(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
16 July 2015 (16.07.2015)

WIPO | PCT

(10) International Publication Number
WO 2015/105904 A1

(51) International Patent Classification:
*H01L 23/62* (2006.01) *H01L 29/423* (2006.01)
*H01L 29/10* (2006.01)

(21) International Application Number:
PCT/US2015/010516

(22) International Filing Date:
7 January 2015 (07.01.2015)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/925,958 10 January 2014 (10.01.2014) US
14/494,917 24 September 2014 (24.09.2014) US

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

(72) Inventors: **PRABHAKAR, Venkatraman**; 3720 Appian St, Pleasanton, California 94588 (US). **KOUZNETSOV, Igor**; 1 Daniel Burnham Ct. Apt. 812, San Francisco, California 94109 (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**: DRAIN EXTENDED MOS TRANSISTORS WITH SPLIT CHANNEL

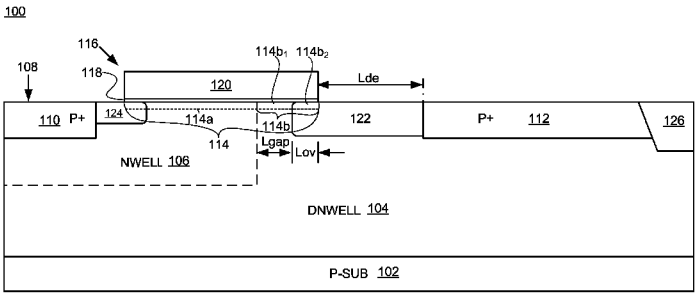



*FIG. 1A*

(57) **Abstract**: A circuit including both drain-extended metal-oxide-semiconductor (DEMOS) and low-voltage metal-oxide-semiconductor (LV_MOS) devices and methods of manufacturing the same are provided. In one embodiment, DEMOS device includes a first channel, a gate, a second channel, and a drain extension, wherein the second channel is split into a first portion and a second portion, and wherein the first portion of the second channel stops under the gate and is spaced away from the drain extension. Other embodiments are also described.

WO 2015/105904 A1

# Drain Extended MOS Transistors With Split Channel

## CROSS-REFERENCE TO RELATED APPLICATIONS

**[0001]** This application is an international application of U.S. Patent Application No. 14/494,917, filed September 24, 2014, which claims the benefit of priority under 35 U.S.C. 119(e) to U.S. Provisional Patent Application No. 61/925,958, filed January 10, 2014, all of which are incorporated by reference herein in their entirety.

## TECHNICAL FIELD

**[0002]** This disclosure relates generally to fabrication of semiconductor devices, and more particularly to drain extended metal-oxide-semiconductor (DEMOS) transistors with split channels integrally fabricated using existing complimentary metal-oxide-semiconductor (CMOS) processes and suitable for use in high-voltage (HV) circuits of devices such as Non-Volatile Memories (NVM).

## BACKGROUND

**[0003]** While many types of integrated circuits may be designed to operate with a single internal voltage, it is often desirable to provide an integrated circuit (IC) including devices (e.g., transistors as well as passive circuit elements) that operate at two or more different voltage levels. Examples of such ICs include a Non-Volatile Memories (NVM) and ICs including a NVM or a flash macro or memory, such as a micro-controller, microprocessor or programmable system on a chip (PSOC). Such a circuit typically includes low-voltage metal-oxide-semiconductor (LV_MOS) transistors used in logic and/or switching applications and designed to operate at a voltage of less than from about

2.5 to about 3.3 volts (V), and other high-voltage metal-oxide-semiconductor (HV_MOS) transistors used in NVM applications such as charge pumps, HV switches, sector selectors, input/output (I/O) cells or drivers, and typically designed to operate at voltages of about 9V or greater.

**[0004]** One approach to integrating a HV_MOS transistor into such a circuit includes introducing drain-extended (DE) architectures to provide transistors having higher breakdown voltages (BV) for use in high power and high voltage applications. Briefly, in a DE transistor the drain is extended by implanting low doped semiconductor region which depletes during reverse biasing, thereby allowing much of voltage to be dropped across a drain extension in the substrate, and thereby reducing the electric field across a gate oxide to a safe level. In existing complementary metal–oxide–semiconductor (CMOS) process flows the DE implants are done using masks and implants borrowed from other devices in baseline process. However, as the size of the transistors advance to less than 65 nanometers (nm) many of these masks and implants are not used. Instead, only well masks are used (including both deep and shallow implants) and halo/tip implants are used control a threshold voltage ($V_T$) of the transistor.

**[0005]** Thus, there is a need for a method of forming DEMOS transistors with a high BV that is compatible with process flows fabricating transistors at 65 nm and beyond. It is further desirable that the method substantially does not introduce any new mask and/or implant steps to the process flow.

## SUMMARY

**[0006]** Methods of forming circuits including non-volatile memory (NVM)

devices, low-voltage metal-oxide-semiconductor (LV_MOS) devices and high-voltage, drain extended metal-oxide-semiconductor (DEMOS) devices and methods of manufacturing the same are described. In one embodiment, the DEMOS device includes a first channel, a gate, a second channel, and a drain extension, wherein the second channel is split into a first portion and a second portion, and wherein the first portion of the second channel stops under the gate and is spaced away from the drain extension.

## 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. 1A is a block diagram illustrating a cross-sectional view of an embodiment of a drain-extended P-type MOS (DEPMOS) transistor or device including a N-type well (NWELL) stopping under a gate of the DEPMOS transistor;

**[0009]** FIG. 1B is a block diagram illustrating a cross-sectional view of another embodiment of a DEPMOS transistor including a NWELL stopping proximal to an edge of the gate of the DEPMOS transistor;

**[0010]** FIG. 2 is a block diagram illustrating a cross-sectional view of an embodiment of a drain-extended N-type MOS (DENMOS) transistor including an PWELL stopping under a gate of the DENMOS transistor;

**[0011]** FIG. 3 is a block diagram of a flash macro or memory illustrating applications for a DEPMOS transistor according to embodiments of the present disclosure;

**[0012]** FIG. 4 is a block diagram of a part of a NVM illustrating applications for a DEPMOS transistor in both a memory array and a bit line driver according to embodiments of the present disclosure;

**[0013]** FIG. 5 is a flowchart illustrating an embodiment of a method for fabricating a circuit including a non-volatile memory (NVM) transistor, a low voltage metal-oxide-semiconductor (LV_MOS) transistor and a drain-extended P-type MOS (DEPMOS) transistor using a separate implant step according to an embodiment of the present disclosure;

**[0014]** FIGs. 6A-6K are block diagrams illustrating cross-sectional views of a part of the circuit formed by the method of the flowchart of FIG. 5;

**[0015]** FIG. 7 is a flowchart illustrating a method for fabricating a circuit including a NVM transistor and a DEMOS transistor using a single implant to concurrently form a channel region and a drain extension for the DEMOS transistor according to another embodiment of the present disclosure;

**[0016]** FIG. 8 is a schematic diagram of an embodiment of an electro-static discharge (ESD) protection circuit showing a discharge path for an ESD pulse through an ESD implant surrounding a drain of an ESD transistor;

**[0017]** FIG. 9 is a flowchart illustrating an embodiment of a method for fabricating a circuit including a NVM transistor, a MOS transistor in an electrostatic discharge (ESD) circuit and a DEMOS transistor using an existing ESD implant step in the CMOS process flow; and

**[0018]** FIGs. 10A-10D are block diagrams illustrating cross-sectional views of a part of the circuit formed by the method of the flowchart of FIG. 9.

## DETAILED DESCRIPTION

**[0019]** The present disclosure is directed generally to drain-extended metal-oxide-semiconductor (DEMOS) transistors and methods for fabricating the same.

**[0020]** Various embodiments of the DEMOS transistors and methods for fabricating the same will now be described with reference to a number of diagrams. The embodiments include methods of concurrently forming a DEMOS transistor, as well as a low-voltage or regular metal-on-semiconductor (MOS) transistor, and a non-volatile memory (NVM) transistor on a single substrate for a number of different circuits and applications. In particular embodiments, the DEMOS transistor may be integrally formed in the same substrate as a NVM transistor and a LV_MOS transistor in an input/output (I/O) cell, or in a driver for the NVM device.

**[0021]** 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.

**[0022]** 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.

**[0023]** 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.

**[0024]** Drain-extended high voltage transistors are useful in applications in which it is desirable to provide higher drain breakdown voltages (BV) across the transistor, and/or in applications in which it is desirable to reduce the amount of voltage dropped across a gate to drain extension edge. Preferably, the transistors and method of present disclosure enable high drain voltage operation without modifying existing processes and device architectures, such as gate oxide thicknesses. More preferably, the DEMOS transistors are fabricated as part of a complimentary-metal-oxide-semiconductor circuit including low-voltage or regular MOS transistors, and NVM transistors fabricated simultaneously on the same semiconductor substrate or wafer using existing CMOS processes.

**[0025]** DEMOS transistors according to various embodiments of the present disclosure will now be described in greater detail with reference to FIGs. 1A to 2. In

particular, FIG. 1A is a block diagram illustrating a cross-sectional view of an embodiment of a drain-extended P-type MOS (DEPMOS) transistor or device including P-type channel (P-channel) and a N-type well (NWELL) in a deep NWELL (DNWELL), in which an interface between the NWELL and the DNWELL near a surface of the substrate stopping under a gate stack of the DEPMOS transistor. FIG. 1B is a block diagram illustrating a cross-sectional view of another embodiment of a DEPMOS transistor in which the interface between the NWELL and the DNWELL stops proximal to an edge of the gate stack of the DEPMOS transistor. FIG. 2 is a block diagram illustrating a cross-sectional view of an embodiment of a drain-extended N-type MOS (DENMOS) transistor including a PWELL in a deep PWELL (DPWELL), in which an interface between the PWELL and the DPWELL near a surface of the substrate stops under a gate stack of the DENMOS transistor.

**[0026]** Referring to FIG. 1A, the DEPMOS transistor **100** is formed in a semiconductor wafer or substrate **102**. The substrate **102** may include any known semiconductor material, such as Silicon, Gallium-arsenide, Germanium, Gallium-nitride, Aluminum-phosphide, and mixtures or alloys thereof. Preferably, the substrate **102** is a doped silicon-based semiconductor substrate, such as a P-type silicon substrate (P-substrate).

**[0027]** The DEPMOS transistor **100** includes or is formed in a deep well implanted with a N-type dopant to form a deep NWELL or DNWELL **104**. The dopants implanted may be of any suitable N-type dopant, implanted at any energy and to any concentration suitable to form a DNWELL for the DEPMOS transistor **100**. For example, in one particular embodiment, phosphorus ions are implanted at a relatively high energy

of between about 800 keV and about 1.5 MeV and to a concentration or dose of between about 0.5E13/cm$^2$ and about 5E13/cm$^2$ to form a lightly doped DNWELL.

**[0028]** The DEPMOS transistor **100** further includes a more heavily doped NWELL **106** formed in the DNWELL **104** near a surface **108** of the substrate **102** by implantation of a N-type dopant, such as arsenic or phosphorus ions, to a concentration or dose greater than that of the lightly doped DNWELL **104**. For example, in one particular embodiment, the NWELL **106** includes phosphorus ions implanted at a relatively low energy of between about 250 keV and about 500 keV and to a concentration or dose of between about 1E13/cm$^2$ and about 7E13/cm$^2$ to form the NWELL.

**[0029]** The DEPMOS transistor **100** further includes a heavily doped P+ source **110** and a heavily doped P+ drain **112** separated by a MOS channel region or channel **114** underlying a gate stack **116** including gate oxide **118** and a polycrystalline silicon (poly) or metal gate electrode **120**. By heavily doped it is meant a concentration of suitable impurities or dopant ions, such as boron ions ($B^+$) of from about 1E15/cm$^2$ to about 5E15/cm$^2$, as compared to the dopant concentration in the NWELL **106** and DNWELL **104**.

**[0030]** In accordance with the present disclosure, the DEPMOS transistor **100** further includes a drain extension **122** between the channel **114** and the drain **112** through which the drain is electrically connected to the channel. The distance or length between an edge of the gate stack **116**, which coincides with the beginning of the channel **114**, and the drain **112** is referred to as the length of the drain extension **122** or $L_{DE}$. Suitable distances or lengths for the $L_{DE}$ include from about 400 to about 1,000 nm. Generally, the drain extension **122** of a DEPMOS transistor **100** is formed by implanting or diffusing a

P-type dopant, such as B+, into the surface **108** of the substrate **102** to achieve depletion during reverse bias across the drain/DNWELL junction, thereby achieving a high junction breakdown or BV. The drain extension **122** is more lightly doped than the highly doped drain **112** so that during reverse bias depletion regions formed in the channel **114** can extend far enough to reduce the electric field per unit length and achieve a high breakdown voltage or BV towards the substrate region. Thus, the lightly doped drain extension **122** gets depleted during reverse biasing to drop most of the voltage applied to the drain **112** of the high voltage DEPMOS transistors **100**.

**[0031]** For example, a lightly doped drain extension **122** can include a concentration of boron ions ($B^+$) of from about 0.5E13/cm$^2$ to about 3E13/cm$^2$, as compared to the dopant concentration in highly doped drain **112**. It has been found that with such a lightly doped drain extension **122** an applied drain voltage of about 9 to 10 volts, approximately 6 volts can be dropped across the drain extension **122**, thereby limiting voltage across the gate oxide to about 3.6V, and enabling use of a standard low I/O voltage gate oxide **118** as isolation layer between the channel **114** and the gate electrode **120**.

**[0032]** In some embodiments, such as that shown in FIG. 1A, the drain extension **122**, overlaps the gate stack **116** and is implanted in such a way that the overlap between the drain extension and the gate stack is reduced to trade off an increase in breakdown voltage against a decrease in on-state current for shorter channel length devices. Suitable distances or lengths for the overlap ($L_{OV}$) include from about 50 to about 100 nm.

**[0033]** Optionally, in some embodiments the DEPMOS transistor **100** may further include a lightly doped diffusion (LDD) **124** between the channel **114** and the source **110**

through which the source is electrically connected to the channel. As with the drain extension **122**, the LDD **124** can be formed either by implanting or diffusing ions of the appropriate dopant species, B+ for the DEPMOS shown, to a suitable concentration and a suitable depth. For example, the LDD **124** can include a concentration of boron ions ($B^+$) of from about 0.5E14/$cm^2$ to about 5E14/$cm^2$, as compared to the dopant concentration in highly doped source **110**, and can extend from edge of the gate stack **116** or the beginning of the channel **114** to the source a distance or length of from about 10 to about 45 nm. Additionally, as with the drain extension **122,** in some embodiments the LDD **124** can overlap the gate stack **116** as shown in FIG. 1A, while in other embodiments it is formed in such a way as to end at the edge of the gate stack **116** or the beginning of the channel **114**.

**[0034]** Generally, the DEPMOS transistor **100** further includes an isolation structure **126**, such as a shallow-trench-isolation (STI) structure, which serves to isolate or separate the DEPMOS transistor from one or more adjacent devices or transistors. Although shown here as a STI structure, it will be appreciated that the isolation structure **126** can also include a field oxidation region or structure (FOX), or a local oxidation of silicon (LOCOS) region or structure.

