PCI EXPRESS TECHNOLOGY

Jim Brewer, Dell Business and Technology Development
Joe Sekel, Dell Server Architecture and Technology

Formerly known as 3GIO, PCI Express is the open standards- based successor to PCI and its variants for server- and client-system I/O interconnects. Unlike PCI and PCI-X, which are based on 32- and 64-bit parallel buses, PCI Express uses high-speed serial link technology similar to that found in Gigabit[1] Ethernet, Serial ATA (SATA), and Serial-Attached SCSI (SAS). PCI Express reflects an industry trend to replace legacy shared parallel buses with high-speed point-to-point serial buses.

The new bus technology is expected to allow the PCI Express transmission rates to keep pace with processor and I/O advances for the next 10 years or more. Systems with PCI Express will begin appearing around the middle of 2004.

PCI Express has the following advantages over PCI:

- Serial technology providing scalable performance.
- High bandwidth—Initially, 5–80 gigabits per second (Gbps) peak theoretical bandwidth, depending on the implementation.
- Point-to-point link dedicated to each device, instead of the PCI shared bus.
- Opportunities for lower latency (or delay) in server architectures, because PCI Express provides a more direct connection to the chip set Northbridge[2] than PCI-X.
- Small connectors and, in many cases, easier implementation for system designers.
- Advanced features—Quality of service (QoS) via isochronous channels for guaranteed bandwidth delivery when required, advanced power management, and native hot plug/hot swap support.

PCI Express will replace the PCI, PCI-X, and AGP parallel buses gradually over the next decade. It will initially replace buses that need the additional performance or features. For instance, PCI Express will initially be deployed as a replacement for the AGP8X graphics bus in client systems, providing high bandwidth and support for multimedia traffic. It will also coexist with and ultimately replace the PCI-X bus in server systems.

In this white paper, we begin with a review of the PCI bus and its variants (PCI-X and AGP) in client and server systems. The paper continues with a discussion of PCI Express technology, including its strengths, advanced features, and form factors. We conclude with its impact on computer system architectures.

Contents

PCI Bus

Since its inception in 1992, the PCI bus has become the I/O backbone of nearly every computing platform. The original 33-MHz, 32-bit implementation delivers a peak theoretical bandwidth of 133 megabytes per second (MB/sec). Over time, the industry has evolved the plat-

1. This term does not connote an actual operating speed of 1 Gbps. For high-speed transmission, connection to a Gigabit Ethernet server and network infrastructure is required.
2. The term, Northbridge, refers to the controller for the processor bus, memory bus, AGP bus, and the link to the Southbridge. The term, Southbridge, refers to the I/O device controller.

form architecture by offloading various functions to higher-bandwidth PCI derivatives, including AGP and PCI-X, both of which are PCI variants. Table 1 presents the peak bandwidth of the PCI, PCI-X, and AGP buses.

Bus and Frequency	Peak 32-Bit Transfer Rate	Peak 64-Bit Transfer Rate
33-MHz PCI	133 MB/sec	266 MB/sec
66-MHz PCI	266 MB/sec	532 MB/sec
100-MHz PCI-X	Not applicable	800 MB/sec
133-MHz PCI-X	Not applicable	1 GB/sec
AGP8X	2.1 GB/sec	Not applicable

Table 1. Bandwidth of PCI, PCI-X, and AGP Buses

A close examination of PCI signaling technology reveals a multidrop,[3] parallel bus that is reaching its performance limits. The PCI bus cannot be easily scaled up in frequency or down in voltage. In addition, the PCI bus does not support features such as advanced power management, native hot plugging/hot swapping of peripherals, or QoS to guarantee bandwidth for real-time operations. Finally, all of the available bandwidth of the PCI bus is limited to one direction (send or receive) at a time. Many communications networks support simultaneous bidirectional traffic, which minimizes message latency.

Client Systems

The original PCI bus was designed to support 2D graphics, higher-performance disk drives, and local area networking. Not long after PCI was introduced, the increasing bandwidth requirements of 3D graphics subsystems outstripped the 32-bit, 33-MHz PCI bus bandwidth. As a result, Intel and several graphics suppliers created the AGP specification, which defined a dedicated high-speed PCI bus for graphics operations. The AGP bus offloaded graphics traffic from the PCI system bus and freed up bandwidth for other communications and I/O operations. In addition, Intel recently added dedicated USB 2.0 and Serial ATA links to the Southbridge in its chip sets, further reducing the I/O demands on the PCI bus. Figure 1 shows the internal architecture of a typical client PC system and the bandwidth of its I/O and graphics buses.

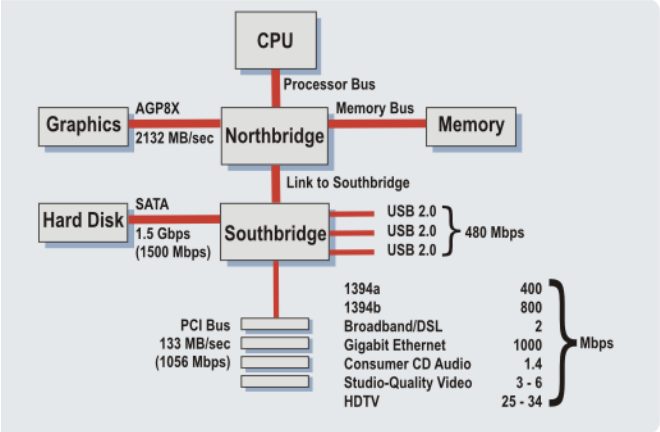


Figure 1. Typical Client System Architecture

Client-System Bottlenecks

Several client-system buses can limit system performance because of CPU, memory, and I/O device advances: the PCI bus, AGP bus, and the link between the Northbridge and Southbridge.

