



(12) 发明专利

(10) 授权公告号 CN 101542449 B

(45) 授权公告日 2012.06.13

(21) 申请号 200780043925.2

G06F 1/26(2006.01)

(22) 申请日 2007.10.09

G06F 11/30(2006.01)

(30) 优先权数据

11/604,631 2006.11.27 US

(85) PCT申请进入国家阶段日

2009.05.26

(86) PCT申请的申请数据

PCT/US2007/021692 2007.10.09

(87) PCT申请的公布数据

W02008/066605 EN 2008.06.05

(73) 专利权人 LSI 公司

地址 美国加利福尼亚州

(72) 发明人 穆罕默德·埃尔-巴塔勒

查尔斯·尼科尔斯 约翰·舍曼

基思·霍尔特 詹森·斯图尔萨茨

(74) 专利代理机构 北京康信知识产权代理有限

责任公司 11240

代理人 余刚 吴孟秋

(51) Int. Cl.

G06F 12/00(2006.01)

(56) 对比文件

US 2006/0265624 A1, 2006.11.23, 全文.

US 2006/0107005 A1, 2006.05.18, 说明书第 5 页第 [0061] - [0062] 段, 第 [0070] - [0071] 段, 第 7 页第 [0087] - [0088] 段、附图 5, 7.

US 2002/0087751 A1, 2002.07.04, 说明书第 4 页第 [0044] - [0046] 段, 第 5 页第 [0052] - [0053] 段, 第 6 页第 [0064] 段, 第 8 页第 [0123] 段、附图 3-5.

US 2002/0087751 A1, 2002.07.04, 说明书第 4 页第 [0044] - [0046] 段, 第 5 页第 [0052] - [0053] 段, 第 6 页第 [0064] 段, 第 8 页第 [0123] 段、附图 3-5.

审查员 陈冬冰

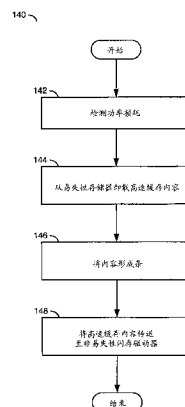
权利要求书 2 页 说明书 5 页 附图 4 页

(54) 发明名称

优化存储控制器高速缓存卸载电路性能和可靠性的系统

(57) 摘要

公开了一种用于对高速缓冲存储器进行卸载的方法 (140)。该方法大体包括以下步骤: (A) 响应于信号的认定对高速缓冲存储器的多条高速缓存线中的所有高速缓存线进行读取, 以对高速缓冲存储器进行卸载 (144), (B) 根据 RAID 配置通过对各高速缓存线进行分割来产生多个块 (146), 以及 (C) 将这些块写入 RAID 配置中的多个非易失性存储器中, 其中, 每个非易失性存储器均具有小于高速缓冲存储器的读取带宽的写入带宽 (148)。



1. 一种用于对高速缓冲存储器进行卸载的方法,包括以下步骤:

(A) 直接通过控制器在所述高速缓冲存储器和主存储器之间交换多条高速缓存线,其中,所述主存储器和所述高速缓冲存储器是易失性的;

(B) 通过所述控制器而不经所述高速缓冲存储器在处理器和所述主存储器之间交换数据;

(C) 响应于表示功率损耗的信号的确,将来自所述高速缓冲存储器的所有所述高速缓存线缓存到所述控制器中;

(D) 根据 RAID 配置,通过对缓存的所述高速缓存线进行分割来在所述控制器中产生多个块;以及

(E) 将来自所述控制器的所述块直接写入所述 RAID 配置中的多个非易失性存储器,其中,每个所述非易失性存储器均具有 (i) 小于所述高速缓冲存储器的读取带宽的写入带宽以及 (ii) 至所述控制器的用于传送所述块的不同的独立路径。

2. 根据权利要求 1 所述的方法,其中,所述块中的至少两个块被基本上同时地写入所述非易失性存储器。

3. 根据权利要求 1 所述的方法,还包括以下步骤:

通过将所述块分成条来产生多个条,其中,每个所述块成为多个条。

4. 根据权利要求 3 所述的方法,其中,所述步骤 (E) 包括以下子步骤:

将 (i) 所述条和 (ii) 相同等级的所述条的奇偶校验写入所述非易失性存储器之中。

5. 根据权利要求 4 所述的方法,其中,所述条中的至少两个条被基本上同时地写入所述非易失性存储器。

6. 根据权利要求 1 所述的方法,其中, (i) 所述信号的所述确表示流入功率电路的功率的损耗,以及 (ii) 所述功率电路被配置为在所述功率的损耗之后向所述控制器、所述高速缓冲存储器和所述非易失性存储器提供功率。

7. 根据权利要求 1 所述的方法,其中, (i) 每个所述非易失性存储器均具有第一存储容量, (ii) 所述高速缓冲存储器具有第二存储容量,以及 (iii) 全部的所述第一存储容量至少与所述第二存储容量一样大。

