



(12) 发明专利

(10) 授权公告号 CN 114631082 B

(45) 授权公告日 2023. 10. 24

(21) 申请号 202080073396.6

(22) 申请日 2020.10.29

(65) 同一申请的已公布的文献号
申请公布号 CN 114631082 A

(43) 申请公布日 2022.06.14

(30) 优先权数据
16/669,973 2019.10.31 US

(85) PCT国际申请进入国家阶段日
2022.04.20

(86) PCT国际申请的申请数据
PCT/US2020/057966 2020.10.29

(87) PCT国际申请的公布数据
W02021/087115 EN 2021.05.06

(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 保罗·莫耶 约翰·凯利

(74) 专利代理机构 上海胜康律师事务所 31263
专利代理师 李献忠 张华

(51) Int.Cl.
G06F 12/126 (2016.01)

(56) 对比文件
CN 101866318 A, 2010.10.20
CN 109478165 A, 2019.03.15
CN 101866322 A, 2010.10.20
US 2017315932 A1, 2017.11.02
US 2008086600 A1, 2008.04.10
US 2019004970 A1, 2019.01.03
WO 2018098225 A2, 2018.05.31

审查员 赵晓敏

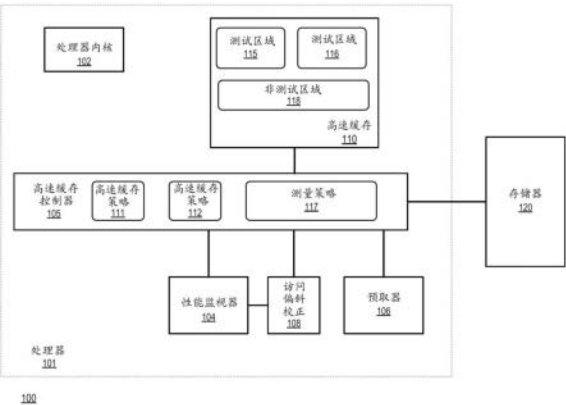
权利要求书3页 说明书11页 附图4页

(54) 发明名称

高速缓存访问测量偏斜校正

(57) 摘要

一种处理器[101],包括高速缓存[110],所述高速缓存具有两个或更多个测试区域[115, 116]和较大的非测试区域[118]。所述处理器还包括:高速缓存控制器[105],所述高速缓存控制器对所述高速缓存的不同测试区域应用不同的高速缓存替换策略[111, 112];和性能监视器[104],所述性能监视器测量所述不同测试区域的性能指标,诸如每个测试区域处的高速缓存命中率。基于所述性能指标,所述高速缓存控制器为所述非测试区域选择高速缓存替换策略,诸如选择与所述不同测试区域中具有较好性能指标的测试区域相关联的替换策略。响应于对所述不同测试区域的访问量的差值超过阈值,所述处理器对存储器访问测量进行偏斜校正。



1. 一种方法,所述方法包括:

监测对高速缓存的第一区域和所述高速缓存的第二区域中的每一者的访问;

基于对所述高速缓存的所述第一区域的受监测访问的第一数量和对所述高速缓存的所述第二区域的受监测访问的第二数量来选择测量策略,其中所选的测量策略控制如何测量对所述高速缓存的所述第一区域和所述第二区域的访问;以及

基于对所述高速缓存的所述第一区域的受监测访问的所述第一数量和对所述高速缓存的所述第二区域的受监测访问的所述第二数量并且基于所述测量策略来设置用于所述高速缓存的第三区域的高速缓存替换策略。

2. 如权利要求1所述的方法,其中选择所述测量策略包括:

确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的差值;

响应于所述差值小于阈值而选择第一测量策略;以及

响应于所述差值大于或等于所述阈值而选择第二测量策略。

3. 如权利要求2所述的方法,其中:

所述第一测量策略包括允许改变所述高速缓存替换策略;并且

所述第二测量策略包括阻止改变所述高速缓存替换策略。

4. 如权利要求2所述的方法,其中:

所述第一测量策略包括将第一权重应用于对所述第一区域的所述受监测访问以及将第二权重应用于对所述第二区域的所述受监测访问;

所述第二测量策略包括将第三权重应用于对所述第一区域的所述受监测访问以及将第四权重应用于对所述第二区域的所述受监测访问;并且

其中设置所述高速缓存替换策略包括基于对所述第一区域的加权的受监测访问和对所述第二区域的加权的受监测访问来设置所述高速缓存替换策略。

5. 如权利要求4所述的方法,其中:

所述第一权重与所述第二权重相同;并且

所述第三权重与所述第四权重不同。

6. 如权利要求1所述的方法,其中选择所述测量策略包括:

确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的第一差值;

基于所述第一差值选择用于所述第一区域的第一权重和用于所述第二区域的第二权重;

基于所述第一权重对所述第一区域的受监测访问进行加权以及基于所述第二权重对所述第二区域的受监测访问进行加权;并且

其中设置所述高速缓存替换策略包括基于对所述第一区域的所述加权的受监测访问和对所述第二区域的所述加权的受监测访问来设置所述高速缓存替换策略。

7. 如权利要求1所述的方法,其中:

监测对所述第一区域和所述第二区域中的每一者的访问包括:

监测对所述第一区域和所述第二区域中的每一者的第一类型的访问以及监测对所述第一区域和所述第二区域中的每一者的第二类型的访问;并且

选择所述测量策略包括基于以下各项选择所述测量策略:

对所述第一区域的所述第一类型的访问次数与对所述第二区域的所述第一类型的访

问次数之间的第一差值;以及

对所述第一区域的所述第二类型的访问次数与对所述第二区域的所述第二类型的访问次数之间的第二差值。

8.如权利要求7所述的方法,其中所述第一类型包括需求访问,并且所述第二类型包括预取访问。

9.如权利要求1所述的方法,其中设置所述高速缓存替换策略包括基于所述高速缓存的所述第一区域处的高速缓存命中率和所述高速缓存的所述第二区域处的高速缓存命中率来设置所述高速缓存替换策略。

10.如权利要求1所述的方法,其中设置所述高速缓存替换策略包括基于所述高速缓存的所述第一区域处的高速缓存未命中率和所述高速缓存的所述第二区域处的高速缓存未命中率来设置所述高速缓存替换策略。

11.一种方法,所述方法包括:

监测对高速缓存的第一区域和所述高速缓存的第二区域中的每一者的访问;

基于对所述高速缓存的所述第一区域的受监测访问的第一数量和对所述高速缓存的所述第二区域的受监测访问的第二数量来选择散列值;以及

基于所述所选的散列值生成所述高速缓存的第三区域的高速缓存地址值;以及

基于所述高速缓存地址值将数据存储在所述高速缓存的所述第三区域处。

12.如权利要求11所述的方法,其中选择所述散列值包括:

确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的差值;

响应于所述差值小于阈值而选择第一散列值;以及

响应于所述差值大于或等于所述阈值而选择第二散列值。

13.如权利要求11所述的方法,所述方法还包括:

基于对所述高速缓存的所述第一区域的受监测访问的所述第一数量和对所述高速缓存的所述第二区域的受监测访问的所述第二数量来改变与所述高速缓存的所述第一区域相关联的地址。

14.一种处理器,所述处理器包括:

高速缓存,所述高速缓存包括第一区域、第二区域和第三区域;以及

高速缓存控制器,所述高速缓存控制被配置为:

监测对所述第一区域和所述第二区域中的每一者的访问;

