(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 103026317 A
(43)申请公布日 2013.04.03

(21)申请号 201080068348.4

(22)申请日 2010.07.30

(85)PCT申请进入国家阶段日
2013.01.30

(86)PCT申请的申请数据
PCT/US2010/043816 2010.07.30

(87)PCT申请的公布数据
WO2012/015418 EN 2012.02.02

(71)申请人 惠普发展公司,有限责任合伙企业
地址 美国德克萨斯州

(72)发明人 A.A. 纳塔拉简

(74)专利代理机构 中国专利代理(香港)有限公司 72001
代理人 马红梅 傅康

(51)Int.Cl.
G06F 1/32(2006.01)
G06F 13/14(2006.01)

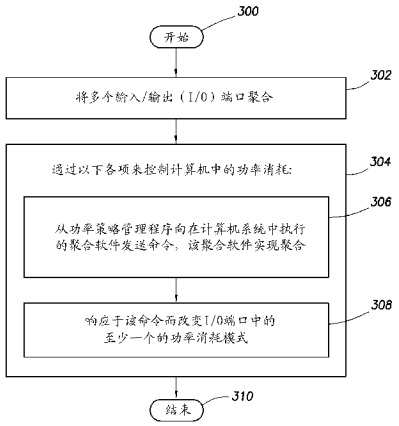
权利要求书 3 页 说明书 8 页 附图 3 页

(54)发明名称

控制聚合 I/O 端口的功率消耗的方法和系统

(57)摘要

控制聚合 I/O 端口的功率消耗。说明性实施例中的至少一些是包括以下各项的方法:将多个输入/输出(I/O)端口聚合;以及控制计算机系统中的功率消耗。控制功率消耗包括:从功率策略管理程序向在计算机系统中执行的聚合软件发送命令,该聚合软件实现聚合;以及响应于该命令而改变 I/O 端口中的至少一个的功率消耗模式。



CN 103026317 A

1. 一种方法，其包括：

将计算机系统的多个输入 / 输出（I/O）聚合；

通过以下各项来控制计算机系统中的功率消耗：

从功率策略管理程序向在计算机系统中执行的聚合软件发送命令，所述聚合软件实现所述聚合；以及

响应于所述命令改变所述多个 I/O 端口中的至少一个的功率消耗模式。

2. 根据权利要求 1 所述的方法，其还包括：

其中，发送所述命令还包括发送用以降低功率消耗的命令；以及

其中，改变所述功率消耗模式还包括将所述多个 I/O 端口的第一 I/O 端口的功率消耗模式从具有第一峰值功率状态的第一功率消耗模式变成具有低于所述第一峰值功率消耗的第二峰值功率状态的第二功率消耗模式。

3. 根据权利要求 1 所述的方法，其还包括：

将所述多个 I/O 端口中的第一 I/O 端口作为主端口来操作，并且将所述多个 I/O 端口中的第二和第三 I/O 端口作为热待机端口来操作；以及

其中，改变所述功率消耗模式还包括：

将所述第三 I/O 端口的功率消耗模式从具有第一峰值功率状态的第一功率消耗模式变成具有低于所述第一峰值功率消耗的第二峰值功率状态的第二功率消耗模式。

4. 根据权利要求 3 所述的方法，其还包括：

确定所述第一 I/O 端口已经经历故障；以及然后

将所述第二 I/O 端口设置为主端口；以及

将所述第三 I/O 端口的功率消耗模式变成所述第一功率消耗模式。

5. 根据权利要求 1 所述的方法，其还包括：

操作所述多个 I/O 端口，使得每个 I/O 端口参与同所述计算机系统耦合到的网络的通信；以及

其中，改变所述功率消耗模式还包括：

将所述多个 I/O 端口的功率消耗模式从具有用于每个 I/O 端口的第一峰值功率状态的第一功率消耗模式变成具有与所述第一峰值功率消耗不同的用于每个 I/O 端口的第二峰值功率状态的第二峰值消耗模式。

6. 根据权利要求 5 所述的方法，其中，改变所述 I/O 端口的功率消耗模式还包括选自由以下各项组成的组的至少一个：变成具有比所述第一峰值功率消耗模式低的峰值功率状态的所述第二峰值功率消耗模式；以及变成具有比所述第一峰值功率消耗模式高的峰值功率状态的所述第二峰值功率消耗模式。

7. 根据权利要求 1 所述的方法，其中，聚合还包括将是选自由以下各项组成的组的至少一个的所述多个 I/O 端口聚合：适配器端口，其被配置成跨因特网与设备通信；以及适配器端口，其被配置成与存储设备通信。

8. 一种计算机系统，其包括：

处理器；

多个输入 / 输出（I/O）适配器，其被耦合到处理器，所述 I/O 适配器被配置成耦合到网络；

存储器，其被耦合到处理器，所述存储器存储指令，所述指令在被处理器执行时促使处理器：

实现软件栈；

将所述多个 I/O 端口聚合；

从功率策略管理程序向执行所述多个 I/O 端口的聚合的聚合软件递送命令；以及

响应于所述命令改变所述多个 I/O 端口中的至少一个的功率消耗模式。

9. 根据权利要求 8 所述的计算机系统，其还包括：

其中，当所述处理器递送命令时，所述指令还促使处理器递送命令以降低功率消耗模式；以及

其中，当所述处理器改变功率消耗模式时，所述指令还促使处理器将所述多个 I/O 端口中的第一 I/O 端口的功率消耗模式从具有第一峰值功率状态的第一功率消耗模式变成具有低于所述第一峰值功率状态的第二峰值功率状态的第二功率消耗模式。

10. 根据权利要求 8 所述的计算机系统，其还包括：

其中，在功率消耗模式变化之前，所述计算机系统将所述多个 I/O 端口中的第一 I/O 端口作为主端口来操作，将所述多个 I/O 端口中的第二 I/O 端口作为热待机端口来操作，并且将所述多个 I/O 端口的其余 I/O 端口作为热待机端口来操作；以及

