[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
*G06F 1/32 (2006.01)*
*G06F 11/30 (2006.01)*

# [12] 发明专利申请公开说明书

[21] 申请号 200510077642.2

[43] 公开日 2006 年 4 月 5 日

[11] 公开号 CN 1755587A

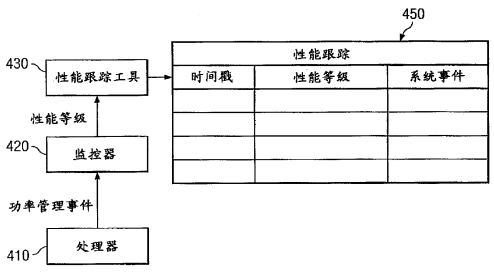
[22] 申请日 2005.6.17
[21] 申请号 200510077642.2
[30] 优先权
[32] 2004. 9.30 [33] US [31] 10/955,182
[71] 申请人 国际商业机器公司
地址 美国纽约阿芒克
[72] 发明人 布雷特·罗纳德·奥尔斯泽斯基
卢斯·雷内·斯莫尔德斯
兰德尔·克雷格·斯旺伯格
[74] 专利代理机构 北京市金杜律师事务所
代理人 王茂华
权利要求书 3 页 说明书 12 页 附图 4 页

[54] 发明名称

用于监控可变速微处理器性能的方法和设备

[57] 摘要

提供了一种功率等级监控器和性能跟踪工具，用于将系统性能与处理器管理事件相关联。 当功率管理请求微处理器的状态发生变化时，将通知软件。 可通知多层软件，包括固件级、操作系统以及应用程序。 性能跟踪工具跟踪功率管理事件的时间以及它们对微处理器性能的影响。 然后，性能跟踪工具可显示或记录处理器性能状态的变化。 这些变化可与其它系统事件相关联，以帮助确定有关功率管理方面的系统性能问题。



1.一种用于对数据处理系统中由功率管理引起的性能变化进行跟踪的方法，该方法包括:

监控处理器中的功率管理事件;

确定该处理器是否响应该功率管理事件，由先前的性能等级变化至新的性能等级；以及

在该处理器变化至该新的性能等级时，记录将该新的性能等级与系统状态信息相关联的性能跟踪。

2.根据权利要求1所述的方法，其中监控功率管理事件包括检测来自于该处理器的中断。

3.根据权利要求1所述的方法，其中该先前的性能等级处于第一范围内，其中确定该处理器是否由该先前的性能等级变化至新的性能等级包括:

确定当前性能等级；以及

确定当前性能等级是否处于第二范围内。

4.根据权利要求1所述的方法，其中记录性能跟踪包括为该功率管理事件产生跟踪条目，以及给该跟踪条目打上时间戳。

5.根据权利要求1所述的方法，其中该系统状态信息包括中断、分页事件和温度值的至少其中之一。

6.根据权利要求1所述的方法，其中操作系统运行于该处理器上，其中记录性能跟踪的步骤由该操作系统完成。

7.根据权利要求6所述的方法，其中该监控功率管理事件的步骤由该操作系统完成。

8.根据权利要求1所述的方法，其中该数据处理系统是多处理器系统。

9.根据权利要求8所述的方法，其中该处理器是第一处理器，该方法进一步包括:

监控第二处理器中的功率管理事件；以及

记录在该第二处理器中的该功率管理事件的性能跟踪条目。

10. 根据权利要求 8 所述的方法，其中该数据处理系统包括系统管理程序。

11. 根据权利要求 10 所述的方法，其中该记录性能跟踪的步骤由该系统管理程序完成。

12. 根据权利要求 10 所述的方法，其中该监控功率管理事件的步骤由该系统管理程序完成。

13. 根据权利要求 1 所述的方法，进一步包括基于该性能跟踪确定计算机使用的费用。

14. 一种用于跟踪数据处理系统中由功率管理引起的性能变化的设备，该设备包括:

用于监控处理器中功率管理事件的装置;

用于确定该处理器是否响应该功率管理事件，由先前的性能等级变化至新的性能等级的装置; 以及

用于在该处理器变化至该新的性能等级时，记录将该新的性能等级与系统状态信息相关联的性能跟踪的装置。

15. 根据权利要求 14 所述的设备，其中该用于监控功率管理事件的装置包括用于检测来自于该处理器的中断的装置。

16. 根据权利要求 14 所述的设备，其中该先前的性能等级处于第一范围内，其中用于确定该处理器是否由该先前性能等级变化至新性能等级的装置包括:

用于确定当前性能等级的装置; 以及

用于确定该当前性能等级是否处于第二范围内的装置。

17. 根据权利要求 14 所述的设备，其中该用于记录性能跟踪的装置包括为该功率管理事件产生跟踪条目的装置，以及给该跟踪条目打上时间戳的装置。

18. 根据权利要求 14 所述的设备，其中该数据处理系统是多处理器系统。

19. 根据权利要求 18 所述的设备，其中该处理器是第一处理器，

该设备进一步包括:

用于监控第二处理器中的功率管理事件的装置;以及

用于记录该第二处理器中的该功率管理事件的性能跟踪条目的装置。

20. 根据权利要求 14 所述的设备,进一步包括用于基于该性能跟踪确定计算机使用的费用的装置。

21. 一种计算机程序产品,用于跟踪数据处理系统中由功率管理引起的性能变化,该计算机程序产品包括:

用于监控处理器中的功率管理事件的指令;

用于确定该处理器是否响应该功率管理事件,由先前的性能等级变化至新的性能等级的指令;以及

用于在该处理器变化至该新的性能等级时,记录将该新的性能等级与系统状态信息相关联的性能跟踪的指令。

# 用于监控可变速微处理器性能的方法和设备

## 技术领域

本发明涉及数据处理，更具体地，涉及微处理器功率管理。更具体地，本发明提供一种方法、设备和程序产品，用于监控功率管理引起的可变速微处理器性能。

## 背景技术

微处理器技术上的进展正在被功率消耗和冷却问题所制约。将来，可能需要将很多高性能微处理器设计为可基于对其自身内部状态或环境的测量自动地改变它们的功率消耗。这意味着，在高度活动或环境压力期间，微处理器的性能可能会受到抑制或者向下调整。

