

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G06F 11/10 (2006.01)

G11B 20/18 (2006.01)



# [12] 发明专利说明书

专利号 ZL 98800381.3

[45] 授权公告日 2006年2月15日

[11] 授权公告号 CN 1242330C

[22] 申请日 1998.2.27 [21] 申请号 98800381.3

[30] 优先权

[32] 1997.2.28 [33] US [31] 08/808,396

[86] 国际申请 PCT/US1998/003833 1998.2.27

[87] 国际公布 WO1998/038576 英 1998.9.3

[85] 进入国家阶段日期 1998.11.27

[71] 专利权人 网络装置公司

地址 美国加利福尼亚州

[72] 发明人 斯蒂芬·R·克雷曼

审查员 张雪梅

[74] 专利代理机构 中国国际贸易促进委员会专利  
商标事务所

代理人 李德山

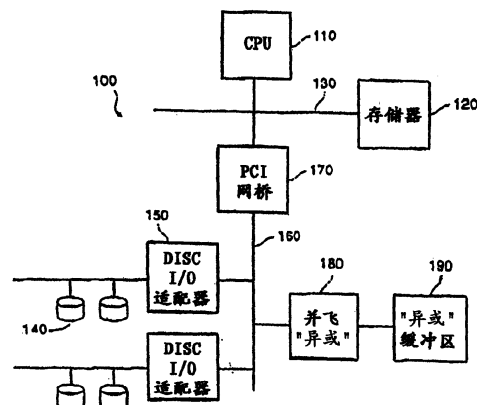
权利要求书 3 页 说明书 15 页 附图 3 页

## [54] 发明名称

执行并飞“异或”运算的方法和系统

## [57] 摘要

本发明提供一种不需消耗大量计算资源的执行“异或”运算的方法和系统。一个专用处理机和一组盘驱动器连接到相同的总线上；该专用处理机观察对和从盘驱动器的数据传送，并在不需要单独的传送下在对和从盘驱动器传送的数据执行“异或”运算。该专用处理机保持一个用于“异或”运算的“异或”累加器，该累加器记录“异或”运算的结果并在处理机的命令下读出该结果。该“异或”累加器包括一组用于一组选定的 RAID 条中的各个 RAID 条的累加器寄存器。一个存储器（例如按内容寻址的存储器）把一组累加器寄存器和各个选定的 RAID 条联系起来。



1. 一种方法，包括步骤：  
启动沿着总线（130、160）向或从多个 RAID 条中的至少一个传送数据的数据传送操作；  
监视所述总线（130、160）上寻址到多个地址中的任意地址的数据；  
根据所述数据的地址从多个“异或”累加器缓冲区（340、440）中选择“异或”累加器缓冲区（340、440）；和  
在所述“异或”累加器缓冲区（340、440）中累加对所述数据执行的“异或”运算（354、453）的结果。
2. 如权利要求 1 中的方法，其中所述数据传送操作向或从所述多个 RAID 条的条的块传送数据。
3. 如权利要求 2 中的方法，其中所述条包含在所述多个 RAID 条的每个上的块。
4. 如权利要求 1 中的方法，其中所述总线（130、160）还包括第一总线（130），和将所述多个 RAID 条连接到所述第一总线（130）的第二总线（160）。
5. 如权利要求 1 中的方法，其中所述数据传送操作是向或从存储器（120）传送数据的直接存储器访问（223）操作。
6. 如权利要求 5 中的方法，其中所述存储器（120）忽略所述地址。
7. 如权利要求 1 中的方法，其中所述累加步骤不干扰所述数据传送操作。
8. 如权利要求 1 中的方法，其中选择所述“异或”累加器缓冲区（340、440）的所述步骤还包括访问使所述地址与所述“异或”累加器缓冲区（340、440）相关的存储器（330、430）。
9. 如权利要求 8 中的方法，其中所述存储器是按内容寻址存储器（330）。
10. 如权利要求 1 中的方法，其中所述数据传送操作是从所述 RAID 条的条的块向所述存储器（120）传送数据的直接存储器访问操作，所述

条包含在所述多个 RAID 条的每个上的块，所述存储器（120）忽略所述地址，并且所述累加步骤不干扰所述数据传送操作。

11. 如权利要求 10 中的方法，其中选择所述“异或”累加器缓冲区（340、440）的所述步骤还包括访问使所述地址与所述“异或”累加器缓冲区（340、440）相关的按内容访问存储器（330）。

12. 一种系统，包括：

处理器（110）；

存储器（120）；

多个 RAID 条；

允许向或从所述多个 RAID 条中的至少一个传送数据的数据传送操作的总线（130、160）；

多个“异或”累加器缓冲区（340，440）；以及

部件（180），该部件（180）监视所述总线（130、160）上寻址到多个地址中的任意地址的数据，根据所述数据的地址从多个“异或”累加器缓冲区（340、440）中选择“异或”累加器缓冲区（340、440），并且在所述“异或”累加器缓冲区（340、440）中累加对所述数据执行的“异或”运算（354、453）的结果。

13. 如权利要求 12 中的系统，其中所述数据传送操作向或从所述多个 RAID 条的条的块传送数据。

14. 如权利要求 13 中的系统，其中所述条包含在所述多个 RAID 条的每个上的块。

15. 如权利要求 12 中的系统，其中所述总线（130、160）还包括第一总线（130），和将所述多个 RAID 条连接到所述第一总线（130）的第二总线（160）。

16. 如权利要求 12 中的系统，其中所述数据传送操作是向或从存储器（120）传送数据的直接存储器访问（223）操作。

17. 如权利要求 16 中的系统，其中所述存储器（120）忽略所述地址。

18. 如权利要求 12 中的系统，其中所述累加步骤不干扰所述数据传

送操作。

19. 如权利要求 12 中的系统, 其中选择所述“异或”累加器缓冲区 (340、440) 的所述步骤还包括访问使所述地址与所述“异或”累加器缓冲区 (340、440) 相关的存储器 (330、430)。

