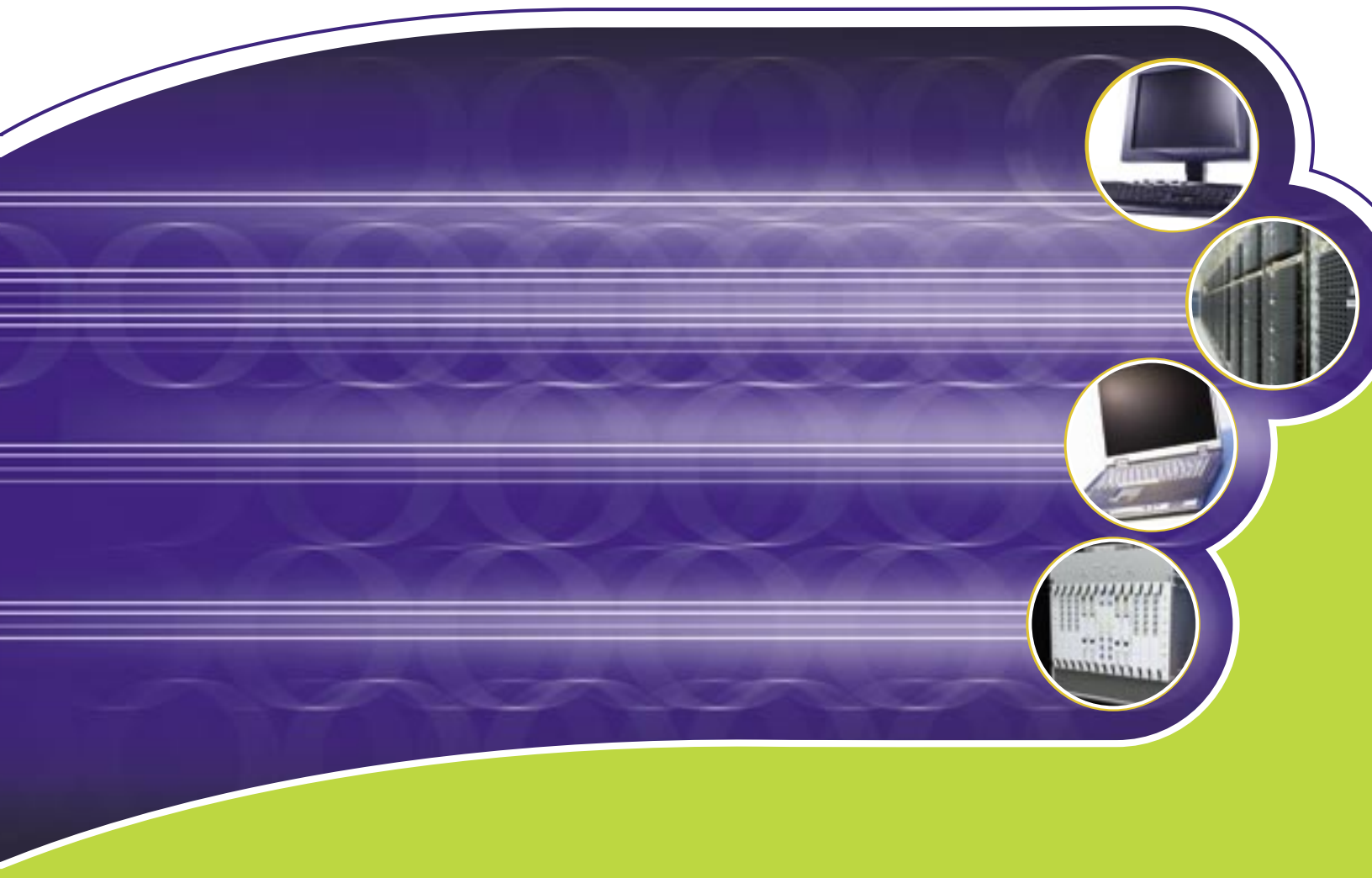


The Benefits of PCI Express Architecture for Components and Systems

PCI EXPRESS® TECHNOLOGY
get in the express lane

desktop • enterprise • mobile • communications

Contents

Introduction

The converging computing and communications industries are on the threshold of a parallel-to-serial interconnect transition. This trend, observed in many application areas, has been gaining momentum since the introduction of the PCI Express Architecture. In this paper, we will examine some of the benefits of designing components and systems using this technology. Some of these benefits include:

- Reduced cost and design complexity
- Robust and feature-rich solutions
- Designs that will meet the demands of applications for the next decade and beyond
- Rapid investment recapture through the marvel of volume economics

The Requirements

Today's computing platforms (desktops, servers, etc.) need an architectural overhaul to mend long-standing performance bottlenecks, prohibitive costs, a dearth of advanced features and implementation complexities. For the past several years, we have witnessed improvements in several areas but these have been isolated to the various components and have not yielded proportional benefits to the platform. The table below illustrates the point:

Platform Components	2000	2003
CPU	733MHz	3+GHz
FSB	133MHz	533+MHz
Memory	133 SDR	266 DDR
Networking	100Mb	1-10Gb
Storage	80MB/s	320MB/s
General-purpose I/O	500MB/s (PCI 64/66)	1GB/s (PCI-X 64/133)

As can be seen, there have been significant improvements in virtually all of the above categories, but general-purpose I/O lags the rest by at least a factor of 2. To reap the benefits of balanced computing, it is crucial to address the cost-effective scalability of general-purpose I/O with a solution that provides performance and advanced features.

The Requirements Answered

PCI Express Architecture is a state-of-the-art serial interconnect technology promoted by the PCI-SIG that delivers performance headroom and advanced platform features to track projected processor and memory subsystem improvements over the next decade. Its release at 0.8V and current 2.5GHz signaling rate supports configurations consisting of 1, 2, 4, 8, 12, 16 and 32 lanes which can yield up to 16 Giga Bytes per second of bandwidth. Future frequency increases promise to scale total bandwidth to the limits of copper. Even higher bandwidth can be achieved with other physical media without impacting levels above the Physical Layer in the protocol stack, thus allowing PCI Express technology to continue to seamlessly evolve well into the next decade and beyond.

The PCI Express Architecture retains the PCI usage model and software interfaces to facilitate a smooth migration from existing PCI based designs. The technology is suitable for multiple market segments and supports chip-to-chip, board-to-board and adapter solutions at an equivalent or lower cost structure than existing PCI designs. Investment preservation is maintained through backwards compatibility to existing PCI software as well as headroom for performance scalability in both interconnect width and frequency as required.

The key features of PCI Express technology are:

- PCI software compatibility
- Scalable performance for multiple computing and communications market segments
- Suitable for chip-to-chip, board-to-board, add-in peripherals and backplane implementations
- Support for end-to-end data integrity to achieve highly-available solutions
- Advanced error reporting and handling for fault isolation and system recovery
- Native power management functions for flexible platform power budgeting
- Inherent hot plug and hot swap capabilities with compatible support for existing PCI hot-plug schemes
- Low-overhead, low-latency data transfers and maximized interconnect efficiency
- Differentiated services through isochronous data delivery for high bandwidth applications
- Multi-hierarchy and advanced peer-to-peer communications across fabric topologies
- High-bandwidth, low pin-count implementations for optimized performance
- Cost effective silicon component designs relative to package and Die Area

Assumptions

This paper highlights the benefits of PCI Express Architecture in the implementation of silicon components and computing systems. When discussing a component design, a 0.13_m process technology is assumed. This paper uses data from laboratory observations, simulations and modeling exercises. The comparisons below are between two devices: one that implements a PCI Express x4 link and the other that implements the legacy parallel PCI-X 2.0 bus operating at 266MHz (DDR). These two interconnects provide roughly the same aggregate bandwidth (2GB/s); however the PCI Express interconnect has many more state-of-the-art features that are simply not available on the legacy parallel PCI-X 2.0 bus. Detailed PCI-X 533 (QDR) comparisons must await the completion of that specification.

Components Designs

In designing components it is useful to characterize three separate interface design parameters such as:

- Package
- Die area
- Power

Note that the overall pin count, die size and power dissipation requirements of a given silicon component depends on its functionality and is outside the scope of this paper. For example, a PCI Express device that supports Gigabit Ethernet will have different interface design parameters than a PCI Express device that supports a graphics interface.

