

(51) International Patent Classification:
G11C 11/34 (2006.01)(21) International Application Number:
PCT/US2016/013258(22) International Filing Date:
13 January 2016 (13.01.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/107,173 23 January 2015 (23.01.2015) US
62/120,184 24 February 2015 (24.02.2015) US
14/858,813 18 September 2015 (18.09.2015) US

(71) Applicant: CYPRESS SEMICONDUCTOR CORPORATION [US/US]; 198 Champion Court, San Jose, California 95134 (US).

(72) Inventors: ASHOKKUMAR, Jayant; 2450 Wimbleton Court, Colorado Springs, Colorado 80920 (US). RAGHAVAN, Vijay; 12513 Tenny Crags Rd, Colorado Springs, Colorado 80921 (US). PRABHAKAR, Venkatarman; 3720 Appian St, Pleasanton, California 94588 (US). SAHA, Swatilekha; 390 Elan Village Lane Unit 226, San Jose, California 95134 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— as to the identity of the inventor (Rule 4.17(i))

Published:

— with international search report (Art. 21(3))

(54) Title: A SPLIT VOLTAGE NON-VOLATILE LATCH CELL

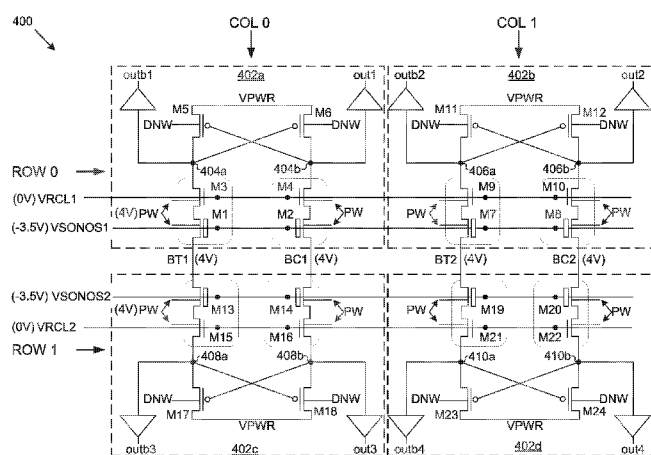


FIG. 4

(57) Abstract: A memory including an array of non-volatile latch (NVL) cells and method of operating the same are provided. In one embodiment, each NVL cell includes a non-volatile portion and a volatile portion. The non-volatile portion includes a first non-volatile memory (NVM) device and a first pass gate transistor coupled in series between a first output node and a bitline true, and a second NVM device and a second pass gate transistor coupled in series between a second output node and a bitline complement. The volatile portion includes cross-coupled first and second field effect transistors (FET), the first FET coupled between a supply voltage (VPWR) and the first output node, and the second FET coupled between VPWR and the second output node. A gate of the first FET is coupled to the second output node, and a gate of the second FET is coupled to the first output node.

A Split Voltage non-volatile latch Cell

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is an International Application of U.S. Application Number 14/858,813, filed September 18, 2015, which claims the benefit of priority under 35 U.S.C. 119(e) to U.S. Provisional Patent Applications Ser. Nos. 62/107,173, filed January 23, 2015, and 62/120,184, filed February 24, 2015, all of which are incorporated by reference herein in their entirety.

TECHNICAL FIELD

[0002] This disclosure relates generally to semiconductor memories and more particularly to a memory including Non-Volatile Latch cells and methods of operating the same.

BACKGROUND

[0003] A non-volatile latch (NVL) is a type of semiconductor memory that uses bistable, cross-coupled floating gate or silicon-oxide-nitride-oxide-silicon (SONOS) memory transistors provided to store each bit of data. Typically, NVL cells are very large, exceeding about 100 μm^2 using current semiconductor technology, and require high-voltages and complicated signals to power-up correctly. In addition, the architecture of a conventional NVL cell can provide indeterminate data on first power up, and the

memory transistors are prone to the well-known read-disturb phenomenon by which repeated read operations can degrade the stored data.

[0004] Thus, there is a need for a non-volatile latch which overcomes the shortcomings of the conventional NVL cell architecture.

SUMMARY

[0005] A memory including an array of non-volatile latch (NVL) cells having fewer transistors and smaller area, and methods of operating the same are provided. Generally, each NVL cell includes a non-volatile portion and a volatile portion. The non-volatile portion includes a series connected first non-volatile memory (NVM) device and a first pass gate transistor coupled between a first output node and a bitline true, and a series connected second NVM device and a second pass gate transistor coupled between a second output node and a bitline complement. The volatile portion includes cross-coupled first and second field effect transistors (FET), the first FET coupled between a supply voltage (VPWR) and the first output node, and the second FET coupled between VPWR and the second output node. A gate of the first FET is coupled to the second output node, and a gate of the second FET is coupled to the first output node.

[0006] In one embodiment, each of the first memory device and second memory device are in a triple well, and gates of the first and second NVM devices and the triple wells are biased separately from a negative voltage source (VNEG) and a positive voltage source (VPOS) during erase and program operations to enable split voltage operation to program and erase the first and second NVM devices.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007] Embodiments of the present invention will be understood more fully from the detailed description that follows and from the accompanying drawings and the appended claims provided below, where:

[0008] FIG. 1 is a schematic diagram of a split voltage non-volatile latch (NVL) cell according to an embodiment of the present disclosure;

[0009] FIG. 2 is a block diagram of a cross-sectional view of a portion of the NVL cell of FIG. 1 illustrating a triple well structure according to an embodiment of the present disclosure;

[0010] FIG. 3 is a block diagram illustrating an erase operation for a SONOS device in a split voltage NVL cell according to an embodiment of the present disclosure;

[0011] FIG. 4 is a schematic diagram of a 2X2 NVL array illustrating the bias conditions used for an erase operation according to an embodiment of the present disclosure;

[0012] FIG. 5 is a block diagram illustrating a program operation for the SONOS device in the split voltage NVL cell according to an embodiment of the present disclosure;

[0013] FIG. 6 is a schematic diagram of a 2X2 NVL array illustrating the bias conditions used for a program operation according to an embodiment of the present disclosure;

[0014] FIG. 7 is a schematic diagram of a 2X2 NVL array illustrating the bias conditions used for a recall or read operation according to an embodiment of the present disclosure;

[0015] FIGs. 8A and 8B are plots illustrating effects of read stress on threshold voltages (V_{TH}) of SONOS devices in a split voltage NVL cell according to an embodiment of the present disclosure at different drain bias voltages; and

[0016] FIG. 9 is a block diagram of a semiconductor memory including a single charge pump integrally formed on a substrate with an array of flash cells, and an array of NVL cells according to an embodiment of the present disclosure.

DETAILED DESCRIPTION

[0017] The present disclosure is directed generally to semiconductor memories and more particularly to a memory including Non-Volatile Latch (NVL) cells and method of operating the same.

[0018] In the following description, for purposes of explanation, numerous specific details are set forth in order to provide a thorough understanding of the present invention. It will be evident, however, to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known structures, and techniques are not shown in detail or are shown in block diagram form in order to avoid unnecessarily obscuring an understanding of this description.

[0019] Reference in the description to “one embodiment” or “an embodiment” means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the invention. The appearances of the phrase “in one embodiment” in various places in the specification do not necessarily all refer to the same embodiment. The term to couple as used herein may include both to directly electrically connect two or more components or elements and to

indirectly connect through one or more intervening components.

[0020] The terms "over," "under," "between," and "on" as used herein refer to a relative position of one layer with respect to other layers. As such, for example, one layer deposited or disposed over or under another layer may be directly in contact with the other layer or may have one or more intervening layers. Moreover, one layer deposited or disposed between layers may be directly in contact with the layers or may have one or more intervening layers. In contrast, a first layer "on" a second layer is in contact with that second layer. Additionally, the relative position of one layer with respect to other layers is provided assuming operations deposit, modify and remove films relative to a starting substrate without consideration of the absolute orientation of the substrate.

[0021] FIG. 1 is a schematic diagram of a split voltage NVL cell according to one embodiment of the present disclosure. Referring to FIG. 1, the NVL cell **100** is a differential cell and includes a non-volatile portion **102**, and a volatile portion **104** or latch portion that is set based on data read from the non-volatile portion **102** upon power up of the cell. The non-volatile portion **102** includes two sets (MT and MC) of series connected non-volatile memory (NVM) devices and pass gate transistors. In the embodiment shown, the first set of series connected NVM devices, MT, consists of a first NVM device M1 and a first pass gate transistor M3 coupled between a first output node 106 and a bitline true BT. The second set of series connected NVM devices, MC, consists of a second NVM device M2 and a second pass gate transistor M4 coupled between a second output node 108 and a bitline complement BC. The first and second NVM devices M1, M2, may include NVM devices or transistors implemented using Silicon-Oxide-

Nitride-Oxide-Silicon (SONOS), floating-gate MOS field-effect transistor (FGMOS), or ferroelectric random access memory (FeRAM) technology.

