(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 July 2007 (19.07.2007)

PCT

(10) International Publication Number
WO 2007/081642 A2

(51) International Patent Classification: Not classified

(21) International Application Number:
PCT/US2006/062188

(22) International Filing Date:
15 December 2006 (15.12.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
11/316,474 21 December 2005 (21.12.2005) US
11/316,654 21 December 2005 (21.12.2005) US

(71) Applicant *(for all designated States except US)*: **SANDISK CORPORATION** [US/US]; 601 McCarthy Blvd., Milpitas, California 95035 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **HIGASHITANI, Masaaki** [JP/US]; 22212 Bitter Oak Street, Cupertino, California 95014 (US).

(74) Agents: **FANUCCI, Allan, A.** et al.; WINSTON & STRAWN LLP, Patent Department, 1700 K Street, N.W., Washington, D.C. 20006-3817 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

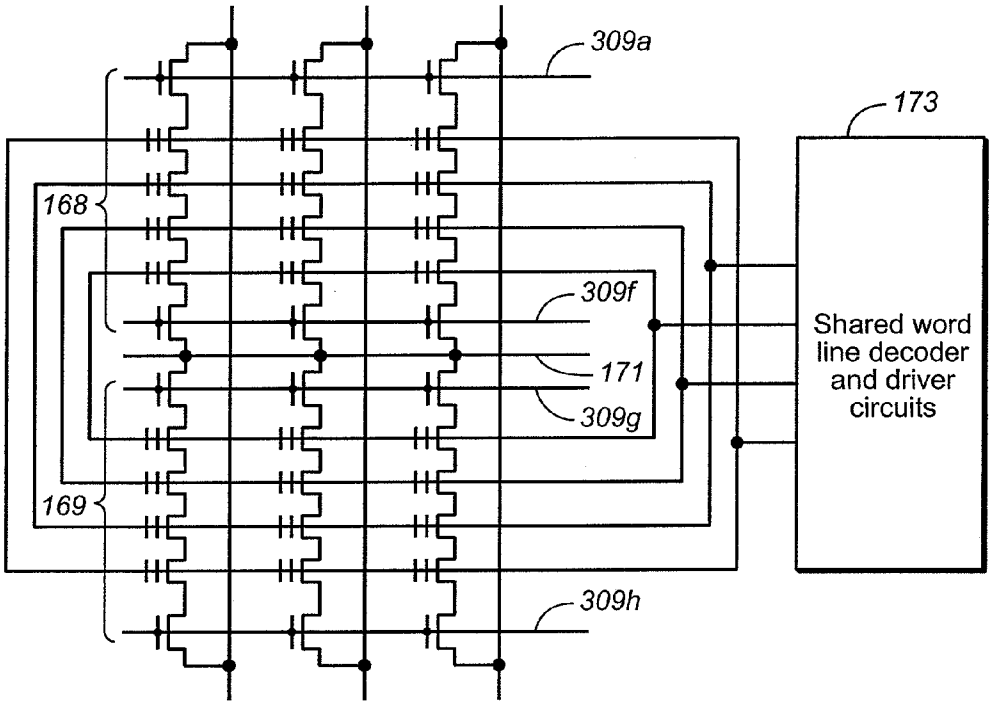
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) **Title:** FLASH DEVICES WITH SHARED WORD LINES



(57) **Abstract:** Word lines of a NAND flash memory array are formed by concentric, rectangular shaped, closed loops that have a width of approximately half the minimum feature size of the patterning process used. The resulting circuits have word lines linked together so that peripheral circuits are shared. Separate erase blocks are established by shield plates.

WO 2007/081642 A2

FLASH DEVICES WITH SHARED WORD LINES

BACKGROUND

[0001] This invention relates to flash memory arrays and in particular to the structures of flash memory arrays and methods of forming them.

[0002] All patents, patent applications, publications and other references cited in the present application are hereby incorporated by reference in their entirety.

[0003] There are many commercially successful nonvolatile memory products being used today, particularly in the form of small form factor cards, which use an array of flash EEPROM (Electrically Erasable and Programmable Read Only Memory) cells. Such cards may be interfaced with a host, for example, by removably inserting a card into a card slot in a host. Some of the commercially available cards are CompactFlash™ (CF) cards, MultiMedia cards (MMC), Secure Digital (SD) cards, Smart Media cards, personnel tags (P-Tag) and Memory Stick cards. Hosts include personal computers, notebook computers, personal digital assistants (PDAs), various data communication devices, digital cameras, cellular telephones, portable audio players, automobile sound systems, and similar types of equipment. In an alternative arrangement to the separate card and host described above, in some examples a memory system is permanently connected to a host providing an embedded memory that is dedicated to the host.

[0004] An example of a prior art memory system 100 is generally illustrated in the block diagram of Figure 1. A large number of individually addressable memory cells are arranged in a regular array 110 of rows and columns, although other physical arrangements of cells are certainly possible. Bit lines, designated herein to extend along columns of the array 110, are electrically connected with a bit line decoder and driver circuit 130 through lines 150. Word lines, which are designated in this description to extend along rows of the array 110, are electrically connected through lines 170 to a word line decoder and driver circuit 190. Each of the decoders 130 and

190 receives memory cell addresses over a bus 160 from a memory controller 180. The decoder and driving circuits are also connected to the controller 180 over respective control and status signal lines 135 and 195.

[0005] The controller 180 is connectable through lines 140 to a host device (not shown). The host may be a personal computer, notebook computer, digital camera, audio player, various other hand held electronic devices, and the like. The memory system 100 of Figure 1 will commonly be implemented in a card according to one of several existing physical and electrical standards, such as one from the PCMCIA, the CompactFlash™ Association, the MMC™ Association, and others. When in a card format, the lines 140 terminate in a connector on the card that interfaces with a complementary connector of the host device. The electrical interface of many cards follows the ATA standard, wherein the memory system appears to the host as if it was a magnetic disk drive. Other memory card interface standards also exist. In some systems, a memory card may not have a controller and the functions of the controller may be carried out by the host. As an alternative to the card format, a memory system of the type shown in Figure 1 may be permanently embedded in the host device.

[0006] The decoder and driver circuits 130 and 190 generate appropriate voltages in their respective lines of the array 110, as addressed over the bus 160, according to control signals in respective control and status lines 135 and 195, to execute programming, reading and erasing functions. Any status signals, including voltage levels and other array parameters, are provided by the array 110 to the controller 180 over the same control and status lines 135 and 195. A plurality of sense amplifiers within the circuit 130 receive current or voltage levels that are indicative of the states of addressed memory cells within the array 110, and provides the controller 180 with information about those states over lines 145 during a read operation. A large number of sense amplifiers are usually used in order to be able to read the states of a large number of memory cells in parallel. During reading and program operations, one row of cells is typically addressed at a time through the circuits 190 for accessing a number of cells in the addressed row that are selected by the circuit 130. During an erase operation, all cells in each of many rows are typically addressed together as a block for simultaneous erasure.

[0007] Two general memory cell array architectures have found commercial application, NOR and NAND. In a typical NOR array, memory cells are connected between adjacent bit line source and drain diffusions that extend in a column direction with control gates connected to word lines extending along rows of cells. A memory cell includes at least one storage element positioned over at least a portion of the cell channel region between the source and drain. A programmed level of charge on the storage elements thus controls an operating characteristic of the cells, which can then be read by applying appropriate voltages to the addressed memory cells. Examples of such cells, their uses in memory systems and methods of manufacturing them are given in the following United States patents: 5,070,032; 5,095,344; 5,313,421; 5,315,541; 5,343,063; 5,661,053 and 6,222,762. These patents, along with all other patents, patent applications and other publications referred to in this application are hereby incorporated by reference in their entirety for all purposes.

[0008] In a NAND array series strings of more than two memory cells, such as 16 or 32, are connected along with one or more select transistors between individual bit lines and a reference potential to form columns of cells. Word lines extend across cells within a large number of these columns. An individual cell within a column is read and verified during programming by causing the remaining cells in the string to be turned on hard so that the current flowing through a string is dependent upon the level of charge stored in the addressed cell. An example of a NAND architecture array and its operation as part of a memory system is found in the following United States patents: 5,570,315; 5,774,397; 6,046,935 and 6,522,580. NAND memory devices have been found to be particularly suitable for mass storage applications such as those using removable memory cards.

[0009] Figure 2A shows a portion of EEPROM cell array 110 of Figure 1 having a NAND structure. Only a small portion of the repetitive structure is shown. NAND strings of memory cells are formed extending in the Y-direction. NAND strings include implanted source/drain regions that connect individual memory cells. A memory cell includes a floating gate overlying a channel region in the substrate. A series of word lines, WL0-WL3 extend across the memory array in the X-direction and overlie floating gates of memory cells of different strings. In addition, select gate

lines (SSL, DSL) extend in the X-direction at either end of the NAND strings and overlie portions of the substrate to form select gates of select transistors that control the connection of NAND strings to memory control circuits. At one end of the NAND strings, a common source line (not shown) connects to each of the NAND strings. At the other end of the NAND strings, connections are made to bit lines (not shown). In a typical NAND memory array, NAND strings that share word lines and select lines form a block in the memory array that is erased as a unit. A typical string may include many memory cells, with 8, 16, 32 or more memory cells in a string being common. Thus, a typical block may have 32 or more word lines extending across the NAND strings of the block. A block may have thousands of strings that are spaced apart in the X-direction. Figure 2B shows a circuit diagram for the physical structure of Figure 2A. Figure 2B includes the common source line connecting the NAND strings at one end. NAND strings are shown extending between bit line connections and common source connections with select transistors controlling these connections.

[0010] Figure 2C shows a cross sectional view of a NAND string of Figure 2A (indicated by A-A in Figure 2A). Figure 2C more clearly shows the structure of individual memory cells having a floating gate (FG) formed from a first polysilicon layer (P1) and a control gate (CG) formed from a second polysilicon layer (P2). The control gate is formed by a portion of a word line that overlies a floating gate. In between a floating gate and a control gate is a dielectric layer 19. In addition, Figure 2C shows implanted source/drain regions connecting adjacent cells in the NAND string. A gate dielectric layer is shown insulating floating gates from the substrate. Metal bit line contact and source contact are shown at either end of the NAND string. A source select transistor and a drain select transistor are shown having portions of both first polysilicon layer P1 and second polysilicon layer P2. For select transistors, these two layers are connected together so that no floating gate is formed. Alternatively, a single polysilicon layer may be used to form select gates.

[0011] The charge storage elements of current flash EEPROM arrays, as discussed in the foregoing referenced patents, are most commonly electrically conductive floating gates, typically formed from conductively doped polysilicon material. An alternate type of memory cell useful in flash EEPROM systems utilizes a non-conductive

dielectric material in place of the conductive floating gate to store charge in a non-volatile manner. A triple layer dielectric formed of silicon dioxide, silicon nitride and silicon oxide (ONO) is sandwiched between a conductive control gate and a surface of a semi-conductive substrate above the memory cell channel. The cell is programmed by injecting electrons from the cell channel into the nitride, where they are trapped and stored in a limited region, and erased by injecting hot holes into the nitride. Several specific cell structures and arrays employing dielectric storage elements and are described in United States patent no. 6,925,007. Thus, while examples given in the present application may refer to floating gates, other charge storage structures may also be used. The present application is not limited to a particular charge storage structure.

[0012] As in most integrated circuit applications, the pressure to shrink the silicon substrate area required to implement some integrated circuit function also exists with flash EEPROM systems. It is continually desired to increase the amount of digital data that can be stored in a given area of a silicon substrate, in order to increase the storage capacity of a given size memory card and other types of packages, or to both increase capacity and decrease size. One way to increase the storage density of data is to store more than one bit of data per memory cell. This is accomplished by dividing a window of a floating gate charge level voltage range into more than two states. The use of four such states allows each cell to store two bits of data, eight states stores three bits of data per cell, and so on. A multiple state flash EEPROM structure and operation is described in United States patents nos. 5,043,940 and 5,172,338, which patents are incorporated herein by this reference.

[0013] Increased data density can also be achieved by reducing the physical size of the memory cells and/or the overall array. Shrinking the size of integrated circuits is commonly performed for all types of circuits as processing techniques improve over time to permit implementing smaller feature sizes. But there are usually limits of how far a given circuit layout can be shrunk in this manner, since there is often at least one feature that is limited as to how much it can be shrunk. When this happens, designers will turn to a new or different layout or architecture of the circuit being implemented in order to reduce the amount of silicon area required to perform its functions. The

shrinking of the above-described flash EEPROM integrated circuit systems can reach such limits.

[0014] One way to form small cells is to use a self-aligned Shallow Trench Isolation (STI) technique. This uses STI structures to isolate adjacent strings of floating gate cells such as those of NAND type memory arrays. According to this technique, a gate dielectric (tunnel dielectric) layer and floating gate polysilicon layer are formed first. Next, STI structures are formed by etching the gate dielectric and floating gate polysilicon layers and the underlying substrate to form trenches. These trenches are then filled with a suitable material (such as oxide) to form STI structures. The portions of the gate dielectric and floating gate polysilicon layers between STI structures are defined by the STI structures and are therefore considered to be self-aligned to the STI structures. Typically, the STI structures have a width that is equal to the minimum feature size that can be produced with the processing technology used. STI structures are also generally spaced apart by the minimum feature size. Thus, the portions of the gate dielectric and floating gate polysilicon layers between STI regions may also have a width that is equal to the minimum feature size. The strips of floating gate polysilicon are further formed into individual floating gates in later steps.

[0015] Another way to form small cells is to reduce the size of the features. However, lithographic processes used to establish the dimensions of devices are generally limited by some minimum feature size. Memory cells are generally designed to have dimensions that are equal to this minimum feature size (F). Thus, in Figure 2A, the width of NAND strings and the separation between adjacent NAND strings is approximately F. Also, the width of the word lines and separation between adjacent word lines is approximately F. In one technique, sidewall spacers are grown that are narrower than F and used to form word lines that are narrower than F. An example of such a technique is described in US Patent No. 6,888,755.

[0016] While memory cells within a memory array may be scaled down in size using various techniques (including providing features that are smaller than the minimum feature size), peripheral circuits may not always be so easily scaled. Peripheral circuits include various circuits that are on the same substrate as a memory array and

are used to manage the memory array. Examples of peripheral devices include word line decoder and driver circuits and bit line decoder and driver circuits. Peripheral circuits may have to withstand relatively large voltages so that they require relatively thick dielectric layers and relatively large device sizes. Because such peripheral circuits are not generally scaled down in size in proportion to the memory array, these peripheral circuits come to occupy an undesirably large area on a substrate.

SUMMARY

[0017] A memory array is formed using sidewall spacers to pattern a masking layer, sidewall spacers having a width that is approximately half the minimum feature size of the lithographic process used. The masking layer pattern is then used to form word lines that overlie and are self aligned to floating gates. Sidewalls are formed as rectangular shaped, closed loops, and so word lines are also formed as portions of rectangular shaped, closed loops that connect word lines of adjacent blocks (or in some cases, two word lines of the same block). Two word lines connected in this manner share word line decoder and driver circuits. The number of decoder and driver circuits may be reduced by half as a result of sharing by two word lines, thus providing a significant space saving.

[0018] A process using sidewalls to form narrow word lines also forms select lines using sidewalls to define the extent of select lines so that select lines are not separately aligned to the memory array but are defined by the same process steps that define word lines. Photoresist portions are used in formation of select lines, but the locations of edges of select lines are not determined by photoresist. Instead, locations of edges are determined by sidewalls of the same masking layer whose sidewalls establish word line locations. While photoresist portions are aligned to the masking layer that defines word lines, the alignment is not critical and a certain amount of misalignment may be tolerated because it does not affect the locations of features formed.

[0019] Forming contacts to narrow word lines (having a thickness less than the minimum feature size) may be difficult. Contact pads may be provided that have larger dimensions than the word line. Contact pads are formed by the same steps used

to form select lines. Thus, the locations of contact pads are determined by sidewalls of the same masking layer used to establish locations of word lines and select lines. Subsequently, photoresist portions are added in a step that is tolerant of misalignment.

[0020] In some examples, all NAND strings that have word lines connected together are erased together and thus form a single block. This results in a block that is similar to two conventional block-like units connected together by word lines. Alternatively, a more conventional block arrangement is achieved by providing shield plates for each block, thus making the blocks separately erasable. Shield plates allow different voltages to be coupled to floating gates of different blocks, even though the word lines of the different blocks are connected together and therefore have the same voltage. Thus, a sufficient voltage difference between floating gates and the substrate can be established for one block to allow erasing, while an adjacent block that shares the same word lines has a lower voltage difference between floating gates and the substrate and therefore does not undergo erasing.

BRIEF DESCRIPTION OF THE DRAWINGS

[0021] Figure 1 shows a memory system of the prior art.

[0022] Figure 2A shows a NAND flash memory array of the prior art.

[0023] Figure 2B shows a circuit diagram for the prior art NAND flash memory array of Figure 2A.

[0024] Figure 2C shows a cross section of a prior art NAND string of Figure 2A.

[0025] Figure 3 shows a cross section of a NAND flash memory array according to an embodiment of the present invention at an intermediate stage of fabrication.

[0026] Figure 4 shows a cross section of the NAND flash memory array of Figure 3 along a direction perpendicular to the cross section of Figure 3 with photoresist portions visible.

[0027] Figure 5 shows the structure of Figure 4 in the same view after photoresist slimming to reduce the width of photoresist portions.

