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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
6 October 2016 (06.10.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/160216 A1

(51) International Patent Classification:
G06F 3/06 (2006.01)

(21) International Application Number:
PCT/US2016/019956

(22) International Filing Date:
26 February 2016 (26.02.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
14/671,663 27 March 2015 (27.03.2015) US

(71) Applicant: **INTEL CORPORATION** [US/US]; 2200 Mission College Boulevard, Santa Clara, California 95054 (US).

(72) Inventors: **FRICKEY, Robert, E., III**; 3101 McKinley Blvd., Sacramento, California 95816 (US). **ZHANG, Ye**; 880 Zixing Road, Zizhu Science Park, Shanghai 200241 (CN).

(74) Agent: **DUTTA, Rabindranath**; Konrad Raynes Davda & Victor LLP, 350 S. Beverly Dr., Suite 360, Beverly Hills, California 90212 (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, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

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

(54) Title: MEASUREMENT AND REPORTING OF THE LATENCY OF INPUT AND OUTPUT OPERATIONS BY A SOLID STATE DRIVE TO A HOST

(57) Abstract: Provided are a method, apparatus, and a system for measuring latency of a storage device. The storage device measures one or more latencies of one or more input/output (I/O) operations received from a host. The storage device transmits information on the one or more latencies to the host.

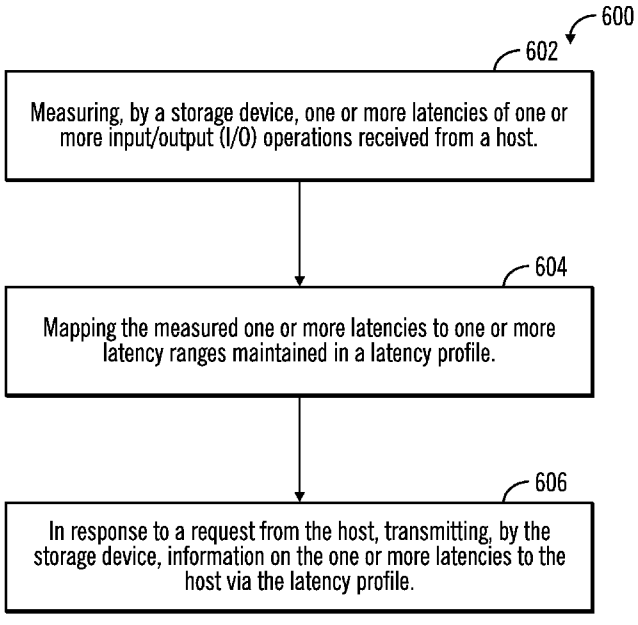


FIG. 6

WO 2016/160216 A1

MEASUREMENT AND REPORTING OF THE LATENCY OF INPUT AND OUTPUT OPERATIONS BY A SOLID STATE DRIVE TO A HOST

BACKGROUND

A solid state drive (SSD) is a data storage device that uses integrated circuit assemblies as memory to store data persistently. The SSD may be coupled to a host computing system, such as a personal computer or a server, where the host computing system performs input/output (I/O) operations on the SSD, and where the I/O operations may include writing data to the SSD and reading data from the SSD.

When an I/O operation is transmitted from the host computing system to the SSD and the results of the processing of the I/O operation are returned back from the SSD to the host computing system, the total time taken for the roundtrip is referred to as the latency of the I/O operation. In other words, the latency refers to the time interval or delay during which the host computing system is waiting for receiving the results of the I/O operation from the SSD. In certain systems, latency measurement mechanisms are implemented within the host computing system. In such mechanisms, the host computing system measures the latency of the I/O operations sent to the SSD or to other storage devices coupled to the host computing system.

BRIEF DESCRIPTION OF THE DRAWINGS

Referring now to the drawings in which like reference numbers represent corresponding parts throughout:

FIG. 1 illustrates a block diagram of a computing environment in which a storage device, such as a SSD, is coupled to a host, in accordance with certain embodiments;

FIG. 2 illustrates a block diagram that shows how the storage device records the latency of an I/O operation, in accordance with certain embodiments;

FIG. 3 illustrates a block diagram that shows how the storage device maintains a latency profile that records the number of I/O operations processed at different intervals of latencies, in accordance with certain embodiments;

FIG. 4 illustrates a block diagram that shows the returning of the latency profile from the storage device to the host, in accordance with certain embodiments;

FIG. 5 illustrates a first flowchart that shows the measurement and reporting of the latency of input and output operations by a storage device to a host, in accordance with certain embodiments;

FIG. 6 illustrates a second flowchart that shows the measurement and reporting of the latency of input and output operations by a storage device to a host, in accordance with certain embodiments; and

FIG. 7 illustrates a block diagram of a device including a solid state drive or a computational device, in accordance with certain embodiments.

DETAILED DESCRIPTION

If a host computing system is used to measure the latencies of I/O operations transmitted from a host computing system to a storage device, then the measured latencies may be affected by the time consumed by a resource intensive program that executes in the host computing system to measure the latencies. Additionally, coupling components that couple the host computing system to the storage device may increase the measured latencies of the I/O operations. It may be difficult to separate latencies due to the storage device from those introduced by the operating system of the host computing system, the storage fabric, the host backplane, etc. Moreover, measuring latency in the host computing system may use software and hardware resources that may be better utilized for other tasks. Such latency measurements by the host computing system may be affected by the system software overhead and the noise introduced outside the storage device.

In certain embodiments, the SSD may record its own latency and this may allow the host computing system to determine the latency of I/O operations without consuming resources and assist in the debugging of latency issues. In certain embodiments, the SSD may record the time at which a read/write command (i.e., an I/O operation) is received and likewise the time at which the command is completed and sent back over a bus to the host computing system. This calculated latency for

each I/O operation may then be placed in various buckets and counted, where each bucket is a latency interval. The latency buckets and I/O counts may be communicated to the host computing system upon request. Thus, in certain embodiments, host resources are not consumed to measure latency.

In the following description, reference is made to the accompanying drawings which form a part hereof and which illustrate several embodiments. It is understood that other embodiments may be utilized and structural and operational changes may be made.

FIG. 1 illustrates a block diagram of a computing environment 100 in which a storage device, such as a SSD 102 with one or more non-volatile memory chips 104a...104n is coupled to a host 106, in accordance with certain embodiments. A non-volatile memory chip (also referred to as a package) may include one or more dies, where a die is the smallest unit that may independently execute commands or report status. Each die may include one or more planes, where identical, concurrent operations may take place on each plane. Each plane may include a number of blocks, which are the smallest unit that may be erased. Each block may include a number of pages, which are the smallest unit that may be programmed.

The solid state drive 102 may be comprised of non-volatile memory, such as NAND memory included in the non-volatile memory chips 104a...104n, NOR memory or some other suitable non-volatile memory. In certain embodiments, the solid state drive 102 may be capable of storing several terabytes of data or more. Certain embodiments may be applied to other types of non-volatile memory, phase change memory (PCM), a three dimensional cross point memory, a resistive memory, nanowire memory, ferro-electric transistor random access memory (FeTRAM), magnetoresistive random access memory (MRAM) memory that incorporates memristor technology, spin transfer torque (STT)-MRAM, byte addressable random access non-volatile memory, etc. In certain embodiments, the storage device 102 may be comprised of some other type of device besides a solid state drive. For example, in certain embodiments the storage device 102 may be comprised of a disk drive, a tape drive, etc.

In certain embodiments, the host 106 may be comprised of any suitable computational device, such as a personal computer, a mainframe, a telephony device, a smart phone, a storage controller, a blade computer, a processor with memory, etc.

