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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 March 2021 (25.03.2021)

(10) International Publication Number
WO 2021/055206 A1

(51) International Patent Classification:
G06F 12/1081 (2016.01) *G06F 21/79* (2013.01)
G06F 12/14 (2006.01) *G11C 7/10* (2006.01)

(21) International Application Number:
PCT/US2020/049938

(22) International Filing Date:
09 September 2020 (09.09.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
16/573,780 17 September 2019 (17.09.2019) US

(71) Applicant: **MICRON TECHNOLOGY, INC.** [US/US]; 8000 South Federal Way, Boise, Idaho 83707 (US).

(72) Inventors: **BRADSHAW, Samuel E.**; 6319 Surfside Way, Sacramento, California 95831 (US). **SWAMI, Shivam**; 200 South Lexington Drive, Apt 816, Folsom, California 95630 (US). **EILERT, Sean S.**; 1721 English Colony Way, Penryn, California 95663 (US). **ENO, Justin M.**; 3844 Yellowstone Lane, El Dorado Hills, California 95762 (US). **AKEL, Ameen D.**; 3408 Grappa Way, Rancho Cordova, California 95670 (US).

(74) Agent: **WARD, John P.** et al.; GREENBERG TRAURIG, LLP, 77 West Wacker Drive, Suite 3100, Chicago, Illinois 60601 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

(54) **Title:** MEMORY CHIP HAVING AN INTEGRATED DATA MOVER

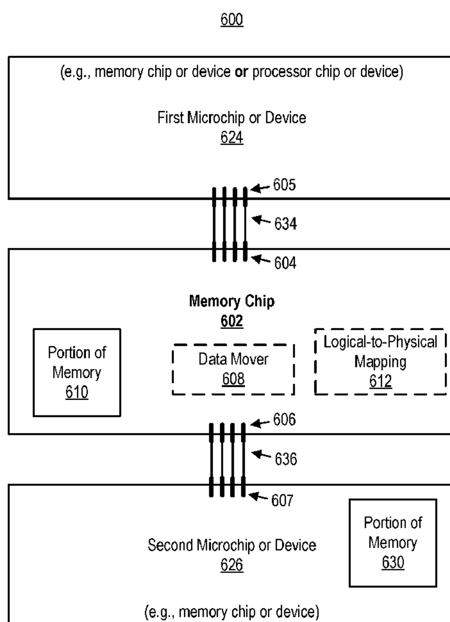
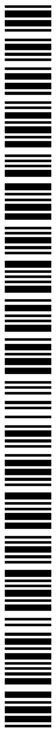


FIG. 6

(57) **Abstract:** A memory chip having a first set of pins configured to allow the memory chip to be coupled to a first microchip or device via first wiring. The memory chip also having a second set of pins configured to allow the memory chip to be coupled to a second microchip or device via second wiring that is separate from the first wiring. The memory chip also having a data mover configured to facilitate access to the second microchip or device, via the second set of pins, to read data from the second microchip or device and write data to the second microchip or device. Also, a system having the memory chip, the first microchip or device, and the second microchip or device.



WO 2021/055206 A1

[Continued on next page]

Published:
— *with international search report (Art. 21(3))*

MEMORY CHIP HAVING AN INTEGRATED DATA MOVER

RELATED APPLICATION

[0001] The present application claims priority to U.S. Pat. App. Ser. No. 16/573,780, filed Sep. 17, 2019 and entitled "MEMORY CHIP HAVING AN INTEGRATED DATA MOVER," the entire disclosure of which is hereby incorporated herein by reference.

FIELD OF THE TECHNOLOGY

[0002] At least some embodiments disclosed herein relate to a memory chip having an integrated data mover. Also, at least some embodiments disclosed herein relate to using such a memory chip in flexible provisioning of a string of memory chips to form a memory.

BACKGROUND

[0003] Memory of a computing system can be hierarchical. Often referred to as memory hierarchy in computer architecture, memory hierarchy can separate computer memory into a hierarchy based on certain factors such as response time, complexity, capacity, persistence and memory bandwidth. Such factors can be related and can often be tradeoffs which further emphasizes the usefulness of a memory hierarchy.

[0004] In general, memory hierarchy affects performance in a computer system. Prioritizing memory bandwidth and speed over other factors can require considering the restrictions of a memory hierarchy, such as response time, complexity, capacity, and persistence. To manage such prioritization, different types of memory chips can be combined to balance chips that are faster with chips that are more reliable or cost effective, etc. Each of the various chips can be viewed as part of a memory hierarchy. And, for example, to reduce latency on faster chips, other chips in a memory chip combination can respond by filling a buffer and then signaling for activating the transfer of data between chips.

[0005] Memory hierarchy can be made of up of chips with different types of memory units. For example, memory units can be dynamic random-access memory (DRAM) units. DRAM is a type of random access semiconductor memory that stores each bit of data in a memory cell, which usually includes a capacitor and a metal-oxide-

semiconductor field-effect transistor (MOSFET). The capacitor can either be charged or discharged which represents the two values of a bit, “0” and “1”. In DRAM, the electric charge on a capacitor leaks off, so DRAM requires an external memory refresh circuit which periodically rewrites the data in the capacitors by restoring the original charge per capacitor. On the other hand, with static random-access memory (SRAM) units a refresh feature is not needed. Also, DRAM is considered volatile memory since it loses its data rapidly when power is removed. This is different from flash memory and other types of non-volatile memory, such as non-volatile random-access memory (NVRAM), in which data storage is more persistent.

[0006] A type of NVRAM is 3D XPoint memory. With 3D XPoint memory, memory units store bits based on a change of bulk resistance in conjunction with a stackable cross-gridded data access array. 3D XPoint memory can be more cost effective than DRAM but less cost effective than flash memory.

[0007] Flash memory is another type of non-volatile memory. An advantage of flash memory is that is can be electrically erased and reprogrammed. Flash memory is considered to have two main types, NAND-type flash memory and NOR-type flash memory, which are named after the NAND and NOR logic gates that can implement the memory units of flash memory. The flash memory units or cells exhibit internal characteristics similar to those of the corresponding gates. A NAND-type flash memory includes NAND gates. A NOR-type flash memory includes NOR gates. NAND-type flash memory may be written and read in blocks which can be smaller than the entire device. NOR-type flash permits a single byte to be written to an erased location or read independently. Because of advantages of NAND-type flash memory, such memory has been often utilized for memory cards, USB flash drives, and solid-state drives. However, a primary tradeoff of using flash memory in general is that it is only capable of a relatively small number of write cycles in a specific block compared to other types of memory such as DRAM and NVRAM.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The present disclosure will be understood more fully from the detailed description given below and from the accompanying drawings of various embodiments of the disclosure.

[0009] FIG. 1 illustrates an example memory system that is configured to provide flexible provisioning of multi-tier memory, in accordance with some embodiments of

the present disclosure.

[0010] FIG. 2 illustrates an example memory system and processor chip configured to provide flexible provisioning of multi-tier memory, in accordance with some embodiments of the present disclosure.

[0011] FIG. 3 illustrates an example memory system and memory controller chip configured to provide flexible provisioning of multi-tier memory, in accordance with some embodiments of the present disclosure.

[0012] FIG. 4 illustrates an example memory system configured to provide flexible provisioning of multi-tier memory with tiers that each include multiple memory chips, in accordance with some embodiments of the present disclosure.

[0013] FIG. 5 illustrates example parts of an example computing device, in accordance with some embodiments of the present disclosure.

[0014] FIG. 6 illustrates an example system that includes a memory chip having an integrated data mover, in accordance with some embodiments of the present disclosure.

[0015] FIG. 7 illustrates an example system that includes the memory chip shown in FIG. 6 as well as a microchip or device that receives data transferred from the data mover and has logical-to-physical mapping for mapping the received data, in accordance with some embodiments of the present disclosure.

[0016] FIG. 8 illustrates an example system that includes the memory chip shown in FIG. 6 which is also shown having encryption and authentication circuitry, in accordance with some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0017] At least some embodiments disclosed herein relate to a memory chip having an integrated data mover (e.g., see FIGS. 6-8 as well as corresponding text herein). Also, at least some embodiments disclosed herein relate to using such a memory chip in flexible provisioning of a string of memory chips to form a memory (e.g., see FIGS. 1-8 as well as corresponding text herein). For example, one or more instances of the memory chip depicted in FIGS. 6-8 can be used as one or more memory chips in flexible provisioning of the string of memory chips 102 shown in FIGS. 1-3 as well as the string of memory chips 402 shown in FIG. 4.

[0018] For the purposes of this disclosure, a data mover is a circuit in a memory chip or device that manages the transfer of data to another memory chip or device.

Such a data mover can be used in a group of memory chips or devices in a memory hierarchy. Thus, a data mover can facilitate movement of data from one memory chip or device to another memory chip or device in a memory hierarchy.

[0019] The memory chip (e.g., see memory chip 602) that includes the integrated data mover (e.g., see data mover 608) can have two separate sets of pins (e.g., see sets of pins 604 and 606 shown in FIGS. 6-8) for respective separate connections to a first microchip or device (e.g., see first microchip or device 624 shown in FIGS. 6-8) and a second microchip or device (e.g., see second microchip or device 626 shown in FIGS. 6-8). The first microchip or device can be a processor, such as a system on a chip (SoC), or another memory chip. The second microchip or device can be another memory chip or a memory device such as a mass storage device.

[0020] In some embodiments, the memory chip can include an encryption engine (e.g., see encryption engine 802 shown in FIG. 8) in addition to the data mover to secure data to be moved to the second microchip or device. Also, in such embodiments and others, the memory chip can include a gatekeeper device (e.g., see gatekeeper 804 shown in FIG. 8) that can perform authentication for access of data stored in the second microchip or device.

[0021] The data mover can combine data stored in the memory chip (e.g., see portion of memory 610 having data accessible by the first microchip or device 624 as shown in FIGS. 6-8) on its way to the second microchip or device using various strategies. This can improve write performance and endurance of the second microchip or device.

[0022] In general, the memory chip can include a first set of pins, a second set of pins, and an integrated data mover. The first set of pins can be configured to allow the memory chip to be coupled to a first microchip or device via first wiring. The second set of pins can be configured to allow the memory chip to be coupled to a second microchip or device via second wiring that is separate from the first wiring. The data mover can be configured to facilitate access to the second microchip or device, via the second set of pins, to read data from the second microchip or device and write data to the second microchip or device.

[0023] The memory chip can be a non-volatile random-access memory (NVRAM) chip in that the memory chip includes a plurality of NVRAM cells. And, in some embodiments, the plurality of NVRAM cells can include a plurality of 3D XPoint memory cells. Also, the memory chip can be a dynamic random-access memory

(DRAM) chip in that the memory chip includes a plurality of DRAM cells. Also, the memory chip can be a flash memory chip in that the memory chip includes a plurality of flash memory cells. The plurality of flash memory cells can include a plurality of NAND-type flash memory cells.

[0024] The first microchip or device can be another memory chip or a memory device or a processor chip or a processor device. In some embodiments, for example, the first microchip or device is a SoC. In some embodiments, for example, the first microchip or device is a DRAM chip. In some embodiments, for example, the first microchip or device is a NVRAM chip. Data stored in a portion of the memory chip can be accessible by or through the first microchip or device via the first set of pins. Also, when the data stored in the portion of the memory chip is accessible through the first microchip or device it is being accessed by another memory chip or device or a processor chip or device. And, the first microchip or device can read data from the memory chip as well as write data to the memory chip.

[0025] The second microchip or device can be another memory chip or a memory device. In some embodiments, for example, the second microchip or device is a DRAM chip. In some embodiments, for example, the second microchip or device is a NVRAM chip. In some embodiments, for example, the second microchip or device is a flash memory chip (e.g., a NAND-type flash memory chip). Data stored in a portion of the second microchip or device can be accessible by or through the memory chip via the second set of pins. Also, when the data stored in the portion of the second microchip or device is accessible through the memory chip it is being accessed by another memory chip or device or a processor chip or device (such as the first microchip or device). And, the memory chip can read data from the second microchip or device as well as write data to the second microchip or device.