[0028] Figure 6 shows the structure of Figure 5 after slimmed photoresist portions are used as an etch mask to pattern an underlying Silicon Nitride layer.

[0029] Figure 7 shows the structure of Figure 6 after deposition of a Silicon dioxide layer that overlies Silicon Nitride portions and exposed polysilicon.

[0030] Figure 8 shows the structure of Figure 7 after deposition of photoresist portions to cover areas of the Silicon dioxide layer.

[0031] Figure 9 shows the structure of Figure 8 after etching of the Silicon dioxide layer to remove portions that are not adjacent to sidewalls of Silicon Nitride portions or covered by photoresist and subsequent removal of photoresist portions.

[0032] Figure 10 shows the structure of Figure 9 after removal of Silicon Nitride portions leaving Silicon dioxide portions on polysilicon.

[0033] Figure 11 shows the structure of Figure 10 after using Silicon dioxide portions as an etch mask to pattern underlying polysilicon layers to form floating gates, select gates and word lines.

[0034] Figure 12 shows the structure of Figures 3 and 4 from above with markings B-B and C-C to show the views of Figures 3 and 4 respectively.

[0035] Figure 13 shows the structure of Figure 12 after photoresist slimming is performed to reduce the width of photoresist portions. Figure 13 shows the same stage of fabrication as Figure 5 from a different perspective.

[0036] Figure 14 shows the structure of Figure 13 after patterning of underlying Silicon Nitride using slimmed photoresist portions as an etch mask. Figure 14 shows the same stage of fabrication as Figure 6 from a different perspective.

[0037] Figure 15 shows the structure of Figure 14 after deposition of a Silicon dioxide layer and formation of photoresist portions to cover parts of the Silicon dioxide layer. Figure 15 shows the same stage of fabrication as Figure 8 from a different perspective.

[0038] Figure 16 shows the structure of Figure 15 after etching of the Silicon dioxide layer, removal of photoresist portions, removal of Silicon Nitride portions and etching of underlying polysilicon layers using Silicon dioxide portions as an etch mask. Figure 16 shows the same stage of fabrication as Figure 11, with additional source and drain contacts.

[0039] Figure 17 shows an equivalent circuit diagram for the structure of Figure 16 with bit lines that are not shown in Figure 17.

[0040] Figure 18 shows an alternative embodiment at an intermediate stage of fabrication including photoresist portions with openings to form contact pads.

[0041] Figure 19 shows the structure of Figure 18 after photoresist slimming and etching of Silicon Nitride using slimmed photoresist portions as an etch mask, deposition of a Silicon dioxide layer and subsequent formation of photoresist portions to cover parts of the Silicon dioxide layer.

[0042] Figure 20 shows the structure of Figure 19 after etching of the Silicon dioxide layer, removal of photoresist portions, removal of Silicon Nitride portions, etching of underlying polysilicon layers to form word lines and floating gates and formation of word line contacts to contact pads.

[0043] Figure 21 shows a cross section of a shield plate formed over a portion of a NAND array according to an embodiment of the present invention.

[0044] Figure 22 shows the portion of a NAND array with shield plate of Figure 21 in plan view and a second shield plate extending over a second portion of the NAND array.

[0045] Figure 23 shows a cross section of an alternative shield plate formed over a portion of a NAND array according to an embodiment of the present invention.

[0046] Figure 24 shows the portion of the NAND array with shield plate of Figure 23 in plan view and a second shield plate extending over a second portion of the NAND array.

[0047] Figure 25 shows a flowchart of a process for fabrication of a NAND flash memory array according to an embodiment of the present invention.

[0048] Figure 26 shows electrical connections to elements of a NAND array according to an embodiment of the present invention.

[0049] Figure 27A shows voltage values applied to elements of the NAND array of Figure 26 during a read operation.

[0050] Figure 27B shows voltage values applied to elements of the NAND array of Figure 26 during a program operation.

[0051] Figure 27C shows voltage values applied to elements of the NAND array of Figure 26 during an erase operation.

[0052] Figure 28 shows electrical connections to elements of a NAND array that has a shield plate according to an embodiment of the present invention.

[0053] Figure 29A shows voltage values applied to elements of the NAND array of Figure 28 during a read operation.

[0054] Figure 29B shows voltage values applied to elements of the NAND array of Figure 28 during a program operation.

[0055] Figure 29C shows voltage values applied to elements of the NAND array of Figure 28 during an erase operation.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

[0056] Figure 3 shows a cross section of a NAND array according to an embodiment of the present invention at an intermediate stage of fabrication. The formation of the NAND array up to this point may follow a conventional technique where a first dielectric layer 301 (gate oxide layer) is formed over a substrate 303 and subsequently a first polysilicon layer 305 is formed over first dielectric layer 301. First polysilicon layer 305 is doped so that it is electrically conductive. STI structures 307a-d are formed by patterning substrate 303 and etching trenches through first polysilicon

layer 305 and through first dielectric layer 301. The trenches also extend into substrate 303. The trenches are filled with STI material (a suitable dielectric material such as Silicon dioxide) to provide electrical insulation between devices. Thus, strips of STI material form STI structures 307a-d that extend across substrate 303 (in a direction perpendicular to the cross section of Figure 3) separated by strips 305a-c of first polysilicon layer 305. Both STI structures 307a-d and strips 305a-c of first polysilicon have a width that is the minimum feature size (F) of the process used for patterning. Subsequently, a second polysilicon layer 309 is deposited that overlies both STI structures 307a-d and strips 305a-c of first polysilicon material. Second polysilicon layer 309 is also doped and electrically conductive. Second polysilicon layer 309 is separated from strips 305a-c of first polysilicon by a second dielectric layer 311. Subsequently, a masking layer 313 is formed over second polysilicon layer 309. In this case, masking layer 313 is formed of a dielectric, Silicon Nitride (SiN), though other suitable masking materials may also be used.

[0057] Figure 4 shows a cross section of the NAND array of Figure 3 along a direction that is at right angles to the cross section of Figure 3. Thus, Figure 4 shows a single strip 305a of first polysilicon material in cross section with second polysilicon layer 309 overlying strip 305a. Figure 4 also shows portions 415a-e of photoresist overlying masking layer 313. Portions of photoresist 415a-e are formed by applying a blanket layer of photoresist and then patterning the photoresist using a lithographic process. Portions of photoresist 415a-e may be formed as strips having a width that is equal to the minimum feature size (F) of the lithographic process used. Portions of photoresist 415a-e may also be spaced apart by a distance that is equal to F. Other dimensions greater than F may also be used. While the present process uses photoresist that is patterned by being exposed to light, other patterning processes may also be used including e-beam lithography.

[0058] Figure 5 shows the NAND array of Figure 4, along the same cross section, after a resist slimming step is performed. Resist slimming involves subjecting portions of photoresist 415a-c to etching to remove at least some photoresist and so make portions of photoresist 415a-e narrower. A conventional etch may be used for this step, such as a dry etch. In the example shown, portions of photoresist 415a-e are

narrowed from an initial width equal to the minimum feature size (F) to about half the initial width (F/2). The distance between portions of photoresist 415a-e increases accordingly from an initial distance (F) to one and a half times the initial distance (3F/2).

[0059] Subsequent to resist slimming, the slimmed portions of photoresist are used to pattern the underlying Silicon Nitride masking layer 313. An etch is performed so that unexposed portions of masking layer 313 are removed, while those portions of masking layer 313 that are covered by portions of photoresist 415a-e are not removed. Portions of photoresist 415a-e are then removed. Figure 6 shows the resulting structure along the same cross section as Figure 5. The etch stops when second polysilicon layer 309 is reached so that second polysilicon layer 309 is not affected by this step. This patterning step transfers the pattern of the portions of photoresist 415a-e to masking layer 313 so that masking portions 313a-e are formed having a width of F/2 that are separated by 3F/2.

[0060] Figure 7 shows the structure of Figure 6, along the same cross section, after formation of a third dielectric layer 717 that overlies masking portions 313a-e and the exposed areas of the second polysilicon layer. In this example third dielectric layer 717 is formed of Silicon dioxide (SiO2 or "oxide"). Third dielectric layer 717 is formed as a blanket layer by a conventional process such as Chemical Vapor Deposition (CVD). Third dielectric layer 717 is generally a thicker dielectric layer than first dielectric layer 301 and second dielectric layer 311. Third dielectric layer 717 extends along second polysilicon layer 309 where it is exposed and extends along the top surfaces and sidewalls of masking portions 313a-e.

[0061] Figure 8 shows the structure of Figure 7, along the same cross section, after formation of photoresist portions 819a, 819b overlying portions of third dielectric layer 717. Photoresist portions 819a, 819b may be formed by covering the structure with photoresist, then patterning the photoresist using a lithographic process by removing unwanted portions of photoresist. Photoresist portions 819a, 819b extend over portions of third dielectric layer 717 that directly overlie second polysilicon layer 309. Subsequent to formation of photoresist portions 819a, 819b, an etch is carried out to remove certain exposed portions of third dielectric layer 717.

[0062] Figure 9 shows the structure of Figure 8, along the same cross section, after an etch step is carried out. The etch step may use anisotropic etching such as Reactive Ion Etching (RIE) so that third dielectric layer 717 is etched through in some places but portions 717a-f of third dielectric layer 717 remain along sidewalls of masking portions 313a-e because of the vertical thickness of third dielectric layer 717 in these locations. Remaining portions 717a-f of third dielectric layer 717 include portions 717b-e referred to as sidewall spacers because they are formed along sidewalls of masking portions 313b-d. The dimensions of sidewall spacers 717b-e are determined by the thickness of third dielectric layer 717 and by the nature of the anisotropic etch used. In this case, sidewall spacers 717b-e have a width of approximately half the minimum feature size (F/2), leaving gaps between sidewall spacers 717b-e that are also approximately half the minimum feature size. After the etch is completed, a photoresist strip step is also performed to remove photoresist portions 819a, 819b. This leaves wide dielectric portion 717a that extends between masking portions 313a and 313b and wide dielectric portion 717f that extends between masking portions 313d and 313e. The dimensions of wide dielectric portions 717a, 717f are determined by the locations of masking portions 313a, 313b, 313d and 313e, not by the dimensions of photoresist portions 819a, 819b. Thus, wide dielectric portions 717a, 717f, which subsequently establish the locations of select gate lines, are aligned with sidewall spacers 717b-e, which subsequently establish the locations of word lines, and do not require separate alignment, unlike many prior art schemes. The precise positioning of photoresist portions 819a, 819b is not critical to positioning of wide dielectric portions 717a, 717f. Photoresist portions 819a, 819b should extend from close to one sidewall to close to an adjacent sidewall but precise alignment is not required. Edges of photoresist portions 819a, 819b do not have to coincide with locations of sidewalls because the thicker dielectric layer in these areas ensures that the dielectric will not be etched through. In the present example, wide dielectric portions 717a, 717f have a width of approximately four times the minimum feature size (4F). Figure 9 shows a masking portion 313d having a width of X. In the present example, X is approximately F/2, though in other examples X may be greater than F/2. The distance X later establishes a distance between a floating gate and a select

gate, so this distance may be chosen separately and is not necessarily the same as the distance between floating gates.

[0063] Figure 10 shows the structure of Figure 9, along the same cross section, after removal of masking portions 313a-e. Sidewall spacers 717b-e and wide dielectric portions 717a, 717f remain in place overlying the second polysilicon layer 309. Subsequently, sidewall spacers 717b-e and wide dielectric portions 717a, 717f are used as an etch mask to pattern underlying layers to form the memory array.

[0064] Figure 11 shows the structure of Figure 10 along the same cross section after an etch step is carried out to etch through polysilicon strip 305a, second polysilicon layer 309 and second dielectric layer 311, stopping on first dielectric layer 301 or on substrate 303. This etch step separates second polysilicon layer 309 into separate word lines 309b-e and strips 309a, 309f. This etch also separates the strips of the first polysilicon layer into separate floating gates 305m-p. Word lines 309b-e form control gates where they overlie floating gates 305m-p. Because word lines 309b-e and floating gates 305m-p are formed by the same etch step they are self aligned. Portions 305*l* and 305q are also formed under strips 309a, 309f. Portion 305*l* is electrically connected to strip 309a to form a first select gate. Similarly, portion 305q is electrically connected to strip 309f to form a second select gate. Such self aligned structures as those shown in Figure 11 provide uniform coupling between floating gates and control gates and simplify fabrication. After the etch step is completed, source/drain regions 111a-e may be formed by implanting dopants into exposed areas of substrate 303. These exposed areas lie between floating gates 308m-p so that the source/drain regions connect memory cells of a string. After source/drain regions 111a-e are formed, the memory array may be covered by a protective layer such as a thick dielectric layer or other protective material. Sidewall spacers 717b-e and wide dielectric portions 717a, 717f may be removed prior to forming the protective layer or may remain in place when the protective layer is formed.

[0065] Figure 12 shows the NAND array of Figures 3-11 in plan view. The cross sections of Figures 3 and 4-11 are indicated in Figure 12 by B-B and C-C respectively. Figure 12 shows the NAND array at a stage of formation that corresponds to that shown in Figures 3 and 4. Photoresist portion 415c is shown

extending across the memory array in the X-direction and also in the Y direction to form a closed loop. In some memory arrays, several similar concentric loops may be used. The width of photoresist portion 415c forming the closed loop is F, the minimum feature size of the lithographic process used to form photoresist portion 415c. Between photoresist portions 415b-d are openings that also have a width of F. An opening 121a is formed between photoresist portions 415a and 415b that is wider than F. A similar opening 121b is formed between photoresist portions 415d and 415e. It can be seen that photoresist portion 415e of Figure 4 is not just a strip extending in the X-direction, but is a strip that extends in both X and Y directions to form a closed, rectangular shaped loop. It can also be seen that the structure of cross section C-C is formed with a mirror-image structure also formed using the same photoresist pattern. Figure 14 includes dotted lines that show the locations of STI structures 307a-d that underlie photoresist portions 415a-e, masking layer 313 and second polysilicon layer 309. STI regions 307a-d, along with first and second polysilicon layers 305, 309 are formed prior to forming photoresist portions 415a-e.

[0066] Figure 13 shows the structure of Figure 12, in the same plan view, after resist slimming. Thus, the structure of Figure 13 corresponds to the cross section shown in Figure 5. As can be seen, photoresist portions 415a-e have become narrower and openings between photoresist portions 415a-e have become correspondingly wider as a result of photoresist slimming. In this example, photoresist portions 415b-d are narrowed to a width of approximately F/2 while the openings between photoresist portions 415b-d are increased to a width of approximately 3F/2. Opening 121a between photoresist portions 415a and 415b has a width of approximately 4F. Opening 121b between photoresist portions 415d and 415e also has a width of 4F.

[0067] Figure 14 shows the structure of Figure 13 after the photoresist pattern of Figure 13 is transferred to masking layer 313 as a result of an etching step using photoresist portions 415a-e as an etch mask. Thus, the structure of Figure 14 corresponds to the cross section shown in Figure 6. Third dielectric layer 717 (not shown) is formed over masking portions 313a-e. Third dielectric layer 717 overlies both masking portions 313a-e and exposed areas of underlying second polysilicon layer 309 in openings between masking portions 313a-e.

[0068] Figure 15 shows the structure of Figure 14 after formation of photoresist portions 819a-d overlying third dielectric layer 717. Thus, the structure of Figure 15 corresponds to the cross section shown in Figure 8. As is shown, photoresist portions 819a-d are somewhat smaller than the openings between masking portions. Photoresist portions 819a-d do not have to be exactly aligned with openings between masking portions. Subsequent to forming photoresist portions 819a-d, an etch is carried out to remove portions of third dielectric layer 717, leaving sidewall spacers 717b-e and portions third dielectric layer 717 covered by photoresist portions 819a-d. Then, photoresist portions 819a, 819b are removed and masking portions 313a-e are removed. Subsequently, an etch is performed to etch the underlying first polysilicon layer 305 and second polysilicon layer 309 in the pattern of the remaining portions of third dielectric layer 717.

[0069] Figure 16 shows the resulting structure after etching first polysilicon layer 305 and second polysilicon layer 309. A series of concentric, rectangular shaped, closed loops are formed in the pattern of sidewall spacers 717b-e. Word lines 309b-e form portions of these loops. Word lines 309b-e have a width of F/2 and are spaced approximately F/2 apart. Underlying word lines 309b-e are STI structures 307a-d and, between STI structures 307a-d, floating gates formed from the first polysilicon layer 305. Because word lines 309b-e and floating gates are formed by the same etch step they are self aligned. Word lines 309b-e form control gates where they overlie floating gates. Word lines are connected together in pairs in the structure shown, with connections formed at either end of word lines by portions of loops that extend in the Y-direction. Thus, for example, word lines 309b and 161b are connected by portions 163b and 167b. Similarly, word lines 309c-e are connected with word lines 161c-e by portions 163c-e and portions 167c-e. The connecting portions 163b-e, 167b-e at word line ends may be formed of portions of both first polysilicon layer 305 and second polysilicon layer 309 joined together in a similar way to select lines. Also shown in Figure 16 are select lines 309a, 309f-h. Select lines 309a, 309f-h have a width that is approximately 4F. A NAND string is formed by a series of floating gate memory cells connected between two select lines. Thus, Figure 16 shows two units 168, 169 of NAND strings that have their word lines connected together but have separate select lines. Source and drain contacts are also shown in Figure 16. In the present example,

source contacts 165a-c are connected together to form a common source contact for all the NAND strings shown. Drain contacts 166a-f are connected to bit lines that run in the Y direction above the word lines.

