Techniques for Increasing PCI Performance

Revision 1.0

Release Date: September 10, 1997

These techniques should be used in implementing PCI devices that require large PCI bandwidth on Intel's Pentium® Pro processor and Pentium® II processor platforms

Implementation of these techniques allow for optimal system performance as well as high PCI to memory bandwidth

Information in this document is provided in connection with Intel products. No license, express or implied, by estoppel or otherwise, to any intellectual property rights is granted by this document. Except as provided in Intel's Terms and Conditions of Sale for such products, Intel assumes no liability whatsoever, and Intel disclaims any express or implied warranty, relating to sale and/or use of Intel products including liability or warranties relating to fitness for a particular purpose, merchantability, or infringement of any patent, copyright or other intellectual property right. Intel products are not intended for use in medical, life saving, or life sustaining applications.

Intel disclaims all liability, including liability for infringement of any proprietary rights, relating to use of information in this specification. Intel does not warrant or represent that such use will not infringe such rights.

Intel may make changes to specifications and product descriptions at any time, without notice.

The 82440FX PCIset and the 82440LX AGPset may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available upon request

Contact your local Intel sales office or your distributor to obtain the latest specifications before placing your product order.

Table Of Contents

Revision History

Date of Revision	Version	Description
September 10, 1997	1.0	Initial Release

Reference Documents And Information Sources

Document Name or Information Source	Available From
“Efficient Use of PCI” Technical Paper	www.intel.com
Pentium ® Processor Data Sheet	www.intel.com
Pentium ® Pro Processor Data Sheet	www.intel.com
Pentium ® II Processor Data Sheet	www.intel.com
440FX Data Sheet	www.intel.com
440LX Data Sheet	www.intel.com
PCI Specification	www.teleport.com/~pcisig/

Overview

The Intel Pentium® Pro and Pentium® II processors implement several new features to maximize available host bus bandwidth including pipelined transactions and support for deferred transactions. In order to optimize the overall system performance with these new features, the 440FX PCIset and 440LX AGPset have enhanced their main memory subsystems over the Intel 430 PCIset family. This enhancement allows for concurrent main memory accesses by both the processor and system buses (PCI and AGP). This concurrent operation requires PCI devices to operate in a more efficient manner than might have been necessary with a Pentium® processor and Intel 430 PCIset based system.

This document is being provided to assist Independent Hardware Vendors (IHVs) in optimizing their PCI devices to operate with the Intel 440FX PCIset and 440LX AGPset products.

PCI Command Usage

For optimal PCI performance it is recommended to always perform a transaction that is a cache line (CL) or larger in size and is CL aligned.

PCI devices should use Memory Read (MR) for PCI reads from main memory that are less than a single CL (8 DWs) and confined within a CL. For reads equal to or larger than a CL (≥ 8 DWs), use Memory Read Multiple (MRM) or Memory Read Line (MRL). These commands are aliased in Intel 440FX PCIset and 440LX AGPset. Additionally, these commands should be used for transfers that are not CL-aligned and cross the CL boundary. A single snoop is performed on the host bus for MR. Processor cache snoop ahead and memory read prefetch is performed for MRM and MRL commands. This results in a minimum of 2 complete CLs being read from memory. If MRM or MRL is used for transfers of less than a CL or back-to-back transactions, then host bandwidth is wasted due to the overhead of snoop aheads and memory efficiency is lowered since a portion of the data will be discarded. In these cases, MR should be used. For longer size bursts it is recommended to use MRM or MRL since snoop ahead would provide better overall bandwidth and data streaming from memory will be performed.

For PCI writes to main memory ensure that the burst size for PCI traffic is a minimum of one CL and is CL-aligned to allow for data streaming and to minimize the overhead of resultant host snoop bandwidth.

PCI Addressing

Non-CL aligned data transfers result in lower bandwidth since the Intel 440FX PCIset and 440LX AGPset are optimized for CL aligned addressing. In the case of PCI writes to memory, the efficiency of the buffering within the 440FX or 440LX can be maximized by doing complete CL transfers.

PCI Data Transfer

PCI devices should use bus mastering to do data transfers rather than using the CPU in programmed IO (PIO) mode. The Intel 440FX PCIset and 440LX AGPset have been optimized to provide high PCI bandwidth for bus mastering devices.

PCI Device Buffering and Local Storage

It is essential that local buffering is used on PCI peripheral devices. This is especially important for devices that are planning to do large isochronous data transfers. The 440FX PCIset and 440LX AGPset provide host and PCI bus concurrency to main memory to allow for an increase in overall system performance. The result is that a PCI peripheral device cannot assume the memory to be a dedicated resource even while it is actively transferring data on PCI. In order to work in a concurrent system, these devices must add local storage in order to avoid over runs and under runs. The amount of local storage required is dictated by the system configuration, active software applications, and application of the PCI peripheral device.

