



(12) 发明专利

(10) 授权公告号 CN 113495854 B

(45) 授权公告日 2024.09.17

(21) 申请号 202110345565.3

(22) 申请日 2021.03.31

(65) 同一申请的已公布的文献号
申请公布号 CN 113495854 A

(43) 申请公布日 2021.10.12

(30) 优先权数据
16/839,894 2020.04.03 US

(73) 专利权人 阿里巴巴集团控股有限公司
地址 英属开曼群岛大开曼

(72) 发明人 段立德 牛迪民 刘宏宇 李双辰
郑宏忠

(74) 专利代理机构 北京三友知识产权代理有限公司 11127
专利代理师 刘爱勤 王小东

(51) Int.Cl.

G06F 12/0815 (2016.01)

G06F 12/0897 (2016.01)

(56) 对比文件

US 2017371787 A1, 2017.12.28

审查员 樊宇

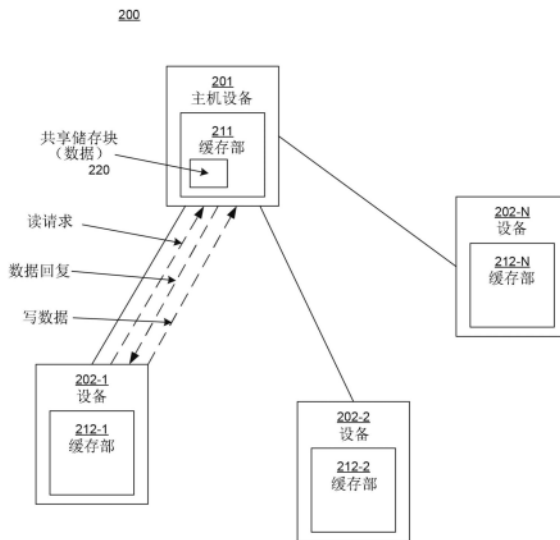
权利要求书4页 说明书7页 附图7页

(54) 发明名称

主机-设备系统中实现或管理缓存一致性的方法和系统

(57) 摘要

本公开提供一种主机-设备系统中实现或管理缓存一致性的方法和系统。缓存一致性模式包括：响应于来自主机-设备系统中的设备的针对共享数据的实例的读请求，将共享数据的实例从主机设备发送至该设备；以及响应于来自设备的写请求，将与该写请求相关联的数据存储在主机设备的缓存部中。共享数据固定在主机设备的缓存部中，而未缓存在主机-设备系统中的任何其它设备中。因为主机-设备系统中只存在共享数据的一个缓存副本，所以该系统中的设备是缓存一致的。



1. 一种在包括多个设备的主机-设备系统中进行缓存一致性管理的方法,所述多个设备包括主机设备,所述方法包括以下步骤:

将共享数据存储在上述主机设备的缓存部中,其中,所述共享数据包括能够由所述多个设备中的至少两个设备访问的数据;以及

当满足第一条件和第二条件两者时,从多个缓存一致性模式中选择第一缓存一致性模式,其中,当所述主机-设备系统中的缓存未命中的数量超过第一阈值时,满足所述第一条件,其中,当与所述共享数据相关联并且由任何两个或更多个设备中的每个设备发出的访问请求的数量在第二阈值与第三阈值之间的范围内时,满足所述第二条件,并且其中,所述第一缓存一致性模式包括:

响应于所述多个设备中的某一设备向所述主机设备发出针对所述共享数据的实例的读请求,将所述共享数据的所述实例从所述主机设备发送至发出所述读请求的所述设备;以及

响应于所述多个设备中的某一设备向所述主机设备发出写请求,将与所述写请求相关联的数据存储在所述主机设备的所述缓存部中。

2. 根据权利要求1所述的方法,其中,所述共享数据未存储在所述多个设备中的除所述主机设备之外的设备上的缓存部中。

3. 根据权利要求1所述的方法,其中,所述多个设备中的设备之间的通信仅通过所述主机设备进行。

4. 根据权利要求1所述的方法,其中,所述多个设备中的各个设备包括各自的处理器和各自的储存器。

5. 根据权利要求1所述的方法,其中,当不满足所述第一条件和所述第二条件两者时,所述方法还包括以下步骤:选择除所述第一缓存一致性模式之外的缓存一致性模式。

6. 根据权利要求5所述的方法,其中,所述访问请求中的各个访问请求包括读请求和写请求两者。

7. 根据权利要求5所述的方法,其中,所述选择除所述第一缓存一致性模式之外的缓存一致性模式的步骤包括:

确定是否满足第三条件,其中,当与所述共享数据相关联并且由所述多个设备中的第一设备发出的访问请求的数量超过第四阈值时,满足所述第三条件;

当满足所述第三条件时,选择第二缓存一致性模式,所述第二缓存一致性模式包括将所述共享数据存储在上述第一设备的缓存部中;以及

当不满足所述第三条件时,选择第三缓存一致性模式,所述第三缓存一致性模式包括将所述共享数据存储在上述多个设备的缓存部中。

8. 根据权利要求7所述的方法,其中,所述第二缓存一致性模式包括:

响应于所述多个设备中的第二设备向所述主机设备发出针对所述共享数据的实例的读请求:从所述主机设备向所述第一设备发送来自所述第二设备的所述读请求,其中,由所述第二设备请求的所述共享数据的所述实例存储在所述第一设备的缓存部中;从所述第一设备向所述主机设备发送由所述第二设备请求的所述共享数据的所述实例;并且从所述主机设备向所述第二设备发送由所述第二设备请求的所述共享数据的所述实例;以及

响应于所述第二设备发出写请求:向所述主机设备发送与来自所述第二设备的所述写

请求相关联的数据;从所述主机设备向所述第一设备发送与来自所述第二设备的所述写请求相关联的数据;并且将与来自所述第二设备的所述写请求相关联的数据存储在所述第一设备的所述缓存部中。

9. 根据权利要求7所述的方法,其中,所述第三缓存一致性模式包括:

由所述第一设备从所述主机设备的所述缓存部访问所述共享数据的实例的版本;

将所述共享数据的所述实例的所述版本存储在所述第一设备的所述缓存部中,其中,所述第一设备的所述缓存部包括脏缓存行,所述脏缓存行包括所述共享数据的所述实例的所述版本;

由所述多个设备中的第二设备向所述主机设备发送缓存读未命中消息;

响应于所述缓存读未命中消息,从所述主机设备向所述多个设备中的其它设备发送读未命中消息;

由所述第一设备监听所述读未命中消息;

响应于所述监听,从所述第一设备向所述主机设备发送所述脏缓存行;

从所述主机设备向所述第二设备发送所述脏缓存行;

将所述脏缓存行存储在所述第二设备的所述缓存部中;

从所述第二设备向所述主机设备发送写无效消息;

从所述主机设备向所述多个设备中的所述其它设备发送所述写无效消息;

响应于所述写无效消息,由所述第一设备使所述第一设备的所述缓存部中的所述脏缓存行无效。

10. 一种在包括多个设备的主机-设备系统中实现缓存一致性的方法,所述多个设备包括主机设备,所述方法包括以下步骤:

将共享数据存储在所述主机设备的缓存部中,其中,所述共享数据包括能够由所述多个设备中的至少两个设备访问的数据,并且其中,所述多个设备中的设备之间的通信通过所述主机设备进行;以及

当满足第一条件和第二条件两者时,执行多个缓存一致性模式中的第一缓存一致性模式,其中,当所述主机-设备系统中的缓存未命中的数量超过第一阈值时,满足所述第一条件,其中,当与所述共享数据相关联并且由任何两个或更多个设备中的每个设备发出的访问请求的数量在第二阈值与第三阈值之间的范围内时,满足所述第二条件,并且其中,所述第一缓存一致性模式包括:

响应于所述多个设备中的某一设备发出针对所述共享数据的实例的读请求,将所述共享数据的所述实例从所述主机设备发送至发出所述读请求的所述设备;以及

响应于所述多个设备中的某一设备发出写请求,将与所述写请求相关联的数据存储在所述主机设备的所述缓存部中。

11. 根据权利要求10所述的方法,其中,所述多个设备中的各个设备包括各自的处理器和各自的储存器。

12. 根据权利要求10所述的方法,其中,所述访问请求中的各个访问请求包括读请求和写请求两者。

13. 根据权利要求10所述的方法,其中,所述多个缓存一致性模式还包括:

第二缓存一致性模式,所述第二缓存一致性模式包括将所述共享数据的实例存储在所

述多个设备中的第三设备的缓存部中,其中,当不满足所述第一条件和所述第二条件并且满足第三条件时,执行所述第二缓存一致性模式;并且其中,当与所述共享数据相关联并且由所述第三设备发出的访问请求的数量超过第四阈值时,满足所述第三条件;以及

第三缓存一致性模式,所述第三缓存一致性模式包括将所述共享数据的实例存储在所述多个设备中的第一设备、第二设备和所述第三设备的缓存部中,其中,当不满足所述第一条件、所述第二条件和所述第三条件时,执行所述第三缓存一致性模式。

14. 根据权利要求10所述的方法,其中,在所述第一缓存一致性模式下,所述共享数据的实例未存储在所述多个设备中的除所述主机设备之外的设备上的缓存部中。

15. 一种用于实现缓存一致性的系统,所述系统包括:

主机设备,所述主机设备包括第一缓存部;以及

多个设备,所述多个设备以通信的方式联接至所述主机设备,所述多个设备中的各个设备包括各自的缓存部;

其中,所述主机设备将共享数据存储在所述缓存部中,并且其中,所述共享数据包括能够由所述多个设备中的至少两个设备访问的数据;

其中,当满足第一条件和第二条件两者时,所述主机设备还能够执行多个缓存一致性模式中的第一缓存一致性模式,其中,当所述系统中的缓存未命中的数量超过第一阈值时,满足所述第一条件;并且其中,当与所述共享数据相关联并且由所述多个设备中的第一设备发出的访问请求的数量以及与所述共享数据相关联并且由所述多个设备中的第二设备发出的访问请求的数量均在第二阈值与第三阈值之间的范围内时,满足所述第二条件,其中,当执行所述第一缓存一致性模式时,所述主机设备:

从所述主机设备向所述多个设备中的发出针对所述共享数据的实例的读请求的设备发送所述共享数据的所述实例;以及

将与来自所述多个设备中的设备的写请求相关联的数据存储在所述第一缓存部中。

16. 根据权利要求15所述的系统,其中,所述多个设备中的设备之间的通信通过所述主机设备进行。

17. 根据权利要求15所述的系统,其中,所述主机设备和所述多个设备中的各个设备包括各自的处理器和各自的存储器。

18. 根据权利要求15所述的系统,其中,所述访问请求中的各个访问请求包括读请求和写请求两者。

19. 根据权利要求15所述的系统,其中,所述主机设备和所述多个设备还能够执行所述多个缓存一致性模式中的第二缓存一致性模式和第三缓存一致性模式;

其中,当不满足所述第一条件和所述第二条件并且满足第三条件时,执行所述第二缓存一致性模式,其中,当与所述共享数据相关联并且由第三设备发出的访问请求的数量超过第四阈值时,满足所述第三条件;

其中,当不满足所述第一条件、所述第二条件和所述第三条件时,执行所述第三缓存一致性模式;

其中,当执行所述第二缓存一致性模式时,所述共享数据的实例存储在所述多个设备中的第三设备的缓存部中;以及

其中,当执行所述第三缓存一致性模式时,所述共享数据的实例存储在所述第一设备、

所述第二设备和所述第三设备的缓存部中。

20. 根据权利要求15所述的系统,其中,所述设备经由网络以通信的方式联接至所述主机设备。

21. 根据权利要求15所述的系统,其中,所述设备经由总线以通信的方式联接至所述主机设备。

22. 根据权利要求15所述的系统,其中,所述共享数据未存储在所述多个设备中的除所述主机设备之外的设备上的缓存部中。

主机-设备系统中实现或管理缓存一致性的方法和系统

技术领域

[0001] 本公开涉及一种在主机-设备系统中进行缓存一致性管理的方法,以及一种在主机-设备系统中实现缓存一致性的方法和系统。

背景技术

[0002] 图1例示了主机-设备系统100,该主机-设备系统100包括可以与N个设备102-1、102-2、...、102-N进行通信的主机设备101。主机设备101和设备102各自分别具有其自己的缓存部(cache)111、112-1、112-2、...、112-N。

[0003] 在系统100中,多个设备可以访问缓存部之一中的同一储存块(memory block)。在这些访问中的各个访问中,设备至少读取一次该储存块,然后至少写入一次该储存块。例如,设备中的所有设备可能递增相同的计数器。而且,一个设备的读访问和写访问可以与其它设备中的一个或更多个其它设备的读访问和写访问同时发生,并且来自不同设备的读访问和写访问可以被任意混合。

[0004] 因为主机设备101和设备102中的每一个在其各自的缓存部111和112中在本地缓存其数据,所以一个缓存部中的数据可以不同于另一个缓存部中的数据,例如,这取决于读和写发生的顺序。

[0005] 因此,需要缓存一致性协议来跨系统100中的设备维持缓存一致性。

发明内容

[0006] 本文公开了一种新颖的缓存一致性协议或模式。

[0007] 在实施方式中,主机-设备系统包括主机设备以及(例如,通过总线和/或在网络中有线地和/或无线地)以通信的方式(communicative)联接至该主机设备的多个其它设备。共享数据(可由主机-设备系统中的设备中的至少两个设备访问的数据)存储在主机设备的缓存部中。主机-设备系统中的设备之间的通信是通过主机设备进行的。在这些实施方式中,缓存一致性模式包括:响应于来自主机-设备系统中的设备的针对共享数据的实例的读请求,将共享数据的实例从主机设备发送至该设备;以及响应于来自设备的写请求,将与该写请求相关联的数据存储在主机设备的缓存部中。在该缓存一致性模式(在本文中可称为第一缓存一致性模式)下,共享数据“固定(pin)”在主机设备的缓存部中,并且未缓存在主机-设备系统中的任何其它设备中。由于主机-设备系统中只存在共享数据的一个缓存副本,因此该系统中的设备是缓存一致的(即,共享数据跨系统中所有缓存部是一致的)。

