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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
10 May 2013 (10.05.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/063777 A1

(51) **International Patent Classification**:
G06F 11/10 (2006.01) *H03M 13/00* (2006.01)
G11C 29/00 (2006.01)

(21) **International Application Number**:
PCT/CN2011/081693

(22) **International Filing Date**:
2 November 2011 (02.11.2011)

(25) **Filing Language**: English

(26) **Publication Language**: English

(71) **Applicant** *(for all designated States except US)*: **EMPIRE TECHNOLOGY DEVELOPMENT LLC** [US/US]; 2711 Centerville Road, Suite 400, Wilmington, DE 19808 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **CAO, Junwei** [CN/CN]; 7A Building #2, No.2 Shangdi Xinxi Road, Haidian District, Beijing 100085 (CN). **LIN, Xiao** [CN/CN]; 7A Building #2, No.2 Shangdi Xinxi Road, Haidian District, Beijing 100085 (CN).

(74) **Agent**: **CHINA SCIENCE PATENT & TRADEMARK AGENT LTD.**; 25/F., Bldg. B, Tsinghua Tongfang Hi-Tech Plaza, No.1, Wangzhuang Rd., Haidian District, Beijing 100083 (CN).

(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, 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, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, 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, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

Published:

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

WO 2013/063777 A1

(54) **Title**: TRIPLE PARITY ENCODING TO FACILITATE DATA RECOVERY

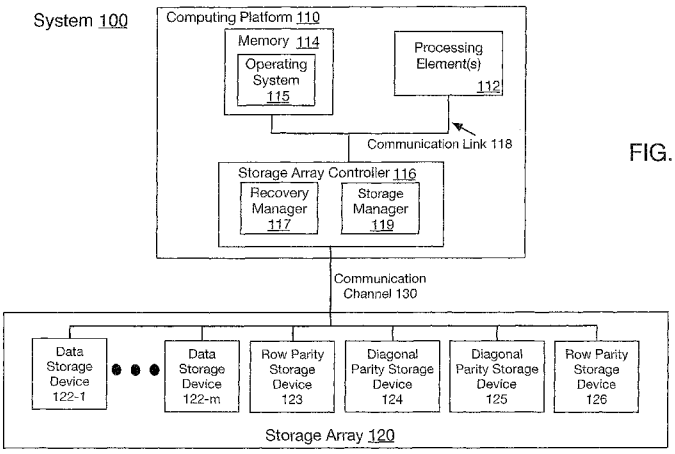


FIG. 1

(57) **Abstract**: Examples are disclosed for facilitating recovery from failures associated with a storage array having a plurality of storage devices.

Triple Parity Encoding to Facilitate Data Recovery

BACKGROUND

[0001] Unless otherwise indicated herein, the approaches described in this section are not prior art to the claims in this application and are not admitted to be prior art by inclusion in this section.

[0002] Storage systems for storing or maintaining data may include various data storage devices such as hard disk drives (HHDs), solid state drives (SSDs) or direct access storage devices (DASDs). These storage systems may be implemented according to one or more types of storage architectures that may include a storage area network-attached storage environment, a storage area network or a disk assembly directly attached to a client or host computer. In some examples, data can be lost when one or more data storage devices fail or a portion of data stored on the one or storage devices becomes corrupted (e.g., checksum or media read errors). In order to avoid the loss of data, one or more schemes may be implemented in a storage system. These one or more schemes may include mirroring, backup or parity protection.

[0003] The present disclosure describes some example methods for facilitating recovery from failures associated with a storage array. The storage array may have a plurality of storage devices. The plurality of storage devices may include a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device. These example methods may include determining first row parities for the plurality of data storage devices. The first row parities may span the plurality of data storage devices and may be stored on the first row parity storage device. First diagonal parities along first diagonal parity sets may then be determined. The first diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the first diagonal parity storage device. The example methods may also include determining second diagonal parities along second diagonal parity sets. The second diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the second diagonal parity storage device. Second row parities for the plurality of data storage devices may then be determined. The second row parities may span the plurality of data storage devices and may be stored on the second row parity storage device. In some examples, at least a portion of one or more rows of the second row parities may include determined second row parities based on a portion of the plurality of data storage devices.

[0004] The present disclosure also describes example devices to facilitate recovery from failures associated with a storage array. The storage array may have a plurality of storage devices. The plurality of storage devices may include a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device. These example devices may

include a recovery manager having logic. The logic may be configured to determine first row parities for the plurality of data storage devices. The first row parities may span the plurality of data storage devices and may be stored on the first row parity storage device. The logic may also be configured to determine a first diagonal parity along first diagonal parity sets. The first diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the first diagonal parity storage device. The logic may also be configured to determine second diagonal parities along second diagonal parity sets. The second diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the second diagonal parity storage device. The logic may also be configured to determine second row parities for the plurality of data storage devices. The second row parities may span the plurality of data storage devices and may be stored on the second row parity storage device. In some examples, at least a portion of one or more rows of the second row parities may include determined second row parities based on a portion of the plurality of data storage devices.

[0005] The present disclosure also describes example systems to facilitate recovery from failures associated with a storage array. These example systems may have a plurality of storage devices. The plurality of storage devices may include a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device. The example systems may also include a recovery manager having logic. The logic may be configured to determine first row parities for the plurality of data storage devices. The first row parities may span the plurality of data storage devices and may be stored on the first row parity storage device. The logic may also be configured to determine first diagonal parities along first diagonal parity sets. The first diagonal parity may span all except for one of the plurality of data

storage devices and the first row parity storage device and may be stored on the first diagonal parity storage device. The logic may also be configured to determine second diagonal parities along second diagonal parity sets. The second diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the second diagonal parity storage device. The logic may also be configured to determine second row parities for the plurality of data storage devices. The second row parities may span the plurality of data storage devices and may be stored on the second row parity storage device. In some examples, at least a portion of one or more rows of the second row parities may include determined second row parities based on a portion of the plurality of data storage devices.

[0006] The present disclosure also describes example computer program products. In some examples, the computer program products may include a non-transitory medium having instructions to facilitate recovery from failures associated with a storage array having a plurality of storage devices. The plurality of storage devices may include a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device. The instructions, which, when executed by logic, may cause the logic to determine first row parities for the plurality of data storage devices. The first row parities may span the plurality of data storage devices and may be stored on the first row parity storage device. The instructions may also cause the logic to determine first diagonal parities along first diagonal parity sets. The first diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the first diagonal parity storage device. The instructions may also cause the logic to determine second diagonal parities along second diagonal parity sets. The second diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the

second diagonal parity storage device. The instructions may also cause the logic to determine second row parities for the plurality of data storage devices. The second row parities may span the plurality of data storage devices and may be stored on the second row parity storage device. In some examples, at least a portion of one or more rows of the second row parities may include a determined second row parity based on a portion of the plurality of data storage devices.

[0007] The foregoing summary is illustrative only and is not intended to be in any way limiting. In addition to the illustrative aspects, embodiments, and features described above, further aspects, embodiments, and features will become apparent by reference to the drawings and the following detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] The foregoing and other features of this disclosure will become more fully apparent from the following description and appended claims, taken in conjunction with the accompanying drawings. Understanding that these drawings depict only several embodiments in accordance with the disclosure and are, therefore, not to be considered limiting of its scope, the disclosure will be described with additional specificity and detail through use of the accompanying drawings.

[0009] **FIG. 1** illustrates an example system to facilitate recovery from failures associated with a storage array;

[0010] **FIG. 2** illustrates a block diagram of an example architecture for a recovery manager;

[0011] **FIGS. 3A-D** illustrates example parity tables associated with triple parity encoding for a storage array;

[0012] **FIG. 4** illustrates a flow chart of example methods for facilitating recovery from failures associated with a storage array;

[0013] **FIG. 5** illustrates a block diagram of an example computer program product; and

[0014] **FIG. 6** illustrates an example computing device, all arranged in accordance with at least some embodiments of the present disclosure.

DETAILED DESCRIPTION

[0015] In the following detailed description, reference is made to the accompanying drawings, which form a part hereof. In the drawings, similar symbols typically identify similar components, unless context dictates otherwise. The illustrative examples or embodiments described in the detailed description, drawings, and claims are not meant to be limiting. Other examples or embodiments may be utilized, and other changes may be made, without departing from the spirit or scope of the subject matter presented here. It will be readily understood that aspects of this disclosure, as generally described herein, and illustrated in the Figures, can be arranged, substituted, combined, and designed in a wide variety of different configurations, all of which are explicitly contemplated and make part of this disclosure.

[0016] This disclosure is drawn, *inter alia*, to methods, apparatus, systems and computer program products related to facilitating recovery from failures associated with a storage array having a plurality of storage devices.

[0017] As contemplated in the present disclosure, one or more schemes may be implemented in a storage system to prevent loss of data should a storage device fail or a portion of data or information stored on the one or storage devices becomes corrupted. Also, as contemplated in the present disclosure, these one or more schemes may include mirroring, backup or parity protection. Mirroring may be an expensive solution in terms of resources needed to mirror data and backup may not protect data modified between backups.

