(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110392092 B
(45)授权公告日 2024.07.12

(21)申请号 201910197641.3

(22)申请日 2019.03.15

(65)同一申请的已公布的文献号
申请公布号 CN 110392092 A

(43)申请公布日 2019.10.29

(30)优先权数据
10-2018-0044420 2018.04.17 KR

(73)专利权人 三星电子株式会社
地址 韩国京畿道

(72)发明人 李昶德 申昇烨 梁耿辅 吴和锡

(74)专利代理机构 北京市立方律师事务所 11330
专利代理师 李娜 赵莎

(51)Int.Cl.
H04L 67/1097(2022.01)
G06F 13/16(2006.01)
G06F 13/10(2006.01)

(56)对比文件
US 2016127492 A1,2016.05.05
CN 107229417 A,2017.10.03
US 2015006663 A1,2015.01.01
CN 103955440 A,2014.07.30
CN 104038550 A,2014.09.10

审查员 翁平

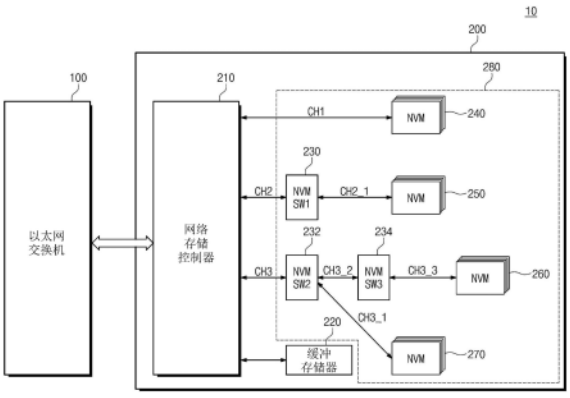
权利要求书3页 说明书10页 附图11页

(54)发明名称

与网络结构连接的网络存储设备

(57)摘要

一种与网络结构连接的网络存储设备包括:网络存储控制器,其执行与所述网络结构的接口连接并转换和处理通过所述网络结构提供的命令;以及非易失性存储器集群,其在所述网络存储控制器的控制下与所述网络存储控制器交换数据。所述非易失性存储器集群包括:通过第一通道与所述网络存储控制器连接的第一非易失性存储器阵列、通过第二通道与所述网络存储控制器连接的非易失性存储器交换机以及第二非易失性存储器阵列,所述第二非易失性存储器阵列在所述非易失性存储器交换机的控制下与所述网络存储控制器通信。



CN 110392092 B

1.一种与网络结构连接的网络存储设备,所述网络存储设备包括:

网络存储控制器,所述网络存储控制器被配置为执行与所述网络结构的接口连接,转换和处理通过所述网络结构提供的命令;以及

非易失性存储器集群,所述非易失性存储器集群被配置为在所述网络存储控制器的控制下与所述网络存储控制器交换数据,

其中,所述非易失性存储器集群包括:

第一非易失性存储器阵列,所述第一非易失性存储器阵列通过第一通道与所述网络存储控制器连接,

非易失性存储器交换机,所述非易失性存储器交换机通过第二通道与所述网络存储控制器连接,以及

第二非易失性存储器阵列,所述第二非易失性存储器阵列被配置为在所述非易失性存储器交换机的控制下与所述网络存储控制器通信,

其中,所述网络存储控制器包括:

主机接口,所述主机接口被配置为接收作为所述命令从所述网络结构提供的具有网络传输格式的第一命令;

存储控制器,所述存储控制器被配置为将所述第一命令转换为具有用于控制所述非易失性存储器集群的格式的第二命令,并基于所述第二命令执行针对所述非易失性存储器集群的操作;以及

接口,所述接口被配置为在所述存储控制器的控制下访问所述非易失性存储器集群,

其中,在所述网络存储控制器中,用于支持以太网协议、NVMe-oF协议和NVMe协议的传输格式在单个层进行处理。

2.根据权利要求1所述的网络存储设备,其中,所述网络存储控制器被配置为将作为所述命令从所述网络结构提供的第一命令转换为用于控制所述非易失性存储器集群的第二命令。

3.根据权利要求1所述的网络存储设备,所述网络存储设备还包括:

缓冲存储器;

第一端口,所述第一端口被配置为与所述网络结构通信;

第二端口,所述第二端口被配置为与所述缓冲存储器通信;以及

第三端口,所述第三端口被配置为与所述非易失性存储器集群通信。

4.根据权利要求3所述的网络存储设备,其中,所述第一端口包括支持网络传输格式的接口。

5.根据权利要求4所述的网络存储设备,其中,所述网络传输格式包括高速以太网系统。

6.根据权利要求1所述的网络存储设备,其中,所述网络存储控制器还包括:

外部计算设备,所述外部计算设备被配置为对所述第一命令或所述第二命令进行处理,以依据从所述网络结构提供的请求执行针对所述非易失性存储器集群的操作。

7.根据权利要求6所述的网络存储设备,其中,所述外部计算设备包括:

一个或更多个处理器,所述一个或更多个处理器被配置为和所述存储控制器一起对所述第一命令或所述第二命令执行分布式处理,或者对所述第一命令或所述第二命令单独执

行处理;或者

加速引擎。

8.一种网络存储设备,所述网络存储设备包括:

网络存储控制器,所述网络存储控制器被配置为将通过网络交换机提供的具有网络传输格式的第一命令转换为具有用于控制非易失性存储器的格式的第二命令,并通过使用接口输出所述第二命令;以及

非易失性存储器集群,所述非易失性存储器集群包括所述非易失性存储器,并且被配置为依据从所述网络存储控制器输出的所述第二命令对写数据进行存储,或者将所存储的数据输出到所述网络存储控制器,

其中,所述网络存储控制器包括一个或更多个外部计算设备,所述一个或更多个外部计算设备被配置为基于所述第二命令执行针对所述非易失性存储器集群的操作,并且

其中,所述网络存储控制器还包括:

主机接口,所述主机接口被配置为接收所述第一命令;

存储控制器,所述存储控制器被配置为将所述第一命令转换为用于控制所述非易失性存储器集群的所述第二命令;以及

接口,所述接口被配置为在所述存储控制器的控制下访问所述非易失性存储器集群,

其中,在所述网络存储控制器中,用于支持以太网协议、NVMe-oF协议和NVMe协议的传输格式在单个层进行处理。

9.根据权利要求8所述的网络存储设备,其中,所述一个或更多个外部计算设备包括一个或更多个处理器,所述一个或更多个处理器根据主机的请求来处理命令,所述命令包括所述第一命令或所述第二命令。

10.根据权利要求9所述的网络存储设备,其中,所述一个或更多个外部计算设备根据所述主机的所述请求来执行网络管理操作。

11.根据权利要求8所述的网络存储设备,其中,所述网络存储控制器还包括嵌入式计算设备,并且

其中,所述一个或更多个外部计算设备包括一个或更多个处理器,所述一个或更多个处理器被配置为和所述嵌入式计算设备一起对所述第一命令或所述第二命令执行分布式处理。

12.根据权利要求8所述的网络存储设备,其中,所述非易失性存储器集群包括第一非易失性存储器阵列,所述第一非易失性存储器阵列通过第一通道与所述网络存储控制器连接并且通过所述接口被控制。

13.根据权利要求12所述的网络存储设备,其中,所述非易失性存储器集群还包括:

第二非易失性存储器阵列;以及

第一非易失性存储器交换机,所述第一非易失性存储器交换机通过第二通道与所述网络存储控制器连接并且被配置为执行所述第二非易失性存储器阵列与所述接口之间的接口连接。

14.根据权利要求13所述的网络存储设备,其中,所述第一非易失性存储器交换机包括:

第一交换机控制器,所述第一交换机控制器被配置为将从所述接口提供的第一地址映

射到所述第二非易失性存储器阵列;以及

第一缓冲器,所述第一缓冲器被配置为对在所述接口与所述第二非易失性存储器阵列之间传输的数据进行缓冲。

15.根据权利要求14所述的网络存储设备,其中,所述非易失性存储器集群还包括:

第三非易失性存储器阵列;以及

第二非易失性存储器交换机,所述第二非易失性存储器交换机被配置为执行所述第一非易失性存储器交换机与所述第三非易失性存储器阵列之间的接口连接。

16.根据权利要求15所述的网络存储设备,其中,所述第一非易失性存储器交换机和所述第二非易失性存储器交换机彼此独立地执行地址映射。

17.一种网络存储设备,所述网络存储设备包括:

网络存储控制器,所述网络存储控制器被配置为将从网络结构提供且具有网络传输格式的第一命令转换为具有用于控制非易失性存储器设备的格式的第二命令;

第一交换机控制器,所述第一交换机控制器被配置为通过第一通道从所述网络存储控制器接收所述第二命令;以及

第一多个原始NAND闪存器件,所述第一多个原始NAND闪存器件通过第二通道连接到所述第一交换机控制器,

其中,所述第一交换机控制器被配置为响应于所述第二命令来访问所述第一多个原始NAND闪存器件,并将从所述网络存储控制器提供的地址映射到所述第一多个原始NAND闪存器件,并且

其中,所述网络存储控制器包括:

主机接口,所述主机接口被配置为接收所述第一命令;

存储控制器,所述存储控制器被配置为将所述第一命令转换为所述第二命令,并基于所述第二命令执行针对所述第一多个原始NAND闪存器件的操作;以及

接口,所述接口被配置为在所述存储控制器的控制下通过所述第一交换机控制器访问所述第一多个原始NAND闪存器件,

其中,在所述网络存储控制器中,用于支持以太网协议、NVMe-oF协议和NVMe协议的传输格式在单个层进行处理。

18.根据权利要求17所述的网络存储设备,所述网络存储设备还包括第二多个原始NAND闪存器件,所述第二多个原始NAND闪存器件通过第三通道连接到所述网络存储控制器。

19.根据权利要求17所述的网络存储设备,所述网络存储设备还包括:

第二多个原始NAND闪存器件;以及

第二交换机控制器,所述第二交换机控制器被配置为提供所述第二多个原始NAND闪存器件与所述第一交换机控制器之间的接口连接。

20.根据权利要求19所述的网络存储设备,其中,所述第二交换机控制器被配置为将从所述第一交换机控制器提供的地址映射到所述第二多个原始NAND闪存器件。

21.根据权利要求17所述的网络存储设备,其中,所述网络传输格式包括高速以太网系统。

与网络结构连接的网络存储设备

[0001] 相关申请的交叉引用

[0002] 要求于2018年4月17日在韩国知识产权局递交的韩国专利申请No.10-2018-0044420,其整体通过引用并入本文。

技术领域

[0003] 本公开在这里涉及半导体存储器设备,更具体地涉及存储大量数据的网络存储设备。

背景技术

[0004] 固态硬盘(以下称为SSD)是基于闪存的大容量存储设备的示例。随着对大容量存储的需求增加,SSD的使用最近变得多样化。例如,在各种其他实现中,SSD可以被表征为细分为实现用作服务器的SSD、实现用于客户端的SSD以及实现用作数据中心的SSD。SSD接口用于提供适合实现的最高速度和可靠性。为了满足高速和可靠性的要求,已积极开发和应用了基于串行高级技术附件(SATA)、串行连接小组件接口(SAS)或外围组件互连快速(PCIe)的非易失性存储器快速(NVMe)接口规范。

