发明专利申请

申请人 HGST 荷兰公司
地址 荷兰阿姆斯特丹

发明人 M・G・卡蒂布 D・C・D・勒莫尔

专利代理机构 北京市柳沈律师事务所
11105
代理人 邱万奎

Int. Cl.
G06F 3/06(2006.01)

发明名称
数据存储设备的性能感知功率封顶控制

摘要

本文提供了用于数据存储设备的功率控制的系统、方法和系统。在一个示例中，提供了数据存储设备。该数据存储设备包括存储控制系统，以标识该数据存储设备的功率阈值。该数据存储设备确定该数据存储设备的功耗特性，并且进入该数据存储设备的功率控制模式。其至少调整该数据存储设备中的存储事务队列深度，以将该功耗特性建立为低于该数据存储设备的功率阈值。
1.一种用于操作主机系统以控制数据存储设备的功率使用的方法，该方法包括：
   标识该数据存储设备的功率阈值；
   确定该数据存储设备的功耗特性；以及
   进入该数据存储设备的功率控制模式，并且作为响应，至少调整该数据存储设备的存
   储事务队列深度，以将功耗特性建立为低于该数据存储设备的功率阈值。

2. 根据权利要求1的方法，进一步包括：提供该数据存储设备的最大存储事务队列深
   度，以降低功耗增加为高于功耗阈值，同时该功耗特性被保持低于该功率阈值。

3. 根据权利要求1的方法，进一步包括：减少该存储事务队列深度，以将延迟降低至延
   迟阈值以下，同时该功耗特性被保持低于该功率阈值。

4. 根据权利要求1的方法，进一步包括：适应性调整该存储事务队列深度，由于实际
   功耗特性和该目标功耗特性的差值比例。

5. 根据权利要求1的方法，进一步包括：利用该数据存储设备中的功率测量系统测量功
   耗特性。

6. 根据权利要求1的方法，进一步包括：利用主机系统自动地和/或连续地传输来自功
   率测量系统的功耗特性。

7. 根据权利要求1的方法，进一步包括以下至少一个：配置采样率，建立测量周期，在功
   率测量系统的基于时间的功能测量和基于事务的功率测量之间选择。

8. 根据权利要求1的方法，进一步包括：通过主机系统控制该数据存储设备的功率测量
   系统的操作。

9. 根据权利要求1的方法，进一步包括：响应于来自主机系统的异步请求，将一组标识
   符写入该数据存储设备的存储介质上的预定位置，并测量该相关功耗。

10. 根据权利要求1的方法，进一步包括：确定存储事务队列的范围大小。

11. 一种用于控制数据存储设备中的功耗的主机系统，包括：
    存储控制系统，被配置为标识该数据存储设备的功率阈值；
    功率测量系统，被配置为监视该数据存储设备的功耗特性；以及
    该存储控制系统被配置为进入该数据存储设备的功率控制模式，并且作为响应，至少
    调整该数据存储设备中的存储事务队列深度，以将功耗特性建立为低于该数据存储设备的
    功率阈值。

12. 根据权利要求11的主机系统，进一步被配置为提供该数据存储设备的最大存储事
    务队列深度，以将吞吐量增加为高于吞吐量阈值，同时该功耗特性被保持低于功率阈值。

13. 根据权利要求11的主机系统，进一步被配置为减少该存储事务队列深度，以将延迟
    降低至延迟阈值以下，同时该功耗特性被保持低于该功率阈值。

14. 根据权利要求11的主机系统，进一步被配置为适应性调整该存储事务队列深度，以
    与该实际功耗特性和该目标功耗特性的差值比例。

15. 根据权利要求11的主机系统，进一步包括功率测量系统。

16. 根据权利要求15的功率测量系统，进一步被配置为利用主机系统自动地和/或连续
    地传输功耗特性。

17. 根据权利要求15的功率测量系统，进一步包括以下至少一个：配置采样率，可配置
    测量周期，基于时间的功能测量，以及基于事务的功能测量。
18. 根据权利要求15的功率测量系统，进一步被配置为接收并响应来自主机系统的控制命令。