[0022] The volatile portion **104** consists of cross-coupled first and second field effect transistors (FET), including a first FET **M5** coupled between a supply voltage (VPWR) and the first output node **106**, and a second FET **M6** coupled between VPWR and the second output node **108**. By cross-coupled it is meant that a gate terminal of the first FET **M5** is coupled to the second output node **108** and through the second output node to a source or drain terminal of the second FET **M6**, and a gate terminal of the second FET **M6** is coupled to the first output node **106** and to a source or drain terminal of the first FET **M5**. The first and second FETs **M5**, **M6**, include channels of a different or opposite type, i.e., with a P or N type dopant, from the first and second pass gate transistors **M3**, **M4** and the first and second NVM devices **M1**, **M2**, which generally include channels of the same dopant type. For example, in the embodiment shown in FIG. 1 and described herein the first and second FETs **M5**, **M6**, are P-channel metal-oxide-semiconductor (PMOS) FETs, while the first and second pass gate transistors **M3**, **M4**, are N-channel metal-oxide-semiconductor (NMOS) FETs, and the first and second NVM devices **M1**, **M2**, are N-channel NVM transistors such as NSONOS N-type FGMOS transistors.

[0023] In certain embodiments, the first and second FETs are weak PMOS transistors having high threshold voltages higher than the threshold voltages of the first and second pass gate transistors so that the NVM devices of the non-volatile portion **102** can pull down on the first and second output nodes **106**, **108**, of the latch portion. For example, the first and second FETs can include PMOS transistors having high threshold

voltages of about 0.75V and a drain saturation current (I_{dsat}) of about 10 μ A for an applied voltage of about 1.2V. Such high threshold voltages can be achieved by providing the weak PMOS transistors with thick gate oxides of from about 25 to about 33 nanometers (nm).

[0024] The NVL cell **100** is placed or formed in a triple well structure (not shown in FIG. 1) in which the first and second FETs M5, M6, are in a deep Nwell in a P-substrate, and the first and second pass gate transistors M3, M4 and the first and second NVM devices M1, M2, are formed in a Pwell inside the deep Nwell, so that during program and erase operations a total high voltage applied across the NVM devices can be split between gates of the NVM devices and the Pwell in which the NVM device is placed or formed. An embodiment of one such triple well structure is shown in FIG. 2.

[0025] FIG. 2 is a block diagram of a cross-sectional view of a portion of the NVL cell **100** of FIG. 1 showing NVM device M1, pass gate transistor M3 and first FET M5. Referring to FIG. 2 it is noted that in an embodiment in which the first FET M5 is a PMOS FET, it is formed or placed in a deep Nwell **110** in a P-substrate **112**, and the first pass gate transistor M3 and the first NVM device M1 are in a Pwell **114** in the deep Nwell. The deep Nwell is biased through a deep Nwell terminal (DNW) while the Pwell is biased through a Pwell terminal (PW) as show in this figure and in FIG. 1. While not shown it will be understood that the second FET M6 is also formed or placed in the deep Nwell **110** in the P-substrate **112**, and the second pass gate transistor M4 and the second NVM device M2 are in a Pwell in the deep Nwell. It will also be understood that the Pwell in which the second pass gate transistor M4 and the second NVM device M2 are

placed or formed can be the same Pwell **114** as shown in FIG. 2, or in a separate Pwell in the deep Nwell.

[0026] Referring again to FIG. 1, the NVL cell may further include a first output buffer **116** coupled to the first output node **106** to provide a first data output (outb), and a second output buffer **118** coupled to the second output node **108** to provide a second data output (out) of which the first output is a complement. The first and second output buffers **116**, **118**, can include either high voltage (HV) or low voltage (LV) output buffers. However, as explained in greater detail below, the output nodes **106**, **108**, of the NVL cell **100** are isolated from high voltages in the NVM portion **102** during program and erase (P/E) operations. The output nodes are at a low voltage and therefore can interface directly with LV devices in a memory incorporating NVL cells **100** of the present disclosure with or without an LV buffer.

[0027] It will be evident, however, to those skilled in the art that the above NVL cell **100** requires substantially fewer transistors or devices as compared to the prior art. Conventional NVL cells typically require eighteen or more transistors per NVL cell as compared to the six in the NVL cells **100** of the present disclosure. Therefore, the area per NVL cell is substantially smaller – less than about $3\ \mu\text{m}^2$ compared to about $125\ \mu\text{m}^2$, thus enabling smaller semiconductor memories or higher density memories. Moreover, because only a single N-type or P-type of NVM devices M1, M2 are used in each cell, fewer masks are needed to fabricate the NVL cells of the present disclosure, as compared to the prior art, lowering costs and fabrication time while also increasing the yield of working memories.

[0028] An erase operation for an NVM device or transistor implemented using SONOS technology will now be described with reference to FIG. 3. FIG. 3 is a block diagram of a cross-sectional view of an embodiment of an N-channel SONOS device or transistor suitable for use in the NVL cell of the present disclosure. Referring to FIG. 3 SONOS device **300** includes a gate stack **302** formed over a substrate **304**. In the embodiment shown, the SONOS device **300** is formed in a first well (Pwell **306**) and formed in a second deep well (deep Nwell **308**) in a P type substrate **304** as described above. SONOS device **300** further includes source/drain regions **310** formed in Pwell **306** on either side of gate stack **302**, which define a channel region **312** underneath gate stack. Gate stack **302** includes an oxide tunnel dielectric layer **314**, one or more nitride or oxynitride charge-trapping layers **316**, a top, blocking oxide layer **318** and a poly-silicon (poly) or metal layer which serves as a control gate or gate **320**.

[0029] When a negative bias is applied to the gate **320** relative to the substrate or Pwell **306**, holes or a positive charge accumulated in the channel region **312** are injected or tunnel through tunnel dielectric layer **314** and are trapped in the charge-trapping layers **316**. The mechanisms by which charge is injected can include both Fowler-Nordheim (FN) tunneling and hot-carrier injection. The charge trapped in the charge-trapping layer **316**, reduces the threshold voltage (V_T) necessary to turn on the device **300**, lowering an energy barrier between the drain and the source, and putting the device in an “erased” or zero data state. An erased NSONOS transistor having an erased threshold voltage (V_{TE}) has a relatively high drain-to-source current (I_{ds}) in the tens of microampere (μA) range with an applied gate-to-source voltage (V_{gs}) of about 0V.

[0030] In another embodiment, the NVM device can be a floating-gate MOS

field-effect transistor (FGMOS) or device. Generally, FGMOS is similar in structure to the SONOS device **300** described above, differing primarily in that a FGMOS includes a poly-silicon (poly) floating gate, which is capacitively coupled to a control gate of the device, rather than a nitride or oxynitride charge-trapping layers **316**. Similar to the SONOS device **300** described above the FGMOS device can be erased by applying an appropriate bias between the control gate and the Pwell terminal (PW) to inject a positive charge (holes) on to the floating gate, lowering the threshold voltage V_T necessary to turn on the FGMOS device.

[0031] In yet another embodiment, the NVM device can be a FeRAM transistor or device. FeRAM is a non-volatile random-access memory technology similar in construction to flash memory, such as DRAM, but uses a ferroelectric layer instead of a dielectric layer to achieve non-volatility. FeRAM is operated similar to flash memory. Erasing is accomplished by applying a field across the ferroelectric layer by charging the plates on either side of it, forcing the atoms inside into the "up" or "down" orientation (depending on the polarity of the charge), thereby storing a "1" or "0".

[0032] An erase operation for an NVL array including NVL cells according to an embodiment of the present disclosure now will be described with reference to FIG. 4. FIG. 4 is a schematic diagram of a two row by two column (2X2) portion of an NVL array illustrating the bias conditions used for an erase operation on the 2X2 NVL array. Referring to FIG. 4, the 2X2 NVL array **400** includes four NVL cells, including a first NVL cell **402a** in a first column (COL. 0) and a first row (ROW. 0), a second NVL cell **402b** in a second column (COL. 1) and the first row (ROW. 0), a third NVL cell **402c** in the first column (COL. 0) and a second row (ROW. 1), and a fourth NVL cell **402d** in the

second column (COL. 1) and the second row (ROW. 1). As with the NVL cell **100** described above with reference to FIG. 1, each NVL cell **402a-d** is a differential cell and including a non-volatile portion having two sets of series NVM devices and pass gate transistors, and a volatile portion or latch portion consisting of cross-coupled first and second FETs coupled between VPWR and first and second output nodes. The latch is set based on data read from the non-volatile portion upon power up. In the embodiment shown, each cell further includes first and second output buffers coupled to first and second nodes respectively.

[0033] In particular, NVL cell **402a** includes first and second NVM devices M1, M2, and first and second pass gate transistors M3, M4. First NVM device M1 and first pass gate transistor M3 are coupled in series between a first node **404a** and a bitline true (BT1). Second NVM device M2 and second pass gate transistor M4 are coupled in series between a second node **404b** and a bitline complement (BC1). The gates of the first and second NVM devices M1, M2, are coupled to a wordline or program/erase line (VSONOS1), and the gates of the first and second pass gate transistors M3, M4, are coupled to a recall line (VRCL1). The volatile portion includes cross-coupled first and second FETs M5, M6, coupled between VPWR and first and second output nodes **404a**, **404b**, with first FET M5 having a gate coupled to the second output node **404b**, and second FET M6 having a gate coupled to the first output node **404a**.