[0005] 目前,正在积极开发能够在诸如大容量存储器中心之类的系统中易于扩展的SSD接口。特别地,NVMe over fabric(NVMe-oF)规范作为在诸如以太网交换机之类的网络结构上安装SSD的标准正得到积极开发。NVMe-oF通过广泛的存储网络结构(例如,以太网、Fibre Channel™和InfiniBand™)来支持NVMe存储协议。

[0006] NVMe存储协议也被应用于NVMe SSD。因此,包括NVMe SSD的存储接口将用于网络结构的协议转换为NVMe-oF协议,或者具有数据缓冲器的功能。然而,在这种情况下,由于需要对应于多个协议层的协议转换,因此延迟的增加是不可避免的。

[0007] 因此,为了减少延迟并且容易地扩展容量,需要一种使用非易失性存储器设备或非易失性存储器阵列作为存储介质并且有效地执行多协议转换的技术,而不是使用NVMe SSD。

发明内容

[0008] 本发明构思的实施例提供了一种以太网存储设备,以用于简化网络存储设备的控制器结构并易于扩展。

[0009] 本发明构思的实施例提供了一种与网络结构连接的网络存储设备。所述网络存储设备包括:网络存储控制器,所述网络存储控制器执行与所述网络结构的接口连接,并转换和处理通过所述网络结构提供的命令;以及非易失性存储器集群,所述非易失性存储器集群在所述网络存储控制器的控制下与所述网络存储控制器交换数据。所述非易失性存储器集群包括通过第一通道与所述网络存储控制器连接的第一非易失性存储器阵列、通过第二通道与所述网络存储控制器连接的非易失性存储器交换机以及第二非易失性存储器阵列,所述第二非易失性存储器阵列在所述非易失性存储器交换机的控制下与所述网络存储控

制器通信。

[0010] 本发明构思的实施例还提供了一种网络存储设备,包括:网络存储控制器,所述网络存储控制器将通过网络交换机提供的第一命令转换为用于控制非易失性存储器的第二命令,并通过使用接口输出所述第二命令;以及非易失性存储器集群,所述非易失性存储器集群包括所述非易失性存储器,并且依据从所述网络存储控制器输出的所述第二命令对写数据进行存储,或者将所存储的数据输出到存储控制器。所述网络存储控制器包括一个或更多个外部计算设备,所述一个或更多个外部计算设备基于所述第二命令执行针对所述非易失性存储器集群的操作。

[0011] 本发明构思的实施例还提供了一种网络存储设备,包括:网络存储控制器,所述网络存储控制器将从网络结构提供且具有网络传输格式的第一命令转换为具有用于控制非易失性存储器设备的格式的第二命令;第一交换机控制器,所述第一交换机控制器通过第一通道从所述网络存储控制器接收所述第二命令;通过第二通道连接到所述第一交换机控制器的第一多个原始NAND闪存器件。所述第一交换机控制器响应于所述第二命令来访问所述第一多个原始NAND闪存器件,并将从所述网络存储控制器提供的地址映射到所述第一多个原始NAND闪存器件。

附图说明

[0012] 通过以下结合附图的详细描述,本发明构思的上述和其他目的和特征将变得容易理解。

[0013] 图1示出了根据本发明构思的实施例的网络存储系统的框图。

[0014] 图2示出了图1的网络存储控制器的实施例的框图。

[0015] 图3示出了图2的存储控制器的配置的框图。

[0016] 图4示出了图1的网络存储控制器的另一实施例的框图。

[0017] 图5示出了根据本发明构思的另一实施例的网络存储控制器的框图。

[0018] 图6示出了根据本发明构思的另一实施例的网络存储控制器的框图。

[0019] 图7示出了本发明构思的NVM集群的一部分的图。

[0020] 图8示出了本发明构思的NVM集群的另一部分的图。

[0021] 图9示出了本发明构思的NVM集群的另一部分的图。

[0022] 图10示出了根据本发明构思的实施例可配置的网络存储系统的框图。

[0023] 图11示出了根据本发明构思的实施例的网络存储系统的框图。

具体实施方式

[0024] 如在本发明构思的领域中的传统,可以根据执行所描述的一个或多个功能的块来描述和说明实施例。这些块(这里可称为单元或模块等)物理地由模拟和/或数字电路(诸如逻辑门、集成电路、微处理器、微控制器、存储器电路、无源电子元件、有源电子元件、光学组件、硬连线电路等)来实现,并且可以可选地由固件和/或软件来驱动。电路例如可以被包括在一个或更多个半导体芯片中,或者被包括在诸如印刷电路板等的基板支撑件上。构成块的电路可以由专用硬件来实现,或者由处理器(例如,一个或更多个经编程的微处理器和相关电路)来实现,或者由执行块的一些功能的专用硬件以及执行块的其他功能的处理器的

组合来实现。在不脱离本发明构思的范围的情况下,实施例的每个块可以被物理地分成两个或更多个交互和离散块。同样地,在不脱离本发明构思的范围的情况下,可以将实施例的块物理地组合成更复杂的块。

[0025] 在下面的描述中,使用闪存器件的固态硬盘(SSD)将被用作用于描述本发明构思的特征和功能的存储设备的示例。然而,本领域技术人员应当依据这里公开的内容,容易地理解本发明构思的其他优点和性能。可以通过其他实施例来实现或应用本发明构思。另外,在不脱离本发明构思的范围和精神以及任何其他目的的情况下,可以根据应用来改变或修改详细描述。

[0026] 图1示出了根据本发明构思的实施例的网络存储系统的框图。参照图1,网络存储系统10包括以太网交换机100和网络存储设备200。主机(未示出)通过以太网交换机100将以太网协议的命令和数据发送到网络存储设备200。网络存储设备200将以太网协议的命令和数据转换成要被直接发送到NVM集群280的命令和数据,而不用转换成中间层(例如,NVMe协议)。

[0027] 以太网交换机100将从主机提供的命令或数据提供给网络存储设备200。或者,以太网交换机100将网络存储设备200依据主机的请求提供的数据发送到主机。也就是说,以太网交换机100可以是使用以太网协议的网络结构或交换机。当向网络存储设备200发送命令和数据时,以太网交换机100可以根据包括NVMe-oF存储协议的以太网协议,发送命令和数据。而且,以太网交换机100可以根据以太网协议与网络存储设备200交换响应或数据。

[0028] 网络存储设备200响应于通过以太网交换机100提供的命令CMD或数据,控制NVM集群280。NVM集群280包括交换机(SW)NVM SW1230、NVM SW2 232和NVM SW3 234(可以在下文中被称为NVM交换机230、232和234)以及非易失性存储器设备240、250、260和270。网络存储设备200可以将从以太网交换机100提供的以太网协议格式的命令或数据直接转换为用于控制非易失性存储器设备(NVM)240、250、260和270的命令或数据格式。为了执行该功能,网络存储设备200包括网络存储控制器210。在网络存储控制器210中,用于支持以太网协议、NVMe-oF协议和NVMe协议的传输格式可以在单层进行处理。由于可以在单个协议层(即,单个层)管理网络传输格式,因此可以对命令路径和数据路径进行分离。网络存储控制器210可以用单个芯片实现。网络存储控制器210提供以太网交换机100与网络存储设备200之间的接口。网络存储控制器210可以将从以太网交换机100提供的网络传输格式(例如,以太网协议格式(例如,包))的命令或数据直接转换为要被应用于非易失性存储器设备240、250、260和270的命令或数据格式。在网络存储控制器210中,用于支持以太网协议、NVMe-oF协议和NVMe协议的传输格式可以在单个层进行处理。稍后将更全面地描述网络存储控制器210的操作。

[0029] 缓冲存储器220可以被用作用于在网络存储设备200与以太网交换机100之间的数据交换的数据缓冲器。例如,缓冲存储器220可以用动态随机存取存储器(DRAM)实现。然而,缓冲存储器220不限于DRAM。例如,缓冲存储器220可以用易失性存储器实现,诸如静态RAM(SRAM)或同步DRAM(SDRAM)、非易失性RAM NV-RAM或他们的组合。

[0030] NVM集群280包括NVM交换机230、232和234以及非易失性存储器设备240、250、260和270。这里,非易失性存储器设备240、250、260和270中的每一个可以包括非易失性存储器阵列(NVM阵列)。也就是说,非易失性存储器设备240可以包括一个或更多个与非(NAND)闪

存器件,并且剩余的非易失性存储器设备250、260和270也可以以类似的形式来实现或进行组合。

[0031] NVM交换机230、232和234被提供为用于将非易失性存储器设备扩展连接到网络存储控制器210的闪存接口的配置。例如,NVM交换机230、232和234中的每一个可以在与其连接的非易失性存储器设备上独立地执行地址映射。将参考以下附图更全面地描述NVM交换机230、232和234的结构以及与NVM交换机230、232和234连接的非易失性存储器设备。

[0032] 根据以上描述,本发明构思的网络存储设备200包括网络存储控制器210,其可以将网络协议直接转换为用于控制非易失性存储器设备的命令或数据格式。因此,网络存储控制器210可以在从网络结构发送的命令和数据通过彼此独立的命令路径和数据路径被处理之后,将该从网络结构发送的命令和数据加载/存储到非易失性存储器设备240、250、260和270。而且,通过使用NVM交换机可以容易地扩展非易失性存储器设备240、250、260和270。

[0033] 图2示出了图1的网络存储控制器的实施例的框图。参照图2,根据实施例的网络存储控制器210a包括存储控制器214。

[0034] 存储控制器214通过使用嵌入式CPU(例如,嵌入式计算设备)211来执行以太网交换机100与NVM集群280之间的接口连接。具体地,存储控制器214通过第一端口PT1与以太网交换机100连接。例如,第一端口PT1可以执行用于与以太网交换机100连接的网络接口功能。因此,第一端口PT1支持网络传输格式。

[0035] 存储控制器214可以通过第二端口PT2与缓冲存储器220交换数据。存储控制器214通过第三端口PT3与NVM集群280连接。这里,第三端口PT3可以包括如图1所示的多个通道CH1、CH2和CH3。通过第三端口PT3提供的通道CH1、CH2和CH3中的至少一个,存储控制器214可以直接与NVM 280的非易失性存储器(NVM)阵列(例如,通过通道CH1的非易失性存储器设备240)连接。或者,通过第三端口PT3提供的通道CH1、CH2和CH3中的至少一个,存储控制器214可以与NVM交换机230和232连接。或者,通过第三端口PT3提供的通道CH1、CH2和CH3中的至少一个,存储控制器214可以与级联NVM交换机(例如,NVM交换机232和234)连接。

