



(12)发明专利

(10)授权公告号 CN 106415522 B

(45)授权公告日 2020.07.21

(21)申请号 201580030595.8

(22)申请日 2015.05.07

(65)同一申请的已公布的文献号
申请公布号 CN 106415522 A

(43)申请公布日 2017.02.15

(30)优先权数据
61/990,199 2014.05.08 US

(85)PCT国际申请进入国家阶段日
2016.12.08

(86)PCT国际申请的申请数据
PCT/US2015/029705 2015.05.07

(87)PCT国际申请的公布数据
W02015/171905 EN 2015.11.12

(73)专利权人 美光科技公司
地址 美国爱达荷州

(72)发明人 理查德·C·墨菲

(74)专利代理机构 北京律盟知识产权代理有限公司
代理人 路勇

(51)Int.Cl.

G06F 15/167(2006.01)

G06F 12/00(2006.01)

(56)对比文件

CN 102157187 A, 2011.08.17, 说明书第14-22, 47-64, 70-102段, 附图2-3, 7-9.

CN 102157187 A, 2011.08.17, 说明书第14-22, 47-64, 70-102段, 附图2-3, 7-9.

US 2013138892 A1, 2013.05.30, 说明书第54-68段.

US 2006129709 A1, 2006.06.15, 全文.

US 2004034747 A1, 2004.02.19, 全文.

US 6415364 B1, 2002.07.02, 全文.

Kim, Gwangsun等.Memory-centric System Interconnect Design with Hybrid Memory Cubes.《International Conference on Parallel Architectures and Compilation Techniques》.2013,

J. Thomas Pawlowski.ybrid memory cube (HMC).《2011 IEEE Hot Chips 23 Symposium (HCS)》.2011, 第1-24页.

审查员 曹青

权利要求书2页 说明书8页 附图7页

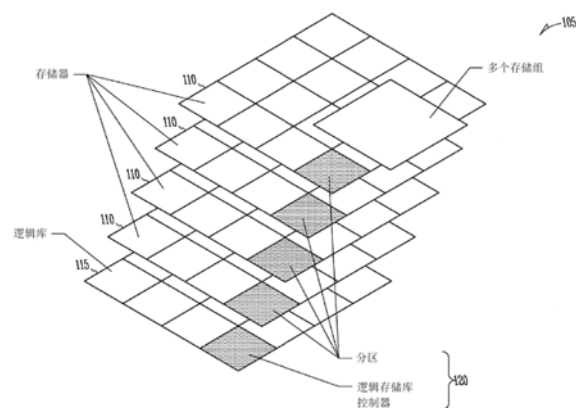
(54)发明名称

存储器内轻量一致性

(57)摘要

一种系统包含多个主机处理器及经配置为用于所述主机处理器的分布式共享存储器的多个HMC装置。HMC装置包含：多个集成电路存储器裸片，其包含布置于第二存储器裸片的顶部上的至少第一存储器裸片，且所述存储器裸片的存储器的至少一部分经映射以包含存储器一致性目录的至少一部分；及逻辑库裸片，其包含：至少一个存储器控制器，其经配置以管理至少一个第二装置对所述多个存储器裸片的存储器的三维3D存取；及逻辑电路，其经配置以确定存储于所述多个存储器裸片的所述存储器中的数据的数据的存储

器一致性状态信息，传递关于对存储器的所述存取的信息且将所述存储器一致性信息包含于所述所传递信息中。



1. 一种计算系统,其包括:

多个主机处理器;

多个存储器裸片,其包含布置于第二存储器裸片的顶部上的至少第一存储器裸片,且所述存储器裸片的存储器的至少一部分经映射以包含存储器一致性目录的至少一部分;及

逻辑库裸片,其与至少所述第一存储器裸片及所述第二存储器裸片封装在一起,所述逻辑库裸片包含:至少一个存储器控制器,其经配置以管理至少一个第二装置对所述多个存储器裸片的存储器的存取;及逻辑电路,其经配置以确定存储于所述多个存储器裸片的所述存储器中的数据的数据一致性状态信息,将所述存储器一致性状态信息和所述数据存储在所述存储器裸片中;及将所述存储器一致性状态信息包含于响应于存取所述存储器裸片的请求而发送的经包化输入/输出I/O协议的经包化消息中,

其中所述计算系统包括多个混合存储器立方体HMC装置,其中HMC装置包含至少所述第一存储器裸片、所述第二存储器裸片及所述逻辑库裸片,及

其中HMC装置的所述逻辑电路经配置以响应于存取所述存储器一致性状态信息的原子存储器请求,授予对存储的所述存储器一致性状态信息的存取为高速缓冲存储器线;及通过重新映射在所述原子存储器请求中接收的地址确定所述高速缓冲存储器线的地址。

2. 根据权利要求1所述的系统,其中HMC装置的所述逻辑电路经配置以作为所述I/O协议的一部分路由经包化信息且检测与所述经包化信息包含在一起的存储器一致性状态信息。

3. 根据权利要求1所述的系统,其中HMC装置的所述逻辑电路经配置以确定所述存储器裸片的存储器字的经修改状态信息、共享状态信息及无效状态信息且与所述存储器裸片的所述存储器字相关联地存储所述状态信息。

4. 根据权利要求1所述的系统,其中HMC装置的所述逻辑电路经配置以确定所述存储器裸片的存储器字的专属权状态信息及所有权状态信息中的至少一者且与共享存储器数据字相关联地存储所述状态信息。

5. 根据权利要求1所述的系统,其包含经配置为集线器HMC装置的至少一个HMC装置,其中所述集线器HMC装置的逻辑电路将路由信息添加到经包化消息且检测所述经包化消息中的存储器一致性状态信息。

6. 根据权利要求1所述的系统,其中HMC的所述逻辑库裸片的所述存储器控制器经配置以管理至少一个处理器及另一HMC装置对所述存储器裸片的存储器的存取。

7. 根据权利要求1所述的系统,其中HMC装置的所述存储器裸片包含布置为多个存储库的存储器,且所述逻辑库裸片包含用于所述存储器裸片的每一存储库的存储器控制器。

8. 根据权利要求1到7中任一权利要求所述的系统,其中所述多个主机处理器的所述处理器是系统单芯片SoC处理器。

9. 一种操作计算机系统的方法,所述方法包括:

管理对所述计算机系统的分布式共享存储器的存取,其中所述共享存储器经配置以用于三维存取;

由所述分布式共享存储器确定由存储器存取产生的存储器一致性状态信息,且将所述存储器一致性状态信息和数据存储在所述分布式共享存储器中;

在所述计算机系统的多个处理器及共享存储器端点当中传递经包化信息且将所述存

存储器一致性状态信息与所传递的经包化信息包含在一起;及

使用所述共享存储器端点来跟踪所述存储器一致性状态信息,

其中使用所述共享存储器端点来跟踪所述存储器一致性状态信息包含:使用包含所述共享存储器的混合存储器立方体HMC装置来跟踪所述存储器一致性状态信息;响应于存取所述存储器一致性状态信息的原子存储器请求,授予对存储的所述存储器一致性状态信息的存取为高速缓冲存储器线;及通过重新映射在所述原子存储器请求中接收的地址确定所述高速缓冲存储器线的地址。