8. 根据权利要求 1 所述的方法,其中,所述非易失性存储器的所述写入带宽的全部至少与所述高速缓冲存储器的所述读取带宽一样大。

9. 根据权利要求 1 所述的方法,其中,以最接近所述高速缓冲存储器的所述读取带宽的第一带宽从所述高速缓冲存储器读取所述高速缓存线。

10. 根据权利要求 9 所述的方法,其中,以最接近所述非易失性存储器的所述写入带宽的第二带宽将所述块写入所述非易失性存储器。

11. 一种系统,包括:

高速缓冲存储器,具有读取带宽并配置为存储多条高速缓存线,所述高速缓冲存储器是易失性的;

多个非易失性存储器,其中每一个均具有小于所述读取带宽的写入带宽;

控制器,配置为 (i) 响应于表示功率损耗的信号的确对来自所述高速缓冲存储器的所有所述高速缓存线进行缓存;

(ii) 根据 RAID 配置通过对缓存的所述高速缓存线进行分割来产生多个块;以及 (iii)

在所述功率的损耗之后,将所述块直接写入所述 RAID 配置中的所述非易失性存储器,其中,每个所述非易失性存储器均具有至所述控制器的用于传送所述块的不同的独立路径;主存储器,配置为直接通过所述控制器与所述高速缓冲存储器交换所述高速缓存线,所述主存储器是易失性的;以及

处理器,配置为通过所述控制器而不经所述高速缓冲存储器与所述主存储器交换第一数据。

12. 根据权利要求 11 所述的系统,其中,所述块中的至少两个块被基本上同时地写入所述非易失性存储器。

13. 根据权利要求 11 所述的系统,其中所述处理器还被配置为通过所述控制器与所述高速缓冲存储器交换第二数据。

14. 根据权利要求 11 所述的系统,还包括功率电路,所述功率电路配置为认定所述信号。

15. 根据权利要求 14 所述的系统,其中,所述信号的所述认定表示流入到所述功率电路的功率的损耗,以及所述功率电路还被配置为在所述功率的损耗之后向所述控制器、所述高速缓冲存储器和所述非易失性存储器提供功率。

16. 根据权利要求 14 所述的系统,其中,所述功率电路包括超电容器和超级电容器中的至少一个,以在源功率释放之后传送功率。

17. 根据权利要求 11 所述的系统,还包括至少四个插口,每个所述插口均配置为连接至一个所述非易失性存储器。

18. 根据权利要求 17 所述的系统,其中,至少一个所述插口在所述系统的至少一种配置中为空。

19. 根据权利要求 17 所述的系统,其中,所述 RAID 配置包括以下各项之一:(i) RAID 0 配置,(ii) RAID 1 配置,以及 (iii) RAID 5 配置。

20. 一种系统,包括:

用于易失性存储的装置,具有读取带宽且配置为存储多条高速缓存线;

用于非易失性存储的多个装置,其中每一个均具有小于所述读取带宽的写入带宽;

用于控制包括以下各项的装置,(i) 响应于表示功率损耗的信号的认识对来自所述用于易失性存储的装置的所有所述高速缓存线进行缓存;(ii) 根据 RAID 配置通过对缓存的所述高速缓存线进行分割来产生多个块;以及 (iii) 在所述功率的损耗之后,将所述块直接写入所述 RAID 配置中的所述用于非易失性存储的装置,其中,所述每个所述用于非易失性存储的装置具有至所述用于控制的装置用来传送所述块的不同的独立路径;

用于主存储器的装置,配置为直接通过所述用于控制的装置与所述用于易失性存储的装置交换所述高速缓存线;以及

用于处理的装置,配置为通过所述用于控制的装置而不经所述用于易失性存储的装置与所述用户主存储器的装置交换数据。

优化存储控制器高速缓存卸载电路性能和可靠性的系统

技术领域

[0001] 本发明大体涉及存储控制器,且更具体地,涉及用于优化存储控制器高速缓存卸载电路的性能和可靠性的方法和/或设备。

背景技术

[0002] 一旦 AC 电源有功率损耗,传统存储控制器就被迫使用来自储量有限的电池备用单元的功率来尽可能快且可靠地将高速缓冲内容从高速缓冲存储器卸载到本地持久性存储装置。该持久性存储装置 (i) 通常是本地的以避免依赖于要被加电的远程装置,以及 (ii) 利用非常少量的功率来避免大的电池。非常少的功率导致持久性存储装置具有有限的访问带宽。大的电池非常昂贵且具有随着时间而降低的可靠性。

发明内容

[0003] 本发明涉及用于对高速缓冲存储器进行卸载的方法。该方法大体包括以下步骤:(A) 响应于信号的认定(assert)对高速缓冲存储器的多条高速缓存线中的所有高速缓存线进行读取,以对高速缓冲存储器进行卸载,(B) 根据 RAID 配置通过对各高速缓存线进行分割来产生多个块,以及(C) 将这些块写入 RAID 配置中的多个非易失性存储器中,其中,每个非易失性存储器都具有小于高速缓冲存储器的读取带宽的写入带宽。