基于对所述第一区域的受监测访问的第一数量和对所述第二区域的受监测访问的第二数量来选择测量策略,其中所选的测量策略控制如何测量对所述第一区域和所述第二区域的访问;并且

基于对所述第一区域的受监测访问的所述第一数量和对所述第二区域的受监测访问的所述第二数量并且基于所述测量策略来设置用于所述第三区域的高速缓存替换策略。

15.如权利要求14所述的处理器,其中所述高速缓存控制器被配置为通过以下步骤来选择所述测量策略:

确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的差值;

响应于所述差值小于阈值而选择第一测量策略;以及

响应于所述差值大于或等于所述阈值而选择第二测量策略。

16. 如权利要求15所述的处理器,其中:

所述第一测量策略包括允许改变所述高速缓存替换策略;并且

所述第二测量策略包括阻止改变所述高速缓存替换策略。

17. 如权利要求15所述的处理器,其中:

所述第一测量策略包括将第一权重应用于对所述第一区域的所述受监测访问以及将第二权重应用于对所述第二区域的所述受监测访问;

所述第二测量策略包括将第三权重应用于对所述第一区域的所述受监测访问以及将第四权重应用于对所述第二区域的所述受监测访问;并且

其中设置所述高速缓存替换策略包括基于对所述第一区域的加权的受监测访问和对所述第二区域的加权的受监测访问来设置所述高速缓存替换策略。

18. 如权利要求17所述的处理器,其中:

所述第一权重与所述第二权重相同;并且

所述第三权重与所述第四权重不同。

19. 如权利要求14所述的处理器,其中所述高速缓存控制器被配置为通过以下步骤来选择所述测量策略:

确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的第一差值;

基于所述第一差值选择用于所述第一区域和所述第二区域的第一权重和第二权重;

基于所述第一权重对所述第一区域的受监测访问进行加权以及基于所述第二权重对所述第二区域的受监测访问进行加权;并且

其中设置所述高速缓存替换策略包括基于对所述第一区域的所述加权的受监测访问和对所述第二区域的所述加权的受监测访问来设置所述高速缓存替换策略。

20. 如权利要求14所述的处理器,其中所述高速缓存替换策略设置所述高速缓存的所述第三区域的预取策略。

高速缓存访问测量偏斜校正

背景技术

[0001] 为了提高存储器访问效率,许多处理器采用包括一个或多个高速缓存的存储器层次结构来存储存储在主存储器中的数据子集。处理器基于存储器管理方案来在高速缓存与主存储器之间以及在各个高速缓存之间移动数据,所述存储器管理方案取决于处理器设计,但通常优先考虑最近访问的数据的高速缓存。例如,一些处理器采用一个或多个预取器来预测在不久的将来可能被处理器内核访问的数据并将该数据从主存储器复制到一个或多个高速缓存。为了将新数据移动到一个或多个高速缓存中,处理器通常基于指定的替换方案(例如最近最少使用(LRU)替换方案)来选择先前存储的数据以进行逐出,从而处理器逐出存储不被处理器内核访问的时间量最长的数据的高速缓存条目。然而,在许多情况下,LRU替换方案与在处理器内核处执行的指令的存储器访问模式不对应,导致存储器访问效率较低,而这是不理想的。

附图说明

[0002] 通过参看附图,可更好地理解本公开,并且它的许多特征和优点对本领域技术人员来说变得显而易见。在不同附图中使用相同附图标记指示类似或相同项。

[0003] 图1是根据一些实施方案的处理系统的框图,所述处理系统对高速缓存的不同测试区域的访问测量进行偏斜校正以允许访问指标确定用于高速缓存的非测试区域的高速缓存替换策略。

[0004] 图2是根据一些实施方案的图1的处理系统的访问偏斜校正模块的框图,所述访问偏斜校正模块对高速缓存的不同测试区域的访问测量进行偏斜校正。

[0005] 图3是根据一些实施方案的图1的处理系统的访问偏斜校正模块的另一个示例的框图,所述访问偏斜校正模块对高速缓存的不同测试区域的访问测量进行偏斜校正。

[0006] 图4是示出根据一些实施方案的图1的处理系统的示例的图,所述处理系统对高速缓存的测试区域的访问测量进行偏斜校正。

[0007] 图5是根据一些实施方案的用于高速缓存的不同测试区域的访问测量的方法的流程图。

[0008] 图6是根据一些实施方案的处理器框图,所述处理器基于高速缓存的测试区域处的访问测量的差值来改变高速缓存的地址散列。

具体实施方式

[0009] 图1至图6示出用于对高速缓存的不同测试区域的访问测量进行偏斜校正的技术,由此支持对每个测试区域的性能指标的准确测量,其中所述性能指标又支持在高速缓存的非测试区域处应用高速缓存替换策略。为了说明,在一些实施方案中,处理器包括具有两个或更多个测试区域和较大的非测试区域的高速缓存。所述处理器还包括:高速缓存控制器,所述高速缓存控制器对高速缓存的不同测试区域应用不同的高速缓存替换策略;和性能监视器,所述性能监视器测量不同测试区域的性能指标,诸如每个测试区域处的高速缓存命

中率。基于性能指标,高速缓存控制器为非测试区域选择高速缓存替换策略,诸如选择与不同测试区域中具有较好性能的测试区域相关联的替换策略。然而,在一些情况下,存储器访问模式导致不同测试区域之间的偏斜的存储器访问,并且这些偏斜的访问又使性能指标偏斜,对于非测试区域,可能会导致不那么有效的高速缓存替换策略。通过如本文进一步描述对存储器访问测量进行偏斜校正,处理器用于选择在更多情况中更有效的高速缓存替换策略,从而提高总体高速缓存效率。

[0010] 为了进一步举例说明,在一些实施方案中,处理器测量两个测试区域中的每一者的高速缓存命中率并选择与具有较高高速缓存命中率的测试区域相关联的高速缓存替换策略。然而,如果测试区域中的给定测试区域接收比另一个测试区域明显更多的存储器访问,则不同测试区域的高速缓存命中率偏斜,即,比它们在大多数存储器访问模式下的高速缓存命中率高或低。特别地,在一些情况下,处理器执行指令集的子集,使得执行该子集导致对高速缓存的访问模式,所述访问模式不代表与整个指令集相关联的存储器访问模式。这种对高速缓存的访问模式在本文中被称作非代表性存储器访问模式。在一些情况下,非代表性存储器访问模式会导致与平均不那么有效的高速缓存替换策略相关联的高速缓存测试区域,暂时具有比与更有效的高速缓存替换策略相关联的高速缓存测试区域高的高速缓存命中率。通过对不同高速缓存区域的访问测量进行偏斜校正,如本文进一步描述,处理器阻止由于非代表性存储器访问模式的出现而选择不那么有效的高速缓存替换策略。

[0011] 在一些实施方案中,为了对访问测量进行偏斜校正,高速缓存控制器将在滑动时间窗口内对两个测试区域中的每一者的访问次数进行比较。响应于访问次数的差值超过阈值,处理器改变访问测量策略,所述访问测量策略控制如何测量对每个测试区域的访问以选择高速缓存替换策略。例如,在一些实施方案中,改变访问测量策略,使得在指定的时间量中从替换策略选择过程省去对测试区域中的一者的访问,从而对访问测量进行偏斜校正。在其他实施方案中,对高速缓存替换策略的选择是基于不同的性能标准的加权比较,其中每个测试区域的权重是由访问测量策略指示。为了对访问测量进行偏斜校正,高速缓存控制器改变所指示权重中的一者或多者。

