Intel® PCI and PCI Express*

PCI Express* keeps in step with an evolving industry

The technology vision for PCI and PCI Express*

From the first Peripheral Component Interconnect (PCI) specification through the upcoming PCI Express 3.0, Intel has spearheaded innovations that make the PC platform more functional, performance-balanced and responsive for a variety of applications. Because of Intel's investment in PCI and PCIe, hard drive access is faster, video displays are faster, network access is faster, and I/O capacity has stayed in step with ever-increasing processor speeds. And with the latest releases of PCI Express, Intel has also contributed to innovations that help I/O capabilities support new data center and computing usage models.

> "Silicon products, notably high performance servers, workstations, PCs and visual computing platforms, continue to push the limits of I/O bandwidth as they scale with Moore's Law. PCIe 3.0 will deliver I/O headroom for the next generation of products and compute-intensive applications in these segments. to provide end users with higher responsiveness, faster data transfers and more realistic graphics."
>
> — Stephen Whalley, Manager, I/O Technology Initiatives, Intel Corporation

A Short History of PCI

The Peripheral Component Interconnect (PCI) specification was used by the computing industry from 1992 until 2004 as the primary local bus system within a computer. The PCI specification standardized how PCI expansion cards, such as a network card or modem, install themselves and exchange information with the CPU. But over the years, CPU frequencies rose from 66 MHz in 1993 to over 3 GHz in 2003 (two orders of magnitude larger). The existing PCI bandwidth could hardly begin to feed the I/O processing capability of the new CPUs, so Intel also spearheaded the industry's next effort and helped create the PCI Express specification so that users could get full value from their new, faster processors.

Appearing in systems starting in 2004, PCI Express was technically not a new generation of PCI architecture, but an architectural leap. It kept the core of PCI's software infrastructure, but completely replaced the hardware infrastructure with a radically new forward-looking architecture that put I/O back into the express lane of performance. PCI Express is not only designed to replace the PCI bus for devices such as modem and network cards, but also the Advanced Graphics Port (AGP) used for desktop graphics cards since 1997. Unlike PCI, its parallel predecessor, PCI Express is a serial point-to-point interconnect that is capable of extremely high-bandwidth transfers, with performance ranges in the first generation up to 4 gigabytes (GBps) per direction for a 16 "lane" implementation.

<PCI timeline graphic here / Slide 4 of Mahesh's Taipei presentation>

While other technologies have come and gone, PCI has remained an industry standard for 17 years. Over time, the PCI Express standard has wound its way through the computing eco-system, but another of its strengths is that it is not one-

size-fits-all. Manufacturers can implement the features that fit with their application, whether that's a server, high-end workstation, laptop, graphics card, NIC, USB host controller or embedded device.

PCI's longevity can be attributed to industry support by the PCI-SIG and to several characteristics of the standard itself:

- PCI is processor-agnostic in both frequency and voltage, so it can function in the desktop, mobile and server markets with little or no change. It isn't wired into a specific processor, so computer manufacturers can standardize their I/O across multiple product groups and generations, and they don't have to redesign their PC architecture every time Intel comes out with a new chip.
- PCI is flexible in its ability to support multiple form factors. PCI-SIG members have been able to define motherboard connectors, add-in cards and brackets to standardize the I/O back panel and form factors for the server, desktop and mobile markets. This standardization made the distribution of PCI-based add-in cards and form factor-based computer chassis possible through the consumer channel and in sufficient volumes to meet consumer price targets.
- The PCI Express specification has been able to double bandwidth approximately every four years, to keep pace with increasing processor performance, and it has been able to provide feature improvements to support new computing models.
- PCI Express specifications have been designed for backward compatibility thereby extending the life of useful peripheral hardware and recovery of industry development costs.

These capabilities have made PCI one of the most (if not the most) successful chip and board interconnect technologies in history. Now Intel is helping lead innovations in new generations of PCI Express that will continue its long and fruitful life.

PCI Express, The Next Generation

As fast as the new PCI Express standard was, processor architectures have continued to accelerate to meet the demands of emerging applications, and PCI Express needed to keep up. For example, added bandwidth was needed to improve the performance of data-intensive graphic workloads and advanced gaming applications. New emerging I/O developments such as increased speeds and feeds of network and storage applications could also directly benefit from increased bandwidth.

To meet new bandwidth requirements, PCI-SIG announced the availability of the PCI Express Base 2.0 specification in January of 2007. PCI Express 2.0 doubles the transmission speed of PCI Express 1.1 to 5.0 GT/s, thus doubling throughput of PCI Express x16 (the typical graphics card interface) to 16 GB/s.

Analogous to the "intelligent performance" and energy efficiency features of new Intel architectures, PCI Express 2.0 offers more flexibility to hardware manufacturers, because a PCI Express 2.0 interface with 4 lanes delivers the same bandwidth as a PCI Express 1.1 interface with 8 lanes. So manufacturers can either design to double the throughput, or they can optimize power requirements by switching from 1.1 to 2.0 mode and use (and power) half the number of lanes. PCI Express 2.0 still supports PCI Express 1.1 speeds, so manufacturers can also design to save energy by reducing speeds during periods of lower throughput requirements. PCI Express 2.0 is capable of automatically negotiating link width (from a few to 16

links) and link speed (2.5 or 5 GT/s). It also supports up to 300W of power, for high-end graphics applications.