[0070] While the structure of Figure 16 shows portions 163b-e, 167b-e that extend in the Y-direction to connect word lines 309b-e and word lines 161b-e together, the space occupied by such portions is generally not significant. In particular, a block generally extends much further in the X-direction than in the Y-direction so that extending a block by a small amount in the X-direction will not greatly affect the area occupied by the block.

[0071] Figure 17 shows a circuit diagram for the structure of Figure 16. Figure 17 shows three NAND strings connected above a common source line 171 to form the first unit 168, and three NAND strings connected below common source line 171 to form the second unit 169. In some cases, first unit 168 and second unit 169 are separately erasable, and so may be considered as separate blocks. Source select lines 309f, 309g and drain select lines 309a, 309h are connected to circuits that enable portions of the memory array to be separately accessed. However, these connections are not shown for clarity in Figure 17. In addition, word lines are connected to shared word line decoder and driver circuits 173 used to access the memory array. Because word lines of first unit 168 and second unit 169 are connected together, shared word line decoder and driver circuits 173 serve both units. In this way, the amount of space on a memory die that is devoted to word line decoder and driver circuits may be reduced by half compared with a memory die that does not share circuits in this manner. While word lines are connected between adjacent units 168 and 169, and word line decoder and driver circuits 173 are also shared between adjacent units 168 and 169, select gate driver circuits are not shared so that select lines 309a, 309f of unit 168 and select lines 309g, 309h of unit 169 are separately controlled.

Contact Pads

[0072] In some cases, forming good connections to word lines may be difficult because of the small size of the word lines. For example, where it is desired to form a plug to connect to a word line in the vertical direction, the plug will generally have a

diameter of F and so extends beyond a word line having a width of F/2 and may electrically contact a neighboring word line if there is any misalignment.

[0073] In order to provide good contacts to narrow word lines an embodiment of the present invention provides contact pads that are formed integrally with the word lines, the pads having dimensions that are greater than the width of the word line (F/2) and may be greater than the minimum feature size. Figure 18 shows a pattern of photoresist portions 181a-c according to this embodiment including additional openings 183a-d in photoresist portions 181a-c that are used to form contact pads. Figure 18 is similar to Figure 12 apart from the addition of openings 183a-d.

[0074] Figure 19 shows the structure of Figure 18 after resist slimming, transfer of the resist pattern to a masking layer and deposition of a third dielectric layer overlying the masking layer and a second polysilicon layer. Unlike the prior embodiment, Figure 19 shows photoresist portions 192 a-d overlying additional openings 183a-d. Figure 19 also shows photoresist portions 192e-h deposited over openings in the inner photoresist portion 181a and outer photoresist portion 181c as before. Photoresist portions 192a-h may be formed together in a single patterning step. Openings and additional openings are protected by photoresist portions 192a-h during subsequent etching of the third dielectric layer. The result is that, when remaining portions of the third dielectric layer are used to pattern a second polysilicon layer to form word lines, contact pads are formed that are connected to the word lines. Contact pads have dimensions that are greater than F/2 and may have dimensions greater than F.

[0075] Figure 20 shows the structure of Figure 19 after etching of a third dielectric layer, removal of photoresist portions 192a-h and etching of first and second polysilicon layers. Contact pads 201a-d are shown with plugs 203a-d formed in the vertical direction to connect word lines 205a-d, 207a-d to word line decoder and driver circuits. Plugs 203a-d may connect to conductive lines that are later formed at a higher level over the memory array. While the space occupied by the contact pads 201a-d appears significant in Figure 20, this drawing is not to scale. In real NAND memory arrays, a block extends much farther in the X-direction than the Y direction so that an increase of a few times F in the X-direction may not greatly increase the overall size of a block.

Shield Plate

[0076] In some embodiments, shield plates are formed subsequent to formation of separate word lines and floating gates. Sidewall spacers and wide dielectric portions are generally removed prior to formation of shield plates. Figure 21 shows the structure of Figure 11 along the same cross section after removal of sidewall spacers 717b-e and wide dielectric portions 717a, 717f and formation of a shield plate 211. A dielectric layer 213 is first formed that overlies source/drain regions, floating gates and word lines. Dielectric layer 213 provides electrical insulation on the surfaces of floating gates and word lines. Subsequently, conductive shield plate 211 is formed. In this example, conductive shield plate 211 is formed of doped polysilicon, though other conductive materials may also be used. Figure 22 shows the structure of Figure 21 in plan view. Separate conductive shield plates 211, 215 are formed corresponding to unit 168 and unit 169 respectively. Shield plates 211, 215 are not in electrical contact with floating gates or word lines because dielectric layer 213 separates shield plates 211, 215 from floating gates and word lines. However, a shield plate is capacitively coupled to floating gates so that the voltage of a shield plate may be used to modify the voltage of a floating gate. Shield plates 211, 215 allow separate erasing of unit 168 and unit 169 and thus define blocks in the memory array. In an alternative embodiment, unit 168 and unit 169 may only be erasable together and so form a single block.

[0077] A shield plate may extend over and between adjacent word lines and floating gates as shown in Figures 21-22 or may extend between adjacent word lines and floating gates but not extend over them as shown in Figure 23. The structure of Figure 23 may be achieved by applying Chemical Mechanical Polishing (CMP) or an etch-back process to the structure shown in Figure 21. Figure 24 shows the structure of Figure 23 in plan view. Separate conductive portions 233a-e shown in Figure 24 may be electrically connected together so that they form a single conductive unit. Formation and use of shield plates in NAND memory arrays are described in US Patent Application Publication No. 2005/0180186 entitled, "Shield plate for limiting cross coupling between floating gates." Shielding in NAND arrays is also described in US Patent Application Publication No. 2005/0072999, entitled "Bitline direction

shielding to avoid cross coupling between adjacent cells for NAND flash memory." In the present application, shield plates are not connected to source/drain regions and are separately controlled so that a voltage may be applied to a shield plate. Thus, a shield plate generally has a connection to a shield plate driver circuit that controls the voltage of the shield plate.

[0078] Figure 25 shows a flowchart for fabricating a memory array according to an embodiment of the present invention. A first dielectric layer and first polysilicon layer are formed on a substrate surface 255a. The polysilicon is generally deposited so that it is doped and therefore electrically conductive. In some cases, the first polysilicon layer may be deposited undoped and later doped. STI structures are formed 255b by patterning the substrate, forming trenches that extend into the substrate. The trenches are filled with STI material such as Silicon dioxide. Subsequently, a second doped polysilicon layer is deposited 255c that overlies both STI structures and first polysilicon portions. The first and second polysilicon portions are separated by a second dielectric layer. A Silicon Nitride masking layer is formed over the second polysilicon layer. Next, a photoresist layer is formed over the Silicon Nitride layer 255d and is patterned into photoresist portions that include one or more concentric, rectangular shaped strips. A slimming process is performed 255e on the photoresist portions so that the thickness of the strips is reduced to about half the minimum feature size of the lithographic process used. Then the slimmed photoresist portions are used to pattern the Silicon Nitride layer into portions that also have a width that is approximately half the minimum feature size 255f. A Silicon dioxide layer is then formed 255g over both the Silicon Nitride portions and the exposed portions of the second polysilicon layer. Photoresist portions are formed 255h to cover parts of the Silicon dioxide layer that are to be protected. Then an anisotropic etch is performed 255i that removes Silicon dioxide except in locations close to sidewalls of Silicon Nitride portions and locations covered by photoresist portions. Photoresist portions cover areas that later become source select lines and drain select lines. In some examples, contact pad areas are also covered by photoresist portions. Photoresist portions are then removed 255j and Silicon Nitride portions are also removed 255k leaving Silicon dioxide portions on the second polysilicon layer. These Silicon dioxide portions are then used as a mask to etch the first and second polysilicon layers

to form word lines and floating gates 255l. In some cases, shield plates are then formed 255m by depositing a dielectric layer followed by a third polysilicon layer to form a conductive polysilicon plate.

Memory Operation

[0079] The operation of a memory array according to an embodiment of the present invention that does not use a shield plate will now be described. Figure 26 shows a portion of a NAND memory array made up of four units, unit A, A+1, A+2, and A+3. Each of the units A through A+3 contain many NAND strings with 32 memory cells in each NAND string. Unit A and unit A+1 are similar in structure to units 168, 169 of Figure 16 except that they have 32 word lines instead of 4, and have many NAND strings. The NAND strings of a unit extend between source select gates and drain select gates. Thus, unit A comprises many NAND strings extending between a drain select gate line (SGD Unit A) and a source select gate line (SGS Unit A). In many prior NAND arrays, a block consisted of a similar grouping of NAND strings to that of Unit A. However, the term "block" is generally used to describe the minimum unit of erase of a memory array and because unit A is not separately erasable, the term "unit" is used instead of the term "block." However, it will be understood that unit A resembles a block of a prior art NAND array and has a similar structure. Unit A has 32 word lines, WL0-WL31 extending in the X-direction. Only a few word lines (WL0, WLn, WLn+1 and WL31) are shown for clarity, with WLn being a representative word line. As is shown in Figure 26, word lines WL0-WL31 of unit A are connected to WL0-WL31 of unit A+1. Such connections may be formed as described earlier in the present application, or in some other manner. Word lines WL0-WL31 may be connected at both ends, forming concentric, closed, rectangular shaped loops. For clarity, connecting portions between units are only shown at one end. Adjacent units, such as units A and A+1, are separately selectable by using select gates, however, because word lines of adjacent units are joined together, such adjacent units are not separately erasable in the present example. Thus, unit A and unit A+1 are erased together and so form a block. Similarly, unit A+2 and unit A+3 are erased together and so form a block. A common source line 261 extends between adjacent units A and A+1, and another common source line 263 extends between units

A+2 and A+3. Bit line connections 265 are formed between unit A+1 and A+2 and also between unit A and an adjacent unit (not shown).

[0080] Figure 27A shows the voltages that are applied to various elements of the memory array of Figure 26 in order to perform a read operation. A read operation determines the logical states of memory cells based on the amount of charge stored in floating gates. In this case, the read operation is performed on cells of word line WLn of unit A+1. Thus, only floating gate memory cells in unit A+1 that underlie word line WLn are read in this operation. Other floating gate memory cells are not read at this time. The drain and source select gates (SGD and SGS) are set to VSS (0volts) for all units except unit A+1 so that select transistors are turned off for all units except unit A+1. Drain and source select gates for unit A+1 are set to VSG (4.5volts) to turn on select transistors for unit A+1. In this way, bit lines that serve units A through A+3 (and other units) are electrically connected only to NAND strings of unit A+1. Word lines WLn of units A and A+1 have VSS (approximately 0volts) applied, while all other word lines (WL0 to WLn-1 and WLn+1 to WL31) of units A and A+1 have VREAD (approximately 4.5volts) applied. The result is that floating gate transistors of all word lines except WLn are turned on so that NAND strings are conductive apart from floating gate transistors underlying WLn. The state of floating gate transistors under WLn may thus be read by passing a current through NAND strings and measuring the effect of charge in the floating gates under WLn. Voltages on word lines of units other than units A and A+1 may be allowed to float during reading of unit A+1. Data is read out through bit lines with the common source lines ("Array source") held at VSS (0volts) and the P-well in the substrate also at VSS.

[0081] Figure 27B shows the voltages that are applied to various elements of the memory array of Figure 26 in order to perform a program operation. The program operation adds charge to floating gates of memory cells to change the logical state of the memory cell. In this case, the programming operation is performed on memory cells of WLn of unit A+1. This means that the floating gate memory cells underlying WLn, in unit A+1 are programmed, while other cells are not programmed. The drain and source select gates (SGD and SGS) are set to VSS (0volts) for all units except unit A+1 so that select transistors are turned off for all units except unit A+1. The

source select gate for unit A+1 is also set to 0volts. The drain select gate for unit A+1 is set to Vdd (approximately 2.5volts). All word lines except for WLn are set to a voltage of VPASS (approximately 10volts). WLn receives a voltage of VPGM (approximately 20volts). This relatively high voltage may cause electrons to enter the floating gate from the substrate depending on a programming voltage supplied by a bit line. In a common arrangement, programming voltages are applied by bit lines as a series of pulses. In one example, 0volts applied to a bit line causes programming of a cell. Cells that have reached their desired voltages receive a bit line voltage of Vdd to inhibit further programming. Programming may involve multiple pulsing and verifying steps. Word lines of units other than units A and A+1 are allowed to float during programming of unit A+1. In some examples, bit lines may be used to enable or inhibit programming of individual cells along WLn as cells reach their desired state. Cells of a word line may be programmed together, or in groups.

[0082] Figure 27C shows the voltages that are applied to various elements of the memory array of Figure 26 during an erase operation. The erase operation removes charge from floating gates to return them to a base level of charge that allows them to be programmed again. Charge is removed by creating an electrical field between word lines and the substrate that causes charge to flow from the floating gate to the substrate. An appropriate electrical field is created by appropriately biasing word lines and the underlying P-well in the substrate. Because word lines of units A and A+1 are connected together in Figure 26, the memory cells of units A and A+1 are erased together. Drain and source select gates are allowed to float during erase. All word lines WL0-WL31 are set to VSS (approximately 0volts), while the voltage of the P well in the substrate is set to VERA (approximately 20volts). Thus, a 20volt difference exists between a word line above a floating gate and the substrate below the floating gate. Floating gates in both units A and A+1 are subject to this electrical field and so they are both erased as a single block. Word lines of other units such as units A+2 and A+3 are allowed to float so no erase occurs in units A+2 and A+3.

[0083] Figure 28 shows an alternative embodiment where shield plates are used to establish two separately erasable blocks that share the same word lines. Figure 28 shows blocks B, B+1, B+2 and B+3 connected in a manner similar to units A, A+1,

A+2 and A+3 of Figure 26, but with shield plates provided. The addition of shield plates allows separate erasing of blocks because an electric field sufficient to cause charge to be removed from a floating gate may be created by applying suitable voltages to the separate shield plates of different blocks.

[0084] Figure 29A shows the voltages that are applied to various elements of the memory array of Figure 28 during a read operation. These voltages are the same as those given in Figure 27A except for the addition of shield plate voltages for each block. The shield plate voltage for block n that is not being read is VRSP (approximately 4.0volts). The shield plate voltage for block n+1, which is being read, is also VRSP. While a value of VRSP in this example is approximately 4.0volts, in other examples, VRSP may be set at a different voltage, or may be allowed to float. The shield plate voltages for all other blocks, such as blocks n+2 and n+3 may be allowed to float.

[0085] Figure 29B shows the voltages that are applied to various elements of the memory array of Figure 28 during a program operation. These voltages are the same as those given in Figure 27B except for the addition of shield plate voltages for each block. The shield plate voltage for block B, which is not being programmed, is set at VPSP (approximately 10volts). The voltage shield plate voltage for block B+1, which is being programmed, is also set at VPSP (approximately 10.0volts). Other values of VPSP may also be used. Shield plates of other blocks, such as blocks B+2 and B+3 may be allowed to float during programming of block B+1.

[0086] Figure 29C shows the voltages that are applied to various elements of the memory of Figure 28 during an erase operation. These voltages are the same as those given in Figure 27C except for the addition of shield plate voltages. Significantly, the addition of shield plates and the application of shield plate voltages allows individual blocks to be erased in the present embodiment, where in the embodiment of Figure 27C both units A and A+1 were used together as a single block. A voltage of VEISP (approximately 18volts) is applied to the shield plate of block B, which is not being erased. The application of this high voltage inhibits erase of block B by coupling a high voltage to floating gates and thereby reducing the voltage difference between floating gates and the substrate that would cause charge flow. A voltage of VESSP

(approximately 5volts) is applied to the shield plate of block B+1, which is being erased. This relatively low voltage couples to the floating gates, keeping floating gate voltage relatively low, so a large voltage difference exists between floating gates and the P-well in the substrate (at 20volts). This large voltage difference causes charge to flow from the floating gate to the substrate. Thus, the addition of a shield plate allows separate erase of different blocks that share the same word lines.

[0087] Although the various aspects of the present invention have been described with respect to exemplary embodiments thereof, it will be understood that the present invention is entitled to protection within the full scope of the appended claims.

CLAIMS:

1. A method of forming a nonvolatile memory array on a substrate surface comprising:

forming a plurality of shallow trench isolation structures on a substrate, individual ones of the plurality of shallow trench isolation structures extending in a first direction, ones of the plurality of shallow trench isolation structures spaced apart in a second direction; and

subsequently forming a plurality continuous conductive strips, an individual continuous conductive strip including first, second, third and fourth conductive portions, the first conductive portion extending in the second direction and overlying the plurality of shallow trench isolation structures, the second conductive portion extending in the second direction and overlying the plurality of shallow trench isolation structures, the first and second conductive strips spaced apart in the first direction, the third and fourth conductive portions extending in the first direction to connect the first and second conductive portions.

2. The method of claim 1 wherein the individual conductive strip forms a closed rectangle, and the first, second, third and fourth conductive portions form the sides of the rectangle.

3. The method of claim 1 wherein the first conductive portion forms a first word line and the second conductive portion forms a second word line.

4. The method of claim 1 wherein a width of an individual one of the plurality of continuous conductive strips is established by spacers formed on sidewalls.

