(19) **United States**
(12) **Patent Application Publication**
Chang et al.
(10) **Pub. No.: US 2007/0155087 A1**
(43) **Pub. Date: Jul. 5, 2007**

(54) **METHOD OF MANUFACTURING SPLIT GATE FLASH MEMORY**

(75) Inventors: **Ko-Hsing Chang**, Hsinchu (TW); **Wu-Tsung Chung**, Miaoli County (TW); **Tsung-Cheng Huang**, Hsinchu City (TW)

Correspondence Address:
JIANQ CHYUN INTELLECTUAL PROPERTY OFFICE
7 FLOOR-1, NO. 100
ROOSEVELT ROAD, SECTION 2
TAIPEI 100 (TW)

(73) Assignee: **POWERCHIP SEMICONDUCTOR CORP.**, Hsinchu (TW)

(21) Appl. No.: **11/683,439**

(22) Filed: **Mar. 8, 2007**

Related U.S. Application Data

(62) Division of application No. 11/163,223, filed on Oct. 11, 2005, now Pat. No. 7,208,796.

(30) **Foreign Application Priority Data**

Mar. 21, 2005 (TW) 94108557

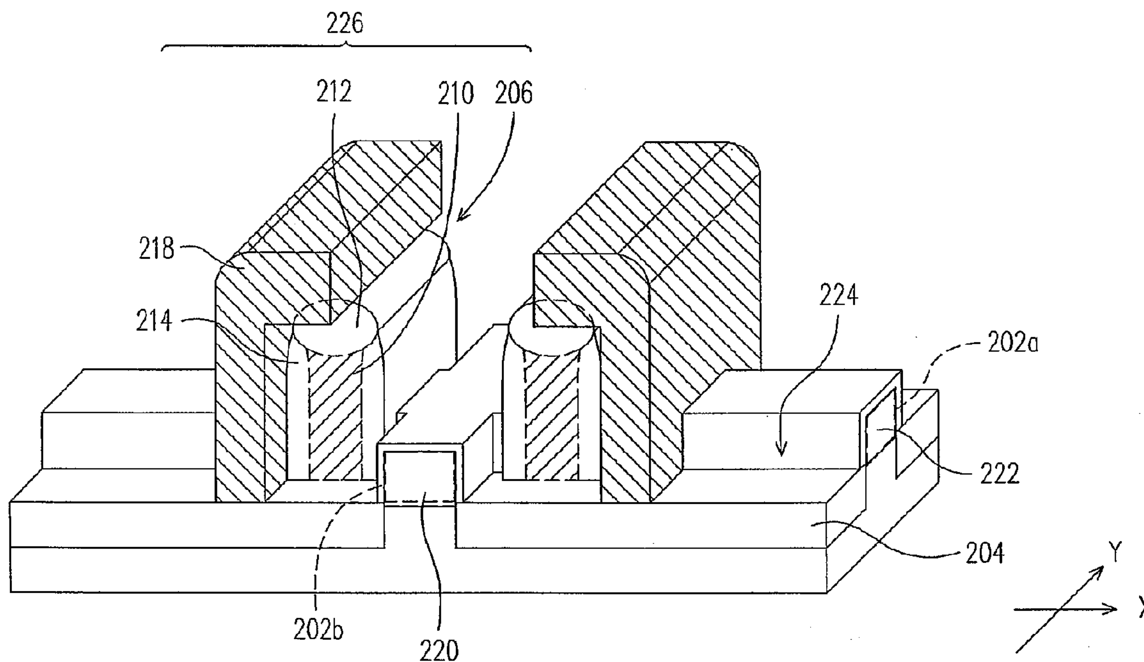
Publication Classification

(51) **Int. Cl.**
H01L 21/8238 (2006.01)

(52) **U.S. Cl.** **438/201**; 257/E21

(57) **ABSTRACT**

A split gate flash memory is provided. Trenches are formed in the substrate to define active layers. The device isolation layers are formed in the trenches. The surface of the device isolation layers is lower than the surface of the active layers. The stacked gate structures each including a tunneling dielectric layer, a floating gate and a cap layer are formed on the active layers. The inter-gate dielectric layers are formed on the sidewalls of the stacked gate structures. The select gates are formed on one side of the stacked gate structure and across the active layer. The select gate dielectric layers are formed between the select gates and the active layers. The source regions are formed in the active layers on the other side of the stacked gate structures. The drain regions are formed in the active layers on one side of the select gates.



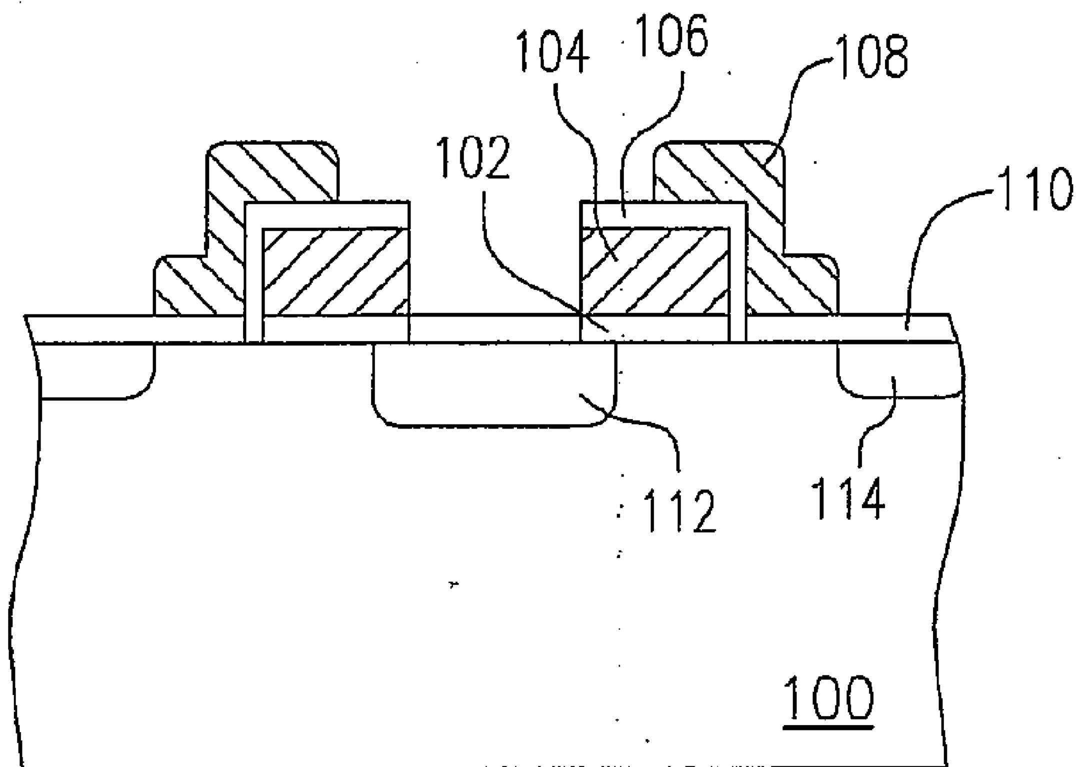


FIG. 1 (PRIOR ART)

