



(12) 发明专利

(10) 授权公告号 CN 101356486 B

(45) 授权公告日 2011. 10. 19

(21) 申请号 200780001298. 6

(22) 申请日 2007. 06. 13

(30) 优先权数据

11/425, 462 2006. 06. 21 US

(85) PCT申请进入国家阶段日

2008. 05. 06

(86) PCT申请的申请数据

PCT/EP2007/055819 2007. 06. 13

(87) PCT申请的公布数据

W02007/147761 EN 2007. 12. 27

(73) 专利权人 国际商业机器公司

地址 美国纽约

(72) 发明人 C·R·约翰斯 M·F·王

(74) 专利代理机构 北京市金杜律师事务所

11256

代理人 王茂华 李辉

(51) Int. Cl.

G06F 1/20(2006. 01)

G06F 1/32(2006. 01)

(56) 对比文件

WO 01/31495 A1, 2001. 05. 03, 全文.

CN 1573655 A, 2005. 02. 02, 全文.

CN 1153563 A, 1997. 07. 02, 全文.

CN 1148435 A, 1997. 04. 23, 全文.

审查员 石颖伟

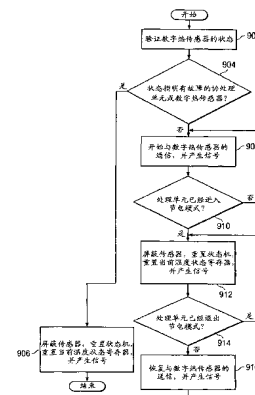
权利要求书 2 页 说明书 23 页 附图 10 页

(54) 发明名称

集成电路热管理系统

(57) 摘要

提供了用于管理热管理系统的状态机、计算机实现的方法、数据处理系统以及处理器。确定多个数字热传感器是有故障的还是工作的。响应于多个数字热传感器中的至少一个是工作的, 监控与工作的数字热传感器相关联的集成电路内的至少一个单元的节电模式。响应于该至少一个单元处于节电模式, 禁用工作的数字热传感器。



1. 一种用于管理集成电路中的热管理系统的计算机实现的方法,所述方法由驻留在集成电路中的热管理控制状态机执行,包括:

确定多个数字热传感器中的任意传感器是被指示为有故障的还是工作的;

响应于所述多个数字热传感器中的至少一个数字热传感器是工作的,监控与所述工作的数字热传感器相关联的集成电路中的至少一个单元的节电模式;以及

响应于所述至少一个单元处于节电模式,禁用所述工作的数字热传感器,其中,通过屏蔽所述工作的数字热传感器来禁用所述工作的数字热传感器。

2. 根据权利要求 1 所述的方法,还包括:

响应于所述多个数字热传感器的至少一个数字热传感器是工作的,确定与所述工作的数字热传感器相关联的所述至少一个单元是否有故障的;以及

响应于与所述工作的数字热传感器相关联的所述至少一个单元是有故障的,禁用所述工作的数字热传感器。

3. 根据权利要求 1 所述的方法,其中,屏蔽所述工作的数字热传感器包括:

重置与所述工作的数字热传感器相关联的当前温度状态寄存器;

停止所述状态寄存器与所述工作的数字热传感器之间的通信;以及

产生信号,以指明所述工作的数字热传感器当前被屏蔽并且不应当参与热管理任务。

4. 根据权利要求 1 所述的方法,还包括:

监控所述至少一个单元的所述节电模式;以及

响应于所述至少一个单元退出所述节电模式,启用所述工作的数字热传感器。

5. 根据权利要求 4 所述的方法,其中,启用所述工作的数字热传感器还包括:

恢复当前温度状态寄存器与所述工作的数字热传感器之间的通信;

将与所述工作的数字热传感器相关联的所述状态寄存器重置为当前温度;以及

产生信号,以指明所述工作的数字热传感器应当参与热管理任务。

6. 根据权利要求 1 所述的方法,还包括:

响应于所述多个数字热传感器中的至少一个数字热传感器是有故障的,屏蔽所述有故障的数字热传感器。

7. 根据权利要求 6 所述的方法,其中,屏蔽所述有故障的数字热传感器包括:

重置与所述有故障的数字热传感器相关联的当前温度状态寄存器;

停止所述状态寄存器与所述有故障的数字热传感器之间的通信;以及

产生信号,以指明所述有故障的数字热传感器当前被屏蔽并且不应当参与热管理任务。

8. 一种用于管理集成电路中的热管理系统的设备,包括:

用于确定多个数字热传感器中的任意传感器是被指示为有故障的还是工作的装置;

用于响应于所述多个数字热传感器中的至少一个数字热传感器是工作的、监控与所述工作的数字热传感器相关联的集成电路中的至少一个单元的节电模式的装置;以及

用于响应于所述至少一个单元处于节电模式、禁用所述工作的数字热传感器的装置,其中通过屏蔽所述工作的数字热传感器来禁用所述工作的数字热传感器。

9. 根据权利要求 8 所述的设备,还包括:

用于响应于所述多个数字热传感器的至少一个数字热传感器是工作的,确定与所述工

作的数字热传感器相关联的所述至少一个单元是否有故障的装置 ;以及

用于响应于与所述工作的数字热传感器相关联的所述至少一个单元是有故障的,禁用所述工作的数字热传感器的装置。

10. 根据权利要求 8 所述的设备,其中,屏蔽所述工作的数字热传感器包括:

用于重置与所述工作的数字热传感器相关联的当前温度状态寄存器的装置;

用于停止所述状态寄存器与所述工作的数字热传感器之间的通信的装置 ;以及

用于产生信号以指明所述工作的数字热传感器当前被屏蔽并且不应当参与热管理任务的装置。

11. 根据权利要求 8 所述的设备,还包括:

用于监控所述至少一个单元的所述节电模式的装置 ;以及

用于响应于所述至少一个单元退出所述节电模式,启用所述工作的数字热传感器的装置。

12. 根据权利要求 11 所述的设备,其中,用于启用所述工作的数字热传感器的装置还包括:

用于恢复当前温度状态寄存器与所述工作的数字热传感器之间的通信的装置;

用于将与所述工作的数字热传感器相关联的所述状态寄存器重置为当前温度的装置 ;

以及

用于产生信号以指明所述工作的数字热传感器应当参与热管理任务的装置。

13. 根据权利要求 8 所述的设备,还包括:

用于响应于所述多个数字热传感器中的至少一个数字热传感器是有故障的,屏蔽所述有故障的数字热传感器的装置。

14. 根据权利要求 13 所述的设备,其中,用于屏蔽所述有故障的数字热传感器的装置包括:

用于重置与所述有故障的数字热传感器相关联的当前温度状态寄存器的装置;

用于停止所述状态寄存器与所述有故障的数字热传感器之间的通信的装置 ;以及

用于产生信号以指明所述有故障的数字热传感器当前被屏蔽并且不应当参与热管理任务的装置。

## 集成电路热管理系统

### 技术领域

[0001] 本发明一般涉及热管理在集成电路中的使用。

### 背景技术

[0002] 第一代异类 Cell 宽频引擎™(BE) 处理器是包括 64 位 Power·PC®处理器内核和 8 个单指令多数据 (SIMD) 协处理器内核的多核芯片,能够进行大量浮点处理,针对计算加强负荷和宽带富媒体 (rich media) 应用进行了最优化。高速存储器控制器和高带宽总线接口也集成在芯片上。Cell BE 突破性的多核架构和超高速通信能力使实时响应取得了巨大的进步,在一些情况下是最新 PC 处理器性能的 10 倍。Cell BE 是操作系统中立的,并且同时支持多个操作系统。这种类型处理器的应用范围包括具有显著增强真实感的下一代游戏系统、构成家庭中数字媒体和流式内容的网络中心 (hub) 的系统、用以开发和分发数字内容的系统、以及加速可视化和超级计算应用的系统。

[0003] 当今的多核处理器经常受限于热考虑。典型的解决方案包括冷却和功率管理。冷却可能是昂贵和 / 或难以封装的。功率管理通常是粗略的动作,其对大量 (如果不是全部的话) 处理器进行“调节 (throttling)”,以响应于所达到的热限制。例如热管理的其他技术通过仅对超过给定温度的单元进行调节来帮助解决这些粗略的动作。然而,大多数热管理技术影响应用的实时保证。因此,提供一种热管理解决方案将是有益的,该热管理解决方案为处理器提供即使在需要处理器调节的热条件下也能保证应用的实时特性的方法。在不满足实时保证的情况下,通知应用管理者,使得可以实现修正动作。

### 发明内容

[0004] 说明性实施方式的不同方面提供了用于管理热管理系统的计算机实现的方法、数据处理系统和处理器。说明性实施方式确定多个数字热传感器被指示为是故障的还是工作的 (functional)。响应于所述多个数字热传感器中的至少一个数字热传感器是工作的,说明性实施方式对与该工作的数字热传感器相关联的集成电路中的至少一个单元的节电模式进行监控。响应于至少一个单元处于节电模式,说明性实施方式禁用该工作的数字热传感器。

### 附图说明

[0005] 现在将结合附图仅通过示例的方式来描述本发明的实施方式,其中:

[0006] 图 1 描述了其中可以实现说明性实施方式的方面的数据处理系统网络的图示;

[0007] 图 2 描述了其中可以实现说明性实施方式的方面的数据处理系统的框图;

[0008] 图 3 描述了其中可以实现说明性实施方式的方面的 Cell BE 芯片的示例性框图;

[0009] 图 4 示出了根据说明性实施方式的示例性热管理系统;

[0010] 图 5 示出了根据说明性实施方式的温度图以及可能发生中断和动态调节的各点;

[0011] 图 6 示出了根据说明性实施方式的用于记录最大温度的操作的流程图;

[0012] 图 7 示出了根据其他说明性实施方式的用于通过性能监控来跟踪热数据的操作的流程图；

[0013] 图 8A 和图 8B 示出了根据其他说明性实施方式的用于高级热中断产生的操作的流程图；

[0014] 图 9 示出了根据其他说明性实施方式的用于在热管理系统中支持深度节电模式和部分良好的操作的流程图；

[0015] 图 10 示出了根据其他说明性实施方式的用于热调节控制特征的操作的流程图，该热调整控制特征支持独立于温度的热敏软件应用的实时测试；

[0016] 图 11 示出了根据其他说明性实施方式的用于对中断延迟影响最小的热调节控制实现的操作的流程图；

[0017] 图 12 示出了根据其他说明性实施方式的用于热调节滞后的操作的流程图；以及

[0018] 图 13 示出了根据其他说明性实施方式的用于热调节逻辑实现的操作的流程图。

### 具体实施方式

[0019] 说明性实施方式涉及管理热管理系统。提供图 1- 图 2，作为其中可以实现说明性实施方式的数据处理环境的示例性框图。应当意识到，关于可以实现本发明方面或实施方式的环境，图 1- 图 2 只是示例性的，而无意断言或暗示任何限制。可以对所描述的环境进行多种修改，而没有背离说明性实施方式的精神和范围。

[0020] 现在参考附图，图 1 描述了其中可以实现说明性实施方式的方面的数据处理系统网络的图示。网络数据处理系统 100 是其中可以实现说明性实施方式的计算机网络。网络数据处理系统 100 包括网络 102，其是用以在网络数据处理系统 100 内连接在一起的各种设备和计算机之间提供通信链路的媒介。网络 102 可以包括连接，例如有线、无线通信链路，或者光纤缆线。

[0021] 在所示的例子中，服务器 104 和服务器 106 以及存储单元 108 连接至网络 102。此外，客户端 110、112 和 114 连接至网络 102。这些客户端 110、112 和 114 可以例如是个人计算机或者网络计算机。在所描述的例子中，服务器 104 为客户端 110、112 和 114 提供数据，例如引导文件、操作系统镜像以及应用。在此例子中，客户端 110、112 和 114 是服务器 104 的客户端。网络数据处理系统 100 可以包括其他服务器、客户端以及未示出的其他设备。

[0022] 在所描述的例子中，网络数据处理系统 100 是具有网络 102 的互联网，其中网络 102 表示使用传输控制协议 / 互联网协议 (TCP/IP) 协议族相互通信的网络和网关的全球集合。互联网的核心是主节点或主机计算机之间的高速数据通信线路的主干，包括成千上万对数据和消息进行路由的商业、政府、教育和其他计算机系统。当然，网络数据处理系统 100 还可以实现为多种不同类型的网络，例如内联网、局域网 (LAN) 或广域网 (WAN)。图 1 意在示例，并非作为对不同说明性实施方式的架构限制。

[0023] 现在参考图 2，示出了其中可以实现说明性实施方式的若干方面的数据处理系统的框图。数据处理系统 200 是计算机的示例，例如图 1 中的服务器 104 或者客户端 110，实现说明性实施方式过程的计算机可用代码或指令可以位于所述计算机中。

[0024] 在所描述的例子中，数据处理系统 200 使用中心架构，包括北桥和存储器控制器中心 (MCH) 202 以及南桥和输出 / 输出 (I/O) 控制器中心 (ICH) 204。处理单元 206、主存储

器 208 以及图形处理器 210 连接至北桥和存储器控制器中心 202。图形处理器 210 可以通过加速图形端口 (AGP) 连接至北桥和存储器控制器中心 202。

[0025] 在所描述的例子中, LAN 适配器 212 连接至南桥和 I/O 控制器中心 204。音频适配器 216、键盘和鼠标适配器 220、调制解调器 222、只读存储器 (ROM) 224、硬盘驱动器 (HDD) 226、CD-ROM 驱动器 230、通用串行总线 (USB) 端口和其他通信端口 232 以及 PCI/PCIe 设备 234 通过总线 238 和总线 240 连接至南桥和 I/O 控制器中心 204。PCI/PCIe 设备可以包括例如用于笔记本电脑的以太网适配器、附加卡和 PC 卡。PCI 使用卡式总线控制器, 而 PCIe 不使用。ROM 224 例如可以是闪速二进制输入 / 输出系统 (BIOS)。

[0026] 硬盘驱动器 226 和 CD-ROM 驱动器 230 通过总线 240 连接至南桥和 I/O 控制器中心 204。硬盘驱动器 226 和 CD-ROM 驱动器 230 例如可以使用集成驱动电子设备 (IDE) 或者串行高级技术附件 (SATA) 接口。超级 I/O (SIO) 设备 236 可以连接至南桥和 I/O 控制器中心 204。操作系统在处理单元 206 上运行, 并且协调和提供对图 2 中数据处理系统 200 内的各种组件的控制。作为客户端, 操作系统可以是商业上可获得的操作系统, 例如 **Microsoft®**

**Windows®** XP (Microsoft 和 Windows 是微软公司在美国、其他国家、或两者的商标)。面向对象的编程系统, 例如 Java™ 编程系统, 可以与操作系统结合运行, 并提供从数据处理系统 200 上执行的 Java 程序或者应用对操作系统的调用 (Java 是 Sun Microsystems 公司在美国、其他国家、或两者的商标)。

[0027] 作为服务器, 数据处理系统 200 例如可以是 IBM eServer™ **pSeries®** 计算机系统, 运行高级交互式可执行 (**AIX®**) 操作系统或者 LINUX 操作系统 (eServer、pSeries 和 AIX 是国际商业机器公司在美国、其他国家、或两者的商标, 而 Linux 是 Linus Torvalds 在美国、其他国家、或两者的商标)。数据处理系统 200 可以是在处理单元 206 中包括多个处理器的对称多处理器 (SMP) 系统。可选地, 可以使用单处理器系统。

[0028] 用于操作系统、面向对象的编程系统以及应用或程序的指令位于例如硬盘驱动器 226 的存储设备上, 并可以载入主存储器 208, 用以由处理单元 206 执行。处理单元 206 使用计算机可用的程序代码来执行用于说明性实施方式的过程, 其中程序代码可以位于诸如主存储器 208、只读存储器 224 或者一个或多个外围设备 226 和 230 的存储器中。

[0029] 本领域的普通技术人员将会意识到, 图 1-图 2 中的硬件可以依赖于实现而变化。除了图 1-图 2 中所描述的硬件之外或是取代这些硬件, 可以使用其他内部硬件或者外围设备, 例如闪存、等同的非易失性存储器或者光盘驱动器等等。而且, 说明性实施方式的过程可以应用于多处理器数据处理系统。

[0030] 在一些说明性实施方式中, 数据处理系统 200 可以是个人数字助理 (PDA), 所述个人数字助理被配置有提供非易失性存储器的闪存, 用于存储操作系统文件和 / 或用户产生的数据。

[0031] 总线系统可以包括一个或多个总线, 例如图 2 中所示的总线 238 或者总线 240。当然, 可以使用任意类型的通信构造或架构来实现总线系统, 该构造或架构在与其附接的不同组件或设备之间提供数据传送。通信单元可以包括一个或多个用以传输和接收数据的设备, 例如图 2 的调制解调器 222 或者网络适配器 212。存储器例如可以是主存储器 208、只读存储器 224 或者例如可存在于图 2 中的北桥和存储器控制器中心 202 中的高速缓存。图