19. 根据权利要求11的主机系统，进一步被配置为响应于来自主机系统的异步请求，将一组标识符写入数据存储设备的存储介质上的预定位置，并测量相关功耗。

20. 根据权利要求11的主机系统，进一步被配置为确定该存储事务队列的范围大小。
数据存储设备的性能感知功率封顶控制

背景技术
[0001] 计算机和网络系统诸如个人计算机、工作站、服务器系统和云存储系统，典型地包括数据存储系统，用于存储和检索数据。这些数据存储系统可包括数据存储设备，诸如硬盘驱动器、固态存储设备、磁带存储设备和其他大容量存储设备。
[0002] 近年来对数字内容和大数据的需求促进了对更多、更密集和更大的数据中心的需要。数据中心的功率消耗至关重要的。数据中心频繁地被其布局中的允许峰值功率所约束。峰值功率容量是重要的，因为功率调节和分布基础设施以及冷却系统必须基于其分配的峰值功率需求来确定大小。
[0003] 数据中心操作者过度润湿具有于在任何时刻最大化其利用率所允许的最大值更多的机器服务器和用户的数据中心电源。功率封顶被用于数据中心来设置包含于数据中心的多种计算系统中最大功率使用率。然而，计算系统的各个组件的功率消耗限于非常粗糙的调整，诸如未使用的计算元件的睡眠模式或电源关闭。

发明内容
[0004] 本文提供了一种用于数据存储设备的功率控制的系统、方法和固件。在一个示例中，提供了数据存储设备。该数据存储设备包括存储控制系统，以标识数据存储设备的功率阈值。数据存储设备确定数据存储设备的功耗特性并且进入数据存储设备的功率控制模式，其至少调整数据存储设备中的存储事项列深度，以将功耗特性建立为低于数据存储设备的功率阈值。

附图说明
[0005] 图1说明在数据存储设备中用于性能感知功率封顶（performance-aware power capping）的存储系统。
[0006] 图2说明操作用于在数据存储设备中性能感知功率封顶的存储系统的方法。
[0007] 图3A说明操作在数据存储设备中性能感知功率封顶的存储系统的方法。
[0008] 图3B说明操作在数据存储设备中性能感知功率封顶的存储系统的方法。
[0009] 图4说明在数据存储设备中性能感知功率封顶的存储系统的可选实施例。
[0010] 图5说明操作用于在数据存储设备中性能感知功率封顶的存储系统的方法。
[0011] 图6说明操作用于在数据存储设备中性能感知功率封顶的存储系统的方法。
[0012] 图7说明用于在数据存储设备中性能感知功率封顶的数据存储设备。
[0013] 图8说明用于在数据存储设备中性能感知功率封顶的体系结构。
[0014] 图9说明用于在数据存储设备中性能感知功率封顶的计算系统。

具体实施方式
[0015] 数据存储系统应用多种大容量存储设备，诸如硬盘驱动器、固态驱动器，以及其他存储设备。然而，这些存储设备可使用高级别的功率，其可导致由聚集多个数据存储系统的
数据中心的过度功耗。在以下示例讨论中，多种方法，系统，装置和原件允许在用于数据服务器和数据存储系统的多种大容量存储设备中对功耗的精细化控制。例如，硬盘驱动器可置于功率封顶模式，其修改硬盘驱动器的参数以使存储操作低于该硬盘驱动器的功率阈值。数据存储设备的功率封顶和功率控制的其他示例在下文讨论。

[0016] 如第一示例，图1被呈现。图1说明了用于在数据存储设备中性能感知功率封顶的存储系统100。存储系统100包括数据存储设备102和主机系统120。数据存储设备102和主机系统120通过链路130通信。数据存储设备102包括存储介质112，存储控制系统114，以及队列116。存储介质112包括任何物理介质，其可用于存储希望的信息并且可由指令执行系统访问，诸如存储控制系统114。存储控制系统114包括处理电路和存储器，用于访问存储介质112。队列116包括数据存储设备102的命令的数据结构，诸如读或写命令。主机系统120包括一个或多个计算和网络系统，诸如个人计算机，服务器或其他计算机和网络系统。在操作中，主机系统120传输功率阈值110和/或队列深度140至数据存储设备102。数据存储设备102从主机系统120接收功率阈值110和/或队列深度140。

[0017] 图2说明了操作用于在数据存储设备中性能感知功率封顶的存储系统100的方法。该操作可由数据存储设备102或由主机系统120独立地执行或由数据存储设备102和主机系统120同时执行。附带地，图2的操作在下文被参考。

[0018] 在图2的第一示例，数据存储设备102执行图2的操作，但应当理解，图2中的操作可由主机系统120执行，如上述。数据存储设备102标识数据存储设备102的功率阈值110 (202)。功率阈值110表示数据存储设备102的可允许的最大或最小功耗特性。功率阈值110以瓦特或功率或能量消耗的其他单位限制功耗阈值，其可基于数据存储设备102的电流和供电电压。数据存储设备102确定数据存储设备102的功耗特性(204)。功率特性可以多种方式定义。例如，功耗特性可表示在数据存储设备102的每个时间单位或每个事件中可允许的功率消耗。数据存储设备102进入数据存储设备102的功率控制模式(206)。

