



(12) 发明专利申请

(10) 申请公布号 CN 103119568 A

(43) 申请公布日 2013. 05. 22

(21) 申请号 201180045186. 7

(51) Int. Cl.

(22) 申请日 2011. 08. 03

G06F 12/00(2006. 01)

(30) 优先权数据

G06F 13/38(2006. 01)

12/860, 340 2010. 08. 20 US

G06F 13/14(2006. 01)

(85) PCT申请进入国家阶段日

2013. 03. 20

(86) PCT申请的申请数据

PCT/US2011/046499 2011. 08. 03

(87) PCT申请的公布数据

W02012/024090 EN 2012. 02. 23

(71) 申请人 英特尔公司

地址 美国加利福尼亚州

(72) 发明人 S·科塔帕里 H·G·尼弗斯

R·帕尔 M·K·阿罗拉 D·纳加拉

(74) 专利代理机构 上海专利商标事务所有限公

司 31100

代理人 邢德杰

权利要求书2页 说明书6页 附图7页

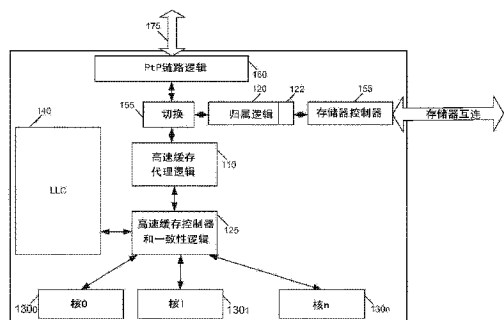
(54) 发明名称

利用目录信息扩展高速缓存一致性监听广播协议

(57) 摘要

在一个实施例中,一种方法包括从第一高速缓存代理接收读取请求,确定与存储器位置相关联的目录条目是否指示信息不存在于远程高速缓存代理中,并且如果是,则在关于读取请求的监听处理完成之前将来自存储器位置的信息发送到第一高速缓存代理。描述并要求保护其他的实施例。

100



1. 一种方法,包括:

在具有监听广播协议的系统的归属代理中接收来自第一高速缓存代理的读取请求,所述读取请求针对存储在与所述归属代理相关联的存储器位置处的信息;

确定与所述存储器位置相关联的目录条目是否指示所述信息不存在于所述系统的远程高速缓存代理中;以及

如果是,则在关于所述读取请求的监听处理完成之前将来自所述存储器位置的所述信息发送至所述第一高速缓存代理。

2. 如权利要求 1 所述的方法,其特征在于,还包括如果确定所述目录条目指示所述信息存在于所述系统的远程高速缓存代理中,则在所述监听处理完成之前防止来自所述存储器位置的信息发送至所述第一高速缓存代理。

3. 如权利要求 1 所述的方法,其特征在于,还包括与向所述归属代理的所述读取请求并行地将来自所述第一高速缓存代理的监听广播请求发送至所述系统的多个高速缓存代理。

4. 如权利要求 3 所述的方法,其特征在于,还包括在所述归属代理中接收对于所述监听广播请求的多个监听响应,并且响应于所述多个监听响应完成所述监听处理。

5. 如权利要求 4 所述的方法,其特征在于,还包括在完成所述监听处理之后将完成消息发送至所述第一高速缓存代理,使得所述读取请求可被解除分配。

6. 如权利要求 1 所述的方法,其特征在于,还包括在所述归属代理中接收来自第二高速缓存代理的针对与来自所述第一高速缓存代理的所述读取请求所针对信息相同的信息的第二读取请求,其中所述第二高速缓存代理属于远程节点,以及更新与所述存储器位置相关联的目录条目以指示所述第二高速缓存代理包括所述信息的副本。

7. 如权利要求 1 所述的方法,其特征在于,还包括响应于所述读取请求,将来自所述归属代理的本地监听请求发送到至少一个本地高速缓存代理,其中所述目录条目不包括关于所述本地高速缓存代理的状态信息。

8. 一种系统,包括:

第一处理器,包括第一高速缓存代理和用于保护存储器的第一本地部分的第一归属代理;

第二处理器,所述第二处理器经由具有监听广播协议的互连耦合至所述第一处理器,所述第二处理器包括第二高速缓存代理和用于保护所述存储器的第二本地区域的第二归属代理,其中所述第二归属代理用于从所述第一处理器接收读取请求,所述读取请求针对存储在所述存储器的所述第二本地区域中的数据,并且如果与所述数据相关联的目录条目指示所述数据的副本不存在于所述系统的远程高速缓存代理中则在关于所述读取请求的监听处理结束之前将所述数据发送至所述第一处理器;以及

所述存储器耦合至第一处理器和所述第二处理器,其中所述存储器用于存储所述目录条目。

9. 如权利要求 8 所述的系统,其特征在于,如果所述目录条目指示所述数据不存在于远程高速缓存代理中,则所述第二归属代理更新与所述数据相关联的目录条目。

10. 如权利要求 8 所述的系统,其特征在于,所述第二归属代理用于响应于所述读取请求产生广播监听并且无需等待对所述目录条目的访问。

11. 如权利要求 8 所述的系统,其特征在于,如果确定所述目录条目指示所述数据存在于远程高速缓存代理中,则所述第二归属代理在所述监听处理完成之前防止所述数据发送至所述第一高速缓存代理。

12. 如权利要求 8 所述的系统,其特征在于,所述第一处理器用于与所述读取请求并行地发送监听广播请求至所述系统的多个高速缓存代理。

13. 如权利要求 12 所述的系统,其特征在于,所述第二归属代理用于接收对于所述监听广播请求的多个监听响应,并且响应于所述多个监听响应完成所述监听处理。

14. 如权利要求 13 所述的系统,其特征在于,所述第二归属代理用于在将所述数据发送到所述第一处理器之后且响应于完成所述监听处理,将不具有所述数据的完成消息发送到所述第一处理器。

15. 如权利要求 8 所述的系统,其特征在于,还包括第一计数器和第二计数器,所述第一计数器用于存储与所述第一本地部分相关联的存储器请求的计数,并且所述第二计数器用于存储与所述第二本地部分相关联的存储器请求的计数,并且其中如果所述第一计数器大于阈值水平则所述第二归属代理用于将一致性处理协议从目录辅助的监听广播协议切换到纯监听广播协议。

16. 如权利要求 8 所述的系统,其特征在于,所述第二归属代理响应于所述读取请求将本地监听请求从所述第二归属代理发送到所述第二高速缓存代理,其中所述目录条目不包括关于所述第二高速缓存代理的状态信息,所述目录条目被存储在所述第二本地部分中。

17. 一种多核处理器,包括:

多个核,每个核与第一高速缓存代理或第二高速缓存代理相关联;