20. 如权利要求 19 中的系统, 其中所述存储器是按内容寻址存储器 (330)。

21. 如权利要求 12 中的系统, 其中所述数据传送操作是从所述 RAID 条的条的块向所述存储器 (120) 传送数据的直接存储器访问操作, 所述条包含在所述多个 RAID 条的每个上的块, 所述存储器 (120) 忽略所述地址, 并且所述累加步骤不干扰所述数据传送操作。

22. 如权利要求 21 中的系统, 其中选择所述“异或”累加器缓冲区 (340、440) 的所述步骤还包括访问使所述地址与所述“异或”累加器缓冲区 (340、440) 相关的按内容访问存储器 (330)。

## 执行并飞“异或”运算的方法和系统

### 技术领域

本发明涉及存储系统。

### 背景技术

一种已知的存储系统称为“RAID”，在该系统中跨越多个盘驱动器分布并存储数据，从而在这些盘驱动器中的一个或者一个或多个存储块出故障而丢失数据时能恢复数据。为了存储以后可恢复的数据，RAID系统同时记录数据以及奇偶信息，从而可由奇偶信息恢复数据。例如，在一种类型的RAID系统中，为了在称为“条”的单元中记录四块数据和一块奇偶信息使用一组为五个的盘。

现有技术中出现的一个问题是记录奇偶信息的RAID系统需要快速地和经常地计算奇偶信息。例如，写一个条的操作需要计算至少一组一奇偶信息。计算奇偶信息主要采用XOR(“异或”)操作。这需要分配相当多的计算资源，例如处理机周期、超高速缓存单元以及存储器存取周期。

一种计算奇偶信息的可能技术是设置用来控制计算“异或”操作的处理机的软件。这种技术具有几种设计选择方式，例如设法调度处理机指令以避免寄存器冲突、避免不必要的等待并且减少未击中超高速缓存，从而避免不必要的存储内容转储以及处理机的不必要等待。尽管该技术达到计算奇偶信息的目标，但它具有需要相当多的计算资源(处理机周期、超高速单元、转储)的缺点。

另一种可能的计算奇偶信息的技术是设置一个带有独立存储器的第二处理机，以执行“异或”运算。该第二处理机利用DMA连接和主存储器(即，第一处理机使用的存储器)耦合，该DMA连接利用DMA操作从存储器中检索“异或”运算的输入，从而使第一处理机去执行其它的任务。虽然该技术达到计算奇偶信息的目标，即使它使第一处理机

摆脱了一些负担，它仍具有需要相当多的转储和使用系统总线的缺点。

### 发明内容

因此，提供一种消耗较少计算资源的执行“异或”运算的方法和系统应是有好处的。本发明的实施方式可达到这种好处，在本发明的实施方式中一个专用处理机观察传送到系统总线上的存储器存取和盘存取，执行“异或”运算，记录“异或”运算的结果并根据处理机命令输出这些“异或”运算的结果。

根据本发明的一个方面，提供一种方法，包括步骤：启动沿着总线向或从多个 RAID 条中的至少一个传送数据的数据传送操作；监视所述总线上寻址到多个地址中的任意地址的数据；根据所述数据的地址从多个“异或”累加器缓冲区中选择“异或”累加器缓冲区；和在所述“异或”累加器缓冲区中累加对所述数据执行的“异或”运算的结果。

根据本发明的另一方面，提供一种系统，包括：处理器；存储器；多个 RAID 条；允许向或从所述多个 RAID 条中的至少一个传送数据的数据传送操作的总线；多个“异或”累加器缓冲区；以及部件，该部件监视所述总线上寻址到多个地址中的任意地址的数据，根据所述数据的地址从多个“异或”累加器缓冲区中选择“异或”累加器缓冲区，并且在所述“异或”累加器缓冲区中累加对所述数据执行的“异或”运算的结果。

### 附图说明

图 1 是执行并飞“异或”的系统的方块图。

图 2 是操作执行并飞“异或”的系统的方法的流程图。

图 3 是第一并飞“异或”部件的方块图。

图 4 是第二并飞“异或”部件的方块图。

### 具体实施方式

在下述说明中，就最佳处理步骤及数据结构说明本发明的一种最佳实施方式。然而，熟练的技术人员应理解，在研读本申请后，可以用一个或多个在程序控制下运行的通用处理机（或者适宜于具体处理步骤及数据结构的专用处理机）实现本发明的实施方式，并且利用这样的设备实现本文中所说明的最佳处理步骤以及数据结构并不需要过多的实验或

其它发明。

### 系统部件

图 1 是执行并飞(fly-by)“异或”的系统的方块图。

系统的 100 至少包括一个处理机 110、一个存储器 120 以及一条用于连接处理机 110 和存储器 120 的系统总线 130。虽然只示出一个处理机 110, 在对于某些应用为最佳的一些实施方式中, 系统总线 130 可连接多于一个的处理机 110 并且在操作中这些处理机 100 相互协调。

系统 100 还包括多个盘驱动器 140 以及至少一个和盘驱动器 140 连接的盘 I/O 适配器 150, 适配器 150 用于控制盘驱动器 140 并利用底层总线 160 耦合盘驱动器 140 和系统总线 130 之间的数据。

在一种最佳实施方式中, 底层总线 160 是 PCI 总线, PCI 总线是周知的标准并利用 PCI 网桥 170 和系统总线 130 连接。

虽然在最佳实施方式中底层总线 160 是 PCI 总线, 但本发明并不限制于使用 PCI 总线或者限制于使用任何其它特定类型的总线。类似地, 虽然在本文中进一步把底层总线 160 描述为一个 32 位的总线, 本发明并不受限于使用 32 位的总线, 并且也可以用其它的总线宽度, 例如 64 位的总线实现。这样的替代不需要过多的经验或创造。

一个并飞“异或”部件 180 和底层总线 160 连接, 从而从利用底层总线 160 的数据传送搜集数据。并飞“异或”部件利用“异或”缓冲器 190 保存“异或”运算的部分或最终结果。

当处理机 110 控制存储器 120 和一个或多个盘驱动器 140 间的数据传送时, 利用系统总线 130、PCI 网桥 170 以及底层总线 160 传送数据。当利用底层总线 160 传送数据时, 并飞“异或”部件 180 读数据(但不干扰数据传送)并且把数据添加到“异或”缓冲器 190 里的多个“异或”累加器(见图 2)的一个中。当整条地或部分地读或写一个 RAID 条时, 并飞“异或”部件 180 动态地为该 RAID 条计算一组奇偶信息, 不再需要从系统 100 的任何部分得到其它资源。不再需要额外的数据传送, 并且不需要处理机 110 和存储器 120 执行“异或”运算。

