

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
4 December 2008 (04.12.2008)

PCT

(10) International Publication Number
WO 2008/147385 A1

(51) International Patent Classification:
H01L 29/792 (2006.01)

(21) International Application Number:
PCT/US2007/020965

(22) International Filing Date:
28 September 2007 (28.09.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/931,905 25 May 2007 (25.05.2007) US
11/904,475 26 September 2007 (26.09.2007) US

(71) Applicant (for all designated States except US): **CY-PRESS SEMICONDUCTRO CORPORATION**
[US/US]; 198 Champion Court, San Jose, CA 95134-1709 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **LEVY, Sagy** [IL/US]; 1139 Parkington Avenue, Sunnvale, CA 94087 (US). **JENNE, Fredrick, B.** [US/US]; 790 Lynxwood Court, Sunnyvale, CA 94086 (US). **RAMKUMAR, Krishnaswamy** [US/US]; 1193 Lynbrook Way, San Jose, CA 95129 (US).

(74) Agents: **BERNADICOU, Michael** et al.; Blakely, Sokoloff, Taylor & Zafman LLP, 1279 Oakmead Parkway, Sunnyvale, CA 94085-4040 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, 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, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, 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, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: NONVOLATILE CHARGE TRAP MEMORY DEVICE HAVING A DEUTERATED LAYER IN A MULTI-LAYER CHARGE-TRAPPING REGION

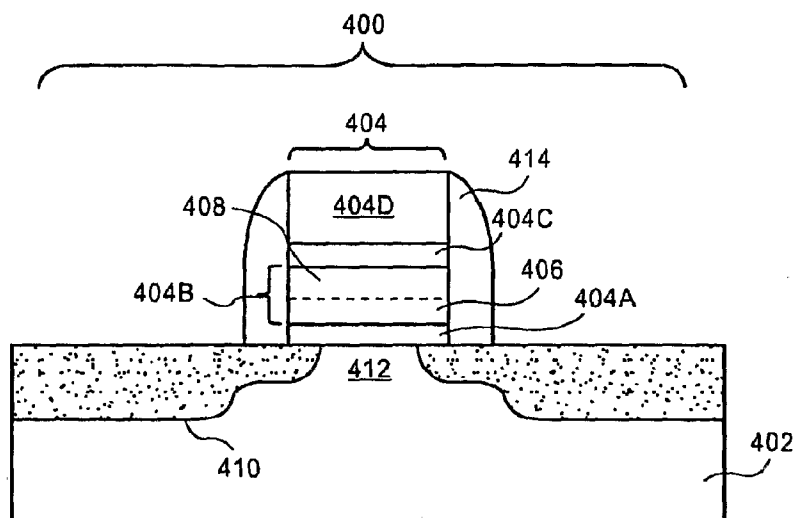


FIG. 4

(57) Abstract: A nonvolatile charge trap memory device is described. The device includes a substrate having a channel region and a pair of source/drain regions. A gate stack is above the substrate over the channel region and between the pair of source/drain regions. The gate stack includes a multi-layer charge-trapping region having a first deuterated layer. The multi-layer charge-trapping region may further include a deuterium-free charge-trapping layer.

WO 2008/147385 A1



Published:

— *with international search report*

Nonvolatile Charge Trap Memory Device Having a Deuterated Layer in a Multi-Layer Charge-Trapping Region

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 60/931,905, filed May 25, 2007, which is hereby incorporated by reference.

TECHNICAL FIELD

[0002] The invention is in the field of Semiconductor Devices.

BACKGROUND

[0003] For the past several decades, the scaling of features in integrated circuits has been a driving force behind an ever-growing semiconductor industry. Scaling to smaller and smaller features enables increased densities of functional units on the limited real estate of semiconductor chips. For example, shrinking transistor size allows for the incorporation of an increased number of memory devices on a chip, lending to the fabrication of products with increased capacity. The drive for ever-more capacity, however, is not without issue. The necessity to optimize the performance of each device becomes increasingly significant.

[0004] Non-volatile semiconductor memories typically use stacked floating gate type field-effect-transistors. In such transistors, electrons are injected into a floating gate of a memory cell to be programmed by biasing a control gate and grounding a body region of a substrate on which the memory cell is formed. An oxide-nitride-oxide (ONO) stack is used as either a charge storing layer, as in a semiconductor-oxide-nitride-oxide-semiconductor (SONOS) transistor, or as an isolation layer between the floating gate and control gate, as in a split gate flash transistor. Figure 1 illustrates a cross-sectional view of a conventional nonvolatile charge trap memory device.

[0005] Referring to Figure 1, semiconductor device 100 includes a SONOS gate stack 104 including a conventional ONO portion 106 formed over a silicon substrate 102. Semiconductor device 100 further includes source and drain regions 110 on either side of SONOS gate stack 104 to define a channel region 112. SONOS gate stack 104 includes a poly-silicon gate layer 108 formed above and in contact with ONO portion 106. Poly-silicon gate layer 108 is electrically isolated from silicon

substrate 102 by ONO portion 106. ONO portion 106 typically includes a tunnel oxide layer 106A, a nitride or oxy-nitride charge-trapping layer 106B, and a top oxide layer 106C overlying nitride or oxy-nitride layer 106B.

[0006] One problem with conventional SONOS transistors is the poor data retention in the nitride or oxy-nitride layer 106B that limits semiconductor device 100 lifetime and its use in several applications due to leakage current through the layer. One attempt to address this problem focused on the use of silicon-rich SONOS layers, which enable a large initial separation between program and erase voltages at the beginning of life but result a rapid deterioration of charge storing ability. Another attempt focused on oxygen-rich layers, which enable a reduced rate of deterioration of charge storing ability, but also reduce the initial separation between program and erase voltages. The effect of both of these approaches on data retention over time may be shown graphically. Figures 2 and 3 are plots of Threshold Voltage (V) as a function of Retention Time (Sec) for conventional nonvolatile charge trap memory devices.

[0007] Referring to Figure 2, rapid deterioration of charge storing ability for a silicon-rich layer is indicated by the convergence of the programming threshold voltage (VTP) 202 and erase threshold voltage (VTE) 204 to a specified minimum 206. Referring to Figure 3, a reduced separation between VTP 302 and VTE 304 is obtained for an oxygen-rich layer. As indicated by line 306, the overall useful lifetime of device is not appreciably extended by this approach.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Figure 1 illustrates a cross-sectional view of a conventional nonvolatile charge trap memory device.

[0009] Figure 2 is a plot of Threshold Voltage (V) as a function of Retention Time (Sec) for a conventional nonvolatile charge trap memory device.

[0010] Figure 3 is a plot of Threshold Voltage (V) as a function of Retention Time (Sec) for a conventional nonvolatile charge trap memory device.

[0011] Figure 4 illustrates a cross-sectional view of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0012] Figure 5 illustrates a cross-sectional view of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0013] Figure 6A illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0014] Figure 6B illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0015] Figure 6C illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0016] Figure 6D illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0017] Figure 6E illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0018] Figure 6F illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0019] Figure 6G illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0020] Figure 6H illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0021] Figure 6I illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0022] Figure 7A illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0023] Figure 7B illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0024] Figure 7C illustrates a cross-sectional view representing a step in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION

[0025] A nonvolatile charge trap memory device and a method to form the same is described herein. In the following description, numerous specific details are set forth, such as specific dimensions, in order to provide a thorough understanding of the present invention. It will be apparent to one skilled in the art that the present invention may be practiced without these specific details. In other instances, well-known processing steps, such as patterning steps or wet chemical cleans, are not described in detail in order to not unnecessarily obscure the present invention. Furthermore, it is to be understood that the various embodiments shown in the Figures are illustrative representations and are not necessarily drawn to scale.