其中，当所述处理器改变功率消耗模式时，所述指令促使处理器：

将其余 I/O 端口的功率消耗模式从具有第一峰值功率状态的第一功率消耗模式变成具有低于所述第一峰值功率状态的第二峰值功率状态的第二功率消耗模式。

11. 根据权利要求 10 所述的计算机系统，其中，所述聚合软件的指令还促使所述处理器：

确定所述第一 I/O 端口已经经历故障；以及然后

将所述第二 I/O 端口设置为主端口；以及

将所述多个 I/O 端口中的第三 I/O 端口的功率消耗模式变成所述第一功率消耗模式。

12. 根据权利要求 8 所述的计算机系统，其还包括：

其中，在功率消耗模式变化之前，所述计算机系统操作所述多个 I/O 端口，使得每个 I/O 端口参与同网络的通信；以及

其中，当所述处理器改变功率消耗模式时，所述指令促使处理器：

将所述多个 I/O 端口的功率消耗模式从具有用于每个 I/O 端口的第一峰值功率状态的第一功率消耗模式变成具有用于每个 I/O 端口的第二峰值功率状态的第二功率消耗模式，所述第二峰值功率状态不同于所述第一峰值功率状态。

13. 根据权利要求 12 所述的计算机系统，其中，当所述处理器改变所述多个 I/O 端口的功率消耗模式时，所述指令促使所述处理器进行选自由以下各项组成的组的至少一个：变成第二功率消耗模式，其中，所述第二峰值功率状态低于所述第一峰值功率状态；以及变成所述第二峰值功率消耗模式，其中，所述第二峰值功率状态高于所述第一峰值功率状态。

14. 根据权利要求 8 所述的计算机系统，其中，所述多个 I/O 端口还包括选自由以下各项组成的组的至少一个：适配器端口，其被配置成跨因特网进行通信；以及适配器端口，其被配置成与存储设备通信。

15. 根据权利要求 8 所述的计算机系统，其中，所述存储器是选自由以下各项组成的组

的一个或多个：随机存取存储器（RAM）；只读存储器（ROM）；硬盘驱动器；以及光盘驱动器。

16. 一种存储指令的非暂时性计算机可读介质，所述指令在被处理器执行时促使处理器：

将被耦合到所述处理器的多个 I/O 端口聚合；

从功率策略管理程序接收命令，所述命令是关于用于所述多个 I/O 端口的功率策略；以及

响应于所述命令改变所述多个 I/O 端口中的至少一个的功率消耗模式。

17. 根据权利要求 16 所述的非暂时性计算机可读介质，其还包括：

其中，当处理器将所述多个 I/O 端口聚合时，所述指令促使所述处理器将所述多个 I/O 端口中的第一 I/O 端口作为主端口来操作；将所述多个 I/O 端口中的第二 I/O 端口作为热待机端口来操作，并且将所述多个 I/O 端口中的其余 I/O 端口作为热待机端口来操作；以及

其中，当处理器改变功率消耗模式时，所述指令促使处理器：

将所述其余 I/O 端口的功率消耗模式从具有第一峰值功率状态的第一功率消耗模式变成具有低于所述第一峰值功率消耗的第二峰值功率状态的第二功率消耗模式。

18. 根据权利要求 17 所述的非暂时性计算机可读介质，其中，所述指令还促使处理器：

确定所述第一 I/O 端口已经经历故障；以及然后

将所述第二 I/O 端口设置为主端口；以及

将所述多个 I/O 端口中的第三 I/O 端口的功率消耗模式变成第一功率消耗模式。

19. 根据权利要求 16 所述的非暂时性计算机可读介质，其还包括：

其中，当处理器将所述多个 I/O 端口聚合时，所述指令促使处理器操作所述多个 I/O 端口，使得每个 I/O 端口参与同网络的通信；以及

其中，当处理器改变功率消耗模式时，所述指令促使处理器：

将所述多个 I/O 端口的功率消耗模式从具有用于每个 I/O 端口的第一峰值功率状态的第一功率消耗模式变成具有用于每个 I/O 端口的第二峰值功率状态的第二功率消耗模式，所述第二峰值功率状态不同于所述第一峰值功率状态。

20. 根据权利要求 19 所述的计算机系统，其中，当处理器改变所述多个 I/O 端口的功率消耗模式时，所述指令促使所述处理器进行选自由以下各项组成的组的至少一个：变成所述第二功率消耗模式，其中，所述第二峰值功率状态低于所述第一峰值功率状态；以及变成所述第二峰值功率消耗模式，其中，所述第二峰值功率状态高于所述第一峰值功率状态。

控制聚合 I/O 端口的功率消耗的方法和系统

背景技术

[0001] 在大型数据中心的操作中，其包括每个耗散数百瓦功率的许多紧密封装服务器计算机系统，保持温度是主要问题和支出。在许多情况下，用以控制数据中心内的温度的设备成本可能达到或超过服务器设备的成本。由于关于温度的问题和成本以及要节能的压力，计算机系统制造商实现了功率降低技术。

[0002] 大多数的功率降低技术已经集中于服务器的处理器和存储器。例如，在处理负载低时的时间段期间，可以在较低功率消耗模式下（例如，在降低的时钟频率下）操作服务器的一个或多个处理器。同样地，在处理负载是或被预期很低时的时间段期间，主存储器的各部分可以使其内容被重新定位，并且将主存储器的各部分断电。

[0003] 可以进一步降低服务器的功率消耗的任何系统或技术将在市场中提供竞争优势。

附图说明

[0004] 为了示例性实施例的详细描述，现在将对附图进行参考，在所述附图中：

图 1 示出了根据至少一些实施例的计算机系统。

[0005] 图 2 示出了根据至少一些实施例的各组可执行指令与 I/O 端口之间的功能关系；以及

图 3 示出了根据至少一些实施例的方法。

[0006] 注释和命名法

特定术语遍及以下描述和权利要求被用来指代特定系统部件。如本领域的技术人员将认识到的，计算机公司可以用不同的名称来指代部件。本文并不意图对在名称而不是功能方面不同的部件之间进行区分。在以下讨论中和权利要求中，以开放方式来使用术语“包括”和“包含”，并且因此应将其解释为意指“包括但不限于”。并且，术语“耦合”或“耦合”意图意指间接、直接、光学或无线电连接。因此，如果第一设备耦合到第二设备，则该连接可以是通过直接电连接、通过经由其他设备和连接的间接电连接、通过光学连接或通过无线连接。

