SEMICONDUCTOR STRUCTURES IN MEMORY DEVICES

TECHNICAL FIELD

[0001] The present disclosure relates to semiconductor devices, e.g., memory devices.

BACKGROUND

[0002] Semiconductor devices, e.g., memory devices, can have various structures to increase the density of memory cells and lines on a chip. A memory device normally includes a memory cell array of memory cells and peripheral circuits for facilitating operations of the memory cell array.

SUMMARY

[0003] The present disclosure describes managing layouts of semiconductor structures in memory devices.

[0004] One aspect of the present disclosure features a memory device including a first semiconductor structure and a second semiconductor structure. The first semiconductor structure includes a memory cell array, where the memory cell array includes a first memory bank and a second memory bank. The second semiconductor structure includes a peripheral circuit, where the peripheral circuit includes a first bank circuit and a second bank circuit arranged along a first direction. The first bank circuit is coupled to the first memory bank and the second bank circuit is coupled to the second memory bank. The first bank circuit includes a first driver circuit on a boundary of the first bank circuit, and the second bank circuit includes a second driver circuit on a boundary of the second bank circuit. At least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction perpendicular to the first direction.

[0005] In some implementations, the first semiconductor structure and the second semiconductor structure are stacked along a third direction perpendicular to the first direction and the second direction. The first memory bank overlaps with the first bank circuit in a plan view perpendicular to the third direction. The second memory bank overlaps with the second bank circuit in the plan view.

[0006] In some implementations, the first bank circuit includes block circuits arranged in one or more rows. The first driver circuit is comprised in a first block circuit that is closest to the second

bank circuit among one row of block circuits of the first bank circuit, and. The second bank circuit includes block circuits arranged in one or more rows. The second driver circuit is comprised in a second block circuit that is closest to the first bank circuit among one row of block circuits of the second bank circuit.

[0007] In some implementations, the first block circuit further includes first sense amplifiers coupled to bit lines in a memory block of the first memory bank. The second bank circuit further includes second sense amplifiers coupled to bit lines in a memory block of the second memory bank.

[0008] In some implementations, positioning of the first driver circuit and the first sense amplifiers in the first block circuit is identical to positioning of the second driver circuit and the second sense amplifiers in the second block circuit.

[0009] In some implementations, the first memory bank includes a first memory block that is coupled to the first block circuit. The first block circuit further includes a third driver circuit. The third driver circuit overlaps with the first memory block in the plan view, and the first driver circuit does not overlap with the first memory block in the plan view.

[0010] In some implementations, word lines in the first memory block are numbered in a numerical order. The first block circuit further includes a fourth driver circuit and a fifth driver circuit. The first driver circuit and the third driver circuit are on a first boundary of the first block circuit and include a first set of word line drivers. The fourth driver circuit and the fifth driver circuit are on a second boundary of the first block circuit opposite to the first boundary, and include a second set of word line drivers. An even-numbered word line in the first memory block is coupled to a corresponding word line driver of the first set of word line drivers, and an odd-numbered word line in the first memory block is coupled to a corresponding word line driver of the second set of word line drivers.

[0011] In some implementations, the peripheral circuit includes a third bank circuit adjacent to the second bank circuit along the first direction. The third bank circuit is coupled to a third memory bank of the memory cell array, and row decoders between the second bank circuit and the third bank circuit.

[0012] In some implementations, the peripheral circuit includes a fourth bank circuit adjacent to the third bank circuit along the first direction. The fourth bank circuit is coupled to a fourth memory bank of the memory cell array. The third bank circuit includes a sixth driver circuit on a boundary

of the third bank circuit, and the fourth bank circuit includes a seventh driver circuit on a boundary of the fourth bank circuit. At least a portion of the sixth driver circuit is adjacent to and overlaps with at least a portion of the seventh driver circuit along the second direction.

[0013] In some implementations, the memory cell array includes DRAM memory cells.

[0014] In some implementations, the first semiconductor structure and the second semiconductor structure include bonding contacts that bond the first semiconductor structure and the second semiconductor structure together. The bonding contacts are isolated by an isolating material.

[0015] Another aspect of the present disclosure features a method of forming a memory device. The method includes forming a first semiconductor structure and forming a second semiconductor structure. The first semiconductor structure includes a memory cell array, where the memory cell array includes a first memory bank and a second memory bank. The second semiconductor structure includes a peripheral circuit, where the peripheral circuit includes a first bank circuit and a second bank circuit arranged along a first direction. The first bank circuit is coupled to the first memory bank and the second bank circuit is coupled to the second memory bank. The first bank circuit includes a first driver circuit on a boundary of the first bank circuit, and the second bank circuit includes a second driver circuit on a boundary of the second bank circuit. At least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction perpendicular to the first direction.

[0016] In some implementations, the method includes stacking the first semiconductor structure and the second semiconductor structure along a third direction perpendicular to the first direction and the second direction. The first memory bank overlaps with the first bank circuit in a plan view perpendicular to the third direction. The second memory bank overlaps with the second bank circuit in the plan view.

[0017] In some implementations, the first bank circuit includes block circuits arranged in one or more rows. The first driver circuit is comprised in a first block circuit that is closest to the second bank circuit among one row of block circuits of the first bank circuit, and. The second bank circuit includes block circuits arranged in one or more rows. The second driver circuit is comprised in a second block circuit that is closest to the first bank circuit among one row of block circuits of the second bank circuit.

[0018] In some implementations, the method includes forming first sense amplifiers in the first bank circuit, where the first sense amplifiers are coupled to bit lines in a memory block of the first memory bank. The method further includes forming second sense amplifiers in a second bank circuit, where the second sense amplifiers are coupled to bit lines in a memory block of the second memory bank.

[0019] In some implementations, positioning of the first driver circuit and the first sense amplifiers in the first block circuit is identical to positioning of the second driver circuit and the second sense amplifiers in the second block circuit.

[0020] In some implementations, the first memory bank includes a first memory block that is coupled to the first block circuit. The first block circuit further includes a third driver circuit. The third driver circuit overlaps with the first memory block in the plan view, and the first driver circuit does not overlap with the first memory block in the plan view.

[0021] In some implementations, word lines in the first memory block are numbered in a numerical order. The first block circuit further includes a fourth driver circuit and a fifth driver circuit. The first driver circuit and the third driver circuit are on a first boundary of the first block circuit and include a first set of word line drivers. The fourth driver circuit and the fifth driver circuit are on a second boundary of the first block circuit opposite to the first boundary, and include a second set of word line drivers. An even-numbered word line in the first memory block is coupled to a corresponding word line driver of the first set of word line drivers, and an odd-numbered word line in the first memory block is coupled to a corresponding word line driver of the second set of word line drivers.

[0022] In some implementations, forming the second semiconductor structure includes forming a third bank circuit adjacent to the second bank circuit along the first direction. The third bank circuit is coupled to a third memory bank of the memory cell array. Forming the second semiconductor structure further includes forming row decoders between the second bank circuit and the third bank circuit.

[0023] A further aspect of the present disclosure features a memory system. The memory system includes a memory device and a memory controller coupled to the memory device and configured to control the memory device. The memory device includes a first semiconductor structure and a second semiconductor structure. The first semiconductor structure includes a memory cell array, where the memory cell array includes a first memory bank and a second memory bank. The second

semiconductor structure includes a peripheral circuit, where the peripheral circuit includes a first bank circuit and a second bank circuit arranged along a first direction. The first bank circuit is coupled to the first memory bank and the second bank circuit is coupled to the second memory bank. The first bank circuit includes a first driver circuit on a boundary of the first bank circuit, and the second bank circuit includes a second driver circuit on a boundary of the second bank circuit. At least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction perpendicular to the first direction.

[0024] A yet further aspect of the present disclosure features a semiconductor structure. The semiconductor structure includes a first bank circuit and a second bank circuit arranged along a first direction. The first bank circuit includes a first driver circuit on a boundary of the first bank circuit, and the second bank circuit includes a second driver circuit on a boundary of the second bank circuit. At least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction perpendicular to the first direction.

[0025] An additional aspect of the present disclosure features a semiconductor structure. The semiconductor structure includes a first memory bank and a second memory block arranged along a first direction. A portion of first memory cells included in the first memory bank are on a boundary of the first memory bank, and a portion of second memory cells included in the second memory bank are on a boundary of the second memory bank. The portion of first memory cells are adjacent to and overlap with the portion of second memory cells along a second direction perpendicular to the first direction.

[0026] In some implementations, the first memory cells and the second memory cells are DRAM cells.

[0027] The details of one or more implementations of the subject matter of this present disclosure are set forth in the accompanying drawings and the description below. Other features, aspects, and advantages of the subject matter will become apparent from the description, the drawings, and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] The accompanying drawings, which are incorporated herein and form a part of the present disclosure, illustrate aspects of the present disclosure and, together with the description,

further serve to explain the principles of the present disclosure and to enable a person of ordinary skill in the pertinent art to make and use the present disclosure.

[0029] FIG. 1A illustrates a schematic view of a cross-section of an example memory device.

[0030] FIG. 1B illustrates a schematic view of a cross-section of another example memory device.

[0031] FIG. 2 illustrates a cross-sectional view of an example memory device.

[0032] FIG. 3 illustrates a schematic diagram of an example memory device including a peripheral circuit and an array of memory cells each having a vertical transistor.

[0033] FIG. 4 illustrates example peripheral circuits.

[0034] FIG. 5 illustrates a layout view of example peripheral circuits included in a second semiconductor structure.

[0035] FIG. 6 illustrates a plan view of an example memory device.

[0036] FIG. 7 illustrates a plan view of an example bank circuit.

[0037] FIG. 8 illustrates a plan view of example block circuits.

[0038] FIGS. 9A-9D illustrate layout view of example bank circuits that are adjacent to each other.

[0039] FIG. 10 illustrates a plan view of an example memory device.

[0040] FIG. 11 illustrates an example memory device including example memory blocks of the first semiconductor structure and example word line drivers of the second semiconductor structure.

[0041] FIG. 12 illustrates schematic circuit diagrams of example word line drivers.

[0042] FIG. 13 is a flow chart of an example process of forming a memory device.

[0043] FIG. 14 illustrates a block diagram of an example system having one or more memory devices.

[0044] FIGS. 15A-15B illustrate block diagrams of example systems having one or more memory devices.

[0045] Like reference numbers and designations in the various drawings indicate like elements. It is also to be understood that the various exemplary implementations shown in the figures are merely illustrative representations and are not necessarily drawn to scale.

DETAILED DESCRIPTION

[0046] As performance requirements for Dynamic Random Access Memory (DRAM) continue to escalate, DRAM structures have evolved from 2D to 3D. Under a 2D DRAM architecture, a memory cell array of the DRAM device and peripheral circuits (e.g., including word line drivers, sense amplifiers, and other peripheral circuits) controlling the memory cell array are formed on the same wafer. Under a 3D DRAM architecture, the memory device can be a bonded memory chip including a first semiconductor structure that includes the memory cell array and a second semiconductor structure that includes the peripheral circuits. The peripheral circuit can include bank circuits configured to control corresponding memory banks of the memory cell array. The first semiconductor structure and the second semiconductor structure can be formed separately on different wafers, and then stacked together to form the bonded memory chip.

[0047] In some implementations, a bank circuit can include block circuits configured to control corresponding memory blocks of the memory bank. Driver circuits including word line drivers can be arranged on a boundary of the block circuit. For example, one driver circuit can overlap with the corresponding memory block in a plan view perpendicular to the stacking direction (e.g., the Z direction), while another driver circuit does not overlap with the corresponding memory block in the plan view. As such, each block circuit on the boundary of the bank circuit can include a driver circuit that does not overlap with the corresponding memory bank in the plan view, therefore forming a zig-zag pattern on the boundary of the bank circuit.

[0048] In some cases, two adjacent bank circuits arranged along a first direction (e.g., the X direction) do not overlap with each other along a second direction (e.g., the Y direction) perpendicular to the first direction, such that boundaries of the two adjacent bank circuits are offset from each other. As such, spacing between two adjacent bank circuits may increase the die size of the bonded memory chip.

