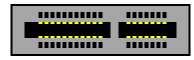
EXPANSION CARDS - III



In the first two installments of this series of Tech Tips, we took a look at PCI and AGP, undoubtedly the most common expansion slots in a computer today. With a few key improvements over both of these, PCI Express is destined to replace both and offer a whole new level of computer performance.

As with AGP and PCI, the development of PCI Express can be attributed to Intel. This time, however, they partnered with some other heavy hitters in the industry, such as Microsoft, IBM, and Dell. Although it is now known as PCI Express, that was not their initial choice for its name. If it wasn't for PCI-SIG, the committee that oversees the PCI standard, we might be referring to this new format at 3GIO (Third GenerationInput / Output).

PCI Express development finds its roots in the PCI and AGP standards, but the physical connections are not interchangeable, and we will see that this is not the only difference. In the PCI standard, data from the various devices travels over a shared bus to the system. In the AGP standard, a dedicated, point-to-point interface transmits the data from the graphics card to the system. The PCI Express approach to data transfer involves a collection of two-way, serial connections that carries data in packets, similar to the way a network connection operates.

The data from a PCI Express device will no longer have to travel over a single bus, or a single dedicated connection, but can use a combination of these two-way serial connections to optimize throughput. The terms "lane" and "link" don't sound like anything overly technical, but take on special meaning with PCI Express. A link is the physical connection between PCI Express devices, which can consist of multiple lanes that transmit and receive data independently. Links can be composed of 1, 2, 4, 8, 12, 16, or 32 lanes, and the configuration allows flexibility in assigning just as many lanes as needed to any particular device. There are obvious benefits to this approach, and a few of the more significant include the following points…

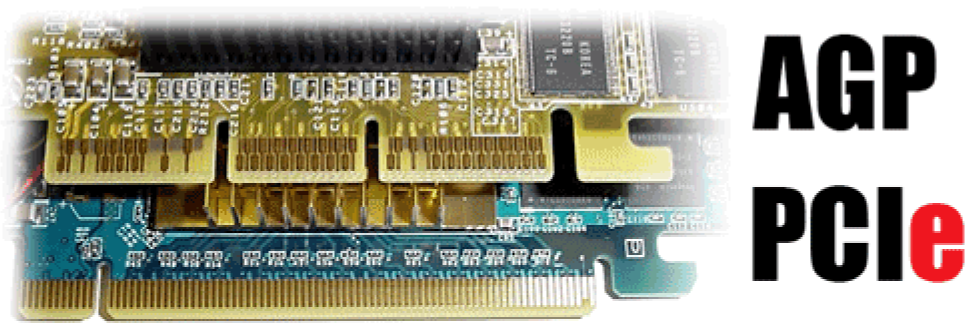
Each lane of PCI Express communication is dedicated between two points, so there is no sharing of bandwidth. PCI's main bottleneck was that all the devices were sharing the equivalent of one lane, and all of the available bandwidth also had to be shared.

Multiple lanes can be assigned to devices whose performance would benefit from the extra speed and bandwidth. A PCI Express graphics card might be assigned 16 lanes (also referred to as x16), while a network adaptor might be assigned just 1 lane. Each lane you make available to a device increases the potential for performance, as the data is sequenced up/down each available lane to optimize throughput. This process of sending the next byte of data down the next available lane is referred to as data striping, and obviously more lanes are better for instances where a good deal of data needs to be transmitted quickly.

Speaking of graphics cards, another benefit is that multiple high performance graphics cards can be installed on one motherboard. The flexibility of PCI Express allows for two x16 PCI Express slots to be included for dual graphics cards, something that in the past required one AGP slot and one PCI slot. And due to the performance limitations, the AGP and PCI combination could not really be considered high performance. In addition to two x16 slots allowing for dual display operation, when incorporating specific graphics cards on a motherboard supporting nVidia's SLi technology, the resources of the two separate cards can be bridged together for even greater performance on one display. An example of such a motherboard can be seen in DFI's LAN Party UT nF4 SLi-D.

Just as motherboards supported both AGP and PCI as a means of allowing dual displays, some motherboards offer both an AGP slot and a PCI Express slot. Not only does this allow the user the ability to run dual displays, it provides the added benefit of allowing an upgrade to be completed in stages. If a new PCI Express capable motherboard was just purchased, perhaps in addition to a new processor, the budget conscience user may not want to spring for a new graphics card right away. By making an AGP slot available on boards such as the ECS 915P-A, there is no reason to retire a perfectly good AGP card just because one bought a new motherboard supporting PCI Express.

