

NROM FLASH MEMORY WITH SELF-ALIGNED STRUCTURAL CHARGE SEPARATION

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to memory devices and in particular the present invention relates to nitride read only memory flash memory devices.

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.

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.

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.

Figure 1 illustrates a cross-sectional view of a typical prior art NROM memory cell with a channel length, L , that is greater than 100 nm. This cell is comprised of a control gate 100 formed on top of an oxide-nitride-oxide (ONO) layer. This layer is comprised of an oxide layer 101 on top of a nitride 103 layer upon which the charge is stored for the various states of the cell. In one embodiment, the cell has trapping areas 105 and 106 for storing two bits of data on the nitride layer 103. The nitride layer 103 is deposited on another oxide layer 104 that is on the substrate.

Two source/drain regions 109 and 111 are at either end of the gate 100. The source/drain regions 109 and 111 are connected by a channel area 110 between the two source/drain regions 109 and 111. The function of each source/drain region 109 or 111 (i.e., whether source or drain) depends upon which bit area 105 or 106 is being read or written. For example, in a read operation, if the carrier is input at the left side source/drain region 111 and output from the right side region 109, the left side is the source 111 and the right side is the drain 109 and the data bit charge is stored on the nitride 103 at the source end 111 for bit area 106.

As IC manufacturers try to increase memory density of NROM devices, the channel length is reduced. Figure 2 illustrates a typical prior art planar NROM device that has a channel length that is less than 100 nm. In this case, the channel length is so short that the bit trapping areas 205 and 206 overlap. The overlap may cause data write and read errors.

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 smaller multiple-bit NROM device without trapping area overlap.

SUMMARY

The above-mentioned problems with trapping site overlap and other problems are addressed by the present invention and will be understood by reading and studying the following specification.

The present invention encompasses a nitride read only memory (NROM) flash memory transistor. The transistor is comprised of a substrate with first and second source/drain regions. An oxide layer is deposited on the substrate.

A gate insulator layer is coupled to a portion of the oxide layer. The gate insulator layer is comprised of separate first and second sections. The two sections are structurally separated by a middle section of a polysilicon gate structure. Each section is capable of storing a separate charge.

The middle section of the gate structure is separated from outer sections of the gate structure by the gate insulator layer. The top of the gate structure, as well as the portion of the gate insulator deposited on the top of the gate structure, is planarized and a metal contact

coupled to the three sections of the gate structure as well as the ends of each section of the gate insulator.

Further embodiments of the invention include methods and apparatus of varying scope.

5

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a cross-sectional view of a typical prior art NROM cell with a channel greater than 100 nm.

10 Figure 2 shows a cross-sectional view of a typical prior art NROM cell with a channel less than 100 nm.

Figure 3 shows a cross-sectional view of one embodiment of an NROM cell of the present invention.

Figure 4 shows a resulting charge separation and distribution diagram of the present invention in accordance with the embodiment of Figure 3.

15 Figure 5 shows a cross-sectional view of a detail of the charge storage regions in accordance with the embodiment of Figure 3.

Figure 6 shows a cross-sectional view of one embodiment of a step in fabrication of the NROM cell of the present invention.

20 Figure 7 shows a cross-sectional view of one embodiment of a subsequent step of NROM cell fabrication of the present invention.

Figure 8 shows a cross-sectional view of one embodiment of a subsequent step of NROM cell fabrication of the present invention.

Figure 9 shows a cross-sectional view of one embodiment of a subsequent step of NROM cell fabrication of the present invention.

25 Figure 10 shows a cross-sectional view of one embodiment of a subsequent step of NROM cell fabrication of the present invention.

Figure 11 shows a cross-sectional view of one embodiment for programming the NROM cell of the present invention using substrate enhanced hot electron injection.

Figure 12 shows a block diagram of an electronic system 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 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 and equivalents thereof.

Figure 3 illustrates a cross-sectional view of one embodiment of an NROM cell of the present invention. The cell is comprised of two charge storage areas 301 and 302 that are discussed subsequently in greater detail with reference to Figure 5. In this embodiment, a nitride layer is not placed under the center of the transistor channel as in the prior art.

