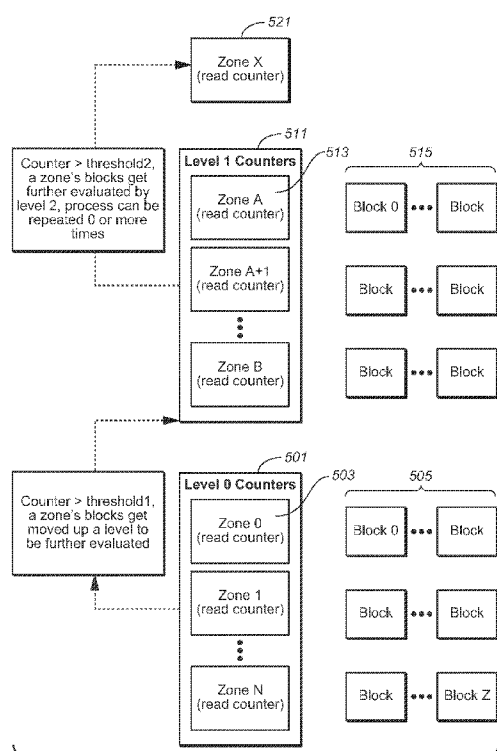




- (51) **International Patent Classification:**
G11C 16/34 (2006.01) *G06F 12/02* (2006.01)
- (21) **International Application Number:**
PCT/US2015/018527
- (22) **International Filing Date:**
3 March 2015 (03.03.2015)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
14/215,924 17 March 2014 (17.03.2014) US
- (71) **Applicant:** SANDISK TECHNOLOGIES INC.
[US/US]; Two Legacy Town Center, 6900 North Dallas Parkway, Plano, Texas 75024 (US).
- (72) **Inventors:** TUERS, Daniel Edward; 1414 Kanepoonui Road, Unit A, Kapaa, Hawaii 96746 (US). ATAKLTI, Yosief; 40878 Robin St., Fremont, California 94538 (US). MANOHAR, Abhijeet; 277/B, 37th A Cross, 8th Block Jayanagar, Bangalore, Karnataka 560082 (IN).
- (74) **Agents:** CLEVELAND, Michael et al.; c/o IP Docketing Department, Davis Wright Tremaine LLP, 1201 Third Avenue Suite 2200, Seattle, Washington 98101-3045 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) **Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, 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,

[Continued on next page]

(54) **Title:** DETECTION OF READ DISTURBANCES ON NON-VOLATILE MEMORIES THROUGH COUNTING OF READ ACCESSES WITHIN DIVISIONS OF THE MEMORY



(57) **Abstract:** In non-volatile memory devices, the accessing of data on word line can degrade the data quality on a neighboring word line, in what is called a read disturb. Techniques are presented for determining word lines likely to suffer read disturbs by use of a hash tree for tracking the number of reads. Read counters are maintained for memory units at a relatively coarse granularity, such as a die or block. When the counter for one of these units reaches a certain level, it is subdivided into sub-units, each with their own read counter, in a process that be repeated to determine frequently read word lines with a fine level of granularity while only using a relatively modest amount of RAM on the controller to store the counters.

WO 2015/142513 A1



SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, **Published:**
GW, KM, ML, MR, NE, SN, TD, TG).

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

DETECTION OF READ DISTURBANCES ON NON-VOLATILE MEMORIES THROUGH COUNTING OF READ ACCESSES WITHIN DIVISIONS OF THE MEMORY

BACKGROUND

5 [0001] The following relates to the operation of re-programmable non-volatile memory systems such as semiconductor flash memory, and, more specifically, to the determination and management of error on such memories.

[0002] Solid-state memory capable of nonvolatile storage of charge, particularly in the form of EEPROM and flash EEPROM packaged as a small form factor card, has recently become the storage of choice in a variety of mobile and handheld devices, notably information appliances and consumer electronics products. Unlike RAM (random access memory) that is also solid-state memory, flash memory is non-volatile, and retaining its stored data even after power is turned off. Also, unlike ROM (read only memory), flash memory is rewritable similar to a disk storage device. In spite of the higher cost, flash memory is increasingly being used in mass storage applications. Conventional mass storage, based on rotating magnetic medium such as hard drives and floppy disks, is unsuitable for the mobile and handheld environment. This is because disk drives tend to be bulky, are prone to mechanical failure and have high latency and high power requirements. These undesirable attributes make disk-based storage impractical in most mobile and portable applications. On the other hand, flash memory, whether embedded, a solid state drive (SSD) and in the form of a removable card is ideally suited in the mobile and handheld environment because of its small size, low power consumption, high speed and high reliability features.

25 [0003] Flash EEPROM is similar to EEPROM (electrically erasable and programmable read-only memory) in that it is a non-volatile memory that can be erased and have new data written or “programmed” into their memory cells. Both utilize a floating (unconnected) conductive gate, in a field effect transistor structure, positioned over a channel region in a semiconductor substrate, between source and drain regions. A control gate is then provided over the floating gate. The threshold voltage characteristic of the transistor is controlled by the amount of charge that is

30

retained on the floating gate. That is, for a given level of charge on the floating gate, there is a corresponding voltage (threshold) that must be applied to the control gate before the transistor is turned "on" to permit conduction between its source and drain regions. In particular, flash memory such as Flash EEPROM allows entire blocks of memory cells to be erased at the same time.

[0004] The floating gate can hold a range of charges and therefore can be programmed to any threshold voltage level within a threshold voltage window. The size of the threshold voltage window is delimited by the minimum and maximum threshold levels of the device, which in turn correspond to the range of the charges that can be programmed onto the floating gate. The threshold window generally depends on the memory device's characteristics, operating conditions and history. Each distinct, resolvable threshold voltage level range within the window may, in principle, be used to designate a definite memory state of the cell.

[0005] The transistor serving as a memory cell is typically programmed to a "programmed" state by one of two mechanisms. In "hot electron injection," a high voltage applied to the drain accelerates electrons across the substrate channel region. At the same time a high voltage applied to the control gate pulls the hot electrons through a thin gate dielectric onto the floating gate. In "tunneling injection," a high voltage is applied to the control gate relative to the substrate. In this way, electrons are pulled from the substrate to the intervening floating gate. While the term "program" has been used historically to describe writing to a memory by injecting electrons to an initially erased charge storage unit of the memory cell so as to alter the memory state, it has now been used interchangeable with more common terms such as "write" or "record."

[0006] The memory device may be erased by a number of mechanisms. For EEPROM, a memory cell is electrically erasable, by applying a high voltage to the substrate relative to the control gate so as to induce electrons in the floating gate to tunnel through a thin oxide to the substrate channel region (i.e., Fowler-Nordheim tunneling.) Typically, the EEPROM is erasable byte by byte. For flash EEPROM, the memory is electrically erasable either all at once or one or more minimum erasable blocks at a time, where a minimum erasable block may consist of one or

more sectors and each sector may store 512 bytes or more of data.

[0007] The memory device typically comprises one or more memory chips that may be mounted on a card or in a solid state drive (SSD). Each memory chip comprises an array of memory cells supported by peripheral circuits such as decoders and erase,
5 write and read circuits. The more sophisticated memory devices also come with a controller that performs intelligent and higher level memory operations and interfacing.

[0008] There are many commercially successful non-volatile solid-state memory devices being used today. These memory devices may be flash EEPROM or may
10 employ other types of nonvolatile memory cells. Examples of flash memory and systems and methods of manufacturing them are given in United States patents nos. 5,070,032, 5,095,344, 5,315,541, 5,343,063, and 5,661,053, 5,313,421 and 6,222,762. In particular, flash memory devices with NAND string structures are described in United States patent nos. 5,570,315, 5,903,495, 6,046,935. Also nonvolatile memory
15 devices are also manufactured from memory cells with a dielectric layer for storing charge. Instead of the conductive floating gate elements described earlier, a dielectric layer is used. Such memory devices utilizing dielectric storage element have been described by Eitan et al., "NROM: A Novel Localized Trapping, 2-Bit Nonvolatile Memory Cell," IEEE Electron Device Letters, vol. 21, no. 11, November 2000, pp.
20 543-545. An ONO dielectric layer extends across the channel between source and drain diffusions. The charge for one data bit is localized in the dielectric layer adjacent to the drain, and the charge for the other data bit is localized in the dielectric layer adjacent to the source. For example, United States patents nos. 5,768,192 and 6,011,725 disclose a nonvolatile memory cell having a trapping dielectric sandwiched
25 between two silicon dioxide layers. Multi-state data storage is implemented by separately reading the binary states of the spatially separated charge storage regions within the dielectric.

[0009] In order to improve read and program performance, multiple charge storage elements or memory transistors in an array are read or programmed in parallel. Thus,
30 a "page" of memory elements are read or programmed together. In existing memory architectures, a row typically contains several interleaved pages or it may constitute

one page. All memory elements of a page will be read or programmed together.

[0010] In flash memory systems, erase operation may take as much as an order of magnitude longer than read and program operations. Thus, it is desirable to have the erase block of substantial size. In this way, the erase time is amortized over a large
5 aggregate of memory cells.

[0011] The nature of flash memory predicates that data must be written to an erased memory location. If data of a certain logical address from a host is to be updated, one way is rewrite the update data in the same physical memory location. That is, the logical to physical address mapping is unchanged. However, this will mean the entire
10 erase block contain that physical location will have to be first erased and then rewritten with the updated data. This method of update is inefficient, as it requires an entire erase block to be erased and rewritten, especially if the data to be updated only occupies a small portion of the erase block. It will also result in a higher frequency of erase recycling of the memory block, which is undesirable in view of the limited
15 endurance of this type of memory device.

[0012] Data communicated through external interfaces of host systems, memory systems and other electronic systems are addressed and mapped into the physical locations of a flash memory system. Typically, addresses of data files generated or received by the system are mapped into distinct ranges of a continuous logical address
20 space established for the system in terms of logical blocks of data (hereinafter the "LBA interface"). The extent of the address space is typically sufficient to cover the full range of addresses that the system is capable of handling. In one example, magnetic disk storage drives communicate with computers or other host systems through such a logical address space. This address space has an extent sufficient to
25 address the entire data storage capacity of the disk drive.

[0013] Flash memory systems are most commonly provided in the form of a memory card or flash drive that is removably connected with a variety of hosts such as a personal computer, a camera or the like, but may also be embedded within such host systems or be in the form of a solid state drive. When writing data to the memory, the
30 host typically assigns unique logical addresses to sectors, clusters or other units of data within a continuous virtual address space of the memory system. Like a disk

operating system (DOS), the host writes data to, and reads data from, addresses within the logical address space of the memory system. A controller within the memory system translates logical addresses received from the host into physical addresses within the memory array, where the data are actually stored, and then keeps track of these address translations. The data storage capacity of the memory system is at least as large as the amount of data that is addressable over the entire logical address space defined for the memory system.

[0014] In current commercial flash memory systems, the size of the erase unit has been increased to a block of enough memory cells to store multiple sectors of data. Indeed, many pages of data are stored in one block, and a page may store multiple sectors of data. Further, two or more blocks are often operated together as metablocks, and the pages of such blocks logically linked together as metapages. A page or metapage of data are written and read together, which can include many sectors of data, thus increasing the parallelism of the operation. Along with such large capacity operating units the challenge is to operate them efficiently.

[0015] For ease of explanation, unless otherwise specified, it is intended that the term “block” as used herein refer to either the block unit of erase or a multiple block “metablock,” depending upon whether metablocks are being used in a specific system. Similarly, reference to a “page” herein may refer to a unit of programming within a single block or a “metapage” within a metablock, depending upon the system configuration.

[0016] In charge storing non-volatile memory devices, such as those based on EEPROM or dielectric storage, the non-volatility is non-perfect and the data value stored in an element can be degraded by leakage over time. Operations in one part of the memory circuit can also affect data values in storage elements not being directly acted upon. For example, capacitive coupling between memory cells on adjacent word lines can lead “read disturbs” or “write disturbs” when a neighboring word line is read or written. (Erase disturbs may also occur, depending on the granularity at which cells are erased.) To maintain the integrity of data in such memory systems, the effects of such disturbs need to be considered.

SUMMARY OF THE INVENTION

[0017] According to a general aspect of the invention, a method is presented for the operating of a non-volatile memory system having one or more memory circuits and a controller circuit. The memory circuits each include one or more arrays of non-volatile memory cells formed along word lines and the controller circuit manages the storage of data on the memory circuit. For each of a first plurality of distinct divisions of the one or more memory arrays the controller circuit maintains a count of the number of times word lines of the corresponding division is accessed for a read operation. In response to one of the counts for a corresponding division reaching a first threshold value, subdividing the corresponding division is subdivide into a second plurality of distinct first subdivisions. For each of the first subdivisions the controller circuit subsequently maintains a count of the number of times word lines of the subdivision is accessed for read operations.

[0018] Various aspects, advantages, features and embodiments of the present invention are included in the following description of exemplary examples thereof, which description should be taken in conjunction with the accompanying drawings. All patents, patent applications, articles, other publications, documents and things referenced herein are hereby incorporated herein by this reference in their entirety for all purposes. To the extent of any inconsistency or conflict in the definition or use of terms between any of the incorporated publications, documents or things and the present application, those of the present application shall prevail.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] **FIG. 1** illustrates schematically the main hardware components of a memory system suitable for implementing the present invention.

[0020] **FIG. 2** illustrates schematically a non-volatile memory cell.

[0021] **FIG. 3** illustrates the relation between the source-drain current I_D and the control gate voltage V_{CG} for four different charges Q1-Q4 that the floating gate may be selectively storing at any one time.

[0022] **FIG. 4A** illustrates schematically a string of memory cells organized into an

NAND string.

[0023] **FIG. 4B** illustrates an example of an NAND array 210 of memory cells, constituted from NAND strings 50 such as that shown in **FIG. 4A**.

5 [0024] **FIG. 5** illustrates a page of memory cells, organized for example in the NAND configuration, being sensed or programmed in parallel.

[0025] **FIG. 6(0) - 6(2)** illustrate an example of programming a population of 4-state memory cells.

[0026] **FIGS. 7A-7E** illustrate the programming and reading of the 4-state memory encoded with a given 2-bit code.

10 [0027] **FIG. 7F** illustrates a foggy-fine programming for an 8-state memory encoded with a given 3-bit code.

[0028] **FIG. 8** illustrates the memory being managed by a memory manager with is a software component that resides in the controller.

