

(51) International Patent Classification:
G06F 11/20 (2006.01) **G06F 12/08** (2006.01)(21) International Application Number:
PCT/JP2008/068690(22) International Filing Date:
8 October 2008 (08.10.2008)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **HITACHI, LTD.** [JP/JP]; 6-6, Marunouchi 1-chome, Chiyoda-ku, Tokyo, 1008280 (JP).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **FUJII, Masanori** [JP/JP]; c/o Hitachi, Ltd., Disk Array Systems Division, 322-2 Nakazato, Odawara-shi, Kanagawa, 2500872 (JP). **NISHIMURA, Tsukasa** [JP/JP]; c/o Hitachi, Ltd., Disk Array Systems Division, 322-2 Nakazato, Odawara-shi, Kanagawa, 2500872 (JP). **MIURA, Sumihiro** [JP/JP]; c/o Hitachi, Ltd., Disk Array Systems Division, 322-2 Nakazato, Odawara-shi, Kanagawa, 2500872 (JP). **OKUBO, Yoshinori** [JP/JP]; c/o Hitachi Software Engineering Co., Ltd., 12-7, Higashishinagawa 4-chome, Shinagawa-ku, Tokyo, 1400002 (JP).(74) Agent: **GOTO, Masaki**; Shoyu-Kaikan, 3-1, Kasumigaseki 3-chome, Chiyoda-ku, Tokyo, 1000013 (JP).

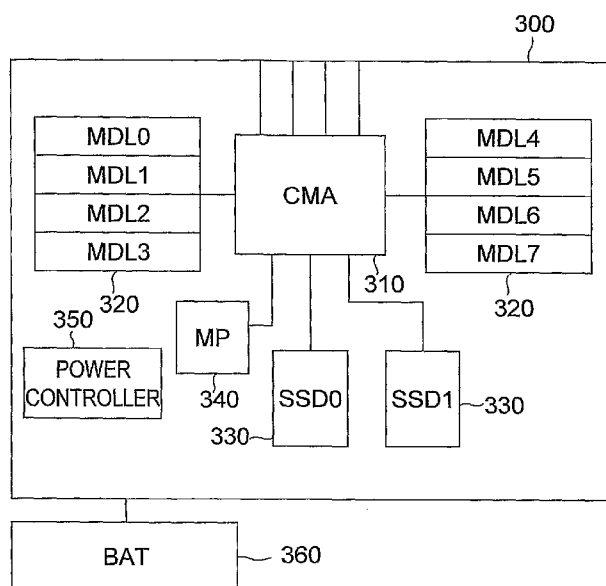
(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM,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

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

[Continued on next page]

(54) Title: STORAGE SYSTEM AND DATA MANAGEMENT METHOD

Fig. 2



(57) Abstract: A storage system, which is coupled to a computer, includes a storage device, a controller, a plurality of cache memory units, and a connecting unit. Each of the plurality of cache memory units includes: a cache memory for storing data; an auxiliary storage device for holding a content of data even after shutdown of power; and a cache controller for controlling an input/output of data to/from the cache memory and the auxiliary storage device. The cache controller store data stored in the cache memory, which is divided into a plurality of parts, into a plurality of the auxiliary storage devices included in the plurality of cache memory units.



Published:

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

DESCRIPTION

STORAGE SYSTEM AND DATA MANAGEMENT METHOD

5 Technical Field

This invention relates to a storage system for storing data according to an instruction from a computer, in particular, a technology of managing data stored in a cache memory.

10 Background Art

In a storage system according to a conventional technology, data written from a host computer is written into a cache memory before being written onto a disk drive. The moment that the data is written into the cache memory, the storage system notifies the host computer that a write
15 process has been completed, which improves responsiveness to the host computer.

A DRAM is generally used as the cache memory in terms of the number of times that a read/write is performed and an access speed. Therefore, the data stored in the cache memory volatilizes (disappears) when
20 power to the storage system is shut down, which necessitates backup of data (dirty data) before being written onto the disk drive by supplying power from a battery in case of service interruption.

Further, JP 2004-21811 A discloses a storage system in which a non-volatile memory is used as a cache memory to thereby prevent data
25 from volatilizing (disappearing) even when power is shut down.

Disclosure of the Invention

According to the above-mentioned conventional technology, data stored in a cache memory can be read by detaching the cache memory or a non-volatile memory (cache memory package) which is subjected to backup by a battery, leading to leakage of the data stored in the cache memory,
5 which raises a problem in terms of security.

Further, if dirty data stored in a DRAM is backed up by supplying power from a battery, a capacity of the battery needs to be increased depending on the period of time necessary for backup in case of service interruption, which raises a problem of an increase in cost of the battery.

10 Further, in a case where a flash memory is used as the cache memory, the flash memory is slower in access speed than the DRAM, which raises a problem of a decrease in response (response to host computer of storage system) of the cache memory.

Therefore, there is desired a storage system capable of backing up
15 dirty data even in case of long-term service interruption without decreasing the response to the host computer of the storage system.

A representative example of this invention will be described as follows. Specifically, a storage system, which is coupled to a computer, includes: a storage device for storing data for which write is requested by the computer;
20 a controller for controlling transfer of data in the storage system; a plurality of cache memory units for temporarily storing data to be stored into the storage device; and a connecting unit for connecting at least the plurality of cache memory units. Each of the plurality of cache memory units includes: a cache memory for storing data; an auxiliary storage device for holding data
25 even after shutdown of power; and a cache controller for controlling an input/output of data to/from the cache memory and the auxiliary storage device. The cache controller stores data stored in the cache memory, which

is divided into a plurality of parts, into a plurality of the auxiliary storage devices included in the plurality of cache memory units.

Further, according to another example of this invention, the storage system further includes a watch unit for watching power supplied to the plurality of cache memory packages. The cache controller stores at least one of configuration information and control information stored in the cache memory, which is divided into a plurality of parts, into the plurality of the auxiliary storage devices before the watch unit detects an abnormality in the power supplied to the plurality of cache memory units.

According to an embodiment of this invention, data stored in the original cache memory cannot be restored only with data stored in one of the auxiliary storage devices, which can ensure secrecy of data.

Further, during a normal operation, the configuration information and control information that are stored in the cache memory are stored in the auxiliary storage device, and hence it is possible to reduce an amount of data to be backed up to the auxiliary storage device when power is shut down, and to reduce a period of time required to store data in the auxiliary storage device. Accordingly, it is possible to reduce the capacity of the battery for backup of data.

Brief Description of the Drawings

Fig. 1 is a block diagram showing a configuration of a storage system in accordance with a first embodiment of this invention.

Fig. 2 is a block diagram showing a configuration of a cache memory package in accordance with the first embodiment of this invention.

Fig. 3 is an explanatory diagram of a storage area provided to a backup memory in accordance with the first embodiment of this invention.

Fig. 4 is a block diagram showing a configuration of the storage system around a switch unit in accordance with the first embodiment of this invention.

5 Fig. 5 is a block diagram showing a configuration of the storage system around a power supply line in accordance with the first embodiment of this invention.

Fig. 6 is an explanatory diagram of SSD memory address allocation information in accordance with the first embodiment of this invention.

10 Fig. 7 is an explanatory diagram of destination address information in accordance with the first embodiment of this invention.

Fig. 8 is an explanatory diagram of source address information in accordance with the first embodiment of this invention.

Fig. 9 is a flowchart showing a saving process of writing data into the backup memory in accordance with the first embodiment of this invention.

15 Fig. 10 is a flowchart showing a recovery process of writing data stored in the backup memories into a cache memory in accordance with the first embodiment of this invention.

20 Fig. 11 is a flowchart showing the recovery process of writing data stored in the backup memories into the cache memory in accordance with the first embodiment of this invention.

Fig. 12 is a block diagram showing a configuration of a storage system around a switch unit in accordance with a modified example of the first embodiment of this invention.

25 Fig. 13 is an explanatory diagram of an operation of a backup memory during a normal operation in accordance with a second embodiment of this invention.

Fig. 14 is a flowchart showing a process of writing secondary cache

data into the backup memory in accordance with the second embodiment of this invention.

Best Mode for carrying out the Invention

5 (First Embodiment)

Fig. 1 is a block diagram showing a configuration of a storage system in accordance with a first embodiment of this invention.

The storage system according to this embodiment comprises a plurality of channel adapters (CHAs) 100, a plurality of disk adapters (DKAs)
10 200, a plurality of cache memory packages (CACHEs) 300, and a processor package (MP) 400. The channel adapters 100, the disk adapters 200, the cache memory packages 300, and the processor package 400 are coupled to one another via a switch package (SW) 500. As shown in Fig. 1, the storage system generally includes the plurality of channel adapters 100, the plurality
15 of disk adapters 200, and the plurality of cache memory packages 300, but may be provided with only one channel adapter 100, one disk adapter 200, and one cache memory package 300.

The channel adapter 100 comprises a protocol controller 110 for controlling a protocol for communication with a host computer 10 and a
20 transfer controller 120 for controlling an address of data transferred from the channel adapter 100 to the cache memory package 300. The channel adapter 100 is coupled to the host computer 10 via, for example, a storage area network (SAN) that communicate by using a fibre channel protocol. It should be noted that the channel adapter 100 and the host computer 10
25 may be coupled to each other via Ethernet. In this case, the channel adapter 100 and the host computer 10 communicate with each other via an I-SCSI protocol.

The disk adapter 200 comprises a protocol controller 210 for controlling a protocol for communication with a disk drive 20 and a transfer controller 220 for controlling an address of data transferred from the disk adapter 200. The disk adapter 200 is coupled to the disk drive 20.

5 The disk drive 20 may be any one of a magnetic disk, an optical disk, a semiconductor storage device, and a magnetic tape. An interface of the disk drive 20 may be one of a fibre channel and an SATA. It should be noted that, though the disk drive 20 is shown in Fig. 1, it does not matter whether the storage system includes the disk drive 20 or the storage system
10 includes no disk drive 20 (so-called NAS).