[0008] 根据本发明的实施方式减少了与读取数据相关联的等待时间;减少了设备之间传输的数据量,从而节省了带宽;减少了与维持缓存一致性相关联的消息的数量,从而也减少了等待时间并节省了带宽;并且不是基于监听(snoop)的(例如,设备确实需要监听由其它设备发送的消息)。

[0009] 本文还公开了一种新颖的缓存一致性管理的方法,其在不同缓存一致性模式之间自适应地且动态地切换。

[0010] 在实施方式中,在类似于上述主机-设备系统的主机-设备系统中,当满足第一条件和第二条件两者时,从多个缓存一致性模式中选择前述的第一缓存一致性模式。例如,当主机-设备系统中存在许多缓存未命中时,或者例如当主机-设备系统中的各个设备经历了许多缓存未命中时,满足第一条件。当与共享数据相关联的读请求和写请求(访问请求)的数量在主机-设备系统中的不同(例如,至少两个)设备中大约相同时,满足第二条件。

[0011] 如果不满足第一条件和第二条件两者,则选择不同的缓存一致性模式。例如,在实施方式中,当不满足第一条件和第二条件但是满足第三条件时,则选择第二缓存一致性模式。例如,当主机-设备系统中的设备之一主导(dominate)访问请求的数量时,满足第三条件。在第二缓存一致性模式下,共享数据存储(固定)在主导设备的缓存部中,而未缓存在系统中的任何其它设备中。如果不满足第一条件、第二条件和第三条件,则选择第三缓存一致性模式(例如,基于监听的协议)。

[0012] 本领域普通技术人员在阅读了各个附图中例示出的实施方式的以下详细描述之后,将认识到本发明的各个实施方式的上述以及其它目的和优点。

附图说明

[0013] 并入本说明书并形成本说明书的一部分的附图例示了本公开的实施方式,在附图中,相同的附图标记表示相同的元素,并且附图与详细描述一起用于解释本公开的原理。

[0014] 图1是例示了传统主机-设备系统的框图。

[0015] 图2例示了根据本发明的实施方式中的主机-设备系统以及可以在主机-设备系统中执行的第一缓存一致性模式。

[0016] 图3是例示了根据本发明的实施方式中的第一缓存一致性模式下的操作的流程图。

[0017] 图4例示了根据本发明的实施方式中的可以在主机-设备系统中执行的第二缓存一致性模式。

[0018] 图5例示了根据本发明的实施方式中的可以在主机-设备系统中执行的第三缓存一致性模式。

[0019] 图6是例示了根据本发明的实施方式中的缓存一致性管理方法中的操作的流程图。

[0020] 图7是例示了可以在其上实现根据本发明的实施方式的示例计算系统或设备的框图。

具体实施方式

[0021] 现在将详细参照本公开的各种实施方式,其示例在附图中例示出。尽管结合这些实施方式进行了描述,但是将理解,并不旨在将本公开限制于这些实施方式。相反,本公开旨在涵盖另选例、修改例和等同例,所述另选例、修改例和等同例可以包括在由所附权利要求书限定的本公开的精神和范围内。此外,在本公开的以下详细描述中,阐述了许多具体细节以提供对本公开的透彻理解。然而,将理解,可以在不具有这些具体细节的情况下实践本公开。在其它情况下,未详细描述公知方法、过程、部件和电路,以免不必要地模糊本公开的各方面。

[0022] 下面的详细描述的一些部分以过程、逻辑块、处理以及与计算机储存器内的数据位有关的操作的其它符号表示的方式呈现。这些描述和表示是数据处理领域的技术人员用来将其工作的实质最有效地传达给本领域其它技术人员的手段。在本申请中,过程、逻辑块、处理等被认为是导致期望结果的步骤或指令的自洽序列。这些步骤是利用物理量的物理操纵的步骤。通常,尽管不是必须的,但是这些量采取能够在计算机系统中存储、传输、组合、比较和以其它方式操纵的电或磁信号的形式。主要出于通用原因,已证明有时将这些信号称为事务、位、值、元素、符号、字符、样本、像素等是方便的。

[0023] 然而,应牢记,所有这些术语和类似术语将与适当物理量相关联,并且仅仅是应用于这些量的方便标签。除非根据下面的讨论明显地另外明确指出,否则应理解,在整个本公开中,利用诸如“接收”、“发送”、“访问”、“确定”、“使用”、“存储”、“选择”、“发出”、“缓存”、“读”、“写”、“执行”、“监听”、“无效”等的术语的讨论是指装置或计算机系统或类似电子计算设备或处理器(例如,图7的系统700)的动作和处理(例如,分别是图3的流程图300和图6的流程图600)。计算机系统或类似电子计算设备在储存器、寄存器或其它这种信息存储、传输或显示设备内操纵和转换表示为物理(电子)量的数据。

[0024] 可以在驻留于由一个或更多个计算机或其它设备执行的某种形式的计算机可读存储(storage)介质(诸如程序模块)上的计算机可执行指令的总体上下文中讨论本文描述的实施方式。通过示例的方式但非限制性地,计算机可读存储介质可以包括非暂时性计算机存储介质和通信介质。通常,程序模块包括执行特定任务或实现特定抽象数据类型的例程、程序、对象、部件、数据结构等。在各种实施方式中,可以根据需要组合或分布程序模块的功能。

[0025] 计算机存储介质包括以用于存储诸如计算机可读指令、数据结构、程序模块或其它数据的信息的任何方法或技术实现的易失性和非易失性、可移除和不可移除介质。计算机存储介质包括但不限于随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程ROM(EEPROM)、闪存存储器(例如,SSD)或其它存储器技术、光盘ROM(CD-ROM)、数字通用盘(DVD)或其它光学存储部、磁带盒、磁带、磁盘存储部或其它磁存储设备或可以用于存储所需信息并且可以被访问以检索该信息的任何其它介质。

[0026] 通信介质可以体现计算机可执行指令、数据结构和程序模块,并且包括任何信息传递介质。通过示例的方式但非限制性地,通信介质包括诸如有线网络或直接有线连接的有线介质,以及诸如声学、射频(RF)、红外和其它无线介质的无线介质。上述项中的任何项的组合也可以包括在计算机可读介质的范围内。

[0027] 图2例示了根据本发明的实施方式中的包括主机设备201和N个设备202-1、202-2、...、202-N(N是整数)的主机-设备系统200。设备202-1、202-2、...、202-N在本文中可以单独地称为设备202或者统称为设备202。主机设备201具有其自己的缓存部211,并且设备202分别各自具有其自己的缓存部212-1、212-2、...、212-N。缓存部212-1、212-2、...、212-N在本文中可以单独地称为缓存部212或者可以统称为缓存部212。

[0028] 设备202(例如通过总线地和/或通过网络有线地和/或无线地)以通信的方式联接至主机设备201。主机-设备系统200中的设备202之间的通信通过主机设备201进行。在实施方式中,主机-设备系统200中的设备202之间的通信仅通过主机设备201进行;设备202不会彼此直接进行通信。

[0029] 在之后的讨论中,共享储存块220位于主机设备201的缓存部211中;然而,共享储存块220可以位于系统200中的任何设备中。主机设备201还可以访问(从共享储存块220读取和写入到)共享储存块220。

