



(12) 发明专利

(10) 授权公告号 CN 102971687 B

(45) 授权公告日 2016. 08. 03

(21) 申请号 201180033354. 0

(22) 申请日 2011. 06. 21

(30) 优先权数据

12/819, 777 2010. 06. 21 US

(85) PCT国际申请进入国家阶段日

2013. 01. 05

(86) PCT国际申请的申请数据

PCT/US2011/041291 2011. 06. 21

(87) PCT国际申请的公布数据

W02011/163261 EN 2011. 12. 29

(73) 专利权人 超威半导体公司

地址 美国加利福尼亚州

(72) 发明人 塞缪尔·D·纳夫齐格

约翰·P·佩特里 威廉·A·休斯

(74) 专利代理机构 上海胜康律师事务所 31263

代理人 李献忠

(51) Int. Cl.

G06F 1/32(2006. 01)

(56) 对比文件

WO 2007/024396 A1, 2007. 01. 03,

US 7681054 B2, 2010. 03. 16,

US 2003/125900 A1, 2003. 07. 03,

US 2004/268166 A1, 2004. 12. 30,

Charles Lefurgy. Server-level Power Control. 《Autonomic Computing》. 2007,

审查员 陈清华

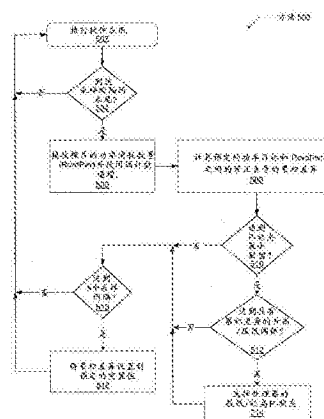
权利要求书2页 说明书8页 附图5页

(54) 发明名称

管理多个分立操作点以产生稳定虚拟操作点的方法及计算系统

(57) 摘要

一种用于管理多个分立操作点以产生稳定的虚拟操作点的系统和方法。处理器内的一个或多个功能块产生对应于与相应功能块相关的活动水平的数据。功率管理器在每个给定的采样间隔基于数据确定功率消耗值一次。此外,功率管理器确定随着时间的流逝在热设计功率(TDP)和功率消耗值之间的带正负号的累积差异。功率管理器基于带正负号的累积差异和给定阈值的比较来选择下一功率-性能状态(P-状态)。以这种方式在P-状态之间转变而工作负荷不明显改变使处理器在所支持的分立操作点之间的虚拟操作点处操作。



1. 一种计算系统,其包括:

一个或多个位于处理器核心内的功能块,每个功能块都被配置成产生对应于相应块的活动水平的数据,其中每个功能块配置为在一个或多个分立的功率-性能状态操作;以及功率管理器,其中所述功率管理器被配置成:

至少部分地基于所述数据为所述一个或多个功能块确定在给定时间间隔期间的平均功率消耗;

随着时间的流逝在平均功率消耗和期望功率消耗之间确定带正负号的累积差异;

响应于至少确定所述带正负号的累积差异大于第一阈值来选择比当前功率-性能状态高的功率-性能状态;

响应于至少确定所述带正负号的累积差异小于第二阈值来选择比当前功率-性能状态低的功率-性能状态,其中所述第二阈值小于所述第一阈值。

2. 如权利要求1所述的计算系统,其中所述功率管理器进一步被配置成:

确定对应于期望功率消耗的虚拟功率-性能状态,

其中所述虚拟功率-性能状态不是所支持的分立功率-性能状态以及

确定用于第一功率-性能状态和第二功率-性能状态的每一个的相应的最小驻留时间,所述第一功率-性能状态低于所述虚拟功率-性能状态,所述第二功率-性能状态高于所述虚拟功率-性能状态。

3. 如权利要求1所述的计算系统,其中所述功率管理器被配置成:响应于进一步确定相应的最小驻留时间消逝以便随着时间的流逝产生比对应于第一功率-性能状态或第二功率-性能状态的功率消耗更紧密地对应于所述期望功率消耗的平均功率消耗,交替地选择和传送所述第一功率-性能状态和第二功率-性能状态。

4. 如权利要求3所述的计算系统,其中所述期望功率消耗对应于所述一个或多个功能块的热设计功率值。

5. 如权利要求2所述的计算系统,其中所述功率管理器还被配置成基于达到对应阈值的速率选择远离所述当前功率-性能状态的功率-性能状态的两个或多个状态。

6. 如权利要求1所述的计算系统,其中所述相应块的活动水平是基于相应块中的采样信号而非电流或电压测量。

7. 如权利要求1所述的计算系统,其中所述功率管理器还被配置成:

在N次功能块采样间隔计数结束时,重置带正负号的累积差异;以及

响应于探测到所述计数达到N,重置所述计数。

8. 如权利要求7所述的计算系统,其中所述功率管理器还被配置成:

在N次功能块采样间隔计数结束时,重置带正负号的累积差异;以及

响应于探测到所述带正负号的累积差异已改变,重置所述计数。

9. 一种用于管理多个分立操作点以产生稳定的虚拟操作点的方法,所述方法包括:

产生对应于一个或多个功能块的活动水平的数据,其中每个功能块配置为在一个或多个分立的功率-性能状态操作;

至少部分地基于所述数据为所述一个或多个功能块确定在给定时间间隔期间的平均功率消耗;

随着时间的流逝在平均功率消耗和期望功率消耗之间确定带正负号的累积差异;

响应于至少确定所述带正负号的累积差异大于第一阈值来选择比当前功率-性能状态高的功率-性能状态;以及

响应于至少确定所述带正负号的累积差异小于第二阈值来选择比当前功率-性能状态低的功率-性能状态,其中所述第二阈值小于所述第一阈值。

10.如权利要求9所述的方法,进一步包括:

确定对应于期望功率消耗的虚拟功率-性能状态,

其中所述虚拟功率-性能状态不是所支持的分立功率-性能状态以及

确定用于第一功率-性能状态和第二功率-性能状态的每一个的相应的最小驻留时间,所述第一功率-性能状态低于所述虚拟功率-性能状态,所述第二功率-性能状态高于所述虚拟功率-性能状态。

11.如权利要求9所述的方法,还包括:

响应于进一步确定相应的最小驻留时间消逝以便随着时间的流逝产生比对应于第一功率-性能状态或第二功率-性能状态的功率消耗更紧密地对应于所述期望功率消耗的平均功率消耗,交替地选择和传送所述第一功率-性能状态和第二功率-性能状态。

12.如权利要求9所述的方法,其中在选择新的功率-性能状态之前,所述方法还包括确定所述一个或多个功能块已在当前功率-性能状态中操作了给定的时间量。

13.如权利要求9所述的方法,还包括:

相应块的活动水平是基于相应块中的采样信号而非电流或电压测量。

管理多个分立操作点以产生稳定虚拟操作点的方法及计算系统

技术领域

[0001] 本发明涉及计算系统,且更具体地涉及处理器的分立操作点的有效管理。

背景技术

[0002] 现代集成电路(IC)的功率消耗变成对半导体芯片的每次生成的日益增大的设计问题。当功率消耗增加时,更昂贵的冷却系统例如较大的风扇和散热器用于移除额外的热并防止IC故障。然而,冷却系统增加了系统成本。IC功率损耗限制不仅对便携式计算机和移动通信设备而且对高性能超标量微处理器也是问题,该微处理器可包括多个处理器核心或核心以及在核心内的多个管线。

[0003] IC例如现代互补型金属氧化物半导体(CMOS)芯片的功率消耗至少与表达式 fV^2 成比例。符号 f 是芯片的操作频率。符号 V 是芯片的操作电压。在现代微处理器中,参数 f 和 V 都可在IC的操作期间变化。例如,在操作期间,现代处理器允许用户在最大性能状态和最小功率状态之间选择一个或多个中间功率性能状态。最大性能状态包括最大操作频率,而最小功率状态包括最小操作频率。中间分立的功率-性能状态(P-状态)包括对操作频率和操作电压的组合的给定折合值。

[0004] 软件例如操作系统或固件或硬件可至少基于改变状态的投影时间、选定的功率限制、工作负荷特征和对应于当前工作负荷的来自片上功率监测器的输入来选择特定的P-状态。然而,操作频率和操作电压的所计算的组合一般不匹配对应于分立的给定P-状态的组合。因此,紧密匹配的给定P-状态被选择。一般,这个选定的P-状态可对应于比所计算的功率限制低的功率消耗。因此,选定的P-状态的性能比所计算的功率限制低。如果几个更分立P-状态被添加到处理器以提供操作频率和电压的更精细的粒度组合,则处理器的设计和测试成本增加。

发明内容

[0005] 鉴于上述内容,需要用于管理处理器的分立操作点的有效方法和机构。

[0006] 设想用于管理多个分立操作点以产生稳定的虚拟操作点的系统和方法。

[0007] 在一个实施方案中,处理器包括几个功能块和功率管理器。每个功能块产生对应于与相应的功能块相关的活动水平的数据。功率管理器在每个给定的采样间隔基于数据来确定功率消耗值一次。此外,功率管理器确定随着时间的流逝在给定的功率目标和功率消耗值之间的带正负号的累积差异。在一个实施方案中,功率目标可对应于处理器的热设计功率(TDP)。在各种实施方案中,滞后作用可用于避免无益的P-状态转变。在这样的实施方案中,如果带正负号的累积差异比负阈值小的量大于给定 $\delta(\text{delta})$,则功率管理器选择比当前P-状态低的功率性能状态(P-状态)。如果带正负号的累积差异比正阈值大的量大于给定 δ ,则功率管理器选择比当前P-状态高的P-状态。还设想在到另一P-状态的转变被允许之前可能需要在P-状态中的最小驻留的实施方案。

[0008] 当参考下面的描述和附图时将进一步认识到这些和其它实施方案。

附图说明

[0009] 图1是半导体芯片的功率-性能状态转变的一个实施方案的一般化方框图。

[0010] 图2是核心功率管理的一个实施方案的一般化方框图。

[0011] 图3是用于管理多个分立操作点以产生稳定的虚拟操作点的方法的一个实施方案的流程图。

[0012] 图4是核心功率管理系统的一个实施方案的一般化方框图。

[0013] 图5是用于管理多个分立操作点以产生稳定的虚拟操作点的方法的一个实施方案的流程图。

[0014] 虽然本发明容许各种修改和可选的形式,但是特定的实施方案作为例子在附图中示出并在本文被详细描述。然而应理解,附图和对其的详细描述并不是用来将本发明限制到所公开的特定形式,而是相反,本发明涵盖落在如所附权利要求所限定的本发明的精神和范围内的所有修改、等效和可选形式。

具体实施方式

[0015] 在下面的描述中,阐述了很多特定的细节以提供对本发明的彻底理解。然而,本领域中具有普通技能的人应认识到,可在没有这些特定细节的情况下实践本发明。在一些实例中,没有详细示出公知的电路、结构和技术以避免不使本发明难理解。

[0016] 参考图1,示出了半导体芯片的功率-性能状态转变100的一个实施方案。在说明功率与电压以及频率与电压之间的非线性(例如三次或二次)关系的图中示出了两条曲线。五个分立的功率-性能状态(P-状态)在图中被表示为P0到P4。示出了少量分立P-状态以简化该图。虽然只示出五个分立P-状态,但可支持其它数量的分立P-状态。

[0017] 在所示的图中,P-状态P4可对应于具有所有支持的分立状态的最低性能的分立状态,并包括最低操作频率。此外,P-状态P4可对应于具有所有支持的分立状态的最低功率消耗的分立状态,并包括最低操作电压。另一方面,P-状态P0可对应于具有所有支持的分立状态的最高性能的分立状态,并包括最高操作频率。此外,P-状态P0可对应于具有所有支持的分立状态的最高功率消耗的分立状态,并包括最高操作电压。一般,端点分立状态由P-状态P0和P4表示,并限定可预测性能的区域。因此,配置处理器以支持沿着频率与电压的非线性关系曲线的多个P-状态或操作点可为半导体芯片例如处理器提供功率的稳定的最佳利用和性能的输送。P-状态的管理可符合工业标准,例如最初由英特尔公司、微软公司和东芝公司发展的高级配置和电源接口(ACPI)标准。

[0018] 如在图中所示的,可为芯片选择功率目标₁(例如,期望功率消耗水平)。在一个实施方案中,选定的功率目标₁可对应于芯片的热设计功率(TDP)。也可以称为热设计点的热设计功率(TDP)表示计算机中的冷却系统能够消耗的最大数量的功率。可以对20瓦TDP设计膝上型处理器的冷却系统。因此,已确定冷却系统能够消耗20瓦而不超过处理器内的晶体管的最大结温度。TDP值可以不同,取决于生产芯片的芯片制造商。例如,一个制造商可将TDP值定义为在给定的最坏情况温度条件下在默认的电压电平处测量的功率值。另一制造商可将TDP值定义为当芯片执行与高功率病毒应用相对的一般应用时在给定的间隔上测量

的最大功率值。其它测量定义是可能的并被设想。

[0019] 在一个实施方案中,在裸片102的硅前模型上执行的功率模型可执行功率测量。在设计周期中的后期,可在测试阶段和调试阶段期间在实际制造的硅裸片上执行功率测量。在一个实施方案中,芯片的峰值功率值可由在核心上执行高功率应用的芯片的功能故障定义。TDP值一般小于峰值功率值。TDP值可用于为了分仓(bin)目的而设定芯片的操作电压和操作频率。