[0049] Implementations of the present disclosure provide techniques for managing layouts of semiconductor structures (e.g., the first semiconductor structure and the second semiconductor structure) in a memory device. In some implementations, a first bank circuit and a second bank circuit are adjacent to each other along the first direction. Driver circuits on the boundary of the first bank circuit can be adjacent to, and at least partially overlaps with driver circuits on the boundary of the second bank circuit along the second direction. As such, the zig-zag pattern of the

boundary of the first bank circuit can fit with the zig-zag pattern of the boundary of the second bank circuit, so that the two adjacent bank circuits can be tightly arranged with each other.

[0050] Implementations of the present disclosure can provide one or more of the following technical advantages. For example, by tightly arranging adjacent bank circuits, the spacing between adjacent bank circuits can be reduced, so that the size of the bonded chip can be reduced. For another example, techniques in the present disclosure do not require changing circuit layout within a bank circuit, which is a cost-effective way to reduce die size. In some implementations, different or additional technical advantages can be achieved.

[0051] FIG. 1A illustrates a schematic view of a cross-section of a memory device 100, according to some aspects of the present disclosure. The memory device 100 represents an example of a bonded chip. The components of the memory device 100 (e.g., memory cell array and peripheral circuits) can be formed separately on different substrates and then joined to form a bonded chip. The memory device 100 can include a first semiconductor structure 102 including memory cell array. The memory device 100 can also include a second semiconductor structure 104 including peripheral circuits. The peripheral circuits (e.g., control and sensing circuits) can include any suitable digital, analog, and/or mixed-signal circuits used for facilitating the operations of the memory cell array. For example, the peripheral circuits can include one or more of a page buffer, a decoder (e.g., a row decoder and a column decoder), a sense amplifier, a driver (e.g., a word line driver), an input/output (I/O) circuit, a charge pump, a voltage source or generator, a current or voltage reference, any portions (e.g., a subcircuit) of the functional circuits mentioned above, or any active or passive components of the circuit (e.g., transistors, diodes, resistors, or capacitors). The peripheral circuits in the second semiconductor structure 104 use complementary metal-oxide-semiconductor (CMOS) technology, which can be implemented, for example, with logic processes (e.g., technology nodes of 90 nm, 65 nm, 60 nm, 45 nm, 32 nm, 28 nm, 22 nm, 20 nm, 16 nm, 14 nm, 10 nm, 7 nm, 5 nm, 3 nm, 2 nm, etc.), according to some implementations.

[0052] As shown in FIG. 1A, the memory device 100 can also include the first semiconductor structure 102 including an array of memory cells (memory cell array) that can use transistors as the switch and selecting devices. In some implementations, the memory cell array includes an array of DRAM cells. For ease of description, a DRAM cell array may be used as an example for describing the memory cell array in the present disclosure. But it is understood that the memory cell array is not limited to DRAM cell array and may include any other suitable types of memory

cell arrays that can use transistors as the switch and selecting devices, such as phase-change memory (PCM) cell array, static random-access memory (SRAM) cell array, FRAM cell array, resistive memory cell array, magnetic memory cell array, spin transfer torque (STT) memory cell array, to name a few, or any combination thereof.

[0053] In some implementations, the memory cell array of the first semiconductor structure 102 can include a plurality of memory banks arranged in rows and in columns. As an example, a memory cell array can include two rows of memory banks, where each row includes eight memory banks. Further, in some implementations, each memory bank can include memory blocks arranged in rows and in columns. As an example, a memory bank can include *n* rows and *m* columns of memory blocks, where each memory block can include memory cells.

[0054] The first semiconductor structure 102 can be a DRAM device in which memory cells are provided in the form of an array of DRAM cells. In some implementations, each DRAM cell includes a capacitor for storing a bit of data as a positive or negative electrical charge as well as one or more transistors (e.g., pass transistors) that control (e.g., switch and select) access to it. In some implementations, each DRAM cell is a one-transistor, one-capacitor (1TlC) cell. Since transistors can leak a small amount of charge, the capacitors can slowly discharge, causing information stored in them to drain. As such, a DRAM cell can be refreshed to retain data, for example, by the peripheral circuit in the second semiconductor structure 104, according to some implementations.

[0055] As shown in FIG. 1A, the memory device 100 further includes a bonding interface 106 vertically between (in the vertical direction, e.g., the z-direction in FIG. 1A) the first semiconductor structure 102 and the second semiconductor structure 104. As described below in detail, the first and the second semiconductor structures 102 and 104 can be fabricated separately (and in parallel in some implementations) such that the thermal budget of fabricating one of the first and the second semiconductor structures 102 and 104 does not limit the processes of fabricating another one of the first and the second semiconductor structures 102 and 104. Moreover, a large number of interconnects (e.g., bonding contacts) can be formed through the bonding interface 106 to make direct, short-distance (e.g., micron-level) electrical connections between the first semiconductor structure 102 and the second semiconductor structure 104, as opposed to the long-distance (e.g., millimeter or centimeter-level) chip-to-chip data bus on the circuit board, such as printed circuit board (PCB), thereby eliminating chip interface delay and

achieving high-speed I/O throughput with reduced power consumption. Data transfer between the memory cell array in the first semiconductor structure 102 and the peripheral circuits in the second semiconductor structure 104 can be performed through the interconnects (e.g., bonding contacts) across the bonding interface 106. For example, the first semiconductor structure 102 can include first bonding contacts isolated by an isolating material (e.g., SiO2), and the second semiconductor structure 104 can include second bonding contacts isolated by an isolating material (e.g., SiO2). The first semiconductor structure 102 and the second semiconductor structure 104 can be bonded together by the first bonding contacts being in contact with the second bonding contacts. By vertically integrating the first and the second semiconductor structures 102 and 104, the chip size can be reduced, and the memory cell density can be increased.

[0056] It is understood that the relative positions of stacked first and second semiconductor structures 102 and 104 are not limited. FIG. 1B illustrates a schematic view of a cross-section of another memory device 101, according to some aspects of the present disclosure. Different from the memory device 100 in FIG. 1A in which the first semiconductor structure 102 including the memory cell array is above the second semiconductor structure 104 including the peripheral circuits, in the memory device 101 in FIG. 1B, the second semiconductor structure 104 including the peripheral circuit is above the first semiconductor structure 102 including the memory cell array. Nevertheless, the bonding interface 106 is formed vertically between the first and the second semiconductor structures 102 and 104 in the memory device 101, and the first and the second semiconductor structures 102 and 104 are jointed vertically through bonding (e.g., hybrid bonding) according to some implementations. Hybrid bonding, also known as "metal/dielectric hybrid bonding," is a direct bonding technology (e.g., forming bonding between surfaces without using intermediate layers, such as solder or adhesives) and can obtain metal-metal (e.g., copper-to-copper) bonding and dielectric-dielectric (e.g., silicon oxide-to-silicon oxide) bonding simultaneously. Data transfer between the memory cell array in the first semiconductor structure 102 and the peripheral circuits in the second semiconductor structure 104 can be performed through the interconnects (e.g., bonding contacts) across bonding interface 106.

[0057] It is noted that x, y, and z axes are included in FIGs. 1A and 1B to further illustrate the spatial relationship of the components in memory devices 100 and 101. The substrate of the memory device includes two lateral surfaces extending laterally in the x-y plane: a top surface on the front side of the wafer on which the semiconductor devices can be formed, and a bottom surface

on the backside opposite to the front side of the wafer. The z-axis is perpendicular to both the x and y axes. As used herein, whether one component (e.g., a layer or a device) is "on," "above," or "below" another component (e.g., a layer or a device) of the memory device is determined relative to the substrate of the memory device in the z-direction (the vertical direction perpendicular to the x-y plane, e.g., the thickness direction of the substrate) when the substrate is positioned in the lowest plane of the memory device in the z-direction. The same notion for describing the spatial relationships is applied throughout the present disclosure.

[0058] FIG. 2 illustrates a side view of a cross-section of an example memory device 200, according to some aspects of the present disclosure. The memory device 200 can be a dynamic random-access memory (DRAM). In some implementations, the memory device 200 is a bonded chip including a first semiconductor structure 202 (e.g., the first semiconductor structure 102 of FIGs. 1A-1B) and a second semiconductor structure 204 (e.g., the second semiconductor structure 104 of FIGs. 1A-1B). The first semiconductor structure 202 can be stacked over the second semiconductor structure 204. The first and the second semiconductor structures 202 and 204 can be jointed at a bonding interface 206 therebetween.

[0059] As shown in FIG. 2, the second semiconductor structure 204 can include a substrate 210, which can include silicon (e.g., single crystalline silicon, c-Si), SiGe, GaAs, Ge, SOI, or any other suitable materials. The second semiconductor structure 204 can include peripheral circuits 212 on and/or in the substrate 210. In some implementations, the peripheral circuits 212 include a plurality of transistors 214 (e.g., planar transistors and/or 3D transistors). Trench isolations (e.g., shallow trench isolations (STIs)) and doped regions (e.g., wells, sources, and drains of transistors 214) can be formed on or in the substrate 210 as well. In some examples, the peripheral circuits 212 are formed using complementary metal-oxide-semiconductor (CMOS) technology, and the second semiconductor structure 204 can be also formed on a semiconductor die that can be referred to as a control die or a CMOS die.

[0060] In some implementations, the second semiconductor structure 204 further includes an interconnect layer 216 above the peripheral circuits 212 to transfer electrical signals to and from the peripheral circuits 212. The interconnect layer 216 can include a plurality of interconnects (also referred to herein as “contacts”), including lateral interconnect lines and VIA contacts. The interconnect layer 216 can include one or more metal layers separated by interlay dielectric (ILD) layers. The interconnect lines and via contacts can form in the ILD layer to form electric contact

between different metal layers. That is, the interconnect layer 216 can include interconnect lines and via contacts in multiple ILD layers. In some implementations, peripheral circuits 212 are coupled to one another through the interconnects in the interconnect layer 216. The interconnects in interconnect layer 216 can include conductive materials including, but not limited to, W, Co, Cu, Al, doped silicon, silicides, or any combination thereof. The ILD layers can be formed with dielectric materials including, but not limited to, silicon oxide, silicon nitride, silicon oxynitride, low-k dielectrics, or any combination thereof.

[0061] As shown in FIG. 2, the second semiconductor structure 204 has a front side and a back side, and the second semiconductor structure 204 can further include a bonding layer 218 at the back side at the bonding interface 206 and above the interconnect layer 216 and the peripheral circuits 212. The bonding layer 218 can include a plurality of bonding contacts 219 and dielectrics electrically isolating the bonding contacts 219. The bonding contacts 219 can include conductive materials, such as Cu. The remaining area of the bonding layer 218 can be formed with dielectric materials, such as silicon oxide. The bonding contacts 219 and surrounding dielectrics in the bonding layer 218 can be used for hybrid bonding. Similarly, as shown in FIG. 2, the first semiconductor structure 202 can also include a bonding layer 220 at the bonding interface 206 and above the bonding layer 218 of the second semiconductor structure 204. The bonding layer 220 can include a plurality of bonding contacts 221 and dielectrics electrically isolating the bonding contacts 221. The bonding contacts 221 can include conductive materials, such as Cu. The remaining area of the bonding layer 220 can be formed with dielectric materials, such as silicon oxide. The bonding contacts 221 and surrounding dielectrics in the bonding layer 220 can be used for hybrid bonding. The bonding contacts 221 can be in contact with the bonding contacts 219 at the bonding interface 206. In some implementations, the bonding layer 220 includes a dielectric layer opposing memory cells (e.g., DRAM cells) 224 with a bit line 223 positioned between the dielectric layer and the memory cells 224, as shown in FIG. 2. The dielectric layer can include the bonding interface 206 having the bonding contacts 221.

