[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G06F 12/08

[12] 发明专利说明书

[21] ZL 专利号 99101099.X

[45] 授权公告日 2003 年 12 月 17 日

[11] 授权公告号 CN 1131481C

[22] 申请日 1999.1.15 [21] 申请号 99101099.X

[30] 优先权

[32] 1998.2.17 [33] US [31] 024612

[71] 专利权人 国际商业机器公司

地址 美国纽约州

[72] 发明人 R·K·阿里米里 J·S·多森
J·D·刘易斯

审查员 邹 斌

[74] 专利代理机构 中国专利代理(香港)有限公司

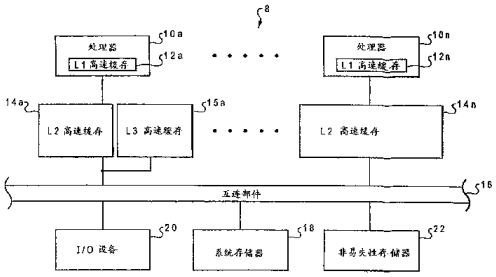
代理人 王 勇 张志醒

权利要求书 6 页 说明书 12 页 附图 3 页

[54] 发明名称 包含一具有精确模式和非精确模式的悬停状态的高速缓存相关协议

[57] 摘要

第一数据项与指示该数据项的地址的地址标记相关联地存储在第一高速缓存中，在第一高速缓存中的一相关指示符被设置至指示所述第一数据项有效的第一状态，作为响应，该相关指示符被修改至指示该地址标记有效而第一数据项无效的第二状态。 此后，根据该第一高速缓存的操作模式，判定是否修改该第一高速缓存。 响应对该第一高速缓存进行修改的判定，存储该第二数据项并且修改该相关指示符至指示该第二数据项有效的第三状态。



ISSN1008-4274

1.一种在含有多个处理器的数据处理系统中保持高速缓存相关的方法，所述多个处理器耦合至一互连部件并且每个处理器与多个高速缓存中相应的一个相关联，所述方法的特征在于包括:

在所述多个高速缓存的第一高速缓存中，与一地址标记相关联地存储第一数据项，该地址标记指示所述第一数据项的地址;

设置所述第一高速缓存中的一相关指示符至第一状态，该第一状态指示所述第一数据项有效;

在所述相关指示符被设置为所述第一状态的时候，响应所述多个高速缓存中的另一个指示想要存储由所述地址标记指示的所述地址，修改所述第一高速缓存中的所述相关指示符至第二状态，该第二状态指示所述地址标记有效而在所述第一高速缓存中的所述数据项无效;

在所述相关指示符被设置为所述第二状态的时候，响应在所述互连部件上的与由所述地址标记指示的所述地址相关联的一数据传送的检测，所述数据传送由所述多个高速缓存中的另一个启动并且包括第二数据项，根据所述第一高速缓存的操作模式，判定是否修改所述第一高速缓存; 以及

响应对所述第一高速缓存进行修改的判定，通过在所述第一高速缓存中与所述地址标记相关联地存储所述第二数据项并且通过修改所述相关指示符至第三状态来替换所述第一数据项，所述第三状态指示所述第二数据项有效。

2.如权利要求 1 的方法，其特征在于所述的根据所述第一高速缓存的操作模式判定是否修改所述第一高速缓存的步骤包括步骤:

如果所述第一高速缓存操作在第一模式，则总是决定修改所述第一高速缓存; 以及

如果所述第一高速缓存操作在第二模式，则仅根据在所述第一高速缓存中不存在选择的条件来决定修改所述第一高速缓存。

3.如权利要求 2 的方法，其特征在于所述第一高速缓存包括一从其中进行对所述第一高速缓存的修改的队列，其中所述仅根据在所述第一高速缓存中不存在选择的条件来决定修改所述第一高速缓存的步骤包括只有当所述队列含有少于一阈值数的项时才决定修改所述第一高速缓存的步骤。

4.如权利要求 1 的方法，其特征在于还包括根据由所述多个处理器中的第一处理器执行的一指令来设置所述第一高速缓存的所述操作模式至第一模式和第二模式的其中之一的步骤。

5.如权利要求 1 的方法，其特征在于所述数据处理系统包括监视一个或多个选择的事件的性能监视器硬件，并且所述的方法还包括利用所述性能监视器硬件设置所述第一高速缓存的所述操作模式的步骤。

6.如权利要求 1 的方法，其特征在于所述修改所述相关指示符至指示所述第二数据项有效的第三状态的步骤包括修改所述相关指示符至一共享状态的步骤，该共享状态指示所述第二数据项存储在所述第一高速缓存和所述多个高速缓存的另一个中。

7.如权利要求 1 的方法，其特征在于，所述设置所述第一高速缓存中的一相关指示符至指示所述数据项有效的第一状态的步骤包括设置所述第一高速缓存中的所述相关指示符至修改状态、共享状态和专用状态的其中之一的步骤。

8.如权利要求 1 的方法，其特征在于所述数据处理系统还包括一个低级存储器，所述第一高速缓存可从该低级存储器检索数据，所述多个处理器包括与所述第一高速缓存相关的第一处理器，所述数据项包括一个第一数据数据项，所述方法还包括步骤:

在所述相关指示符设置为所述第二状态时，响应所述第一处理器对与由所述地址标志指示的所述地址相关的数据的请求，从所述多个高速缓存的另一个中而不是从所述低级存储器中获得与所述地址有关的一个有效的第二数据项。

9.如权利要求 1 的方法，其特征在于还包括设置所述相关指示符为一个无效状态以表示所述地址标志和所述数据项均无效的步骤。

10.一种用于在含有多个处理器的数据处理系统中支持高速缓存相关的高速缓存，所述多个处理器中的每一个与多个高速缓存中相应的一个相关联，所述高速缓存的特征在于包括:

存储数据项的数据存储器;

标记存储器，它存储指示包含在所述数据存储器中的所述数据项的地址的地址标记;

相关指示符，响应所述数据项存入所述数据存储器，该相关指示

符被设置至表示所述数据项有效的第一状态，而根据在所述相关指示符被设置为所述第一状态的时候，该多个高速缓存中的另一个指示想要存储由所述地址标记指示的所述地址，该相关指示符被设置至第二状态，所述相关指示符的所述第二状态指示所述地址标记有效而在所述数据存储器中的所述数据项无效；

一装置，在所述相关指示符被设置为所述第二状态的时候，该装置响应在所述互连部件上的与由所述地址标记指示的所述地址相关联的一数据传送的检测，所述数据传送由所述多个高速缓存中的另一个启动并且包括第二数据项，用于根据所述第一高速缓存的操作模式，判定是否修改所述第一高速缓存；以及

一装置，该装置响应对所述第一高速缓存进行修改的判定，用于通过在所述第一高速缓存中与所述地址标记相关联地存储所述第二数据项替换所述第一数据项，并且修改所述相关指示符至第三状态，所述第三状态指示所述第二数据项有效。

