US 20110286283A1

(19) United States

(12) Patent Application Publication

LUNG et al.

(10) Pub. No.: US 2011/0286283 A1

(43) Pub. Date: Nov. 24, 2011

(54) 3D TWO-BIT-PER-CELL NAND FLASH MEMORY

(75) Inventors: **HSIANG-LAN LUNG**, Dobbs Ferry, NY (US); **Hang-Ting Lue**, Hsinchu (TW); **Yen-Hao Shih**, Elmsford, NY (US); **Erh-Kun Lai**, Elmsford, NY (US); **Ming Hsiu Lee**, Hsinchu (TW); **Tien-Yen Wang**, Taipei (TW)

(73) Assignee: **Macronix International Co., Ltd.**, Hsinchu (TW)

(21) Appl. No.: **12/785,291**

(22) Filed: **May 21, 2010**

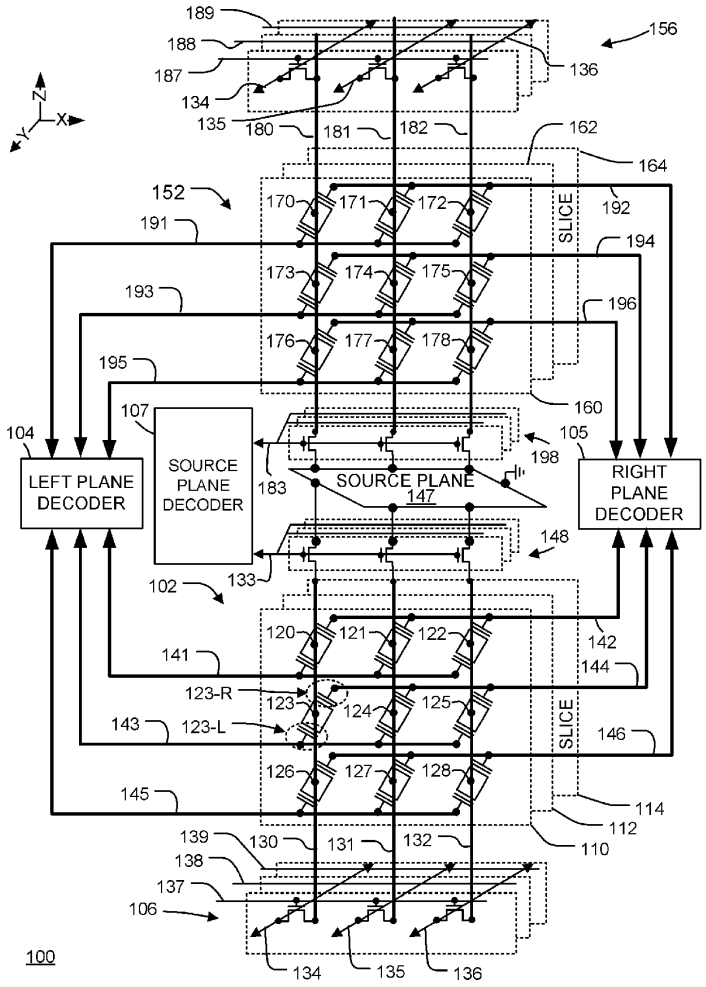
Publication Classification

(51) **Int. Cl.**
G11C 16/04 (2006.01)
H01L 21/336 (2006.01)

(52) **U.S. Cl.** **365/185.28**; 365/185.17; 438/257; 257/E21.409

(57) **ABSTRACT**

A 3D memory device is described which includes bottom and top memory cubes having respective arrays of vertical NAND string structures. A common source plane comprising a layer of conductive material is between the top and bottom memory cubes. The source plane is supplied a bias voltage such as ground, and is selectively coupled to an end of the vertical NAND string structures of the bottom and top memory cubes. Memory cells in a particular memory cube are read using current through the particular vertical NAND string between the source plane and a corresponding bit line coupled to another end of the particular vertical NAND string.



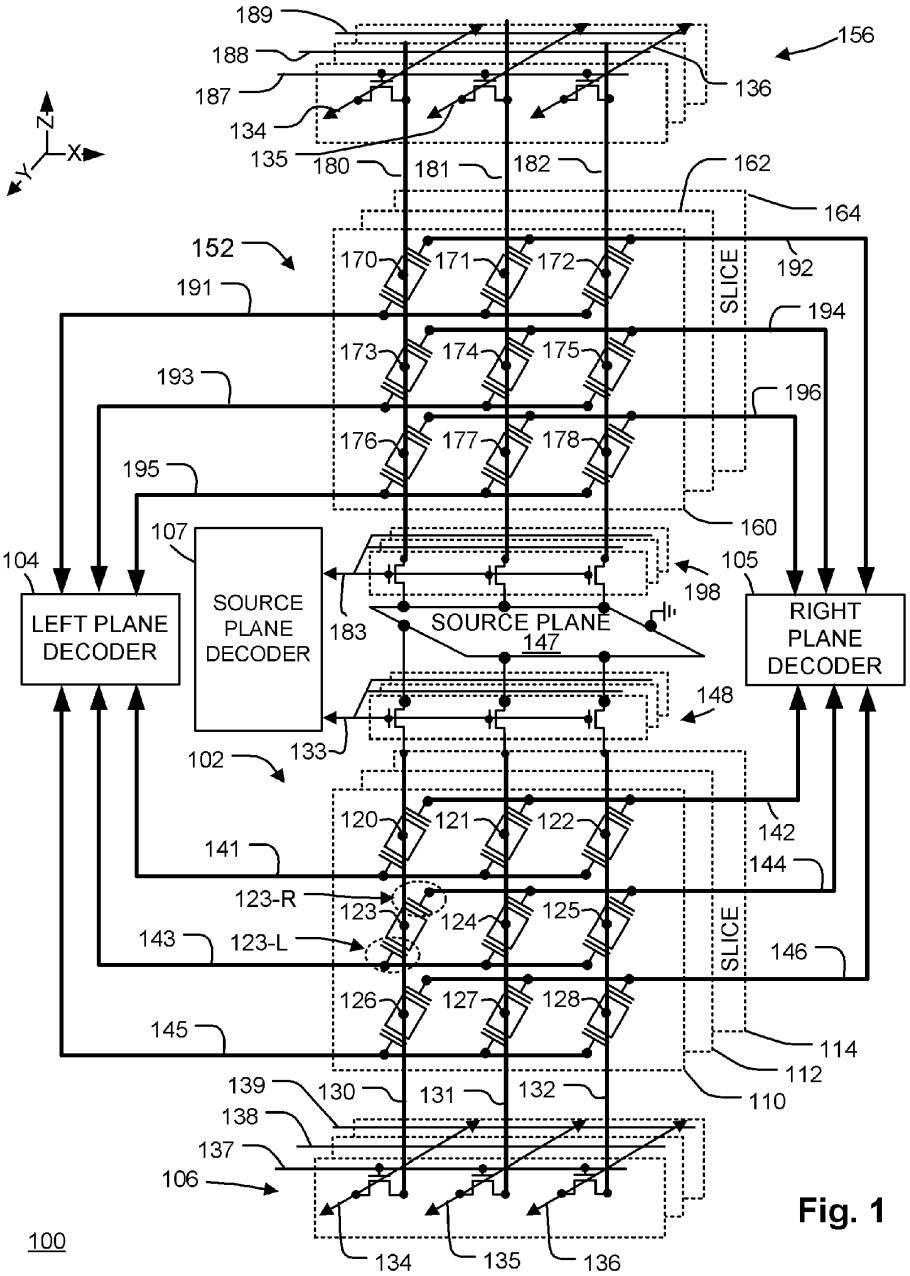

189
188
187
134
135
180
181
182
156
136
162
164
152
170
171
172
192
SLICE
191
194
173
174
175
193
196
176
177
178
195
160
107
104
105
198
LEFT PLANE DECODER
SOURCE PLANE DECODER
183
SOURCE PLANE 147
RIGHT PLANE DECODER
148
133
102
120
121
122
142
141
144
123-R
123
124
125
143
123-L
SLICE
146
126
127
128
145
114
112
110
139
130
131
132
138
137
106
100
134
135
136

Fig. 1

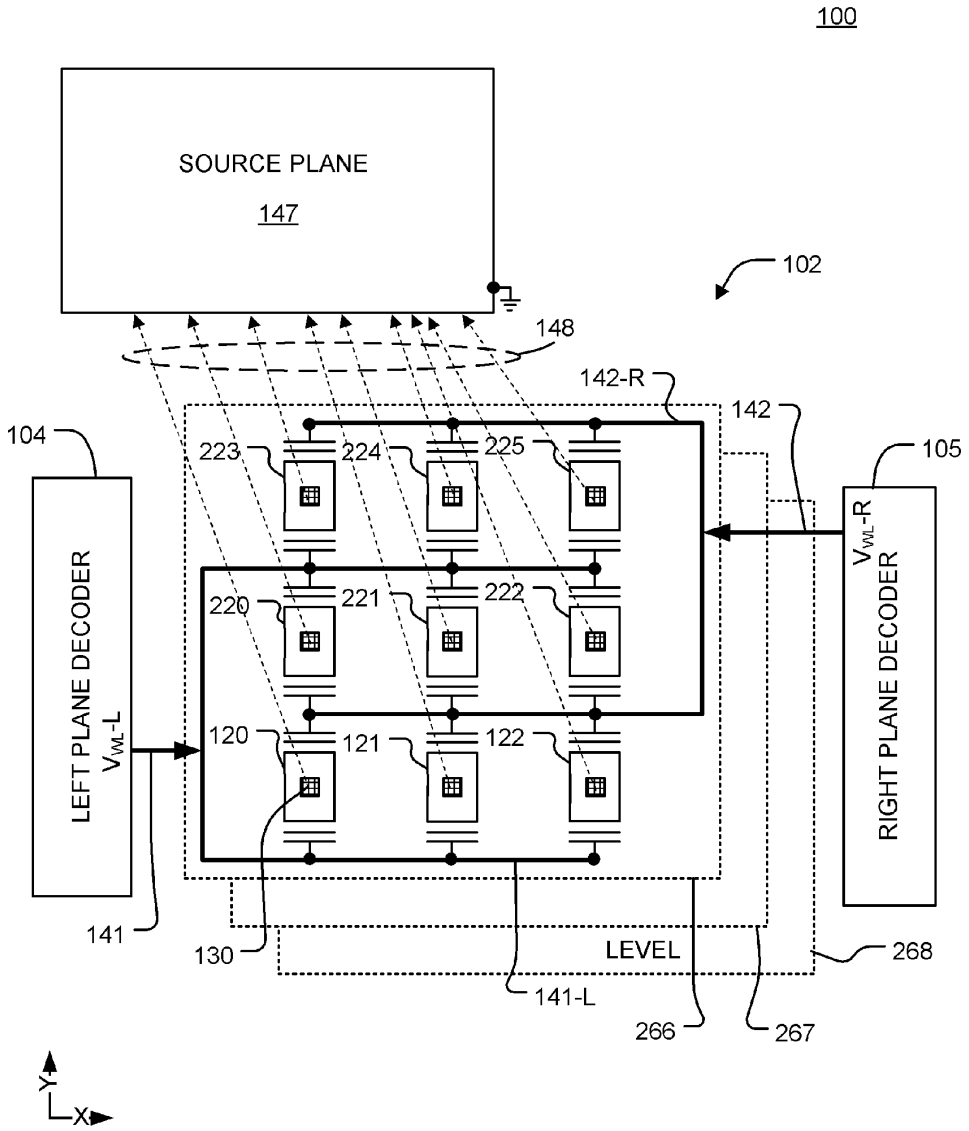

100
SOURCE PLANE
147
102
148
142-R
142
104
105
223
224
225
220
221
222
120
121
122
LEFT PLANE DECODER
V_{WL}-L
V_{WL}-R
RIGHT PLANE DECODER
141
130
LEVEL
268
141-L
266
267
Y
X

Fig. 2

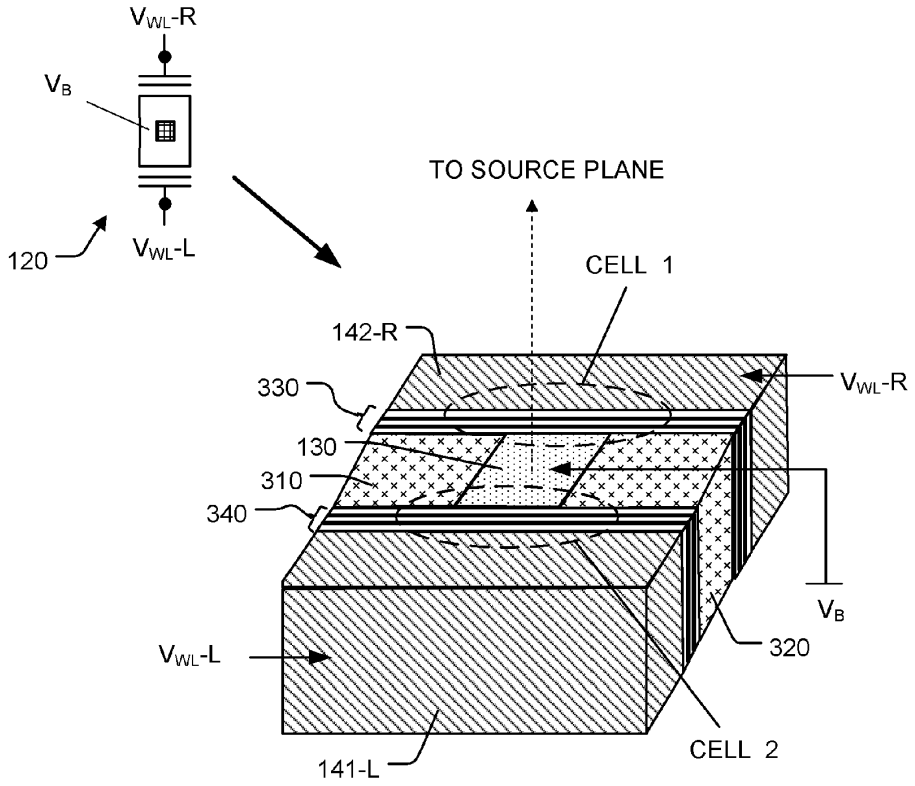

V_{WL}-R
V_B
120
V_{WL}-L
TO SOURCE PLANE
CELL 1
142-R
330
130
310
340
V_{WL}-R
V_B
320
V_{WL}-L
CELL 2
141-L

Fig. 3

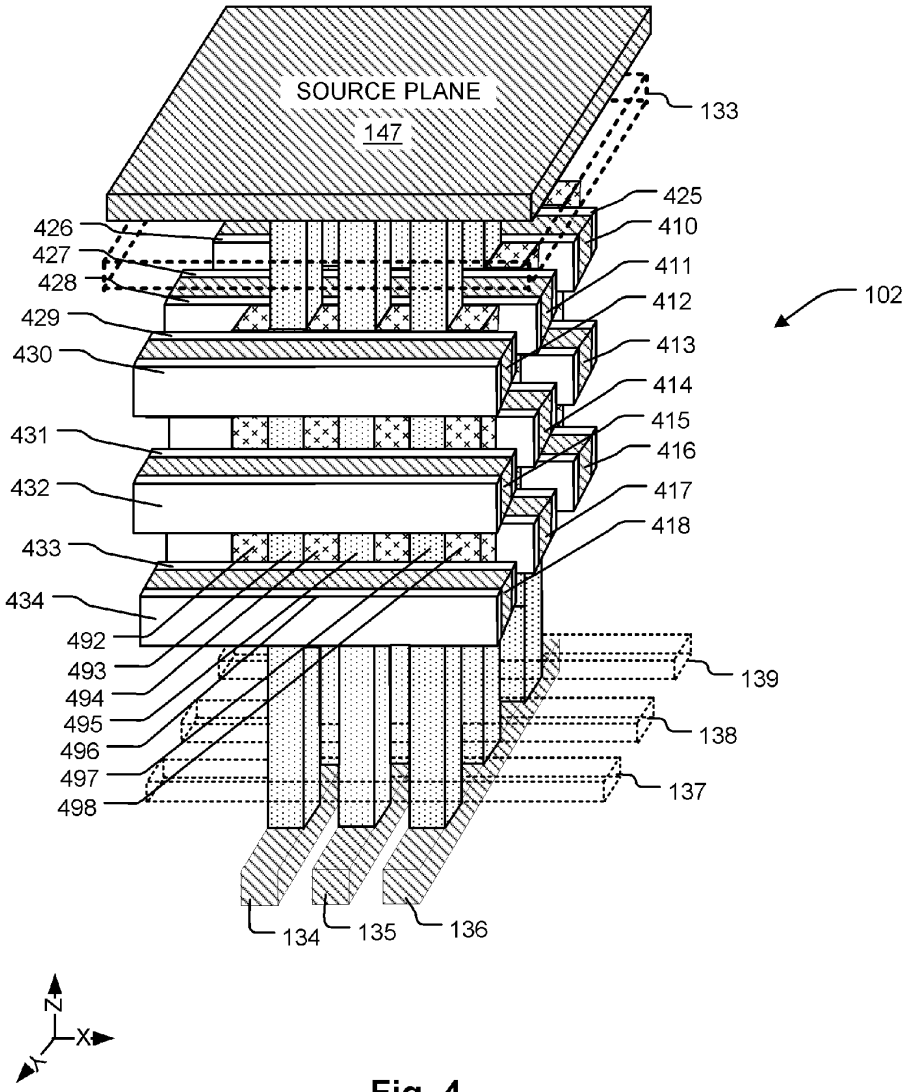

SOURCE PLANE
147
133
425
410
411
412
413
414
415
416
417
418
102
426
427
428
429
430
431
432
433
434
492
493
494
495
496
497
498
139
138
137
134
135
136
Z
X
Y