[0034] The second NVL cell **402b** in the first row (ROW.0) and second column (COL.1) includes first and second NVM devices M7, M8, and first and second pass gate transistors M9, M10. First NVM device M7 and first pass gate transistor M9 are coupled in series between a first node **406a** and a bitline true (BT2). Second NVM device M8 and

second pass gate transistor M10 are coupled in series between a second node **406b** and a bitline complement (BC2). The gates of the first and second NVM devices M7, M8, are coupled to VSONOS1, and the gates of the first and second pass gate transistors M9, M10, are coupled to VRCL1. The volatile portion includes cross-coupled first and second FETs M11, M12, coupled between VPWR and first and second output nodes **406a**, **406b**, with first FET M11 having a gate coupled to the second output node **406b**, and second FET M12 having a gate coupled to the first output node **406a**.

[0035] The third NVL cell **402c** in the second row (ROW.1) and first column (COL. 0) includes first and second NVM devices M13, M14, and first and second pass gate transistors M15, M16. First NVM device M13 and first pass gate transistor M15 are coupled in series between a first node **408a** and a bitline true (BT1). Second NVM device M14 and second pass gate transistor M16 are coupled in series between a second node **408b** and a bitline complement (BC1). The gates of the first and second NVM devices M13, M14, are coupled to a program/erase line (VSONOS2), and the gates of the first and second pass gate transistors M15, M16, are coupled to a recall line (VRCL2). The volatile portion includes cross-coupled first and second FETs M17, M18, coupled between VPWR and first and second output nodes **408a**, **408b**, with first FET M17 having a gate coupled to the second output node **408b**, and second FET M18 having a gate coupled to the first output node **408a**.

[0036] Finally, the fourth NVL cell **402d** in the second row (ROW.1) and second column (COL. 1) includes first and second NVM devices M19, M20, and first and second pass gate transistors M21, M22. First NVM device M19 and first pass gate transistor M21 are coupled in series between a first node **410a** and a bitline true (BT2). Second NVM

device M20 and second pass gate transistor M22 are coupled in series between a second node **410b** and a bitline complement (BC2). The gates of the first and second NVM devices M19, M20, are coupled to VSONOS2, and the gates of the first and second pass gate transistors M21, M22, are coupled to VRCL2. The volatile portion includes cross-coupled first and second FETs M23, M24, coupled between VPWR and first and second output nodes **410a**, **410b**, with first FET M23 having a gate coupled to the second output node **410b**, and second FET M24 having a gate coupled to the first output node **410a**.

[0037] Referring to FIG. 4, the erase operation uses a split voltage concept wherein the gates of the NVM devices M1, M2, M7, M8, M13, M14, M19, M20, coupled to VSONOS1 and VSONOS2 go to a negative value (VNEG) of -3.5V while the Pwells (PW) go to a positive value (VPOS) of 4V. Thus, the total voltage across the NVM devices is about 7.5V, which is sufficient to enable holes or a positive charge to tunnel or be injected from the substrate into a nitride charge trapping layer of the NVM devices. The pass transistors M3, M4, M9, M10, M15, M16, M21, M22, whose gates are connected to VRCL1 and VRCL2, are held at ground during an erase operation thus isolating the volatile latch from the non-volatile portion of the NVL cells **402a-d**. Throughout the erase operation the true bitlines (BT1 and BT2) and complement bitlines (BC1 and BC2) and the deep Nwells (DNW) are held at VPOS or 4V. VPWR is from about 1.2V to about 3.6V.

[0038] A program operation for an NVM device or transistor implemented using SONOS technology will now be described with reference to FIG. 5. FIG. 5 is a block diagram of a cross-sectional view of an embodiment of an N-channel SONOS device or transistor suitable for use in the NVL cell of the present disclosure. The structure of the

SONOS device is identical to that described in reference to the SONOS device **300** above. Referring to FIG. 5 SONOS device **500** includes a gate stack **502** formed over a substrate **504**. In the embodiment shown, the SONOS device **500** is formed in a first well (Pwell **506**) formed in a second deep well (deep Nwell **508**) in a P type substrate **504** as described above. SONOS device **500** further includes source/drain regions **510** formed in Pwell **506** on either side of gate stack **502**, which define a channel region **512** underneath gate stack. Gate stack **502** includes an oxide tunnel dielectric layer **514**, one or more nitride or oxynitride charge-trapping layers **516**, a top blocking oxide layer **518** and a poly-silicon (poly) or metal layer which serves as a control gate or gate **520**.

[0039] When a positive bias is applied to the gate **520** relative to the substrate or Pwell **506**, electrons or a negative charge accumulated in the channel region **512** are injected or tunnel through tunnel dielectric layer **514** and are trapped in the charge-trapping layers **516**. The negative charge trapped in the charge-trapping layer **516**, increases the threshold voltage (V_T) necessary to turn on the device **500**, thus a programmed NSONOS transistor has a more positive programmed threshold voltage (V_{TP}), raising an energy barrier between the drain and the source, turning the transistor “off” and putting it in a programmed or one data state. A programmed NSONOS transistor has a low drain-to-source current (I_{ds}) of nano-ampere (nA) range with an applied gate-to-source voltage (V_{gs}) of about 0V.

[0040] A program operation for an NVL array including NVL cells according to an embodiment of the present disclosure will now be described with reference to FIG. 6. FIG. 6 is a schematic diagram of a 2X2 portion of an NVL array illustrating the bias conditions used for a program operation on the 2X2 NVL array, and the structure of the

array and the NVL cells in the array is identical to that described above with reference to FIG. 4. Referring to FIG. 6, the program operation uses a split voltage concept to program the selected NVM devices. In particular, throughout the program operation voltage on recall lines (VRCL1 and VRCL2) are held to -3.5V to isolate the volatile latch from the non-volatile portion of the NVL cells **402a-d**; the program/erase line (VSONOS1) in the first row (ROW.0) coupled to the gates of NVM devices M1, M2, M7, M8, is held to VPOS (4V); the true bitlines (BT1 and BT2) coupled to NVM devices M1, M13, M7, M19, are held to VNEG (-3.5V); complement bitlines (BC1 and BC2) coupled to NVM devices M2, M14, M8, M20, are held to a static write inhibit voltage of about 1.2V; the Pwells are held to VNEG (-3.5V); and the deep Nwells (DNW) are held at VPOS or 4V. VPWR is from about 1.2V to about 3.6V.

[0041] With these bias conditions, the NVM devices M1 and M7 are programmed because the total voltage across VSONOS1 to BT1/BT2 is about 7.5V, which is sufficient to enable electrons or a negative charge to tunnel or be injected from the substrate into a nitride charge trapping layer of the NVM devices. The voltage across VSONOS1 to BC1/BC2 is around 2.8V, i.e., $4V - 1.2V = 2.8V$, which is not sufficient to enable electrons or a negative charge to tunnel or be injected from the substrate into a nitride charge trapping layer of the NVM devices, and thus NVM devices M2 and M8 which are connected to BC1/BC2 are program inhibited. For the second row (ROW.1), which is not being programmed in FIG. 6, VSONOS2 and VRCL2 are held to VNEG (-3.5V). Thus, none of the NVM devices M13, M14, M19, M20 in the un-selected row will be programmed. The pass transistors M3, M4, M9, M10, M15, M16, M21, M22, whose gates are connected to VRCL1 and VRCL2, are held at VNEG during the program

operation thus isolating the volatile latch from the non-volatile portion of the NVL cells **402a-d**.

[0042] A recall or read operation for an NVL array including NVL cells according to an embodiment of the present disclosure will now be described with reference to FIG. 7. FIG. 7 is a schematic diagram of a 2X2 portion of an NVL array illustrating the bias conditions used for a program operation on the 2X2 NVL array, and the structure of the array and the NVL cells in the array is identical to that described above with reference to FIG. 4. Referring to FIG. 7, in this mode the contents from the non-volatile portion are recalled into the volatile portion and latched. Erased NVM devices M2, M8 has an erased threshold voltage (V_{TE}) and are on. Thus, the second node **404b** in NVL cell **402a** discharges to bitline complement BC1, turns on first FET M5 and latches the data. Similarly, the second node **406b** in NVL cell **402b** discharges to bitline complement BC2, turning on first FET M11 and latching the data.

[0043] To recall data to the first row (ROW.0) in the NVL array in FIG. 7, true bitlines (BT1 and BT2), complement bitlines (BC1 and BC2) and VSONOS1 are held at 0V. Next, VRCL1 is held to VPWR (1.2V-3.6V). As the power (VPWR) supplied to the array from an external source slowly ramps up from a minimum operating voltage of 1.2V to 3.6V. As VPWR ramps up, both first output nodes **404a**, **406a**, and second output nodes **404b**, **406b**, ramp up also, but because the VSONOS1 is at VGND, the erased NVM devices M2, M8 are ON and conducting current and discharging second output nodes **404b**, **406b** to VGND which in turn ON the first FETs M5, M11 and put the output nodes **404a**, **406a** nodes to VPWR thus latching the recalled data from the non-volatile portion.