第一归属代理,用于保护存储器的第一本地部分;以及

第二归属代理,用于保护所述存储器的第二本地部分,其中如果应用的请求模式侧重于本地存储器请求则所述第二归属代理用于根据目录辅助的监听广播协议执行所述存储器的所述第二本地部分的一致性处理,否则根据非目录辅助的监听广播协议执行一致性处理。

18. 如权利要求 17 所述的多核处理器,其特征在于,根据所述目录辅助的监听广播协议,所述第二归属代理用于确定与存储器请求相关联的目录条目是否指示所述存储器请求的数据不存在于远程高速缓存代理中,并且如果是,则在关于所述读取请求的监听处理完成之前将所述数据发送至请求方。

19. 如权利要求 18 所述的多核处理器,其特征在于,如果所述目录条目指示所述数据不存在于远程高速缓存代理中,则所述第二归属代理更新所述目录条目,并且所述请求方是远程高速缓存代理。

20. 如权利要求 17 所述的多核处理器,其特征在于,所述第二归属代理与第一计数器相关联以存储与所述第二本地部分相关联的存储器请求的计数,以及与第二计数器相关联以存储与所述存储器的其它本地部分相关联的存储器请求的计数,并且基于所述第一计数器与所述第二计数器之间的比较从目录辅助的监听广播协议切换至非目录辅助的监听广播协议。

利用目录信息扩展高速缓存一致性监听广播协议

[0001] 背景

[0002] 现代计算机系统通常由许多组件形成,这些组件耦合在一起以在中央处理器(一般被称为中央处理单元,CPU)的控制下交互操作和执行多种处理功能。除了作为系统存储器来提供对处理单元所使用的信息的存储的附加半导体设备之外,大多数系统通常包括半导体设备的集合,这些半导体设备包括处理器、外围设备、专用处理单元等。在许多系统中,存在多个存储器,每个存储器可与给定处理器关联,从而用作相应处理器的本地存储器。然而,位于与第一处理器相关联的存储器中的数据可能需要由另一处理器访问。

[0003] 为了维持跨系统的数据一致性,可在系统内实现高速缓存一致性协议,诸如基于目录的协议或基于监听广播的协议。在基于目录的高速缓存一致性协议中,保护存储器的代理(通常称为归属代理,HA)共同地维护目录,该目录跟踪在系统中每个高速缓存线在何地以及在何种状态下被高速缓存。寻求获得高速缓存线的高速缓存代理(CA)发送请求至HA,HA查找目录并仅仅向被目录指示为可具有线的高速缓存副本的那些CA发送消息(一般被称为监听)。然而,该查找会增加等待时间,因为归属代理直至目录查找结束时才会发起监听。

[0004] 替代地,在广播协议中,通常不存在目录,并且CA或HA向所有对等高速缓存代理广播监听。在高速缓存代理之一具有所请求数据的最新副本的情况下,这类协议可改进等待时间。然而,如果监听等待时间长于存储器读取等待时间,那么当从存储器获取数据时等待时间可能更长,因为在这种情况下数据被保持直至所有监听响应已被接收且已由HA处理。即,来自存储器的数据返回被保持直至接收到来自所有代理的监听响应。

[0005] 附图简述

[0006] 图1是根据本发明一个实施例的处理器的框图。

[0007] 图2是根据本发明一个实施例的系统的框图。

[0008] 图3是根据本发明实施例的群集系统的框图。

[0009] 图4是根据本发明一个实施例的方法的流程图。

[0010] 图5是根据本发明实施例执行的操作的流程图。

[0011] 图6是根据本发明另一实施例的目录的一部分的框图。

[0012] 图7是根据本发明另一实施例的处理器的框图。

[0013] 详细描述

[0014] 在多个实施例中,可实现进一步包括目录以使等待时间减少的基于监听广播协议的系统,特别是在数据相对于请求方存在于本地的情况。即,在多个实施例中,除了目录信息之外,可实现监听广播协议,使得当基于目录信息可确定所请求的数据不存在于远程节点中时,可提供与监听相关联的存储器请求(例如,读取请求)的服务并且在监听处理结束之前返回所请求的数据。

[0015] 一般地,对于监听广播协议,响应于存储器请求,利用与存储器请求并行进行的监听请求探测可具有线副本的所有对等高速缓存代理。一般地,在对等高速缓存代理具有线的最新副本的情况下,这种协议可提供更快响应(例如,更低等待时间)。如以下将要进一步

讨论的, 监听广播可由高速缓存代理发送(源监听请求), 或者由归属代理响应于从高速缓存代理接收到读取请求而发送。一般地, 源监听协议可改进高速缓存到高速缓存转移的等待时间, 而源于归属代理的监听协议可允许将监听广播和其它归属代理事务配对。然而, 在大的系统中, 一致性处理所花的时间可能比从存储器取出所请求的数据的等待时间要长。因此从存储器取出的信息可被保持直至所有响应被收集, 如此存储器等待时间可由远程监听响应延迟周期选通, 这尤其会影响本地存储器请求。

[0016] 替代地, 在实现基于目录和非监听广播的流程的系统中, 存储器请求被发送至归属代理, 归属代理查找目录以确定所请求的高速缓存线是否存在于对等代理中。如果是, 那么经由监听请求来探测对等代理。如果替代地目录信息指示没有对等高速缓存代理包括所请求数据的副本, 那么不需要发送监听请求。这样可减小互连带宽。

[0017] 根据多个实施例, 可实现这两个系统的混合。即, 可实现进一步利用存在于目录中的信息的基于监听广播的协议。尽管本发明的范围在这方面不受限制, 但是目录信息可对应于最少量的信息, 例如存储在与其请求的数据相对应的存储器条目的纠错编码(ECC)位中的信息。该系统可被称为目录辅助监听(DAS)广播协议。因此基于存在于目录条目中的最少信息, 当可确定没有远程代理保持所请求数据的副本时, 归属代理可减小与存储器请求相关联的等待时间。

[0018] 现在参见图 1, 所示为根据本发明的一个实施例的处理器框图。如图 1 所示, 处理器 100 包括至少一个高速缓存代理逻辑 110 (一般被称为“高速缓存代理”) 以及一个归属逻辑 120 (一般被称为“归属代理”)。注意术语“设备”或“代理”是通用的且可用于描述耦合到链路的任何电气组件。“链路”或“互连”一般被定义为建立用于消息(即置于预定格式的信息)的通信路径的信息承载介质。链路或互连可以是有线物理介质(例如, 总线、一条或多条电线、迹线、电缆等)或无线介质(例如, 结合无线信令技术的空中)。