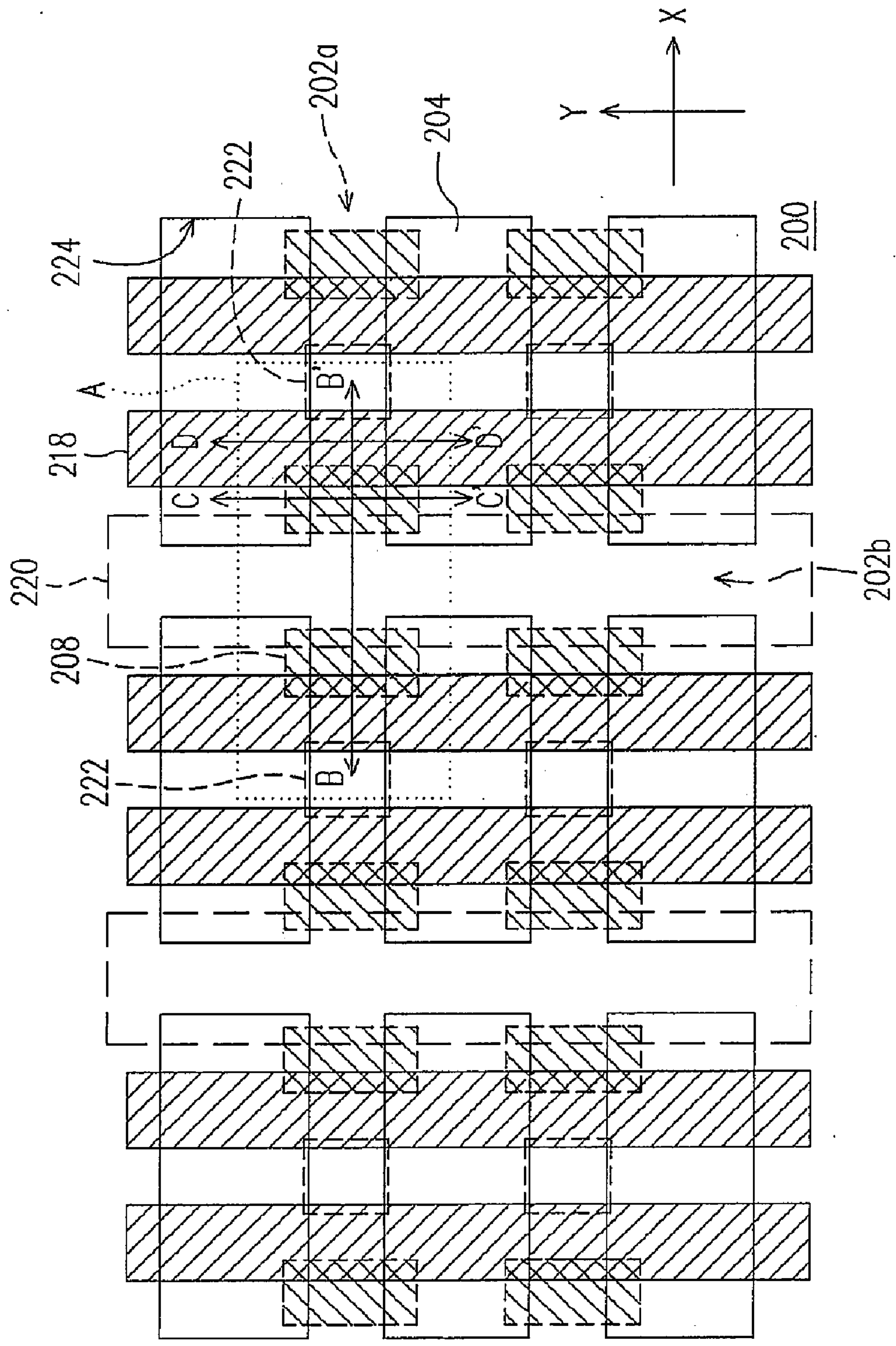


FIG. 2A

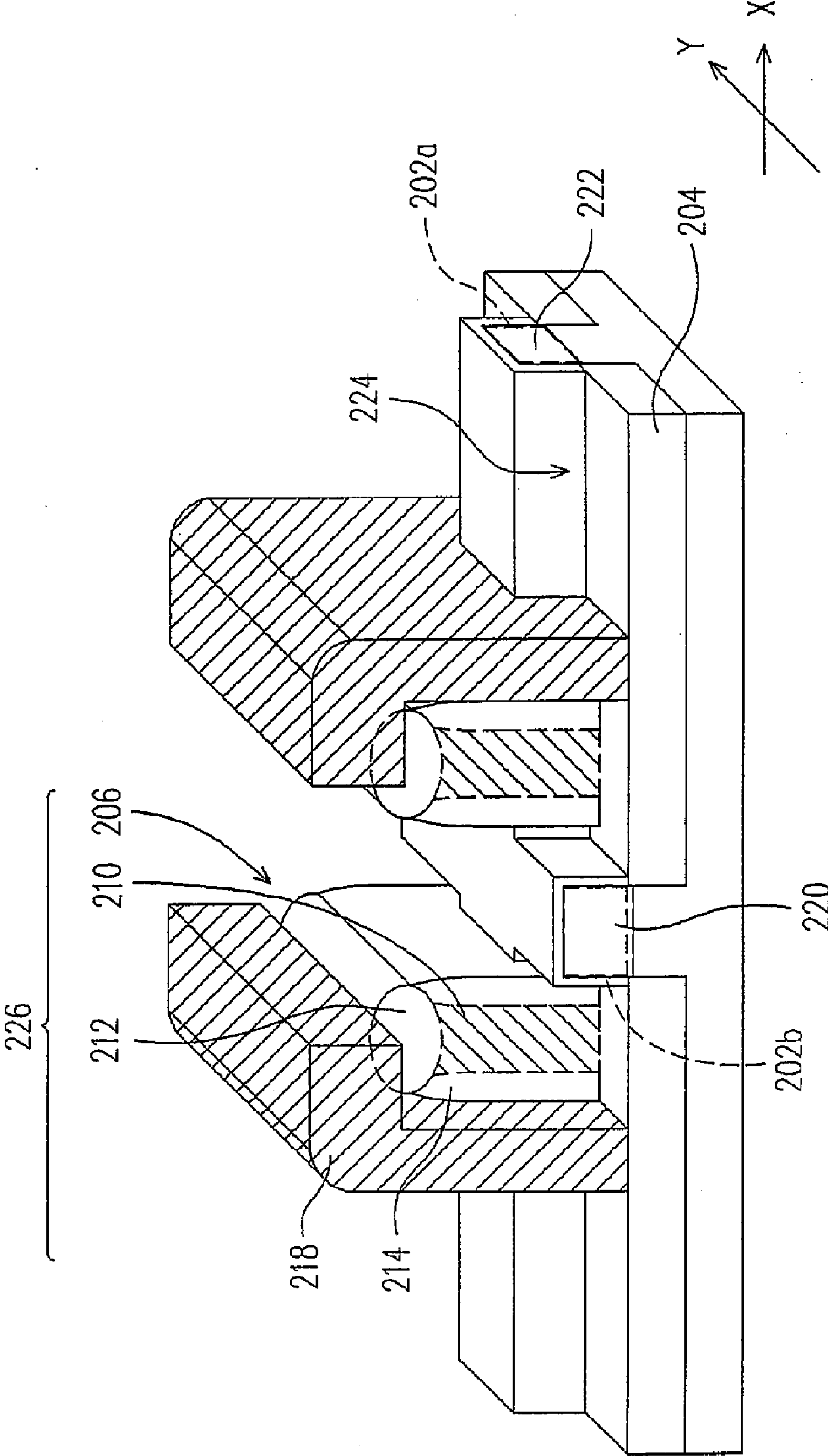


FIG. 2B

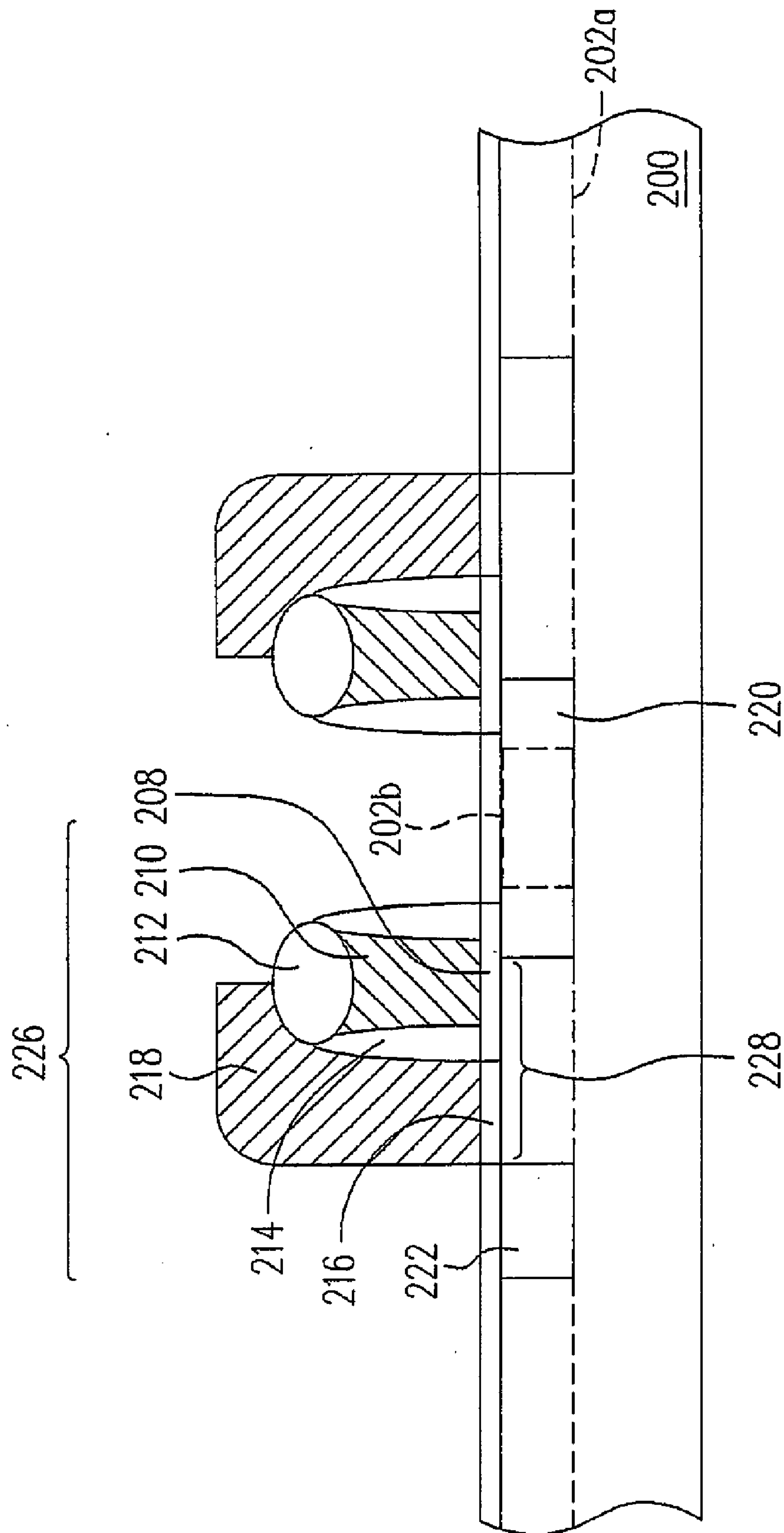


FIG. 2C

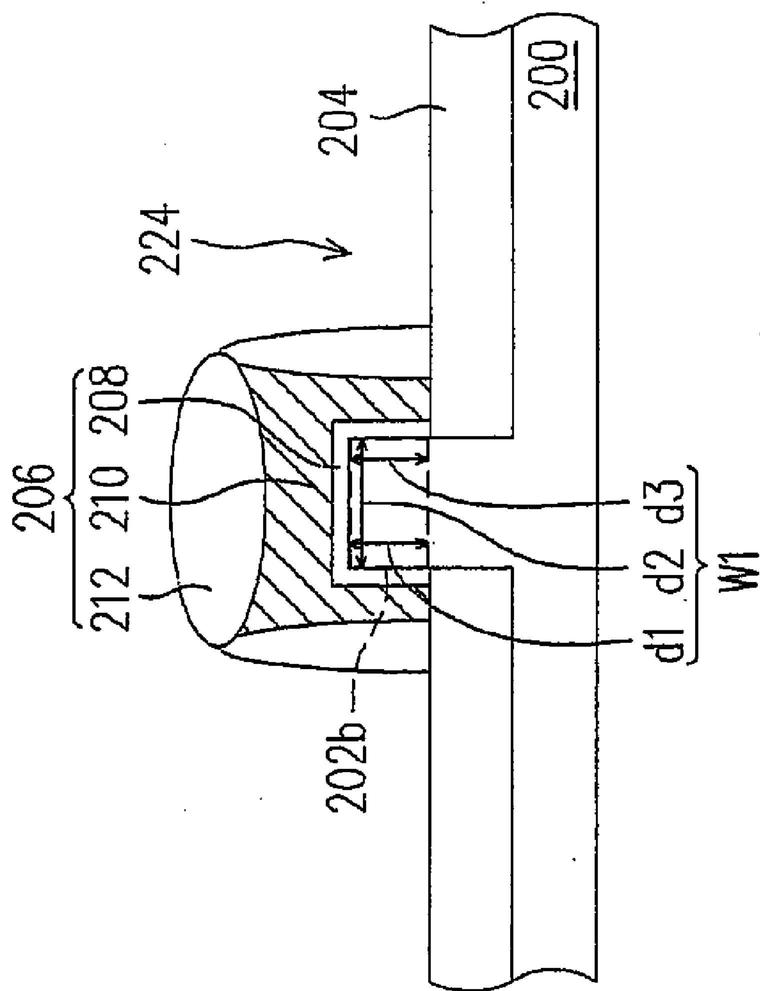


FIG. 2D

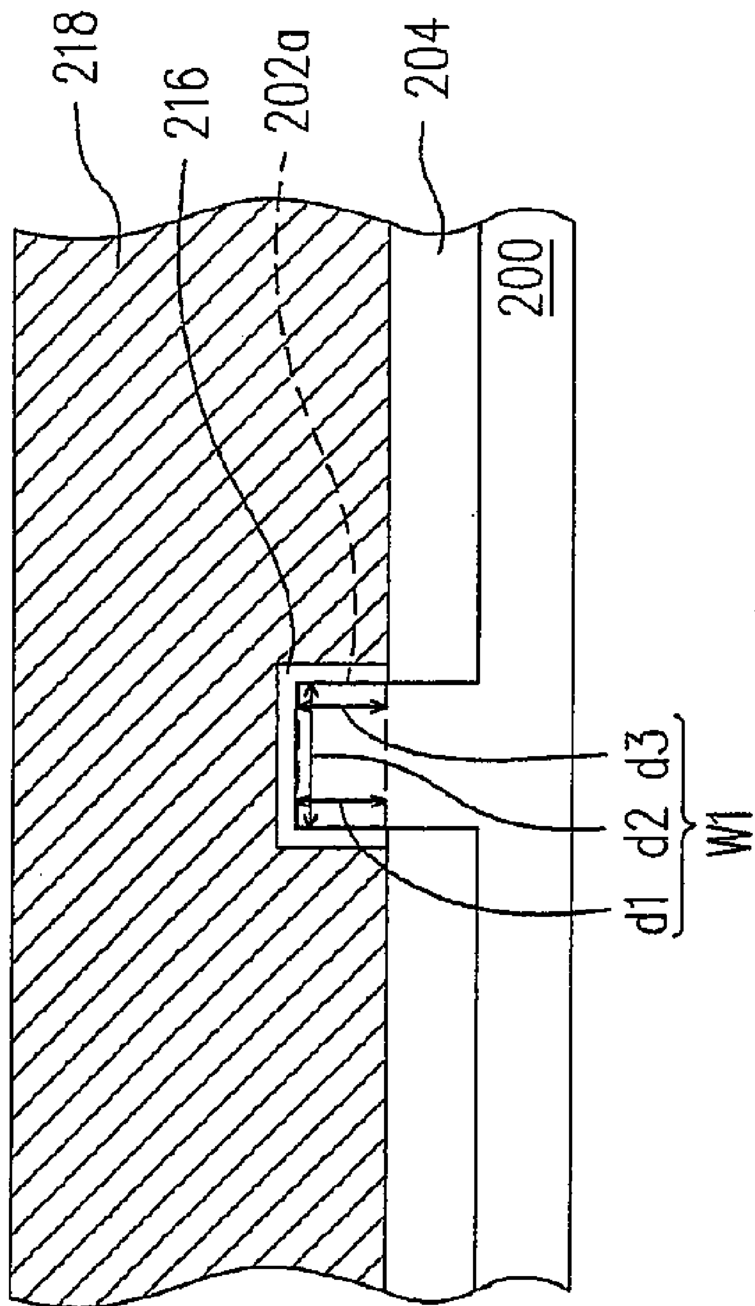


FIG. 2E

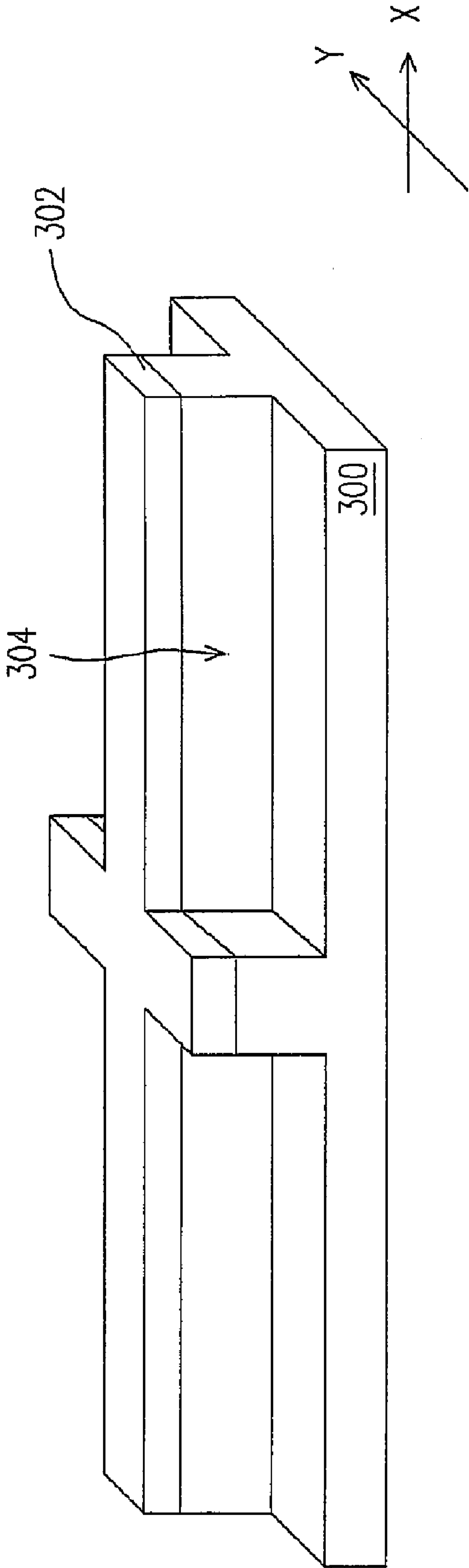


FIG. 3A

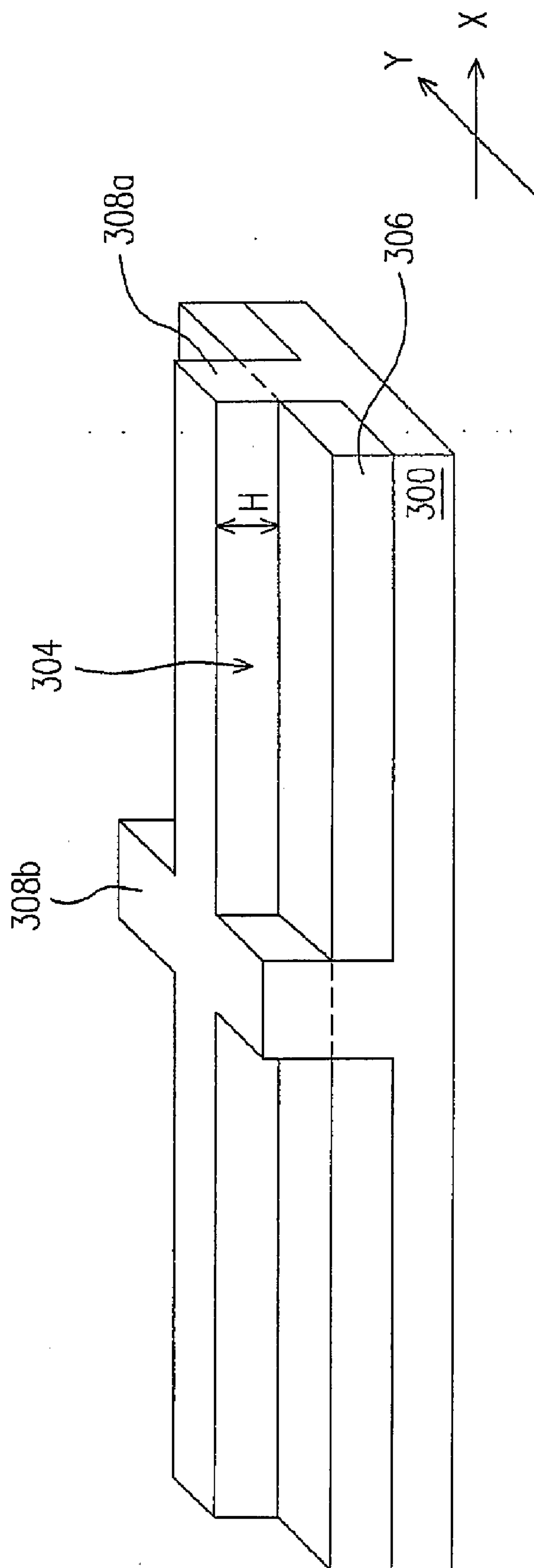


FIG. 3B

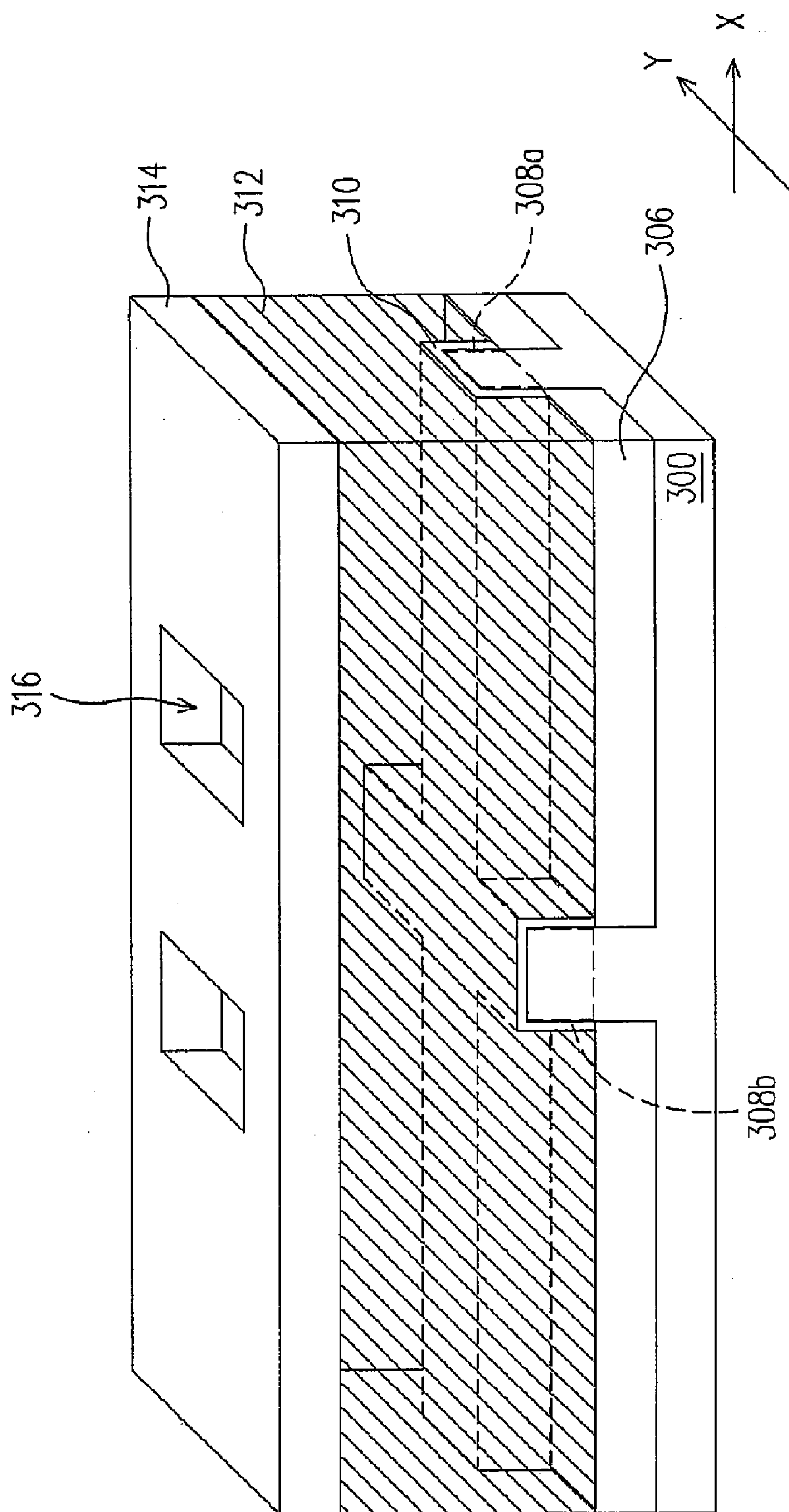


FIG. 3C

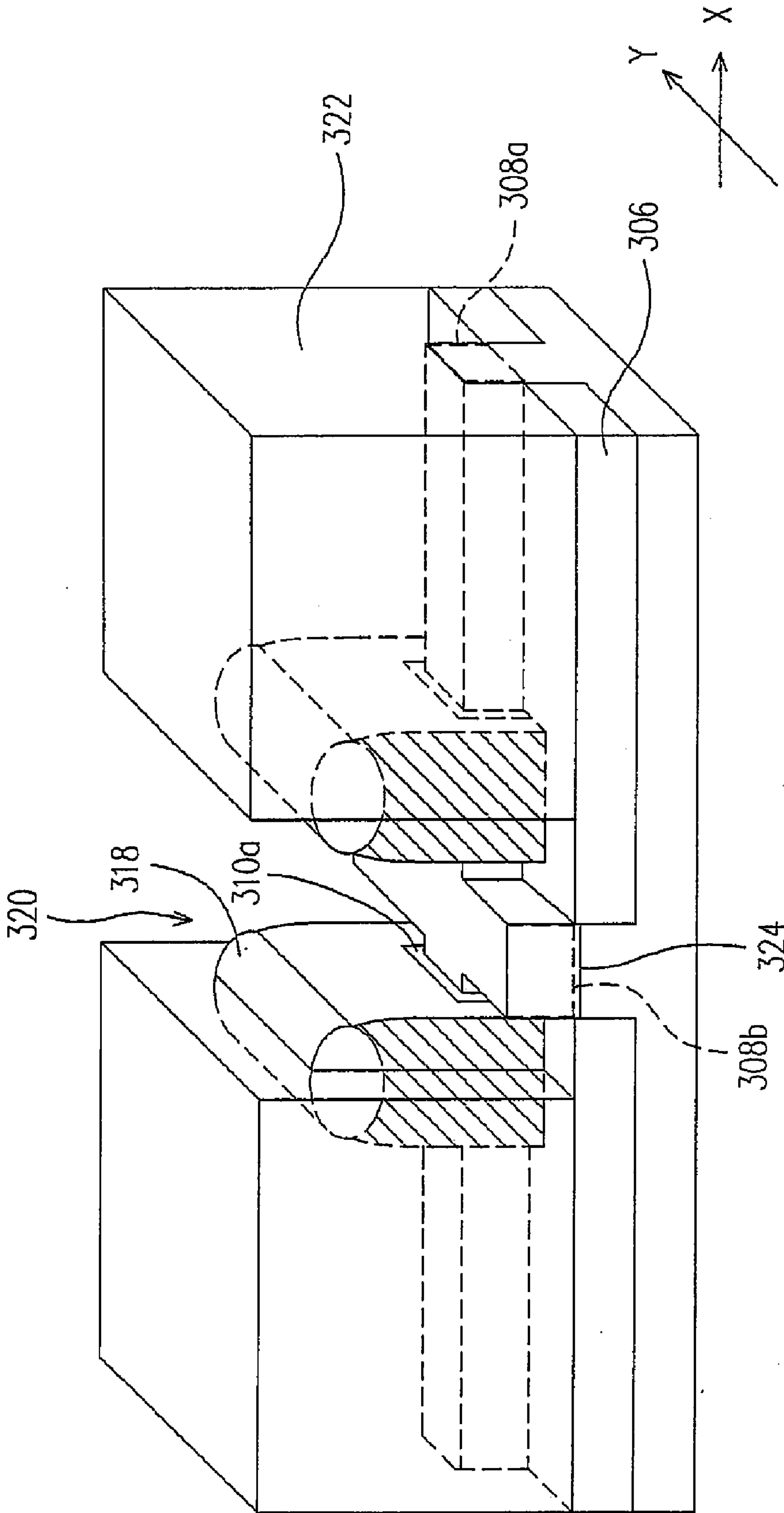


FIG. 3D

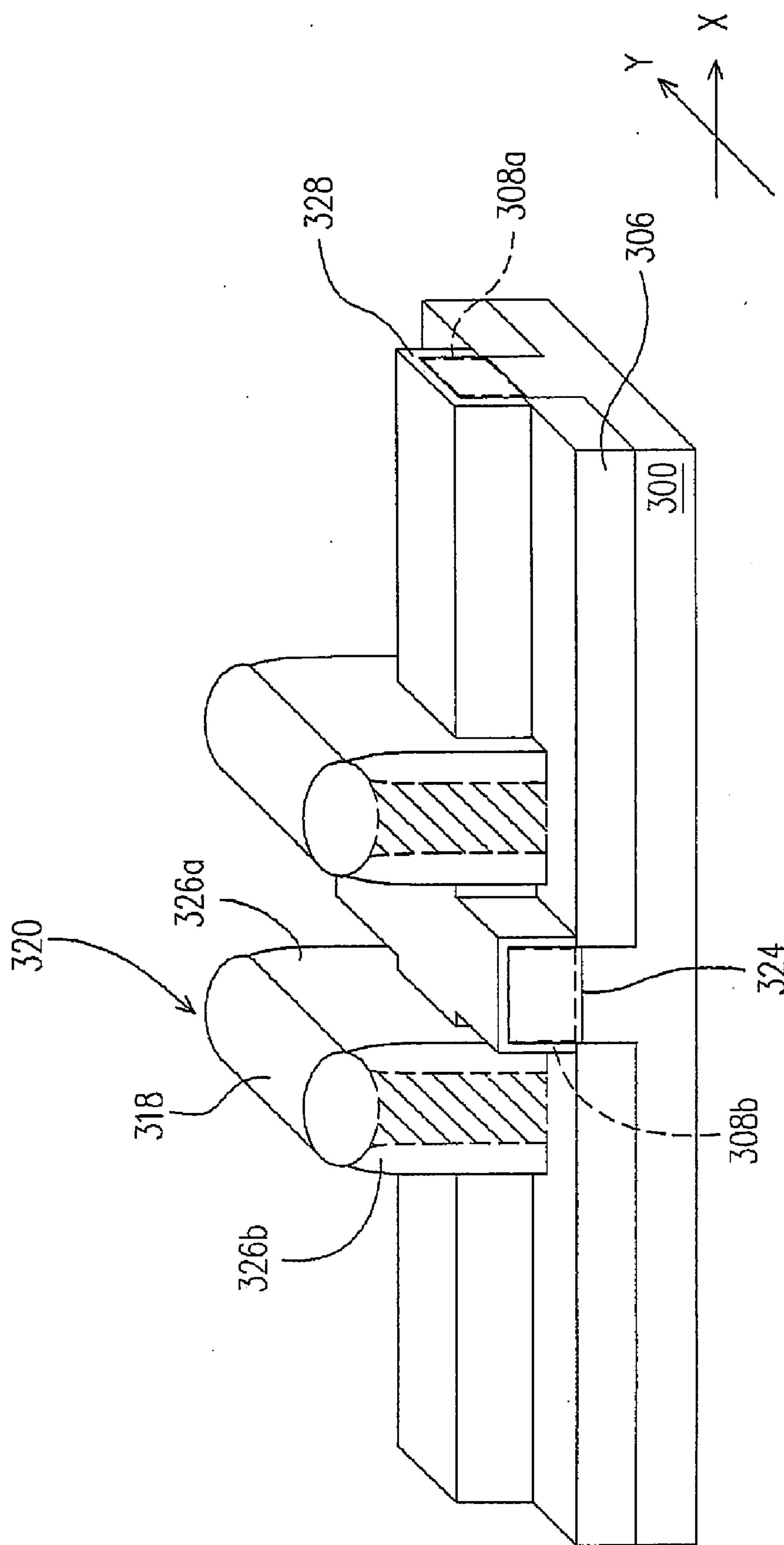


FIG. 3E

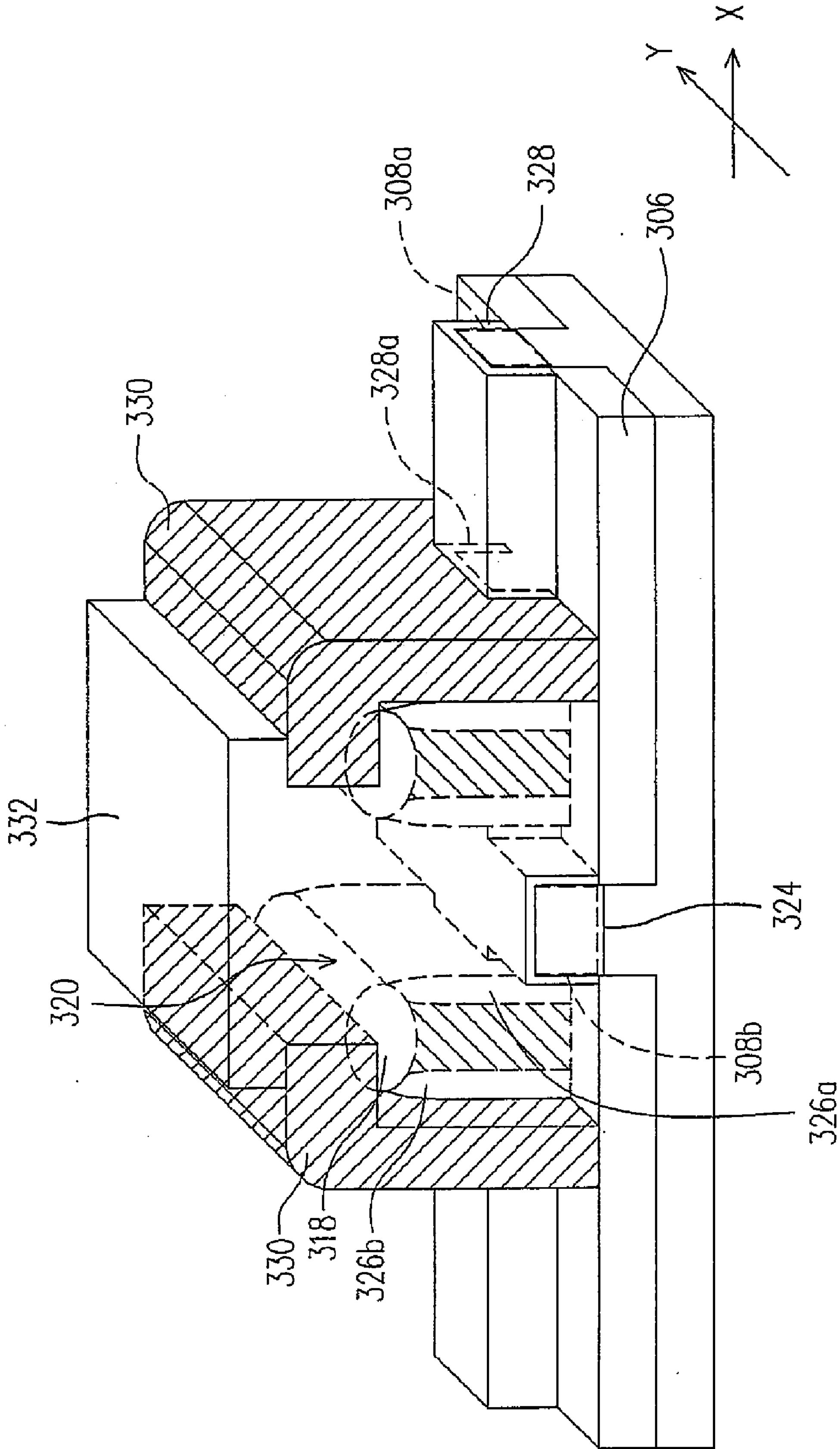


FIG. 3F

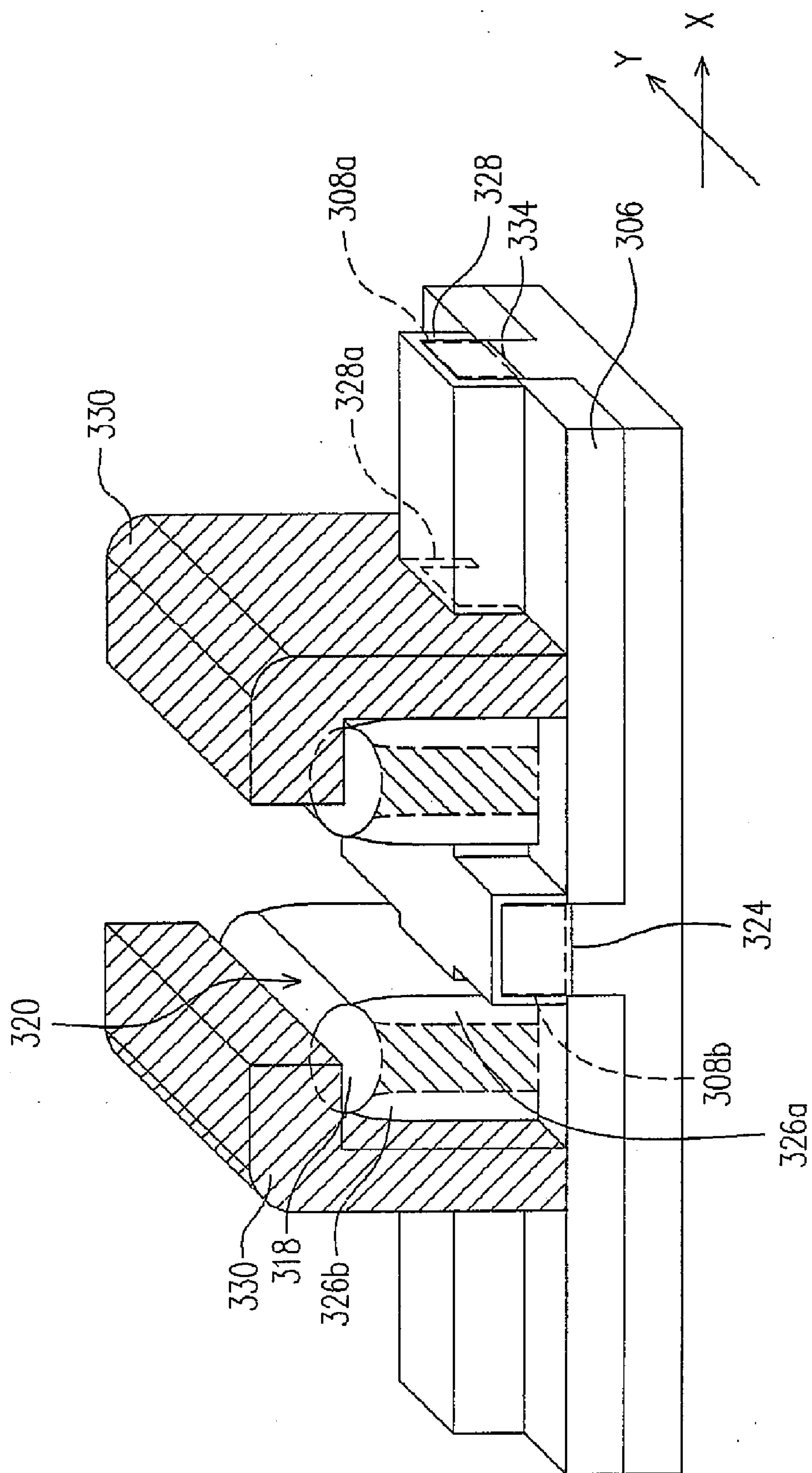


FIG. 3G

