(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 109240944 B
(45)授权公告日 2021.02.19

(21)申请号 201810931880.2
(22)申请日 2018.08.16
(65)同一申请的已公布的文献号
申请公布号 CN 109240944 A
(43)申请公布日 2019.01.18
(73)专利权人 上海天数智芯半导体有限公司
地址 201100 上海市闵行区苏召路1628号
(72)发明人 倪岭 李云鹏 王永流 郑成根 郑金山 邵平平
(74)专利代理机构 南京钟山专利代理有限公司 32252
代理人 戴朝荣
(51)Int.Cl.
G06F 12/0804(2016.01)
G06F 12/0871(2016.01)

(56)对比文件
US 9336152 B1,2016.05.10
CN 103077133 A,2013.05.01
CN 103605485 A,2014.02.26

审查员 顾兰

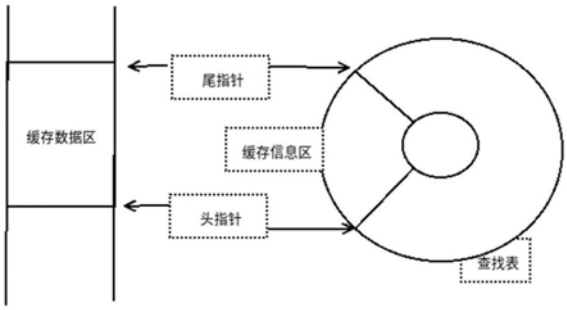
权利要求书1页 说明书3页 附图1页

(54)发明名称

一种基于可变长缓存行的数据读写方法

(57)摘要

本发明公开了一种基于可变长缓存行的数据读写方法,查找表来存储每个请求的缓存行信息,当一个读取任务到达缓存时,根据请求索引获取缓存线信息,若请求命中了缓存,根据缓存线信息中的缓存偏移量和缓存线大小,缓存中的数据在多个周期中被读取并发送给请求者,否则请求不在缓存中,一些读取请求被创建并发送出去,把偏移量、标签和缓存行大小记到查找表的记录中,并将请求发送到DRAM,一旦所有数据返回并写入缓存,查找表相应的记录就被设置为有效。本发明只读写有效数据,只存储有效数据,无效数据丢弃,这样节约读写带宽和存储空间。



CN 109240944 B

1.一种基于可变长缓存行的数据读写方法,其特征在于包含以下步骤:

步骤一:数据在缓存中以环形缓冲方式排列,标记与数据之间设置查找表,查找表也以环形缓冲方式管理;

步骤二:若缓存收到读请求,首先检查该请求是否命中查找表中已有的有效记录;

步骤三:若在查找表中找到相同标记的记录,该读请求命中缓存,从命中的记录中读出数据偏移和数据大小,然后读出数据缓存中相应的数据并返回给请求者;

步骤四:若没有命中,就添加一个新记录到查找表;

步骤五:移动头指针在查找表中获取条目;

步骤六:若这是一个有效的入口,则在数据缓存中释放其数据,并在数据缓存中分配所需的大小;

步骤七:若数据缓存中的可用大小小于要求的大小,则在查找表中按顺序释放更多的入口,直到有足够的空间;并将请求发送到DRAM,一旦所有数据返回并写入缓存,查找表相应的记录就被设置为有效;

步骤八:若缓存收到写请求,则添加一个新记录到查找表;

步骤九:移动头指针并在查找表中获取条目;

步骤十:若这是一个有效的入口,则在数据缓存中释放其数据,并在数据缓存中分配所需的大小;

步骤十一:若数据缓存中可用的大小小于要求的大小,则在查找表中按顺序释放更多的入口,直到有足够的空间;

步骤十二:然后用数据缓存更新查找表中相应记录的信息、偏移量和请求大小,并将数据写入数据缓存,使缓存行有效。

2.按照权利要求1所述的一种基于可变长缓存行的数据读写方法,其特征在于:所述步骤一中,若头指针到尾指针大于1,则数据缓存有头指针-尾指针-1个有效入口;若头指针等于尾指针,则数据缓存是空的。