[0044] The programmed first NVM device M1 and first pass gate transistor M3, and first NVM device M7 and first pass gate transistor M9 in NVL cells 402a and 402b are OFF and no current flows through those series connections. Thus, the volatile data stays latched as long as power (VPWR) is ON or above the minimum operating voltage of 1.2V.

[0045] The same procedure may be followed to recall data for the second row (ROW.1). That is true bitlines (BT1 and BT2), complement bitlines (BC1 and BC2) and VSONOS2 are held at 0V. Next, VRCL2 is held to VPWR (1.2V-3.6V).

[0046] NVM Devices M1, M7, in the first row (ROW.0) can see a read disturb during a recall operation because both these NVM devices see an erase voltage of -3V between the gate (0V) and the drain coupled to the pass gate transistors M3, M9. This can limit the reliability of the NVM devices over time. To understand this a Time Dependent Dielectric Breakdown (TDDB) analysis was done by looking at data from actual NVL cells and the results are shown in FIGs. 8A and 8B.

[0047] FIG. 8A is a plot illustrating effects of read stress on threshold voltages (V_{TH}) of SONOS type NVM devices in a split voltage NVL cell for drain bias (V_D) of 3.6V, 3.0V and 2.0V for 120 seconds. Referring to FIG. 8A it is seen that, as expected, the NVL cell cannot survive a V_D stress of 3.6V, exhibiting a drop in V_{TH} of greater than 1.5V. It is further noted that a V_D stress of 3.0V resulted in a drop in V_{TH} from about 1.3V to about 0.2V in about 120 seconds.

[0048] FIG. 8B is a plot illustrating effects of read stress on V_{TH} of SONOS type NVM devices in a split voltage NVL cell for V_D of 2.4V, 2.0V and 1.8V for a time in excess of $1.0E+05$ seconds. Referring to FIG. 8B it is seen that for a V_D of 1.8V the V_{TH}

of the NVM device degrades by less than about 0.4 to a V_{TH} of greater than 1.0V in a time of about $1.0E+05$ seconds or about 27 hours, which is acceptable. For drain bias of 2.4 V and 2.0 V, the V_{TH} degradation is worse than 1.0V.

[0049] Because of this TDDB data, it is desirable in certain embodiments to restrict operation of the NVL cell to a maximum voltage of 1.8V, even though the external supply to the semiconductor memory can as high as 3.6V. Thus, in one embodiment a semiconductor memory including NVL cells according to the present disclosure, would further include an internally regulated supply voltage (VPWR) of 1.2V as soon as power up is finished so that the NVM devices will not see a high disturb voltage or V_D exceeding 1.8V.

[0050] In another embodiment, shown in FIG. 9, the array of NVL cells **902** can be part of a semiconductor memory **900** further including an array of flash cells **904** integrally formed on a common or shared substrate **906** with the array of NVL cells. In certain versions of this embodiment the semiconductor memory **900** may further include a single charge pump **908**, also integrally formed on the same substrate **906** with the array of flash cells **904** and the array of NVL cells **902**, and configured to be used to program and erase both the flash cells and the NVL cells, thus reducing the number of charge pumps need in the semiconductor memory **900**, and further reducing the area needed for the memory. It will be understood that the ability to use a single charge pump **908** is made possible by the split-voltage architecture and method of operation, which enable the NVM devices in the NVL cells of the present disclosure to be programmed and erased with lower magnitudes of voltage, and by the output nodes **106**, **108**, of the NVL cell **100** are isolated from high voltages in the NVM portion **102** during program and erase.

[0051] Thus, embodiments of semiconductor memories including Non-Volatile Latch (NVL) cells and method of the same have been described. Although the present disclosure has been described with reference to specific exemplary embodiments, it will be evident that various modifications and changes may be made to these embodiments without departing from the broader spirit and scope of the disclosure. Accordingly, the specification and drawings are to be regarded as an illustrative rather than a restrictive sense.

[0052] The Abstract of the Disclosure is provided to comply with 37 C.F.R. §1.72(b), requiring an abstract that will allow the reader to quickly ascertain the nature of one or more embodiments of the technical disclosure. It is submitted with the understanding that it will not be used to interpret or limit the scope or meaning of the claims. In addition, in the foregoing Detailed Description, it can be seen that various features are grouped together in a single embodiment for the purpose of streamlining the disclosure. This method of disclosure is not to be interpreted as reflecting an intention that the claimed embodiments require more features than are expressly recited in each claim. Rather, as the following claims reflect, inventive subject matter lies in less than all features of a single disclosed embodiment. Thus, the following claims are hereby incorporated into the Detailed Description, with each claim standing on its own as a separate embodiment.

[0053] Reference in the description to one embodiment or an embodiment means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the circuit or method. The appearances of the phrase one embodiment in various places in the specification do not

necessarily all refer to the same embodiment.

IN THE CLAIMS

WHAT IS CLAIMED IS:

1. A memory comprising:
an array of non-volatile latch (NVL) cells, each NVL cell comprising:
a non-volatile portion including:
a first non-volatile memory (NVM) device and a first pass gate transistor coupled in series between a first output node and a bitline true ; and
a second NVM device and a second pass gate transistor coupled in series between a second output node and a bitline complement; and
a volatile portion including cross-coupled first and second field effect transistors (FET), the first FET coupled between a supply voltage and the first output node, and the second FET coupled supply voltage and the second output node,
wherein a gate of the first FET is coupled to the second output node and a gate of the second FET is coupled to the first output node.
2. The memory of claim 1, wherein each of the first NVM device and the second NVM device is in a triple well.
3. The memory of claim 2, wherein gates of the first and second NVM devices and the triple wells are biased separately from a negative voltage source (VNEG) and a positive voltage source (VPOS) during erase and program operations to enable split

voltage operation to program and erase the first and second NVM devices.

4. The memory of claim 3, further comprising low voltage (LV) devices coupled to the first and second output nodes, wherein the first and second pass gate transistors are configured to isolate the volatile portion of the NVL cell from the split voltage used to erase and program the non-volatile portion during erase and program operations.

5. The memory of claim 3, further comprising high voltage (HV) buffers coupled to the first and second output nodes.

6. The memory of claim 1, wherein the first NVM device and the second NVM device comprise transistors having channels of the same type.

7. The memory of claim 6, wherein the first and second pass gate transistors comprise channels of the same type as the first and second NVM devices, and wherein the first and second FETs comprise channels of a different type from the first and second NVM devices.

8. The memory of claim 7, wherein the first FET and the second FET are in a deep well of a first type in a substrate of a second type, and wherein each of the first NVM device and the second NVM device is in a well of the second type in the deep well.

9. The memory of claim 1, wherein the first NVM device and the second NVM

device comprise silicon-oxide-nitride-oxide-silicon (SONOS), floating gate or FeRAM type NVM devices.

10. The memory of claim 1, wherein the first FET and the second FET are P-channel FETs in a deep Nwell in a P-type substrate, and the first NVM device and the second NVM device comprise N-channel silicon-oxide-nitride-oxide-silicon (SONOS) devices in separate Pwells in the deep Nwell.

11. The memory of claim 10, further comprising an array of flash cells, and a single charge pump integrally formed on a substrate with the array of NVL cells, wherein the charge pump is configured to be used to program and erase both the flash cells and the NVL cells.

12. The memory of claim 1, wherein the first and second field effect FETs comprise weak P-channel metal-oxide-semiconductors (PMOS) transistors having threshold voltages of higher magnitude than the first and second pass gate transistors.

13. A memory comprising:

- an array of flash cells;

- an array of non-volatile latch (NVL) cells, each NVL cell comprising:

- a non-volatile portion including:

- a series connected first non-volatile memory (NVM) device and a

- first pass gate transistor coupled between a first output node and a bitline

true; and

a series connected second NVM device and a second pass gate transistor coupled between a second output node and a bitline complement;

a volatile portion including cross-coupled first and second field effect transistors (FET), the first FET coupled between a supply voltage and the first output node, and the second FET coupled between the supply voltage and the second output node; and

a single charge pump configured to be used to program and erase both the flash cells and the NVL cells.

14. The memory of claim 13, wherein each of the first NVM device and second NVM device is in a triple well.

15. The memory of claim 14, wherein gates of the first and the second NVM devices and the triple wells are biased separately from a negative voltage source (VNEG) and a positive voltage source (VPOS) during erase and program operations to enable split voltage operation to program and erase the first and second NVM devices.

16. The memory of claim 14, wherein the first and second pass gate transistors and the first and second NVM devices comprise transistors having channels of the same type, and wherein the first and second FETs comprise channels of a different type.

17. The memory of claim 16, wherein the first FET and the second FET are in a deep well of a first type in a substrate of a second type, and wherein each of the first NVM device and the second NVM device is in a well of the second type in the deep well.