**[0035]** In accordance with another aspect of the present disclosure shown in FIG. 1A, the DEPMOS transistor **100** includes an interface between the NWELL **106** and the DNWELL **104** near the surface **108** of the substrate **102** that underlies the gate stack **116** of the DEPMOS transistor to split the channel **114** into a first channel **114a** encompassed by or within the NWELL and a second channel **114b** encompassed by or within the DNWELL. In certain embodiments, such as that shown in FIG. 1A, the second channel

**114b** is further split into a first portion **$114b_1$** encompassed by or within only the DNWELL **104** and a second portion **$114b_2$** encompassed by or within the drain extension **122** in the DNWELL. The embodiment will enable fabrication of a DEPMOS transistor **100** even with a baseline CMOS process that typically produces low junction BV. The BV of the DEPMOS **100** will be substantially higher than for devices and transistors produced by the baseline process since the drain **112** and the drain extension **122** are in the DNWELL **104** rather than NWELL as in the prior art approach.

**[0036]** In another embodiment, shown in FIG. 1B, the interface between the NWELL **106** and the DNWELL **104** does not underlie the gate stack **116** of the DEPMOS transistor **100**, such that substantially the entire channel **114** consists of the first channel **114a** encompassed by or within the NWELL **106**. It is noted that this embodiment also will still provide a BV of the DEPMOS **100** significantly higher than for a baseline CMOS process since the drain **112** is spaced away from the NWELL **106**, although the drain extension **122** remains touching the NWELL.

**[0037]** FIG. 2 is a block diagram illustrating a cross-sectional view of an embodiment of a drain-extended N-type MOS (DENMOS) transistor including a PWELL in a deep PWELL (DPWELL), in which an interface between the PWELL and the DPWELL near a surface of the substrate stops under a gate of the DENMOS transistor.

**[0038]** Referring to FIG. 2, the DENMOS transistor **200** is formed in a semiconductor wafer or substrate **202**. As with the DEPMOS transistor **100** described above, the substrate **202** may include any known semiconductor material, such as Silicon, Gallium-arsenide, Germanium, Gallium-nitride, Aluminum-phosphide, and mixtures or alloys thereof. Preferably, the substrate **202** is a doped silicon-based semiconductor

substrate, such as a N-type silicon substrate (n-substrate).

**[0039]** The DENMOS transistor **200** includes or is formed in a deep well implanted with a P-type dopant to form a deep PWELL or DPWELL **204**. The dopants implanted may be of any suitable P-type dopant, implanted at any energy and to any concentration suitable to form a DPWELL for the DENMOS transistor **200**. For example, in one particular embodiment, boron ions (B+) are implanted at a relatively high energy of between about 300 keV and about 650 keV and to a concentration or dose of between about 0.5E13/cm$^2$ and about 5 E13/cm$^2$ to form a lightly doped DPWELL.

**[0040]** The DENMOS transistor **200** further includes a more heavily doped PWELL **206** formed in the DPWELL **204** near a surface **208** of the substrate **202** by implantation of a P-type dopant, such as boron ions, to a concentration or dose greater than that of the lightly doped DPWELL 2**04**. For example, in one particular embodiment, the PWELL **206** includes B+ ions implanted at a relatively low energy of between about 80 keV and about 400 keV and to a concentration or dose of between about 1E13/cm$^2$ and about 7E13/cm$^2$ to form the PWELL.

**[0041]** The DENMOS transistor **200** further includes a heavily doped N+ source **210** and a heavily doped N+ drain **212** separated by a channel **214** underlying a gate stack **216** including gate oxide **218** and a polycrystalline silicon (poly) or metal gate electrode **220**. By heavily doped it is meant a concentration of suitable impurities or dopant ions, such as arsenic or phosphorous ions of from about 1E15/cm$^2$ to about 5E15/cm$^2$, as compared to the dopant concentration in the PWELL **206** and DPWELL **204**.

**[0042]** In accordance with the present disclosure, the DENMOS transistor **200** further includes a drain extension **222** between the channel **214** and the drain **212** through

which the drain is electrically connected to the channel. The distance or length between an edge of the gate stack **216**, which coincides with the beginning of the channel **214**, and the drain **212** is referred to as the length of the drain extension **222** or $L_{DE}$. Suitable distances or lengths for the $L_{DE}$ include from about 400 to about 1,000 nm. Generally, the drain extension **222** of a DENMOS transistor **200** is formed by implanting or diffusing a N-type dopant, such as arsenic or phosphorous ions, into the surface **208** of the substrate **202** to achieve depletion during reverse bias, thereby achieving a high junction breakdown or BV. The drain extension **222** is more lightly doped than the highly doped drain **212** so that during reverse bias across the N+/DPWELL junction, depletion regions formed in the channel **214** can extend far enough to reduce the electric field per unit length and achieve a high breakdown voltage or BV towards the substrate region. Thus, the lightly doped drain extension **222** gets depleted during reverse biasing to drop most of the voltage applied to the drain **212** of the high voltage DENMOS transistors 2**00**.

**[0043]** For example, a lightly doped drain extension **222** can include a concentration of arsenic or phosphorous ions of from about 0.5E13/$cm^2$ to about 3E13/$cm^2$, as compared to the dopant concentration in highly doped drain **212**. It has been found that with such a lightly doped drain extension **222** an applied drain voltage of about 9 to 10 volts, approximately 6 volts can be dropped across the drain extension **222**, thereby limiting voltage across the gate oxide to about 3.6V, and enabling use of a standard low I/O voltage gate oxide **218** as isolation layer between the channel **214** and the gate electrode **220**.

**[0044]** In some embodiments, such as that shown in FIG. 2, the drain extension **222**, overlaps the gate stack **216** and is implanted in such a way that the overlap between

the drain extension and the gate stack is reduced to trade off an increase in breakdown voltage against a decrease in on-state current for shorter channel length devices. Suitable distances or lengths for the overlap ($L_{OV}$) include from about 50 to about _ 100 nm.

**[0045]** Optionally, in some embodiments the DENMOS transistor **200** may further include a lightly doped diffusion (LDD) **224** between the channel 2**14** and the source **210** through which the source is electrically connected to the channel. As with the drain extension **222**, the LDD **224** can be formed either by implanting or diffusing ions of the appropriate dopant species, arsenic or phosphorous ions for the DENMOS shown, to a suitable concentration and a suitable depth. For example, the LDD **224** can include a concentration of arsenic or phosphorous ions of from about 0.3E14/$cm^2$ to about 3E14/$cm^2$, as compared to the dopant concentration in highly doped source **210**, and can extend from edge of the gate stack **216** or the beginning of the channel **214** to the source a distance or length of from about 10 to about 45 nm. Additionally, as with the drain extension **222,** in some embodiments the LDD **224** can overlap the gate stack **216** as shown in FIG. 2, while in other embodiments it is formed in such a way as to end at the edge of the gate stack **216** or the beginning of the channel **214**.

**[0046]** Generally, the DENMOS transistor **200** further includes an isolation structure **226**, such as a shallow-trench-isolation (STI) structure, which serves to isolate or separate the DENMOS transistor from one or more adjacent devices or transistors. Although shown here as a STI structure, it will be appreciated that the isolation structure **226** can also include a field oxidation region or structure (FOX), or a local oxidation of silicon (LOCOS) region or structure.

**[0047]** In accordance with another aspect of the present disclosure shown in FIG.

2, the DENMOS transistor **200** includes an interface between the PWELL **206** and the DPWELL **204** near the surface **208** of the substrate **202** that underlies the gate stack **216** of the DENMOS transistor to split the channel **214** into a first channel **214a** encompassed by or within the PWELL and a second channel **214b** encompassed by or within the DPWELL. In certain embodiments, such as that shown in FIG. 2, the second channel **214b** is further split into a first portion $\mathbf{214b_1}$ encompassed by or within only the DPWELL **204** and a second portion $\mathbf{214b_2}$ encompassed by or within the drain extension **222** in the DPWELL. The embodiment will enable fabrication of a DENMOS transistor **200** even with a baseline CMOS process that typically produces low junction BV. The BV of the DENMOS **200** will be substantially higher than for devices and transistors produced by the baseline process since the drain **212** and the drain extension **222** are in the DPWELL **204** rather than PWELL as in the prior art approach.

**[0048]** In another embodiment, not shown, the interface between the PWELL **206** and the DPWELL **204** does not underlie the gate stack **216** of the DENMOS transistor **200**, such that substantially the entire channel **214** consists of the first channel **214a** encompassed by or within the PWELL **206**. It is noted that this embodiment also will still provide a BV of the DENMOS **200** significantly higher than for the baseline process since the drain **212** is spaced away from the PWELL **206**, but the drain extension **222** remains touching the PWELL.

**[0049]** FIG. 3 is a block diagram of a flash memory or flash macro **300** illustrating applications for a DEMOS transistors according to embodiments of the present disclosure. Referring to FIG. 3, it is noted that high voltage DEMOS transistors can be used in at least three separate sub-circuits. In particular, it has been found that HV

capabilities of the DEMOS transistors can be advantageously used in HV sector select/CMUXs **302**, in HV page latches **304** and in HV row drivers **306**. As shown in FIG. 3, the HV page latches **304** include latches coupled to multiple global bit lines (GBL) from GBL 0 to GBL N-1, where N is the number of columns in the flash macro **300**. The flash macro **300** further includes multiple sectors 308, only two of which are shown, each having a width of N columns and coupled to HV sector select/CMUXs **302**, HV page latches **304** and HV row drivers **306**.

**[0050]** FIG. 4 is a layout of a part of a NVM **400** illustrating applications for a DEMOS transistor **402** according to embodiments of the present disclosure in a bit line driver **404**. The bit line driver **404** is laid out on-pitch with a memory array **406** including a plurality of cells **408**.

**[0051]** An embodiment of a method for fabricating drain-extended metal-oxide-semiconductor (DEMOS) transistors having a multi-well and/or split channel architecture will now be described in detail with reference to FIG. 5 and FIGs. 6A through 6K. FIG. 5 is a flowchart illustrating an embodiment of a method or process flow for fabricating a memory or circuit including a non-volatile memory (NVM) transistor, a metal-oxide-semiconductor (MOS) transistor and a DEMOS transistor using a separate implant step to form the drain extension. FIGs. 6A through 6K are block diagrams illustrating cross-sectional views of a part of the circuit or memory cell **600** formed by the method of the flowchart of FIG. 5. Generally, the DEMOS transistor can be fabricated with little to no impact on an existing CMOS process flow used to fabricate a NVM transistor and/or other MOS transistors (not shown in these figures). The NVM transistor may include memory transistors using Silicon-Oxide-Nitride-Oxide-Silicon (SONOS) or Metal-

Oxide-Nitride-Oxide-Silicon (MONOS) technology.

**[0052]** Referring to FIG. 5 and FIG. 6A, the process begins with forming a number of isolation structures **602** in a wafer or substrate **604** (step **502**). The isolation structures **602** isolate the memory cell being formed from other memory cells formed in adjoining areas (not shown) of the substrate **604** and/or isolate the NVM transistor being formed in an NVM region **608** of the substrate from one or more MOS transistors, including the DEMOS transistor of the present disclosure, being formed in one or more adjoining MOS regions **612**. The isolation structures **602** include a dielectric material, such as an oxide or nitride, and may be formed by any conventional technique, including but not limited to shallow trench isolation (STI) or local oxidation of silicon (LOCOS).

**[0053]** The substrate **604** may be a bulk wafer composed of any single crystal or polycrystalline material suitable for semiconductor device fabrication, or may include a top epitaxial layer of a suitable material formed on a substrate. Suitable materials include, but are not limited to, silicon, germanium, silicon-germanium or a III-V compound semiconductor material.

**[0054]** A pad oxide **614** is formed over a surface **616** of the substrate **604** in both the NVM region **608** and the MOS region **612**. The pad oxide **614** can be silicon dioxide ($SiO_2$) having a thickness of from about 10 nanometers (nm) to about 20 nm and can be grown by a thermal oxidation process or in-situ steam generation (ISSG).

**[0055]** Referring again to FIG. 5 and FIG. 6B, dopants are then implanted into substrate **604** through the pad oxide **614** to form a lightly doped deep well (DWELL **618**), a more heavily doped well (WELL **620**) doped with the same type of dopant as the DWELL, and channels **622** and **624** for the NVM transistor and the DEMOS transistor

that will be formed in the NVM region **608** and MOS region **612** (step **504**). Generally, this involves several separate deposition, lithography, implant and stripping processes to implant wells and channels for different types of devices, formed in different areas, i.e., the core or I/O circuit of the NVM region **608** and the MOS region **612**.

**[0056]** To form the DWELL **618** a photoresist layer is deposited and patterned using standard lithographic techniques to expose the MOS region **612** and an appropriate ion species implanted at an appropriate energy and concentration. For example, to form the DWELL **618** of a P-type DEMOS (DEPMOS) transistor in the MOS region **612**, arsenic or phosphorous ions can be implanted at an energy of from about 800 to about 1,500 kilo-electron volts (keV), and a dose of from about 0.5E13cm$^2$ to about 5E13/cm$^2$ to form a deep well. A DWELL **618** of a N-type DEMOS (DENMOS) transistor may likewise be formed by implantation of boron or $BF_2$ ions at a suitable dose and energy. Optionally, as in the embodiment shown, the DWELL **618** may be concurrently formed in the NVM region **608** and/or in other MOS regions (not shown).

**[0057]** Next, channels **622** and **624** for the NVM transistor and the DEMOS transistor are formed. The channels **622** and **624** can be formed immediately following or preceding the formation of the DWELL **618** using the same mask used for forming the DWELL, or a using separate mask formed using standard lithographic techniques as described above. The channels **622** and **624** are then formed by implanting an appropriate ion species at an appropriate energy to an appropriate concentration. For example, to perform implants of a P-type DEMOS (DEPMOS) transistor boron or $BF_2$ ions can be implanted at an energy of from about 10 to about 100 kilo-electron volts (keV), and a dose of from about 1E12/cm$^2$ to about 1E14/cm$^2$. Similarly, a N-type DEMOS

(DENMOS) transistor may likewise be formed by implantation of arsenic or phosphorous ions at a suitable dose and energy. It is to be understood that implantation can be used to form channels, in both the MOS region **612** and in the NVM region **608** at the same time, or at separate times using standard lithographic techniques, including a patterned photoresist layer to mask one of the MOS regions.

**[0058]** Following the implant of the DWELL **618**, and either before or after the implants forming the channels **622** and **624** for the NVM transistor and the DEMOS transistor, the WELL **620** is formed by implanting an appropriate ion species at an appropriate energy and concentration. For example, to form the WELL **620** of a P-type DEMOS (DEPMOS) transistor in the MOS region **612**, arsenic or phosphorous ions can be implanted at an energy of from about 250 to about 500 kilo-electron volts (keV), and a dose of from about 1E13/$cm^2$ to about 7E13/$cm^2$ to form an NWELL. A NWELL **620** of a N-type DEMOS (DENMOS) transistor may likewise be formed by implantation of boron or $BF_2$ ions at a suitable dose and energy.

**[0059]** In the embodiment shown in FIG. 6B, and described above, following the WELL and channel implants, the channel **624** for the DEMOS to be formed in the MOS region **612** includes a first channel **624a** encompassed by or within the WELL **620** and a second channel **624b** encompassed solely by or within the DNWELL **618**.