The cache memory package 300 temporarily stores data for which write has been requested, and temporarily stores data read from the disk drive 20. Fig. 2 will be used to later describe a detailed configuration of the cache memory package 300.

15 The processor package 400 comprises a microprocessor 410. The microprocessor 410 controls an operation of the storage system (for example, transfer of data from channel adapter 100 to cache memory package 300), and manages configuration information on the storage system.

20 The switch package 500 comprises a switch for coupling among the channel adapter 100, the disk adapter 200, the cache memory package 300, and the processor package 400. Examples of the switch can be configured by a crossbar switch. The switch package 500 constitutes an internal network for coupling the channel adapter 100, the disk adapter 200, the cache memory package 300, and the processor package 400 with one
25 another. The internal network is used to transfer data among those components.

The storage system may comprise an interface (SVP) to which a

maintenance terminal is coupled.

When the channel adapter 100 receives a write request for data from the host computer 10, the protocol controller 110 converts the received write request into a protocol within the storage system, and transmits the write
5 request with the protocol having been converted, to the transfer controller 120. The transfer controller 120 analyzes an address of the received write request, and transmits results of the analysis to the processor package 400. The processor package 400 determines a transfer destination of the write request (data) on the basis of the results of analyzing the address, and
10 instructs the transfer destination to the switch package 500. The switch package 500 changes over the switch so that the data is transferred to the transfer destination specified by the processor package 400.

The write request (data) transmitted from the host computer 10 is transferred from the channel adapter 100 to the cache memory package 300,
15 and the transferred data is temporarily stored into a cache memory 320 by a cache memory adapter 310. When the data is stored into the cache memory 320, a notification of write completion is responded to the host computer 10, which can improve performance of response to the host computer 10.

The data stored in the cache memory 320 is transferred to the disk
20 adapter 200 by the cache memory adapter 310. The protocol controller 210 converts the transferred data into a protocol (for example, fibre channel protocol) used for transfer to the disk drive 20. The data with the protocol having been converted is written onto the disk drive 20.

Fig. 2 is a block diagram showing a configuration of the cache
25 memory package 300 in accordance with the first embodiment of this invention.

The cache memory package 300 comprises the cache memory adapter

310, the cache memory 320, a backup memory 330, a backup memory controller 340, a power controller 350, and a battery 360.

The cache memory adapter 310 controls an input/output of data to/from the cache memory 320.

5 The cache memory 320 is a memory for temporarily storing data input to/output from the disk drive 20, and is generally configured by a DRAM whose access speed is high. Therefore, when power is shut down, data stored in the cache memory 320 is erased unless the cache memory 320 is backed up by supplying power from the battery 360 to the cache
10 memory 320.

The cache memory adapter 310 stores data to be written onto the disk drive 20, into the cache memory 320. Among the data stored in the cache memory 320, data that is not written on the disk drive 20 is managed as dirty data. Accordingly, for each item of data stored in the cache memory
15 320, the cache memory adapter 310 manages cache memory management information in which an attribute of data (dirty data or clean data), an address thereof on the disk drive, and an address thereof in the cache memory are associated with one another.

The dirty data is stored only in the cache memory 320 within the
20 storage system, and therefore needs to be prevented from being lost when power is shut down.

The backup memory 330 is a backup-purpose memory for preventing data stored in the cache memory 320 from being lost even when power is shut down, and is generally configured by a non-volatile memory (for
25 example, flash memory). It should be noted that, instead of the non-volatile memory, a volatile memory (such as DRAM) to which backup power is supplied may be used as the backup memory 330. In addition, as shown in

Fig. 2, a packaged solid-state disk (SSD) is preferably used as the backup memory 330.

It is preferable that a single cache memory package 300 be provided with a plurality of backup memories 330. This is because resistance to a
5 fault is improved by making the backup memory 330 redundant, and data is transferred in parallel at high speed by defining a backup memory in which a data is to be saved for each cache memory 320.

The backup memory controller 340 controls an input/output of data to/from the backup memory 330. It should be noted that, in this
10 embodiment, when power is shut down, the backup memory controller 340 is supplied with the power from the battery 360, and transfers data stored in the cache memory 320 for backup.

The power controller 350 watches a power supply voltage supplied to the cache memory package 300 to thereby detect shutdown of power. The
15 battery 360 supplies power necessary for an operation of the cache memory package 300 upon the shutdown of power. It should be noted that, though the battery 360 is provided outside the cache memory package 300 in Fig. 2, the battery 360 may be provided inside the cache memory package 300.

Upon detection of the shutdown of power, the power controller 350
20 switches over the power so that the power from the battery 360 is supplied to each component of the cache memory package 300. It should be noted that, even if the power controller 350 does not switch over the power to the one supplied from the battery 360 when power is shut down, the power from the battery 360 may be automatically supplied to a power supply line when
25 power is shut down with the battery 360 having such a floating configuration as to be coupled to the power supply line.

Fig. 3 is an explanatory diagram of a storage area provided to the

backup memory 330 in accordance with the first embodiment of this invention.

The storage area of the backup memory 330 is divided into a backup address information storage area 331, a configuration information storage area 332, a data storage area 333, and a CRC area 334.

The backup address information storage area 331 stores information indicating a correlation between an address in the cache memory 320 and an address in the backup memory 330, in other words, a pair of destination address information 1100 shown in Fig. 7 and source address information 1200 shown in Fig. 8. The configuration information storage area 332 stores configuration information on the storage system. The configuration information on the storage system includes the number of respective packages implemented in the storage system, an identifier thereof, performance information thereon, and a layout of data.

The data storage area 333 stores, for the purpose of backup, user data and control information that are stored in the cache memory 320. The CRC area 334 stores an error-checking code. It should be noted that the error-checking code may be another type than a CRC type, and an error-correcting code may be stored instead of the error-checking code.

Fig. 4 is a block diagram showing a configuration of the storage system around the switch package 500 in accordance with the first embodiment of this invention.

The storage system according to this embodiment comprises the plurality of cache memory packages 300 (300A, 300B, and 300C). The cache memory packages 300A, 300B, and 300C each comprise a plurality of ports, and each of the ports is coupled to the switch package 500.

Further, in the storage system according to this embodiment, as

described later, the data stored in the cache memory 320 is written into the backup memory 330 when power is shut down, which means that the switch package 500 needs to operate for a predetermined period of time after the shutdown of power. Therefore, a battery 510 is coupled to the switch package 500, and even after the power to the storage system has been shut down, power is supplied from the battery 510 to the switch package 500. It should be noted that, though the battery 510 is provided outside the switch package 500 in Fig. 4, the battery 510 may be provided inside the switch package 500.

It should be noted that, as described above, each component of the cache memory package 300 is supplied with the power from the battery 360 when power is shut down.

Those backup power supplies 360 and 510 allow the cache memory package 300 and the switch package 500, respectively, to operate for a predetermined period of time after the shutdown of power to the storage system. Accordingly, the data stored in the cache memory 320 can be transferred to another cache memory package 300 via the switch package 500, and stored into the backup memory 330 of the another cache memory package 300.

Next, description will be made of redundancy of a cache memory package.

The cache memory packages 300A and 300B form a pair. In other words, the same data is stored in the cache memory 320 of the cache memory package 300A (CMPK00) and the cache memory 320 of the cache memory package 300B (CMPK01). The cache memory package 300 is thus made redundant to thereby achieve safety against data loss. In addition, loads on access are balanced to thereby improve performance of response to

the host computer 10.

In this case, as described later, the data stored in the cache memory 320 of any one of the cache memory packages 300A and 300B is the only data to be stored into the backup memory 330 for backup when power is shut down.

Fig. 5 is a block diagram showing a configuration of the storage system around the power supply line in accordance with the first embodiment of this invention.

During a normal operation, power is supplied from a power supply 600 to the cache memory package 300 and the switch package 500. The power supply 600 converts utility power supply of alternating current to the storage system into DC power having a voltage necessary for each component of the storage system, and outputs the DC power obtained after conversion to a power supply line 620.

If a power switch 610 of the power supply 600 is shut down, if the supply of AC power to the power supply 600 is interrupted, or if the power supply 600 breaks down, a voltage applied to the power supply line 620 is decreased. The power controller 350 of the cache memory package 300 watches a power supply voltage supplied to the cache memory package 300 to thereby detect lowering of the power supply voltage.

It should be noted that, instead of the power controller 350 watching a power supply voltage, in such a configuration that the power supply 600 outputs a signal indicating a power supply abnormality, the power controller 350 may detect the decreasing of the power supply voltage by receiving the signal indicating a power supply abnormality output from the power supply 600.

Fig. 6 is an explanatory diagram of SSD address allocation

information 1000 in accordance with the first embodiment of this invention.

The SSD address allocation information 1000 is one item of the configuration information on the storage system, and is defined for each cache memory package 300. The SSD address allocation information 1000
5 is managed by the processor package 400.

The SSD address allocation information 1000 is information indicating an area of the backup memory 330 provided to each cache memory package 300, which is allocated to the cache memory package 300, and includes an identifier 1001 of a backup memory and an allocation
10 address area 1002 thereof.

The identifier 1001 of a backup memory is an identifier of the backup memory 330 which is unique across the entire storage system. The allocation address area 1002 indicates an address range of the backup memory 330, which is allocated to the cache memory package 300 that uses
15 the SSD address allocation information 1000.

In other words, each cache memory package 300 can use an address in the backup memory 330 within a range defined by the SSD address allocation information 1000 to save data stored in its own cache memory
320.

20 Fig. 7 is an explanatory diagram of the destination address information 1100 in accordance with the first embodiment of this invention.

The destination address information 1100 is generated along with the source address information 1200 shown in Fig. 8 in Step S104 of a saving process shown in Fig. 9. In addition, a pair of the destination address
25 information 1100 and the source address information 1200 constitute backup address information.