Fig. 4

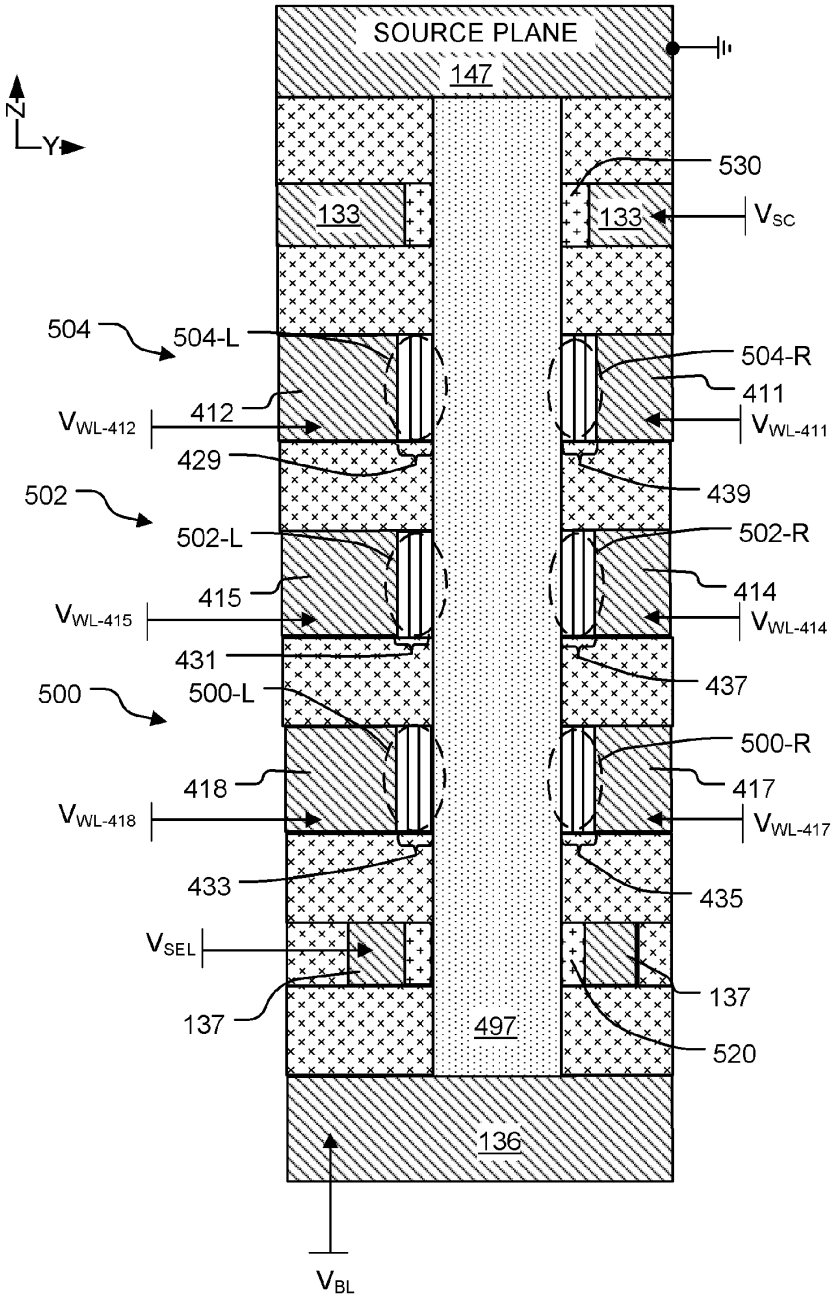
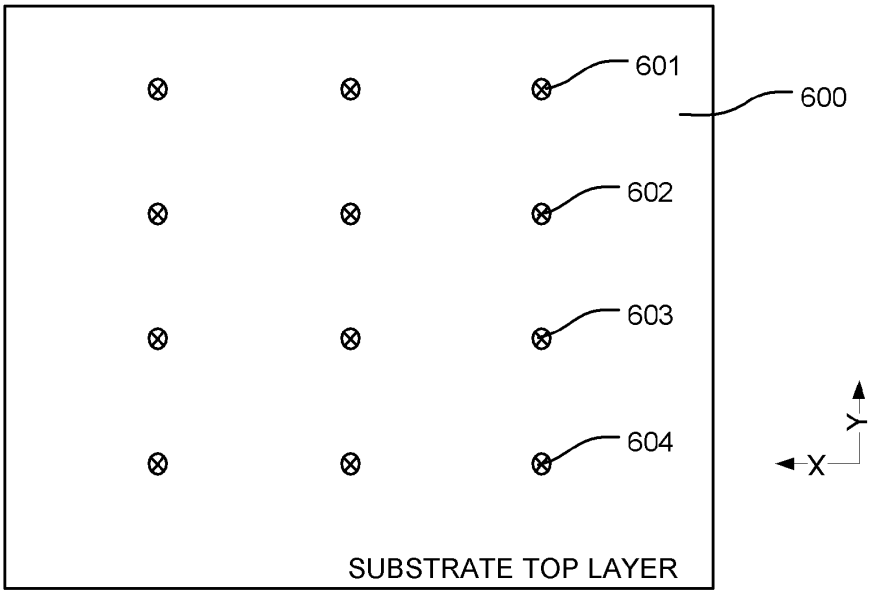

SOURCE PLANE
147
Z
Y
530
133
133
V_{SC}
504
504-L
504-R
411
412
V_{WL-412}
V_{WL-411}
429
439
502
502-L
502-R
415
414
V_{WL-415}
V_{WL-414}
431
437
500
500-L
500-R
418
417
V_{WL-418}
V_{WL-417}
433
435
V_{SEL}
137
137
497
520
136
V_{BL}

