MULTI-STATE NROM DEVICE

RELATED APPLICATIONS

This application is a continuation-in-part of United States Patent Application Serial Number 10/738,408 filed December 16, 2003 that is assigned to the assignee of the present invention and incorporated herein by reference in its entirety.

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to memory devices and in particular the present invention relates to NROM flash memory devices with high storage density.

BACKGROUND OF THE INVENTION

Memory devices are typically provided as internal, semiconductor, integrated circuits in computers or other electronic devices. There are many different types of memory including random-access memory (RAM), read only memory (ROM), dynamic random access memory (DRAM), synchronous dynamic random access memory (SDRAM), and flash memory. One type of flash memory is a nitride read only memory (NROM). NROM has some of the characteristics of flash memory but does not require the special fabrication processes of flash memory. NROM integrated circuits can be implemented using a standard CMOS process.

Flash memory devices have developed into a popular source of non-volatile memory for a wide range of electronic applications. Flash memory devices typically use a one-transistor memory cell that allows for high memory densities, high reliability, and low power consumption. Common uses for flash memory include personal computers, personal digital assistants (PDAs), digital cameras, and cellular telephones. Program code and system data such as a basic input/output system (BIOS) are typically stored in flash memory devices for use in personal computer systems.

As computers and software become more complex, greater amounts of memory are required to store data. Memory capacity can be increased by reducing transistor size (e.g.,

feature size "F") and/or storing multiple bits in one cell. Performing both of these options simultaneously greatly increases memory capacity while increasing the speed and decreasing the power requirements of the memory device. However, a problem with decreased NROM flash memory size is that NROM flash memory cell technologies have some scaling limitations. As dimensions are scaled down it becomes difficult to maintain adequate separation between multiple charge storage regions of the NROM cell.

For the reasons stated above, and for other reasons stated below which will become apparent to those skilled in the art upon reading and understanding the present specification, there is a need in the art for a higher performance flash memory transistor that can store multiple bits per cell.

SUMMARY

The embodiments of the present invention encompass a nitride read only memory device comprising a substrate with a plurality of vertical pillars, each pillar having an upper doped region. A gate insulator layer is formed along facing sides of a first pillar and a second pillar of the plurality of vertical pillars. A control gate is formed overlying the gate insulator layers and the pillars. A lower doped region is formed under a trench located between the first and second pillars. During operation of the transistor, the lower doped region couples a first channel that forms along the facing side of the first pillar and a second channel that forms along the facing side of the second pillar. In one embodiment, the lower doped region is not connected to an electrical contact.

These and other embodiments, aspects, advantages, and features of the present invention will be set forth in part in the description which follows, and in part will become apparent to those skilled in the art by reference to the following description of the invention and referenced drawings or by practice of the invention. The aspects, advantages, and features of the invention are realized and attained by means of the instrumentalities, procedures, and combinations particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a cross-sectional view of a semiconductor substrate portion at one stage in processing in accordance with an embodiment of the present invention.

Figure 2 is a cross-sectional view of one embodiment of the substrate portion of Figure 1 at a later stage in processing.

Figure 3 is a cross-sectional view of one embodiment of the substrate portion of Figure 2 at a later stage in processing.

Figure 4 is a simplified plan view of a substrate portion showing a portion of a memory cell array, in accordance with an embodiment of the present invention.

Figure 5 is a cross-sectional view illustrating a relationship between the structures of Figures 1-3 and the plan view of Figure 4, in accordance with an embodiment of the present invention.

Figure 6 is a simplified plan view of a memory cell array illustrating an interconnection arrangement for the memory cell array of Figure 4, in accordance with an embodiment of the present invention.

Figure 7 is a cross-sectional view, taken along section lines 7-7 of Figure 6, illustrating part of an interconnection arrangement in accordance with an embodiment of the present invention.

Figure 8 is a cross-sectional view, taken along section lines 8-8 of Figure 6, illustrating part of an interconnection arrangement in accordance with an embodiment of the present invention.

Figure 9A is a block diagram of a metal oxide semiconductor field effect transistor (MOSFET) in a substrate according to the teachings of the prior art.

Figure 9B illustrates the MOSFET of Figure 9A operated in the forward direction showing some degree of device degradation due to electrons being trapped in the gate oxide near the drain region over gradual use.

Figure 9C is a graph showing the square root of the current signal (Ids) taken at the drain region of the conventional MOSFET versus the voltage potential (VGS) established between the gate and the source region.

Figure 10A is a diagram of a programmed MOSFET which can be used as a multi-state cell in accordance with an embodiment of the present invention.

Figure 10B is a diagram suitable for explaining the method by which the MOSFET of the multi-state cell of the present invention can be programmed to achieve the embodiments of the present invention.

Figure 10C is a graph plotting the current signal (Ids) detected at the drain region versus a voltage potential, or drain voltage, (VDS) set up between the drain region and the source region (Ids vs. VDS) in accordance with an embodiment of the present invention.

Figure 11 illustrates a vertical nitride read only memory cell that is part of a memory array of the present invention.

Figure 12 illustrates an electrical equivalent circuit for the portion of the memory array shown in Figure 11.

Figure 13 is another electrical equivalent circuit useful in illustrating a read operation on the novel multi-state cell in accordance with an embodiment of the present invention.

Figure 14 illustrates a portion of a memory array in accordance with an embodiment of the present invention.

Figure 15A illustrates one embodiment of the gate insulator for the embodiments of the present invention having a number of layers.

Figure 15B illustrates the conduction behavior of the multi-state cell of the embodiments of the present invention.

Figure 16A illustrates the operation and programming of the multi-state cell in the reverse direction.

Figure 16B illustrates the now programmed multi-state cell's operation in the forward direction and differential read occurring in this differential cell embodiment, e.g., 2 transistors in each cell.

Figure 17 illustrates a cross-sectional view of one embodiment of an NROM split channel flash memory cell of the embodiments of the present invention.

Figure 18 illustrates a cross-sectional view of another embodiment of an NROM flash memory cell of the embodiments of the present invention.

Figure 19 illustrates an electrical schematic equivalent of the embodiments of Figures 17 and 18.

Figure 20 illustrates one embodiment for a memory device in accordance with an embodiment of the present invention.

Figure 21 is a block diagram of one embodiment of an electronic system, or processor-based system, utilizing a multi-state cell constructed in accordance with the embodiments of the present invention.

DETAILED DESCRIPTION

In the following detailed description of the invention, reference is made to the accompanying drawings that form a part hereof, and in which is shown, by way of illustration, specific embodiments in which the invention may be practiced. In the drawings, like numerals describe substantially similar components throughout the several views. These embodiments are described in sufficient detail to enable those skilled in the art to practice the invention. Other embodiments may be utilized and structural, logical, and electrical changes may be made without departing from the scope of the present invention.

The terms wafer and substrate used in the following description include any structure having an exposed surface with which to form the integrated circuit (IC) structure of the invention. The term substrate is understood to include semiconductor wafers. The term substrate is also used to refer to semiconductor structures during processing, and may include other layers that have been fabricated thereupon. Both wafer and substrate include doped and undoped semiconductors, epitaxial semiconductor layers supported by a base semiconductor or insulator, as well as other semiconductor structures well known to one skilled in the art. The term conductor is understood to include semiconductors, and the term insulator is defined to include any material that is less electrically conductive than the materials referred to as conductors. The following detailed description is, therefore, not to be taken in a limiting sense, and the scope of the present invention is defined only by the appended claims, along with the full scope of equivalents to which such claims are entitled.

Figure 1 is a cross-sectional view of a semiconductor substrate portion 20 at one stage in processing, in accordance with an embodiment of the present invention. The portion 20 includes etched or incised recesses 22, doped regions 24 and 26 and caps 28. The etched recesses 22 form trenches extending along an axis into and out of the page of Figure 1.