1-图 2 所描述的示例和上述示例并不意味着隐含架构限制。例如,除了采取 PDA 形式外,数据处理系统 200 还可以是平板式计算机、膝上型计算机或者电话设备。

[0032] 图 3 示出了其中可以实现说明性实施方式的若干方面的 Cell BE 芯片的示例性框图。Cell BE 芯片 300 是单片多处理器实现,其针对以例如游戏控制台、桌面系统和服务器的富媒体应用为目标的分布式处理。

[0033] 可以在逻辑上将 Cell BE 芯片 300 划分为以下功能性组件:Power PC<sup>®</sup>处理器元件 (PPE) 301,协处理器单元 (SPU) 310、311 和 312,以及存储器流控制器 (MFC) 305、306 和 307。尽管示出了协处理器元件 (SPE) 302、303 和 304 以及 PPE 301 作为例子,但可以支持任意类型的处理器元件。示例性 Cell BE 芯片 300 实现包括 1 个 PPE 301 和 8 个 SPE,尽管图 3 只示出了 3 个 SPE 302、303 和 304。CELL 处理器的 SPE 是被设计用以加速媒体和数据流负荷的新处理器架构的首次实现。

[0034] Cell BE 芯片 300 可以是片上系统,使得在单个微处理器芯片上提供图 3 中所描述的每个元件。此外,Cell BE 芯片 300 是异类处理环境,其中 SPU 310、311 和 312 中的每一个可以从系统中其他 SPU 中的每一个接收不同的指令。此外,用于 SPU 310、311 和 312 的指令集不同于 Power PC<sup>®</sup>处理器单元 (PPU) 308 的指令集,例如,PPU308 可以执行 Power<sup>™</sup>架构中基于精简指令集计算机 (RISC) 的指令,而 SPU 310、311 和 312 执行向量化指令。

[0035] 每个 SPE 包括一个 SPU 310、311 或 312,其具有自己的本地存储 (LS) 区域 313、314 或 315 以及专用 MFC 305、306 或 307,其中专用 MFC305、306 或 307 具有相关联的存储器管理单元 (MMU) 316、317 或 318,以保持和处理存储器保护和访问允许信息。尽管再次示出 SPU 作为例子,但可以支持任意类型的处理器单元。而且,Cell BE 芯片 300 实现了元件互连总线 (EIB) 319 和其他 I/O 结构,以便支持片上和外部数据流。

[0036] EIB 319 的作用是针对 PPE 301 和 SPE 302、303 和 304 的主片上总线。此外,EIB 319 接口连接至专用于片外访问的其他片上接口控制器。片上接口控制器包括存储器接口控制器 (MIC) 320 和 Cell BE 接口单元 (BEI) 323,其中 MIC 320 提供 2 个极限数据率 I/O(XIO) 存储器通道 321 和 322,以及 BEI 323 提供 2 个高速外部 I/O 通道以及用于 Cell BE 300 的内部中断控制。BEI 323 被实现为总线接口控制器 (BIC,标识为 BIC0 和 BIC1) 324 和 325 以及 I/O 接口控制器 (IOC) 326。两个高速外部 I/O 通道连接至 Redwood

**Rambus<sup>®</sup>** Asic Cell(RRAC) 接口的一个极 (polarity),为 Cell BE 300 提供灵活的输入和输出 (FlexIO\_0 和 FlexIO\_1) 353。

[0037] 每个 SPU 310、311 或 312 具有相应的 LS 区域 313、314 或 315 以及协执行单元 (SXU) 354、355 或 356。每个单个 SPU 310、311 或 312 可以执行仅仅来自其关联 LS 区域 313、314 或 315 内的指令 (包括数据装载和存储操作)。为此原因,经由 SPU 310、311 和 312 专用的 MFC 305、306 和 307,MFC 直接存储器访问 (DMA) 操作执行去往或来自系统中别处存储设备的全部所需数据传送。

[0038] 在 SPU 310、311 或 312 上运行的程序只使用 LS 地址来参考其自己的 LS 区域 313、314 或 315。然而,还针对每个 SPU 的 LS 区域 313、314 或 315 指派全部系统的存储器映射内的实际地址 (RA)。RA 是设备将对其进行响应的地址。在 Power PC<sup>®</sup>中,应用通过有效地址 (EA) 来参考存储器位置 (或设备),EA 继而被映射为针对该存储器位置 (或设备) 的

虚拟地址 (VA), VA 继而被映射为 RA。EA 是应用用来参考存储器和 / 或设备的地址。该映射允许操作系统分配比系统中物理存在的存储器更多的存储器 ( 也即, VA 所参考的术语虚拟存储器)。存储器映射是系统中的所有设备 ( 包括存储器) 及其相应 RA 的列表。存储器映射是实际地址空间的映射, 其中该映射识别设备或存储器将对其进行响应的 RA。

[0039] 这允许授权软件将 LS 区域映射为过程的 EA, 以实现一个 SPU 的 LS 和其他 SPU 的 LS 区域之间的直接存储器访问传送。PPE 301 还可以使用 EA 来直接访问任何 SPU 的 LS 区域。在 Power PC<sup>®</sup> 中, 存在 3 个状态 ( 问题、授权和管理者)。授权软件是在授权状态或管理者状态中运行的软件。这些状态具有不同的访问权限。例如, 授权软件可以访问用于将实际存储器映射为应用的 EA 的数据结构寄存器。问题状态是处理器在运行应用时通常处于的状态, 并通常被禁止访问系统管理资源 ( 例如用于映射实际存储器的数据结构)。

[0040] MFC DMA 数据命令总是包括一个 LS 地址和一个 EA。DMA 命令将存储器从一个位置复制到其他位置。在这种情况下, MFC DMA 命令在 EA 和 LS 地址之间复制数据。LS 地址直接寻址对应于 MFC 命令队列的相关联 SPU 310、311 或 312 的 LS 区域 313、314 或 315。命令队列是 MFC 命令的队列。存在保持来自 SPU 的命令的一个队列, 以及保持来自 PXU 或其他设备的命令的一个队列。然而, 可以将 EA 布置或映射为访问系统中的任何其他存储器存储区域, 包括其他 SPE302、303 和 304 的 LS 区域 313、314 和 315。

[0041] 主存储设备 ( 未示出) 由系统 ( 例如图 2 中所示系统) 中的 PPU308, PPE 301, SPE 302、303 和 304 以及 I/O 设备共享。保持在主存储器中的所有信息对于系统中的所有处理器和设备是可见的。程序使用 EA 来参考主存储器。由于 MFC 代理命令队列、控制和状态工具具有 RA, 并且使用 EA 来映射 RA, 因此 Power 处理器元件可以使用主存储设备和相关联 SPE 302、303 和 304 的本地存储设备之间的 EA 来发起 DMA 操作。

[0042] 作为示例, 当 SPU 310、311 或 312 上运行的程序需要访问主存储器时, SPU 程序产生具有适当 EA 和 LS 地址的 DMA 命令, 并将该 DMA 命令放置到 MFC 305、306 或 307 命令队列中。在 SPU 程序将命令放置到队列中之后, MFC 305、306 或 307 执行该命令, 并在 LS 区域和主存储器之间传送所需数据。MFC 305、306 或 307 为例如 PPE 301 的其他设备所产生的命令提供第二代理命令队列。MFC 代理命令队列通常用来在开始 SPU 之前将程序存储在本地存储设备中。MFC 代理命令还可以用于上下文存储操作。

[0043] EA 地址向 MFC 提供可由 MMU 转换为 RA 的地址。转换过程允许系统存储器的虚拟化, 以及实际地址空间中存储器和设备的访问保护。由于 LS 区域被映射为实际地址空间, 因此 EA 还可以寻址所有 SPU LS 区域。

[0044] Cell BE 芯片 300 上的 PPE 301 包括 64 位 PPU 308 和 Power PC<sup>®</sup> 存储子系统 (PPSS) 309。PPU 308 包含处理器执行单元 (PXU) 329、一级 (L1) 高速缓存 330、MMU 331 以及替换管理表 (RMT) 332。PPSS 309 包括可高速缓存接口单元 (CIU) 333、不可高速缓存单元 (NCU) 334、二级 (L2) 高速缓存 328、RMT 335 以及总线接口单元 (BIU) 327。BIU 327 将 PPSS 309 连接至 EIB 319。

[0045] SPU 310、311 或 312 和 MFC 305、306 和 307 通过具有容量的单向通道彼此通信。通道基本上是使用 34 个 SPU 指令之一来访问的 FIFO ; 读通道 (RDCH)、写通道 (WRCH) 和读通道计数 (RDHCNT)。RDHCNT 返回通道中的信息量。容量是 FIFO 的深度。通道传输去往



和来自 MFC 305、306 和 307 以及 SPU 310、311 和 312 的数据。BIU 339、340 和 341 将 MFC 305、306 和 307 连接至 EIB 319。

[0046] MFC 305、306 和 307 为 SPU 310、311 和 312 提供两个重要功能。MFC 305、306 和 307 在 SPU 310、311 或 312, LS 区域 313、314 或 315, 以及主存储器之间移动数据。此外, MFC 305、306 和 307 在 SPU 310、311 和 312 与系统中的其他设备之间提供同步工具。

[0047] MFC 305、306 和 307 实现具有 4 个功能性单元: 直接存储器访问控制器 (DMAC) 336、337 和 338 ; MMU 316、317 和 318 ; 原子单元 (ATO) 342、343 和 344 ; RMT 345、346 和 347 ; 以及 BIU 339、340 和 341。DMAC 336、337 和 338 维护和处理 MFC 命令队列 (MFCCMDQ) (未示出), 其包括 MFC SPU 命令队列 (MFC SPUQ) 和 MFC 代理命令队列 (MFC PrxyQ)。16 条目的 MFC SPUQ 处理从 SPU 通道接口接收的 MFC 命令。8 条目的 MFC PrxyQ 通过存储器映射输入和输出 (MMIO) 装载和存储操作来处理来自其他设备 (例如, PPR301 或 SPE 302、303 和 304) 的 MFC 命令。典型的直接存储器访问命令在 LS 区域 313、314 或 315 与主存储器之间移动数据。使用 MFC DMA 命令的 EA 参数对包括主存储器、本地存储设备以及具有 RA 的所有设备的主存储设备进行寻址。使用 MFC DMA 命令的本地存储参数来寻址相关联的逻辑存储设备。

[0048] 在虚拟模式中, MMU 316、317 和 318 提供地址转换和存储器保护工具, 以便处理来自 DMAC 336、337 和 338 的 EA 转换请求, 并发回转换后的地址。每个 SPE 的 MMU 维护分段旁视缓冲器 (SLB) 和转换旁视缓冲器 (TLB)。SLB 将 EA 转换为 VA, 并且 TLB 将 SLB 所产生的 VA 转换为 RA。EA 由应用使用, 并且通常是 32 位或者 64 位地址。不同的应用或者应用的多个副本可以使用相同的 EA 来参考不同的存储位置 (例如, 应用的两个副本, 每个具有相同的 EA, 将需要两个不同的物理存储器位置)。为了实现它, 首先将 EA 转换到大的多的 VA 空间中, VA 空间对于操作系统下运行的所有应用是公共的。由 SLB 来执行 EA 到 VA 的转换。继而使用 TLB 将 VA 转换为 RA, 其中 TLB 是包含 VA 到 RA 映射的映射表或者页表的高速缓存。该表由操作系统维护。

[0049] ATO 342、343 和 344 提供维护与系统中其他处理单元的同步所需的数据高速缓存级别。原子直接存储器访问命令为协处理器元件提供用以执行与其他单元的同步的装置。

[0050] BIU 339、340 和 341 的主要功能是为 SPE 302、303 和 304 提供到 EIB 的接口。EIB 319 在 Cell BE 芯片 300 上的所有处理器内核与附接至 EIB 319 的外部接口控制器之间提供通信路径。

[0051] MIC 320 提供 EIB 319 与 XIO 321 和 322 中一个或两个之间的接口。极限数据率 (XDR™) 动态随机访问存储器 (DRAM) 是由 **Rambus**® 提供的高速、高串行存储器。由 Rambus 提供的宏访问极限数据率动态随机访问存储器, 在此文档中称为 XIO 321 和 322。

[0052] MIC 320 在 EIB 319 上只是从属的。MIC 320 确认在其配置地址范围内的命令, 该地址范围对应于所支持的中心中的存储器。

[0053] BIC 324 和 325 管理从 EIB 319 到两个外部设备任意一个的片上或片外数据传送。BIC 324 和 325 可以与 I/O 设备交换非一致的业务量, 或者它可以将 EIB 319 扩展至其他设备, 该设备甚至可以是其他 Cell BE 芯片。在用来扩展 EIB 319 时, 总线协议维护 Cell BE 芯片 300 中的高速缓存和附接的外部设备中高速缓存之间的一致性, 其中附接的外部设备可以是其他 Cell BE 芯片。

[0054] IOC 326 处理在 I/O 接口设备中发起的、并以一致 EIB 319 为目的地的命令。I/O 接口设备可以是附接至 I/O 接口的任何设备,例如附接多个 I/O 设备或者以非一致方式访问的其他 Cell BE 芯片 300 的 I/O 桥芯片。IOC 326 还截取 EIB 319 上以位于 I/O 桥芯片或非一致 Cell BE 芯片 300 之中或之后的存储器映射寄存器为目的地的访问,并将它们路由到适当的 I/O 接口。IOC 326 还包括内部中断控制器 (IIC) 349 和 I/O 地址转换单元 (I/O Trans) 350。

[0055] 普适逻辑 351 是为 Cell BE 芯片 300 提供时钟管理、测试特征和加电序列的控制器。普适逻辑可以提供用于处理器的热管理系统。普适逻辑包含通过联合测试行动组 (JTAG) 或串行外围接口 (SPI) 接口与系统中其他设备的连接,其中 JTAG 和 SPI 是本领域公知的。

[0056] 尽管已经提供了如何可以实现不同组件的特定示例,但是这并不意味着限制其中可以使用说明性实施方式方面的架构。说明性实施方式的方面可以与任何多核处理器系统结合使用。

[0057] 在应用或软件的执行期间,Cell BE 芯片中区域的温度可能升高。若保持不加限制,温度可能升高到最大特定结温度之上,这导致不正确的操作或物理损坏。为了避免这些情况,Cell BE 芯片的数字热管理单元在操作期间监控并试图控制 Cell BE 芯片内的温度。数字热管理单元包括在此描述的热管理控制单元 (TMCU) 和 10 个分布式数字热传感器 (DTS)。

[0058] 8 个 SPE 的每一个中都具有一个传感器,一个传感器位于 PPE 中,并且一个靠近线性热二极管。线性热二极管是计算温度的片上二极管。将这些传感器定位在相关联单元内通常在大多数应用的执行期间经历最大温度升高的区域附近。热控制单元监控来自这些传感器的每一个的反馈。如果传感器的温度升到高于可编程点之上,则可以将热控制单元配置为导致对 PPE 或者一个或多个 SPE 的中断,并动态地对相关联 PPE 或 SPE 的执行进行调节。

[0059] 在可编程数目的周期内停止和运行 PPE 或 SPE 提供了所必需的调节。中断允许授权软件采取修正动作,同时动态调节试图在没有软件干预的情况下将宽带引擎芯片内的温度保持在可编程水平之下。授权软件将调节水平设置为等于或低于应用提供的推荐设置。每个应用可以是不同的。

[0060] 如果对 PPE 或 SPE 进行调节没有有效地管理温度并且温度继续升高,普适逻辑 351 在温度达到热过载温度(由可编程配置数据定义)时停止 Cell BE 芯片的时钟。热过载特征保护 Cell BE 芯片不受物理损坏。从这种情况中恢复需要硬重置。DTS 所监控的区域的温度不一定是相关联 PPE 或 SPE 内最热的点。

[0061] 图 4 示出了根据说明性实施方式的示例性热管理系统。可以将热管理系统实现为集成电路,例如图 3 的普适逻辑单元 351 提供的集成电路。热管理系统可以是专用集成电路、处理器、多处理器或者异类多核处理器。热管理系统在 10 个分布式 DTS(为了简单仅示出了 DTS404、406、408 和 410) 与热管理控制单元 (TMCU) 402 之间划分。SPU 传感器 440 中的 DTS 404 和 406 的每一个、PPU 传感器 442 中的 DTS 408、以及靠近线性热二极管(未示出)的传感器 444 中的 DTS 410 提供当前温度检测信号。该信号指明温度等于或低于 TMCU 402 所设置的当前温度检测范围。TMCU 402 使用来自 DTS 404、406、408 和 410 的信号的状态

态来继续跟踪每个 PPE 或 SPE 的 DTS 404、406、408 和 410 的温度。由于对温度进行跟踪，TMCU 402 将当前温度提供为数值，其代表相关联 PPE 或 SPE 内的温度。对各个传感器进行校准的制造设置内部校准存储 428。

