

PCI Express 总线中消息中断的研究

唐雷雷, 贺占庄

(西安微电子技术研究所, 陕西 西安 710065)

摘 要: 本文主要讲述了在第三代 IO 总线 PCI Express 总线中的一种新型的中断机制, 即消息中断机制。这种中断机制不再采用 INTx 并边带信号提交中断, 而是采用带内方式实现的。带内提交中断的方式有两种, 一种是消息信号中断 MSI, 另一种是虚拟 INTx 信号中断。其高效性、可扩展性, 在实际应用中有着非常重要的意义, 也为 PCI Express 成为第三代 IO 总线的主流增色不少。

关键词: PCI Express; 中断; 消息; MSI; INTx

中图分类号: TP336

文献标识码: A

文章编号: 1000-7180(2013)07-0137-04

Research of Message Interrupt for the PCI Express Bus

TANG Lei-lei, HE Zhan-zhuang

(Xi'an Institute Microelectronics Technology, Xi'an 710065, China)

Abstract: This paper describes a new interrupt mechanism named message interrupt in the third generation IO bus called PCI Express. The message interrupt mechanism includes Message Signaled Interrupt (MSI) and Virtual INTx Signaled Interrupt which adopt In-Band way instead of Side-Band way. In addition, because of its high efficiency and extensibility, the message interrupt mechanism is of important significance in practice and enhances PCI Express bus to make it become the mainstream in the IO bus.

Key words: PCI Express; interrupt; message; MSI; INTx

1 引言

中断是为计算机处理紧急事件或非预测事件而设计的, 是重要的 IO 系统, 也是整个计算机系统的必要组成部分^[1]。优良的中断设计, 不仅可以提高计算机的可靠性和安全性, 而且可以提高计算机的工作效率。

PCI Express 总线是继第一代 IO 总线 ISA、EISA、MC 和 VESA 和第二代 IO 总线 PCI、AGP 和 PCI-X 之后的第三代 IO 总线。PCI Express 总线, 简称 PCIe、PCI-E 或 PCI-XP, 是采用点到点互连的串行总线, 与第一代和第二代并行 IO 总线存在根本性的变革^[2-3]。随着 PCIe 总线的提出和广泛应用, 一种新型的中断机制得到了充分认可, 这就是消息中断机制。

2 PCI Express

PCIe 总线与 PCI 总线不同, 在总线技术和结构上有很大的改进, 传输速率更高, 性能更好。它采用点到点的、全双工、串行、差分信号链路来互联设备^[2], 在保证总线带宽的前提下, 比起并行数据信号更加节省信号线。PCIe 总线的点对点连接的串行结构, 使得每一条 PCIe 链路中只能连接两个设备, 这两个设备互为数据发送端和数据接收端。在 PCIe 总线的物理链路的一个数据通路 (Lane) 中, 由发送端和接收端两组差分信号组成, 共四根信号线。一个 PCIe 链路可以有多个 Lane 组成, 目前 PCIe 链路可以支持 1、2、4、8、12、16 和 32 个 Lane, 七种 PCIe 链路^[4-6]。在 PCIe3.0 版本中, 时钟频率提升到 8GHz, 比 1.0 版本的 2.5GHz 和 2.0 版本的 5GHz 提高很多; 同时采用 128b/130b 的编码技术, 取代了 8b/

10b 编码技术,传输效率几乎达到 100%,单通道(x1)单向带宽接近 1GB/s,双向带宽接近 2GB/s. PCIe3.0 各个模式下的双向峰值总带宽可由下面公式计算得出:

峰值总带宽(GBps)= $\frac{8\text{ Gbps} * 2 * \text{链路宽度}}{8\text{ b/B}}$ *

128 b/130 b

3 存储器写事务和消息事务

在 PCIe 事务层生成的数据包叫事务层包(TLP),TLP 是发送设备根据设备核提供的信息在事务层生成,穿过根复合体或交换开关,到达最终目标接收设备的事务层后才得以处理.在传递 TLP 并执行事务时,使用四种独立的地址空间:存储器、IO、配置和消息.

本节主要介绍与消息中断密切相关的总线事务:存储器写事务和消息事务.