[0007] “功率消耗模式”应指代一种设备的操作模式，其将上限设置为设备在被利用时可能消耗的功率的量，但是不应指代设备的利用状态。例如，应将在“全开”功率消耗模式下操作的设备视为处于“全开”功率消耗模式，无论设备正在被其最大可能地利用（且正在吸取较高功率）还是设备根本并未被利用（和吸取较低功率）。换言之，不应将仅仅基于设备的利用率的变化的功率使用变化视为功率消耗模式的变化。

[0008] “聚合”和“聚合”应意指相对于通信网络的输入 / 输出（I/O）端口而言，该 I/O 端口表现为到软件栈的单个逻辑 I/O 端口。“聚合”和“聚合”应意指相对于存储网络的 I/O 端口而言，该 I/O 端口表示到存储设备的冗余路径。

具体实施方式

[0009] 以下讨论针对本发明的各种实施例。虽然这些实施例中的一个或多个可能是优

选的，但不应将公开的实施例解释为或以其他方式用作限制本公开的范围，包括权利要求。另外，本领域的技术人员将理解的是以下描述具有广泛的应用，并且任何实施例的讨论仅仅意图是该实施例的示例，并且并不意图暗示本公开的范围（包括权利要求）局限于该实施例。

[0010] 图 1 图示出根据至少一些实施例的计算机系统 100。特别地，计算机系统 100 包括通过集成主桥 14 被耦合到主存储器阵列 12 和各种其他外围计算机系统部件的主处理器 10。计算机系统 100 可以实现多个主处理器 10。主处理器 10 经由主机总线 16 耦合到主桥 14，或者可以将主桥 14 集成到主处理器 10 中。因此，除图 1 中所示的那些或作为其替代，计算机系统 100 可以实现其他总线配置或总线桥。

[0011] 主存储器 12 通过存储器总线 18 耦合到主桥 14。因此，主桥 14 包括通过维护（asserting）控制信号以用于存储器访问来控制到主存储器 12 的事务处理的存储器控制单元。在其他实施例中，主处理器 10 直接地实现存储器控制单元，并且主存储器 12 可以直接地耦合到主处理器 10。主存储器 12 充当用于主处理器 10 的工作存储器，并且包括其中存储程序、指令和数据的存储器设备或存储器设备阵列。主存储器 12 可以包括任何适当类型的存储器，诸如动态随机存取存储器（DRAM）或各种类型的 DRAM 设备中的任何一个，诸如同步 DRAM （SDRAM）、扩展数据输出 DRAM （EDODRAM）或兰巴斯（Rambus） DRAM （RDRAM）。主存储器 12 是存储程序和指令的非暂时性计算机可读介质的示例，并且其他示例是磁盘驱动器和闪速存储器设备。

[0012] 在一些实施例中，由在处理器 10 上执行的软件生成的文本和视频被提供给经由高级图形端口总线 22、串行总线（PCI Express）或其他适当类型的总线耦合到主桥 14 的图形处理单元（GPU）20。替换地，显示驱动器设备可以耦合到主扩展总线 26 或辅助扩展总线（即，外围部件互连（PCI）总线 32）中的一个。图形处理单元 20 耦合到的显示设备 24 可以包括能够在其上面表示任何图像或文本的任何适当电子显示设备。在其中计算机系统 100 是服务器系统（例如，在具有多个其他服务器系统的机架安装外壳中）的实施例中，可以省略图形处理单元 20 和显示设备 24。

[0013] 仍参考图 1，说明性计算机系统 100 还包括将主扩展总线 26 桥接至各种辅助扩展总线的第二桥接器 28，所述辅助扩展总线诸如低管脚计数（LPC）总线 30 和外围部件互连（PCI）总线 32。可以由诸如通用串行总线（USB）的桥设备 28 来支持各种其他辅助扩展总线。然而，计算机系统 100 不限于任何特定的芯片组制造商，并且因此可以等价地使用来自多种制造商中的任何一个的桥设备和扩展总线协议。

[0014] 固件集线器 34 经由 LPC 总线 30 耦合到桥设备 28。固件集线器 34 包括包含可由主处理器 10 执行的软件程序的只读存储器（ROM）。该软件程序包括在加电自我测试（POST）期间和刚好在其之后执行的过程（procedure）以及存储器参考代码的程序。POST 过程和存储器参考代码在计算机系统的控制被移交至操作系统之前执行计算机系统内的各种功能。

[0015] 计算机系统 100 还包括被说明性地耦合到 PCI 总线 32 的多个输入 / 输出（I/O）端口设备 36。I/O 端口设备 36 耦合到一个或多个网络类型。例如，在特定实施例中，I/O 端口设备 36 是网络接口卡（NIC），其将计算机系统 100 耦合到通信网络，诸如局域网（LAN）、广域网（WAN）和 / 或因特网。在又一其他实施例中，I/O 端口设备是存储适配器卡，其经由存储网络（例如，光纤信道）将计算机系统 100 耦合到一个或多个远程定位长期存储设备（例

如，硬盘、光盘）。可以等价地使用其他类型的 I/O 端口设备。虽然图 1 图示出被耦合到同一 PCI 总线 32 的 I/O 端口设备 36，但在其他实施例中，I/O 端口设备 36 可以耦合到不同的 PCI 总线，或者具有不同的通信协议的总线（例如，一个 I/O 端口卡被耦合到 PCI 总线且第二 I/O 端口设备被耦合到主扩展总线）。

[0016] 仍参考图 1，计算机系统 100 还可以包括经由 LPC 总线 30 被耦合到桥接器 28 的超级 I/O 控制器 38。超级 I/O 控制器 38 控制许多计算机系统功能，例如与诸如“软”盘驱动器 40 和“软”盘 42、光盘驱动器 44 和光盘 46、键盘 48 以及定点设备 50（例如鼠标）的各种输入和输出设备对接。超级 I/O 控制器 38 常常由于执行许多计算机系统功能而被称为“超级”。

[0017] 计算机系统 100 还可以包括长期数据存储设备，诸如经由说明性 PCI 总线 32（未示出总线适配器以免使图过于复杂）耦合到桥接器 28 的磁盘驱动系统 52。磁盘驱动系统 52 可以是单个驱动器或作为独立（或廉价）磁盘（RAID）系统的冗余阵列操作的驱动器阵列。虽然说明性磁盘驱动系统 52 被示为被耦合到 PCI 总线 24，但磁盘驱动系统可以等价地耦合至其他总线，诸如主扩展总线 26 或其他辅助扩展总线。