Fig. 5

Fig. 6


601
600
602
603
604
Y
X
SUBSTRATE TOP LAYER

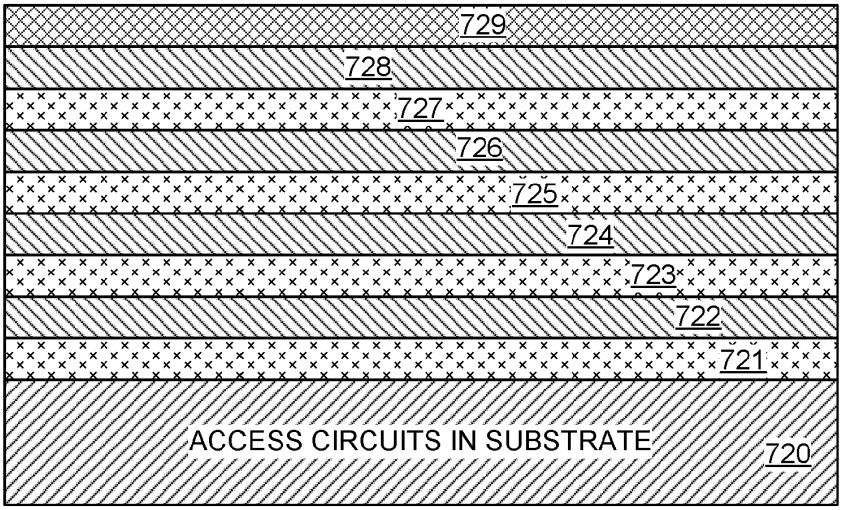

729
728
727
726
725
724
723
722
721
ACCESS CIRCUITS IN SUBSTRATE
720

Fig. 7

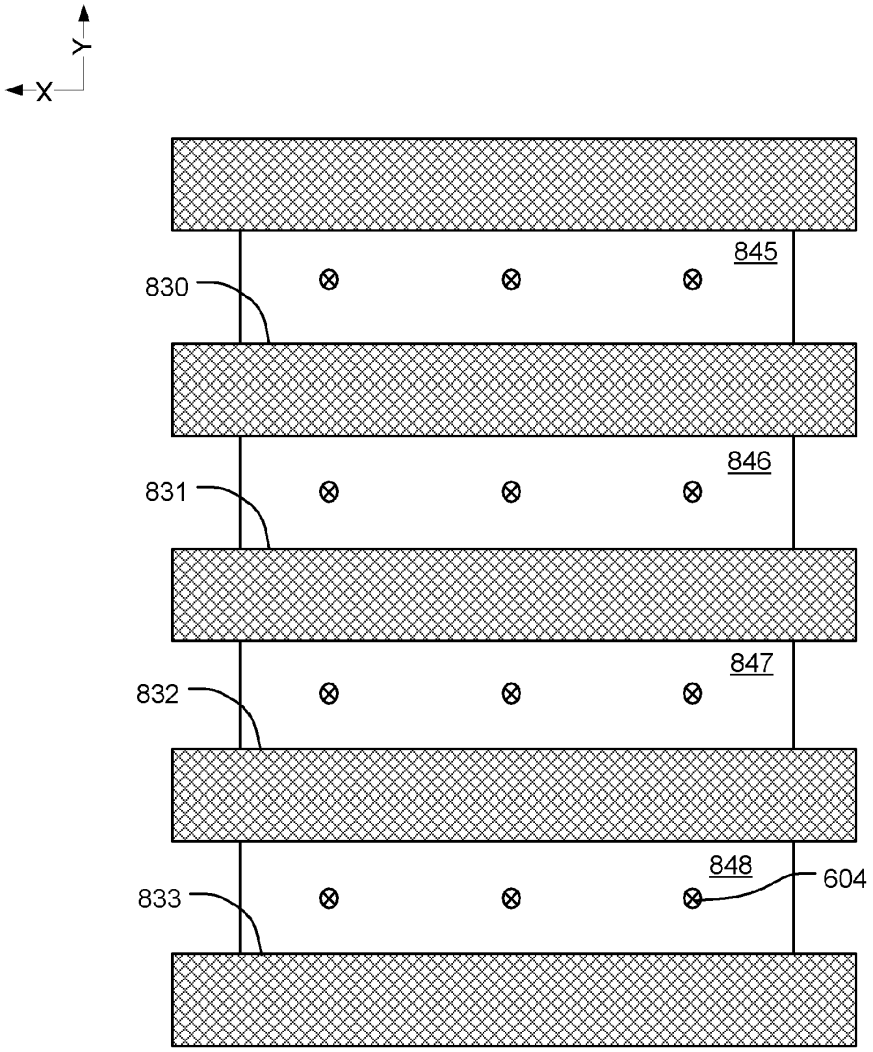


Fig. 8

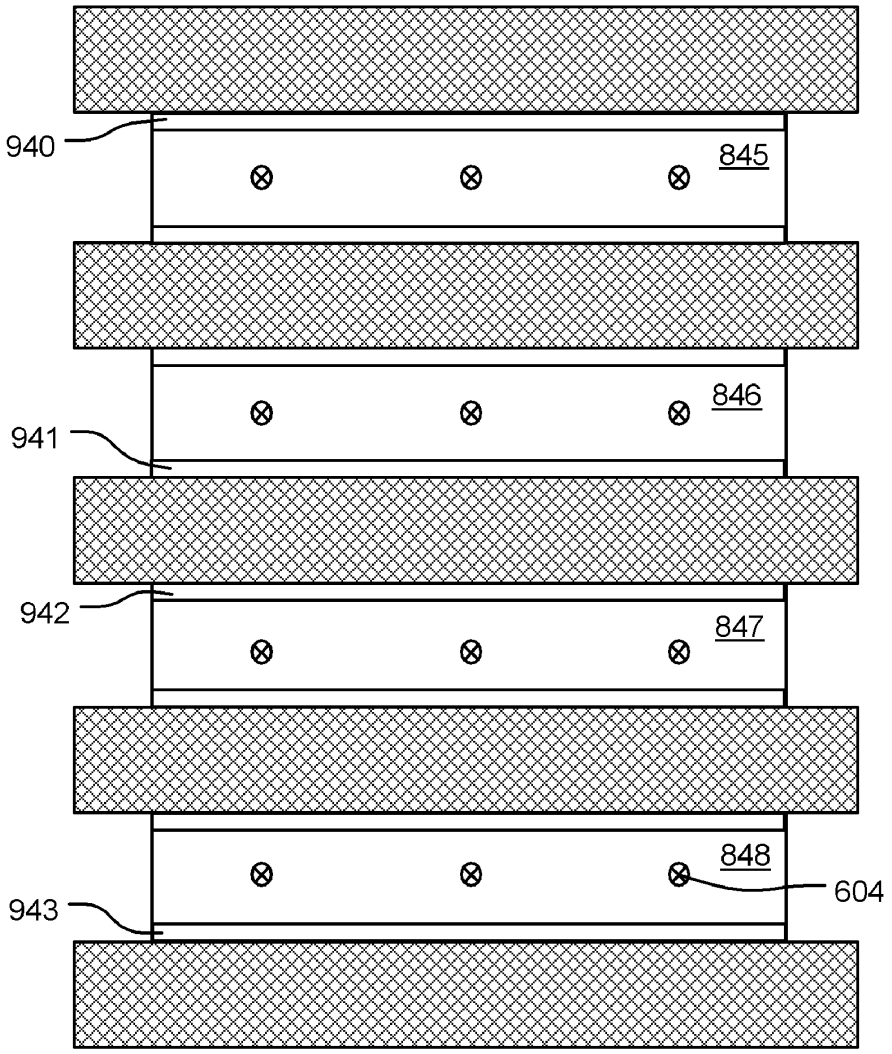


Fig. 9

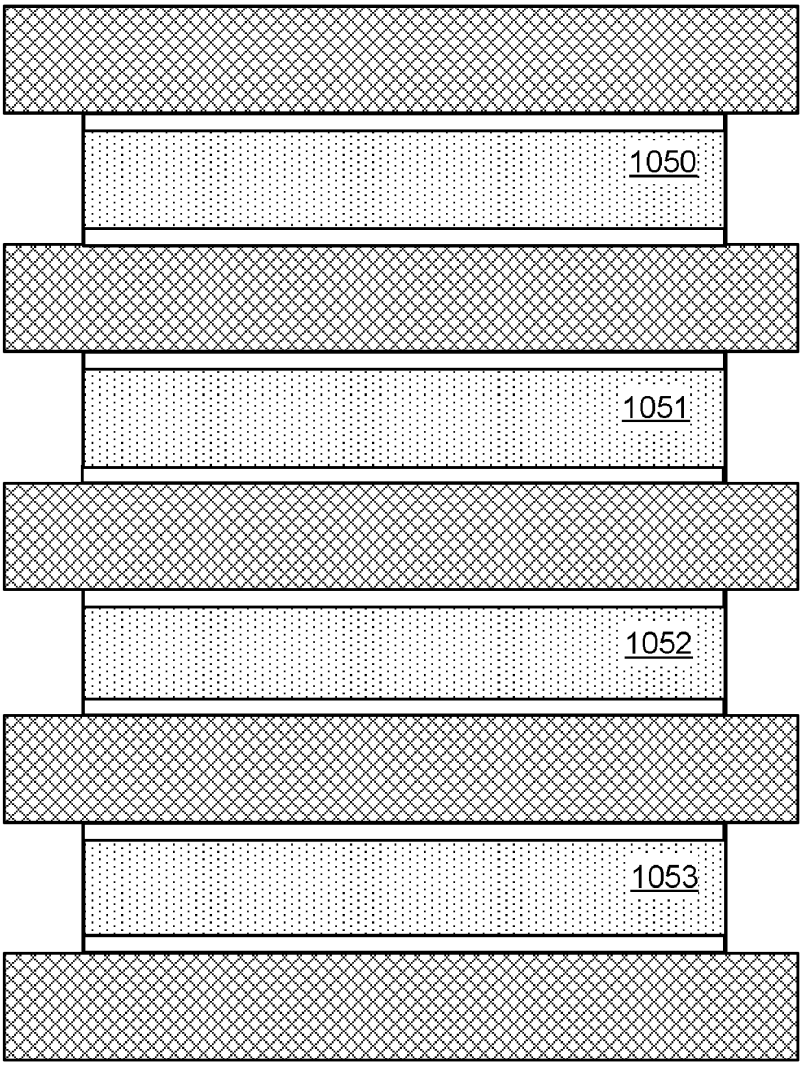


Fig. 10

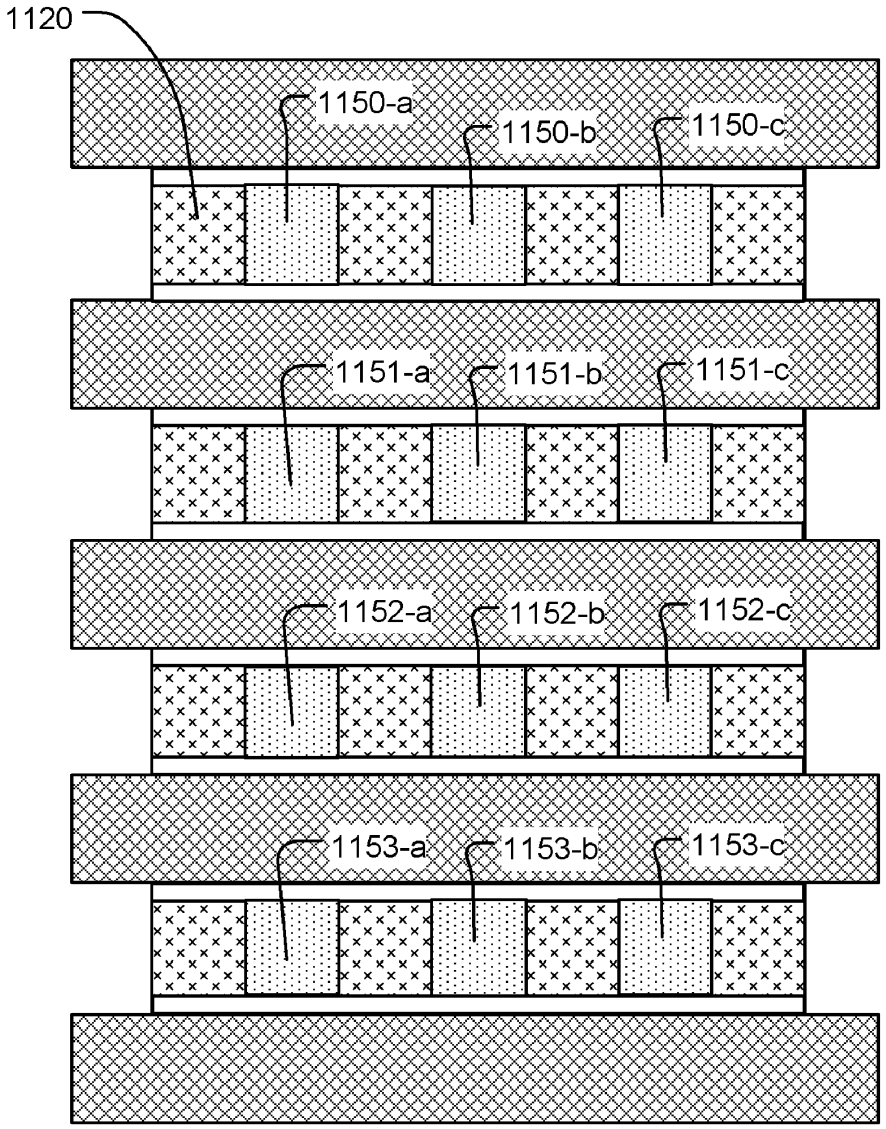

1120
1150-a
1150-b
1150-c
1151-a
1151-b
1151-c
1152-a
1152-b
1152-c
1153-a
1153-b
1153-c

Fig. 11

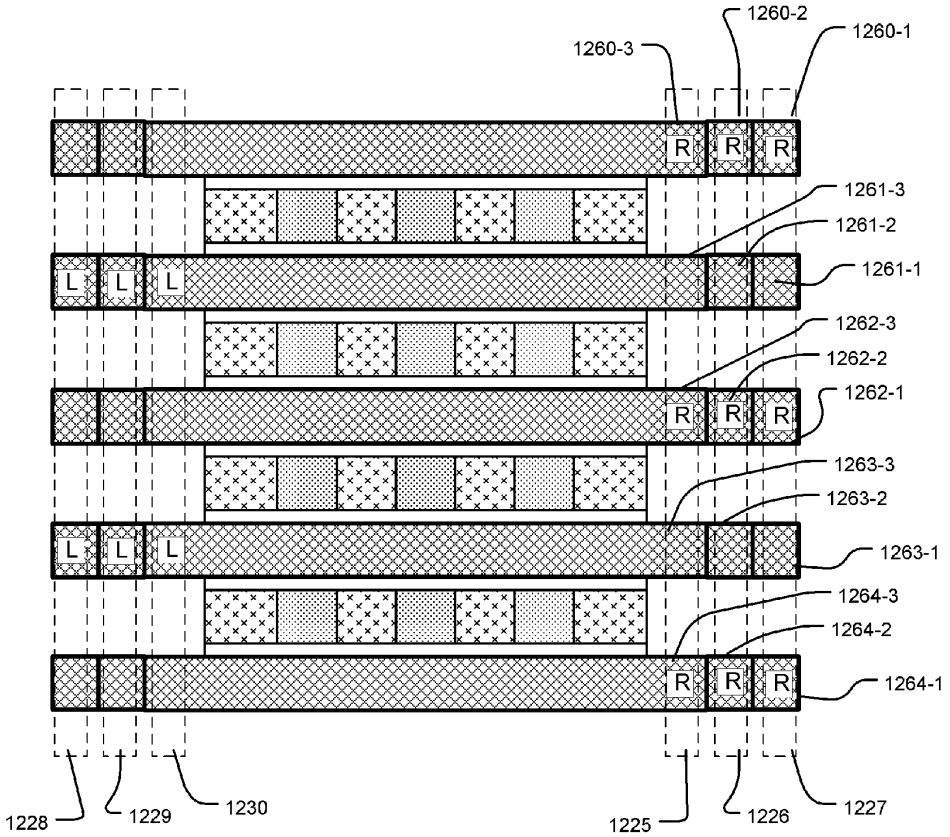

1260-2
1260-1
1260-3
R
R
R
1261-3
1261-2
L
L
L
1261-1
1262-3
1262-2
R
R
R
1262-1
1263-3
1263-2
L
L
L
1263-1
1264-3
1264-2
R
R
R
1264-1
1228
1229
1230
1225
1226
1227