[0004] 本发明的目的、特征和优点包括:提供用于优化存储器控制器高速缓存卸载电路的性能和可靠性的方法和/或设备,该方法和/或设备可以 (i) 在 RAID 配置中布置多个非易失性存储器,(ii) 基本上同时地对两个或更多的非易失性存储器进行写入,(iii) 使得能够通过添加更多的存储器电路来进行非易失性存储器的容量扩展,(iv) 相比于传统方法允许更小的电池备用单元大小,和/或(v) 允许使用超电容器技术作为传统电池单元的替代。

附图说明

[0005] 本发明的这些和其他目的、特征和优点将从下面的详细描述和所附权利要求以及附图中变得显而易见,附图中:

[0006] 图 1 是根据本发明优选实施例的系统的框图;

[0007] 图 2 是非易失性存储器电路的示例性实现的示图;

[0008] 图 3 是用于对高速缓冲存储器进行卸载的示例性方法的流程图;

[0009] 图 4 是示例性 RAID 0 配置的示图;

[0010] 图 5 是示例性 RAID 1 配置的示图;以及

[0011] 图 6 是示例性 RAID 5 配置的示图。

具体实施方式

[0012] 本发明大体实现了并行使用多个非易失性驱动器的快速高速缓存卸载架构。这些

非易失性驱动器可以被安排在 RAID 配置中,诸如 RAID 0 配置、RAID 1 配置、或 RAID 5 配置。还可以实现其他的 RAID 配置以满足特定应用的标准。相比于传统技术,多个 RAID 配置的并行写入特性通常考虑了高速缓存卸载接口上更高的性能和更高的可靠性。

[0013] 参照图 1,示出了根据本发明一个优选实施例的系统 100 的框图。该系统(或设备)100 可以实现为一个基于高速缓存的处理系统。该系统 100 通常包括:电路(或模块)102、电路(或模块)104、电路(或模块)106、电路(或模块)108、电路(或模块)110、和电路(或模块)112。可以通过电路 110 接收信号(例如,PWR)。可以由电路 110 产生一个信号(例如,OFFLOAD)并将其提供给电路 104。接口 114 可以使得电路 102 和电路 104 能够彼此进行通信。电路 104 可以通过接口 116 与电路 106 进行通信。接口 118 可以允许电路 104 与电路 108 进行通信。电路 104 可以通过接口 120 与电路 112 进行通信。

[0014] 电路 102 可以实现为处理器电路。电路 102 可以用于通过执行软件程序来执行各种功能。电路 102 可以通过电路 104 从电路 106、108 和 112 读取软件程序的指令和/或数据,以及将软件程序的指令和/或数据写入电路 106、108 和 112。

[0015] 电路 104 可以实现为存储器控制器电路。电路 104 可以用于控制电路 106、电路 108 和电路 112。电路 104 可以通过处理器接口 114 与电路 102 交换软件程序的指令和数据。这些数据和指令可以(i)通过高速缓存接口 116 在电路 104 和电路 106 之间、(ii)通过闪存接口 118 在电路 104 和电路 108 之间、以及通过存储器接口 120 在电路 104 和电路 112 之间进行交换。电路 104 可以进一步用于响应于信号 OFFLOAD 的认定状态(例如,逻辑低)来通过接口 118 将存储在电路 106 中的所有信息(例如,数据和指令)卸载到电路 108 中(见箭头 128)。

[0016] 电路 106 可以实现为易失性存储器。具体地,电路 106 可以被实现为易失性高速缓存存储器。电路 106 通常可用于缓冲由在电路 102 中执行的软件产生和使用的数据和指令。存储在电路 106 中的信息可以被安排为高速缓存线 124a-124n。高速缓存线 124a-124n 中的每一个都可以基于高速缓存命中和高速缓存未命中来与电路 112 进行交换。这些高速缓存线可以被以第一读取带宽从电路 106 被读取,而以第一写入带宽被写入。

[0017] 电路 108 可以实现为非易失性存储器 126a-126d 的阵列。存储器(或部件)126a-126d 可以被安排在 RAID(独立磁盘的冗余阵列)配置中。在一些实施例中,电路 108 的每个存储器“磁盘”126a-126d 都可以实现为闪存。还可以实现其他的非易失性存储器技术来满足特定应用的标准。信息可以以第二写入带宽被写入到存储器 126a-126d 中的每一个,以及以第二读取带宽被读取。

[0018] 电路 110 可以实现为备用功率单元。电路 110 可以用于将在信号 PWR 中接收到的电功率存储、变换、调整和/或滤波成适于电路 102、104、106、108 和 112 使用的一个或多个功率网络。电路 110 还可用于在有限的时间内提供电功率,以适于使电路 104、106 和 108 至少工作足够的时间以将信息从电路 106 卸载到电路 108 中。此外,电路 110 可以监控通过信号 PWR 流入的功率状况,以及响应于信号 PWR 中功率的严重下降和/或完全损失来认定信号 OFFLINE。在一些实施例中,电路 110 可以实现为一个或多个电池。在至少一个实施例中,电路 110 可以实现为一个或多个超电容器或超级电容器(super-capacitor 或 ultra-capacitor)。