18. The memory of claim 14, wherein the first FET and the second FET are P-channel FETs in a deep Nwell in a P-type substrate, and the first NVM device and the second NVM device comprise N-channel devices in separate Pwells in the deep Nwell.

19. The NVL cell of claim 18, wherein the first FET and the second FET are weak P-channel FETs having threshold voltages of higher magnitude than the first and second pass gate transistors.

20. A method of programming an array of non-volatile latch (NVL) cells comprising:
providing an array of NVL cells, each including a first non-volatile memory (NVM) device and a first pass gate transistor formed in a PWELL of a triple well and coupled in series between a first output node and a bitline true, a second NVM device and a second pass gate transistor formed in the PWELL and coupled in series between a second output node and a bitline complement, and a volatile latch including cross-coupled first and second field effect transistors coupled between a supply voltage and the output nodes;

isolating the first and second NVM devices in all NVL cells in the array from the output nodes by applying a negative voltage (VNEG) to gates of the first and second pass gate transistors;

applying VNEG to the PWELLS in all NVL cells in the array;

applying VNEG to the bitline true while applying an inhibit voltage (VINH) to the bitline complement in all NVL cells in the array; and

applying a programming voltage (VPOS) to gates of the first and second NVM devices in NVL cells in a selected row of the array while applying VNEG to the first and second NVM devices in NVL cells in a deselected row,

wherein a difference between VPOS and VNEG raises a threshold voltage (V_{TP}) of the first NVM devices in NVL cells on the selected row, and the difference between VPOS and VNEG applied to the bitline true is sufficient to program first NVM device in NVL cells on the selected row, and wherein the difference between VPOS and VINH applied to the bitline complement is sufficient to inhibit the second NVM devices in NVL cells on the selected row from getting programmed.