Package

The high speed nature of the PCI Express interconnect provides the same level of bandwidth as the legacy parallel PCI-X 2.0 bus with only a third of the pin count in a typical implementation. Figure 1 shows the pin counts in a sample implementation. The pin count comparison is drawn from three categories: signal pins, supply voltage pins and supply ground pins. Generally, the PCI Express device requires 55% fewer pins.

Devices based on PCI-X 266 require about 96 signal pins plus an additional 50% for power and other side-band pins to provide performance that is comparable to a x4 PCI Express device which requires only 19 pins—a substantial cost savings for the component and a significant reduction in overall system design complexity. The PCI Express device uses far fewer pins—1/5th in this example—to deliver the same performance. Also, typical side-band functions are signaled in-band, greatly reducing pin count and overall component cost. Further, a PCI Express device does not have any multi-voltage requirements, as does a PCI-X 266 device that must support both 3.3V and 1.5V rails, and hence needs fewer supply pins. Figure 2 shows the pin counts in a typical implementation with the voltage pin requirements for a x4 PCI Express device and a PCI-X 266 device. The ground pin requirements for both implementations are included. It is assumed that the PCI Express device is powered by a 1.5V supply rail. Also, a 2-to-1 signal-to-ground ratio is used.

Figure 1

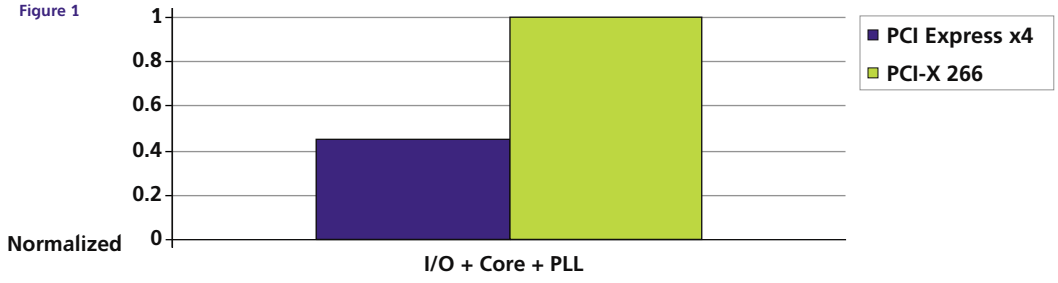
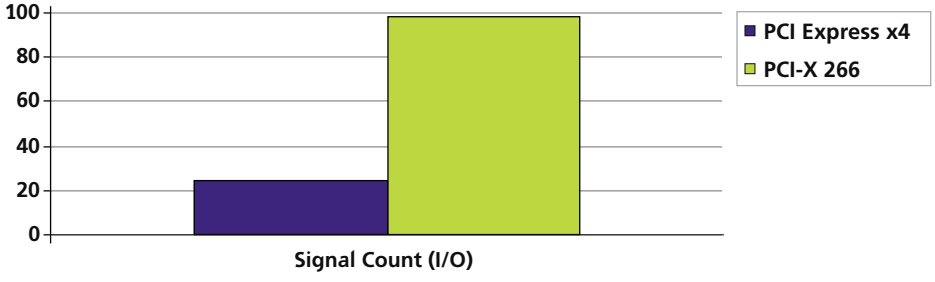


Figure 2



Die Area

There are three different contributors to the die area analysis considered in this paper:

- I/O area
- Core logic area (synthesized portion + RAMs)
- PLL area

Figure 3 demonstrates the impact of the above parameters given a 0.13_m process. The PLL area is approximately the same for both the PCI Express and PCI-X 266 devices and is not shown.

The I/O area is the overall layout area consumed by all the I/O cells put together for a particular interface. Each I/O cell contains the driver and receiver circuits required per signal on the bus interface. PCI Express designs consume substantially less total area for I/Os than do PCI-X 2.0 designs because they require only about a third of the total I/O cell count. After accounting for the area differences in the I/O cell between a PCI Express design and one based on the legacy parallel PCI-X 2.0 bus, the overall I/O area efficiency of the PCI Express technology is well over 100%. Larger I/O counts and the resultant larger aggregate I/O area typically lead to I/O limited designs, which put a limit on the lower bound of the die size of the device for a given package. For PCI-X 2.0 devices, this problem will only get worse as future lower voltage processes will require more silicon die size to support higher voltage signaling. The core logic area is a combination of the logic required to implement the interface protocol and the buffering required for supporting full bandwidth on the interface. The logic area required for a x4 PCI Express device includes logic to support both x4 and narrower link width operations as well as lane reversal and polarity inversion. It is assumed that 50% of the core area is consumed in device buffering with interface buffering being a function of the interface bandwidth and not just the frequency.