[0026] The data mover can be configured to combine the data stored in a portion of the memory chip by moving the data in blocks to the second microchip or device. For example, the data mover can be configured to combine the data stored in the portion of the memory chip that is accessible by or through the first microchip or device via the first set of pins by moving the data in blocks to the second microchip or device. The data mover by moving data in blocks can increase write performance and endurance of the second microchip or device, and sequential or block access on memory chips is orders of magnitude faster than random access on memory chips.

[0027] In some embodiments, the blocks are at a granularity that is coarser than

the data initially stored in a portion of the memory chip. For example, the blocks are at a granularity that is coarser than the data initially stored in the portion of the memory chip that is accessible by or through the first microchip or device via the first set of pins. The blocks being at a granularity that is coarser than the pre-blocked data in the memory chip, such as the data to be accessed by first microchip or device, can reduce the frequency of data writes to the second microchip or device.

[0028] The data mover can also be configured to buffer movement of changes to the data stored in a portion of the memory chip, such as data stored in the portion of the memory chip that is accessible to the first microchip or device. And, in such embodiments, the data mover can also be configured to send write requests to the second microchip or device in a suitable size due to the buffering by the data mover. When the second microchip or device is a second microchip or device and a write to the second microchip or device is in the suitable size due to the buffering by the data mover, the second microchip or device can erase a block and program the block in the second microchip or device according to the write without further processing or minimal processing in the second microchip or device. This is one example way that the data mover integrated in the memory chip can improve write performance and endurance of the second microchip or device. Also, with the buffering by the data mover and when frequent and/or random changes are made to the data in a portion of the memory chip (such as the data in the portion of the chip accessible by the first microchip or device), the second microchip or device does not have to be frequently erased and reprogrammed in a corresponding way as the changes occurring in the portion of the memory chip.

[0029] The buffering by the data mover is even more beneficial when the second microchip or device is a flash memory chip because the buffering can remove or at least limit the effects of write amplification that occurs in flash memory. Write amplification can be reduced or even eliminated by the buffering because, with the buffering, a write request sent by the memory chip can be modified to a suitable size or granularity expected by the receiving flash memory chip. Thus, the flash memory chip can erase a block and program the block according to the write request without possible duplication of the write and thus avoid further processing in the second microchip or device.

[0030] The data mover can also be configured to bundle changed addresses in the changes to the data stored in a portion of the memory chip (such as the data in the

portion of the memory chip accessible to the first microchip or device). In such embodiments, the data mover can also be configured to write bundled changed addresses into another portion of the memory chip to be moved to the second microchip or device via a write request to the second microchip or device. These features of the data mover can improve the buffering by the data mover and the sending of write requests according to the buffering since the bundling of changed addresses can be controlled by the data mover to correspond to a suitable size or granularity expected by the receiving second microchip or device.

[0031] In some embodiments, the memory chip can also include logical-to-physical mapping for the second microchip or device (e.g., see logical-to-physical mapping 612 shown in FIG. 6). And, the logical-to-physical mapping for the second microchip or device can be configured to use the bundled changed addresses as input. In some other embodiments, the second microchip or device can include logical-to-physical mapping for itself (e.g., see logical-to-physical mapping 712 shown in FIG. 7) that is configured to use the bundled changed addresses as input once the bundled changed addresses are sent in a write request to the second microchip or device.

[0032] Some embodiments described herein can include a system that includes an intermediate memory chip (e.g., see memory chip 602), a first memory chip (e.g., see first microchip or device 624), and a second memory chip (e.g., see second microchip or device 626). In such embodiments, the intermediate memory chip can be a NVRAM chip (e.g., a 3D XPoint memory chip), the second memory chip can be a flash memory chip (e.g., a NAND-type flash memory chip), and the first memory chip can be a DRAM chip. Some other embodiments described herein can include a system that includes an intermediate memory chip (e.g., see memory chip 602), a second memory chip (e.g., see second microchip or device 626), and a processor chip, such as a SoC (e.g., see first microchip or device 624). In such embodiments, the intermediate memory chip can be a NVRAM chip (e.g., a 3D XPoint memory chip) or a DRAM chip and the second memory chip can be a flash memory chip (e.g., a NAND-type flash memory chip) or a NVRAM chip.

[0033] Also, at least some aspects of the present disclosure are directed to flexible provisioning of a string of memory chips to form a memory for a processor chip or system on a chip (SoC), e.g., see FIGS. 1-5 as well as corresponding text herein. From the perspective of the processor chip or SoC wired to the memory, the string of memory chips of the memory appears no different from a single memory chip

implementation; however, with the flexible provisioning, benefits of using a string of memory chips is achieved. For example, with the flexible provisioning, benefits of using a string of memory chips with a memory hierarchy can be achieved.

[0034] The processor chip or SoC can be directly wired to a first memory chip in the string and can interact with the first memory chip without perceiving the memory chips in the string downstream of the first memory chip. In the memory, the first memory chip can be directly wired to a second memory chip and can interact with the second memory chip such that the processor chip or SoC gains the benefits of the string of the first and second memory chips without perceiving the second memory chip. And, the second memory chip can be directly wired to a third memory chip and so forth such that the processor chip or SoC gains benefits of the string of multiple memory chips without perceiving and interacting with the multiple memory chips downstream of the first memory chip. Also, in some embodiments, each chip in the string perceives and interacts with the immediate upstream chip and downstream chip in the string without perceiving chips in the string further upstream or downstream.

[0035] In some embodiments, the first memory chip in the string can be a DRAM chip. The second memory chip in the string immediately downstream of the first chip can be a NVRAM chip (e.g., a 3D XPoint memory chip). The third memory chip in the string immediately downstream of the second chip can be a flash memory chip (e.g., a NAND-type flash memory chip). Also, for example, the string can be DRAM to DRAM to NVRAM, or DRAM to NVRAM to NVRAM, or DRAM to flash memory to flash memory; although, DRAM to NVRAM to flash memory may provide a more effective solution for a string of memory chips being flexibly provisioned as multi-tier memory. Also, for the sake of understanding the flexible provisioning of a string of memory chips disclosed herein, examples will often refer to a three-chip string of memory chips; however, it is to be understood that the string of memory chips can include more than three memory chips.

[0036] Also, for the purposes of this disclosure, it is to be understood that that DRAM, NVRAM, 3D XPoint memory, and flash memory are techniques for individual memory units, and that a memory chip for any one of the memory chips described herein can include a logic circuit for command and address decoding as well as arrays of memory units of DRAM, NVRAM, 3D XPoint memory, or flash memory. For example, a DRAM chip described herein includes a logic circuit for command and address decoding as well as an array of memory units of DRAM. Also, for example, a

NVRAM chip described herein includes a logic circuit for command and address decoding as well as an array of memory units of NVRAM. And, for example, a flash memory chip described herein includes a logic circuit for command and address decoding as well as an array of memory units of flash memory.

[0037] Also, a memory chip for any one of the memory chips described herein can include a cache or buffer memory for incoming and/or outgoing data. In some embodiments, the memory units that implement the cache or buffer memory may be different from the units on the chip hosting the cache or buffer memory. For example, the memory units that implement the cache or buffer memory can be memory units of SRAM.

[0038] Each of the chips in the string of memory chips can be connected to the immediate downstream and/or upstream chip via wiring, e.g., peripheral component interconnect express (PCIe) or serial advanced technology attachment (SATA). Each of the connections between the chips in the string of memory chips can be connected sequentially with wiring and the connections can be separate from each other. Each chip in the string of memory chips can include one or more sets of pins for connecting to an upstream chip and/or downstream chip in the string. In some embodiments, each chip in the string of memory chips can include a single integrated circuit (IC) enclosed within an IC package. In such embodiments, the IC package can include the sets of pins on the boundaries of the package.

[0039] The first memory chip (e.g., DRAM chip) in the string of memory chips of the memory for the processor chip or the SoC can include a portion that can be configured, such as by the processor chip or SoC, as the cache for the second memory chip (e.g., NVRAM chip) in the string of memory chips. A portion of the memory units in the first memory chip can be used as the cache memory for the second memory chip.

[0040] The second memory chip in the string of memory chips of the memory for the processor chip or the SoC can include a portion that can be configured, such as by the fist memory chip directly and the processor chip or SoC indirectly, as the buffer for accessing the third memory chip (e.g., flash memory chip) in the string of memory chips. A portion of the memory units in the second memory chip can be used as the buffer for accessing the third memory chip. Also, the second memory chip can include a portion that can be configured, such as by the fist memory chip directly and the processor chip or SoC indirectly, as a table for logical-to-physical address mapping (logical-to-physical table) or as logical-to-physical address mapping in general. A

portion of the memory units in the second memory chip can be used for the logical-to-physical address mapping.

[0041] The third memory chip in the string of memory chips of the memory for the processor chip or the SoC can include a controller that can use the logical-to-physical address mapping in the second memory chip to manage a translation layer (e.g., flash translation layer function) of the third memory chip. The translation layer of the third memory chip can include logical-to-physical address mapping such as a copy or derivative of the logical-to-physical address mapping in the second memory chip.

[0042] Also, in some embodiments, the processor chip or SoC connected to the memory can configure the locations and the sizes of the cache in the first memory chip, the buffer and the logical-to-physical address mapping in the second memory chip, as well as cache policy parameters (e.g., write through vs write back) in the first chip by writing data into the first memory chip. And, the aforesaid configurations and settings by the processor chip or SoC can be delegated to a second data processing chip so that such tasks are removed from the processor chip or SoC. For example, the memory having the string of memory chips can have a dedicated controller separate from the processor chip or SoC configured to provide and control the aforesaid configurations and settings for the memory.

[0043] In general, with the techniques described herein to provide flexible provisioning of multi-tier memory, the flexibility to allocate a portion of memory units on certain memory chips in the string of chips as a cache or a buffer is how the memory chips (e.g., the DRAM, NVRAM, and flash memory chips) are configured to make the connectivity workable and flexible. The cache and buffer operations allow downstream memory devices of different sizes and/or different types to be connected to the upstream devices, and vice versa. In a sense, some functionalities of a memory controller are implemented in the memory chips to enable the operations of cache and buffer in the memory chips.

[0044] FIG. 1 illustrates an example memory system 100 that is configured to provide flexible provisioning of multi-tier memory, in accordance with some embodiments of the present disclosure. The memory system 100 includes a first memory chip 104 in a string of memory chips 102 of a memory. The memory system 100 also includes a second memory chip 106 in the string of memory chips 102 and a third memory chip 108 in the string of memory chips.

[0045] In FIG. 1, the first memory chip 104 is directly wired to the second memory

chip 106 (e.g., see wiring 124) and is configured to interact directly with the second memory chip. Also, the second memory chip 106 is directly wired to the third memory chip 108 (e.g., see wiring 126) and is configured to interact directly with the third memory chip.

[0046] Also, each chip in the string of memory chips 102 can include one or more sets of pins for connecting to an upstream chip and/or downstream chip in the string (e.g., see sets of pins 132, 134, 136, and 138). In some embodiments, each chip in the string of memory chips (e.g., see string of memory chips 102 or string of groups of memory chips 402 shown in FIG. 4) can include a single IC enclosed within a IC package. For example, set of pins 132 is part of first memory chip 104 and connects first memory chip 104 to second memory chip 106 via wiring 124 and set of pins 134 that is part of second memory chip 106. The wiring 124 connects the two sets of pins 132 and 134. Also, for example, set of pins 136 is part of second memory chip 106 and connects second memory chip 106 to third memory chip 108 via wiring 126 and set of pins 138 that is part of third memory chip 108. The wiring 126 connects the two sets of pins 136 and 138.

[0047] Also, as shown, the first memory chip 104 includes a cache 114 for the second memory chip 106. And, the second memory chip 106 includes a buffer 116 for the third memory chip 108 as well as logical-to-physical mapping 118 for the third memory chip 108.

[0048] The cache 114 for the second memory chip 106 can be configured by a processor chip or a memory controller chip (e.g., see processor chip 202 shown in FIG. 2 and memory controller chip 302 shown in FIG. 3). Locations and the sizes of the cache 114 in the first memory chip 104 can be configured by the processor chip or memory controller chip by corresponding data being written into the first memory chip by the processor or memory controller chip. Also, cache policy parameters of the cache 114 in the first memory chip 104 can be configured by the processor or memory controller chip by corresponding data being written into the first memory chip by the processor or the memory controller chip.