The destination address information 1100 is information indicating

an address of the transfer destination of the data stored in the cache memory 320, and includes an identifier 1101 of a cache memory package, an identifier 1102 of a backup memory, and an address 1103 in the backup memory.

5 The identifier 1101 of a cache memory package is a unique identifier of the cache memory package 300 at the transfer destination of the data. The identifier 1102 of a backup memory is a unique identifier of the backup memory 330 at the transfer destination of the data. The address 1103 in the backup memory is an address of the transfer destination of the data in
10 the backup memory 330.

Fig. 8 is an explanatory diagram of the source address information 1200 in accordance with the first embodiment of this invention.

The source address information 1200 is information indicating an address at which data to be transferred is stored in the cache memory 320,
15 and includes an identifier 1201 of a cache memory package, an identifier 1202 of a cache memory, and an address 1203 in the cache memory.

The identifier 1201 of a cache memory package is a unique identifier of the cache memory package 300 in which data to be transferred to the backup memory 330 is stored. The identifier 1202 of a cache memory is a
20 unique identifier of the cache memory 320 in which the data to be transferred is stored. The address 1203 in the cache memory is an address at which the data to be transferred is stored in the cache memory 320.

Fig. 9 is a flowchart showing the saving process of writing data into the backup memory 330 in accordance with the first embodiment of this
25 invention.

First, at startup of the storage system, the microprocessor 410 stores the backup memory address allocation information 1000 into a configuration

information storage register within the cache memory adapter 310. It should be noted that, also when the storage system has its configuration changed, the microprocessor 410 generates the SSD address allocation information 1000, and stores the generated SSD address allocation information 1000 into the configuration information storage register (S101).
5 It should be noted that the configuration information storage register is a storage area provided to the cache memory adapter 310.

After that, when the power controller 350 detects shutdown of power (S102), the power is switched over so that the power from the battery 360 is
10 supplied to each component of the cache memory package 300 (S103).

Then, the backup memory controller 340 reads the SSD address allocation information 1000 from the configuration information storage register, and generates backup address information for storing the data stored in the cache memory 320 into the backup memory 330 (S104). The
15 data stored in the cache memory 320 is divided into a predetermined number of blocks (for example, write blocks) to be stored into different backup memories 330, and hence the backup address information (pair of destination address information 1100 and source address information 1200) is generated for each block of the divided data. It should be noted that the
20 data can be divided into not only blocks but also bits, sectors, or other various units.

It should be noted that, in this embodiment, cache data is desirably stored into the backup memories 330 provided to different cache memory packages 300, but instead, the cache data may be divided to be stored into a
25 plurality of backup memories 330 provided to one cache memory package 300.

Further, as described above, in a case where two or more cache

memory packages 300 form a pair, the same data is stored in the cache memories 320 of the cache memory packages 300 that form the pair, and hence the data stored in the cache memory 320 of one of the cache memory packages 300 that form the pair may be stored into the backup memory 330.

5 After that, the backup memory controller 340 reads the configuration information on the storage system from the configuration information storage register, and stores the read configuration information into the configuration information storage area 332 of the backup memory 330. Further, the backup memory controller 340 stores the generated destination
10 address information 1100 and source address information 1200 into the backup address information storage area 331 of the backup memory 330 (S105).

 In addition, the backup memory controller 340 reads data from the cache memory 320 on an address basis, and on the basis of the destination
15 address information 1100 corresponding to an address at which the data has been read, specifies an address at which the read data is written into the backup memory 330, and writes the read data into the data storage area 333 of the backup memory 330 at the specified address (S106).

 After Step S106 has been executed on all of storage areas (at all of
20 addresses) of the cache memory 320, the saving of the data stored in the cache memory 320 completes normally, which is a completion of the saving process (S107). On the other hand, if Step S106 cannot be executed on a part of storage areas (at part of addresses) of the cache memory 320, the saving of the data stored in the cache memory 320 stops abnormally, which
25 is a failure of the saving process (S108).

 After that, the backup memory controller 340 stores a saving process execution status regarding the data stored in the cache memory 320 into the

configuration information storage area 332 of the backup memory 330 (S109). The saving process execution status has values such as "normal end", "abnormal end", and "unprocessed". In other words, if the saving of data ends with regard to all of storage areas of the cache memory 320, the saving process execution status becomes "normal end". On the other hand, if the saving of data cannot be executed on a part of storage areas of the cache memory 320, the saving process execution status becomes "abnormal end".

After that, the backup memory controller 340 instructs the power controller 350 to shut down the supply of power from the battery 360. The power controller 350 shuts down the supply of power from the battery 360 to each component of the cache memory package 300 (S110). Then, the storage system stops its operation (S111).

As described above, when the power is abnormal, the data stored in the cache memory 320 is stored into the backup memory 330, which holds stored data even when the power is shut down. Accordingly, it is not necessary to provide a large-capacity battery in order to back up data during long-term shutdown of power, which can reduce the cost of the battery.

Further, the data stored in the cache memory 320 of one of the cache memory packages 300 is divided to be stored into the backup memory 330 of a plurality of cache memory packages 300. Accordingly, the data stored in the original cache memory 320 cannot be restored only with the data stored in one of the backup memories 330, which can ensure the secrecy of data.

For example, according to the conventional technology, in a state in which data is stored in the backup memory 330 (for example, in case of service interruption), the cache memory package 300 (or only backup memory 330) is removed from the storage device, and then the data stored in

the removed backup memory 330 is read, thereby allowing data that was stored in the cache memory 320 to be known. However, by dividing, as in this embodiment, the data stored in the cache memory 320 of one of the cache memory packages 300 so as to be stored into the backup memory 330 of a plurality of cache memory packages 300, it is impossible to know the data that was stored in the cache memory 320 even if the data stored in one of the backup memories 330 can be read.

In addition, only the data stored in the cache memory 320 of a part of the cache memory packages 300 (one of cache memory packages 300 that form pair) is stored into the backup memory 330, thereby reducing the amount of cache data to be saved upon the shutdown of power, which can reduce the period of time necessary for the saving process. Accordingly, it is possible to reduce the capacity of the battery used for supplying power after the shutdown of power. Further, it is possible to reduce the period of time necessary for a recovery process of recovering the data stored in the backup memory 330 into the cache memory 320. In addition, a general flash memory has a limited number of rewrites, and hence, in a case where the flash memory is used as the backup memory 330, it is possible to reduce the number of rewrites of the backup memory 330, suppress deterioration of the flash memory, and decrease an exchange cycle of the backup memory 330.

Figs. 10 and 11 are flowcharts showing the recovery process of writing the data stored in the backup memories 330 into the cache memory 320 in accordance with the first embodiment of this invention. It should be noted that Fig. 10 shows a process executed by the processor package 400, while Fig. 11 shows a process executed by the cache memory package 300.

First, description will be made of the process shown in Fig. 10 which

is executed by the microprocessor 410 of the processor package 400.

When the power switch 610 of the power supply 600 of the storage system is operated, power starts to be supplied to the storage system (S201). After that, the microprocessor 410 reads the configuration information from
5 the configuration information storage area 332 of the backup memory 330, and initializes each component of the storage system (S202).

After that, the microprocessor 410 reads the saving process execution status from the configuration information storage area 332 of the backup memory 330 (S203). Then, if the read saving process execution status is
10 "normal end" (YES in Step S204), the procedure advances to Step S205 to recover the cache data into the cache memory 320. On the other hand, if the read saving process execution status is not "normal end" (NO in Step S204), the procedure advances to Step S207 to end a device startup process without the need to recover the cache data into the cache memory 320.

15 In other words, if the read saving process execution status is not "normal end", the data stored in the backup memory 330 is not the one that has been obtained by normally backing up the cache data that was stored in the cache memory 320, and hence the cache memory 320 is judged to be in a volatile state, which leads to the startup of the storage system.

20 In Step S205, the microprocessor 410 issues an instruction to recover the cache data. To be specific, the microprocessor 410 writes a recovery command into a predetermined area (register) of the cache memory adapter 310. The written recovery command is read by the cache memory package 300 (Step S301 of Fig. 11).

25 After that, the microprocessor 410 watches the end of the recovery process. To be specific, the microprocessor 410 uses a polling method to watch recovery completion status information stored in the predetermined

area (internal register) of the cache memory adapter 310 (S206).

When the microprocessor 410 detects that the recovery process has come to an end on the basis of a value (recovery process execution status) of an inner register written in Step S305 of Fig. 11, the device startup process comes to an end (S207).

Next, description will be made of the process shown in Fig. 11 which is executed by the backup memory controller 340 of the cache memory package 300.

When the backup memory controller 340 is started and initialized, the backup memory controller 340 waits for the instruction to recover the cache data (issued in Step S205 of Fig. 10) from the microprocessor 410 (S301). Then, if the backup memory controller 340 receives the instruction to recover the cache data (if the recovery command is written into the predetermined area (internal register of the backup memory controller 340)), the backup memory controller 340 generates recovery-purpose address information for storing the data stored in the backup memory 330 into the cache memory 320 (S302). To be specific, the destination address information 1100 as shown in Fig. 7 and the source address information 1200 as shown in Fig. 8, which have been stored in Step S105, are read from the backup address information storage area 331 of the backup memory 330. In other words, in the recovery-purpose address information, the destination address information 1100 is set as a read source address, and the source address information 1200 is set as a write destination address.

Subsequently, the backup memory controller 340 read the saved data from the backup memory 330 according to the generated recovery-purpose address information (destination address information 1100), and store the read data into the cache memory 320 according to the generated

recovery-purpose address information (source address information 1200) (S303).

Then, after data has been written for all of the storage areas of the cache memory 320, the backup memory controller 340 completes the data
5 recovery into the cache memory 320 (S304).

Subsequently, the backup memory controller 340 stores the recovery process execution status for the data recovery into the cache memory 320, into the configuration information storage area 332 of the backup memory 330 (S305). The recovery process execution status has values such as
10 "normal end" and "abnormal end". For example, if the data recovery has been completed for all of the storage areas of the cache memory 320, the recovery process execution status becomes "normal end".