The cell has a polysilicon gate structure 313 – 315 that is comprised of a middle section 315 and two outer sections 313 and 314. A gate insulator is formed on the sides of the middle section of the gate structure 315 such that the insulator separates the middle section 315 from the two gate outer sections 313 and 314. A control gate metal contact 312 is formed over all three sections of the gate structure 313 – 315.

The middle gate section 315 has only an oxide insulator 320 and does not trap injected electrons in the NROM device structure. In one embodiment, the gate insulator is a composite insulator that is comprised of an oxide-nitride-oxide (ONO) configuration where the charge trapping is accomplished in the nitride layer 305 and 306. In one embodiment, the top oxide layer 301 and 302 is part of the oxide fill 303 and 304 respectively.

Alternate embodiments use other gate insulators besides the ONO structure shown. These structures include oxide-nitride-aluminum oxide composite layers, oxide- aluminum oxide-oxide composite layers, oxide, silicon oxycarbide-oxide composite layers as well as other composite layers.

In still other alternate embodiments, the gate insulator could include thicker than normal silicon oxides formed by wet oxidation and not annealed, silicon rich oxides with

inclusions of nanoparticles of silicon, silicon oxynitride layer that are not composite layers, silicon rich aluminum oxide insulators that are not composite layers, silicon oxycarbide insulators that are not composite layers, silicon oxide insulators with inclusions of nanoparticles of silicon carbide, in addition to other non-stoichiometric single layers of gate insulators of two or more commonly used insulator materials such as Si, N, Al, Ti, Ta, Hf, Zr, and La.

The embodiment of Figure 3 is also comprised of two source/drain regions 310 and 311. In the illustrated embodiment, these regions are n+ type semiconductor material while the substrate is a p+ type semiconductor material. In an alternate embodiment, the source/drain regions may use p+ type semiconductor material while the substrate is n+.

The function of each source/drain region 310 or 311 depends on which bit area 301 or 302 is being read or written. For example, in a read operation, if the carrier is input at the left side source/drain region 311 and output from the right side region 310, the left side is the source 311 and the right side is the drain 310 and the data bit charge is stored on the nitride layer 306 at the source end 311 in bit area 302.

Figure 4 illustrates one embodiment of a charge separation and distribution diagram in relation to the Figure 3 embodiment of the NROM cell of the present invention. This diagram indicates the charge storage density in the vertical direction and the distance along the cell in the horizontal direction. The length of the channel between the source/drain areas of Figure 3 is illustrated as L.

The two charges 401 and 402 stored in the NROM cell are indicated on the charge separation and distribution diagram as coinciding with the charge storage areas 301 and 302 of Figure 3. The diagram also illustrates that there is no charge present 405 in the middle of the cell.

Figure 5 illustrates a more detailed cross-sectional view of the charge storage area 302 of the embodiment of Figure 3. This view clearly shows the oxide 304-nitride 306-oxide 320 composite insulator as shown on the left side of the NROM cell of Figure 3. The charge storage area 302 as well as one source/drain region 311 and part of the polysilicon gate structure 313 are also shown.

The above-described embodiments show one section of each side of the gate insulator layer substantially horizontal and a second section of each side substantially vertical

and extending upwards through the gate structure. The present invention, however, is not limited to any one angle between the substantially horizontal section and the substantially vertical section. In other words, the “horizontal” and “vertical” sections may not be horizontal and vertical. There is also no limitation that each side of the gate insulator layer be symmetrical with the other side.

Figure 6 illustrates a cross-sectional view of one embodiment for a step in fabrication of the NROM cell of Figure 3. A thicker gate oxide 601 is grown on the substrate 600. The source/drain regions 604 and 605 are implanted. The polysilicon gate electrode 610 is also defined using conventional techniques that are well known in the art.

The gate oxide 601 is then removed in areas 602 and 603 outside the polysilicon gate area by an etch process to define the polysilicon gate structure 610. The oxide can then be regrown to a new desired thickness.

Figure 7 illustrates the regrown oxide areas 720 and 721 outside the polysilicon gate electrode. This structure is then covered with the composite insulators 701 and 703 such as nitride or other insulator as described previously.