[0049] The buffer 116 for the third memory chip 108 can be configured by a processor chip or a memory controller chip (e.g., see processor chip 202 shown in FIG. 2 and memory controller chip 302 shown in FIG. 3). Locations and the sizes of the buffer 116 in the second memory chip 106 can be configured by the processor chip or memory controller chip by corresponding data being written into the second

memory chip by the processor or memory controller chip, such as indirectly via the first memory chip 104. Also, buffer policy parameters of the buffer 116 in the second memory chip 106 can be configured by the processor or memory controller chip by corresponding data being written into the second memory chip by the processor or the memory controller chip, such as indirectly via the first memory chip 104.

[0050] The logical-to-physical mapping 118 for the third memory chip 108 can be configured by a processor chip or a memory controller chip (e.g., see processor chip 202 shown in FIG. 2 and memory controller chip 302 shown in FIG. 3). Locations and the sizes of the logical-to-physical mapping 118 in the second memory chip 106 can be configured by the processor chip or memory controller chip by corresponding data being written into the second memory chip by the processor or memory controller chip, such as indirectly via the first memory chip 104. Also, buffer policy parameters of the logical-to-physical mapping 118 in the second memory chip 106 can be configured by the processor or memory controller chip by corresponding data being written into the second memory chip by the processor or the memory controller chip, such as indirectly via the first memory chip 104.

[0051] In some embodiments, the third memory chip 108 can have a lowest memory bandwidth of the chips in the string. In some embodiments, the first memory chip 104 can have a highest memory bandwidth of the chips in the string. In such embodiments, the second memory chip 106 can have a next highest memory bandwidth of the chips in the string, such that the first memory chip 104 has a highest memory bandwidth of the chips in the string and the third memory chip 108 has a lowest memory bandwidth of the chips in the string.

[0052] In some embodiments, the first memory chip 104 is or includes a DRAM chip. In some embodiments, the first memory chip 104 is or includes a NVRAM chip. In some embodiments, the second memory chip 106 is or includes a DRAM chip. In some embodiments, the second memory chip 106 is or includes a NVRAM chip. In some embodiments, the third memory chip 108 is or includes a DRAM chip. In some embodiments, the third memory chip 108 is or includes a NVRAM chip. And, in some embodiments, the third memory chip 108 is or includes a flash memory chip.

[0053] In embodiments having one or more DRAM chips, a DRAM chip can include a logic circuit for command and address decoding as well as arrays of memory units of DRAM. Also, a DRAM chip described herein can include a cache or buffer memory for incoming and/or outgoing data. In some embodiments, the memory units that

implement the cache or buffer memory can be different from the DRAM units on the chip hosting the cache or buffer memory. For example, the memory units that implement the cache or buffer memory on the DRAM chip can be memory units of SRAM.

[0054] In embodiments having one or more NVRAM chips, a NVRAM chip can include a logic circuit for command and address decoding as well as arrays of memory units of NVRAM such as units of 3D XPoint memory. Also, a NVRAM chip described herein can include a cache or buffer memory for incoming and/or outgoing data. In some embodiments, the memory units that implement the cache or buffer memory can be different from the NVRAM units on the chip hosting the cache or buffer memory. For example, the memory units that implement the cache or buffer memory on the NVRAM chip can be memory units of SRAM.

[0055] In some embodiments, NVRAM chips can include a cross-point array of non-volatile memory cells. A cross-point array of non-volatile memory can perform bit storage based on a change of bulk resistance, in conjunction with a stackable cross-gridded data access array. Additionally, in contrast to many flash-based memories, cross-point non-volatile memory can perform a write in-place operation, where a non-volatile memory cell can be programmed without the non-volatile memory cell being previously erased.

[0056] As mentioned herein, NVRAM chips can be or include cross point storage and memory devices (e.g., 3D XPoint memory). A cross point memory device uses transistor-less memory elements, each of which has a memory cell and a selector that are stacked together as a column. Memory element columns are connected via two perpendicular lays of wires, where one lay is above the memory element columns and the other lay below the memory element columns. Each memory element can be individually selected at a cross point of one wire on each of the two layers. Cross point memory devices are fast and non-volatile and can be used as a unified memory pool for processing and storage.

[0057] In embodiments having one or more flash memory chips, a flash memory chip can include a logic circuit for command and address decoding as well as arrays of memory units of flash memory such as units of NAND-type flash memory. Also, a flash memory chip described herein can include a cache or buffer memory for incoming and/or outgoing data. In some embodiments, the memory units that implement the cache or buffer memory can be different from the flash memory units

on the chip hosting the cache or buffer memory. For example, the memory units that implement the cache or buffer memory on the flash memory chip can be memory units of SRAM.

[0058] Also, for example, an embodiment of the string of memory chips can include DRAM to DRAM to NVRAM, or DRAM to NVRAM to NVRAM, or DRAM to flash memory to flash memory; however, DRAM to NVRAM to flash memory may provide a more effective solution for a string of memory chips being flexibly provisioned as multi-tier memory.

[0059] Also, for the purposes of this disclosure, it is to be understood that that DRAM, NVRAM, 3D XPoint memory, and flash memory are techniques for individual memory units, and that a memory chip for any one of the memory chips described herein can include a logic circuit for command and address decoding as well as arrays of memory units of DRAM, NVRAM, 3D XPoint memory, or flash memory. For example, a DRAM chip described herein includes a logic circuit for command and address decoding as well as an array of memory units of DRAM. For example, a NVRAM chip described herein includes a logic circuit for command and address decoding as well as an array of memory units of NVRAM. For example, a flash memory chip described herein includes a logic circuit for command and address decoding as well as an array of memory units of flash memory.

[0060] Also, a memory chip for any one of the memory chips described herein can include a cache or buffer memory for incoming and/or outgoing data. In some embodiments, the memory units that implement the cache or buffer memory may be different from the units on the chip hosting the cache or buffer memory. For example, the memory units that implement the cache or buffer memory can be memory units of SRAM.

[0061] FIG. 2 illustrates the example memory system 100 and processor chip 202 configured to provide flexible provisioning of multi-tier memory, in accordance with some embodiments of the present disclosure. In FIG. 2, the processor chip 202 is directly wired (e.g., see wiring 204) to the first memory chip 104 and is configured to interact directly with the first memory chip.

[0062] In some embodiments, the processor chip 202 includes or is a SoC. A SoC describe herein can be or include an integrated circuit or chip that integrates any two or more components of a computing device. The two or more components can include at least one or more of a central processing unit (CPU), graphics processing unit

(GPU), memory, input/output ports, and secondary storage. For example, an SoC described herein can also include a CPU, a GPU, graphics and memory interfaces, hard-disk, USB connectivity, random-access memory, read-only memory, secondary storage, or any combination thereof on a single circuit die. Also, where the processor chip 202 is a SoC, the SoC includes at least a CPU and/or a GPU.

[0063] For an SoC described herein, the two or more components can be embedded on a single substrate or microchip (chip). In general, a SoC is different from a conventional motherboard-based architecture in that the SoC integrates all of its components into a single integrated circuit; whereas a motherboard houses and connects detachable or replaceable components. Because the two or more components are integrated on a single substrate or chip, SoCs consume less power and take up much less area than multi-chip designs with equivalent functionality. Thus, in some embodiments, the memory systems described herein can be connected with or be a part of SoCs in mobile computing devices (such as in smartphones), embedded systems, and the Internet of Things devices.

[0064] The processor chip 202 can be configured to configure the cache 114 for the second memory chip 106. The processor chip 202 can also be configured to configure locations and the sizes of the cache 114 by writing corresponding data into the first memory chip 104. The processor chip 202 can also be configured to configure cache policy parameters by writing corresponding data into the first memory chip 104.

[0065] Also, the processor chip 202 can configured to configure the buffer 116 for the third memory chip 108 and/or the logical-to-physical mapping 118 for the third memory chip. The processor chip 202 can also be configured to configure locations and sizes of the buffer 116 by writing corresponding data into the first memory chip 104. The processor chip 202 can also be configured to configure locations and the sizes of the logical-to-physical mapping 118 by writing corresponding data into the first memory chip 104.

[0066] FIG. 3 illustrates the example memory system 100 and memory controller chip 302 configured to provide flexible provisioning of multi-tier memory, in accordance with some embodiments of the present disclosure. In FIG. 3, the memory controller chip 302 is directly wired (e.g., see wiring 304) to the first memory chip 104 and is configured to interact directly with the first memory chip.

[0067] In some embodiments, the memory controller chip 302 includes or is a SoC.

Such a SoC can be or include an integrated circuit or chip that integrates any two or more components of a computing device. The two or more components can include at least one or more of a separate memory, input/output ports, and separate secondary storage. For example, the SoC can include memory interfaces, hard-disk, USB connectivity, random-access memory, read-only memory, secondary storage, or any combination thereof on a single circuit die. Also, where the memory controller chip 302 is a SoC, the SoC includes at least a data processing unit.

[0068] The memory controller chip 302 can be configured to configure the cache 114 for the second memory chip 106. The memory controller chip 302 can also be configured to configure locations and the sizes of the cache 114 by writing corresponding data into the first memory chip 104. The memory controller chip 302 can also be configured to configure cache policy parameters by writing corresponding data into the first memory chip 104.

[0069] Also, the memory controller chip 302 can configured to configure the buffer 116 for the third memory chip 108 and/or the logical-to-physical mapping 118 for the third memory chip. The memory controller chip 302 can also be configured to configure locations and sizes of the buffer 116 by writing corresponding data into the first memory chip 104. The memory controller chip 302 can also be configured to configure locations and the sizes of the logical-to-physical mapping 118 by writing corresponding data into the first memory chip 104.

[0070] FIG. 4 illustrates an example memory system 400 configured to provide flexible provisioning of multi-tier memory with tiers that each include multiple memory chips, in accordance with some embodiments of the present disclosure. The memory system 400 includes a string of groups of memory chips 402. The string of groups of memory chips 402 includes a first group of memory chips including a first type of memory chips (e.g., see memory chips 404a and 404b which are the same type of chips). The string of groups of memory chips 402 includes a second group of memory chips including the first type of memory chips or a second type of memory chips (e.g., see memory chips 406a and 406b which are the same type of chips). The string of groups of memory chips 402 also includes a third group of memory chips including a first type of memory chips, a second type of memory chips, or a third type of memory chips (e.g., see memory chips 408a and 408b which are the same type of chips). The first type of memory chips can be or include DRAM chips. The second type of memory chips can be or include NVRAM chips. The third type of memory chips can

be or include flash memory chips.

[0071] Also, as shown in FIG. 4, the chips in the first group of memory chips are directly wired to the chips in the second group of memory chips via wiring 424 and are configured to interact directly with one or more of the chips in the second group of memory chips. Also, as shown in FIG. 4, the chips in the second group of memory chips are directly wired to the chips in the third group of memory chips via wiring 426 and are configured to interact directly with one or more of the chips in the third group of memory chips.

[0072] Also, as shown in FIG. 4, each chip in the first group of memory chips includes a cache (e.g., see cache 414) for the second group of memory chips. And, each chip in the second group of memory chips includes a buffer 416 for the third group of memory chips as well as logical-to-physical mapping 418 for the third group of memory chips.

[0073] In some embodiments, each chip in the third group of memory chips (e.g., see memory chips 408a and 408b) can have a lowest memory bandwidth relative to the other chips in the string of groups of memory chips 402. In some embodiments, each chip in the first group of memory chips (e.g., see memory chips 404a and 404b) can have a highest memory bandwidth relative to the other chips in the string of groups of memory chips 402. In such embodiments, each chip in the second group of memory chips (e.g., see memory chips 406a and 406b) can have a next highest memory bandwidth relative to other chips in the string of groups of memory chips 402, such that each chip in the first group of memory chips has a highest memory bandwidth and each chip in the third group of memory chips has a lowest memory bandwidth.

[0074] In some embodiments, the first group of memory chips (e.g., see memory chips 404a and 404b) can include DRAM chips or NVRAM chips. In some embodiments, the second group of memory chips (e.g., see memory chips 406a and 406b) can include DRAM chips or NVRAM chips. In some embodiments, the third group of memory chips (e.g., see memory chips 408a and 408b) can include DRAM chips, NVRAM chips, or flash memory chips.

