[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G06F 13/28 (2006.01)
G06F 13/40 (2006.01)
G06F 13/16 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200780021182.9

[43] 公开日 2009 年 6 月 24 日

[11] 公开号 CN 101467136A

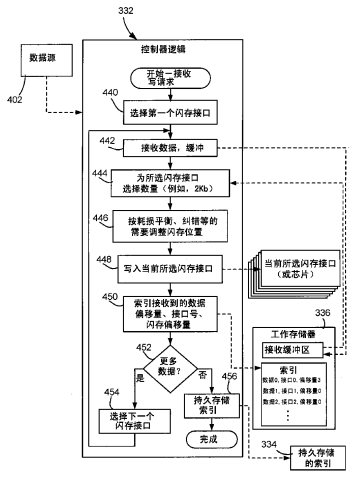
[22] 申请日 2007.6.1
[21] 申请号 200780021182.9
[30] 优先权
[32] 2006. 6. 9 [33] US [31] 11/450,015
[86] 国际申请 PCT/US2007/013072 2007.6.1
[87] 国际公布 WO2007/145869 英 2007.12.21
[85] 进入国家阶段日期 2008.12.8
[71] 申请人 微软公司
地址 美国华盛顿州
[72] 发明人 R·帕纳巴克
[74] 专利代理机构 上海专利商标事务所有限公司
代理人 陈 斌
权利要求书 4 页 说明书 14 页 附图 5 页

[54] 发明名称

高速非易失性存储器设备

[57] 摘要

所描述的是高速非易失性存储器设备和技术，其包括经由接口耦合到诸如单独的闪存芯片或单个芯片的单独区域等各组非易失性存储的控制器。该控制器包括通过在接口之间交错写入，包括通过在接口之间并行写入，来处理任意大小的请求的逻辑。例如，数据可经由直接存取访问(DMA)传输来接收。该控制器维护信息以允许交错的数据在诸如通过 DMA 读回时被重新组装到其正确的相对位置。高速非易失性存储器设备由此提供允许个人计算机快速引导或从诸如休眠等降低功率状态中恢复的硬件设备和软件解决方案。高速非易失性存储器设备还可出于诸如高速缓存和文件存储等其他数据存储目的来使用。



1. 一种在计算环境中的方法，包括：

接收涉及将数据写入非易失性存储设备的数据传输请求，其中要传输的数据无需匹配所述非易失性存储设备的数据格式要求；

通过数据传输机制接收对应于所述请求的数据；

转换所述数据以便写入多个非易失性存储设备接口，包括经由一个接口写入所述数据的一个部分，并且在经由所述一个接口写入所述数据的所述一个部分的同时，经由至少一个其它接口写入所述数据的另一个部分；以及

维护与所述数据相关联的、跟踪所述数据的每一部分被写入哪个非易失性存储设备接口的信息。

2. 如权利要求 1 所述的方法，其特征在于，接收所述数据包括通过直接存储器存取传输机制来接收所述数据。

3. 如权利要求 1 所述的方法，其特征在于，转换所述数据包括缓冲所述数据并基于非易失性存储块大小写入一定量的所述数据。

4. 如权利要求 1 所述的方法，其特征在于，还包括使用所维护的信息来读取所述数据，包括经由通过其写入所述数据的至少一个部分的接口读取所述数据的所述部分。

5. 如权利要求 4 所述的方法，其特征在于，接收所述数据包括结合进入计算机系统的降低功率状态接收对应于存储器内容的数据，并且其中，使用所维护的信息来读取所述数据包括在稍后时间还原所述存储器内容。

6. 如权利要求 4 所述的方法，其特征在于，接收所述数据包括接收计算机系统的引导相关信息，并且还包括使用所读取的数据的至少一部分来以引导计算机系统。

7. 如权利要求 1 所述的方法，其特征在于，维护所述信息包括维护与所写入的数据的相应的至少一部分相关联的至少一个偏移值。

8. 如权利要求 5 所述的方法，其特征在于，还包括确定对应于所述数据的所述部分的块的偏移值，并且在对应于所述偏移值的位置写入所述块。

9. 一种在计算环境中的系统，包括：

多个非易失性存储设备接口，每一接口都对应于可经由其相应的接口来访问的一组非易失性存储；以及

经由所述接口耦合到所述非易失性存储的控制器，所述控制器包括用于执行以下功能的逻辑：

a)通过在所述接口的至少两个之间交错写入，以及维护与所述数据相关联的、可用于随后确定数据的每一部分经由哪个接口写入的信息，来处理对应于一组数据的写请求，其中所述一组数据被允许比非易失性存储块写大小大，所述交错写入包括在经由一个接口写入所述数据的一个部分的同时经由至少一个其它接口写入所述数据的另一个部分，以及

b) 通过使用所维护的信息来确定数据的每一部分经由哪个接口写入，以及，并且对于每个部分，经由该接口读取所述数据的该部分，来处理对应于所述一组数据的读请求。

10. 如权利要求 9 所述的系统，其特征在于，所述控制器耦合到主存储器，并且通过直接存储器存取传输从所述主存储器接收所述数据以供写入，并且通过直接存储器存取传输将所读取的数据传输到所述主存储器。

11. 如权利要求 9 所述的系统，其特征在于，所述非易失性存储包括闪存。

12. 如权利要求 9 所述的系统，其特征在于，所述控制器通过 PCI Express®机制耦合到计算机系统。

13. 如权利要求 9 所述的系统，其特征在于，所述控制器耦合到计算机系统的主存储器，并且其中，所述控制器结合进入计算机系统的降低功率状态处理持久存储对应于主存储器内容的数据的写请求，并且其中，所述控制器处理在稍后时间还原所述主存储器内容的读请求。

14. 如权利要求 9 所述的系统，其特征在于，所述控制器处理输出用于引导计算机系统的信息的读请求。

15. 一种具有计算机可执行指令的计算机可读介质，所述指令在被执行时执行以下步骤：

接收将一组数据存储在非易失性存储器中的写请求，其中数据量对于用于写入所述非易失性存储器的块大小是任意的；

处理所述写请求，包括将所述一组数据的块大小的子集交错到对各组非易失性存储器的多个接口，包括通过经由至少两个接口同时写入；

维护关于所述块大小的子集如何在所述各组非易失性存储器之间定位的信息；

通过发出所述写请求完成的信号完成所述写入请求；

接收在完成所述写入请求后输出所述一组数据的至少一部分的读请求；

响应于所述读请求使用所维护的信息来输出所请求的数据，以使得所输出的所请求的数据不会因在处理所述写请求时发生的交错而改变；以及

通过发出所述读请求完成的信号完成所述读请求。

16. 如权利要求 15 所述的计算机可读介质，其特征在于，处理所述写请求包括通过直接存储器存取传输机制接收数据，并且其中，使用所维护的信息来输出所请求的数据包括通过直接存储器存取传输机制输出数据。

17. 如权利要求 15 所述的计算机可读介质，其特征在于，处理所述写请求对应于结合进入计算机系统的降低功率状态存储对应于存储器内容的数

据，并且其中，使用所维护的信息来输出所请求的数据对应于在恢复时还原所述存储器内容。

18. 如权利要求 15 所述的计算机可读介质，其特征在于，使用所维护的信息来输出所请求的数据对应于提供用于引导计算机系统的数据。

19. 如权利要求 15 所述的计算机可读介质，其特征在于，处理所述写请求对应于存储高速缓存的数据和/或存储文件。

高速非易失性存储器设备

背景

当引导时，当代个人计算机在用户能够对一任务进行操作之前经常花费大约二十秒来加载操作系统。该长延迟使用户感到烦恼，并且有时使得用户在可以使用完成该任务的其他手段时却不能打断计算机的启动，由此限制了个人计算机的有用性。