Power

In a typical implementation, there are three different contributors to the power dissipation of a given device:

- I/O power
- Core power (synthesized portion + RAMs)
- PLL power

Figure 4 shows the contrast between the PCI Express and PCI-X 266 devices—PLL power being roughly the same for both. The power value noted in the figure is the typical power. A logic activity factor of 50% is assumed for these typical power calculations. In general, the PCI Express device requires approximately 56% less power.

As noted in the previous section, the core logic and hence core power are most significantly functions of the interface bandwidth. A PCI Express implementation supporting a x4 link width consumes about the same amount of core power as the legacy parallel PCI-X 2.0 bus. As can be seen, the I/O power requirement for the x4 PCI Express device is roughly 1/4th of that of the PCI-X 266 device. This is because of the large number of I/Os required for implementing the legacy parallel PCI-X 2.0 bus.

Figure 3

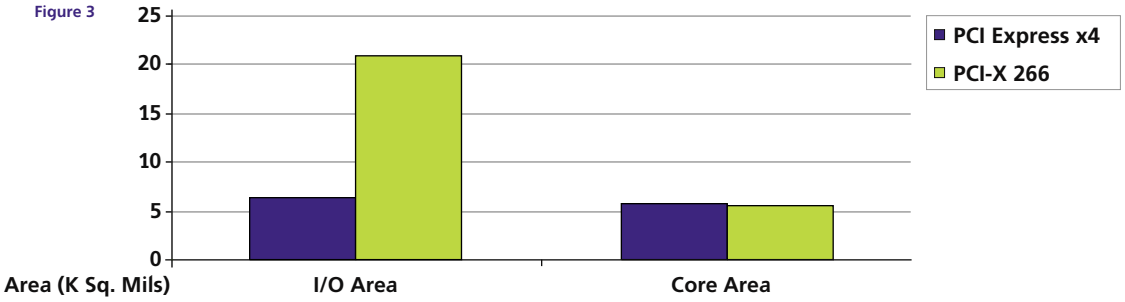
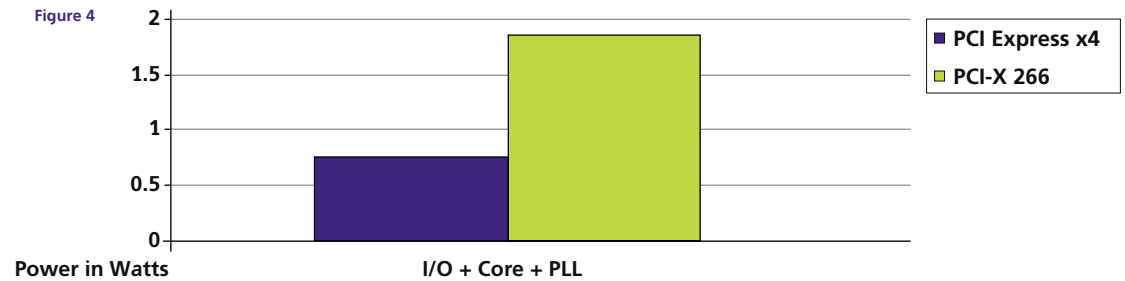


Figure 4



Systems Designs

We will now examine the benefits of the PCI Express Architecture in the design of systems. There are a number of attributes of importance to platform designers, including:

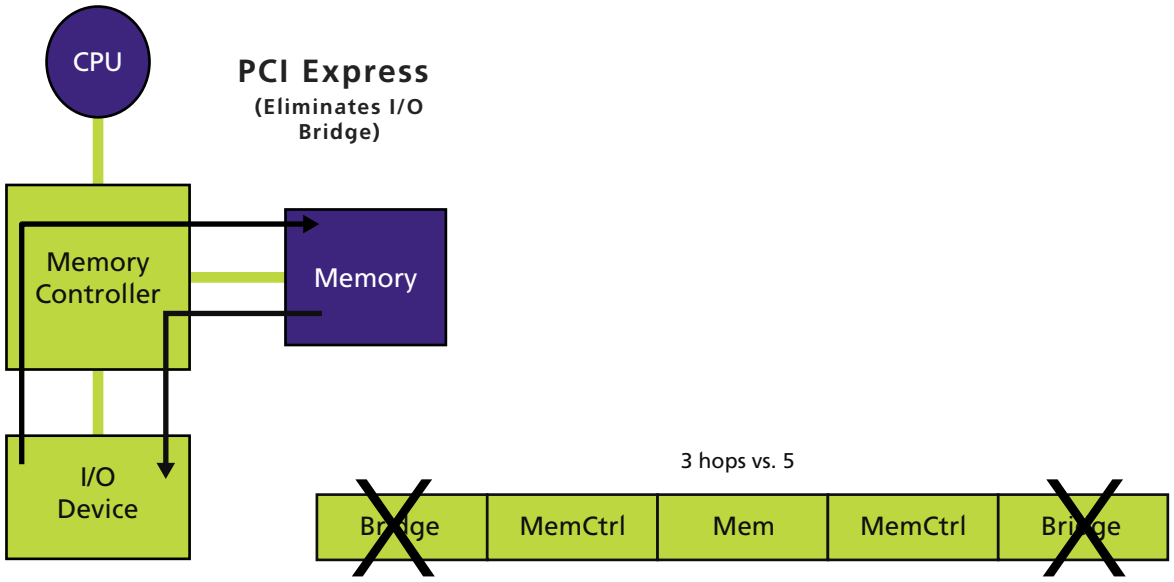
- Performance
- Connectivity and fan-out
- Scalability
- Reliability, Availability and Serviceability (RAS)

These design requirements are discussed below.

Performance

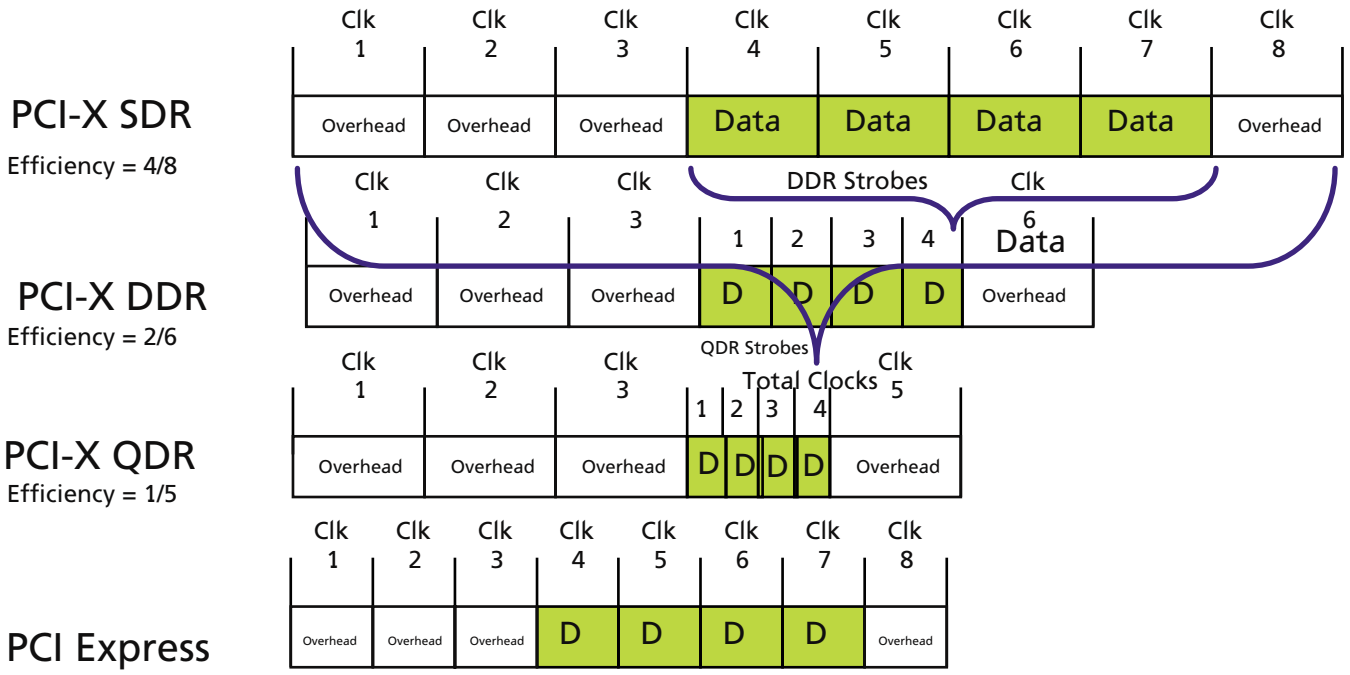
The performance of an I/O interface is typically measured with two attributes: latency and bandwidth. High latency has a negative impact on sustained throughput and both of these parameters will be considered. PCI Express provides substantial benefits by decreasing latency and increasing effective bandwidth.