[0062] 除了上述 TMCU 402 的元件之外，TMCU 402 还包含多路复用器 446 和 450、工作寄存器 448、比较器 452 和 454、串行器 456、热管理控制状态机 458 以及数据流 (DF) 单元 460。多路复用器 446 和 450 对各种传出和传入信号进行合成，以用于在单个媒介上的传输。工作寄存器 448 保持 TMCU 402 中执行的乘法的结果。比较器 452 和 454 提供两个输入的比较功能。比较器 452 是大于等于比较器。比较器 454 是大于比较器。串行器 456 将来自源的低速并行数据转化为高速串行数据以用于传输。串行器 456 与 SPU 传感器 440 上的解串行器 462 和 464 结合工作。解串行器 462 和 464 将接收到的高速串行数据转化为低速并行数据。热管理控制状态机 558 启动 TMCU 402 的内部初始化。DF 单元 460 控制去往和来自热管理控制状态机 458 的数据。

[0063] 可以将 TMCU 402 配置为使用中断逻辑 416 来导致 PPE 的中断，以便使用调节逻辑 418 来动态地对 PPE 或 SPE 的执行进行调节。

[0064] TMCU 402 对代表温度的数值与可编程中断温度以及可编程调节点进行比较。每个 DTS 具有独立的可编程中断温度。如果温度在所编程的中断温度范围内，如果 PPE 是启用的，则 TMCU 402 产生对 PPE 的中断。取决于方向位，在温度高于或低于所编程水平的情况下产生中断，这将在稍后讨论。此外，第二可编程中断温度可以导致到系统控制器的注意信号的断言。系统控制器在系统平面上，并通过 SPI 端口与 Cell BE 连接。

[0065] 如果与 PPE 或 SPE 相关联的 DTS 所检测到的温度等于或高于调节点，则 TMCU 402 对 PPE 或者一个或多个 SPE 的执行进行调节，这是通过独立地启动和停止该 PPE 或 SPE 来完成的。软件可以使用热管理寄存器来控制调节的比率和频率，热管理寄存器例如为热管理停止时间寄存器和热管理比例寄存器。

[0066] 图 5 示出了根据说明性实施方式的温度图以及可能发生中断和动态调节的各点。在图 5 中，线 500 可以代表 PPE 或 SPE 的温度。如果 PPE 或 SPE 正常运行，则在标记为“N”的区域中没有调节。当 PPE 或 SPE 的温度达到调节点时，TMCU 开始对相关 PPE 或 SPE 的执行进行调节。将发生调节的区域标记为“T”。当 PPE 或 SPE 的温度下降到末端调节点之下时，执行返回正常操作。

[0067] 如果温度出于任何原因而继续上高并达到完全调节点或之上的温度，则 TMCU 402 停止 PPE 或 SPE，直到温度下降到完全调节点之下。将 PPE 或 SPE 停止的区域标记为“S”。将在温度达到或超过完全调节点时停止 PPE 或 SPE 称为内核停止保护。

[0068] 在此示例性说明中，中断温度设置在调节点之上；因此，TMCU402 产生中断，其是对软件的通知：因为温度处于或者仍然在内核停止温度之上，因此停止相应的 PPE 或 SPE；假设热中断屏蔽寄存器 (TM\_ISR) 被设为活动，参见图 4 中的 422，则允许 PPE 或 SPE 在未决中断期间继续。如果动态调节被禁用，则授权软件管理热条件。不对热条件进行管理可能导致相关联 PPE 或 SPE 的不适当操作或由热过载功能引起的热关机。

[0069] 返回图 4，热传感器状态寄存器包括热传感器当前温度状态寄存器 412 和热传感器最大温度状态寄存器 414。这些寄存器允许软件读取每个 DTS 的当前状态，确定在一段时间中达到的最高温度，并在温度达到可编程温度时导致中断。热传感器状态寄存器具有相

关联的实际地址分页,该分页可被标记为管理者授权的。

[0070] 热传感器当前温度状态寄存器 412 包含针对每个 DTS 的当前温度的编码或数字值。由于传感器的温度检测的延迟、读取这些寄存器的延迟、以及正常的温度波动,这些寄存器中所报告的温度是较早时间点的温度,并且可能没有反映软件接收到数据时的实际温度。由于每个传感器具有专用的控制逻辑,DTS 404、406、408 和 410 中的控制逻辑并行地采样所有传感器。TMCU 402 在采样周期的结尾更新热传感器当前温度状态寄存器 412 的内容。TMCU 402 将热传感器当前温度状态寄存器 412 中的值改变为当前温度。TMCU 402 每个 SenSampTime 周期对新的当前温度进行轮询。SenSampTime 配置字段控制采样周期的长度。

[0071] 热传感器最大温度状态寄存器 414 包含每个传感器从上次读取热传感器最大温度状态寄存器 414 开始所达到的经过数字编码的最大温度。由软件或者任何片外设备(例如片外设备 472 或片外 I/O 设备 474) 读取这些寄存器,这使 TMCU 402 将每个传感器的当前温度复制到寄存器中。在读取之后,TMCU 402 从这点开始继续跟踪最大温度。每个寄存器的读取是独立的。一个寄存器的读取不影响其他寄存器的内容。

[0072] 每个传感器具有专用控制逻辑,因此 DTS 404、406、408 和 410 中的控制逻辑并行地采样所有传感器。TMCU 402 将热传感器最大温度状态寄存器 414 中的值改变为当前温度。TMCU 402 每个 SenSampTime 周期对新的当前温度进行轮询。SenSampTime 配置字段控制采样周期的长度。

[0073] 中断逻辑 416 中的热传感器中断寄存器控制对 PPE 的热管理中断的产生。这组寄存器包括热传感器中断温度寄存器 420 (TS\_ITR1 和 TS\_ITR2)、热传感器中断状态寄存器 422 (TS\_ISR)、热传感器中断屏蔽寄存器 424 (TS\_IMR) 以及热传感器全局中断温度寄存器 426 (TS\_GITR)。热传感器中断温度寄存器 420 和热传感器全局中断温度寄存器 426 包含针对导致对 PPE 的热管理中断的温度的编码。

[0074] 当针对传感器的热传感器当前温度状态寄存器 412 中以数字格式经过编码的温度大于或等于热传感器中断温度寄存器 420 中相应传感器的中断温度编码时,TMCU 402 设置热传感器中断状态寄存器 422 中的相应状态位 (TS\_ISR[Sx])。当针对任何传感器的热传感器当前温度状态寄存器 412 中的温度编码大于或等于热传感器全局中断温度寄存器 426 中的全局中断温度编码时,TMCU 402 设置热传感器中断状态寄存器 422 中的相应状态位 (TS\_ISR[Gx])。

[0075] 如果任何热传感器中断温度状态寄存器 422 的位 (TS\_ISR[Sx]) 被设置,并且热传感器中断屏蔽寄存器 424 中的相应屏蔽位 (TS\_IMR[Mx]) 也被设置,则 TMCU 402 断言对 PPE 的热管理中断信号。如果任何热传感器中断状态寄存器 422 的位 (TS\_ISR[Gx]) 被设置,并且热传感器中断屏蔽寄存器 424 中的相应屏蔽位 (TS\_IMR[Cx]) 也被设置,则 TMCU 402 断言对 PPE 的热管理中断信号。

[0076] 为了清除中断条件,授权软件应当将热传感器中断屏蔽寄存器中的任何相应屏蔽位设置为“0”。为了启用热管理中断,授权软件保证温度低于对于相应的寄存器的中断温度,并继而执行以下序列。在温度不低于中断温度时启动中断可能导致产生立刻的热管理中断。

[0077] 1. 将“1”写到热传感器中断状态寄存器 422 中的相应状态位。

[0078] 2. 将“1”写到热传感器中断屏蔽寄存器 424 中的相应状态位。

[0079] 热传感器中断温度寄存器 420 包含用于位于 SPE、PPE 中以及靠近线性热二极管的传感器的中断温度水平。TMCU 402 对该寄存器中经过编码的中断温度水平与热传感器当前温度状态寄存器 412 中的相应中断温度编码进行比较。这些比较的结果产生热管理中断。每个传感器的中断温度水平是独立的。

[0080] 除了在热传感器中断温度寄存器 420 中设置的独立的中断温度水平之外，热传感器全局中断温度寄存器 426 包含第二中断温度水平。该水平适用于 Ce11 BE 芯片中的所有传感器。TMCU 402 对该寄存器中经过编码的全局中断温度水平与针对每个传感器编码的当前温度编码进行比较。这些比较的结果产生热管理中断。

[0081] 全局中断温度的目的是提供 Ce11 BE 芯片中温度升高的早期指示。授权软件和系统控制器可以使用该信息来启动控制温度的动作，例如，提高风扇速度，跨过单元对应用软件进行再平衡，等等。

[0082] 热传感器中断状态寄存器 422 标识哪些传感器满足中断条件。中断条件是指每个热传感器中断状态寄存器 422 位所具有的特定条件，在满足该条件时可能发生中断。仅在相应的屏蔽位被设置的情况下向 PPE 给出实际中断。

[0083] 热传感器中断状态寄存器 422 包含三组状态位：数字传感器全局阈值中断状态位 (TS\_ISR[Gx])、数字传感器阈值中断状态位 (TS\_ISR[Sx]) 以及数字传感器全局低于阈值中断状态位 (TS\_ISR[Gb])。

[0084] 当热传感器当前温度状态寄存器 412 中针对传感器编码的温度大于或等于热传感器中断温度寄存器 420 中相应传感器的中断温度编码、并且相应的方向位热传感器中断屏蔽寄存器 424  $TM\_IMR[Bx] = '0'$  时，TMCU 402 设置热传感器中断状态寄存器 422 中的状态位 (TS\_ISR[Sx])。此外，当热传感器当前温度状态寄存器 412 中针对传感器的温度编码低于热传感器中断温度寄存器 420 中相应寄存器的中断温度编码、并且相应的方向位热传感器中断屏蔽寄存器 424  $TM\_IMR[Bx] = '1'$  时，TMCU 402 设置热传感器中断状态寄存器 422  $TS\_ISR[Sx]$ 。

[0085] 当任何参与传感器的当前温度大于或等于热传感器全局中断温度寄存器 426 以及热传感器中断屏蔽寄存器 424  $TS\_IMR[B_c]$  的温度时，TMCU 402 将热传感器中断状态寄存器 422  $TS\_ISR[Gx]$  设置为“0”。各个热传感器中断状态寄存器 422  $TS\_ISR[Gx]$  位指明哪些相应的传感器满足这些条件。

[0086] 当热传感器中断屏蔽寄存器 424  $TS\_IMR[C_x]$  中的所有参与传感器的当前温度都低于热传感器全局中断温度寄存器 426 以及热传感器中断屏蔽寄存器 424  $TS\_IMR[B_c]$  的温度时，TMCU 402 将热传感器中断状态寄存器 422  $TS\_ISR[Gb]$  设置为“1”。由于所有参与传感器的当前温度都低于热传感器全局中断温度寄存器 426 的温度，因此只存在一个状态位热传感器中断状态寄存器 422 (TS\_ISR[Gb]) 用于全局低于阈值中断条件。

[0087] 一旦热传感器中断状态寄存器 422 中的状态位 (TS\_ISR[Sx]、[Gx] 或 [Gb]) 被设置为“1”，则 TMCU 402 维持该状态，直到授权软件将该状态位重置为“0”。授权软件通过将“1”写到热传感器中断状态寄存器 422 中的相应位来将状态位重置为“0”。

[0088] 热传感器中断屏蔽寄存器 424 包含用于各个传感器的两个字段以及用于全局中断条件的多个字段。中断条件是指每个热传感器中断屏蔽寄存器 424 位所具有的特定条

件,在满足该条件时可能发生中断。仅在相应的屏蔽位被设置的情况下向 PPE 给出实际中断。

[0089] 用于各个传感器的两个热传感器中断屏蔽寄存器数字热阈值中断字段是 TS\_IMR[Mx] 和 TS\_IMR[Bx]。热传感器中断屏蔽寄存器 424TS\_IMR[Mx] 屏蔽位阻止中断状态位产生对 PPE 的热管理中断。热传感器中断屏蔽寄存器 424 TS\_IMR[Bx] 方向位设置用于中断条件的温度方向,其在热传感器中断温度寄存器 420 中的相应温度之上或之下。通过将热传感器中断屏蔽寄存器 424 TS\_IMR[Bx] 设置为“1”来将用于中断条件的温度设置为低于热传感器中断温度寄存器 420 中的相应温度。通过将热传感器中断屏蔽寄存器 424 TS\_IMR[Bx] 设置为“0”来将用于中断条件的温度设置为等于或高于热传感器中断温度寄存器 420 中的相应温度。

[0090] 用于全局中断条件的热传感器中断屏蔽寄存器 424 字段是 TS\_IMR[Cx]、TS\_IMR[B<sub>c</sub>]、TS\_IMR[Cgb] 和 TS\_IMR[A]。热传感器中断屏蔽寄存器 424 TS\_IMR[Cx] 屏蔽位阻止全局阈值中断,并选择哪些传感器参与全局低于阈值中断条件。热传感器中断屏蔽寄存器 424 TS\_IMR[B<sub>c</sub>] 方向位选择用于全局中断条件的温度方向。热传感器中断屏蔽寄存器 424 TS\_IMR[Cgb] 屏蔽位阻止全局低于阈值中断。热传感器中断屏蔽寄存器 424 TS\_IMR[A] 断言向系统控制器的注意。注意是给系统控制器的信号,其指明普适逻辑需要注意或者具有针对系统控制器的状态。可以将注意映射为系统控制器中的中断。系统控制器在系统平面上,并且通过 SPI 端口与 Cell 宽带引擎连接。

[0091] 通过将热传感器中断屏蔽寄存器 424 TS\_IMR[B<sub>c</sub>] 设置为“1”来设置在热传感器中断屏蔽寄存器 424 TS\_IMR[Cx] 中设置的所有参与传感器的温度低于全局中断温度水平时发生全局中断条件的温度范围。通过将热传感器中断屏蔽寄存器 424 TS\_IMR[B<sub>c</sub>] 设置为“0”来设置在任意参与传感器的温度大于或等于热传感器全局中断温度寄存器 426 中的相应温度时发生全局中断条件的温度范围。如果将热传感器中断屏蔽寄存器 424 TS\_IMR[A] 设置为“1”,则当任意热传感器中断屏蔽寄存器 424 TS\_IMR[Cx] 位及其相应的热传感器中断状态寄存器 422 状态位 (TS\_ISR[Gx]) 都被设置位“1”时,TMCU 402 断言注意 (attention)。此外,当热传感器中断屏蔽寄存器 424TS\_IMR[Cgb] 和热传感器中断状态寄存器 422 TS\_ISR[Gb] 都被设置为“1”时,TMCU 402 断言注意。

[0092] 当任意热传感器中断屏蔽寄存器 424 TS\_IMR[Mx] 位及其相应的热传感器中断状态寄存器 422 状态位 (TS\_ISR[Sx]) 都被设置为“1”时,TMCU 402 向 PPE 给出热管理中断。当任意热传感器中断屏蔽寄存器 424 TS\_IMR[Cx] 位及其相应的热传感器中断状态寄存器 422 状态位 TS\_ISR[Gx] 都被设置为“1”时,TMCU 402 产生热管理中断。此外,当热传感器中断屏蔽寄存器 424 TS\_IMR[Cgb] 和热传感器中断状态寄存器 422 TS\_TSR[Gb] 都被设置为“1”时,TMCU 402 向 PPE 给出热管理中断。

[0093] 调节逻辑 418 中的动态热管理寄存器包含用于控制 PPE 或 SPE 的执行调节的参数。动态热管理寄存器是一组寄存器,包含热管理控制寄存器 430(TM\_CR1 和 TM\_CR2)、热管理调节点寄存器 432(TM\_TPR)、热管理停止时间寄存器 434(TM\_STR1 和 TM\_STR2)、热管理调节比例寄存器 436(TM\_TSR) 以及热管理系统中断屏蔽寄存器 438(TM\_SIMR)。

[0094] 热管理调节点寄存器 432 设置用于传感器的调节温度点。在热管理调节点寄存器 432 中可以设置两个独立的调节温度点,ThrottlePPE 和 ThrottleSPE,一个用于 PPE,

一个用于 SPE。而且,该寄存器中还包含用于禁用调节和停止 PPE 或 SPE 的温度点。当温度等于或高于调节点时,PPE 或 SPE 的执行调节开始。当温度下降到禁用调节的温度 (TM\_TPR[EndThrottlePPE/EndThrottleSPE]) 之下时,调节停止。如果温度达到完全调节或停止温度 (TM\_TPR[FullThrottlePPE/FullThrottleSPE]),TMCU 402 停止 PPE 或 SPE 的执行。热管理控制寄存器 430 控制调节行为。

[0095] 热管理停止时间寄存器 434 和热管理调节比例寄存器 436 控制调节的频率和量。当温度达到调节点时,TMCU 402 在热管理停止时间寄存器 434 中的相应值中的停止时间乘以热管理比例寄存器 436 中的相应比例值所指定的时钟数目内停止相应的 PPE 或 SPE。TMCU 402 继而允许 PPE 或 SPE 在运行时间乘以相应比例值所指定的时钟数目内运行,其中运行时间是基于实现的固定时间量减去停止时间的差。热管理比例寄存器 436 中的可编程比例值是用于停止时间和运行时间两者的乘数。一个例子可以是 (停止 × 比例)/(运行 × 比例)。内核停止的时间百分比保持相同,但是可以增加周期或者降低频率。该序列继续,直到温度下降到禁用调节 (TM\_TPR[EndThrottlePPE/EndThrottleSPE]) 之下。