[0018] 每个 I/O 端口设备 36 实现至少一个通信端口，并且每个 I/O 端口设备 36 可以实现多个通信端口。例如，在采用 NIC 形式的 I/O 端口设备 36 的说明性情况下可以实现四个或八个通信端口，并且因此可以实现四个或八个单独可控接口。作为另一示例，在采用存储适配器卡形式的 I/O 端口设备 36 的说明性情况下，每个存储适配器可以实现四个或八个通信端口，并且因此可以实现到远程定位存储器的四个或八个单独可控接口。根据各种实施例，可以出于故障容忍度和 / 或增加通信吞吐量的目的将两个或更多端口分组、分队或聚合。可以在相同的 I/O 端口设备上实现聚合端口，或者端口可以跨越多个 I/O 端口设备。此外，计算机系统 100 可以实现多个聚合队。

[0019] 图 2 示出了根据至少一些实施例的被处理器 10 执行以实现聚合的各种软件的说明性关系。特别地，计算机系统实现了支持远程通信的操作系统（O/S）60。可以使用支持远程通信的任何当前可用或后来开发的操作系统。在图 2 的说明性情况下，操作系统 60 支持软件栈 62。在其中 I/O 端口设备 36 是通信网络接口设备的情况下，软件栈 62 可以是传输控制协议 / 网际协议（TCP/IP）栈，但可以同时地或替换地实现其他通信协议（例如，IPX、NetBEUI）。在其中 I/O 端口设备 36 是存储网络设备的情况下，软件栈 62 是存储网络栈，诸如 SCSI 栈。操作系统 60 且特别是说明性软件栈 62 使得一个或多个应用程序 63 能够通过网络和 / 或远程定位存储设备向例如其他计算机系统进行通信。

[0020] 每个端口 64 具有与之相关联的驱动器 68（在一些情况下，可以替换地将每个单独驱动器称为小端口驱动器）。在其中每个 I/O 端口设备 36 由同一卖方制造且具有相同能力的情况下，驱动器 68 可以是重复程序。然而，I/O 端口设备不需要由同一卖方制造或者具有相同的能力。例如，在 I/O 端口设备 36 是 NIC 的情况下，一个 NIC 可以实现 100 兆位每秒（Mbps）数据吞吐量，同时另一 NIC 可以实现 1000 Mbps（千兆位）吞吐量，并且在这些替换实施例中驱动器可以是卖方和 / 或能力特定的。尽管具有不同的卖方和 / 或不同的能力，根据本发明的实施例，仍可以将各种 I/O 端口设备或其端口聚合。

[0021] 在其中每个端口 64 独立地操作的情况下，说明性栈软件 62 直接与每个驱动器 68 通信；然而，根据各种实施例，端口 64 被聚合。为了使得能够实现聚合，聚合软件 70 在说明

性软件栈 62 与各种驱动器 68 之间对接。虽然图 2 将软件栈 62 和聚合软件 70 示为单独的软件片段(pieces),但在一些情况下,可以将该功能在单个程序中组合,如虚线 71 所图示的。更特别地,在通信网络的情况下,聚合软件 70 与软件栈 42 通信且向软件栈呈现用于每组聚合端口的单个逻辑端口(即,看起来像单个驱动器)。同样地,在通信网络的情况下,聚合软件 70 表现为每个驱动器 68 的软件栈。在 I/O 端口设备是 NIC 的情况下,可使用多个市售聚合软件产品,诸如可从加利福尼亚州帕洛阿尔托市的惠普公司获得的自动端口聚合(APA)LAN 监视软件。

[0022] 在存储网络的情况下,聚合以作为到端存储设备(end-storage device)的冗余链路的 I/O 端口形式出现。针对到存储设备的冗余链路,可以从任何聚合端口发送出存储通信,并且其仍到达端存储设备。由此可见,在其中所有 I/O 端口 64 被聚合的图 2 的说明性情况下,可以从 I/O 端口 64 中任何一个发送出存储通信,并且其仍到达特定端存储设备(例如,硬盘)。在存储网络的情况下,软件栈 62 可以知道用聚合 I/O 端口 64 表示的冗余链路,但是聚合软件 70 进行关于存储通信从哪个 I/O 端口流出的判定,并且因此执行聚合功能。这里再次地,虽然说明性图 2 将软件栈 62 和聚合软件示为单独软件片段,但可以将该功能组合成单个程序,如虚线 71 所图示的。

[0023] 然而,除端口 64 的聚合之外,根据各种实施例的聚合软件 70 还在 I/O 端口级执行关于功率管理的任务。在深入研究功率管理功能之前,并且为了更好地理解各种实施例,本说明书首先描述相关技术的功率管理的缺点。

[0024] 在其中端口 64 独立地操作的情况下(即,无组队或聚合),一些相关技术系统实现基于端口的不活动性操作的功率节省特征。例如,如果端口空闲达到预定时间量,则该端口可以被其各自的驱动器置于较低功率消耗模式。然而,并未跨越端口协调相关技术中的关于功率消耗模式的判定。此外,在其中端口被组队或聚合的情况下,相对于单个端口进行功率消耗模式修改可能对总体操作不利。考虑其中一个端口充当主端口的情况(所有通信被从该主端口发送出,并且通过该主端口接收所有通信)以及充当热待机的第二端口。如果第二端口基于不活动性而被置于降低功率消耗模式,则第二端口将不会具有在主端口发生故障的情况下快速地且无缝地载送负载的能力。由于不能跨组队或聚合端口应用协调控制,所以相关技术设备在此类情况下不能实现功率消耗模式控制(即,功率消耗模式控制被关掉,并且端口始终在峰值功率消耗模式下运行)。

[0025] 根据各种实施例,除实现用于端口 64 的聚合策略之外,聚合软件 70 还实现每个 I/O 端口 64 的功率消耗模式的协调控制,其中功率消耗模式控制取决于所实现的聚合的类型。本说明书首先转到功率消耗模式(和相关峰值功率状态)的描述,然后至在本文中称为“活动 - 待机”情况下的功率消耗模式控制,后面是在本文中称为“活动 - 活动”的情况下的功率消耗模式控制。

[0026] I/O 端口设备具有各种功率消耗模式。在其中在经由 PCI 总线耦合到计算机系统部件的 I/O 端口设备 36 上实现端口 64 的说明性情况下,可能的功率消耗模式包括 D0“全开”模式、$D3_{hot}$“关”模式和两个中间功率消耗模式 D1 和 D2。此外,根据串行总线