5. The method of claim 4 wherein the width is less than a minimum feature size of a lithographic process used to form the nonvolatile memory array.

6. The method of claim 5 wherein the spacers are slimmed to be less than the minimum feature size of the lithographic process used.

7. The method of claim 1 further comprising forming select gate lines that extend in the first direction, the select gate lines being at least as wide as the minimum feature size.

8. The method of claim 1 further comprising forming floating gates underlying the plurality of first conductive portions so that a first plurality of NAND strings is formed between the plurality of shallow trench isolation structures.

9. The method of claim 8 wherein the first plurality of NAND strings extend between a first select gate line and a second select gate line.

10. The method of claim 9 further comprising forming a first conductive shield plate that extends over the first plurality of NAND strings.

11. The method of claim 10 further comprising forming floating gates underlying the plurality of second conductive portions so that a second plurality of NAND strings is formed between the plurality of shallow trench isolation structures, the second plurality of NAND strings extending between a third select gate line and a fourth select gate line and a second conductive shield plate extending over the second plurality of NAND strings.

12. The method of claim 1 further comprising forming a conductive shield plate that extends between the plurality of continuous conductive strips.

13. A method of forming a nonvolatile memory array comprising:

forming a masking layer that includes a plurality of masking portions extending over a substrate;

forming a plurality of sidewall spacers along sidewalls of the plurality of masking portions;

forming a plurality of word lines extending across the nonvolatile memory array in a first direction and spaced apart in a second direction, the second direction being perpendicular to the first direction, individual word lines having a first width in

the second direction that is less than a minimum feature size of a lithographic process used to form the nonvolatile memory array, the first width established by the width of a sidewall spacer of a plurality of sidewall spacers, an individual word line overlying floating gates of the nonvolatile memory array, individual ones of the plurality of word lines separated from adjacent ones of the plurality of word lines by a distance that is less than the minimum feature size produced by a lithographic process used to form the masking layer; and

forming a plurality of select gate lines extending across the nonvolatile memory array in the first direction, individual ones of the plurality of select gate lines having a second width that is at least as great as the minimum feature size, the second width established by the distance between two sidewall spacers of the plurality of sidewall spacers.

14. The method of claim 13 wherein individual word lines are linked in pairs by conductive elements extending in the second direction.

15. The method of claim 13 further comprising a conductive shield plate that extends between the plurality of word lines.

16. The method of claim 13 wherein the masking portions are formed by lithographically patterning a photoresist layer to form photoresist portions having a width that is equal to a minimum feature size and subsequently slimming the photoresist portions.

17. A method of operating a NAND flash memory array that has word lines of a first block connected to word lines of a second block, the first block and the second block sharing a plurality of bit lines, the first block having a first shield plate and the second block having a second shield plate, comprising:

erasing data in the first block while maintaining data in the second block by applying a first voltage to the first shield plate while applying a second voltage to the second shield plate.

18. The method of claim 17 further comprising maintaining a common voltage on the word lines of the first block and word lines of the second block during the erasing.

19. The method of claim 17 wherein a portion of the substrate underlying the first block and the second block is maintained at a third voltage during erasing, the third voltage being higher than the first or second voltages.

20. The method of claim 19 wherein the first voltage is approximately 5volts, the second voltage is approximately 18volts and the third voltage is approximately 20volts.

21. The method of claim 17 further comprising writing data in the first block while not writing data in the second block by turning on select transistors of the first block while turning off select transistors of the second block and applying programming voltages to the plurality of bit lines.

22. The method of claim 21 further comprising maintaining the first shield plate and the second shield plate at the same voltage during the writing data in the first block.

23. The method of claim 17 further comprising reading data from the first block while not reading data from the second block by turning on select transistors of the first block while turning off select transistors of the second block and reading data through the plurality of bit lines.

24. The method of claim 23 further comprising maintaining the first shield plate and the second shield plate at the same voltage during the reading data from the first block.

25. A nonvolatile floating gate memory array comprising:

a first plurality of floating gate memory cells having a first plurality of word lines extending in a first direction;

a second plurality of floating gate memory cells having a second plurality of word lines extending in the first direction;

the first plurality of floating gate memory cells separated from the second plurality of memory cells in a second direction that is perpendicular to the first direction; and

the first plurality of word lines electrically connected to the second plurality of word lines by conductive portions extending in the second direction.

26. The nonvolatile floating gate memory array of claim 25 wherein the first plurality of word lines, the second plurality of word lines and the conductive portions form a plurality of concentric rectangles.

27. The nonvolatile floating gate memory array of claim 26 wherein individual ones of the first plurality of word lines and the second plurality of word lines have a width that is less than a minimum feature size that is produced by a lithographic process used to form the memory array.

28. The nonvolatile floating gate memory array of claim 25 further comprising select gate lines to separately select the first plurality of floating gate memory cells and the second plurality of floating gate memory cells.

29. The nonvolatile floating gate memory array of claim 25 further comprising a first shield plate overlying the first plurality of floating gate memory cells and a second shield plate overlying the second plurality of floating gate memory cells.

30. The nonvolatile floating gate memory array of claim 29 further comprising a first driver circuit connected to the first shield plate and a second driver circuit connected to the second shield plate.

31. The nonvolatile floating gate memory array of claim 25 wherein the first plurality of word lines and the second plurality of word lines share word line decoder and driver circuits.

32. A nonvolatile floating gate memory array comprising:

a first plurality of strings of floating gate memory cells, the first plurality of strings having a first source select line connecting source select gates of the first plurality of strings and having a first drain select line connecting drain select gates of the first plurality of strings;

a second plurality of strings of floating gate memory cells, the second plurality of strings having a second source select line connecting source select gates of the second plurality of strings and having a second drain select line connecting drain select gates of the second plurality of strings;

a plurality of conductive portions extending across the first plurality of strings to connect control gates of the first plurality of strings and extending across the second plurality of strings to connect control gates of the second plurality of strings; and

a plurality of bit lines shared by the first plurality of strings and the second plurality of strings so that an individual bit line of the plurality of bit lines connects to a string of the first plurality of strings and connects to a string of the second plurality of strings.

33. The nonvolatile floating gate memory array of claim 32 wherein the plurality of conductive portions form a plurality of concentric rectangles.

34. The nonvolatile floating gate memory array of claim 32 wherein individual ones of the plurality of conductive portions have a width that is less than a minimum feature size of a lithographic process used to form the memory array.

35. The nonvolatile floating gate memory array of claim 32 wherein the plurality of conductive portions are connected to driver circuits that are shared by the first plurality of strings and the second plurality of strings.

36. The nonvolatile floating gate memory array of claim 32 wherein applying voltages to erase data in the first plurality of strings of floating gate memory cells causes data in the second plurality of strings to be erased also.

37. The nonvolatile floating gate memory array of claim 32 wherein a first shield plate is capacitively coupled to floating gates of the first plurality of strings of floating gate memory cells and a second shield plate is capacitively coupled to floating gates of the second plurality of strings of floating gate memory cells, and wherein data is erased from the first plurality of strings of floating gate memory cells without erasing data from the second plurality of strings of floating gate memory cells by applying a first voltage to the first shield plate while applying a second voltage to the second shield plate.

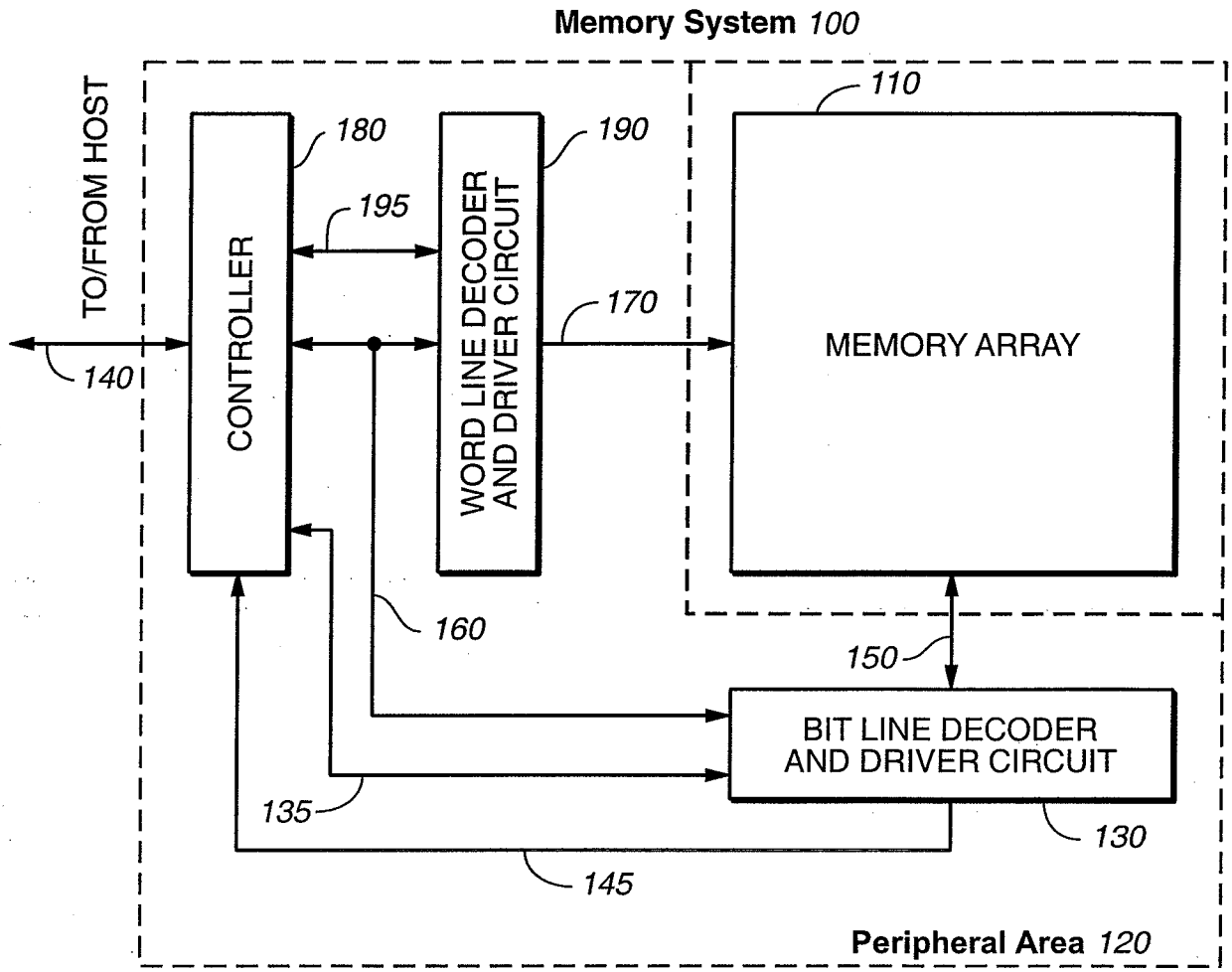


FIG. 1
(PRIOR ART)

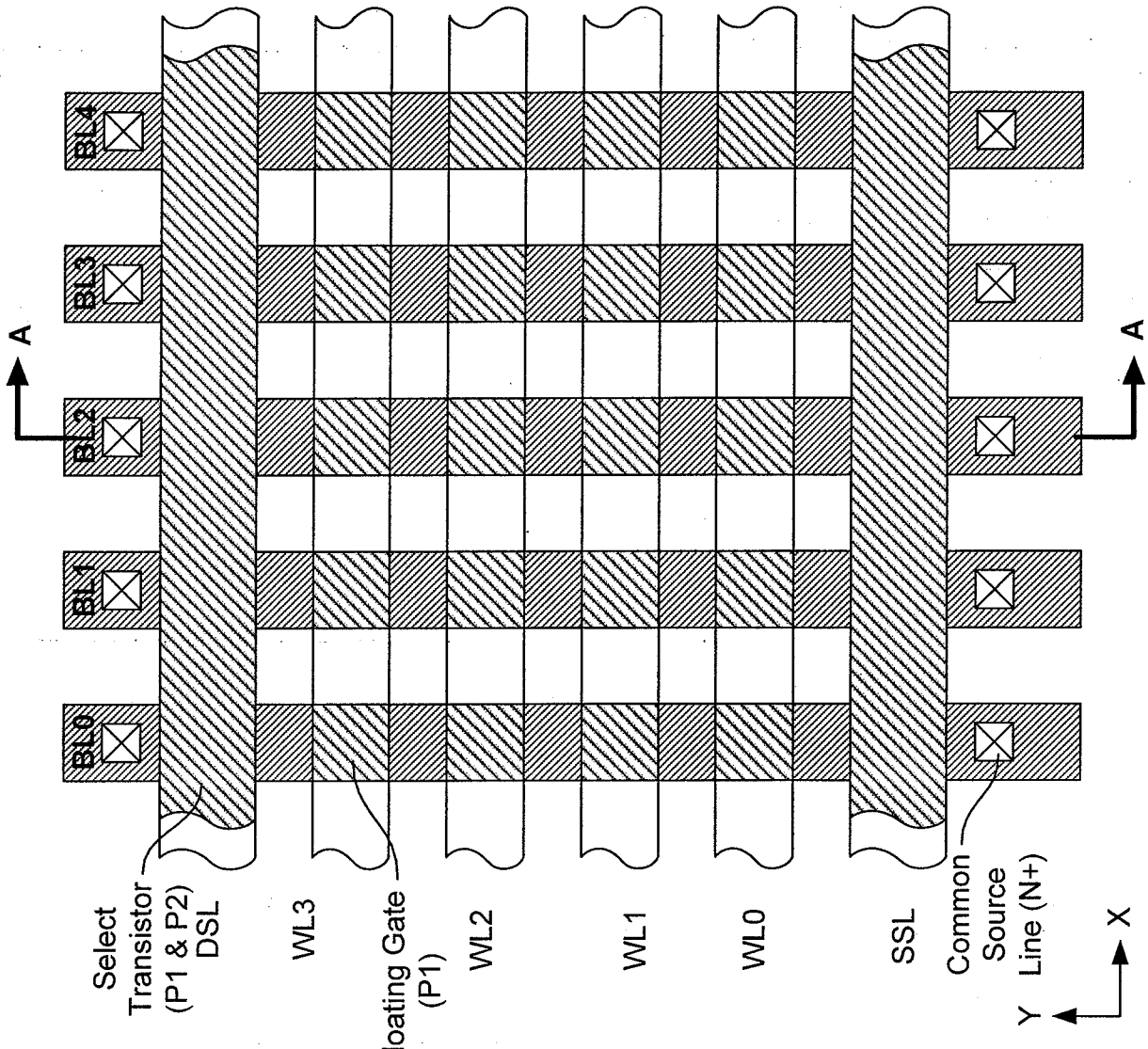


FIG. 2A
(PRIOR ART)

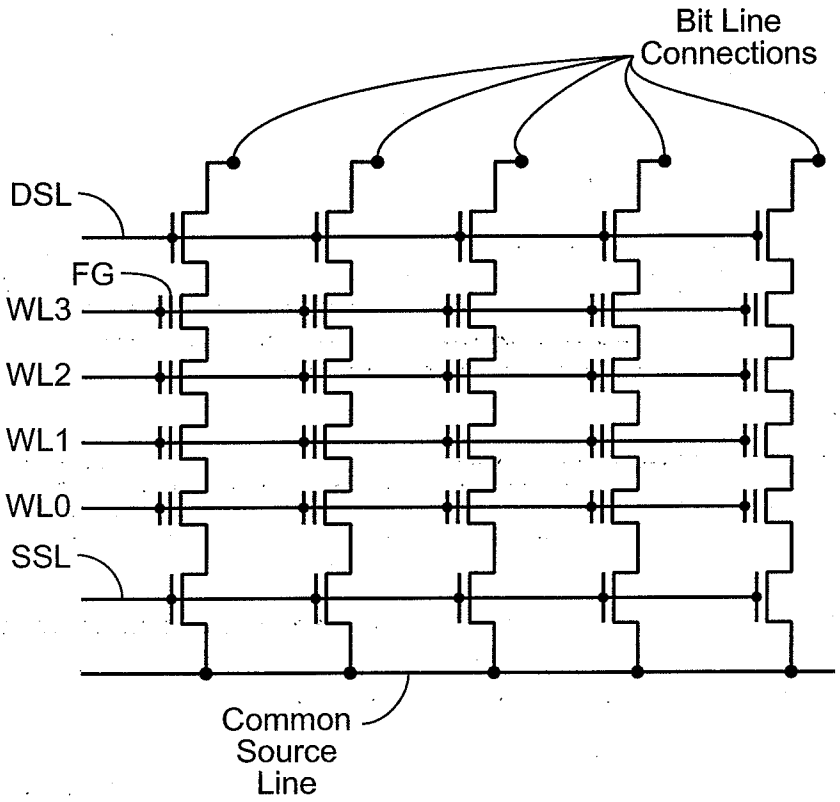


FIG. 2B
(PRIOR ART)