The host 106 may be referred to as a host computing system or as a computational device. The host 106 may communicate with the SSD 102 over a bus (such as Peripheral Component Interconnect (PCIe), Serial Advanced Technology Attachment (SATA), Serial Attached Small Computer System Interface (SAS)) or a network, such as the Internet, a storage area network (SAN), a local area network (LAN), etc. Further details of the SATA specification may be found in the publication titled "Serial ATA Specification, Revision 3.2," released August 2013, by SATA International Organization (SATA-IO), Beaverton, OR. In another example, the interface and/or interconnect protocol may comply and/or be compatible with an NVMe (Non-Volatile Memory Host Controller Interface Express). Further details of NVMe may be found in the publication titled "NVM Express™, Revision 1.2," released November 3, 2014 by NVM Express™ Work Group, and/or earlier and/or later versions of this specification (NVM Express is a trademark of NVM Express, Inc.).

In FIG. 1 the solid state drive 102 has been shown as being external to the host 106. In alternative embodiments, the solid state drive 102 may be included within a chassis of the host 106. The host 106 may have an operating system 108 that interacts with the solid state drive 102 to perform I/O operations.

The solid state drive 102 includes a controller 110 implemented in firmware, hardware, software or any combination thereof. The controller 110 a latency measurement and reporting logic 112 that may be implemented in firmware, hardware, software of any combination thereof.

In certain embodiments illustrated in FIG. 1, the latency measurement and reporting logic 112 measures the latencies of I/O operations processed by the solid state drive 102 and reports the latencies of the I/O operations to the host 106, either at periodic intervals or in response to a request from the host 106. The I/O operations may include a request to perform read operations on one or more of the non-volatile chips 104a...104n, or a request to perform write operations on one or more of the non-volatile memory chips 104a...104n.

FIG. 2 illustrates a block diagram 200 that shows how a storage device records the latency of an I/O operation, in accordance with certain embodiments. In certain embodiments, the latency measurement and reporting logic 112 that executes within the storage device performs the operations shown in FIG. 2.

The host (shown via reference numeral 202) may send a read or a write command to the storage device (shown via reference numeral 204), where the transmission of the read or the write command is shown via reference numeral 206. The host (shown via reference numeral 202) may send the read or write command to the storage device (shown via reference numeral 204) over a bus, such as, PCIe, SATA, SAS, etc., or over a network, such as the Internet, a SAN, a LAN, etc. For example, PCIe is a high speed I/O bus in which multiple lanes may combine their data transfer capability to send the read or write command from the host to the storage device, whereas SATA may use a high-speed serial cable to send the read or write command from the host to the storage device. In certain embodiments that use interfaces in accordance with the SATA Express specification Revision 3.2, an interface that combines both SATA and PCIe buses may be used to send the read or write commands from the host to the storage device. The storage device records a time stamp 208 to indicate the time at which the read or the write command was received by the storage device from the host. In certain embodiments, the time stamp 208 indicates the time when the entirety of the read or the write command is received by the storage device. For example, in the case of a read command the time stamp 208 may record when the last byte of the read command is received by the storage device, and in the case of a write command the time stamp 208 may record when the last byte of data to be written is received by the storage device. In alternative embodiments, other indications related to the receiving of the read or write command may be used to record the time of receiving the read or write command in the time stamp 208. Since the storage device has an internal clock the internal clock may be used to determine the time for the recording the time stamp.

Then the storage device processes the read or the write command, and subsequently the storage device (shown via reference numeral 210) sends the response 214 to the read or the write command to the host (shown via reference numeral 212). The response 214 includes the data generated via the processing of the read or the write command. The data included in the response 214 may be of a considerable size (e.g., in the case of a read command). For example, in certain embodiments, the response includes a plurality of bytes of data (e.g., one or more kilobytes of data, one or more megabytes of data, etc., in response to a read command). In responding to a read command, the storage device records a time stamp

216 at the time at which the last byte of the data included in the response to the read command from host is transmitted by the storage device. In responding to the write command from the host, the storage device records in the timestamp 216, the time at which the last byte of data is sent to the host to inform the host that that data requested to be written by the write command has been written to the non-volatile memory chips 104a…104n. The number of bytes of data sent in response to a write request is usually much smaller than the number of bytes of data sent in response to a read request. In response to a write request, the number of bytes of data that are sent may just inform the host that the write operation has completed in the storage device, whereas in response to a read request, the number of bytes of data that are sent may include the data that is read from the storage device.

FIG. 3 illustrates a block diagram 300 that shows how the storage device maintains a latency profile that records the number of I/O operations processed at different intervals of latencies, in accordance with certain embodiments. In certain embodiments, the latency measurement and reporting logic 112 that executes within the storage device performs the operations shown in FIG. 3.

The difference 302 between the time stamp 216 and the time stamp 208 is the latency of the read or the write command. Since the timestamp 216 is the time at which the last byte of the data included in the response to the host is transmitted by the storage device, therefore, latency is not only dependent on the time to perform the reading of data from the non-volatile memory chips 104a..104n or the writing of the data to the non-volatile memory chips 104a…104n, but is also dependent on the size of the data transfer to the host. Latency may also differ between read and write operations depending on the amount of data that is read or written. In the above example if the time stamp 208 is “6:32” and the “time stamp” 208 is “5:56” (as shown earlier in FIG. 2), then the latency for the read or the write operation is “0:36” as shown via reference numeral 302 and this latency is calculated by the latency measurement and reporting logic 112 that executes within the storage device. The units for the timestamps 208, 214 and the difference 302 may be any unit of time such as microseconds, milliseconds, nanoseconds, etc.

Since there may be many thousands, hundreds of thousands, or more of I/O operations, storing separate latencies for each I/O operation may need a significant amount of storage space in the storage device. To reduce the amount of storage space

needed for recording latencies, the storage device may maintain a latency profile 304 of a plurality of I/O operations in the form of a table or some other data structure. For example, latencies may be maintained in certain latency intervals (e.g. microseconds, milliseconds or some other unit). In FIG. 3, in the latency profile 304 there are ten latency intervals each of 10 units of time (e.g. milliseconds or microseconds or some other unit of time) between 0 and 99, and another latency interval that captures all latencies over 100 units of time. The latency intervals may also be referred to as latency ranges.

The latency measurement and reporting logic 112 may map the time stamp difference (i.e., the latency shown via reference numeral 306) of "0:36" to the latency interval "30-39" shown via reference numeral 308 and increment the command count 310 by 1 for this latency interval "30-39" and the incremented command count is shown via reference numeral 312 in the updated latency profile 314. Thus for each latency interval (i.e., latency range), the number of host I/O operations that fall within that latency interval is recorded in the latency profile.

As more and more I/O operations are received by the storage device, the latency profile is updated. At any point in time, the latency profile indicates the number of I/O commands that had latencies that fell within each latency interval.

FIG. 4 illustrates a block diagram 400 that shows the returning of the latency profile from the storage device to the host, in accordance with certain embodiments. In certain embodiments, the latency measurement and reporting logic 112 that executes with the storage device performs the operations shown in FIG. 4.

In certain embodiments, after the host (reference numeral 402) has sent (reference numeral 404) a plurality of I/O operations 404 to the storage device (reference numeral 406) the host may request a latency profile from the storage device. In response to receiving the request, the storage device (reference numeral 408) returns (shown via reference numeral 410) the latency profile 412 to the host (reference numeral 414). The latency profile 412 shows that 20,120 I/O commands had latencies that fell between 0 and 9 units of time (reference numeral 416), 103,130 I/O commands had latencies that fell between 40-49 units of time (reference numeral 418) , and that no I/O command had a latency of 100 or more units of time (reference numeral 420).

Thus FIGs. 1-4 illustrate certain embodiments in which a storage device computes the latencies of a plurality of I/O operations, and populates a latency profile in which a count of the number of I/O operations at different intervals of latencies are maintained. In response to a request from the host, the storage device transmits the latency profile to the host.

FIG. 5 illustrates a flowchart 500 that shows the measurement and reporting of the latency of input and output operations by a storage device 102, in accordance with certain embodiments. In certain embodiments, the latency measurement and reporting logic 112 that executes with the storage device 102 performs the operations shown in FIG. 5.