[0019] 电路 112 可以实现为主存储器电路。具体地,电路 112 可以实现为易失性随机存

取存储器。电路 112 可以用于存储在电路 102 上执行的软件的指令和数据。如由电路 104 所确定的,电路 112 可以为电路 106 提供高速缓存线以及接收来自电路 106 的高速缓存线。

[0020] 参照图 2,示出了电路 108 的示例性实现的示图。除了存储器部件 126a-126d 以外,电路 108 还可以包括多个插口 130a-130d。这些插口(或端口)130a-130d 中的每一个都通常被安排为连接至单个的存储器 126a-126d。该连接可以包括物理连接、电功率连接和通信连接。在系统 100 的至少一个配置中,插口 130a-130d 可以由单个存储器部件(例如,126a)总装(populate)。在系统 100 的其他配置中,两个或更多的存储器 126a-126d 可以被安装在插口 130a-130d 中。

[0021] 参照图 3,示出了用于对电路 106 进行卸载的示例性方法 140 的流程图。该方法 140 大体实现了将数据从电路 106 移至电路 108 的快速卸载方法。方法 140 大体包括步骤(或块)142、步骤(或块)144、可选步骤(或块)146 和步骤(或块)148。

[0022] 该方法 140 可以利用信号 OFFLOAD 的认定来触发。诸如来自电路 102 的命令的其他触发也可以启动该方法 140。在步骤 142 中,一旦检测到信号 PWR 中的电功率损失,则电路 110 可认定信号 OFFLOAD。信号 OFFLOAD 的认定可以由电路 104 感测。作为响应,在步骤 144 中,电路 104 可以读取(卸载)来自电路 106 的高速缓存线 124a-124n。信息从电路 106 到电路 104 的传送速度可以由电路 106 的读取带宽控制。

[0023] 根据在电路 108 中实现的具体 RAID 配置,在步骤 146 中,电路 104 可以/不可以对高速缓存线 124a-124n 中的信息分条(stripe)。之后在步骤 148 中,信息块/信息条以及误差校正信息(如果有的话)可以由电路 104 写入到存储器 126a-126d 中。这些块/条从电路 104 到电路 108 的传送速度可以由存储器 126a-126d 的写入带宽确定。

[0024] 由于信息可以被基本同时地沿多条并行路径从电路 104 写入到存储器 126a-126d,因此到存储器 126a-126d 的组合写入带宽可以大于(快于)从电路 106 的读取带宽。相比于传统技术,更高的组合写入带宽通常减少执行传送所耗费的时间。系统 100 的架构可以以低成本利用可移动非易失性存储器部件 126a-126d。示例性的存储器部件 126a-126d 可以包括但不限于安全数字(SD)闪存卡和 USB 闪存驱动器。

[0025] 近年来,用于电路 106 的用户指定的高速缓存大小已经变得很大。因此,由于较慢的写入时间和更小的容量,低成本的非易失性存储器选择通常是不可用的。本发明通常使用多个非易失性存储器,从而可以使用 RAID 技术建立比单个普通非易失性存储器元件更大且更快的虚拟非易失性存储器(电路 108),来提高非易失性存储器的容量和速度。

[0026] 通过使用多个存储器 126a-126d,电路 104 和电路 108 可以与由用户定制的高速缓存的数量成比例的缩放。例如,电路 104 可以支持电路 106 中的 8 十亿字节(GB)、16 十亿字节(GB)和 32 十亿字节(GB)的高速缓存大小选项。电路 104 可以被配置为控制电路 108 中的多个(例如,四个)存储器部件 126a-126d,这些部件中的每一个都为 8GB 大小。如此,8GB 的高速缓存系统 100 可以利用单个 8GB 的存储器(例如,126a)来构建。16GB 的高速缓存系统 100 可以利用两个 8GB 的存储器(例如,126a 和 126b)来构建。32GB 的高速缓存系统可以利用四个 8GB 的存储器(例如,126a-126d)来构建。

[0027] 考虑了存储器 126a-126d 中的每一个都具有 20 兆字节每秒(MB/sec)的示例性写速度。8GB 高速缓存系统 100 可以使用约 $8GB/(20MB/sec) = 400$ 秒将 8GB 易失性电路 106 卸载至 8GB 的非易失性电路 108。对于 16GB 的高速缓存系统 100,由于使用了 RAID 技术来

配置两个存储器（例如，126a 和 126b），到电路 108 的写入带宽通常被加倍。用于将信息从 16GB 电路 106 移出的全部卸载时间可以为 $16GB / (2 \times 20MB/sec) = 400$ 秒。32GB 的高速缓存系统 100 可以使用四个存储器元件 126a-126d，来提供 $4 \times 20MB/sec = 80MB/sec$ 的有效带宽。较大的写入带宽可以允许 $32 GB / (4 \times 20MB/sec) = 400$ 秒的高速缓存卸载时间。在所有的三个实例中，高速缓存卸载时间都可以保持在约 400 秒。较大数量的存储器部件 126a-126d 可以用于减少卸载时间，允许更大的高速缓存大小和 / 或实现其他的 RAID 配置。