(PCIe)动态功率分配(DPA)标准,D0 模式具有多个子状态,其中,每个子状态定义所消耗功率、性能和 / 或其他特性之间的权衡。每个功率消耗模式可以具有不同的峰值功率状态。例如,说明性 D0“全开”模式具有第一峰值功率状态,并且说明性 D2 中间模式具有第

二峰值功率状态，其中，第二峰值功率状态低于第一峰值功率状态。可以用多种操作技术中的任何一个来实现较低峰值功率状态，诸如用于 I/O 端口设备上的电路的较低时钟速率、由 I/O 端口设备降低的通信频率以及 I/O 端口设备的降低的存储器使用。因此，为了使设备在其功率消耗模式下被最大可能地利用，功率消耗模式的变化导致功率消耗的变化。

[0027] 虽然每个功率消耗模式具有峰值功率状态，但设备不需要在峰值功率状态下操作—峰值功率状态仅仅是可以基于特定模式下的利用率被吸取的峰值功率。例如，在说明性 D0 “全开”状态中操作的设备（但该设备不在被利用）吸取特定量的功率以保持设备上的各种电路活动，但是吸取或耗散比设备被完全利用的情况下少的功率。然而，即使是针对未被利用的设备，功率消耗模式的变化也可能导致较低功率使用。例如，其功率消耗模式从说明性 D0 “全开”变成 $D3_{hot}$ “关”状态的空闲设备在 $D3_{hot}$ “关”状态下将吸取比 D0 “全开”更少的功率，即使未发生利用率的变化。为了本说明书的平衡，对功率消耗模式的变化的参考隐含地包括峰值功率状态的变化，再次地，理解成功率变化可以来自变化的利用率（当在功率消耗模式的极限处操作时）、甚至在不存在用于特定模式的完全利用的变化功率消耗或同时来自两者。

[0028] 在说明性活动 - 待机模式下操作的端口 64 表示其中端口 64 被聚合的情况，并且通过聚合软件 70 的操作表现为到软件栈 62 的单个端口。然而，在活动 - 待机模式下，一个端口（例如，端口 64A）被指定为主端口，并且其余端口（例如，端口 64B—64D）被指定为辅助端口。说明性主端口 64A 发送和接收所有通信，并且其余端口作为待机端口进行操作，准备在主端口故障的情况下接管作为主端口的职责。为了快速地且无缝地接管作为主端口的职责，在其中在几乎没有延迟的情况下端口能够接管有故障主端口（即，热待机）的功率消耗模式下操作辅助端口。

[0029] 根据在活动 - 待机模式下操作的实施例，聚合软件 70 实现跨所有聚合端口的功率控制策略，其并未负面地影响活动 - 待机操作模式。例如，考虑其中所有端口 64 处于其最高功率消耗模式且其中端口 64A 是主端口的情况。如果聚合软件 70 判定或被命令降低端口 64 的功率消耗，则聚合软件 70 可以将端口中的一个选择为唯一热待机端口（例如，端口 64B）。还可预期一个以上的热待机端口，但是为了方便起见，本讨论采取单个热待机端口。所选热待机被留在其中能够通过热待机来快速地且无缝地接管通信的功率消耗模式。换言之，聚合软件 70 避免改变热待机端口和主端口的功率消耗模式。其余端口（例如，端口 64C 和 64D）被置于降低功率消耗模式。在实现端口 64 的 PCI 设备的说明性情况下，主端口 64A 和热待机端口 64B 两者被置于或留在 D0 “全开”模式，而其余端口 64C 和 64D 被置于降低功率消耗模式（例如，$D3_{hot}$ “关”模式）。

[0030] 如果在说明性活动 - 待机情况下期望进一步的功率降低，并且能够容忍数据通信吞吐量降低，则聚合软件 70 可以改变主端口 64A 和热待机端口 64B 的功率消耗模式（例如，D0 的子状态）。

[0031] 在主端口故障的情况下，聚合软件 70 将热待机设置为主端口；另外，知道用于活动 - 待机情况的功率策略的聚合软件 70 可以选择另一（未发生故障）端口并提高所选端口的功率消耗模式以变成新的热待机。例如，在初始主端口 64A 故障时，聚合软件将端口 64B 设置为主端口，将另一端口（例如，端口 64C）选择为热待机端口，并且将新的热待机为端口 64C 置于用于热待机操作的适当功率消耗模式。在一些情况下，用于热待机端口的适当功率

消耗模式将是与主端口相同的功率消耗模式。

[0032] 仍参考图 2，本说明书现在转到端口的操作的说明性活动－活动模式。在说明性活动－活动模式下操作的端口 64 表示其中端口 64 被聚合且通过聚合软件 70 的操作对软件栈 62 表现为单个端口的情况。然而，在活动－活动模式下，每个端口参与同端口附着到的网络的通信。例如，在实现到存储网络的通信的端口的说明性情况下，每个端口 64 在到网络附着存储设备的通信中担任活动角色。同样地，在实现到通信网络的通信的端口的说明性情况下，每个端口在通信中担任活动角色。然而，活动角色不需要跨所有端口是相同的。例如，在实现到通信网络的通信的端口 64 的说明性情况下，一个端口(例如，端口 64A)可以既发送又接收消息分组，并且其余端口(例如，端口 64B—64D)可以仅从网络发送消息分组。在其他情况下，每个端口 64 可以既发送又接收消息分组。

[0033] 无论在活动－活动模式下实现的精确机制如何，聚合软件 70 跨所有聚合端口实现功率控制策略，其并不负面地影响活动－活动操作模式。例如，考虑其中所有端口 64 处于其最高功率消耗模式且所有端口以某种形式参与到网络的通信的情况。如果聚合软件 70 判定或被命令降低端口 64 的功率消耗，则聚合软件 70 通过降下(lowering)功率消耗模式或一个或多个端口 64 的功率状态来降低功率消耗。在一个情况下，聚合软件可以均匀地降低所有端口 64 的功率消耗模式。例如，聚合软件可以将所有端口的功率消耗模式从 D0“全开”模式变成 D0 的子状态中的一个。