[0019] 根据一个实施例, 术语“归属代理”用于指示表示系统存储器的一部分的代理, 并且可被宽泛地定义为提供资源用于高速缓存代理访问存储器、以及基于来自高速缓存代理的请求解决冲突、维持排序等的设备。如此, 归属代理可以是存储器的一部分的指定拥有方并且充当关联存储器的一致性处理和请求的最终仲裁方。归属代理包括跟踪器 122 和用于如下所述的每个高速缓存代理的数据缓冲器。“跟踪器”是可缓冲与各种代理相关联的请求的存储。如以下进一步讨论的, 归属代理还可包括一个或多个计数器或者可与其相关联。根据本发明的一个实施例, “高速缓存代理”一般表示可请求和高速缓存存储器数据副本(以及修改该数据)的高速缓存逻辑。这种高速缓存代理可涵盖适于将存储器请求路由至归属代理的高速缓存控制器。在一个实施例中, 归属代理 120 可支持多达三个高速缓存代理, 一个高速缓存代理用于多个核 130 中的每一个, 并且一个高速缓存代理用于输入/输出中枢(I/OH)写高速缓存(未在图 1 中示出)。

[0020] 核 130 可使用管芯上互连并且经由高速缓存控制器 125 连接至末级高速缓存(LLC)140。包括 LLC 和核高速缓存的高速缓存域由高速缓存代理逻辑 110 和高速缓存控制器 125 封装。该逻辑将底层高速缓存域实现为针对高速缓存一致性系统的单个高速缓存代理。保护与本地存储器控制器 155 连接的存储器的归属代理逻辑在归属逻辑 120 中实现。使用切换逻辑 155 将高速缓存代理逻辑 110 和归属代理逻辑 120 连接至链路逻辑 160。链路逻辑 160 可支持一条或多条点对点(PtP)链路 175。链路逻辑 160 可例如根据源监听广

播协议或基于归属代理的广播监听协议发送基于广播的监听至其它处理器。

[0021] 各实施例可在许多不同的系统类型中使用。在某些实现中,除了目录信息之外还包括广播监听的基于广播的监听协议的混合处理可能尤其适于具有非均匀存储器体系结构(NUMA)的系统。现在参见图 2,所示为根据本发明一个实施例的系统的框图。如图 2 中所示的,系统 200 包括多个插槽 210₀-210₃。每个插槽可包括诸如以上关于图 1 所述的处理器,但是其它实现的确也是可能的。每个插槽可通过 PtP 链路耦合至其它插槽。在一个实施例中,PtP 链路可提供根据 Intel® 快速路径互连(QPI)协议的通信。QPI 协议是包括多层——包括物理层、链路层以及协议层——的高速缓存一致性协议。通过使用该协议,一致性通信可在包括多个高速缓存代理的系统中进行。该协议提供了多个通道和虚拟网络上沿低等待时间链路的各种通信,从而提供了经由 PtP 链路耦合在一起的设备之间的通信。当然,本发明的范围在这方面不受限制,并且在其它实施例中,PtP 链路可遵循另一通信协议。

[0022] 此外,每个插槽可耦合至系统存储器的本地部分(例如,由可包括动态随机存取存储器(DRAM)的多个双列直插式存储器模块(DIMM)构成)。特别地,如图 2 中所示的,每个插槽 210 可耦合至关联的存储器控制器 220₀ - 220₃,存储器控制器 220₀ - 220₃ 又可耦合至系统存储器 230₀ - 230₃ 的相应本地部分。每个插槽可具有被配置成本地存储器的系统存储器的一部分。通过使用本发明的各实施例,当基于目录信息(例如,存储在系统存储器的相应本地部分中)可确定远程节点(例如,其它插槽或其它系统组件中的一个或多个)中不存在请求的数据的副本时,可直接履行待决请求(例如,读取请求)而不等待监听处理完成。

[0023] 注意,根据本发明的实施例的基于 DAS 监听的一致性处理可基于应用利用来控制。即,在一些系统中,当 NUMA 类应用正在执行使得大多数存储器请求去往本地存储器时,可将该混合方法用于广播和目录协议系统。替代地,在基于应用利用可确定许多存储器请求以远程节点为目标的情况下,可作出判定以替代地根据例如纯监听广播协议执行一致性处理。然而,注意即使当没有目录辅助的该纯监听广播流程发生时,如果目录辅助的处理将稍后发生,那么目录可经由对目录的更新消息来维持一致。

[0024] 在一些实施例中,计数器可与本地存储器请求以及远程存储器请求相关联,并且当远程存储器请求的数量超过特定阈值时(例如,依据总存储器请求的预定百分比),一致性处理可从 DAS 监听广播协议切换至纯监听广播协议或纯目录协议。往回参见图 1,每个归属代理可包括计数器,该计数器用于其相应本地存储器部分来存储针对该部分的存储器请求的计数。基于这些计数器的数值,其中这些数值可指示应用的请求模式是否针侧重本地或远程存储器请求,归属代理可根据目录辅助的监听广播协议和非目录辅助的监听广播协议中所选的一个来执行一致性处理。

[0025] 在一个实施例中,存储器 230₀ 可以是主存储器,而存储器 230₃ 可以是例如用于存储镜像数据的辅存储器。尽管在图 2 的实施例中示为独立的存储器控制器组件,但是在某些实现中存储器控制器可被集成在插槽内,诸如在图 1 的处理器插槽中作为示例示出的那些。诸如图 2 中所示的实现由此可用于可实现镜像的插槽无胶配置。利用存储器控制器 220 (和存储器 230₀)分配的数据可通过 PtP 链路被镜像至存储器控制器 230₃ (和存储器 230₃)。然而,注意镜像与目录辅助的监听协议不相关,其也可在非镜像系统中实现。

[0026] 其中可使用各实施例的另一实现在图 3 中示出。在图 3 的实施例中,提供了群集系统。特别地,系统 300 包括多个群集 310₀ - 310₃。每个群集可包括相应的处理器、存储器

控制器和存储器。进而,每个群集可耦合至节点控制器(NC)320。因此,两个插槽群集经由节点控制器来连接。作为示例,数据可从群集 310₀ (群集 0-CPU0) 被镜像到群集 310₃ (群集 1-CPU1)。

[0027] 在图 3 所示的实现中,当使存储器请求成为本地的(例如位于群集内)时,可减少等待时间。这样,当本地请求方寻求访问本地存储器时,所请求的信息可在监听处理完成之前被直接提供给请求方。该存储器等待时间也可存在于作出远程读取请求的情形中,并且可基于目录信息确定没有远程节点包括所请求数据的副本。

[0028] 现在参照图 4,示出的是根据本发明一个实施例的方法的流程图。如图 4 所示,可例如通过使用归属代理逻辑来实现方法 300,归属代理逻辑用于检测传入的对于本地存储器的请求并且确定在监听处理完成之前是否可能取出所请求的数据并将其提供给请求方。