[0018] In some examples, parity protection may be a preferred alternative to mirroring or backups. One type of parity protection typically referred to as single parity may provide redundant encoding of data that allows for recovery of lost data from a single storage device failure or corrupt data on a single storage device. Another type of parity protection typically referred to as double parity may provide redundant encoding of data that allows for recovery

of lost data in a storage system having two storage devices failures, corrupt data on two storage devices or a combination of a storage device failure and corrupt data on a storage device. However, some storage systems may include large numbers of storage devices and the larger the system, the greater the probability of three storage device failures or corrupt data on three storage devices. Thus, a need exists for parity protection that allows for recovery of lost data in a storage system having three storage device failures, corrupt data on three storage devices or a combination of storage device failures and corrupt data on storage devices.

[0019] In some examples, recovery from failures associated with a storage array may be facilitated by a triple parity encoding scheme. For these examples, the storage array may have a plurality of storage devices. The plurality of storage devices may include a plurality of data storage devices and plurality of parity storage devices. These parity storage devices may include a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device. For these examples, first row parities for the plurality of data storage devices may be determined. The first row parities may span the plurality of data storage devices and may be stored on the first row parity storage device. First diagonal parities along first diagonal parity sets may then be determined. The first diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the first diagonal parity storage device. Second diagonal parities along second diagonal parity sets may also be determined. The second diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the second diagonal parity storage device. Second row parities for the plurality of data storage devices may then be determined. The second row parities may span the plurality of data storage devices and may be stored on the second row parity storage device. For these examples, at

least a portion of one or more rows of the second row parities may include determined second row parities based on a portion of the plurality of data storage devices.

[0020] **FIG. 1** illustrates an example system to facilitate recovery from failures associated with a storage array. As shown in FIG. 1, system 100 includes a computing platform 110 coupled to a storage array 120 via communication channel 130. Also, as shown in FIG. 1, computing platform 110 includes a processing element(s) 112, a memory 114 and a storage array controller 116. In some examples, as depicted in FIG. 1, processing element(s) 112, system memory 114 and storage array controller 116 may communicatively couple via a communication link 118.

[0021] In some examples, as shown in FIG. 1, storage array controller 116 may include a recovery manager 117 and a storage manager 119. System memory 114 is also shown in FIG. 1 as including an operating system 115. For some examples, operating system 115 may be maintained in memory 114 and may be configured to be executed by processing element(s) 112. Also, operating system 115 may work in cooperation with storage manager 119 to read from and write to storage array 120.

[0022] Also as shown in FIG. 1, storage array 120 may include data storage devices 122-1 to 122-m, where m represents any positive integer greater than 1. Storage array 120 may also include four parity storage devices shown in FIG. 1 as row parity storage device 123, diagonal parity storage device 124, diagonal parity storage device 125 and row parity storage device 126. The plurality of storage devices of storage array 120, as shown in FIG. 1, may be interconnected and communicatively coupled to storage array controller 116 via communication channel 130.

[0023] According to some examples, storage array controller 116 may cooperate with storage array 120 to access information or data requested by a user (or client). The data may be stored on any type of attached storage device such as video tape, optical, digital versatile

disks (DVDs), magnetic tape, bubble memory, electronic random access memory, micro-electro mechanical or any other similar storage media configured to store data or parity information. Although not shown in FIG. 1, storage array controller 116 and storage array 120 may separately include interface circuitry to communicatively couple computing platform 110 to storage array 120 via communication channel 130. For these examples, communication channel 130 may operate according to various communication protocols to include, but not limited to, Universal Serial Bus (USB), PCI-Express, Serial Advanced Technology Attachment (SATA), Parallel Advanced Technology Attachment (PATA) or small computer system interface (SCSI).

[0024] In some examples, as described more below, recovery manager 117 may include logic and/or features to facilitate recovery from failures associated with storage array 120. The failures may include a complete failure of one or more storage devices from storage array 120 and/or corrupt data or information (e.g., checksum error) on one or more storage devices from storage array 120. For these examples, recovery manager 117 may implement a triple parity encoding scheme that uses the four parity storage devices of storage array 120.

[0025] According to some examples, when implementing the triple parity encoding scheme, storage array 120 may include n data storage devices and four parity storage devices, where n = p - 1 and p is a prime number greater than 3. For these examples, each of the plurality of storage devices of storage array 120 may be divided by recovery manager 117 into chunks. Each of the plurality of storage devices may then be further organized by recovery manager 117 into stripes that include a same number of chunks in each storage device. Also, each of the plurality of storage devices may be striped into $(p-1)^2$ chunks. In some examples, the plurality of storage devices may be striped according to a Redundant Array of Independent Disks (RAID) scheme (e.g., RAID 6), although this disclosure is not limited to only a RAID striping scheme.

[0026] **FIG. 2** illustrates a block diagram of an example architecture for a recovery manager. As described above for system 100 in FIG. 1, storage array controller 116 includes recovery manager 117. In some examples, recovery manager 117 includes features and/or logic configured or arranged to facilitate recovery from failures associated with a storage array such as storage array 120.

[0027] The example recovery manager 117 of FIG. 2 includes recover logic 210, control logic 220, memory 230 and input/output (I/O) interfaces 240. As illustrated in FIG. 2, recover logic 210 is coupled to control logic 220, memory 230 and I/O interfaces 240. Recover logic 210 may further include one or more of a stripe feature 212, a row feature 214, a diagonal feature 216, or a recovery feature 218, or any reasonable combination thereof.

[0028] In some examples, the elements portrayed in FIG. 2's block diagram are configured to support or enable recovery manager 117 as described in this disclosure. A given recovery manager 117 may include some, all or more elements than those depicted in FIG. 2. For example, recover logic 210 and control logic 220 may separately or collectively represent a wide variety of logic device(s) to implement the features of recovery manager 117. An example logic device may include one or more of a computer, a microprocessor, a microcontroller, a field programmable gate array (FPGA), an application specific integrated circuit (ASIC), a sequestered thread or a core of a multi-core/multi-threaded microprocessor or a combination thereof.

[0029] In some examples, as shown in FIG. 2, recover logic 210 includes stripe feature 212, row feature 214, diagonal feature 216, or recovery feature 218. Recover logic 210 may be configured to use one or more of these features to perform operations. As described in more detail below, example operations may include organizing storage devices into stripes according to a striping scheme. Example operations may also include determining row and diagonal parity sets to be stored according to the striping scheme and facilitating a recovery

from failures associated with a storage array (e.g., storage array 120) using the determined row and diagonal parity sets.

[0030] In some examples, control logic 220 may be configured to control the overall operation of recovery manager 117. As mentioned above, control logic 220 may represent any of a wide variety of logic device(s) configured to operate in conjunction with executable content or instructions to implement the control of recovery manager 117. In some alternate examples, the features and functionality of control logic 220 may be implemented within recover logic 210.

[0031] According to some examples, memory 230 is arranged to store executable content or instructions. The executable content or instructions may be used by control logic 220 and/or recover logic 210 to implement or activate features or elements of recovery manager 117. Memory 230 may also be arranged to at least temporarily maintain information associated with determining row or diagonal parities (e.g., construction algorithms) and storing the row or diagonal parities according to a striping scheme.

[0032] Memory 230 may include a wide variety of memory media including, but not limited to, one or more of volatile memory, non-volatile memory, flash memory, programmable variables or states, random access memory (RAM), read-only memory (ROM), or other static or dynamic storage media.

[0033] In some examples, I/O interfaces 240 may provide an interface between recovery manager 117 and elements or devices that may be communicatively coupled to storage array controller 116. For example, as mentioned above for FIG. 1, storage array controller 116 may be configured to communicatively couple to processing element(s) 112 or memory 114 via communication link 118 or may be configured to communicatively couple to storage array 120 via communication channel 130. The I/O interfaces 240, for example, may include an interface configured to operate according to various wireless and/or wired communication

protocols to allow recovery manager 117 to communicate over communication link 118 or communication channel 130 (e.g., Inter-Integrated Circuit (I^2C), System Management Bus (SMBus), Serial Peripheral Interface Bus (SPI)USB, SATA, PATA, SCSI, eSATA, PCI-Express, IEEE 802.1, IEEE 802.11, IEEE 802.16, GSM, GPRS, EDGE, W-CDMA, HSPA, LTE, CDMA-2000, EV-DO, etc.).

[0034] **FIGS. 3A-D** illustrate example parity tables associated with triple parity encoding for a storage array. In some examples, the storage array may be similar to storage array 120 as shown in FIG. 1. As shown in FIGS. 3A-D, table 310, table 320, table 330 and table 340 may be constructed based, at least in part, on the use of four parity storage devices. These parity tables provide an example of a triple parity encoding scheme. This triple parity encoding scheme may facilitate recovery from failures associated with up to three failures (complete device failure(s) or data corruption(s)) of storage array 120. For these examples, in addition to the four parity storage devices depicted in FIG. 1, storage array also includes four data storage devices. As shown in FIGS. 3A-D, parities associated with the four data storage devices are shown as column headers 122-1 to 122-4 and parities associated with the four parity storage devices are shown as column headers 123 – 126.

[0035] According to some examples, as described above, the storage devices of storage array 120 may be divided into $(p - 1)^2$ chunks, where p is a prime number greater than 3. So for the example storage array mentioned above for FIGS. 3A-D, $p = 5$. Thus, for these examples, the storage devices of storage array 120 may be divided into $(5 – 1)^2 = 16$ chucks. As a result of being divided into 16 chunks, each of the parity tables shown in FIGS. 3A-D includes 16 rows.