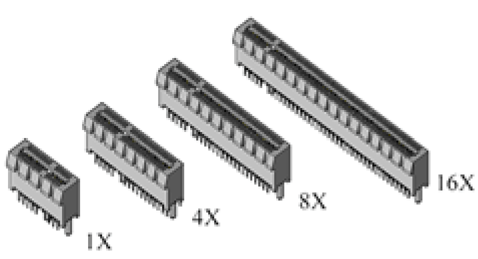
PCI Express graphics cards are quite similar to AGP cards, except for the connector configuration. The physical size and layout are comparable, and even the prices are not that different. The current selection of graphics cards at Geeks.com doesn't allow you to compare apples to apples in any one card, but one may find many of the same AGP cards available in PCI Express format for roughly the same price (or for even less money). For the time being, the markets seem to be running in parallel, but in time a shift will occur in favor of PCI Express dominating the market.



Minimizing the cost involved in motherboard fabrication could be another benefit. Let's look at the example of a network adaptor requiring just 1 lane to operate. If this was a PCI based network adaptor, traces for the standard 32-bit bus would need to reach this device, instead of the four traces required for 1 PCI Express lane. Motherboard design will obviously weigh heavily on this benefit ever being realized, and it is possible that higher-end boards might actually require more traces.

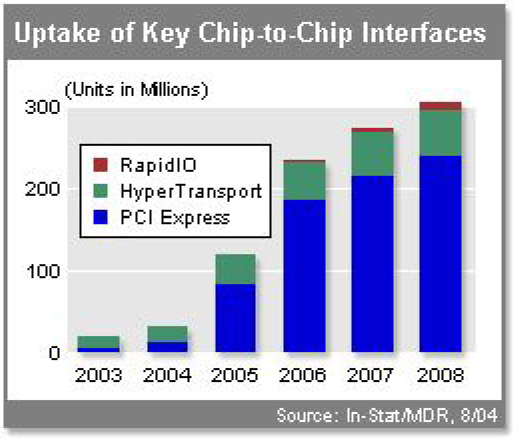
Before taking a look at the ultimate benefit of PCI Express, the performance, let's have a refresher on the capabilities of PCI and AGP. The standard PCI bus has a width of 32-bit, operates at 33 MHz, and provides a maximum bandwidth of 132 MB/s (which has to be shared by all devices connected). AGP 8x has a 32-bit bus width, operates at 533 MHz, and provides a maximum (dedicated) bandwidth of 2.1 GB/s.

Each PCI Express lane is capable of 250 MB/s in each direction, and as advances in the necessary silicon technologies are realized, that number can be expected to quadruple. Presently, a 164-pin x16 slot can be expected to provide around 4GB/s of usable bandwidth in either direction, which is almost double the 2.1GB/s bandwidth that AGP 8x could offer! Definitely an impressive increase, and as the technology is refined, it will be very interesting to see the performance scale up.



In the previous paragraph, I mentioned that the x16 slot features 164 pins. Each of the different lane configurations is accompanied by a different physical connector, and a sampling of an x16, x8, x4, and x1 can be seen here. For a real world example, the Chaintech VNF4 Ultra Athlon 64 Socket 939 motherboard shows an actual installation of one x16 slot and two x1 slots.

Graphics cards are obviously going to benefit the most from the power and performance available with PCI Express, but as mentioned, other devices will also be able to take advantage of this new standard. The example of a network adaptor is just one that not only can use PCI Express, but will also see performance benefits. A Gigabit Ethernet adaptor will be more likely to actually achieve its rated speed thanks to the main bottleneck being removed in the form of the slower, narrower PCI Bus. Other bandwidth intensive devices, such as RAID controllers, can also be expected to jump off of the slower PCI Bus and find a smoother ride on PCI Express. Although PCI devices requiring less bandwidth may not see any performance benefits from going to PCI Express, as the standard achieves greater mainstream acceptance, the cost implications may find these devices shifting over anyway, just as happened with the transition from ISA to PCI.



Source : http://www.geeks.com/techtips/2005/techtips-031005.htm