METHOD OF MANUFACTURING SPLIT GATE FLASH MEMORY

CROSS-REFERENCE TO RELATED APPLICATION

[0001] This application is a divisional of a prior application Ser. No. 11/163,223, filed on Oct. 11, 2005, which claims the priority benefit of Taiwan application serial no. 94108557, filed on Mar. 21, 2005 and now allowed. The entirety of each of the above-mentioned patent applications is hereby incorporated by reference herein and made a part of this specification

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The present invention relates to a method of manufacturing a semiconductor device. More particularly, the present invention relates to a method of manufacturing a split gate flash memory.

[0004] 2. Description of the Related Art

[0005] Among the various types of non-volatile memory products, flash memory device is a memory device that has been widely used inside personal computer systems and electron equipment. In the flash memory, data can be stored, read out or erased many numerous times and stored data are retained even after power is cut off.

[0006] Typically, the floating gate and the control gate of a flash memory cell are fabricated by doped polysilicon. Furthermore, the floating gate and the control gate are isolated from each other through a dielectric layer and the floating gate and the substrate are isolated from each other through a tunneling oxide layer. To perform a data write/erase operation on the flash memory, a biased voltage is applied to the control gate and the source/drain regions. As a result, electrons are injected into the floating gate or pulled out from the floating gate. To read data from a flash memory, an operating voltage is applied to the control gate so that the charging state of the floating gate will effect the 'on' or 'off' state of the channel underneath. Consequently, the 'on' or 'off' state of the channel can be used to determine if a '0' or '1' data bit is read out.