通常,一个事务包 TLP 由 TLP Head、Data Payload 和 TLP Digest 组成,其中 TLP Head 格式长 4DW,其通用字段如图 1 所示.

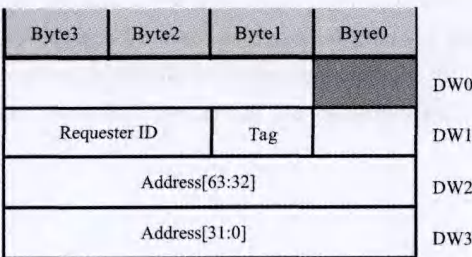


图 1 TLP Head 4DW 头标格式

头标中的 DW0 的 32 位所表示的字段如下: DW0[9:0]为 Length 字段,DW0[11:10]为 AT 字段,DW0[18,13,12]为 Attr 字段,DW0[14]为 EP 字段,DW0[15]为 TD 字段,DW0[16]为 TH 字段,DW0[22:20]为 TC 字段,DW0[28:24]为 Type 字段,DW0[32:29]为 Fmt 字段,其余位保留. DW1[3:0]为 First DW BE 字段,DW1[7:4]为 Last DW BE 字段.

对于消息事务包 TLP,DW1[7:0]为 Message Code 字段,而且对于隐含路由方式,消息事务包头标中的 Address 字段保留.如果此消息事务是 INTx 中断消息,则 Message Code 字段编码如表 1 所示.

另外,INTx 消息包不含有 Data Payload 字段,头标中的 Length 字段保留;TLP Digest 字段是对 TLP 进行 ECRC 计算生成的,用于检查所发送的

TLP 在请求者到完成者的路途是否发生变化.

表 1 INTx 中断消息的编码

Message Code	INTx 中断消息说明
0010 0000b	Assert_INTA
0010 0001b	Assert_INTB
0010 0010b	Assert_INTC
0010 0011b	Assert_INTD
0010 0100b	Deassert_INTA
0010 0101b	Deassert_INTB
0010 0110b	Deassert_INTC
0010 0111b	Deassert_INTD

4 MSI 和虚拟 INTx 信号中断

MSI 即消息信号中断(Message Signaled Interrupt),其本质是存储器写事务.这种中断机制使用存储器写请求 TLP 向 CPU 提交中断请求,无需返回完成 TLP.即 PCIe 设备向指定的一个存储器地址写指定的数据,所写的就是提交中断请求的那个设备所对应的中断向量^[3].指定的存储器地址和数据被分别存放在 MSI 能力寄存器组中的消息地址寄存器和消息数据寄存器中.在启动存储器写事务时,将地址和数据又分别放在图 1 中 TLP Head 和 Data Payload 里面.

PCIe 总线不同于 PCI 总线的共享总线方式,它采用交换式结构,每个 PCIe 设备都设有独立的带宽,极大地提高了总线的利用率. MSI 中断机制不仅可以解决异步中断机制 INTx 所带来的数据完整性问题,消除了虚设读操作,而且由于 MSI 中断机制不必再次识别中断源,在一定程度上提高了中断效率.

虚拟 INTx 信号中断,又叫 Legacy INTx Emulation,即采用 Legacy INTx 方式来模拟 INTx# 信号,用以兼容 PCI 设备^[2].这种中断方式将 INTx# 这个边带信号转换成了带内方式,通过中断消息事务向跟复合体内的中断控制器提交中断请求,以中断消息的形式模仿 PCI 兼容的中断信号 INTx#,传达 INTx# 信号的跳变情况.在 INTx 信号有效时,PCI Express 桥将该信号转化为 Assert_INTx 消息报文;当其无效时,再转化为 Deassert_INTx 消息报文.

图 2 是 PCIe 设备中 MSI 和虚拟 INTx 信号中断的提交路径.