当今的计算机系统总体上设计为提供一个使所有微处理器运行在相同的基本性能等级的环境。在某些情况下，可以对微处理器类型进行混合和匹配，使得一些微处理器运行在较快的速度，而其他的则运行在较慢的速度。但在总体上，软件却认为每个处理器的性能等级是固定不变的。

随着加大对功率管理的强调，软件可能在这样一种环境中执行，其中，复杂系统内的微处理器性能可以从处理器到处理器，或从时间段到时间段剧烈变化。这种变化可以引起关于一些客户的工作量的性能问题。这些问题可包括这样一些变化的征兆，例如处理器消耗的运行到运行（run-to-run）易变性，可扩缩性问题，以及外部中断的丢失。

## 发明内容

本发明认识到现有技术的一些问题，并提供一种功率等级监控器和性能跟踪工具，用于将系统性能和处理器管理事件相关联。当功

率管理请求对微处理器状态的变化时，就通知软件。可能会通知到多层软件，包括固件级、操作系统以及应用软件。性能跟踪工具跟踪功率管理事件的时间以及它们对微处理器性能的影响。性能跟踪工具可显示或记录处理器性能的状态变化。这些变化可与其它系统事件相关联，以帮助确定关于功率管理方面的系统性能问题。

## 附图说明

被认为是本发明特征的新颖特征在所附权利要求书中给出。但是结合附图阅读下列示意性实施方式的详细描述，可以对本发明本身、其优选使用模式及其进一步的目的和优势有更好的理解，其中：

图 1 是一个数据处理系统示意性实施方式的框图，利用这个系统可以有利地使用本发明；

图 2 是一个可在其中实现本发明的示意性逻辑分区平台的框图；

图 3 是一个可在其中实现本发明的单处理器数据处理系统的框图；

图 4 是表示依照本发明一个示意性实施方式的性能跟踪系统的框图；

图 5 是表示依照本发明一个示意性实施方式的监控器的操作的流程图；以及

图 6 是表示依照本发明一个示意性实施方式的性能跟踪工具的操作的流程图。

## 具体实施方式

本发明提供一种方法、设备和计算机程序产品，用于监控功率管理引起的可变速微处理器性能。数据处理设备可以是单处理器计算设备、多处理数据处理系统、或者是虚拟处理器环境，在其中可利用多个处理器和多层软件来完成本发明的各方面。因此，提供下列图 1-3 作为可在其中实现本发明的数据处理环境的示意图。应当理解，图 1-3 仅仅是示例性的，并不旨在表明或暗示关于本发明可在

其中实现的环境方面的任何限制。可以在不偏离本发明的精神和范围的基础上，对所描述环境进行各种修改。

现在参照附图，更具体地，参照图 1，其描述了一个数据处理系统示意性实施方式的框图，可在这个系统有利地使用本发明。如图所示，数据处理系统 100 包括处理器卡 111a-111n。处理器卡 111a-111n 中的每个包括处理器和高速缓冲存储器。例如，处理器卡 111a 包含处理器 112a 和高速缓冲存储器 113a，处理器卡 111n 包含处理器 112n 和高速缓冲存储器 113n。

处理器卡 111a-111n 连接到主总线 115 上。主总线 115 支持包含处理器卡 111a-111n 和存储卡 123 在内的系统平面 120。该系统平面还包含数据交换器 121 和存储控制器/高速缓冲存储器 122。存储控制器/高速缓冲存储器 122 支持存储卡 123，该存储卡 123 包括带有多个双嵌入存储模块（DIMM）的本地存储器 116。

数据交换器 121 连接到位于本地 I/O （NIO）平面 124 内的总线桥 117 和总线桥 118。如图所示，总线桥 118 经由系统总线 119 连接到互连外围设备（PCI）桥 125 和 126。PCI 桥 125 经由 PCI 总线 128 连接到多个 I/O 设备。如图所示，硬盘 136 可经由小型计算机系统接口（SCSI）主机适配器 130 连接到 PCI 总线 128。图形适配器 131 可直接或间接地连接到 PCI 总线 128。PCI 桥 126 通过 PCI 总线 127 向经由网络适配器 134 和适配器卡插槽 135a-135n 的外部数据流提供连接。

工业标准结构（ISA）总线 129 经由 ISA 桥 132 连接到 PCI 总线 128。ISA 桥 132 通过具有串行连接串口 1 和串口 2 的 NIO 控制器 133 提供互连能力。软盘驱动器连接 137、键盘连接 138 和鼠标连接 139 由 NIO 控制器 133 提供，以允许数据处理系统 100 接收来自用户的通过相应的输入设备输入的数据输入。此外，非易失性 RAM（NVRAM）140 提供非易失性存储，用于保存来自例如电源问题等系统中断或系统故障的特定类型的数据。系统固件 141 也连接到 ISA 总线 129，用于实现初始的基本输入/输出系统（BIOS）功能。服务处理器 144 连

接到 ISA 总线 129，以提供用于系统诊断或系统维修的功能。

操作系统（OS）可存储在硬盘 136 上，硬盘 136 也可为由数据处理系统执行的另外应用软件提供存储。NVRAM 140 用于存储系统变量和错误信息，用于现场可更换单元（FRU）隔离。在系统启动期间，引导程序加载操作系统并初始化操作系统的执行。为了加载操作系统，引导程序首先从硬盘 136 定位操作系统内核类型，将 OS 加载到存储器中，并跳转至由操作系统内核提供的初始地址。典型地，操作系统加载到数据处理系统内的随机存取存储器（RAM）。一旦完成加载和初始化，操作系统控制程序的执行并可提供各种服务，例如资源分配，调度，输入/输出控制和数据管理。