[0012] 在一些实施方案中,处理器采用对测试区域的访问次数的比较来平衡对高速缓存的不同区域的访问。为了说明,在一些实施方案中,高速缓存控制器在访问高速缓存(从高速缓存读取或写入到高速缓存)时使用指定的散列表,其中散列表预期会分散访问,从而提高总体高速缓存利用率。然而,对测试区域的相对较高的访问次数表明散列值不足以分散访问。因此,响应于对高速缓存的测试区域的访问次数的差值超过阈值,高速缓存控制器将散列值改变为不同的指定值,从而提高高速缓存利用率和存储器访问效率。

[0013] 图1示出根据一些实施方案的处理系统100,所述处理系统基于高速缓存的不同测试区域的访问指标对高速缓存的一部分应用高速缓存老化策略。处理系统100通常执行按计算机程序的形式组织的指令集以便代表电子装置实施任务。因此,在不同实施方案中,处理系统100在各种电子装置(诸如桌上型或膝上型计算机、服务器、智能电话、平板计算机、游戏控制台等)中的任一者中使用。

[0014] 为了便于指令的执行,处理系统100包括具有处理器内核102的处理器101、高速缓存控制器105、访问偏斜校正模块108、高速缓存110,并且还包括存储器120。在一些实施方案中,处理器内核102、高速缓存控制器105和高速缓存110形成于相同半导体裸片上以至少

形成处理器101的一部分,并且存储器120形成于不同的半导体裸片上并且经由一个或多个总线或信号线连接到处理器101。在一些实施方案中,存储器120的至少一部分与处理器101的模块形成于同一个半导体裸片上。

[0015] 高速缓存110和存储器120各自是存储数据的存储器装置,并且在不同实施方案中,是随机存取存储器(RAM)存储器模块、非易失性存储器装置(例如,快闪存储器)等中的一者或多者。如本文进一步描述的,高速缓存110和存储器120形成处理系统100的存储器层次结构。此外,在一些实施方案中,处理系统100的存储器层次结构包括其他存储器模块,诸如图1未示出的额外高速缓存。

[0016] 处理器内核102包括指令管线,所述指令管线具有例如用于快取指令的快取级、用于将每个快取的指令解码为一个或多个操作的解码级、用于执行操作的执行级和用于退出操作已完成执行的指令的退出级。基于指令生成的一组操作在本文中被称作存储器访问指令,并且是从处理系统100的存储器层次结构中加载数据的请求(加载操作),以及在处理系统100的存储器层次结构中存储数据的请求。由处理器内核102生成的存储器访问请求表示加载或存储正在使用或已知被执行程序需要的数据的请求,并且在本文中被称作需求请求。

[0017] 高速缓存控制器105是通过在高速缓存110与存储器120之间移动数据来管理处理系统100的存储器层次结构的模块。特别地,高速缓存110是包括多个条目的存储器装置,其中每个条目与存储器120的一个或多个地址相关联。响应于存储器访问请求,高速缓存控制器105查阅存储于高速缓存110的一个或多个条目处的数据以确定所述条目中的任一者是否存储与存储器访问请求所针对的存储器地址相关联的数据。如果是这样,高速缓存控制器105指示高速缓存命中并且通过将数据存储在条目处(在存储操作的情况下)或通过所识别条目处的数据提供到存储器内核102(在加载操作的情况下)在所识别条目处满足存储器访问请求。如果与存储器地址相关联的数据未存储在高速缓存110的条目处,则高速缓存控制器105指示高速缓存未命中并且向存储器120发出对所述存储器地址处的数据的请求。高速缓存控制器105随后将数据存储在高速缓存110的条目处并且满足所述条目处的存储器访问请求。

[0018] 为了支持提高的存储器访问效率,处理器101包括预取器106,所述预取器监测向高速缓存控制器105发出的需求请求以识别模式并基于所识别的模式预测在不久的将来将是处理器内核102的需求请求的目标的存储器地址。预取器106向高速缓存控制器发出以所识别的存储器地址为目标的存储器访问请求,本文中被称为预取请求。高速缓存控制器105通过以下步骤类似于需求请求来处理预取请求:首先识别与预取请求所针对的存储器地址相关联的数据是否已存储在高速缓存110的条目处,并且如果还没有存储,则向存储器120请求数据并将所述数据存储在高速缓存110的条目处。

[0019] 在一些实施方案中,高速缓存110的每个条目与存储器地址的子集相关联,并且仅与该子集相关联的数据被准许存储在所述条目处。例如,在一些实施方案中,高速缓存是具有M个组并且所述M个组中的每一者包括N个条目的N路组关联高速缓存。M个组中的每一者被指派存储器地址的不同子集,使得如果数据与某条目被指派的子集中的存储器地址相关联,则高速缓存控制器105仅将数据存储在所述条目处。当一组的所有条目存储可能是存储器访问的目标的数据(本文中被称为“有效数据”)时,该组在本文中被称作满的。此外,当高

速缓存控制器105接收到将存储在满组处的数据时,它选择该组的一个条目来逐出(即,从该组移除)并用接收到的数据替换。高速缓存控制器105用来选择要逐出的条目的方案被称为高速缓存110的替换策略。在不同的实施例方案中,替换策略结合了许多标准中的任一者,诸如在存储条目时指派给条目的初始年龄值、年龄值的变化、是否响应于需求请求或预取请求来取出条目处的数据等。

[0020] 在一些实施方案中,高速缓存控制器105对高速缓存110的不同区域应用不同的替换策略,其中每个区域包括多个条目。在至少一个实施方案中,每个区域仅在其替换策略方面有所不同。因此,一个区域的条目能够用于任何组,并且不同的区域能够包括同一组中的不同条目。在图1所示的示例中,高速缓存110包括三个区域:测试区域115和116以及非测试区域118。此外,高速缓存控制器105存储在至少一个方面不同的替换策略111和112,诸如指派给条目的初始年龄值、响应于指定标准调整年龄值的量等。在一些实施方案中,替换策略111为响应于预取请求而存储数据的条目规定指定的初始年龄,并且为响应于非预取请求而存储数据的条目规定不同的指定初始年龄。如本文所使用的,非预取请求是指不是以应用老化策略的高速缓存层级为目标的预取请求的存储器访问。非预取请求的示例包括需求请求、响应于指令高速缓存未命中而生成的存储器访问请求、硬件页表遍历请求、由处理器内核102生成的推测性存储器访问请求(例如,在分支指令之后的指令的推测性执行期间生成的存储器访问请求)以及在其他高速缓存层级处生成的预取请求。相反地,替换策略112为响应于预取请求而存储数据的条目和响应于非预取请求而存储数据的条目规定相同的初始年龄。在一些实施方案中,替换策略111因此对于更频繁地访问预取数据的存储器访问模式更有效,并且替换策略112对于以大约相同频率访问预取数据和非预取数据的存储器访问模式更有效。然而,处理器内核102的存储器访问模式随时间而变化,使得高速缓存110在不同时间在不同老化策略下更有效地操作。因此,对于非测试区域118,高速缓存控制器105取决于测试区域115和116的性能特征在老化策略111和112之间进行选择。