[0075] As shown in FIGS. 1-4, the present disclosure is directed to flexible provisioning of a string of memory chips (e.g., see string of memory chips 102 shown in FIGS. 1-3 or string of groups of memory chips 402 shown in FIG. 4). And, the flexible provisioning of the string of memory chips forms a memory (e.g., see memory system 100 shown in FIG. 2 or memory system 400 shown in FIG. 4).

[0076] A memory system disclosed herein, such as memory system 100 or 400, can be its own apparatus or within its own packaging.

[0077] In some embodiments, a memory system disclosed herein, such as memory system 100 or 400, can be combined with and for a processor chip or SoC (e.g., see FIG. 2). When combined with and for a processor chip or SoC, the memory system and the processor chip or SoC can be a part of a single apparatus and/or combined into a single packaging.

[0078] Also, in some embodiments, a memory system disclosed herein, such as memory system 100 or 400, can be combined with a memory controller chip (e.g., see FIG. 3). When combined with a memory controller chip, the memory system and the memory controller chip can be a part of a single apparatus and/or combined into a single packaging. Alternatively, each chip in the string of chips, or at least the first memory chip and the second memory chip, can include a respective memory controller providing similar functionality to the memory controller chip shown in FIG. 3.

[0079] From the perspective of the processor chip or SoC wired to the memory (e.g., see processor chip 202 shown in FIG. 2) or the memory controller chip (e.g., see memory controller chip 302 shown in FIG. 3), the string of memory chips of the memory appears no different from a single memory chip implementation; however, with the flexible provisioning, benefits of using a string of memory chips is achieved. In such embodiments, the processor chip or SoC--or the memory controller chip--can be directly wired (e.g., see wiring 204 shown in FIG. 2 or wiring 304 shown in FIG. 3) to a first memory chip (e.g., see first memory chip 104) in the string of memory chips 102 and can interact with the first memory chip without perceiving the memory chips in the string downstream of the first memory chip (e.g., see second memory chip 106 and third memory chip 108 which are downstream of the first memory chip 104).

[0080] In the memory (e.g., see memory system 100 or 400), the first memory chip (e.g., see first memory chip 104 or one of memory chips 404a or 404b) can be directly wired to a second memory chip (e.g., see second memory chip 106 or one of memory chips 406a or 406b) and can interact with the second memory chip such that the processor chip, SoC, or memory controller chip (e.g., see processor chip 202 and memory controller chip 302) gains the benefits of the string of the first and second memory chips without perceiving the second memory chip. And, the second memory chip (e.g., see first memory chip 104 or one of memory chips 404a or 404b) can be directly wired to a third memory chip (e.g., see third memory chip 108 or one of

memory chips 408a or 408b) and so forth such that the processor chip, SoC, or memory controller chip gains benefits of the string of multiple memory chips (e.g., see string of memory chips 102 or string of groups of memory chips 402) without perceiving and interacting with the multiple memory chips downstream of the first memory chip. Also, in some embodiments, each chip in the string perceives and interacts with an immediate upstream chip and downstream chip in the string without perceiving chips in the string further upstream or downstream.

[0081] As mentioned, with the flexible provisioning, benefits of using a string of memory chips with a memory hierarchy can be achieved. Thus, for example, in some embodiments, the first memory chip (e.g., see first memory chip 104) in the string can be a chip with the highest memory bandwidth in the memory. The second memory chip (e.g., see second memory chip 106) in the string immediately downstream of the first chip can be a chip with next highest memory bandwidth of the memory (which may have other benefits such as being cheaper to manufacture than the first chip or be more reliable or persistent at storing data than the first chip). The third memory chip (e.g., see third memory chip 108) in the string immediately downstream of the second chip (or the final downstream chip in the string where the string has more than three memory chips) can have the lowest memory bandwidth. The third memory chip in such examples (or the final downstream chip in other examples with more than three memory chips) can be the most cost-effective chip or most reliable or persistent chip for storing data.

[0082] In some embodiments, the first memory chip in the string can be a DRAM chip. In such embodiments, the second memory chip in the string immediately downstream of the first chip can be a NVRAM chip (e.g., a 3D XPoint memory chip). And, in such embodiments, the third memory chip in the string immediately downstream of the second chip can be a flash memory chip (e.g., a NAND-type flash memory chip).

[0083] As mentioned, for the sake of understanding the flexible provisioning of a string of memory chips disclosed here, examples often refer to a three-chip string of memory chips (e.g., see string of memory chips 102 shown in FIGS. 1-3 and string of groups of memory chips 402 shown in FIG. 4); however, it is to be understood that the string of memory chips can include more than three memory chips or more than three groups of chips where each of the groups is a tier of chips.

[0084] As mentioned, some embodiments of string of memory chips can include a

DRAM memory chip that is the first chip in the string, a NVRAM chip that is the second chip in the string, and a flash memory chip (e.g., NAND-type flash memory chip) that is the third chip in the string and can be used as the bulk memory chip in the string. In such embodiments and in other embodiments with other arrangements of memory chip types, each of the chips in the string of memory chips are connected to the immediate downstream and/or upstream chip via wiring (e.g., PCIe or SATA). Each of the connections between the chips in the string of memory chips can be connected sequentially with wiring and the connections can be separate from each other (e.g., see wiring 124 and 126 as well as wiring 424 and 426). Also, each chip in the string of memory chips can include one or more sets of pins for connecting to an upstream chip and/or downstream chip in the string (e.g., see sets of pins 132, 134, 136, and 138 depicted in FIG. 1). In some embodiments, each chip in the string of memory chips (e.g., see string of memory chips 102 or string of groups of memory chips 402) can include a single IC enclosed within a IC package. In such embodiments, the IC package can include the sets of pins on the boundaries of the package (such as sets of pins 132, 134, 136, and 138).

[0085] The first memory chip (e.g., DRAM chip) in the string of memory chips of the memory for the processor chip or the SoC can include a portion that can be configured, such as by the processor chip or SoC, as the cache for the second memory chip (e.g., NVRAM chip) in the string (e.g., see cache 114 for the second memory chip). A portion of the memory units in the first memory chip can be used as the cache memory for the second memory chip.

[0086] The second memory chip in the string of memory chips of the memory for the processor chip or the SoC can include a portion that can be configured, such as by the fist memory chip directly and the processor chip or SoC indirectly, as the buffer for accessing the third memory chip (e.g., flash memory chip) in the string (e.g., see buffer for the third memory chip 116). A portion of the memory units in the second memory chip can be used as the buffer for accessing the third memory chip. Also, the second memory chip can include a portion that can be configured, such as by the fist memory chip directly and the processor chip or SoC indirectly, as a table for logical-to-physical address mapping (logical-to-physical table) or as logical-to-physical address mapping in general (e.g., see logical-to-physical mapping 118). A portion of the memory units in the second memory chip can be used for the logical-to-physical address mapping.

[0087] The third memory chip in the string of memory chips of the memory for the processor chip or the SoC can include a controller (e.g., see controller 128) that can use the logical-to-physical address mapping in the second memory chip to manage a translation layer (e.g., flash translation layer function) of the third memory chip (e.g., see translation layer 130). The translation layer of the third memory chip can include logical-to-physical address mapping such as a copy or derivative of the logical-to-physical address mapping in the second memory chip.

[0088] Also, in some embodiments, the processor chip or SoC connected to the memory (e.g., see processor chip 202) can configure the locations and the sizes of the cache in the first memory chip, the buffer and the logical-to-physical address mapping in the second memory chip, as well as cache policy parameters (e.g., write through vs write back) in the first chip by writing data into the first memory chip (e.g., see first memory chip 104). And, the aforesaid configurations and settings by the processor chip or SoC can be delegated to a second data processing chip so that such tasks are removed from the processor chip or SoC (e.g., see memory controller chip 302 shown in FIG. 3). For example, the memory having the string of memory chips can have a dedicated controller separate from the processor chip or SoC configured to provide and control the aforesaid configurations and settings for the memory (e.g., see memory controller chip 302).

[0089] For the purposes of this disclosure it is to be understood that a memory chip in the string of memory chips can be replaced by a group of similar memory chips, such that the string includes a string of groups of similar chips (e.g., see string of groups of memory chips 402 shown in FIG. 4). In such examples, each group of similar chips is a node in the string. Also, in some embodiments, the nodes of the string of memory chips can be made up of a combination of single chip nodes and multiple chip nodes (not depicted in the drawings). For example, in the string of memory chips, the first memory chip (e.g., DRAM chip) can be replaced by a group of similar memory chips (e.g., a group of DRAM chips), the second memory chip (e.g., NVRAM chip) can be replaced by a group of similar memory chips (e.g., a group of NVRAM chips), the third memory chip (e.g., flash memory chip) can be replaced by a group of similar memory chips (e.g., a group of flash memory chips), or some combination thereof.

[0090] FIG. 5 illustrates example parts of an example computing device 500, in accordance with some embodiments of the present disclosure. The computing device 500 can be communicatively coupled to other computing devices via the

computer network 502 as shown in FIG. 5. The computing device 500 includes at least a bus 504, a processor 506 (such as a CPU and/or the processor chip 202 shown in FIG. 2), a main memory 508, a network interface 510, and a data storage system 512. The bus 504 communicatively couples the processor 506, the main memory 508, the network interface 510, and the data storage system 512. The computing device 500 includes a computer system that includes at least processor 506, main memory 508 (e.g., read-only memory (ROM), flash memory, DRAM such as synchronous DRAM (SDRAM) or Rambus DRAM (RDRAM), NVRAM, SRAM, etc.), and data storage system 512, which communicate with each other via bus 504 (which can include multiple buses and wirings).

[0091] The main memory 508 can include the memory system 100 depicted in FIG 1. Also, the main memory 508 can include the memory system 400 depicted in FIG 4. In some embodiments, the data storage system 512 can include the memory system 100 depicted in FIG 1. And, the data storage system 512 can include the memory system 400 depicted in FIG 4.

[0092] Processor 506 can represent one or more general-purpose processing devices such as a microprocessor, a central processing unit, or the like. The processor 506 can be or include the processor 202 depicted in FIG. 2. The processor 506 can be a complex instruction set computing (CISC) microprocessor, reduced instruction set computing (RISC) microprocessor, very long instruction word (VLIW) microprocessor, or a processor implementing other instruction sets, or processors implementing a combination of instruction sets. Processor 506 can also be one or more special-purpose processing devices such as an application specific integrated circuit (ASIC), a field programmable gate array (FPGA), a digital signal processor (DSP), network processor, a processor in memory (PIM), or the like. Processor 506 can be configured to execute instructions for performing the operations and steps discussed herein. Processor 506 can further include a network interface device such as network interface 510 to communicate over one or more communications network such as network 502.

[0093] The data storage system 512 can include a machine-readable storage medium (also known as a computer-readable medium) on which is stored one or more sets of instructions or software embodying any one or more of the methodologies or functions described herein. The instructions can also reside, completely or at least partially, within the main memory 508 and/or within the processor 506 during execution

thereof by the computer system, the main memory 508 and the processor 506 also constituting machine-readable storage media.

[0094] While the memory, processor, and data storage parts are shown in the example embodiment to each be a single part, each part should be taken to include a single part or multiple parts that can store the instructions and perform their respective operations. The term “machine-readable storage medium” shall also be taken to include any medium that is capable of storing or encoding a set of instructions for execution by the machine and that cause the machine to perform any one or more of the methodologies of the present disclosure. The term “machine-readable storage medium” shall accordingly be taken to include, but not be limited to, solid-state memories, optical media, and magnetic media.

[0095] Referring back to the memory chip having the data mover, FIG. 6 illustrates an example system 600 that includes a memory chip 602 having an integrated data mover 608, in accordance with some embodiments of the present disclosure. In FIG. 6, the memory chip 602 includes a first set of pins 604 configured to allow the memory chip 602 to be coupled to a first microchip or device 624 via first wiring 634. Also shown, the first microchip or device 624 has a set of pins 605 configured to allow the first microchip or device 624 to be coupled to the memory chip 602 via first wiring 634. The memory chip 602 also includes a second set of pins 606 configured to allow the memory chip 602 to be coupled to a second microchip or device 626 via second wiring 636 that is separate from the first wiring 634. Also shown, the second microchip or device 626 has a set of pins 607 configured to allow the second microchip or device 626 to be coupled to the memory chip 602 via second wiring 636. The memory chip 602 also includes a data mover 608 configured to facilitate access to the second microchip or device 626, via the second set of pins 606, to read data from the second microchip or device 626 and write data to the second microchip or device 626.