10. 根据权利要求9所述的方法,其包含将所述存储器一致性状态信息与共享存储器数据字存储在一起。

11. 根据权利要求9所述的方法,其包含使用包含用于所述共享存储器的至少一个存储器控制器的逻辑库层来修改所述存储器一致性状态信息。

12. 根据权利要求9到11中任一权利要求所述的方法,其中将存储器一致性状态信息存储于所述共享存储器中包含:将经修改位、共享位及无效位与所述共享存储器数据字存储在一起。

13. 一种具有单电子装置封装的电子装置,所述电子装置包含:

多个存储器裸片,其经配置为系统存储器,且包含布置于第二存储器裸片的顶部上的至少第一存储器裸片,且所述存储器裸片的存储器的至少一部分经映射以包含存储器一致性目录的至少一部分;及

逻辑库裸片,其与所述多个存储器裸片封装在一起,所述逻辑库裸片包含:至少一个存储器控制器,其经配置以管理至少一个第二装置对所述多个存储器裸片的存储器的存取;及逻辑电路,其经配置以确定存储于所述多个存储器裸片的所述存储器中的数据的数据的存储器一致性状态信息,将所述存储器一致性状态信息和所述数据存储在所述存储器裸片中,响应于存取所述存储器一致性状态信息的原子存储器请求,授予对存储的所述存储器一致性状态信息的存取为高速缓冲存储器线,通过重新映射在所述原子存储器请求中接收的地址确定所述高速缓冲存储器线的地址;及将所述存储器一致性状态信息包含于响应于存取所述存储器裸片的请求而发送的经包化输入/输出I/O协议的经包化消息中,

其中所述电子装置是混合存储器立方体HMC装置。

14. 根据权利要求13所述的电子装置,其中所述逻辑库裸片的所述逻辑电路经配置以使用用于存取所述存储器裸片的输入/输出I/O协议来传递经包化信息且作为所述I/O协议的一部分实施存储器一致性协议。

15. 根据权利要求14所述的电子装置,其中所述逻辑库裸片的所述逻辑电路经配置以作为所述I/O协议的一部分路由经包化信息且检测与所述经包化信息包含在一起的存储器一致性状态信息。

16. 根据权利要求13所述的电子装置,其中所述逻辑库裸片的所述逻辑电路经配置以确定所述存储器裸片的存储器字的经修改状态信息、共享状态信息及无效状态信息且与所述存储器裸片的所述存储器字相关联地存储所述状态信息。

17. 根据权利要求13到16中任一权利要求所述的电子装置,其中所述逻辑库裸片的所述逻辑电路经配置以确定所述存储器裸片的存储器字的专属权状态信息及所有权状态信息中的至少一者且与共享存储器数据字相关联地存储所述状态信息。

存储器内轻量一致性

[0001] 相关申请案

[0002] 此美国专利申请案主张2014年5月8日提出申请的序列号为61/990,199的美国临时申请案的优先权权益,所述临时申请案的内容以全文引用的方式并入。

背景技术

[0003] 多处理器系统可包含允许多个处理器对同一数据执行操作的共享存储器。对共享存储器的存取可为缓慢的且可导致系统等待时间。为改善存储器存取时间,多处理器共享存储器系统可包含相对较小但较快的存取高速缓冲存储器。高速缓冲存储器存储可由处理器频繁存取的数据的副本。为维持经高速缓存数据的正确性,后续接着高速缓冲存储器一致性协议。高速缓冲存储器一致性协议对通过处理器的对经高速缓存数据的存取强加规则以确保尽管存在数据的多个副本,但数据跨越所有处理器是正确的。然而,高速缓冲存储器一致性协议可消耗宝贵系统带宽且产生数据存取的不均匀等待时间。本发明人已认识到在多处理器共享存储器系统中需要经改善数据管理。

附图说明

[0004] 图1图解说明可用于实践本发明的一些实施例的混合存储器立方体的表示。

[0005] 图2是图解说明可用于实践本发明的一些实施例的混合存储器立方体的四分之一的框图。

[0006] 图3展示可用于实践本发明的一些实施例的混合存储器立方体的四个可能装置拓扑的框图。

[0007] 图4展示可用于实践本发明的一些实施例的一种一致性目录存储器格式的实例。

[0008] 图5展示可用于实践本发明的一些实施例的用HMC装置实施的计算机系统的实例的框图。

[0009] 图6展示操作计算机系统的方法的实例的流程图。

[0010] 图7展示操作计算机系统的方法的另一实例的流程图。

具体实施方式

[0011] 在本发明的实例性实施例的以下详细描述中,通过图式及图解说明方式而参考本发明的特定实例性实施例。充分详细地描述这些实例以使得所属领域的技术人员能够实践本发明,且所述实例用于图解说明本发明可如何应用于各种目的或实施例。存在本发明的其它实施例且所述实施例在本发明的范围内,且可在不背离本发明的标的或范围的情况下做出逻辑、机械、电及其它改变。本文中所描述的本发明的各种实施例的特征或限制虽然对其中并入所述特征或限制的实例性实施例是必要的,但其并不限制本发明的其它实施例或作为整体的本发明,且对本发明的任何参考(其元件、操作及应用)并不限制作为整体的本发明,而是仅用于定义这些实例性实施例。因此,以下详细描述并不限制仅由所附权利要求书界定的本发明的各种实施例的范围。

[0012] 在常规计算机系统中,存储器以二维方式组织成若干行及若干列。为改善存储器存取等待时间做出的努力已包含开发具有高带宽接口的双数据速率类型三 (DDR3) 或DDR4。DDR3及DDR4使用传统行/列方法。

[0013] 图1图解说明混合存储器立方体105 (HMC) 的表示。HMC通过堆叠集成电路 (IC) 存储器裸片110且使用实现在IC裸片的顶部及底部两者上的存取的穿硅通孔 (TSV) 结构 (未展示) 来互连裸片而包含第三维度。TSV制造技术实现多个裸片层的互连,以便构造三维裸片。互连多个裸片层的此能力准许建造具有存储器存储层与一或多个逻辑层的组合的存储器装置。以此方式,所述装置在单个电子装置封装中提供物理存储器存储及逻辑存储器事务处理。图1中所展示的布置是用以图解说明概念且未必表示实际裸片布局。

[0014] HMC (例如,如混合存储器立方体规范版本1.0中所规定) 可提供具有高达每装置320GB/s的可用带宽容量的极紧凑且功率高效封装。HMC装置能够经由阶层式且并行设计方法而具有高带宽。装置阶层跨越逻辑层而垂直地发生且硬件并行跨越给定裸片层而发生。在HMC内,存储器组织成若干存储库。每一存储库120使用穿硅通孔在裸片内垂直地横跨存储器层中的每一者。在图1的经简化实例中,HMC组织成16个存储库,其中每一存储库具有逻辑库裸片层115或逻辑库。