**[0060]** After the implants have been performed, any remaining photoresist (PR) or mask is stripped either in an ashing process using oxygen plasma, or in a photoresist strip using a commercially available strip or solvent.

**[0061]** Next, referring to FIG. 5 and FIGs. 6C-6D a hardmask (HM) **626** is deposited and patterned to expose the surface **616** of the substrate **604** in the NVM region

**608**, the pad oxide **614** in the NVM region removed and the HM removed (step **506**).The HM **626** can include one or more layers of material that can be patterned or opened using photoresist and standard lithographic techniques, but which is not itself photosensitive and protects underlying surface and structures formed therein from the photoresist and lithographic processes as well as from implants and etch process performed through openings formed in the hardmask. Suitable materials for the HM **626** include, for example, a layer of from about 5 to about 20 nm of silicon nitride ($Si_xN_y$), or silicon oxynitride (SiON) deposited by any known nitride deposition process. A photoresist layer (not shown), commonly referred to as a tunnel mask or TUNM is formed over substantially an entire surface of the HM **626** and patterned using standard lithographic techniques. The exposed part of the HM **626** is etched or removed using any suitable wet or dry etching technique, depending on the material of the hardmask and the underlying structures or layers. For example, in those embodiments in which the HM **626** includes a layer of silicon nitride overlying a pad oxide **614**, the HM can be etched using a standard low pressure nitride etch at a moderate power (about 500W) in a plasma of a fluorine containing gas, such as $CF_4$, or $CHF_3$, which exhibits good selectivity to silicon oxides, such as the silicon dioxide (SiO2) of the underlying pad oxide and/or the STI **602** structures.

**[0062]** Next, the pad oxide **614** over both the NVM region **608** is removed or etched as shown in FIG. 6C. The etch can be accomplished, for example, in a wet clean process using a 10:1 buffered oxide etch (BOE) containing a surfactant. Alternatively, the wet clean process can be performed using a 20:1 BOE wet etch, a 50:1 hydrofluoric (HF) wet etch, a pad etch, or any other similar hydrofluoric-based wet etching chemistry.

**[0063]** Referring to FIG. 6D, the HM **626** is substantially entirely stripped or removed. The HM **626** can be removed using the same process and chemistry previously used to form openings in the hardmask. Alternatively, the silicon nitride can also be removed by a wet etch using Phosphoric acid ($H_3PO_4$) at a temperature of about 150°C to 160°C.

**[0064]** Referring to FIG. 5 and FIGs. 6E and 6F, a number of dielectric or oxide-nitride-oxide (ONO) layers, shown collectively as dielectric layers **628** in FIG. 6F, are formed or deposited over the surface **616** of the substrate **604**, a mask (not shown) formed on or overlying the dielectric layers, and the dielectric layers etched to form ONO or dielectric stacks **630** of the NVM transistor to be formed in the NVM region **608** (step **508**).

**[0065]** Referring to FIG. 6F, the number of dielectric layers **628** includes a tunneling layer **632** overlying the surface **616** of the substrate **604**, a charge-trapping layer **634** overlying the tunneling layer and a blocking layer **636** overlying the charge-trapping layer. The tunneling layer **632** may be any material and have any thickness suitable to allow charge carriers to tunnel into an overlying charge-trapping layer under an applied gate bias while maintaining a suitable barrier to leakage when the NVM transistor is unbiased. In certain embodiments, tunneling layer **632** is silicon dioxide, silicon oxy-nitride, or a combination thereof and can be grown by a thermal oxidation process, using ISSG or radical oxidation, and having a thickness of from about 1.0 nanometers (nm) to about 3.0 nm by oxidation and consumption of the exposed surface of substrate.

**[0066]** Referring again to FIG. 6F, a charge-trapping layer **634** is formed on or

overlying the tunneling layer **632**. Generally, as in the embodiment shown, the charge-trapping layer is a multilayer charge-trapping layer having multiple layers including at least an oxygen-rich, substantially charge trap free lower or first charge-trapping layer **634a** closer to the tunneling layer **632**, and an upper or second charge-trapping layer **634b** that is silicon-rich and oxygen-lean relative to the first charge-trapping layer and comprises a majority of a charge traps distributed in multilayer charge-trapping layer.

**[0067]** The first charge-trapping layer **634a** of a multilayer charge-trapping layer **634** can include a silicon nitride ($Si_3N_4$), silicon-rich silicon nitride or a silicon oxy-nitride ($SiO_xN_y$). For example, the first charge-trapping layer **634a** can include a silicon oxynitride layer having a thickness of between about 1.5 nm and about 4.0 nm formed by a CVD process using dichlorosilane (DCS)/ammonia ($NH_3$) and nitrous oxide ($N_2O$)/$NH_3$ gas mixtures in ratios and at flow rates tailored to provide a silicon-rich and oxygen-rich oxynitride layer.

**[0068]** The second charge-trapping layer **634b** of the multilayer charge-trapping layer is then formed over the first charge-trapping layer **634a**. The second charge-trapping layer **634b** can include a silicon nitride and silicon oxynitride layer having a stoichiometric composition of oxygen, nitrogen and/or silicon different from that of the first charge-trapping layer **634a**. The second charge-trapping layer **634b** can include a silicon oxynitride layer having a thickness of between about 2.0 nm and about 10.0 nm, and may be formed or deposited by a CVD process using a process gas including DCS/$NH_3$ and $N_2O$/$NH_3$ gas mixtures in ratios and at flow rates tailored to provide a silicon-rich, oxygen-lean top nitride layer.

**[0069]** As used herein, the terms "oxygen-rich" and "silicon-rich" are relative to a stoichiometric silicon nitride, or "nitride," commonly employed in the art having a composition of ($Si_3N_4$) and with a refractive index (RI) of approximately 2.0. Thus, "oxygen-rich" silicon oxynitride entails a shift from stoichiometric silicon nitride toward a higher wt. % of silicon and oxygen (i.e. reduction of nitrogen). An oxygen-rich silicon oxynitride film is therefore more like silicon dioxide and the RI is reduced toward the 1.45 RI of pure silicon dioxide. Similarly, films described herein as "silicon-rich" entail a shift from stoichiometric silicon nitride toward a higher wt. % of silicon with less oxygen than an "oxygen-rich" film. A silicon-rich silicon oxynitride film is therefore more like silicon and the RI is increased toward the 3.5 RI of pure silicon.

**[0070]** Referring again to FIG. 6F, the number of dielectric layers further includes a blocking dielectric layer or blocking layer **636** that is formed on or overlying the charge-trapping layer **634**. In one embodiment, the blocking layer **636** can include an oxidized part of the silicon nitride of the underlying second charge-trapping layer **634b**, which is subsequently oxidized by in-situ-steam-generation (ISSG), or radical oxidation to form the blocking layer **636**. In other embodiments, the blocking layer **636** can include a silicon oxide ($SiO_2$) or a silicon oxynitride (SiON), deposited by CVD, and performed in a batch or single substrate processing chamber with or without an ignition event such as plasma. In one embodiment, the blocking layer **636** can include a silicon nitride, a silicon-rich silicon nitride or a silicon-rich silicon oxynitride layer having a thickness of between 2.0 nm and 4.0 nm formed by a CVD process using $N_2O/NH_3$ and $DCS/NH_3$ gas mixtures.

**[0071]** Referring to FIGs. 5 and 6G, a gate oxide or GOX preclean is performed, and a gate oxide **638** formed for the DEMOS transistor to be formed in the MOS region **612** (step **510**). Referring to FIG. 6G, the GOX preclean removes the remaining pad oxides **614** in the MOS region **612** in a highly selective cleaning process. This cleaning process prepares the substrate **604** in the MOS region **612** for gate oxide growth. In one exemplary implementation the pad oxide **614** is removed in a wet clean process. Alternatively, the wet clean process can be performed using a 20:1 BOE wet etch, a 50:1 hydrofluoric (HF) wet etch, a pad etch, or any other similar hydrofluoric-based wet etching chemistry. In other embodiments, the cleaning process chemistry is chosen so as to remove only a negligible portion of the blocking layer **636** of the dielectric stack **630**.

**[0072]** In one embodiment, the oxidation process to form gate oxide **638** involves forming a thick oxide over substantially all of the MOS region **612**, using any known oxidation process in accordance with the methods described herein, forming a patterned photoresist mask using standard lithographic techniques and etching the oxide using a 10:1 buffered oxide etch (BOE), after which the photoresist mask is stripped or removed. The gate oxide **638** can have a thickness from about 1 nm to about 7 nm.

**[0073]** Referring to FIGs. 5 and 6H, a gate layer (not shown) is deposited and patterned to concurrently form gates **640** for the DEMOS transistor, and the NVM transistor (step **512**). Generally, the gate layer is a conductive, conformal layer deposited over substantially the entire surface **616** of the substrate **604** and all layers and structures formed thereon. A patterned photoresist mask (not shown) is then formed using standard lithographic techniques and the gate layer etched to remove the gate layer from areas not protected by the mask.

**[0074]** In one embodiment, the gate layer includes a doped polysilicon or poly layer deposited using chemical vapor deposition (CVD) to a thickness of from about 30 nm to about 100 nm, and etched using standard polysilicon etch chemistries, such as $CHF_3$ or $C_2H_2$ or $HBr/O_2$ which are highly selective to the underlying material of the gate oxide **638** and the dielectric stack **630**. The polysilicon can be doped using phosphorus implant for NMOS and boron implant for PMOS transistors. The implant doses are in the range of 1E15 to 1E16/$cm^2$ at energies of 2 to 50 keV.

**[0075]** Next, referring to FIGs. 5 and 6I in accordance with the present disclosure a drain extension **642** is formed in the MOS region **612** adjacent to where a drain for a DEMOS transistor will be formed by implanting or diffusing ions **644** of the appropriate dopant species to a suitable concentration and a suitable depth (step **514**). For example, the drain extension **642** can be formed by implanting $BF_2$ ions ($BF_2^+$) or boron ions **644** at an energy of from about 20 to about 50 keV, to a concentration of from about 0.5E13 to about 5E13/$cm^2$. In some embodiments, such as that shown, the drain extension implant can be angled relative to the surface **616** of the substrate **604** in such a way that the drain extension **642**, overlaps the gate stack **638/640** of the DEMOS. Suitable distances or lengths for the overlap include from about 30 to about 80 nm. In certain embodiments, such as that shown in FIG. 6I, the drain extension implant can split the second channel **624b** into a first portion **$624b_1$** encompassed by or within only the DNWELL **612** and a second portion **$624b_2$** encompassed by or within the drain extension **642** in the DNWELL.

**[0076]** Referring to FIGs. 5 and 6J, optionally in some embodiments another or second drain extension **646** can be formed in the NVM region **608** adjacent to where a

drain for a NVM transistor will be formed, and lightly doped diffusion (LDD) regions **648** can be formed in one or both of the MOS region **612** and the NVM region adjacent to where sources for the DEMOS transistor **650** and NVM transistor **652** will be formed (step **516**). As with the drain extension **642** for the DEMOS transistor the drain extension **646** for the NVM transistor can be formed by implanting or diffusing ions of the appropriate species and energy to the appropriate concentration.

**[0077]** For example, the drain extension **646** for the NVM transistor can be formed by implanting or diffusing phosphorus ions ($P^+$) at an energy of from about 4 to about 10 keV, to a concentration of from about 3E13 to about 3E14/cm$^2$.

**[0078]** Similarly, the LDD regions **648** can include a concentration of $BF_2$ ions ($BF_2^+$) of from about 0.5E14 to about 5E14/cm$^2$, as compared to the dopant concentration in highly doped source, and can extend from edge of the gate stacks or the beginning of the channels **622** and **624** for the NVM transistor and the DEMOS transistor to the sources distance or length of from about 20 to about 80 nm. Additionally, as with the drain extension **642** and **646** in some embodiments the LDD **648** can overlap the gate stacks as shown in FIG. 6K, while in other embodiments it is formed in such a way as to end at the edge of the gate stacks or the beginning of the channels.

**[0079]** Referring again to FIGs. 5 and 6K, a patterned mask (not shown) is formed and source and drain (S/D) implants are performed to form source and drain (S/D) regions **654** for both the DEMOS transistor **650** and NVM transistor **652** (step **518**). The patterned mask can include a photoresist mask or a hardmask patterned to expose only the S/D regions the DEMOS transistor **650** and NVM transistor **652**. S/D regions **654** can be formed by implanting or diffusing ions of the appropriate species and

energy to the appropriate concentration. For example, the S/D regions **654** for a P-type DEMOS transistor **650** and NVM transistor **652** can be formed by implanting or diffusing boron ions ($B^+$) or $BF_2$ ions at an energy of from about 2 to about 10 keV, to a concentration of from about 1E15 to about 1E16/$cm^2$.

**[0080]** After the S/D implants have been performed, the standard or baseline CMOS process flow is continued to substantially complete the front end fabrication for both the DEMOS transistor **650** and NVM transistor **652** (step **520**). The CMOS process flow can include spacer deposition, etch and silicidation to form contacts to the S/D regions **654**, and the formation of one or more interlayer dielectrics (ILD) as well as local interconnects between devices and transistors on a single level and vertical or interlayer contacts or vias.

**[0081]** In an alternative to the embodiment of FIG. 5, a step of forming the channel for either or both of the DEMOS transistor and the NVM transistor can be performed concurrently with a step of forming the drain extension for the DEMOS transistor. More particularly, a MOS channel region for the DEMOS transistor, including both first and the second channels, can be formed concurrently with an implant to form a MOS drain extension using a single implant, such as threshold voltage ($V_T$) implant, so that the MOS channel region and MOS drain extension include dopants of the same conductivity-type and concentration. An embodiment of this method will now be described with reference to FIG. 7. FIG. 7 is a flowchart illustrating an embodiment of a method for fabricating a circuit including a NVM transistor and a DEMOS transistor using a single implant to concurrently form a channel region and a drain extension for the DEMOS transistor.

**[0082]** Referring to FIG. 7, the process begins with forming a number of isolation structures in a wafer or substrate (step **702**). The isolation structures isolate a NVM transistor being formed in an NVM region of the substrate from the DEMOS transistor being formed in one or more adjoining MOS regions. A pad oxide is formed over a surface of the substrate in both the NVM region and the MOS region. The isolation structures are formed in substantially the same manner as that described above in connection with step **502**.

**[0083]** Next, dopants are then implanted into substrate through the pad oxide to form a lightly doped deep well (DWELL), a more heavily doped well (WELL) doped with the same type of dopant as the DWELL for the NVM transistor and the DEMOS transistor that will be formed in the NVM region and MOS regions (step **704**). The DWELL and WELL are formed in substantially the same manner as that described above in connection with step **504**.

**[0084]** In accordance with this embodiment a single implant is performed to concurrently form a channel region and a drain extension for the DEMOS transistor (step **706**). For example, the drain extension and channel region can be formed by implanting $BF_2$ ions ($BF_2^+$) or boron ions at an energy of from about 20 to about 50 keV, to a concentration of from about 0.5E13 to about 5E13/cm$^2$. In some embodiments, such as that shown, the drain extension implant can be angled relative to the surface of the substrate in such a way that the drain extension overlaps the gate stack of the DEMOS. Suitable distances or lengths for the overlap include from about 30 to about 80 nm. In certain embodiments, the drain extension implant can split the second channel into a first

portion encompassed by or within only the DNWELL and a second portion encompassed by or within the drain extension in the DNWELL.