11.如权利要求 10 的高速缓存，其特征在于所述的用于根据所述第一高速缓存的操作模式判定是否修改所述第一高速缓存的装置包括：

一装置，该装置响应所述第一高速缓存操作在第一模式，用于总是决定修改所述第一高速缓存；以及

一装置，该装置响应所述第一高速缓存操作在第二模式，用于仅根据在所述第一高速缓存中不存在选择的条件来决定修改所述第一高速缓存。

12.如权利要求 11 的高速缓存，其特征在于所述第一高速缓存包括一从其中进行对所述第一高速缓存的修改的队列，其中所述的用于仅根据在所述第一高速缓存中不存在选择的条件来决定修改所述第一高速缓存的装置包括用于只有当所述队列含有少于一阈值数的项时才决定修改所述第一高速缓存的装置。

13.如权利要求 10 的高速缓存，其特征在于还包括用于根据由所述多个处理器中的第一处理器执行的一指令来设置所述第一高速缓存的所述操作模式至第一模式和第二模式的其中之一的装置。

14.如权利要求 10 的高速缓存，其特征在于还包括性能监视器硬件，该性能监视器硬件监视一个或多个选择的事件并且根据所述一个

或多个选择的事件中至少一个事件的发生来设置所述第一高速缓存的所述操作模式。

15. 如权利要求 10 的高速缓存，其特征在于所述第三状态是一共享状态，该共享状态指示所述第二数据项存储在所述第一高速缓存和所述多个高速缓存的另一个中。

16. 如权利要求 10 的高速缓存，其特征在于所述第一状态包括修改状态、共享状态和专用状态的其中之一。

17. 如权利要求 10 的高速缓存，其特征在于所述高速缓存是一个第一高速缓存，所述多个处理器包括与所述第一高速缓存相关的第一处理器，所述数据处理系统还包括一个低级存储器，所述第一高速缓存可从该低级存储器检索数据，所述高速缓存还包括：

在所述相关指示符设置为所述第二状态时，响应所述第一处理器对与由所述地址标志指示的所述地址相关的数据的请求，从所述多个高速缓存的另一个中而不是从所述低级存储器中获得与所述地址有关的有效数据的装置。

18. 如权利要求 10 的高速缓存，其特征在于所述相关指示符还包括一个无效状态以表示所述地址标志和所述数据项均无效。

19. 一种数据处理系统，包括：

一互连部件；

耦合至所述互连部件的多个处理器；

多个高速缓存，其中每个高速缓存与所述多个处理器中相应的一个相关联，其中，在所述多个高速缓存中的第一高速缓存包括：

存储数据项的数据存储器；

标记存储器，它存储指示包含在所述数据存储器中的所述数据项的地址的地址标记；

相关指示符，响应所述数据项存入所述数据存储器，该相关指示符被设置至表示所述数据项有效的第一状态，而根据在所述相关指示符被设置为所述第一状态的时候，该多个高速缓存中的另一个指示想要存储由所述地址标记指示的所述地址，该相关指示符被设置至第二状态，所述相关指示符的所述第二状态指示所述地址标记有效而在所述数据存储器中的所述数据项无效；

一装置，在所述相关指示符被设置为所述第二状态的时候，该装

置响应在所述互连部件上的与由所述地址标记指示的所述地址相关联的一数据传送的检测，所述数据传送由所述多个高速缓存中的另一个启动并且包括第二数据项，用于根据所述第一高速缓存的操作模式，判定是否修改所述第一高速缓存；以及

一装置，该装置响应对所述第一高速缓存进行修改的判定，用于通过在所述第一高速缓存中与所述地址标记相关联地存储所述第二数据项替换所述第一数据项，并且修改所述相关指示符至第三状态，所述第三状态指示所述第二数据项有效。

20.如权利要求 19 的数据处理系统，其特征在于所述的用于根据所述第一高速缓存的操作模式判定是否修改所述第一高速缓存的装置包括:

一装置，该装置响应所述第一高速缓存操作在第一模式，用于总是决定修改所述第一高速缓存；以及

一装置，该装置响应所述第一高速缓存操作在第二模式，用于仅根据在所述第一高速缓存中不存在选择的条件来决定修改所述第一高速缓存。

21.如权利要求 20 的数据处理系统，其特征在于所述第一高速缓存包括一从其中进行对所述第一高速缓存的修改的队列，其中所述的用于仅根据在所述第一高速缓存中不存在选择的条件来决定修改所述第一高速缓存的装置包括用于只有当所述队列含有少于一阈值数的项时才决定修改所述第一高速缓存的装置。

22.如权利要求 19 的数据处理系统，其特征在于还包括用于根据由所述多个处理器中的第一处理器执行的一指令来设置所述第一高速缓存的所述操作模式至第一模式和第二模式的其中之一的装置。

23.如权利要求 19 的数据处理系统，其特征在于还包括性能监视器硬件，该性能监视器硬件监视一个或多个选择的事件并且根据所述一个或多个选择的事件中至少一个事件的发生来设置所述第一高速缓存的所述操作模式。

24.如权利要求 19 的数据处理系统，其特征在于所述第三状态是一共享状态，该共享状态指示所述第二数据项存储在所述第一高速缓存和所述多个高速缓存的另一个中。

25.如权利要求 19 的数据处理系统，其特征在于所述第一状态包

括修改状态、共享状态和专用状态的其中之一。

26.如权利要求 19 的数据处理系统，其特征在于所述多个处理器包括与所述第一高速缓存相关的第一处理器，所述数据处理系统还包括：

一个低级存储器，所述第一高速缓存可从该低级存储器检索数据；

在所述第一高速缓存的所述相关指示符设置为所述第二状态时，响应所述第一处理器对与由所述地址标志指示的所述地址相关的数据的请求，从所述多个高速缓存的另一个中而不是从所述低级存储器中获得与所述地址有关的有效数据的装置。

27.如权利要求 19 的数据处理系统，其特征在于所述相关指示符还包括一个无效状态以表示所述地址标志和所述数据项均无效。

包含一具有精确模式和非精确模式的悬停状态的高速缓存相关协议

本发明一般涉及一种用于数据处理的方法和系统，具体涉及一种在多处理器数据处理系统中保持高速缓存相关的方法和系统。而更具体地，本发明涉及一种用于多处理器数据处理系统的高速缓存相关协议，其中包括一悬停（H）状态，该状态允许以有效数据修改第一高速缓存，以响应第二高速缓存独立发送该有效数据至一耦合第一和第二高速缓存的互连部件上。