[0015] 图2是图解说明HMC 205的四分之一的框图。展示HMC的四个存储库,其中每一存储库220包含多个存储器层。逻辑库层215包含多个组件。例如,每一存储库在逻辑库中具有管理用于所述存储库的所有存储器参考操作的存储器控制器。存储器控制器提供对3D存储层的灵活存取。3D分层方法准许存储器事务不仅跨越目标存储阵列内的存储组,且还跨越并行存储阵列并行地存在。逻辑库层215可包含用以实施所描述功能的一或多个处理器,且HMC可为存储器中处理器 (PIM) 装置。

[0016] 逻辑库层215提供对HMC装置的外部链路存取以及内部路由及事务逻辑两者。外部I/O链路由四个或八个逻辑链路提供。每一链路可包含十六个或八个串行I/O或SERDES双向链路的群组。四链路装置具有在10Gbps、12.5Gbps及15Gbps下操作的能力。八链路装置具有在10Gbps下操作的能力。在内部,链路附接到路由逻辑以便引导控制每一垂直存储器存储单元的逻辑装置处的事务。在图2的实例中,路由逻辑包含纵横开关逻辑。

[0017] HMC链路结构支持将装置附接到主机 (处理器) 或其它HMC装置两者的能力。链接 (chaining) 的此概念准许构造需要比单个装置大的容量而不扰乱链路结构及经包化事务协议的存储器子系统。链路可以众多拓扑配置为主机装置链路或直通链路。

[0018] 图3展示基于库四链路HMC配置的四个可能装置拓扑的框图。HMC存储器装置305可以传统网络拓扑 (例如网状、环面或纵横状) 配置。将多个HMC装置链接在一起增加可用于主机的总存储器容量。

[0019] 返回到图2,针对每一独立存储库,存储库的逻辑库层类似于DIMM控制器单元。在图2的实例中,存储库逻辑块225及其相应存储库存储单元220可组织成四重单元。每一四重单元表示四个存储库单元。每一四重单元与最接近物理链路块松散地相关联。以此方式,主机装置具有以下能力:通过逻辑上将请求包发送到其相关联四重单元物理上最接近于所需存储库的链路而使HMC装置的整个逻辑库层的等待时间最小化。

[0020] 一旦在目标存储器存储库内,存储器存储装置便可分解成传统概念的存储组及动态随机存取存储器 (DRAM)。通过经堆叠存储器层的垂直存取类似于选择适当存储器存储

组。下部存储组可经配置于下部裸片层中,而垂直上升(vertical ascension)选择后续存储组。一旦在存储组层内,便可使用行及列以传统方式组织DRAM。存储库控制器将DRAM分解成各自寻址16个字节的一百万位(1Mb)块。针对每一列提取,可以32个字节来执行到目标存储组的读取或写入请求。

[0021] 针对多主机多HMC系统,使用经包化格式来执行主机装置与HMC装置之间的带内通信。此经包化格式可包含三个主要包分类:请求包、响应包及流控制包。包可经配置为单个16字节流单位或FLIT的倍数。最大包大小含有9个FLIT或144个字节。最小16字节(一个FLIT)包含有包标头及包尾标。用于所有存储器有效负载大小的存储器读取请求包可仅需要包标头、包尾标及相应物理存储器地址。因此,读取请求可使用单个FLIT配置。然而,写入请求及原子请求包还含有分别用于写入及读取-修改-写入操作的所需输入数据。因此,这些请求类型具有2个FLIT到9个FLIT的包宽度。

[0022] HMC系统实施方案可具有多个包重新排序点。目的地是辅助装置的到达包可通过等待本地存储库存取的那些辅助装置。本地存储库还可重新排序已排队包以便最高效地利用去往及来自相应存储库存储组的带宽。然而,存在于给定HMC实施方案中的重新排序点维持包流从特定链路到存储库内的特定存储组的次序。此确保后续接着存储器读取请求的存储器写入请求递送正确且确定性存储器行为。

[0023] 如先前所解释,HMC提供高带宽装置。然而,本发明系统互连规范通常缺少维持系统(其主存储器基于高带宽装置,例如HMC)中的适当平衡必需的充分功能性、灵活性及性能。例如,系统架构可包含用以维持系统单芯片(SoC)与本地存储器之间的每秒数百十亿字节带宽的一或多个HMC,但所述系统架构可提供仅具有此可用带宽的一小部分(例如,十分之一)的系统互连。所述结果可为高度不平衡系统。

[0024] 此外,如果附接到系统互连的处理器提供高速缓冲存储器一致性机构,那么可必须将单独逻辑块添加到所述系统以提供充分带宽及功能性以维持共享存储器及高速缓冲存储器子系统两者中的并发性及正确性。此产生与一组复杂外围逻辑块的处理器互连、存储器互连及系统互连及具有不同程度的等待时间的复杂协议规范。

[0025] 低效高速缓冲存储器一致性尤其在具有分布于多个SoC装置或插座当中的多个执行线程(或任务)的应用中是一担忧。如果核心SoC或处理器支持功能数据高速缓存,那么SoC插座之间的高速缓冲存储器一致性机构必须支持可为比可用于系统互连上的带宽高的数量级的本地存储器带宽。再次,所述结果是高度不平衡系统。

[0026] 用以改善系统层级等待时间的方法是提供具有大体上高带宽存储器装置的带宽的存储器一致性子系统。用以提供此目的的一种方式是使用一或多个HMC装置实施系统存储器一致性。由于HMC具有执行原位原子存储器操作的能力,因此HMC可经延伸以提供高带宽及低等待时间存储器一致性机构。

[0027] 在常规系统中,用于高速缓冲存储器一致性的一种方法是使用基于目录的高速缓冲存储器一致性协议。一致性目录充当筛选器,处理器通过所述筛选器寻求许可以将条目从主要存储器加载到其高速缓冲存储器。当条目改变时,所述目录更新具有所述条目的其它高速缓冲存储器或使所述其它高速缓冲存储器失效。基于目录的高速缓冲存储器一致性包含对将状态指派到高速缓冲存储器线的“目录”的维持。每一高速缓冲存储器线可被指派“家节点(home node)”作为高速缓冲存储器线的状态的最终仲裁者。家节点拥有地址空间

且准许修改高速缓冲存储器线。目录指派状态且基于目录的协议接着使用主机当中的点对点通信来处理元数据。

[0028] 处理器通常存储目录的显著量的高速缓冲存储器线状态信息且依赖于快速高速缓冲存储器存取以读取高速缓冲存储器元数据且确定高速缓冲存储器线的状态。高速缓冲存储器控制器保持跟踪哪一主机签出高速缓冲存储器线。在此配置中,高速缓冲存储器目录本身必须被高速缓存(通常完全相关联地),从而形成维持于最后层级高速缓冲存储器或存储器控制器处的复杂且高耗电数据结构。针对大的对称多处理器(SMP)配置,这些数据结构通常不适合于芯片上且必须由DRAM支持。