In one embodiment, the doped regions 24 are implanted n+ regions. In one embodiment, the doped regions 24 are formed by a blanket implant. In one embodiment, the caps 28 are dielectric caps and may be formed using conventional silicon nitride and conventional patterning techniques. In one embodiment, the etched recesses 22 are then etched using conventional plasma etching techniques. In one embodiment, the doped regions 26 are then doped by implantation to form n+ regions. The etched or incised recesses 22 may be formed by plasma etching, laser-assisted techniques or any other method presently known or that may be developed. In one embodiment, the recesses 22 are formed to have substantially vertical sidewalls relative to a top surface of the substrate portion 20. In one embodiment, substantially vertical means at 90 degrees to the substrate surface, plus or minus ten degrees.

Figure 2 provides a cross-sectional view of the substrate portion 20 of Figure 1 at a later stage in processing, in accordance with an embodiment of the present invention. The portion 20 of Figure 2 includes thick oxide regions 32, ONO regions 34 formed on sidewalls 36 of the recesses 22, gate material 38 and a conductive layer 40. In one embodiment, the gate material 38 comprises conductively-doped polycrystalline silicon.

In one embodiment, conventional techniques are employed to oxidize the doped regions 24 and 26 preferentially with respect to sidewalls 36. As a result, the thick oxide regions 32 are formed at the same time as a thinner oxide 42 on the sidewalls 36. These oxides also serve to isolate the doped regions 24 and 26 from what will become transistor channels along the sidewalls 36. Other techniques for isolation may be employed. For example, in one embodiment, high density plasma grown oxides may be employed. In one embodiment, spacers may be employed.

In one embodiment, conventional techniques are then employed to provide a nitride layer 44 and an oxide layer 46, as is described, for example, in "NROM: A Novel Localized Trapping, 2-Bit Nonvolatile Memory Cell", by Boaz Eitan et al., IEEE Electron Device Letters, Vol. 21, No. 11, November 2000, pp. 543-545, IEEE Catalogue No. 0741-3106/00, or in "A True Single-Transistor Oxide-Nitride-Oxide EEPROM Device" by T.Y.

Chan et al., IEEE Electron Device Letters, Vol. EDL-8, No. 3, March, 1987, pp. 93-95, IEEE Catalogue No. 0741-3106/87/0300-0093.

In one embodiment, the thin oxide 42, nitride layer 44 and oxide layer 46 combine to form the ONO layer 34, such as is employed in SONOS devices, while the polysilicon 38 forms a control gate. In operation, application of suitable electrical biases to the doped regions 24, 26 and the control gate 38 cause hot majority charge carriers to be injected into the nitride layer 44 and become trapped, providing a threshold voltage shift and thus providing multiple, alternative, measurable electrical states representing stored data. "Hot" charge carriers are not in thermal equilibrium with their environment. In other words, hot charge carriers represent a situation where a population of high kinetic energy charge carriers exist. Hot charge carriers may be electrons or holes.

SONOS devices are capable of storing more than one bit per gate 38. Typically, the hot carriers are injected into one side 47 or 47′ of the ONO layer 34, adjacent a contact, such as the region 24 or the region 26, that provides a high electrical field.

By reversing the polarity of the potentials applied to the regions 24 and 26, charge may be injected into the other side 47′ or 47 of the ONO layer 34. Thus, four electronically-discriminable and distinct states can be easily provided with a single gate 38. As a result, the structure shown in Figure 2 is capable of storing at least four bits per gate 38.

Figure 3 is a cross-sectional view of the substrate portion 20 of Figure 1 at an alternative stage in processing, in accordance with an embodiment of the present invention. The embodiment shown in Figure 3 includes the oxide regions 32 and 42, but a floating gate 48 is formed on the thin oxide region 42. A conventional oxide or nitride insulator 49 is formed on the floating gate 48, followed by deposition of gate material 38. Floating gate devices are known and operate by injecting hot charge carriers, which may comprise electrons or holes, into the floating gate 48.

Floating gate devices can be programmed to different charge levels that can be electrically distinct and distinguishable. As a result, it is possible to program more data than one bit into each floating gate device, and each externally addressable gate 38 thus corresponds to more than one stored bit. Typically, charge levels of 0, Q, 2Q and 3Q might be employed, where Q represents some amount of charge corresponding to a reliably distinguishable output signal.

Figure 4 is a simplified plan view of a substrate portion showing a portion of a memory cell array 50, in accordance with an embodiment of the present invention. Figure 4 also provides examples of pitch P, width W, space S and minimum feature size F, as described in the Background. An exemplary memory cell area 52, the physical area of a single transistor, can be seen to be about one F^2. Wordlines 54 are formed from the conductive layer 40, and bitlines 56 and 58 are formed.

Figure 5 is a simplified side view, in section, illustrating a relationship between the structures of Figures 1-3 and the plan view of Figure 4, in accordance with an embodiment of the present invention. The trenches 22 correspond to bitlines 56 and 58, as is explained below in more detail with reference to Figures 6-8.

The density of memory arrays such as that described with reference to Figures 1-5 can require interconnection arrangements that differ from prior art memory arrays. One embodiment of a new type of interconnection arrangement useful with such memory systems is described below with reference to Figures 6-8.

Figure 6 is a simplified plan view illustrating an interconnection arrangement 60 for the memory cell array 50 of Figure 4, in accordance with an embodiment of the present invention. The interconnection arrangement 60 includes multiple patterned conductive layers 62 and 64, separated by conventional interlevel dielectric material 65 (Figures 7 and 8). The views in Figure 6-8 have been simplified to show correspondence with the other Figures and to avoid undue complexity. Shallow trench isolation regions 67 isolate selected portions from one another.

Figure 7 is a cross-sectional view, taken along section lines 7-7 of Figure 6, illustrating part of an interconnection arrangement in accordance with an embodiment of the present invention.

Figure 8 is a cross-sectional view, taken along section lines 8-8 of Figure 6, illustrating part of an interconnection arrangement in accordance with an embodiment of the present invention.

With reference to Figures 6-8, the patterned conductive layer 62 extends upward to nodes 70, 70′, 70″ and establishes electrical communication between the conductive layers 62 and selected portions of the doped region 24. The patterned conductive layer 62 stops at the line denoted 72, 72′.

Similarly, other portions of the patterned conductive layer 62 extend from the line denoted 74, 74′ and extend upward, providing electrical communication from nodes 76, 76′, 76″ to other circuit elements. The nodes 76, 76′, 76″ provide contact to selected portions of the doped region 24.

In contrast, patterned conductive layers 64 extend from top to bottom of Figure 6 and electrically couple to nodes 78, 78″ and thus to doped region 26.

Such is but on example of a simplified interconnection arrangement suitable for use with the memory devices of Figures 1-5. Other arrangements are possible.

Figure 9A is useful in illustrating the conventional operation of a MOSFET such as can be used in a DRAM array. Figure 9A illustrates the normal hot electron injection and degradation of devices operated in the forward direction. As is explained below, since the electrons are trapped near the drain they are not very effective in changing the device characteristics.

Figure 9A is a block diagram of a metal oxide semiconductor field effect transistor (MOSFET) 101 in a substrate 100. The MOSFET 101 includes a source region 102, a drain region 104, a channel region 106 in the substrate 100 between the source region 102 and the drain region 104. A gate 108 is separated from the channel region 108 by a gate oxide 110. A sourceline 112 is coupled to the source region 102. A bitline 114 is coupled to the drain region 104. A wordline 116 is coupled to the gate 108.

In conventional operation, a drain to source voltage potential (Vds) is set up between the drain region 104 and the source region 102. A voltage potential is then applied to the gate 108 via a wordline 116. Once the voltage potential applied to the gate 108 surpasses the characteristic voltage threshold (Vt) of the MOSFET a channel 106 forms in the substrate 100 between the drain region 104 and the source region 102. Formation of the channel 106 permits conduction between the drain region 104 and the source region 102, and a current signal (Ids) can be detected at the drain region 104.

