How PCI Express Works (by Tracy V. Wilson)

http://computer.howstuffworks.com/pci-express.htm

Peripheral Component Interconnect (PCI) slots are such an integral part of a computer's architecture that most people take them for granted. For years, PCI has been a versatile, functional way to connect sound, video and network cards to a motherboard.

But PCI has some shortcomings. As processors, video cards, sound cards and networks have gotten faster and more powerful, PCI has stayed the same. It has a fixed width of 32 bits and can handle only 5 devices at a time. The newer, 64-bit PCI-X bus provides more bandwidth, but its greater width compounds some of PCI's other issues.

A new protocol called PCI Express (PCIe) eliminates a lot of these shortcomings, provides more bandwidth and is compatible with existing operating systems. In this article, we'll examine what makes PCIe different from PCI. We'll also look at how PCI Express makes a computer faster, can potentially add graphics performance, and can replace the AGP slot.

High-Speed Serial Connection

In the early days of computing, a vast amount of data moved over serial connections. Computers separated data into packets and then moved the packets from one place to another one at a time. Serial connections were reliable but slow, so manufacturers began using parallel connections to send multiple pieces of data simultaneously.

It turns out that parallel connections have their own problems as speeds get higher and higher -- for example, wires can interfere with each other electromagnetically -- so now the pendulum is swinging back toward highly-optimized serial connections. Improvements to hardware and to the process of dividing, labeling and reassembling packets have led to much faster serial connections, such as USB 2.0 and FireWire.

PCI Express is a serial connection that operates more like a network than a bus. Instead of one bus that handles data from multiple sources, PCIe has a switch that controls several point-to-point serial connections. (See How LAN Switches Work for details.) These connections fan out from the switch, leading directly to the devices where the data needs to go. Every device has its own dedicated connection, so devices no longer share bandwidth like they do on a normal bus.

PCI Express Lanes

When the computer starts up, PCIe determines which devices are plugged into the motherboard. It then identifies the links between the devices, creating a map of where traffic will go and negotiating the width of each link. This identification of devices and connections is the same protocol PCI uses, so PCIe does not require any changes to software or operating systems.
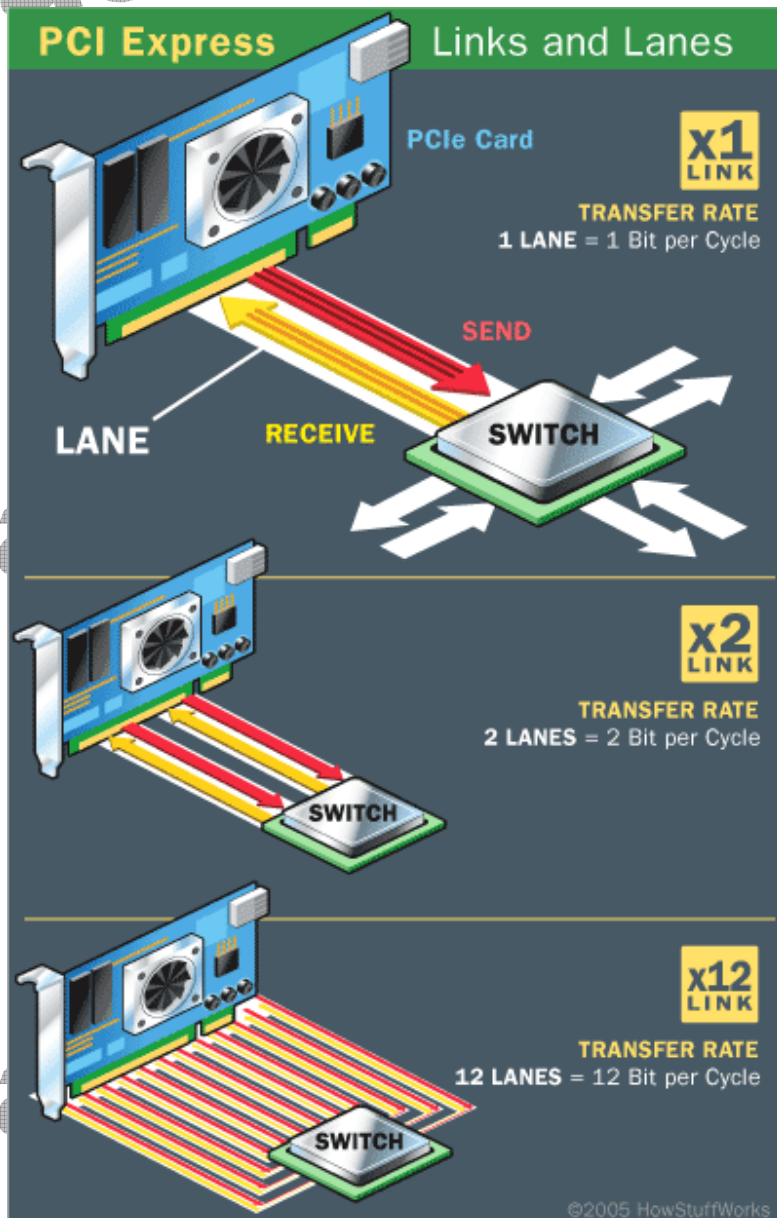
Each lane of a PCI Express connection contains two pairs of wires -- one to send and one to receive. Packets of data move across the lane at a rate of one bit per cycle. A x1 connection, the smallest PCIe connection, has one lane made up of four wires. It carries one bit per cycle in each direction. A x2 link contains eight wires and transmits two bits at once, a x4 link transmits four bits, and so on. Other configurations are x12, x16 and x32.