[0028] 参照图 4，示出了示例性 RAID 0 配置的示图。RAID 0 配置可以实现由存储器部件 126a-126d 构成的条阵列。电路 104 可以将从电路 106 读取的高速缓存线 124a-124n 分组成多个块（例如，A-H）。各块 A-H 中的每一个都可以被写至单个存储器 126a-126d，其中多个块被基本同时地沿并行路径 150a-150d 写入。例如，电路 104 可以并行或以交错起始顺序将块 A 写至存储器 126a，将块 B 写至存储器 126b，将块 C 写至存储器 126c，将块 D 写至存储器 126d。在交错起始序列中，电路 104 可以开始对块 A 进行写入，同时仍然从高速缓存线 124a-124n 来组装块 B。一旦块 B 准备好，电路 104 就可以开始对块 B 进行写入，继续块 A 的写入以及开始组装块 C。RAID 0 配置通常以存储器 126a-126d 中的至少两个实现。

[0029] 参照图 5，示出了 RAID 1 配置的示图。RAID 1 配置通常使用多个（例如，8 个）存储器 126a-126h 来实现镜像对的双工。电路 104 可以将从电路 106 读取的高速缓存线 124a-124n 分组成多个块 A-H。各块 A-H 中的每一个都可以被写至存储器 126a-126h 中的两个，其中多个块被基本同时地沿并行路径 150a-150h 写入。例如，块 A 可以被写至存储器 126a 和 126b 两者，块 B 可以被写至存储器 126c 和 126d 两者，等等。RAID 1 配置通常提供所存信息的容错。对于每个存储器对，即使这些存储器部件之一失效，写入到该对中的块也可以被恢复。RAID 1 配置以存储器 126a-126h 中的至少四个实现。

[0030] 参照图 6，示出了示例性 RAID 5 配置的框图。RAID 5 配置可以实现具有分布式奇偶校验（parity）的数据条。如前所述，电路 104 可以响应于信号 OFFLOAD 的认定读取来自电路 106 的高速缓存线 124a-124n。读取的信息可以被组装成块 A-H。然后块 A-H 中的每一个都可以被形成条。例如，块 A 可以变成条 A0、A1 和 A2，块 B 可以变成条 B0、B1 和 B3，块 C 可以变成条 C0、C2 和 C3，块 D 可以变成条 D1、D2 和 D3 等等。给定块的各条可以被按顺序写至单个存储器 126a-126d。

[0031] 可以由电路 104 以相同的等级对所有的条计算奇偶校验条，然后将其写入到单个存储器 126a-126d 中。例如，0 级奇偶校验（例如，0 PARITY）可以由条 A0、条 B0 和条 C0 产生，并被写入到存储器 126d 中。第一级奇偶校验（例如，1 PARITY）可以用条 A1、条 B1 和条 D1 计算，并被写入到存储器 126c 中。在写入每个新的级时都可以继续奇偶校验计算。RAID 5 配置通常提供在单个存储器部件 126a-126d 失效的情况下恢复所存信息的能力。分布式奇偶校验的使用可以允许存储器 126a-126d 的有效使用。RAID 5 配置可以以存储器 126a-126d 中的三个或更多个来实现。可以在电路 108 中实现其他的 RAID 配置来满足特定应用的标准。

[0032] 由图 1 和图 3 的示图执行的功能可以使用根据本说明书的教导编程的传统的通用数字计算机实现，这对本领域技术人员将是显而易见的。有经验的程序员基于本公开的教导可以容易地制备适当的软件代码，这对于本领域技术人员也将是显而易见的。

[0033] 本发明还可以通过制备 ASIC、FPGA，或通过传统部件电路的适当网络进行互联

来实现,如在本文中所述,其改进对于本领域技术人员也是显而易见的。

[0034] 因此,本发明还可以包括计算机产品,该计算机产品可以是存储介质,该存储介质包括可以被用于对计算机进行编程以执行根据本发明的过程的指令。存储介质可以包括但不限于任何类型的磁盘,包括软磁盘、光盘、CD-ROM、磁光盘、ROM、RAM、EPROM、EEPROM、闪存、磁或光卡、或适于存储电子指令的任何类型的介质。如在本文中使用的,术语“同时”是指描述共享一些公共时间段的一些事件,但是该术语不意味着限制于以同一时间点开始、以同一时间点结束、或具有相同的持续时间的事件。