[0096] In general, the data mover 608 can combine data stored in the memory chip 602 on its way to the second microchip or device 626 using various strategies. This can improve write performance and endurance of the second microchip or device 626. For example, sequential or block access on memory chips is orders of magnitude faster than random access on memory chips. In some embodiments, data stored in a portion of memory 610 of the memory chip 602 is accessible by or through the first microchip or device 624 via the first set of pins 604. And, in such embodiments, the data mover 608 is configured to combine the data stored in the portion of memory 610

of the memory chip 602 by moving the data in blocks to the second microchip or device 626. Thus, write performance and endurance of the second microchip or device 626 is improved.

[0097] Also, in such embodiments and others, the blocks can be at a granularity that is coarser than the data stored in the portion of memory 610 of the memory chip 602. The blocks being at a granularity that is coarser than the pre-blocked data in the memory chip 602, such as the data to be accessed by first microchip or device, can reduce the frequency of data writes to the second microchip or device.

[0098] In some embodiments, such as shown in FIG. 6, data stored in a portion of memory 610 of the memory chip 602 can be accessible by or through the first microchip or device 624 via the first set of pins 604. Also, when the data stored in the portion of memory 610 of the memory chip 602 is accessible through the first microchip or device 624 it is being accessed by another memory chip or device or a processor chip or device (which not depicted in FIG. 6 but is shown in FIGS. 2 and 3 to some extent wherein the second memory chip 106 is acting as the memory chip 602). And, the first microchip or device 624 can read data from the memory chip 602 as well as write data to the memory chip 602.

[0099] In some embodiments, such as shown in FIG. 6, data stored in a portion of memory 630 of the second microchip or device 626 can be accessible by or through the memory chip 602 via the second set of pins 606. Also, when the data stored in the portion of memory 630 of the second microchip or device 626 is accessible through the memory chip 602 it is being accessed by another memory chip or device or a processor chip or device (such as the first microchip or device 624). And, the memory chip 602 can read data from the second microchip or device 626 as well as write data to the second microchip or device 626.

[0100] In some embodiments, such as shown in FIG. 6, the data mover 608 is configured to buffer movement of changes to the data stored in the portion of memory 610 of the memory chip 602. Also, the data mover 608 is configured to send write requests to the second microchip or device 626 in a suitable size due to the buffering by the data mover 608.

[0101] When the write to the second microchip or device 626 is in the suitable size due to the buffering by the data mover 608, the second microchip or device can erase a block and program the block in the second microchip or device according to the write without further processing or minimal processing in the second microchip or

device. This is one example way that the data mover 608 integrated in the memory chip 602 can improve write performance and endurance of the second microchip or device. Also, with the buffering by the data mover 608 and when frequent and/or random changes are made to the data in a portion of the memory chip 602 (such as the data in the portion of memory 610 accessible by the first microchip or device), the second microchip or device 626 does not have to be frequently erased and reprogrammed in a corresponding way as the changes occurring in the memory chip 602.

[0102] The buffering by the data mover 608 is even more beneficial when the second microchip or device 626 is a flash memory chip because the buffering can remove or at least limit the effects of write amplification that occurs in flash memory. Write amplification can be reduced or even eliminated by the buffering because, with the buffering, a write request sent by the memory chip 602 can be modified to a suitable size or granularity expected by the receiving flash memory chip. Thus, the flash memory chip can erase a block and program the block according to the write request without possible duplication of the write and thus avoid further processing in the second microchip or device 626.

[0103] Also, the data mover 608 can be configured to bundle changed addresses in the changes to the data stored in the portion of memory 610 of the memory chip 602. And, the data mover 608 can be configured to write bundled changed addresses into another portion of the memory chip 602 to be moved to the second microchip or device 626 via a write request to the second microchip or device 626. The bundling by the data mover 608 can improve the buffering by the data mover and the sending of write requests according to the buffering since the bundling of changed addresses can be controlled by the data mover to correspond to a suitable size or granularity expected by the receiving second microchip or device 626.

[0104] In some embodiments, such as shown in FIG. 6, the memory chip 602 includes logical-to-physical mapping 612 for the second microchip or device 626 that is configured to use the bundled changed addresses as input.

[0105] FIG. 7 illustrates an example system 700 that includes the memory chip 602 shown in FIG. 6 as well as a microchip or device that receives data transferred from the data mover and has logical-to-physical mapping for mapping the received data (e.g., see logical-to-physical mapping 712), in accordance with some

embodiments of the present disclosure. The system 700 is similar to the system 600, except in system 700, the logical-to-physical mapping 712 for the second microchip or device 626 is provided in the second microchip or device 626. This is instead of providing the logical-to-physical mapping for the second microchip or device on the memory chip 602 (e.g., see logical-to-physical mapping 612 depicted in FIG. 6).

[0106] With embodiments similar to system 700, the second microchip or device 626 includes logical-to-physical mapping 712 for itself. And, the logical-to-physical mapping 712 is configured to use the bundled changed addresses as input once the bundled changed addresses are sent from the data mover 608 of the memory chip 602 in a write request to the second microchip or device 626. To put it another way, once the bundled changed addresses are received by the second microchip or device 626 from the data mover 608 of the memory chip 602 in a write request to the second microchip or device 626, the logical-to-physical mapping 712 is configured to use the bundled changed addresses as input.

[0107] FIG. 8 illustrates an example system 800 that includes the memory chip 602 shown in FIG. 6 which is also shown having encryption and authentication circuitry (e.g., see encryption engine 802 and gatekeeper 804), in accordance with some embodiments of the present disclosure. As shown in FIG. 8, in some embodiments, the memory chip 602 includes an encryption engine 802 configured to secure data to be moved to the second microchip or device 626 for long term storage. Also, as shown, in such embodiments, the memory chip 602 includes a gatekeeper 804, configured to provide an authentication process for access of the secured data stored in the second microchip or device 626. Also, the gatekeeper 804 can be configured to decrypt the secured data according to the authentication process. In some other embodiments, the encryption engine can be hosted on another chip or device instead of being hosted on the memory chip 602.

[0108] FIGS. 7 and 8 also depict other components that are illustrated in FIG. 6. For example, FIGS. 7 and 8 depict the memory chip 602 that includes a first set of pins 604 configured to allow the memory chip 602 to be coupled to a first microchip or device 624 via first wiring 634. Also shown in FIGS. 7 and 8, the first microchip or device 624 has a set of pins 605 configured to allow the first microchip or device 624 to be coupled to the memory chip 602 via first wiring 634. Also shown in FIGS. 7 and 8, the memory chip 602 also includes a second set of pins 606 configured to

allow the memory chip 602 to be coupled to a second microchip or device 626 via second wiring 636 that is separate from the first wiring 634. Also shown, the second microchip or device 626 has a set of pins 607 configured to allow the second microchip or device 626 to be coupled to the memory chip 602 via second wiring 636. The memory chip 602 also includes a data mover 608 configured to facilitate access to the second microchip or device 626, via the second set of pins 606, to read data from the second microchip or device 626 and write data to the second microchip or device 626.

[0109] With respect to the systems shown in FIGS. 6-8, the memory chip 602 can be a NVRAM chip, and with such embodiments, the memory chip 602 includes a plurality of NVRAM cells. The plurality of NVRAM cells in such examples can be or include a plurality of 3D XPoint memory cells. Alternatively, the memory chip 602 can be a dynamic DRAM chip, and with such embodiments, the memory chip 602 includes a plurality of DRAM cells. Also, the memory chip 602 can be a flash memory chip, and with such embodiments, the memory chip 602 includes a plurality of flash memory cells. The plurality of flash memory cells in such embodiments can be or include a plurality of NAND-type flash memory cells.

[0110] The second microchip or device 626 can be another memory chip or a memory device. In some embodiments, for example, the second microchip or device 626 is a DRAM chip. In some embodiments, for example, the second microchip or device 626 is a NVRAM chip. In some embodiments, for example, the second microchip or device 626 is a flash memory chip (e.g., a NAND-type flash memory chip).

[0111] Some embodiments can include a system having an intermediate memory chip (e.g., see memory chip 602 shown in FIG. 6), a first memory chip (e.g., see first microchip or device 624), and a second memory chip (e.g., see second microchip or device 626). And, in such embodiments, the intermediate memory chip includes a first set of pins (e.g., see first set of pins 604) configured to allow the intermediate memory chip to be coupled to the first memory chip via first wiring (e.g., see wiring 634). The intermediate memory chip can also include a second set of pins (e.g., see second set of pins 606) configured to allow the intermediate memory chip to be coupled to the second memory chip via second wiring (e.g., see wiring 636) that is separate from the first wiring. And, the intermediate memory chip can include a data mover (e.g., see data mover 608) configured to facilitate access to the second

memory chip, via the second set of pins, to read data from the second memory chip and write data to the second memory chip.

[0112] In such embodiments, the intermediate memory chip can be a NVRAM chip, wherein the NVRAM chip includes a plurality of NVRAM cells. Alternatively, the intermediate memory chip can be a DRAM chip, wherein the DRAM chip includes a plurality of DRAM cells. The plurality of NVRAM cells in such examples can be or include a plurality of 3D XPoint memory cells.

[0113] Also, in such embodiments, the second memory chip can be a flash memory chip, wherein the flash memory chip includes a plurality of flash memory cells. The plurality of flash memory cells in such embodiments can be or include a plurality of NAND-type flash memory cells. Alternatively, the second memory chip can be a NVRAM chip, wherein the NVRAM chip includes a plurality of NVRAM cells. The plurality of NVRAM cells in such examples can be or include a plurality of 3D XPoint memory cells. Also, the second memory chip can be a DRAM chip, wherein the DRAM chip includes a plurality of DRAM cells.

[0114] Also, in such embodiments, the first memory chip can be a DRAM chip, wherein the DRAM chip includes a plurality of DRAM cells. Alternatively, the first memory chip can be a NVRAM chip, wherein the NVRAM chip includes a plurality of NVRAM cells. The plurality of NVRAM cells in such examples can be or include a plurality of 3D XPoint memory cells.

[0115] Some embodiments can include a system having an intermediate memory chip (e.g., see memory chip 602 shown in FIG. 6), a second memory chip (e.g., see second microchip or device 626), and a processor chip, such as a SoC (e.g., see first microchip or device 624). In such embodiments, the intermediate memory chip includes a first set of pins (e.g., see first set of pins 604) configured to allow the intermediate memory chip to be coupled to the processor chip via first wiring (e.g., see wiring 634). Also, the intermediate memory chip can include a second set of pins (e.g., see second set of pins 606) configured to allow the intermediate memory chip to be coupled to the second memory chip via second wiring (e.g., see wiring 636) that is separate from the first wiring. And, the intermediate memory chip can include a data mover (e.g., see data mover 608) configured to facilitate access to the second memory chip, via the second set of pins and the second wiring, to read data from the second memory chip and write data to the second memory chip.

[0116] In some embodiments, one or more instances of the memory chip 602

(such as the different instances of the memory chip 602 shown in FIGS. 6-8) can be used as one or more memory chips in flexible provisioning of the string of memory chips 102 shown in FIGS. 1-3. Also, one or more instances of the memory chip 602 can be used as one or more memory chips in flexible provisioning of the string of memory chips 402 shown in FIGS. 4.

[0117] Also, the system having the intermediate memory chip (e.g., see memory chip 602 shown in FIG. 6), the first memory chip (e.g., see first microchip or device 624), and the second memory chip (e.g., see second microchip or device 626) can be included in the string of memory chips 102 shown in FIGS. 1-3. For example, the memory chip 602 can be or include the second memory chip 106, the first microchip or device 624 can be or include the first memory chip 104 and the second microchip or device 626 can be or include the third memory chip 108.

[0118] In some embodiments, the memory chip 602 can be or include the first memory chip 104 in the string of memory chips 102. In such embodiments, for example, the first microchip or device 624 can be or include the processor chip 202 shown in FIG. 2 or the memory controller chip 302 shown in FIG. 3. Also, in such embodiments, the second microchip or device 626 can be or include the second memory chip 106 in the string of memory chips 102.

[0119] In some embodiments, any one or more of the systems 600, 700, and 800 can be included in the main memory 508 and/or the data storage system 512 shown in FIG. 5. In such embodiments and others, the first microchip or device 624 can be, be a part of, or include the processor 506.

