

PCI Express Overview

Introduction

This paper is intended to introduce design engineers, system architects and business managers to the PCI Express protocol and how this interconnect technology fits into today's environment.

This technology brief will first introduce the problems that exist with interconnects today, and then discuss how the PCI Express protocol addresses these issues. We then will present an overview on the PCI Express architecture, and give an example of how PCI Express would operate. We will then wrap up with how this protocol addresses various market needs, as well as how to rationalize when to use PCI Express and Advanced Switching.

The Problem Today

Today's system designers are facing a challenging dilemma. They are expected to produce a higher performance solution with more features at the same cost as their current solutions.

And, by the way, they need to do it in less time.

Traditionally, engineers could satisfy their bandwidth requirements by migrating to the next generation PCI. This was possible because the PCI governing body, the PCI SIG, actively managed the roll out of higher performance PCI revisions to meet market demand. Further, because the next generation PCI was code compatible with legacy PCI, engineers could reuse their PCI code to reduce their design time. This was an effective strategy for several generations.

Recently, however, designers have been running into a bottleneck. In order to achieve greater throughput, the PCI community defined wider busses. Including the data bus and sideband control signal, 64-bit PCI exceeds 100 pins. Even at 64-bits and 133 MHz, architectural problems remained because the higher clock rates came at the cost of reduced fanout and the bandwidth still needed to be shared over multiple devices running at ever higher data rates.

As a result many vendors began migrating to either proprietary or marginally supported protocol standards in search of a switched interconnect. While this addressed the architectural problems, it forced engineers to create new software, which slowed development. Further, because of the smaller market adoption, aggressive cost reductions were not possible.

PCI Express Architecture

To address these needs, the PCI SIG adopted a specification called 3GIO from the Arapahoe group, and renamed it "PCI Express".

The PCI Express protocol was created to address the following design goals:

- Support multiple market segments
- Boot existing OS without change
- Scalable performance
- Advanced features including QoS, Power management, and data integrity

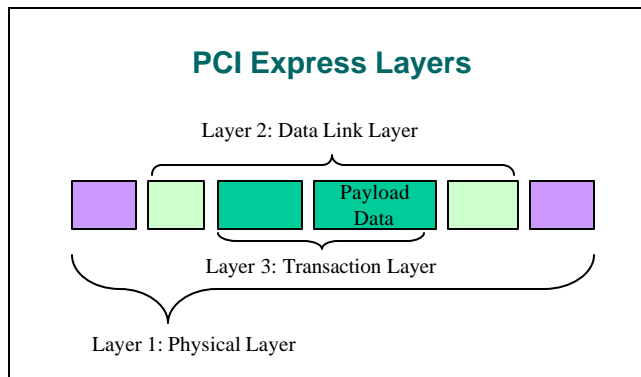
PCI Express technology is a low cost, highly scalable, switched, point-to-point, serial I/O interconnect that maintains complete software compatibility with PCI. It transfers data at 2.5 Gb/s¹ per lane, per direction, and will scale

¹ Like all protocols, those using PCI Express will need to allocate some bandwidth to overhead, including management structures such as DLP (discussed below), as well as packet specific structure such as CRCs, 8b10b encoding, and header processing. For a full dissertation on this topic, please go to www.plxtech.com and download our whitepaper on the topic.

proportionately for performance by adding lanes to the link.

Explaining the Layers

To achieve code compatibility with PCI, PCI Express does not modify the transaction layer. This is significant, because it allows vendors to leverage their existing PCI code to achieve not only a faster time to market, but by using their proven design, provides for a more stable and mature platform. PCI Express does, however, modify layers 1 and 2 of the OSI model.



Layer 1

Layer 1, or the Physical Layer, defines the electrical characteristics of PCI Express. The basic transmission unit consists of two pairs of wires, called a “lane”. Each pair allows for unidirectional data transmission 2.5 Gbps, so the two pairs combined provides 2.5 Gbps full duplex communication, without the risk of transmission collision.

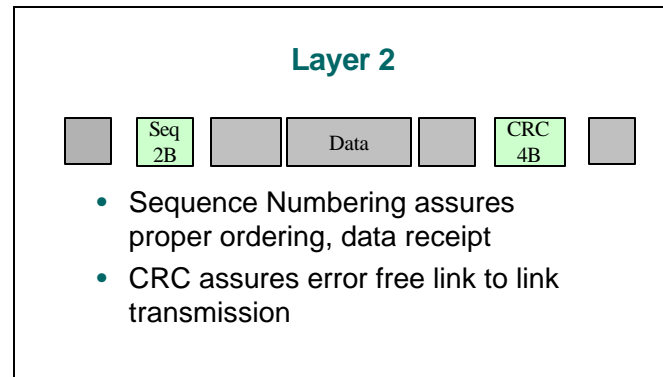
Lanes can be concatenated to provide scalable performance; so for example, combining two lanes will yield 5 Gbps throughput. The PCI Express specification allows for x1, x2, x4, x8, x12, x16, and x32 lanes, which means that the current PCI Express scales up to 80 Gbps per port in each direction.