在常规的对称多处理器（SMP）数据处理系统中，所有的处理器通常是相同的，即，这些处理器都采用公共的指令集和通信协议，具有相似的硬件结构，并且一般配备有类似的存储器层次。例如，一常规 SMP 数据处理系统可以包括一系统存贮器、多个处理部件和一系统总线，其中每个处理部件包括一处理器和一级或多级高速缓冲存贮器，该系统总线将一处理部件耦合至每一其它处理部件和系统存贮器。为了在 SMP 数据处理系统中获得有效的执行结果，重要的是保持相关存贮器层次，即，为所有的处理部件提供存贮器内容的单一视图。

通过使用一选择的存贮器相关协议，例如 MESI 协议的应用来保持相关存贮器层次。在该 MESI 协议中，一相关状态的指示被与至少所有较高级的（高速缓冲）存贮器的每一相关区组（例如，高速缓存行或区段）相关联地存储，每个相关区组可以具有四个状态中的一个状态，即修改（M）、专用（E）、共享（S）或无效（I），它们在高速缓存目录中由两位表示。该修改状态表示一相关区组仅在存储该被修改相关区组的高速缓存中有效并且该被修改的相关区组的值还未写到系统存贮器。当一相关区组被指示为专用状态时，则在存贮器层次的这一级的所有高速缓存中，该相关区组仅驻留在具有该处于专用状态的相关区组的高速缓存中。但是在专用状态中的数据与系统存贮器一致。如果在高速缓存目录中一相关区组被标记为共享，则该相关区组驻留在其相关联的高速缓存中以及至少一个位于存贮器层次同一级的其它高速缓存中，该相关区组的所有这些拷贝与系统存贮器一致。最后，所述无效状态指示与一相关区组相关联的数据和地址标记都是无效的。

每个相关区组（例如高速缓存行）被设置的状态依赖于该高速缓存行的在先状态和请求处理器想要的存贮器访问的类型．因此，在多处理器数据处理系统中保持存贮器相关就需要这些处理器经由系统总线传送指示它们读或写存储单元的意图的消息．例如，当一处理器想要将数据写到一存储单元时，为了实现该写操作，该处理器必须首先将其把数据写到该存贮单元的意图通知所有其它处理部件并且从所有其它处理部件接收许可．由该请求处理器接收的这些许可消息表示该存贮单元内容的所有其它的高速缓存拷贝已经被无效，由此保证其它处理器将不访问失效的局部数据．这种消息交换就是公知的交叉－无效（cross-invalidation）（XI）．

本发明知道的是，虽然高速缓存项的交叉－无效足以维持 SMP 数据处理系统中的存贮器相关，但是，远程处理器的高速缓存项的无效因降低局部高速缓存中的命中率而对数据处理系统的性能产生不利影响．因此，即使配置大的局部高速缓存，但是当一处理部件从另一处理部件的远程高速缓存或者系统存贮器中检索曾经驻留在局部高速缓存中的数据时，前述欲检索数据的处理部件也要承担长的访问等待时间．因此，如将会明显地看到的那样，最好是提供一种方法和系统，用于在 SMP 数据处理系统中保持存贮器相关，该方法和系统减小了由于高速缓存项的交叉－无效的结果而招致的性能恶化．

因此，本发明的一个目的是提供一种改进的用于数据处理的方法和系统．

本发明的另一个目的是提供一种改进的用于在多处理器数据处理系统中保持高速缓存相关的方法和系统．

本发明的再一个目的是提供一种用于多处理器数据处理系统的高速缓存相关协议，该协议包括一悬停（H）状态，该状态允许第一高速缓存被以有效数据修改，以响应第二高速缓存独立发送该有效数据至耦合第一和第二高速缓存的互连部件上．

前面的目的的实现将如下所述．一数据处理系统包括多个处理器，其中每个处理器与多个高速缓存中相应的一个相关联．根据本发明的方法，第一数据项与表示该第一数据项的地址的一地址标记相关联地存储在第一高速缓存中．在第一高速缓存中的一相关指示符被设置为第一状态，该第一状态指示第一数据项有效．在该相关指示符被设置为第一状

态的时候，响应另一高速缓存指示想要存储由该地址标记指示的地址，在第一高速缓存中的相关指示符被修改至第二状态，该第二状态指示该地址标记有效而第一数据项无效。此后，响应检测到一远程发送的数据传送，该数据传送与由地址标记指示的地址相关联并且包括第二数据项，根据第一高速缓存的一种操作模式，产生一是否修改第一高速缓存的判定。根据对第一高速缓存进行修改的判定，第一数据项通过存储与该地址标记相关联的第二数据项而被替换，并且该相关状态指示符被修改至第三状态，该第三状态指示第二数据项有效。在一个实施例中，第一高速缓存的操作模式包括一精确模式和一非精确模式，在该精确模式中高速缓存的修改总是被执行，而在该非精确模式中高速缓存的修改被选择地执行。第一高速缓存操作于其中的模式可以由硬件或软件设置。

本发明的上述及其它的目的、特性和优点将在下面的详细描述中变得明白。

新的性能所认为的本发明的特征描述于后附的权利要求中。但是，本发明本身，所用的最佳实施例以及其它的目的和优点将最好是结合附图并参照下面的说明性实施例的详细描述来理解，其中：

图 1 描述了根据本发明的多处理器数据处理系统的一说明性实施例；

图 2 是描述根据本发明的高速缓存的说明性实施例的方框图；

图 3 是描述本发明的 H－MESI 存贮器相关协议的说明性实施例的状态图。

现在参照附图，特别参照图 1，图 1 说明了根据本发明的多处理器数据处理系统的高级方框图。如图所示，数据处理系统 8 包括多个处理器 10a-10n，其中每个处理器最好由来自 IBM 公司的 Power PC™系列处理器的其中之一构成。除了常规寄存器、指令流逻辑和用于执行程序指令的执行部件外，每个处理器 10a－10n 还包括板上一级（L1）高速缓存 12a－12n 中相关联的一个高速缓存，这个高速缓存临时存储很可能被相关联的处理器访问的指令和数据。虽然在图 1 中 L1 高速缓存 12a-12n 被图解为存储指令和数据（此后这二者被简单地称作数据）的一体化的高速缓存，但是本领域的技术人员将明白的是，L1 高速缓存 12a－12n 中的每一个都能替换地实现为分开的指令和数据高速缓存。

为了最小化数据访问等待时间，数据处理系统 8 还包括一级或多级

附加的高速缓存，例如二级（L2）高速缓存 14a－14n，它们用于分级至 L1 高速缓存 12a-12n 的数据．换句话说，L2 高速缓存 14a－14n 用作系统存贮器 18 和 L1 高速缓存 12a－12n 的中间存储器，并且它们通常存储比 L1 高速缓存 12a－12n 大的多的数据量，但需要较长的访问等待时间．例如，L2 高速缓存 14a－14n 可以具有 256 或 512KB 的存贮容量，而 L1 高速缓存可以具有 64 或 128KB 的存贮容量．如上所述，虽然图 1 中仅示出了两级高速缓存，但是，数据处理系统 8 的存贮层次可以被扩展至包括串联的附加级（L3、L4，等等）高速缓存或者后备高速缓存．