[0120] Like the memory chip 602, any other of the memory chips described herein (e.g., see memory chips 104, 106, and 108) can include an encryption engine and/or an authentication gatekeeper for securing data moved to or through the memory chip (e.g., see encryption engine 802 and gatekeeper 804 shown in FIG. 8). Also, like the memory chip 602, any other of the memory chips described herein (e.g., see memory chips 104, 106, and 108) can include a data mover, such as the data mover 608.

[0121] In the foregoing specification, embodiments of the disclosure have been described with reference to specific example embodiments thereof. It will be evident that various modifications can be made thereto without departing from the broader spirit and scope of embodiments of the disclosure as set forth in the following claims. The specification and drawings are, accordingly, to be regarded in

an illustrative sense rather than a restrictive sense.

CLAIMS

What is claimed is:

1. A memory chip, comprising:
a first set of pins configured to allow the memory chip to be coupled to a first microchip or device via first wiring;
a second set of pins configured to allow the memory chip to be coupled to a second microchip or device via second wiring that is separate from the first wiring; and
a data mover configured to facilitate access to the second microchip or device, via the second set of pins, to read data from the second microchip or device and write data to the second microchip or device.

2. The memory chip of claim 1, wherein data stored in a portion of the memory chip is accessible by or through the first microchip or device via the first set of pins.

3. The memory chip of claim 2, wherein the data mover is configured to combine the data stored in the portion of the memory chip by moving the data in blocks to the second microchip or device.

4. The memory chip of claim 3, wherein the blocks are at a granularity that is coarser than the data stored in the portion of the memory chip.

5. The memory chip of claim 4, wherein the data mover is configured to:
buffer movement of changes to the data stored in the portion of the memory chip; and
send write requests to the second microchip or device in a suitable size due to the buffering by the data mover.

6. The memory chip of claim 5, wherein the data mover is configured to:
bundle changed addresses in the changes to the data stored in the portion of the memory chip; and

write bundled changed addresses into another portion of the memory chip to be moved to the second microchip or device via a write request to the second microchip or device.

7. The memory chip of claim 5, comprising logical-to-physical mapping for the second microchip or device that is configured to use the bundled changed addresses as input.

8. The memory chip of claim 5, wherein the second microchip or device comprises logical-to-physical mapping for itself that is configured to use the bundled changed addresses as input once the bundled changed addresses are sent in a write request to the second microchip or device.

9. The memory chip of claim 1, wherein the memory chip is a non-volatile random-access memory (NVRAM) chip, and wherein the memory chip comprises a plurality of NVRAM cells.

10. The memory chip of claim 9, wherein the plurality of NVRAM cells is a plurality of 3D XPoint memory cells.

11. The memory chip of claim 1, wherein the memory chip is a dynamic random-access memory (DRAM) chip, and wherein the memory chip comprises a plurality of DRAM cells.

12. The memory chip of claim 1, wherein the memory chip is a flash memory chip, and wherein the memory chip comprises a plurality of flash memory cells.

13. The memory chip of claim 12, wherein the plurality of flash memory cells is a plurality of NAND-type flash memory cells.

14. The memory chip of claim 1, comprising an encryption engine configured to secure data to be moved to the second microchip or device for long term storage.

15. The memory chip of claim 14, comprising a gatekeeper, configured to:
provide an authentication process for access of the secured data stored in the second microchip or device; and
decrypt the secured data according to the authentication process.

16. A system, comprising:
an intermediate memory chip;
a first memory chip; and
a second memory chip, and wherein the intermediate memory chip comprises:
a first set of pins configured to allow the intermediate memory chip to be coupled to the first memory chip via first wiring;
a second set of pins configured to allow the intermediate memory chip to be coupled to the second memory chip via second wiring that is separate from the first wiring; and
a data mover configured to facilitate access to the second memory chip, via the second set of pins, to read data from the second memory chip and write data to the second memory chip.

17. The system of claim 16, wherein the intermediate memory chip is a non-volatile random-access memory (NVRAM) chip, and wherein the NVRAM chip comprises a plurality of NVRAM cells.

18. The system of claim 17, wherein the second memory chip is a flash memory chip, and wherein the flash memory chip comprises a plurality of flash memory cells.

19. The system of claim 17, wherein the first memory chip is a dynamic random-access memory (DRAM) chip, and wherein the DRAM chip comprises a plurality of DRAM cells.

20. A system, comprising:
an intermediate memory chip;
a second memory chip; and
a processor chip, and wherein the intermediate memory chip comprises:

a first set of pins configured to allow the intermediate memory chip to be coupled to the processor chip via first wiring;

a second set of pins configured to allow the intermediate memory chip to be coupled to the second memory chip via second wiring that is separate from the first wiring; and

a data mover configured to facilitate access to the second memory chip, via the second set of pins and the second wiring, to read data from the second memory chip and write data to the second memory chip.