[0021] 为了说明,处理器101包括性能监视器104,所述性能监视器监测和记录测试区域115和116中的每一者的性能特征,例如高速缓存命中率、高速缓存未命中率等。基于测试区域115和116的一个或多个性能特征或其组合,高速缓存控制器选择替换策略111和112中的一者并将其应用于非测试区域118。例如,如果测试区域115的高速缓存命中率高于测试区域116的高速缓存命中率,则高速缓存控制器105选择测试区域115的替换策略(例如,替换策略111)并将其应用于非测试区域118。在一些实施方案中,非测试区域118包括高速缓存110的大部分条目。因此,通过选择非测试区域118的老化策略作为更有效的老化策略,如测试区域115和116的性能所示,高速缓存控制器105提高高速缓存110的总体效率。此外,随着测试区域115和116的性能特征随时间改变,高速缓存控制器105将替换策略改变为对于给定的一组性能特征更有效的策略。

[0022] 性能监视器104基于测量策略117测量测试区域115和116的性能特征。在一些实施方案中,测量策略117管理对测试区域115和116的访问的测量如何影响对非测试区域118的替换策略的选择。例如,在一些实施方案中,测量策略117由高速缓存控制器设置为多种状态中的任何一种,其中不同状态被分类为未偏斜状态和一个或多个偏斜校正状态。当测量策略117处于未偏斜状态时,高速缓存控制器105基于测试区域115和116的标称的、未修改的性能特征来选择非测试区域的高速缓存策略。当测量策略117处于一个偏斜校正状态时,

为了选择非测试区域118的替换策略,修改策略选择过程中的一个或多个,或者性能特征本身的值。特别地,修改策略选择过程或性能特征值以校正对不同测试区域115和116的访问次数的相对较大差异。

[0023] 例如,在一些实施方案中,对于每个测试区域,性能监视器104记录高速缓存命中率中的加权表示,表示为WCHR,其中WCHR通过以下公式计算:

$$[0024] \quad WCHR = W_{TR} * CHR$$

[0025] 其中 W_{TR} 是与测试区域相关联的权重,并且CHR是测试区域的标称(未加权)高速缓存命中率。当测量策略117处于未偏斜状态时,性能监视器104将测试区域115和116中的每一者的性能特征的权重设置为第一指定值(例如,每个权重的值为1)。当测量策略117处于一个偏斜校正状态时,性能监视器104将与测试区域115和116中的至少一者相关联的权重设置为第二指定值,从而对与测试区域中的一者相关联的访问测量进行偏斜校正,如本文进一步所述。例如,在一些实施方案中,当测量策略117处于非偏斜状态时,与测试区域115和116中的每一者相关联的权重相等,并且当测量策略117处于一个偏斜校正状态时,与测试区域115相关联的权重不同于与测试区域116相关联的权重。

[0026] 在一些实施方案中,测量策略117的状态不直接决定性能特征的测量,而是决定高速缓存控制器105是否被准许改变非测试区域118的替换策略。因此,当测量策略117处于未偏斜状态时,高速缓存控制器105准许基于测试区域115和116的性能特征改变非测试区域118的替换策略。当测量策略117处于偏斜校正状态时,高速缓存控制器105暂停对非测试区域118的替换策略的改变,直到测量策略117返回到未偏斜状态。

[0027] 如上所述,高速缓存所经历的一些存储器访问模式是非代表性模式,它们不代表与处理器101正在执行的给定应用程序或其他指令集相关联的存储器访问模式。在某些情况下,这种非代表性的存储器访问模式会导致性能监视器104记录的性能特征出现异常,这些异常又会导致高速缓存控制器为非测试区域118选择高速缓存替换策略111和112中效率较低的一个。为了减轻这些非代表性存储器访问模式的影响,访问偏斜校正模块108在滑动时间窗口内监测对测试区域115的存储器访问次数与对测试区域116的存储器访问次数之间的差值。响应于所述差值超过指定阈值,访问偏斜校正模块108将测量策略117从未偏斜状态改变为偏斜校正状态,以校正每个测试区域的访问测量的影响。例如,在一些实施方案中,响应于存储器访问次数的差值超过阈值,访问偏斜校正模块108将测量策略117改变为偏斜校正状态,从而暂停改变非测试区域118的替换策略。响应于在滑动时间窗口内测试区域115和116的访问测量的差值下降到阈值以下,访问偏斜校正模块108将测量策略117返回到未偏斜状态,从而恢复非测试区域118的改变(当这种变化由上面讨论的相对性能特征指示时)。

[0028] 在其他实施方案中,响应于存储器访问次数的差值超过阈值,访问偏斜校正模块108将测量策略117改变为偏斜校正状态,从而改变测试区域115和116的权重(W_{TR} 值)以降低访问次数较多的测试区域的相对权重。例如,在一些实施方案中,响应于在滑动时间窗口内对测试区域115的访问次数与对测试区域116的访问次数的差值大于阈值,访问偏斜校正模块108将测量策略117改变为减少与测试区域115相关联的权重(或增加与测试区域116相关联的权重,或其组合)的状态,从而降低非代表性存储器访问模式将导致非测试区域118的替换策略改变的可能性。响应于在滑动时间窗口内测试区域115和116的访问测量的差值下

降到阈值以下,访问偏斜校正模块108将测量策略117设置为其未偏斜状态,从而使测试区域的权重返回到其标称值。

[0029] 在一些实施方案中,为将要进行偏斜校正的测试区域选择的权重是基于对不同测试区域115和116的访问之间的差值的大小。因此,例如,在一些实施方案中,测试区域115和116之间的访问差值的大小由以下公式给出:

$$[0030] \quad \text{Access}_{\text{差值}} = |\text{Accesses}_{115} - \text{Accesses}_{116}|$$

[0031] 其中 Accesses_{115} 是在滑动时间窗口内对测试区域115的访问次数,并且 Accesses_{116} 是在滑动时间窗口内对测试区域116的访问次数。在一些实施方案中,应用于每个测试区域的性能特征的权重是基于 $\text{Access}_{\text{差值}}$ 的值。

[0032] 图2示出根据一些实施方案的偏斜校正模块108的框图。偏斜校正模块108包括移位寄存器222和比较模块225。移位寄存器222包括多个字段(例如,字段226)。对于对测试区域的每次访问,移位寄存器222 1)将每个字段的值向左移位一个位置,在移位之前最左边的字段被丢弃;和2)响应于对测试区域115的存储器访问,在最右边的位位置存储值1,并且响应于对测试区域116的存储器访问,在最右边的位位置存储值0。因此,存储在移位寄存器222处的1和0的数量分别表示在滑动时间窗口内测试区域115和116处的存储器访问次数。

[0033] 比较模块225确定存储在移位寄存器222处的1和0的数量之间的差值,并确定差值的大小是否超过阈值228。在一些实施方案中,阈值228是在处理器101或处理系统100的表征或测试过程期间设置。在其他实施方案中,阈值228是基于由正在执行的程序或由编译程序的编译器指示的值来设置。从而根据在处理器101处执行的特定指令集来设置阈值228。

[0034] 比较模块225基于差值大小与阈值228的比较来设置输出信号(表示为DESKEW)的状态。因此,例如,响应于差值的大小等于或低于阈值228,比较模块225将DESKEW信号设置为对应于未偏斜访问测量的第一状态,并且响应于大小高于阈值228,比较模块225将DESKEW信号设置为对应于偏斜访问测量的第二状态。基于DESKEW信号的状态,高速缓存控制器105设置测量策略117。因此,例如,响应于DESKEW信号指示未偏斜访问测量,高速缓存控制器105将测量策略117设置为其未偏斜状态。响应于DESKEW信号指示未偏斜访问测量,高速缓存控制器105将测量策略117设置为偏斜校正状态,以便如上所述校正性能特征的测量,诸如通过暂停改变高速缓存替换策略或通过改变与测试区域115和116中的每一者相关的性能特征权重。