In operation of the conventional MOSFET of Figure 9A, some degree of device degradation does gradually occur for MOSFETs operated in the forward direction by electrons 117 becoming trapped in the gate oxide 110 near the drain region 104. This effect is illustrated in Figure 9B. However, since the electrons 117 are trapped near the drain region 104 they are not very effective in changing the MOSFET characteristics.

Figure 9C illustrates this point. Figure 9C is a graph showing the square root of the current signal (Ids) taken at the drain region versus the voltage potential (VGS) established between the gate 108 and the source region 102. The change in the slope of the plot of $\sqrt{Ids}$ versus VGS represents the change in the charge carrier mobility in the channel 106.

In Figure 9C, ΔVT represents the minimal change in the MOSFET's threshold voltage resulting from electrons gradually being trapped in the gate oxide 110 near the drain region 104, under normal operation, due to device degradation. This results in a fixed trapped charge in the gate oxide 110 near the drain region 104. Slope 1 represents the charge carrier mobility in the channel 106 for Figure 9A having no electrons trapped in the gate oxide 110. Slope 2 represents the charge mobility in the channel 106 for the conventional MOSFET of Figure 9B having electrons 117 trapped in the gate oxide 110 near the drain region 104. As shown by a comparison of slope 1 and slope 2 in Figure 9C, the electrons 117 trapped in the gate oxide 110 near the drain region 104 of the conventional MOSFET do not significantly change the charge mobility in the channel 106.

There are two components to the effects of stress and hot electron injection. One component includes a threshold voltage shift due to the trapped electrons and a second component includes mobility degradation due to additional scattering of carrier electrons caused by this trapped charge and additional surface states. When a conventional MOSFET degrades, or is "stressed," over operation in the forward direction, electrons do gradually get injected and become trapped in the gate oxide near the drain. In this portion of the conventional MOSFET there is virtually no channel underneath the gate oxide. Thus the trapped charge modulates the threshold voltage and charge mobility only slightly.

Applicant has previously described programmable memory devices and functions based on the reverse stressing of MOSFET's in a conventional CMOS process and technology in order to form programmable address decode and correction. (See generally, L. Forbes, W.P. Noble and E.H. Cloud, "MOSFET technology for programmable address decode and correction," United States Patent Application Serial Number 09/383,804). That disclosure, however, did not describe multi-state memory cell solutions, but rather address decode and correction issues.

According to the teachings of the present invention, normal MOSFETs, including split-channel NROM devices, can be programmed by operation in the reverse direction and utilizing avalanche hot electron injection to trap electrons in the gate oxide of the MOSFET. When the programmed MOSFET is subsequently operated in the forward direction the electrons trapped in the oxide are near the source and cause the channel to have two different threshold voltage regions. The novel programmed MOSFETs of the present invention conduct significantly less current than conventional MOSFETs, particularly at low drain voltages. These electrons will remain trapped in the gate oxide unless negative gate voltages are applied. The electrons will not be removed from the gate oxide when positive or zero gate voltages are applied. Erasure can be accomplished by applying negative gate voltages and/or increasing the temperature with negative gate bias applied to cause the trapped electrons to be re-emitted back into the silicon channel of the MOSFET. (See generally, L. Forbes, E. Sun, R. Alders and J. Moll, "Field induced re-emission of electrons trapped in SiO_2," IEEE Trans. Electron Device, vol. ED-26, no. 11, pp. 1816-1818 (Nov. 1979); S.S.B. Or, N. Hwang, and L. Forbes, "Tunneling and Thermal emission from a distribution of deep traps in SiO_2," IEEE Trans. on Electron Devices, vol. 40, no. 6, pp. 1100-1103 (June 1993); S.A. Abbas and R.C. Dockerty, "N-channel IGFET design limitations due to hot electron trapping," IEEE Int. Electron Devices Mtg., Washington D.C., Dec. 1975, pp. 35-38).

Figures 10A-10C are useful in illustrating the present invention in which a much larger change in device characteristics is obtained by programming the device in the reverse direction and subsequently reading the device by operating it in the forward direction.

Figure 10A is a diagram of a programmed MOSFET that can be used as a multi-state cell according to the teachings of the present invention. As shown in Figure 10A the multi-state cell 201 includes a MOSFET in a substrate 200 which has a first source/drain region 202, a second source/drain region 204, and a channel region 206 between the first and second source/drain regions, 202 and 204. In one embodiment, the first source/drain region 202 includes a source region 202 for the MOSFET and the second source/drain region 204 includes a drain region 204 for the MOSFET. Figure 10A further illustrates a gate 208 separated from the channel region 206 by a gate oxide 210. A first transmission line 212 is coupled to the first source/drain region 202 and a second transmission line 214 is coupled to the second source/drain region 204. In one embodiment, the first transmission line includes a sourceline 212 and the second transmission line includes a bit line 214.

As stated above, multi-state cell 201 is comprised of a programmed MOSFET. This programmed MOSFET has a charge 217 trapped in the gate oxide 210 adjacent to the first source/drain region 202 such that the channel region 206 has a first voltage threshold region (Vt1) and a second voltage threshold region (Vt2) in the channel 206. In one embodiment, the charge 217 trapped in the gate oxide 210 adjacent to the first source/drain region 202 includes a trapped electron charge 217. According to the teachings of the present invention and as described in more detail below, the multi-state cell can be programmed to have one of a number of charge levels trapped in the gate insulator adjacent to the first source/drain region 202 such that the channel region 206 will have a first voltage threshold region (Vt1) and a second voltage threshold region (Vt2) and such that the programmed multi-state cell operates at reduced drain source current.

Figure 10A illustrates the Vt2 in the channel 206 is adjacent the first source/drain region 202 and that the Vt1 in the channel 206 is adjacent the second source/drain region 204. According to the teachings of the present invention, Vt2 has a higher voltage threshold than Vt1 due to the charge 217 trapped in the gate oxide 217 adjacent to the first source/drain region 202. Multiple bits can be stored on the multi-state cell 201.

Figure 10B is a diagram suitable for explaining the method by which the MOSFET of the multi-state cell 201 of the present invention can be programmed to achieve the embodiments of the present invention. As shown in Figure 10B the method includes programming the MOSFET in a reverse direction. Programming the MOSFET in the reverse direction includes applying a first voltage potential V1 to a drain region 204 of the MOSFET. In one embodiment, applying a first voltage potential V1 to the drain region 204 of the MOSFET includes grounding the drain region 204 of the MOSFET as shown in Figure 10B. A second voltage potential V2 is applied to a source region 202 of the MOSFET. In one embodiment, applying a second voltage potential V2 to the source region 202 includes applying a high positive voltage potential (VDD) to the source region 202 of the MOSFET, as shown in Figure 10B. A gate potential VGS is applied to a gate 208 of the MOSFET. In one embodiment, the gate potential VGS includes a voltage potential which is less than the second voltage potential V2, but which is sufficient to establish conduction in the channel 206 of the MOSFET between the drain region 204 and the source region 202. As shown in Figure 10B, applying the first, second and gate potentials (V1, V2, and VGS respectively) to the MOSFET creates a hot electron injection into a gate oxide 210 of the MOSFET adjacent

to the source region 202. In other words, applying the first, second and gate potentials (V1, V2, and VGS respectively) provides enough energy to the charge carriers, e.g. electrons, being conducted across the channel 206 that, once the charge carriers are near the source region 202, a number of the charge carriers get excited into the gate oxide 210 adjacent to the source region 202. Here the charge carriers become trapped.

In one embodiment of the present invention, the method is continued by subsequently operating the MOSFET in the forward direction in its programmed state during a read operation. Accordingly, the read operation includes grounding the source region 202 and precharging the drain region a fractional voltage of VDD. If the device is addressed by a wordline coupled to the gate, then its conductivity will be determined by the presence or absence of stored charge in the gate insulator. That is, a gate potential can be applied to the gate 208 by a wordline 216 in an effort to form a conduction channel between the source and the drain regions as done with addressing and reading conventional DRAM cells.

