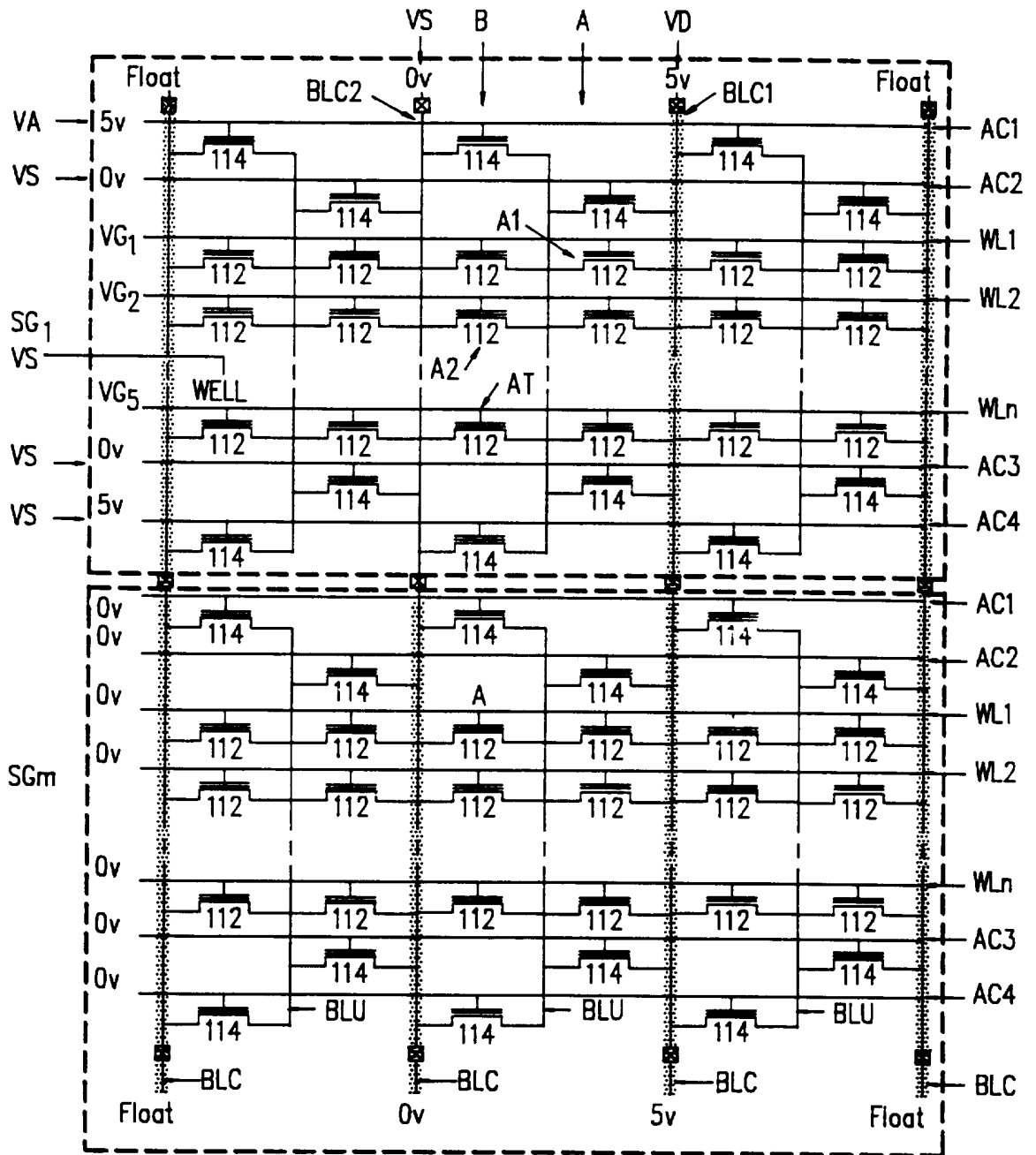


FIG. 2

SUBSTITUTE SHEET (RULE 26)