To increase noise resistance, PCI Express utilizes differential current mode signaling, CML. The clock is embedded using familiar 8b10b encoding.

Early on, the PCI Express architects realized that performance increases to the individual lanes would also be necessary. They therefore created the architecture so that the PHY layer could be easily swapped out for higher speeds without seriously affecting the upper layers.

Layer 2

Layer 2, or the Data Link Layer, defines the data control for PCI Express. The primary task of the link layer is to provide link management and data integrity, including error detection and correction. The function of this layer is to calculate and append a Cyclic Redundancy Check (CRC) and sequence number to the information sent from the data packet. The sequence number allows proper ordering of the data packets. The CRC verifies that data from link to link has been correctly transmitted².



Layer 3

Layer 3, or the transaction layer, connects the lower protocols to the upper layers. The primary value to end users is that this layer appears to the upper layers to be PCI.

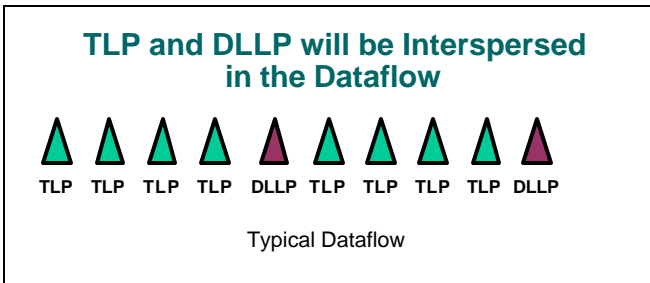
² For those customers who wish to ensure end to end data integrity, there is an optional feature in the specification to achieve this. This feature is called “TLP Digest” and is defined in Layer 3.

The Transaction layer packetizes and then prepends a header to the payload data. Read and write commands, as well as prior side-band signals, like interrupts and power management requests, are also included in this layer.

PCI Express Packets in Action-DLLP and TLP

One of the key advantages is that unlike PCI, PCI Express does not utilize sideband signaling; rather, all information is transmitted in-band which reduces pin count. This includes the clock, system status, and power management.

As mentioned earlier, the links carry both payload data (called Transaction Layer Packet, or TLP) as well as management data (called Data Link Layer Packet, or DLLP). Both payload data and management data will be interspersed within the link.

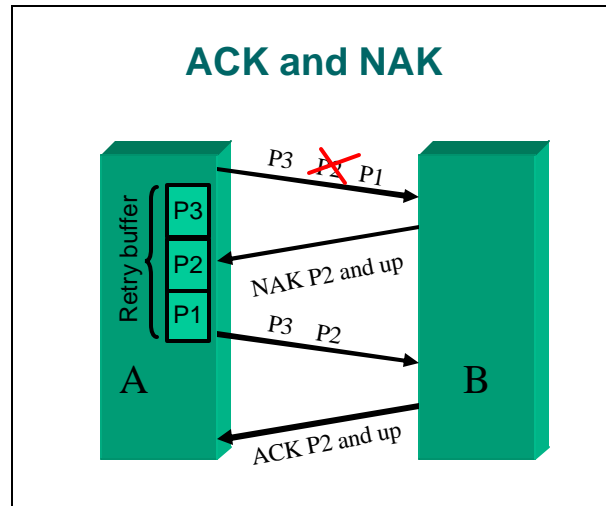


Management data contained in the DLLP includes traditional PCI functions such as power state control as well as new PCI Express specific system functions such as flow control, and packet acknowledgement.

In order to maximize interconnect efficiency, the PCI Express architecture employs a mechanism called Flow Control. Using the DLLP, each link communicates the amount of resources it has available to receive data. The transmitting device will only send its

data when the receiving device has enough resources available to accept the entire packet.

In addition, PCI Express verifies data integrity at each hop. This is accomplished through packet acknowledgement, or ACK/NAK. The receiving device will, using the DLLP packets, inform the sending device whether it has correctly received the packets. The transmitting device will then resend those packets not correctly received.

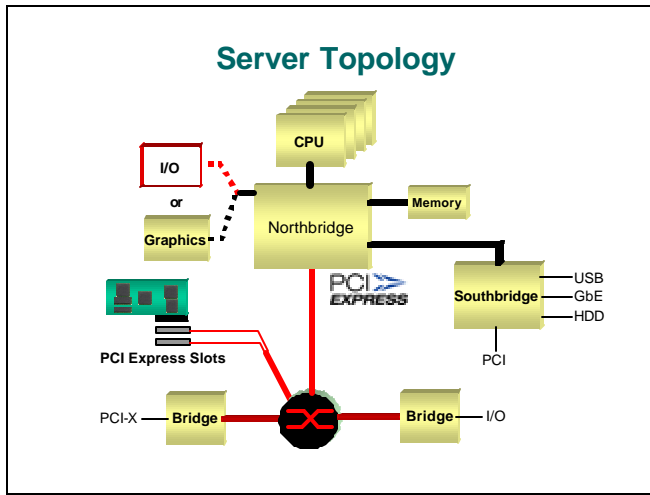


The PCI Express protocol specifies a special packet to transmit payload data. This packet is called Transaction Layer Protocol or TLP. In addition to the actual data, the TLP adds a header that carries information such as packet size, message type (memory, I/O, or configuration), traffic class for QoS and any modifications to the default handling of the transaction (for example, relaxed ordering, or snooping)