[0062] The first semiconductor structure 202 can be bonded on top of the second semiconductor structure 204 in a face-to-face manner at the bonding interface 206. In some implementations, the bonding interface 206 is disposed between the bonding layers 220 and 218 as a result of hybrid bonding. In some implementations, the bonding interface 206 is the place at which bonding layers 220 and 218 are met and bonded. In some examples, the bonding interface 206 can be a layer with

a certain thickness that includes the top surface of the bonding layer 218 of the second semiconductor structure 204 and the bottom surface of the bonding layer 220 of the first semiconductor structure 202.

[0063] In some implementations, the first semiconductor structure 202 further includes an interconnect layer 222 including bit lines 223 above the bonding layer 220 to transfer electrical signals. The interconnect layer 222 can include a plurality of interconnects, such as mid end of line (MEOL) interconnects and back end of line (BEOL) interconnects. In some implementations, the interconnects in interconnect layer 222 also include local interconnects, such as the bit lines 223 and word line contacts (not shown). The interconnect layer 222 can include one or more metal layers separated by interlay dielectric (ILD) layers. The interconnect lines and via contacts can form in the ILD layers to form electric contact between different metal layers. The interconnects in the interconnect layer 222 can include conductive materials including, but not limited to, W, Co, Cu, Al, doped silicon, silicides, or any combination thereof. The ILD layers can be formed with dielectric materials including, but not limited to, silicon oxide, silicon nitride, silicon oxynitride, low-k dielectrics, or any combination thereof.

[0064] In some implementations, the peripheral circuits 212 include a word line driver/row decoder coupled to the word line contacts in the interconnect layer 222 through the bonding contacts 221 and 219 in the bonding layers 220 and 218 and the interconnect layer 216. In some implementations, the peripheral circuits 212 include a bit line driver/column decoder coupled to the bit lines 223 and bit line contacts in the interconnect layer 222 through the bonding contacts 221 and 219 in the bonding layers 220 and 218 and the interconnect layer 216. In some implementations, the bit line 223 is a metal bit line, as opposed to semiconductor bit lines (e.g., doped silicon bit lines). For example, the bit line 223 may include W, Co, Cu, Al, or any other suitable metals having higher conductivities than doped silicon. In some implementations, the bit line contact is an ohmic contact as opposed to a Schottky contact.

[0065] In some implementations, the bit line 223 is made of a composite conductive material that can be based on a metallic material (e.g., W, Co, Cu, Al) and a semiconductor material (e.g., Si). For example, the composite conductive material can include metal silicide, e.g., such as WSi, CoSi, CuSi, AlSi, or any other suitable metal silicides having higher conductivities than doped silicon.

[0066] In some implementations, the first semiconductor structure 202 includes DRAM cells 224 provided in the form of a memory cell array above the interconnect layer 222 and the bonding layer 220. That is, the interconnect layer 222 including the bit lines 223 can be disposed between bonding layer 220 and array of DRAM cells 224. A bit line 223 in the interconnect layer 222 can be coupled to a string of DRAM cells 224. In some implementations, the first semiconductor structure 202 is formed on a semiconductor die and can be referred to as array die.

[0067] In some implementations, a semiconductor device can include multiple array dies (e.g., the first semiconductor structure 202) and a CMOS die (e.g., the second semiconductor structure 204). The multiple array dies and the CMOS die can be stacked and bonded together. The CMOS die can be respectively coupled to each of the multiple array dies, and can respectively drive each of the multiple array dies to operate in the similar manner as the semiconductor device. The semiconductor device can be any suitable device. In some examples, the semiconductor device includes at least a first wafer and a second wafer bonded face to face. The array die can be disposed with other array dies on the first wafer, and the CMOS die can be disposed with other CMOS dies on the second wafer. The first wafer and the second wafer can be bonded together. As such, the array dies on the first wafer can be bonded with corresponding CMOS dies on the second wafer. In some examples, the semiconductor device is a chip with at least the array die and the CMOS die bonded together. In an example, the chip is diced from wafers that are bonded together. In another example, the semiconductor device is a semiconductor package that includes one or more semiconductor chips assembled on a package substrate.

[0068] Each DRAM cell 224 can include a vertical transistor 226 and a capacitor 228 coupled to the vertical transistor 226. DRAM cell 224 can be a 1T1C cell consisting of one transistor and one capacitor. It is understood that DRAM cell 224 may be of any suitable configurations, such as 2T1C cell, 3T1C cell, etc. The vertical transistor 226 can be a MOSFET used to switch a respective DRAM cell 224. In some implementations, the vertical transistor 226 includes a semiconductor body 230 (the active region in which a channel can form) extending vertically (in the z-direction), and a gate structure 236 in contact with one side of semiconductor body 230. In a single-gate vertical transistor, the semiconductor body 230 can have a cuboid shape or a cylinder shape, and the gate structure 236 can abut a single side of semiconductor body 230 in a plane view, e.g., as shown in FIG. 2. In some implementations, the vertical transistor 226 has a structure including two or more gates, e.g., a two-gates structure, a three-gates structure, or a gate all around (GAA)

structure. In some implementations, the gate structure 236 includes a gate electrode 234 and a gate dielectric 232 laterally between the gate electrode 234 and the semiconductor body 230 in a bit line direction (e.g., in the Y direction). In some implementations, the gate dielectric 232 abuts one side of the semiconductor body 230, and the gate electrode 234 abuts the gate dielectric 232.

[0069] As shown in FIG. 2, in some implementations, the semiconductor body 230 has two ends (the upper end and lower end in FIG. 2) in the vertical direction (the z-direction), and at least one end (e.g., the lower end) extends beyond gate dielectric 232 in the vertical direction (the z-direction) into the ILD layers. In some implementations, one end (e.g., the upper end) of the semiconductor body 230 is flush with the respective end (e.g., the upper end) of the gate dielectric 232. In some implementations, both ends (the upper end and lower end) of the semiconductor body 230 extend beyond the gate electrode 234, respectively, in the vertical direction (the z-direction) into ILD layers. That is, the semiconductor body 230 can have a larger vertical dimension (e.g., the depth) than that of the gate electrode 234 (e.g., in the z-direction), and neither the upper end nor the lower end of semiconductor body 230 is flush with the respective end of the gate electrode 234. Thus, short circuits between the bit lines 223 and the word lines 235 or between the word lines 235 and the capacitors 228 can be avoided. The vertical transistor 226 can further include a source and a drain (both referred to as 238 as their locations may be interchangeable) disposed at the two ends (the upper end and lower end) of the semiconductor body 230, respectively, in the vertical direction (the z-direction). In some implementations, one of the source or drain 238 (e.g., at the upper end in FIG. 2) is coupled to the capacitor 228, and the other one of source and drain 238 (e.g., at the lower end in FIG. 2) is coupled to the bit line 223. That is, the vertical transistor 226 can have a first terminal in the positive z-direction and a second terminal opposite the first terminal in the negative z-direction, as shown in FIG. 2.

[0070] In some implementations, the semiconductor body 230 includes semiconductor materials, such as single crystalline silicon, polysilicon, amorphous silicon, Ge, any other semiconductor materials, or any combinations thereof. In one example, semiconductor body 230 may include single crystalline silicon. Source and drain 238 can be doped with N+ type dopants (e.g., Phosphorus (P) or Arsenic (As)) or P-type dopants (e.g., Boron (B) or Gallium (Ga)) at a desired doping level. In some implementations, a silicide layer, such as a metal silicide layer, is formed between source/drain 238 of the vertical transistor 226 and the bit line 223 as the bit line contact or between source/drain 238 of the vertical transistor 226 and the first electrode of the

capacitor 228 as capacitor contact 242 to reduce the contact resistance. In some implementations, gate dielectric 232 includes dielectric materials, such as silicon oxide, silicon nitride, or high-k dielectrics including, but not limited to, Al_2O_3, HfO_2, Ta_2O_5, ZrO_2, TiO_2, or any combination thereof. In some implementations, gate electrode 234 includes a conductive material including, but not limited to W, Co, Cu, Al, TiN, TaN, polysilicon, silicide, or any combination thereof. In some implementations, the gate electrode 234 includes multiple conductive layers, such as a W layer over a TiN layer. In one example, the gate structure 236 may be a "gate oxide/gate poly" gate in which the gate dielectric 232 includes silicon oxide and gate electrode 234 includes doped polysilicon. In another example, gate structure 236 may be an HKMG in which gate dielectric 232 includes a high-k dielectric and gate electrode 234 includes a metal.

[0071] As described above, since the gate electrode 234 may be part of a word line or extend in the word line direction (e.g., the X direction) as a word line, the first semiconductor structure 202 of the memory device 200 can also include a plurality of word lines each extending in the word line direction. Each word line 235 can be coupled to a row of DRAM cells 224. That is, the bit line 223 and the word line 235 can extend in two perpendicular lateral directions, and the semiconductor body 230 of the vertical transistor 226 can extend in the vertical direction perpendicular to the two lateral directions in which the bit line 223 and the word line 235 extend. Word lines 235 are in contact with word line contacts (not shown). In some implementations, the word lines 235 include conductive materials including, but not limited to W, Co, Cu, Al, TiN, TaN, polysilicon, silicides, or any combination thereof. In some implementations, the word line 235 includes multiple conductive layers, such as a W layer over a TiN layer.

[0072] In some implementations, the vertical transistor 226 extends vertically through and contacts the word lines 235, and the source or drain 238 of vertical transistor 226 at the lower end thereof is in contact with the bit line 223 (or bit line contact if any). Accordingly, the word lines 235 and the bit lines 223 can be disposed in different planes in the vertical direction due to the vertical arrangement of vertical transistor 226, which simplifies the routing of the word lines 235 and the bit lines 223. In some implementations, the bit lines 223 are disposed vertically between the bonding layer 220 and the word lines 235, and the word lines 235 are disposed vertically between the bit lines 223 and the capacitors 228. The word lines 235 can be coupled to the peripheral circuits 212 in the second semiconductor structure 204 through word line contacts (not shown) in the interconnect layer 222, the bonding contacts 221 and 219 in the bonding layers 220

and 218, and the interconnects in the interconnect layer 216. Similarly, the bit lines 223 in the interconnect layer 222 can be coupled to the peripheral circuits 212 in the second semiconductor structure 204 through the bonding contacts 221 and 219 in the bonding layers 220 and 218 and the interconnects in the interconnect layer 216.

[0073] In some implementations, the vertical transistors 226 can be arranged in a mirror-symmetric manner to increase the density of DRAM cells 224 in the bit line direction (the Y direction). As shown in FIG. 2, two adjacent vertical transistors 226 in the bit line direction are mirror-symmetric to one another with respect to a trench isolation 260. That is, the first semiconductor structure 202 can include a plurality of trench isolations 260 each extending in the word line direction (the X direction) in parallel with word lines 235 and disposed between vertical gate electrodes 234 of two adjacent rows of the vertical transistors 226. In some implementations, the rows of vertical transistors 226 separated by the trench isolation 260 are mirror-symmetric to one another with respect to the trench isolation 260. The trench isolation 260 can be formed with dielectric materials including, but not limited to, silicon oxide, silicon nitride, silicon oxynitride, low-k dielectrics, or any combination thereof. It is understood that the trench isolation 260 may include an air gap each disposed laterally between adjacent vertical gate electrodes 234. Air gaps may be formed due to the relatively small pitches of vertical transistors 226 in the bit line direction (e.g., the Y direction). On the other hand, the relatively large dielectric constant of air in air gaps (e.g., about 4 times of the dielectric constant of silicon oxide) can improve the insulation effect between vertical transistors 226 (and rows of DRAM cells 224) compared with some dielectrics (e.g., silicon oxide). Similarly, in some implementations, air gaps are formed laterally between word lines 235 in the bit line direction as well, depending on the pitches of word lines 235 in the bit line direction. In some implementations, instead of the trench isolation 260 having the air gap being disposed between adjacent vertical gate electrodes 234 of two adjacent rows of the vertical transistors 226, a conductive structure (e.g., including metal such as W) is disposed between adjacent semiconductor bodies 230 of two adjacent rows of vertical transistors 226.