[0096] 热管理系统中断屏蔽寄存器 438 选择哪些 PPE 中断将导致 TMCU402 禁用调节。当这些中断仍然未决并且屏蔽仍然选择该未决中断时,TMCU 402 将继续阻止调节。如果屏蔽被解选择或中断不再未决,TMCU 402 将不再阻止调节。

[0097] 热管理控制寄存器 430 针对每个 PPE 或 SPE 独立地设置调节模式。控制位在两个寄存器之间分割。下面是可以针对每个 PPE 或 SPE 独立设置的 5 种不同模式。

[0098] 禁用动态调节 (包括内核停止保护)。

[0099] 正常操作 (启用动态调节和内核停止保护)。

[0100] PPE 或 SPE 始终调节 (启用内核停止保护)。

[0101] 禁用内核停止保护 (启用动态调节,并且禁用内核停止保护)。

[0102] PPE 或 SPE 始终调节,并且禁用内核停止保护。

[0103] 对于正在运行应用或操作系统的 PPE 或 SPE,授权软件应当将控制位设置为正常操作。如果 PPE 或 SPE 没有运行应用代码,授权软件应当将控制位设置为禁用。“PPE 或 SPE 总是调节”模式意在应用开发。这些模式有助于确定应用是否可以在极限调节条件下进行操作。只有当授权软件主动地管理热事件时,才能够允许 PPE 或 SPE 与动态调节或内核停止保护禁用结合执行。

[0104] 热管理系统中断屏蔽寄存器 438 控制哪些 PPE 中断导致热管理逻辑临时停止对 PPE 进行调节。在中断未决时,TMCU 402 临时挂起针对两个线程的调节,而不论中断的目标线程。当线程不再未决时,只要调节条件仍然存在,则可以恢复调节。从不根据系统中断条件来禁用 SPE 的调节。可以覆盖调节条件的 PPE 中断条件如下:

[0105] 外部

[0106] 递减

[0107] 管理者递减

[0108] 系统错误

[0109] 热管理

[0110] 热管理调节点寄存器 432 包含 PPE 或 SPE 的执行调节开始和结束的已编码温度点。该寄存器还包含对 PPE 或 SPE 的执行完全调节的已编码温度点。

[0111] 软件使用热管理调节点寄存器中的值来设置三个温度点,以用于在三个热管理状态之间改变:正常运行(N)、PPE或SPE调节(T)以及PPE或SPE停止(S)。TMCU 402支持用于PPE和SPE的独立温度点。

[0112] 当热传感器当前温度状态寄存器412中经过编码的传感器当前温度等于或大于调节温度(ThrottlePPE/ThrottleSPE)时,相应PPE或SPE的执行调节开始(如果启用的话)。执行调节继续,直到相应传感器的已编码当前温度小于结束调节的已编码温度(EndThrottlePPE/EndThrottleSPE)。作为保护测量,如果经过编码的当前温度等于或大于完全调节点(FullThrottlePPE/FullThrottleSPE),则TMCU 402停止相应的PPE或SPE。

[0113] 热管理停止时间寄存器434控制应用于热管理调节状态中的指定PPE或SPE的调节量。热管理停止时间寄存器434中由软件设置的值代表停止内核的时间量相对于允许内核运行的时间量(停止/运行)或者说停止内核的时间百分比。热管理调节比例寄存器436控制PPE或SPE停止和运行的实际时钟数目(NC1k)。

[0114] 热管理调节比例寄存器436控制PPE或SPE在热管理调节状态期间停止和运行的实际周期数。该寄存器中的值是配置环设置TM\_Config[MinStopSPE]的倍数。以下等式计算停止和运行周期的实际数目:

[0115] SPE运行和停止时间:

[0116] 
$$PEStopTime = (TMSTR1[StopCore(x)] * TM\_Config[MinStopSPE]) * mTM\_TSR[ScaleSPE]$$

[0117] 
$$SPERunTime = (32 - TM\_STR1[StopCore(x)]) * TM\_Config[MinStopSPE] * TM\_TSR[ScaleSPE]$$

[0118] Power PC<sup>®</sup>元件运行和停止时间:

[0119] 
$$PPEStopTime = (TM\_STR2[StopCore(8)] * TM\_Config[MinStopPPE]) * TM\_TSR[ScalePPE]$$

[0120] 
$$PPERunTime = (32 - TM\_STR2[StopCore(8)]) * TM\_Config[MinStopPPE] * TM\_TSR[ScalePPE]$$

[0121] 运行和停止时间可以由中断以及对各种热管理寄存器进行写操作的授权软件变更。

[0122] 片上性能监控器466可以提供性能监控,其可以跟踪由例如DTS404、406、408和410的温度传感设备所提供的热数据。热数据可以存储在存储器470中,或是写入片外设备472(例如图2的主存储器208)或片外I/O设备474(例如图2的南桥和输入/输出(I/O)控制器中心(ICH)204)。位于性能监控器466中的控制器468控制将热数据发送到哪里的决定。

[0123] 尽管以下描述涉及一个指令流和一个处理器,但是,指令流可以是指令流的集合,并且处理器可以是处理器的集合。换言之,集合可以是刚好单个指令流和单个处理器,或者是两个或更多指令流和处理器。

[0124] 使用上文描述的架构,可以对Cell BE芯片的热管理和热调节进行多种改进以及添加可编程性。这些改进中的一些和所添加的可编程性支持关键的特征以及其他增强可用性的特征。

[0125] 图6示出了根据说明性实施方式的用于记录最大温度的操作的流程图。操作开始



时,包含 Ce11 BE 芯片(例如图 3 的 Ce11 BE 芯片 300)的计算机系统启动或重置(步骤 602)。如前所述,Ce11 BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。对于每个 DTS,例如图 4 的 DTS 404、406、408 和 410,热管理系统包括一组最大温度状态寄存器和一组当前温度状态寄存器,例如图 4 的最大温度状态寄存器 414 和当前温度状态寄存器 412。当前温度状态寄存器存储从热管理控制状态机(例如图 4 的热管理控制状态机 458)上次感知其目标 DTS 以来该 DTS 的当前温度。最大温度状态寄存器存储从上次计算机系统最大温度状态寄存器中进行读取或者计算机系统重置以来其目标 DTS 的最大温度。可以使用任意数目的例如处理器、集成电路的设备来读取最大温度状态寄存器,或者通过使用串行外围接口(SPI)端口或联合测试行动组(JTAG)端口的设备来读取。虽然这样,通过 JTAG 端口读取没有引起重置。

[0126] 将下面的讨论说明性地限制为一个 DTS,计算机系统启动或重置(步骤 602)后的最大温度是零。只要热管理控制状态机感知到 DTS 的温度,热管理控制状态机就将所感知的 DTS 温度发送给比较器,例如图 4 的比较器 454(步骤 604)。比较器对所感知的温度与存储在针对该 DTS 的最大温度状态寄存器中的当前最大温度进行比较(步骤 606)。如果在步骤 606 处所感知的温度高于最大温度状态寄存器中存储的当前最大温度,则所感知的最大温度变为新的最大温度,并且热管理控制状态机将该新的最大温度记录在最大温度状态寄存器中(步骤 608)。换言之,热管理控制状态机重写或替换存储在最大温度状态寄存器中的当前最大温度。如果在步骤 606 所感知的温度低于或等于最大温度状态寄存器中所存储的当前最大温度,则最大温度状态寄存器保持最大温度状态寄存器中存在的当前最大温度(步骤 610)。

[0127] 最大温度状态寄存器中的当前最大温度停留在最大温度,直到计算机系统以读取请求的形式读取最大温度状态寄存器(步骤 612)或是计算机系统重置。如果当前最大温度未被读取,则操作返回步骤 604。如果在步骤 612 计算机读取了当前最大温度,则热管理控制状态机将当前最大温度重置为当前温度状态寄存器中的当前温度(步骤 614),操作返回步骤 604。

[0128] 对于该操作的示例,如果例如处理器内核或处理器本身的特定单元的 DTS 在一段时间内感知到的温度是:67°C、70°C、75°C、72°C 和 74°C,则最大温度状态寄存器中的最大温度将是 75°C。如果在 DTS 的第四感知之后,计算机系统发出读取请求,则返回的最大温度将是 75°C。然而,这时热管理控制状态机将最大温度重置为当前温度,并且在 DTS 执行的最后的感知之后,最大温度状态寄存器中的最大温度将是 74°C。

[0129] 因此,最大温度状态寄存器的目的是记录从最大温度寄存器最后被读取以来 DTS 所达到的最大温度。该最大温度信息辅助操作系统来确定在应用或程序的执行期间 DTS 所达到的最大温度,而无需持续对当前温度寄存器进行轮询。持续的轮询将影响系统的性能并因此可能影响最大温度。此外,轮询当前温度并未保证读取了最大温度。如果最大温度出现在对当前温度的读取之间,将出现这种情况。

[0130] 图 7 示出了根据其他说明性实施方式的通过性能监控来跟踪热数据的操作的流程图。如前所述,Ce11 BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。可以通过性能监控器,例如图 4 的性能监控器 466,来提供性能监控。性能监控可以对由温度感知设备(例如图 4 的 DTS 404、406、408 和 410)提供的热数据、其内部存储器(例如图 4 的

存储器 470) 中的热数据、写入主存储器 (例如如图 2 的主存储器 208 或图 4 的片外设备 472) 或写入 I/O 设备 (例如如图 2 的南桥和输入 / 输出 (I/O) 控制器中心 (ICH) 204 或图 4 的片外 I/O 设备 474) 的热数据进行跟踪。

[0131] 性能监控支持两种主要跟踪模式:固定时间段的跟踪或持续跟踪。热性能的跟踪可以是迹线,例如图 5 的迹线 500。性能监控还可以提供采样频率的配置,以控制两次连续采样之间的时间段。此外,可以使用热信息的压缩以增加采样间隔。一种压缩技术是仅在改变发生时存储热信息。也可以将相同的热采样的数目的计数随同热信息一起存储。这是有益的技术,因为热信息通常改变较慢。

[0132] 通过性能监控器的跟踪热数据操作开始时,热管理控制状态机,例如图 4 的热管理控制状态机 458,将性能监控器设置为跟踪模式(步骤 702)。说明性地,将以下讨论限制为一个 DTS,热管理控制状态机感知 DTS 的温度(步骤 704)并将所感知的 DTS 温度发送给当前温度状态寄存器和 / 或其他数据结构以便存储(步骤 706)。此时,热管理控制状态机确定性能监控器是否还在运行(步骤 708)。一旦性能监控器在步骤 702 中启动,性能监控器将在用户指定的时间段内运行,或是运行直到用户通过用户输入将其停止。然而,性能监控器也可以根据特定的热条件而停止。特定的热条件称为触发器,例如,在一组信号中查找特定条件的逻辑分析器。在软件调试中使用触发器是有益的。例如,用户可以建立性能监控器,以便在达到热条件时停止或检查停止系统。这可以允许用户准确地确定哪段代码或代码组合导致了热条件。在步骤 708,如果性能监控器还在运行,则操作返回步骤 704。

[0133] 返回步骤 708,如果性能监控器不再运行,则热管理控制状态机读取存储在存储器中的温度信息,并将所存储的信息图形化地显示给用户(步骤 710),此后操作结束。如箭头 712 所示,还可以在操作仍然进行时同时显示在步骤 706 发送给当前温度状态寄存器和 / 或其他数据结构的所感知的温度(步骤 710),而不是等待跟踪结束。

[0134] 因此,性能监控器跟踪由 DTS 提供的热数据。自动跟踪热数据消除了对软件持续轮询当前温度寄存器的需要。性能监控对于收集负荷的热数据而言是重要的,因为性能监控无需插入附加代码来轮询热数据,而插入附加代码来轮询热数据可能改变负荷的行为。换言之,性能监控提供了实时跟踪软件应用的热属性的非入侵式方法。将热信息发送给性能监控器的其他益处是能够根据预指定的热条件来触发或停止记录热信息。此外,还可以使用性能监控器在满足热条件时停止系统(或者检查停止)。这么做允许用户确定哪个代码段或代码段组合产生了热条件。用户继而可以重写该代码段或避免特定的组合,从而避免热事件。

[0135] 图 8A 和图 8B 示出了根据其他说明性实施方式的用于高级热中断产生的操作的流程图。如前所述,Cell BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。高级热中断产生是帮助操作系统处理热事件的又一特征。高级热中断逻辑是热管理控制单元(例如图 4 的 TMCU 402)的部分。热中断在存在热条件(也即,芯片温度升高到特定阈值以上)时提示操作系统。在这样的事件中,操作系统将采取修正动作以降低芯片温度。修正动作可以由软件中断处理机来处理,软件中断处理机是处理热条件并发起修正动作的代码段。操作系统继而在恢复正常操作之前等待热条件消失。这通常需要操作系统等待特定的时间量,继而轮询处理器温度以确定恢复正常操作是否安全。通过高级热中断产生,操作系统可以将中断设置为检测温度何时下降到特定阈值以下,由此消除轮询当前温度寄存器

的需要。图 4 的组合热传感器中断屏蔽寄存器 424(TS\_IMR) 和热传感器中断状态寄存器 422(TS\_ISR) 使操作系统处理热事件容易很多。

[0136] 可以在本地级别和全局级别执行高级热中断产生。换言之,高级热中断产生可以在特定 DTS 上单独地执行(本地),或是在所有 DTS 例如图 4 的 DTS 404、406、408 和 410 上执行(全局)。热传感器中断屏蔽寄存器的方向位是  $B_c$  和  $B_x$ 。中断方向限定产生中断的条件。中断可以在温度从低于中断温度变为等于或高于中断温度时产生,或者在温度从高于或等于中断温度变为低于中断温度时产生。热管理控制状态机通过中断屏蔽寄存器中的方向位  $B_c$  和  $B_x$  来标识条件。 $B_c$  是全局方向位。当  $B_c$  设置为“0”时,热管理控制状态机在任意 DTS 的温度大于等于全局中断温度时产生中断。当  $B_c$  设置为“1”时,热管理控制状态机在所有 DTS 的温度低于全局中断温度时产生中断。 $B_x$  是本地方向位,其中 X 是各个相关联的 DTS 的数目。当  $B_x$  设置为“0”时,热管理控制状态机在各个 DTS 的温度大于等于 DTS 中断温度时产生中断。当  $B_x$  设置为“1”时,热管理控制状态机在各个 DTS 的温度低于 DTS 中断温度时产生中断。热中断状态寄存器(TS\_ISR)记录哪个传感器导致了高级热中断。软件读取该寄存器,以确定发生了哪个条件、以及哪个或哪些传感器导致了中断。一旦软件读取了热中断状态寄存器中的状态位,热管理控制状态机便重置该状态位。

[0137] 因此,可以从全局和本地的视角示出用于高级热中断产生的操作。图 8A 描述全局高级热中断产生,以及图 8B 描述了本地高级热中断产生。在图 8A 的全局高级热中断产生中的操作开始时,热管理控制状态机将全局中断温度 T 设置为温度 T1,并将全局中断方向  $B_c$  设置为“0”(步骤 802)。热管理控制状态机感知 DTS 的温度(步骤 804)。热管理控制状态机确定是否有任何从 DTS 感知到的温度大于或等于温度 T1(步骤 806)。如果没有所感知的温度大于或等于温度 T1,则操作返回步骤 804。如果在步骤 806 任何一个所感知的温度大于或等于温度 T1,则热管理控制状态机产生中断,并设置热中断状态寄存器中相应的状态位,以记录哪个或哪些传感器导致了中断(步骤 808)。操作系统继而将服务该中断,并可以减慢处理器上的负荷或将处理器的一些负荷卸载(offload)给系统中的其他处理器。

[0138] 在产生中断之后,热管理控制状态机将全局中断温度 T 设置为温度 T2,并且将全局中断方向  $B_c$  设置为“1”(步骤 810)。应当将温度 T2 设置为小于等于温度 T1。热管理控制状态机再次感知 DTS 的温度(步骤 812)。热管理控制状态机确定来自 DTS 的所有感知到的温度是否都低于温度 T2(步骤 814)。如果没有所感知的温度低于温度 T2,则操作返回步骤 812。如果在步骤 814 处所有所感知的温度都低于温度 T2,则热管理控制状态机产生中断,并设置热中断状态寄存器中相应的状态位,以记录哪个或哪些传感器导致了中断(步骤 816)。这时,操作系统恢复正常操作现在是安全的。操作系统继而将服务该中断,并将系统恢复到正常操作。接下来,操作返回步骤 802,在此,将全局中断温度 T 设置为温度 T1,并将全局中断方向  $B_c$  设置为“0”。

[0139] 该操作的示例是:如果所有 DTS 都具有 80°C 的全局中断温度以及全局中断方向“0”。一旦例如处理器内核或处理器本身的相关单元的任何 DTS 感知到温度大于或等于 80°C,热管理控制状态机将产生中断,并设置热中断状态寄存器中相应的状态位,以记录哪个或哪些传感器导致了中断。操作系统继而将服务该中断,并且可以减慢处理器上的负荷或是将处理器的一些负荷卸载给系统中的其他处理器。而且,此时热管理控制状态机可以将全局中断温度重置为示例性的 77°C,并将全局中断方向设置为“1”。负荷将继续以慢模