当处理机 110 如此控制时, 并飞“异或”部件 180 将从“异或”缓

冲器 190 里的一个“异或”累加器中读出累加的奇偶信息，以把奇偶信息写到刚被读或写的该 RAID 条中的存储块上。来自“异或”缓冲器 190 的累加奇偶信息还可在该 RAID 条中的一个盘驱动器 140 丢失该 RAID 条的整个存储块或部分存储块时，重构或修复该 RAID 条中的数据。

#### 系统操作

图 2 是一种执行并飞“异或”的系统的操作方法流程图。

并飞“异或”部件 180 的操作方法 200 包括一系列由处理机 110、盘驱动器 140 以及并飞“异或”部件 180 执行的步骤。

在点 210 处，系统 100 将对其 RAID 条执行“异或”运算。

在步骤 221，处理机 110 在存储器 120 中分配一个存储缓冲区，并且建立要向或从该存储缓冲区和一个盘驱动器 140 进行 DMA 传送的数据。

在步骤 222，处理机 110 向并飞“异或”部件 180 发送该存储缓冲区的存储器地址，并控制并飞“异或”部件 180 识别向该存储缓冲区传送或从其传来的数据。处理机 110 还分配“异或”缓冲器 190 里各“异或”累加器缓冲区中的一个缓冲区以累加“异或”运算的结果。

在步骤 223，执行 DMA 传送。并飞“异或”部件 180 监视底层总线 160 并试图识别向或来自该存储缓冲区的传送。当并飞“异或”部件 180 识别出这种传送时，它复制正在传送的数据并且(通过执行“异或”运算)把数据加到相关“异或”累加器缓冲区中的累加奇偶信息上。当该识别出的存储缓冲区的所有数据全都传送了时，该相关的“异或”累加器缓冲区持有用于该 RAID 条的计算出的奇偶信息。

在步骤 224，处理机 110 控制并飞“异或”部件把累加的奇偶信息从该相关的“异或”累加器缓冲区读到存储器 110 里的某存储缓冲区中或者读到一个盘驱动器 140 上的某存储块中。在这二种情况下，从并飞“异或”部件 180 的读出是通过处理机读或通过对一个盘驱动器的 DMA 传送进行的。在一种最佳实施方式中，对累加的奇偶信息的读出造成把相关的“异或”累加器缓冲区复位到零。

在点 230，完成该 RAID 条的“异或”运算，而系统 100 可以继续处理另一个 RAID 条。



在一种最佳实施方式中，处理机 110 可以实际上在不把数据从一个盘驱动器传送到存储器 120 的某存储缓冲区下执行“异或”运算。这在本文中称为对“空”存储缓冲区的传送，因为该运算各方面上都象是对某存储缓冲区进行，唯独实际上不把数据存储在存储器 120 中。能向一个空存储缓冲区传送是有好处的，这是因为存在用于 RAID 条的盘操作，在盘操作中需要读或写少于 RAID 条的全部存储块的存储块(仿佛存储在盘驱动器 140 上)，但是其中又必须为整个 RAID 条计算奇偶信息。

例如，当只需要写某 RAID 条的一个存储块时，必须写该存储块以及奇偶信息存储块两者。在已知的 RAID 存储系统中，必须首先从盘中读出当前奇偶信息以及数据块二者，从而处理机可接着计算新的奇偶信息。然而借助本发明，可把这些块传送到一个空存储缓冲区，从而不使用处理机 110 或存储器 120 的资源，只利用并飞“异或”部件 180 计算奇偶信息。

此外，在一种最佳实施方式中，以较少的步骤执行只写 RAID 条的一个存储块：(1)把老的存储块数据以及老的奇偶块数据传送到空存储缓冲区，以利用并飞“异或”部件 180 累加奇偶信息。(2)把新的存储块数据写入到存储块，从而利用并飞“异或”部件 180 在写存储块的同时累加奇偶信息。(3)把累加的奇偶信息从“异或”缓冲器 190 中写入到奇偶存储块里。这些运算不需要处理机 110 以及存储器 120 的资源。

为执行对空存储缓冲区的传送，在步骤 221 中，处理机 110 不在存储器 120 中分配一个存储缓冲区，并且建立要在一个盘驱动器 140 和某不为存储器 120 识别的存储器地址之间进行 DMA 传送的数据。在步骤 223，执行 DMA 传送并且并飞“异或”部件 180 监视底层总线 160，但存储器 120 不捕获任何“传送”到该空存储缓冲区的数据。

第一并飞“异或”部件

图 3 是第一并飞“异或”部件的方块图。

第一并飞“异或”部件包括一个总线接口 310、一个定时及控制部件 320、一个按内容寻址存储器(CAM)330 以及其它有关部件。“异或”缓冲器 190 和并飞“异或”部件 180 连接，并且包括一个随机存取存储器

(RAM)340、一个 FIFO(先进先出)队列 350 以及其它有关部件。虽然概念上“异或”缓冲器 190 包括 RAM340, 在一种最佳实施方式中 RAM340 是用外部 RAM 电路实现的, 而 FIFO350 以及其它有关部件是在 ASIC 下用并飞“异或”部件 180 的其它电路实现的。

总线接口 310 和定时及控制部件 320 提供并飞“异或”部件 180 和底层总线 160 之间的接口。该接口包括一组通路: 由底层总线 160 使用的传递控制信号的控制通路 311, 传递具有位 0 到位 31 的总线地址值的总线地址(BA)通路 312 以及传递具有位 0 到位 31 的总线数据值的总线数据(BD)通路 313。采用 PCI 总线的通信在计算机系统中是周知的。

在备择的实施方式中, 底层总线 160 不是 PCI 总线或不为 32 位宽, 总线数据值具有其它位的标识符(例如位 0 到位 63)并利用非 PCI 的通信协议进行通信。