The benefits of low latency are measured not only in performance but also in reduced platform and adapter costs. System latency decreases with PCI Express because bridging components are no longer required. Effectively, the number of "hops" is reduced in a platform using PCI Express Architecture. By reducing latency, not only does the bandwidth increase but the adapter costs decrease. The buffering requirements of the adapter are typically determined by the roundtrip latency that the device expects. If the latency decreases the pipelining requirements decrease, thereby reducing adapter costs.

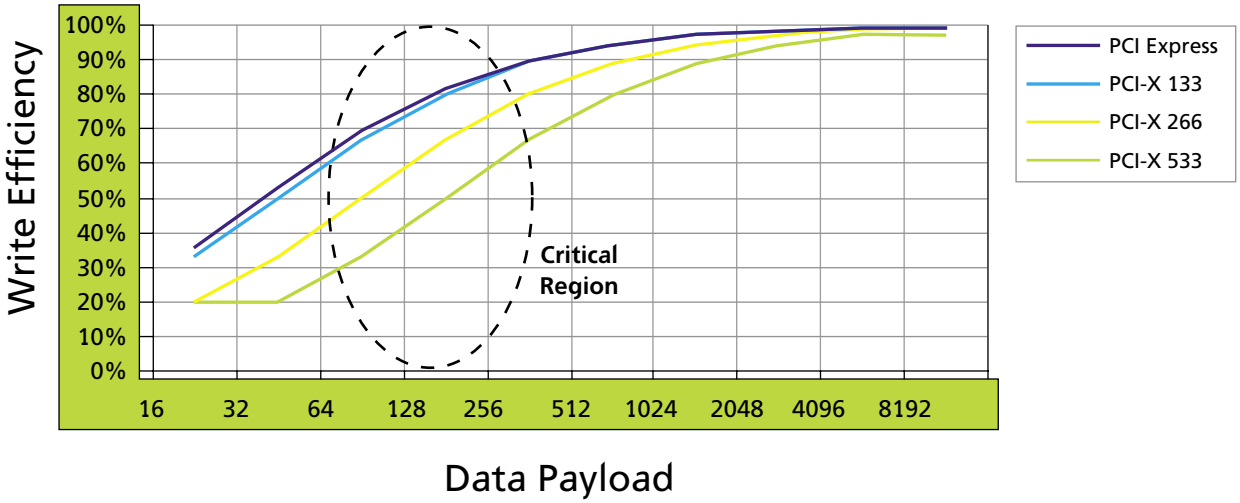


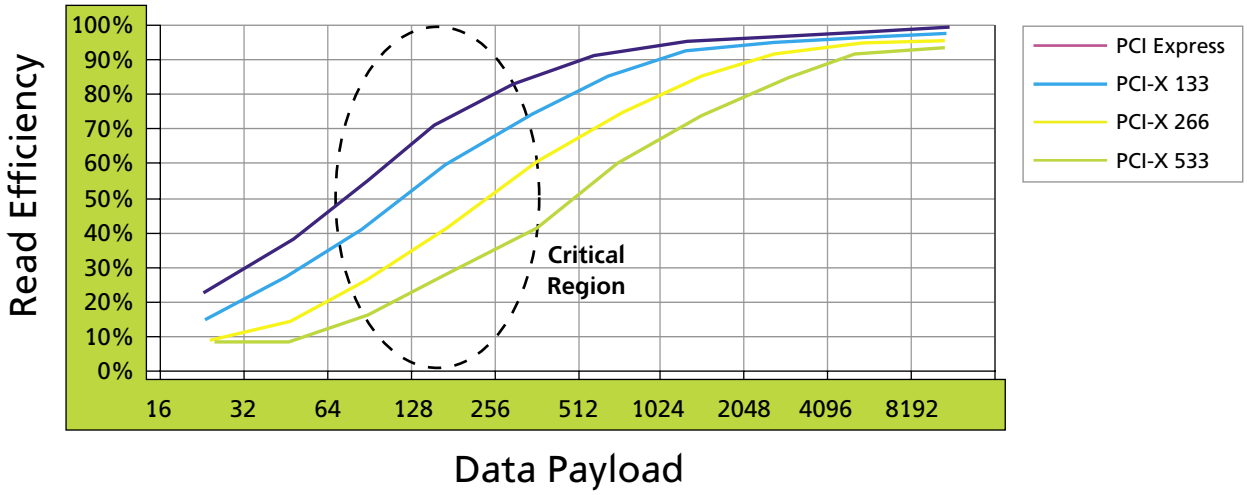
Peak bandwidth is governed by frequency and width. For example, the peak bandwidth of a x4 PCI Express interface is 2GB/s (setting aside the 8b/10b encoding overhead). However, the effective bandwidth takes into account frequency, width and protocol efficiency. All I/O protocols implementing multiplexed address/command/data use some of the interface bandwidth to transmit header information. Interface efficiency is equal to Payload / (Payload + Overhead). The effective bandwidth is equal to Efficiency * Peak Bandwidth.

Another important factor to consider is the payload size. A 4KB packet has high efficiency on any interface. However, the important sizes to consider are packets of 64B, 128B, and 256B. Descriptors used in DMA transfers, doorbell writes, status updates, and so on are all on the order of 64B or less. Moreover, cache line sizes are typically 64B in some systems. Large packets delivered on half duplex multi-drop busses, such as legacy parallel PCI-X 2.0, must arbitrate for the shared bus and consequently they have to be broken up, thereby impacting overall efficiency. In the figures on page 9, the read, write and bandwidth efficiencies of PCI Express Architecture are compared