可在利用多个不同硬件配置和软件，例如引导程序和操作系统的数据处理系统中执行本发明。例如，数据处理系统 100 可以是独立系统，或者是局域网（LAN）或广域网（WAN）等网络的一部分。

依照本发明的一个优选实施方式，处理器 112a-112n 包括功率管理能力。处理器 112a-112n 可基于环境影响等调节功率等级。如上所述，服务处理器 144 可提供用于系统诊断或系统维修的功能。例如，服务处理器 144 可确定温度情况或功率消耗事件。处理器 112a-112n 可响应这样的事件，向下或向上调节性能（速度）。功率调节的种类可包括单一低性能等级、多性能等级，甚至无限的可变性能等级。

监控器识别功率管理事件在何时发生并将功率状态通知给性能跟踪工具。性能跟踪工具记录或显示状态变化，并将这些变化与其它系统事件相关联，以帮助确定系统性能问题。例如，性能跟踪工具可利用跟踪来识别其它系统事件。

在本发明的一个示例性实施方式中，监控器和性能跟踪工具可作为软件存在于操作系统上或作为操作系统的一部分。或者，监控器和性能跟踪工具两者或其中一个可存在于固件中，例如系统固件 141，或者服务处理器 144。例如，系统固件 141 可包括在逻辑分区（LPAR）数据处理系统中管理分区的系统管理程序（hypervisor）。

大型对称多处理器数据处理系统，例如可从国际商业机器公司得到的 IBM eServer P690，可从惠普公司得到的 DHP9000 Superdome 企业服务器，以及可从 SUN 微系统有限公司得到的 Sunfire 15k 服务器，可进行分区并用作多个小型系统。这些系统通常被称为逻辑分区（LPAR）数据处理系统。数据处理系统内的逻辑分区功能允许在一个单数据处理系统平台上同时运行单操作系统的多个拷贝和多个不同的操作系统。分区，操作系统映像运行于其内，可以被分配一个平台物理资源的非重叠子集或重叠的资源，由固件管理。这些平台可分配资源包括一个或多个架构截然不同的处理器以及它们的中断管理区域、系统存储器区和输入/输出（I/O）适配器总线插槽。分区资源由平台的固件提供给操作系统映像。

关于逻辑分区数据处理系统中的硬件资源，这些资源在多个分区之间共享。例如，这些资源可包括输入/输出（I/O）适配器，存储模块，非易失性随机存取存储器（NVRAM），以及硬盘驱动器。可反复地引导和关闭 LPAR 数据处理系统内的每个分区，不需要重新启动（power-cycle）整个数据处理系统。

现在参照图 2，其描述了一个可在其中实现本发明的示例性逻辑分区平台的框图。例如，逻辑分区平台 200 中的硬件可实现图 1 中的数据处理系统 100。逻辑分区平台 200 包括分区硬件 230、操作系统 202、204、206、208 和系统管理程序 210。操作系统 202、204、206 和 208 可以是同时运行于平台 200 上的单操作系统的多个拷贝或多个不同的操作系统。这些操作系统可使用设计用来与系统管理程序通信的 OS/400、AIX 或 Linux™ 操作系统实现。操作系统 202、204、206 和 208 位于分区 203、205、207 和 209 中。

此外，这些分区还包括固件加载器 211、213、215 和 217。固件加载器 211、213、215 和 217 可使用可从国际商业机器公司得到的运行时间提取软件（RTAS）以及 IEEE-1275 标准的开放固件来实现。在将分区 203、205、207 和 209 实例化时，系统管理程序的分区管理器将开放固件的拷贝加载到每个分区里。然后，将与各分区相关

的或分配给各分区的处理器调度至分区的存储器中，以执行分区固件。

分区硬件 230 包括多个处理器 232-238，多个系统存储单元 240-246，多个输入/输出（I/O）适配器 248-262，以及存储单元 270。分区硬件 230 还包括服务处理器 290，可用于提供多种服务，例如处理分区内的错误，或者处理来自每个分区内处理器的功率管理事件。可将处理器 232-238、存储单元 240-246，NVRAM 存储器 298，以及输入/输出（I/O）适配器 248-262 中的每一个分配给逻辑分区平台 200 中多个分区中的一个，每个分区对应于操作系统 202、204、206 和 208 中的一个。

系统管理程序固件 210 为分区 203、205、207 和 209 完成多种功能和服务，以产生和执行对逻辑分区平台 200 的分区。系统管理程序 210 是一个与下层硬件相同的固件实现的虚拟机。系统管理程序软件可从国际商业机器公司得到。固件是存储在只读存储器（ROM）、可编程 ROM（PROM）、可擦除可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）以及非易失性随机存取存储器（非易失性 RAM）等不需要电源就能保存其中内容的存储芯片中的“软件”。从而，系统管理程序 210 通过把逻辑分区平台 200 的所有硬件资源虚拟化允许同时执行独立的操作系统映像 202、204、206 和 208。

根据本发明的一个示例性实施方式，在虚拟处理器环境，例如 LPAR 平台 200 中，系统管理程序固件 210 监控每个微处理器的状态，将当前状态传送至每个分区，也就是所谓的虚拟处理器。系统管理程序固件 210 还可维护每个虚拟处理器的统计数据，该统计数据跟踪了功率管理事件的数量以及在每个性能阈值的执行时间。如果微处理器有三个性能等级（例如，正常的、降低的、超级降低的），则分别记录在每种模式中执行的周期。如果微处理器的性能等级变化范围很大，则记录可以是以范围为单位来进行。

操作系统 202、204、206、208 中的每个可使用基于跟踪的方法来将微处理器性能与其它系统事件相关联。同时，调度分区，关于

微处理器当前性能等级的每个虚拟处理器信息可为其相应的操作系统所使用。然后，操作系统采用这个性能等级，直至向操作系统表示有功率管理事件发生，这时候，就定义一个新状态。从而，操作系统可以获知任意一个在其实例中的虚拟处理器在任意时间的当前性能等级。通过将状态压入到一个打上时间戳的跟踪中，操作系统事件，例如调度、中断、分页等可与当前处理器性能等级相关联。因此，可以识别因功率管理而引起性能变化的性能问题。