“异或”缓冲器 190 包括一个随机存取存储器(RAM)340, 后者提供一组 32 个的“异或”累加器缓冲区, 每一个缓冲区为 1024 入口长 4 字节(32 位)宽, 即尺寸为 4K 字节。从而并飞“异或”部件 180 可以同时执行 32 个不同的 RAID 条的“异或”累加。

“异或”缓冲器 190 还包括一个 FIFO 队列 350, 它输入总线数据值的各顺序字并根据这些顺序字执行“异或”运算, 从而在 RAM340 里的 32 个“异或”累加器缓冲区中的一个里累加奇偶信息。

CAM330 提供一个 256 个项的组, 每个项把 20 位的总线地址值和 RAM340 里 32 个“异或”累加器缓冲区中的一个缓冲区的 5 位相关地址联系起来。CAM330 包括一个地址输入端口 331, 用于选择 CAM330 里 256 个项中的一项; 一个 20 位的标记端口 332, 用于提供由 CAM330 去匹配的 20 位的标记; 一个 5 位的值的端口 333, 用于提供 5 位的值, 以由 CAM330 把该值和 20 位的标记联系起来; 一个写控制端口 34, 用于接收写控制信号; 以及一个匹配控制端口 335, 用于提供匹配控制信号。

定时及控制部件 320 还从匹配控制端口 335 接收匹配控制信号, 并且提供用于对 RAM340 寻址的 15 位宽的 RAM 地址通路(RA)314 以及用于控制 CAM330 的写控制通路 315。

总线地址通路 312 的位 12 至位 31 以及总线数据通路 313 的位 12 至位 31(即,各情况下的 20 位高位)和 CAM 多路复用器 336 的输入端连接,该多路复用器的一个输出端和 CAM330 的标记端口 332 连接。标记端口 332 还提供一个 20 位的标记,利用激励器 337 可把该标记耦合到总线数据通路 313。值端口处 333 的相关 5 位值被耦合到“异或”缓冲器 190,并且利用一对激励器 338 还耦合到总线数据通路 313。

“异或”缓冲器 190 包括 RAM340, RAM340 包括一个用于接收 15 位 RAM 地址值的 RAM 地址端口 341 以及一个 RAM 数据端口 342, RAM 数据端口 342 用于提供存储在 RAM340 中相关单元里的 32 位 RAM 数据值。在概念上把 RAM340 划分成 32 个独立的“异或”累加器缓冲区,每个缓冲区包含 1K 的 32 位的字(即, 4K 字节)并且占据 RAM340 中的一组单元,每个单元具有其 5 位的“异或”累加器缓冲区号(0 至 31)以充当 RAM 地址值的(五个)高位。

值端口 333 处的相关 5 位值被耦合到 RAM 地址端口 341, 以提供 RAM340 中寻址字的 5 个高位。总线地址通路 312 的位 2 至位 11(即用于寻址某 4 字节字的 10 个低位)提供 RAM340 中的寻址字的 10 个低位,并且也被耦合到 RAM 地址端口 341。这样, RAM340 在 RAM 地址端口 341 接收 15 位的地址,并在 RAM 数据端口 342 提供 32 位的输出数据字。

RAM 数据端口 342 利用激励器 351 和总线数据通路 313(以及 FIFO 队列 350 的一个输入端)连接,并且把 FIFO 多路复用器 352 的一个输出端以及一个 FIFO 寄存器 353 的一个输入端连接。FIFO 队列 350 具有和一个“异或”运算器 354 连接的输出端,“异或”运算器 354 包括一组 32 个的“异或”门,用于在一个完整的 32 位字上并行地执行“异或”运算。FIFO 多路复用器 352 和“异或”运算器 354 的输出端以及一个“零”输入端 355 连接。

为了如在步骤 221 和步骤 222 中为新的 RAID 条分配“异或”累加器缓冲区,定时及控制部件 320 使 CAM330 装入一个用于存储器 120 中某存储缓冲区的 20 位的存储器地址(或者用于某空存储缓冲区的 20 位的存储器地址),并把该 20 位的存储器地址和一个用于 32 个“异或”

累加器缓冲区中的一个的 5 位标识符联系起来。把总线地址通路 312 的位 2 至位 9 耦合到地址输入端口 331 以便为 CAM330 提供新项的地址。在定时及控制部件 320 的控制下, CAM 多路复用器 336 选择用于连接到标记端口 332 的总线数据通路 313。CAM330 把来自总线数据通路 313 的位 12 至位 31 的总线数据值插入到地址输入端口 331 处规定的地址中作为新的 20 位的标记, 并且把来自总线数据通路 313 的位 0 至位 4 的总线数据值插入为新的 5 位的值。处理机 110 出于诊断目的也可以读 CAM330 中某选出的项。

为了如在步骤 223 中那样监视底层总线 223, 定时及控制部件 320 把总线地址通路 312 选择成和标记端口 332 连接。当底层总线 160 上出现的某存储器地址(或者空存储缓冲区内的某地址)的 20 个高位和 CAM330 的一个 20 位的标记匹配时, CAM330 把总线地址的位 12 至 31 以及相关的 5 位值和该 20 位的标记匹配。

相关的 5 位值用于选择 RAM340 里的 32 个“异或”累加器缓冲器中的一个, 而总线地址通路 312 的位 2 至位 11 用于选择 RAM340 里该“异或”累加器缓冲区中的某个别字, 从而把底层总线 160 上的数据值逐字地读入到 FIFO 队列 350 中。定时及控制部件 320 对 FIFO 队列 350 中的各字定序, 连同从 RAM340 读到 FIFO 寄存器 353 的某对应字一起通过“异或”运算器 354, 从而对 RAM340 里相关“异或”累加器缓冲区中的数据以及 FIFO 队列 350 中的数据执行“异或”运算。定时及控制部件 320 然后使“异或”运算的结果被回写到 RAM340 里的相关“异或”累加器缓冲区中。