[0019] 功率控制模式为数据存储设备102的操作模式，其通过控制队列深度140经由数据存储设备102的功耗特性提供精细的控制度。两个可能的功率控制模式包括通过调整队列116获取的延迟模式和吞吐量模式。延迟模式通过减少队列深度140提供了快速响应时间。吞吐量模式通过增加队列深度140提供数据流的有效性能。作为响应，数据存储设备102至少调整数据存储设备102中的队列深度140 (208)。队列深度140可由存储控制系统114或主机系统120控制。延迟模式和吞吐量模式可通过控制队列深度140实现。数据存储设备102将功耗特性建立为低于数据存储设备102的功率阈值(210)。数据存储设备102监视数据存储设备102的功耗特性并且作为响应调整功率控制模式和队列深度140，以建立和保持希望的功耗特性。功耗特性监视可为连续或离散的。

[0020] 作为控制数据存储驱动的功耗的进一步的示例，图3A和图3B被呈现。图3A和图3B进一步说明了操作用于在数据存储设备中性能感知功率封顶的存储系统100的方法。图3A说明了由数据存储设备102执行的操作用于在数据存储设备中性能感知功率封顶的存储系统100的方法。图3B说明了由主机系统120执行的操作用于在数据存储设备中性能感知功率封顶的存储系统100的方法。

[0021] 图3A说明了由数据存储设备102执行的操作用于在数据存储设备中性能感知功率封顶的存储系统100的方法。主机系统120标识用于数据存储设备102的功率阈值110。主机
系统120将功率阈值110传输至数据存储设备102(302)。数据存储设备102确定功耗特性。数据存储设备102进入功率控制模式，并且作为响应，至少调整队列深度140。数据存储设备102将功耗特性建立为低于功率阈值。

【0022】图3B说明了由主机系统120执行的操作用于在数据存储设备中性能感知功率封顶的存储系统100的方法。主机系统120标识数据存储设备102的功率阈值110。主机系统120确定数据存储设备102的功耗特性。主机系统120进入功率控制模式，并且作为响应，至少调整队列深度140。主机系统120发送队列深度信息至数据存储设备102，并且将数据存储设备102的功耗特性建立为低于功率阈值110(304)。

【0023】回到图1，存储系统100的元件进一步如下描述。数据存储设备102可为硬盘驱动器（HDD）、固态存储驱动器（SSD）、闪存驱动器，相变存储器驱动器、光存储驱动器或某些其他类型的存储设备，包括其组合和变型。数据存储设备102包括静态组件，诸如主机接口、处理电路、数据缓冲器、存储器和读/写头。

【0024】存储介质112包括任何物理介质，其可用于存储希望的信息并且可由指令执行系统访问，诸如存储控制系统114。存储介质112的示例包括随机存取存储器，只读存储器，磁盘，光盘，闪存，相变存储器，虚拟存储器和非虚拟存储器，磁盘盒，磁性带，磁盘存储或其他磁存储设备，或可用于存储希望的信息并可由指令执行系统访问的任何其他介质，以及任何其组合或变型或其他任何类型的存储介质。

【0025】存储控制系统114包括控制电路和存储器，用于访问存储介质112。存储控制系统114其进一步包括处理电路，存储器，主机接口和缓冲器。存储器还包括固件，其至少由存储控制系统114执行，以操作并且响应通过链路130接收的读和写命令。

【0026】主机系统120可包括一个或多个计算和网络系统，诸如个人计算机，服务器计算机，机架服务器，网络服务器，云计算平台，报文网络，管理系统和数据中心装置，以及其他任何类型的物理或虚拟服务器机器和其任何变型或组合。在操作中，主机系统120通过链路330发送读和写命令至数据存储设备102。在进一步示例中，主机系统120可发送数据存储设备102的一个或多个命令以进入功率控制模式。

【0027】链路130可包括一个或多个链路。链路130包括金属的，无线的，光学的，或包括其组合的某些其他网络链路。链路130可包括任何存储或磁盘接口，诸如串行附接ATA(SATA)、串行附接SCSI(SAS)、光纤通道，通用串行总线(USB)、SCSI、InfiniBand、外围组件高速互联(PCIe)、以太网，互联网协议(IP)或其他并行或串行存储或外围接口，包括变型及其组合。链路130利用多种通信和控制协议，诸如LTE，GSM，CDMA，互联网协议(IP)，以太网，会话发起协议(SIP)，直径(Diameter)，实时传输协议(RTP)和/或某些其他格式，包括其组合。

【0028】图4说明了用于在数据存储设备中性能感知功率封顶的存储系统400。存储系统400为存储系统100的示例，但存储系统100可具有可选的配置和操作。存储系统400包括数据存储设备402和主机系统420。数据存储设备402和主机系统420经由链路430通信。