[0029] 如图 4 所示,方法 300 可通过从请求方接收读取请求开始(框 310)。注意,该读取请求可在系统的多个代理中接收并且转发给与要读取的存储器位置相关联的归属代理。该读取请求可与请求方自己根据源监听广播协议所发送的监听广播相关联。否则,归属代理可响应于接收到读取请求而生成监听广播请求并将其发送至多个系统代理。在任一事件中,由此发送与读取请求相关联的监听广播。

[0030] 如图 4 所进一步示出的,在框 320 归属代理可访问与监听请求的存储器位置相关联的目录条目。尽管在一些实施例中归属代理可包括存储最近或频繁访问的目录条目的小高速缓存存储器或者可与其相关联,但是在许多实现中该目录条目可对应于存储在例如 ECC 或存储器位置的其它可用位的存储器内信息。注意,这种对目录条目的访问可与对本地代理的监听请求和取回数据自身两者并行,因为目录条目可以是存储器位置自身的一部分。注意,尽管在图 4 的实现中,对本地代理的本地监听请求可并行发生,但是这并非必须的。替代地,可在针对 NUMA 类系统优化的实现中执行这种本地监听请求,其中大多数存储器请求将针对本地存储器。这样,如以下将进一步讨论的,可避免将信息存储在对给定节点而言是本地的代理的有限目录条目中的需要。这不仅减小了目录条目的大小,而且减小了存储器互连带宽。

[0031] 仍然参见图 4,可基于所接收的信息(例如,目录条目和任何本地监听响应)来确定所请求的存储器位置是否没有存在于远程节点(菱形框 330)。即,在许多实施例中,目录信息可用于指示与目录条目相关联的数据是否存在于系统的远程节点中。如以上所讨论的,在特定优化中,目录条目可能无法指示数据是否存在于本地代理中。然而对于本地监听响应,存在最小等待时间,因此这类优化可能是适当的,尤其是对于基于 NUMA 的应用。

[0032] 如果确定位置的副本不存在于远程节点中,那么控制传递至框 340。在框 340,所请求的数据可被转发到请求方。在该数据转移之后的某时间,监听处理可以完成(框 350)。即,在数据已被发送至请求方之后,可执行和完成用于处理来自不同代理的传入监听响应的多种监听处理。最后在框 360,可将完成消息发送至请求方。注意,该完成消息不具有数据,因为数据先前已被发送。因此,在框 340 发送数据和在框 360 发送该完成消息之间的时间可对应于用于处理该存储器请求的减小的等待时间。注意,请求方在接收到完成消息之后可执行多种操作,例如从请求队列等清除该请求。

[0033] 仍参见图 4,如果替代地在菱形框 330 确定所请求信息的副本存在于远程节点中,那么控制替代地传递至框 370,其中监听处理可以完成。最后在框 380,可将带有数据的完

成消息发送至请求方。因此, 去往请求方的带有数据的该完成可发生在比以上所讨论的在框 340 中执行的更晚的时间。尽管在图 4 的实施例中示出了该特定实现, 但应理解本发明的范围在这方面不受限制。

[0034] 现在参见图 5, 所示为根据本发明实施例示出在处理读取请求中执行的操作的视图。如图 5 所示, 存在多个代理, 即多个高速缓存代理 CA1 和 CA2、节点控制器(NC)、归属代理(HA)以及存储器(MEM)。如所示, 事务流可通过高速缓存代理 1 发出读取请求来开始。注意, 与该读取请求并行, 高速缓存代理还发送监听请求, 由此图 5 所示视图用于源监听广播协议。如所示, 监听请求可涉及系统的其它高速缓存代理, 即 CA2 和 NC, 而读取请求可涉及归属代理。

[0035] 响应于该读取请求, 归属代理可执行存储器读取(MemRd)。注意, 除了从存储器位置读取所请求的数据之外, 还可访问相应的目录条目, 该目录条目可以是与所请求的数据相关联且指示远程节点中存在数据副本的一个或多个位。如所示, 发生存储器取出的返回, 使得归属代理接收数据(MemData)和指示, 该指示通过目录条目的一个或多个位来指示数据是干净的(Clean_Dir), 其在该实施例中意味着数据副本不存在于远程节点中。尽管图 5 中未示出, 应当理解除了执行这种存储器读取和目录读取之外, 归属代理还可并行地监听本地节点中的数据副本。

[0036] 因此由于目录信息指示数据是干净的, 归属代理可在响应数据返回(Data)中将数据直接发送至请求方。在该时刻之后进一步看到, 归属代理可从 CA2 和 NC 接收监听响应(Rsp)。在处理指示这些代理不具有所请求数据的副本的这些响应之后, 归属代理可发送完成消息(Cmp)至请求方。如所示, 第一高速缓存代理对数据的接收与对完成消息的稍后接收之间的时间可对应于根据本发明实施例的等待时间减少。

[0037] 注意, 尽管在图 5 的实施例中示出了该特定实现, 但本发明的范围在这方面不受限制。例如, 假设来自对等高速缓存代理的响应之一指示高速缓存代理包括所请求数据的副本(例如, 处于排他状态或修改状态)。在这种情况下, 在目录信息和监听响应信息之间存在不一致。因此, 可能会产生故障并且使处理机来处理这种一致不一致性, 这可能是对错误的指示, 例如关于在有问题的存储器位置之一的单个位错误。

[0038] 现在参见图 6, 所示出的是可利用具有对目录的最少目录写入的实现根据本发明实施例来执行的多个事务的视图。如所示, 示出了多个事务 410-450。对于每个事务, 具有在事务开始处的当前目录状态, 请求方的节点类型(例如, 本地(L)或远程(R)), 以及请求类型(其中请求类型 G0 指示数据请求不具有对唯一拥有的请求, 例如对于共享数据副本的只读请求, 以及 G1 请求指示对写入线或以其它方式修改线的拥有(例如 RFO)请求)。此外, 每个事务示出目录写入操作是否要发生, 还示出确实发生此类写入操作的目录的新状态。通过减少目录写入操作的数量, 可减少存储器互连带宽。

[0039] 如所示, 对于本地请求方发起的第一事务 410, 其具有的当前目录状态是空闲, 这意味着数据副本不存在于任何远程节点, 无论请求类型如何, 目录条目的状态都不改变并且因而不会发生目录写入。对于由远程请求方发起的请求 420, 无论请求类型如何, 目录写入都会发生, 并且新的目录状态是现在远程拥有(R-Own)以指示数据副本存在于远程节点中。对于下一本地事务 430, 其中当前目录状态是 R-Owned, 当请求类型是 G0 (指示数据请求不具有对于唯一拥有的请求)时, 没有目录写入发生。在事务 440 处, 接收到其中当前目

录状态是 R-Owned 的本地请求,但是在这里请求是寻求拥有的 G1 类型请求。因此,发生目录写入,并且新的目录状态返回到空闲,从而指示没有数据的远程副本。最后在由远程请求方发起的事务 450 中,其中当前目录状态是 R-Own,无论请求类型如何,都没有发生目录写入并且目录条目的状态被维持。