Applications Using PCI Express

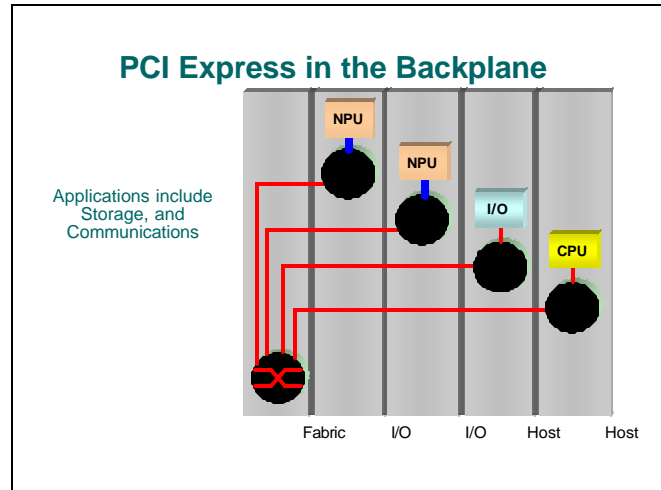
PCI Express presents a powerful value proposition. It provides not only a scalable, full featured, high performance platform, but also one designed to thrive in the high volume, rapid turning, and cost sensitive markets.

Certainly, one of the first segments to adopt PCI Express will be the desktop and server marketplaces. System integrators in this space will value PCI Express' bandwidth flexibility and low pin count. Most importantly, however, these vendors will value the PCI code compatibility, which will reduce their development cost.



Protocol bridges will be introduced that will allow OEMs to leverage their legacy designs into the new PCI Architecture and therefore gain a time to market advantage.

These benefits are not limited to desktop and servers. Other markets, including storage, communication and industrial control will also value these features. These markets will implement PCI Express, not only on the motherboard but also into the backplane. Many of these applications will migrate not only to utilize the PCI code compatibility and low pin count, but also to make use of the bandwidth, scalability, and advanced features.



Using Non-Transparent Bridging for Multiprocessor System

Multiprocessor systems have become omnipresent in today's products. Almost every application requires multiple processors, either to allow for intelligent add-in cards, or greater processing power, or higher reliability or some combination of all three.

There are three major applications for non-transparent bridging: intelligent adapters, host failover and multiprocessor systems.

Intelligent adapters are typically add in cards that employ a processor on the card to offload the host. A good example of an intelligent add in card would be a storage HBA card, but certainly almost every market niche requires that some cards would have a support microprocessor on board. In order for PCI Express to be successful, it needs to provide a mechanism for intelligent adapters.

Another popular usage for multiprocessor systems would be to provide a high reliability system. High reliability is by achieved with a secondary host processor. After enumeration, the secondary host would monitor the system state, including the health of the primary processor. When the secondary host determines that the primary host

has failed, it would assume control of the system by promoting itself to the primary host, and restoring the system to an operational state.

Applications that would need this include storage and communications systems, where 99.999% uptimes are required.

Another popular application for multiprocessor systems is for computational bandwidth. Blade servers exemplify this need. By providing a mechanism to pass data among arrays of processors, PCI Express provides both the performance and economic features necessary for this market segment.

To accommodate multihost systems, many will utilize non-transparent bridging. Non-transparent bridges logically isolate processors. This allows multiple processors to co-exist in the system. While there are many means to implement multiprocessor systems, non-transparent bridging is attractive because it is a methodology proven in the PCI space, and is non-proprietary.

PCI Express and Advanced Switching

There has been quite a bit of confusion on the features and capabilities of PCI Express and Advanced Switching, so it makes sense to compare and contrast the two protocols.

The goal of the Advanced Switching specification is to build upon the PCI Express specification and provide new features not previously possible with the standard PCI architecture. This is accomplished by relaxing the requirement to be PCI compatible. These new features include multicast as well as congestion management, both of which are desirable as

the system bandwidth utilization begins to reach 100%.

Furthermore, Advanced Switching provides protocol encapsulation. Like its name implies, protocol encapsulation allows Advanced Switching to route any protocol by wrapping it in an AS packet.

These features make AS an ideal protocol for those who are designing large systems, and who require those benefits not found in PCI Express.

When Advanced Switching products begin to roll out, it is expected that system designers will mix AS with PCI Express in their systems. Specifically, they will employ AS only where they require the additional functionality. They will reuse their PCI Express designs to take advantage of the low cost, and proven designs. Because PCI Express and Advanced Switching share the lower layers, one can expect an easy migration.

Conclusion

PCI Express continues to grow in interest and maturity. This phenomenon even will accelerate past its current frenetic pace as products begin to roll out. PLX hopes that this document is a helpful introduction to this exciting technology.

For more detailed information, please visit us at www.plxtech.com or send us an email at pciexpress@plxtech.com.