[0034] 然而，降低活动－活动模式下的功率不需要功率消耗模式下的均匀降低。例如，在一些情况下，聚合软件可以在聚合端口 64 的较小子集上降低功率消耗模式以实现功率降低。例如，在其中一个端口既发送又接收消息分组(例如，端口 64A)且其余端口被用作只发送(例如，端口 64B 和 64C)的情况下，聚合软件可以降低只发送端口中的一个或多个的功率消耗模式以降低功率消耗。在一些情况下，功率消耗模式的降低包括将端口置于关闭状态。

[0035] 除降低功率之外，聚合软件 70 还可以负责增加功率。例如，聚合软件 70 可以判定或被命令以基于通信负载的实际或预期增加来增加可用通信容量。聚合软件 70 可以在预期利用率增加之前提高功率消耗模式，诸如网络服务器的用于端口 64 是通信网络端口的一天中的忙碌时间，或者可预期大的数据备份的用于端口 64 是存储网络端口的一天中的预期时间。就像功率消耗模式的降低，可以均匀地施加(例如，所有端口被置于 D0“全开”模式)或者非均匀地施加功率消耗模式的增加。

[0036] 而且，聚合软件 70 可以考虑活动－活动和活动－待机说明性模式下的端口特定参数。如上所述，可以在每个 I/O 端口设备上实现一个端口 64，可以在单个 I/O 端口设备上实现多个端口，并且 I/O 端口设备不需要具有相同的制造商、品牌或型号。在这些实施例中，聚合软件 70 可以考虑特定参数进行功率消耗模式修改。作为示例，考虑在单个 I/O 端口设备上实现端口 64C 和 64D(用虚线 72 图示出)。在这样的情况下，不可能不同地或者至少明显不同地设置端口 64C 和 64D 的功率消耗模式(例如，虽然两个端口可以在不同的“可操作”功率消耗模式下操作，但不可能将一个端口关闭而留下第二个可操作)。无论是在活动－待机还是在活动－活动操作中，在选择要关闭的端口时(例如，说明性 D3“关”状态)，聚合软件 70 可以选择关掉同一 I/O 端口设备 72 上的端口，其可以包括将特定功能移动至其他端口(例如，分别将主和热待机移动至端口 64A 和 64B)。如果可以使得所有其端口 64 不活动并设置在较低功率消耗模式，则可以将 I/O 端口设备 72 断电。这样做导致显著的功率节省。

[0037] 此外，给定端口 64 可以跨越不同的卖方和能力，端口 64 可以具有用于特定利用率的变化功率消耗。聚合软件 70 在实现特定功率策略时可以选择最高效的（从功率角度出发）一个或多个端口以用于预期利用率。例如，端口 64A 和 64B 可以具有类似的峰值消息分组操作速率，但是可以在功率消耗方面不同（例如，相对于较旧的设备而言较新的硬件设备）。因此，在实现功率策略时（包括改变功率消耗模式以降低 I/O 的总功率使用）时，聚合软件可以选择在最低功率消耗下提供期望功能的设备或设备组。当需要附加功能或容量时，不那么高效的端口可以将其功率消耗模式提高，使得端口参与或更多地参与总体通信。基于与特定端口相关联的参数来改变功率策略可在活动 - 活动和活动 - 待机情况下应用。

[0038] 仍参考图 2，根据至少一些实施例，聚合软件 70 仅负责知道和实现用于计算机系统 100 的聚合 I/O 的功率策略。例如，聚合软件可以始终尝试实现用于 I/O 端口的功率消耗模式，其在最低功率消耗下提供足够量的带宽或吞吐量能力。然而，在至少一些实施例中，聚合软件 70 是被频繁地计划用于执行且具有管理员权限的内核级软件。然而，在一些情况下，不需要如计划和执行聚合软件 70 那样频繁地进行功率策略判定，并且因此为了降低聚合软件的复杂性，一些实施例实现功率策略管理程序 74。

[0039] 功率策略管理程序 74 至少与聚合软件 70 通信，并且命令聚合软件 70 实现关于 I/O 端口的功率策略的变化。在特定实施例中，功率策略管理程序 74 是用户级程序，并且因此与聚合软件 70 相比不那么频繁地运行且具有较低权限。在一些情况下，在计算机系统 100 内执行功率策略管理程序 74，但是在其他情况下在不同的计算机系统中执行。功率策略管理程序 74 可以独立地进行关于功率策略的判定（在一些情况下基于从聚合软件接收到的数据，诸如利用率），或者功率策略管理程序 74 可以从其他程序接收命令（未明确地示出），并且基于从聚合软件 70 学习的计算机系统 100 的当前状态和较高级命令来设计特殊化策略。

[0040] 图 3 示出了根据至少一些实施例的方法（例如，软件）。特别地，该方法开始（方框 300）并前进至：将多个输入 / 输出（I/O）端口聚合（方框 302）；以及控制计算机系统中的功率消耗（方框 304）。控制功率消耗包括：从在计算机系统中执行的功率策略管理程序向同样在计算机系统中执行的聚合软件发送命令（方框 306），聚合软件实现聚合；并响应于该命令而改变 I/O 端口中的至少一个的功率消耗模式（方框 308）。其后，该方法结束（方框 310）。

[0041] 根据在本文中提供的描述，本领域的技术人员很容易能够将如所述地创建的软件与适当的通用或专用计算机硬件组合以根据各种实施例来创建计算机系统和 / 或计算机子部件，以创建用于执行各种实施例的方法的计算机系统和 / 或计算机子部件，以及和 / 或创建用于存储软件或程序以实现各种实施例的方法方面的非暂时性计算机可读存储介质。

[0042] 以上讨论意图说明本发明的原理和各种实施例。一旦完全认识到以上公开，许多变更和修改对于本领域的技术人员而言将是显而易见的。例如，每个部件具有最大数目的功率循环，在那之后，设备很可能将发生故障。在一些实施例中，在本文中所讨论的协调控制在做出关于用于设备的功率消耗模式的判定时将设备的功率循环的数目考虑在内。此外，本发明人使用不同的术语“程序”或“软件”来帮助读者区分可由处理器执行的指令的各种功能单元，并且并不意味着存在基本差异（除被编码成执行不同任务之外）。事实上，可以用相同的编程语言来编写每个程序和 / 或软件且其具有被共享或重叠的许多元素。意图