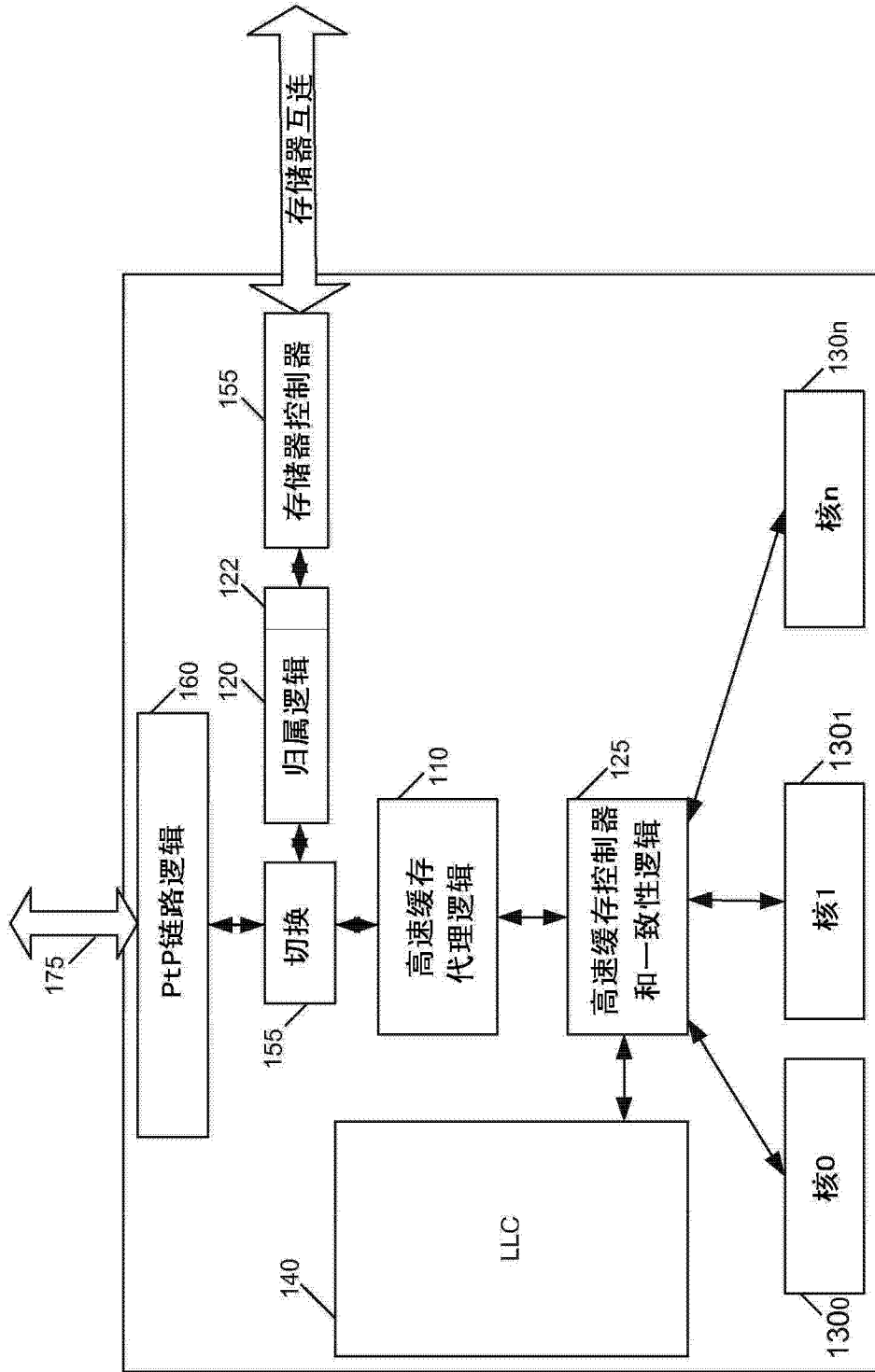
[0040] 注意,关于所有这些事务,当目录处于空闲状态时可实现根据本发明实施例的等待时间改善。或者,在副本存在于共享状态下的情况下,本地或远程读取请求还可受益于等待时间改善。

[0041] 如上所述,特别是在服务器环境中,实施例可能特别适于基于 NUMA 的应用,其中一些应用变成主要负荷。现在参见图 7,所示为根据本发明的另一个实施例的框图。具体地,图 7 示出处理器 500,其是多核处理器并且可能特别适于基于服务器的应用。如所示,处理器 500 包括多个核 510₀ - 510₉。尽管在图 7 的实施例中示出具有特定数量的核,但是应理解本发明的范围在这方面不受限制。每个核可与私有存储(例如,一级或多级高速缓存存储器)相关联。另外,每个核被示为与一片共享的高速缓存存储器相耦合,例如由多个片 520₀ - 520₉ 形成的末级高速缓存(LLC)。如所示,可经由基于环的互连发生经由不同核和高速缓存的通信,所述基于环的互连可以是双向可伸缩环互连 560。为了提供芯片外通信,可存在多个不同的端口和代理。具体地如所示,除了其它输入输出(I/O)端口 575₀ - 575₁ 之外,可存在多个 PtP 端口 570₀ - 570₃。

[0042] 此外,为了提供尤其是用于 NUMA 应用的改善的存储器等待时间,实施例可提供利用多种处理器代理的高速缓存一致性协议。具体地如所示,路由器 530 耦合至一对高速缓存代理 535₀ - 535₁。这些可表示管芯上核种类的高速缓存代理可耦合至相应的归属代理 540₀ - 540₁,并且又可与相应的存储器控制器 545₀ - 545₁ 通信。进而,这些存储器控制器可例如经由芯片外存储器互连耦合至系统存储器的本地部分,例如与处理器相耦合的一个或多个 DRAM。注意,芯片外存储器可包括根据本发明实施例的目录条目,如图 6 所示。在一些实施例中,每个归属代理还可包括小的存储部分,例如用于存储特别热的目录信息的高速缓存,这些目录信息用于被频繁访问或较近访问的少量目录条目。尽管在图 7 的实施例中示出了该特定实现,但应理解本发明的范围在这方面不受限制。

[0043] 各实施例可以用代码实现,并可存储在其上存储有指令的存储介质上,这些指令可用来对系统编程以执行这些指令。存储介质可包括但不限于诸如盘之类的任何类型非瞬态存储介质,盘包括软盘、光盘、光碟、固态驱动器(SSD)、紧凑盘只读存储器(CD-ROM)、可重写紧凑盘(CD-RW)和磁光盘、诸如只读存储器(ROM)的半导体器件、诸如动态随机存取存储器(DRAM)、静态随机存取存储器(SRAM)之类的随机存取存储器(RAM)、可擦除可编程只读存储器(EPROM)、闪存、电可擦除可编程只读存储器(EEPROM)、磁卡或者光学卡、或适于存储电子指令的任意其它类型的介质。