如图所示，数据处理系统 8 还包括 I/O 设备 20、系统存贮器 18、以及非易失性存贮器 22，它们都耦合至互连部件 16．I/O 设备 20 包括诸如显示设备、键盘和图形指示器等常规外围设备，它们都通过常规适配器连接至互连部件 16．非易失性存储器 22 存储操作系统和其它软件，响应数据处理系统 8 被加电它们被加载到易失性的系统存贮器 18．当然，本领域的技术人员将明白的是，数据处理系统 8 能够包括许多图 1 中未示出的其它部件，诸如用于至网络或所接设备的连接的串并口、管理对系统存储器 18 的访问的存贮控制器，等等．

互连部件 16 可由一个或多个总线或一交叉点开关组成，该互连部件作为用于在 L2 高速缓存 14a－14n 系统存贮器 18、输入/输出（I/O）设备 20、和非易失性存贮器之间的通信事务的管道．在互连部件 16 上的一个典型通信事务包括指示该事务的源的一个源标记、指定该事务的预计的接受者的一个目的标记、一地址和/或数据．耦合到互连部件 16 的每个设备最好都探听在互连部件 16 上的所有通信事务．

现在参见图 2，图 2 描述了根据本发明的 L2 高速缓存 14 的一说明性实施例的较详细的方框图．在该说明性实施例中，L2 高速缓存 14 是采用 32 位地址的四路组相关高速缓存．因此，L2 高速缓存 14 的数据阵列 34 包括许多同余类（congruence class），每个同余类含有用于存储高速缓存行的 4 个路．如常规组相关高速缓存那样，采用在存贮单元地址范围内的索引位（例如 32 位地址的 20－26 位），系统存贮器 18 的存贮单元被映射到特定的同余类．

在数据阵列 34 内存储的高速缓存行记录在高速缓存目录 32 中，该目录包括用于数据阵列 34 的每一路的一个目录项．每个目录项包括标记字段 40、相关状态字段 42、最近最少使用（LRU）字段 44 和包含字

段 46. 标记字段 40 通过存储该高速缓存行的系统存贮器地址的标记位（例如 0－19 位）来确定哪个高速缓存行存储在数据阵列 34 的相应路中. 如下面将参照图 3 详细讨论的那样，相关状态字段 42 利用预定义的位组合来表示存储在数据阵列 34 的相应路中的数据的相关状态. LRU 字段 44 指示最近数据阵列 34 的相应路相对于其同余类的其它路已经如何被访问，由此指示哪个高速缓存行应从该同余类中舍去以响应一高速缓存未命中. 最后，包含字段 46 指示存储在数据阵列 34 的相应路中的高速缓存行是否还存储在相关联的 L1 高速缓存 12 中.

再参照图 2，L2 高速缓存 14 还包括高速缓存控制器 36，它根据从相关联的 L1 高速缓存 12 接收的信号以及在互连部件 16 上探听的事务而管理在数据阵列 34 中的数据的存储和检索以及对高速缓存目录 32 的修改. 如图所示，高速缓存控制器 36 含有一读队列 50 和一写队列 52，从这两个队列中高速缓存控制器 36 执行对高速缓存目录 32 的修改以及对数据阵列 34 的访问. 例如，响应从关联的 L1 高速缓存 12 接收一读操作，高速缓存控制器 36 将该读操作放在读队列 50 的一个项中. 高速缓存控制器 36 通过提供所请求的数据至关联的 L1 高速缓存 12 而满足该读请求的需要，并且然后将该读请求从读队列 50 中删除. 作为另一个例子，高速缓存控制器 36 可以探听到一由 L2 高速缓存 14a－14n 中另一个启动的表示一远程处理器 10 想要修改其本地的一特定高速缓存行的拷贝的事务. 响应于该探听的事务，高速缓存控制器 36 在读队列 50 中放入一个读高速缓存目录 32 的请求以便确定该特定高速缓存行是否驻留在数据阵列 34 中. 如果是那样的话，高速缓存控制器 36 发出一适当的响应到互连部件 16 上，并且如果需要，插入一目录写请求到写队列 52 中，当其被服务时，修改与该特定高速缓存行相关联的相关状态字段. 虽然图 2 示出了在其中仅使用一个读队列和一个写队列的实施例，但是应该明白，高速缓存控制器 36 采用的队列数只是设计选择的问题，并且高速缓存控制器 36 可以为高速缓存目录访问和数据阵列访问采用单独的队列.

高速缓存控制器 36 还包括模式寄存器 60，如下面更详细地的描述那样，该模式寄存器由一位或多位组成，所述一位或多位的设置控制高速缓存控制器 36 的操作. 另外，高速缓存控制器 36 包括性能监视器 70. 性能监视器 70 装配有性能视计数器（PMC0－PMCn）72，当启动时，这

些计数器递增以响应由一个或多个控制寄存器（CR0－CR1n）74 确定的一个事件或事件的组合的每一次发生。可由 PMC72 计数以响应 CR74 的设置的这些事件包括高速缓存命中、高速缓存未命中、特定队列中的项数、L2 高速缓存命中的访问等待时间、L2 高速缓存未命中的访问等待时间，等等。PMC72 和 CR74 中的每一个最好是可由相关联的处理器 10 通过加载和存储指令来读和写的存贮映象寄存器。

现在参见图 3，图中描述了根据本发明的 H－MESI 存贮器相关协议的一说明性实施例。该 H－MESI 协议最好只由存贮层次中最低级的高速缓存（例如，在图 1 的数据处理系统 8 的实施例中的 L2 高速缓存 14a－14n）实现，而较高级的高速缓存最好实现常规的 MESI 协议。但是，在数据处理系统 8 的另一实施例中，H－MESI 协议能够以额外的高速缓存之间的通信量为代价而在存贮层次的每一级高速缓存中实现。

如图 3 所示，H－MESI 存贮器相关协议包括常规的 MESI 协议的修改（M）、专用（E）、共享（S）和无效（I）状态，它们分别由参考号 80、82、84、86 标识。另外，本发明的 H－MESI 存贮器相关协议包括悬停（H）状态 90，H 状态指示存储在相关联的标记字段 40 中的地址标记有效而存储在数据阵列 34 的相应路中的数据项（例如高速缓存行或高速缓存区段）无效。

在一最佳实施例中，任一 L2 高速缓存目录 32 的每一项的相关状态字段 42 在加电时被初始化为 I 状态 86，以便指示标记字段 40 和存储在数据阵列 34 的相应路中的数据都是无效的。类似地，根据常规 MESI 协议，L1 高速缓存目录项也被初始化为无效状态。此后，根据由处理器 10a－10n 产生的存贮器请求的类型和存贮层次对这些请求的响应，处于无效状态 86 的存储于 L2 高速缓存 14a－14n 的其中之一的一高速缓存行（或高速缓存区段）的相关状态能够修改为 M 状态 80、E 状态 82 或 S 状态 84 的其中之一。