为了避免必须引导计算机以便使用其功能，已经引入了各种解决方案，诸如使得计算机在其先前使用之后进入待机状态（例如，ACPI S3 睡眠状态）。在待机状态中，向系统存储器提供某些电量以保留存储器内容。虽然待机允许个人计算机相对快速地恢复到其有用的状态，但待机具有当在睡眠模式时耗尽电池的缺点，并因此并非始终是合乎需要的。待机模式还会在掉电的情况下丢失数据，这意味着即使台式机或插电式笔记本也可能在电源故障期间丢失数据。

提供快速启动的另一解决方案被称为休眠（例如，ACPI S4 状态），其中存储器的内容在休眠时被传送到硬盘休眠文件，并且当恢复到正常操作状态时从休眠文件中读回。该解决方案也有缺点，包括它花费相对较长的时间来恢复，这是因为休眠文件需要从相对较慢的硬盘驱动器传送回到系统存储器并解包。

概述

提供本概述以便以简化形式介绍将在以下的详细描述中进一步描述的一些代表性概念。本概述并不旨在标识出所要求保护的主题的关键特征或必要特征，也不旨在用于以任何方式帮助确定所要求保护的主题的范围。

简言之，此处所描述的主题的各方面涉及一种高速非易失性存储器设备，其包括经由接口耦合到诸如单独的闪存芯片或一个闪存芯片的单独区域等各组 非易失性存储的控制器。该控制器包括通过在接口之间交错写入，包括通过在适当时在接口之间并行写入来处理基本上任意大小的请求的逻辑。例如，数据可经由直接存储器存取（DMA）传输来接收，并且作为块写来写入闪存。

控制器维护与写请求相关联的信息以允许交错的数据在稍后被读回时被重新组装到其正确的相对位置。

当控制器接收到读请求时，该控制器通过使用所维护的信息来确定对应于该请求的数据是如何交错和存储的来处理该读请求。该数据然后通过从用于写入该数据的适当的接口读回每个块（或其他子集）来返回。例如，数据可经由DMA 传输来返回。

特别地，该高速非易失性存储器设备由此提供一种允许个人计算机快速引导或从休眠或其他降低功率状态中恢复，由此使得个人计算机就例如启动时间而言更像消费电子设备的硬件设备和软件解决方案。该高速非易失性存储器设备还可出于其他数据存储目的来使用，诸如在正常操作期间的高速缓存和文件存储，诸如用于提供比硬盘交换更快的额外的存储器。因为控制器处理非易失性存储器和计算机系统的剩余部分之间的转换，所以可经由单个请求来保存任意量的数据，而无需由数据源来特殊格式化，由此便于快速操作。

结合附图阅读以下详细描述，本发明的其他优点会变得显而易见。

附图简述

作为示例而非限制，在附图中示出了本发明，附图中相同的附图标记指示相同或相似的元素，附图中：

图 1 示出了可以将本发明的各方面并入其中的通用计算环境的说明性示例。

图 2 是其中存在高速非易失性存储器设备的示例个人计算机系统体系结构的表示。

图 3 是示例高速非易失性存储器设备的表示。

图 4 是示例高速非易失性存储器设备的表示，包括由其中的控制器逻辑执行以便将数据写入非易失性存储的各示例步骤。

图 5 是示例高速非易失性存储器设备的表示，包括由其中的控制器逻辑执行以便从非易失性存储读取数据的各示例步骤。

详细描述

示例性操作环境

图 1 示出了可在其上实现本发明的合适的计算系统环境 100 的示例。计算系统环境 100 只是合适计算环境的一个示例，而非意在暗示对本发明使用范围或功能有任何限制。也不应该把计算环境 100 解释为对示例性操作环境 100 中示出的任一组件或其组合有任何依赖性或要求。

本发明可用各种其它通用或专用计算系统环境或配置来操作。适用于本发明的公知的计算系统、环境和/或配置的示例包括，但不限于：个人计算机、服务器计算机、手持式或膝上型设备、图形输入板设备、多处理器系统、基于微处理器的系统、机顶盒、可编程消费者电子产品、网络 PC、小型机、大型计算机、包括上述系统或设备中的任一个的分布式计算机环境等。

本发明可在诸如程序模块等由计算机执行的计算机可执行指令的通用上下文中描述。一般而言，程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件、数据结构等等。本发明也可以在其中任务由通过通信网络链接的远程处理设备执行的分布式计算环境中实施。在分布式计算环境中，程序模块可以位于包括存储器存储设备在内的本地和/或远程计算机存储介质中。

参考图 1，用于实现本发明的示例性系统包括计算机 110 形式的通用计算设备。计算机 110 的组件可以包括但不限于：处理单元 120、系统存储器 130 和将包括系统存储器在内的各种系统组件耦合至处理单元 120 的系统总线 121。系统总线 121 可以是几种类型的总线结构中的任何一种，包括存储器总线或存储控制器、外围总线、以及使用各种总线体系结构中的任一种的局部总线。作为示例，而非限制，这样的体系结构包括工业标准体系结构（ISA）总线、微通道体系结构（MCA）总线、增强型 ISA（EISA）总线、和外围部件互连（PCI）总线（也称为夹层（Mezzanine）总线）。

计算机 110 通常包括各种计算机可读介质。计算机可读介质可以是能由计算机 110 访问的任何可用介质，而且包含易失性和非易失性介质以及可移动、不可移动介质。作为示例而非限制，计算机可读介质可以包括计算机存储介质和通信介质。计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其它数据这样的信息的任何方法或技术来实现的易失性和非易失

性、可移动和不可移动介质。计算机存储介质包括，但不限于，RAM、ROM、EEPROM、闪存或其它存储器技术、CD-ROM、数字多功能盘（DVD）或其它光盘存储、磁带盒、磁带、磁盘存储或其它磁性存储设备、或能用于存储所需信息且可以由计算机 100 访问的任何其它介质。通信介质通常以诸如载波或其它传输机制等已调制数据信号来体现计算机可读指令、数据结构、程序模块或其它数据，并包括任一信息传送介质。术语“已调制数据信号”指的是其一个或多个特征以在信号中编码信息的方式被设定或更改的信号。作为示例而非限制，通信介质包括有线介质，诸如有线网络或直接线连接，以及无线介质，诸如声学、RF、红外线和其它无线介质。上述中的任意组合也应包括在计算机可读介质的范围之内。

系统存储器 130 包括易失性和/或非易失性存储器形式的计算机存储介质，如只读存储器（ROM）131 和随机存取存储器（RAM）132。基本输入/输出系统 133（BIOS）包含有助于诸如启动时在计算机 110 中的元件之间传递信息的基本例程，它通常被存储在 ROM 131 中。RAM 132 通常包含处理单元 120 可以立即访问和/或目前正在其上操作的数据和/或程序模块。作为示例而非局限，图 1 示出了操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137。