However, now in its programmed state, the conduction channel 206 of the MOSFET will have a first voltage threshold region (Vt1) adjacent to the drain region 204 and a second voltage threshold region (Vt2) adjacent to the source region 202, as explained and described in detail in connection with Figure 10A. According to the teachings of the present invention, the Vt2 has a greater voltage threshold than the Vt1 due to the hot electron injection 217 into a gate oxide 210 of the MOSFET adjacent to the source region 202.

Figure 10C is a graph plotting a current signal (Ids) detected at the second source/drain region 204 versus a voltage potential, or drain voltage, (VDS) set up between the second source/drain region 204 and the first source/drain region 202 (Ids vs. VDS). In one embodiment, VDS represents the voltage potential set up between the drain region 204 and the source region 202. In Figure 10C, the curve plotted as D1 represents the conduction behavior of a conventional MOSFET which is not programmed according to the teachings of the present invention. The curve D2 represents the conduction behavior of the programmed MOSFET, described above in connection with Figure 10A, according to the teachings of the present invention. As shown in Figure 10C, for a particular drain voltage, VDS, the current signal (IDS2) detected at the second source/drain region 204 for the programmed MOSFET (curve D2) is significantly lower than the current signal (IDS1) detected at the second source/drain region 204 for the conventional MOSFET which is not programmed according to the teachings of the present invention. Again, this is attributed to the fact that the channel

206 in the programmed MOSFET of the present invention has two voltage threshold regions and that the voltage threshold, Vt2, near the first source/drain region 202 has a higher voltage threshold than Vt1 near the second source/drain region due to the charge 217 trapped in the gate oxide 217 adjacent to the first source/drain region 202.

Some of these effects have recently been described for use in a different device structure, called an NROM, for flash memories. This latter work in Israel and Germany is based on employing charge trapping in a silicon nitride layer in a non-conventional flash memory device structure. (See generally, B. Eitan et al., "Characterization of Channel Hot Electron Injection by the Subthreshold Slope of NROM device," IEEE Electron Device Lett., Vol. 22, No. 11, pp. 556-558, (Nov. 2001); B. Etian et al., "NROM: A novel localized Trapping, 2-Bit Nonvolatile Memory Cell," IEEE Electron Device Lett., Vol. 21, No. 11, pp. 543-545, (Nov. 2000)). Charge trapping in silicon nitride gate insulators was the basic mechanism used in MNOS memory devices (see generally, S. Sze, Physics of Semiconductor Devices, Wiley, N.Y., 1981, pp. 504-506), charge trapping in aluminum oxide gates was the mechanism used in MIOS memory devices (see generally, S. Sze, Physics of Semiconductor Devices, Wiley, N.Y., 1981, pp. 504-506), and Applicant has previously disclosed charge trapping at isolated point defects in gate insulators (see generally, L. Forbes and J. Geusic, "Memory using insulator traps," US Patent Number 6,140,181, issued October 31, 2000).

In contrast to the above work, the present invention discloses programming a MOSFET in a reverse direction to trap one of a number of charge levels near the source region and reading the device in a forward direction to form a multi-state memory cell based on a modification of DRAM technology.

Prior art DRAM technology generally employs silicon oxide as the gate insulator. Further the emphasis in conventional DRAM devices is placed on trying to minimize charge trapping in the silicon oxide gate insulator. According to the teachings of the present invention, a variety of insulators are used to trap electrons more efficiently than in silicon oxide. That is, in the present invention, the multi-state memory cell employs charge trapping in gate insulators such as, wet silicon oxide, silicon nitride, silicon oxynitride SON, silicon rich oxide SRO, aluminum oxide Al_2O_3, composite layers of these insulators such as oxide and then silicon nitride, or oxide and then aluminum oxide, or multiple layers as oxide-nitride-oxide. While the charge trapping efficiency of silicon oxide may be low such is not the case for silicon nitride or composite layers of silicon oxide and nitride.

Figure 11 illustrates a vertical NROM memory cell that is part of a memory array according to the teachings of the present invention. The memory in Figure 11 is shown illustrating a number of vertical pillars, or multi-state cells, 301-1 and 301-2 formed according to the teachings of the present invention. As one of ordinary skill in the art will appreciate upon reading this disclosure, the number of vertical pillars are formed in rows and columns extending outwardly from a substrate 303.

As shown in Figure 11, the number of vertical pillars, 301-1 and 301-2 are separated by a number of trenches 340. According to the teachings of the present invention, the number of vertical pillars, 301-1 and 301-2, serve as transistors including a first source/drain region, 302-1 and 302-2, respectively. The first source/drain region, 302-1 and 302-2, is coupled to a sourceline 304. As shown in Figure 11, the sourceline 304 is formed in a bottom of the trenches 340 between rows of the vertical pillars, 301-1 and 301-2. In one embodiment, according to the teachings of the present invention, the sourceline 304 is formed from a doped region implanted in the bottom of the trench. A second source/drain region, 306-1 and 306-2 respectively, is coupled to a bitline (not shown). A channel region 305 is located between the first and the second source/drain regions.

As shown in Figure 11, a gate 309 is separated from the channel region 305 by a gate insulator 307 in the trenches 340 along rows of the vertical pillars, 301-1 and 301-2. In one embodiment, according to the teachings of the present invention, the gate insulator 307 includes a gate insulator 307 selected from the group of silicon dioxide (SiO_2) formed by wet oxidation, silicon oxynitride (SON), silicon rich oxide (SRO), and aluminum oxide (Al_2O_3). In another embodiment, according to the teachings of the present invention, the gate insulator 307 includes a gate insulator 307 selected from the group of silicon rich aluminum oxide insulators, silicon rich oxides with inclusions of nanoparticles of silicon, silicon oxide insulators with inclusions of nanoparticles of silicon carbide, and silicon oxycarbide insulators. In another embodiment, according to the teachings of the present invention, the gate insulator 307 includes a composite layer 307. In this embodiment, the composite layer 307 includes a composite layer 307 selected from the group of an oxide-aluminum oxide (Al_2O_3)-oxide composite layer, and oxide-silicon oxycarbide-oxide composite layer. In another embodiment, the composite layer 307 includes a composite layer 307, or a non-stoichiometric single layer, of two or more materials selected from the group of silicon (Si), titanium (Ti), and tantalum (Ta). In another embodiment, according to the teachings of the

present invention, the gate insulator 307 includes an oxide-nitride-oxide (ONO) gate insulator 307.

Figure 12 illustrates an electrical equivalent circuit 400 for the portion of the memory array shown in Figure 11. As shown in Figure 12, a number of vertical multi-state cells, 401-1 and 401-2, are provided. Each vertical multi-state cell, 401-1 and 401-2, includes a first source/drain region, 402-1 and 402-2, a second source/drain region 406-1 and 406-2, a channel region 405 between the first and the second source/drain regions, and a gate 409 separated from the channel region by a gate insulator 407.

Figure 12 further illustrates a number of bit lines, 411-1 and 411-2, coupled to the second source/drain region 406-1 and 406-2 of each multi-state cell. In one embodiment, as shown in Figure 12, the number of bit lines, 411-1 and 411-2, are coupled to the second source/drain region 406-1 and 406-2 along rows of the memory array. A number of word lines, such as wordline 413 in Figure 12, are coupled to the gate 409 of each multi-state cell along columns of the memory array. A number of sourcelines, such as common sourceline 415, are coupled to the first source/drain regions, e.g. 402-1 and 402-2, along columns of the vertical multi-state cells, 401-1 and 401-2, such that adjacent pillars containing these transistors share the common sourceline 415.