Figure 8 illustrates a cross-sectional view of the NROM cell with a deposited polysilicon layer 801 on top of the composite insulators of Figure 7. This second polysilicon layer 801 is then directionally etched to leave only the sidewalls 901 and 902 as illustrated in Figure 9. This provides a structure with a composite gate insulator 905 under the polysilicon gate and along the sidewalls 901 and 902. A single gate oxide 910 is under the central polysilicon gate region 903.

Figure 10 illustrates the NROM cell with a deposited oxide filler 1001 and 1002. The top of the structure is planarized by chemical mechanical polishing (CMP). This removes the insulators from the top 1005 of the central polysilicon gate. A patterned metal contact that selectively adheres to the polysilicon is deposited on top of the gate structures 1006 – 1008. The electrical gate provides contact to all three gate regions 1006 – 1008.

In one embodiment, the NROM flash memory cell of the present invention operates by conventional tunnel injection with positive gate voltages with respect to the substrate/p-well. In another embodiment, channel hot electron injection (HEI) can be used for programming. Such an embodiment uses conventional positive gate voltages with respect to the substrate/p-well. Tunneling can be used for erase operations.

Using HEI, the NROM device of the present invention provides two bit storage as in prior art NROM devices. Charges are stored near the drain and the device is read in the reverse direction. Either end of the channel can be used as a drain and a charge stored at both ends of the channel near the surface of the n+ regions.

5 Figure 11 illustrates an embodiment for programming the NROM flash memory cell. In this embodiment, a negative substrate bias, V_{SUB} , is applied to the p-type substrate 1100. This bias increases the surface lateral field near a source/drain region 1101 or 1102 (depending on which direction the cell is operated), thus increasing the number of hot electrons. Such a substrate enhanced hot electron (SEHE) injection embodiment would
10 require a lower drain voltage during programming operations. In one embodiment, the negative substrate bias is in the range of 0V to -3V. Alternate embodiments may use other voltage ranges.

As is well known in the art, applying a drain voltage to a first source/drain region 1101 and grounding the second source/drain region 1102 creates a hot electron injection into
15 the gate insulator of the charge storage region 1105 nearest the drain region 1101. The second charge storage region 1106 is programmed by performing the same biasing of the source/drain regions 1101 and 1102 in the reverse direction.

For an erase operation, substrate enhanced band-to-band tunneling induced hot hole injection (SEBBHH) can be used. SEBBHH and SEHE are both well known in the art and are
20 not discussed further.

Figure 12 illustrates a functional block diagram of a memory device 1200 that can incorporate the NAND flash memory cells of the present invention. The memory device 1200 is coupled to a processor 1210. The processor 1210 may be a microprocessor or some other type of controlling circuitry. The memory device 1200 and the processor 1210 form part of an
25 electronic system 1220. The memory device 1200 has been simplified to focus on features of the memory that are helpful in understanding the present invention.

The memory device includes an array of NROM flash memory cells 1230. In one embodiment, the memory cells are NROM flash memory cells and the memory array 1230 is arranged in banks of rows and columns. The control gates of each row of memory cells is
30 coupled with a wordline while the drain and source connections of the memory cells are

coupled to bitlines. As is well known in the art, the connection of the cells to the bitlines depends on whether the array is a NAND architecture or a NOR architecture.

An address buffer circuit 1240 is provided to latch address signals provided on address input connections A0-Ax 1242. Address signals are received and decoded by a row decoder 1244 and a column decoder 1246 to access the memory array 1230. It will be appreciated by those skilled in the art, with the benefit of the present description, that the number of address input connections depends on the density and architecture of the memory array 1230. That is, the number of addresses increases with both increased memory cell counts and increased bank and block counts.

The memory device 1200 reads data in the memory array 1230 by sensing voltage or current changes in the memory array columns using sense/buffer circuitry 1250. The sense/buffer circuitry, in one embodiment, is coupled to read and latch a row of data from the memory array 1230. Data input and output buffer circuitry 1260 is included for bi-directional data communication over a plurality of data connections 1262 with the controller 1210).

Write circuitry 1255 is provided to write data to the memory array.

Control circuitry 1270 decodes signals provided on control connections 1272 from the processor 1210. These signals are used to control the operations on the memory array 1230, including data read, data write, and erase operations. The control circuitry 1270 may be a state machine, a sequencer, or some other type of controller.