【0029】数据存储设备402包括存储介质412，存储控制系统414，队列416，以及功率测量系统418。数据存储设备402可为硬盘驱动器，可移除磁盘存储驱动器，固态驱动器，磁带驱动器，或具有队列深度影响的功耗特性的某些其他存储设备。

【0030】存储控制系统414包括处理电路和存储器。处理电路可包括一个或多个微处理器和其他电路，其从存储器取回并执行固件。处理电路可实现于单个处理设备，还分布于多
个处理设备或子系统，其协同执行程序指令。处理电路的示例包括通用中央处理单元、专用处理模块和逻辑设备，以及其他任何类型的处理设备，组合或其变形。存储控制系统414可增加队列深度440以将吞吐量增加至吞吐量阈值之上，同时功耗特性被保持低于功率阈值410。可选地，存储控制系统414可减少队列深度440以将延迟降低至延迟阈值以下，同时功耗特性被保持低于功率阈值。存储控制系统414可配置采样率，建立测量周期，并且在基于时间的功率测量和基于事务的功率测量中选择。主机系统420能够执行与存储控制系统414相同的操作。

[0031] 存储介质412包括随机存取存储器、只读存储器、磁盘、光盘、闪存、相变存储器、磁性带或其他存储介质，包括其组合和变形。在多个示例中，数据存储设备402包括用于读取和写入存储介质412的多种机械、电气和机电器件，诸如读/写头、电枢、光读/写元件、伺服、前端放大器、放大器或其他组件，包括其组合。

[0032] 功率测量系统418至少测量数据存储设备402使用的功率。功率测量系统可立即，通过指定时间段或在每个数据存储设备402中进行功率测量。数据存储设备402的功率测量系统418的操作可由主机系统420控制。功率测量系统可连续地和周期性地与主机系统420相通信，用于功率测量系统418设置。

[0033] 主机系统420包括处理器424和存储器426。主机系统420可包括一个或多个计算和网络系统，诸如个人计算机、服务器、云存储系统、报文网络、管理系统或其他计算机和网络系统，包括其组合和变形。在操作中，主机系统420通过链路430发送读和写命令至数据存储设备402。在某些示例中，主机系统420发送命令至数据存储设备402。例如，主机系统420可使用或禁止功率测量系统418，以及发送测量命令至功率测量系统418。主机系统420还可命令数据存储设备进入功率控制模式。主机系统420还能够调整队列深度440，以建立数据存储设备402的希望的功耗特性。主机系统420传输功率阈值410和/或队列深度440至数据存储设备402。主机系统配置采样率，建立测量周期，并在基于时间的功率测量和基于事务的功率测量之间选择。

[0034] 处理器424包括一个或多个微处理器和其他电路，其从存储器取回并执行指令。处理电路可实现于单个处理设备，还可分布于多个处理设备或子系统，其协同执行程序指令。处理电路的示例包括通用中央处理单元、专用处理器和逻辑设备，以及其他任何类型的处理设备，其组合或变形。

[0035] 存储器426包括文件系统428。存储器426的示例包括随机存取存储器、只读存储器、磁盘、光盘、闪存、虚拟存储器和非虚拟存储器、磁带盒、磁性带、磁盘存储或其他磁存储设备，或其他任何可用于存储信息的介质并可由指令执行系统访问的介质，以及任何组合或其变形或其他任何类型的存储介质。

[0036] 文件系统428可为磁盘文件系统、光盘文件系统、闪存文件系统、磁带文件系统、事务性文件系统、联机文件系统、共享磁盘文件系统、设备文件系统和其他文件系统。磁盘文件系统可包括FAT（FAT12, FAT16, FAT32）、exFAT、NTFS、HFS, HFS+, HPFS、UFS、ext2, ext4, XFS, btrfs, ISO 9660, vfat, replace文件系统、VMFS、ZFS, ReiserFS, UDF或某些其他文件系统。光盘文件系统可包括ISO 9660,通用盘格式（UDF）、Mount Rainier或某些其他光盘文件系统。磁带文件系统可包括IBM Linear Tape文件系统-单机驱动版（LTFS-SDE）或某些其他磁带文件系统。事务性文件系统可包括事务性NTFS、Valor、Amino、LFS, 事务性ext3或
某些其他事务性文件系统。某些联机文件系统的示例可包括NFS，AFS，SMD协议客户端，以及
类文件系统客户端，用于FTP和WEBDAV。某些共享磁盘文件系统的示例包括GFS2，GPFS，SFS，
CXFS，StorNext或其他某些其他共享磁盘文件系统。

功率阈值410表示数据存储设备402的可允许最大或最小功耗特性。功率阈值410
以瓦特或功率或能量消耗的其他单位指定功耗阈值。功率阈值410可基于时间或事务限制
功耗。例如，功率阈值410可指示数据存储设备402可在特定时间段内使用峰值功率量。
可选地，功率阈值410可指定数据存储设备402限制功耗至每个事务的指定功率量。功率阈
值410可连续或可变或恒定。