[0074] In some implementations, a capacitor 228 includes a first electrode 244 above and coupled to the source or drain 238 of vertical transistor 226, e.g., the upper end of the semiconductor body 230, via a capacitor contact 242. In some implementations, the capacitor contact 242 is an ohmic contact, such as a metal silicide contact, as opposed to a Schottky contact. For example, the capacitor contact 242 may include metal silicides, such as WSi, CoSi, CuSi, AlSi,

or any other suitable metal silicides having higher conductivities than doped silicon. The capacitor 228 can also include a capacitor dielectric above and in contact with the first electrode 244, and a second electrode above and in contact with the capacitor dielectric. That is, the capacitor 228 can be a vertical capacitor in which the electrodes and capacitor dielectric are stacked vertically (in the z-direction), and the capacitor dielectric can be sandwiched between the electrodes. In some implementations, each first electrode is coupled to source or drain 238 of a respective vertical transistor 226 in the same DRAM cell, while all second electrodes are coupled to a common plate 246 coupled to the ground, e.g., a common ground. The capacitor 228 can have a first end in the negative z-direction and a second end opposite the first end in the positive z-direction, as shown in FIG. 2. In some implementations, the first end of the capacitor 228 is coupled to the first terminal of the vertical transistor 226 via an ohmic contact (e.g., the capacitor contact 242 made of a metal silicide material). As shown in FIG. 2, the first semiconductor structure 202 can further include a capacitor contact 247 (e.g., a conductor) in contact with a common plate 246 for coupling the capacitors 228 to the peripheral circuits 212 or to the ground directly. In some implementations, the capacitor contact 247 (e.g., a conductor) extends in the z-direction from the dielectric layer of the bonding layer 220 to couple to the second end of the capacitor 228 via common plate 246, as shown in FIG. 2. In some implementations, the ILD layer in which the capacitors 228 are formed has the same dielectric material as the two ILD layers into which the semiconductor body 230 extends, such as silicon oxide.

[0075] It is understood that the structure and configuration of a capacitor 228 are not limited to the example in FIG. 2 and may include any suitable structure and configuration, such as a planar capacitor, a stack capacitor, a multi-fins capacitor, a cylinder capacitor, a trench capacitor, or a substrate-plate capacitor. In some implementations, the capacitor dielectric includes dielectric materials, such as silicon oxide, silicon nitride, or high-k dielectrics including, but not limited to, Al_2O_3, HfO_2, Ta_2O_5, ZrO_2, TiO_2, or any combination thereof. It is understood that in some examples, a capacitor 228 may be a ferroelectric capacitor used in a FRAM cell, and the capacitor dielectric may be replaced by a ferroelectric layer having ferroelectric materials, such as PZT or SBT. In some implementations, the electrodes include conductive materials including, but not limited to W, Co, Cu, Al, TiN, TaN, polysilicon, silicides, or any combination thereof.

[0076] As shown in FIG. 2, vertical transistor 226 extends vertically through and contacts the word lines 235, source or drain 238 of vertical transistor 226 at the lower end thereof is in contact

with the bit line 223, and source or drain 238 of vertical transistor 226 at the upper end thereof is coupled to the capacitor 228. That is, the bit line 223 and the capacitor 228 can be disposed in different planes in the vertical direction and coupled to opposite ends of vertical transistor 226 of DRAM cell 224 in the vertical direction due to the vertical arrangement of vertical transistor 226. In some implementations, the bit line 223 and the capacitor 228 are disposed on opposite sides of the vertical transistor 226 in the vertical direction, which simplifies the routing of the bit lines 223 and reduces the coupling capacitance between the bit lines 223 and the capacitors 228 compared with DRAM cells in which the bit lines and capacitors are disposed on the same side of the planar transistors.

[0077] As shown in FIG. 2, in some implementations, the vertical transistors 226 are disposed vertically between the capacitors 228 and the bonding interface 206. That is, the vertical transistors 226 can be arranged closer to the peripheral circuits 212 of the second semiconductor structure 204 and the bonding interface 206 than the capacitors 228. Since the bit lines 223 and the capacitors 228 are coupled to opposite ends of the vertical transistors 226, the bit lines 223 (as part of the interconnect layer 222) are disposed vertically between the vertical transistors 226 and the bonding interface 206 As a result, the interconnect layer 222 including bit lines 223 can be arranged close to the bonding interface 206 to reduce the interconnect routing distance and complexity.

[0078] In some implementations, the first semiconductor structure 202 further includes a substrate 248 disposed above the DRAM cells 224. The substrate 248 can be part of a carrier wafer. It is understood that in some examples, the substrate 248 may not be included in the first semiconductor structure 202.

[0079] In some implementations, the first semiconductor structure 202 can further include a pad-out interconnect layer 250 above the substrate 248 and the DRAM cells 224. The pad-out interconnect layer 250 can include interconnects, e.g., contact pads 254, in one or more ILD layers. The pad-out interconnect layer 250 and the interconnect layer 222 can be formed on opposite sides of the DRAM cells 224. The capacitors 228 can be disposed vertically between the vertical transistors 226 and the pad-out interconnect layer 250. In some implementations, the interconnects in pad-out interconnect layer 250 can transfer electrical signals between the memory device 200 and outside circuits, e.g., for pad-out purposes.

[0080] In some implementations, the first semiconductor structure 202 further includes one or more contacts 252 extending through the substrate 248 and part of the pad-out interconnect layer

250 to couple the pad-out interconnect layer 250 to the DRAM cells 224 and the interconnect layer 222. As a result, the peripheral circuits 212 can be coupled to the DRAM cells 224 through the interconnect layers 216 and 222 as well as the bonding layers 220 and 218, and the peripheral circuits 212 and the DRAM cells 224 can be coupled to outside circuits through contacts 252 and pad-out interconnect layer 250. Contact pads 254 and contacts 252 can include conductive materials including, but not limited to, W, Co, Cu, Al, silicides, or any combination thereof. In one example, the contact pad 254 may include Al, and the contact 252 may include W. In some implementations, the contact 252 includes a via surrounded by a dielectric spacer (e.g., having silicon oxide) to electrically separate the via from substrate 248. Depending on the thickness of substrate 248, contact 252 can be an ILV having a depth in the submicron level (e.g., between 20 nm and 1 μm), or a TSV having a depth in the micron- or tens micron- level (e.g., between 1 μm and 200 μm).

[0081] Although not shown, it is understood that the pad-out of memory devices is not limited to from the first semiconductor structure 202 having DRAM cells 224 as shown in FIG. 2 and may be from the second semiconductor structure 204 having peripheral circuit 212. Although not shown, it is also understood that the air gaps between word lines 235 and/or between semiconductor bodies 230 may be partially or fully filled with dielectrics. Although not shown, it is further understood that more than one array of DRAM cells 224 may be stacked over one another to vertically scale up the number of DRAM cells 224.

[0082] In some implementations, instead of having the substrate 248 above the DRAM cells 224 as shown in FIG. 2, the first semiconductor structure 202 includes a substrate disposed below the DRAM cells 224. The substrate can be part of a carrier wafer. The DRAM cells 224 can be formed in a front side of the substrate, and the bit lines 223 can be formed in a back side of the substrate. The bit lines 223 can be conductively coupled to the DRAM cells 224 (e.g., the source/drain 238 of the vertical transistors 226) through the substrate.

[0083] FIG. 3 illustrates a schematic diagram of a memory device 300 including peripheral circuits and an array of memory cells each having a vertical transistor, according to some aspects of the present disclosure. The memory device 300 can include a memory cell array 301 and peripheral circuits 302 coupled to memory cell array 301. With reference to FIGs. 1A-1B and FIG. 2, memory devices 100, 101, 200 may be examples of the memory device 300 in which memory cell array 301 and peripheral circuits 302 may be included in the first and the second semiconductor

structures 102 and 104, respectively. The memory cell array 301 can be any suitable memory cell array in which each memory cell 308 includes a vertical transistor 310 and a storage unit 312 coupled to vertical transistor 310. In some implementations, the memory cell array 301 is a DRAM cell array, and the storage unit 312 is a capacitor for storing charge as the binary information stored by the respective DRAM cell. In some implementations, the memory cell array 301 is a PCM cell array, and storage unit 312 is a PCM element (e.g., including chalcogenide alloys) for storing binary information of the respective PCM cell based on the different resistivities of the PCM element in the amorphous phase and the crystalline phase. In some implementations, the memory cell array 301 is a FRAM cell array, and the storage unit 312 is a ferroelectric capacitor for storing binary information of the respective FRAM cell based on the switch between two polarization states of ferroelectric materials under an external electric field.

[0084] As shown in FIG. 3, memory cells 308 can be arranged in a two-dimensional (2D) array having rows and columns. Memory device 300 can include word lines 304 coupling peripheral circuits 302 and memory cell array 301 for controlling the switch of vertical transistors 310 in memory cells 308 located in a row, as well as bit lines 306 coupling peripheral circuits 302 and memory cell array 301 for sending data to and/or receiving data from memory cells 308 located in a column. That is, each word line 304 is coupled to a respective row of memory cells 308, and each bit line is coupled to a respective column of memory cells 308 (e.g., a string of memory cells 308).

[0085] In some implementations, a memory cell 308 can include a vertical transistor 310, such as a vertical metal-oxide-semiconductor field-effect transistor (MOSFET), instead of a planar transistor as a pass transistor, to reduce the area occupied by the pass transistors of the memory cells 308, reduce the coupling capacitance, as well as reduce the interconnect routing complexity. As shown in FIG. 3, in some implementations, different from planar transistors in which the active regions are formed in the substrates, vertical transistor 310 includes a semiconductor body 314 extending vertically (in the z direction) above the substrate (not shown). That is, semiconductor body 314 can extend above the top surface of the substrate to expose not only the top surface of semiconductor body 314, but also one or more side surfaces thereof. As shown in FIG. 3, for example, semiconductor body 314 can have a cuboid shape to expose four sides thereof. It is understood that semiconductor body 314 may have any suitable 3D shape, such as polyhedron shapes or a cylinder shape. That is, the cross-section of semiconductor body 314 in the plan view

(e.g., in the x-y plane) can have a square shape, a rectangular shape (or a trapezoidal shape), a circular (or an oval shape), or any other suitable shapes.

[0086] As shown in FIG. 3, vertical transistor 310 can also include a gate structure 316 in contact with one or more sides of semiconductor body 314, i.e., in one or more planes of the side surface(s) of the active region. In other words, the active region of vertical transistor 310, i.e., semiconductor body 314, can be at least partially surrounded by gate structure 316. Gate structure 316 can include a gate dielectric 318 over one or more sides of semiconductor body 314, e.g., in contact with four side surfaces of semiconductor body 314 as shown in FIG. 3. Gate structure 316 can also include a gate electrode 320 over and in contact with gate dielectric 318. Gate dielectric 318 can include any suitable dielectric materials, such as silicon oxide, silicon nitride, silicon oxynitride, or high-k dielectrics. For example, gate dielectric 318 may include silicon oxide, i.e., gate oxide. Gate electrode 320 can include any suitable conductive materials, such as polysilicon, metals (e.g., tungsten (W), copper (Cu), aluminum (Al), etc.), metal compounds (e.g., titanium nitride (TiN), tantalum nitride (TaN), etc.), or silicides. For example, gate electrode 320 may include doped polysilicon, i.e., a gate poly. In some implementations, gate electrode 320 includes multiple conductive layers, such as a W layer over a TiN layer. It is understood that gate electrode 320 and word line 304 may be a continuous conductive structure in some examples. In other words, gate electrode 320 may be viewed as part of word line 304 that forms gate structure 316, or word line 304 may be viewed as the extension of gate electrode 320 to be coupled to peripheral circuits 302.

[0087] As shown in FIG. 3, vertical transistor 310 can further include a pair of a source and a drain (SID, dope regions, a.k.a., source electrode and drain electrode) formed at the two ends of semiconductor body 314 in the vertical direction (the z-direction), respectively. The source and drain can be doped with any suitable P-type dopants, such as boron (B) or Gallium (Ga), or any suitable N-type dopants, such as phosphorus (P) or arsenic (As). The source and drain can be separated by gate structure 316 in the vertical direction (the z-direction). In other words, gate structure 316 is formed vertically between the source and drain. As a result, one or more channels (not shown) of vertical transistor 310 can be formed in semiconductor body 314 vertically between the source and drain when a gate voltage applied to gate electrode 320 of gate structure 316 is above the threshold voltage of vertical transistor 310. That is, each channel of vertical transistors 310 is also formed in the vertical direction along which semiconductor body 314 extends, according to some implementations.