[0029] **FIG. 9** illustrates the software modules of the back-end system.

15 [0030] **FIGS. 10A(i) – 10A(iii)** illustrate schematically the mapping between a logical group and a metablock. **FIG. 10B** illustrates schematically the mapping between logical groups and metablocks.

[0031] **FIG. 11** is a schematic representation of the process for a two level example.

[0032] **FIGS. 12 and 13** provide further illustration of a hash tree to track read zones.

20

DETAILED DESCRIPTION

MEMORY SYSTEM

[0033] **FIG. 1** to **FIG. 10B** provide example memory systems in which the various aspects of the present invention may be implemented or illustrated.

[0034] FIG. 1 illustrates schematically the main hardware components of a memory system suitable for implementing the present invention. The memory system 90 typically operates with a host 80 through a host interface. The memory system is typically in the form of a memory card, solid state drive (SSD), or an embedded memory system. The memory system 90 includes a memory 200 whose operations are controlled by a controller 100. The memory 200 comprises of one or more array of non-volatile memory cells distributed over one or more integrated circuit chip. The controller 100 includes an interface 110, a processor 120, an optional coprocessor 121, ROM 122 (read-only-memory), RAM 130 (random access memory) and optionally programmable nonvolatile memory 124. The interface 110 has one component interfacing the controller to a host and another component interfacing to the memory 200. Firmware stored in nonvolatile ROM 122 and/or the optional nonvolatile memory 124 provides codes for the processor 120 to implement the functions of the controller 100. Error correction codes may be processed by the processor 120 or the optional coprocessor 121. In an alternative embodiment, the controller 100 is implemented by a state machine (not shown.) In yet another embodiment, the controller 100 is implemented within the host.

Physical Memory Structure

[0035] With respect to the memory section 200, memory devices include volatile memory devices, such as dynamic random access memory (“DRAM”) or static random access memory (“SRAM”) devices, non-volatile memory devices, such as resistive random access memory (“ReRAM”), electrically erasable programmable read only memory (“EEPROM”), flash memory (which can also be considered a subset of EEPROM), ferroelectric random access memory (“FRAM”), and magnetoresistive random access memory (“MRAM”), and other semiconductor elements capable of storing information. Furthermore, each type of memory device may have different configurations. For example, flash memory devices may be configured in a NAND or a NOR configuration.

[0036] The memory devices can be formed from passive and/or active elements, in any combinations. By way of non-limiting example, passive semiconductor memory elements include ReRAM device elements, which in some embodiments include a

resistivity switching storage element, such as an anti-fuse, phase change material, etc., and optionally a steering element, such as a diode, *etc.* Further by way of non-limiting example, active semiconductor memory elements include EEPROM and flash memory device elements, which in some embodiments include elements containing a charge storage region, such as a floating gate, conductive nanoparticles or a charge storage dielectric material.

[0037] Multiple memory elements may be configured so that they are connected in series or such that each element is individually accessible. By way of non-limiting example, NAND devices contain memory elements (*e.g.*, devices containing a charge storage region) connected in series. For example, a NAND memory array may be configured so that the array is composed of multiple strings of memory in which each string is composed of multiple memory elements sharing a single bit line and accessed as a group. In contrast, memory elements may be configured so that each element is individually accessible, *e.g.*, a NOR memory array. One of skill in the art will recognize that the NAND and NOR memory configurations are exemplary, and memory elements may be otherwise configured.

[0038] The semiconductor memory elements of a single device, such as elements located within and/or over the same substrate or in a single die, may be distributed in two or three dimensions, such as a two dimensional array structure or a three dimensional array structure.

[0039] In a two dimensional memory structure, the semiconductor memory elements are arranged in a single plane or single memory device level. Typically, in a two dimensional memory structure, memory elements are located in a plane (*e.g.*, in an x-z direction plane) which extends substantially parallel to a major surface of a substrate that supports the memory elements. The substrate may be a wafer over which the layers of the memory elements are deposited and/or in which memory elements are formed or it may be a carrier substrate which is attached to the memory elements after they are formed. As a non-limiting example, the substrate may include a semiconductor such as silicon.

[0040] The memory elements may be arranged in the single memory device level in an ordered array, such as in a plurality of rows and/or columns. However, the

memory elements may be arranged in non-regular or non-orthogonal configurations as understood by one of skill in the art. The memory elements may each have two or more electrodes or contact lines, such as bit lines and word lines.

5 [0041] A three dimensional memory array is organized so that memory elements occupy multiple planes or multiple device levels, forming a structure in three dimensions (*i.e.*, in the x, y and z directions, where the y direction is substantially perpendicular and the x and z directions are substantially parallel to the major surface of the substrate).

10 [0042] As a non-limiting example, each plane in a three dimensional memory array structure may be physically located in two dimensions (one memory level) with multiple two dimensional memory levels to form a three dimensional memory array structure. As another non-limiting example, a three dimensional memory array may be physically structured as multiple vertical columns (*e.g.*, columns extending substantially perpendicular to the major surface of the substrate in the y direction)
15 having multiple elements in each column and therefore having elements spanning several vertically stacked memory planes. The columns may be arranged in a two dimensional configuration, *e.g.*, in an x-z plane, thereby resulting in a three dimensional arrangement of memory elements. One of skill in the art will understand that other configurations of memory elements in three dimensions will also constitute
20 a three dimensional memory array.

[0043] By way of non-limiting example, in a three dimensional NAND memory array, the memory elements may be connected together to form a NAND string within a single horizontal (*e.g.*, x-z) plane. Alternatively, the memory elements may be connected together to extend through multiple horizontal planes. Other three
25 dimensional configurations can be envisioned wherein some NAND strings contain memory elements in a single memory level while other strings contain memory elements which extend through multiple memory levels. Three dimensional memory arrays may also be designed in a NOR configuration and in a ReRAM configuration.

[0044] A monolithic three dimensional memory array is one in which multiple
30 memory levels are formed above and/or within a *single* substrate, such as a semiconductor wafer. In a monolithic three dimensional array the layers of each

level of the array are formed on the layers of each underlying level of the array. One of skill in the art will understand that layers of adjacent levels of a monolithic three dimensional memory array may be shared or have intervening layers between memory levels. In contrast, two dimensional arrays may be formed separately and then packaged together to form a non-monolithic memory device. For example, non-monolithic stacked memories have been constructed by forming memory levels on separate substrates and adhering the memory levels atop each other. The substrates may be thinned or removed from the memory levels before bonding, but as the memory levels are initially formed over separate substrates, such memories are not monolithic three dimensional memory arrays. Further, multiple two dimensional memory arrays or three dimensional memory arrays (monolithic or non-monolithic) may be formed separately and then packaged together to form a stacked-chip memory device.

[0045] Associated circuitry is typically required for proper operation of the memory elements and for proper communication with the memory elements. This associated circuitry may be on the same substrate as the memory array and/or on a separate substrate. As non-limiting examples, the memory devices may have driver circuitry and control circuitry used in the programming and reading of the memory elements.

[0046] One of skill in the art will recognize that this invention is not limited to the two dimensional and three dimensional exemplary structures described but cover all relevant memory structures within the spirit and scope of the invention as described herein and as understood by one of skill in the art.

[0047] FIG. 2 illustrates schematically a non-volatile memory cell. The memory cell 10 can be implemented by a field-effect transistor having a charge storage unit 20, such as a floating gate or a dielectric layer. The memory cell 10 also includes a source 14, a drain 16, and a control gate 30.

[0048] There are many commercially successful non-volatile solid-state memory devices being used today. These memory devices may employ different types of memory cells, each type having one or more charge storage element.

[0049] Typical non-volatile memory cells include EEPROM and flash EEPROM.

Examples of EEPROM cells and methods of manufacturing them are given in United States patent no. 5,595,924. Examples of flash EEPROM cells, their uses in memory systems and methods of manufacturing them are given in United States patents nos. 5,070,032, 5,095,344, 5,315,541, 5,343,063, 5,661,053, 5,313,421 and 6,222,762. In particular, examples of memory devices with NAND cell structures are described in United States patent nos. 5,570,315, 5,903,495, 6,046,935. Also, examples of memory devices utilizing dielectric storage element have been described by Eitan et al., "NROM: A Novel Localized Trapping, 2-Bit Nonvolatile Memory Cell," IEEE Electron Device Letters, vol. 21, no. 11, November 2000, pp. 543-545, and in United States patents nos. 5,768,192 and 6,011,725.

[0050] In practice, the memory state of a cell is usually read by sensing the conduction current across the source and drain electrodes of the cell when a reference voltage is applied to the control gate. Thus, for each given charge on the floating gate of a cell, a corresponding conduction current with respect to a fixed reference control gate voltage may be detected. Similarly, the range of charge programmable onto the floating gate defines a corresponding threshold voltage window or a corresponding conduction current window.

[0051] Alternatively, instead of detecting the conduction current among a partitioned current window, it is possible to set the threshold voltage for a given memory state under test at the control gate and detect if the conduction current is lower or higher than a threshold current. In one implementation the detection of the conduction current relative to a threshold current is accomplished by examining the rate the conduction current is discharging through the capacitance of the bit line.

[0052] FIG. 3 illustrates the relation between the source-drain current I_D and the control gate voltage V_{CG} for four different charges Q_1 - Q_4 that the floating gate may be selectively storing at any one time. The four solid I_D versus V_{CG} curves represent four possible charge levels that can be programmed on a floating gate of a memory cell, respectively corresponding to four possible memory states. As an example, the threshold voltage window of a population of cells may range from 0.5V to 3.5V. Seven possible memory states "0", "1", "2", "3", "4", "5", "6", respectively representing one erased and six programmed states may be demarcated by partitioning

the threshold window into five regions in interval of 0.5V each. For example, if a reference current, I_{REF} of 2 μA is used as shown, then the cell programmed with Q1 may be considered to be in a memory state “1” since its curve intersects with I_{REF} in the region of the threshold window demarcated by $V_{CG} = 0.5V$ and $1.0V$. Similarly,
5 Q4 is in a memory state “5”.

[0053] As can be seen from the description above, the more states a memory cell is made to store, the more finely divided is its threshold window. For example, a memory device may have memory cells having a threshold window that ranges from $-1.5V$ to $5V$. This provides a maximum width of $6.5V$. If the memory cell is to store
10 16 states, each state may occupy from $200mV$ to $300mV$ in the threshold window. This will require higher precision in programming and reading operations in order to be able to achieve the required resolution.

[0054] FIG. 4A illustrates schematically a string of memory cells organized into an NAND string. An NAND string 50 comprises of a series of memory transistors M1, M2, ... Mn (e.g., $n = 4, 8, 16$ or higher) daisy-chained by their sources and drains. A pair of select transistors S1, S2 controls the memory transistors chain's connection to the external via the NAND string's source terminal 54 and drain terminal 56 respectively. In a memory array, when the source select transistor S1 is turned on, the source terminal is coupled to a source line (see FIG. 4B). Similarly, when the drain
15 select transistor S2 is turned on, the drain terminal of the NAND string is coupled to a bit line of the memory array. Each memory transistor 10 in the chain acts as a memory cell. It has a charge storage element 20 to store a given amount of charge so as to represent an intended memory state. A control gate 30 of each memory transistor allows control over read and write operations. As will be seen in FIG. 4B,
20 the control gates 30 of corresponding memory transistors of a row of NAND string are all connected to the same word line. Similarly, a control gate 32 of each of the select transistors S1, S2 provides control access to the NAND string via its source terminal 54 and drain terminal 56 respectively. Likewise, the control gates 32 of corresponding select transistors of a row of NAND string are all connected to the
25 same select line.
30

[0055] When an addressed memory transistor 10 within an NAND string is read or is

verified during programming, its control gate 30 is supplied with an appropriate voltage. At the same time, the rest of the non-addressed memory transistors in the NAND string 50 are fully turned on by application of sufficient voltage on their control gates. In this way, a conductive path is effectively created from the source of the individual memory transistor to the source terminal 54 of the NAND string and likewise for the drain of the individual memory transistor to the drain terminal 56 of the cell. Memory devices with such NAND string structures are described in United States patent nos. 5,570,315, 5,903,495, 6,046,935.

[0056] **FIG. 4B** illustrates an example of an NAND array 210 of memory cells, constituted from NAND strings 50 such as that shown in **FIG. 4A**. Along each column of NAND strings, a bit line such as bit line 36 is coupled to the drain terminal 56 of each NAND string. Along each bank of NAND strings, a source line such as source line 34 is coupled to the source terminals 54 of each NAND string. Also the control gates along a row of memory cells in a bank of NAND strings are connected to a word line such as word line 42. The control gates along a row of select transistors in a bank of NAND strings are connected to a select line such as select line 44. An entire row of memory cells in a bank of NAND strings can be addressed by appropriate voltages on the word lines and select lines of the bank of NAND strings. When a memory transistor within a NAND string is being read, the remaining memory transistors in the string are turned on hard via their associated word lines so that the current flowing through the string is essentially dependent upon the level of charge stored in the cell being read.

[0057] **FIG. 5** illustrates a page of memory cells, organized for example in the NAND configuration, being sensed or programmed in parallel. **FIG. 5** essentially shows a bank of NAND strings 50 in the memory array 210 of **FIG. 4B**, where the detail of each NAND string is shown explicitly as in **FIG. 4A**. A “page” such as the page 60, is a group of memory cells enabled to be sensed or programmed in parallel. This is accomplished by a corresponding page of sense amplifiers 212. The sensed results are latched in a corresponding set of latches 214. Each sense amplifier can be coupled to a NAND string via a bit line. The page is enabled by the control gates of the cells of the page connected in common to a word line 42 and each cell accessible by a sense amplifier accessible via a bit line 36. As an example, when respectively

sensing or programming the page of cells 60, a sensing voltage or a programming voltage is respectively applied to the common word line WL3 together with appropriate voltages on the bit lines.

Physical Organization of the Memory

5 [0058] One important difference between flash memory and of type of memory is that a cell must be programmed from the erased state. That is the floating gate must first be emptied of charge. Programming then adds a desired amount of charge back to the floating gate. It does not support removing a portion of the charge from the floating
10 to go from a more programmed state to a lesser one. This means that update data cannot overwrite existing one and must be written to a previous unwritten location.