计算机 110 还可以包括其它可移动/不可移动、易失性/非易失性计算机存储介质。仅作为示例，图 1 示出了从不可移动、非易失性磁介质中读取或向其写入的硬盘驱动器 141，从可移动、非易失性磁盘 152 中读取或向其写入的磁盘驱动器 151，以及从诸如 CD ROM 或其它光学介质等可移动、非易失性光盘 156 中读取或向其写入的光盘驱动器 155。可以在示例性操作环境中使用的其它可移动/不可移动、易失性/非易失性计算机存储介质包括但不限于，磁带盒、闪存卡、数字多功能盘、数字录像带、固态 RAM、固态 ROM 等等。硬盘驱动器 141 通常由不可移动存储器接口，诸如接口 140 连接至系统总线 121，磁盘驱动器 151 和光盘驱动器 155 通常由可移动存储器接口，诸如接口 150 连接至系统总线 121。

以上描述并在图 1 中示出的驱动器及其相关联的计算机存储介质为计算机 110 提供了对计算机可读指令、数据结构、程序模块和其它数据的存储。例如，在图 1 中，硬盘驱动器 141 被示为存储操作系统 144、应用程序 145、其

它程序模块 146 和程序数据 147。注意，这些组件可以与操作系统 134、应用程序 135、其它程序模块 136 和程序数据 137 相同，也可以与它们不同。操作系统 144、应用程序 145、其它程序模块 146 和程序数据 147 在这里被标注了不同的标号是为了说明至少它们是不同的副本。用户可通过诸如图形输入板或者电子数字化仪 164、话筒 163、键盘 162 和定点设备 161（通常指的是鼠标、跟踪球或触摸垫）等输入设备向计算机 110 输入命令和信息。图 1 中未示出的其它输入设备可以包括操纵杆、游戏手柄、圆盘式卫星天线、扫描仪等。这些和其它输入设备通常由耦合至系统总线的用户输入接口 160 连接至处理单元 120，但也可以由其它接口或总线结构，诸如并行端口、游戏端口或通用串行总线（USB）连接。监视器 191 或其它类型的显示设备也经由接口，诸如视频接口 190 连接至系统总线 121。监视器 191 也可以与触摸屏面板等集成。注意到监视器和/或触摸屏面板可以在物理上耦合至其中包括计算设备 110 的外壳，诸如在图形输入板型个人计算机中。此外，诸如计算设备 110 等计算机还可以包括其它外围输出设备，诸如扬声器 195 和打印机 196，它们可以通过输出外围接口 194 等连接。

计算机 110 可使用至一个或多个远程计算机，如远程计算机 180 的逻辑连接在网络化环境中操作。远程计算机 180 可以是个人计算机、服务器、路由器、网络 PC、对等设备或其它常见的网络节点，并且通常包括许多或所有以上相对于计算机 110 所描述的元件，尽管在图 1 中仅示出了存储器存储设备 181。图 1 中所示的逻辑连接包括局域网（LAN）171 和广域网（WAN）173，但也可以包括其它网络。这样的联网环境在办公室、企业范围计算机网络、内联网和因特网中是常见的。

当在 LAN 联网环境中使用时，计算机 110 通过网络接口或适配器 170 连接至 LAN 171。当在 WAN 联网环境中使用时，计算机 110 通常包括调制解调器 172 或用于通过诸如因特网等 WAN 173 建立通信的其它装置。调制解调器 172 可以是内置或外置的，它可以通过用户输入接口 160 或其它合适的机制连接至系统总线 121。在网络化环境中，相对于计算机 110 所描述的程序模块或其部分可被储存在远程存储器存储设备中。作为示例而非局限，图 1 示出远程应用程序 185 驻留在存储器设备 181 上。可以理解，所示的网络连接是示例性

的，也可以使用在计算机之间建立通信链路的其他手段。

辅助显示子系统 199 可经由用户接口 160 连接以允许诸如程序内容、系统状态和事件通知等数据被提供给用户，即使计算机系统的主要部分处于低功率状态中。辅助显示子系统 199 可连接至调制解调器 172 和/或网络接口 170 以允许在主处理单元 120 处于低功率状态中时在这些系统之间进行通信。

高速非易失性存储器设备

此处所描述的技术的各方面涉及一种提供快速引导或从休眠中恢复以及其他用途的非易失性存储器设备。然而，如可以理解的，此处所描述的技术并不限于任何特定用途或类型的睡眠状态，例如，完全通电和完全断电之间的其他状态可从这一设备中获益，并且在操作状态中时的一般使用以便有助于性能也是可能的。由此，本发明不限于此处所描述的示例、使用模型、结构或功能。相反，此处所描述的任何使用模型、示例、结构或功能都是非限制性的，并且本发明一般能够以在计算和数据存储方面提供好处和优点的各种方式来使用。

在图 2 所一般表示的一个示例实现中，此处所描述的技术的一部分被结合到耦合到典型的北桥/南桥芯片组的南桥组件 202 的高速非易失性存储器设备 200 中。如在这一体系结构中已知的，（可对应于图 1 的计算机系统 110）CPU 204 通过总线和北桥组件 206 耦合到动态 ram 208。北桥组件 206 进而通过另一总线耦合到南桥组件 202，南桥组件 202 耦合到 I/O 设备。硬盘驱动器 210 连同高速非易失性存储器设备 200 一起被例示为连接的 I/O 设备，尽管可连接众多其他类型的设备。例如，出于可扩展性的目的，在某些示例体系结构中，南桥 202 具有用于将 PCI Express® (PCIe) 组件耦合到计算机系统的接口，并且这是可以耦合诸如设备 200 等高速非易失性存储器设备的一种方式。用于耦合高速非易失性存储器设备的桌上型（例如，PCIe 卡）的替换方案包括诸如迷你 pci、PCMCIA、和 Express 卡等组件、被制成插入专用连接器，或甚至更直接地将设备耦合（例如，焊接）到主板的组件封装。本质上，对于所需高速是足够的将非易失性存储器设备耦合到计算机系统的任何方式和/或手段是等效的。图 3 示出了高速非易失性存储器设备 200 的一个示例，其包括闪存设备 320-325。如可以理解的，该设备通过允许发生基本上并行的读和写操作的交

错技术比常规闪存设备更快。尽管示出了闪存（例如，基于NAND或基于NOR的），但可以理解，可以使用任何非易失性存储装置来替换闪存或作为其补充，包括备有电池的RAM。尽管在图3中例示了六个这样的闪存设备320-325，但可以理解，存储器设备200可包含任何实用数量的闪存芯片等等。

此外，可以理解，设备制造商可以在单个集成电路封装中实现该设备，和/或还可提供到更多组合的闪存组的并行接口，例如，具有到单独区域的六个并行接口的一组闪存本质上等价于各自具有其自己的接口的六个独立的闪存芯片。换言之，代替具有多个单独的闪存设备320-325等具有单个存储器设备，其具有可各自经由单独接口等被同时访问的内部并行区域本质上是等效的。例如，具有相对快得多的接口的NOR类型的设备可提供该芯片上可被同时读写的多个区域。如此处所使用的，关于闪存的术语“接口”包括用于与一组闪存进行通信的任何机制，该组闪存包括包含独立闪存设备的一组闪存或包含闪存设备的可单独访问区域的一组闪存。

为了实现所需的高速操作，存储器设备200包括控制器330。控制器330包括逻辑332，其特别地懂得如何以对应于在其上接收发往该闪存的数据并且在其上发送从该闪存中读取的数据的闪存接口/协议和外部接口/协议的方式从该闪存中读取并向其写入。换言之，控制器逻辑部分地担当传输机制和非易失性存储器之间的转换器。在图3的示例中，外部接口/协议基于PCIe标准，然而，可容易地理解，实际上可以使用任何适合的数据通信机制和相应的协议，例如，基于SATA（串行高级技术附件）的总线接口和协议。此外，注意，需要至少一条数据线，但是如由去往和发自控制器330的虚线所指示的，在给定配置中可存在更多数据线，例如，多条PCIe线可传输数据。