**[0085]** Next, a hardmask is deposited and patterned to expose the surface of the substrate in the NVM region, the pad oxide in the NVM region removed and the hardmask removed (step **708**). The deposition, patterning and removal of the hardmask and the pad oxide is accomplished in substantially the same manner as that described above in connection with step **506**.

**[0086]** A number of dielectric or ONO layers are formed or deposited over the surface of the substrate, a mask formed on or overlying the dielectric layers, and the dielectric layers etched to form an ONO stacks of the NVM transistor to be formed in the NVM region (step **710**). The deposition and patterning removal of the dielectric layers is accomplished in the same manner as that described in connection with step **508** above. A gate oxide or GOX preclean is performed, and a gate oxide formed for the DEMOS transistor to be formed in the MOS region (step **712**). The GOX preclean and deposition and patterning of the gate oxide and is accomplished in substantially the same manner as that described above in connection with step **510**.

**[0087]** Next, a gate layer is deposited and patterned to concurrently form a gates for the NVM and DEMOS transistor (step **714**). The deposition and patterning of the gate layer is accomplished in substantially the same manner as that described above in connection with step **512**.

**[0088]** LDD regions are formed between the source and channel of the DEMOS transistor and NVM transistor, and, optionally, a second drain extension formed between the drain of the NVM transistor and channel (step **716**) as described above in connection

with step **516**. A patterned mask is formed and source and drain (S/D) implants are performed to form source and drain (S/D) regions for the DEMOS transistor and NVM transistor (step **718**). The formation of the S/D regions is accomplished in substantially the same manner as that described above in connection with step **518**. The standard or baseline CMOS process flow is then continued to substantially complete the front end fabrication the DEMOS transistor and NVM transistor (step **720**) in substantially the same manner as that described above in connection with step **520**.

**[0089]** In another embodiment, the drain extension of a DEMOS transistor according to the present disclosure can be formed concurrently in a single implant with an existing ESD implant step to form an electro-static discharge (ESD) diffusion region of an ESD device or circuit in a standard CMOS process flow.

**[0090]** Referring to FIG. 8, ESD protection device or circuit **800** protects circuits of an IC, such as the circuit or memory cell **600** described above, from voltage pulses or spikes generated or applied to I/O pads **802** of the circuit during normal operation due buildup of static electric charges. Referring to FIG. 8 an exemplary ESD protection circuit includes a first MOS transistor **804** or a first or P-type with a source coupled to a $V_{CC}$ pin of the IC and a drain coupled to the I/O pad **802** and to a second MOS transistor **806** of a second or N-type. The second MOS transistor **806** includes a drain coupled to the I/O pad **802** and a source coupled to a $V_{SS}$ pin of the IC. A deep P-type ESD implant **808** under the drain of second MOS transistor **806** reduces a drain junction breakdown voltage of the transistor. When an ESD pulse that is positive with respect to $V_{SS}$ is applied to the I/O pad **802** the drain junction breaks down earlier than other devices in the circuit triggering a parasitic bulk bipolar transistor **810** associated with the second MOS

transistor, providing a current path **812** through a bulk of the transistor to $V_{SS}$.

**[0091]** An embodiment of a method for fabricating DEMOS transistors having a multi-well and/or split channel architecture using an existing ESD implant will now be described in detail with reference to FIG. 9 and FIGs. 10A through 10D. FIG. 9 is a flowchart illustrating an embodiment of a method for fabricating a circuit including a NVM transistor, a MOS transistor in an ESD circuit and a DEMOS transistor using an existing ESD implant step in the CMOS process flow. FIGs. 10A through 10D are block diagrams illustrating cross-sectional views of a part of the circuit or memory cell **1000** formed by the method of the flowchart of FIG. 9.

**[0092]** Referring to FIG. 9 and FIG. 10A, the process begins with forming a number of isolation structures **1002** in a wafer or substrate **1004** (step **902**). The isolation structures **1002** isolate a NVM transistor being formed in an NVM region **1006** of the substrate from the DEMOS transistor and ESD transistor being formed in one or more adjoining MOS regions **1008** and **1010**. A pad oxide **1014** is formed over a surface **1016** of the substrate **1004** in both the NVM region **1006** and the MOS regions **1008** and **1010**.

**[0093]** Referring again to FIG. 9 and FIG. 10B, dopants are then implanted into substrate **1004** through the pad oxide **1014** to form a lightly doped deep well (DWELL **1018**), a more heavily doped well (WELL **1020**) doped with the same type of dopant as the DWELL, and channels **1022, 1024a, 1024b** and **1025** for the NVM transistor, the DEMOS transistor and the ESD transistor that will be formed in the NVM region **1006** and MOS regions **1008** and **1010** (step **904**).

**[0094]** Next, referring to FIG. 9 and FIG. 10B, a hardmask is deposited and patterned to expose the surface **1016** of the substrate **1004** in the NVM region **1006**, the

pad oxide **1014** in the NVM region removed and the hardmask removed (step **906**). The deposition, patterning and removal of the hardmask and the pad oxide **1016** is accomplished in the same manner as that described in connection with step **506** above.

**[0095]** A number of dielectric or ONO layers are formed or deposited over the surface **1016** of the substrate **1004**, a mask formed on or overlying the dielectric layers, and the dielectric layers etched to form an ONO stacks of the NVM transistor to be formed in the NVM region **1006** (step **908**). The deposition and patterning removal of the dielectric layers is accomplished in the same manner as that described in connection with step **508** above.

**[0096]** Next, referring to FIG. 9 and FIG. 10C, a gate oxide or GOX preclean is performed, and a gate oxides **1030**, **1038** and **1039** formed for the NVM transistor, the DEMOS transistor and the ESD transistor to be formed in the MOS regions **1008** and **1010** (step **910**). The GOX preclean and deposition and patterning of the gate oxides **1038** and **1039** is accomplished in the same manner as that described in connection with step **510** above.

**[0097]** A gate layer is deposited and patterned to concurrently form gates **1040** for the NVM transistor, the DEMOS transistor and the ESD transistor (step **912**). The deposition and patterning of the gate layer **1040** is accomplished in the same manner as that described in connection with step **512** above.

**[0098]** Referring again to FIGs. 9 and 10C, a patterned mask (not shown) is formed and source and drain (S/D) implants are performed to form source and drain (S/D) regions **1054** for the DEMOS transistor 10**50**, ESD transistor **1051** and NVM transistor **1052** (step **914**).

**[0099]** Referring again to FIG. 9 and FIG. 10D, in accordance with the present disclosure a drain extension **1042** is formed in the MOS region **1008** adjacent to where a drain for a DEMOS transistor will be formed concurrent with an ESD diffusion region **1043** of the ESD transistor by implanting or diffusing ions **1044** of the appropriate dopant species to a suitable concentration and a suitable depth (step **914**). For example, the drain extension **1042** and ESD diffusion region **1043** can be formed by implanting boron ions ($B^+$) ions **1044** at an energy of from about 40 to about 100 keV, to a concentration of from about 1E13 to about 1E14/cm$^2$. In some embodiments, such as that shown, the drain extension implant can be angled relative to the surface **1016** of the substrate **1004** in such a way that the drain extension **1042**, overlaps the gate stack **1038/1040** of the DEMOS. Suitable distances or lengths for the overlap include from about 20 to about 80 nm.

**[00100]** The standard or baseline CMOS process flow is then continued to substantially complete the front end fabrication the DEMOS transistor, ESD transistor and NVM transistor (step **918**). The S/D implants and deposition and completion of the baseline CMOS process flow is accomplished in the same manner as that described in connection with steps **518** and **520** above.

**[00101]** Although not shown or described in connection with this embodiment, it will be understood that the method can further include the formation of LDD regions between the source and channel of the DEMOS transistor **1050**, ESD transistor **1051** and NVM transistor **1052**, and a second drain extension between the drain of the NVM transistor and channel as described in connection with steps **514** and **516** above.

**[00102]** Thus, embodiments of drain-extended metal-oxide-semiconductor (DEMOS) transistors and methods for fabricating 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 in an illustrative rather than a restrictive sense.

**[00103]** 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.

**[00104]** 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 drain extended metal–oxide–semiconductor (DEMOS) device comprising a first channel, a gate, a second channel, and a drain extension, wherein the second channel is split into a first portion and a second portion, and wherein the first portion of the second channel stops under the gate and is spaced away from the drain extension.

2. The DEMOS device of Claim 1, wherein doping in the second portion of the second channel is higher than doping in the first portion of the second channel.

3. The DEMOS device of Claim 2, wherein a drain junction of the device is formed between the first channel and the second portion of the second channel.

4. The DEMOS device of Claim 3, wherein the doping in the second portion of the second channel is such that a breakdown voltage of the drain junction is above 9V.

5. The DEMOS device of Claim 2, wherein the drain extension is at least partially under the gate and within the second portion of the second channel.

6. The DEMOS device of Claim 1, wherein the device is a P-type MOS (DEPMOS) device, and the first channel is in a N-type well (NWELL), and the second

channel is in a deep NNWELL (DNWELL).

7. The DEMOS device of Claim 1, wherein the device is used to program or erase one or more cells in a flash memory.

8. The DEMOS device of Claim 1, wherein doping in the second channel is lower than doping in the first channel.

9. The DEMOS device of Claim 1, wherein the drain extension is formed concurrently with a MOS channel region comprising the first and second channels in a single threshold voltage ($V_T$) implant, and comprise dopants of the same conductivity-type and concentration.

10. A drain extended metal–oxide–semiconductor (DEMOS) device comprising:

a well of a first conductivity-type in a deep well of the first conductivity-type in a substrate of a second conductivity-type;

a channel underlying a gate stack on a substrate through which a source in the well and drain in the deep well are electrically coupled, the channel including at least a first channel in the well a second channel in the deep well; and

a drain extension through which the drain is electrically connected to the channel.

11. The DEMOS device of claim 10, wherein the drain extension extends into

the second channel of the channel underlying the gate stack.

12. The DEMOS device of claim 11, wherein the second channel comprises a first portion between the first channel and the drain extension, and a second portion in the drain extension underlying the gate stack.

13. The DEMOS device of claim 11, wherein the drain extension extends from the drain to the gate stack, without extending into the second channel of the channel underlying the gate stack.

14. The DEMOS device of claim 10, wherein an interface between the well and the deep well at a surface of the substrate coincides with an edge of the gate stack, and wherein the channel underlying the gate stack comprises only the first channel in the well.

15. The DEMOS device of claim 14, wherein the drain extension extends into the first channel of the channel underlying the gate stack.

16. The DEMOS device of claim 14, wherein the drain extension extends from the drain to the edge of the gate stack, without extending into the first channel of the channel underlying the gate stack.

17. The DEMOS device of claim 14, wherein the first conductivity-type is a

N-type conductivity to form a N-well and deep N-well in a P-type substrate to form a DEPMOS device.

18. The DEMOS device of claim 14, wherein the drain extension has a doping concentration lower than the drain to achieve depletion during reverse biasing of the drain junction of the device, thereby raising a breakdown voltage thereof.

19. A circuit formed in a semiconductor substrate, the circuit comprising:

at least one drain extended metal–oxide–semiconductor (DEMOS) device including:

a well of a first dopant-type in a deep well of the first dopant-type in a substrate of a second dopant-type;

a channel underlying a gate stack on the substrate through which a source in the well and drain in the deep well are electrically coupled, the channel including at least a first channel in the well; and

a drain extension between the channel and the drain through which the drain is electrically connected to the channel; and

an ESD protection circuit comprising an ESD diffusion region formed in the substrate adjacent a source or drain diffusion region of a MOS transistor in the ESD protection circuit,

wherein the drain extension and the ESD diffusion region are formed concurrently in a single ESD implant step and comprise dopants of a same conductivity-type and concentration.

20. The circuit of claim 19, wherein the DEMOS device further comprises a second channel of the channel in the deep well underlying the gate stack.

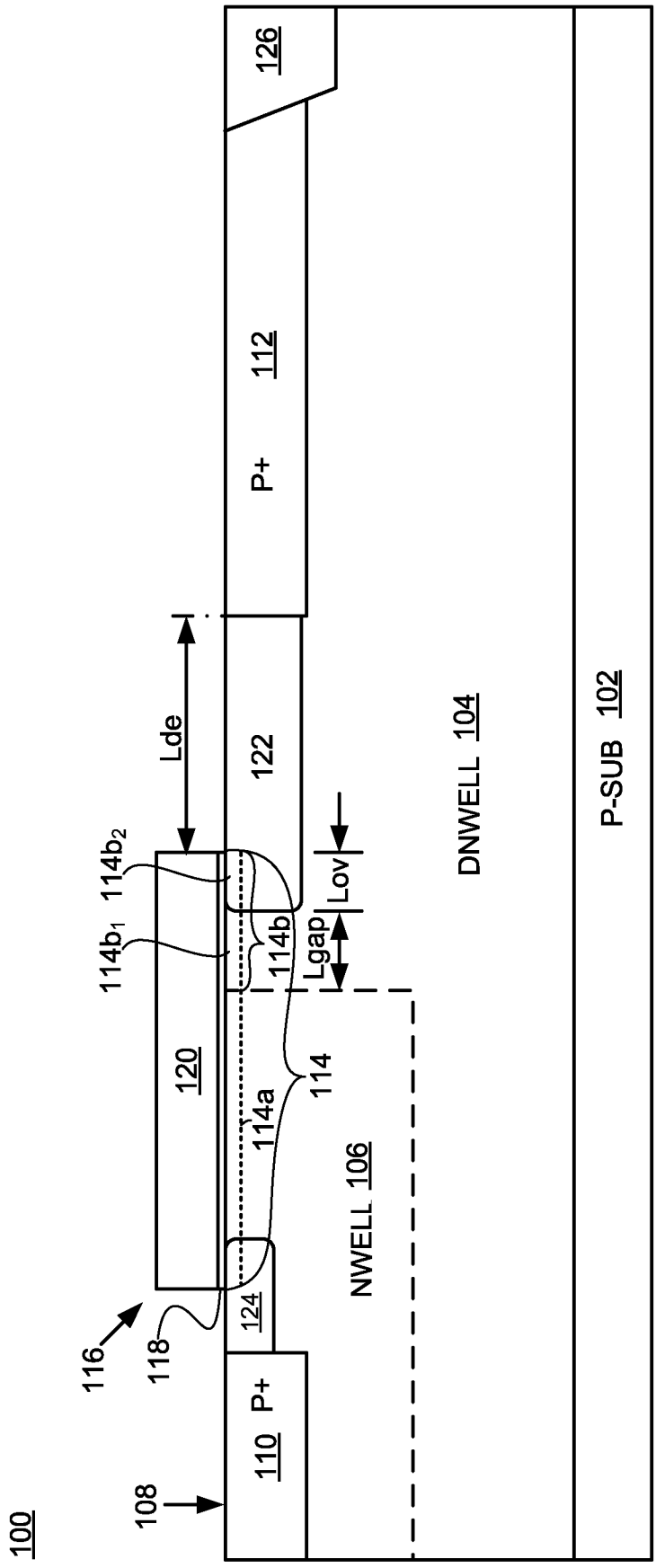

100
108
110
P+
116
118
124
120
114a
114
114b
Lgap
Lov
114b1
114b2
Lde
122
112
P+
126
NWELL 106
DNWELL 104
P-SUB 102