式操作或是保持脱离处理器,直到 DTS 感知到所有 DTS 的温度低于 77°C。一旦热管理控制状态机确定所感知的温度低于 77°C,热管理控制状态机产生其他中断。热管理控制状态机将全局中断温度设置为 80°C,将全局中断方向设置为“0”,并且操作系统继而恢复负荷的正常操作。

[0140] 转到图 8B,将说明性实施方式限制为一个 DTS,但是说明针对每个 DTS 是一样的。在针对本地高级热中断产生的操作开始时,热管理控制状态机将本地中断温度 T 设置为温度 T3,并将本地中断方向 B<sub>x</sub> 设置为“0”(步骤 852)。热管理控制状态机感知 DTS 的温度(步骤 854)。热管理控制状态机确定来自 DTS 的所感知温度是否大于或等于温度 T3(步骤 856)。如果所感知温度没有大于或等于温度 T3,则操作返回步骤 854。如果所感知温度大于或等于温度 T3,则热管理控制状态机产生中断,并设置热中断状态寄存器中相应的状态位,以记录哪个或哪些传感器导致了中断(步骤 858)。操作系统继而将服务该中断,并可以减慢处理器上的负荷或将处理器的一些负荷卸载给系统中的其他处理器。

[0141] 在热管理控制状态机产生中断之后,热管理控制状态机将本地中断温度 T 设置为温度 T4,并且将本地中断方向 B<sub>x</sub> 设置为“1”(步骤 860)。应当将温度 T4 设置为小于或等于温度 T3。热管理控制状态机再次感知 DTS 的温度(步骤 862)。热管理控制状态机确定来自 DTS 的所感知温度是否低于温度 T4(步骤 864)。如果所感知的温度没有低于温度 T4,则操作返回步骤 862。如果所感知的温度低于温度 T4,则热管理控制状态机产生中断,并设置热中断状态寄存器中相应的状态位,以记录哪个或哪些传感器导致了中断(步骤 866)。此时,操作系统现在恢复正常操作是安全的。操作系统继而将服务该中断,并将系统恢复到正常操作。接下来,操作返回步骤 852,在此,热管理控制状态机将本地中断温度 T 设置为温度 T3,并将本地中断方向 B<sub>x</sub> 设置为“0”。

[0142] 该操作的示例是:如果给定的 DTS 具有 80°C 的本地中断温度以及本地中断方向“0”。一旦相关联单元的 DTS 感知到温度大于或等于 80°C,热管理控制状态机将产生中断,并设置热中断状态寄存器中相应的状态位,以记录哪个或哪些传感器导致了中断。操作系统继而将服务该中断,并且可以减慢处理器上的负荷或是将处理器的一些负荷卸载给系统中的其他处理器。而且,此时热管理控制状态机可以将本地中断温度重置为示例性的 77°C,并将本地中断方向设置为“1”。负荷将继续以慢模式操作或是保持脱离经历热条件的处理器单元或者处理器,直到 DTS 感知到温度低于 77°C。一旦热管理控制状态机确定所感知的温度低于 77°C,热管理控制状态机产生其他中断。热管理控制状态机将本地中断温度设置为 80°C,将本地中断方向设置为“0”,并且操作系统继而恢复负荷的正常操作。

[0143] 因此,高级热中断产生允许操作系统对中断产生进行编程,以跟随温度改变的方向并消除在热中断情况下中断处理机持续轮询当前温度的需要。

[0144] 图 9 示出了根据其他说明性实施方式的在热管理系统中支持深度节电模式和部分良好的操作的流程图。如前所述,Ce11 BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。在图 3 的 Ce11 BE 芯片 300 中,存在多种节电模式。依赖于每种节电模式的实现,一些可能限制 DTS(例如图 4 的 DTS 404、406、408 和 410)的可访问性。例如,如果 SPU(例如图 3 的 SPU 310、311 和 312)处于节电模式,其中时钟被关闭,即解串行器(例如图 4 的解串行器 462)被禁用,串行器(例如图 4 的串行器 456)和 DTS(例如图 4 的 DTS 404)之间的通路将不起作用。节电模式的另一示例可以是关闭电源。在这种情况下,可能

禁用实际的 DTS。另一示例是热管理控制状态机在制造检测过程中将处理器中的传感器或单元确定为损坏。如果该传感器或单元是冗余的,则制造可以将该传感器或单元标记为故障,从而创建部分良好的处理器,该处理器仍将通过有限数目的单元或处理器起作用。在任一情况下,热管理控制状态机(例如图 4 的热管理控制状态机 458)需要监控这些节电模式的状态,并屏蔽不工作的 DTS 参与热管理任务(例如,调节、中断等)。

[0145] 返回图 9,其描述了在热传感和热管理系统中支持深度节电模式和部分良好的操作的流程图。操作开始时,热管理控制状态机使用来自各个 DTS 的数据跟踪 DTS 的状态(步骤 902)。热管理控制状态机将数据存储在内在校准存储设备中,例如图 4 的内在校准存储设备 428 中。如前所述,节电模式、故障 DTS、或者通过数据流(例如图 4 的数据流 460)与热管理控制状态机通信的 SPU 可能禁止特定 DTS 的操作。由制造过程报告的部分良好情况的效果与节电模式类似,只是部分良好是永久性情况,并且 DTS 将被永久屏蔽。在 SPU 被标记为故障的情况下,热管理控制状态机关闭整个 SPU,并禁用串行器。在 DTS 被标记为故障的情况下,热管理控制状态机屏蔽该 DTS。热管理控制状态机确定 DTS 或 SPU 是有故障的还是工作的(步骤 904)。如果 DTS 或 SPU 是有故障的,则热管理控制状态机屏蔽该 DTS(步骤 906),此后操作结束。

[0146] 为了屏蔽处于功率管理状态的 DTS,热管理控制状态机将当前温度状态寄存器(例如图 4 的当前温度状态寄存器 412)的相关当前温度状态寄存器重置为最低温度设置 0x0。可选方法还可以是:通过设置状态位来分配相关当前温度状态寄存器的编码,以指明该 DTS 被屏蔽,这可能比仅仅重置传感器读取更为精确。热管理控制状态机继而停止从当前温度状态寄存器到和来自 DTS 的通信。停止通信是可选步骤,主要是为了节电以及不执行无用的开销工作。热管理控制状态机继而产生信号,以指明 DTS 当前被屏蔽并且不应当参与热管理任务。最后,热管理控制状态机重置 DTS 的状态。当与该 DTS 相关的单元,例如处理器内核或处理器本身,退出节电模式时,热管理控制状态机恢复与 DTS 的通信、恢复当前温度状态寄存器的更新、并发送 DTS 可以参与热管理任务的信号。

[0147] 返回步骤 904,如果 DTS 和 SPU 都是工作的,则热管理控制状态机开始与 DTS 的通信(步骤 908)。热管理控制状态机监控 SPU 的功率管理状态,以确定 SPU 何时进入节电模式(步骤 910)。直到 SPU 进入节电模式,操作返回步骤 908。如果 SPU 进入节电模式并且 DTS 被禁用,则热管理控制状态机通过上文关于步骤 906 讨论的方法屏蔽该 DTS(步骤 912)。由于将 DTS 指明为禁用并工作的,因此热管理控制状态机继续监控 SPU 的功率管理状态(步骤 914)。直到 SPU 退出节电模式,操作返回步骤 912。当 SPU 退出节电模式、并且 DTS 不再被禁用时,热控制状态机开始与 DTS 的通信、恢复当前温度状态寄存器的更新、并发送 DTS 可以参与热管理任务的信号(步骤 916),操作返回步骤 908。

[0148] 因此,屏蔽部分良好、故障、或者是处于节电模式的 DTS 的温度读取将不工作的或被禁用的 DTS 与参与热管理任务相隔离开。

[0149] 图 10 示出了根据其他说明性实施方式的支持不依赖于温度的热敏软件应用的实时测试的热调节控制特征的操作的流程图。如上所述,Ce11 BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。热管理控制寄存器(例如图 4 的热管理控制寄存器 430),提供对各种热调节控制特征的访问和配置。将热调节设计为通过在热事件的情况下使用调节以削减性能来降低温度。

[0150] 热管理停止时间寄存器（例如图 4 的热管理停止时间寄存器 434）和热管理调节比例寄存器（例如图 4 的热管理调节比例寄存器 436）一起设置调节量和调节行为。在实时系统中，需要保证实时期限。对于软件开发者和质量保证团队来说，了解和测试最大调节量是很重要的，其中，最大调节量是程序或代码段可以容忍并仍然保证实时系统的实时期限的热管理停止时间寄存器和热管理调节比例寄存器的最大设置。热管理控制状态机提供不论温度总是提供调节的模式，而不是调整导致热事件并由此触发调节条件的硬件的实际温度。热管理控制状态机在热管理控制寄存器中设置该模式，其将芯片设置为恒定调节状态。该特征帮助软件开发者测试其代码并使其代码满足实时标准。

[0151] 操作开始时，接收热管理停止时间寄存器和热管理调节比例寄存器热控制设置（步骤 1002）。热管理控制状态机使用热管理停止时间寄存器和热管理调节比例寄存器的设置来确定将如何执行调节。接着，热管理控制状态机设置测试模式，并将热管理控制寄存器设置为始终调节设置（步骤 1004）。接着，程序运行，以用于实时确认：软件或程序在热管理停止时间寄存器和热管理调节比例寄存器热控制设置下将满足实时期限（步骤 1006）。检测模式可以是任意类型的调节模式，例如始终调节或随机调节。接着，热管理控制状态机确定是否满足实时期限（步骤 1008）。如果没有满足实时期限，则热管理控制状态机将当前的热管理停止时间寄存器和热管理调节比例寄存器热控制设置记录为失败（步骤 1010）。热管理控制状态机继而确定是否存在将减小调节量的任何新的热管理停止时间寄存器和热管理调节比例寄存器热控制设置（步骤 1012）。如果存在新的热管理停止时间寄存器和热管理调节比例寄存器热控制设置，则操作返回步骤 1002。如果在步骤 1012 不存在任何新的热管理停止时间寄存器和热管理调节比例寄存器热控制设置，则操作结束。

[0152] 返回步骤 1008，如果满足实时期限，则热管理控制状态机将当前热管理停止时间寄存器和热管理调节比例寄存器热控制设置记录为通过（步骤 1014）。热管理控制状态机确定是否存在将提高调节量的任何新的热管理停止时间寄存器和热管理调节比例寄存器热控制设置（步骤 1016）。如果存在新的热管理停止时间寄存器和热管理调节比例寄存器热控制设置，则操作返回步骤 1002。如果在步骤 1016，没有任何新的热管理停止时间寄存器和热管理调节比例寄存器热控制设置，则操作结束。

[0153] 因此，提供始终调节的操作模式帮助软件开发者检测其代码并使其代码在最差情况的热条件下满足实时期限。软件开发者和质量保证团队还可以使用该特征来确定程序或代码段可以容忍并仍保证满足实时系统的实时期限的最大调节量。一旦热管理控制状态机确定并确认最大调节量，软件可以设置在完全调节发生的条件下发生的中断。如果热管理控制状态机曾经产生该中断，则热管理控制状态机将存在违反或不满足实时保证的可能性通知应用。

[0154] 除了始终调节控制设置之外，实现还可以提供这样的模式：发出随机热事件或受控随机热事件，以便仿真更真实的调节交互和软件执行。该技术类似于在总线上随机发出错误以测试错误恢复码。

[0155] 图 11 示出了根据其他说明性实施方式的用于对中断延迟影响最小的热调节控制实现的操作的流程图。如上所述，Cell BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。当计算机系统的任意部分处于调节条件中时，调节条件降低整个系统的性能。按照多久之后可以服务于中断以及将花费多长时间来服务该中断，性能的降低增加了中断

的延迟。中断延迟的增加严重牵连整个系统,并因此希望并需要使热调节对中断延迟的影响最小化。最小化由于中断延迟而对热调节的影响是一个针对 PPU(例如图 3 的 PPU 308)调节控制的特征。SPU,例如图 3 的 SPU 310、311 和 312 不进行中断,并因此不受该特征的影响。

[0156] 操作开始时,热管理控制状态机(例如图 4 的热管理控制状态机 458)监控所有 PPU 中断状态位以及热管理系统中断屏蔽寄存器,例如图 4 的热管理系统中断屏蔽寄存器 438(步骤 1102)。热管理系统中断屏蔽寄存器控制中断的屏蔽。热管理控制状态机确定是否存在任何未屏蔽的未决中断(步骤 1104)。如果没有未决中断或是存在未决中断但是受到屏蔽,则操作返回步骤 1102。

[0157] 如果在步骤 1104 处存在未屏蔽的未决中断,则热管理控制状态机临时禁用任何调节模式,不论是部分调节还是完全调节状态(步骤 1106)。禁用调节模式允许 PPU 临时以完全性能进行操作并处理任何未决中断,而没有由热调节效果导致的任何延时。热管理控制状态机再次监控所有 PPU 中断状态和热管理系统中断屏蔽寄存器(步骤 1108)。热管理控制状态机确定是否存在任何未屏蔽的未决中断(步骤 1110)。如果没有未决中断或是存在未决中断但是受到屏蔽,操作返回步骤 1108。当在步骤 1110 中断状态清除时,热管理控制状态机将 PPU 恢复到原始调节模式(步骤 1112),并且操作返回步骤 1102。

[0158] 中断处理机可以选择在中断处理机子程序的开始还是子程序的结束处来清除中断状态位。中断处理机可以位于 Power 处理器元件(例如图 3 的 Power 处理器元件 301)中,或由 Power 处理器元件执行的软件中。如果中断处理机选择在开始处清除中断状态位并且还希望避免任何 PPU 的性能下降,则中断处理机可以在清除中断状态位之前禁用热调节。换言之,中断不会导致控制寄存器的改变。因此,在存在未屏蔽中断时,调节仍然是启用的,但是由热管理控制单元(例如图 4 的 TMCU 402)挂起。如果中断处理机应当在处理中断之前重置中断状态,则处理机应当将控制寄存器设置为禁用调节(或者将调节量减小到可接受的水平)、重置中断、服务中断、并继而重新启用调节或者将调节量设置回先前的水平。可以通过将热管理控制寄存器(例如图 4 的热管理控制寄存器 430)设置为 0XX 来执行示例性的热调节禁用,其中 X 无关紧要。在中断子程序的结束,中断处理机应当将热管理控制寄存器设置回其原始值。如果中断处理机在中断子程序结束处清除中断状态位,则无需其他工作,并且只要中断状态位是活动的,热管理控制状态机就将保持 PPU 不进入调节模式。

[0159] 图 12 示出了根据其他说明性实施方式的用于热调节滞后的操作的流程图。如前所述,Ce11 BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。热调节滞后是进行改变(例如调节和结束调节)与该改变的响应或效果之间的延迟。例如,如果将调节点设置为 75°C 并将结束调节点设置为 72°C,则滞后的范围是从 75°C 到 72°C。图 5 描述了热调节滞后。

[0160] 热管理调节点寄存器(例如图 4 的热管理调节点寄存器 432)提供两个温度设置:调节温度和结束调节温度。应当将调节温度设置为高于结束调节温度。温度差限定了调节温度和结束调节温度之间的滞后量,由此提供了可编程的滞后量。

[0161] 说明性地将以下讨论限制为一个 DTS,在滞后热调节操作开始时,热管理控制状态机设置热管理调节点寄存器中的调节温度和结束调节温度(步骤 1202)。热管理控制状态机感知 DTS 的温度(步骤 1204)。热管理控制状态机确定来自 DTS 的感知温度是否大于

或等于调节温度（步骤 1206）。如果所感知温度不大于或等于调节温度，则操作返回步骤 1204。如果在步骤 1206 处所感知的温度大于或等于调节温度，则热管理控制状态机启动调节模式（步骤 1208）。

[0162] 热管理控制状态机再次感知 DTS 的温度（步骤 1210）。热管理控制状态机确定来自 DTS 的所感知温度是否大于或等于调节温度（步骤 1212）。如果所感知温度不小于结束调节温度，则操作返回步骤 1210。如果在步骤 1212 中 DTS 小于结束调节温度，则热管理控制状态机禁用调节模式（步骤 1214），操作返回步骤 1204。

[0163] 因此，假设热管理控制寄存器被适当配置为允许调节模式，则当温度上升到等于或高于调节温度时，热管理控制状态机将单元置于调节模式中。热管理控制状态机保持该单元为调节模式，直到温度下降到结束调节温度之下。如果结束调节温度小于调节温度，则所识别的滞后在禁用调节模式之前允许单元充分冷却。如果没有滞后，则单元可能非常频繁地进入和离开调节模式，并降低调节的总体效率和处理器的效率。

[0164] 处理器调节的示例性方法可以通过阻塞指令分派来实现。如果非常频繁地启用和禁用调节，则可能很经常地排空 (flush) 处理器的流水线，因此降低了处理器的能力。处理器调节的另一示例性方法可以通过减慢时钟频率来实现。