PCI Express 2.1 and 3.0

After PCI Express 2.0, the PCI-SIG began to consider how PCI Express could address new usage models such as multi-tasking, graphics I/O bandwidth, specialized applications using co-processors, new data center imperatives such as improved power management and server and client virtualization. A number of protocol changes were proposed to provide greater throughput, to fit specialized hybrid-computing models, to transfer data more efficiently for HPC-type applications, and to support device-based power management.

Some of these protocol changes have already been released as the PCI Express 2.1 specification.

Extensions	Explanation	Benefits	Application Class
Transaction Layer Packet (TLP) Processing Hints	Request hints to enable optimized processing within host memory/cache hierarchy	Reduce access latency to system memory. Reduce System Interconnect and memory bandwidth and associated power consumption	NIC, storage, accelerators/GP-GPU
Latency Tolerance Reporting	Mechanisms for platform to tune PM	Reducing platform power based on device service requirements	All devices and segments
Opportunistic Buffer Flush and Fill	Platform mechanisms to tune power management and to align device activities	Reducing platform power based on aligning device activity with platform power management events to further reduce platform power	All devices and segments
Atomics	Atomic Read-Modify-Write mechanism	Reduced synchronization overhead Software library algorithm and data structure re-use across core and accelerators/devices	Graphics, accelerators/GP-GPU
Resizable BAR	Mechanism to negotiate BAR size	System resource optimizations - breakaway from “all-or-nothing” device address space allocation	Any Device with large local memory (e.g., Graphics)
Multicast	Address Based Multicast	Significant gain in efficiency compared to multiple unicasts	Embedded, storage, multiple graphics adapters

I/O Page Faults	Extends IO address remapping for page faults – (Address Translation Services 1.1)	System Memory Management optimizations	Accelerators, GP-GPU usage models
Ordering Enhancements	New ordering semantic to improve performance	Improved performance (latency reduction) ~ (IDO) 20% read latency improvement by permitting unrelated reads to pass writes.	All devices with two party communication
Dynamic Power Allocation (DPA)	Mechanisms for dynamic power/performance management of D0 (active) sub-states.	Dynamic component power/thermal control, manage endpoint function power usage to meet new customer or regulatory operation requirements	GP-GPU
Internal Error Reporting	Extend AER to report component internal errors (Correctable/ uncorrectable) and multiple error logs	Enables software to implement common and interoperable error handling services Improved error containment and recovery	RAS for switches
TLP Prefix	Mechanism to extend TLP headers	Scalable architecture headroom for TLP headers to grow with minimal impact to routing elements. Support vendor-specific header formats	MR-IOV, large topologies, provisioning for future use models
I/O Page Faults	Extends I/O address remapping for page faults (Address Translation Services 1.1)	System memory management optimizations	Accelerators, GP-GPU usage models

PCI Express 3.0 will carry a data rate of 8 GT/s. Following a six-month technical analysis of the feasibility of scaling the PCIe interconnect bandwidth, the PCI technologists concluded that 8 GT/s can be manufactured in mainstream silicon process technology. They feel it can be deployed with existing low-cost system materials and infrastructure, and with negligible impact, while maintaining full compatibility to the PCIe protocol stack.

PCIe 2.0 delivers 5 GT/s but it employed an 8b/10b encoding scheme which took 20 percent overhead from the overall raw bit rate. PCIe 3.0 removes the requirement for the 8b/10b encoding scheme and instead uses a scrambling only scheme. By removing the overhead, it delivers the more cost-effective 8 GT/s data rate and still manages to effectively double the PCIe 2.0 bandwidth.

In addition to speed increases, the PCI Express 3.0 specification will include optimizations for enhanced signaling and data integrity, such as transmitter and receiver equalization, PLL improvements for clock data recovery, and channel enhancements for currently supported topologies. PCI-SIG expects the PCIe 3.0 specifications to undergo rigorous technical review and validation before being released to the industry in 2011. The additional bandwidth provided by PCI Express 3.0 will make it possible for users to enjoy benefits such as higher responsiveness, more realistic graphics, faster file transfer, and a richer networking experience.

What lies ahead?

After all the improvements coming in PCI Express 3.0, what will the next generation of PCI Express look like? Based on the industry's processor technology and application roadmaps, doubling bandwidth continues to be important. [Mahesh] Intel and other companies within the PCIe ecosystem have expressed interest in increasing interconnect bandwidth and are exploring ways to do this in the most cost and power-efficient manner.

While research continues, Intel will continue to work with other industry leaders to ensure that future generations of PCI Express stay true to the standard's core vision: delivering cost-effective interconnect technology that keeps pace with evolving customer requirements, emerging applications, and advances in processor technology, while allowing manufacturers the flexibility to implement the changes that their customers demand.