[0030] 主机-设备系统200中的多个设备可以访问同一(共享)储存块220。共享储存块220中的数据在本文中称为共享数据或共享数据的实例。在设备对共享储存块220的每次访问中,设备至少一次读取共享储存块220,并且至少一次写入该储存块。而且,一个设备对共享储存块220的读访问和写访问可以与其它设备中的一个或更多个其它设备对共享储存块的读访问和写访问同时发生,并且不同设备的读访问和写访问可以任意混合。对共享储存块220的读访问和写访问可以是原子事务。

[0031] 对共享储存块220的访问可以是真共享(true sharing)或伪共享(false sharing)。本文如本领域普通技术人员理解的那样使用术语“真共享”和“伪共享”。例如,在真共享中,不同设备访问缓存行中的数据的实例;而在伪共享中,不同设备可以访问数据的不同实例,但是数据的不同实例位于同一缓存行中。

[0032] 在根据本发明的实施方式中,系统200执行在本文中称为第一缓存一致性模式的缓存一致性模式。系统200还可以执行将在下面进一步描述的其它缓存一致性模式。在第一缓存一致性模式下,共享数据的实例未存储在除主机设备201之外的设备202上的缓存部中。结合图3进一步描述第一缓存一致性模式。

[0033] 图3是例示了根据本发明的实施方式中的用于缓存一致性的方法中的操作的流程图300。具体地,流程图300例示了前述的第一缓存一致性模式。

[0034] 在框302中,还参照图2,设备202向主机设备201发送针对共享数据的实例的读请求。

[0035] 在框304中,响应于发出读请求的设备202,主机设备201向所述设备202发送共享数据的实例。

[0036] 在框306中,设备202向主机设备201发送具有要写入的数据的写请求。数据可以是在框304中发送的共享数据的更新版本或改变版本。

[0037] 在框308中,主机设备201将与写请求相关联的数据存储在缓存部211中。

[0038] 针对设备202对共享数据的其它访问,重复刚刚描述的处理。

[0039] 在第一缓存一致性模式下,共享数据“固定”在主机设备201的缓存部211中,并且未缓存在设备202中。术语“固定”在本文中用于表示在第一缓存一致性模式下,主机-设备系统200中的设备识别出共享数据的唯一正确版本驻留在主机设备201的缓存部211中。由于主机-设备系统200中只存在共享数据的一个缓存副本,所以设备201和设备202是缓存一致的(即,共享数据跨系统200中的所有缓存部211和212是一致的)。

[0040] 如上所述,主机-设备系统200可以执行除第一缓存一致性模式之外的其它缓存一致性模式。图4例示了根据本发明的实施方式中的可以由主机-设备系统200执行的第二缓存一致性模式。在第二缓存一致性模式下,共享数据存储(固定)在设备202之一中的缓存部212中,并且未缓存在主机-设备系统200中的任何其它设备中。在实施方式中,选择发出了最多访问请求的设备作为存储共享数据的设备。

[0041] 在图4的示例中,共享数据存储(固定)在设备202-2的缓存部212-2中,并且设备202-1将访问共享数据的实例。为此,设备202-1向主机设备201发送读请求,该主机设备201将读请

求转发至设备202-2。设备202-2向主机设备201发送共享数据的所请求的实例,该主机设备201将该数据转发至设备202-1。然后,设备202-1向主机设备201发送要写入的数据,该主机设备201将该数据转发至设备202-2。设备202-2将数据存储在缓存部212-2中。

[0042] 针对设备202对共享数据的其它访问,重复刚刚描述的处理。

[0043] 在第二缓存一致性模式下,主机-设备系统200中仅存在共享数据的一个缓存副本,因此设备201和设备202是缓存一致的(即,共享数据跨系统200中的所有缓存部211和212是一致的)。相对于第一缓存一致性模式,第二缓存一致性模式:具有更长的读取等待时间;传输更多的数据并因此使用更多的带宽;并且需要在系统中的设备之间交换更多缓存一致性相关消息,这也增加了等待时间并消耗了带宽。下面进一步提供附加信息(参见表1)。

[0044] 图5例示了在根据本发明的实施方式中可以由主机-设备系统200执行的第三缓存一致性模式。在第三缓存一致性模式下,共享数据在本地存储在系统200中的设备201和设备202上。第三缓存一致性模式可以称为基于监听的一致性协议。

[0045] 在图5的示例中,设备202-2刚刚完成了对保存共享数据的共享储存块的访问,因此缓存部212-2具有包括数据的共享实例的当前版本的脏缓存行(dirty cache line)。在该示例中,设备202-1将访问共享储存块。首先,设备202-1向主机设备201发送缓存读未命中消息。响应于读未命中消息,主机设备201向系统200中的其它设备202-2、...、202-N广播读未命中消息。设备202-2监听到读未命中消息,并且作为响应向主机设备201发送脏缓存行。主机设备201向设备202-1发送脏缓存行,因此此时,设备202-1和设备202-2两者具有共享数据的实例的当前版本。设备202-1可以更新共享数据的实例,因此缓存部212-1现在具有包括共享数据的实例的最新版本的脏缓存行。然后,设备202-1向主机设备201发送写无效消息,并且主机设备向系统200中的其它设备202-2、...、202-N广播该消息。设备202-2监听到写无效消息,并且使包括数据的共享实例的缓存部212-2中的缓存行无效。

[0046] 针对设备202对共享数据的其它访问,重复刚刚描述的处理。

[0047] 相对于第一缓存一致性模式,第三缓存一致性模式:具有更长的读等待时间;传输更多的数据并因此使用更多的带宽;需要在系统中的设备之间交换更多缓存一致性相关消息,这也增加了等待时间并消耗了带宽;并且是基于监听的(例如,设备需要监听由其它设备发送的消息)。下面在表1中提供附加信息,表1提供了第一缓存一致性模式、第二缓存一致性模式和第三缓存一致性模式的性能指标的比较。

[0048] 表1性能指标的比较

模式	数据读等待时间	总数据传输量	一致性消息	设备监听?
第三	4个跃程(hop)(2跃程读未命中和2跃程数据回复)	2*(缓存行大小)* (设备访问)	(N个读未命中)+(N个写未命中)	是
第二	固定设备的本地读/写命中;所有其它设	4*(读/写数据量)* (设备访问)*	(2个读请求)+(2个写请求)	否

[0049]

	备的 4 个跃程	(N-1/N)	
[0050]	第一 2 个跃程	2*(读/写数据量)	(1 个读请求) + 否 * (设备访问) (1 个写请求)

[0051] 在表1中,N是系统200中的设备的数量。在第二缓存一致性模式下,因子(N-1/N)考虑了固定了共享数据的设备不传输数据的事实。而且,在第一缓存一致性模式和第二缓存一致性模式下,读/写数据量可以小于或等于缓存行的大小。

[0052] 如表1中的信息所示,第一缓存一致性模式减少了与读数据相关联的等待时间;减少了在设备之间传输的数据量,从而节省了带宽;减少了与维持缓存一致性相关联的消息的数量,因此也减少了等待时间并节省了带宽;并且不是基于监听的(例如,设备确实需要监听由其它设备发送的消息)。在所有性能指标方面,第一缓存一致性模式优于第二缓存一致性模式和第三缓存一致性模式。随着系统200中的设备202的数量N的增加,与第一缓存一致性模式相关联的益处变得更加显著。