PCI Bus. The PCI bus provides up to 133 MB/sec to connected I/O devices. A number of I/O devices can saturate or consume a high percentage of this bandwidth. When more than one of these devices are active, the shared PCI bus is quickly stressed beyond its limits.

Figure 2 shows many of the contributors to the PCI bus bottleneck. This figure shows the bandwidth required by various communications, video, and external devices that are serviced by the PCI bus. It can be seen that the multidrop, shared PCI bus is hard pressed to keep up with today's devices. This situation worsens with upcoming peripheral devices with even higher data rates. For example, Gigabit Ethernet requires a bandwidth of 125 MB/sec, which effectively saturates the 133-MB/sec PCI bus. The IEEE 1394b bus has a maximum bandwidth of 100 MB/sec, which can saturate a standard PCI bus.

AGP. Over the past decade, video performance requirements have approximately doubled every two years. During this time, the graphics bus has transitioned from PCI to AGP, and from AGP to AGP2X, AGP4X, and finally

3. On a multidrop bus, all devices attached to it are connected to the same set of wires. When a device is using the PCI bus, no other device can communicate over the bus. All connected devices must share the bus and wait their turn before sending or receiving data.

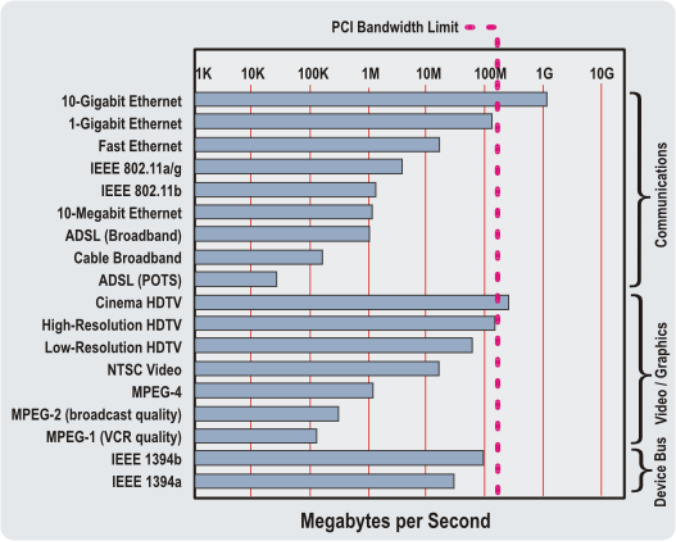


Figure 2. Bandwidth of Devices Serviced by PCI Bus

today's AGP8X. AGP8X operates at 2.134 gigabytes per second (GB/sec). Despite this bandwidth, the progressive performance demands on the AGP bus are putting considerable pressure on board design and interconnection costs. Like the PCI bus, extending the AGP bus becomes more difficult and expensive as frequencies increase.

Link Between Northbridge and Southbridge. Congestion on the PCI bus also affects the link between the Northbridge and the Southbridge. SATA drives and USB devices further stress this link. A higher-bandwidth link will be required in the future.

Server Systems

In servers, the original 32-bit, 33-MHz PCI bus was extended to a 64-bit, 66-MHz bus with a bandwidth of 532 MB/sec. The 64-bit bus was recently extended to 100 and 133 MHz, referred to as PCI-X. The PCI-X bus connects the server-system (and the high-end, dual-processor workstation-system) chip set to expansion slots, Gigabit Ethernet controllers, and Ultra320 SCSI controllers embedded on the system board. A 64-bit PCI-X bus at 133 MHz delivers 1 GB/sec of peak bandwidth between the system chip set and the I/O device. This is sufficient bandwidth for the majority of immediate server I/O requirements, including Gigabit Ethernet, Ultra320 SCSI, and 2-GB/sec Fibre Channel. However, like PCI, PCI-X is a shared bus and is likely to require a higher-bandwidth alternative in 2004.

The PCI Special Interest Group (PCI SIG) has been developing the PCI-X 2.0 specification, which will effectively create a 64-bit, 266-MHz PCI-X bus with double the data rate of the 133-MHz PCI-X bus. However, there are significant design issues associated with extending these parallel PCI-X bus variants. The connectors are large and expensive, and stringent design requirements drive up the cost of system boards significantly as frequencies are increased. In addition, to avoid excessive electrical loading at the higher speeds of PCI-X 2.0, only one I/O device can be attached in a point-to-point configuration to the PCI-X bus. It cannot be implemented as a shared bus.

Server-System Bottlenecks

Figure 3 shows the internal system interconnects in a typical dual-processor server system. In this architecture, high-bandwidth expansion is provided via a proprietary interface between the Northbridge and PCI-X bridge chips. Multiple PCI-X buses connect to high-speed expansion slots, 10-Gigabit Ethernet, and SAS/SATA drives. This architecture has some drawbacks. The proprietary PCI-X bridge chips connect multiple parallel PCI-X buses to the chip set's proprietary serial interconnect. This approach is expensive, inefficient, and introduces latency between I/O devices and the Northbridge. For example, the approach connects a serial 10-Gbps fabric to a point-to-point, 64-bit parallel bus that is, in turn, connected via a proprietary PCI-X bridge chip to a proprietary serial interconnect into the Northbridge.