[0026] Disclosed herein is a nonvolatile charge trap memory device. The device may include a substrate having a channel region and a pair of source and drain regions. A gate stack may be formed above the substrate over the channel region and between the pair of source and drain regions. In one embodiment, the gate stack includes a multi-layer charge-trapping region having a first deuterated layer. The multi-layer charge-trapping region may further include a deuterium-free charge-trapping layer. Alternatively, the multi-layer charge-trapping region may include a partially deuterated charge-trapping layer having a deuterium concentration less than that of the first deuterated layer.

[0027] A nonvolatile charge trap memory device including a multi-layer charge-trapping region having a deuterated layer may exhibit improved programming and erase speed and data retention. In accordance with an embodiment of the present invention, a deuterated layer is formed between the charge-trapping layer of the multi-layer charge-trapping region and the tunnel dielectric layer. In one embodiment, the deuterated layer is essentially trap-free and mitigates hot electron degradation during erase and program cycles. By incorporating a trap-free layer between the tunnel dielectric layer and the charge-trapping layer of a multi-layer charge-trapping region, the V_t shift from erase and program cycles may be reduced and the retention may be increased. In accordance with another embodiment of the present invention, a second

deuterated layer is also formed between the charge-trapping layer of the multi-layer charge-trapping region and a top dielectric layer of the gate stack.

[0028] A nonvolatile charge trap memory device may include a multi-layer charge-trapping region having a deuterated layer. Figure 4 illustrates a cross-sectional view of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0029] Referring to Figure 4, semiconductor device 400 includes a gate stack 404 formed over a substrate 402. Semiconductor device 400 further includes source and drain regions 410 in substrate 402 on either side of gate stack 404, defining a channel region 412 in substrate 402 underneath gate stack 404. Gate stack 404 includes a tunnel dielectric layer 404A, a multi-layer charge-trapping region 404B, a top dielectric layer 404C and a gate layer 404D. Thus, gate layer 404D is electrically isolated from substrate 402. Multi-layer charge-trapping region 404B includes a deuterated layer 406 between a charge-trapping layer 408 of multi-layer charge-trapping region 404B and tunnel dielectric layer 404A. A pair of dielectric spacers 414 isolates the sidewalls of gate stack 404.

[0030] Semiconductor device 400 may be any nonvolatile charge trap memory device. In one embodiment, semiconductor device 400 is a Flash-type device wherein the charge-trapping layer is a conductor layer or a semiconductor layer. In accordance with another embodiment of the present invention, semiconductor device 400 is a SONOS-type device wherein the charge-trapping layer is an insulator layer. By convention, SONOS stands for "Semiconductor-Oxide-Nitride-Oxide-Semiconductor," where the first "Semiconductor" refers to the channel region material, the first "Oxide" refers to the tunnel dielectric layer, "Nitride" refers to the charge-trapping dielectric layer, the second "Oxide" refers to the top dielectric layer (also known as a blocking dielectric layer) and the second "Semiconductor" refers to the gate layer. A SONOS-type device, however, is not limited to these specific materials, as described below.

[0031] Substrate 402 and, hence, channel region 412, may be composed of any material suitable for semiconductor device fabrication. In one embodiment, substrate 402 is a bulk substrate composed of a single crystal of a material which may include, but is not limited to, silicon, germanium, silicon-germanium or a III-V compound semiconductor material. In another embodiment, substrate 402 includes a bulk layer with a top epitaxial layer. In a specific embodiment, the bulk layer is composed of a

single crystal of a material which may include, but is not limited to, silicon, germanium, silicon-germanium, a III-V compound semiconductor material and quartz, while the top epitaxial layer is composed of a single crystal layer which may include, but is not limited to, silicon, germanium, silicon-germanium and a III-V compound semiconductor material. In another embodiment, substrate 402 includes a top epitaxial layer on a middle insulator layer which is above a lower bulk layer. The top epitaxial layer is composed of a single crystal layer which may include, but is not limited to, silicon (i.e. to form a silicon-on-insulator (SOI) semiconductor substrate), germanium, silicon-germanium and a III-V compound semiconductor material. The insulator layer is composed of a material which may include, but is not limited to, silicon dioxide, silicon nitride and silicon oxy-nitride. The lower bulk layer is composed of a single crystal which may include, but is not limited to, silicon, germanium, silicon-germanium, a III-V compound semiconductor material and quartz. Substrate 402 and, hence, channel region 412, may include dopant impurity atoms. In a specific embodiment, channel region 412 is doped P-type and, in an alternative embodiment, channel region 412 is doped N-type.

[0032] Source and drain regions 410 in substrate 402 may be any regions having opposite conductivity to channel region 412. For example, in accordance with an embodiment of the present invention, source and drain regions 410 are N-type doped regions while channel region 412 is a P-type doped region. In one embodiment, substrate 402 and, hence, channel region 412, is composed of boron-doped single-crystal silicon having a boron concentration in the range of $1 \times 10^{15} - 1 \times 10^{19}$ atoms/cm³. Source and drain regions 410 are composed of phosphorous- or arsenic-doped regions having a concentration of N-type dopants in the range of $5 \times 10^{16} - 5 \times 10^{19}$ atoms/cm³. In a specific embodiment, source and drain regions 410 have a depth in substrate 402 in the range of 80-200 nanometers. In accordance with an alternative embodiment of the present invention, source and drain regions 410 are P-type doped regions while channel region 412 is an N-type doped region.

[0033] Tunnel dielectric layer 404A may be any material and have any thickness suitable to allow charge carriers to tunnel into the charge-trapping layer under an applied gate bias while maintaining a suitable barrier to leakage when the device is unbiased. In one embodiment, tunnel dielectric layer 404A is formed by a thermal oxidation process and is composed of silicon dioxide or silicon oxy-nitride, or a combination thereof. In another embodiment, tunnel dielectric layer 404A is formed

by chemical vapor deposition or atomic layer deposition and is composed of a dielectric layer which may include, but is not limited to, silicon nitride, hafnium oxide, zirconium oxide, hafnium silicate, hafnium oxy-nitride, hafnium zirconium oxide and lanthanum oxide. In a specific embodiment, tunnel dielectric layer 404A has a thickness in the range of 1-10 nanometers. In a particular embodiment, tunnel dielectric layer 404A has a thickness of approximately 2 nanometers.

[0034] Multi-layer charge-trapping region 404B may be composed of any material and have any thickness suitable to store charge and, hence, raise the threshold voltage of gate stack 404. In one embodiment, multi-layer charge-trapping region 404B is formed by a chemical vapor deposition process and is composed of a dielectric material which may include, but is not limited to, stoichiometric silicon nitride, silicon-rich silicon nitride and silicon oxy-nitride. In accordance with an embodiment of the present invention, multi-layer charge-trapping region 404B includes a deuterated layer 406 between tunnel dielectric layer 404A and charge trapping layer 408, as depicted in Figure 4. Deuterated layer 406 and charge-trapping layer 408 may be composed of a deuterated derivative and a non-deuterated derivative, respectively, of the same material. For example, in accordance with an embodiment of the present invention, deuterated layer 406 is a deuterated derivative of silicon oxy-nitride, while charge-trapping layer 408 is formed from the hydrogenated derivative of silicon oxy-nitride. In one embodiment, the total thickness of multi-layer charge-trapping region 404B is in the range of 5 - 10 nanometers. In a specific embodiment, the ratio of thicknesses of deuterated layer 506:charge-trapping layer 508 is approximately 1:1, respectively.