[0036] In some examples, the triple parity encoding scheme shown in FIGS. 3A-D may be based, at least in part, on implementing one or more example equations or algorithms as

shown below. The example equations may be based, at least in part, on the following assumptions for each of the data storage devices 122-1 to 122-4:

p is a prime number greater than 3,

data storage devices are represented as D,

D(k,j) and (k = 0,1,...p-2) and (j = 0,1,...,$(p-1)^2-1$),

where j is the j^{th} chunk on data storage device k.

The example equations may also be based, at least in part, on the following assumptions for each of the parity storage devices 123-126:

p is a prime number greater than 3,

parity storage devices are represented as DP,

DP(i,j) and (i = 1, 2, 3, 4) and (j = 0,1,...,$(p-1)^2-1$),

where j is the j^{th} chunk on parity storage device i.

[0037] In some examples, first row parities spanning data storage devices 122-1 to 122-4 may be determined according to example equation (1). For these examples, the first row parities are identified in the far right column of table 310 of FIG. 3A and may be stored in row parity storage device 123.

(1) $DP(1,j)=\bigoplus_{k=0}^{p-2} D(k,j)$

[0038] In some examples, first diagonal parities along first diagonal parity sets may be determined according to example equation (2). As described more below, the first diagonal parities may span all except for one of data storage devices 122-1 to 122-4 and row parity storage device 123. For these examples, the first diagonal parities are identified in the far right column of table 320 of FIG. 3B and may be stored in diagonal parity storage device 124.

(2) If (j%(p-1)==(p-2) $DP(2,j)=\bigoplus_{k=0}^{j\%(p-1)} D(k,j-k)\oplus\bigoplus_{k=j\%(p-1)+2}^{p-2} D(k,j+p-k)$

Else $DP(2,j) = \bigoplus_{k=0}^{j\%(p-1)} D(k, j-k) \oplus \bigoplus_{k=j\%(p-1)+2}^{p-2} D(k, j+p-k) \oplus DP(1, j+1)$

[0039] According to some examples, as shown in table 320, the first diagonal parities along the first diagonal parity sets may be defined based, at least in part, on the first diagonal parity sets wrapping around within $(p-1)^2$ rows so that all chunks belonging to a respective first diagonal parity set of a given stripe are stored in the given stripe. For example, the first diagonal parity set identified as "0" in the far right column of table 320 includes chunks from data storage device 122-1 at row 1, data storage device 122-3 at row 4, data storage device 122-4 at row 3 and row parity storage device 123 at row 2. Also, note that the first diagonal parity set identified as "0" spans all except for one of data storage devices 122-1 to 122-4 and row parity storage device 123. The one storage device excepted is data storage device 122-2. Further, note that other first diagonal parity sets also span all except for one of data storage devices 122-1 to 122-4 and row parity storage device 123.

[0040] In some examples, second diagonal parities along second diagonal parity sets may be determined according to example equation (3). As described more below, the first diagonal parities may span all except for one of data storage devices 122-1 to 122-4 and row parity storage device 123. For these examples, the second diagonal parities are identified in the far right column of table 330 of FIG. 3C and may be stored in diagonal parity storage device 125.

(3) If (j/(p-1)==(p-2)) $DP(3,j) = \bigoplus_{k=0}^{j/(p-1)} D(k, j-k(p-1)) \oplus \bigoplus_{k=j/(p-1)+2}^{p-2} D(k, j+(p-k)(p-1))$

Else

$DP(3,j) = \bigoplus_{k=0}^{j/(p-1)} D(k, j-k(p-1)) \oplus \bigoplus_{k=j/(p-1)+2}^{p-2} D(k, j+(p-k)(p-1) \oplus DP(1, (j+1)(p-1))$

[0041] According to some examples, as shown in table 330, the second diagonal parities along the second diagonal parity sets may be defined based, at least in part, on the second diagonal parity sets wrapping around within $(p-1)^2$ rows so that all chunks belonging to a

respective first diagonal parity set of a given stripe are stored in the given stripe. Also, similar to what was described above for table 320, the first diagonal parity sets span all except one of data storage devices 122-1 to 122-4 and row parity storage device 123.

[0042] In some examples, second row parities spanning data storage devices 122-1 to 122-4 may be determined according to example equation (4). For these examples, the second row parities are identified in the far right column of table 340 of FIG. 3D and may be stored in row parity storage device 126.

(4) If (j%(p-1) + 1 > j/(p-1) + 1) $DP(4,j) = \bigoplus_{k=0}^{j\%(p-1)} D(k, j-k(p-1)-k)) \oplus \bigoplus_{k=j\%(p-1)+2}^{j/(p-1)}$

$$D(k, j+p-k-k(p-1)) \oplus DP(1,(j+1)(p-1)) \oplus \bigoplus_{k=j/(p-1)+2}^{p-2} D(k, j+p-(p-k)(p-1))$$

Else $DP(4,j) = \bigoplus_{k=0}^{j/(p-1)} D(k, j-k(p-1)-k)) \oplus \bigoplus_{k=j/(p-1)+2}^{j\%(p-1)} D(k, j-k-(p-k)(p-1)) \oplus$

$$\bigoplus_{k=j\%(p-1)+2}^{p-2} D(k, j+p-(p-k)(p-1))$$

[0043] According to some examples, as shown in table 340, the second row parities may span data storage devices 122-1 to 122-4. Also, for these examples, at least some of the determined second row parities may be based on a portion of the data storage devices rather than all four of the data storage devices. For example, a row parity at row 2 of table 340 indicates a determined row parity based on three of the four data storage devices and a row parity at row 12 indicates a determined row parity based on two of the four data storage devices.

[0044] According to some examples, recovery manager 117 may include logic and/or features configured to recover from various different types of failures associated with storage array 120 (e.g., via recovery feature 218). For example, if three data storage devices fail, recovery manager 117 may implement a first row-diagonal parity (RDP) scheme that may utilize one or more row parity values associated with the second row parities stored in row

parity storage device 126 and one or more diagonal parity values associated with the second diagonal parities stored in diagonal parity storage device 125. Recovery manager 117 may also implement a second RDP scheme that utilizes one or more diagonal parity values associated with the first diagonal parities stored in diagonal parity storage device 124 and one or more row parity values associated with the first row parities stored in row parity storage device 123. For these examples, by implementing the first and second RDP schemes, the data maintained in the three failed storage devices may be recovered.

[0045] In some examples, a combination of data storage devices and parity storage devices may fail or have data corruption. For example, one of the data storage devices and both row parity storage device 123 and diagonal parity storage device 124 may fail or have a data corruption. Recovery manager 117 may implement a single RDP scheme that utilizes one or more diagonal parity values associated with the first diagonal parities stored in diagonal parity storage device 124 and one or more one or more row parity values associated with the second row parities stored in row parity storage device 126. For these examples, as a result of implementing the single RDP scheme, the data maintained in the failed data storage devices may be recovered as well as the parity values stored in row parity storage device 123 and diagonal parity storage device 124.

[0046] The example failures described above are just two examples of how recovery manager 117 may implement one or more RDP schemes to recover data or parity values stored in failed or data corrupted storage devices of storage array 120. Various other RDP schemes may be implemented by using row and diagonal parity values stored in row parity storage devices 123, 126 or diagonal parity storage devices 124, 125 to recover data or parity values.

[0047] **FIG. 4** illustrates a flow chart of example methods for facilitating recovery from failures associated with a storage array. In some examples, system 100 as shown in FIG. 1, is

used to illustrate example methods related to the flow chart depicted in FIG. 4. A recovery manager 117 as shown in FIG. 2 may also be used to illustrate the example methods. But the described methods are not limited to implementations as shown in or described for FIGS. 1 or 2. The example methods may be implemented on other systems or managers having one or more of the elements depicted in FIGS. 1 or 2.

[0048] Beginning at block 410 (Stripe Storage Devices), recovery manager 117 may include logic and/or features configured to stripe the storage devices of storage array 120 (e.g., via stripe feature 212). In some examples, the storage devices may be divided into chunks and then organized into stripes that include a same number of chunks in each storage device. As described previously, the storage devices of storage array 120 may be divided into $(p-1)^2$ and with a value of 5 for p this equates to $(5-1)^2$ or 16 chunks.

[0049] Continuing from block 410 to block 420 (Determine First Row Parities), recovery manager 117 may include logic and/or features configured to determine first row parities (e.g., via row feature 214). In some examples, row feature 214 may implement example equation (1) to determine the first row parities. For these examples, the first row parities may be stored in row parity storage device 123.

[0050] Continuing from block 420 to block 430 (Determine First Diagonal Parities), recovery manager 117 may include logic and/or features configured to determine first diagonal parities (e.g., via diagonal feature 216). In some examples, diagonal feature 214 may implement example equation (2) to determine the first diagonal parities. For these examples, the first diagonal parities may be stored in diagonal parity storage device 124.

[0051] Continuing from block 430 to block 440 (Determine Second Diagonal Parities), recovery manager 117 may include logic and/or features configured to determine second diagonal parities (e.g., via diagonal feature 216). In some examples, diagonal feature 216

may implement example equation (3) to determine the second diagonal parities. For these examples, the second diagonal parities may be stored in diagonal parity storage device 125.

[0052] Continuing from block 440 to block 450 (Determine Second Row Parities), recovery manager 117 may include logic and/or features configured to determine second row parities (e.g., via row feature 214). In some examples, row feature 214 may implement example equation (4) to determine the second row parities. For these examples, the second row parities may be stored in row parity storage device 126.