[0053] 图6是例示了根据本发明的实施方式中的缓存一致性管理方法中的操作的流程图600。具体地,流程图600例示了在不同缓存一致性模式之间自适应且动态地切换的一致性管理方法的实施方式。还参照图2、图4和图5讨论图6。

[0054] 在图6的框602中,确定是否满足第一条件。在实施方式中,当主机-设备系统200中的缓存读未命中的数量超过第一阈值时,满足第一条件。例如,当主机-设备系统200中的缓存读未命中的总数超过第一阈值时,或者当主机-设备系统中的设备202中的各个设备经历了超过第一阈值的多次缓存读未命中时,可以满足第一条件。可以使用一个或多个计数器来对缓存读未命中的数量进行计数。如果满足第一条件,则流程图600进行至框604;否则,流程图进行至框608。

[0055] 在框604中,确定是否满足第二条件。在实施方式中,当由主机-设备系统200中的设备202中的任何两个或更多个设备发出的针对共享数据的访问请求的数量大约相同时,满足第二条件。例如,如果由设备202-1发出的访问请求的数量和由设备202-2发出的访问请求的数量大约相同,则满足第二条件。在实施方式中,当由主机-设备系统200中的设备202中的任何两个或更多个设备中的每个设备发出的针对共享数据的访问请求的数量在指定范围内时,可以满足第二条件。指定范围可以由第二(例如,较低的)阈值和第三(例如,较高的)阈值界定。可以使用一个或多个计数器来对访问请求的数量进行计数。如果满足第二条件,则流程图600进行至框606;否则,流程图进行至框608。

[0056] 在框606中,由主机-设备系统200选择并执行第一缓存一致性模式。在第一缓存一致性模式下,共享数据存储(固定)在主机设备201的缓存部中,并且未缓存在主机-设备系统200中的任何其它设备中。

[0057] 在框608中,确定是否满足第三条件。在实施方式中,当由设备202中的任何一个设备发出的针对共享数据的访问请求的数量超过第四阈值时,满足第三条件。换句话说,当针对共享数据的访问请求的数量由主机-设备系统200中的单个设备主导时,满足第三条件。可以使用一个或多个计数器来对访问请求的数量进行计数。如果满足第三条件,则流程图600进行至框610;否则,流程图进行至框612。

[0058] 在框610中,由主机-设备系统200选择并执行第二缓存一致性模式。在第二缓存一

致性模式下,共享数据存储(固定)在主导设备的缓存部中,并且未缓存在主机-设备系统200中的任何其它设备中。

[0059] 在框612中,由主机-设备系统200选择并执行第三缓存一致性模式。在第三缓存一致性模式下,共享数据在本地存储在主机-设备系统200中的设备上。

[0060] 在指定的时间段之后,可以将计数器重设为初始值(例如,零),并且可以重复流程图600的操作。

[0061] 图7是例示了可以在其上实现根据本发明的实施方式的示例计算系统或设备700的框图。通常,设备700是在主机-设备系统(诸如上述的主机-设备系统)中使用的设备的类型的示例。可以使用设备700来执行上述方法(例如,图3和图6的方法)。

[0062] 在图7的示例中,设备700可以包括一个或更多个存储单元(计算机存储介质)710以及通过通信接口730以通信的方式联接的处理器720。处理器720可以包括一个或更多个内核(core),并且各个内核可以包括:可以在内核之间分布的本地存储器;一个或更多个缓冲部(Buffer)和/或缓存部;以及一个或更多个执行引擎。本文中可以使用术语“存储器”来指代缓冲部和/或缓存部和/或本地存储器。通信接口730可以包括有线网络通信链路、无线网络通信链路、数据总线等。

[0063] 尽管前述公开使用特定框图、流程图和示例阐述了各种实施方式,但是本文描述和/或例示的各个框图部件、流程图步骤、操作和/或部件可以使用各种硬件配置、软件配置或固件配置(或其任何组合)单独地和/或共同地实现。另外,包含在其它部件内的部件的任何公开都应视为示例,因为可以实现许多其它架构来实现相同功能。

[0064] 本文描述和/或例示的处理参数和步骤次序仅通过示例的方式给出,并且可以根据需要改变。例如,尽管可以以特定顺序示出或讨论本文例示和/或描述的步骤,但是这些步骤不一定需要以所例示或所讨论的顺序执行。本文描述和/或例示的各种示例方法还可以省略本文描述或例示的步骤中的一个或更多个步骤,或者除了所公开的步骤之外还包括附加步骤。

[0065] 尽管本文已在全功能计算系统或设备的上下文中描述和/或例示了各种实施方式,但是这些示例实施方式中的一者或更多者可以以各种形式作为程序产品分布,而与用于实际执行分布的计算机可读介质的特定类型无关。还可以使用执行某些任务的软件模块来实现本文公开的实施方式。这些软件模块可以包括可以存储在计算机可读存储介质上或计算系统中的脚本、批处理或其它可执行文件。这些软件模块可以配置计算系统或设备以执行本文公开的示例实施方式中的一个或更多个示例实施方式。

[0066] 可以在云计算环境中实现软件模块中的一个或更多个软件模块。云计算环境可以经由互联网提供各种服务和应用。这些基于云的服务(例如,软件即服务、平台即服务、基础架构即服务等)可以通过Web浏览器或其它远程接口来访问。可以通过远程桌面环境或任何其它基于云的计算环境来提供本文描述的各种功能。

[0067] 尽管已经以特定于结构特征和/或方法动作的语言描述了主题,但是将理解,本公开中限定的主题不必限于上述特定特征或动作。相反,上文描述的特定特征和动作作为实现本公开的示例形式被公开。

[0068] 因此描述了根据本发明的实施方式。尽管已经在特定实施方式中描述了本公开,但是应理解,本发明不应被解释为受这些实施方式的限制,而应根据所附权利要求书来解释。