诸如控制器330等控制器可以按各种方式来实现。例如，控制器可使其逻辑硬连线，诸如在具有对大块传输有效的简单交错策略的相对直接的高速非易失性存储器设备中。可优化这一控制器以增进对于类似休眠文件存储和还原的任务的性能。能灵活地存储引导数据、休眠文件、常规文件和/或可担当高速缓存（并且可能可执行诸如耗损平衡等存储器管理技术）的更复杂的控制器可在闪存中编码以使得可以在必要时或需要时对该控制器逻辑做出更新。控制器能够同时满足许多读或写操作。此外，控制器可以是动态的，这表现在它可检测

或被通知正在使用多少 I/O 数据线（例如，PCIe 线），并相应地修改其操作。类似地，可编码控制器以使其适应有多少闪存接口和/或多大的闪存是可用的，使得制造商能够使用具有不同闪存配置的相同的控制器，包括用户可通过添加（或通过移除来修改）闪存设备来扩展的配置。控制器还可检测或以其他方式被通知至少一个其他高速非易失性存储器设备，并且能够与该其他设备的控制器传达并协调数据读和写，诸如用于允许通过简单地添加第二块卡等来扩展高速非易失性存储器的数量。例如，在两个设备的系统中，一个控制器能够让另一个控制器处理预定的（例如，协商的）一半发往和来自其闪存的 DMA 传输，由此使总体速度翻倍（假设 DMA 通道未满）。

本质上控制器 330 并行地向闪存设备 320-325 写入并从中读取，同时维护索引 334 以便跟踪哪些数据被写到哪些闪存位置。注意，索引 334（以及可能地，逻辑 330）可以在闪存中维护，诸如在闪存设备 320-325 中的一个的某一位置中。控制器 330 还可具有用作正常操作时的临时索引的高速缓冲工作存储器 336，且内容在需要时被持久存储到闪存索引 334 以防止数据丢失。例如，高速读可通过首先将索引信息从闪存复制到高速缓冲区/工作存储器 336，并且然后访问该高速缓冲区/工作存储器 336 以便建立数据传输而不是对于每次数据传输从较慢的闪存中读取来实现。同样，如将在以下描述的，该索引数据可被保留在工作存储器 336 中，直到在将要写入整个数据集的成功写入时，而不是在每次部分写入或某一组更小的部分写入时被持久存储到闪存为止。

如图 3 所表示的，每个存储器设备 320-325 可以基本上同时由设备的控制器 330 来访问，这允许累积数据率变得非常高，包括当设备 200 使用 DMA 技术来将大数据块传输到主系统存储器 208（图 2）中时。注意，写入设备 200 比起读取可能要慢得多，因为非易失性存储器技术当前在写入方面较慢。例如，这对于 NAND 闪存以及读取比标准 NAND 快得多的 OneNAND 类型混合存储器而言都是如此。

在一个实现中，存储器设备 200 被配置成快访问设备，且控制器逻辑 332 被配置成尽可能快地传输尽可能多的数据，例如，通过 DMA。为此，控制器在索引 334 中跟踪哪些块的每个存储器设备中。此外，控制器 330 可管理哪个设备得到数据的哪个部分，由此数据提供者无需关注格式化块大小，将数据匹

配到分配单元边界等等。相反，控制器逻辑 332 在需要时中断大的写请求，诸如最大化并行写和读回，以及执行诸如耗损平衡等其他存储器管理技术。

作为示例逻辑，考虑相对较大的（例如，两兆字节）数据写请求由控制器 330 从某一数据源 402（图 4）接收，且某一数量的闪存接口（例如，对于设备 320-325 中的每一个有一个闪存接口）可用，每个闪存接口都被配置成一次写入两千字节的块。尽管图 4 未明确示出，但控制器逻辑 332 可执行检查等，诸如用于确保写请求在给定可用存储器的量的情况下不至于太大，和/或还可将任何高速缓存的数据转储清除到硬盘驱动器以便腾出空间，例如，程序的临时高速缓存数据可被转储清除到硬盘驱动器以便为休眠文件腾出空间。

如图 4 所例示的，控制器逻辑 332 交错数据以使得块被并行写入到可用的单独闪存接口，例如，选择（步骤 440）第一闪存接口以便将前两千字节写入到其中，将后两千字节写入到闪存芯片 321，以此类推。为此，控制器可以在接收数据时缓冲数据（步骤 442 和 444），这可对于剩余步骤独立（且并行）地发生直到缓冲区满。如可容易地理解的，通过具有足够大以处理对应于每个块的数据的缓冲区，例如，对于一次写入两个块的六个闪存芯片的至少一万两千字节，加上也许额外的数量以便当正在发生全组写入时开始收集下一个块，并行写的数量基本上被最大化。控制器在其缓冲区满时输出忙碌等。

在步骤 448 处，当缓冲了至少块大小数量的数据时，该逻辑将适当大小的块从缓冲区写入到当前所选闪存接口中的位置，并且然后在当前选择下一个闪存接口时将下一个块写入到该接口中的位置，以此类推。每次成功写入块时，在步骤 450 处更新索引 334 以使得该数据在稍后接收到读请求时可被重新组装。例如，接收到的数据偏移量或相应的排序信息，接口标识符（例如，哪个闪存芯片）和对闪存的位置偏移量将足够作为对于所写入的每一块的基本索引记录；还与该索引相关联的是将该数据返回到其适当的位置所需的任何信息，例如，在保存文件情况下的文件属性。注意，如果不需要对该数据的随机访问，例如，该数据只允许被循序读回，则数据偏移量/排序信息可以是固有的，例如，如果对索引 334 进行排序以使得按序对数据块进行重新排序，则无需维护所接收到的数据偏移量或排序信息。然而，这将意味着控制器在例如由于坏写而时序改变的情况下可能必须调整排序。

此外，可以使用预定约定以替换索引模式或与之相结合。例如，可以将前2千字节的块按照一个启始偏移量写入到一个闪存设备，将接下来的2千字节的块按照其启始偏移量写入到下一个闪存设备，以此类推。有了这一约定，只需记录启始闪存设备和每个闪存设备的启始偏移量，这可例如记录在数据开始处的首部等中（例如，作为六个指针，对每个设备有一个指针）。这将为整个索引模式节省空间。可以在需要修改预定约定的情况下记录异常，例如，为了耗损平衡、纠错等。例如，可以使用压缩类型的索引模式，其中只要偏移量满足约定，即，除非它是除了来自前一偏移量的两千字节之外的某个偏移量，该偏移量可被留空，而不是索引每个偏移量。

注意，图4中的示例逻辑包括步骤446，通过该步骤可以改变到当前所选闪存接口中的偏移位置（以及如果有必要的话该闪存接口本身）。不让位置线性前进的原因可包括执行耗损平衡技术，执行任何纠错（例如，为了绕过已知的坏块），以及还为了确保诸如为持久存储索引而保留的以及可能地用于存储逻辑（或对以其他方式持久存储的逻辑的更新/扩展）的任何保留的闪存区域不被盖写。