Since the NROM memory cells of the present invention use a CMOS compatible process, the memory device 1200 of Figure 12 may be an embedded device with a CMOS processor.

The flash memory device illustrated in Figure 12 has been simplified to facilitate a basic understanding of the features of the memory. A more detailed understanding of internal circuitry and functions of flash memories are known to those skilled in the art.

CONCLUSION

In summary, the NROM flash memory transistor of the present invention provides a self-aligned, structural charge separation that allows smaller cells to be manufactured without bit area overlap. This cell provides a low initial threshold voltage, fast operation, and low power consumption in addition to the high memory density. The NROM cell can be used in NOR-type memory arrays, NAND-type memory arrays, or other memory array architectures.

Although specific embodiments have been illustrated and described herein, it will be appreciated by those of ordinary skill in the art that any arrangement that is calculated to achieve the same purpose may be substituted for the specific embodiments shown. Many adaptations of the invention will be apparent to those of ordinary skill in the art. Accordingly, 5 this application is intended to cover any adaptations or variations of the invention. It is manifestly intended that this invention be limited only by the following claims and equivalents thereof.

What is claimed is:

1. An NROM flash memory transistor comprising:
a substrate comprising first and second source/drain regions;
an oxide layer on the substrate;
a gate insulator layer coupled to a portion of the oxide layer, the gate insulator layer comprising a first section and a separate second section; and
a gate structure comprising a plurality of sections, a middle section coupled to the oxide layer and first and second outer sections each coupled to the gate insulator layer such that the gate insulator layer separates the middle section from the first and second outer sections.
2. The transistor of claim 1 wherein the gate insulator layer is comprised of a composite oxide-nitride-oxide layer.
3. The transistor of claim 1 wherein the gate insulator layer is a composite layer comprised of one of an oxide-nitride-aluminum oxide composite layer, an oxide-aluminum oxide-oxide composite layer, or an oxide-silicon oxycarbide-oxide composite layer.
4. The transistor of claim 1 wherein the gate insulator layer is a non-composite layer comprised of one of silicon oxides formed by wet oxidation and not annealed, silicon-rich oxides with inclusions of nanoparticles of silicon, silicon oxynitride layers, silicon-rich aluminum oxide insulators, silicon oxycarbide insulators, or silicon oxide insulators with inclusions of nanoparticles of silicon carbide.
5. The transistor of claim 1 wherein the gate insulator is comprised of non-stoichiometric single layers of two or more of silicon, nitrogen, aluminum, titanium, tantalum, hafnium, lanthanum, or zirconium.
6. The transistor of claim 1 wherein a first charge is stored on the first section of the gate insulator layer and a second charge is stored on the second section of the gate insulator layer.

7. The transistor of claim 1 and further comprising an oxide fill layer coupled to at least a portion of the first and second gate insulator sections and the first and second outer sections of the gate structure.
8. The transistor of claim 1 and further including a metal contact coupled to the plurality of sections of the gate structure.
9. The transistor of claim 1 wherein the substrate is a p+ material and the first and second source/drain regions are n+ material.
10. An NROM flash memory transistor comprising:
a substrate comprising first and second source/drain regions located laterally such that a channel area is formed between them;
an oxide layer on the substrate above the first and second source/drain regions and the channel area;
a gate insulator layer comprising a first section and a separate second section, a substantially horizontal portion of the first and second sections coupled to separate areas of the oxide layer; and
a gate structure comprising a middle section and first and second outer sections, the middle section coupled to the oxide layer and separating the first and second gate insulator sections, the first and second outer sections each separated from the middle section by portions of the gate insulator.
11. The transistor of claim 10 wherein the substrate is coupled to a negative bias that enhances hot electron injection.
12. The transistor of claim 10 wherein the gate structure is comprised of a polysilicon material.
13. The transistor of claim 10 and further including an oxide material coupled to the first and second outer sections of the gate structure and portions of the gate insulator layer not within the gate structure.