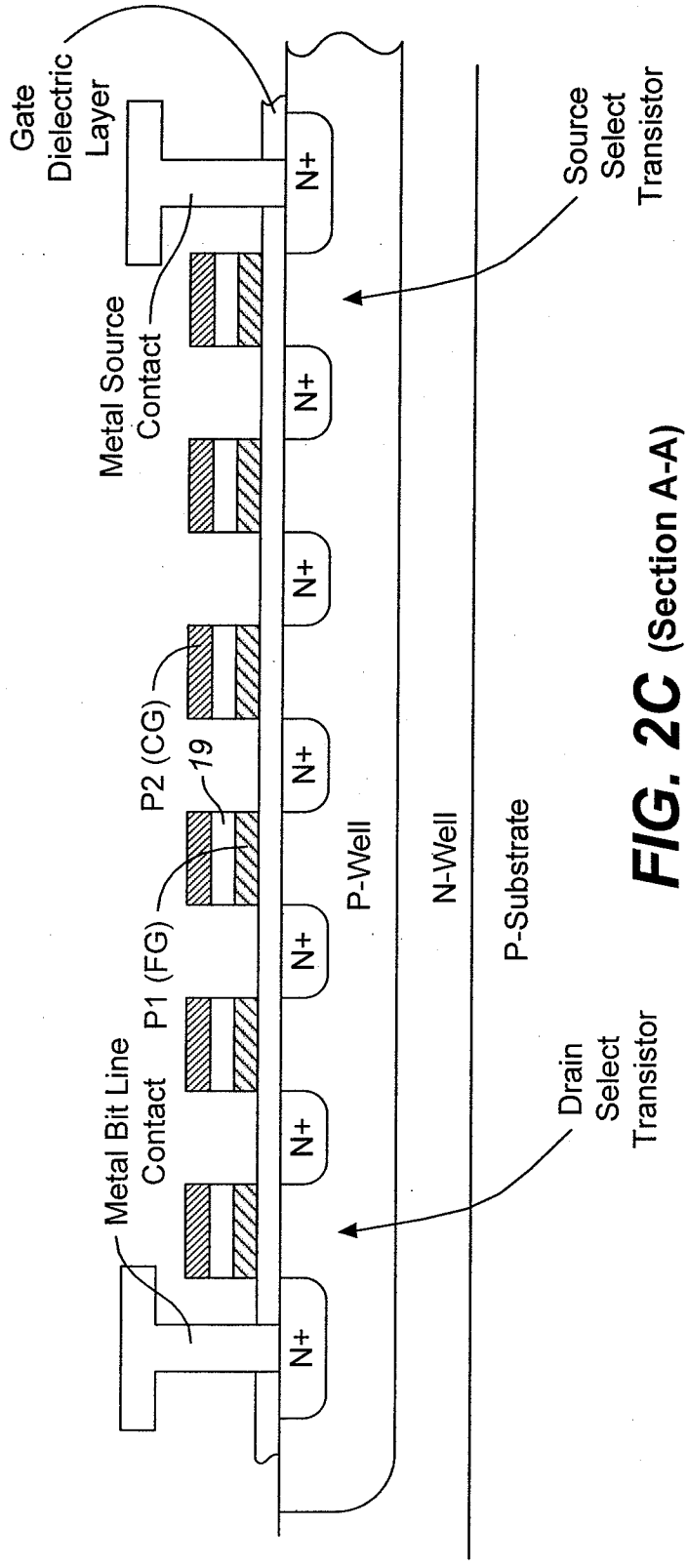


FIG. 2C (Section A-A)
(PRIOR ART)