[0088] In some implementations, as shown in FIG. 3, vertical transistor 310 is a multi-gate transistor. That is, gate structure 316 can be in contact with more than one side of semiconductor body 314 (e.g., four sides in FIG. 3) to form more than one gate, such that more than one channel can be formed between the source and drain in operation. That is, different from the planar transistor that includes only a single planar gate (and resulting in a single planar channel), vertical transistor 310 shown in FIG. 3 can include multiple vertical gates on multiple sides of semiconductor body 314 due to the 3D structure of semiconductor body 314 and gate structure 316 that surrounds the multiple sides of semiconductor body 314. As a result, compared with planar transistors, vertical transistor 310 shown in FIG. 3 can have a larger gate control area to achieve better channel control with a smaller subthreshold swing. During the off state, since the channel is fully depleted, the leakage current of vertical transistor 310 can be significantly reduced as well. The multi-gate vertical transistors can include double-gate vertical transistors (e.g., dual-side gate vertical transistors), tri-gate vertical transistors (e.g., tri-side gate vertical transistors), and GAA vertical transistors.

[0089] It is understood that although vertical transistor 310 is shown as a multi-gate transistor in FIG. 3, the vertical transistors disclosed herein may also include single-gate transistors as described below in detail. That is, gate structure 316 may be in contact with a single side of semiconductor body 314, for example, for the purpose of increasing the transistor and memory cell density. It is also understood that although gate dielectric 318 is shown as being separate (i.e., a separate structure) from other gate dielectrics of adjacent vertical transistors (not shown), gate dielectric 318 may be part of a continuous dielectric layer having multiple gate dielectrics of vertical transistors.

[0090] In planar transistors and some lateral multiple-gate transistors (e.g., FinFET), the active regions, such as semiconductor bodies (e.g., Fins), extend laterally (in the x-y plane), and the source and the drain are disposed at different locations in the same lateral plane (the x-y plane). In contrast, in vertical transistor 310, semiconductor body 314 extends vertically (in the z-direction), and the source and the drain are disposed in the different lateral planes, according to some implementations. In some implementations, the source and the drain are formed at two ends of semiconductor body 314 in the vertical direction (the z direction), respectively, thereby being overlapped in the plan view. As a result, the area (in the x-y plane) occupied by vertical transistor 310 can be reduced compared with planar transistor and lateral multiple-gate transistors. Also, the

metal wiring coupled to vertical transistors 310 can be simplified since the interconnects can be routed in different planes. For example, bit lines 306 and storage units 312 may be formed on opposite sides of vertical transistor 310. In one example, bit line 306 may be coupled to the source or the drain at the upper end of semiconductor body 314, while storage unit 312 may be coupled to the other source or the drain at the lower end of semiconductor body 314.

[0091] FIG. 4 illustrates example peripheral circuits 302, according to some aspects of the present disclosure. The peripheral circuits 302 can be coupled to the memory cell array 301 through bit lines 306 and word lines 304. The peripheral circuits 302 can include any suitable analog, digital, and mixed-signal circuits for facilitating the operations of the memory cell array 301. The peripheral circuits 302 can include various types of peripheral circuits formed using metal-oxide-semiconductor (MOS) technologies. The example peripheral circuits include a row decoder/word line driver 402, a sense amplifier 404, a column decoder/data line driver 406, control logic 408, a command decoder 410, a mode register set/extended mode register set (MRS/EMRS) circuit 412, an address buffer 414, and a data input/output circuit 416. In some examples, additional peripheral circuits not shown in FIG. 3 may be included as well.

[0092] The sense amplifier 404 can sense and amplify data of a memory cell and can store data in the memory cell. The sense amplifier 404 can be implemented by a cross-coupled amplifier connected between a bit line and a complementary bit line, which are included in the memory cell array 301.

[0093] The data input/output circuit 416 can write input data to the memory cell array 301 based on an address signal (ADD), and can read output data from the memory cell array 301 based on address signal (ADD) and output the data to the outside of the memory device 400. To designate a memory cell for data to be written to or to be read from, the address signal (ADD) can be input to the address buffer 414, which can temporarily store the address signal (ADD).

[0094] The row decoder/word line driver 402 can decode a row address in the address signal (ADD) output from the address buffer 414, to designate a word line connected to a memory cell for data to be written to or to be read from. For example, in a data write or read mode, the row decoder/word line driver 402 can decode a row address output from the address buffer 414 and thus enable a word line corresponding to the row address. In addition, in a self-refresh mode, the row decoder/word line driver 402 can decode a row address generated by an address counter and thus enable a word line corresponding to the row address.

[0095] The column decoder/data line driver 406 can decode a column address in the address signal (ADD), which is output from the address buffer 414, to designate a bit line connected to a memory cell for data to be written to or to be read from. The memory cell array 301 can read data from or write data to a memory cell designated by the row and column addresses.

[0096] The command decoder 410 can receive a command signal (CMD) from a host or a memory controller, and can internally generate a decoded command signal by decoding such signals.

[0097] The MRS/EMRS circuit 412 can set a mode register in response to an MRS/EMRS command for designating an operation mode of the memory device 400.

[0098] The peripheral circuits may further include a clock circuit for generating a clock signal, a power supply circuit generating or distributing internal voltages by receiving power supply voltages applied from outside thereof, or the like.

[0099] The control logic 408 can be coupled to each peripheral circuit described above and configured to control operations of each peripheral circuit.

[0100] FIG. 5 illustrates a layout view of example peripheral circuits included in a second semiconductor structure 500. The second semiconductor structure 500 can be an example of the second semiconductor structure 104 of FIGS. 1A-1B, or the second semiconductor structure 204 of FIG. 2.

[0101] Peripheral circuits can be arranged in different regions of the second semiconductor structure 500. In some implementations, the peripheral circuits include bank circuits 502 that are each coupled to and configured to control a respective memory bank of the memory cell array of a first semiconductor structure (e.g., the first semiconductor structure 102 of FIGS. 1A-1B, or the second semiconductor structure 202 of FIG. 2). As an example, the peripheral circuits can include 16 bank circuits arranged in two rows. When the second semiconductor structure 500 is stacked with a corresponding first semiconductor structure, each bank circuit 502 can overlap, or partially overlap with a respective memory bank of the memory cell array in the plan view perpendicular to the stacking direction (e.g., in the X-Y plane).

[0102] In some implementations, a bank circuit 502 can include circuits that are configured to control a specific memory bank. For example, with reference to FIG. 4, each bank circuit 502 can include a row decoder/word line driver 402, a sense amplifier 404, and a column decoder/data line driver 406. In some implementations, two adjacent bank circuits 502 can share the use of row

decoders 508. As shown in FIG. 5, Bank Circuit 15 and Bank Circuit 13 share the use the row decoders 508 that are arranged between the two bank circuits; Bank Circuit 14 and Bank Circuit 12 share the use the row decoders 508 that are arranged between the two bank circuits; Bank Circuit 11 and Bank Circuit 9 share the use the row decoders 508 that are arranged between the two bank circuits; and so on. In some implementations, the row decoders 508 do not overlap with memory banks of the first semiconductor structure in the plan view perpendicular to the stacking direction.

[0103] In some implementations, the second semiconductor structure 500 can further include a side peripheral region 504 arranged to one side of the bank circuits 502, and/or a middle peripheral region 506 arranged between two rows of bank circuits 502. Circuits arranged in the side peripheral region 504 and the middle peripheral region 506 can be configured to control more than one memory bank of the memory cell array. For example, with reference to FIG. 4, one or more of the control logic 408, command decoder 410, the MRS/EMRS circuit 412, the address buffer 414, or the data input/output circuit 416 can be arranged in the side peripheral region 504 or the middle peripheral region 506.

[0104] FIG. 6 illustrates an example memory device 600 (e.g., memory device 300 of FIGS. 3-4). Compared with the memory device 100 of FIG. 1A, the memory device 101 of FIG. 1B, or the memory device 200 of FIG. 2, where the memory cell array and the peripheral circuits are formed on different semiconductor structures, in the memory device 600, the memory cell array of the memory device 600 and the peripheral circuits of the memory device 600 can be formed on the same semiconductor structure (e.g., on the same wafer).

[0105] As an example, the memory cell array of the memory device 600 includes 16 memory banks 606 (BANK 0 to 15) arranged in two rows. The peripheral circuits of the memory device 400 can include 16 bank circuits 608 and circuits arranged in the side peripheral region 602 and the middle peripheral region 604. The bank circuit 608 can also be arranged in two rows. The side peripheral region 602 can be arranged to one side of the bank circuits 608, and the middle peripheral region 604 can be arranged between the two rows of bank circuits 608.

[0106] As shown in FIG. 6, the memory cell array and the peripheral circuits of the memory device 400 can be formed on the same wafer. The memory bank 606 coupled to and controlled by each bank circuit 608 can be integrally formed and arranged in the same area as the bank circuit 608. As an example, bank circuits 608 can be formed surrounding respective memory banks 406.

[0107] FIG. 7 illustrates a plan view (e.g., in X-Y plane) of an example bank circuit 700 (e.g., bank circuit 502 of FIG. 5, or bank circuit 608 of FIG. 6).

[0108] In some implementations, the bank circuit 700 can include a plurality block circuits 702a, 702b (collectively 702). Each block circuit 702 can be coupled to a corresponding memory block of the memory bank, and configured to control the corresponding memory block. As an example, the bank circuit 700 can include n rows of block circuits 702, where each row includes m block circuits 702 along a first direction (e.g., the X direction) from a first boundary 701 (e.g., the left boundary) of the bank circuit 700 to a second boundary 703 (e.g., the right boundary) of the bank circuit 700. In some implementations, each block circuit 702 is stacked with a corresponding memory block along the vertical direction (e.g., the Z direction).

[0109] With reference to FIG. 8, a block circuit 702a that is on the first boundary 701 or the second boundary 703 of the bank circuit 700 (e.g., Block Circuit 00 of FIG. 7) can include three driver circuits 712a, 712b and 712c. The driver circuit 712a includes word line drivers (e.g., word line drives coupled to even-numbered word lines), the driver circuit 712b includes word line drivers (e.g., word line drivers coupled to even-numbered word lines), and the driver circuit 712c includes word line drivers (e.g., word line drivers coupled to odd-numbered word lines). In some implementations, as shown in FIG. 8, the boundaries of the block circuit 702a can include three linear structures, and a zig-zag structure (e.g., on the second boundary 703 of the bank circuit 700). As such, the boundary 701 and the boundary 703 of the bank circuit 700 can include zig-zag structures.

[0110] Further, a block circuit 702b that is on neither the first boundary 701 nor the second boundary 703 of the bank circuit 700 can include two driver circuits 712d, 712e. The driver circuit 712d includes word line drivers (e.g., word line drives coupled to odd-numbered word lines), the driver circuit 712e includes word line drivers (e.g., word line drivers coupled to even-numbered word lines). In some implementations, as shown in FIG. 8, the boundaries of the block circuit 702b can include four linear structures.

[0111] In some implementation, a driver circuit 712 can be shared by two adjacent block circuits 702 of the same bank circuit 700. For example, word line drivers in the driver circuits 712c and 712d can be coupled to odd-numbered word lines in both the block circuit 702a and the block circuit 702b; word line drivers in the driver circuits 712e and 712f can be coupled to even-numbered word lines in both the block circuit 702b and the block circuit adjacent to the block

circuit 702b. The driver circuit 712f can be included in the block circuit adjacent to the block circuit 702b.