FIG. 1A

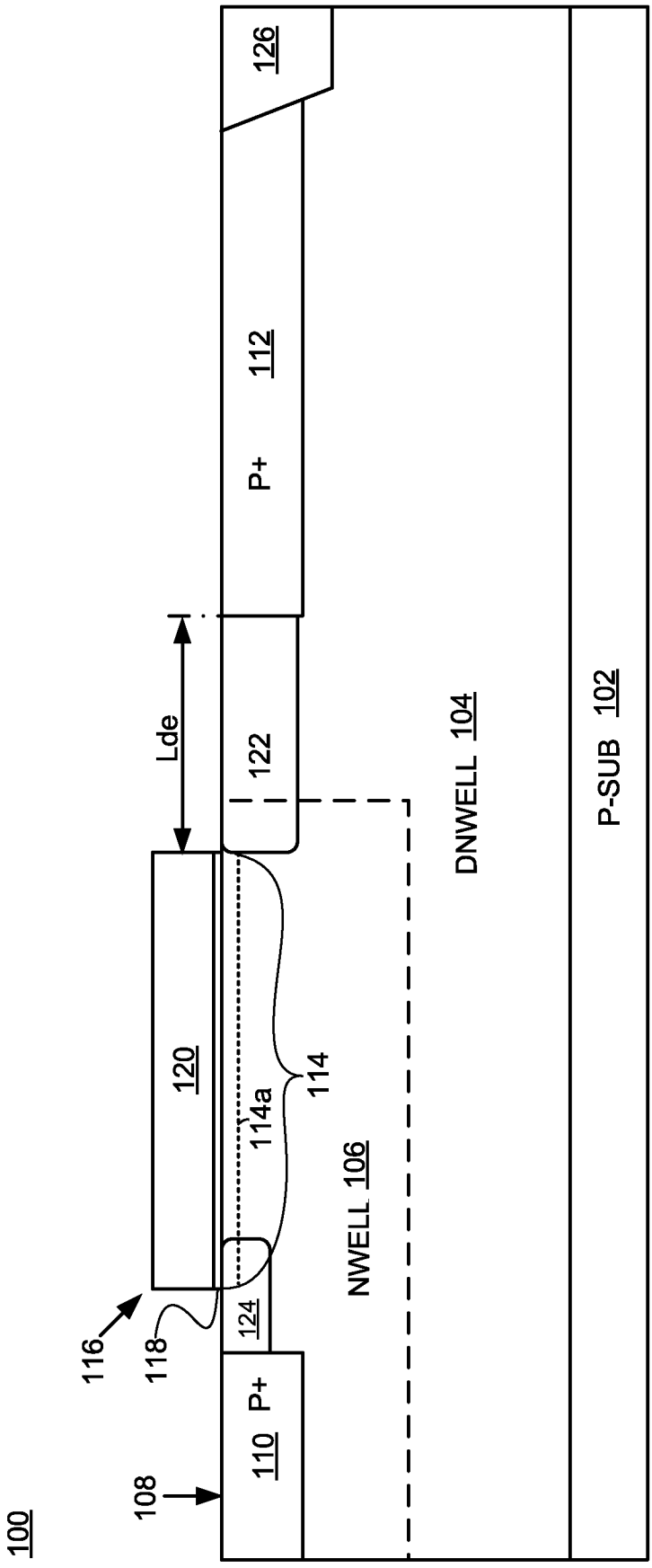

100
108
110 P+
116
118
124
120
114a
114
122
Lde
112
P+
126
NWELL 106
DNWELL 104
P-SUB 102

FIG. 1B

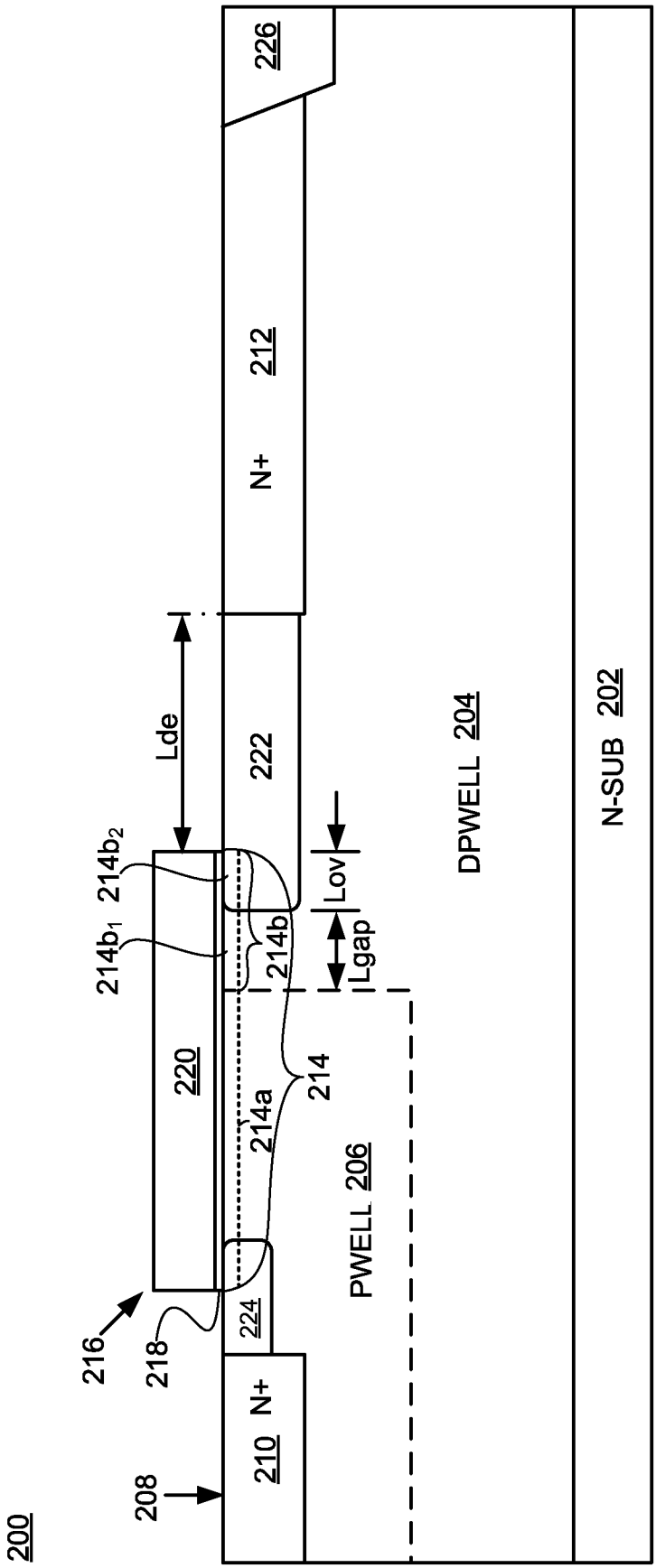

200
208
210 N+
216
218
224
220
214
214a
214b
214b1
214b2
Lde
222
Lov
Lgap
212 N+
226
PWELL 206
DPWELL 204
N-SUB 202

FIG. 2

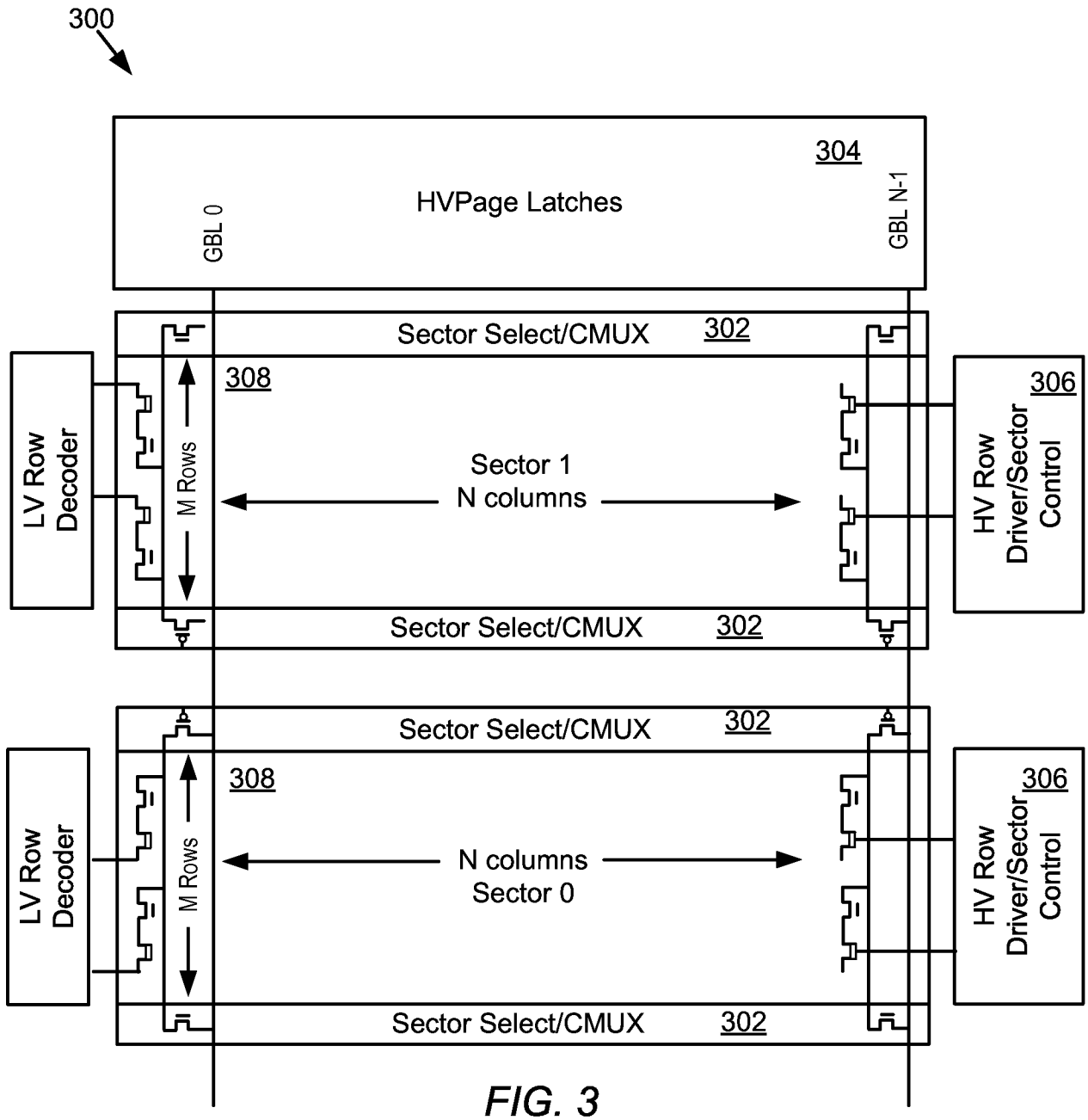

300
304
HVPage Latches
GBL 0
GBL N-1
Sector Select/CMUX
302
308
Sector 1
N columns
M Rows
LV Row Decoder
HV Row Driver/Sector Control
306
Sector Select/CMUX
302
Sector Select/CMUX
302
308
N columns
Sector 0
M Rows
LV Row Decoder
HV Row Driver/Sector Control
306
Sector Select/CMUX
302