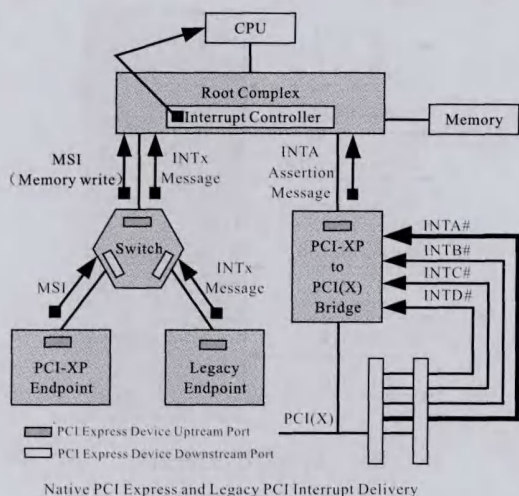


图2 MSI和INTx Message的提交路径

MSI中断机制在纯粹的PCIe设备中必须被使用,在传统端点的PCIe设备中必须被支持,但不一定被使用;而虚拟INTx消息中断在传统端点的PCIe设备中可选,在PCIe-PCI(X)桥中必须被支持.传统端点的PCIe设备是指在对系统初始化和加载OS的过程中所涉及到的PCIe设备,在引导过程中,MSI子系统尚未初始化,不能被使用,所以在OS和引导设备驱动程序安装之前,它们只能使用传统中断方式. PCIe-PCI(X)桥下的PCI(X)设备只能用INTx#信号提交中断,INTx#信号提交给PCIe-PCI(X)以后,由桥再转化为INTx消息^[3].

5 高效性和可扩展性

5.1 高效性

对于MSI中断机制,其高效性体现在中断源和中断请求同时提交给处理器,这样处理器在执行中断服务例程时,不需要再次确认中断源,与传统INTx中断机制相比,提高了中断效率.

对于虚拟INTx信号中断机制,需要两条INTx消息Assert_INTx和Deassert_INTx^[1].设备先发送Assert_INTx消息,指示虚拟INTx信号由高到低的跳变,告知上游设备INTx中断已被肯定;进入中断服务例程以后,设备在发送Deassert_INTx消息,指示虚拟INTx信号由低到高的跳变,告知上游设备INTx中断已被否定,即被释放.虚拟INTx信号中断在消息事务TLP头标的Message Code字段已经指出了中断源,所以也不需要再次确认中断源,同样可以提高中断效率.

PCIe3.0的时钟频率达到8GHz,比PCIe2.0

提升了将近1倍,比PCI和PCI-X提高了几十倍甚至几百倍,所以,不管是MSI还是虚拟INTx信号中断,都大大缩短了中断响应时间,高效性显而易见.

5.2 可扩展性

MSI中断机制最多支持32个中断请求,而且要求中断向量连续,为了扩展中断向量的个数,PCIe总线引出了MSI-X机制. MSI-X中断机制不仅可以支持更多中断请求,更重要是不要求中断向量连续.与MSI中断机制相比,MSI-X中断机制更为合理.

当一个PCIe设备需要多个MSI中断请求时,虽然中断控制器中有足够的中断向量号,但却无法满足这些中断向量号都连续,这就造成了有足够的资源但却无法使用的窘境.在MSI-X中断机制中,每一个中断请求都使用独立的Message Address字段和Message Data字段,可以解决上述问题.

图3(a)和图3(b)分别是MSI Capability的结构图和MSI-X Capability的结构图.

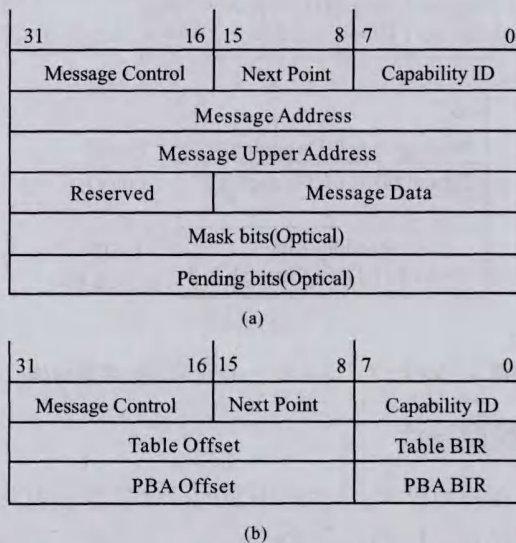


图3 MSI Capability和MSI-X Capability结构

在实际设计中,MSI-X Capability比MSI Capability结构复杂,同时设置了MSI-X Table和Pending Table.在图3(b)中,Table BIR和Table Offset分别表示MSI-X Table存放在哪一个BAR空间中,以及在该BAR中的偏移;PBA BIR和PBA Offset分别表示Pending Table存放在哪一个BAR空间中,以及在该BAR中的偏移.其中Table BIR和PBA BIR由3位组成,000-101与BAR0-BAR5对应.