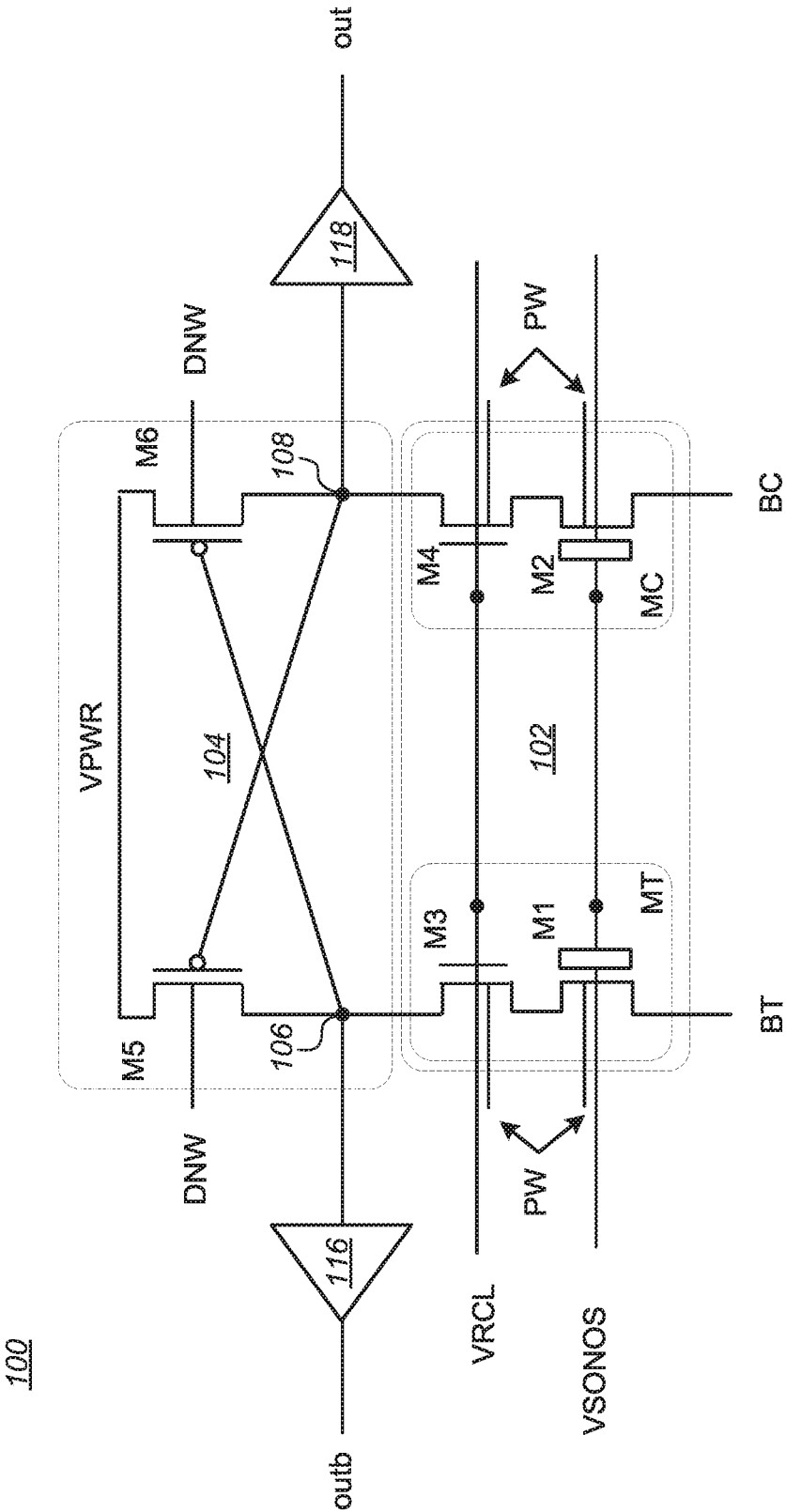


FIG. 1

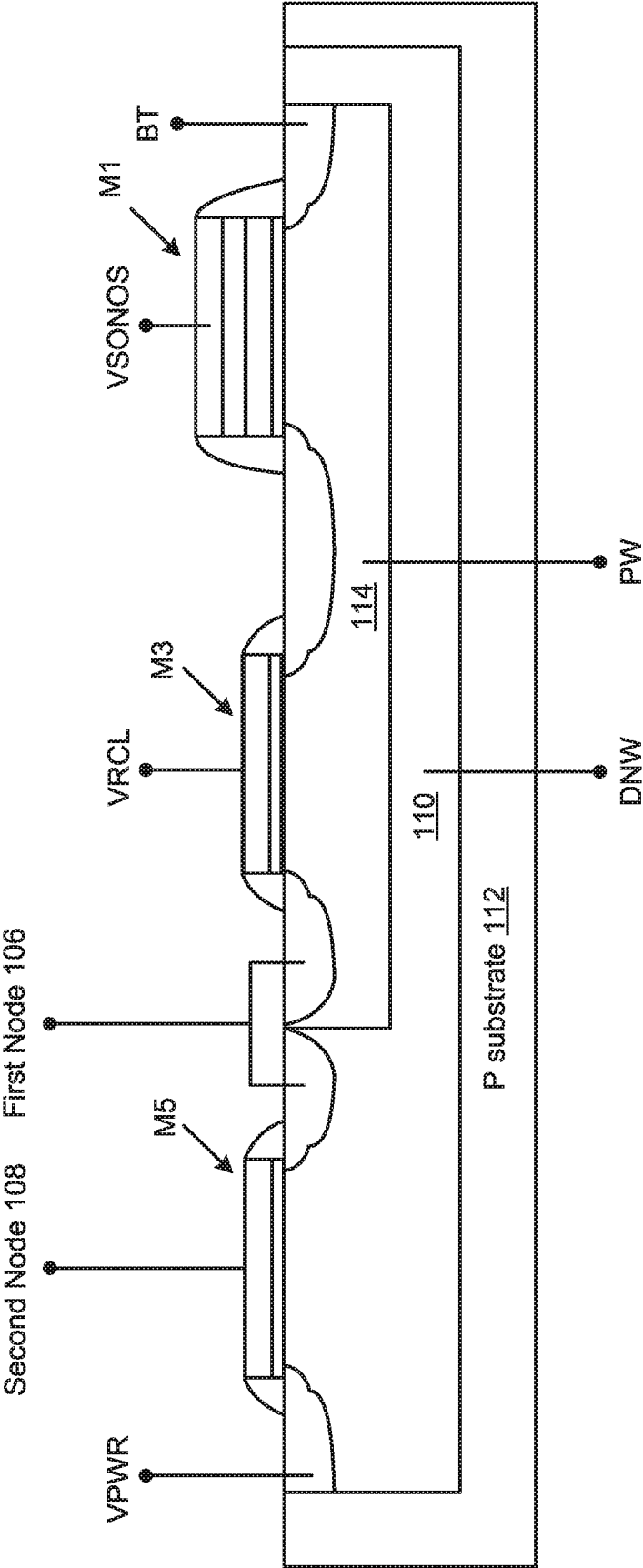


FIG. 2

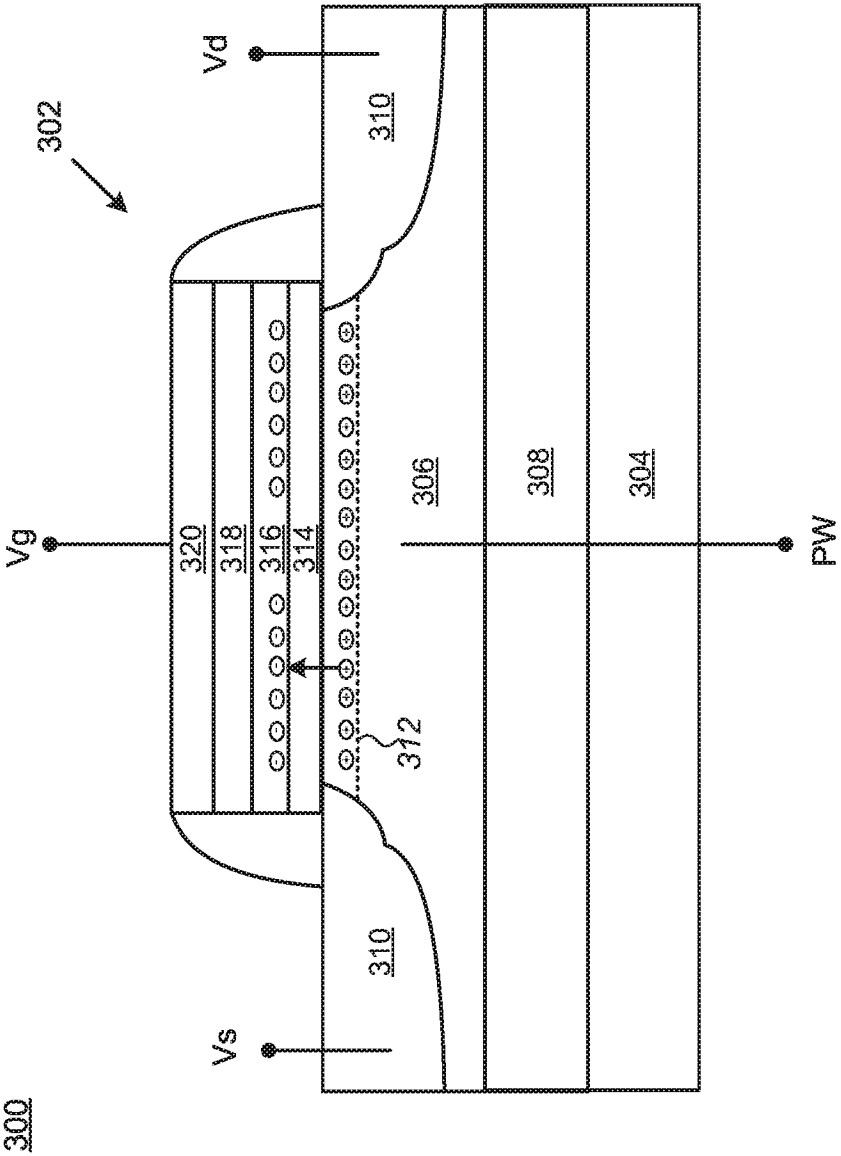