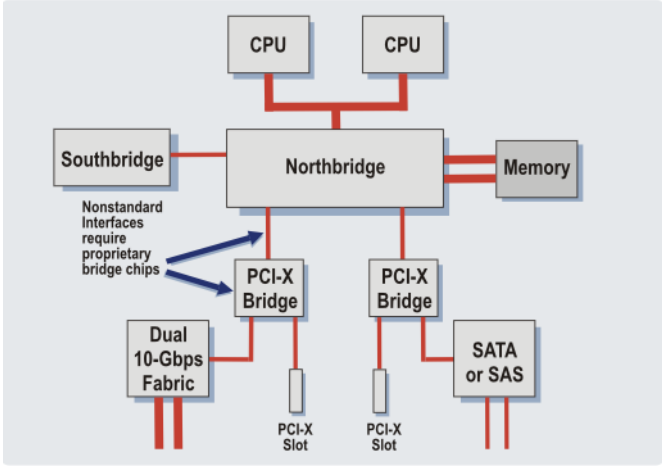


Figure 3. Current Dual-Processor Server Architecture

In addition, next-generation external server I/O technologies are expected to require much greater bandwidth than a 133-MHz PCI-X bus can provide. These technologies include system-area fabrics such as 10-Gigabit Ethernet, 10-Gbps Fibre Channel, and 4x Infiniband. They also include future higher-speed hard-drive interfaces such as 3-Gbps SATA and SAS. In the case of a 10-Gbps fabric, each 10-Gbps port will be able to transmit bidirectional data at a peak bandwidth of 2 GB/sec. The 133-MHz PCI-X bus delivers a maximum of 1 GB/sec in one direction at a time. This suggests that the 133-MHz PCI-X bus could throttle the peak bandwidth of these fabrics by as much as 50 percent. Although PCI-X 2.0 at 266 MHz would double the PCI-X peak bandwidth to 2 GB/sec, it would still fall short of the total 4 GB/sec required by a dual-ported 10-Gbps fabric controller.

Dell believes that client and server systems require a replacement bus for the parallel PCI bus and its variants.

PCI Express Technology

PCI Express provides a scalable, high-speed, serial I/O bus that maintains backward compatibility with PCI applications and drivers. The PCI Express layered architecture supports existing PCI applications and drivers by maintaining compatibility with the existing PCI load-store (and flat address space) model. The layered architecture is discussed in the sidebar below, "PCI Express Layered Architecture."

The PCI Express architecture defines a high-performance, point-to-point, scalable, serial bus. A PCI Express link consists of dual simplex channels, each implemented as a transmit pair and a receive pair for simultaneous transmission in each direction. Each pair consists of two low-voltage, differentially driven pairs of signals. A data clock is embedded in each pair, using an 8b/10b clock-encoding scheme to achieve very high data rates. Figure 4 compares the PCI and PCI Express links.

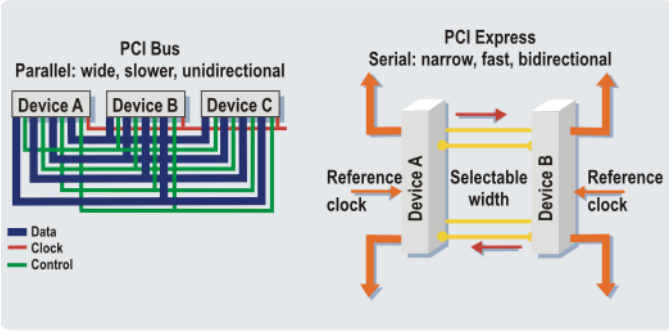
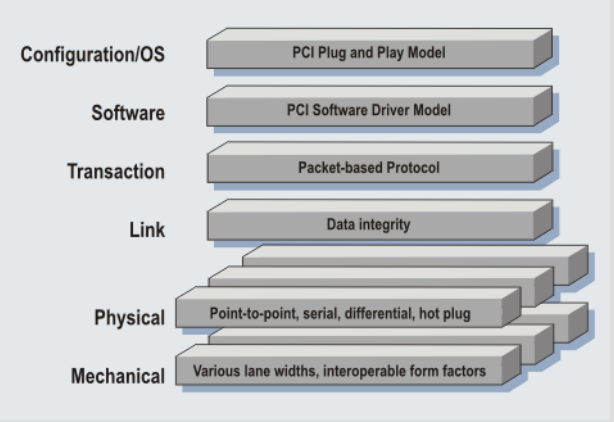


Figure 4. PCI Versus PCI Express

PCI Express Layered Architecture



Configuration/Operating System Layer—Leverages the standard mechanisms defined in the PCI Plug-and-Play specification for device initialization, enumeration, and configuration. This layer communicates with the software layer by initiating a data transfer between peripherals or receiving data from an attached peripheral. PCI Express is designed to be compatible with existing operating systems, but future operating system support is required for many of the technology's advanced features.

Software Layer—Generates read and write requests to peripheral devices. PCI Express maintains initialization and runtime software compatibility with PCI. Like PCI, the PCI Express initialization model allows the operating system to discover add-in hardware devices and allocate system resources. PCI Express retains the PCI configuration space and the programmability of I/O devices. In fact, all operating systems will boot without modification on a PCI Express system. The PCI runtime software model is also preserved, enabling existing software to execute unchanged.