3.按照权利要求1所述的一种基于可变长缓存行的数据读写方法,其特征在于:所述步骤二检查该请求是否命中查找表中已有的有效记录是通过比较读请求的标记和查找表中所有有效的记录中的标记来判断。

4.按照权利要求1所述的一种基于可变长缓存行的数据读写方法,其特征在于:为了在数据缓存中写入数据,将数据对应的缓存行大小加到尾指针;为了更新数据缓存,将数据写入数据缓存,并将它的缓存行大小加到头指针;为了检查是否有k个可用条目,确保头指针-尾指针-k>1。

5.按照权利要求1所述的一种基于可变长缓存行的数据读写方法,其特征在于:所述查找表存储每个请求的缓存行信息,缓存行信息包括有效位、缓存偏移量、缓存行大小和请求标记。

6.按照权利要求5所述的一种基于可变长缓存行的数据读写方法,其特征在于:所述请求包含对查找表的标记和索引,它有可变的有效数据长度,该数据长度是根据该请求的元数据计算的。

一种基于可变长缓存行的数据读写方法

技术领域

[0001] 本发明涉及一种数据读写方法,特别是一种基于可变长缓存行的数据读写方法。

背景技术

[0002] 通常情况,出于控制和管理的方便,缓存线是等长,但实际情况中,等长数据中有效数据的比例是变化的,意味着总是从总线读取/向总线写入部分无效数据,总线带宽被浪费;在缓存中存放部分无效数据,缓存利用率降低。每笔数据的有效缓存线的个数也不一样,如用等长得缓存线,缓存就有很多的无效数据。可变的缓存线可以节省每笔数据缓存线的个数。如果不用可变长得缓存线,我们就需要数据中最长的缓存线数来存储每笔数据。假设有四笔数据。第一、二、三笔需要两个缓存线,但第四笔数据要八个缓存线,如果用等长策略,就需要32个缓存线,若用不等长策略,只需要14个缓存线。节约了18个缓存线。这些缓存线就可以存其他的数据。因此需要提供一种不等长策略的数据读写方法来提高数据读写效率。

发明内容

[0003] 本发明所要解决的技术问题是提供一种基于可变长缓存行的数据读写方法。

[0004] 为解决上述技术问题,本发明所采用的技术方案是:

[0005] 一种基于可变长缓存行的数据读写方法,其特征在于包含以下步骤:

[0006] 步骤一:数据在缓存中以环形缓冲方式排列,标志信息与与数据之间设置查找表,查找表也以环形缓冲方式管理;

[0007] 步骤二:若缓存收到读请求,首先检查该请求是否命中查找表中已有的有效记录;

[0008] 步骤三:若在查找表中找到相同标记的记录,该读请求命中缓存,从命中的记录中读出数据偏移和数据大小,然后读出数据缓存中相应的数据并返回给请求者;

[0009] 步骤四:若没有命中,就添加一个新记录到查找表;

[0010] 步骤五:移动头指针在查找表中获取条目;

[0011] 步骤六:若这是一个有效的入口,则在数据缓存中释放其数据,并在数据缓存中分配所需的大小;

[0012] 步骤七:若数据缓存中的可用大小小于要求的大小,则在查找表中按顺序释放更多的入口,直到有足够的空间;并将请求发送到DRAM,一旦所有数据返回并写入缓存,查找表相应的记录就被设置为有效;

[0013] 步骤八:若缓存收到写请求,则添加一个新记录到查找表;

[0014] 步骤九:移动头指针并在查阅表中获取条目;

[0015] 步骤十:若这是一个有效的入口,则在数据缓存中释放其数据,并在数据缓存中分配所需的大小;

[0016] 步骤十一:若数据缓存中可用的大小小于要求的大小,则在查找表中按顺序释放更多的入口,直到有足够的空间;