to legacy parallel busses, such as PCI-X, running at different speeds. As can be seen, the protocol efficiency for PCI-X **decreases** as the speed increases. More significantly, PCI Express Architecture provides the best read and write efficiencies for packet sizes in the above mentioned critical region.



PCI Express Architecture provides optimum efficiency when considering the "sweet spot" of 64B, 128B, and 256B packets. A protocol that provides good efficiency also provides good effective bandwidth.





Connectivity and fan-out

Low pin count interfaces enable high connectivity options. For instance, a x1 PCI Express link only requires 4 signals for 500 MB/s total peak bandwidth. There are no sideband signals. Since all of the data and control information is carried on the same 4 pins, components can support many PCI Express interfaces. In cases where platform architects must design multiple slots for I/O expansion, the pin savings of PCI Express technology bring significant benefits to the whole system by reducing cost and complexity. For example, an enterprise platform with six PCI Express x4 slots will require less than 200 pins whereas the same system with six PCI-X 266 slots will require over 1,000 pins, giving the PCI Express design a significant advantage in cost savings, reduction in design complexity and board routing.

Lower pin counts facilitate inexpensive board solutions with fewer signal layers. Also, routing distances are more flexible in designs using PCI Express technology. Blade servers and other configurations can take advantage of the reduced board routing requirements and its related benefits. In addition, polarity and lane reversal—features that simply don't exist in PCI-X 2.0—promote simple, inexpensive topologies. These and other attributes enable newer connectivity options such as cabling and modular form factors for PCI Express.