In one embodiment, column adjacent pillars include a transistor which operates as a vertical multi-state cell, e.g. 401-1, on one side of a shared trench, the shared trench separating rows of the pillars as described in connection with Figure 11, and a transistor which operates as a reference cell, e.g. 401-2, having a programmed conductivity state on the opposite side of the shared trench. In this manner, according to the teachings of the present invention and as described in more detail below, at least one of multi-state cells can be programmed to have one of a number of charge levels trapped in the gate insulator, shown generally as 417, adjacent to the first source/drain region, e.g. 402-1, such that the channel region 405 will have a first voltage threshold region (Vt1) and a second voltage threshold region (Vt2) and such that the programmed multi-state cell operates at reduced drain source current.

Figure 13 is another electrical equivalent circuit useful in illustrating a read operation on the novel multi-state cell 500 according to the teachings of the present invention. The electrical equivalent circuit in Figure 13 represents a programmed vertical multi-state cell. As explained in detail in connection with Figure 11, the programmed vertical

multi-state cell 500 includes a vertical metal oxide semiconductor field effect transistor (MOSFET) 500 extending outwardly from a substrate. The MOSFET has a source region 502, a drain region 506, a channel region 505 between the source region 502 and the drain region 506, and a gate 509 separated from the channel region 505 by a gate insulator, shown generally as 507.

As shown in Figure 13 a wordline 513 is coupled to the gate 509. A sourceline 504, formed in a trench adjacent to the vertical MOSFET as described in connection with Figure 11, is coupled to the source region 502. A bit line, or data line 511 is coupled to the drain region 506. The multi-state cell 500 shown in Figure 13 is an example of a programmed multi-state cell 500 having one of a number of charge levels trapped in the gate insulator, shown generally as 517, adjacent to the first source/drain region, 502, such that the channel region 505 will have a first voltage threshold region (Vt1) and a second voltage threshold region (Vt2) and such that the programmed multi-state cell 500 operates at reduced drain source current. According to the teachings of the present invention, the second voltage threshold region (Vt2) is now a high voltage threshold region that is greater than the first voltage threshold region (Vt1).

Figure 14 illustrates a portion of a memory array 600 according to the teachings of the present invention. The memory in Figure 14 is shown illustrating a pair of multi-state cells 601-1 and 601-2 formed according to the teachings of the present invention. As one of ordinary skill in the art will understand upon reading this disclosure, any number of multi-state cells can be organized in an array, but for ease of illustration only two are displayed in Figure 14.

As shown in Figure 14, a first source/drain region, 602-1 and 602-2 respectively, is coupled to a sourceline 604. A second source/drain region, 606-1 and 606-2 respectively, is coupled to a bitline, 608-1 and 608-2 respectively. Each of the bitlines, 608-1 and 608-2, couple to a sense amplifier, shown generally at 610. A wordline, 612-1 and 612-2 respectively, is couple to a gate, 614-1 and 614-2 respectively, for each of the multi-state cells, 601-1 and 601-2. According to the teachings of the present invention, the wordlines, 612-1 and 612-2, run across or are perpendicular to the rows of the memory array 600.

Finally, a write data/precharge circuit is shown at 624 for coupling a first or a second potential to bitline 608-1. As one of ordinary skill in the art will understand upon reading this disclosure, the write data/precharge circuit 624 is adapted to couple either a

ground to the bitline 608-1 during a write operation in the reverse direction, or alternatively to precharge the bitline 608-1 to fractional voltage of V_{DD} during a read operation in the forward direction. As one of ordinary skill in the art will understand upon reading this disclosure, the sourceline 604 can be biased to a voltage higher than V_{DD} during a write operation in the reverse direction, or alternatively grounded during a read operation in the forward direction.

As shown in Figure 14, the array structure 600, including multi-state cells 601-1 and 601-2, has no capacitors. Instead, according to the teachings of the present invention, the first source/drain region or source region, 602-1 and 602-2, are coupled directly to the sourceline 604. In order to write, the sourceline 604 is biased to voltage higher than VDD and the devices stressed in the reverse direction by grounding the data or bit line, 608-1 or 608-2. If the multi-state cell, 601-1 or 601-2, is selected by a word line address, 612-1 or 612-2, then the multi-state cell, 601-1 or 601-2, will conduct and be stressed with accompanying hot electron injection into the cells gate insulator adjacent to the source region, 602-1 or 602-2. As one of ordinary skill in the art will understand upon reading this disclosure, a number of different charge levels can be programmed into the gate insulator adjacent to source region such that the cells is used as a differential cell and/or the cell is compared to a reference or dummy cell, as shown in Figure 14, and multiple bits can be stored on the multi-state cell.

During read the multi-state cell, 601-1 or 601-2, is operated in the forward direction with the sourceline 604 grounded and the bit line, 608-1 or 608-2, and respective second source/drain region or drain region, 606-1 and 606-2, of the cells precharged to some fractional voltage of VDD. If the device is addressed by the word line, 612-1 or 612-2, then its conductivity will be determined by the presence or absence of the amount of stored charge trapped in the gate insulator as measured or compared to the reference or dummy cell and so detected using the sense amplifier 610. The operation of DRAM sense amplifiers is described, for example, in U.S. Pat. Nos. 5,627,785; 5,280,205; and 5,042,011, all assigned to Micron Technology Inc., and incorporated by reference herein. The array would thus be addressed and read in the conventional manner used in DRAM's, but programmed as multi-state cells in a novel fashion.

In operation the devices would be subjected to hot electron stress in the reverse direction by biasing the sourceline 604, and read while grounding the sourceline 604 to compare a stressed multi-state cell, e.g. cell 601-1, to an unstressed dummy device/cell, e.g.

601-2, as shown in Figure 14. The write and possible erase feature could be used during manufacture and test to initially program all cells or devices to have similar or matching conductivity before use in the field. Likewise, the transistors in the reference or dummy cells, e.g. 601-2, can all initially be programmed to have the same conductivity states. According to the teachings of the present invention, the sense amplifier 610 can then detect small differences in cell or device characteristics due to stress induced changes in device characteristics during the write operation.

As one of ordinary skill in the art will understand upon reading this disclosure such arrays of multi-state cells are conveniently realized by a modification of DRAM technology. According to the teachings of the present invention a gate insulator of the multi-state cell includes gate insulators selected from the group of thicker layers of SiO_2 formed by wet oxidation, SON silicon oxynitride, SRO silicon rich oxide, Al_2O_3 aluminum oxide, composite layers and implanted oxides with traps (L. Forbes and J. Geusic, "Memory using insulator traps," US Patent Number 6,140,181, issued October 31, 2000). Conventional transistors for address decode and sense amplifiers can be fabricated after this step with normal thin gate insulators of silicon oxide.

Figures 15A-15B and 16A-16B are useful in illustrating the use of charge storage in the gate insulator to modulate the conductivity of the multi-state cell according to the teachings of the present invention. That is, Figures 15A-16B illustrate the operation of the novel multi-state cell 701 formed according to the teachings of the present invention. As shown in Figure 15A, the gate insulator 707 has a number of layers, e.g. an ONO stack, where layer 707A is the oxide layer closest to the channel 705 and a nitride layer 707B is formed thereon.

In the embodiment shown in Figure 15A the oxide layer 707A is illustrated having a thickness of approximately 6.7 nm or 67 Å (roughly 10^{-6} cm). In the embodiment shown in Figure 15A a multi-state cell is illustrated having dimensions of 0.1 μm (10^{-5} cm) by 0.1 μm. For purposes of illustration, the charge storage region near the source can reasonably have dimensions of 0.1 micron (1000 Å) by 0.02 micron (200 Å) in a 0.1 micron technology. If the gate oxide 707A nearest the channel 705 is 67 Å then a charge of 100 electrons will cause a threshold voltage shift in this region of 1.6 Volts since the oxide capacitance is about 0.5 micro-Farad (μF) per square centimeter. If the transistor has a total effective oxide thickness of 200 Å then a change in the threshold voltage of only 0.16 Volts near the source,

corresponding to 10 electrons, is estimated to change the transistor current by 4 micro Amperes (μA). The sense amplifier described in connection with Figure 14, which is similar to a DRAM sense amplifier, can easily sense this charge difference on the data or bitlines. In this embodiment, the sensed charge difference on the data or bitlines will be 40 femto Coulombs (fC) over a sense period of 10 nano seconds (nS).