Indication of Transfer Complete

Many PCI peripheral device software drivers today use CPU polling of a busy bit that is an I/O space address or is located in non-cacheable PCI memory space (memory-mapped I/O) to determine when the PCI device has completed its transfer. In a concurrent environment, this polling could itself interfere with the completion of the PCI data transfer. The 440FX PCIset and 440LX AGPset allow the processor a time "window" within which to enter its memory and PCI usage requests between PCI requests. Frequent polling cycles accessing non-cacheable memory (main memory or PCI memory) or IO space (PCI) can severely limit the effective PCI bandwidth available to the PCI peripheral device by toggling the accesses between the processor and PCI peripheral device.

An improved method is to use a semaphore located in cacheable memory. Using this semaphore requires that the PCI peripheral device be capable of writing a DWORD to memory indicating completion of the data transfer. This semaphore must be located in the cacheable memory space. This allows the processor to poll this bit in its cache as long as the transfer is in progress without wasting any host or PCI bandwidth in the process. When the transfer is completed, the PCI peripheral device will set the semaphore (write the DWORD to memory) and a snoop on the host bus will update the cache and consequently indicate to the processor that the transaction has completed.

Another method is for the PCI peripheral device to assert an interrupt upon completion of the data transfer.

Back-to-Back Transfers

The Intel 440FX PCIset and 440LX AGPset incorporate a Multiple Transaction Timer (MTT) which allows multiple back-to-back transfers to non-consecutive address locations. System PCI performance can be "tuned" by programming the MTT to a system optimized value. Using the MTT requires that the PCI peripheral device hold its PCI request active. The 440FX and 440LX also park the PCI bus at the last PCI peripheral device that completed its transfer if no other PCI bus requests are active. This allows for expedited back-to-back requests when the host bus is idle. System OEMs and motherboard manufacturers typically set the MTT to a default value. The BIOS Setup routine is the most likely method to be used to change the system MTT. Some manufacturer's may not allow changes to the MTT value to be made. Contact your motherboard or system manufacturer for information concerning changing the system's MTT.

High Host Bus Utilization

The Pentium Pro and Pentium II processor's caches allow for a high cache hit rate for normal workloads and provide very high throughputs while maintaining a low host bus utilization. This increases the availability of the host bus for cache snoops caused by PCI transfers to or from main memory. Consequently, a high bandwidth is supported on both the host and PCI bus.

In some cases, very high demand for main memory may be observed by both the processor and the PCI bus. In these cases, the 440FX and 440LX allow the processor to access memory concurrently with the PCI devices resulting in throttling of the PCI accesses until the host bus utilization is reduced. The processor performance allowed for a given load of PCI transactions is much higher in 440FX and 440LX than in the Intel 430 PCIset family due to this concurrency.

Write Combining and Graphic Accelerators

The Pentium Pro and Pentium II processors support different memory types. The memory type can be defined by programming an associated Memory Type Range Register (MTRR) in the processor. The Write-Combine (WC) memory type allows speculative reads and the memory model assumes weak ordering. Writes to WC memory can be buffered and combined in the processor's internal write-combining buffers. WC writes would result in cache line transfers on the host bus and allows for data streaming on the PCI bus. This transfer type is optimal for frame buffer write accesses and allows for very high throughput from the processor to the linear frame buffer (typically located on PCI for 440FX based systems and AGP for 440LX based systems). Other PCI devices that can accommodate out-of-order transactions should define their local on-board memory to WC type to take advantage of bursting on the host bus and the PCI bus.

Master Initiated Wait States

The PCI master must start an access after its GNT# signal has been asserted and the bus is in an idle state. The PCI specification, Rev.2.1, allows the PCI arbiter (located in 440FX or 440LX) to remove the device's GNT# signal at any time after GNT# has been asserted and the PCI master has not started a transaction. This is done in order to service a higher priority PCI device.

A PCI master that has requested use of the bus and does not assert FRAME# when the bus is in the idle state and its GNT# is asserted faces the possibility of losing its turn on the bus. The Intel 430 PCIset family removes GNT# in order to service a higher priority agent if the master does not assert FRAME# within 8 PCI clocks. This has been reduced to 5 PCI clocks on Intel 440FX and 440LX products.

If PCI masters require the insertion of wait states, they should use IRDY# to indicate master initiated wait states rather than delaying the assertion of FRAME# upon receipt of GNT# from the arbiter. The excessive use of PCI master initiated wait states will negatively impact the performance of the system however and should be avoided.

Summary

New applications are continually being developed which require higher system performance. As the system performance requirements increase, all components of that system must operate in a more efficient manner to ensure that they can meet the needs of the application programs. Intel continues to make enhancements to its products to increase this overall system performance. These guidelines are provided to assist other component manufacturer's in their development of products for these higher performance systems.