[0044] 尽管本发明已针对有限数量的实施例作了描述,然而本领域技术人员将会从其中领会到许多修改和变型。所附权利要求旨在覆盖所有这样的修改和变型,只要其落在本发明的真实精神和范围内。



100

图 1

200

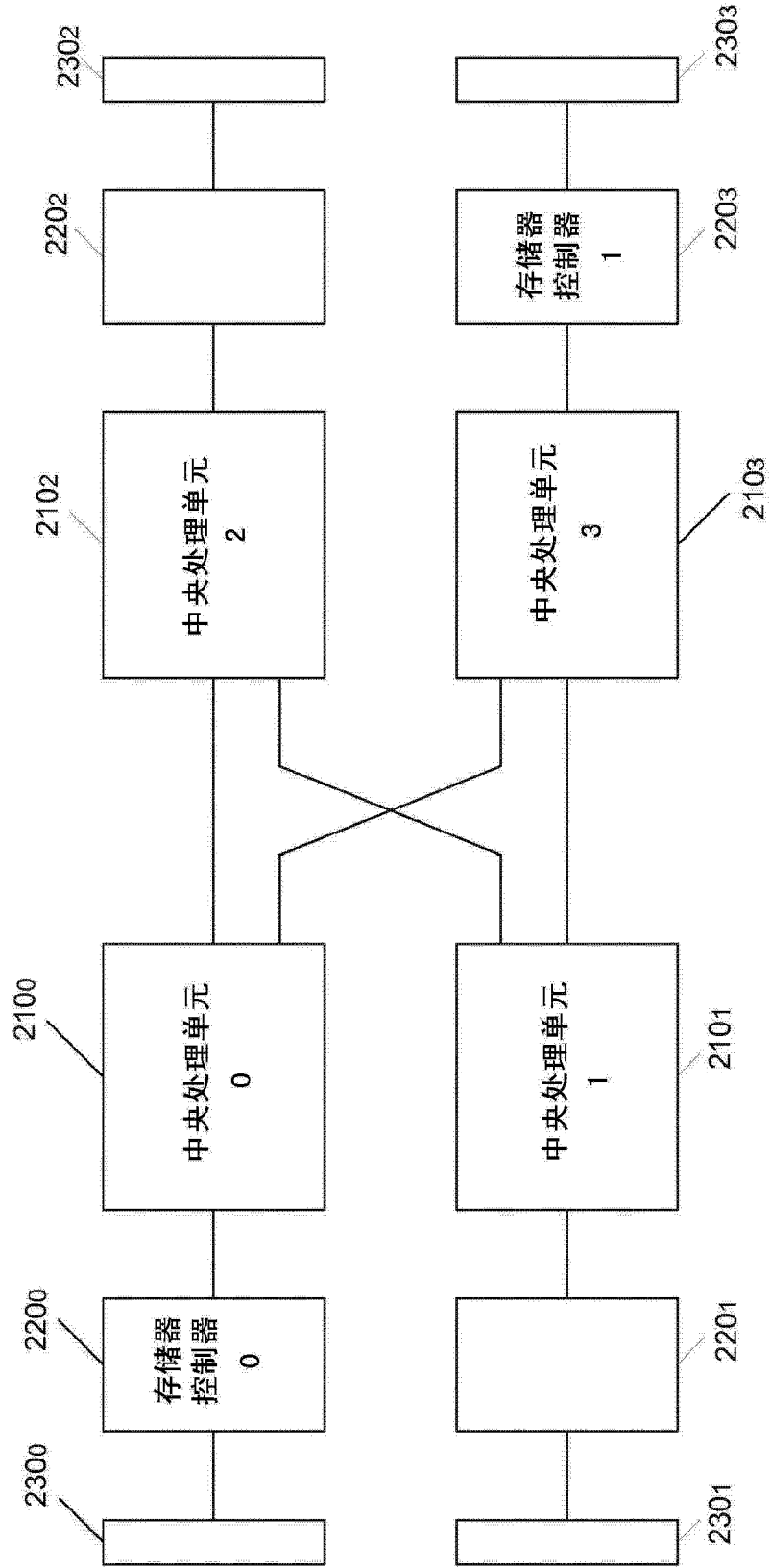


图 2

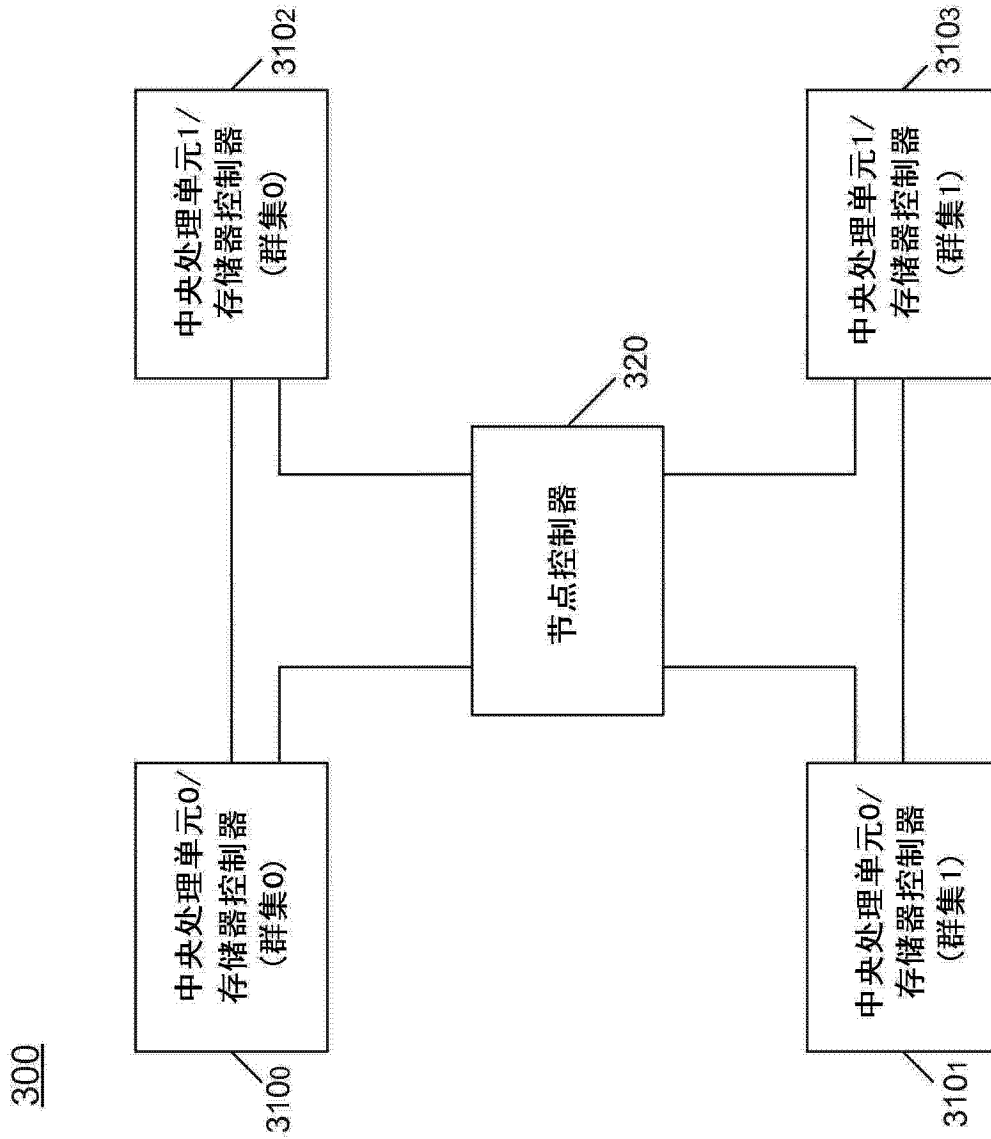


图 3

300