After that, the backup memory controller 340 deletes the saving process execution status regarding the data written in Step S109 of Fig. 9
15 (S306). This is because, if the saving process execution status regarding the data cannot be stored next time power is shut down, the storage system is prevented from starting with erroneous information at the subsequent startup.

Fig. 12 is a block diagram showing a configuration of a storage
20 system around the switch package 500 in accordance with a modified example of the first embodiment of this invention.

As described above, in the storage system according to this embodiment, the data stored in the cache memory 320 is written into the backup memory 330 upon the shutdown of power. Therefore, even after the
25 power supplied to the storage system has been shut down, data needs to be transferred between the cache memory packages 300.

In the modified example shown in Fig. 12, in addition to the switch

package 500 (not shown), there is provided a communication line 700 for transferring data between the cache memory packages 300 during the shutdown of power. For example, a PCI Express can be used as the communication line 700.

5 Accordingly, in the same manner as the case shown in Fig. 4 where the switch package 500 is operated by the power supplied from the battery 510, data can be transferred between the cache memory packages 300 for a predetermined period after the power supplied to the storage system has been shut down. Therefore, this modified example eliminates the need to
10 provide the switch package 500 with the battery 510.

(Second Embodiment)

In the first embodiment described above, the backup memory 330 is used only when the power is shut down, but is not used during the normal operation. However, in the second embodiment, in addition to the operation
15 upon the shutdown of power according to the first embodiment described above, the backup memory 330 is used as a secondary cache during the normal operation.

Fig. 13 is an explanatory diagram of the operation of the backup memory 330 during the normal operation in accordance with the second
20 embodiment of this invention.

When the configuration information and control information on the device are received from the processor package 400, the cache memory package 300 stores the received configuration information and control information into the cache memory 320 as well as user data written from the
25 host computer 10.

When the configuration information and the control information are stored into the cache memory 320, the backup memory controller 340

divides the configuration information and the control information that are stored in the cache memory 320, and stores the divided data into the backup memories 330 of a plurality of cache memory packages 300.

The configuration information and the control information that are stored in the backup memories 330 are held even during the shutdown of power, and therefore do not need to be stored onto the disk drive 20. In this case, after having been stored into the backup memories 330, the configuration information and the control information become clean data, and can be deleted from the cache memory 320.

In a case where the backup memory 330 and the disk drive 20 are configured by a non-volatile semiconductor storage device (such as flash memory) and a magnetic disk, respectively, the backup memory 330 can have data input to/output from the disk drive 20 at high speed, which can improve accessibility to the configuration information and the control information.

It should be noted that the configuration information and the control information that are stored in the backup memories 330 may be stored onto the disk drive 20. In this case, after having been stored onto the disk drive 20, the configuration information and the control information become clean data, and can be deleted from the cache memory 320.

The configuration information and the control information that are stored in the cache memory 320 do not need to be processed at high speed in the cache memory package 300 unlike the user data written from the host computer 10. Therefore, even if the configuration information and the control information that are stored in the cache memory 320 are stored into the backup memory 330 during the normal operation, little influence is exerted upon the response to the host computer 10.

The configuration information and the control information may be divided into a predetermined number of units (for example, bits, sectors, or blocks) as in the manner described above in the first embodiment that the data stored in the cache memory 320 is divided upon the shutdown of power.

Fig. 14 is a flowchart showing a process of writing secondary cache data into the backup memory 330 in accordance with the second embodiment of this invention.

First, the microprocessor 410 stores the SSD address allocation information 1000 into the configuration information storage register of the cache memory adapter 310. In the second embodiment, for each of the backup memories 330, an address at which the configuration information is stored into the configuration information storage area 332 is set, and an address at which the control information is stored into the data storage area 333 is set (S401).

After that, the microprocessor 410 writes data into a predetermined area of the backup memory controller 340 (for example, enables store bit of backup memory controller 340). When the store bit is enabled, the backup memory controller 340 stores the configuration information and the control information into the backup memory 330 (S402).

Subsequently, if the backup memory controller 340 confirms that the store bit has been enabled, the backup memory controller 340 reads the configuration information from the configuration information storage register, generates the backup address information that stores the read configuration information, and further generates the backup address information that stores the read control information (S403). The configuration information and the control information that are stored in the cache memory 320 are

divided into a predetermined number of blocks (for example, write blocks) to be stored into different backup memories 330, and hence the backup address information (pair of destination address information 1100 and source address information 1200) is generated for each item of the divided
5 data. It should be noted that, also in the second embodiment, the data can be divided into not only blocks but also bits, sectors, or other various units.

After that, the backup memory controller 340 stores the configuration information read from the configuration information storage register into the configuration information storage area 332 of the backup memory 330.
10 Further, the backup memory controller 340 stores the destination address information 1100 and the source address information 1200 that are included in the generated configuration information into the backup address information storage area 331 of the backup memory 330 (S404).

Further, the backup memory controller 340 stores the control
15 information into the data storage area 333 of the backup memory 330. In addition, the backup memory controller 340 stores the destination address information 1100 and the source address information 1200 that are included in the generated control information into the backup address information storage area 331 of the backup memory 330 (S405).

20 After Step S404 has been executed on all of the configuration information storage registers (all of configuration information) and control information of the cache memories 320, the storing of the configuration information and the control information that are stored in the cache memory 320 into the backup memory 330 is completed (S406).

25 After that, the backup memory controller 340 stores the saving process execution status regarding the configuration information and the control information that are stored in the cache memory 320 into a

predetermined area (inner register of backup memory controller 340) (S407), and the backup memory controller 340 ends the saving process for the control information.

On the other hand, the microprocessor 410 watches the end of the
5 recovery process at such intervals as to exert no influence on the response. To be specific, the microprocessor 410 uses the polling method to watch information on the saving process execution status stored in the predetermined area (inner register of backup memory controller 340). Then, upon detection of the end of the recovery process based on the value
10 (information on saving process execution status) of the inner register, the microprocessor 410 ends the saving process for the control information (S408).

As has been described above, according to the second embodiment of this invention, the configuration information and the control information
15 that are stored in the cache memory 320 are stored into the backup memory 330 during the normal operation, and hence it is possible to reduce the amount of data to be saved upon the shutdown of power, which can reduce the period of time necessary for the saving process. It should be noted that the second embodiment has been described by taking the example case
20 where both the configuration information and the control information are stored into the backup memory 330, but only one of the configuration information and the control information may be stored into the backup memory 330.

Further, in the second embodiment, the configuration information
25 and the control information are described as examples of such data as to exert no influence on the response to the host computer 10, but the second embodiment can be applied to other such data as to exert no influence on

the response to the host computer 10.

Further, in the case where the backup memory 330, in which the configuration information and the control information are stored, and the disk drive 20 are configured by a non-volatile semiconductor storage device (such as flash memory) and a magnetic disk, respectively, it is possible to improve the access speed with respect to the configuration information and the control information that are stored in the backup memories 330. In addition, the configuration information and the control information are divided to be stored in a plurality of backup memories 330, and hence it is possible to further improve the accessibility than in the case where the configuration information and the control information are stored in one backup memory 330.

Industrial Applicability

This invention can be applied to a storage system including a cache memory, in which data stored in the cache memory does not volatilize (disappear) even when power is shut down.

CLAIMS

1. A storage system, which is coupled to a computer, comprising:

a storage device for storing data for which write is requested by the

5 computer;

a controller for controlling transfer of data in the storage system;

a plurality of cache memory units for temporarily storing data to be
stored into the storage device; and

a connecting unit for connecting at least the plurality of cache
10 memory units,

each of the plurality of cache memory units comprising:

a cache memory for storing data;

an auxiliary storage device for holding data even after
shutdown of power; and

15 a cache controller for controlling an input/output of
data to/from the cache memory and the auxiliary storage device,

wherein the cache controller is configured to store data stored in the
cache memory, which is divided into a plurality of parts, into a plurality of
the auxiliary storage devices included in the plurality of cache memory units.

20

2. The storage system according to Claim 1, further comprising a
watching unit for watching power supplied to the plurality of cache memory
units,

wherein the cache controller is configured to store the data stored in
25 the cache memory into the plurality of the auxiliary storage devices in case
of which the watching unit detects an abnormality in the power supplied to
the plurality of cache memory units.

3. The storage system according to Claim 1, wherein each of the plurality of the auxiliary storage devices comprises a non-volatile memory for holding data even after the shutdown of the power.

5

4. The storage system according to Claim 1, wherein:
the plurality of cache memory units comprise:

a first cache memory unit; and

a second cache memory unit for storing the same data as

10 stored in the first cache memory unit; and

the cache controller is configured to:

store data stored in the cache memory of the first cache memory unit
into the plurality of the auxiliary storage devices; and

prevent to store data stored in the cache memory of the second cache
15 memory unit into the plurality of the auxiliary storage devices.

5. The storage system according to Claim 1, further comprising an auxiliary connecting unit for connecting the plurality of cache memory units, wherein:

20 the connecting unit is configured to stop the transfer of data when power supplied to the connecting unit stops; and

the auxiliary connecting unit is configured to enable the transfer of data between the plurality of cache memory units when the power supplied to the connecting unit stops.

25

6. The storage system according to Claim 1, further comprising a watch unit for watching power supplied to the plurality of cache memory

units,

wherein the cache controller is configured to stores at least one of configuration information and control information stored in the cache memory, which is divided into a plurality of parts, into the plurality of the auxiliary storage devices before the watch unit detects an abnormality in the power supplied to the plurality of cache memory units.