[0035] 在一些实施方案中,偏斜校正模块108监测不同类型的访问,并且响应于对测试区域115和116的给定类型的访问数量的差值超过对应阈值来指示测量策略117的变化。图3示出根据一些实施方案的示例。特别地,在图3的示例中,偏差校正模块108包括两个移位寄存器322和324以及比较模块325。移位寄存器322和324中的每一者追踪测试区域115和116处的不同类型的存储器访问。

[0036] 为了说明,在一些实施方案中,移位寄存器322跟踪对测试区域115和116的需求访问。对于对测试区域的每次需求访问,移位寄存器322 1)将每个字段的值向左移位一个位置,在移位之前最左边的字段被丢弃;和2)响应于对测试区域115的需求访问,在最右边的位位置存储值1,并且响应于对测试区域116的存储器访问,在最右边的位位置(字段326)存储值0。因此,存储在移位寄存器322处的1和0的数量分别表示在滑动时间窗口内测试区域115和116处的需求访问次数。移位寄存器324以类似方式跟踪对测试区域115和116的预取

访问。

[0037] 比较模块325基于两种类型的存储器访问的差值大小中的每一者与阈值328的比较来设置DESKEW信号的状态。因此,例如,响应于对测试区域115和116的需求访问的差值大小等于或低于阈值328,比较模块325将DESKEW信号设置为对应于未偏斜访问测量的第一状态,并且响应于大小高于阈值328,比较模块325将DESKEW信号设置为对应于偏斜访问测量的第二状态。类似地,响应于对测试区域115和116的预取访问的差值大小等于或低于阈值328,比较模块325将DESKEW信号设置为对应于未偏斜访问测量的第一状态,并且响应于大小高于阈值328,比较模块325将DESKEW信号设置为对应于偏斜访问测量的第二状态。

[0038] 应当理解,在一些实施方案中,偏斜校正模块108存储用于需求访问和预取访问的具有不同值的不同阈值。此外,在一些实施方案中,除了移位寄存器322和324之外或代替移位寄存器322和324,偏斜校正模块108还包括移位寄存器以跟踪其他类型的存储器访问,例如对不同高速缓存层级的存储器访问。此外,在一些实施方案中,高速缓存控制器105将测量策略117改变为不同的状态,这取决于哪些类型的存储器访问导致了DESKEW信号的断言。

[0039] 图4示出根据一些实施方案的绘示改变处理器101的测量策略117的示例的图400。图400包括代表时间的x轴和代表测试区域115和116中的每一者的存储器访问次数的差值大小的y轴。图400处示出的图401因此代表测试区域115和116中的每一者的存储器访问次数的差值大小。

[0040] 在所描绘的示例中,在时间430之前,存储器访问次数的差值大小低于阈值228。因此,在时间430之前,高速缓存控制器105将测量策略117维持在偏斜校正状态,其中与测试区域115和116相关联的性能特征没有被偏斜校正。在时间430,存储器访问次数的差值大小达到阈值228并随后超过阈值。作为响应,高速缓存控制器105将测量策略117改变为偏斜校正状态以对性能特征进行偏斜校正,诸如通过改变应用于性能特征的一个或多个权重、通过暂停改变高速缓存110的非测试区域118的高速缓存替换策略等。在时间432时(在时间430之后),存储器访问次数的差值大小降到阈值228以下。作为响应,高速缓存控制器105将测量策略117返回到标称状态。

[0041] 图5示出根据一些实施方案的在处理器处设置存储器访问测量策略的方法500的流程图。关于处理器101处的示例实现方式来描述方法500。在框502,访问偏斜校正模块108记录在滑动时间窗口内对测试区域115和116中的每一者的访问次数。在框504,访问偏斜校正模块108确定对测试区域115的访问次数与对测试区域116的访问次数之间的差值大小并且将所确定的大小与阈值228进行比较。响应于所述差值大小等于或小于阈值228,所述方法流程移动到框506,并且访问偏斜校正模块108将测量策略117设置为其未偏斜状态,其中例如,基于测试区域115和116的等权重性能特征准许改变非测试区域118的替换策略。方法流程返回到框502。

[0042] 返回到框504,响应于对测试区域115的访问次数与对测试区域116的访问次数之间的差值大小大于阈值228,方法流程移动到框408,并且访问偏斜校正模块108将测量策略117设置为其偏斜校正状态。在一些实施方案中,当测量策略117处于偏斜校正状态时,高速缓存控制器105基于测试区域115和116的性能特征而不会改变非测试区域118的替换策略。在其他实施方案中,当测量策略117处于偏斜校正状态时,高速缓存控制器105对测试区域115和116的性能特征应用不同的权重并且基于加权的性能特征来选择非测试区域118的替

换策略。

[0043] 如上所述,在一些实施方案中,替代用于设置处理器的访问测量策略或除了用于设置处理器的访问测量策略之外,对测试区域115和116的访问的差值大小还用于其他目的。图6示出一个示例,该图绘示了根据一些实施方案的处理器600,所述处理器基于对高速缓存测试区域的访问的差值大小来设置地址散列。处理器600包括处理器内核602、高速缓存控制器605、性能监视器604和高速缓存610。这些模块中的每一者与上述的图1的对应模块类似地操作。特别地,高速缓存610包括两个测试区域,分别表示为测试区域615和616。

[0044] 此外,高速缓存控制器605使用高速缓存地址散列619来访问高速缓存610。响应于接收到对高速缓存610的存储器访问,高速缓存控制器605将高速缓存地址散列619应用于存储器访问的存储器地址以生成散列地址值。高速缓存控制器605使用散列地址值来访问高速缓存610。也就是说,高速缓存控制器605访问由散列地址值指示的高速缓存610的位置,而不是由存储器访问提供的原始的未散列地址值。

[0045] 在一些实施方案中,高速缓存地址散列619被设置为标称值,所述标称值预期会促进使用相对较大百分数的高速缓存610的高速缓存位置。也就是说,高速缓存地址散列619的标称值预期会通过“分散”对相对较大百分数的高速缓存位置的访问来促进对高速缓存610的有效使用,即便与所述访问相关联的非散列地址落入相对较窄的存储器地址范围内。

[0046] 然而,对于某些存储器访问模式,散列的标称值不会充分地改变散列地址值,使得所得的存储器访问倾向于访问相对较小的高速缓存位置集。这些存储器访问模式通过对测试区域615和616中的一者比对另一个测试区域的访问更频繁来指示。因此,处理器101包括散列控制模块608,所述散列控制模块以与图2的访问偏斜校正模块108类似的方式对在滑动时间窗口内对测试区域615的存储器访问次数与对测试区域616的存储器访问次数之间的差值进行比较。响应于差值的大小超过阈值,散列控制模块608将高速缓存访问散列619改变为不同的值。在一些实施方案中,散列控制模块608继续改变高速缓存地址散列619(例如,通过连续地改变散列值的不同位的值),直到差值大小低于阈值。散列控制模块608由此促进高速缓存110对各种不同存储器访问模式的有效使用。

[0047] 在一些实施方案中,除了改变高速缓存地址散列619的值之外或替代改变高速缓存地址散列619的值,高速缓存控制器605改变与测试区域615和616中的一者或多者相关联的存储器地址,由此使测试区域转移到高速缓存610的不同区域。因此,例如,响应于对测试区域615和616的访问的差值大小超过阈值,高速缓存控制器改变与测试区域615相关联的地址,由此使测试区域615转移到高速缓存610的不同部分。