Transaction Layer—Transports read and write requests from the software layer to the link layer using a packet-based protocol, and matches response packets to the original software requests. The transaction layer supports 32-bit and extended 64-bit memory addressing. It also supports PCI memory, I/O, and configuration address spaces, as well as a new message space for in-band messages such as interrupts and resets. This message space eliminates the need for numerous PCI and PCI-X sideband signals.

Link Layer—Adds sequencing and error detection cyclic redundancy codes (CRCs) to the data packets to create a reliable data transfer mechanism between the system chip set and the I/O controller.

Physical Layer—Implements the dual simplex PCI Express channels. Implementations are flexible and various technologies and frequencies may be used. In this way, initial silicon technology can be replaced easily with future implementations that are backward compatible. For example, fiber-optic technology might be used to increase the data transfer rate.

Mechanical Layer—Defines various form factors for peripheral devices.

The bandwidth of a PCI Express link can be scaled by adding signal pairs to form multiple lanes between the two devices. The specification supports x1, x4, x8, and x16 lane widths and stripes the byte data across the links accordingly. Once the two agents at each end of the PCI Express link negotiate lane widths and frequency of operation, the striped data bytes are transmitted with 8b/10b encoding.

The basic "x1" link has a peak raw bandwidth of 2.5 Gbps. Because the bus is bidirectional (that is, data can be transferred in both directions simultaneously), the effective raw data transfer rate is 5 Gbps. Table 2 summarizes the encoded and unencoded data rates (see sidebar) of x1, x4, x8, and x16 implementations, which are defined in the initial generation of PCI Express.

PCI Express "Coded" and "Unencoded" Bandwidth

PCI Express bandwidth is commonly expressed as "encoded" bandwidth. PCI Express uses 8b/10b encoding, which encodes 8-bit data bytes into 10-bit transmission characters. This approach improves the physical signal so that bit synchronization is easier, design of receivers and transmitters is simplified, error detection is improved, and control characters can be distinguished from data characters.

The "encoded" bandwidth of a basic x1 PCI Express lane is 5 Gbps. However, a more accurate bandwidth figure is the "unencoded" bandwidth, which is 80 percent of 5 Gbps or 4 Gbps. Table 2 presents both encoded and unencoded PCI Express bandwidth. In this paper, we follow the common industry practice of citing the higher encoded bandwidth figures.

PCI Express Implementation	Encoded Data Rate	Unencoded Data Rate
x1	5 Gbps	4 Gbps (500 MB/sec)
x4	20 Gbps	16 Gbps (2 GB/sec)
x8	40 Gbps	32 Gbps (4 GB/sec)
x16	80 Gbps	64 Gbps (8 GB/sec)

Table 2. PCI Express Bandwidth

Future implementations of PCI Express will raise the channel communication frequency to even higher levels. For example, a second generation of PCI Express could increase the communication frequency by a factor of 2 or more.

Because it is a point-to-point architecture, the entire bandwidth of each PCI Express bus is dedicated to the device at the end of the link. Multiple PCI Express devices can be active without interfering with each other.

In contrast to PCI, PCI Express has minimal sideband signals and the clocks and addressing information are embedded in the data. Because PCI Express is a serial technology with few sideband signals, it provides a very high bandwidth per I/O connector pin compared to PCI. This is designed to provide more efficient, smaller, and cheaper connectors. Figure 5 compares the bandwidth per I/O connector pin of PCI, PCI-X, AGP, and PCI Express.

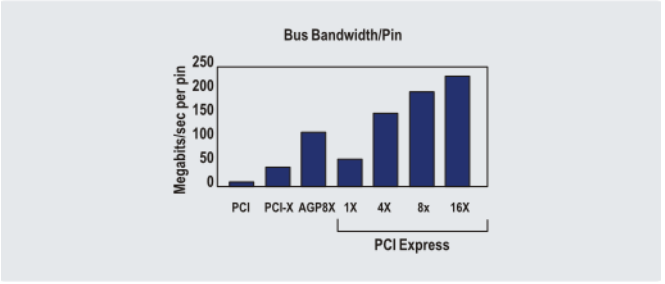


Figure 5. Comparison of I/O Bus Bandwidth Per Pin

PCI Express technology achieves high data rates reliably by using low-voltage differential signaling. In this approach, the signal is sent from the source to the receiver over two lines. One contains a "positive" image and the other, a "negative" or "inverted" image of the signal. The lines are routed using strict routing rules so that any noise that affects one line also affects the other line. The receiver collects both signals, inverts the negative version back to the positive and sums the two collected signals, which effectively removes the noise.

The original PCI Express specification defines graphics cards with up to 75 watts of power. In addition, a new high-end PCI Express graphics specification is under development that defines cards of up to 150 watts. These higher power levels accommodate the requirements of graphics adapters, which currently peak at 41 watts for mainstream AGP cards and 110 watts for AGP Pro 110 cards.

PCI Express Advanced Features

PCI Express has advanced features that will be phased in as operating system and device support is developed and as customer applications require them:

- Advanced power management
- Support for real-time data traffic
- Hot plug and hot swap
- Data integrity and error handling

Advanced Power Management

PCI Express has "active-state" power management, which lowers power consumption when the bus is not active (that is, no data is being sent between components or peripherals). On a parallel interface such as PCI, no transitions occur on the interface until data needs to be sent. In contrast, high-speed serial interfaces such as PCI Express require that the interface be active at all times so that the transmitter and receiver can maintain synchronization. This is accomplished by continuously sending idle characters when there is no data to send. The receiver decodes and discards the idle characters. This process consumes additional power, which impacts battery life on portable and handheld computers.