[0029] 由于HMC装置的物理存储器存储装置的阶层式性质,因此HMC装置相比于传统分组DRAM装置(例如,DDR3或DDR4装置)具有不同物理寻址与交错模型。HMC装置的物理地址编码到含有存储库、存储组及地址位的34位字段中。具有四链路装置的实施方案利用所述字段的下部32个位,且具有八链路装置的实施方案利用所述字段的下部33个位。

[0030] 与DDR3或DDR4装置相比,HMC装置可廉价地实施抽象化存储器存储接口。例如针对DDR类型装置,当处理器期望存取地址时,真实或实际物理地址被置于总线上且被存取。在HMC协议中,物理地址可在其到达HMC处时重新映射到不同存储器地址。因此,对于HMC,不存在对所写入的实际存储器位置的保证。另外,不需要HMC在设置时间产生响应。

[0031] 可利用HMC的此地址重新映射能力以实施基于目录的存储器一致性。为提供对目录存储器一致性的支持,HMC的逻辑库层可经配置以将给定存储器块的位置抽象化成相关联存储器字或“高速缓冲存储器线”。因此,可将存储器一致性目录存储于HMC存储装置存储库内。此提供使用相同HMC协议管理传统存储器业务及高速缓冲存储器一致性业务两者的方法;借此提供本地存储器数据存取及系统存储器数据存取的统一性。

[0032] 图4展示一种一致性目录存储器格式的实例。可在包含于库物理地址字段中的物理地址方面描述HMC的直接存储器空间。在所展示的实例中,高达34个位的物理地址存储是可用的。库物理地址用作存储器操作的目标且可包含存储库、存储组及DRAM地址位。地址扩展立方体ID(CUB)字段指示支持存储器字的存储器驻存于其中的HMC装置。目录位字段提供存储器一致性位以如同存储器字是高速缓冲存储器线一般地维持存储器字的状态信息。如果存储器一致性目录使用MESI一致性协议,那么目录位包含经修改位、专属位、共享位及无效位。通过存取与库物理地址相关联的存储器字加上CUB偏移而执行将数据字作为高速缓冲存储器线而存取。如果数据条目包含八个字节,那么单个8GB HMC装置可支持超过十亿个并发目录条目。

[0033] 在存储器一致性以此方式组织的情况下,可使用HMC原子存储器操作对存储器执行查找及修改操作。在使用HMC原生的事务协议与目录交互时维持目录可变动性。此使得存储器存取的类型较统一。

[0034] 可使用原子请求执行对目录条目的存取。原子请求涉及:从DRAM读取16个字节的数据(如由请求地址字段存取);通过使用16字节操作数(也包含于请求包中)而对数据执行操作;及接着将结果写回到DRAM中的相同位置。读取-更新-写入序列以原子方式发生,此意指直到原子请求的写入完成为止,无其它请求可存取相同存储组。原子请求类似于16字节写入请求,在于所述原子请求在请求包中具有16字节数据有效负载,且可或可不返回写入响应(取决于所发出的请求是张贴(posted)请求还是非张贴请求)。由原子操作产生的数据

在响应命令中不被返回。可使用原子存储器操作(例如(举例来说)HMC协议的ADD即时请求命令)执行对目录条目的修改。还可使用原子存储器操作(例如(举例来说)其中目标即时值为零的ADD即时请求命令)执行对目录条目的查找。

[0035] 图5展示用HMC装置实施以包含基于HMC的存储器一致性目录的计算机系统500的实例的框图。所述系统包含多个主机处理器。所述处理器可实施为SoC,其中每一处理器由SoC插座502表示。所述实例展示四个处理器,但所述系统灵活地包含其它数目个处理器。

[0036] 所述系统也包含经配置为用于主机处理器的分布式共享存储器的多个HMC装置。所述HMC装置中的一些HMC装置用作端点装置且一些HMC装置用作集线器装置。集线器装置可提供直通通信能力以提供系统内通信。

[0037] 在给出用具有以三维方式布置的存储器裸片的HMC装置实施存储器时,共享存储器经配置以用于3D存储器存取。HMC装置505的存储器裸片的存储器的至少一部分经映射以包含存储器一致性目录的至少一部分。此暗示常规高速缓冲存储器一致性数据或元数据可存储于存储器字驻存于其中的存储器中。HMC装置的逻辑库裸片包含管理至少一个第二装置(例如(举例来说)处理器或另一HMC装置)对HMC装置存储器的三维(3D)存取的至少一个存储器控制器。逻辑库裸片还包含针对存储于多个存储器裸片的存储器中的数据实施存储器一致性协议的逻辑电路。

[0038] 由于一致性目录存储于HMC装置中,因此HMC装置本身可作用于其自身的数据的家节点。存储器一致性状态信息可与所述数据存储在一起。逻辑库层在不将所述数据存储于单独高速缓冲存储器中的情况下允许存储器一致性状态信息在处理器当中传递。存储于共享存储器中的数据基本上是高速缓冲存储器线。另外,在给出由HMC装置实施存储器一致性的情况下,与维持存储器一致性状态信息相关联的处理开销被从处理器移除。

[0039] 如先前所解释,针对多主机多HMC系统,使用经包化协议来执行主机装置与HMC装置之间的I/O通信。存储器一致性协议可包含于此I/O协议中。经包化信息可包含存储器一致性状态信息。在一些实例中,存储器一致性状态信息可如图4的实例中一般与地址信息包含在一起。将存储器一致性协议与I/O协议包含在一起将两种功能组合到统一高效通信系统中。此使本地存储器请求及系统存储器请求两者中的等待时间最小化且消除本地接口与系统接口之间的协议翻译。

[0040] 状态信息确定处理器是否可安全地将数据写入到存储器字、存储器字当前是否保持有效数据且确定可如何将存储器字转换为其中可能进行对存储器字的读取或写入的状态。在一些实例中,HMC装置的逻辑电路实施MSI高速缓冲存储器一致性协议且确定存储器裸片的存储器字的“经修改”状态信息、“共享”状态信息及“无效”状态信息且与存储器裸片的存储器字相关联地存储状态信息。在特定变化中,经修改状态信息、共享状态信息及无效状态信息分别包含经修改位、共享位及无效位。在特定实例中,HMC装置的逻辑电路实施确定专属权状态信息的MESI高速缓冲存储器。在特定实例中,HMC装置的逻辑电路实施确定存储器裸片的存储器字的所有权状态信息的MOSI高速缓冲存储器。逻辑电路与共享存储器数据字相关联地存储专属权状态信息及所有权状态信息中的一者或两者。在一些实例中,集线器HMC装置将路由信息添加到经包化消息。路由信息可用于(例如,通过接收HMC装置)确定存储器一致性状态信息。