Scalable PCI Express slots.

PCI Express is available for desktop and laptop PCs. Its use may lead to lower cost of motherboard production, since its connections contain fewer pins than PCI connections do. It also has the potential to support many devices, including Ethernet cards, USB 2 and video cards.

PCI Express Connection Speeds

The 32-bit PCI bus has a maximum speed of 33 MHz, which allows a maximum of 133 MB of data to pass through the bus per second. The 64-bit PCI-X bus has twice the bus width of PCI. Different PCI-X specifications allow different rates of data transfer, anywhere from 512 MB to 1 GB of data per second.
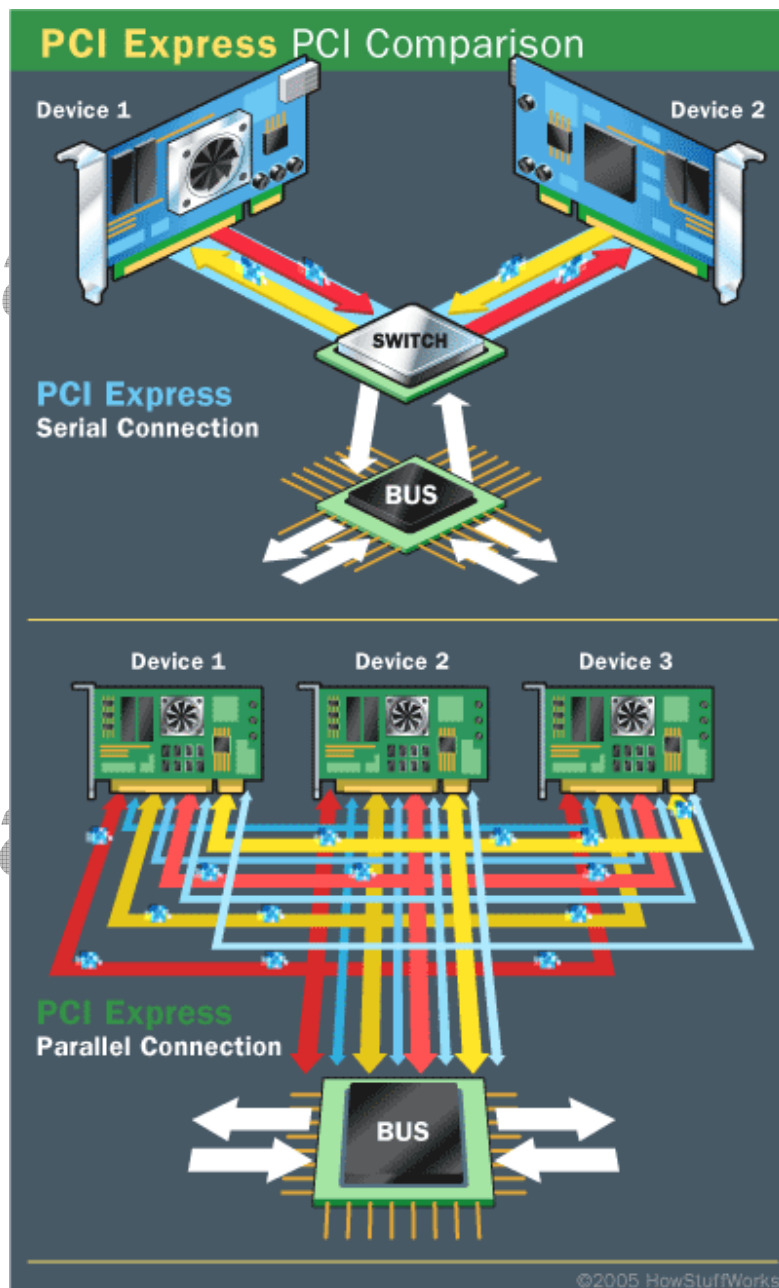
Devices using PCI share a common bus, but each device using PCI Express has its own dedicated connection to the switch.