队列深度440可被调整以建立数据存储设备402的逻辑功耗特性。数据存储设备
402被配置以进入功率控制模式，并且作为响应，至少调整队列深度440，以将功耗特性建立
为低于功率阈值410。功率控制模式包括吞吐量模式和延迟模式。吞吐量功率控制模式增加
队列深度440，以导致高于吞吐量阈值的高吞吐量，同时功耗特性被保持低于功率阈值。延
迟功率控制模式减少了队列深度440，以降低延迟至延迟阈值以下，同时功耗特性保持低于功
耗阈值。

尽管图4示出了单个链路，链路430可包括一个或多个链路。链路可包括任何存储
或磁盘接口，诸如串行附接ATA（SATA），串行附接SCSI（SAS），光纤通道，通用串行总线
（USB），SCSI，InfiniBand，外部组件高速互联（PCIe），以太网，互联网协议（IP）或其
他并行或串行存储或外围接口，包括其变型及组合。

图5说明了用于在数据存储设备中性能感知功率封顶的示例性功率控制模式。数
据存储设备402基于操作模式具有不同的或功耗特性。队列416的操作可建立包括功率控制模
式的不同的操作模式。某些功率控制模式的示例包括吞吐量模式和延迟模式。例如，增加队
列深度440将牺牲延迟性能来改进数据存储设备402的吞吐量性能。可选地，减少队列深度
440将牺牲吞吐量性能来改进数据存储设备402的延迟。这些功率控制模式可用于调整数据
存储设备402的功耗特性，以使数据存储设备402不比功率阈值410更多的功率。

图5包括与队列416共同工作的数据存储设备402的正常操作模式的说明。对于该
示例，数据存储设备102为硬盘驱动器。虽然未示出，硬盘驱动器典型地包括伺服，用于移动
固定读/写头的电机。当在存储介质412表面定位读/写头至不同的角度位置时，伺服/电枢
机构消耗功率。

在正常操作模式中，读/写头开始于初始位置“A”，并且必须以第一方向行进以到
达第二位置“B”，然后以相反方向移动以到达第三位置“C”。注意读/写头位置未相对于时间
有效控制，并且使用了比所需更多的功率。队列416不具有足够的深度，以由存储控制系统
112或主机系统120操作来用于多个有效读/写头定位。功耗特性因此关联于队列深度440。

图5包括描述用于在数据存储设备中性能感知功率封顶的数据存储设备402的吞
吐量模式的说明。吞吐量模式通过增加队列深度440实现。当队列深度440足够深时，正在进行
的请求可被组织，用于更多的有效读写次数，如图5的吞吐量模式所示。增加队列深度440
允许正在进行的请求在队列416中组织，用于更有效的读/写头定位。注意，以吞吐量模式，
相同的请求比在正常操作模式中更快地处理。如果队列416深度较浅，如在正常操作模式，
则存储控制系统414或主机系统120不具有资源以吞吐量模式的有效顺序来设置正在进行
的请求。队列416将被强制操作于类似于正常操作模式的模式。
图5说明了用于在数据存储设备中性能感知功率封顶的数据存储设备402的延迟模式。延迟可包括事务存储于队列416的时间，直到其由读/写头读取或写入存储介质112。读取延迟包括读取命令由主机420发送的时间和数据响应于读取返回主机系统所花费的时间。写入延迟类似，不同之处在于端点为写入头的数据提交到存储介质112和表示写入完成的信号。较大队列深度440增加延迟，因为读/写命令将必须等待队列416中要完成的正在进行的请求。如果数据存储设备102或主机系统120建立小队列深度440，则延迟被降低，因为存储控制系统414或主机系统420不需要等待较大数据的要完成的正在进行的请求。

图6说明了操作用于操作于在数据存储设备402中性能感知功率封顶的存储系统400的方法。操作可通过数据存储设备402或主机系统420独立地或通过数据存储设备402和主机系统420同时执行。附带地，图6的操作在下文参考。