[0020] 图1中的值功率目标₁可表示所分配的TDP值。如图1所示,功率目标₁对应于在功率与电压的非线性关系曲线上的数据点A。数据点A对应于操作电压 V_2 。关于操作电压 V_2 将数据点A投影到频率与电压的非线性关系曲线上提供了数据点A'。数据点A'对应于操作频率 F_2 。由操作电压 V_2 和操作频率 F_2 的组合表示的操作点可为芯片提供功率的最佳利用和性能的输送。

[0021] 如上所述和在图中示出的,功率目标₁的操作点由数据点A'标识。然而,这个操作点并不由功率与频率的关系曲线上的分立P-状态表示。数据点A'位于P-状态 P_1 和 P_2 之间。为了减少功率消耗,P-状态 P_2 可被选择为相应芯片的初始操作点。操作电压 V_1 和操作频率 F_1 的相应组合可以是因而产生的选定的操作点。这个操作点对应于比值功率目标₁低的功率消耗值。值功率 P_2 指示对应于P-状态 P_2 的操作点的较低功率消耗值。

[0022] 芯片例如处理器可继续利用最初分配的P-状态处理工作负荷,直到(i)工作负荷明显改变,这引起所报告的活动水平中的明显变化,或(ii)初始TDP值改变,例如由功率监测软件或固件调节,这改变图中所示的功率目标值。例如,如果处理器正对在给定的时间点减半的功率负荷执行指令,则因而产生的总的消耗的电流和热能将明显减少。在一个实施方案中,可位于处理器内的功率管理器可探测这个条件,并相应地选择对应于较高的功率-性能操作点的不同P-状态。例如,功率管理器可确定将当前的P-状态 P_2 增高或升高到较高性能P-状态 P_1 。为了讨论的目的,较高性能P-状态可具有较低的号数。例如, P_0 可表示比 P_1 高的功率-性能状态。然而,符号表示可以是相反的,使得 P_0 表示比 P_1 低的功率-性能状态。用于符号表示的选定方法可以仅仅是关于偏好的问题。

[0023] P-状态的“压低(throttle)”包括将当前选择的P-状态减少一个(或多个)P-状态到较低功率消耗P-状态。相反,P-状态的“升高(boost)”包括将当前选择的P-状态增加一个(或多个)P-状态到较高性能P-状态。压低P-状态 P_2 包括将当前选择的P-状态 P_2 转变到较低功率-性能状态 P_3 。在图1中作为例子示出了升高和压低给定的P-状态例如P-状态 P_2 的简单图示。在一些实施方案中,当逻辑支持这个增加的复杂性时,每个升高操作和每个压低操作可使当前选择的P-状态转变两个或多个P-状态。P-状态转变可由处理器内的逻辑控制,并且从而是完备的系统。然而,在处理器或位于处理器外部的机架控制器上运行的功率管理软件可改变处理器的TDP值,这改变了图中的功率目标值。

[0024] 虽然由图1中的P-状态 P_2 表示的操作点消耗比功率目标₁值小的功率,由 P_2 表示的操作点也具有较小的性能。不是维持较低性能P-状态直到在所测量的活动水平中的明显变化,处理器可在两个分立P-状态之间转换,以便实现当前工作负荷的平均“虚拟”操作点。例如,功率管理器可对相同的工作负荷确定在升高到P-状态 P_1 之前利用P-状态 P_2 的时间的量。类似地,功率管理器可对相同的工作负荷确定在压低到P-状态 P_2 之前利用P-状态 P_1 的时间的量。这个过程可出现多次,直到工作负荷明显改变。在与操作电压和频率结合的特定P-状

态中花费的时间的平均值可具有处理器正利用由图1中的数据点A'表示的“虚拟”操作点的效应。这样的方法将使具有任意功率限制设置的灵活的功率管理能够通过建立在已经存在的分立P-状态上来为任何工作负荷实现稳定的虚拟操作点(或虚拟P-状态)。

[0025] 现在转到图2,示出了核心功率管理系统200的一个实施方案。示出了随着时间的流逝对操作电压210和所测量的活动水平220的变化。此外,示出了随着时间的流逝的功率目标235和平均功率与功率目标之比230。如图2所示,当活动水平经历明显的变化时和当活动水平是不变的时,都存在P-状态转变。在图中,P-状态转变与操作电压210中的变化相关。在图中标出一些P-状态值。例如,使用不变的活动水平220,P-状态 P_1 转变到具有如图所示的较低操作电压的P-状态 P_2 。当所测量的活动水平不变时在P-状态值之间转换可以被称为P-状态抖动(dithering)。P-状态抖动可用于维持在芯片上消耗的平均功率与功率目标例如TDP值之间的接近于一的比。通过维持接近于一的这样的比,芯片可试图最大化性能,同时仍然消耗期望量的功率(例如,由相应的冷却系统消耗的功率的量)。

[0026] 现在转到图3,示出了用于管理多个分立操作点以产生相对稳定的虚拟操作点的方法300的一个实施方案。为了讨论的目的,以连续的顺序示出了在本实施方案中和稍后描述的方法的随后的实施方案中的步骤。然而,一些步骤可以按与所示的不同的顺序出现,一些步骤可以同时被执行,一些步骤可以与其它步骤结合,且一些步骤可以在另一实施方案中缺少。

[0027] 在块302中,初始化芯片的裸片的功率使用目标。可使用用于选择功率使用目标的各种方法中的任一种,包括早些时候描述的方法。在块304中,在给定的工作负荷处确定裸片的初始分立的功率-性能状态(P-状态)。软件例如固件和/或硬件可确定P-状态。例如图中所述的过程可用于确定P-状态。接着在块306中测量裸片的平均功率消耗。稍后提供这样的测量的另外的细节。在块308中,所测量的平均功率消耗与功率目标比较,且差异被确定。在一个实施方案中,这个差异可与其它确定的差异累积。例如,可维持累积的差异值。虽然响应于探测到目标和所测量的功率之间的差异来发起到另一P-状态的转变,但是这样的方法可引起P-状态之间的无益的转变。因此,在一个实施方案中,各种技术用于防止这样的不需要的状态。在一个实施方案中,给定 δ 用于确定累积的差异何时足以引起P-状态转变的目的。例如,如果所测量的平均功率超过目标功率的量多于给定 δ ,则到较高P-状态的转变可被发起或以另外方式被允许。类似地,如果所测量的平均功率落在目标功率之下的量至少为给定 δ ,则到较低P-状态的转变可被发起或以另外方式被允许。注意,上述 δ 可用作按需要必须被超过或仅仅被满足的值。在一些情况下,到较高或较低功率状态的转变的给定 δ 可以是相同的(按照绝对值)。在其它实施方案中,对到较高状态的转变可使用与对到较低状态的转变不同 δ 值。所有这样的实施方案被设想。此外,在各种实施方案中, δ 值可以或可以不是可编程的。除了上述内容以外,另外的条件可用于确定P-状态转变是否可出现。例如,在一些实施方案中,在转变被允许之前,可能需要在P-状态中的最小驻留。在图5的讨论中将提供这样的最小驻留的进一步讨论。