[0165] 图 13 示出了根据其他说明性实施方式的热调节逻辑实现的操作的流程图。图 13 表示以上附图中描述的完整的热管理解决方案。如前所述，Ce11 BE 芯片包括通过图 3 的普适逻辑单元 351 提供的热管理系统。TCMU，例如图 4 的 TCMU 402，包括多个动态热管理寄存器。动态热管理寄存器是：热管理控制寄存器、热管理调节点寄存器、热管理停止时间寄存器、热管理调节比例寄存器以及热管理系统中断屏蔽寄存器，例如图 4 的热管理控制寄存器 430 (TM\_CR1 和 TM\_CR2)、热管理调节点寄存器 432 (TM\_TPR)、热管理停止时间寄存器 434 (TM\_STR1 和 TM\_STR2)、热管理调节比例寄存器 436 (TM\_TSR) 以及热管理系统中断屏蔽寄存器 438 (TM\_SIMR)。

[0166] 热管理调节点寄存器设置 DTS 的调节点。可以在热管理调节点寄存器中设置两个独立的调节点，一个用于 PPE，一个用于 SPE。该寄存器中还包含用于启用调节和禁用调节或是停止 PPE 或 SPE 的温度点。当温度等于或高于调节点时，PPE 或 SPE 的调节执行开始。当温度下降到禁用调节的温度之下时，调节停止。如果温度达到了完全调节温度或停止温度，则停止 PPE 或 SPE 的执行。

[0167] 热管理控制状态机使用热管理停止时间寄存器和热管理调节比例寄存器来控制调节的频率和量。当温度达到调节点时，热管理控制状态机在热管理调节比例寄存器中的相应比例值所指定的时钟数目内停止相应的 PPE 或 SPE。继而，热管理控制状态机在热管理停止时间寄存器中的运行值乘以相应比例值所指定的时钟数目内允许 PPE 或 SPE 运行。该序列继续，直到温度下降到禁用调节以下。

[0168] 热管理控制状态机使用热管理系统中断屏蔽寄存器来选择在中断未决时哪些中断禁用 PPE 的调节。

[0169] 热管理控制寄存器针对每个 PPE 或 SPE 独立地设置调节模式。下面是针对每个 PPE 或 SPE 独立设置的 5 种不同模式：

[0170] 禁用动态调节（包括内核停止保护）

[0171] 正常操作（启用动态调节和内核停止保护）



[0172] PPE 或 SPE 始终调节（启用内核停止保护）

[0173] 禁用内核停止保护（启用动态调节，并且禁用内核停止保护）

[0174] PPE 或 SPE 始终调节，并且禁用内核停止保护

[0175] 作为实现热调节逻辑的操作，热管理控制状态机设置热管理调节点寄存器中的调节温度和结束调节温度（步骤 1302）。热管理控制状态机感知 DTS 的温度（步骤 1304）。热管理控制状态机确定来自 DTS 的所感知温度是否大于或等于调节温度（步骤 1306）。如果所感知温度没有大于或等于调节温度，则操作返回步骤 1304。如果所感知的温度大于或等于调节温度，则热管理控制状态机启动调节模式（步骤 1308）。

[0176] 接着，热管理控制状态机通过热管理控制寄存器中指明的值所指出的调节类型来控制调节（步骤 1310）。一旦指出了调节类型，热管理控制状态机就通过热管理停止时间寄存器中指明的调节量来对调节进行限制（步骤 1312）。停止时间寄存器设置处理器将停止多长时间与将允许处理器运行多长时间的比率，或者说调节百分比。最后，热管理控制状态机通过热管理比例寄存器中所指定的值来缩放（scale）停止和运行时间的持续时间（步骤 1314）。此时，操作分开为并发的操作：步骤 1316 和 1322。在步骤 1316，热管理控制状态机感知 DTS 的温度。热管理控制状态机确定来自 DTS 的所感知温度是否大于或等于调节温度（步骤 1318）。如果所感知温度不小于结束调节温度，则操作返回步骤 1316。如果 DTS 小于结束调节温度，则热管理控制状态机禁用调节模式（步骤 1320），操作返回步骤 1304。

[0177] 返回步骤 1314，在实现最终调节限制之后，热管理控制状态机并发地针对任何未决中断而监控所有 PPU 中断状态（步骤 1322）。如果在实现调节时遇到中断，则热管理控制状态机临时禁用任何调节模式，直到该中断已得到处理，此时，启用调节（不论是部分调节或完全调节状态）并且操作返回步骤 1308。关于图 11 给出了中断状态监控的深入讨论。

[0178] 因此，Cell BE 芯片中所包括的热管理系统的热中断逻辑提供了管理 Cell BE 芯片的热条件以及保护 Cell BE 芯片及其组件的动态方式。

[0179] 说明性实施方式的形式可以采用全硬件实施方式，全软件实施方式，或者包含硬件和软件元件两者的实施方式。说明性实施方式在软件中实现，软件包括但不限于固件、驻留软件、微代码等等。

[0180] 此外，说明性实施方式的形式可以采用计算机程序产品，其可以从计算机可用或计算机可读介质访问，提供由计算机或任意指令执行系统使用或与之结合的程序代码。为了该描述的目的，计算机可用或计算机可读介质可以是包含、存储、通信、传播或传输指令执行系统、装置或设备使用或与之结合的程序的任何有形装置。

[0181] 介质可以是电、磁、光、电磁、红外或半导体系统（或装置或设备）或传播介质。计算机可读介质的例子包括半导体或固态存储器、磁带、可移动计算机盘、随机访问存储器（RAM）、只读存储器（ROM）、刚性磁盘和光盘。光盘的当前示例包括致密盘 - 只读存储器（CD-ROM）、致密盘 - 读 / 写（CD-R/W）以及 DVD。

[0182] 适于存储和 / 或执行程序代码的数据处理系统将包括至少一个处理器，其通过系统总线直接或间接地耦合至存储器元件。存储器元件可以包括在程序代码执行期间使用的本地存储器，大容量存储设备，以及高速缓存存储器，其中高速缓存存储器提供至少一些程序代码的临时存储以降低在执行期间从大容量存储设备获取代码的次数。

[0183] 输入 / 输出或者 I/O 设备（包括但不限于键盘、显示器、指示设备等）可以直接或

者通过中介 I/O 控制器耦合至系统。

[0184] 网络适配器也可以耦合至系统,以使数据处理系统能够通过中介的私有或公共网络变为耦合至其他数据处理系统或远程打印机或存储设备。调制解调器、缆线调制解调器和以太网卡仅是网络适配器的当前可用类型中的几个。

[0185] 出于说明和描述的目的已经给出了对说明性实施方式的描述,但这并非意在详尽或是限制为所公开形式的说明性实施方式。对于本领域的普通技术人员来说,多种修改和变形将是显然的。选择和描述实施方式是为了最好地阐释说明性实施方式的原理、实践应用,以及使本领域的其他普通技术人员能够理解到,说明性实施方式具有各种变形的多种实施方式也适合于预期的特定用途。