100

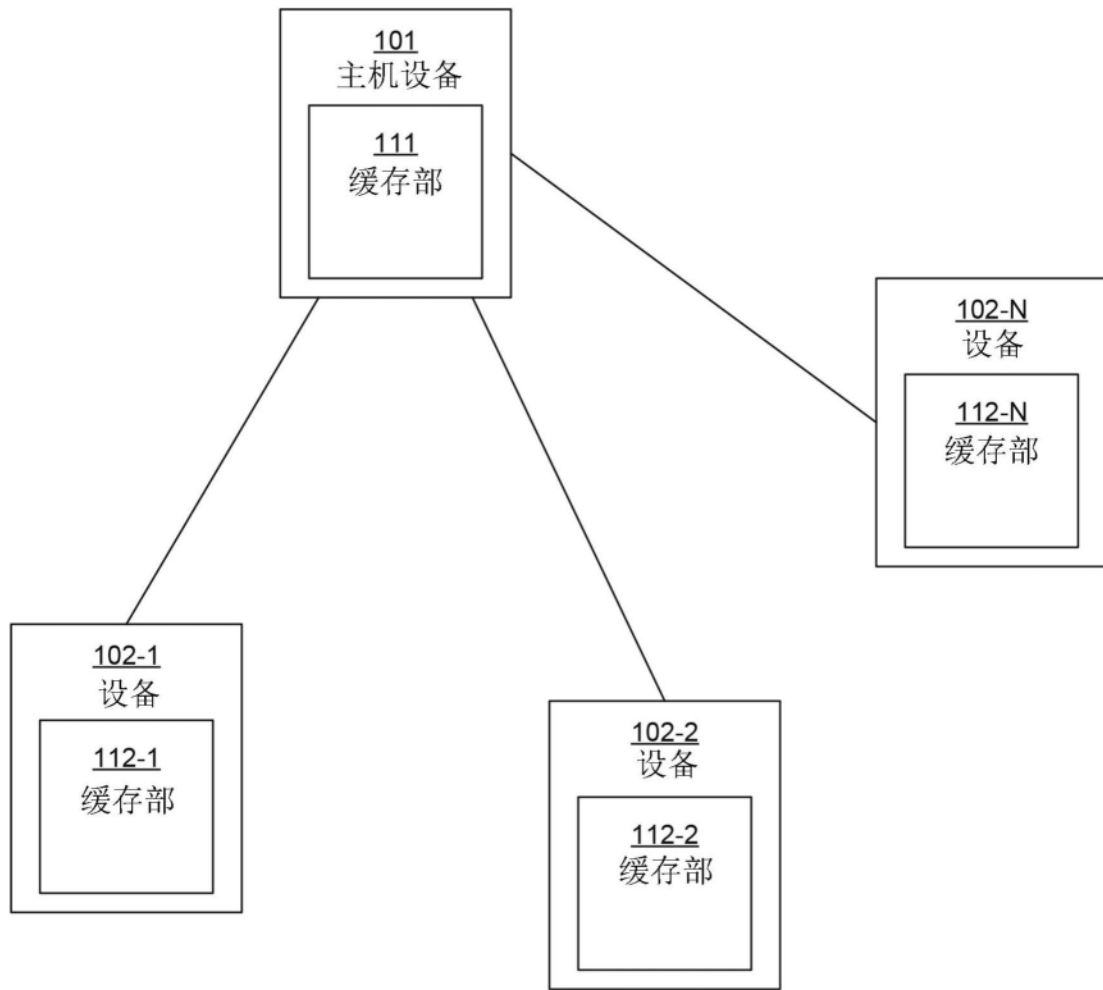


图1

200

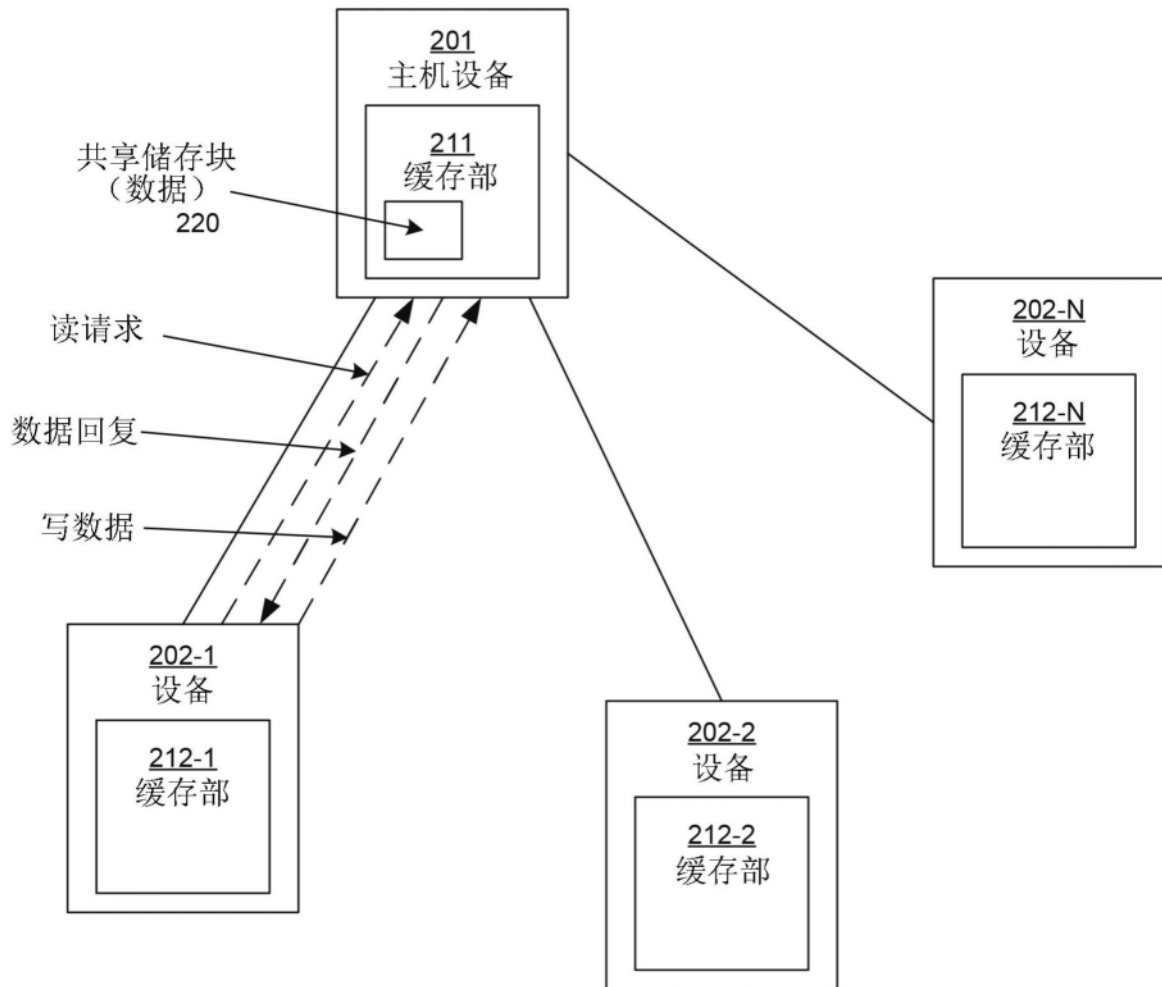


图2

300