To illustrate these numbers, the capacitance, Ci, of the structure depends on the dielectric constant, $\in i$, (which for silicon dioxide SiO_2 equates to $1.06/3 \times 10^{-12}$ F/cm), and the thickness of the insulating layers, t, (given here as 6.7×10^{-7} cm), such that $Ci = \in i/t = ((1.06 \times 10^{-12}$ F/cm/$(3 \times 6.7 \times 10^{-7}$ cm)) $= 0.5 \times 10^{-6}$ Farads/cm^2 (F/cm^2). This value taken over the charge storage region near the source, e.g. 20 nm x 100 nm or 2×10^{-11} cm^2, results in a capacitance value of $Ci = 10^{-17}$ Farads. Thus, for a change in the threshold voltage of $\Delta V = 1.6$ Volts the stored charge must be $Q = C \times \Delta V = (10^{-17}$ Farads x 1.6 Volts) $= 1.6 \times 10^{-17}$ Coulombs. Since $Q = Nq$, the number of electrons stored is approximately $Q/q = (1.6 \times 10^{-17}$ Coulombs/1.6×10^{-19} Coulombs) or 100 electrons.

In effect, the programmed multi-state cell, or modified MOSFET is a programmed MOSFET having a charge trapped in the gate insulator adjacent to a first source/drain region, or source region, such that the channel region has a first voltage threshold region (Vt1) and a second voltage threshold region (Vt2), where Vt2 is greater than Vt1, and Vt2 is adjacent the source region such that the programmed MOSFET operates at reduced drain source current. For $\Delta Q = 100$ electrons in the dimensions given above, if the transistor has a total effective oxide thickness of 200 Å then a change in the threshold voltage of only 0.16 Volts near the source, corresponding to 10 electrons, is estimated to change the transistor current by 4 micro Amperes (⌋A). As stated above, the sense amplifier described in connection with Figure 14, which is similar to a DRAM sense amplifier, can easily sense this charge difference on the data or bitlines. The sensed charge difference on the data or bitlines will be 40 femto Coulombs (fC) over a sense period of 10 nano seconds (nS) for this representative one of a number of stored charge levels according to the teachings of the present invention. A number of different charge levels can be programmed into the gate insulator adjacent to source region such that the cell is used as a differential cell and/or the cell is compared to a reference or dummy cell, as shown in Figure 14, and multiple bits can be stored on the multi-state cell of the present invention.

Figure 15B aids to further illustrate the conduction behavior of the novel multi-state cell of the present invention. The electrical equivalent circuit shown in Figure 15B illustrates a multi-state cell 701 having an equivalent oxide thickness of 200 Å. The charge storage region near the source 702 can reasonably have a length dimension of 0.02 micron (20 nm) in a 0.1 micron technology with a width dimension of 0.1 micron (100 nm). Therefore, for a change in the drain source voltage (Δ VDS) in this region an electric field of $E = (0.1\ V/2 \times 10^{-6}\ cm) = 0.5 \times 105\ V/cm$ or $5 \times 104\ V/cm$ is provided. The drain current is calculated using the formula $ID = \mu Cox \times (W/L) \times (Vgs - Vt) \times \Delta VDS$. In this example, $\mu Cox = \mu Ci$ is taken as 50 μA/V2 and W/L = 5. Appropriate substitution into the drain current provides ID = (50 μA/V2 x 5 x 0.16 Volts x 0.1 Volts) = 2.5 x 1.6 μA = 4 μA. As noted above this drain current ID corresponds to 10 electrons trapped in the gate insulator, or charge storage region 707 near the source 702. Sensed over a period of 10 nanoseconds (ns) produces a current on the bitline of 40 fC (e.g. $4\ \mu A \times 10\ nS = 40 \times 10^{-15}$ Coulombs).

Figures 16A and 16B illustrate the operation and programming of the novel multi-state cell as described above. However, Figures 16A and 16B also help illustrate an alternative array configuration where adjacent devices are compared and one of the devices on the opposite side of a shared trench is used as a dummy cell transistor or reference device. Again, the reference devices can all be programmed to have the same initial conductivity state.

Figure 16A illustrates the operation and programming of the novel multi-state cell in the reverse direction. A transistor 801-1 on one side of the trench (as described in connection with Figure 11) is stressed by grounding its respective drain line, e.g. 811-1. The drain line 811-2 for the transistor 801-2 on the opposite side of the trench is left floating. A voltage is applied to the shared sourceline 804 located at the bottom of the trench (as described in connection with Figure 11) that now acts as a drain. The neighboring (shared trench)/column adjacent transistors, 801-1 and 801-2, share a gate 807 and the wordline 813, e.g. polysilicon gate lines, coupling thereto run across or are perpendicular to the rows containing the bit and source lines, e.g. 811-1, 811-2, and 804. A gate voltage is applied to the gates 807. Here the multi-state cell 801-1 will conduct and be stressed with accompanying hot electron injection into the cells gate insulator 817 adjacent to the source region 802-1.

Figure 16B illustrates the now programmed multi-state cell's operation in the forward direction and differential read occurring in a this differential cell embodiment, e.g. 2 transistors in each cell. To read this state the drain and source (or ground) have the normal connections and the conductivity of the multi-state cell is determined. The drain lines 811-1 and 811-2 have the normal forward direction potential applied. The shared sourceline 804 located at the bottom of the trench (as described in connection with Figure 11) is grounded and once again acts as a source. And, a gate voltage is applied to the gates 807.

As one of ordinary skill in the art will understand upon reading this disclosure, a number of different charge levels can be programmed into the gate insulator 817 adjacent to source region 802-1 and compared to the reference or dummy cell, 802-2. Thus, according to the teachings of present invention multiple bits can be stored on the multi-state cell.

As stated above, these novel multi-state cells can be used in a DRAM-like array. Two transistors can occupy an area of $4F^2$ (F = the minimum lithographic feature size) when viewed from above, or each memory cell consisting of one transistor utilizing an area of $2F^2$. Each transistor can now, however, store many bits so the data storage density is much higher than one bit for each $1F^2$ unit area. Using a reference or dummy cell for each memory transistor where the reference transistor is in close proximity, e.g. the embodiment shown in Figures 16A and 16B vs. that shown in Figure 12, results in better matching characteristics of transistors, but a lower memory density.

Figure 17 illustrates a cross-sectional view of an embodiment of a vertical NROM flash memory transistor of the present invention. Use of vertical device structure increases the channel length while keeping the area occupied by the cell to four square feature sizes (i.e., $4F^2$).

This embodiment is comprised of one control gate 1704 and two split channels 1710 and 1711 along the sides of two pillars 1701 and 1702 respectively. An n+ region 1703 under the trench connects the two channel segments 1710 and 1711 during transistor operation so that the structure acts like two transistors in series. In the present embodiment, the two transistors in series have at least two charge storage areas. Alternate embodiments may include different quantities of storage areas.

The transistors each have a nitride storage region 1706 and 1708 that, in one embodiment, is part of an ONO gate insulator layer. Charge can be stored in the gate

insulator in either or both channel segments 1710 and 1711. The n+ regions 1720 and 1721 in the upper portions of adjacent pillars 1701 and 1702 act as either a source region or a drain region, depending on the direction of operation of the transistors. The source/drain regions are coupled by data/bit lines that extend along the z-axis, substantially perpendicular to the wordline/control gate 1704.

The bottom of the trench and the tops of the pillars have an oxide dielectric material between the substrate and the control gate 1704. Alternate embodiments may use other types of dielectric materials.