例如，如果处理器 10a 产生一读请求以响应一加载指令，则 L1 高速缓存 12a 首先确定所请求的数据是否驻留在 L1 高速缓存 12a 中。响应在 L1 高速缓存 12a 中的命中，L1 高速缓存 12a 简单地将所请求数据提供给处理器 10a。但是，响应在 L1 高速缓存 12a 中的未命中，L1 高速缓存 12a 通过高速缓存之间的连接发送该读请求至 L2 高速缓存 14a。响应在 L2 高速缓存 14a 中的命中，所请求的数据被 L2 高速缓存 14a 提

供至 L1 高速缓存 12a，L1 高速缓存 12a 与适当的 MESI 相关状态结合地存储所请求数据并且发送所请求数据至处理器 10a。但是，如果该读请求在 L1 高速缓存 12a 和 L2 高速缓存 14a 中都未命中，则 L2 高速缓存 14a 的高速缓存控制器 36 作为一事务发出该读请求到互连部件 16 上，该事务被每个 L2 高速缓存 14b－14n 探听。

根据探听在互连部件 16 上的该读请求，在每个 L2 高速缓存 14b－14n 中的高速缓存控制器 36 确定所请求数据是否驻留在其数据阵列 34 或者 L2 高速缓存 12b－12n 中相关联的一个中。如果 L2 高速缓存 14b－14n 或者 L1 高速缓存 12b-12n 都未存储所请求数据，则每个 L2 高速缓存 14a-14n 返回一空响应至 L2 高速缓存 14a，然后 L2 高速缓存 14a 从系统存贮器 18 中请求该数据。当所请求数据从系统存储器 18 返回到 L2 高速缓存 14a 时，高速缓存控制器 36 发送所请求数据至 L1 高速缓存 12a，存储所请求数据到其数据阵列 34 中，并且如参考号 100 所示，修改与存储所请求数据的路相关联的相关状态字段 42 从 I 状态 86 至 E 状态 82。如在常规 MESI 协议中那样，E 状态 82 表示关联的高速缓存行有效并且没有驻留在存贮层次第二级的任何其它高速缓存中。

类似地，如果任一 L1 高速缓存 12b－12n 或者 L2 高速缓存 14b-14n 存储所请求数据在 E 状态 82 或 S 状态 84，并且因此对由 L2 高速缓存发送到互连部件 16 上的该读请求指示“共享”响应，则 L2 高速缓存 14a 从系统存贮器 18 中检索所请求数据。但是，在这种情况下，在存储所请求数据的 L2 高速缓存 14a 中的路的相关状态从 I 状态 86 变换到 S 状态 84，如参考号 102 所示。存储所请求数据于 E 状态 82 的其它 L2 高速缓存 14 也修改至 S 状态 84，如参考号 104 所示。

如果处理器 10a 请求的数据没有驻留在 L1 高速缓存 12a 和 L2 高速缓存 14a 中，而是例如在 L1 高速缓存 12n 中存储为 M 状态 80，则 L2 高速缓存 12n 的高速缓存控制器 36 用一重试回答该读请求并且发信号通知 L1 高速缓存 12n 将所请求数据推入存贮器。然后，在 L1 高速缓存 12n 和 L2 高速缓存 14n 中所请求数据的相关状态被修改至 S 状态 84，如参考号 106 所示。此后，当 L2 高速缓存 14a 重试该读请求到互连部件 16 上时，如上所述，L2 高速缓存 14n 回答一共享响应并且 L2 高速缓存 14a 从系统存贮器 18 中获取所请求数据。在支持所谓的修改干预的另一实施例中，所请求数据由 L2 高速缓存 14n 的高速缓存控制器 36 而

不是系统存贮器 18 来发送，因而减少了访问等待时间。

如果 L1 高速缓存 12a 不是发出一读请求，而是发出一表示处理器 10a 想要获得一存贮单元的专用权以便修改该单元的“想要修改的读”请求，则就会接着发生上述的获得含有该特定存贮单元的高速缓存行的处理。但是，当获得所请求的高速缓存行时，L1 高速缓存 12a 存储所请求高速缓存行在修改状态。另外，由于该“想要修改的读”事务表示所请求的高速缓存行的其它拷贝将变为陈旧的，因此，其它的 L1 和 L2 高速缓存必须将它们的所请求高速缓存行的拷贝表示为无效。在 L1 高速缓存 12b-12n 中，所请求高速缓存行的任何拷贝都简单地标记为无效。但是，存储在 L2 高速缓存 14b－14n 中的该请求高速缓存行的拷贝的相关状态不是象常规的利用交叉－无效（XI）的多处理器数据处理系统那样被修改为 I 状态 86。但是，根据本发明的一个重要方面，存储该请求高速缓存行的拷贝的各个 L2 高速缓存 14b－14n 将与其拷贝相关联的相关状态字段 42 从任意的 S 状态 84、M 状态 80、或者 E 状态 82 分别修改为 H 状态 90，如参考号 110、112、114 所示。如上所述，H 状态 90 指示存储在标记字段 40 中的标记保持有效，但在数据阵列 34 中相关联的高速缓存行无效。类似地，响应其它被探听的要求使数据无效的事务，在高速缓存目录 32 中的项也被修改至 H 状态 90，前述的事务包括清除（Kill）（即，明确地使一特定数据块无效的事务）、清空（flush）（即，使一特定数据块无效并且拷贝任一修改数据至系统存贮器的事务）、dclaim（即，根据一高速缓存行的局部拷贝变成修改状态以响应一存储，使在远程高速缓存中标记为共享的该高速缓存行的拷贝无效的事务），等等。

如参考号 116、118、120 所示，根据一高速缓存接收的事务类型，该高速缓存的目录项可以从 H 状态 90 分别变换到 E 状态 82、M 状态 80、或者 S 状态 84。例如，根据产生一读请求的处理器 10a（在未命中于 L1 高速缓存 12a 和 L2 高速缓存 14a 之后）接收一来自 L2 高速缓存 14b-14n 的空响应，处理 H 状态 90 的 L2 高速缓存 14a 的一目录项产生一个至 E 状态 82 的转换（如参考号 86 所示），这是因为所有的 L2 高速缓存 14a－14n 当中，从系统存贮器 18 检索到的数据仅存储在 L2 高速缓存 14a 中。另一方面，如果处理器 10a 指示想要存储数据到处于 H 状态 90 的 L1 高速缓存 12a 的一路中，L1 高速缓存 12a 指示这个意图至 L2 高速缓存