FIG. 3

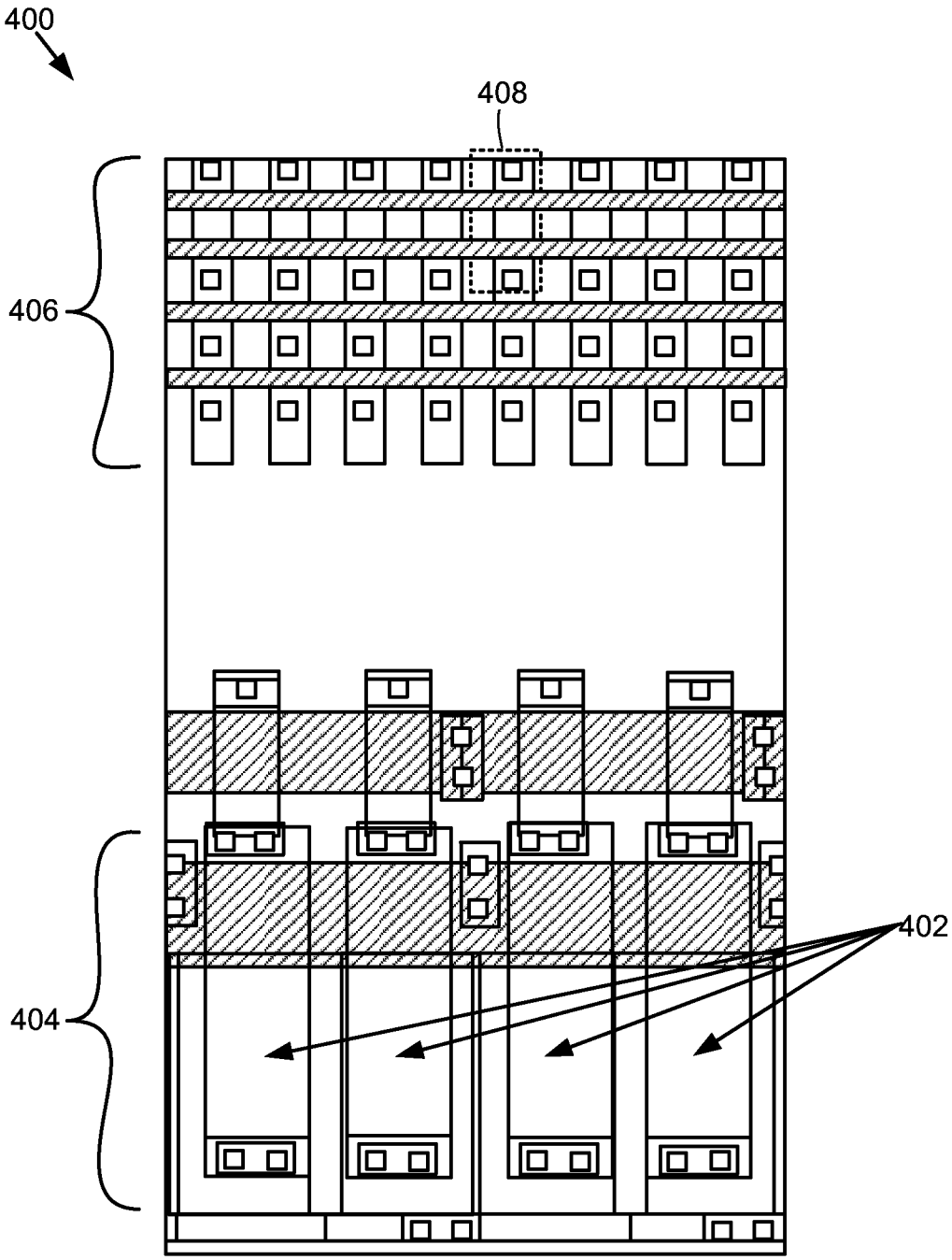



FIG. 4

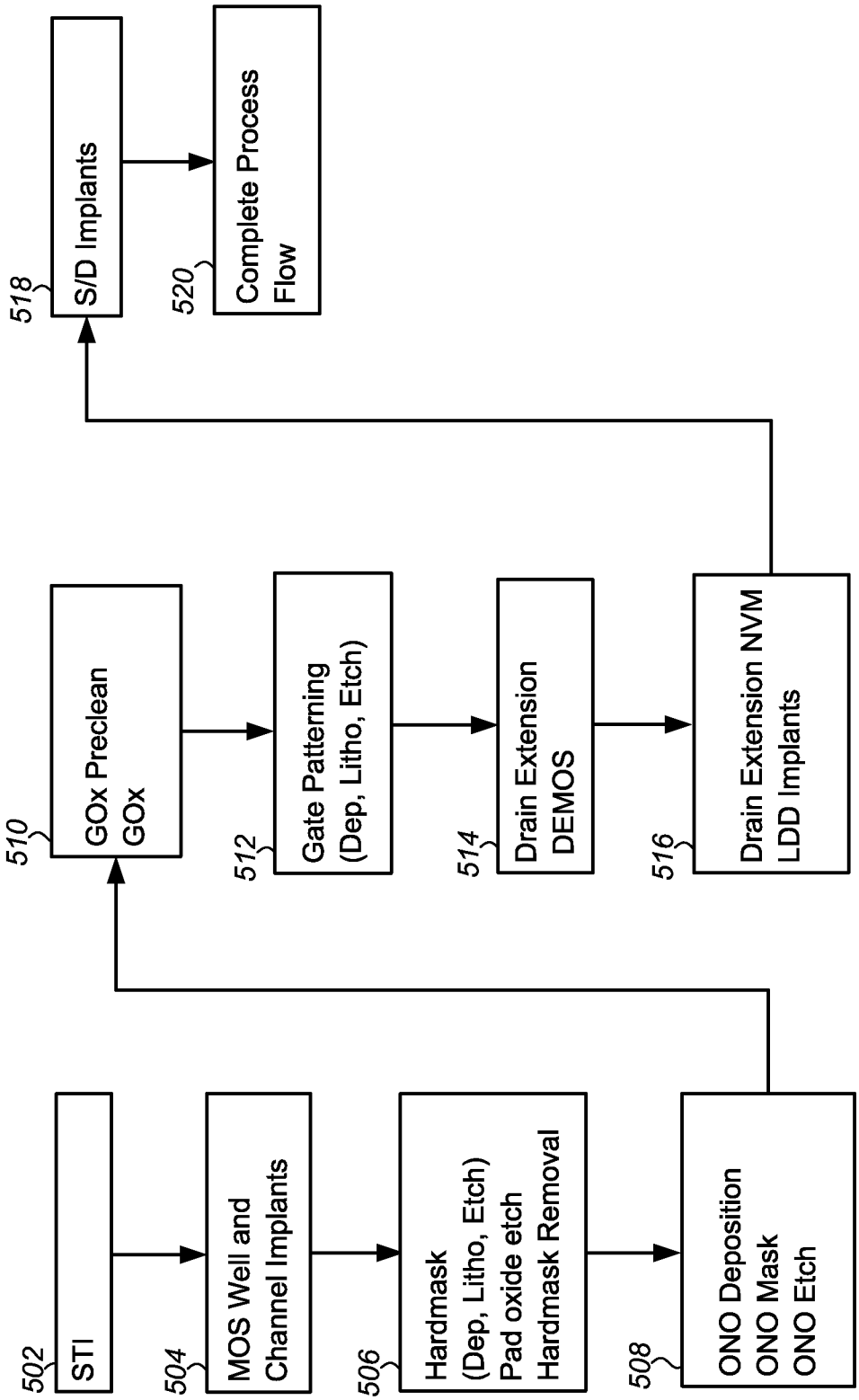

502
STI
504
MOS Well and Channel Implants
506
Hardmask (Dep, Litho, Etch) Pad oxide etch Hardmask Removal
508
ONO Deposition ONO Mask ONO Etch
510
GOx Preclean GOx
512
Gate Patterning (Dep, Litho, Etch)
514
Drain Extension DEMOS
516
Drain Extension NVM LDD Implants
518
S/D Implants
520
Complete Process Flow