为了如在步骤 224 中读出累加的奇偶信息, 定时和控制部件 320 把总线地址通路 312 选择成和 RAM 地址端口 341 连接, 以便按顺序地把 RAM340 里的“异或”累加器缓冲区中的各个字读出到总线数据通路 313 上, 同时或者对存储器 120 或者对一个盘驱动器 140 发生 DMA 传送。DMA 传送最常见是朝向一个盘驱动器的, 但若处理机 110 必须运行奇偶信息则也可以是朝向存储器 120 的。出于诊断目的或者如果出于其它原因要读出或操作位于“异或”缓冲器 190 中的奇偶信息时, 处理机 110

也可以直接读 RAM340 中的项。定时和控制部件 320 同时向 RAM340 里的“异或”累加器缓冲区中的每个字写入零值，从而把“异或”累加器缓冲区复位成统一的零值。

处理机 110 也可以直接对 RAM340 中选定的单元进行写，从而在“异或”缓冲器 190 和选定的信息之间直接执行“异或”运算。这类似于在步骤 223 中执行的“异或”操作。但是，不是把从 CAM330 的值端口 333 输出的 5 位的值而是把总线数据通路 313 的位 12 至位 16 连接到 RAM 地址端口 341 上。

在一种最佳实施方式，按如表 3-1 中所示使用总线地址通路 312 上的以及总线数据通路 313 上的值。表 3-1 中，地址区间值是用十六进制表示的，并且为总线地址(BA)通路 312 以及总线数据(BD)通路 313 显示选定的位区间。

表 3 - 1

地址区间	寄存器	读/写	数据/运算
000000..0000FF	PCI 总线寄存器	读写	按 PCI 总线协议规定
010000..0103FF	CAM 项	读写	读或写 CAM 项 BA<2 : 9> CAM 项号 BD<0 : 4> “异”累加器号 BD<12 : 31> 存储缓冲区地址
100000..10FFFF	“异”累加器缓冲区	读	读并清除“异”累加器缓冲区 BA<12 : 16> “异”累加器缓冲区号 BD<0 : 31> 结果
100000..10FFFF	“异”累加器缓冲区	写	对“异”累加器缓冲区的直接“异” BA<12 : 16> “异”累加器缓冲区号 BD<0 : 31> 用于直接“异”的数据
000000..FFFFFF	“异”累加器缓冲区	X	并飞“异”，清除存储缓冲区 BA<12 : 31> 要匹配的存储器地址 BD<0 : 31> 用于并飞“异”的数据
200000..7FFFFFFF	空存储缓冲区	写	并飞“异”，清除空存储缓冲区 BA<12 : 31> 要匹配的存储器地址 BD<0 : 31> 用于并飞“异”的数据

### 第二并飞“异或”部件

图 4 是第二种并飞“异或”部件的方块图。

第二种并飞“异或”部件 180 及“异或”缓冲器 190 类似于第一种并飞“异或”部件 180 及“异或”缓冲器 190。但是，用第一 SRAM430 代替 CAM330，并且用第二 SRAM440 代替 RAM340。类似于第一种并飞“异或”部件 180 及“异或”缓冲器 190，虽然在概念上“异或”缓冲器 190 包含 SRAM430，但在一种最佳实施方式中 SRAM430 是利用 SRAM 电路实现的，而“异或”缓冲器 190 的其它部件是在 ASIC 下用并飞“异或”部件 180 的其它电路实现的。

第二种并飞“异或”部件 180 的总线接口 410 类似于第一种并飞“异或”部件 180 的总线接口 310, 第二种并飞“异或”部件 180 的定时及控制部件 420 类似于第一种并飞“异或”部件 180 的定时及控制部件 320。类似地, 总线接口 410 和定时及控制部件 420 提供并飞“异或”部件 180 和底层总线 160 之间的接口。该接口包括一组通路: 由底层总线 160 使用的传递控制信号的控制通路 411, 传递具有位 0 到位 31 的总线地址值的总线地址(BA)通路 412, 以及传递具有位 0 到位 31 的总线数据值的总线数据(BD)通路 413。

第一 SRAM430 包括每字为 5 位的 1 兆字的存储器, 第一 SRAM430 还包括: 第一 SRAM 地址输入端口 431; 一个写控制输入端口 432, 用于接收写控制信号; 一个 5 位的第一 RAM 数据端口 433; 以及一个有效控制输出端口 434(一个用于 SRAM430 的附加数据位输出端口)用于提供一个充当有效控制信号的输出数据位。总线地址通路 412 的位 2 至位 21 和 SRAM 多路复用器 435 的第一输入端连接, 总线地址通路 412 的位 12 至位 31 和 SRAM 多路复用器 435 的第二输入端连接, 而 SRAM 多路复用器的输出端 435 和第一 SRAM 地址输入端口 431 连接。

定时及控制部件 420 还从有效控制输出端口 434 接收有效控制信号, 并且提供用于在第二 SRAM440 中寻址的 15 位的 RAM 地址通路(RA)414 和用于控制第一 SRAM430 的写控制通路 415。

“异或”缓冲器 190 包括第二 SRAM440, 后者包括: 一个用于接收 15 位的第二 SRAM 地址值的第二 SRAM 地址输入端口 441, 以及一个提供存储在在第二 SRAM440 的相关单元处的 32 位 SRAM 数据值的第二 SRAM 数据端口 442。类似于 RAM340, 在概念上把第二 SRAM440 划分成 32 个独立的“异或”累加器缓冲区每个缓冲区包含 1K 个 32 位的字(即, 4K 字节)并且在第二 SRAM440 中占据一组单元, 每个单元具有 5 位的“异或”累加器缓冲区号(从 0 到 31), 该号是第二 SRAM 地址的高位。

和使 256 个 20 位的标记组中的每个和一个 5 位的值关联起来不同, 第一 SRAM430 保持一个表, 对于每个可能的 20 位的标记该表具有一条

项、各项的一个有效位以及用于这些带有有效位组的项的 5 位的值。每个 20 位的标记代表存储器 120 里某存储缓冲区中一个可能地址(或空存储缓冲区的地址)的 20 个高位。每个 5 位的值代表一个可能的“异或”累加器缓冲区号。