[0059] Furthermore erasing is to empty all the charges from the floating gate and generally takes appreciably time. For that reason, it will be cumbersome and very slow to erase cell by cell or even page by page. In practice, the array of memory cells is divided into a large number of blocks of memory cells. As is common for flash
15 EEPROM systems, the block is the unit of erase. That is, each block contains the minimum number of memory cells that are erased together. While aggregating a large number of cells in a block to be erased in parallel will improve erase performance, a large size block also entails dealing with a larger number of update and obsolete data. Just before the block is erased, a garbage collection is required to salvage the non-
20 obsolete data in the block.

[0060] Each block is typically divided into a number of pages. A page is a unit of programming or reading. In one embodiment, the individual pages may be divided into segments and the segments may contain the fewest number of cells that are written at one time as a basic programming operation. One or more pages of data are
25 typically stored in one row of memory cells. A page can store one or more sectors. A sector includes user data and overhead data. Multiple blocks and pages distributed across multiple arrays can also be operated together as metablocks and metapages. If they are distributed over multiple chips, they can be operated together as megablocks and megapage.

30 Examples of Multi-level Cell ("MLC") Memory Partitioning

[0061] A nonvolatile memory in which the memory cells each stores multiple bits of data has already been described in connection with **FIG. 3**. A particular example is a memory formed from an array of field-effect transistors, each having a charge storage layer between its channel region and its control gate. The charge storage layer or unit can store a range of charges, giving rise to a range of threshold voltages for each field-effect transistor. The range of possible threshold voltages spans a threshold window. When the threshold window is partitioned into multiple sub-ranges or zones of threshold voltages, each resolvable zone is used to represent a different memory states for a memory cell. The multiple memory states can be coded by one or more binary bits. For example, a memory cell partitioned into four zones can support four states which can be coded as 2-bit data. Similarly, a memory cell partitioned into eight zones can support eight memory states which can be coded as 3-bit data, etc.

All-bit, Full-Sequence MLC Programming

[0062] **FIG. 6(0) - 6(2)** illustrate an example of programming a population of 4-state memory cells. **FIG. 6(0)** illustrates the population of memory cells programmable into four distinct distributions of threshold voltages respectively representing memory states “0”, “1”, “2” and “3”. **FIG. 6(1)** illustrates the initial distribution of “erased” threshold voltages for an erased memory. **FIG. 6(2)** illustrates an example of the memory after many of the memory cells have been programmed. Essentially, a cell initially has an “erased” threshold voltage and programming will move it to a higher value into one of the three zones demarcated by verify levels vV_1 , vV_2 and vV_3 . In this way, each memory cell can be programmed to one of the three programmed state “1”, “2” and “3” or remain un-programmed in the “erased” state. As the memory gets more programming, the initial distribution of the “erased” state as shown in **FIG. 6(1)** will become narrower and the erased state is represented by the “0” state.

[0063] A 2-bit code having a lower bit and an upper bit can be used to represent each of the four memory states. For example, the “0”, “1”, “2” and “3” states are respectively represented by “11”, “01”, “00” and “10”. The 2-bit data may be read from the memory by sensing in “full-sequence” mode where the two bits are sensed together by sensing relative to the read demarcation threshold values rV_1 , rV_2 and rV_3

in three sub-passes respectively.

Bit-by-Bit MLC Programming and Reading

[0064] FIGs. 7A-7E illustrate the programming and reading of the 4-state memory encoded with a given 2-bit code. FIG. 7A illustrates threshold voltage distributions of the 4-state memory array when each memory cell stores two bits of data using the 2-bit code. Such a 2-bit code has been disclosed in US Patent Application No. 10/830,824 filed April 24, 2004 by Li et al., entitled "NON-VOLATILE MEMORY AND CONTROL WITH IMPROVED PARTIAL PAGE PROGRAM CAPABILITY".

10 [0065] FIG. 7B illustrates the lower page programming (lower bit) in a 2-pass programming scheme using the 2-bit code. The fault-tolerant LM New code essentially avoids any upper page programming to transit through any intermediate states. Thus, the first pass lower page programming has the logical state (upper bit, lower bit) = (1, 1) transits to some intermediate state (x, 0) as represented by programming the "unprogrammed" memory state "0" to the "intermediate" state designated by (x, 0) with a programmed threshold voltage greater than D_A but less than D_C .

[0066] FIG. 7C illustrates the upper page programming (upper bit) in the 2-pass programming scheme using the 2-bit code. In the second pass of programming the upper page bit to "0", if the lower page bit is at "1", the logical state (1, 1) transits to (0, 1) as represented by programming the "unprogrammed" memory state "0" to "1". If the lower page bit is at "0", the logical state (0, 0) is obtained by programming from the "intermediate" state to "3". Similarly, if the upper page is to remain at "1", while the lower page has been programmed to "0", it will require a transition from the "intermediate" state to (1, 0) as represented by programming the "intermediate" state to "2".

[0067] FIG. 7D illustrates the read operation that is required to discern the lower bit of the 4-state memory encoded with the 2-bit code. A readB operation is first performed to determine if the LM flag can be read. If so, the upper page has been programmed and the readB operation will yield the lower page data correctly. On the

other hand, if the upper page has not yet been programmed, the lower page data will be read by a readA operation.

[0068] FIG. 7E illustrates the read operation that is required to discern the upper bit of the 4-state memory encoded with the 2-bit code. As is clear from the figure, the upper page read will require a 3-pass read of readA, readB and readC, respectively relative to the demarcation threshold voltages D_A , D_B and D_C .

[0069] In the bit-by-bit scheme for a 2-bit memory, a physical page of memory cells will store two logical data pages, a lower data page corresponding to the lower bit and an upper data page corresponding to the upper bit.

10 Foggy-Fine Programming

[0070] Another variation on multi-state programming employs a foggy-fine algorithm, as is illustrated in FIG. 7F for a 3-bit memory example. As shown there, this another multi-phase programming operation. A first programming operation is performed as shown in the top line, followed the foggy programming stage. The foggy phase is a full 3-bit programming operation from the first phase using all eight of the final states. At the end of the foggy, though, the data in these states is not yet fully resolved into well-defined distributions for each of the 8 states (hence, the “foggy” name) and is not readily extractable.

[0071] As each cell is, however, programmed to near its eventual target state, the sort of neighboring cell to cell couplings, or “Yupin” effect, described in US patent number 6,870,768 are presenting most of their effect. Because of this, when the fine program phase (shown on the bottom line) is executed, these couplings have largely been factored in to this final phase so the cell distributions are more accurately resolved to their target ranges. More detail on these subjects is given in US patents numbers 6,870,768 and 6,657,891 and in the US patent application entitled “Atomic Program Sequence and Write Abort Detection” by Gorobets et al. having attorney application number 12/642,740, which was filed December 18, 2009, and which presents a “diagonal” first-foggy-fine method.

Binary and MLC Memory Partitioning

[0072] FIG. 6 and FIG. 7 illustrate examples of a 2-bit (also referred to as “D2”) memory. As can be seen, a D2 memory has its threshold range or window partitioned into 4 regions, designating 4 states. Similarly, in D3, each cell stores 3 bits (low, middle and upper bits) and there are 8 regions. In D4, there are 4 bits and 16 regions, etc. As the memory’s finite threshold window is partitioned into more regions, the resolution and for programming and reading will necessarily become finer. Two issues arise as the memory cell is configured to store more bits.

[0073] First, programming or reading will be slower when the threshold of a cell must be more accurately programmed or read. In fact in practice the sensing time (needed in programming and reading) tends to increase as the square of the number of partitioning levels.

[0074] Secondly, flash memory has an endurance problem as it ages with use. When a cell is repeatedly programmed and erased, charges is shuttled in and out of the floating gate 20 (see FIG. 2) by tunneling across a dielectric. Each time some charges may become trapped in the dielectric and will modify the threshold of the cell. In fact over use, the threshold window will progressively narrow. Thus, MLC memory generally is designed with tradeoffs between capacity, performance and reliability.

[0075] Conversely, it will be seen for a binary memory, the memory’s threshold window is only partitioned into two regions. This will allow a maximum margin of errors. Thus, binary partitioning while diminished in storage capacity will provide maximum performance and reliability.

[0076] The multi-pass, bit-by-bit programming and reading technique described in connection with FIG. 7 provides a smooth transition between MLC and binary partitioning. In this case, if the memory is programmed with only the lower bit, it is effectively a binary partitioned memory. While this approach does not fully optimize the range of the threshold window as in the case of a single-level cell (“SLC”) memory, it has the advantage of using the same demarcation or sensing level as in the operations of the lower bit of the MLC memory. As will be described later, this approach allows a MLC memory to be “expropriated” for use as a binary memory, or vice versa. How it should be understood that MLC memory tends to have more

stringent specification for usage.

Binary Memory and Partial Page Programming

[0077] The charge programmed into the charge storage element of one memory cell produces an electric field that perturbs the electric field of a neighboring memory cell.

- 5 This will affect the characteristics of the neighboring memory cell which essentially is a field-effect transistor with a charge storage element. In particular, when sensed the memory cell will appear to have a higher threshold level (or more programmed) than when it is less perturbed.

- [0078] In general, if a memory cell is program-verified under a first field environment and later is read again under a different field environment due to neighboring cells subsequently being programmed with different charges, the read accuracy may be affected due to coupling between neighboring floating gates in what is referred to as the “Yupin Effect”. With ever higher integration in semiconductor memories, the perturbation of the electric field due to the stored charges between memory cells
- 10 (Yupin effect) becomes increasingly appreciable as the inter-cellular spacing shrinks.

- [0079] The Bit-by-Bit MLC Programming technique described in connection with **FIG. 7** above is designed to minimize program disturb from cells along the same word line. As can be seen from **FIG. 7B**, in a first of the two programming passes, the thresholds of the cells are moved at most half way up the threshold window. The effect of the first pass is overtaken by the final pass. In the final pass, the thresholds are only moved a quarter of the way. In other words, for D2, the charge difference among neighboring cells is limited to a quarter of its maximum. For D3, with three
- 20 passes, the final pass will limit the charge difference to one-eighth of its maximum.

- [0080] However, the bit-by-bit multi-pass programming technique will be compromised by partial-page programming. A page is a group of memory cells, typically along a row or word line, that is programmed together as a unit. It is possible to program non overlapping portions of a page individually over multiple programming passes. However, owing to not all the cells of the page are programmed in a final pass together, it could create large difference in charges
- 25 programmed among the cells after the page is done. Thus partial-page programming
- 30

would result in more program disturb and would require a larger margin for sensing accuracy.

[0081] In the case the memory is configured as binary memory, the margin of operation is wider than that of MLC. In the preferred embodiment, the binary
5 memory is configured to support partial-page programming in which non-overlapping portions of a page may be programmed individually in one of the multiple programming passes on the page. The programming and reading performance can be improved by operating with a page of large size. However, when the page size is much larger than the host's unit of write (typically a 512-byte sector), its usage will be
10 inefficient. Operating with finer granularity than a page allows more efficient usage of such a page.

[0082] The example given has been between binary versus MLC. It should be understood that in general the same principles apply between a first memory with a first number of levels and a second memory with a second number of levels more than
15 the first memory.

LOGICAL AND PHYSICAL BLOCK STRUCTURES

[0083] FIG. 8 illustrates the memory being managed by a memory manager with is a software component that resides in the controller. The memory 200 is organized into blocks, each block of cells being a minimum unit of erase. Depending on
20 implementation, the memory system may operate with even large units of erase formed by an aggregate of blocks into "metablocks" and also "megablocks". For convenience the description will refer to a unit of erase as a metablock although it will be understood that some systems operate with even larger unit of erase such as a "megablock" formed by an aggregate of metablocks.

[0084] The host 80 accesses the memory 200 when running an application under a file system or operating system. Typically, the host system addresses data in units of logical sectors where, for example, each sector may contain 512 bytes of data. Also, it is usual for the host to read or write to the memory system in unit of logical clusters, each consisting of one or more logical sectors. In some host systems, an optional
25
30 host-side memory manager may exist to perform lower level memory management at

the host. In most cases during read or write operations, the host 80 essentially issues a command to the memory system 90 to read or write a segment containing a string of logical sectors of data with contiguous addresses.

[0085] A memory-side memory manager 300 is implemented in the controller 100 of the memory system 90 to manage the storage and retrieval of the data of host logical sectors among metablocks of the flash memory 200. The memory manager comprises a front-end system 310 and a back-end system 320. The front-end system 310 includes a host interface 312. The back-end system 320 includes a number of software modules for managing erase, read and write operations of the metablocks. The memory manager also maintains system control data and directory data associated with its operations among the flash memory 200 and the controller RAM 130.

[0086] FIG. 9 illustrates the software modules of the back-end system. The Back-End System mainly comprises two functional modules: a Media Management Layer 330 and a Dataflow and Sequencing Layer 340.

[0087] The media management layer 330 is responsible for the organization of logical data storage within a flash memory meta-block structure. More details will be provided later in the section on “Media management Layer”.

[0088] The dataflow and sequencing layer 340 is responsible for the sequencing and transfer of sectors of data between a front-end system and a flash memory. This layer includes a command sequencer 342, a low-level sequencer 344 and a flash Control layer 346. More details will be provided later in the section on “Low Level System Spec”.

[0089] The memory manager 300 is preferably implemented in the controller 100. It translates logical addresses received from the host into physical addresses within the memory array, where the data are actually stored, and then keeps track of these address translations.

[0090] FIGs. 10A(i) – 10A(iii) illustrate schematically the mapping between a logical group and a metablock. The metablock of the physical memory has N physical

sectors for storing N logical sectors of data of a logical group. **FIG. 10A(i)** shows the data from a logical group LG_i , where the logical sectors are in contiguous logical order 0, 1, ..., $N-1$. **FIG. 10A(ii)** shows the same data being stored in the metablock in the same logical order. The metablock when stored in this manner is said to be
5 “sequential.” In general, the metablock may have data stored in a different order, in which case the metablock is said to be “non-sequential” or “chaotic.”

[0091] There may be an offset between the lowest address of a logical group and the lowest address of the metablock to which it is mapped. In this case, logical sector address wraps round as a loop from bottom back to top of the logical group within the
10 metablock. For example, in **FIG. 10A(iii)**, the metablock stores in its first location beginning with the data of logical sector k . When the last logical sector $N-1$ is reached, it wraps around to sector 0 and finally storing data associated with logical sector $k-1$ in its last physical sector. In the preferred embodiment, a page tag is used to identify any offset, such as identifying the starting logical sector address of the data
15 stored in the first physical sector of the metablock. Two blocks will be considered to have their logical sectors stored in similar order when they only differ by a page tag.