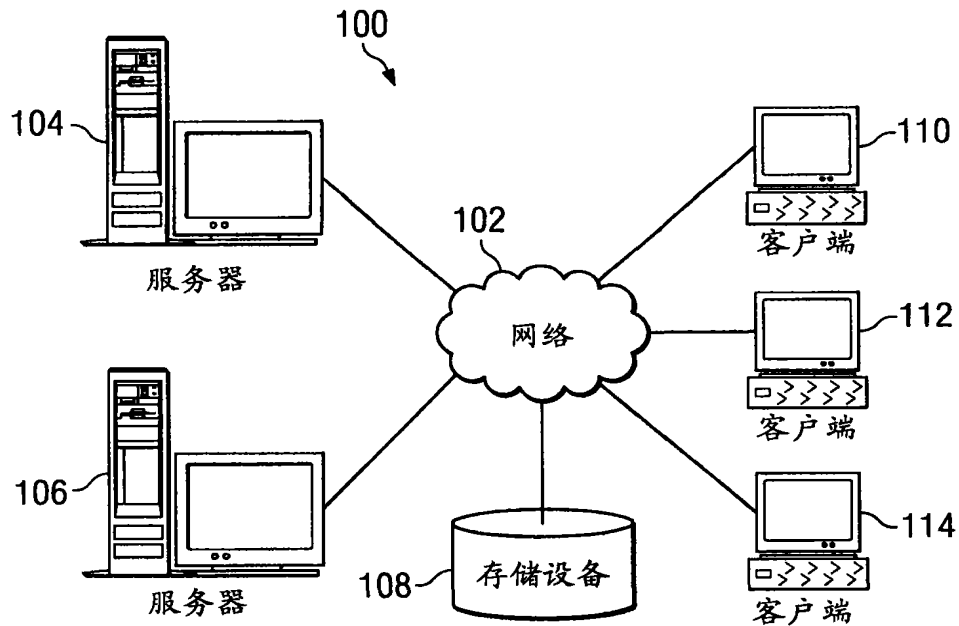


图 1

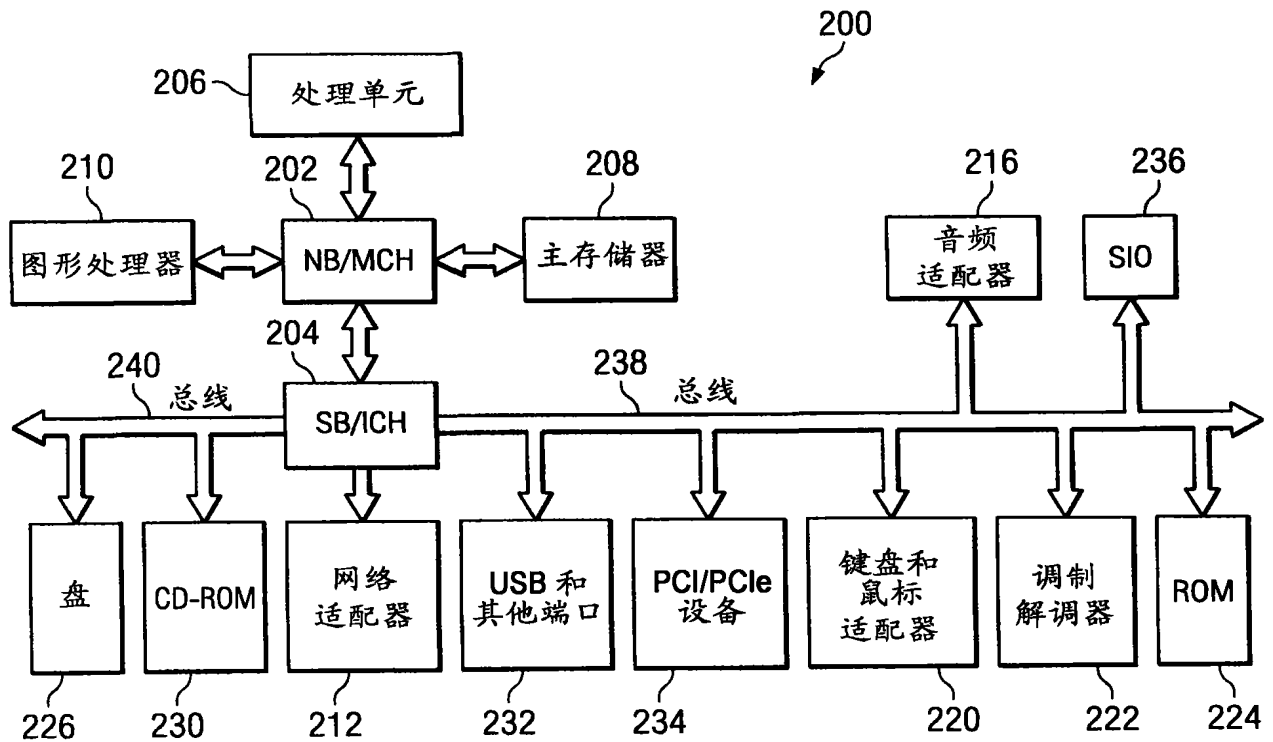


图 2

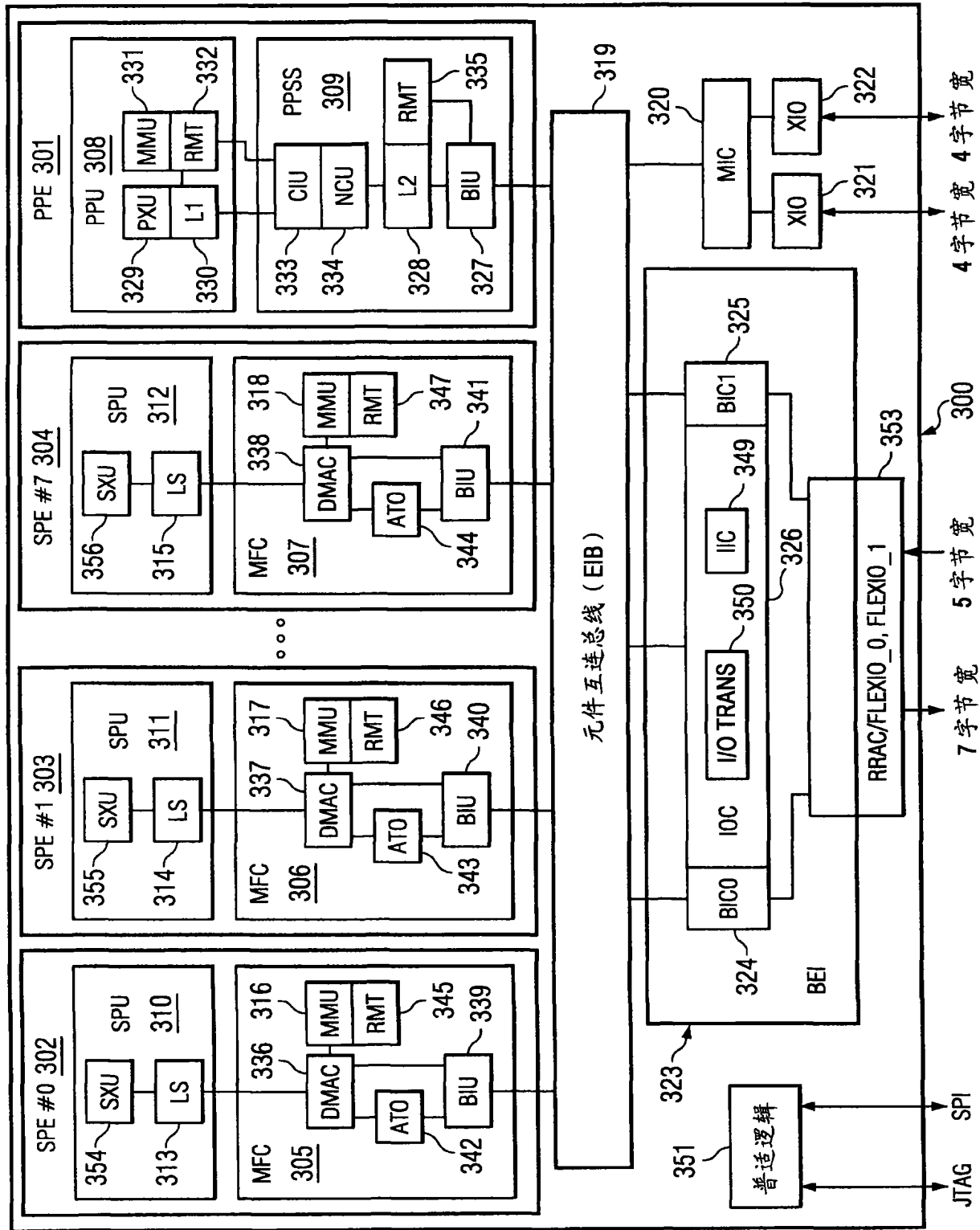


图 3

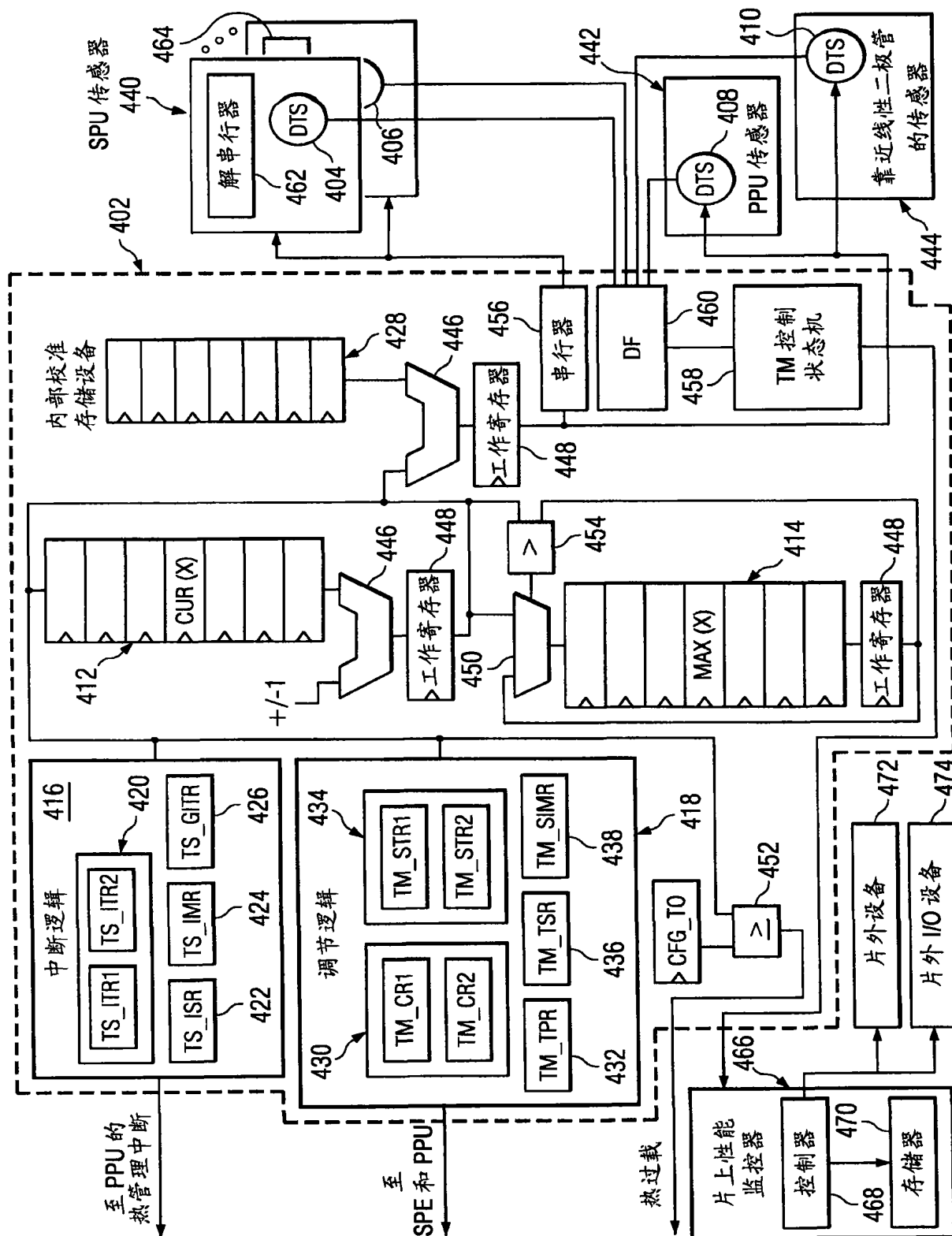
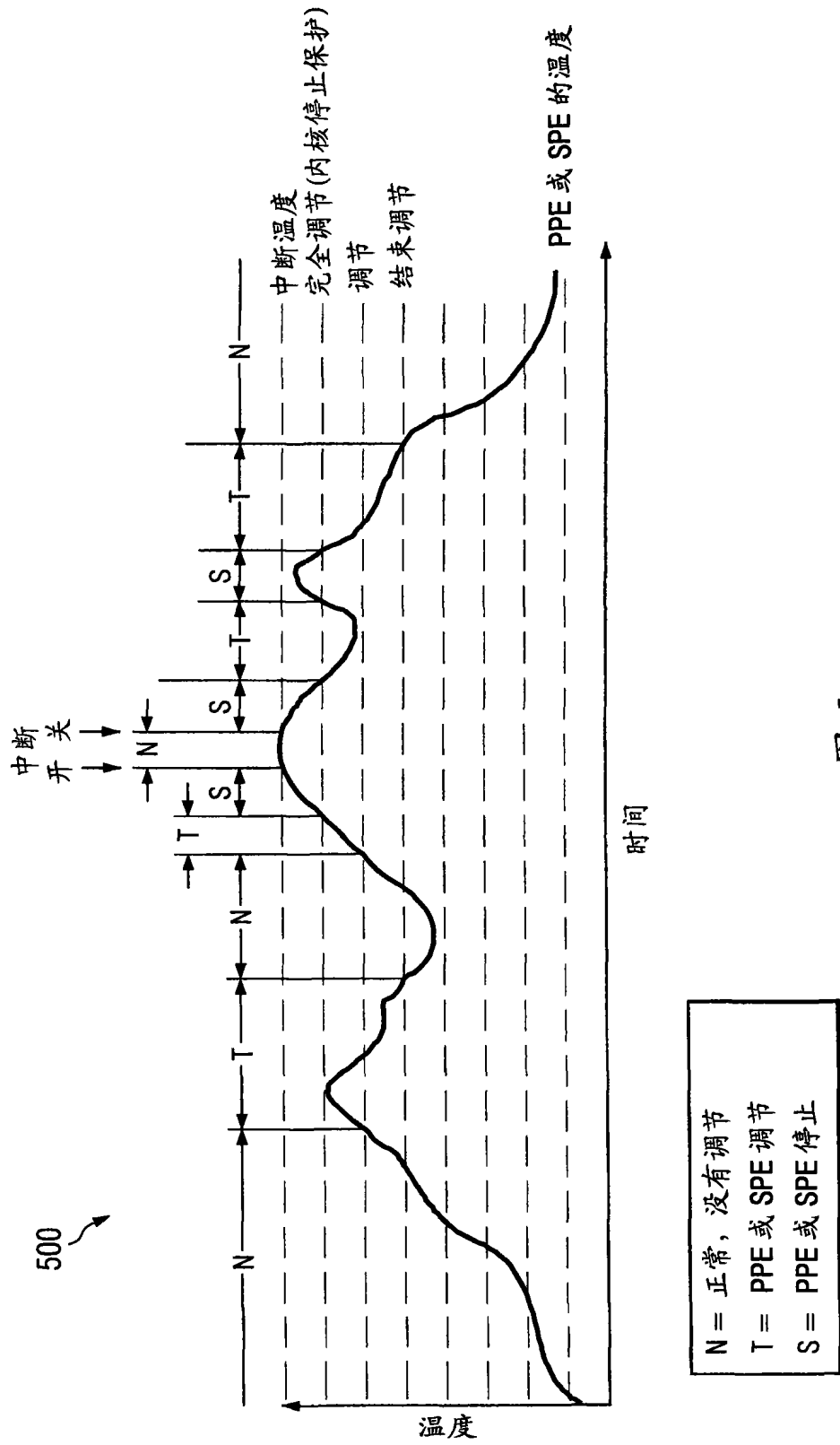