7. A data management method, which is used for a storage system, the storage system including: a storage device for storing data for which write is requested by a computer; a controller for controlling transfer of data in the storage system; a plurality of cache memory units for temporarily storing data to be stored into the storage device; and a connecting unit for connecting at least the plurality of cache memory units,

each of the plurality of cache memory units including: a cache memory for storing data; an auxiliary storage device for holding data even after shutdown of power; and a cache controller for controlling an input/output of data to/from the cache memory and the auxiliary storage device,

the data management method including the steps of:
reading, by the cache controller, the data stored in the cache memory;

dividing, by the cache controller, the read data into a plurality of parts; and

controlling, by the cache controller, each of the parts of the divided data to be stored into any one of a plurality of the auxiliary storage devices included in the plurality of cache memory units.

8. The data management method according to Claim 7, further including the step of storing, by the cache controller, the data stored in the cache memory into the plurality of the auxiliary storage devices in the case of which an abnormality is detected in power supplied to the plurality of
5 cache memory units.

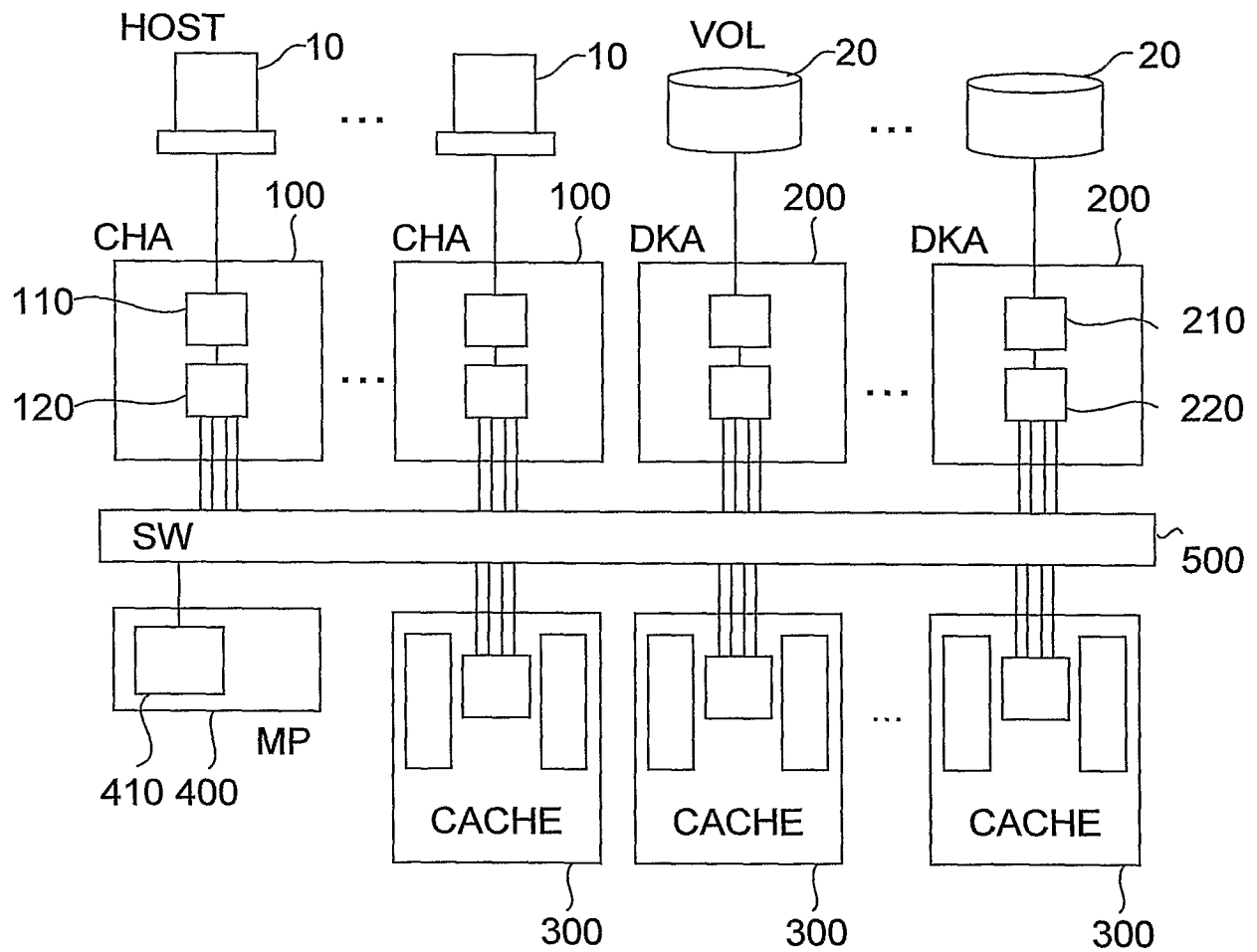
9. The data management method according to Claim 7, wherein:
the plurality of cache memory units include: a first cache memory unit; and a second cache memory unit for storing the same data as stored in
10 the first cache memory unit; and

the data management method further including the steps of:
storing, by the cache controller, data stored in the cache memory of the first cache memory unit into the plurality of the auxiliary storage devices; and
15 preventing, by the cache controller, to store data stored in the cache memory of the second cache memory unit from being stored into the plurality of the auxiliary storage devices.

10. The data management method according to Claim 7, further
20 including the step of storing, by the cache controller, at least one of configuration information and control information stored in the cache memory, which is divided into a plurality of parts, into the plurality of the auxiliary storage devices before an abnormality is detected in power supplied to the plurality of cache memory units.