在图6的示例中，数据存储设备402执行图6的操作，但应当理解，图6中的操作可执行于主机系统120，如上述。数据存储设备402标识数据存储设备402的功率阈值410（602）。功率阈值410表示数据存储设备402可允许的最大或最小功耗特性。功率阈值410以千瓦或功率或能量消耗的其他单位标识功耗阈值。数据存储402配置采样率，建立测量周期，以及在基于时间的或基于事务的功率测量中选择，以确定功耗特性（604）。功耗特性可利用多种方式测量。数据存储设备402和主机系统420能够配置采样率，建立测量周期，并且在基于时间的或基于事务的功率测量中选择，提供高度的灵活性和通过功率测量的控制。在某些示例中，数据存储设备402包括功率测量系统418。数据存储设备402从功率测量系统418确定数据存储设备402的功耗特性（606）。功耗特性可以多种方式定义。例如，功耗特性可以表示数据存储设备402的每个单位或时间或每事务的最大可允许的功率消耗。进入数据存储设备402的功率控制模式，以将功耗特性建立为低于功率阈值410（608）。功率控制模式为数据存储设备402的操作模式，其通过控制队列深度440提供数据存储设备402的功耗特性的精细的控制度。

吞吐量模式和延迟模式为存储设备在被功率控制时可工作的两个面向性能的模式。本文这里讨论这些模式。在吞吐量模式，存储控制系统414或主机系统420提供最大队列深度440，以将吞吐量增加至吞吐量阈值以上，同时功耗特性被保持低于功率阈值（610）。在吞吐量模式，存储控制系统414或主机系统420提供队列深度440，以降低延迟至延迟阈值以下，同时功耗特性被保持低于功率阈值（612）。数据存储设备402和主机系统420可在吞吐量模式和延迟模式之间变化，以实现数据存储设备402的功耗特性。存储控制系统414或主机系统420可将功耗特性提供给主机系统420。在主机系统420控制功率测量系统418的操作（618）。功率测量系统418被配置以从主机系统接受命令。队列（620）中的大小范围可被设置，代替单个队列深度。

图7说明了用于在数据存储设备中性能感知功率封顶的硬盘驱动器702。硬盘驱动器702为数据存储设备102和数据存储设备402的示例，但数据存储设备102和数据存储设备402可具有可选的配置和操作方法。硬盘驱动器702包括存储介质712、队列716和功率测量系统718。存储介质712利用标识符ID1-1D6序列说明。在操作中，硬盘驱动器702的功耗特性
可通过对响应于来自主机系统的访问请求将一组标识符（由ID1-ID6表示）写入存储介质712的预定位置并测量相关性能来确定。

图8说明了用于在数据存储设备中性能感知功率封顶的存储系统800。存储系统800为存储系统100和存储系统400的示例，但存储系统100和存储系统400可以具有可选的配置和操作的方法。存储系统800包括数据存储设备802，文件系统804和块I/O调度器806。数据存储设备802可为硬盘驱动器、固态存储驱动器、闪存驱动器或某些其他类型的存储设备，包括其组合和变型。

文件系统802可为磁盘文件系统、光盘文件系统、闪存文件系统、磁带文件系统、事务性文件系统、网络文件系统、共享磁盘文件系统，设备文件系统和其他文件系统。磁盘文件系统可包括FAT(FAT12,FAT16,FAT32), exFAT, NTFS, HFS, HFS+, HPFS, UFS, ext2, ext4, XFS,btrfs, ISO 9660, File4, 1, Verita文件系统, VMFS, ZFS, ReiserFS, UDF或某些其他文件系统。光盘文件系统可包括ISO 9660, 通用磁盘格式(UDF), Mount Rainier或某些其他光盘文件系统。磁带文件系统可包括IBM线性磁带文件系统-单机驱动版(LTFS-SDE)或某些其他磁带文件系统。事务性文件系统可包括事务性NTFS, Valor, Amino, LFS, 事务性ext3或某些其他事务性文件系统。某些联网文件系统的示例可包括NFS, AFS, SMD协议的客户端和类似文件系统客户端，用于FTP和WebDAV。某些共享磁盘文件系统的示例包括GFS2, GPFS, SFS, CXFS, StorNext或某些其他共享磁盘文件系统。文件系统804包括功率封顶块I/O接口模块808。

功率封顶块I/O接口808包括功率封顶监视器812，性能监视器814，以及I/O执行控制816。性能监视器814与I/O执行控制816和功率封顶监视器812相通信。在操作中，功率封顶监视器812标识功率阈值810，并将功率阈值810传输至性能监视器814。性能监视器814基于功率阈值810选择并执行数据存储设备802的功率控制模式。性能监视器814发送命令至I/O执行控制816。I/O执行控制816提供反馈至功率封顶监视器812，并且性能监视器814提供精细化调整系统操作的能力。

I/O执行控制816基于接收自性能监视器814的命令来控制块I/O请求818。I/O执行控制816将反馈传输至性能监视器814和功率封顶监视器812，以允许对块I/O请求812的精微调整。I/O执行控制816对块I/O调度器806发送指令，以调整队列深度并且优化块I/O请求。块I/O调度器806传输命令和数据至存储设备802。数据存储设备802提供反馈信息至功率封顶模块。