[0053] Continuing from block 450 to decision block 460 (Failure or Data Corruption), recovery manager 117 may include logic and/or features configured to determine whether failures or data corruption has occurred on one or more of the storage devices of storage array 120 (e.g., via recovery feature 218). If failure or data corruption has occurred, the process moves to block 470. Otherwise, the process moves to block 480.

[0054] Moving from decision block 460 to block 470 (Implement Data/Parity Recovery), recovery manager 117 may include logic and/or features configured to implement one or more recovery schemes (e.g., via recovery feature 218). In some examples, the one or more recovery schemes may include RDP schemes to recover data or parity information lost from device failures and/or data corruption associated with the storage devices of storage array 120.

[0055] Continuing from block 470 to decision block 480 (New Info. to be Striped?), recovery manager 117 may include logic and/or features configured to determine whether new information or data needs to be striped on the storage devices of storage array 120 (e.g., via stripe feature 212). If new information or data is to be striped, the process moves to block 420 and row/diagonal parities may be determined. Otherwise the process moves to decision block 460.

[0056] **FIG. 5** illustrates a block diagram of an example computer program product 500. In some examples, as shown in FIG. 5, computer program product 500 includes a signal bearing

medium 502 that may also include instructions 504 to facilitate recovery from failures associated with a storage array (e.g., storage array 120). The storage array may have a plurality of storage devices that includes a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device. Instructions 504, which, when executed by logic (e.g., recover logic 210), may cause the logic to determine first row parities for the plurality of data storage devices. The first row parities may span the plurality of data storage devices and may be stored on the first row parity storage device. Instructions 504 may also cause the logic to determine first diagonal parities along first diagonal parity sets. The first diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the first diagonal parity storage device. Instructions 504 may also cause the logic to determine second diagonal parities along second diagonal parity sets. The second diagonal parities may span all except for one of the plurality of data storage devices and the first row parity storage device and may be stored on the second diagonal parity storage device. Instructions 504 may also cause the logic to determine second row parities for the plurality of data storage devices. The second row parities may span the plurality of data storage devices and may be stored on the second row parity storage device. In some examples, at least a portion of one or more rows of the second row parities may include a determined second row parity based on a portion of the plurality of data storage devices.

[0057] Also depicted in FIG. 5, in some examples, computer product 500 may include one or more of a computer readable medium 506, a recordable medium 508 and a communications medium 510. The dotted boxes around these elements depict different types of mediums included within, but not limited to, signal bearing medium 502. These types of mediums may distribute instructions 504 to be executed by logic (e.g., recover logic 210). Computer

readable medium 506 and recordable medium 508 may include, but are not limited to, a flexible disk, a hard disk drive (HDD), a Compact Disc (CD), a Digital Versatile Disk (DVD), a digital tape, a computer memory, etc. Communications medium 510 may include, but is not limited to, a digital and/or an analog communication medium (e.g., a fiber optic cable, a waveguide, a wired communication link, a wireless communication link, etc.).

[0058] **FIG. 6** illustrates an example computing device 600. In some examples, recovery manager 117 depicted in FIGS. 1-2 may be implemented on computing device 600. In these examples, elements of computing device 600 may be arranged or configured for allocating a block of persistent storage or for a file system or application to access a block of persistent storage. In a very basic configuration 601, computing device 600 typically includes one or more processors 610 and system memory 620. A memory bus 630 can be used for communicating between the processor 610 and the system memory 620.

[0059] Depending on the desired configuration, processor 610 can be of any type including but not limited to a microprocessor (μP), a microcontroller (μC), a digital signal processor (DSP), or any combination thereof. Processor 610 can include one or more levels of caching, such as a level one cache 611 and a level two cache 612, a processor core 613, and registers 614. The processor core 613 can include an arithmetic logic unit (ALU), a floating point unit (FPU), a digital signal processing core (DSP Core), or any combination thereof. A memory controller 615 can also be used with the processor 610, or in some implementations the memory controller 615 can be an internal part of the processor 610.

[0060] Depending on the desired configuration, the system memory 620 can be of any type including but not limited to volatile memory (such as RAM), non-volatile memory (such as ROM, flash memory, etc.) or any combination thereof. System memory 620 typically includes an operating system 621, one or more applications 622, and program data 624. Application 622 includes instructions 623 that are arranged to perform the functions as

described herein including the actions described with respect to the recovery manager 117 architecture shown in FIG. 2. Program Data 624 includes recovery data 625 that is useful for implementing instructions 623 (e.g., determining row/diagonal parities). In some examples, application 622 can be arranged to operate with program data 624 on an operating system 621 such that implementations for facilitating recovery from failures associated with a storage array may be provided as described herein. This described basic configuration is illustrated in FIG. 6 by those components within dashed line 601.

[0061] Computing device 600 can have additional features or functionality, and additional interfaces to facilitate communications between the basic configuration 601 and any required devices and interfaces. For example, a bus/interface controller 640 can be used to facilitate communications between the basic configuration 601 and one or more data storage devices 650 via a storage interface bus 641. The data storage devices 650 can be removable storage devices 651, non-removable storage devices 652, or a combination thereof. Examples of removable storage and non-removable storage devices include magnetic disk devices such as flexible disk drives and hard-disk drives (HDD), optical disk drives such as compact disk (CD) drives or digital versatile disk (DVD) drives, solid state drives (SSD), and tape drives to name a few. Example computer storage media can include volatile and nonvolatile, removable and non-removable media implemented in any method or technology for storage of information, such as computer readable instructions, data structures, program modules, or other data.

[0062] System memory 620, removable storage 651 and non-removable storage 652 are all examples of computer storage media. Computer storage media includes, but is not limited to, RAM, ROM, EEPROM, flash memory or other memory technology, CD-ROM, digital versatile disks (DVD) or other optical storage, magnetic cassettes, magnetic tape, magnetic disk storage or other magnetic storage devices, or any other medium which can be used to

store the desired information and which can be accessed by computing device 600. Any such computer storage media can be part of device 600.

[0063] Computing device 600 can also include an interface bus 642 for facilitating communication from various interface devices (e.g., output interfaces, peripheral interfaces, and communication interfaces) to the basic configuration 601 via the bus/interface controller 640. Example output interfaces 660 include a graphics processing unit 661 and an audio processing unit 662, which can be configured to communicate to various external devices such as a display or speakers via one or more A/V ports 663. Example peripheral interfaces 670 include a serial interface controller 671 or a parallel interface controller 672, which can be configured to communicate with external devices such as input devices (e.g., keyboard, mouse, pen, voice input device, touch input device, etc.) or other peripheral devices (e.g., printer, scanner, etc.) via one or more I/O ports 673. An example communication interface 680 includes a network controller 681, which can be arranged to facilitate communications with one or more other computing devices 690 over a network communication via one or more communication ports 682. A network communication connection is one example of a communication media. Communication media may typically be embodied by computer readable instructions, data structures, program modules, or other data in a modulated data signal, such as a carrier wave or other transport mechanism, and includes any information delivery media. A "modulated data signal" can be a signal that has one or more of its characteristics set or changed in such a manner as to encode information in the signal. By way of example, and not limitation, communication media can include wired media such as a wired network or direct-wired connection, and wireless media such as acoustic, radio frequency (RF), infrared (IR) and other wireless media. The term computer readable media as used herein can include both storage media and communication media.

[0064] Computing device 600 can be implemented as a portion of a small-form factor

portable (or mobile) electronic device such as a cell phone, smart phone, a personal data assistant (PDA), a personal media player device, a wireless web-watch device, a personal headset device, an application specific device, or a hybrid device that include any of the above functions. Computing device 600 can also be implemented as a personal computer including both laptop computer and non-laptop computer configurations or implemented in a workstation or a server configuration.

[0065] References made in this disclosure to the term "responsive to" or "in response to" are not limited to responsiveness to a particular feature and/or structure. A feature may also be responsive to another feature and/or structure and also be located within that feature and/or structure. Moreover, when terms or phrases such as "coupled" or "responsive" or "in response to" or "in communication with", etc. are used herein or in the claims that follow, these terms should be interpreted broadly. For example, the phrase "coupled to" may refer to being communicatively, electrically and/or operatively coupled as appropriate for the context in which the phrase is used.

[0066] Those skilled in the art will recognize that it is common within the art to describe devices and/or processes in the fashion set forth herein, and thereafter use engineering practices to integrate such described devices (e.g., transmitters, receivers, wireless devices, computing platforms, computing devices, etc.) and/or methods into data processing systems. That is, at least a portion of the devices and/or methods described herein can be integrated into a data processing system via a reasonable amount of experimentation. Those having skill in the art will recognize that a typical data processing system generally includes one or more of a system unit housing, a video display device, a memory such as volatile and non-volatile memory, processors such as microprocessors and digital signal processors, computational entities such as operating systems, drivers, graphical user interfaces, and applications programs, one or more interaction devices, such as a touch pad or screen, and/or

control systems including feedback loops and control motors (e.g., feedback for sensing position and/or velocity; control motors for moving and/or adjusting components and/or quantities). A typical data processing system may be implemented utilizing any suitable commercially available component, such as those typically found in data computing/communication and/or network computing/communication systems.