[0092] **FIG. 10B** illustrates schematically the mapping between logical groups and metablocks. Each logical group 380 is mapped to a unique metablock 370, except for a small number of logical groups in which data is currently being updated. After a
20 logical group has been updated, it may be mapped to a different metablock. The mapping information is maintained in a set of logical to physical directories, which will be described in more detail later.

FINDING READ DISTURBS

[0093] Due to capacitive coupling between memory cells on adjacent word lines,
25 voltage levels used while accessing data on one word line can affect data quality on the neighboring word line, non-accessed word. This can occur in both write operations, leading to “write disturbs”, and in read operation, leading to “read disturbs”. There are a number of techniques are known for dealing with program disturbs, including altering voltage ramp rates (see, for example, US patent number
30 6,717,851), altering the order in which word lines are written (see, for example, US patent number 7,986,554), using foggy-fine algorithms (discussed above), or various

post-write read verifications (see, for example, US patents and patent publication numbers 8,634,240 and 2013-0028021). This section considers techniques for determining word lines more likely to be suffering from a read disturbs, where reads to a localized zone of the memory can cause undetected disturbances in neighboring word lines within a block.

[0094] A number of approaches can be used to address this problem. One is to use block read counters in RAM on the controller, which can give a fair approximation when a block may become disturbed by reads.. This solution consumes a pretty fair amount of memory space (at least 4 bytes per block times the number of blocks in the system). In an SSD with 256K blocks, this can equate to about a 1MB of space to track simple read counters, which is a significant amount of relatively expensive controller RAM. Even when using this amount of RAM, a major problem with block counters is that they do not provide the granularity to know which word lines caused the problem and that the threshold is set extremely low. When the threshold for the block trips, the whole block is scrubbed (i.e. copied), which leads to premature cycling. With the scheme presented here, the system can detect specifically which word line was read, moves that data and neighbor data, but can leave the rest of the block intact.

[0095] Another approach is a read patrol (see US patents and patent publication numbers 7,012,835; 7,477,547; 7,616,484; and 2009/0172258), either randomly or serially sampling written blocks with valid data to look for read disturbs. This method uses extra background power and may not find hot zones in time (as this is dependent upon the amount of background time) or at all. The read patrol takes a significant amount of time to get coverage of the full device. Also, the read patrol itself contributes to the read disturb, the problem it is trying to solve. Reading neighboring word lines periodically for host reads can help to detect potential read disturbs, but adds additional power and latency to host operations. Additionally, reading neighbor word lines is subject to noise error and becomes cumbersome in trying to detect and identify hot read data addresses.

[0096] The exemplary embodiments presented here use a hash tree to track read zones and find hot read areas of the flash. The hash tree uses a small footprint of memory

area and can provide more accuracy than a simple block counter, depending on the tree's design level. The hash tree can help isolate regions of the flash that the host is reading. Depending on the embodiment, the tree can be designed to be tall or short, with more branches or more singular, the tree can be symmetric or asymmetrical.

5 [0097] Initially, when the tree starts, only one level is used, level 0. Level 0 can be a single set of N 4-byte counters. As a zone is hit by reads, the read counter is incremented. When the zone reaches a certain threshold, the zone is elevated to level 1. When the zone is elevated to level 1, its space is broken into further zones in level 1. This provides more accuracy for detecting the hot space of a zone. A pointer can
10 be used to indicate which zones have been elevated. After a zone is elevated, its root can still increment to help keep track of its read count. The threshold to elevate can be modified as the tree fills up.

[0098] The elevation of the zones can continue up the tree, where each level hash can have one or more children nodes. Because space is limited in an embedded device,
15 when a different zone in a lower node becomes hotter, an eviction should take place at some point so that the tree cannot grow without bounds. When an eviction happens, the zone is collapsed down and a new zone is elevated in the tree for more analysis. During an eviction the zone read count is still preserved, but the tree information is lost.

20 [0099] The number of levels will depend on the implementation. As zones reach the tree tips, and reaches the highest threshold, the zone can placed on a read patrol list for further evaluation for read scrub. More detail on scrub, refresh, and patrol operations can be found in US patents and patent publication numbers 7,012,835; 7,477,547; 7,616,484; and 2009/0172258.

25 [0100] To maintain the tree structure and level counts, periodically the tree and its parameters can be saved to the non-volatile memory on the memory section. Then on power up the table can be loaded from the non-volatile memory back to controller RAM.

[0101] To avoid counter saturation, the tree can be pruned back by some amount
30 periodically. The pruning will occur with block erases, hot read data migration, and,

alternatively, when certain hot count (number of erases) milestones are reached. This process can be referred to as a “tree leveling”, where when the program/erase counter reaches a certain level, the whole tree can be pared back. If the device is evenly wear leveled, then all the blocks in the system shall vary in hot count by no more than 5 percentage variation. (More detail on where leveling is given in US patents and patent publication numbers 7,353,325; 7,441,067; 7,120,729; and 2010-0174845.) Using this knowledge, certain hot count checkpoints can be established to prune down the tree. The pruning of the tree can be a global reset, or pairing the counters by a percentage, or pairing the tree by a fixed amount. The tree can be pruned whenever 10 the stem threshold is crossed (causing a block, word line, logical group, or other structure corresponding to that level’s zone to reach the scrub list).

[0102] When an area is detected hot, the hot area (and its neighboring word lines) can either be refreshed to a new block (leaving the rest of the block intact), or the whole block can be scrubbed. The hot data can either be refreshed to a separate zone, so as 15 to not trip the tree up as the reads continue, or it can just migrate naturally. If migrating naturally then nothing special needs to be done in the event that a hot read zone becomes cold. The separate zone for hot data can be managed in various ways to reduce the resultant amount read disturbs. For example, the data can be stored in binary format, use different margins, be stored with one or more unused word lines 20 between the word lines storing data, and so on. For any of these arrangements, the detection of potential read disturbs can lead to the eviction of a partial block of data, including the immediate neighbor word lines, or recycling of the entire block with possible special handling of the hot read data. If the hot read data cools later, it can rejoin the regular pool.

25 [0103] In one set of embodiments, after a zone is detected hot, and the neighbors are checked and scrubbed, the hot data can be tracked logically using a separate table. Tracking this data separately can provide the advantage of detecting when the data becomes cold, and can help from having the same data trigger branching in the tree after scrubbing. The hot data would still need to be tracked for future read disturbs.

30 [0104] If the tree is used to track the device physically and an erase occurs within a zone, the read counters can be rolled back by a defined algorithm amount to account

for the fact that a portion of that zone has been refreshed. Depending on the embodiment and the level, a branch of the tree can represent a logical group (4KB, for example), a die's WL, a series of WLs, or a series of blocks. As erases occur on units of blocks, the branch that represents that block would need to collapse down to the block level if necessary. At the time of collapse, a new branch can be elevated or the elevation can occur on the next read to that set.

[0105] As read disturbs are the result of operations on physically adjacent word lines, the exemplary embodiments are based on tracking physical addresses of the zones at each level. An alternative embodiment is to have the tree track the addresses logically instead of physically, with any writes to a zone/branch should have a decrementing effect on the counters.

[0106] Some of these concepts are illustrated with respect **FIGs. 11-13**. **FIG. 11** is a schematic representation of the process for a two level example. Generally, level 0 zones can be die, portion of a die, number of blocks, and so on down to a group of word lines. The zones at the highest level (level 1 in this example) can be as small as individual word lines. (In this discussion, "zone" is not meant, or at least not necessarily meant, to correspond to the sort of zone structure described in US patent number 6,901,498, for example, but is being used more generally for physical memory regions such as die, block, word line and soon.) In the example of **FIG. 11** the level 0 zone is taken as some number of blocks and the level 1 zone is taken as some fraction of the blocks of the level 0 zone.

[0107] At lower part of **FIG. 11** are the level 0 counters 501, each of the read counters corresponding to one or zones 0 to N. For example, counter 503 tracks the number of times the blocks 505 of zone 0 are accessed for a read. In this example, Zone 0's blocks are split into level 1 zones for further analysis, there can be several level 1 counters. As shown to the left, when a counter reaches a threshold for the level the corresponding zone's block are moved up a level for further evaluation. When a zone moves up in levels, the count can either be evenly distributed to its subparts, reset to 0, or a separate master zone counter maintained.

[0108] At level 1, the elevated level 0 zone is sub-divided into a set of level 1 zones with corresponding counters 511: for example, Zone A of blocks 515 is monitored

using counter 513. The counters of level 1 are checked against the level's threshold and moved up as needed to the level or on a list for read scan, where, depending on the implementation, the system can repeat the narrowing process to get down to a group of blocks, a group of WLs, or even to the smallest read/write granularity as based on the system's ECC unit (referred to variously as a flash management unit, FMU, a codeword, or ECC page). When a block is erased, and the block is at a higher level (such as level 1 in **FIG. 11**), the zone can be shifted down a level. The collapsing of elements can be based on the comparison of other counters, block erasures, or as part of a tree levelling procedure. A collapsing of one branch can cause division in another branch. The root nodes can be evaluated on a tree collapse, or if no free zone counters are available, then the whole tree could be re-evaluated on any collapse, where priority can be given to the zone with the highest counter or the most recent counter.

[0109] From level 1, a zone Zone X with counter 521 is moved to the candidate list for read scan to check for read disturb effects. If this zone is narrowed down to a word line or series of word lines, these can be placed on a "burning zone" list for frequently accessed zones.

[0110] **FIG. 12** looks the situation in more detail. At bottom is the set of zone counters for Level 0, where Zones 0-N can represent the entire plane, die, chip, or device capacity. Based on the count values, as these reach the corresponding thresholds the counters of the set can be partitioned into equal sized zones for level 1. **FIG. 12** shows the case when Q zones (two of which are shown) from level 0 have been elevated to level 1, each subdivided into X zones with a corresponding counter. For example, Zone 0 from level 0 is here broken into multiple smaller zones Zone A to Zone A+X at level 1. The counters for the zones of level 1 are similarly used to determine when to move a zone of level 1 to level. For example, Zone A+1 is here broken into multiple (X again in this example) smaller zones in level 2. If, say, Zone 'E+1' then has a high count value, it can then be sent to the read scan to check for read disturb effects.

[0111] **FIG. 13** is similar to **FIG. 12**, but for a slightly different scenario. **FIG. 13** shows two of the level 0 zones (zones 0 and N) each subdivided into a set of level 1

zones, each of which then has two zones (zones A+1 and A+X, zones B and B+1) further split into sets of level 2 zones. A symmetrical hash tree may be easier for firmware/hardware to maintain, although asymmetrical trees can be used as well.

- 5 [0112] For any of the embodiments, the techniques of this section can help to determine locations of possible read disturbs with fewer reads to the device and less use of power. Hot read zones can be found accurately with a smaller memory footprint than in other counter solutions. This can help to avoid un-necessary scrubbing of data and blind background patrolling.

Conclusion

- 10 [0113] The foregoing detailed description of the invention has been presented for purposes of illustration and description. It is not intended to be exhaustive or to limit the invention to the precise form disclosed. Many modifications and variations are possible in light of the above teaching. The described embodiments were chosen in order to best explain the principles of the invention and its practical application, to
15 thereby enable others skilled in the art to best utilize the invention in various embodiments and with various modifications as are suited to the particular use contemplated. It is intended that the scope of the invention be defined by the claims appended hereto.

IT IS CLAIMED:

1. A method of operating a non-volatile memory system having one or more memory circuits and a controller circuit, the memory circuits each including one or more arrays of non-volatile memory cells formed along word lines and the controller circuit managing the storage of data on the memory circuit, the method comprising:
- 5 for each of a first plurality of distinct divisions of the one or more memory arrays maintaining by the controller circuit a count of the number of times word lines of the corresponding division is accessed for a read operation;
- 10 in response to one of the counts for a corresponding division reaching a first threshold value, subdividing the corresponding division into a second plurality of distinct first subdivisions; and
- 15 subsequently maintaining by the controller circuit for each of the first subdivisions a count of the number of times word lines of the subdivision is accessed for read operations.
2. The method of claim 1, further comprising:
- 20 in response to one of the counts for a corresponding subdivision reaching a second threshold value, further subdividing the corresponding subdivision into a third plurality of distinct first sub-subdivisions; and
- 25 subsequently maintaining by the controller circuit for each of the first sub-subdivisions a count of the number of times word lines of the sub subdivision is accessed for read operations.
3. The method of claim 1, wherein the first plurality of distinct divisions were previously formed by subdividing one of a plurality of structures formed of such divisions in response to a number of accesses for read operations.
4. The method of claim 1, further comprising:
- 30 in response to one of the counts for a corresponding subdivision reaching a second threshold value, placing the corresponding one of the subdivisions a list of subdivisions on which to perform a data scrub operation.
5. The method of claim 1, further comprising:

in response to one of the counts for a corresponding subdivision reaching a second threshold value, performing a scrub operation on the corresponding one of the subdivisions.

5 6. The method of claim 1, further comprising:

in response to one of the counts for a corresponding subdivision reaching a second threshold value, relocating the data from the corresponding one of the subdivisions to a different location on the memory circuits.

10 7. The method of claim 6, wherein the different location on the memory circuits is reserved by the controller circuit for frequently accessed data.

8. The method of claim 7, wherein the location reserved by the controller circuit for frequently accessed data is managed differently by the controller circuit
15 than other portions of the memory circuits.

9. The method of claim 8, wherein data is stored in a binary format in the location reserved for frequently accessed data.

20 10. The method of claim 8, wherein data is stored on non-adjacent word lines in the location reserved for frequently accessed data.

11. The method of claim 7, wherein the relocated data is subsequently evicted from the location reserved for frequently accessed data in response to determining that
25 the frequency of being accessed for the relocated data is decreased.

12. The method of claim 6, wherein the data relocated includes one or more frequently accessed word lines and one or more word lines adjacent the frequently accessed word lines.

30

13. The method of claim 6, wherein the relocating the data from the corresponding one of the subdivisions includes the relocation of an entire erase block.

14. The method of claim 1, wherein the controller circuit includes a volatile memory in which are maintained the values of the counts.

15. The method of claim 14, wherein the controller circuit periodically saves
5 the count values to non-volatile memory on the memory circuits.

16. The method of claim 15, wherein on power up, the controller circuit loads the count values saved in non-volatile memory into the volatile memory on the controller circuit.
10

17. The method of claim 1, wherein the memory circuits are flash memory circuit and the divisions correspond to a plurality of erase blocks.