[0036] 图3示出了图2的存储控制器的示例性配置的框图。参照图3,存储控制器214包括嵌入式CPU 211、工作存储器213、主机接口(I/F)215、缓冲管理器217和闪存接口(I/F)219。

[0037] 嵌入式CPU 211可以将执行与NVM集群280相关联的读/写操作所需的各种控制信息发送到主机接口215和闪存接口219的寄存器。嵌入式CPU 211可以基于为网络存储控制器210的各种控制操作而提供的固件或操作系统OS来操作。例如,嵌入式CPU 211可以在其他各种控制操作中执行用于管理NVM集群280的垃圾回收,或者用于执行地址管理和磨损均衡的闪存转换层(FTL)。具体地,嵌入式CPU 211可以调用并执行加载到工作存储器213的存储管理器216。

[0038] 当执行存储管理器216时,嵌入式CPU 211可以在单层上关于通过以太网交换机100提供的命令或数据来处理支持以太网协议、NVMe-oF协议和NVMe协议的传输格式。另外,嵌入式CPU 211可以在从网络结构发送的命令和数据通过彼此独立的命令路径和数据路径被处理之后,将该从网络结构发送的命令和数据加载/存储到NVM集群280。

[0039] 工作存储器213可以被用作嵌入式CPU 211的操作存储器、高速缓冲存储器或缓冲存储器。工作存储器213可以存储嵌入式CPU 211执行的代码和命令。工作存储器213可以存储由嵌入式CPU 211处理的数据。在实施例中,工作存储器213例如可以用静态随机存取存

储器(SRAM)实现。具体地,存储管理器216可以被加载到工作存储器213。当存储管理器216由嵌入式CPU 211执行时,存储管理器216可以在单层处理从以太网交换机100提供的命令或数据的传输格式的转换。另外,存储管理器216可以在命令路径和数据路径是分离的状态下,处理从网络结构发送的命令或数据。另外,闪存转换层FTL或各种存储器管理模块可以被存储在工作存储器213中。

[0040] 存储管理器216可以收集和调整关于NVM集群280的总体信息。即,存储管理器216可以维护和更新存储在NVM集群280中的数据的状态或映射信息。因此,即使从以太网交换机100进行访问请求,存储管理器216也可以以高速向以太网交换机100提供所请求的数据,或者可以以高速写入请求写入的数据。另外,由于存储管理器216具有管理用于管理数据地址的映射表的权限,因此存储管理器216可以在NVM集群280内执行数据迁移或者在必要时校正映射信息。

[0041] 主机接口215(例如,第一端口PT1)可以与以太网交换机100通信。主机接口215可以识别从以太网交换机100提供的网络传输格式的命令。例如,主机接口215可以通过使用诸如Fiber Channel™或InfiniBand™的高速以太网系统的通信协议与以太网交换机100通信。主机接口215可以转换从以太网交换机100发送的网络传输格式的命令,并且可以将转换后的命令提供给闪存接口219。或者,主机接口215可以转换通过闪存接口219提供的数据,以便包括网络传输格式,并且可以将转换后的数据发送到以太网交换机100。

[0042] 缓冲管理器217可以控制缓冲存储器220的读取和写入操作(参照图1)。例如,缓冲管理器217将写数据或读数据临时存储在缓冲存储器220中。缓冲管理器217可以在嵌入式CPU 211的控制下以流为单位对缓冲存储器220的存储区进行分类和管理。

[0043] 闪存接口219与NVM集群280交换数据。闪存接口219通过相应的存储器通道CH1、CH2和CH3将从缓冲存储器220发送的数据写入NVM集群280。可以通过闪存接口219收集从NVM集群280读出并通过存储器通道提供的读取数据。之后,可以将收集的数据存储到缓冲存储器220中。

[0044] 上述结构的存储控制器214可以将用于与以太网交换机100通信的网络协议直接转换为闪存级别的命令或数据。因此,通过网络结构提供的命令或数据不会经历多个相继的转换过程,这些转换过程例如通过以太网网络接口卡(NIC)、TCP/IP卸载引擎和PCIe交换机来执行。根据上述特征,从以太网交换机100发送的命令或数据可以在通过彼此分离的命令路径和数据路径处理之后被加载/存储到NVM集群280。

[0045] 具体地,存储控制器214可以用单个芯片实现。由于存储控制器214用单个芯片实现,所以本发明构思的网络存储设备200可以是轻量级、薄型和小型的。因此,本发明构思的网络存储设备200在网络结构上提供了低延迟、高经济可行性和高可扩展性。

[0046] 图4示出了图1的网络存储控制器的另一实施例的框图。参照图4,根据另一实施例的网络存储控制器210b包括外部处理器211b和存储控制器214b。

[0047] 存储控制器214b通过使用外部处理器(例如,外部计算设备)211b来执行以太网交换机100与NVM集群280之间的接口连接。外部处理器211b可以转换从以太网交换机100发送的以太网协议的命令或数据。另外,外部处理器211b可以控制存储控制器214b,以便依据通过以太网交换机100提供的请求来搜索存储在NVM集群280中的数据或向以太网交换机100输出所发现的数据。外部处理器211b可以包括专用于网络管理的一个或更多个处理器。网

络存储控制器210b可以包括一个或更多个外部处理器211b。

[0048] 外部处理器211b可以执行包括在存储控制器214b中的嵌入式处理器(未示出)的功能,并且可以和嵌入式处理器一起来执行工作负载的分布式处理。另外,外部处理器211b可以执行从主机请求的单独操作的处理,例如网络管理。

[0049] 存储控制器214b通过第一端口PT1与以太网交换机100电连接。存储控制器214b可以通过第二端口PT2与缓冲存储器220交换数据。存储控制器214b通过第三端口PT3与NVM集群280连接。这里,第三端口PT3可以包括如图1所示的多个通道CH1、CH2和CH3。另外,存储控制器214b通过第四端口PT4与外部处理器211b连接。存储控制器214b例如可以通过使用外部处理器211b执行转换协议的操作(例如,以便提供第二命令)、搜索存储在NVM集群280中的数据的操作或者(例如,基于转换的协议)将数据加载/存储到NVM集群280的操作。第四端口PT4可以用PCIe接口实现,该PCIe接口可以与多个处理器连接。然而,本发明构思不限于此。除了存储控制器214b由外部处理器211b控制之外,存储控制器214b可以具有与图3的存储控制器214的配置和功能相类似的配置和功能。这里,存储控制器214b可以用一个芯片实现,网络存储控制器210b可以在一个PCB板上实现。

[0050] 图5示出了根据本发明构思的另一实施例的网络存储控制器210c的框图。参照图5,网络存储控制器210c可以用一个芯片实现。

[0051] 网络存储控制器210c包括与以太网交换机100连接的第一端口PT1。第一端口PT1可以是用于将网络存储控制器210c与诸如Fibre channel™或InfiniBand™的高速以太网系统连接的硬件端口。第一端口PT1可以将以太网协议的命令或数据发送到主机接口215。主机接口215可以根据以太网协议来执行与以太网交换机100的通信。例如,高速以太网系统(诸如Fibre channel™或InfiniBand™)的通信协议可以通过主机接口215被转换为用于控制NVM集群280的传输格式。

[0052] 缓冲管理器217可以通过第二端口PT2来控制缓冲存储器220(参照图1)的读取和写入操作。例如,缓冲管理器217将写数据或读数据临时存储到缓冲存储器220。缓冲管理器217可以在嵌入式CPU 211c的控制下以流为单位对缓冲存储器220的存储区进行分类和管理。

[0053] 闪存接口219通过第三端口PT3与NVM集群280交换数据。闪存接口219通过各自的通道将从缓冲存储器220发送的数据写入NVM集群280。可以通过闪存接口219收集从NVM集群280读出并且通过存储器通道提供的读取数据。之后,可以将收集了的数据存储在缓冲存储器220中。

[0054] 嵌入式CPU 211c可以将执行与NVM集群280相关联的读/写操作所需的各种控制信息发送到主机接口215和闪存接口219的寄存器。嵌入式CPU 211c可以基于为网络存储控制器210c的各种控制操作提供的固件或操作系统OS来操作。例如,嵌入式CPU 211c可以在各种其他控制操作中执行用于管理NVM集群280的垃圾回收,或者用于执行地址管理和磨损均衡的闪存转换层(FTL)。具体地,嵌入式CPU 211c可以调用并执行加载到工作存储器213的操作系统OS或管理模块。嵌入式CPU 211c可以在单层上关于提供给网络存储控制器210c的命令或数据来处理用于支持以太网协议、NVMe-oF协议和NVMe协议的传输格式。

[0055] 如上所述,根据另一实施例的网络存储控制器210c可以直接执行网络接口,并且可以包括嵌入式CPU 211c。因此,网络存储控制器210c可以直接连接到以太网交换机100,

并且可以控制NVM集群280。

[0056] 图6示出了根据本发明构思的另一实施例的网络存储控制器210d的框图。参照图6,网络存储控制器210d可以通过使用外部处理器(例如,外部计算设备)211d来执行NVM集群280与以太网交换机100之间的接口连接。在实施例中,网络存储控制器210d可以通过使用一个或更多个外部处理器(例如,外部计算设备)211d来执行NVM集群280与以太网交换机100之间的接口连接。

[0057] 网络存储控制器210d包括与以太网交换机100连接的第一端口PT1。第一端口PT1可以是用于将网络存储控制器210d与诸如Fibre channel™或InfiniBand™的高速以太网系统连接的硬件端口。第一端口PT1可以从以太网交换机100接收以太网协议的命令或数据。第一端口PT1将接收到的命令或数据发送到主机接口215。这里,嵌入式CPU 211c、工作存储器213、主机接口215、缓冲管理器217、闪存接口219、第一端口PT1、第二端口PT2和第三端口PT3可以被配置为与图5中的相同。

[0058] 相反,网络存储控制器210d可以通过使用外部处理器211d来执行以太网交换机100与NVM集群280之间的接口连接。外部处理器211d可以将从以太网交换机100发送的以太网协议的命令(例如,第一命令)或数据转换为第二命令。另外,外部处理器211d可以控制网络存储控制器210d,以便依据从以太网交换机100提供的请求来搜索存储在NVM集群280中的数据或者将找到的数据输出到以太网交换机100。外部处理器211d可以和嵌入式CPU 211c一起对工作负载执行分布式处理。在实施例中,外部处理器211d可以在嵌入式CPU 211c的控制下处理所分配的命令(例如,第二命令)。另外,外部处理器211d可以执行从主机请求的单独操作的处理,例如网络管理。外部处理器211d可以包括一个或更多个处理器。提供用于连接外部处理器211d和网络存储控制器210d的第四端口PT4。外部处理器211d和网络存储控制器210d可以通过例如支持PCIe接口的第四端口PT4彼此通信。