[0041] 根据一些实例,“签出/签入”机制由希望更新特定存储器字的处理器使用。在此情

形中,处理器请求对存储器位置X的专属存取。一旦所述存取由存储器字驻存于其中的HMC装置授予,存储器字便可标记为“经签出”(在装置上的小相关联表中或作为存储器中的状态的额外位)。当数据由处理器往回签入时,HMC装置推定所述数据被更新。因此,存储器字本身将由处理器写回(或“清空”)且存储器中的空间可经重新使用于存储额外状态信息(例如,指示签出存储器字的装置的状态信息)。对存储器字X的进一步请求可从组合两条信息(例如,高速缓冲存储器线由处理器P签出的状态信息)的HMC装置接收响应。基于目录的一致性协议接着通过使用存储器一致性状态更新(例如,可要求处理器P清空存储器内容等)而继续进行。当处理器P将新数据清空回到家节点(存储器字的HMC装置)时,可签入存储器字且清除状态信息。

[0042] 在一些实例中,HMC装置响应于对目标存储器字的特定状态请求。例如,目标HMC装置可从处理器接收对对存储器字的专属存取的请求(例如,对专属状态的请求)。HMC装置的逻辑电路可起始对存储器字的状态信息的读取。如果状态信息指示第二装置已具有对存储器字的专属存取及所有权,那么逻辑电路可使所述存储器字无效且等待来自拥有的装置的无效确认。当接收到确认时,HMC装置可将专属存取赋予原始请求装置。

[0043] 图6展示操作计算机系统(例如图5中所展示的实例性系统)的方法600的实例的流程图。在605处,例如通过HMC协议而管理对所述计算机系统的分布式共享存储器的存取。所述共享存储器经配置以用于三维存取,且包含一或多个HMC装置。

[0044] 在610处,将存储器一致性状态信息作为存储器一致性目录存储于分布式共享存储器中。在一些实例中,存储器一致性目录分布于计算机系统的多个HMC装置上。

[0045] 在615处,在计算系统的多个处理器与共享存储器之间传递经包化信息,且在620处,将高速缓冲存储器一致性状态信息与所传递经包化信息包含在一起。在一些实例中,在620处在计算机系统的处理器之间点对点地传递高速缓冲存储器一致性状态信息。

[0046] 本文中所描述的实例性系统及方法提供优于常规多处理器分布式共享存储器系统的数个优点。由HMC装置提供的高带宽产生高带宽存储器及高带宽系统互连。HMC装置在系统架构中的分布使本地存储器之间以及与HMC装置互连的其它系统层级存储器之间的带宽及系统等待时间特性平衡。将存储器一致性协议并入到系统I/O协议中提供维持本地存储器与系统层级存储器之间的存储器请求的原子性的一个统一协议。所述统一协议通过减少或消除本地存储器存取与系统层级存储器存取之间的协议翻译且通过使用用于本地存储器存取及系统层级存储器存取两者的高效HMC包规范而使等待时间最小化。通过卸载存储器一致性目录的维持且通过由SoC简化一致性目录查找而简化SoC设计。

[0047] 用于高速缓冲存储器一致性的另一方法是使用窥探协议。在窥探协议中,计算机系统的处理器监视使用可提供处理器当中的点对点通信的存储器一致性接口发射的存储器业务。与基于目录的方法相比,处理器从经由一致性接口的通信跟踪状态信息而非通过存取目录来确定状态信息。针对MSI窥探协议,存储器一致性接口广播特定高速缓冲存储器线是否处于经修改、共享或无效状态中,从而允许高速缓冲存储器及处理器相应地更新共享副本的状态。例如,当修改高速缓冲存储器线时,必须使系统中的高速缓冲存储器线的其它副本无效,使得可维持一致数据。

[0048] 计算机系统的HMC装置可经配置以从处理器将窥探功能卸载(例如通过将存储器一致性广播协议并入到系统的I/O协议中)。保持跟踪存储器字的状态(或窥探)的HMC装置

可响应于对共享存储器字的请求,犹如任何常规高速缓冲存储器响应于对共享高速缓冲存储器线的请求一样。

[0049] 例如,HMC装置的存储库的逻辑库层可包含确定驻存于其装置中的存储器字的状态信息的逻辑电路。当对存储器字进行读取存取时,逻辑库层的逻辑电路可将所述存储器字的状态信息设置为“专属的”以将专属存取赋予请求者。可借助对读取存取的经包化响应而返回状态信息。其它HMC装置可跟踪包含于经包化通信中的状态信息(例如,通过使用经扩展HMC协议)。计算机系统中的一些HMC装置(例如,集线器HMC装置)包含直通链路以促进包路由。直通链路可用于跟踪在响应包中传递的状态信息。端点(例如HMC主机链路或处理器)可通过将非修改请求发射到存储器字而从存储器字的拥有者查询状态信息。可以对非修改请求的经包化响应通信来返回状态信息且其它HMC装置可从响应包检测状态信息。

[0050] 图7展示操作计算机系统的方法700的另一实例的流程图。所述计算机系统包含经配置为共享存储器的多个HMC装置。在705处,管理对所述计算机系统的分布式共享存储器的存取(例如,使用经扩展HMC协议)。在710处,分布式共享存储器确定由存储器存取产生的存储器一致性状态信息。在一些实例中,共享存储器分布于多个HMC装置上且使用HMC装置来确定存储器一致性状态信息。

[0051] 在715处,在计算系统的多个处理器及共享存储器端点当中传递经包化信息。将存储器一致性状态信息与所传递经包化信息包含在一起。在一些实例中,使用经扩展HMC协议来传递经包化信息。在720处,使用共享存储器端点来跟踪或窥探所广播存储器一致性状态信息。

[0052] 经窥探无效及写回将使数据的一致副本能够维持于存储器中且可产生系统协议的总体简化。将协议并入到存储器接口的经包化通信中可增加响应代理(或响应端点)的数目,但HMC装置的高带宽还可能实现较快且较具可扩缩性响应。

[0053] 虽然本文中已图解说明且描述了特定实施例,但所属领域的技术人员将了解,实现相同目的、结构或功能的任何布置可替代所展示的特定实施例。

[0054] 举例来说,以上描述特定来说是指HMC装置。然而,HMC装置仅是实施计算机系统以利用所描述实施例的一个说明性实例。所属领域的技术人员将理解,在阅读本说明书后,所述实施例可应用于利用其它装置实施的计算机系统。例如,可利用存储器中处理器(PIM)装置及DRAM中的一者或两者来实施所述实施例。PIM装置是指紧密地耦合到存储器的一或多个处理器。有时,PIM装置的一或多个处理器与存储器装置包含于同一IC中。PIM装置也可包含逻辑电路,所述逻辑电路用以实施用于存储于一或若干PIM装置的存储器中的数据的数据的存储器一致性协议。