[0067] The herein described subject matter sometimes illustrates different components or elements contained within, or connected with, different other components or elements. It is to be understood that such depicted architectures are merely examples, and that in fact many other architectures can be implemented which achieve the same functionality. In a conceptual sense, any arrangement of components to achieve the same functionality is effectively "associated" such that the desired functionality is achieved. Hence, any two components herein combined to achieve a particular functionality can be seen as "associated with" each other such that the desired functionality is achieved, irrespective of architectures or intermedial components. Likewise, any two components so associated can also be viewed as being "operably connected", or "operably coupled", to each other to achieve the desired functionality, and any two components capable of being so associated can also be viewed as being "operably couplable", to each other to achieve the desired functionality. Specific examples of operably couplable include but are not limited to physically mateable and/or physically interacting components and/or wirelessly interactable and/or wirelessly interacting components and/or logically interacting and/or logically interactable components.

[0068] With respect to the use of substantially any plural and/or singular terms herein, those having skill in the art can translate from the plural to the singular and/or from the singular to the plural as is appropriate to the context and/or application. The various singular/plural permutations may be expressly set forth herein for sake of clarity.

[0069] It will be understood by those within the art that, in general, terms used herein, and

especially in the appended claims (e.g., bodies of the appended claims) are generally intended as "open" terms (e.g., the term "including" should be interpreted as "including but not limited to," the term "having" should be interpreted as "having at least," the term "includes" should be interpreted as "includes but is not limited to," etc.). It will be further understood by those within the art that if a specific number of an introduced claim recitation is intended, such an intent will be explicitly recited in the claim, and in the absence of such recitation no such intent is present. For example, as an aid to understanding, the following appended claims may contain usage of the introductory phrases "at least one" and "one or more" to introduce claim recitations. However, the use of such phrases should not be construed to imply that the introduction of a claim recitation by the indefinite articles "a" or "an" limits any particular claim containing such introduced claim recitation to inventions containing only one such recitation, even when the same claim includes the introductory phrases "one or more" or "at least one" and indefinite articles such as "a" or "an" (e.g., "a" and/or "an" should typically be interpreted to mean "at least one" or "one or more"); the same holds true for the use of definite articles used to introduce claim recitations. In addition, even if a specific number of an introduced claim recitation *is* explicitly recited, those skilled in the art will recognize that such recitation should typically be interpreted to mean *at least* the recited number (e.g., the bare recitation of "two recitations," without other modifiers, typically means *at least* two recitations, or *two or more* recitations). Furthermore, in those instances where a convention analogous to "at least one of A, B, and C, etc." is used, in general such a construction is intended in the sense one having skill in the art would understand the convention (e.g., "a system having at least one of A, B, and C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). In those instances where a convention analogous to "at least one of A, B, or C, etc." is used, in general such a construction is intended in the sense one having

skill in the art would understand the convention (e.g., "a system having at least one of A, B, or C" would include but not be limited to systems that have A alone, B alone, C alone, A and B together, A and C together, B and C together, and/or A, B, and C together, etc.). It will be further understood by those within the art that virtually any disjunctive word and/or phrase presenting two or more alternative terms, whether in the description, claims, or drawings, should be understood to contemplate the possibilities of including one of the terms, either of the terms, or both terms. For example, the phrase "A or B" will be understood to include the possibilities of "A" or "B" or "A and B."

[0070] While various aspects and embodiments have been disclosed herein, other aspects and embodiments will be apparent to those skilled in the art. The various aspects and embodiments disclosed herein are for purposes of illustration and are not intended to be limiting, with the true scope and spirit being indicated by the following claims.

CLAIMS:

What is claimed is:

1. A method for facilitating recovery from failures associated with a storage array having a plurality of storage devices that includes a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device, the method comprising:

determining first row parities for the plurality of data storage devices, the first row parities spanning the plurality of data storage devices and being stored on the first row parity storage device;

determining first diagonal parities along first diagonal parity sets, the first diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined first diagonal parities being stored on the first diagonal parity storage device;

determining second diagonal parities along second diagonal parity sets, the second diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined second diagonal parities being stored on the second diagonal parity storage device; and

determining second row parities for the plurality of data storage devices, the second row parities spanning the plurality of data storage devices and being stored on the second row parity storage device, wherein at least a portion of one or more rows of the second row parities includes determined second row parities based on a portion of the plurality of data storage devices.

2. The method of claim 1, wherein the plurality of storage devices comprises n data storage devices, where $n = p - 1$ and p is a prime number greater than 3.

3. The method of claim 2, further comprising:

dividing each of the plurality of storage devices into chunks; and

organizing each of the plurality of storage devices into stripes that include a same number of chunks in each storage device, wherein each of the plurality of storage devices are striped into $(p-1)^2$ chunks.

4. The method of claim 3, wherein determining the first row parities and the second row parities includes the first row parities and the second row parities being separately defined based, at least in part, on determined first row parities and determined second row parities that separately include $(p-1)^2$ rows.

5. The method of claim 4, wherein determining the first diagonal parities along the first diagonal parity sets includes the first diagonal parities being defined based, at least in part, on the first diagonal parity sets wrapping around within $(p-1)^2$ rows so that all chunks belonging to a respective first diagonal parity set of a given stripe are stored in the given stripe.

6. The method of claim 5, wherein determining the second diagonal parities along the second diagonal parity sets includes the second diagonal parities being defined based, at least in part, on the second diagonal parity sets partially wrapping around within the $(p-1)^2$ rows so that all chunks belonging to a respective second diagonal parity set of a given stripe are stored in the given stripe.

7. The method of claim 6, wherein facilitating recovery from failures associated with the storage array includes enabling an implementation of a first row-diagonal parity (RDP)

scheme that utilizes one or more row parity values associated with the second row parities and one or more diagonal parity values associated with the second diagonal parities and then enabling an implementation of a second RDP scheme that utilizes one or more diagonal parity values associated with the first diagonal parities and one or more row parity values associated with the first row parities.

8. The method of claim 3, wherein each of the plurality of storage devices are striped according to a Redundant Array of Independent Disks (RAID) scheme.

9. The method of claim 1, wherein the plurality of storage devices includes at least one of a hard disk drive, a solid state disk drive, a direct access storage device, magnetic tape, bubble memory, optical storage media, electronic random access or a digital versatile disk.

10. The method of claim 1, wherein failures associated with the storage array comprises three concurrent failures associated with the storage array that include a combination of failures in one or more of the plurality of storage devices or checksum errors associated with one or more of the plurality of storage devices.

11. An apparatus to facilitate recovery from failures associated with a storage array having a plurality of storage devices that includes a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device, the apparatus comprising:

a recovery manager having logic, the logic configured to:

determine first row parities for the plurality of data storage devices, the first row parities spanning the plurality of data storage devices and being stored on the first row parity storage device;

determine first diagonal parities along first diagonal parity sets, the first diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined first diagonal parities being stored on the first diagonal parity storage device;

determine second diagonal parities along second diagonal parity sets, the second diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined second diagonal parities being stored on the second diagonal parity storage device; and

determine second row parities for the plurality of data storage devices, the second row parities spanning the plurality of data storage devices and being stored on the second row parity storage device, wherein at least a portion of one or more rows of the second row parities includes determined second row parities based on a portion of the plurality of data storage devices.

12. The apparatus of claim 11, wherein the plurality of storage devices comprises n data storage devices, where $n = p - 1$ and p is a prime number greater than 3.

13. The apparatus of claim 12, further comprising the logic configured to:

divide each of the plurality of storage devices into chunks; and

organize each of the plurality of storage devices into stripes that include a same number of chunks in each storage device, wherein each of the plurality of storage devices are striped into $(p-1)^2$ chunks.

14. The apparatus of claim 13, wherein to determine the first row parities and the second row parities includes the first row parities and the second row parities being separately defined based, at least in part, on determined first row parities and determined second row parities that separately include $(p-1)^2$ rows.

15. The apparatus of claim 14, wherein to determine the first diagonal parities along the first diagonal parity sets includes the first diagonal parities being defined based, at least in part, on the first diagonal parity sets wrapping around within $(p-1)^2$ rows so that all chunks belonging to a respective first diagonal parity set of a given stripe are stored in the given stripe.

16. The apparatus of claim 15, wherein to determine the second diagonal parities along the second diagonal parity sets includes the second diagonal parities being defined based, at least in part, on the second diagonal parity sets partially wrapping around within the $(p-1)^2$ rows so that all chunks belonging to a respective second diagonal parity set of a given stripe are stored in the given stripe.

17. The apparatus of claim 16, wherein to facilitate recovery from failures associated with the storage array includes enabling an implementation of a first row-diagonal parity (RDP)

scheme that utilizes one or more row parity values associated with the second row parities and one or more diagonal parity values associated with the second diagonal parities and then enabling an implementation of a second RDP scheme that utilizes one or more diagonal parity values associated with the first diagonal parities and one or more row parity values associated with the first row parities.

18. The apparatus of claim 13, wherein each of the plurality of storage devices are striped according to a Redundant Array of Independent Disks (RAID) scheme.

19. The apparatus of claim 11, wherein failures associated with the storage array comprises three concurrent failures associated with the storage array that include a combination of failures in one or more of the plurality of storage devices or checksum errors associated with one or more of the plurality of storage devices.