Fig. 12

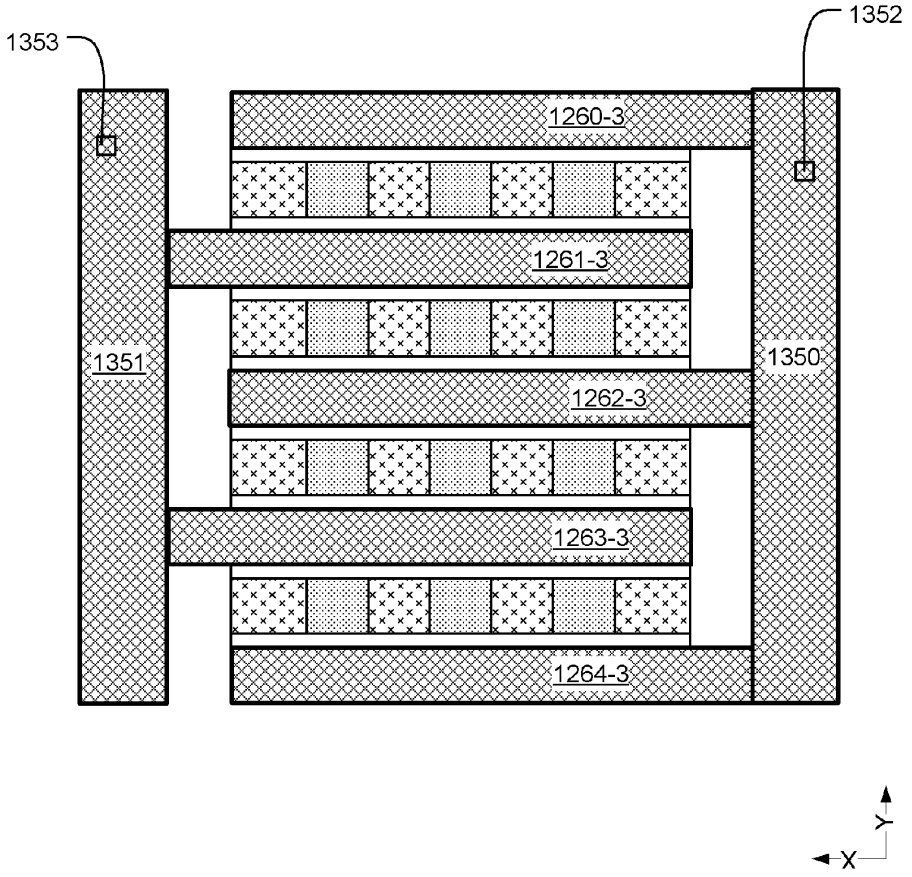


Fig. 13

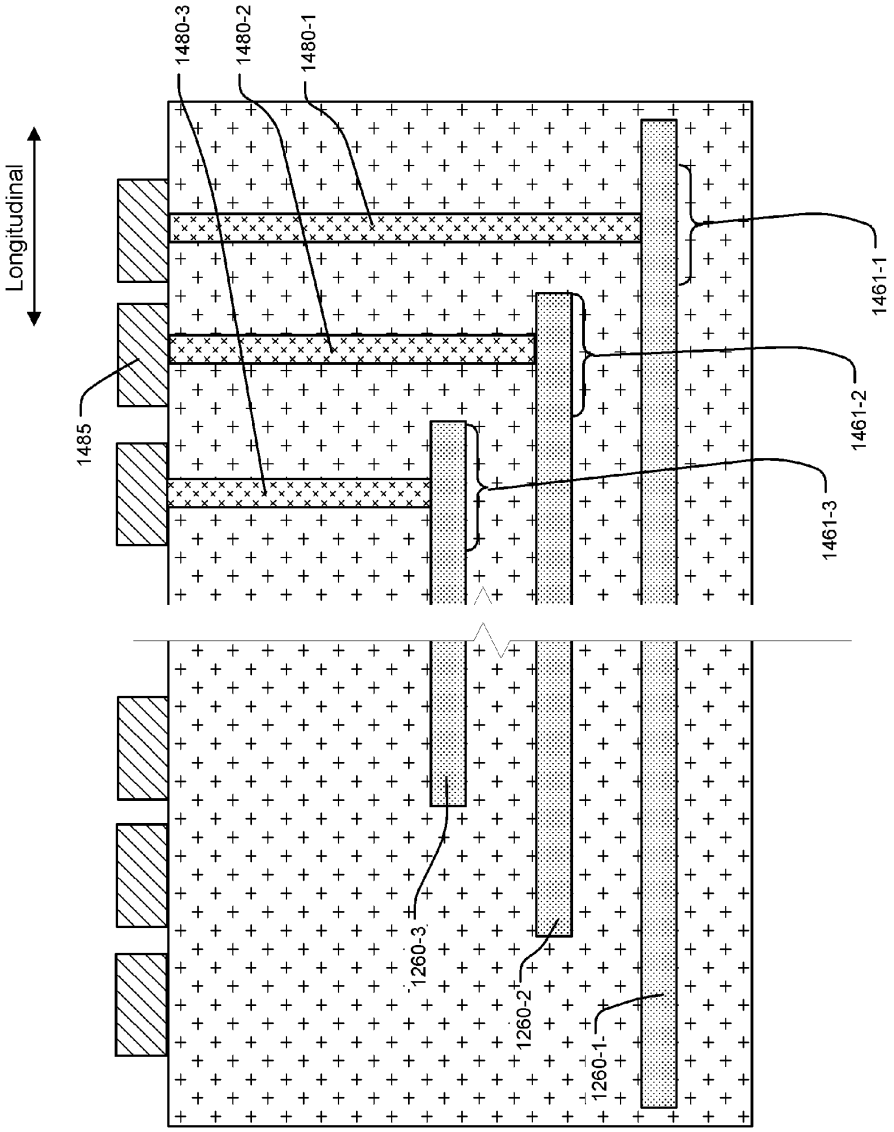

Longitudinal
1485
1480-3
1480-2
1480-1
1461-3
1461-2
1461-1
1260-3
1260-2
1260-1

Fig. 14

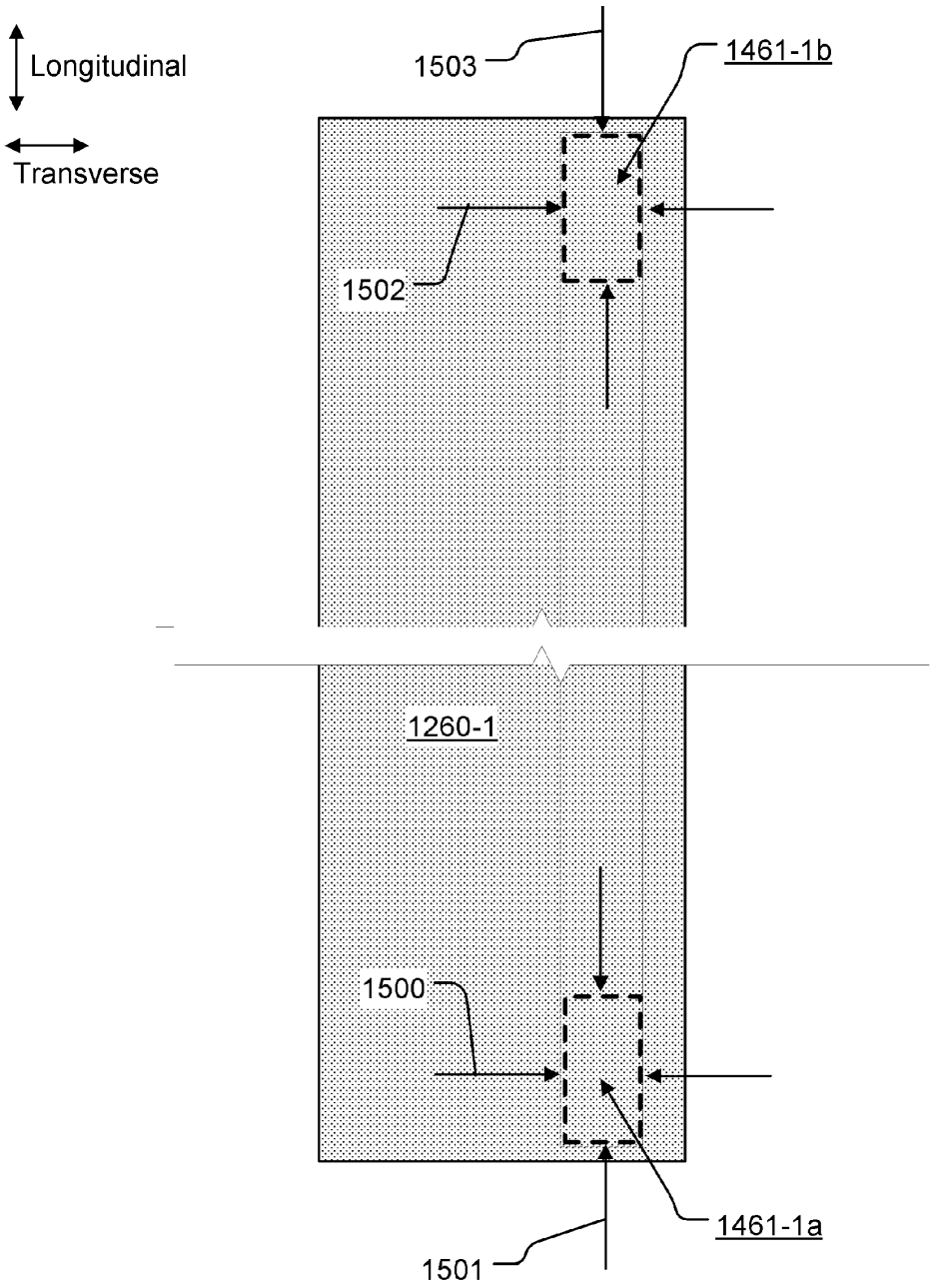

Longitudinal
Transverse
1503
1461-1b
1502
1260-1
1500
1461-1a
1501

Fig. 15A

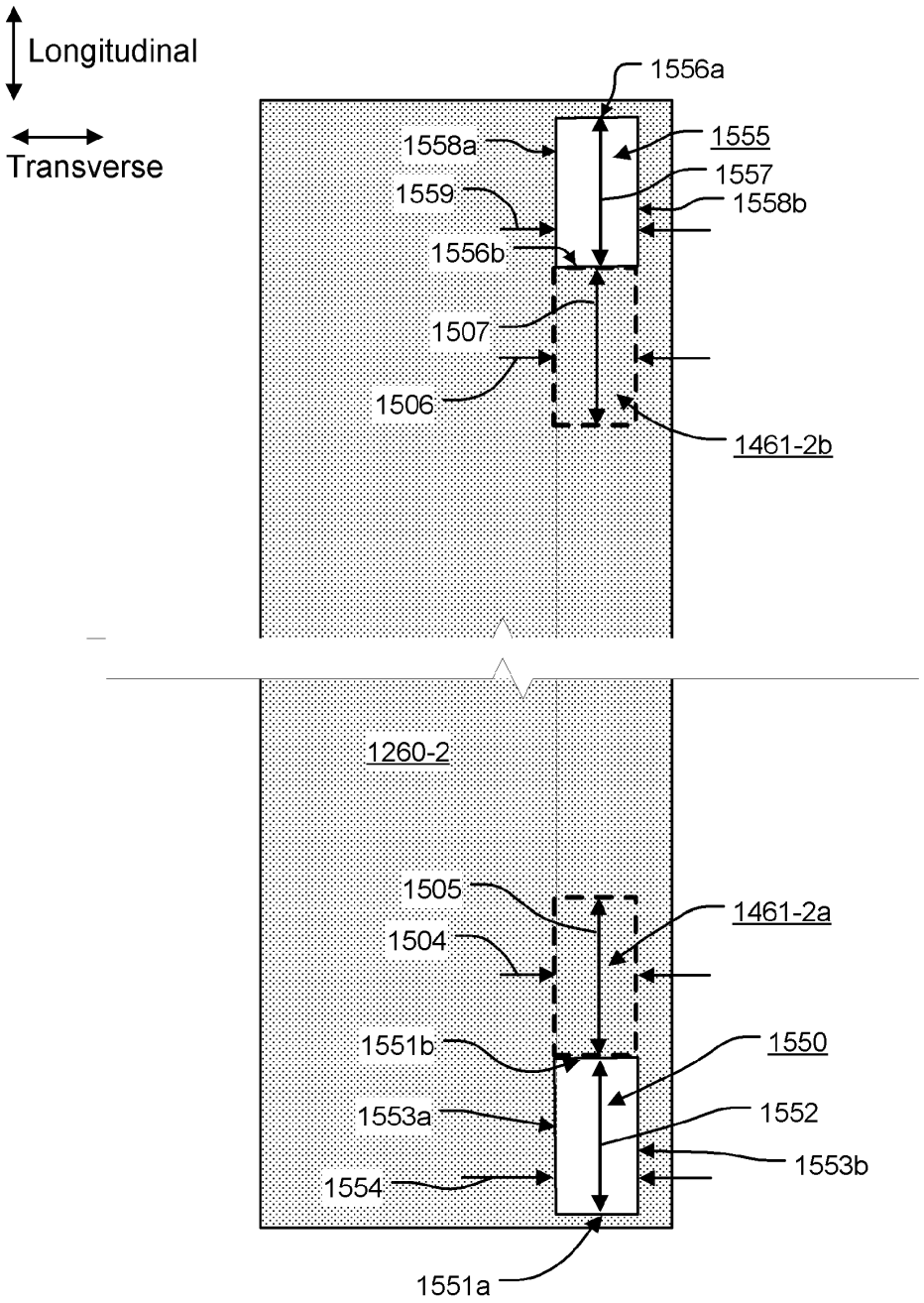

Longitudinal
Transverse
1556a
1558a
1555
1557
1559
1558b
1556b
1507
1506
1461-2b
1260-2
1505
1461-2a
1504
1551b
1550
1553a
1552
1553b
1554
1551a

Fig. 15B

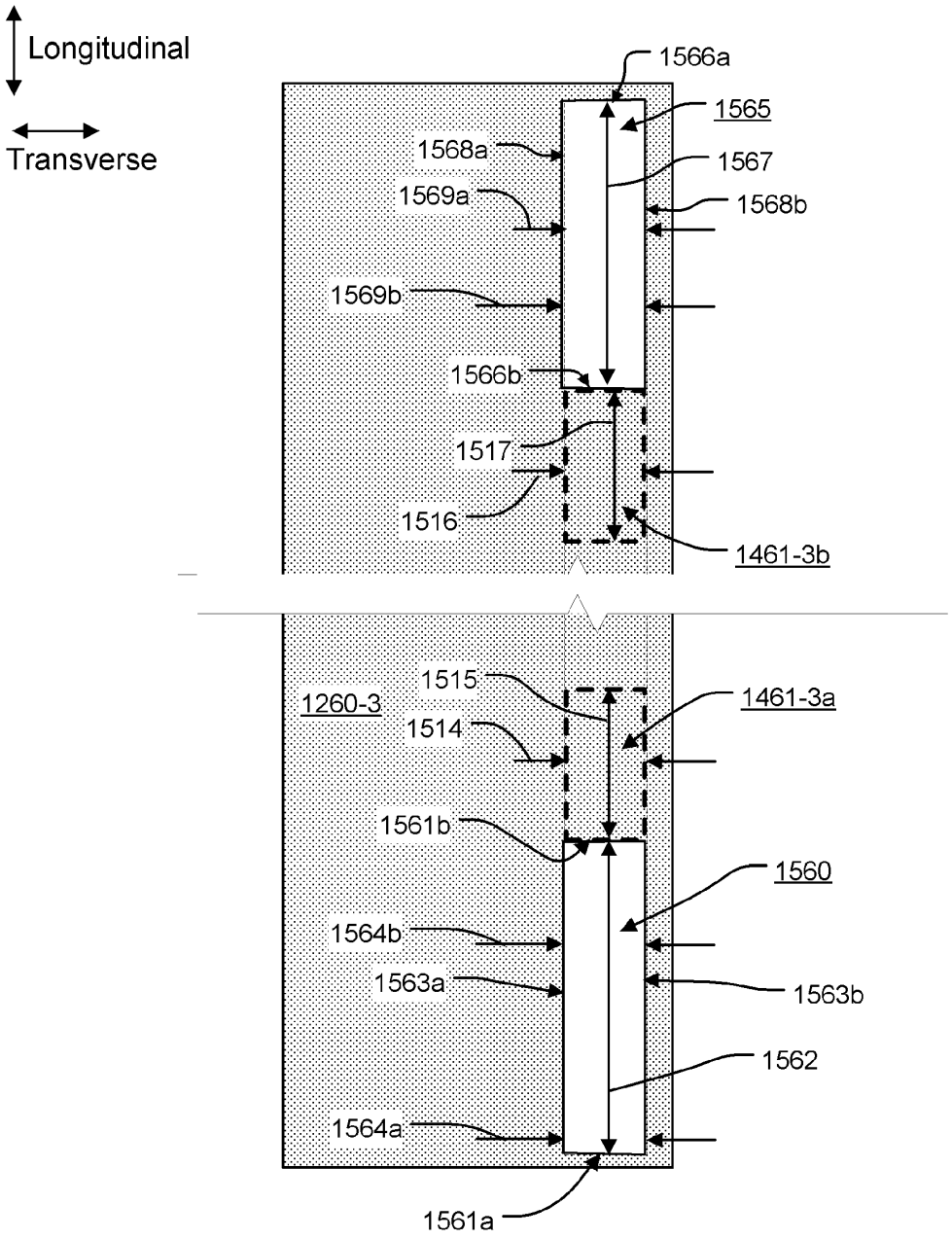

Longitudinal
Transverse
1566a
1565
1568a
1567
1569a
1568b
1569b
1566b
1517
1516
1461-3b
1515
1260-3
1461-3a
1514
1561b
1560
1564b
1563a
1563b
1562
1564a
1561a

Fig. 15C

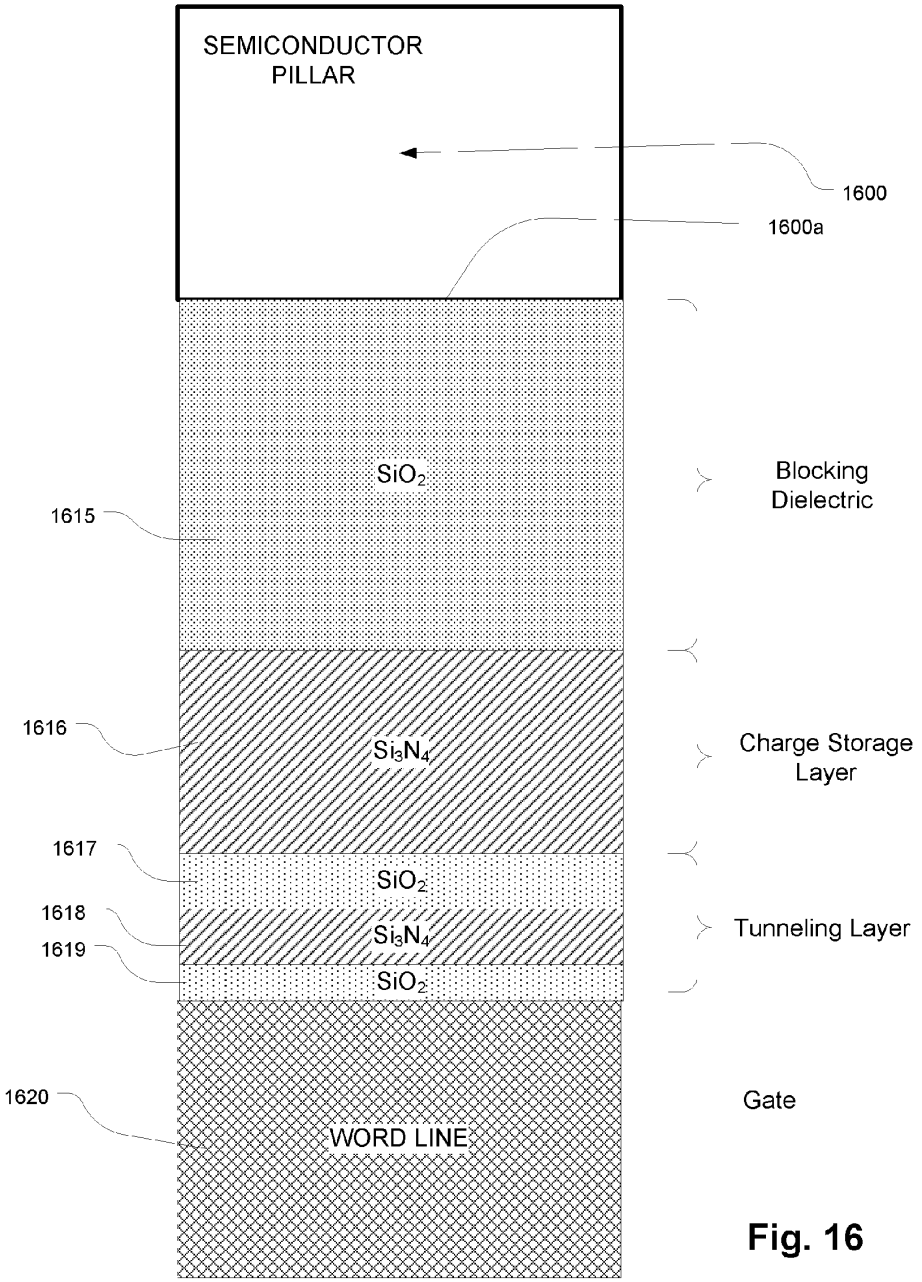

SEMICONDUCTOR
PILLAR
1600
1600a
SiO_2
Blocking
Dielectric
1615
1616
Si_3N_4
Charge Storage
Layer
1617
SiO_2
1618
Si_3N_4
Tunneling Layer
1619
SiO_2
1620
WORD LINE
Gate

Fig. 16

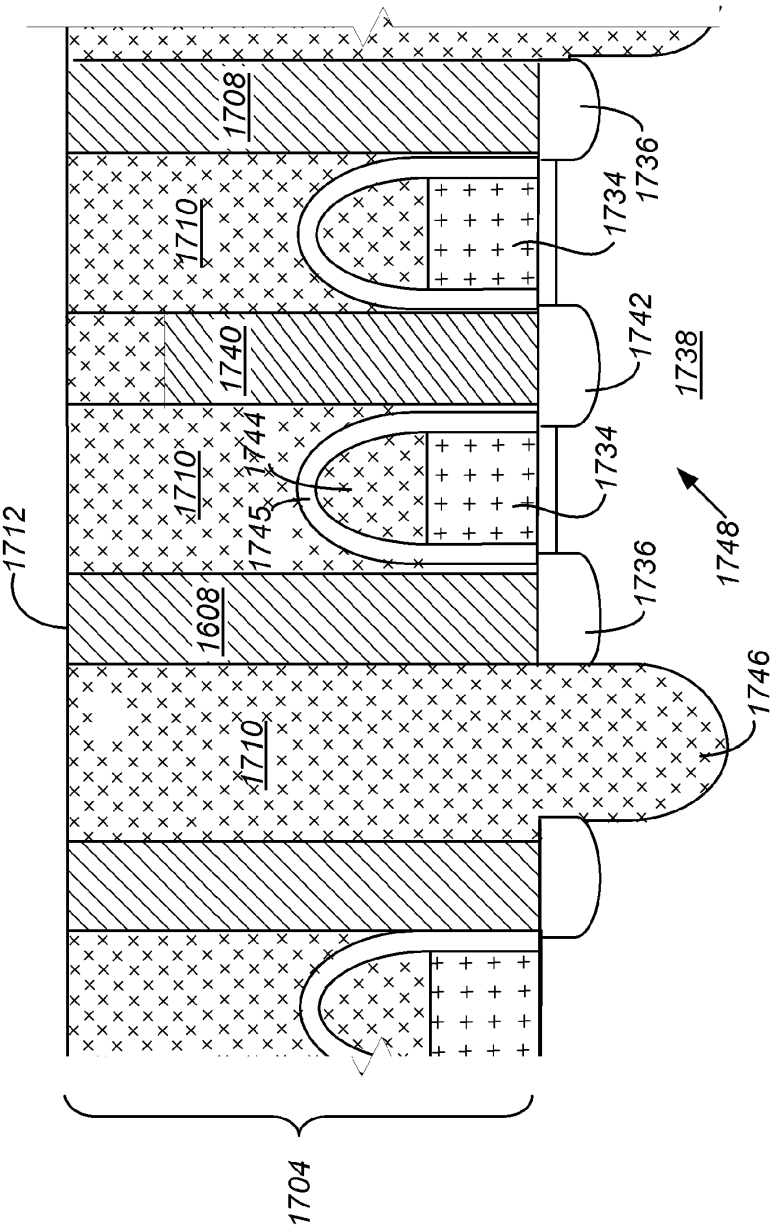

1708
1710
1740
1710
1744
1745
1608
1710
1712
1736
1734
1742
1738
1734
1736
1748
1746
1704