[0055] 在本文献中,就像在专利文献中常见的一样,术语“一(a或an)”用来包含一或多个,而不管“至少一个”或“一或多个”的任何其它实例或使用。在本文献中,术语“或(or)”用来指代非排他性或,使得除非另有指示,否则“A或B”包含“A但非B”、“B但非A”以及“A及B”。在所附权利要求书中,术语“包含(including)”及“其中(in which)”用作相应术语“包括(comprising)”及“其中(wherein)”的通俗英语等效物。并且,在所附权利要求书中,术语“包含”及“包括”为开放式的,即,在一技术方案中除在此术语后列示的那些要素外还包含其它要素的系统、装置、物品或过程仍视为属于所述技术方案的范围。此外,在所附权利要求书中,术语“第一”、“第二”及“第三”等仅用作标签,且不打算对其对象强加数值要求。

[0056] 本申请案打算涵盖对本文中所描述的本发明的实例性实施例的任何更改或变化形式。打算本发明仅由权利要求书及其等效物的完全范围限制。

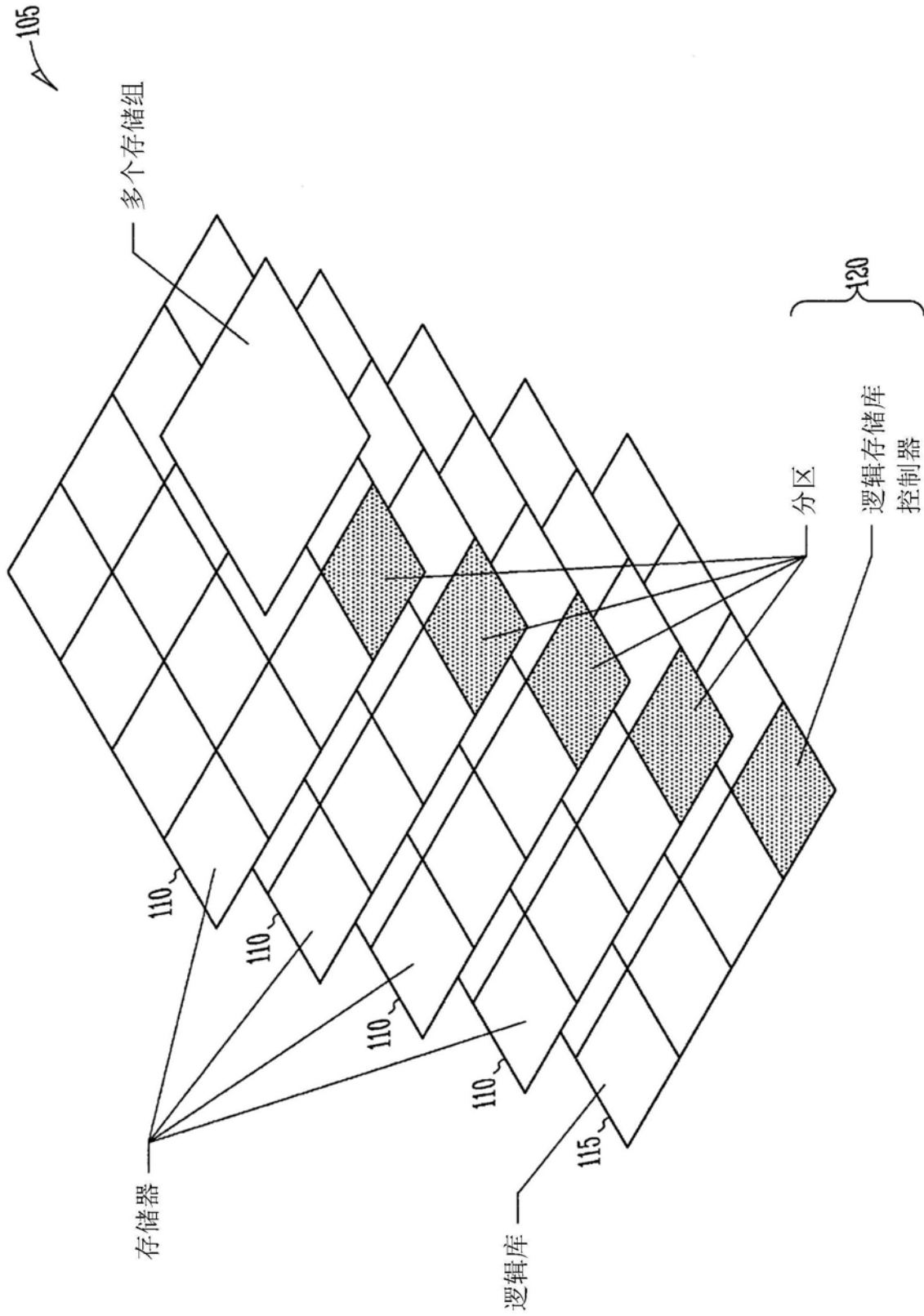


图1

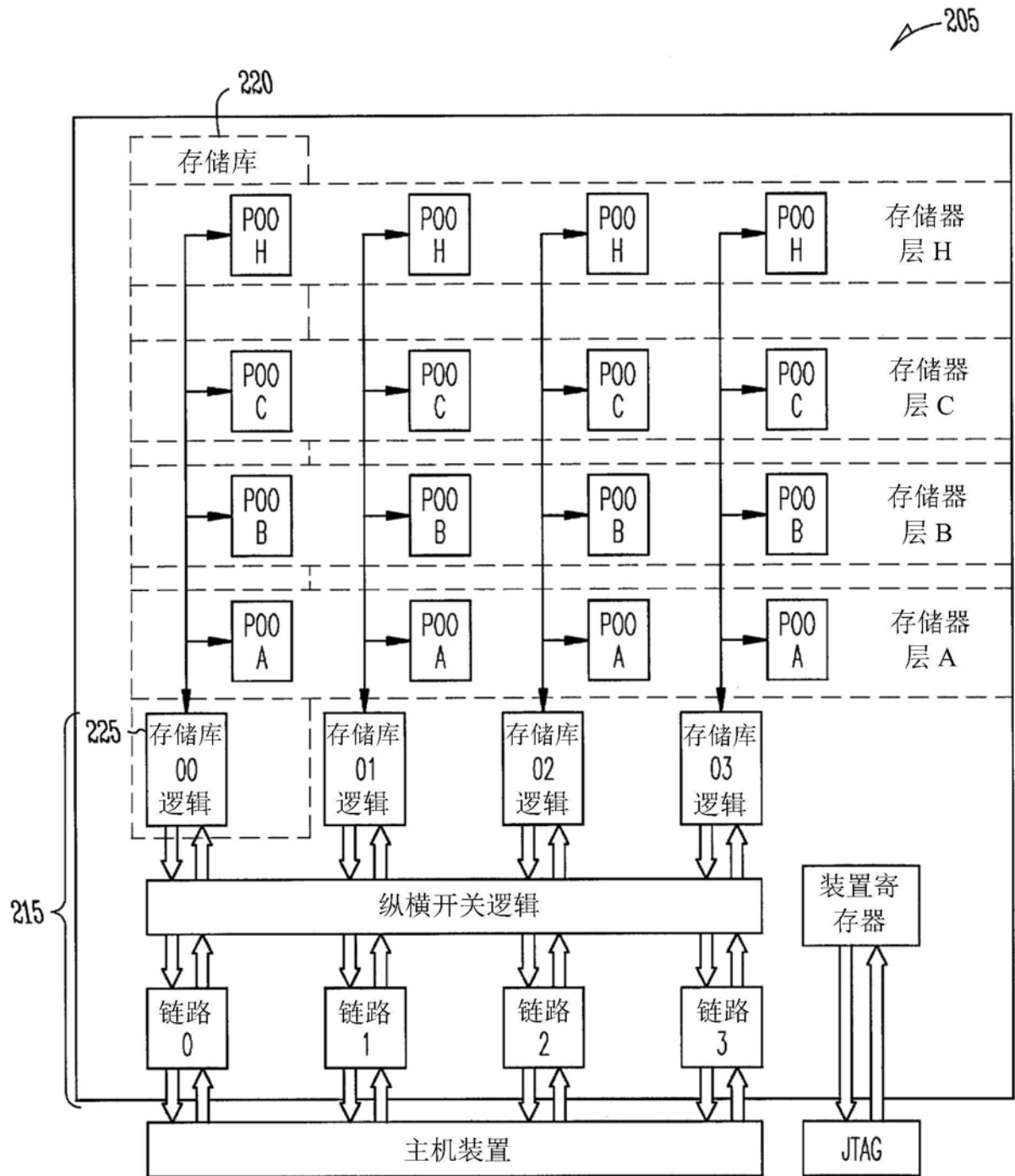


图2

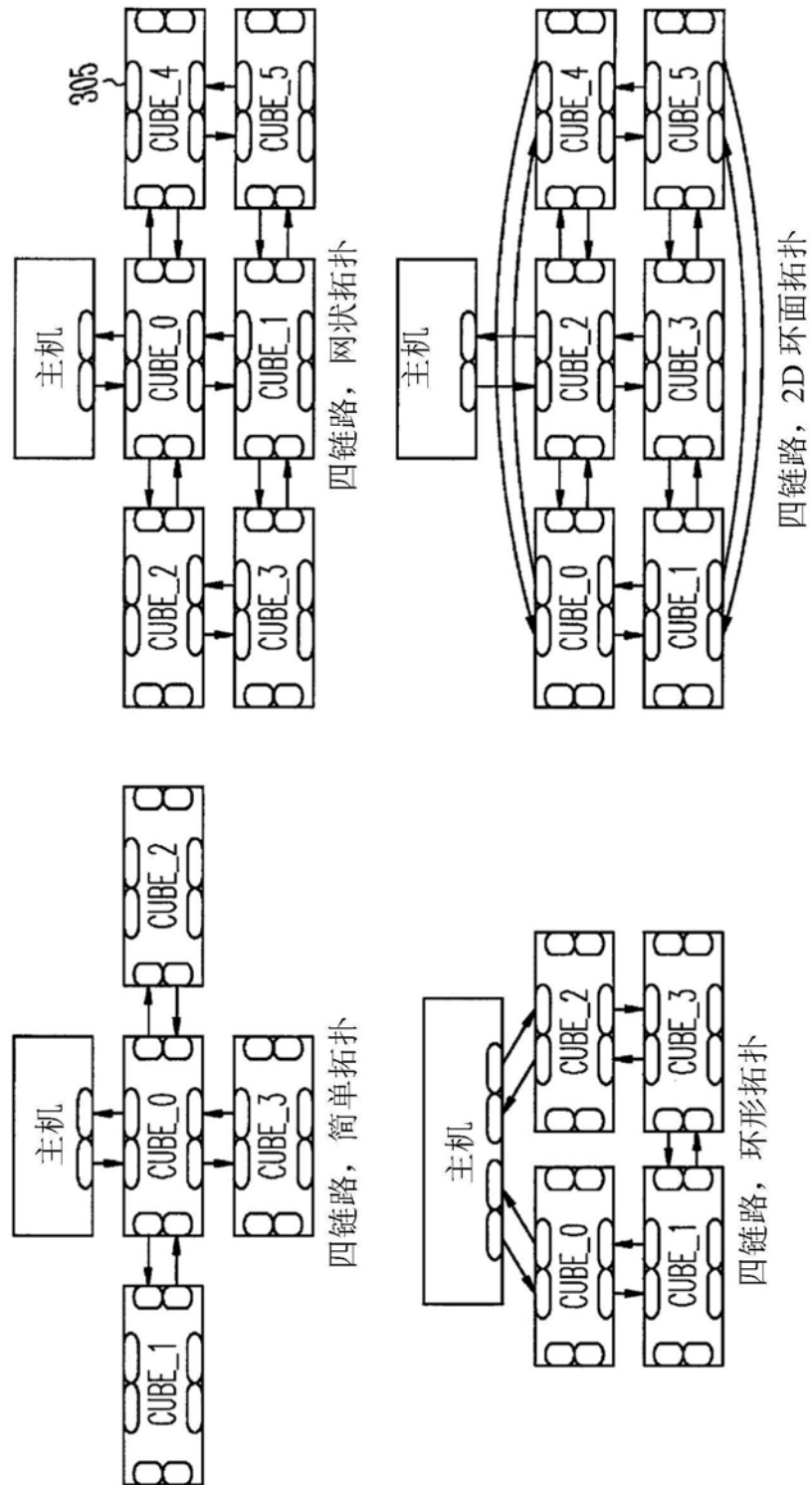


图3

未使用 [63:46]	目录位 [45:42]	CUB [41:34]	库物理 地址 [33:0]
-------------	----------------	-------------	------------------

