A THREE-DIMENSIONAL CONTACTLESS NON-VOLATILE MEMORY CELL AND PROCESS FOR FABRICATING THE SAME

FIELD OF THE INVENTION

The invention relates to the field of semiconductor memory devices employing floating gates and the processes and methods for fabricating these devices.

BACKGROUND OF THE INVENTION

Memory cells which have members which may be electrically charged are well-known in the prior art. Most often, these cells employ polysilicon floating gates which are completely surrounded by insulation (e.g., silicon dioxide). These gates are electrically charged by electron transfer from the substrate through a variety of mechanisms such as avalanche injection, channel injection, tunneling, etc. The presence or absence of this charge represents a stored, binary information. An early example of such a device is shown in U. S. Patent No. 3,500,142.

One category of semiconductor floating gate memory devices are both electrically programmable and electrically erasable (EEPROMs). Such a device is shown in U. S. Patent No. 4,203,158. These cells are characterized by a substrate region including source and drain regions which define a channel therebetween. Disposed above this channel is a floating gate separated from the substrate region by a relatively thin gate insulative material. Typically, a layer of silicon dioxide is employed as the insulator. A control gate disposed above, and insulated from, the floating gate. Normally, the control gate is also fabricated of polysilicon.

A more recent category of floating gate memory devices uses channel injection for charging the floating gate and tunnelling for removing charge. For these devices the memory cell comprises only a single device and the entire memory array is erased at one time. That is, individual cells or groups of cells are not separately erasable as in current EEPROMs. These memories are sometimes referred to as flash EPROMs or flash EEPROMs.

An example of these devices is disclosed in pending US application serial number 892,446, filed August 4, 1986, entitled "Low Voltage EEPROM Cell", which application is assigned to the assignee of the present application. U. S. Patent No. 4,698,787 of Mukherjee et al., also discloses an electrically erasable programmable memory device which is programmed by hot-electron injection from the channel onto the floating gate, and erased by these Fowler-Nordheim tunnelling from the floating gate to the substrate.

The principle upon which these EEPROM cells operate is that electrons (i.e., charge) are stored on the floating gate in a capacitive manner. By way of example, during programming of an EEPROM device, the control gate is usually taken to a high positive potential ranging between 12 and 20 volts. The source is grounded and the drain is taken to an intermediate potential of approximately 7 volts. This creates a high lateral electrical field within the channel region nearest to the drain. The high lateral electric field accelerates electrons along the channel region to the point where they become "hot". These hot-electrons create additional electron-hole pairs through impact ionization. A large number of these electrons are attracted to the floating gate by the large positive potential on the control gate.

During erasing of an EPROM device, the control gate is usually grounded and the drain is left unconnected. The source is taken to a high positive potential, creating a high vertical electric field from the source to the control gate. Charge is erased from the floating gate by the mechanism of Fowler-Nordheim tunnelling of electrons through the gate oxide region between the source and the floating gate in the presence of such a field.

In certain instances, floating gate memory devices are fabricated in arrays where each device or device pair is separated from other devices by thick field oxide regions. An example of this is shown in U. S. Patent No. 4,114,255. In these arrays a metal contact is ordinarily required for each device or device pair. Obviously, these metal contacts limit the reduction of device area.

Other architectures substantially reduce the number of metal contacts required by employing elongated source/drain regions disposed beneath oxide regions. These arrays sometimes are referred to as having "buried bit lines" or using "contactless cells" and need virtual ground circuitry for sensing and programming. An example of this type of array is shown in U. S. Patent No. 4,267,632; virtual ground circuitry is shown in U. S. Patent No. 4,460,981. A process for fabricating an electrically alterable contactless memory array is described in U. S. Patent No. 4,780,424 of Holler et al., which is assigned to the assignee of the present invention.

Although the contactless memory array architecture provides an increase in device density, it is not without certain drawbacks. For instance, prior art floating gate memory devices generally require a relatively thick oxide for capacitive coupling concerns. These thick oxide regions contribute significantly to the overall cell size. Perhaps more importantly, prior art floating gate memory devices are characterized by a floating gate having a large horizontal dimension. Often, as is the case in U. S. Patent No. 4,780,424, the floating gate extends beyond the channel regions to overlap the isolation regions of the device. This increased lateral dimension is necessary to increase the capacitive coupling between the floating gate and the control gate. This large lateral floating gate dimension further increases the minimal cell dimension geometry for this technology.