[0028] 在块310中,当累积的差异达到给定 δ 时,到另一P-状态的转变可出现。例如,如果对应于上面的比较的带正负号的累积差异落在给定阈值之下,则功率管理器可选择比当前P-状态低的功率-性能状态(P-状态)。较低的P-状态可通常对应于小于功率目标的功率消耗值。类似地,如果随着时间的流逝对应于上面的比较的带正负号的累积差异超过给定阈

值,则功率管理器可选择比当前的P-状态高的P-状态。

[0029] 再次参考图2,活动水平220可跟踪工作负荷。如图2的实施方案中所示的,当平均功率与功率目标之比从一改变时,P-状态转变出现。在各种实施方案中,给定阈值偏差可用于确定P-状态转变何时出现。可选地,累积的带正负号的差异而不是比可存在于所测量的平均功率和功率目标之间。在提供另外的细节之前,描述了测量活动水平以跟踪工作负荷的一个实施方案。

[0030] 参考图4,示出了核心功率管理400的一个实施方案。在这里,核心102可以是任何集成电路(IC)。在一个实施方案中,核心102可以是处理器核心。处理器核心可具有裸片上指令和数据高速缓冲存储器。处理器核心可以是具有单个管线或多个管线的超标量处理器。在另一实施方案中,核心102可以是专用IC(ASIC)。任何晶体管系列可用于实现核心102。例子包括金属氧化物半导体场效应晶体管(MOSFET)和双极结晶体管(BJT)。

[0031] 功能块110可包括被配置成执行逻辑功能、数据存储器等晶体管。为了功率管理目的,功能块110可分成单元132a-132d。如本文使用的,由跟随有字母的参考数字表示的元件可共同由仅仅数字表示。例如,单元132a-132d可共同称为单元132。在一个实施方案中,单元132可以不对应于处理器的功能部件,例如记录器缓冲器、存储管理单元、执行单元等。更确切地,为了功率管理目的,可基于待采样的信号的类型来选择单元132。例如,在一个实施方案中,被选择来采样的信号包括按规定路线发送到本地时钟分布块的时钟启动信号。

[0032] 在特定的时钟周期期间选择哪些信号来采样可对应于该选择与单元132内的开关节点电容相关得多么好。待采样的选定信号例如时钟启动信号可与平面布置图中的功能块重叠。因此,分开例如单元132a和132b的划分可以不对应于平面布置图中的划分。单元132是消耗功率的单元,且该功率被实时测量。与当前的工作负荷相关的裸片的活动水平可对应于与待采样的选定信号相关的值或权重。

[0033] 在一个实施方案中,功率监测器130可用于收集来自单元132的数据,例如所有给定的采样信号的逻辑值。在一个实施方案中,可以用连续的方式扫描输出所采样的信号的值。因此,选定的信号可在单个时钟周期中从每个单元112被采样,并在下一采样被执行之前被连续地扫描输出。在收集数据之后,功率监测器130可计算功率消耗估计。一个监测器控制装置132可对应于每个单元132。在可选的实施方案中,监测器控制装置132可收集一个或多个单元112的数据,并计算那些单元112的总功率消耗估计。在又一实施方案中,一个监测器控制装置132(即,控制装置132a)可具有与一个或多个其它监测器控制装置132(即,控制装置132b-132d)的信号接口,以便收集来自一个或多个监测器控制装置132(即,控制装置132b-132d)的数据。于是,可计算一个或多个监测器控制装置132的功率消耗估计。

[0034] 信号样本120和输出数据122可以是用于功率管理目的的控制和数据信号。在功率监测器110和功能块130之间的接口信号可包括任何必要数量的信号和通信协议。在一个实施方案中,控制信号样本120可以只在选定的重复间隔期间例如每100个时钟周期在单个时钟周期内是有效的。在一个实施方案中,在控制信号样本120是有效的之后的给定数量的时钟周期,数据信号输出数据122可开始提供在每个时钟周期的不同采样信号的逻辑值。换句话说,数据信号输出数据122可用于扫描输出一系列值,其包括在特定的周期的所采样的信号的逻辑值。此外,在其它实施方案中,可能没有在每个监测器控制装置132和单元112对之间的单个信号对。在可选的实施方案中,额外的信号可被包括,以便使监测器控制装置132

轮询单元112,使单元112向监测器控制装置132承认它准备传送输出数据。

[0035] 在给定的时间间隔期间可采用多个样本。在计算活动水平之前使用的间歇的时钟周期的数量的确定可取决于所采样的数据的期望准确度和置信度。电子数据表或查找表可使用应用的真实功率消耗和来自采样的所估计的功率消耗的统计分析和测量来产生。置信水平和误差率可被选择成进一步发展统计分析。实时功率估计方法的例子包括2008年4月11日提交的标题为“Sampling Chip Activity for Real Time Power Estimation”的申请序列号12/101,598,该申请通过引用以其整体并入本文。

[0036] 当功率监测器130在重复的间隔上从接收自功能块110的数据计算功率消耗估计时,功率监测器130确定当前运行的应用的功率分布。这个确定被传送到功率管理器140。功率管理器140可改变功能块110的操作点,以便在应用高于(低于)阈值限制时减少(或增加)功率。例如,功率管理器140可引起当前P-状态转变到另一给定P-状态的升高或压低。

[0037] 在一个实施方案中,在上面指出的规定时间段期间,功率管理器140可计算功率监测器130所提供的功率分布和功率目标之间的带正负号的连续累积差异。再次,功率目标可以是热设计点(TDP)。累积差异可在每个给定的时间间隔结束时被计算为 $AccTdpDelta = AccTdpDelta + (TDP - RcvdPwr)$ 。在这里,变量 $AccTdpDelta$ 是带正负号的连续累积差异。变量TDP是带正负号的热设计功率或相等数量的热信用(thermal credit)。变量 $RcvdPwr$ 是从功率监测器130接收的功率消耗估计。该值可通过测量功能块110中的所采样的信号来跟踪活动水平。

[0038] 如果由变量 $RcvdPwr$ 表示的所测量的活动水平高于TDP,则累积值 $AccTdpDelta$ 朝着负值漂移。当累积值达到负的给定阈值时,功率管理器可确定压低当前P-状态。再次参考图1,压低当前P-状态的例子将是P-状态 P_1 转变到P-状态 P_2 。当活动水平在核心内为高时,这样的条件可出现。如果活动水平保持在高值,则随着时间的流逝,功率管理器可继续压低当前P-状态。