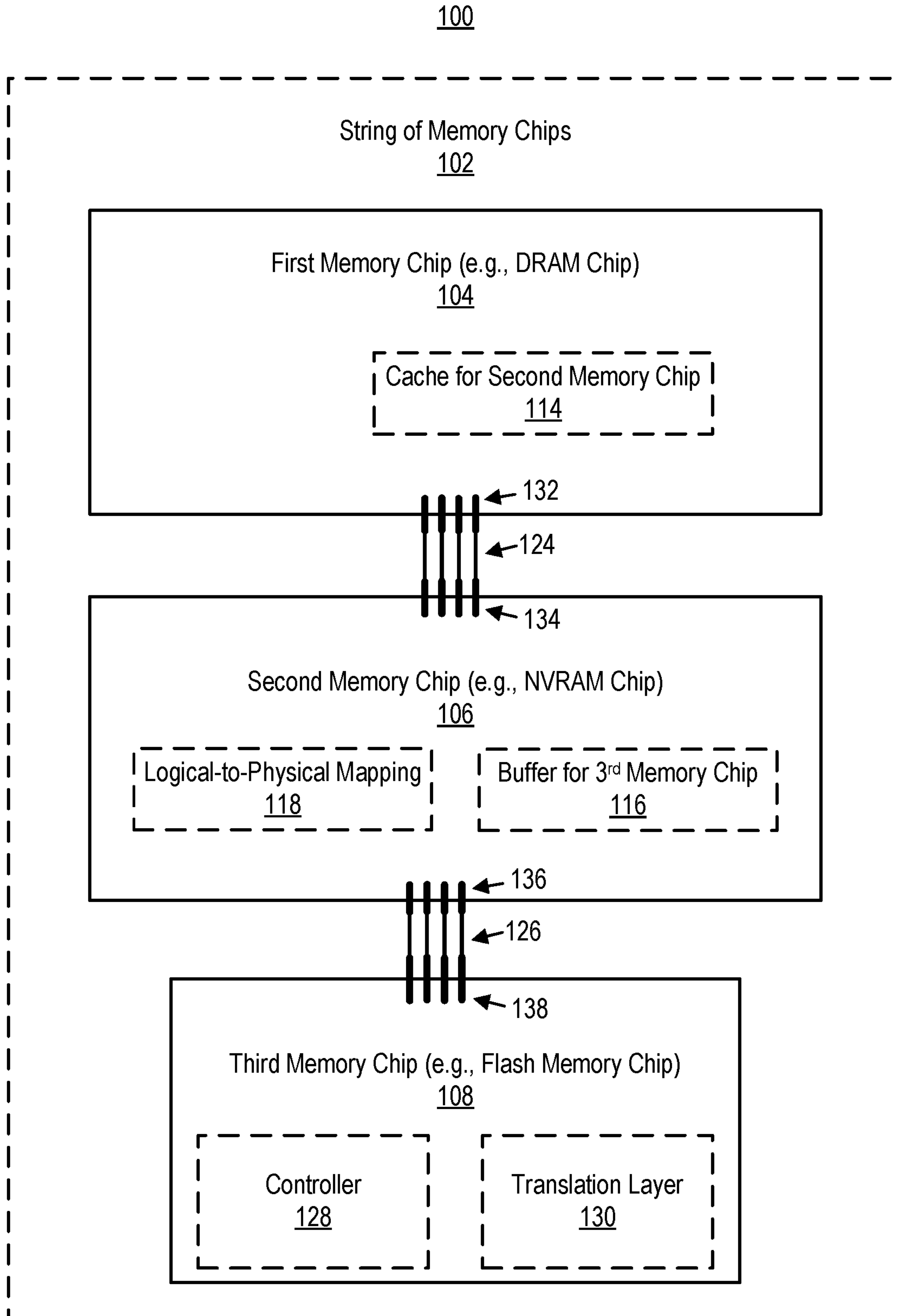


FIG. 1

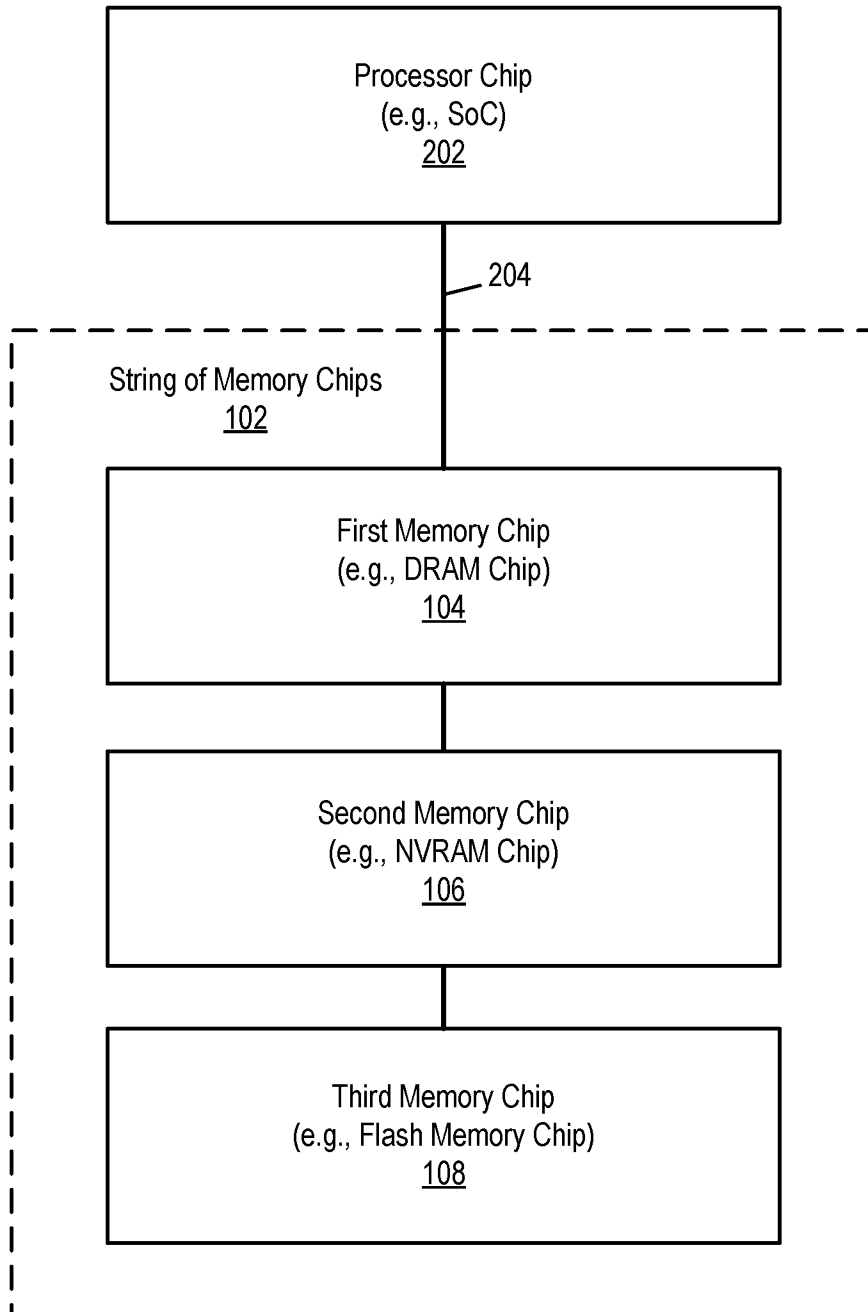


FIG. 2

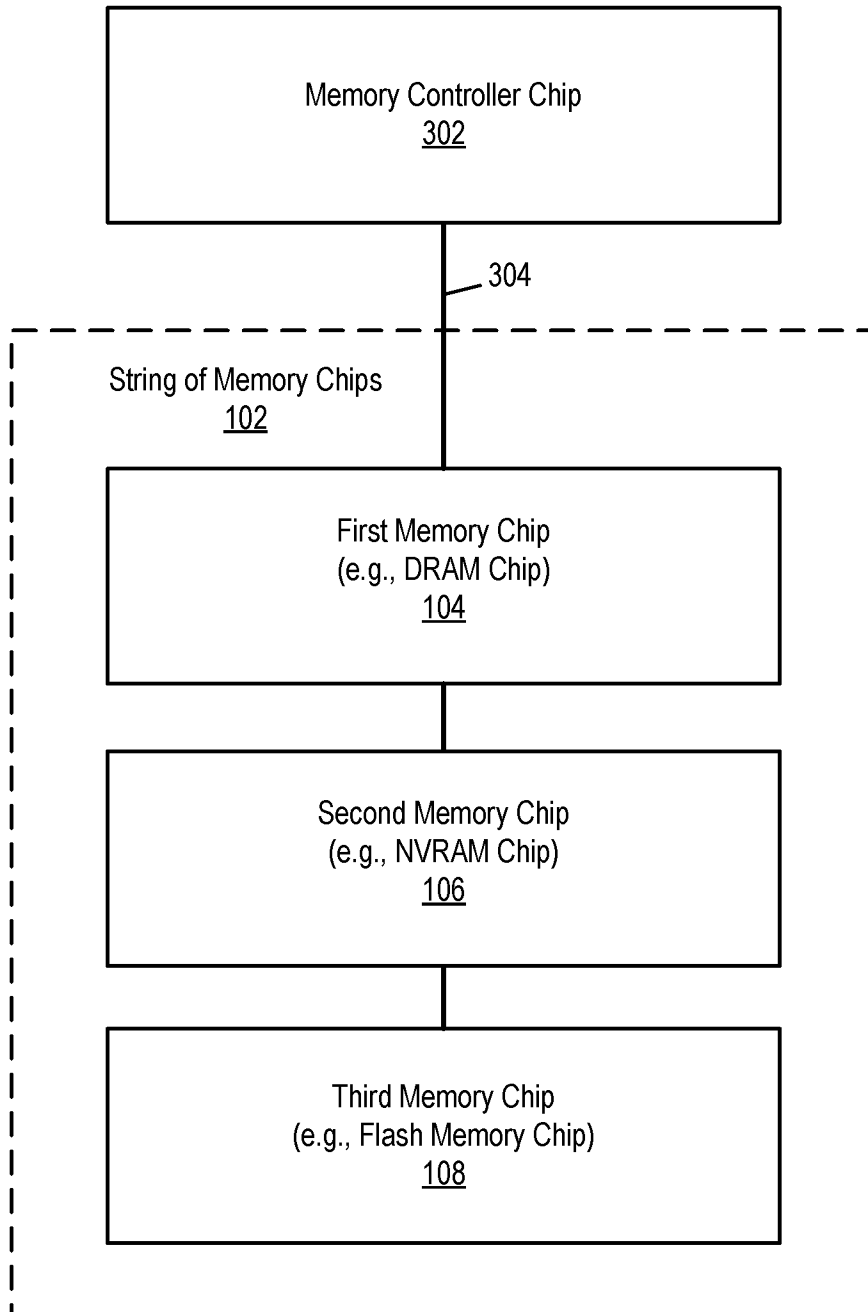


FIG. 3

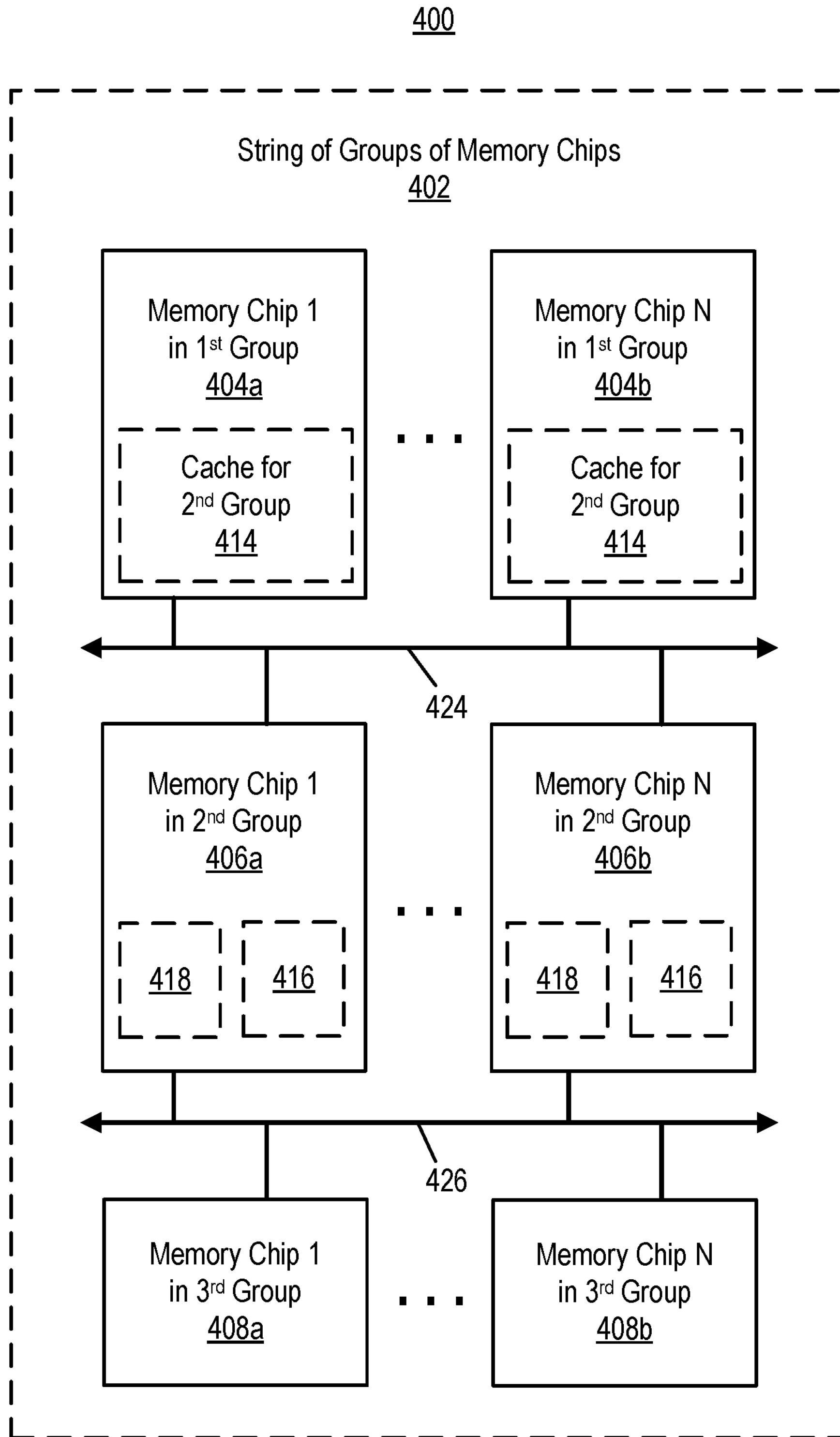


FIG. 4

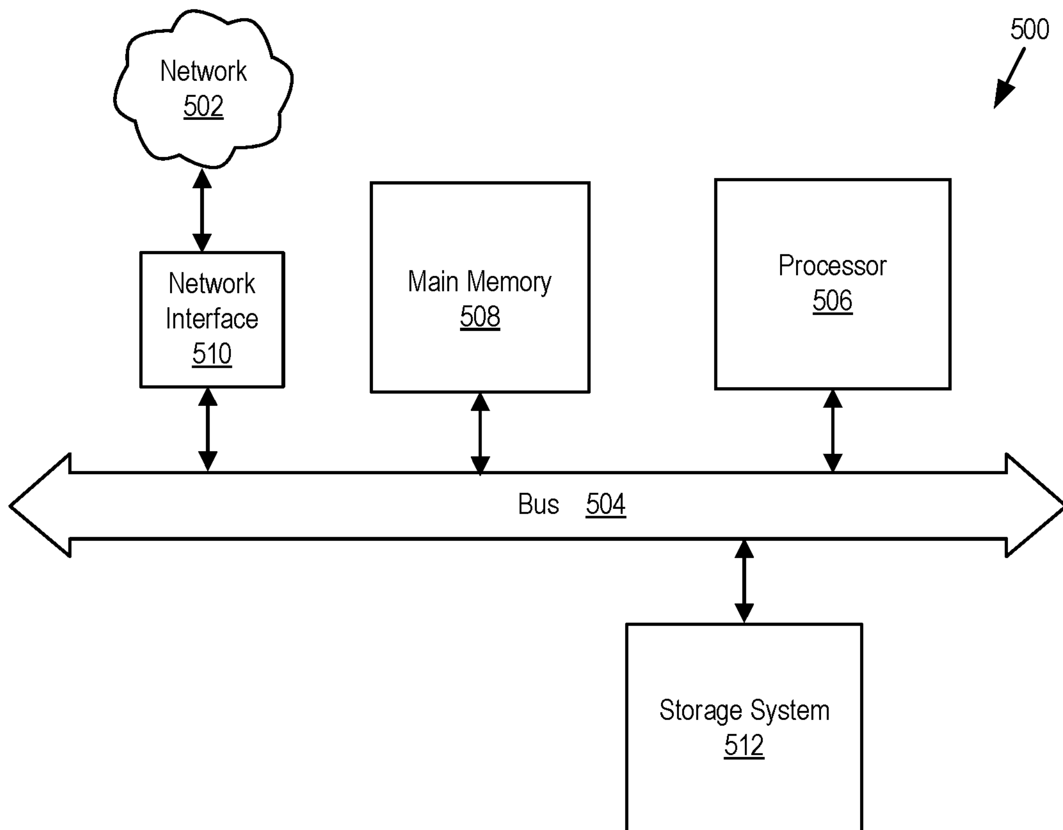


FIG. 5

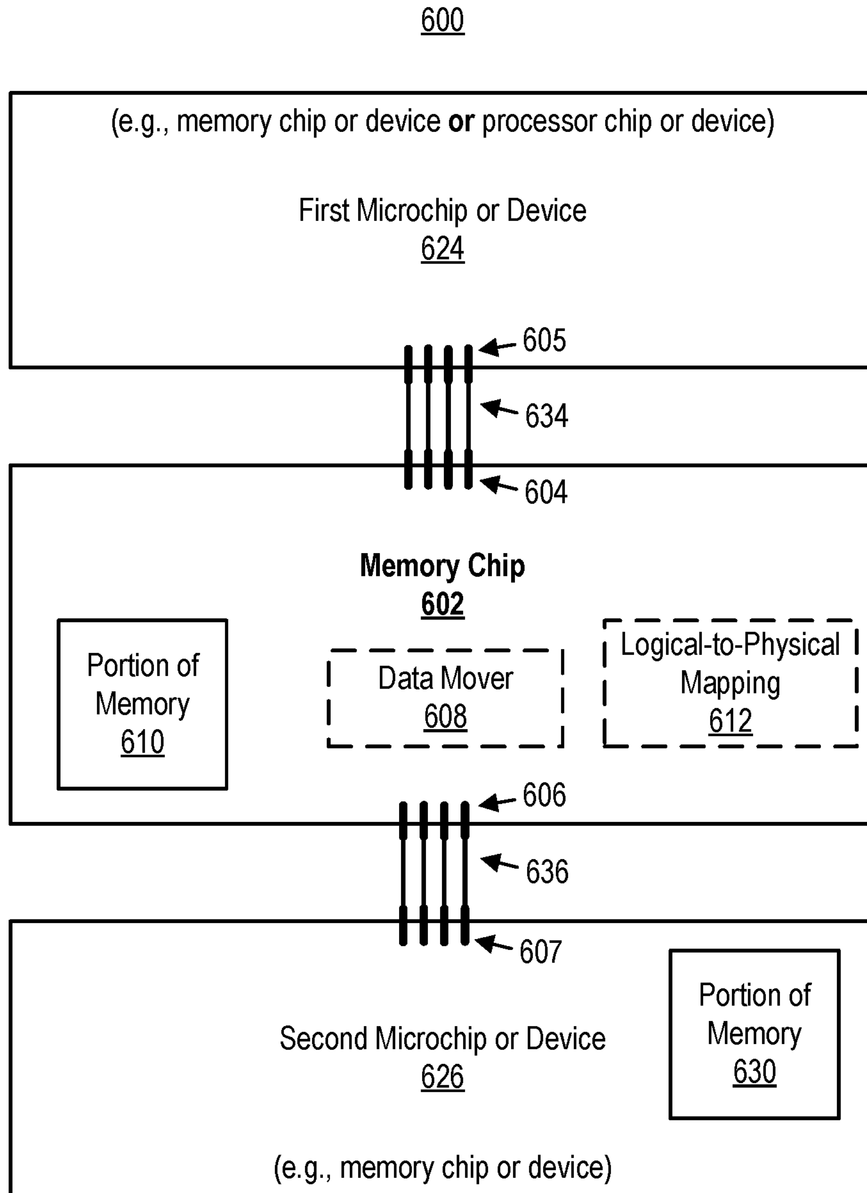


FIG. 6

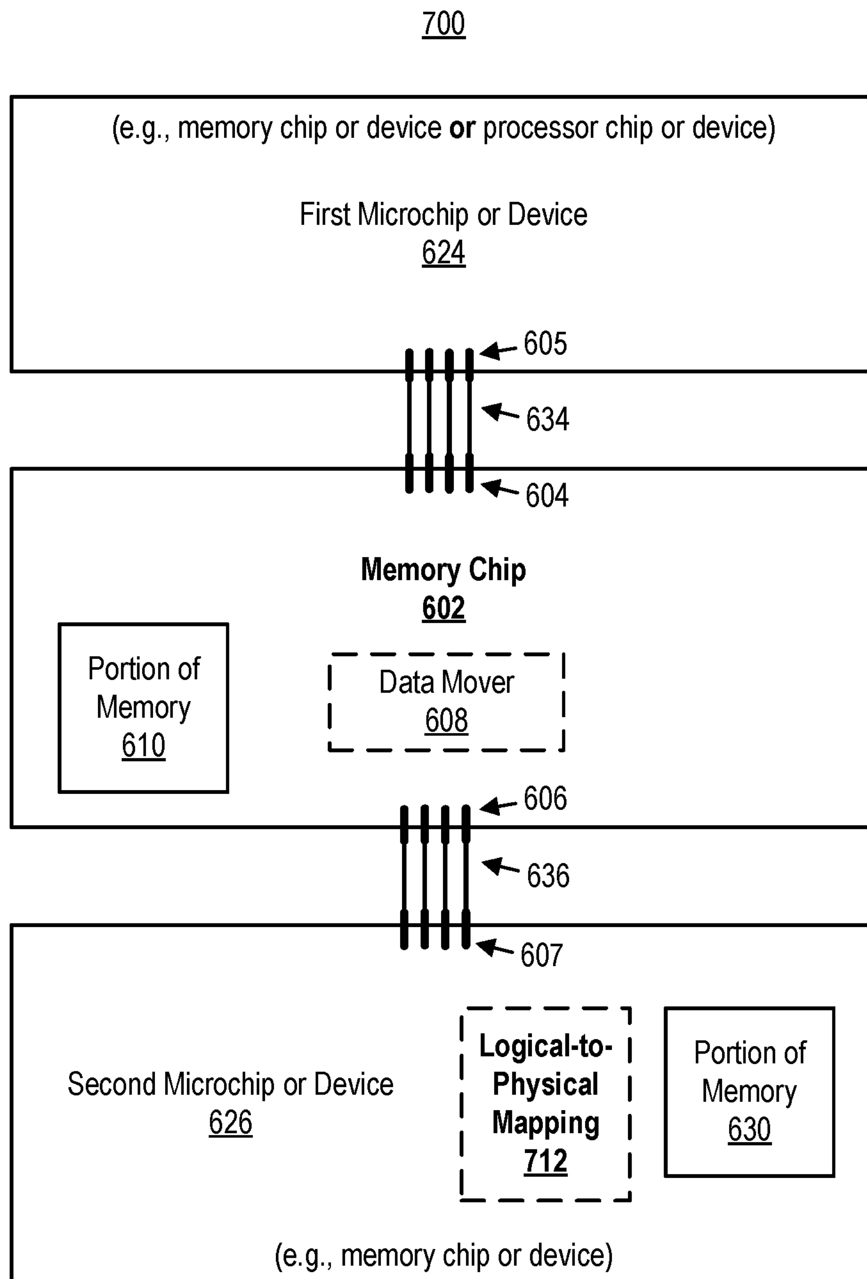


FIG. 7

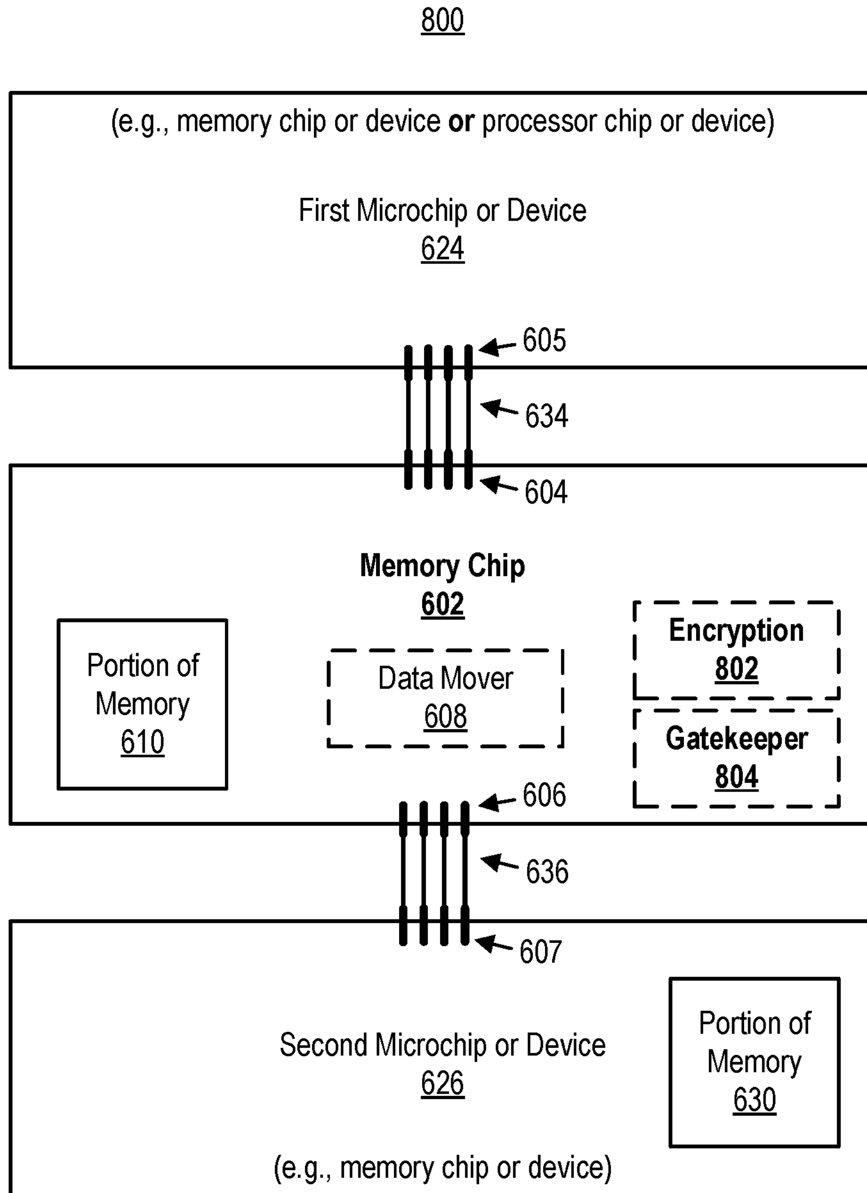


FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2020/049938

A. CLASSIFICATION OF SUBJECT MATTER

G06F 12/1081(2016.01)i, G06F 12/14(2006.01)i, G06F 21/79(2013.01)i, G11C 7/10(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
G06F 12/1081; G06F 012/00; G06F 12/00; G06F 12/08; G06F 12/14; G06F 13/14; G11C 011/406; H04L 9/14; G06F 21/79; G11C 7/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
Korean utility models and applications for utility models
Japanese utility models and applications for utility models

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
eKOMPASS(KIPO internal) & keywords: set of pins, wiring, memory chip, microchip, processor chip, data mover

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2009-0063786 A1 (HAKJUNE OH) 05 March 2009 paragraphs [0030], [0036], [0046], [0101], [0106], [0149]; claims 1, 3; and figures 1A-1B, 3-4, 7	1-13,16-20
Y		14-15
Y	US 2016-0062921 A1 (SAMSUNG ELECTRONICS CO., LTD.) 03 March 2016 paragraph [0064]; and figure 2	14-15
A	US 2010-0030951 A1 (JIN-KI KIM) 04 February 2010 paragraphs [0026]-[0057]; and figures 2-6	1-20
A	US 6928512 B2 (KAZUSHIGE AYUKAWA et al.) 09 August 2005 column 3, line 52 - column 8, line 26; column 17, line 42 - column 18, line 32; and figures 1-2, 19A-20B	1-20
A	US 9195602 B2 (CRAIG HAMPEL et al.) 24 November 2015 column 2, line 53 - column 8, line 4; and figures 1-2B	1-20