在一个可供选择的实施方式中，本发明可应用于单处理器数据处理系统。图 3 是可在其中实现本发明的单处理器数据处理系统的框图。数据处理系统 300 可以是一个计算机，例如图 1 中的客户机 108 的例子，实现本发明各个过程的代码或指令位于其中。在所描述的例子中，数据处理系统 300 采用中心（HUB）架构，包括北桥和存储控制中心（MCH）308 以及南桥和输入/输出（I/O）控制中心（ICH）310。处理器 302、主存储器 304 和图形处理器 318 连接到 MCH 308。例如，图形处理器 318 可通过加速图形端口（AGP）连接到 MCH。

在所描述的例子中，局域网（LAN）适配器 312、音频适配器 316、键盘和鼠标适配器 320、调制解调器 322、只读存储器（ROM）324、硬盘驱动器（HDD）326、CD-ROM 驱动器 330、通用串行总线（USB）端口和其它通信端口 332，以及 PCI/PCIe 设备 334 可连接到 ICH 310。例如，PCI/PCIe 设备可包括以太网适配器、插卡、用于笔记本计算机的 PC 卡等。PCI 使用卡总线控制器，但 PCIe 不使用。例如，ROM 324 可以是闪速二进制输入/输出系统（BIOS）。例如，硬盘驱动器 326 和 CD-ROM 驱动器 330 可使用电子集成驱动器（IDE）或串行高级技术附加装置（SATA）接口。超级 I/O（SIO）设备 336 可连接到 ICH 310。

操作系统运行于处理器 302 上，用于协调图 3 中数据处理系统 300 内的各种组件并提供对这些组件的控制。操作系统可以是商业上可得到的操作系统，例如可从微软公司得到的 Windows XP™。面向对象的编程系统，例如 Java™编程系统，可结合操作系统运行，并提供从执行于数据处理系统 300 上的 JAVA™程序或应用程序到操作系

统的调用。“JAVA”是SUN微系统有限公司的商标。

用于操作系统、面向对象编程系统以及应用程序或程序的指令位于硬盘驱动器326等存储器设备中，并且可加载到主存储器304中，用于由处理器302执行。由处理器302使用计算机实现的指令来完成本发明的各个过程，这些计算机实现的指令可以位于存储器中，例如主存储器304、存储器324或者一个或多个外围设备326和330。

本领域普通技术人员可以理解，图3中的硬件可基于不同的实现方式而有所变化。除了图3所述的硬件之外，还可使用其他内部硬件或外围设备，例如闪存、等效非易失性存储器、或光盘驱动器等，也可用这些来代替图3所述的硬件。本发明的各个过程还可应用于多处理器数据处理系统。

例如，数据处理系统300可以是个人数字助理（PDA），其配置有闪存，以提供非易失性存储，用于存储操作系统文件和/或用户生成的数据。图3中所描述的例子和上述各例子并不意味着有架构上的限制。例如，除了PDA形式，数据处理系统300也可以是板块（tablet）计算机、膝上型计算机，或者电话设备。膝上型计算机等移动数据处理系统可以包括功率管理功能。

在一个示例性实施方式中，处理器302能够自动调节功率等级，以适应环境的即时变化。当功率管理请求一个处理器状态的变化时，将通知软件。例如，处理器302可产生一个中断来指示功率管理事件。监控器检测该功率管理事件，并向性能跟踪工具提供状态变化的指示，性能跟踪工具跟踪功率管理事件的时间以及它们对微处理器性能的影响。然后，性能跟踪工具可显示或记录处理器性能状态的变化。这些变化可以与其它系统事件相关联，以帮助确定关于功率管理方面的系统性能问题。然后，例如，操作系统或系统管理程序可使用关于可变处理器性能的信息进行计费。

图4是表示依照本发明一个示例性实施方式的性能跟踪系统的框图。当对性能进行调节以防止功率消耗或冷却问题时，处理器410生成功率管理事件。例如，处理器410可通过产生中断来生成这一

事件。

监控器 420 从处理器 410 处接收功率管理事件，确定处理器 410 的性能等级。在多处理器数据处理系统中，可以对于所有的处理器有一个监控器，或者，选择地，可以对于每个处理器存在一个监控器。在处理器 410 初始化时，操作系统可以采用默认的性能等级。操作系统可提供包含可由性能跟踪工具 430 提取的状态的结构。

对于每个功率管理事件，性能跟踪工具 430 在性能跟踪 450 中存储一个条目。例如，性能跟踪 450 中的条目可包括时间戳、处理器性能等级、以及系统事件。例如，性能跟踪 450 可以是表格、数据库或其他数据结构。例如，系统事件可包括中断、分页等等。操作系统或系统管理程序可跟踪统计数据，以将其包括在系统事件信息中。例如，操作系统或系统管理程序可跟踪由每个性能等级执行的周期数。性能跟踪工具 430 可为多个处理器中的每个单独提供性能跟踪。图 4 所示的性能跟踪是示例性的，且基于不同的实现方式可以有所变化。例如，性能跟踪 450 可包括硬件事件，例如温度测量等。例如，性能跟踪 450 还可在单一跟踪中将多个处理器的性能管理事件与系统事件相关联。

然后，操作系统或系统管理程序（未示出）可使用关于可变处理器性能的信息进行计费。例如，如果是基于程序所用的执行时间量对客户收费，则可以基于客户实际从处理器得到的性能对费用设定比例或进行加权。例如，考虑下列公式：

$$\cos t = CPU\cos t * ((s1 * scale1) + (s2 * scale2) + (s3 * scale3))$$

其中，CPUcost 是 CPU 使用的总费用，s1 是处理器处于第一性能等级的秒数，scale1 是第一性能等级的权重，s2 是处理器处于第二性能等级的秒数，scale2 是第二性能等级的权重，s3 是处理器处于第三性能等级的秒数，scale3 是第三性能等级的权重。在这个例子中，第一性能等级可以是“正常的”性能等级，对于这个性能等级的权重可以是 1 或 100%。第二性能等级可以是“降低的”性能等级，对于这个性能等级的权重可以是 2/3 或 66%。第三性能等级可以是