[0059] 图7示出了本发明构思的NVM集群的一部分的图。参照图7,NVM阵列240可以直接与诸如图1中所示的网络存储控制器210连接并且可以交换数据。

[0060] NVM阵列240例如包括NAND闪存器件原始(Raw)NAND_11、原始NAND_12、原始NAND_13、……、原始NAND_1i(i是大于3的自然数)。NAND闪存器件原始NAND_11、原始NAND_12、原始NAND_13、……、原始NAND_1i可以是单个器件的组合,或者可以以一个包括多个器件的封装的形式进行提供。NAND闪存器件原始NAND_11、原始NAND_12、原始NAND_13、……、原始NAND_1i中的每一个不包括第一通道CH1与器件之间的单独接口或交换机。而且,与被管理的NAND闪存器件相比,原始NAND闪存器件不执行存储器管理功能,该被管理的NAND闪存器件可以包括执行一些存储器管理功能的存储控制器。

[0061] NAND闪存器件原始NAND_11、原始NAND_12、原始NAND_13、……、原始NAND_1i中的每一个与第一通道CH1连接。例如,第一通道CH1可以是网络存储控制器210与共享相同数据线DQ的NAND闪存器件原始NAND_11、原始NAND_12、原始NAND_13、……、原始NAND_1i之间的数据路径。也就是说,与第一通道CH1连接的NAND闪存器件原始NAND_11、原始NAND_12、原始NAND_13、……、原始NAND_1i可以共享相同的数据线DQ。通过第一通道CH1,网络存储控制器210的闪存接口219(例如,参照图3)可以选择NAND闪存器件原始NAND_11、原始NAND_12、原始NAND_13、……、原始NAND_1i中的任何一个或者可以控制交错的通道。

[0062] 图8示出了本发明构思的NVM集群的另一部分的图。参照图8,NVM阵列250通过诸如

图1中所示的NVM交换机230与网络存储控制器210连接。

[0063] NVM交换机230通过第二通道CH2与网络存储控制器210通信。例如,NVM交换机230可以通过第二通道CH2与网络存储控制器210的闪存接口219(例如,参照图3)连接。闪存接口219可以通过使用独立于第一通道CH1的第二通道CH2来访问NVM交换机230和NVM阵列250。

[0064] 在本发明构思的实施例中,NVM交换机230包括交换机(SW)控制器231和缓冲器233。交换机控制器231可以执行地址映射功能,以用于将通过第二通道CH2提供的地址映射到NVM阵列250的实际地址。交换机控制器231可以将与存储器通道CH2_1连接的NVM阵列250的地址映射到闪存接口219提供的地址上。当设备被添加到NVM阵列250或从NVM阵列250移除时,交换机控制器231也可以更新地址映射。缓冲器233可以临时存储在网络存储控制器210与NVM阵列250之间交换的数据。缓冲器233可以用诸如DRAM的易失性存储器来实现。

[0065] NVM阵列250例如包括NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j(j是大于3的自然数)。可以在交换机控制器231的控制下选择NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j中的每一个。NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j中的每一个在存储器通道CH2_1与器件之间不包括单独的接口或交换机。NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j可以是单个器件的组合,或者可以以一个包括多个器件的封装的形式进行提供。

[0066] NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j中的每一个与存储器通道CH2_1连接。例如,存储器通道CH2_1可以是NVM交换机230与共享相同数据线DQ的NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j之间的数据路径。也就是说,与存储器通道CH2_1连接的NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j可以共享相同的数据线DQ。

[0067] 以上描述了NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j通过NVM交换机230与网络存储控制器210连接的示例。这里,给出了在NVM交换机230中提供一个存储器通道CH2_1的描述。然而,应当理解,本发明构思不限于此。也就是说,应当理解,在其他实施例中,两个或更多个存储器通道可以与NVM交换机230连接,并且多个NVM阵列可以与存储器通道连接。

[0068] 通过PCIe交换机对SSD的扩展存在实质限制。然而,理论上,对通过NVM交换机230来扩展设备单元的NAND闪存器件原始NAND_21、原始NAND_22、原始NAND_23、……、原始NAND_2j没有限制。因此,在使用本发明构思的技术的情况下,可以容易地扩展网络存储设备200的容量。

[0069] 图9示出了本发明构思的NVM集群的另一部分的图。参照图9,NVM阵列260和270通过诸如图1所示的NVM交换机232和234与网络存储控制器210连接。这里,NVM交换机232和234可以具有级联连接结构。

[0070] NVM交换机232通过第三通道CH3与网络存储控制器210连接。NVM交换机232通过第三通道CH3与网络存储控制器210的闪存接口219(参照图3)连接。闪存接口219可以通过使用独立于第一通道CH1和第二通道CH2的第三通道CH3来访问NVM交换机232。

[0071] 在本发明构思的实施例中,NVM交换机232包括交换机控制器235和缓冲器237。交

换机控制器235可以执行地址映射功能,以用于将通过第三通道CH3提供的地址映射到以级联方式与NVM交换机232连接的NVM阵列270和NVM交换机234的地址。交换机控制器235可以将与存储器通道CH3_1连接的NVM阵列270的地址映射到闪存接口219提供的地址。交换机控制器235可以将与存储器通道CH3_2连接的NVM交换机234的地址映射到闪存接口219提供的地址。

[0072] NVM交换机234通过存储器通道CH3_2与NVM交换机232连接。NVM交换机234通过存储器通道CH3_3与NVM阵列260连接。NVM交换机234包括交换机控制器236和缓冲器238。

[0073] 交换机控制器235或交换机控制器236可以通过已添加的存储器端口(存储器通道)来识别另外的NVM阵列或NVM交换机,并且可以执行接口连接。这里,构成NVM阵列260和270的NAND闪存器件原始NAND_31、原始NAND_32、原始NAND_33、……、原始NAND_3k和原始NAND_41、原始NAND_42、原始NAND_43、……、原始NAND_4m可以是单个器件的组合或者可以以包括多个器件的封装的形式进行设置。NAND闪存器件原始NAND_31、原始NAND_32、原始NAND_33、……、原始NAND_3k和原始NAND_41、原始NAND_42、原始NAND_43、……、原始NAND_4m中的每一个在相应的存储器通道CH3_1和存储器通道CH3_3与器件之间不包括单独的接口或交换机。

[0074] 以上描述了NVM阵列240、250、260和270通过NVM交换机230、232和234与网络存储控制器210连接的示例。这里,NVM交换机230、232和234可以级联连接,并且可以与单个NVM产品或阵列单元连接,而不是与SSD单元连接。通过PCIe交换机来扩展SSD存在限制。然而,理论上,对通过NVM交换机230、232和234来扩展设备单元中的存储容量没有限制。因此,在使用本发明构思的技术的情况下,网络存储设备200的容量可以很容易被扩大。

[0075] 图10示出了根据本发明构思的实施例可配置的网络存储系统的框图。参照图10,网络存储设备300包括用于执行与以太网交换机100、外部处理器305、缓冲存储器320和NVM集群380接口连接的网络存储控制器310。外部处理器(例如,外部计算设备)305可以用为特定目的而提供的处理器或计算设备或者加速引擎来实现。

[0076] 诸如服务器或数据中心的主机(未示出)通过以太网交换机100将以太网协议的命令和数据发送到网络存储设备300。网络存储设备300可以将以太网协议的命令和数据转换为要被发送到NVM集群380的命令和数据,而无需转换为PCIe协议。

[0077] 网络存储控制器310响应于通过以太网交换机100提供的以太网协议的命令或数据来控制直连NVM集群380。具体地,网络存储控制器310可以由外部处理器305来控制。应当理解,外部处理器305可以用一个或更多个处理器来实现,以便与高容量NVM集群380交换数据并处理数据或命令。也就是说,网络存储设备300可以包括一个或更多个外部处理器305。外部处理器305可以利用网络存储控制器310内的嵌入式CPU来实现。

[0078] 在网络存储控制器310中,可以在单个层处理用于支持以太网协议、NVMe-oF协议和NVMe协议的传输格式。网络存储控制器310提供以太网交换机100和网络存储设备300之间的接口连接。网络存储控制器310可以直接将从以太网交换机100提供的以太网协议格式(例如,包)的命令或数据转换为要应用于NVM集群380的命令或数据格式。

[0079] 缓冲存储器320可以用作用于网络存储设备300与以太网交换机100之间的数据交换的数据缓冲器。例如,缓冲存储器320可以用DRAM来实现。然而,缓冲存储器320不限于DRAM。例如,缓冲存储器320可以用诸如SRAM或SDRAM的易失性存储器、非易失性RAM NV-RAM

或他们的组合来实现。

[0080] NVM集群380可以包括一个或更多个NVM交换机330、332、334、336和338以及一个或更多个NVM阵列340、350、352、354、360、362、370、372和374。NVM阵列340、350、352、354、360、362、370、372和374中的每一个可以用非易失性存储器设备阵列来实现。也就是说,NVM阵列340、350、352、354、360、362、370、372和374中的每一个可以包括多个NAND闪存器件。

[0081] 提供NVM交换机330、332、334、336和338作为用于将非易失性存储器设备扩展连接到网络存储控制器310的闪存接口的配置。例如,每个NVM交换机330、332、334、336和338可以独立地在与其连接的非易失性存储器设备或非易失性存储器设备阵列上执行地址映射。

[0082] NVM交换机330与三个NVM阵列350、352和354连接,并且可以执行与网络存储控制器310的接口连接。NVM交换机332可以执行NVM交换机334与网络存储控制器310之间的接口连接。NVM交换机334可以控制NVM交换机336和NVM阵列360。NVM交换机336可以控制NVM交换机338和NVM阵列362。NVM交换机338与三个NVM阵列370、372和374连接,并且可以执行与NVM交换机336的接口连接。

[0083] 以上简要描述了可以与网络存储控制器310和NVM阵列340、350、352、354、360、362、370、372和374连接的NVM交换机330、332、334、336和338的连接形式。应当很好地理解,可以在没有NVM交换机的情况下进行与网络存储控制器310的连接,例如连接到网络存储控制器310的NVM阵列340。

[0084] 图11示出了根据本发明构思的实施例的网络存储系统的框图。参照图11,本发明构思的网络存储系统1000包括服务器1100、网络结构1200和多个以太网SSD 1300和1400。