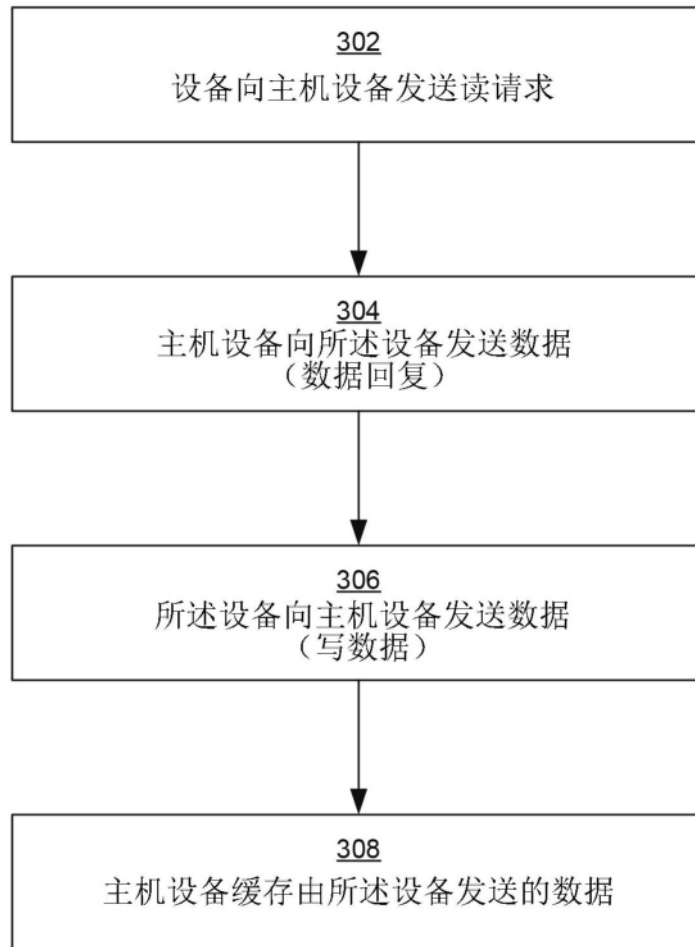


图3

200

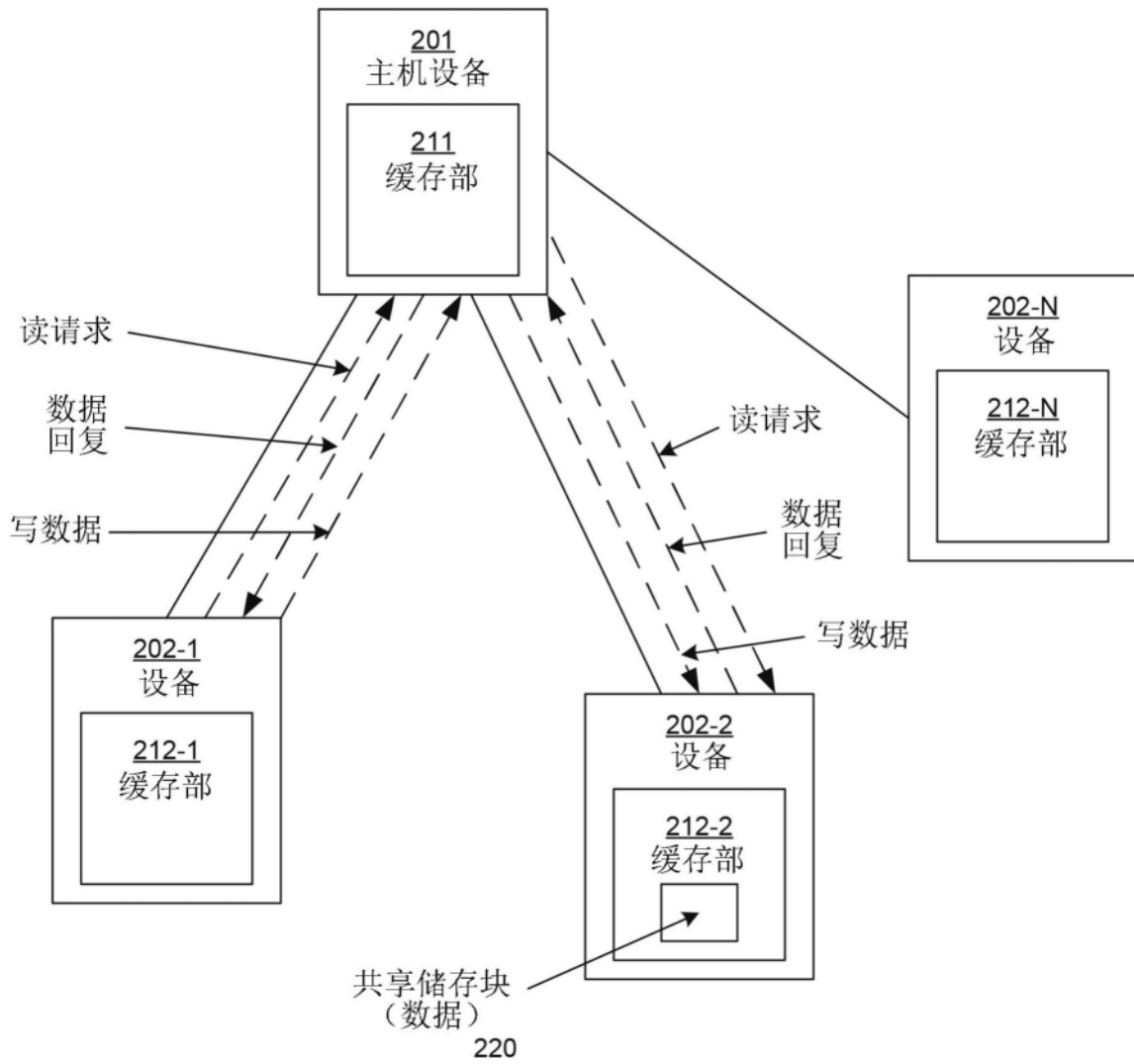


图4

200

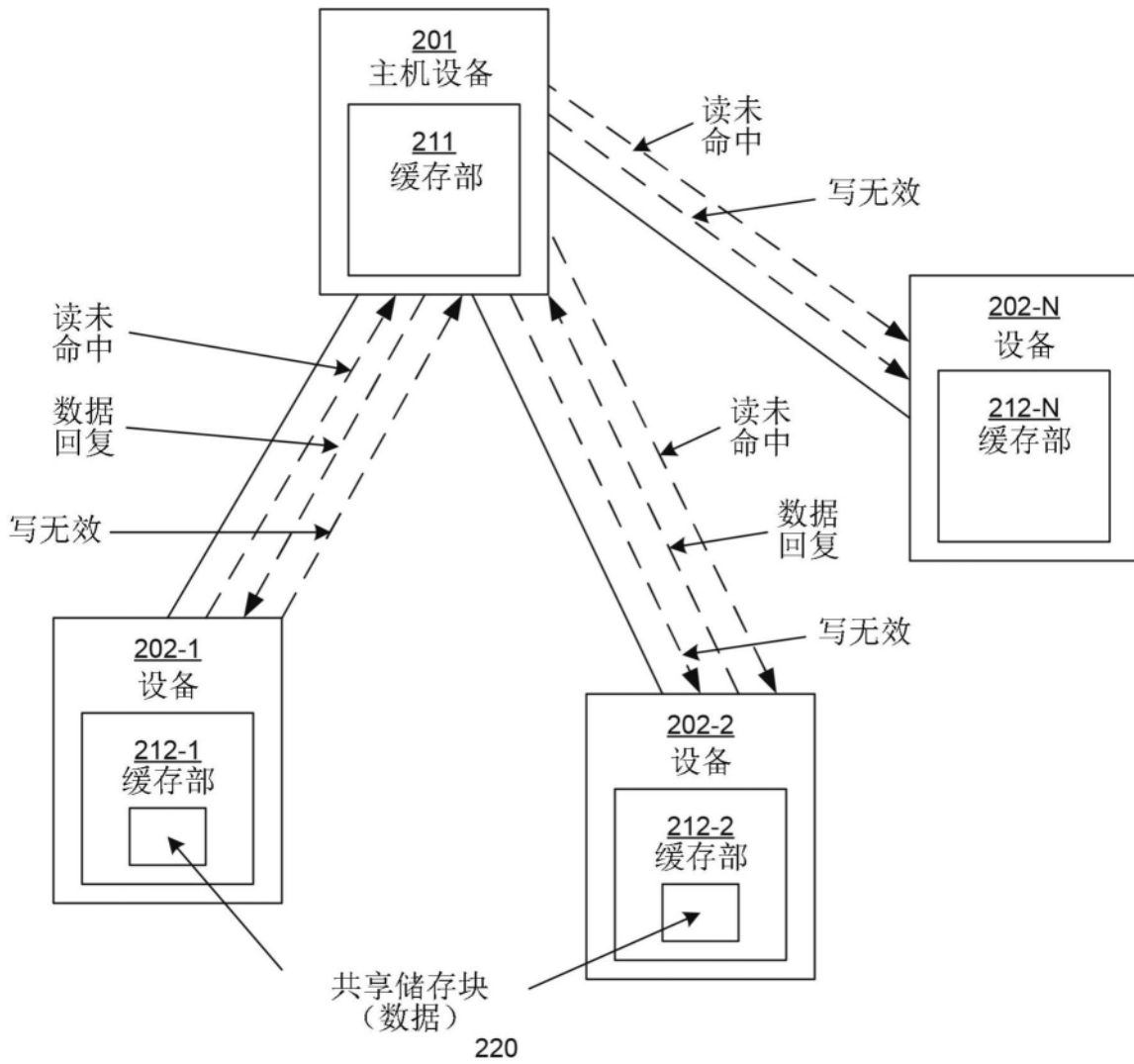