[0007] Because the quantity of electric charges expelled from the floating gate when erasing data from the aforementioned flash memory is difficult to control, too many charges may be expelled from the floating gate leading to the floating gate having a net positive charge, the so-called 'over-erasing'. If the degree of over-erasing is severe, the channel underneath the floating gate may continue to be conductive even though no operation voltage is applied to the control gate. As a result, errors in reading from the flash memory may occur.

[0008] To resolve the over-erasing problem, a flash memory with a split gate structure has been developed. FIG. **1** is a schematic cross-sectional view showing the structure of a conventional split-gate flash memory cell. The flash memory cell in FIG. **1** has a structure including a tunneling dielectric layer **102**, a floating gate **104**, an inter-gate dielectric layer **106** and a select gate **108** sequentially formed over a substrate **100**. Aside from covering the floating gate **104**, a portion of the select gate **108** also extends to cover an area above the substrate **100**. The select gate **108** is isolated from the substrate **100** through a select gate dielectric layer **110**. The source region **112** is formed in the substrate **100** on one side of the floating gate **104**. The drain region **114** is formed in the substrate **100** on the same side as the extension of the select gate **108**. With this setup, even when the over-erasing problem is so severe that the channel underneath the floating gate **104** remains open in the absence of an operating voltage to the select gate **108**, the channel underneath the select gate **108** is still maintained in the shut-down state. Thus, the drain region **114** and the source region **112** are cut off from each other and reading error is prevented.

[0009] However, a split gate structure needs a larger area to accommodate the split gate so that the size of each memory cell has to increase. Thus, the memory cell with a split gate structure must occupy a larger area compared with the memory cell with a stacked gate structure. In other words, the level of integration of the devices can hardly increase.

[0010] Furthermore, as the level of integration of integrated circuits continues to increase through miniaturization, the dimension of each memory cell can be reduced by shortening the length of the gate. Yet, a shorter gate length will lead to a reduction of the channel length underneath the gate. With a shorter channel, the chance of having an abnormal punch through between the drain region and the source region is increased during memory cell programming. Ultimately, the electrical performance of the memory cell will be seriously affected.

SUMMARY OF THE INVENTION

[0011] Accordingly, at least one objective of the present invention is to provide a split gate flash memory cell and manufacturing method thereof. The split gate memory cell has a three-dimensional channel path and can provide a longer channel length. Thus, the level of integration of the memory devices can be increased.

[0012] To achieve these and other advantages and in accordance with the purpose of the invention, as embodied and broadly described herein, the invention provides a split gate flash memory. The split gate flash memory includes a substrate, at least an active layer, a plurality of device isolation layers and at least a memory cell. The active layer is formed on the substrate to protrude above the surface of the substrate. The device isolation layers are formed on the sides of the active layer. Furthermore, the surface of the device isolation layer is lower than the surface of the active layer. The memory cell is formed on the substrate. The memory cell includes a stacked gate structure, a select gate, a source region and a drain region. The stacked gate structure includes at least a floating gate that crosses over the active layer. The select gate is formed on a first side of the stacked gate structure across the active layer. The source region is formed in the active layer on a second side of the stacked gate structure. The source region extends to an area underneath the floating gate. The second side and the first side are reciprocal to each other. The drain region is formed in the active layer on one side of the select gate.

[0013] The aforementioned split gate flash memory further includes an inter-gate dielectric layer formed on the sidewall of the stacked gate structure. The inter-gate dielectric layer can be fabricated by silicon oxide, for example.

[0014] The aforementioned split gate flash memory further includes a select gate dielectric layer formed between the select gate and the active layer. The select gate dielectric layer is fabricated by silicon oxide, for example.

[0015] In the aforementioned split gate flash memory, the stacked gate structure further includes a tunneling dielectric layer formed between the floating gate and the active layer. The tunneling dielectric layer can be fabricated by silicon oxide, for example.

[0016] In the aforementioned split gate flash memory, the stacked gate structure further includes a cap layer formed between the floating gate and the select gate. The floating gate and the select gate can be fabricated by doped polysilicon, for example.

[0017] The present invention also provides an alternative split gate flash memory. The split gate flash memory includes a substrate, a plurality of first active layers, a plurality of second active layers, a plurality of device isolation structures, a first memory cell and a second memory cell. The first active layers are formed on the substrate. The first active layers are aligned in a first direction parallel to one another to protrude above the surface of the substrate. The second active layers are formed on the substrate. The second active layers are aligned in a second direction parallel to one another such that the second direction and the first direction are formed across each other. The device isolation layers are formed on the sides of the first active layers. The surface of the device isolation layers is lower than the surface of the first active layers. The first memory cell is formed on the substrate across the first active layer. The first memory cell includes a stacked gate structure, a select gate, a source region and a drain region. The stacked gate structure includes at least a floating gate that formed across the first active layer. The select gate is formed on a first side of the stacked gate structure across the first active layer. The source region is formed in the first active layer on a second side of the floating gate. The second side and the first side are reciprocal to each other. The drain region is formed in the first active layer on one side of the select gate. The second memory cell is formed on the substrate across the first active layer. The second memory cell has a structure identical and symmetry to the first memory cell. The second and the first memory cell share the same source region.

[0018] In the aforementioned split gate flash memory, the source region of the second memory cell and the first memory cell is formed in the second active layer. Furthermore, the select gates in the second direction are serially connected together.

[0019] The aforementioned split gate flash memory further includes an inter-gate dielectric layer formed on the sidewalls of the stacked gate structures. The inter-gate dielectric layer can be fabricated by silicon oxide, for example.

[0020] The aforementioned split gate flash memory further includes a select gate dielectric layer formed between the select gate and the first active layer. The select gate dielectric layer can be fabricated by silicon oxide, for example.

[0021] In the aforementioned split gate flash memory, the stacked gate structure further includes a tunneling dielectric layer formed between the floating gate and the first active layer. The tunneling dielectric layer can be fabricated by silicon oxide, for example.

[0022] In the aforementioned split gate flash memory, the stacked gate structure further includes a cap layer formed between the floating gate and the select gate.

[0023] Because the memory cell of the split gate flash memory in the present invention has a three-dimensional channel path, the channel length is increased. As a result, the size of each memory cell is reduced and the level of integration of the devices is increased.

[0024] The present invention also provides a method of manufacturing a split gate flash memory. First, a substrate is provided. Then, an active layer protruding above the surface of the substrate is formed over the substrate. Thereafter, a plurality of device isolation layers is formed on the sides of the active layer. The surface of the device isolation layers is lower than the surface of the active layer. After that, at least a stacked gate structure having a floating gate therein is formed over the substrate. The floating gate is formed across the active layer. Then, a source region is formed in the active layer on a first side of the stacked gate structure and a select gate is formed on the sidewall of a second side of the stacked gate structure and on the active layer. Thereafter, a drain region is formed in the active layer on one side of the select gate.

[0025] In the aforementioned method of manufacturing a split gate flash memory, the stacked gate structure further includes a tunneling dielectric layer formed between the floating gate and the active layer and a cap layer formed on the floating gate.

[0026] In the aforementioned method of manufacturing a split gate flash memory, the process of forming the stacked gate structure on the substrate includes the following steps. First, a dielectric layer, a first conductive layer and a mask layer are sequentially formed over the substrate. Then, the mask layer is patterned to form an opening that exposes a portion of the first conductive layer. After forming a cap layer over the exposed first conductive layer, the mask layer is removed. Thereafter, using the cap layer as a mask, the first conductive layer and the dielectric layer are etched to form the stacked gate structure.

[0027] In the aforementioned method of manufacturing a split gate flash memory, the process of forming the select gate on the sidewall on the second side of the stacked gate structure and on the active layer includes forming a second conductive layer over the substrate. Thereafter, the second conductive layer is patterned to form the select gate on the sidewall on the second side of the stacked gate structure and on the active layer.

[0028] In the aforementioned method of manufacturing a split gate flash memory, an inter-gate dielectric layer is formed on the sidewalls of the stacked gate structure and a select gate dielectric layer is formed on the active layer on the second side of the stacked gate structure after forming the stacked gate structure on the substrate. The method of forming the inter-gate dielectric layer on the sidewalls of the stacked gate structure includes performing a thermal oxidation process.

[0029] In the aforementioned method of manufacturing a split gate flash memory, the process of forming the active layer on the substrate includes forming a plurality of trenches in the substrate.

[0030] The present invention also provides an alternative method of manufacturing a split gate flash memory. First, a substrate is provided. Then, a plurality of first active layers aligned in a first direction parallel to one another and a plurality of second active layers aligned in a second direction parallel to one another are formed on the substrate such that the first direction and the second direction are formed across each other. Furthermore, the first active layers and the second active layers protrude above the surface of the substrate. Thereafter, a plurality of device isolation layers is formed on the respective sides of the first active layers and the second active layers. The surface of the device isolation layers is lower than the surface of the first active layers and the second active layers. After that, a plurality of stacked gate structures is formed on the substrate. The stacked gate structures are aligned to form an array. Each stacked gate structure includes at least a floating gate. The floating gate is formed across the first active layer. Then, a plurality of source regions is formed in a portion of the first active layer and the second active layer between two adjacent stacked gate structures. The source regions in the second direction are serially connected together through the second active layers. A plurality of select gates is formed on one sidewall of the stacked gate structures and on the first active layers respectively. The select gates in the second direction are electrically connected together. After that, a plurality of drain regions is formed in the respective first active layers on a reciprocal side of the select gate and the source region.

[0031] In the aforementioned method of manufacturing a split gate flash memory, each stacked gate structure includes a tunneling dielectric layer formed between the floating gate and the active layer and a cap layer formed on the floating gate.

[0032] In the aforementioned method of manufacturing a split gate flash memory, the process of forming the stacked gate structures on the substrate includes the following steps. First, a dielectric layer, a first conductive layer and a mask layer are sequentially formed over the substrate. Then, the mask layer is patterned to form a plurality of openings that exposes a portion of the first conductive layer. After forming a cap layer over the exposed first conductive layer, the mask layer is removed. Thereafter, using the cap layer as a mask, the first conductive layer and the dielectric layer are etched to form the stacked gate structures.