[0035] Multi-layer charge-trapping region 404B may have an abrupt interface between deuterated layer 406 and charge-trapping layer 408. That is, in accordance with an embodiment of the present invention, charge-trapping layer 408 is deuterium-free. Alternatively, a gradient of deuterium atom concentration moving from high concentration of deuterium in deuterated layer 406 ranging to low concentration of deuterium in charge-trapping layer 408 may be formed. Thus, in accordance with an alternative embodiment of the present invention, charge-trapping layer 408 is a partially deuterated layer, but having a deuterium concentration less than that of deuterated layer 406.

[0036] Top dielectric layer 404C may be any material and have any thickness suitable to maintain a barrier to charge leakage without significantly decreasing the

capacitance of gate stack 404. In one embodiment, top dielectric layer 404C is formed by a chemical vapor deposition process and is composed of silicon dioxide, silicon oxy-nitride, silicon nitride, or a combination thereof. In another embodiment, top dielectric layer 404C is formed by atomic layer deposition and is composed of a high-k dielectric layer which may include, but is not limited to, hafnium oxide, zirconium oxide, hafnium silicate, hafnium oxy-nitride, hafnium zirconium oxide and lanthanum oxide. In a specific embodiment, top dielectric layer 404C has a thickness in the range of 1-20 nanometers.

[0037] Gate layer 404D may be composed of any conductor or semiconductor material suitable for accommodating a bias during operation of a SONOS-type transistor. In accordance with an embodiment of the present invention, gate layer 404D is formed by a chemical vapor deposition process and is composed of doped poly-crystalline silicon. In another embodiment, gate layer 404D is formed by physical vapor deposition and is composed of a metal-containing material which may include, but is not limited to, metal nitrides, metal carbides, metal silicides, hafnium, zirconium, titanium, tantalum, aluminum, ruthenium, palladium, platinum, cobalt and nickel.

[0038] A nonvolatile charge trap memory device may include a multi-layer charge-trapping region having more than one deuterated layer. Figure 5 illustrates a cross-sectional view of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0039] Referring to Figure 5, semiconductor device 500 includes a gate stack 504 formed over a substrate 502. Semiconductor device 500 further includes source and drain regions 510 in substrate 502 on either side of gate stack 504, defining a channel region 512 in substrate 502 underneath gate stack 504. Gate stack 504 includes a tunnel dielectric layer 504A, a multi-layer charge-trapping region 504B, a top dielectric layer 504C and a gate layer 504D. Thus, gate layer 504D is electrically isolated from substrate 502. Multi-layer charge-trapping region 504B includes a first deuterated layer 506 and a second deuterated layer 516 sandwiching charge-trapping layer 508 of multi-layer charge-trapping region 504B. A pair of dielectric spacers 514 isolates the sidewalls of gate stack 504.

[0040] Semiconductor device 500 may be any semiconductor device described in association with semiconductor device 400 from Figure 4. Substrate 502, source and drain regions 510 and channel region 512 may be composed of any material and

dopant impurity atoms described in association with substrate 402, source and drain regions 410 and channel region 412, respectively, from Figure 4. Tunnel dielectric layer 504A, top dielectric layer 504C and gate layer 504D may be composed of any material described in association with tunnel dielectric layer 404A, top dielectric layer 404C and gate layer 404D, respectively, from Figure 4.

[0041] However, in contrast to semiconductor device 400, semiconductor device includes a multi-layer charge-trapping region 504B having second deuterated layer 516 above charge trapping layer 508, as depicted in Figure 5. First deuterated layer 506 and charge-trapping layer 508 may be composed of any material described in association with deuterated layer 406 and charge-trapping layer 408, respectively, from Figure 4. Additionally, second deuterated layer 516 may also be composed of any material described in association with deuterated layer 406 from Figure 4.

However, in accordance with an embodiment of the present invention, the total thickness of multi-layer charge-trapping region 504B is in the range of 5 - 10 nanometers, i.e. multi-layer charge-trapping region 504B has a thickness in the same range as multi-layer charge-trapping region 404B from Figure 4. Thus, the relative ratios of thicknesses of deuterated layers and the charge-trapping layer may differ from those of semiconductor device 400. For example, in one embodiment, the ratio of thicknesses of first deuterated layer 506:charge-trapping layer 508:second deuterated layer 516 is approximately 1:2:1, respectively.

[0042] As with multi-layer charge-trapping region 404B from Figure 4, multi-layer charge-trapping region 504B may have an abrupt interface between first deuterated layer 506 and charge-trapping layer 508. Likewise, as second abrupt interface may exist between second deuterated layer 516 and charge-trapping layer 508. That is, in accordance with an embodiment of the present invention, charge-trapping layer 508 is deuterium-free. Alternatively, a gradient of deuterium atom concentration moving from high concentration of deuterium in first and second deuterated layers 506 and 516 ranging to low concentration of deuterium in charge-trapping layer 508 may be formed. Thus, in accordance with an alternative embodiment of the present invention, charge-trapping layer 508 is a partially deuterated layer, but having a deuterium concentration less than that of deuterated layers 506 and 516.

[0043] A nonvolatile charge trap memory device may be fabricated to include a multi-layer charge-trapping region having a deuterated layer. Figures 6A-I illustrate

cross-sectional views representing steps in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0044] Referring to Figure 6A, a substrate 602 is provided. Substrate 602 may be composed of any material and have any characteristics described in association with substrates 402 and 502 from Figures 4 and 5, respectively.

[0045] Referring to Figure 6B, a tunnel dielectric layer 620 is formed on the top surface of substrate 602. Tunnel dielectric layer 620 may be formed from any material, from any process, and have any thickness described in association with tunnel dielectric layers 404A and 504A from Figures 4 and 5, respectively.

[0046] Referring to Figure 6C, a multi-layer charge-trapping region 622 is formed on the top surface of tunnel dielectric layer 620. In accordance with an embodiment of the present invention, multi-layer charge-trapping region 622 includes a deuterated layer 624 between tunnel dielectric layer 620 and a charge-trapping layer 626, as depicted in Figure 6C. Deuterated layer 624 and charge-trapping layer 626 may be composed of any materials and have any thicknesses described in association with deuterated layer 406 and charge-trapping layer 408, respectively, from Figure 4. Multi-layer charge-trapping region 622 and, hence, deuterated layer 624 and charge-trapping layer 626 may be formed by any process suitable to provide substantially uniform coverage above tunnel dielectric layer 620. In accordance with an embodiment of the present invention, multi-layer charge-trapping region 622 is formed by a chemical vapor deposition process. In one embodiment, deuterated layer 624 is formed first using deuterated formation gases and, subsequently, charge-trapping layer 626 is formed next using non-deuterated formation gases. In a specific embodiment, multi-layer charge-trapping region 622 is composed substantially of silicon oxy-nitride, wherein deuterated layer 624 is first formed using formation gases such as, but not limited to, deuterated silane (SiD_4), deuterated dichlorosilane (SiD_2Cl_2), nitrous oxide (N_2O), deuterated ammonia (ND_3) and oxygen (O_2). Charge-trapping layer 626 is then formed using formation gases such as, but not limited to, non-deuterated-bis(tert-butylamino)silane (non-deuterated-BTBAS), silane (SiH_4), dichlorosilane (SiH_2Cl_2), N_2O , ammonia (NH_3) and O_2 . In a particular embodiment, deuterated layer 624 and charge-trapping layer 626 are formed in the same process step, i.e. they are formed in the same process chamber with a seamless transition from deuterated formation gases to non-deuterated formation gases.