Control starts at block 502 in which a storage device 102 (e.g., a SSD) initializes a latency profile with latency intervals and a corresponding number of I/O operations that have latencies falling within the latency intervals, where the number of I/O operations for each latency interval is initially set to zero (as shown in FIG. 3 in the latency profile 304).

The storage device 102 receives (at block 504) an I/O command (i.e., an I/O operation) from the host 106. The storage device 102 computes (at block 506) the latency of the I/O command by recording the time of receiving the I/O command (referred to as a first recorded time) and by recording the time at which the result of the I/O command is transmitted to the host 106 (referred to as the second recorded time), and computing the time difference from the two recorded times. If the I/O command is a read operation then the second of the two recorded times is the time at which the final block of data requested by the host is sent from the storage device to the host. If the I/O command is a write command, then the second of the two recorded times is the time at which the storage device responds to the host that the write operation is completed in the storage device. The first recorded time is the time at which the entirety of the I/O command is received by the storage device. Thus the latency is the difference between the second recorded time and the first recorded time. In alternative embodiments other times may be recorded for measuring the latency of the storage device.

Control proceeds to block 508 in which the storage device 102 determines the latency interval within which the computed latency falls and increments the number of I/O operations for the latency interval by 1 (as shown in FIG. 3). Subsequently, the

storage device 102 determines (at block 510) whether the host 106 has requested the latency profile. If so, then the storage device 102 transmits (at block 512) the latency profile to the host 106. If not, then control proceeds to block 504 in which the storage device 102 receives another I/O command from the host 106.

It should be noted that the storage device 102 may process more than one I/O operation in parallel and may compute latencies in parallel for a plurality of I/O operations and may update the latency profile in parallel.

FIG. 6 illustrates a second flowchart 600 that shows the measurement and reporting of the latency of input and output operations by a storage device 102, in accordance with certain embodiments. In certain embodiments, the latency measurement and reporting logic 112 that executes with the storage device performs the operations shown in FIG. 6.

Control starts at block 602 in which a storage device 102 measures one or more latencies of one or more input/output (I/O) operations received from a host 106. The storage device 102 maps (at block 604) the measured one or more latencies to one or more latency ranges maintained in a latency profile. In response to a request from the host 106, the storage device 102 transmits (at block 606) the information on the one or more latencies to the host by via the latency profile.

Therefore, FIGs. 1-6 illustrate certain embodiments in which a latency profile is generated for a plurality of I/O operations in a storage device 102. In response to a request from a host 106 (e.g., a request sent over one or more lanes of a PCIe interface, or over a SATA interface from the host 106 to the storage device 102), the latency profile is transmitted over the one or more lanes of a PCIe interface or over the SATA interface by the storage device to the host 106. Other interfaces may be used to communicate between the host 106 and the storage device 102. As a result, the latency profile is not affected by delays in the host 106 or via coupling elements between the host 106 and storage device 102.

The described operations may be implemented as a method, apparatus or computer program product using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof. The described operations may be implemented as code maintained in a "computer readable storage medium", where a processor may read and execute the code from the computer storage readable medium. The computer readable storage medium includes

at least one of electronic circuitry, storage materials, inorganic materials, organic materials, biological materials, a casing, a housing, a coating, and hardware. A computer readable storage medium may comprise, but is not limited to, a magnetic storage medium (e.g., hard drive drives, floppy disks, tape, etc.), optical storage (CD-ROMs, DVDs, optical disks, etc.), volatile and non-volatile memory devices (e.g., EEPROMs, ROMs, PROMs, RAMs, DRAMs, SRAMs, Flash Memory, firmware, programmable logic, etc.), Solid State Devices (SSD), etc. The code implementing the described operations may further be implemented in hardware logic implemented in a hardware device (e.g., an integrated circuit chip, Programmable Gate Array (PGA), Application Specific Integrated Circuit (ASIC), etc.). Still further, the code implementing the described operations may be implemented in "transmission signals", where transmission signals may propagate through space or through a transmission media, such as an optical fiber, copper wire, etc. The transmission signals in which the code or logic is encoded may further comprise a wireless signal, satellite transmission, radio waves, infrared signals, Bluetooth, etc. The program code embedded on a computer readable storage medium may be transmitted as transmission signals from a transmitting station or computer to a receiving station or computer. A computer readable storage medium is not comprised solely of transmission signals. Those skilled in the art will recognize that many modifications may be made to this configuration, and that the article of manufacture may comprise suitable information bearing medium known in the art.

Computer program code for carrying out operations for aspects of the certain embodiments may be written in any combination of one or more programming languages. Blocks of the flowchart and block diagrams may be implemented by computer program instructions.

FIG. 7 illustrates a block diagram of a system 700 that includes both the host 106 (the host 106 comprises at least a processor) and the solid state drive 102, in accordance with certain embodiments. For example, in certain embodiments the system 700 may be a computer (e.g., a laptop computer, a desktop computer, a tablet, a cell phone or any other suitable computational device) that has the host 106 and the solid state drive 102 both included in the system 700. For example, in certain embodiments the system 700 may be a laptop computer that includes the solid state drive 102. The system 700 may include a circuitry 702 that may in certain

embodiments include at least a processor 704. The system 700 may also include a memory 706 (e.g., a volatile memory device), and storage 708. The storage 708 may include the solid state drive 102 or other drives or devices including a non-volatile memory device (e.g., EEPROM, ROM, PROM, flash, firmware, programmable logic, etc.). The storage 708 may also include a magnetic disk drive, an optical disk drive, a tape drive, etc. The storage 708 may comprise an internal storage device, an attached storage device and/or a network accessible storage device. The system 700 may include a program logic 710 including code 712 that may be loaded into the memory 706 and executed by the processor 704 or circuitry 702. In certain embodiments, the program logic 710 including code 712 may be stored in the storage 708. In certain other embodiments, the program logic 710 may be implemented in the circuitry 702. Therefore, while FIG. 7 shows the program logic 710 separately from the other elements, the program logic 710 may be implemented in the memory 706 and/or the circuitry 702. The system 700 may also include a display 714 (e.g., an liquid crystal display (LCD), a light emitting diode (LED) display, a cathode ray tube (CRT) display, a touchscreen display, or any other suitable display). The system 700 may also include one or more input devices 716, such as, a keyboard, a mouse, a joystick, a trackpad, or any other suitable input devices). Other components or devices beyond those shown in FIG. 7 may also be found in the system 700.

Certain embodiments may be directed to a method for deploying computing instruction by a person or automated processing integrating computer-readable code into a computing system, wherein the code in combination with the computing system is enabled to perform the operations of the described embodiments.

The terms "an embodiment", "embodiment", "embodiments", "the embodiment", "the embodiments", "one or more embodiments", "some embodiments", and "one embodiment" mean "one or more (but not all) embodiments" unless expressly specified otherwise.

The terms "including", "comprising", “having” and variations thereof mean "including but not limited to", unless expressly specified otherwise.

The enumerated listing of items does not imply that any or all of the items are mutually exclusive, unless expressly specified otherwise.

The terms "a", "an" and "the" mean "one or more", unless expressly specified otherwise.

Devices that are in communication with each other need not be in continuous communication with each other, unless expressly specified otherwise. In addition, devices that are in communication with each other may communicate directly or indirectly through one or more intermediaries.

A description of an embodiment with several components in communication with each other does not imply that all such components are required. On the contrary a variety of optional components are described to illustrate the wide variety of possible embodiments.

Further, although process steps, method steps, algorithms or the like may be described in a sequential order, such processes, methods and algorithms may be configured to work in alternate orders. In other words, any sequence or order of steps that may be described does not necessarily indicate a requirement that the steps be performed in that order. The steps of processes described herein may be performed in any order practical. Further, some steps may be performed simultaneously.

When a single device or article is described herein, it will be readily apparent that more than one device/article (whether or not they cooperate) may be used in place of a single device/article. Similarly, where more than one device or article is described herein (whether or not they cooperate), it will be readily apparent that a single device/article may be used in place of the more than one device or article or a different number of devices/articles may be used instead of the shown number of devices or programs. The functionality and/or the features of a device may be alternatively embodied by one or more other devices which are not explicitly described as having such functionality/features. Thus, other embodiments need not include the device itself.