图9说明了计算系统902，其表示任何系统或系统集合，其中本文公开的多种操作系统体系结构、环境和过程可被实现。计算系统902是存储控制系统114,主机系统120,存储控制系统414,主机系统420的示例，但这些示例可具有可选的配置和/或操作。计算系统902的示例包括但不限于：智能电话，膝上型计算机，平板计算机，桌面计算机，混合计算机，游戏机，虚拟机，智能电视，手表和其他可穿戴设备，以及其任何变型或组合。其他示例包括服务器计算机器，机架服务器，网络服务器，云计算平台和数据中心设备，以及其他任何类型的物理或虚拟服务器机器和其任何变型或组合。

计算系统902可实现为单个装置、系统或设备，或可以分布式方式实现多个装置、系统或设备。计算系统902包括通信收发器901和处理系统903。处理系统903包括处理器电路905和存储器系统906。存储器系统包括存储单元907，其包括写入模式模块908、延迟模式模
块909，以及功耗模块910。

【0055】处理系统903在处理器电路905中加载并执行来自存储器系统906的软件907。当由用于数据存储设备的性能感知功率封顶的处理系统905执行时，软件907指示处理电路905按照本文描述的在前讨论的数据存储设备的性能感知功率封顶实现的至少多种过程，操作环境和顺序操作。计算系统902可选地包括附加的设备，特征或功能，为简洁起见不再讨论。

【0056】处理电路905可包括微处理器和其他电路，其从存储器906取回并执行软件907。处理电路905可实现于单个处理器设备，还可分布于多个处理器设备或子系统，其协同执行程序指令。处理电路905的示例包括通用中央处理单元，专用处理器和逻辑设备，以及其他任何类型的处理设备，组合或其变形。

【0057】存储器系统906可包括任何非暂存计算机可读存储介质，由处理系统905可读并且能够存储软件907。存储器系统906可包括易失性和非易失性，可移除和非可移除介质，实现于任何用于信息存储的方法或技术，诸如计算机可读指令，数据结构，程序模块或其他数据。存储介质的示例包括随机存取存储器，只读存储器，磁盘，光盘，闪存，虚拟存储器和非虚拟存储器。存储介质可以被配用于存储其他信息，包括但不限于存储介质和传输介质。存储介质可以是可写入的或不可写入的。

【0058】除计算机可读存储介质外，在某些实现中，存储器系统906还可包括计算机可读通信介质。软件907可通过其进行内部或外部通信。存储器系统906可实现为单个存储设备，但还可实现于共同定位或相对于彼此分布的多个存储设备或子系统中。存储器系统906可包括附加的元件，诸如控制器，能够利用处理系统903或可能的其他系统进行通信。

【0059】软件907可实现于程序指令以及其他功能，可当由处理系统903执行时，指示处理系统903根据描述操作于本文说明的多种操作环境，序列和过程。例如，软件907可包括程序指令，用于实现代码服务和其相关功能。

【0060】特别地，程序指令可包括多种组件或模块，其协同操作或以其他方式交互，以执行本文描述的多种过程和操作环境。多种组件或模块可以编译或解释指令或某些其他指令变形或组合来实现。多种组件或模块可执行于同步或异步方式，串行或并行，单线程环境或多线程环境。软件907可包括附加的过程，程序或组件，诸如操作系统软件或其它应用软件。软件907还可包括可由处理系统903执行的固件或某些其他形式的机器可读处理指令。

【0061】总体上，软件907可当载入处理系统903并且由处理电路905执行时，总体上从通用计算机系统转换合适的装置，系统或设备（表示其计算机系统902）至定做的专用计算机系统，以实现增强组件互操作性。事实上，存储器系统906的编码软件907可转换存储器系统906的物理结构。物理结构的特定转换可取决于所描述的不同实现中的多种因素。这些因素的示例可包括但不限于：用于实现存储器系统906的存储介质的技术，以及计算机存储介质是否特定在于主存储或辅存储，以及其他因素。

【0062】例如，如果计算机可读存储介质实现为基于半导体的存储器，软件907可当程序指令被编码其中时，转换半导体存储器的物理状态，诸如通过转换组成半导体存储器的晶体管，电容器或其他离散的电路元件的状态。类似转换可相对于磁或光介质产生。物理介质的其他转换是可能的，而不偏离本说明书的范围。在前的示例仅被提供以实现本讨论。