☐ Further documents are listed in the continuation of Box C.   ☒ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents,such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
21 December 2020 (21.12.2020)	**21 December 2020 (21.12.2020)**

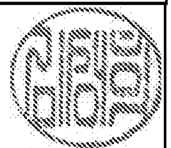
Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsa-ro, Seo-gu, Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

KIM, Sung Hee



Telephone No. +82-42-481-3516

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2009-0063786 A1	05/03/2009	CN 101836258 A	15/09/2010
		EP 2183748 A1	12/05/2010
		JP 2010-537326 A	02/12/2010
		JP 2013-225352 A	31/10/2013
		KR 10-1507192 B1	31/03/2015
		KR 10-2010-0075860 A	05/07/2010
		WO 2009-026696 A1	05/03/2009
US 2016-0062921 A1	03/03/2016	KR 10-2016-0026563 A	09/03/2016
		US 9977749 B2	22/05/2018
US 2010-0030951 A1	04/02/2010	CN 101410814 A	15/04/2009
		CN 103714841 A	09/04/2014
		EP 1999601 A1	10/12/2008
		EP 2031516 A2	04/03/2009
		EP 2348510 A1	27/07/2011
		JP 2009-531746 A	03/09/2009
		JP 2013-037712 A	21/02/2013
		JP 5189072 B2	24/04/2013
		JP 5575856 B2	20/08/2014
		KR 10-1314893 B1	04/10/2013
		KR 10-1365827 B1	20/02/2014
		KR 10-2009-0007280 A	16/01/2009
		KR 10-2013-0073991 A	03/07/2013
		US 2007-0076479 A1	05/04/2007
		US 2007-0076502 A1	05/04/2007
		US 2007-0165457 A1	19/07/2007
		US 7652922 B2	26/01/2010
		WO 2007-109888 A1	04/10/2007
US 6928512 B2	09/08/2005	JP 2003-006041 A	10/01/2003
		KR 10-0884157 B1	17/02/2009
		KR 10-2003-0011231 A	07/02/2003
		US 2002-0199056 A1	26/12/2002
US 9195602 B2	24/11/2015	CN 101689145 A	31/03/2010
		EP 2132635 B1	16/08/2017
		EP 3279798 B1	29/07/2020
		JP 2010-524059 A	15/07/2010
		JP 5401444 B2	29/01/2014
		US 10755794 B2	25/08/2020
		US 2010-0115191 A1	06/05/2010
		US 2016-0098354 A1	07/04/2016
		US 2017-0025187 A1	26/01/2017
		US 2017-0365354 A1	21/12/2017
		US 9460021 B2	04/10/2016
		US 9767918 B2	19/09/2017
		WO 2008-121559 A1	09/10/2008