At least certain operations that may have been illustrated in the figures show certain events occurring in a certain order. In alternative embodiments, certain operations may be performed in a different order, modified or removed. Moreover, steps may be added to the above described logic and still conform to the described embodiments. Further, operations described herein may occur sequentially or certain operations may be processed in parallel. Yet further, operations may be performed by a single processing unit or by distributed processing units.

The foregoing description of various embodiments has been presented for the purposes of illustration and description. It is not intended to be exhaustive or to be

limited to the precise forms disclosed. Many modifications and variations are possible in light of the above teaching.

Examples

The following examples pertain to further embodiment.

Example 1 is a method for measuring latency of a storage device. A storage devices measures one or more latencies of one or more input/output (I/O) operations received from a host. The storage device transmits information on the one or more latencies to the host.

In example 2, the subject matter of example 1 may include that the storage device is a solid state drive (SSD).

In example 3, the subject matter of example 2 may include recording, by the SSD, a first time at which an I/O operation is received from the host; recording, by the SSD, a second time at which the SSD transmits a response to the I/O operation to the host; and determining a latency of the I/O operation to be a time duration between the second time and the first time.

In example 4, the subject matter of example 3 may include maintaining a plurality of latency ranges; and determining how many of the one or more I/O operations have latencies that fall within each of the plurality of latency ranges, to generate a latency profile for the SSD.

In example 5, the subject matter of example 4 may include that in response to a request from the host, the SSD transmits the latency profile to the host.

In example 6, the subject matter of example 1 may include that the measured one or more latencies are mapped to one or more latency ranges maintained in a latency profile, wherein the measured one or more latencies are transmitted to the host via the latency profile to transmit the information on the one or more latencies to the host.

In example 7, the subject matter of example 1 may include that all measurements of the one or more latencies are performed in the storage device and no measurements of the one or more latencies are performed in the host.

In example 8, the subject matter of example 7 may include that by measuring the one or more latencies in the storage device and not in the host, the measurements of the one or more latencies do not include delays introduced by extraneous effects of processes that execute in the host.

Example 9 is an apparatus comprising a plurality of non-volatile memory chips;

and a controller coupled to the plurality of non-volatile memory chips, wherein the controller is operable to: measure one or more latencies of one or more input/output (I/O) operations received from a host; and transmit information on the one or more latencies to the host.

In example 10, the subject matter of example 9 may include that the apparatus comprises a solid state drive (SSD).

In example 11, the subject matter of example 10 may include that the controller is further operable to: record a first time at which an I/O operation is received from the host; recording a second time at which the SSD transmits a response to the I/O operation to the host; and determine a latency of the I/O operation to be a time duration between the second time and the first time.

In example 12, the subject matter of example 11 may include that the controller is further operable to: maintain a plurality of latency ranges; and determine how many of the one or more I/O operations have latencies that fall within each of the plurality of latency ranges, to generate a latency profile for the SSD.

In example 13, the subject matter of example 12 may include that the controller is further operable to: in response to a request from the host, transmit the latency profile to the host.

In example 14, the subject matter of example 9 may include that the measured one or more latencies are mapped to one or more latency ranges maintained in a latency profile, wherein the measured one or more latencies are transmitted to the host via the latency profile to transmit the information on the one or more latencies to the host.

In example 15, the subject matter of example 9 may include that all measurements of the one or more latencies are performed in the storage device and no measurements of the one or more latencies are performed in the host.

In example 16, the subject matter of example 15 may include that by measuring the one or more latencies in the storage device and not in the host, the measurements of the one or more latencies do not include delays introduced by extraneous effects of processes that execute in the host.

Example 17 is a system comprising a solid state drive (SSD) comprising a plurality of non-volatile memory chips; a display; and a processor coupled to the solid state drive and the display, wherein, the SSD is operable to: measure one or more latencies of one or more input/output (I/O) operations received from a host; and

transmit information on the one or more latencies to the host.

In example 18, the subject matter of example 9 may include that host is included in the system and coupled to the SSD, and the processor is included in the host.

In example 19, the subject matter of example 18 may include that the SSD is further operable to: record a first time at which an I/O operation is received from the host; recording a second time at which the SSD transmits a response to the I/O operation to the host; and determine a latency of the I/O operation to be a time duration between the second time and the first time.

In example 20, the subject matter of example 19 may include that the SSD is further operable to: maintain a plurality of latency ranges; and determine how many of the one or more I/O operations have latencies that fall within each of the plurality of latency ranges, to generate a latency profile for the SSD.

In example 21, the subject matter of example 20 may include that the SSD is further operable to: in response to a request from the host, transmit the latency profile to the host.

In example 22, the subject matter of example 17 may include that the measured one or more latencies are mapped to one or more latency ranges maintained in a latency profile, wherein the measured one or more latencies are transmitted to the host via the latency profile to transmit the information on the one or more latencies to the host.

In example 23, the subject matter of example 17 may include that all measurements of the one or more latencies are performed in the storage device and no measurements of the one or more latencies are performed in the host.

In example 24, the subject matter of example 23 may include that by measuring the one or more latencies in the storage device and not in the host, the measurements of the one or more latencies do not include delays introduced by extraneous effects of processes that execute in the host.

WHAT IS CLAIMED IS

1. A method for measuring latency of a storage device, the method comprising:

measuring, by the storage device, one or more latencies of one or more input/output (I/O) operations received from a host; and

transmitting, by the storage device, information on the one or more latencies to the host.

2. The method of claim 1, wherein the storage device is a solid state drive (SSD).

3. The method of claim 2, the method further comprising:

recording, by the SSD, a first time at which an I/O operation is received from the host;

recording, by the SSD, a second time at which the SSD transmits a response to the I/O operation to the host; and

determining a latency of the I/O operation to be a time duration between the second time and the first time.

4. The method of claim 3, the method further comprising:

maintaining a plurality of latency ranges; and

determining how many of the one or more I/O operations have latencies that fall within each of the plurality of latency ranges, to generate a latency profile for the SSD.

5. The method of claim 4, the method further comprising:

in response to a request from the host, transmitting, by the SSD, the latency profile to the host.

6. The method of claim 1, wherein the measured one or more latencies are mapped to one or more latency ranges maintained in a latency profile, and wherein the measured one or more latencies are transmitted to the host via the latency profile to transmit the information on the one or more latencies to the host.

7. The method of claim 1, wherein all measurements of the one or more latencies are

performed in the storage device and no measurements of the one or more latencies are performed in the host.

8. The method of claim 7, wherein by measuring the one or more latencies in the storage device and not in the host, the measurements of the one or more latencies do not include delays introduced by extraneous effects of processes that execute in the host.

9. An apparatus, comprising:

a plurality of non-volatile memory chips; and

a controller coupled to the plurality of non-volatile memory chips, wherein the controller is operable to:

measure one or more latencies of one or more input/output (I/O) operations received from a host; and

transmit information on the one or more latencies to the host.

10. The apparatus of claim 9, wherein the apparatus comprises a solid state drive (SSD).

11. The apparatus of claim 10, wherein the controller is further operable to:

record a first time at which an I/O operation is received from the host;

record a second time at which the SSD transmits a response to the I/O operation to the host; and

determine a latency of the I/O operation to be a time duration between the second time and the first time.

12. The apparatus of claim 11, wherein the controller is further operable to:

maintain a plurality of latency ranges; and

determine how many of the one or more I/O operations have latencies that fall within each of the plurality of latency ranges, to generate a latency profile for the SSD.

13. The apparatus of claim 12, wherein the controller is further operable to:

in response to a request from the host, transmit the latency profile to the host.

14. The apparatus of claim 9, wherein the measured one or more latencies are mapped to one or more latency ranges maintained in a latency profile, and wherein the measured one or more latencies are transmitted to the host via the latency profile to transmit the information on the one or more latencies to the host.