[0085] 服务器1100通过网络结构1200与多个以太网SSD 1300和1400连接。服务器1100可以通过使用以太网协议将命令和数据发送到多个以太网SSD 1300和1400。服务器1100可以接收从多个以太网SSD 1300和1400中的至少一个提供的以太网协议的数据。网络结构1200可以是网络交换机。

[0086] 多个以太网SSD 1300和1400中的每一个可以用图1至图10的存储设备来实现。即,在以太网SSD 1300中,网络存储控制器1310可以直接与NVM阵列1320连接,而不通过NVM交换机。相反,包括在以太网SSD 1400中的网络存储控制器1410可以通过NVM交换机1412和1414来控制NVM阵列1422和1424。而且,以太网SSD控制器1410可以直接控制NVM阵列1420而不使用NVM交换机。

[0087] 根据本发明构思的实施例,可以提供网络存储设备,其容量可以以非易失性存储器设备或诸如NAND闪存器件的非易失性存储器阵列为单位进行扩展。因此,本发明构思的网络存储设备可以与网络结构连接,以针对多个协议执行单层的协议转换,从而减少延迟并且能够扩展存储器容量。

[0088] 尽管已经参考本发明构思的示例性实施例描述了本发明构思,但是对于本领域普通技术人员显而易见的是,在不脱离所附权利要求中阐述的发明构思的精神和范围的情况下,可以对其进行各种改变和修改。

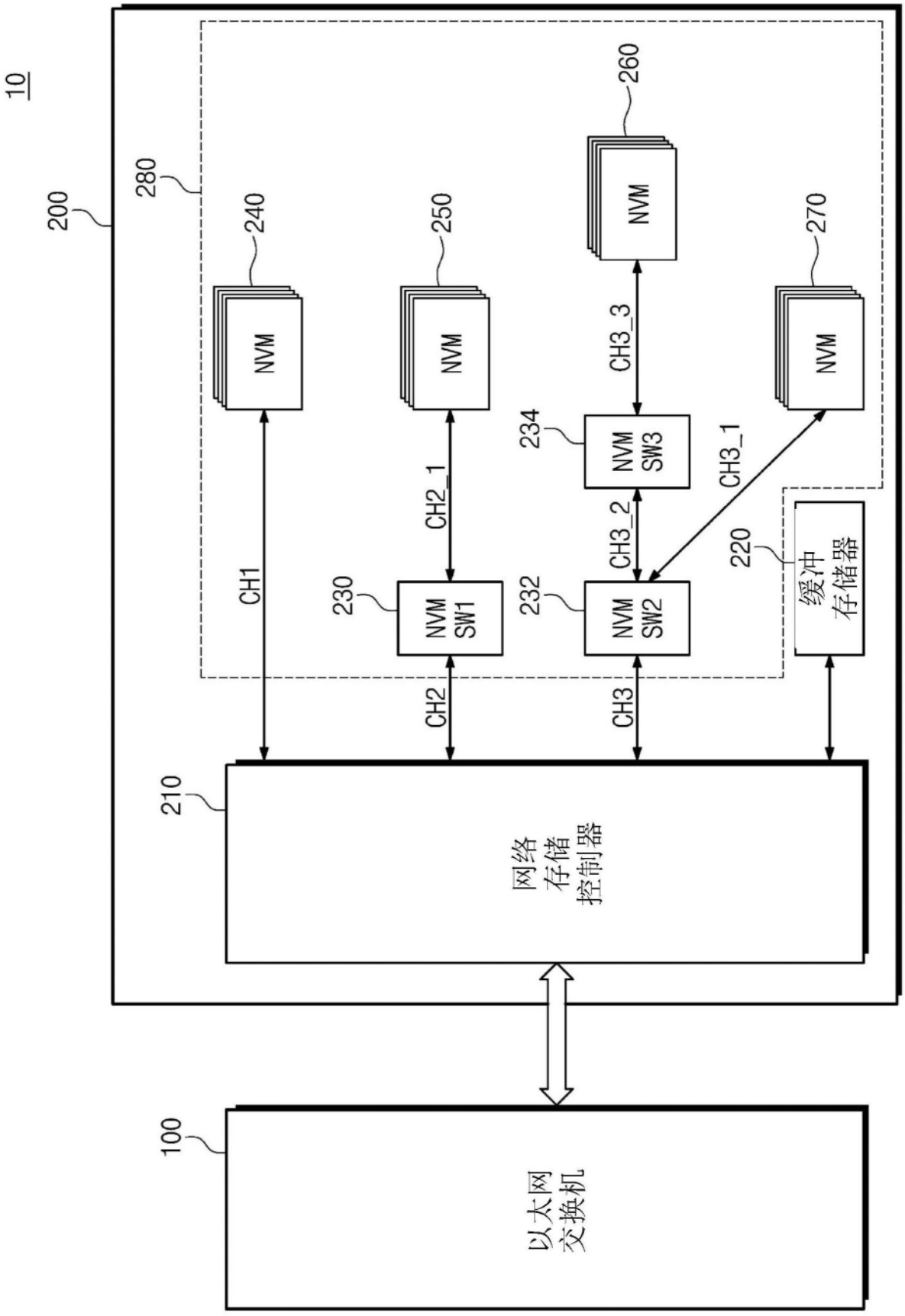

10
200
280
240
250
260
270
NVM
NVM
NVM
NVM
CH3_3
CH3_1
234
NVM SW3
CH2_1
CH3_2
CH1
230
NVM SW1
232
NVM SW2
220
缓冲存储器
CH2
CH3
210
网络存储控制器
100
以太网交换机

图1

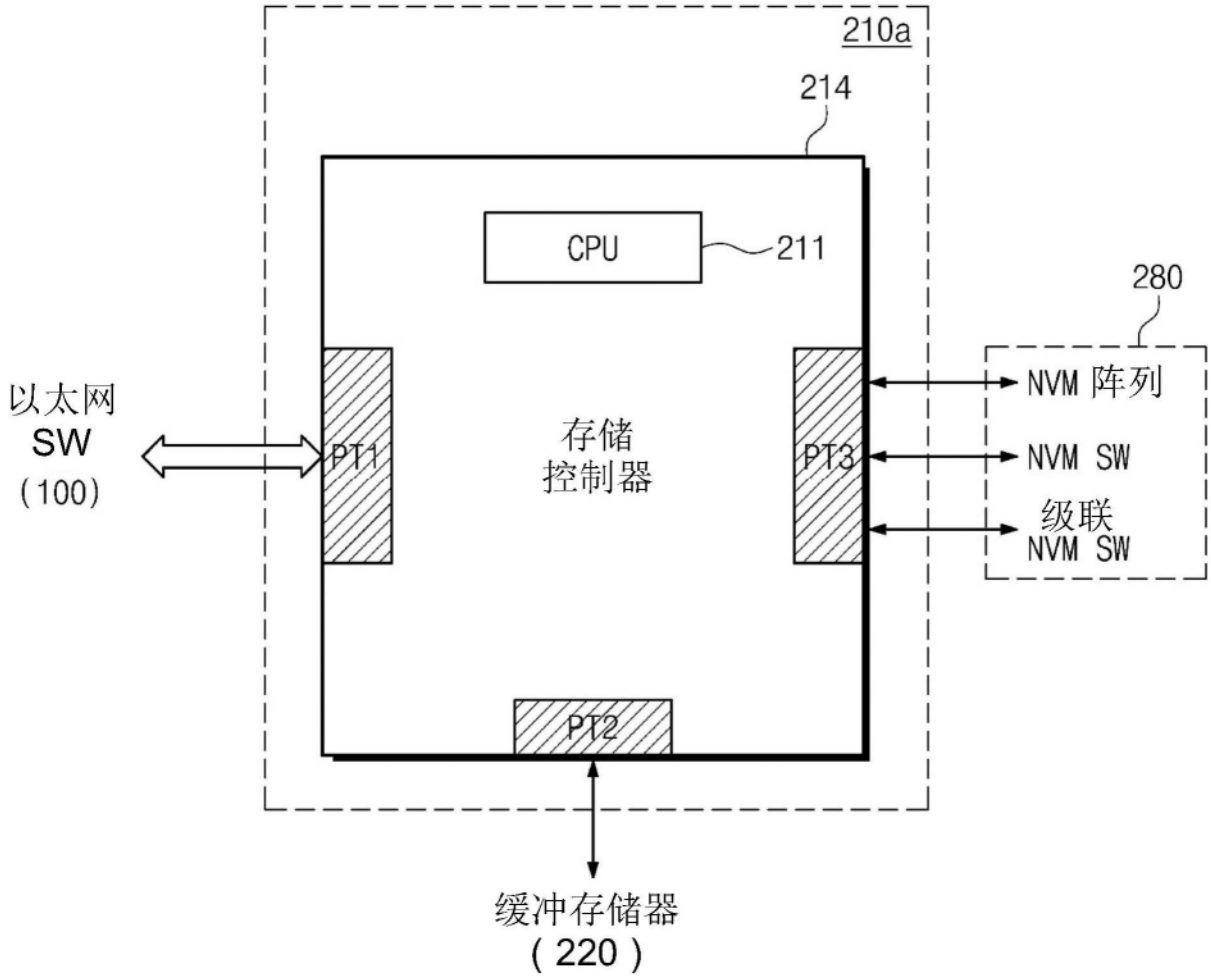

210a
214
CPU
211
280
以太网
SW
(100)
PT1
存储
控制器
PT3
NVM 阵列
NVM SW
级联
NVM SW
PT2
缓冲存储器
(220)