[0047] An abrupt deuterated and non-deuterated junction may be present at the interface of deuterated layer 624 and charge-trapping layer 626. Thus, in accordance with an embodiment of the present invention, charge-trapping layer 626 remains deuterium-free. Alternatively, some of the deuterium present in deuterated layer 624 may migrate to charge-trapping layer 626 during the deposition of charge-trapping layer 626 or during subsequent high temperature process steps. That is, a gradient of deuterium atom concentration moving from high concentration of deuterium in deuterated layer 624 ranging to low concentration of deuterium in charge-trapping layer 626 may be formed. Thus, in accordance with an alternative embodiment of the present invention, charge-trapping layer 626 becomes a partially deuterated layer, but having a deuterium concentration less than that of deuterated layer 624. In a specific embodiment, deuterated formation gases are employed to form a partially deuterated charge-trapping layer 626 having a deuterium concentration less than that of deuterated layer 624.

[0048] Referring to Figure 6D, a top dielectric layer 628 is formed on the top surface of multi-layer charge-trapping region 622. Top dielectric layer 628 may be formed from any material, from any process, and have any thickness described in association with top dielectric layers 404C and 504C from Figures 4 and 5, respectively. In accordance with an alternative embodiment of the present invention, top dielectric layer 628 is formed by using deuterated formation gases. In such an embodiment, deuterated top dielectric layer 628 subsequently acts as a source of deuterium to form a trap-free layer in multi-layer charge-trapping region 622 during a subsequent anneal process. In a specific alternative embodiment, deuterated top dielectric layer 628 is formed using formation gases such as, but not limited to, SiD_4 , SiD_2Cl_2 and N_2O .

[0049] Referring to Figure 6E, a gate layer 630 is formed on the top surface of top dielectric layer 628. Gate layer 630 may be formed from any material and from any process described in association with gate layers 404D and 504D from Figures 4 and 5, respectively. Thus, a gate stack 632 may be formed above substrate 602.

[0050] Referring to Figure 6F, gate stack 632 is patterned to form a patterned gate stack 604 above substrate 602. Patterned gate stack 604 includes a patterned tunnel dielectric layer 604A, a patterned multi-layer charge-trapping region 604B, a patterned top dielectric layer 604C, and a patterned gate layer 604D. Patterned multi-layer charge-trapping region 604B includes a patterned deuterated layer 606 and a

patterned charge-trapping layer 608. Gate stack 632 may be patterned to form patterned gate stack 604 by any process suitable to provide substantially vertical sidewalls for gate stack 604 with high selectivity to substrate 602. In accordance with an embodiment of the present invention, gate stack 632 is patterned to form patterned gate stack 604 by a lithography and etch process. In a specific embodiment, the etch process is an anisotropic etch process utilizing gases such as, but not limited to, carbon tetrafluoride (CF₄), O₂, hydrogen bromide (HBr) and chlorine (Cl₂).

[0051] Referring to Figure 6G, it may be desirable to implant dopant impurity atoms 640 into the exposed portions of substrate 604 to form source and drain tip extension regions 650. Source and drain tip extension regions 650 will ultimately become part of source and drain regions subsequently formed, as described below. Thus, by forming source and drain tip extension regions 650 as defined by the location of patterned gate stack 604, channel region 612 may be defined, as depicted in Figure 6G. In one embodiment, the conductivity type and the concentration of dopant impurity atoms used to form source and drain tip extension regions 650 are substantially the same as those used to form source and drain regions, described below.

[0052] Referring to Figure 6H, it may be desirable to form a pair of dielectric spacers 614 on the sidewalls of patterned gate stack 604. Finally, referring to Figure 6I, source and drain regions 610 are formed by implanting dopant impurity atoms 660 into the exposed portions of substrate 604. Source and drain regions 610 may have any characteristics as those described in association with source and drain regions 410 and 510 from Figures 4 and 5, respectively. In accordance with an embodiment of the present invention, the profile of source and drain regions 610 is defined by dielectric spacers 614, patterned gate stack 604 and source and drain tip extension regions 650, as depicted in Figure 6I.

[0053] A nonvolatile charge trap memory device may be fabricated to include a multi-layer charge-trapping region having more than one deuterated layer. Figures 7A-C illustrate cross-sectional views representing steps in the formation of a nonvolatile charge trap memory device, in accordance with an embodiment of the present invention.

[0054] Referring to Figure 7A, a tunnel dielectric layer 720 formed on the top surface of a substrate 702 is provided. Substrate 702 may be composed of any material and have any characteristics described in association with substrates 402 and

502 from Figures 4 and 5, respectively. Tunnel dielectric layer 720 may be formed from any material, from any process, and have any thickness described in association with tunnel dielectric layers 404A and 504A from Figures 4 and 5, respectively.

[0055] Referring to Figure 7B, a multi-layer charge-trapping region 722 is formed on the top surface of tunnel dielectric layer 720. In accordance with an embodiment of the present invention, multi-layer charge-trapping region 722 includes a first deuterated layer 724 between tunnel dielectric layer 720 and a charge-trapping layer 726. Additionally multi-layer charge-trapping region 722 includes a second deuterated layer 727 on the top surface of charge-trapping layer 726, as depicted in Figure 7B. First deuterated layer 724, charge-trapping layer 726, and second deuterated layer 727 may be composed of any materials and have any thicknesses described in association with first deuterated layer 506, charge-trapping layer 508, and second deuterated layer 516, respectively, from Figure 5. Multi-layer charge-trapping region 722 and, hence, first and second deuterated layers 724 and 727 and charge-trapping layer 726 may be formed by any process suitable to provide substantially uniform coverage above tunnel dielectric layer 720. In accordance with an embodiment of the present invention, multi-layer charge-trapping region 722 is formed by a chemical vapor deposition process. In one embodiment, first deuterated layer 724 is formed first using deuterated formation gases, charge-trapping layer 726 is formed next using non-deuterated formation gases and, finally, second deuterated layer 727 is formed using deuterated formation gases. In a specific embodiment, multi-layer charge-trapping region 722 is composed substantially of silicon oxynitride, wherein first deuterated layer 724 is formed first using formation gases such as, but not limited to, SiD_4 , SiD_2Cl_2 , N_2O , ND_3 and O_2 . Charge-trapping layer 626 is then formed using formation gases such as, but not limited to, non-deuterated-BTBAS, SiH_4 , SiH_2Cl_2 , N_2O , NH_3 and O_2 . Finally, second deuterated layer 727 is formed using formation gases such as, but not limited to, SiD_4 , SiD_2Cl_2 , N_2O , ND_3 and O_2 . In a particular embodiment, first deuterated layer 724, charge-trapping layer 726 and second deuterated layer 727 are formed in the same process step, i.e. in the same process chamber with a seamless transition from deuterated formation gases to non-deuterated formation gases and back to deuterated formation gases.