“超级降低的”性能等级，对于第三性能等级的权重可以是 1/3 或 33%。基于性能跟踪统计数据的计费可在操作系统中完成，或者，选择地，当对固件编程以运行于方式，也可在系统管理程序或服务处理器中完成计费。

监控器 420 和性能跟踪工具 430 可存在于在操作系统控制下运行的软件中或者本身是操作系统的一部分。监控器 420 和性能跟踪工具 430 还可集成到单个软件组件中。可供选择地，监控器 420 和/或性能跟踪工具 430 可存在于固件或硬件中。在一个示例性实施方式中，在 LPAR 数据处理系统中，监控器 420 可存在于系统管理程序固件中。在一个可供选择的实施方式中，监控器 420 可存在于服务处理器中。在 LPAR 数据处理系统中，性能跟踪工具 430 可存在于系统管理程序固件中或操作系统中。从而，操作系统的每个实例，即虚拟处理器，可具有其各自的性能跟踪工具，用于跟踪其运行于的处理器。

图 5 是表示依照本发明一个示意性实施方式的监控器的操作的流程图。操作开始，监控器确定是否存在退出情况（块 502）。例如，当数据处理系统关闭，处理器不再分配给指定的逻辑分区，或者处理器变得不可操作时，可存在退出情况。如果存在退出情况，操作结束。

如果块 502 中不存在退出情况，则监控器确定是否发生功率管理事件（块 504）。例如，处理器可通过中断来表示功率管理事件。如果没有发生功率管理事件，操作返回块 502，以确定是否存在退出情况。如果块 504 中发生功率管理事件，则监控器确定处理器的当前性能等级（块 506），并向性能跟踪工具提供性能等级（块 508）。接下来，操作返回块 502，以确定是否存在退出情况。

监控器可通过判断自从最近一个性能等级开始性能等级是否发生变化，来确定该性能等级。例如，监控器可确定性能等级是否从“正常的”变为“降低的”。更特别地，如果处理器具有很大范围的性能等级，监控器可确定性能等级是否从一个范围变化到另一个

范围。例如，处理器可能具有 100 个离散的性能等级。在这种情况下，监控器可以包括一个或多个阈值，用于将性能等级分组到多个不同的范围中。从而，0 至 33 之间的性能等级可认为是超级降低的，34 至 66 之间的性能等级可认为是降低的，67 至 100 之间的性能等级可认为是正常的运行范围。在这种情况下，对于每个性能等级变化，或者，可供选择地，只有当性能等级进入到一个新的范围时，监控器将性能等级发送至性能跟踪工具。

图 6 是表示依照本发明一个示意性实施方式的性能跟踪工具的操作的流程图。操作开始，性能跟踪工具确定是否存在退出情况（块 602）。例如，当数据处理系统关闭，或者指定的逻辑分区终止时，可存在退出情况。如果存在退出情况，则操作结束。

如果块 602 中不存在退出情况，性能跟踪工具确定处理器是否变化到一个新的性能等级（块 604）。例如，可通过确定是否从监控器接收到新性能等级来进行上述确定。性能跟踪工具可从多个这样的监控器接收性能等级信息。如果处理器没有变化到新的性能等级，操作返回块 602，以确定是否存在退出情况。

性能跟踪工具可通过判断自从最近一个性能等级开始性能等级是否发生变化，来确定处理器是否变化到新的性能等级。例如，性能跟踪工具可确定性能等级是否从“正常的”变化为“降低的”。更特别地，如果处理器具有很大范围的性能等级，性能跟踪工具可确定性能等级是否从一个范围变化到另一个范围。例如，处理器可能具有 100 个离散的性能等级。在这种情况下，性能跟踪工具可确定处理器的性能等级是否进入到一个新的范围。或者，可供选择地，可由监控器进行该确定，再将变化通知给性能跟踪工具。

如果块 604 中处理器变化到一个新的性能等级，则性能跟踪工具确定系统状态（块 606），记录一个将性能等级与系统状态相关联的跟踪条目（块 608），并对跟踪条目打上时间戳（块 610）。之后，操作返回块 602，以确定是否存在退出情况。

从而，本发明通过提供一个将功率管理事件和作为结果的功率等

级与系统状态信息相关联的性能跟踪，解决了现有技术的缺点。使用性能跟踪，可以识别由功率管理产生的性能变化而引起的性能问题。此外，系统管理程序或操作系统通过跟踪功率管理事件的数量以及在每个性能阈值上的执行时间，为每个虚拟处理器或物理处理器维护统计数据。如果微处理器具有三种性能等级，可以分别记录每种模式的执行周期或其它统计数据。如果微处理器具有很大范围的性能，可按照范围来进行记录。

非常重要的是应该注意到，虽然本发明是针对完全功能的数据处理系统进行描述的，本领域普通技术人员应当理解，本发明的各个过程可分布于计算机可读的指令媒体或其它多种形式中，而且，不管实际用于实现这种分布的信号承载媒体的特定类型，本发明同样适用。计算机可读媒体的例子包括可记录类型媒体，例如，软盘、硬盘驱动器、RAM、CD-ROM、DVD-ROM，以及传输类型媒体，例如数字和模拟通信链路、使用射频和光波传输等传输形式的有线或无线通信链路。计算机可读媒体可采用编码的格式的形式，其中当在特定数据处理系统中实际使用时进行解码。

提供本发明的说明书的目的是为了说明和描述，而不是用来穷举或将本发明限制为所公开的形式。对本领域的一般技术人员而言，许多修改和变更都是显而易见的。选择并描述实施方式是为了更好地解释本发明的原理，其实际应用，并使本领域的其他一般技术人员理解带有各种修改的各种实施方式的本发明同样适用于设想的特定用途。