图2

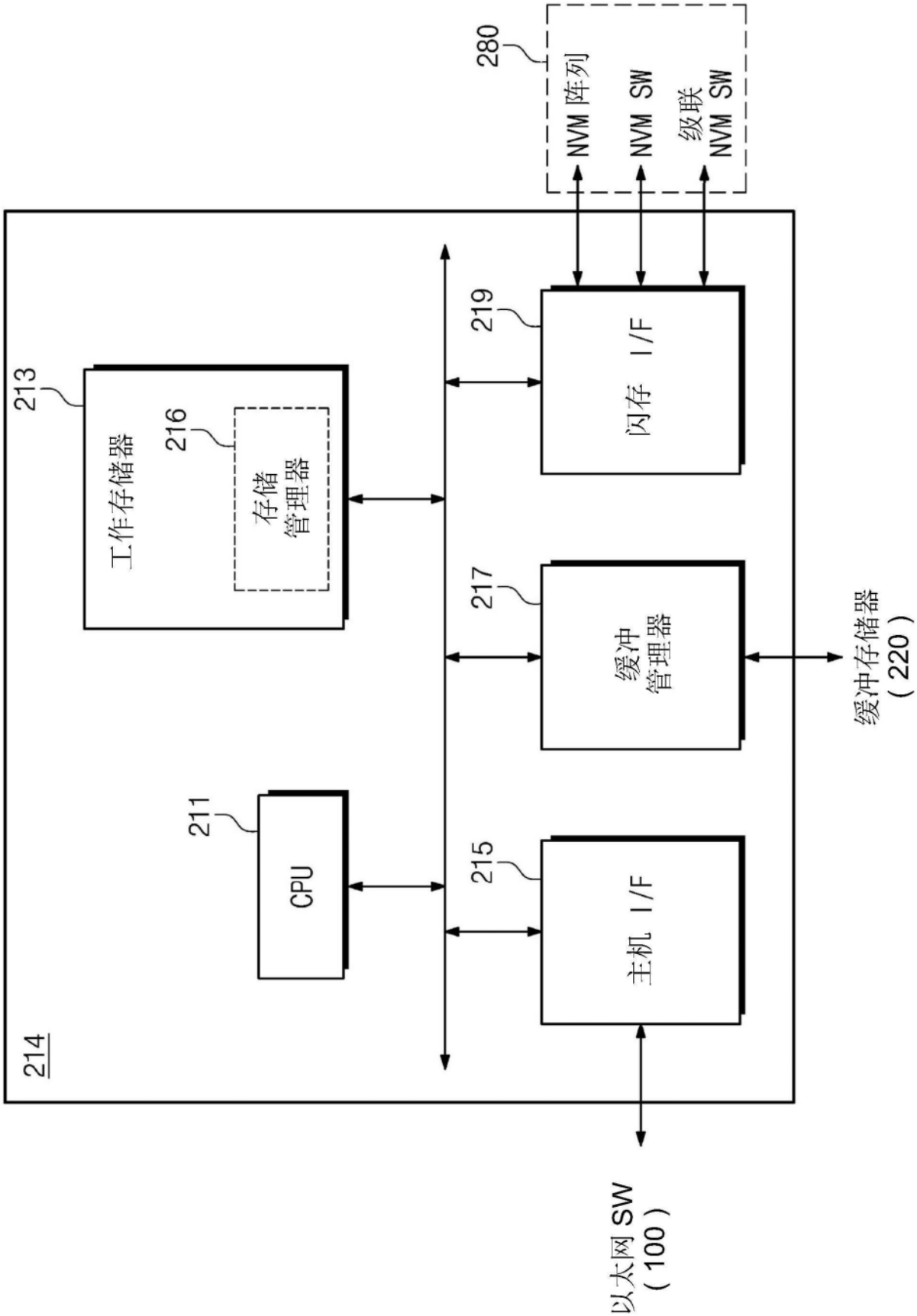

280
NVM 阵列
NVM SW
级联
NVM SW
219
闪存 I/F
213
工作存储器
216
存储
管理器
217
缓冲
管理器
缓冲存储器
(220)
211
CPU
215
主机 I/F
214
以太网 SW
(100)

图3

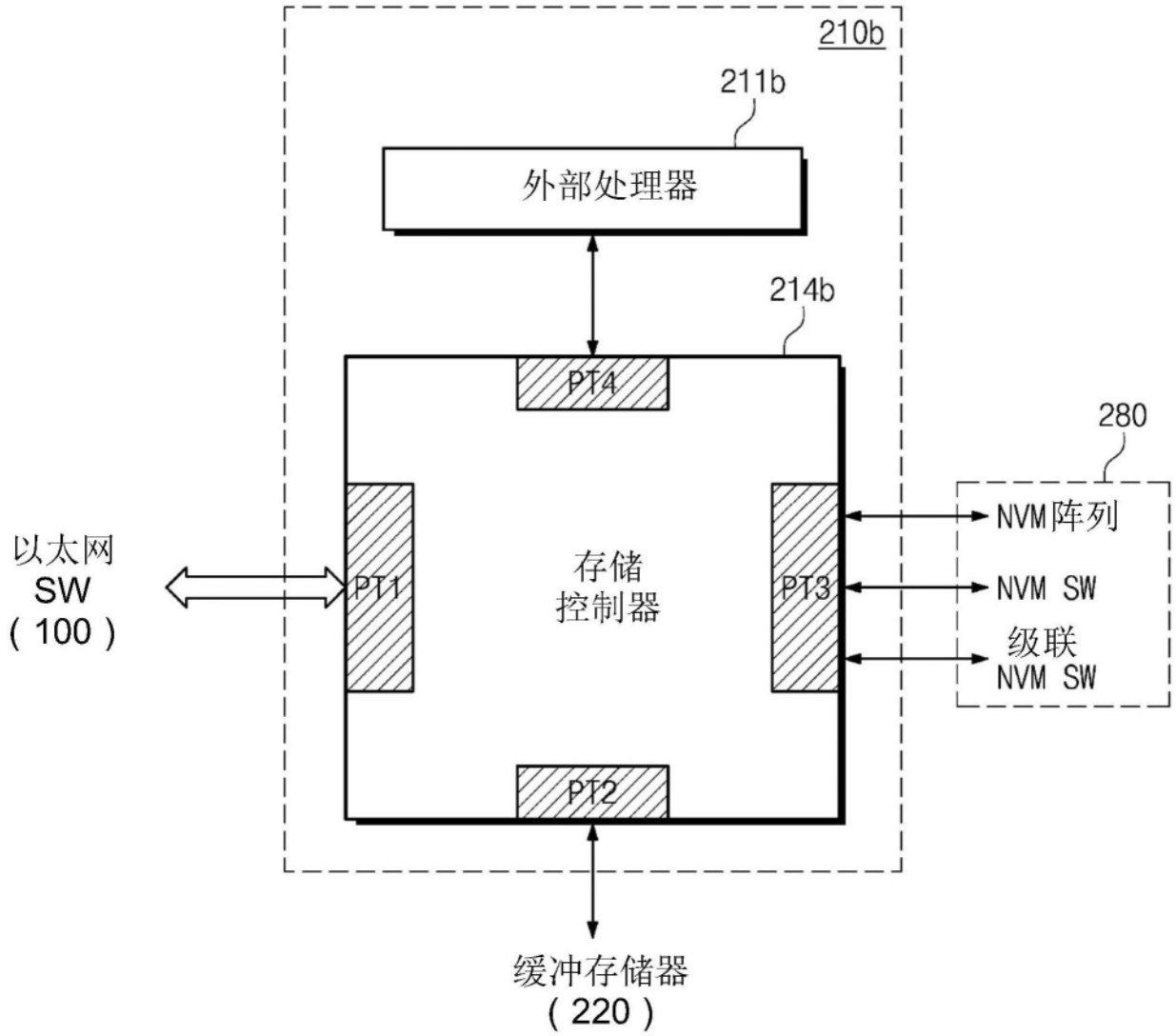

210b
211b
外部处理器
214b
PT4
280
以太网
SW
(100)
PT1
存储
控制器
PT3
NVM阵列
NVM SW
级联
NVM SW
PT2
缓冲存储器
(220)

图4

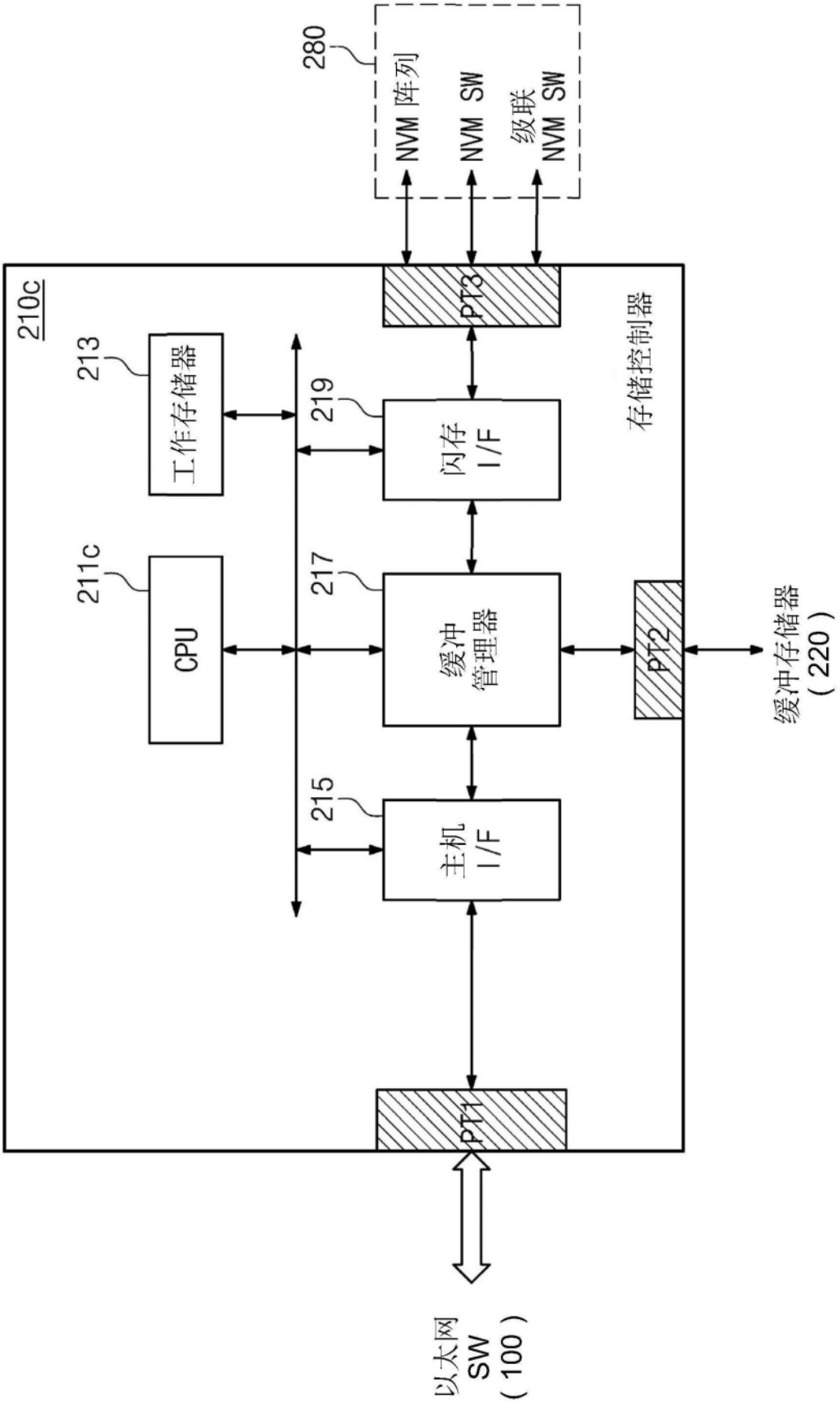

280
NVM 阵列
NVM SW
级联 NVM SW
210c
PT3
存储控制器
213
工作存储器
219
闪存 I/F
211c
CPU
217
缓冲管理器
PT2
缓冲存储器 (220)
215
主机 I/F
PT1
以太网 SW (100)