FIG. 5

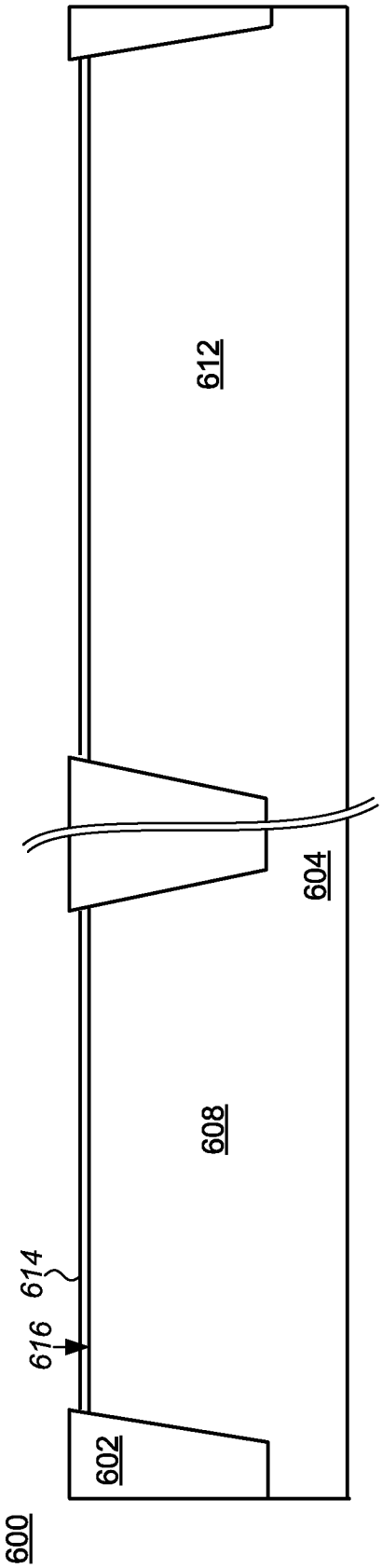

600
602
604
608
612
614
616

FIG. 6A

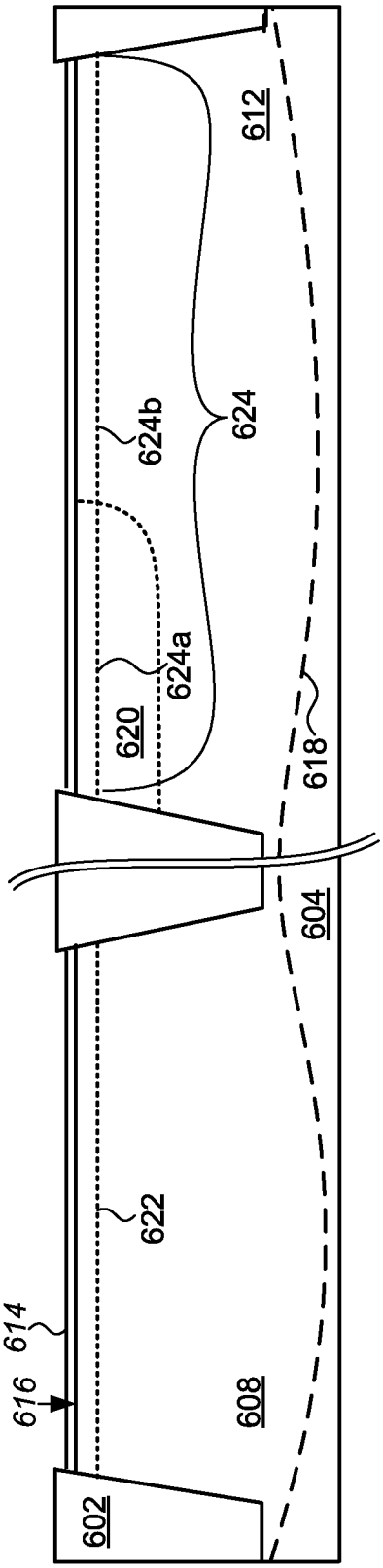

602
604
608
612
614
616
618
620
622
624
624a
624b

FIG. 6B

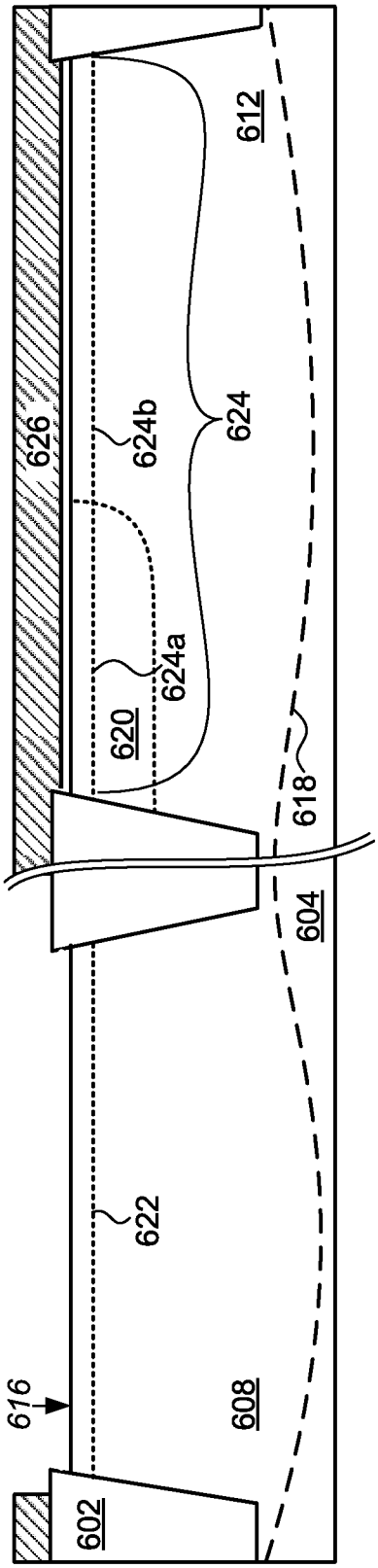

612
626
624b
624
624a
620
618
604
622
616
608
602

FIG. 6C

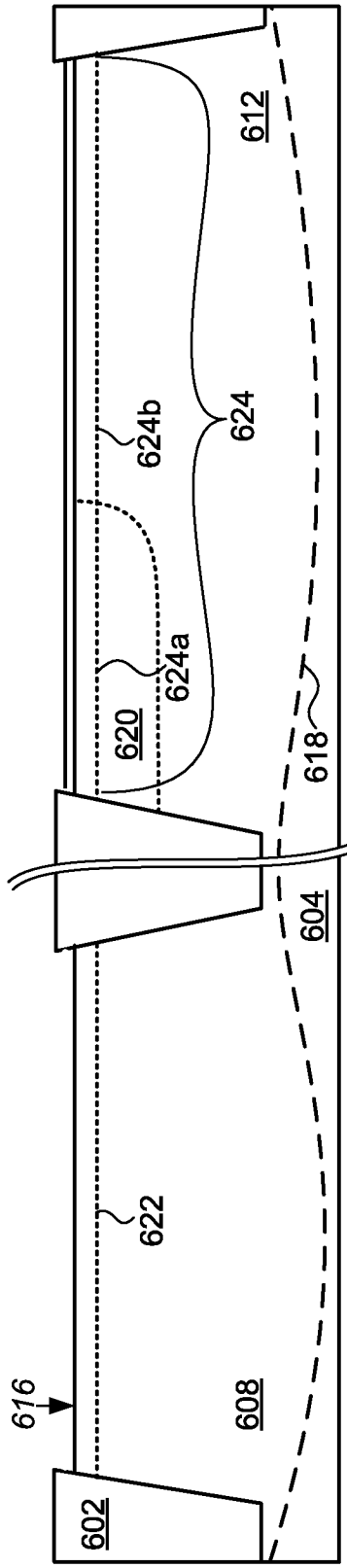

612
624b
624
624a
620
618
604
622
616
608
602

FIG. 6D

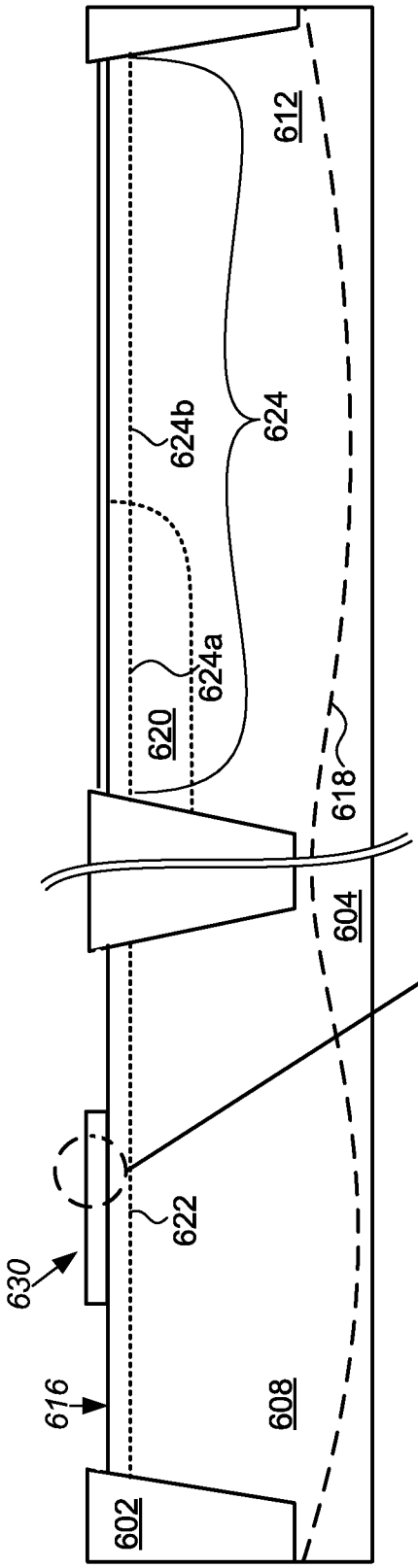

612
624b
624
624a
620
618
604
622
630
616
608
602

FIG. 6E

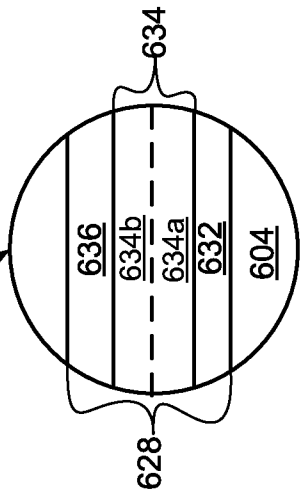

634
636
634b
634a
632
604
628

FIG. 6F

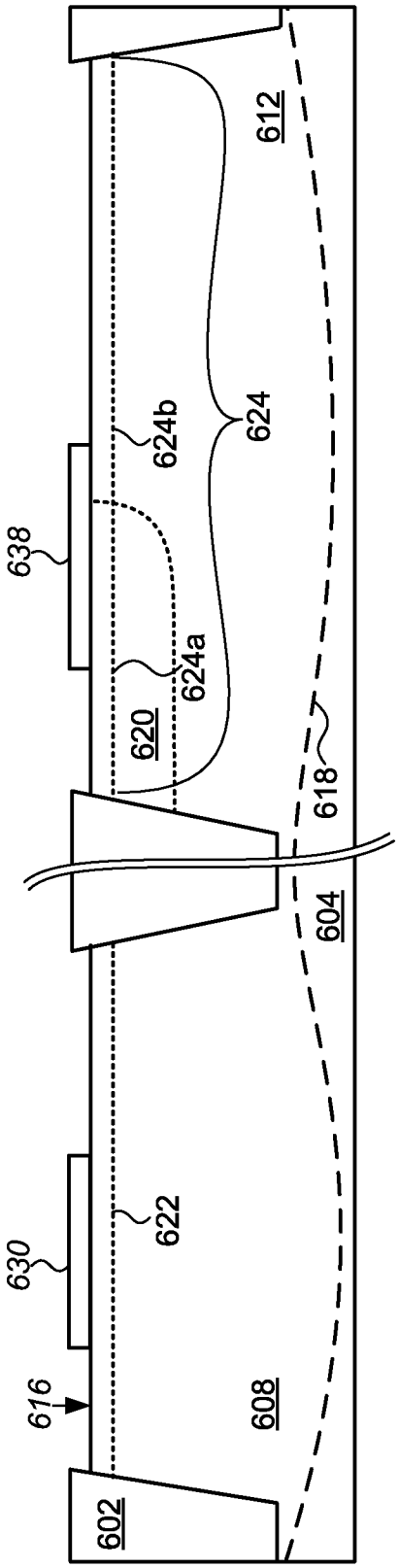

612
624
624b
638
624a
620
618
604
630
622
616
608
602

FIG. 6G

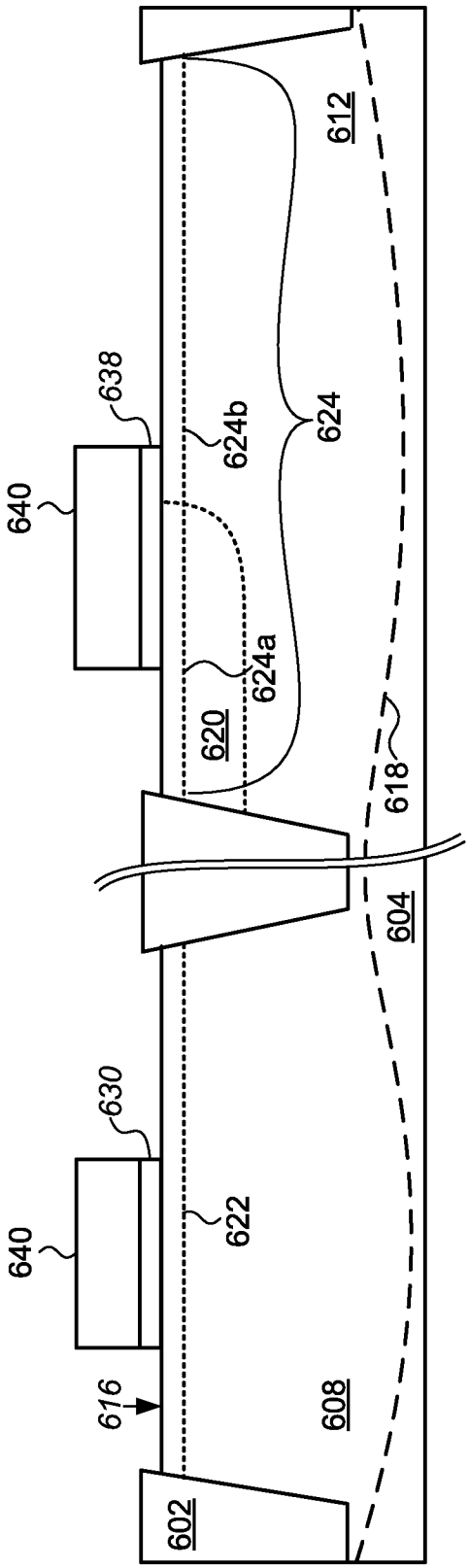

612
638
624b
624
640
624a
620
618
604
630
640
622
616
608
602

FIG. 6H

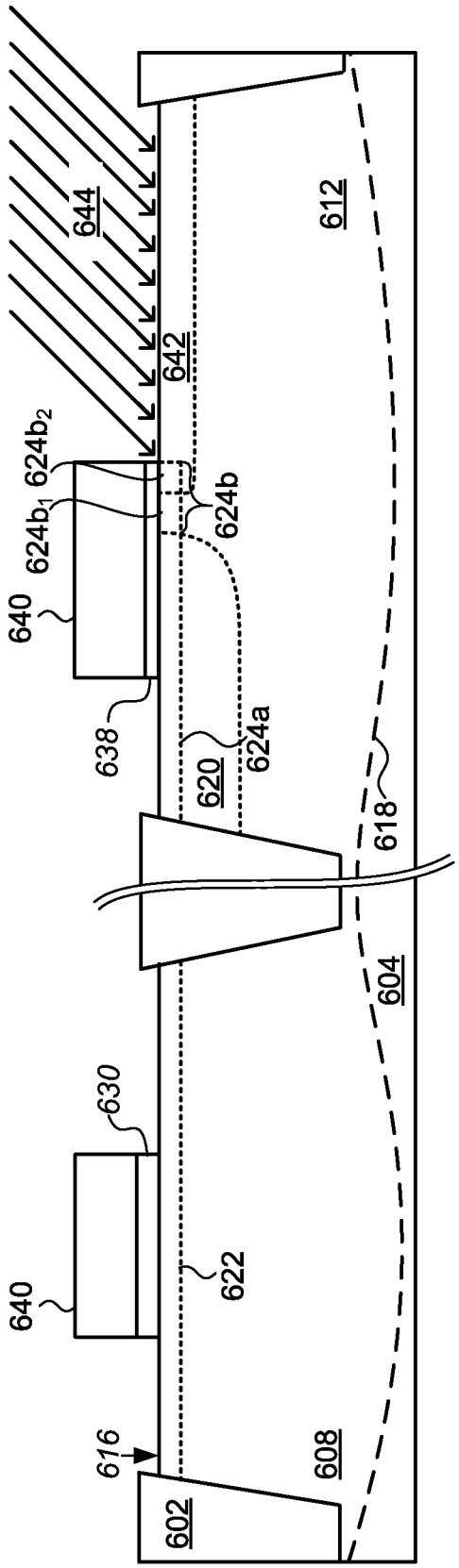

644
642
612
624b2
624b1
624b
640
638
624a
620
618
604
630
640
622
616
608
602

FIG. 6I

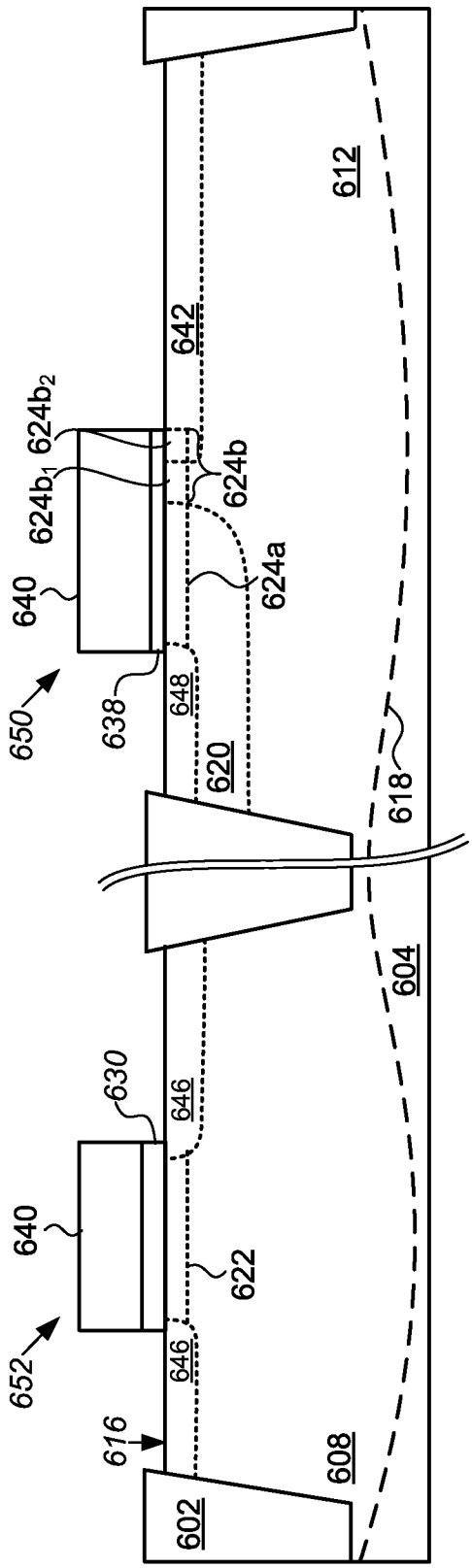

642
612
624b2
624b1
624b
640
624a
650
638
648
620
618
604
630
646
640
622
652
646
616
608
602

FIG. 6J

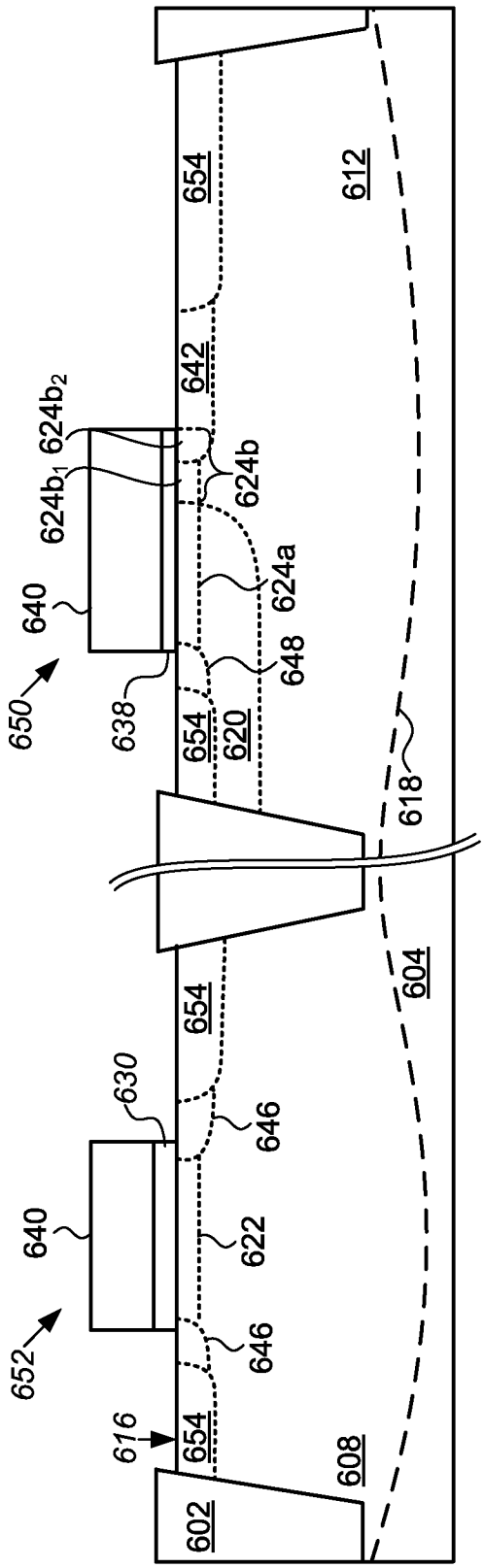

654
612
642
624b2
624b1
624b
624a
640
650
638
648
654
620
618
604
654
630
646
640
622
652
646
616
654
608
602

FIG. 6K

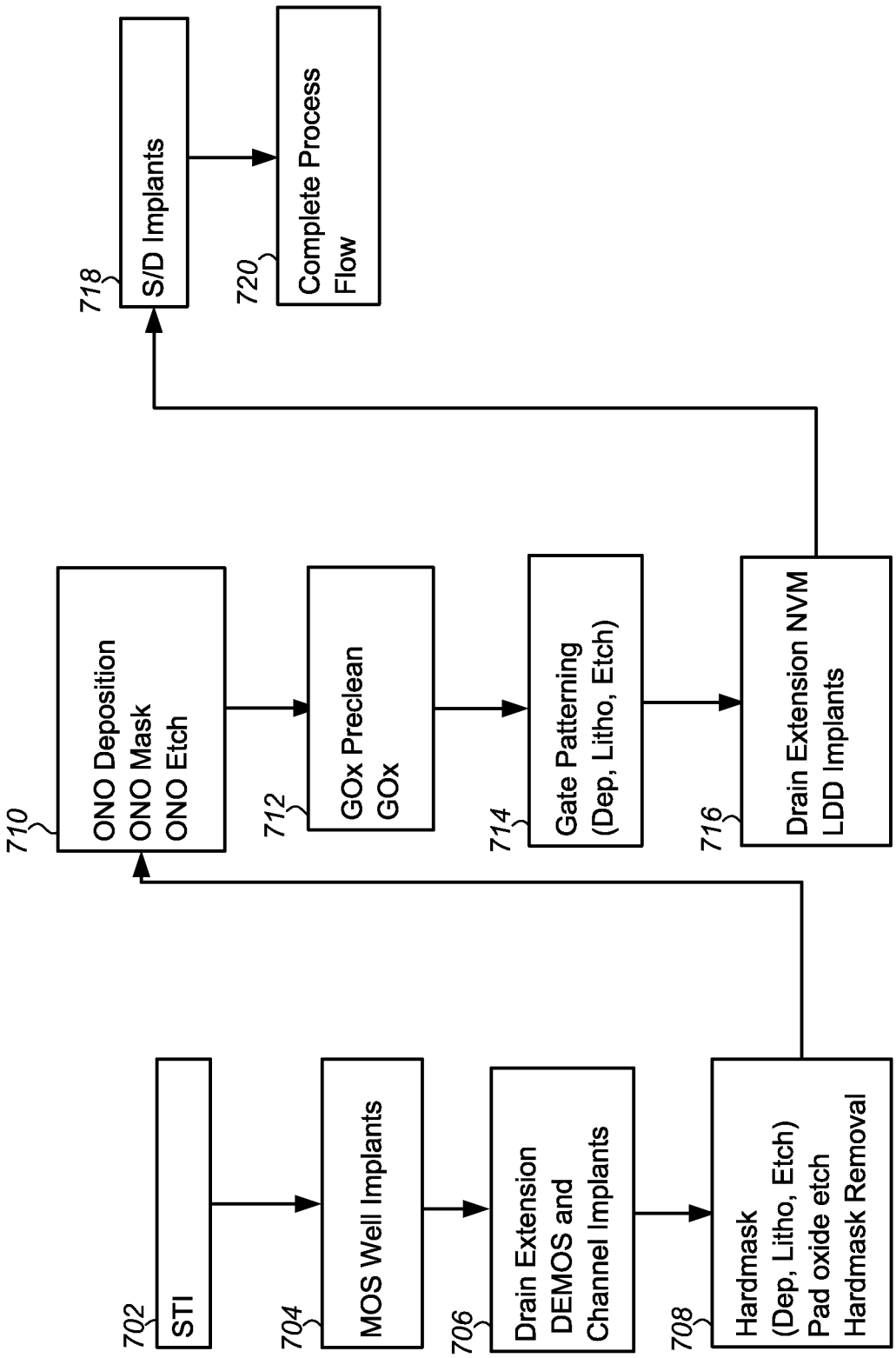

702
STI
704
MOS Well Implants
706
Drain Extension
DEMOS and
Channel Implants
708
Hardmask
(Dep, Litho, Etch)
Pad oxide etch
Hardmask Removal
710
ONO Deposition
ONO Mask
ONO Etch
712
GOx Preclean
GOx
714
Gate Patterning
(Dep, Litho, Etch)
716
Drain Extension NVM
LDD Implants
718
S/D Implants
720
Complete Process
Flow