Fig. 17

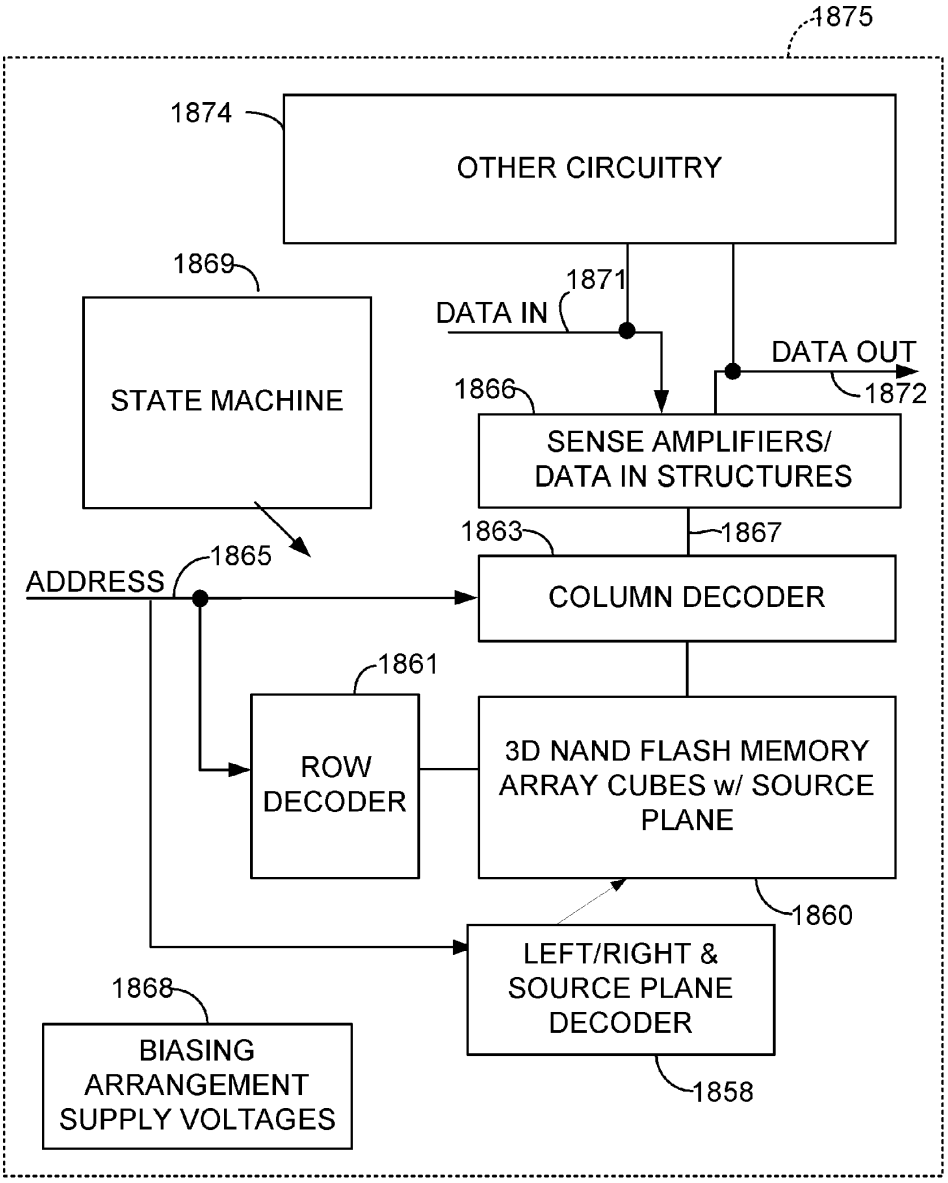
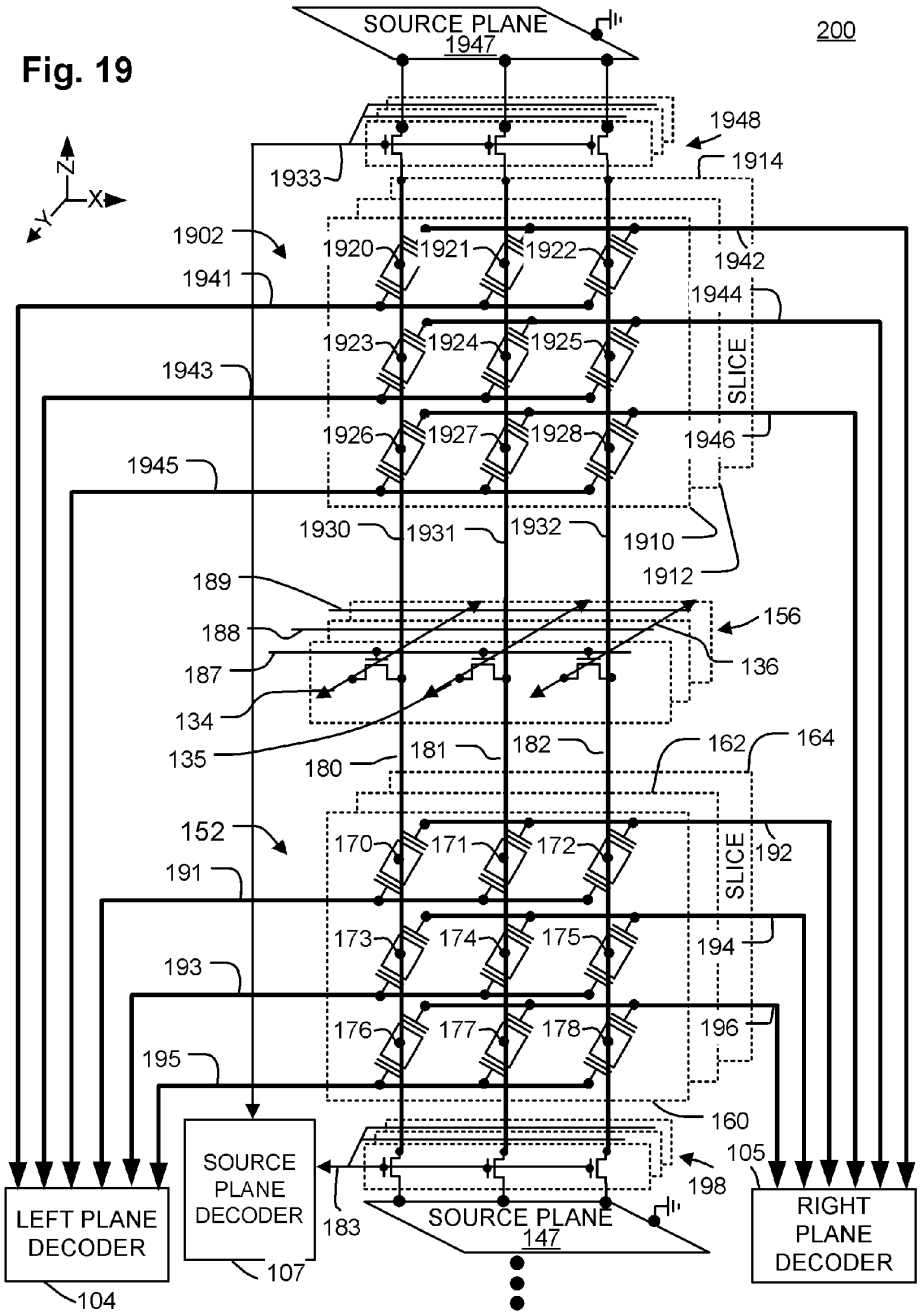

1875
1874
OTHER CIRCUITRY
1869
STATE MACHINE
1871
DATA IN
DATA OUT
1872
1866
SENSE AMPLIFIERS/
DATA IN STRUCTURES
1863
1867
1865
ADDRESS
COLUMN DECODER
1861
ROW
DECODER
3D NAND FLASH MEMORY
ARRAY CUBES w/ SOURCE
PLANE
1860
LEFT/RIGHT &
SOURCE PLANE
DECODER
1868
BIASING
ARRANGEMENT
SUPPLY VOLTAGES
1858

Fig. 18

Fig. 19


SOURCE PLANE
1947
200
1948
1914
1933
1902
1920
1921
1922
1942
1941
1944
1923
1924
1925
SLICE
1943
1926
1927
1928
1946
1945
1930
1931
1932
1910
1912
189
156
188
136
187
134
135
180
181
182
162
164
152
170
171
172
SLICE
192
191
173
174
175
194
193
176
177
178
196
195
160
105
198
SOURCE PLANE DECODER
183
SOURCE PLANE
147
LEFT PLANE DECODER
RIGHT PLANE DECODER
107
104
Z
X
Y

3D TWO-BIT-PER-CELL NAND FLASH MEMORY

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] The present invention relates to high density memory devices, and particularly to memory devices in which multiple planes of memory cells are arranged to provide a three-dimensional 3D array.

[0003] 2. Description of Related Art

[0004] As critical dimensions of devices in integrated circuits shrink to the limits of common memory cell technologies, designers have been looking to techniques for stacking multiple planes of memory cells to achieve greater storage capacity, and to achieve lower costs per bit. For example, thin film transistor techniques are applied to charge trapping memory technologies in Lai, et al., "A Multi-Layer Stackable Thin-Film Transistor (TFT) NAND-Type Flash Memory", IEEE Int'l Electron Devices Meeting, 11-13 Dec. 2006; and in Jung et al., "Three Dimensionally Stacked NAND Flash Memory Technology Using Stacking Single Crystal Si Layers on ILD and TANOS Structure for Beyond 30 nm Node", IEEE Int'l Electron Devices Meeting, 11-13 Dec. 2006.

[0005] Another structure that provides vertical NAND cells in a charge trapping memory technology is described in Tanaka et al., "Bit Cost Scalable Technology with Punch and Plug Process for Ultra High Density Flash Memory", 2007 Symposium on VLSI Technology Digest of Technical Papers; 12-14 Jun. 2007, pages: 14-15. Also, Jin et al., U.S. Patent Application Publication No. US 2009/0184360, describe a vertical NAND string structure using SONOS charge storage technology to implement data storage cites on two sides of each vertical channel structure.

[0006] However, there is a limit on the number of layers of cells that can be implemented using vertical NAND string structures, which is imposed by the impedance of the vertical string of cells in each NAND string. Thus, the number of layers that can be practically implemented in this structure is relatively small.

[0007] As the need for higher and higher memory capacity in integrated circuit memory devices continues to increase, it is desirable to provide a structure for three-dimensional integrated circuit memory with a low manufacturing cost, including reliable, very small memory elements that can be erased and programmed.

SUMMARY OF THE INVENTION

[0008] A memory device comprises a plurality of bit lines over a substrate; a plurality of memory cells over the plurality of bit lines, wherein at least one memory cell of the plurality of memory cells is over another one memory cell of the plurality of memory cells; and a source plane over the plurality of memory cells. This structure is suitable for use in forming a memory device that includes a first set of select transistors; a first 3D array of memory cells coupled to the first set of select transistors; a common source coupled to the first 3D array of memory cells; a second 3D array of memory cells coupled to the common source, wherein the common source is between the first and second 3D array of memory cells; and a second set of select transistors coupled to the second 3D array of memory cells.

[0009] A 3D memory device is described which includes bottom and top memory cubes having respective arrays of vertical NAND string structures. A common source plane comprising a layer of conductive material is between the top and bottom memory cubes. The source plane is supplied a bias voltage such as ground, and is selectively coupled to an end of the vertical NAND string structures of the bottom and top memory cubes. Memory cells in a particular memory cube are read using current through the particular vertical NAND string between the source plane and a corresponding bit line coupled to another end of the particular vertical NAND string. The vertical NAND strings in the bottom and top memory cubes provide a large number of stackable layers of cells, resulting a high density NAND architecture flash memory device which overcomes the prior art limitation on the number of stackable layers of cells imposed by the impedance of a vertical string of cells of a single NAND string.

[0010] A 3D array of memory cells described herein includes bottom and top memory cubes including respective arrays of semiconductor pillars. A given semiconductor pillar in each array of semiconductor pillars includes a plurality of dielectric charge trapping structures in a NAND string along a side of the given semiconductor pillar. Word lines are orthogonal to the semiconductor pillars and adjacent the charge trapping structures of the semiconductor pillars. A source plane is between the bottom and top memory cubes. The source plane is selectively coupled to top ends of each of the semiconductor pillars of the bottom memory cube by a bottom set of source plane select transistors, and selectively coupled to bottom ends of each of the semiconductor pillars of the top memory cube by a top set of source plane select transistors. The 3D array further includes a bottom set of pillar select transistors to selectively couple bottom ends of the semiconductor pillars of the bottom memory cube to corresponding bit lines in a plurality of bottom bit lines. A top set of pillar select transistors selectively couple top ends of the semiconductor pillars of the top memory cube to corresponding bit lines in a plurality of top bit lines.

[0011] Memory cubes described herein are based on a plurality of two-cell unit structures. The two-cell unit structures of a given semiconductor pillar include a first or left dielectric charge trapping structure along a first side of the given semiconductor pillar, a second or right dielectric charge trapping structure along a second side of the given semiconductor pillar. The first charge trapping structures in the two-cell unit structures of the given semiconductor pillar are arranged in a first NAND string, and the second charge trapping structures in the two-cell unit structures of the given semiconductor pillar are arranged in a second NAND string. Multiple levels of cells are implemented on the array of semiconductor pillars using word line structures on each level, so that cells are formed at the cross-points of the sides of the semiconductor pillars and word line structures in each level, with dielectric charge trapping structures in between, providing the 3D array of memory cells. Examples of the dielectric charge trapping structures include SONOS-type ONO structures and BE-SONOS-type ONONO structure.

[0012] The word line structures include a first set of word lines, coupled in common to a first driver on for example the left side of the structure, arranged adjacent the dielectric charge trapping structures between alternating pairs of rows of semiconductor pillars, and a second set of word lines interleaved with the first set, and coupled in common to a second driver on for example the right side of the structure, arranged adjacent the dielectric charge trapping structures between interleaved alternating pairs of rows of semiconduc-

tor pillars. This results in word lines that provide gates adjacent the dielectric charge trapping structures on the first and second sides of the semiconductor pillars, and thereby provides two independently addressable memory cells in each level on each semiconductor pillar. A controller is coupled to the array, and arranged to program and erase selected memory cells two-cell unit structures by biasing corresponding semiconductor pillars and one of the first or second sets of word lines to induce gate side Fowler-Nordheim (FN) programming and gate side Fowler-Nordheim erasing operations.

[0013] Methods for manufacturing a 3D array of memory cells including bottom and top memory cubes and a source plane between them as described herein require few lithographic steps, and are therefore practical and low-cost compared to other 3D memory processes.

[0014] Other aspects and advantages of the present invention can be seen on review of the drawings, the detailed description and the claims, which follow.

BRIEF DESCRIPTION OF THE DRAWINGS

[0015] FIG. **1** is a schematic diagram of a 3D memory device including a common source plane arranged between a bottom memory cube of memory cells and a top memory cube of memory cells.

[0016] FIG. **2** is a schematic diagrams of the 3D memory device, showing levels which lie in the X-Y planes of the bottom memory cube.

[0017] FIG. **3** shows the structure of a 2-bit unit cell.

[0018] FIG. **4** is a perspective drawing of a portion of a 3D memory device described herein.

[0019] FIG. **5** is a cross-sectional view in the Y-Z plane of the structure in FIG. **4**.

[0020] FIGS. **6-11** show a sequence of stages of a manufacturing process for making the 3D memory structure described herein.

[0021] FIGS. **12**, **13**, **14** and **15**A-**15**C illustrate 3D interconnect structures for coupling the word line levels to decoding circuitry.

[0022] FIG. **16** illustrates a BE-SONOS charge storage structure utilized in an embodiment of the 3D memory structure described herein.

[0023] FIG. **17** illustrates an implementation for an array of access devices suitable for use as the bottom set of pillar select transistors.

[0024] FIG. **18** is a simplified block diagram of an integrated circuit including a 3D NAND-type charge trapping memory array as described herein.

[0025] FIG. **19** is a schematic diagram of a second embodiment of a 3D memory device as described herein.

DETAILED DESCRIPTION

[0026] A detailed description of embodiments of the present invention is provided with reference to the FIGS. **1-19**.

[0027] FIG. **1** is a schematic diagram of a 3D memory device **100** including a common source plane **147** arranged between a bottom memory cube **102** of memory cells and a top memory cube **152** of memory cells.

[0028] The schematic diagram of FIG. **1** shows "slices" **110**, **112**, **114** which lie in X-Z planes of the bottom memory cube **102**. In the illustrated schematic, there are nine two-cell unit structures **120-128** in slice **110**, each two-cell unit structure **120-128** having two memory cells including having separate charge storage structures and left and right gates. Embodiments can include many two-cell unit structures per slice.

[0029] The bottom memory cube **102** includes an array of memory cells arranged for NAND-type decoding, using a left plane decoder **104**, right plane decoder **105**, a bottom set of pillar select transistors **106**, and source plane decoder **107**. The left and right side memory cells of the two-cell unit structures **120-128** in a Z-direction column (e.g. **120**, **123**, **126**) are arranged in left and right NAND strings respectively. More details of the two-cell unit structure and the NAND strings are provided below.

[0030] Bottom ends of the array of semiconductor pillars of the two-cell unit structures **120-128** in a Z-direction column (e.g. **120**, **123**, **126**) are coupled via a corresponding semiconductor pillar **130**, **131**, **132** to a corresponding access device in the bottom set of pillar select transistors **106**, implemented for example in the integrated circuit substrate beneath the structure.

[0031] The access devices in the bottom set of pillar select transistors **106** selectively couple a Z-direction column of the two-cell unit structures **120-128** to a corresponding bit line in a plurality of bit lines **134**, **135**, **136** extending in the Y-direction. The plurality of bit lines **134**, **135**, **136** are coupled to a column decoder (not shown).

[0032] The gates of the bottom set of pillar select transistors **106** are coupled to select lines **137**, **138**, **139** extending in the X-direction. The select lines **137**, **138**, **139** are coupled to a row decoder (not shown).