图5

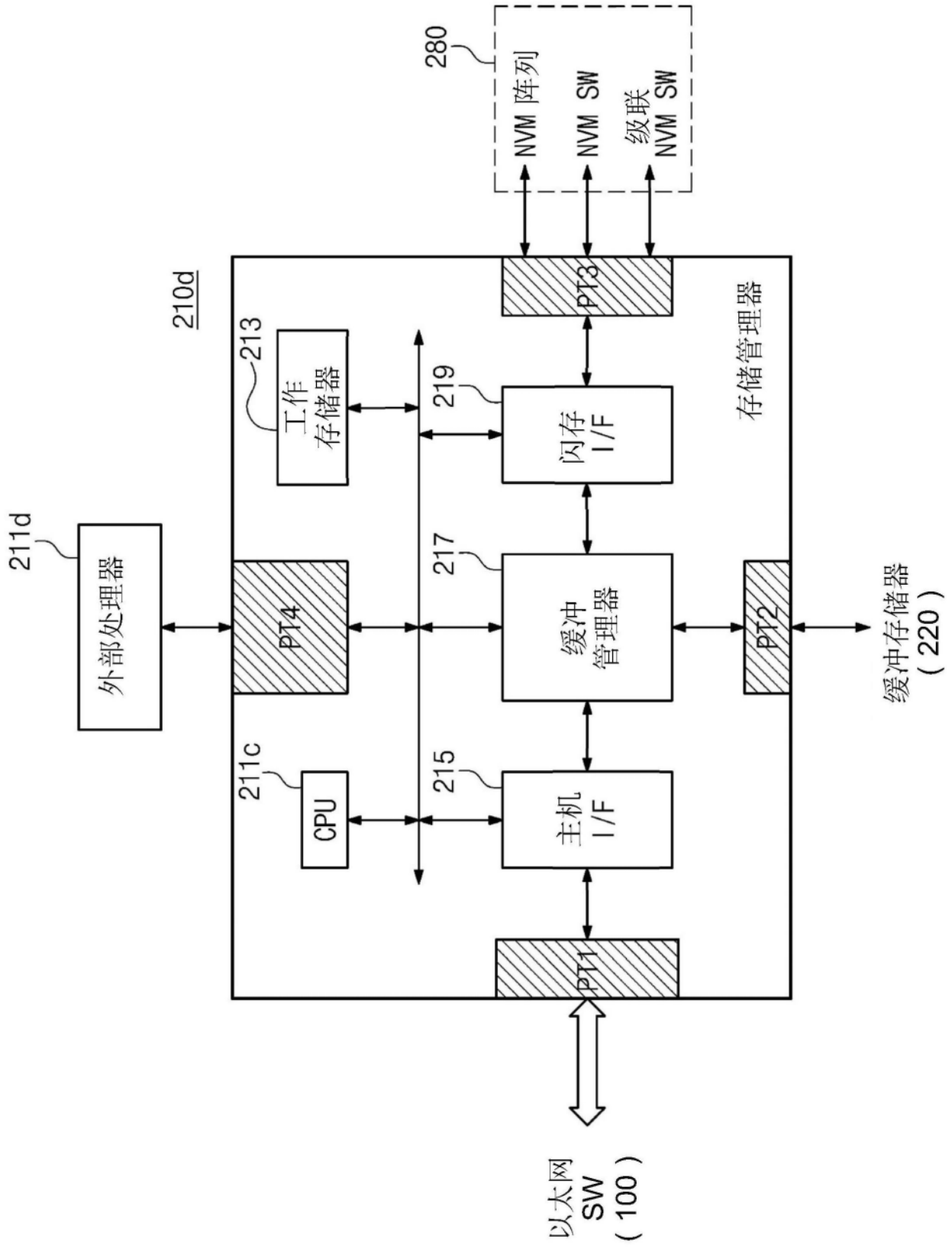

280
NVM 阵列
NVM SW
级联
NVM SW
210d
213
工作
存储器
219
闪存
I/F
PT3
存储管理器
211d
外部处理器
PT4
217
缓冲
管理器
PT2
缓冲存储器
(220)
211c
CPU
215
主机
I/F
PT1
以太网
SW
(100)