14a，则 L2 高速缓存 14a 将发送一“想要修改的读”事务到互连部件 16 上．如上所述，响应探听该“想要修改的读”事务，存储在 L2 高速缓存 14b-14n 中的该请求高速缓存行的拷贝被修改为 H 状态 90，而存储在 L1 高速缓存 12b－12n 中的该请求高速缓存行的拷贝被标记为无效．一旦该请求高速缓存行返回到 L1 高速缓存 12a 并且处理器 10a 修改该高速缓存行，在 L1 高速缓存 12a 中该高速缓存行就被标记为修改状态以表示该高速缓存行有效，但是与系统存贮器 18 不相关．根据该实现，其后该修改的高速缓存行可以存储到 L2 高速缓存 14a（例如，响应一 L1 的舍去）而不需要该修改的高速缓存行写回到系统存贮器 18．如果这样，则在 L2 高速缓存 14a 中与该修改的高速缓存行相关联的相关状态字段被修改为 M 状态 80，如参考号 118 所示．最后，根据一些不同的请求响应情况，处于 H 状态 90 的一 L2 高速缓存目录项被修改为 S 状态 84．

首先，当关联的处理器 10 发出一读请求至由标记字段 40 中的（有效）地址指示的地址并且至少一个 L2 高速缓存 14 回答一共享响应时，处于 H 状态 90 的一 L2 目录项变换到 S 状态 84．更重要地，处于 H 状态 90 的一 L2 高速缓存目录项能够修改至 S 状态 84 而不需要关联的处理器 10 发出一数据请求或者 L2 高速缓存 14 产生一事务到互连部件 16 上．如上所述，每个 L2 高速缓存 14a－14n 探听发送在互连部件 16 上的所有事务．如果 L2 高速缓存 14a－14n 的其中之一，例如 L2 高速缓存 14a 探听由 L2 高速缓存 14b－14n 的另一个发出的一事务，该事务包括在 L2 高速缓存 14a 中存储为 H 状态 90 的数据的一修改（即，有效）拷贝，则 L2 高速缓存 14a 的高速缓存控制器 36 从互连部件 16 上采样该数据，存储所探听的数据到数据阵列 34 中，并且将相关联的相关状态字段 42 从 H 状态 90 修改至 S 状态 84．当然，如果需要一响应以保持相关，则 L2 高速缓存 14a 还提供一响应至该探听的事务．例如，如果该探听的事务是一读请求，则 L2 高速缓存 14a 必须提供一表示其想要采样所请求数据的共享响应，使得发出请求的 L2 高速缓存存储所请求数据在 S 状态 84 而不是 E 状态 82．这样，在互连部件 16 上的能够被探听而刷新与一有效地址标记相关联的无效数据的事务包括读事务、写事务、由于高速缓存行的舍去而导致的数据回写至系统存贮器 18，等等．

在图 3 所示的 H－MESI 存贮器相关协议的说明性实施例中可能产生的状态变换概括于下面的表 I 中．

表 I

状态变换	原因	备注
I→E	具有空响应的 CPU 读	
I→S	具有共享或修改响应的 CPU 读	
I→M	CPU“想要修改的读”（rwitm）	
E→S	探听的读	
E→M	CPU rwitm	
E→H	探听的数据无效	探听的数据无效＝rwitm，dclaim，消除，清空，等等
S→M	CPU rwitm	发送 dclaim 到互连部件上
S→H	探听的数据无效	
M→S	探听的读	如果支持修改干预则提供数据
M→H	探听的数据无效	如果探听到的事务是 rwitm，则如果支持修改干预就提供数据
H→E	具有空响应的 CPU 读	
H→S	具有共享或修改响应的 CPU 读；探听的读或写	
H→M	CPU rwitm	

根据本发明的一个重要方面，H－MESI 协议可以精确地或者不精确地实现。H－MESI 协议的精确实现要求 L2 高速缓存 14a-14n 总是采样可在互连部件 16 上得到的数据以便刷新处于 H 状态 90 的无效高速缓存行。相反，非精确实现允许高速缓存 14a-14n 选择地采样互连部件 16 上的数据以便刷新处于 H 状态 90 的高速缓存行。在图 2 所示的说明性实施例中，根据在其模式寄存器 60 中的模式位 62 的状态，每个 L2 高速缓存 14 能够独立于其它 L2 高速缓存地操作于精确模式或者非精确模式。

当在对软件进行调试或性能调整时，由于操作的精确模式促进了更加可预测的工作状态以及一致的软件定时，因此在精确模式中操作 L2

高速缓存 14a-14n 具有特别的优点．另外，在该精确模式中，在两级局部高速缓存中未命中（并且要求局部 L2 高速缓存 14 发出一事务到互连部件 16 上）的数据请求通常很少，因此这种数据请求在软件中可用作一可能“故障”（bugs）的指示．而且，在支持修改干预的本发明的实施例中，该精确 H－MESI 协议保证由处理器 10 请求并且在局部 L2 高速缓存 14 中存储为 H 状态 90 的数据将总是通过修改干预（即，很快地）发送．在精确模式中操作 L2 高速缓存 14 的主要缺点是，对于探听到的能够修改处于 H 状态 90 的 L2 高速缓存行的事务，如果例如由于 L2 高速缓存 14 的写队列 52 满（即，忙）而不能执行该修改，则必须重试该事务．

由于最好不重试必须的操作，例如读操作，以便执行处于 H 状态 90 的选择修改，因此，通常较好的是在正常操作期间使 L2 高速缓存 14a－14n 处于非精确模式．如上所述，操作的非精确模式允许对处于 H 状态 90 的高速缓存行的修改选择性地执行．在一最佳实施例中，当 L2 高速缓存 14 处于非精确模式中时，只有当写队列 52（或者一专用目录写队列，如果有的话）少于一阈值数量的项时才执行对处于 H 状态 90 的高速缓存行的修改．因此，根据写队列 52 中的项数超过一预定阈值，在 L2 高速缓存 14 中的硬件或者由相关联的处理器 10 执行的软件能够用于设置模式位 62 至与该非精确模式对应的状态．但是，如下所述，本发明的其它实施例可以根据其它的标准来选择地执行对处于 H 状态 90 的 L2 高速缓存的修改．

在图 2 所示的数据处理系统 8 的说明性实施例中，每个 L2 高速缓存 14a－14n 能够通过软件或者硬件或者二者的结合而独立地设置为精确模式或非精确模式．例如，如果需要对 L2 高速缓存 14a 操作于其中的模式进行软件控制，则处理器 10a 能够通过执行一以模式寄存器 60 为目标的存储指令而简单地设置模式位 62．另外，软件能够存储值至 CR74，使得 PMC72 对所关心的事件的发生，诸如在写队列 52 中插入并移去项、L2 访问、L2 高速缓存未命中、在 L2 高速缓存未命中时的访问等待时间等等计数．然后软件能够通过执行加载指令访问在所关心的 PMC 72 中的值．根据 PMC72 的一个值或几个值的组合超过软件定义的阈值，该软件能够设置模式位 62 以选择该精确和非精确模式中适当的一个．例如，如果 L2 高速缓存 14a 操作于非精确模式并且 L2 高速缓存未命中