As will be seen, the present invention provides an electrically erasable and programmable memory cell which utilizes both the vertical and the planar dimension of the floating gate to achieve a minimal area memory cell. When fabricated in the contactless array, the present invention yields the absolute smallest cell size with any given design rule.

SUMMARY OF THE INVENTION

A novel three-dimensional floating gate memory cell is described along with the process for fabricating the same within an array. In one embodiment, a gate oxide is first formed on a silicon substrate. Over the gate oxide is deposited a layer of polysilicon. This polysilicon layer is etched to form elongated, parallel, spaced-apart polysilicon strips. These strips are used to define the active areas for the memory cells. The polysilicon strips will later comprise the floating gates of the devices. Next, dopants are implanted into the substrate to form source and drain regions in alignment with the floating gate members. By extending the full length of a column, the source/drain regions function as buried bit lines for the array.

Immediately after the removal of the photoresist and after deposition of an interpoly dielectric, a second layer of polysilicon is deposited over the substrate. The second poly layer is formed above, and is insulated from, the first polysilicon layer. This poly2/dielectric/poly1 stack is then etched to define the floating gates and control gates for each of the cells. The control gate is formed by a continuous wordline which extends across a row of the array. Each wordline is patterned so as to be generally perpendicular to the parallel source and drain regions.

The crux of the present invention is the formation of the first polysilicon layer to a sufficient thickness so as to allow capacitive coupling to the control gate member along the vertical (i.e., sidewall) regions of the floating gate member. In this manner, the vertical dimension of the floating gate contributes to the total capacitive coupling area between the floating and control gates. This permits a substantial increase in device density within the array. Currently, a 50% reduction in total array area may be realized.

Further features of the present invention include an improved gate oxide quality due to elimination of thick oxide regions. The gate oxide is also immediately capped by a protective polysilicon layer following its formation to prevent exposure to damaging effects of subsequent

processing. The present invention also provides continuous source/drain regions in the form of buried bit lines to eliminate the traditional requirement for a single contact per two cells. Additionally, the present invention saves one critical masking step by utilizing a single masking step to define both the floating gate and the active area for the device.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a cross-sectional elevation view of a substrate which includes a gate oxide over which has been deposited a polysilicon layer and an additional oxide layer.

Figure 2 illustrates the substrate of Figure 1 after photoresist masking members have been formed and the underlying oxide, polysilicon layers have been etched. This view also illustrates the ion implantation of arsenic into the regions between the masked members.

Figure 3 illustrates the substrate of Figure 2 during another implantation step.

Figure 4 illustrates the substrate of Figure 3 following deposition of an interpoly dielectric layer followed by a second polysilicon layer.

Figure 5 is a perspective view of the cross-sectional elevation view of Figure 4 following the etching of the second polysilicon layer. Figure 5 shows the location of buried bit lines and wordline.

Figure 6 shows a cross-sectional view of an alternative method of fabrication. Following ion implantation steps, reoxidation and planarization steps are performed.

Figure 7 illustrates the substrate of Figure 6 following etchback of the upper insulation layers.

Figure 8 illustrates the substrate of Figure 7 after deposition of an interpoly dielectric layer followed by a second polysilicon layer.

DESCRIPTION OF THE PREFERRED EMBODIMENT(S)

A novel three-dimensional non-volatile memory cell is disclosed which utilizes both the vertical and the planar gate dimensions to achieve a minimal area. In the following description, numerous specific details are set forth such as specific doping levels, dimensions, etc., in order to provide a thorough understanding of the present invention. It will be obvious, however, to one skilled in the art that these specific details need not be employed to practice the present invention. In other instances, well-known processing steps are not described in detail in order not to unnecessarily obscure the present invention.

The memory cells of the present invention are fabricated using standard metal-oxide-semiconductor (MOS) processing. The array which contains the cells, in the currently preferred embodiment is fabricated of n-channel devices. A peripheral circuit can employ either n-channel devices or complimentary MOS devices.

With reference to Figure 1, initially the entire substrate is subjected to a thermal oxidation cycle to thermally grow a gate oxide 11 over p-type monocrystalline silicon substrate 10. Gate oxide 11 is a high-grade thermal oxide which in the currently preferred embodiment assumes a thickness of approximately 110Å. Note that this thickness is typical for flash applications but may vary for other types of device structures or applications.