[0048] 如本文所述,在一些实施方案中,一种方法包括:监测对高速缓存的第一区域和所述高速缓存的第二区域中的每一者的访问;基于对所述高速缓存的所述第一区域的受监测访问的第一数量和对所述高速缓存的所述第二区域的受监测访问的第二数量来选择测量策略;以及基于对所述高速缓存的所述第一区域的受监测访问的所述第一数量和对所述高速缓存的所述第二区域的受监测访问的所述第二数量并且基于所述测量策略来设置用于所述高速缓存的第三区域的高速缓存替换策略。在一个方面中,选择所述测量策略包括:确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的第一差值;响应于所述差值小于阈值而选择第一测量策略;以及响应于所述差值大于或等于所述阈值而选择第二测量策略。在另一方面中,所述第一测量策略包括允许改变所述高速缓存替换策略;并且所

述第二测量策略包括阻止改变所述高速缓存替换策略。在又一方面中,所述第一测量策略包括将第一权重应用于对所述第一区域的受监测访问以及将第二权重应用于对所述第二区域的所述受监测访问;所述第二测量策略包括将第三权重应用于对所述第一区域的受监测访问以及将第四权重应用于对所述第二区域的所述受监测访问;并且其中选择所述高速缓存策略包括基于对所述第一区域的加权的受监测访问和对所述第二区域的加权的受监测访问来选择所述高速缓存策略。在一个方面中,所述第一权重与所述第二权重是相同的权重;并且所述第三权重与所述第四权重不同。

[0049] 在一个方面中,选择所述测量策略包括:确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的第一差值;基于所述第一差值选择用于所述第一区域的第一权重和用于所述第二区域的第二权重;基于所述第一权重对所述第一区域的访问进行加权以及基于所述第二权重对所述第二高速缓存区域的访问进行加权;并且其中选择所述高速缓存策略包括基于对所述第一区域的加权的受监测访问和对所述第二区域的加权的受监测访问来选择所述高速缓存策略。在另一个方面中,监测对所述第一区域和所述第二区域中的每一者的访问包括:监测对所述第一区域和所述第二区域中的每一者的第一类型的访问以及监测对所述第一区域和所述第二区域中的每一者的第二类型的访问;并且选择所述测量策略包括基于以下各项选择所述测量策略:对所述第一区域的所述第一类型的访问次数与对所述第二区域的所述第一类型的访问次数之间的第一差值;以及对所述第一区域的所述第二类型的访问次数与对所述第二区域的所述第二类型的访问次数之间的第二差值。在又一个方面中,所述第一类型包括需求访问,并且所述第二类型包括预取访问。

[0050] 在一个方面中,设置所述高速缓存策略包括基于所述高速缓存的所述第一区域处的高速缓存命中率和所述高速缓存的所述第二区域处的高速缓存命中率来设置所述高速缓存策略。在另一个方面中,设置所述高速缓存策略包括基于所述高速缓存的所述第一区域处的高速缓存未命中率和所述高速缓存的所述第二区域处的高速缓存未命中率来设置所述高速缓存策略。

[0051] 在一些实施方案中,一种方法包括:监测对高速缓存的第一区域和所述高速缓存的第二区域中的每一者的访问;基于对所述高速缓存的所述第一区域的受监测访问的第一数量和对所述高速缓存的所述第二区域的受监测访问的第二数量来选择散列值;以及基于所述所选的散列值生成所述高速缓存的第三区域的高速缓存地址值;以及基于所述高速缓存地址值将数据存储于所述高速缓存的所述第三区域处。在一个方面中,选择所述散列值包括:确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的第一差值;响应于所述差值小于阈值而选择第一散列值;以及响应于所述差值大于或等于所述阈值而选择散列值。在另一个方面中,所述方法包括:基于对所述高速缓存的所述第一区域的受监测访问的所述第一数量和对所述高速缓存的所述第二区域的受监测访问的所述第二数量来改变与所述高速缓存的所述第一区域相关联的地址。

[0052] 在一些实施方案中,一种处理器包括:高速缓存,所述高速缓存包括第一区域、第二区域和第三区域;以及高速缓存控制器,所述高速缓存控制被配置为:监测对所述第一区域和所述第二区域中的每一者的访问;基于对所述第一区域的受监测访问的第一数量和对所述第二区域的受监测访问的第二数量来选择测量策略;并且基于对所述第一区域的受监测访问的所述第一数量和对所述第二区域的受监测访问的所述第二数量并且基于所述测

量策略来选择用于所述第三区域的高速缓存替换策略。在一个方面中,所述高速缓存控制器被配置为通过以下步骤来选择所述测量策略:确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的第一差值;响应于所述差值小于阈值而选择第一测量策略;以及响应于所述差值大于或等于所述阈值而选择第二测量策略。

[0053] 在一个方面中,所述第一测量策略包括允许改变所述高速缓存替换策略;并且所述第二测量策略包括阻止改变所述高速缓存替换策略。在另一个方面中,所述第一测量策略包括将第一权重应用于对所述第一区域的受监测访问以及将第二权重应用于对所述第二区域的所述受监测访问;所述第二测量策略包括将第三权重应用于对所述第一区域的受监测访问以及将第四权重应用于对所述第二区域的所述受监测访问;并且其中选择所述高速缓存策略包括基于对所述第一区域的加权的受监测访问和对所述第二区域的加权的受监测访问来选择所述高速缓存策略。在另一个方面中,所述第一权重与所述第二权重是相同的权重;并且所述第三权重与所述第四权重不同。

[0054] 在一个方面中,所述高速缓存控制器被配置为通过以下步骤来选择所述测量策略:确定受监测访问的所述第一数量与受监测访问的所述第二数量之间的第一差值;基于所述第一差值选择用于所述第一高速缓存区域和所述第二高速缓存区域的第一权重和第二权重;基于所述第一权重对所述第一区域的访问进行加权以及基于所述第二权重对所述第二高速缓存区域的访问进行加权;并且其中选择所述高速缓存策略包括基于对所述第一区域的加权的受监测访问和对所述第二区域的加权的受监测访问来选择所述高速缓存策略。在另一个方面中,所述高速缓存策略设置所述高速缓存的所述第三区域的预取策略。

[0055] 在一些实施方案中,上文所描述的技术的某些方面可由执行软件的处理系统的一个或多个处理器实现。软件包括存储或以其他方式有形地体现在非暂时性计算机可读存储介质上的一个或多个可执行指令集。软件可包括指令和某些数据,当由一个或多个处理器执行时,所述指令和某些数据操纵所述一个或多个处理器来执行上述技术的一个或多个方面。非暂时性计算机可读存储介质可以包括例如磁盘或光盘存储装置、例如闪存存储器的固态存储装置、高速缓存、随机存取存储器(RAM)或其他一个或多个非易失性存储器装置等。存储在非暂时性计算机可读存储介质上的可执行指令可呈现源代码、汇编语言代码、对象代码或由一个或多个处理器解译或以其他方式执行的其他指令格式。

[0056] 应注意,并不需要上文在一般描述中所描述的所有活动或元件,特定活动或装置的一部分可能是不需要的,并且可以执行一个或多个其他活动,或者可以包括除所描述的那些元件之外的元件。此外,列出活动的顺序不一定是执行活动的顺序。另外,参考具体实施方案描述了概念。然而,本领域普通技术人员应了解,在不脱离如所附权利要求书中所阐述的本公开范围的情况下,可作出各种修改和改变。因此,本说明书和附图将被视为说明性而非限制性的,并且所有此类修改意欲包括在本公开的范围之内。