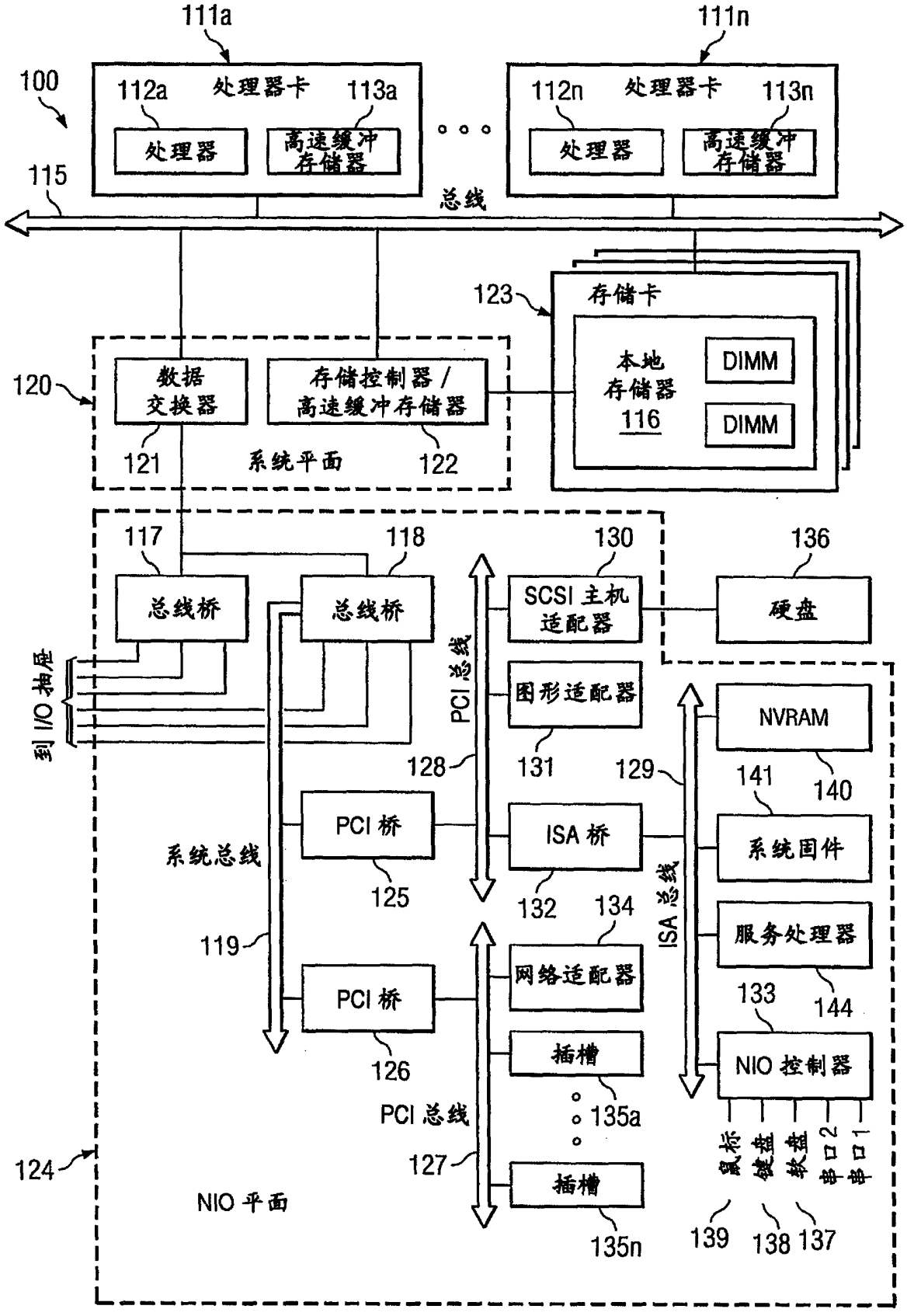

111a
111n
100
112a
处理器卡
113a
处理器
高速缓冲存储器
112n
处理器卡
113n
处理器
高速缓冲存储器
115
总线
123
存储卡
本地存储器
116
DIMM
DIMM
120
数据交换器
存储控制器/高速缓冲存储器
121
系统平面
122
117
118
130
136
总线桥
总线桥
SCSI 主机适配器
硬盘
到 I/O 插座
PCI 总线
图形适配器
NVRAM
128
131
129
141
140
PCI 桥
ISA 桥
系统固件
系统总线
125
132
134
ISA 总线
119
网络适配器
服务处理器
PCI 桥
133
144
插槽
NIO 控制器
126
PCI 总线
135a
串口2
串口1
124
127
插槽
NIO 平面
139
138
137
135n