[0112] The block circuit 702a can further include a sensing circuit 814a including sense amplifiers (e.g., sense amplifiers coupled to even-numbered bit lines), a sensing circuit 814b including sense amplifiers (e.g., sense amplifiers coupled to odd-numbered bit lines), column decoders 816, data line drivers 818, and remaining circuits 828. In some implementations, each sense amplifier is coupled to a bit line in the first semiconductor structure, where the bit line is coupled to a column of memory cells in the memory block stacked with the block circuit 702a. The sense amplifiers can be coupled to column decoders 816 to decode column addresses, and coupled to data line drivers 818 to receive signals to select/deselect bit lines. The remaining circuits 828 can include other circuits such as bit line regulators, word line regulators, etc.

[0113] In some implementations, the block circuit 702a occupies a larger area in XY plane than the corresponding memory block 802a. For example, when the first semiconductor structure and the second semiconductor structure are stacked together along the Z direction, while the driver circuit 712b and the driver circuit 712c may overlap with the memory block 802a in the plan view perpendicular to the Z direction, the driver circuit 712a and the driver circuit 712d may not overlap with the memory block 802a in the plan view. For example, the driver circuit 712a may be arranged outside of the projection of the memory bank corresponding to the bank circuit 700 in the X-Y plane. In some other implementations, the corresponding memory block 802a can occupy the same, or substantially same area in the XY plane as the block circuit 702a. For example, the memory bank 802a can also include memory cells overlapping with the driver circuit 712a in the plan view, such that the memory bank 802 can include a boundary having a zig-zag structure.

[0114] In some implementations, the driver circuits 712d and 712e of the driver circuit 702b can overlap with the memory block 802b in the plan view perpendicular to the Z direction.

[0115] Further, in some implementations, the data line drivers 818 do not overlap with the corresponding memory block 802a in the plan view. For example, the data line drivers 818 can be arranged between two adjacent memory blocks (e.g., along Y direction) in the plan view. Other block circuits (e.g., Block Circuits 10, Block Circuit 20, …, Block Circuit n0; Block Circuit m0, Block Circuit m1, …, Block Circuit mn of FIG. 7) that are on the first boundary 701 or the second boundary 703 of the bank circuit 700 may have similar structure as the block circuit 702a.

[0116] Referring back to FIG. 7, each block circuit 702a can have a driver circuit 712a offset from the rest of the circuits in the block circuit 702a. For example, each block circuit 702a can have a driver circuit 712a that is outside of the projection of the memory bank that corresponds to the bank circuit 700 in the X-Y plane. As such, the first boundary 701 and the second boundary 703 of the bank circuit 700 can have a zig-zag pattern.

[0117] With reference to FIG. 8, a block circuit 702b adjacent to the block circuit 702a along the first direction (e.g., the X direction) can share the use of the driver circuits 712c and 712d with the block circuit 702a. For example, odd-numbered word lines of the memory block 802b corresponding to the block circuit 702b can be coupled to word line drivers in the driver circuits 712c and 712d. Even-numbered word lines of the memory block 802b can be coupled to word line drivers in driver circuits 712e and 712f, which are shared with another block circuit adjacent to the block circuit 702b.

[0118] Similar to the block circuit 702a, the block circuit 702b can also include a sensing circuit including sense amplifiers coupled to even-numbered bit lines, a sensing circuit including sense amplifiers coupled to odd-numbered bit lines, column decoders, data line drivers, and remaining circuits. Other block circuits (e.g., Block Circuits 01, Block Circuit 11, Block Circuit 21, … of FIG. 7) that are not on the first boundary 701 or the second boundary 703 of the bank circuit 700 may have a similar structure as the block circuit 702b.

[0119] FIGS. 9A-9D illustrate layout view of example bank circuits that are adjacent to each other. For illustration, the bank circuit 700 (e.g., Bank Circuit 13 of FIG. 5) is adjacent to the bank circuit 900 (e.g., Bank Circuit 11 of FIG. 5) along the X direction.

[0120] For illustration purpose, the block circuit 702a that is closest to the bank circuit 900 among one row of block circuits is used as an example block circuit in the bank circuit 700. The block circuit 902a that is closest to the bank circuit 700 among one row of block circuits is used as an example block circuit in the bank circuit 900. The block circuit 702a includes driver circuits 712a-712c and sensing circuits 814a, 814b. The bank circuit 700 can further include a driver circuit 712d included in a bank circuit adjacent to bank circuit 702a. The block circuit 902a includes driver circuits 912a-912c and sensing circuits 914a, 914b. The bank circuit 900 can further include a driver circuit 912d included in a bank circuit adjacent to bank circuit 702a.

[0121] In some implementations, positioning of the circuits in the bank circuit 700 and positioning of the circuits in the bank circuit 900 are in mirror symmetry. As shown in FIG. 9A,

the positioning of driver circuits 712a-712d in relation to the sensing circuits 814a, 814b are in mirror symmetry as the positioning of driver circuits 912a-912d in relation to the sensing circuits 914a, 914b. As such, the bank circuit 700 and the bank circuit 900 may not be tightly arranged with one another. For example, the driver circuit 712b and the driver circuit 912b may be distanced from each other by a gap, and the boundary of the bank circuit 700 and the boundary of the bank circuit 900 may not fit with each other.

[0122] In some implementations, positioning of the circuits in the bank circuit 700 and positioning of the circuits in the bank circuit 900 are identical. For example, as shown in FIG. 9B, the positioning of driver circuits 712a-712d in relation to the sensing circuits 814a, 814b are identical to the positioning of driver circuits 912a-912d in relation to the sensing circuits 914a, 914b.

[0123] As shown in FIG. 9C, in some implementations, positioning of the circuits in the bank circuit 700 and positioning of the circuits in the bank circuit 900 are identical, so that the bank circuit 700 and the bank circuit 900 can be tightly arranged with one another. For example, at least a portion of the driver circuit 712a is adjacent to and overlaps with at least a portion of the driver circuit 912a along the Y direction. In some implementations, the driver circuit 712a and the driver circuit 912a can be aligned with each other along the Y direction. As such, the bank circuit 700 and the bank circuit 900 can be tightly arranged with one another. As such, the distance between the boundary of the bank circuit 700 and the boundary of the bank circuit 900 can be reduced, as shown by an common boundary line 930. In this way, a die size of the memory device can be reduced. For example, the die size of the memory device along the X direction can be reduced.

[0124] It should be noted that in some implementations, the driver circuit 712a and the driver circuit 912a can be immediately adjacent to each other along the Y direction. In some other implementations, the driver circuit 712a and the driver circuit 912a can be distanced from each other along the Y direction by a gap.

[0125] In some implementations, as shown in FIG. 9D, a boundary of the bank circuit 700 having a zig-zag pattern and a boundary of the bank circuit 900 having a zig-zag pattern can fit with each other. For example, driver circuits of bank circuits that are on the boundary of the bank circuit 700 and driver circuits of bank circuits that are on the boundary of the bank circuit 900 can be alternatively arranged with one another along the Y direction. With reference to FIG. 5, as an example, the bank circuit 700 can be Bank Circuit 13, and the bank circuit 900 can be Bank Circuit

11; or the bank circuit 700 can be Bank Circuit 9, and the bank circuit 900 can be Bank Circuit 7. In some implementations, row decoders are arranged between the driver circuits on the boundary of Bank Circuit 11 and driver circuits on the boundary of the BANK Circuit 9, such that the driver circuits of BANK Circuit 11 are not adjacent to the driver circuits of Bank Circuit 9 along the Y direction.

[0126] FIG. 10 illustrates a plan view of an example memory device 1000. The memory device 1000 can include a memory bank including memory blocks 1024, 1022, 1026 arranged along the X direction, and a bank circuit (e.g., bank circuit 700 of FIG. 7-8).

[0127] In some implementations, the bank circuit can include a first block circuit 1002 (e.g., coupled to and stacked with a first memory block 1022 in the memory bank), a second block circuit 1004 (e.g., coupled to and stacked with a second memory block 1024 in the memory bank), and a third block circuit 1006 (e.g., coupled to and stacked with a third memory block 1026 of the memory bank). The first block circuit 1002 can be arranged between the second block circuit 1004 and the third block circuit 1006 along the X direction,

[0128] In some implementations, adjacent memory blocks can share word lines (e.g., local word lines). For example, word lines are coupled to corresponding word line drivers in an interleaving way. Word lines coupled to memory cells in the first memory block 1022 can be grouped into two groups. For example, the word lines coupled to memory cells in the first memory block 1022 are numbered in order (e.g., consecutively from 0 to n). A first group of word lines include even-numbered word lines (e.g., WL0, WL2, WL4, …), and a second group of word lines include odd-number word lines (e.g., WL1, WL3, WL5, …). In some implementations, the first group of word lines are also coupled to memory cells in the second memory block 1024. The second group of the word lines are also coupled to memory cells in the third memory block 1026.

[0129] In some implementations, the first block circuit 1002 shares the use of word line drivers with the second block circuit 1004 and the third block circuit 1006. For example, the word line drivers coupled to even-numbered word lines in the upper half of the first memory block 1022 are included in driver circuit 1012; the word line drivers coupled to even-numbered word lines in the lower half of the first memory block1022 are included in the driver circuit 1014; the word line drivers coupled to odd-numbered word lines in the upper half of the first memory block 1022 are included in the driver circuit 1016; and the word line drivers coupled to odd-numbered word lines in the lower half of the first memory block 1022 are included in driver circuit 1018.

[0130] As shown in FIG. 11, in some implementations, word line drivers that each drive a word line coupled to memory cells in the memory block 1022 can be arranged into different groups. As an example, each group of word line drivers can include eight word line drivers that each drive one of eight consecutively-numbered word lines. As shown in FIG. 11, memory cells in the memory block 1022 are coupled to a total of 832 word lines numbered from 0 to 831. Odd-numbered word lines (e.g., WL1, WL3, WL7, …, WL831) are also coupled to memory cells in the third memory block 1026, while even-numbered word lines (e.g., WL0, WL2, WL4, …, WL830) are also coupled to memory cells in the second memory block (not shown in FIG. 11). The first group of word line drivers 1102 can include word line drivers that are coupled to WL0-7, the second group of word line drivers include word line drivers that are coupled to WL8-15, …, and the last group of word line drivers 1104 can include word line drivers that are coupled to WL824-831. In some implementations, half of each group of word line drivers 1102, 1104 are positioned on one side of the memory block 1022 in the plan view, and half of the each group of word line driver 1102, 1104 are positioned on the other side of the memory block 1022.

[0131] In some implementations, the word line drivers are configured to select/deselect word lines based on two driving signals. A first driving signal 1130 can be configured to select a group of word line drivers (e.g., a first group of word line drivers 1102), and a second driving signal 1132 can be configured to select one or more word line drivers (e.g., word line driver coupled to WL1) in the selected group of word line drivers. In some implementations, each memory bank can have a corresponding set of global word lines and row decoders. The first driving signal 1130 is generated by a global word line coupled to local word lines (e.g., WL0-8) of memory blocks 1022, 1026 in the memory bank, and the second driving signal 1132 can be generated by the row decoder.

[0132] FIG. 12 illustrates schematic circuit diagrams of example word line drivers. As illustrated in FIG. 11, half of the first group of word line drivers 1102 are coupled to odd-number word lines (e.g., WL1, WL3, WL5, WL7). A word line driver 1210 (e.g., word line driver coupled to WL7) can include a P-channel metal oxide semiconductor (PMOS) transistor 1212, a N-channel metal oxide semiconductor (NMOS) transistor 1214, and a keeping NMOS transistor 1216. The gate of the PMOS transistor 1212 receives the first driving signal 1130, a first terminal (e.g., source) of the PMOS transistor 1212 receives the second driving signal 1132, and a second terminal (e.g., drain) of the PMOS transistor 1212 is coupled to the local word line (e.g., WL7). The gate of the NMOS transistor 1214 can receive the first driving signal 1130, a first terminal (e.g., source) of

the NMOS transistor 1214 can receive a negative voltage VWLN, and a second terminal (e.g., drain) of the NMOS transistor 1214 is coupled to the local word line driver. The gate of the keeping NMOS transistor 1216 can receive an inverted signal of the second driving signal 1132, a first terminal (e.g., source) of the keeping NMOS transistor 1216 can receive the negative voltage VWLN, and a second terminal (e.g., drain) of the keeping NMOS transistor 1216 can be coupled to the local word line.