的次数大于 L2 访问总次数的一预定比例，则软件能够设置模式位 62 至与精确模式对应的状态。

类似地，性能监视器 70 能够实现 L2 高速缓存 14a-14n 的操作模式的硬件控制。在一说明性的实施例中，每个性能监视器 70 包括用于根据在一个或多个 PMC72 中累积的一选择事件或多个事件的组合发生的次数超过一预定阈值而产生一设置模式位 62 至特定状态的信号的逻辑。通过性能监视器 70 的缺省设置或者通过相关联处理器 10 执行的软件能够确定 PMC 72 的启动和所述关心的一个事件或多个事件的选择。而在另一个实施例中，根据一选择事件或多个事件的组合发生的次数超过一预定阈值，能够设置性能监视器 70 以产生一性能监视器中断（PMI）。该 PMI 由相关联的处理器 10 提供服务，该关联的处理器 10 执行一改变模式位 62 的状态的中断处理程序。

如上所述，本发明提供了一种用于在多处理器数据处理系统中保持存贮器相关的改进的方法和系统。本发明提供的改进的存贮器相关协议允许在相关联的处理器不发出一明确的读或写请求的情况下，与一有效地址标记相关联的存储在一高速缓存中的一无效数据项被自动地修改为有效数据。这样，因远程处理器的活动而被无效的数据能够在该数据被本地处理器访问之前刷新，从而通过消除从一远程高速缓存或系统存贮器中检索该数据的需求而实质上减少了访问等待时间。由于在不访问存贮器或不请求一锁定的情况下就能修改高速缓存行，因此也实质上减少了对存贮器访问的争用和系统范围内的锁定。

虽然本发明已经参照说明性的实施例而具体地示出和描述，但是应该明白，在不脱离本发明的精神和范围的前提下，本领域的技术人员可以在形式和细节上作出各种改变。例如，在图 3 所示的存贮器相关协议的说明性实施例中，由于 I 状态 86 仅用于在加电时初始化目录项并且决不会从另一状态重新进入，因此可通过删除 I 状态 86 而修改该实施例。如果 I 状态 86 被删除，则在加电时每个 L2 目录项的相关状态字段被初始化为 H 状态 90，并且每个 L2 目录项的标记字段初始化为一标记值，该标记值至少在同一同余类中是唯一的。另外，应该明白的是，图 2 的性能监视器 70 能够可替换地实现为耦合至互连部件 16 的单个系统范围内的性能监视器，而不是在每个 L2 高速缓存 14 内部的多个分离的性能监视器。

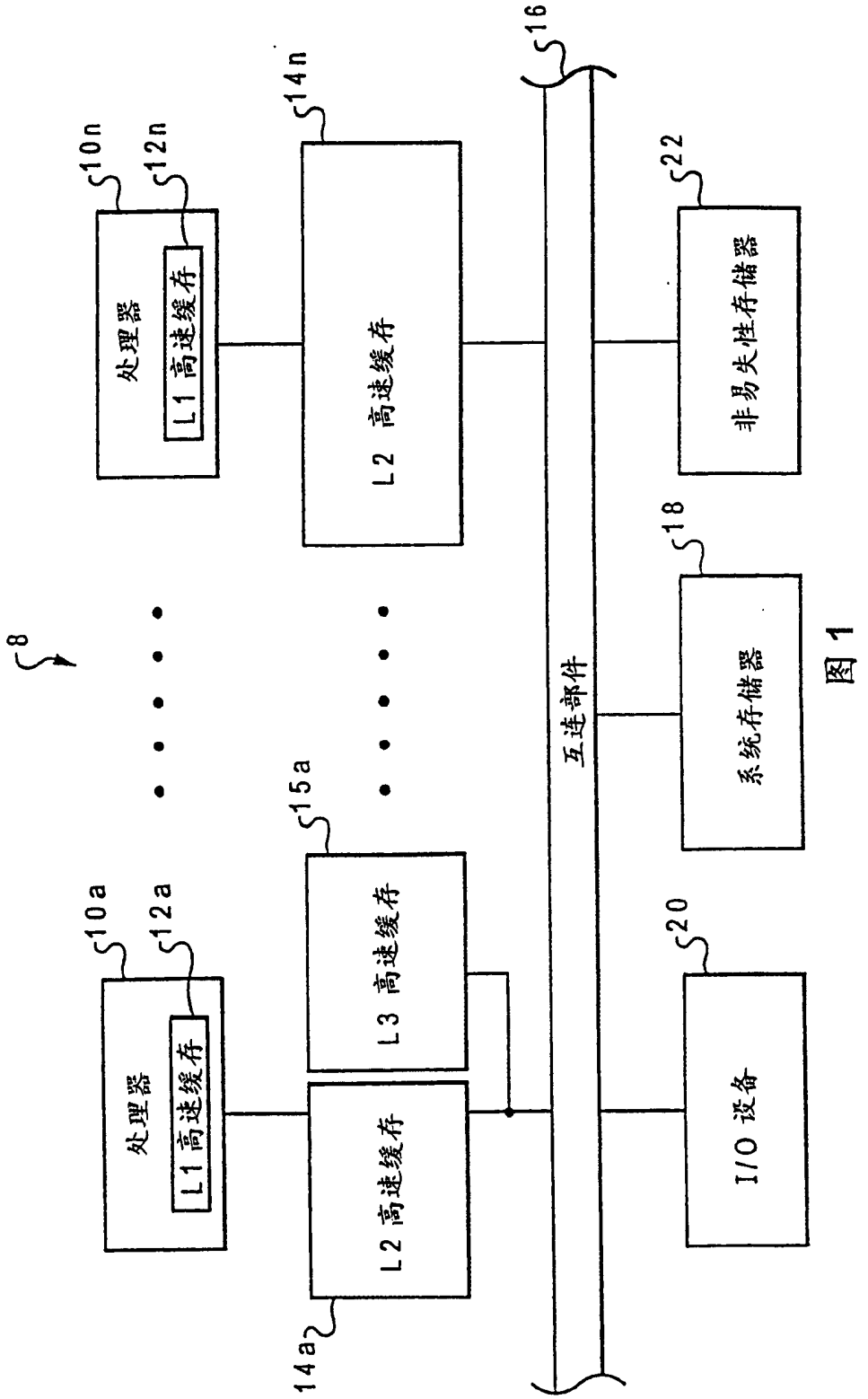

8
处理器
L1 高速缓存
10a
12a
L2 高速缓存
14a
L3 高速缓存
15a
处理器
L1 高速缓存
10n
12n
L2 高速缓存
14n
16
互连部件
系统存储器
18
I/O 设备
20
非易失性存储器
22