[0039] 如果所测量的活动水平低于TDP,则累积值 $AccTdpDelta$ 朝着正值漂移。当累积值达到正的给定阈值时,功率管理器140可确定升高当前P-状态。再次参考图1,升高当前P-状态的例子将是P-状态 P_2 转变到P-状态 P_1 。当活动水平在核心内为低时,这样的条件可出现。如果活动水平保持在低值,则随着时间的流逝,功率管理器可继续升高当前P-状态。

[0040] 当从功率监测器130发送的信息对应于实际活动水平和核心102内的功率消耗而不是温度信息时,功率管理器140可能提供对核心102中的潜在热问题的较快响应。在半导体芯片裸片的整个裸片上放置的模拟或数字热传感器可确定随着时间的流逝的温度波形。热传感器提供关于何时裸片由于增加的计算活动而在特定的区域中变热的信息。然而,这些传感器对上升热气流中的每个变化作出响应,而不管它是通过核心102中的功率消耗的计算相关的增加还是通过外部环境因素例如环境温度的升高来驱动。例如,在数据中心中的机架系统中的周围服务器可引起环境温度的升高。在特定核心内的开关电容的量可以不在时间间隔上改变,但传感器可报告由于环境温度的升高而引起的较高热能消耗。此外,在功率消耗的计算相关的增加和温度增加之间存在时间延迟。因此,虽然试图维持平均功率消耗和功率目标之间的一的比,与核心内的活动水平和开关电容相关的测量与温度的测量的关系可提供较好的结果。

[0041] 现在转到图5,示出了用于管理多个分立操作点以产生稳定的虚拟操作点的方法

500的另一实施方案。为了讨论的目的,以连续的顺序示出了在本实施方案中和稍后描述的方法的随后的实施方案中的步骤。然而,一些步骤可以按与所示的不同的顺序出现,一些步骤可以同时被执行,一些步骤可以与其它步骤结合,且一些步骤可以在另一实施方案中缺少。

[0042] 在块502中,半导体芯片执行一个或多个软件应用的指令。如果给定采样间隔的末尾被到达(条件块504),则在块506中,功率消耗估计被确定并传送到功率管理器。可通过对功能块中的选定信号进行采样并使相应的权重与如早些时候描述的所采样的信号相关来找到功率消耗估计。对应于所确定的功率估计值的计数的计数器可递增。在块508中,功率管理器可计算在给定的功率目标例如TDP和所接收的功率消耗值之间的带正负号的累积差异。如早些时候描述的计算可以是 $AccTdpDelta = AccTdpDelta + (TDP - RcvdPwr)$,其中变量RcvdPwr表示所接收的功率消耗值。

[0043] 注意,当P-状态转变出现时,带正负号的累积差异AccTdpDelta可在另一采样间隔出现的时为止仍然超过给定阈值。例如,从P-状态 P_2 到P-状态 P_1 的升高可由于值AccTdpDelta大于正升高阈值而出现。在P-状态转变之前,在下一时间间隔,带正负号的累积差异AccTdpDelta可仍然大于正升高阈值。可能没有使所测量的功率消耗值超过TDP值的足够时间。因此,在确定当前的P-状态在给定的时间提供最佳功率-性能操作点之前,可能快速继续升高当前P-状态。

[0044] 为了避免如上所述的快速P-状态转变,随后的P-状态转变可以不被允许在当前P-状态转变之后的给定时间出现。给定时间可以被称为最小驻留。在一个实施方案中,计数器可用于确定许可数量的时间是否在当前P-状态转变之后消逝。计数器值可与给定阈值比较。可选地,计数器可装有给定阈值并递减到零的值。当这个许可数量的时间消逝时,另一P-状态转变可出现。

[0045] 此外,可使用累积差异AccTdpDelta的值老化。值老化帮助防止芯片裸片的过热。老化的累积差异可保持在长时间内在高功率P-状态之间的转变或抖动。当P-状态被压低时,累积差异AccTdpDelta可继续累积。在一段时间之后,累积差异可以不再表示对升高可采用的实际热能净空。因此,不时地,累积差异AccTdpDelta可在给定的时间段结束时被设置到重置值。重置值可从累积差异AccTdpDelta的当前值的一部分改变到零。存储在配置寄存器中的值可用于确定重置值。例如,第一存储值可对应于等于累积差异AccTdpDelta的重置值。第二存储值可对应于等于改变的累积差异AccTdpDelta(例如,除以2、除以4等)的重置值。第三存储值可对应于等于零的重置值。

[0046] 在一个例子中可以递减的计数器可设置上面提到的给定时间段。计数器可装入存储在配置寄存器中的值N。在N次采样出现之后,其中累积差异AccTdpDelta在每次采样或时间间隔结束时被更新,累积差异AccTdpDelta可被重置。每当累积差异AccTdpDelta改变符号时,装有值N的计数器可被重置。

[0047] 再次参考图5中的方法500,如果P-状态最小驻留时间段没有达到(条件块510),则执行关于N次采样是否出现的检查。在每次采样间隔结束时,确定裸片的新的功率消耗值,并对累积差异AccTdpDelta计算新值。如果N次采样的计数出现(条件块516),则在块518中,累积差异AccTdpDelta被设置为如上所述的给定重置值。