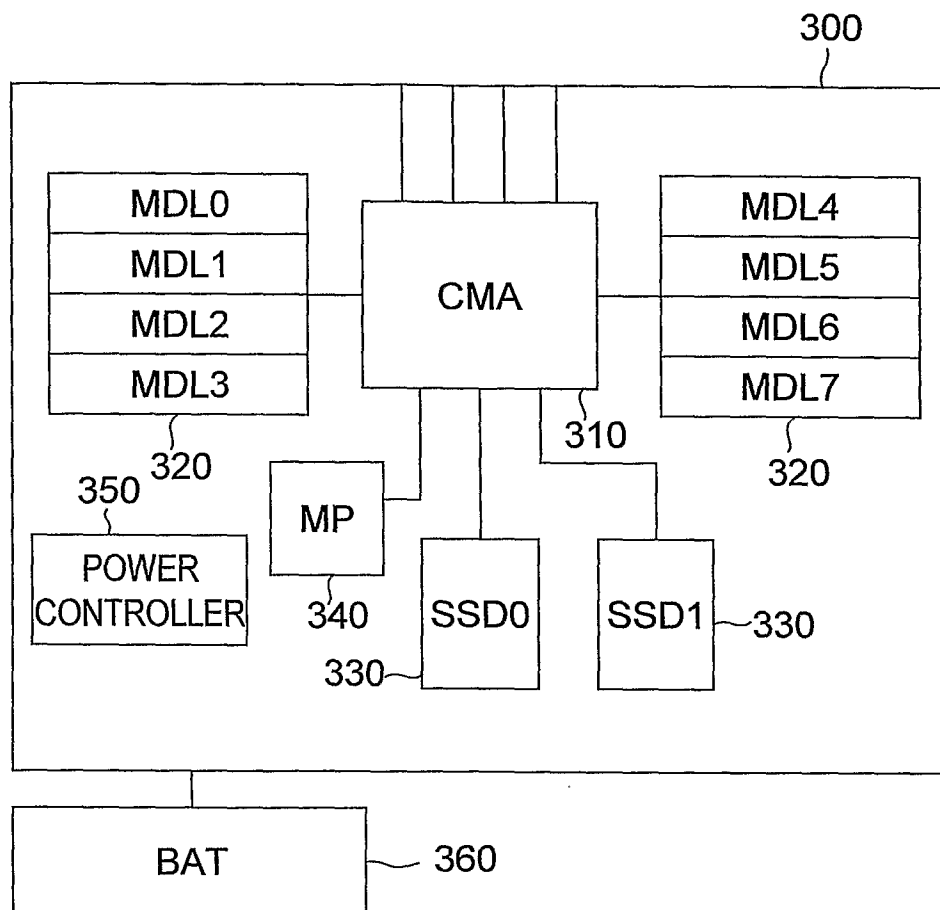
1/12

Fig. 1



2/12

Fig. 2



3/12

Fig. 3

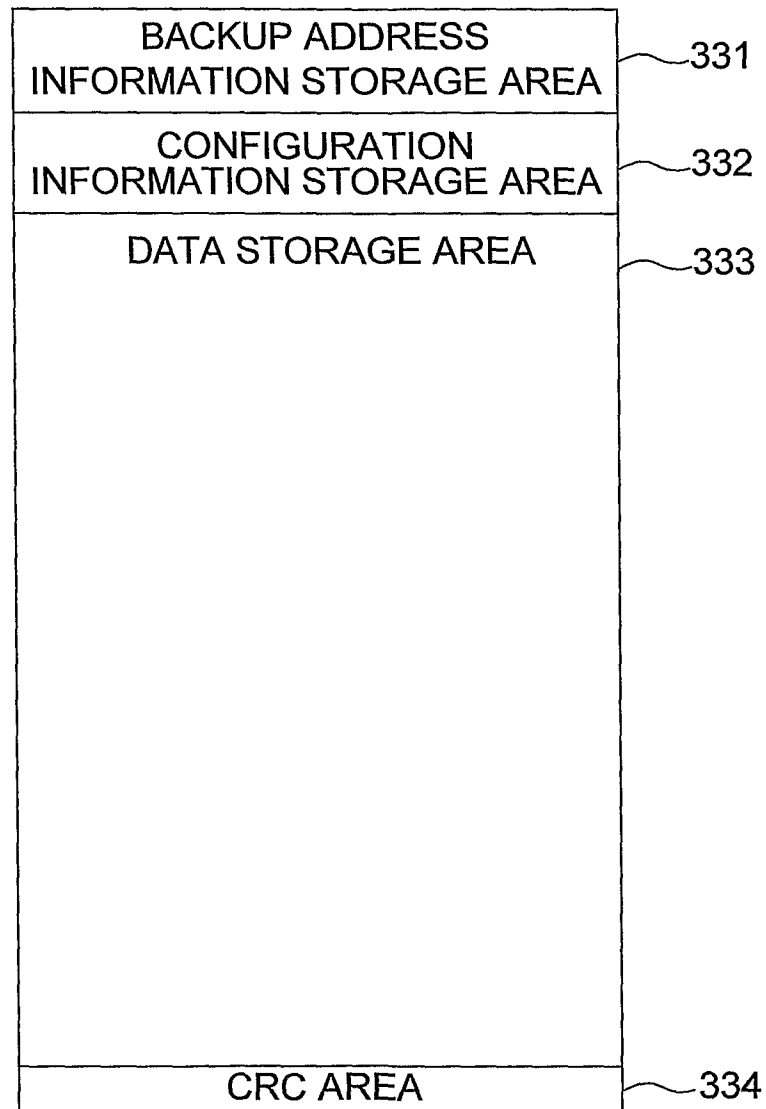
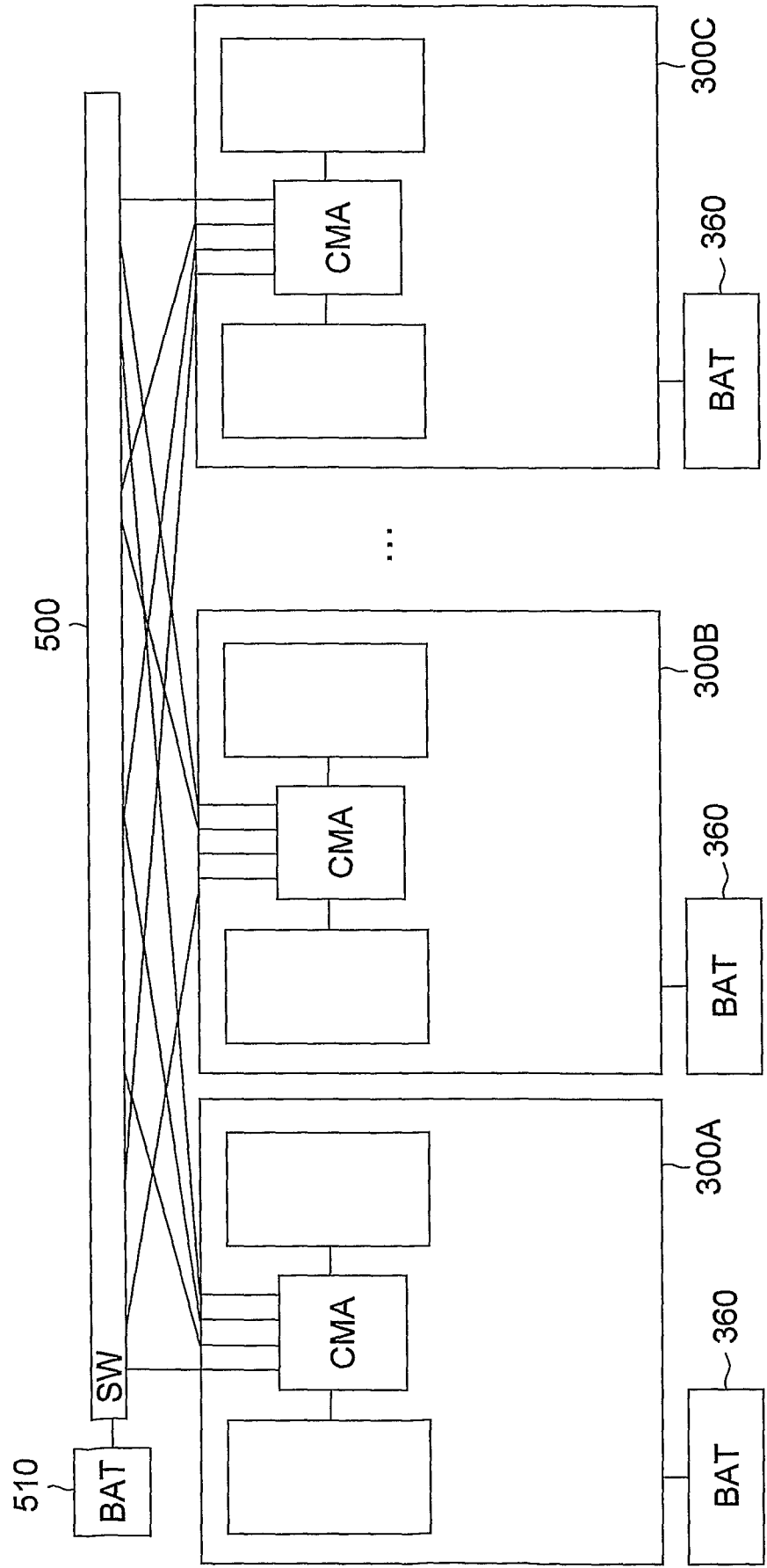


Fig. 4



5/12

Fig. 5

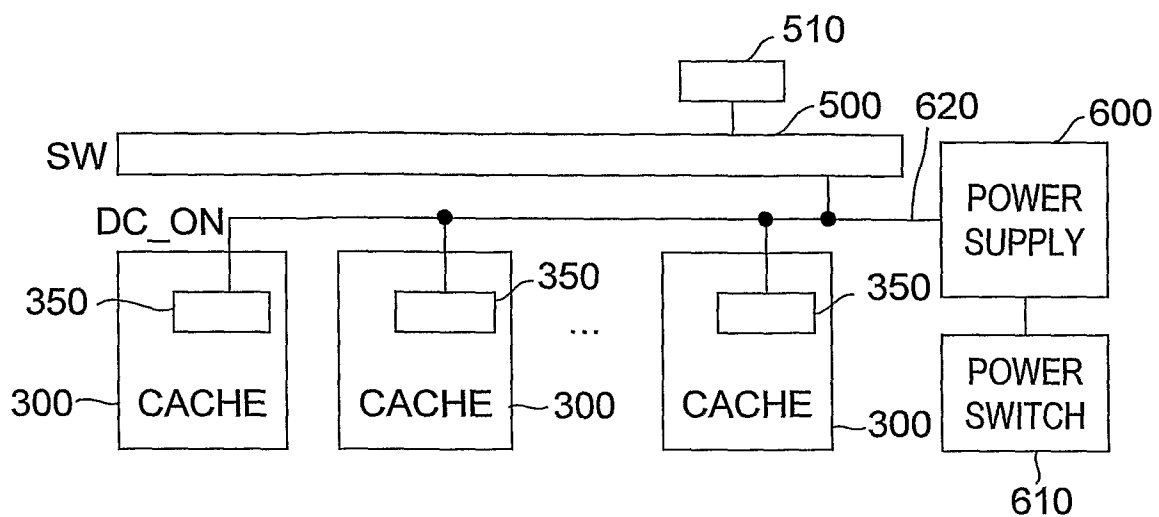


Fig. 6

SSD ADDRESS ALLOCATION INFORMATION		1000
1001	1002	
SSD 0	ADDRESS AREA	
SSD 1	ADDRESS AREA	
...	...	
SSD N	ADDRESS AREA	

6/12

Fig. 7

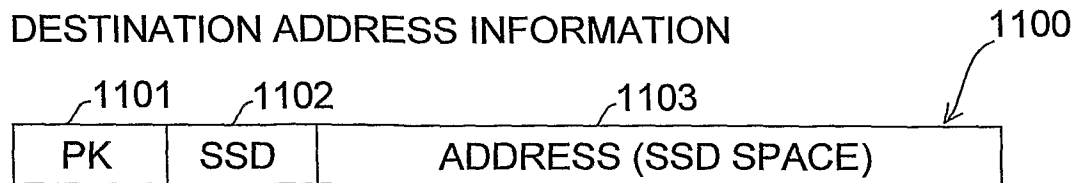
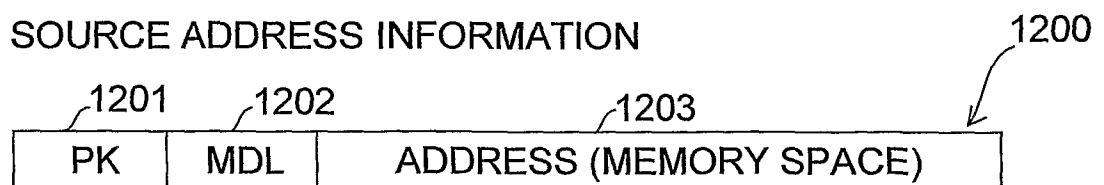
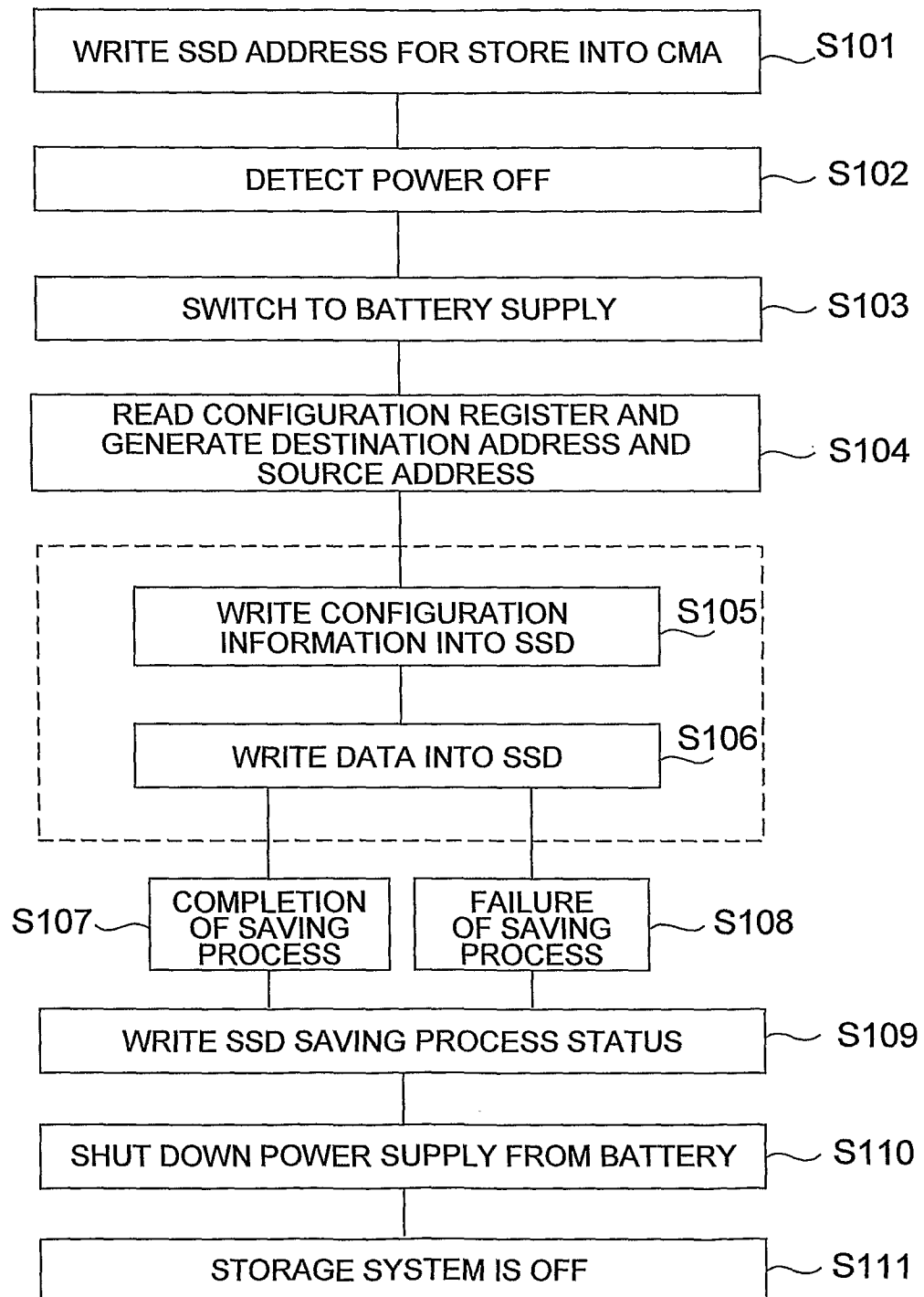