【0063】在至少一个实现中，程序指令可包括吞吐量模式模块908，延迟模式模块909，以及功耗模块910。吞吐量模式模块908包括建立存储驱动的队列深度，以封顶存储驱动的功耗，
同时保持希望的能率。延迟模式模块909包括建立存储驱动的队列深度，以封锁存储驱动的功耗，同时保持存储于存储驱动的数据的希望的读/写延迟。功耗模块910测量和监视存储驱动的功耗。功耗模块910可从外部系统接收功率测量数据，诸如实现于存储驱动的功率监视电路。功率测量数据可包括存储驱动当前功率的数字化表示。功耗模块910可通过通信收发器901接收功率测量数据。功耗模块910可传输功率测量数据至吞吐量模式模块908，延迟模块909的任一个，用于确定队列深度，其建立存储驱动的操作的吞吐率模式或延迟模式。

【0064】 可以理解，计算系统902总体上意图表示计算系统或软件907可部署其中并且执行以实现增强组件间操作性的系统。然而，计算系统902还可适用于任何计算系统，其中软件907可被展现，并且一个或两者可被分发、传输、下载或以其他形式提供至另一计算系统用于部署和执行或附加的分发。

【0065】通信收发器901可包括通信连接和设备，其允许通过通信网络(未示出)与其他计算系统(未示出)的通信。同时允许系统同通信的连接和设备的示例可包括网络接口卡、天线、功率放大器、RF电路、收发器和其他通信电路。连接和设备可通过通信介质通信以与其他计算系统或网络交换通信，诸如金属、玻璃、空气或其他任何合适的通信介质。前述介质、连接和设备为公知，不需要在此过多讨论。

【0066】用户界面系统911是选的，并可包括键盘、鼠标、语音输入设备，用于从用户接收触摸手世的触摸输入设备、用于检查非触摸手世和用户其他运动的运动输入设备、和和或其他可比较输入设备和能够从用户接受输入的相关处理元件。输出设备诸如显示器、扬声器、触觉设备和其他类型的输出设备还可包括于用户界面系统911。在某些情况下，输入和输出设备可合并于单个设备，诸如能够显示图像和接收触摸手世的显示器。前述用户输入和输出设备为本领域公知，不需要在此过多讨论。

【0067】用户界面系统911还可包括可由处理器电路905在以上讨论的多种用户输入和输出设备的支持下执行的相关用户界面软件。独立地或相互和其他硬件和软件元件结合，用户界面软件和用户界面设备可支持图形用户界面、自然用户界面或其他任何类型的用户界面。

【0068】在计算系统902和其他计算系统(未示出)之间的通信，可通过通信网络或网络根据多种通信协议、协议的组合或其变型发生。示例包括内联网、互联网、因特网、局域网、广域网、无线网络、有线网络、虚拟网络、软件定义网络、数据中心总线、计算背板或其他任何类型的网络、网络的组合或其变型。前述通信网络和协议为公知，不需要在此过多讨论。然而，可使用的某些通信协议包括但不限于：互联网协议(IP，IPv4，IPv6等)，传输控制协议(TCP)，以及用户数据报协议(UDP)，以及任何其他合适的通信协议，其变型或组合。

【0069】在数据、内容或其他任何类型的信息被交换的任何前述示例中，信息交换可根据任何多种协议发生，包括FTP(文件传输协议)，HTTP(超文本传输协议)，REST(表示状态传输)，WebSocket，DOM(文档对象模型)、HTML(超文本标记语言)、CSS(级联样式表)、HTML5、XML(可扩展标记语言)、JavaScript、JSON(JavaScript对象表示)和AJAX(异步JavaScript和XML)，以及其他任何合适的协议，其变型或组合。

【0070】所包括的描述和附图描述了具体实施例以教导本领域技术人员如何制造和使用最佳模式。为教导创造性原理，某些传统方面被简化或省略。本领域技术人员将理解，来自
这实施例的变型属于本发明的范围。本领域技术人员将还理解，上述特征可以多种方式合并以形成多个实施例。结果，本发明不被上述具体实施例限制，而是仅通过权利要求和其等价物限定。
图1
标识数据存储设备的功率阈值

确定数据存储设备的功耗特性

进入数据存储设备的功率控制模式

作为响应，至少调整数据存储设备的存储事务队列深度

将功耗特性建立为低于数据存储设备的功率阈值

图2
图3A

图3B
标识数据存储设备的功率阈值

配置采样率，建立测量周期，以及在基于时间或基于事务的功率测量中选择

从功率测量系统确定数据存储设备的功率特性

进入数据存储设备的功率控制模式，以将功率特性建立为低于数据存储设备的功率阈值

提供最大队列深度，以将吞吐量增加为吞吐量阈值以上，同时功耗特性保持低于功率阈值

降低队列深度，以减少延迟至低于延迟阈值，同时功耗特性保持低于功率阈值

适应性调整队列深度，以与实际功耗特性和目标功耗特性的差成比例

利用主机系统自动和/或连续传输来自功率测量系统的功率特性

通过主机系统控制功率测量系统的操作

图6
图7
图9