To address this issue, the PCI Express specification creates two low-power link states and the active-state power management (ASPM) protocol. When the PCI Express link goes idle, the link can transition to one of the two low-power states. These states save power when the link is idle, but require a recovery time to resynchronize the transmitter and receiver when data needs to be transmitted. The longer the recovery time (or latency), the lower the power usage. The most frequent implementation will be the low-power state with the shortest recovery time.

Support for Real-Time Data Traffic

Unlike PCI, PCI Express includes native support for isochronous (or time-dependent) data transfers and various QoS levels. These features are implemented via "virtual channels" that are designed to guarantee that particular data packets arrive at their destination in a given period of time. PCI Express supports multiple isochronous virtual channels—each an independent communications session—per lane. Each channel may have a different QoS level. This end-to-end solution is designed for applications that require real-time delivery such as real-time voice and video.

Hot Plug and Hot Swap

PCI-based systems do not have native (or built-in) support for hot plugging or hot swapping I/O cards. Instead, a few limited server and PC Card hot plug, hot swap implementations were developed as add-ons to PCI after the original bus definition. These solutions addressed pressing requirements of server and portable computer platforms:

- It is often difficult or impossible to schedule downtime on a server to replace or install peripheral cards. The ability to hot plug I/O devices minimizes downtime.
- Portable computer users need the ability to hot plug cards that provide I/O functions such as mobile disk drives and communications.

PCI Express has native support for hot plugging and hot swapping I/O peripherals. No sideband signals are required and a unified software model can be used for all PCI Express form factors.

Data Integrity and Error Handling

PCI Express supports link-level data integrity for all types of transaction- and data-link packets. Thus, it is suitable for end-to-end data integrity for high-availability applications, particularly those running on server systems. PCI Express also supports PCI error handling and has advanced error reporting and handling to help improve fault isolation and recovery solutions.

PCI Express Form Factors

A number of PCI Express form factors address the requirements of client, server, and portable computer platforms:

- Standard and low-profile cards: desktops, workstations, and servers
- Mini card: portable computers
- ExpressCard: portable computers and desktops
- Server I/O module (SIOM) that is currently being defined by PCI SIG

PCI Express Standard and Low-Profile Cards

Current PCI standard and low-profile cards are used in a variety of platforms, including servers, workstations, and desktops. PCI Express also defines standard and low-profile cards that can replace or coexist with legacy PCI cards. These cards have the same dimensions as PCI cards and are equipped with a rear bracket to accommodate external cable connections.

The differences between the PCI and PCI Express cards lie in their I/O connectors. A x1 PCI Express connector has 36 pins, compared to the 120 pins on a standard PCI

connector. Figure 6 compares PCI and PCI Express low-profile cards. The x1 PCI Express connector shown is much smaller than the connector on the PCI card. Next to the PCI Express connector is a small tab that precludes it from being inserted into a PCI slot. The standard and low-profile form factors also support x4, x8, and x16 implementations.

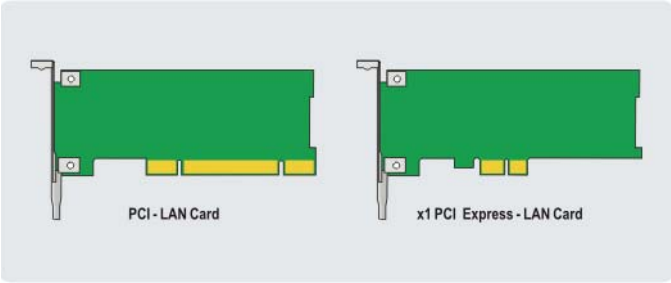


Figure 6. Comparison of PCI Express and PCI Low-Profile Cards

Figure 7 compares the size of PCI Express connectors to the PCI, AGP8X, and PCI-X connectors they will replace on the system board.

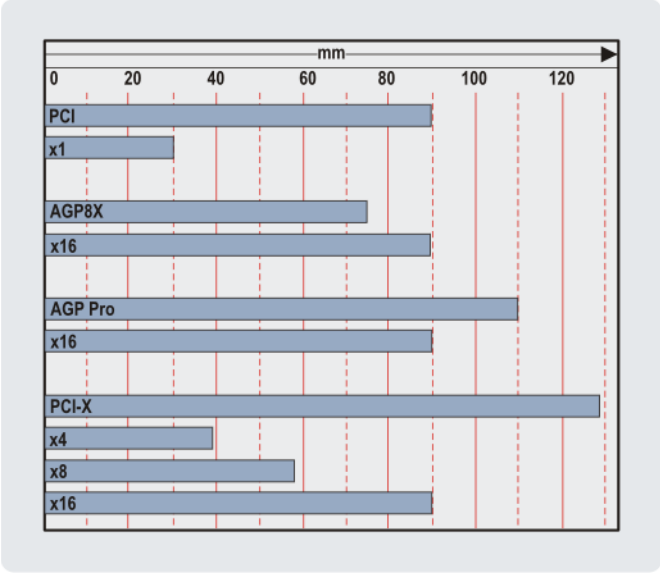


Figure 7. PCI Express System Board Connector Size for Standard and Low-Profile Cards

Table 3 shows the interoperability requirements of standard and low-profile PCI Express cards. A x1 card can be used in all four system board slots: x1, x4, x8, and x16. When a x1 card is inserted into a higher-bandwidth slot, the link layer negotiates the link down to the x1 data transfer rate.