18. The method of claim 1, wherein the memory circuits are flash memory
15 circuit and the subdivisions correspond to a plurality of erase blocks.

19. The method of claim 1, wherein the memory circuits are flash memory circuit and the divisions correspond to an erase blocks.

20. The method of claim 1, further comprising:
20 subsequently discontinuing of the maintaining by the controller circuit for one or more of the first subdivisions the count of the number of times word lines of the subdivision is accessed for read operations.

21. The method of claim 20, wherein the discontinuing is in response to the
25 value of the counts being discontinued relative to others of the counts of the first subdivision.

22. The method of claim 20, wherein the discontinuing is in response to the
30 corresponding division being erased.

23. The method of claim 20, wherein the discontinuing is in response to the number of counts being maintained.

24. The method of claim 1, wherein the memory circuits are flash memory circuit and the subdivisions correspond to an erase blocks.

5 25. The method of claim 1, wherein the divisions correspond to a set of word lines.

26. The method of claim 1, wherein the subdivisions correspond to a set of word lines.

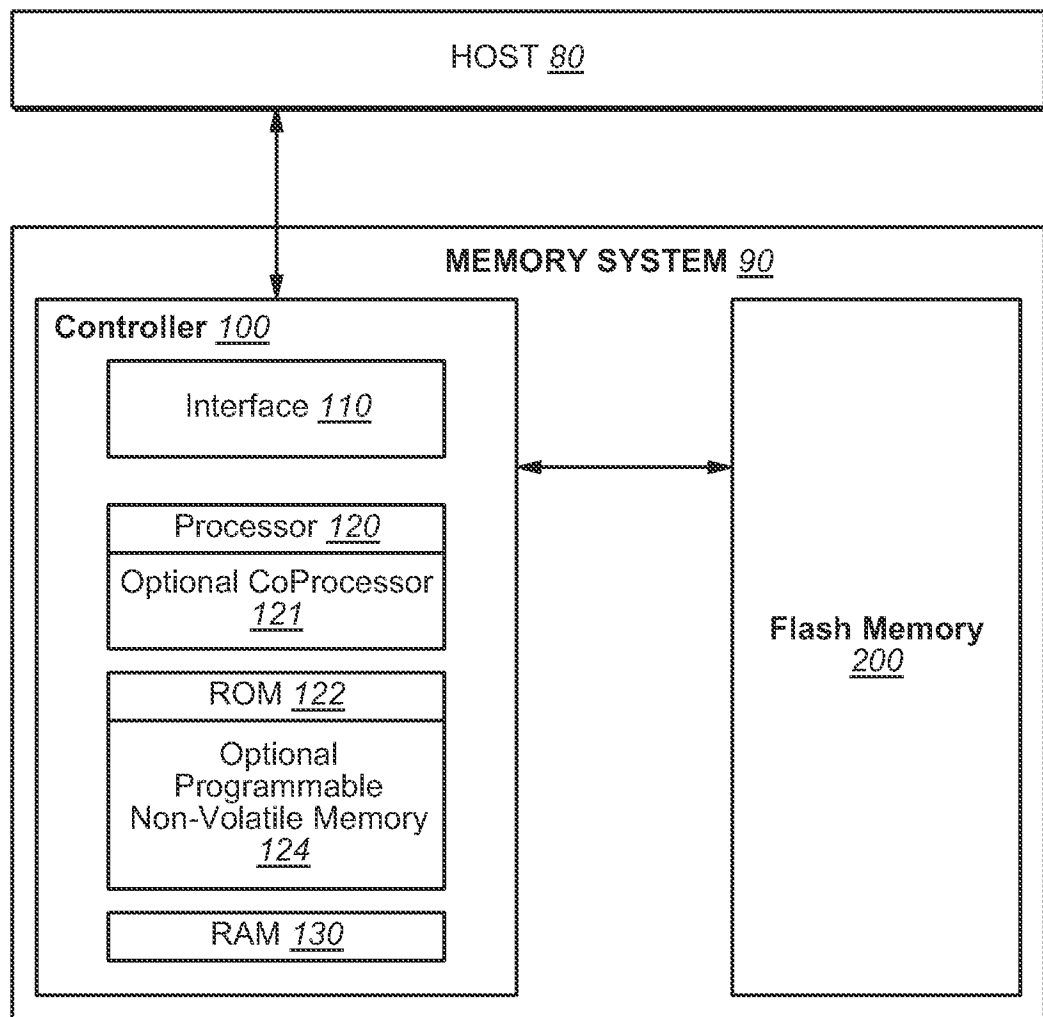
10

27. The method of claim 1, wherein the memory circuits include a plurality of dies and the divisions correspond to a die.

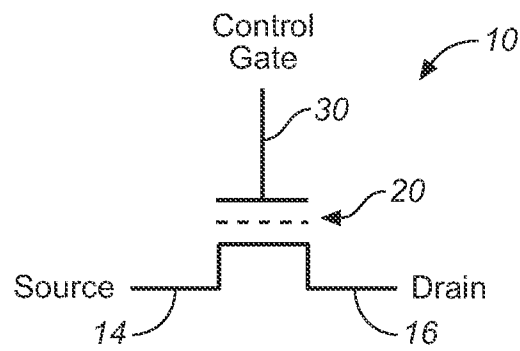
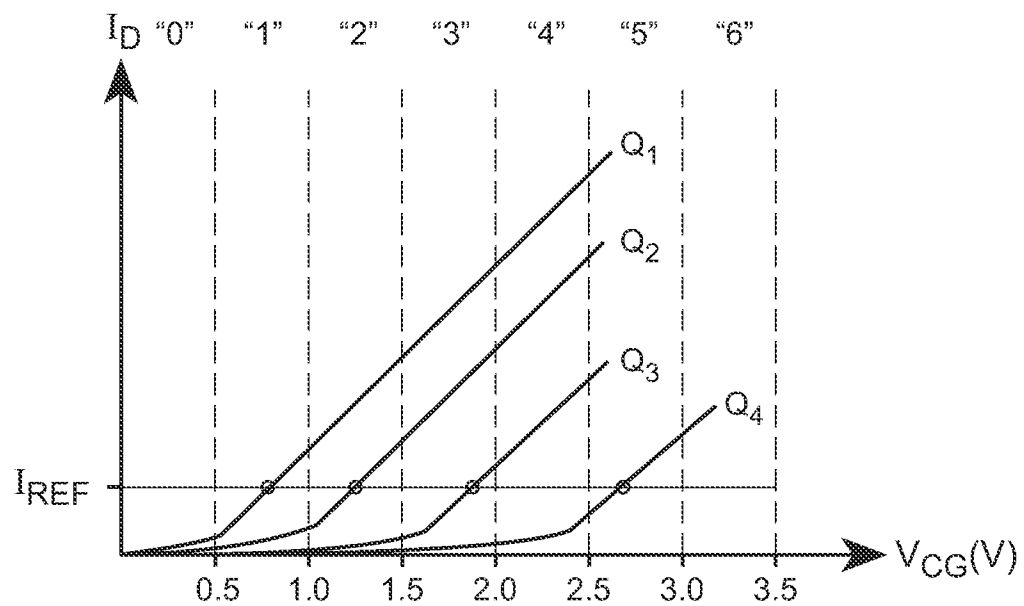
15 28. The method of claim 1, wherein the arrays are of a NAND-type of architecture.

29. The memory of claim 1, wherein the arrays are of an architecture having a three dimensional memory array that is monolithically formed as a plurality of physical levels of memory cells, a physical level having an active layer disposed
20 above a silicon substrate, the memory cells in communication with operating circuitry.