[0033] The left gates on the two-cell unit structures in a particular level (e.g. two unit-cell structures **120**, **121**, **122**) in all of the slices **110**, **112**, **114** are coupled via a word line element **141** to a driver selected by left plane decoder **104**. Likewise, the right gates on the two-cell unit structures in a particular level (e.g. **120**, **121**, **122**) in all of the slices **110**, **112**, **114** are coupled via word line element **142**. The left gates and right gates on the level including two-cell unit structures **123**, **124**, **125** are coupled via the word line element **143** to the left plane decoder **104** and via word line element **144** to the right plane decoder **105**. The left and right gates on the level including two-cell unit structures **126**, **127**, **128** are coupled via the word line element **145** to the left plane decoder **104** and via word line element **146** to the right plane decoder **105**.

[0034] Top ends of the array of semiconductor pillars of the two-cell unit structures **120-128** in a Z-direction column (e.g. **120**, **123**, **126**) are coupled via a corresponding select transistor in a bottom set of source plane select transistors **148** to the source plane **147**. The source plane **147** is a layer of conductive material between the bottom and top memory cubes **102**, **152**. In the illustrated example, the source plane **147** is coupled to a ground terminal. Alternatively, the source plane **147** may be coupled to a voltage source other than ground.

[0035] Common conductive plane **133**, acting as the gate of each of the access devices in the bottom set of source plane select transistors **148**, is coupled to a driver selected by source plane decoder **107**.

[0036] The top memory cube **152** overlies and is selectively coupled to the source plane **147** by a top set of source plane select transistors **198**. The arrangement of the memory cells in the top memory cube **152** is similar to that of the bottom memory cube **102**. The top memory cube **152** includes "slices" **160**, **162**, **164** which lie in the X-Z planes of the top memory cube **152**. In the illustrated schematic, there are nine

two-cell unit structures **170-178** in slice **160**, each unit structure **170-178** having two memory cells including having separate charge storage structures and left and right gates. Embodiments can include many two-cell structures per slice.

[0037] The memory cells in the top memory cube **152** are arranged for NAND-type decoding, using the left plane decoder **104**, right plane decoder **105**, the top set of pillar select transistors **156**, and source plane decoder **107**. As discussed in more detail below, the left and right side memory cells of the two-unit cell structures **170-178** in a Z-direction column (e.g. **170**, **173**, **176**) are arranged in left and right NAND strings respectively.

[0038] Top ends of the semiconductor pillars of the two-cell unit structures **170-178** in a Z-direction column (e.g. **170**, **173**, **176**) are coupled via a corresponding semiconductor pillar **180**, **181**, **182** to a corresponding access device in the top set of pillar select transistors **156**, implemented for example in the integrated circuit above the top memory cube **152**.

[0039] The access devices in the top set of pillar select transistors **156** selectively couple the semiconductor pillars of the two-cell unit structures **170-178** to a corresponding bit line in a plurality of bit lines **184**, **185**, **186** extending in the Y-direction and coupled to the column decoder (not shown).

[0040] The gates of the top set of pillar select transistors **156** are coupled to select lines **187**, **188**, **189** extending in the X-direction. The select lines **187**, **188**, **189** are coupled to a row decoder (not shown).

[0041] The left gates on the two-cell unit structures in a particular level (e.g. **170**, **171**, **172**) in all of the slices **160**, **162**, **164** are coupled via a word line element **191**, **193**, **195** to the left plane decoder **104**. Likewise, the right gates on the two-cell unit structures in a particular level (e.g. **170**, **171**, **172**) in all of the slices **160**, **162**, **164** are coupled to the right plane decoder **105** via word line element **192**, **194**, **196**.

[0042] Bottom ends of the semiconductor pillars of the two-cell unit structures **120-128** in a Z-direction column (e.g. **120**, **123**, **126**) are coupled via a corresponding select transistor in the top set of source plane select transistors **198** to the source plane **147**.

[0043] Common conductive plane **183**, acting as the gate of each of the access devices in the bottom set of source plane select transistors **148**, is coupled to a driver selected by source plane decoder **107**.

[0044] The two-cell unit structure **123** in the bottom memory cube **102** is representative of the two-cell unit structures in the 3D memory device **100**. The two-cell unit structure includes a left side memory cell **123**-L and a right side memory cell **123**-R.

[0045] In a read operation of a selected one of the left and right side memory cells **123**-L, **123**-R, a current path is established between the bit line **134** and the source plane **147** via the particular NAND string to which the selected memory cell belongs.

[0046] For example, in a read operation of the left side memory cell **123**-L, voltages are applied to the word line elements **141** and **145** sufficient to turn on the left side memory cells of the two-cell unit structures **120** and **126**. An X-direction row on the select line **137** is selected using a row (X) decoder (not shown), and a Y-direction column on the bit line **134** using a column (Y) decoder (not shown). Voltages are applied to the common conductive planes **133** and **183** using the source plane decoder **107** to turn on the bottom set of source plane select transistors **148**, and turn off the top set of source plane select transistors **198** respectively. As a result the source plane **147** is coupled to the top ends of the array of semiconductor pillars of the bottom memory cube **102**.

[0047] A read voltage is applied to the word line element **143** coupled to the left side memory cell **123**-L in the unit-cell structure **123**. The current on the bit line **134** and through the left NAND string of the left side memory cells of the two-unit cell structures **120**, **123**, **126** indicates the data value stored in the left side memory cell **123**-L in the unit-cell structure **123**.

[0048] Three-dimensional stacking as described herein using top and bottom memory cubes **152**, **102** and the source plane **147** between them is an efficient way to reduce the cost per bit for semiconductor memory by providing a large number of stackable layers of cells, which overcome the limitation imposed by the impedance of a vertical string of cells in a given NAND string.

[0049] Two memory cubes **102**, **152** are illustrated in FIG. **1**. The 3D memory device **100** can also include more than two memory cubes. For example, the structure of a pair of memory cubes separated by a source plane can be repeated a number of times along the Z-direction. In some embodiments, such as explained below with reference to FIG. **19**, an additional memory cube overlying the top memory cube **162** shares the top set of pillar select transistors **156** with the top memory cube **152**, including the bit lines **134**, **135**, **136** and select lines **187**, **188**, **189**.

[0050] FIG. **2** is a schematic diagram of the 3D memory device **100**, showing "levels" **266**, **267**, **268** which lie in the X-Y planes of the bottom memory cube **102**. The levels **266**, **267**, **268** lie in the X-Y planes, and thus are perpendicular to the "slices" (See, FIG. **1**, reference numerals **110**,**112**,**114**) which lie in the X-Z planes. The left and right plane decoders **104**, **105** and the source plane **147** are illustrated in the figure. Each level in the schematic includes nine two-cell unit structures. Embodiments can include many cells per level. The front row of two-unit structures in level **266** includes two-unit cell structures **120**, **121** and **122**, corresponding to the top row in the slice **110** shown in FIG. **1**. The two-cell unit structures **120-122** and **220-225** define a 3-by-3, X-Y arrangement of unit structures on the level **266**. As shown in FIG. **2**, the left word line element **141** is arranged to connect to the left side gates between alternating pairs of rows using a forked word line element **141**-L. Likewise, the right word line element **142** is interleaved with the left word line element **141**, and arranged to connect to the right side gates between other alternating pairs of rows using forked word line element **142**-R. As described below, the left and right side word lines may be separated from one another in each level, and connected by vias to overlying conductors (rather than forked and connected together in each level as shown).

[0051] The two-cell unit structure is shown in FIG. **3**. Reference number **120** which is utilized in FIG. **1** and FIG. **2** representing the unit structure can be represented by the structure shown, including word line element **141**-L, word line element **142**-R, and the semiconductor pillar **130**. Dielectric insulators **310**, **320** separate the pillars. Dielectric charge trapping structures **330**, **340** lie on opposing sides of the semiconductor pillar **130** and the corresponding gates provided by the word line elements, **142**-R or **141**-L. Thus, two memory cells are provided by this structure, including CELL **1** and CELL **2** as labeled in the drawing, each cell including a charge trapping element and a gate.

[0052] The charge trapping structures **330**, **340** include a tunneling layer in contact with the word line material, the

charge trapping layer over the tunneling layer, and a blocking layer over the charge trapping layer, such as used in typical charge trapping memory devices. For example, the tunneling layer may comprise a layer of silicon dioxide or silicon oxide nitride, a charge trapping layer may comprise a layer of silicon nitride or other charge trapping material, and the blocking layer may comprise a layer of silicon dioxide or a high-k material typical in SONOS type memory device. In alternatives, as explained below with reference to FIG. **16**, a bandgap engineered charge trapping structure can be utilized.

[0053] Bias voltages applied to the unit structures include the right word line voltage V_{WL-R}, the left word line voltage V_{WL-L}, and the pillar voltage V_B. Self boosting caused by capacitive coupling of voltages from word lines of unselected, floating semiconductor pillars helps prevent program disturb conditions. The voltages applied depend upon the operation to be performed (e.g. read, program, and erase), and will vary from embodiment to embodiment. The read, program and erase operations of a selected memory cell in the two-unit cell structure are discussed in more detail below.

[0054] FIG. **4** shows a portion of the 3D memory device **100** including the bottom memory cube **102** of memory cells, common conductive plane **133** for the bottom set of source plane select transistors **148**, the source plane **147**, and bit lines **134**, **135**, **136** and select lines **137**, **138**, **139** for the bottom set of pillar select transistors **106** as described above with reference to FIGS. **1-3**. Three levels of word lines are illustrated, where a top level includes word lines **410**, **411** and **412** extending in the X-direction, a next lower level includes word lines **413**, **414** and **415**, and a bottom level includes word lines **416**, **417** and **418**. Charge storage structures **425-430** are formed on opposing sides of the word lines **410-412** on the top level. Charge storage structures **431-432** are formed on opposing sides of word line **415**, and charge storage structures **433-434** are formed on opposing sides of word line **418**. Similar charge storage structures are formed on the sides of the other word lines in the structure as well. The structure includes an array of semiconductor pillars, including pillars **493**, **495**, and **497** in the front of the structure shown. Between and on opposing sides of the semiconductor pillars, insulating pillars are formed. Thus, insulating pillars **492**, **494**, **496** and **498** are shown on opposing sides of the semiconductor pillars **493**, **495**, and **497**.

[0055] The select lines **137**, **138**, **139**, acting as gates for select transistors, underlie the bottom memory cube **102** and extend in the X-direction. The semiconductor pillars extend through the select lines **137**, **138** and **139** to the bit lines **134**, **135** and **136** extending in the Y-direction.

[0056] Common conductive plane **133**, acting as gates for each of the bottom set of source plane select transistors **148**, overlies the bottom memory cube **102**. The semiconductor pillars extend through the common conductive plane **133** to the source plane **147**.

[0057] Thus, FIG. **4** shows a memory device that comprises a plurality of bit lines **134**, **135**, **136** over a substrate. A plurality of memory cells is formed at the cross points of the semiconductor pillars **493**, **495**, **497** and the charge storage structures **425-430** on the opposing sides of the word lines (e.g. **410-412**), and lies over the plurality of bit lines. In this structure at least one memory cell of the plurality of memory cells is over another one memory cell of the plurality of memory cells, such as by being on the same semiconductor pillar as a lower cell. A source plane **147** is over the plurality of memory cells. This structure is suitable for use in forming a memory device that includes a first set of select transistors; a first 3D array of memory cells coupled to the first set of select transistors; a common source coupled to the first 3D array of memory cells; a second 3D array of memory cells coupled to the common source, wherein the common source is between the first and second 3D array of memory cells; and a second set of select transistors coupled to the second 3D array of memory cells, like that shown in FIG. **1**. Alternatively, this structure can be used as a top memory cube in a memory device including a top and a bottom memory cube, and in which the plurality of bit lines is shared between the top and bottom memory cubes.

[0058] FIG. **5** is a cross-sectional view in the Y-Z plane of the structure in FIG. **4** showing the two-cell unit structures **500**, **502**, **504** along a Z-direction column which includes the semiconductor pillar **497**. The reference numerals in FIG. **4** are repeated in FIG. **5** where appropriate.

[0059] The two-cell unit structure **500** includes a left cell **500**-L and a right cell **500**-R. The left cell **500**-L includes word line **418** as the gate and the charge storage structure **433** as the charge storage element. The right cell **500**-L includes word line **417** as the gate and the charge storage structure **435** as the charge storage element.

[0060] The two-cell unit structure **502** includes a left cell **502**-L and a right cell **502**-R. The left cell **502**-L includes word line **415** as the gate and the charge storage structure **431** as the charge storage element. The right cell **502**-L includes word line **414** as the gate and the charge storage structure **437** as the charge storage element.

[0061] The two-cell unit structure **504** includes a left cell **504**-L and a right cell **504**-R. The left cell **504**-L includes word line **412** as the gate and the charge storage structure **429** as the charge storage element. The right cell **504**-L includes word line **411** as the gate and the charge storage structure **439** as the charge storage element.

[0062] Each of the levels of word lines are separated by insulating material, such as silicon nitride or silicon dioxide. Thus, two Z-direction NAND strings are provided by the two-cell unit structures **500**, **502**, **504**. A left NAND string includes the left cells **500**-L, **502**-L, and **504**-L. A right NAND includes the right cells **500**-R, **502**-R, and **504**-R.

[0063] The select line **137** surrounds the semiconductor pillar **497**, and extends into and out of the cross-section illustrated in FIG. **5**. Gate dielectric **520** separates the select line **137** from the semiconductor pillar **497**. The common conductive plane **133** is separated from the semiconductor pillar **497** by gate dielectric **530**.

[0064] Bias voltages applied to the structures include right word line voltages V_{WL} applied to the word lines **411**, **414** and **417**, left word line voltages V_{WL} applied to the word lines **412**, **415**, and **418**, select line voltage V_{SEL} applied to the select line **137**, bit line voltage V_{BL} applied to the bit line **136**, source control voltage V_{SC} applied to the common conductive plane **133**, and source plane voltage V_sp applied to the source plane **147**. In the illustrated example the source plane is coupled to a ground terminal, so the source plane voltage V_{SP} is 0 Volts. Alternatively, voltages other than ground can be used for the source plane voltage V_{SP}.

[0065] The following table shows representative operating voltages for a read operation of the left memory cell **502**-L, and for a read operation of the right memory cell **502**-R. Of course, the voltage levels will vary from embodiment to embodiment.

Example Operating Voltages for Read Operation		
	Read Cell 502-L	Read Cell 502-R
V_{WL-412}	5 V	floating
V_{WL-415}	2 V	floating
V_{WL-418}	5 V	floating
V_{WL-411}	floating	5 V
V_{WL-414}	floating	2 V
V_{WL-417}	floating	5 V
V_{SEL}	3 V	3 V
V_{SC}	3 V	3 V
V_{SP}	0 V	0 V
V_{BL}	1 V	1 V

[0066] The following table shows representative operating voltages for a gate injection program via +FN programming of the left memory cell **502**-L, and for programming of the right memory cell **502**-R. Of course, the voltage levels will be adjusted as suits a particular implementation and programming method, and thus will vary from embodiment to embodiment.

Example Operating Voltages for Program Operation		
	Prog Cell 502-L	Prog Cell 502-R
V_{WL-412}	floating	floating
V_{WL-415}	+10 V	+20 V
V_{WL-418}	floating	floating
V_{WL-411}	floating	floating
V_{WL-414}	+20 V	+10 V
V_{WL-417}	floating	floating
V_{SEL}	5 V	5 V
V_{SC}	0 V	0 V
V_{SP}	0 V	0 V
V_{BL}	0 V	0 V

[0067] In an erase operation, all of the memory cells of the two-cell unit structures **500**, **502**, **504** are erased via an −FN erasing operation. The following table shows representative operating voltages for gate injection erase of all of the memory cells of the two-cell unit structures **500**, **502**, **504**.

Example Operating Voltages for Erase Operation	
V_{WL-412}	−20 V
V_{WL-415}	−20 V
V_{WL-418}	−20 V
V_{WL-411}	−20 V
V_{WL-414}	−20 V
V_{WL-417}	−20 V
V_{SEL}	3 V
V_{SC}	3 V
V_{SP}	0 V
V_{BL}	0 V

[0068] FIGS. **6-11** illustrate stages in a process for manufacturing a memory cube as discussed above. In FIG. **6**, a surface **600** of an integrated circuit substrate is illustrated with an array of contacts for connection to the 3D structure. The array of contacts includes a first set of contacts, including contacts **601-604** which are coupled to individual access devices, and adapted for connection to the semiconductor pillars in the 3D structure. The individual access devices can be formed in the substrate, and may include for example MOS transistors having gates coupled to select lines arranged to the X-direction, sources coupled to the bit lines arranged in the Y-direction, and drains connected to the contacts (e.g. **201-204**). The individual access devices are selected by biasing the select lines and bit lines as appropriate for the particular operation.