FIG. 7

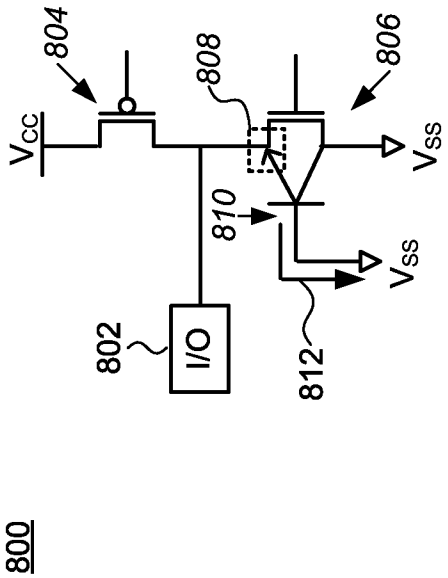



FIG. 8

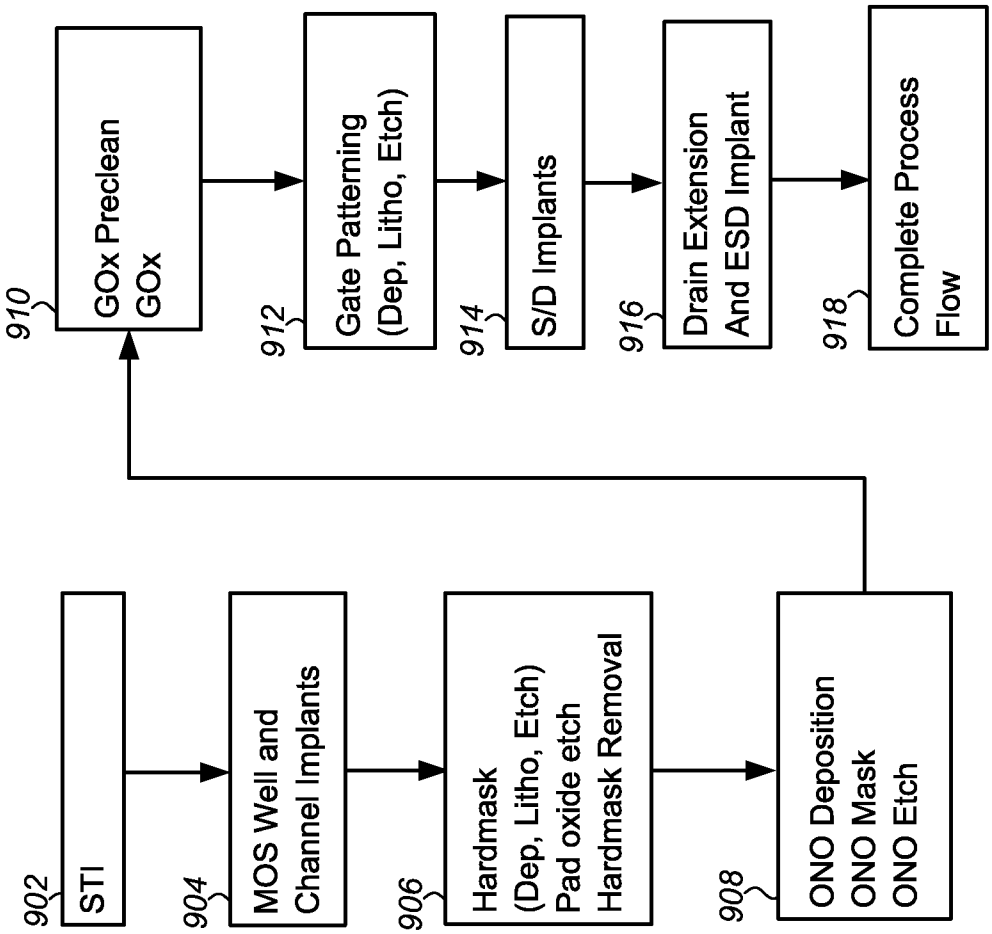

902
STI
904
MOS Well and Channel Implants
906
Hardmask (Dep, Litho, Etch) Pad oxide etch Hardmask Removal
908
ONO Deposition ONO Mask ONO Etch
910
GOx Preclean GOx
912
Gate Patterning (Dep, Litho, Etch)
914
S/D Implants
916
Drain Extension And ESD Implant
918
Complete Process Flow

FIG. 9

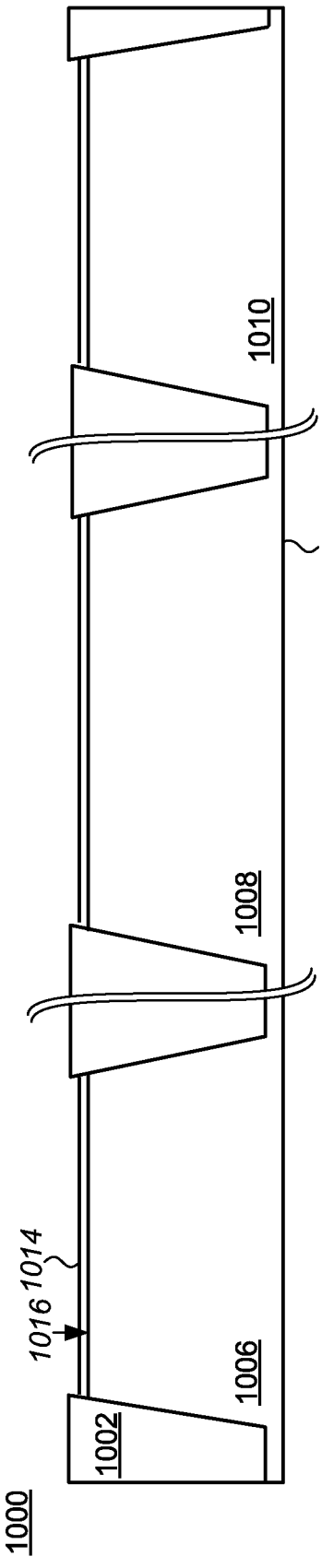



FIG. 10A

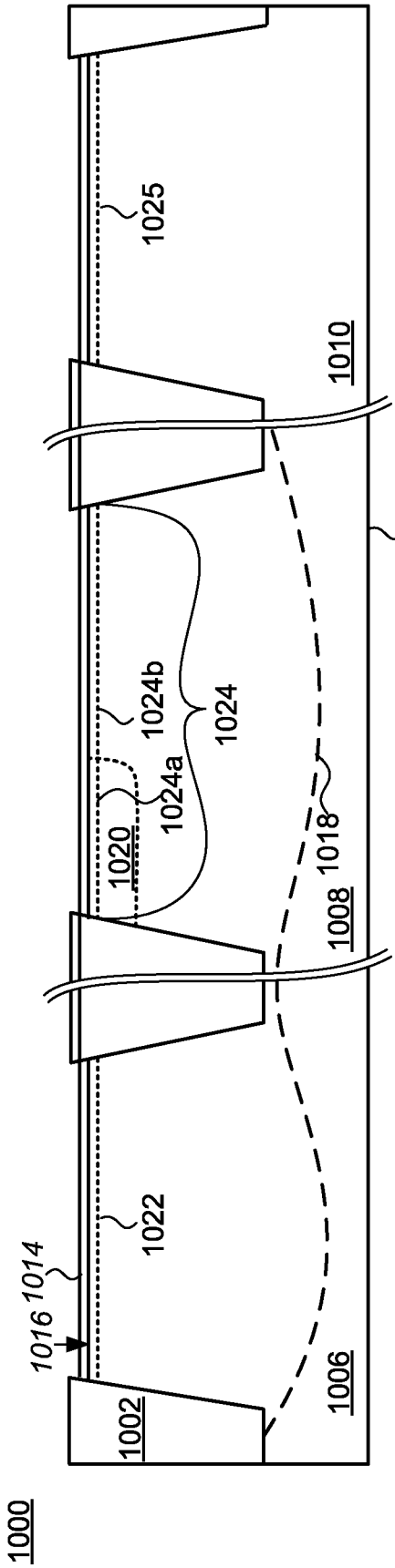



FIG. 10B

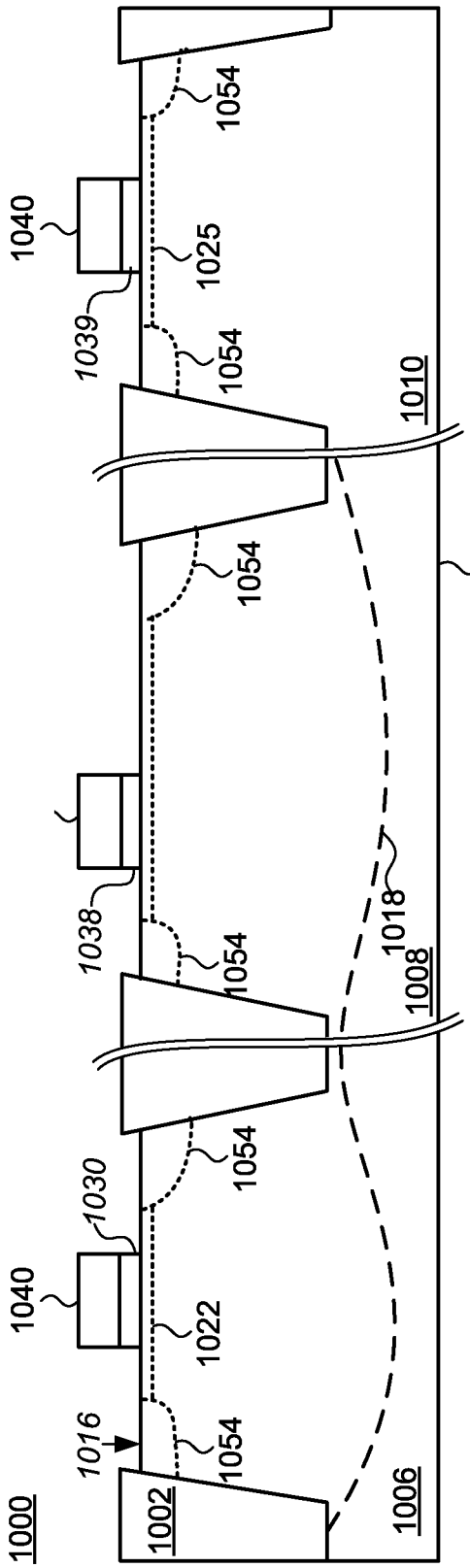



FIG. 10C

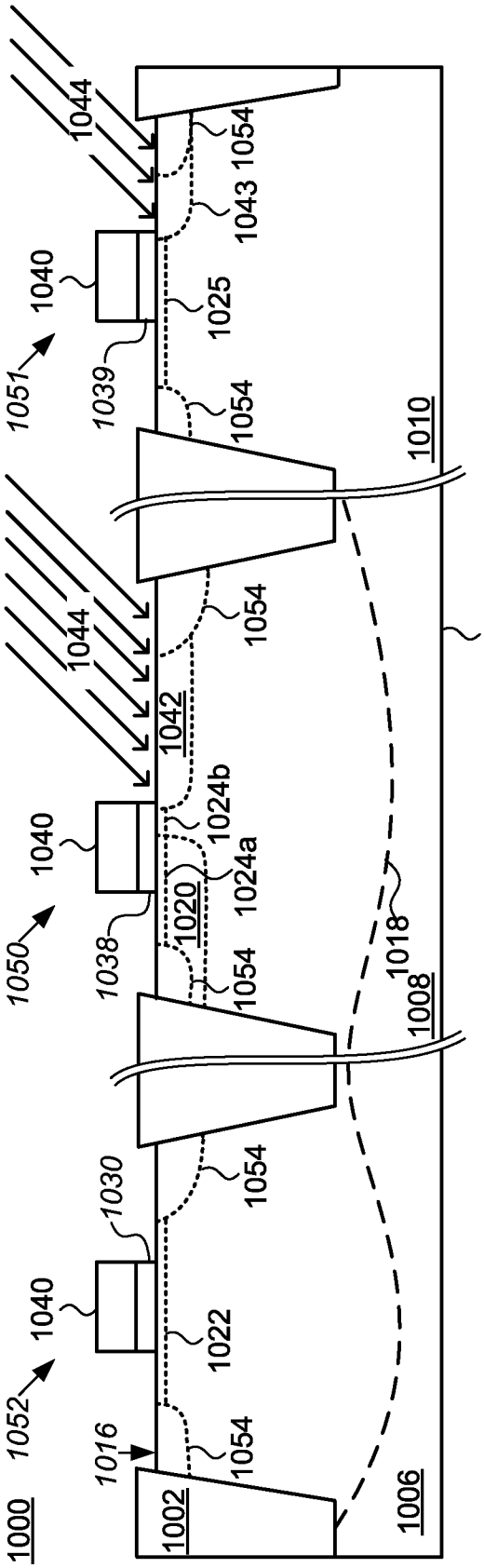



FIG. 10D

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US15/10516

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - H01L 23/62, 29/10, 29/423 (2015.01)
CPC - H01L 23/62, 29/10, 29/423
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(8): H01L 23/58, 23/60, 23/62, 29/06, 29/08, 29/10, 29/423 (2015.01); CPC: H01L 23/58, 23/60, 23/62, 29/06, 29/08, 29/10, 29/423; USPC: 257/288, 355, 356, 357, 358, 360, 363

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)
PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); Google; Google Scholar; ProQuest; KEYWORDS: drain extended metal oxide semiconductor DEMOS first second channel drain extension gate N-well deep well p-type n-type breakdown voltage

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X<br>---<br>Y | US 7,838,937 B1 (WALKER, A et al.) November 23, 2010; figures 1, 4; column 2, lines 43-56; column 5, line 50 to column 6, line 8; column 7, lines 7-25; column 11, line 62 to column 12, line 28 | 1, 6, 8-20<br>---<br>2-5, 7 |
| Y | US 8,111,558 B2 (PESAVENTO, A et al.) February 7, 2012; figure 1C; column 4, lines 34-56; column 8, lines 52-63 | 2-5, 7 |
| A | US 8,283,727 B1 (WALKER, A et al.) October 9, 2012; entire document | 1-20 |

☐ 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 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 | Date of mailing of the international search report |
|---|---|
| 06 March 2015 (06.03.2015) | 13 APR 2015 |
| Name and mailing address of the ISA/<br>Mail Stop PCT, Attn: ISA/US, Commissioner for Patents<br>P.O. Box 1450, Alexandria, Virginia 22313-1450<br>Facsimile No. 571-273-3201 | Authorized officer<br>Shane Thomas<br>PCT Helpdesk: 571-272-4300<br>PCT OSP: 571-272-7774 |