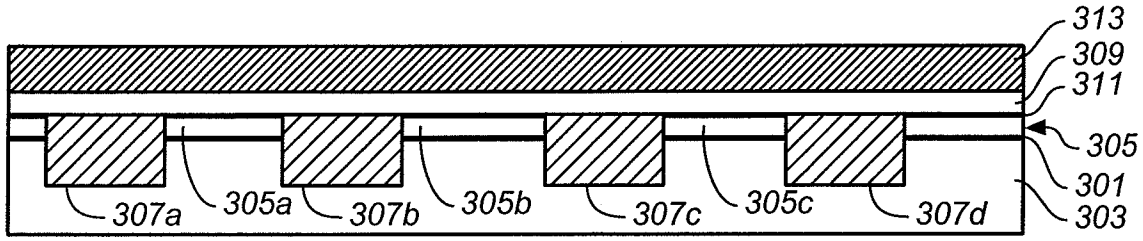


FIG. 3

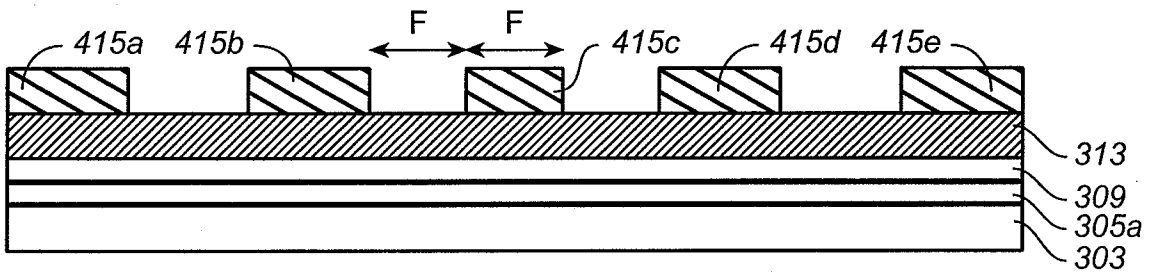


FIG. 4

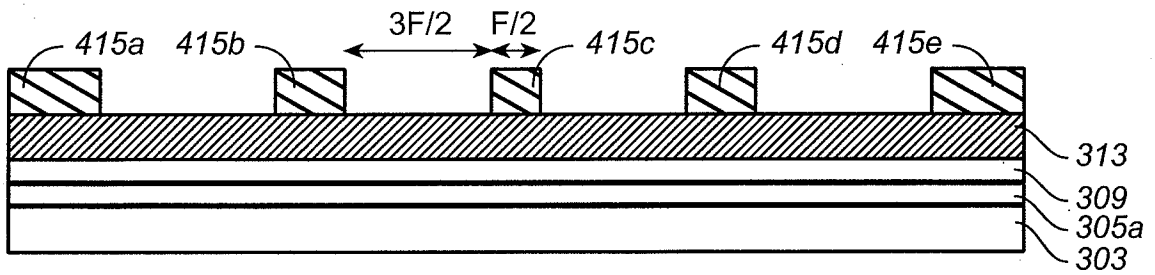


FIG. 5

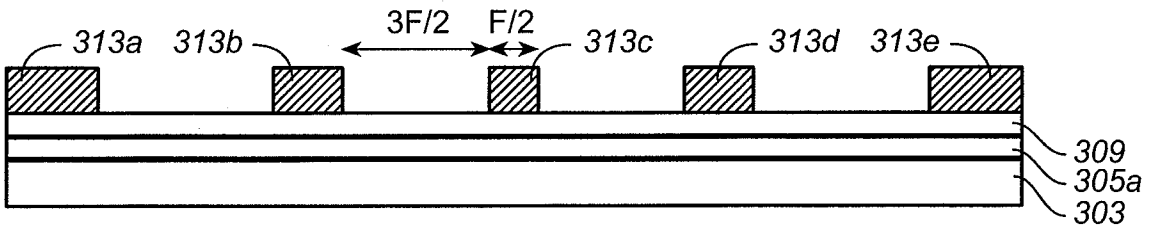


FIG. 6

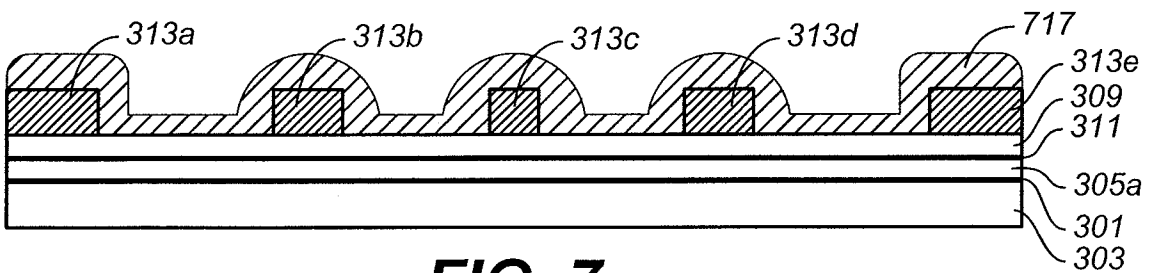


FIG. 7

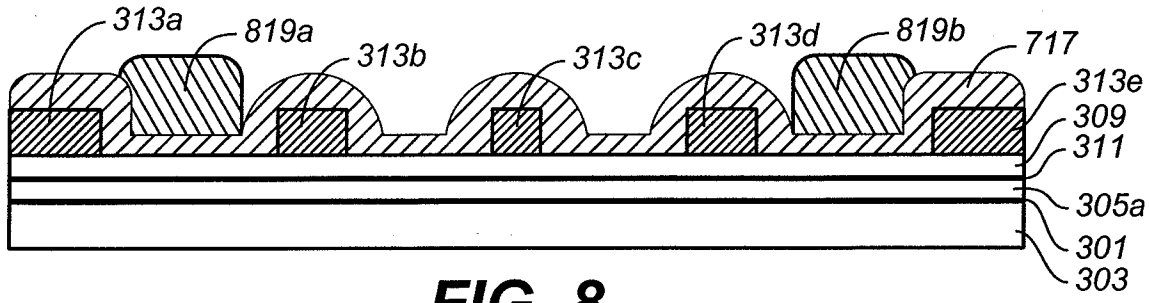


FIG. 8

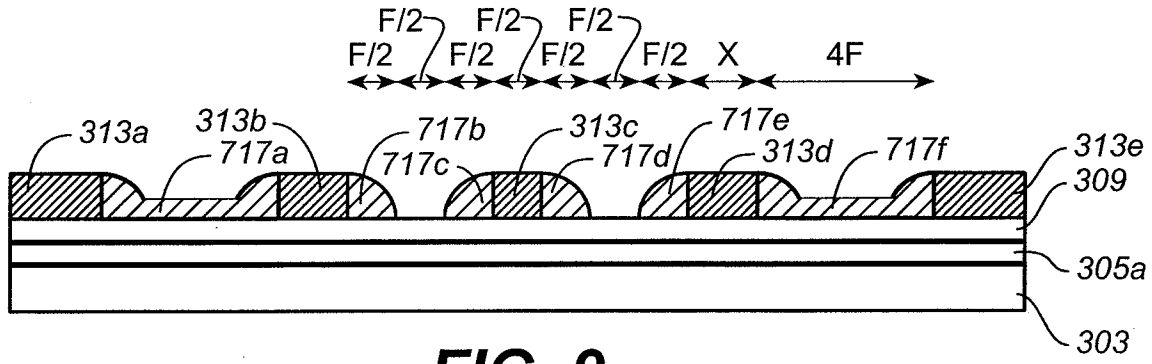


FIG. 9

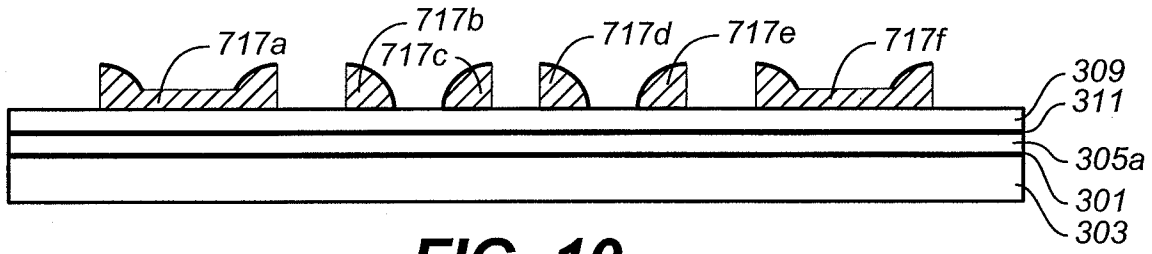


FIG. 10

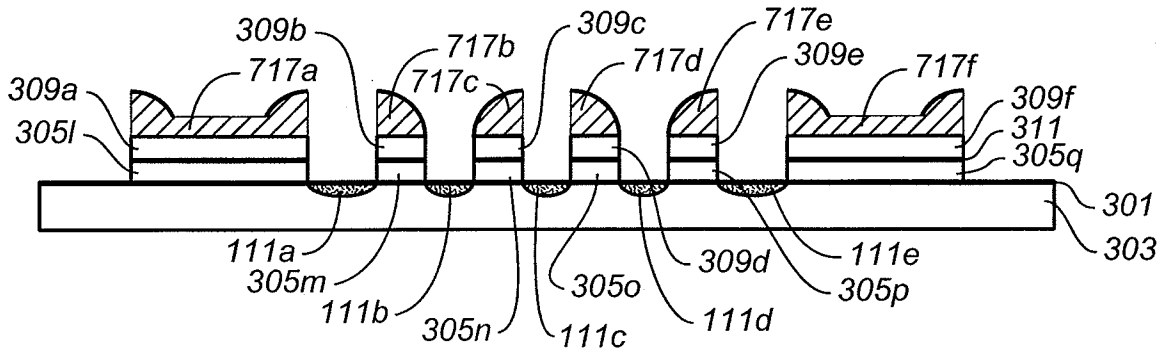


FIG. 11

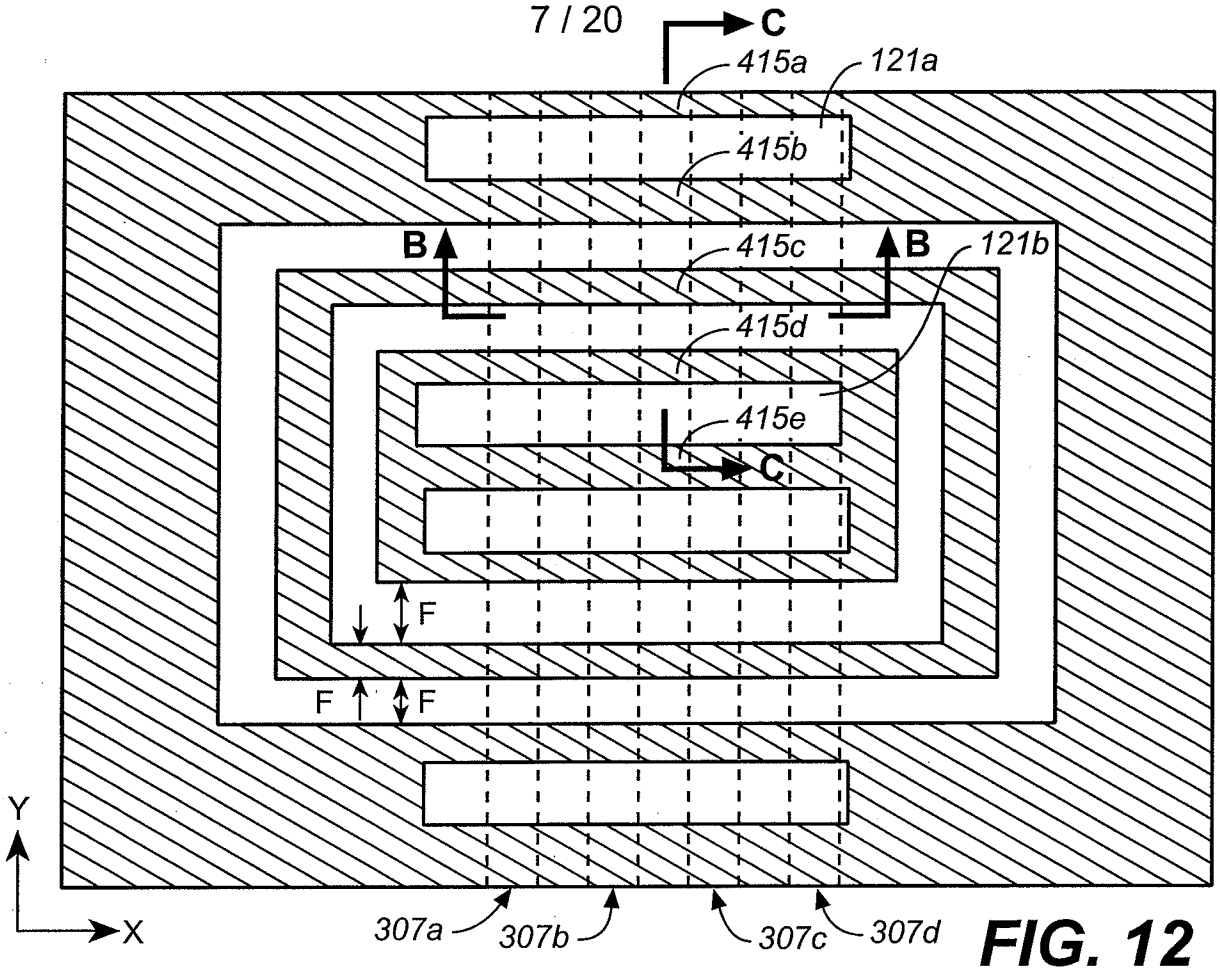

C
415a
121a
415b
B
B
415c
121b
415d
415e
C
F
F
F
Y
X
307a
307b
307c
307d

FIG. 12

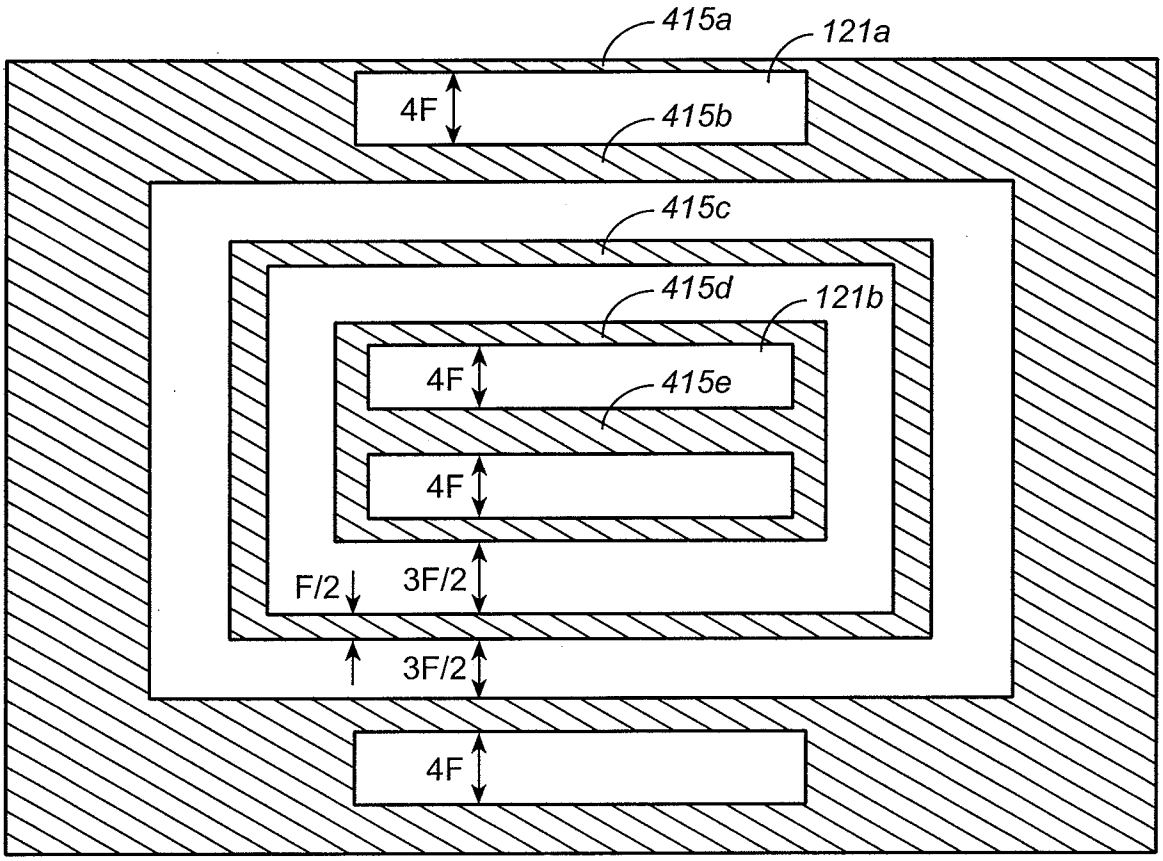

415a
121a
4F
415b
415c
415d
121b
4F
415e
4F
F/2
3F/2
3F/2
4F

FIG. 13

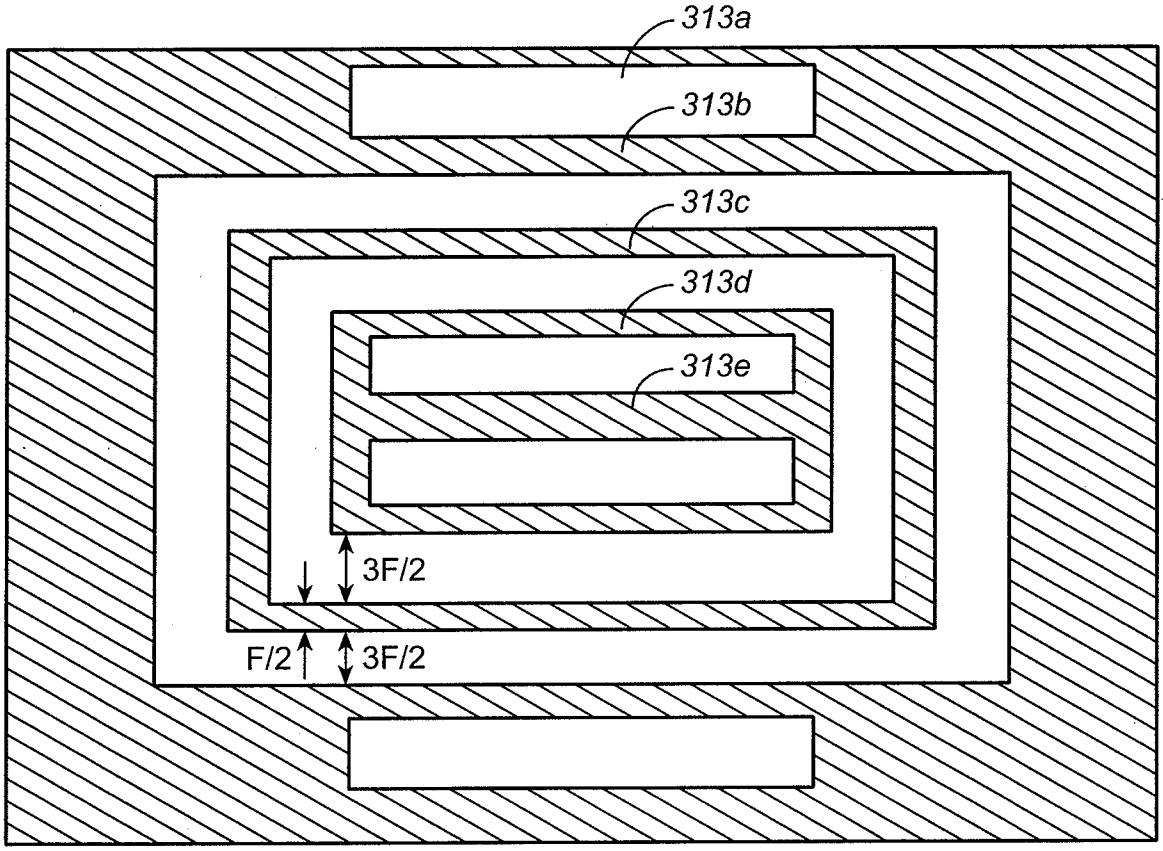


FIG. 14

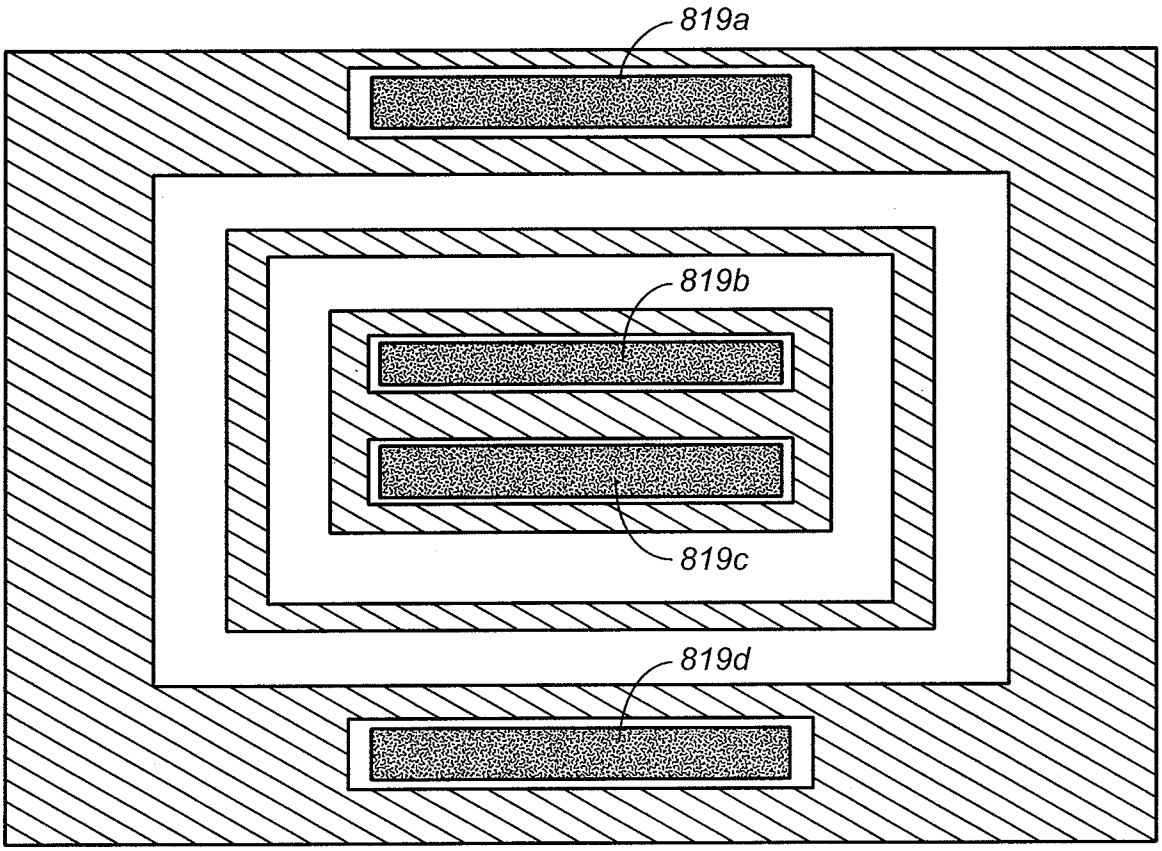


FIG. 15

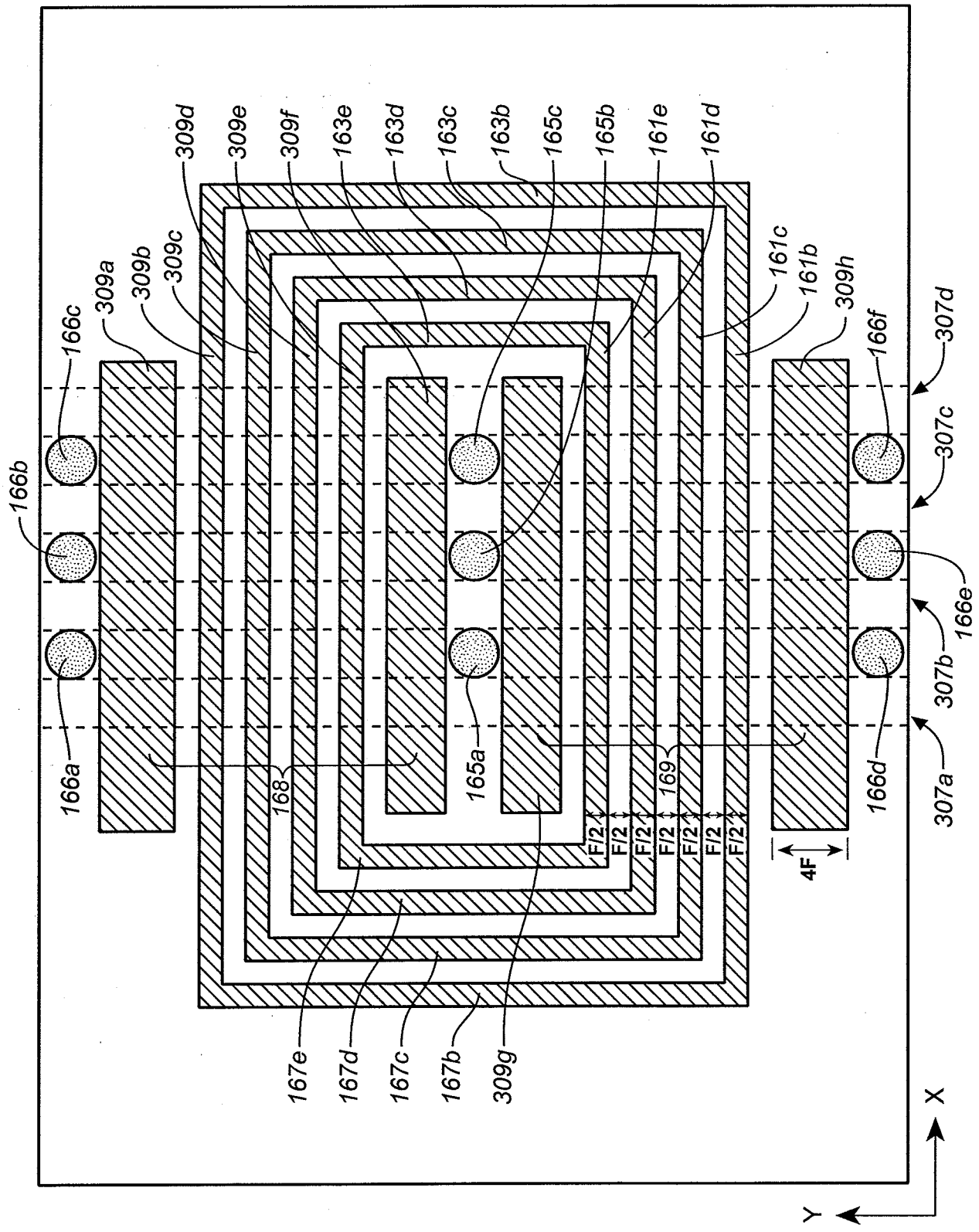

309d
309e
309f
163e
163d
163c
163b
165c
165b
161e
161d
161c
161b
309h
166f
307d
307c
166e
307b
307a
166d
309a
309b
309c
166c
166b
166a
168
165a
169
167e
167d
167c
167b
309g
F/2
4F
X
Y