[0069] FIG. **7** is a side cross-section showing a multilayer stack of materials at a first stage in the manufacturing process, after forming alternating layers **721**, **723**, **725**, **727** of insulating material, such as silicon dioxide or silicon nitride, and layers **722**, **724**, **726**, **728** of word line material, such as n+-polysilicon, on top of the substrate **220**. In a representative structure, the thicknesses of the alternating layers of insulating material can be about 50 nanometers, and the thicknesses of the alternating layers of word line material can be about 50 nanometers. Over the top of the alternating layers, a layer **729** of hard mask material, such as silicon nitride, can be formed.

[0070] FIG. **8** is a layout view from a perspective over layer **729** showing the results using a first lithographic process to define a pattern for the trenches, and a patterned etch of the stack to form trenches **845-848** through the multilayer stack of materials shown in FIG. **7**, exposing contacts, such as contact **604**, coupled to individual access devices in the body line access circuits. Anisotropic reactive ion etching techniques can be used to etch through the polysilicon and silicon oxide or silicon nitride layers, with a high aspect ratio. The trenches have sidewalls **830-833** on which the layers of word line material are exposed at each level of the structure. The widths of the trenches **845-848** in a representative structure can be about 50 nanometers.

[0071] FIG. **9** shows a later stage in the process after deposition of the multilayer charge trapping structures (**940-943**) over and on the sidewalls of the trenches (**845-848**) contacting the layers of word line material. A representative structure and process is described below with reference to FIG. **16**. After deposition of the multilayer charge trapping structures, the process includes depositing a thin protective layer, such as p-type polysilicon over the multilayer charge trapping structures, and etching the resulting formation using an anisotropic process to remove the material of the multilayer charge trapping structures (**940-943**) from the bottom of the trenches, **845-848**, and exposing the contacts (e.g. **604**).

[0072] The charge trapping structures **940-943** include a tunneling layer in contact with the word line material, the charge trapping layer over the tunneling layer, and a blocking layer over the charge trapping layer, such as used in typical charge trapping memory devices. For example, the tunneling layer may comprise a layer of silicon dioxide or silicon oxide nitride, a charge trapping layer may comprise a layer of silicon nitride or other charge trapping material, and the blocking layer may comprise a layer of silicon dioxide or a high-κ material typical in SONOS type memory devices. In alternatives, as explained below with reference to FIG. **16**, a bandgap engineered charge trapping structure can be utilized.

[0073] FIG. **10** shows a next stage in the process after filling the trenches with the material to be used for the semiconductor pillars, such as p-type polysilicon, to form filled trenches **250-253**. The semiconductor material contacts the blocking layers of the charge trapping structures **940-943**.

[0074] FIG. **11** shows the result of using a second lithographic process to define a pattern for the semiconductor pillars. A patterned etch of the filled trenches is applied using an anisotropic etch process that is selective for the material of

the semiconductor pillars, to define the semiconductor pillars (**1150**-*a*, **1150**-*b*, **1150**-*c*, **1151**-*a*, **1151**-*b*, **1151**-*c*, **1152**-*a*, **1152**-*b*, **1152**-*c*, **1153**-*a*, **1153**-*b*, **1153**-*c*) contacting the contacts, including contact **604** (not shown) to the underlying individual access devices, and to create vertical openings between the semiconductor pillars. Next, dielectric insulating material, such as silicon dioxide, is filled in between the pillars to form insulator columns (e.g. insulator **1120**) between the semiconductor pillars.

[0075] The bottom set of source plane select transistors **148**, including the common conductive plane **133**, and the source plane **147** are then formed on the top of the structure, arranged as discussed above for connection to the top ends of the semiconductor pillars. The top set of source plane select transistors **198** are then formed over the source plane **133**, and the process for manufacturing a memory cube as discussed above with respect to FIGS. **6-11** can then be repeated to form a second memory cube.

[0076] FIG. **12** illustrates a top view of a configuration for making contact the left side and right side word lines in the plurality of levels. The left side word lines **1261-1**, **1261-2**, **1261-3** and **1263-1**, **1263-2**, **1263-3** and right side word lines **1262-1**, **1262-2**, **1262-3** and **1264-1**, **1264-2**, and **1264-3** in each layer have landing areas (labeled "L" or "R") arranged in a stair-step pattern (or other pattern) so that the landing areas in each level are not overlaid by any of the left side and right side word lines in the overlying levels. Contact plugs or other conductive lines (not shown) extend through the plurality of levels and contact the landing areas. An overlying patterned connection layer includes left side connectors **1228**, **1229**, **1230** and right side connectors **1225**, **1226**, **1227** over the plurality of patterned layers and in contact with the conductive lines contacting the landing areas of left and right sides conductors. The left and right side connectors are routed to left and right plane decoding circuits (not shown).

[0077] FIG. **13** is a layout view of a level in an alternative embodiment, showing left side word lines **1261-3**, **1263-3** and right side word lines **1260-3**, **1262-3**, **1264-3** from the top level of the structure in FIG. **12** and with extensions **1350**, **1351** for connection of the left side and right side word lines to the left and right plane decoders. The reference numerals used in FIG. **12** are repeated in FIG. **13** where appropriate. As can be seen, the left side word lines **1261-3**, **1263-3** are coupled to an extension **1351** which is adapted for connection to a contact plug on a landing area **1353**, through which connection to a decoder circuit on the integrated circuit substrate can be made. Likewise, right side word lines **1260-3**, **1262-3**, **1264-3** are coupled to an extension **1350** which is adapted for connection to a contact plug on landing area **1352**, through which connection to a decoder circuit on the integrated circuit substrate can be made.

[0078] FIG. **14** is a cross-sectional view of the interconnect structure taken through word lines **1260-1**, **1260-2**, **1260-3** of FIG. **12**, where conductors **1480-1**, **1480-2**, and **1480-3** extend through vias to landing areas on the right side word lines **1260-1**, **1260-2**, **1260-3** in respective levels. The conductors **1480-1**, **1480-2**, **1480-3** are used in this example for coupling the levels to interconnect lines (e.g. line **1485**) in a wiring layer, such as a patterned metallization layer for connection to decoding or bias circuits.

[0079] The landing areas are portions of patterned word lines **1260-1**, **1260-2**, **1260-3** used for contact with conductors **1480-1**, **1480-2**, **1480-3**. The sizes of the landing areas are large enough to provide room for the conductors **1480-1**, **1480-2**, **1480-3** to adequately couple the word lines **1260-1**, **1260-2**, **1260-3** on the various levels to the overlying interconnect lines (e.g. **1485**), as well as to address issues such as alignment tolerances.

[0080] The size of a landing area thus depends on a number of factors, including the size and number of conductors used, and will vary from embodiment to embodiment.

[0081] For the purpose of this description, the direction in which the word lines **1260-1** to **1260-3** extend is referred to herein as the "longitudinal" direction. The "transverse" direction is perpendicular to the longitudinal direction, and is into and out of the cross-section illustrated in FIG. **14**. Both the longitudinal and transverse directions are considered to be "lateral dimensions," meaning a direction that is in a 2-dimensional area of a plan view of the word lines **1260-1** to **1260-3** on the various levels. The "length" of structures or features is its length in the longitudinal direction, and its "width" is its width in the transverse direction.

[0082] Word line **1260-1** is the lowest level in the plurality of levels. The word line **1260-1** includes landing area **1461-1**. The word line **1260-2** includes landing area **1461-2**. The word line **1260-3** includes landing area **1461-3**.

[0083] In FIG. **14**, the right side word line **1260-1** includes landing area **1461-1** on the right end. Left side word lines, such as word line **1261-1** in FIG. **12**, include landing areas on the left end. In some alternative embodiments additional landing areas may be defined, such as a landing area on the opposite end of the word line. This structure can be made using a first mask to define a longitudinal opening over the left side and right side word lines, and a second mask to define an opening over a landing area to the lowest layer through all the overlying layers, etching the opening, and then trimming the mask to define a next opening, etching the opening, and trimming the mask and so on until openings to all the layers are formed, and landing areas are created in a stair-step configuration, that are aligned on the conductor lines, and that are not overlaid by conductor lines on any overlying layer. A more detailed description of a process for manufacturing this structure is shown in my co-pending U.S. patent application entitled 3D INTEGRATED CIRCUIT LAYER INTERCONNECT, application Ser. No. 12/579,192, filed 14 Oct. 2009, which is incorporated by reference as if fully set forth herein.

[0084] FIG. **15**A is a plan view of a portion of word line **1260-1**, including the landing areas **1461-1***a*, **1461-1***b*, one on each end of the word line, so that all the left side and right side conductors can be formed in the same process. The width of the word line **1260-1** in this drawing is exaggerated for the purpose of clarity in the figure. As shown in FIG. **15**A, landing area **1461-1***a* has a width **1500** in the transverse direction and a length **1501** in the longitudinal direction. Landing area **1461-1***b* has a width **1502** in the transverse direction and a length **1503** in the longitudinal direction. In the embodiment of FIG. **15**A the landing areas **1461-1***a*, **1461-1***b* each have a rectangular cross-section. In embodiments, the landing areas **1461-1***a*, **1461-1***b* may each have a cross-section that is circular, elliptical, square, rectangular, or somewhat irregularly shaped.

[0085] Because word line **1260-1** is in the lowest level, the vertical conductors (like conductors **1480-1**, **1480-2**, **1480-3** in FIG. **14**) in the vias need not pass through the word line **1260-1** to underlying levels. Thus, in this example conductor line **1260-1** does not have openings.

[0086] FIG. **15**B is a plan view of word line **1260-2**. As illustrated in FIG. **14**, word line **1260-2** overlies word line

1260-1. Word line **1260-2** includes an opening **1550** overlying the landing area **1461-1***a* on word line **1260-1**. The opening **1550** has a distal longitudinal sidewall **1551***a* and a proximal longitudinal sidewall **1551***b* defining the length **1552** of the opening **1550**. The length **1552** of the opening **1550** is at least as large as the length **1505** of the underlying landing area **1461-1***a*, so that the conductor **1480-1** for the landing area **1461-1***a* can pass through the word line **1260-2**.

[0087] The word line **1260-2** also includes opening **1555** overlying the landing area **1461-1***b*. The opening **1555** has distal and proximal longitudinal sidewalls **1556***a*, **1556***b* defining the length **1557** of the opening **1555**. The length **1557** of the opening **1555** is at least as large as the length **1507** of the underlying landing area **1461-1***b*, so that the conductors for the landing area **1461-1***b* can pass through the word line **1260-2**.

[0088] The word line **1260-2** also includes first and second landing areas **1461-2***a*, **1461-2***b* adjacent the openings **1550**, **1555** respectively. The first and second landing areas **1461-2***a*, **1461-2***b* are the portions of word line **1260-2** used for contact with the vertical conductors.

[0089] FIG. **15**C is a plan view of a portion of word line **1260-3**, including the first and second landing areas **1461-3***a*, **1461-3***b* and the openings **1560**, **1565** within the interconnect structure. As shown in FIG. **14**C, word line **1260-3** includes an opening **1560**, which is arranged to overlie landing area **1461-1***a* on word line **1260-1** and landing area **1461-2***a* on word line **1260-2**. The opening **1560** has distal and proximal longitudinal sidewalls **1561***a*, **1561***b* defining the length **1562** of the opening **1560**. The length **1562** of the opening **1560** is at least as large as the sum of the lengths **1501** and **1505** of the underlying landing areas **1461-1***a* and **1461-2***a*, so that the conductors **1480** for the landing areas **1461-1***a* and **1461-2***a* can pass through the word line **1260-3**.

[0090] The distal longitudinal sidewall **1561***a* of opening **1560** is vertically aligned with the distal longitudinal sidewall **1551***a* of the underlying opening **1550**. As mentioned above, the openings can be formed using the opening in a single etch mask and one additional mask formed over the opening in the single etch mask, as well as processes for etching the additional mask without a critical alignment step, resulting in the formation of openings having distal longitudinal sidewalls (**1561***a*, **1551***a*,) along the perimeter of the single etch mask that are vertically aligned.

[0091] The word line **1260-3** also includes opening **1565** overlying the landing area **1461-1***b* on conductor line **1260-1** and landing area **1461-2***b* on word line **1260-2**. The opening **1565** has outside and inside longitudinal sidewalls **1566***a*, **1566***b* defining the length **1567** of the opening **1565**. The outside longitudinal sidewall **1566***a* of opening **1565** is vertically aligned with the outside longitudinal sidewall **1556***a* of the underlying opening **1555**.

[0092] The length **1567** of the opening **1565** is at least as large as the sum of the lengths of the underlying landing areas and openings, so that the conductors **1480** for the landing areas can pass through.

[0093] The word line **1260-3** also includes first and second landing areas **1461-3***a*, **1461-3***b* adjacent the openings **1560**, **1565** respectively. The first and second landing areas **1461-3***a*, **1461-3***b* are the portions of word line **1260-3** used for contact with the conductors **1480**. As shown in FIG. **15**C, landing area **1461-3***a* is adjacent the opening **1560** and has a width **1514** in the transverse direction and a length **1515** in the longitudinal direction. Landing area **1461-3***b* is adjacent the opening **1565** having a width **1516** in the transverse direction and a length **1517** in the longitudinal direction.

[0094] In the illustrated embodiment, the openings in the various word lines **1260-1** to **1260-3** have substantially the same width in the transverse direction. Alternatively, the width of the openings can vary along the longitudinal direction, for example in a step-like manner, in order to accommodate landing areas having different widths.

[0095] In the cross-section of FIG. **12**, the openings within the interconnect structure result in the levels having a staircase-like pattern on both ends. That is, the two openings in each level are symmetrical about an axis perpendicular to both the longitudinal and transverse directions, and the two landing areas of each level are also symmetrical about that axis. As used herein, the term "symmetrical" is intended to accommodate manufacturing tolerances in the formation of the openings using the opening in a single etch mask and multiple etch processes which may cause variations in the dimensions of the openings.

[0096] In alternative embodiments in which each level includes a single opening and a single landing area, the levels have a staircase-like pattern on only one side.

[0097] FIG. **16** is a simplified diagram of a charge storage structure suitable for use in the memory cells described herein, employing a bandgap engineered dielectric tunneling layer (BE-SONOS- type). The memory cell includes a channel surface **1600***a* in a semiconductor pillar **1600**. Not shown in FIG. **16** are the dielectric insulators separating the pillars.

[0098] A gate **1620** in this embodiment comprises n+ polysilicon. P+ polysilicon may also be used. Other embodiments employ metals, metal compounds or combinations of metals and metal compounds for the gate **1620**, such as platinum, tantalum nitride, metal silicides, aluminum or other metal or metal compound gate materials. For some applications, it is preferable to use materials having work functions higher than 4.5 eV. A variety of high work function materials suitable for use as a gate terminal are described in U.S. Pat. No. 6,912,163. Such materials are typically deposited using sputtering and physical vapor deposition technologies, and can be patterned using reactive ion etching.

[0099] In the embodiment illustrated in FIG. **16**, the dielectric tunneling layer on the gate side comprises a composite of materials, including a first layer **1619**, referred to as a hole tunneling layer, of silicon dioxide on the surface of the gate **1620** formed for example using in-situ steam generation ISSG with optional nitridation by either a post deposition NO anneal or by addition of NO to the ambient during deposition. The thickness of the first layer **1619** of silicon dioxide is less than 2 nm, and preferably 1.5 nm or less.

[0100] A second layer **1618**, referred to as a band offset layer, of silicon nitride lies on the first layer **1619** of silicon oxide formed for example using low-pressure chemical vapor deposition LPCVD, using for example dichlorosilane DCS and NH3 precursors at 680° C. In alternative processes, the band offset layer comprises silicon oxynitride, made using a similar process with an N2O precursor. The thickness of the layer **1618** of silicon nitride is less than 3 nm, and preferably 2.5 nm or less.

[0101] A third layer **1617** of silicon dioxide, referred to as an isolation layer, lies on the layer **1618** of silicon nitride formed for example using LPCVD high temperature oxide HTO deposition. The layer **1617** may also be implemented using silicon oxynitride, or other suitable material having a

relatively large bandgap. The thickness of the layer **1617** is less than 4 nm , and preferably 3.5 nm or less.

[0102] A charge trapping layer **1616** in this embodiment comprises silicon nitride having a thickness greater than 5 nm, including for example about 7 nm in this embodiment formed for example using LPCVD. Other charge trapping materials and structures may be employed, including for example silicon oxynitride (SixOyNz), silicon-rich nitride, silicon-rich oxide, trapping layers including embedded nanoparticles and so on.