[0056] An abrupt deuterated and non-deuterated junction may be present at the interfaces of first deuterated layer 724, second deuterated layer 727 and charge-trapping layer 726. Thus, in accordance with an embodiment of the present invention,

charge-trapping layer 726 remains deuterium-free. Alternatively, some of the deuterium present in first and second deuterated layers 724 and 727 may migrate to charge-trapping layer 726 during the deposition of charge-trapping layer 726 and second deuterated layer 727 or during subsequent high temperature process steps. That is, a gradient of deuterium atom concentration moving from high concentration of deuterium in first and second deuterated layers 724 and 727 ranging to low concentration of deuterium in charge-trapping layer 726 may be formed. Thus, in accordance with an alternative embodiment of the present invention, charge-trapping layer 726 becomes a partially deuterated layer, but having a deuterium concentration less than that of first and second deuterated layers 724. In a specific embodiment, deuterated formation gases are employed to form a partially deuterated charge-trapping layer 726 having a deuterium concentration less than that of deuterated layer 724.

[0057] Referring to Figure 7C, process steps similar to those described in association with Figures 6D-I are carried out to form a nonvolatile charge trap memory device having more than one deuterated layer. Thus, a patterned gate stack 704 is formed over a substrate 702. Source and drain regions 710 are formed on either side of patterned gate stack 704, defining a channel region 712. Patterned gate stack 704 includes a patterned tunnel dielectric layer 704A, a patterned multi-layer charge-trapping region 704B, a patterned top dielectric layer 704C and a patterned gate layer 704D. Patterned multi-layer charge-trapping region 704B includes a patterned first deuterated layer 706 and a patterned second deuterated layer 716 sandwiching patterned charge-trapping layer 708.

[0058] Thus, a nonvolatile charge trap memory device has been disclosed. The device includes a substrate having a channel region and a pair of source and drain regions. A gate stack is above the substrate over the channel region and between the pair of source and drain regions. In accordance with an embodiment of the present invention, the gate stack includes a multi-layer charge-trapping region having a first deuterated layer. In one embodiment, the multi-layer charge-trapping region further includes a deuterium-free charge-trapping layer. In an alternative embodiment, the multi-layer charge-trapping region includes a partially deuterated charge-trapping layer having a deuterium concentration less than that of the first deuterated layer.

CLAIMS

What is claimed is:

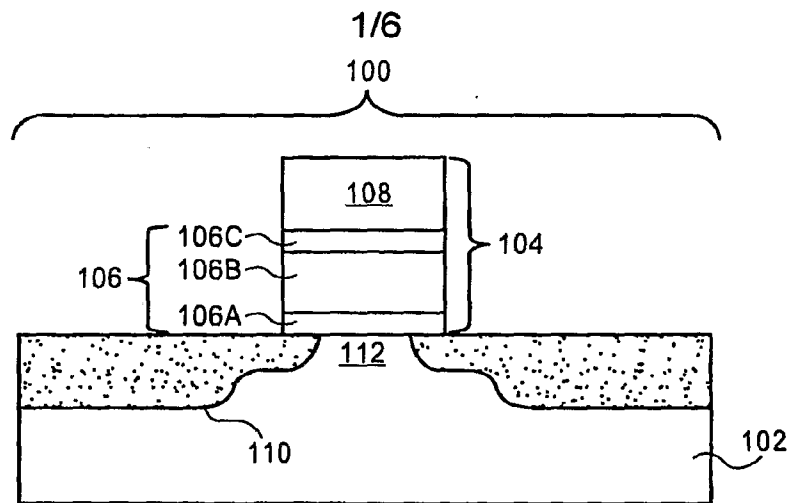
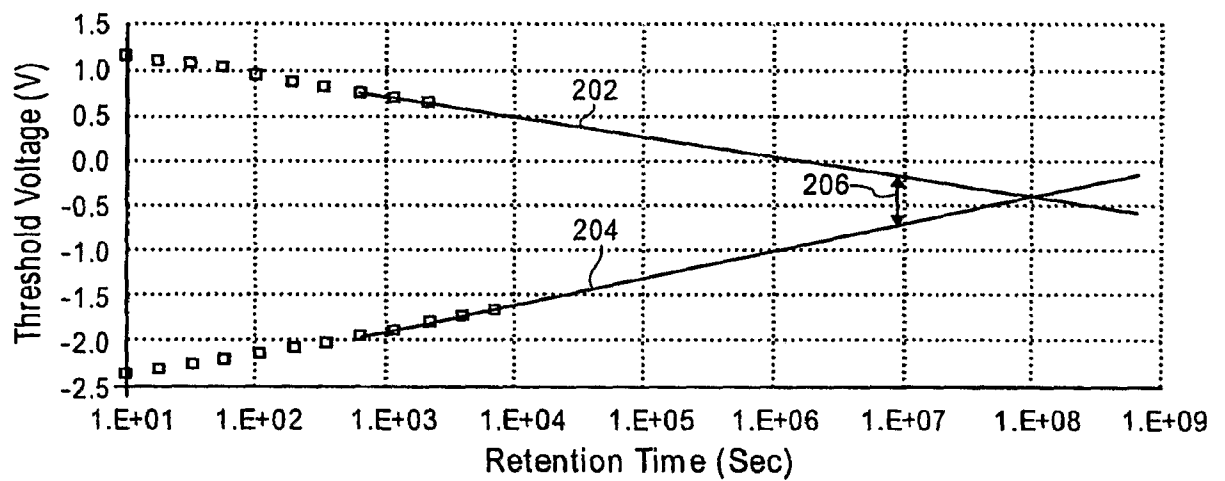
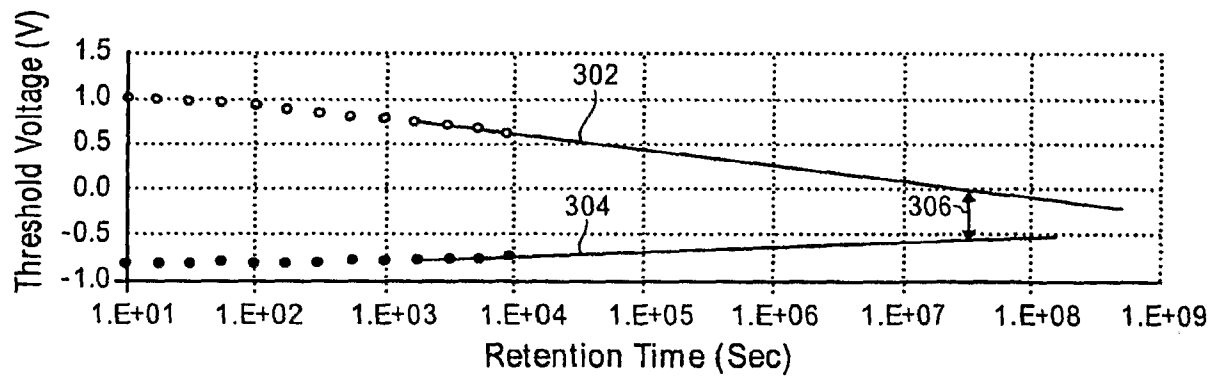
1. A nonvolatile charge trap memory device, comprising:
a substrate having a channel region, a source region and a drain region; and
a gate stack disposed above the substrate over the channel region and between the source region and the drain region, wherein the gate stack comprises a multi-layer charge-trapping region having at least a deuterated layer.
2. The nonvolatile charge trap memory device of claim 1 wherein the multi-layer charge-trapping region further comprises a second deuterated layer having a deuterium concentration less than that of the deuterated layer.
3. The nonvolatile charge trap memory device of claim 2 wherein a total thickness of the multi-layer charge-trapping region is approximately in a range of 5 - 10 nanometers.
4. The nonvolatile charge trap memory device of claim 3 wherein a ratio of a thickness of the deuterated layer and a thickness of the second deuterated layer is approximately 1:1, respectively.
5. The nonvolatile charge trap memory device of claim 1 wherein the multi-layer charge-trapping region further comprises a deuterium-free layer.
6. The nonvolatile charge trap memory device of claim 5 wherein a total thickness of the multi-layer charge-trapping region is approximately in a range of 5 - 10 nanometers.
7. The nonvolatile charge trap memory device of claim 6 wherein a ratio of a thickness of the deuterated layer and a thickness of the deuterium-free layer is approximately 1:1, respectively.
8. A nonvolatile charge trap memory device, comprising:
a substrate having a channel region, a source region and a drain region; and

a gate stack disposed above the substrate over the channel region and between the source region and the drain region, wherein the gate stack comprises:

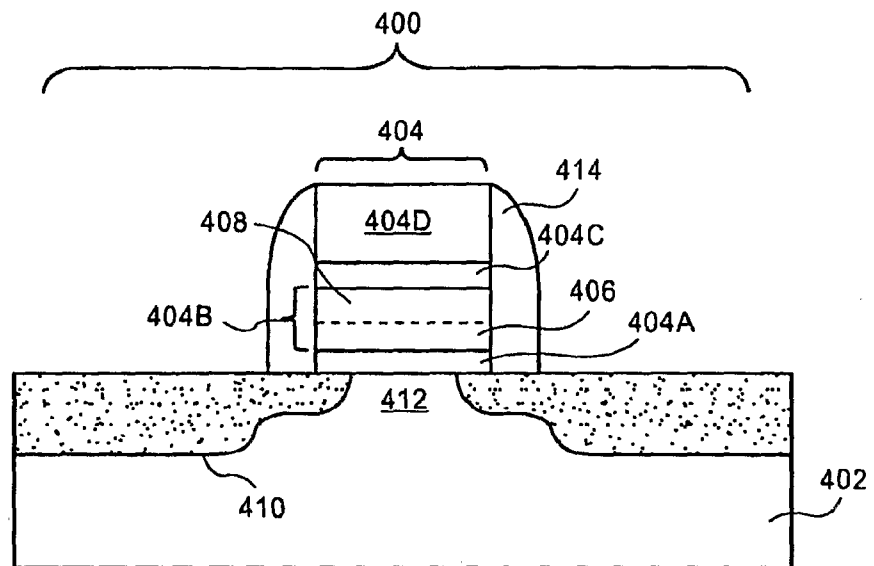
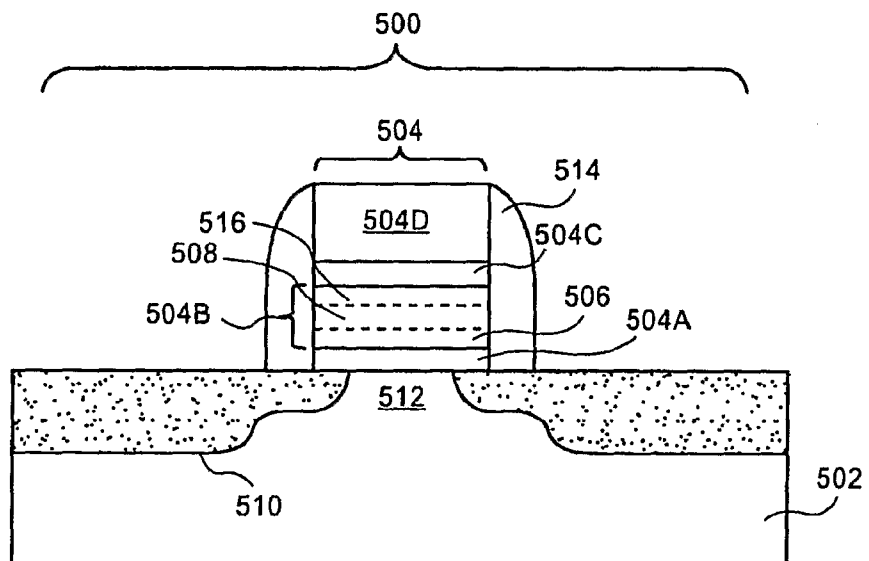
- a tunnel dielectric layer disposed above the channel region;
- a multi-layer charge-trapping region disposed above the tunnel dielectric layer, wherein the multi-layer charge-trapping region comprises a first deuterated layer above the tunnel dielectric layer, a charge-trapping layer disposed above the first deuterated layer, and a second deuterated layer disposed above the charge-trapping layer;
- a top dielectric layer disposed above the second deuterated layer of the multi-layer charge-trapping region; and
- a gate layer disposed above the top dielectric layer.

9. The nonvolatile charge trap memory device of claim 8 wherein the charge-trapping layer of the multi-layer charge-trapping region has a concentration of deuterium less than that of the first and the second deuterated layers.
10. The nonvolatile charge trap memory device of claim 9 wherein a total thickness of the multi-layer charge-trapping region is approximately in a range of 5 - 10 nanometers.
11. The nonvolatile charge trap memory device of claim 10 wherein a ratio of a thickness of the first deuterated layer, a thickness of the charge-trapping layer, and a thickness of the second deuterated layer is approximately 1:2:1, respectively.
12. The nonvolatile charge trap memory device of claim 8 wherein the charge-trapping layer of the multi-layer charge-trapping region is deuterium-free.
13. The nonvolatile charge trap memory device of claim 12 wherein a total thickness of the multi-layer charge-trapping region is approximately in a range of 5 - 10 nanometers.
14. The nonvolatile charge trap memory device of claim 13 wherein a ratio of a thickness of the first deuterated layer, a thickness of the charge-trapping layer, and a thickness of the second deuterated layer is approximately 1:2:1, respectively.

15. A method of fabricating a nonvolatile charge trap memory device, comprising:
providing a substrate;
forming a gate stack above the substrate, wherein forming the gate stack comprises:
forming a tunnel dielectric layer above the substrate;
forming a multi-layer charge-trapping region above the tunnel dielectric layer, wherein the multi-layer charge-trapping region comprises a deuterated layer;
forming a top dielectric layer above the multi-layer charge-trapping region;
forming a gate layer above the top dielectric layer; and
patterning the tunnel dielectric layer, the multi-layer charge-trapping region, the top dielectric layer, and the gate layer; and
forming a source region and a drain region in the substrate and on other side of the gate stack to provide a channel region in the substrate and below the gate stack.
16. The method of claim 15 wherein forming the deuterated layer of the multi-layer charge-trapping region comprises a deposition process utilizing a formation gas selected from the group consisting of SiD_4 , SiD_2Cl_2 and ND_3 .
17. The method of claim 15 wherein the multi-layer charge-trapping region is formed in a single process step.
18. The method of claim 15 wherein forming the multi-layer charge-trapping region further comprises forming a deuterium-free layer.
19. The method claim 18 wherein a ratio of a thickness of the deuterated layer and a thickness of the deuterium-free layer is approximately 1:1, respectively.
20. The method of claim 19 wherein forming the top dielectric layer further comprises forming a deuterated top dielectric layer with a gas selected from the group consisting of SiD_4 , SiD_2Cl_2 and ND_3 .