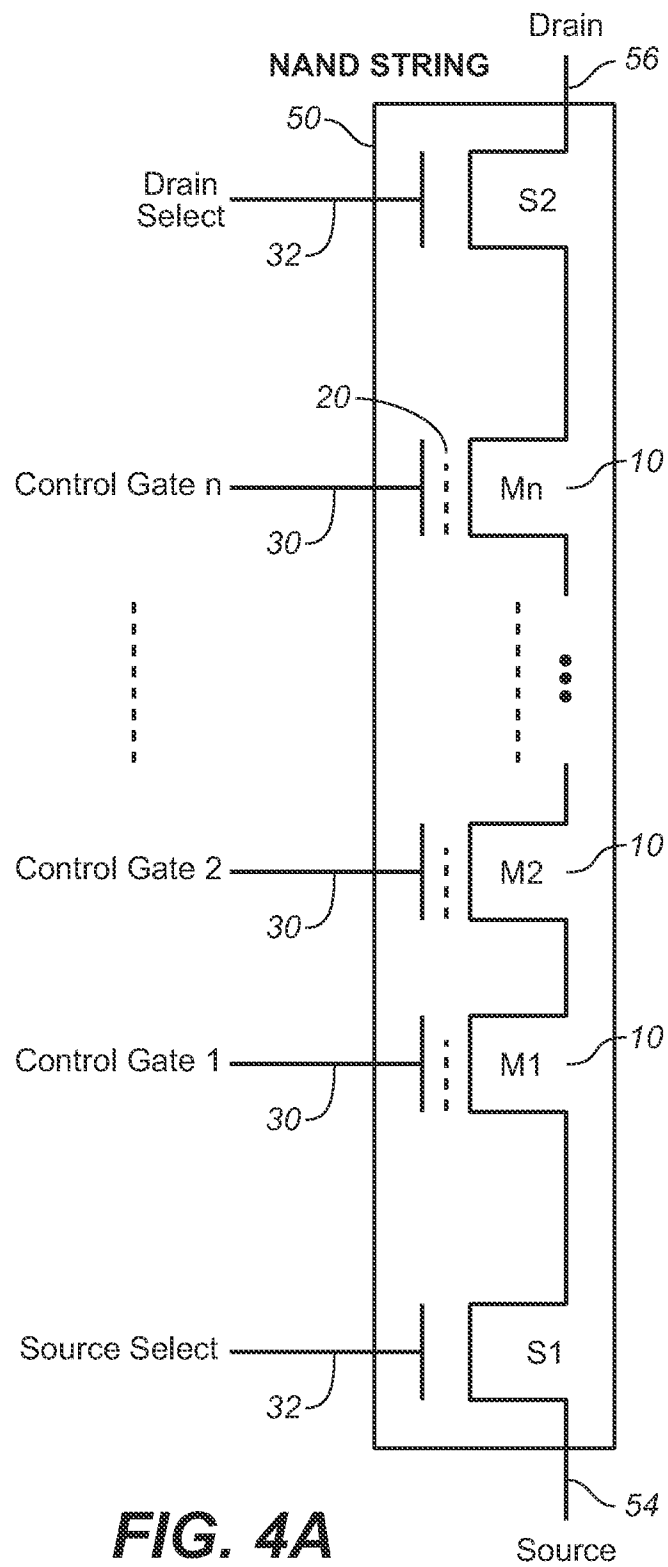
1 / 15

**FIG. 1**

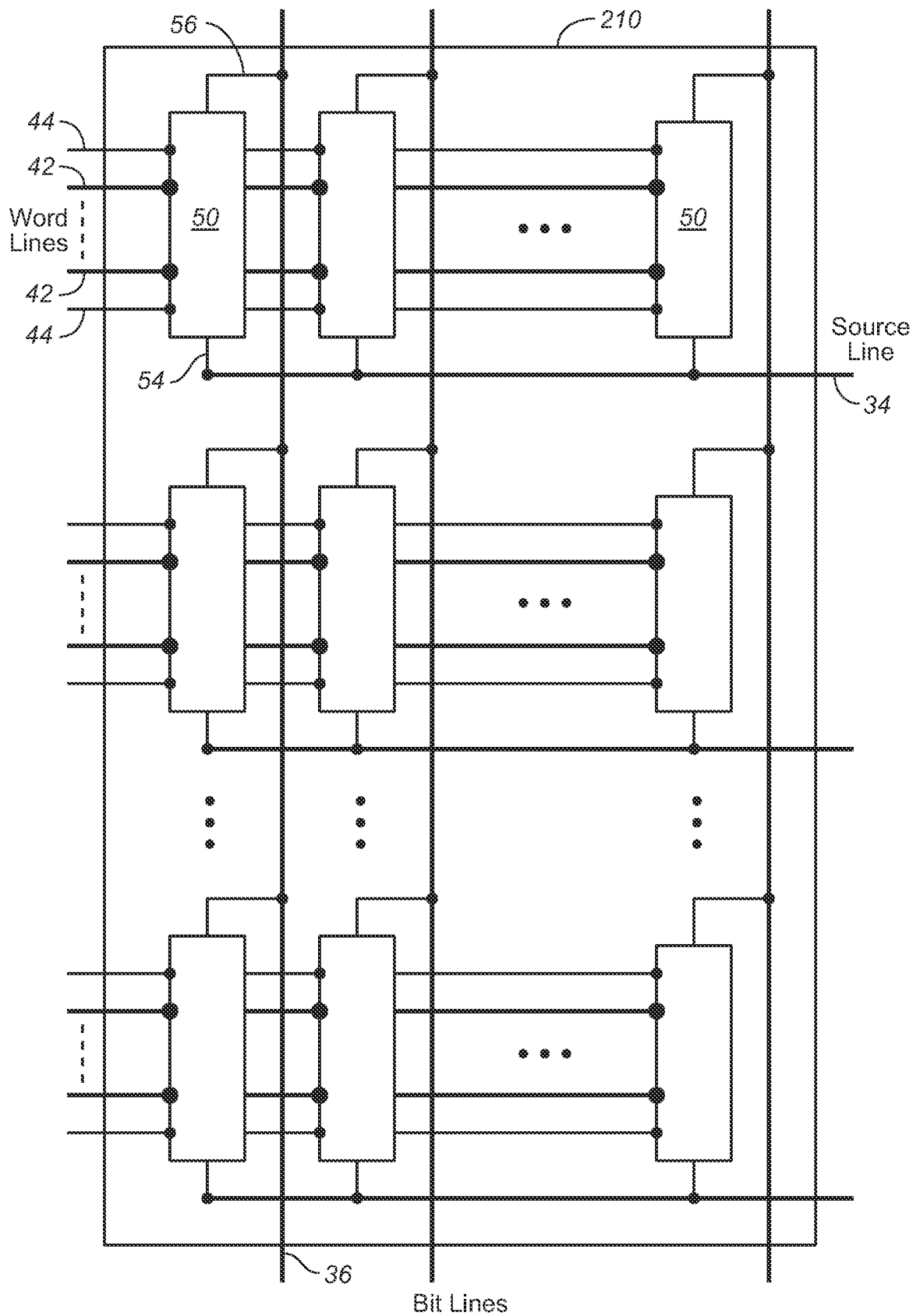
2 / 15

**FIG. 2****FIG. 3**

3 / 15



4 / 15

**FIG. 4B**

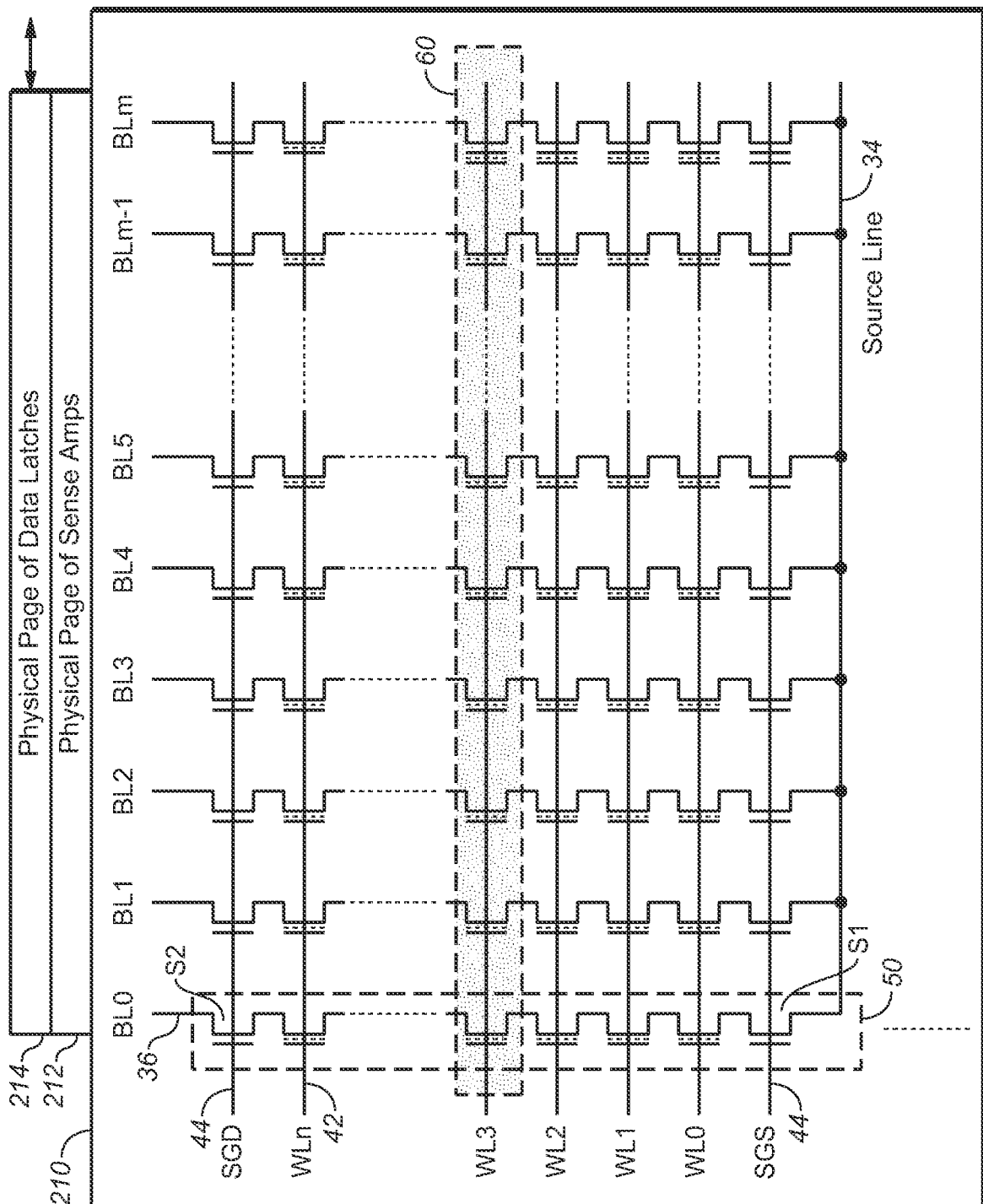
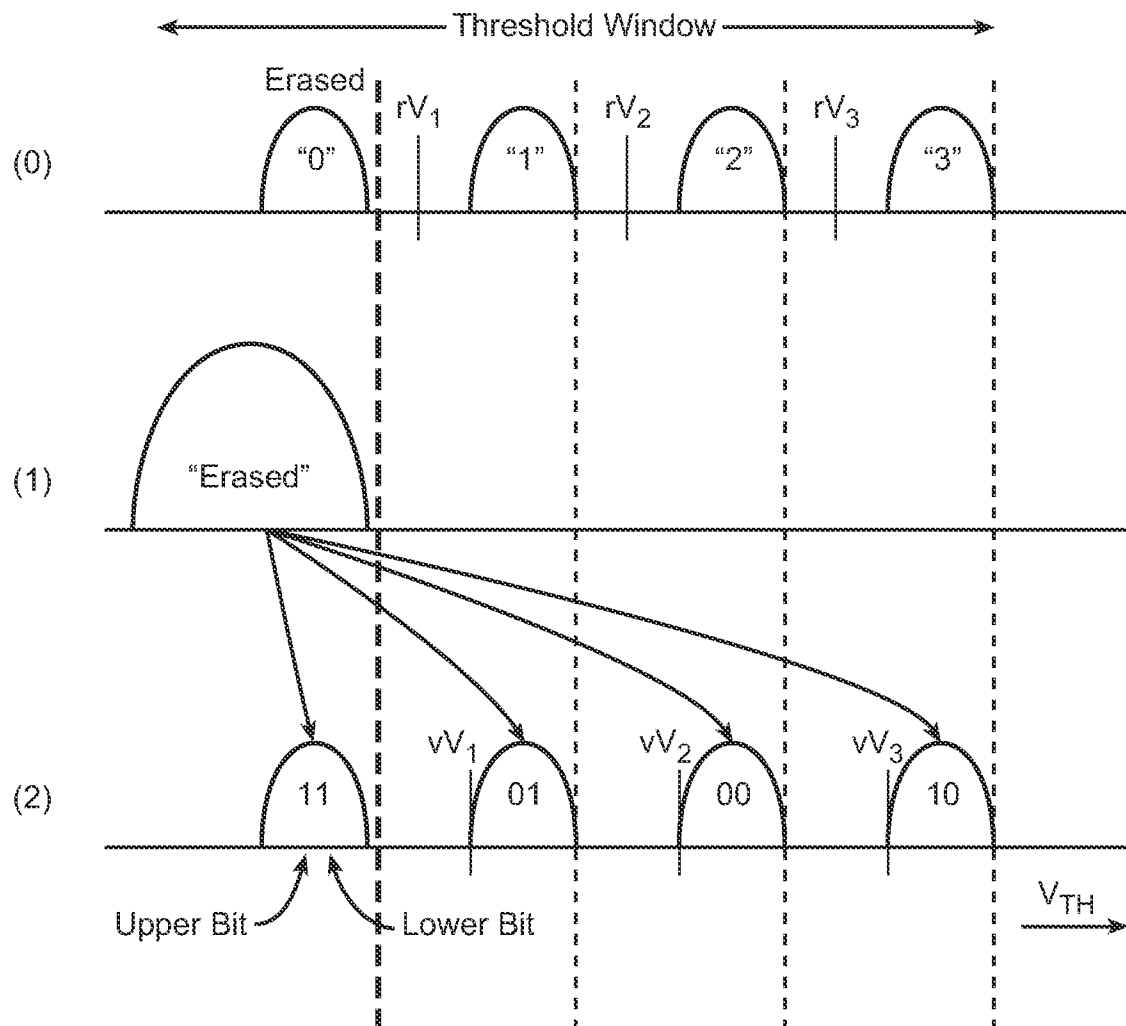


Fig. 5

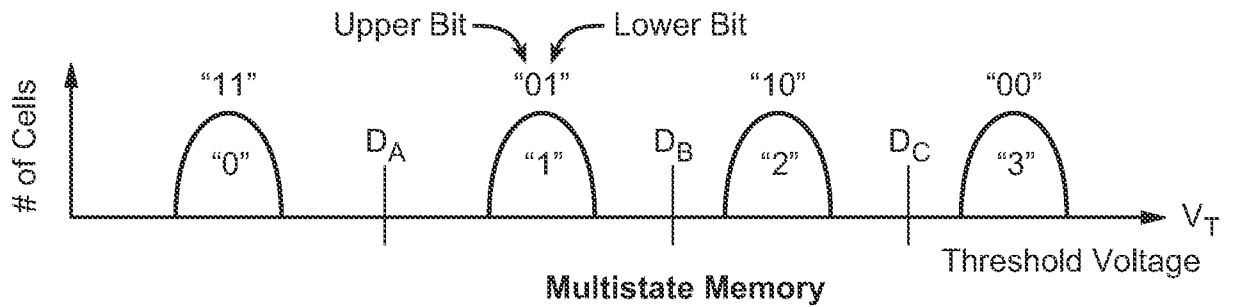
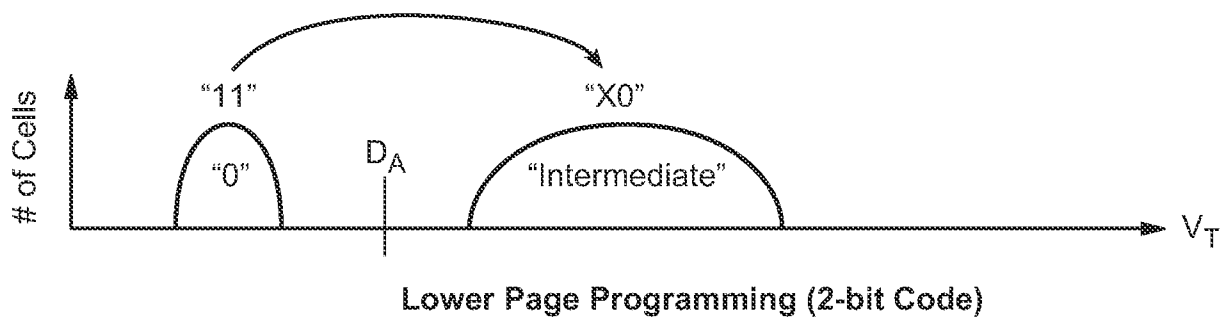
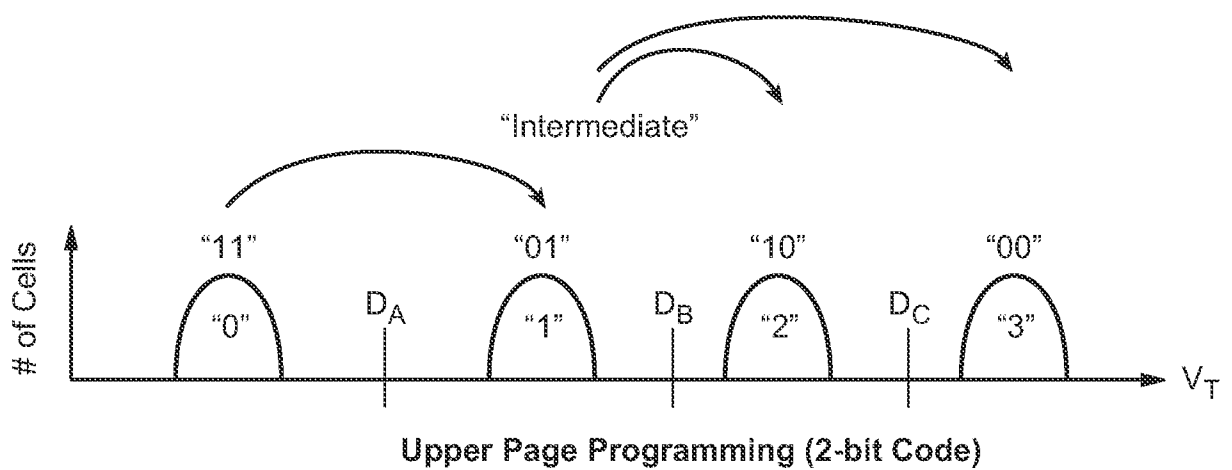
6 / 15



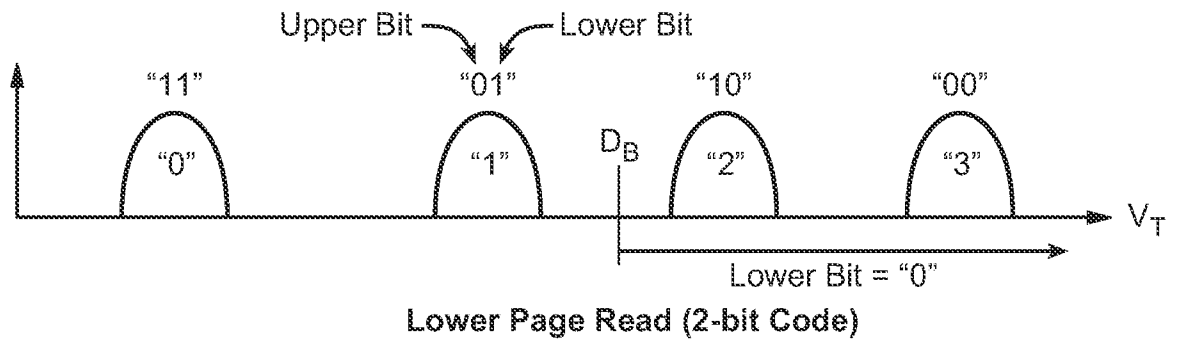
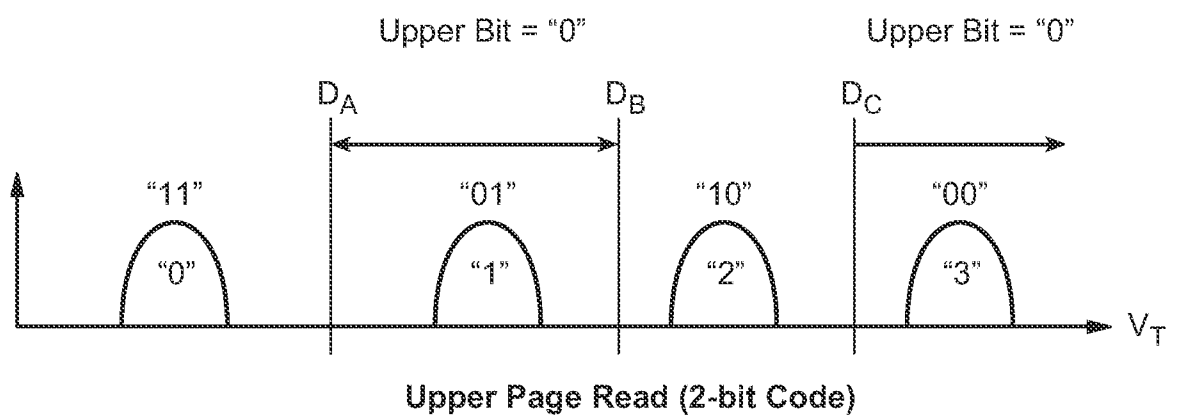
Programming into Four States Represented by a 2-bit Code