[0133] In some implementations, in order to select a word line (e.g., WL7), the first driving signal 1130 for the first group of word line drivers 1102 can be set to a low voltage, so that the NMOS transistors 1214 are switched off. The second driving signal 1132 for word line driver of the selected word line can be set to a high voltage (e.g., Vpp), so that the PMOS transistor 1212 is switched on, and the keeping NMOS transistor 1216 is switched off. As such, Vpp is applied to the word line to select/enable the word line. The second driving signal 1132 for word line drivers of unselected word lines in the first group of word lines can be set to a low voltage, so that the PMOS transistor 1212 is switched off, and the keeping NMOS transistor 1216 is switched on. As such, the word lines are deselected/disabled.

[0134] FIG. 13 is a flow chart of an example process 1300 of forming a memory device, according to some aspects of the present disclosure. The memory device can be the memory device 100 of FIG. 1A, the memory device 101 of FIG. 1B, or the memory device 200 of FIG. 2, the memory device 300 of FIGS. 3-4, the memory device 1000 of FIG, 10. The process 1300 includes operations (or steps) that can be performed with any suitable order and/or any combination.

[0135] At 1302, a first semiconductor structure (e.g., the first semiconductor structure 102 of FIG. 1A or 1B, or the first semiconductor structure 202 of FIG. 2) is formed. The first semiconductor structure can include a memory cell array that includes a first memory bank and a second memory bank. The memory cell array can be the memory cell array 301 of FIG. 3 that includes DRAM memory cells.

[0136] At 1304, a second semiconductor structure (e.g., the second semiconductor structure 104 of FIGs. 1A-1B, the second semiconductor structure 204 of FIG. 2, the second semiconductor structure 500 of FIG. 5) is formed. The second semiconductor structure includes a peripheral circuit (e.g., the peripheral circuit 302 of FIGS. 3-4). The peripheral circuit includes a first bank circuit (e.g., bank circuit 700 of FIG. 9D) and a second bank circuit (e.g., bank circuit 900 of FIG. 9D arranged along a first direction (e.g., the X direction). The first bank circuit includes a first

driver circuit (e.g., the driver circuit 712a of FIG. 9C) on a boundary of the first bank circuit, and the second bank circuit includes a second driver circuit (e.g., the driver circuit 912d of FIG. 9C) on a boundary of the second bank circuit. At least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction (e.g., the Y direction) perpendicular to the first direction.

[0137] In some implementations, the first bank circuit and the second bank circuit each include a plurality of block circuits arranged in one or more rows. The first driver circuit is included in a first block circuit (e.g., block circuit 702a of FIGS. 7-8 and 9C) that is on a boundary of the first bank circuit, and the second driver circuit is included in a second block circuit (e.g., block circuit 902a of FIG. 9C) that is on a boundary of the second bank circuit. The first block circuit can include first sense amplifiers (e.g., sense amplifiers in sensing circuits 814a, 814b of FIG. 9C) that are coupled to bit lines in a memory block of the first memory bank. The second block circuit can include second sense amplifiers (e.g., sense amplifiers in sensing circuits 914a, 914b of FIG. 9C) that are coupled to bit lines in a memory block of the second memory bank.

[0138] At 1306, the first semiconductor structure and the second semiconductor structure are stacked together along a third direction (e.g., the Z direction) to form a memory device (e.g., a bonded chip). In some implementations, after stacking and bonding the first semiconductor structure and the second semiconductor structure, the first bank circuit is coupled to the first memory bank, and the second bank circuit is coupled to the second memory bank.

[0139] In some implementations, the first block circuit further includes a third driver circuit (e.g., driver circuit 712b of FIG. 9C) including word line drivers. The third driver circuit overlaps with a first memory block in the plan view (e.g., in the XY plane), and the first driver circuit does not overlap with the first memory block in the plan view.

[0140] In some implementations, word lines are coupled to corresponding word line drivers in the driver circuits in an interleaving way. For example, word lines in the first memory block are numbered in a numerical order, e.g., from 0 to n. The first block circuit further includes a fourth driver circuit (e.g., driver circuit 712c of FIG. 9C) and a fifth driver circuit (e.g., driver circuit 712d of FIG. 9C). The first driver circuit and the third driver circuit are on a first boundary of the first block circuit and include a first set of word line drivers. The fourth driver circuit and the fifth driver circuit are on a second boundary of the first block circuit opposite to the first boundary, and include a second set of word line drivers. An even-numbered word line in the first memory block

is coupled to a corresponding word line driver of the first set of word line drivers, and an odd-numbered word line in the first memory block is coupled to a corresponding word line driver of the second set of word line drivers.

[0141] In some implementations, the second semiconductor structure includes a third bank circuit (e.g., Bank Circuit 9 of FIG. 5) adjacent to the second bank circuit (e.g., Bank Circuit 11 of FIG. 5) along the first direction. The third bank circuit is coupled to a third memory bank of the memory cell array. The second semiconductor structure includes row decoders (e.g., row decoders 508 of FIG. 5) between the second bank circuit and the third bank circuit.

[0142] In some implementations, with reference to FIG. 5, the second semiconductor structure includes a fourth bank circuit (e.g., the Bank Circuit 7 of FIG. 5) adjacent to the third bank circuit along the first direction, where the fourth bank circuit is coupled to a fourth memory bank of the memory cell array. The third bank circuit includes a sixth driver circuit on a boundary of the third bank circuit, and the fourth bank circuit includes a seventh driver circuit on a boundary of the fourth bank circuit. Similar to the first driver circuit and the second driver circuit, at least a portion of the sixth driver circuit is adjacent to and overlaps with at least a portion of the seventh driver circuit along the second direction.

[0143] FIG. 14 illustrates a block diagram of a system 1400 having one or more semiconductor devices (e.g., memory devices), according to one or more implementations of the present disclosure. The system 1400 can be a mobile phone, a desktop computer, a laptop computer, a tablet, a server, a vehicle computer, a gaming console, a printer, a positioning device, a wearable electronic device, a smart sensor, a virtual reality (VR) device, an argument reality (AR) device, or any other suitable electronic devices having storage therein. As shown in FIG. 14, the system 1400 can include a host device 1408 and a memory system 1402 having one or more memory devices 1404 and a memory controller 1406. Host device 1408 can include a processor of an electronic device, such as a central processing unit (CPU), or a system-on-chip (SoC), such as an application processor (AP). Host device 1408 can be configured to send or receive data to or from the one or more memory devices 1404.

[0144] A memory device 1404 can be any memory device disclosed herein, such as memory device depicted in any one of FIGs. 1-14. In some implementations, a memory device 1404 includes a DRAM memory. Memory controller 1406 (a.k.a., a controller circuit) is coupled to memory device 1404 and host device 1408. Consistent with implementations of the present

disclosure, memory device 1404 can include a plurality of conductive interconnections through a cover layer that are in contact with conductive pads in a conductive pad layer, and memory controller 1406 can be coupled to memory device 1404 through at least one of the plurality of conductive interconnections. Memory controller 1406 is configured to control memory device 1404. Memory controller 1406 can manage data stored in memory device 1404 and communicate with host device 1408. In some implementations, the memory system 1402 may not include the memory controller 1406, and the host device 1408 can function as a controller that controls operations of the memory device 1404.

[0145] FIG. 15A illustrate block diagram of an example system 1501 having one or more memory devices, according to one or more implementations of the present disclosure. The system 1501 can include a host device 1508, and a flash memory device 1504 coupled to the host device through a memory controller 1506. As shown in FIG. 15A, a DRAM memory device 1510 can be coupled to the host device 1508. The memory controller 1506 can be configured to receive data and instructions from the host device 1508, and control operations of memory devices (e.g., flash memory device 1504 and/or DRAM memory device 1510). In some implementations, the flash memory device 1504 can be a NAND memory device or a NOR memory device. The memory controller 1506 can include an interface 1512 configured to communicate with the host device 1508 according to a particular communication protocol. For example, the interface 1512 may communicate with the host device 1508 through at least one of various interface protocols, such as a USB protocol, a peripheral component interconnection (PCI) protocol, a PCIexpress (PCI-E) protocol, an advanced technology attachment (ATA) protocol, a serial-ATA (SATA) protocol, a parallel-ATA protocol, a small computer small interface (SCSI) protocol, an enhanced small disk interface (ESDI) protocol, an integrated drive electronics (IDE) protocol, a Firewire protocol, etc. The memory controller 1506 can further include a CPU 1514 configured to execute instructions, perform calculations, and control operations across the memory controller 1506. The memory controller 1506 can further include a Static Random-Access Memory (SRAM) 1516 for data storage, a flash controller 1518 configured to control operations (e.g., read, write, and erase operations) of one or more flash memory devices 1504, and a CPU 1514 configured to execute instructions, perform calculations, and control operations across the memory controller 1506.

[0146] FIG. 15B illustrates a block diagram of another example system 1502 having one or more memory devices, according to one or more implementations of the present disclosure. In some

implementations, as shown in FIG. 15B, the DRAM memory device 1510 can be coupled to the memory controller 1506. The memory controller 1506 can include a DRAM controller 1520 configured to control operations of the DRAM memory device 1510. In some implementations, the memory controller 1506 may not include the DRAM controller 1520, and the CPU 1514 can function as a controller that controls operations of the DRAM memory device 1510.

[0147] It is noted that references in the present disclosure to "one embodiment," "an embodiment," "an example embodiment," "some implementations," "some implementations," etc., indicate that the embodiment described can include a particular feature, structure, or characteristic, but every embodiment can not necessarily include the particular feature, structure, or characteristic. Moreover, such phrases do not necessarily refer to the same embodiment. Further, when a particular feature, structure or characteristic is described in connection with an embodiment, it would be within the knowledge of a person skilled in the pertinent art to affect such feature, structure or characteristic in connection with other implementations whether or not explicitly described.

[0148] In general, terminology can be understood at least in part from usage in context. For example, the term "one or more" as used herein, depending at least in part upon context, can be used to describe any feature, structure, or characteristic in a singular sense or can be used to describe combinations of features, structures or characteristics in a plural sense. Similarly, terms, such as "a," "an," or "the," again, can be understood to convey a singular usage or to convey a plural usage, depending at least in part upon context. In addition, the term "based on" can be understood as not necessarily intended to convey an exclusive set of factors and may, instead, allow for existence of additional factors not necessarily expressly described, again, depending at least in part on context.

[0149] It should be readily understood that the meaning of "on," "above," and "over" in the present disclosure should be interpreted in the broadest manner such that "on" not only means "directly on" something, but also includes the meaning of "on" something with an intermediate feature or a layer therebetween. Moreover, "above" or "over" not only means "above" or "over" something, but can also include the meaning it is "above" or "over" something with no intermediate feature or layer therebetween (i.e., directly on something).

[0150] Further, spatially relative terms, such as "beneath," "below," "lower," "above," "upper," and the like, can be used herein for ease of description to describe one element or feature's

relationship to another element(s) or feature(s) as illustrated in the figures. The spatially relative terms are intended to encompass different orientations of the device in use or process step in addition to the orientation depicted in the figures. The apparatus can be otherwise oriented (rotated 90 degrees or at other orientations) and the spatially relative descriptors used herein can likewise be interpreted accordingly.

[0151] As used herein, the term "substrate" refers to a material onto which subsequent material layers are added. The substrate includes a "top" surface and a "bottom" surface. The top surface of the substrate is typically where a semiconductor device is formed, and therefore the semiconductor device is formed at a top side of the substrate unless stated otherwise. The bottom surface is opposite to the top surface and therefore a bottom side of the substrate is opposite to the top side of the substrate. The substrate itself can be patterned. Materials added on top of the substrate can be patterned or can remain unpatterned. Furthermore, the substrate can include a wide array of semiconductor materials, such as silicon, germanium, gallium arsenide, indium phosphide, etc. Alternatively, the substrate can be made from an electrically non-conductive material, such as glass, plastic, or sapphire wafer.