15. The apparatus of claim 9, wherein all measurements of the one or more latencies are performed in the apparatus and no measurements of the one or more latencies are performed in the host.

16. The apparatus of claim 15, wherein by measuring the one or more latencies in the apparatus and not in the host, the measurements of the one or more latencies do not include delays introduced by extraneous effects of processes that execute in the host.

17. A system, comprising:

a solid state drive (SSD) comprising a plurality of non-volatile memory chips;

a display; and

a processor coupled to the solid state drive and the display, wherein, the SSD is operable to:

measure one or more latencies of one or more input/output (I/O) operations received from a host; and

transmit information on the one or more latencies to the host.

18. The system of claim 17, wherein the host is included in the system and coupled to the SSD, and the processor is included in the host.

19. The system of claim 18, wherein the SSD is further operable to:

record a first time at which an I/O operation is received from the host;

record a second time at which the SSD transmits a response to the I/O operation to the host; and

determine a latency of the I/O operation to be a time duration between the second time and the first time.

20. The system of claim 19, wherein the SSD is further operable to:

maintain a plurality of latency ranges; and

determine how many of the one or more I/O operations have latencies that fall within each of the plurality of latency ranges, to generate a latency profile for the SSD.

21. The system of claim 20, wherein the SSD is further operable to:

in response to a request from the host, transmit the latency profile to the host.

22. The system of claim 17, wherein the measured one or more latencies are mapped to one or more latency ranges maintained in a latency profile, and wherein the measured one or more latencies are transmitted to the host via the latency profile to transmit the information on the one or more latencies to the host.

23. The system of claim 17, wherein all measurements of the one or more latencies are performed in the SSD and no measurements of the one or more latencies are performed in the host.

24. The system of claim 23, wherein by measuring the one or more latencies in the SSD and not in the host, the measurements of the one or more latencies do not include delays introduced by extraneous effects of processes that execute in the host.

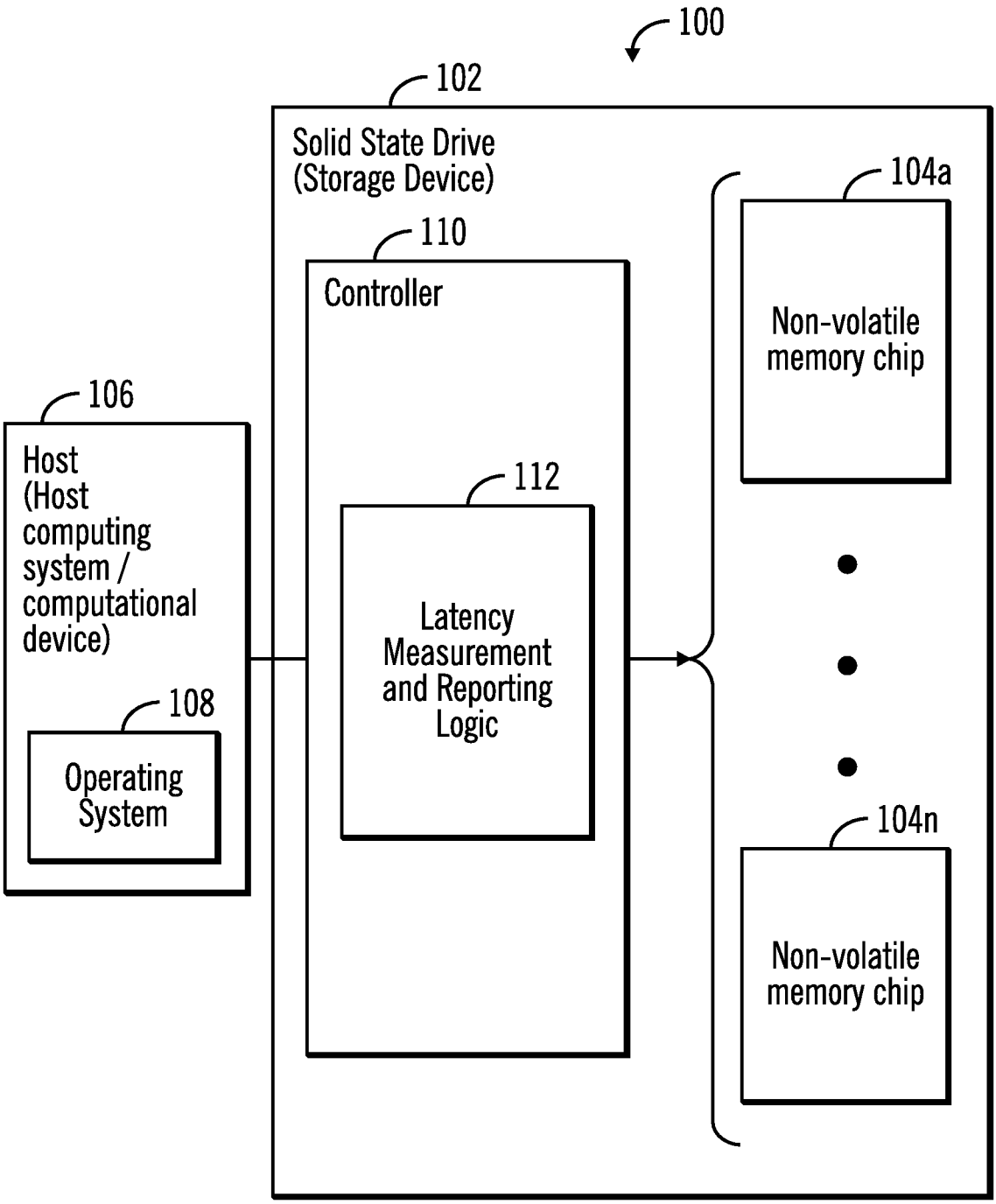

100
102
Solid State Drive
(Storage Device)
104a
110
Controller
Non-volatile
memory chip
106
Host
(Host
computing
system /
computational
device)
112
Latency
Measurement
and Reporting
Logic
108
Operating
System
104n
Non-volatile
memory chip