FIG. 16

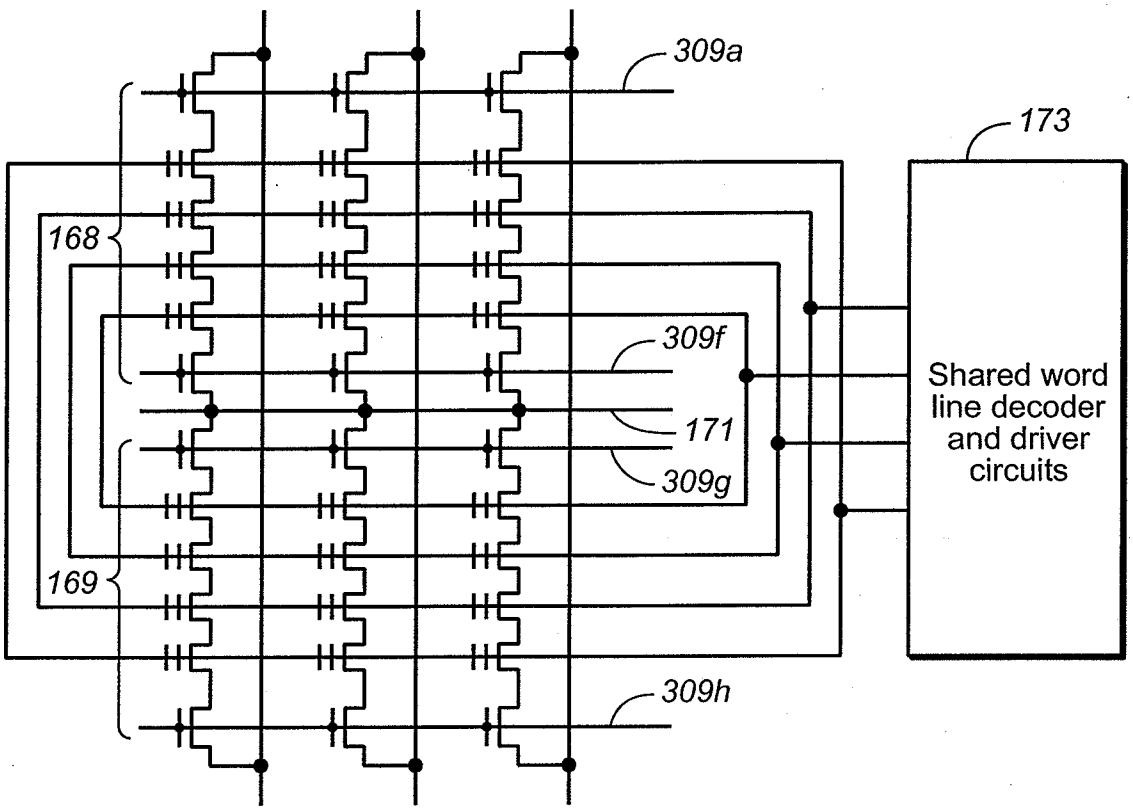

309a
173
168
309f
Shared word line decoder and driver circuits
171
309g
169
309h

FIG. 17

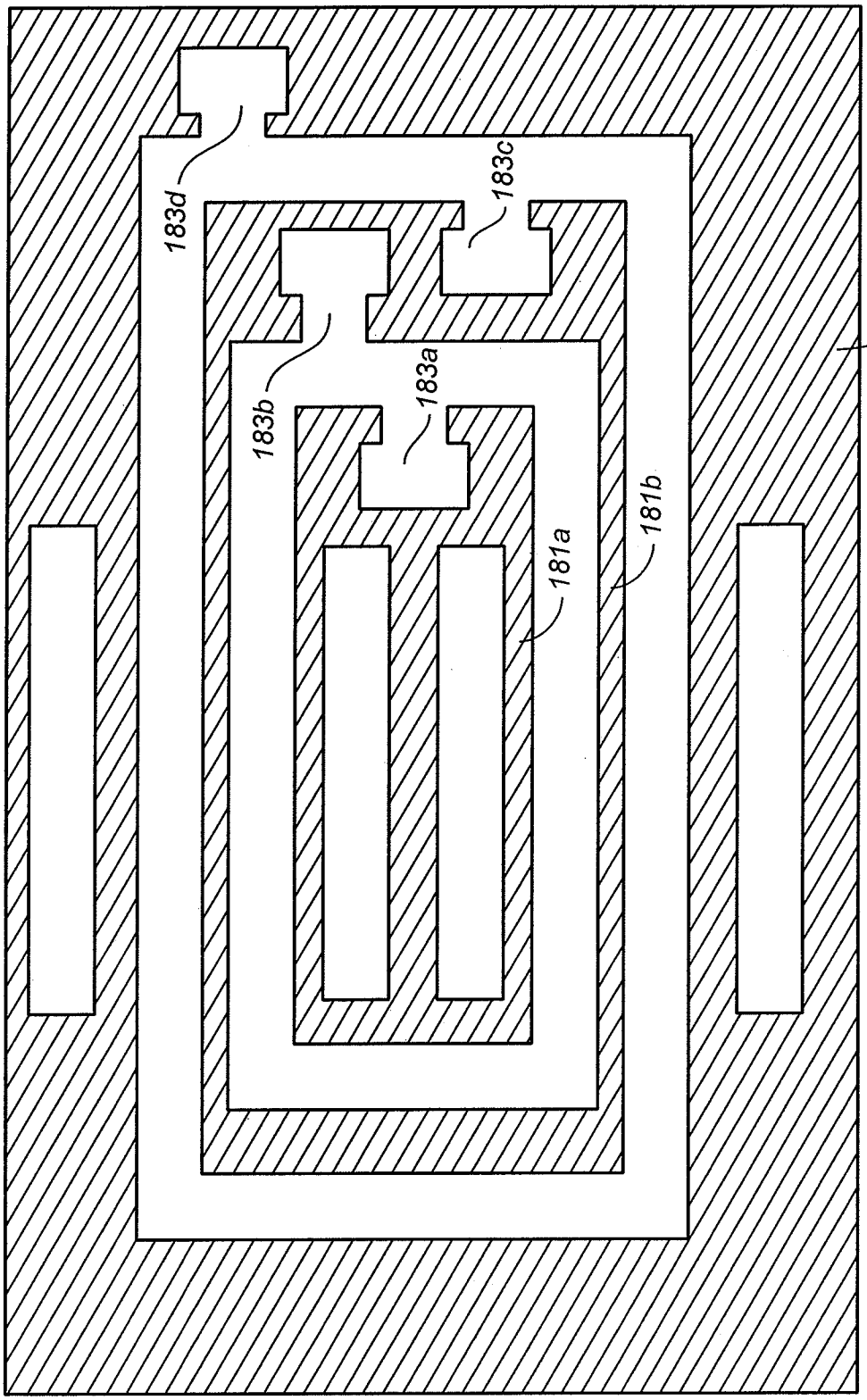


FIG. 18

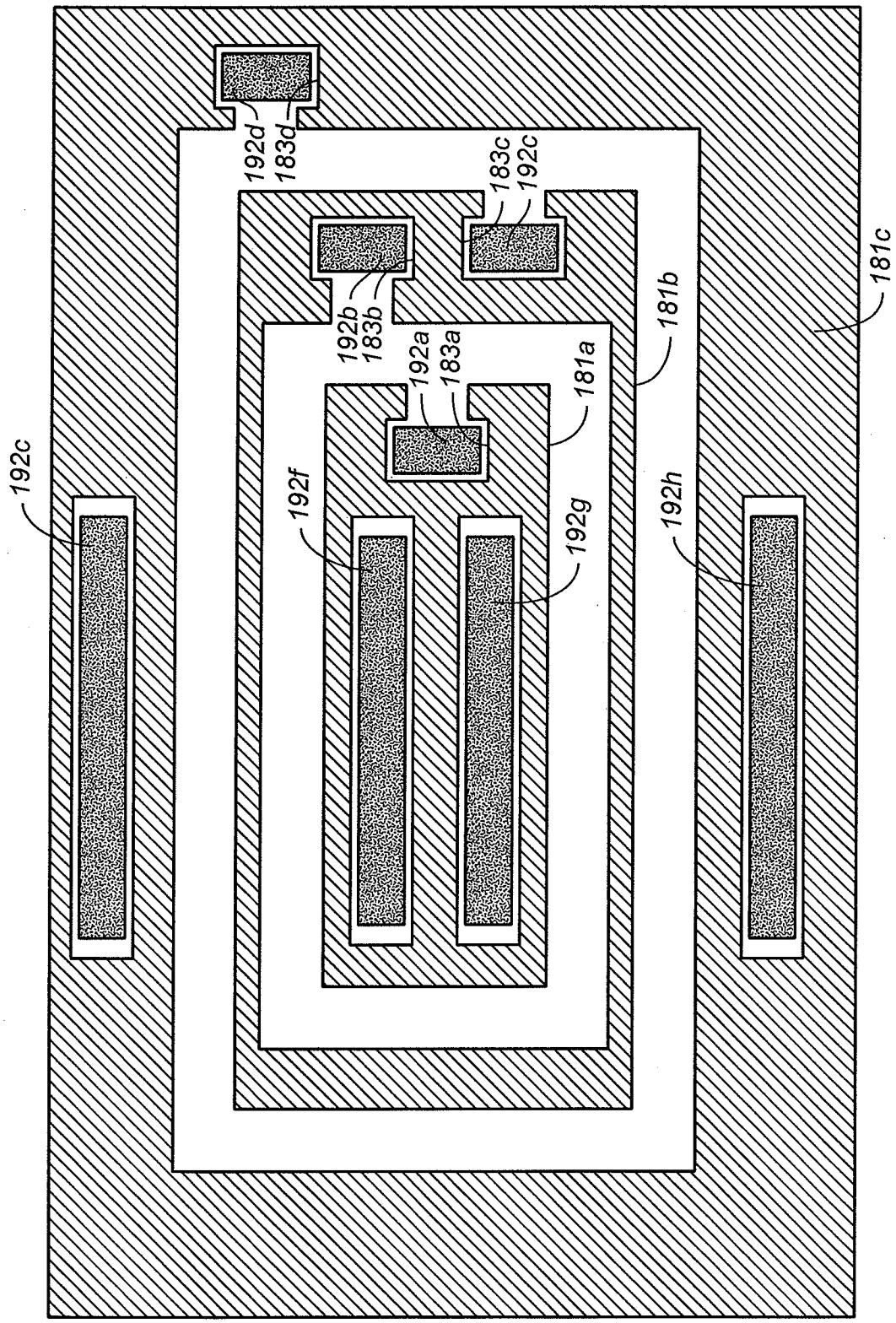

192d
183d
183c
192c
192b
183b
192a
183a
181a
181b
181c
192c
192f
192g
192h

FIG. 19

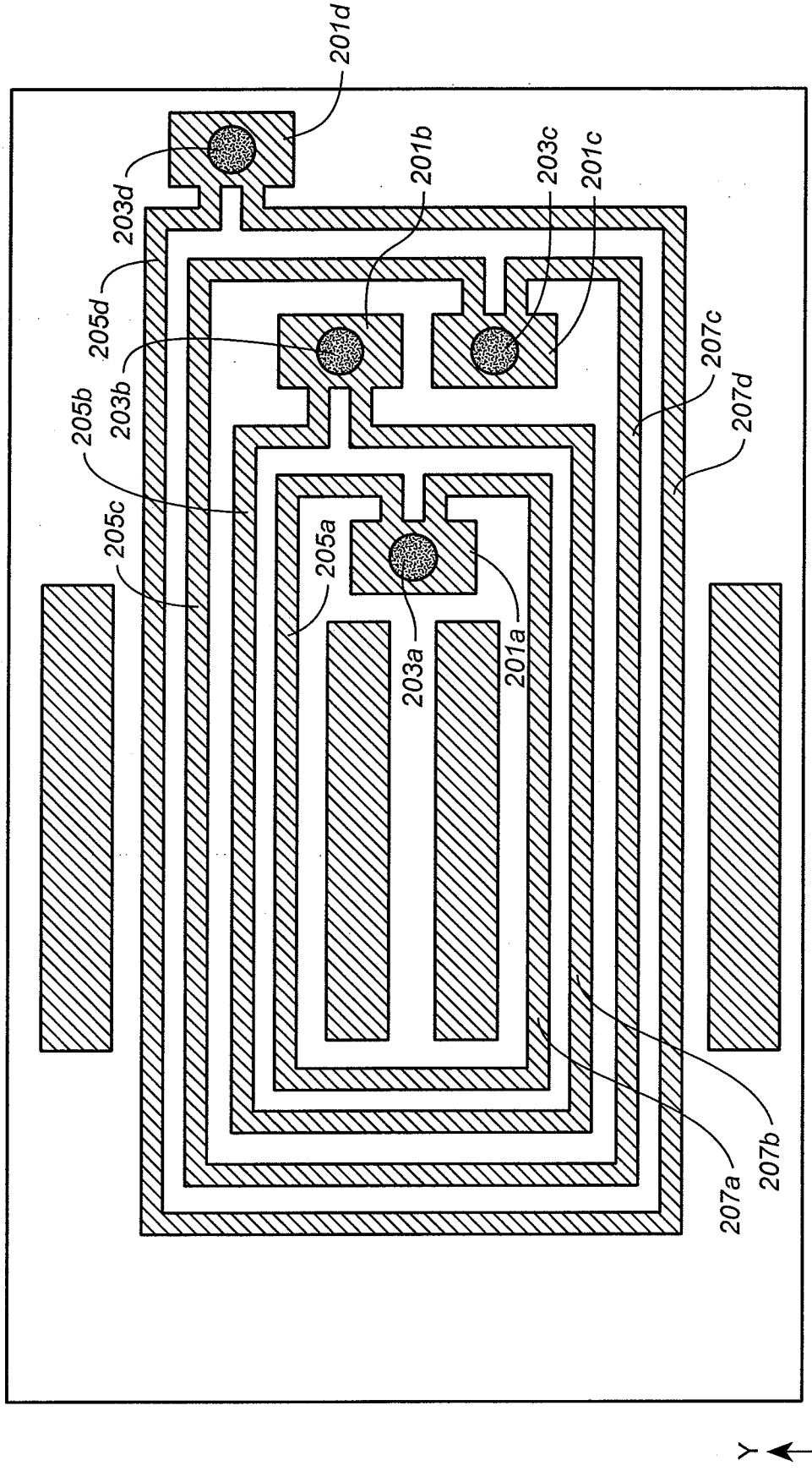

201d
203d
201b
203c
201c
205d
205b
203b
207c
207d
205c
205a
203a
201a
207a
207b
X
Y