如上所述，步骤448表示写入，且步骤450表示索引。这两个步骤本质上基本是事务性的，使得例如对应于写入的数据直到该写入成功才被提交给索引，例如，在该写入由于检测到坏块而需要被重新尝试到不同的块的情况下。注意，当正发生该写入时，控制器在更多数据可用时下不等待（步骤452），而是选择下一个闪存接口（步骤454）以供写入下一个数据块。本质上，控制器收集数据直到其接收缓冲区满，并且执行对闪存的写入直到所有（或某一期望数量的）闪存接口被占用，且当其期望数量的闪存接口正在使用时只延迟写入。

当没有剩下要写入的数据时，步骤452分支到步骤456，其中索引334与现在写入的数据相关联地持久存储。由此可通过在持久存储之前丢弃索引来防止不完整的写入。这还提供了更快的速度，因为索引334可被暂时保留在超高速SRAM或DRAM存储器中，直到当完成完整写入时被持久存储。如果被成功地持久存储，则返回成功等，否则返回错误代码。注意，可通过在写入所有数据之前持久存储任何成功写入的索引信息来允许不完整的写入，尽管比每整

个数据写入一次（例如，每个块写入一次）更频繁地将索引持久存储到闪存将减缓设备。

图 5 表示用于处理来自某一数据请求者的读取请求的示例逻辑。该请求本质上可以是任何种类，例如，读取文件、从高速缓存中读取、读取“打开”文件的一部分等等，但出于此示例的目的，此示例将被一般描述为顺序块数据传输，诸如用于读出从开始到结束的休眠文件。

步骤 550 表示定位该请求的相应的（例如，持久存储的）索引，并且如果该索引尚未在工作存储器中，则可将该索引 334 读入工作存储器 336。注意，取决于存储了什么数据可能存在不同的索引，例如，每个文件一个索引。

步骤 552 表示为该请求选择第一个索引条目。注意，所索引的可能已经以反映接收数据的次序的方式保存了，在这种情况下，没有理由排序或以其他方式确定哪个索引条目是第一个。在从某一偏移量的随机访问读取的情况下，要读取的第一组数据可由启始偏移值来确定，由此控制器逻辑 332 可扫描该索引的数据以便精确定位哪两千字节的块包含第一组所请求的数据，并执行计算以确定该块中的确切字节以便开始返回。

步骤 554 和 556 涉及在正确位置从正确闪存接口（或设备）读取，这通过索引数据来确定。该数据从闪存中被读入输出缓冲区中（例如，工作存储器 336 中），且缓冲区输出的内容（例如，经由设备 200 或每个接口的 DMA 引擎）在步骤 558 处作为数据经由闪存读取变得可用。因为闪存读取相对于经由 DMA/PCIe 输出数据花费较长的时间，所以控制器在没有数据可用时就不输出数据（并且如果有必要可发出忙碌信号以便于异步操作）。注意，从其他闪存接口读取经由步骤 560 和 562 并行执行，这通过循环返回到步骤或以其他方式移动通过索引（步骤 552 和 554）以将正确的数据读回到输出缓冲区以便传输回到正确的位置来实现。注意，多个 DMA 引擎 570（例如，每个闪存设备一个）可被设置成控制总线以实现传输，并且可驻留在设备 200 中、南桥 202 中、和/或作为独立组件，并且这些引擎可由设备控制器 330、BIOS 和/或 CPU 204 来控制。例如，有了图 3 的高速存储器设备 200，可以在任何给定时刻建立并发生六个单独的、并行的 DMA 传输，直到所请求的数据被完全传输回到主存储器（例如，图 2 的 RAM 208）中。

当所有数据都已例如经由 DMA 被传输回去时，控制器可在步骤 564 处发出“完成”状态的信号，诸如经由“成功”错误代码等。对于多个请求，可返回对应于每个请求的标识符（虚拟块），例如，读请求 X（对应于如在写入时所标识的写请求 X）被成功地传输回到存储器中。以此方式，请求者知道正确的数据现在在 RAM 中，而不管闪存是如何访问的，例如，按照从闪存读回的次序、设备之间不同的读取速度等等。此时，数据请求者 502 知道所有请求的数据都在正确的 RAM 存储器位置中。超时或不成功错误代码可以在失败的情况下被请求者检测到。

转向使用所例示的高速存储器设备 200 的硬件实现和软件解决方案来在从完全断电（例如，ACPI S5 状态）冷引导的情况下更快地引导计算机系统的示例，设备 200 可保存引导所需的所有（或大多数）文件及其他信息。这些文件可通过 DMA（直接存储器存取）被传输到被配置在系统存储器中的 RAM 盘中，且计算机系统从该 RAM 盘中引导。如可容易理解的，这允许比从常规硬盘驱动器访问快得多地访问所需文件。

一种替换即时引导解决方案也可用于诸如吉比特以太网等相对高速网络。在该替换方案中，网络设备被要求访问特定网络资源，并经由 DMA 将其传输到存储器中，如在以上解决方案中所一般描述的。

对于进入休眠状态，包含在休眠时刻的 RAM 内容的休眠文件被写入到闪存中。例如，BIOS 可将 RAM 的内容配置到诸如 RAM 中的压缩休眠文件中，并且然后经由对控制器 330 的单个请求，可通过 DMA 传输该休眠文件以便持久存储在高速非易失性存储设备 200 中。注意，大规模传输就开销而言节省大量时间；例如，有了单条 2.6 吉比特 PCIe 线，从写入者的观点来看，以单个请求可将 200 兆字节传输到两千字节、六接口的闪存设备。如上所述，控制器并行写入六个闪存接口，这基本上以写入单个设备六倍的速度写入。

对于从休眠状态中恢复，例如，当计算机系统从 ACPI S4 睡眠状态中恢复时，包含休眠时刻的 RAM 内容的休眠文件被传输回到主存储器中。这通过使得设备控制器对于例如每个存储器设备 320-325 启动 DMA 引擎来非常快速地（相对于硬盘读取）实现。

注意，一替换非易失性设备可具有一般保存主系统存储器并然后还原它的

接口，而不是被配置成用于休眠的块模式设备。例如，可实现协议以使得系统 BIOS 调用该接口，导致主存储器的快照被传输到该非易失性设备和从该非易失性设备传输，从而本质上使得该 BIOS 执行“准 S3”恢复。在一个示例准 S3 场景中，用户或系统可进入待机，并具有由 BIOS 自动保存到高速非易失性存储器设备的（可能采用休眠文件的形式的）存储器内容的快照。BIOS 然后可执行操作，诸如在某一时间到期后关闭系统电源，并且在电源被关闭或以其他方式丢失的情况下从高速非易失性存储器设备恢复，或者在电源未被关闭的情况下从待机中恢复而不从高速非易失性存储器设备读回。通过 BIOS，用户或系统还可执行标准 S4 休眠，例如，通过直接进入休眠并从休眠中恢复。

尽管在设备中可能存在任何实用数量的非易失性存储器，但为了优化引导和休眠，期望具有足够的存储器容量以便保存整个休眠文件和/或所有或大多数引导文件。例如，在设备上可能存在大约 128 或 256 兆可用字节，但对于休眠，非易失性数量可对应于易失性存储器内容在被压缩时的大小。较大容量的非易失性设备可持久存储引导和恢复文件。