[0035] 尽管已经参照优选实施例具体示出并描述了本发明,但是应该理解,在不背离本发明范围的条件下,本领域技术人员可以进行形式和细节上的各种改变。

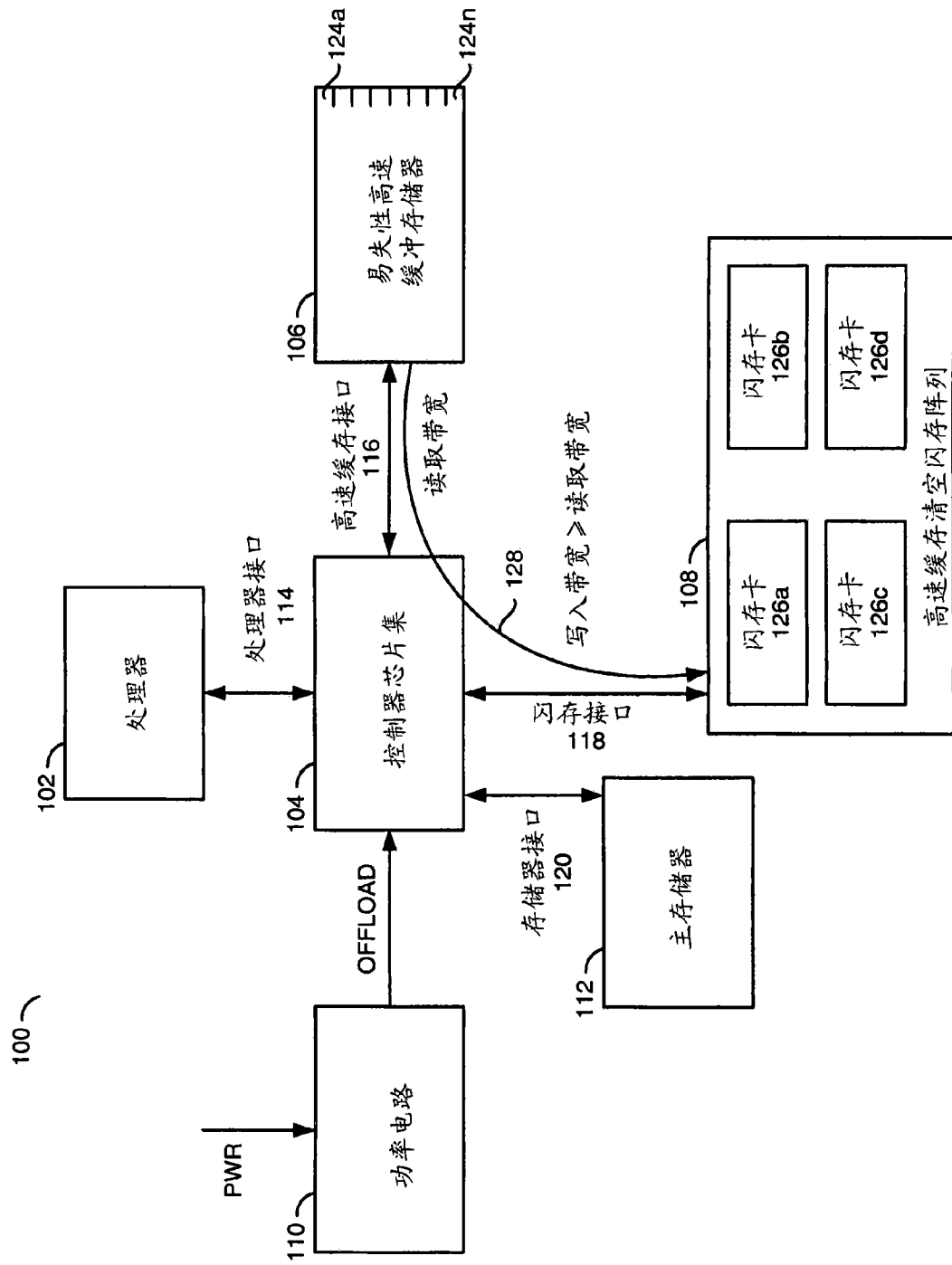