图5

600

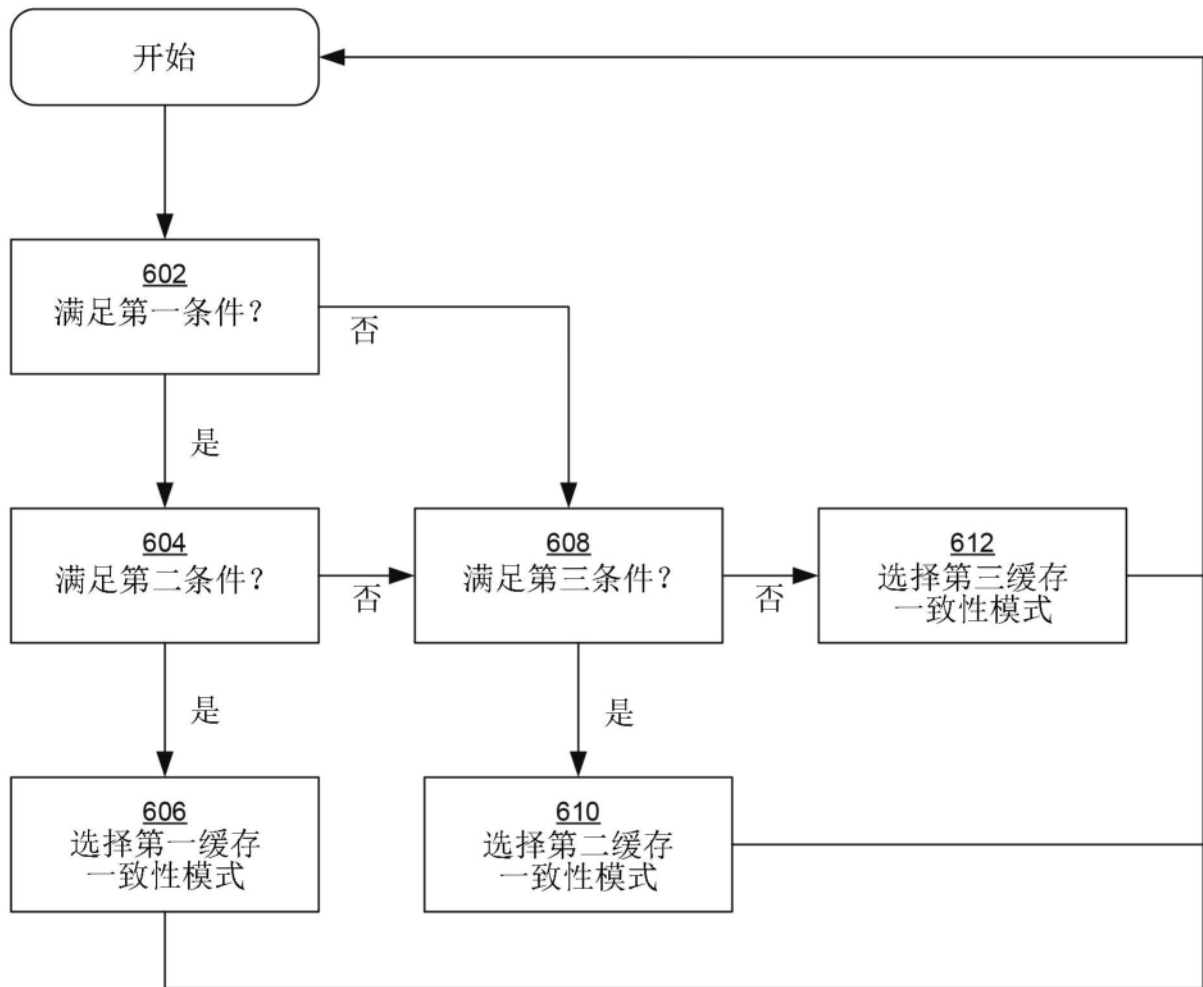


图6

700

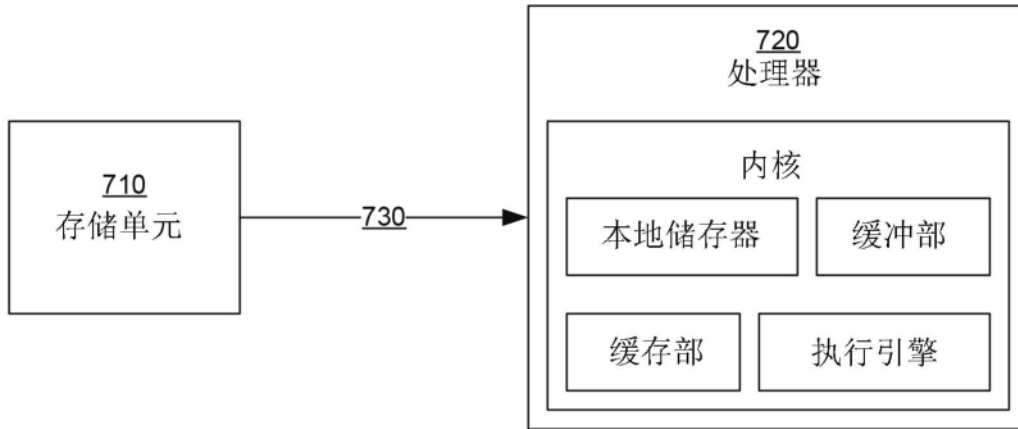


图7