[0152] As used herein, the term "layer" refers to a material portion including a region with a thickness. A layer has a top side and a bottom side where the bottom side of the layer is relatively close to the substrate and the top side is relatively away from the substrate. A layer can extend over the entirety of an underlying or overlying structure, or can have an extent less than the extent of an underlying or overlying structure. Further, a layer can be a region of a homogeneous or inhomogeneous continuous structure that has a thickness less than the thickness of the continuous structure. For example, a layer can be located between any set of horizontal planes between, or at, a top surface and a bottom surface of the continuous structure. A layer can extend horizontally, vertically, and/or along a tapered surface. A substrate can be a layer, can include one or more layers therein, and/or can have one or more layer thereupon, thereabove, and/or therebelow. A layer can include multiple layers. For example, an interconnect layer can include one or more conductive and contact layers (in which contacts, interconnect lines, and/or vertical interconnect accesses (VIAs) are formed) and one or more dielectric layers.

[0153] As used herein, the term "nominal/nominally" refers to a desired, or target, value of a characteristic or parameter for a component or a process step, set during the design phase of a product or a process, together with a range of values above and/or below the desired value. As

used herein, the range of values can be due to slight variations in manufacturing processes or tolerances. As used herein, the term "about" indicates the value of a given quantity that can vary based on a particular technology node associated with the subject semiconductor device. Based on the particular technology node, the term "about" can indicate a value of a given quantity that varies within, for example, 10-30% of the value (e.g., ±10%, ±20%, or ±30% of the value).

[0154] In the present disclosure, the term "horizontal/horizontally/lateral/laterally" means nominally parallel to a lateral surface of a substrate, and the term "vertical" or "vertically" means nominally perpendicular to the lateral surface of a substrate. The terms "operation" and "step" can be used interchangeably to describe a process.

[0155] The present disclosure provides many different implementations, or examples, for implementing different features of the provided subject matter. Specific examples of components and arrangements are described below to simplify the present disclosure. These are, of course, merely examples and are not intended to be limiting. For example, the formation of a first feature over or on a second feature in the description that follows may include implementations in which the first and second features may be in direct contact, and may also include implementations in which additional features may be formed between the first and second features, such that the first and second features may not be in direct contact. In addition, the present disclosure may repeat reference numerals and/or letters in the various examples. This repetition is for the purpose of simplicity and clarity and does not in itself dictate a relationship between the various implementations and/or configurations discussed.

[0156] The foregoing description of the specific implementations can be readily modified and/or adapted for various applications. Therefore, such adaptations and modifications are intended to be within the meaning and range of equivalents of the disclosed implementations, based on the teaching and guidance presented herein.

[0157] While the present disclosure contains many specific implementation details, these should not be construed as limitations on the scope of what is being claimed, which is defined by the claims themselves, but rather as descriptions of features that may be specific to particular implementations. Certain features that are described in this present disclosure in the context of separate implementations can also be implemented in combination in a single implementation. Conversely, various features that are described in the context of a single implementation can also be implemented in multiple implementations separately or in any suitable sub-combination.

Moreover, although features may be described above as acting in certain combinations and even initially be claimed as such, one or more features from a claimed combination can in some cases be excised from the combination, and the claim may be directed to a sub-combination or variation of a sub-combination.

[0158] Similarly, while operations are depicted in the drawings and recited in the claims in a particular order, this should not be understood as requiring that such operations be performed in the particular order shown or in sequential order, or that all illustrated operations be performed, to achieve desirable results. In certain circumstances, multitasking and parallel processing may be advantageous. Moreover, the separation of various system modules and components in the implementations described above should not be understood as requiring such separation in all implementations, and it should be understood that the described program components and systems can generally be integrated together in a single software product or packaged into multiple software products.

[0159] Particular implementations of the subject matter have been described. Other implementations also are within the scope of the following claims. For example, the actions recited in the claims can be performed in a different order and still achieve desirable results. As one example, the processes depicted in the accompanying figures do not necessarily require the particular order shown, or sequential order, to achieve desirable results. In some cases, multitasking and parallel processing may be advantageous.

[0160] The breadth and scope of the present disclosure should not be limited by any of the above-described exemplary implementations, but should be defined only in accordance with the following claims and their equivalents.

CLAIMS

What is claimed is:

1. A memory device, comprising:

a first semiconductor structure comprising a memory cell array, wherein the memory cell array comprises a first memory bank and a second memory bank; and

a second semiconductor structure comprising a peripheral circuit, wherein the peripheral circuit comprises a first bank circuit and a second bank circuit arranged along a first direction, wherein the first bank circuit is coupled to the first memory bank and the second bank circuit is coupled to the second memory bank,

wherein the first bank circuit comprises a first driver circuit on a boundary of the first bank circuit, and the second bank circuit comprises a second driver circuit on a boundary of the second bank circuit, and

wherein at least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction perpendicular to the first direction.

2. The memory device of claim 1, wherein the first semiconductor structure and the second semiconductor structure are stacked along a third direction perpendicular to the first direction and the second direction,

wherein the first memory bank overlaps with the first bank circuit in a plan view perpendicular to the third direction, and

wherein the second memory bank overlaps with the second bank circuit in the plan view.

3. The memory device of claim 2, wherein the first bank circuit comprises block circuits arranged in one or more rows, wherein the first driver circuit is comprised in a first block circuit that is closest to the second bank circuit among one row of block circuits of the first bank circuit, and

wherein the second bank circuit comprises block circuits arranged in one or more rows, wherein the second driver circuit is comprised in a second block circuit that is closest to the first bank circuit among one row of block circuits of the second bank circuit.

4. The memory device of claim 3, wherein the first block circuit further comprises first sense amplifiers coupled to bit lines in a memory block of the first memory bank, and

wherein the second bank circuit further comprises second sense amplifiers coupled to bit lines in a memory block of the second memory bank.

5. The memory device of claim 4, wherein positioning of the first driver circuit and the first sense amplifiers in the first block circuit is identical to positioning of the second driver circuit and the second sense amplifiers in the second block circuit.

6. The memory device of any one of claims 3 to 5, wherein the first memory bank comprises a first memory block that is coupled to the first block circuit,

wherein the first block circuit further comprises a third driver circuit,

wherein the third driver circuit overlaps with the first memory block in the plan view, and the first driver circuit does not overlap with the first memory block in the plan view.

7. The memory device of claim 6, wherein word lines in the first memory block are numbered in a numerical order,

wherein the first bank circuit further comprises a fourth driver circuit and a fifth driver circuit,

wherein the first driver circuit and the third driver circuit are on a first boundary of the first block circuit and comprise a first set of word line drivers,

wherein the fourth driver circuit and the fifth driver circuit are on a second boundary of the first block circuit opposite to the first boundary, and comprise a second set of word line drivers, and

wherein an even-numbered word line in the first memory block is coupled to a corresponding word line driver of the first set of word line drivers, and an odd-numbered word line in the first memory block is coupled to a corresponding word line driver of the second set of word line drivers.

8. The memory device of any one of claims 1 to 7, wherein the peripheral circuit comprises:

a third bank circuit adjacent to the second bank circuit along the first direction, wherein the third bank circuit is coupled to a third memory bank of the memory cell array; and

row decoders between the second bank circuit and the third bank circuit.

9. The memory device of claim 8, wherein the peripheral circuit comprises a fourth bank circuit adjacent to the third bank circuit along the first direction, wherein the fourth bank circuit is coupled to a fourth memory bank of the memory cell array,

wherein the third bank circuit comprises a sixth driver circuit on a boundary of the third bank circuit, and the fourth bank circuit comprises a seventh driver circuit on a boundary of the fourth bank circuit, and

wherein at least a portion of the sixth driver circuit is adjacent to and overlaps with at least a portion of the seventh driver circuit along the second direction.

10. The memory device of any one of claims 1 to 9, wherein the memory cell array comprises DRAM memory cells.

11. The memory device of any one of claims 1 to 10, wherein the first semiconductor structure and the second semiconductor structure comprise bonding contacts that bond the first semiconductor structure and the second semiconductor structure together, wherein the bonding contacts are isolated by an isolating material.

12. A method of forming a memory device, comprising:

forming a first semiconductor structure comprising a memory cell array, wherein the memory cell array comprises a first memory bank and a second memory bank; and

forming a second semiconductor structure comprising a peripheral circuit, wherein the peripheral circuit comprises a first bank circuit and a second bank circuit arranged along a first direction, wherein the first bank circuit is coupled to the first memory bank and the second bank circuit is coupled to the second memory bank,

wherein the first bank circuit comprises a first driver circuit on a boundary of the first bank circuit, and the second bank circuit comprises a second driver circuit on a boundary of the second bank circuit, and

wherein at least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction perpendicular to the first direction.

13. The method of claim 12, comprising:

stacking the first semiconductor structure and the second semiconductor structure along a third direction perpendicular to the first direction and the second direction,

wherein the first memory bank overlaps with the first bank circuit in a plan view perpendicular to the third direction, and

wherein the second memory bank overlaps with the second bank circuit in the plan view.

14. The method of claim 13, wherein the first bank circuit comprises block circuits arranged in one or more rows, wherein the first driver circuit is comprised in a first block circuit that is closest to the second bank circuit among one row of block circuits of the first bank circuit, and

wherein the second bank circuit comprises block circuits arranged in one or more rows, wherein the second driver circuit is comprised in a second block circuit that is closest to the first bank circuit among one row of block circuits of the second bank circuit.

15. The method of claim 14, comprising:

forming first sense amplifiers in the first bank circuit, wherein the first sense amplifiers are coupled to bit lines in a memory block of the first memory bank; and

forming second sense amplifiers in the second bank circuit, wherein the second sense amplifiers are coupled to bit lines in a memory block of the second memory bank.

16. The method of claim 15, wherein positioning of the first driver circuit and the first sense amplifiers in the first block circuit is identical to positioning of the second driver circuit and the second sense amplifiers in the second block circuit.

17. The method of any one of claims 14 to 16, wherein the first memory bank comprises a first memory block that is coupled to the first block circuit,

wherein the first block circuit further comprises a third driver circuit,

wherein the third driver circuit overlaps with the first memory block in the plan view, and the first driver circuit does not overlap with the first memory block in the plan view.

18. The method of claim 17, wherein word lines in the first memory block are numbered in a numerical order,

wherein the first bank circuit further comprises a fourth driver circuit and a fifth driver circuit,

wherein the first driver circuit and the third driver circuit are on a first boundary of the first block circuit and comprise a first set of word line drivers,

wherein the fourth driver circuit and the fifth driver circuit are on a second boundary of the first block circuit opposite to the first boundary, and comprise a second set of word line drivers, and

wherein an even-numbered word line in the first memory block is coupled to a corresponding word line driver of the first set of word line drivers, and an odd-numbered word line in the first memory block is coupled to a corresponding word line driver of the second set of word line drivers.

19. The method of any one of claims 12 to 18, wherein forming the second semiconductor structure comprises:

forming a third bank circuit adjacent to the second bank circuit along the first direction, wherein the third bank circuit is coupled to a third memory bank of the memory cell array; and

forming row decoders between the second bank circuit and the third bank circuit.

20. A memory system, comprising:

a memory device comprising:

a first semiconductor structure comprising a memory cell array, wherein the memory cell array comprises a first memory bank and a second memory bank; and

a second semiconductor structure comprising a peripheral circuit, wherein the peripheral circuit comprises a first bank circuit and a second bank circuit arranged along a first direction, wherein the first bank circuit is coupled to the first memory bank and the second bank circuit is coupled to the second memory bank,

wherein the first bank circuit comprises a first driver circuit on a boundary of the first bank circuit, and the second bank circuit comprises a second driver circuit on a boundary of the second bank circuit, and

wherein at least a portion of the first driver circuit is adjacent to and overlaps with at least a portion of the second driver circuit along a second direction perpendicular to the first direction; and

a controller coupled to the memory device and configured to control the memory device.