PCI Express Implementation	x1 Slot	x4 Slot	x8 Slot	x16 Slot
x1 Card	Required	Required	Required	Required
x4 Card	No	Required	Allowed	Allowed
x8 Card	No	Allowed*	Required	Allowed
x16 Card	No	No	No	Required

*These implementations will have an x8 connector on a wired x4 slot. This means that the slot will accept x8 cards, but run at x4 speeds.

Table 3. PCI Express Card Interoperability

Transition to PCI Express Cards

Client system boards will gradually migrate from the PCI connector to the x1 PCI Express connector. Workstations will migrate from PCI to x1 PCI Express connectors, and from PCI-X to x4 PCI Express connectors. The AGP8X connector will be replaced with a x16 PCI Express connector. Unlike AGP, this connector can be used for other PCI Express cards if a PCI Express graphics card is not required.

Servers will gradually migrate from PCI-X connectors to primarily x4 and x8 connectors. Beginning in 2004, customers should expect a mix of PCI Express and PCI/PCI-X slots in server systems. This approach will allow customers to adopt new technology, while maintaining legacy support.

Figure 8 compares the I/O connectors on a typical current client system board to those on a transitional PCI Express system board. The PCI system board contains five standard PCI slots and one AGP slot. The PCI Express system board also has six I/O slots, but only three are PCI slots. Two are x1 PCI Express connectors and one is a x16 PCI Express connector that replaces the AGP8X slot. The PCI Express connectors on the system board are black to distinguish them from off-white PCI and brown AGP slots.

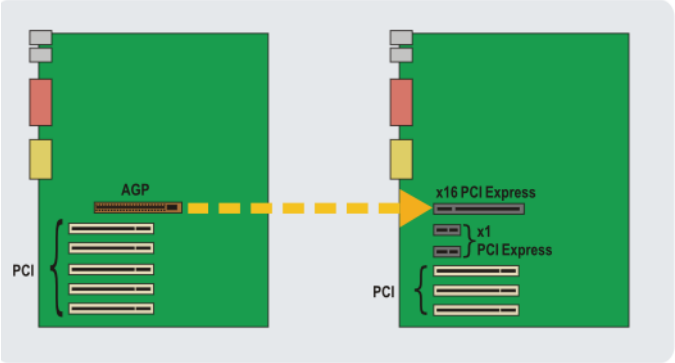
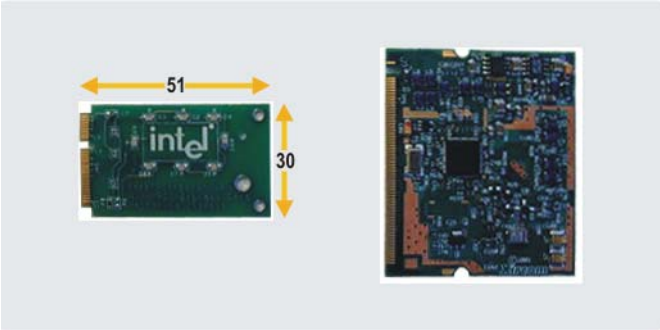


Figure 8. Comparison of PCI and Transitional PCI Express System Boards



(Source: Intel)

Figure 9. PCI Express Mini Versus Mini PCI

The first devices that will migrate to PCI Express cards will be those that require the bandwidth. For client systems, these devices will include graphics, 1394, Gigabit Ethernet, and TV tuner cards. For server systems, Ultra320 SCSI RAID cards, Fibre Channel host bus adapters (HBAs), and 1- and 10-Gigabit Ethernet cards will be available initially. The cost of these cards is expected to be comparable to (and, in some cases, lower) than PCI-X alternatives. Other cards are expected to gradually migrate to PCI Express, but it will be many years before inexpensive and low-bandwidth cards such as modems are migrated. Similar to the transition from the ISA to PCI bus, systems with both PCI and PCI Express will exist for many years.

PCI Express Mini Card

The PCI Express Mini Card replaces the Mini PCI card, which is a small internal card functionally identical to standard desktop computer PCI cards. Mini PCI cards are used mainly to add communications functions to portable computers that are built- or customized-to-order. The PCI Express Mini Card is half the size of the Mini PCI card as shown in Figure 9. This allows system designers to include one or two cards, depending on the size constraints of a particular portable computer.

A PCI Express Mini Card socket on the system board must support both a x1 PCI Express link and a USB 2.0 link. A PCI Express Mini Card can use either PCI Express or USB 2.0 (or both). USB 2.0 support will help during the transition to PCI Express, because peripheral vendors will need time to design PCI Express into their chip sets. During the transition, PCI Express Mini Cards can be quickly implemented using USB 2.0.

ExpressCard

ExpressCard is a small, modular add-in card designed to replace the PC Card over the next few years. The ExpressCard specification was developed by the Personal Computer Memory Card International Association (PCMCIA). The ExpressCard form factors shown in Figure 10 are designed to provide a small, less-expensive, and higher-bandwidth replacement for the PC Card. Like the PCI Express Mini Card, an ExpressCard module can support a x1 PCI Express and a USB 2.0 link. Its low cost also makes it feasible for small form-factor desktop systems. The ExpressCard module also has low power requirements and is hot pluggable. It is likely to be used for communications, hard-disk storage, and emerging I/O technologies. ExpressCard modules are expected in the second half of 2004.