通过提供并行地管理接口的控制器，同时使得控制器划分一请求，由此允许仅单个请求来处理任何任意大小的写或读并由此消除主处理器方的开销，实现了显著的速度增益。然而，主处理器可以在需要时将请求分成各虚拟块以便诸如在从休眠中恢复时写入及稍后读取，以便在其他虚拟块还正在被传输的同时开始执行首先恢复的某些代码。可以维护任何实用数量的虚拟块；然而，注意，与文件系统不同，虚拟块大小是可对应于请求而变化的。由此，写和读请求者无需关注格式化数据以供闪存写和读。

高速非易失性存储器设备的另一用途模型包括补充主系统易失性（例如，DRAM）存储器。例如，相对较大的高速缓存可由设备来提供以便将数据交换进出主存储器。在该示例中，高速、非易失性存储器设备本质上可用作诸如减少（并且有时可消除）访问硬盘以便进行虚拟存储器交换的需求的中间高速缓存。文件系统也可将文件写入高速、非易失性存储器设备。如可容易理解的，这一示例提高了应用程序和其他程序的性能。

高速非易失性存储器设备的又一用途模型是当主系统 CPU 未通电，或以其他方式被占用时为系统组件提供存储。例如，网卡、传真卡、辅助设备等各

自都可被配置成对高速非易失性存储器设备读写数据而无需 CPU 协助。

尽管本发明易于作出各种修改和替换构造，其某些说明性实施例在附图中示出并在上面被详细地描述。然而应当了解，这不旨在将本发明限于所公开的具体形式，而是相反地，旨在覆盖落入本发明的精神和范围之内的所有修改、替换构造和等效方案。

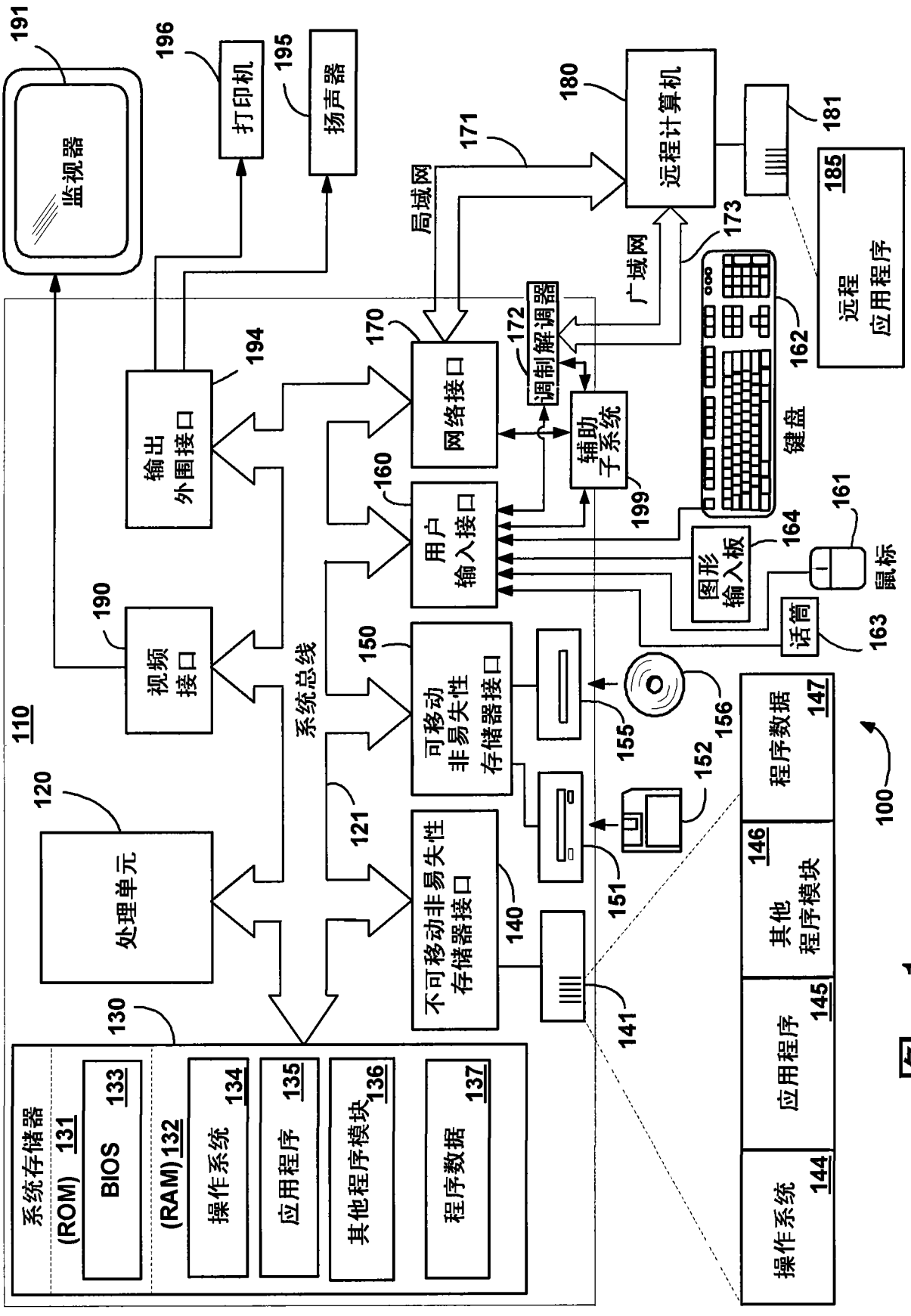

监视器
191
打印机
196
扬声器
195
局域网
171
远程计算机
180
广域网
173
181
185
远程
应用程序
输出
外围接口
194
网络接口
170
调制解调器
172
辅助
子系统
199
键盘
162
图形
输入板
164
161
鼠标
话筒
163
用户
输入接口
160
视频
接口
190
系统总线
110
可移动
非易失性
存储器接口
150
处理单元
120
121
不可移动非易失性
存储器接口
140
155
156
152
151
141
130
系统存储器
(ROM) 131
BIOS 133
(RAM) 132
操作系统 134
应用程序 135
其他程序模块 136
程序数据 137
操作系统
144
应用程序
145
其他
程序模块
146
程序数据
147
100