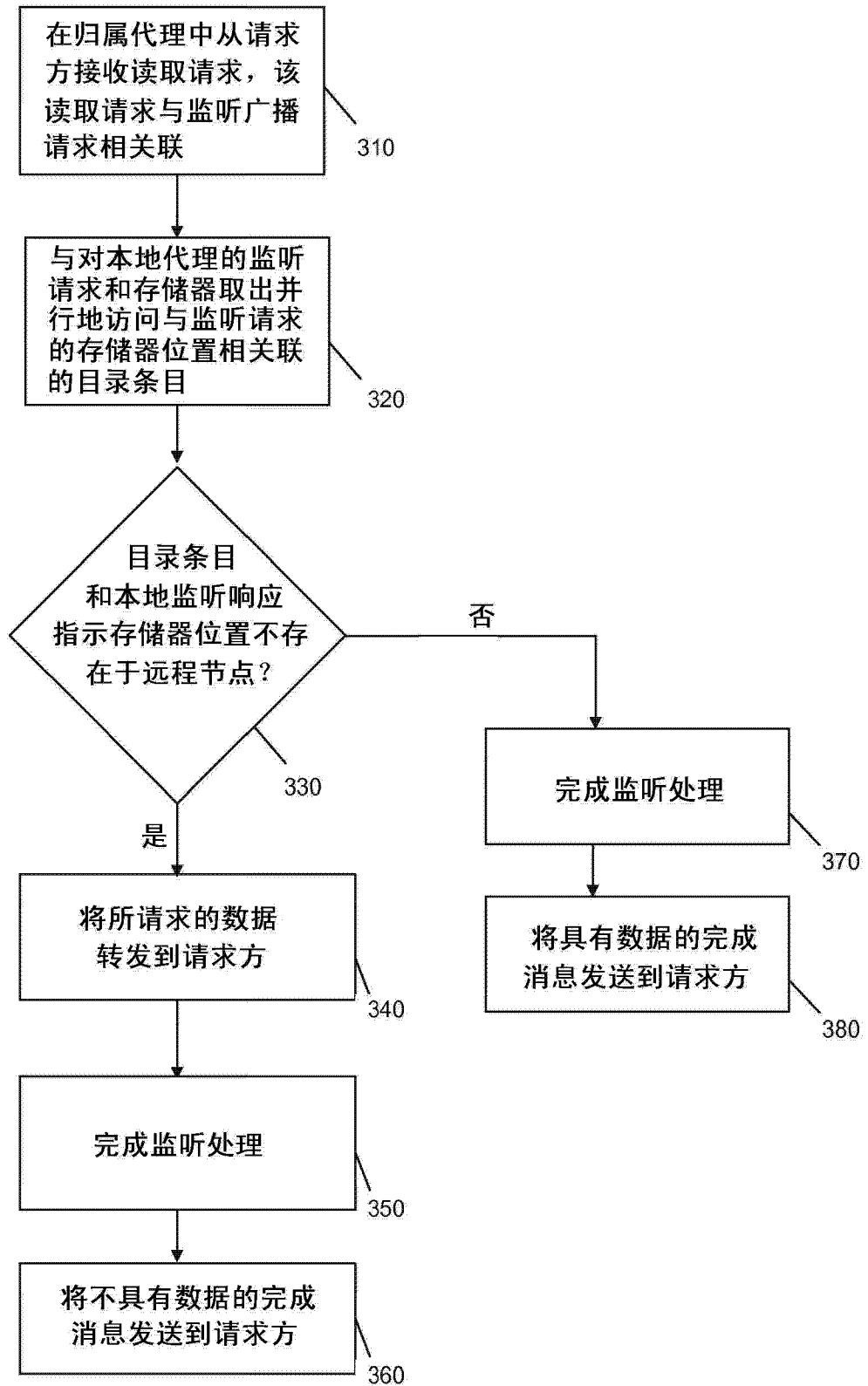


图 4

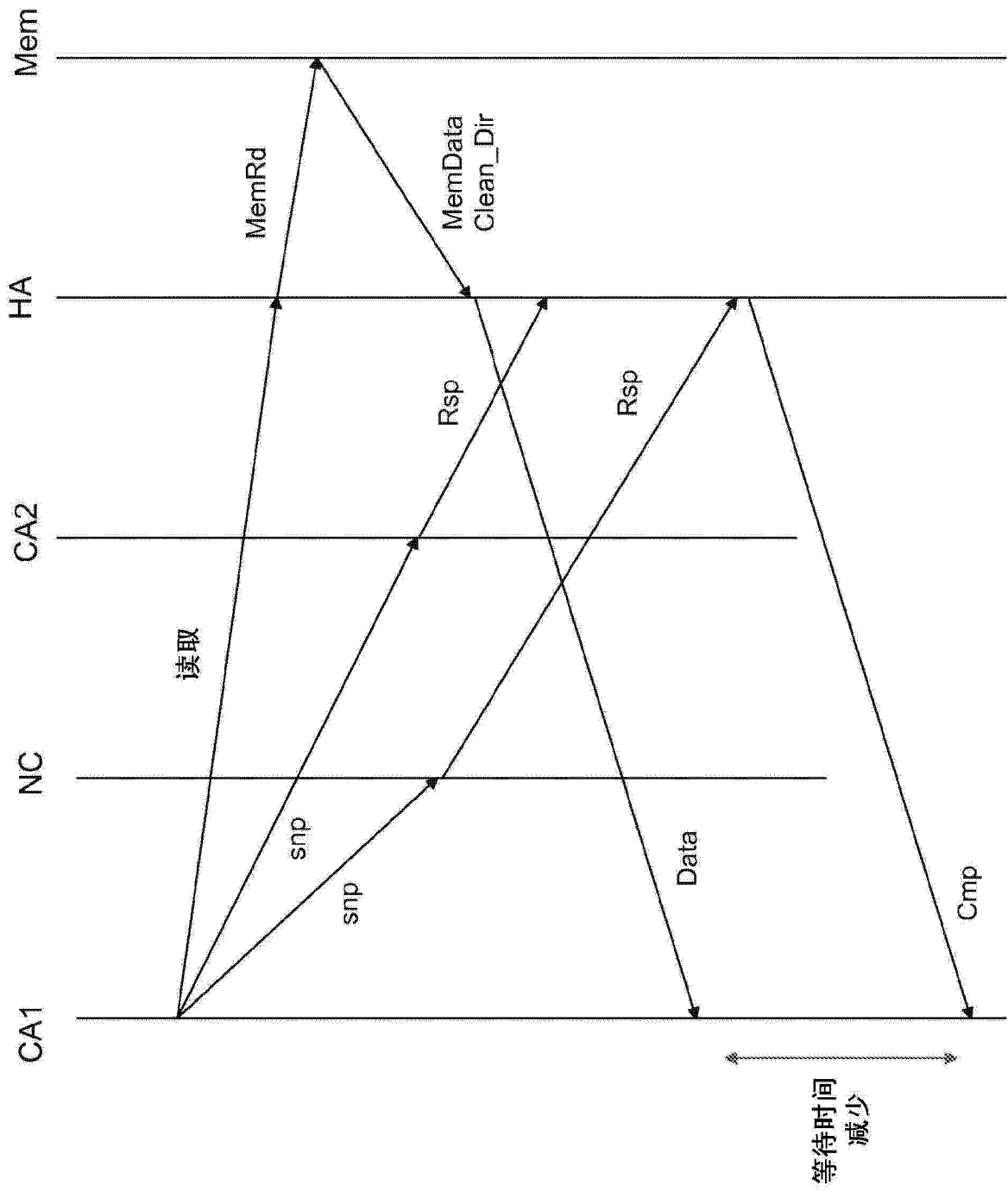


图 5

| 当前目录状态 | 请求方节点类型 | 请求类型 | 目录写入 | 新的目录状态 |
|--------|---------|-------|------|--------|
| 空闲 | L | G0/G1 | 否 | 空闲 |