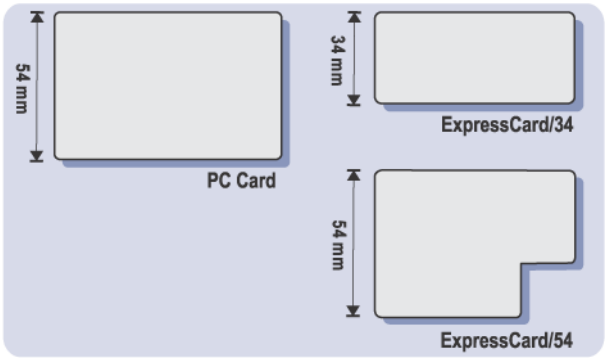


Figure 10. ExpressCard Modules

PCI Express Server I/O Module

The SIOM specification is currently being defined. SIOMs are expected with the second generation of the PCI Express technology. The PCI Express SIOM will provide a robust form factor that can be easily installed or replaced. It will be modular, allowing I/O cards to be installed and serviced in a system while it is still operating and without opening the chassis.

The SIOM is a more radical form factor change than other PCI Express form factors. It will solve many of the problems with PCI and PCI-X cards in servers. It will be hot pluggable and its cover will protect the internal components. These features are designed to make the cards more reliable in data center environments where many people handle cards.

The module is also designed with forced-air cooling in mind because high-speed server devices tend to generate a lot of heat. The cooling air can originate from the back, top, or bottom of the module. This flexibility offers system designers more options when evaluating thermal solutions for rack-mounted systems equipped with SIOMs.

The largest SIOM form factor will accommodate relatively complicated functions and should be able to leverage the full range of PCI Express links.

Sample PCI Express System Architectures

The following sections present examples of PCI Express architectures for client and server systems.

Client Systems

Figure 11 shows how PCI Express could be implemented in a client system. Initially, a x16 PCI Express link will replace the AGP bus between the graphics subsystem and the Northbridge. A PCI Express variant could also replace the link between the Northbridge and Southbridge, relieving the bottleneck between peripheral I/O devices and the Northbridge. There will also be multiple PCI Express links off the Southbridge for the network interface controller (NIC), 1394 devices, and other peripherals. The Southbridge will continue to support legacy PCI slots.

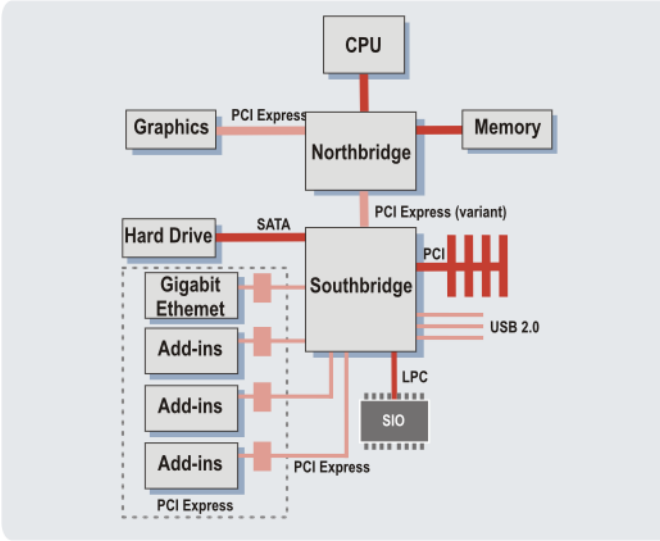


Figure 11. Sample Desktop Architecture

This architecture highlights several key implications for customers. Desktop systems will have both PCI and PCI Express buses for a long time. To minimize confusion during the transition, PCI cards cannot be inadvertently inserted into PCI Express slots, nor can PCI Express cards be inserted into legacy PCI slots. In addition, PCI Express enables widespread adoption of Gigabit[4] Ethernet, 10-Gigabit Ethernet, 1394b, or other high speed devices in client systems. It will also support in-

4. This term does not connote an actual operating speed of 1 Gbps. For high-speed transmission, connection to a Gigabit Ethernet server and network infrastructure is required.

creasing bandwidth requirements of graphics subsystems.

Portable Computers

Figure 12 shows how PCI Express could be implemented in a portable computer system. Like desktop systems, PCI Express will replace the AGP bus, and a PCI Express variant is a candidate to replace the link between the Northbridge and Southbridge. In addition, PCI Express could be used to replace the PCI bus between the Northbridge and the build-to-order/customize-to-order (BTO/CTO) slot. This slot currently accommodates Mini PCI cards, but in new systems it may be used for PCI Express Mini Cards.

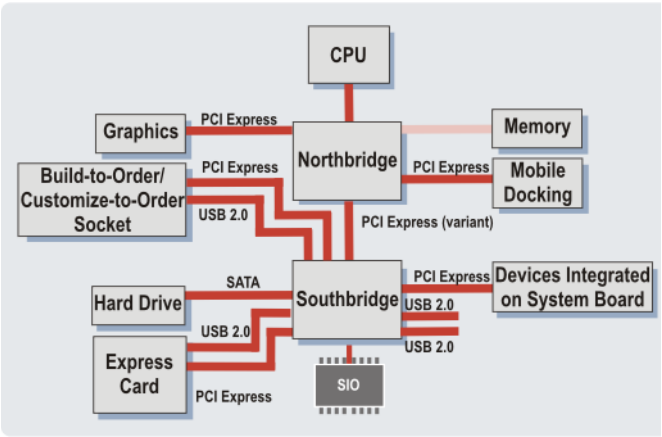


Figure 12. Sample Portable Computer Architecture

The PCI bus between the Northbridge and the docking station could also migrate to PCI Express. A x1 ExpressCard slot that uses a USB 2.0 link will replace the PC Card slot. Finally, individual PCI Express links will replace the PCI bus that supports integrated peripheral devices such as Gigabit Ethernet, audio, and graphics.