图 1

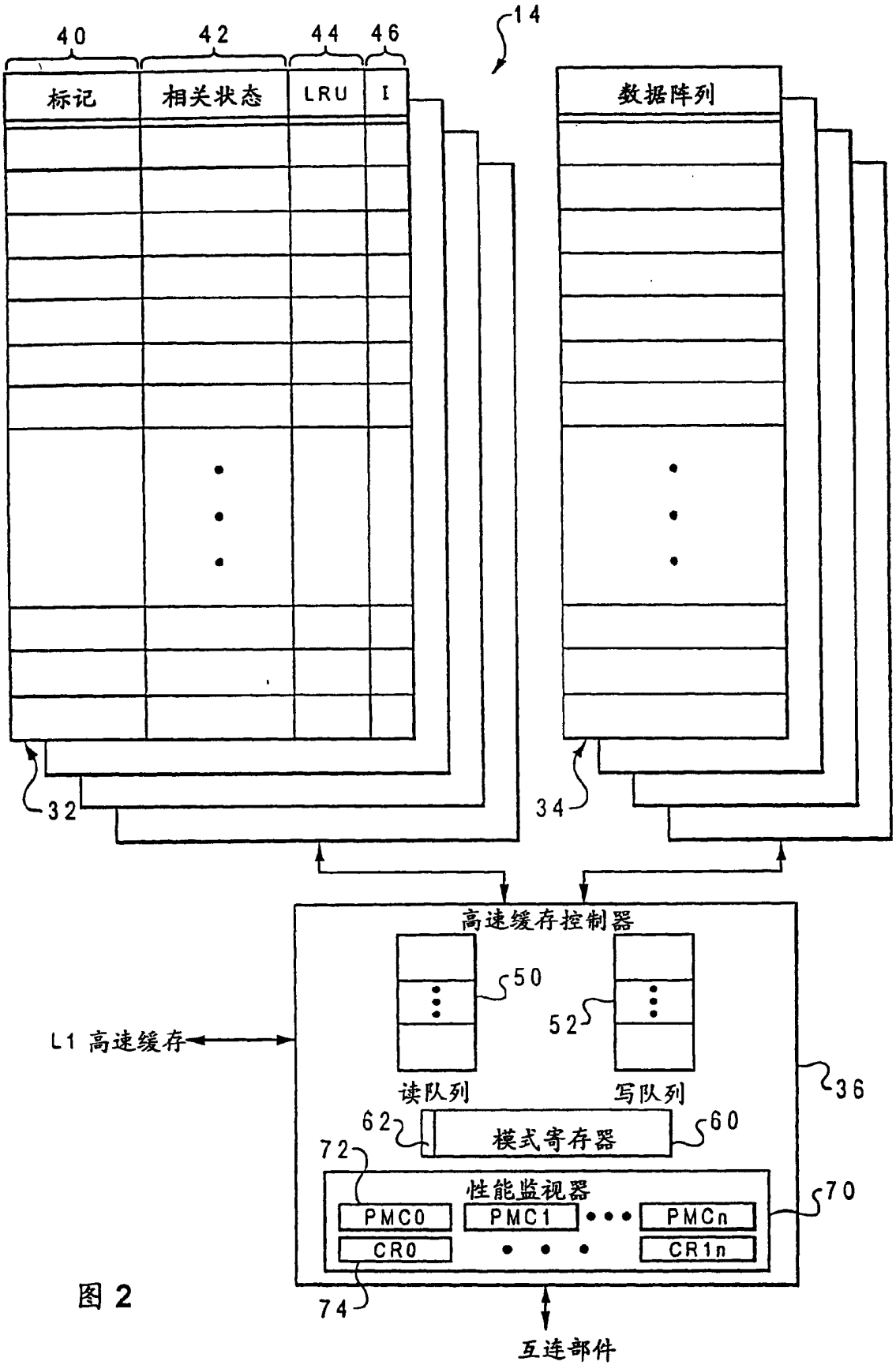

40
42
44
46
14
标记
相关状态
LRU
I
数据阵列
32
34
高速缓存控制器
50
52
L1 高速缓存
读队列
写队列
36
62
模式寄存器
60
72
性能监视器
PMC0
PMC1
PMCn
CR0
CR1n
70
74
互连部件

图 2

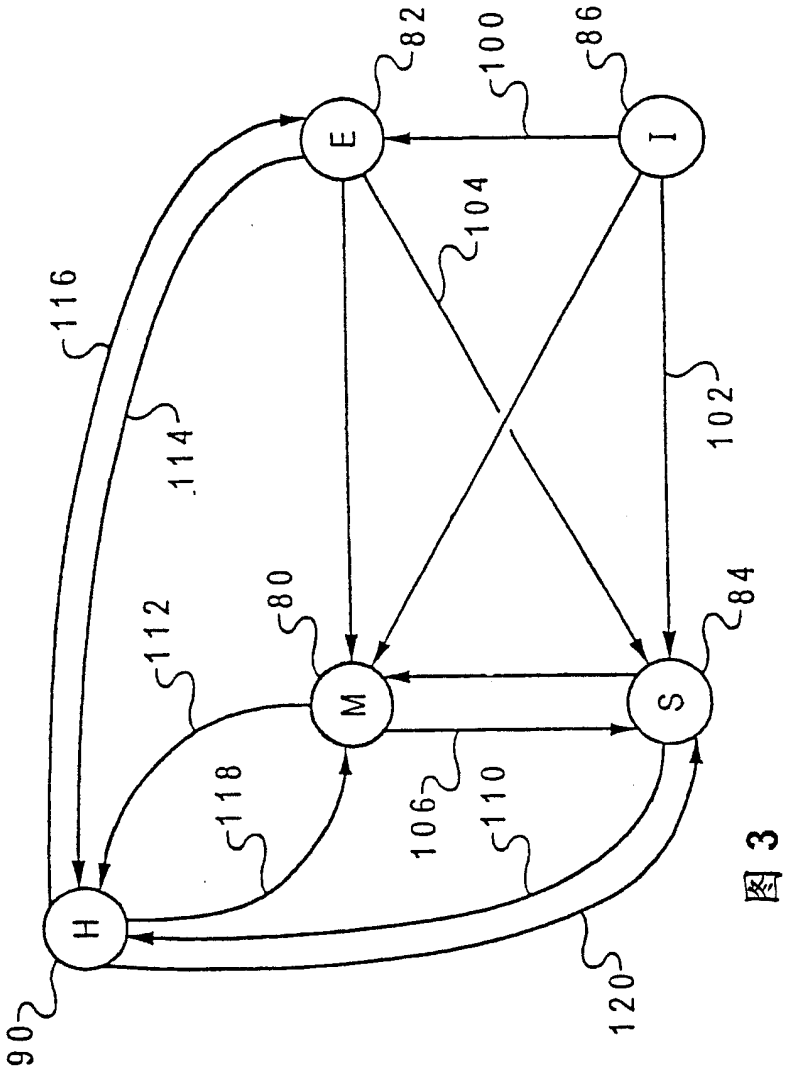


图 3