第一 SRAM 数据端口 433 处的相关 5 位的值耦合到第二 SRAM 地址端口 41, 并提供第二 SRAM440 中寻址字的 5 个高位。总线地址通路 412 的位 2 至位 11(即, 用于对某 4 字节字寻址的 10 个低位)提供第二 SRAM440 中寻址字的 10 个低位, 并且还耦合到第二 SRAM 地址端口 441 上。从而, 第二 SRAM440 在第二 SRAM 地址端口 441 上接收 15 位的地址, 并提供第二 SRAM 数据端口 442 处输出的 32 位数据字。

第二 SRAM 数据端口 442 和第一寄存器 450 连接, 并利用激励器 451 和总线数据通路 413(以及还和第二寄存器 452)连接。第一寄存器 450 以及第二寄存器 452 和“异或”运算器 453 的各输入端连接, “异或”运算器包括一组 32 个的“异或”门, 用于并行地在整个 32 位的字上执行“异或”运算。“异或”运算器 453 的输出端以及一个“零”输入端 454 和“异或”多路复用器 455 的输入端连接; “异或”多路复用器 455 的输出端还和第二 SRAM 数据端口 442 连接。

在备择实施方式中, 第二寄存器 452 可用 FIFO 或其它电路结构替代, 尤其当第二 SRAM440 不能足够快地及时向第二寄存器 452 准备好结果时。

为了如在步骤 221 和步骤 222 中向“异或”累加器缓冲区提供新的 RAID 条, 定时及控制部件 420 使第一 SRAM430 和存储器 120 里某存储缓冲区的某 20 位的存储器地址(或者某空存储缓冲区的某 20 位的存储器地址)连接, 并且把某用于 32 个“异或”累加器缓冲区中的一个缓冲区的 5 位标识符存储到由该 20 位的存储器地址规定的单元中。

在定时及控制部件 420 的控制下, SRAM 多路复用器 435 选出总线地址通路 412 的位 2 至位 21 以连接到第一 SRAM 地址输入端口 431, 存储来自总线数据通路 413 的位 0 至位 4 的总线数据值以作为该单元的新的 5 位值, 并且为该单元设置有效位。处理机 110 也可以出于诊断目的



读第一 SRAM430 中选定的项。

为了如步骤 223 中那样监视底层总线 160, 定时及控制部件 420 选出用于连接到第一 SRAM 输入端口 431 的总线地址通路 412。底层总线 160 上出现的存储器地址(或空存储缓冲区内的地址)的 20 个高位用于对第一 SRAM430 寻址; 当为该单元设置了有效位时, 第一 SRAM 数据端口 433 对和该 20 位的存储器地址关联的某“异或”累加器缓冲区提供 5 位的值。

相关的 5 位值用于选择第二 SRAM440 里 32 个“异或”累加器缓冲区中的一个缓冲区, 而总线地址通路 412 的位 2 至位 11 用于选择第二 SRAM440 里该“异或”累加器缓冲区中的某特定字。逐字地把第二 SRAM440 中的数据值读到第一寄存器, 同时逐字地从底层总线 160 把数据值复制到第二寄存器 452。“异或”运算器 453 对带有从底层总线 160 移动过来的数据的“异或”累加器缓冲区中的数据执行“异或”操作。利用第二 SRAM 数据端口 442 把“异或”运算器 453 的输出回写到第二 SRAM440。

类似于第一种并飞“异或”部件 180, 为了如步骤 224 中那样读出累加的奇偶信息, 定时及控制部件 420 把总线地址通路选择成和第二 SRAM 地址端口 441 连接, 从而顺序地把第二 SRAM440 里的“异或”累加器缓冲区中的每个字读到总线数据通路 413 上, 同时对存储器 120 或者对一个盘驱动器 140 进行 DMA 传送。类似于第一种并飞“异或”部件 180, 出于诊断目的或者如果出于其他原因要读出或操作位于“异或”缓冲器 190 中的奇偶信息, 处理机 110 也可以直接读第二 SRAM440 的项。类似于第一种并飞“异或”部件 180, 定时及控制部件 420 同时把零值写入到第二 SRAM440 里的“异或”累加器缓冲区的各字中, 从而把“异或”累加器缓冲区复位成统一的零值。

类似于第一种并飞“异或”部件 180, 处理机 110 还可以直接对第二 SRAM440 中的选定单元进行写, 从而在“异或”缓冲器 190 的内容和选定的信息之间直接执行“异或”运算。

第二种并飞“异或”部件 180 具有这样的额外特性, 即, 存储器 120 中和“异或”累加器缓冲区相关的存储缓冲区(或空存储缓冲区)的数量不

受限于 CAM330 中项的数量。这允许使用任意宽度的 RAID 条，并且其中可同时使用多达 32 个 RAID 条。通过改变第一 SRAM 数据端口 433 的宽度并改变第二 SRAM 的尺寸，可以修改第二种并飞“异或”部件 180 以改变“异或”累加器缓冲区的数量。

在一种最佳实施方式中，对总线地址通路 412 和总线数据通路 413 所采用的数值如表 4-1 中所示。在表 4-1 中，以十六进制表示地址区间值，并且示出总线地址(BA)通路 312 和总线数据(BD)通路 313 的选定的位区间。

表 4 - 1

地址区间	寄存器	读/写	数据/操作
000000..0000FF	PCI 总线寄存器	读写	按 PCI 总线协议规定
010000..0103FF	表项	读写	读或写表项 BD<0 : 4> “异”累加器号 BD<8> 表项有效 BD<12 : 31> 存储缓冲区地址
100000..10FFFF	“异”累加器缓冲区	读	读并清除“异”累加器缓冲区 BA<12 : 16> “异”累加器缓冲区号 BD<0 : 31> 结果
100000..10FFFF	“异”累加器缓冲区	写	对“异”累加器缓冲区的直接“异” BA<12 : 16> “异”累加器缓冲区号 BD<0 : 31> 用于直接“异”的数据
000000..FFFFFF	“异”累加器缓冲区	X	并飞“异”，清除存储缓冲区

		BA<12 : 31> 要匹配的存储器地址
		BD<0 : 31> 用于并飞“异”的数据
200000..7FFFFFFF	空存储缓冲区	写 并飞“异”，清除空存储缓冲区
		BA<12 : 31> 要匹配的存储器地址
		BD<0 : 31> 用于并飞“异”的数据