| 空闲 | R | G0/G1 | 是 | R-Own |
| R-Own | L | G0 | 否 | R-Own |
| R-Own | L | G1 | 是 | 空闲 |
| R-Own | R | G0/G1 | 否 | R-Own |

410

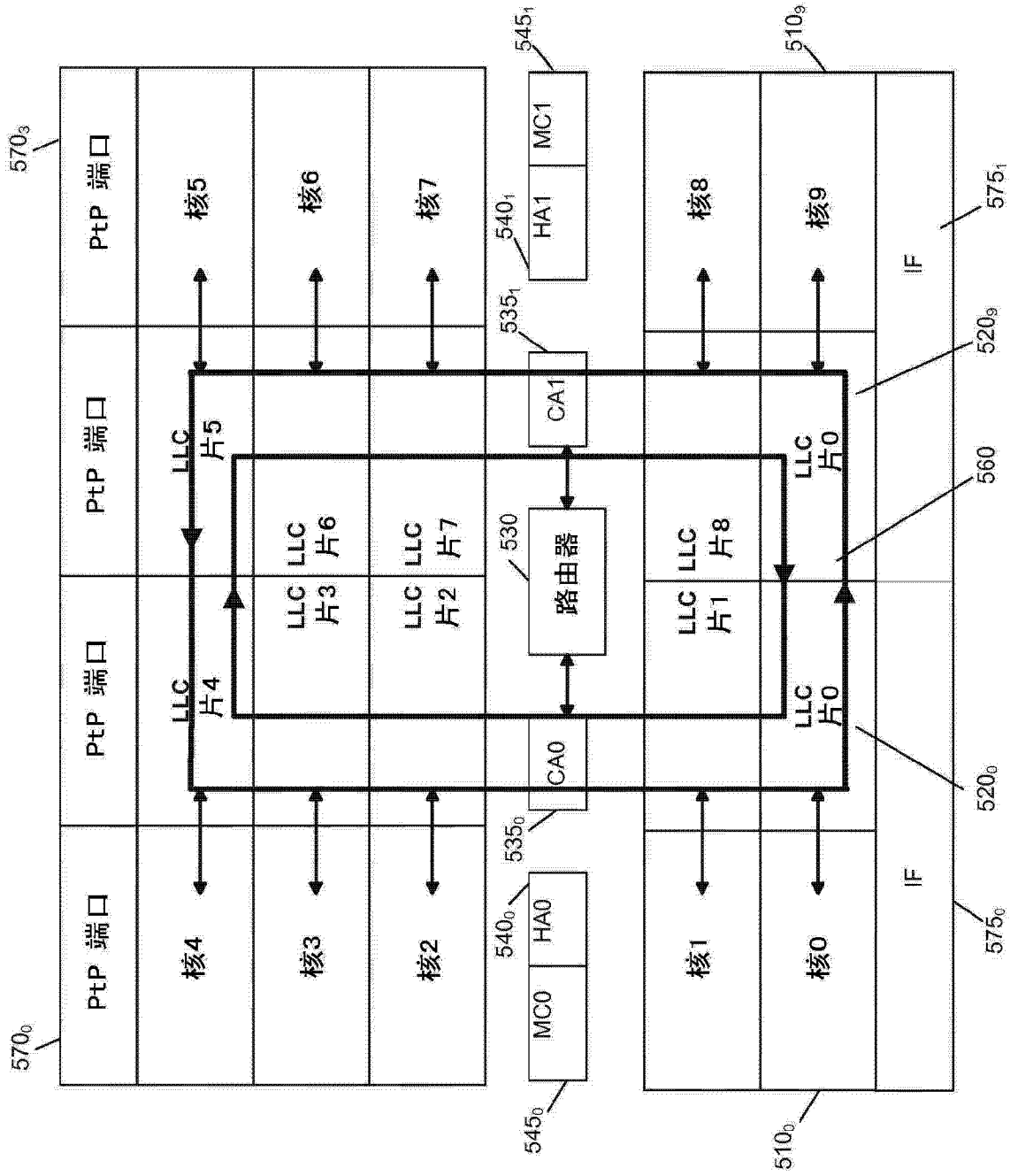
420

430

440

450

图 6



500

图 7