FIG. 1

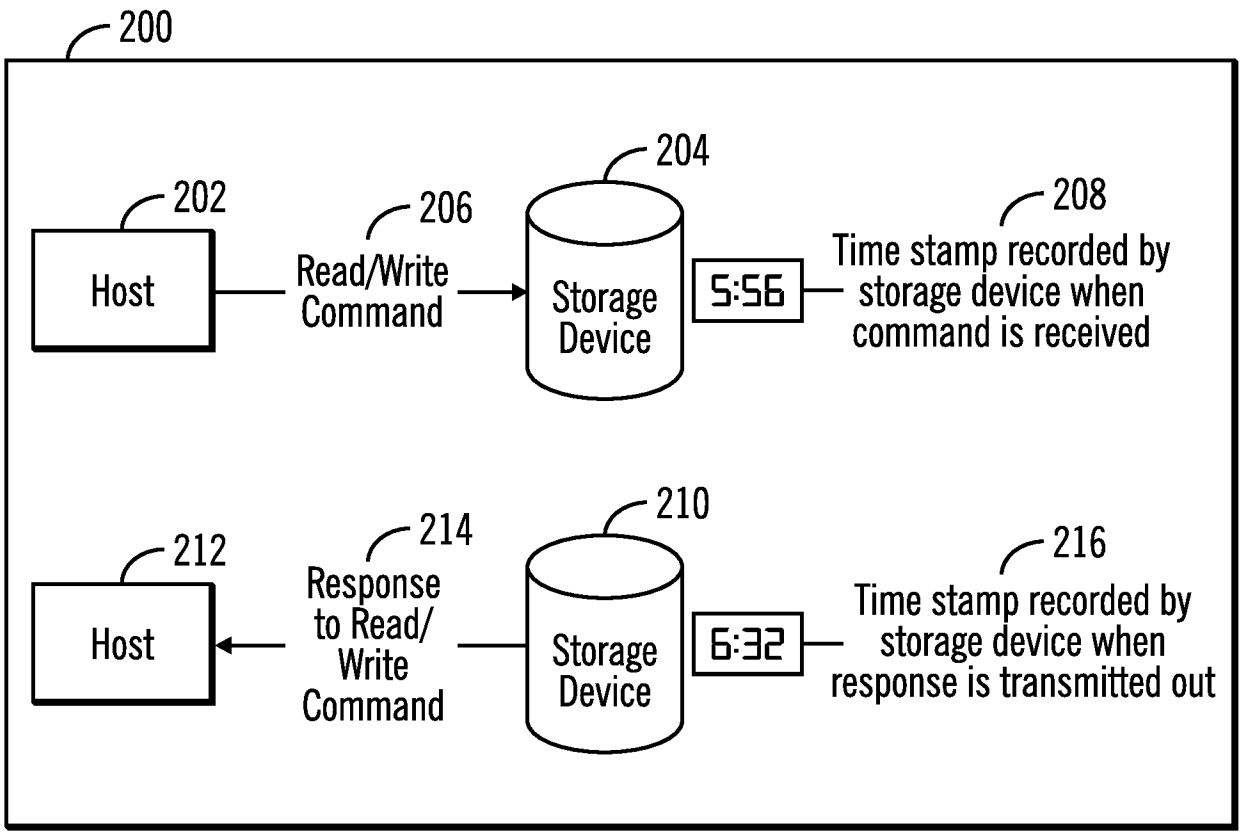

200
202
Host
206
Read/Write Command
204
Storage Device
5:56
208
Time stamp recorded by storage device when command is received
212
Host
214
Response to Read/ Write Command
210
Storage Device
6:32
216
Time stamp recorded by storage device when response is transmitted out

FIG. 2

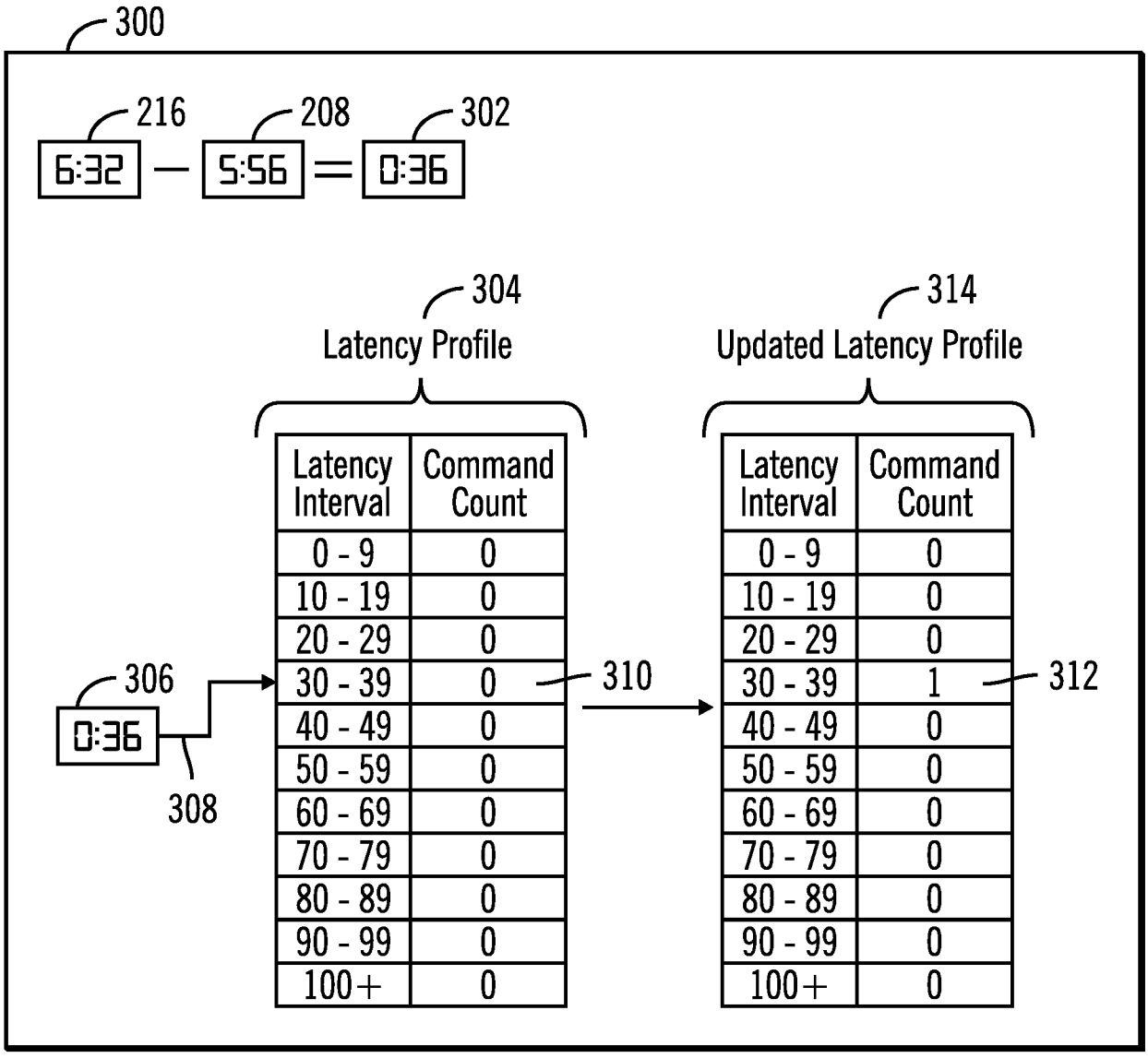
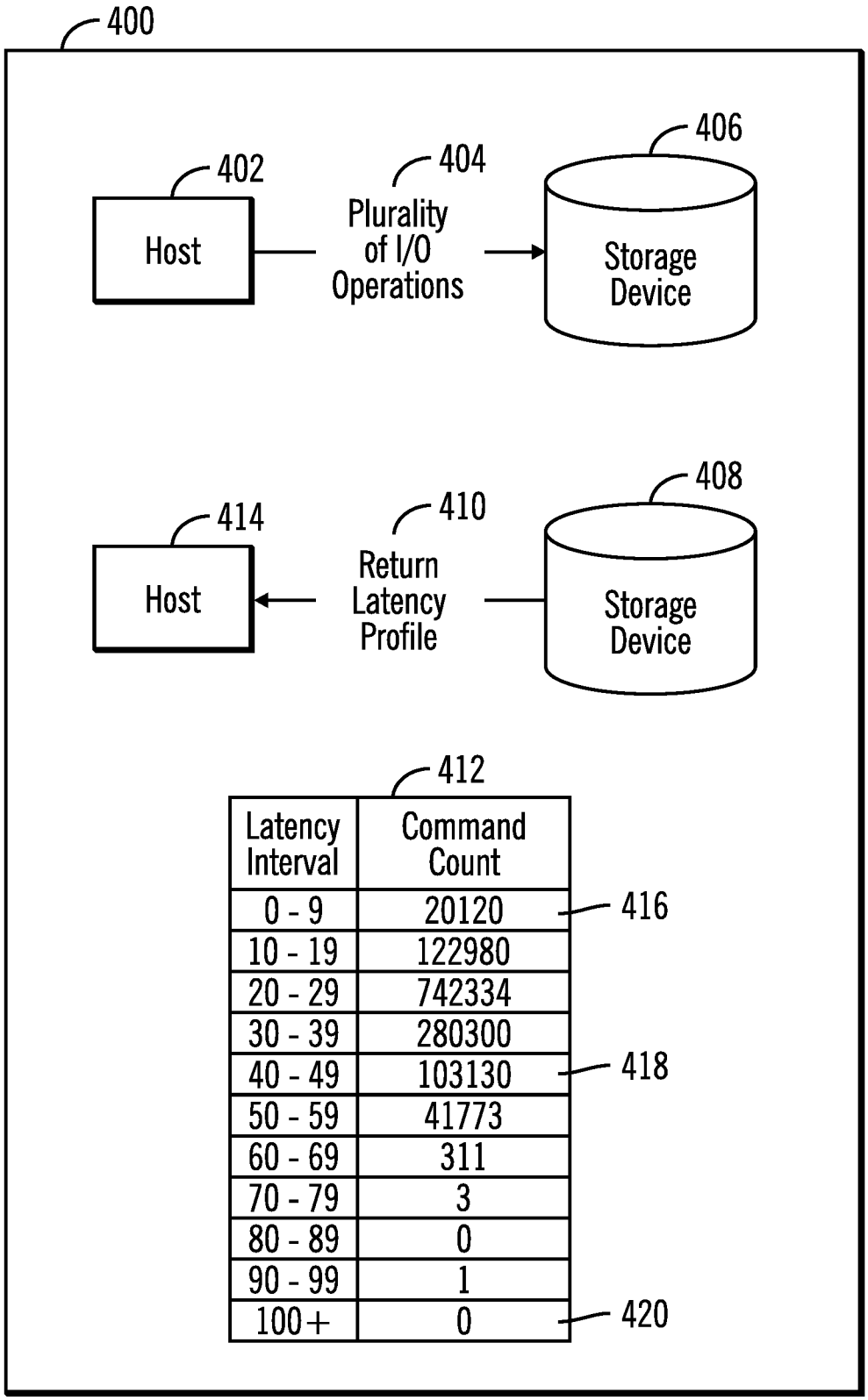