The embodiment of Figure 17 illustrates n+ regions being doped into a p-type the substrate. However, alternate embodiments may dope p+ regions into an n-type substrate.

Figure 18 illustrates a cross-sectional view of another embodiment of a vertical NROM flash memory transistor of the present invention. Charge can be stored at either end of the channel 1801. As in the embodiment of Figure 17, the n+ regions 1803 and 1804 act as source/drain regions and their function depends on the direction of operation of the transistor.

Figure 19 illustrates a schematic diagram of the electrical equivalent of the embodiments of Figures 17 and 18. The transistors illustrated in Figures 17 and 18 are shown as two field effect transistors (FETs) operating in series with the drain of one coupled to the source of the other.

The reference numbers of Figure 17 have been used in Figure 19 to illustrate the relation of the components of Figure 19 to those of Figure 17. While Figure 19 illustrates the drain 1721 and source 1720 being a certain orientation, if the transistor is operated in the opposite direction, the drain and source regions are opposite.

The floating n+ diffusion area 1703 couples the separate parts 1710 and 1711 of the channel. There is no electrical contact on the n+ region 1703. The single gate 1704 couples the two transistors.

The flash memory cells of the embodiments of Figures 17 and 18 can be fabricated using modifications to the fabrication techniques discussed previously. The structures of Figures 17 and 18 use the same etched vertical pillars but the NROM flash memory structure forms two channels along the sidewalls of the adjacent pillars and an n+ region forms a transistor channel along the bottom of the trench. The single control gate is formed in the trenches between the pillars and the n+ source/drain regions at the tops of the

pillars form the data/bit lines. In the embodiment of Figure 18, a gate insulator and the control gate form a part of the channel across the bottom of the trench.

Conventional channel hot electron injection can be used for programming in which a source region is grounded and a drain region is biased with a positive voltage while the control gate has a positive programming voltage applied. Conventional negative gate Fowler-Nordheim tunneling can be used for erasing the cells. In the present embodiments, the device can be used for two bit storage. The charge is stored near the drain and the device is read in the reverse direction. Either end of the channel can be used as a drain in response to the direction of operation and a charge stored at both ends of the channel near the n+ regions at the surface.

In alternate embodiments, substrate enhanced hot electron injection can be used for programming the NROM cells of the present invention. Additionally, substrate enhanced band-to-band tunneling induced hot hole injection can be used for erasing the cells.

The ONO layer is only one embodiment for a gate insulator of the NROM cells of the present invention. Additional gate insulator compositions include: oxide – nitride – aluminum oxide composite layers, oxide – aluminum oxide – oxide composite insulators, oxide – silicon oxycarbide – oxide composite layers, as well as other composite layers. Additionally, the gate insulator may be thicker than normal silicon oxides formed by wet oxidation and not annealed, silicon rich oxides with inclusions of nanoparticles of silicon, silicon oxynitride layer (not composite layers), silicon rich aluminum oxide insulators (not composite layers), silicon oxycarbide insulators (not composite layers), silicon oxide insulators with inclusions of nanoparticles of silicon carbide, as well as other non-stoichiometric single layers of gate insulators of two or more commonly used insulator materials including, but not limited to, Si, N, Al, Ti, Ta, Hf, and La.

In Figure 20 a memory device is illustrated according to the teachings of the present invention. In one embodiment, the device is an NROM device of the present invention. In an alternate embodiment, it can be a DRAM device of the present invention.

The memory device 940 contains a memory array 942, row and column decoders 944, 948 and a sense amplifier circuit 946. The memory array 942 consists of a plurality of multi-state cells 900, formed according to the teachings of the present invention whose word lines 980 and bit lines 960 are commonly arranged into rows and columns, respectively. The

bit lines 960 of the memory array 942 are connected to the sense amplifier circuit 946, while its word lines 980 are connected to the row decoder 944. Address and control signals are input on address/control lines 961 into the memory device 940 and connected to the column decoder 948, sense amplifier circuit 946 and row decoder 944 and are used to gain read and write access, among other things, to the memory array 942.

The column decoder 948 is connected to the sense amplifier circuit 946 via control and column select signals on column select lines 962. The sense amplifier circuit 946 receives input data destined for the memory array 942 and outputs data read from the memory array 942 over input/output (I/O) data lines 963. Data is read from the cells of the memory array 942 by activating a word line 980 (via the row decoder 944), which couples all of the memory cells corresponding to that word line to respective bit lines 960, which define the columns of the array. One or more bit lines 960 are also activated. When a particular word line 980 and bit lines 960 are activated, the sense amplifier circuit 946 connected to a bit line column detects and amplifies the conduction sensed through a given multi-state cell, where in the read operation the source region of a given cell is couple to a grounded array plate (not shown), and transferred its bit line 960 by measuring the potential difference between the activated bit line 960 and a reference line which may be an inactive bit line. The operation of Memory device sense amplifiers is described, for example, in US Patent Numbers 5,627,785; 5,280,205; and 5,042,011, all assigned to Micron Technology Inc., and incorporated by reference herein.

Figure 21 is a block diagram of an electronic system, or processor-based system, 1000 utilizing multi-state memory cells 1012 constructed in accordance with the embodiments of the present invention. That is, the multi-state memory cells 1012 utilize the DRAM or NROM flash memory cells as described previously.

The processor-based system 1000 may be a computer system, a process control system, or any other system employing a processor and associated memory. The system 1000 includes a central processing unit (CPU) 1002 or other controller circuit (e.g., a microprocessor) that communicates with the multi-state memory 1012 and an I/O device 1008 over a bus 1020. The bus 1020 may be a series of buses and bridges commonly used in a processor-based system, but for convenience purposes only, the bus 1020 has been illustrated as a single bus. A second I/O device 1010 is illustrated, but is not necessary to practice the invention. The processor-based system 1000 can also includes read-only memory

(ROM) 1014 and may include peripheral devices such as a floppy disk drive 1004 and a compact disk (CD) ROM drive 1006 that also communicates with the CPU 1002 over the bus 1020 as is well known in the art.

It will be appreciated by those skilled in the art that additional circuitry and control signals can be provided, and that the memory device 1000 has been simplified to help focus on the invention. At least one of the multi-state cells in NROM 1012 includes a programmed MOSFET having a charge trapped in the gate insulator adjacent to a first source/drain region, or source region, such that the channel region has a first voltage threshold region (Vt1) and a second voltage threshold region (Vt2), where Vt2 is greater than Vt1, and Vt2 is adjacent the source region such that the programmed MOSFET operates at reduced drain source current.

It will be understood that the embodiment shown in Figure 21 illustrates an embodiment for electronic system circuitry in which the novel memory cells of the present invention are used. The illustration of system 1000, as shown in Figure 21, is intended to provide a general understanding of one application for the structure and circuitry of the present invention, and is not intended to serve as a complete description of all the elements and features of an electronic system using the novel memory cell structures. Further, the invention is equally applicable to any size and type of memory device 1000 using the novel memory cells of the present invention and is not intended to be limited to that described above. As one of ordinary skill in the art will understand, such an electronic system can be fabricated in single-package processing units, or even on a single semiconductor chip, in order to reduce the communication time between the processor and the memory device.

Applications containing the novel memory cell of the present invention as described in this disclosure include electronic systems for use in memory modules, device drivers, power modules, communication modems, processor modules, and application-specific modules, and may include multilayer, multichip modules. Such circuitry can further be a subcomponent of a variety of electronic systems, such as a clock, a television, a cell phone, a personal computer, an automobile, an industrial control system, an aircraft, and others.

Conclusion

The novel multi-state cells of the present invention can be used in an NROM flash memory array. Two transistors can occupy an area of $4F^2$ when viewed from above. Each such transistor can now, however, store multiple bits so that the data storage density is much higher than one bit for each $1F^2$ unit area.

It is to be understood that the above description is intended to be illustrative, and not restrictive. Many other embodiments will be apparent to those of skill in the art upon reviewing the above description. The scope of the invention should, therefore, be determined with reference to the appended claims, along with the full scope of equivalents to which such claims are entitled.