图 1

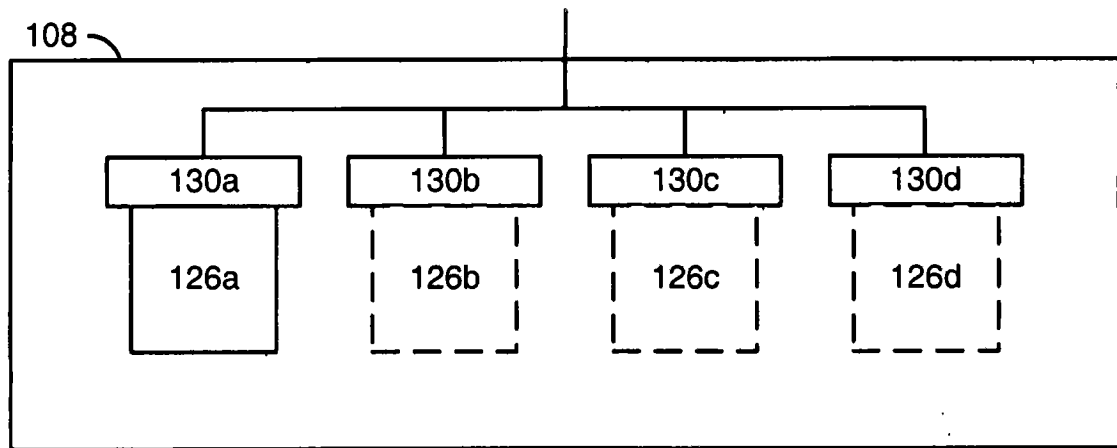


图 2

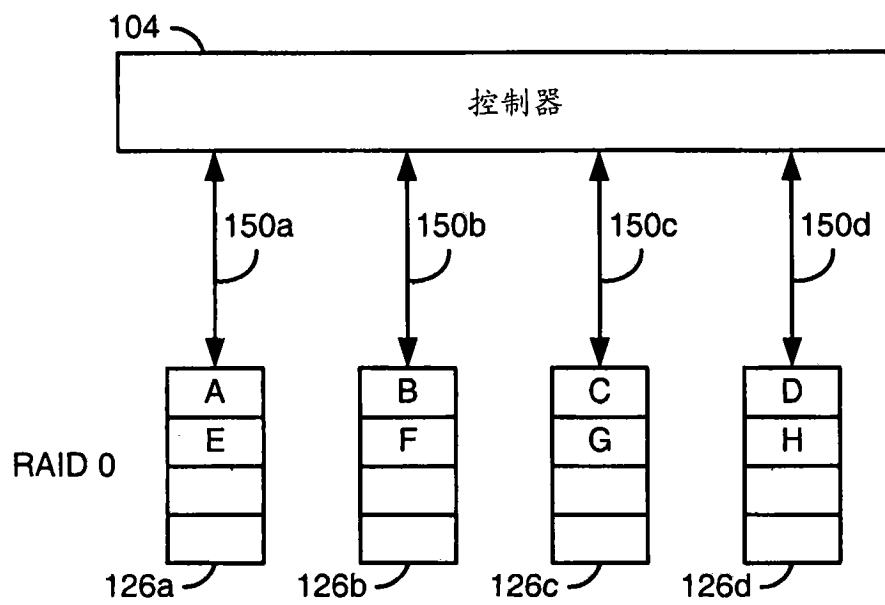


图 4

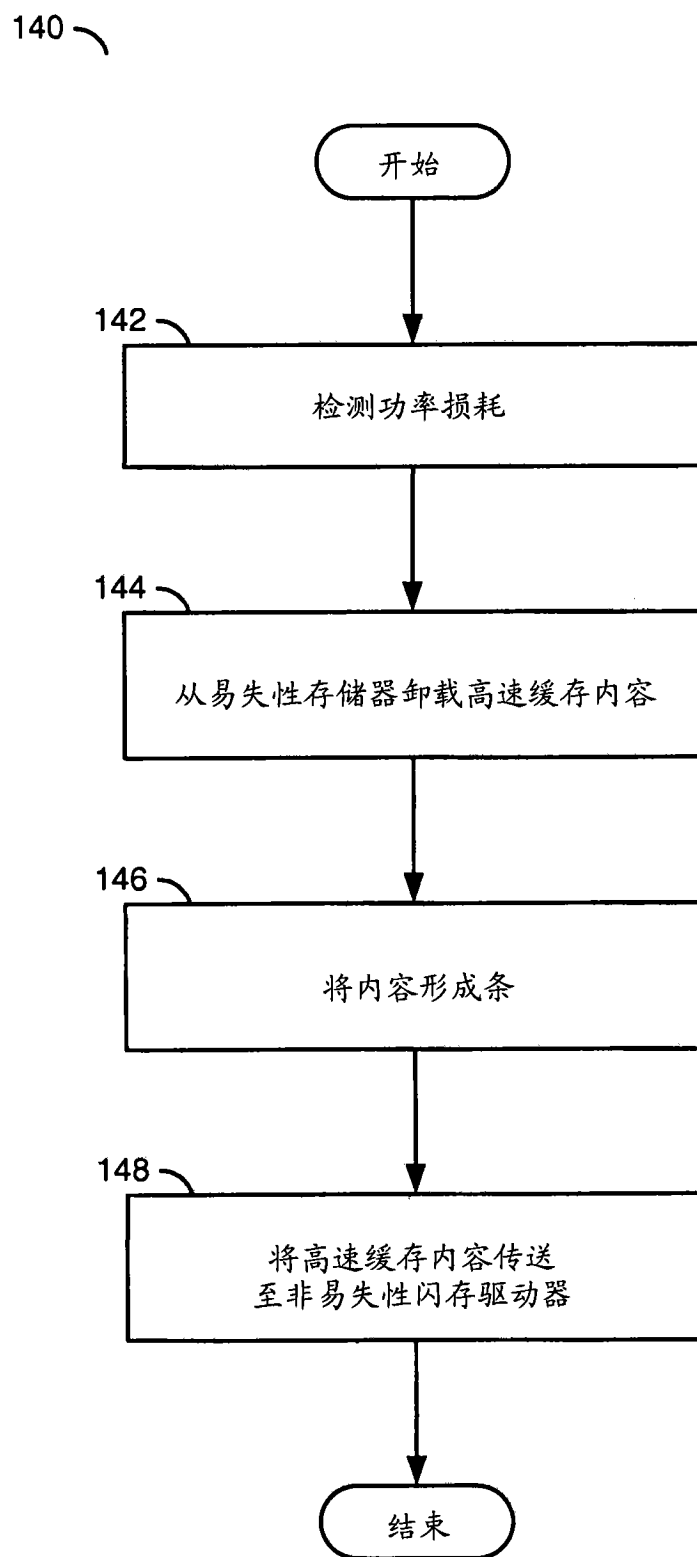