**FIG. 1****FIG. 2****FIG. 3**

2/6

**FIG. 4****FIG. 5**

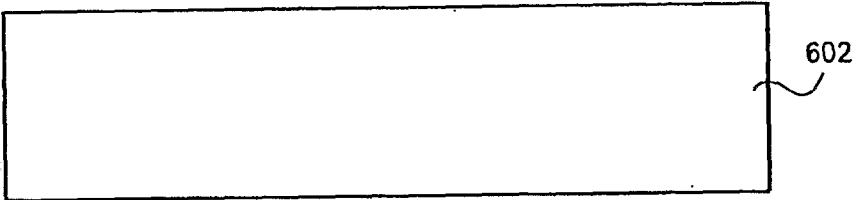


FIG. 6A

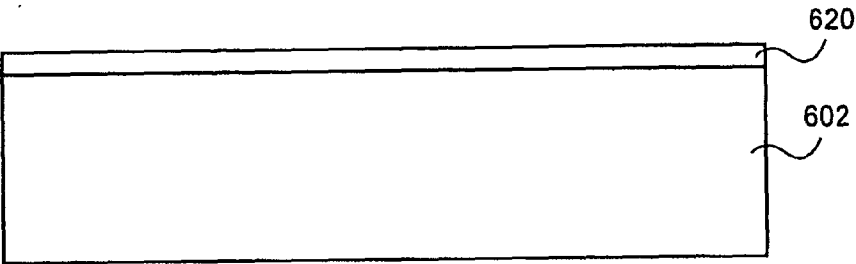


FIG. 6B

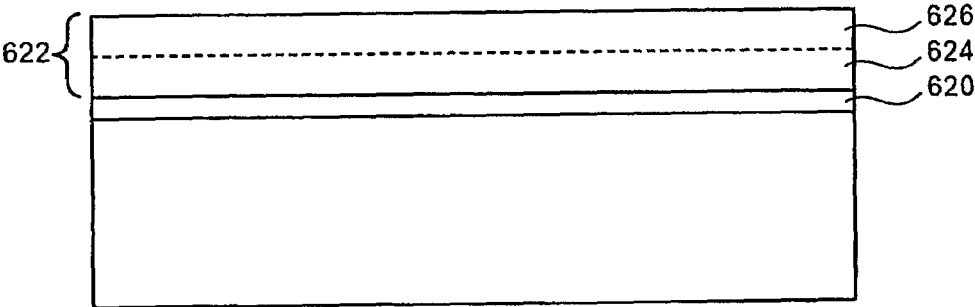


FIG. 6C

4/6

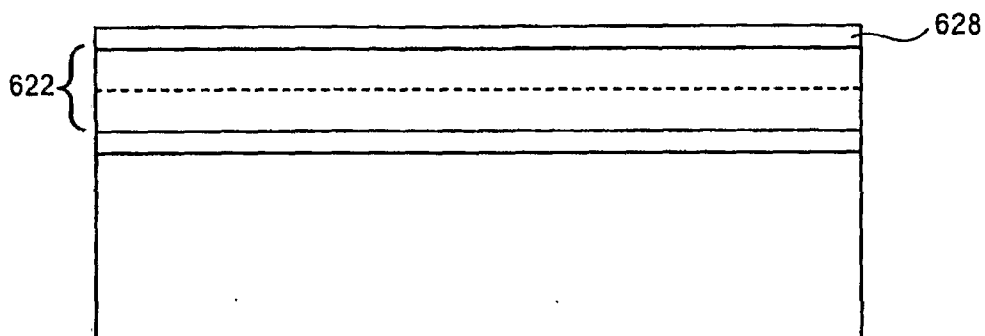


FIG. 6D

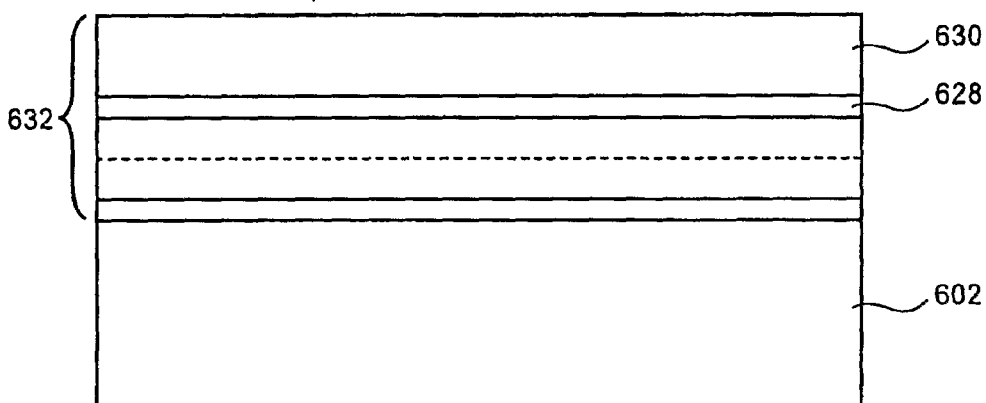


FIG. 6E

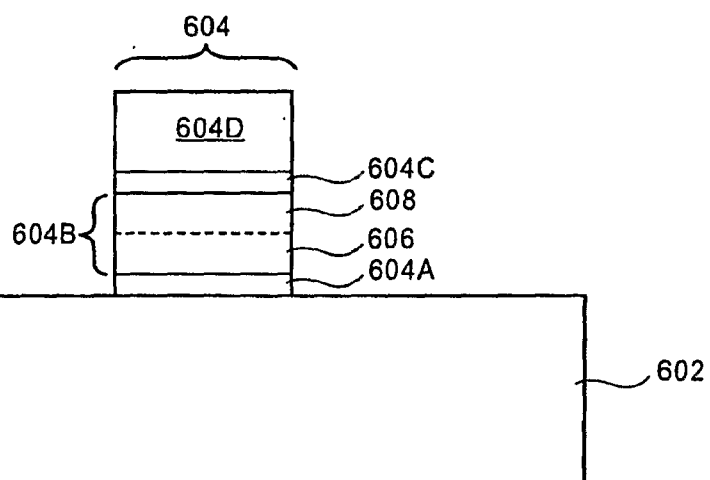


FIG. 6F

5/6

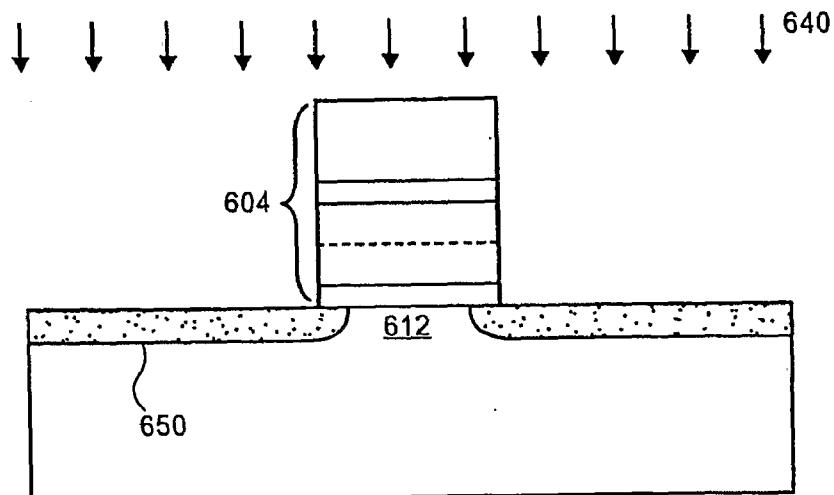


FIG. 6G

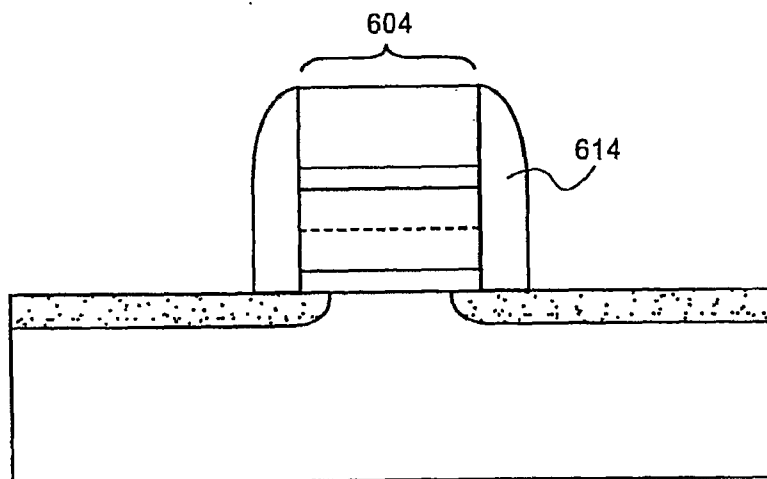


FIG. 6H

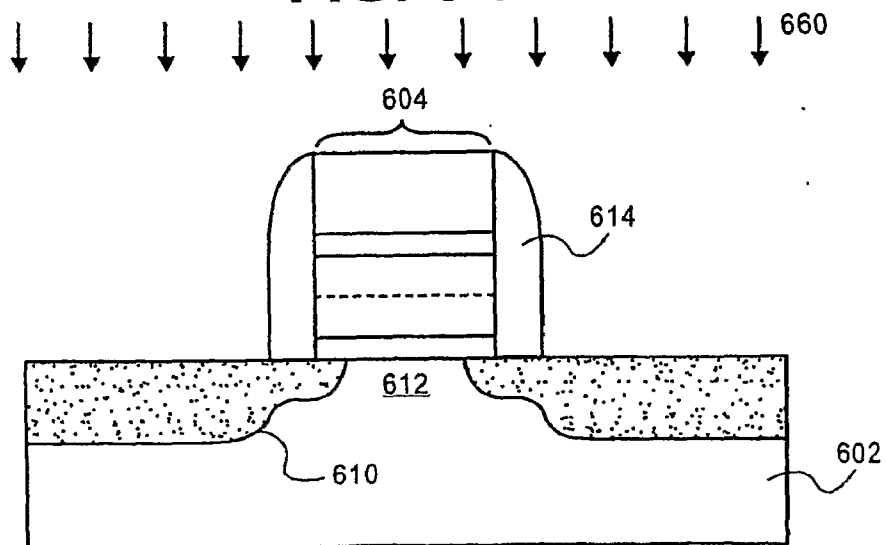


FIG. 6I

6/6

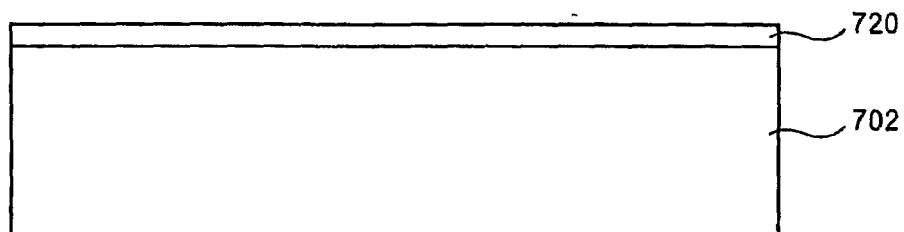


FIG. 7A

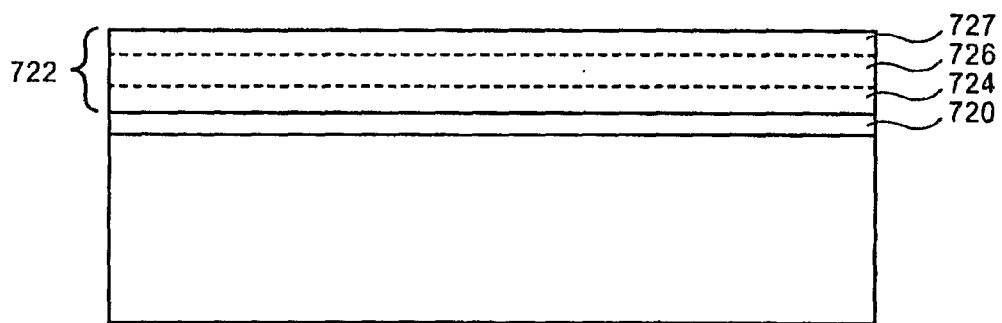


FIG. 7B

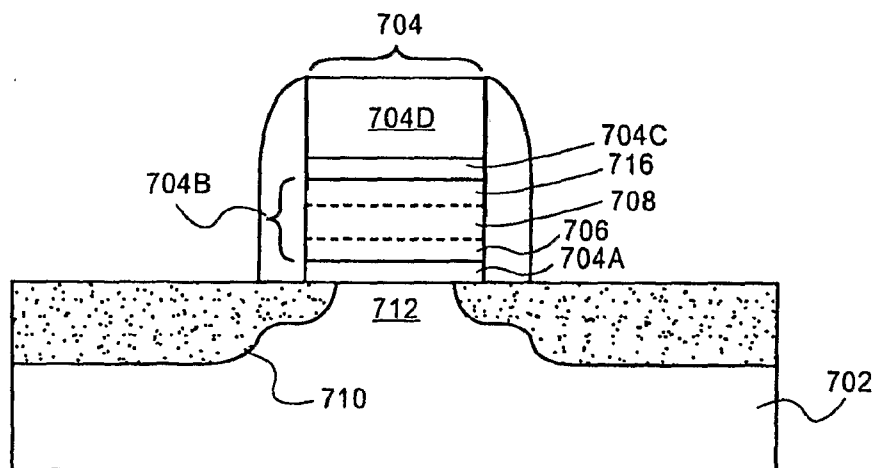


FIG. 7C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 07/20965

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - H01L 29/792 (2008.01)

USPC - 257/324

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)
USPC - 257/324

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 257/288, 314, 324, E21.423, E29.309; 438/287, 475; 365/129 -- text search, see search terms below

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

PubWEST(PGPB,USPT,USOC,EPAB,JPAB); DialogPRO(Engineering); Google Scholar

Search Terms Used: charge, trap, gate, stack, deuterat, deuterium, deuter, layer, multi, two, second, plurality, dielectric, tunnel, top, free, thickness, ratio, nanometer, substrate, concentrat

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2006/0192248 A1 (WANG) 31 August 2006 (31.08.2006), especially Figs 2-4; para [0009]-[0010], [0013], [0020]-[0025].	1-20
Y	US 2004/0251521 A1 (TANAKA et al.) 16 December 2004 (16.12.2004), especially Fig 17; para [0009], [0016], [0084], [0118]-[0121], [0143], [0168]-[0170], [0176]-[0180].	1-20
Y	US 6,670,241 B1 (KAMAL et al.) 30 December 2003 (30.12.2003), especially Fig 4; col 1, ln 8-9; col 2, ln 25-36, 60-63; col 4, ln 62-66; col 5, ln 8-10, 15-17.	5-7, 9-14, 18-20

☐ Further documents are listed in the continuation of Box C.

* 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

17 January 2008 (17.01.2008)

Date of mailing of the international search report

21 APR 2008

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:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774