FIG. 3

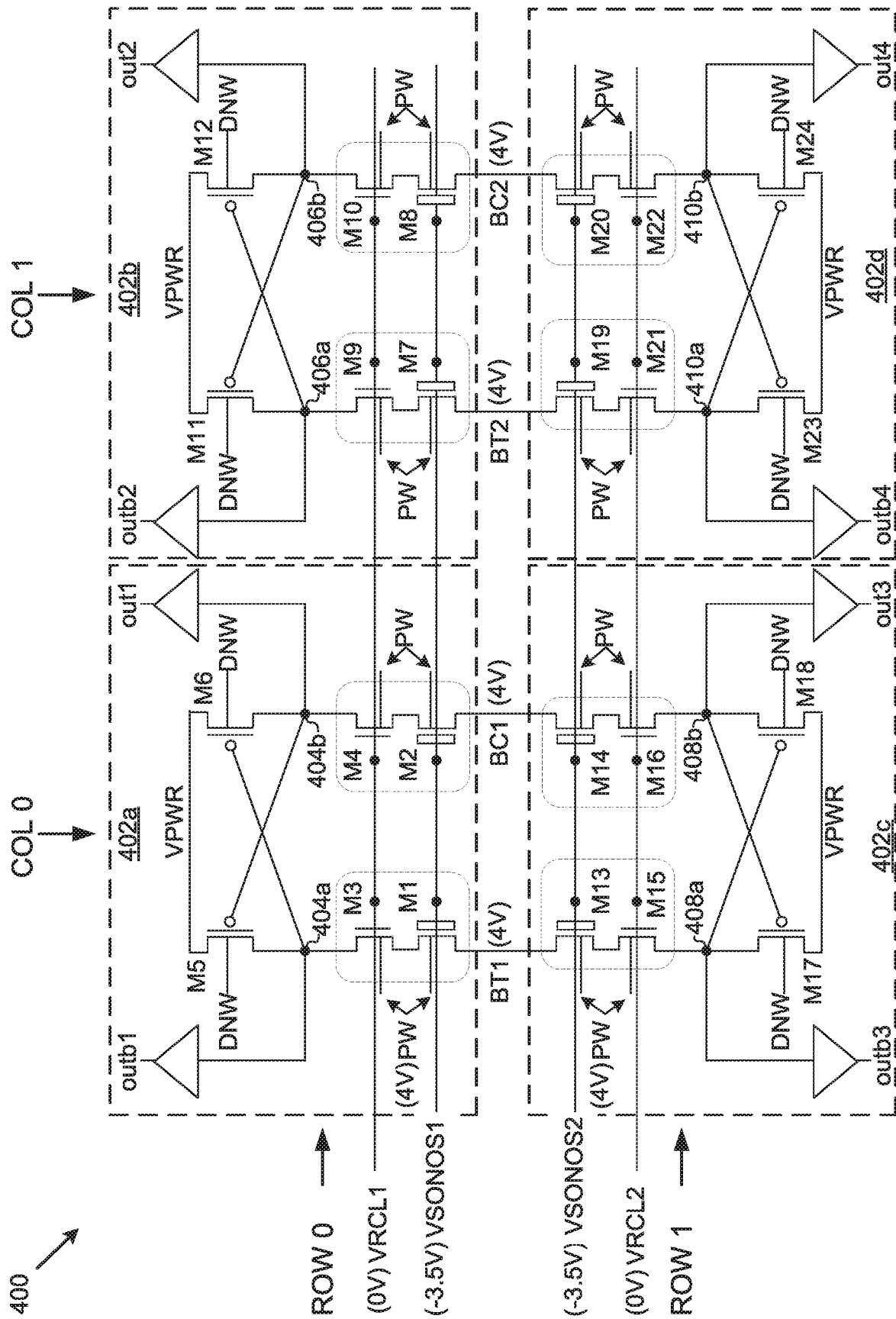


FIG. 4

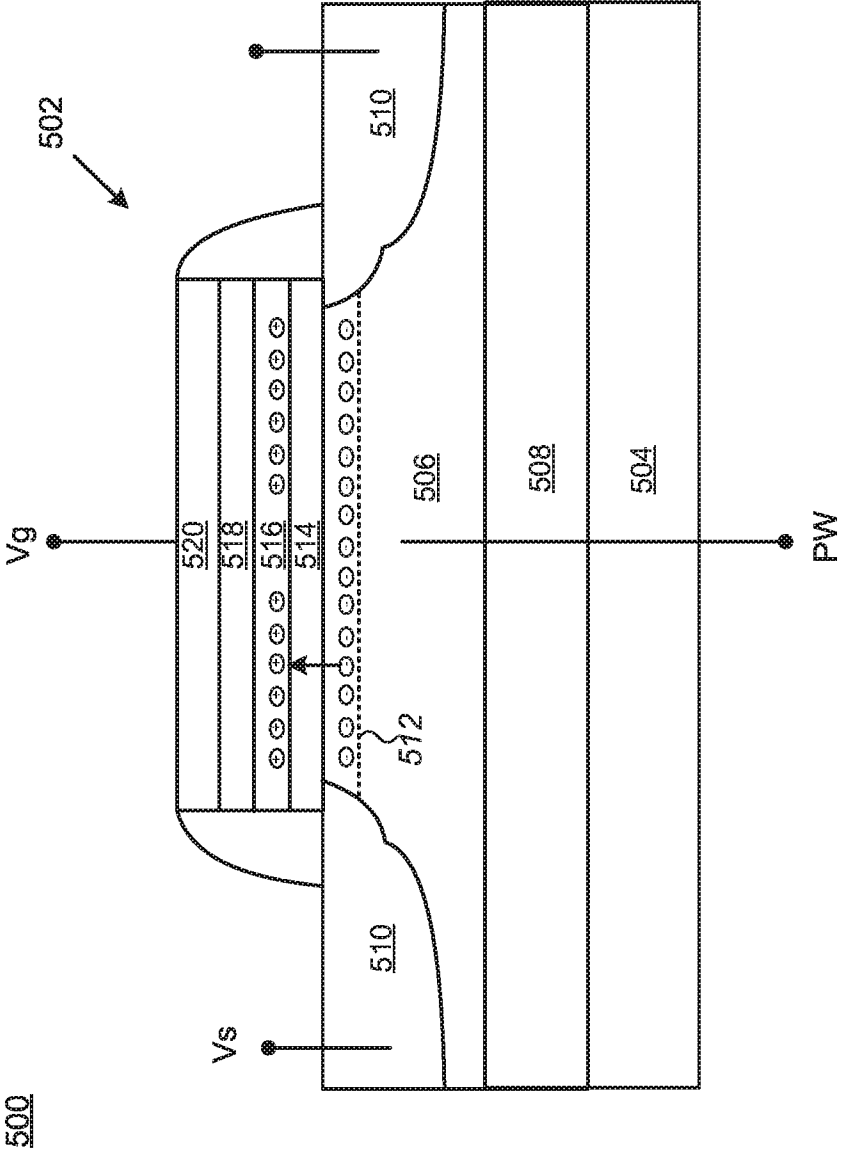
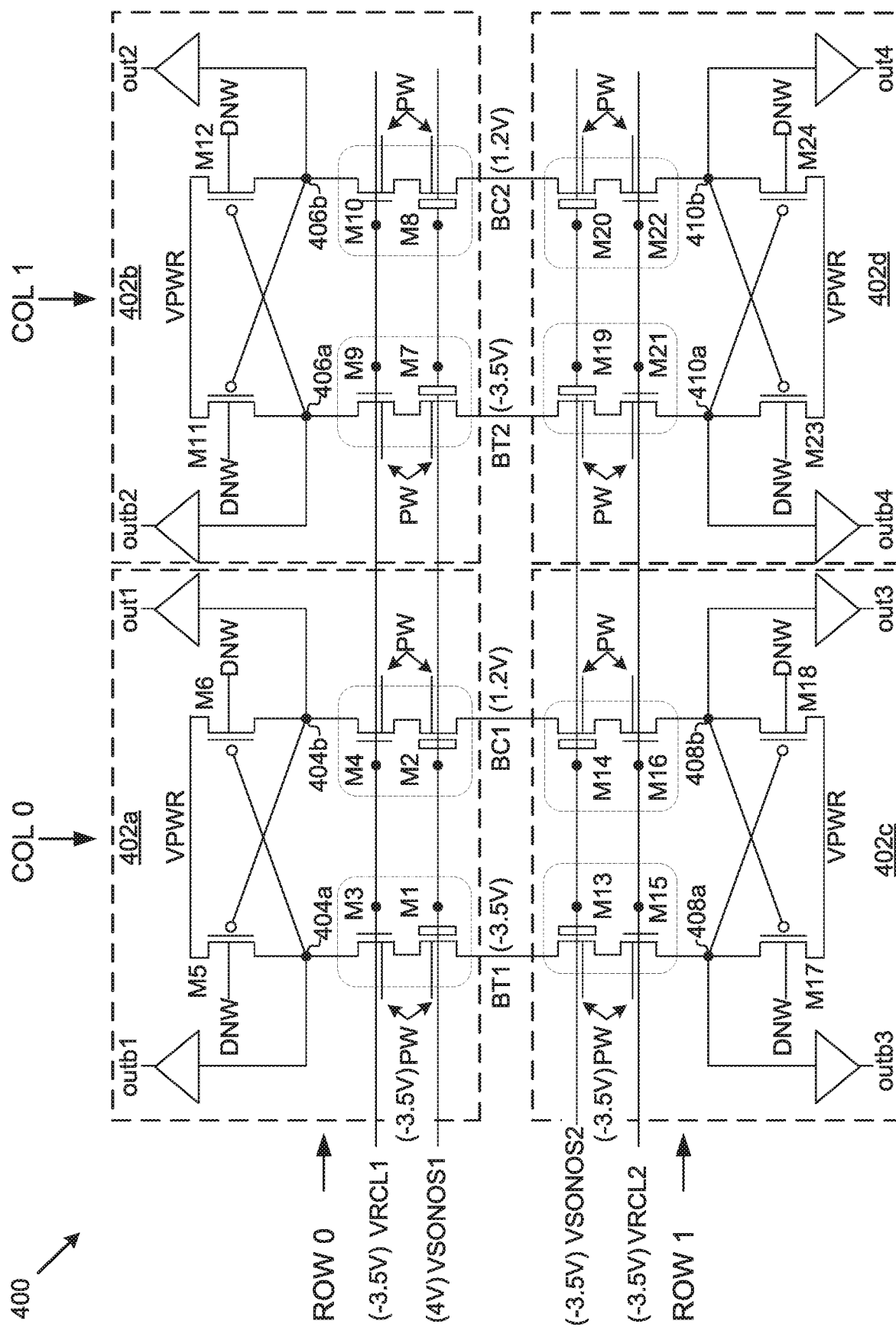