图 1

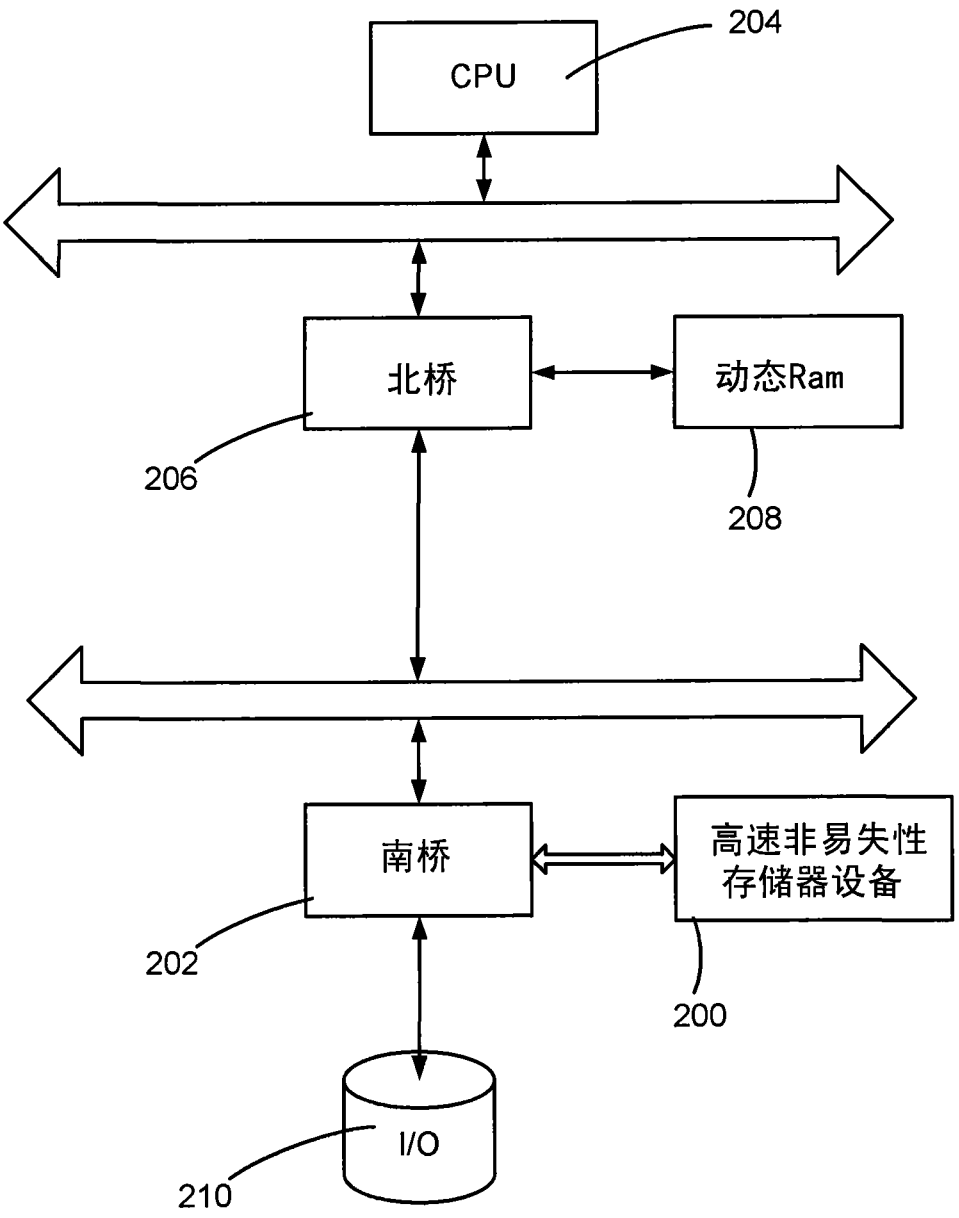


图 2

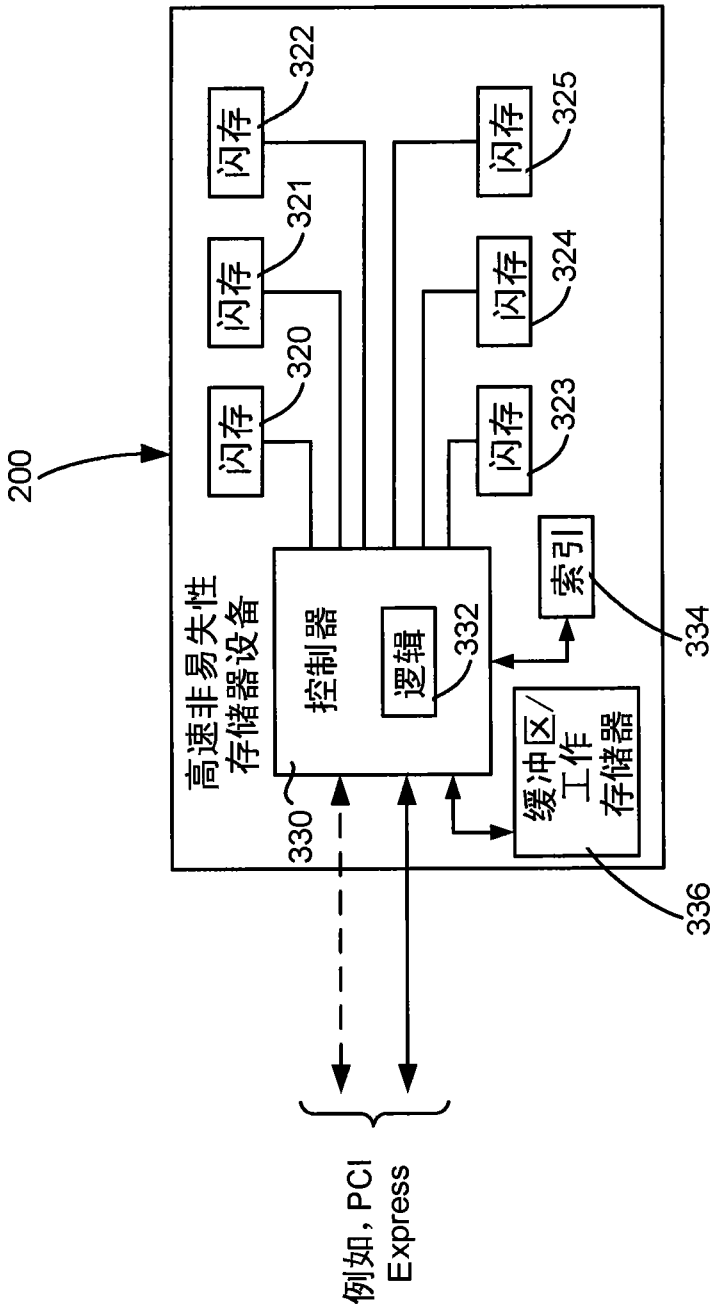

200
高速非易失性存储器设备
闪存
322
闪存
321
闪存
320
闪存
325
闪存
324
闪存
323
控制器
逻辑
332
索引
334
缓冲区/工作存储器
336
330
例如，PCI Express