What is claimed is:

1. A multi-state NROM transistor comprising:
a substrate having a plurality of vertical pillars, each pillar comprising an upper doped region;
a gate insulator layer formed along facing sides of a first pillar and a second pillar of the plurality of vertical pillars;
a control gate formed overlying the gate insulator layers and the pillars; and
a lower doped region formed under a trench located between the first and second pillars, wherein during transistor operation the lower doped region couples a first channel that forms along the facing side of the first pillar and a second channel that forms along the facing side of the second pillar.

2. The transistor of claim 1 wherein the upper and lower doped regions are n+ doped regions in a p-type substrate and the upper doped region is located substantially at the top of each pillar.

3. The transistor of claim 1 wherein the lower doped region is not coupled to an electrical contact.

4. The transistor of claim 1 wherein the transistor operates as equivalent to a transistor having a size of less than 1.0 lithographic feature squared ($1F^2$).

5. The transistor of claim 1 wherein each gate insulator defines a charge storage region.

6. The transistor of claim 1 wherein the first and second channels are formed between the upper doped region and the lower doped region during a program operation.

7. The transistor of claim 1 wherein the gate insulator layer is comprised of an oxide – nitride – oxide composite structure.

8. The transistor of claim 1 wherein the gate insulator layer is comprised of a composite structure of one of: oxide – nitride – aluminum oxide, oxide – aluminum oxide – oxide, or oxide – silicon oxycarbide – oxide.

9. The transistor of claim 1 wherein the gate insulator layer is comprised of one of the following non-composite structures: silicon oxides formed by wet oxidation and not annealed, silicon rich oxides with inclusions of nanoparticles of silicon, silicon oxynitride layer, silicon rich aluminum oxide insulators, silicon oxycarbide insulators, or silicon oxide insulators with inclusions of nanoparticles of silicon carbide.

10. The transistor of claim 1 wherein the gate insulator layer is comprised of a non-stoichiometric single layer.

11. The transistor of claim 10 wherein the non-stoichiometric single layer comprises one of Si, N, Al, Ti, Ta, Hf, or La.

12. An array of multi-state NROM transistors comprising:

a substrate having a plurality of vertical pillars each separated by a trench, each pillar comprising an upper doped region, each upper doped region coupled to a first bitline of the array;

a plurality of gate insulator layers, each layer formed along facing sides of adjacent pillars of the plurality of vertical pillars;

a control gate formed in the trenches and overlying the plurality of vertical pillars, the control gate forming a wordline between a row of NROM transistors of the array of multi-state NROM transistors; and

a plurality of lower doped regions, each region formed under each trench, wherein during operation of the transistors each lower doped region couples a first channel that forms in a first pillar along a first side of a first trench and a second channel that forms in a second pillar along a second side of the first trench, each lower doped region coupled to a second bitline of the array.

13. The transistor of claim 12 wherein the substrate is comprised of a silicon material and the control gate is comprised of a polysilicon.

14. The transistor of claim 12 wherein the plurality of lower doped regions are not accessible by electrical contacts.

15. The transistor of claim 12 wherein the gate insulator is comprised of one of a composite structure or a non-stoichiometric single layer.

16. The transistor of claim 12 wherein, during transistor operation, the first and second channels act as series coupled transistors having at least two charge storage areas.

17. An electronic system comprising:
a processor circuit that generates memory control signals; and
an NROM flash memory device coupled to the processor circuit, the flash memory device having a plurality of multi-state NROM transistors, each transistor comprising:
a substrate having a plurality of vertical pillars, each pillar comprising an upper doped region and separated by a trench from adjacent pillars;
a first gate insulator layer formed along a first side of a first trench;
a second gate insulator layer formed along a second opposing side of the first trench;
a control gate formed in the first trench and overlying the plurality of vertical pillars; and
a lower doped region formed under the first trench, wherein during transistor operation the lower doped region couples a first channel that forms adjacent to the first gate insulator and a second channel that forms adjacent to the second gate insulator.

18. The system of claim 17 wherein during operation of the transistor, the first and second gate insulators each comprise a charge storage area.

19. The system of claim 17 wherein the first and second gate insulator layers are comprised of an oxide – nitride – oxide composite.

20. The system of claim 17 wherein an upper doped region of a first pillar adjacent the first trench acts as a drain region and an upper doped region of a second pillar adjacent the first trench acts as a source region.

21. A multi-state NROM transistor comprising:
a substrate having a plurality of vertical pillars, each pillar comprising a doped region;
a gate insulator layer formed along facing sides of a first pillar and a second pillar of the plurality of vertical pillars; and
a control gate formed overlying the gate insulator layers and the pillars wherein during transistor operation a channel forms between a doped region of the first pillar and the doped region of the second pillar.

22. The transistor of claim 21 wherein during transistor operation a first doped region operates as a source region and a second doped region operates a drain region.

23. The transistor of claim 21 wherein the transistor operation is a programming operation.

24. An array of multi-state NROM transistors comprising:
a substrate having a plurality of vertical pillars with a trench between each pillar, each pillar comprising a doped region;
a plurality of gate insulator layers, each layer formed along opposing sides of each trench; and
a control gate formed in each trench and overlying the plurality of vertical pillars to form a wordline wherein during transistor operation a channel forms between a doped region of the first pillar and the doped region of the second pillar.

25. The array of claim 24 wherein the doped regions are coupled to bitlines that underlie and are substantially perpendicular to the wordline.

26. A method for manufacturing a split-channel transistor, the method comprising:
incising a substrate to form a plurality of trenches, each pair of trenches defining a pillar;

doping a region in an upper portion of each pillar;

doping a lower region under each trench;

forming a nitride storage region on each facing side of adjacent pillars; and

forming a control gate overlying the pillars and in the plurality of trenches wherein during a programming operation of the transistor, a channel is formed along the facing sides of adjacent pillars and are connected by the lower doped region, the lower doped region not being coupled to an electrical contact.

27. The method of claim 26 and further including forming a dielectric material in each trench between the substrate and the control gate.

28. The method of claim 26 wherein the control gate is polysilicon.

29. The method of claim 26 wherein the split-channel transistor forms two field effect transistors connected in series.

30. The method of claim 26 wherein a first doped region in a first pillar is a drain region and a second doped region in a second, adjacent pillar is a source region.

31. The method of claim 26 wherein the doping comprises creating n+ regions in a p-type substrate.

32. A method for programming a split-channel having a pair of pillars forming a trench, each pillar having a source/drain region and the trench having a floating n+ diffusion region that has no electrical contact, nitride charge storage regions formed along facing sides of the trench and a control gate overlying the nitride charge storage regions and the pair of pillars, the method comprising:

grounding a first source/drain region;

applying a gate voltage to the control gate; and

applying a drain voltage to a second source/drain region such that a channel forms along the facing sides of the trench between the first and second source/drain regions and under the trench.

33. The method of claim 32 wherein the channels along each side of the trench are connected by the floating n+ diffusion region.

34. A method for programming a split-channel having a pair of pillars forming a trench, each pillar having a source/drain region, nitride charge storage regions formed along facing sides of the trench and a control gate overlying the nitride charge storage regions and the pair of pillars, the method comprising:
grounding a first source/drain region;
applying a gate voltage to the control gate; and
applying a drain voltage to a second source/drain region such that a channel forms along the facing sides of the trench between the first and second source/drain regions and under the trench.

35. A method for manufacturing a split-channel transistor, the method comprising:
incising a substrate to form a plurality of trenches, each pair of trenches defining a pillar;
doping a region in an upper portion of each pillar;
forming a nitride storage region on each facing side of adjacent pillars; and
forming a control gate overlying the pillars and in the plurality of trenches wherein during a programming operation of the transistor, a channel is formed along the facing sides of adjacent pillars and are connected together under each trench.