[0017] 步骤十二:然后用数据缓存更新查找表中相应记录的信息、偏移量和请求大小,并将数据写入数据缓存,使缓存行有效。

[0018] 进一步地,所述步骤一中,若头指针到尾指针大于1,则数据缓存有(头指针-尾指针-1)个有效入口;若头指针等于尾指针,则数据缓存是空的。

[0019] 进一步地,所述步骤二检查该请求是否命中查找表中已有的有效记录是通过比较读请求的标记和查找表中所有有效的记录中的标记来判断。

[0020] 进一步地,为了在数据缓存中释放数据,将数据对应的缓存线大小加到尾指针;为了更新数据缓存,将数据写入数据缓存,并将它的缓存线大小加到头指针;为了检查是否有k个可用条目,确保(头指针-尾指针-k)>1。

[0021] 进一步地,所述查找表存储每个请求的缓存行信息,缓存行信息包括有效位、缓存偏移量、缓存行大小和请求标记。

[0022] 进一步地,所述请求包含对查找表的标记和索引,它有可变的有效数据长度,该数据长度是根据该请求的元数据计算的。

[0023] 本发明与现有技术相比,具有以下优点和效果:本发明根据数据特点,只读取、写入、存放有效数据,有效的利用总线带宽以及缓存空间,从而增加了总线位宽,提高了频率,增大了缓存物理容量。

附图说明

[0024] 图1是本发明的一种基于可变长缓存行的数据读写方法的缓存结构示意图。

[0025] 图2是本发明的一种基于可变长缓存行的数据读写方法的示意图。

具体实施方式

[0026] 下面结合附图并通过实施例对本发明作进一步的详细说明,以下实施例是对本发明的解释而本发明并不局限于以下实施例。

[0027] 如图1所示,本发明的一种基于可变长缓存行的数据读写方法,其特征在于包含以下步骤:

[0028] 步骤一:数据在缓存中以环形缓冲方式排列,标志信息与与数据之间设置查找表,查找表也以环形缓冲方式管理;

[0029] 由于数据长度不固定,因此需要增加一个查找表(LUT),来连接标志和数据。数据在缓存中以环形缓冲的方式排列。查找表也是用环形缓冲的方式管理。如果头指针到尾指针大于1,则数据缓存有(头指针-尾指针-1)个有效入口。如果头指针等于尾指针,则数据缓存是空的。查找表也是以这种循环缓冲区的方式更新的。如果缓存收到写请求,移动头指针并在查找表中获取条目。如果这是一个有效的入口,则在数据缓存中释放其数据。并在数据缓存中分配所需的大小。如果数据缓存中可用的大小小于要求,则在查找表中按顺序释放更多的入口,直到有足够的空间。然后更新查找表相应记录的信息,偏移量和请求大小,还将数据写入数据缓存,并使该记录有效。

[0030] 为了在数据缓存中释放数据,我们将数据对应的缓存线大小加到尾指针。为了更新数据缓存,我们将数据写入数据缓存,并将它的缓存线大小加到头指针。为了检查是否有k个可用条目,我们确保(头指针-尾指针-k)>1。

[0031] 步骤二:若缓存收到读请求,首先检查该请求是否命中查找表中已有的有效记录;这是通过比较读请求的标记和查找表中所有有效的记录中的标记来判断。

[0032] 步骤三:若在查找表中找到相同标记的记录,该读请求命中缓存,从命中的记录中读出数据偏移和数据大小,然后读出数据缓存中相应的数据并返回给请求者;

[0033] 步骤四:若没有命中,就添加一个新记录到查找表;

[0034] 步骤五:移动头指针在查找表中获取条目;

[0035] 步骤六:若这是一个有效的入口,则在数据缓存中释放其数据,并在数据缓存中分配所需的大小;