20. A system to facilitate recovery from failures associated with a storage array, the system comprising:

a plurality of storage devices, including a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device; and

a recovery manager having logic, the logic configured to:

determine first row parities for the plurality of data storage devices, the first row parities spanning the plurality of data storage devices and being stored on the first row parity storage device;

determine first diagonal parities along first diagonal parity sets, the first diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined first diagonal parities being stored on the first diagonal parity storage device;

determine second diagonal parities along second diagonal parity sets, the second diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined second diagonal parities being stored on the second diagonal parity storage device; and

determine second row parities for the plurality of data storage devices, the second row parities spanning the plurality of data storage devices and being stored on the second row parity storage device, wherein at least a portion of one or more rows of the second row parities includes determined second row parities based on a portion of the plurality of data storage devices.

21. The system of claim 20, wherein the plurality of storage devices comprises n data storage devices, where $n = p - 1$ and p is a prime number greater than 3.

22. The system of claim 21, further comprising the logic configured to:

divide each of the plurality of storage devices into chunks; and

organize each of the plurality of storage devices into stripes that include a same number of chunks in each storage device, wherein each of the plurality of storage devices are striped into $(p-1)^2$ chunks.

23. The system of claim 22, wherein to determine the first row parities and the second row parities includes the first row parities and the second row parities being separately defined based, at least in part, on determined first row parities and determined second row parities that separately include $(p-1)^2$ rows.

24. The system of claim 23, wherein to determine the first diagonal parities along the first diagonal parity sets includes the first diagonal parities being defined based, at least in part, on the first diagonal parity sets wrapping around within $(p-1)^2$ rows so that all chunks belonging to a respective first diagonal parity set of a given stripe are stored in the given stripe.

25. The system of claim 24, wherein to determine the second diagonal parities along the second diagonal parity sets includes the second diagonal parities being defined based, at least in part, on the second diagonal parity sets partially wrapping around within the $(p-1)^2$ rows so that all chunks belonging to a respective second diagonal parity set of a given stripe are stored in the given stripe.

26. The system of claim 25, wherein to facilitate recovery from failures associated with the storage array includes enabling an implementation of a first row-diagonal parity (RDP)

scheme that utilizes one or more row parity values associated with the second row parities and one or more diagonal parity values associated with the second diagonal parities and then enabling an implementation of a second RDP scheme that utilizes one or more diagonal parity values associated with the first diagonal parities and one or more row parity values associated with the first row parities.

27. The system of claim 20, wherein each of the plurality of storage devices are striped according to a Redundant Array of Independent Disks (RAID) scheme.

28. The system of claim 20, wherein the plurality of storage devices includes at least one of a hard disk drive, a solid state disk drive, a direct access storage device, magnetic tape, bubble memory, optical storage media, electronic random access or a digital versatile disks.

29. A computer program product comprising a non-transitory medium having instructions to facilitate recovery from failures associated with a storage array having a plurality of storage devices that includes a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device, which, when executed by logic, cause the logic to:

determine first row parities for the plurality of data storage devices, the first row parities spanning the plurality of data storage devices and being stored on the first row parity storage device;

determine first diagonal parities along first diagonal parity sets, the first diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined first diagonal parities being stored on the first diagonal parity storage device;

determine second diagonal parities along second diagonal parity sets, the second diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined second diagonal parities being stored on the second diagonal parity storage device; and

determine second row parities for the plurality of data storage devices, the second row parities spanning the plurality of data storage devices and being stored on the second row parity storage device, wherein at least a portion of one or more rows of the second row parities includes determined second row parities based on a portion of the plurality of data storage devices.

30. The computer program product of claim 29, wherein the plurality of storage devices comprises n data storage devices, where n = p - 1 and p is a prime number greater than 3.

31. The computer program product of claim 30, further comprising the logic configured to:

divide each of the plurality of storage devices into chunks; and

organize each of the plurality of storage devices into stripes that include a same number of chunks in each storage device, wherein each of the plurality of storage devices are striped into $(p-1)^2$ chunks.

32. The computer program product of claim 31, wherein the instructions to facilitate recovery from failures associated with the storage array comprises instructions to further cause the logic to enable an implementation of a first row-diagonal parity (RDP) scheme that utilizes one or more row parity values associated with the second row parities and one or more diagonal parity values associated with the second diagonal parities and then enabling an

implementation of a second RDP scheme that utilizes one or more diagonal parity values associated with the first diagonal parities and one or more row parity values associated with the first row parities.

33. The computer program product of claim 31, wherein each of the plurality of storage devices are striped according to a Redundant Array of Independent Disks (RAID) scheme.

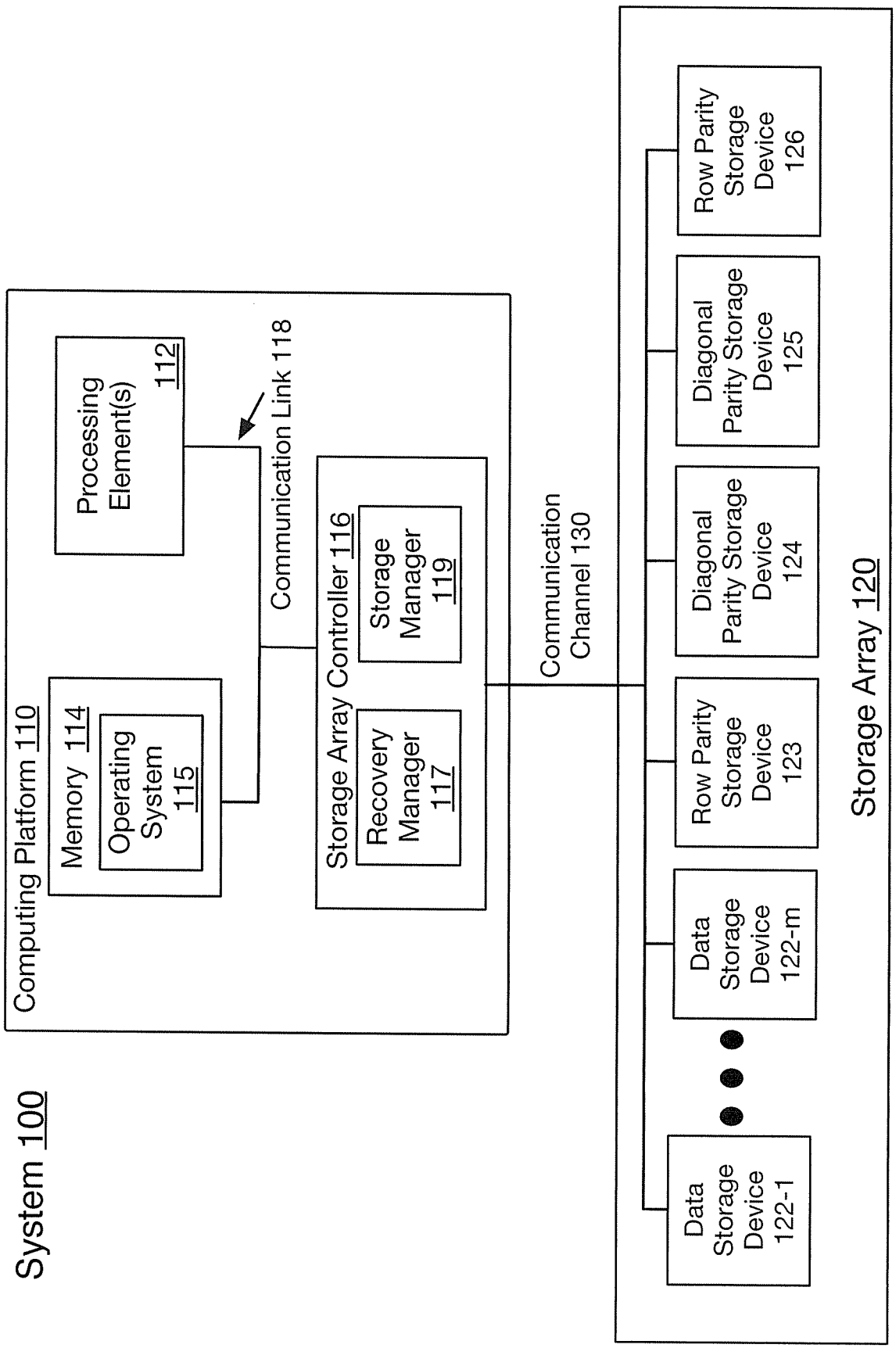

System 100
Computing Platform 110
Memory 114
Operating System 115
Processing Element(s) 112
Communication Link 118
Storage Array Controller 116
Recovery Manager 117
Storage Manager 119
Communication Channel 130
Storage Array 120
Data Storage Device 122-1
Data Storage Device 122-m
Row Parity Storage Device 123
Diagonal Parity Storage Device 124
Diagonal Parity Storage Device 125
Row Parity Storage Device 126