图4是MSI-X Table和Pending Table组成

结构. 在图 4(a)中, MSI-X Table 存放有 N 组 Entry, 每个 Entry 包含 4 个 DW, 一共 128 位, 存放着每一个对应的 MSI-X 报文使用的数据及其地址, Vector Control 字段只有第 0 位有效, 0 和 1 表示 PCIe 设备能否使用该 Entry 提交中断. 在图 4(b)中, Pending Table 有 $N/64$ 个 Entry 组成, 每个 Entry 有 64 位, 一共有 N 位, 对应 MSI-X Table 中 N 个 Entry. Pending 位要与 Per Vector Mask 位配合使用, 当 Per Vector Mask 位为 1 时, PCIe 设备不会立即提交 MSI-X 中断请求, 而是将对应的 Pending 位置 1; 当系统软件将 Per Vector Mask 位清零时, PCIe 设备需要提交中断请求, 同时将 Pending 位清零^[7-9].

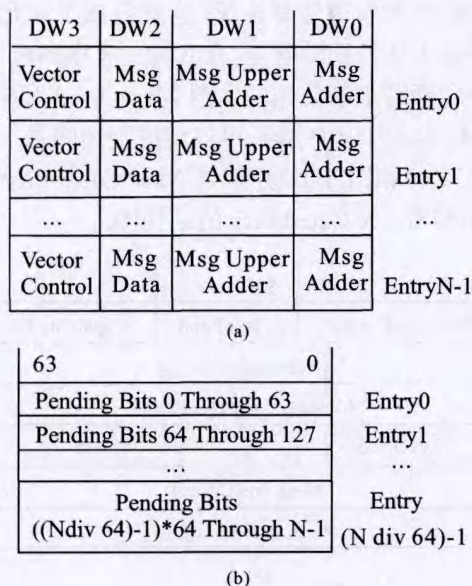


图 4 MSI-X Table 和 Pending Table 组成结构

6 模型设计

在模型设计上, 消息中断的设计模型分为两个部分: MSI Handler 和 INTx Handler. 如图 5 所示.

msi_req 和 msi_ack 分别是 MSI 中断请求和响应信号; msi_tc[2:0] 表示传输类别, 决定了给定事务在 PCIe 线路结构中的相对优先级; msi_num_app[4:0] 是 MSI 的中断向量; msi_num_pex[4:0] 是电源管理和热插拔引起的 MSI 的中断向量, 是一个固定值; msi_cfg[31:0] 用于配置 MSI 中断控制状态寄存器; int_sts 用于指示 INTx 中断的肯定和否定, 它将导致 Assert_INTx 消息 TLP 和 Deassert_INTx 消息 TLP 的产生; int_ack 是 INTx 中断响应信号; int_code[7:0] 用于传送 INTx 中断消息代码. 图 6(a)和图 6(b)分别是 MSI 和 INTx 中断时序波形图.