A single PCI Express lane, however, can handle 200 MB of traffic in each direction per second. A x16 PCIe connector can move an amazing 6.4 GB of data per second in each direction. At these speeds, a x1 connection can easily handle a gigabit Ethernet connection as well as audio and storage applications. A x16 connection can easily handle powerful graphics adapters.

How is this possible? A few simple advances have contributed to this massive jump in serial connection speed:

- Prioritization of data, which allows the system to move the most important data first and helps prevent bottlenecks
- Time-dependent (real-time) data transfers
- Improvements in the physical materials used to make the connections
- Better handshaking and error detection
- Better methods for breaking data into packets and putting the packets together again. Also, since each device has its own dedicated, point-to-point connection to the switch, signals from multiple sources no longer have to work their way through the same bus.

PCI Express and Advanced Graphics

We've established that PCIe can eliminate the need for an [AGP](#) connection. A x16 PCIe slot can accommodate far more data per second than current AGP 8x connections allow. In addition, a x16 PCIe slot can supply 75 watts of power to the video card, as opposed to the 25watt/42 watt AGP 8x connection. But PCIe has even more impressive potential in store for the future of graphics technology.

With the right hardware, a motherboard with two x16 PCIe connections can support two graphics adapters at the same time. Several manufacturers are developing and releasing systems to take advantage of this feature:

- NVIDIA Scalable Link Interface (SLI): With an SLI-certified motherboard, two SLI graphics cards and an SLI connector, a user can put two video cards into the same system. The cards work together by splitting the screen in half. Each card controls half of the screen, and the connector makes sure that everything stays synchronized.
- ATI CrossFire: Two ATI Radeon® video cards, one with a "compositing engine" chip, plug into a compatible motherboard. ATI's technology focuses on image quality and does not require identical video cards, although high-performance systems must have identical cards. Crossfire divides up the work of rendering in one of three ways:
  - splitting the screen in half and assigning one half to each card (called "scissoring")
  - dividing up the screen into tiles (like a checkerboard) and having one card render the "white" tiles and the other render the "black" tiles
  - having each card render alternate frames
- Alienware Video Array: Two off-the-shelf video cards combine with a Video Merger Hub and proprietary software. This system will use specialized cooling and power systems to handle all the extra heat and energy from the video cards. Alienware's technology may eventually support as many as four video cards.

Since PCI, PCI-X and PCI Express are all compatible, all three can coexist indefinitely. So far, video cards have made the fastest transition to the PCIe format. Network and sound adapters, as well as other peripherals, have been slower in development. But since PCIe is compatible with current operating systems and can provide faster speeds, it is likely that it will eventually replace PCI as a PC standard. Gradually, PCI-based cards will become obsolete.