[0033] In the aforementioned method of manufacturing a split gate flash memory, the process of forming the select gates on one sidewall of the stacked gate structures and on the first active layers includes forming a second conductive layer over the substrate. Thereafter, the second conductive layer is patterned to form the select gates on the sidewall of the stacked gate structures and on the first active layers.

[0034] In the aforementioned method of manufacturing a split gate flash memory, inter-gate dielectric layers are formed on the sidewalls of the stacked gate structures and select gate dielectric layers are formed on the first active layers on one side of the stacked gate structures after forming the stacked gate structures on the substrate.

[0035] In the aforementioned method of manufacturing a split gate flash memory, the method of forming the inter-gate dielectric layers on the sidewalls of the stacked gate structures includes performing a thermal oxidation process.

[0036] In the aforementioned method of manufacturing a split gate flash memory, the process of forming the first active layer and the second active layers on the substrate includes forming a plurality of trenches in the substrate.

[0037] In the method of manufacturing a split gate flash memory according to the present invention, the floating gates and the select gates are formed on and across the active layers. Therefore, the memory cells have three-dimensional channel so that the channel length is increased. As a result, the size of each memory cell is reduced and the level of integration of the devices is increased.

[0038] Furthermore, the width of the channel can be determined by the depth of the trench and the thickness of the device isolation layer filling the trench.

[0039] It is to be understood that both the foregoing general description and the following detailed description are exemplary, and are intended to provide further explanation of the invention as claimed.

BRIEF DESCRIPTION OF THE DRAWINGS

[0040] The accompanying drawings are included to provide a further understanding of the invention, and are incorporated in and constitute a part of this specification. The drawings illustrate embodiments of the invention and, together with the description, serve to explain the principles of the invention.

[0041] FIG. **1** is a schematic cross-sectional view showing the structure of a conventional split-gate flash memory cell.

[0042] FIG. **2**A is a top view of a split gate flash memory according to one preferred embodiment of the present invention.

[0043] FIG. **2**B is a perspective view showing the 'A' portion of the structure in FIG. **2**A.

[0044] FIG. **2**C is a cross-sectional view showing the structure along line B-B' of FIG. **2**A.

[0045] FIG. **2**D is a cross-sectional view showing the structure along line C-C' of FIG. **2**A.

[0046] FIG. **2**E is a cross-sectional view showing the structure along line D-D' of FIG. **2**A.

[0047] FIGS. **3**A through **3**G are schematic cross-sectional views showing the steps for fabricating a split gate flash memory according to one preferred embodiment of the present invention.

DESCRIPTION OF THE PREFERRED EMBODIMENTS

[0048] Reference will now be made in detail to the present preferred embodiments of the invention, examples of which are illustrated in the accompanying drawings. Wherever possible, the same reference numbers are used in the drawings and the description to refer to the same or like parts.

[0049] FIG. **2**A is a top view of a split gate flash memory according to one preferred embodiment of the present invention. FIG. **2**B is a perspective view showing the 'A' portion of the structure in FIG. **2**A. FIG. **2**C is a cross-sectional view

showing the structure along line B-B' of FIG. **2**A. FIG. **2**D is a cross-sectional view showing the structure along line C-C' of FIG. **2**A. FIG. **2**E is a cross-sectional view showing the structure along line D-D' of FIG. **2**A.

[**0050**] FIGS. **2**A through **2**E are used to describe the structure of a split gate flash memory according to the present invention. The flash memory cell of the present invention includes at least a substrate **200**, an active layer **202***a*, another active layer **202***b*, a device isolation layer **204**, a stacked gate structure **206**, an inter-gate dielectric layer **214**, a select gate dielectric layer **216**, a select gate **218**, a source region **220** and a drain region **222**.

[**0051**] The substrate **200** is a silicon substrate, for example. The active layers **202***a* and **202***b* protrude above the surface of the substrate **200**. The active layers **202***a* and **202***b* are defined by forming a plurality of trenches **224** in the substrate **200**, for example. The active layers **202***a* are aligned, for example, in an X direction parallel to one another. Moreover, the active layers **202***a* extend in the X direction in the form of linear strips. The active layers **202***b* are aligned, for example, in a Y direction parallel to one another. The X direction and the Y direction cross over each other. In other words, the active layers **202***a* are formed across the active layers **202***b*.

[**0052**] The device isolation layers **204** are disposed on the respective sides of the active layers **202***a* and the active layers **202***b*. In other words, the device isolation layers **204** are disposed in the trenches **224** to isolate the substrate **200** from the stacked gate structures **206** and the select gates **218**. Furthermore, the surface of the device isolation layers **204** is lower than the surface of the active layers **202***a* and the active layers **202***b*. That is, the active layers **202***a* and **202***b* protrude above the surface of the device isolation layers **204** to form fin-like structures. The device isolation layers **204** are fabricated by silicon oxide, for example.

[**0053**] Each stacked gate structure **206** includes, for example, a tunneling dielectric layer **208**, a floating gate **210** and a cap layer **212**. The floating gate **210** is formed on the substrate **200** across the active layer **202***a* and the tunneling dielectric layer **208** is formed between the floating gate **204** and the active layer **202***a*. The floating gate **210** is fabricated by doped polysilicon, for example. The tunneling dielectric layer is formed between the floating gate **204** and the substrate **200** and is fabricated by silicon oxide, for example. The cap layer **206** is formed on the floating gate **204** and is fabricated by silicon oxide, for example.

[**0054**] The inter-gate dielectric layers **214** are formed on the sidewalls of the stacked gate structures **206**. The inter-gate dielectric layers **214** are fabricated by silicon oxide, for example.

[**0055**] The select gates **218** are formed on one side of the stacked gate structures **206** across the active layer **202***a*. As shown in FIG. **2**A, the select gates **218** are serially connected in the Y direction, for example. The select gates **218** are fabricated by doped polysilicon, for example.

[**0056**] The select gate dielectric layers **216** are formed between the select gates **218** and the active layers **202***a*. The select gate dielectric layers **216** are fabricated by silicon oxide, for example. The source region **220** is formed in the active layer **202***a* and the active layer **202***b* on another side of the of the stacked gate structure **206**. The source regions **220** in the Y direction are connected together through the active layers **202***b*, for example. The source regions **220** extend to an area underneath the floating gate **210** of various memory cells. The drain regions **222** are formed in the active layers **202***a* on one side of the select gates **218**, for example. The source region **220** and the drain region **222** are disposed on reciprocal locations separated from each other by the stacked gate structure **206** and the select gate **218**.

[**0057**] The stacked gate structure **206** (including the tunneling dielectric layer **208**, the floating gate **210** and the cap layer **212**), the select gate **218**, the select gate dielectric layer **216**, the inter-gate dielectric layer **214**, the source region **220** and the drain region **222** together form a memory cell **226**. A plurality of the memory cells **226** is formed over the substrate **200**. The memory cells **226** are aligned to form an array. As shown in FIG. **2**B, every pair of adjacent memory cells **226** has an identical structure that shares a source region **220**.

[**0058**] As shown in FIGS. **2**B, **2**C, **2**D and **2**E, both the floating gate **210** and the select gate **218** of the memory cell **226** are formed across the fin-like active layer **202***a*. Therefore, the memory cell **226** has a three-dimensional channel path leading to a greater channel **228** width. For example, as shown in FIG. **2**D, the channel width W**1** underneath the floating gate **210** is the sum of the distances d**1**, d**2** and d**3**, and as shown in FIG. **2**E, the channel width W**2** underneath the select gate **210** is the sum of the distances d**1**, d**2** and d**3**. That is why the dimension of each memory cell can be reduced to increase the overall level of integration of the devices.

[**0059**] Furthermore, the total width of the channel **228** is the sum of the distances d**1**, d**2** and d**3** and the distance d**1** and d**3** is determined by the depth of the trenches **224** and the thickness of the device isolation layer **204** filling the trenches **224**. Hence, the width of the channel **228** can be easily adjusted whenever it is required.

[**0060**] FIGS. **3**A through **3**G are schematic cross-sectional views showing the steps for fabricating a split gate flash memory according to one preferred embodiment of the present invention. In fact, FIGS. **3**A through **3**G are perspective views of the 'A' portion of the FIG. **2**A for showing the steps in the fabrication process.

[**0061**] First, as shown in FIG. **3**A, a substrate such as a silicon substrate is provided. Then, a mask layer **302** is formed on the substrate **300**. The mask layer **302** is a silicon nitride layer formed, for example, by performing a chemical vapor deposition process. Obviously, a pad oxide layer (not shown) can be formed between the mask layer **302** and the substrate **300**. The pad oxide layer is a silicon oxide layer formed, for example, by performing a thermal oxidation process. After patterning the mask layer **302**, a portion of the substrate **300** is removed to form trenches **304** in the substrate and define an active region using the mask layer **302** as a mask. The process of removing a portion of the substrate **300** includes performing a reactive ion etching operation, for example.

[**0062**] As shown in FIG. **3**B, device isolation layers **306** are formed in the trenches **304**. The top surface of the device isolation layers **306** is lower than the top surface of the substrate **300**. That portion of the substrate **300** protruding above the device isolation layers **306** is called the active

layers **308***a* and **308***b*. The active layer **308***a* appearing as a linear strip extends in the X direction, for example. The active layer **308***b* appearing as a linear strip extends in the Y direction, for example. In other words, the active layers **308***a* and **308***b* cross over each other. In the present embodiment, only one single active layer **308***a* and one active layer **308***b* are shown. In practice, a plurality of active layers **308***a* aligned in the X direction parallel to one another and a plurality of active layers **308***b* aligned in the Y direction parallel to one another may be formed as shown in FIG. **2**A.

[**0063**] The method of forming the device isolation layers **306** includes forming an insulating material layer (not shown) over the substrate **300**. The insulating material layer has a thickness greater than the sum of the depth of the trench **304** and the thickness of the mask layer **302**. Thereafter, a planarization operation is carried out to planarize the upper surface of the insulating material layer. The method of planarizing the insulating material layer includes performing a chemical-mechanical polishing operation, for example. After that, a portion of the insulating material layer is removed so that the top surface of the insulating material layer is lower than the top surface of the substrate **300**, thereby forming the device isolation layers **306**. The method of removing a portion of the insulating material layer includes performing an etching back operation, for example. The height H of the active layer **308** is determined by the depth of the trench **304** and the thickness of the device isolation layer **206**. The height H of the active layer **308** will directly affect the channel width of a subsequently formed memory cell.