图 4



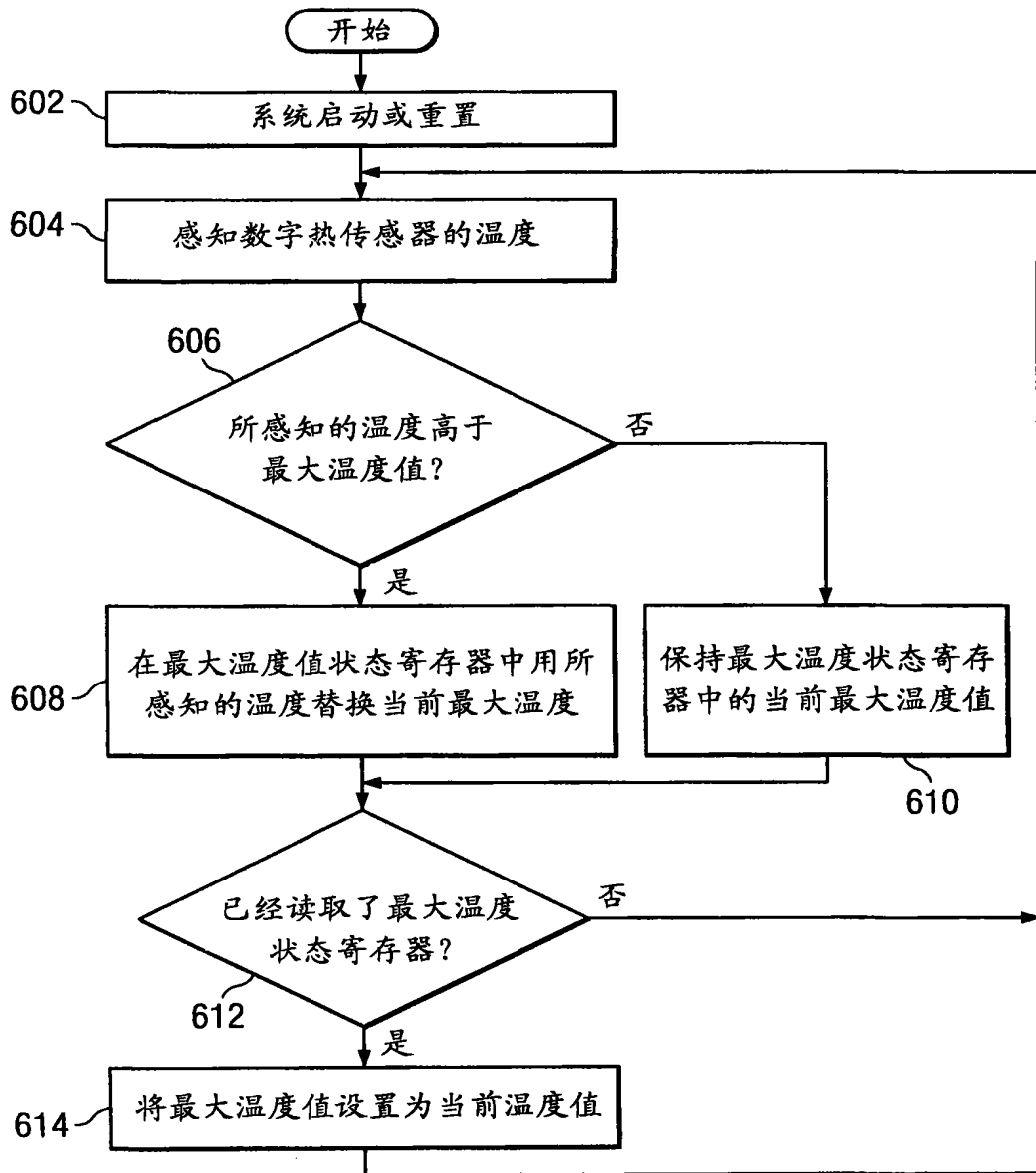


图 6

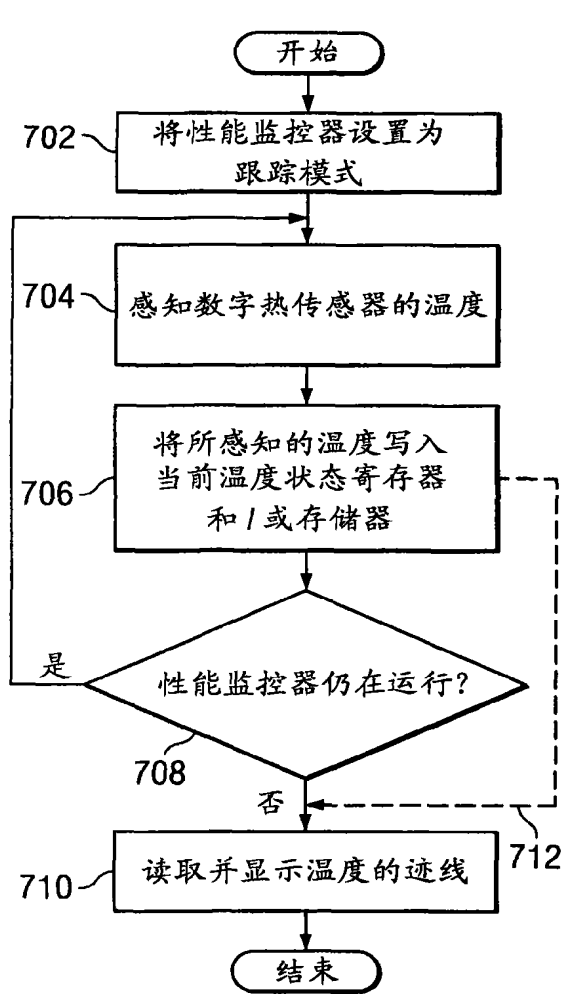


图 7

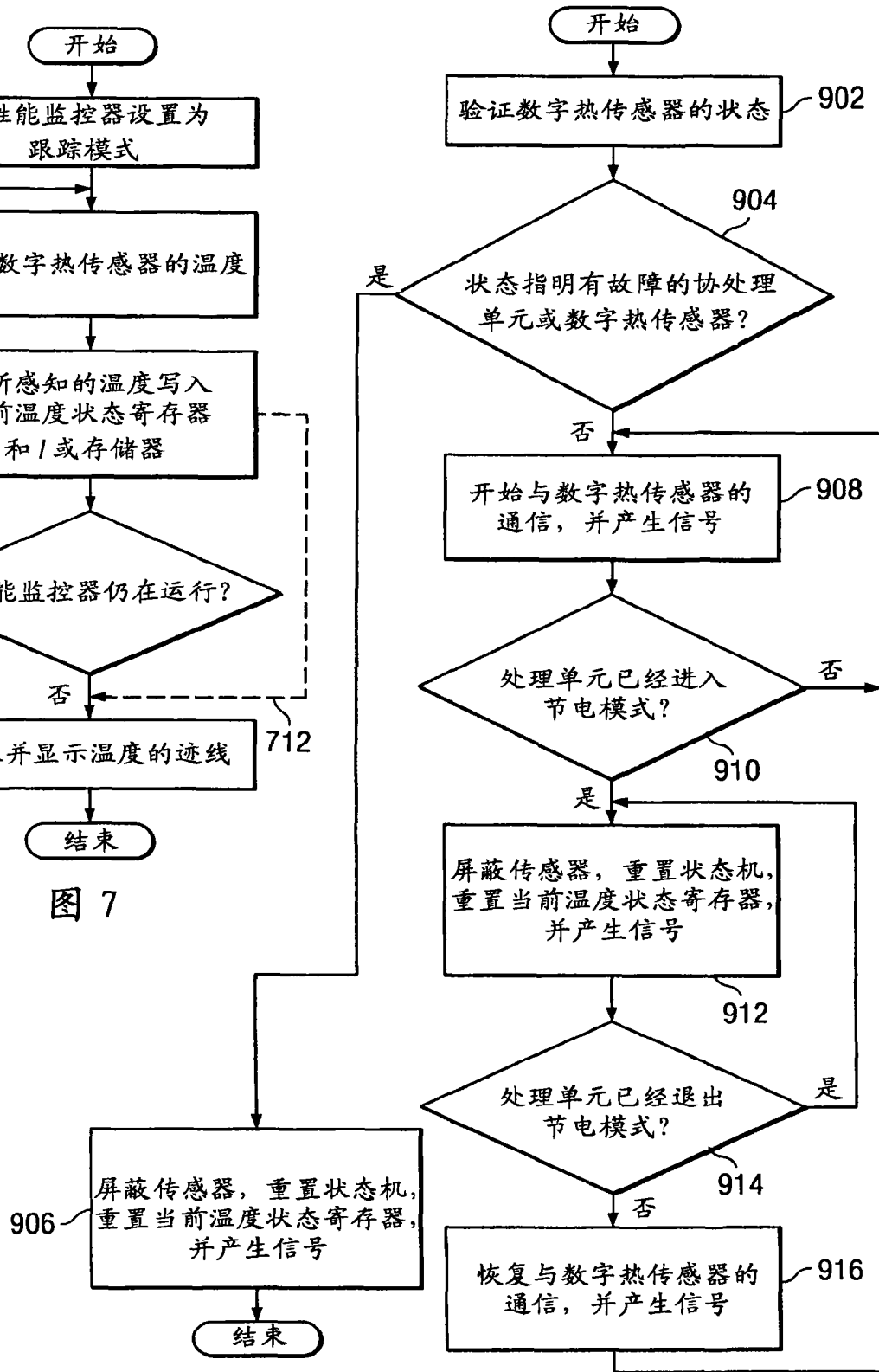


图 9



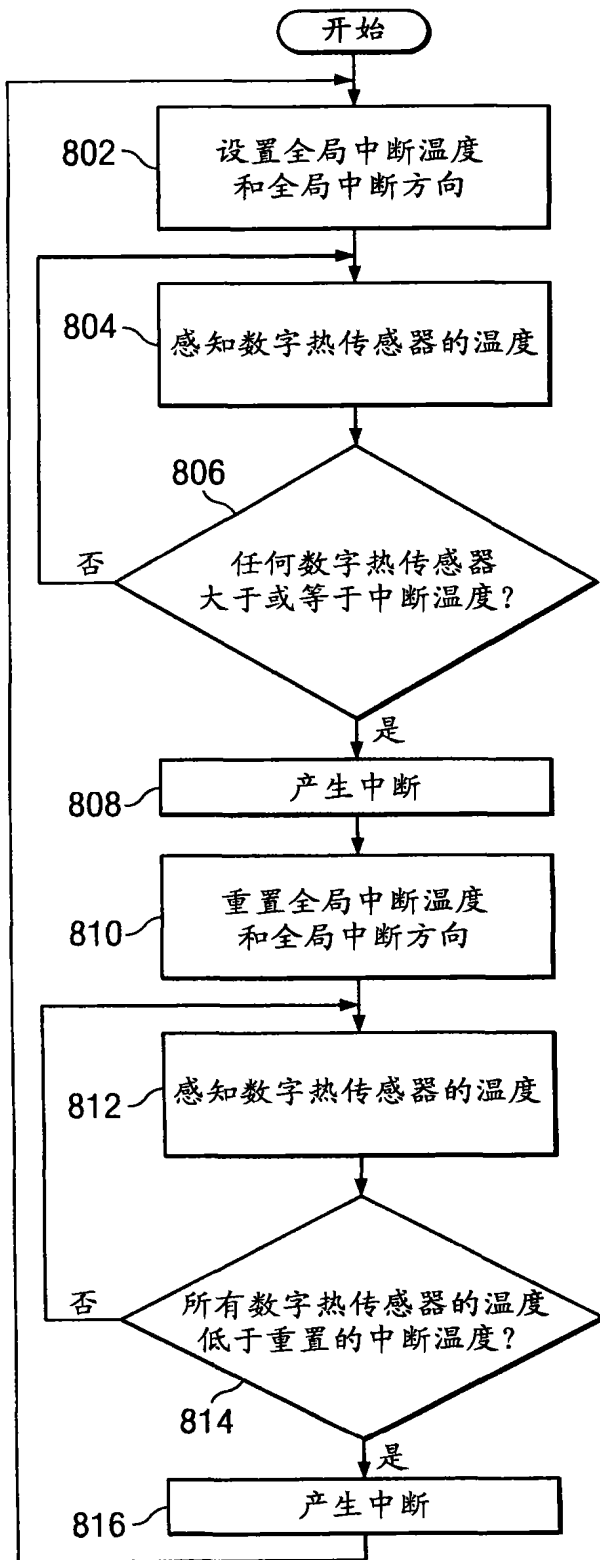


图 8A

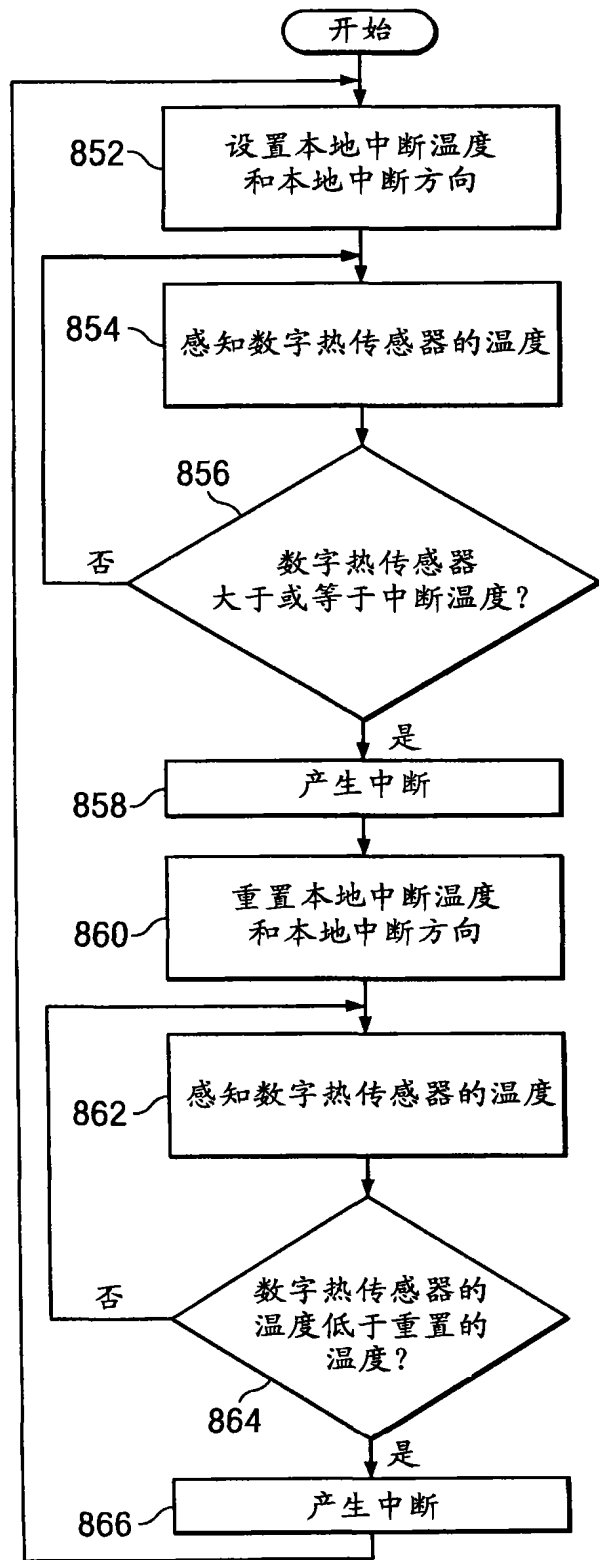


图 8B

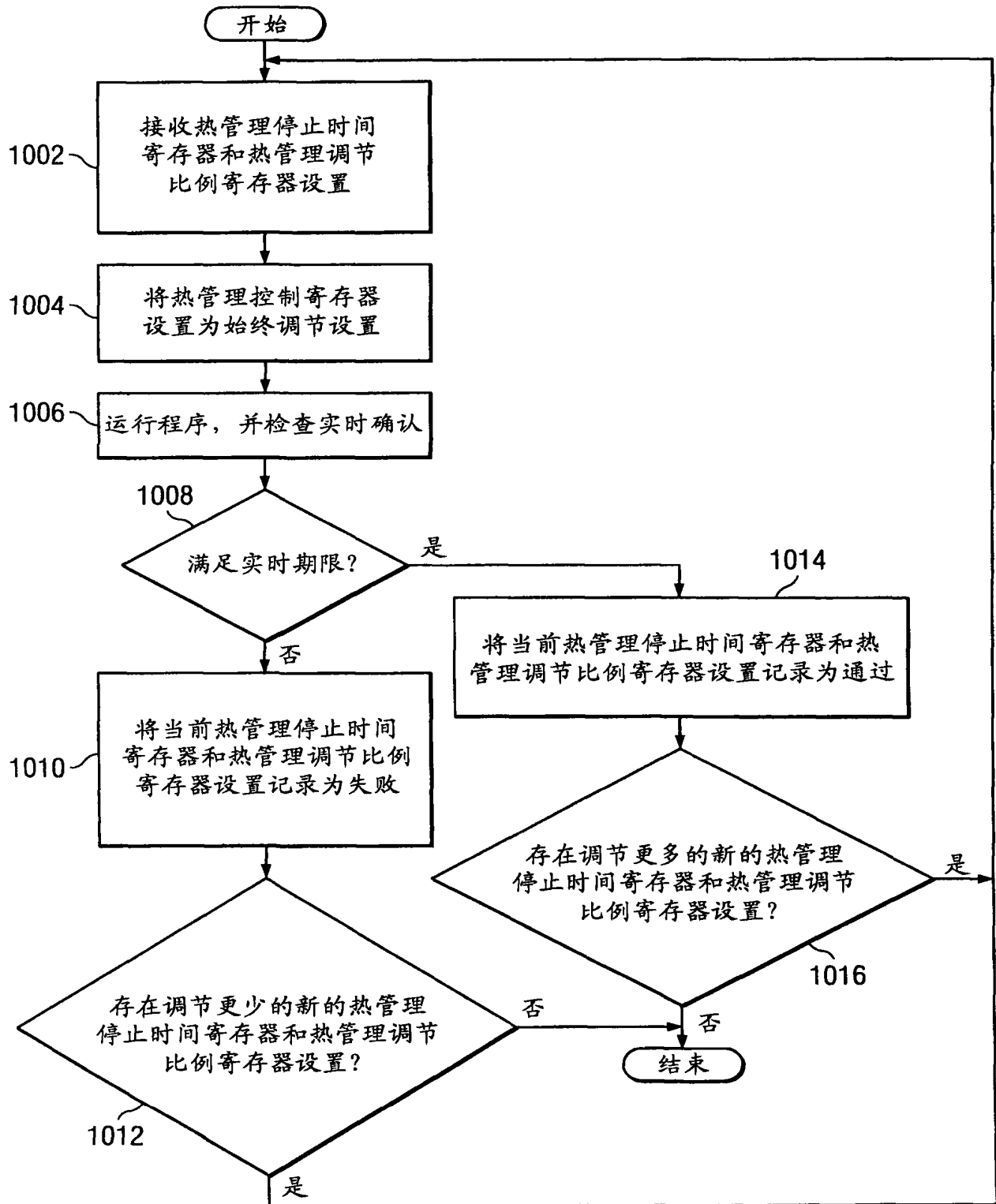


图 10

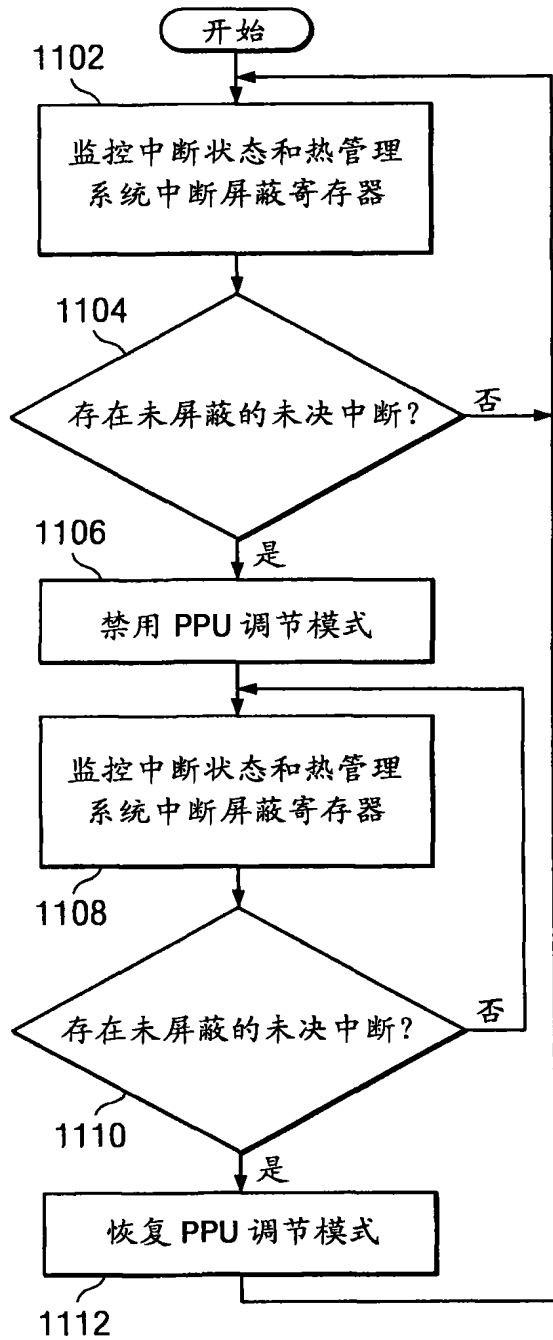


图 11

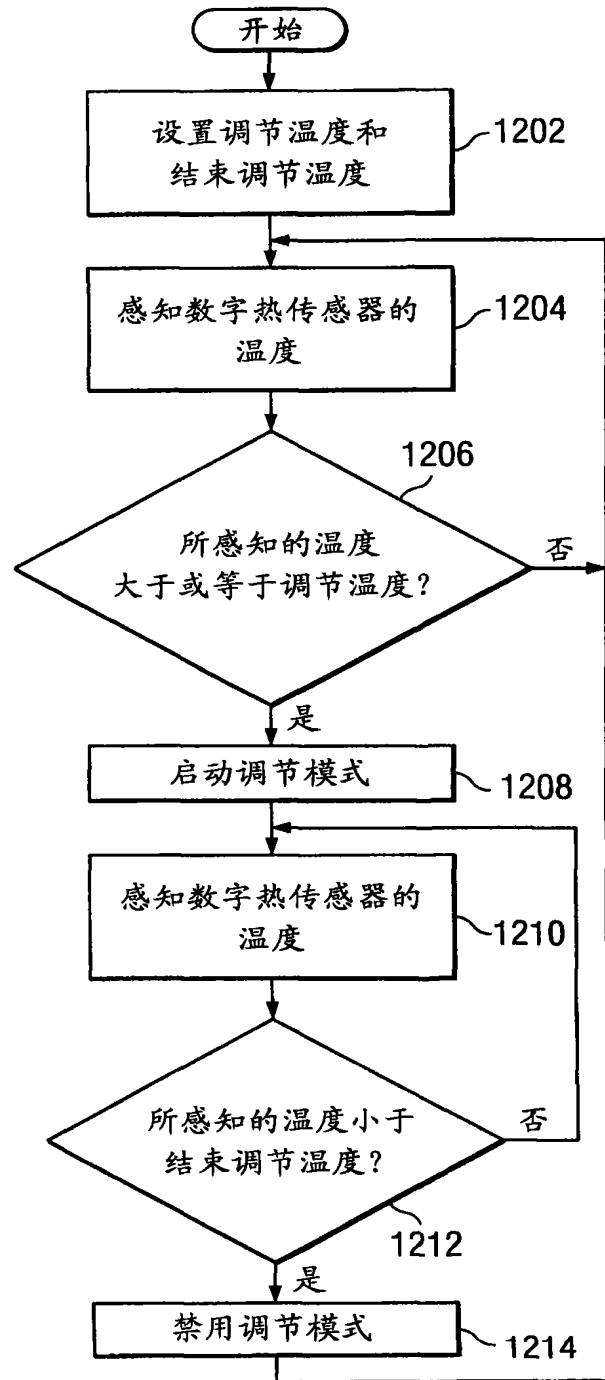


图 12

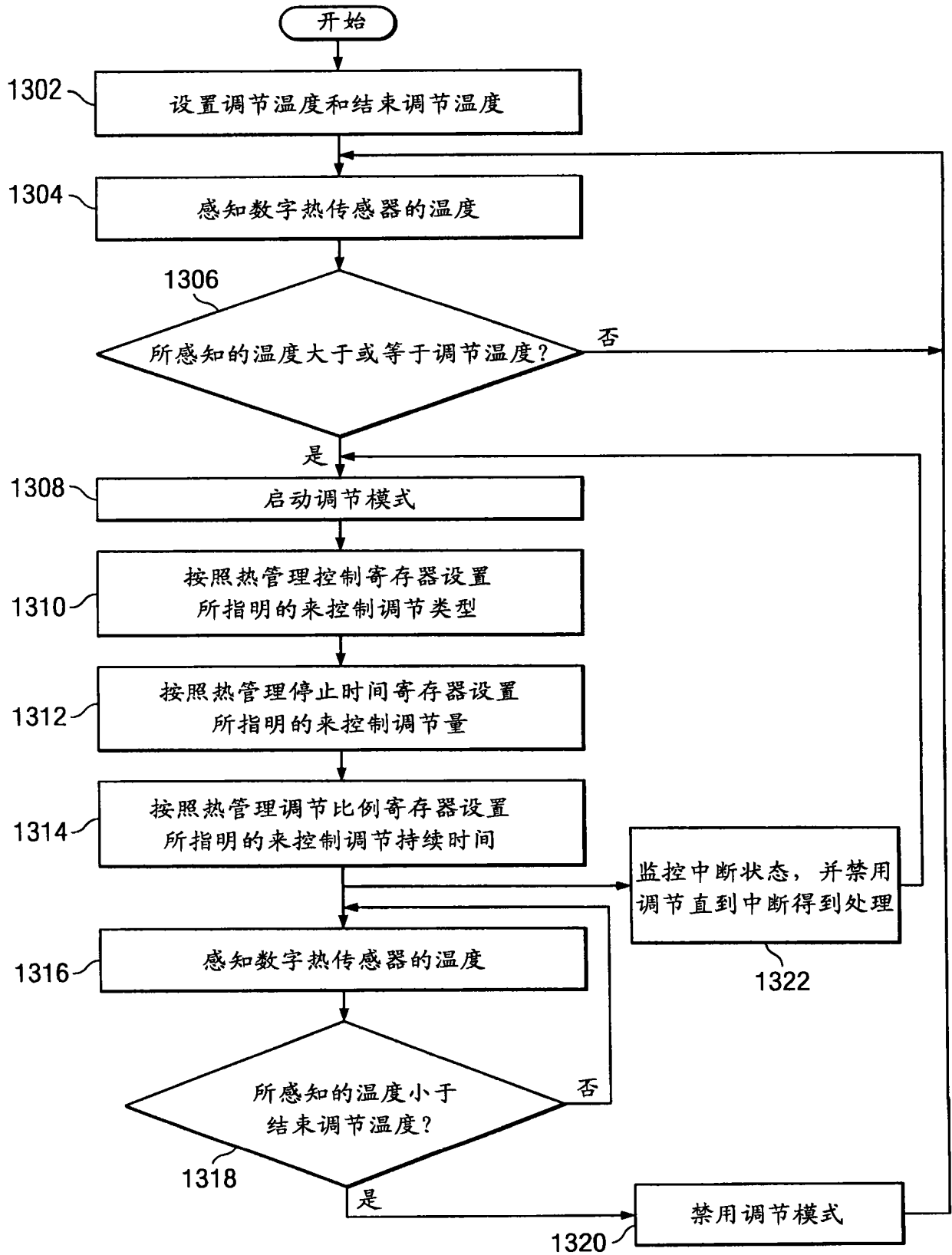


图 13