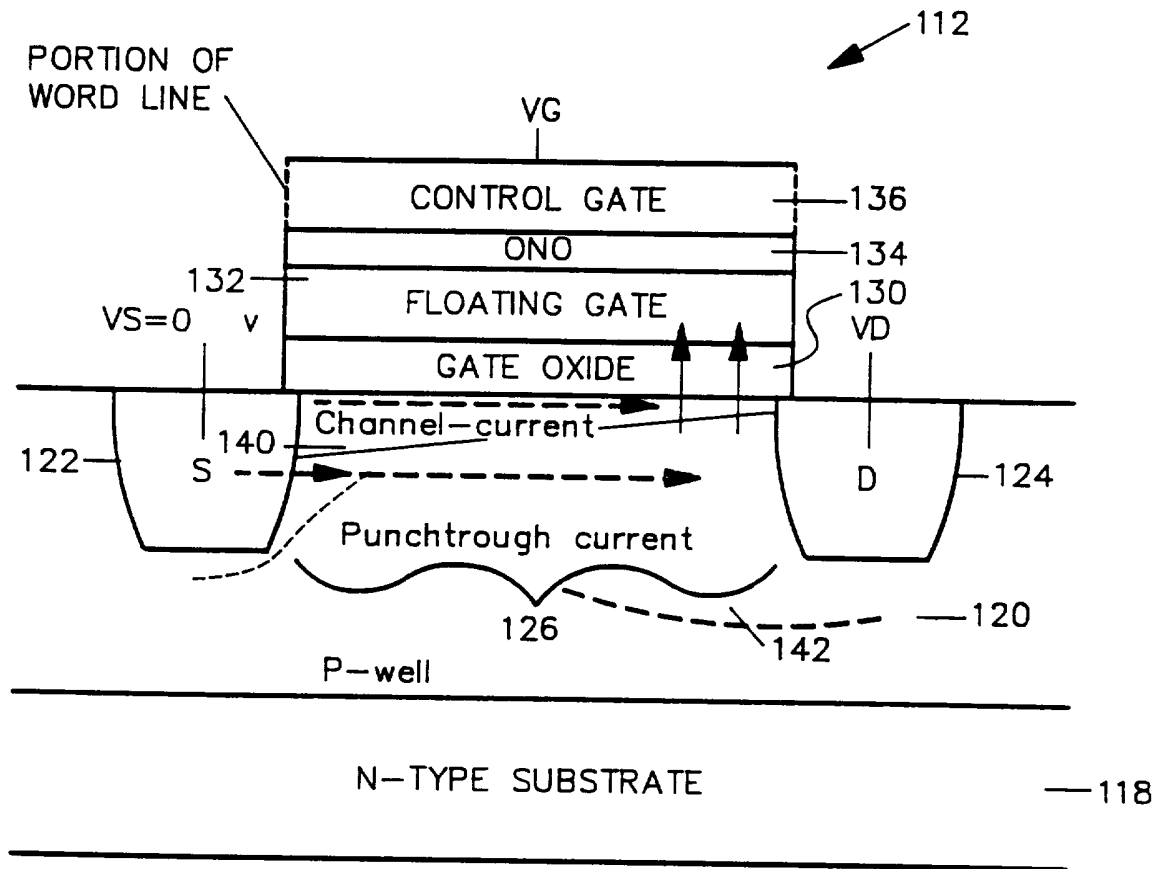


FIG. 3

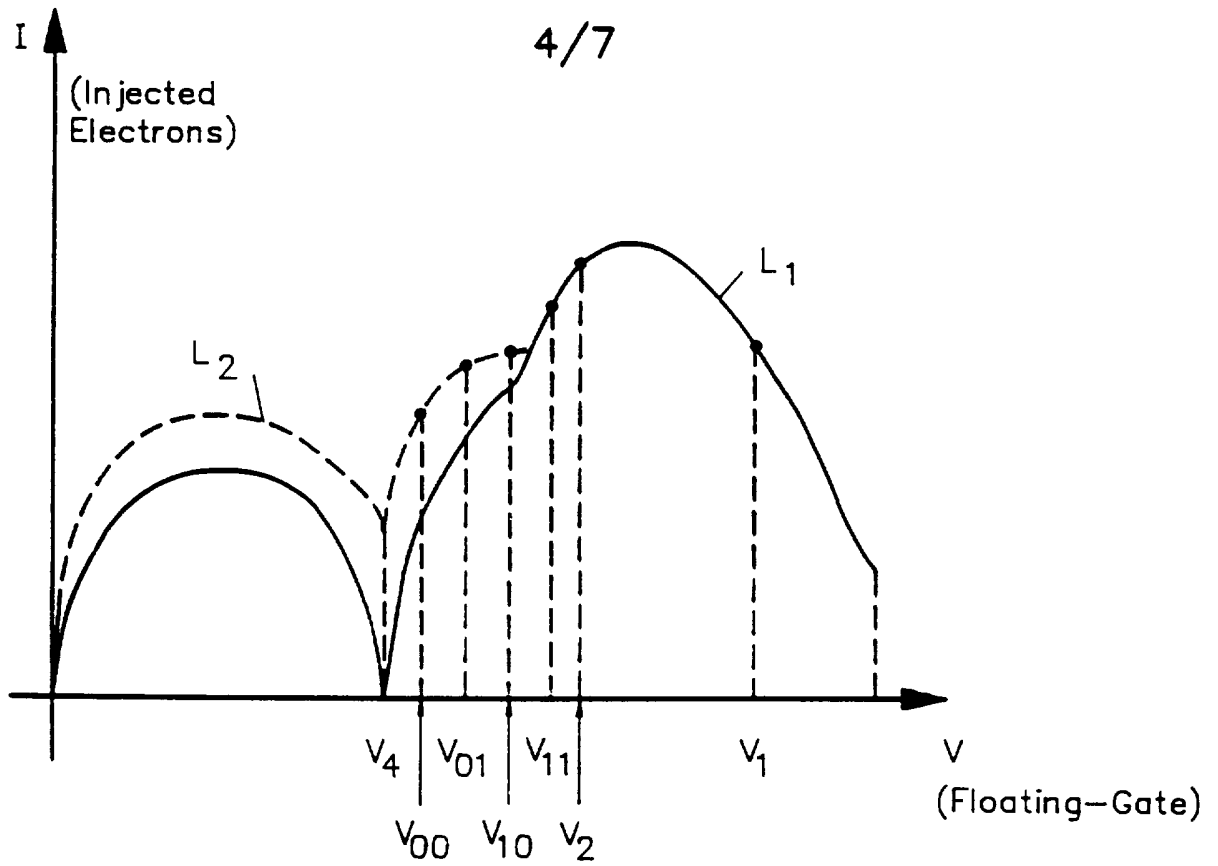


FIG. 4

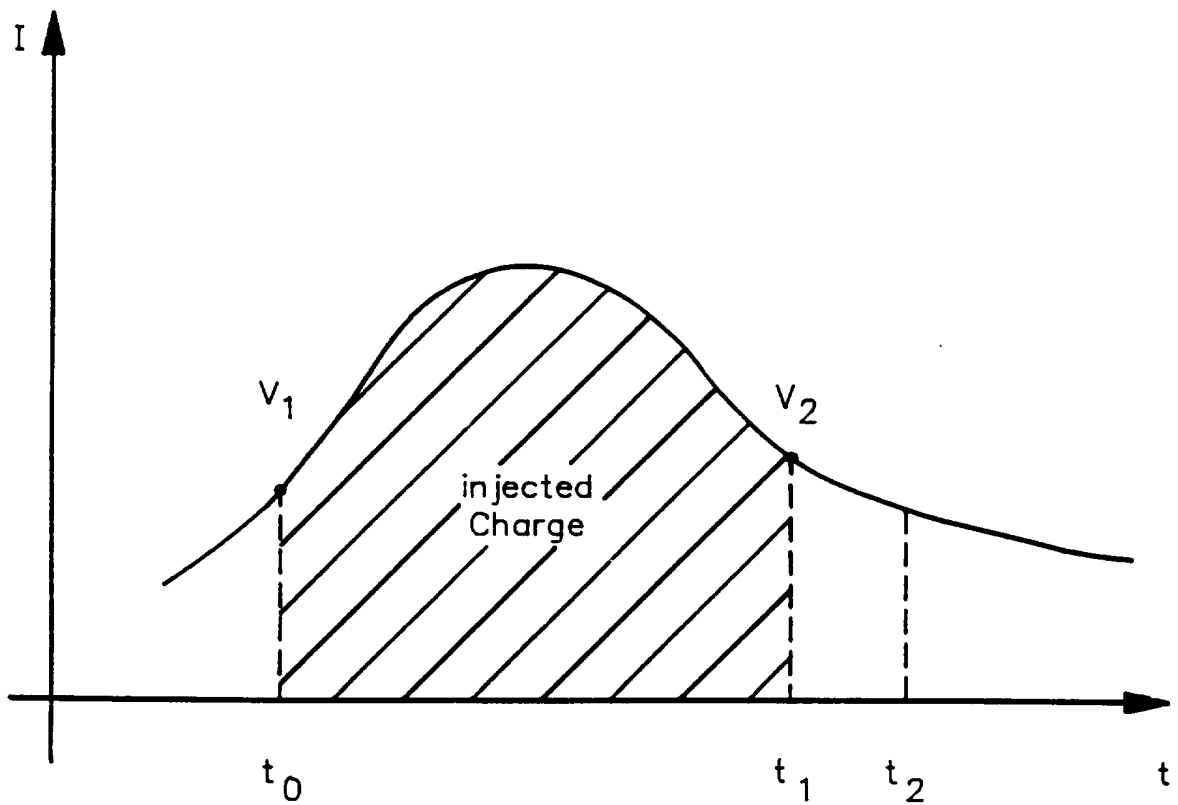


FIG. 5

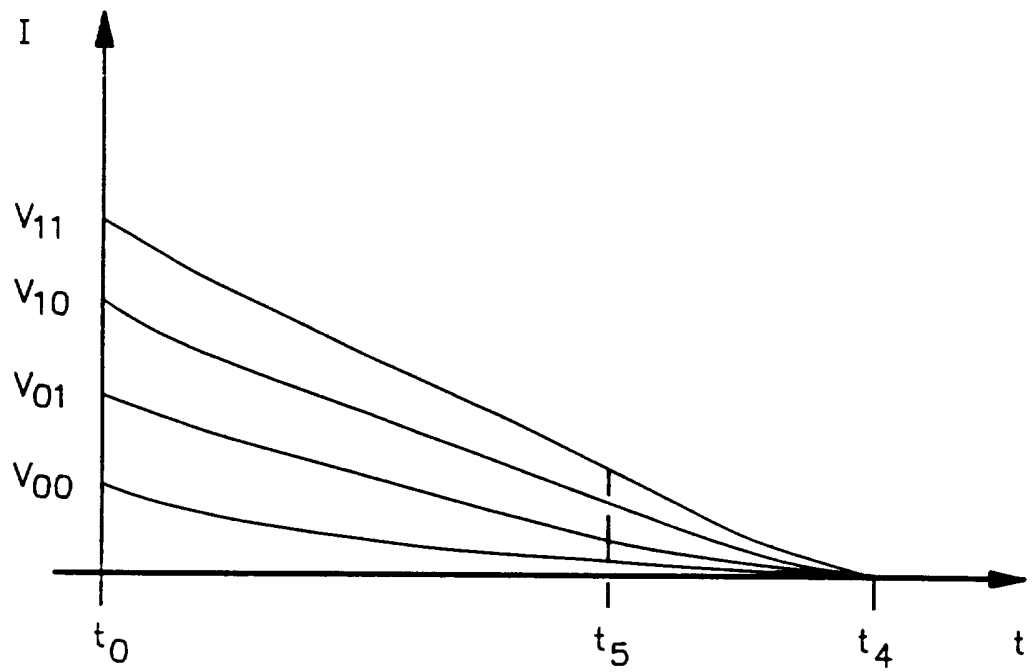


FIG. 6

6/7

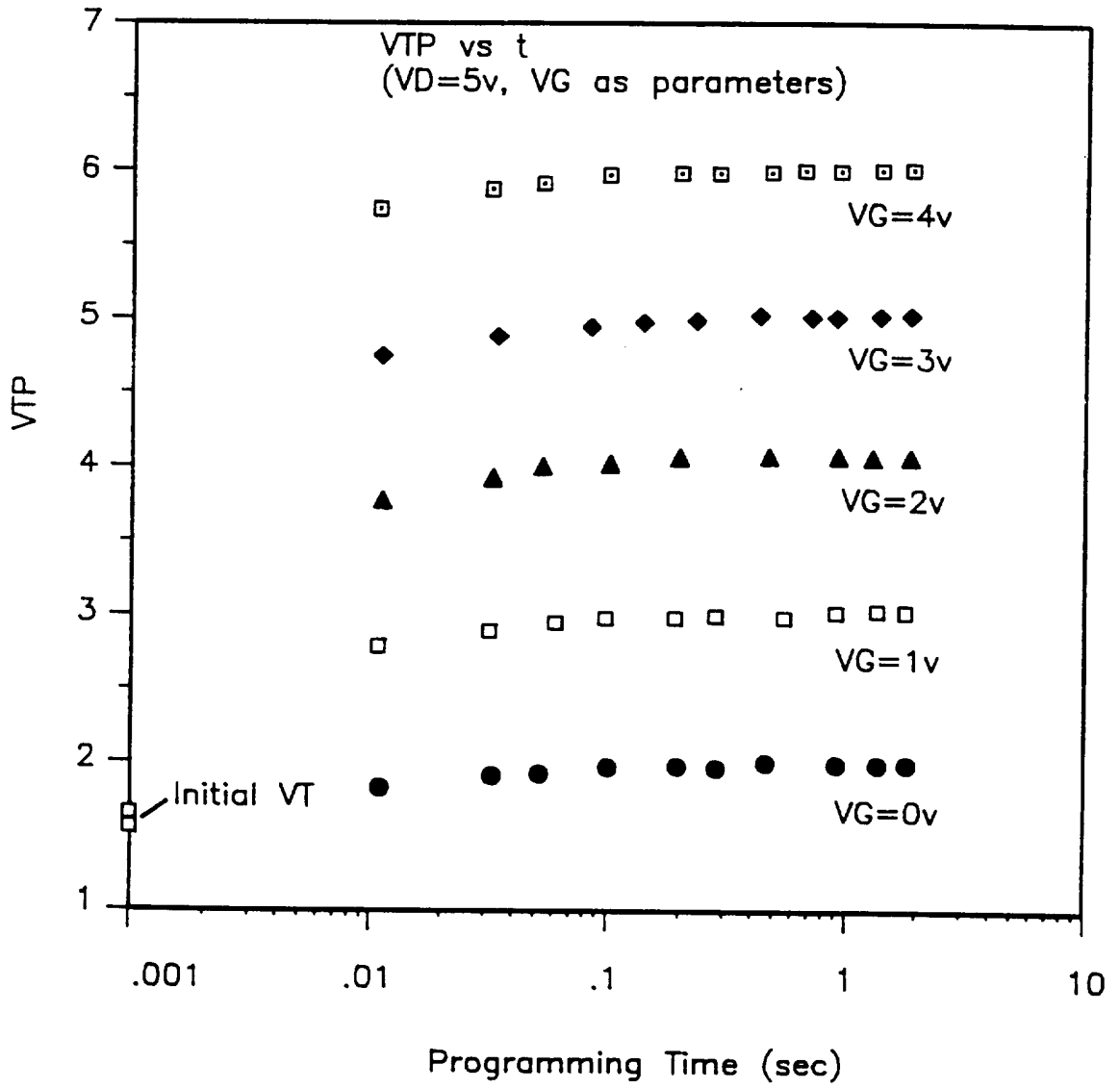


FIG. 7

7/7

100

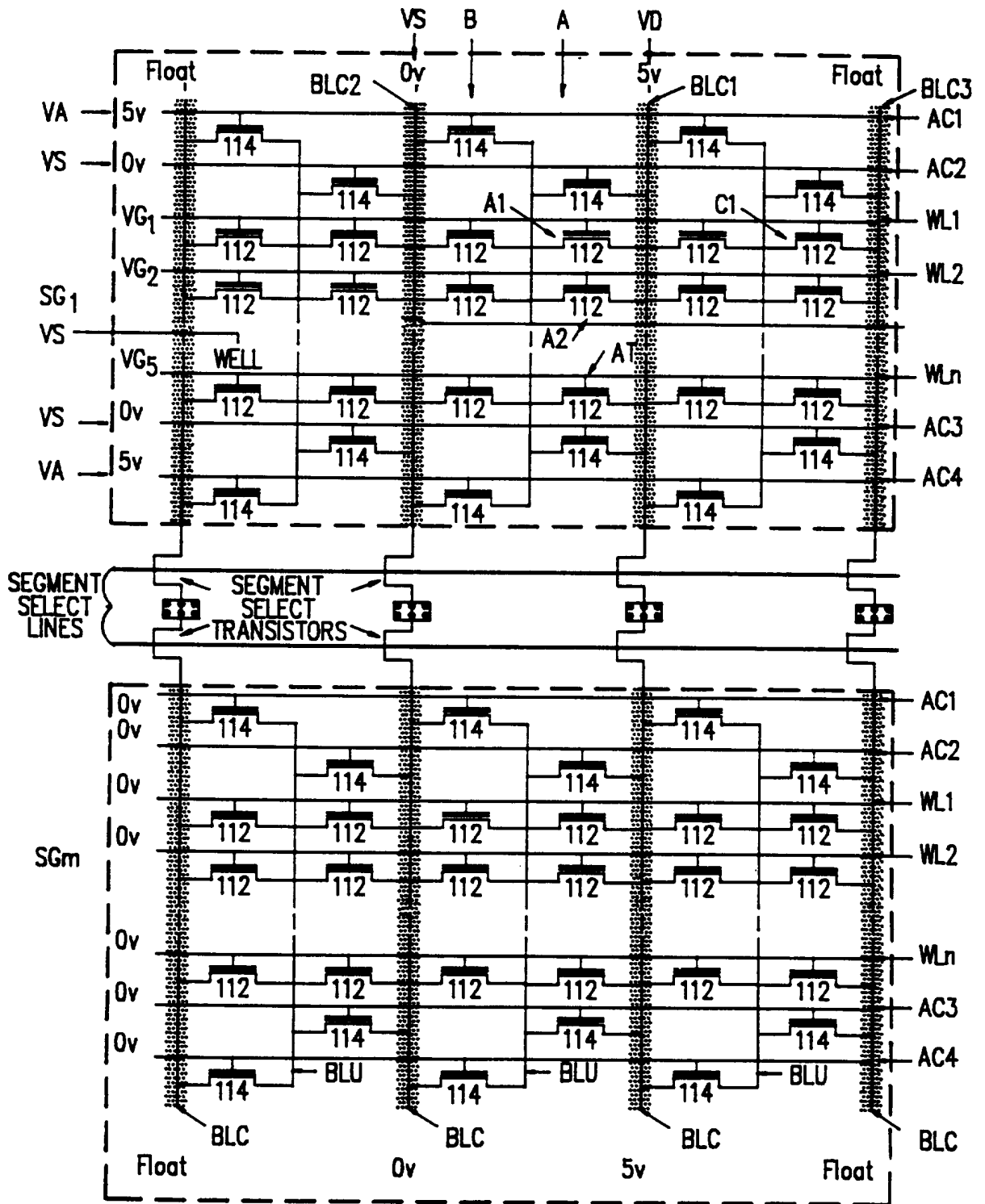


FIG. 8

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/04843

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G11C11/56

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G11C