Fig. 8



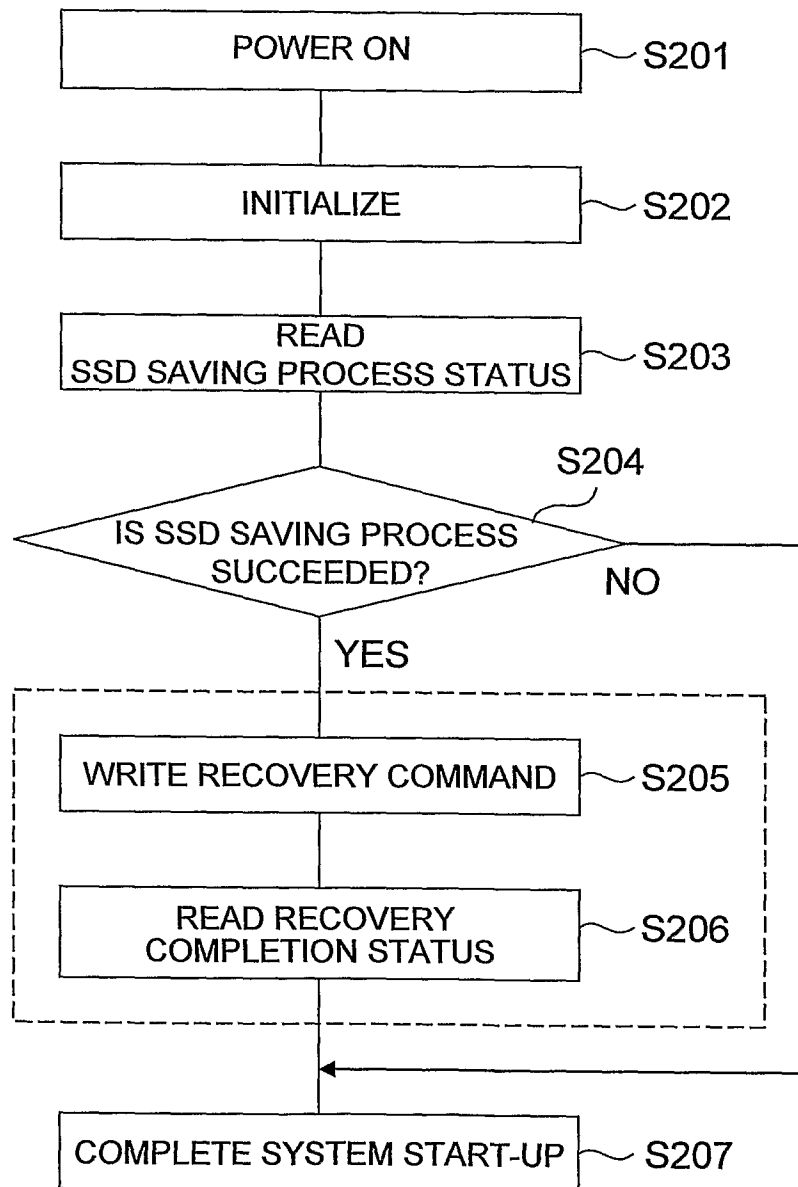
7/12

Fig. 9



8/12

Fig. 10



9/12

Fig. 11

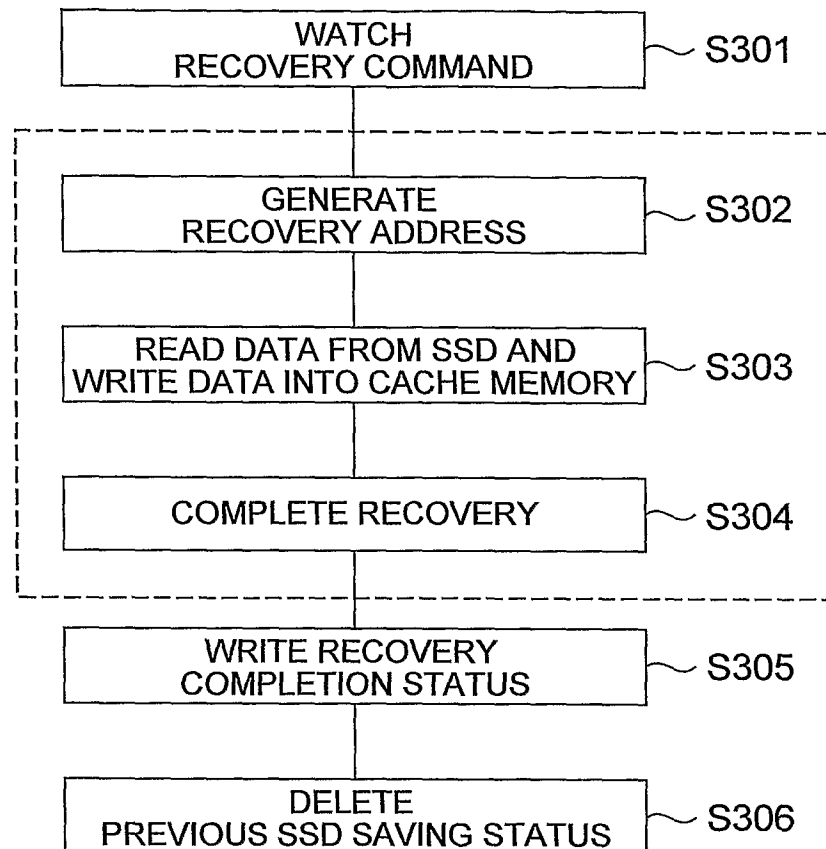
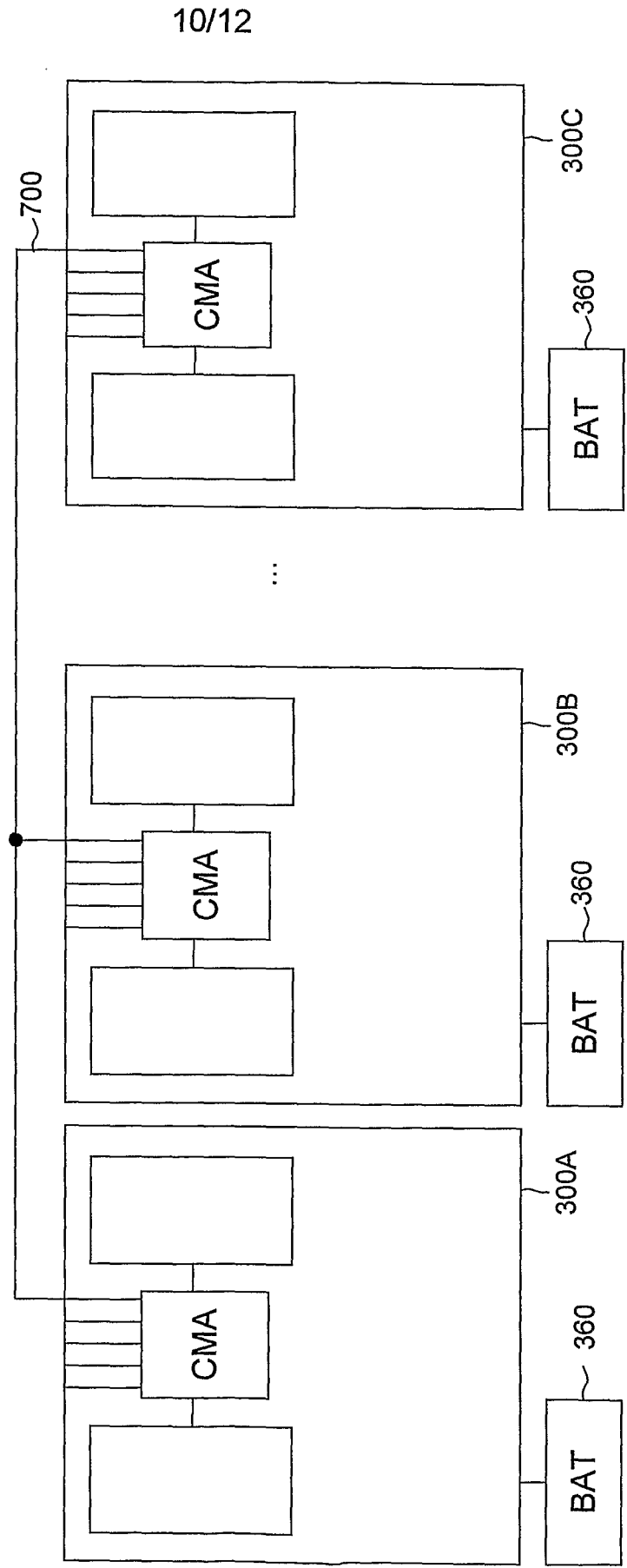
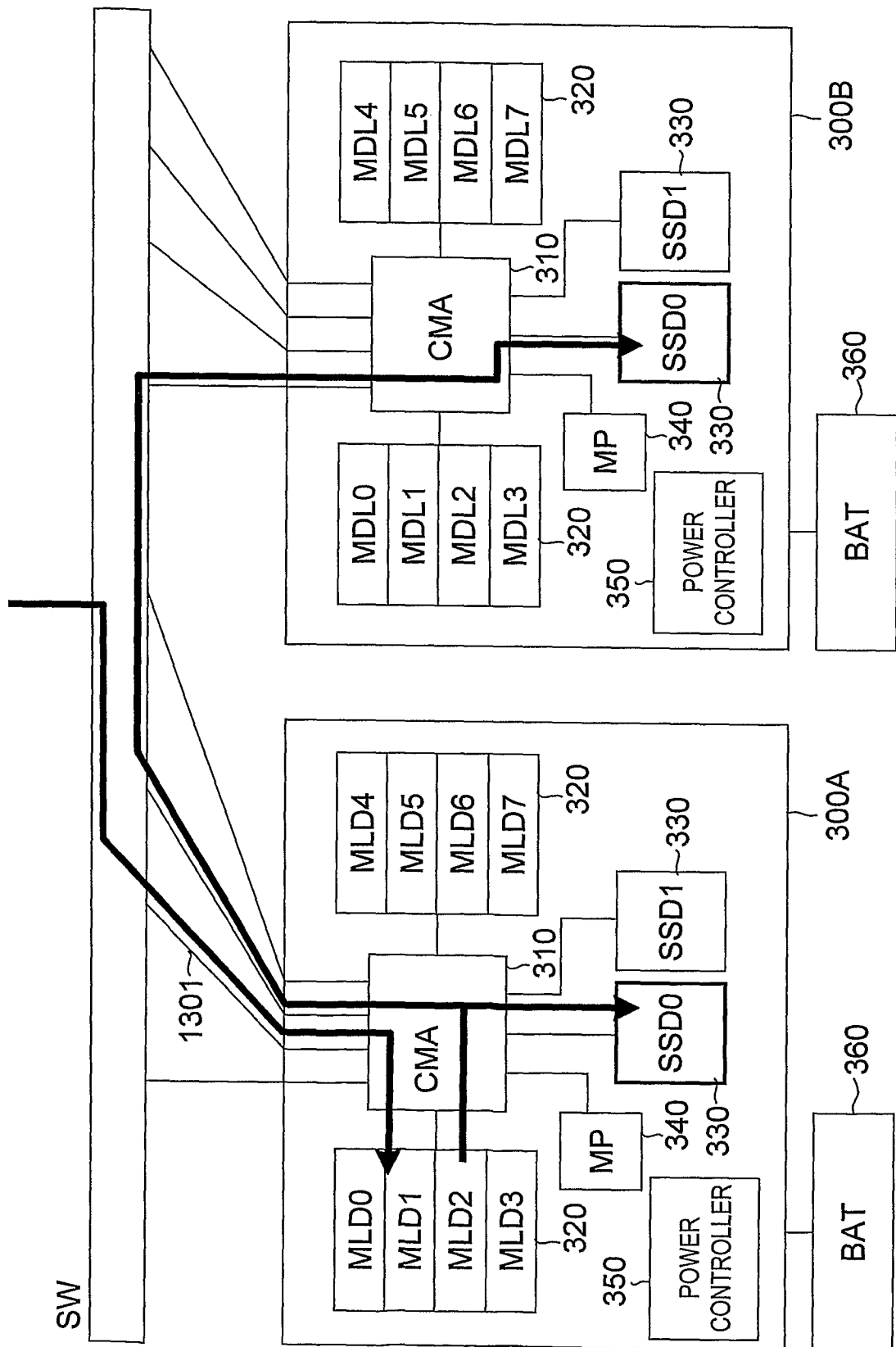


Fig. 12



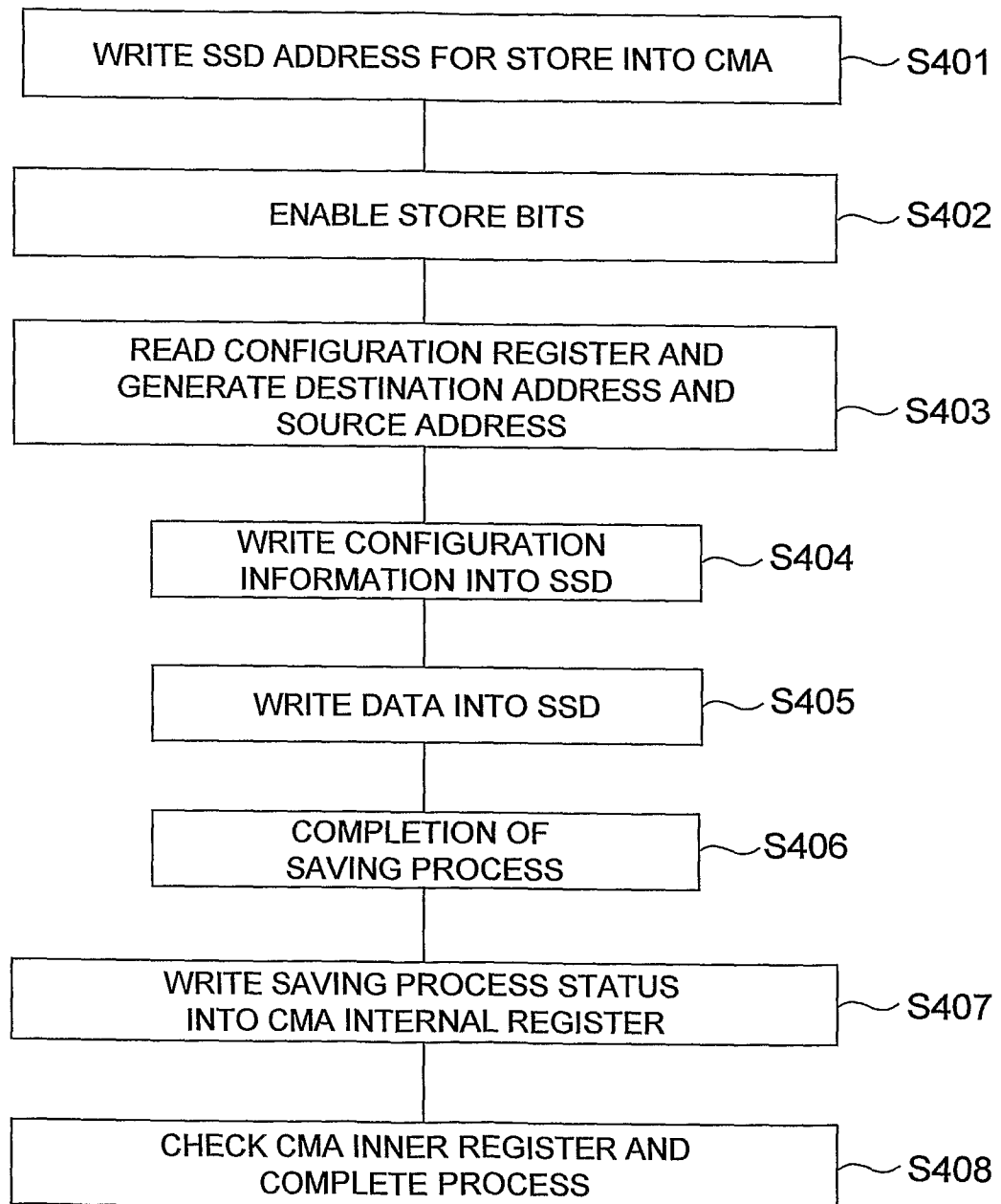
11/12

Fig. 13



12/12

Fig. 14



INTERNATIONAL SEARCH REPORT

International application No

PCT/JP2008/068690

A. CLASSIFICATION OF SUBJECT MATTER
 INV. G06F11/20 G06F12/08

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

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 practical, search terms used)

EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	GB 2 256 735 A (INTEL CORP [US]) 16 December 1992 (1992-12-16) figures 1,3 page 6, lines 14-16 page 9, lines 9-13 page 10, lines 10-12 page 10, lines 20-22 page 17, lines 12-14 claim 1	1-10
Y	US 2003/041215 A1 (GEORGE ROBERT T [US] ET AL) 27 February 2003 (2003-02-27) figure 1 paragraphs [0014], [0015]	1-10



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 document 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

16 June 2009

Date of mailing of the international search report

23/06/2009

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

Filip, Liviu

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/JP2008/068690

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
GB 2256735	A	16-12-1992	US	5519831 A	21-05-1996
US 2003041215	A1	27-02-2003	CN	1549973 A	24-11-2004
			EP	1421499 A1	26-05-2004
			WO	03019384 A1	06-03-2003