[0103] A blocking dielectric layer **1615** in this embodiment comprises silicon oxide, which can be formed by wet conversion from the nitride by a wet furnace oxidation process. Other embodiments may be implemented using high temperature oxide (HTO) or LPCVD SiO2. The thickness of layer **1615** of silicon oxide can be for example in the range of about 5 to 8 nanometers, while the thickness of the layer **1616** of silicon nitride can be for example in the range of 5 to 7 nanometers. The layer **1615** of silicon oxide in one example is about 7 nm. Alternatively, other materials such as high-κ metal oxides like aluminum oxide, hafnium oxide and so on, or combinations of materials, can be used for the blocking dielectric layer **415**.

[0104] In a representative embodiment, the first layer **1619** can be 1.3 nm of silicon dioxide; the band offset layer **1618** can be 2 nm of silicon nitride; the isolation layer **1617** can be 2.5 nm of silicon dioxide; the charge trapping layer **1616** can be 7 nm of silicon nitride; and the blocking dielectric layer **1615** can be 7 nm of silicon oxide. The gate material can be p+ polysilicon.

[0105] The stack of layers **1619-1617** of FIG. **16** under a low electric field, has a "U-shaped" conduction band and an "inverted U-shaped" valence band. The dielectric tunneling layer described herein therefore is characterized by band offset characteristics, include a relatively large hole tunneling barrier height in a thin region (first layer **1619**) at the interface with the semiconductor body, and an increase in valence band energy level at a first offset less than 2 nm from the channel surface. The band offset characteristics also include a decrease in valence band energy level at a second offset (second layer **1618**) from the channel by providing a thin layer of relatively high tunneling barrier height material (third layer **1617**), resulting in the inverted U-shaped valence band shape. Likewise, the conduction band has a U-shape caused by the same selection of materials.

[0106] The valence band energy level at the first location is such that an electric field sufficient to induce hole tunneling through the thin region between the interface with the semiconductor body and the first location, is also sufficient to raise the valence band energy level after the first location to a level that effectively eliminates the hole tunneling barrier in the composite tunneling dielectric after the first location. This structure enables electric field assisted hole tunneling at high speeds while effectively preventing charge leakage through the composite tunneling dielectric in the absence of electric fields or in the presence of smaller electric fields induced for the purpose of other operations, such as reading data from the cell or programming adjacent cells.

[0107] FIG. **17** shows one example implementation for an array of access devices suitable for use as the bottom set of pillar select transistors **106** shown in FIG. **1**. As shown in FIG. **17**, an access layer **1704** is implemented in a substrate including insulating material **1710**, having a top surface with an array of contacts (e.g. contact **1712**) exposed thereon. The contacts for individual pillars are provided at top surfaces of drain contacts **1708**, which are coupled to the drain terminals of MOS transistors in the access layer **1704**. The access layer **1704** includes a semiconductor body having source regions **1742** and drain regions **1736** therein. Polysilicon select lines **1734** are provided over dielectric layers and between the source regions **1742** and drain regions **1736**. In the embodiment shown, the source regions **1742** are shared by adjacent MOS transistors, making two-transistor structures **1748**. Source contacts **1740** are positioned between word lines **1734** and contact source regions **1742** within substrate **1738**. The source contacts **1740** can be connected to bit lines (not shown) in a metal layer, which run perpendicular to the word lines and between the columns of drain contacts **1708**. Select lines **1734** are covered by silicide caps **1744**. Select lines **1734** and caps **1744** are covered by a dielectric layer **1745**. Isolation trenches **1746** separate the two-transistor structures **1748** from the adjacent two-transistor structures. In this example transistors act as the access devices. Individual pillars can be coupled to the contacts **1712**, and selected individually by controlling the biasing of the source contacts **1740** and the word lines **1734**. Of course other structures may be used to implement the access device array, including for example, vertical MOS device arrays.

[0108] FIG. **18** is a simplified block diagram of an integrated circuit **1875** according to an embodiment of the present invention. The integrated circuit **1875** including a 3D NAND flash memory array **1860** with a source plane, implemented as described herein, on a semiconductor substrate. Addresses are supplied on bus **1865** to column decoder **1863**, row decoder **1861** and left/right plane and source plane decoders **1858**. Arrays of access devices for individual pillars underlie and overlie the array **1860**, and are coupled to the row decoder **1861** and the column decoder **1863**, for array embodiments like that shown in FIG. **1**. Sense amplifiers and data-in structures in block **1866** are coupled to the array in this example via data bus **1867**. Data is supplied via the data-in line **1871** from input/output ports on the integrated circuit **1875** or from other data sources internal or external to the integrated circuit **1875**, to the data-in structures in block **1866**. In the illustrated embodiment, other circuitry **1874** is included on the integrated circuit, such as a general purpose processor or special purpose application circuitry, or a combination of modules providing system-on-a-chip functionality supported by the memory cell array. Data is supplied via the data-out line **1872** from the sense amplifiers in block **1866** to input/output ports on the integrated circuit **1875**, or to other data destinations internal or external to the integrated circuit **1875**.

[0109] A controller implemented in this example using bias arrangement state machine **1869** controls the application of bias arrangement supply voltages generated or provided through the voltage supply or supplies in block **1868**, such as read, erase, program, erase verify and program verify voltages. The controller can be implemented using special-purpose logic circuitry as known in the art. In alternative embodiments, the controller comprises a general-purpose processor, which may be implemented on the same integrated circuit, which executes a computer program to control the operations of the device. In yet other embodiments, a combination of special-purpose logic circuitry and a general-purpose processor may be utilized for implementation of the controller.

[0110] Three-dimensional stacking is an efficient way to reduce the cost per bit for semiconductor memory, particu-

larly when physical limitations in the size of the memory elements is reached for a given plane. Prior art technology addressed to 3D arrays requires several critical lithography steps to make minimum feature size elements in each stack layer. Also, driver transistors used for the memory array multiplied in number by the number of planes.

[0111] Technology described here includes a high density 3D array in which only one critical layer lithography step is required to pattern all the layers of a memory cube. The memory via and layer interconnect via patterning steps shared by each layer in a memory cube. Also, the layers can share the word line and bit line decoders to reduce the area penalty of prior art multilevel structures. Also, a unique two-cell unit structure is described for charge trapping memory in which independently addressable data sites are provided on each of two sides of a semiconductor pillar. An array of access devices is used to select individual semiconductor pillars. Left and right word lines are used to select individual cells on selected levels, and common conductive planes are used to select individual memory cubes.

[0112] FIG. **19** illustrates a schematic diagram of a second embodiment of the 3D memory device **200** which includes a third memory cube **1902** overlying the structure illustrated in FIG. **1**.

[0113] The third memory cube **1902** includes slices **1910**, **1912**, **1914** which lie in X-Z planes. In the illustrated schematic, there are nine two-cell unit structures **1920-1928** in slice **1910**. The third memory cube **1902** shares the top set of pillar select transistors **156** with the memory cube **152**, including the bit lines **134**, **135**, **136** and select lines **187**, **188**, **189**. Semiconductor pillars **1930**, **1931**, **1932** couple the two-cell unit structures **1920-1928** in a Z-direction column to the top set of pillar select transistors **156**.

[0114] The left gates on the two-cell unit structures in a particular level in all of the slices **1910**, **1912**, **1914** are coupled via a word line element **1941**, **1943**, **1945** to the left plane decoder **104**. Likewise, the right gates on the two-cell unit structures in a particular level in all of the slices **1910**, **1912**, **1914** are coupled via a word line element **1942**, **1944**, **1945** to the right plane decoder **105**.

[0115] Top ends of the array of semiconductor pillars of the two-cell unit structures **1920-1928** in a Z-direction column are coupled via a corresponding select transistor in a set of source plane select transistors **1948** to a source plane **1947**. The source plane **147** is a layer of conductive material overlying the memory cube **1902**. In the illustrated example, the source plane **1947** is coupled to a ground terminal. Alternatively, the source plane **1947** may be coupled to a voltage source for applying a voltage other than ground.

[0116] The gates of the access devices in the set of source plane select transistors **1948** are all coupled via common conductive plane **1933** to a driver selected by source plane decoder **107**.

[0117] In some embodiments, the 3D memory device **200** may include one or more sets of select transistors between the top set of pillar select transistors **156** and the pillars in the third memory cube **1902** and/or the pillars in the memory cube **152**.

[0118] While the present invention is disclosed by reference to the preferred embodiments and examples detailed above, it is to be understood that these examples are intended in an illustrative rather than in a limiting sense. It is contemplated that modifications and combinations will readily occur to those skilled in the art, which modifications and combinations will be within the spirit of the invention and the scope of the following claims.

What is claimed is:

1. A memory device, comprising:

a plurality of bit lines over a substrate;

a plurality of memory cells over the plurality of bit lines, wherein at least one memory cell of the plurality of memory cells is over another one memory cell of the plurality of memory cells; and

a source plane over the plurality of memory cells.

2. The memory device of claim **1**, comprising:

bottom and top memory cubes including respective arrays of semiconductor pillars, a given semiconductor pillar in each array of semiconductor pillars including a plurality of dielectric charge trapping structures in a NAND string along a side of the given semiconductor pillar, and word lines orthogonal to the semiconductor pillars and adjacent the charge trapping structures of the semiconductor pillars, the bottom memory cube including said plurality of memory cells;

the source plane being between the bottom and top memory cubes, the source plane selectively coupled to top ends of each of the semiconductor pillars of the bottom memory cube by a bottom set of source plane select transistors, and selectively coupled to bottom ends of each of the semiconductor body pillars of the top memory cube by a top set of source plane select transistors;

a bottom set of pillar select transistors to selectively couple bottom ends of the semiconductor pillars of the bottom memory cube to corresponding bit lines in said plurality of bit lines; and

a top set of pillar select transistors to selectively couple top ends of the semiconductor pillars of the top memory cube to corresponding bit lines in a plurality of top bit lines.

3. The memory device of claim **2**, wherein the given semiconductor pillar in each array of semiconductor pillars comprises a plurality of two-cell unit structures, the two-cell unit structures of the given semiconductor pillar including a first dielectric change trapping structure along a first side of the given semiconductor pillar, and a second dielectric charge trapping structure along a second opposing side of the given semiconductor pillar, such that the first charge trapping structures in the two-cell unit structures of the given semiconductor pillar are arranged in a first NAND string, and the second charge trapping structures in the two-cell unit structures of the given semiconductor pillar are arranged in a second NAND string.

4. The memory device of claim **3**, wherein said word lines include a plurality of pairs of word lines arranged orthogonal to one of the arrays of semiconductor pillars, a given pair of word lines including:

a first set of word lines commonly coupled to one another and adjacent the first dielectric charge trapping structures between alternating rows of semiconductor pillars in said one of the arrays of semiconductor pillars; and

a second set of word lines commonly coupled to one another and interleaved with the word lines in the first set of word lines, the second set of word lines adjacent the second dielectric charge trapping structures between interleaved alternative rows of semiconductor pillars.

5. The memory device of claim **2**, wherein the dielectric charge trapping structures in the plurality of dielectric charge trapping structures including a tunneling layer adjacent a corresponding word line in said word lines, and a dielectric charge trapping layer between the tunneling layer and a blocking layer, the blocking layer between the dielectric charge trapping layer and the given semiconductor pillar.

6. The memory device of claim **5**, further comprising a controller to program and erase selected memory cells by biasing corresponding semiconductor pillars and corresponding word lines to induce Fowler-Nordheim tunneling from the word lines through the tunneling layer to the dielectric charge trapping structures.

7. The memory device of claim **2**, including decoder circuitry, the decoder circuitry including:

a row decoder coupled to the bottom and top bit lines and arranged to access individual rows of semiconductor pillars;

a column decoder coupled to the bottom and top sets of pillar select transistors and arranged to access individual columns of semiconductor pillars;

a source plane decoder coupled to the bottom and top sets of source plane select transistors and arranged to access one of the bottom and top memory cubes; and

a fifth decoder coupled to the word lines, and arranged to drive word lines on selected levels in the semiconductor pillars.

8. The memory device of claim **2**, further comprising a third memory cube overlying the top memory cube, the top set of pillar select transistors and the plurality of top bit lines, the third memory cube including a third array of semiconductor pillars including a plurality of dielectric charge trapping structures arranged in NAND strings, and word lines orthogonal to the semiconductor pillars in the third array of semiconductor pillars, and wherein:

the top set of pillar select transistors selectively couple bottom ends of the semiconductor pillars of the third memory cube to corresponding bit lines in the plurality of top bit lines.

9. The memory device of claim **1**, comprising:

bottom and top memory cubes including respective arrays of semiconductor pillars, a given semiconductor pillar in each array of semiconductor pillars including a plurality of dielectric charge trapping structures in a NAND string along a side of the given semiconductor pillar, and word lines orthogonal to the semiconductor pillars and adjacent the charge trapping structures of the semiconductor pillars, the top memory cube including said plurality of memory cells;

the source plane overlying the top memory cube, the source plane selectively couple to top ends of the semiconductor pillars of the top memory cube by a top set of source plane select transistors;

a bottom source plane underlying the bottom memory cube, the bottom source plane selectively coupled to bottom ends of the semiconductor pillars of the bottom memory cube by a bottom set of source plane select transistors; and

a set of shared pillar select transistors to selectively couple top ends of the semiconductor pillars of the bottom memory cube and bottom ends of the semiconductor pillars of the top memory cube to corresponding bit lines in the plurality of bit lines.

10. A memory device, comprising:

a first set of select transistors;

a first 3D array of memory cells coupled to the first set of select transistors;

a common source coupled to the first 3D array of memory cells;

a second 3D array of memory cells coupled to the common source, wherein the common source is between the first and second 3D array of memory cells; and

a second set of select transistors coupled to the second 3D array of memory cells.

11. A method for manufacturing a memory device, comprising:

forming a memory cube, comprising:

providing a substrate including a set of pillar select transistors and a plurality of bit lines coupled to the set of pillar select transistors, the substrate having an array of contacts coupled to the set of pillar select transistors;

forming a stack of alternating layers of word line material and insulating material over the array of contacts;

forming trenches in the stack, the trenches exposing respective rows of contacts on the surface of the substrate coupled to the set of pillar select transistors, and having sidewalls exposing word line material in the layers of word line material in the stack;

forming a charge trapping structure, lining the sidewalls of the trenches at least on word line material exposed on sidewalls of the trenches;

forming semiconductor pillars within the trenches over the charge trapping structure, the semiconductor pillars contacting respective contacts in the rows of contacts in the trenches; and

forming insulator columns within the trenches on first and second opposing sides of the semiconductor pillars;

forming a set of source plane select transistors having first conduction terminals coupled to top ends of each the semiconductor pillars of the memory cube, and having gate terminals commonly coupled to one another; and

forming a source plane coupled to second conduction terminals of each source plane select transistor in the set of source plane select transistors.

12. The method of claim **11**, wherein said forming semiconductor pillars includes filling the trenches with doped semiconductor material over the charge trapping structure, and then applying a patterned etch process to remove the doped semiconductor material within the trenches to leave said semiconductor pillars.

13. The method of claim **12**, wherein said forming insulator columns includes filling regions left by said patterned etch process with insulator material to provide said insulator columns.

14. The method of claim **11**, wherein said forming a charge trapping structure includes forming a multilayer stack of dielectric materials on the sidewalls of the trenches, the multilayer stack including a tunneling layer, a charge trapping layer and a blocking layer.

15. The method of claim **14**, wherein said tunneling layer is formed adjacent the sidewalls of the trenches.

16. The method of claim **11**, wherein said forming trenches includes using a first lithographic mask to define a pattern for the trenches, and said forming semiconductor pillars includes

using a second lithographic mask to define a pattern for the semiconductor pillars in the trenches.

17. The method of claim **11**, including etching the stack to define left side word line structures including a plurality of word lines between alternating pairs of trenches, and right side word line structures including a second plurality of word lines interleaved with the first plurality of word lines between alternating pairs of trenches.

18. The method of claim **11**, further comprising:

forming a second set of source plane select transistors having first conduction terminals to the source plane, having gate terminals having gate terminals commonly coupled to one another, and having second conduction terminals coupled to a second array of contacts; and

forming a second memory cube overlying the second set of source plane select transistors, comprising:

forming a second stack of alternating layers of word line material and insulating material over the second array of contacts;

forming second trenches in the second stack, the second trenches exposing respective rows of contacts in the second array of contacts coupled to the second set of pillar select transistors, and having sidewalls exposing word line material in the layers of word line material in the second stack;

forming a second charge trapping structure, lining the sidewalls of the second trenches at least on word line material exposed on sidewalls of the second trenches;

forming second semiconductor pillars within the second trenches over the second charge trapping structure, the second semiconductor pillars contacting respective contacts in the rows of contacts in the second trenches; and

forming second insulator columns within the second trenches on first and second opposing sides of the second semiconductor pillars.

* * * * *