图 1

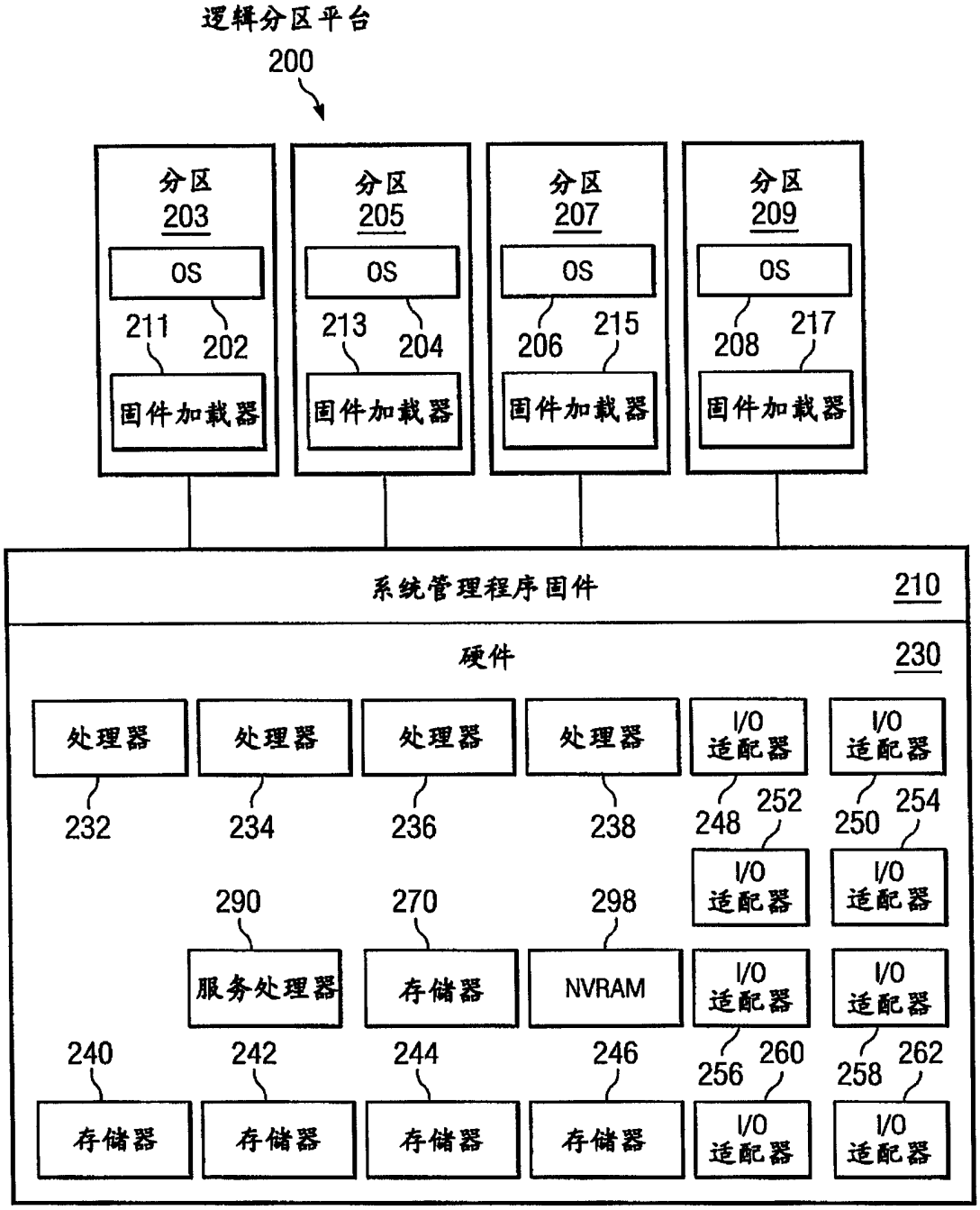

逻辑分区平台
200
分区
203
OS
211
202
固件加载器
分区
205
OS
213
204
固件加载器
分区
207
OS
206
215
固件加载器
分区
209
OS
208
217
固件加载器
系统管理程序固件
210
硬件
230
处理器
232
处理器
234
处理器
236
处理器
238
I/O
适配器
248
252
I/O
适配器
I/O
适配器
250
254
I/O
适配器
290
服务处理器
270
存储器
298
NVRAM
I/O
适配器
256
260
I/O
适配器
I/O
适配器
258
262
I/O
适配器
240
存储器
242
存储器
244
存储器
246
存储器

图 2

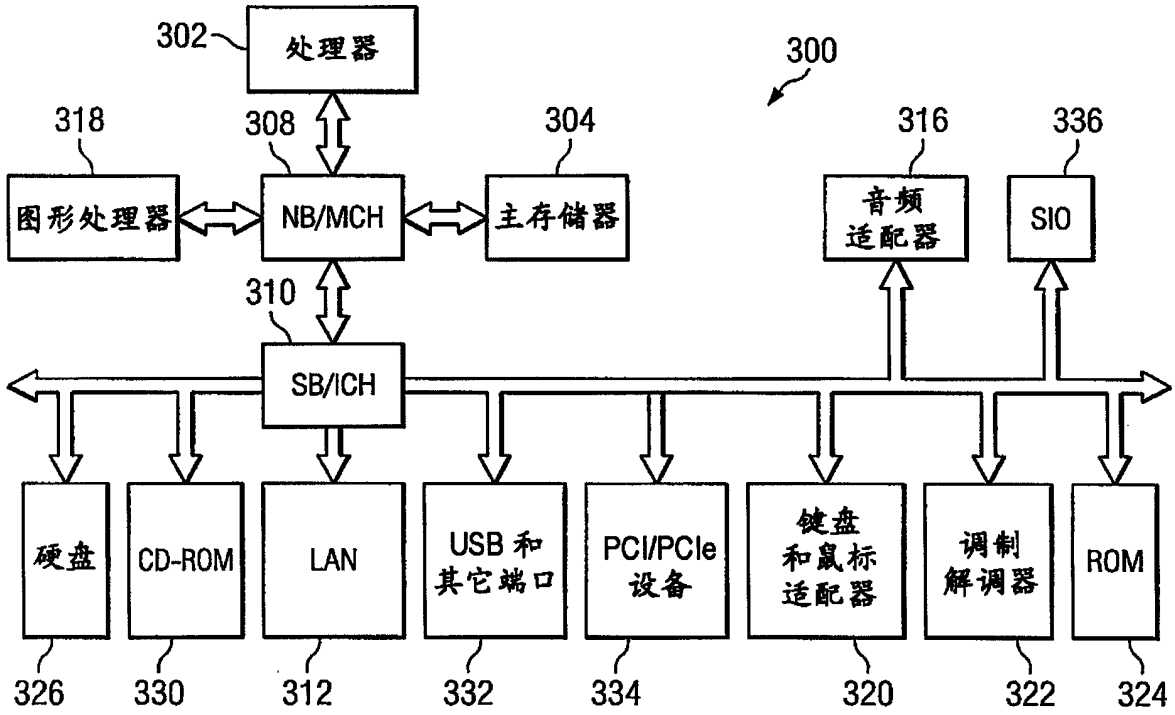

302
处理器
300
318
308
304
316
336
图形处理器
NB/MCH
主存储器
音频适配器
SIO
310
SB/ICH
硬盘
CD-ROM
LAN
USB 和其它端口
PCI/PCIe 设备
键盘和鼠标适配器
调制解调器
ROM
326
330
312
332
334
320
322
324