图 3

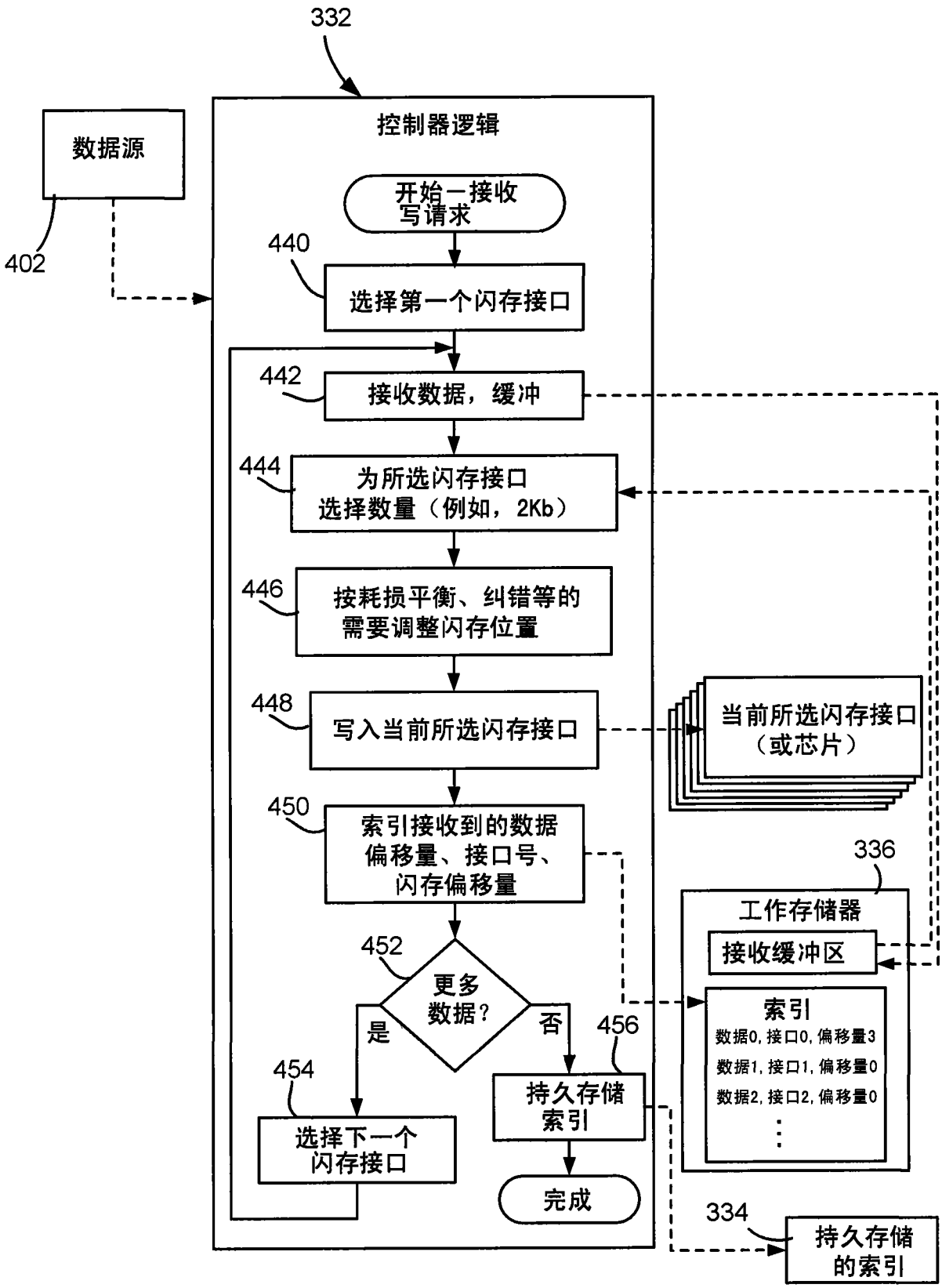

332
控制器逻辑
数据源
402
开始—接收写请求
440
选择第一个闪存接口
442
接收数据，缓冲
444
为所选闪存接口选择数量（例如，2Kb）
446
按耗损平衡、纠错等的需要调整闪存位置
448
写入当前所选闪存接口
当前所选闪存接口（或芯片）
450
索引接收到的数据偏移量、接口号、闪存偏移量
336
工作存储器
接收缓冲区
索引
数据0, 接口0, 偏移量3
数据1, 接口1, 偏移量0
数据2, 接口2, 偏移量0
452
更多数据?
是
否
454
选择下一个闪存接口
456
持久存储索引
完成
334
持久存储的索引

图 4

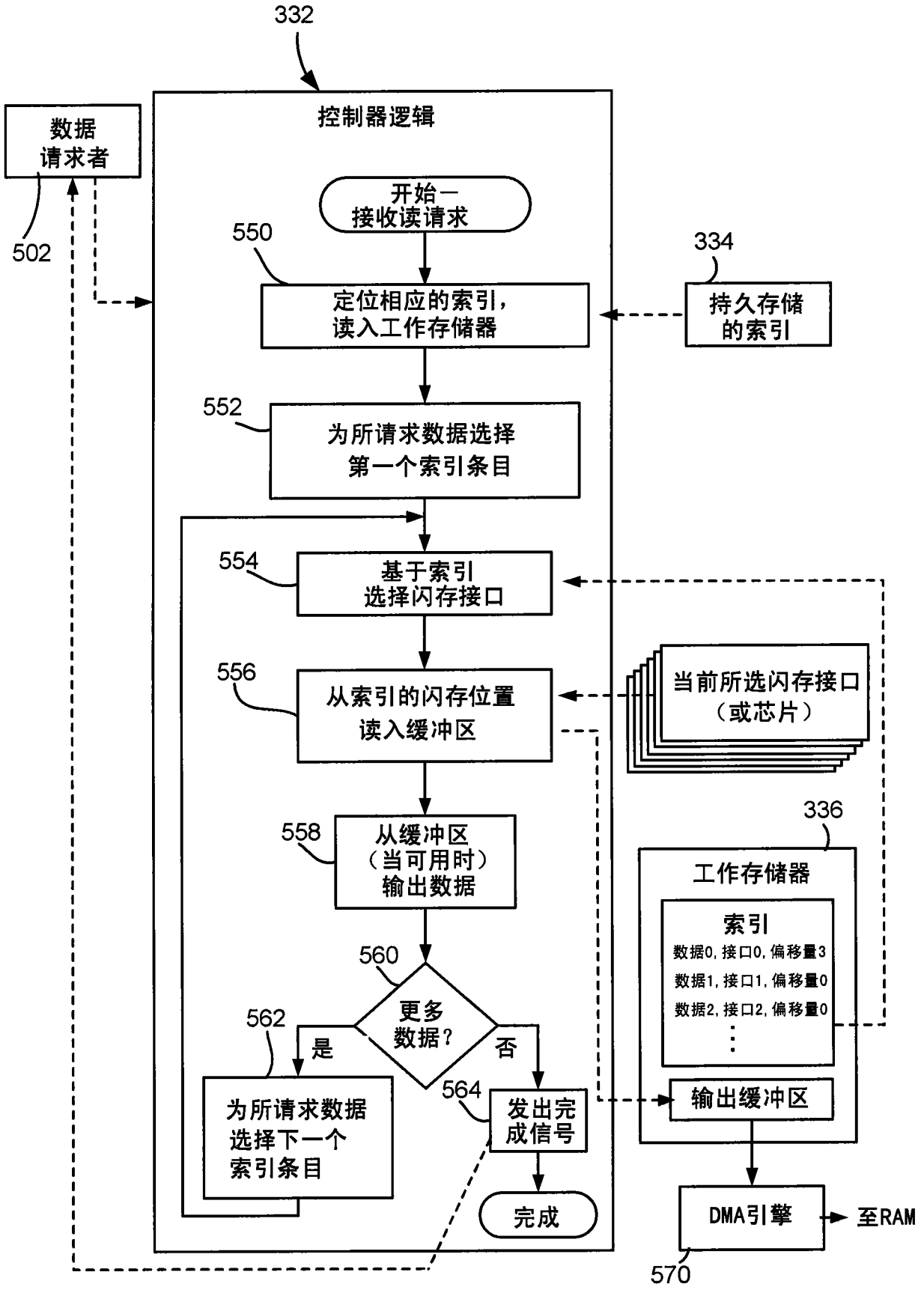

332
控制器逻辑
数据
请求者
502
开始—
接收读请求
550
定位相应的索引，
读入工作存储器
334
持久存储
的索引
552
为所请求数据选择
第一个索引条目
554
基于索引
选择闪存接口
556
从索引的闪存位置
读入缓冲区
当前所选闪存接口
（或芯片）
558
从缓冲区
（当可用时）
输出数据
336
工作存储器
索引
数据0，接口0，偏移量3
数据1，接口1，偏移量0
数据2，接口2，偏移量0
560
更多
数据？
562
是
否
为所请求数据
选择下一个
索引条目
564
发出完
成信号
输出缓冲区
完成
DMA引擎
至RAM
570

图 5