300
216
208
302
6:32
5:56
0:36
304
Latency Profile
314
Updated Latency Profile
Latency Interval
Command Count
0 - 9
10 - 19
20 - 29
30 - 39
40 - 49
50 - 59
60 - 69
70 - 79
80 - 89
90 - 99
100+
306
0:36
308
310
312

FIG. 3



412

Latency Interval	Command Count
0 - 9	20120
10 - 19	122980
20 - 29	742334
30 - 39	280300
40 - 49	103130
50 - 59	41773
60 - 69	311
70 - 79	3
80 - 89	0
90 - 99	1
100+	0

416 418 420

FIG. 4

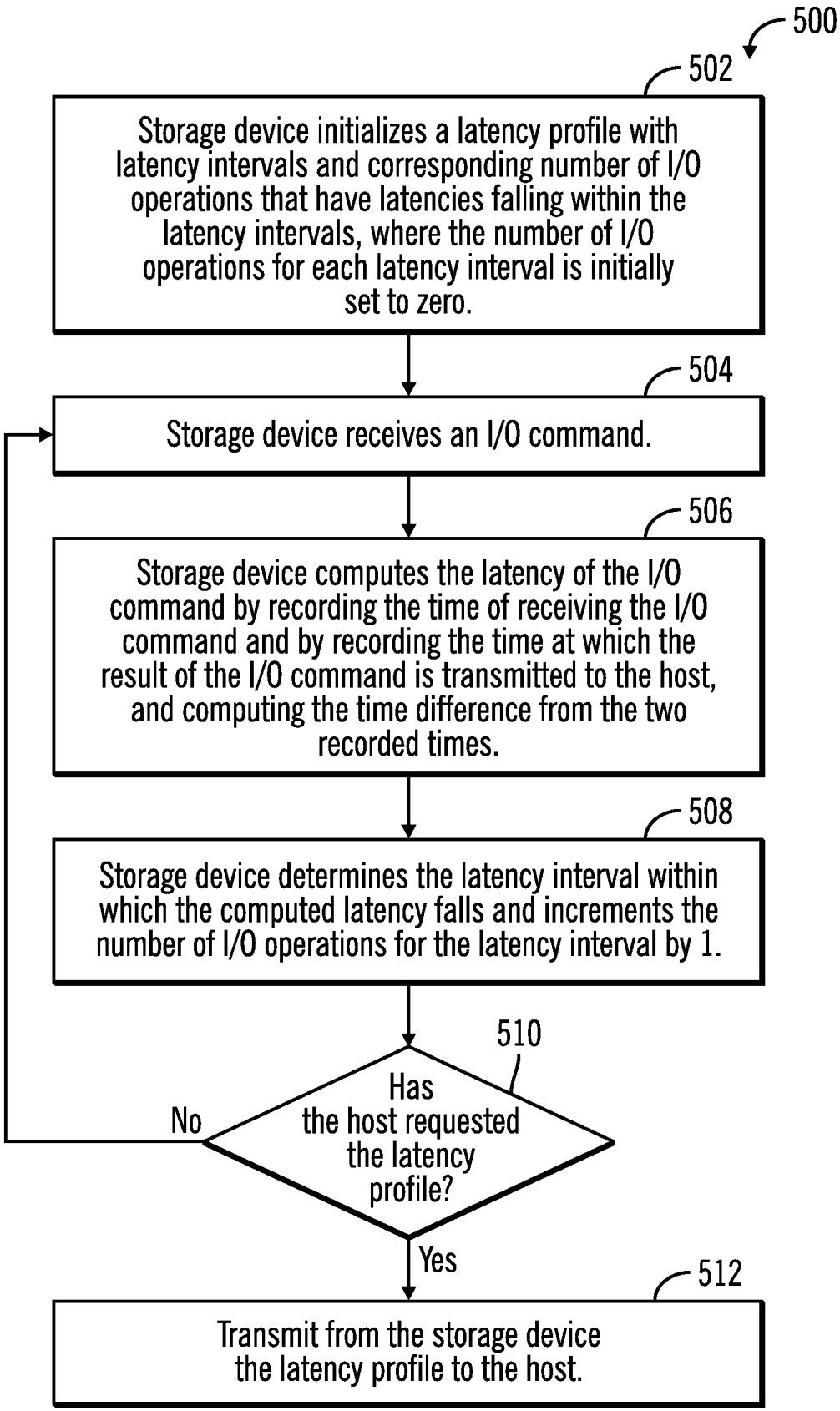

500
502
Storage device initializes a latency profile with latency intervals and corresponding number of I/O operations that have latencies falling within the latency intervals, where the number of I/O operations for each latency interval is initially set to zero.
504
Storage device receives an I/O command.
506
Storage device computes the latency of the I/O command by recording the time of receiving the I/O command and by recording the time at which the result of the I/O command is transmitted to the host, and computing the time difference from the two recorded times.
508
Storage device determines the latency interval within which the computed latency falls and increments the number of I/O operations for the latency interval by 1.
510
Has the host requested the latency profile?
No
Yes
512
Transmit from the storage device the latency profile to the host.