FIG. 6

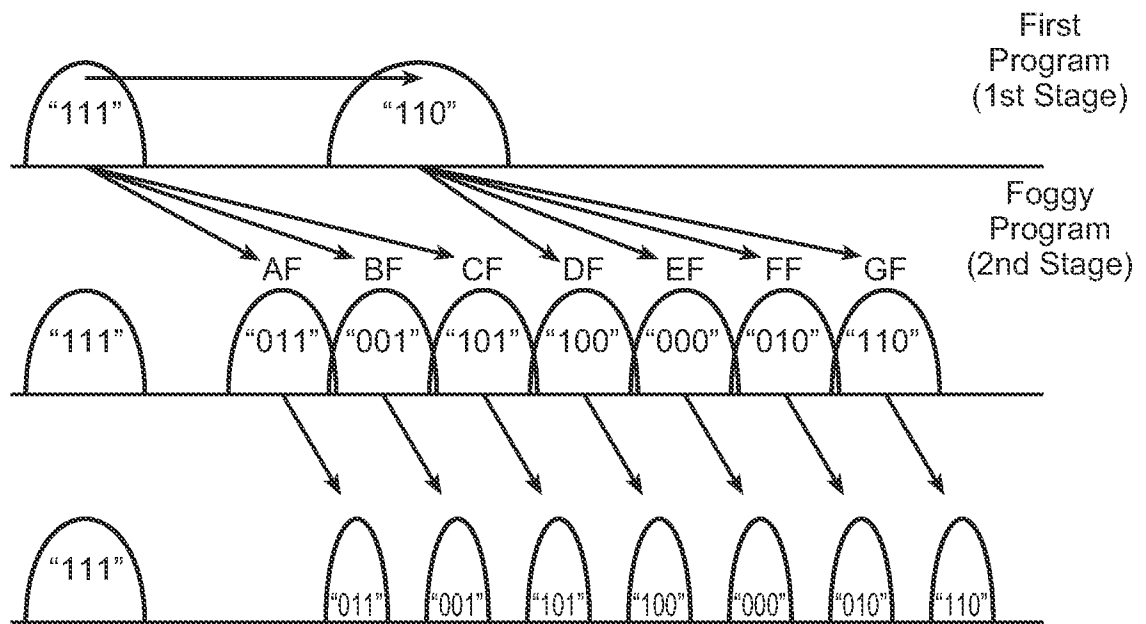
7 / 15

**FIG. 7A****FIG. 7B****FIG. 7C**

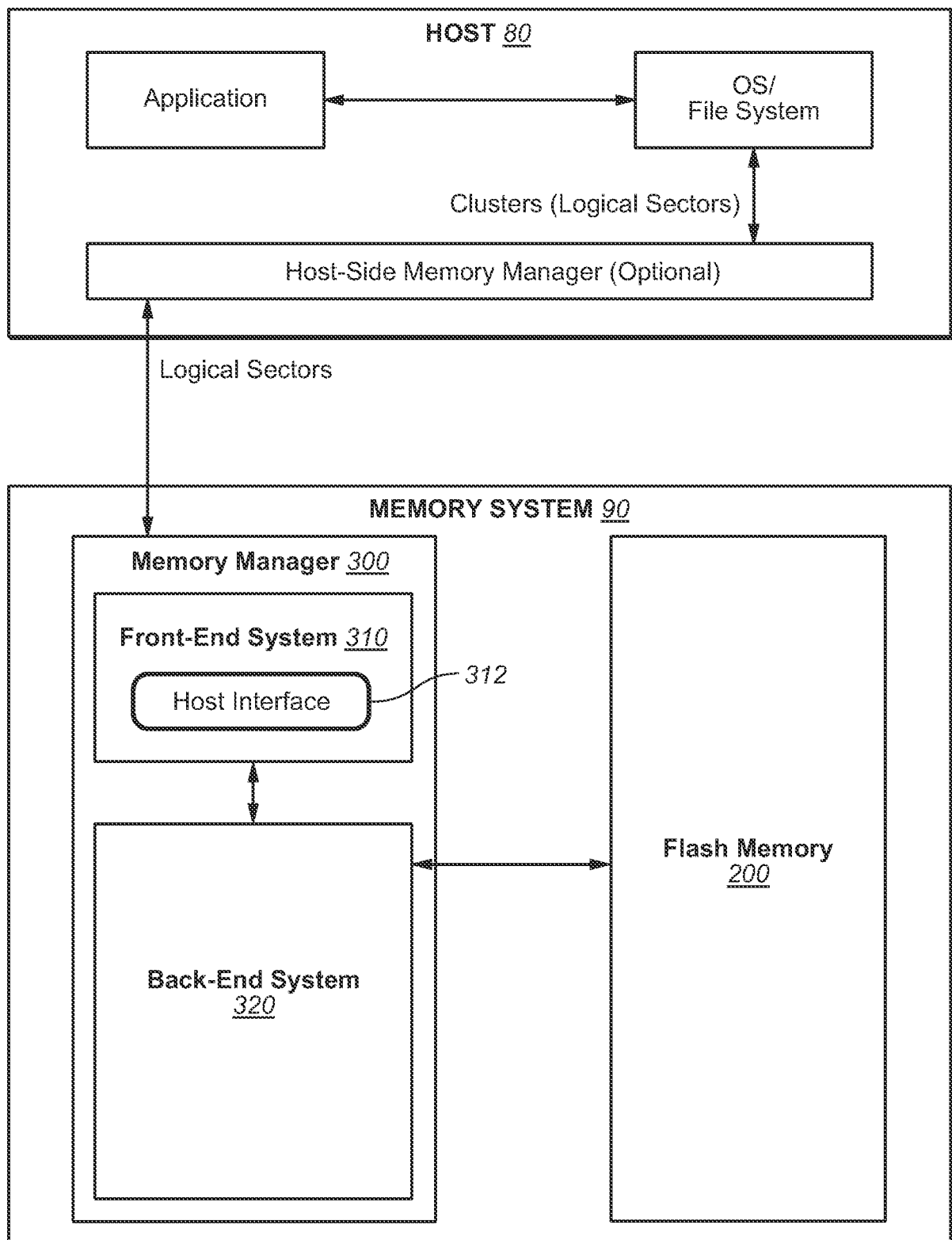
8 / 15

**FIG. 7D****FIG. 7E**

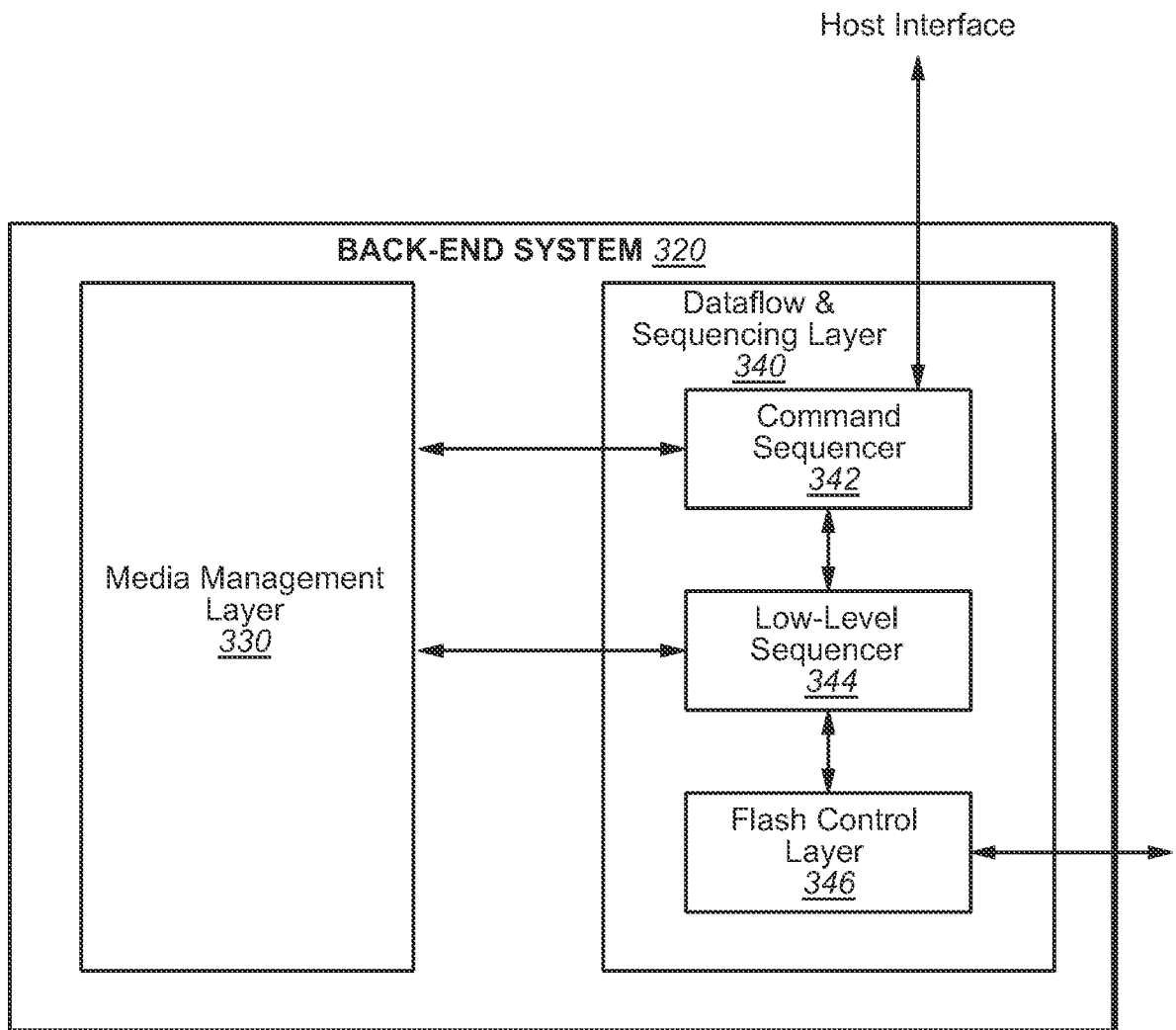
9 / 15

**FIG. 7F**

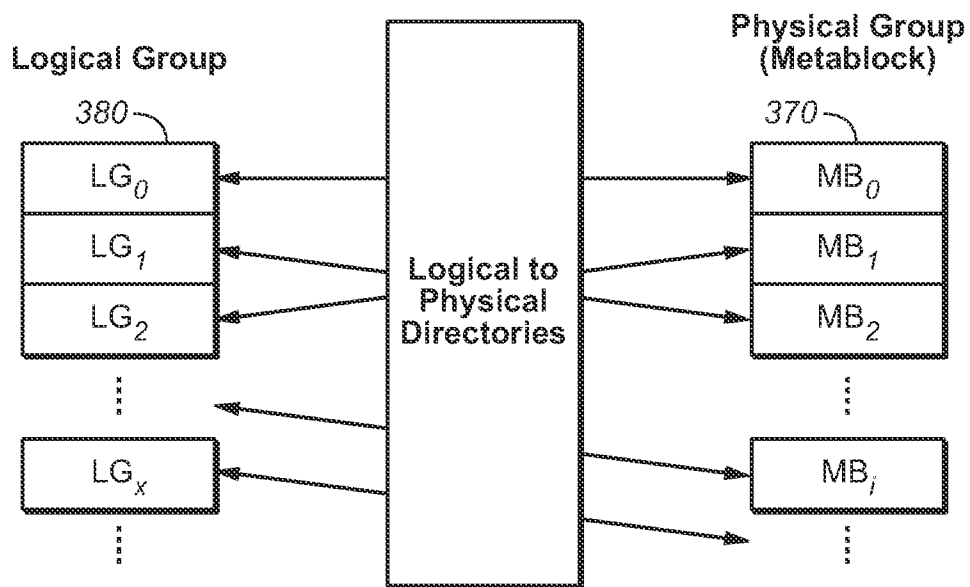
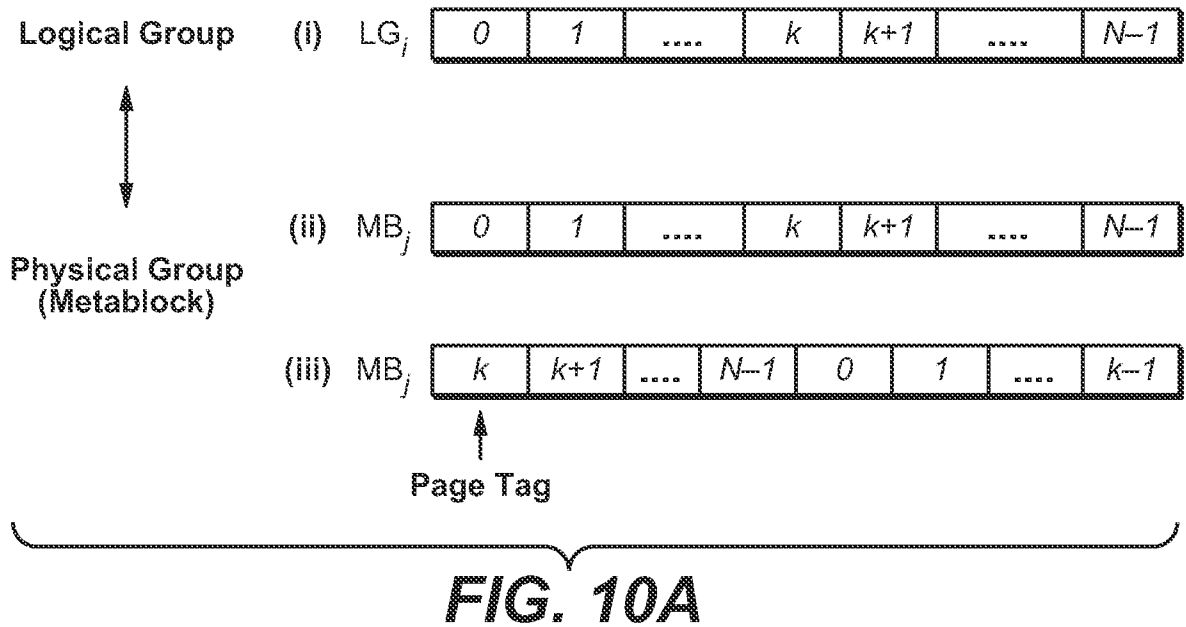
10 / 15