[0036] 步骤七:若数据缓存中的可用大小小于要求的大小,则在查找表中按顺序释放更多的入口,直到有足够的空间;并将请求发送到DRAM,一旦所有数据返回并写入缓存,查找表相应的记录就被设置为有效;

[0037] 步骤八:若缓存收到写请求,则添加一个新记录到查找表;

[0038] 步骤九:移动头指针并在查阅表中获取条目;

[0039] 步骤十:若这是一个有效的入口,则在数据缓存中释放其数据,并在数据缓存中分配所需的大小;

[0040] 步骤十一:若数据缓存中可用的大小小于要求的大小,则在查找表中按顺序释放更多的入口,直到有足够的空间;

[0041] 步骤十二:然后用数据缓存更新查找表中相应记录的信息、偏移量和请求大小,并将数据写入数据缓存,使缓存行有效。

[0042] 本发明的工作原理为,在带有可变缓存行的缓存中,有一个查找表来存储每个请求的缓存行信息。此信息包括有效位、缓存偏移量、缓存行大小和请求标记.请求分为两部分:对查找表的标记和索引。它有可变的有效数据长度,该数据长度是根据该请求的元数据计算的。当一个读取任务到达缓存时,根据请求索引获取缓存线信息。如果这是一个有效的缓存线,而这个缓存线的标签等于请求的标签,这意味着这个请求命中了缓存,根据缓存线信息中的缓存偏移量和缓存线大小,缓存中的数据在多个周期中被读取并发送给请求者。否则,请求不在缓存中,一些读取请求被创建并发送出去。要发出这些请求,我们需要在缓存中找到一些空间。空间的数量是请求的数据长度。如果我们没有足够的空间,一些缓存线被设成无效,它们占用的空间被释放。我们把偏移量、标签和缓存行大小记到查找表的记录中,并将请求发送到DRAM,一旦所有数据返回并写入缓存,查找表相应的记录就被设置为有效。

[0043] 本说明书中所描述的以上内容仅仅是对本发明所作的举例说明。本发明所属技术领域的技术人员可以对所描述的具体实施例做各种修改或补充或采用类似的方式替代,只要不偏离本发明说明书的内容或者超越本权利要求书所定义的范围,均应属于本发明的保护范围。