14. An NROM flash memory transistor comprising:
 - a substrate comprising first and second source/drain regions located laterally such that a channel area is formed between them;
 - an oxide layer on the substrate above the first and second source/drain regions and the channel area;
 - a composite gate insulator layer comprising a first section and a separate second section, a substantially horizontal portion of each section coupled to separate areas of the oxide layer and a substantially vertical portion of each section extending from the oxide layer, the composite gate insulator layer having a nitride layer for storing a charge on each of the first and second sections;
 - a polysilicon gate structure comprising a middle section and first and second outer sections, the middle section coupled to the oxide layer and separated from the first and second outer sections by the substantially vertical portions of the gate insulator layer; and
 - a gate metal contact coupled to the outer and middle sections of the gate structure and each of the substantially vertical portions of the gate insulator.
15. The transistor of claim 14 and further including an oxide material deposited on the composite gate insulator and the outer sections of the gate structure.
16. The transistor of claim 14 wherein the transistor is operated with either the first source/drain region or the second source/drain region serving as the source region in response to a direction of operation of the transistor.
17. A method for fabricating an NROM flash memory cell, the method comprising:
 - doping a substrate to form first and second source/drain regions positioned laterally on the substrate and separated by a channel area;
 - depositing an oxide layer over the substrate including the first and second source/drain regions and the channel area;
 - forming a polysilicon middle gate region on the oxide layer above the channel area;
 - depositing a gate insulator layer over the oxide layer;
 - depositing a polysilicon layer over the gate insulator layer;

etching the polysilicon layer such that two outer gate regions are left from the polysilicon layer thus forming a gate structure with a middle gate region and two outer gate regions separated from the middle by the gate insulator layer; planarizing a top of the gate structure such that the gate insulator layer is removed from the top of the gate structure; and forming a contact over the gate structure that is coupled to each region of the gate structure and remaining end portions of the gate insulator layer.

18. The method of claim 17 and further including etching the oxide layer, prior to depositing the gate insulator, such that the channel area silicon on either side of the middle gate region is substantially exposed.
19. The method of claim 17 wherein planarizing comprises using chemical mechanical polishing.
20. An NROM flash memory array comprising:
 - a plurality of NROM flash memory cells arranged in rows and columns, each cell comprising:
 - a substrate comprising first and second source/drain regions;
 - an oxide layer on the substrate;
 - a gate insulator layer coupled to a portion of the oxide layer, the gate insulator layer comprising a first section and a separate second section; and
 - a gate structure comprising a plurality of sections, a middle section coupled to the oxide layer and first and second outer sections each coupled to the gate insulator layer such that the gate insulator layer separates the middle section from the first and second outer sections;
 - a plurality of wordlines, each wordline coupling the gate structures of the rows of cells; and
 - a plurality of bit lines that couple the columns of the cells.
21. The memory array of claim 20 wherein the plurality of NROM flash memory cells are arranged in a NAND flash memory architecture.

22. The memory array of claim 20 wherein the plurality of NROM flash memory cells are arranged in a NOR flash memory architecture.
23. An electronic system comprising:
a processor that generates control signals for the system; and
an NROM flash memory array coupled to the processor that operates in response to the control signals, the array comprising:
a plurality of NROM flash memory cells arranged in rows and columns, each cell comprising:
a substrate comprising first and second source/drain regions;
an oxide layer on the substrate;
a gate insulator layer coupled to a portion of the oxide layer, the gate insulator layer comprising a first section and a separate second section; and
a gate structure comprising a plurality of sections, a middle section coupled to the oxide layer and first and second outer sections each coupled to the gate insulator layer such that the gate insulator layer separates the middle section from the first and second outer sections;
a plurality of wordlines, each wordline coupling the gate structures of the rows of cells; and
a plurality of bit lines that couple the columns of the cells.
24. A method for programming an NROM flash memory cell having a substrate, two structurally separated charge trapping areas, and two source/drain regions, the method comprising:
biasing the two source/drain regions in response to the charge trapping area to be programmed;
applying a gate voltage to a gate of the memory cell; and
applying a negative voltage to the substrate of the memory cell such that biasing the source/drain regions and applying the gate and negative substrate voltages to the memory cell creates substrate enhanced hot electron injection into the gate insulator to program a charge onto a first of the structurally separated charge trapping areas in the gate insulator.

25. The method of claim 24 wherein the first structurally separated charge trapping area is programmed when the source/drain region nearest the first charge trapping area is biased with a larger voltage than the remaining source/drain region.