**FIG. 8**

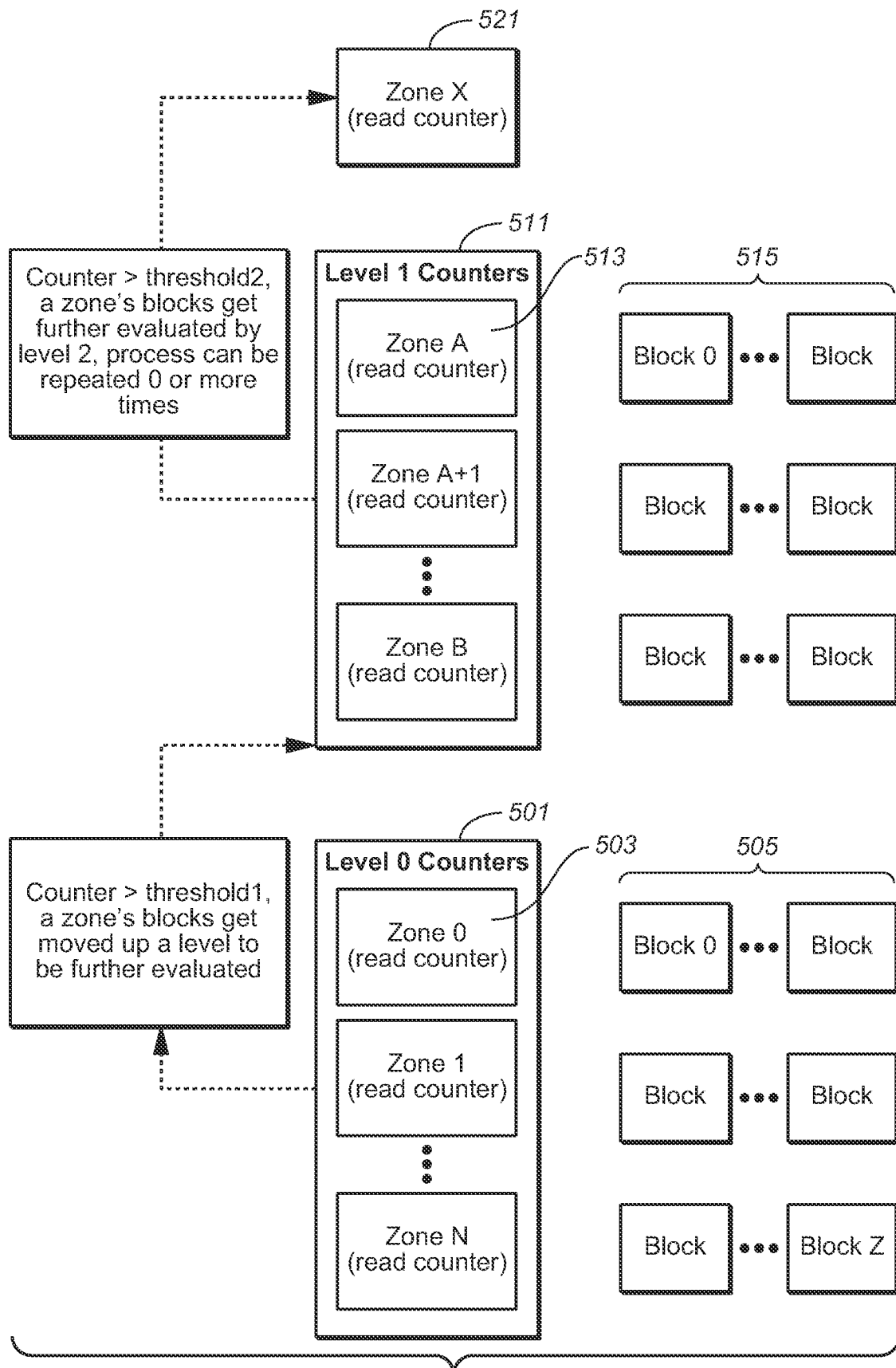
11 / 15

**FIG. 9**

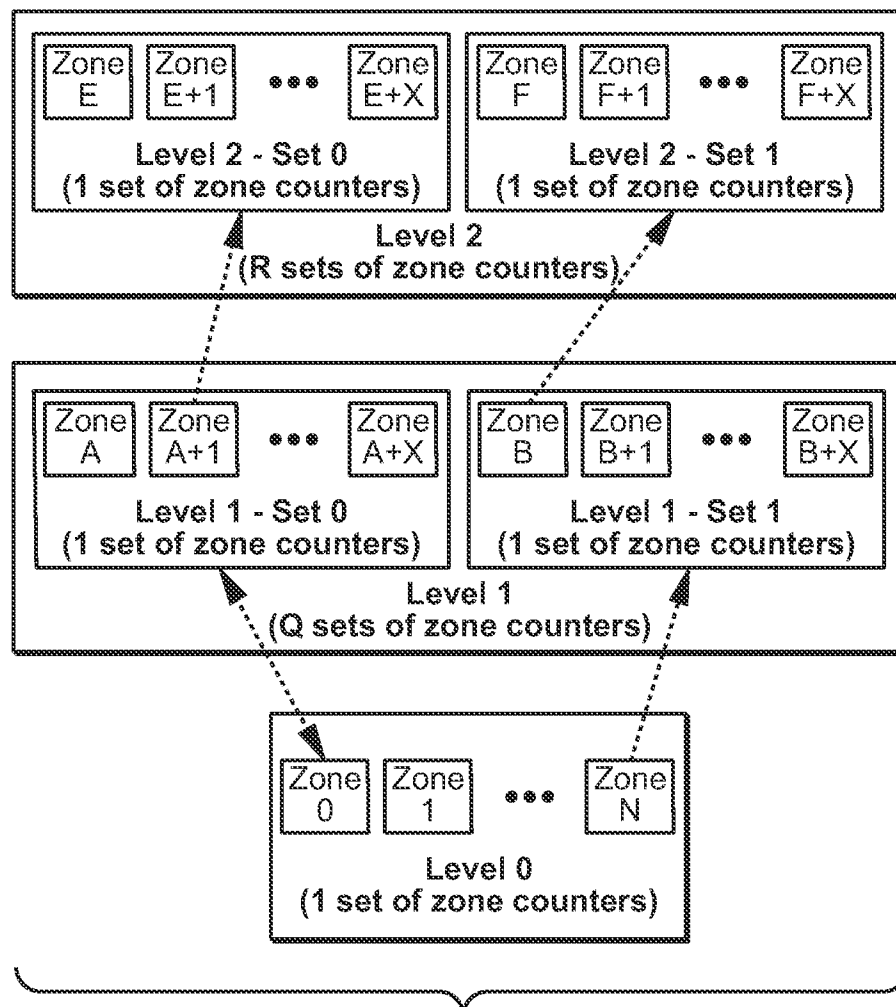
12 / 15

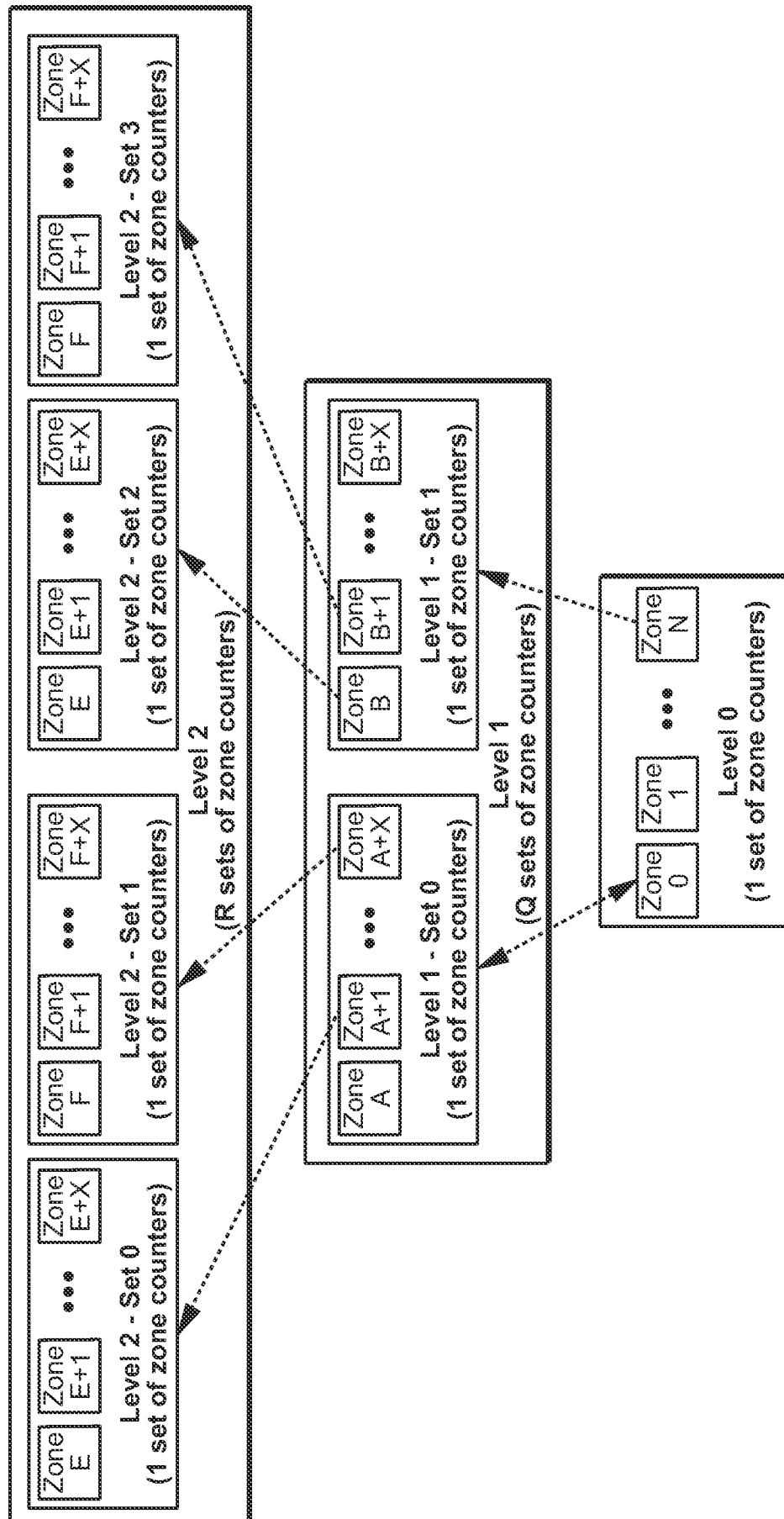
**FIG. 10B**

13 / 15

**FIG. 11**

14 / 15



**FIG. 13**

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2015/018527

A. CLASSIFICATION OF SUBJECT MATTER
INV. G11C16/34 G06F12/02
ADD.

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)
G11C G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010/217919 A1 (SUKEGAWA HIROSHI [JP] ET AL) 26 August 2010 (2010-08-26)	1,4,5, 7-9,12, 14-20, 24,27-29
A	page 3, paragraph 39; claim 1; figures 1,4	2,3,11, 21-23
Y	----- US 2010/229055 A1 (MUKHERJEE NILANJAN [US] ET AL) 9 September 2010 (2010-09-09)	1,4-10, 12-20, 24-29
A	page 2, paragraph 26; claim 1 ----- -/-	2,3,11, 21-23



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

"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

15 May 2015

Date of mailing of the international search report

22/05/2015

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Kontogiannis, K

INTERNATIONAL SEARCH REPORT

International application No

PCT/US2015/018527

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2009/034328 A1 (SEOL BONG-GWAN [KR]) 5 February 2009 (2009-02-05)	1,7-10, 12, 14-19, 24,27-29
A	claims 1,18-19,20	2,3,11, 21-23
Y	----- US 7 818 525 B1 (FROST HOLLOWAY H [US] ET AL) 19 October 2010 (2010-10-19)	1,6-10, 12,13, 17-19, 24-29
A	claim 1; figures 3C,6	2,3,11, 21-23
A	----- US 2013/128666 A1 (AVILA CHRIS [US] ET AL) 23 May 2013 (2013-05-23) page 2, paragraph 14-15	4,5
A	----- US 2013/028021 A1 (SHARON ERAN [IL] ET AL) 31 January 2013 (2013-01-31) cited in the application figures 16,33 -----	7-9

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2015/018527

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2010217919 A1	26-08-2010	JP 5185156 B2 JP 2010198219 A KR 20100097041 A US 2010217919 A1	17-04-2013 09-09-2010 02-09-2010 26-08-2010
US 2010229055 A1	09-09-2010	US 2010229055 A1 WO 2010102235 A1	09-09-2010 10-09-2010
US 2009034328 A1	05-02-2009	JP 5505922 B2 JP 2009037619 A KR 20090014036 A US 2009034328 A1 US 2010241796 A1	28-05-2014 19-02-2009 06-02-2009 05-02-2009 23-09-2010
US 7818525 B1	19-10-2010	US 7818525 B1 US 2011040932 A1 US 2012233391 A1 US 2015113341 A1	19-10-2010 17-02-2011 13-09-2012 23-04-2015
US 2013128666 A1	23-05-2013	CN 103946805 A KR 20140096059 A TW 201337559 A US 2013128666 A1 WO 2013077962 A1	23-07-2014 04-08-2014 16-09-2013 23-05-2013 30-05-2013
US 2013028021 A1	31-01-2013	EP 2737487 A1 KR 20140065393 A US 2013028021 A1 US 2014254272 A1 WO 2013016401 A1	04-06-2014 29-05-2014 31-01-2013 11-09-2014 31-01-2013