Server Systems

Figure 13 shows how PCI Express could be implemented in a dual-processor server architecture. PCI Express can help to significantly reduce server system complexity. PCI Express links for I/O devices and slots are placed directly off the Northbridge. This approach is expected to provide the following potential advantages:

- **Higher bandwidth for next-generation I/O such as 10-Gbps Ethernet and x4 Infiniband fabrics.** For example, a x8 PCI Express link can accommodate the peak bandwidth required for a dual-ported 10-Gbps controller.
- **Lower implementation cost.** More slots and embedded I/O devices can be connected to the system chip set with fewer bridge chips and fewer signal routing requirements on the system board.
- **Lower latency.** Transmission latency between I/O devices and the CPU and memory can be reduced by eliminating the PCI-X bridge chip.

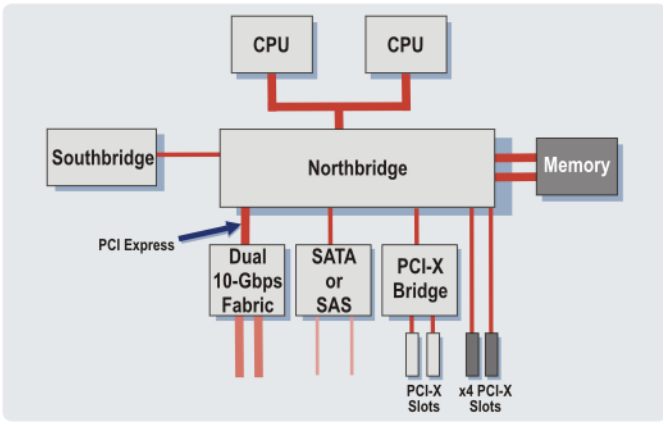


Figure 13. Sample Server Architecture

Initial generations of PCI Express servers will also include PCI-X slots for legacy PCI-X cards.

Enabling Future Modular Designs

The PCI SIG is also working on the PCI Express cable specification. Because PCI Express has high data rates and low-pin-count connectors, it is likely to be used as a high-speed interconnect between components in client and server systems. Modular systems with separate high-speed components can be connected with PCI Express cables. Figure 14 illustrates the concept of a "split" system that separates components that generate heat such as processors, memory, and graphics from other components such as removable storage, display devices, and I/O ports. It may also make sense to separate high-end graphics subsystems, which require more power and generate heat, from the main processor chassis. This approach would make it easier to deliver appropriate power and cooling to the graphics subsystem.

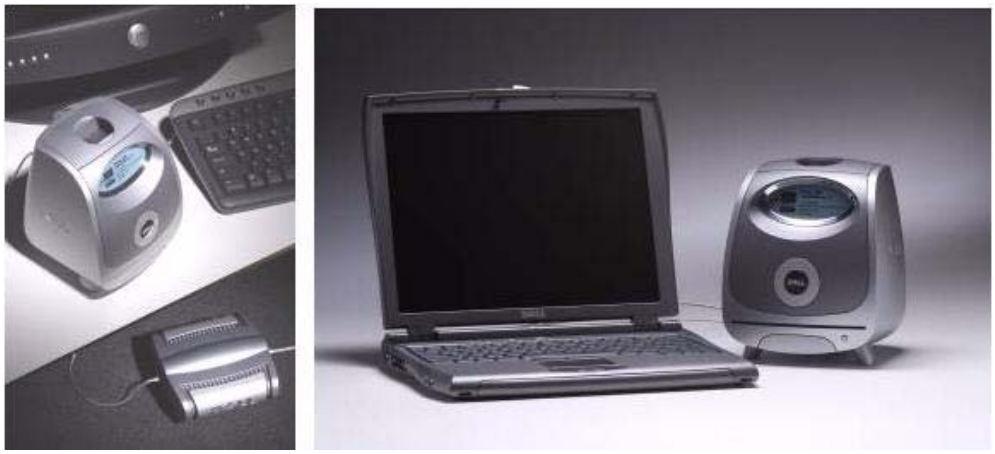


Figure 14. Examples of Split Systems That Separate Processor From I/O

Conclusion

PCI Express is the open standards-based successor to PCI. It is designed to provide a reliable and scalable high-speed serial interconnect that maintains backward compatibility with PCI. Like PCI, it will be implemented in a broad range of existing platforms, including servers, portable computers, desktop systems, and workstations. It will also enable innovative modular computer system designs.

Dell has been a strong supporter of PCI Express technology, participating fully in the development of the specification and planning for its inclusion in Dell™ products. Dell will begin transitioning Dell platforms to PCI Express in 2004 when chip set support and PCI Express devices are introduced.

For More Information

- Intel white paper: "Creating a Third Generation I/O Interconnect," **www.intel.com/technology/pciexpress/devnet/docs/WhatisPCIExpress.pdf**
- PCI SIG: **www.pcisig.com**