FIG. 20

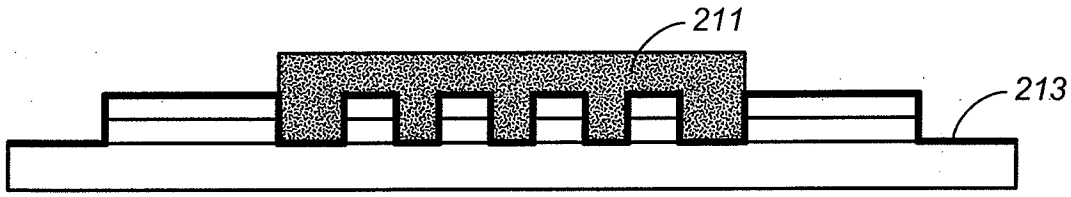


FIG. 21

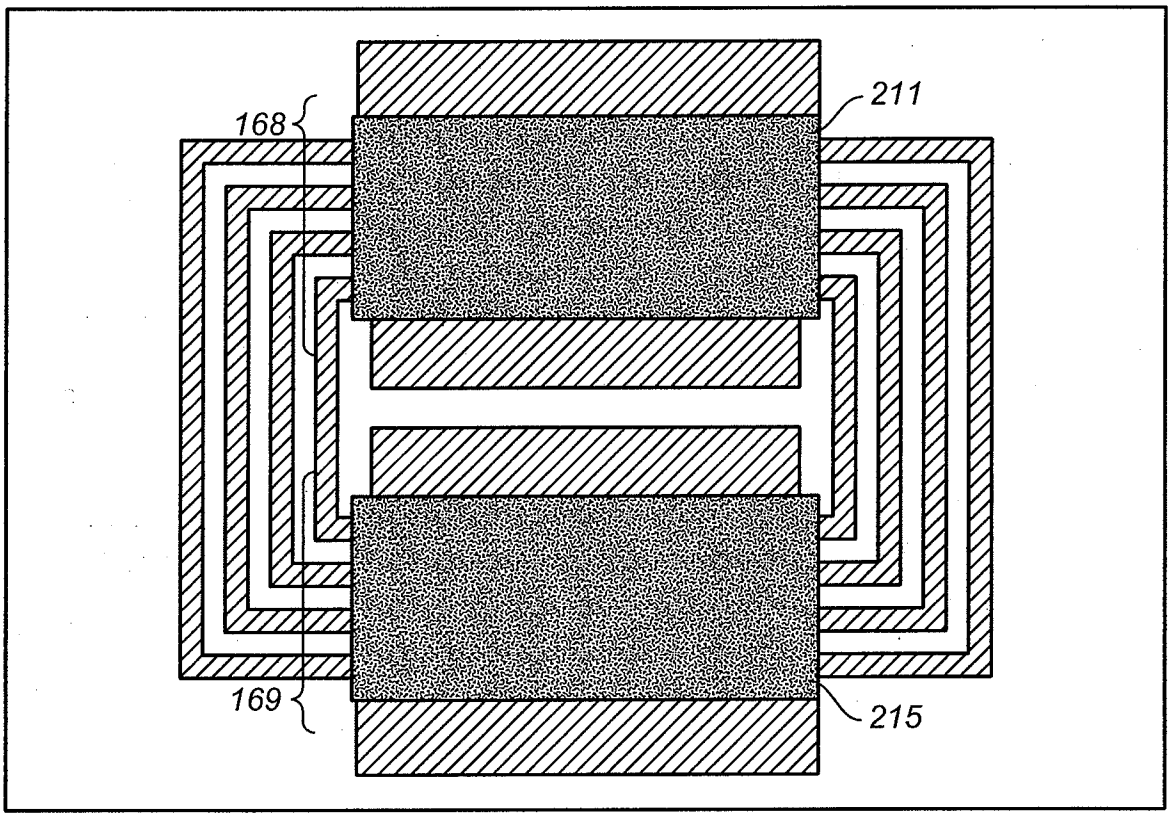


FIG. 22

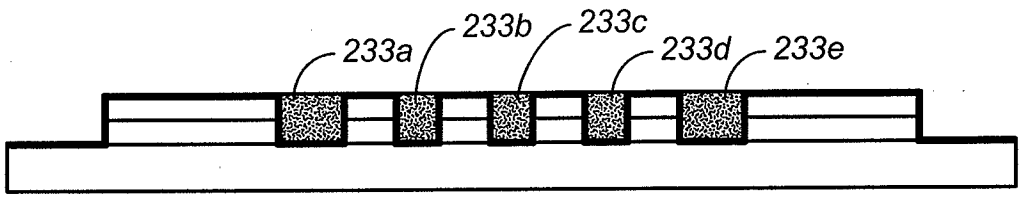


FIG. 23

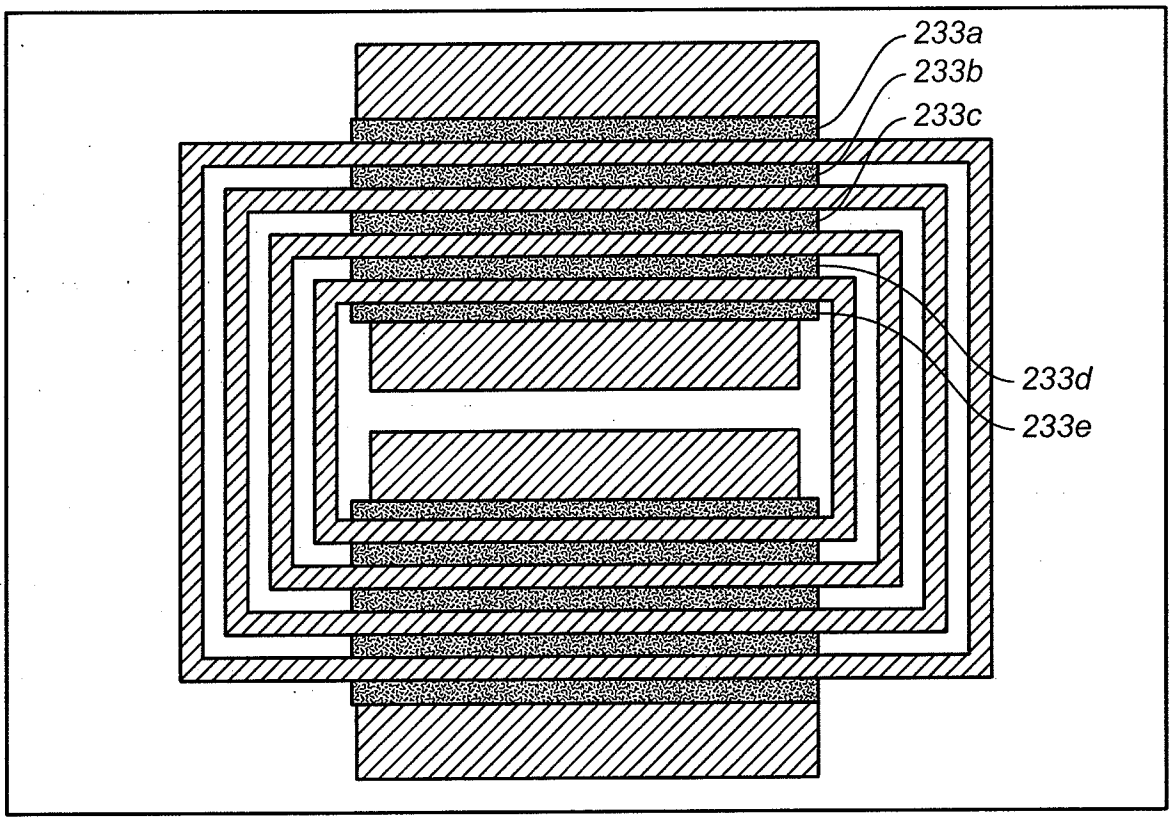


FIG. 24

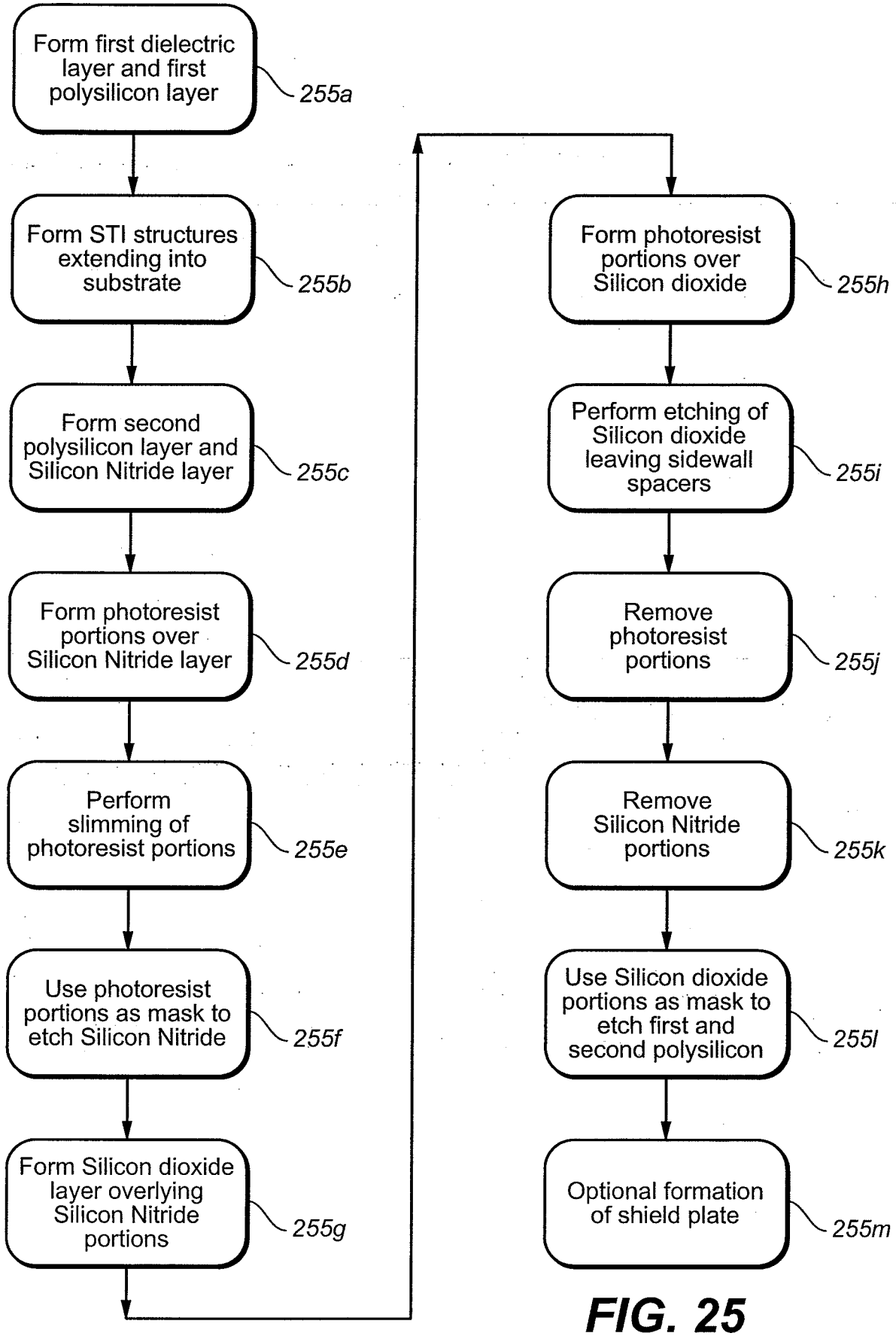

Form first dielectric layer and first polysilicon layer
255a
Form STI structures extending into substrate
255b
Form second polysilicon layer and Silicon Nitride layer
255c
Form photoresist portions over Silicon Nitride layer
255d
Perform slimming of photoresist portions
255e
Use photoresist portions as mask to etch Silicon Nitride
255f
Form Silicon dioxide layer overlying Silicon Nitride portions
255g
Form photoresist portions over Silicon dioxide
255h
Perform etching of Silicon dioxide leaving sidewall spacers
255i
Remove photoresist portions
255j
Remove Silicon Nitride portions
255k
Use Silicon dioxide portions as mask to etch first and second polysilicon
255l
Optional formation of shield plate
255m

FIG. 25

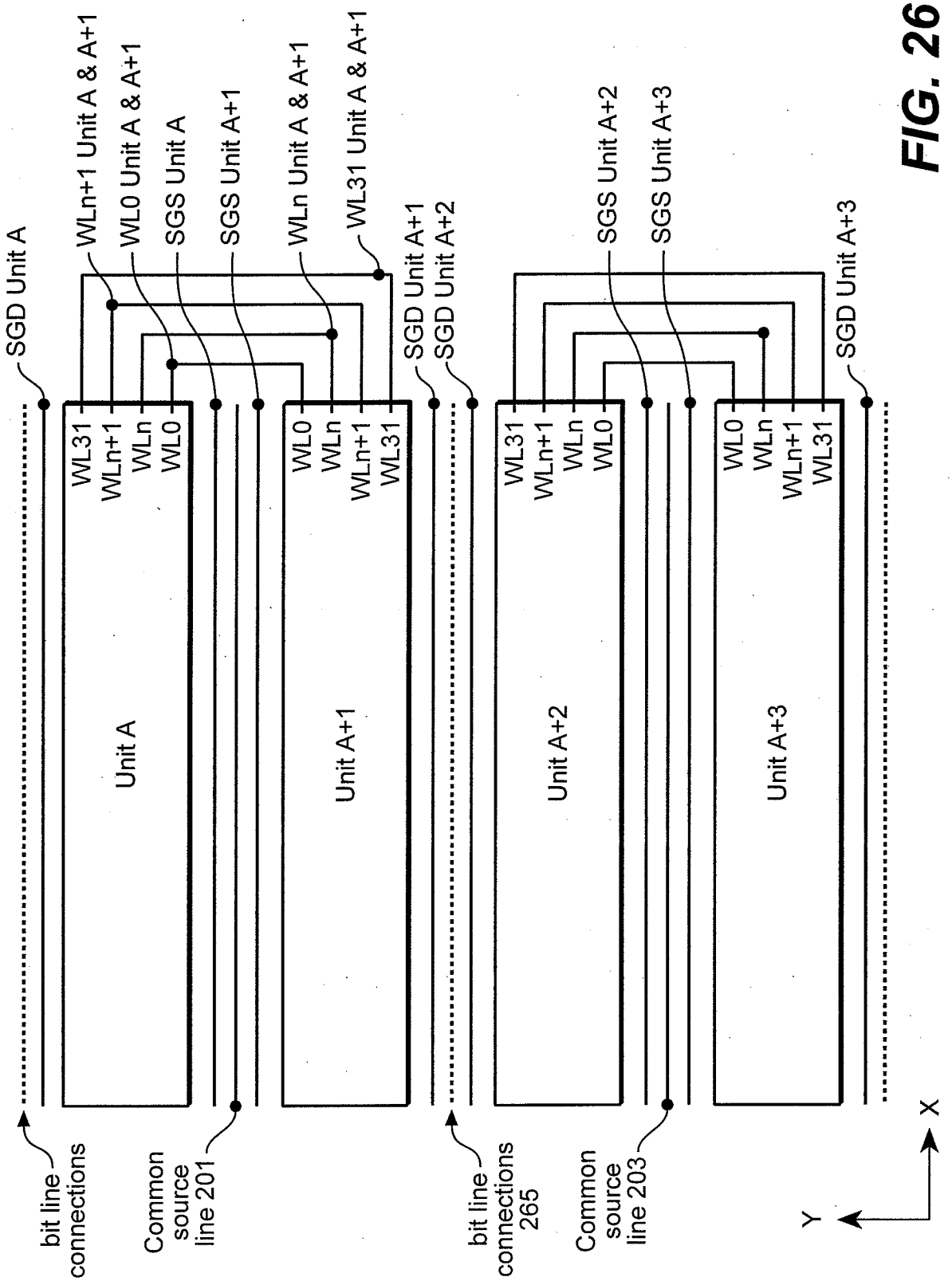

SGD Unit A
WLn+1 Unit A & A+1
WL0 Unit A & A+1
SGS Unit A
SGS Unit A+1
WLn Unit A & A+1
WL31 Unit A & A+1
SGD Unit A+1
SGD Unit A+2
SGS Unit A+2
SGS Unit A+3
SGD Unit A+3
WL31
WLn+1
WLn
WL0
Unit A
Unit A+1
Unit A+2
Unit A+3
bit line connections
Common source line 201
bit line connections 265
Common source line 203
Y
X

FIG. 26

Read Unit A+1, WLn

	Unit A	Unit A+1	Unit A+2	Unit A+3
SGD	VSS	VSG	VSS	VSS
WL31	VREAD		Float	
WL30	VREAD		Float	
WL n+1	VREAD		Float	
WL n	VSS		Float	
WL n–1	VREAD		Float	
WL1	VREAD		Float	
WL0	VREAD		Float	
SGS	VSS	VSG	VSS	VSS
Array Source	VSS			
Pwell	VSS			

FIG. 27A

VSG 4.5v
VREAD ~4.5v
VSS 0v

Program Unit A+1, WLn

	Unit A	Unit A+1	Unit A+2	Unit A+3
SGD	VSS	Vdd	VSS	VSS
WL31	VPASS		Float	
WL30	VPASS		Float	
WL n+1	VPASS		Float	
WL n	VPGM		Float	
WL n–1	VPASS		Float	
WL1	VPASS		Float	
WL0	VPASS		Float	
SGS	VSS	0	VSS	VSS
Array Source	~Vdd			
Pwell	VSS			

FIG. 27B

VPGM ~20v
VPASS ~10v
Vdd ~2v

Erase Unit A / Unit A+1

	Unit A	Unit A+1	Unit A+2	Unit A+3
SGD	Float	Float	Float	Float
WL31	VSS		Float	
WL30	VSS		Float	
WL n+1	VSS		Float	
WL n	VSS		Float	
WL n–1	VSS		Float	
WL1	VSS		Float	
WL0	VSS		Float	
SGS	Float	Float	Float	Float
Array Source	Float			
Pwell	VERA			

FIG. 27C

VERA ~20v

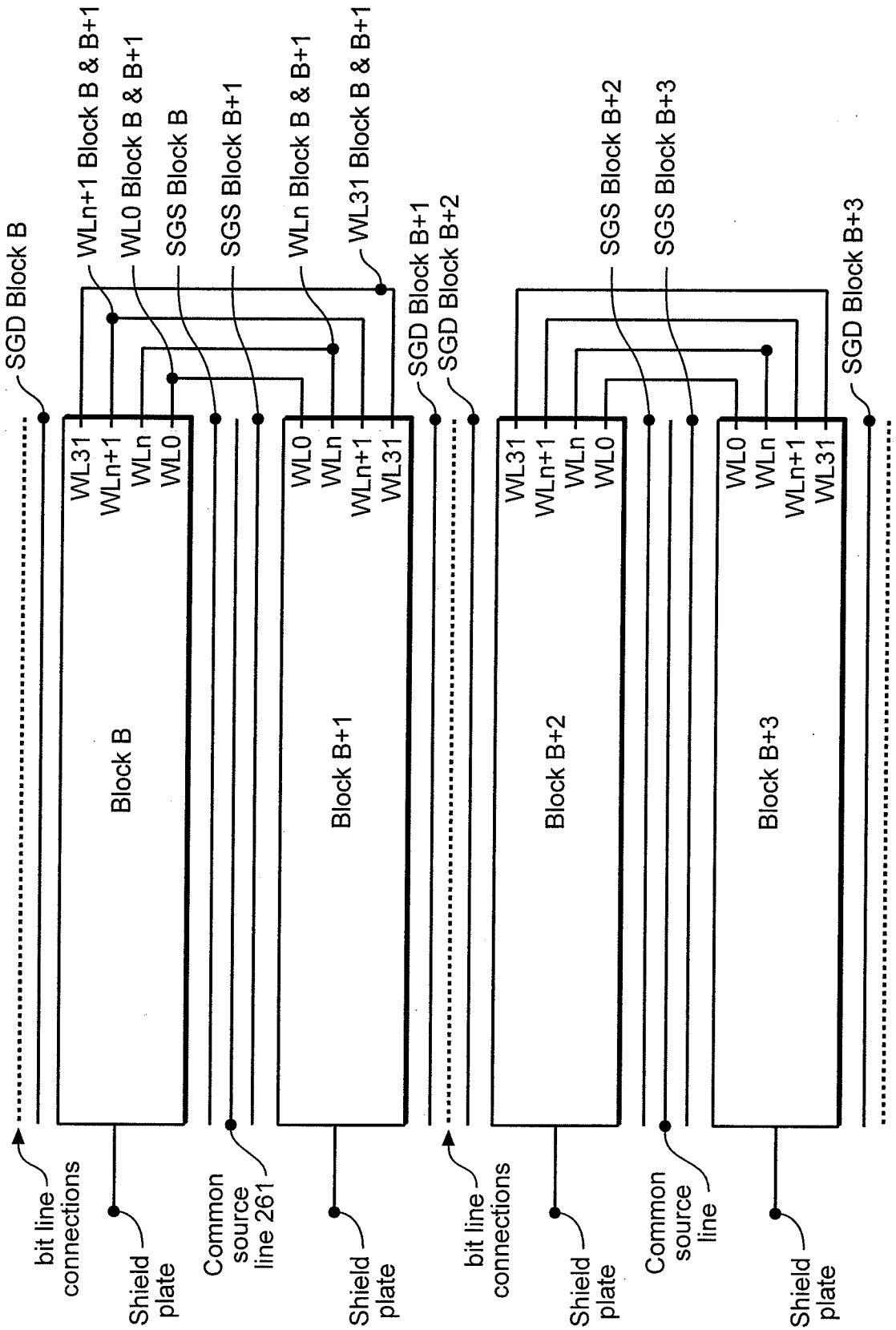

SGD Block B
WLn+1 Block B & B+1
WL0 Block B & B+1
SGS Block B
SGS Block B+1
WLn Block B & B+1
WL31 Block B & B+1
SGD Block B+1
SGD Block B+2
SGS Block B+2
SGS Block B+3
SGD Block B+3
WL31
WLn+1
WLn
WL0
Block B
Block B+1
Block B+2
Block B+3
bit line connections
Shield plate
Common source line 261
Common source line

FIG. 28

Read Block B+1, WLn

	Block B	Block B+1	Block B+2	Block B+3
SGD	VSS	VSG	VSS	VSS
WL31	VREAD		Float	
WL30	VREAD		Float	
WL n+1	VREAD		Float	
WL n	VSS		Float	
WL n–1	VREAD		Float	
WL1	VREAD		Float	
WL0	VREAD		Float	
SGS	VSS	VSG	VSS	VSS
Shield Plate	VRSP	VRSP	Float	Float
Array Source	VSS			
Pwell	VSS			

FIG. 29A

VSG	~4.5v
VREAD	~4.5v
VRSP	~4.0v
VSS	0v

Program Block B+1, WLn

	Block B	Block B+1	Block B+2	Block B+3
SGD	VSS	Vdd	VSS	VSS
WL31	VPASS		Float	
WL30	VPASS		Float	
WL n+1	VPASS		Float	
WL n	VPGM		Float	
WL n–1	VPASS		Float	
WL1	VPASS		Float	
WL0	VPASS		Float	
SGS	VSS	0	VSS	VSS
Shield Plate	VPSP	VPSP	Float	Float
Array Source	~Vdd			
Pwell	VSS			

FIG. 29B

VPGM	~20v
VPASS	~10v
Vdd	~2v
VPSP	~10v

Erase Block B+1

	Block B	Block B+1	Block B+2	Block B+3
SGD	Float	Float	Float	Float
WL31	VSS		Float	
WL30	VSS		Float	
WL n+1	VSS		Float	
WL n	VSS		Float	
WL n–1	VSS		Float	
WL1	VSS		Float	
WL0	VSS		Float	
SGS	Float	Float	Float	Float
Shield Plate	VEISP	VESSP	Float	Float
Array Source	Float			
Pwell	VERA			

FIG. 29C

VERA	~20v
VEISP	~18v
VESSP	~5v