图 3

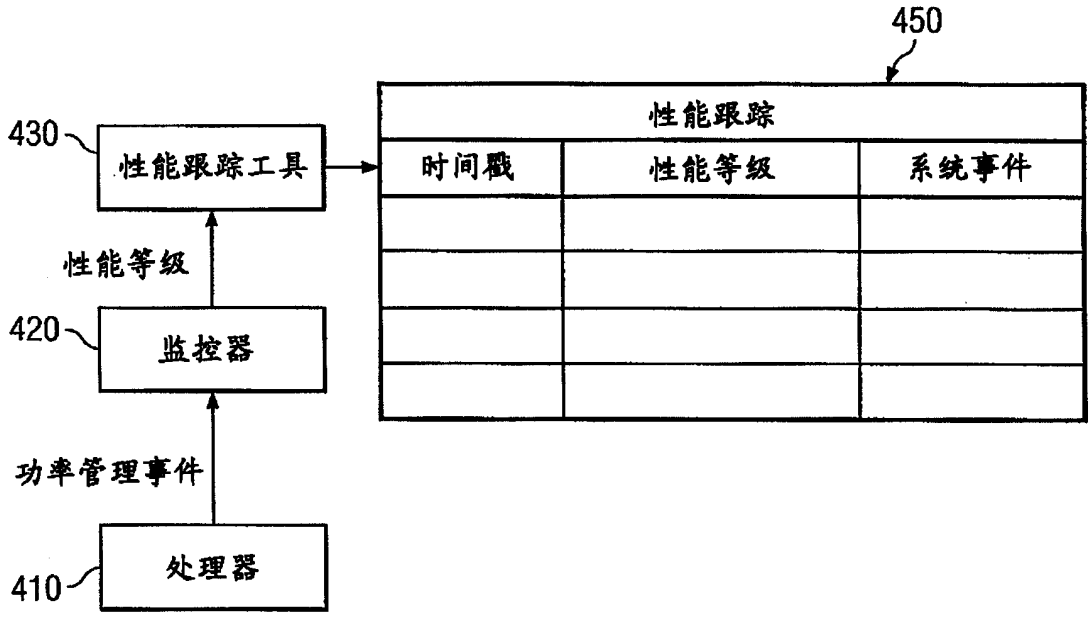

450
性能跟踪
时间戳
性能等级
系统事件
430
性能跟踪工具
性能等级
420
监控器
功率管理事件
410
处理器

图 4

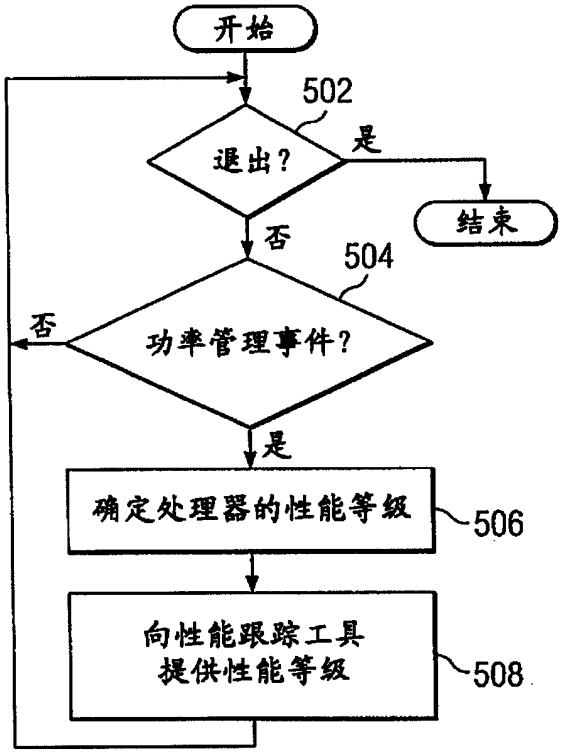

开始
502
退出?
是
结束
否
504
功率管理事件?
否
是
确定处理器的性能等级
506
向性能跟踪工具
提供性能等级
508

图 5

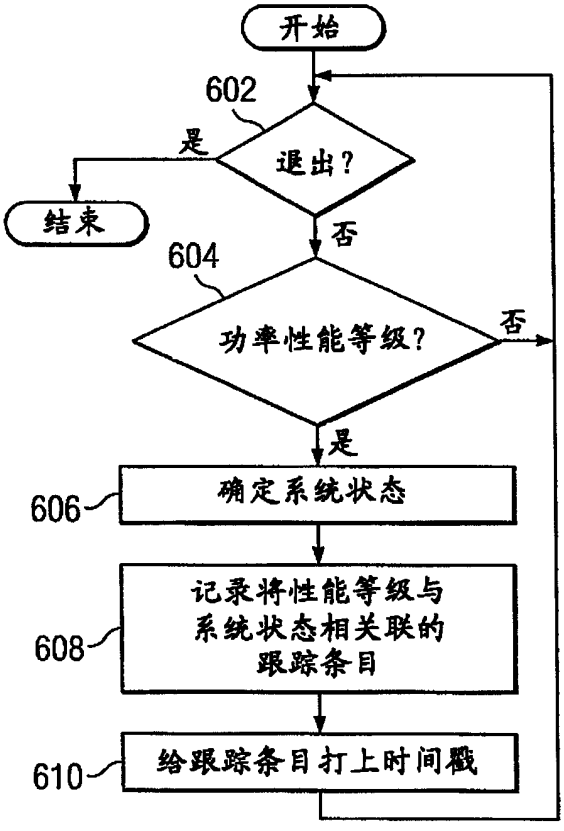

开始
602
退出?
是
结束
否
604
功率性能等级?
否
是
606
确定系统状态
608
记录将性能等级与
系统状态相关联的
跟踪条目
610
给跟踪条目打上时间戳

图 6