[**0064**] After forming the device isolation layers **306**, the mask layer **302** is removed. The method of removing the mask layer **302** includes performing a wet etching operation, for example.

[**0065**] As shown in FIG. **3**C, a dielectric layer **310** is formed over the substrate **300**. The dielectric layer **310** is a silicon oxide layer formed, for example, by performing a thermal oxidation process.

[**0066**] Thereafter, a conductive layer **312** fabricated by doped polysilicon is formed over the dielectric layer **310**. The method of forming the conductive layer **312** includes depositing undoped polysilicon in a chemical vapor deposition to form an undoped polysilicon layer and performing an ion implantation on the undoped polysilicon layer thereafter. Alternatively, the conductive layer **312** is formed in a chemical vapor deposition process with in-situ ion implantation.

[**0067**] After that, a mask layer **314** is formed over the conductive layer **312**. The mask layer **314** is a silicon nitride layer formed, for example, by performing a chemical vapor deposition process. The mask layer **314** is patterned to form a plurality of openings **316** that exposes the conductive layer **312**.

[**0068**] As shown in FIG. **3**D, a cap layer **318** is formed on the exposed conductive layer **312** within the openings **316**. The cap layer **318** is a silicon oxide layer formed, for example, by performing a thermal oxidation process. After forming the cap layer **310**, the mask layer **314** is removed. Thereafter, using the cap layer **318** as a self-aligning mask, the conductive layer **312** and the dielectric layer **310** are etched until the substrate **300** is exposed to form a conductive layer **312***a* and a tunneling dielectric layer **310***a*. The cap layer **318**, the conductive layer **312***a* and the tunneling dielectric layer **310***a* together form a stacked gate structure **312** with the conductive layer **312***a* serving as the floating gate of a memory cell.

[**0069**] Thereafter, a patterned photoresist layer **322** is formed over the substrate **300**. The patterned photoresist layer **322** exposes the areas for forming the source regions. An ion implant process is carried out to form a source region **324** in the substrate **300** on one side of the stacked gate structure **320**. The source region **324** is formed in the active layers **308***a* and **308***b* between two adjacent stacked gate structures **320**. The source regions **324** in the Y direction are connected through the active layer **308***b*, for example. In other words, the two subsequently formed adjacent memory cells will share the same source region **324**.

[**0070**] As shown in FIG. **3**E, the patterned photoresist layer **322** is removed. Thereafter, inter-gate dielectric layers **326***a* and **326***b* are formed on the respective sidewalls of the stacked gate structures **320**. The inter-gate dielectric layers **326***a* and **326***b* are silicon oxide layers or silicon oxide/nitride composite layers formed, for example, by performing a thermal oxidation process. Obviously, the process of forming the inter-gate dielectric layers **326***a* and **326***b* may include performing a thermal oxidation process to form a silicon oxide layer. Thereafter, a chemical vapor deposition process is performed to deposit a silicon oxide layer or a silicon nitride layer. Finally an anisotropic etching operation is carried out to remove a portion of the silicon oxide layer and the silicon nitride layer.

[**0071**] Thereafter, a dielectric layer **328** is formed over the substrate **300** (the active layers **308***a* and **308***b*). The dielectric layer **328** is a silicon oxide layer formed, for example, by performing a thermal oxidation process. Obviously, the dielectric layer **328** will also form over the source region **324**. In addition, a thin silicon oxide layer will also form on the surface of the inter-gate dielectric layers **326***a* and **326***b*. If the inter-gate dielectric layers **326***a* and **326***b* are silicon oxide/nitride composite layers, then the inter-gate dielectric layers **326***a* and **326***b* are converted to oxide/nitride/oxide structures after forming the dielectric layer **328**.

[**0072**] As shown in FIG. **3**F, a select gate **330** is formed on that side of the sidewall of the stacked gate structure **312** where the inter-gate dielectric layer **326***b* are formed thereon. The select gate **330** is formed, for example, by depositing a conductive material layer (not shown) on the substrate **300** and patterning the conductive material layer thereafter. The select gate **330** can be a doped polysilicon layer formed, for example, by depositing doped polysilicon in a chemical vapor deposition and performing an ion implant process, or performing an in-situ dopant implant in a chemical vapor deposition. The dielectric layer **328** between the select gate **330** and the active layer **308***a* serves as a select gate dielectric layer **328***a*.

[**0073**] Thereafter, another patterned photoresist layer **332** is formed over the substrate **300**. The patterned photoresist layer **332** covers the area above the source regions **324** and exposes the areas for forming the drain regions.

[**0074**] As shown in FIG. **3**G, an ion implantation is carried out to form a drain region **334** in the substrate **300** (the active layer **308***a*) on one side of the select gate **330**. Thereafter, the patterned photoresist layer **332** is removed and then the remaining steps necessary for forming a complete split gate flash memory cell performed. Since these steps involve familiar conventional processes, a detailed description is omitted.

[0075] In the aforementioned embodiment, the floating gate **312***a* and the select gate **330** are formed on and across the active layer **308***a*. Therefore, the memory cells have a three-dimensional channel path with a greater channel width. Furthermore, the width of the channel can be determined through the depth of the trench **304** and the thickness of the device isolation layer **306** filling the trench **304**.

[0076] It will be apparent to those skilled in the art that various modifications and variations can be made to the structure of the present invention without departing from the scope or spirit of the invention. In view of the foregoing, it is intended that the present invention cover modifications and variations of this invention provided they fall within the scope of the following claims and their equivalents.

What is claimed is:

1. A method of manufacturing a split gate flash memory, comprising:

providing a substrate;

forming an active layer on the substrate, wherein the active layer protrudes from the surface of the substrate;

forming a plurality of device isolation layers on the sides of the active layer, wherein the surface of the device isolation layers is lower than the surface of the active layer;

forming a stacked gate structure on the substrate, wherein the stacked gate structure comprises at least a floating gate that formed across the active layer;

forming a source region in the active layer on a first side of the stacked gate structure;

forming a select gate on the sidewall of a second side of the stacked gate structure and on the active layer; and

forming a drain region in the active layer on one side of the select gate.

2. The method of claim 1, wherein the stack gate structure comprises:

a tunneling dielectric layer formed between the floating gate and the active layer; and

a cap layer formed on the floating gate.

3. The method of claim 2, wherein the step of forming the stacked gate structure on the substrate comprises:

forming a dielectric layer, a first conductive layer and a mask layer sequentially over the substrate;

patterning the mask layer to form an opening that exposes a portion of the first conductive layer;

forming a cap layer on the exposed first conductive layer;

removing the mask layer; and

etching the first conductive layer and the dielectric layer using the cap layer as a mask to form the stacked gate structure.

4. The method of claim 1, wherein the step of forming the select gate on the sidewall on the second side of the stacked gate structure and on the first active layer comprises:

forming a second conductive layer over the substrate; and

patterning the second conductive layer to form the select gate on the sidewall on the second side of the stacked gate structure and on the first active layer.

5. The method of claim 1, wherein after forming the stacked gate structure on the substrate, further comprises:

forming an inter-gate dielectric layer on the sidewalls of the stacked gate structure; and

forming a select gate dielectric layer on the active layer on the second side of the stacked gate structure.

6. The method of claim 5, wherein the step of forming the inter-gate dielectric layer on the sidewalls of the stacked gate structure comprises performing a thermal oxidation process.

7. The method of claim 1, wherein the step of forming the active layers on the substrate comprises forming a plurality of trenches in the substrate.

8. A method of manufacturing a split gate flash memory, comprising:

providing a substrate;

forming a plurality of first active layers and a plurality of second active layers on the substrate, wherein the first active layers are aligned in a first direction parallel to one another and the second active layers are aligned in a second direction parallel to one another such that the first direction crosses the second direction, and the first active layers and the second active layers protrude from the surface of the substrate;

forming a plurality of device isolation layers on the sides of the first active layers and the second active layers, wherein the surface of the device isolation layers is lower than the surface of the first active layers and the second active layers;

forming a plurality of stacked gate structures on the substrate, wherein the stacked gate structures are aligned to form an array and each stacked gate structure comprises at least a floating gate is formed across the first active layer;

forming a plurality of source regions in the first active layer and the second active layer between two adjacent stacked gate structures, wherein the source regions in the second direction are serially connected together through the respective second active layers;

forming a plurality of select gates on one sidewall of the stacked gate structures and on the first active layers, wherein the select gates in the second direction are connected to one another; and

forming a plurality of drain regions in the respective first active layers on a reciprocal side of the select gates and the source regions.

9. The method of claim 8, wherein the stacked gate structure further comprises:

a tunneling dielectric layer formed between the floating gate and the active layer, and

a cap layer formed on the floating gate.

10. The method of claim 9, wherein the step of forming the stacked gate structures on the substrate comprises:

forming a dielectric layer, a first conductive layer and a mask layer sequentially over the substrate;

patterning the mask layer to form a plurality of openings that exposes a portion of the first conductive layer;

forming a cap layer on the exposed first conductive layer;

removing the mask layer;

etching the first conductive layer and the dielectric layer using the cap layer as a mask to form the stacked gate structures.

11. The method of claim 8, wherein the step of forming the select gates on one sidewall of the stacked gate structures and on the first active layers comprises:

forming a second conductive layer over the substrate; and

patterning the second conductive layer to form the select gates on the sidewalls of the stacked gate structures and on the first active layers.

12. The method of claim 8, wherein after forming the stacked gate structures on the substrate, further comprises:

forming an inter-gate dielectric layer on the respective sidewalls of the stacked gate structures; and

forming a select gate dielectric layer on one side of the stacked gate structures and on the first active layers.

13. The method of claim 12, wherein the step of forming the inter-gate dielectric layer on the sidewalls of the stacked gate structure comprises performing a thermal oxidation process.

14. The method of claim 8, wherein the step of forming the first active layers and the second active layers on the substrate comprises forming a plurality of trenches in the substrate.

* * * * *