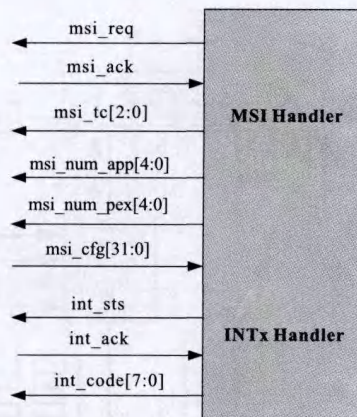


图 5 消息中断模型设计

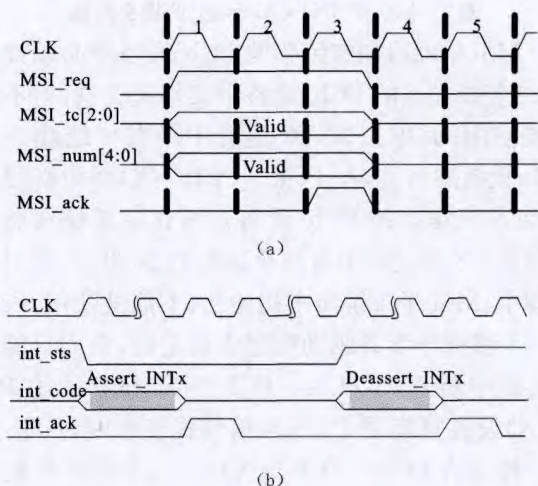


图 6 MSI 和 INTx 中断时序波形图

7 结束语

消息中断机制虽然产生于 PCI 总线中, 但其作用却在 PCIe 总线中得到了充分发挥, PCIe 总线所蕴藏的强大生命力, 使得消息中断的研究意义更加长远.

参考文献:

- [1] 仇玉章. 32 位微型计算机原理与接口技术[M]. 北京: 清华大学出版社, 2000.
- [2] 王齐. PCI Express 体系结构导读[M]. 机械工业出版社, 2010.
- [3] 马鸣锦, 朱剑冰, 何红旗, 等. PCI、PCI-X 和 PCI Express 的原理及体系结构[M]. 北京: 清华大学出版社, 2007.
- [4] Altera Corporation. PCI express compiler user guide 7.1[S]. 2007.

(下转第 144 页)

人脸识别的影响^[8],将其用于人脸识别能显著提高系统的识别率和鲁棒性,有效解决了训练样本较少的问题,当每人的训练样本数为 1 时,仍能达到 90% 的识别率和 172ms 的识别时间,满足多数实际应用需求,具有一定的实用价值。

参考文献:

- [1] 苏剑波,徐波.应用模式识别技术导论:人脸识别与语音识别[M].上海:上海交通大学出版社,2001.
- [2] 王衍,胡金演,杨慧,等.基于 Gabor 小波变换复频域相应的人脸识别研究[J].视频应用与工程,2010,34(07):104-10.
- [3] 俞燕.基于特征的 Gabor 小波人脸识别算法改进[D].镇江:江苏大学,2010.
- [4] 胡月.基于主成分分析和独立成分分析的人脸识别研究[D].长春:吉林大学,2010.
- [5] Lades M, Vorbruggen J C, Buhmann J. Face recognition by independent component analysis [J]. IEEE Transactions on Neural Networks,2002,13(6): 1450-1464.
- [6] Phillip s J P, Moon H, Rizvi A S, et al. The FERET evaluation methodology for face recognition algorithms [J]. IEEE Trans Pattern Analysis and Machine Intelligence,2000, 22 (10):1090-1104.
- [7] Qin Zhao, Jun Gao. Image Matching based on Hybrid Genetic Algorithm[C]// Proceedings of The National Conference On Neural Network, He Fei, IEEE,2005, 174-178.
- [8] 田晓亮.基于 Gabor 小波变换和子空间的人脸识别技术研究[D].太原:太原理工大学,2010.

作者简介:

吴延海 男,(1957—),教授.研究方向为信号与信息处理.
梁文莉 女,(1987—),硕士研究生.研究方向为数字图像处理.

(上接第 140 页)

- [5] Altera Corporation. IP compiler for PCI express user guide 11.0[S]. 2011.
- [6] PCISIG. PCI express base specification revision 3.0 [S]. 2010.
- [7] Pavi. Budruk, Don. Anderson, Tom. Shanley. PCI express system architecture[M]. 2003.
- [8] 田玉敏,王岗,张波,译. PCI Express 系统体系结构 [M]. 北京:电子工业出版社,2005.
- [9] 刘晖,冀然然,夏意军,译. PCI 系统结构[M]. 北京:电子工业出版社 2000.

作者简介:

唐雷雷 男,(1986—),硕士研究生.研究方向为计算机控制技术.
贺占庄 男,(1962—),研究员,博士生导师.研究方向为计算机控制技术.