[0057] 上文已经参照具体实施方案描述了益处、其他优点以及问题的解决方案。然而,所述益处、优点、问题解决方案以及可使任何益处、优点或问题解决方案出现或变得更突出的任何特征都不应被解释为是任何或所有权利要求的关键、必需或必要特征。此外,以上所公开的特定实施方案只是说明性的,因为所公开的主题可以按受益于本文教义的本领域的技术人员显而易见的不同但等效的方式来修改和实践。除了如所附权利要求中所描述的之

外,并不意欲限于本文所示的构造或设计的细节。因此,明显的是,上文所公开的特定实施方案可被更改或修改,并且所有此类变化都被认为是在所公开的主题的范围内。因此,本文所寻求的保护如所附权利要求中所阐述。

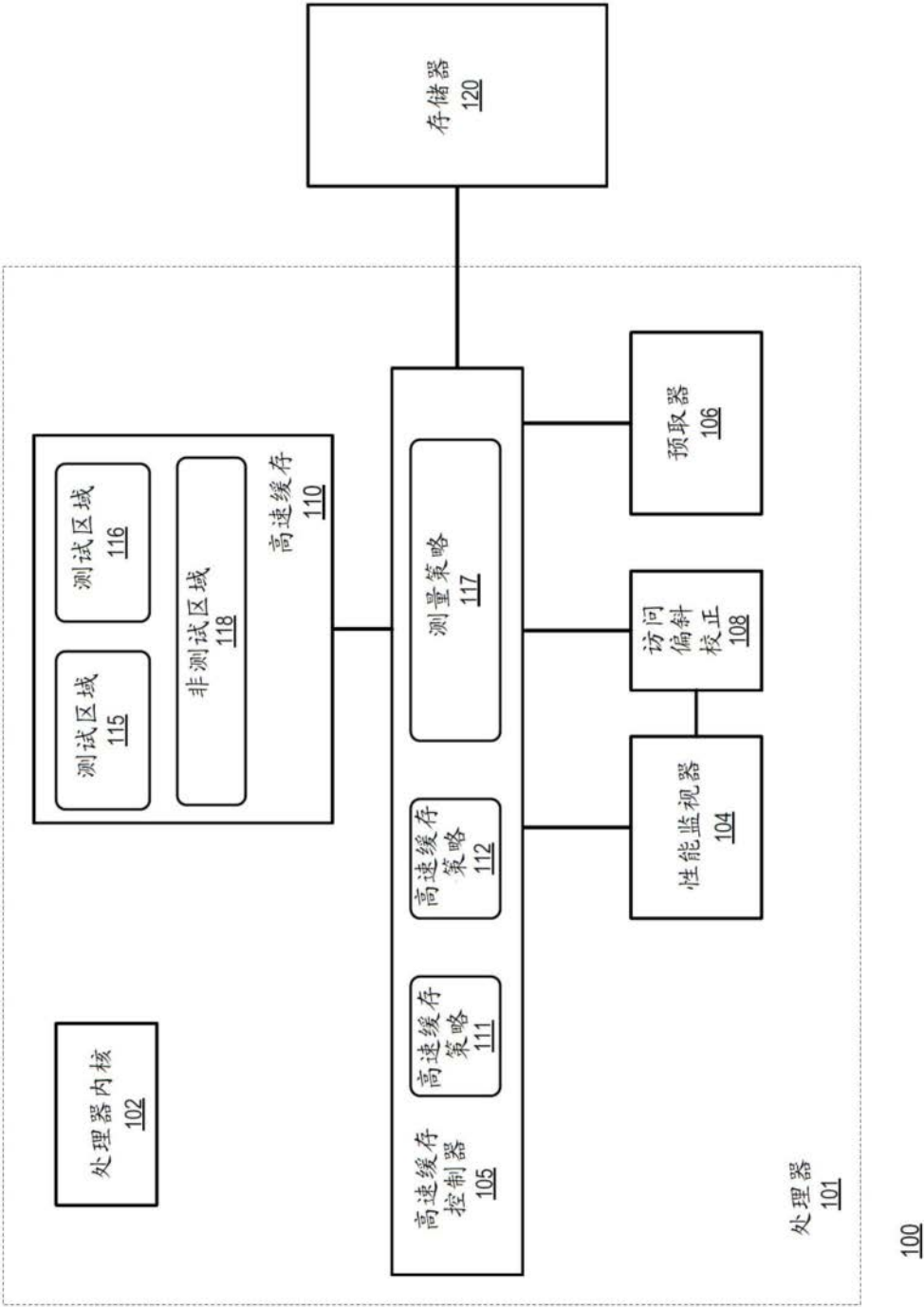


图1

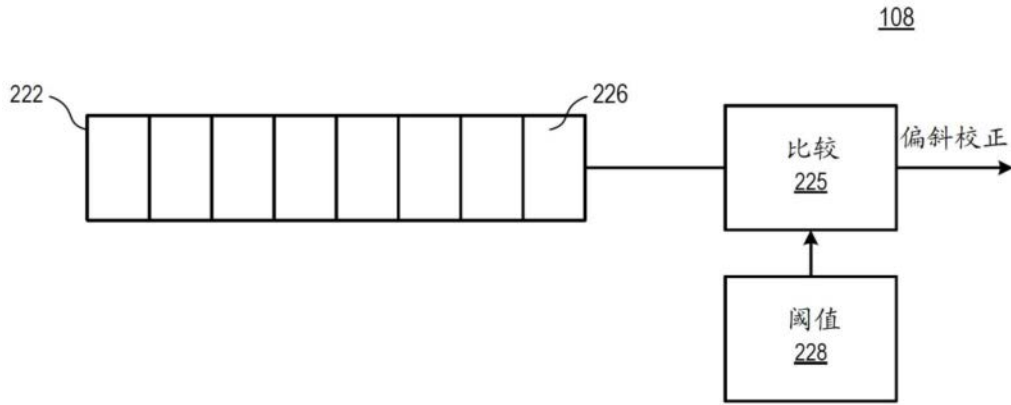


图2

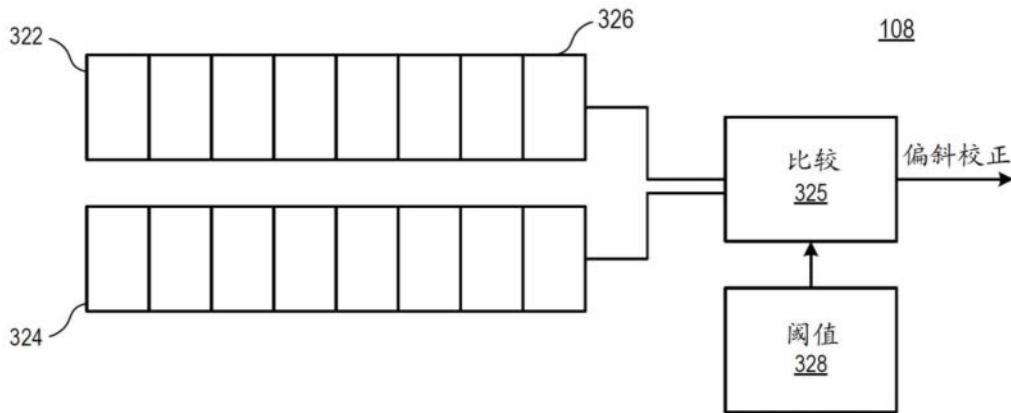


图3

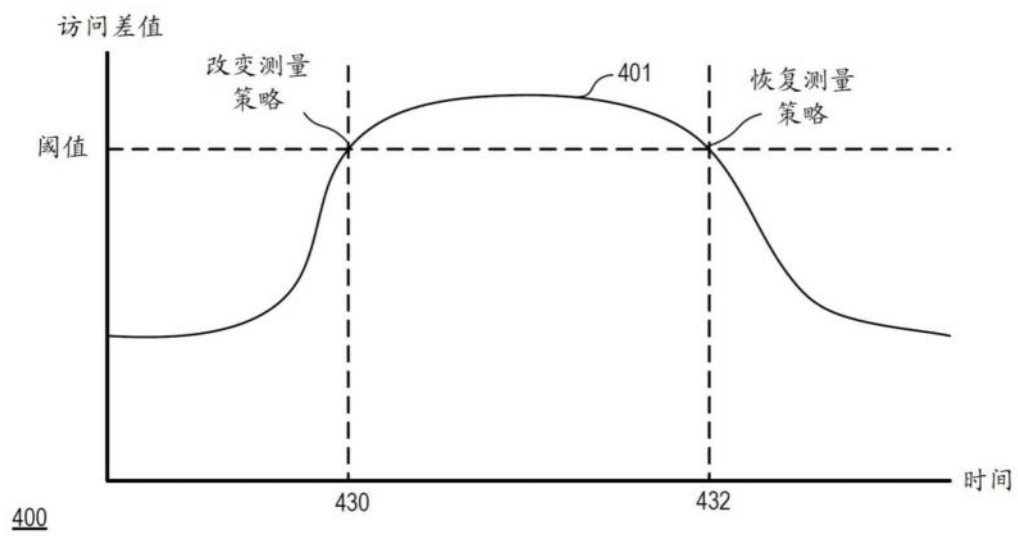


图4

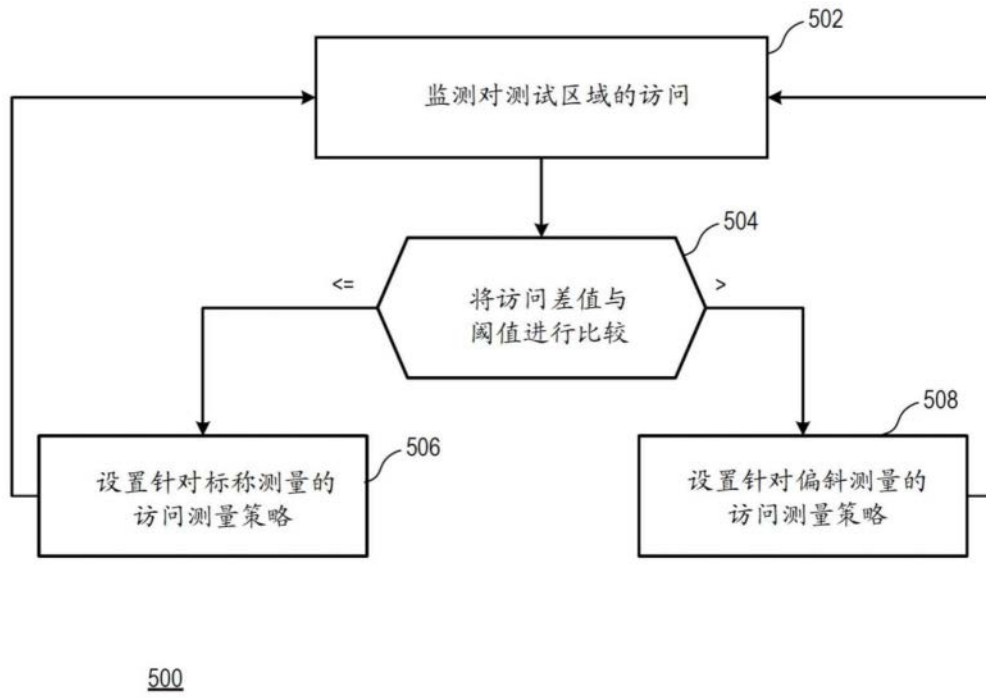
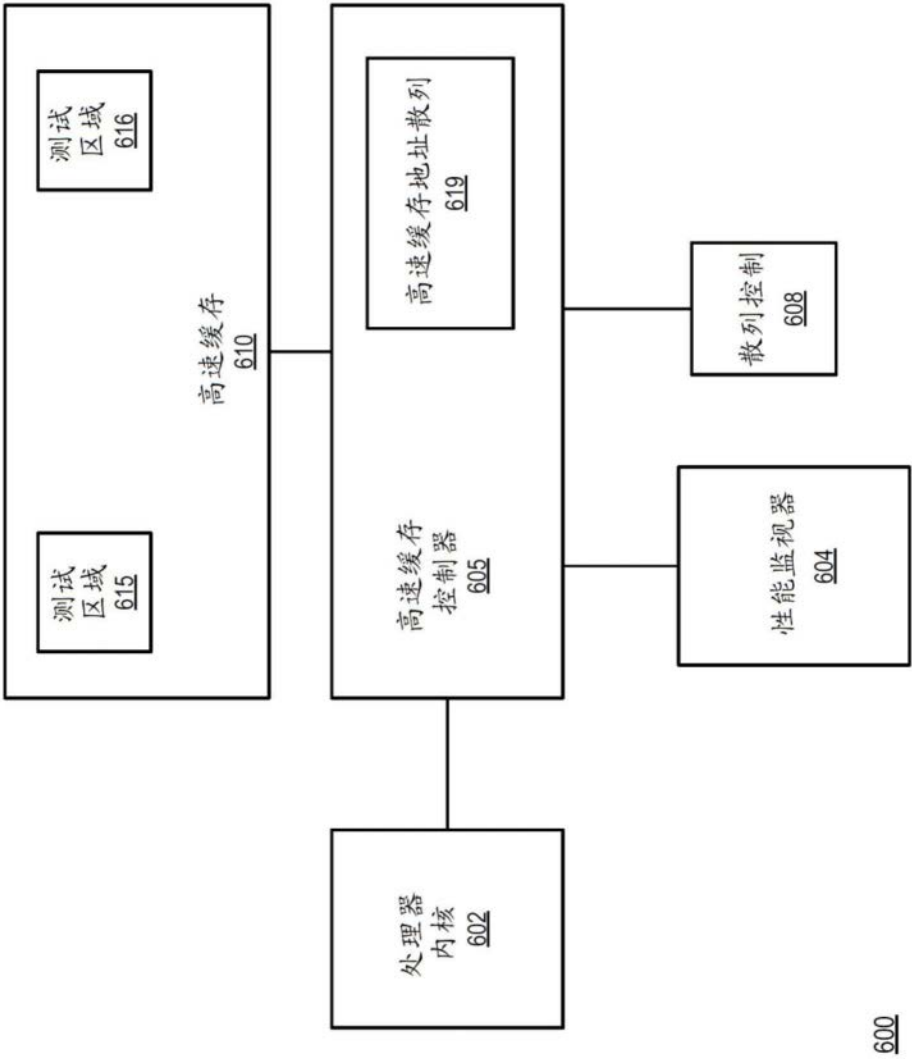


图5



600

图6