[0048] 如果P-状态最小驻留时间段已经达到(条件块510),则执行关于升高和压低阈值

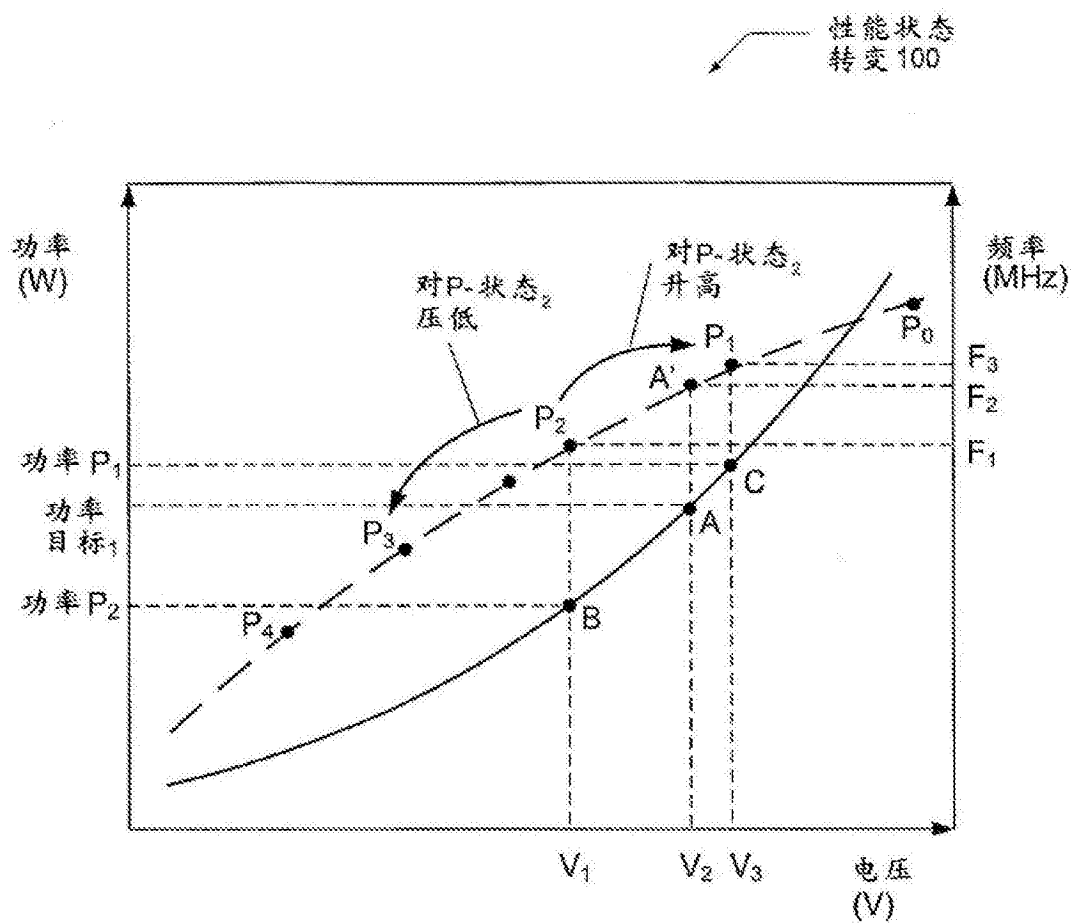
的检查。可执行比较以确定带正负号的累积差异AccTdpDelta是否超过阈值。例如,累积差异AccTdpDelta可大于正升高阈值。可选地,累积差异AccTdpDelta可小于负压低阈值。如果累积差异AccTdpDelta超过阈值(条件块512),则在块514中,为芯片裸片选择相应的下一P-状态。例如,如果带正负号的累积差异AccTdpDelta大于正(升高)阈值,则到比当前P-状态高的功率-性能P-状态的转变可能出现。可选地,如果带正负号的累积差异AccTdpDelta小于负压低阈值,则到比当前P-状态低的功率-性能P-状态的转变可能出现。

[0049] 在一个实施方案中,如果处理器包括多个核心,则功率消耗估计可在每个核心内被计算。此外,每个核心可确定带正负号的累积差异AccTdpDelta。当多个核心中的任一个超过升高或压低阈值时,整个处理器的P-状态可相应地转变。接着方法500的控制流程移动到条件块516。

[0050] 注意,上述实施方案可包括软件。在这样的实施方案中,表示所述系统和/或方法的程序指令和/或数据库(都被称为“指令”)可被输送或存储在计算机可读介质上。一般来说,计算机可访问存储介质可包括由计算机在使用期间可访问的任何存储介质,以向计算机提供指令和/或数据。例如,计算机可访问存储介质可包括存储介质,例如磁性或光学介质,例如磁盘(固定或可移动)、磁带、CD-ROM或DVD-ROM、CD-R、CD-RW、DVD-R、DVD-RW或蓝光。存储介质还可包括经由外围接口例如通用串行总线(USB)接口等可访问的易失性或非易失性存储介质,例如RAM(例如同步动态RAM(SDRAM)、双数据率(DDR、DDR2、DDR3等)SDRAM、低功率DDR(LPDDR2等)SDRAM、Rambus DRAM(RDRAM)、静态RAM(SRAM)等)、ROM、闪存、非易失性存储器(例如,闪存)。存储介质可包括微机电系统(MEMS)以及经由通信介质例如网络和/或无线链路可访问的存储介质。

[0051] 此外,指令可包括以编程语言例如C或设计语言(例如HDL)如Verilog、VHDL或数据库格式例如GDS II流格式(GDSII)的硬件功能的行为级描述或寄存器传输级(RTL)描述。这些指令可被读取并用于制造包括系统(或系统的部分)的硬件。在一些情况下,描述可由合成工具(例如,在计算设备上运行的程序代码)读取以形成设计的实现。例如,这样的工具可用于合成描述以产生包括来自合成库的一系列门的网络表。网络表可通常包括也表示包括系统的硬件的功能的一组门。网络表可接着被放置并按规定路线发送以产生描述将应用于掩模的几何形状的数据集。掩模可接着在各种半导体制造步骤中用于产生对应于系统的一个或多个半导体电路。可选地,在计算机可访问存储介质上的指令按需要可以是网络表(有或没有合成库)或数据集。此外,指令可通过基于硬件型的仿真器例如来自厂商Cadence®、EVE®和Mentor Graphics®的仿真器用于仿真的目的。例如,在这样的实施方案中,指令可用于配置基于FPGA的硬件以根据设计来执行。很多这样的实施方案是可能的并被设想。

[0052] 虽然相当详细地描述了上面的实施方案,但是一旦上面的公开被充分认识到,很多变化和修改就将对本领域技术人员变得明显。意图是下面的权利要求被解释为包括所有这样的变化和修改。



关键字:

—— 功率与电压的关系曲线

- - - 频率与电压的关系曲线

图1

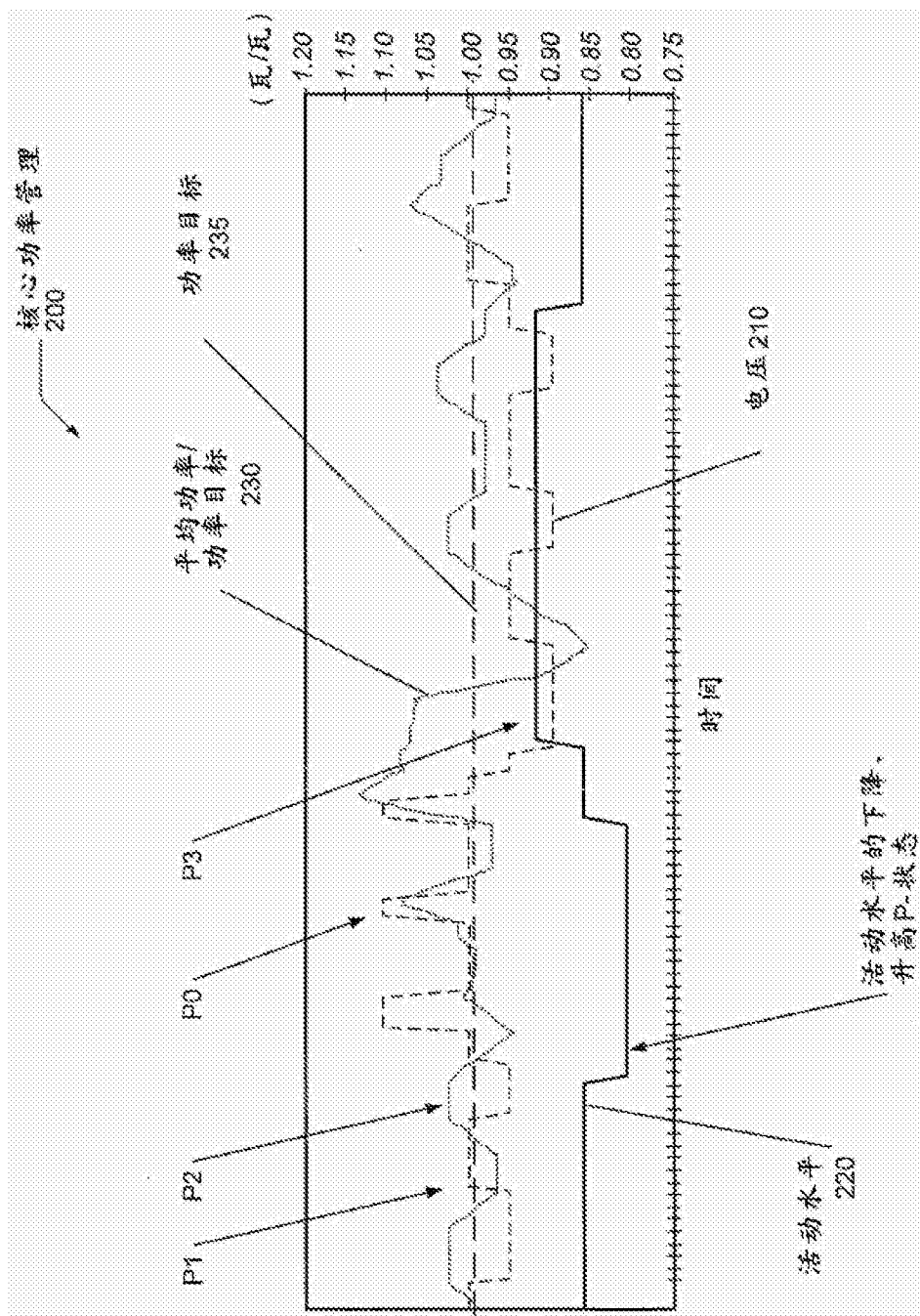


图2

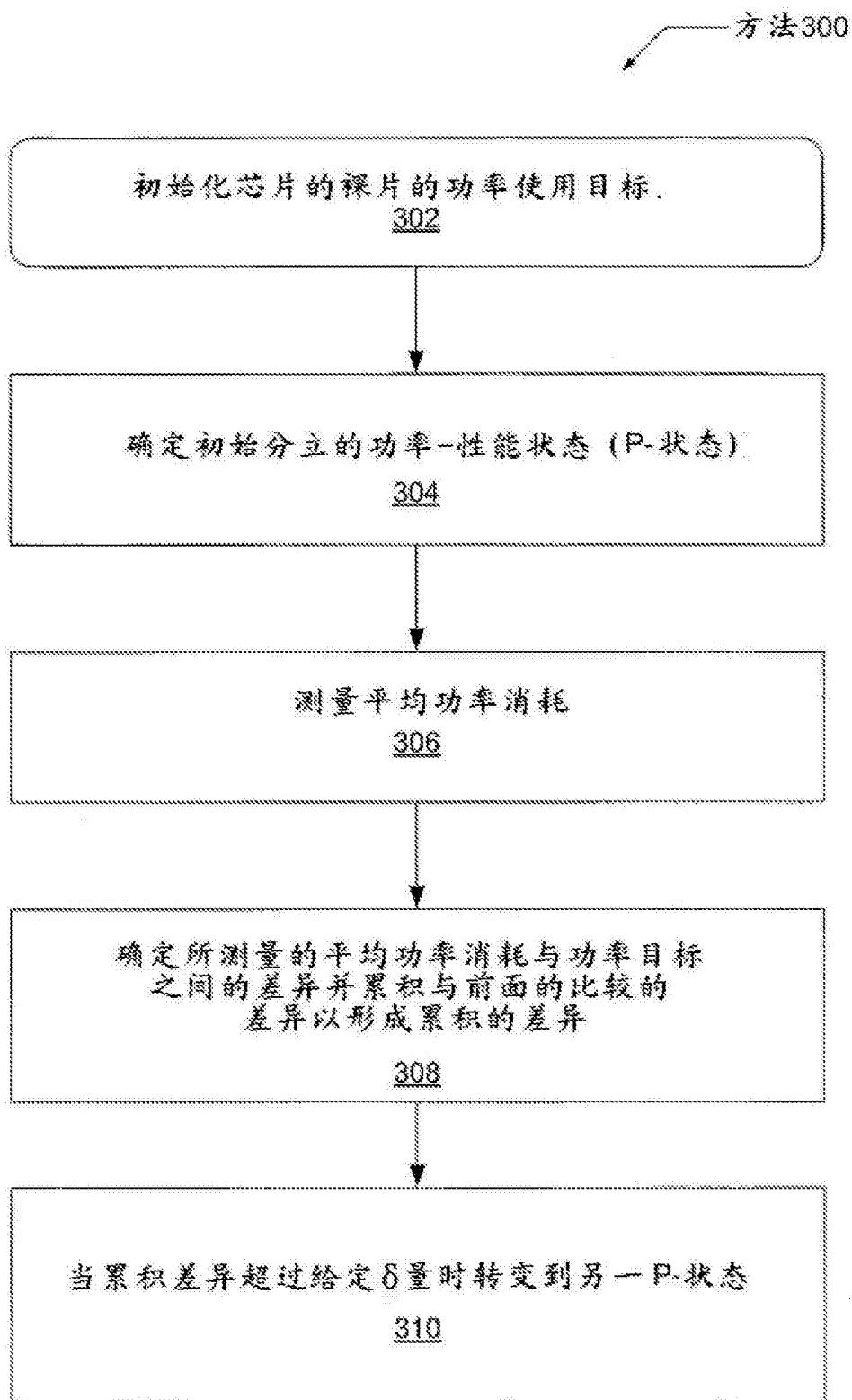