图4

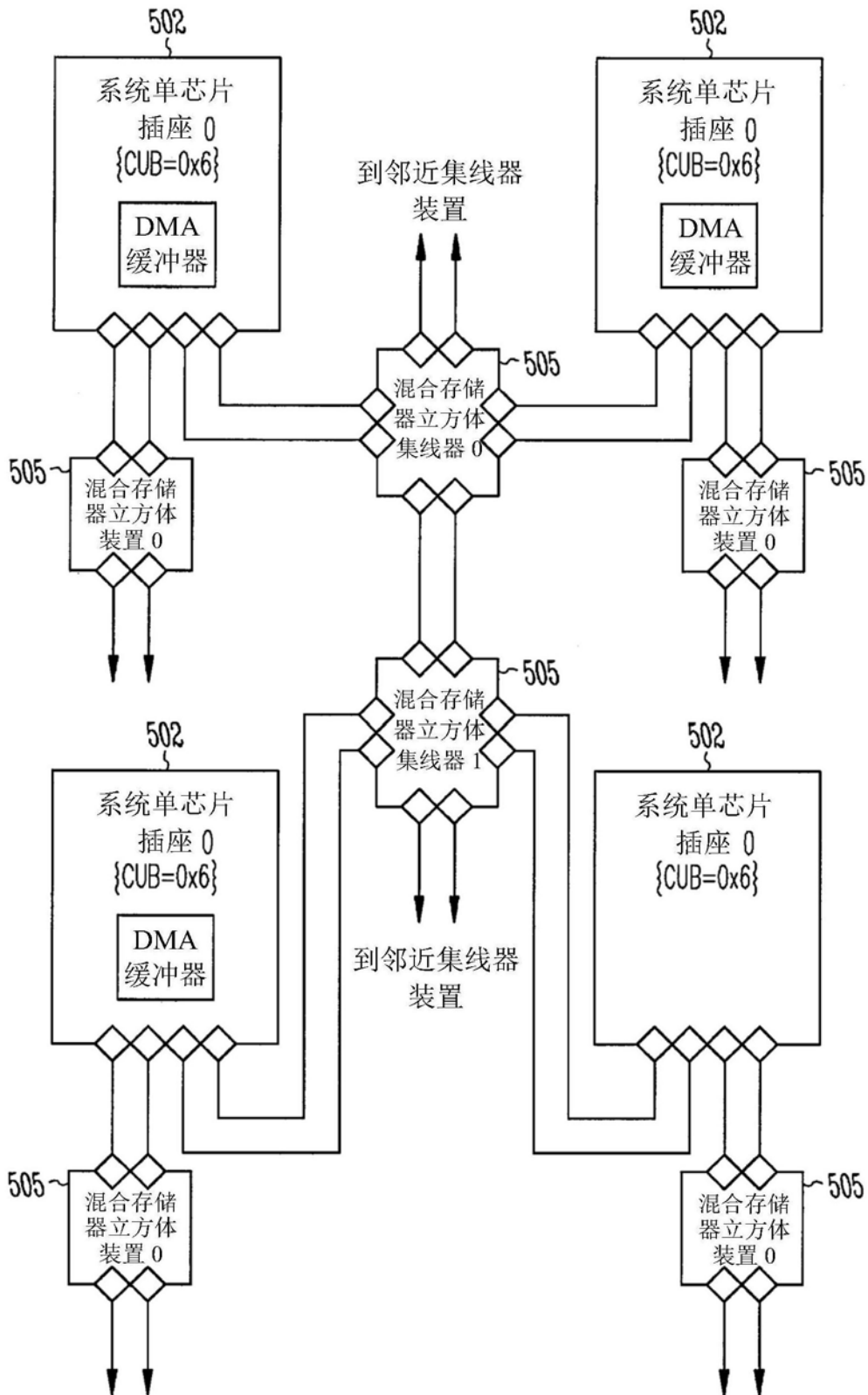


图5

600 ↗

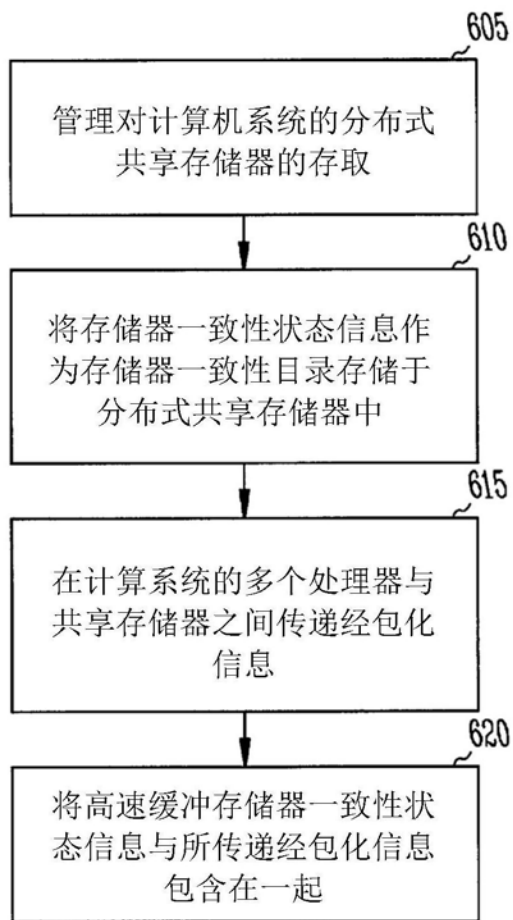


图6

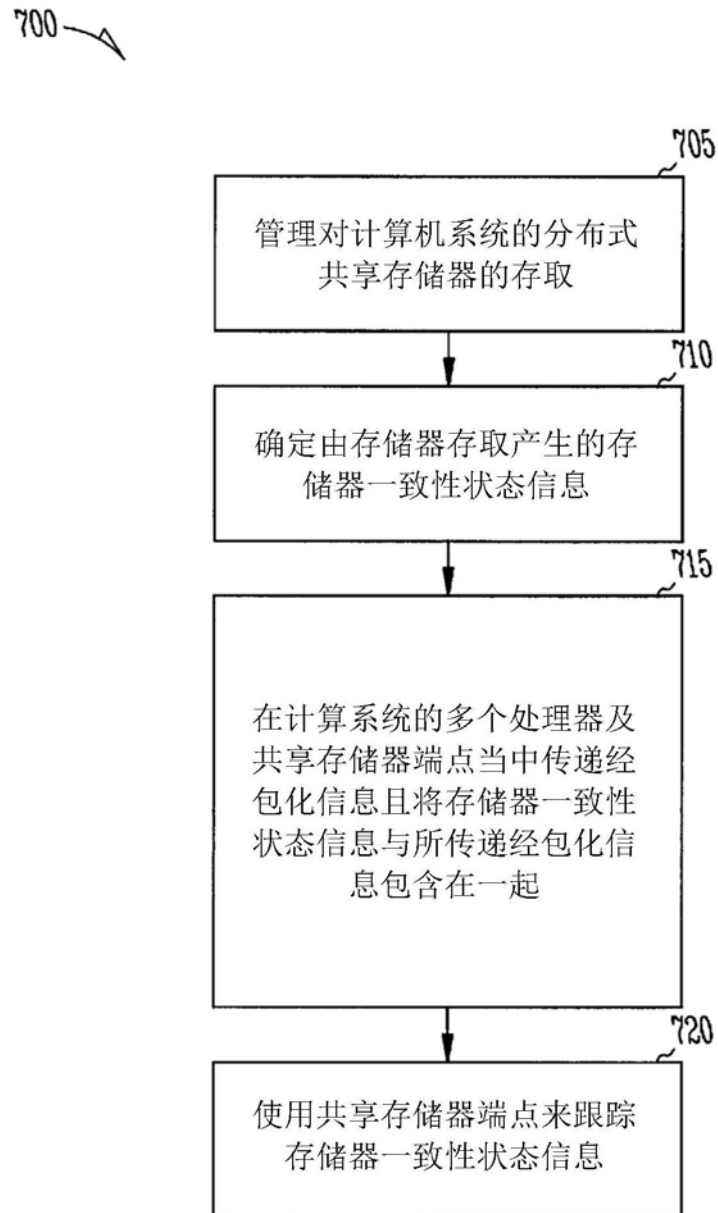


图7