FIG. 5

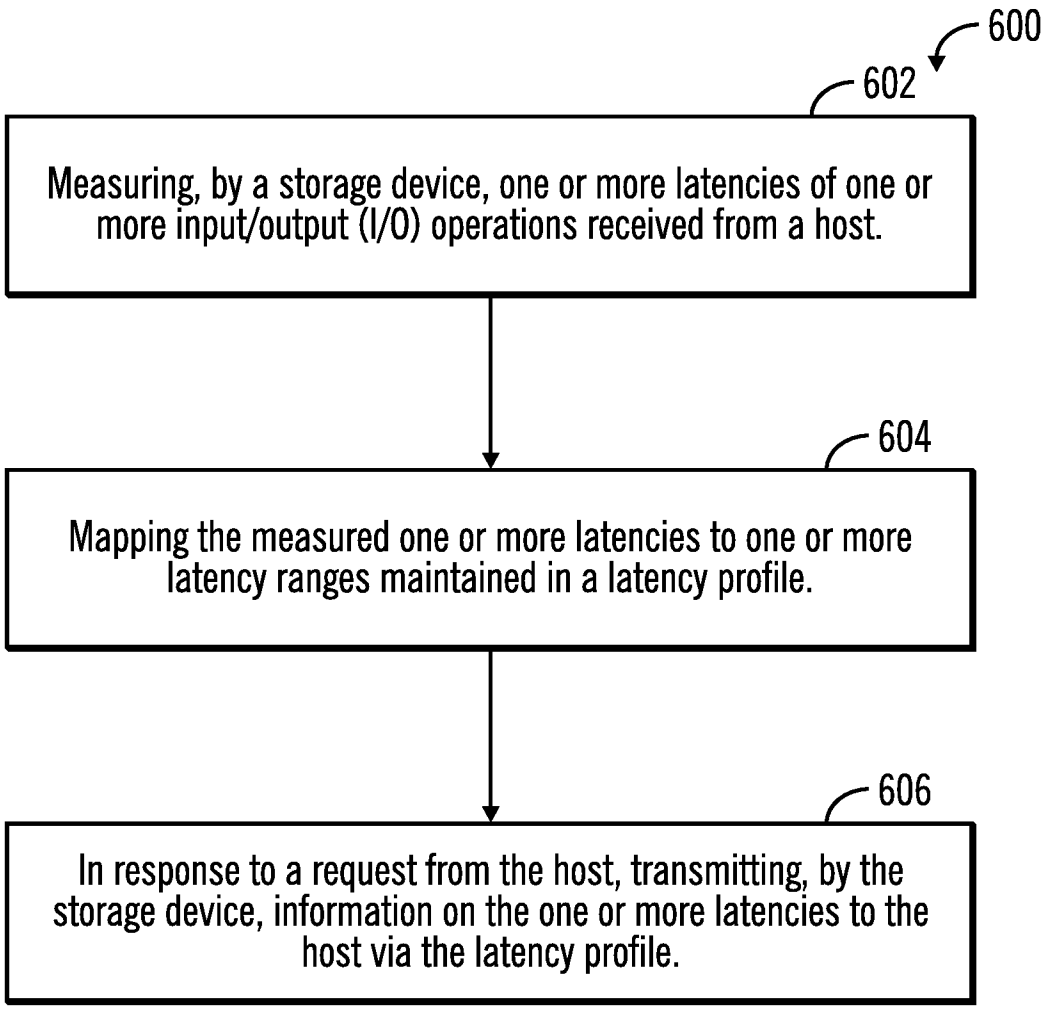

600
602
Measuring, by a storage device, one or more latencies of one or more input/output (I/O) operations received from a host.
604
Mapping the measured one or more latencies to one or more latency ranges maintained in a latency profile.
606
In response to a request from the host, transmitting, by the storage device, information on the one or more latencies to the host via the latency profile.

FIG. 6

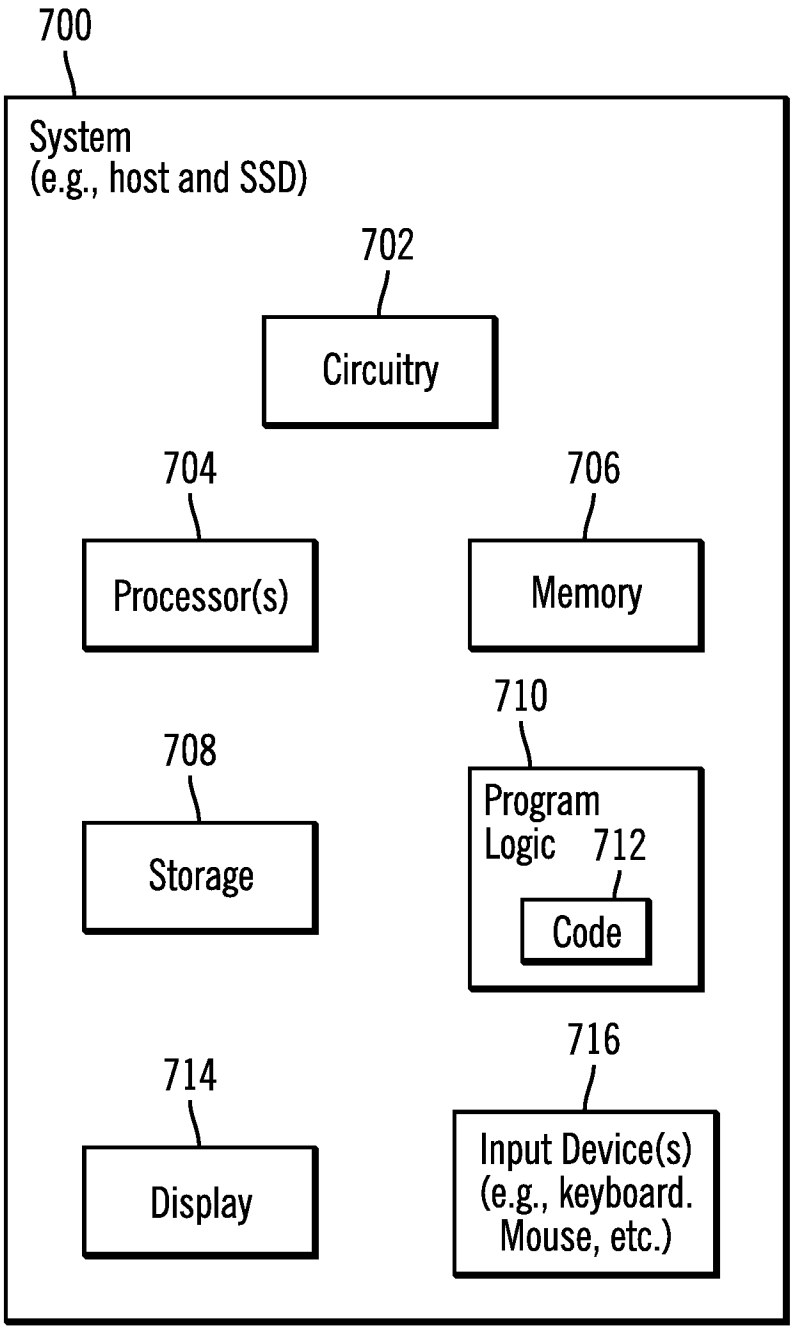

700
System
(e.g., host and SSD)
702
Circuitry
704
Processor(s)
706
Memory
708
Storage
710
Program
Logic 712
Code
714
Display
716
Input Device(s)
(e.g., keyboard.
Mouse, etc.)

FIG. 7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2016/019956

A. CLASSIFICATION OF SUBJECT MATTER

G06F 3/06(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 3/06; G06F 1/06; G06F 13/16; G06F 12/00; G06F 12/08; G06F 13/00

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: measure, latency, solid state drive(SSD), host, report, profile, and similar terms.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2014-126263 A1 (KABUSHIKI KAISHA TOSHIBA) 21 August 2014 See page 4, line 36 - page 7, line 29; page 11, line 11 - page 13, line 18; and figures 1, 3, 10.	1-24
Y	US 2012-0246435 A1 (AVRAHAM MEIR et al.) 27 September 2012 See paragraphs [0029]-[0052]; claims 1, 8; and figure 1.	1-24
A	US 8566483 B1 (XIANGPING CHEN et al.) 22 October 2013 See column 4, line 1 - column 6, line 44; claim 1; and figure 1	1-24
A	US 8972689 B1 (ERIK DE LA IGLESIA) 03 March 2015 See column 7, line 4 - column 9, line 38; and figures 7-9.	1-24
A	US 2011-0238928 A1 (MASAHIRO ABE et al.) 29 September 2011 See paragraphs [0016]-[0026]; and figure 1.	1-24

☐ 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	Date of mailing of the international search report
10 June 2016 (10.06.2016)	**10 June 2016 (10.06.2016)**

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
BYUN, Sung Cheal
Telephone No. +82-42-481-8262

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.
PCT/US2016/019956

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2014-126263 A1	21/08/2014	JP 2014-154023 A US 2015-0347033 A1	25/08/2014 03/12/2015
US 2012-0246435 A1	27/09/2012	US 9021215 B2	28/04/2015
US 8566483 B1	22/10/2013	None	
US 8972689 B1	03/03/2015	None	
US 2011-0238928 A1	29/09/2011	JP 2011-203905 A JP 5159817 B2 US 8671260 B2	13/10/2011 13/03/2013 11/03/2014