By growing the gate oxide (also known as the tunnel oxide) at the beginning of the process, erasing operations are improved as compared to prior art approaches. In the prior art cells the tunnel oxide is usually grown after the formation of field oxidation regions or after the source/drain implants have been performed. Both of these processing steps tend to degrade the quality of the gate oxide. Flash erasure operation is improved by growing the gate oxide early on in the process flow since an exceedingly flat silicon surface is still available. In other words, the substrate surface has not yet been subjected to several implantation or oxidation steps. For example, a high temperature thermal oxidation cycle for growing field oxide

regions normally produces a great deal of stress in the neighboring gate oxide regions. Field oxide growth also destroys the planarity of the substrate surface at the ends of the active regions.

Optionally, a voltage threshold adjusting implant may be performed before growth of the gate oxide 11. This threshold adjusting implant may consist of a combination of low energy and high energy boron implants.

After the gate oxide is formed, a layer of polysilicon 12 is then deposited over the substrate. In the currently preferred embodiment, the thickness of polysilicon layer 12 is on the order of 3500Å. However, thicker polysilicon layers (up to 7500Å and higher) may also be employed depending on specific design considerations. In other words, the height of the first polysilicon layer 12 may be increased to provide greater capacitive coupling to the control gate member up to the step coverage limit of the particular process, as will be discussed in more detail later.

Next, a layer of oxide 13 is formed over first polysilicon layer 12 either by oxidation or chemical vapor deposition (CVD) techniques. The thickness of layer 13 is generally about 400Å. The purpose of layer 13 is to prevent polysilicon layer 12 from being doped during subsequent implantation steps. For those applications in which subsequent doping of the polysilicon layer 12 is immaterial (such as is typically the case for EPROM devices) this oxidation step may be omitted from the process flow without departing from the scope of the present invention.

Referring now to Figure 2, a single photoresist masking layer 15 is used to define the active channel area for the devices. The portions of oxide layer 13 and polysilicon layer 12 which are not covered by photoresist layer 15 are removed during an etching step. The polysilicon regions underlying masking members 15 are protected during the etching step, with the result being that elongated, parallel, spaced-apart poly 1 strips are formed.

Afterwards substrate 10 is subjected to an arsenic implant as indicated by arrows 16. This implant penetrates through tunnel oxide

layer 11 into p-type substrate 10, but not into the polysilicon strips or the underlying substrate regions protected by resist members 15. In the presently preferred embodiment, arsenic is implanted to a level of $1 \times 10^{15}/cm^2$ to $5 \times 10^{15}/cm^2$. This implantation step forms elongated, parallel, spaced-apart doped regions in the substrate. Note that not only do masking members 15 serve as the active channel definition mask, but also as the polysilicon gate mask. This approach saves one critical mask in the process flow and also eliminates registration variations between masking steps.

Next, as illustrated in Figure 3, alternate ones of the elongated regions between the polysilicon members 12 are covered with a photoresist member 20. The regions directly under photoresist member 20 are thereby protected from exposure during the subsequent ion implant step. The regions 18 are then subjected to a phosphorous implant as shown by arrows 21 in Figure 3. In the currently preferred embodiment, regions 18 are implanted with phosphorous to a level of approximately $0.5 \times 10^{15}/cm^2$ to $1 \times 10^{15}/cm^2$. Note that this processing step is only required for flash devices and is unnecessary when fabricating ordinary EEPROMs.

Following the phosphorous implantation step, three separate and distinct regions exist within p-type substrate 10. Those regions include n-type doped regions 18 and 19 (region 18 receiving the additional phosphorous implant) and channel regions 17 formed directly beneath first polysilicon members 12. Note that during the phosphorous implant polysilicon members 12 are protected by oxide layer 13. (It should be appreciated that the additional phosphorous implant 21 is generally only required when fabricating flash-type devices, and may be eliminated for ordinary EPROMs or EEPROMs).

After the phosphorous implant has been performed, photoresist member 20 is removed and oxide layer 13 is etched back along with the exposed regions of gate oxide 11. The gate or tunnel oxide 11 directly under polysilicon members 12 is preserved.