图6

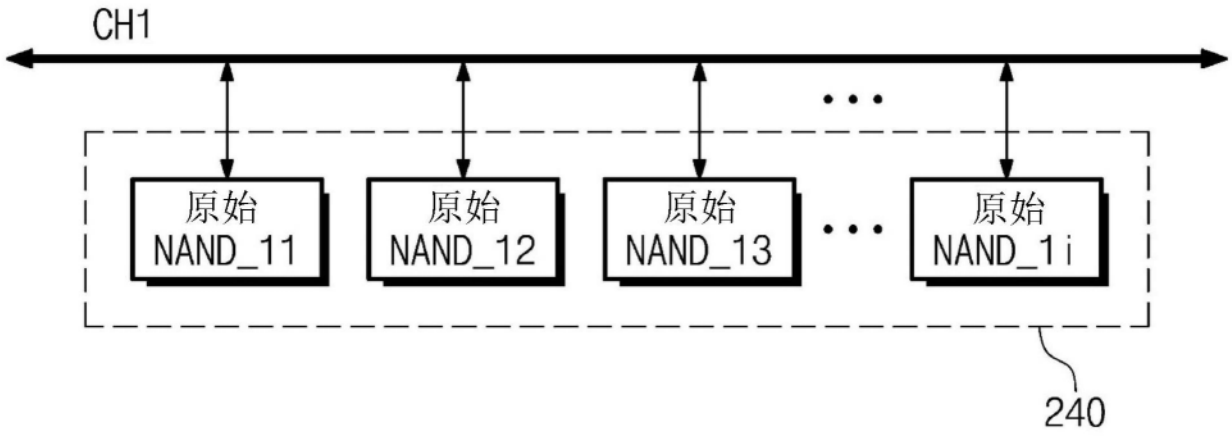


图7

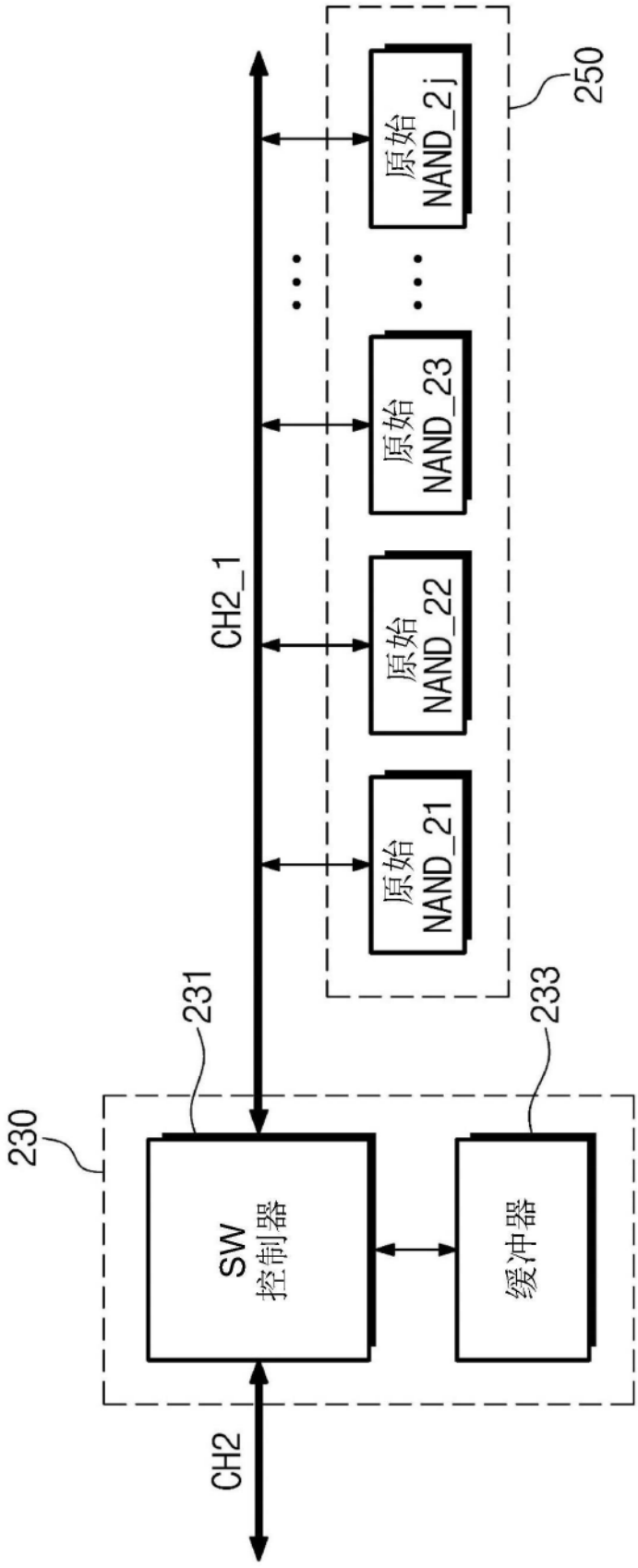

250
原始 NAND_2j
原始 NAND_23
原始 NAND_22
原始 NAND_21
CH2_1
231
233
230
SW 控制器
缓冲器
CH2

图8

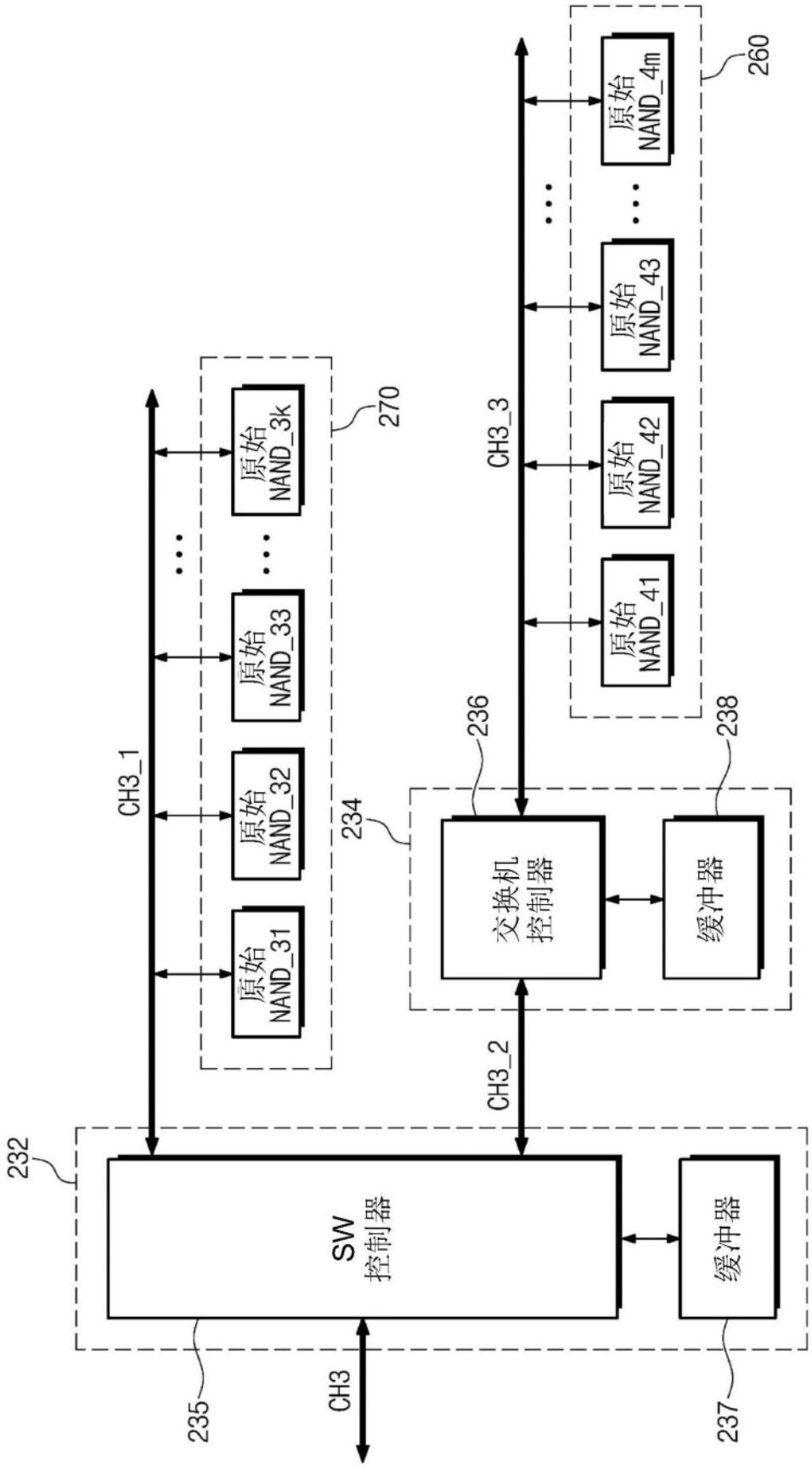

232
SW
控制器
缓冲器
235
CH3
237
CH3_2
234
交换机
控制器
236
缓冲器
238
CH3_1
CH3_3
270
原始
NAND_31
原始
NAND_32
原始
NAND_33
原始
NAND_3k
260
原始
NAND_41
原始
NAND_42
原始
NAND_43
原始
NAND_4m

图9

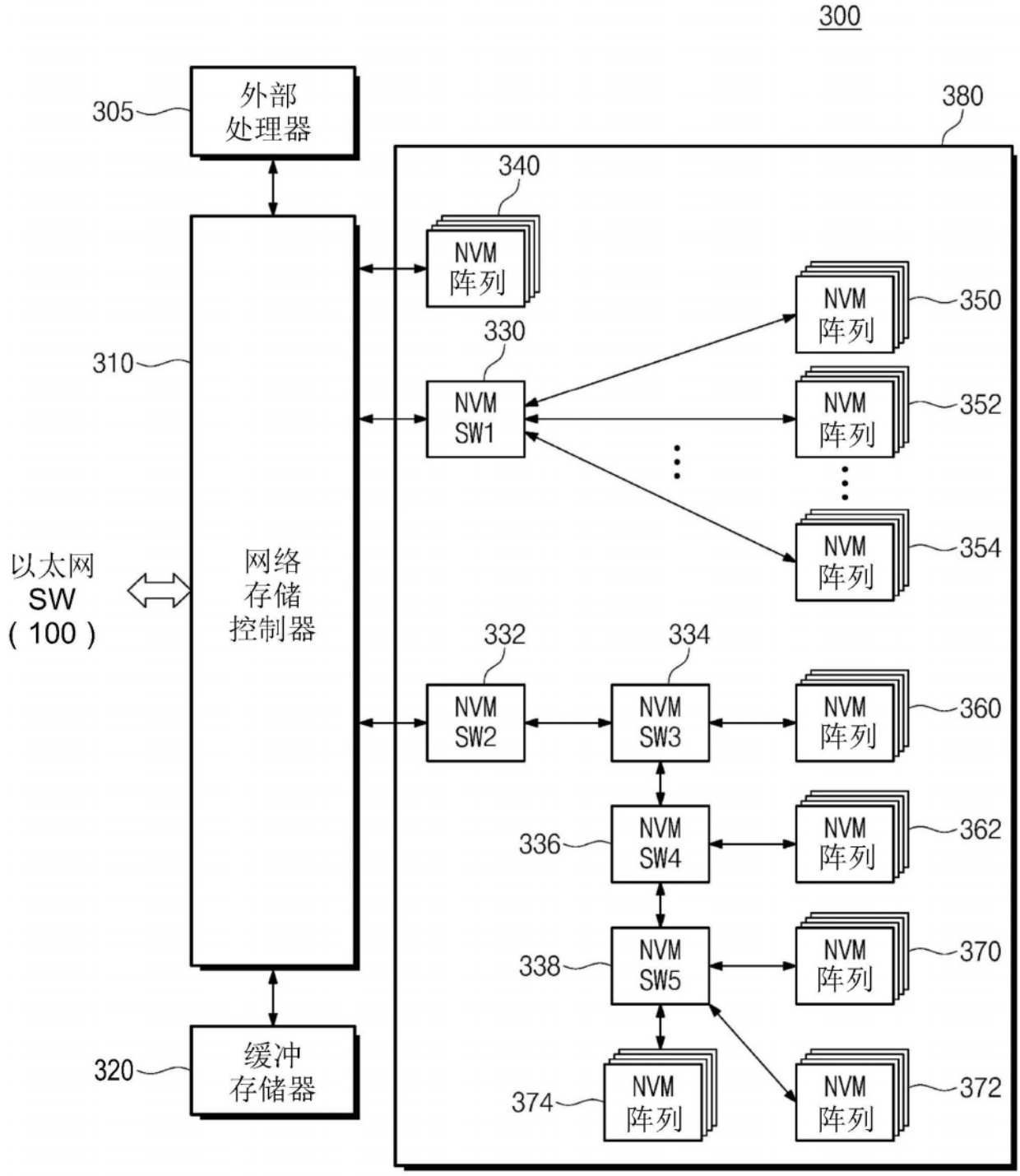

300
305
外部
处理器
380
340
NVM
阵列
330
310
NVM
SW1
350
352
354
以太网
SW
(100)
网络
存储
控制器
332
334
NVM
SW2
NVM
SW3
360
336
NVM
SW4
362
338
NVM
SW5
370
374
372
320
缓冲
存储器

图10

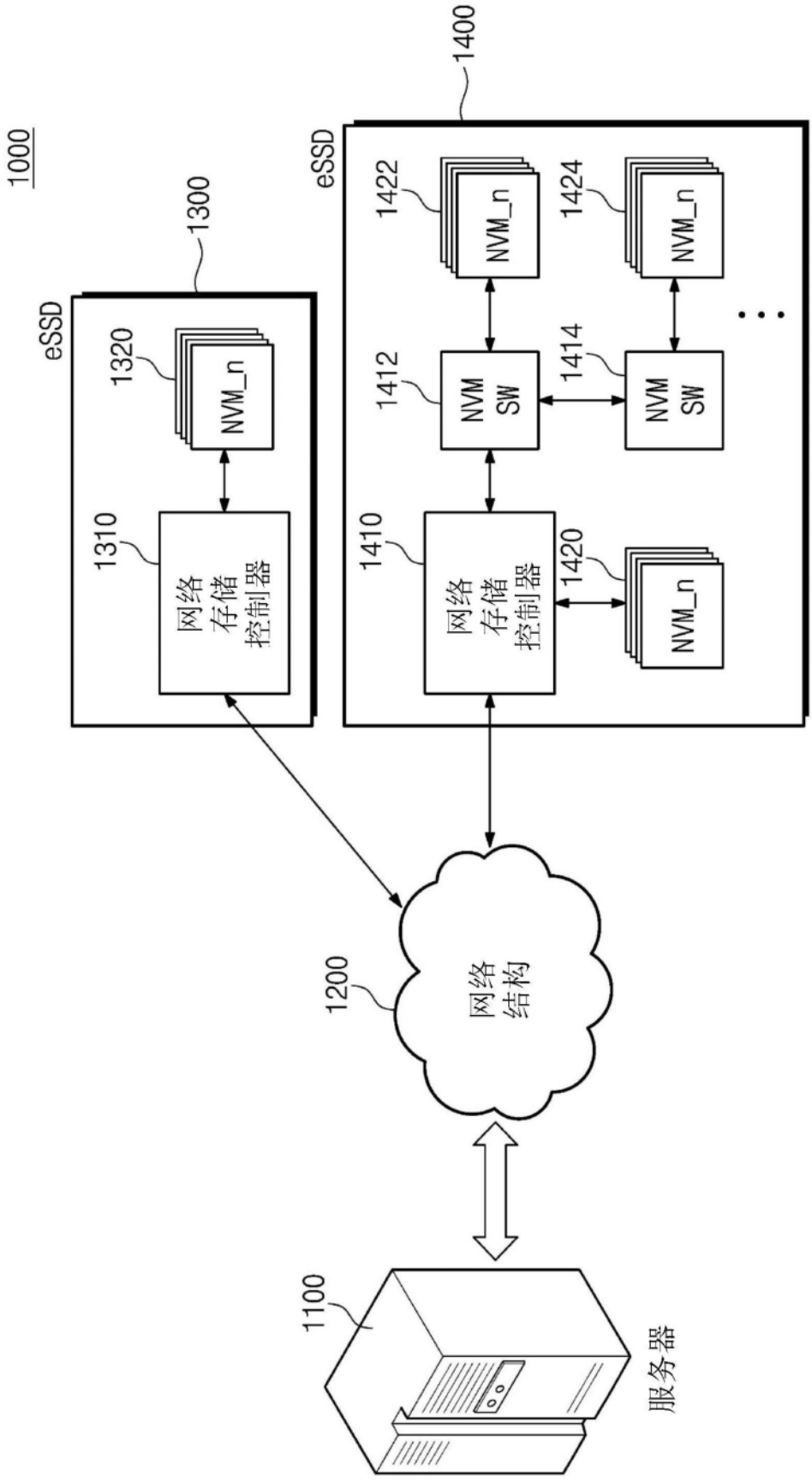

1000
1400
eSSD
1422
NVM_n
1424
NVM_n
1300
eSSD
1320
NVM_n
1412
NVM SW
1414
NVM SW
1310
网络存储控制器
1410
网络存储控制器
1420
NVM_n
1200
网络结构
1100
服务器

图11