FIG. 5



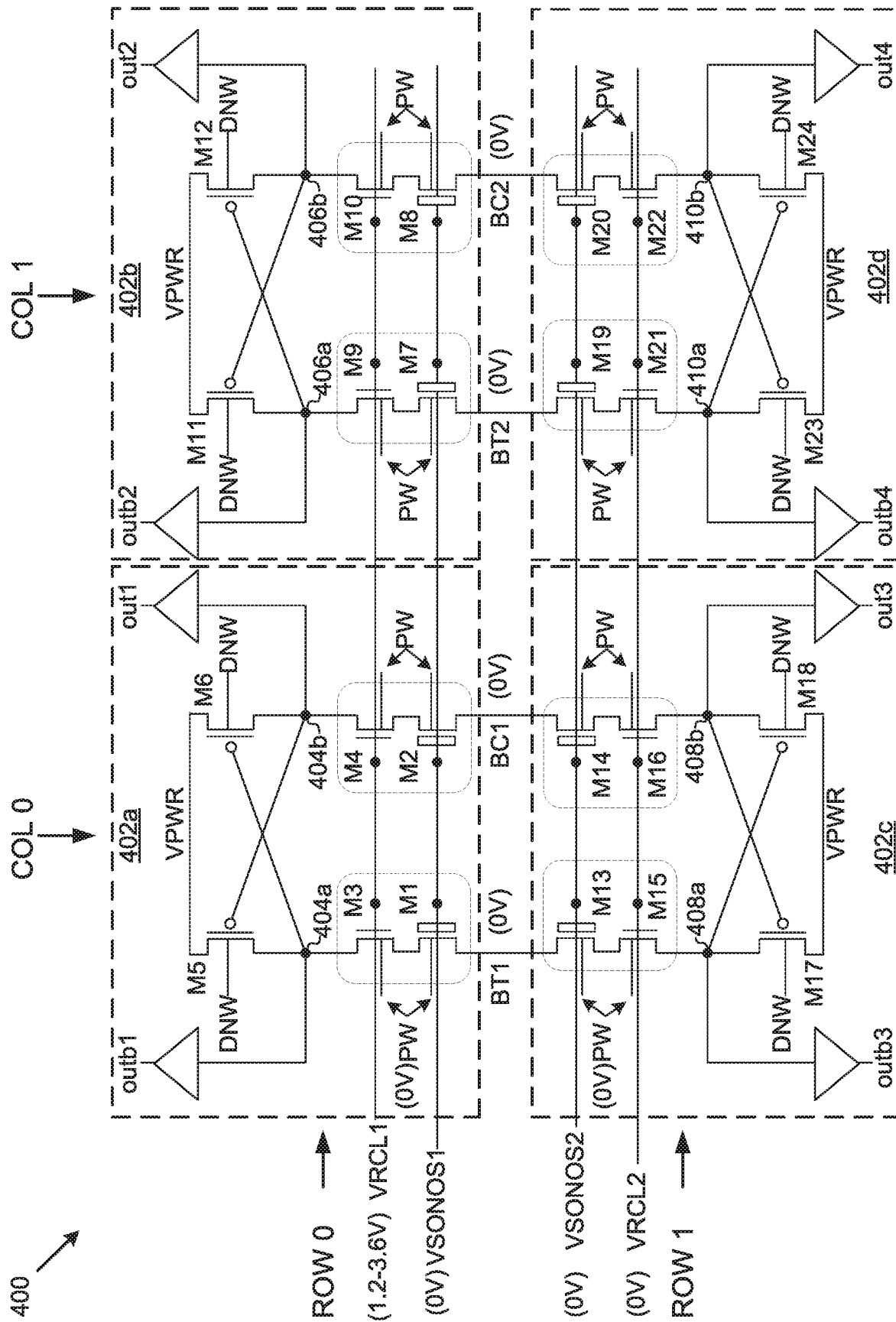


FIG. 7

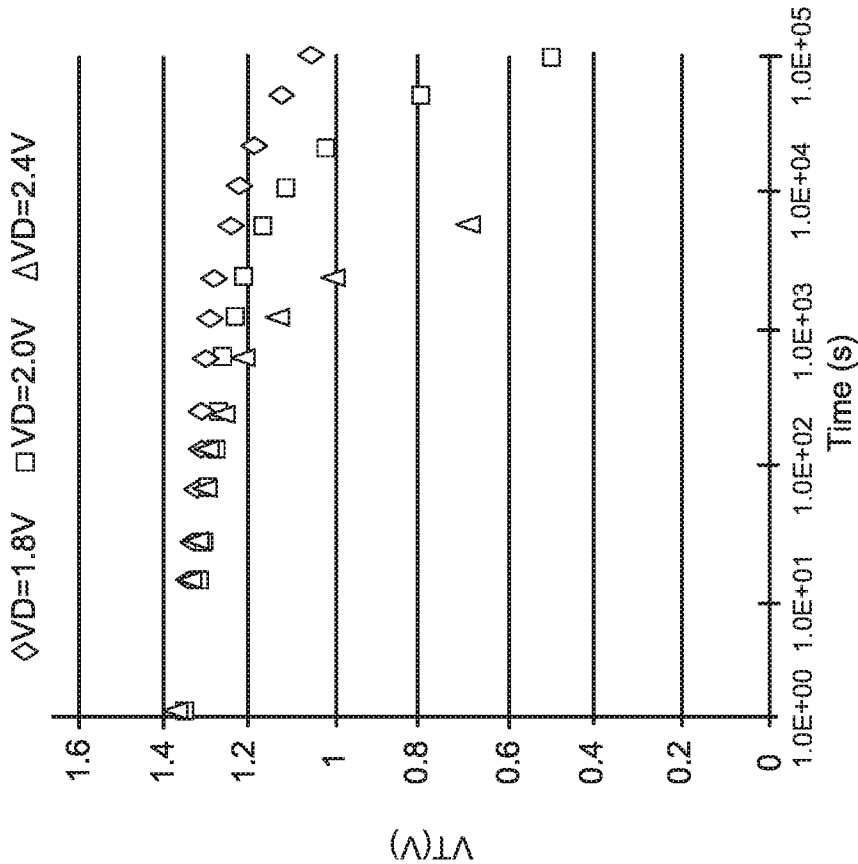


FIG. 8B

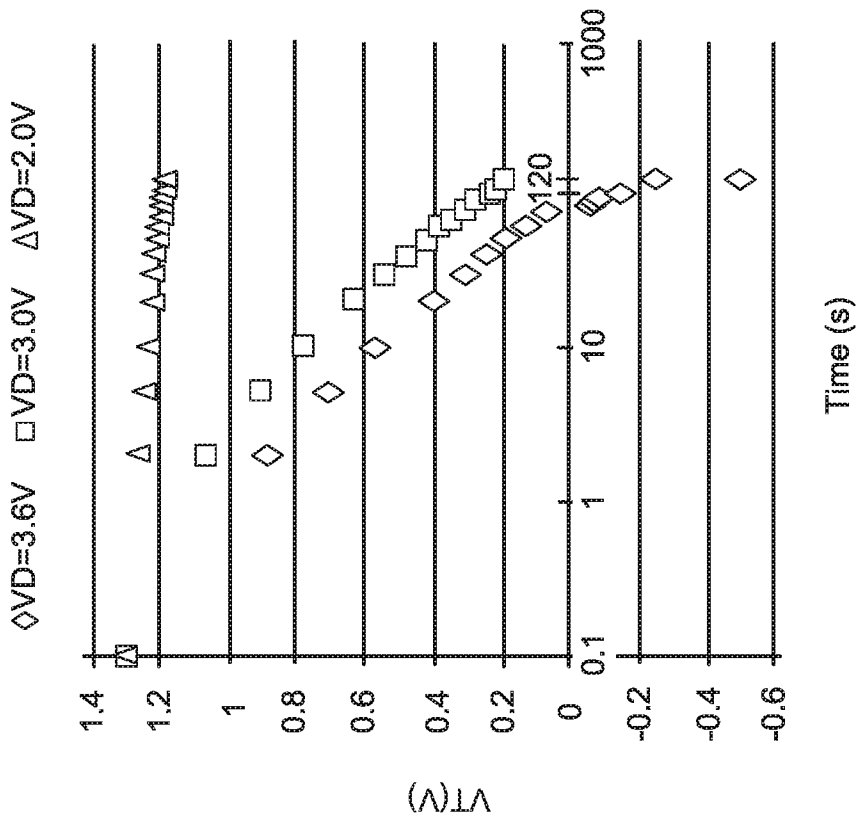


FIG. 8A

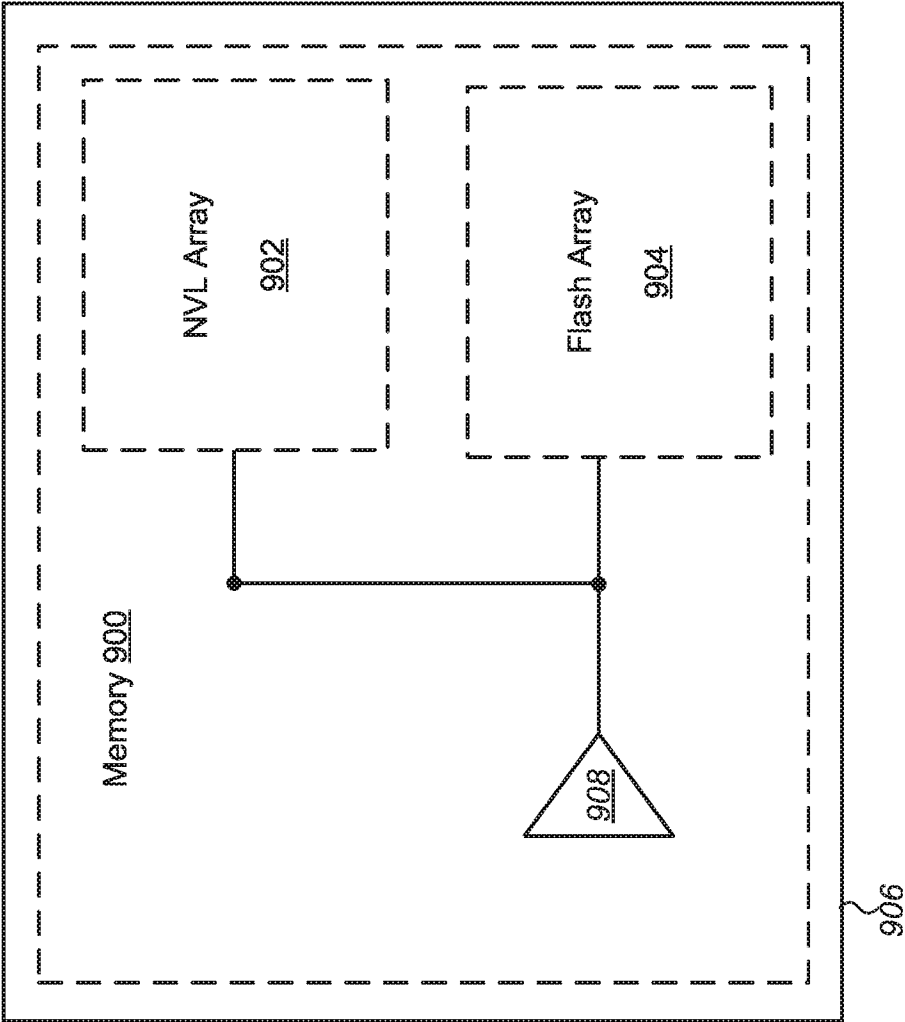


FIG. 9

INTERNATIONAL SEARCH REPORT	International application No. PCT/US16/13258	
A. CLASSIFICATION OF SUBJECT MATTER IPC: G11C 11/34(2006.01) USPC: 365/185.08 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) U.S. : 365/185.08		
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 practicable, 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.
X	US 2011/0280073 A1 (Chiu et al.) 17 November 2011, see entire document.	1
X	US 6,556,487 B1 (Ratnakumar et al.) 29 April 2003, see entire document.	1, 9
---		-----
Y		2-8, 10-19
Y	US 2006/0023503 A1 (Lee) 02 February 2006, see entire document.	2-19
Y	US 2014/0197864 A1 (Prabhakar) 17 July 2014, see entire document.	4-5, 12, 19

☐ Further documents are listed in the continuation of Box C.		☐ See patent family 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 application or patent 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 21 March 2016 (21.03.2016)		Date of mailing of the international search report 25 MAR 2016	
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (571) 273-3201		Authorized officer Ham Seungsook Telephone No. 571-272-2405	

INTERNATIONAL SEARCH REPORT	International application No. PCT/US16/13258
Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)	
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:	
1.	☐ Claims Nos.: because they relate to subject matter not required to be searched by this Authority, namely:
2.	☐ Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.	☐ Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).
Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)	
This International Searching Authority found multiple inventions in this international application, as follows:	
1.	☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.	☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of any additional fees.
3.	☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.	☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest	☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee. ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation. ☐ No protest accompanied the payment of additional search fees.