在于将以下权利要求解释为涵盖所有此类变更和修改。

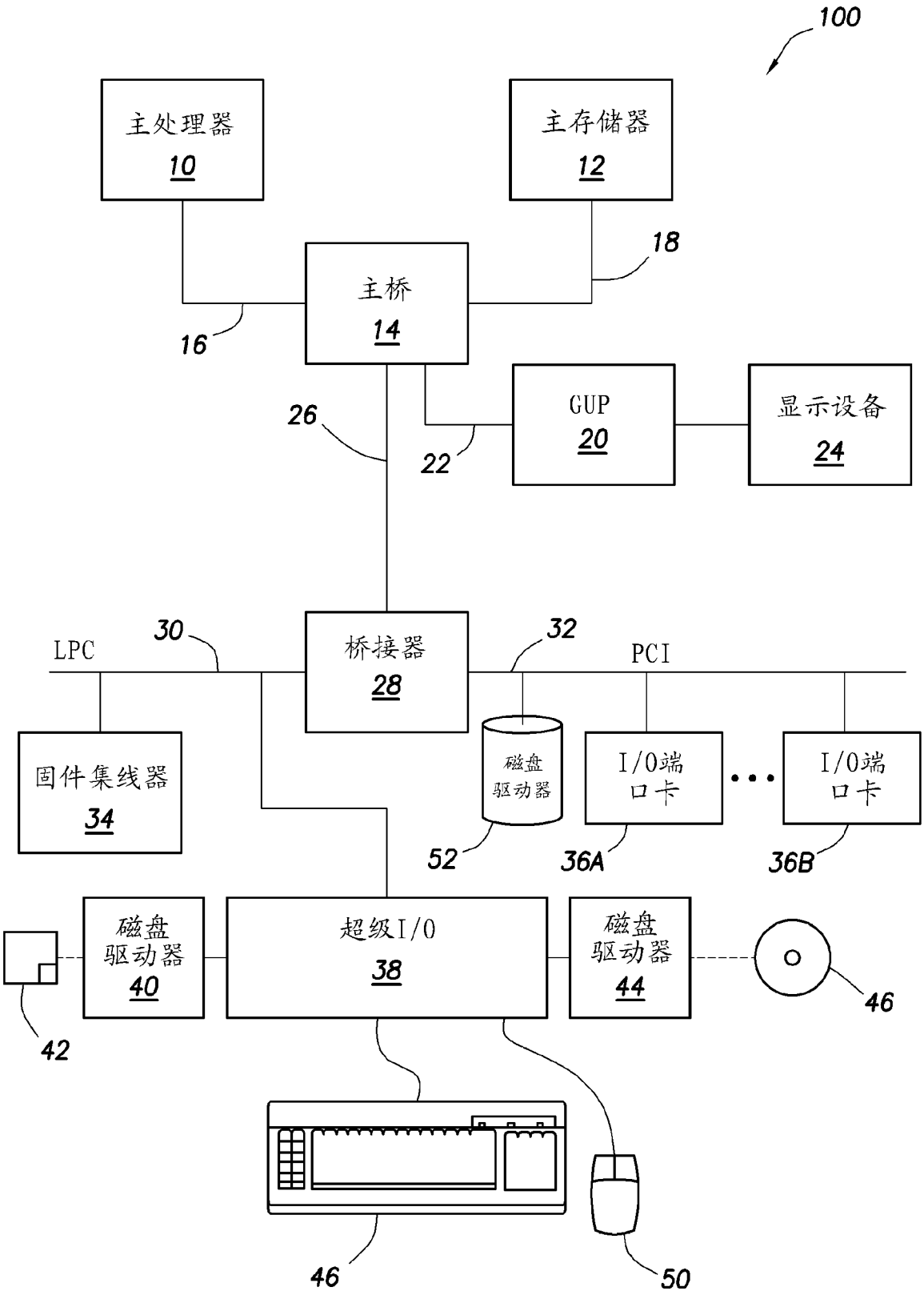

100
主处理器
10
主存储器
12
18
主桥
14
16
26
22
GUP
20
显示设备
24
30
LPC
桥接器
28
32
PCI
固件集线器
34
磁盘
驱动器
52
I/O端
口卡
36A
I/O端
口卡
36B
磁盘
驱动器
40
超级I/O
38
磁盘
驱动器
44
46
42
46
50