### 最佳实施方式

第一种并飞“异或”部件和其相关的“异或”缓冲器 190 以及第二种并飞“异或”部件和其相关的“异或”缓冲器 190 是最佳的实施方式。在研读本申请后，熟练的技术人员清楚，在某些设计限制下第一种并飞“异或”部件 180 和其相关的“异或”缓冲器 190 是最佳的，而在另一些设计限制下第二种并飞“异或”部件 180 和其相关的“异或”缓冲器 190 是最佳的。

例如，在相对小型的但复杂的 CAM330 比额外的相对大的外部 SRAM430 便宜的情况下，第一种并飞“异或”部件 180 和其相关的“异或”缓冲器 190 是最佳的；而在相反情况下或者 RAID 子系统的各条的宽度相对宽的情况下，第二种并飞“异或”部件 180 和其相关的“异或”缓冲器 190 是最佳的。

### 其它的实施方式

虽然本文公开最佳的实施方式，在保持在本发明的内容、范围和精神之内下各种修改是可能的，并且在研读本申请后，对于熟练的技术人员这些修改是清楚的。

图1

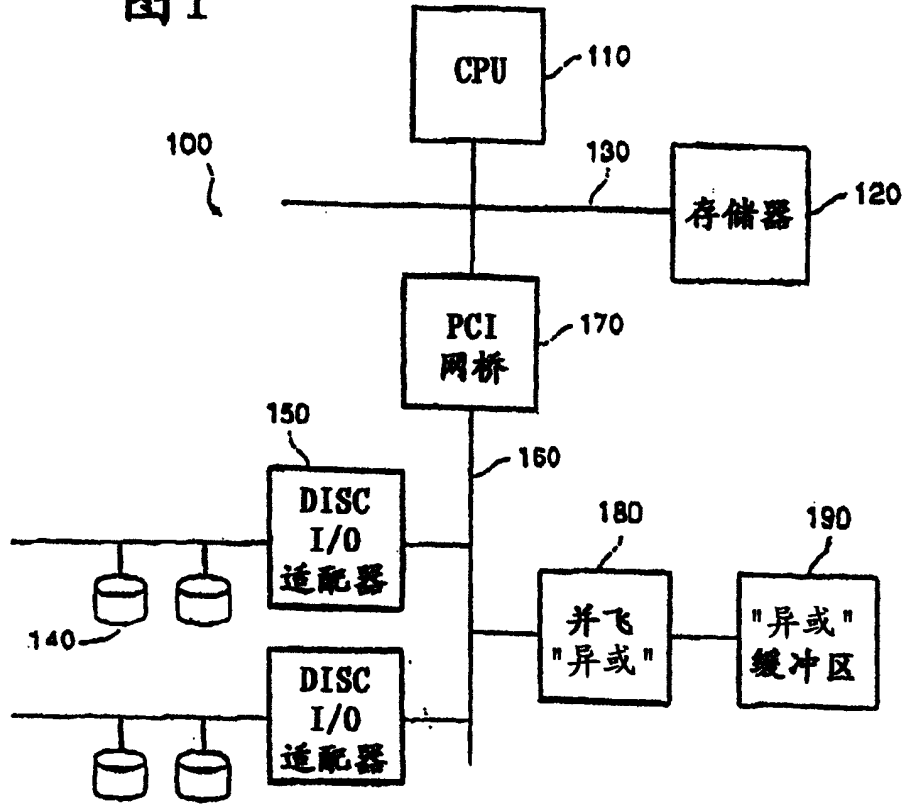
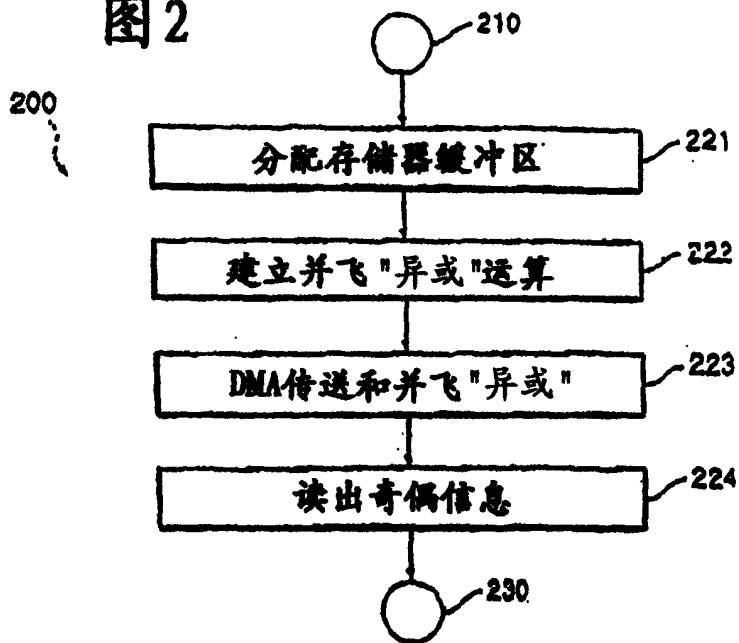