FIG. 1

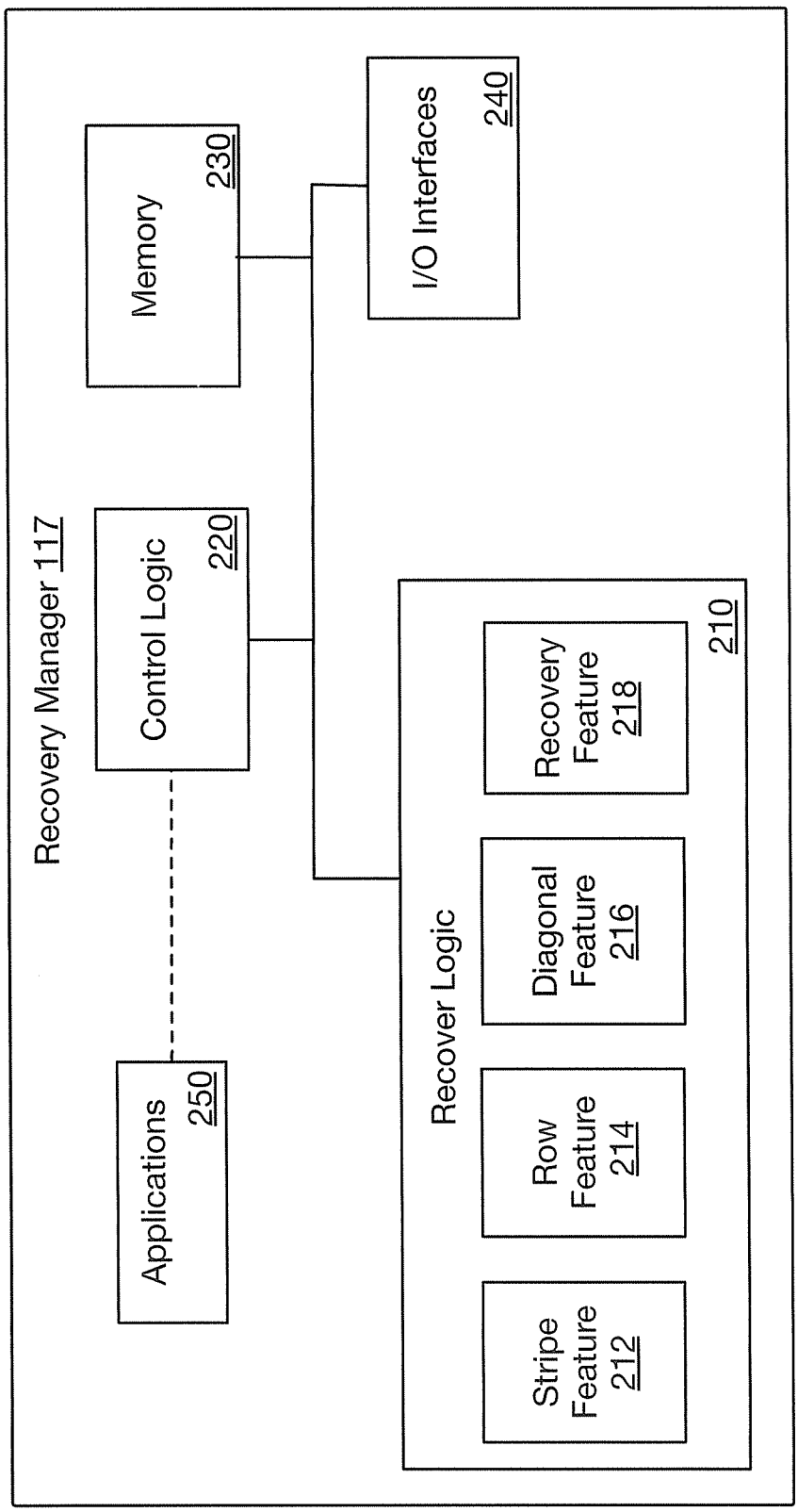

Recovery Manager 117
Memory 230
I/O Interfaces 240
Control Logic 220
Applications 250
Recover Logic 210
Recovery Feature 218
Diagonal Feature 216
Row Feature 214
Stripe Feature 212

FIG. 2

Table 310

Row #	122-1	122-2	122-3	122-4	123
1	0	0	0	0	0
2	1	1	1	1	1
3	2	2	2	2	2
4	3	3	3	3	3
5	4	4	4	4	4
6	5	5	5	5	5
7	6	6	6	6	6
8	7	7	7	7	7
9	8	8	8	8	8
10	9	9	9	9	9
11	10	10	10	10	10
12	11	11	11	11	11
13	12	12	12	12	12
14	13	13	13	13	13
15	14	14	14	14	14
16	15	15	15	15	15

FIG. 3A

Table 320

Row #	122-1	122-2	122-3	122-4	123	124
1	0	1	2	3		0
2	1	2	3		0	1
3	2	3		0	1	2
4	3		0	1	2	3
5	4	5	6	7		4
6	5	6	7		4	5
7	6	7		4	5	6
8	7		4	5	6	7
9	8	9	10	11		8
10	9	10	11		8	9
11	10	11		8	9	10
12	11		8	9	10	11
13	12	13	14	15		12
14	13	14	15		12	13
15	14	15		12	13	14
16	15		12	13	14	15

FIG. 3B

Table 330

Row #	122-1	122-2	122-3	122-4	123	125
1	0	4	8	12		0
2	1	5	9	13		1
3	2	6	10	14		2
4	3	7	11	15		3
5	4	8	12		0	4
6	5	9	13		1	5
7	6	10	14		2	6
8	7	11	15		3	7
9	8	12		0	4	8
10	9	13		1	5	9
11	10	14		2	6	10
12	11	15		3	7	11
13	12		0	4	8	12
14	13		1	5	9	13
15	14		2	6	10	14
16	15		3	7	11	15

FIG. 3C

Table 340

Row #	122-1	122-2	122-3	122-4	126
1	0	5	10	15	0
2	1	6	11		1
3	2	7		12	2
4	3		8	13	3
5	4	9	14		4
6	5	10	15		5
7	6	11			6
8	7		12		7
9	8	13		3	8
10	9	14			9
11	10	15		0	10
12	11			1	11
13	12		2	7	12
14	13		3		13
15	14			4	14
16	15		0	5	15

FIG. 3D

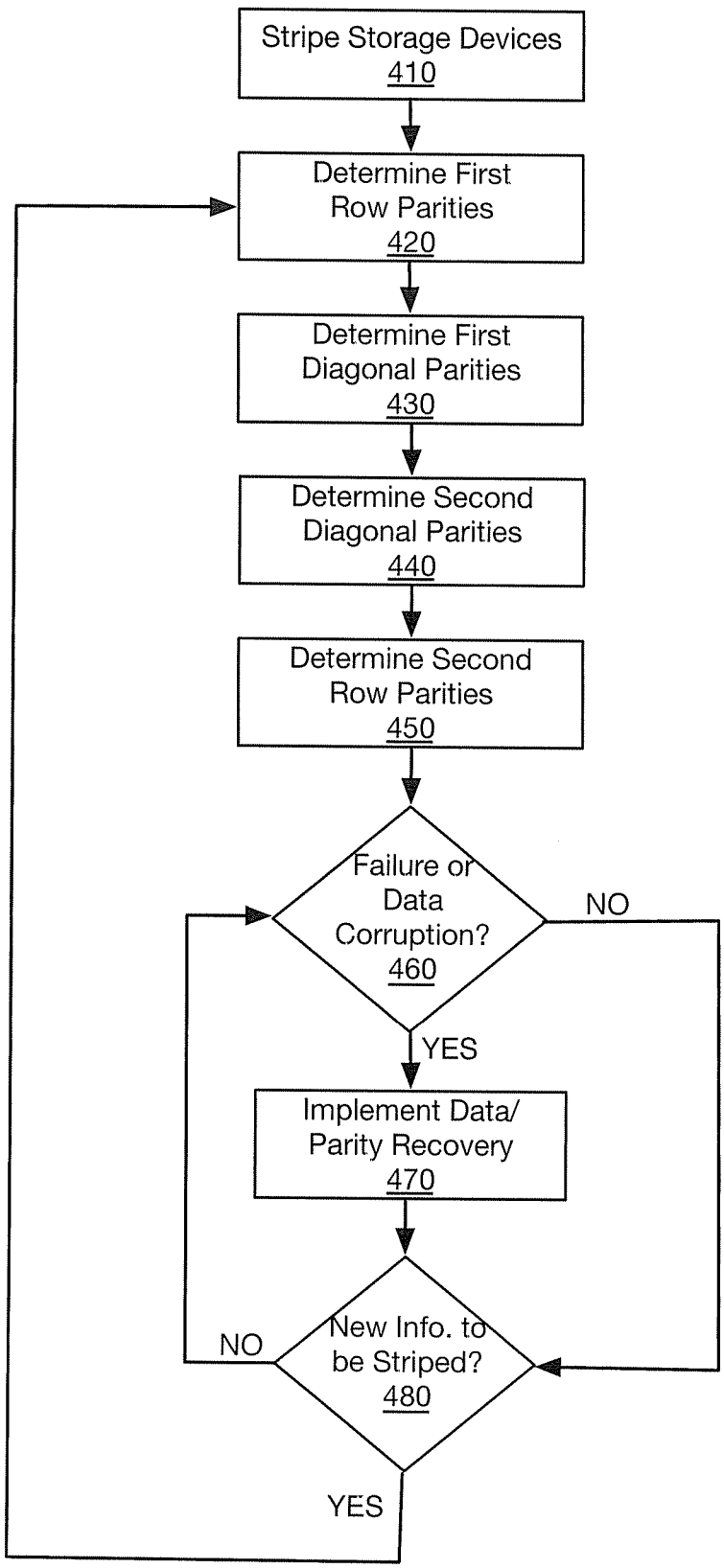

Stripe Storage Devices
410
Determine First
Row Parities
420
Determine First
Diagonal Parities
430
Determine Second
Diagonal Parities
440
Determine Second
Row Parities
450
Failure or
Data
Corruption?
460
NO
YES
Implement Data/
Parity Recovery
470
New Info. to
be Striped?
480
NO
YES

FIG.4

500 A computer program product.

502 A signal bearing medium.

504 instructions to facilitate recovery from failures associated with a storage array having a plurality of storage devices that includes a plurality of data storage devices, a first row parity storage device, a second row parity storage device, a first diagonal parity storage device and a second diagonal parity storage device, which, when executed by logic, cause the logic to:

determine first row parities for the plurality of data storage devices, the first row parities spanning the plurality of data storage devices and being stored on the first row parity storage device;

determine first diagonal parities along first diagonal parity sets, the first diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined first diagonal parities being stored on the first diagonal parity storage device;

determine second diagonal parities along second diagonal parity sets, the second diagonal parities spanning all except for one of the plurality of data storage devices and the first row parity storage device, the determined second diagonal parities being stored on the second diagonal parity storage device; or

determine second row parities for the plurality of data storage devices, the second row parities spanning the plurality of data storage devices and being stored on the second row parity storage device, wherein at least a portion of one or more rows of the second row parities includes determined second row parities based on a portion of the plurality of data storage devices.