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP,A,0 555 039 (NATIONAL SEMICONDUCTOR) 11 August 1993 see column 7, line 32 - column 7, line 56; figures 11,12	1
A	--- EP,A,0 461 764 (WAFERSCALE INTEGRATION) 18 December 1991 see column 7, line 10 - column 14, line 18; figure 6	1,2
P,A	--- EP,A,0 660 413 (WAFERSCALE INTEGRATION) 28 June 1995 see page 5, line 6 - page 7, line 32; figures 4-6	1,5,6
A	--- US,A,4 821 236 (HAYASHI ET AL) 11 April 1989 see the whole document ---	1
-/--		

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *&* document member of the same patent family

Date of the actual completion of the international search

29 August 1996

Date of mailing of the international search report

12 09. 96

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+ 31-70) 340-3016

Authorized officer

Degraeve, L

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 96/04843

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>IEICE TRANSACTIONS ON ELECTRONICS, vol. 77, no. 4, April 1994, TOKYO JP, pages 601-606, XP000460064 KABUTA ET AL: "a proposal of new multiple valued mask-rom design" see page 602, column 1, line 4 - page 603, column 1, line 30</p> <p style="text-align: center;">---</p>	1
A	<p>PATENT ABSTRACTS OF JAPAN vol. 9, no. 159 (E-326), 4 July 1985 & JP,A,60 038881 (KOGYO GIJUTSUIN) see abstract</p> <p style="text-align: center;">---</p>	1
A	<p>PATENT ABSTRACTS OF JAPAN vol. 12, no. 440 (E-684), 18 November 1988 & JP,A,63 172471 (AGENCY FOR IND SCIENCE & TECHNOL.) see abstract</p> <p style="text-align: center;">-----</p>	1

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No
PCT/US 96/04843

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP-A-555039	11-08-93	US-A- 5346842	13-09-94
		JP-A- 7086530	31-03-95
		US-A- 5397726	14-03-95
		US-A- 5464999	07-11-95
		US-A- 5481493	02-01-96

EP-A-461764	18-12-91	US-A- 5204835	20-04-93
		JP-A- 4230079	19-08-92

EP-A-660413	28-06-95	US-A- 5432730	11-07-95

US-A-4821236	11-04-89	JP-C- 1802660	26-11-93
		JP-B- 4081346	22-12-92
		JP-A- 59161873	12-09-84