图2



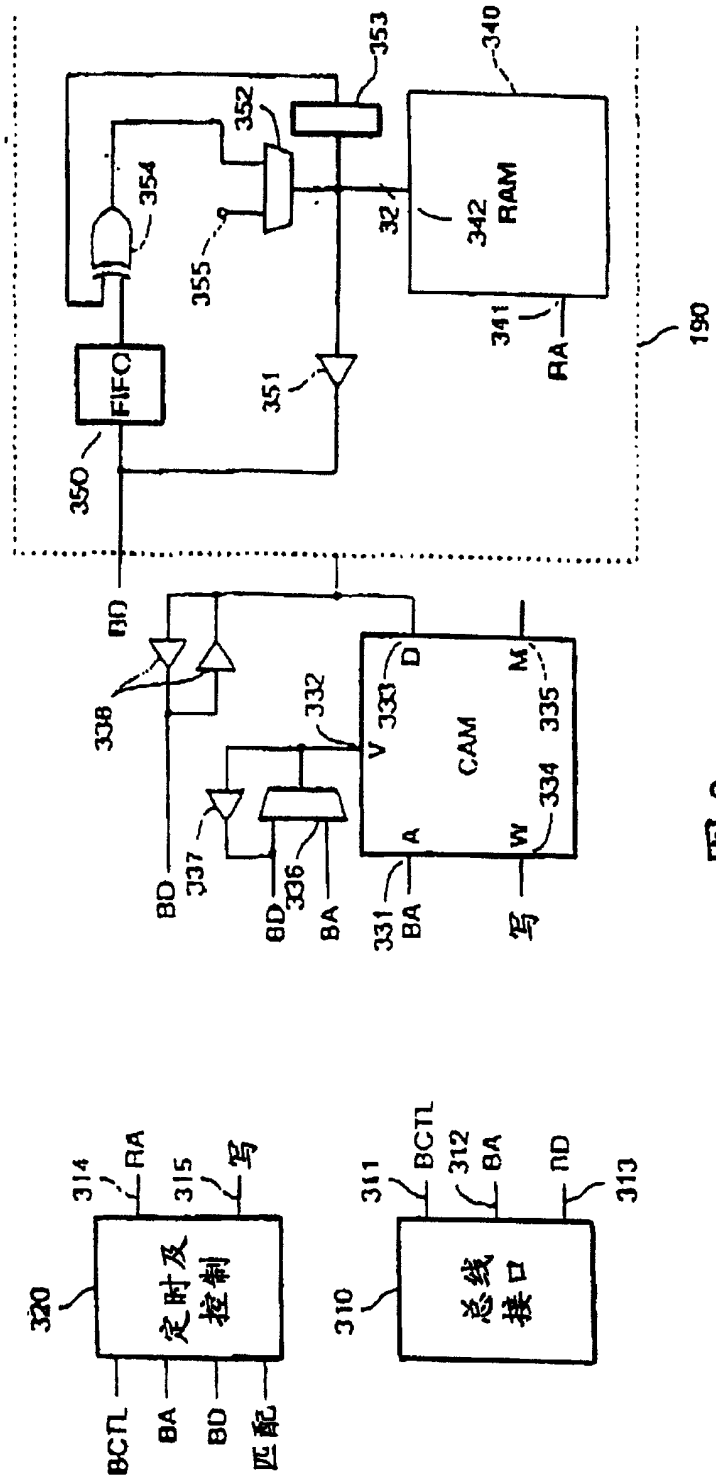


图 3

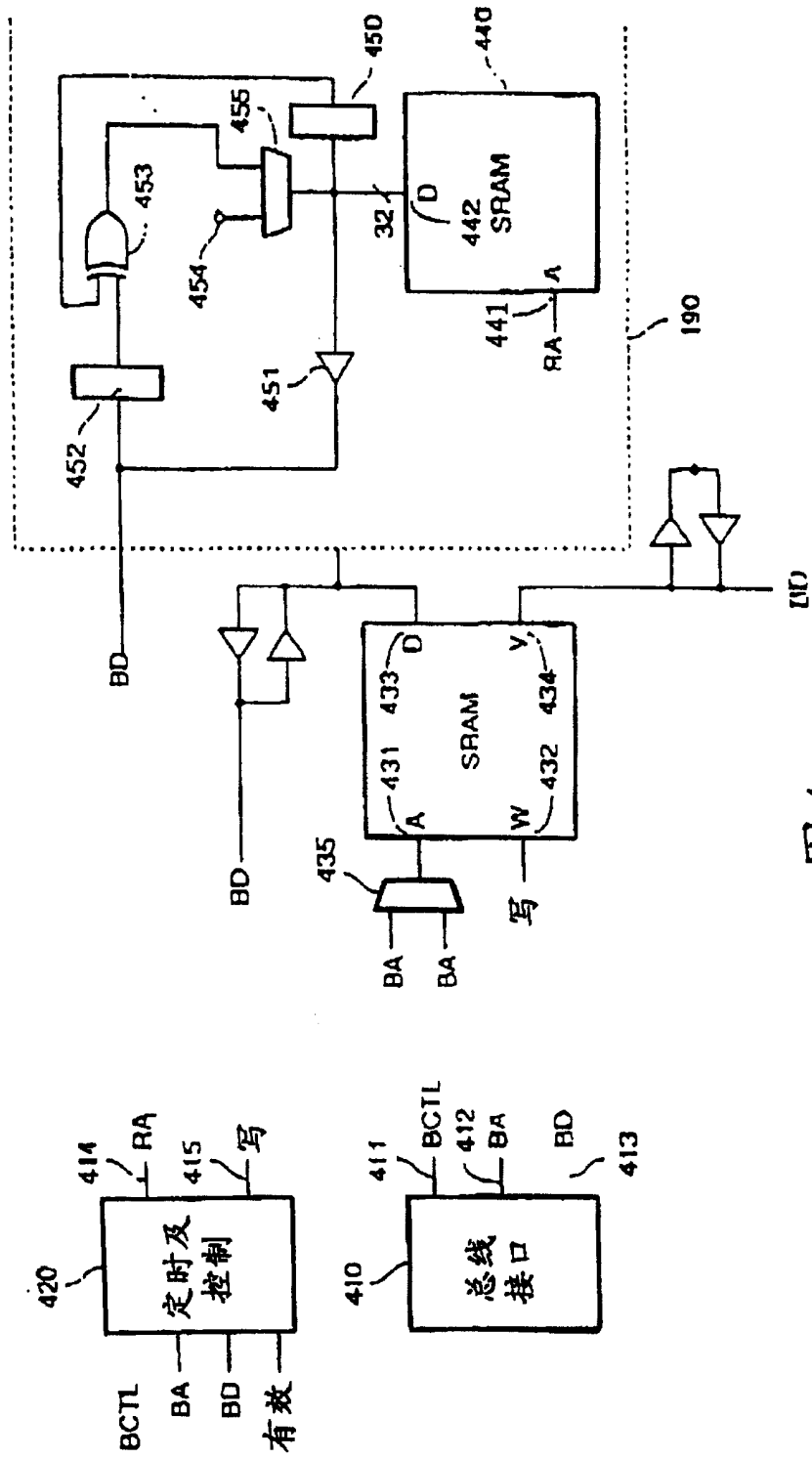


图4