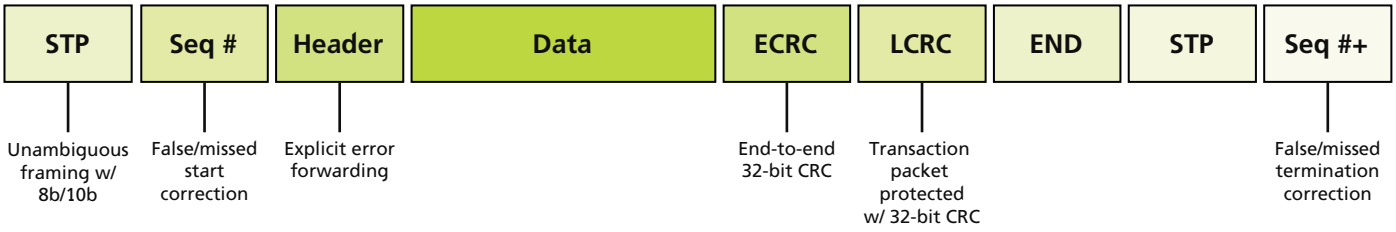
Scalability

The PCI Express Architecture currently supports a wide array of interconnect widths including 1, 2, 4, 8, 12, 16, and 32 lanes. Future improvements will yield even higher scalability in both frequency and width. Aside from the ability to linearly scale required bandwidth by increasing the interconnect width, the PCI Express technology also supports port bifurcation. Since all the information is carried in-band, a PCI Express port is capable of bifurcating into narrower links.

RAS

The PCI Express Architecture provides an advanced error reporting and correction capability. High reliability is achieved through its mechanisms to detect and correct packet errors. Packets are framed by start (STP) and end (END) symbols. The framing is unambiguous and by default 8b/10b encoding offers protection of symbols. The PCI Express Physical layer is responsible for generating and detecting packet framing. Each packet is identified with a sequence number which enables the receiver to detect a missed or dropped packet. The receiving PCI Express Link layer tracks the expected sequence number. The receiver also calculates a 32-bit CRC on the Transaction layer information. If the sequence number is not what is expected or the calculated CRC mismatches the transmitted CRC, the packet is retried. This correctable condition gives the interface an opportunity to "fix" the problem through packet reissue. At the PCI Express Transaction layer, the packet header contains explicit mechanisms to forward poisoned data errors. Data poisoning enables the possibility of tracking down the path of a packet in error. Finally, the optional end-to-end CRC (ECRC) is a code applied to the invariant elements of the PCI Express packet. The ECRC information enables protection on those elements from source to destination without concern for internal component errors going undetected.

These error protection and correction capabilities distinguish the PCI Express Architecture as valuable in the design of reliable and highly availability platforms and solutions. Further, the PCI Express hot plug mechanisms provide seamless operation. External switches and FETs are not required and all hot plug messaging is communicated in-band. This design philosophy facilitates newer and evolving form factors of the future in a consistent manner.



Summary

The PCI Express Architecture provides design advantages for both components and systems. For components, benefits in package, die area and power provide a compelling value proposition in reducing cost and design complexity. The PCI Express interconnect is at least 3 times more efficient than the legacy parallel PCI-X 2.0 bus with respect to pin count. This advantage in a package can mean the difference between a 4 Layer 421-pin Plastic Ball Grid Array (PBGA) and a 8 Layer 512-pin Flip-Chip Ball Grid Array (FCBGA), which has a typical cost differential of approximately $4. The PCI Express interconnect is at least 1.5 times more efficient in terms of die-size than PCI-X 2.0. Smaller die sizes lead to lower cost designs. And the PCI Express interconnect is at least 1.8 times more efficient than PCI-X 2.0 relative to device power dissipation. This provides for more degrees of freedom in device thermal solutions.

In addition to the component advantages mentioned above, the PCI Express Architecture provides many design

benefits for platform architects, ranging from reduced latency and increased effective peak bandwidth for higher performance, to superior connectivity and fan-out with low pin count, facilitating relaxed board routing, excellent scalability in both interconnect width and future frequency increases, as well as advanced features to enable the implementation of reliable and highly available systems and solutions. At the same time, this technology is fully compatible with the ubiquitous PCI software programming model and device driver interfaces for investment protection and a smooth industry transition. The PCI Express technology is proven and silicon-ready.

For more information on the PCI Express Architecture, please visit www.pcisig.com. Further information is available at http://developer.intel.com/technology/pciexpress/devnet/index.htm.

Author Biographies

Ken Creta is a senior staff server chipset architect with Intel Corporation. Ken has architected Intel's highest end server chipsets with experience in defining a wide variety of both industry and proprietary coherent and non-coherent interconnects.

Sridhar Muthrasanallur is an I/O Bridge Architect with Intel Corporation. Sridhar has experience in architecting and designing Intel's Server Chipsets.