图 3

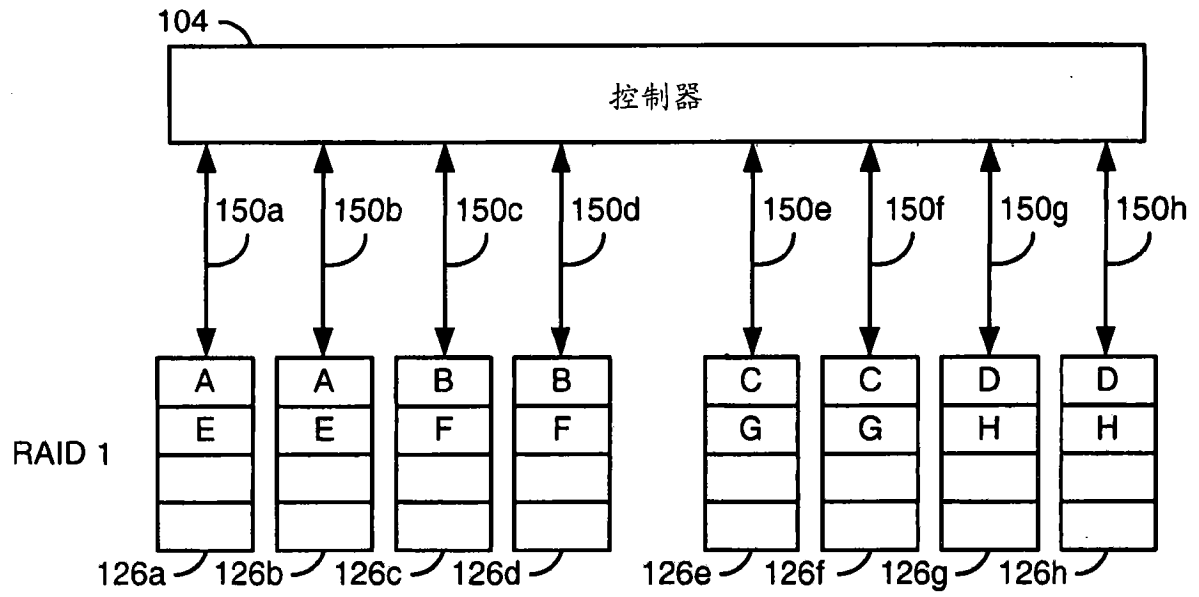


图 5

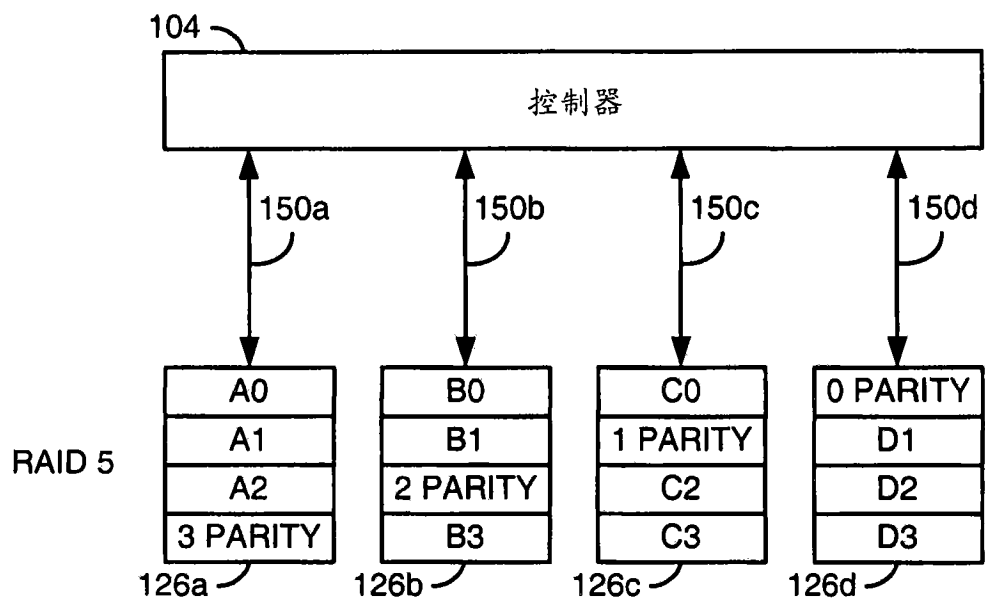


图 6