A layer of silicon dioxide, or preferably interpoly dielectric layer 26

comprising an oxide/nitride/oxide (ONO) stack is then deposited over first polysilicon members 12 as shown in Figure 4. The thickness of interpoly dielectric 26 is about 200Å in the currently preferred embodiment. A second polysilicon layer 30 is then deposited over dielectric layer 26. The thickness of layer 30 is approximately 3000Å in the presently preferred embodiment. This second polysilicon (poly 2) layer is masked and layers 30, 26 and 12 are etched to define the control gate and floating gate members for the array. The poly 2 layer is patterned into wordlines which completely cover the floating gates as they extend across a row of the array.

Note that in Figure 4, the elongated parallel, spaced-apart implanted substrate regions are now shown as source and drain regions 25 and 24, respectively.

A key feature of the present invention is the vertical dimension of the polysilicon floating gate members 12. This vertical dimension, which is typically between 3000-7500Å thick, increases the capacitive coupling between the floating gate and the control gate. By increasing the vertical height of the floating gate the coupling is enhanced without a corresponding increase in the lateral dimension of the device. Thus, overall cell density is greatly improved. An additional feature characteristic of the cell of Figure 4 is the absence of field oxide or other thick oxide (e.g., SATO) regions separating each of the devices within the array. In the prior art, thick oxide regions are required to minimize first polysilicon (poly 1) to source/drain capacitance. However, in obviating the large lateral first polysilicon dimension, the corresponding need for thick oxide isolation regions is likewise eliminated. Of course, isolation in the form of field oxide may still be required in the periphery cells, even though it is unnecessary in the central array. Also note that within the array, adjacent cells share common bit lines. For example, in Figure 4 adjacent memory cells share a common n+ drain 24.

Capacitive coupling between the floating gate and control gate members may be increased in the process simply by increasing the vertical

dimension of first polysilicon member 12. As explained above, the increased coupling capacitance is achieved without a corresponding increase in the lateral dimension of the device. In the currently preferred embodiment, channel widths for individual memory cells are on the order of 1.0 microns in width. Furthermore, the elimination of field oxide regions from the memory of the present invention markedly reduces stress in the tunnel oxide regions; thereby providing superior programming and erase performance.

Figure 5 provides a perspective view of the cross-sectional elevation of Figure 4. In particular, Figure 5 more clearly illustrates the location of poly 2 wordline 30 in relation to the underlying source/drain buried bit lines. Wordline 30 is patterned generally perpendicular to the underlying buried bit lines and extends completely across a row of the array. Poly wordlines 30 are defined using an ordinary poly 2/ONO/poly 1 etch cycle. Planarization, passivation, contact formation and metal line formation steps are then performed in an ordinary manner to complete the memory array.

Figures 6-8 show cross-sectional views of an alternative method of fabrication. In Figure 6, a high temperature reoxidation follows removal of photoresist member 20 to form oxide layer 32. Oxide 32 protects polysilicon members 12 from later processing steps and increases the oxide thickness over source/drain regions 25 and 24, respectively. A BPSG or TEOS oxide 33 is then deposited to smooth the topology and better planarize the upper surface of the wafer. A typical thickness for oxide 33 is 1 micron.

An etchback process is then performed so that the oxide regions are recessed from the upper portions of polysilicon members 12 as shown in Figure 7. As is evident, the etchback process leaves a portion of oxide 32 over the source/drain regions of the device while exposing the vertical extent of polysilicon member 12.

Figure 8 shows the completed device which further includes interpoly dielectric layer 35 and a second polysilicon layer 38. Polysilicon layer 38

is etched along with layers 35 and 12 to define wordline segments for the devices in accordance with the memory array of Figure 4-5.

Thus, a three-dimensional contactless non-volatile memory device has been described.

* C L M S *

CLAIMS

1. A process for fabricating an array of floating gate memory devices on a substrate comprising the steps of:

forming a gate oxide on said substrate;

forming a plurality of first gate members on said gate oxide from a first layer of polysilicon;

introducing dopants into said substrate to form elongated, parallel, spaced-apart source and drain regions in alignment with said first gate members said spaced-apart source and drain regions forming a channel therebetween; and

forming elongated second gate members from a second layer of polysilicon, said second gate members being formed over and insulated from said first gate members, said first gate members, having a vertical thickness which is at least thirty percent as large as the width of said channel so as to allow enhanced capacitive coupling to said second gate members along the side wall regions of said first gate members, said side wall regions being disposed over said source/drain regions.