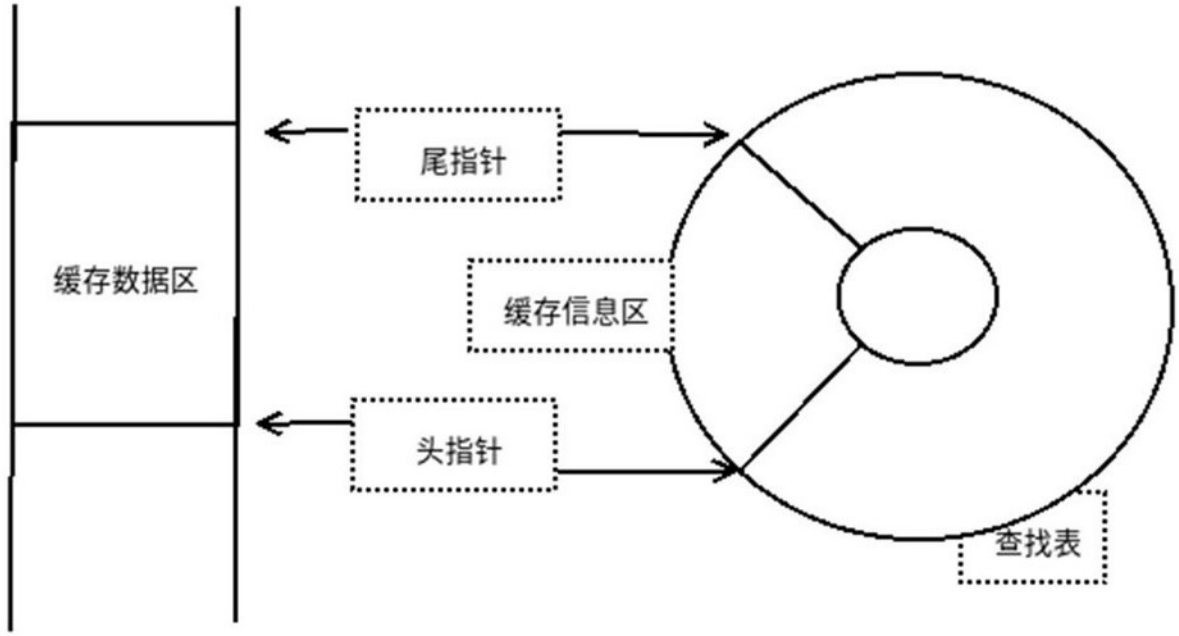


图1

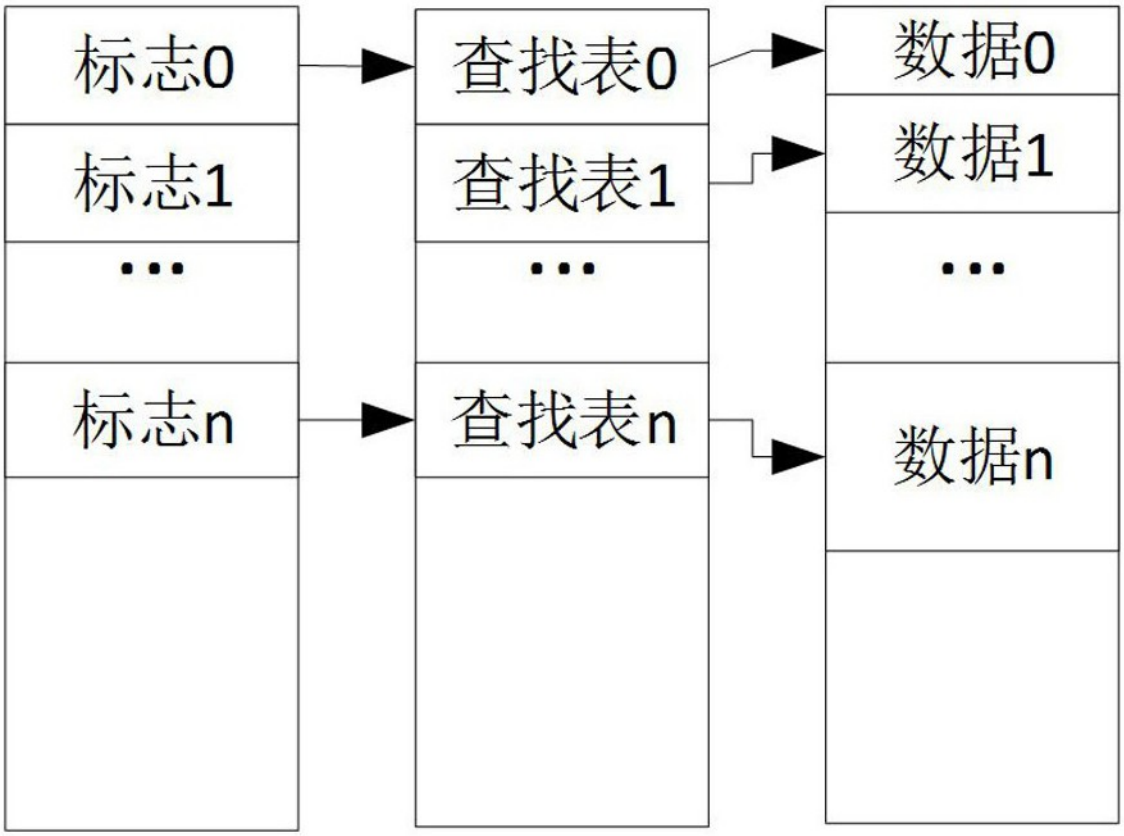


图2