图3

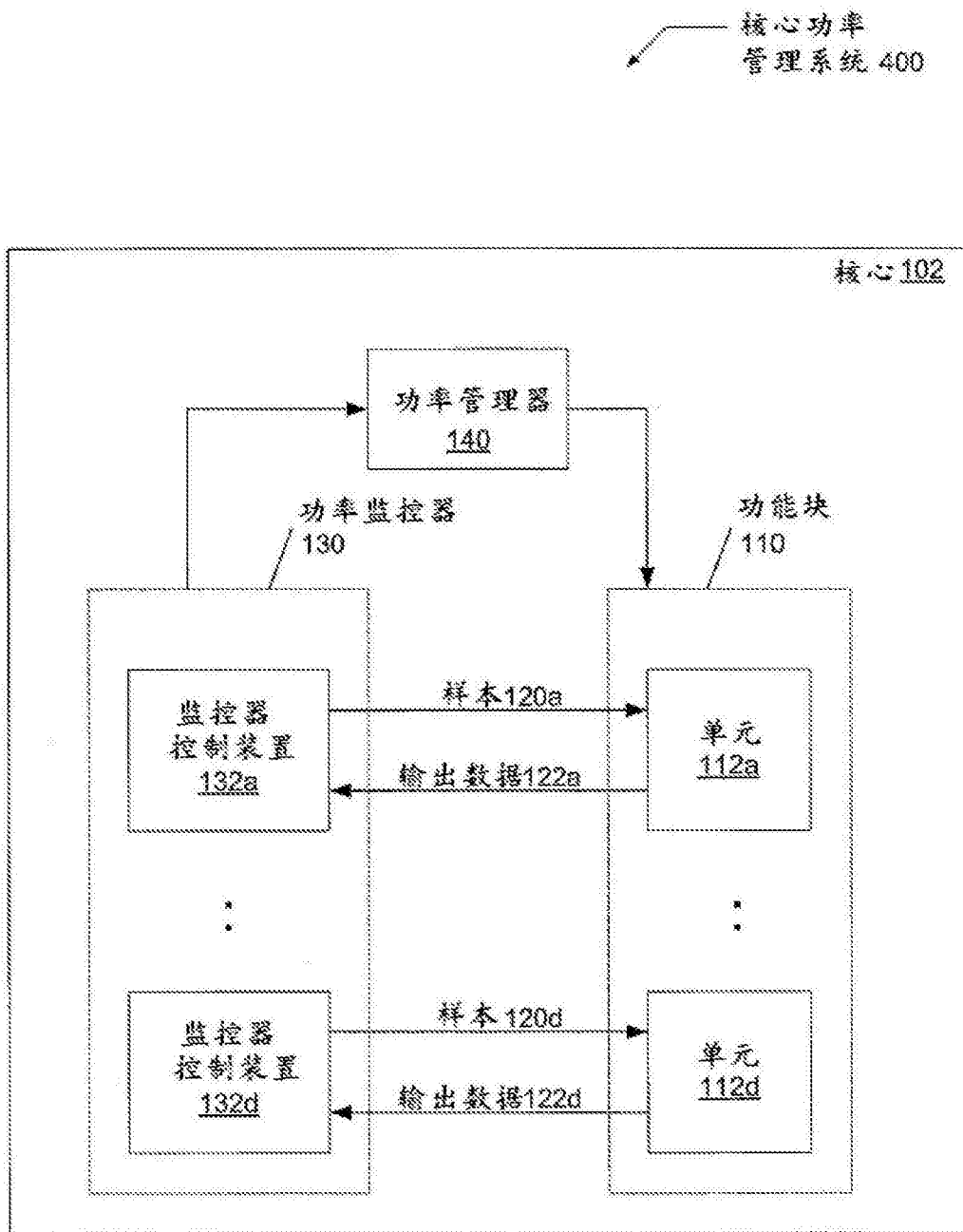


图4

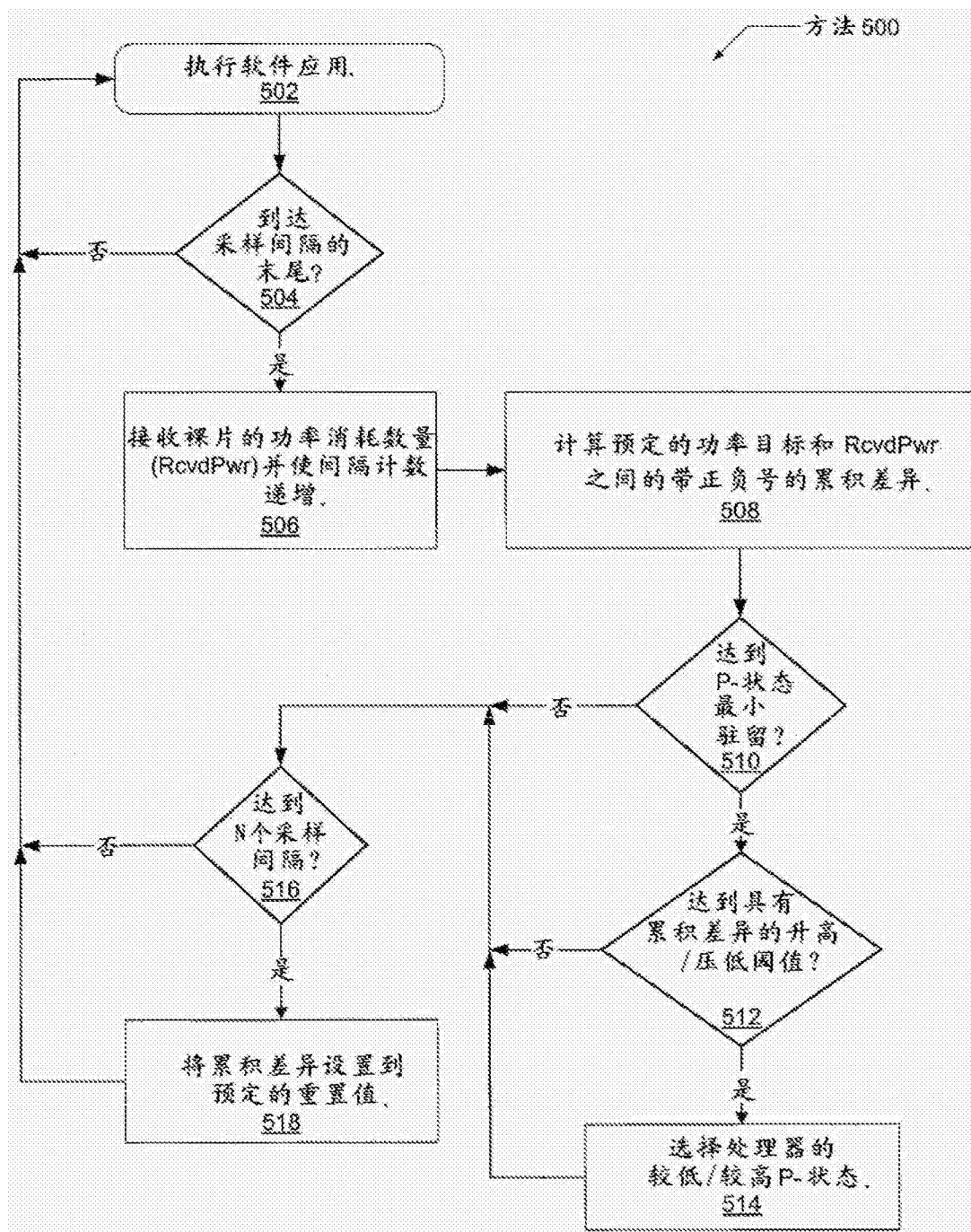


图5