2. The process according to claim 1 wherein said second gate members are generally formed perpendicular to said parallel spaced-apart source and drain regions.

3. The process according to claim 2 wherein said first gate members have a thickness of about 3000Å to 7500Å.

4. The process according to claim 3 wherein said gate oxide is approximately 110Å thick.

5. The process according to claim 4 further comprising the step of performing a threshold adjustment implant prior to the formation of said gate oxide.

6. A process for fabricating an array of floating gate memory devices on a substrate comprising the steps of:

forming a gate oxide on said substrate;

depositing a first layer of polysilicon over said gate oxide;

patterning said first polysilicon layer to form a plurality of first gate members;

introducing dopants into said substrate to form elongated, parallel, spaced-apart source and drain regions aligned to said first gate members, said spaced-apart source/drain regions forming a channel therebetween;

forming an interpoly dielectric over said first gate members;

forming elongated second gate members from a second layer of polysilicon, said second gate members being formed over said first gate members and said second gate members being insulated from said first gate members by said interpoly dielectric, said first gate members having a vertical dimension which is at least thirty percent as large as the width of said channel such that sidewall regions of said first gate members contribute to the total capacitive area between said first and second gate members, said sidewall regions being disposed over said source/drain regions.

7. The process according to claim 6 wherein said second gate members are generally formed perpendicular to said parallel spaced-apart source and drain regions.

8. The process according to claim 6 wherein said first gate members have a thickness of about 3000Å to 7500Å.

9. The process according to claim 8 wherein said gate oxide is approximately 110Å thick.

10. The process according to claim 9 further comprising the step of performing a threshold adjustment implant prior to the formation of said gate oxide.

11. A process for fabricating a floating gate memory device on a silicon substrate comprising the steps of:

growing a gate oxide over said substrate;

forming a floating gate member from a first layer of polysilicon over said gate oxide;

implanting n-type dopants into said substrate in alignment with said floating gate member to form source and drain regions separated by a channel region;

forming an interpoly dielectric over said floating gate member;

forming a control gate member from a second layer of polysilicon deposited over said interpoly dielectric, said first layer of polysilicon having a vertical dimension which is at least thirty percent as large as the width of said channel such that the sidewall region of said floating gate member contribute to the total capacitive area between said floating gate and said control gate members, said sidewall regions being disposed over said source/drain regions.

12. The process according to claim 11 further comprising the following steps after said implanting step:

masking said drain regions; and

implanting said source regions with an n-type

dopant so that said source regions are more heavily doped than said drain regions.

13. The process according to claim 12 wherein said floating gate member has a thickness approximately 7500Å.

14. The process according to claim 13 wherein said gate oxide is approximately 110Å thick.

15. A three dimensional non-volatile memory cell comprising:

a substrate of a first conductivity type;

spaced-apart regions of a second conductivity type disposed in said substrate regions defining a channel therebetween;

a first gate member disposed above and insulated from said channel, said first gate member storing a charge affecting the conductivity within said channel;

a second gate member disposed above and insulated from said first gate member, said first gate member having a vertical dimension which is at least thirty percent as large as the width of said channel such that the sidewall regions of said first gate member contribute to the total capacitive area between said first and said second gate members, said sidewall regions being disposed over said source/drain regions, said second gate member controlling the programming and erasing of said cell.

16. The memory cell of claim 15 wherein said cell is programmed by hot electron injection of electrons from said substrate to said first gate member and erased by Fowler-Nordheim tunnelling of electrons from said first gate member to said substrate.

17. The memory cell of claim 15 wherein said first gate member comprises polysilicon.

* C L M S *

18. The memory cell of claim 17 wherein said vertical dimension is between 3000 and 7500Å thick and said channel is approximately 1 micron wide.

19. The memory cell of claim 18 wherein said second conductivity type is n-type.

20. A process for fabricating an array of floating gate memory devices on a substrate substantially as hereinbefore described with reference to the accompanying drawings.

21. A three dimensional non-volatile memory cell as claimed in claims 15 to 19 substantially as hereinbefore described and illustrated.