506 a computer-readable medium.

508 a recordable medium.

510 a communications medium.

FIG. 5

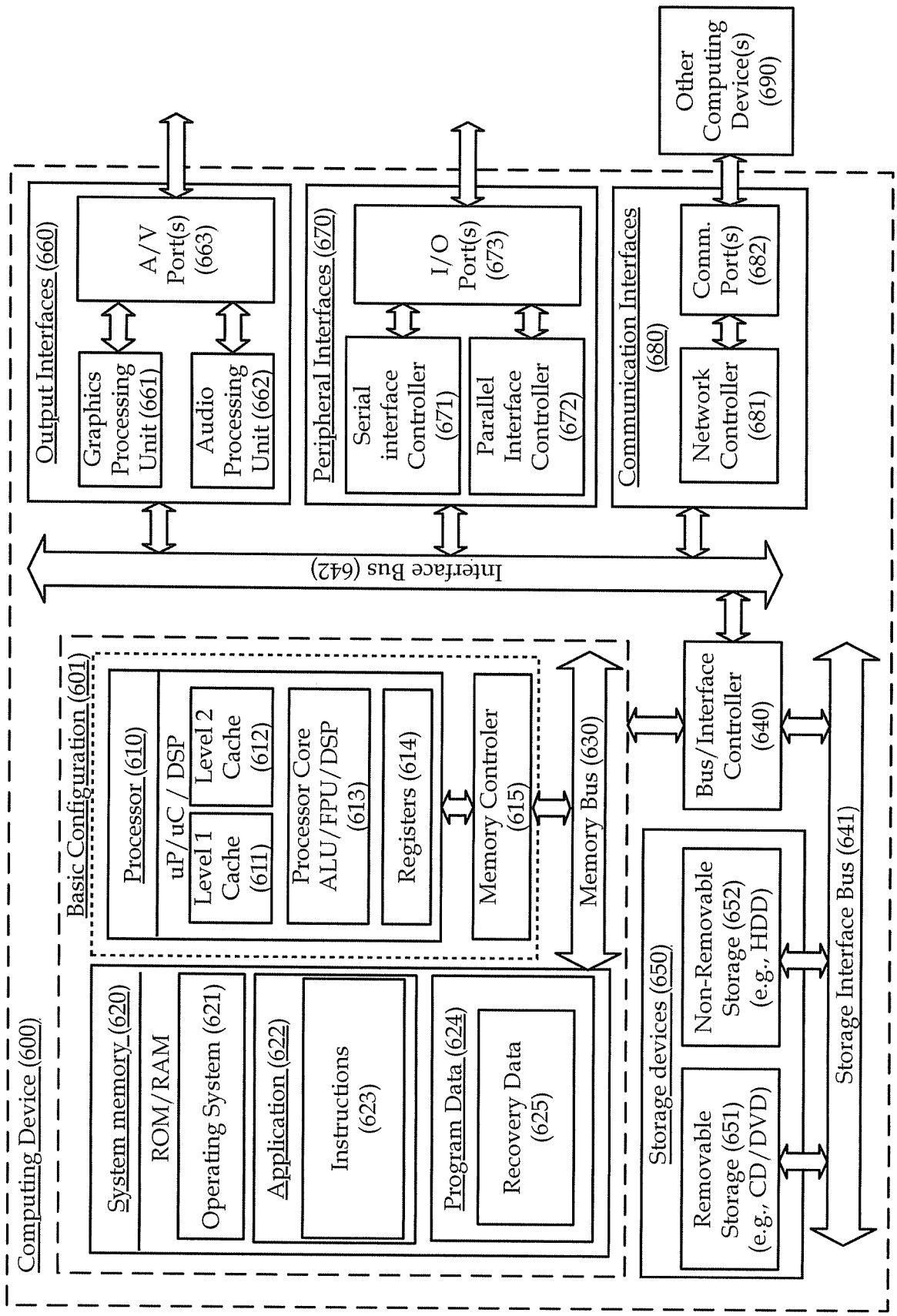

Computing Device (600)
Basic Configuration (601)
Processor (610)
uP/uC / DSP
Level 1 Cache (611)
Level 2 Cache (612)
Processor Core ALU/FPU/DSP (613)
Registers (614)
Memory Controler (615)
System memory (620)
ROM/RAM
Operating System (621)
Application (622)
Instructions (623)
Program Data (624)
Recovery Data (625)
Memory Bus (630)
Bus/Interface Controller (640)
Storage devices (650)
Removable Storage (651) (e.g., CD/DVD)
Non-Removable Storage (652) (e.g., HDD)
Storage Interface Bus (641)
Interface Bus (642)
Output Interfaces (660)
Graphics Processing Unit (661)
Audio Processing Unit (662)
A/V Port(s) (663)
Peripheral Interfaces (670)
Serial interface Controller (671)
Parallel Interface Controller (672)
I/O Port(s) (673)
Communication Interfaces (680)
Network Controller (681)
Comm. Port(s) (682)
Other Computing Device(s) (690)

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2011/081693

A. CLASSIFICATION OF SUBJECT MATTER

See the extra sheet

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)

IPC: G11C29/-; G06F11/- ; H03M13/- ; G06F12/-;

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

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

CNABS, CNTXT; VEN; CNKI; triple parity parities encod+ data recover+ row diagonal span+ storage prime chunk? strip+ HHD RAID SSD set? failure error

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN101512492A, (NETWORK APPLIANCE INC), 19 Aug. 2009(19.08.2009), the whole document	1-33
A	US7613984B2, (NET APP INC), 03 Nov. 2009(03.11.2009), the whole document	1-33
A	CN1692335A, (IBM CORP), 02 Nov. 2005(02.11.2005), the whole document	1-33
A	EP1324200A2, (NET APP INC), 02 Jul. 2003(02.07.2003), the whole document	1-33
A	US2005102470A1, (NEC CORP), 12 May 2005(12.05.2005), the whole document	1-33

☐ 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
27 Jul. 2012(27.07.2012)	**16 Aug. 2012 (16.08.2012)**
Name and mailing address of the ISA/CN The State Intellectual Property Office, the P.R.China 6 Xitucheng Rd., Jimen Bridge, Haidian District, Beijing, China 100088 Facsimile No. 86-10-62019451	Authorized officer ZHAO, Xiaoning Telephone No. (86-10)62412071

Form PCT/ISA /210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2011/081693

Patent Documents referred in the Report	Publication Date	Patent Family	Publication Date
CN101512492A	19.08.2009	WO2007078803A3	07.05.2009
		EP1960883A2	27.08.2008
		JP2009524124A	25.06.2009
		US7640484B2	29.12.2009
		WO2007078803A2	12.07.2007
		CN101512492B	22.02.2012
		US2007089045A1	19.04.2007
US7613984B2	03.11.2009	WO2008082584A2	10.07.2008
		US2008016435A1	17.01.2008
		EP2122467A2	25.11.2009
		WO2008082584A3	05.11.2009
CN1692335A	02.11.2005	KR100690429B1	12.03.2007
		US7085953B1	01.08.2006
		CA2739639A1	13.05.2004
		EP1556763A1	27.07.2005
		KR20050062629A	23.06.2005
		JP2006505035A	09.02.2006
		EP1556763B1	22.03.2006
		DE60304194D1	11.05.2006
		CN100337209C	12.09.2007
		JP4516846B2	04.08.2010
		CA2503129C	09.08.2011
		WO2004040450A1	13.05.2004
		DE60304194T2	28.12.2006
		AU2003271982A1	25.05.2004
EP1324200A2	02.07.2003	JP2003233468A	22.08.2003
		US7409625B2	05.08.2008
		US6993701B2	31.01.2006
		US2007180348A1	02.08.2007
		US2003126523A1	03.07.2003
		JP4823468B2B2	24.11.2011
		US2006107135A1	18.05.2006
		US7203892B2	10.04.2007
US2005102470A1	12.05.2005	JP2005166016A	23.06.2005

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2011/081693

continuation of second sheet

CLASSIFICATION OF SUBJECT MATTER

G06F11/10 (2006.01) i
G11C29/00 (2006.01) i
H03M13/00 (2006.01) n