图 1

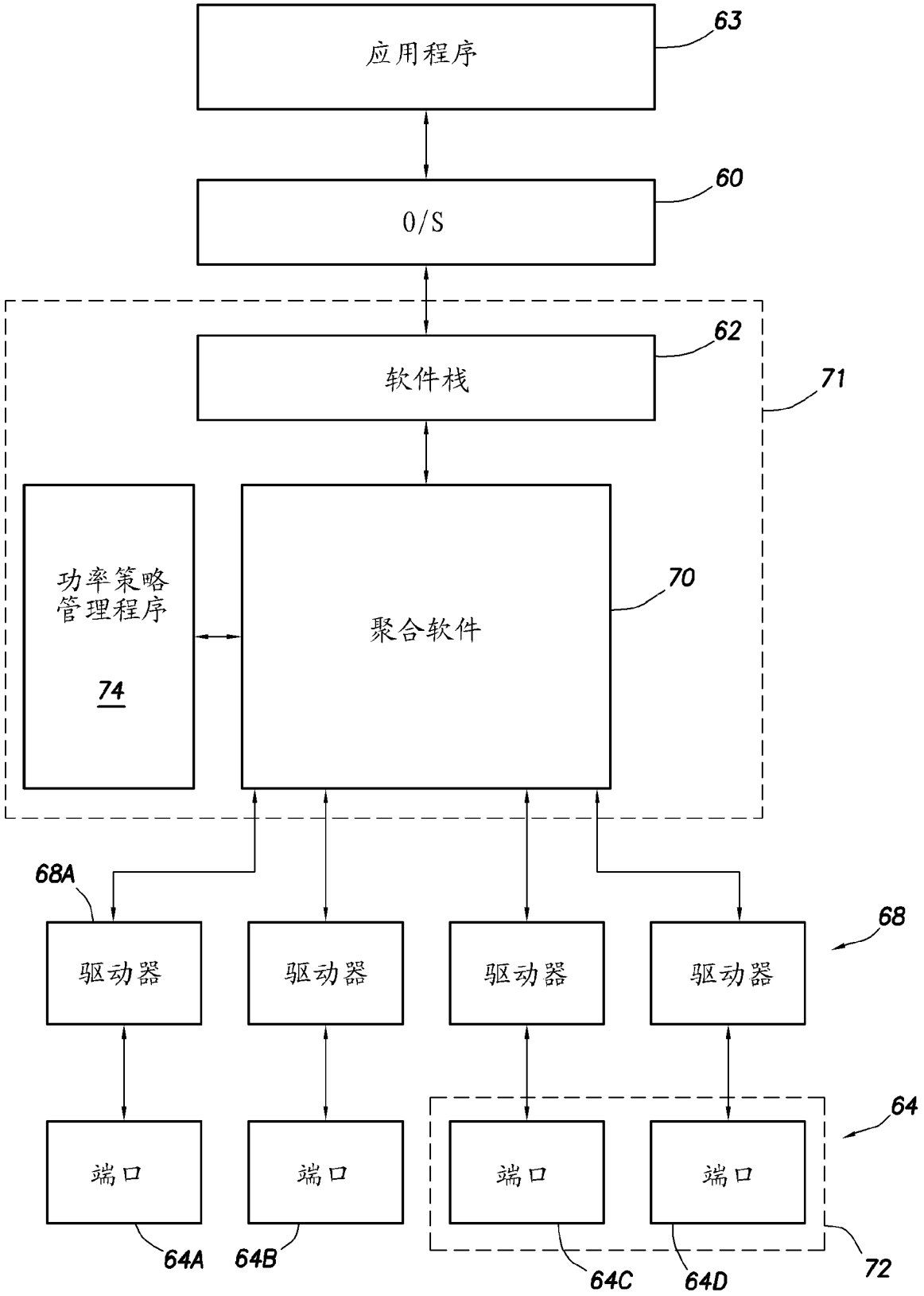

应用程序
63
O/S
60
软件栈
62
71
功率策略
管理程序
74
聚合软件
70
68A
驱动器
驱动器
驱动器
驱动器
68
端口
端口
端口
端口
64
64A
64B
64C
64D
72

图 2

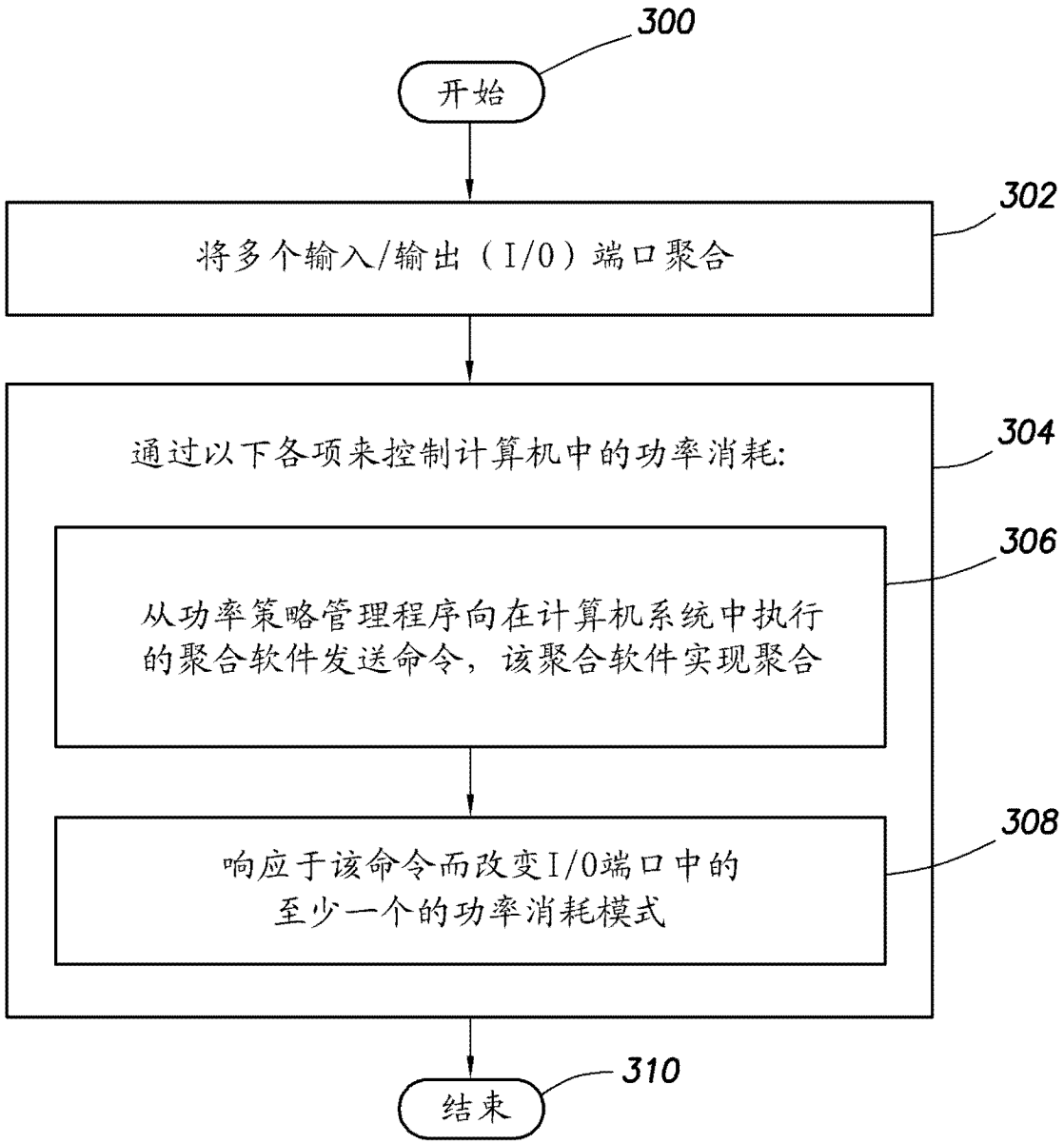

300
开始
302
将多个输入/输出（I/O）端口聚合
304
通过以下各项来控制计算机中的功率消耗:
306
从功率策略管理程序向在计算机系统中执行的聚合软件发送命令，该聚合软件实现聚合
308
响应于该命令而改变I/O端口中的至少一个的功率消耗模式
310
结束

图 3